Deep Learning course

The report, homework 3, part 1

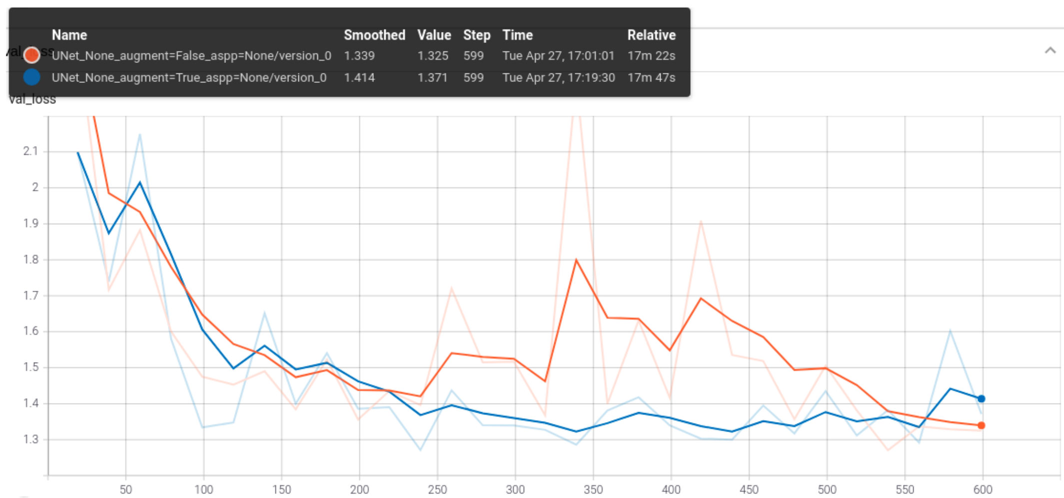
Skoltech, 2021

Task 1.

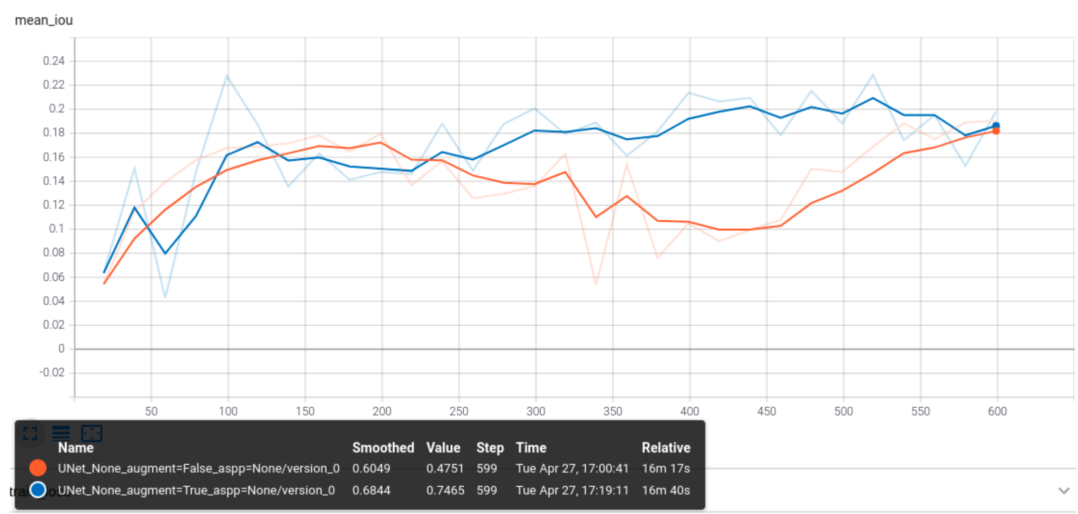
Visualize training loss and validation loss curves for UNet trained with and without data augmentation. What are the differences in the behaviour of these curves between these experiments, and what are the reasons?

Red - without augmentation, Blue - with augmentation

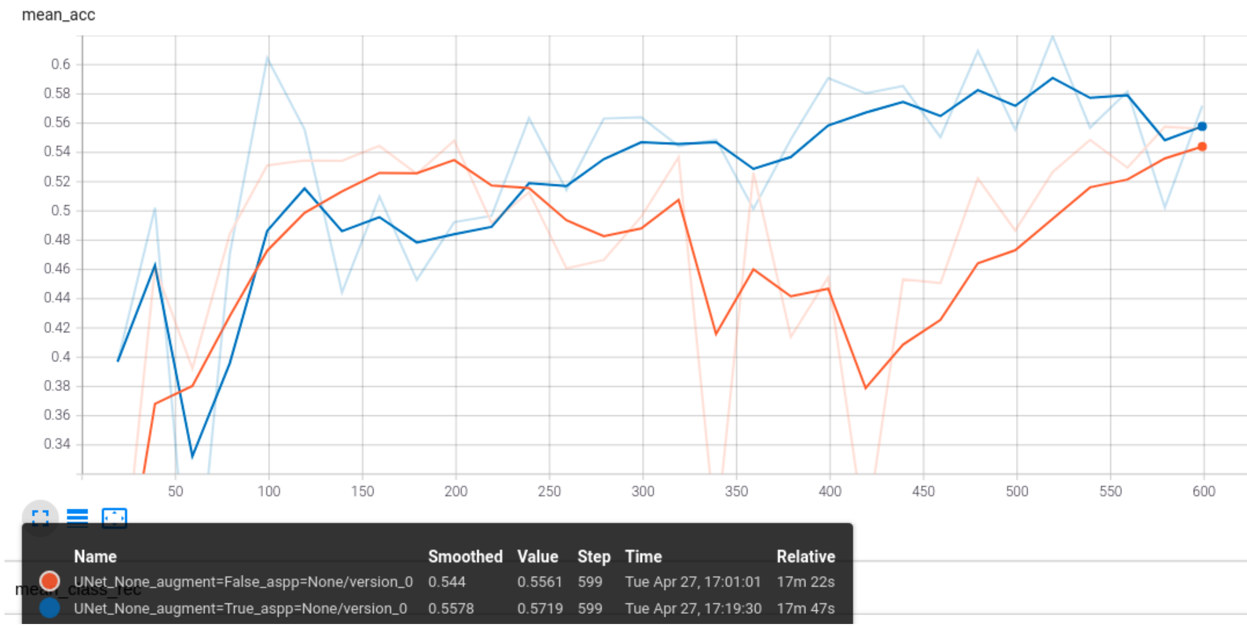
Cross-entropy loss:



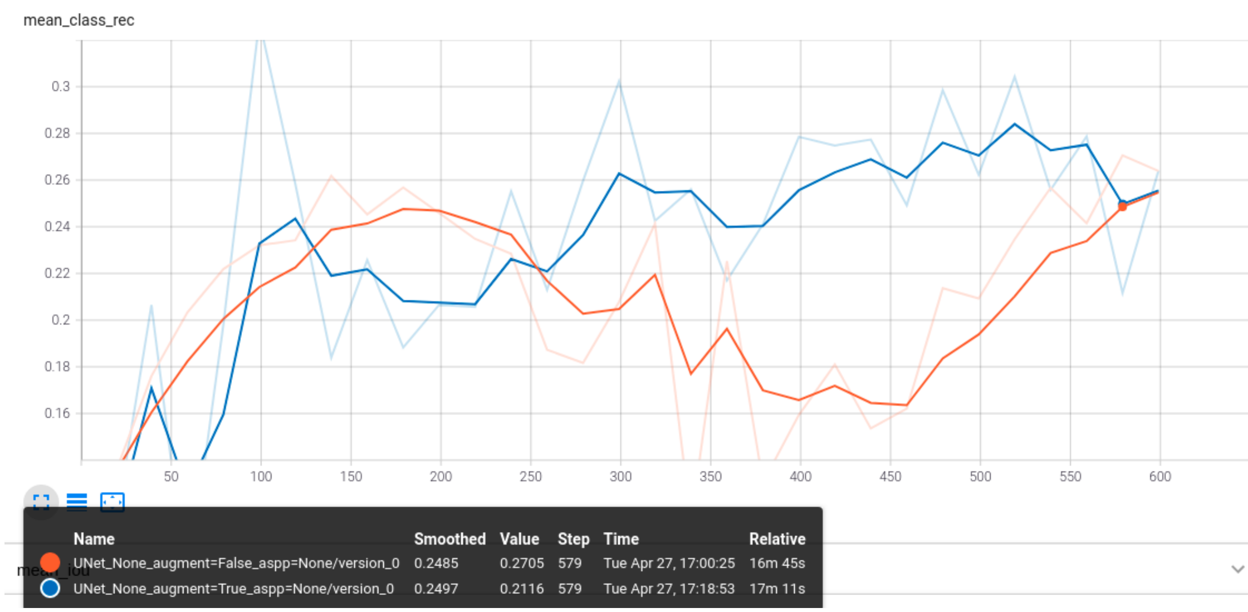
Mean IoU:



Mean accuracy:



Mean class recall:



For the network without data augmentation, the metrics are worse after several epochs and loss doesn't decrease all the time. Perhaps overfitting is observed here, but to be sure, it is better to use a training loss graph (this graph wasn't added here, but on this graph, loss decreases all the time, therefore we can observe overfitting). After using augmentation loss is calmer, there is no overfitting here.

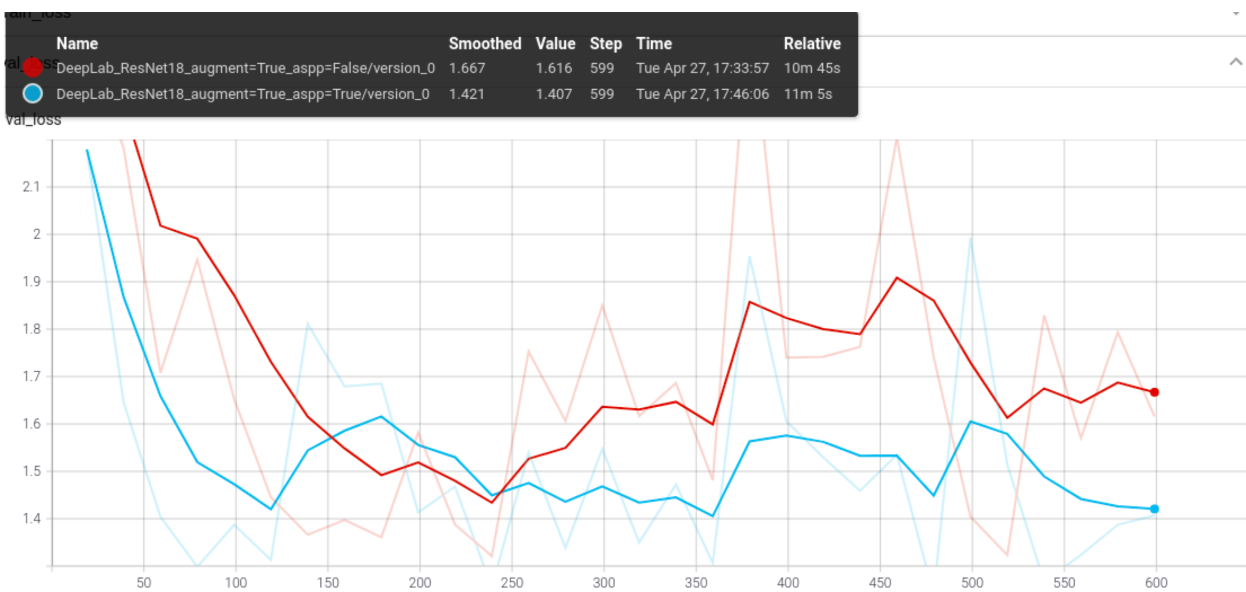
Task 2.

Visualize training and validation loss curves for ResNet18 trained with and without ASPP. Which model performs better?

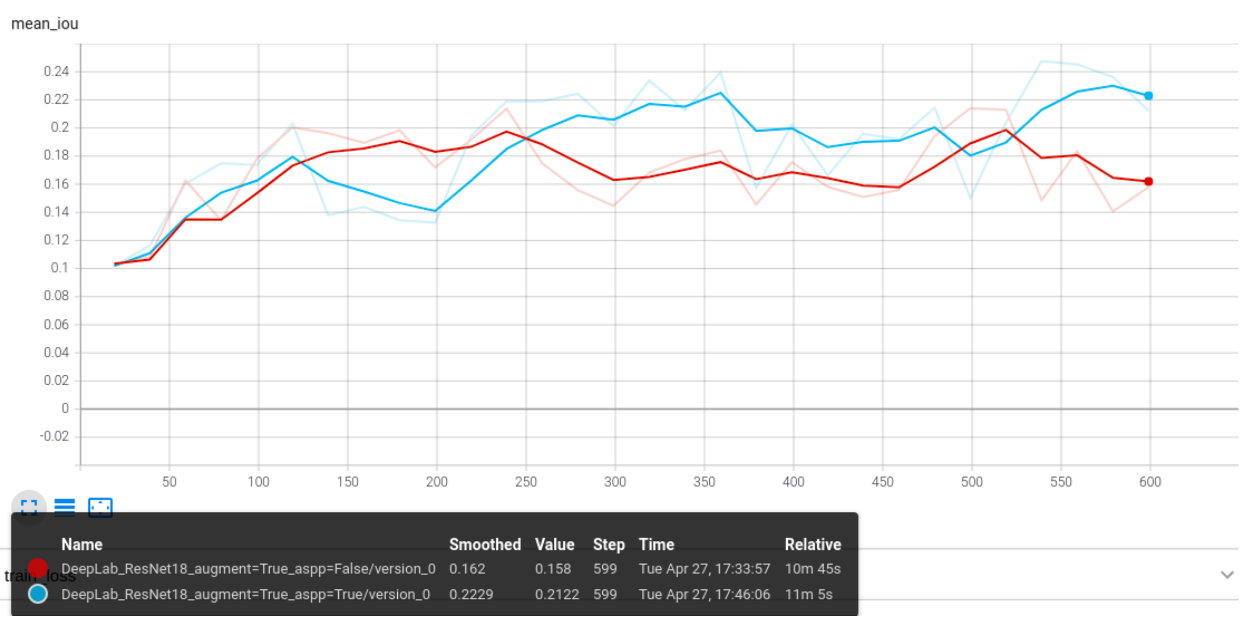
Red - without ASPP

Blue - with ASPP

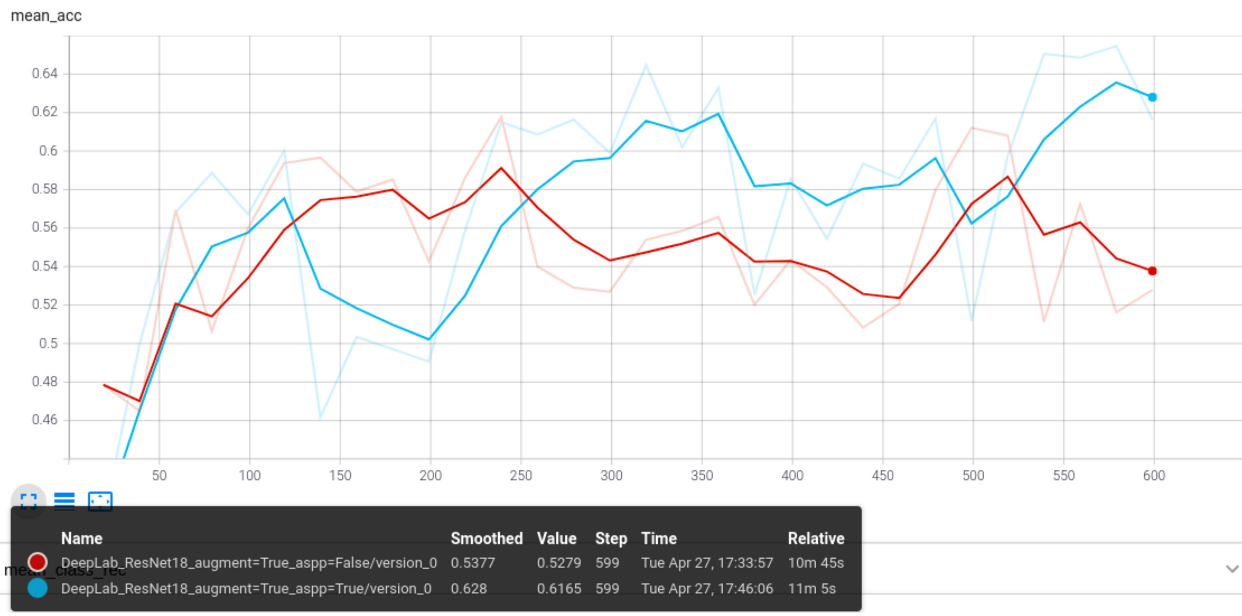
Cross-entropy loss:



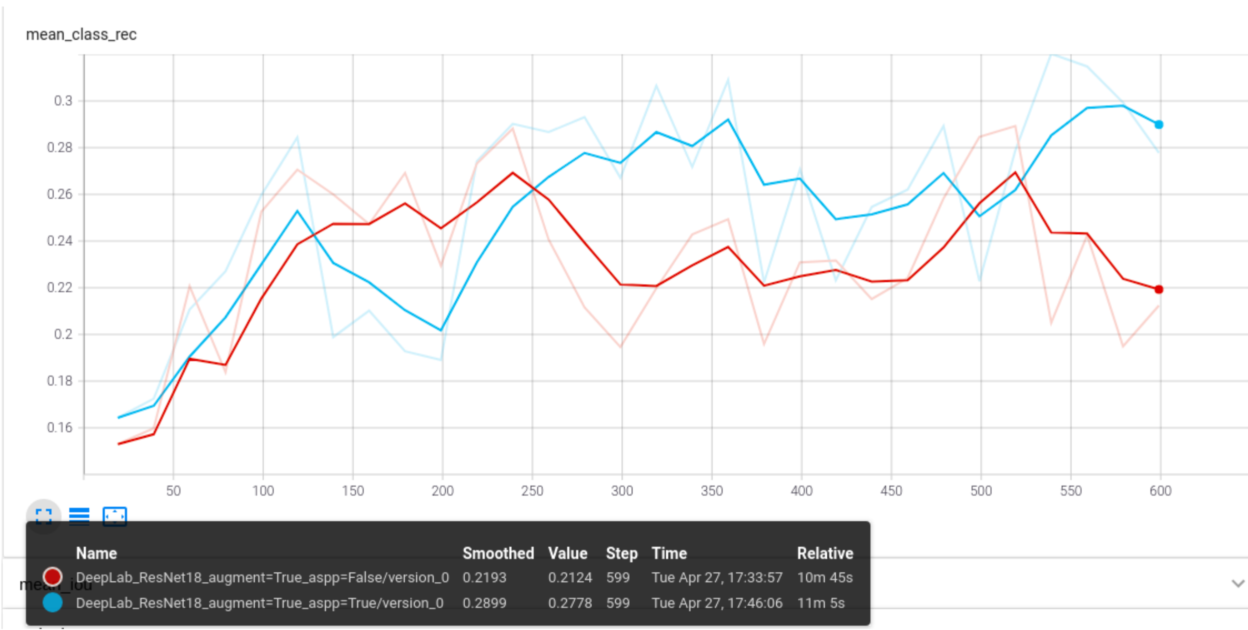
Mean IoU:



Mean accuracy:



Mean class recall:



ASPP model performs better, especially on the last epochs (in accordance with all metrics).

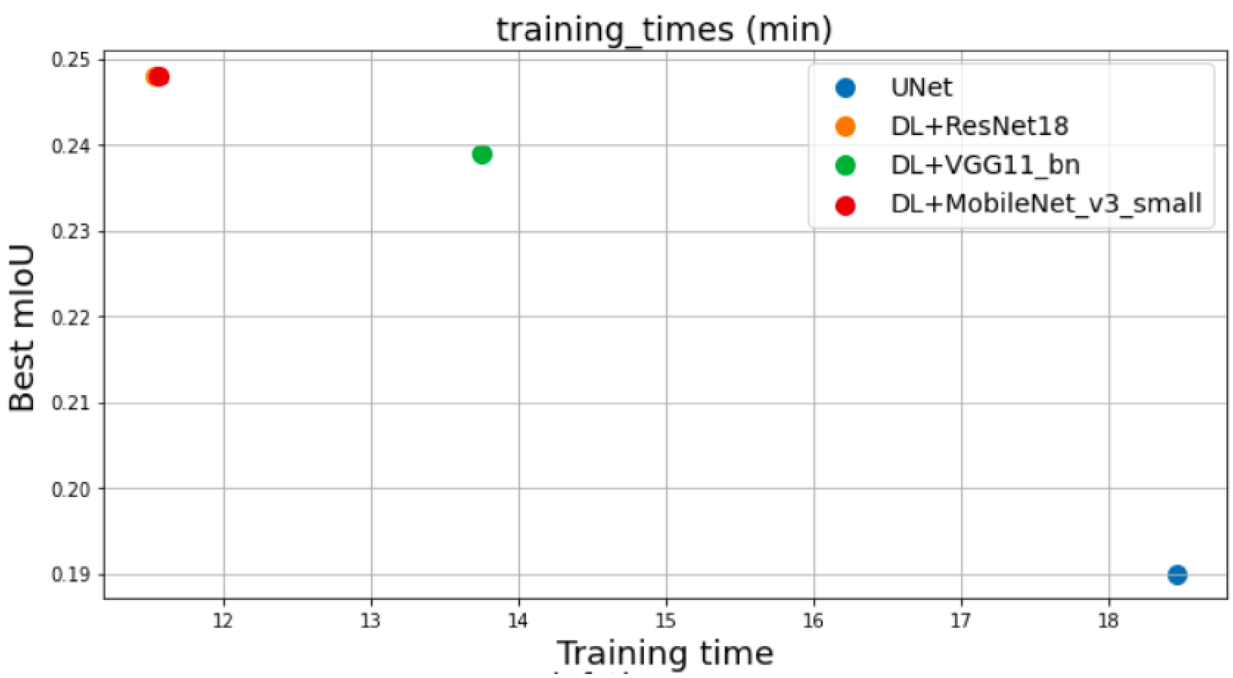
Task 3.

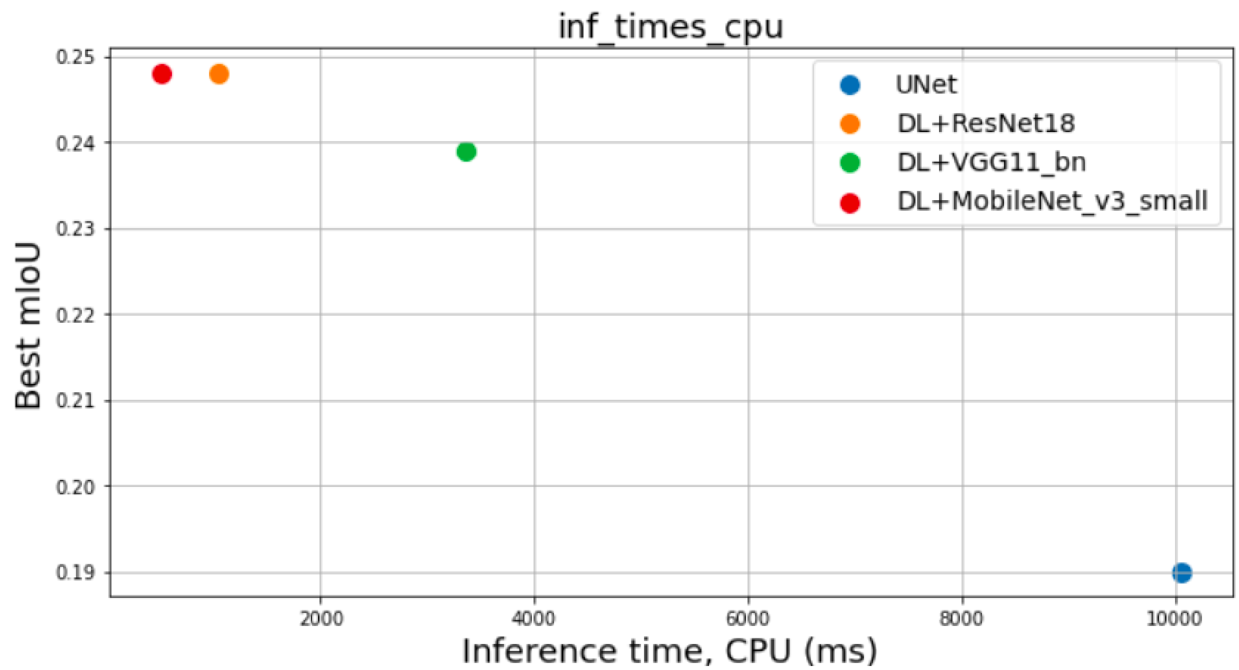
*Compare **UNet** with augmentations and **DeepLab** with all backbones (only experiments with **ASPP**). To do that, put these models on three scatter plots. For the first plot, the x-axis is **training time** (in minutes), for the second plot, the x-axis is **inference time** (in milliseconds), and for the third plot, the x-axis is **model size** (in megabytes). The size of each model is printed by PyTorch Lightning. For all plots, the y-axis is the best **mIoU**. To clarify, each of the **4** requested models should be a single dot on each of these plots.*

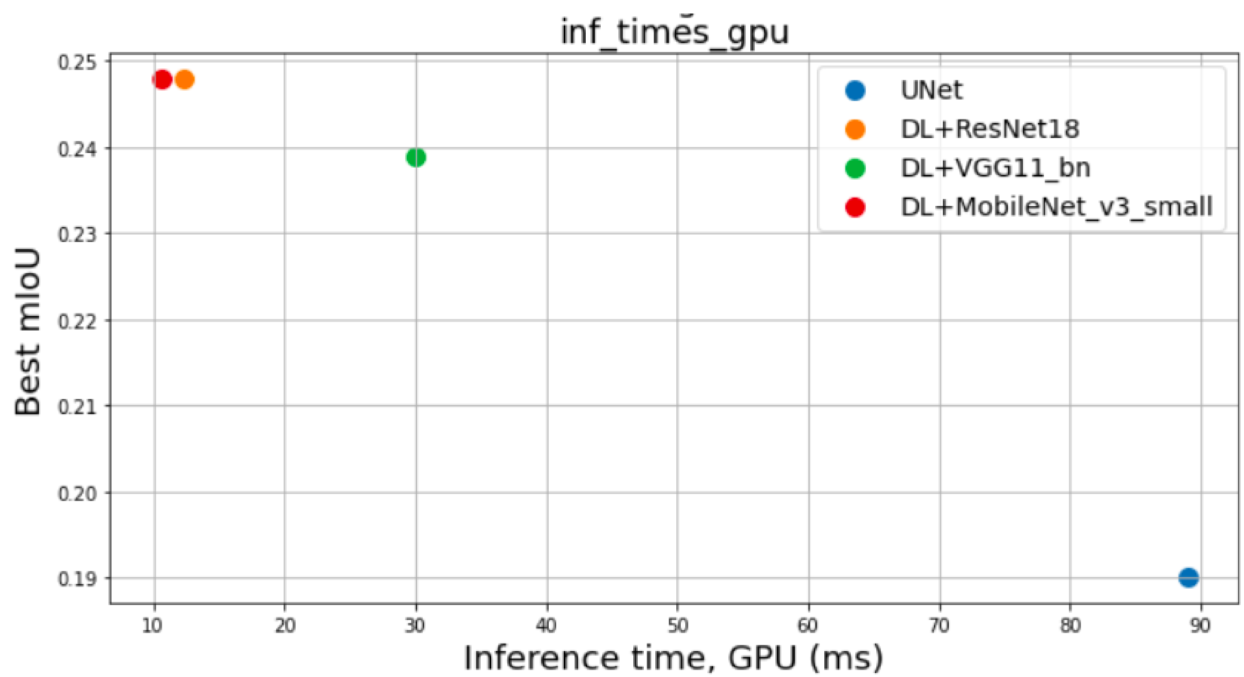
Which models are the most efficient with respect to each metric on the x-axes? For each of the evaluated models, rate its performance using their validation metrics, training and inference time, and model size. Also for each model explain what are its advantages, and how its performance could be improved?

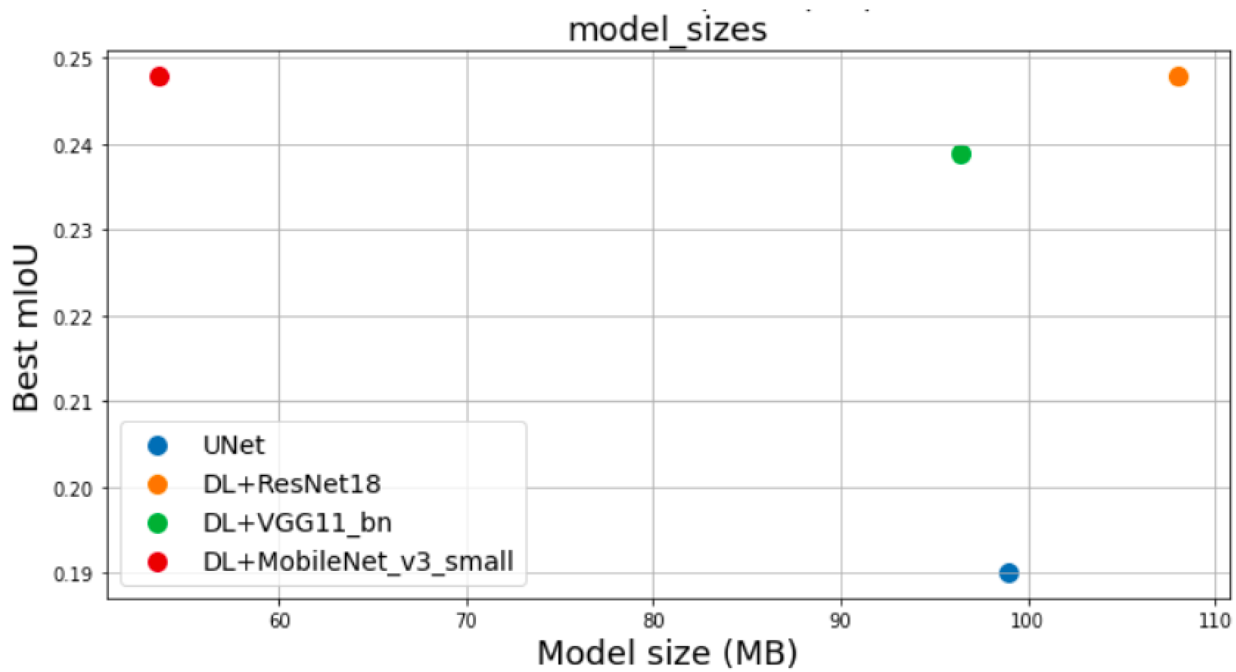
I've plotted a bit more (inference time for both GPU and CPU).

In this case, mIoU and training time are almost the same for 1 and 3 motions, so they merged into 1 point:




inf_times_cpu
Best mIoU
Inference time, CPU (ms)
0.25
0.24
0.23
0.22
0.21
0.20
0.19
2000
4000
6000
8000
10000
UNet
DL+ResNet18
DL+VGG11_bn
DL+MobileNet_v3_small


inf_times_gpu
Best mIoU
Inference time, GPU (ms)
0.25
0.24
0.23
0.22
0.21
0.20
0.19
10
20
30
40
50
60
70
80
90
UNet
DL+ResNet18
DL+VGG11_bn
DL+MobileNet_v3_small



Model's performance (descending order). The 1st in the list - the most effective with respect to each metric:

From a model size point of view:

1)DeepLab+MobileNet_v3_small

2)DeepLab+VGG11_bn

3)UNet

4)DeepLab+ResNet18

From GPU and CPU inference time point of view:

1)DeepLab+MobileNet_v3_small

2)DeepLab+ResNet18

3)DeepLab+VGG11_bn

4)UNet

From a training time point of view:

1)DeepLab+MobileNet_v3_small **and** DeepLab+ResNet18 (almost the same)

2)DeepLab+VGG11_bn

3)UNet

From the best mIoU point of view:

1)DeepLab+MobileNet_v3_small **and** DeepLab+ResNet18 (almost the same)

2)DeepLab+VGG11_bn

3)UNet

Model's advantages and how to improve performance:

1)DeepLab+MobileNet_v3_small - has shown almost the best result in all metrics, which is an undoubted advantage.

2)DeepLab+ResNet18 - almost the same performance as in the 1st point. But the model's size is larger. To improve performance from this point of view it would be better to revise architecture and to do it less memory consuming.

3)DeepLab+VGG11_bn - lower inference and training time than Unet has. The performance can be increased by a complication of the model.

4)UNet - there is a single advantage: the model size is lower than DeepLab+ResNet18 (almost the same). The performance can be increased by adding a pre-trained network to the architecture or other complication of the model.

Task 4.

Pick the best model according to ***mIoU*** *and look at the visualized predictions on the validation set in the TensorBoard. For each segmentation class, find the good examples (if they are available), and the failure cases. Provide the zoomed-in examples and their analysis below. Please do not attach full validation images, only the areas of interest which you should crop manually.*

There are 2 models with the best mIoU:

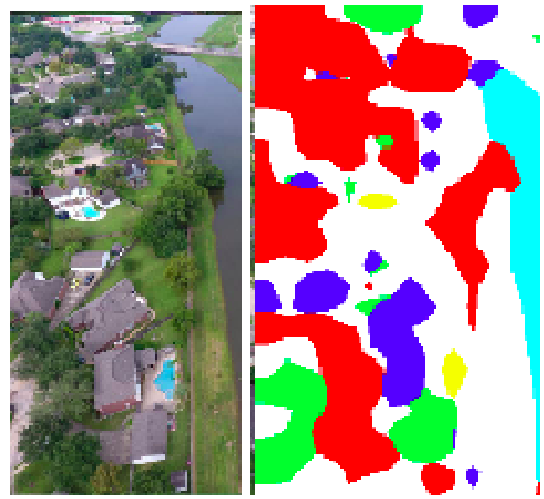
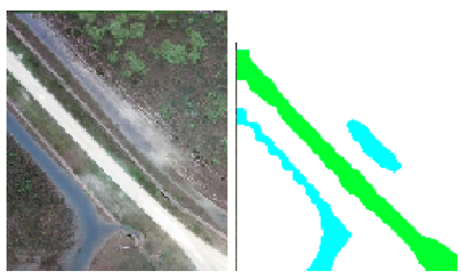
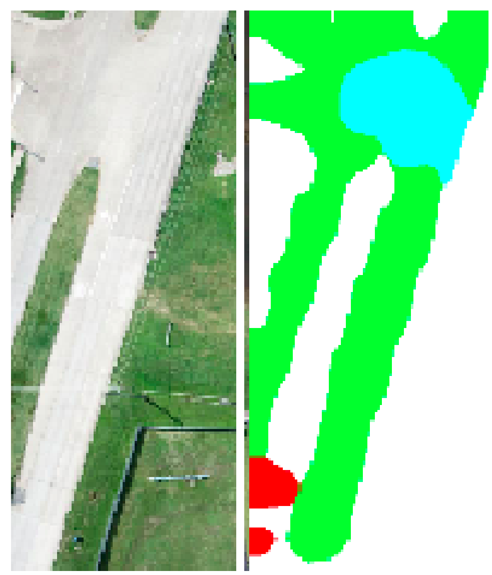
- `DeepLab_MobileNet_v3_small_augment=True_aspp=True/epoch=23-mean_iou=0.248.ckpt`
- `DeepLab_ResNet18_augment=True_aspp=True/epoch=26-mean_iou=0.248.ckpt`

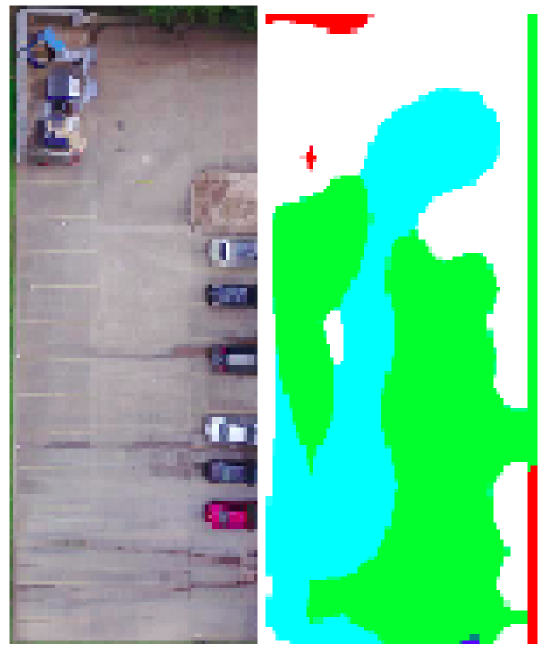
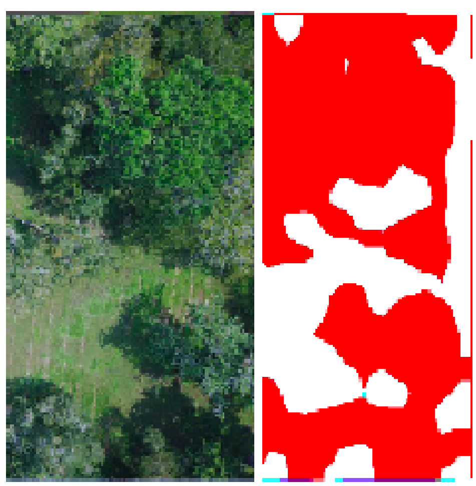
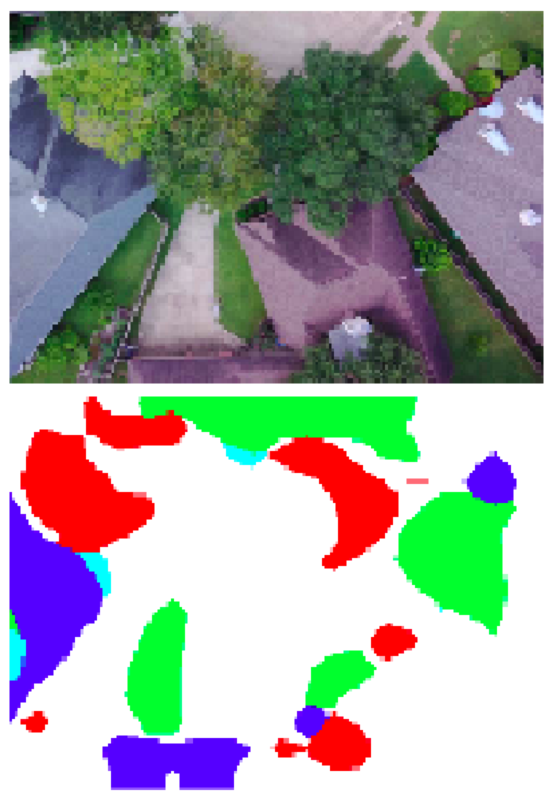
There are 8 classes:

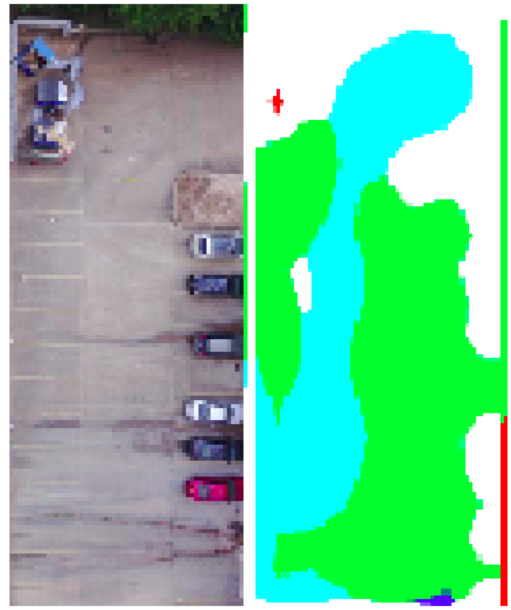
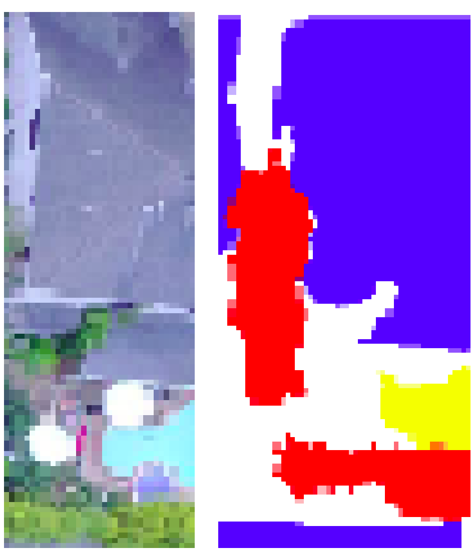
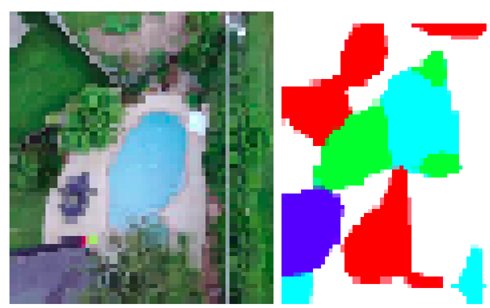
`0: Background, 1: Building, 2: Road, 3: Water, 4: Tree, 5: Vehicle, 6: Pool, 7: Grass`

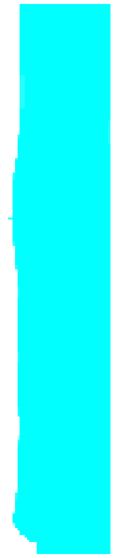
So, I've chosen the first model. Validation results:

Class	Good example	Bad example
Background (black)	NONE	
The background was detected as water. This looks like an anomaly because the similarity of water and background is quite low, at least visually.		

<table>
<tr><td>Building (blue)</td><td></td><td></td></tr>
<tr><td colspan="3">Large objects are detected well, so the house on the right was recognized correctly, but in the right picture, all objects are smaller and presented in large numbers, which causes rather random results for this class.</td></tr>
<tr><td>Road (green)</td><td></td><td></td></tr>
<tr><td colspan="3">On the left, the road is well marked, but on the right, the road is moaning like water. Perhaps this is due to the proximity of their colours.</td></tr>
</table>

<table>
<tr><td>Water (cyan)</td><td></td><td></td></tr>
<tr><td colspan="3">On the left, the river was recognized well, but in the right picture it was confused with the road (parking). Perhaps this is due to the proximity of their colours.</td></tr>
<tr><td>Tree (red)</td><td></td><td></td></tr>
</table>

<table>
<tr><td colspan="3">Trees are recognized well mainly on grass, while in other places the recognition quality deteriorates.</td></tr>
<tr><td>Vehicle (purple)</td><td>NONE</td><td></td></tr>
<tr><td colspan="3">The cars are small and often blend in with the terrain, so no incident has been reported.</td></tr>
<tr><td>Pool (yellow)</td><td></td><td></td></tr>
<tr><td colspan="3">On the left, the pool was recognized well, and on the right as water. This is most likely due to the fact that it is dimmer in the right picture. On the left, there is a pronounced cyan colour.</td></tr>
</table>

Grass (white)	 	NONE
The grass is well recognized everywhere since it usually occupies a large area and has a sharp contrast with other objects.		