Deep Semantic Image Segmentation summary

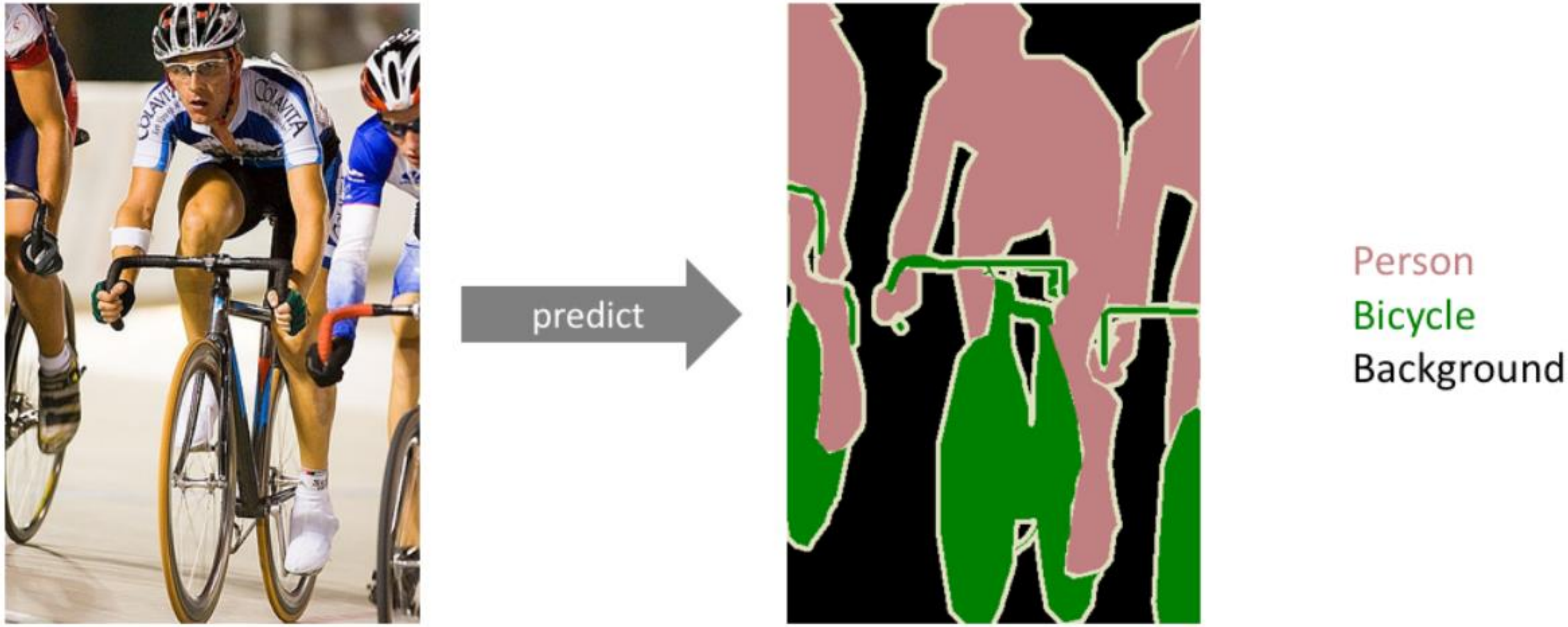
C. Papaioannidis, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 3.1.1

Artificial Intelligence &
Information Analysis Lab

Deep semantic image segmentation

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

Deep semantic image segmentation



Semantic image segmentation of a sports event [EVE2011].

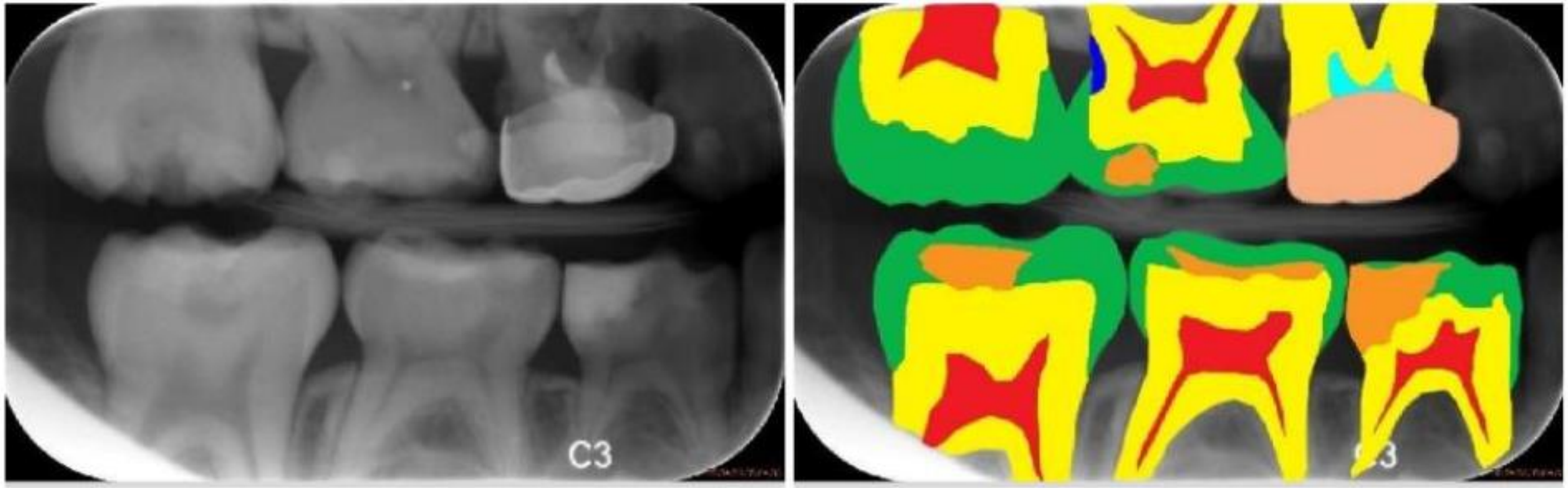
Deep semantic image segmentation

- Autonomous driving.

Semantic image segmentation for autonomous driving [COR2016].

Deep semantic image segmentation

- Medical purposes.



Semantic dental Xray segmentation [TOR2014].

Deep semantic image segmentation

- An image domain $\mathcal{X}$ must be segmented 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:

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

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

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

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

Deep semantic image segmentation

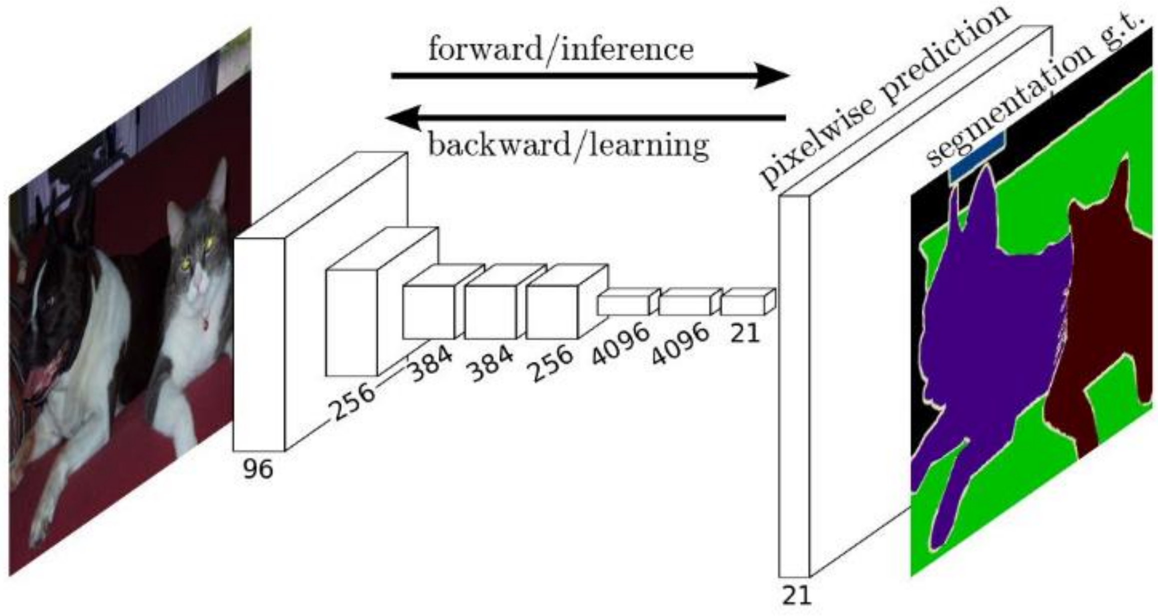
- Introduction
- **Classical image segmentation techniques.**
- Deep semantic image segmentation.
- Applications

Deep semantic image segmentation

- Recent semantic image segmentation methods classify each pixel of an input image to an object class using DNNs.
- ***Dense prediction***: DNN predictions are made at pixel level.

Deep semantic image segmentation

- Fully convolutional network for semantic segmentation.

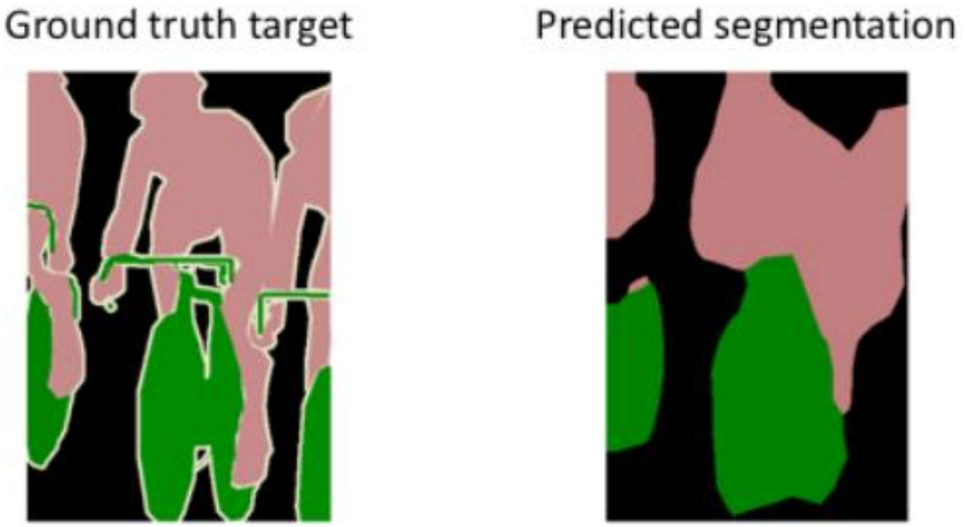


End-to-end CNN training for semantic image segmentation [LON2015].

Artificial Intelligence & Information Analysis Lab

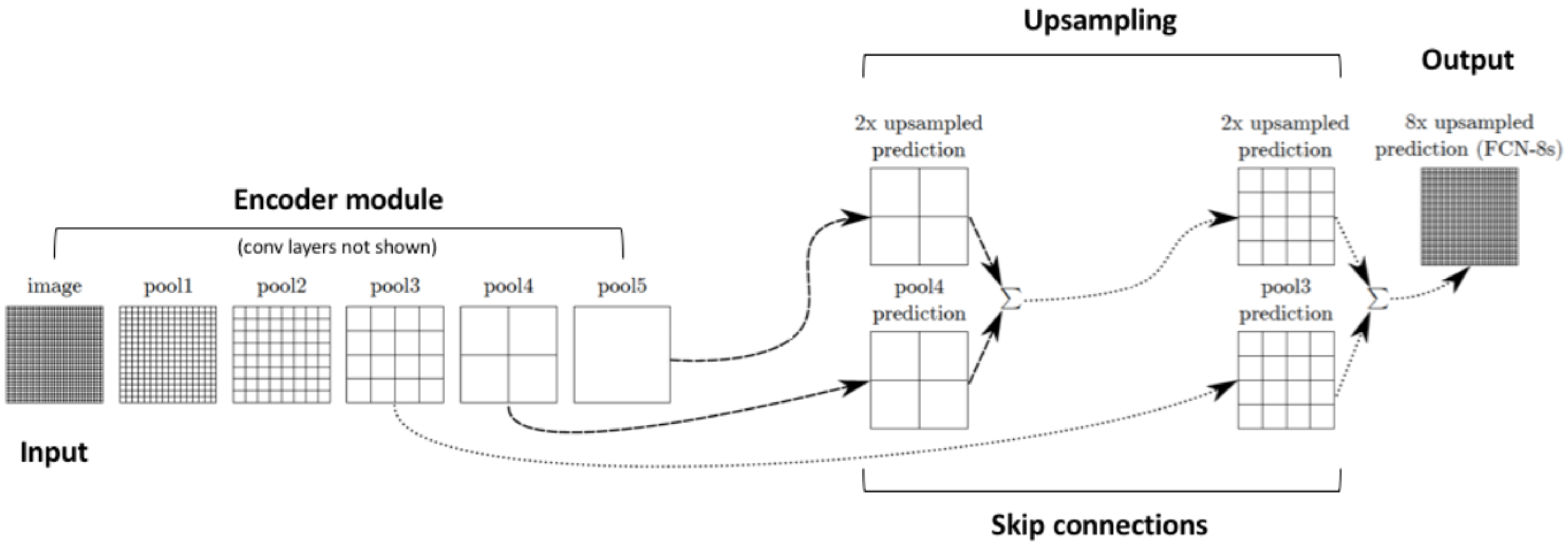
Deep semantic image segmentation

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



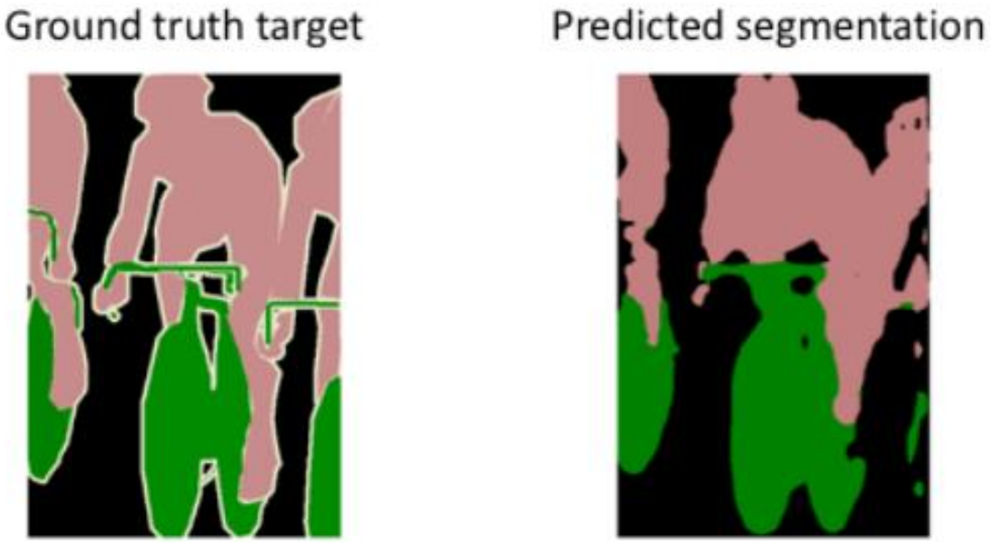
Coarse image segmentation [LON2015].

Deep semantic image segmentation



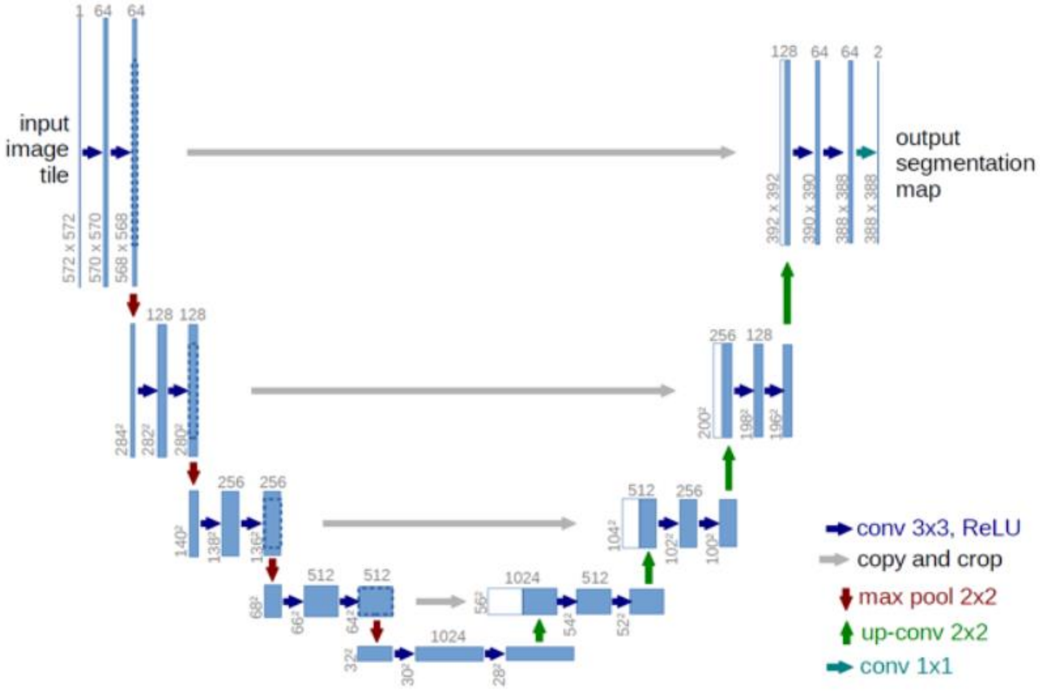
Skip CNN layer connections [LON2015].

Deep semantic image segmentation



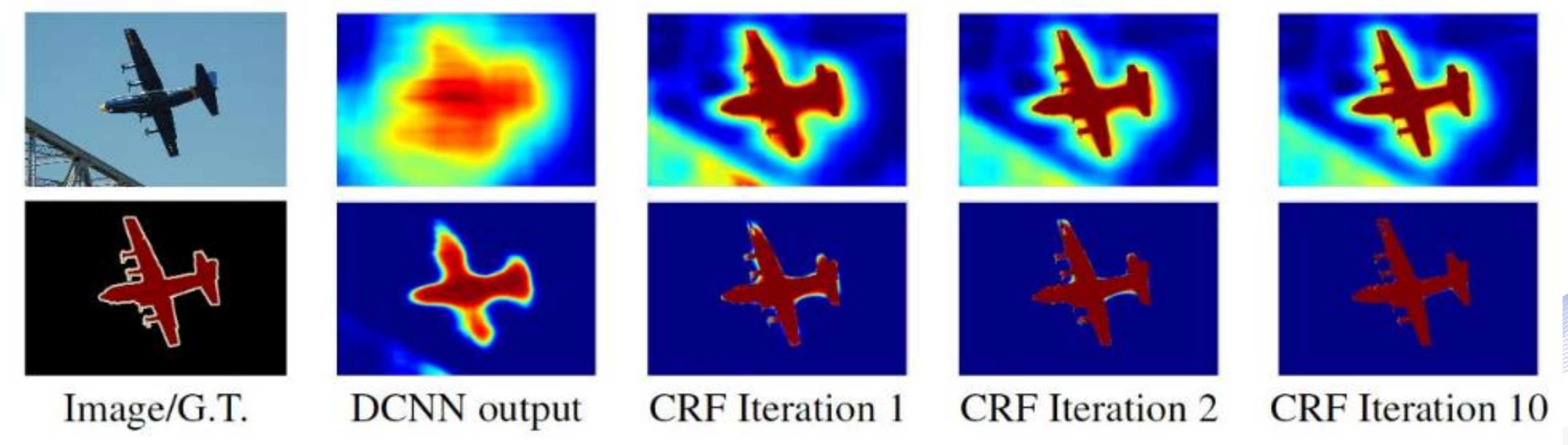
Improved segmentation results with skip connections [LON2015].

Deep semantic image segmentation



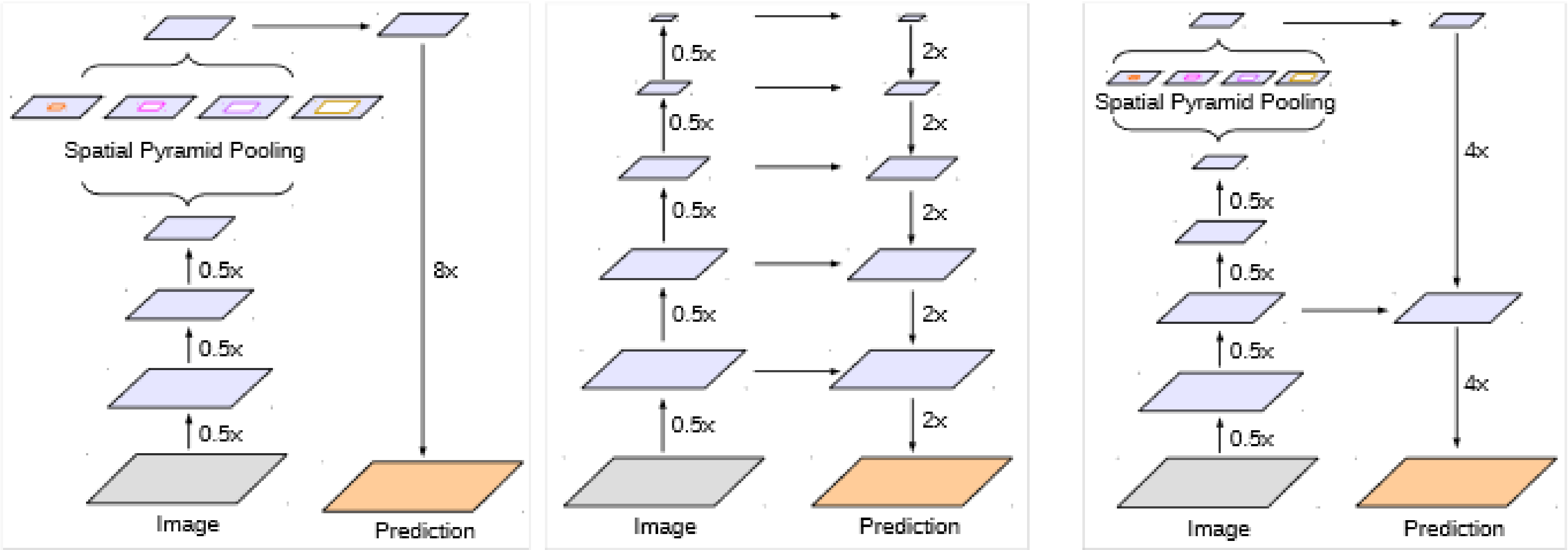
U-Net network architecture [RON 2015].

Deep semantic image segmentation



Fully convolutional networks with CRFs [CHE2017].

Deep semantic image segmentation



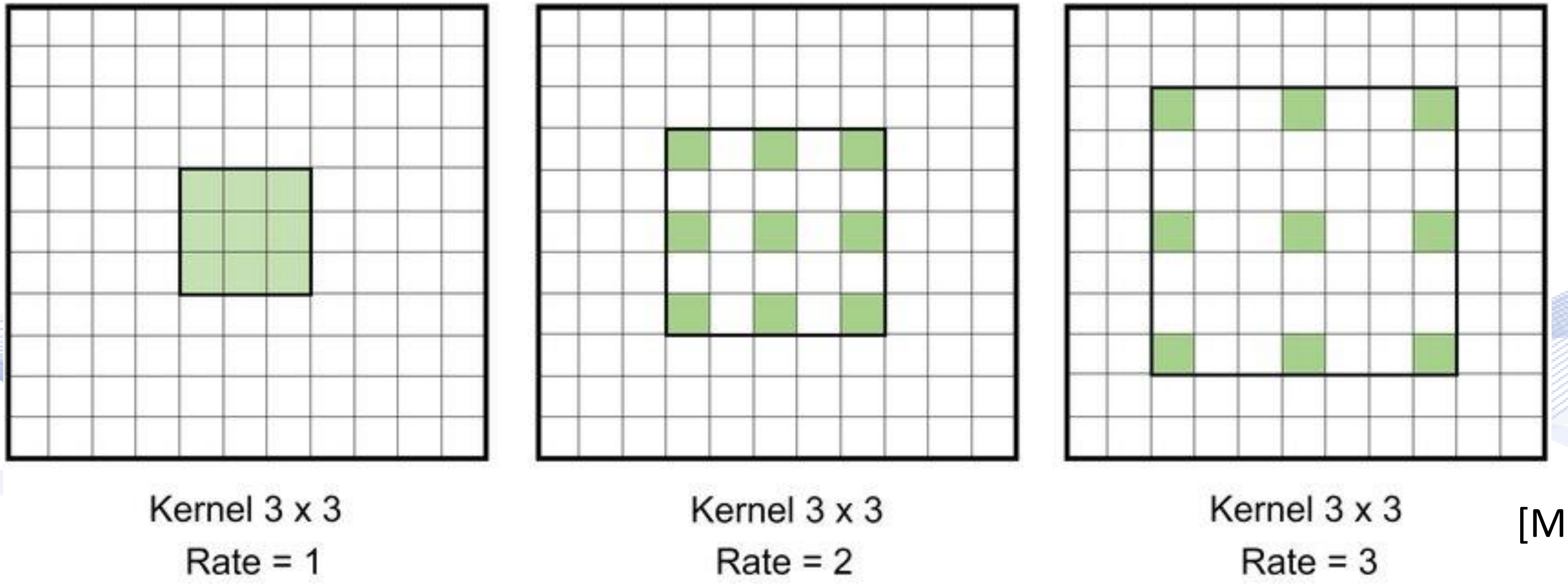
Spatial Pyramid Pooling Encoder-Decoder Combined [CHE2018]

Deep semantic image segmentation

- Encoder-decoder architectures can be very slow due to the operations required by the convolution layers.
- In this direction, **atrous** or **dilated** convolution has been introduced as a replacement to the convolution layer.
- Atrous convolution offers a wider field of view at the same computational cost.

Deep semantic image segmentation

- The **dilation rate** defines a spacing between the values in a convolving kernel.



[MOR2018]

Deep semantic image segmentation

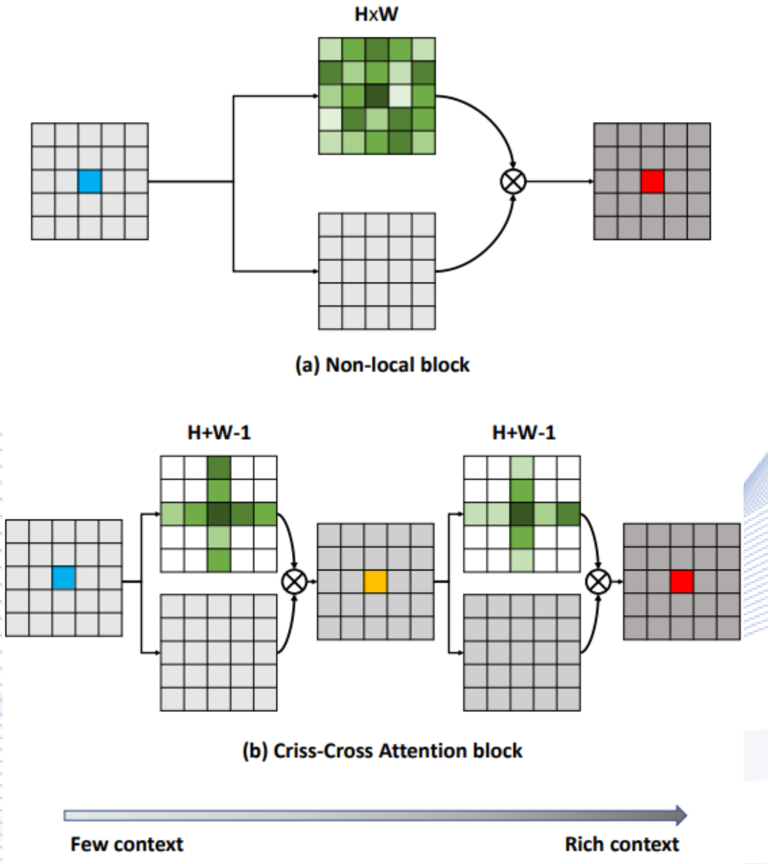
(a) ASPP (b) OCR

a) ASPP (using dilated convolution): The context of a pixel (red box) is the set of sparsely sampled pixels around it (blue and ochre boxes). b) OCR: The context is a set of pixels residing in the object it belongs to.

Deep semantic image segmentation

Learnable ***attention matrices*** can help semantic segmentation by eliminating spurious regions/noise [HUA2019] .

- ***Criss-cross attention mechanism*** is a better alternative to regular ***attention matrix*** for semantic segmentation [HUA2019] .
- It adaptively captures contextual information for each pixel on the vertical and horizontal axes.



Joint 3D Scene Geometry and Semantics Estimation

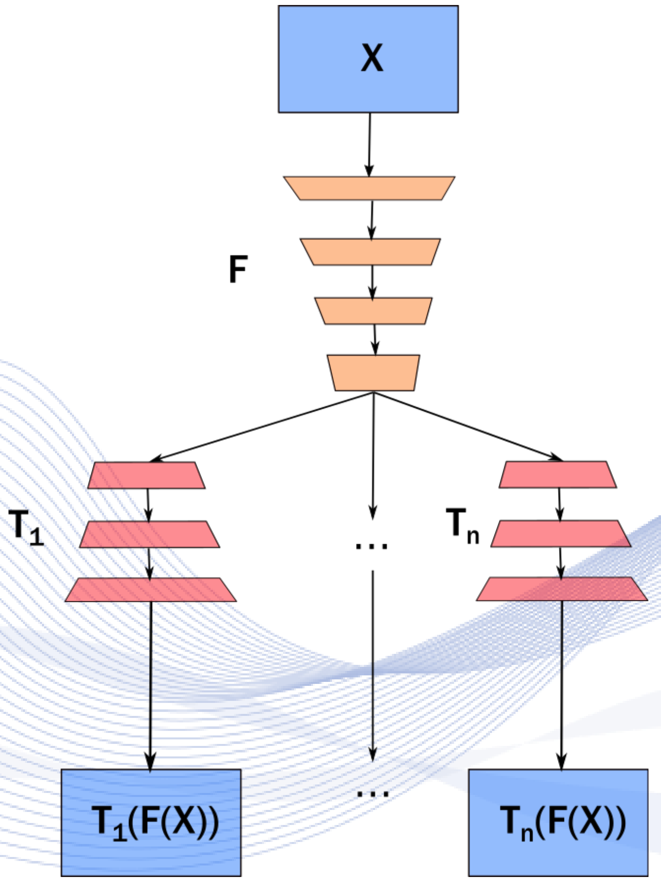
Typical multitask networks have:

- Common input $\mathbf{X}$.
- Common feature extraction operator $\mathbf{F}$.
- n concurrent task operators:

$$\mathbf{T}_1, \ldots, \mathbf{T}_n,\ n \geq 2.$$

- The **multitask network** output is the set:

$$\mathcal{T} = \{\mathbf{T}_1(\mathbf{F}(\mathbf{X})), \ldots, \mathbf{T}_n(\mathbf{F}(\mathbf{X}))\}.$$



Deep semantic image segmentation

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

Crowd detection using semantic image segmentation

- The crowd detection problem is effectively approached using semantic image segmentation.
- If only two object classes are considered (i.e., crowd, no-crowd), semantic image segmentation corresponds to crowd detection.

Crowd detection using semantic image segmentation

Crowd detection as semantic image segmentation.

Crowd detection using semantic image segmentation

- The semantic image segmentation branch is trained using the following loss function:

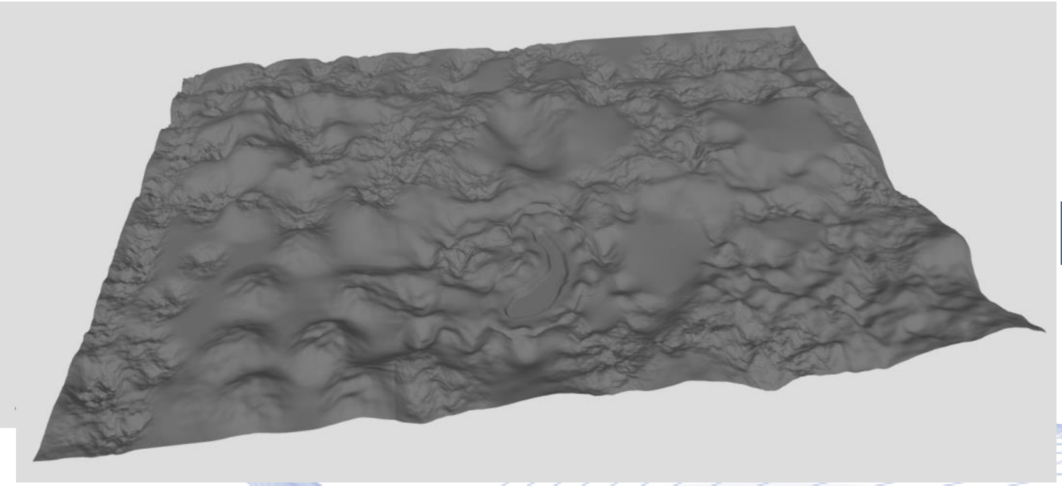
$$J_s = J_p + \alpha \sum_{i=2}^{3} J_{a_i},$$

- J_p: principal segmentation loss.
- J_{a_2},J_{a_3}: auxiliary loss for stage $2, 3$ of the base model, respectively.
- Both J_p, J_{a_i} are standard softmax loss functions.

Crowd detection using semantic image segmentation

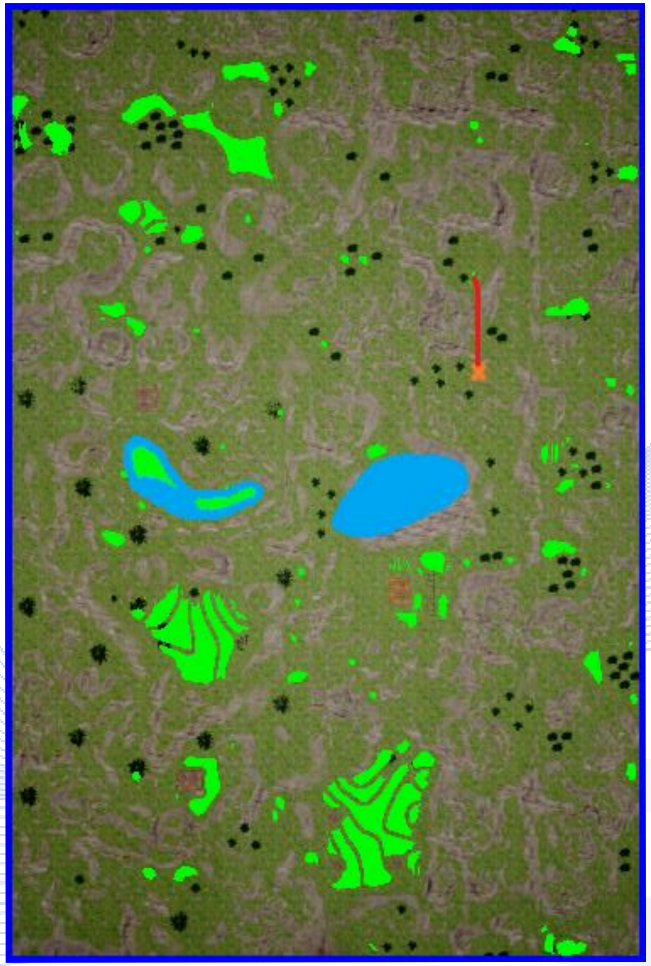
Crowd detection results on a UAV-captured video.

Semantic segmentation for landing site detection



3D projection

INPUT : 2D projection



OUTPUT : Safe Landing Areas in Green Color

Landing site detection results on a synthetic image. Green pixels correspond to landing zones.

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.

Q & A

Thank you very much for your attention!

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