DIABETIC RETINOPATHY DETECTION

Team 10: Shuai ZHANG, Peng GU

Institute of Signal Processing and System Theory, University of Stuttgart, Germany

Universität Stuttgart

ISS

1. Input Pipeline

- IDRID Dataset
  - a dataset of retinal images with and without signs of diabetic retinopathy
- Preprocessing and Augmentation
  - For each image:
    - normalize to [0, 1]
    - resize to (512, 512, 3)
    - augmentation
  - For dataset:
    - load csv file
    - create tfrecords
    - balance dataset

2. Model and Training

Three models, namely vgg-like, resnet-like and transfer model (densenet121) were implemented. At the end, ensemble learning was applied by combining three models.

- Overview of Model Architectures
  The architectures of all three models are shown in fig 1.
  - Vgg-like model
    - base filters 16
    - 4 vgg blocks
    - dropout rate 0.5
  - Resnet-like model
    - base filters 16

- 6 basic blocks
- dropout rate 0.5
- Transfer model
  - GAP layer, dropout layer and a classification output layer on top of the model
  - dropout rate 0.4
  - first 88 convolutional layers frozen, the other layers trainable
- Ensemble learning
  - Averaging
  - Voting
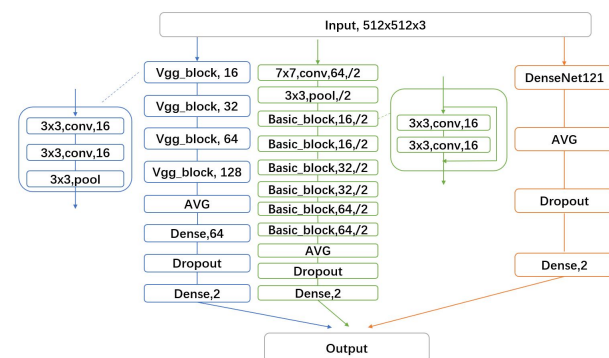
Figure 1: *Illustration of the ensemble model*

- Training Details
  - Class weighted balanced loss and accuracy
    - Multiply the loss and accuracy of each sample by the inverse of the class weight.
  - Optimizer
    - Adam
  - Learning rate
    - vgg-like model: 1e-3
    - resnet-like model: 1e-3

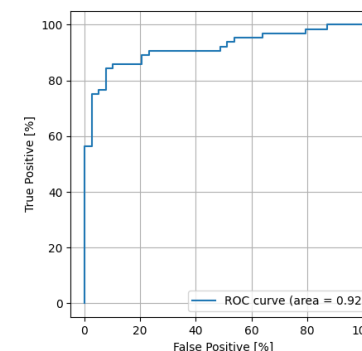- transfer model: 1e-6

3. Evaluation

- Confusion matrix
  - The confusion matrix test on vgg-like model is shown in fig2.
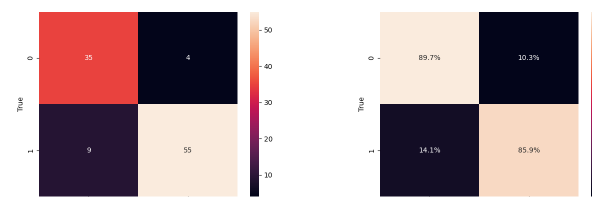
Figure 2: *Confusion Matrix*

- ROC curve
  - The ROC curve test on vgg-like model is shown in fig3.
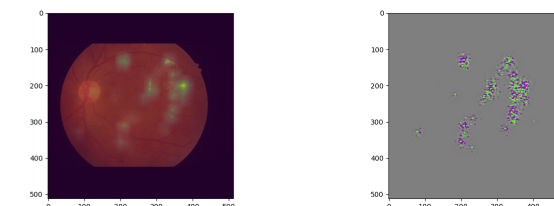

Figure 3: *ROC curve*

- F1 score
  - harmonic mean of the precision and recall, shown in fig4

- more suitable to unbalanced dataset
- reached about 0.9

$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision}$$

Figure 4: *Definition of F1 score*

- Deep visualization
  - The performance in form of grad-CAM test on vgg-like model is shown in fig5.

Figure 5: *Deep visualization*

4. Results

Overall, vgg-like model performs slightly better. After ensemble learning, the accuracy is further improved.

Architecture	F1 Score	Accuracy
Vgg-like	0.89	87.4%
Resnet-like	0.89	87.4%
Transfer(densenet121)	0.88	84.5%
Ensemble	0.90	88.3%

Table 1: *Results for binary classification*