

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one partially covering the green one.

Semantic Segmentation

By: Humza Syed and Yuri Yakovlev

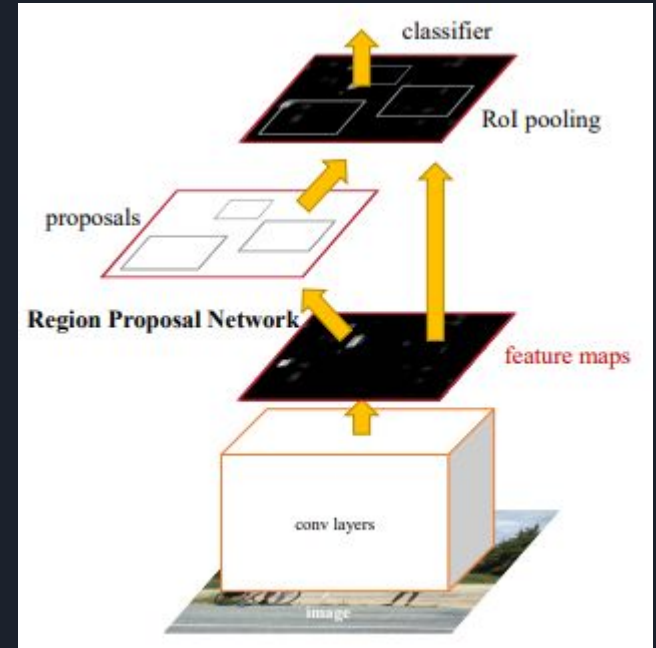


Objective

- Detect labeled objects in an image
- Pixel-wise segmentation of objects relative to these labels
- Achieve an IoU of above 0.25

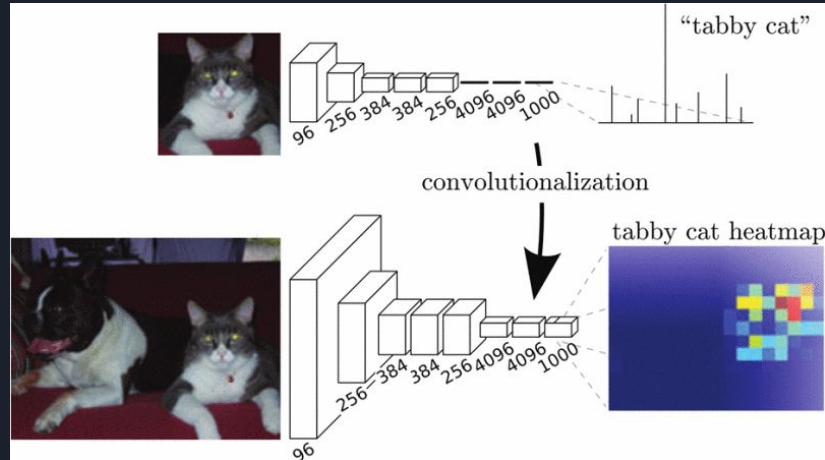
Related Work: Faster R-CNN

- Faster Region-based Convolutional Neural Network
- Region Proposal Network defines regions of interest
- Convolutional Neural Network detects objects in the proposed regions
- Regions where objects are best detected become the bounding boxes for said objects
- Leads to bounding box localization and classification



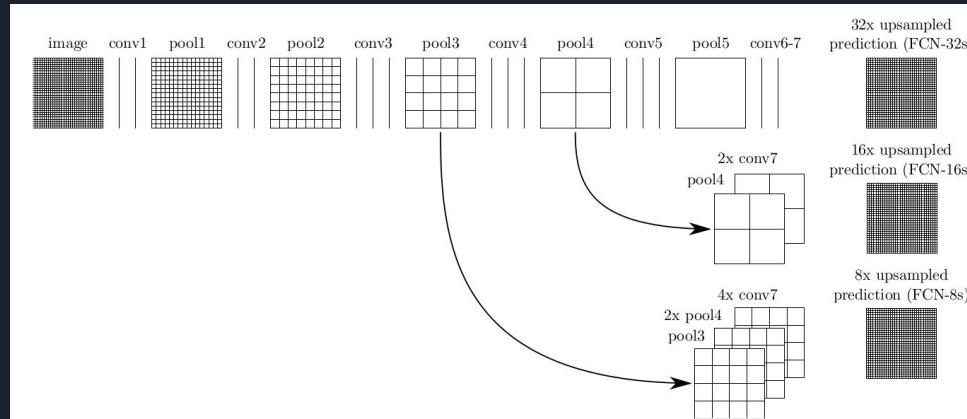
Related Work: FCN

- Fully Convolutional Network
- Change fully connected layers in a given architecture into convolutional layers
- Utilizes a convolutional encoder network, such as VGG16, to extract features from an image
- Upscales outputs from the encoder via deconvolutional layers
- Leads to pixel-wise localization and classification



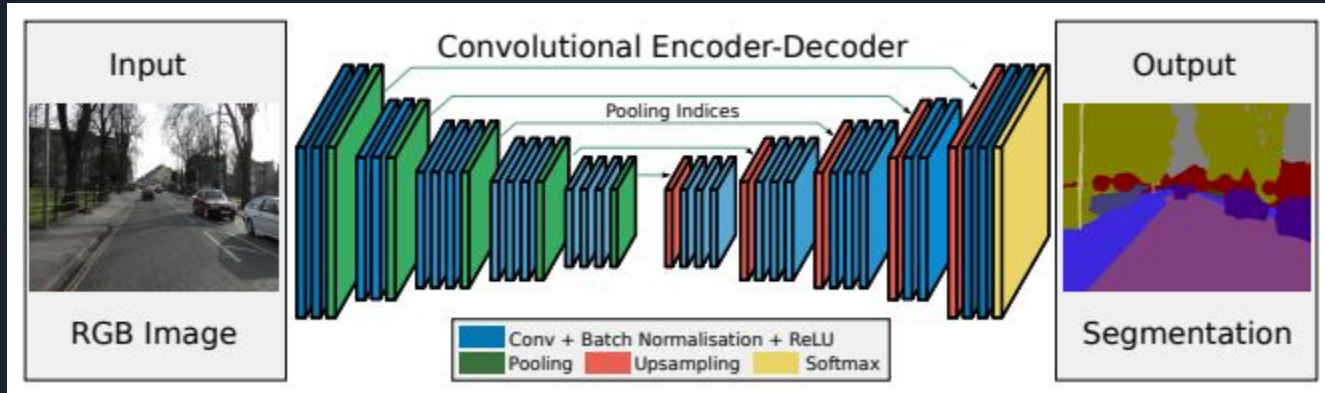
Methods: FCN

- Takes intermediate layers and output layers of a given encoder, such as VGG16, and upsamples them using deconvolution
- For example, FCN32 uses the final output layer and the last pooling layer and performs element wise addition
- In this work, we use the outputs from each pooling layer and the output layer and perform element wise addition on them



Methods: SegNet

- Uses an encoder and decoder system, where the encoder is a VGG16 network
- From each pooling layer, the pooling indices are recorded from the encoder such that sparse upsampled feature maps can be created through the decoder using unpooling rather than deconvolution

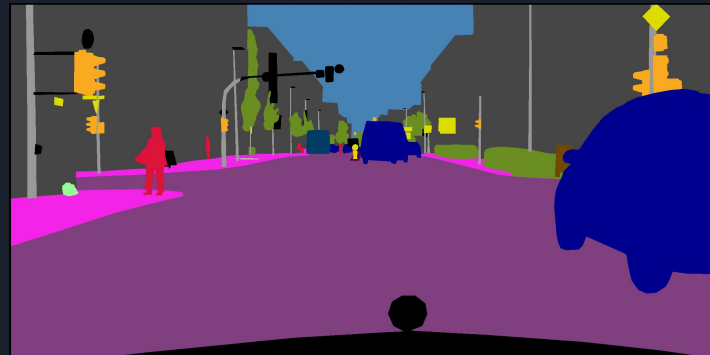


Experimental Methods: Datasets

- NYUv2
- Consists of 1,449 RGB 640x480 images with 14 classes
- Train/Test split was approximately 50/50
- Cityscapes
- Consists of 5,000 RGB 2048x1024 images with 19 classes
- Train/Val/Test split was 60/10/30



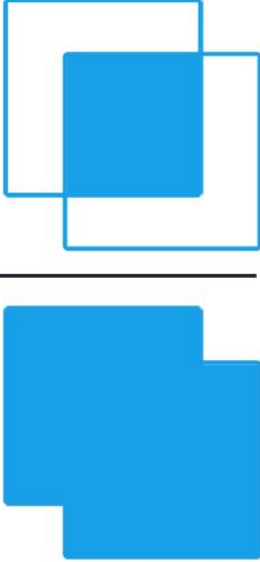
Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In ECCV, 2012.



Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Experimental Methods: Metrics and Hyperparameters

- Intersection over Union (IoU)
 - Pixel Accuracy
 - Mean Pixel Accuracy
-
- Stochastic Gradient Descent
 - Number of Epochs = 100
 - Learning Rate = 1e-3
 - Weight Decay = 0.0016

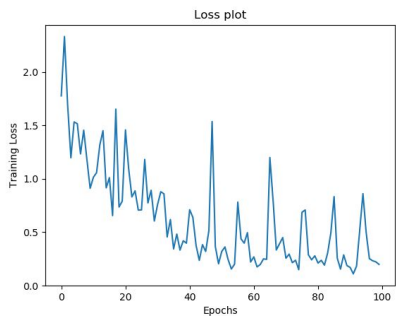

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Results: Metrics

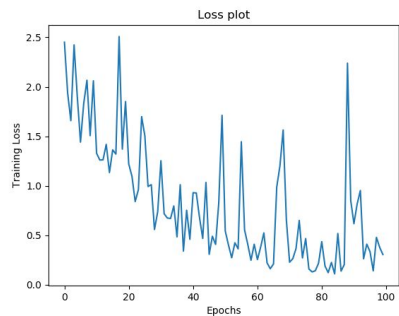
Table 1: Results using each model on each dataset.

Dataset	Model	Pixel Accuracy (%)	Mean Pixel Accuracy (%)	IoU (%)
Cityscapes	FCN	91.58	56.92	49.68
Cityscapes	SegNet	90.16	48.59	40.68
NYUDv2	FCN	59.77	48.64	36.18
NYUDv2	SegNet	50.15	33.17	22.81

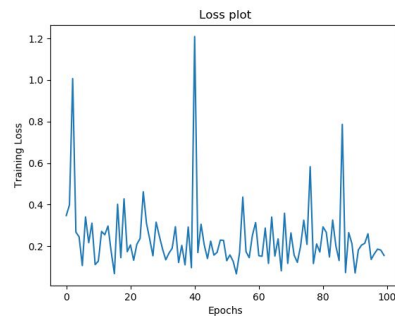
FCN NYUv2



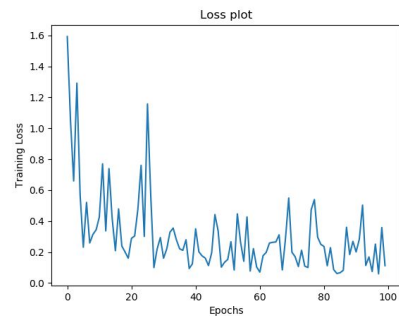
SegNet NYUv2



FCN Cityscapes



SegNet Cityscapes



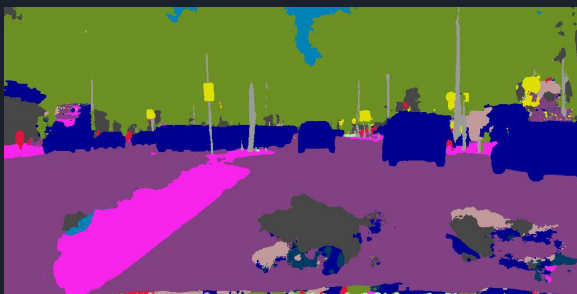
Results: Segmentations



Test
Image



FCN



SegNet





Critical Evaluation

- FCN performed better than SegNet across both datasets
- NYUv2's train/test split was 50/50 with 1,499 images in total which was not ideal
- Cityscapes performed better as more data was available and the train/val/test split left more images for training
- A reduced learning rate or a learning rate scheduler would likely improve the results for both networks as well as additional training epochs



Conclusions

- We successfully performed semantic segmentation using FCN and SegNet on two datasets
- Given the time spent on training and the hyperparameters used the results were acceptable