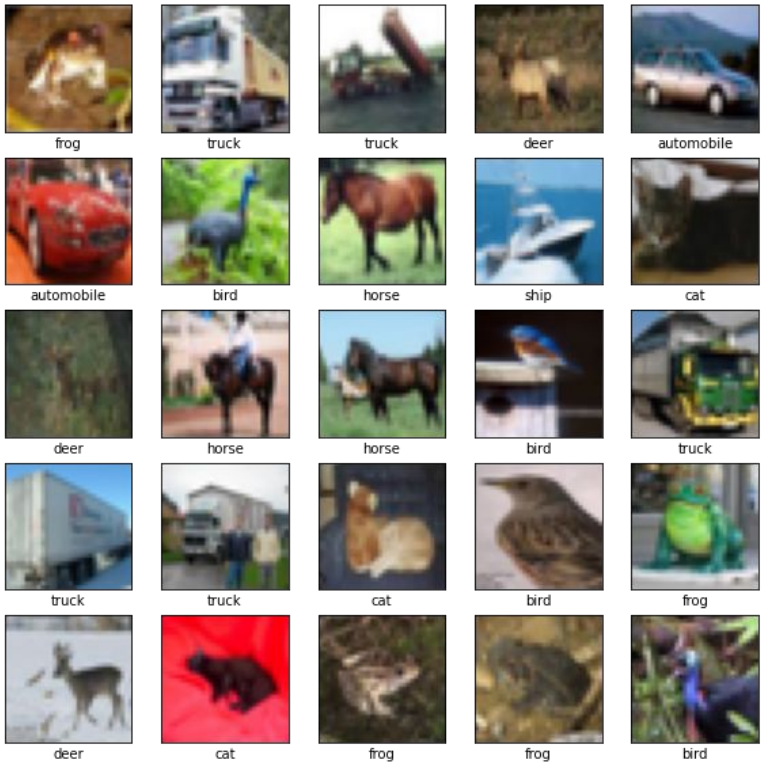
Deep Learning for ECE
EECE-580G

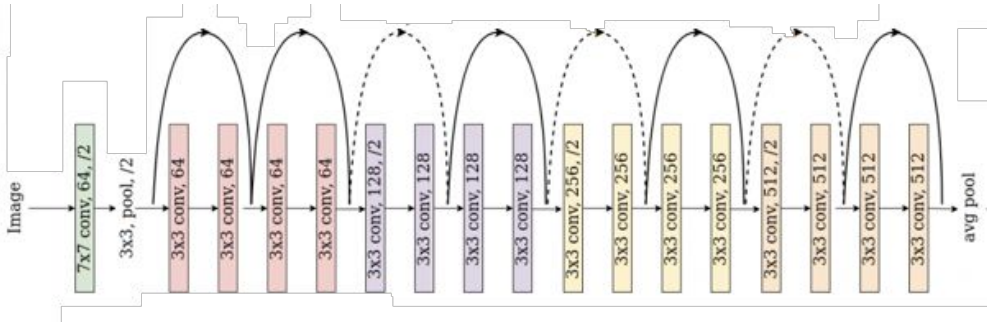
Image Segmentation

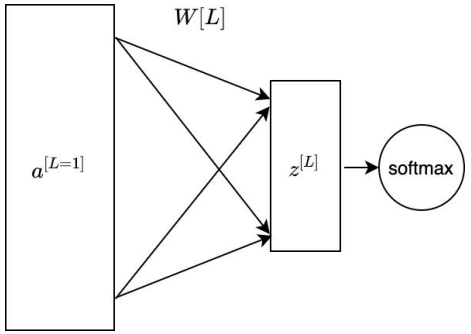
Recap

Single label classification



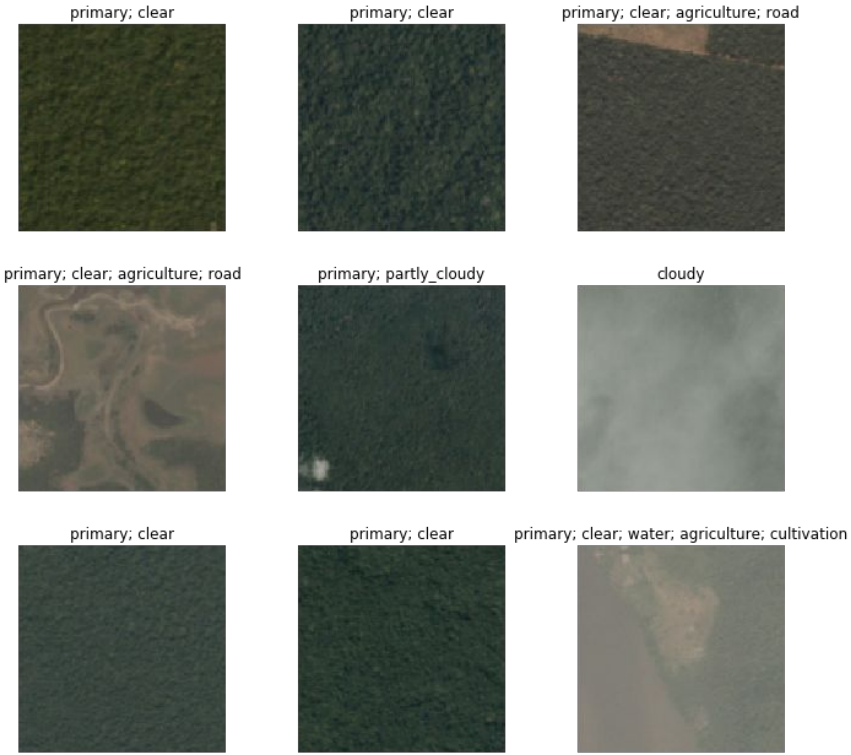
Single label classification





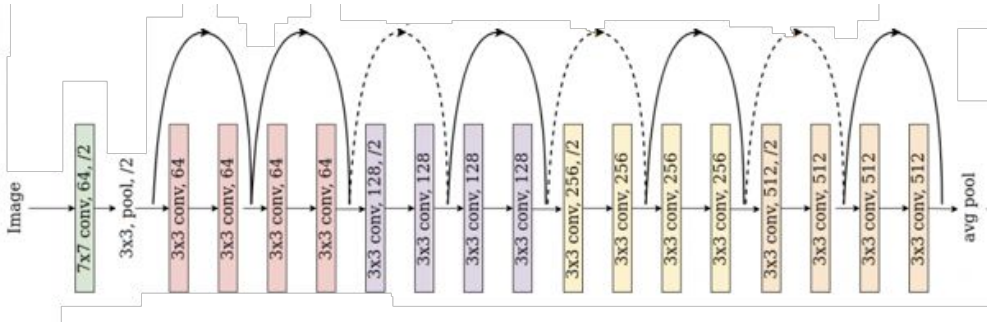
$$\text{output}_i = \frac{e^{z_i^{[L]}}}{\sum_{j=1}^{K} e^{z_j^{[L]}}}$$

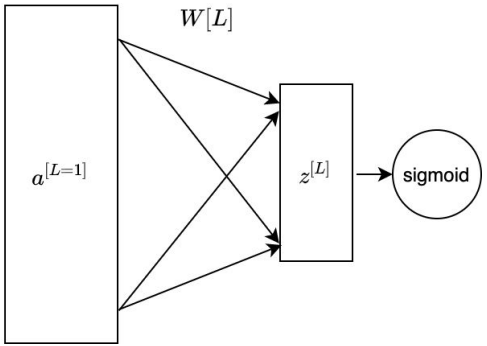
Multi label classification



https://www.kaggle.com/c/planet-understanding-the-amazon-from-space

Multi label classification

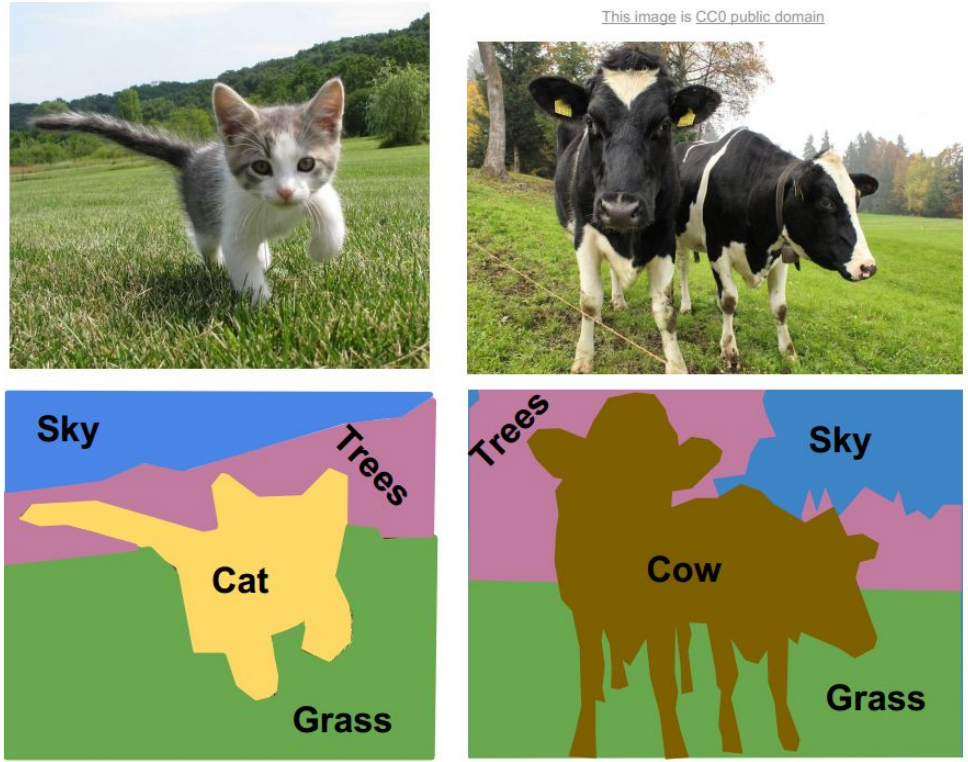




$$\text{output}_i = \frac{1}{1 + e^{-z_i^{[L]}}}$$

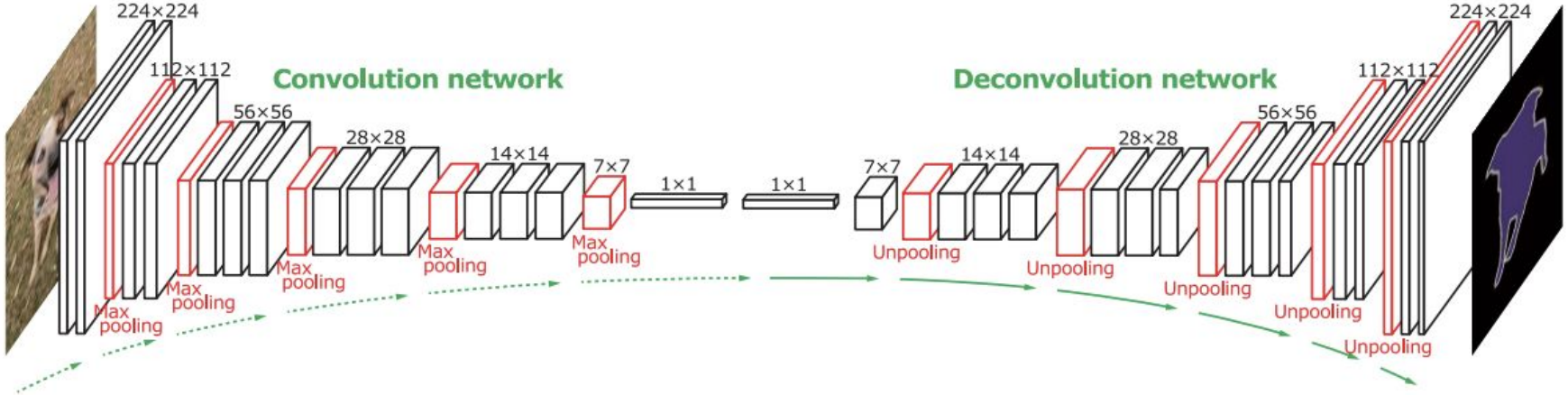
Semantic segmentation

Segmentation = Pixel level classification



Semantic segmentation

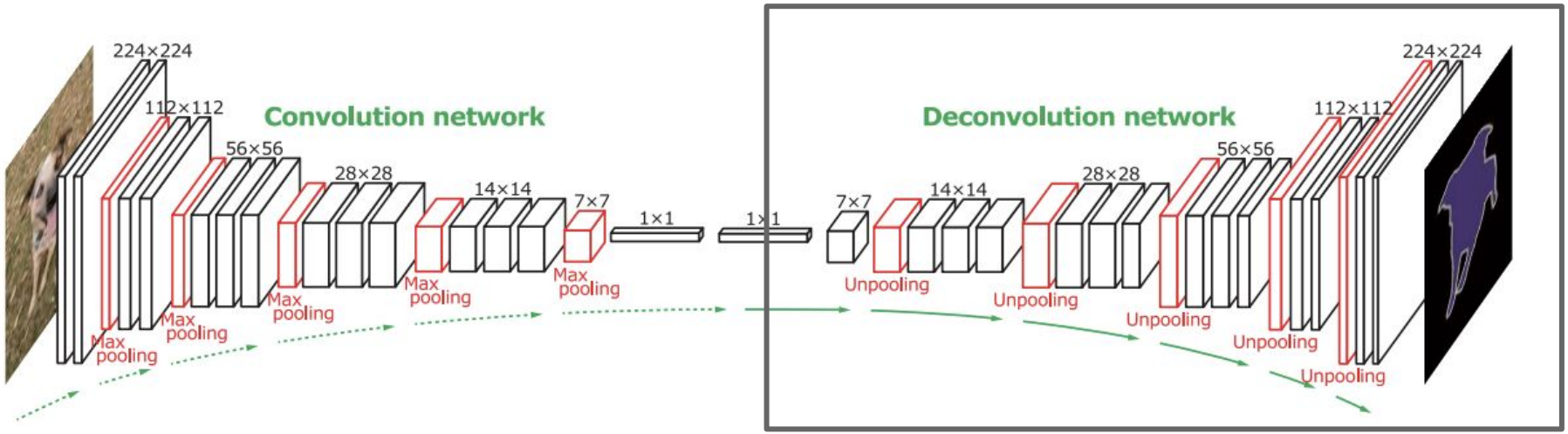
- Output of CNN of shape: **HxWxK**
- Where K is the number of classes



Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation." Proceedings of the IEEE international conference on computer vision. 2015.

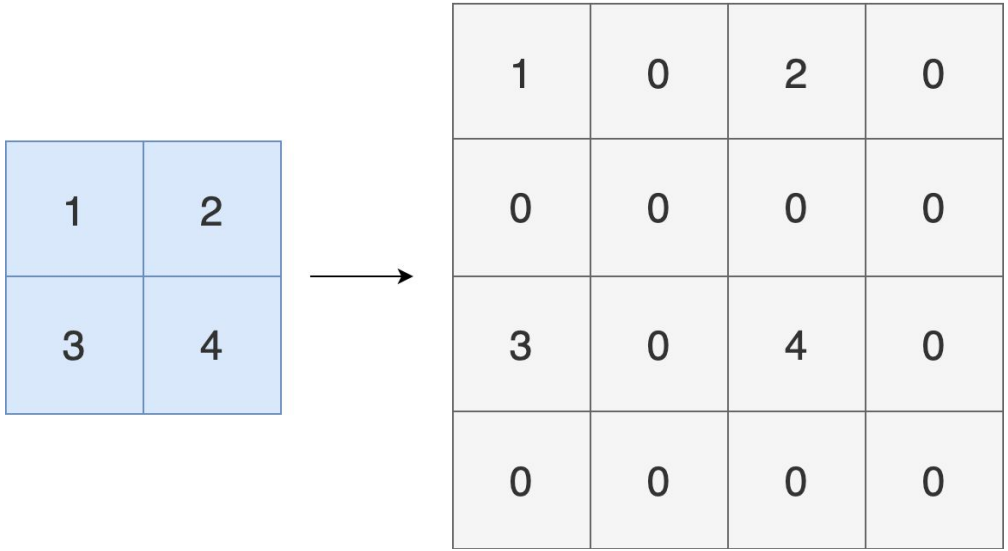
Semantic segmentation

- Output of CNN of shape: **HxWxK**
- Where K is the number of classes



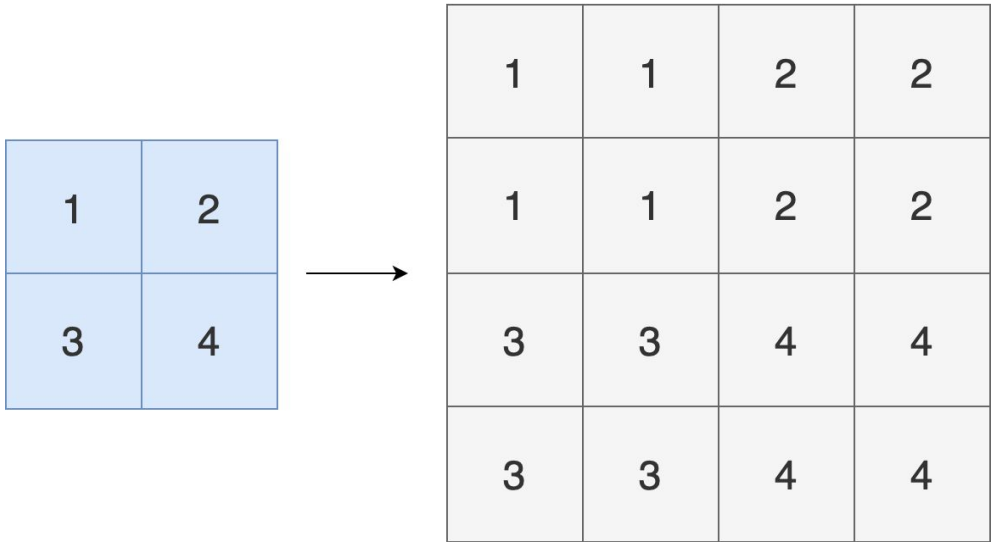
Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation." Proceedings of the IEEE international conference on computer vision. 2015.

Upsampling



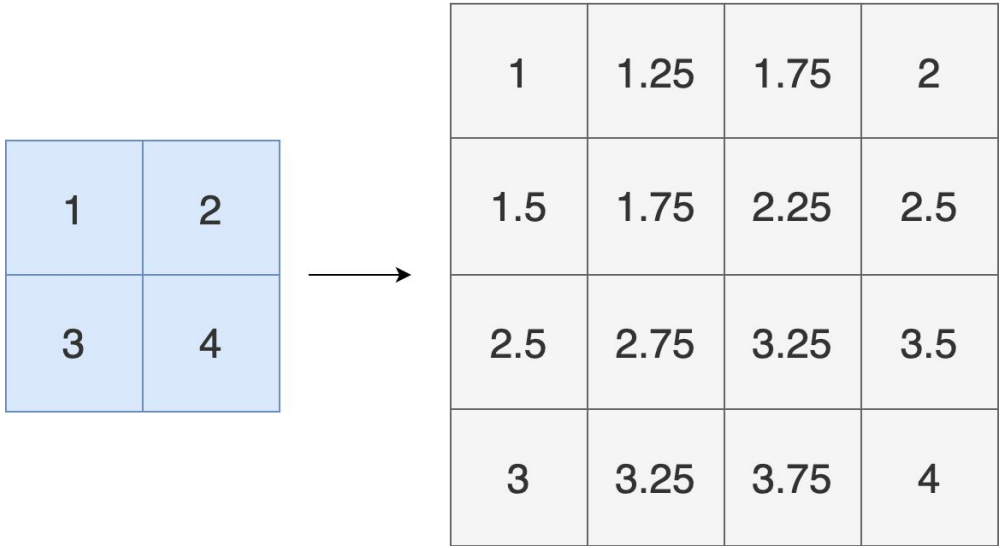
“Bed of nails” upsampling

Upsampling



Nearest Neighbor upsampling

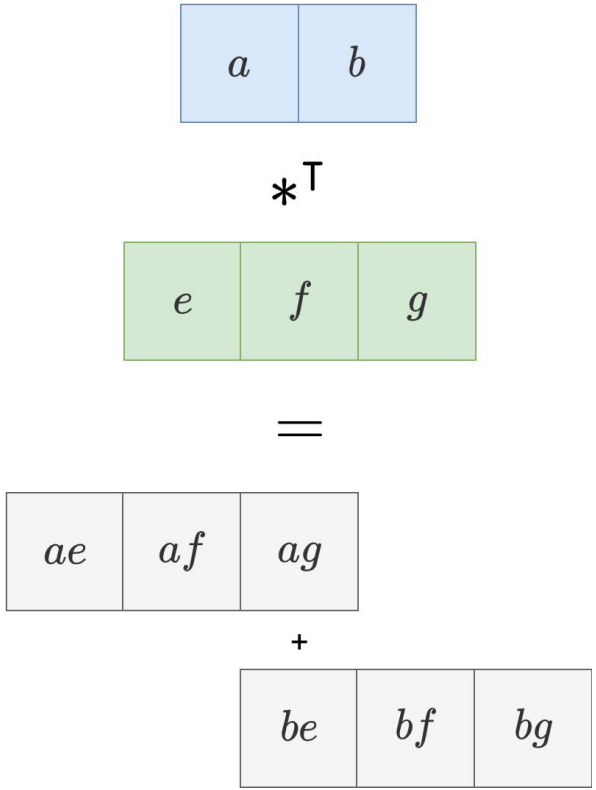
Upsampling



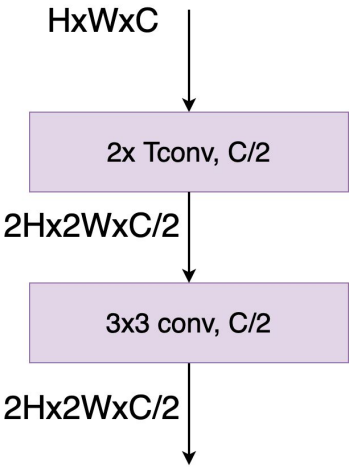
Bilinear upsampling

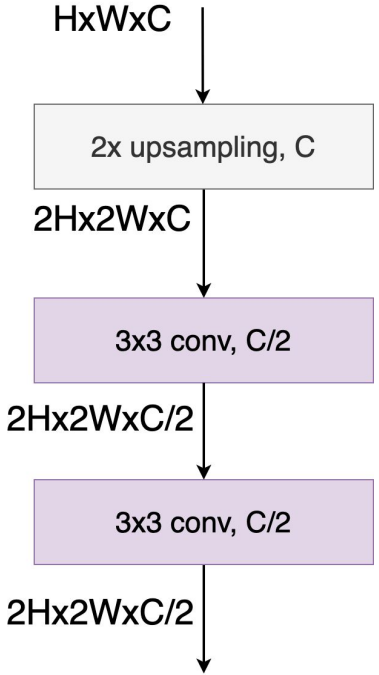
Transposed Convolution

- **Transposed Convolution** = Learnable upsampling
- Already implemented in TF >>>>
- Also called: Deconvolution - Upconvolution - Backward strided convolution - Fractionally strided convolution



Upsampling blocks





Other tricks:

- Subpixel convolution + ICNR initialization

Losses and metrics

Pixel Cross-Entropy and Focal Loss

- Simply flatten w.r.t. **HxW** and compute the average (or total) cross-entropy loss for each pixel

$$L(y, \hat{y}) = -y \cdot \log(\hat{y})$$

- **Focal Loss:** In image segmentation: class imbalance is usually a problem (i.e. a lot of “background” pixels, small objects…)

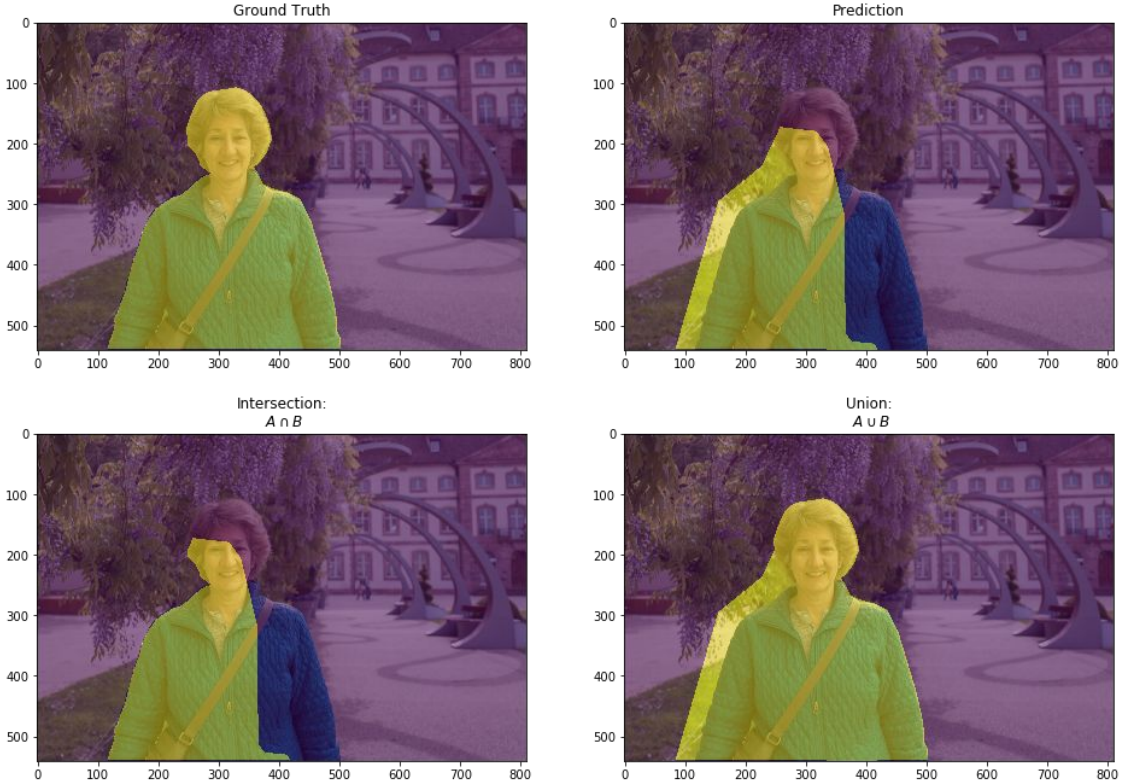
$$L(y, \hat{y}) = -y \cdot (1 - \hat{y})^{\gamma} \cdot \log(\hat{y})$$

- \gamma = 2 is a good starting point but need to be tuned

Pixel Accuracy

- Simply flatten w.r.t. **HxW** and compute the rate of classifying a pixel correctly
- (nothing special here)

Intersection over union (IoU or Jaccard)

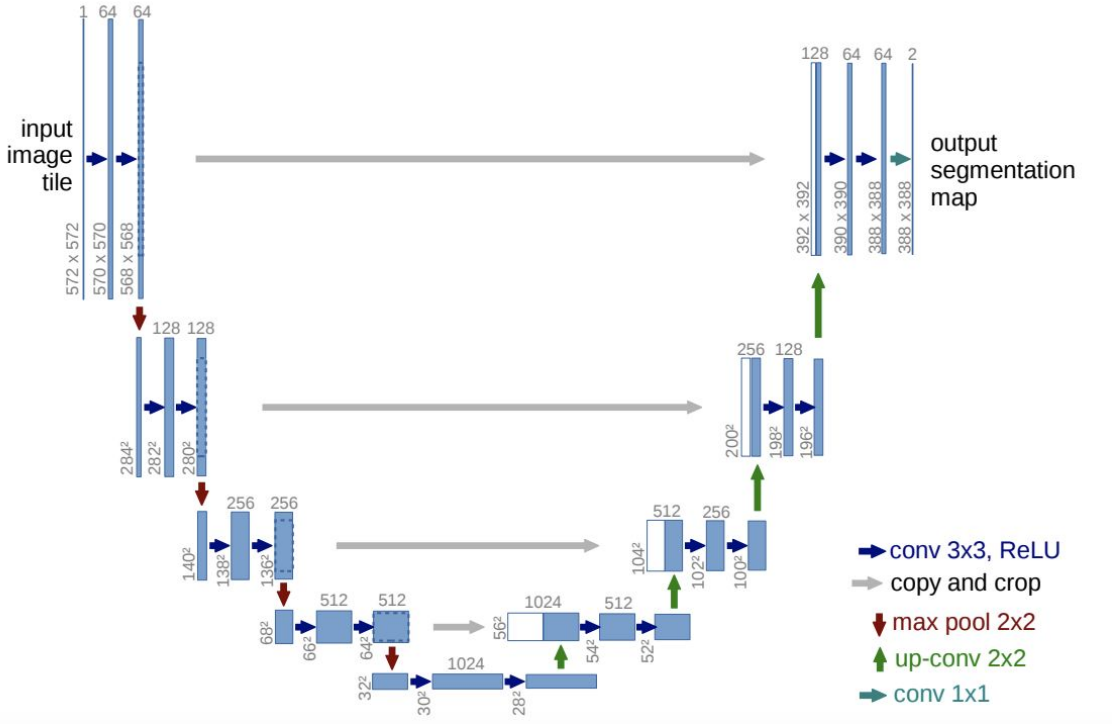


$$IoU = \frac{target \cap prediction}{target \cup prediction}$$

https://www.jeremyjordan.me/evaluating-image-segmentation-models/

U-Net

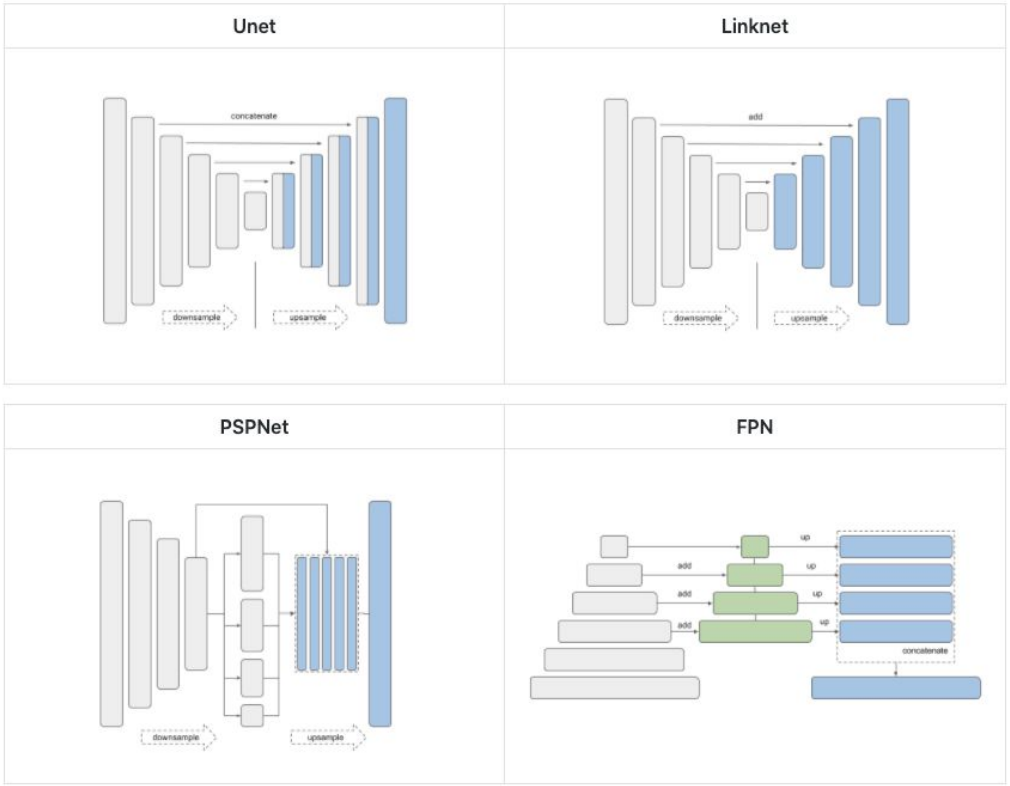
U-Net



- "Skip" connections through resolutions sometimes called "cross connections"
- Cross connections are done as concat
- Bottleneck layer is not 1x1
- The U-Net structure generalizes to other **CNN architectures** + **Transfer Learning**

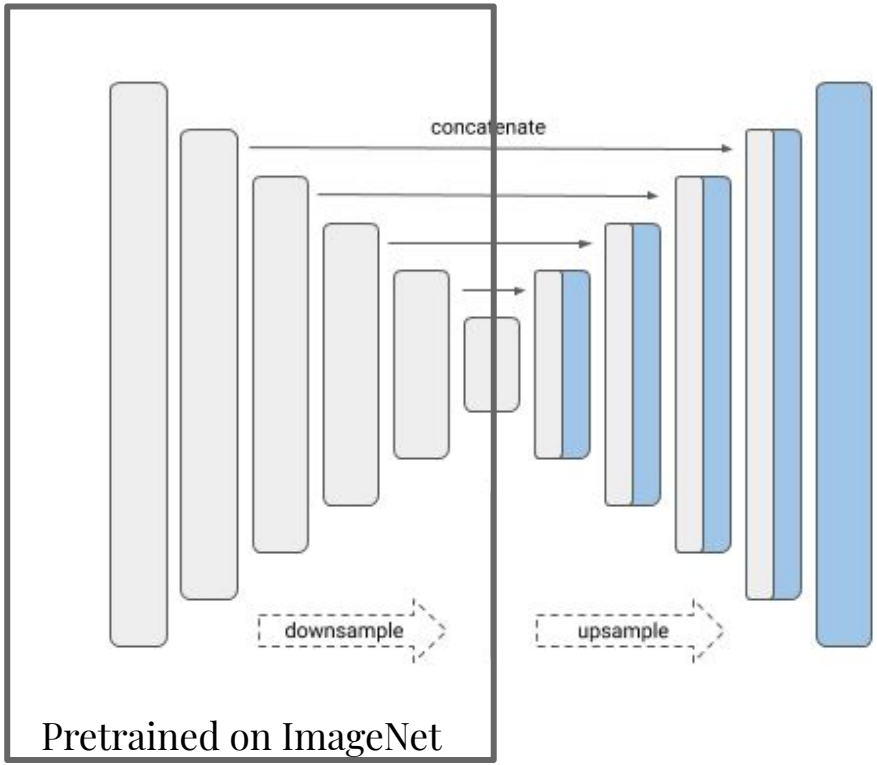
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

Other up-sampling architectures

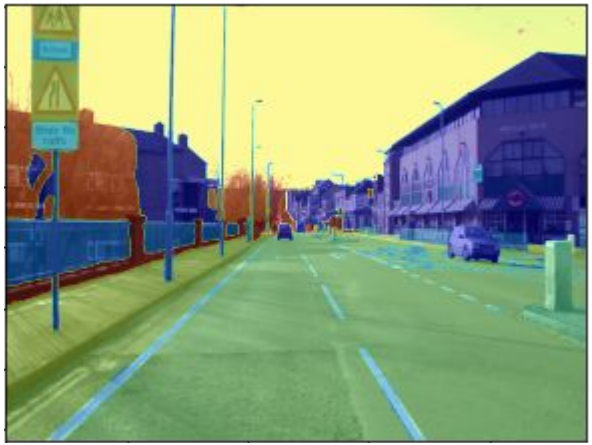


https://github.com/qubvel/segmentation_models

Transfer Learning + Segmentation



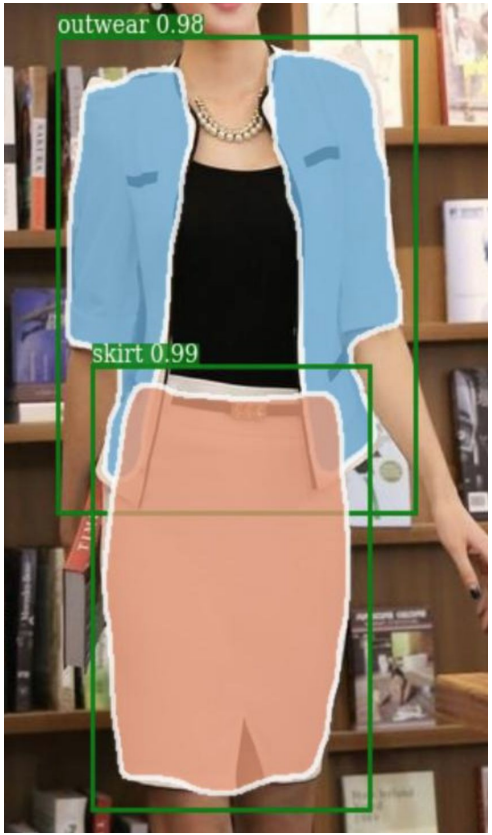
Example - Self driving cars



Conclusions

Conclusions

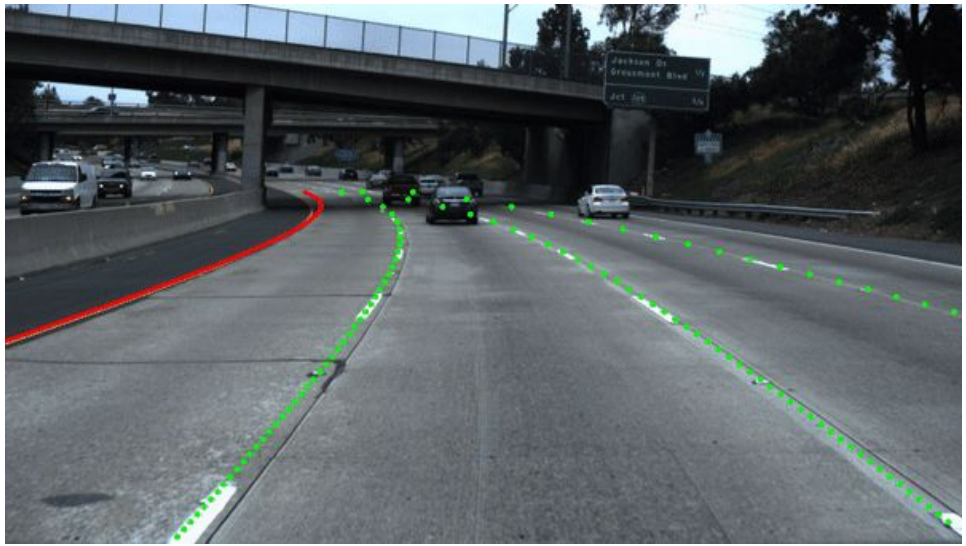
- Segmentation as a pixel classification problem
- Can be a multi label pixel classification problem if the **masks are overlapping**



https://github.com/switchablenorms/DeepFashion2/issues/26

Conclusions

- Doesn’t have to segment all pixels of the image
- E.g. Lane detection



Van Gansbeke, Wouter, et al. "End-to-end lane detection through differentiable least-squares fitting." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2019.

TU-Simple Benchmark dataset

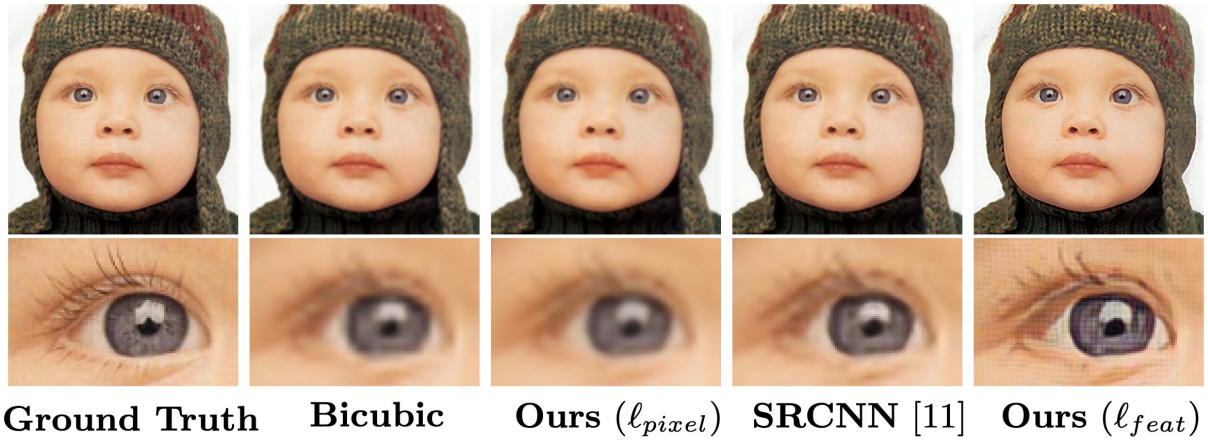
Conclusions

- Another related task = Instance segmentation = detect and segment (car A, car B, etc.)

Arnab, Anurag, and Philip HS Torr. "Pixelwise instance segmentation with a dynamically instantiated network." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

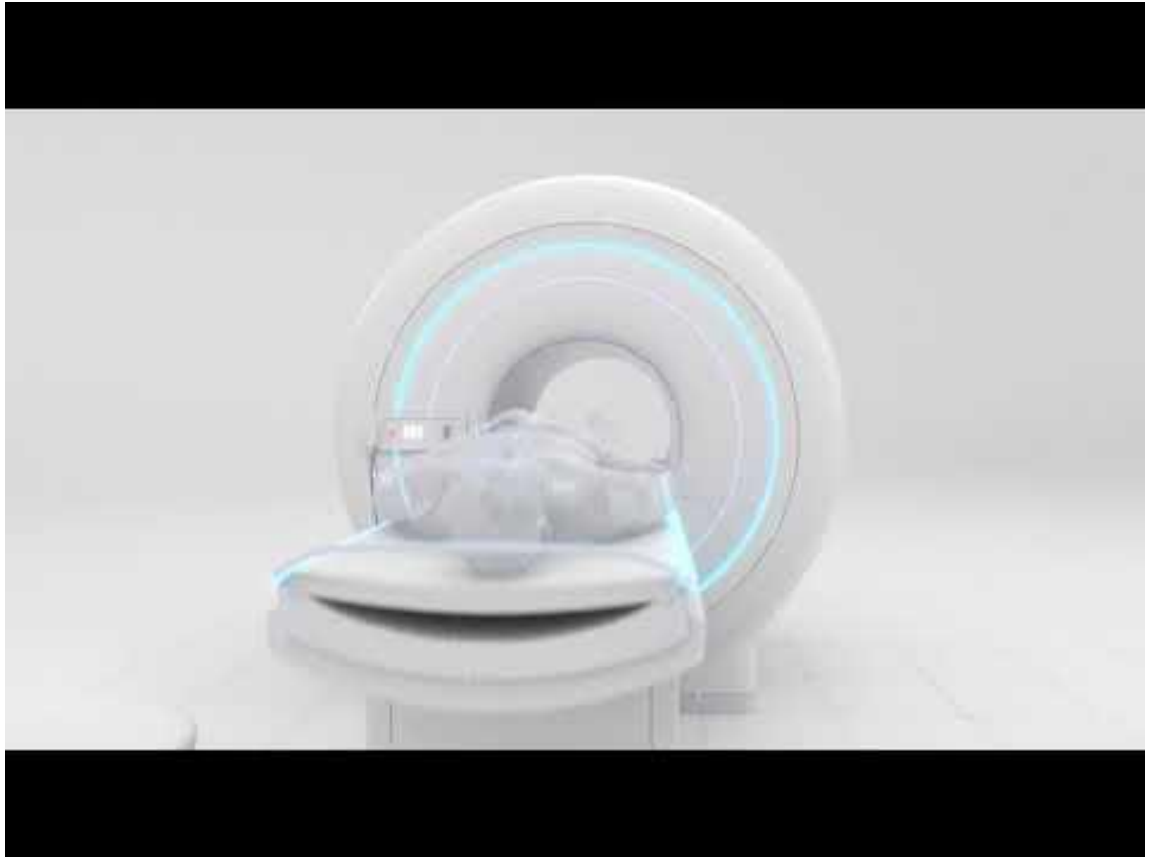
Conclusions

- In practice, U-Net works well for segmentation
- Progressive resizing is commonly used
- Can be used for other tasks: **image super resolution**, **image restoration**, ...



Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." *European conference on computer vision*. Springer, Cham, 2016.

Conclusions



End