Fully convolutional networks for Image Segmentation

Laura Munar Acosta
Universidad de los Andes
l.munar10@uniandes.edu.co

Maria Ana Ortiz
Universidad de los Andes
ma.ortiz1@uniandes.edu.co

1. Methodology

For this report we run FCN-32s from scratch and from VGG weights and FCN-16s from VGG weights and from 32s weights. It is important to notice that only 6 epochs were tried in all the models.

2. Results

Table 1 summarizes the metrics obtained in each of the trained methods.

Net	Pixel acc	mean acc	mean IU	f.w
FCN-32s-Vgg	0.89	0.72	0.59	0.82
FCN-32s-scratch	0/72	0.04	0.034	0.53
FCN-16s-Vgg	0.905	0.63	0.59	0.834
FCN-16s-FCN32s	0.91	0.76	0.65	0.84

Table 1. Metrics obtained from models over val set. All models were trained for 6 epochs

2.1. FCN 32s - from VGG weigths

Results from training process show on validation set over iterations can be observed on Figures 1-2 and a result on demo image can be observed on page 3

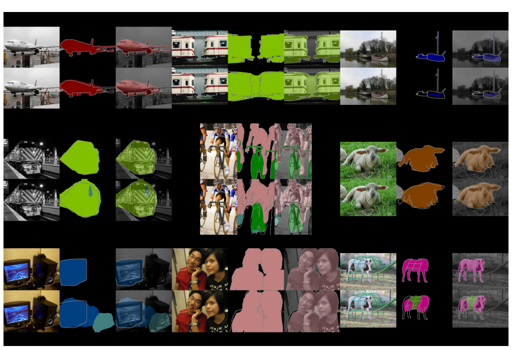

Figure 1. Model FCN-32s-from vgg weights half of training process

2.2. FCN 32s - from scratch

Results from training process show on validation set over iterations can be observed on Figures 4-5.

Figure 2. Model FCN-32s-from vgg weights last model

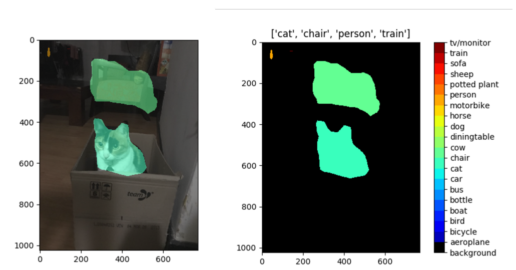

Figure 3. Test of image on the wild

Figure 4. Model FCN-32s-from scratch half of training process

2.3. FCN 16s - from VGG weigths

The Figure 6 shows some examples of the model at half way of training. Where it can be seen that the model still had some flaws performing segmentation.

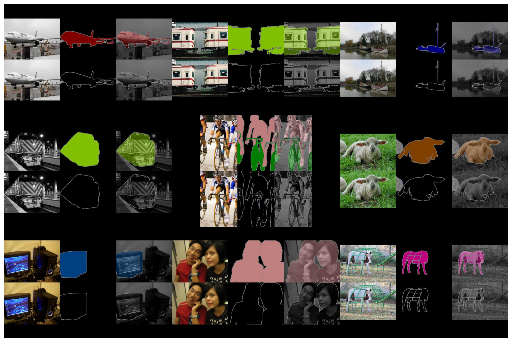

Figure 5. Model FCN-32s-from scratch last model

On the other hand, Figure 7 presents the final segmentation achieved by the model.

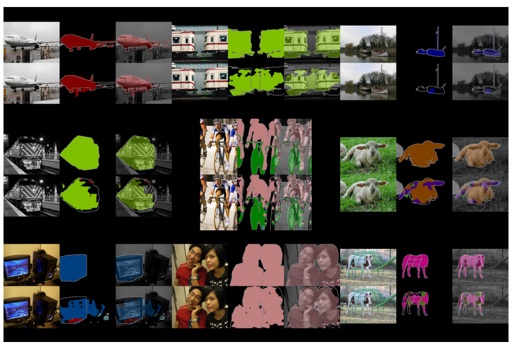

Figure 6. Model FCN-16s-from vgg weights at half of training process

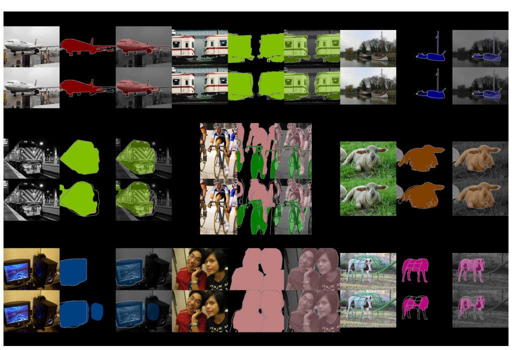

Figure 7. Last model FCN-16s-from vgg weights

2.4. FCN 16s - from 32FCNs

In the Figure 8 is presented the types of results that this model provides when its half way its training process. On the other hand, Figure 9 presents the segmentation produced at the final level of the training.

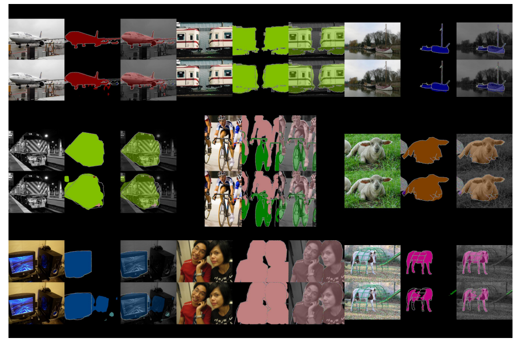

Figure 8. Model FCN-16s-from FCN32s at half of training process

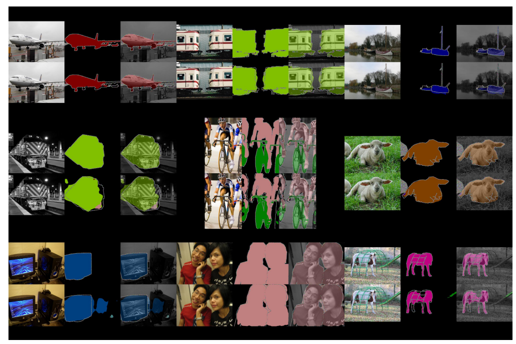

Figure 9. Last model FCN-16s-from FCN32s

3. Proof of concept: Best model with in the wild pictures

In the Figures 10, 11 and 12 are presented some examples of the best model (FCN16s pretrained with FCN32s). In these pictures, it can be seen that the model mostly recognizes the animals but makes some mistakes identifying what it is on the background.

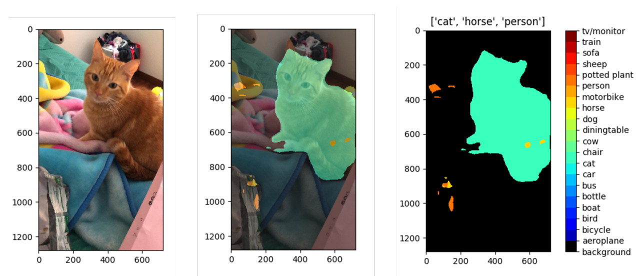

Figure 10. In the wild image segmentated with the best model: Example 1

4. Discussion

As it can be seen in Table 1, the models that used FCN16s presented better results. The latter is explained because of the combination of the final layer with lower layers. This represents more information to the final upsampling, which results in better definition in the final mask [1]. This

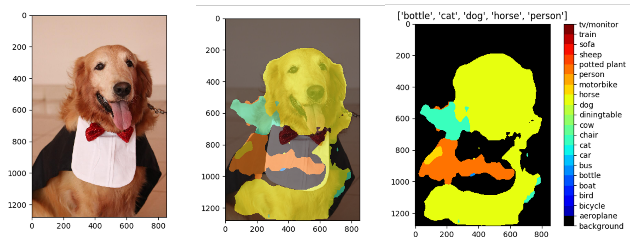

Figure 11. In the wild image segmentated with the best model: Example 2

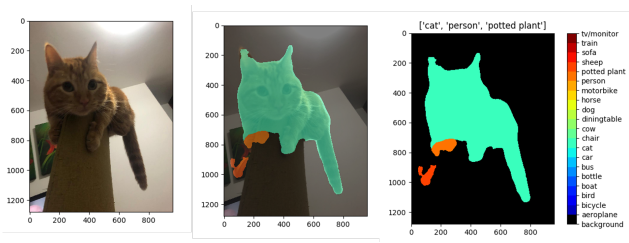

Figure 12. In the wild image segmentated with the best model: Example 3

merge of layers from high level with layers from low levels makes that the final mask to have details that a FCN32s model had lost in poolings [1].
Training a net with pretrained weights results on a better performance, because of the fact that improves the convergence of the models. These pretrained weights are closer to the answer than random weights. Thus, the convergence can be improved. To illustrate that, Table 1 shows that using the pretrained model from VGG provided better results than training from scratch. It is possible that six epochs were not enough for the from scratch model to be trained as well as the other models. On the other hand, the best performance was the model FCN16s pretrained with the FCN32s weights. This was a expected behaviour because the pretrained model used is more similar to the FCN16s than the VGG model. That is, because both solve the same task. This means that the convergence of the models is faster than the one trained with VGG weights, because the initialization is closer to the final outcome.

References

[1] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.