TEMA

TRUSTED
EXTREMELY PRECISE
MAPPING AND PREDICTION
FOR EMERGENCY
MANAGEMENT

D3.4

Report on AI model adaptability to extreme data

Funded by
the European Union

TEMA
TRUSTED
EXTREMELY PRECISE
MAPPING AND PREDICTION
FOR EMERGENCY
MANAGEMENT

Project Information

Project acronym:	**TEMA**
Project full title:	Trusted Extremely Precise Mapping and Prediction for Emergency Management
Call identifier:	HORIZON-CL4-2022-DATA-01
Type of action:	HORIZON Research and Innovation Actions
Start date:	1 December 2022
End date:	30 November 2026
Grant agreement no:	101093003

D3.4- Report on AI model adaptability to extreme data			
Executive Summary:	**Deliverable D3.4 Report on AI model adaptability to extreme data, is the fourth deliverable of Work Package 3 (WP3) within the TEMA project. This document encapsulates the research results of Task T3.5 over the months M13-M36 of the project. See the Section Executive Summary.**		
WP:	**3**		
Author(s):	**See table below for a full list of authors**		
Editor:	**Eleonor Diaz (ATOS)**		
Leading Partner:	**ATOS**		
Participating Partners:	**ATOS, AUTH, FHHI, IT:U**		
Version:	**1.0**	Status:	**Final**
Deliverable Type:	**R Document, report**	Dissemination Level:	**Public**
Official Submission Date:	**30 Nov 2025**	Actual Submission Date:	**30 Nov 2025**

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Disclaimer

The TEMA consortium consists of the following partners:

No.	Partner Organization Name	Partner Organization Short Name	Country
1	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTH	GR
2	DEUTSCHES ZENTRUM FUR LUFT UND RAUMFAHRT EV	DLR	DE
3	ENGINEERING - INGEGNERIA INFORMATICA SPA	ENG	IT
4	ATOS IT SOLUTIONS AND SERVICES IBERIA SL	ATOS IT	ES
4.1	ATOS SPAIN SA	ATOS SP	ES
5	UNIVERSIDAD DE SEVILLA	USE	ES
6	TECNOSYLVA SL	TSYL	ES
7	NORTHDOCKS GMBH	ND	DE
9	THE LISBON COUNCIL FOR ECONOMIC COMPETITIVENESS ASBL	LC	BE
10	LATITUDO 40 SRL	LAT40	IT
11	NELEN & SCHUURMANS TECHNOLOGY BV	NS	NL
11.1	NELEN & SCHUURMANS CONSULTANCY BV	NS C	NL
12	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV	FHHI	DE
13	UNIVERSITA DEGLI STUDI DI MESSINA	UNIME	IT
14	KAJAANIN AMMATTIKORKEAKOULU OY	KAMK	FI
16	KENTRO MELETON ASFALEIAS	KEMEA	GR
17	DIMOS MANTOUDIOU - LIMNIS - AGIAS ANNAS	D.MALIAN	GR
18	REGIONE AUTONOMA DELLA SARDEGNA	RAS	IT

This project has received funding from the European Union – European Commission under grant agreement No 101093003

19	BAYERISCHES ROTES KREUZ	BRK	DE
20	KAINUUN HYVINVOINTIALUE	KAHY	FI
21	INTERDISCIPLINARY TRANSFORMATION UNIVERSITY	IT:U	AT

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Document Revision History

Version	Description	Contributions
0.1	ToC	Eleonor Diaz
0.2	First draft	Eleonor Diaz, David Hanny, Ehsaneddin Jalilian, Shaily Gandhi, Bernd Resch, Sebastian Schmidt, Dorian Arifi, Burcu Bilgic, Marco-Constantin Badici, Afzal Ahmad, Valsamara, Leila Arras
1.0	Final version ready for submission.	-, -

Authors

Name	Partner
Eleonor Diaz	ATOS
David Hanny	IT:U
Ehsaneddin Jalilian	IT:U
Shaily Gandhi	IT:U
Bernd Resch	IT:U
Sebastian Schmidt	IT:U
Dorian Arifi	IT:U
Burcu Bilgic	IT:U
Marco-Constantin Badici	IT:U
Afzal Ahmad	IT:U
Ioanna Valsamara	AUTH
Ioannis Pitas	AUTH
Vasileios Mygdalis	AUTH
Filippos Kitsos	AUTH
Michael Siavrakas	AUTH
Dimitrios Papaioannou	AUTH
Eugenios Vlachos	AUTH
Evangelos Charalampakis	AUTH
Dimitrios Fotiou	AUTH
Apostolis Apostolidis	AUTH
Evangelos Spatharis	AUTH
Pantelis Mentesidis	AUTH
Leila Arras	FHHI

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Reviewers

Name	Partner
Bernd Resch	IT:U
David Hanny	IT:U

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table of Contents

This project has received funding from the European Union – European Commission under grant agreement No 101093003

This project has received funding from the European Union – European Commission under grant agreement No 101093003

List of Figures

This project has received funding from the European Union – European Commission under grant agreement No 101093003

This project has received funding from the European Union – European Commission under grant agreement No 101093003

This project has received funding from the European Union – European Commission under grant agreement No 101093003

List of Tables

This project has received funding from the European Union – European Commission under grant agreement No 101093003

List of Terms and Abbreviations

Abbreviation	Meaning
AA	Authorship Attribution
AI	Artificial Intelligence
AKSA	Agent Knowledge Self-Assessment
API	Application Programming Interface
AUW	Artificial UnrealFire Wildfire Dataset
BERT	Bidirectional Encoder Representations from Transformers
BLAZE	Blaze Wildfire Dataset
BiseNet	Bilateral Segmentation Network
C-LENC	Cloud Learning-by-Education Node Community
CIFAR-10	Canadian Institute For Advanced Research Dataset (10 classes)
CIFAR100	Canadian Institute For Advanced Research Dataset (100 classes)
CLIP	Contrastive Language–Image Pre-training
CNN	Convolutional Neural Network
CNN-I2I	Convolutional Neural Network Image-to-Image
CovGI	Covariance-based Gradient Œ Input
CovLRP	Covariance-based Layer-wise Relevance Propagation
CVPR	Computer Vision and Pattern Recognition
DAAM	Diffusion Attentive Attribution Maps
DNN	Deep Neural Network
DualXDA	Dual eXplainable Data Attribution
ECCV	European Conference on Computer Vision
EUSIPCO	European Signal Processing Conference
ExCEL	Extreme Class Embedding Learning
EWS	Extreme Weakly Supervised
EWS-M	Extreme Weakly Supervised Multiclass
FCL-ViT	Feedback Continual Learning Vision Transformer
FCMA	Fire Classification Multi-Agent
FID	Fréchet Inception Distance
FL	Federated Learning
FLAME	Fire Luminosity and Aerial Monitoring Evaluation Dataset
Flux / Flux1.dev	Flux Diffusion Model (Version 1.dev)
GAN	Generative Adversarial Network
GI	Gradient Œ Input
GPT	Generative Pre-trained Transformer
Grounding DINO	Grounding Detection with DINO

This project has received funding from the European Union – European Commission under grant agreement No 101093003

ICCV	International Conference on Computer Vision
IEEE	Institute of Electrical and Electronics Engineers
IoU	Intersection over Union
ISCC	IEEE Symposium on Computers and Communications
KD	Knowledge Distillation
KPIs	Key Performance Indicators
LENC	Learning-by-Education Node Community
LLM	Large Language Model
LRP	Layer-wise Relevance Propagation
mAP	mean Average Precision
Mistral7B	Mistral 7-Billion Parameter Model
MM-Grounding-DINO	Multi-Modal Grounding DINO
Molmo	Molmo Vision-Language Model
mIoU	mean Intersection over Union
NAS	Neural Architecture Search
NASKD	Neural Architecture Search and Knowledge Distillation
NDM	Natural Disaster Management
OSM	OpenStreetMap
OVD	Open-Vocabulary Detection
OVS	Open-Vocabulary Segmentation
PIDNet	Pixel-level Interaction-based Dual Branch Network
Proto-SVDD	Prototype Support Vector Data Description
RoBERTa	Robustly Optimised BERT Pre-training Approach
RGB	RedGreenBlue
SAM-HQ	Segment Anything Model High Quality
SDXL	Stable Diffusion XL
SEEM	Segment Everything Everywhere, All at Once
SVDD	Support Vector Data Description
SVM	Support Vector Machine
SAM-HQ	Segment Anything Model – High Quality
TAB	Tunable self-Attention Block
TCP	Transmission Control Protocol
TSB	Task-Specific Block
TPAMI	IEEE Transactions on Pattern Analysis and Machine Intelligence
UaV	Unmanned Aerial Vehicle
UI	User Interface
UNet++	Nested U-Net
Unreal Engine 5	Unreal Engine (Version 5)
UnrealFire	UnrealFire Synthetic Data Pipeline

This project has received funding from the European Union – European Commission under grant agreement No 101093003

UMAP	Uniform Manifold Approximation and Projection
VAE	Variational Autoencoder
ViT	Vision Transformer
VisDrone-DET2019	Visual Object Detection Dataset for Drones (2019)
VLM	Vision-Language Model
XAI	Explainable Artificial Intelligence
YOLOv6	You Only Look Once Version 6
YOLOv8	You Only Look Once Version 8

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Executive Summary

Deliverable D3.4 "Final report on algorithms for extreme data analytics" is the fourth deliverable of Work Package 3 (WP3) within the TEMA project. This document encapsulates the research results of Task T3.5, "Study AI model adaptability to extreme data conditions," conducted between months M13 and M36. The primary focus of this task is to address the challenges faced in Natural Disaster Management (NDM) by exploring the adaptability of AI models under extreme data conditions, while also linking to previous research conducted in Tasks T3.1, T3.2, and T3.3.

Significant advancements have been made in several key areas:

- **Synthetic Data Generation:** The TEMA project focused on generating high-quality synthetic images to augment datasets for various tasks, including wildfire and flood detection. This was achieved through the use of advanced diffusion models, which iteratively refine noise into coherent images, greatly enhancing the representativeness of limited datasets.These advancements directly support OA2 increasing the accuracy of extreme data analysis algorithms. By enriching training datasets with realistic synthetic imagery, TEMA strengthens the robustness and precision of detection and segmentation models.
- **Zero-shot Learning:** The integration of zero-shot models for automatic dataset labelling has drastically reduced manual annotation requirements while improving the accuracy and reliability of disaster analysis. This approach facilitates scalable and efficient labelling processes, particularly important in the context of rapidly evolving disaster scenarios.These advancements directly contribute to OA2 increasing the accuracy of extreme data analysis algorithms. By improving the precision and consistency of labelled datasets.
- **Graph-based Frameworks:** A novel graph-based learning framework was developed to unify heterogeneous social media data, improving real-time disaster analysis. This framework allows for better situational awareness by integrating data from various platforms, such as Twitter and TikTok, to capture the dynamic nature of disasters. These developments directly support OA2 increasing the accuracy of extreme data analysis algorithms. Through the structured representation and fusion of multi-platform data, the graph-based approach significantly improves the precision and contextual relevance of semantic analysis.
- **Explainable AI Methods:** The development of explainable AI (XAI) methods tailored for extreme data conditions has enhanced model interpretability and trustworthiness. These methods include frameworks that guide users in understanding AI decisions, thereby improving reliability in critical scenarios. This work directly contributes to OA1 increasing the trustworthiness of extreme data analysis algorithms. By delivering advanced, trustworthy AI methods capable of providing rapid and accurate explanations not only for model predictions but also for the input modalities that most influenced the outcomes.
- **Continuous Learning Frameworks:** A feedback-based continuous learning system was introduced to enable Vision Transformers to adapt to evolving disaster data. This system ensures that models retain knowledge from past data while adapting to new incoming information without performance degradation. These advancements directly contribute to Objectives OA2 and OA3 increasing both the accuracy and responsiveness of extreme data analysis algorithms. By maintaining high model precision through adaptive learning and enabling faster adjustment to changing conditions, the TEMA framework significantly enhances analytical performance.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Overall, the deliverable meets or exceeds the key performance indicators (KPIs) and objectives set for WP3. The results of the research have led to the publication of nine peer-reviewed articles, and generation of five datasets, contributing to the scientific discourse on the development of trustworthy, adaptive, and efficient AI systems for extreme data analytics in NDM. As a public deliverable, D3.4 not only documents significant research outputs but also plays a vital role in disseminating TEMA's achievements in the domain of AI adaptability to extreme data conditions.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

1. Introduction

1.1. Purpose and Scope of the Document

Deliverable D3.4 “Final report on algorithms for extreme data analytics” is the fourth Deliverable of the third Work Package (WP3) of the TEMA project. The main purpose of this document is to report the research results of Task T3.5 “Study AI model adaptability to extreme data conditions” between M13-M36 and their links to the work done on Tasks T3.1 “Explainable and robust analytics“, T3.2 ”Real-time semantic visual analysis and remote sensing”, and T3.3 ”Social media and text semantic analysis”.
The TEMA research efforts in the time from M13 to M36 were focused on the following areas in order to improve AI models adaptability and reliability under scarce data scenarios:

- Synthetic image generation for NDM Dataset creation and augmentation.
- Automatic Dataset labelling with Zero-Shot Models.
- Graph-based framework to integrate multi-platform social media data for better real-time disaster analysis.
- Ground Truth generation on synthetic images through Explainable AI.
- Explainable AI methods to improve model reliability and interpretability under extreme data conditions.
- Federated learning and knowledge distillation to enable adaptive, collaborative deep learning for real-time disaster management.
- Feedback-based continuous learning framework that enables Vision Transformers to adapt to evolving disaster data without forgetting past knowledge.
- Automated Neural Architecture Search and Knowledge Distillation pipeline that builds lightweight, high-accuracy models for wildfire segmentation under limited data conditions.
- Weakly supervised multi-class segmentation framework that enables accurate disaster scene analysis using minimal annotations and limited data.

1.2. Structure of the Document

Each subsection of this document outlines the research progress achieved within the TEMA project between months 13 and 36 (M13M36) for the respective research tasks. Each subsection 1.) briefly describes the state of the artboth internationally and in relation to TEMAs own research developmentsand 2.) summarises the research progress made in the TEMA project during the second reporting period (M13M36).
The remainder of this document is structured as follows: Section 2 summarises the main research efforts and key outputs in relation to the TEMA objectives. Section 3 presents novel methods addressing data scarcity in NDM. Section 4 describes the development of explainable AI methods for extreme disaster data. Section 5 outlines several advanced learning paradigms designed to enhance the efficiency of AI systems under extreme data conditions. Finally, conclusions are presented in Section 6.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

2. Summary of the work carried out

2.1. Facing Data Scarcity on Natural Disaster Management

Diffusion models have emerged as a key solution to data scarcity in natural disaster scenarios, enabling the generation of diverse, high-quality synthetic imagery for dataset augmentation and scenario simulation. By iteratively refining noise into coherent images, they expand the representativeness of limited datasets. Integrated within frameworks such as ComfyUI, these models support flexible pipeline design, parameter optimization, and efficient management of large-scale image generation processes.

State-of-the-art diffusion models, including Stable Diffusion XL and Flux1.dev, demonstrate strong performance in producing high-fidelity, contextually rich imagery even from sparse data. Large and vision-language models such as Mistral7b and Molmo further enhance this process by generating descriptive prompts that guide diffusion-based synthesis, ensuring alignment with real-world disaster conditions. ATOS leverages both text-to-image and image-to-image pipelines: the former generates extensive synthetic datasets for tasks like flood or wildfire detection, while the latter modifies existing data to simulate varied conditionsfires, floods, or trapped individualswhile preserving contextual metadata. Through iterative diffusion, inpainting, and upscaling, these pipelines yield diverse, high-resolution datasets that strengthen model robustness and reduce the need for risky data collection.

To overcome wildfire segmentation data scarcity, AUTH developed UnrealFire, a synthetic image generation and automatic annotation pipeline built on Unreal Engine 5 and AirSim. Real datasets such as FLAME and Corsican are constrained by safety and environmental variability; UnrealFire fills this gap by generating photo-realistic UAV-perspective wildfire imagery with pixel-accurate segmentation masks under diverse conditions. Its key innovation the particle segmentation camera pluginenables accurate 2D projection of dynamic fire particles in segmentation maps, a feature absent from prior tools. Using UnrealFire, AUTH produced the AUW dataset of 1,700 RGB-mask pairs, achieving a 91.98% mIoU when combined with only 10% of real data, outperforming Corsican-trained baselines. Style transfer experiments further improved cross-dataset generalization by up to +19.82% mean Intersection over Union (mIoU) through rapid synthetic data generation and retraining cycles.

The work done by ATOS complements this efforts by labelling data using zero-shot and open-vocabulary models for automatic annotation. Models such as Grounding DINO, MM-Grounding-DINO, SAM-HQ, and SEEM provide robust, reproducible baselines for scalable, automated labelling of fire and flood imagery. For fire detection, ATOS employs a hybrid approach combining proprietary YOLOv8 models with open-vocabulary detectors, fusing outputs through Weighted Boxes Fusion for consistent annotations. For flood segmentation, multi-model fusion generates precise masks for water, mud, and river regions, mitigating model-specific weaknesses. This integration of zero-shot labelling drastically reduces manual annotation needs while improving dataset accuracy and consistency, thereby enhancing predictive model performance and the scal-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

ability of disaster-response systems.

In parallel, IT:U addresses the challenge of processing heterogeneous and noisy user-generated content from platforms such as Bluesky, TikTok, Reddit, Telegram, and Mastodon. Traditional keyword-based or topic-modelling approaches often fail to capture the spatiotemporal and semantic complexity essential for real-time disaster management. To overcome this, IT:U developed a graph-based framework that unifies semantic and geospatial information, enabling end-to-end learning of clusters that are both semantically coherent and spatially interpretable. This cross-platform approach mitigates data fragmentation, enhancing the robustness of social media-based disaster monitoring.

Using Hurricane Ian as a case study, IT:U demonstrated that integrating data from multiple platforms improves situational awareness, even as no single source can fully substitute Twitter. Spatial context-dependent topic modelling revealed stable disaster-relevant themes and their regional evolution over time, while embedding-based clustering uncovered fine-grained distinctions in the informativeness and actionability of posts. Together, these advances establish a comprehensive framework for analysing heterogeneous, sparse, and noisy user-generated data, enriching disaster analysis with nuanced, geospatially grounded insights.

2.2. Explainable AI Methods for Extreme Data Conditions

Diffusion models have demonstrated strong capabilities for generating realistic synthetic images, particularly for natural disaster scenarios such as forest fires. Understanding and interpreting these models is crucial, as explainability provides a form of pseudo ground-truth for objects like fire and smoke. By mapping the influence of textual tokens onto image regions, explainable diffusion techniques allow the automatic extraction of bounding boxes and segmentation masks, supporting the creation of annotated datasets without manual labelling.

State-of-the-art explainability approaches, including DAAM (Diffusion Attentive Attribution Maps) and Attention-Map-Diffusers, convert cross-attention mechanisms into spatial maps that localize textual concepts in generated images. DAAM offers a lightweight, training-free solution suitable for Stable Diffusion and SDXL pipelines, while Attention-Map-Diffusers provides fine-grained token-to-region attribution compatible with newer high-fidelity models like Flux. Both tools enable precise localization of key elements, although Attention-Map-Diffusers demonstrates improved accuracy in segmenting fire and smoke regions.

The explainability pipelines involve processing attention maps with Gaussian smoothing, K-means clustering, morphological operations, and contour extraction to highlight and isolate relevant regions. These can be used to generate segmentation masks and bounding boxes. These methods allow the systematic identification of relevant regions for each token, producing reliable annotations for fire detection datasets. While DAAM provides a straightforward integration and works well for simple tokens, Attention-Map-Diffusers delivers higher-quality, more localized ground-truth suitable for advanced model outputs.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

By leveraging explainability in diffusion models, the work done by ATOS establishes a methodology for generating synthetic disaster imagery with automated annotations. This approach enhances dataset quality, reduces reliance on manual labelling, and enables the creation of training data that accurately reflects object locations and shapes, forming a foundation for subsequent detection and segmentation tasks.

Additionally to the work carried out by ATOS for generating synthetic data using XAI on Diffusion models, FHHI pursued its work on generic XAI methods started in Task3.1, while focusing on XAI methods for handling extreme data conditions as they occur in the context of natural disaster management. In particular for handling data scarcity, and consequently risks of overfitting, FHHI proposed a novel regularization technique based on explanation-guided data augmentation. Similarly, in order to reveal under-represented features due to covariate shift and data scarcity and retrain a model on consolidated data, FHHI proposes to explain predictive uncertainty of deep learning models by decomposing the uncertainty estimated via a variance over an ensemble of predictions and taking into account second-order effects, thereby extending classical first-order explanation techniques, such as Layer-wise Relevance Propagation (LRP) and Gradient×Input (GI) into more powerful second-order uncertainty explainers (CovLRP, CovGI, etc.)

Further, in order to improve the detection of data-related issues inside the training data leading to suboptimal performance, and allow for model debugging, FHHI improved over standard data-attributions XAI methods based on Influence Functions, through proposing DualXDA, a method that combines sparse attributions using a linear multiclass SVM, together with feature-attribution, in order to explain why training samples are relevant for the prediction of a test sample in terms of impactful features. Besides its high performance on downstream tasks, DualXDA dramatically improves in terms of XAI computation time w.r.t. state-of-the-art data-attribution methods.

Moreover, FHHI evaluated the impact of colormap choices for the cognitive load on end users when visualizing heatmaps, underscoring the fact that interactive interfaces are better adapted to minimize this load, which is especially helpful for high-stake scenarios, where human decisions need to be taken rapidly, such as in natural disaster management.

Lastly, FHHI made progress in interpretability-driven shortcut detection and bias mitigation using concept-based explanations, and improved its concept-based XAI generation pipeline for delivering near real-time explanations inside the TEMA software platform, fulfilling all target values that were defined by the KPIS of the TEMA Objective OA1 "Increase trustworthiness of extreme data analysis algorithms".

2.3. Multiple Learning Paradigms

To enhance adaptability and collaboration among distributed DNN systems, AUTH proposed the Fire Classification Multi-Agent (FCMA) framework [3]. This architecture integrates peer-to-peer Knowledge Distillation (KD) and Federated Learning (FL) for collective intelligence across autonomous DNN agents operating in Natural Disaster Management (NDM) scenarios. The system includes an Agent Knowledge Self-Assessment (AKSA) module employing Likelihood Regret-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

based Out-of-Distribution detection to evaluate each agents competence and trigger collaborative knowledge transfer when encountering unfamiliar or insufficient amounts of data. Experiments on the Blaze dataset demonstrate that peer-to-peer KD improves classification accuracy by +1.19 percentage points (from 76.31% to 77.50%), surpassing the FL approach. The complete study was presented as a conference paper [3] and is discussed in Section 5.

AUTH addressed the persistent problem of data fragmentation and incremental availability in real-world disasters through the Feedback Continual Learning Vision Transformer (FCL-ViT) framework [4]. Unlike conventional feed-forward continual learning approaches, FCL-ViT employs Tunable self-Attention Blocks (TABs) and Task-Specific Blocks (TSBs) that dynamically adjust attention across tasks via a feedback mechanism, eliminating the need for stored exemplars or network expansion. This approach enables learning from sequential or limited datasets without catastrophic forgettingan essential capability when labeled data are sparse or collected incrementally during emergencies. Experimental evaluations on the BLAZE wildfire classification and CIFAR100 datasets demonstrate that FCL-ViT maintains stable accuracy across sequential learning tasks, effectively mitigating catastrophic forgetting. This confirms its capability for sustained adaptability and long-term generalization under real-world conditions. Detailed results are provided in Section 5 and the corresponding journal publication [4].

To overcome the constraints of distributed data availability and isolated local learning, AUTH developed the Cloud Learning-by-Education Node Community (C-LENC) framework. Building upon the LENC paradigm, the C-LENC framework enables scalable, distributed AI collaboration across networked nodes. Implemented using containerized Docker-based nodes, C-LENC supports diverse workflows including federated learning, multi-teacher knowledge distillation, and distributed inference, all operating in real-time cloud environments. A secure service discovery protocol allows nodes to autonomously identify peers and exchange knowledge via TCP-based communication channels. The framework was evaluated on the CIFAR-10 and BLAZE datasets across four workflows, demonstrating successful distributed learning and inference. The method allows AI systems to learn collaboratively with reduced computation latency in decentralized, data-scarce environments. The method is described in Section 5 and in an accepted conference publication.

To tackle the dual challenges of non-IID data distribution and communication constraints in decentralized environments, AUTH introduced Proto-SVDD, a prototype-based federated learning framework for object detection in UAV-based disaster monitoring. Unlike traditional FL methods that exchange entire model weights, Proto-SVDD transmits compact, class-wise Support Vector Data Description (SVDD) prototypes derived from YOLOv6 classification heads, significantly reducing communication overhead and enhancing privacy. Each client exchanges only SVDD centers with peers and uses a prototype alignment loss to maintain consistency across decentralized learners. Evaluations on the VisDrone-DET2019 dataset under non-IID splits show that Proto-SVDD achieves superior or comparable mAP performance while cutting communication to two vectors per client per round. The framework is presented in Section 5 and detailed in a technical report.

To address the limited data and overfitting challenges that arise when training deep models for wildfire burnt area segmentation, AUTH developed a Neural Architecture Search and Knowledge Distillation (NASKD) pipeline [5]. Large neural networks trained on limited disaster datasets often memorize training samples and fail to generalize to unseen events, reducing their reliability in real-world operations. To overcome this, AUTH combined automated architecture optimization

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(NAS) with teacher-student knowledge distillation (KD) to design compact, high-performing segmentation models tailored to the available data. The proposed pipeline systematically searches across multiple DNN familiesincluding PIDNet, UNet++, and CNN-I2I BiseNetoptimizing for mean Intersection over Union (mIoU) while minimizing computational cost. Through successive distillation steps, smaller student models inherit the spatial precision and semantic consistency of larger teacher models. When evaluated on the Blaze dataset [1], the final NASKD configuration achieved a 62.3% reduction in trainable parameters and a +1.02% mIoU improvement over baseline architectures, enabling real-time deployment in UAV-based wildfire monitoring systems. The complete method and results are described in Section 5 and in a conference paper.

To address the annotation limitations and data scarcity that hinder large-scale semantic segmentation in Natural Disaster Management (NDM), AUTH developed the Multiclass Extreme Weak Supervision (MEWS) framework. Traditional weakly supervised segmentation approaches, such as CLIP-ES and ExCEL, rely on text-based supervision and prompt-derived embeddings, which often suffer from ambiguity and poor generalization. Building upon the binary Extreme Weakly Supervised (EWS) paradigm, AUTH extended this approach to multiclass segmentation, enabling the discrimination of multiple disaster-related categories - such as flooded regions, debris, vegetation, and infrastructure - using only a few annotated pixels per class. MEWS integrates self-supervised DINO feature representations with sparse pixel annotations treated as class prototypes, and introduces a novel margin triplet loss that explicitly models inter-class dependencies while maintaining intra-class consistency. This design ensures stable training and effective feature separation even with extremely limited supervision. Evaluated on the Cityscapes benchmark, MEWS achieved a mean IoU of 61.19%, outperforming both the DINO+Prototypes and other state-of-the-art weakly supervised baselines. This work facilitates rapid model training and deployment under limited supervision. The method is detailed in Section 5 and in a technical report.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

3. Facing Data Scarcity on Natural Disaster Management

3.1. Diffusion Models for Natural Disaster Management

Diffusion models have emerged as a powerful class of generative models [6], particularly for image synthesis and transformation. In the context of natural disaster research, these models address data scarcity by generating synthetic data [7], augmenting limited datasets, and simulating diverse disaster scenarios. As of early 2024, state-of-the-art (SOTA) diffusion models demonstrate remarkable capabilities, producing high-quality images by iteratively refining random noise. Their ability to create diverse and realistic outputs often surpasses traditional generative adversarial networks (GANs) in both fidelity and variety. Diffusion models can generate synthetic data, augment small datasets [8], and simulate diverse disaster scenarios, thereby enhancing the effectiveness of predictive models and preparedness strategies.

To illustrate their role in disaster-related applications, diffusion models support three main functions:

- Generating Synthetic Data: Diffusion models can produce realistic, high-quality images of natural disasters from limited datasets. This synthetic imagery expands available data, enabling broader coverage of scenarios such as floods, fires, and other natural hazards. By filling these gaps, synthetic data strengthens predictive models and improves their robustness.
- Data Augmentation: Diffusion models augment small datasets [8] [9] by generating new samples consistent with the statistical properties of the original data. Augmented datasets can also lead to more accurate and reliable predictive models for disaster response, enhancing their ability to forecast and mitigate the impacts of floods, fires, and related events.
- Simulation and Scenario Planning: Diffusion models simulate diverse disaster situations, providing researchers and emergency responders with visualizations that aid training, preparedness, and response planning. These simulations help organizations anticipate potential challenges and develop effective mitigation strategies.

3.1.1. Study of the SOTA

Diffusion Models SOTA early 2024

The field of diffusion models advanced significantly throughout 2023 and early 2024, with several state-of-the-art (SOTA) models emerging as leaders in image generation. This section reviews these advancements, focusing on performance metrics and licensing considerations for Stable Diffusion XL [10], DALL-E 2 [11], Midjourney V5, and Imagen [12].

This project has received funding from the European Union – European Commission under grant agreement No 101093003

During this period, research efforts concentrated on improving the quality, efficiency, and adaptability of diffusion models. The primary objectives were to increase image fidelity and diversity while addressing practical concerns such as training time and performance under limited data conditions. Each model offers distinct strengths and limitations, which are reflected in both their technical benchmarks and their licensing terms. Comparing these models provides valuable insights into their practical applications and suitability across different use cases.

To evaluate these advancements, ATOS conducted a comparative study of the leading models. Our assessment examined key performance indicators that capture image quality and diversity, as well as licensing terms that determine their accessibility and use in our use case.

Table 1. Performance Metrics and Licensing of Leading Diffusion Models as of Early 2024

Metric	Stable Diffusion XL	DALL-E 2	Midjourney V5	Imagen
FID Score	8.4	10.2	9.0	11.5
Inception Score	9.1	8.8	9.0	8.5
Training Time (hours)	12	20	15	25
Adaptability to Scarcity	High	Moderate	Moderate	Low
License	Creative ML OpenRAIL-M License	OpenAI API Terms of Service	Midjourney Terms of Service	Research and Non-Commercial Use License

Key Performance Indicators:

- FID Score (Fréchet Inception Distance): Lower scores indicate better quality and diversity of generated images.
- Inception Score: Higher values reflect better quality in terms of the recognition of generated images.
- Training Time: Represents the computational resources required to train the model effectively.
- Adaptability to Scarcity: Qualitative assessment of how well the model performs when trained on limited datasets.

Licensing Information for Leading Diffusion Models

- Stable Diffusion XL: Released under the Creative ML OpenRAIL-M License, which allows both commercial and non-commercial use, with restrictions on harmful applications. Attribution is required, and sharing modifications under the same license is encouraged.
- DALL-E 2: Distributed via the OpenAI API Terms of Service. The model is accessible only through OpenAIs API. Commercial use is generally allowed, but content restrictions apply (e.g., no adult or hateful content). Attribution to OpenAI is required.
- Midjourney V5: Operates under the Midjourney Terms of Service. It follows a subscription model, where personal use is allowed but commercial use requires a separate license. All users must comply with community guidelines.
- Imagen: Governed by a Research and Non-Commercial Use License. Developed by Google, Imagen is limited to research contexts and cannot be used commercially without explicit permission. Attribution to Google is required.

Among these models, ATOS selected Stable Diffusion XL as it stands out as the premier choice due to its balance of strong performance and flexible licensing. With a Fréchet Inception Distance (FID) of 8.4 and an Inception Score of 9.1, it consistently delivers high-quality, diverse image outputs, outperforming many competitors. Its adaptability to limited datasets is a crucial

This project has received funding from the European Union – European Commission under grant agreement No 101093003

advantage in real-world applications where data scarcity is common. Furthermore, its permissive yet responsible licensing framework allows both commercial and non-commercial use while enforcing ethical standards and requiring attribution. This combination of superior performance and licensing flexibility makes Stable Diffusion XL the most practical and reliable choice for integrating cutting-edge diffusion models into disaster data augmentation workflows.

Advances in Diffusion Models SOTA

From 2024 to 2025, diffusion models continued to advance rapidly, introducing innovative techniques and new architectures that significantly enhanced image generation capabilitiesdevelopments particularly relevant for disaster response applications. This section reviews recent models such as Flux1.dev [13], Stable Diffusion 3 [14], DreamFusion [15], DeepArt, and SynGen, with a focus on efficiency, scalability, and the ability to produce increasingly complex and high-fidelity images.

During this period, these models were extensively evaluated and benchmarked. Each introduced distinct improvements, reflected in their performance metrics and licensing frameworks, offering insights into their strengths and suitability for different use cases.

Table 2 summarizes the key indicators and licensing terms of these leading models.

Table 2. Performance Metrics and Licensing of Leading Diffusion Models as of Early 2024

Metric	Flux1.dev	Stable Diffusion 3	DreamFusion	DeepArt	SynGen
FID Score	7.2	8.0	8.0	9.1	10.0
Inception Score	9.4	9.3	9.2	8.9	8.7
Training Time (hours)	10	14	18	22	20
Adaptability to Scarcity	Very High	High	High	Moderate	Low
License	Open Source License	Creative ML OpenRAIL-M License	Commercial License	Academic Use License	Proprietary License

Key Performance Indicators:

- FID Score (Fréchet Inception Distance): Lower scores indicate better quality and diversity of generated images.
- Inception Score: Higher values reflect better quality in terms of the recognition of generated images.
- Training Time: Represents the computational resources required to train the model effectively.
- Adaptability to Scarcity: Qualitative assessment of how well the model performs when trained on limited datasets.

Licensing Information for Leading Diffusion Models

- Flux1.dev: Open Source License, Released under an open-source license supporting commercial and non-commercial use, with emphasis on ethical practices and community contribution. Attribution is required, and modifications are encouraged to be shared.
- Stable Diffusion 3: Creative ML OpenRAIL-M License, Permits both commercial and non-commercial use, subject to attribution, ethical guidelines, and restrictions on harmful applications.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

- DreamFusion: Commercial License, Access requires a paid subscription. Designed for professional use with commercial provisions, support services, and attribution requirements.
- DeepArt: Academic Use License, Intended for research and academic purposes. Commercial use requires explicit permission. Attribution and community sharing of modifications are mandatory.
- SynGen: Proprietary License, Commercially oriented with strict access controls and usage restrictions to ensure ethical deployment.

ATOS evaluated these newer models, and Flux1.dev emerged as the optimal choice for our applications. With a FID score of 7.2 and an Inception Score of 9.4, it delivers superior image quality and diversity while maintaining strong adaptability under data-scarcity conditions. These capabilities are essential in real-world scenarios where disaster datasets are often limited. Furthermore, its open-source license promotes transparency, ethical use, and community-driven innovation. This combination of high performance, robustness, and permissive licensing makes Flux1.dev the premier model for efficient and responsible integration into our workflows.

LLM SOTA early 2024

In the latter part of 2023 and early 2024, several leading Large Language Models (LLMs) emerged as key contenders for generating prompts to guide diffusion models in image creation. This section evaluates the capabilities of OpenAIs GPT-4 [16], Googles PaLM 2 [17], Anthropics Claude [18], and Mistral [19], focusing on their technical strengths and licensing frameworks.

- **GPT-4 (OpenAI):** A frontrunner due to its vast training corpus and sophisticated architecture. GPT-4 excels at producing detailed, contextually relevant prompts that enhance diffusion model outputs. It is offered under a proprietary license, requiring compliance with usage guidelines and subscription-based access.
- **PaLM 2 (Google):** Known for its multilingual breadth and nuanced prompt generation. PaLM 2 supports diverse and complex image descriptions. It is licensed under a mix of proprietary and limited open-access terms, which broaden accessibility while restricting certain commercial uses.
- **Claude (Anthropic):** Emphasizes safety, interpretability, and ethical alignment. Claudes prompt generation is designed to reduce harmful or biased outputs, ensuring that diffusion models align with societal norms. It is distributed under a proprietary license with strict ethical usage conditions.
- **Mistral:** A newer but impactful entrant, valued for its lightweight architecture that balances performance with computational efficiency. Mistral delivers strong prompt-generation capabilities without requiring extensive resources. Released under an open-source license, it supports transparency, collaboration, and broad accessibility.

These LLMs stand out for their ability to produce precise, contextually rich prompts that significantly enhance diffusion model performance. Their licensing terms reflect a spectrum between proprietary control and open accessibility, shaping how they can be adopted in practice.
ATOS selected Mistral 7B as the preferred model. Its open-source license provides transparency and deployment flexibility, while its architecture offers a strong balance of model size and performance. Crucially, Mistral can be deployed on-premise, aligning with requirements for efficiency, security, and responsible innovation.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 3. Comparison of LLMs for Image Generation Prompting

Feature	GPT-4	PaLM 2	Claude	Mistral
Developer	OpenAI	Google	Anthropic	Mistral AI
Release Year	2023	2023	2023	2024
Prompt Generation Quality	High	High	High	High
Multilingual Support	Yes	Yes	Yes	Yes
Contextual Understanding	Excellent	Excellent	Excellent	Very Good
Ethical Considerations	Moderate	Moderate	High	Moderate
Resource Efficiency	Moderate	Moderate	Moderate	High

Table 4. Licensing Information of LLMs for Image Generation Prompting

License Aspect	GPT-4	PaLM 2	Claude	Mistral
Type of License	Proprietary	Proprietary/ Open Access	Proprietary	Open Source
Commercial Use	Subscription Plan	Restricted	Restricted	Permitted
Ethical Guidelines	Moderate	Moderate	Strict	Moderate
Access Restrictions	Yes	Partial	Yes	No
User Guidelines	Detailed	Detailed	Extensive	Community-based

Advances in LLM and VLM SOTA

From 2024 to 2025, the landscape of Large Language Models (LLMs) and Vision-Language Models (VLMs) advanced considerably, improving prompt generation, image description quality, and overall applicability across diverse domains. This section reviews key modelsincluding Molmo [20], GPT-5 [21], PaLM 3, VisualGPT [22], and ClipText [23]attention to their performance, capabilities, and licensing frameworks.

- **Molmo:** A recently developed model with an innovative architecture and advanced training techniques. Molmo excels at understanding complex prompts and generating contextually accurate responses. Its strong image description capabilities make it highly effective for real-world applications. Molmo is released under an open-source license, supporting transparency, community-driven innovation, and responsible use.
- **GPT-5 (OpenAI):** The latest iteration from OpenAI, building on its predecessors with expansive training data and enhanced reasoning. GPT-5 is particularly strong at producing detailed, contextually rich prompts for image generation. It is available under a proprietary license, requiring adherence to usage policies and subscription-based access.
- **PaLM 3 (Google):** Advances multilingual comprehension and prompt generation beyond its predecessors. PaLM 3 produces diverse and intricate image descriptions, making it a powerful LLM for cross-lingual applications. It is licensed under a hybrid proprietary/open-access model, balancing accessibility with commercial restrictions.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

- **VisualGPT:** Combines LLM and VLM capabilities, excelling in multimodal tasks where both text understanding and visual generation are required. It can both describe images accurately and generate visual outputs from prompts. VisualGPT is distributed under a proprietary license, with usage governed by specific guidelines.
- **ClipText:** Integrates text and image embeddings, enabling seamless interaction between modalities. ClipText is particularly effective for generating descriptive content and enhancing prompt-image alignment. It is licensed under a proprietary framework, requiring compliance with defined protocols.

Table 5 outlines the performance metrics of leading LLMs and VLMs, showcasing their respective strengths in prompt generation and image description.

Table 5. Performance Metrics of Leading LLMs and VLMs (2024–2025)

Metric	Molmo	GPT-5	PaLM 3	VisualGPT	ClipText
Developer	MolmoAI	OpenAI	Google	VisualAI	ClipAI
Release Year	2025	2024	2024	2025	2024
Prompt Generation Quality	Very High	Very High	High	High	High
Image Description Quality	Very High	High	Very High	High	High
Multilingual Support	Yes	Yes	Yes	Yes	Yes
Contextual Understanding	Excellent	Excellent	Excellent	Very Good	Very Good
Ethical Considerations	High	Moderate	High	Moderate	Moderate
Resource Efficiency	High	Moderate	Moderate	High	Moderate

Table 6. Licensing Information of Leading LLMs and VLMs (2024–2025)

Metric	Molmo	GPT-5	PaLM 3	VisualGPT	ClipText
Type of License	Open Source	Proprietary	Proprietary	Commercial	Proprietary
Commercial Use	Permitted	Subscription	Restricted	Subscription	Subscription
Ethical Guidelines	Strict	Moderate	High	Moderate	Moderate
Access Restrictions	No	Yes	Partial	Yes	Yes
User Guidelines	Community-based	Detailed	Detailed	Community-based	Extensive

After evaluating these advancements, ATOS selected Molmo as the optimal model for our applications. Molmo combines superior performance in both prompt generation and image description with a permissive open-source license, enabling transparent, ethical, and community-driven innovation. Its multilingual capabilities, contextual understanding, and resource efficiency make it highly versatile for diverse real-world use cases.

The combination of technical excellence, adaptability, and open licensing positions Molmo as the best choice for deploying next-generation LLM and VLM technologies effectively and responsibly.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

3.1.2. ComfyUI Image Generation Pipelines

To address the scarcity of natural disaster datasets, ATOS employs ComfyUI[24], an open-source, node-based interface for designing and managing Diffusion models workflows, to develop advanced diffusion-based image generation pipelines. Integrating ComfyUI into our workflow enables efficient creation and management of complex image generation processes tailored to disaster data augmentation and simulation.

ComfyUI provides a versatile interface for constructing and controlling image generation pipelines. Its visual, modular design enables users to configure each stage, adjust parameters, and customize outputs with flexibility.

The interface provides a clear visual representation of pipeline components and their interactions. This visualization clarifies the role of each block, highlights the effects of adjustments, and helps achieve desired synthetic data outcomes.

- Pipeline Creation and Management: ComfyUI enables us to design and manage custom pipelines for generating synthetic disaster data. We are able to configure various stages of the diffusion model process, from initial noise generation to iterative refinement, ensuring realistic and diverse output images [25].
- Parameter Optimization: The intuitive interface of ComfyUI allows us to experiment with different parameters and settings, optimizing diffusion models to produce high-quality synthetic data that reflects diverse disaster scenarios, thereby enhancing the robustness of predictive models.
- Integration and Flexibility: ComfyUI's modular design and support for multiple frameworks support integration with diverse image generation tools and techniques. This flexibility enables continuous pipeline improvements by incorporating new advancements in diffusion models and related generative technologies [26].

3.1.3. Text-to-Image Dataset Generation

Text-to-image generation using diffusion models enables the creation of synthetic datasets tailored to specific disaster scenarios. These models gradually transform random noise into coherent images, guided by semantic information extracted from textual prompts. By learning from large textimage pairs, diffusion models can generate realistic or stylized images aligned with descriptive language.

For disaster response, this capability is particularly valuable: synthetic datasets can augment real-world data, address underrepresented conditions, and simulate controlled scenarios that would otherwise be costly or impossible to capture. By expanding datasets efficiently, diffusion-based generation ensures strong textimage alignment while reducing reliance on manual collection.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Mistral prompt + Stable Diffusion XL

To generate synthetic disaster imagery, ATOS implemented a pipeline that integrates Mistral 7B (LLM) with Stable Diffusion XL (text-to-image diffusion model). Mistral generates rich prompts describing trial scenarios, such as Finnish forests or Greek and Austrian floods, which are then transformed into high-quality images by Stable Diffusion XL. A detailed pipeline image can be found in Appendix A.

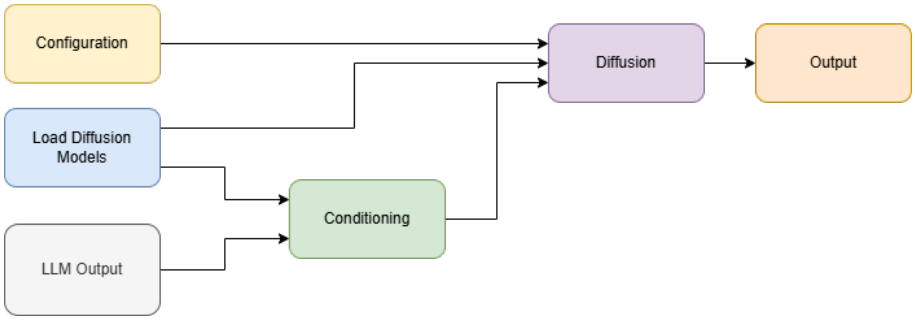


Figure 1. Mistral prompt + Stable Diffusion XL pipeline. [credit ATOS]

Pipeline description (Figure 1):

1. **Load Diffusion Models (Blue):** Load Checkpoint of stable diffusion XL CLIP and VAE, sends model and VAE to the diffusion block, and CLIP to the conditioning block.
2. **LLM Output (Grey):** Mistral 7B generates descriptive text from the prompt (Different prompts in Appendix B) forwarded to the Conditioning block.
3. **Configuration (Yellow):** Seed and Empty Latent Image define size and randomization.
4. **Conditioning (Green):**
   - Positive: Mistral-generated text concatenated with fixed keywords (e.g., realistic, highly detailed, birds-eye view).
   - Negative: constant keywords remove unwanted artifacts (cropped, glitch, bad anatomy).
5. **Diffusion (Purple):** KSampler integrates model, seed, latent image, and conditioning to synthesize images. VAE Decode transforms latent outputs into final images
6. **Output (Orange):** Images are previewed and stored in the dataset.

The pipeline produced two synthetic datasets:

- **Forest Fire Images**: 600 images depicting various forest fire scenarios. These images capture different stages and intensities of fires, providing a diverse dataset that can be used for training and testing fire prediction and management models.
- **Flood Images**: 760 images of floods in different environments, including rural and urban settings. These images illustrate various flooding conditions, helping improve flood detection and response strategies.

These synthetic datasets significantly enrich the available training data, enhancing robustness and accuracy in disaster-response AI. By simulating a wide range of environmental conditions, they fill critical gaps in real-world datasets and ensure preparedness across diverse scenarios.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 2. Forest Fire image generated with Stable Diffusion XL using the prompt: "Come up with a description of a realistic finnish forest fire at dawn." [image generated by ATOS]

Figure 3. Flooded town image generated with Stable Diffusion XL using the prompt: "Come up with a description of a street in Austria in a rainy day after heavy rain and floods, damaged buildings, trapped cars." [image generated by ATOS]

Molmo Description + Flux dev

In this second pipeline, ATOS used a combination of Molmo (VLM) with Flux.1-dev (diffusion model) to generate synthetic datasets based on real-world imagery. Historical images of natural disasters (floods and fires), provided by partners, were used as the basis. Molmo extracted detailed textual descriptions of each image, which were then used as prompts for Flux.1-dev to synthesize new images.

For each description, 50 synthetic images were generated. In total, the pipeline produced 550 forest fire images and 500 flood images, recreating and extending the real-world disaster sce-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

narios. A detailed pipeline diagram is provided in Appendix C).

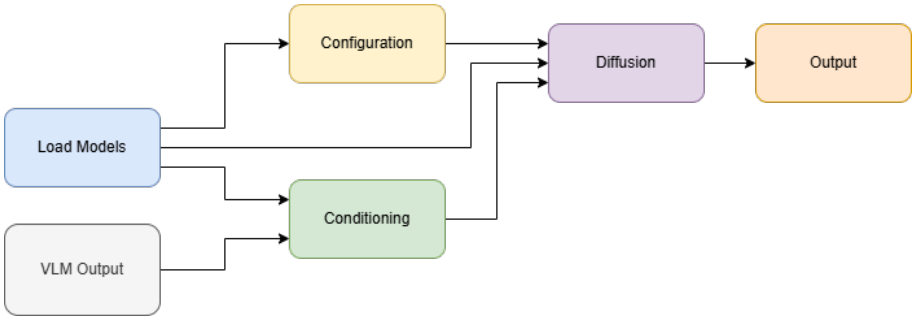


Figure 4. Molmo Description + Flux dev pipeline. [credit ATOS]

Pipeline description (Figure 4):

1. **Load Models (Blue):** Load Checkpoint of Flux.1-dev CLIP and VAE. The model and VAE are directed to the Configuration and Diffusion blocks, while is sent CLIP to the Conditioning block.
2. **VLM Output (Grey):** Molmo-1b analyzes historical disaster images and generates descriptive text prompts (details inAppendix D). These prompts are forwarded to the Conditioning block.
3. **Configuration (Yellow):** Seed, Random Noise, Scheduler, and the Empty Latent Image are defined and sent to the Diffusion block.
4. **Conditioning (Green):** Converts the text prompt into a guidance parameter for Flux.1-dev.
5. **Diffusion (Purple):** The sampler integrates model, seed, scheduler, latent image, and conditioning to synthesize outputs. The VAE Decode transforms latents into final images.
6. **Output (Orange):** Generated images are previewed and stored.

Image description forest fire of Figure 5: The image shows a striking aerial view of a forest after a recent fire. The landscape is dominated by a vast expanse of green forest, with the trees appearing lush and vibrant, especially in the foreground. In the center of the image, theres a large, dense cloud of smoke rising from the forest. This smoke is thick and billowing, creating a stark contrast against the green backdrop. The smoke appears to be spreading across the image, likely due to wind currents. To the left side of the image, a body of water is visible. It could be a lake or a river, and its surface is reflecting the smoke from the forest. This adds another layer of visual interest to the scene. The layout of the image is quite interesting. The forest dominates the lower portion, while the smoke rises prominently in the center. The upper part of the image shows more of the surrounding area, including what appears to be a clear sky. The scene is devoid of any visible human structures or people, which emphasizes the raw, untouched nature of the forest. The color palette is predominantly green, with the smoke providing a stark contrast. Overall, this image captures a powerful post-fire landscape, showcasing natures resilience and the immediate impact of wildfires on forest ecosystems.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a prescribed fire in Finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 5. Real origin fire image and synthetic resulting image comparison

Image description flood of Figure 6: The image shows a drone's-eye view of a flooded rural landscape after heavy rainfall. The scene is dominated by a muddy river that has swollen to an unusually wide width, covering much of the visible terrain. The water has a thick, brownish-yellow hue, indicating significant pollution or sediment content. The river's banks are lined with dense green trees on both sides, creating a stark contrast between the lush vegetation and the muddy waters. The trees appear to be thriving, suggesting this flooding may be a recurring event in the area. To the right of the river, there's a grassy field that's partially submerged. This field is dotted with small puddles and patches of standing water, indicating the extent of water saturation in the soil. The sky above is overcast with gray clouds, which adds to the somber and dramatic atmosphere of the scene. The lack of direct sunlight suggests that the image was taken during a cloudy day, possibly in the afternoon. The overall layout of the scene is characterized by the wide, muddy river running through the center of the image, flanked by trees on both sides and a grassy field to the right. The color palette is predominantly earthy, with various shades of brown, green, gray, and white dominating the view. This image captures the immediate aftermath of a flood event, showcasing the dramatic impact of heavy rainfall on rural landscapes. It provides a unique perspective on how water can transform natural environments, temporarily turning everyday fields into vast, muddy waters.

The pipeline produced two synthetic datasets:

- **Forest Fire Dataset**: 550 images recreating historical fire events, capturing environmental diversity and varying fire intensities.
- **Flood Dataset**: 500 images simulating historical flood scenarios across rural, urban, and mixed landscapes.

These datasets not only extend the real-world imagery but also enable the generation of controlled variations of past disaster events. This approach provides valuable synthetic augmentation, improving model resilience by bridging gaps in historical data while maintaining strong visual realism.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a flood from the ahrtal region [Credit: Bavarian Red Cross 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 6. Real origin flood image and synthetic resulting image comparison

3.1.4. Dataset Augmentation Image to Image

Image-to-image generation modifies or enhances an existing image according to a target specification while preserving the underlying scene structure. Unlike text-to-image, the model starts from a real image and applies learned transformations, adding objects, changing weather, or simulating damage. Diffusion-based approaches perform this by iterative denoising conditioned on the input image, with optional guidance from a text prompt or an auxiliary image.

For disaster-response applications, image-to-image models are especially valuable: they can synthesize realistic variations of scenes that are rare, hazardous, or impractical to capture in the field. Examples include injecting fire or smoke, simulating flood extent, or altering environmental conditions. This method augments datasets without risky data collection, exposes models to rare yet critical conditions, and keeps the augmented outputs closely aligned with the original real-world context (including geometry, lighting, and metadata). The following subsections detail our pipelines and resulting datasets.

Montiferru drone image fire augmentation

The pipeline created by ATOS augments a real dataset of 134 drone flight images from the Montiferru region by adding synthetic fires, while preserving all associated metadata. The approach ensures the resulting dataset retains geographical and contextual integrity, making it highly suitable for location-specific disaster simulations. The full pipeline includes resizing, inpainting, refinement, upscaling, and metadata transfer. The location of all the images can be seen on Figure 7.

Step 1 Image Resizing:

This project has received funding from the European Union – European Commission under grant agreement No 101093003

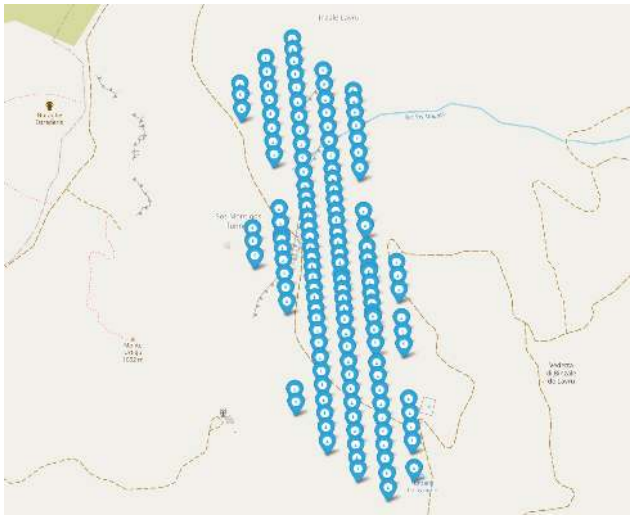


Figure 7. Map of the original dataset in the Montiferru Region [credit ATOS]

To accommodate hardware constrains during the inpainting model each original drone image (5280Œ3956 pixels) was reduced to a quarter resolution (1329Œ990).

Steps 2 and 3 Fire Inpaint and Refinement:

The resized images are sent as input to the pipeline (Figure 8) which will inpaint fires in a random crop of the image and then refined to avoid artifacts. (Detailed pipeline image in Appendix E).

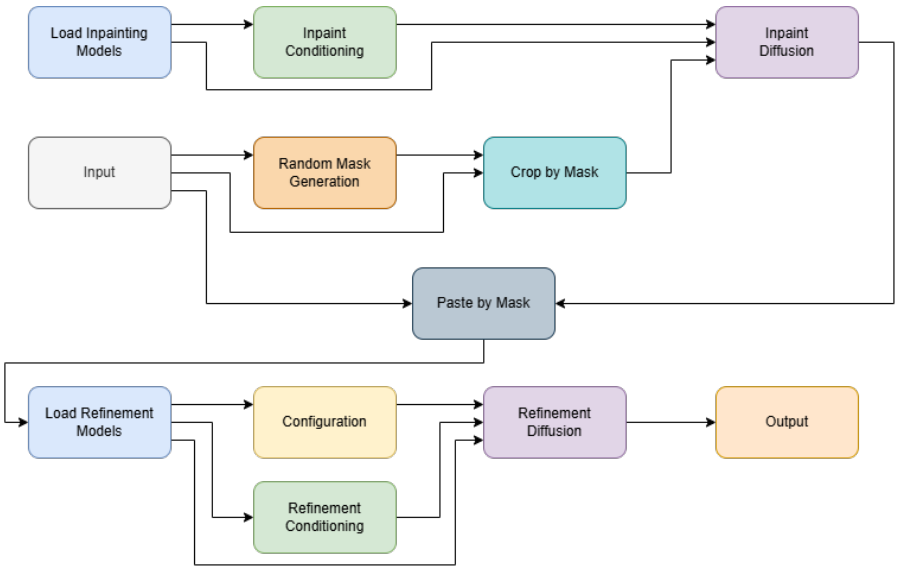


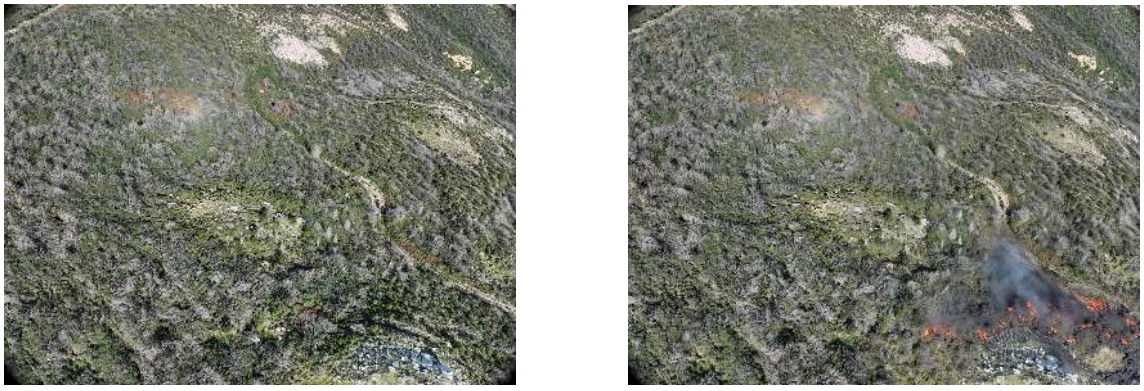
Figure 8. Stable Diffusion XL Inpaint + Flux.1 dev Refinement pipeline [credit ATOS]

Pipeline description for fire inpainting (Figure 8):

1. **Load Inpainting Models (Blue):** Loads the checkpoints for Stable Diffusion XL, CLIP, VAE and Foocus Inpaint [27] patch. Send the checkpoints into the Inpaint diffusion block, and CLIP to the Inpaint conditioning block.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

2. **Input (Grey):** Resized images are passed into the pipeline.
3. **Inpaint Conditioning (Green):**
   - Positive: "Drone image of a burnt area, fire spots, smoke rising, burnt brushes".
   - Negative: "Text, watermark, artifacts".
4. **Random Mask Generation (Orange):** Random square mask that is 10% the size of the image where the fire will be inpainted, this mask is sent to the Crop by Mask block.
5. **Crop by Mask (Teal):** Extracts the masked crop for the inpainting, this crop is sent to the Inpaint Diffusion block.
6. **Inpaint Diffusion (Purple):** The sampler adds synthetic fire elements. VAE Decode converts latent to actual image.
7. **Paste by Mask (Dark Blue):** Reinserts the inpainted crop into the original image.
8. **Load Refinement Models (Blue):** The loading checkpoint of Flux.1-dev CLIP and VAE sends the model to the Configuration and Refinement Diffusion blocks, VAE to the Refinement Diffusion block, and CLIP to the Refinement Conditioning block.
9. **Configuration (Yellow):** Seed, noise, scheduler, and latent image defined, and sent to the Refinement Diffusion block.
10. **Refinement Conditioning (Green):** Converts the text prompt into a guidance parameter for the refinement model.
11. **Refinement Diffusion (Purple):** The Sampler uses model, seed, guider, sampler, latent image, and conditioning to refine the input image. VAE Decode converts latent to actual image.
12. **Output (Orange):** Preview and store the refined image.



(a) Real image taken from a drone in the Montiferru region [provided by RAS]

(b) Augement image with fire added using the prompt: "Drone image of a burnt area fire spots, smoke rising, burnt bushes, burned and black ground,flames" [image augmented by ATOS]

Figure 9. Comparison between real drone image and augmented one

Step 4 Upscaling:
The refined images are upscaled back to the original resolution $(5280 \times 3956) using Real ESRGAN x4 [28] to reduce the$

Pipeline description of the Upscale step (Figure 10):

This project has received funding from the European Union – European Commission under grant agreement No 101093003

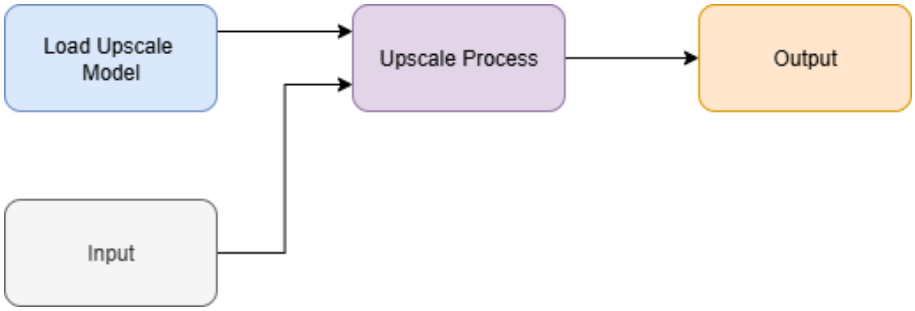


Figure 10. Image Upscale RealESRGAN pipeline [credit ATOS]

1. **Load Upscale Models (Blue):** Loads the checkpoints for RealESRGAN x4 and send it to the Upscale process block
2. **Input (Grey):** Load the input images with synthetic fire to upscale
3. **Upscale Process (Purple):** this block extracts multi-scale features from the low-res image using deep residual blocks, then progressively upsamples ($\times 2$ twice) with sub-pixel convolutions. It refines textures with adversarial + perceptual losses, producing a $4\times$ sharper, more natural-looking image with reduced noise and artifacts.
4. **Output (Orange):** Preview and store the refined image.

Step 5 Metadata preservation:
Original drone metadata (geolocation, altitude, timestamp, camera parameters) was copied to the corresponding synthetic image, preserving essential contextual information.

The augmentation process not only increases the volume of available data but also ensures that the data is contextually rich and relevant for real-world applications, while reducing the necessity to do prescribed fires to generate a dataset. This approach enhances the ability of AI models to generalize from synthetic to real-world scenarios, improving their reliability and effectiveness.

- **Augmented Fire Images**: Each of the 134 images was enhanced with synthetic fire elements, resulting in a comprehensive dataset that demonstrates various fire behaviors and effects on the landscape. This augmented dataset is crucial for improving fire detection algorithms and emergency response planning.
- **Metadata Preservation**: By maintaining the original metadata, the augmented images retain their geographical and contextual information, making them highly valuable for location-specific trials and testing of the platform.

This activity directly advances OA2 by providing high-quality, semantically meaningful datasets for training and validating advanced AI models. The augmented fire imagery introduces realistic and variable conditionssuch as different smoke densities, flame patterns, and terrain interactionsthat strengthen the capacity of semantic analysis algorithms to accurately interpret visual data under complex and extreme circumstances. By exposing models to these diverse and lifelike visual scenarios, the resulting algorithms are expected to achieve significantly higher accuracy in detecting and classifying wildfire events.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

The augmented fire datasets are particularly relevant to the Mediterranean Forest Fires pilot trial in Sardinia. By replicating realistic wildfire conditions, including smoke, reduced visibility, and varying terrain effects, the datasets enable more accurate testing and validation of the TEMA platform. These enhanced data resources help improve situational awareness and decision support for Civil Protection Authorities (CPAs) and First Responders (FRs), leading to better prediction, detection, and management of wildfire events.

Ahrtal drone image flood augmentation

The pipeline developed by ATOS augments a real drone dataset captured in the Ahrtal region (Altenahr town) after the floods, enriching it with synthetic floodwaters, debris, and trapped persons while preserving all associated metadata. The dataset consists of 218 real images, whose locations are shown in Figure 11.

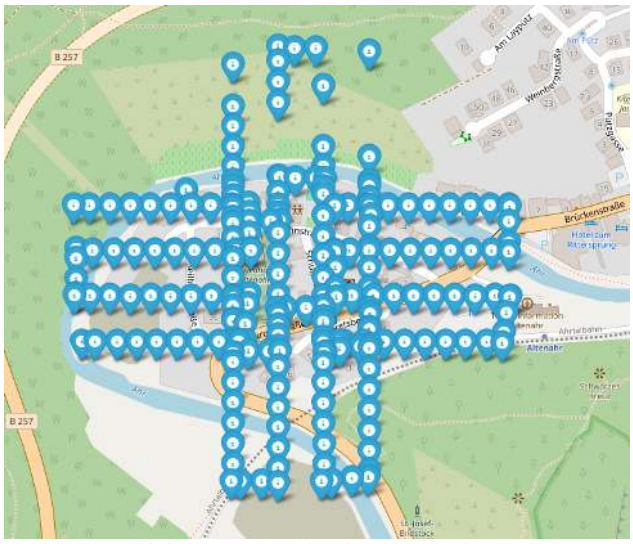


Figure 11. Map of the original dataset in the Ahrtal Region on the Altenahr town [credit ATOS]

To enable processing on available hardware, the images were first resized from $5184 \times 3888 to 1296 \times 972. The resized inputs were then passed through the flood augmentation pipeline (Figure 12, detailed in Appendix G). For t kontext-dev model, which allows image manipulation through guided text prompts.$

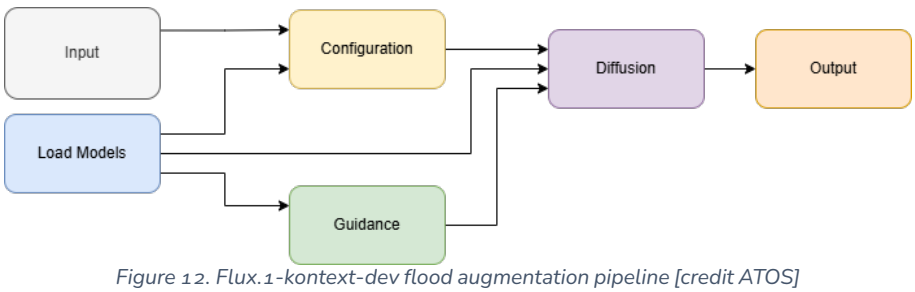
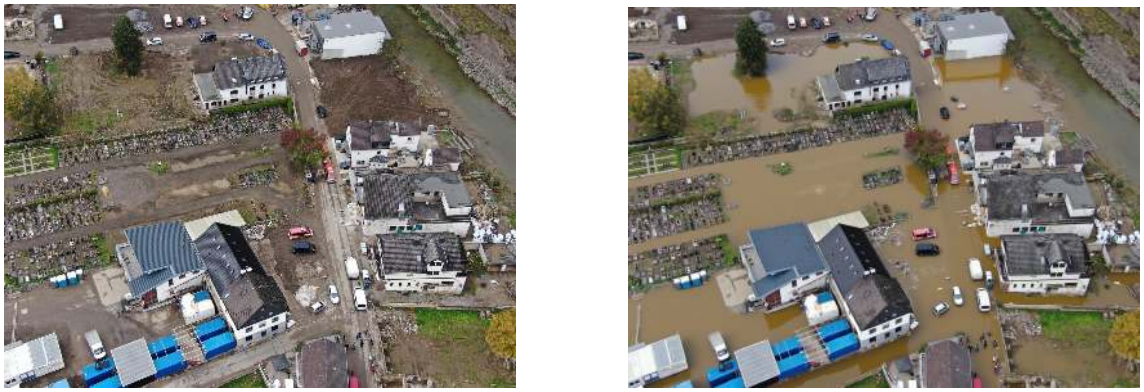


Figure 12. Flux.1-kontext-dev flood augmentation pipeline [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Pipeline description for flood augmentation (Figure 12):

1. **Input (Grey):** Load each resized real image into the pipeline.
2. **Load Models (Blue):** The loading checkpoint of Flux.1-kontext-dev, CLIP and VAE sends the model to the configuration and diffusion blocks, VAE to the diffusion block, and CLIP to the guidance block.
3. **Configuration (Yellow):** Define seed, random noise, scheduler, and latent image from input pixels, forwarding to the Diffusion block.
4. **Guidance (Green):** Apply the prompt: The image is seen from above. Change the ground and streets to be flooded with murky moving water. Change the water in the river to be murky. Add debris to the water. Add the destruction after a flood.
5. **Diffusion (Purple):** The sampler applies the model, seed, guider, and latent conditioning to alter the input image. VAE Decode reconstructs the final image.
6. **Output (Orange):** Preview and store the augmented image.



(a) Real image taken from a drone in the Ahrtal region [provided by DLR]

(b) Augement image with flood added using the prompt: "The image is seen from above. Change the ground and streets image to be flooded with murky moving water. Change the water in the river to be murky. Add debris to the water. Add the destruction after a flood." [image augmented by ATOS]

Figure 13. Comparison between real drone image and augmented one

After augmenting all 218 images with floodwaters and debris, the outputs were passed through the same Upscale pipeline used in the fire augmentation (Figure 10) to restore them to their original resolution.

The next augmentation step introduced trapped persons on rooftops to simulate realistic rescue scenarios. High-resolution crops of random roofs were extracted and modified with Flux.1-fill-dev. The processed crops were then reintegrated into the full image (Figure 14, detailed in Appendix H). An example can be seen on Figure 15 of a person standing on a roof on the bottom right of the image.

Pipeline description for person inpainting (Figure 14):

This project has received funding from the European Union – European Commission under grant agreement No 101093003

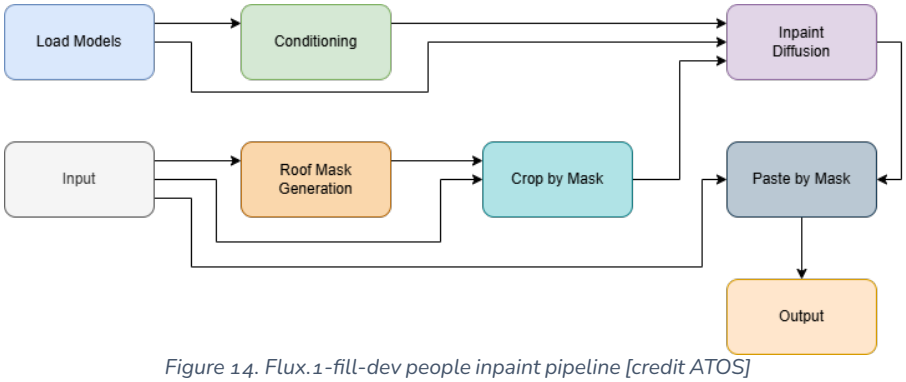


Figure 14. Flux.1-fill-dev people inpaint pipeline [credit ATOS]

1. **Load Models (Blue):** The loading checkpoint of Flux.1-fill-dev, CLIP and VAE sends the model to the configuration and diffusion blocks, VAE to the diffusion block, and CLIP to the guidance block.
2. **Input (Grey):** Load each resized image.
3. **Conditioning (Green):** Apply the prompt: person standing on a roof.
4. **Roof Mask Generation (Orange):** Detect a roof region using a zero-shot detector and generate a square mask.
5. **Crop by Mask (Teal):** Extract the roof crop defined by the mask.
6. **Inpaint Diffusion (Purple):** Use the model, conditioning, and cropped input to inpaint a person onto the roof. Decode with VAE.
7. **Paste by Mask (Dark Blue):** Insert the modified crop back into the original image.
8. **Output (Orange):** Preview and store the refined image.

(a) Real image taken from a drone in the Ahrtal region [provided by DLR]

(b) Augement image with flood and a person traped on a roof using the prompt: "person standing on a roof" [image augmented by ATOS]

Figure 15. Comparison between real drone image and augmented one with inpainted person

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Finally with all the augmented images rescaled back to their original size and added some persons trapped on the buildings, ATOS proceed to copy the metadata from the original images to the upscaled and augmented images.

- **Augmented Flood Images**: Enhanced 218 images with synthetic flood elements, debris, and trapped persons, creating a realistic and diverse dataset that captures various post-flood conditions. These images are instrumental in training models for flood impact assessment and rescue operations.
- **High Resolution and Metadata Integrity**: By maintaining the original metadata, the augmented images retain their geographical and contextual information, making them highly valuable for location-specific trials and testing of the platform.

The augmented datasets provide a rich resource for developing and testing AI models aimed at flood prediction, impact assessment, and emergency response. By simulating realistic disaster scenarios, these datasets help bridge the gap between limited real-world data and the extensive data needs of modern AI systems.

This work directly supports OA2 (Increase accuracy of extreme data analysis algorithms) by providing high-quality, semantically enriched visual datasets that enable the development and validation of novel AI-based semantic analysis algorithms. The inclusion of realistic flood elements, debris, and trapped individuals enhances the representativeness and diversity of the data, allowing models to better learn complex visual patterns associated with extreme events. As a result, these algorithms are expected to significantly improve their accuracy in detecting and interpreting disaster-related visual content from social media and other data sources

The realistic synthetic data enables training and testing of models under conditions similar to those in the first pilot trial and other flood-prone regions of Germany. This supports the flood response teams (FRs) by improving the automatic identification of flooded areas, accessibility of settlements, and detection of affected personscritical aspects for mission planning and timely emergency response during regional flood events.

3.2. UnrealFire: Synthetic annotated image creation pipeline for wildfire segmentation

Wildfire Image Segmentation Datasets SOTA

Existing wildfire image segmentation datasets, such as FLAME [29] or Corsican [30], constitute significant benchmarks in nearly every wildfire image region segmentation or wildfire flame detection system utilizing Deep Neural Networks (DNNs), for Natural Disaster Management (NDM) purposes. While DNN models exhibit commendable accuracy when tested on these datasets, they do not generalize well due to real-world factors, such as the environmental conditions and 3D terrain structure of the aforementioned datasets. Real wildfire image datasets will always be limited by the fact that prescribed fires for data collection purposes must be contained and not be situated near flammable materials, considerably restricting the variety of 3D scenery that can be captured, and limiting their applicability in real-world conditions. The most common ways

This project has received funding from the European Union – European Commission under grant agreement No 101093003

for synthetic image creation are game engines (e.g., Unreal Engine, Unity) or generative machine learning models [31] (e.g. GANs, Diffusion models). These methods can produce high-quality images, but many of them cannot produce accurate segmentation maps or any at all. Existing plugins for game engines like AirSim [32] or CARLA [33] come with some limited segmentation map generation functionalities. With current methods, it is not possible to project particle or transparent objects into the 2D annotation maps. This severely limits the use cases and types of synthetic images that can be created, as most liquid, gas, or plasma objects need to be rendered as particle objects to capture their swiftly changing shapes.

Advances beyond SOTA

AUTH proposes UnrealFire, a method that leverages the cutting-edge capabilities of Unreal Engine (UE) 5, along with AirSim, and Procedural Content Generation tools (PCG). Through UnrealFire, AUTH can generate 'unlimited' pre-annotated wildfire images from a UAV perspective. The method with more details is described in a conference paper [34]:

Evangelos Spatharis, Christos Papaioannidis and Ioannis Pitas, "UnrealFire: Synthetic annotated image creation pipeline for wildfire segmentation", IEEE International Conference on Image Processing Workshop, IEEE ICIPW 2025

The main contribution of the paper is the creation of the particle segmentation camera plugin by modifying the source code of AirSim. Traditionally, fire objects are not registered in the default segmentation camera of AirSim, as it does not support the 2D projection of transparent objects without a set 3D mesh. To this end, AUTH created a particle segmentation AirSim camera. This process involved creating a new custom Post Process Material to get the semantic segmentation map of fire particles. The new Post Process Material loads the captured image from the G-Buffer at a stage where transparent objects (e.g., fire) were not rendered, as well as the final image RGB scene. Subtracting the first from the latter results in an image where the only colored pixels are those from transparent objects, in this case, fire particles. Adding this Post Process Material to a new AirSim camera positioned at the exact point where the RGB camera enables us to generate the binary segmentation map, containing pixels where there is fire and pixels that are not colored accordingly. The complete pipeline is illustrated in Fig. fig. 16.
AUTH has used it to create 1700 RGB image - segmentation mask pairs. In our first set of experiments, we train the PIDNet [35] medium with AUW data, while at times augmenting with images from the Corsican dataset at various percentages (1%, 2%, 5%, 10%, 25%, 50%, 75%). Testing was performed on the Corsican test set, and the results are illustrated in Table 7. Results highlighted in red correspond to experiments where combining AUW with a subset of Corsican training data yields a higher mIoU / Fire IoU than using only Corsican data. Bolded entries denote the top-performing result in each experiment. When the training set consists of AUW synthetic images and more than 10% of Corsican images, the Mean Intersection over Union (MIoU) and Fire Intersection over Union (IoU) are better than the benchmark Corsican. Using both the AUW and Corsican training sets results in a 2% increase in test accuracy vs the benchmark one.
In another experiment, AUTH tested the potential of FLAME, Corsican, and AUW for style transferring and cross-dataset accuracy. To perform the style transfer, we used StyleID [36]. In table 8, we see that style transferring between the 2 real datasets did not improve the cross-dataset accuracy. However, styling either the real dataset with an image from AUW, increased the achieved MIoU. The achieved MIoU with the styled FLAME test set increased 4%, surpassing even the best mIoU that PIDNet trained on Corsican could achieve when tested on FLAME. In the case of the

This project has received funding from the European Union – European Commission under grant agreement No 101093003

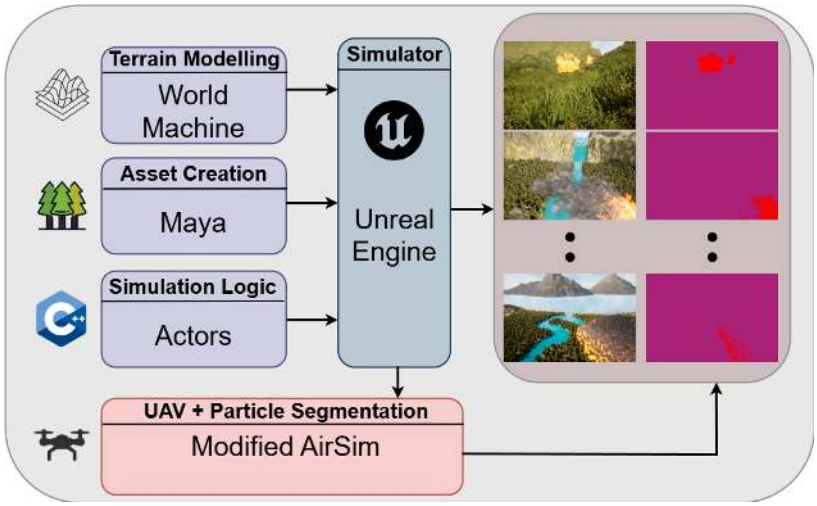


Figure 16. Our proposed UnrealFire pipeline for the creation of the annotated synthetic images using the modified version of AirSim with our particle segmentation camera.

Table 7. Results on the test set of Corsican. AUW represents our synthetic training dataset. AUW + X% represents training with our synthetic training set combined with images from the Corsican training set. 75% represents the whole Corsican training set.

Training Dataset	*Fire IoU*	*mIoU*
Corsican	86.95	91.60
AUW	45.07	65.86
AUW + 1%	78.17	86.06
AUW + 2%	81.33	87.99
AUW + 5%	83.80	89.49
AUW + 10%	87.52	91.98
AUW + 25%	87.22	91.82
AUW + 50%	87.80	92.14
AUW + 75%	**89.35**	**93.18**

styled Corsican test set, there is an 18% increase in Fire IoU, surpassing the results of training with FLAME. The ability of our synthetic dataset to improve through styling stems from the more general nature of the images, which include better-structured wildfires without any highly specific effects (e.g., saturation, focal length, etc.), and is able to adapt better to other domains.

3.3. Automatic Data Labelling using Zero shot models

As highlighted in previous sections, addressing data scarcity in natural disaster scenarios requires not only generating datasets but also ensuring that these datasets are properly labelled to maximize their usefulness. High-quality annotations are crucial for enhancing the performance of predictive models, yet the lack of labelled data remains a significant barrier to progress. To

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 8. Results on cross-dataset validation accuracy. + styling represents that the validation set was styled according to the training set.

	FLAME Validation		Corsican Validation	
Training Dataset	Fire IoU	mIoU	Fire IoU	mIoU
FLAME	83.40	91.66	58.05	69.38
Corsican	26.63	62.88	91.60	86.95
AUW	46.81	73.27	45.07	65.86
FLAME + styling	-		43.63	65.03
Corsican + styling	11.34	55.45	-	
AUW + styling	**50.82**	**75.27**	**63.79**	**77.01**

overcome this challenge, this chapter explores the use of zero-shot models as a means of automating the data labelling process. Specifically, ATOS investigated open-vocabulary detectors for fire detection and open-vocabulary segmenters for flood segmentation, both of which can accurately identify and annotate relevant features without requiring prior task-specific examples.

Zero-shot learning (ZSL) provides a promising pathway by enabling models to generalize and recognize classes they have not explicitly encountered during training. Leveraging this capability allows us to automate the labelling process and enrich the quality of datasets generated with the methods introduced in earlier chapters. Furthermore, because labelling tasks do not require real-time inference, we can prioritize models that maximize output quality over processing speed.

By integrating zero-shot models into the data preparation pipeline, ATOS reduces the dependence on costly manual annotation and enabling more scalable, efficient, and reliable disaster response systems.

3.3.1. Study of the SOTA

Since 2023, research in open-vocabulary detection (OVD) and segmentation (OVS) has advanced rapidly, driven by large-scale pretraining and promptable architectures. For open-vocabulary detection, Grounding DINO [37] has emerged as the dominant baseline. Accepted to ECCV 2024, it integrates grounding pretraining with strong detection backbones, achieving state-of-the-art results across benchmarks such as LVIS and ODinW. Its reproducibility and extensibility have been enhanced through the MMDetection [38] ecosystem, particularly via MM-Grounding-DINO, which provides standardized training and inference recipes. Together, Grounding DINO and MMDetection represent the core reference point for OVD research since 2023.

On the segmentation side, multiple frameworks have established themselves as benchmarks for open-vocabulary and promptable segmentation. The Segment Anything Model (SAM) [39] demonstrated universal, prompt-driven segmentation, and its high-quality extension, SAM-HQ [40], presented at NeurIPS 2023, has become widely used for more precise mask generation. Building on the trend of generality, SEEM (Segment Everything Everywhere, All at Once) [39] introduced a multi-modal and multi-task segmentation framework. The initial release, SEEM

This project has received funding from the European Union – European Commission under grant agreement No 101093003

vo, already showed strong zero-shot segmentation performance, while SEEM v1 added multi-object interactive segmentation, consolidating its role as a strong open-vocabulary segmentation benchmark.

Beyond these, two other frameworks frequently appear in recent comparisons. OpenSeeD [41] (ICCV 2023) presented a unified approach to both open-vocabulary detection and segmentation, simplifying pipelines by sharing representations. ODISE [42] (CVPR 2023) leveraged diffusion models to push forward open-vocabulary panoptic segmentation, achieving strong results on ADE20K and COCO Panoptic. Collectively, SEEM v1, SAM-HQ, OpenSeeD, and ODISE represent the central reference points for OVS benchmarks since late 2023.

Finally, comprehensive analyses such as the TPAMI 2024 [43] survey by Wang et al. have consolidated the landscape of OVD and OVS, systematizing taxonomies, benchmarks, and evaluation practices. In practice, researchers reporting new results since 2023 are expected to compare against Grounding DINO for detection, and SEEM v1 and SAM-HQ (often in conjunction with OpenSeeD or ODISE) for segmentation. This combination defines the state of the art in open-vocabulary vision systems.

Table 9. Summary of representative open-vocabulary detection (OVD) and segmentation (OVS) models since 2023, including task, benchmarks, highlights, and license.

Metric	Grounding DINO (ECCV 2024)[1]	MM-Grounding-DINO (MMDet)[1]	SAM-HQ (NeurIPS 2023)[1]	SEEM vo / v1 (2023)[1]	OpenSeeD (ICCV 2023)[1]	ODISE (CVPR 2023)[2]
Task	OVD	OVD	OVS	OVS	OVD + OVS	OVS
Benchmarks	LVIS, ODinW, COCO	LVIS, ODinW	COCO, ADE20K	COCO, RefCOCO, ADE20K	LVIS, COCO-Panoptic	ADE20K, COCO-Panoptic
Highlights	SOTA zero-shot detection	Reproducible config pipeline	High-quality promptable masks	Promptable, multi-object interactive segmentation	Unified detection & segmentation framework	Diffusion-aided panoptic segmentation

1. Apache 2.0: permits commercial use, modification, and distribution.
2. CC BY-NC-SA 4.0: allows non-commercial use with attribution and share-alike requirements.

For our experiments on forest fire and flood imagery, ATOS selected SEEM v1, SAM-HQ, Grounding DINO, and MM-Grounding-DINO for generating detections and segmentations. These models were chosen both for their strong zero-shot and open-vocabulary performance, which ensures robust labelling across diverse and previously unseen environmental scenarios, and for their permissive open-source licenses (Apache 2.0). This combination allows us to leverage state-of-the-art capabilities while maintaining transparency and reproducibility. By integrating these models, ATOS can achieve high-quality annotations with minimal manual effort, ensuring consistency and reliability across our dataset of natural disaster imagery.

3.3.2. Labelling Process

The open sources models for the labelling process are shown on the Table 10, Additionally we employ our current proprietary fire & smoke, and forest fire YOLOv8 models. on the fire labelling task.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 10. Summary of selected models for labelling detections and segmentations, including checkpoint, download URL, and model size.

Metric	Grounding DINO	SAM-HQ	MM-Grounding-DINO	SEEM v1	SEEM vo
Checkpoint	groundingdino_swinb _cogcoor.pth	sam_hq_vit_h.pth	mm_grounding_dino _large_o365v2_oiv6 _goldg.safetensors	seem_focall_v1.pt	seem_focall_vo.pt
Download URL	Grounding DINO releases	SAM-HQ vitH	MM-Grounding-DINO large	SEEM v1 focall	SEEM vo focall
Size	900 MB	2.6 GB	1.4 GB	1.4 GB	1.4 GB

Fire Detection Labels

To label the fire in the generated images, ATOS employs a combination of proprietary models specifically designed for fire detection and Open vocabulary models. In the initial phase, we utilize our Fire and Smoke general detector alongside our proprietary Forest Fire detector models, both of which are built upon the YOLOv8 architecture. These models are tailored to identify various fire-related features in images, ensuring a robust detection capability.

Following the initial detection process, ATOS incorporate two open vocabulary detection models, Grounding Dino and MMDET. These models enhance our labelling process by providing additional insights and improving the accuracy of the fire detection across a broader range of scenarios. Since this entire labelling process is conducted offline, we take advantage of the flexibility to implement slower and multiple models, allowing for a more comprehensive and thorough dataset labelling.

Finally, to consolidate the results obtained from each model, ATOS utilizes the repository Weighted Boxes Fusion [44]. This method allows us to effectively combine the outputs from our proprietary models and the open vocabulary detectors into a cohesive final output merging the bounding boxes from the different models based on their confidence scores. By fusing the results, ATOS ensures that the labelled dataset is both accurate and reliable, and also addressing the gaps that each model could have in their training data. This approach enhances the overall quality of the dataset while being completely autonomous, eliminating the need for human labelling and significantly increasing the efficiency of the data preparation process.

This labelling process was applied to the forest fire dataset generated in Section 3.1.3, the output labels for each image are stored in COCO format [45].The Figure 17 shows the image #520 of the forest fire dataset, in Figure 18 on it are different labelling outputs provided by each of the models. In this case all of the models properly detect most of the fires in the image, but combining the results ATOS makes sure that every fire on the image is detected as seen in Figure 19.

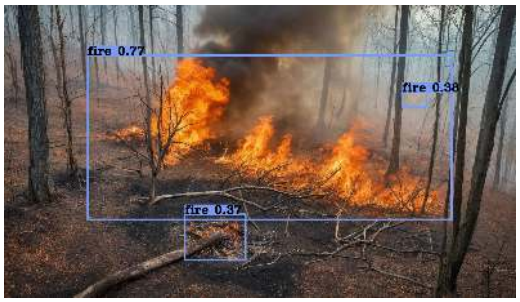
Other example of the labelling process can be seen on Figure 20 of image #4 of the same dataset, in this case each of the models detect a different subset of the fires on the images as seen on Figure 21. This case demonstrates the usefulness of the approach on filling the gaps of the detections of each of the models. The resulting fusion of labels can be seen on Figure 22.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

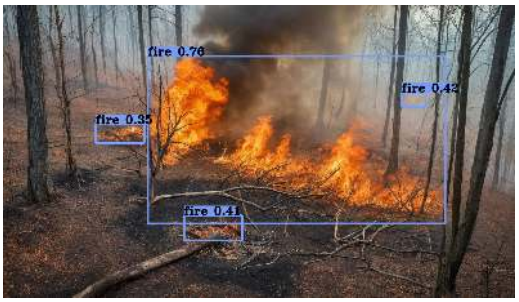
Figure 17. Image #520 of the Forest Fire dataset [image created by ATOS]

(a) Fire detections using Grounding Dino model [image created by ATOS]

(b) Fire detections using MMDET model [image created by ATOS]



(c) Fire Detections using Proprietary General Fire and Smoke Model [image created by ATOS]



(d) Fire Detections using Proprietary Forest Fire Model [image created by ATOS]

Figure 18. Fire Detections provided by the different models for Image #520 of the Dataset

This project has received funding from the European Union – European Commission under grant agreement No 101093003

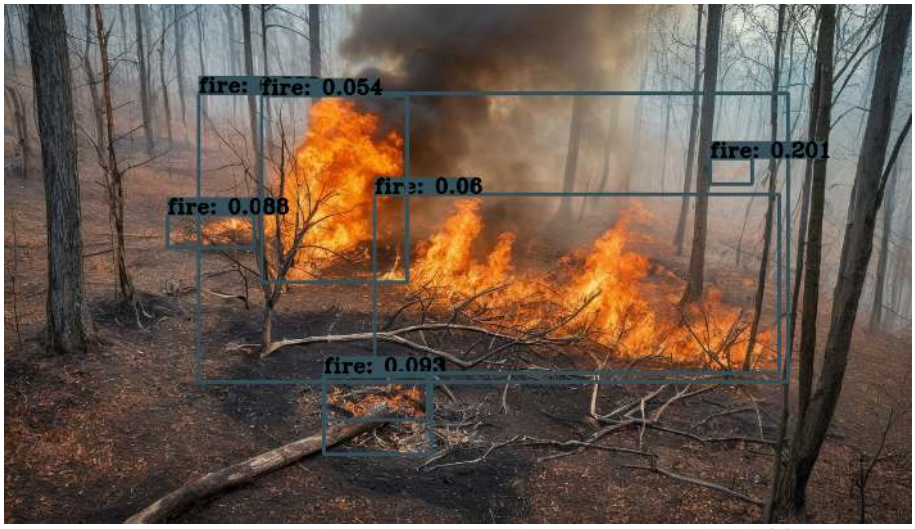
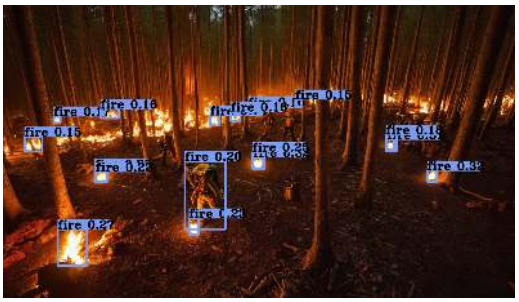


Figure 19. Fused fire detections for image #520 [image created by ATOS]

Figure 20. Image #4 of the Forest Fire dataset [image created by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Fire detections using Grounding Dino model [image created by ATOS]



(b) Fire detections using MMDET model [image created by ATOS]

(c) Fire Detections using Proprietary General Fire and Smoke Model [image created by ATOS]

(d) Fire Detections using Proprietary Forest Fire Model [image created by ATOS]

Figure 21. Fire Detections provided by the different models for Image #4 of the Dataset

Figure 22. Fused fire detections for image #4 [image created by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Flood Segmentation Masks

To generate the segmentation labelling on the generated images, ATOS employs a combination of open vocabulary models to increase the accuracy of the results by merging their outputs. The first pair of models we apply are Grounding DINO, which acts as an open vocabulary detector able to identify bounding boxes around the regions that correspond to the desired class; and SAM-HQ, which effectively segments the desired classes related to flooding around the bounding boxes. With these two models ATOS obtains an initial segmentation.

Then we apply two different versions of SEEM (vo and v1), multimodal open vocabulary detectors and segmentators. These models are able to provide segmentation masks for the desired classes. On all of the models we select the classes "water," "flood," "mud," and "river" that represent the various concepts of the floods present on the images. This multi-modal approach combined with the proper class selection permits us to generate several segmentation masks for the floods.

The final step involves merging the segmentation mask produced by each model and class to address the gaps in the outputs of each individual model. This fusion process enhances the accuracy and detail of the final segmentation mask, while reducing the influence of each model gaps. Although these open vocabulary models have higher inference times, this limitation does not affect us during the labelling process, as it is performed offline. By leveraging this collaborative approach, ATOS achieve a higher quality segmentation that is both reliable and efficient, significantly improving our ability to generate a flood scenarios dataset without the need for human intervention during the labelling process.

The labelling process for flood segmentation was applied to the flood dataset generated in Section 3.1.3, for each image a black and white image was generated, with the white region detailing the presence of flood on the image. On Figure 23 there is an image #124 from the flood dataset, an example of the process is shown. On Figure 24 are the results for each class using Grounding Dino + SAM-HQ, for all of the four classes part of the desired region is segmented. Similar outputs can be seen on both version of SEEM as shown on Figure 25 and Figure 26. The mask generated from the merger of results can be seen on Figure 27.

Another example of the flood segmentation can be seen on image #217 of the dataset Figure 28. In this case for the Grounding Dino + SAM-HQ only the class mud produced a segmentation for part of the flood on the bottom left of the image Figure 29b. The SEEM vo segmentations for all the classes cover most of the flood on the image, but leave some gaps on the small parts on the right side Figure 30. Lastly the SEEM v1 segmentation for the class water did cover the small regions on the right side, however the other classes did not produce any segmentation Figure 31. The final merged mask can be seen on Figure 32, which shows the benefits of using this approach to get a detailed and accurate segmentation.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 23. Image #124 of the Flood dataset [image created by ATOS]

(a) Flood Segmentation for the word "water". [image created by ATOS]

(b) Flood Segmentation for the word "flood". [image created by ATOS]

(c) Flood Segmentation for the word "mud". [image created by ATOS]

(d) Flood Segmentation for the word "river". [image created by ATOS]

Figure 24. Segmentation masks using Grounding Dino and SAM for the image #124 of the Dataset

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Flood Segmentation for the word "water". [image created by ATOS]

(b) Flood Segmentation for the word "flood". [image created by ATOS]

(c) Flood Segmentation for the word "mud". [image created by ATOS]

(d) Flood Segmentation for the word "river". [image created by ATOS]

Figure 25. Segmentation masks using SEEM vo for the image #124 of the Dataset

(a) Flood Segmentation for the word "water". [image created by ATOS]

(b) Flood Segmentation for the word "river". [image created by ATOS]

Figure 26. Segmentation masks using SEEM v1 for the image #124 of the Dataset

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Fused segmentation masks for Image #124 of the Flood dataset. [image created by ATOS]

(b) Segmentation mas for Image #124 of the flood dataset. [image created by ATOS]

Figure 27. Overlay of the segmentation mask and output mask for Image #124

Figure 28. Image #217 of the Flood dataset [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Flood Segmentation for the word "water". [image created by ATOS]

(b) Flood Segmentation for the word "flood". [image created by ATOS]

(c) Flood Segmentation for the word "mud". [image created by ATOS]

(d) Flood Segmentation for the word "river". [image created by ATOS]

Figure 29. Segmentation masks using Grounding Dino and SAM for the image #217 of the Dataset

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Flood Segmentation for the word "water". [image created by ATOS]

(b) Flood Segmentation for the word "flood". [image created by ATOS]

(c) Flood Segmentation for the word "mud". [image created by ATOS]

(d) Flood Segmentation for the word "river". [image created by ATOS]

Figure 30. Segmentation masks using SEEM vo for the image #124 of the Dataset

(a) Flood Segmentation for the word "water". [image created by ATOS]

(b) Flood Segmentation for the word "river". [image created by ATOS]

Figure 31. Segmentation masks using SEEM v1 for the image #217 of the Dataset

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Fused segmentation masks for Image #217 of the Flood dataset. [image created by ATOS]

(b) Segmentation mas for Image # 217 of the flood dataset. [image created by ATOS]

Figure 32. Overlay of the segmentation mask and output mask for Image #217

This project has received funding from the European Union – European Commission under grant agreement No 101093003

3.4. Handling Extreme Data Conditions in User-generated Data

As per the project's necessity to stringently connect the work carried out in T3.3 (geo-social media analytics) with T3.5 (adaptation to extreme data conditions), IT:U investigated the adaptability of the geo-social media analysis methodologies to extreme data conditions. User-generated content in disaster scenarios often exhibits extreme variability, sparsity, and noise. Addressing these challenges is critical for high-quality outputs from the TEMA GeoAI research activities. It is also crucial to provide useful information targeted to the needs of the TEMA users in the trials, as well as to future users of the TEMA platform.
To handle these issues of extreme variability, sparsity, and noise, IT:U has investigated methods that explicitly leverage the latent semantic structure, spatiotemporal context, and cross-platform diversity of social media posts, enabling a more nuanced understanding, extraction and prediction of relevant disaster information.

3.4.1. Study of the SOTA

User-generated data, particularly from social media platforms, presents unique challenges for disaster management due to its inherent heterogeneity, noise, and extreme sparsity in certain contexts. State-of-the-art methods for analysing such data have leveraged a combination of textual, visual, spatial, and temporal information to enhance situational awareness [46].
A range of scholarly works have utilized topic modelling to reveal latent thematic structures within large volumes of user-generated texts (e.g., social media data) [47, 48]. However, the resulting topic representations are rarely contextualized with additional data modalities. Early efforts to analyze geo-social textual media in a multimodal setting typically employed sequential workflows, wherein each modality was processed in isolation through distinct modules [49]. While these methods have demonstrated utility in various applications, their modular and linear design often hinders the ability to capture complex interdependencies between modalities. Furthermore, this sequential structure can introduce dependencies between processing stages, rendering the overall output sensitive to the order in which analyses are performed. More recent approaches introduced advancements toward integrating multiple modalities such as time, geographic space, semantic topics, and sentiments in the analysis of social media [50]. However, they still depend on numerous isolated components and fragmented intermediate processing steps, falling short of delivering a truly end-to-end processing methodology. Nonetheless, the inherent heterogeneity of the input modalities adds to the complexity of the task: text is symbolic and sequential, while location data is numeric and geospatial, often resulting in misaligned feature representations and inconsistencies during fusion. The lack of a unified machine learning model constrains cross-modal interactions and introduces fragile interdependencies, whereby variations in preprocessing or processing order can significantly impact both performance and interpretability.
Relevance classification has also emerged as a critical step in disaster-related geo-social media analysis, aiming to distinguish informative posts from noise [51]. Traditional approaches rely on keyword-based filtering [52, 53] or unsupervised topic modelling [54, 55], while more recent methods use supervised deep learning models such as Convolutional Neural Networks (CNNs), transformer architectures including Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimised BERT Pre-training Approach (RoBERTa), and graph-based neural networks [56, 57, 58, 59, 60]. Multimodal approaches that integrate textual and

This project has received funding from the European Union – European Commission under grant agreement No 101093003

visual content have further improved the accuracy of relevance and informativeness classification, particularly in cases where text alone may be ambiguous or sparse [58, 61, 60]. These techniques, however, typically consider spatial and temporal context only as post-processing or secondary features, limiting their ability to fully capture the dynamics of disaster-related user-generated content.

In parallel, while geo-social media has become an important tool for disaster management, research has overwhelmingly focused on Twitter, with far fewer studies examining other social media platforms such as TikTok, Instagram, Facebook, or Telegram [62, 63, e.g.]. Existing work on these alternative platforms typically relies on small datasets, qualitative methods, or manual annotation, often overlooking spatial or temporal dynamics [64]. Consequently, there remains a significant gap in understanding how non-Twitter platforms can support disaster response and recovery at scale.

Furthermore, recent research has increasingly emphasized the need to contextualize geo-social media analyses within physically meaningful spatial frameworks. Most studies on flood-related geo-social media have relied on administrative boundaries, which only loosely align with hydrological processes and thus limit spatial interpretability. For instance, Havas and Resch demonstrated that integrating semantic features with spatial and temporal context enhances the interpretability of social media data and supports situational awareness during disasters [65]. Further, Hanny and Resch introduced a unified multimodal framework that jointly embeds textual and geospatial information, enabling the identification of spatially coherent and semantically meaningful clusters in disaster-related content [66]. Yet, these approaches generally remained confined to national or urban scales and did not address transboundary dynamics or the alignment of online discourse with physical flood basins.

IT:U has addressed the above limitations by integrating spatiotemporal context with textual features in a unified graph-based framework, enabling the joint analysis of semantic, spatial, and temporal dependencies in disaster-related online communication. In addition, IT:U developed methods for collecting and analysing disaster-related data across multiple social media platforms, improving the handling of heterogeneous user-generated content. IT:U also introduced a watershed-based analytical framework that aligns online discourse with the physical processes of the 2021 Central European floods by linking BERTopic-derived semantic clusters with hydrological and socio-environmental indicators. The research further explores substructures within relevance classes and semantic topics, revealing distinctions in the informativeness and actionability of posts.

3.4.2. Graph-based Learning for Social Media Data

Previous research has emphasised the significant role of geo-location data in social media analysis, particularly in understanding and monitoring large-scale phenomena like natural disasters through geo-referenced content from micro-blogging platforms [65]. While semantic topic modelling has been widely adopted to uncover latent thematic structures in large volumes of user-generated text, these methods often lack integration with additional data modalities such as geo-location. Early multimodal approaches employed modular, sequential workflows, processing each modality independently through distinct analysis stages. While these workflows have been useful, they often fail to capture the complex interdependencies between modalities and are highly sensitive to the order in which each modality is processed, limiting their ability to offer a truly unified, end-to-end solution [49].

To overcome these limitations, IT:U investigated a novel methodology for multimodal geospa-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

tial machine learning that jointly embeds both semantic and geospatial information in a unified graph-based space. This approach employs end-to-end learning with a composite loss function to generate clusters that are both semantically coherent and spatially interpretable. The model constructs a unified multimodal embedding space where diverse data types, including text, geolocation, and optional temporal or emotional signals, are integrated into a shared representation. Nodes in the graph represent embeddings of user-generated posts, while edges capture relational signals such as semantic similarity, geographic proximity, and user connections. This design allows the model to process multimodal interactions in a context-aware manner, ensuring that the relationships between the different data types are effectively captured.
At the heart of IT:U's methodology is an unsupervised artificial neural network that aggregates information from neighbouring nodes via a modified message-passing mechanism, with the aggregation process scaled according to edge weights. A key innovation of our approach is the composite loss function, which is inspired by reference-less clustering evaluation techniques [66]. This loss function integrates three objectivescontrastive, coherence, and alignment. The contrastive component ensures that semantically similar nodes are pulled closer in the embedding space, while dissimilar ones are pushed apart. The coherence objective encourages the formation of well-defined clusters by maximising intra-cluster similarity and inter-cluster separation. The alignment term stabilises the learning process by ensuring that each node remains close to the centroid of its assigned cluster, thereby enhancing representational consistency. By incorporating these objectives directly into the learning process, the model is able to discover clusters that are not only semantically meaningful but also spatially coherent. The approach outperforms existing baselines in terms of topic coherence, spatial compactness, and interpretability, providing a scalable and principled framework for multimodal social media analysis. Moreover, the geographically fine-grained multimodal clusters generated by the model align closely with observed disaster impacts, showcasing the practical value of this methodology in real-world applications.

3.4.3. Data Acquisition from Heterogeneous Social Media Platforms

IT:U explored diverse social media data sources to address the growing fragmentation of online platforms and the need for multi-social media platform integration in disaster management. While X (Twitter) has historically dominated geo-social media research, amongst others due to its openly accessible Application Programming Interfaces (APIs) [67], recent restrictions on data access have made it necessary to investigate alternatives. To address this reduced data availability, IT:U has developed a cross-platform framework incorporating Bluesky, TikTok, Reddit, Telegram, and Mastodon to evaluate their suitability for disaster-related situational awareness. This work directly responds to three key research gaps: the limited assessment of multilingual keyword taxonomies for consistent crawling across heterogeneous platforms, the unclear platform-specific differences in content relevance and geospatial granularity, and the underexplored potential of cross-platform integration for generating better information for disaster management.
The social media platforms differ substantially in both crawling requirements and geotagging capabilities. TikTok provides a research API with rate limitations, Reddit is primarily accessed through archival datasets [68, 69], Telegram requires explicit following of channels for crawling [70], Mastodon demands distributed queries across decentralised instances [71], and Bluesky relies on the federated AT Protocol [72]. Geotagging support is similarly uneven: TikTok allows manual location tagging of videos [73], Telegram enables static and live geolocation sharing [70], and Mastodon supports geotags through the ActivityPub protocol [71], whereas Reddit

This project has received funding from the European Union – European Commission under grant agreement No 101093003

and Bluesky lack native geotagging, requiring reliance on geoparsing [68, 72, 74]. Accordingly, IT:U reseaching on a combination of keyword-based queries, crawling strategies, and text-based geoparsing to extract location mentions.
This heterogeneous landscape illustrates the methodological challenges of collecting and harmonising user-generated content across structurally diverse platforms, while also pointing to opportunities for data fusion for identifying signals to enhance the robustness of disaster monitoring.

3.4.4. Analysing Alternative Geo-social Media Data Sources

As an initial step toward the cross-platform data acquisition framework described above, IT:U first conducted an empirical study to evaluate the suitability of alternative social media platforms for disaster-related GeoAI research. This early work provided the conceptual and methodological basis for the later multi-platform integration approach. The study presented in [64] analysed four platforms - TikTok, Reddit, Telegram, and Mastodon - by developing workflows to extract and geoparse posts and comparing their spatial and temporal patterns to established Twitter data. For this, Hurricane Ian was used as a case study. Figure 33 illustrates the data collection and processing workflow employed in the study.

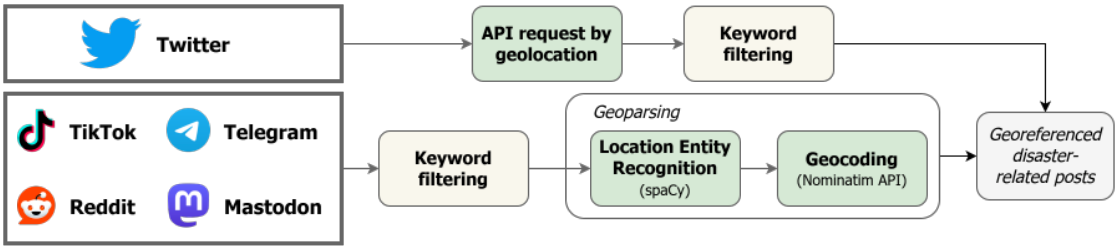


Figure 33. Overview of study workflow. The logos of the social media platforms are taken from Wikimedia Commons.

Data collection for Hurricane Ian spanned September 20 to October 10, 2022, with Twitter serving as the benchmark due to its historical dominance in geo-social media research. TikTok provided 168,761 posts via its Research API, though data access was limited by rate restrictions, delayed availability, and exclusion of videos and most comments. Reddit data was retrieved from archival dumps, yielding 31,187 submissions and 169,861 comments filtered by hurricane-related content. Telegram and Mastodon, accessed through snowball crawling and API queries respectively, produced very large, highly noisy datasets (over 50 million Telegram messages and more than 1 million Mastodon posts), though both platforms posed challenges due to chat-based structures and limited filtering options. Since none of the platforms provided explicitly geotagged posts, a geoparsing approach was applied to extract location mentions from text content via spaCy and subsequently geocode them via Nominatim, an open-source geocoder based on OpenStreetMap (OSM). This further reduced dataset sizes, often returning coarse or inaccurate locations, though most posts still concentrated on Florida and other southeastern U.S. states.
The results showed that while no single platform can fully replace Twitter due to limited spatial accuracy and real-time coverage, TikTok and Reddit in particular can provide valuable complementary data. Correlation analyses showed that TikTok and Reddit data were most similar to Twitters spatial distribution (see Figure 34), while Telegram and Mastodon had weaker coverage and lower overlap. Temporally, all platforms peaked in activity on September 28-29, aligning

This project has received funding from the European Union – European Commission under grant agreement No 101093003

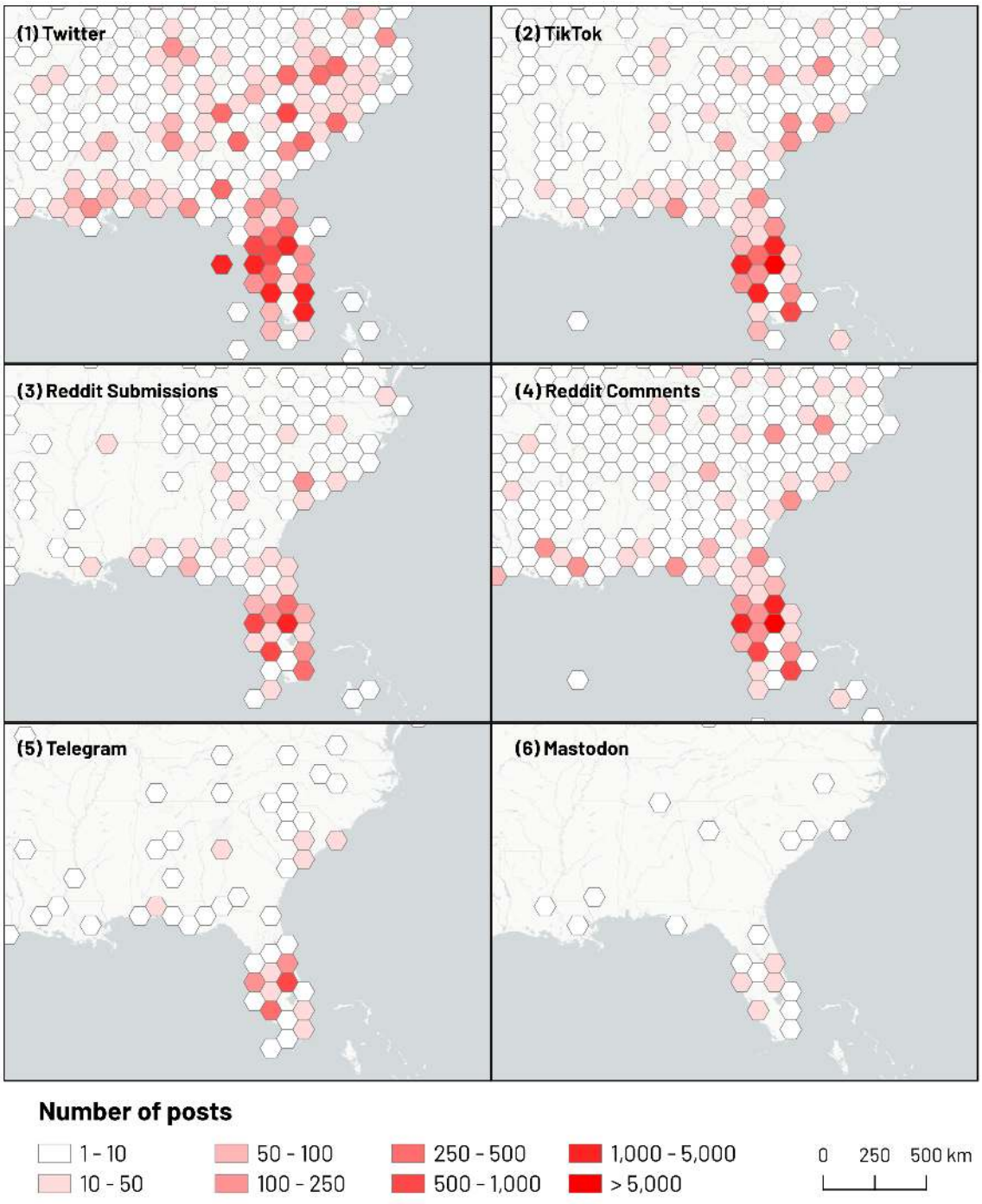


Figure 34. Comparison of spatial distribution of georeferenced posts on 100km hexagonal grid.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

with the hurricane's landfall, with high cross-platform correlations confirming relatively consistent disaster-related posting patterns. The study thus shows that combining multiple platforms can improve situational awareness in disaster management and calls for the development of more accurate geoparsing methods to make non-geotagged platforms more useful for future research.

3.4.5. Spatial Context-dependent Topic Modelling

To support TEMAs objectives of improving the trustworthiness, accuracy, and responsiveness of extreme data analytics, IT:U explored how spatially contextualised topic modelling can enhance the interpretation of social media data during disasters. Traditional topic modelling approaches often ignore spatial context, causing semantically similar discussions from different locations or phases of an event to be mixed and reducing interpretability. By linking semantic information to hydrological and environmental conditions, the approach yields faster, more reliable insights into public reactions and needs throughout flood events. Specifically, IT:U processed more than 14,400 georeferenced tweets through a pipeline of language translation, disaster-related classification, and machine-learning-based topic modelling (BERTopic). This enabled the extraction of stable, flood-relevant themes such as Heavy Rain, Damage, Help to Victims, and Climate Crisis, and allowed us to trace their temporal evolution across varying environmental and societal contexts. Figure 35 illustrates how dominant topics shifted across regions, reflecting both the immediate physical impacts of the floods and the subsequent social responses.
By aligning these semantic patterns with the timing of the flood event and with watershed-level characteristics, we were able to systematically reduce noise and highlight meaningful spatiotemporal structures in the data. This integration revealed where conversations were driven by acute environmental stressors, such as heavy rainfall or infrastructure disruption, and where they instead reflected post-disaster solidarity or broader political discourse. Figure 36 further demonstrates how topic distributions varied with hydrological and socio-environmental factors, underscoring the value of contextualisation in making sense of highly variable social media data. The results demonstrate how immediate, event-driven topics closely aligned with rainfall and flood extent in upstream regions, while post-disaster topics reflecting solidarity and volunteering arose mainly in downstream or less-affected areas. Basin-specific topics such as Rhine Flood or Meuse Flood persisted across borders, while broader political or climate discussions concentrated in urban downstream environments. This shows how integrating semantic and spatial analysis helps to disentangle noisy user-generated data streams and translate them into actionable insights.
In doing so, the study illustrates how extreme variability in social media data can be handled not by suppressing it, but by contextualising it-linking digital traces to hydrological and socio-environmental conditions. This methodological advance provides a pathway toward more robust monitoring and coordination during extreme weather events, particularly in transboundary river systems where risks and responses are inherently interconnected.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

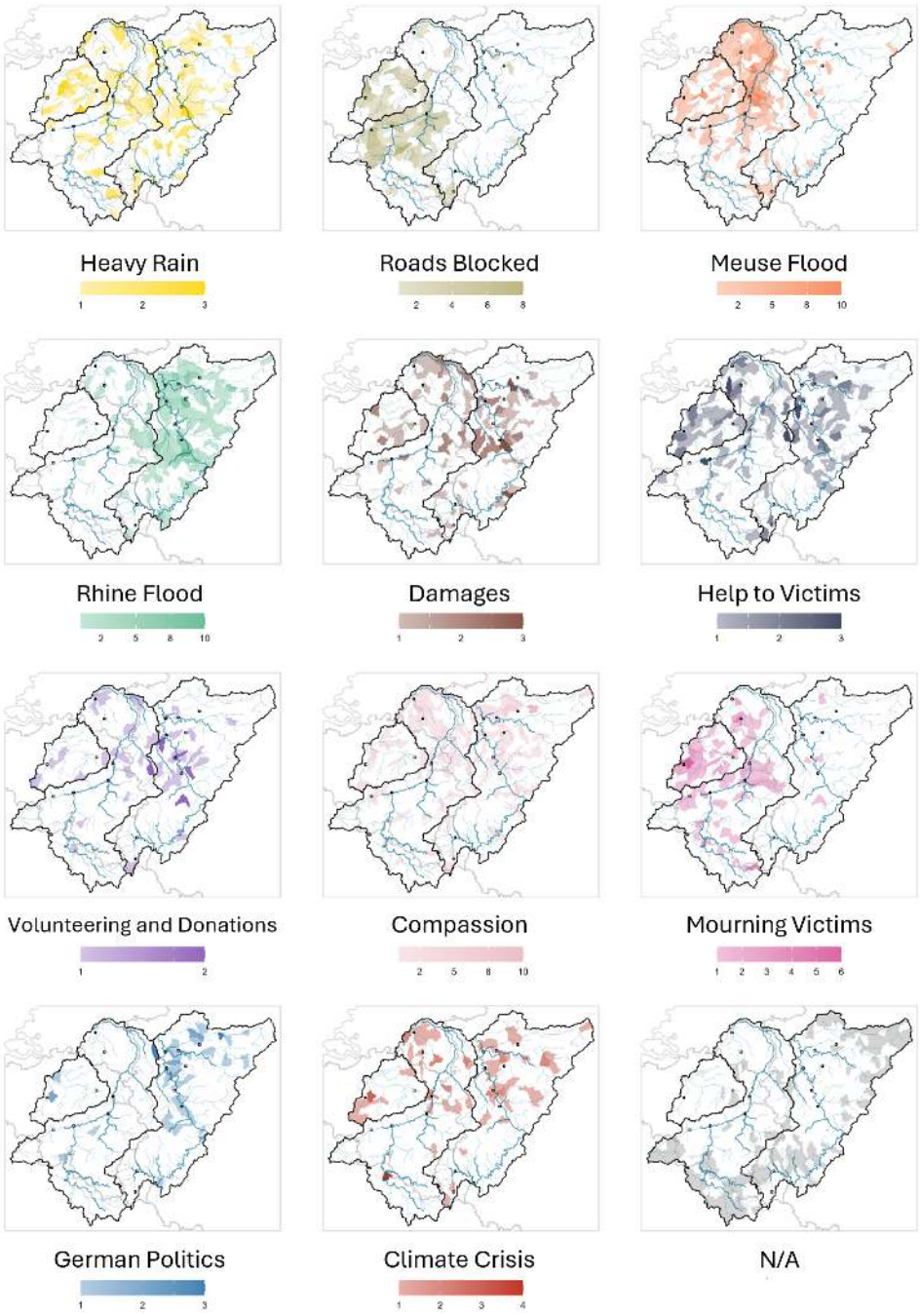


Figure 35. Dominant topics per region

This project has received funding from the European Union – European Commission under grant agreement No 101093003

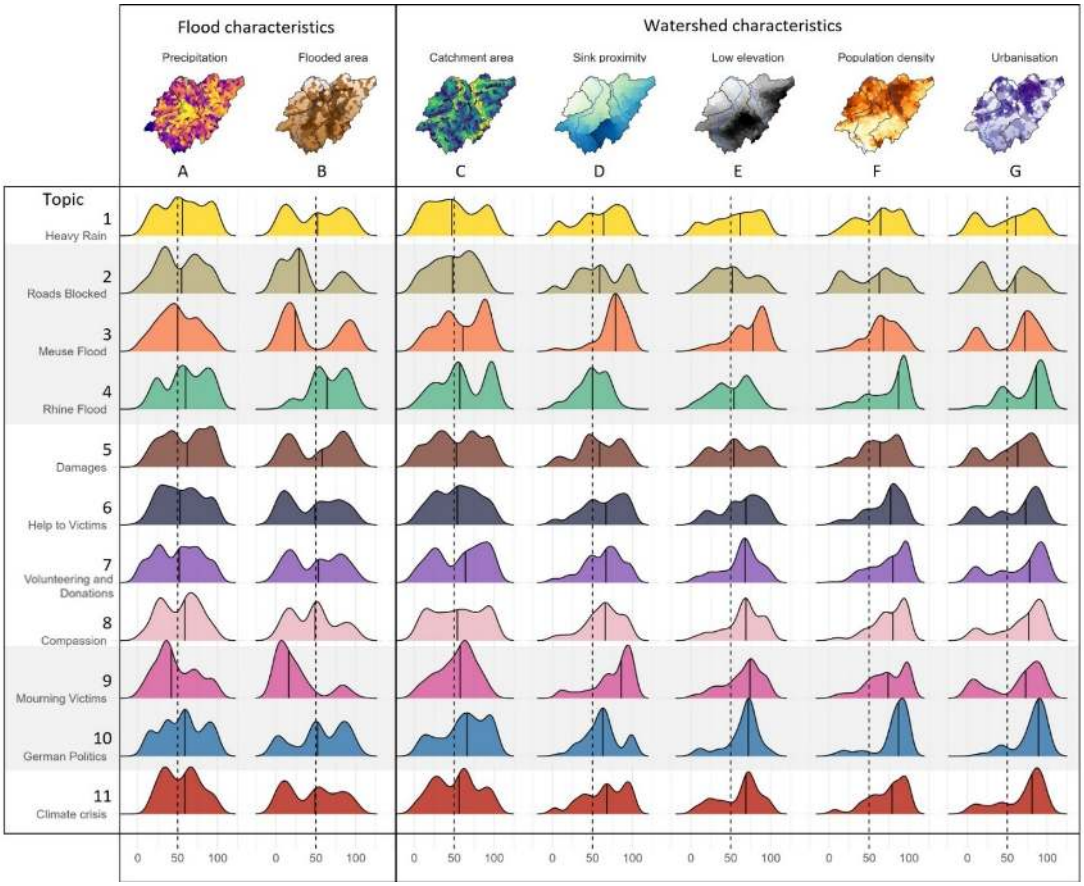
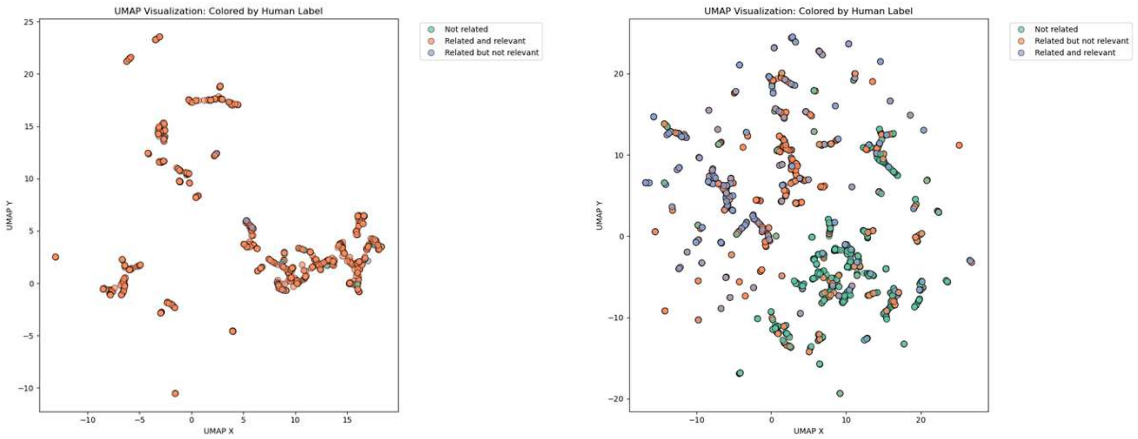


Figure 36. Geo-social media topic occurrences across different societal and environmental contexts.

3.4.6. Substructures of Relevant Disaster Content

Building on the state-of-the-art, IT:U has further advanced methods for assessing the relevance of disaster-related social media content, which is often noisy, multilingual, and unevenly distributed. The focus was on uncovering the underlying semantic structure of such data to better distinguish relevant from irrelevant information and support more accurate event understanding. Using a multilingual, georeferenced dataset spanning five major disasters, IT:U generated multiple embedding configurations with TwHIN-BERT-base [75] and applied diverse pooling and dimensionality-reduction strategies. Embedding diagnostics, measured via cosine similarity and the Hopkins statistic [76], revealed that CLS-token-pooled embeddings with Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction [77] exhibited the strongest clusterability, reflecting a coherent semantic organisation suitable for downstream analysis. These configurations enabled the application of seven complementary clustering algorithms, uncovering both *pure* clusters dominated by a single relevance label and *mixed* clusters containing semantically similar posts spanning multiple labels (Fig. 37). This dual pattern illustrates how conventional relevance classes may obscure finer distinctions in content, highlighting the value of unsupervised latent structure analysis in extreme, sparse data settings.
Additionally, IT:U introduced a second-stage subtopic induction pipeline that combines BERTopic

This project has received funding from the European Union – European Commission under grant agreement No 101093003



(a) Pure cluster: dominated by a single label, with subtle internal variation.

(b) Mixed cluster: label boundaries less distinct, with cross-label semantic groupings.

Figure 37. 2D UMAP visualisations of two clusters from the best-performing configuration (CLS+UMAP + spectral, $k = 5$) colored by human annotations. These plots illustrate our motivation for applying BERTopic in the next step to uncover coherent subtopics within each cluster.

with Generative Pre-trained Transformer (GPT)-based labelling to produce interpretable, human-readable summaries of cluster substructures. This workflow revealed coherent semantic subtopics within both pure and mixed clusters, capturing nuanced differences in informativeness and actionability that standard relevance labels fail to distinguish. In parallel, our experiments with spatiotemporal decay transformations demonstrated that inverse temporal decay applied to proximity features improves relevance prediction metrics, including macro F1 and accuracy. Collectively, these methodological advances provide a robust framework for processing heterogeneous, noisy, and sparse user-generated data, enabling more granular semantic interpretation, enhanced predictive performance, and better handling of extreme data conditions in disaster scenarios.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4. Explainable AI Methods for Extreme Data Conditions

4.1. XAI of Diffusion models as Ground Truth

As discussed on 3.1, diffusion models generate highly realistic images through the iterative denoising of random noise. These models have demonstrated strong potential for creating synthetic datasets for natural disasters, such as forest fire detection.

A key requirement for their practical adoption is explainability: the ability to interpret the mechanisms behind image generation and to identify where and how specific concepts (e.g., fire, smoke, flood) are represented in the outputs. In this context, explainability acts as a ground truth provider, automatically localizing objects within generated images. Because diffusion models rely on well-defined token-to-visual mappings, explainability techniques can extract bounding boxes or segmentation masks directly from the generative process, reducing the need for manual annotation.

4.1.1. Study of the SOTA

Explainability on Diffusion models SOTA early 2024

In 2023, explainability for text-to-image diffusion models advanced significantly with the introduction of DAAM (Diffusion Attentive Attribution Maps). DAAM transforms cross-attention signals into token-level heatmaps, localizing where each prompt word is represented in the image. By aggregating and upsampling attention across the U-Net denoising process, DAAM produces word-region attributions competitive with segmentation-style probes.
Integrated seamlessly with Stable Diffusion, DAAM serves as a practical baseline for prompt explainability. For wildfire imagery, it allows the automatic extraction of masks or bounding boxes for the token fire, enabling dataset annotation with minimal human effort. Importantly, cross-attention maps stabilize early in the sampling process, meaning that single-step or lightly aggregated maps are sufficient for reliable results [78].

Alongside DAAM, 2023 early 2024 saw a wave of cross-attention based control mechanisms that also provide explainability signals. GLIGEN incorporates explicit grounding (text plus bounding boxes) into pre-trained diffusion backbones, while training-free layout guidance methods bias cross-attention during sampling to place tokens within user-specified boxes. Both approaches reveal strong wordregion alignment in attention maps [79]. Attend-and-Excite employs attention interventions to correct missing objects and improve attribute binding, again relying on the tokenpixel channels visualized by DAAM [80]. In practice, these tools form a spectrum: DAAM provides attribution and bounding boxes from a plain prompt (fire); GLIGEN or layout guidance enables explicit placement of objects; and attention-control methods are most useful for complex prompts or underrepresented objects. Collectively, they represent the state-of-the-art toolkit for generating wildfire datasets and extracting reliable fire bounding boxes directly from the gener-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

ative process.

Table 11. Comparison of Explainability Methods for Diffusion Models: 2023 to Early 2024

Metric	DAAM	GLIGEN	Attend-and-Excite
Explainability Approach	Cross-attention heatmaps	Text + bounding box grounding	Attention interventions during sampling
Strengths	Direct token-to-region mapping; simple, training-free	Explicit control over object placement; strong word-region alignment	Fixes missing objects; improves attribute binding; interpretable via attention channels
Use Case for Wildfire Imagery	Extract masks/bounding boxes for "fire" and "smoke"; baseline attribution	Place fire explicitly in a scene; useful for layout-controlled datasets	Ensure accurate generation of fire regions in complex prompts; improve coverage of underrepresented objects

Among the three approaches, DAAM stands out as the most suitable method for ATOS purpose of generating synthetic wildfire datasets. Its main advantages are:

1. Training-free and lightweight: DAAM does not require additional training or architecture modifications, which simplifies integration with pre-trained Stable Diffusion models.
2. Direct explainability: By converting cross-attention into token-level heatmaps, DAAM directly identifies regions corresponding to fire or smoke. This makes it possible to extract masks or bounding boxes without manual annotation.
3. Sufficient precision: Cross-attention maps stabilize early during sampling, so single-step or lightly aggregated maps are accurate enough for dataset generation, reducing computational overhead.
4. Scalable for multiple objects: DAAM can handle multiple simple tokens (like fire and smoke) in the same prompt, providing separate bounding boxes for each concept, which is crucial for building diverse datasets.

While GLIGEN and Attend-and-Excite provide stronger control over layout, this is less critical for our task, where the goal is annotating generated objects for detection models rather than enforcing composition. Thus, DAAM is the most efficient and reliable method for extracting ground-truth annotations from synthetic wildfire imagery.

Advances in Explainability of Diffusion Models SOTA

Explainability methods continued to evolve through 2024, focusing on enhanced visualization, interpretability, and control of cross-attention mechanisms. These advancements improved transparency in image generation and allowed closer inspection of how textual prompts map onto visual features.

Among the current tools, attention-map-diffusers [81] stands out for its practical approach to cross-attention visualization. Built on Hugging Faces Diffusers library (v0.32.0), it enables detailed inspection of how individual tokens influence different regions of generated images. Its

This project has received funding from the European Union – European Commission under grant agreement No 101093003

support for multiple models including Stable Diffusion 3.5, Flux-dev, and Flux-schnellcombined with batch processing and the ability to save attention maps at specific layers and timesteps, makes it particularly well-suited for tasks requiring precise localization of generated features.

Conversely, Diffusers-Interpret [82] enhances the Diffusers framework by offering a comprehensive, step-by-step analysis of the diffusion process. While it also visualizes attention maps, its primary focus is on interpreting the contributions of different model components to the final output. This makes Diffusers-Interpret especially valuable for researchers seeking a more holistic understanding of model behavior, rather than extracting spatially precise features.

Self-Attention Diffusion Guidance [83] takes a different approach by using attention maps to actively guide the generation process. By leveraging self-attention signals to influence the diffusion steps, it aims to improve the fidelity and coherence of generated images. Unlike attention-map-diffusers or Diffusers-Interpret, which are primarily diagnostic, this method directly intervenes in generation, making it particularly useful when high-quality synthesis is critical.

Finally, InterpretDiffusion [84] introduces architectural modifications to allow the model to self-discover interpretable components within the diffusion process. By identifying and visualizing meaningful structures, it provides insight into the models internal decision-making, offering a complementary perspective to attention-based analysis tools. While less focused on token-level spatial mapping, it excels in revealing high-level structural reasoning within the model.

Table 12. Comparison of Explainability Tools for Diffusion Models SOTA

Metric	Attention-Map-Diffusers	Diffusers-Interpret	Self-Attention Diffusion Guidance	InterpretDiffusion
Focus	Cross-attention visualization	Model interpretability	Sample quality enhancement	Self-discovery of interpretable components
Key Features	Batch operations, timestep/layer specific maps, model compatibility	Step-by-step diffusion analysis, attention visualization	Utilizes self-attention maps for guidance	Architecture adaptation for interpretability
Use Case in Wildfire Detection	Extracting bounding boxes for "fire" and "smoke" tokens	Understanding diffusion process for synthetic dataset generation	Improving fidelity in generated wildfire images	Identifying meaningful structures in wildfire imagery

Among the available explainability solutions, Attention-Map-Diffusers emerges as the most suitable choice for ATOS wildfire detection dataset generation. Built on Hugging Faces Diffusers library (v0.32.0), it ensures compatibility with the latest features and optimizations. Its support for multiple models including Stable Diffusion 3.5, Flux-dev, Flux-schnell, and SDXLprovides flexibility in model selection, which is particularly relevant for our work as discussed in previous chapters. Batch processing capabilities allow for efficient handling of multiple images simultaneously.

A key advantage of Attention-Map-Diffusers is its ability to extract attention maps at specific layers and timesteps, offering detailed insights into the spatial alignment between textual tokens and generated image regions. Additionally, its design emphasizes ease of integration with existing workflows, minimizing setup time and complexity. These features collectively make

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Attention-Map-Diffusers especially effective for extracting precise bounding boxes for fire and smoke tokens, directly supporting the creation of annotated datasets for training detection models.

In the broader context of diffusion model explainability, approaches vary in their focus: Attention-Map-Diffusers provides fine-grained token-to-region attribution suitable for localization tasks, Diffusers-Interpret offers stepwise component-level analysis, Self-Attention Diffusion Guidance improves image quality via attention-driven interventions, and InterpretDiffusion uncovers high-level interpretable structures. For our specific application generating wildfire datasets with accurate fire and smoke annotations using Flux and SDXL model families Attention-Map-Diffusers stands out due to its combination of spatial precision, model compatibility, and straightforward integration.

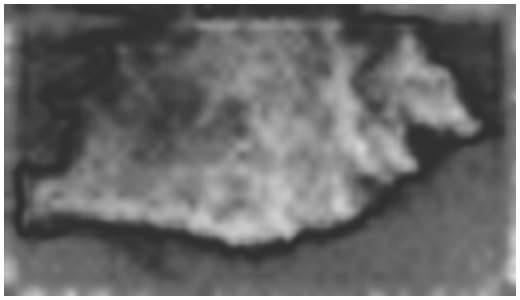
4.1.2. Explanations on Diffusion Models

Explainability on SDXL using DAAM

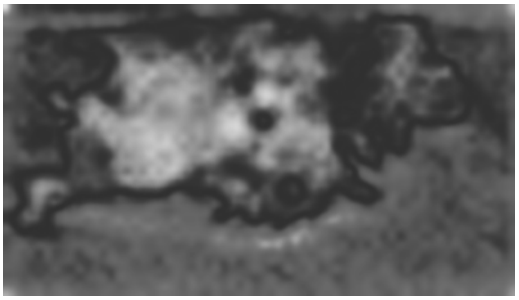
DAAM is compatible with ComfyUI, allowing for seamless integration into the generation pipeline detailed in the previous chapter. To accomplish this integration we changed the Sampler and Clip Text encode blocks with DAAM compatible ones, which provide the heatmaps per tokens generated through the diffusion, Also it is added the DAAM Analyzer block which provide the final explanation per token (Detailed pipeline on Appendix I). Utilizing these pipelines, we can extract explanations for specific tokens, such as 'fire' and 'smoke,' in the context of forest fires. An example is illustrated in Figures fig. 38 and Figure 39, where lighter regions on the heatmap indicate greater relevance of the corresponding token during the diffusion process.

Figure 38. Generated image using SDXL and DAAM explanations with the prompt: "realistic forest fire with rising smoke and flames as seen from the air" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003



(a) Explanation heatmap for the token "fire" on the diffusion steps [image generated by ATOS]



(b) Explanation heatmap for the token "smoke" on the diffusion steps [image generated by ATOS]

Figure 39. Generated explanations using DAAM and SDXL

With the explanation heatmaps ATOS is able to postprocess this images in order to filter out the noise and extract the regions with most impact for the corresponding token and the resulting bounding box. The first step is to suppress high-frequency noise and enhance the robustness of segmentation, a Gaussian blur is applied. This step ensures smoother intensity transitions and mitigates the influence of spurious variations in pixel values. After the noise reduction, the image is transformed into a one-dimensional vector representation of pixel intensities, to then do a K-means clustering with K=2, under the assumption that the image can be decomposed into foreground and background regions. Upon convergence, the cluster assignments are reshaped into the original spatial dimensions, resulting in a preliminary segmentation of the image. The brighter cluster is selected as the foreground, and a binary mask is generated. To refine this mask, morphological operations are employed. A closing operation is utilized to eliminate small holes within the detected regions, and an opening operation removes isolated noise artifacts. Afterwards, a connected component analysis is performed, and regions below a minimum area thresholds are discarded to ensure that only significant regions are kept. Finally Contour extraction is subsequently applied to the refined mask, enabling the computation of a bounding box that encloses all detected regions. The detailed algorithm is in appendix J. The resulting bounding boxes for the previous example can be seen on Figure 40 and Figure 41 for the tokens fire and smoke correspondingly.

(a) Resulting Segmentation for the token "fire" [image created by ATOS]

(b) Resulting Bounding box for the token "fire" [image created by ATOS]

Figure 40. Generated segmentations and bbox for the token "fire"

Using this explanation the fire segmentation encompasses all the flame spots on the generated image, though it also takes part of the smoke, the resulting bounding box encompasses the fire properly if it was considered just one object in the image. The results for the smoke are more

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Resulting Segmentation for the token "smoke" [image created by ATOS]

(b) Resulting Bounding box for the token "smoke" [image created by ATOS]

Figure 41. Generated segmentations and bbox for the token "smoke"

lackluster, the segmentation only covers part of the smoke on the image, and accordingly the resulting bounding box is not useful as ground truth.

The application of DAAM within SDXL generation pipelines effectively elucidates which parts of the image each token influences during generation. However, the resulting segmentations and bounding boxes currently lack the detail and utility to be considered reliable ground truth for synthetic images. Nevertheless, further advancements in this area of explainability could yield the desired results in the near future.

Explainability on Flux.1-dev using Attn-Map-Diffussers

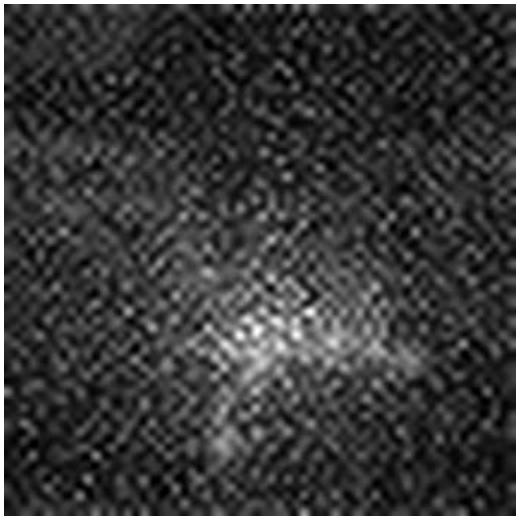
Attn-Map-Diffusers is incompatible with ComfyUI, preventing straightforward integration into the existing pipelines outlined in the previous chapter. The benefits is that ATOS can apply explainability to recent models like Flux.1-dev which provide a higher quality and fidelity of the generated images. To generate the images ATOS created a script where Flux pipeline is defined, the first step is to load the text encoder (T_5EncoderModel) that handles prompt tokenization and representation. Then load the Flux transformer model that would take care of the image generation. ATOS defines the prompt for the image "A fire in a forest with rising smoke as seen from a drone". We generate the image using the pipeline and finally save the attention maps of all the tokens on the input prompt. The resulting image is displayed in Figure Figure 42, while the explanation attention maps for the tokens 'fire' and 'smoke' are shown in Figure 43. In these maps, brighter spots indicate higher attention from the model during the diffusion steps for the respective tokens.

ATOS applied the same algorithm to extract the segmentation and bounding boxes from the previous section (Appendix J) on the generated attention maps for "fire" and "smoke". The results can be seen on fig. 44 for fire, and Figure 45 for the smoke. There is an improvement on the segmentations for the concepts in comparison to the DAAM explanations; the regions for both the fire and smoke are more localized to the actual position on the image where they are present. These results could be used as a ground truth for synthetic images with the only caveat that the prompts that allow this explanations need to be brief and concise in order to get understandable explanations. For longer prompts the outcome of the attention maps are disperse and does not provide useful information.

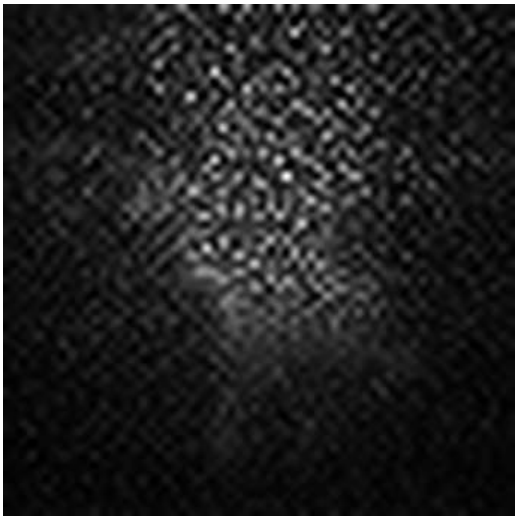
This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 42. Generated image using Flux.1-dev and Attn-Map-Diffussers explanations with the prompt: "A fire in a forest with rising smoke as seen from a drone" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003



(a) Explanation heatmap for the token "fire" on the diffusion steps [image created by ATOS]



(b) Explanation heatmap for the token "smoke" on the diffusion steps [image created by ATOS]

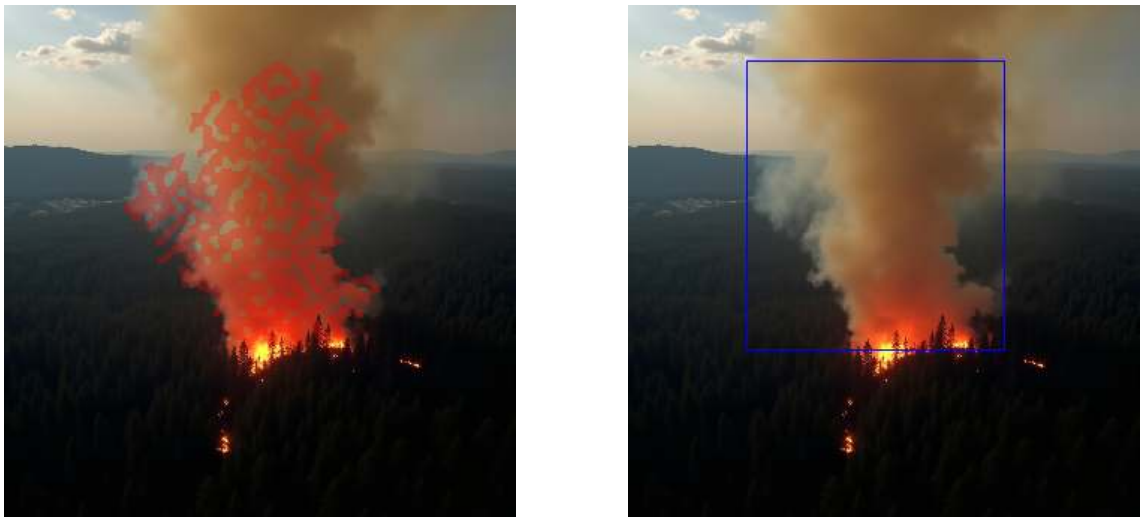
Figure 43. Generated explanations using Flux.1-dev and Attn-Map-Diffussers

(a) Resulting Segmentation for the token "fire" [image created by ATOS]

(b) Resulting Bounding box for the token "fire" [image created by ATOS]

Figure 44. Generated segmentations and bbox for the token "fire"

This project has received funding from the European Union – European Commission under grant agreement No 101093003



(a) Resulting Segmentation for the token "smoke" [image created by ATOS]

(b) Resulting Bounding box for the token "smoke" [image created by ATOS]

Figure 45. Generated segmentations and bbox for the token "smoke"

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4.2. Generic XAI Methods for Extreme Data Conditions

Data available in natural disaster management is often characterized by extreme data conditions, such as visual data scarcity, covariate shift, underrepresented features or biases, mislabeled samples, posing significant challenges for real-time and trustworthy explainable data analytics. To tackle those challenges, FHHI has developed various generic XAI methods that explicitly leverage second-order uncertainty explanations, explanation-guided regularization, dual data attribution using multiclass kernel SVMs as surrogates for deep learning models, along with evaluating the cognitive load of explanation visualizations for end-users in critical scenarios, expending upon its previous work from Tasks T3.1 and T3.3, which was already described in Deliverables D3.1 and D3.2. Hence in this Deliverable D3.4 we mainly focus on novel accepted publications and new preprint works not yet described as such in those previous Deliverables, and corresponding to the current time period M31-M36. Additionally, during this time period, FHHI further optimized its concept-based explanation pipeline to enable near real-time explanations of the predictions of segmentation and detection models inside the TEMA platform, matching all XAI-related KPIs as defined by Objective OA1 "Increase trustworthiness of extreme data analysis algorithms".

4.2.1. A second-order XAI method for explaining predictive uncertainty

SOTA

High predictive uncertainty occurs for instance when using an ensemble of predictors in the context of data scarcity, which prevents the models in the ensemble from reaching a consensus on what the actual prediction should be [85]. While understanding a models prediction in terms of input features has been tackled extensively within the field of Explainable AI, with many successes, e.g., in image classification with methods such as Grad-CAM [86] and LRP [87], the explanation of predictive uncertainty has received little attention so far.

Advances beyond SOTA

FHHI contributes a new method to the problem of explaining uncertainty for the common case where it is estimated as the variance over an ensemble of predictions. This novel method proposed by FHHI, which was published in the Journal of Pattern Recognition in 2025 [88], accounts for second-order effects and is applicable to general neural network structures, including highly nonlinear ones, and integrates with existing explanation frameworks such as LRP [87], Integrated Gradients (IG) [89] and Shapley Values [90]. High predictive uncertainty commonly arises when a model makes predictions for data points dissimilar from the observed training data [85]. Such a covariate shift may be caused by measurement biases or insufficient and unrepresentative training data collection, i.e., in the case of data scarcity. As a consequence of insufficient training data collection, some input features may remain underrepresented at training time. When these features appear at test time, the model is ill-prepared to interpret the end-user to precisely diagnose what is missing in the current data and, their effect on the prediction task. Thus, the model prediction is unreliable, and predictive uncertainty is high. In this case, explaining predictive uncertainty in terms of underrepresented features can enable subsequently, gather additional training data to

This project has received funding from the European Union – European Commission under grant agreement No 101093003

improve the model. In the FHHI's publication [88] it was shown that the new uncertainty explanation is able to **reveal underrepresented high-level features at test time** and that **retraining on a consolidated dataset reduces uncertainty** attributed to the originally underrepresented feature. The theoretical derivation of the proposed method leads to a general scheme for computing uncertainty explanations, namely a covariance over an ensembles individual classical explanations. Thus it allows to systematically transform classical first-order explanation techniques (such as LRP, GI, etc.) into more powerful second-order uncertainty explainers (CovLRP, CovGI, etc.). In a quantitative evaluation [88] the high performance of the proposed approach was demonstrated, with CovLRP achieving the highest explanation accuracy as evaluated by a feature-flipping experiment, outperforming classical LRP as well as a number of other competitive baselines.

4.2.2. Explanation-guided regularization as a novel data augmentation

SOTA

Training deep learning models generally requires vast amounts of high-quality data, which may be unavailable in real-world applications such as natural disaster management, due to high annotation costs. Consequently, overfitting is a common challenge, where biases or regularities specific to the training set are reflected by the model, resulting in bad generalization to unseen examples. However, acquiring more training data tends to be difficult and expensive. For this reason, data augmentation techniques are commonly employed to mitigate overfitting. These techniques introduce *variations* in training data, thereby *artificially* enlarging the available data and effectively regularizing the model to learn more robust representations. Among the various data augmentation strategies, "occlusion" is a prominent technique that typically focuses on *randomly* masking regions of the input during training. Besides, most of the existing methods such as Random Erasing [91], Mixup [92], CutMix [93] emphasizes *randomness* in selecting and modifying the input features, instead of regions that strongly *influence* the model decisions.

Advances beyond SOTA

FHHI therefore introduces "Relevance-driven Input Dropout" (RelDrop), a novel XAI-guided technique that augments data by masking (currently) *relevant* input features. The proposed method can be efficiently applied during training, requiring only one additional backward pass per batch. In a recent preprint by FHHI from 2025 [94], its effectiveness was shown in the context of 2D image and 3D point cloud classification. In particular, it **doubles the improvement in average test accuracies** over various random data augmentation baselines, demonstrating the increased robustness and generalization ability of the resulting model. The key idea behind the new method is to leverage XAI attributions as a signal to guide data augmentation, as these attributions provide a more informed approach to examine how augmentations affect model predictions. By removing or occluding (currently) important features, RelDrop indeed aims to force the model to base its inference on a larger number of features, and thus increases its robustness and generalization ability.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4.2.3. A framework for sparse and efficient explainable data attribution

SOTA

Data Attribution (DA) has emerged as a promising paradigm that shifts the focus from features attribution to data provenance. With the insights gained on the level of (training) data points, DA provides transparency about the model and individual predictions, e.g. for model debugging, identifying data-related causes of suboptimal performance, such as mislabeled instances, dataset distillation or knowledge discovery purposes. However, existing DA approaches such as Influence Functions suffer from prohibitively high computational costs and memory demands [95]. Additionally, current attribution methods exhibit low sparsity, resulting in non-negligible attribution scores across a high number of training examples, hindering the discovery of decisive patterns in the data.

Advances beyond SOTA

To address those challenges FHHI introduces DualDA which is a surrogate-based DA method, which replaces the final linear layer of the deep learning model with a linear multiclass SVM. The learned weights of an SVM in feature space can be exactly expressed as a weighted sum of representations over the training datapoints by solving the corresponding dual problem. Using an SVM also presents the advantage of exhibiting an implicit bias towards sparsity in the attribution values as, under constrained conditions, only a limited number of support vectors contribute in determining the models decision boundary, which helps to sparsify DA. FHHI further introduces XDA, a method for enhancing DA with capabilities from feature attribution methods to explain why training samples are relevant for the prediction of a test sample in terms of impactful features. Combined, DualDA and XDA constitute the DualXDA framework. In a extended Preprint work [96], including new quantitative evaluations, FHHI demonstrates that DualXDA achieves high attribution quality, excels at solving a series of evaluated downstream tasks, while at the same time **improving explanation time by a factor of up to 4,100,000× compared to the original Influence Functions method** [95], and **up to 11,000× compared to its most efficient approximations from the literature to date [97, 98]**.

4.2.4. A model for Cognitive Understanding of Explanations

SOTA

As machine learning systems increasingly inform critical human-decisions, such as in natural disaster management scenarios, the need for human-understandable explanations grows. Current evaluations of Explainable AI (XAI) however often prioritize technical fidelity [99, 100] over cognitive accessibility which critically affects end-users.

Advances beyond SOTA

To address those limitations, FHHI proposes a structured framework for modeling the "Cognitive Understanding of Explanations" (CUE), building on prior cognitive models of visualization com-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

prehension [101]. CUE uniquely links the external properties of explanations to internal user cognitive processes. More specifically, in the field of data visualizations, legibility refers to the clarity of individual visual elements (such as bars or lines), while readability involves understanding the overall narrative conveyed by these elements. Together, these concepts play a critical role in minimizing cognitive load and maximizing the effectiveness of a visualization by facilitating easier interpretation. The CUE framework was applied and evaluated in a user-study focused on heatmap explanations, **revealing a significant impact of colormap choices** on end-user interaction with heatmap explanations, underscoring the effect of explanation properties on user understanding. This FHHI study was published in the IJCAI'25 International Joint Conference on Artificial Intelligence Workshop on Explainable Artificial Intelligence [102]. The findings of this study underscore the need for explanation systems that go beyond visual optimization and that adaptive, multi-modal interfaces, allowing for interactive and tailored engagement, such as the SmartDesk interface developed inside the TEMA platform, may better support end-user needs.

4.2.5. Concept-based explanations for NDM and beyond

SOTA

While traditional local XAI methods focus on feature importance for individual predictions, global XAI approaches aim to understand overall model behavior by explaining the roles of internal representations and encoded features through concepts [103, 104]. Global explanations facilitate the detection of spurious model behaviors and encompass solutions to mitigate such issues, as well as the identification of outlier (Out-of-Distribution (OOD)) samples, hence making model decisions more robust, which is critically relevant in high-stake scenarios, such as in natural disaster management and medical applications.

Advances beyond SOTA

To provide an overview of state-of-the-art interpretability-driven shortcut detection and bias mitigation methods for high-stake decisions, such as in medicine and NDM, FHHI proposed a comprehensive review published in the Machine Learning Journal in 2025 [105]. This review further extends the Reveal2Revise framework [106] with bias annotation techniques, enabling the (semi-)automated generation of sample- and feature-level bias annotations. Further, in a recent Preprint [107], FHHI proposed a novel framework for interpreting the CLIP model via latent attributions for instance-wise attribution of sparse, interpretable components, enabling a dual perspective on model behavior: understanding what concepts are encoded, and also how they influence predictions. By combining attributions with alignment to expected semantics, the approach automatically identifies reliance on both spurious features and surprising concepts. Lastly in a work published at the ACL Conference 2025 [108], FHHI proposed FADE, a new automated evaluation framework designed to rigorously evaluate the alignment between features and their open-vocabulary feature descriptions. By combining four complementary metrics, namely Clarity, Responsiveness, Purity, and Faithfulness, the approach gives a comprehensive assessment of how a feature reacts to instances of the described concept, an evaluation of the description itself as well as the features causal role in the models outputs.
Besides advancing various generic XAI methods for extreme data conditions, FHHI also implemented and deployed an XAI component (TFA-tech-02) inside the TEMA sofware platform, based on a method that was developed in a previous TEMA reporting period, namely Prototypical

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Concept-based Explanation (PCX) [109]. This XAI method leverages latent relevance distributions using Gaussian Mixture Model clustering of concept-based relevance vectors, which are obtained by summing-up LRP relevances [87] across spatial dimensions inside a deep convolutional layer of a neural network based computer vision model. This allows one to: 1) quantify the similarity between a new prediction and a "prototypical" prediction (i.e., a cluster centroid) in order to validate whether the current prediction is "ordinary" or if it might constitute an "outlier", 2) as well as inspect the semantic of concepts through visualizing relevance-maximizing reference samples from the training data along with concept-conditional heatmaps. While the previously published work on PCX [109] applied the technique only to classification models, a novelty of the work carried out at TEMA is to extend this method to segmentation and object detection models. More precisely, FHHI implemented and deployed PCX inside the TEMA platform to explain the predictions of the YOLOv6s6 model for person and vehicle detection trained by AUTH (and available as component TFA-tech-05 in the TEMA platform), as well as multiple UNet models for flood and fire segmentation trained by AUTH (available as component TFA-tech-06).
Figures 46-52 illustrate various PCX results. Figure 46 illustrates the clusters and prototypes for vehicle detection. One can see through this plot the different types of strategies that the model has learned to detect vehicles. Prototypes 0 to 3 correspond to personal cars, prototypes 4 and 5 to white trucks, and prototypes 7 and 8 to red ambulances, while prototype 6 corresponds to samples with no vehicle detected. Hence the model has implicitly learned to detect various types of vehicles (although it was only trained to detect all types of vehicles) as revealed by PCX. Figure 47 represents the average concept usage per prototype. Concepts are shapes or colors that filters of convolutional layers have learned to use for prediction. So for instance red vehicle parts are used to detect red ambulances. Then Figures 48 and 49 explain the predictions of single samples for vehicle detection with PCX. The explanations contain a heatmap highlighting input regions most decisive for the prediction, together with the prediction result (i.e., the bounding box for detection or the segmentation mask) overlaid on the input image on the left side of the explanation. Further, the most important concepts responsible for this result and their corresponding conditional heatmaps are retrieved in the middle part of the explanation. Concepts are illustrated via reference images extracted from training data. Finally, the right side of the explanation depicts the same results (prediction and heatmaps) for a prototype, which is another input image extracted from the training data that resembles the most to the current prediction. Most importantly the difference to the prototype column (which contains concept-based relevance score differences displayed inside green or red squares) enables one to inspect whether the current prediction is likely to be an ordinary one (in case the concept usage is similar between the current prediction and the prototype), or whether it is likely an outlier (in case the concept usage differs a lot). According to the comparison of concept profiles in Figure 48, this prediction is an ordinary sample, while Figure 49 depicts an outlier sample, since concepts representing green vegetation and persons are over-used for this prediction. Figure 50 shows the clusters and prototypes for AUTH's person detection model. Here again one prototype corresponds to no detection, then prototype 0 corresponds to black persons with relatively sharp silhouette, while prototype 1 corresponds to blurry black persons. Then Figure 51 shows the PCX prototypes for flood segmentation. One can see that each prototype corresponds to differents types of flood the model has implicitly learned to detect. For instance prototype 3 contains low-scale street-level flood, prototype 0 is localized linear flood along river, while prototype 7 corresponds to wide-level entire floodplain inundation. Finally Figure 52 illustrates a PCX explanation on a prediction using a synthetic flood image generated by ATOS. Here all concepts are used in a similar fashion to the nearest prototype, hence the prediction is labeled as an ordinary one. This further validates the usage of synthetic images for making predictions with AUTH's model, as will be done during the BRK trial.
Further, during the current reporting period, FHHI improved its results over all KPIs related to

This project has received funding from the European Union – European Commission under grant agreement No 101093003

computation times of the local and global explanations as defined by the TEMA **Objective OA1 "Increase trustworthiness of extreme data analysis algorithms"**, these results are reported in Table 13. This was achieved by adapting the LRP rule [87] during the relevance backward pass (from using the LRP-gamma rule to using the LRP-alpha1 rule inside convolutional layers) to reduce the attribution time, increasing the caching of intermediate computations for global explanations, and various code optimizations.

Table 13. KPIs for computation time ratios of local and global XAI methods using the AI models from AUTH for segmentation and detection. These time ratios are reported as average over 100 samples and recorded on GPU.

Computation Time Ratio	Model	Dataset Size	KPI	Target Value
$\text{local XAI time}/\text{prediction time}$	PIDNet	955	2.6	4
	YOLOv6s6	695	2.5	4
$\text{global XAI time}/\text{all local XAI time}$	PIDet	955	0.9	10
	YOLOv6s6	695	0.4	10

This project has received funding from the European Union – European Commission under grant agreement No 101093003

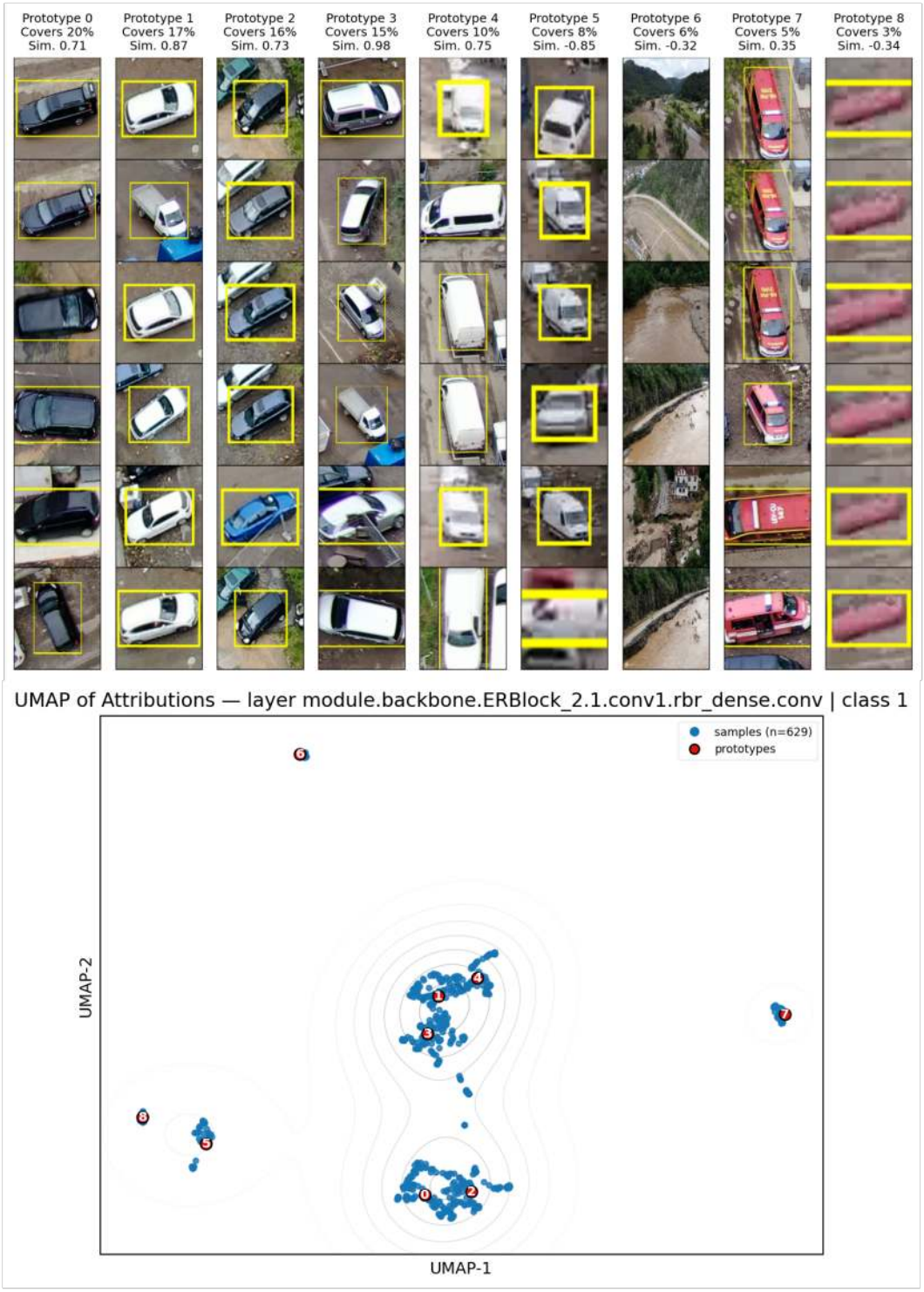


Figure 46. UMAP Visualization of the PCX clusters and associated prototypes for the YOLOv6s6 vehicle detection model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

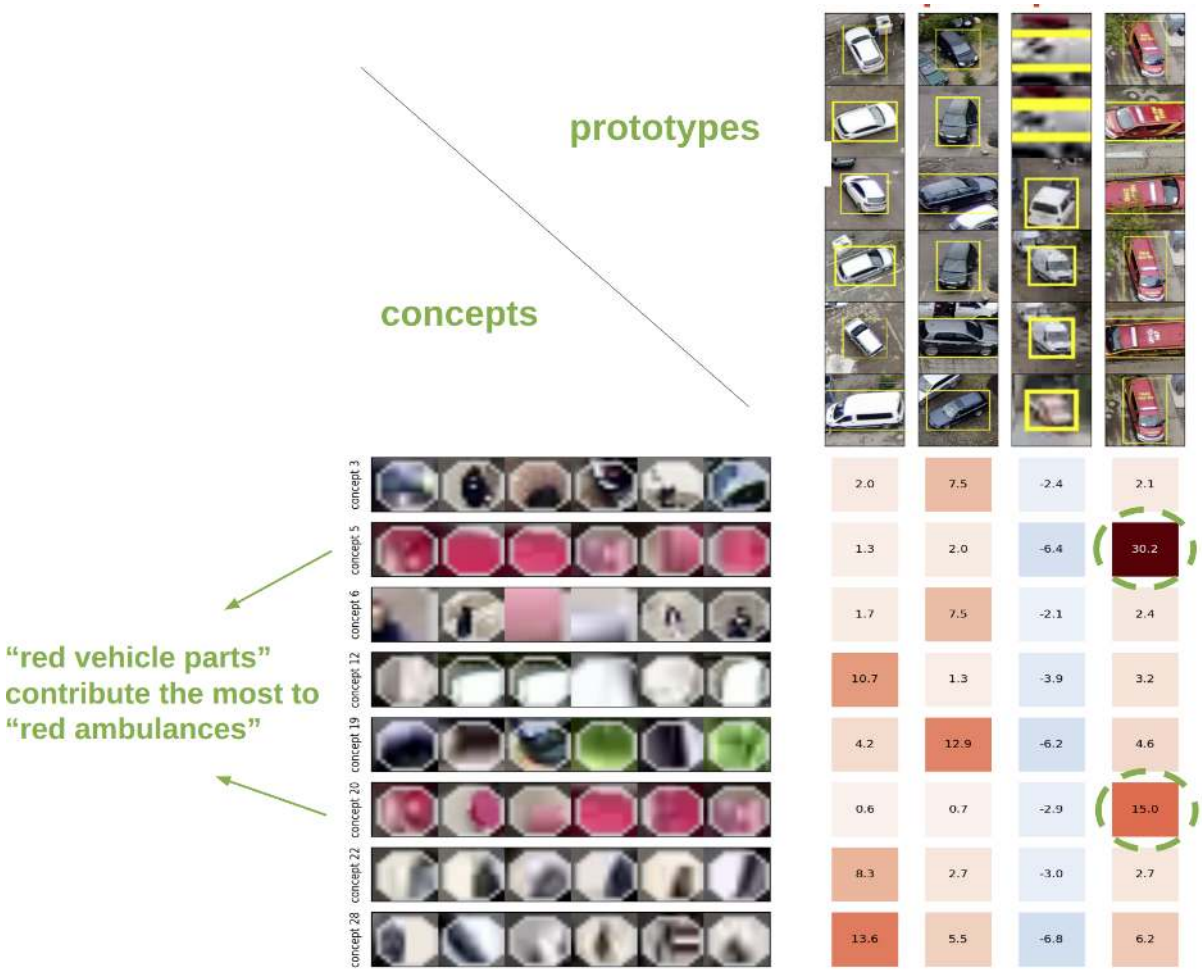


Figure 47. Grid plot of the concept contributions per PCX prototype for the YOLOv6s6 vehicle detection model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

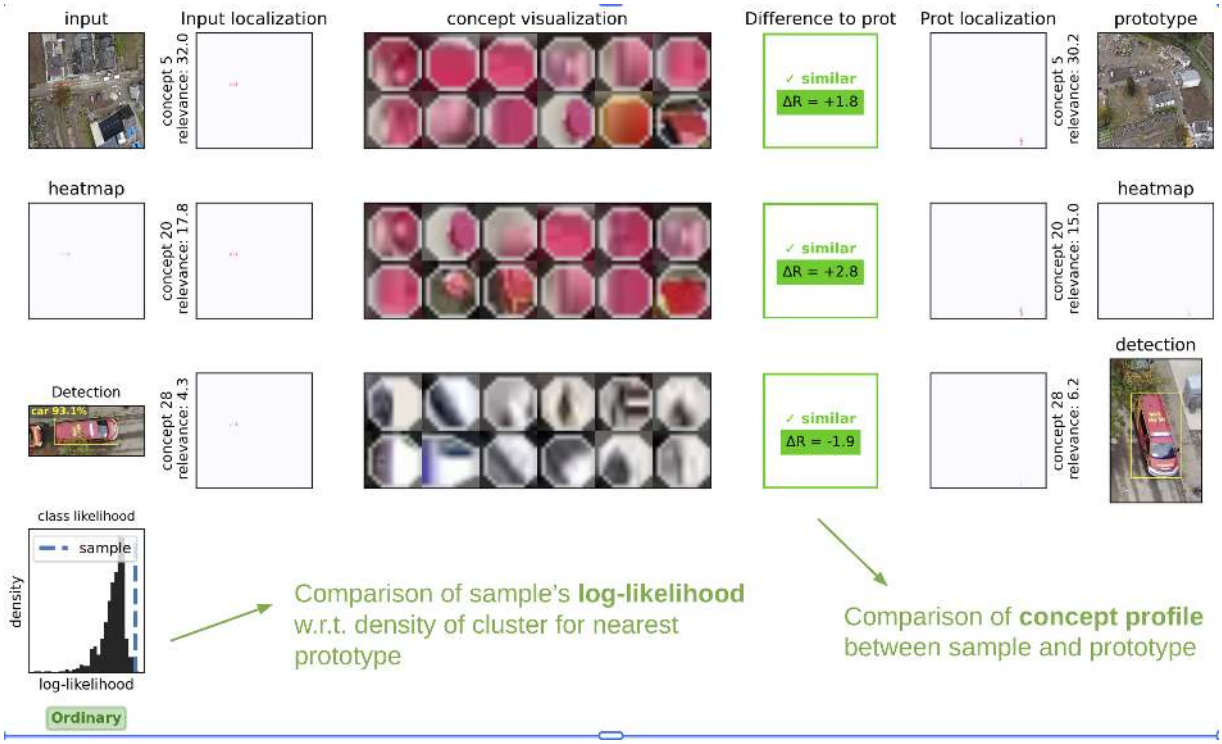


Figure 48. Example PCX explanation of the model's prediction on an ordinary sample using the YOLOv6s6 vehicle detection model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

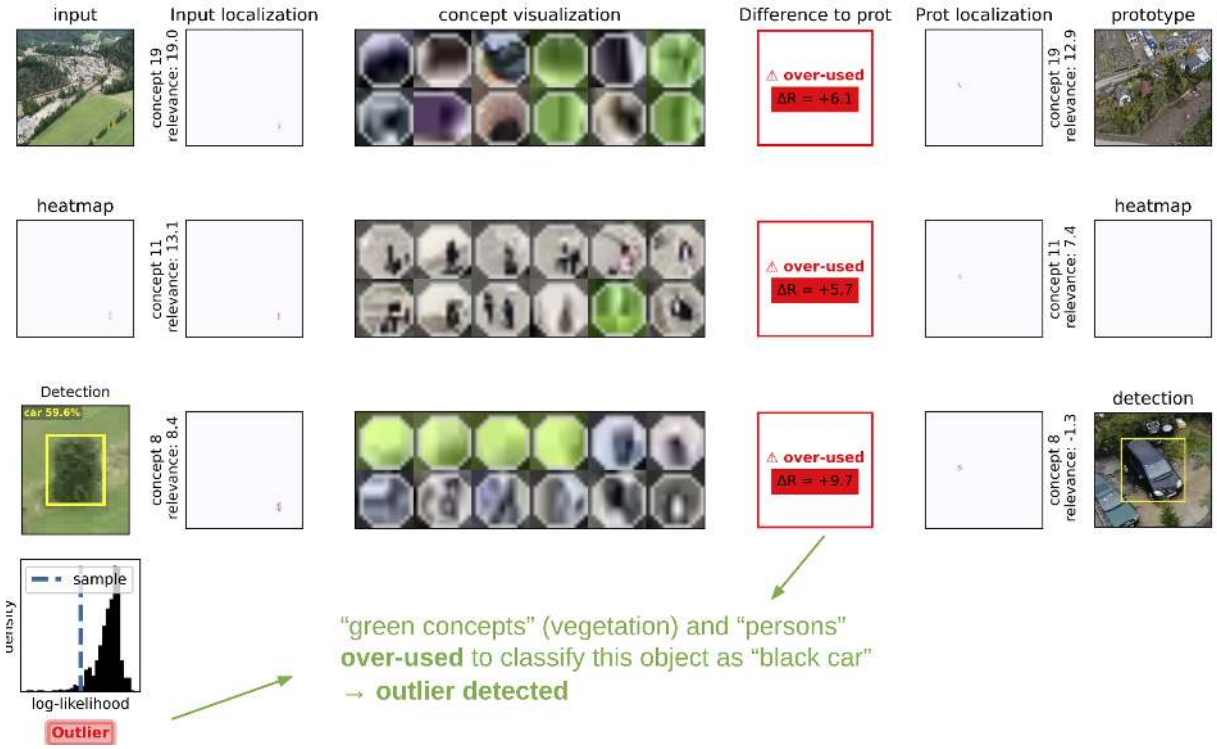


Figure 49. Example PCX explanation of the model's prediction on an outlier sample using the YOLOv6s6 vehicle detection model from AUTH.

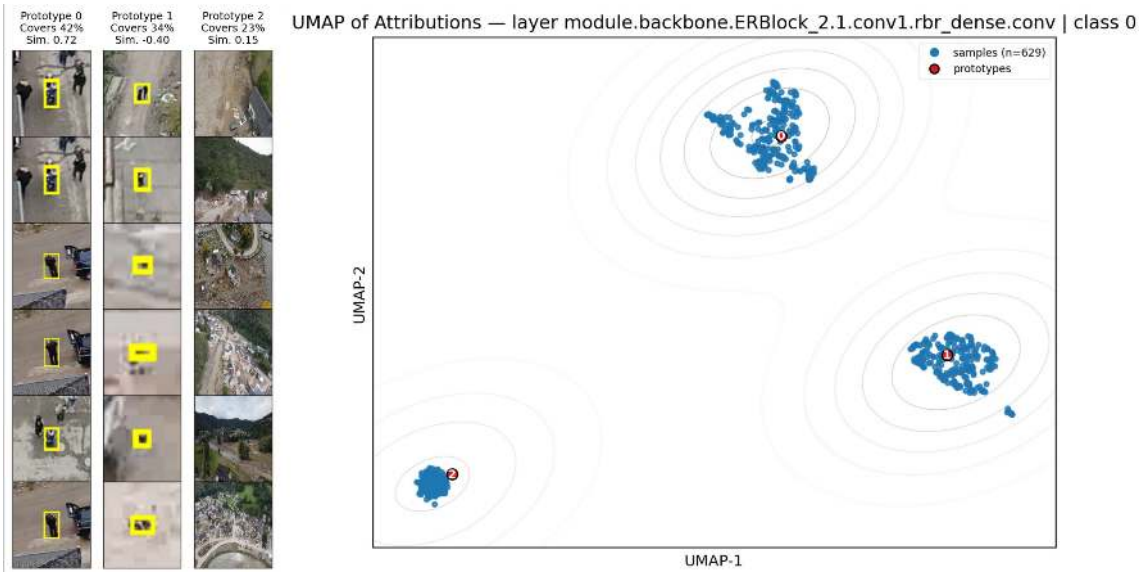


Figure 50. UMAP Visualization of the PCX clusters and associated prototypes for the YOLOv6s6 person detection model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

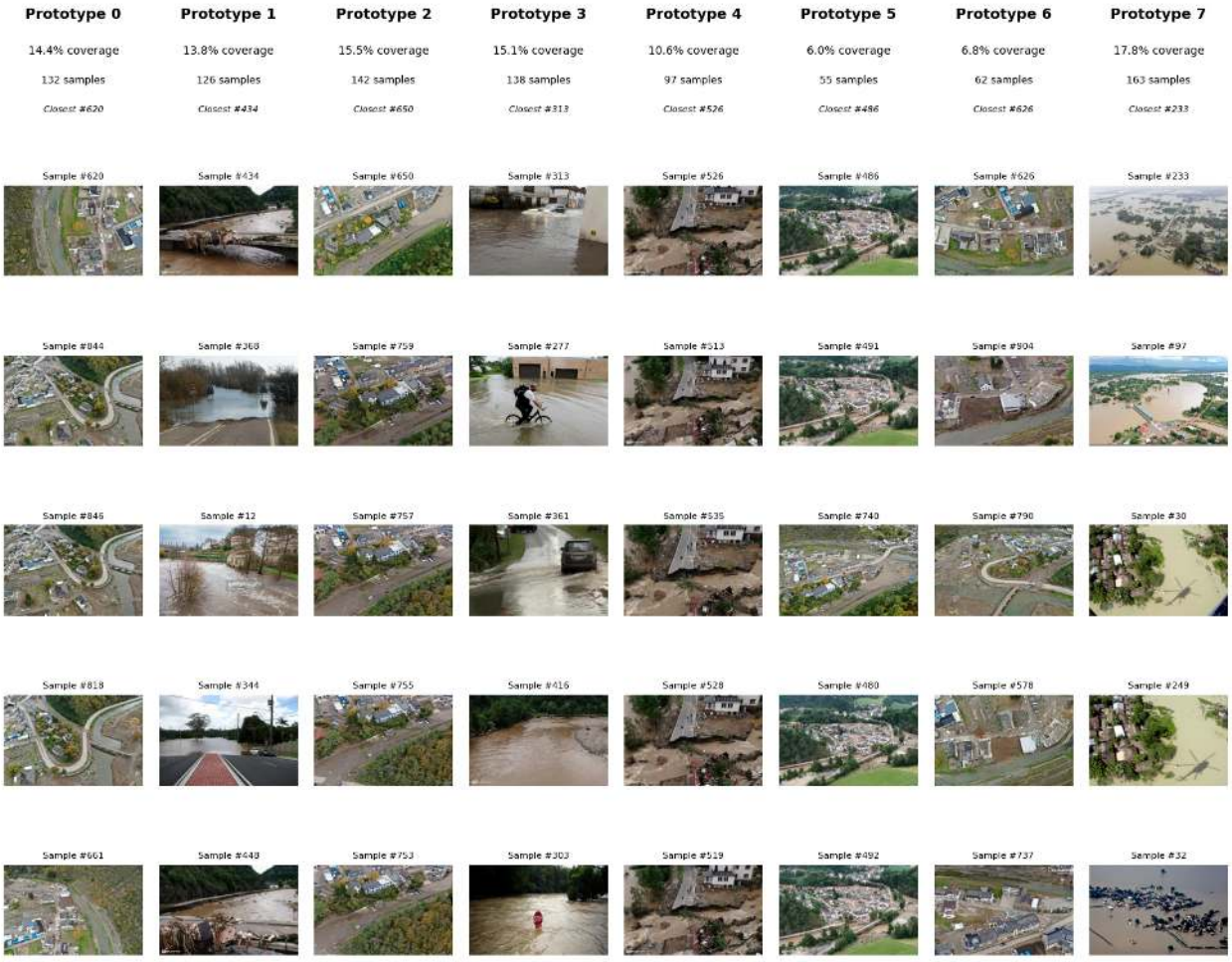


Figure 51. PCX Prototypes for the PIDNet flood segmentation model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

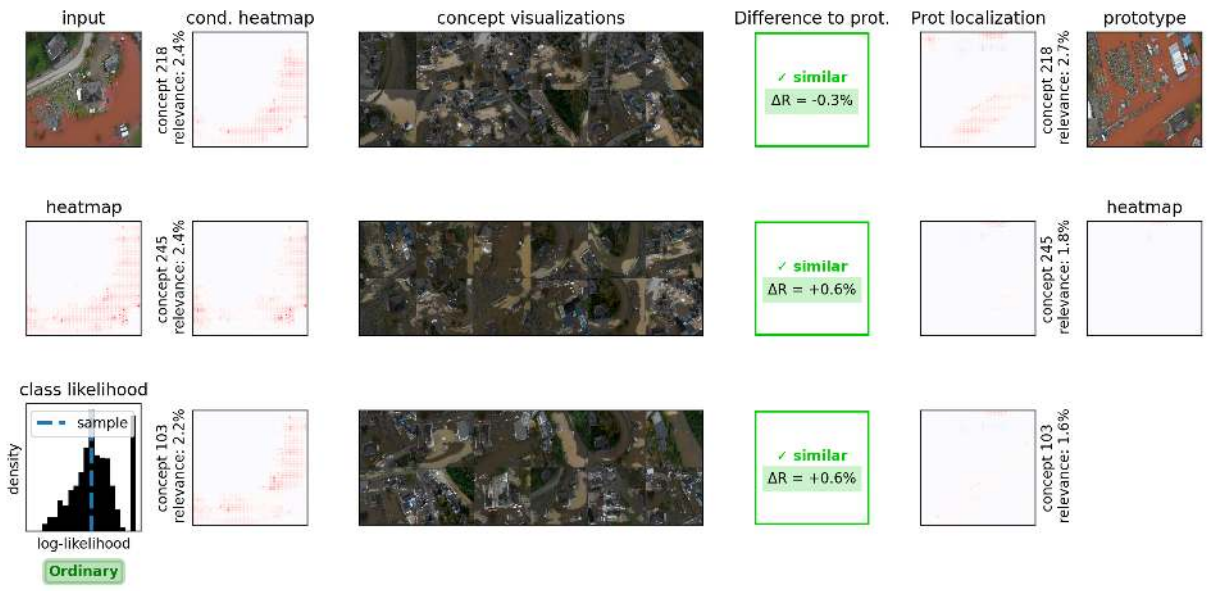


Figure 52. Example PCX explanation of the model's prediction on an ordinary sample's synthetic flood image generated by ATOS using the PIDNet flood segmentation model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

5. Multiple Learning Paradigms

5.1. Introduction

While Deep Neural Networks (DNNs) have become indispensable for perception tasks such as segmentation, classification, and object detection, their effectiveness typically depends on access to large, diverse, and well-annotated datasets. In contrast, real-world Natural Disaster Management (NDM) contexts are often characterized by data scarcity and heterogeneity data may be fragmented across regions, arrive sequentially during an event, or exhibit substantial domain shifts due to environmental or sensor variations.
Under such conditions, conventional centralized training paradigms are insufficient. Models trained in isolation tend to overfit, struggle to generalize across unseen domains, and cannot readily adapt to new or evolving information streams. To address these limitations, this section explores a set of advanced learning paradigms that collectively enable adaptive, distributed, and knowledge-aware AI systems for NDM. This work introduces multiple complementary strategies that enhance model collaboration, adaptability, and resilience in real-time operations. These developments contribute to the overarching objectives of WP3, notably those under Task T3.5, by improving the accuracy, adaptability, and responsiveness of AI-driven perception systems in real-world disaster management pipelines.

5.2. Collective knowledge-based forest fire classification

SOTA

Forest fire classification is a critical task for effective NDM, as rapid and accurate identification of fire conditionssuch as detecting active fires and classifying affected areas (burnt, half-burnt, non-burnt)is essential to minimize damage to ecosystems, human lives, and infrastructure. Timely classification directly informs emergency response strategies, resource allocation, and containment measures, ultimately shaping the effectiveness of disaster mitigation efforts. Existing frameworks for forest fire classification primarily rely on isolated DNN models, which are trained and operate independently, limiting their adaptability when encountering new or evolving tasks in emergency situations [110, 111, 112]. Moreover, while recent works explore Knowledge Distillation (KD) [113, 114] and Federated Learning (FL) [115] methods individually, these approaches have not been extensively integrated into a unified multi-agent architecture specifically tailored for NDM. Consequently, current systems face limitations such as restricted adaptability, scalability, and real-time knowledge sharing among heterogeneous DNN agents.

Advances beyond SOTA

Motivated by the lack of collaboration and autonomous knowledge sharing among agents and the limited adaptability of isolated DNNs, AUTH makes several key contributions. The method with more details is described in a conference paper [3]:

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Anestis Kaimakamidis and Ioannis Pitas, "Leveraging Collective Knowledge for Forest Fire Classification", IEEE Symposium on Computers and Communications (ISCC), 2024

Firstly, AUTH introduces the Fire Classification Multi-Agent (FCMA) framework, a novel multi-agent architecture specifically designed for natural disaster management systems, leveraging peer-to-peer KD and FL to enable effective collaboration and autonomous learning among DNN agents. Secondly, it proposes a specialized Agent Knowledge Self-Assessment (AKSA) module, integrating Out-of-Distribution (OOD) detection using Likelihood Regret (LR) to autonomously evaluate each agent's capability to handle new tasks and trigger collaborative knowledge exchange when necessary. Lastly, AUTH provides comprehensive experimental validation demonstrating the effectiveness of collective knowledge dissemination among diverse agent architectures and comparing the efficiency and accuracy of peer-to-peer and federated learning options within a practical forest fire classification scenario, highlighting the potential for significantly improving performance and adaptability in real-time emergency situations.
In **??**, the comparison of the two methods of knowledge transfer: peer-to-peer KD and FL in terms of average accuracy on the Blaze dataset [1], before and after knowledge dissemination among DNN agents is presented. Examples of the Blaze classification dataset are depicted on fig. 53.

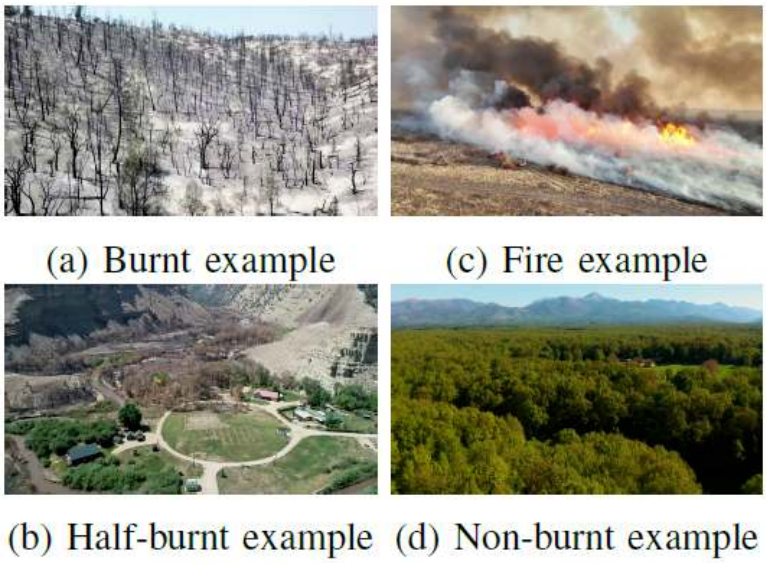


Figure 53. Examples of the Blaze classification dataset [1]

For KD, a ResNet101 [116] is used as teacher while a ResNet50 [116] and an Alexnet [117] are used as students. For FL 3, ResNet50 [116] agents are used.

Table 14. Average accuracy results comparing peer-to-peer Knowledge Distillation (KD) with ResNet101 [116] as teacher and and ResNet50 [116] - Alexnet [117] as students, and Federated Learning (FL) with 3 ResNet50 [116] agents, for forest fire classification using the Blaze dataset [1]. The KD-based approach achieves better accuracy improvement compared to FL.

Method	Initial Accuracy (%)	Final Accuracy (%)	Improvement
Peer-to-peer KD	76.31	**77.50**	+1.19 (+1.6%)
Federated Learning (FL)	76.15	76.56	+0.41 (+0.4%)

This project has received funding from the European Union – European Commission under grant agreement No 101093003

The results demonstrate that the KD-based approach achieves an increase in average accuracy from 76.31% to 77.5% (**+1.19**), outperforming the FL-based approach, which shows a more modest improvement from 76.15% to 76.56% (**+0.41**). This indicates that the KD method is effective in transferring knowledge among heterogeneous agent communities.

5.3. Continual learning for AI algorithms

SOTA

Natural disaster scenarios often present diverse and rapidly changing visual data, such as images of wildfires and floods. Continual Learning (CL) involves adapting the prior DNN knowledge to new tasks, without forgetting the old ones. CL is beneficial for NDM because it enables DNNs to dynamically adapt to new visual patterns while retaining previously learned knowledge. Recent approaches to CL can be grouped into three primary categories: regularization-based, replay-based, and dynamically expandable networks. Regularization methods, such as Elastic Weight Consolidation (EWC) [118] and Learning without Forgetting (LwF) [119], focus on preserving key network parameters across tasks. Replay-based methods, including Incremental Classifier and Representation Learning (iCaRL) [120], mitigate catastrophic forgetting by maintaining a buffer of past samples. Dynamic expandable architectures, like DER [121] and DyTox [122], progressively add new parameters or tokens to accommodate new tasks. Despite their effectiveness, these methods still experience substantial forgetting or incur significant computational overhead. Notably, transformer-based CL methods remain relatively unexplored. Approaches such as MEAT [123] utilize parameter masks, while L2P [124] and DualPrompt [125] rely on soft prompting, which limits their scalability as the number of tasks grows. Consequently, there is a critical gap in developing efficient, memory-conserving transformer architectures that dynamically adapt attention mechanisms to the current task without relying on stored exemplars or extensive parameter growth.

Advances beyond SOTA

AUTH proposes a novel method called Feedback Continual Learning Vision Transformer (FCL-ViT), which significantly advances beyond existing SOTA CL methods by introducing a novel feedback mechanism that dynamically generates task-specific attention features in real-time. The method with more details is described in a journal paper [4]:

Anestis Kaimakamidis and Ioannis Pitas, "FCL-ViT: Task-aware attention tuning for Continual Learning", Pattern Recognition Letters, 2025

Unlike traditional feed-forward CL methods, which often rely on rehearsal memories or continuously expanding model parameters, FCL-ViT leverages Tunable self-Attention Blocks (TABs) and Task-Specific Blocks (TSBs) to adaptively retune attention across tasks without increasing memory demands substantially. Experimental evaluations demonstrate that FCL-ViT achieves superior performance on standard CL benchmarks, notably outperforming methods such as iCaRL [120], DER [121], and DyTox [122]. Table 15 reports the average Top-1 classification accuracy (Avg) after each task, along with the Top-1 accuracy obtained after the final task (Last). The column #TP represents the number of trainable parameters (in millions) for each method.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

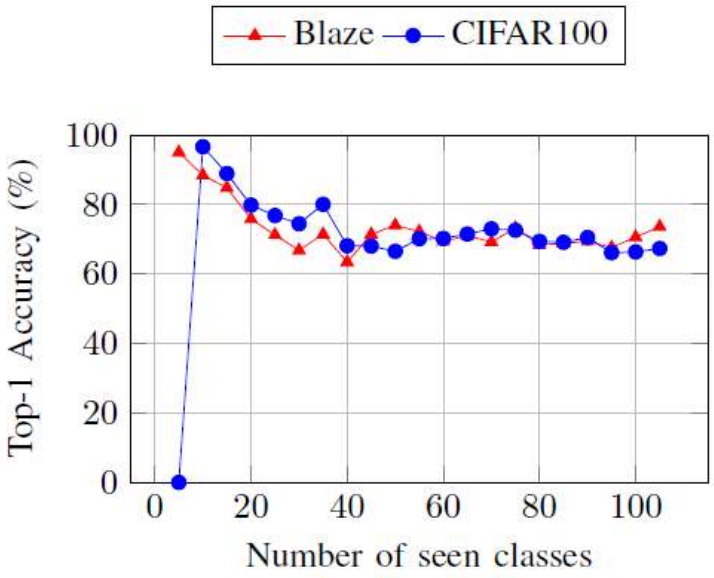


Figure 54. FCL-ViT classification accuracy on the wildfire BLAZE [1] and CIFAR100 [2] datasets.

Table 15. Top-1 accuracy classification results on Imagenet-100 for 10 task splits.

	10 tasks		
Methods	**#TP**	**Avg**	**Last**
iCaRL [120]	11.22	57.84	46.13
DER [121]	112.27	**77.18**	66.70
DyTox+ [122]	11.01	77.15	69.10
FCL-ViT	14.23	75.42	**71.80**

Importantly, FCL-ViT maintains stable accuracy across an increasing number of tasks, effectively mitigating catastrophic forgetting even without rehearsal strategies. This highlights its capability for sustained adaptability and generalization in diverse continual learning scenarios. The performance (Top-1 Accuracy) of FCL-ViT when trained sequentially on two datasets: first, the BLAZE wildfire classification dataset [1], and subsequently the CIFAR100 dataset [126] is demonstrated in fig. 53. Specifically, the figure illustrates that after initially learning the BLAZE dataset, the model maintains stable accuracy on BLAZE even after learning 100 new classes from CIFAR100. Moreover, the model demonstrates minimal degradation in accuracy on the CIFAR100 dataset after initially learning the significantly different BLAZE wildfire dataset. This result highlights the models effectiveness in CL, showing its capability to retain previous knowledge (BLAZE) while sequentially acquiring new knowledge (CIFAR100) without severe catastrophic forgetting.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

5.4. Cloud Learning-by-Education Node Community (C-LENC) framework

SOTA

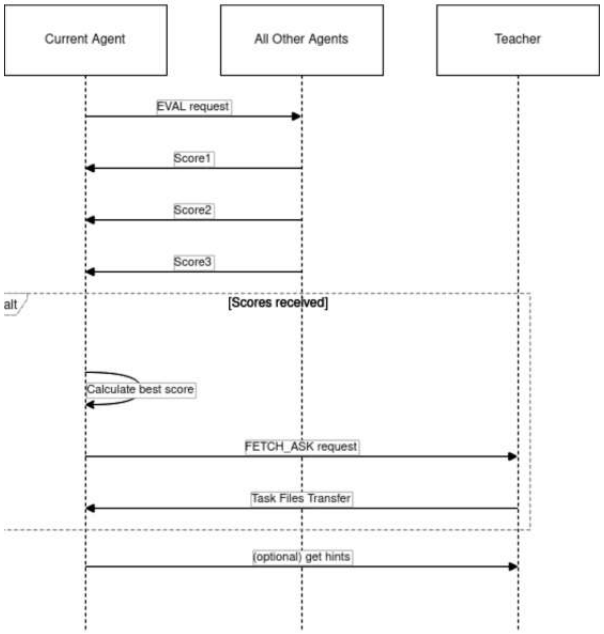


Figure 55. Communication diagram for the workflow of learning a new task.

The Learning by Education Node Community (LENC) [127] paradigm has laid out the foundation for examining the interaction rules between collaborative nodes performing knowledge exchange, based on evaluating their knowledge. As a research prototype, Out Of Distribution detectors were employed for knowledge (self) assessment by using the likelihood regret as an out-of-distribution score in variational autoencoders, and knowledge distillation for model updates, all operating on the same computer. More specifically, when a new data stream appears as input in the Deep Neural Network (DNN) of a LENC node, the out-of-distribution detector examines each sample and categorizes it as in or out of distribution. If a sample is found outside the training distribution, the node sends the sample to all other nodes to find which one has seen similar samples. At this point, the nodes that know the sample become potential teachers, and the one with the best score becomes the teacher, while all nodes that do not know the sample become students. Finally, the knowledge distillation procedure begins.

Advances beyond SOTA

LENC's implementation as a research prototype, and the fact that it operates on a single computer, does not make it a complete framework, as it lacks provisions for connectivity and cloud implementation. In this work, AUTH presents an extension to the LENC paradigm, the novel Cloud Learning by Education Node Community (C-LENC) framework, that can be used to per-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

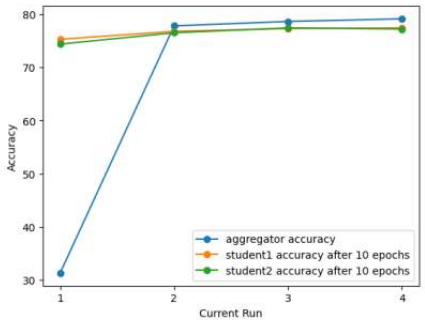


Figure 56. DNN Classification accuracy of the Aggregator and the Student DNNs on the CIFAR-10 test dataset

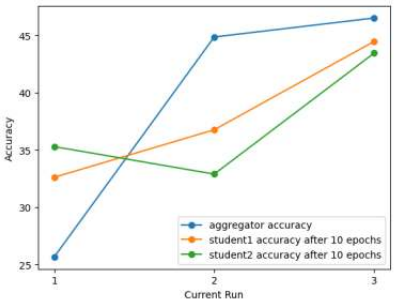


Figure 57. DNN Classification accuracy of the Aggregator and the Student DNNs on the BLAZE test dataset.

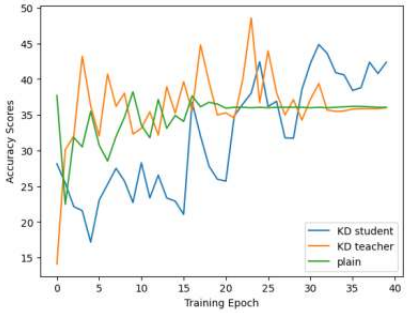


Figure 58. Knowledge Distillation classification accuracy for the Teacher and Student DNN, as well as a plain DNN trained on ground truth data on the BLAZE test dataset.

form a plethora of collaborative distributed machine learning workflows between the nodes of the C-LENC network, operating on the cloud. Extending the original LENC, the proposed C-LENC framework supports all the functionalities of LENC for knowledge assessment and distillation, as well as additional distributed learning workflows such as federated learning and distributed inference. The paper with more details is accepted to a conference:

Nick Tzavidas, Anestis Kaimakamidis, and Ioannis Pitas, "Cloud Learning-by-Education Node Community (C-LENC) framework", technical report, 2025.

Each LENC node is implemented in the form of a containerized software package, with each

This project has received funding from the European Union – European Commission under grant agreement No 101093003

node occupying a Docker container, enabling node management and communication. To allow the nodes to be aware of the network, a simple service discovery-like HTTPS service is used. The communication is achieved through network sockets and the Transmission Control Protocol (TCP). The proposed C-LENC was evaluated on 4 workflows: a.) learning a new task, b.) federated learning, c.) multi-teacher knowledge distillation, and d.) distributed inference (using majority voting), on the CIFAR-10 and BLAZE image datasets, adding the federated learning and distributed inference functions to the LENC framework. A communication diagram for the workflow of learning a new task can be seen in Figure 55. Results for the federated learning task are presented in Figures 56 and 57. For the distillation task, the accuracy scores after training a student with the outputs of a teacher for 40 epochs against a student that is trained for 40 epochs on the ground truth, are presented in Figure 58. For the distributed inference task, the Majority Voting method was employed by training 5 ResNet18 models on the BLAZE dataset, and results on its test set can be found in Table 16.

Table 16. Classification accuracy for the Majority Voting estimator DNNs and the aggregator.

Model	Test Data Accuracy
Model 1	37.78%
Model 2	30.31%
Model 3	25.14%
Model 4	26.80%
Model 5	28.84%
Maj. Vot.	28.97%

5.5. Proto-SVDD: Decentralized Federated Object Detection with Prototype-Based Communication

SOTA

Recent advances in Federated Learning (FL) have focused on classification tasks, while federated object detection remains less explored due to its multitask nature and high communication cost. Baseline approaches such as FedAvg [128] rely on full model weight aggregation, leading to significant communication overhead and potential privacy leakage. Prototype-based FL methods, including FedProto [129], PearFL [130], TurboSVM-FL [131], and knowledge-distillation-based schemes [132], have been proposed to alleviate communication bottlenecks by exchanging class-wise prototypes or informative embeddings. However, these methods were mainly developed for image classification and are not directly tailored to object detection, where both class recognition and bounding box localization must be addressed. Furthermore, prototype averaging strategies fail to capture non-IID data variability across clients, while support vector selection does not fully preserve the semantic representativeness of classes. To the best of AUTH knowledge, no decentralized prototype-based federated object detection study has systematically benchmarked such methods under non-IID splits for UAV datasets.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Advances beyond SOTA

AUTH proposes Proto-SVDD, a fully decentralized prototype-based FL framework for object detection. Instead of exchanging full DNN weights, Proto-SVDD transmits compact class-wise prototypes derived from the classification head of YOLOv6. The paper with more details will be submitted to a conference:

Eugenios Vlachos, Christos Papaioannidis and Ioannis Pitas," Proto-SVDD: Decentralized Federated Object Detection with Prototype-Based Communication", technical report, 2025.

To enhance representativeness, AUTH integrates Support Vector Data Description (SVDD) [133] to select embeddings that lie within class-specific hyperspheres, yielding more semantically consistent prototypes. During training, each client exchanges only SVDD centers with randomly chosen peers, eliminating reliance on a central server. A prototype alignment loss is introduced to guide local embeddings toward aggregated peer prototypes, ensuring robust convergence under non-IID conditions. Extensive experiments on the VisDrone-DET2019 dataset with varying client numbers (3, 5, 8, and 15) demonstrate that Proto-SVDD achieves competitive or superior detection accuracy compared to state-of-the-art prototype-based FL methods, while reducing communication to just two vectors per client per round. This method offers a scalable, communication-efficient, and privacy-preserving solution for decentralized object detection tasks. The evaluation of several state-of-the-art prototype-based methods on the VisDrone-DET2019 with respect to map@0.5 % and mapo.5:0.95 % is presented on Table 17. Furthermore, prototype loss results further validate the benefit of our prototype-based coordination, with results summarized on Table 18.

*Table 17. Comparison of Prototype FL methods on VisDrone-DET2019 across different numbers of clients under non-IID splits. The mAP is reported at Rounds 1, 5, and 9 for $\alpha = 1$. Best results are in **bold**. Note that Local and FedAvg are included only as lower and upper performance references, respectively, and not as prototype-based baselines.*

#Clients	Method	mAP Round 1		mAP Round 5		mAP Round 9	
		mAP@0.5 %	*mAP@0.50:0.95 %*	*mAP@0.5 %*	*mAP@0.50:0.95 %*	*mAP@0.5 %*	*mAP@0.50:0.95 %*
3	Local	37.81	20.31	47.62	25.45	50.54	26.40
	FedAvg	40.17	21.02	50.17	27.27	53.14	29.15
	FedProto	37.43	19.80	45.87	24.67	48.70	26.40
	PearFL	37.36	19.33	46.36	24.73	48.60	26.47
	TurboSVM-FL	37.12	19.05	45.86	24.50	49.23	26.60
	Proto-SVDD	**37.92**	**20.49**	**47.96**	**25.87**	**51.33**	**26.73**
5	Local	28.30	10.90	37.83	19.31	42.05	21.12
	FedAvg	30.85	13.78	42.74	23.20	46.72	25.74
	FedProto	27.55	11.85	35.82	18.72	40.07	20.72
	PearFL	27.58	11.73	36.48	18.88	39.78	20.52
	TurboSVM-FL	27.83	10.21	36.33	18.64	40.97	19.78
	Proto-SVDD	**28.92**	**11.30**	**38.45**	**19.94**	**42.78**	**22.19**
8	Local	21.20	9.85	**30.81**	15.25	34.24	17.25
	FedAvg	23.94	11.68	34.21	18.29	39.16	20.13
	FedProto	20.69	9.56	30.57	15.12	34.22	17.20
	PearFL	20.62	9.52	30.39	15.09	34.14	17.27
	TurboSVM-FL	20.56	9.57	30.60	15.22	34.19	17.30
	Proto-SVDD	**21.93**	**10.77**	30.79	**15.33**	**34.92**	**17.66**
15	Local	11.09	4.75	18.11	8.48	21.37	10.17
	FedAvg 15	13.38	5.05	23.17	11.82	26.90	13.16
	FedProto	11.58	4.97	18.33	10.25	21.66	10.21
	PearFL	11.23	4.80	18.18	8.53	21.48	**10.22**
	TurboSVM-FL	11.33	4.84	18.44	8.54	21.60	10.21
	Proto-SVDD	**12.57**	**4.82**	**19.69**	**10.69**	**22.26**	10.12

This project has received funding from the European Union – European Commission under grant agreement No 101093003

*Table 18. Proto Loss comparison of FL methods on VisDrone-DET2019 across different numbers of clients under non-IID splits. Best results are in **bold**.*

#Clients	Method	Proto Loss Round 1	Proto Loss Round 5	Proto Loss Round 9
3	Local	–	–	–
	FedAvg	–	–	–
	FedProto	0.2221	0.0048	0.0052
	PearFL	0.2863	0.0114	**0.0032**
	TurboSVM-FL	0.1292	0.0042	0.0023
	Proto-SVDD	**0.1156**	**0.0032**	0.0066
5	Local	–	–	–
	FedAvg	–	–	–
	FedProto	0.2168	0.3367	0.2016
	PearFL	0.1738	0.1672	0.1644
	TurboSVM-FL	0.1383	**0.0129**	0.0382
	Proto-SVDD	**0.0685**	0.0216	**0.0194**
8	Local	–	–	–
	FedAvg	–	–	–
	FedProto	0.0969	0.0344	0.0384
	PearFL	0.1714	0.0467	0.0482
	TurboSVM-FL	0.0838	0.0184	0.0092
	Proto-SVDD	**0.0558**	**0.0072**	**0.0047**
15	Local	–	–	–
	FedAvg	–	–	–
	FedProto	0.1290	0.1056	0.1216
	PearFL	0.1352	0.0713	0.0952
	TurboSVM-FL	0.0516	**0.0361**	0.0694
	Proto-SVDD	**0.0271**	0.0398	**0.0481**

5.6. Weakly Supervised Multi-Class Semantic Segmentation

5.6.0.1. SOTA

The problem of weakly supervised semantic segmentation has gained increasing importance, especially in natural disaster management (NDM), where annotated datasets covering multiple classes are scarce and expensive to obtain. Current state-of-the-art approaches often rely on text-based supervision, such as CLIP-ES and ExCEL [134, 135], which use CLIP [136]to generate text embeddings from prompt-engineered phrases and then employ these embeddings to produce class activation maps (CAMs) [137] and Grad-CAMs [138]. While effective in some cases, these methods suffer from query ambiguity and lack robustness, as even small variations in the input text can significantly affect segmentation quality. To overcome these limitations, Extreme Weakly Supervised (EWS) Binary Semantic Segmentation [139] has been introduced as an alternative. This method combines unsupervised feature extractors like DINO [140, 141, 142] with minimal supervisionrequiring only a single annotated pixel per classand uses the STEGO (Self-supervised Transformer with Energy-based Graph Optimization) [143] loss to learn more robust features and to better align sparse annotations with patch-level features, resulting in segmentation maps that are competitive despite limited supervision. However, EWS is currently

This project has received funding from the European Union – European Commission under grant agreement No 101093003

restricted to binary segmentation tasks, which limits its applicability to NDM scenarios where multiple disaster-related categories (e.g., flooded regions, debris, infrastructure damage) must be segmented simultaneously. Extending this framework to multiclass segmentation requires not only adding one-pixel annotations for each class but also designing mechanisms to capture inter-class dependencies, ensuring that annotations for one class contribute to refining the segmentation of others. Developing such an extension would help address the challenges of data scarcity in NDM and improve the reliability and practicality of weakly supervised segmentation methods in real-world disaster response.

5.6.0.2. Advances beyond SOTA

AUTH tackles this problem by extending binary methodology to multiclass, tackling the challenges of inter-class dependencies with the introduction of a novel triplet loss [144, 145, 146]that takes into account the pixel annotations of the other classes. The paper with more details will be submitted to a conference: A. Apostolidis, M. Tzimas, V. Mygdalis, I. Pitas, "Semantic Image Segmentation with Multiclass Extreme Weak Supervision (MEWS)", technical report, 2025. The MEWS method significantly alleviates the challenges posed by the absence of extensive annotated datasets in natural disaster management (NDM). By employing sparse pixel annotations as class prototypes, the approach ensures precise initialization and effective clustering of class-specific features within the self-supervised DINO feature space. Building on this foundation, we extend the framework from binary to multiclass segmentation by incorporating the mean prototypes of each class, the global feature representation of the image, and the relationships with other class prototypes into a margin triplet loss formulation. This extension allows the model to explicitly enforce inter-class separability while maintaining intra-class consistency, a property that is essential in scenarios with multiple disaster-related categories. Furthermore, the dynamic computation of hyperparameters has been generalized to the multiclass setting, reducing the need for manual tuning and improving training stability. As a result, the extended approach not only retains the efficiency and robustness of the binary method but also enables scalable and accurate multiclass segmentation, thereby enhancing its applicability in real-world NDM tasks. Our proposed **MEWS(AUTH)** method achieves the best weakly supervised perfor-

Table 19. Comparison of multiclass semantic segmentation performance across various methods on the benchmark dataset Cityscapes. 4Px8I means 4 annotated pixels per class across 8 images.

Model	Mean	Void	Flat	Construction	Object	Nature	Sky	Human	Vehicle
Stego Supervised	63.46	61.04	88.95	75.25	1.25	79.55	79.60	50.72	71.34
Stego Unsupervised	54.18	60.19	79.68	68.77	0.33	76.76	76.73	8.88	62.07
CLIP-ES	40.14	68.05	22.10	54.12	10.96	70.14	60.18	8.01	27.47
ExCEL	41.90	32.60	80.07	47.80	9.14	55.49	51.85	11.34	46.87
DINO+Prototypes (4Px8I)	57.55 ± 0.78	60.37 ± 0.70	84.81 ± 0.91	70.42 ± 2.52	15.82 ± 1.44	74.36 ± 1.26	76.18 ± 2.08	17.70 ± 3.11	60.78 ± 2.42
MEWS (4Px8I)	61.19 ± 1.71	58.91 ± 1.20	88.37 ± 0.24	73.04 ± 0.97	15.62 ± 1.49	77.63 ± 0.80	77.77 ± 0.95	32.25 ± 9.11	65.92 ± 1.89
DINO+Prototypes (BEST)	58.07 ± 0.65	61.01 ± 0.35	84.31 ± 0.71	71.27 ± 1.75	17.53 ± 1.00	75.82 ± 0.77	76.82 ± 1.32	18.82 ± 5.35	61.59 ± 1.30
MEWS (BEST)	63.27 ± 0.81	60.62 ± 0.43	89.05 ± 0.11	74.69 ± 0.34	17.68 ± 1.61	79.14 ± 0.87	79.04 ± 0.49	40.07 ± 3.72	66.22 ± 1.15

mance on Cityscapes, with a mean IoU of 61.19% in the 4Pixels x 8Images setting and 63.27% in the *BEST* configuration. Compared to the DINO+Prototypes baseline, our method consistently improves segmentation quality, highlighting the benefit of integrating class prototypes, global features, and inter-class relationships.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

5.7. Neural Architecture Search and Knowledge Distillation for Semantic Image Segmentation on Big Wildfire Datasets

SOTA

When trained on limited data, large, complex models with millions of parameters are highly susceptible to overfitting. They may achieve high accuracy on the training set by memorizing specific features but fail to generalize to new, unseen fire incidents. This reduces their reliability for real-world deployment. A common approach to mitigate this is to use smaller, more efficient models, but manually designing an architecture that strikes the perfect balance between model capacity and generalization for a niche task like burnt area segmentation is a non-trivial challenge. Neural Architecture Search (NAS) is a principled and automated method for discovering optimal network architectures tailored to the constraints of any task.
State-of-the-art research on efficient DNN design often relies on either model compression or automated architecture optimization. Knowledge Distillation (KD) has been established as a powerful technique to transfer knowledge from large, complex teacher networks to smaller student networks while preserving segmentation accuracy [114, 147]. Neural Architecture Search (NAS) has also emerged as a method to automatically explore and optimize neural architectures, targeting improved efficiency and accuracy [148, 149]. However, existing NASKD pipelines are primarily developed for image classification tasks and are rarely adapted to dense prediction problems such as burnt area segmentation [150, 151]. Moreover, many studies rely on fixed teacherstudent settings, limiting the preservation of spatial consistency and boundary sharpness that are critical in wildfire monitoring. To the best of AUTHs knowledge, no comprehensive study combining NAS and KD for wildfire burnt area segmentation has been carried out so far.

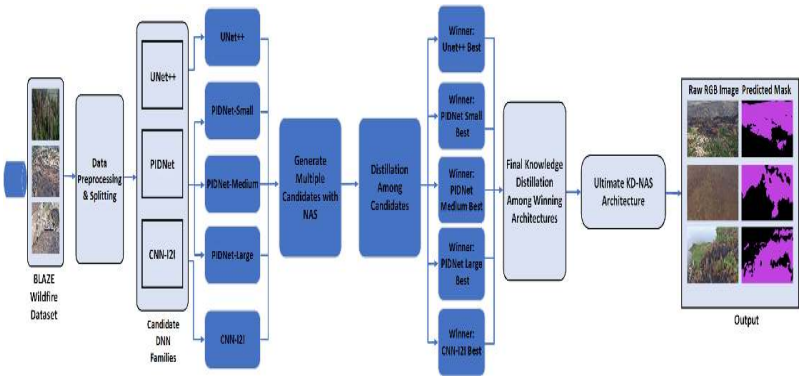


Figure 59. The KD-NAS Pipeline.

Advances beyond SOTA

AUTH proposes a novel KDNAS pipeline tailored for burnt area image segmentation. The approach integrates NAS to systematically search multiple DNN families (PIDNet Small [35], PIDNet Medium [35], PIDNet Large [35], UNet++ [152], and CNN-I2I BiseNet [153]) while optimizing

This project has received funding from the European Union – European Commission under grant agreement No 101093003

directly for the mean Intersection over Union (mIoU). The paper with more details is described in a conference paper [5]:

Eugenios Vlachos, Christos Papaioannidis and Ioannis Pitas (2025), "Neural Architecture Search and Knowledge Distillation for Semantic Image Segmentation on Big Wildfire Datasets", 33rd European Signal Processing Conference (EUSIPCO 2025)

Candidate architectures are first refined through NAS to reduce computational cost while maintaining accuracy. Next, KD is employed within and across model families, transferring rich per-pixel soft predictions from larger teacher models into smaller student models. Finally, the best-performing DNNs are distilled once more to converge on the ultimate winning lightweight architecture. Experimental evaluation on the Blaze dataset [1] demonstrated that this method yields a 62.3% reduction in trainable parameters while improving segmentation accuracy by +1.02% mIoU, enabling efficient deployment of UAV-based wildfire management systems in real-time environments. The details of the proposed method can be found in Figure 59, where the NAS-KD pipeline diagram is presented. The evaluation of several state-of-the-art region segmentation architectures on the Blaze dataset with respect to Mean IoU is presented on Table 20 along with the performance of optimized DNN architectures after KD application.

Table 20. Comparison of segmentation performance across all pipeline stages.

Model	Baseline			NAS			NAS + KD			Final KD-NAS		
	mIoU %	Epochs	Params	mIoU %	Epochs	Params	mIoU %	Epochs	Params	mIoU %	Epochs	Params
CNN-I2I BiseNet	74.82	120	18.4M	74.18	50	17.8M	74.18	50	17.8M	73.92	50	17.8M
PIDNet-Small	73.79	120	7.72M	74.22	50	6.97M	75.04	50	6.94M	**75.84**	50	**6.94M**
PIDNet-Medium	71.27	120	28.8M	71.07	50	19.6M	71.93	50	14.5M	72.94	50	14.5M
PIDNet-Large	70.13	120	37.3M	69.66	50	28.0M	71.08	50	19.3M	71.08	50	19.3M
UNet++	64.11	120	7.76M	65.47	50	1.93M	67.14	50	1.86M	68.21	50	1.86M

This project has received funding from the European Union – European Commission under grant agreement No 101093003

6. Conclusion

Deliverable D3.4 "Final report on algorithms for extreme data analytics" successfully encapsulates the critical advancements and research outcomes achieved in Task T3.5 during the period from M13 to M36 within the TEMA project. The innovative approaches developed throughout this reporting period have significantly contributed to the understanding and adaptability of AI models under extreme data conditions, particularly in the context of Natural Disaster Management (NDM).

The research efforts have yielded substantial progress in several areas, including the generation of synthetic datasets, the implementation of zero-shot learning for automatic labelling, and the establishment of robust frameworks for integrating multi-platform social media data. These advances not only enhance the reliability and interpretability of AI systems, but also ensure that they remain responsive to the dynamically changing landscapes of disaster scenarios. By strengthening data diversity, improving automated labelling accuracy, and enabling seamless integration of heterogeneous information sources, these developments directly contribute to OA2 increasing the accuracy of extreme data analysis algorithms.

Moreover, the introduction of explainable AI methods tailored for extreme data conditions has bolstered user trust and model transparency. By focusing on the cognitive aspects of model explanations, these methods allow end-users to engage more effectively with AI outputs, fostering a deeper understanding of model behaviour and decision-making processes. These advances directly support Objective OA1 increasing the trustworthiness of extreme data analysis algorithms. This contributes to enhanced accountability, transparency, and confidence in AI-assisted analysis of extreme data scenarios.

The continuous learning frameworks established during this period enable the AI models to evolve alongside incoming data, ensuring they maintain high performance even as conditions change. This adaptability is crucial for operational readiness in disaster response efforts, aligning with the overarching goals set forth in the TEMA project. In particular, these developments contribute directly to OA2 and OA3 by enhancing both the accuracy and responsiveness of extreme data analysis algorithms. Through continuous model refinement and real-time data integration, TEMAs approach supports the development of novel semantic analysis methods that surpass the state of the art in precision (OA2) while simultaneously improving processing efficiency and analytical speed on equivalent hardware (OA3).

Overall, the deliverable has fulfilled the key performance indicators (KPIs) and objectives defined for T3.5. Collective research outputs, reflected in nine peer-reviewed publications, not only advance the scientific knowledge base but also contribute significantly to the practical applications of AI in managing extreme data scenarios.

As we move forward, the insights gained from this deliverable will improve the next phases of the TEMA project, enhancing the development of scalable, trustworthy, and efficient AI systems for emergency management. The ongoing dissemination of TEMA's research findings will further promote the integration of advanced analytics into real-world disaster response strategies, ultimately improving resilience and preparedness in the face of natural disasters.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

References

[1] M. Siavrakas, C. Papaioannidis, and I. Pitas, "Blaze: A dataset for wildfire and burnt area uav image classification and segmentation," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*, pp. 1960–1965, 2025.

[2] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.

[3] A. Kaimakamidis and I. Pitas, "Leveraging collective knowledge for forest fire classification," in *2024 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–8, IEEE, 2024.

[4] A. Kaimakamidis and I. Pitas, "Fcl-vit: Task-aware attention tuning for continual learning," *Pattern Recognition Letters*, 2025.

[5] E. Spatharis, C. Papaioannidis, and I. Pitas, "Neural architecture search and knowledge distillation for semantic image segmentation on big wildfire datasets," *3rd European Signal Processing Conference (EUSIPCO)*, 2025.

[6] L. Yang *et al.*, "Diffusion models: A comprehensive survey of methods and applications," 2022. arXiv 2209.00796.

[7] Z. Ma *et al.*, "Generative deep learning for data generation in natural hazard analysis," *Artificial Intelligence Review*, 2024.

[8] P. Alimisis *et al.*, "Advances in diffusion models for image data augmentation: A review ," 2024. arXiv 2407.04103.

[9] Y. Lim *et al.*, "Data augmentation using diffusion models to enhance parameter inference ," *Physical Review E*, 2025.

[10] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023. Available at arXiv.

[11] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, 2022. DALL-E 2 system description on text-to-image generation.

[12] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, K. Seyed Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. Gontijo Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, "Photorealistic text-to-image diffusion models with deep language understanding," in *NeurIPS 2022*, 2022. arXiv preprint arXiv:2205.11487.

[13] Black Forest Labs, "Flux (text-to-image model) hugging face entry," 2025. Model description and licensing information for Flux.1 Kontext [dev].

[14] Stability AI, "Stable diffusion 3: Research paper," 2024. Official announcement and model description for Stable Diffusion 3.

[15] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," in *ICLR 2023*, 2023. Also available as arXiv preprint arXiv:2209.14988.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[16] OpenAI, "Gpt-4 technical report," *CoRR*, vol. abs/2303.08774, 2023.

[17] R. Anil, A. M. Dai, O. Firat, M. Johnson, D. Lepikhin, A. Passos, S. Shakeri, E. Taropa, P. Bailey, Z. Chen, and Y. ..., Wu, "Palm 2 technical report," *arXiv preprint arXiv:2305.10403*, 2023.

[18] Anthropic, "Model card and evaluations for claude models," 2023. Online model documentation.

[19] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de Las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, "Mistral 7b," *CoRR*, vol. abs/2310.06825, 2023.

[20] M. Deitke, C. Clark, S. Lee, R. Tripathi, Y. Yang, J. S. Park, M. Salehi, N. Muennighoff, K. Lo, L. Soldaini, J. Lu, T. Anderson, E. Bransom, K. Ehsani, H. Ngo, Y. Chen, A. Patel, M. Yatskar, C. Callison-Burch, A. Head, R. Hendrix, F. Bastani, E. VanderBilt, N. Lambert, Y. Chou, A. Chheda, J. Sparks, S. Skjonsberg, M. Schmitz, A. Sarnat, B. Bischoff, P. Walsh, C. Newell, P. Wolters, T. Gupta, K.-H. Zeng, J. Borchardt, D. Groeneveld, C. Nam, S. Lebrecht, C. Wittlif, C. Schoenick, O. Michel, R. Krishna, L. Weihs, N. A. Smith, H. Hajishirzi, R. Girshick, A. Farhadi, and A. Kembhavi, "Molmo and pixmo: Open weights and open data for state-of-the-art vision-language models," *arXiv preprint arXiv:2409.17146*, 2024.

[21] OpenAI, "Gpt-5 system card," August 2025. System card describing GPT-5s architecture and capabilities.

[22] J. Chen, H. Guo, K. Yi, B. Li, and M. Elhoseiny, "Visualgpt: Data-efficient adaptation of pretrained language models for image captioning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18030–18040, June 2022.

[23] L. Qin, W. Wang, Q. Chen, and W. Che, "CLIPText: A new paradigm for zero-shot text classification," in *Findings of the Association for Computational Linguistics: ACL 2023*, (Toronto, Canada), pp. 1077–1088, Association for Computational Linguistics, July 2023.

[24] comfyanonymous, "Comfyui: A powerful and modular diffusion model gui, api, and backend with a graph/nodes interface," 2023. Accessed: 2025-11-04.

[25] B. L. Sharma, "An introduction to comfyui for stable diffusion," 2025. LearnOpenCV blog.

[26] M. blog, "Sharing models and custom nodes in comfyui," 2025.

[27] lllyasviel, "Fooocus inpainting model," 2023. Model card and patches for inpainting via Fooocus; open-rail license.

[28] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *ICCV Workshops (AIM Workshop), 2021*, 2021. Also available as arXiv preprint arXiv:2107.10833.

[29] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Fulé, and E. Blasch, "The flame dataset: Aerial imagery pile burn detection using drones (uavs)," 2020.

[30] T. Toulouse, L. Rossi, A. Campana, T. Celik, and M. A. Akhloufi, "Computer vision for wildfire research: An evolving image dataset for processing and analysis," *Fire Safety Journal*, vol. 92, pp. 188–194, 2017.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[31] H. Wang, S. P. H. Boroujeni, X. Chen, A. Bastola, H. Li, W. Zhu, and A. Razi, "Flame diffuser: Wildfire image synthesis using mask guided diffusion," in *2024 IEEE International Conference on Big Data (BigData)*, pp. 6171–6179, IEEE, 2024.

[32] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "Airsim: High-fidelity visual and physical simulation for autonomous vehicles," *CoRR*, vol. abs/1705.05065, 2017.

[33] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*, pp. 1–16, PMLR, 2017.

[34] E. Spatharis, C. Papaioannidis, and I. Pitas, "Unrealfire: Synthetic annotated image creation pipeline for wildfire segmentation," *IEEE International Conference on Image Processing Workshop, IEEE ICIPW*, 2025.

[35] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," 2023.

[36] J. Chung, S. Hyun, and J.-P. Heo, "Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8795–8805, June 2024.

[37] S. Liu, Z. Zhang, F. Zhang, S. Wang, *et al.*, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection," *arXiv preprint arXiv:2303.05499*, 2023.

[38] K. Chen, J. Wang, J. Pang, Y. Cao, *et al.*, "MMDetection: Openmmlab object detection toolbox and benchmark." https://github.com/open-mmlab/mmdetection, 2019–2025.

[39] X. Zou, B. Zhou, S. Wang, F. Zhang, *et al.*, "Segment everything everywhere all at once," *arXiv preprint arXiv:2304.06718*, 2023.

[40] L. Ke, M. Ye, M. Danelljan, Y. Liu, , *et al.*, "Segment anything in high quality," *arXiv preprint arXiv:2306.01567*, 2023.

[41] H. Zhang, E. Xie, Z. Li, P. Sun, *et al.*, "Openseed: Unified open-vocabulary segmentation and detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[42] W. Xu, L. Ke, H. Zhao, and M. Z. Shou, "Odise: Open-vocabulary panoptic segmentation with diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[43] J. Wang, B. Zhou, E. Xie, H. Zhang, *et al.*, "A survey on open-vocabulary detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. arXiv:2305.12346.

[44] R. Solovyev, W. Wang, and T. Gabruseva, "Weighted boxes fusion: Ensembling boxes from different object detection models," *Image and Vision Computing*, pp. 1–6, 2021.

[45] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *European conference on computer vision (ECCV)*, pp. 740–755, Springer, 2014.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[46] Z. Wang and X. Ye, "Social media analytics for natural disaster management," vol. 32, no. 1, pp. 49–72.

[47] B. Resch, F. Usländer, and C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," *Cartography and Geographic Information Science*, vol. 45, no. 4, pp. 362–376, 2018.

[48] T. R. Hannigan and G. Casasnovas, "Topic modeling in management research: Rendering new theory from textual data," *Academy of Management Annals*, vol. 13, no. 2, pp. 586–632, 2019.

[49] Z. Yin, L. Cao, J. Han, C. Zhai, and T. Huang, "A geographical topic model for social media," in *Proceedings of the 20th international conference on World wide web*, pp. 247–256, 2011.

[50] C. González-Pizarro and G. Carenini, "Neural multimodal topic modeling: A comprehensive evaluation," *arXiv preprint arXiv:2403.17308*, 2024.

[51] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pp. 1079–1088, Association for Computing Machinery.

[52] S. A. Shah, S. B. Yahia, K. McBride, A. Jamil, and D. Draheim, "Twitter streaming data analytics for disaster alerts," in *2021 2nd International Informatics and Software Engineering Conference (IISEC)*, pp. 1–6, IEEE.

[53] J. Ray Chowdhury, C. Caragea, and D. Caragea, "On Identifying Hashtags in Disaster Twitter Data," vol. 34, no. 01, pp. 498–506.

[54] C. Havas and B. Resch, "Portability of semantic and spatialtemporal machine learning methods to analyse social media for near-real-time disaster monitoring," vol. 108, no. 3, pp. 2939–2969.

[55] X. Zhou and L. Chen, "Event detection over twitter social media streams," vol. 23, no. 3, pp. 381–400.

[56] L. Huang, P. Shi, H. Zhu, and T. Chen, "Early detection of emergency events from social media: A new text clustering approach," vol. 111, no. 1, pp. 851–875.

[57] C. J. Powers, A. Devaraj, K. Ashqeen, A. Dontula, A. Joshi, J. Shenoy, and D. Murthy, "Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach," vol. 3, no. 1, p. 100164.

[58] R. Koshy and S. Elango, "Multimodal tweet classification in disaster response systems using transformer-based bidirectional attention model," vol. 35, no. 2, pp. 1607–1627.

[59] J. Li, Y. Wang, and W. Li, "MGMP: Multimodal Graph Message Propagation Network for Event Detection," in *MultiMedia Modeling* (B. Þór Jónsson, C. Gurrin, M.-T. Tran, D.-T. Dang-Nguyen, A. M.-C. Hu, B. Huynh Thi Thanh, and B. Huet, eds.), pp. 141–153, Springer International Publishing.

[60] D. Adwaith, A. K. Abishake, S. V. Raghul, and E. Sivasankar, "Enhancing multimodal disaster tweet classification using state-of-the-art deep learning networks," vol. 81, no. 13, pp. 18483–18501.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[61] S. Madichetty, S. M, and S. Madisetty, "A RoBERTa based model for identifying the multi-modal informative tweets during disaster," vol. 82, no. 24, pp. 37615–37633.

[62] S. Z. Razavi and M. Rahbari, "Understanding reactions to natural disasters: A text mining approach to analyze social media content," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, (Paris, France), pp. 1–7, IEEE, 2020.

[63] J. P. Stimpson, A. Srivastava, K. Tamirisa, J. K. Kaholokula, and A. N. Ortega, "Crisis communication about the maui wildfires on tiktok: Content analysis of engagement with maui wildfire–related posts over 1 year," *JMIR Formative Research*, vol. 9, pp. e67515–e67515, 2025.

[64] N. Vračević, S. Schmidt, M. Keskin, D. Hanny, and B. Resch, "More than just tweets: the potential of alternative geo-social media data for disaster management," *Soc. Netw. Anal. Min.*, vol. 15, July 2025.

[65] C. Havas and B. Resch, "Portability of semantic and spatialtemporal machine learning methods to analyse social media for near-real-time disaster monitoring," *Natural Hazards*, vol. 108, pp. 2939–2969, 2021.

[66] D. Hanny and B. Resch, "Multimodal geoai: An integrated spatio-temporal topic-sentiment model for the analysis of geo-social media posts for disaster management," *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, p. 104540, 2025.

[67] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," *ACM Computing Surveys*, vol. 47, no. 4, p. 67, 2015.

[68] Reddit, *Reddit API Documentation*, 2025. Accessed: 12 September 2025.

[69] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn, "The pushshift reddit dataset," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, pp. 830–839, 2020.

[70] Telegram, *Telegram API and Bot Documentation*, 2025. Accessed: 12 September 2025.

[71] Mastodon, *Mastodon API and ActivityPub Documentation*, 2025. Accessed: 12 September 2025.

[72] Bluesky, *AT Protocol Documentation*, 2025. Accessed: 12 September 2025.

[73] TikTok, *TikTok Research API Documentation*, 2025. Accessed: 12 September 2025.

[74] M. Graham and S. De Sabbata, "Mapping information wealth and poverty: The geography of big data," *Dialogues in Human Geography*, vol. 6, no. 3, pp. 300–306, 2016.

[75] X. Zhang, Y. Malkov, O. Florez, S. Park, B. McWilliams, J. Han, and A. El-Kishky, "TwHIN-BERT: A Socially-Enriched Pre-trained Language Model for Multilingual Tweet Representations at Twitter," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, pp. 5597–5607, Association for Computing Machinery.

[76] B. Hopkins and J. G. Skellam, "A new method for determining the type of distribution of plant individuals," *Annals of Botany*, vol. 18, pp. 213–227, 04 1954.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[77] L. McInnes, J. Healy, N. Saul, and L. GroSSberger, "Umap: Uniform manifold approximation and projection," *Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.

[78] R. Tang, L. Liu, A. Pandey, Z. Jiang, G. Yang, K. Kumar, P. Stenetorp, J. Lin, and F. Ture, "What the daam: Interpreting stable diffusion using cross attention," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (Toronto, Canada), Association for Computational Linguistics, July 2023.

[79] Y. Li, H. Liu, Q. Wu, F. Mu, J. Yang, J. Gao, C. Li, and Y. J. Lee, "Gligen: Open-set grounded text-to-image generation," *arXiv*, vol. arXiv:2301.07093, 2023. Preprint.

[80] H. Chefer, Y. Alaluf, Y. Vinker, L. Wolf, and D. Cohen-Or, "Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models," *arXiv*, vol. arXiv:2301.13826, 2023. Preprint.

[81] W. Baek, "Attention-map-diffusers: Cross attention map visualization for hugging face diffusers," 2023. Accessed: 2025-08-20.

[82] J. Lages, "Diffusers-interpret: Model explainability for diffusers," 2023. Accessed: 2025-08-20.

[83] S. Hong, G. Lee, W. Jang, and S. Kim, "Improving sample quality of diffusion models using self-attention guidance," *arXiv preprint arXiv:2210.00939*, 2022. Accessed: 2025-08-20.

[84] H. Li, C. C. Loy, P. H. S. Torr, V. Tresp, and J. Gu, "Self-discovering interpretable diffusion latent directions for responsible text-to-image generation," 2024. Accessed: 2025-08-20.

[85] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[86] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int. J. Comput. Vision*, vol. 128, no. 2, 2020.

[87] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation," *PLOS ONE*, vol. 10, pp. 1–46, 07 2015.

[88] F. Bley, S. Lapuschkin, W. Samek, and G. Montavon, "Explaining predictive uncertainty by exposing second-order effects," *Pattern Recognition*, vol. 160, p. 111171, 2025.

[89] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, 2017.

[90] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[91] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[92] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*, 2018.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[93] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[94] S. Gururaj, L. Grüne, W. Samek, S. Lapuschkin, and L. Weber, "Relevance-driven Input Dropout: an Explanation-guided Regularization Technique," *Preprint arXiv:2505.21595*, 2025.

[95] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, 2017.

[96] G. Ümit Yolcu, M. Weckbecker, T. Wiegand, W. Samek, and S. Lapuschkin, "DualXDA: Towards Sparse, Efficient and Explainable Data Attribution in Large AI Models," *Preprint arXiv:2402.12118v2*, 2025.

[97] A. Schioppa, P. Zablotskaia, D. Vilar, and A. Sokolov, "Scaling up influence functions," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 8179–8186, Jun. 2022.

[98] R. B. Grosse, J. Bae, C. Anil, N. Elhage, A. Tamkin, A. Tajdini, B. Steiner, D. Li, E. Durmus, E. Perez, E. Hubinger, K. Lukovsiut.e, K. Nguyen, N. Joseph, S. McCandlish, J. Kaplan, and S. Bowman, "Studying large language model generalization with influence functions," *ArXiv*, vol. abs/2308.03296, 2023.

[99] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the Visualization of What a Deep Neural Network Has Learned," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, 2017.

[100] C. Agarwal and A. Nguyen, "Explaining image classifiers by removing input features using generative models," in *Computer Vision – ACCV 2020* (H. Ishikawa, C.-L. Liu, T. Pajdla, and J. Shi, eds.), 2021.

[101] A.-F. Cabouat, T. He, P. Isenberg, and T. Isenberg, "Pondering the reading of visual representations." Preprint, 2023.

[102] T. Labarta, N. Hoang, K. Weitz, W. Samek, S. Lapuschkin, and L. Weber, "See What I Mean? CUE: A Cognitive Model of Understanding Explanations," *IJCAI 2025 Workshop on Explainable Artificial Intelligence (XAI)*, 2025.

[103] B. Kim, M. Wattenberg, J. Gilmer, C. J. Cai, J. Wexler, F. B. Viégas, and R. Sayres, "Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV).," in *ICML*, vol. 80, pp. 2673–2682, 2018.

[104] R. Achtibat, M. Dreyer, I. Eisenbraun, S. Bosse, T. Wiegand, W. Samek, and S. Lapuschkin, "From Attribution Maps to Human-Understandable Explanations through Concept Relevance Propagation," *Nature Machine Intelligence*, vol. 5, pp. 1006–1019, 2023.

[105] F. Pahde, T. Wiegand, S. Lapuschkin, and W. Samek, "Ensuring Medical AI Safety: Explainable AI-Driven Detection and Mitigation of Spurious Model Behaviorand Associated Data," *Machine Learning*, 2025.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[106] F. Pahde, M. Dreyer, W. Samek, and S. Lapuschkin, "Reveal to Revise: An Explainable AI Life Cycle for Iterative Bias Correction of Deep Models," in *Medical Image Computing and Computer Assisted Intervention – MICCAI*, 2023.

[107] M. Dreyer, L. Hufe, J. Berend, T. Wiegand, S. Lapuschkin, and W. Samek, "From What to How: Attributing CLIP's Latent Components Reveals Unexpected Semantic Reliance," *Preprint arXiv:22505.20229*, 2025.

[108] B. Puri, A. Jain, E. Golimblevskaia, P. Kahardipraja, T. Wiegand, W. Samek, and S. Lapuschkin, "FADE: Why bad descriptions happen to good features," in *Findings of the Association for Computational Linguistics: ACL 2025*, July 2025.

[109] M. Dreyer, R. Achtibat, W. Samek, and S. Lapuschkin, "Understanding the (Extra-)Ordinary: Validating Deep Model Decisions with Prototypical Concept-based Explanations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024.

[110] K. Vani *et al.*, "Deep learning based forest fire classification and detection in satellite images," in *2019 11th international conference on advanced computing (ICoAC)*, pp. 61–65, IEEE, 2019.

[111] S. Khan and A. Khan, "Ffirenet: Deep learning based forest fire classification and detection in smart cities," *Symmetry*, vol. 14, no. 10, p. 2155, 2022.

[112] A. Akagic and E. Buza, "Lw-fire: A lightweight wildfire image classification with a deep convolutional neural network," *Applied Sciences*, vol. 12, no. 5, p. 2646, 2022.

[113] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," in *International Conference on Learning Representations (ICLR)*, 2015.

[114] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in *NIPS Deep Learning and Representation Learning Workshop*, 2015.

[115] T. R. Gadekallu, Q.-V. Pham, T. Huynh-The, S. Bhattacharya, P. K. R. Maddikunta, and M. Liyanage, "Federated learning for big data: A survey on opportunities, applications, and future directions," *arXiv preprint arXiv:2110.04160*, 2021.

[116] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[117] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[118] J. Kirkpatrick, R. Pascanu, N. C. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, pp. 3521 – 3526, 2016.

[119] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 2935–2947, 2016.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[120] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "icarl: Incremental classifier and representation learning," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5533–5542, 2016.

[121] S. Yan, J. Xie, and X. He, "Der: Dynamically expandable representation for class incremental learning," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3013–3022, 2021.

[122] A. Douillard, A. Ram'e, G. Couairon, and M. Cord, "Dytox: Transformers for continual learning with dynamic token expansion," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9275–9285, 2021.

[123] M. Xue, H. Zhang, J. Song, and M. Song, "Meta-attention for vit-backed continual learning," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 150–159, 2022.

[124] Z. Wang, Z. Zhang, C.-Y. Lee, H. Zhang, R. Sun, X. Ren, G. Su, V. Perot, J. G. Dy, and T. Pfister, "Learning to prompt for continual learning," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 139–149, 2021.

[125] Z. Wang, Z. Zhang, S. Ebrahimi, R. Sun, H. Zhang, C.-Y. Lee, X. Ren, G. Su, V. Perot, J. G. Dy, and T. Pfister, "Dualprompt: Complementary prompting for rehearsal-free continual learning," *ArXiv*, vol. abs/2204.04799, 2022.

[126] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," 2009.

[127] A. Kaimakamidis, I. Mademlis, and I. Pitas, "Collaborative knowledge distillation via a learning-by-education node community," *arXiv preprint arXiv:2410.00074*, 2024.

[128] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, pp. 1273–1282, PMLR, 2017.

[129] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, and C. Zhang, "Fedproto: Federated prototype learning across heterogeneous clients," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, pp. 8432–8440, 2022.

[130] L. Qi, H. Chen, H. Zou, S. Chen, X. Zhang, and H. Chen, "Decentralized federated learning with prototype exchange," *Mathematics*, vol. 13, no. 2, p. 237, 2025.

[131] M. Wang, A. Bodonhelyi, E. Bozkir, and E. Kasneci, "Turbosvm-fl: Boosting federated learning through svm aggregation for lazy clients," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 15546–15554, 2024.

[132] C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie, "Communication-efficient federated learning via knowledge distillation," *Nature communications*, vol. 13, no. 1, p. 2032, 2022.

[133] V. Mygdalis and I. Pitas, "Hyperspherical class prototypes for adversarial robustness," *Pattern Recognition*, vol. 125, p. 108527, 2022.

[134] Y. Lin, M. Chen, W. Wang, B. Wu, K. Li, B. Lin, H. Liu, and X. He, "Clip is also an efficient segmenter: A text-driven approach for weakly supervised semantic segmentation," 2023.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[135] Z. Yang, Y. Meng, K. Fu, F. Tang, S. Wang, and Z. Song, " Exploring CLIPs Dense Knowledge for Weakly Supervised Semantic Segmentation ," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Los Alamitos, CA, USA), pp. 20223–20232, IEEE Computer Society, June 2025.

[136] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," 2021.

[137] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, 2016.

[138] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, p. 336359, Oct. 2019.

[139] M. Tzimas, V. Mygdalis, C. Papaioannidis, and I. Pitas, "Extreme weakly supervised binary semantic image segmentation via one-pixel supervision," *Elsevier Pattern Recognition*, 2025.

[140] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," 2021.

[141] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. HAZIZA, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "DINOv2: Learning robust visual features without supervision," *Transactions on Machine Learning Research*, 2024. Featured Certification.

[142] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, F. Massa, D. Haziza, L. Wehrstedt, J. Wang, T. Darcet, T. Moutakanni, L. Sentana, C. Roberts, A. Vedaldi, J. Tolan, J. Brandt, C. Couprie, J. Mairal, H. Jégou, P. Labatut, and P. Bojanowski, "Dinov3," 2025.

[143] M. Hamilton, Z. Zhang, B. Hariharan, N. Snavely, and W. T. Freeman, "Unsupervised semantic segmentation by distilling feature correspondences," 2022.

[144] D. P. Vassileios Balntas, Edgar Riba and K. Mikolajczyk, "Learning local feature descriptors with triplets and shallow convolutional neural networks," in *Proceedings of the British Machine Vision Conference (BMVC)* (E. R. H. Richard C. Wilson and W. A. P. Smith, eds.), pp. 119.1–119.11, BMVA Press, September 2016.

[145] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 815823, IEEE, June 2015.

[146] E. Paul, M. Rampavan, S. Beerukuri, S. Vinnakota, and V. Nelakurthi, "Person re-identification using vision transformer and centroid triplet loss," *Multimedia Tools and Applications*, vol. 83, pp. 1–12, 08 2024.

[147] L. Wang and K.-J. Yoon, "Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, p. 30483068, June 2022.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[148] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 6105–6114, PMLR, 09–15 Jun 2019.

[149] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," 2019.

[150] C. Li, J. Peng, L. Yuan, G. Wang, X. Liang, L. Lin, and X. Chang, "Blockwisely supervised neural architecture search with knowledge distillation," 2020.

[151] Z. Zheng and G. Kang, "Model compression with nas and knowledge distillation for medical image segmentation," in *2021 4th International Conference on Data Science and Information Technology*, pp. 173–176, 2021.

[152] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[153] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 325–341, 2018.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

A. Stable Diffusion XL Pipeline

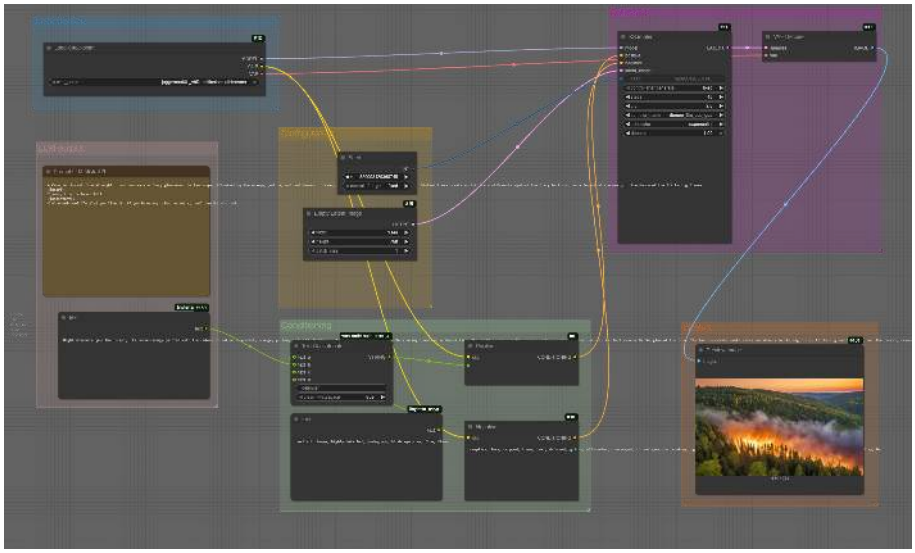


Figure 6o. ComfyUI generation pipeline using Mistral and Stable DiffusionXL [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

B. Stable Diffusion XL prompts

Figure 61. "prompt": "<|system|> You are a Stable Diffusion prompt generator. <|user|> Come up with a description of a realistic finnish forest fire at night. <|assistant|> A night view of a forest fire in the finnish taiga. Many trees are on fire, and the smoke rises <|user|> How about another one? <|assistant|>" [image generated by ATOS]

Figure 62. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a realistic finnish forest fire at dawn. <|assistant|> At dawn the sun rising on a forest fire in the finnish taiga. Many trees are on fire, and the smoke rises. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 63. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a realistic finnish forest fire at sunset. <|assistant|> At sunset the sun is setting on the horizon while there is a fire in the finnish taiga. Many trees are on fire, and the smoke rises. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

Figure 64. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a realistic finnish forest fire while raining. <|assistant|> Is raining heavily on a forest in the finnish taiga while a raging forest fire is consuming the vegetation. Many trees are on fire, and the smoke rises. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 65. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a realistic canadian forest fire at autumn. <|assistant|> In a forest fire in canada, there are several fire hot spots on the forest burning the red, yellow and orange leaves trees. Many trees are on fire, and the smoke rises. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

Figure 66. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a realistic forest fire near a small town at night. <|assistant|> a small town with several houses, there is a forest fire in the background that light up the night sky as the flame rises. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 67. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a realistic forest fire near a small town. <|assistant|> a small town with several houses, there is a forest fire in the background rages while consuming the scenery. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

Figure 68. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of the beginning of a realistic forest fire. <|assistant|> A dense forest with several small spots that are on fire, as a blazing inferno is starting. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 69. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of the beginning of a realistic forest fire at dawn. <|assistant|> the sun is rising while some flickers of fire can be seen on the forest as a forest fire starts. <|user|> How about another one? <|assistant|>" [image generated by ATOS]

Figure 70. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of burned down forest with some flames still active. <|assistant|> there can be seen a charred and burned down forest ravaged by the fire, the trees are filled with ashes and some flames can still be seen <|user|> How about another one? <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 71. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a forest fire in a winter snowy forest. <|assistant|> On a snowy forest scene a wildfire can be seen ravaging the top of the tress <|user|> How about another one? <|assistant|>" [image generated by ATOS]

Figure 72. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of town in Austria after heavy rain and floods, damaged buildings. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 73. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a town in Austria in a rainy day after heavy rain and floods, damaged buildings. <|assistant|>" [image generated by ATOS]

Figure 74. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a street in Austria in a rainy day after heavy rain and floods, damaged buildings, trapped cars. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 75. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a street in Greece on a rainy day after heavy rain and floods, damaged buildings, trapped cars. <|assistant|>" [image generated by ATOS]

Figure 76. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a town in Greece on a sunny day after heavy rain and floods, damaged buildings, trapped cars. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 77. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded river in the greek hillside, after heavy rain and floods. <|assistant|>" [image generated by ATOS]

Figure 78. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded river in the Austrian hillside, after heavy rain and floods. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 79. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded river in the Austrian hillside, during heavy rain. <|assistant|>" [image generated by ATOS]

Figure 80. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a Greek forest with a flooded river, during heavy rain. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 81. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded river at night on a Greek town. <|assistant|>" [image generated by ATOS]

Figure 82. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded river at night on a Austrian town, trapped cards, damaged bridge. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 83. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded river at night on a Greek town, trapped cards, damaged bridge. <|assistant|>" [image generated by ATOS]

Figure 84. "prompt": "<|system|> You are a Stable Diffusion prompt generator.<|user|> Come up with a description of a flooded street at night on a Greek town, trapped cards, debris. <|assistant|>" [image generated by ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

C. Flux.1-dev Pipeline

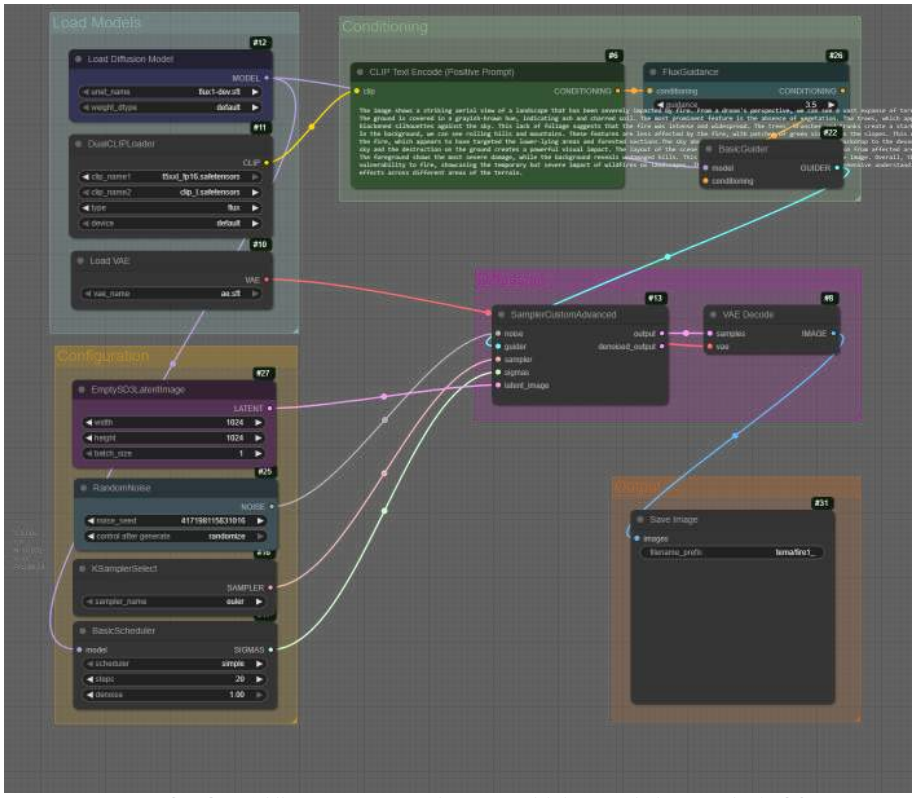


Figure 85. ComfyUI generation pipeline using Molmo and Flux [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

D. Flux.1.dev prompts

Image description of Figure 86

The image shows an aerial view of a forested area following a recent fire. The scene is dominated by a large body of water, likely a lake or river, which occupies the majority of the frame. The water appears dark blue and still, with no visible movement. The most striking feature of the image is the smoke rising from the center of the forest. It's white and billowing upwards, creating a stark contrast against the blue sky and green trees. The smoke suggests that a fire has recently occurred in the area, and it's still active, as evidenced by its intensity and the way it's spreading. The forest itself is predominantly green, with a mix of tall, healthy trees and some that appear to be damaged or dead. This arrangement is typical of many forest ecosystems and provides an interesting visual contrast with the water. On the right side of the image, there's clear evidence of human intervention. A clear-cut area is visible, showing where trees have been removed. This area appears to be in a state of transition, with some regrowth visible but still a significant amount of exposed soil and vegetation. The overall layout of the scene is quite striking. The smoke from the fire adds a dynamic element to the scene, suggesting ongoing changes and adjustments in the environment. This image provides a powerful visual representation of the impact of fire on forest ecosystems and the balance between natural processes and human influence.

(a) Real image of a prescribed burn in finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 86. Real origin fire image and synthetic resulting image comparison num.1

Image description of Figure 87

The image shows a striking aerial view of a forest after a recent fire. The landscape is dominated by a vast expanse of green forest, with the trees appearing lush and vibrant, especially in the foreground. In the center of the image, there's a large, dense cloud of smoke rising from the forest. This smoke is thick and billowing, creating a stark contrast against the green backdrop. The smoke appears to be spreading across the image, likely due to wind currents. To the left side of the image, a body of water is visible. It could be a lake or a river, and its surface is reflecting the smoke from the forest. This adds another layer of visual interest to the scene. The layout of the image is quite interesting. The forest dominates the lower portion, while the smoke rises prominently in

This project has received funding from the European Union – European Commission under grant agreement No 101093003

the center. The upper part of the image shows more of the surrounding area, including what appears to be a clear sky. The scene is devoid of any visible human structures or people, which emphasizes the raw, untouched nature of the forest. The color palette is predominantly green, with the smoke providing a stark contrast. Overall, this image captures a powerful post-fire landscape, showcasing nature's resilience and the immediate impact of wildfires on forest ecosystems.

(a) Real image of a prescribed burn in finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 87. Real origin fire image and synthetic resulting image comparison num.2

Image description of Figure 88

The image shows the aftermath of a forest fire viewed from a drone. The scene is dominated by a large area of smoke rising from the center, creating a hazy atmosphere. The forest is partially burnt, with many trees still standing but showing signs of damage. The landscape is a mix of green and brown, with some areas of grass and exposed soil visible. The smoke is thick and billowing, obscuring parts of the view and creating a stark contrast against the blue sky. In the background, there's a body of water visible, likely a lake or river. The water appears dark blue, and there are some small islands or landmasses visible within it. The drone is positioned above the scene, providing a bird's-eye view of the entire area. The sky is a light blue with scattered white clouds, indicating fair weather conditions. The layout of the scene is as follows: 1. A foreground dominated by the burnt and smoking forest 2. A body of water (lake or river) in the background 3. A sky with scattered white clouds The overall color palette includes shades of blue, white, gray, brown, green, and black, with the smoke adding a hazy, atmospheric quality to the image. This scene captures the immediate aftermath of a forest fire, showcasing the immediate impact on the landscape and the ongoing process of recovery or containment.

Image description of Figure 89

The image shows the aftermath of a forest fire viewed from a drone. The scene is dominated by thick, billowing smoke that obscures much of the view. The fire appears to have burned through a clear-cut area, leaving behind a stark contrast between the charred remains and the surrounding intact forest. In the foreground, we can see the aftermath of the fire. The ground is bare, with scattered patches of burnt vegetation and dirt. This clear-cut area is surrounded by dense, green forest, creating a striking juxtaposition between the damaged and undamaged sections of the forest. The smoke is particularly dense in some areas, making it difficult to see the full extent

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a prescribed burn in Finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 88. Real origin fire image and synthetic resulting image comparison num.3

of the damage. However, we can observe that the fire has affected a wide swath of the forest, leaving behind a path of destruction. The layout of the scene is as follows: 1. A clear-cut area in the foreground, where the fire has burned through 2. Green forest surrounding the cleared area, creating a stark contrast 3. Thick, billowing smoke dominating much of the image 4. The overall scene captures the immediate aftermath of a forest fire, showcasing the long-lasting impact on the landscape This image provides a powerful visual representation of the destructive power of forest fires and the rapid rate at which they can transform a landscape. It also highlights the importance of forest management and fire prevention strategies to protect both human communities and natural ecosystems.

Image description of Figure 90

The drone's camera captures a striking aerial view of a forest fire's aftermath. The scene is dominated by a large plume of smoke rising from the centre, obscuring much of the view. The sky above is a clear blue with scattered white clouds, providing a stark contrast to the smoky atmosphere below. In the foreground, we can see a path cutting through the charred landscape. This path is likely the result of firefighting efforts, possibly a firebreak or access route. The ground is covered with a mix of brown and green grass, with some areas appearing more barren than others, especially towards the edges of the visible area. The forest itself is visible in the background, with a variety of tree species present. Some trees appear to be blackened and charred, while others, particularly those on the edges of the frame, seem to have survived the fire. This mix of affected and unaffected trees is typical of forest fires, where some areas are heavily impacted while others remain largely untouched. The overall layout of the scene suggests this is a relatively large fire, as evidenced by the significant amount of smoke and the extensive area affected. The drone's perspective provides a unique view of how the fire has changed the landscape, highlighting the contrast between the fire's destructive power and the forest's structure. This image offers a powerful visual representation of the impact of wildfires, showcasing both the natural beauty of the forest and the devastating effects of fire on the environment.

Image description of Figure 91

The image shows the aftermath of a forest fire viewed from a drone. The landscape is dominated

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a prescribed burn in Finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 89. Real origin fire image and synthetic resulting image comparison num.4

(a) Real image of a prescribed burn in Finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 90. Real origin fire image and synthetic resulting image comparison num.5

by a large, smoky plume that extends from the bottom right corner towards the top right, creating a hazy atmosphere. In the foreground, we can see the charred remains of a once-thriving forest. The ground is dark and barren, with scattered trees that have been reduced to ash. Some trees still stand, but their leaves are singed and brown, indicating the severity of the fire. A narrow road cuts through the scene, likely used by firefighters during the blaze. This road is one of the few visible structures in the area, emphasizing the vastness of the forest that has been affected. The fire appears to have burned through the center of the image, leaving a clear path of destruction. The smoke from the blaze is thick and billowing, obscuring parts of the sky and

This project has received funding from the European Union – European Commission under grant agreement No 101093003

adding to the overall sense of devastation. In the background, we can see a dense forest of trees, their green foliage a stark contrast to the charred landscape in the foreground. This juxtaposition highlights the immediate impact of the fire on the local ecosystem. The sky is overcast, with a grayish hue that complements the somber mood of the scene. The drone's perspective provides a comprehensive view of the entire area, allowing us to fully appreciate the extent of the fire's damage and the scale of the forest fire's impact on this ecosystem. Overall, the image captures a powerful and poignant scene of nature in the aftermath of a destructive forest fire, showcasing the immediate and long-lasting effects of such events on the environment.

(a) Real image of a prescribed burn in Finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 91. Real origin fire image and synthetic resulting image comparison num.6

Image description of Figure 92

The image shows an aerial view of a forest after a recent fire. The landscape is divided into distinct areas, with dense green forests on both sides of a cleared central strip. This strip appears to be the focus of the fire, as it's covered in brown, charred vegetation. A winding road cuts through the forest, starting from the bottom right corner and curving towards the top left of the image. This road likely served as a firebreak, which is a crucial element in containing wildfires. The right side of the image is particularly striking, with thick white smoke billowing upwards from the fire-damaged area. This smoke is likely filled with ash and other pollutants, creating a stark contrast against the remaining greenery. In the bottom right corner, there are visible spots of orange, which are likely areas of active fire or recently burned vegetation. This creates a vivid contrast with the surrounding green areas. The layout of the scene is quite dramatic. The fire has swept through the center of the image, leaving behind a strip of destruction. The surrounding forests on both sides appear untouched, creating a natural border for the affected area. The winding road serves as a visual link between these two forested regions. The aerial perspective provides a comprehensive view of the fire's impact and the forest's structure. It's a powerful image that illustrates the immediate and long-lasting effects of wildfires on forest ecosystems.

Image description of Figure 93

The image shows an aerial view of a town that has been severely impacted by flooding. The scene is dominated by muddy, brown terrain with scattered patches of green grass and trees. In the upper right corner, there's a body of water, likely a river or lake, that has overflowed its banks

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a prescribed burn in Finland [credit Kainuu wellbeing services county]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 92. Real origin fire image and synthetic resulting image comparison num.7

and inundated the surrounding area. The layout of the town is clearly visible from this high vantage point. In the upper left corner, there's a prominent white house with a gray roof, standing out against the floodwaters. To the right of this house, another white house with a gray roof is partially submerged, with its lower level covered in water. The streets are in a state of disarray. One road runs diagonally from the upper left to the lower right of the image, with another street branching off from it. These roads are likely covered in mud and debris, making them impassable. In the lower left corner of the image, there's a large area that appears to be a parking lot. It's filled with numerous cars, all of which are likely abandoned due to the flooding. The overall scene is one of destruction and chaos. The floodwaters have transformed what was once a peaceful town into a landscape of mud, debris, and abandoned vehicles. The aerial perspective provides a stark, comprehensive view of the extent of the damage, highlighting the vulnerability of urban areas to flooding and the rapid pace at which a town can be destroyed by water.

Image description of Figure 94

The image shows an aerial view of a landscape that has been severely impacted by flooding. The drone is positioned in the upper right corner of the frame, providing a wide-angle perspective of the scene below. The most striking feature is a large river that has swollen beyond its normal banks, inundating the surrounding area. The water appears to be a murky brown color, likely due to sediment and debris suspended in it. The floodwaters have inundated what seems to be a parking area or road, turning it into a vast expanse of water. In the foreground, there's a notable area of bare earth, which could be a cleared space or possibly the foundation of a building that has been damaged by the flood. This area stands out against the surrounding water and vegetation. The landscape is predominantly green, with numerous trees visible in the background and scattered throughout the flooded area. These trees appear to be in various states, possibly affected by the flood or simply part of the natural landscape. On the left side of the image, there's a white building visible. While it's not clear if this structure is still functional or severely impacted by the flood, its presence adds to the sense of scale and human impact on this natural environment. The overall layout of the scene suggests a rural or semi-rural setting, with the river likely being a major water source for the area before its overflow. The aerial view provides a

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a flood from the ahrtal region [Credit: DLR-ZKI 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 93. Real origin flood image and synthetic resulting image comparison num.1

unique perspective on how natural disasters can dramatically alter landscapes and force people to evacuate or face significant damage to their property. This image captures a moment of transition, where the natural landscape is being reshaped by floodwaters, and human structures and communities are being forced to adapt to the new conditions.

(a) Real image of a flood from the Ahrtal region [provided by DLR]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 94. Real origin flood image and synthetic resulting image comparison num.2

Image description of Figure 95

The image shows a drone's-eye view of a flooded landscape, likely captured shortly after a significant flood event. The scene is dominated by a large, murky brown river that has inundated the area, turning what was once a picturesque town into a waterlogged environment. The river is

This project has received funding from the European Union – European Commission under grant agreement No 101093003

filled with debris, including fallen trees and branches, which are scattered throughout its width. This suggests a powerful flood event that has swept through the area, uprooting trees and depositing them in the water. On the right side of the image, we can see the aftermath of the flooding. The land appears to be a mix of dirt and patches of grass, with some areas still showing signs of recent construction. This could indicate that the town is in the process of recovery and reconstruction following the flood. The left side of the image is where the most dramatic changes are visible. There's a large white building that stands out against the surrounding greenery. This structure appears to be intact and unscathed, contrasting sharply with the flooded surroundings. The landscape is framed by trees on both sides of the river, creating a natural border. In the background, rolling hills covered in dense forest can be seen, providing a stark contrast to the flooded town below. The sky above is overcast, with a grayish hue that adds to the somber mood of the scene. This cloudy sky likely contributed to the flood conditions by potentially increasing rainfall. Overall, the image presents a complex scene of destruction and resilience. The flood has transformed the town into a watery landscape, but signs of human presence and activity, such as the nearby building and ongoing construction, suggest that the area is not completely abandoned. The drone's perspective offers a unique view of how nature and human development can be dramatically affected by such events, while also showcasing the town's resilience in the face of adversity.

(a) Real image of a flood from the Ahrtal region [provided by DLR]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 95. Real origin flood image and synthetic resulting image comparison num.3

Image description of Figure 96

The image shows a village in the aftermath of a significant flood event. From a drone's perspective, we can observe: 1. Flooding: The most prominent feature is the extensive flooding of the village. Many buildings are partially or completely submerged, with only rooftops and upper floors visible above the water level. 2. Construction activity: There's active construction work in progress, particularly noticeable in the bottom left of the image. A large area has been cleared of vegetation and dirt, likely to prevent further water damage and create a dry space for future construction. 3. Buildings affected: Most of the visible buildings are white, with some brown structures also visible. The floodwaters have inundated the lower portions of these buildings, leaving only the upper floors exposed above the water. 4. Landscape changes: The flood has

This project has received funding from the European Union – European Commission under grant agreement No 101093003

dramatically altered the landscape. The water level has risen significantly, submerging most of the lower areas of the village. The surrounding terrain, including hills visible in the background, has been affected by the water. 5. Community efforts: Despite the devastation, there are signs of ongoing community action. A white van is parked in front of one of the partially submerged buildings, and a tractor can be seen in the flooded area, likely being used for rescue or recovery operations. 6. Infrastructure: The flood has likely damaged or destroyed much of the village's infrastructure, including roads, power lines, and other utilities. These may be under repair or assessment. 7. Resilience: The presence of construction equipment and the ongoing community efforts suggest that the village is already beginning to recover and rebuild, showcasing the resilience of its residents. This aerial view provides a stark picture of the flood's impact on the village, highlighting both the destruction caused and the ongoing efforts to recover and rebuild.

(a) Real image of a flood from the Ahrtal region [provided by DLR]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 96. Real origin flood image and synthetic resulting image comparison num.4

Image description of Figure 97

The image shows an aerial view of a flooded landscape following a flood. The scene is dominated by a large body of murky, brown water that appears to be a river or creek. The water level is significantly elevated, likely due to heavy rainfall or a river overflowing its banks. In the foreground, there's a notable metal fence or barrier running along the edge of the water. This structure seems to be serving as a flood control measure, attempting to prevent the water from spreading further into the surrounding area. The floodwater has inundated what appears to be a rural or semi-rural setting. In the background, I can see a dense forest with lush green trees, providing a stark contrast to the flooded foreground. To the right side of the image, there's a patch of land that looks like it could be farmland. It appears to be an agricultural area, with rows of crops visible. This suggests that the flood may have affected both natural habitats and cultivated land. The overall layout of the scene is characterized by the juxtaposition of the floodwater, the flood control structure, and the surrounding landscape. The aerial perspective provides a comprehensive view of how the flood has impacted this area, showing the extent of water coverage and the potential long-term consequences for the local environment and ecosystem. This image offers a powerful visual representation of the impact of flooding on rural landscapes and the importance of flood control measures in protecting both natural habitats and human settlements.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a flood from the Ahrtal region [Credit: Arbeiter-Samariter-Bund Deutschland e. V. 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 97. Real origin flood image and synthetic resulting image comparison num.5

Image description of Figure 98

The image shows a drone's-eye view of a severely damaged bridge in the aftermath of a flood. The bridge, which appears to be a railroad bridge based on its structure, is completely destroyed and lying in the water. It's covered in debris, primarily consisting of wood, branches, and other materials that have washed up from the floodwaters. The bridge spans across a body of water that looks murky and brown, likely due to the sediment and debris from the surrounding landscape. The water level appears to be quite high, submerging much of the bridge's structure. On one side of the bridge, there's a small patch of land visible. This area has some grass and a few trees, but it's clear that the flood has affected this section as well, though to a lesser extent than the bridge. The scene is one of widespread destruction. The bridge, which would typically be a vital infrastructure piece, is now rendered useless and unsalvageable. The debris scattered across the water and on the surrounding land emphasizes the force of the flood and the amount of material that was moved by the water. This image provides a stark visual representation of the devastating impact of natural disasters on infrastructure and the environment. It serves as a powerful reminder of the need for preparedness and resilience in the face of extreme events.

Image description of Figure 99

The image shows a devastating scene of a town ravaged by flooding. From a drone's perspective, we can see a landscape transformed by water and destruction. In the foreground, there's a road running diagonally across the image. On this road, we can observe several vehicles: 1. A yellow truck with a large container 2. A black truck 3. A white truck with a red back These vehicles appear to be navigating the flooded road, possibly delivering aid or assessing the damage. The most striking feature of the image is the area of destruction in the middle. This section shows: 1. A house that has been completely destroyed 2. Debris scattered around, including wood and other building materials 3. A muddy, wet ground, indicating recent flooding The destruction extends to multiple buildings in this central area, highlighting the widespread nature of the flood's impact. In the background, we can see: 1. More houses, some still standing but likely damaged 2. Trees, some of which may have been uprooted by the floodwaters 3. A grassy area, possibly a park or open space The overall layout of the scene suggests this is a residential area, with houses

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a flood from the Ahrtal region [Credit: Arbeiter-Samariter-Bund Deutschland e. V. 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 98. Real origin flood image and synthetic resulting image comparison num.6

and green spaces intermingled. However, the flood has transformed this area into a scene of devastation. The drone's perspective provides a comprehensive view of the damage, allowing for an assessment of the scale and nature of the flood's impact on this community. The presence of the vehicles on the road indicates that life is continuing, albeit in a dramatically altered state. This image captures a moment of crisis, showcasing the vulnerability of urban areas to flooding and the immediate human response required in the aftermath of such natural disasters.

(a) Real image of a flood from the Ahrtal region [Credit: Arbeiter-Samariter-Bund Deutschland e. V. 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 99. Real origin flood image and synthetic resulting image comparison num.7

Image description of Figure 100

The image shows a landscape that has been significantly impacted by flooding. In the foreground,

This project has received funding from the European Union – European Commission under grant agreement No 101093003

there's a river that has swollen with water, causing significant erosion along its banks. The river's bed is now filled with debris, including fallen trees and other natural materials, which have been washed down from higher elevations. A notable feature in the scene is a stone bridge that spans across the river. This bridge appears to be partially damaged, with some sections collapsed or damaged, particularly on the right side. The structure of the bridge is still visible, but it's clear that it's struggling to maintain its integrity in the altered landscape. On the left side of the image, there's a large concrete pillar supporting the bridge. This pillar is likely part of the bridge's original support structure, and its presence indicates that the bridge was once a more robust and stable construction. The surrounding area is densely forested, with a thick canopy of trees visible beyond the river. This lush greenery contrasts sharply with the damaged infrastructure in the foreground. The overall scene depicts a landscape that has been dramatically altered by flood waters. The river's increased water volume and debris content, along with the damaged bridge, suggest a significant change in the area's water cycle and possibly climate patterns. The scene captures the immediate aftermath of a major flood event, showcasing both the natural beauty of the area and the destructive power of water.

(a) Real image of a flood from the ahrtal region [Credit: Arbeiter-Samariter-Bund Deutschland e. V. 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 100. Real origin flood image and synthetic resulting image comparison num.8

Image description of Figure 101

The image shows a drone's-eye view of a town that has been severely impacted by flooding. The scene is dominated by a winding river that cuts through the centre of the town, its waters swollen and muddy, indicating recent heavy rainfall or flooding. On the left side of the river, there's clear evidence of flood damage. A large area of land appears to have been submerged, with debris scattered across the water's surface. A bridge spans the river, and interestingly, there's a car partially submerged in the water, a stark reminder of the flood's intensity. The right side of the river presents a stark contrast. The land is mostly bare, with dirt and debris visible. A road runs through this area, and several vehicles can be seen on it - a white van and a black truck are particularly noticeable. The town itself is composed of numerous white houses with black roofs, clustered together in what appears to be a typical suburban layout. However, the flood has disrupted this orderly pattern, with houses and streets likely submerged or damaged. In the background, a large hill covered in dense forest rises up, providing a natural backdrop to the scene. The trees

This project has received funding from the European Union – European Commission under grant agreement No 101093003

on the hillside appear unaffected by the flood, standing tall amidst the chaos below. The sky is not visible in this image, focusing the viewer's attention on the ground-level impact of the flood. The color palette is predominantly earthy, with shades of brown from the dirt and debris, green from the trees, and the muddy water of the river. This aerial view offers a comprehensive look at the town's layout and the devastating effects of the flood. It's a powerful image that captures the vulnerability of urban areas to natural disasters and the significant changes that can occur in a matter of hours.

(a) Real image of a flood from the Ahrtal region [Credit: Arbeiter-Samariter-Bund Deutschland e. V. 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 101. Real origin flood image and synthetic resulting image comparison num.9

Image description of Figure 102 The image shows a drone's-eye view of a flooded rural landscape after heavy rainfall. The scene is dominated by a muddy river that has swollen to an unusually wide width, covering much of the visible terrain. The water has a thick, brownish-yellow hue, indicating significant pollution or sediment content. The river's banks are lined with dense green trees on both sides, creating a stark contrast between the lush vegetation and the muddy waters. The trees appear to be thriving, suggesting this flooding may be a recurring event in the area. To the right of the river, there's a grassy field that's partially submerged. This field is dotted with small puddles and patches of standing water, indicating the extent of water saturation in the soil. The sky above is overcast with gray clouds, which adds to the somber and the dramatic atmosphere of the scene. The lack of direct sunlight suggests that the image was taken during a cloudy day, possibly in the afternoon. The overall layout of the scene is characterized by the wide, muddy river running through the center of the image, flanked by trees on both sides and a grassy field to the right. The color palette is predominantly earthy, with various shades of brown, green, gray, and white dominating the view. This image captures the immediate aftermath of a flood event, showcasing the dramatic impact of heavy rainfall on rural landscapes. It provides a unique perspective on how water can transform natural environments, temporarily turning everyday fields into vast, muddy waters.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(a) Real image of a flood from the Ahrtal region [Credit: Bavarian Red Cross 2021]

(b) Generated image using molmo to get the description and Flux.1-dev [image generated by ATOS]

Figure 102. Real origin flood image and synthetic resulting image comparison num.10

This project has received funding from the European Union – European Commission under grant agreement No 101093003

E. Stable Diffusion XL fire Inpaint + Flux.1 dev Refinement pipeline

Figure 103. ComfyUI generation pipeline Stable Diffusion XL Inpaint + Flux.1 dev Refinement to augment images with fire [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

F. Image Upscale RealESRGAN pipeline

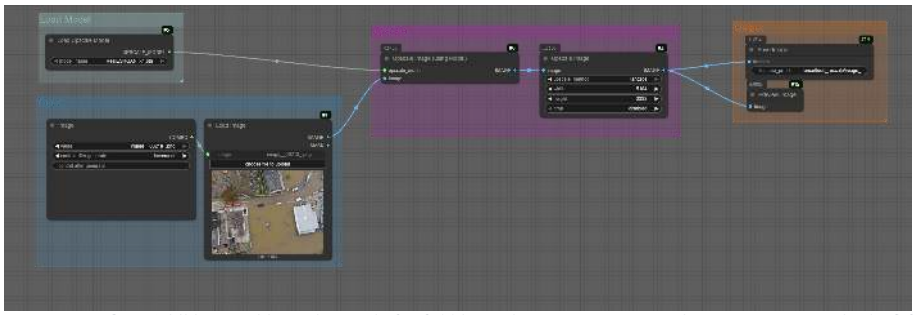


Figure 104. ComfyUI Image Upscale RealESRGAN pipeline on augmented fire images [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

G. Flux.1-kontext-dev flood augmentation pipeline

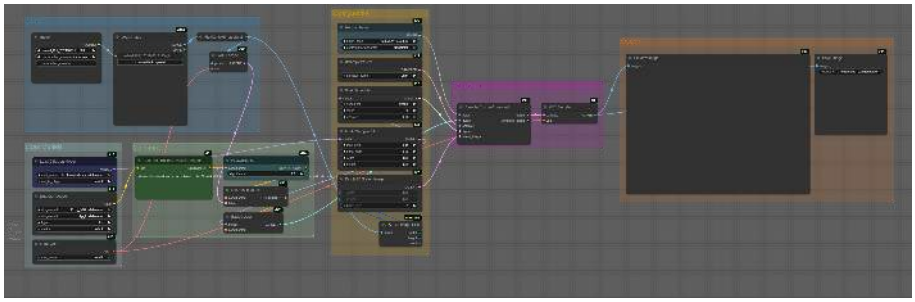


Figure 105. ComfyUI Flux.1-kontext-dev flood augmentation pipeline [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

H. Flux.1-fill-dev people inpaint pipeline

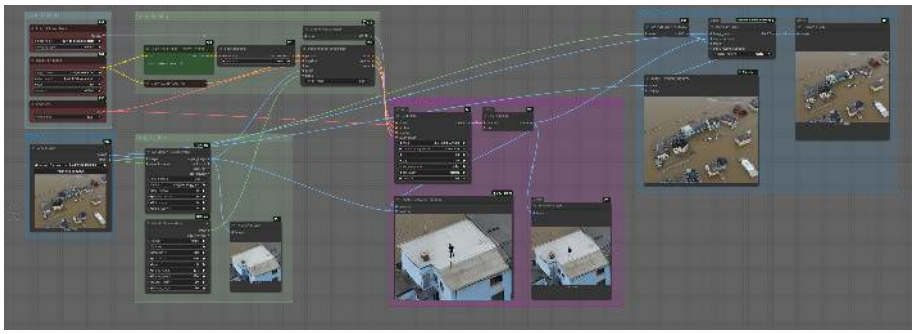


Figure 106. ComfyUI Flux.1-fill-dev people inpaint pipeline [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

I. SDXL DAAM explanation pipeline

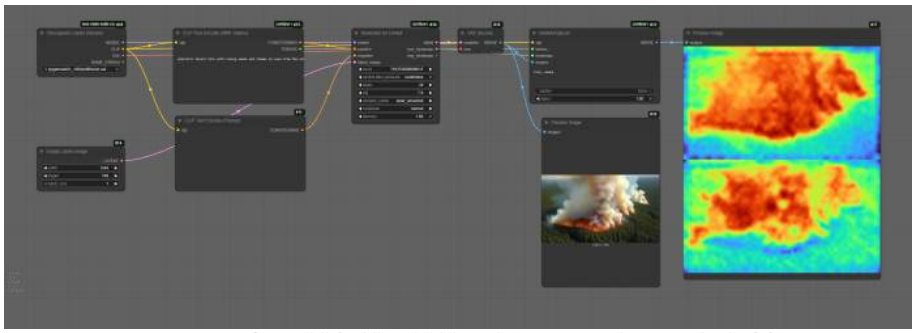


Figure 107. ComfyUI SDXL DAAM explanation pipeline [credit ATOS]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

J. Heatmap analisis Algorithm

Algorithm 1 Heatmap analisis Algorithm

Require: Heatmap grayscale image I, Background image B
Ensure: Binary mask M, Blended overlay O, Bounding box image BB

Load heatmap grayscale image I
Apply Gaussian blur to $I \rightarrow I_{blur}$
Reshape I_{blur} into 1D vector
Apply K-means clustering ($K = 2$) on vector
Reshape cluster labels back to image dimensions $\rightarrow L$
Select brighter cluster as foreground
Generate binary mask M from L (foreground=255, background=0)
Apply morphological closing on M
Apply morphological opening on M
Remove connected components in M with area $<$ threshold
Extract contours from M
Compute bounding box enclosing all contours $\rightarrow BB$
Draw bounding box on background image
Overlay M onto background image B
Blend overlay with alpha transparency $\rightarrow O$
Save M, O, and BB as output files

This project has received funding from the European Union – European Commission under grant agreement No 101093003