Introduction to Semantic Segmentation

Zamaliev Eduard

eduard.zamaliev@intel.com

intel®

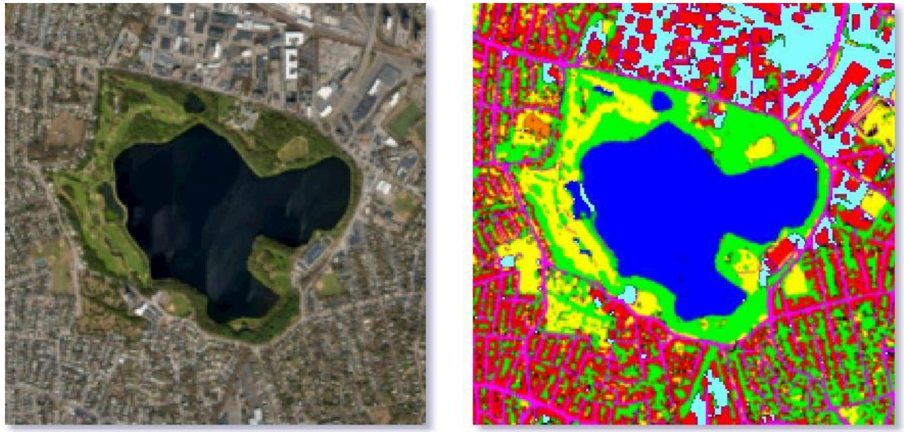
Agenda

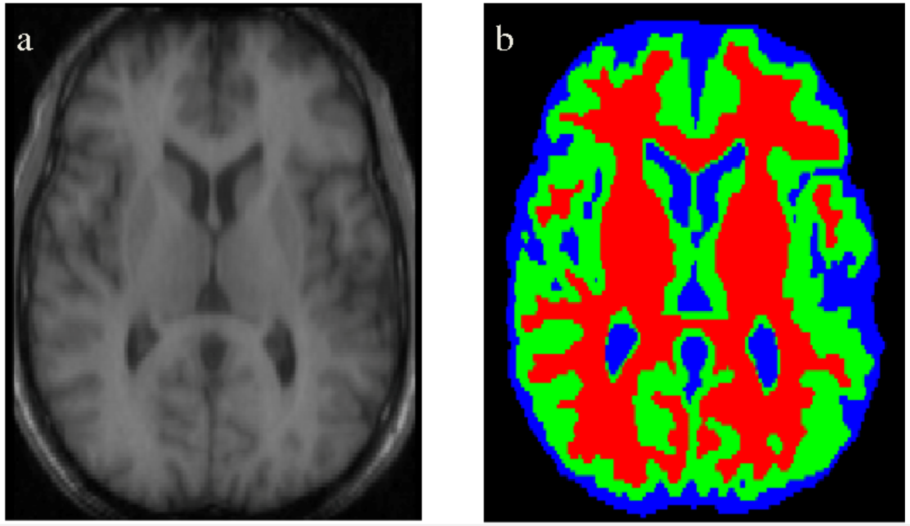
- Problem formulation
- Datasets
- Evaluation metrics
- Architectures
- Loss functions
- Comparison

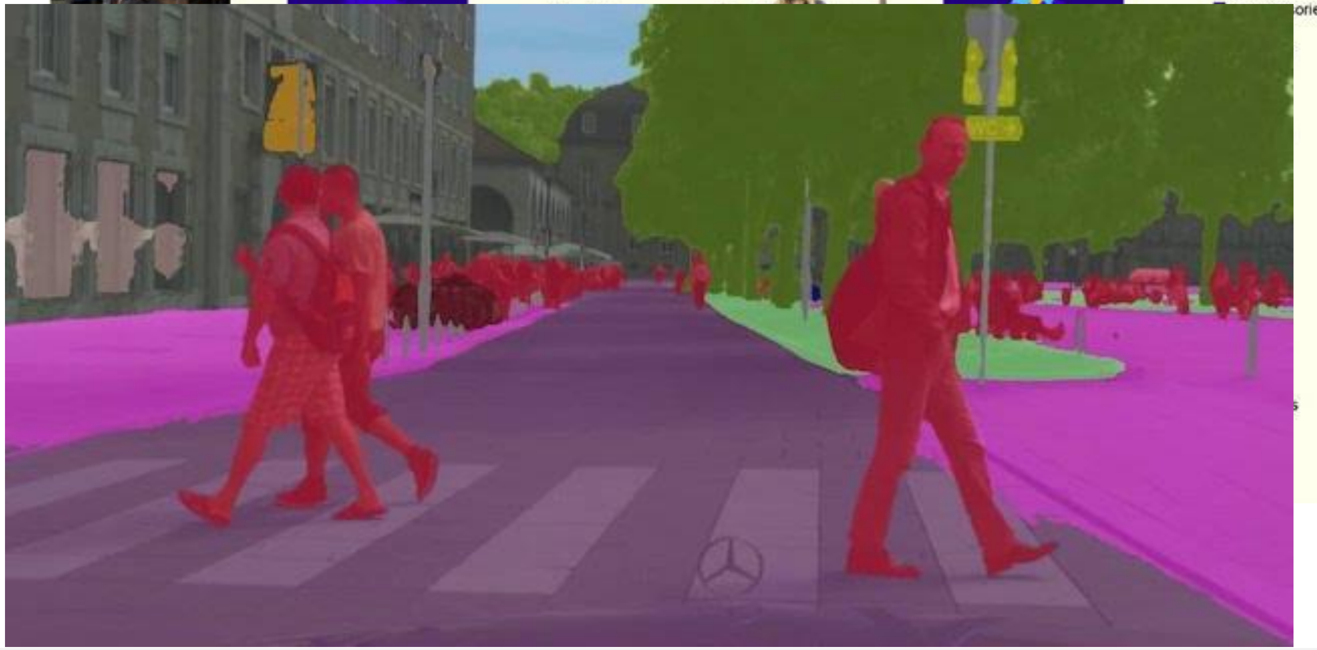
Computer vision problems

- Aerospace photos processing
- Medical scan segmentation
- Autonomous driving

Computer vision problems







Problem formulation

- Input image:

$$I = \{I_{ij}\}_{\substack{0 \le i < w \\ 0 \le j < h}}, I_{ij} \in R^c$$

- Set of classes:

$$C = \{0, 1, \dots, N-1\}$$

- Mask:

$$M = \{M_{ij}\}_{\substack{0 \le i < w \\ 0 \le j < h}}, M_{ij} \in C$$

- Segmentation function:

$$\varphi(R^c) \to C$$

Datasets

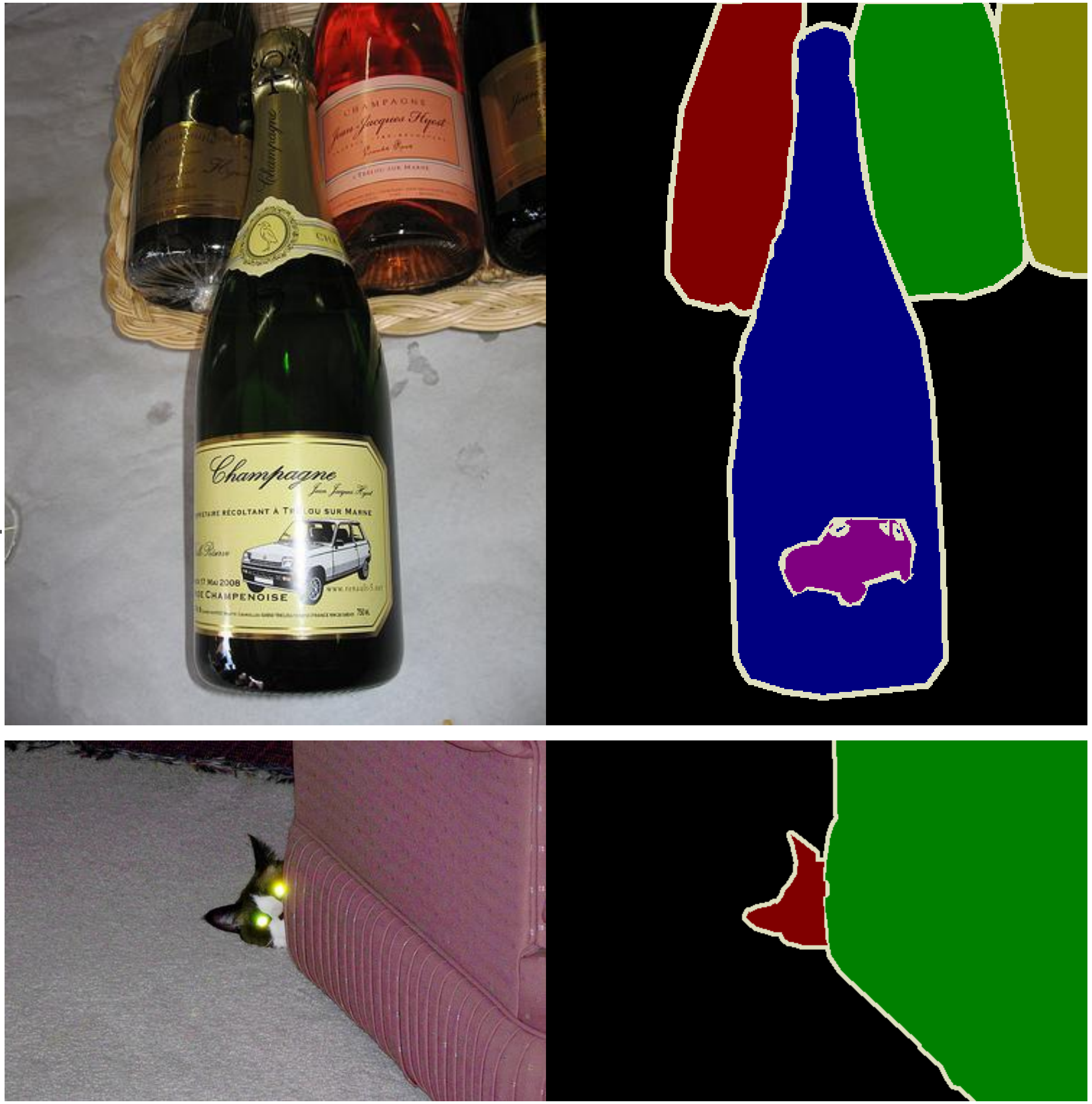
Dataset	Train subset	Test subset	Classes
Common objects			
PASCAL VOC 2012 [http://host.robots.ox.ac.uk/pascal/VOC/voc2012]	9 963	1 447	20
ADE20K [http://groups.csail.mit.edu/vision/datasets/ADE20K]	20 210	2 000	150
MS COCO'15 [http://mscoco.org]	80 000	40 000	80

Datasets

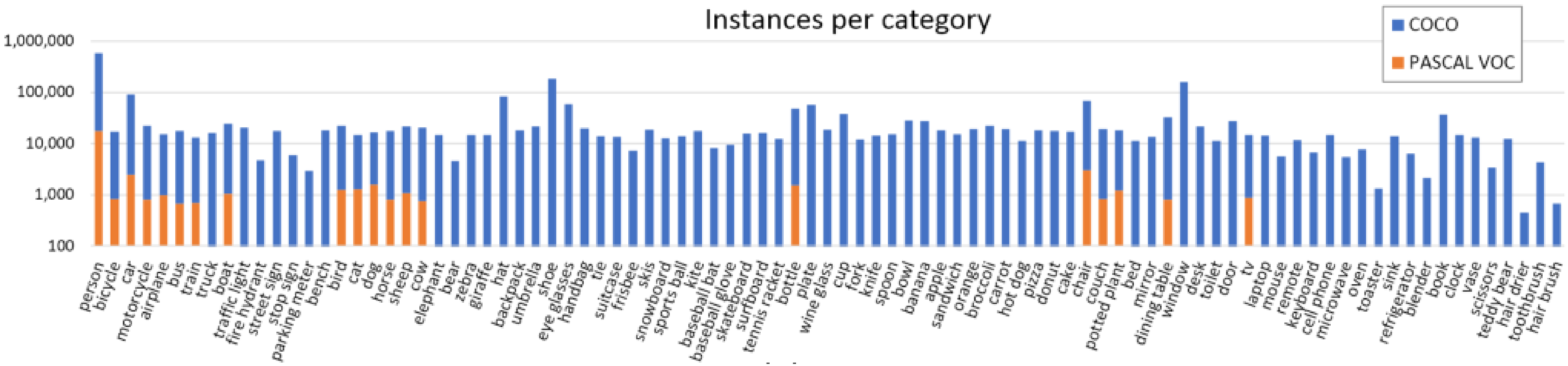
Dataset	Train subset	Test subset	Classes
City, streets, cars			
CamVid [http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid]	468	233	11
Cityscapes [https://www.cityscapes-dataset.com]	2 975	500	19
KITTI [http://www.cvlibs.net/datasets/kitti]	200	200	4
Interiors			
Sun-RGBD [http://rgbd.cs.princeton.edu]	10 355	2 860	37
NYUDv2 [http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html]	795	645	40

Datasets: Pascal VOC2012

- Airplane
- Bicycle
- Bird
- Boat
- Bottle
- Bus
- Car
- Cat
- Chair
- Cow
- dining table
- Dog
- Horse
- Motorbike
- Person
- potted plant
- Sheep
- Sofa
- Train
- tv/monitor



Datasets: MS COCO



Lin T.Y., et al. Microsoft COCO: Common objects in context // Lecture Notes in Computer Science. – Vol. 8693. – 2014. – P. 740-755. [https://arxiv.org/pdf/1405.0312].

Datasets: Citiscapes

- 50 cities
- 5 000 fine annotations
- 20 000 coarse annotations
- 30 classes, 8 groups
- Diversity: daytime, season, weather conditions

The Cityscapes Dataset Homepage [https://www.cityscapes-dataset.com/examples].

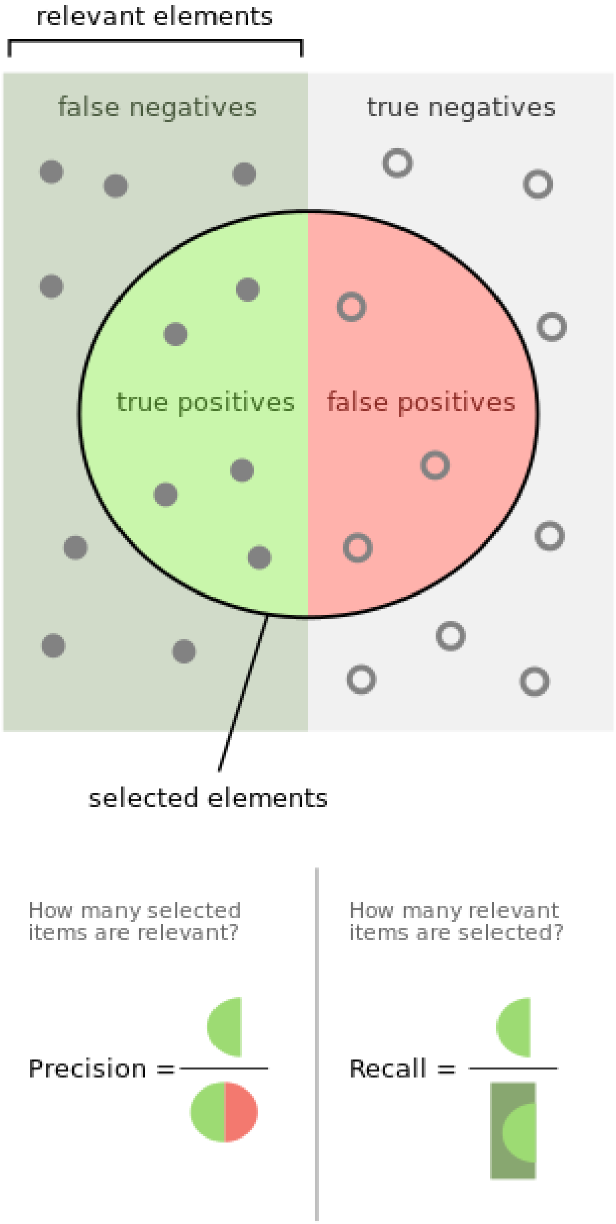
Evaluation metrics

- Pixel accurary
- Mean pixel accuracy over classes
- Jaccard index
- Dice index

Pixel accuracy

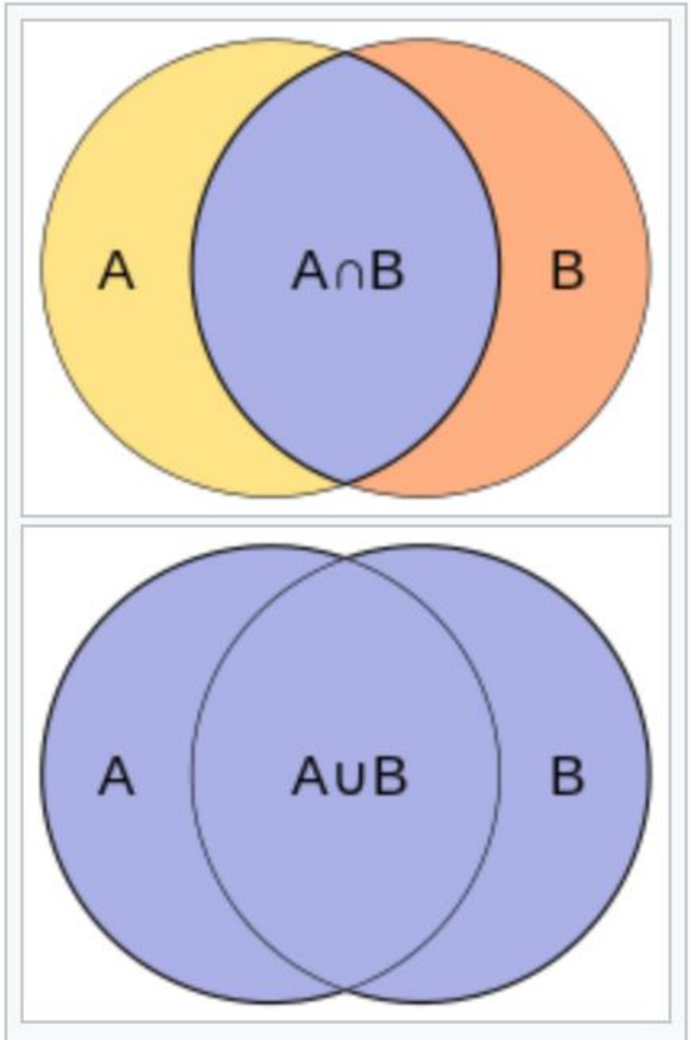
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

	Prediction	
Ground Truth	True	False
True	TP	FN
False	FP	TN

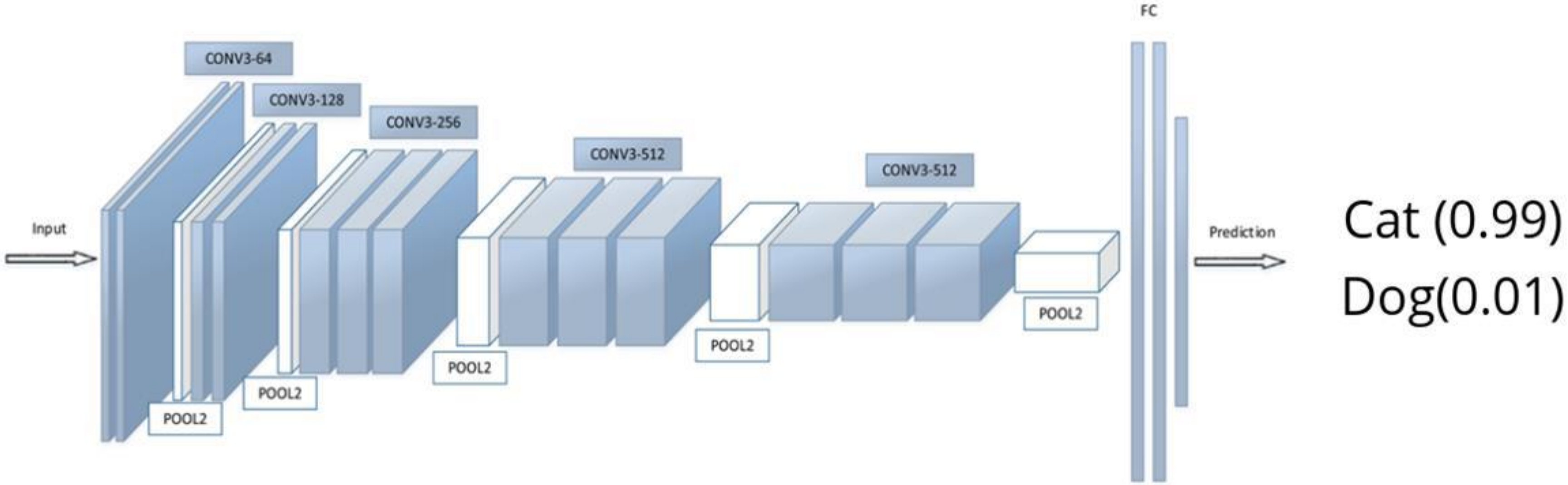


IoU and Jaccard index

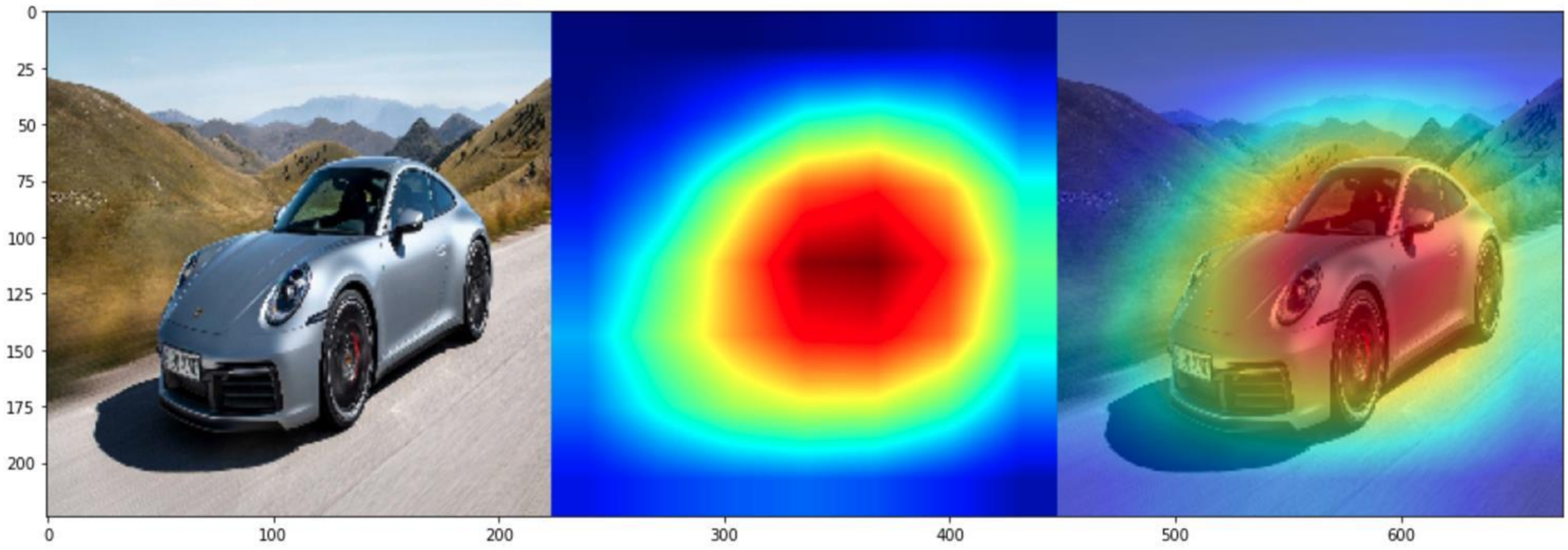
- $IoU(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP+FN+FP}$
- $J(A,B) = 2\frac{|A \cap B|}{|A|+|B|} = \frac{2TP}{2TP+FN+FP}$
- $IoU = \frac{J}{2-J}$



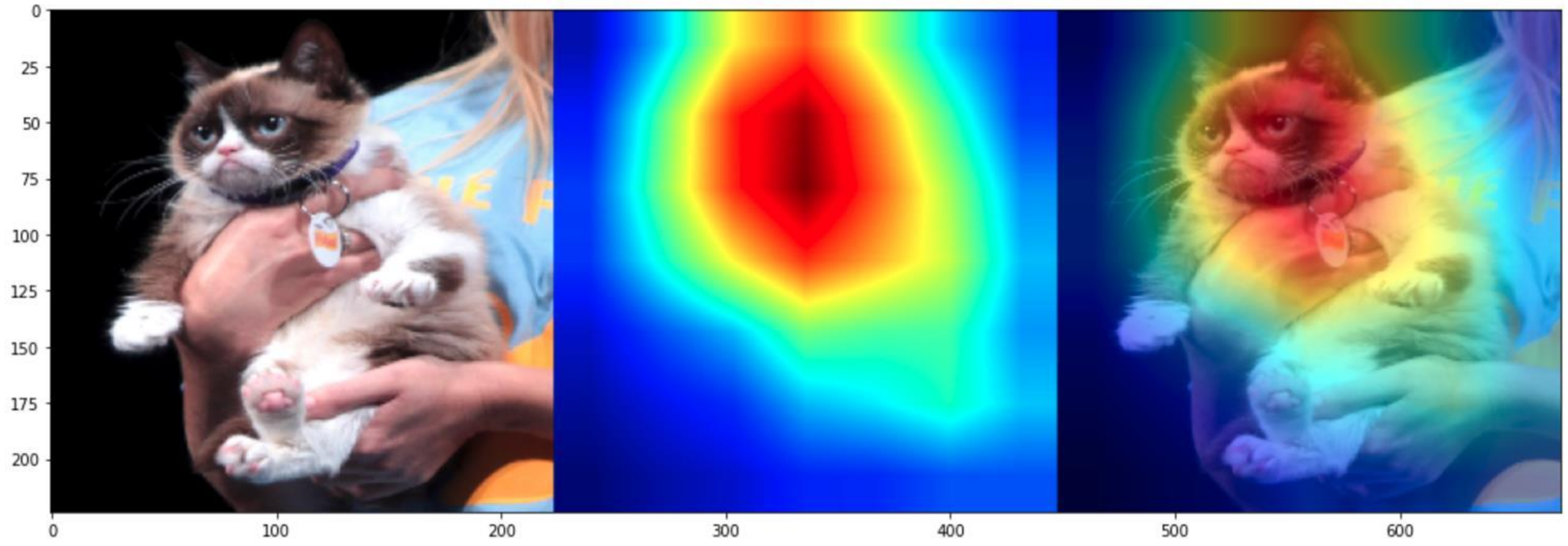
Architectures: CNN



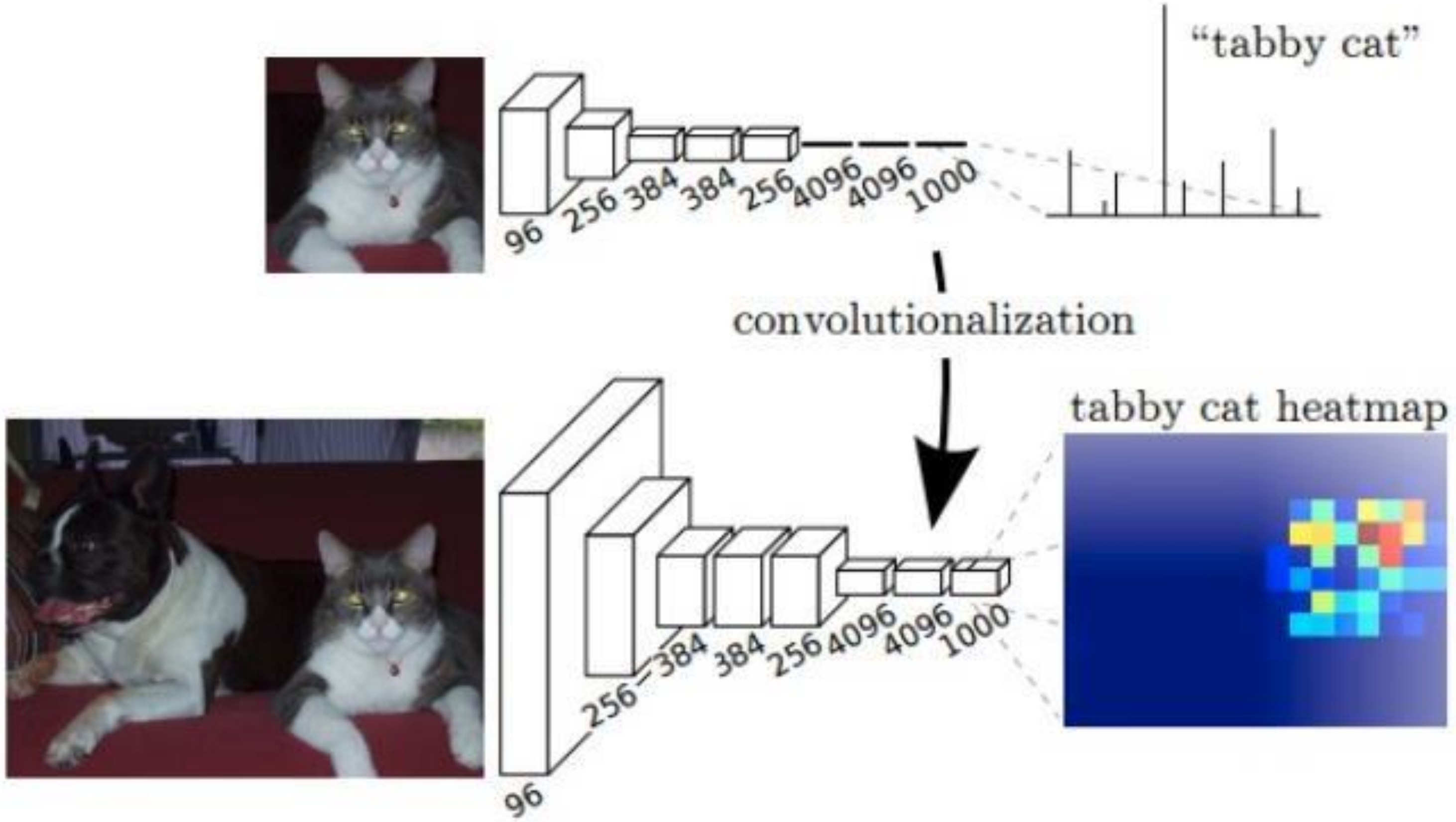
Architectures: CNN



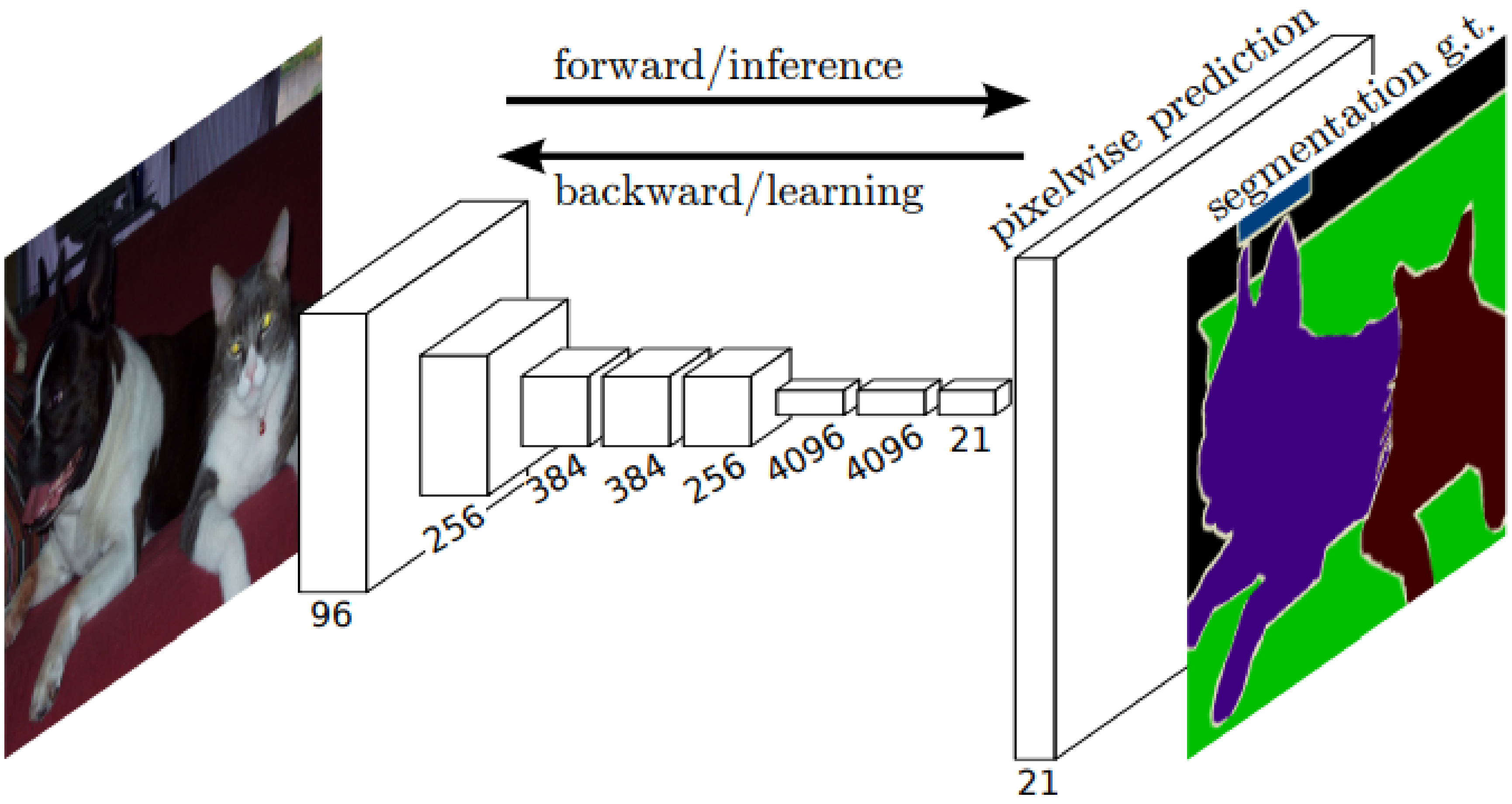
Architectures: CNN



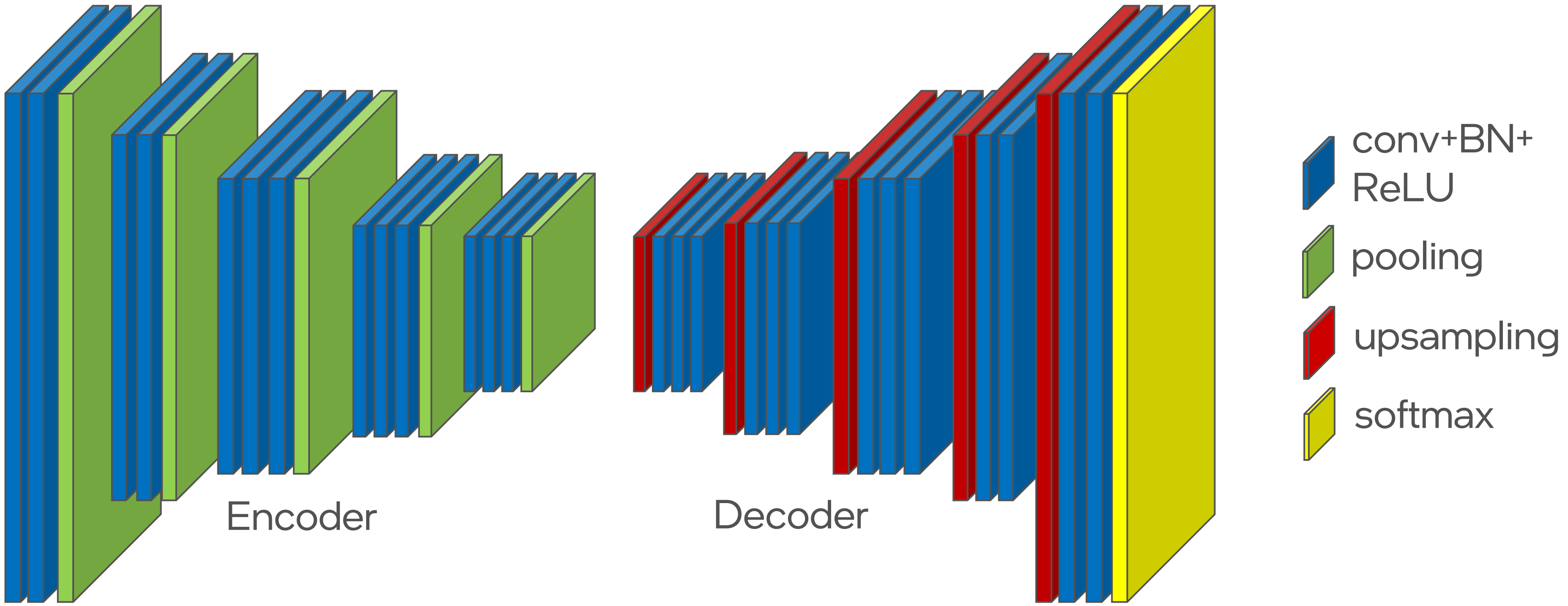
Architecture: FCN



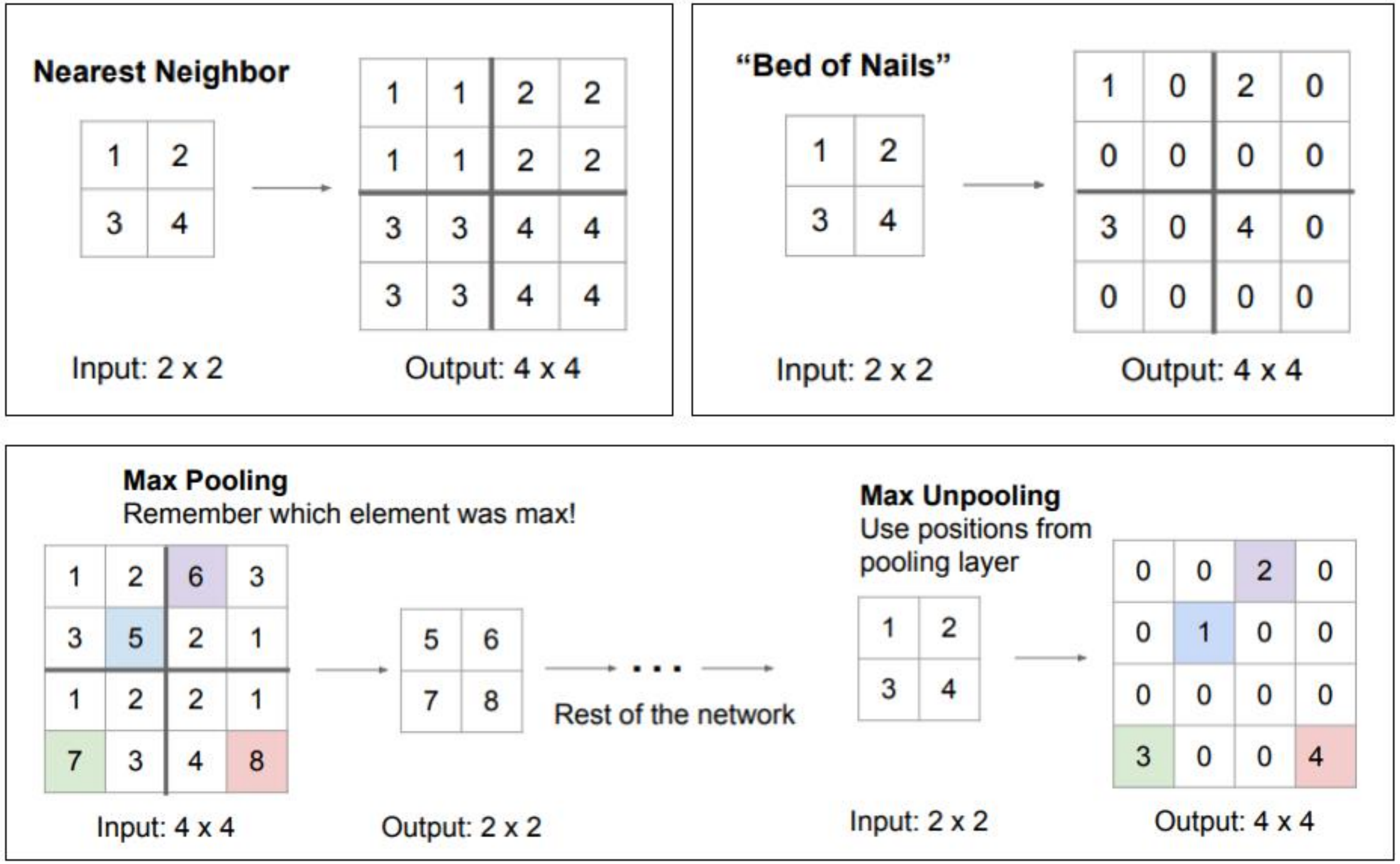
Architecture: FCN



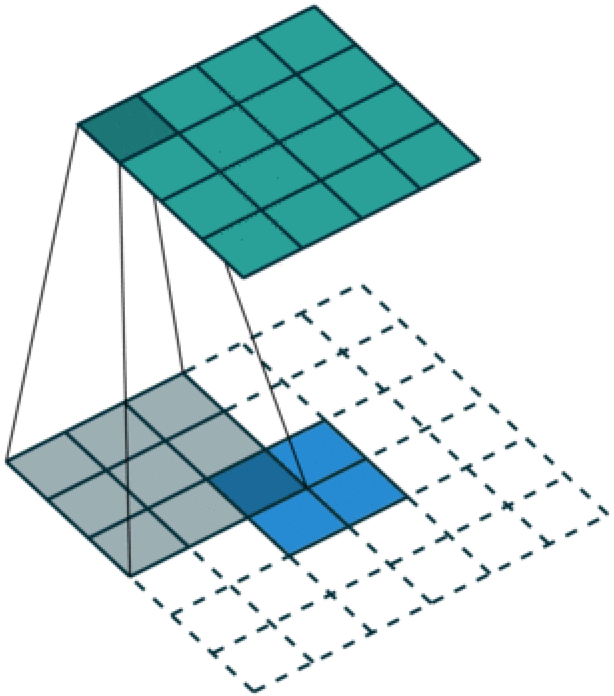
Architectures: SegNet



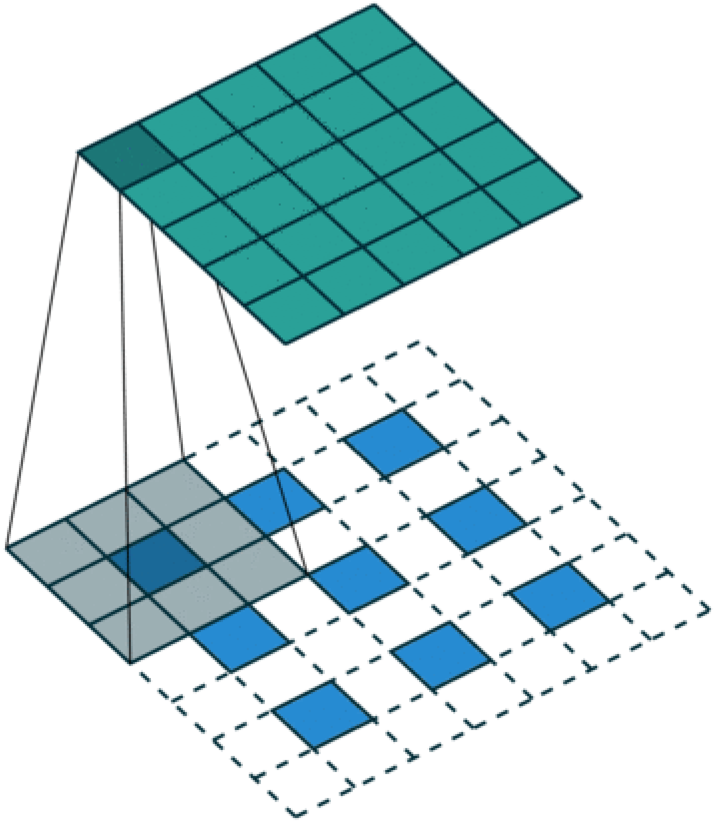
Architectures: Upsampling



Architectures: Deconvolution

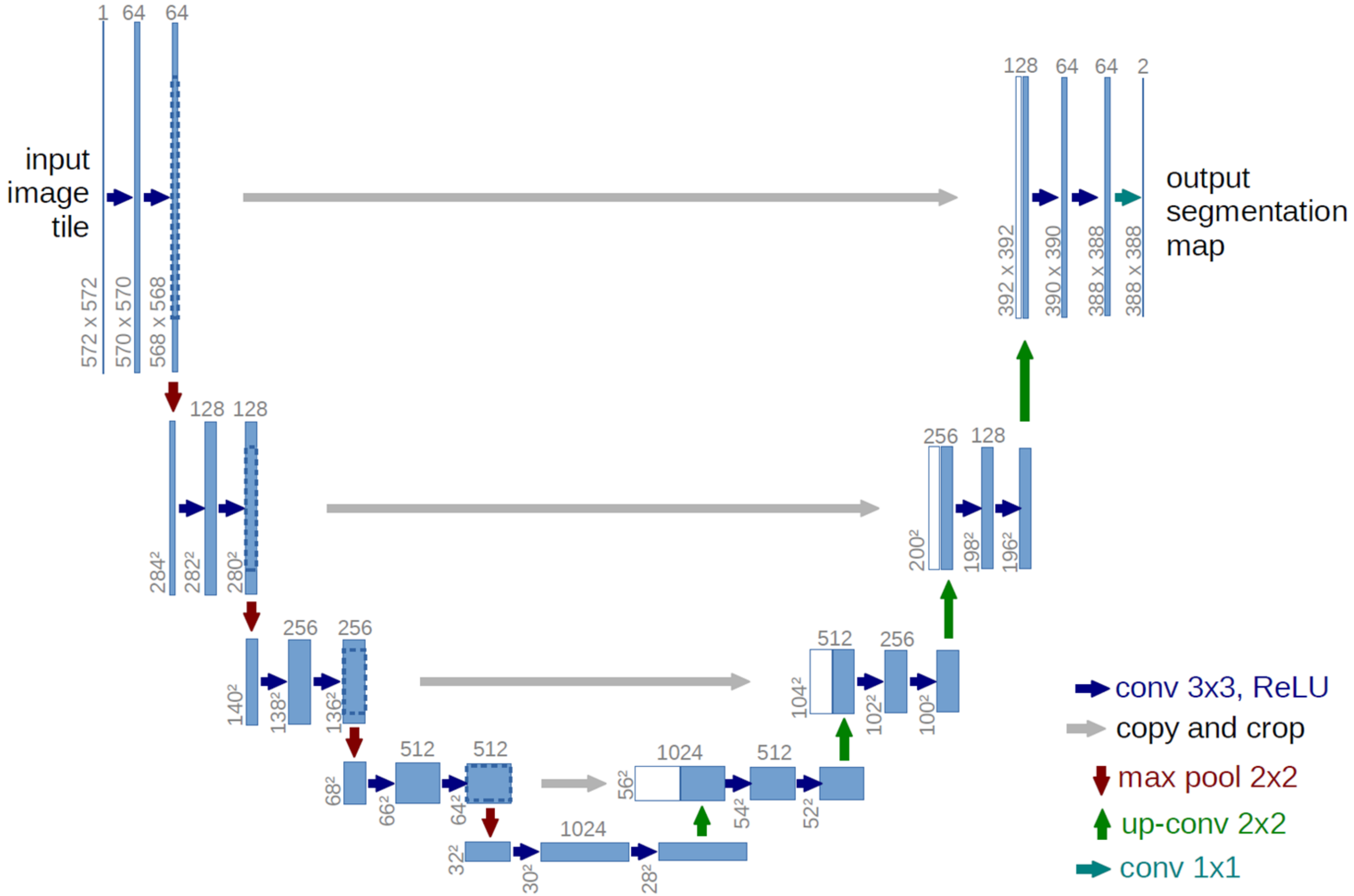


stride=0, padding=0

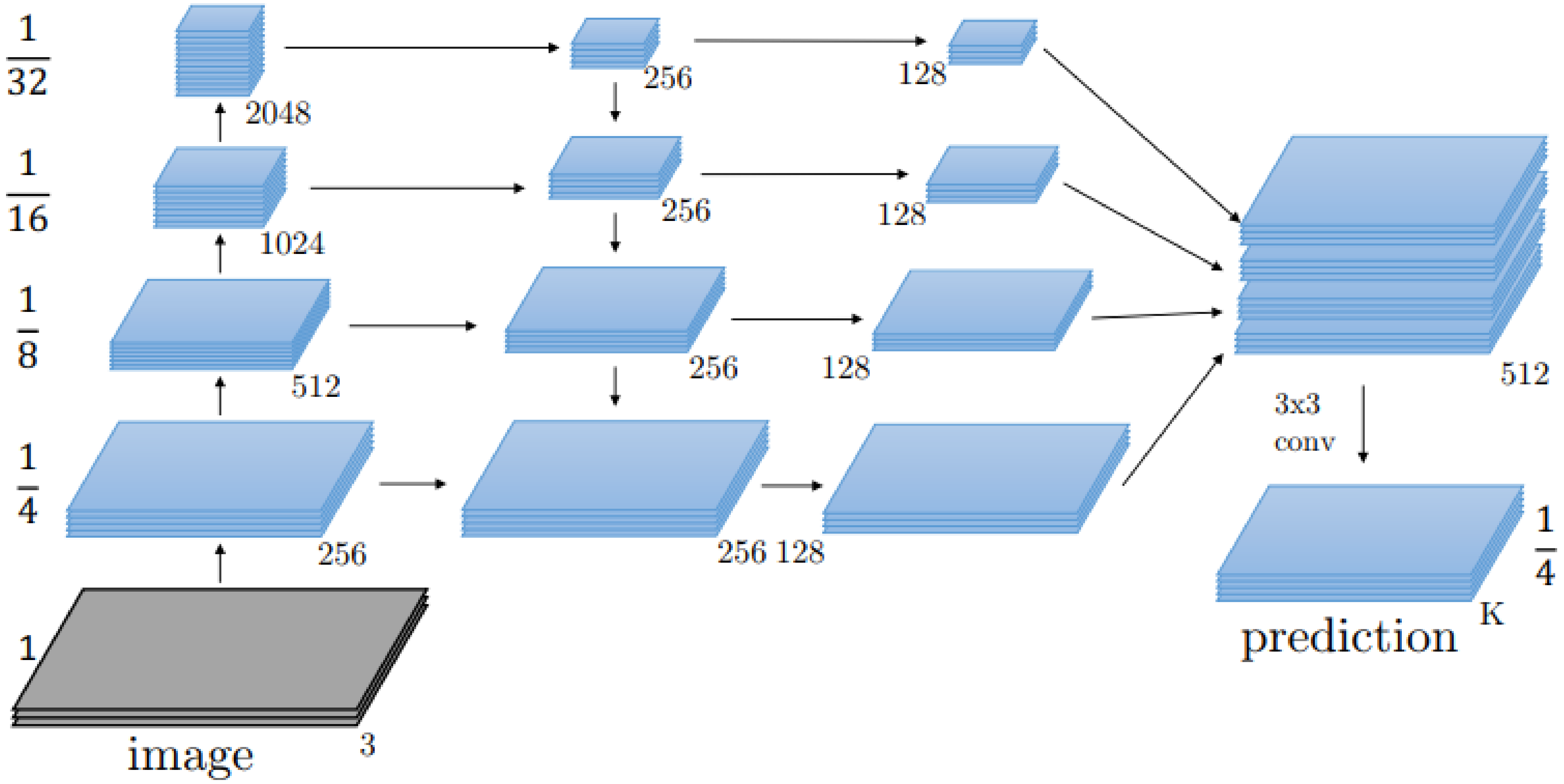


stride=0, padding=1

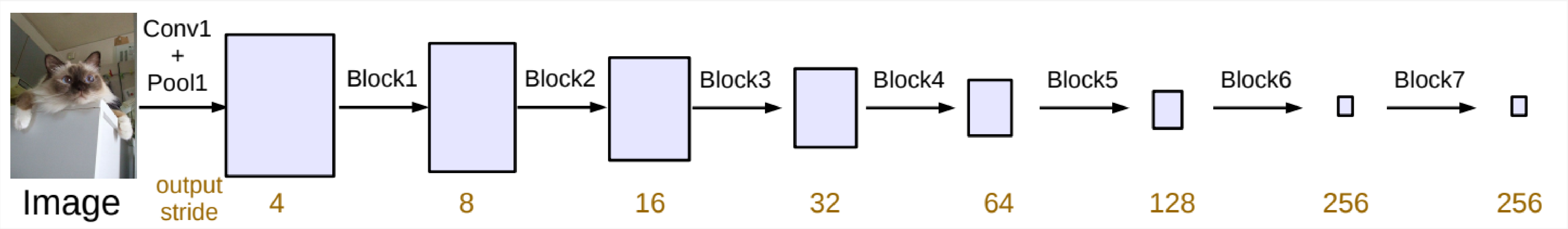
Architectures: UNet



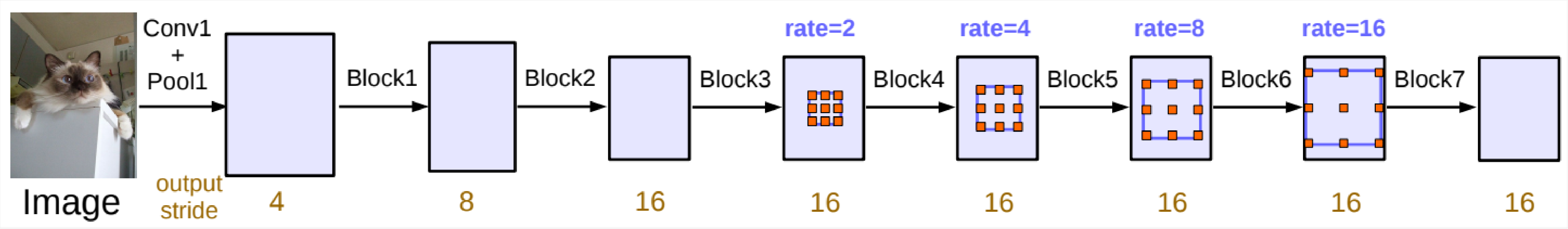
Architectures: Feature Pyramid Network



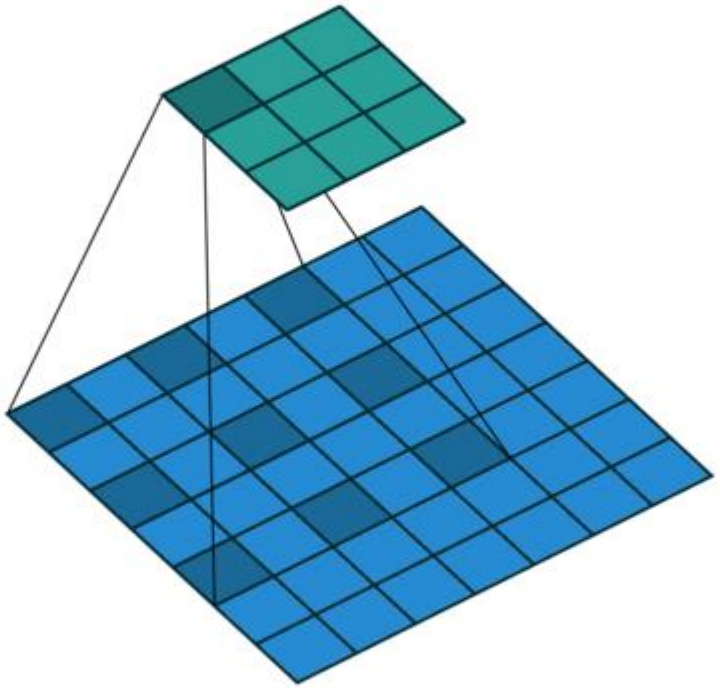
Architectures: DeepLab v1

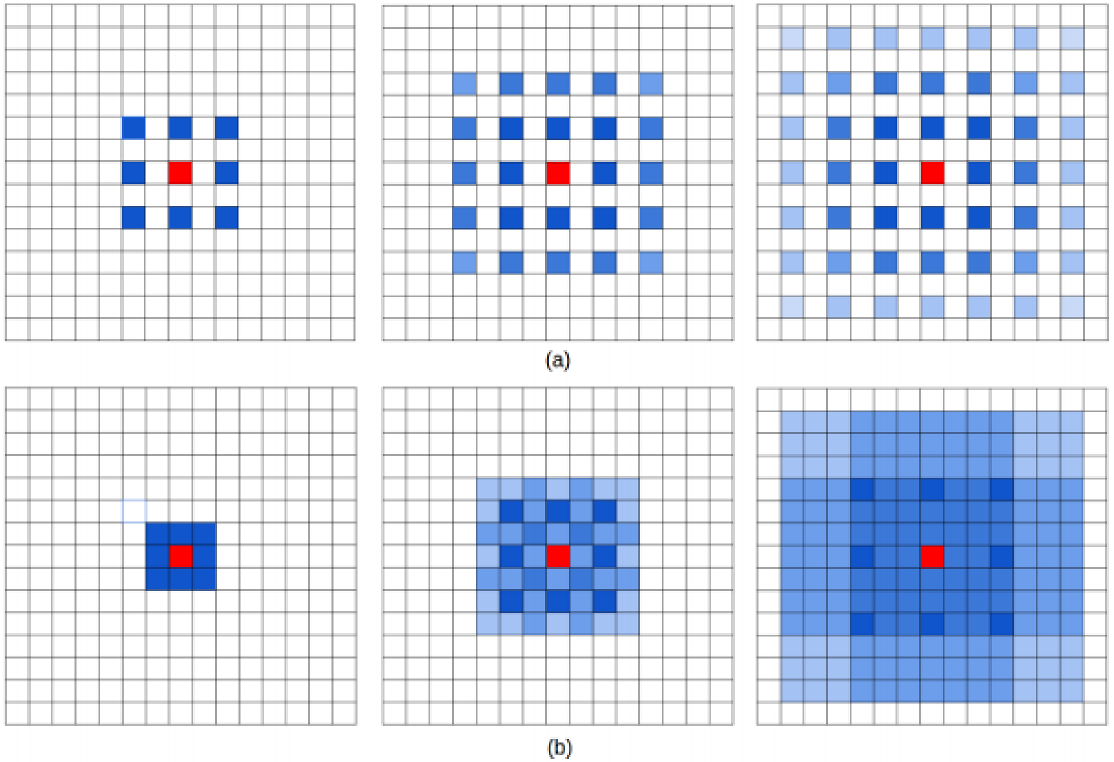


(a) Going deeper without atrous convolution.

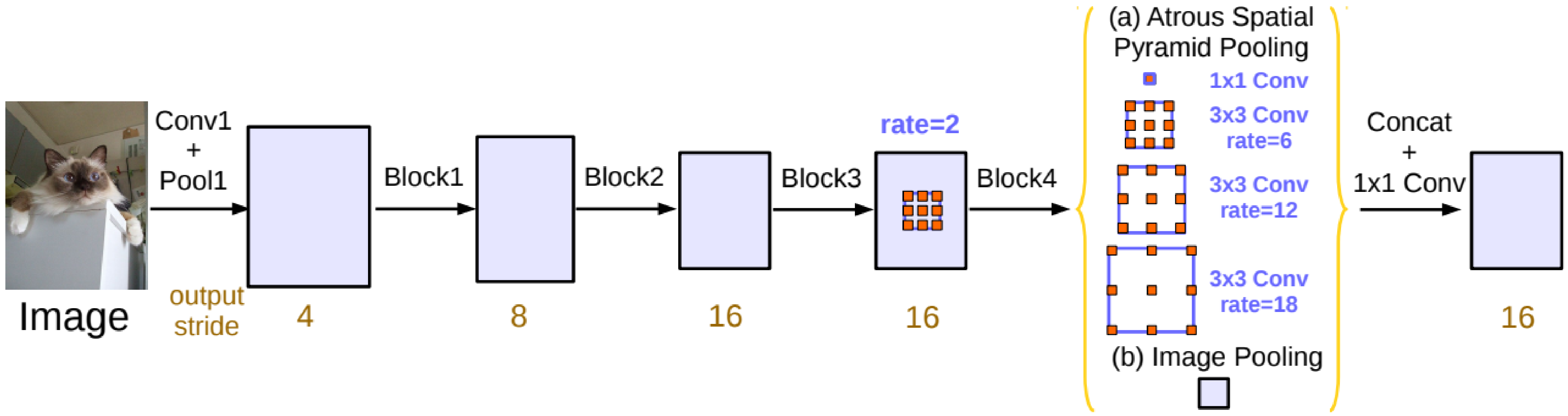


Architectures: Atrous convolutions

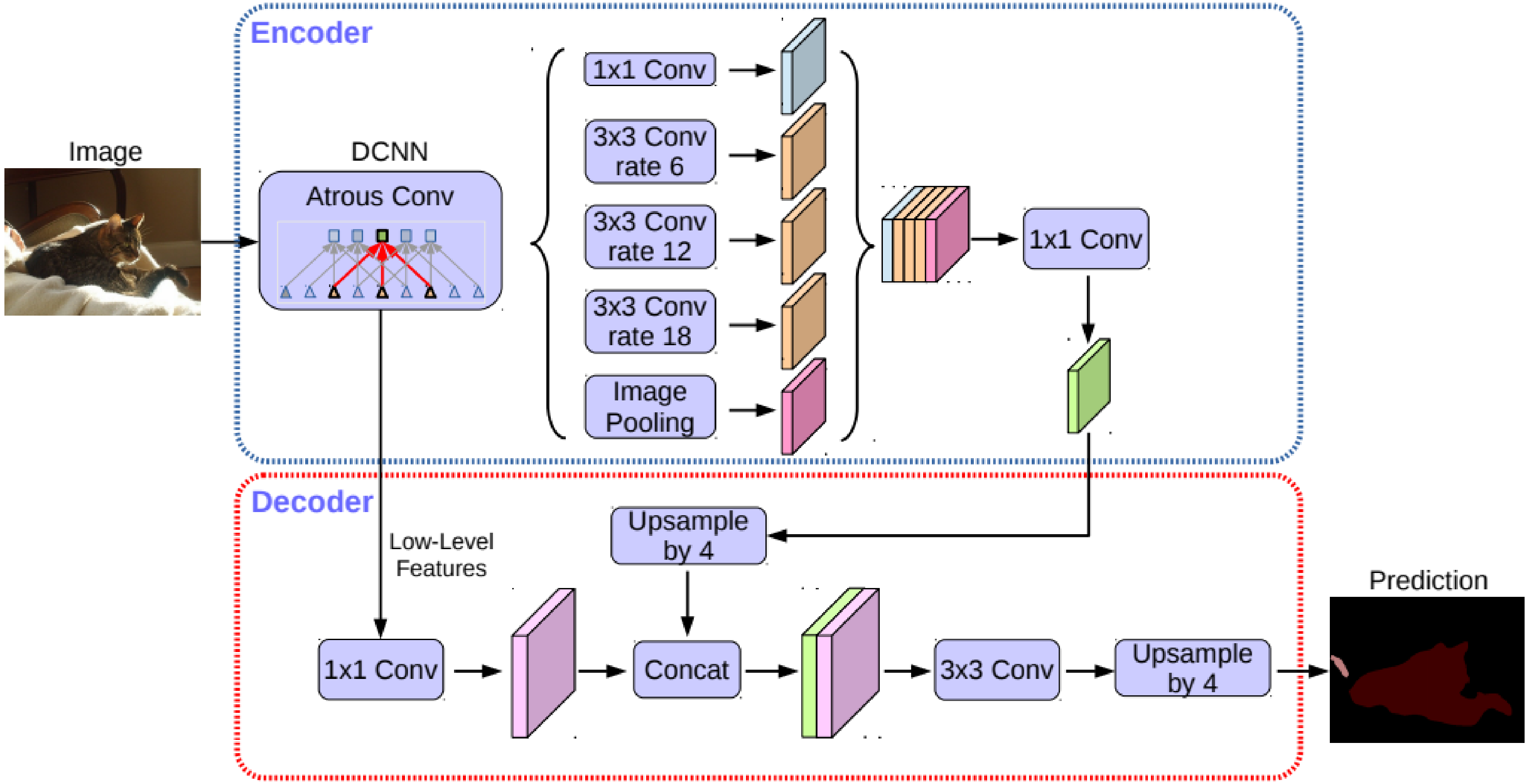




Architectures: DeepLab v2



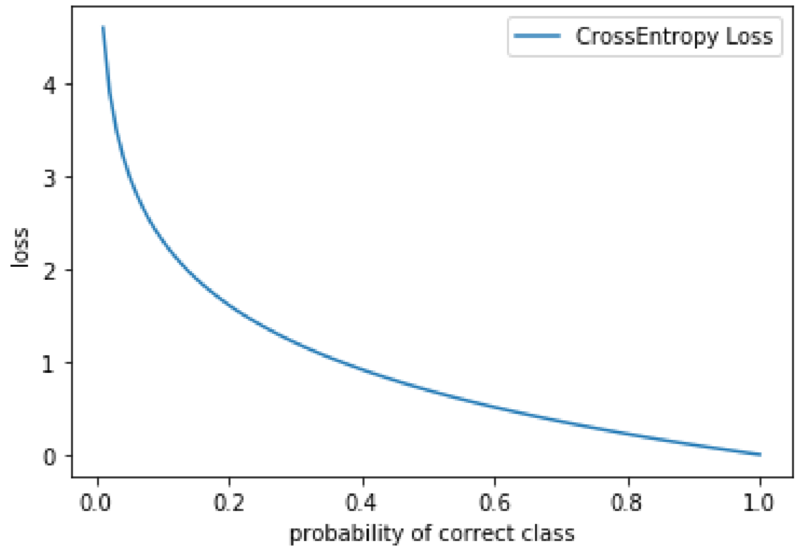
Architectures: DeepLab v3+

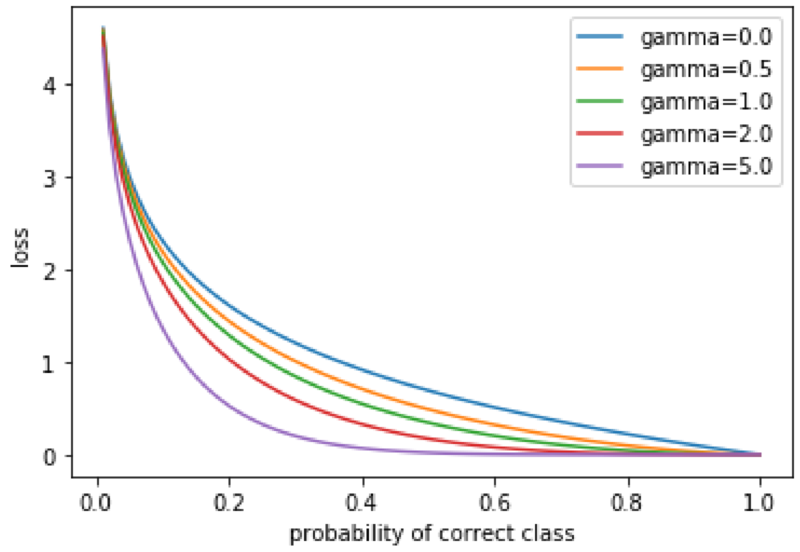


Loss functions: Cross entropy

$$L_{CE}(p,y) = -\sum_{c=1}^{M} y_c \log(p_c)$$

$$L_{CE}(p,y) = -\sum_{c=1}^{M} y_c \,(1-p_c)^{\gamma} \log(p_c)$$



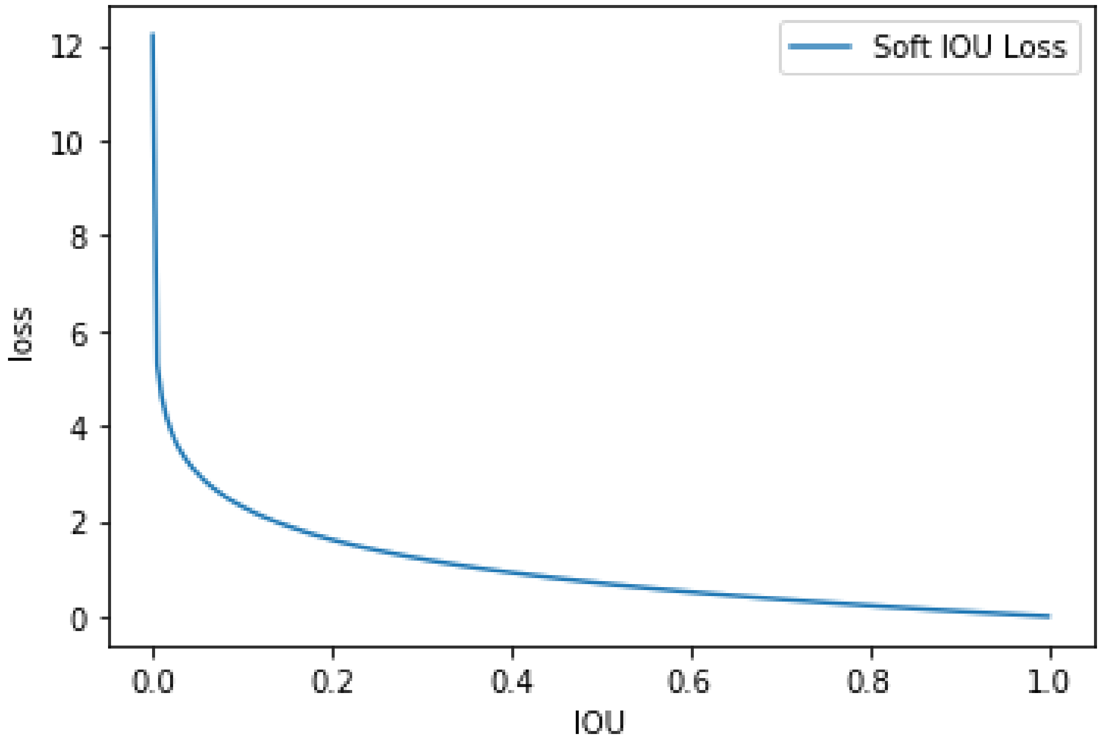


Loss functions: IoU

$$IoU(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$IoU(p,y) = \frac{\sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} p_i y_i}$$

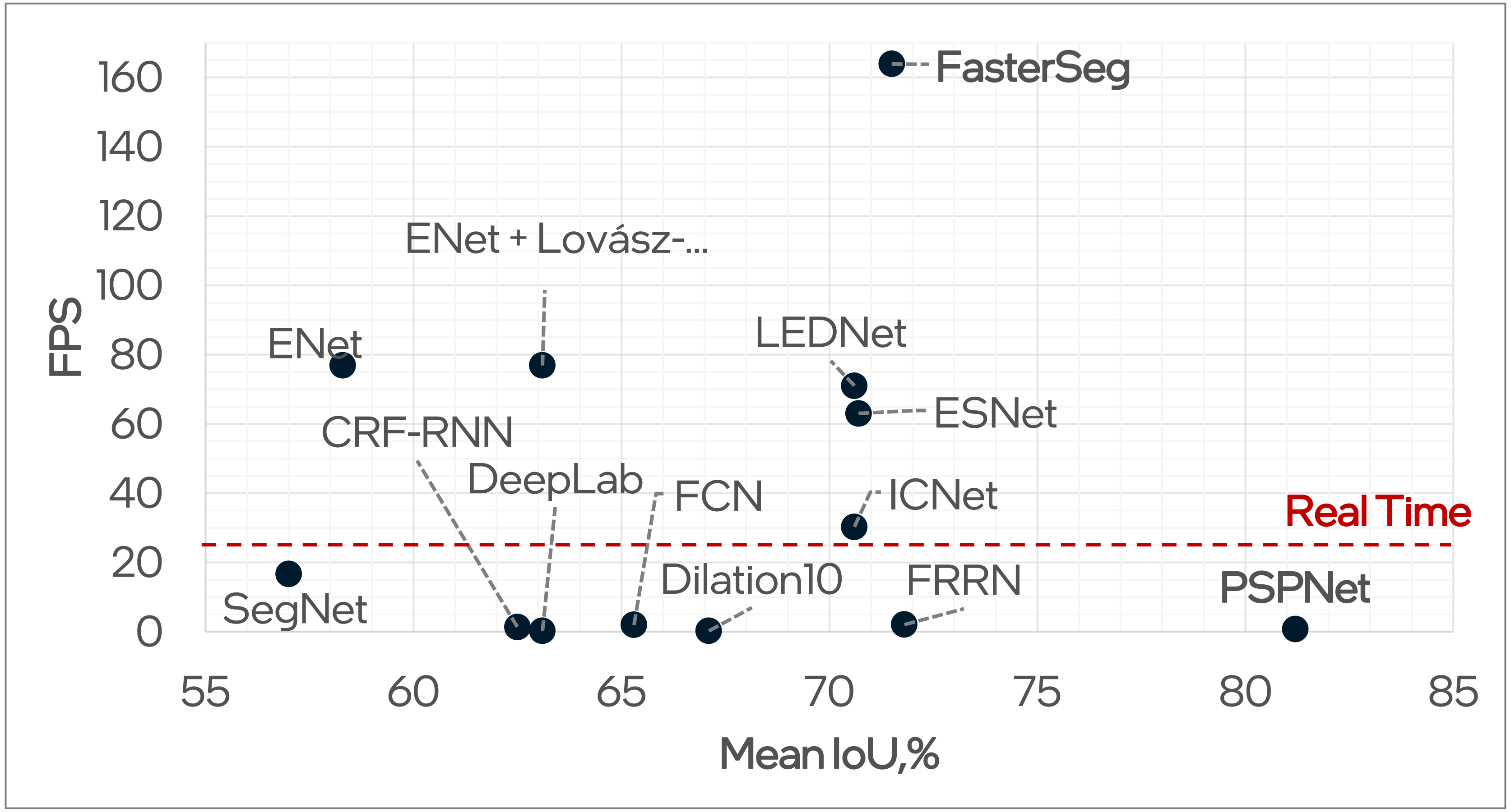
$$L_{IoU} = -\log\left(\frac{\sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} p_i y_i}\right)$$



Comparison

Model	Year	Mean IoU, %	FPS	Latency, ms
DeepLab	2014	63.1	0.25	4000
SegNet	2015	57.0	16.7	60
CRF-RNN	2015	62.5	1.4	700
Dilation10	2015	67.1	0.25	4000
ENet	2016	58.3	76.9	13
FCN	2016	65.3	2	500
FRRN	2016	71.8	2.1	469
ICNet	2017	70.6	30.3	33
PSPNet	2017	81.2	0.78	1288
ENet + Lovász-Softmax	2018	63.1	76.9	13
LEDNet	2019	70.6	71	14
ESNet	2019	70.7	63	16
FasterSeg	2019	71.5	163.9	6.1

Comparison


FasterSeg
ENet + Lovász-...
ENet
LEDNet
ESNet
CRF-RNN
DeepLab
FCN
ICNet
Real Time
SegNet
Dilation10
FRRN
PSPNet
160
140
120
100
80
60
40
20
0
FPS
55
60
65
70
75
80
85
Mean IoU,%

Useful links

- UNet: https://arxiv.org/abs/1505.04597
- DeepLab: https://arxiv.org/abs/1606.00915
- DeepLabV3: https://arxiv.org/abs/1706.05587
- DeepLabV3+: https://arxiv.org/abs/1802.02611
- SegNet: https://arxiv.org/abs/1511.00561
- FCN: https://arxiv.org/abs/1411.4038
- Grad-CAM: https://arxiv.org/abs/1610.02391

- https://github.com/mrgloom/awesome-semantic-segmentation
- Kaggle: https://www.kaggle.com/
- ODS (@bes): https://ods.ai/ https://opendatascience.slack.com
- Deep Learning Book: https://www.deeplearningbook.org/

intel®