Flood Image Analysis

E. Vlachos, D. Papaioannou, Prof. I. Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.0

Artificial Intelligence &
Information Analysis Lab

Flood Image Analysis

- **Introduction**
- Deep Semantic Segmentation
- Flood Region Segmentation
- Object Detection
- Person/Vehicle Detection in Flooded Regions
- House-Roof Detection in Flooded Regions
- Flood Monitoring System

Natural Disaster Management

Recent catastrophic flood in Thessaly (2023).

- Due to climate change, **flash floods** are more usual than ever, affecting the lives of millions of people.

- There is an imminent need for cutting-edge Natural Disaster Management systems (NDM)

Natural Disaster Management

- ***Unmanned Aerial Vehicles (UAVs)*** can fly over areas which humans cannot access and capture useful footage.
- State-of-the-art computer vision models can perform tasks like ***flood segmentation*** or ***person detection*** and produce valuable insights from multimedia.
- Evolution on ***edge computing*** enables us to deploy computer vision models on UAVs.

UAV monitors flood damage in Peru.

Artificial Intelligence & Information Analysis Lab

Flood Region Segmentation on Drone Images

Prediction

Flood Region Segmentation of a flood event on the use case of Arthal, Western Germany.

Image Segmentation

- An image domain $\mathcal{X}$ must be segment in N different regions $\mathcal{R}_1 \dots, \mathcal{R}_N$.
- The segmentation rule is a logical predicate of the form $P(\mathcal{R})$.
- Image segmentation partitions the set $\mathcal{X}$ into the subsets $\mathcal{R}_i, i = 1, \dots, N$, having the following properties:

$$\mathcal{X} = \bigcup_{i=1}^{N} \mathcal{R}_i\,,$$

$$\mathcal{R}_i \cap \mathcal{R}_j = \emptyset, \quad i \neq j,$$

$$P(\mathcal{R}_i) = TRUE, \qquad i = 1, \dots, N,$$

$$P(\mathcal{R}_i \cup \mathcal{R}_j) = FALSE, \qquad i \neq j$$

Flood Region Segmentation on Drone Images

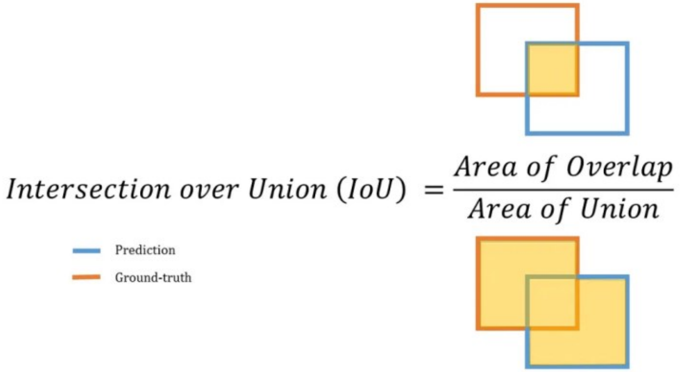
- Flood region segmentation is a task where each pixel in an image domain $\mathcal{X}$ is classified as either belonging to a flooded region (e.g., foreground) or to a non-flooded region (e.g., background).
- Given a function f, a segmentation mask is produced as $S = f(\mathcal{X})$.
- $f(\mathcal{X})$ generates a probability map $P(x, y)$ where $P(x, y) = P(S(x, y) = 1|\mathcal{X})$ is the probability that the pixel (x, y) belongs to the flooded region.
- Given a threshold T the segmentation mask is given as:

$$S(x,y) = \begin{cases} 1, & if\ P(x,y) \geq T \\ 0, & otherwise \end{cases}.$$

Evaluating Performance with IoU

- ***Intersection over Union (IoU)***:

For each class i: $$IoU_i = \frac{\sum_{k=1}^{N} \left| A_k^i \cap B_k^i \right|}{\sum_{k=1}^{N} \left| A_k^i \cup B_k^i \right|}$$



- A_k^i , B_k^i are the predicted mask for class i in image k and the ground truth mask for class i in image k respectively.
- Also called ***Jaccard Similarity Coefficient*** or ***Overlap Score.***

- ***Mean Intersection over Union (IoU)***:

$$\mathrm{mIoU}(\mathcal{A}, \mathcal{B}) = \frac{1}{C} \sum_{i=0}^{C-1} IoU_i$$

- Average all IoU's across all C classes.

Flood Image Analysis

- Introduction
- **Deep Semantic Segmentation**
- Flood Region Segmentation
- Object Detection
- Person/Vehicle Detection in Flooded Regions
- House-Roof Detection in Flooded Regions
- Flood Monitoring System

Deep Semantic Segmentation

- Transforming the fully connected layers of image classification networks into convolution layers enables the transformed network to output **heatmaps**.

- End-to-end dense prediction learning is possible by adding extra layers and using an appropriate loss function.

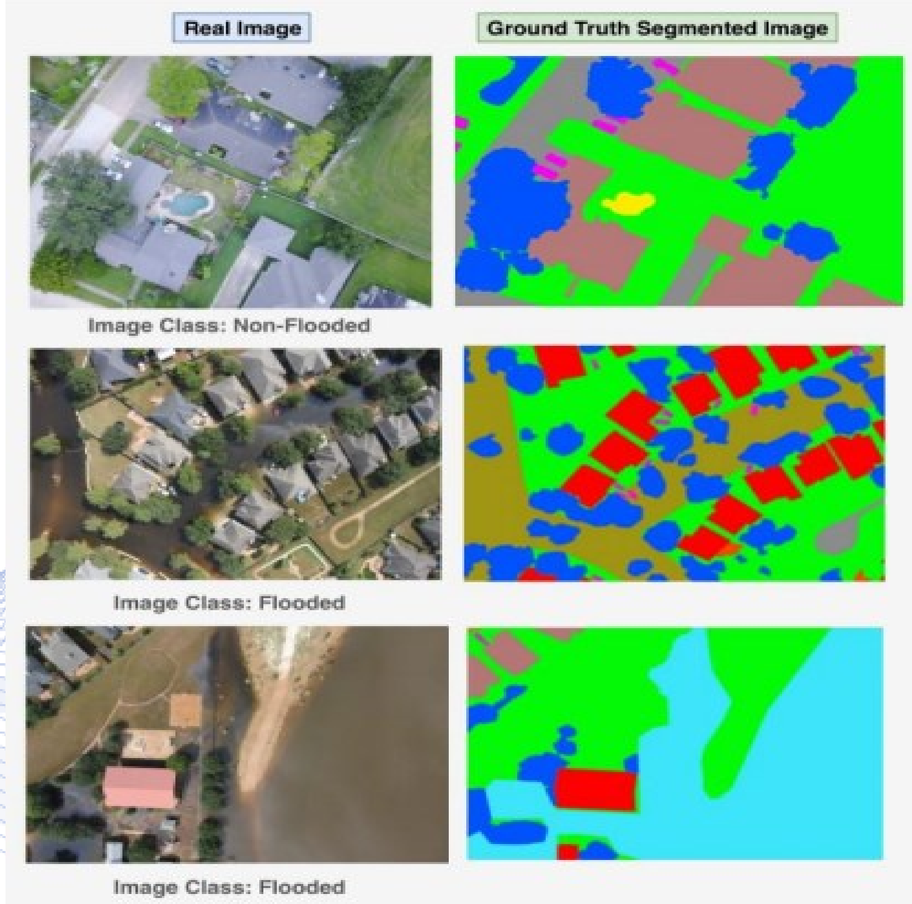
- Encoder-Decoder network architecture.

Flood Image Analysis

- Introduction
- Deep Semantic Segmentation
- **Flood Region Segmentation**
- Object Detection
- Person/Vehicle Detection in Flooded Regions
- House-Roof Detection in Flooded Regions
- Flood Monitoring System

Flood Region Segmentation

- FloodNet Dataset, consisting of UAV imagery for post-flood semantic segmentation.
- This dataset however is more suitable for post destruction assessment applications.
- V-floodnet includes thousands of images depicting regular waterbodies in their dataset, thus making the flood a "minority".



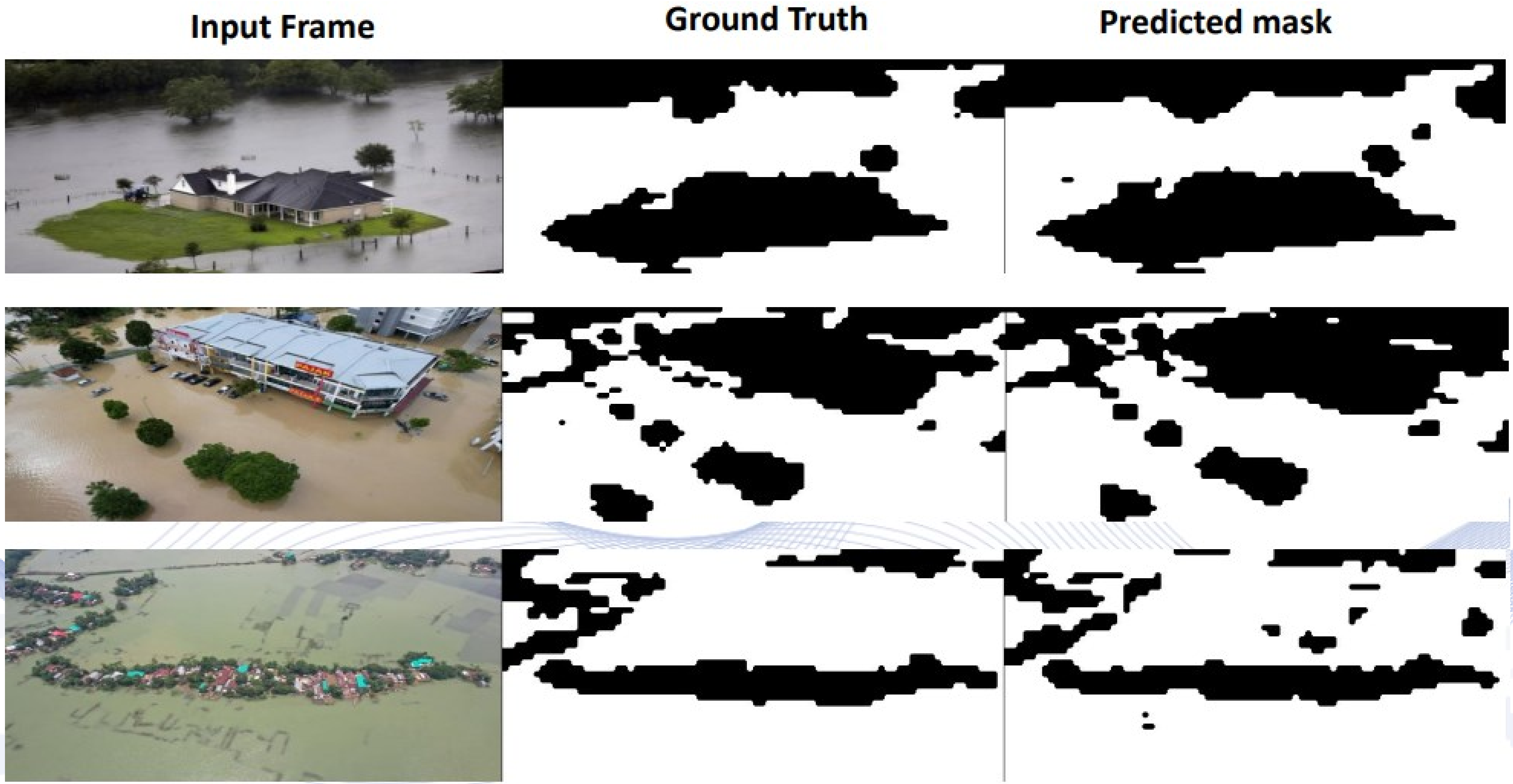
Sample images and ground truth masks from FloodNet Dataset.

Flood Region Segmentation



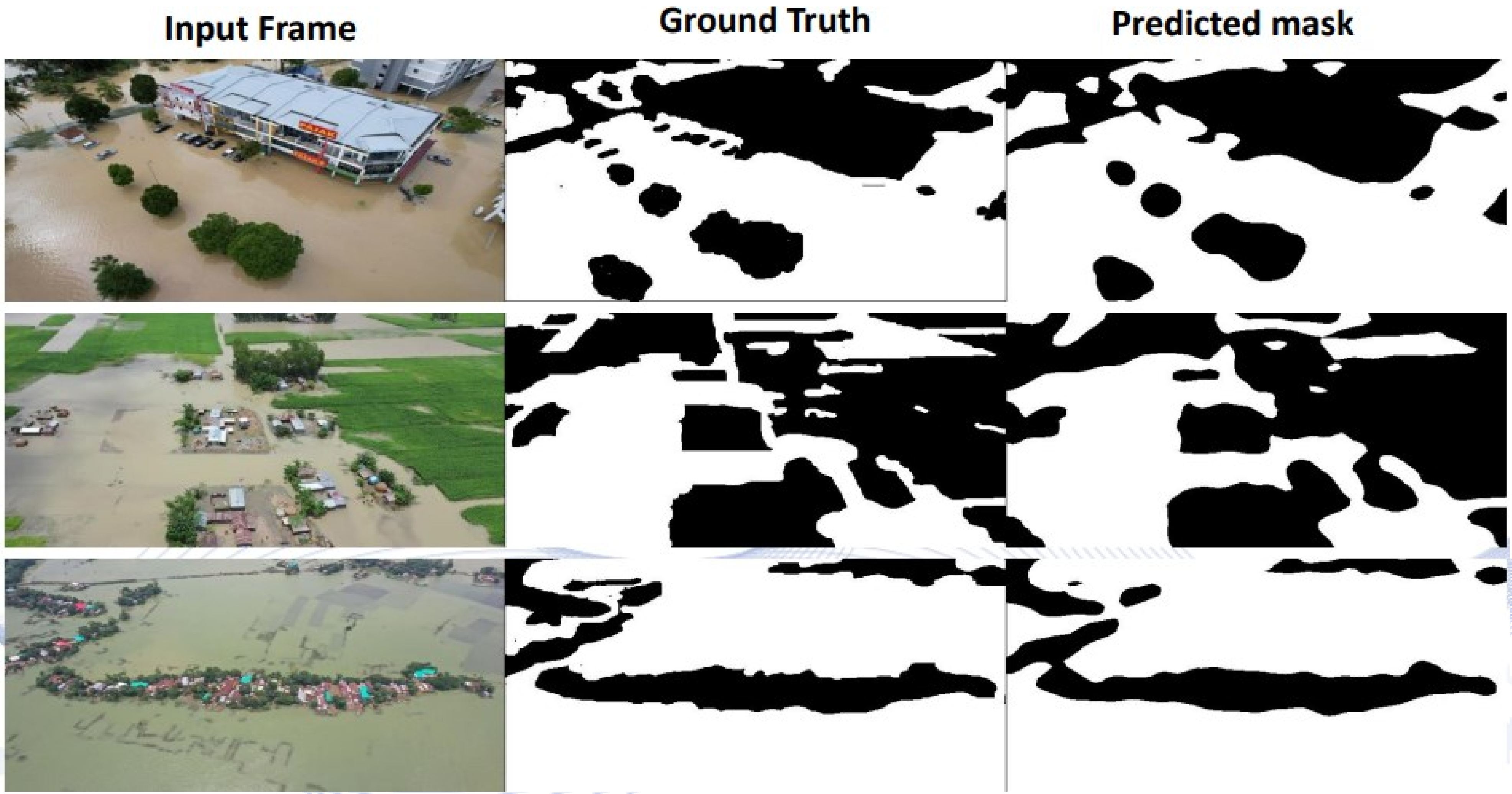
Sample images from the FloodSeg Training Dataset, each one from a different source.

Flood Segmentation



Segmented outputs with **CNN-i2i**.

Flood Segmentation



Segmented outputs with **PSPnet**.

Flood Segmentation

Segmented Results of Arthal West Germany flash flood case using CNN-i2i Architecture.

Artificial Intelligence & Information Analysis Lab

Visualized Segmentation Results

Visualized flood segmentation masks for two sample input frames, from the real-world testing videos that we annotated.

Flood Image Analysis

- Introduction
- Deep Semantic Segmentation
- Flood Region Segmentation
- **Object Detection**
- Person/Vehicle Detection in Flooded Regions
- House-Roof Detection in Flooded Regions
- Flood Monitoring System

Object Detection

- Object detection = Classification + Localization:
- Find ***what*** is in a picture as well as ***where*** it is.

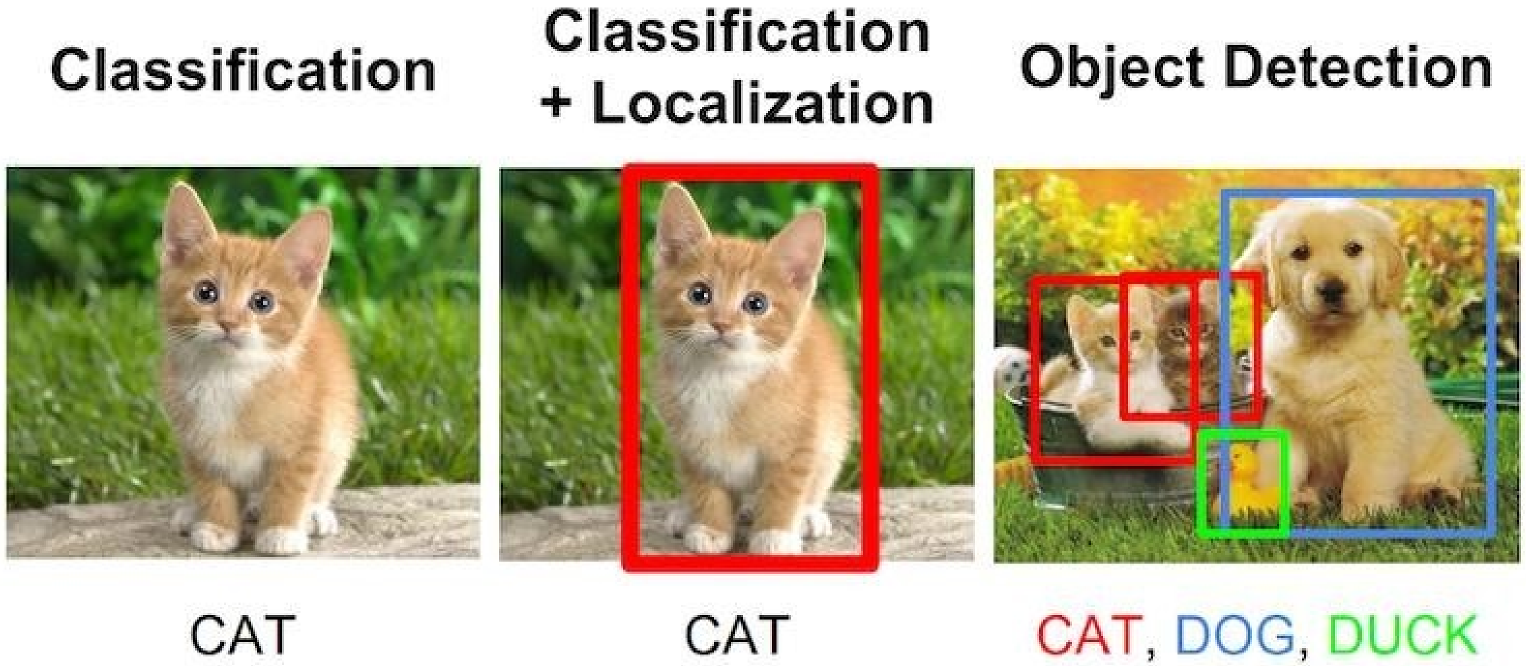


Figure: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Object Detection

- ***Input***: an image.
- ***Output***: ***bounding boxes*** containing depicted objects.
  - Each image may contain a different number of detected objects.
- Old approach: train a specialized classifier and deploy in ***sliding-window style*** to detect all object of that class.
  - Very inefficient, quite ineffective.
- ***Goal***: combine classification and localization into a ***single architecture for multiple, multiclass object detection***.

Object Detection

Object detection is a ***multitask machine learning*** problem:

- combination of classification and regression.
- Given a set of classes $\mathcal{C} = \{\mathcal{C}_i, i = 1, \dots, m\}$ and an image sample $\mathbf{x} \in \mathbb{R}^n$, the model predicts (for one object instance only) an output vector $\hat{\boldsymbol{y}} = [\hat{\boldsymbol{y}}_1^T | \hat{\boldsymbol{y}}_2^T]^T$ consisting of:
  - A class vector $\hat{\boldsymbol{y}}_1 \in [0, 1]^m$ and
  - A bounding box parameter vector $\hat{\boldsymbol{y}}_2 = [x, y, w, h]^T$ corresponding to object ROI.
- Optimization of a joint cost function:

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \alpha_1 J_1(\boldsymbol{y}_1, \hat{\boldsymbol{y}}_1) + \alpha_2 J_2(\boldsymbol{y}_2, \hat{\boldsymbol{y}}_2)$$

- The above vector pair will be computed for every possible target detected in the image sample $\mathbf{x}$.

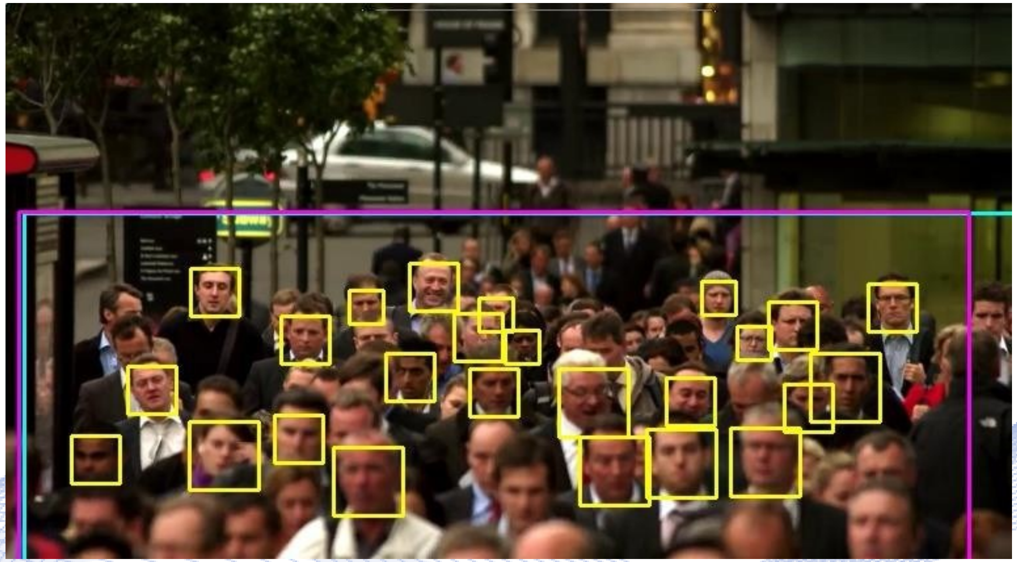
Non-Maximum Suppression (NMS)

- **Challenge:** Object detectors often output many overlapping detections.

- **Solution:** Post process raw detections using NMS
  - For each category:
    - Select next highest-scoring box
    - Eliminate lower-scoring boxes with IoU>threshold(e.g. 0.5)
    - If remaining boxes, go to first step.

Object Detection Inference Examples

Bicycle Detection.



Face Detection.

CNN-Object Detection

Region proposal-based detectors

- R-CNN, Fast R-CNN, Faster R-CNN
- R-FCN

Single Stage Detectors

- YOLO
- SSD
- YOLO v2, v3, v4
- RetinaNet, RBFnet
- CornerNet, CenterNet

Transformer Detectors

- DETR.

CNN-Object Detection Architectures

- **Backbone** refers to a CNN used for image feature extraction:
  - ResNet, MobileNet, VGG etc.
- **Neck** is an extra object detector layer that goes on top of the backbone. It extracts different feature maps from different stages of the backbone.
  - FPN, PANet, Bi-FPN etc.
- ***Head*** network performs actual object detection: classification (probability of $m+1$ classes) and regression of RoI parameters (x, y, h, w).

YOLO Models

YOLO versions have evolved significantly, each uniquely enhancements to the single-stage detector architecture:

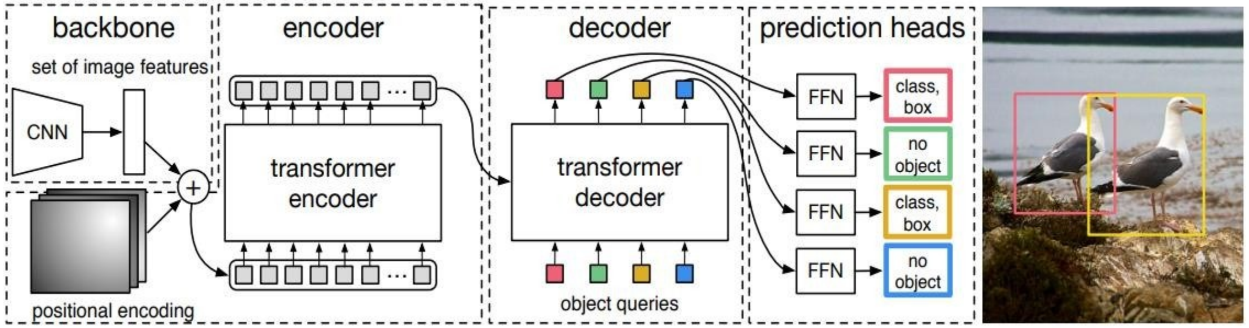
- ***YOLOX***
- ***YOLOv7***
- ***YOLOv8***
- ***YOLOv9***
- ***YOLOv10***
- ***YOLOv11.***

Each model improves detection accuracy, speed, and robustness for various applications.

DETR

DETR architecture has three main components:

- A RestNet50/101 CNN backbone for feature extraction.
- An encoder-decoder transformer model.
- A feed-forward head network makes the final detection predictions.



DETR architecture [CAR2020].

Object Detection Performance Metrics

For M ground truth object ROIs on all N_t images:

- Let $n_{ij} = 1$ for a successful classification at ***confidence threshold*** t $(s_{ij} \geq t)$:
- ***Recall, Precision*** definitions (modified):

$$r(t) = \frac{\sigma_{ij}\, n_{ij} z_{ij}}{M},$$

$$p(t) = \frac{\sigma_{ij}\, n_{ij} z_{ij}}{\sigma_{ij}\, n_{ij}}.$$

Object Detection Performance Metrics

Mean Average Precision (***mAP***):

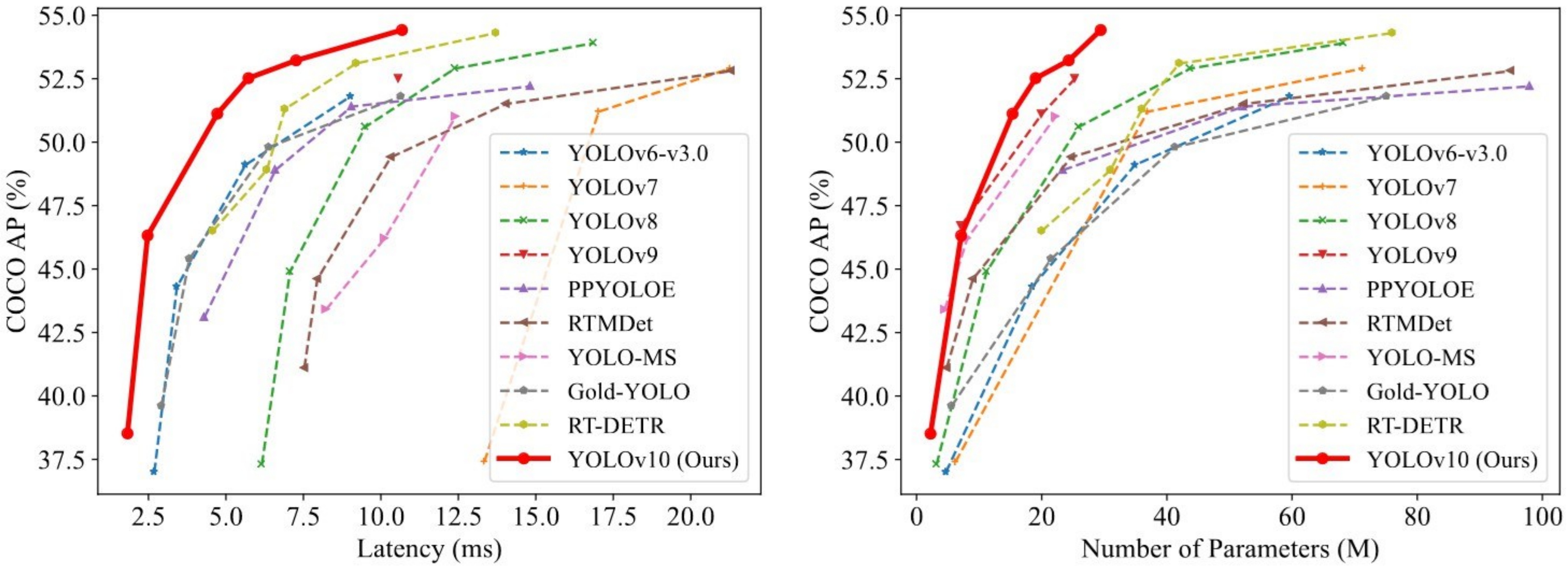
- It is calculated over N levels of confidence threshold $t_n, n = 1, \dots, N$:

$$mAP = \frac{1}{N} \Sigma_n \, p(t_n).$$

Object Localization Performance Metrics Example

Object localization performance: a) mAP($\mathcal{A}, \mathcal{B}$) = 0.67, b) mAP(A,B) = 0.27

Real-Time Object Detectors



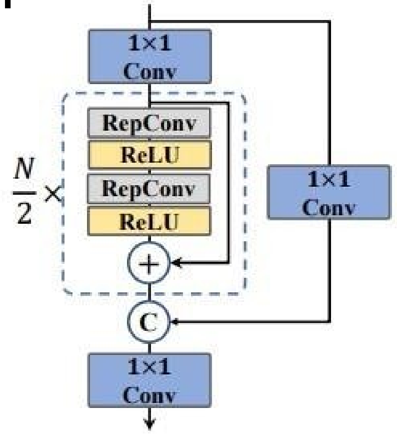
Real time object detectors comparison on COCO dataset [WAN2024].

Flood Image Analysis

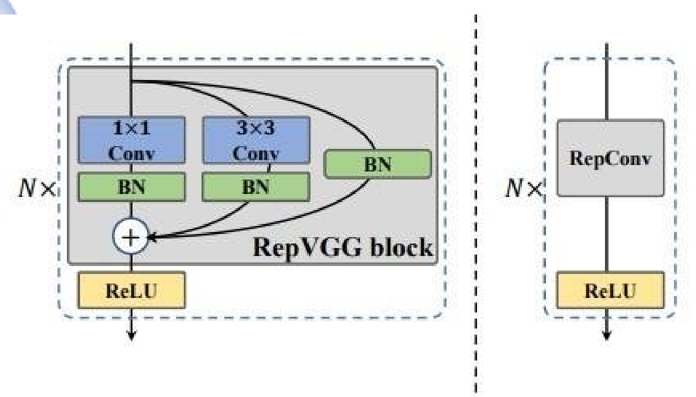
- Introduction
- Deep Semantic Segmentation
- Flood Region Segmentation
- Object Detection
- **Person/Vehicle Detection in Flooded Regions**
- House-Roof Detection in Flooded Regions
- Flood Monitoring System

Person/Vehicle Detection

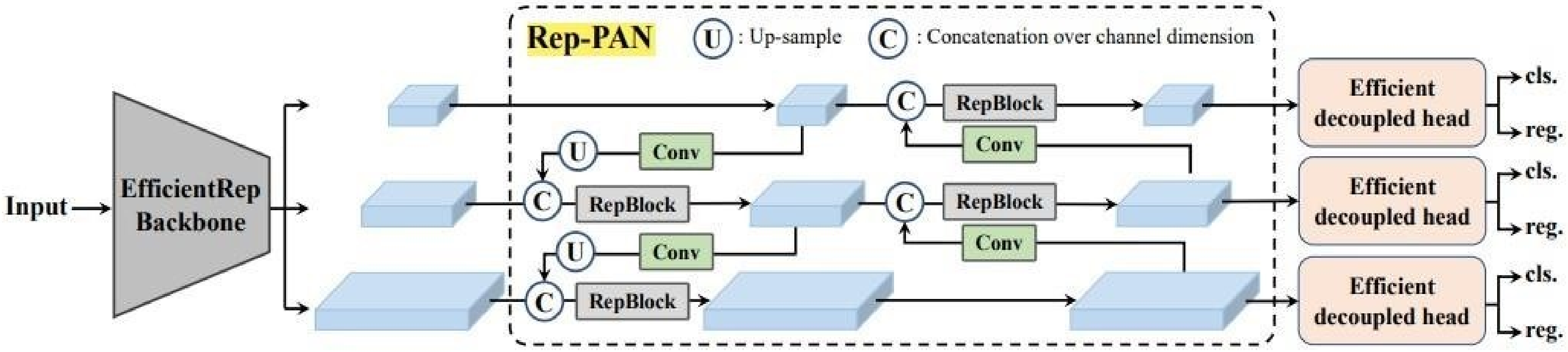
- Opted for the vanilla approach of using a model pretrained on
- COCO dataset.
- Yolov6 [LI2023] is a state-of-the-art model in real-time object
- detection.
- Small and Large models were selected and compared.



CSPStackRep building block.



RepVGG [DING2021] building block.



Yolov6 small model architecture.

Object Detection and Tracking in Floods

DNN models, pretrained on COCO dataset were used to detect classes of interest (***cars, persons***) that may be in danger).

YOLOv6 4.0 small version in person, car detection in Thessaly floods, Greece (September 2023).

Artificial Intelligence & Information Analysis Lab

Detection Results Visualized

Sample images from our detection test set, with visualized bounding boxes (red boxes for **persons**, green boxes for **cars**).

Flood Image Analysis

- Introduction
- Deep Semantic Segmentation
- Flood Region Segmentation
- Object Detection
- Person/Vehicle Detection in Flooded Regions
- **House-Roof Detection in Flooded Regions**
- Flood Monitoring System

Flooded House Detection on Drone Images

- **House-roof detection** is a task where the goal is to locate and classify bounding boxes around house roofs in an image domain X.
- Given a detection function f, the output is a set of bounding boxes $\mathrm{B}=\{b_1, b_2, \ldots, b_n\}$ where each $b_i = (\mathbf{x}_i, \mathbf{y}_i, \mathbf{W}_i, \mathbf{h}_i)$ defines the coordinates, width, and height of a bounding box that is predicted to contain a house roof.
- Each bounding box b_i is associated with a confidence score $\mathrm{P}(\boldsymbol{b}_i|\mathbf{X})$ representing the probability that b_i contains a house roof.
- Given a confidence threshold T, a bounding box b_i is considered to contain a house roof if: $\boldsymbol{b}_i = \begin{cases} 1, & if\ P(\boldsymbol{bi}|\boldsymbol{X}) \geq T \\ 0, & otherwise \end{cases}$

Artificial Intelligence & Information Analysis Lab

Datasets

Three main datasets have been used.

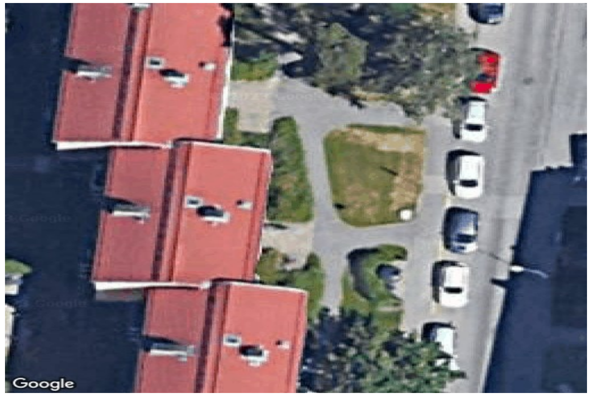
- FloodNet
- Giannitsa (images of the city of Giannitsa, Macedonia, Greece)
- RedRoofs (Google Maps Images)

The following datasets were created:

- FRG (Floodnet + RedRoofs + Giannitsa)
- MixedAreas (images from the FRG dataset and internet-sourced images depicting flooded houses)
- NoWater (
  - images without flooded regions for training
  - Images with flooded regions for testing)

Datasets

FloodNet | Giannitsa | RedRoofs



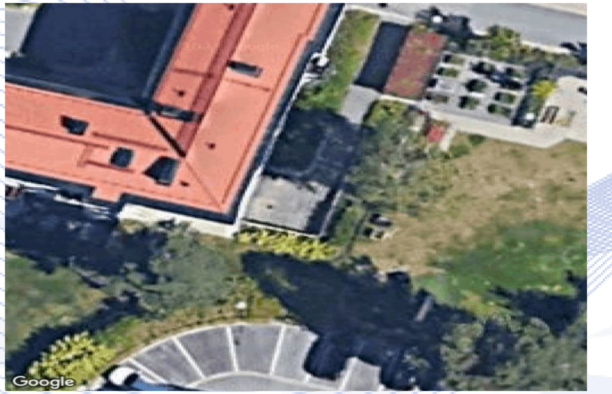


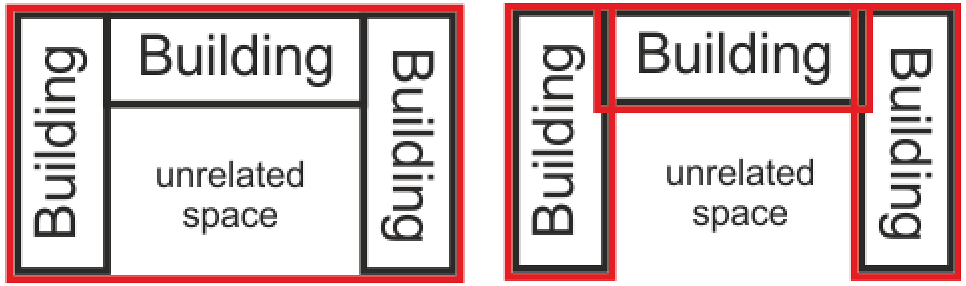
Image samples from the three main datasets.

Datasets

Annotation:

Two ways for image annotation:

a) One bounding box
b) Multiple bounding boxes



(a) (b)

Example of annotating a C-shaped building.

Advantages of (b):

- Improved Object Description
- Fewer unrelated pixels
- Reduced false detections

Disadvantages of (b):

- Higher number of bounding boxes
- Longer annotation time

Inference Example

DETR

YOLOv6m

Comparison of DETR and YOLOv6m Inference Results.

Artificial Intelligence & Information Analysis Lab

Flood Image Analysis

- Introduction
- Deep Semantic Segmentation
- Flood Region Segmentation
- Object Detection
- Person/Vehicle Detection in Flooded Regions
- House-Roof Detection in Flooded Regions
- **Flood Monitoring System**

Flood Monitoring System

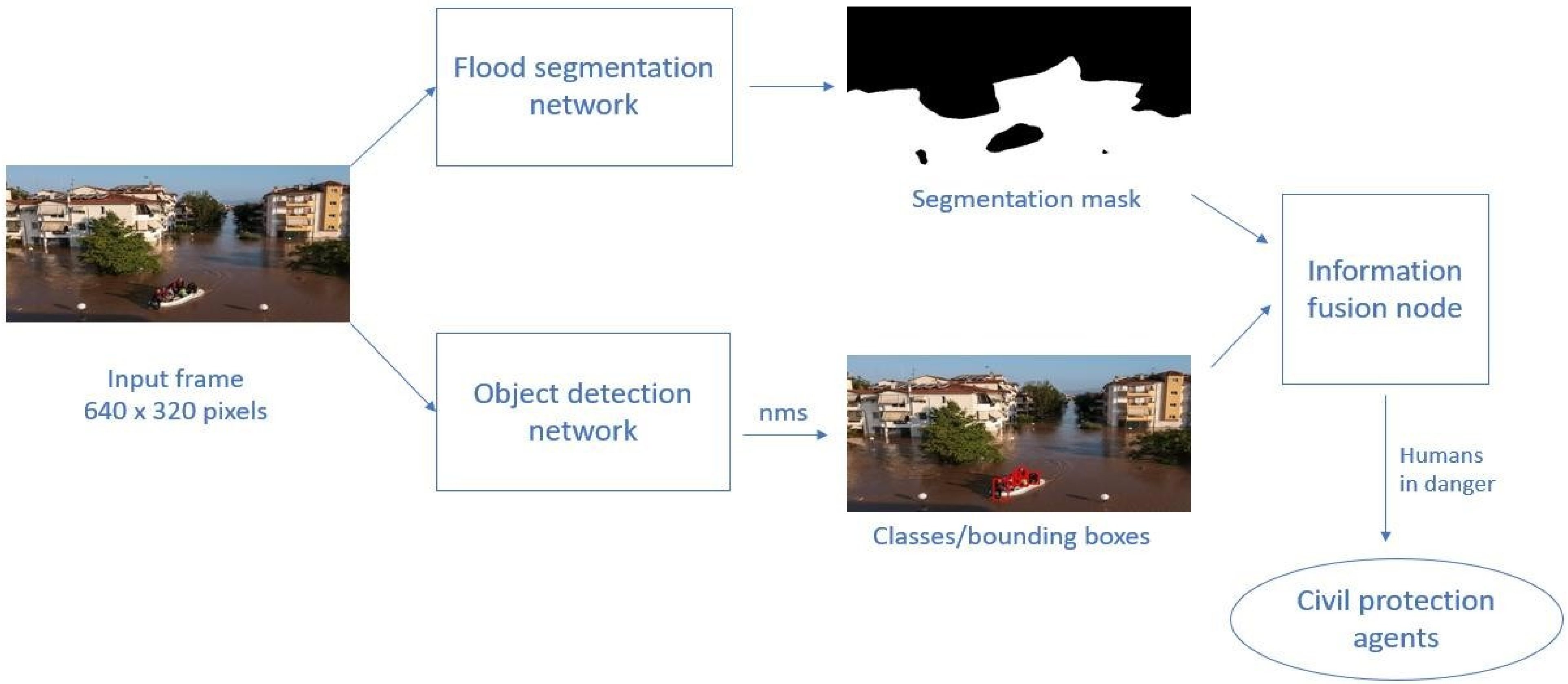
We developed a Deep Learning based Flood Monitoring System, capable of being deployed on the edge.

This system can:

- Segment the flood water, providing its end users with a precise flood mapping tool.
- Detect humans or vehicles in the flood context, even when they are partially submerged in the water.
- Fuse the two outputs and generate rescue alert if the detected object is in danger of the flood.

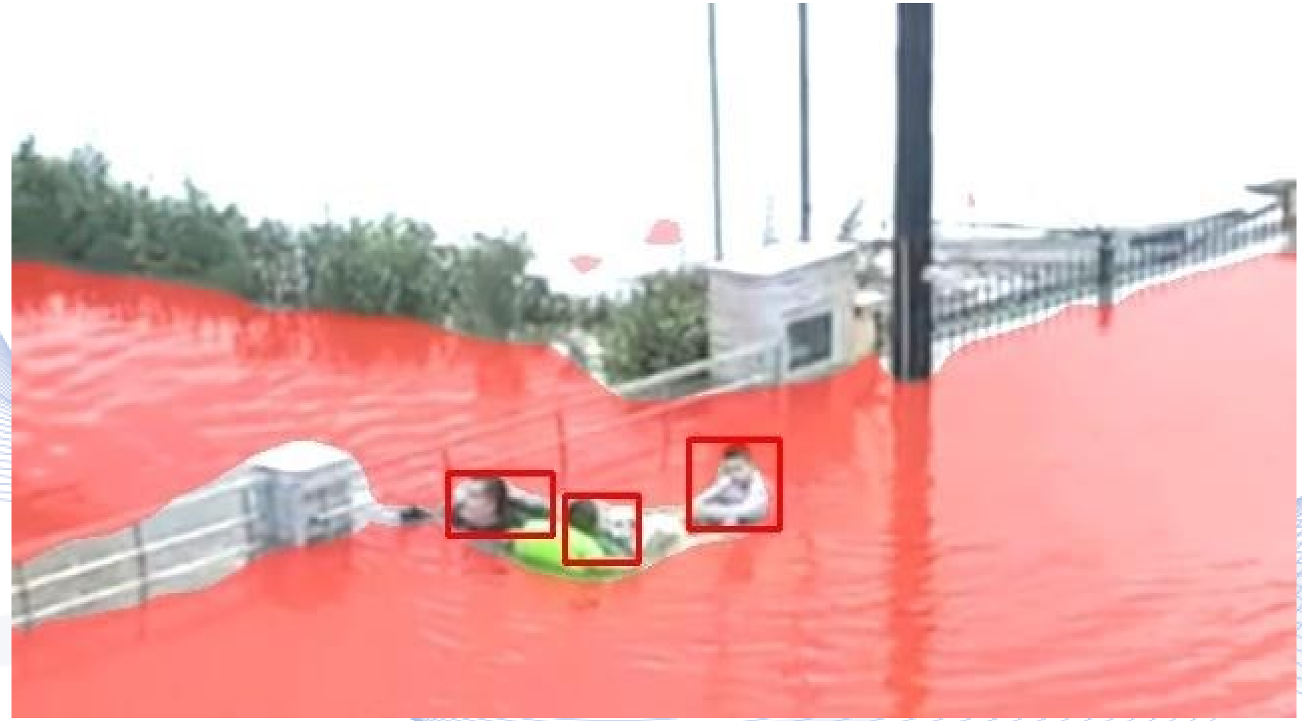
Our system can process big visual data in near real time and send meaningful information to the end users (e.g., the Civil Protection's agents).

Unified system



Graphical depiction of our system’s abstract architecture.

Rescue Alert examples: Case 1



Images from the recent floods in Thessaly.
Our proposed system would generate alerts due to the presence of water pixels inside the bounding boxes.

Rescue Alert examples: Case 2

Sample images displaying the extended bounding boxes.
Our proposed system would generate alerts due to the percentage of water inside the surrounding area.

References

[LI2022] C.Li, L.Li, H.Jiang, et al. “YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications”. arXiv, 2022.

[VAS2017] A. Vaswani, Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. “Attention is all you need”. NeurIPS (2017).

[CAR2020] N. Carion, F. Massa, G. Synnaeve, N. Usanier, A. Kirillov, S. Zagoruyko. “End-to-end object detection with transformers”. arXiv, 2020.

[WAN2024] A.Wang, H.Chen, et al. “YOLOv10: Real-Time End-to-End Object Detection”. arXiv, 2024.

[LI2023] C. Li, L. Li, Y. Geng, H. Jiang, M. Cheng, B. Zhang, Z. Ke, X. Xu, and X. Chu, “Yolov6 v3. 0: A fullscale reloading,” arXiv preprint arXiv:2301.05586, 2023.

[DING2021] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, “Repvgg: Making vgg-style convnets great again,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 13 733–13 742.

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr