Improving Image Semantic Segmentation with RGB-Depth

Mengxin Zhang, Jiayuan Li, Xinyu Huang, Jiqian Dong

{mengxinz, jiayuanl, xinyuh1, jiqiand}@andrew.cmu.edu

Problem settings

Specific Goals:

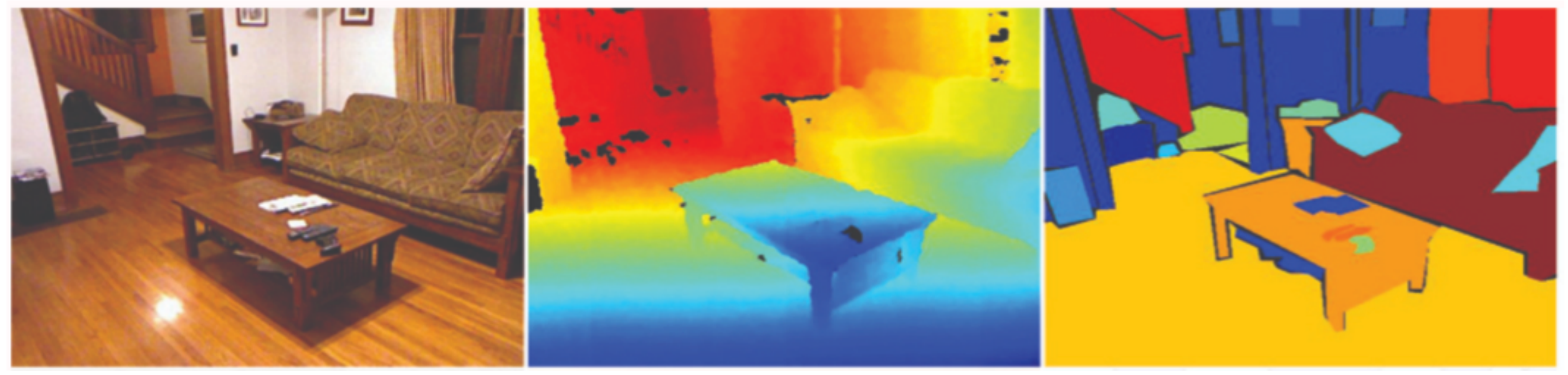
- Comparing performance of image semantic segmentation between RGB and RGBD dataset.
- Test out the performance of an existing method (U-Net) on large number class segmentation.
- Improve the segmentation by fine tuning and changing number of classes.
- Post processing by showing depth visual effects.

Challenges:

- The fusion of depth channel with original RGB channels.
- The less training images but large number of classes.
- Unbalanced labeled pixels, most are background.
- Model complexity, memory consuming.

Dataset

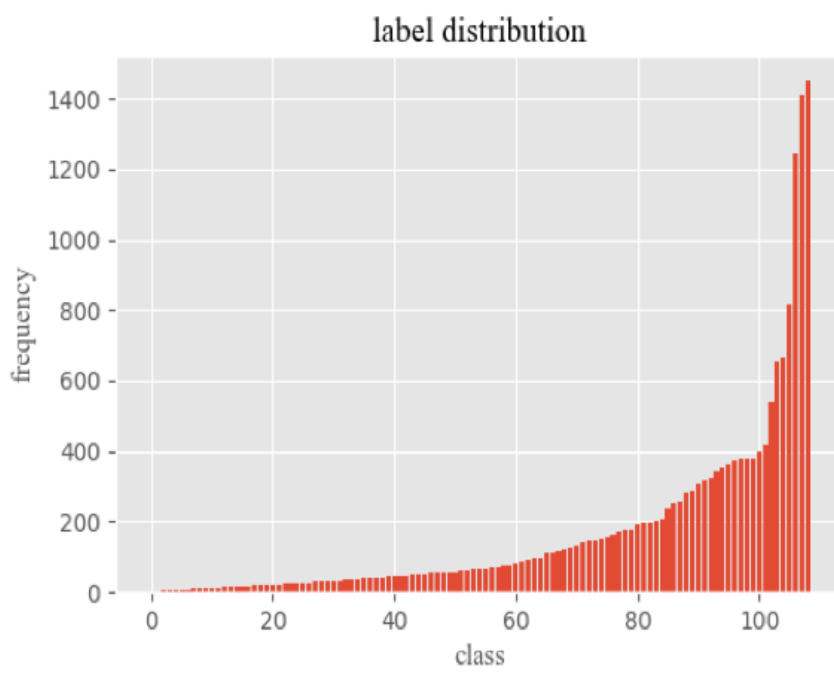
NYU Depth V2



RGB image depth image Labels

Features:

- 1449 RGB-D images of size 480 (height)×640 (width).
- Composed of video sequences from indoor scenes.
- Pixel level labels with totally 894 different classes.



Top 10 classes

Rank	Frequency	Item
1	1449	background
2	1409	wall
3	1246	floor
4	818	picture
5	664	cabinet
6	654	chair
7	539	table
8	417	window
9	400	door
10	380	pillow

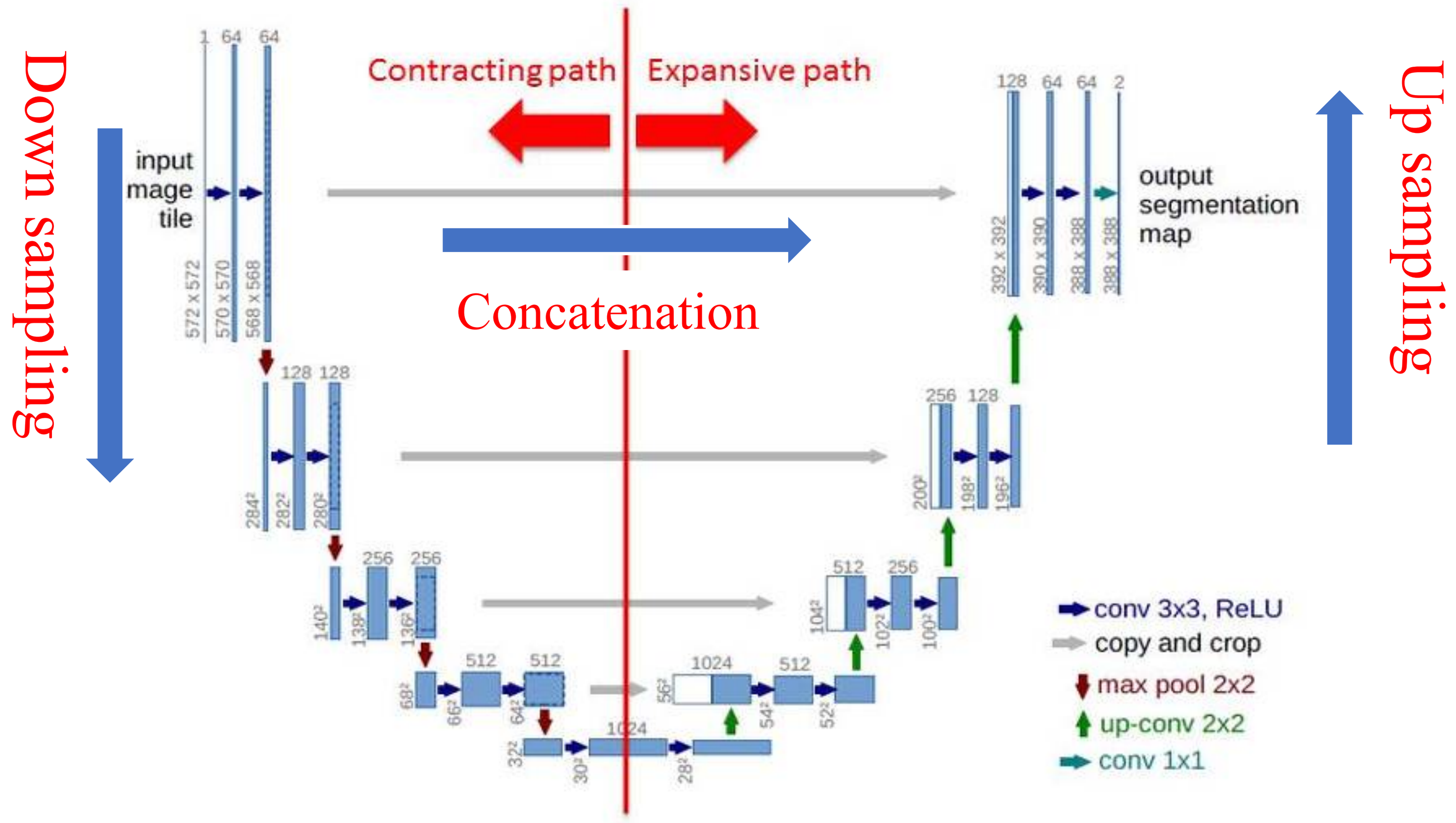
Method

U-Net

Features:

- Layers: Fully convolutional network
- 3 paths: Down sample + Up sample + Concatenation
- Output: Pixel wise feature map, # channel = # classes

Architecture:



Evaluation Metrics:

- Weighted IoU: IoU weighted by the frequency of classes in the original image labels.
- F1 score + AUC
- Pixel-wise accuracy

Results

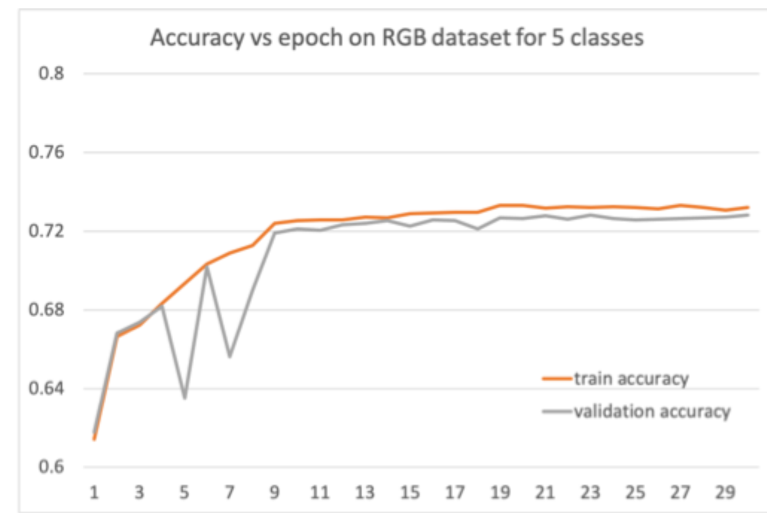
Overall settings:

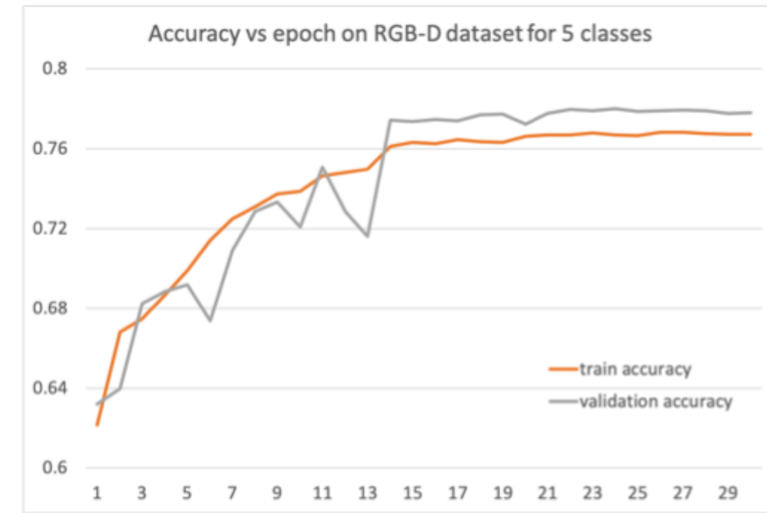
- Comparing raw performance on RGB (baseline) and RGBD (test case).
- Concatenate RGBD into 4 channels
- Train test 4:1, SGD + Adam
- Select 10, 5 most frequent classes for prediction

Table 1. Performance over different settings

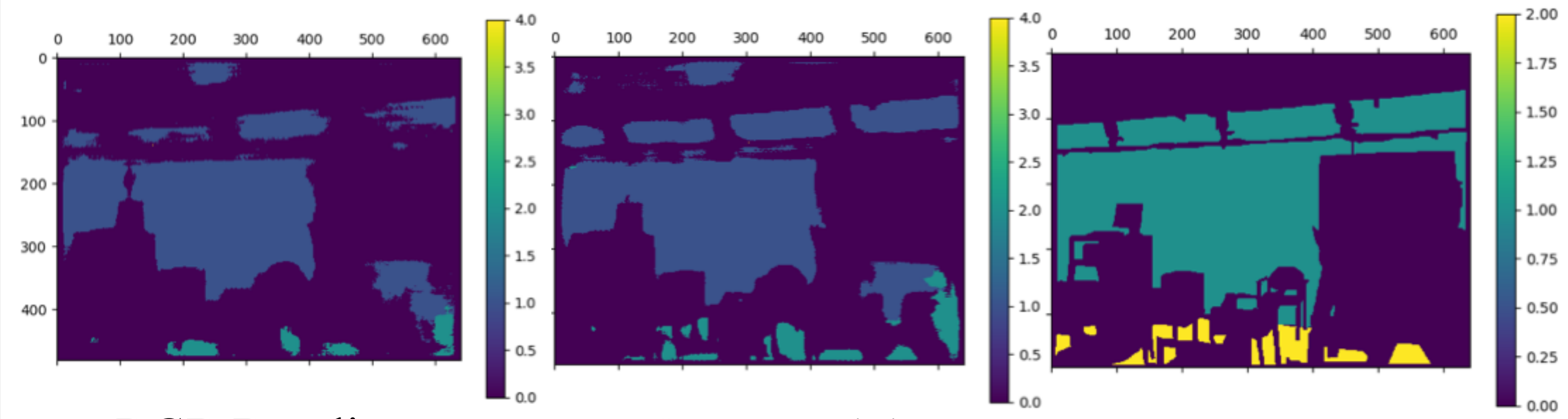
Image set	IoU		F1 Score		AUC	Accuracy
	mean	std	mean	std		
RGB 5	0.63	0.16	0.75	0.12	0.944	0.73
RGBD 5	0.67	0.17	0.79	0.12	0.958	0.78
RGB 10	0.55	0.18	0.69	0.14	0.945	0.64
RGBD 10	0.55	0.19	0.69	0.15	0.944	0.68
RGB 895	-	-	-	-	-	0.42
RGBD 895	-	-	-	-	-	0.46

Accuracy comparison:





Sample prediction:



RGB Baseline RGBD Model Ground truth

Conclusion

- U-Net is suitable for multi class semantic segmentation.
- Performance drops when class number increases.
- Add depth information makes the boundary clear, enhance the performance.

Bonus: Post processing

Application of RGB-D segmentation:

- Generate depth-of-field effect more natural based on depth change per pixel
- Ability to select which object we want to show clearly, and how much blurring we want to apply