Real-Time Image segmentation

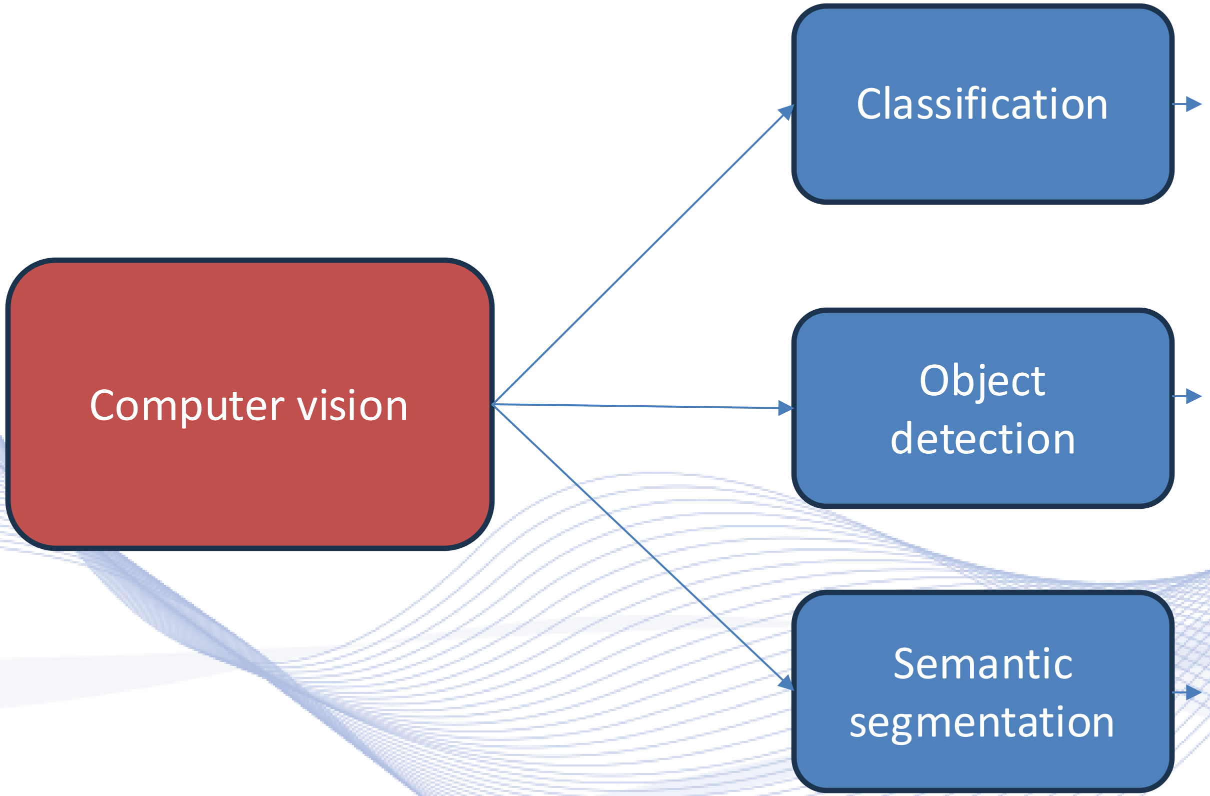
M. Tzimas, D. Fotiou, P. Dinopoulos, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Artificial Intelligence & Information Analysis Lab

Real-Time Image Segmentation

- **Computer Vision**
  - **Classification**
  - **Object detection**
  - **Semantic Segmentation**
- Classical image segmentation techniques
- Deep semantic image segmentation
- Fire Detection
- Fire Segmentation

Computer vision



Classify an image into a specific class

Classify and localize objects within an image

Classify each pixel in an image into a specific class

Artificial Intelligence & Information Analysis Lab

Computer Vision

Classification

Fire / No Fire

Smoke / No Smoke

Burnt area / No Burnt area

Computer Vision

Object Detection

Fire detection

Smoke Detection

Computer Vision

- Object Detection = classification + localization
- Find **what** is in a picture as well as **where** it is

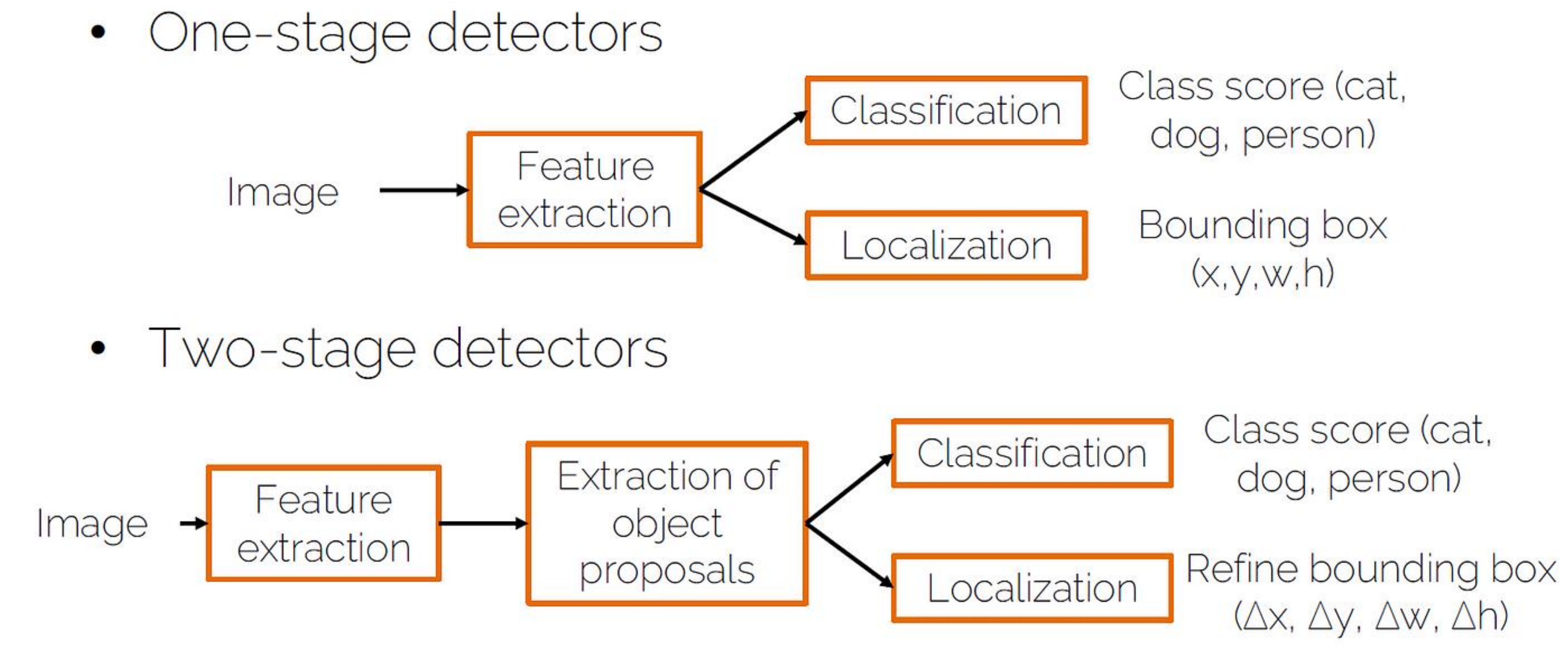
Computer Vision

Classification – Regression

- Given a training set of **images annotated with bounding boxes** (coordinates and class per depicted object)
  - Classification: predict probabilities that each box belongs to each of the classes present in the dataset
  - Regression: for each depicted object predict bounding box coordinates in some predefined format, e.g., coordinates of the bounding box center along with its width and height (x, y, w, h)

Computer Vision

One stage vs Two stage object detection architectures

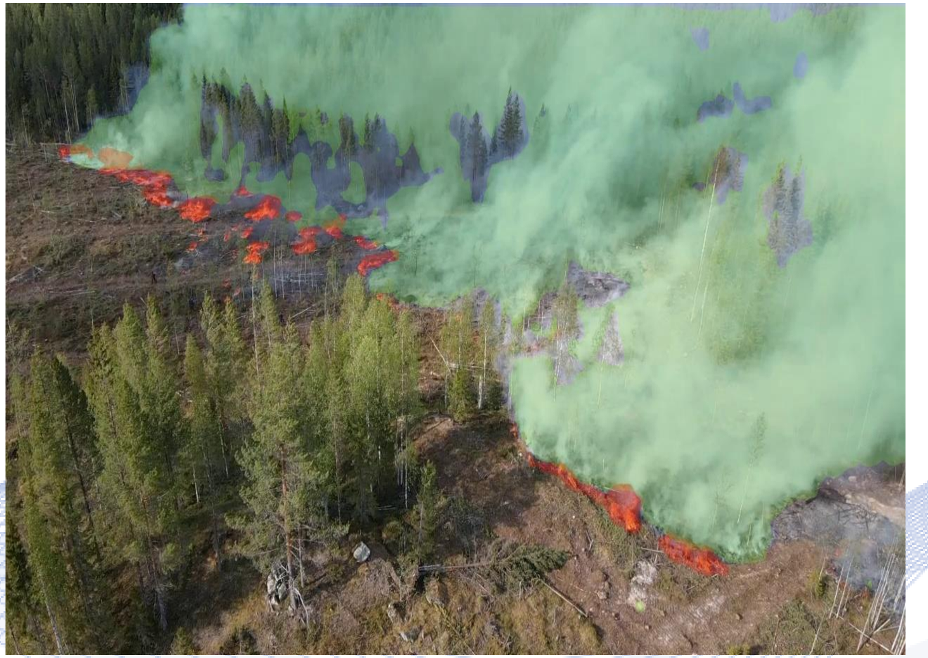


[THA2023]

Computer Vision

Semantic Segmentation

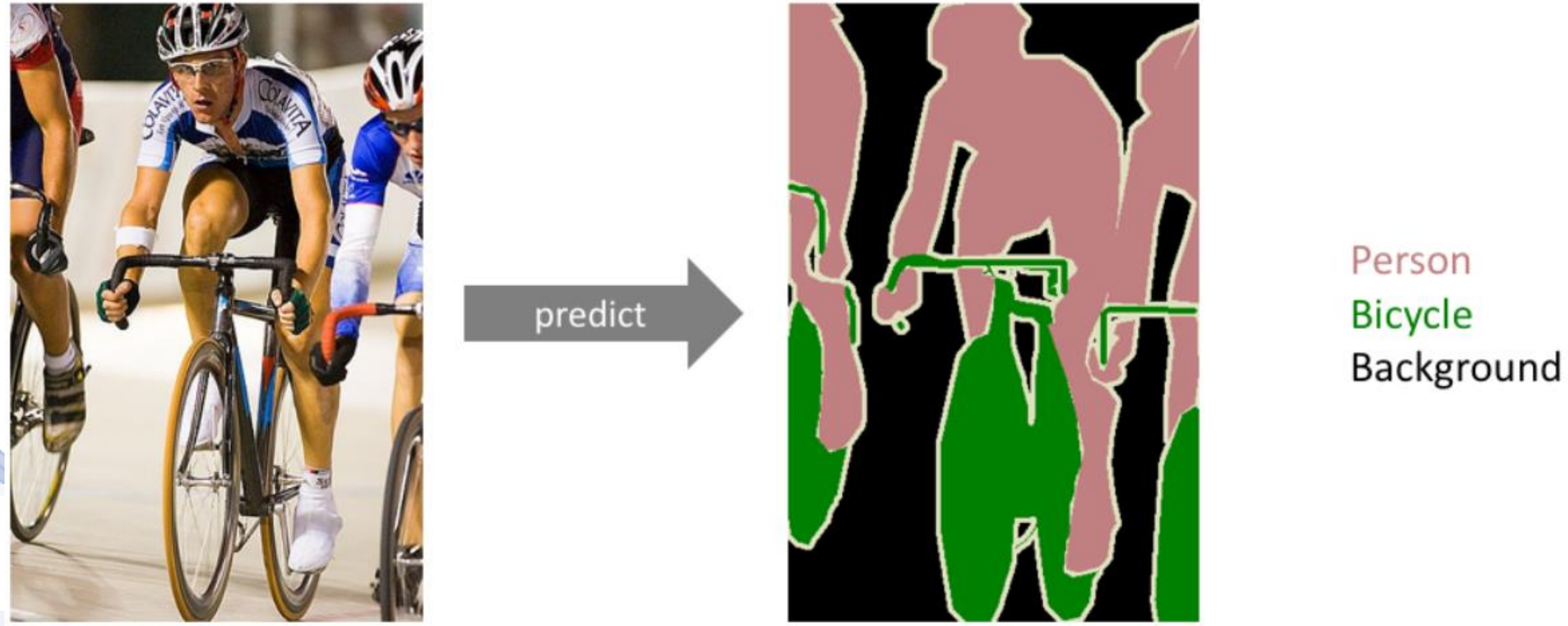
Fire Segmentation



Fire/Smoke Segmentation

Computer Vision

Semantic Segmentation

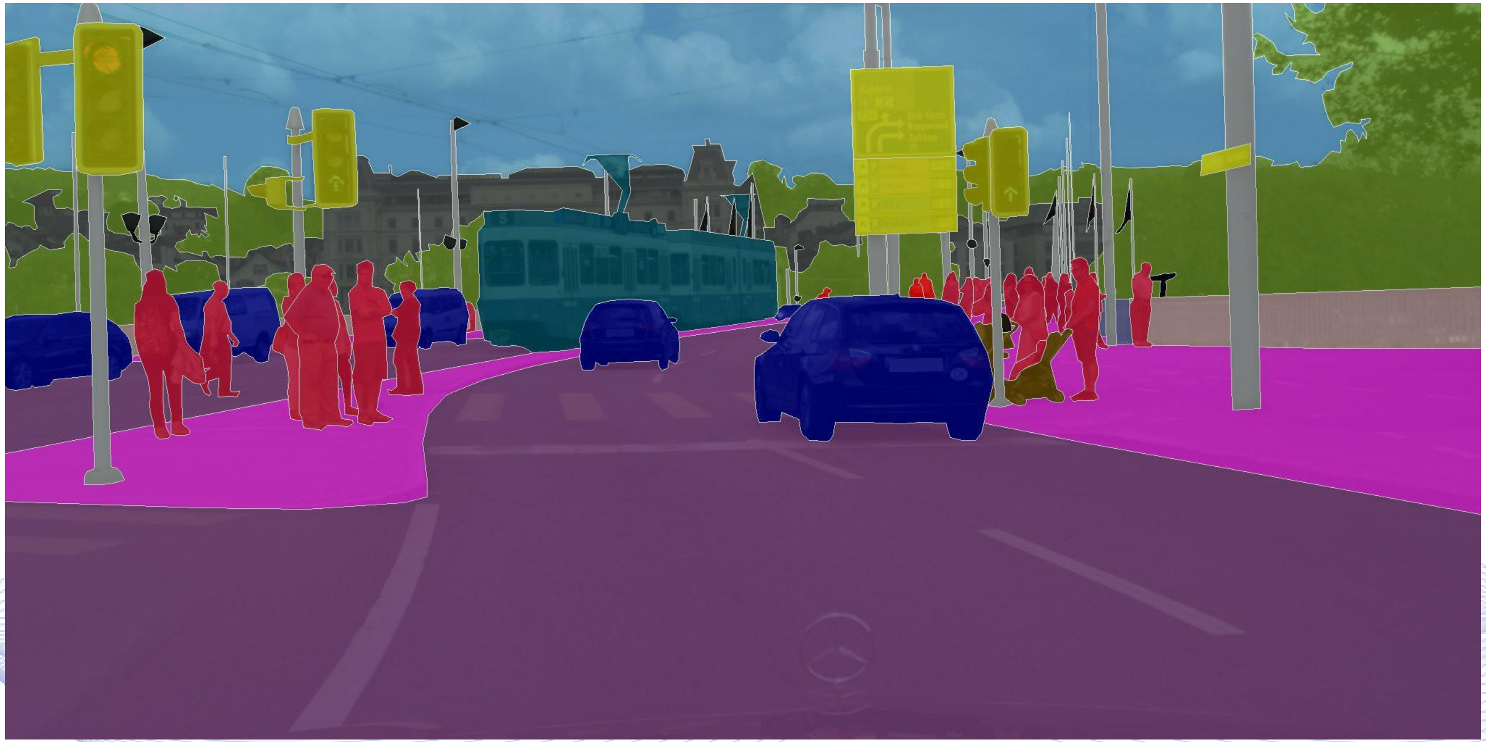


Semantic image segmentation of a sports event [EVE2011].

Computer Vision

Semantic Segmentation

- Autonomous driving.

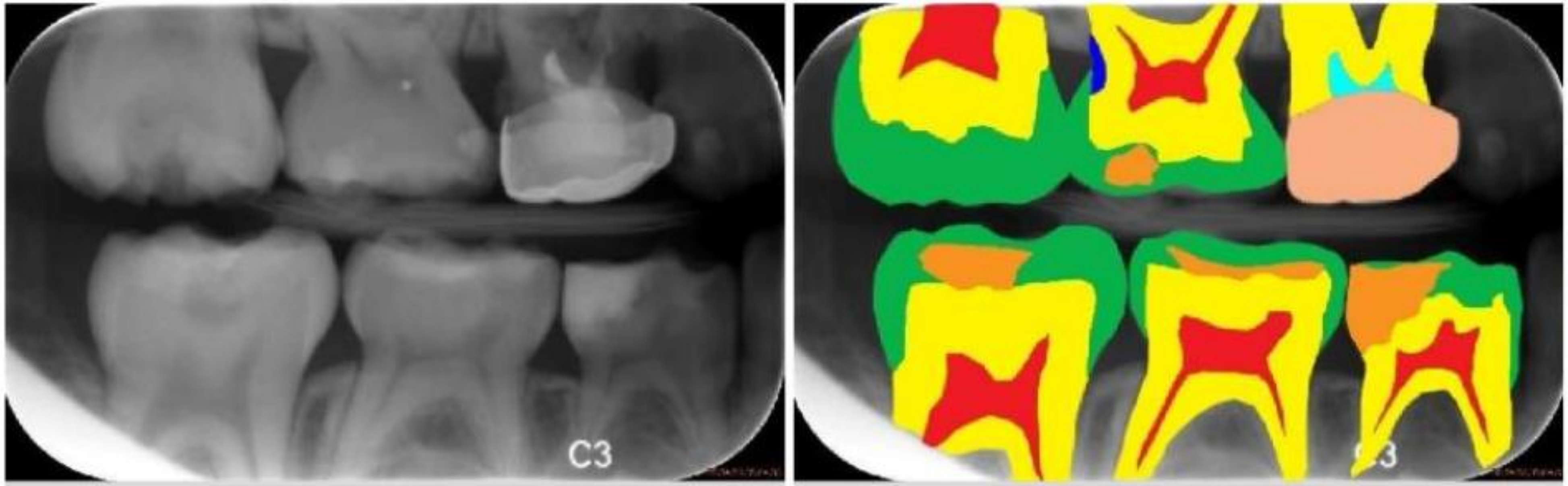


Semantic image segmentation for autonomous driving [COR2016].

Computer Vision

Semantic Segmentation

- Medical purposes.



Semantic dental Xray segmentation [TOR2014].

Computer Vision

Semantic Segmentation

- An image domain $\mathcal{X}$ must be segmented in N different regions $R_1, \dots, R_N$.
- The segmentation rule is a logical predicate of the form $P(\mathcal{R})$
- Image segmentation partitions the set $\mathcal{X}$ into the subsets R_i, $i = 1, \dots, N$, having the following properties:

$$\mathcal{X} = \cup_{i=1}^{N} R_i,$$
$$R_i \cap R_j = \emptyset, \qquad i \neq j,$$
$$P(R_i) = TRUE, \qquad i = 1, \dots, N,$$
$$P\left(R_i \cup R_j\right) = FALSE, \qquad i \neq j,$$

Real-Time Image Segmentation

- Computer Vision
- **Classical image segmentation techniques**
- Deep semantic image segmentation
- Fire Detection
- Fire Segmentation

Image thresholding

- The simplest image segmentation problem occurs when an image contains.
  - an object having homogenous intensity.
  - a background with a different intensity level.
- Such an image can be segmented in two regions by simple thresholding:

$$g(x,y) = \begin{cases} 1, if\ f(x,y) \\ 0, otherwise \end{cases}$$

- The choice of threshold T can be based on the image histogram.

Image thresholding

Image thresholding.

Image thresholding

(a)

(b)

a) Original image; b) Image segmentation in four equirange regions.

Region Growing

- The pixel seeds are chosen in a supervised mode.
- At least one seed s_i, $i = 1, \ldots, N$ is chosen per image region R_i.
- In order to implement region growing, we need a rule describing a growth mechanism and a rule checking the homogeneity of the regions after each growth step.

Region Growing

- The growth mechanism is simple: at each stage (k) and for each region $R_I^{(k)}, i = 1, \dots, N$, we check if there are unclassified pixels in the 8-neighbourhood of each pixel of the region border.
- Before assigning such a pixel $\mathbf{x}$ to a region $R_I^{(k)}$, we check the region homogeneity:

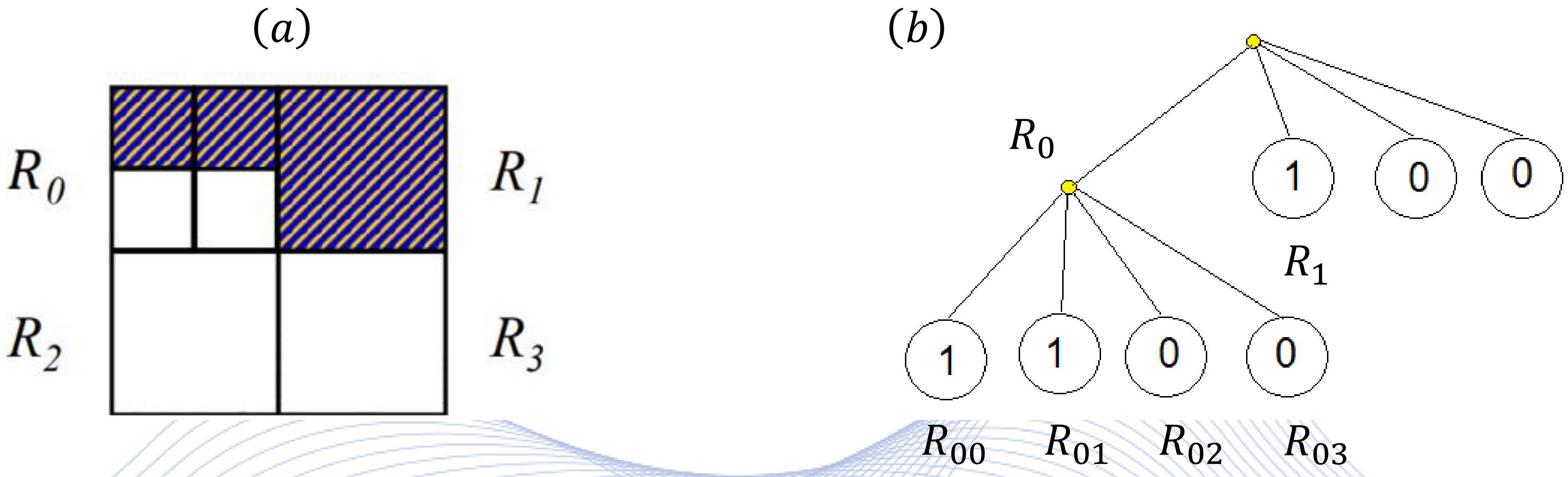
$$P\left(R_I^{(k)} \cup \{\mathbf{x}\}\right) = TRUE$$

is still valid.

Split/merge algorithm

- If the original image is square $N \times N$, having dimensions that are powers of 2($N = 2^n$):
  - All regions produced by the splitting algorithm are squares having dimensions $M \times M$, where M is a power of 2 as well $(M = 2^m, m \leq n)$.
  - Since the procedure is recursive, it produces an image presentation that can be described by a tree whose nodes have four sons each.
  - Such a tree is called a quadtree and is a very convenient region representation scheme.

Split/merge algorithm



a) Image segmentation by region splitting; b) Quadtree.

Deep Semantic Image segmentation

- Introduction
- Classical image segmentation techniques
- **Deep semantic image segmentation**
- Applications

Deep Semantic Image segmentation

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).

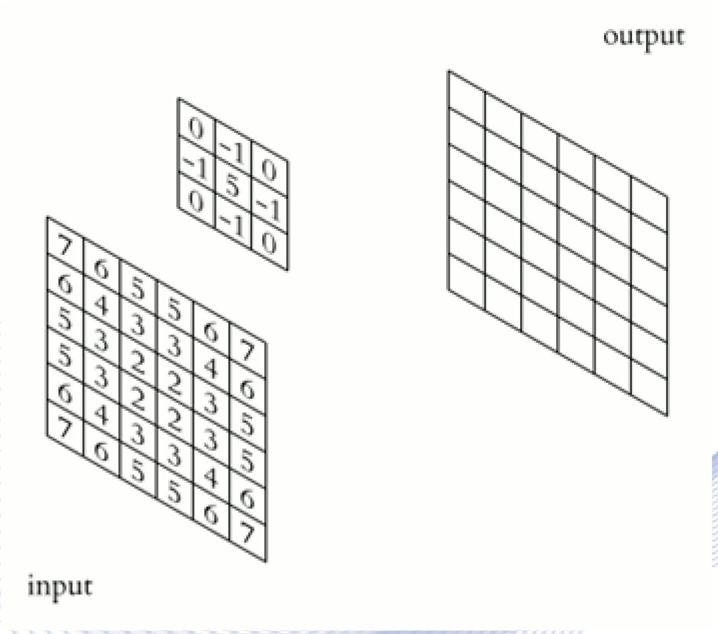
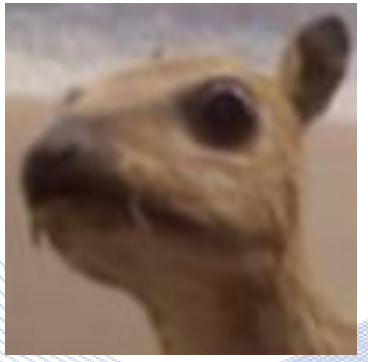


Image blurring

Original image

$$\mathbf{W} = \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

Convolution output

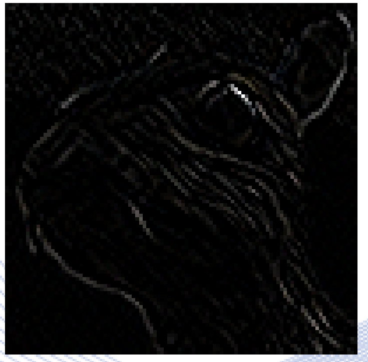


Edge detection

Original image

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

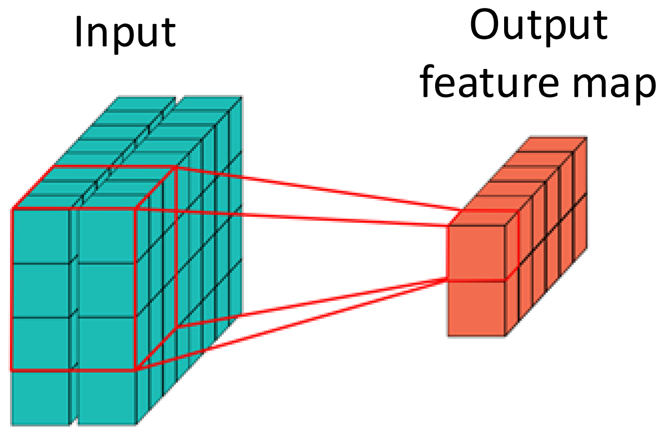
Convolution output



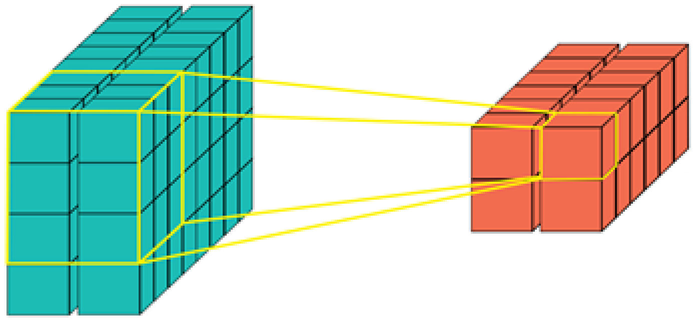
Deep Semantic Image segmentation

Convolutional layer

Kernels can have more than just two dimensions; they may also include depth.



Multiple convolutional kernels can be applied to the same input simultaneously



Deep Semantic Image segmentation

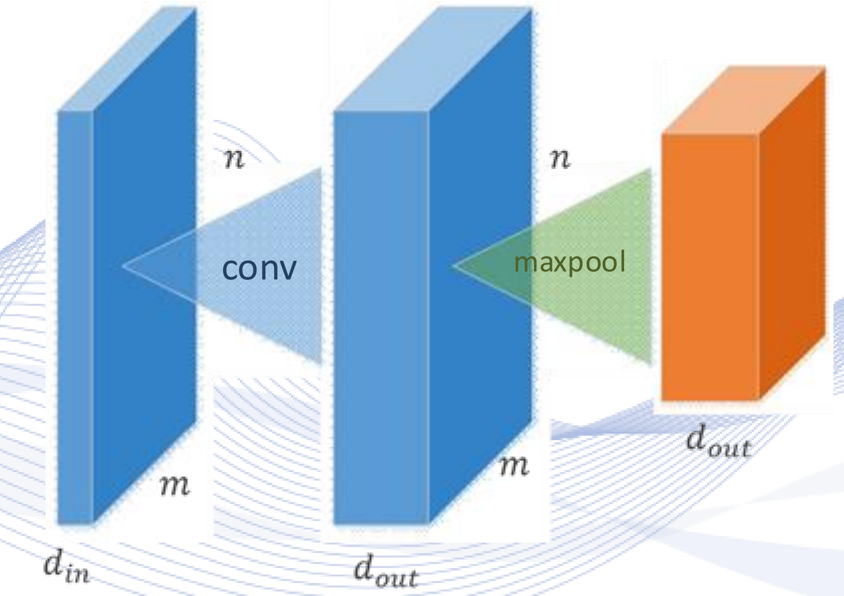
Deep semantic image segmentation architectures

Composed of multiple convolution layers.

Convolution Layer:

• Performs feature extraction using convolution operations.

• Often is followed by a **max-pooling** step to reduce spatial dimensions and retain important features.

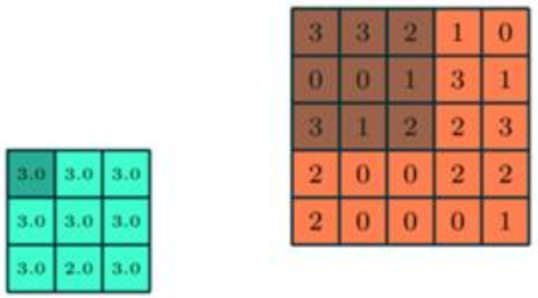


Deep Semantic Image segmentation

Max pooling keeps the strongest activation in a $n \times m$ region of an activation map.

- Edges between high and low activations could be lost.
- Downsampling is preferred to be done in max pooling layers and not in convolutional layers.

- No formal justification for the benefits of keeping the strongest activation.



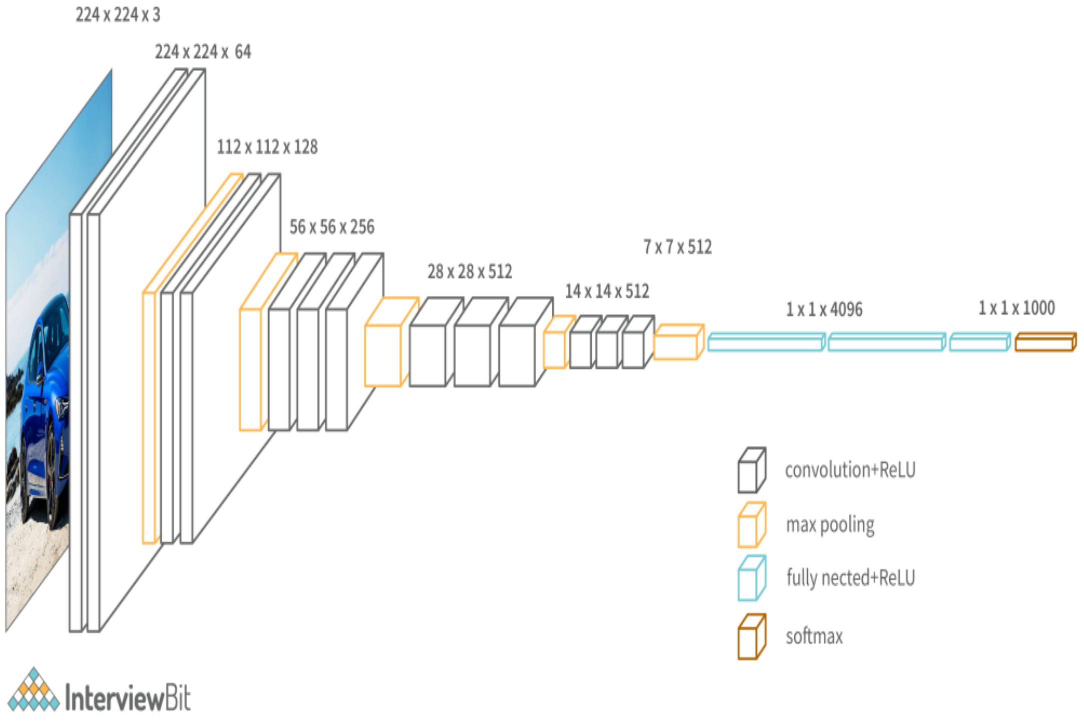
Deep Semantic Image segmentation

Image Classification

- The final feature map is **flattened** into a 1D vector.

Fully Connected Layers:

- Reduce dimensionality to match the number of classes in the dataset.
- Perform the final classification by mapping features to class probabilities.



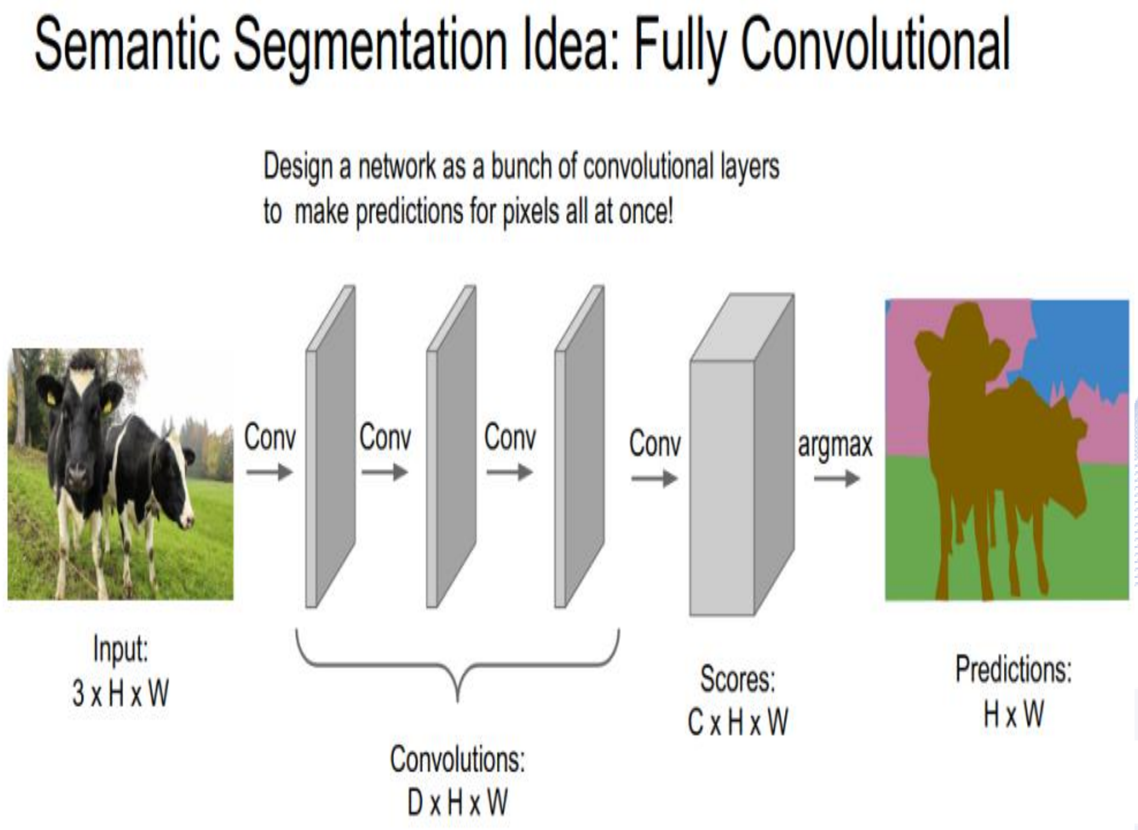
[BIT2024]

Deep Semantic Image segmentation

Semantic Segmentation

In contrast to Image classification, in segmentation the final feature map has dimensions **C × H × W**, where:

- **C**: Number of classes in the dataset.
- **H, W**: Height and width of the image.

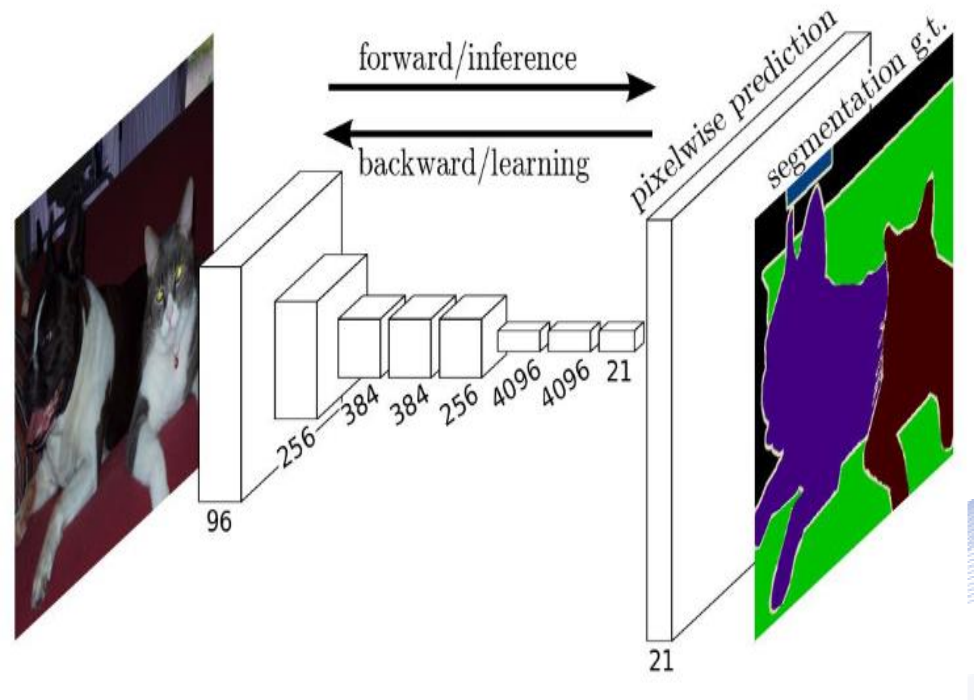


[SUP2024]

Deep Semantic Image segmentation

Semantic Segmentation

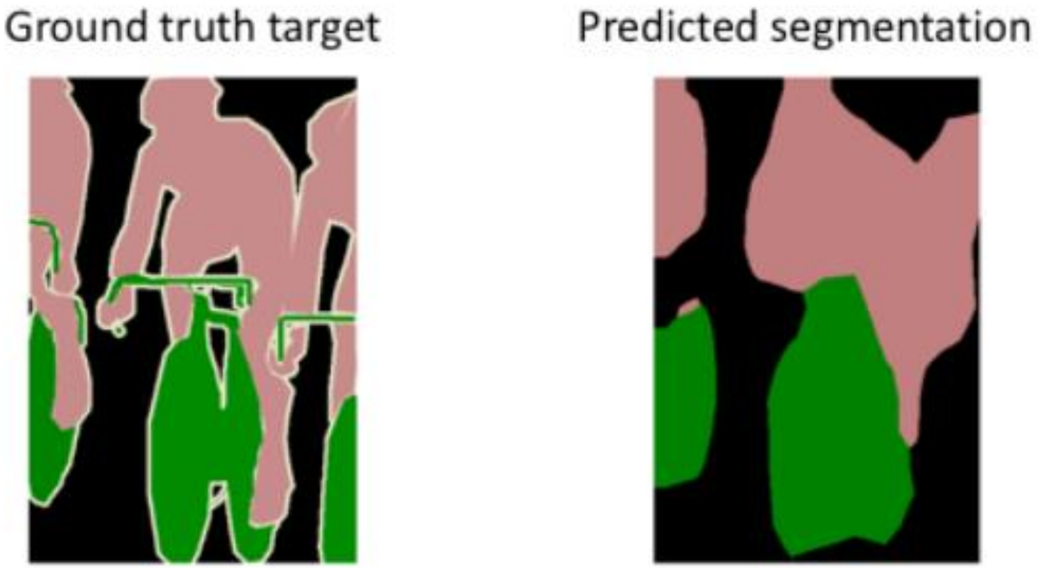
- Fully convolutional network for semantic segmentation.
- Usually, the final feature map is upsampled to match the resolution of the input image.



End-to-end CNN training for semantic image segmentation [LON2015].

Deep Semantic Image segmentation

- However, as the model radically reduces the resolution of the input image, it fails to produce fine-grained segmentations.



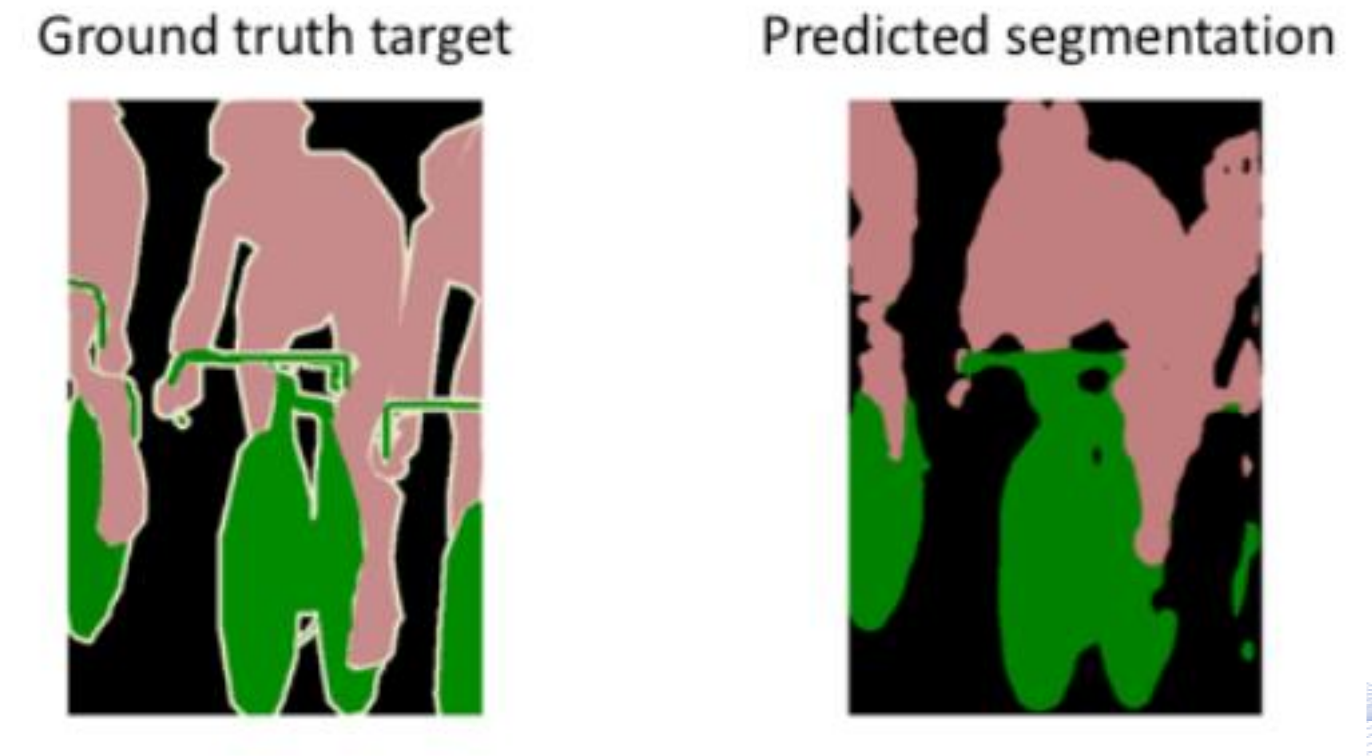
Coarse image segmentation [LON2015].

Deep Semantic Image segmentation

- To address this problem, ***skip network connections*** are added in fully convolutional network that combine the final prediction layer with previous fine-grained layers.

- Combining fine layers and coarse layers allows the model to make local predictions that respect global structure.

Deep Semantic Image segmentation



Improved segmentation results with skip connections [LON2015].

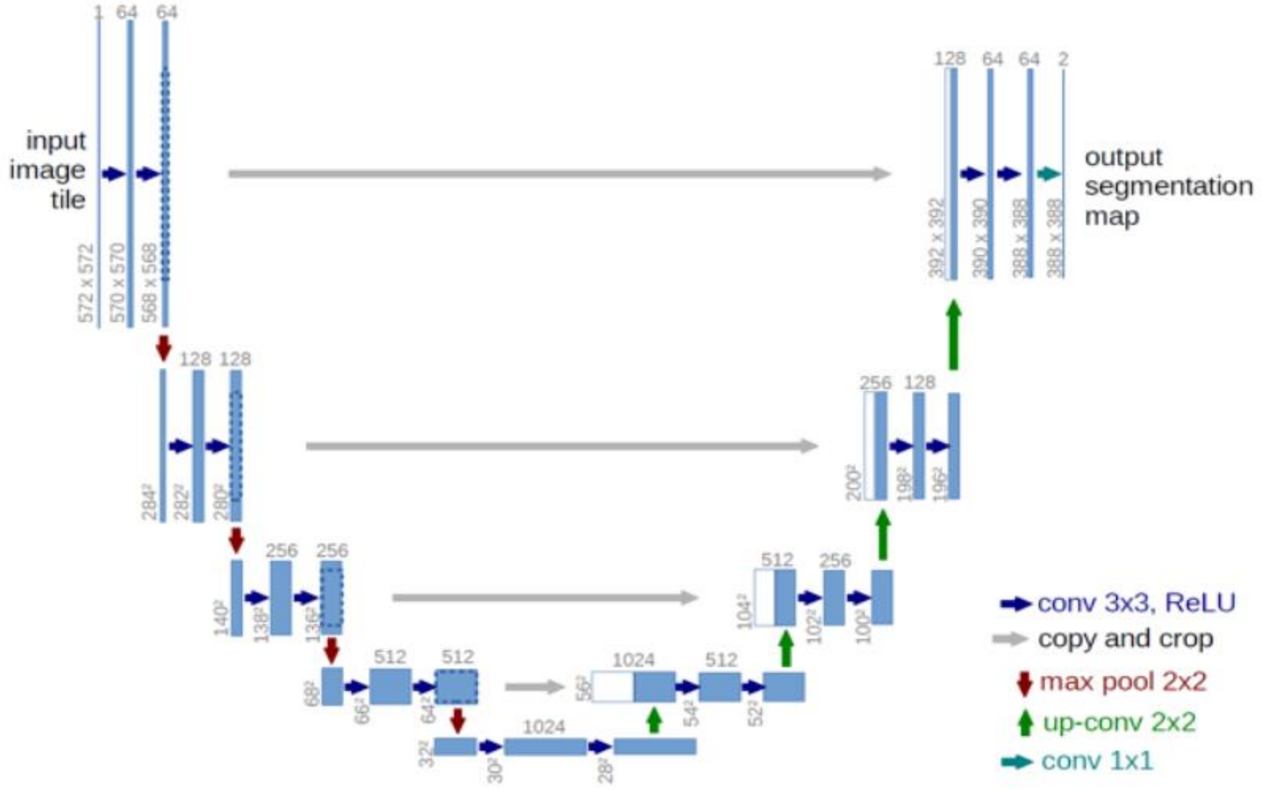
Deep Semantic Image segmentation

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.

Deep Semantic Image segmentation

U-Net architecture



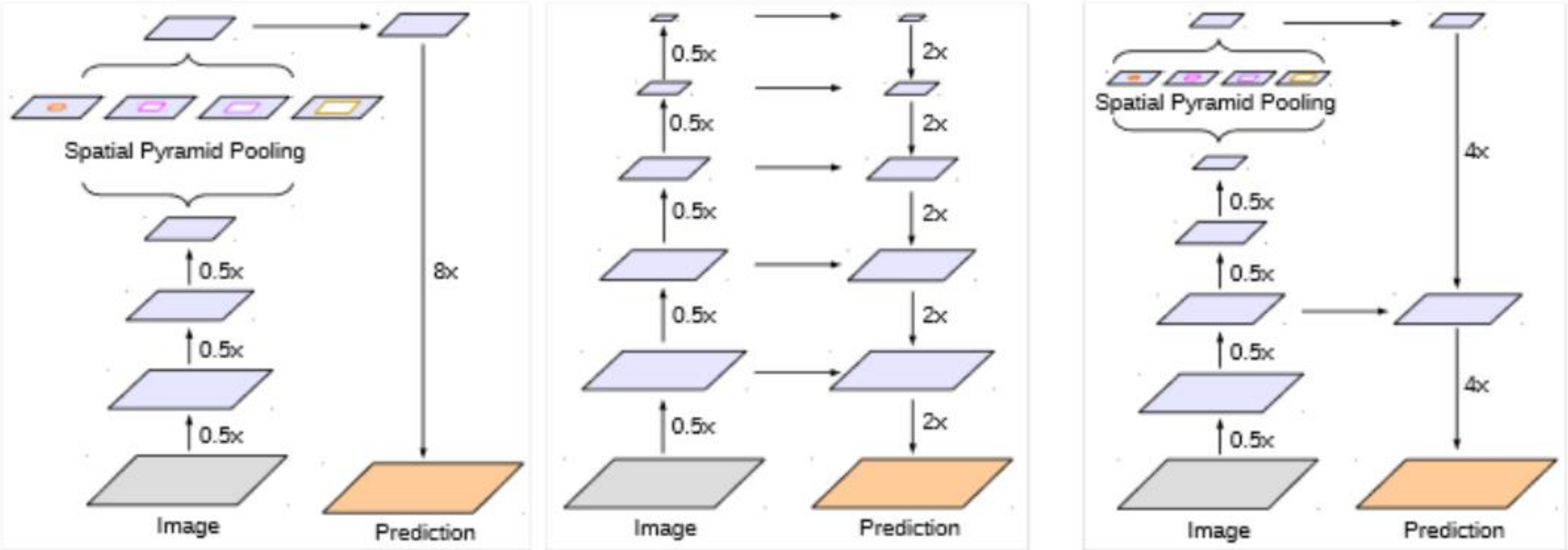
U-Net network architecture [RON 2015]

Deep Semantic Image segmentation

Spatial Pyramid Pooling

- Semantic image segmentation performance was also increased by combining the advantages of a **Spatial Pyramid Pooling (SPP)** [ZHA2017] module and the encoder-decoder architecture.
- SPP module can encode multi-scale contextual information, by probing the incoming features with filters or pooling operations at multiple rates and multiple effective fields-of-view.

Deep Semantic Image segmentation



Spatial Pyramid Pooling. Encoder-Decoder. Combined approach [CHE2018].

Deep Semantic Image segmentation

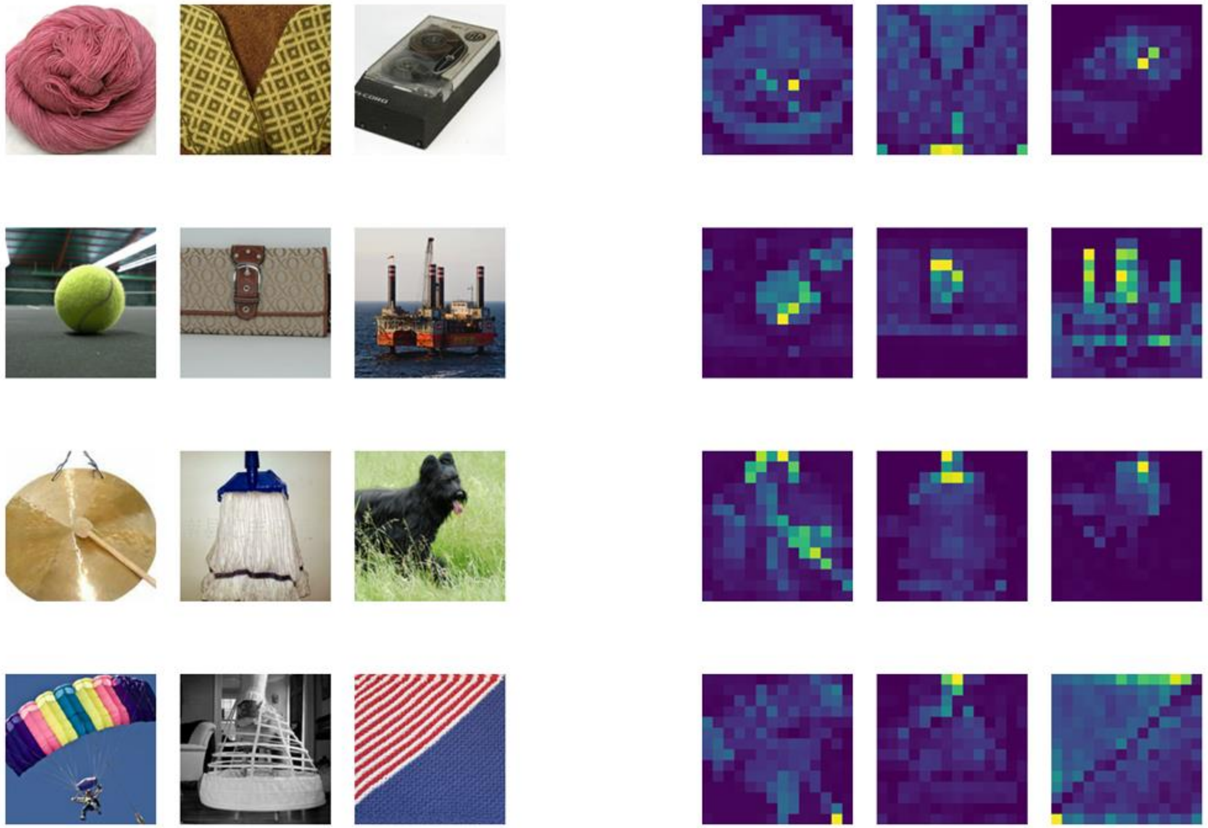
Vision Transformer (ViT) [DOS2020].

- 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

Deep Semantic Image segmentation

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.



Deep Semantic Image segmentation

Semantic Segmentation loss functions

Categorical cross entropy: $$L_{cce} = \sum_{i=1}^{h} \sum_{j=1}^{w} \sum_{k=1}^{c} y_{i,j,k} \log(p_{i,j,k})$$

- h, w are the spatial dimensions of the feature map.
- c is the number of classes.
- $y_{i,j,k}$ is the one-hot encoded ground truth label for the k-th class at position (i, j).
- $p_{i,j,k}$ is the predicted probability for the k-th class at position (i, j).

Deep Semantic Image segmentation

Semantic Segmentation loss functions

Dice Loss : Focuses on maximizing overlap between predicted and ground truth masks, commonly used for imbalanced datasets.

IoU Loss : Optimizes the intersection over union between predicted and actual regions, improving pixel-level accuracy.

Focal Loss : Addresses class imbalance by down-weighting easy examples and focusing on hard-to-classify pixels.

Real-Time Image Segmentation

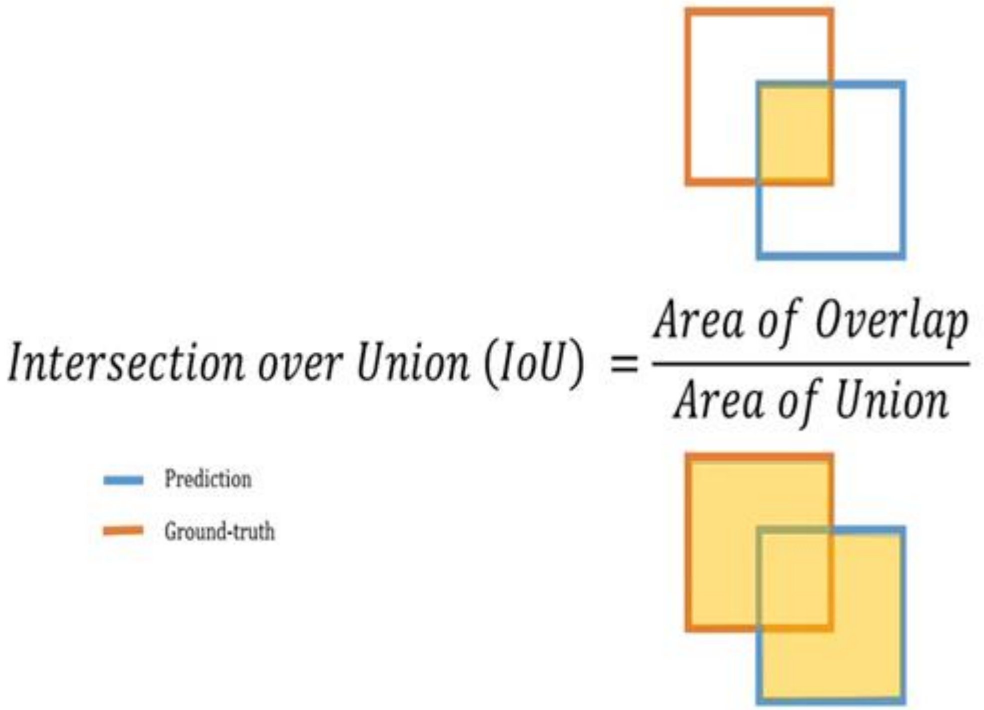
- Computer Vision
- Classical image segmentation techniques
- Deep semantic image segmentation
- **Fire Detection**
  - **Fire detection evaluation metric**
  - Fire detection localization loss
- Fire Segmentation

Fire Detection Evaluation metric

Intersection Over Union (IoU)

The overlapping area between a predicted bounding box (P) and a ground truth bounding box (G) is measured using the Intersection over Union (IoU) method, which is formulated as follows:

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



Fire Detection Evaluation metric

mAP (Object Detection):

- Combines precision and recall to evaluate detection accuracy.
- Uses Intersection over Union (IoU) to match predicted and ground-truth bounding boxes.
- Calculates the average precision for each class and averages across all classes.
- Rewards precise alignment and penalizes missing or incorrect predictions.

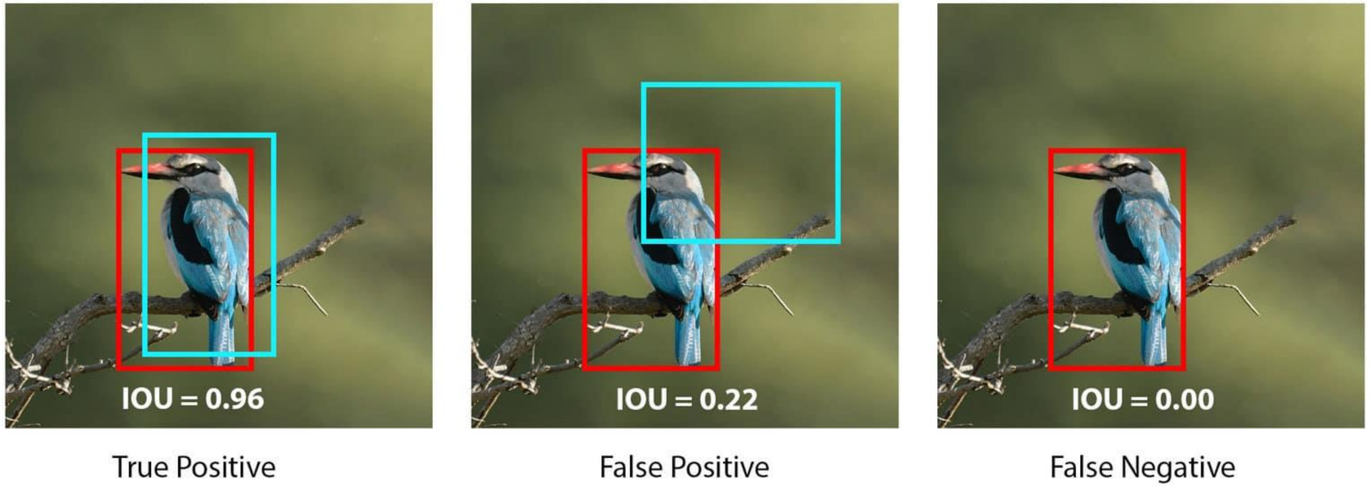
mIoU (Segmentation):

- Averages IoU across all pixels and classes to assess segmentation quality.

Fire Detection Evaluation metric

Predicted Bound box evaluation :

TP : IOU > 0.5
FP : IOU < 0.5



Bounding box evaluation [KUK2023]

Artificial Intelligence & Information Analysis Lab

Fire Detection Evaluation metric

3 TP – 0 FP

1TP – 2 FP

0 TP – 3 FP

BASED ON IOU

Raw image

Possible ground-truth annotation strategies

Predictions

Fire Detection Evaluation metric

Challenges with mAP for Fire Detection: Unlike most objects, fires consist of “children” objects (flames) that belong to the same class as the “parent” object (fire), making it uncertain how many bounding boxes are needed for accurate representation.

Limitations of mAP: Inconsistent annotation styles for fire objects can misalign with predicted bounding boxes, leading to mAP scores that do not accurately reflect model performance.

Fire Detection Evaluation metric

- **Proposed Solution – ImAP** [TZI2023] : The Image-level mean Average Precision (ImAP) metric evaluates fire detection models based on their ability to predict bounding boxes for the entire image rather than individual boxes.

- **Experiments and Results**: ImAP demonstrates greater suitability than mAP for evaluating object detectors in fire detection tasks, addressing the unique properties of fire entities.

Fire Detection Evaluation metric

ImAP utilize Image Level Intersection Over Union (**ImIOU**) instead of **IOU** in order to evaluate fire detection in the entire image

ImIOU : Intersection over Union between all predictions and all ground truth bounding boxes of the same image

Given then bounding box ground truths $\mathcal{G} = \{G_i\}_{i=1,\ldots,N}$ of an image and their corresponding predictions $\mathcal{P} = \{P_i\}_{i=1,\ldots,M}$ then ImIoU is formulated as :

$$ImIoU\ (\mathcal{P},\mathcal{G}) = \frac{\left|\left(\bigcup_{i=1}^{|\mathcal{P}|} P_i\right) \cap \left(\bigcup_{i=1}^{|\mathcal{G}|} G_i\right)\right|}{\left|\left(\bigcup_{i=1}^{|\mathcal{P}|} P_i\right) \cup \left(\bigcup_{i=1}^{|\mathcal{G}|} G_i\right)\right|}$$

Artificial Intelligence & Information Analysis Lab

Fire Detection Evaluation metric

BASED ON IOU 3 TP – 0 FP 1TP – 2 FP 0 TP – 3 FP

BASED ON IMIOU IMIOU=0.78 -> TP IMIOU=0.6 -> TP IMIOU=0.51 -> TP

Artificial Intelligence & Information Analysis Lab

Real-Time Image Segmentation

- Computer Vision
- Classical image segmentation techniques
- Deep semantic image segmentation
- **Fire Detection**
  - Fire detection evaluation metric
  - **Fire detection localization loss**
- Fire Segmentation

Fire Detection Localization Loss

Regression Losses for the localization task of the object detection:

- L1
- IOU based

The state of the art object detection model RTDETR [ZHA2024] combines the L_1 loss with L_{IoU} to improve the detection of object of interest.

$$L_{loc} = \lambda_1 \cdot L_1 + \lambda_{IoU} \cdot L_{IoU}$$

Fire Detection Localization Loss

For an image with N bounding box ground truths described by set $\mathcal{G} = \{G_i\}_{i=1,\ldots,N} = \{\{G_{i,x}, G_{i,y}, G_{i,w}, G_{i,h}\}\}_{i=1,\ldots,N}$ and their matched predictions by $\mathcal{P} = \{P_i\}_{i=1,\ldots,N} = \{\{P_{i,x}, P_{i,y}, P_{i,w}, P_{i,h}\}\}_{i=1,\ldots,N}$ the L_1 and L_{IoU} are formulated as :

$$L_1(\mathcal{P},\mathcal{G}) = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j\in\{x,y,w,h\}} \left|P_{i,j} - G_{i,j}\right|\right)$$

$$L_{IoU}(\mathcal{P},\mathcal{G}) = \frac{1}{N}\sum_{i=1}^{N}\left(1 - IoU(P_i, G_i)\right)$$

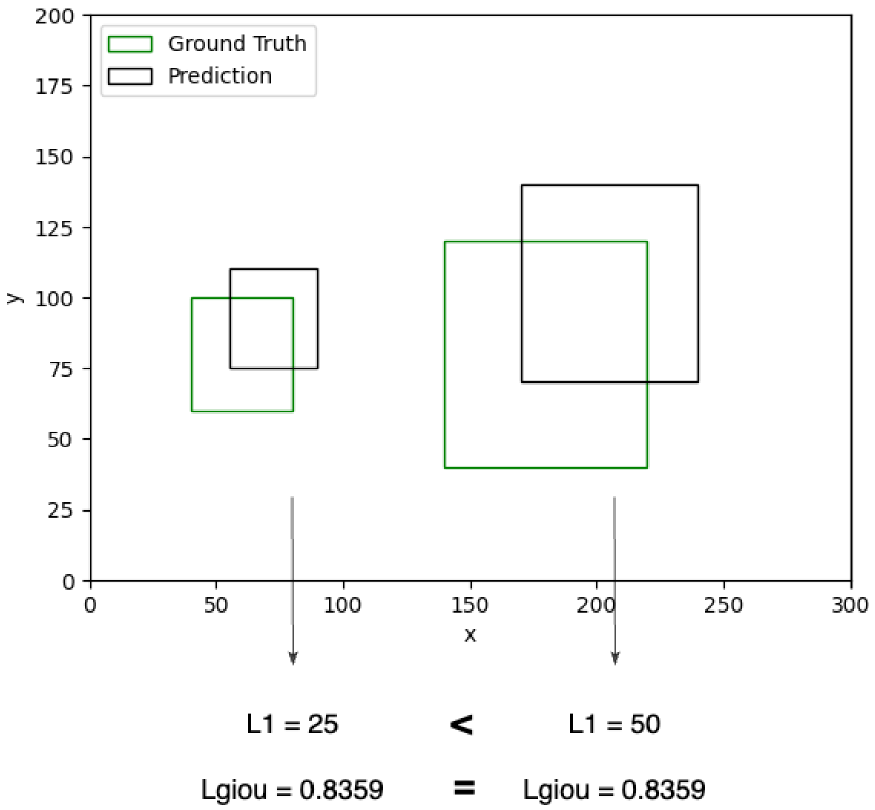
where N the number of bounding boxes in the image

Fire Detection Localization Loss

In fire detection there are many Scenarios where in the same image can appear small and large flames.

In this case the larger prediction boxes have larger error with their corresponding due to the l1 loss

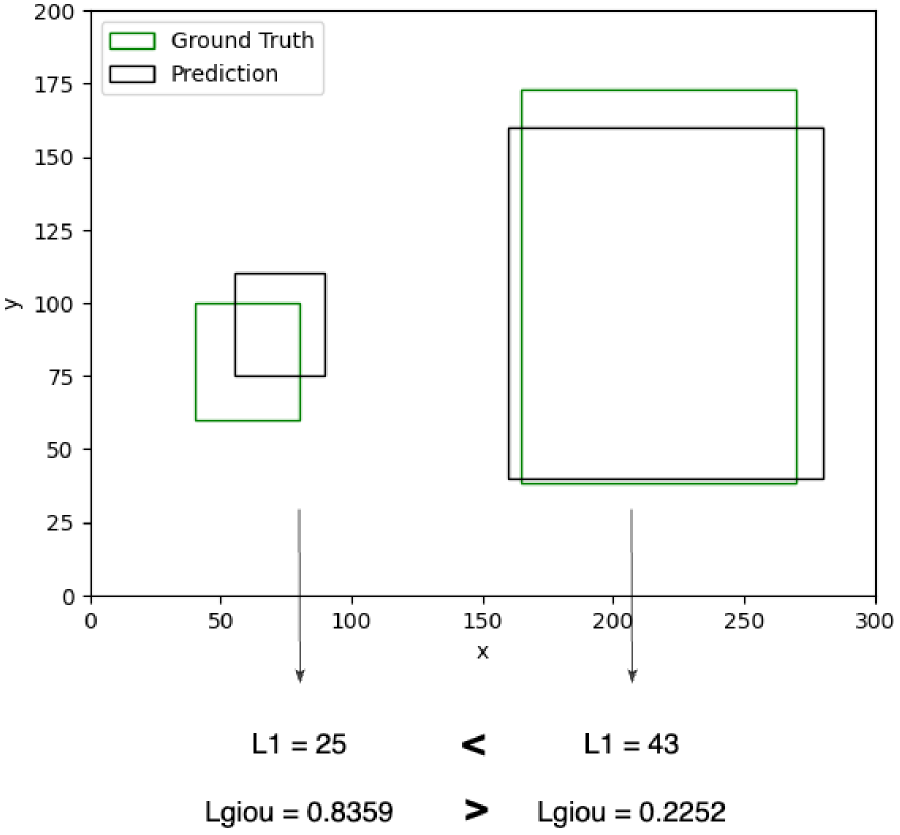
IoU based losses are invariant of the bounding box sizes



Artificial Intelligence & Information Analysis Lab

Fire Detection Localization Loss

In many cases, there may be disagreements between the L1 and IoU losses, which can affect training, as the two losses may not share the same local minimums.



Fire Detection Localization Loss

Solution : adding a weighting mechanism on the L1 loss based on the ground-truth bounding box size.

Size balanced L1 loss L_{SB}:

$$L_{SB}(\mathcal{P},\mathcal{G}) = \sum_{i=1}^{N} \boldsymbol{W_i} \left(\sum_{j\in\{x,y,w,h\}} \left|P_{i,j} - G_{i,j}\right| \right)$$

Experiments on fire detection datasets demonstrate +2% improvment over mAP and lmAP

Artificial Intelligence & Information Analysis Lab

Real-Time Image Segmentation

- Computer Vision
- Classical image segmentation techniques
- Deep semantic image segmentation
- Fire detection
- **Fire segmentation**
  - RGB/IR Fire segmentation
  - Unsupervised fire segmentation

Fire Segmentation

In this approach 3 deep neural network based semantic segmentation architectures were trained on the flame dataset :

- BiSeNet (backbones: ResNet18, ResNet101) [CYO2018]
- I2I-CNN (backbones: ResNet18) [PAP2021]
- PIDNet (backbone: ResNet18) [JXU2023]

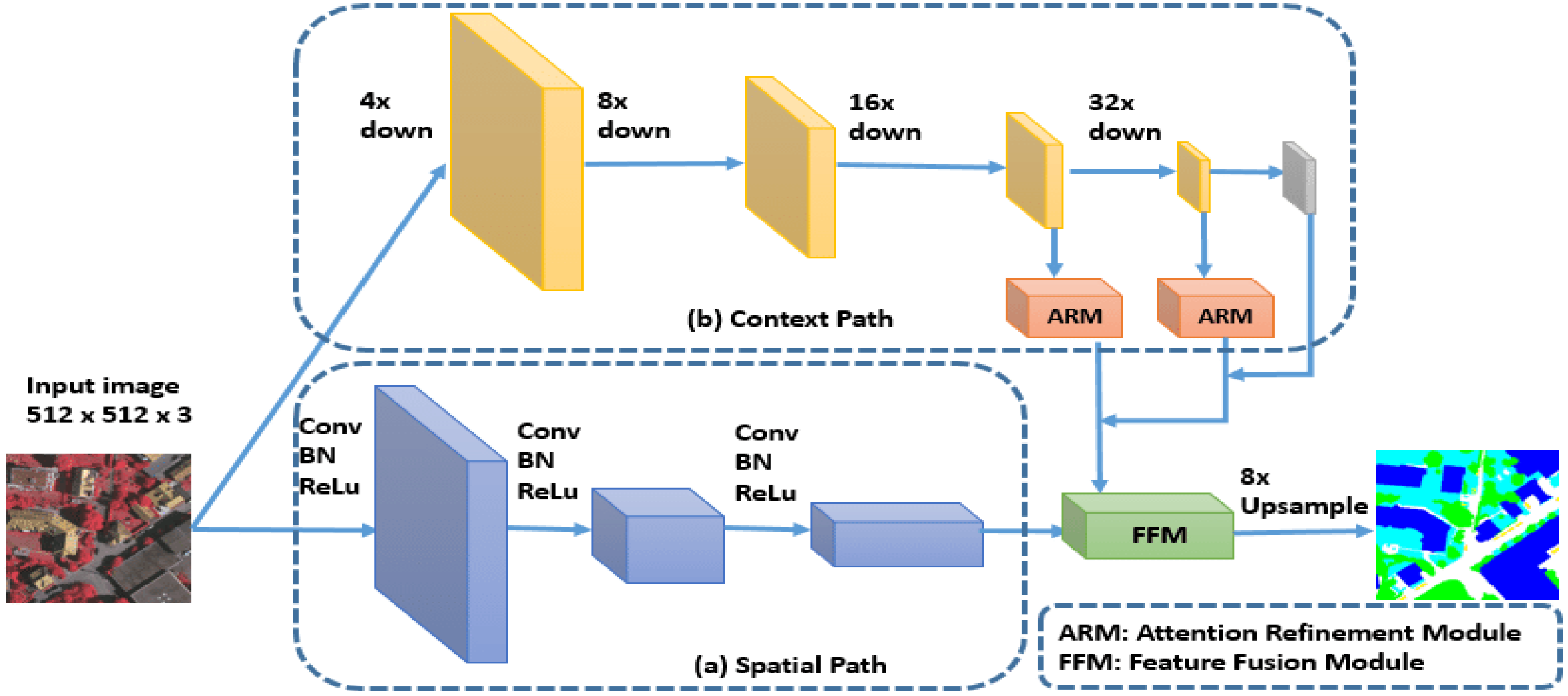
flame dataset [SHA2021]

Fire 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.

Fire Segmentation



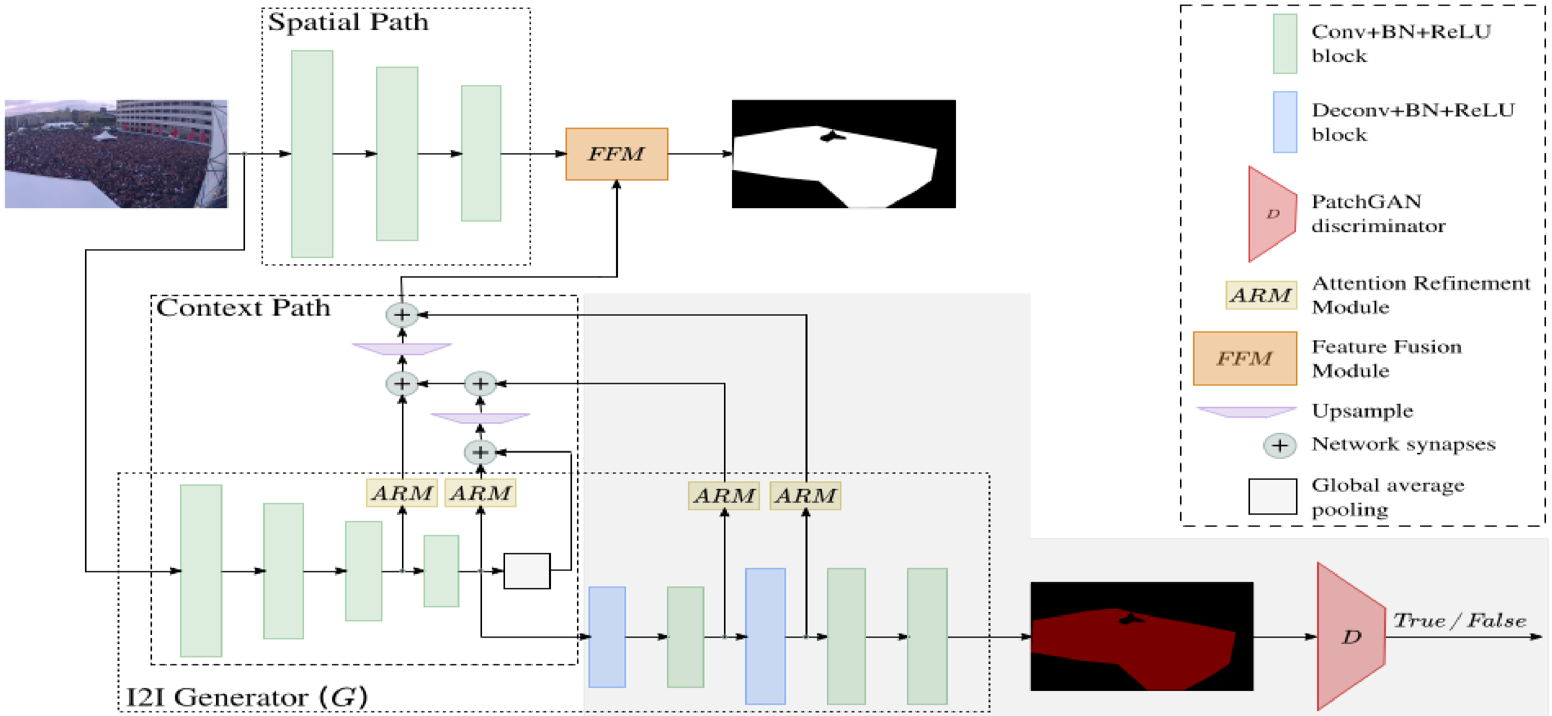
BiSeNet architecture [CYO2018]

Fire 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.

Fire Segmentation



I2I-CNN architecture [PAP2021]

Fire Segmentation

PIDNet architecture

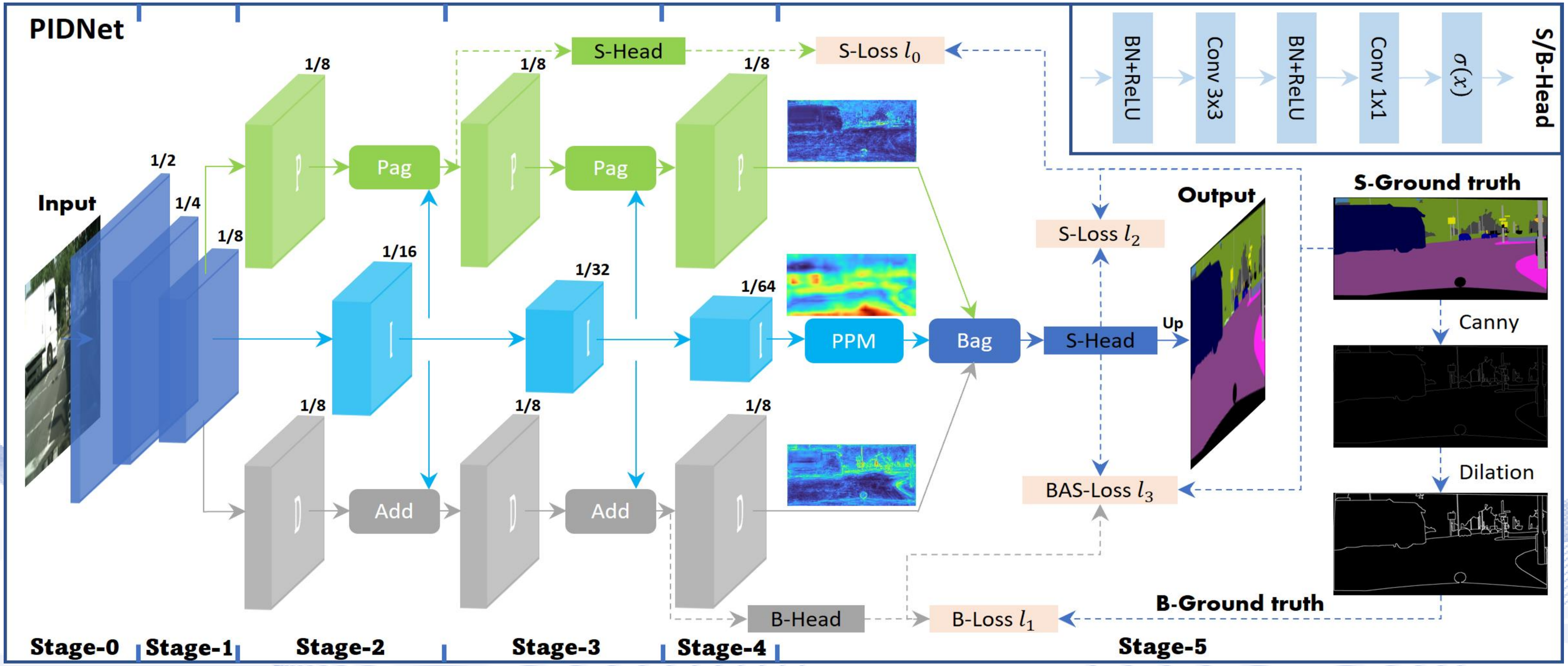
• **Triple-Branch Design:** Uses three branches—Proportional (P), integral (I), and derivative (D)—to balance accuracy and efficiency.

• **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.

Fire Segmentation



PIDNet architecture [JXU2023]

Fire Segmentation

The DNNs where studied with respect to the given input. The input was fed to the networks in the 3 following forms :

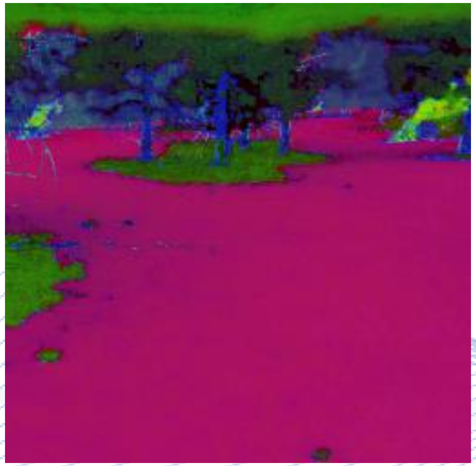
- RGB (3 channels)
- RGB+HSV (6 channels)
- RGBS (4 channels)

Where S in RGBS input image represents the processed saturation channel of HSV image transform and is used to suggest potential fire regions. This mask is then concatenated with the RGB image to form a new 4-channel input.

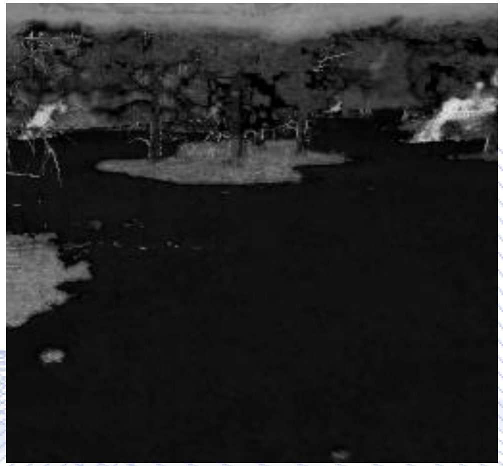
Fire Segmentation

Process of creating the S channel (visualization)

RGB input image



HSV transformation of RGB image



Saturation channel (S)

Thresholding of channel S

Fire Segmentation

BiSeNet, I2I-CNN and PID-Net were evaluated using mIoU and novel fire region segmentation metrics based on :

- *The* number N of fire instances (D_N)
- *The* average fire region area in pixels (D_A)
- The spatial dispersion of fire region instances (D_S)

These metrics extract meaningful information about the extend of a forest fire and target the explainability to the end-user

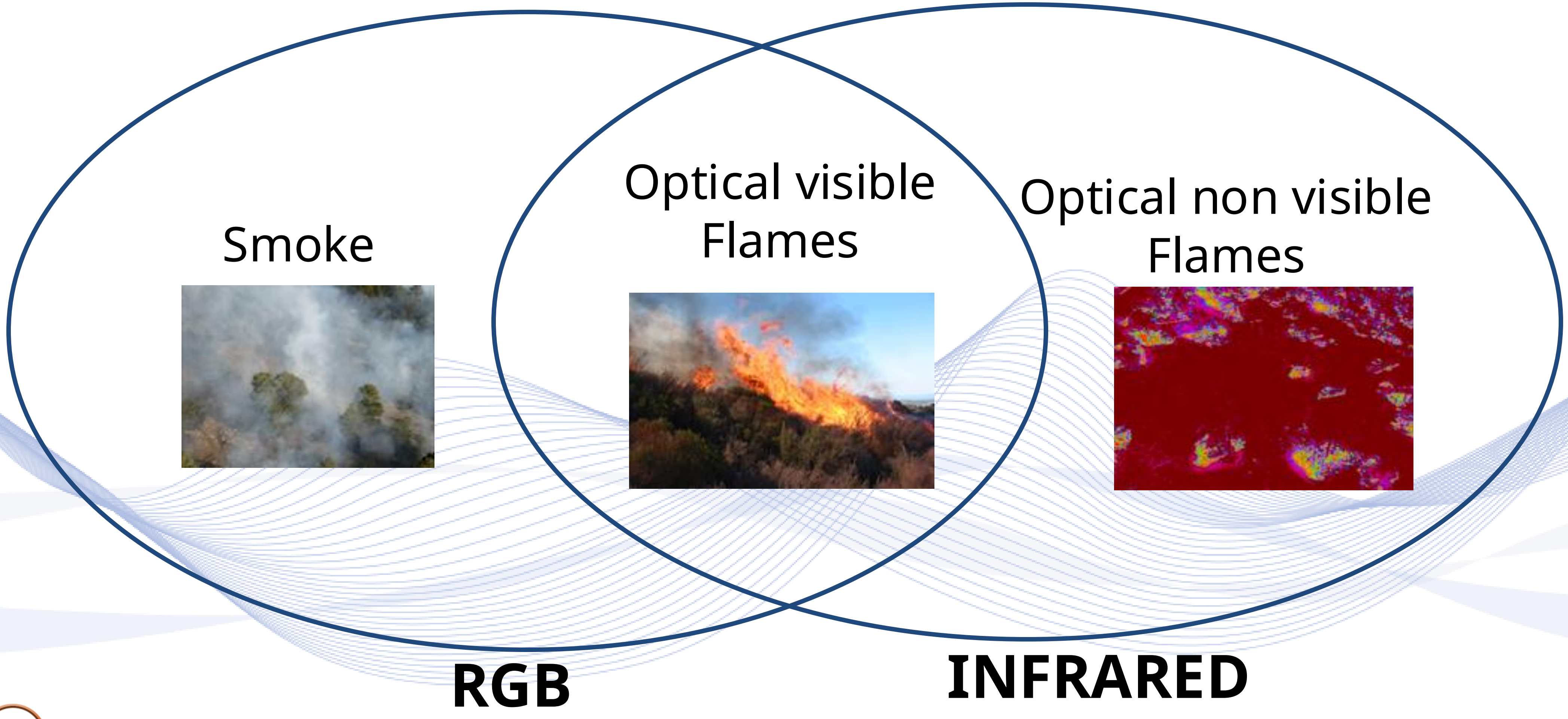
The experiments on the flame dataset demonstrate that the PIDNet with RGB+S as input achieve the best mIoU among all the other configurations

Real-Time Image Segmentation

- Computer Vision
- Classical image segmentation techniques
- Deep semantic image segmentation
- Fire detection
- Fire segmentation
  - **RGB/IR Fire segmentation**
  - Unsupervised fire segmentation

RGB/IR Fire Segmentation

A Venn Diagram of RGB and IR Capabilities



RGB/IR Fire Segmentation

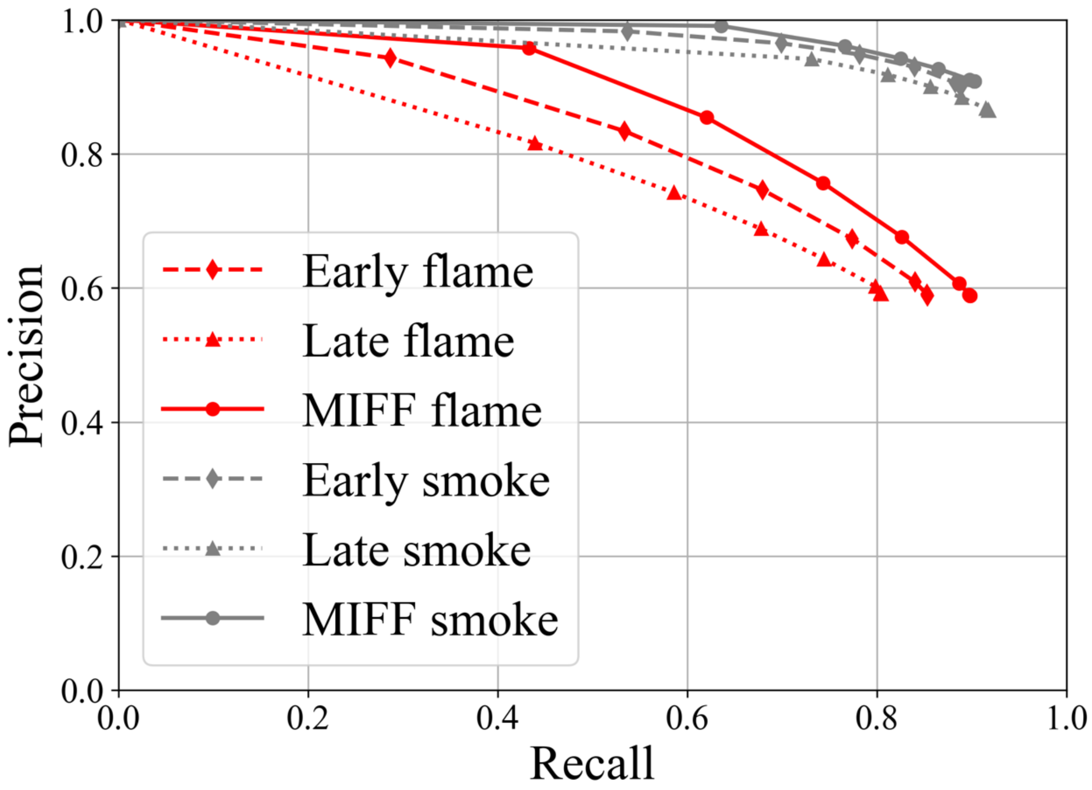
Combining IR and RGB:

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.

RGB/IR Fire Segmentation



RGB/IR Fire Segmentation

Real-Time Image Segmentation

- Computer Vision
- Classical image segmentation techniques
- Deep semantic image segmentation
- Fire detection
- Fire segmentation
  - RGB/IR Fire segmentation
  - **Unsupervised fire segmentation**

Unsupervised Fire Segmentation

The natural disaster management field requires an enormous amount of labeled data to train deep learning models to detect objects of interest. Annotating datasets is a time-consuming and expensive task.

Raw images and the corresponding labels

Unsupervised Fire Segmentation

Unsupervised semantic segmentation 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

Unsupervised Fire Segmentation

Raw Images

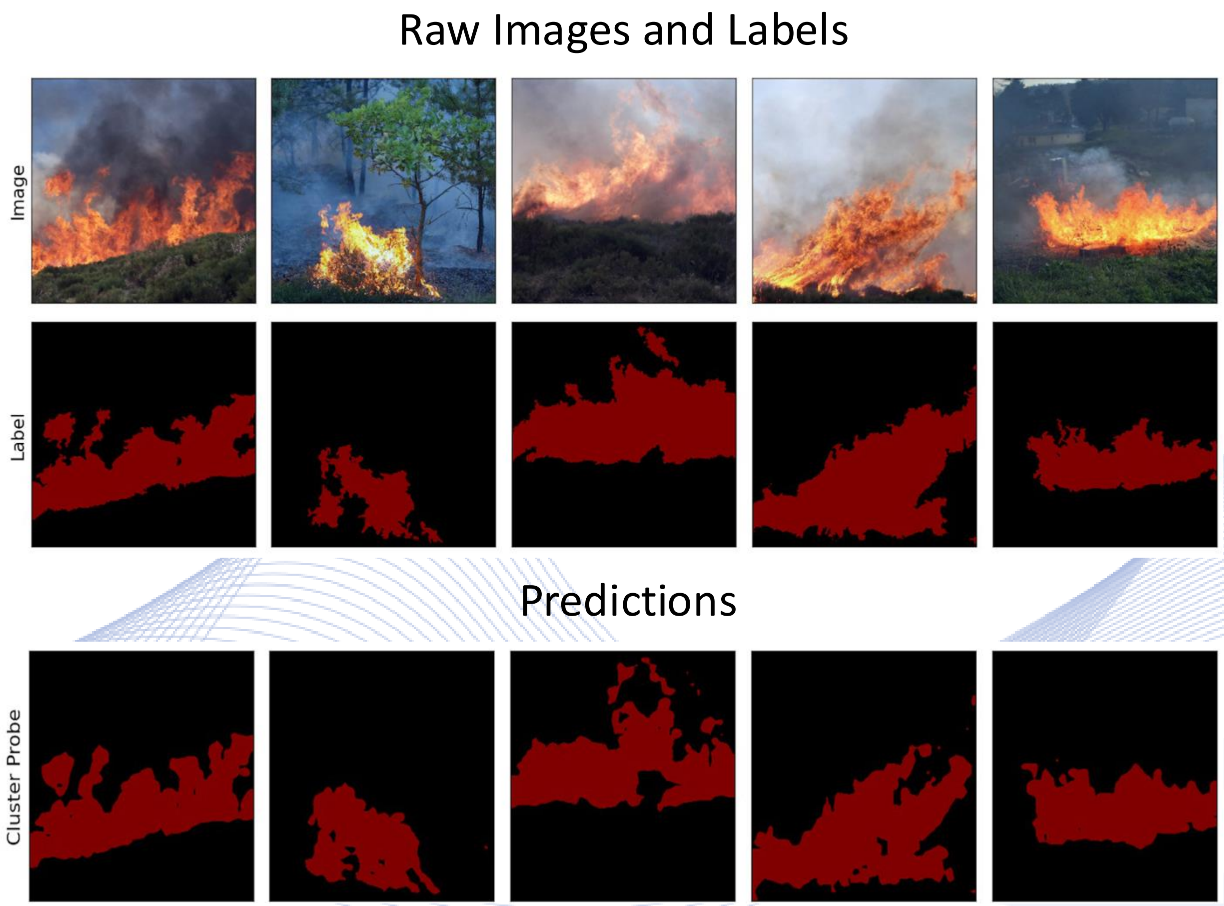
We select a single image from the dataset and specify only one point where our object of interest is located.

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

Unsupervised Fire Segmentation

- Unsupervised performance : 40 % mIoU
- Our performance : 80 % mIoU
- Our approach achieves a 40% 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



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, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, “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. Kosećka, “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. https://www.superannotate.com/blog/guide-to-semantic-segmentation#:~:text=Semantic%20segmentation%20is%20simply%20the,Image%20Source

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/

Q & A

Thank you very much for your attention!

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