Image Segmentation Using Deep Learning: A Survey

Nupur Yadav

San Jose State University

Deep Learning

Introduction

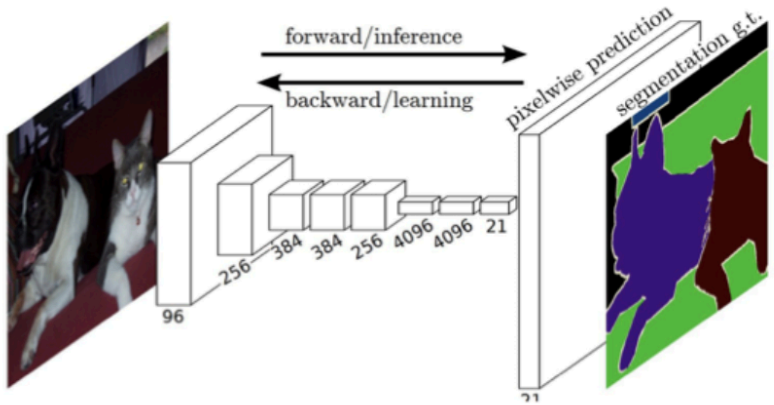
- Image segmentation helps us understand the content of the image and is a very important topic in image processing and computer vision.
- Wide variety of applications such as image compression, scene understanding, locating objects in satellite images.
- Many deep learning models for image segmentation have also emerged.

DL-based image segmentation models

1. Fully Convolutional networks
2. Convolutional Models with Graphical Models
3. Encoder-Decoder Based Models
4. Multi-Scale and Pyramid Network Based Models
5. R-CNN Based Models (for Instance Segmentation)
6. Dilated Convolutional Models and DeepLab Family
7. Recurrent Neural Network Based Models
8. Attention-Based Models
9. Generative Models and Adversarial Training
10. CNN Models with Active Contour Models

Fully Convolutional networks

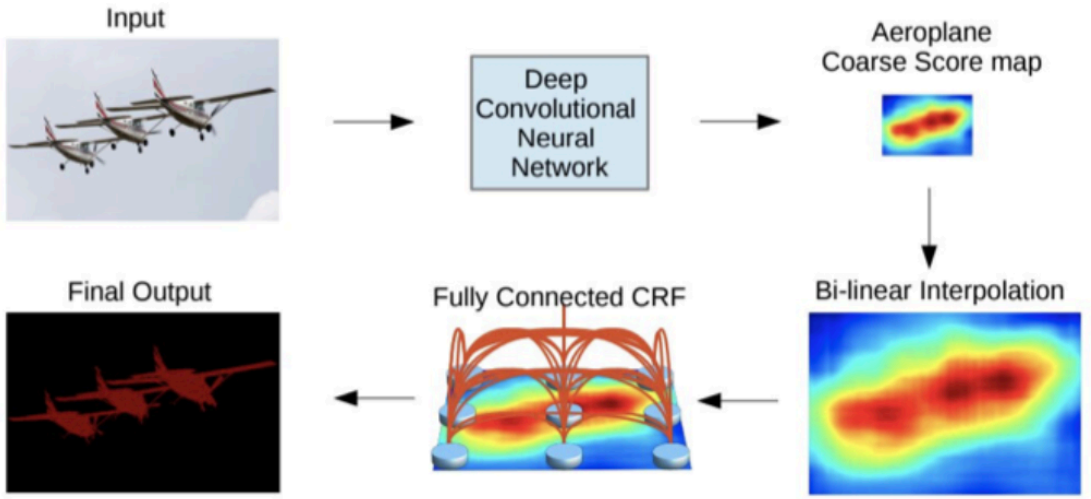
- Consists of only convolutional layers where features are extracted by convolving a kernel/filter of weights.
- It takes any image of arbitrary size and produces a segmentation map of the same size.
- It uses skip connections which allows feature maps from final layers to be up-sampled and fused with features maps of earlier layers.
- And helps the model to produce a very accurate and detailed segmentation by combining the semantic information from the deep and coarse layers with the appearance information from the shallow and fine layers.
- Few notable applications of FCNs are iris segmentation and brain tumor segmentation.



A fully convolutional image segmentation network. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Convolutional Models with Graphical Models

- Deep CNNs have poor localization property, which means the responses at the final layers of CNNs are insufficiently localized to produce accurate object segmentation
- The responses at the final CNN layer were then combined to a fully connected Conditional Random Field (CRF).
- Achieved a higher accuracy rate than the previous FCN methods.

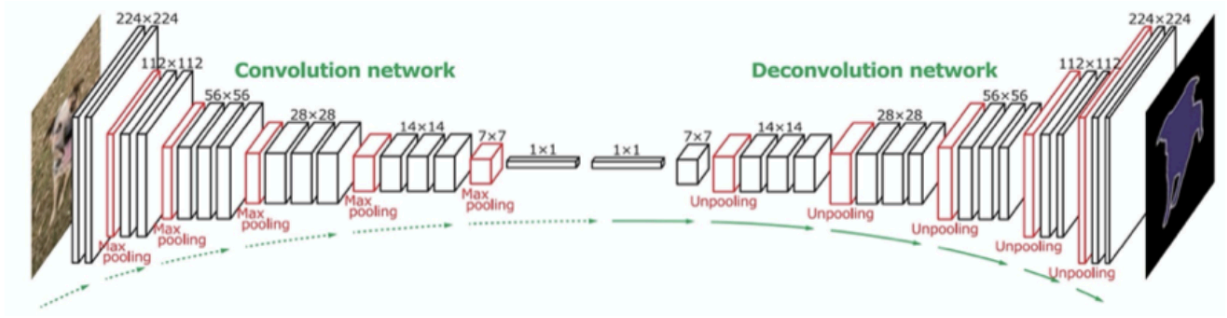


CNN + CRF model. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Encoder-Decoder Based Models

A. Encoder-Decoder Models for General Segmentation:

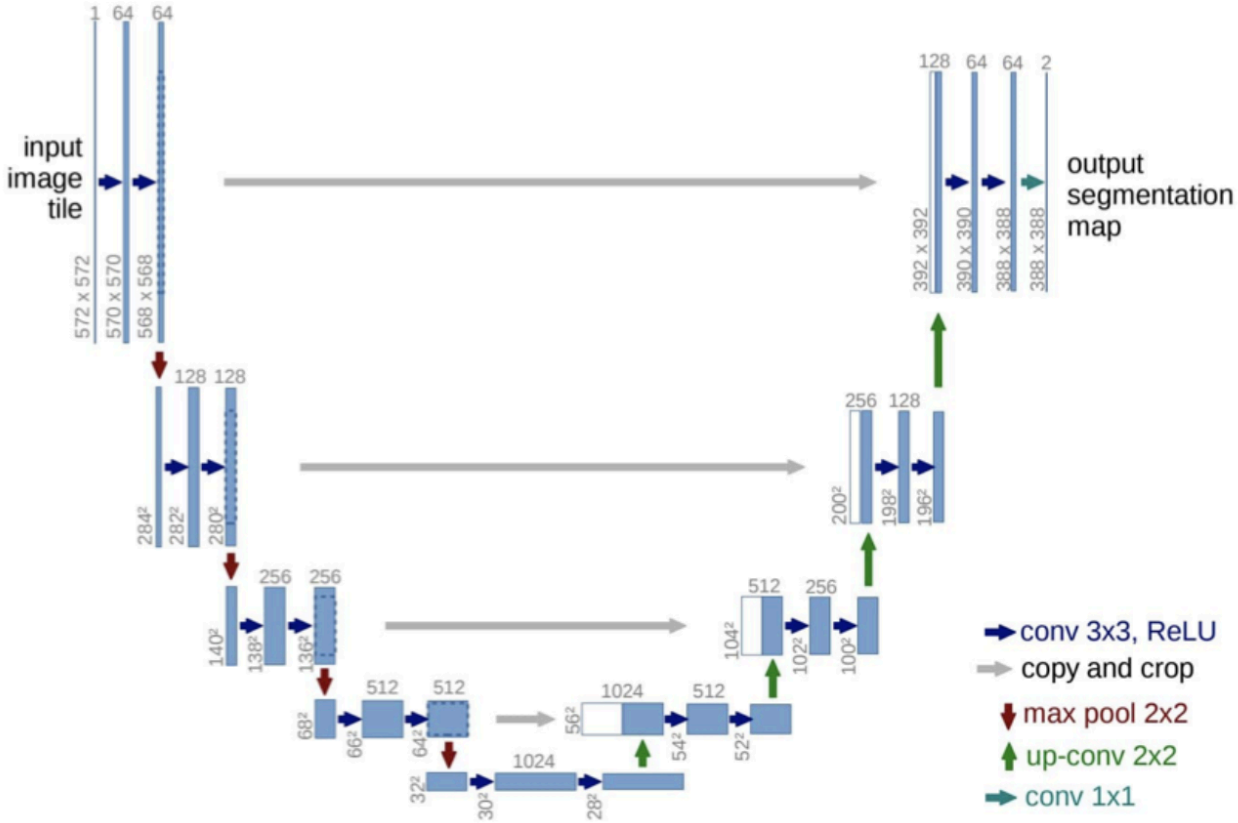
- *Consists of* an encoder and a decoder. An encoder uses convolutional layers whereas a decoder uses a deconvolutional network which generates a map of pixel-wise class probabilities based on the input feature vector. Ex: SegNet and HRNet



Deconvolutional semantic segmentation followed by a VGG-16 layer convolutional network. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

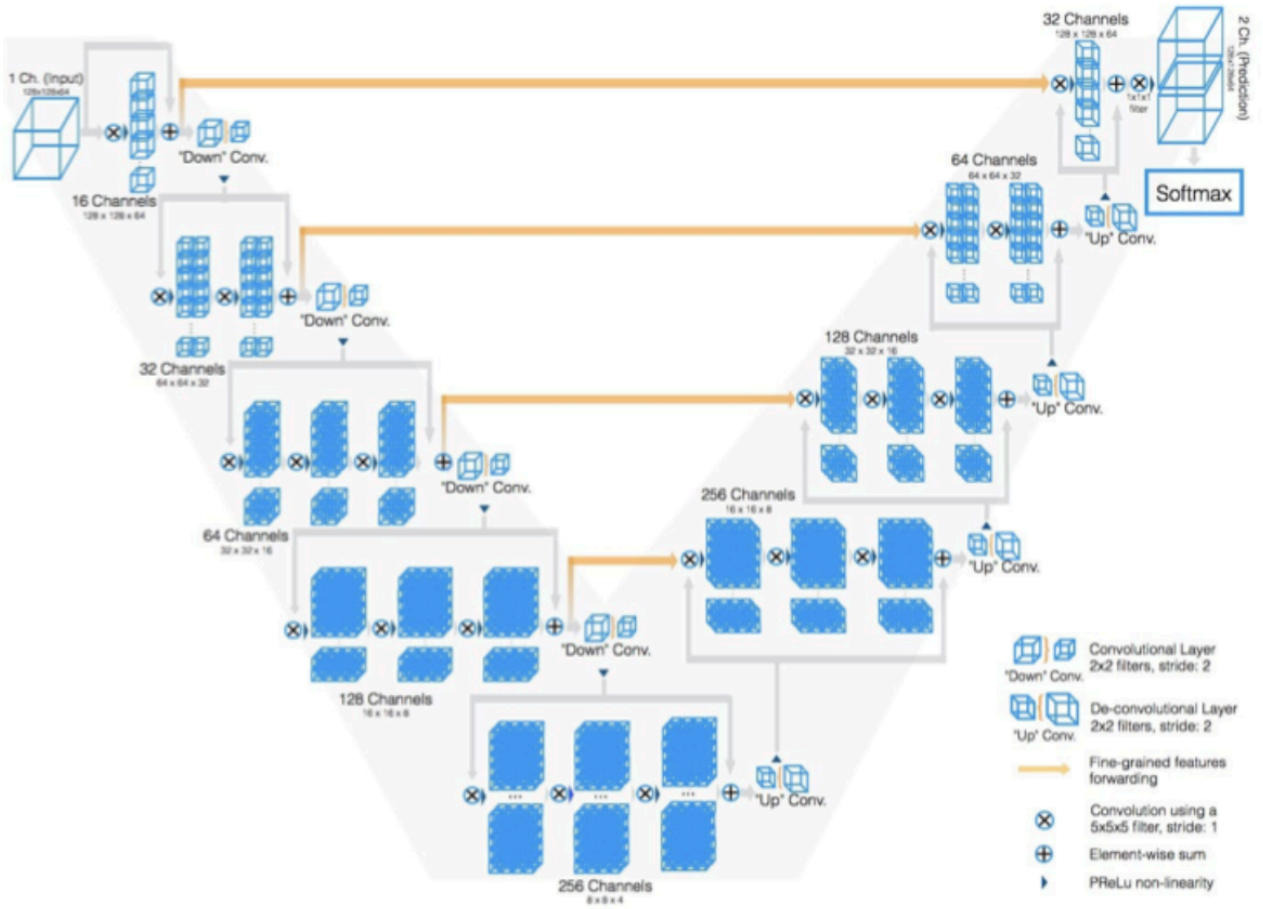
B. Encoder-Decoder Models for Medical and Biomedical Image Segmentation:

- U-Net and V-Net are the two most popular architectures used in medical/biomedical image segmentation.
- U-Net is basically used for the segmentation of biological microscopy images. It uses data augmentation techniques to learn from the available annotated images.
- U-Net architecture consists of two parts: a contracting part and a symmetric expanding path, for capturing context and enabling precise localization, respectively.



U-Net model. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

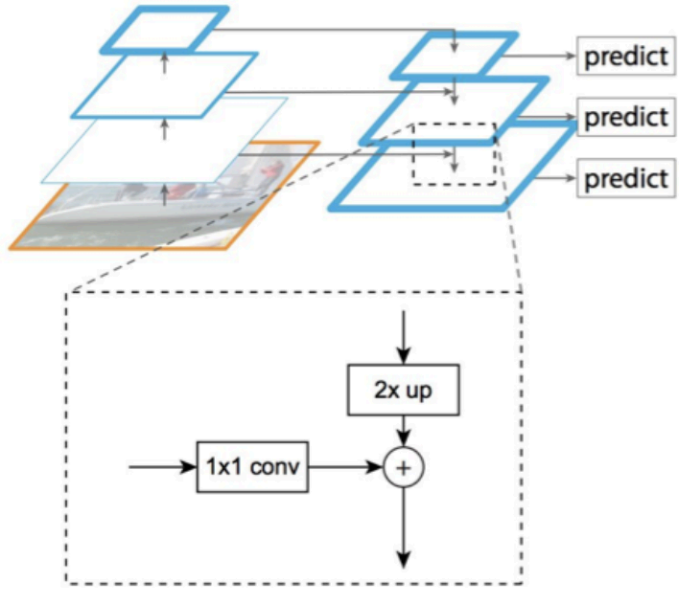
- V-Net is another popular model used for 3D medical image segmentation. It uses a new objective function for model training which is based on Dice coefficient.
- V-Net model is trained on MRI volumes and predicts the segmentation for the whole MRI volume at once.



V-Net model. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Multi-Scale and Pyramid Network Based Models

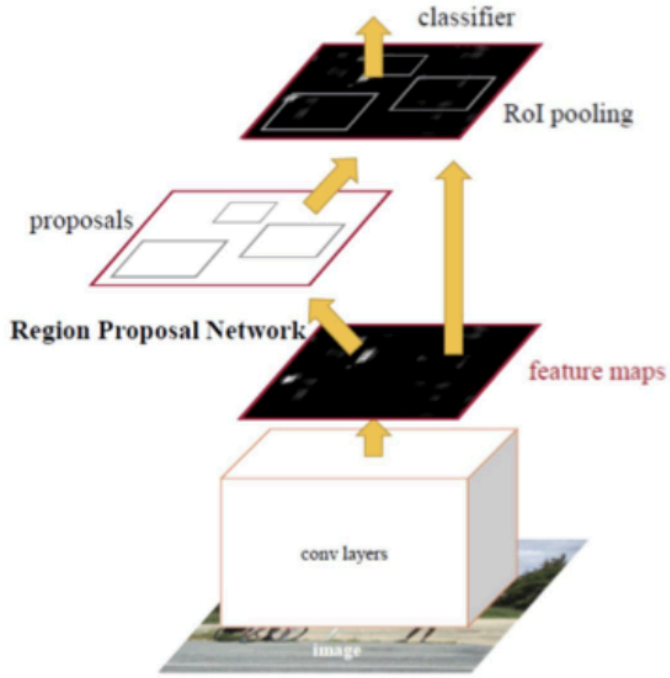
- Feature Pyramid Network (FPN) is the most popular model in this category. Initially it was developed for object detection but later was used for image segmentation as well.
- It constructs pyramid of features and uses a bottom-up pathway, a top-down pathway and lateral connections to merge low- and high-resolution features.
- It then uses a 3 × 3 convolution on concatenated feature maps to produce the output of each stage. Finally, each stage of the top-down pathway generates a prediction to detect an object



A building block illustrating the lateral connection and the top-down pathway, merged by addition. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

R-CNN Based Models (for Instance Segmentation)

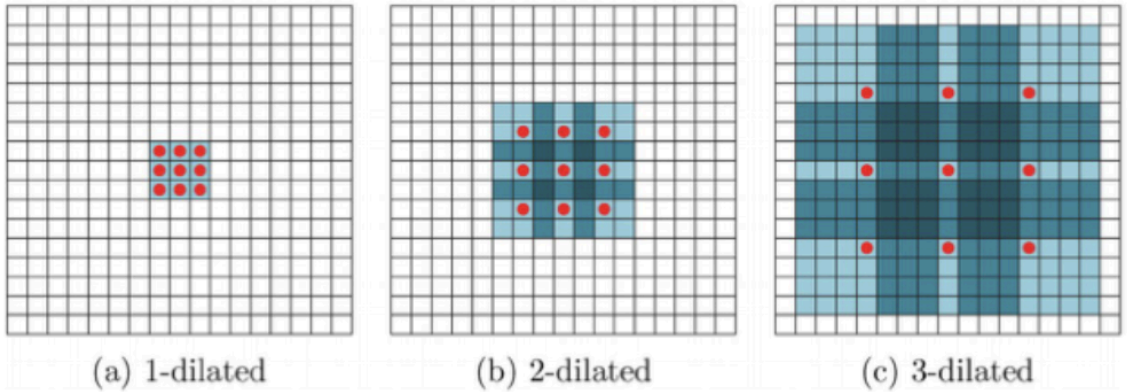
- The regional convolutional network (RCN) is a very popular model addressing the problem of instance segmentation.
- It performs the tasks of object detection and semantic segmentation simultaneously.
- Its extension Faster R-CNN uses a regional proposal network (RPN) to extract a Region of Interest (RoI) and then uses a RoIPool layer for feature computation from these proposals and infers the bounding box coordinates and class of the object.



Faster R-CNN architecture. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Dilated Convolutional Models and DeepLab Family

- In Dilated Convolutional models an additional parameter is added to convolutional layers known as dilation rate which defines a spacing between the weights of the kernel.
- They are very popular for real-time segmentation.



Dilated convolution. A 3 × 3 kernel at different dilation rates. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

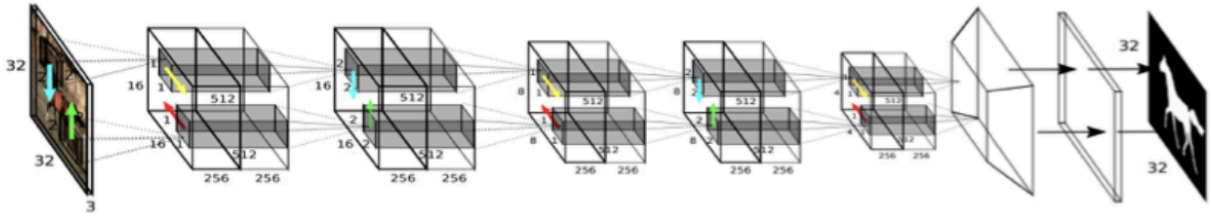
Among the DeepLab Family DeepLab v1, DeepLab v2 and DeepLab v3 are the state-of-the-art models for image segmentation approaches with DeepLap v3+ being the latest one. The DeepLab v2 has three key features.

- Use of dilated convolution to address the decreasing resolution in the network (caused by max-pooling and striding).
- Atrous Spatial Pyramid Pooling (ASPP), which probes an incoming convolutional feature layer with filters at multiple sampling rates, thus capturing objects as well as image context at multiple scales to robustly segment objects at multiple scales.
- Improved localization of object boundaries by combining methods from deep CNNs and probabilistic graphical models.

The best DeepLab (using a ResNet-101 as backbone) has reached a 70.4% mIoU score on the Cityscapes challenge.

Recurrent Neural Network Based Models

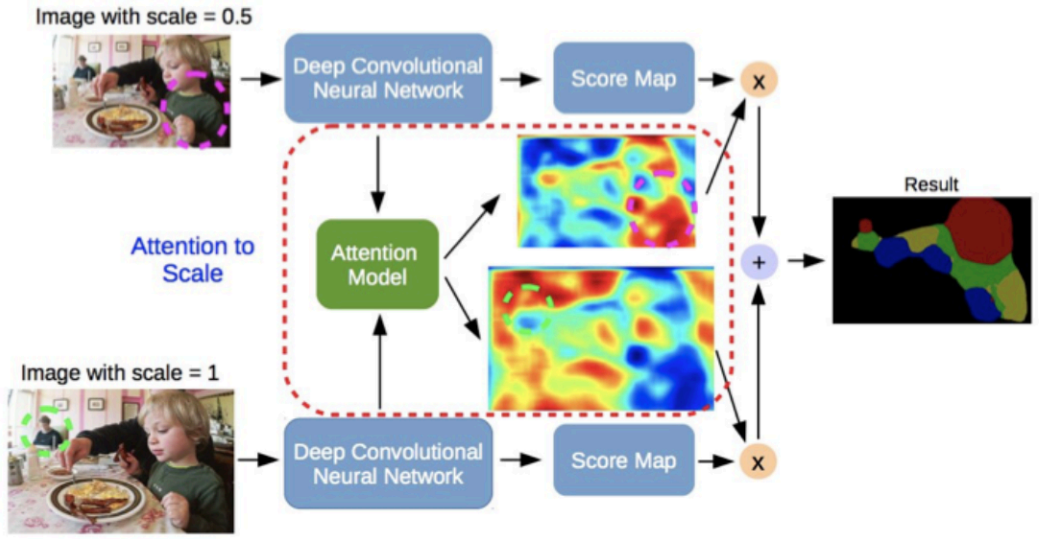
- They potentially improve the estimation of the segmentation map by modeling the short- and long-term dependencies among pixels.
- ReSeg was the first RNN-based model used for image segmentation. It was developed from ReNet which was used for image classification.
- ReSeg model uses ReNet layers which are stacked on top of the pre-trained VGG-16 convolutional layers that extract generic local features to perform image segmentation.



ReSeg Model. The pre-trained VGG-16 feature extractor network is not shown. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Attention-Based Models

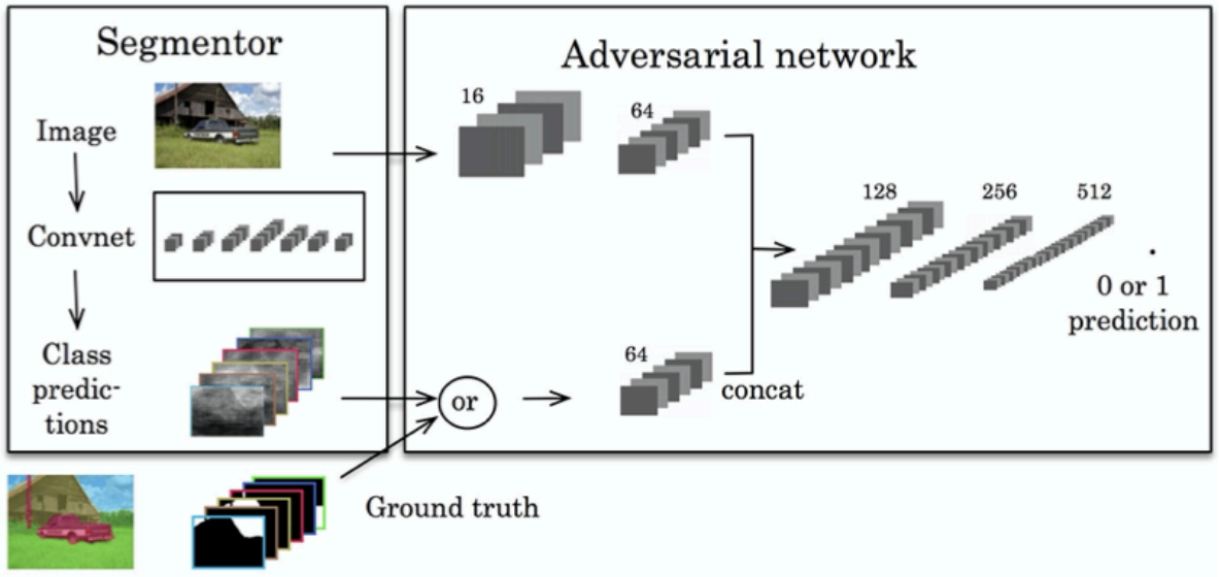
- The attention mechanism outperforms average and max pooling, and it enables the model to assess the importance of features at different positions and scales.
- Unlike CNN models, where convolutional classifiers are trained to learn the representative semantic features of labeled objects, the Reverse Attention Network (RAN) architecture trains the model to capture the features that are not associated with a target class.
- The RAN is a three-branch network that performs the direct, and reverse-attention learning processes simultaneously.



Attention-based semantic segmentation model. The attention model learns to assign different weights to objects of different scales; e.g., the model assigns large weights on the small person (green dashed circle) for features from scale 1.0, and large weights on the large child (magenta dashed circle) for features from scale 0.5. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Generative Models and Adversarial Training

- In adversarial training approach a convolutional semantic segmentation network is trained along with an adversarial network that discriminates ground-truth segmentation maps from those generated by the segmentation network.
- This approach has showed improved accuracy on the Stanford Background and PASCAL VOC 2012 datasets.

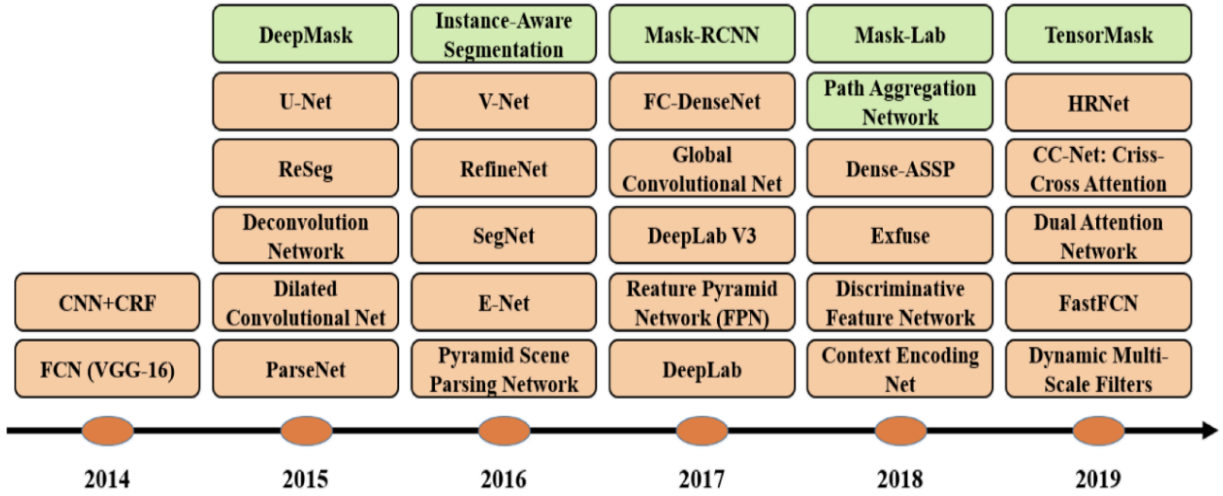


Adversarial model for image segmentation. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

CNN Models With Active Contour Models

- The FCNs along with Active Contour Models (ACMs) have recently gained interest and it is an ongoing research.
- One of its approach involves formulating new loss functions inspired by various ACM principles
- Other approach utilizes ACM merely as a post-processor of the output of an FCN and several efforts attempted modest co-learning by pre-training the FCN.

Following diagram shows the the timeline of some of the most popular DL-based works for semantic segmentation, as well as instance segmentation since 2014.



The timeline of DL-based segmentation algorithms for 2D images. Orange and green blocks refer to semantic, and instance segmentation algorithms respectively. Photo Credit : https://arxiv.org/pdf/2001.05566.pdf

Image Segmentation Datasets

The Image Segmentation datasets are divided into 3 categories: 2D images, 2.5D RGB-D (color + depth) images, and 3D images. The most popular in each of these categories include:

- 2D – PASCAL Visual Object Classes (VOC), PASCAL Context, Microsoft Common Objects in Context (MS COCO), Cityscapes
- 2.5 D– NYU-D V2, SUN-3D, SUN RGB-D, UW RGB-D Object Dataset, ScanNet
- 3D – Stanford 2D-3D, ShapeNet Core, Sydney Urban Objects Dataset.

Following table shows the accuracies of different models on cityscapes dataset using mIoU (mean Intersection over Union)as evaluation metric.

Method	Backbone	mIoU
SegNet basic [44]	-	57.0
FCN-8s [32]	-	65.3
DPN [42]	-	66.8
Dilation10 [79]	-	67.1
DeeplabV2 [78]	ResNet-101	70.4
RefineNet [117]	ResNet-101	73.6
FoveaNet [126]	ResNet-101	74.1
Ladder DenseNet [127]	Ladder DenseNet-169	73.7
GCN [121]	ResNet-101	76.9
DUC-HDC [80]	ResNet-101	77.6
Wide ResNet [122]	WideResNet-38	78.4
PSPNet [57]	ResNet-101	85.4
BiSeNet [128]	ResNet-101	78.9
DFN [99]	ResNet-101	79.3
PSANet [98]	ResNet-101	80.1
DenseASPP [81]	DenseNet-161	80.6
SPGNet [129]	2xResNet-50	81.1
DANet [93]	ResNet-101	81.5
CCNet [96]	ResNet-101	81.4
DeeplabV3 [12]	ResNet-101	81.3
DeeplabV3 [83]	Xception-71	82.1
AC-Net [131]	ResNet-101	82.3
OCR [119]	ResNet-101	82.4
GS-CNN [130]	WideResNet	82.8
HRNetV2+OCR (w/ASPP) [119]	HRNetV2-W48	83.7

Conclusion

- We discussed about various state-of-the-art models for image segmentation using deep learning and performance characteristics of different models on cityscapes dataset.
- Deep learning for image segmentation have proved to be very powerful so far but as most of the segmentation networks require large amount of memory for training and inference, these models are bounded by this constraint. Extensive research is ongoing to tackle this problem and we can expect a flurry of innovation and unique research lines in the upcoming years.