DIABETIC RETINOPATHY DETECTION

Ayush Mittal and John Pravin Arockiasamy (Team 13)

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

1. Input Pipeline

Dataset

The Indian Diabetic Retinopathy Image Dataset (IDRID) contains 413 retinal fundus images in training and 103 in the test set. The dataset is labeled with 5 levels of different severity of the disease.

Preprocessing Steps

- Train & Validation data split with 9:1 ratio.
- Oversampling the training set for balancing data.
- 2 labels instead of 5 labels. '0' being no disease and '1' being a positive result.
- Resizing the images to (256, 256, 3).
- The Graham preprocessing (with and without).

Augmentation

Random flip, rotation, brightness, hue, and saturation.

2. Model Architectures

We experimented with custom and standard architectures. For standard architecture, pre-trained ImageNet weights were used whereas for custom models random initialization was used.

Custom Architectures

- VGG-like: 6 vgg blocks, 8 base filters, and 0.6 dropout rate.
- ResNet-like: 4 basic blocks, 64 base filters, 0.7 dropout rate.
- Transformer-like: 1-3 transformer layers, 3 heads, 8 patch size, and 0.7 dropout rate.

Standard Architectures

- ResNet50
- InceptionV3
- Xception

The standard architectures are used for ensemble results (averaging). All the cnn backbones are followed by GAP and then linear layers.

3. Training and Evaluation

- During Training, we used Binary Cross Entropy(BCE) loss, Adam optimizer with a constant learning rate of 1e-4 for custom cnn models, 1e-5 for transformer, and 1e-5 as well for standard pre-trained models.
- For evaluation a confusion matrix, and F1 score were used.
  - **Confusion Matrix** (fig.1) represents counts from predicted and actual values. The True Negative shows the number of negative examples classified accurately. True Positive indicates the number of positive examples classified accurately. False Positive value, i.e., the number of actual negative examples classified as positive; and a False Negative value is the number of actual positive examples classified as negative.
  - **F1 Score** (fig.4) is the harmonic mean of the precision and recall. More suitable for imbalanced datasets.

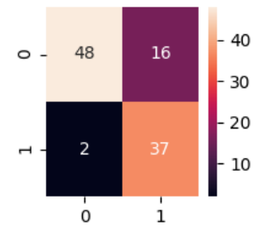

Figure 1: VGG-like confusion matrix

4. Visualization

Grad-Cam weights the 2D activations by the average gradient.

Figure 2: Input Image

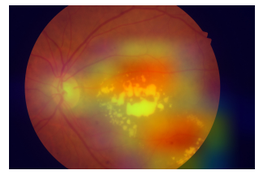

Figure 3: Grad-Cam Image

5. Results Comparison

The ensemble model outperformed other models mentioned in fig.4.

	Model	Accuracy	F1 Score
0	Ensemble	85.11±0.9	86.27±1.26
1	VGG-C	83.94±1.54	84.98±1.33
2	Resnet-C	81.19±3.23	83.09±2.45
3	Transformer-C	77.21±1.77	77.78±2.19

Figure 4: Results Table

Graham Preprocessing improves the accuracy on the IDRiD dataset as shown in fig.5.

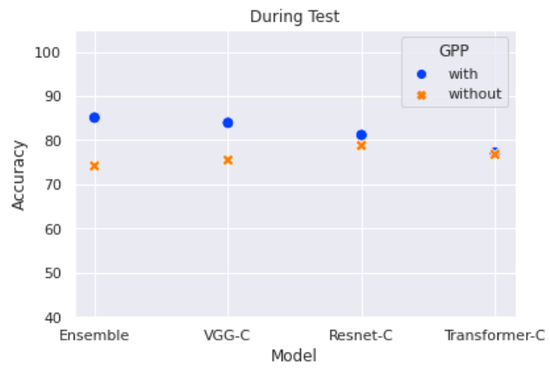

Figure 5: Graham Preprocessing effect

Deep Learning Lab 2020, February 4, Stuttgart, Germany