Wildfire Image Analysis

Dimitrios Fotiou, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Artificial Intelligence &
Information Analysis Lab

Wildfire Image Analysis

- Forest Fire Computer Vision
- Classical Segmentation for Wildfires
- Deep Learning for Fire & Smoke Segmentation
- Real-World Implementation Challenges

Wildfire Image Analysis

- **Forest Fire Computer Vision**
  - **Classification**
  - **Object detection**
  - **Semantic Segmentation**
- Classical Segmentation for Wildfires
- Deep Learning for Fire & Smoke Segmentation
- Real-World Implementation Challenges

Computer Vision

Computer vision

- Classification → Classify an image into specific classes
- Object detection → Classify and localize objects within an image
- Semantic segmentation → Classify each pixel in an image into a specific class

Computer Vision

Classification

Smoke / No Smoke

Burnt area / No Burnt area

Fire / No Fire
Smoke / No Smoke

Computer Vision

Object Detection

Fire detection

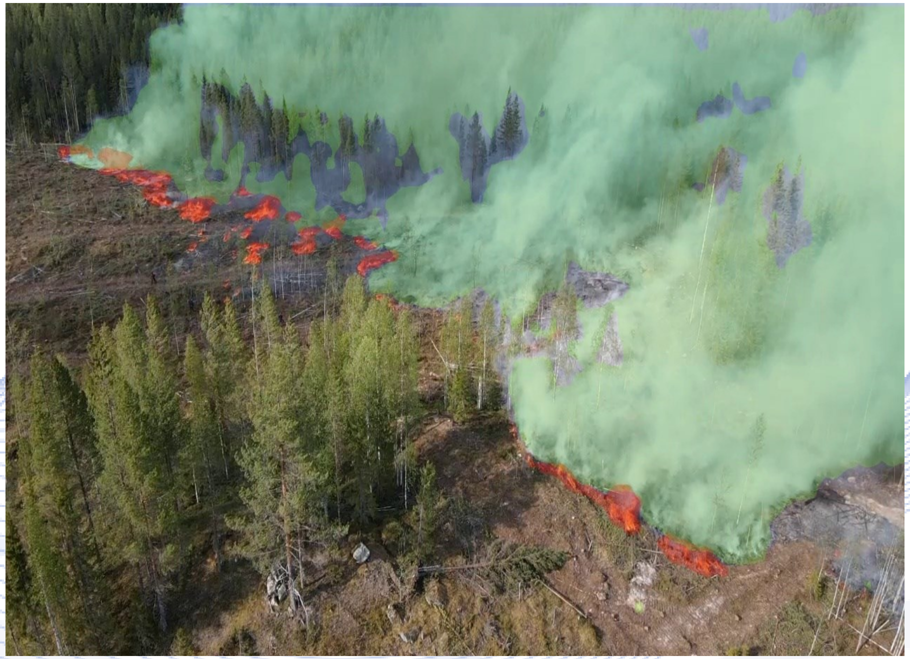
Smoke Detection

Person & Vehicle Detection

Computer Vision

Semantic Segmentation

Fire Segmentation



Fire+Smoke Segmentation

Wildfire Image Analysis

- Computer Vision
- **Classical image segmentation techniques**
  - **Thresholding**
  - **Region Growing**
- Deep semantic image segmentation
- Real Implementation Issues

Image thresholding

The simplest image segmentation problem occurs when an image contains:

- an object having homogenous intensity
- a background with different intensity level

Such image can be segmented in two regions by simple thresholding:

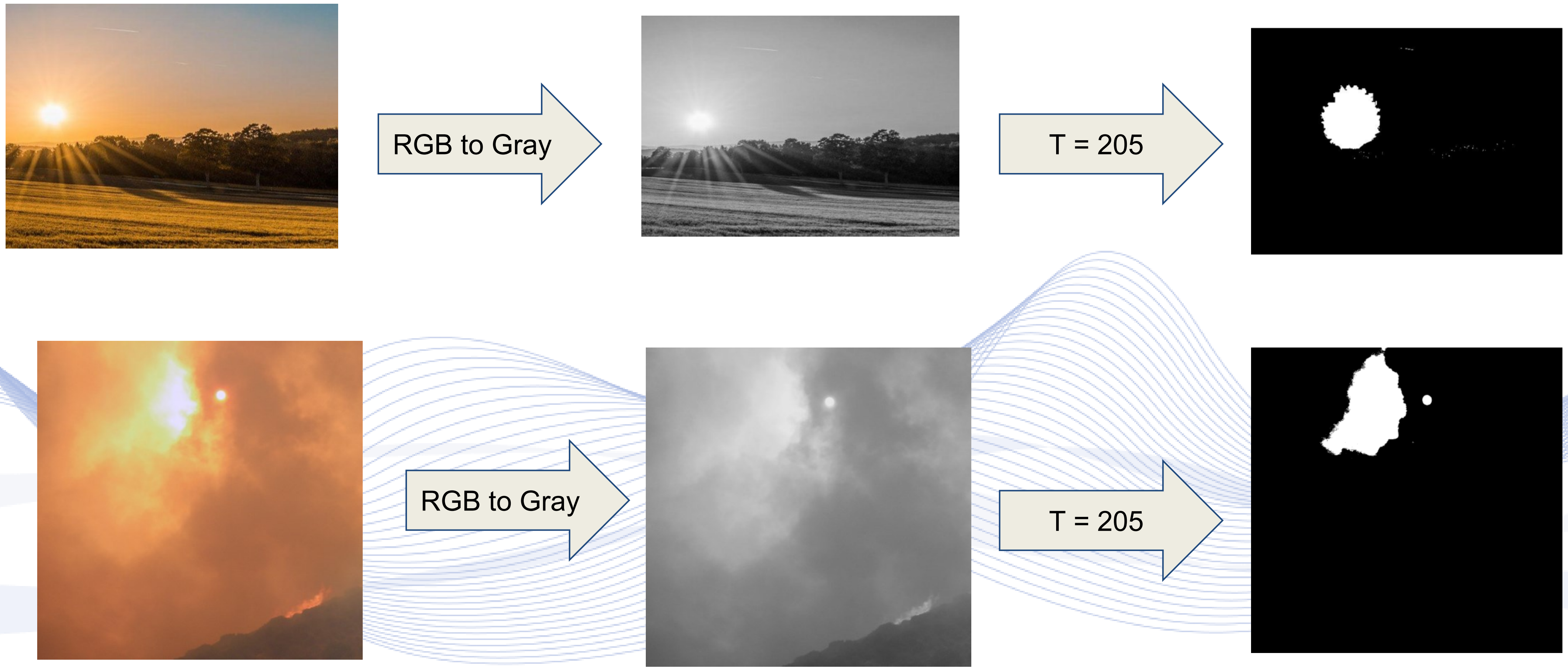
$$g(x, y) = \begin{cases} 1 & f(x, y) \geq T \\ 0 & f(x, y) < T \end{cases}$$

Image thresholding

RGB to Gray

T = 205

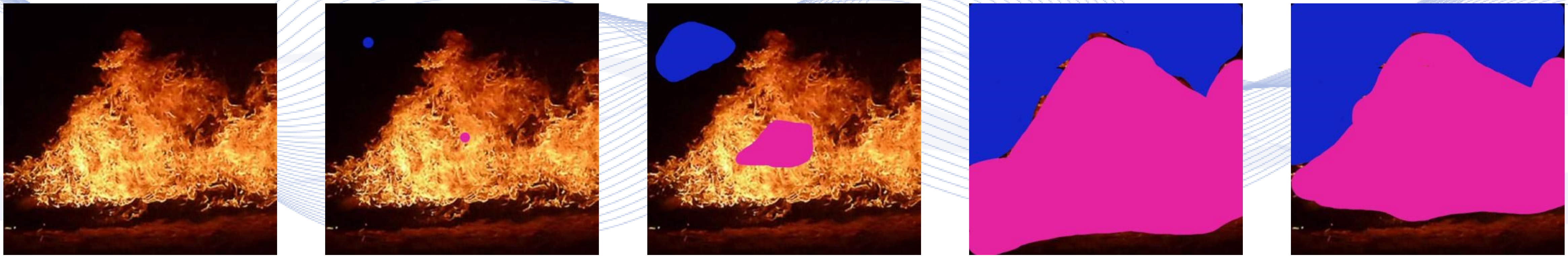
Image thresholding



Region Growing

Basic Steps

- **Seed Selection**: Manually or based on intensity or similar rule
- **Similarity Criteria Definition:** Color, Texture, Intensity
- **Region Growing:** Iteratively add 4- or 8-connected neighboring pixels that meet the criteria.
- **Stopping Condition:** max region size or no other candidate pixels
- **Post-Processing**: Remove noisy segments, merge similar segments



Wildfire Image Analysis

- Forest Fire Computer Vision
- Classical Segmentation for Wildfires
- **Deep Learning for Fire & Smoke Segmentation**
  - Core Concepts
  - RGB Segmentation
  - Multimodal Segmentation
- Real-World Implementation Challenges

Artificial Intelligence & Information Analysis Lab

Core Concepts

Convolution

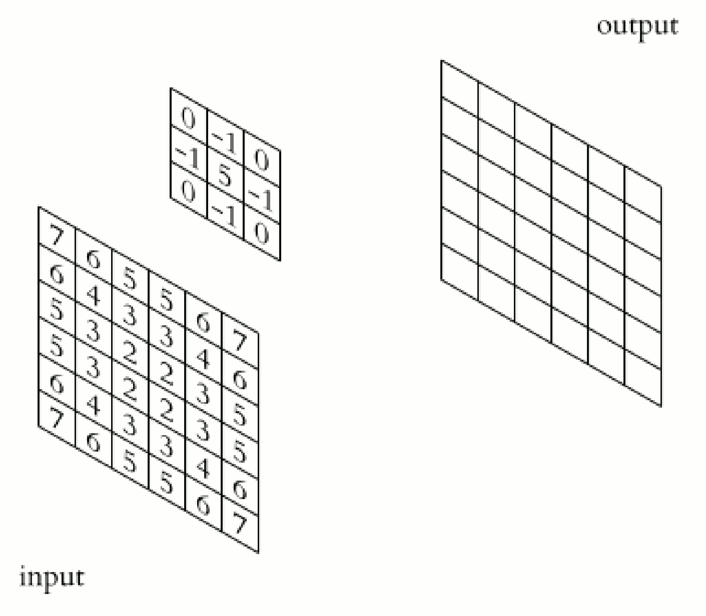
Convolution is a mathematical operation that applies a filter (kernel) to an image to extract specific features like edges, textures, or patterns.

Process:

• A small filter slides over the image.

• The dot product of the filter and overlapping image values is computed.

• The result forms a new, processed image (feature map).



Core Concepts

Edge detection

Original image

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Convolution output

Core Concepts

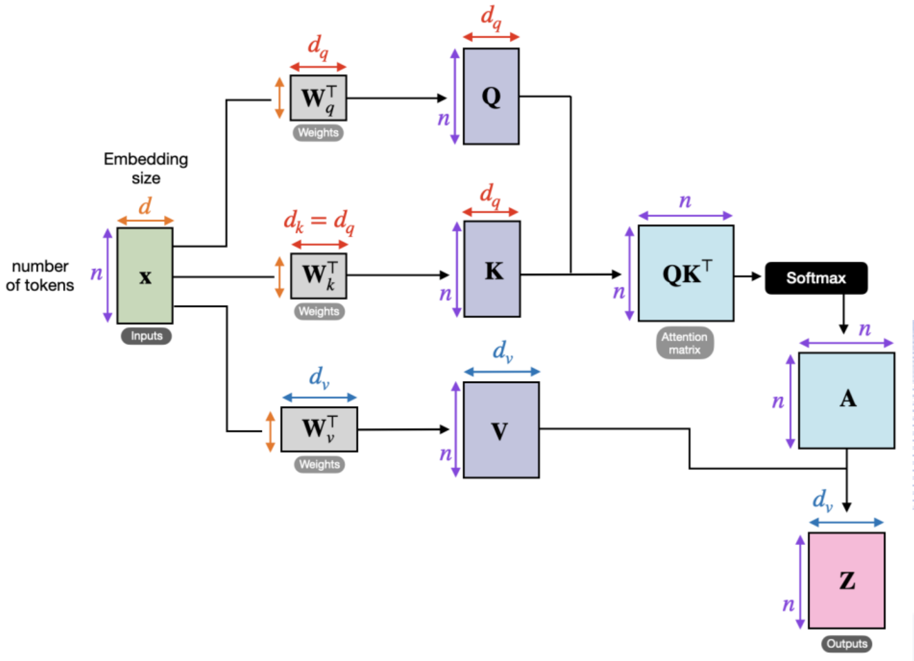
Vision Transformer (ViT) [DOS2020].

- Implementation of transformer architecture in Computer Vision.
- A pure transformer applied directly to sequences of image patches works exceptionally well on image classification, segmentation and object detection tasks.
- Uses self-attention mechanisms to process images

Core Concepts

Self-Attention

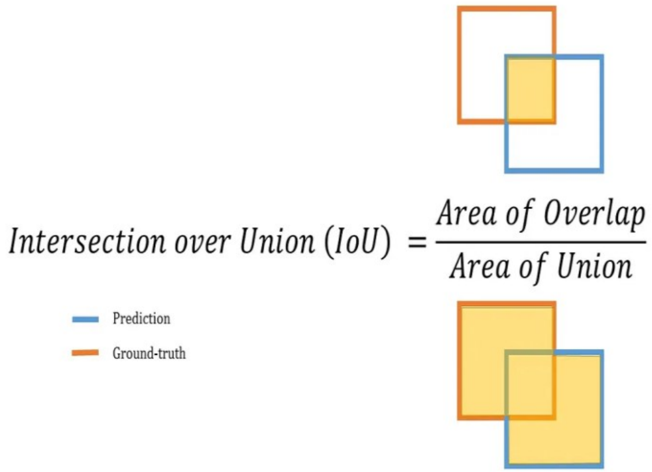
A mechanism which computes a weighted sum of the input data, where the weights are computed based on the similarity between the input features.



Core Concepts

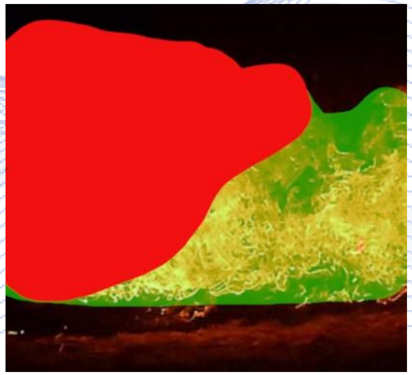
Intersection Over Union (IoU)

The overlap between a predicted bounding box (P) and a ground truth bounding box (G) is measured using IoU:



$$IoU(P, G) = \frac{|P \cap G|}{|P \cup G|}.$$

IoU: 40%



IoU: 85%

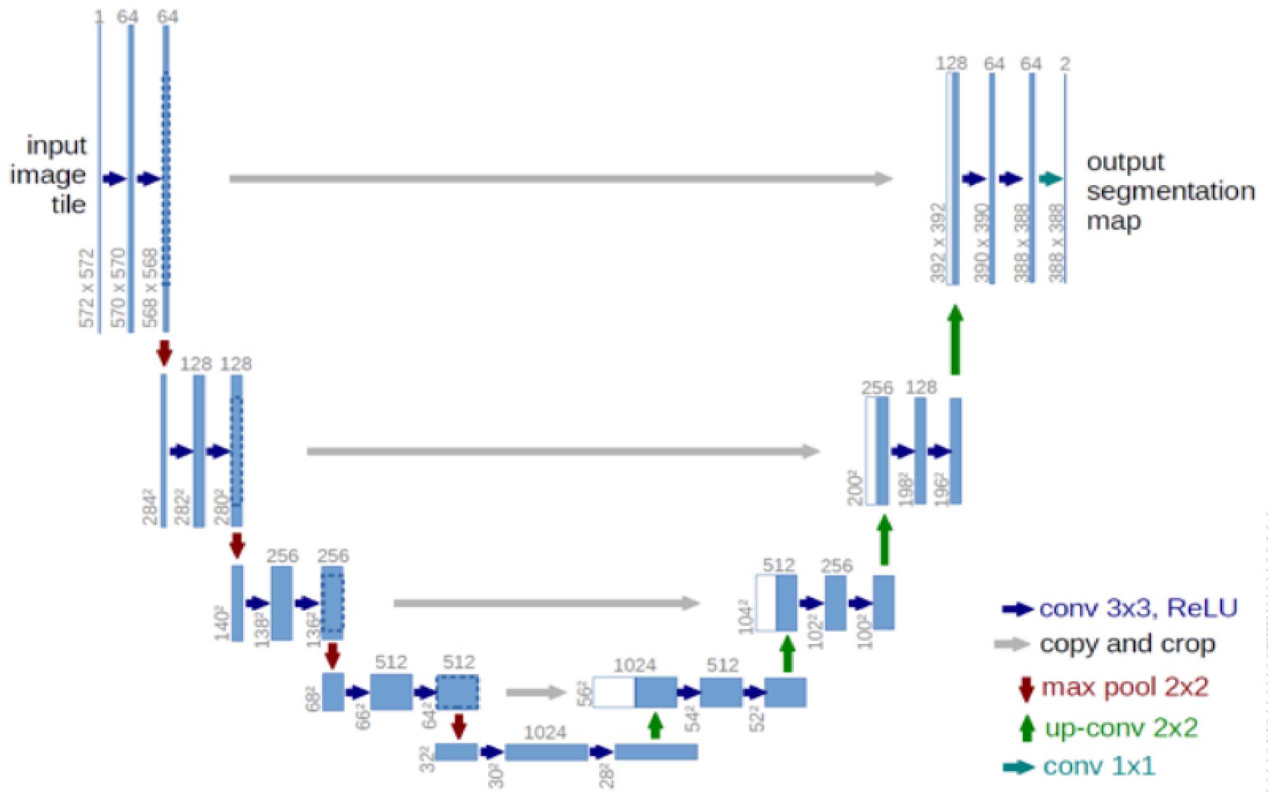
RGB Segmentation

U-Net architecture

- More advanced semantic segmentation network architectures have emerged.
- The capacity of the decoder was expanded by using a ***U-shaped network*** architecture (***U-Net***).
- Consists of a ***contracting path*** to capture context and a ***symmetric expanding path*** that enables precise localization.

RGB Segmentation

U-Net architecture



U-Net network architecture [RON 2015].

RGB Segmentation

BiseNet architecture

• **Two-Stream Network:** Combines spatial and contextual information for high accuracy in segmentation.

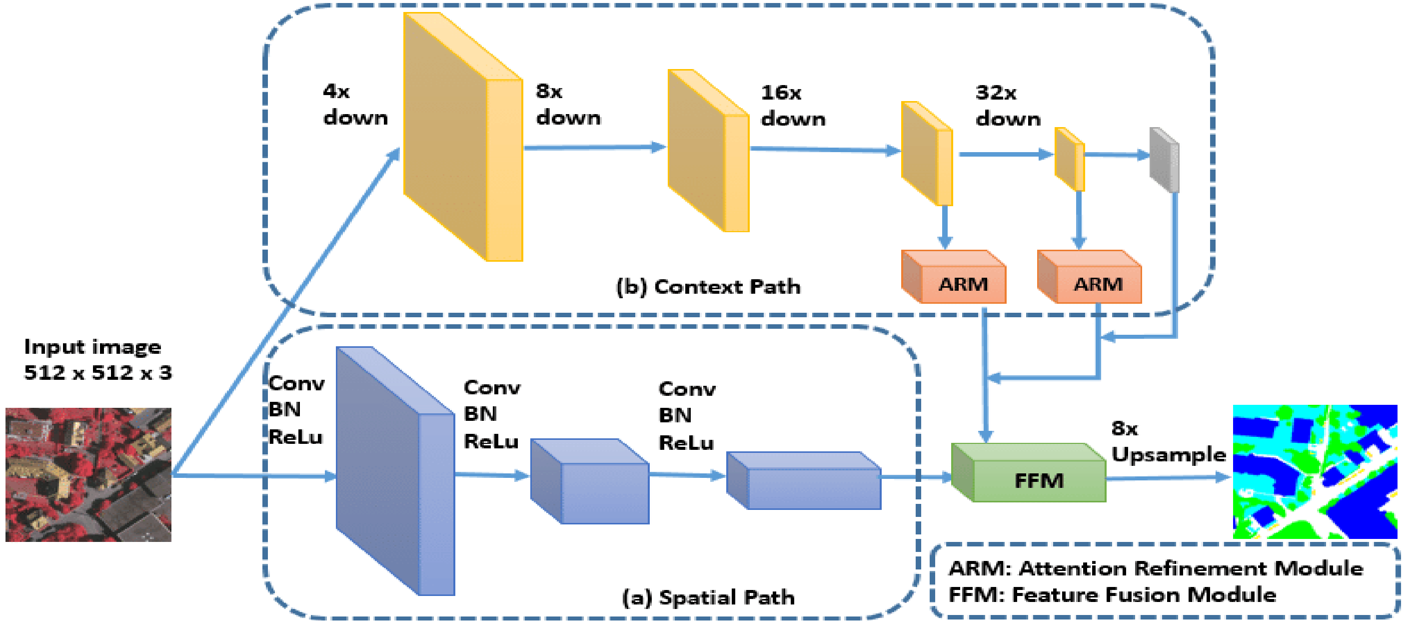
• **Efficient and Fast:** Designed for real-time performance with lightweight structure, ideal for real-time applications like fire detection.

• **Context Path:** Captures large-scale features for better scene understanding.

• **Spatial Path:** Retains high-resolution details for precise boundary segmentation.

RGB Segmentation

BiseNet architecture



BiSeNet architecture [CYO2018]

RGB Segmentation

PIDNet architecture

• ***Triple-Branch Design***: Uses three branches—Proportional (P), integral (I), and derivative (D)—to balance accuracy and efficiency.

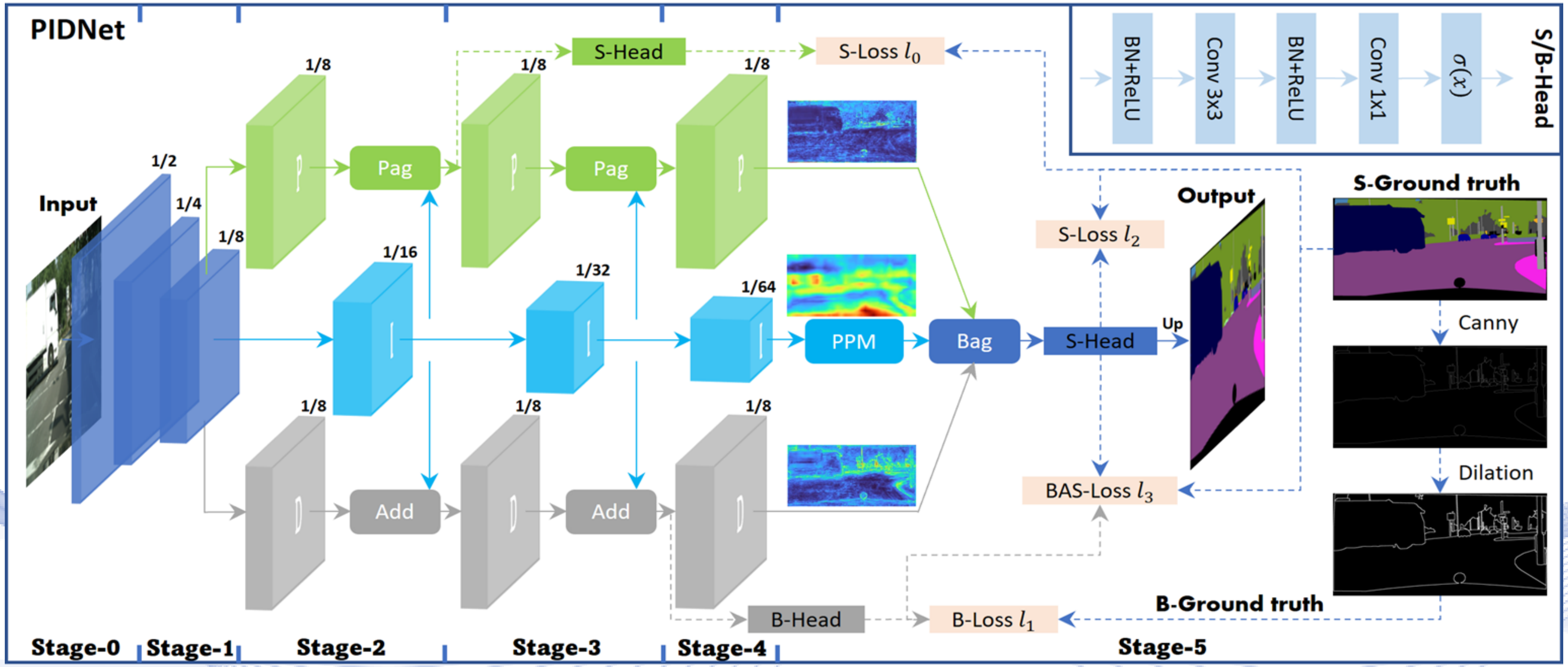
• ***Real-Time Performance***: Optimized for real-time applications, making it suitable for tasks like fire detection in edge environments.

• ***High Precision in Edge Detection***: The Detail branch captures fine edges, crucial for accurately outlining objects in segmentation.

• ***Competitive Accuracy***: Delivers performance close to more complex models, but with much faster inference speeds.

RGB Segmentation

PIDNet architecture



PIDNet architecture [JXU2023]

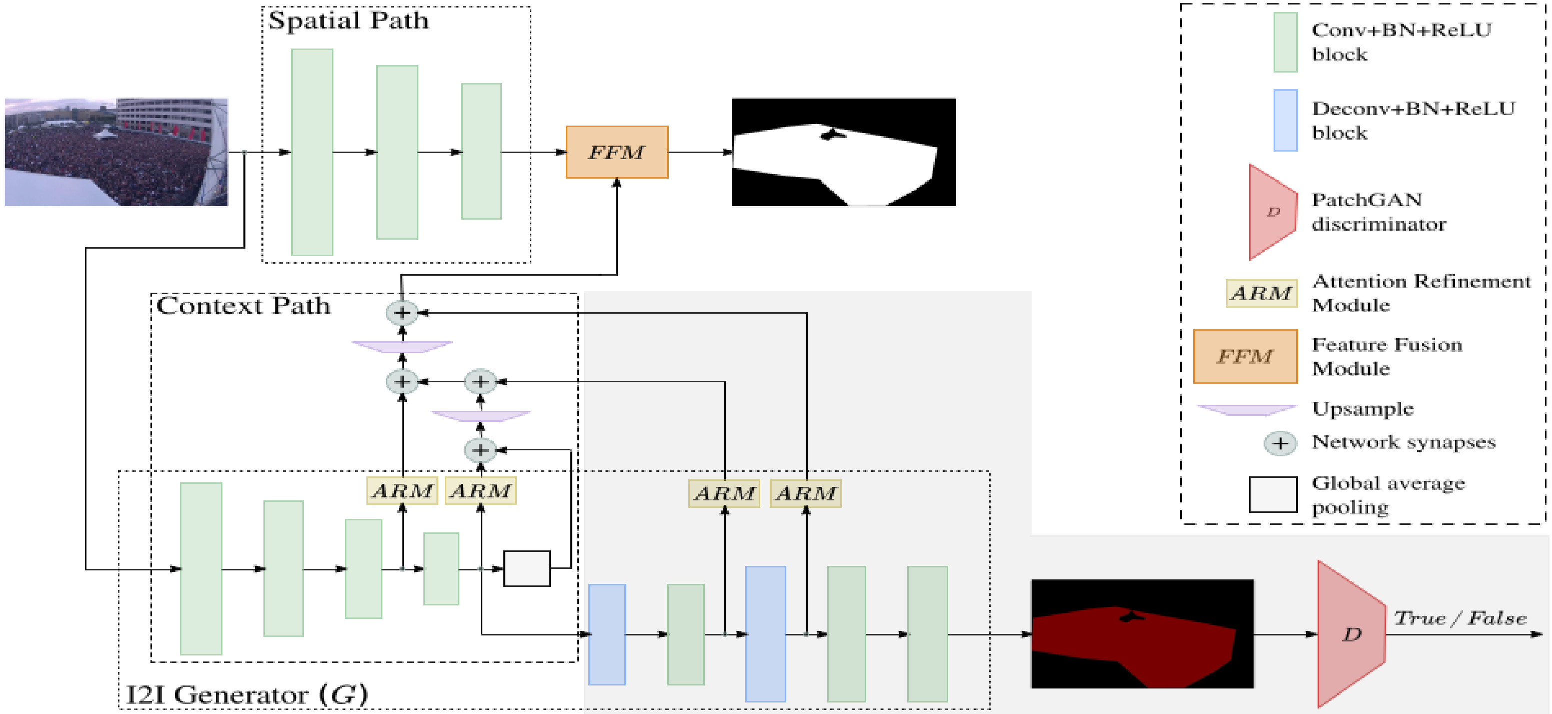
RGB Segmentation

I2I-CNN architecture

- ***Dual-Branch Design***: Adds an auxiliary neural branch to the BiseNet branch for enhanced semantic accuracy without slowing down execution.
- ***GAN-Based Auxiliary Branch***: Trained using a Generative Adversarial Network (GAN) to generate RGB-like segmentation maps, capturing additional semantic information.
- ***Adversarial Training with Discriminator***: The auxiliary branch learns through adversarial loss, where a Discriminator validates its output for improved semantic feature extraction.
- ***Lightweight and Fast***: This network has the same inference speed as Bisenet.

RGB Segmentation

I2I-CNN architecture

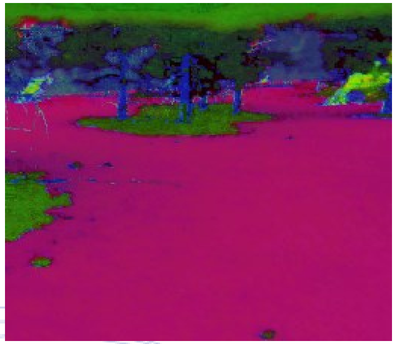


I2I-CNN architecture [PAP2021]

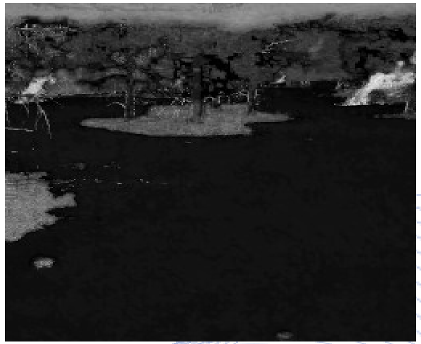
Multimodal Segmentation

Process of creating the S channel (visualization)

RGB



HSV

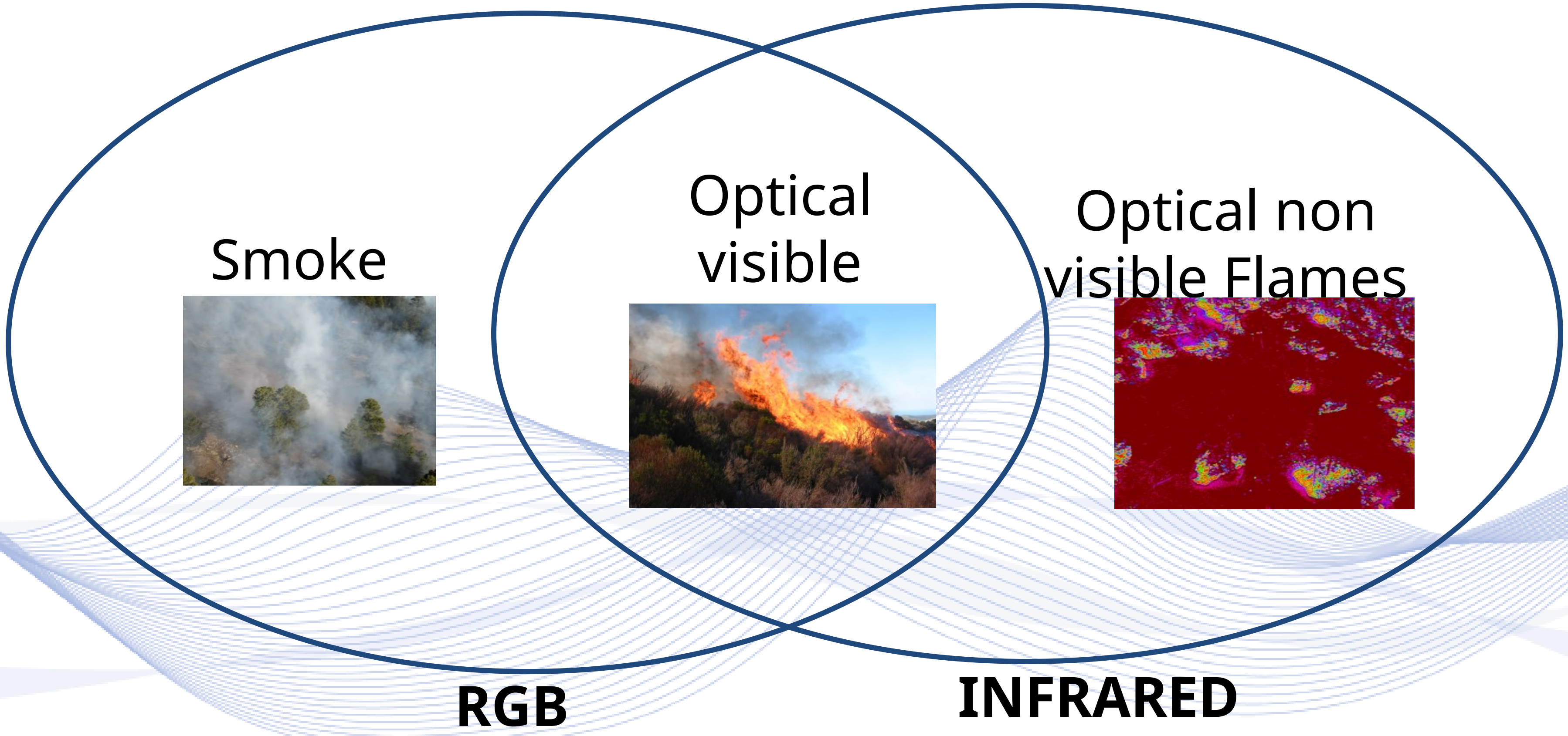


Saturation (S)

Thresholding of S

Multimodal Segmentation

A Venn Diagram of RGB and IR Capabilities



Multimodal Segmentation

Combining IR and RGB:

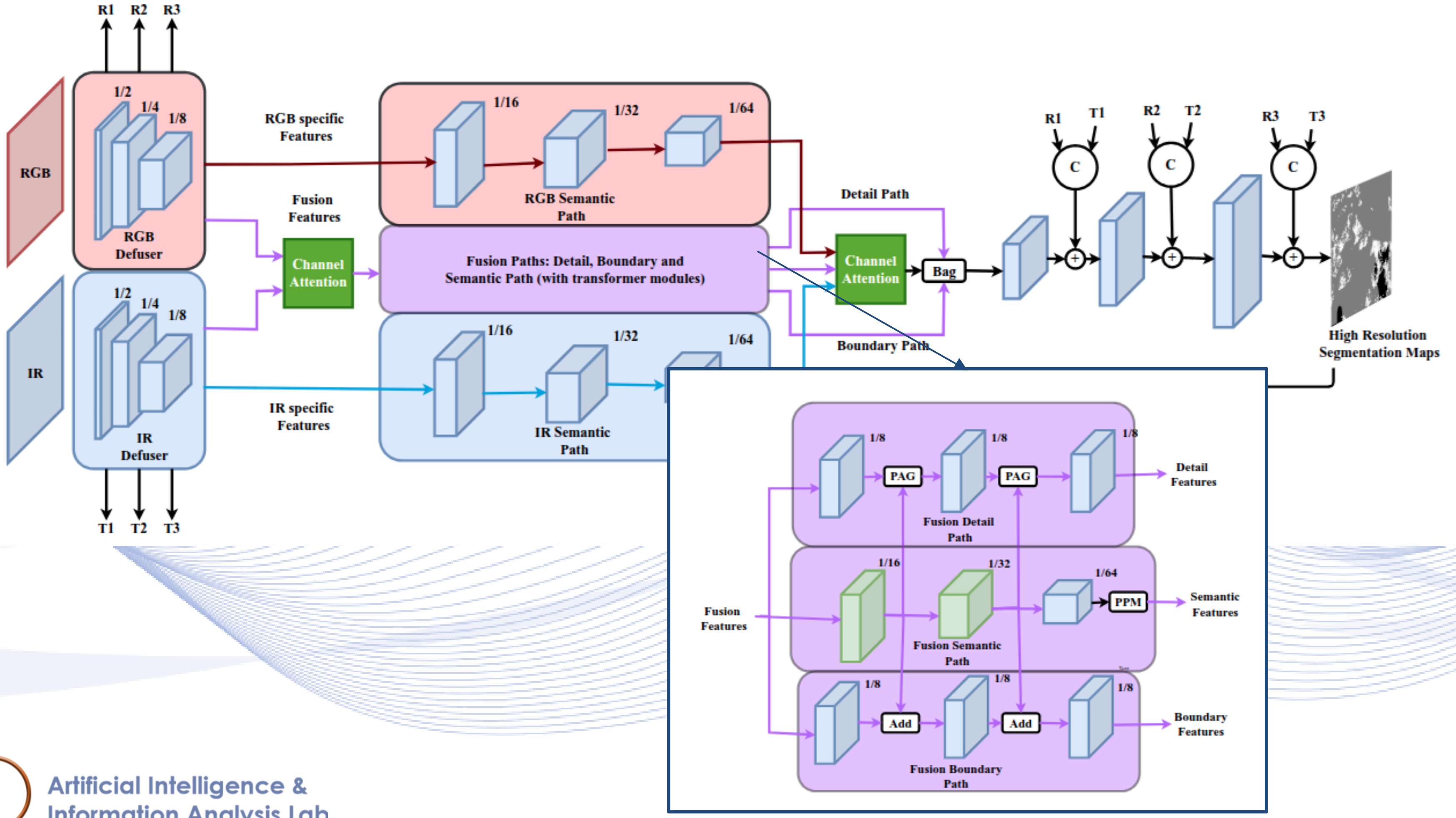
Early Fusion: Concatenate the three RGB channels with the IR image to create a unified 4D input for the DNN.

Intermediate Fusion : Feed the RGB and IR images separately into their respective DNNs, concatenate their intermediate feature maps, and then pass the aggregated map through a common network for further processing.

Late Fusion: Process the RGB and IR images separately through their respective DNNs, then concatenate the segmentation results from both networks to obtain the final output.

Multimodal Segmentation

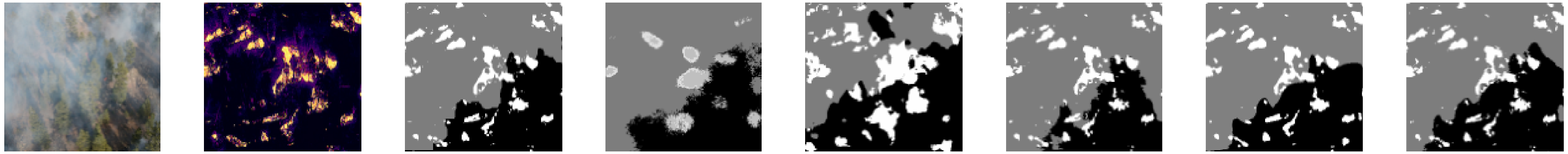
RFFNet



Multimodal Segmentation

RFFNet

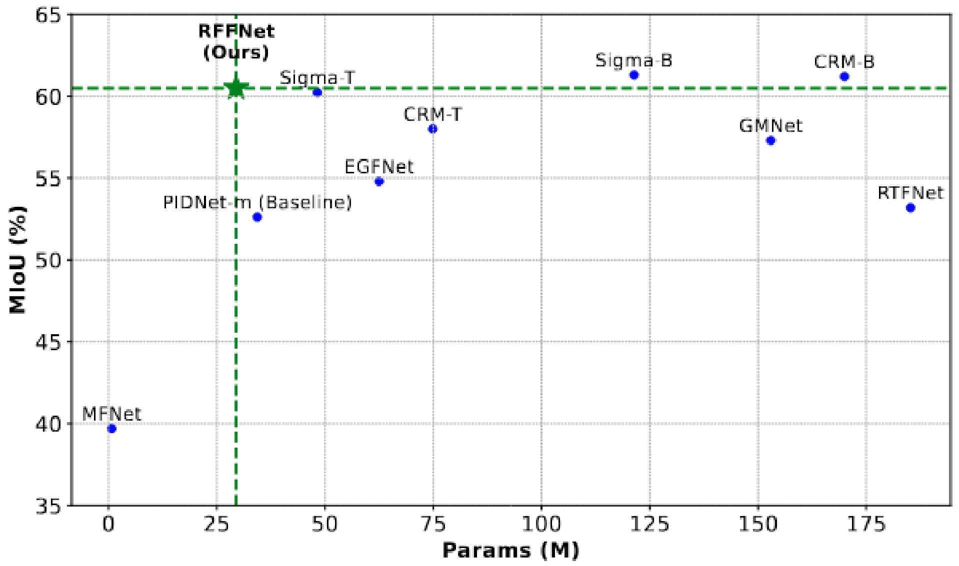
Method	BG Recall (%)	BG IoU (%)	Flame Recall (%)	Flame IoU (%)	Smoke Recall (%)	Smoke IoU (%)	Avg Recall (%)	mIoU (%)
PIDNet-RGB [18]	93.51	90.96	52.13	19.54	81.35	71.14	75.66	61.21
PIDNet-IR [18]	86.76	83.68	89.17	47.66	73.21	44.81	83.05	58.71
PIDNet-Early [18]	95.46	93.65	81.59	52.34	84.64	74.1	88.25	73.90
MFNet [11]	95.45	92.68	**96.08**	65.70	89.06	82.40	93.53	80.26
RTFNet [12]	95.12	76.77	37.36	30.36	89.13	76.25	73.87	65.42
GMNet [32]	88.63	83.65	41.92	15.39	72.04	63.20	67.53	54.08
EGFNet [33]	91.29	88.73	49.59	21.05	81.93	73.16	74.27	60.98
Sigma-T [17]	96.67	93.93	92.54	78.727	88.38	86.12	92.6	86.27
RFFNet (Ours)	**98.2**	**95.08**	93.85	**81.5**	**91.08**	**87.98**	**94.37**	**88.17**



(a) RGB (b) IR (c) GT (d) RTFNet (e) EGFNet (f) MFNet (g) Sigma-T (h) RFFNet

Multimodal Segmentation

RFFNet



Artificial Intelligence &
Information Analysis Lab

Wildfire Image Analysis

- Forest Fire Computer Vision
- Classical Segmentation for Wildfires
- Deep Learning for Fire & Smoke Segmentation
- **Real-World Implementation Challenges**
    - Domain Shift
    - Optimization

Domain Shift

Domain Shift in Natural Disaster Management

- Deep learning models for natural disaster response rely heavily on ***large, labeled datasets***.
- However, annotating such data is both ***time-consuming*** and costly.
- Despite large training datasets, models often encounter ***unseen or novel data*** due to domain shift — changes in conditions, geography, or disaster type.

Artificial Intelligence & Information Analysis Lab

Unsupervised Fire Segmentation

Unsupervised Semantic Segmentation

USS architectures in deep learning do not rely on labeled datasets. However, without prior information about the objects of interest, they struggle to achieve the desired clustering.

Unsupervised segmentation results that correspond to the above raw images

Prompted fire segmentation

Raw Images

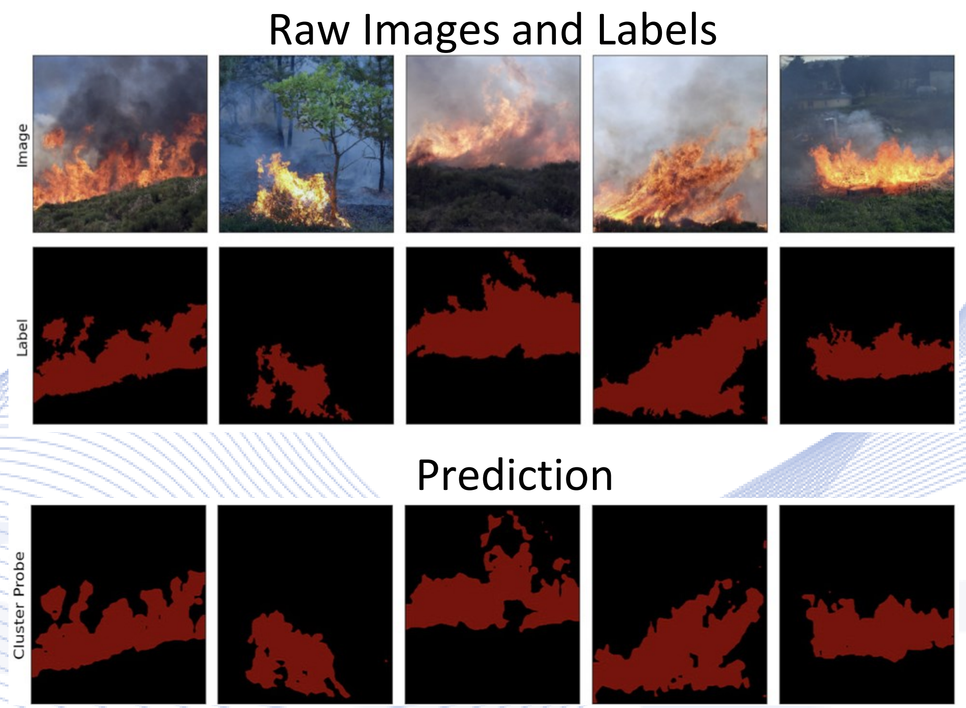
We select a single image from the dataset and specify only one point where our object of interest is located.

1. Combine the raw images with the signal from the annotated point.
2. Push fire representations closer together in the feature space
3. Create a cluster head that separates fire from the background

Artificial Intelligence & Information Analysis Lab

Prompted fire segmentation

- Unsupervised performance : 50 % mIoU
- Our performance : 75 % mIoU
- Our approach achieves a 25% increase in mIoU using only a single point to indicate fire. Visualizations show that our results closely match the actual labels. This method can be extended to other classes, such as smoke, flood, and more



Artificial Intelligence & Information Analysis Lab

Test-Time Domain Adaptation

- **TTDA** is a technique in machine learning where a model adapts to a new (target) domain during inference, without access to target labels during training.
- Unlike traditional domain adaptation, TTDA assumes:
  - No target data during training.
  - Only the trained source model and test samples are available.
  - Adaptation happens on-the-fly during testing.

Test-Time Domain Adaptation

- **Entropy Minimization:** Encourages consistent and confident predictions across augmented test samples by minimizing the uncertainty in model outputs.

- **Pseudo-labels:** Generates stable labels from test-time predictions, guiding refinement during inference without access to ground truth.

- **Memory Banks:** Maintains history features that produce robust feature representations and though consistency and feature alignment try to alleviate domain shift.

Optimization

- **Why Optimization of Fire Segmentation is Essential**
  - Limited Resources
  - Medium-range GPU with limited VRAM
  - Simultaneous frame input from multiple drones
  - Fast response for early fire detection
- **How Optimization is Achieved**
  - Designing a real-time segmentation pipeline
  - Applying quantization to reduce floating-point precision and memory usage
  - Using the TensorRT framework for inference optimization

References

[PIT2000] I. Pitas, "Digital image processing algorithms and applications", Wiley 2000.
[LON2015] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation", In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
[RON 2015] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation", In Proceedings of the International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, 2015.
[CHE2014] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs", arXiv preprint arXiv:1412.7062, 2014.
[TOR2014] O.A.J del Toro, O. Goksel, B.Menze, H. Muller, G. Langs, "VISCERAL–VISual Concept Extraction challenge in RAdioLogy: ISBI 2014 challenge organization", In Proceedings of the VISCERAL Challenge at ISBI, 2014.
[YU2018] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, " Bisenet: Bilateral segmentation network for real-time semantic segmentation", In Proceedings of the European conference on computer vision (ECCV), 2018.
[ZHU2019] J. Zhuang, J. Yang, L. Gu, N. Dvornek, "ShelfNet for Fast Semantic Segmentation", In Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.
[CHE2017] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs", IEEE transactions on pattern analysis and machine intelligence (PAMI), 2017

References

[EVE2011] M. Everingham, John Winn, “The PASCAL visual object classes challenge 2012 (VOC2012) development kit”, Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep, 2011.
[COR2016] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, “The cityscapes dataset for semantic urban scene understanding”, In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016.
[CHE2018] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, A. Hartwig, “Encoder-decoder with atrous separable convolution for semantic image segmentation”, In Proceedings of the European conference on computer vision (ECCV). 2018.
[MOR2018] G. Morales, G. Kemper, G. Sevillano, D. Arteaga, I. Ortega, J. Telles, “Automatic segmentation of Mauritia flexuosa in unmanned aerial vehicle (UAV) imagery using deep learning.” Forests, 2018.
[YUA2019] Y. Yuan, X. Chen, J. Wang, “Object-contextual representations for semantic segmentation.” arXiv preprint arXiv:1909.11065, 2019.
[TAK2019] T. Takikawa, D. Acuna, V. Jampani, S. Fidler, “Gated-SCNN: Gated shape CNNs for semantic segmentation.” In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
[HUA2019] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, W.Liu, “CCNet: Criss-cross attention for semantic segmentation.” In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
[CHE2018] L.-C. Chen, et al, “Encoder-decoder with atrous separable convolution for semantic image segmentation.” In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

References

[MOU2016] A. Mousavian, H. Pirsiavash, J. Kosecḱa, “Joint semantic segmentation and depth estimation with deep convolutional networks.”, In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016.

[LIU2018] J. Liu, Y. Wang, Y. Li, J. Fu, J. Li, H. Lu, “Collaborative deconvolutional neural networks for joint depth estimation and semantic segmentation.”, IEEE transactions on neural networks and learning systems, 2018.

[ALA2020] M. Aladem, S.A. Rawashdeh. “A single-stream segmentation and depth prediction CNN for autonomous driving.”, IEEE Intelligent Systems, 2020.

[CHE2019] P.Y. Chen, A. H Liu, Y.C. Liu, Y.C.F Wang, “Towards scene understanding: Unsupervised monocular depth estimation with semantic-aware representation.”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[QI2017] X. Qi, R. Liao, J. Jia, S. Fidler, R. Urtasun, “3D graph neural networks for RGBD semantic segmentation.”, In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[KEN2018] A. Kendall, Y. Gal, R. Cipolla, “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics.”, In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2018.

[ZHA2017] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, “Pyramid scene parsing network.”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[APOLLO] http://apolloscape.auto/

References

[DOS2020] A. DOSOVITSKIY, An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020.

[KRI2023] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P. Segment anything. InProceedings of the IEEE/CVF International Conference on Computer Vision 2023 (pp. 4015-4026).

[SHA2021] Shamsoshoara, Alireza, et al. "Aerial imagery pile burn detection using deep learning: The FLAME dataset." *Computer Networks* 193 (2021): 108001.

[CYU2018] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in European Conference on Computer Vision. Springer, 2018, pp. 334–349.

[PAP2021] C. Papaioannidis, I. Mademlis, and I. Pitas, "Autonomous uav safety by visual human crowd detection using multi-task deep neural networks," in 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 11 074–11 080.

[JXU2023] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 19 529–19 539.

[BIT2024] *CNN Architecture - Detailed Explanation*. (n.d.). InterviewBit. https://www.interviewbit.com/blog/cnn-architecture/

[SUP2024]SuperAnnotate AI Inc. (n.d.). *Semantic segmentation: Complete guide [Updated 2024] | SuperAnnotate*. SuperAnnotate.

References

[TZI2023] M. D. Tzimas, C. Papaioannidis, V. Mygdalis, and I. Pitas, "Evaluating Deep Neural Network-based Fire Detection for Natural Disaster Management," in 2023 IEEE/ACM 16th International Conference on Utility and Cloud Computing (UCC '23), Taormina (Messina), Italy, Dec. 2023.

[ZHA2024] Zhao, Yian, et al. "Detrs beat yolos on real-time object detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.

[THA2023]Thakur, N. (2023, June 2). A detailed introduction to Two Stage Object Detectors. *Medium*. https://namrata-thakur893.medium.com/a-detailed-introduction-to-two-stage-object-detectors-d4ba0c06b14e

[KUK2023] Kukil, & Kukil. (2023, August 4). *Intersection over union IOU in object detection segmentation*. LearnOpenCV – Learn OpenCV, PyTorch, Keras, Tensorflow With Code, & Tutorials. https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/

Dimitrios Fotiou, Vasileios Mygdalis, Ioannis Pitas. RoboFireFuseNet: Robust Fusion of Visible and Infrared Wildfire Imaging for Real-Time Flame and Smoke Segmentation. TechRxiv. May 08, 2025.

Tzimas, Matthaios Dimitrios and Mygdalis, Vasileios and Papaioannidis, Christos and Pitas, Ioannis, Extreme Weakly Supervised Binary Semantic Image Segmentation Via One-Pixel Supervision. Available at SSRN: https://ssrn.com/abstract=5217487 or http://dx.doi.org/10.2139/ssrn.5217487

NVIDIA Corporation. (2024). NVIDIA TensorRT: High performance deep learning inference. Retrieved July 2, 2025, from https://developer.nvidia.com/tensorrt

Q & A

Thank you very much for your attention!

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