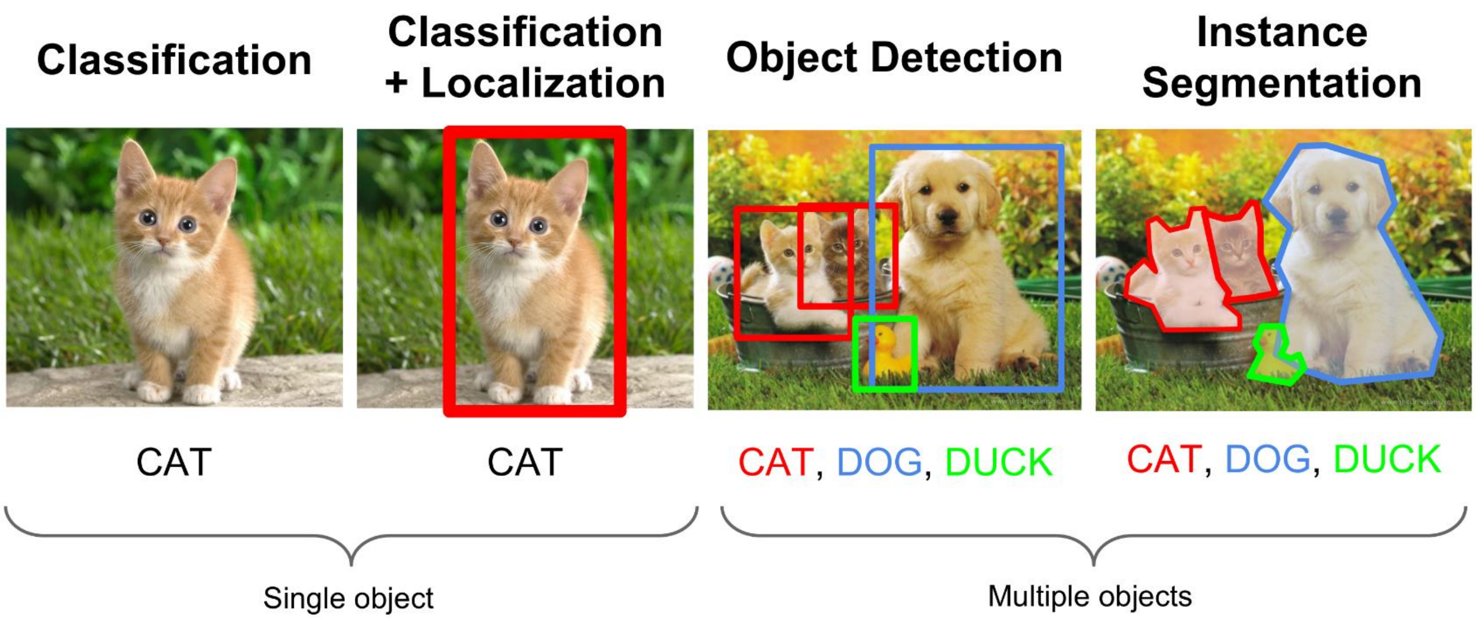
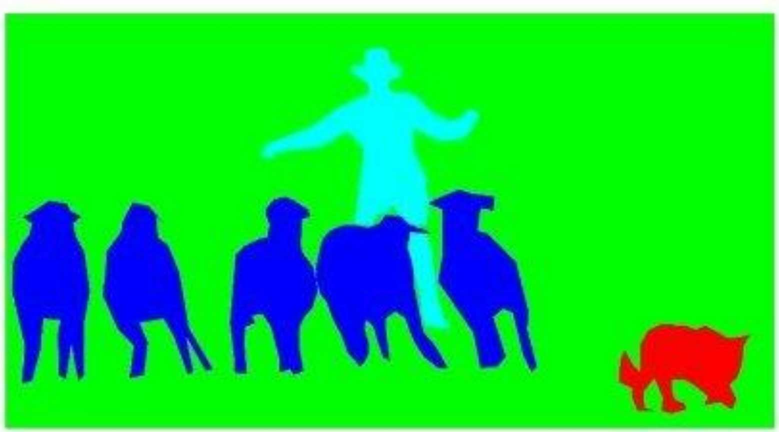
Deep learning in computer vision
- image segmentation

Sean, 201807

Quick Review: Tasks in computer vision



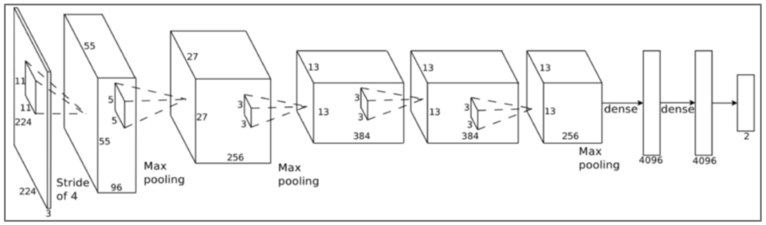


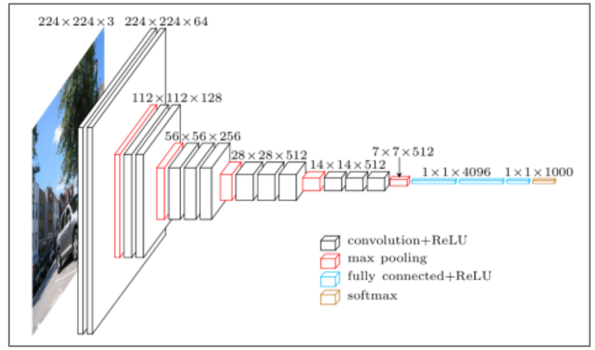
Semantic segmentation

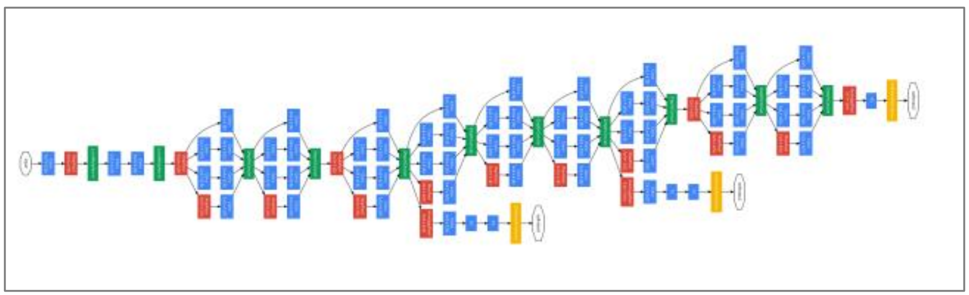
Instance segmentation

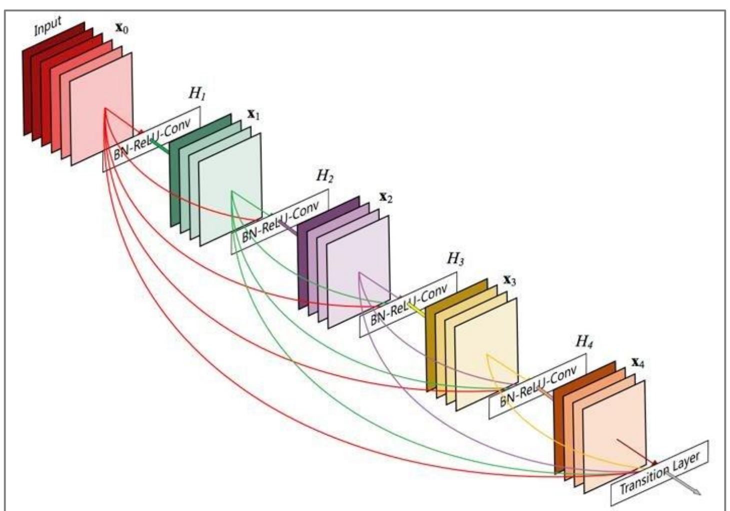
Classification

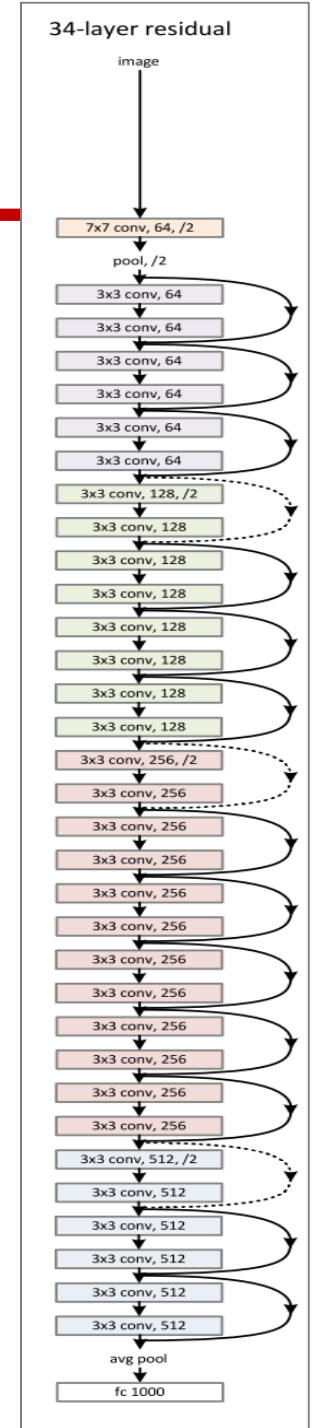
The birth of computer vision：backbone of other tasks









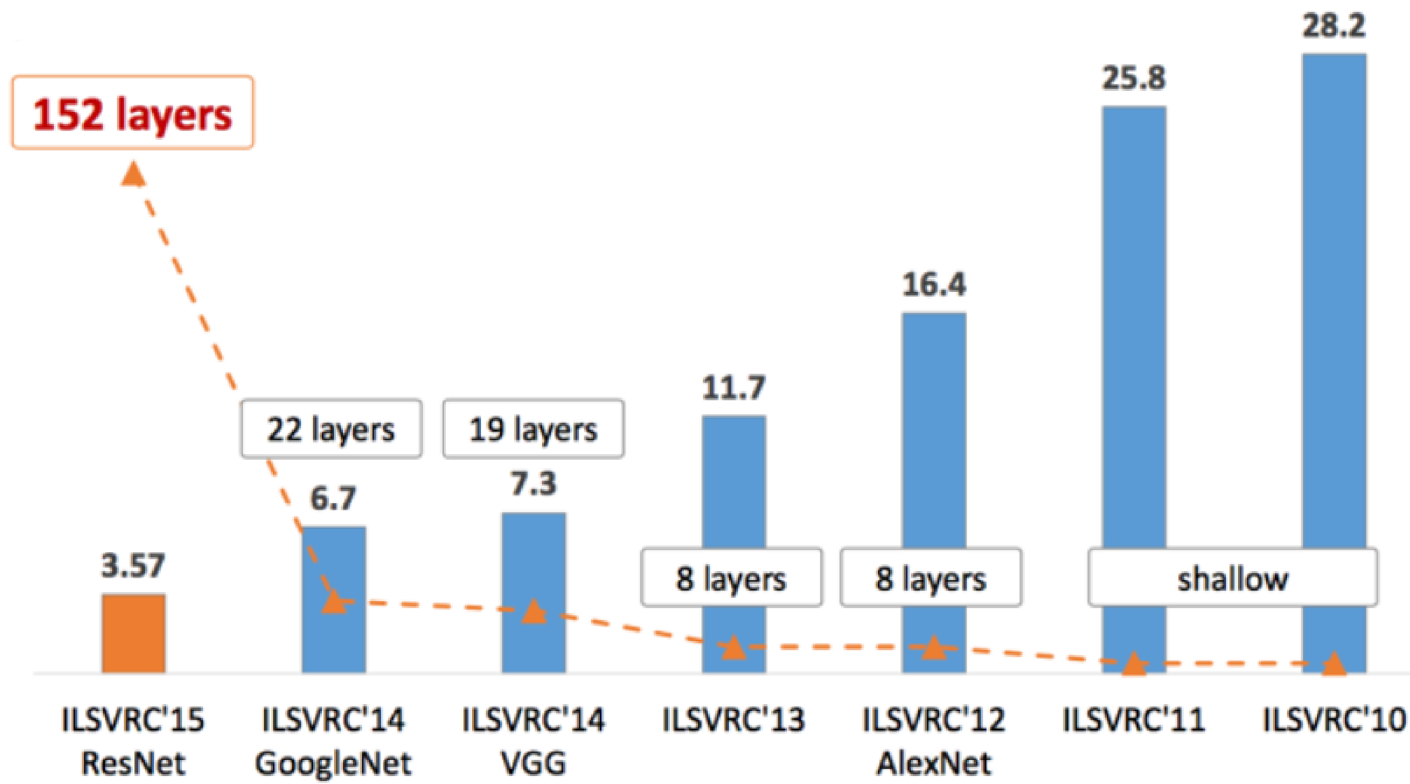


Classification is an easy problem?

NO

Imagenet Large Scale Visual Recognition Challenge
(ILSVRC)
of classification

2010 - 2017



Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

Human performance: ~5% error rate

https://medium.com/@siddharthdas_32104/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5

Classification is an easy problem?

- No, there are lots of problems you’ll met in classification problems.
  - data imbalance
  - noisy label
  - ultra big image
  - 3d image (e.g. PET/CT) or multi-sources (e.g. radiology + pathology)
  - …
- To overcome these problems, you might need to ...
  - Have a better pre-processing pipeline (e.g. generator & multi-processing)
  - Do lots of reasonable augmentation
  - Batching skills
  - Modify network structures
  - Modify loss functions

Classification is an easy problem?

- No, there are lots of problems you'll met in classification problems.
  - data imbalance
  - noisy label
  - ultra big image
  - 3d image (e.g. PET/CT) or multi-sources (e.g. radiology + pathology)
  - …

But these are not problems TODAY!

- To overcome these problems, you might need to ...
  - Have a better pre-processing pipeline (e.g. generator & multi-processing)
  - Do lots of reasonable augmentation
  - Batching skills
  - Modify network structures
  - Modify loss functions

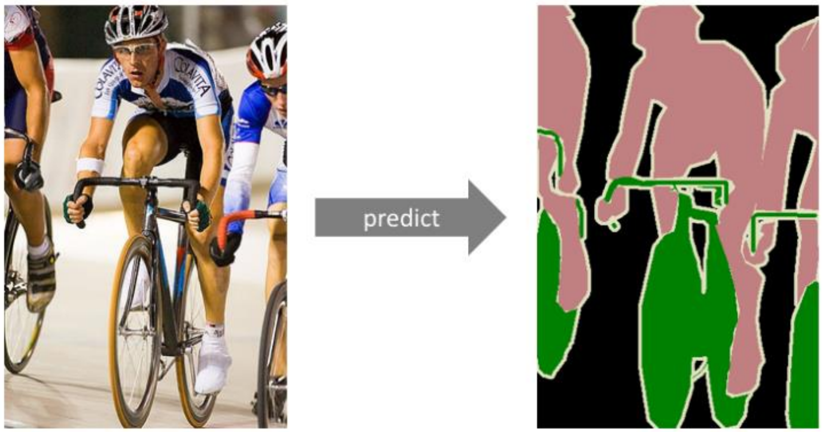
Today’s talk

- Semantic segmentation - Sean
- Instance segmentation - Jimmy

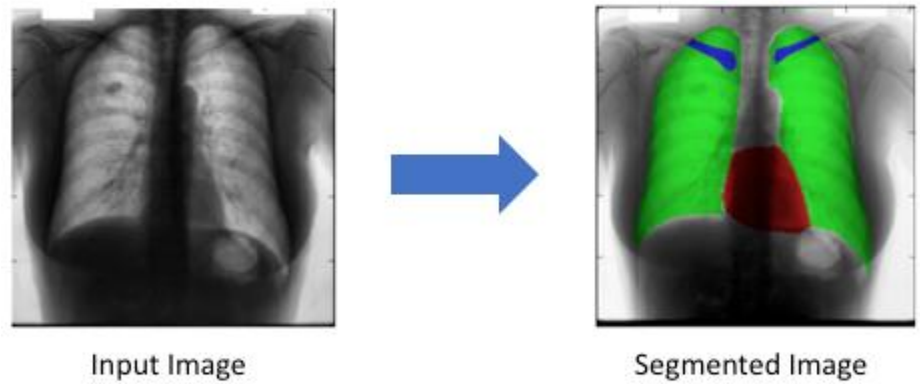
Semantic segmentation

Examples of semantic segmentation

COCO

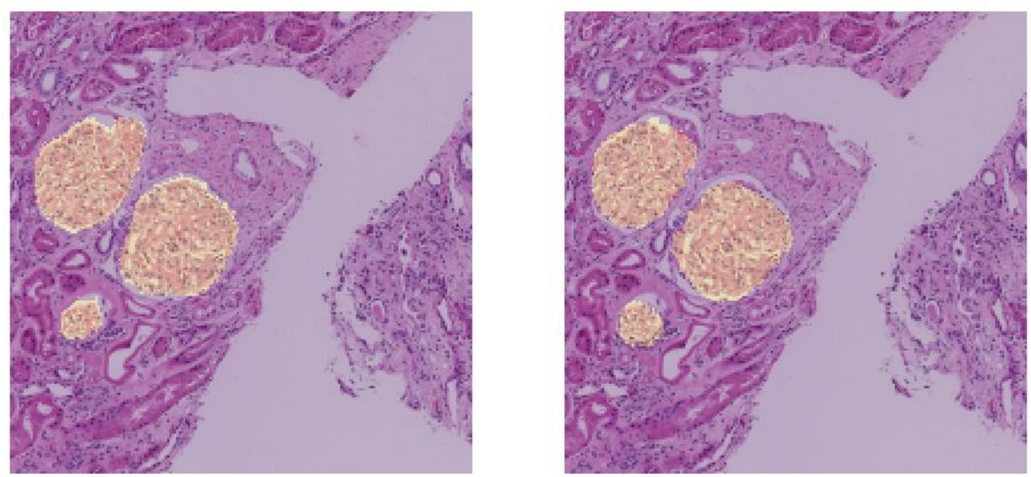


Organ



CityScape

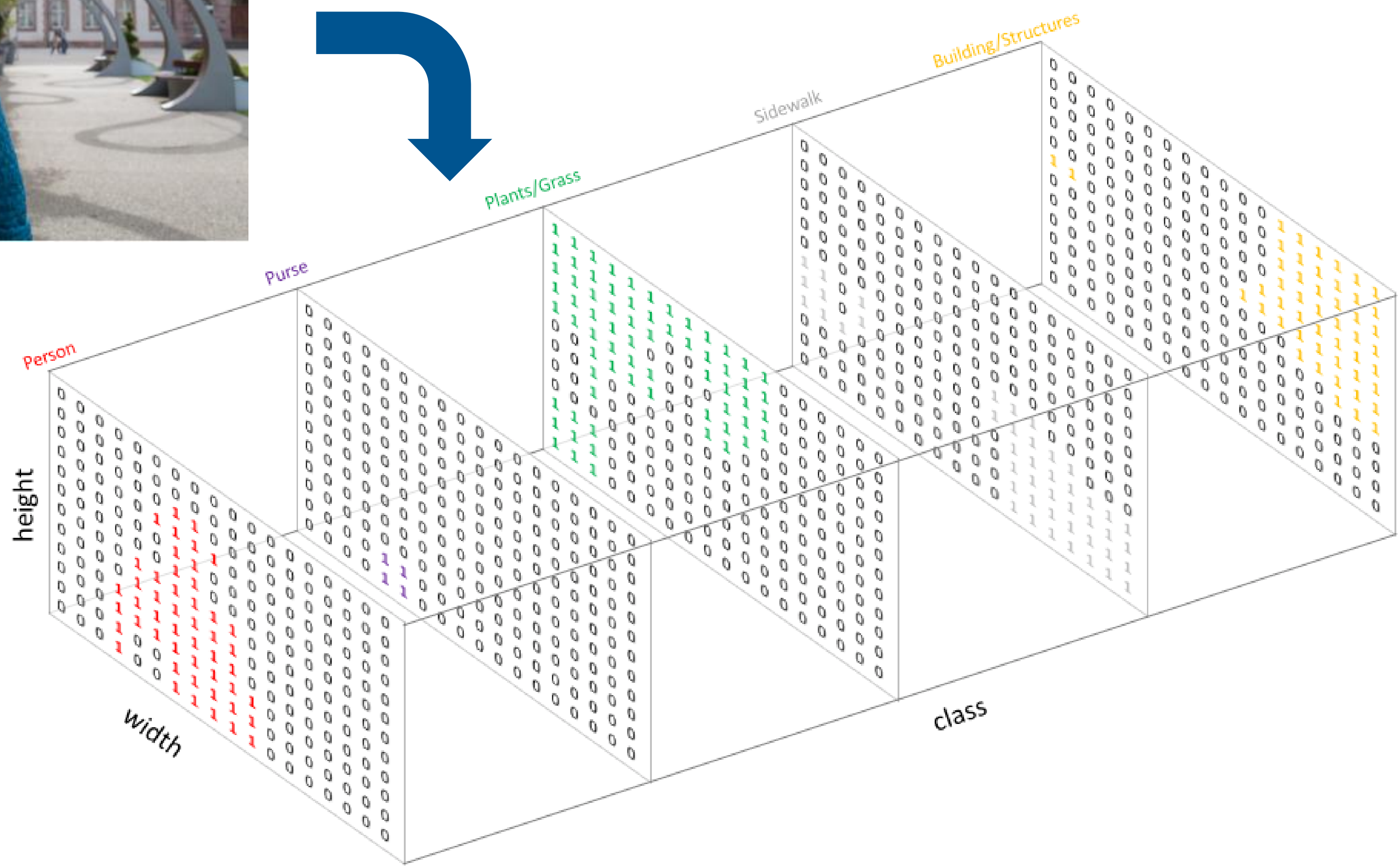
Digital pathology



Ground truth

Prediction

Data labeling

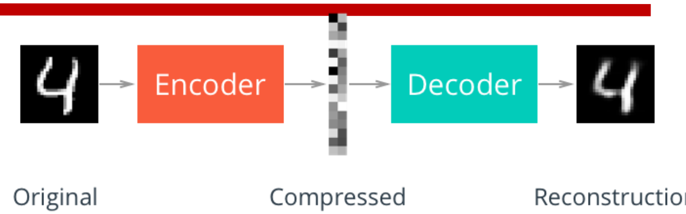


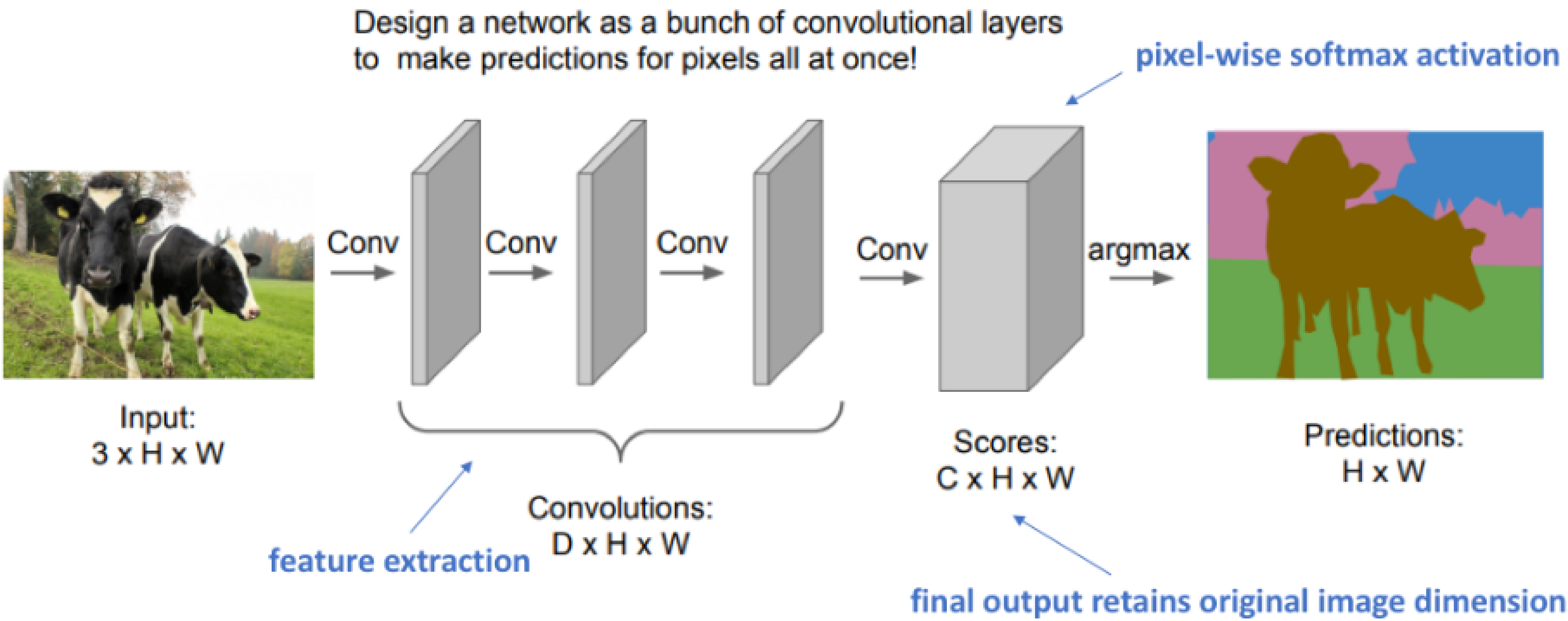
Credit: https://www.jeremyjordan.me/semantic-segmentation/#advanced_unet

Model

Model.01 – AE-like DCNN

- Intuition
  - Input size = output size
  - Output: softmax with n-classes masks
  - Model = feature extractor

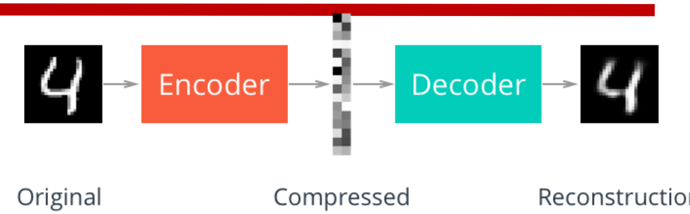


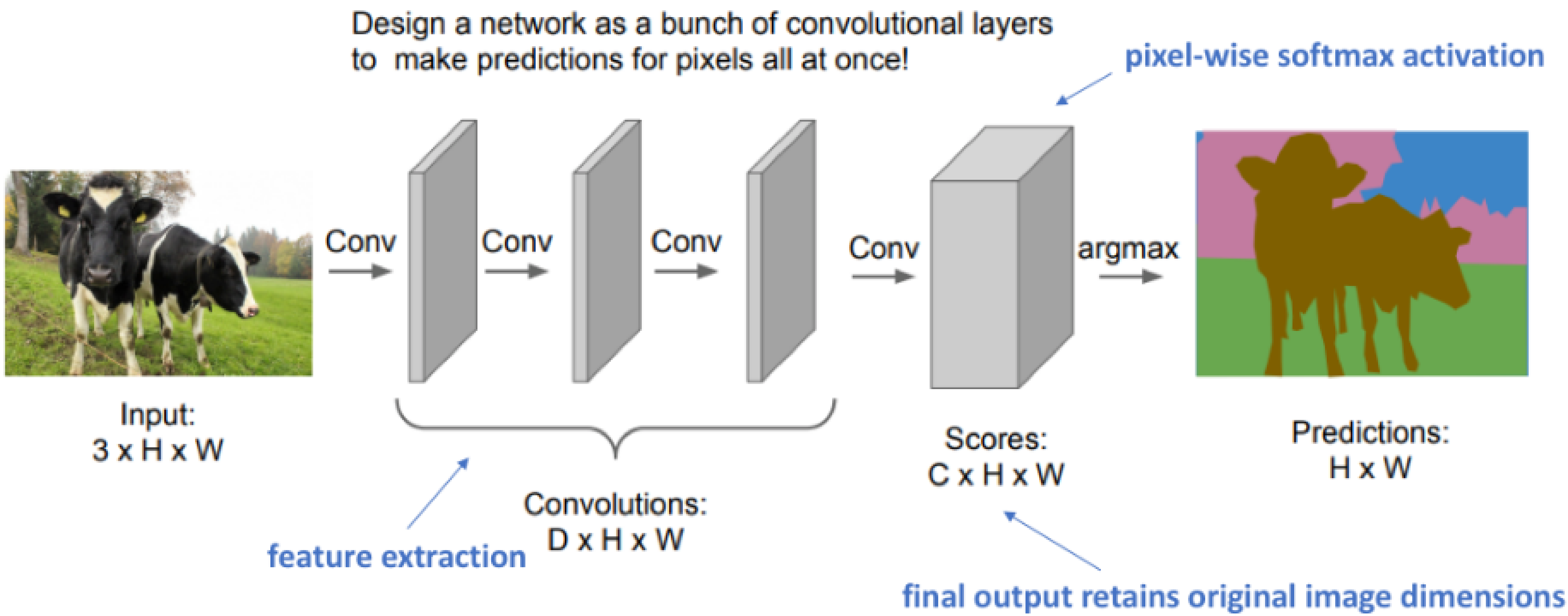


Model

Model.01 – AE-like DCNN

- Intuition
  - Input size = output size
  - Output: softmax with n-classes masks



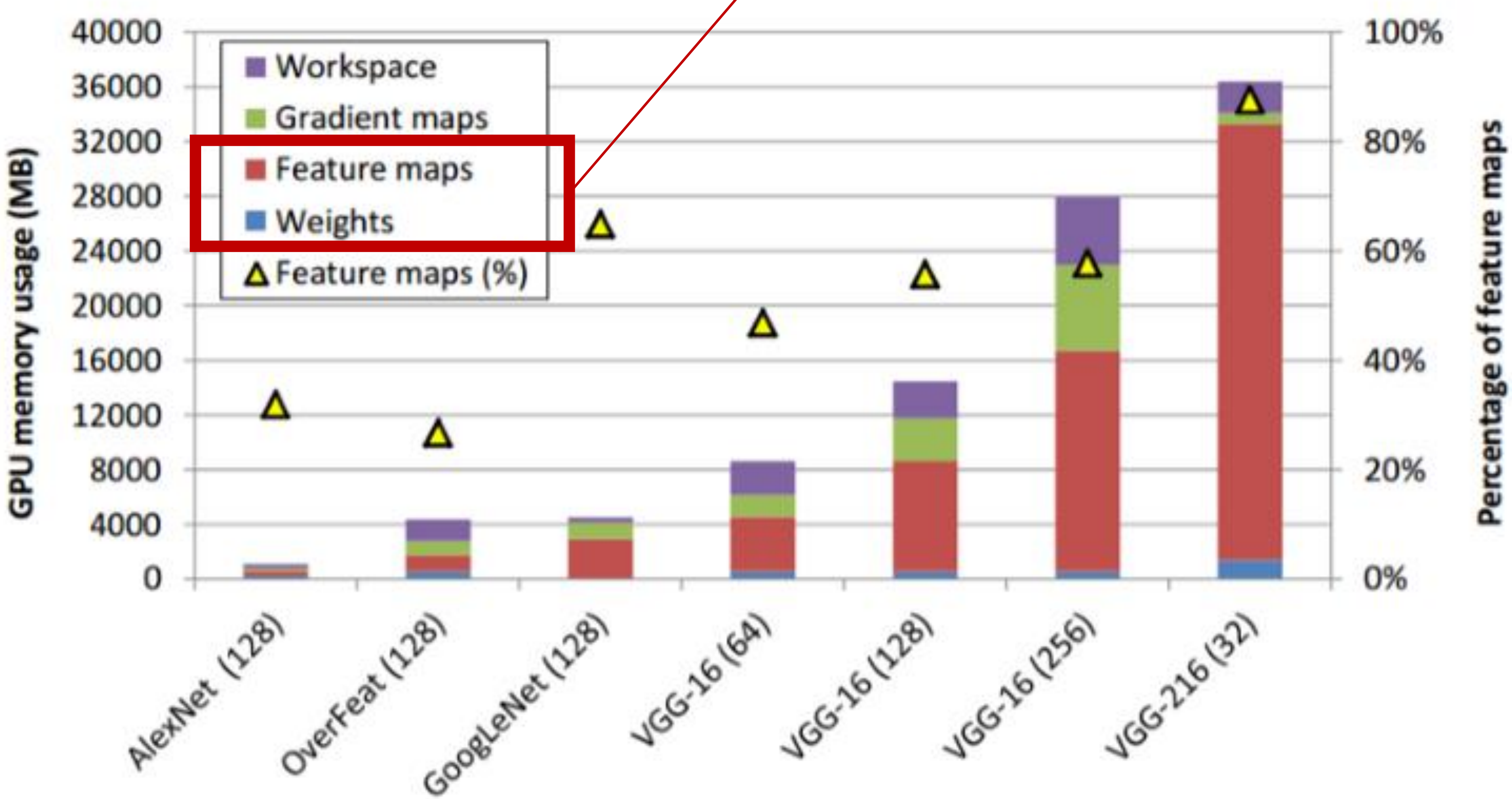


- **OOM! (Out-of-Memory)**
- **Filters cannot build-up larger components (Receptive field)**

OOM issue in DCNN-AE

- Intuition: with all convolution net, parameters are far less than MLP-like network, it should cost less memory

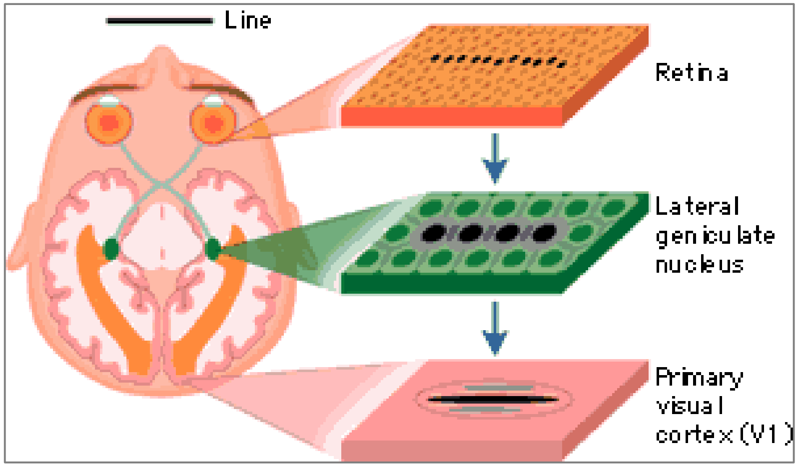
Actually, features map takes most of memory in the model

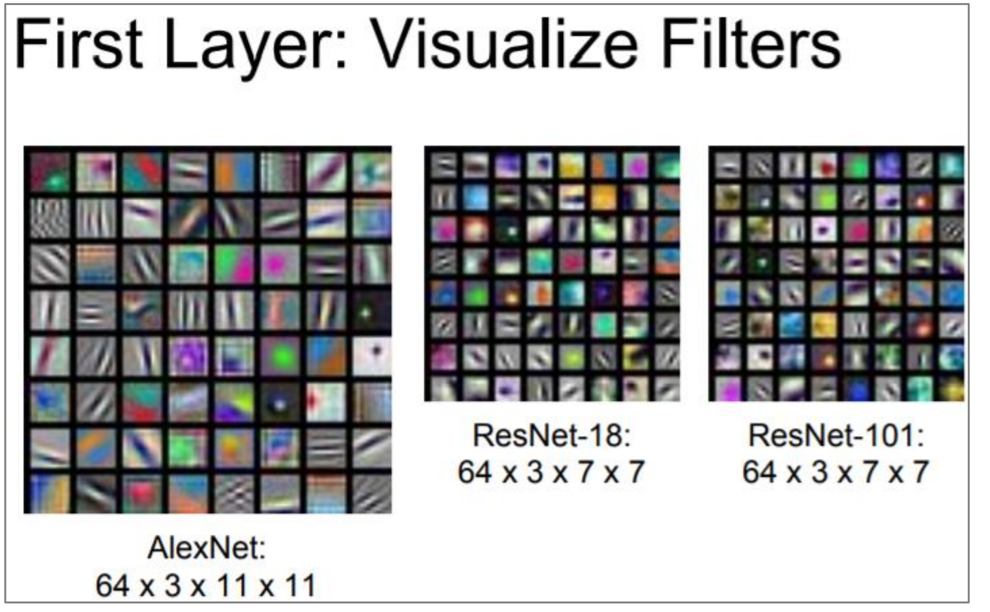


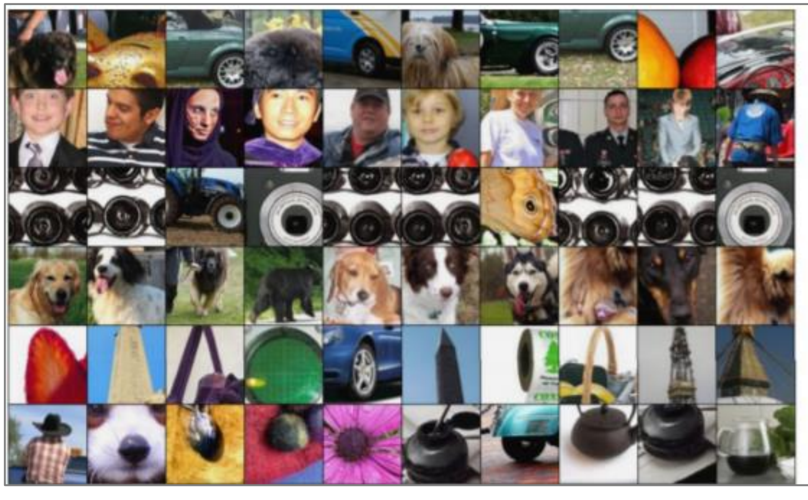
https://medium.com/syncedreview/how-to-train-a-very-large-and-deep-model-on-one-gpu-7b7edfe2d072

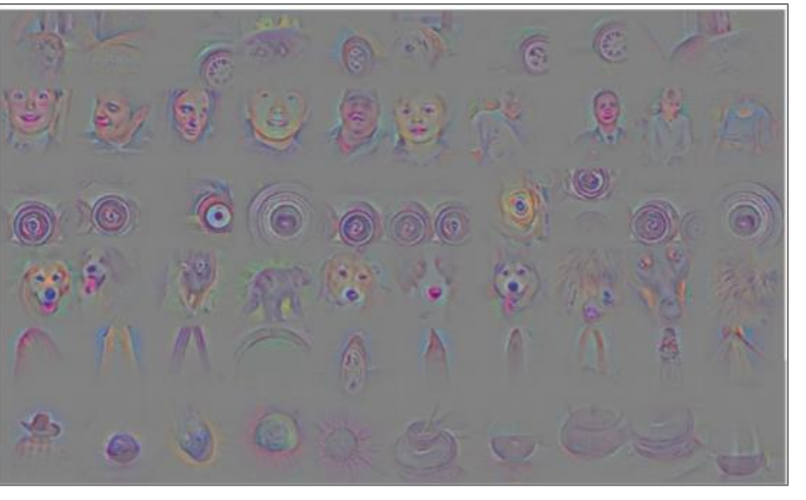
Receptive fields

From small to larger, from simple to complex





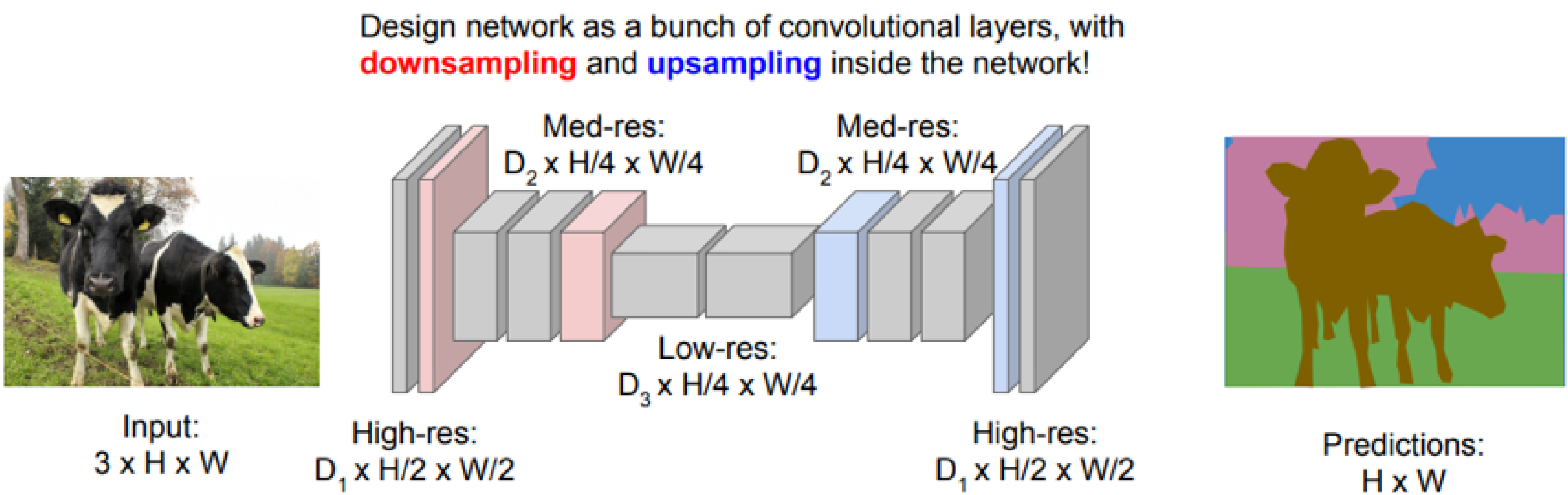




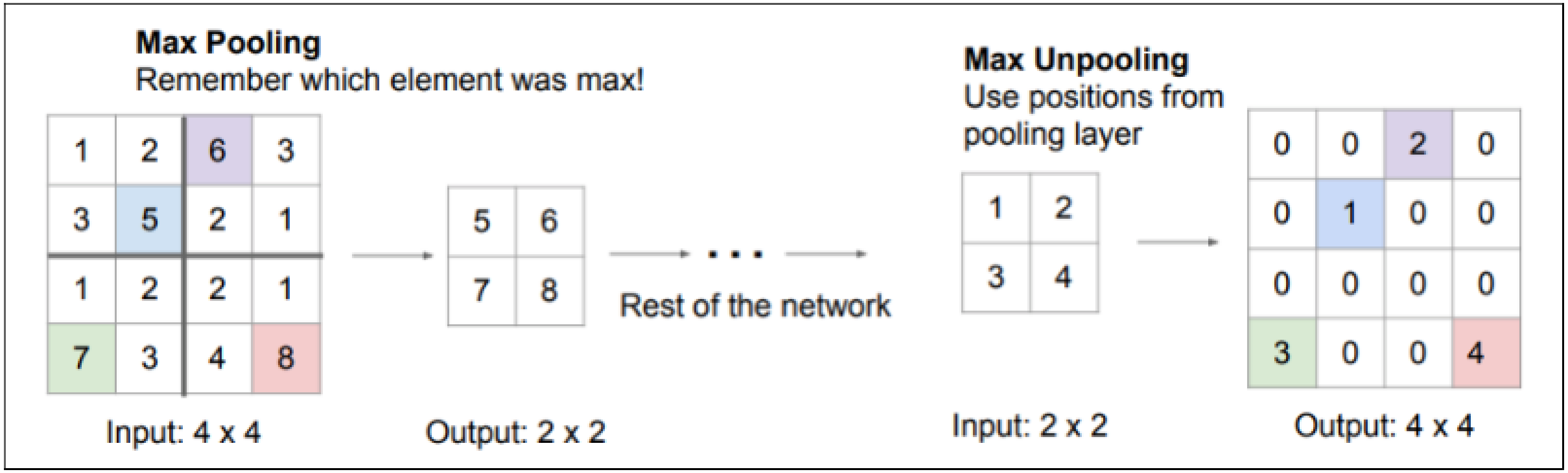
Model

Model.02 – DCNN – down-sampling + up-sampling

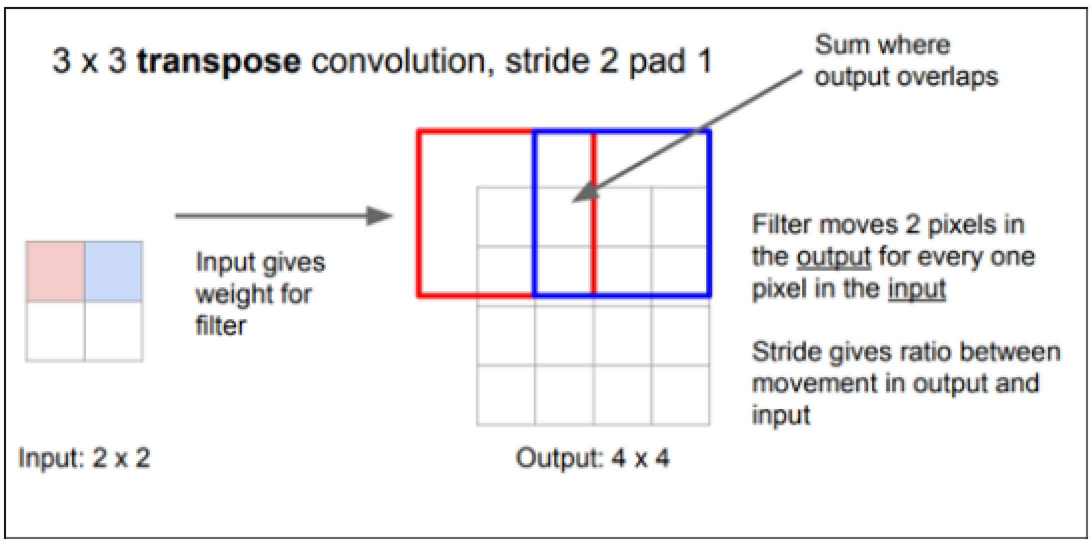
- Intuition
  - Input size = output size
  - Down-sample to encode features and upsampling to build map
  - Output: softmax with n-classes masks



Method to up-sampling images

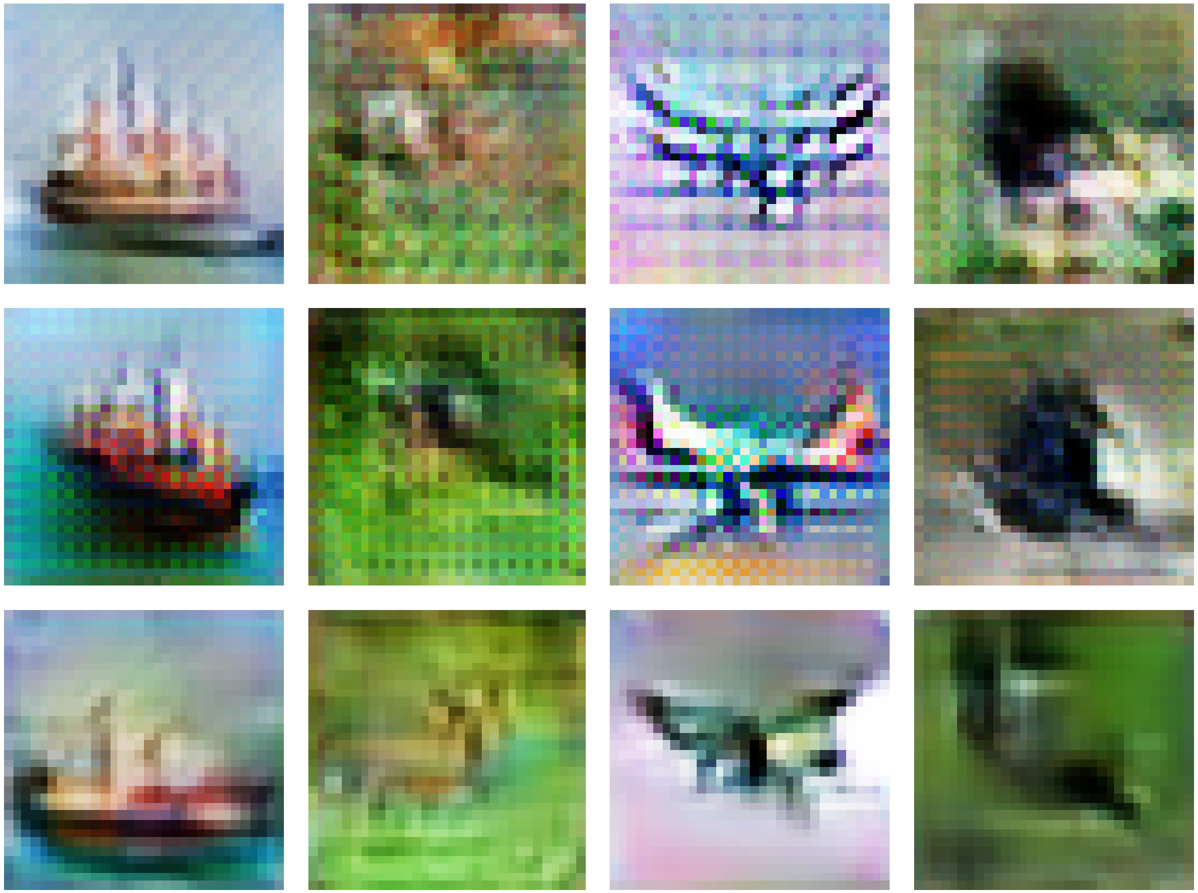


Directly unpool, no params (not learnable)



Unpool with parameters
> There is side effect: if overlap between filters, checkerboard artifact will emerge.

Checkerboard artifacts



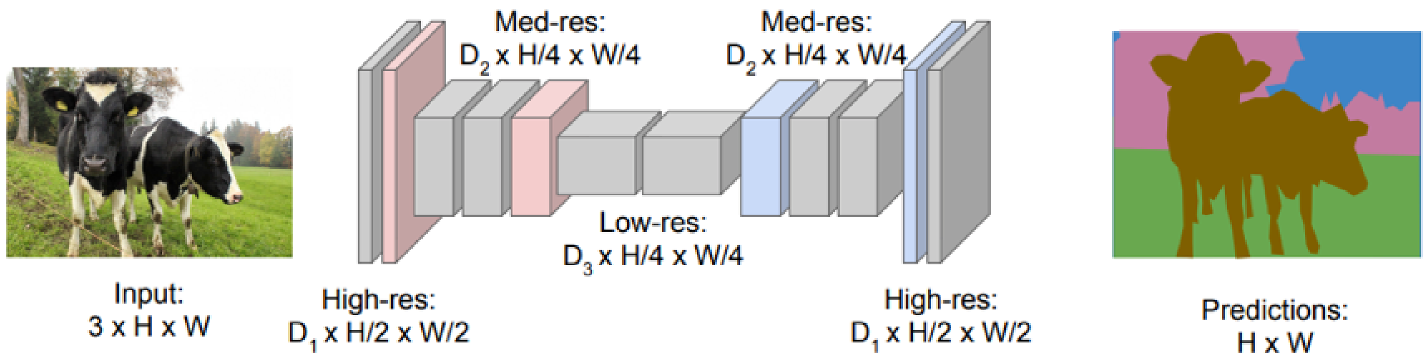
Deconv in last two layers.
Other layers use resize-convolution.
Artifacts of frequency 2 and 4.

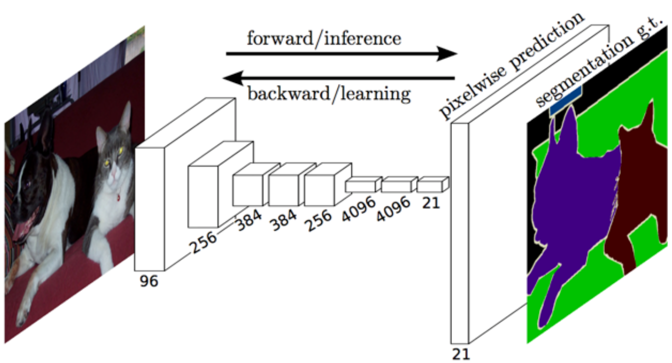
Deconv only in last layer.
Other layers use resize-convolution.
Artifacts of frequency 2.

All layers use resize-convolution.
No artifacts.

https://distill.pub/2016/deconv-checkerboard/

Problems in down-sampling – up-sampling FCN





Ground truth target

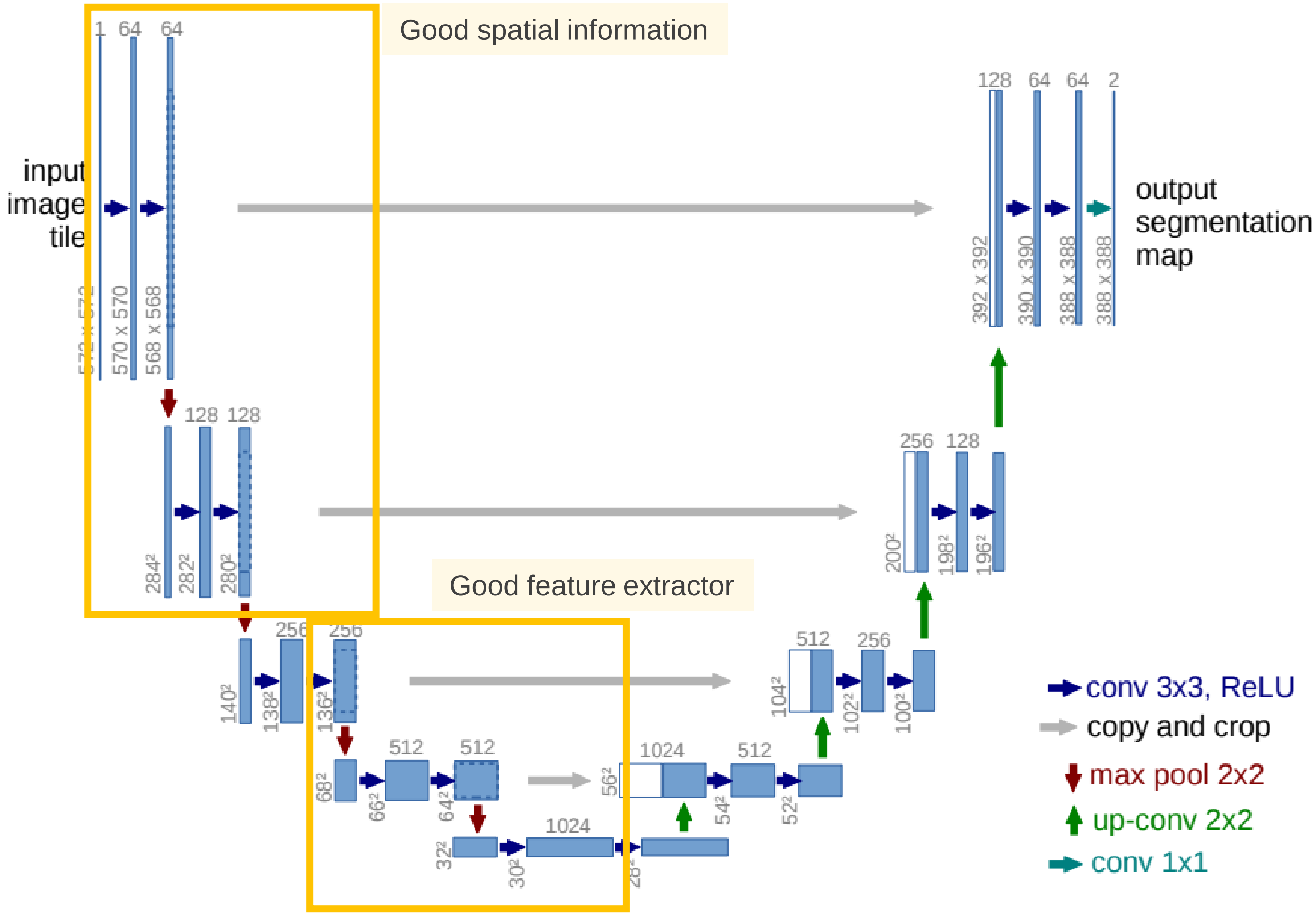
Predicted segmentation

- Spatial resolution loss
  - Cannot do fine-grained segmentations

Model

Model.03 – up/down-sampling + skip connections

- U-Net



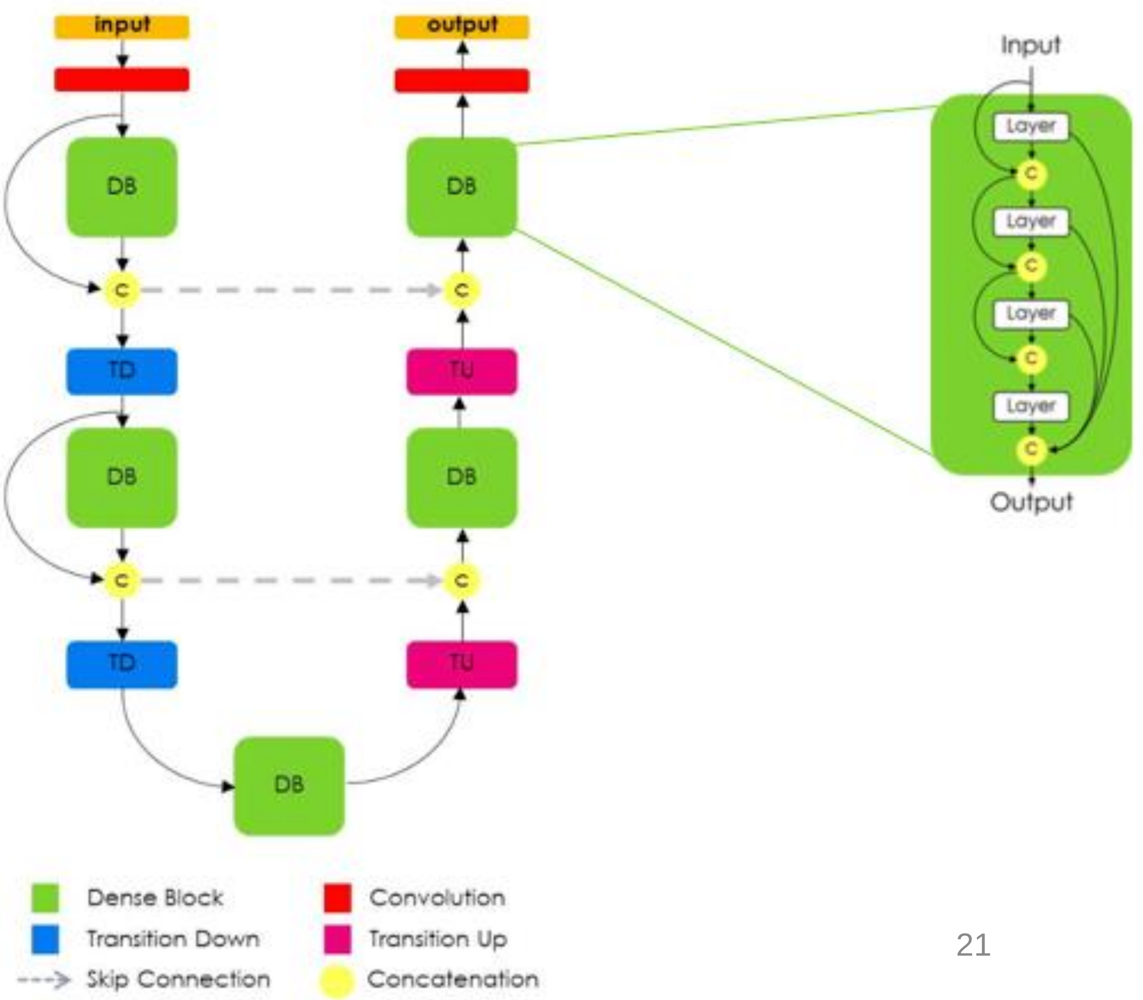
https://arxiv.org/pdf/1505.04597.pdf

https://github.com/zhixuhao/unet/blob/master/model.py

Model

Model.03 – up/down-sampling + skip connections

- Modified U-net: with different backbone or add residual/skips connections

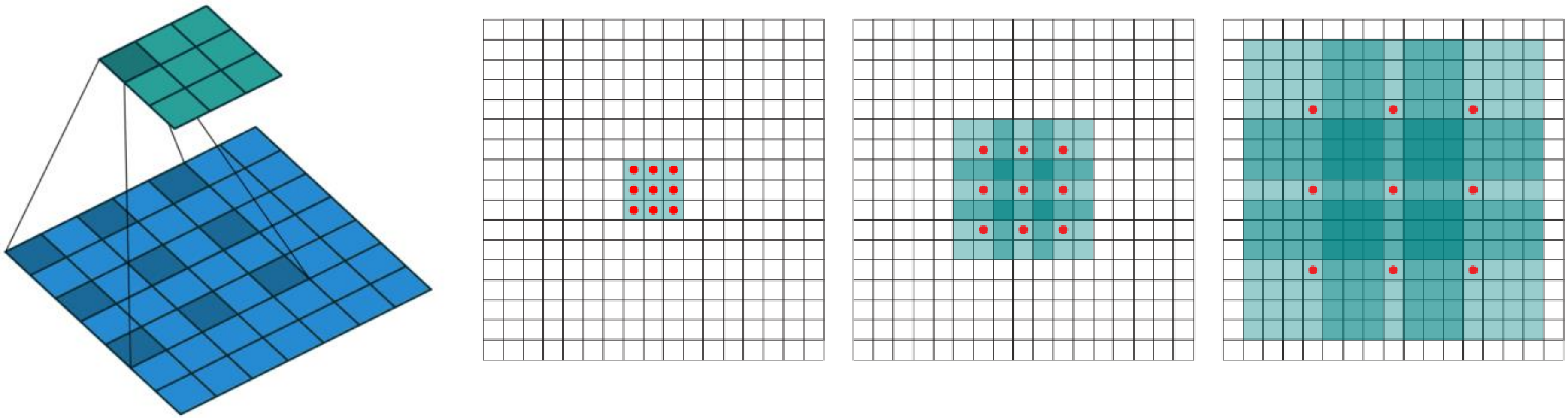


MoNuSeg/Keras_Unet_Stable.ipynb

Model

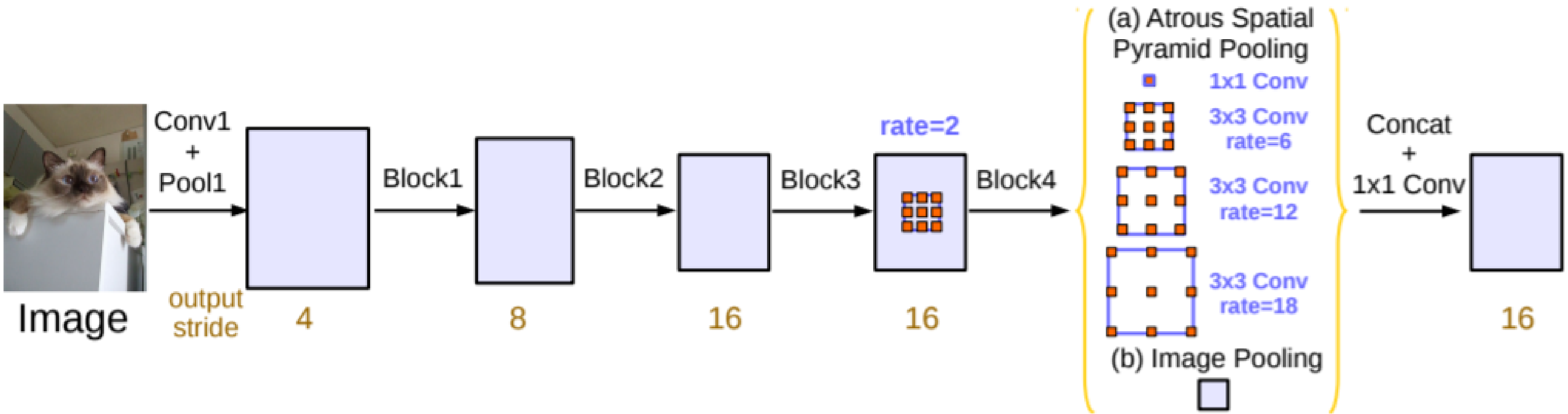
Model.04 – Dilate convolution module

- Convolution multiple times and maxpool/stride
  - Receptive field problem
    - Do we need to use multiple pooling / strides to get larger receptive fields?

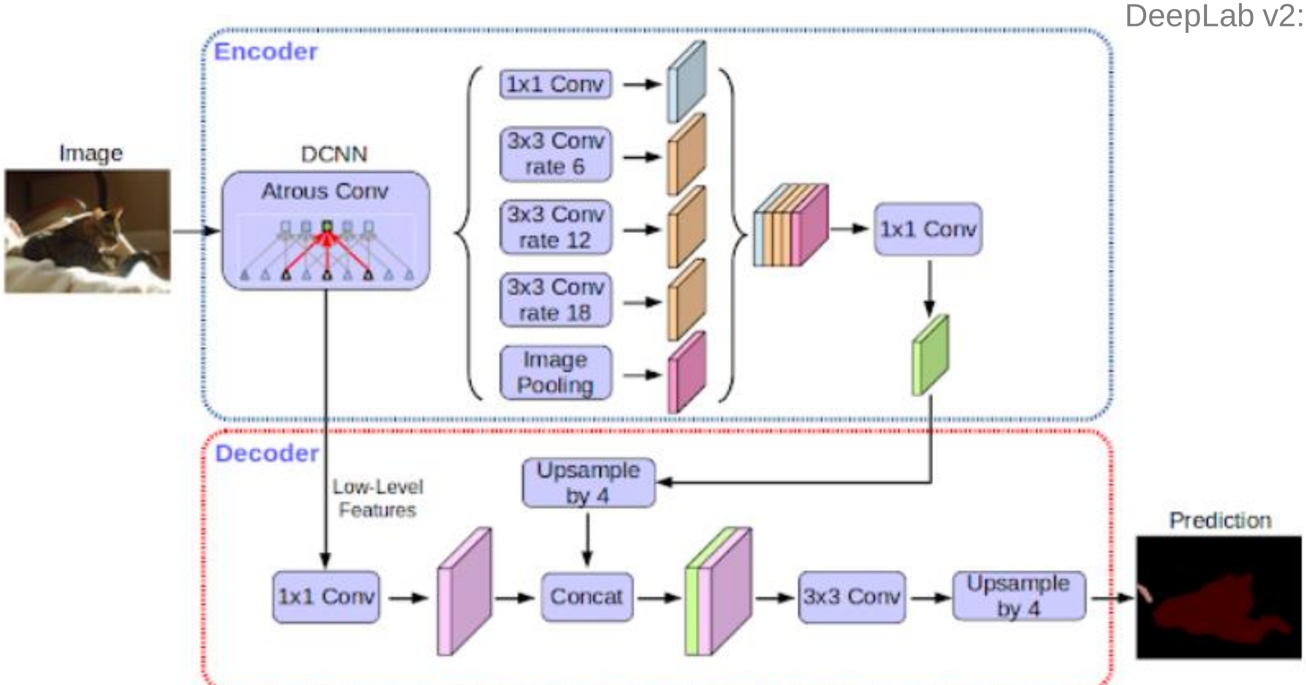


Model

Model.04 – Dilate convolution module



DeepLab v2: https://arxiv.org/pdf/1606.00915.pdf



DeepLab v3+: https://arxiv.org/pdf/1802.03039.pdf

Loss functions

- Do we only have cross-entropy?
  - potential problems
    - if your various classes have unbalanced representation in the image, as training can be dominated by the most prevalent class

$$- y \log \hat{y} - (1 - y) \log(1 - \hat{y})$$

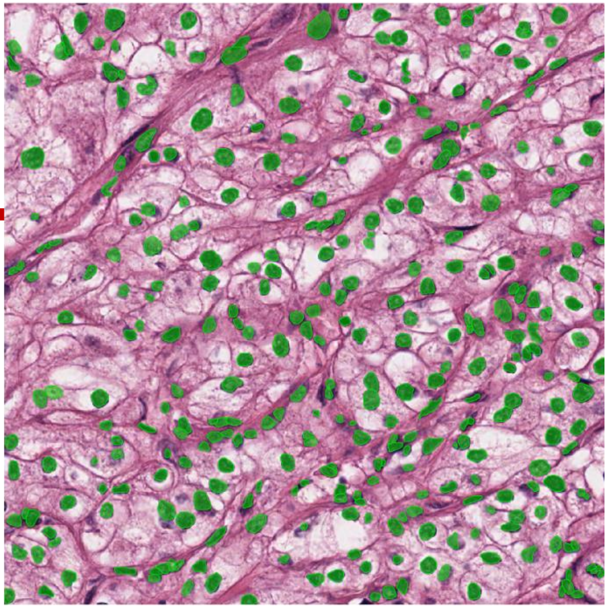
[Fully Convolutional Networks for Semantic Segmentation]

  - Simple method to overcome classes imbalance problem
    - Just evaluate the intersection of union (Dice-loss)
    - It will become very important issue for bio-medical images
    - Jaccard Index

$$Dice = \frac{2 \cdot |mask \cap prediction|}{|mask| + |prediction|}$$

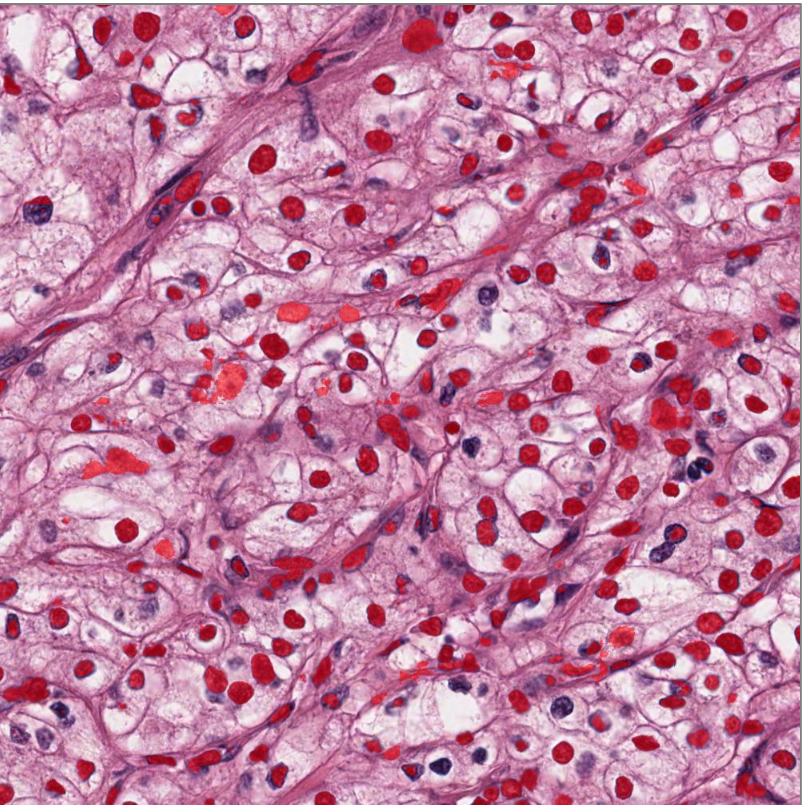
[Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations]

Loss functions

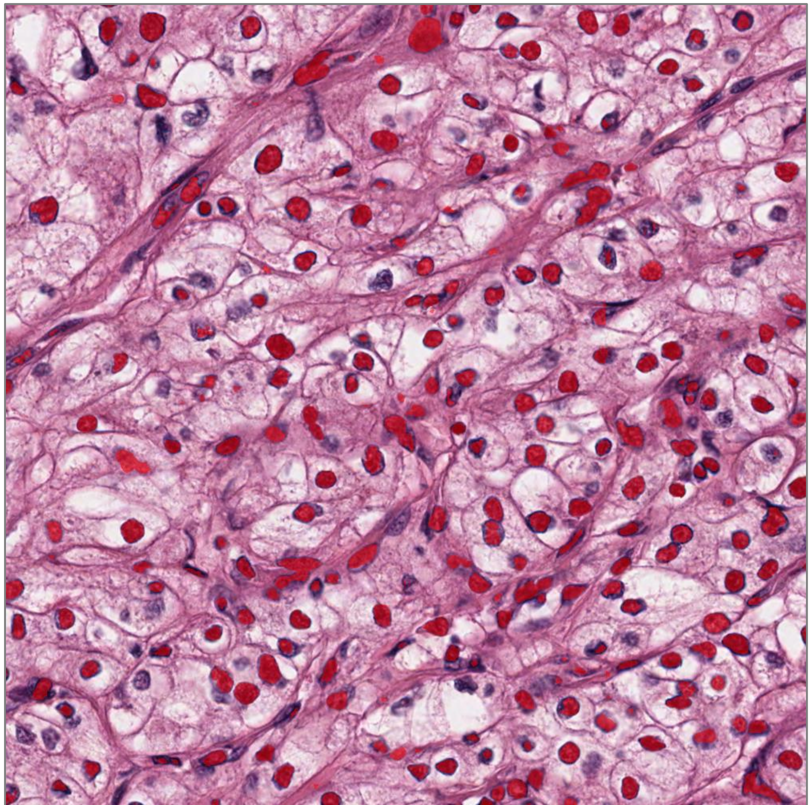


GT

X-loss



X-loss + Jaccard loss

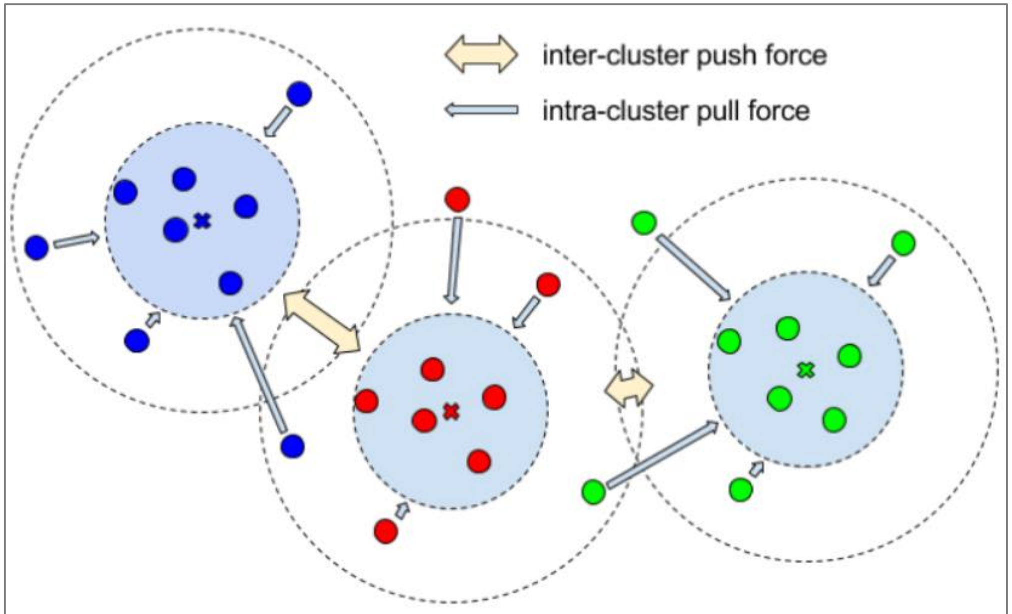


Codes on Multi-Organ-Nuceli-Segmentation

- https://monuseg.grand-challenge.org/
- Instance segmentation task

- Current path
  - Semantic segmentation → Instance segmentation
    - Unet version
    - DeeplabV3+ version
    - ---
      - Loss function modification with metric learning (discriminative loss)
      - Semantic Instance Segmentation via Deep Metric Learning
      - Recurrent Pixel Embedding for Instance Grouping
  - Direct instance segmentation (with Mask-RCNN)

Loss function modification with metric learning



visualizing the contrastive loss.

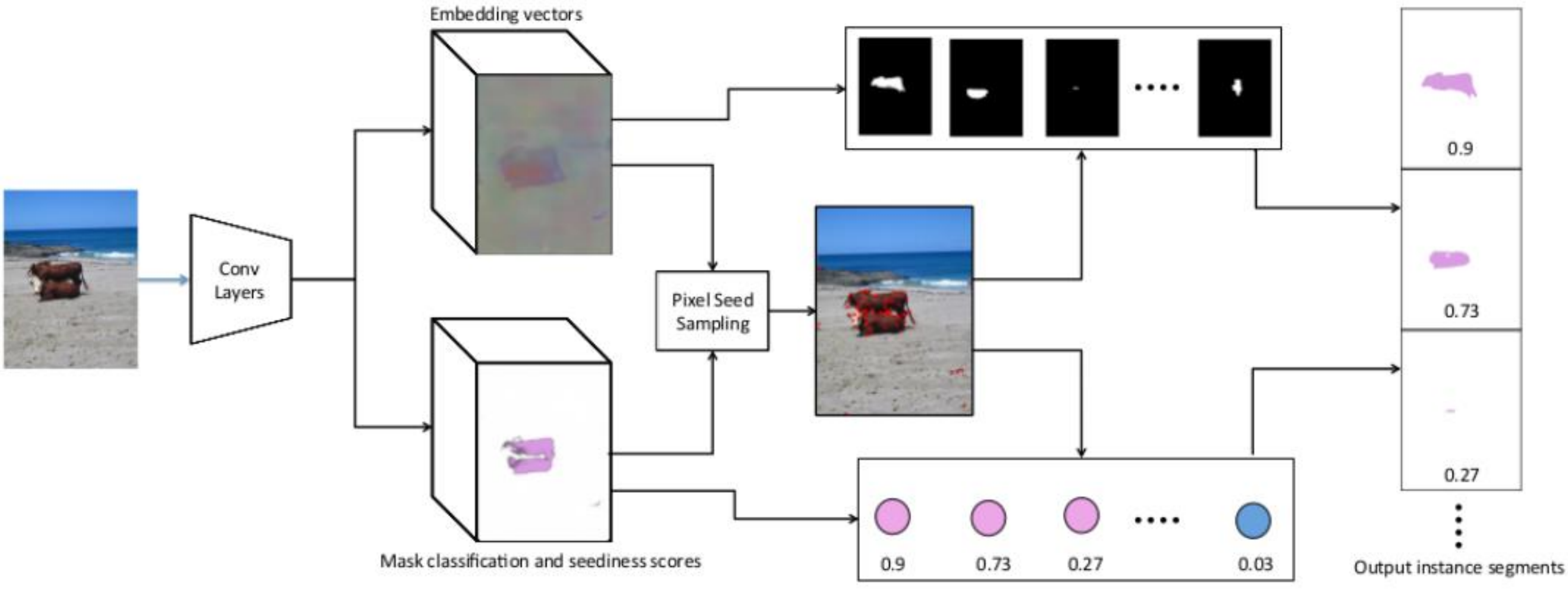
$$L_{var} = \frac{1}{C} \sum_{c=1}^{C} \frac{1}{N_c} \sum_{i=1}^{N_c} \left[\|\mu_c - x_i\| - \delta_{\mathrm{v}} \right]_+^2 \qquad (1)$$

$$L_{dist} = \frac{1}{C(C-1)} \sum_{\substack{c_A=1 \\ c_A \neq c_B}}^{C} \sum_{c_B=1}^{C} \left[2\delta_{\mathrm{d}} - \|\mu_{c_A} - \mu_{c_B}\| \right]_+^2 \qquad (2)$$

$$L_{reg} = \frac{1}{C} \sum_{c=1}^{C} \|\mu_c\| \qquad (3)$$

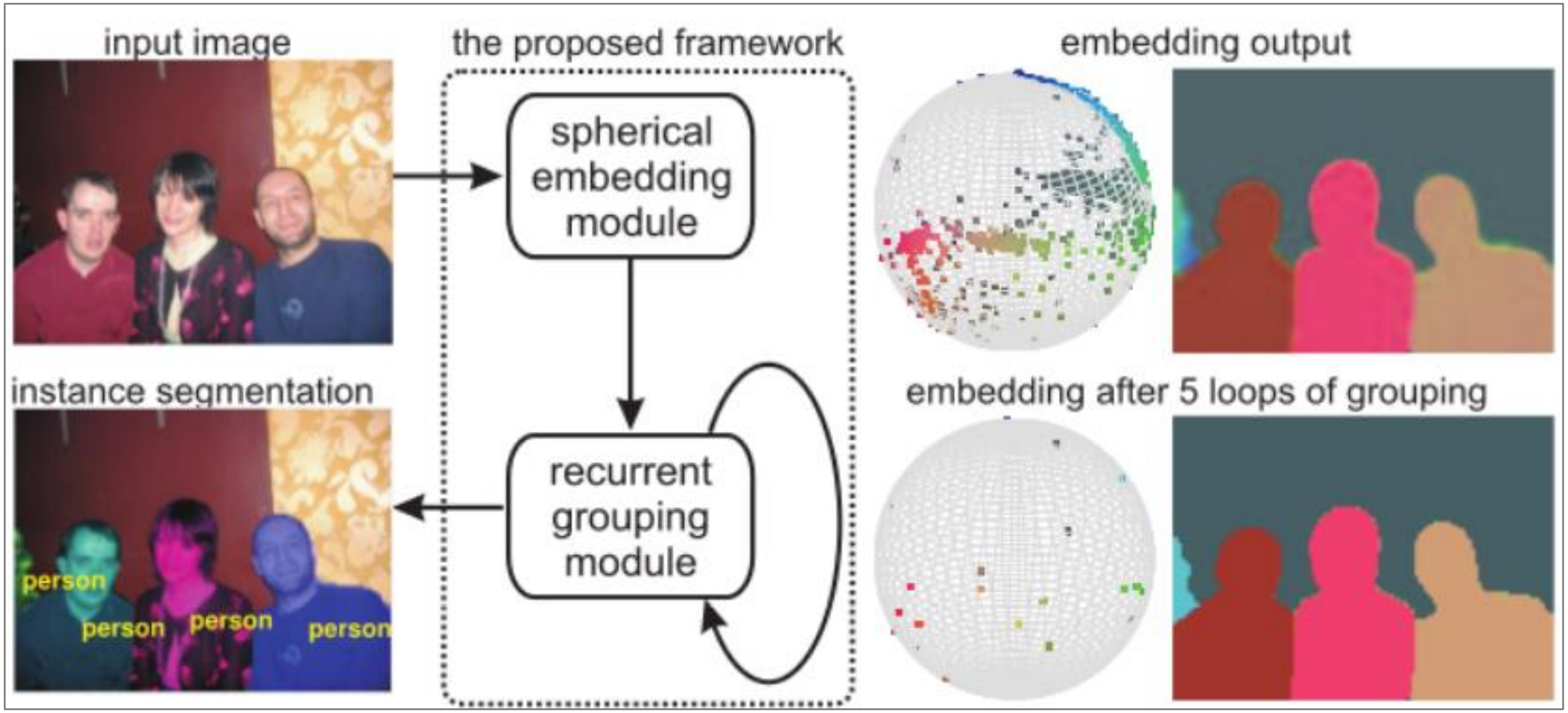
$$L = \alpha \cdot L_{var} + \beta \cdot L_{dist} + \gamma \cdot L_{reg} \qquad (4)$$

Semantic Instance Segmentation via Deep Metric Learning

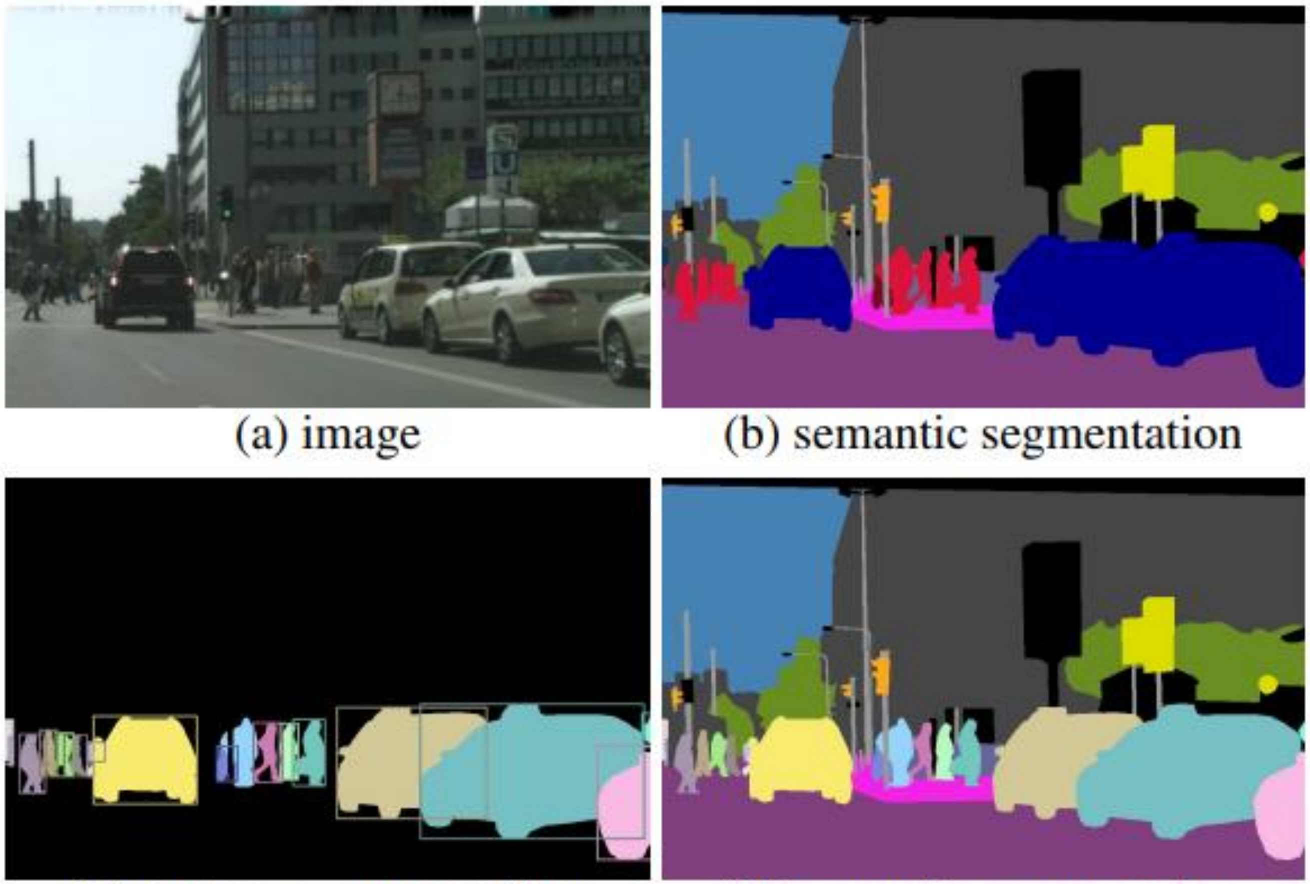


Network architecture proposed in Semantic Instance Segmentation via Deep Metric Learning

Recurrent Pixel Embedding for Instance Grouping



New tasks in computer vision: Pantropic segmentation labeling all the things!



(a) image (b) semantic segmentation

(c) instance segmentation (d) panoptic segmentation

https://arxiv.org/pdf/1801.00868.pdf