Semantic Segmentation

Guowei Wei[1], Rui Wang[1]

[1]Department of Mathematics, Michigan State University

March 18, 2021

- **What is the difference?**

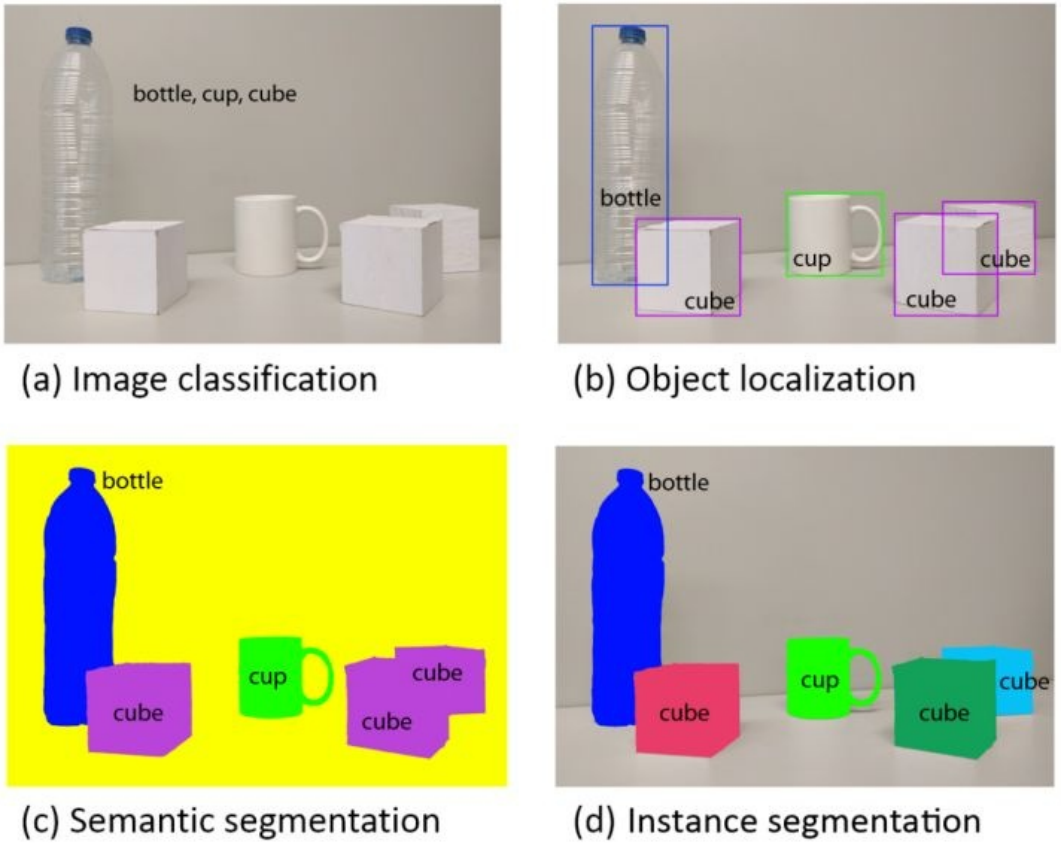


Figure 1: Top 4 tasks in computer vision.

Image Segmentation

- **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

Datasets we have

- **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 number of scene-centric images exhaustively annotated with objects and object parts.
- **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**

Dataset: Cityscapes

(a) Input 1

(b) Input 2

(c) Input 3

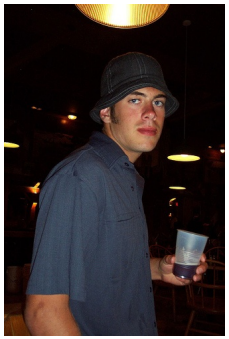
(d) Pred 1

(e) Pred 2

(f) Pred 3

Figure 2: PspNet Cityscapes Example

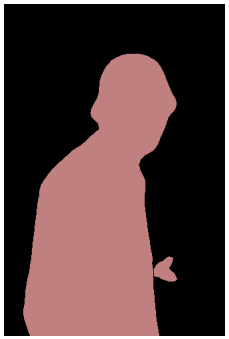
Dataset: VOC2012



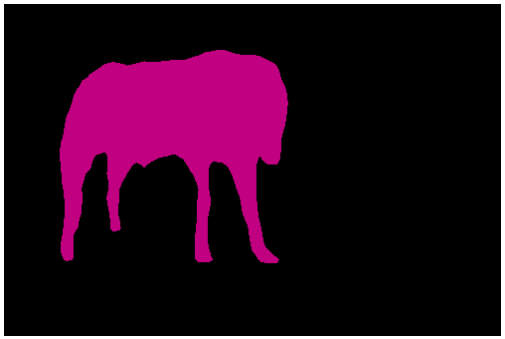
(a) Input 1

(b) Input 2

(c) Input 3



(d) Pred 1

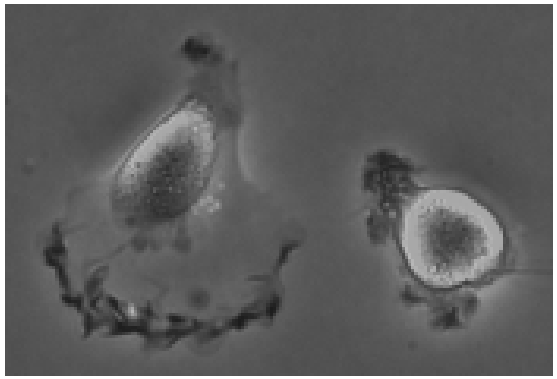


(e) Pred 2

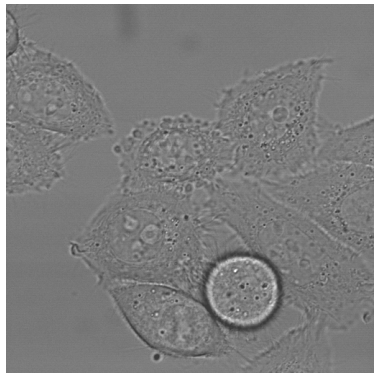
(f) Pred 3

Figure 3: PspNet VOC2012 Example

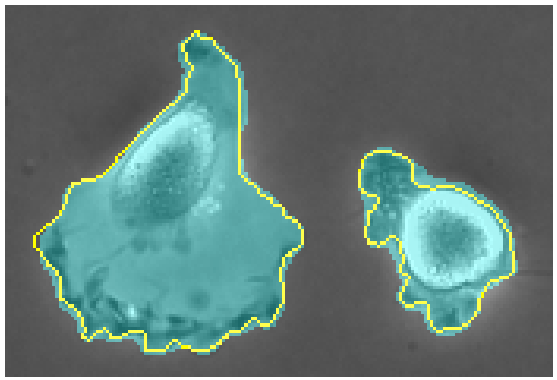
Biomedical Dataset



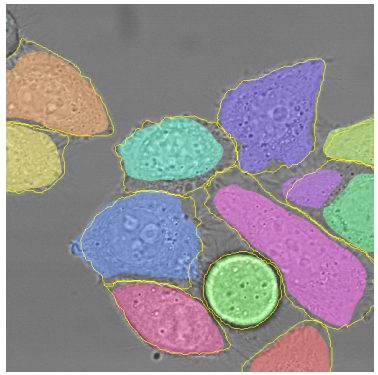
(a) Input 1



(b) Input 2



(c) Pred 1



(d) Pred 2

Figure 4: U-Net Example [3]

Satellite Dataset

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

Satellite Dataset

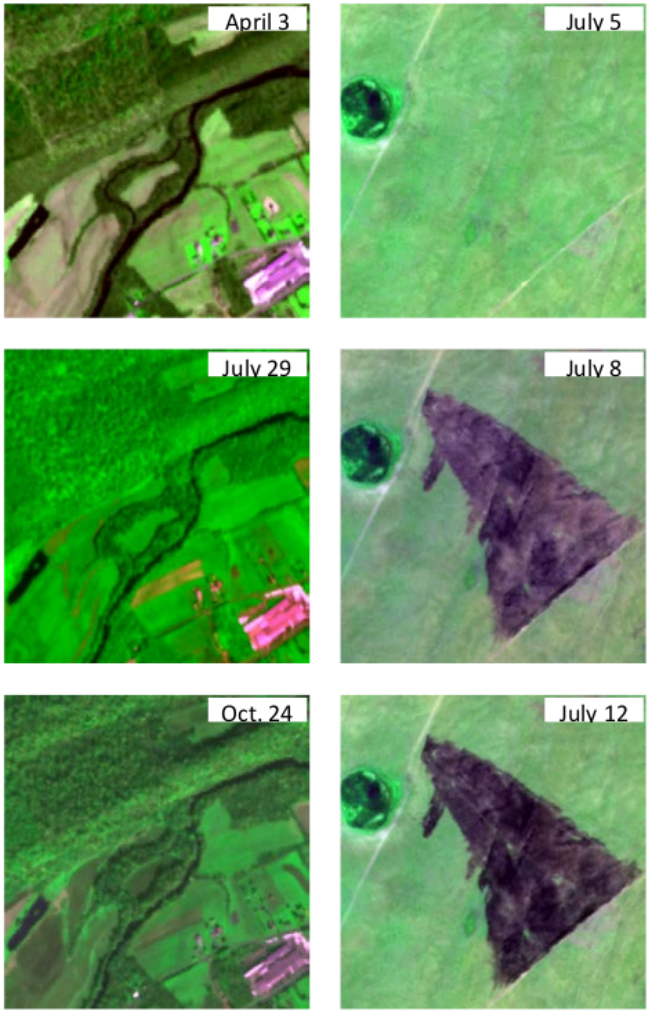


Figure 7: Dove Image Dateset. Left: Tree cover in non-winter seasons (forest and fields in Pennsylvania, USA.) Right: Burned area evolution over a week (grassland in Madagascar.)

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

Basic ideas for segmentation

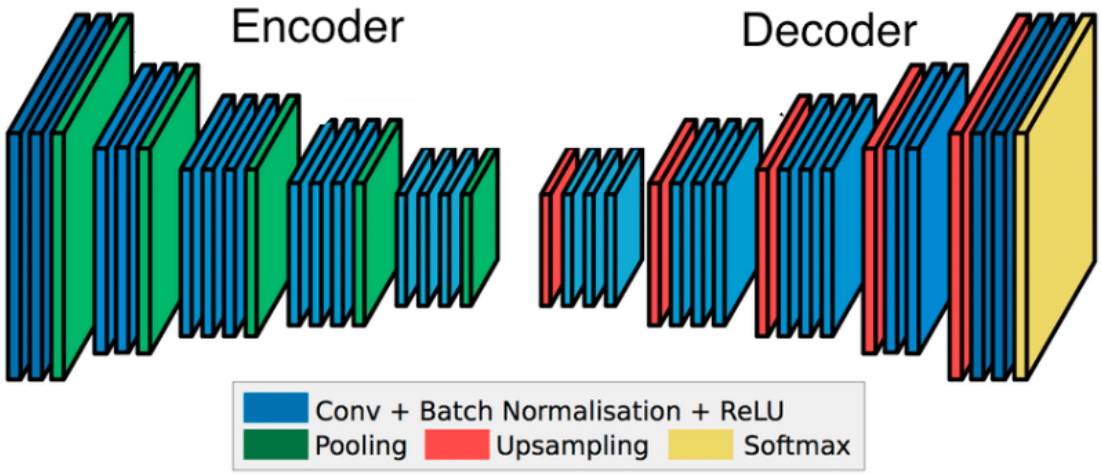


Figure 8: Encoder and Decoder [4]

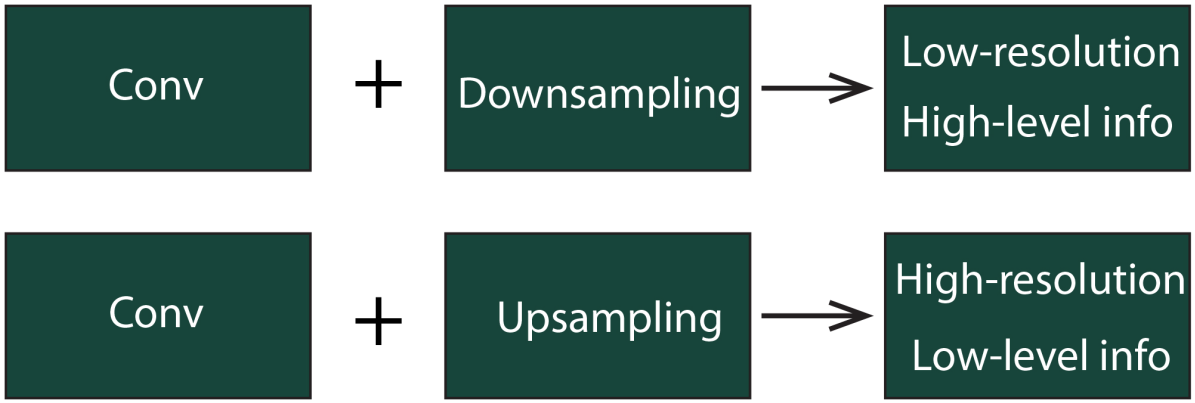


Figure 9: Layers in the encoder and decoder [4]

Basic ideas for segmentation

- 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 down by pooling layers.
- No Fully connected layer needed.
  - Image classification maps the spatial tensor to a vector.

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

  - Image segmentation needs to keep the spatial information.

Basic ideas for segmentation

- Produce high-resolution segmentation outputs by using the low-resolution spatial tensor.
  - Add more convolutional layers
  - Add upsampling layers which increase the sizer 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.

Encoder: Base model

- 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.

Decoder: Upsampling

- 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

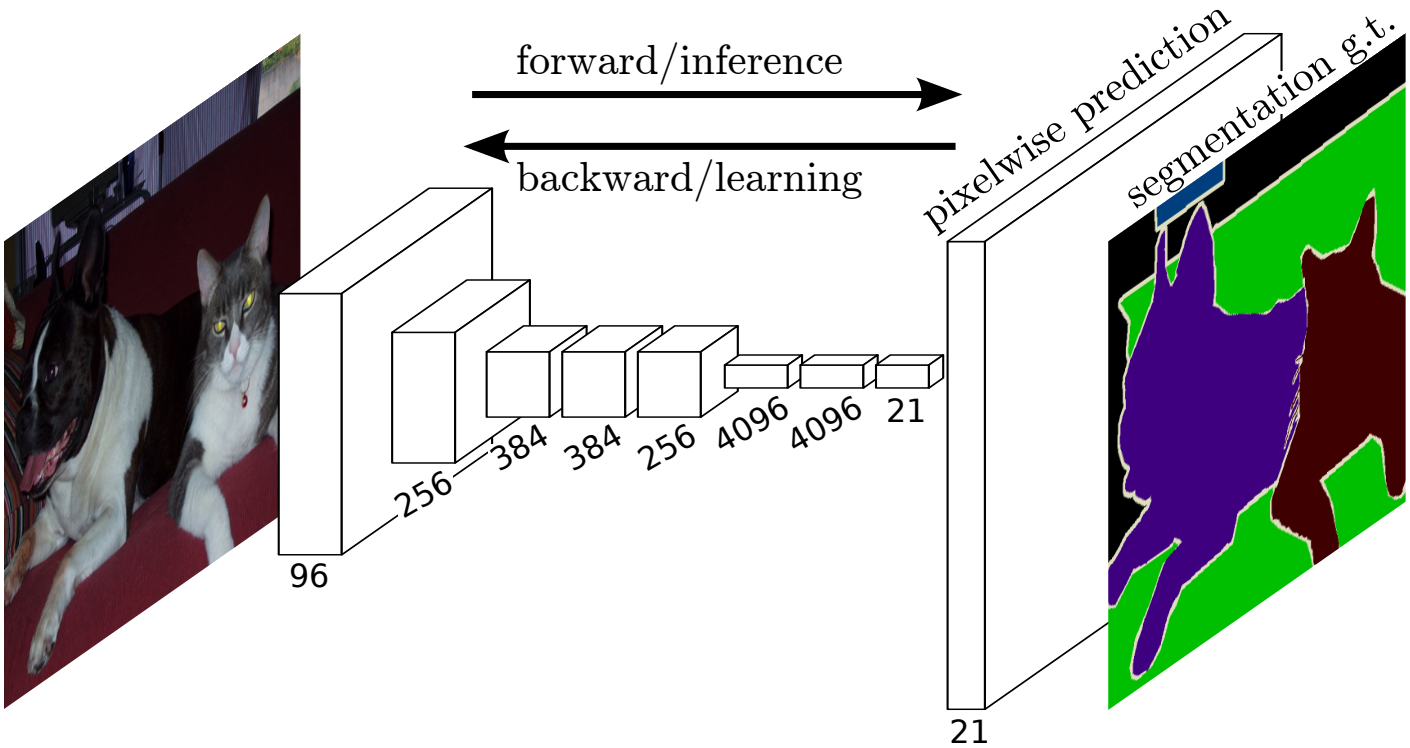


Figure 10: Overview of FCN[1].

- Decoder: Max-Pooling indices

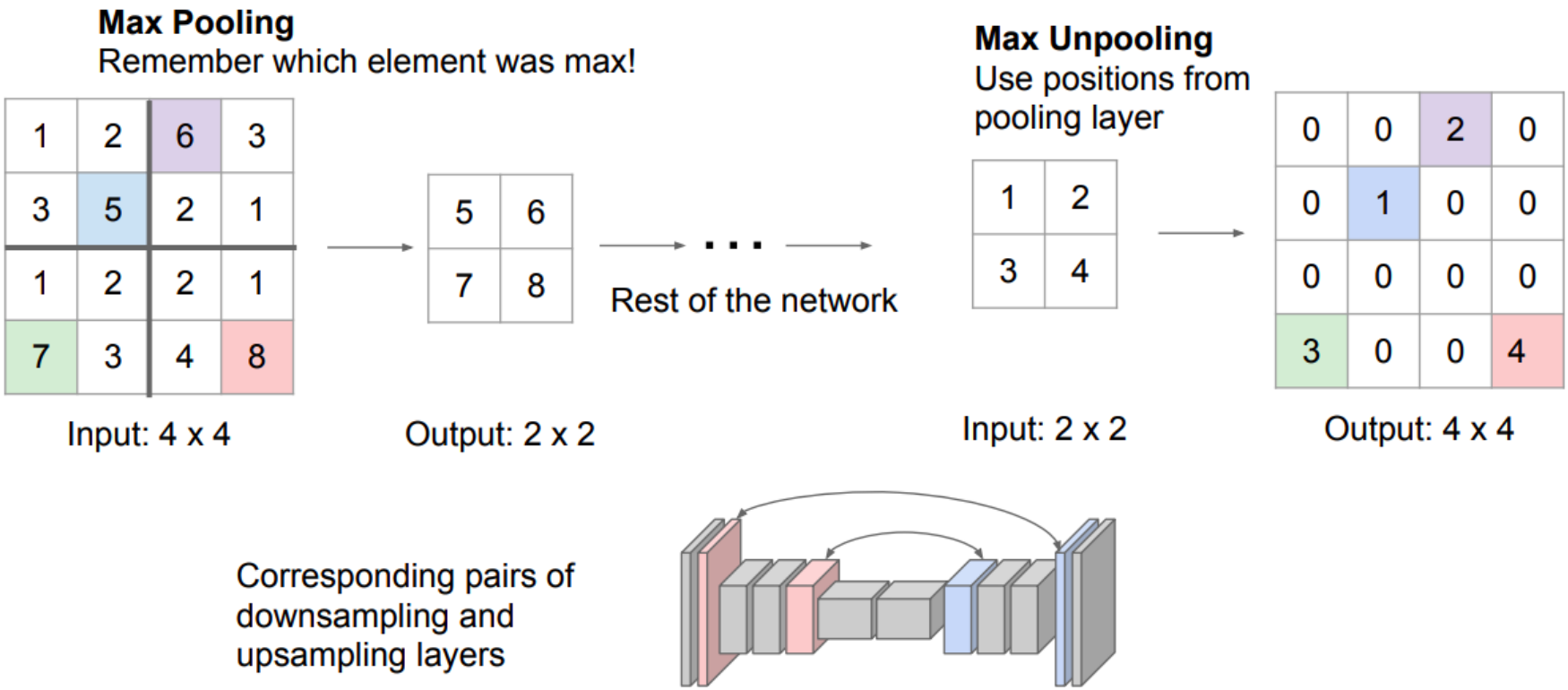


Figure 11: Downsampling and Upsampling.

Segmentation architecture: SegNet

Advantages of using Max-Pooling indices:

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

- Decoder: Bilinear

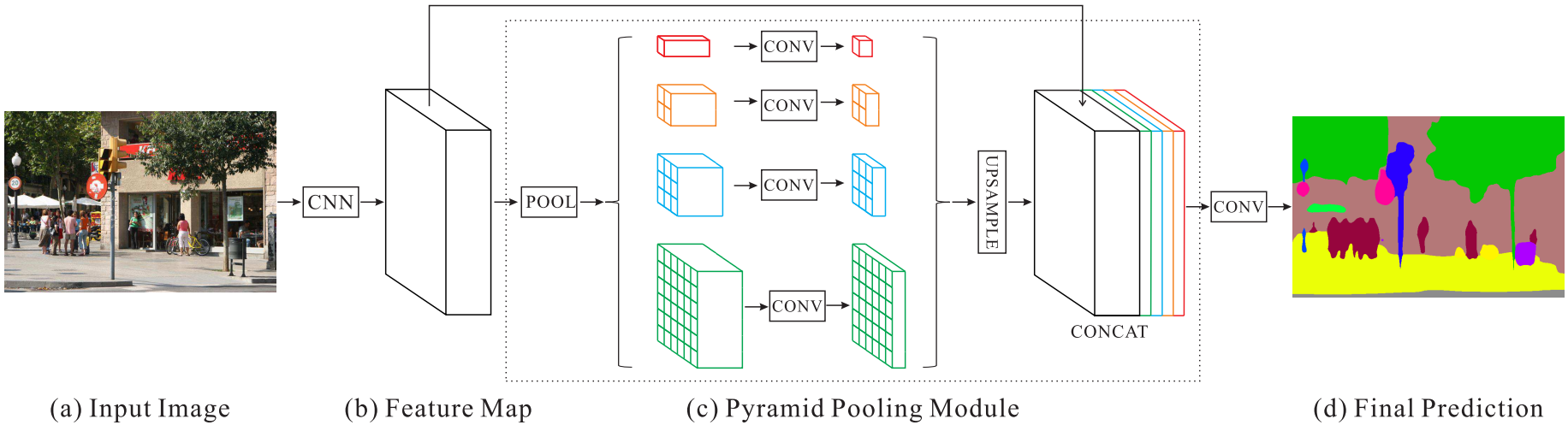


Figure 12: Architecture of PSPNet[2].

- Decoder: Transposed convolution

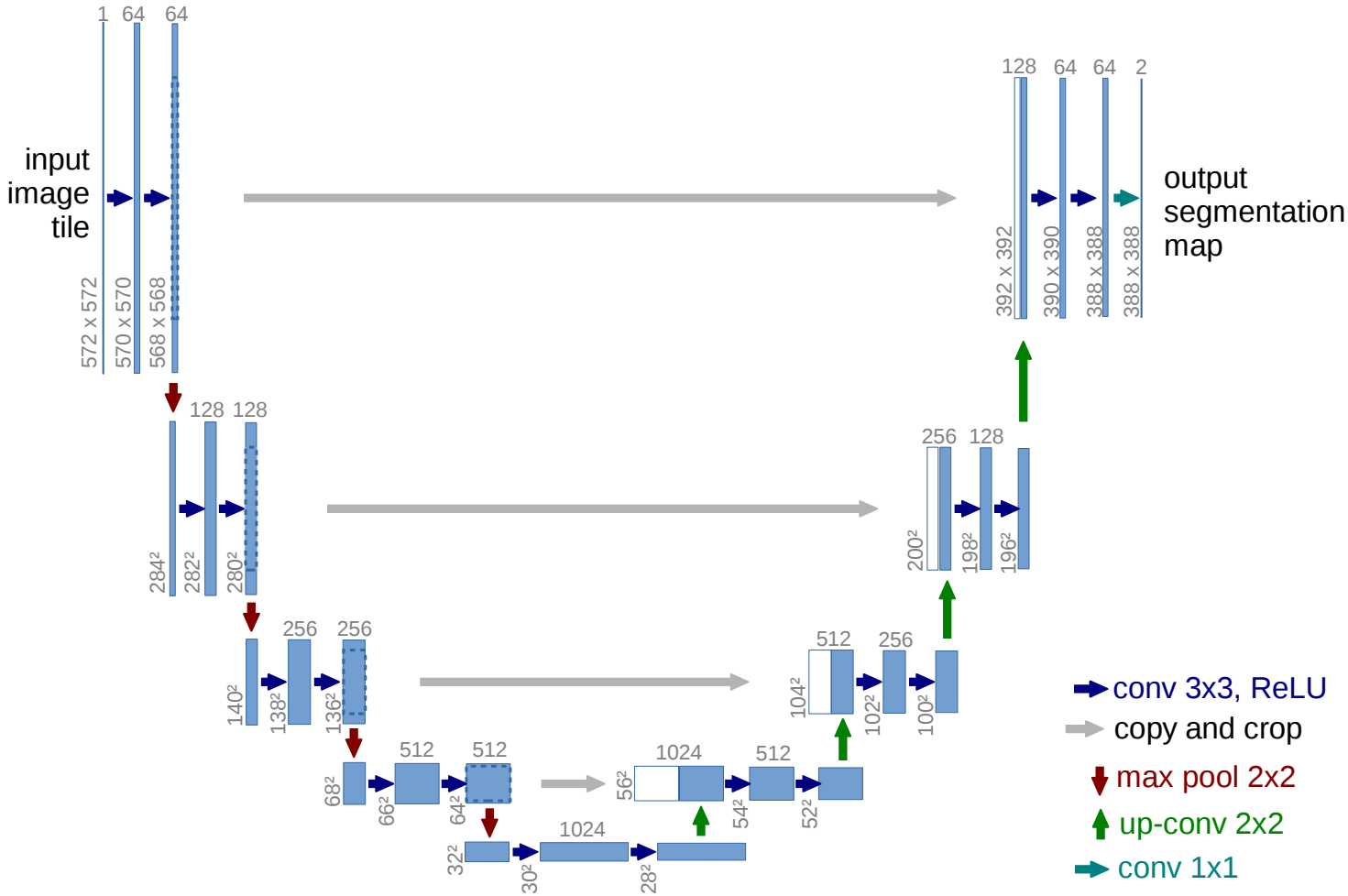
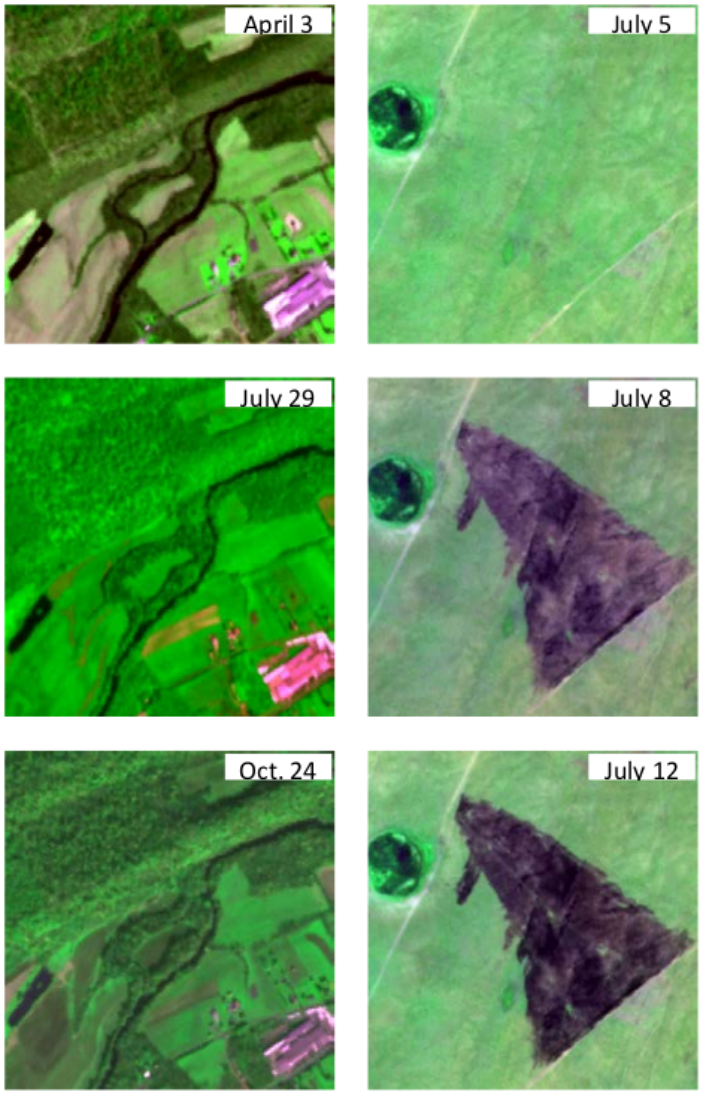


Figure 13: Architecture of U-Net[3].

Segmentation architecture: U-Net

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

Recurrent U-Net



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

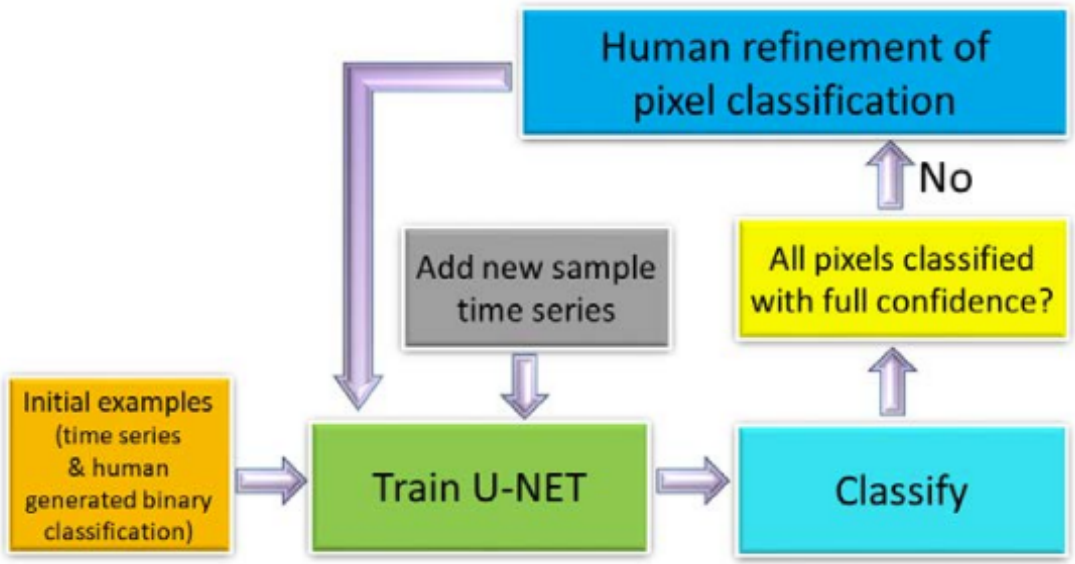


Figure 14: Workflow of the iterative semi-automatic active learning training data set generation process.

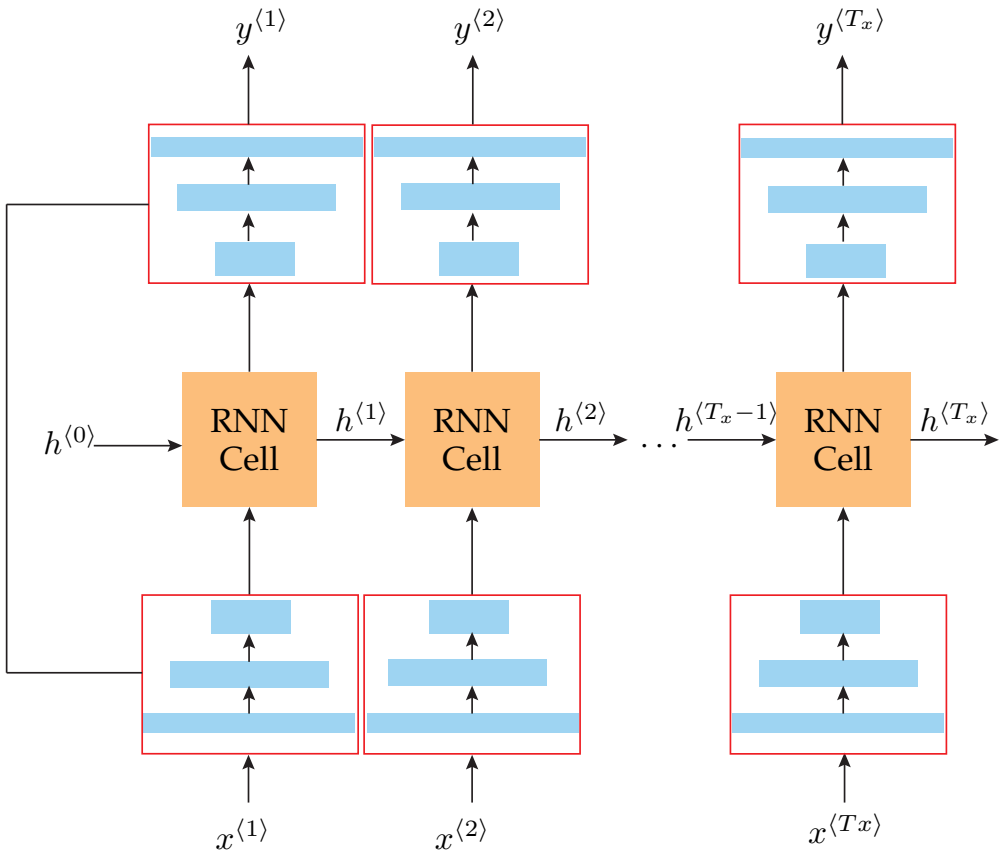


Figure 15: The structure of Recurrent U-Net.

- 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.

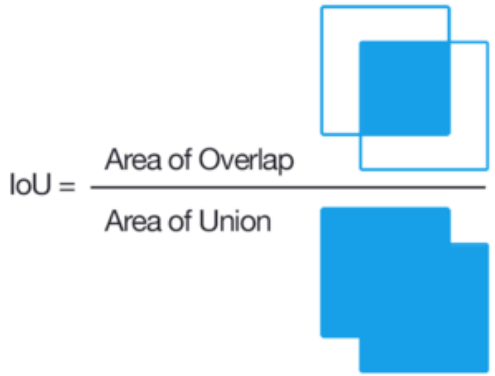


Figure 16: IoU. Source: Wikipedia

Github projects

- 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/zhixuhao/unet
- SegNet: https://github.com/toimcio/SegNet-tensorflow
- Adversarial Models (GAN): https://github.com/albertbou92/Semantic-Segmentation-with-Adversarial-Networks

References

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