Project Report (June 15, 2021)

Ankita Ghosh[1] **and Sahil Khose**[2]

[1]Research Assistant, ghoshankita0907@gmail.com , CSE, MIT Manipal
[2]Research Assistant, sahilkhose18@gmail.com, ICT, MIT Manipal

ABSTRACT

We discuss our implementation and pipeline for fovea segmentation. We implement the DeepLab V3+ model with EfficientNet-B3 backbome and train 5 different models with different values of the learning rate (alpha) through a logarithmic grid search method.

Data and Pre-processing

We have trained the model on two datasets to perform binary segmentation where the model segments the fovea:

1. Drive: 40 images with ground truth
2. Messidor: 180 images with ground truth

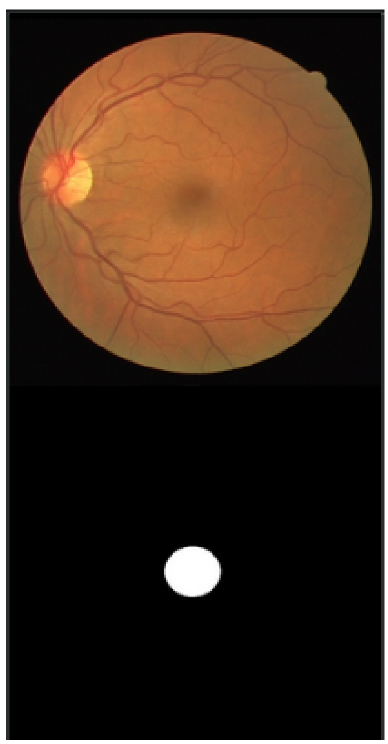
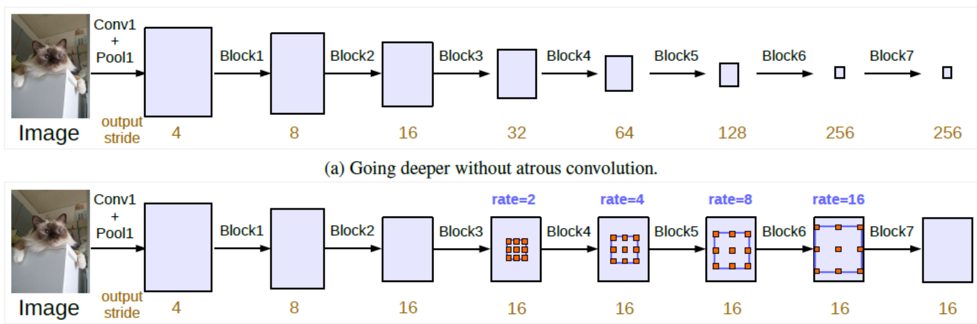


Figure 1. Fovea segmentation

- To expedite the process of feature extraction for the deep learning model, we apply bilateral filter to the image, followed by two iterations of dilation and one iteration of erosion.
- For image augmentation we perform shuffling, rotation, scaling, shifting and brightness contrast.
- The images and masks are resized to 512×512 dimensions while training to strike a balance between processing efficiency gained by the lower dimensional images and information retrieval of the high-resolution images.

Training

- The loss function used is a weighted combination of Binary Cross-Entropy loss (BCE) and Dice loss as it provides visually cleaner results.
- We trained a DeepLabV3+ model with EfficientNet-B3 as the backbone.
  - DeepLabV3+ is a refinement of DeepLabV3 which uses atrous convolution. Atrous convolution is a powerful tool to explicitly adjust the filter's field-of-view as well as control the resolution of feature responses computed by Deep Convolution Neural Network.
  - We use encoder depth of 5 which refers to the number of stages used in encoder. The number of convolution filters (decoder channels) used is 256.



(a) Going deeper without atrous convolution.

(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Figure 2. DeepLab V3+

- Batch size was set to 8. Different learning rates were tested in a logarithmic grid ranging from 1 to 1e-6. The model configuration and hyperparameters used are as shown in Figure 3.

```
# ####### DEEPLAB V3+ CONFIG #######
ENCODER_DEPTH=5
DECODER_CHANNELS=256
BATCH_SIZE= [8]
# BATCH_SIZE= [8, 16, 32, 64, 128]
EPOCHS= 10
# LR = [1e-3]
LR = [1, 1e-2, 1e-4, 1e-6]
ENCODER_NAME= 'efficientnet-b3'
```

Figure 3. Model configuration

- We graphed the results based on the loss, accuracy and based on the metrics MIoU (Mean Intersection over Union). Figure 4 shows our training progress using different learning rates.

Discussion

- We empirically concluded that LR of 1e-4 was the best for DeepLab V3+ with EfficientNet B-3 backbone, we plan to use it for our further implementations.
- Architectures like UNet, UNet+, PSPNet are available for semantic segmentation task which will be experimented in future iterations.
- Backbones like ResNet-18, ResNet-50, ResNet-101, EfficientNet family will be experimented to find the optimal image encoding architecture for the segmentation model.

Figure 4. Training Progress