

Semantic Segmentation

Guowei Wei¹, Rui Wang¹

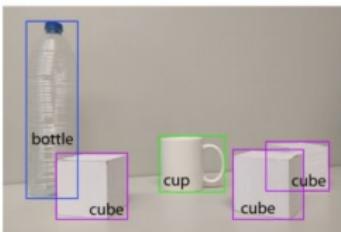
¹Department of Mathematics, Michigan State University

March 18, 2021

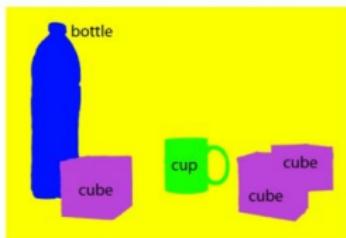
► What is the difference?



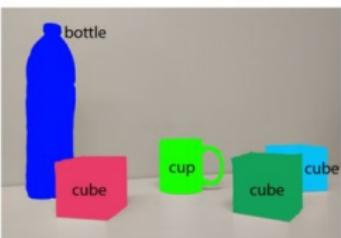
(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation

Figure 1: Top 4 tasks in computer vision.

► What is Image Segmentation?

A fundamental topic in computer vision. The task of semantic image segmentation is to classify each pixel in the image.

► Why Image Segmentation?

- Robot vision and understanding
- Autonomous driving
- Medical purposes ([ISBI Challenge](#))
- Object Detection in Satellite Imagery

► **PASCAL context dataset**

- Include more labels with similar context, such as chair, sofa, horse, and cow.

► **ADE20K dataset**

- The most challenging one with a large and unrestricted open vocabulary and more scene classes.

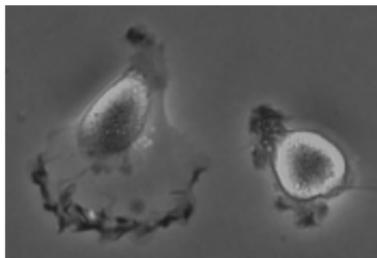
► **Cityscapes dataset**

- For semantic urban scene understanding. Contains 5000 high quality pixel-level finely annotated images collected from 50 cities in different seasons.

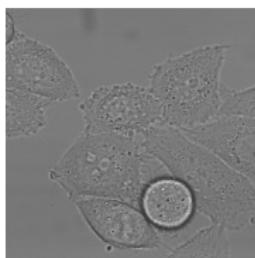
► **ISBI Dataset**

- For segmentation of neuronal structures in electron microscopy images.

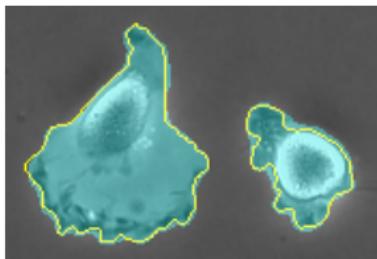
► **Satellite Dataset**



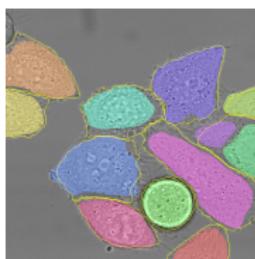
(a) Input 1



(b) Input 2



(c) Pred 1



(d) Pred 2

Figure 2: U-Net Example [3]

Dataset: Cityscapes



(a) Input 1



(b) Input 2



(c) Input 3



(d) Pred 1



(e) Pred 2



(f) Pred 3

Figure 3: PspNet Cityscapes Example

Dataset: VOC2012



(a) Input 1



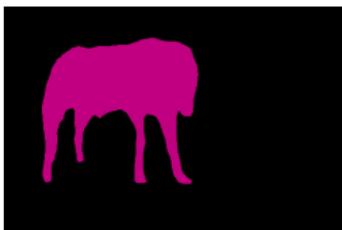
(b) Input 2



(c) Input 3



(d) Pred 1



(e) Pred 2



(f) Pred 3

Figure 4: PspNet VOC2012 Example

Dataset: Eye in the Sky



(a) Input



(b) Ground Truth



(c) Pred

Figure 5: U-Net Training Example

Dataset: Eye in the Sky



(a) Input



(b) Pred

Figure 6: U-Net Testing Example



Figure 7: Dove Image Dateset. Left: Tree cover in non-winter seasons (forest and fields in Pennsylvania, USA.) Right: Burned area evolution over a week (grassland in Madagascar.)

- ▶ **FCNs (Fully Convolutional Networks)**
- ▶ **PSPNet (Pyramid Scene Parsing Network)**
- ▶ **U-Net**
- ▶ **SegNet**
- ▶ **RCNNs (Region-based CNNs)**
- ▶ **Adversarial Models (GAN)**

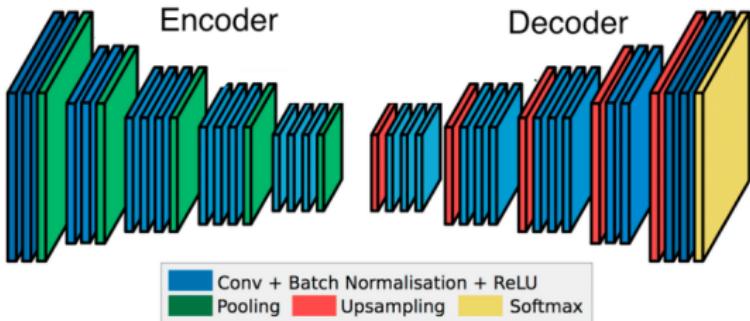


Figure 8: Encoder and Decoder [4]

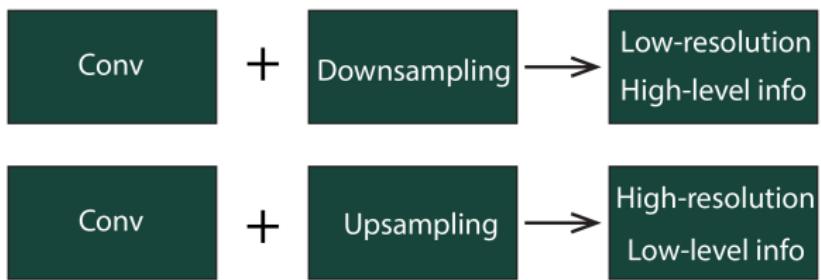


Figure 9: Layers in the encoder and decoder [4]

- ▶ Convolutional layers are involved in the image segmentation.
 - The initial layers learn to detect features like color and edges.
 - Deeper layers learn to detect features such as different objects.
- ▶ Downsampling is done by pooling layers.
- ▶ No Fully connected layer needed.
 - Image classification maps the spatial tensor to a vector.

$$(N, C, H, W) \rightarrow (N, C \times H \times W) \rightarrow (N, H) \rightarrow (N, \# \text{of classes})$$

- Image segmentation needs to keep the spatial information.

- ▶ Produce high-resolution segmentation outputs by using the low-resolution spatial tensor.
 - Add more convolutional layers
 - Add upsampling layers which increase the size of the spatial tensor.
- ▶ Loss function
 - Cross-Entropy on each pixel.
- ▶ Data augmentation
 - If we have limited training sets, the result is not good since the model might overfitting.
 - Increase the size of the dataset by random transformation such as rotation, scale, and flipping on the images to enlarge our dataset.

- ▶ Some initial layers of the base CNN network are used in the encoder.
- ▶ Choices:
 - ResNet: By Microsoft, large number of layers
 - VGG16: By Oxford, lesser layers than ResNet, faster to train.
 - MoboileNet: By Google, for a small model size and faster inference time
- ▶ Depends on the use case:
 - The number of training images.
 - Size of images.
 - The domain of the images.

- ▶ Learnable parameters are used
 - Transposed convolution
- ▶ No learnable parameters are used
 - Nearest Neighbor interpolation
 - Bed of Nails interpolation
 - Bilinear interpolation
 - Max-Pooling indices

Detailed information is listed in the Semantic_Segmentation.pdf

► Decoder: Bilinear

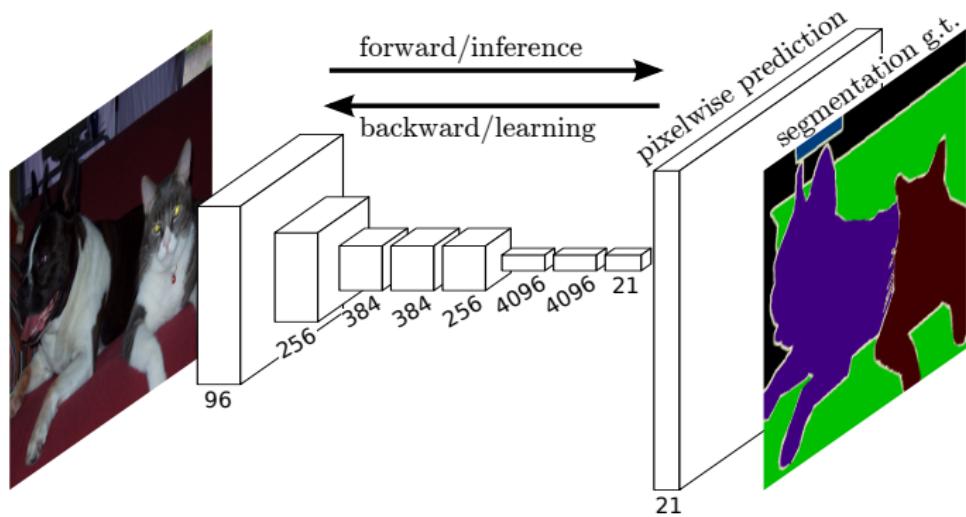


Figure 10: Overview of FCN^[1].

► Decoder: Max-Pooling indices

Max Pooling

Remember which element was max!

1	2	6	3
3	5	2	1
1	2	2	1
7	3	4	8

Input: 4 x 4

5	6
7	8

Output: 2 x 2

Max Unpooling

Use positions from pooling layer

1	2
3	4



0	0	2	0
0	1	0	0
0	0	0	0
3	0	0	4

Output: 4 x 4

Corresponding pairs of
downsampling and
upsampling layers

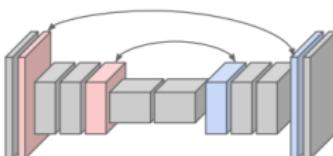


Figure 11: Downsampling and Upsampling.

Advantages of using Max-Pooling indices:

- ▶ It improves boundary delineation
- ▶ Reduces the number of parameters enabling end-to-end training
- ▶ This upsampling method can be incorporated into any encoder-decoder architecture.

► Decoder: Bilinear

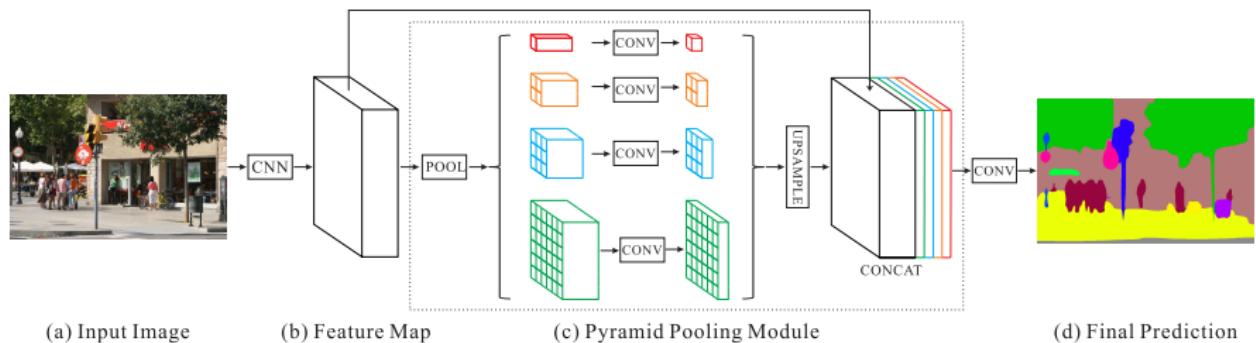


Figure 12: Architecture of PSPNet^[2].

Segmentation architecture: U-Net

► Decoder: Transposed convolution

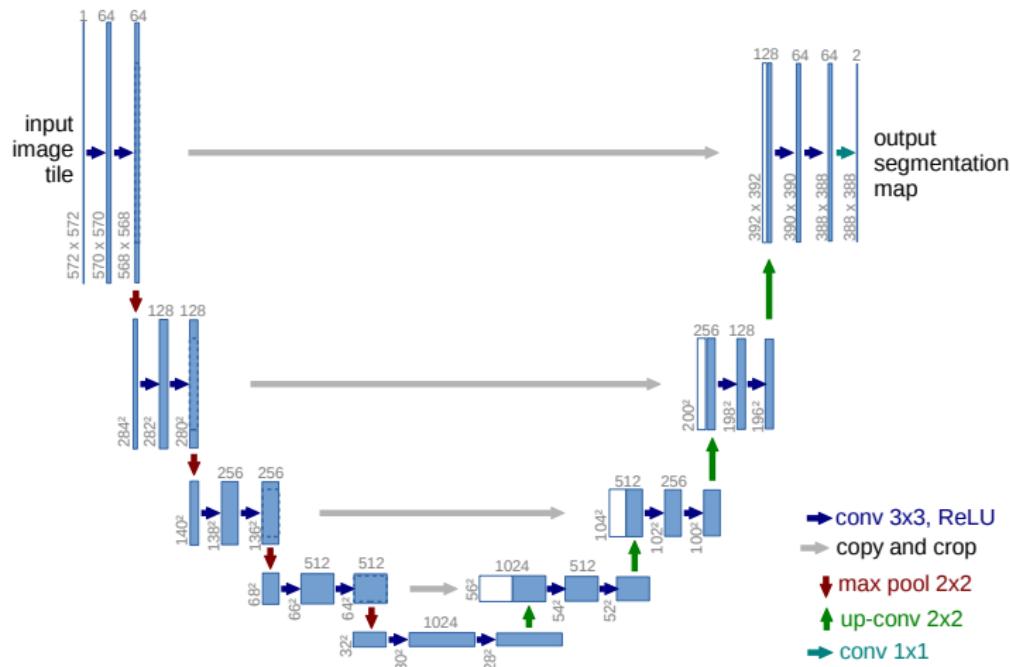


Figure 13: Architecture of U-Net^[3].

- ▶ Used for biomedical image segmentation.
- ▶ Instead of using pooling indices, the entire feature maps are transferred from encoder to decoder, then with concatenation to perform convolution.
- ▶ Require more memory for this large model.



- We propose to develop and apply a multi-temporal deep learning approach to PLANET time series..

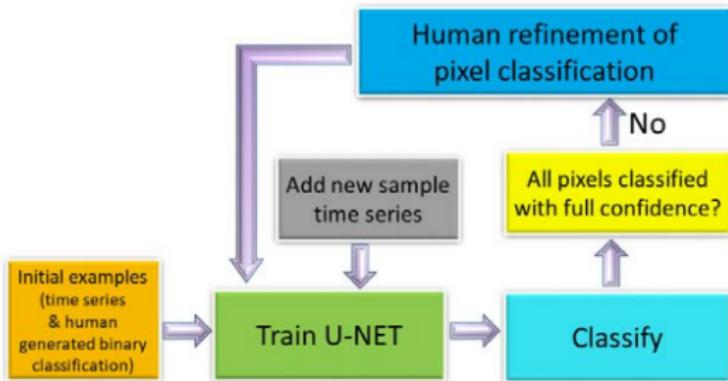


Figure 14: Workflow of the iterative semi-automatic active learning training data set generation process.

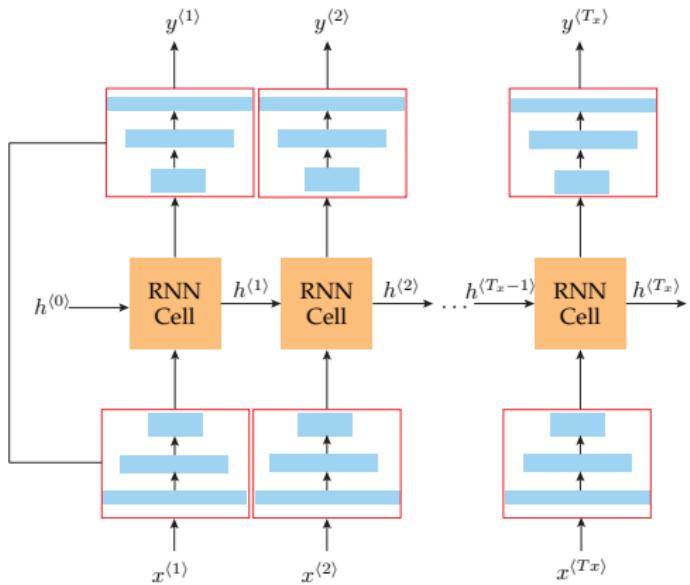


Figure 15: The structure of Recurrent U-Net.

- ▶ Pixel Accuracy: the percent of pixels in our image that are classified correctly.
- ▶ Intersection-Over-Union(IoU,Jaccard index): One of the most commonly used metrics in semantic segmentation.

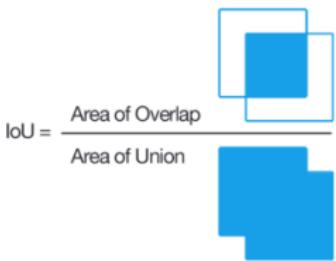


Figure 16: IoU. Source: Wikipedia

- ▶ Fully convolutional networks (FCNs):
<https://github.com/shelhamer/fcn.berkeleyvision.org>
- ▶ Region-based Convolutional Neural Networks (RCNNs):
<https://github.com/rbgirshick/py-faster-rcnn>
- ▶ PspNet: <https://github.com/hszhao/semseg>
- ▶ U-Net: <https://github.com/zhipenghuo/unet>
- ▶ SegNet: <https://github.com/toimcio/SegNet-tensorflow>
- ▶ Adversarial Models (GAN): <https://github.com/albertbou92/Semantic-Segmentation-with-Adversarial-Networks>

- 1 Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. [FCN](#)
- 2 Zhao, Hengshuang, et al. "Pyramid scene parsing network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. [PSPNet](#)
- 3 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. [U-Net](#)
- 4 Syed, Arsal, and Brendan Tran Morris. "SSeg-LSTM: Semantic Scene Segmentation for Trajectory Prediction." 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019. [SSeg-LSTM](#)