

3rd place solution (U-Net + Dilated Conv)

posted in [Carvana Image Masking Challenge](#) 5 months ago



61

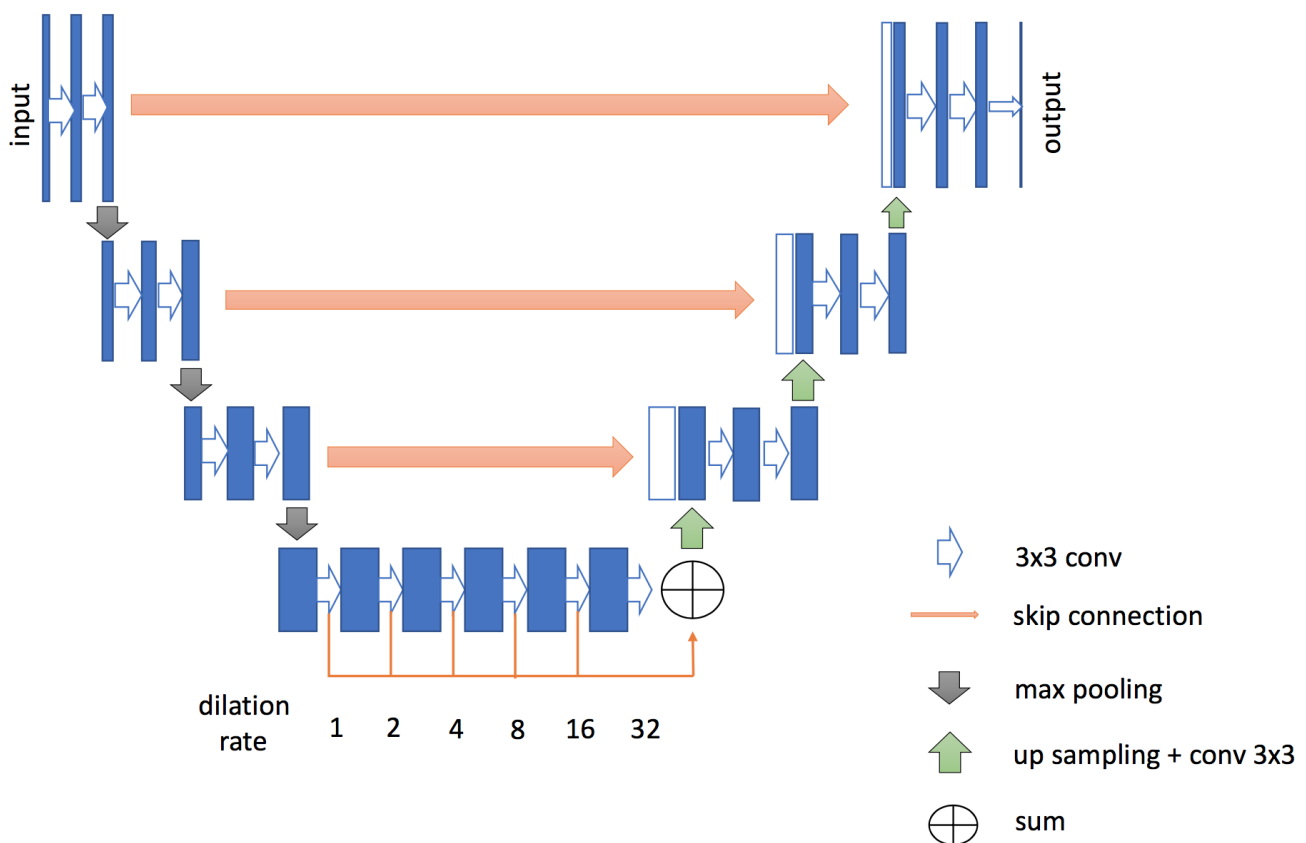
Thanks for hosting such an exciting competition! I was really enthusiastic for spending my time for this competition! And thanks for helpful code by @Peter and excellent ideas by @HengCher Keng. I learned a lot from them.

The competition repository is here. I put two scripts (My network script & loss functions script) in it. <https://github.com/lyakaap/Kaggle-Carvana-3rd-place-solution>

My solution overview

- I used 1536x1024 & 1920x1280 resolution.
- I used modified U-Net. It has several dilated convolution layers in bottleneck block. (i.e. where the resolution of feature maps are lowest)

Detailed figure of my network architecture is here.



The best score of this model is **0.997193** only around 8.5 million parameters. (trained one of 6 folds, no TTA & no ensemble, input resolution: 1920x1280) Averaging two predictions(TTA, original image & flipped image) by 0.997193 model reached **0.997222**. They are ranked 6th place and 5th place on LB respectively!

I tried normal convolution layers instead of dilated convolution layers in bottleneck block, and its score is significantly lower than using dilated convolution. (normal: 0.9905, using dilated conv: 0.9918 @256x256)

I also tried parallelized dilated convolution layers instead of stacking them, but it gave me lower score than stacked architecture.

- Optimizer: RMSprop lr = 0.0002, reducing learning rate by using ReduceLROnPlateau() that is Keras callback function. Reducing factor is 0.2 & 0.5
- Data Augmentation: only horizontal flip. Scaling, Shifting, and Shifting HSV were results of overfitting for me.
- Batchsize: 1, and no BN.
- Training whole time on single model takes around 2 days.
- Pseudo Labeling: learning simultaneously or only using pretraining phase.
- Loss function: bce + dice loss (I also tried weighing boundary pixel loss, it gave similar result. Fear of overfitting, I finally decided not to use it.)
- Ensemble: 5 fold ensemble @1536x1024 + 6 fold ensemble @1920x1280, weighted average. I weighted by LB ranking in my submissions.
- TTA: only horizontal flip.
- Adjusting threshold: I decided threshold which gives best score on validation set. I set the threshold to 0.508. In LB, it makes score improving only 0.000001.

Other

One of the best contributor of improving score is training on pseudo labeling data. I think why pseudo labeling contribute so much is the amount of test data, and we can get predictions close ground truth.

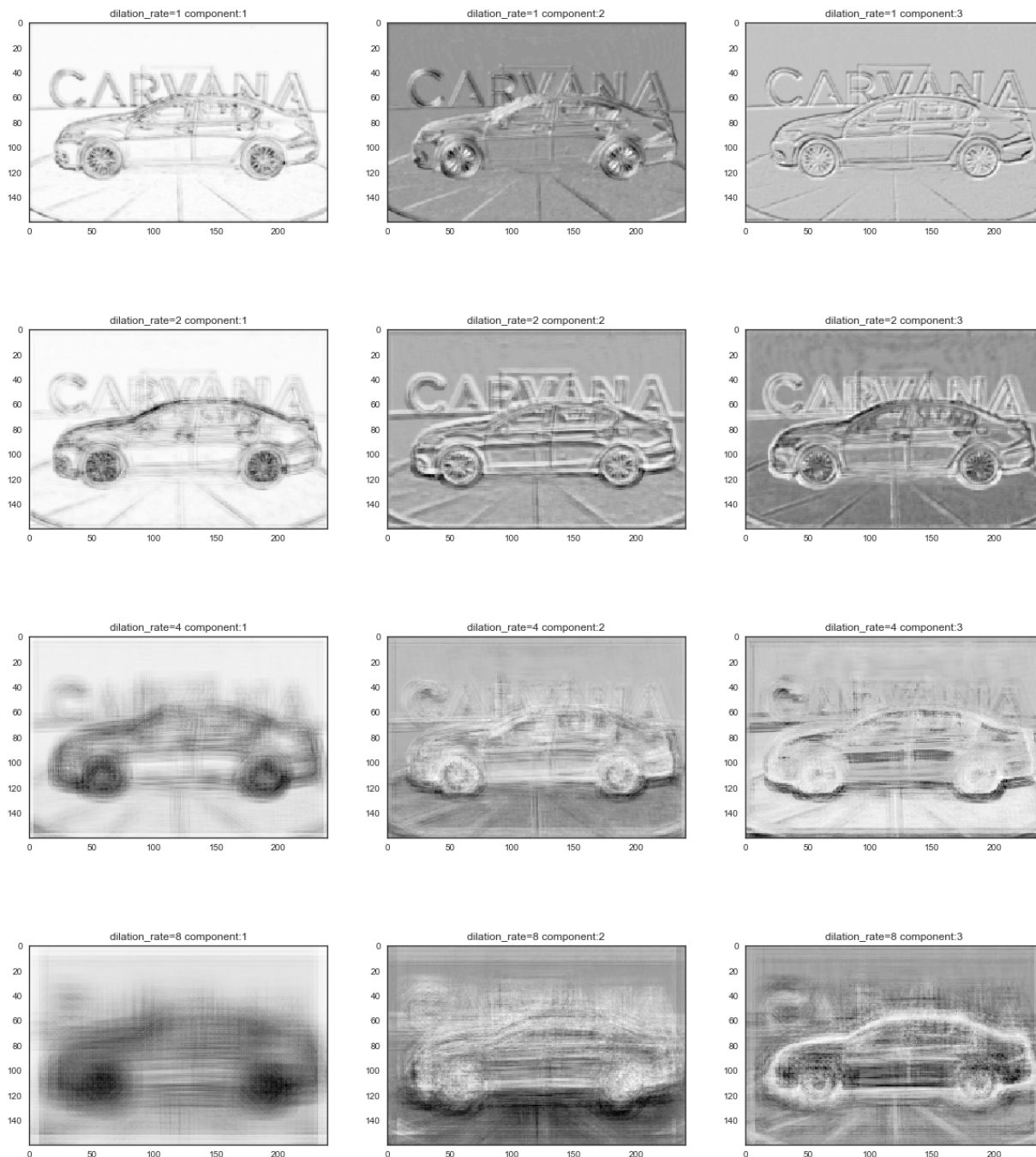
I tried post processing by using pydensecrf for only difficult to mask car images. But it gave me no improvement. As for how to choose "difficult images", I calculated multi class version of dice coefficient (I'm afraid that I shouldn't say so) of predictions by several models.



lyakaap • (3rd in this Competition) • 5 months ago • Options • Reply



Principal Components of feature maps that is output of Dilated Convolution layers each has a different dilation_rate. I think it's interesting.



- [dilation_rate=1.png \(137.44 KB\)](#)
- [dilation_rate=2.png \(152.19 KB\)](#)
- [dilation_rate=8.png \(160.36 KB\)](#)
- [dilation_rate=32.png \(141.42 KB\)](#)
- [dilation_rate=16.png \(153.32 KB\)](#)
- [dilation_rate=4.png \(144.79 KB\)](#)



Heng CherKe... • 5 months ago • Options • Reply

3

Very nice work. By the way i tries residual connections at the center layers. It also improve my results. It seems that the center layers are important but I don't know why. It will be fun to investigate.

Being the smallest, the center layers are the easiest to work with. I am going to try other structure like densenet or senet at the center layers later



lyakaap • (3rd in this Competition) • 5 months ago • Options • Reply



Thank you! I'm looking forward to your experiment results. I'm very curious.



TacoHC • (65th in this Competition) • 5 months ago • Options • Reply



Thanks for sharing! what do you exactly mean by saying pseudo labeling data?



lyakaap • (3rd in this Competition) • 5 months ago • Options • Reply



Using test data predictions by other model as if they were ground truth. In the training phase, you can use (X_train, y_train) and (X_test, y_test) for training.



ZhiyaoTang • 5 months ago • Options • Reply



Congratulation and thanks for sharing! By the way, could you share the setting of ReduceLROnPlateau, e.g. the monitor is val_loss or val_metrics?



lyakaap • (3rd in this Competition) • 5 months ago • Options • Reply



```
ReduceLROnPlateau(monitor='val_dice_coef',  
                  factor=0.2,  
                  patience=3,  
                  verbose=1,  
                  epsilon=1e-4,  
                  mode='max')
```

I used above settings, but I think it is better to use validation loss as metrics.



ironbar • (149th in this Competition) • 5 months ago • Options • Reply

^ 0 v

Thanks, Did you split by car on cross-validation?



lyakaap • (3rd in this Competition) • 5 months ago • Options • Reply

^ 1 v



Yes. I used normal K-fold CV.