

Everywhere ~ 65 epochs

Params: lr 10e-4, scheduler MultiStep, gamma 0.3, steps - 20, 40, 60

1.

UNet without augmentation, Inference time (per frame): 50.640 ms



UNet with augmentation, Inference time (per frame): 49.413 ms

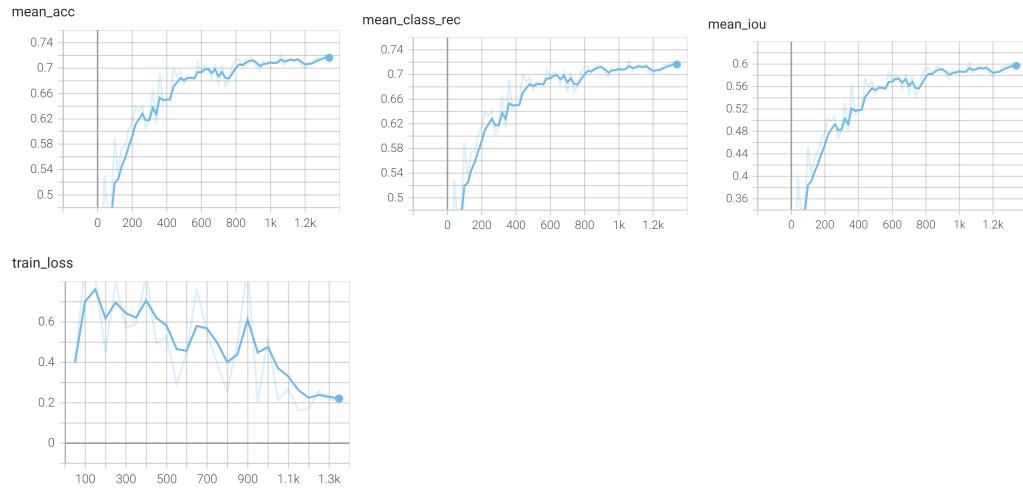
Inference time is roughly the same (which is logically right), accuracy is better in case of augmentation (via all the metrics), because of the idea of augmentations - to provide a greater amount of training examples through different image modifications (better representation in a result) and graphs show that it works, but loss curves are rather similar. Nevertheless, I think that in a perfect experiment loss curve on augmented data will converge faster and to the lower numbers because of the same reasons - more training examples = more predictive capability in bigger number of cases.

2. Deeplab: Resnet without ASPP:

Inference time (per frame): 533.605 ms

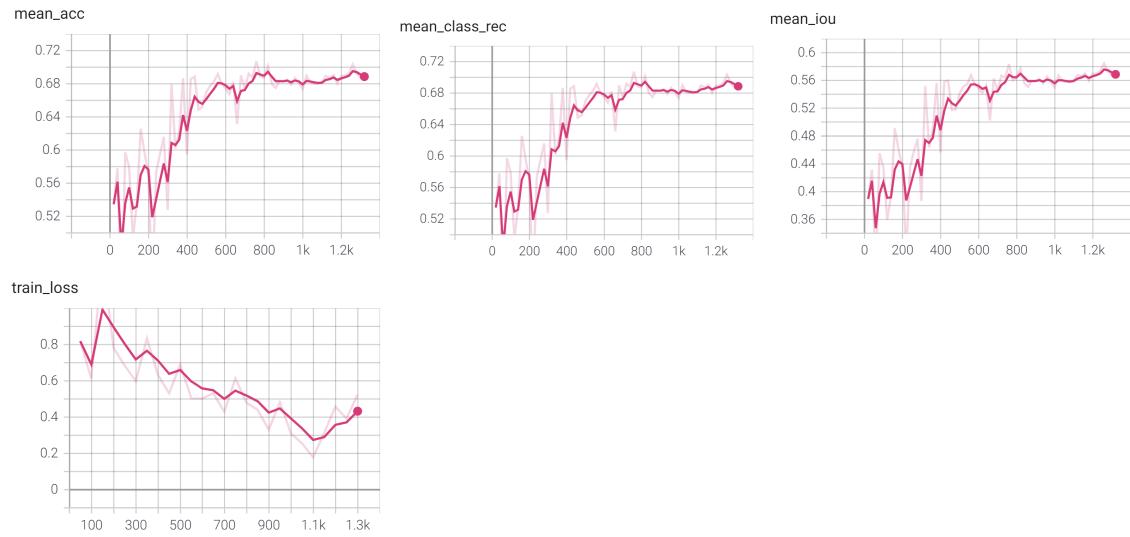
Accuracy is better than in UNet (pretrain really makes sense!), also every experiment here and below with augmentations (which showed itself as a permissible trick). Maybe bilinear upsampling is also a nice thing for accuracy. Loss converges much better and faster as well as accuracy (now it beats 0.7 and in UNet it was nearly 0.5). But surely, inference time is bigger.

As for loss comparison between two models (with and without ASPP) a better loss curve is in case of a model without ASPP.



### Deeplab: Resnet with ASPP

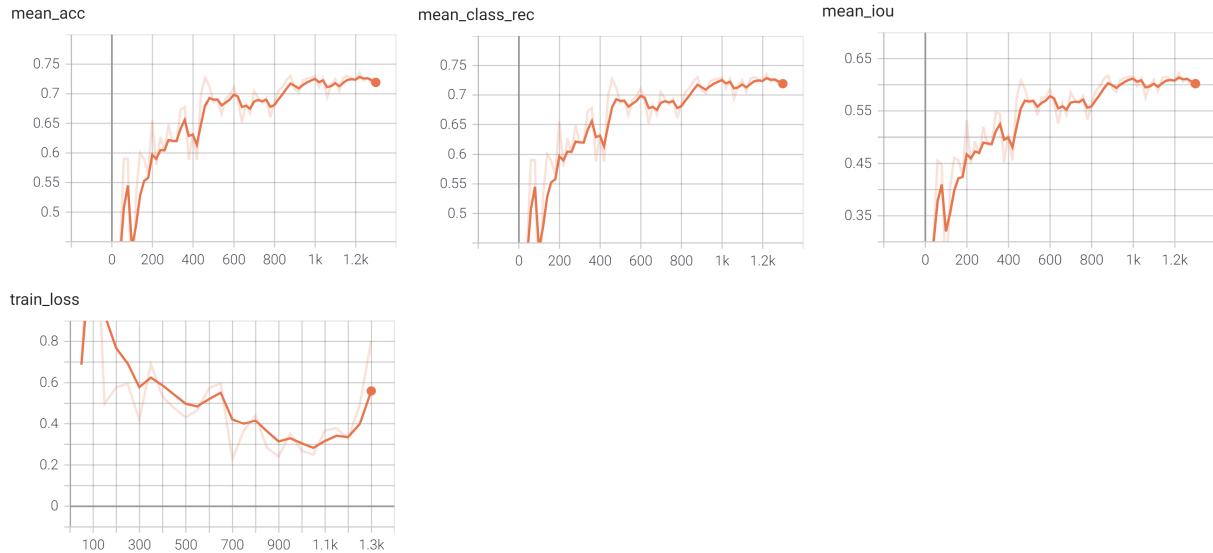
Inference time (per frame): 546.139 ms



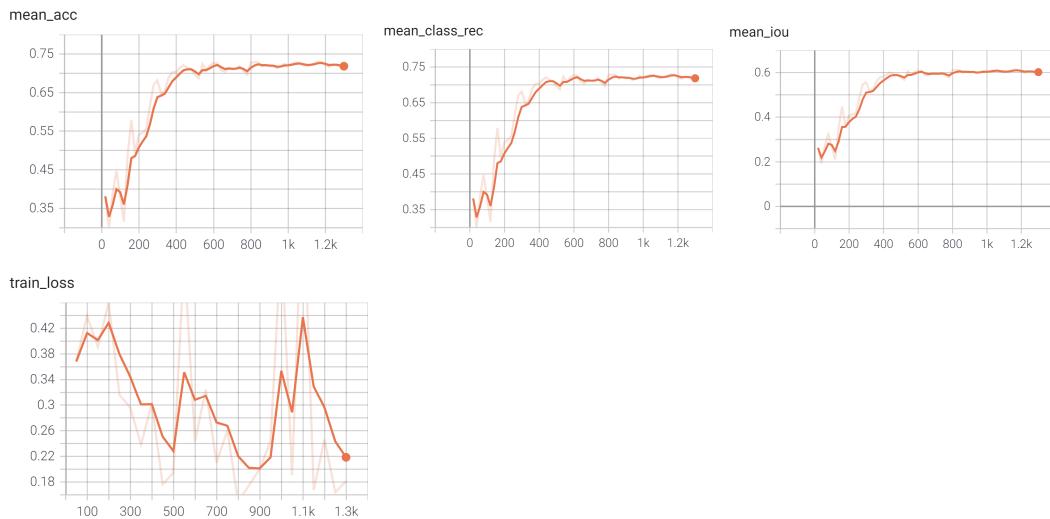
### Loss curves comparison

3. VGG Training time: 18.769 minutes

Inference time (per frame): 1835.703 ms

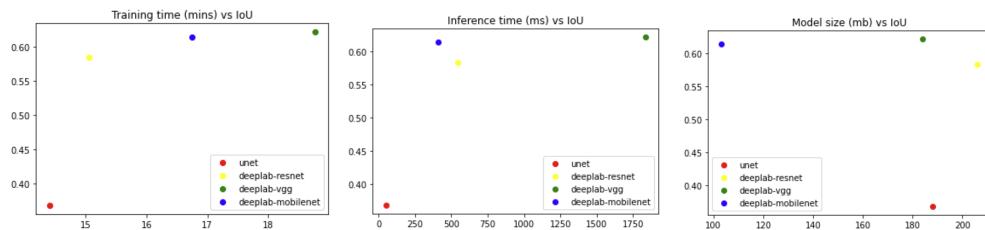


MobileNet Training time: 16.747 minutes  
Inference time (per frame): 409.144 ms



Comparative table (training time, inference time, model size, best IoU)

- Unet with augmentations (14.402 mins, 49.413 ms, 188 mb, 0.368)
- DeepLab ResNet18 with ASPP (15.059 mins, 546.139 ms, 206 mb, 0.584)
- DeepLab VGG with ASPP (18.769 mins, 1835.703 ms, 184 mb, 0.622)
- DeepLab MobileNet with ASPP (16.747 mins, 409.144 ms, 103 mb, 0.614)



In terms of training time Unet is the best, VGG backbone is the worst, ResNet is somewhat optimal in terms of both accuracy and training time.

In terms of inference time UNet with augmentations is the best one, but if to choose using trade-off accuracy-inference time, then MobileNet is the leading one: accuracy is 2 times bigger than for Unet, inference time is 10 times bigger than for Unet, but slightly smaller compare with ResNet backbone and much smaller than with VGG.

In terms of model size Mobile Net is the best one: it has the smallest model size and nice iou compared with VGG backbone and much better than ResNet and Unet.

How models can be improved?

UNet (mostly need it in terms of quality) - attention blocks can increase an accuracy as was mentioned in <https://arxiv.org/abs/1804.03999>

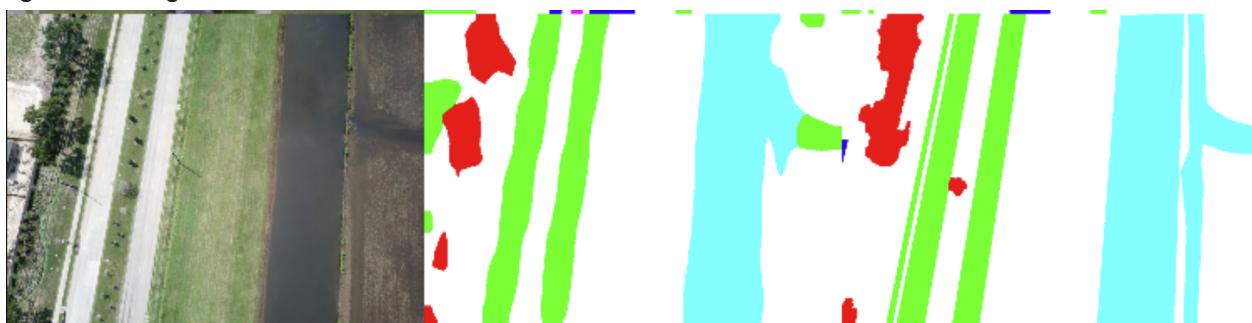
Resnet and VGG - its variations with more layers probably can hold with accuracy (to make internal representations more accurate and meaningful), but it will be also an increase in terms of memory consumption (there can help network compression techniques - pruning, quantization, maybe some extra ways such as matrix factorization <https://openreview.net/pdf?id=B1eHgu-Fim>)

MobileNetV3 - wait for MobileNetV4 :)

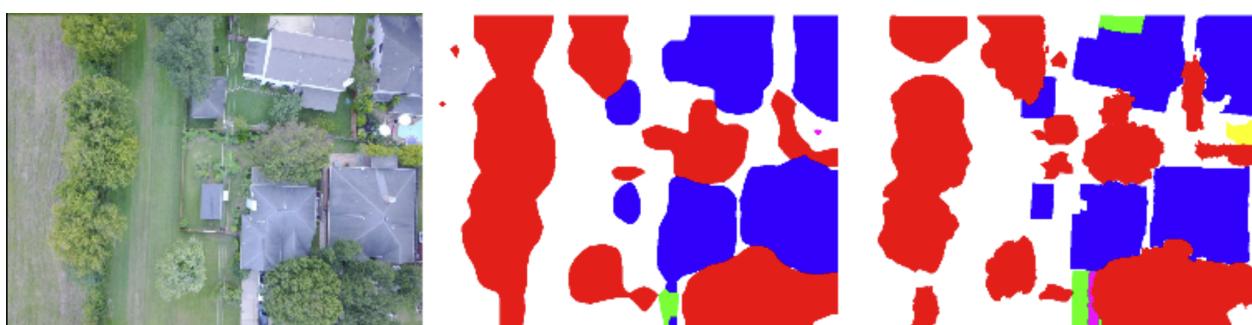
4. I took VGG with the highest mean IoU

Nice examples:

- green and light blue



- red and blue

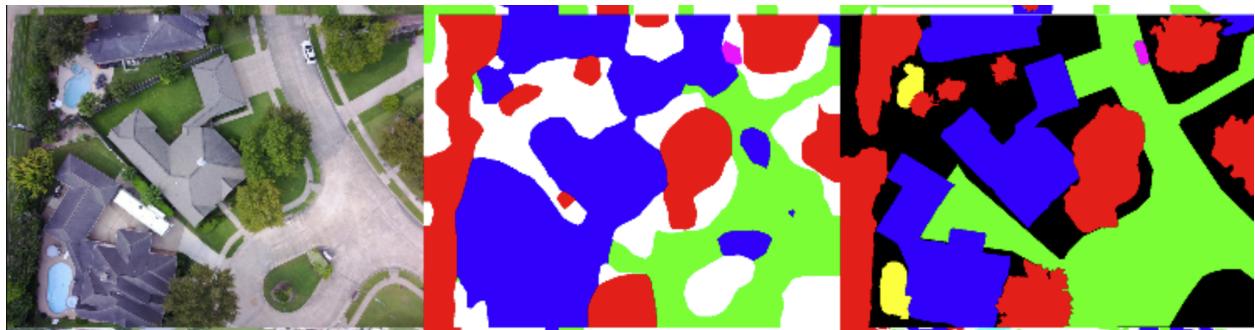


Not so good examples

- purple



- Yellow



Analysis - small images are more difficult to segment (they have small area and little missegmentation gives not so big impact into total IoU)

Bigger images are easier to segment, but they can have more sharp or blurry edges, because the tendency is clear but not so overfitted to make a perfect score on it.