CNN-BASED DIABETIC RETINOPATHY DETECTION

Baran Can Gül and Lydia Schönpflug

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

Universität Stuttgart

1. Introduction

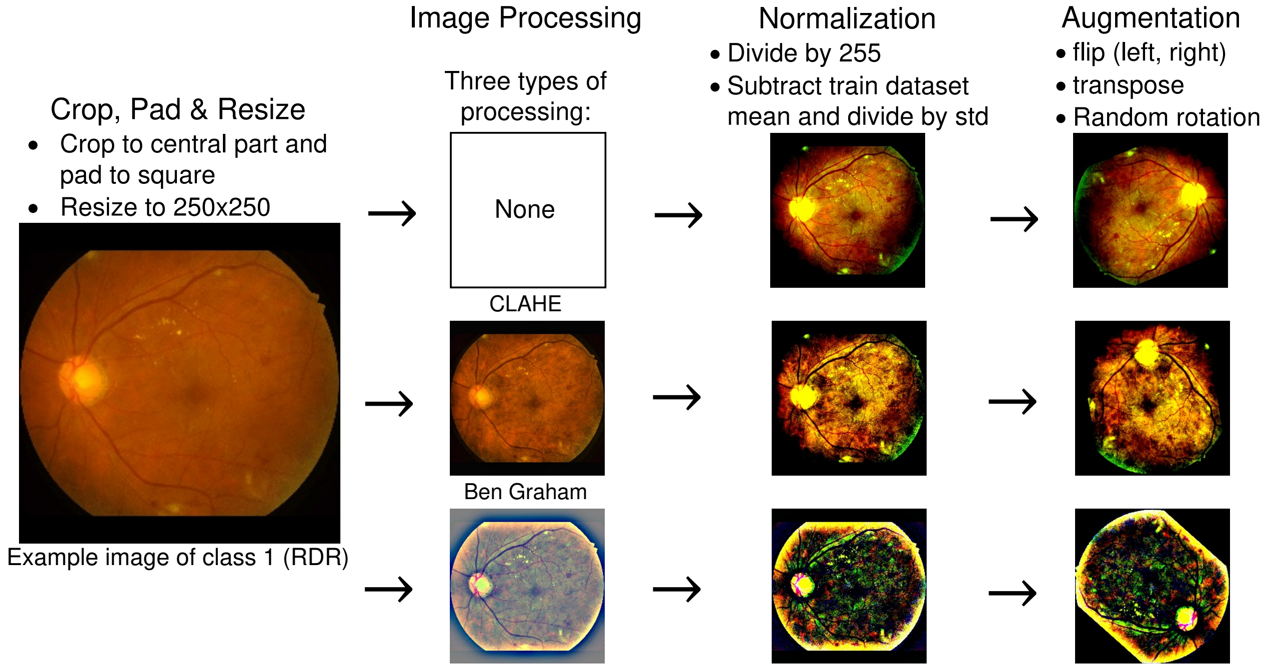
Diabetic retinopathy (DR) is a side effect of diabetes and mainly detected through a time-consuming examination of retina images by trained personal. An automated detection could support doctors in recognizing and categorizing stages of DR, thus fastening the diagnosis and treatment process.

IDRID Dataset			
Total images	516	**Training dataset samples**	413 (80%)
Image resolution	4288 x 2848 pixels	**Test dataset samples**	103 (20%)
Image format	jpg	**Ground truth format**	csv
Labels	0: no apparent DR 1: mild non-proliferative DR 4: proliferative DR	2: moderate non-proliferative DR 3:severe non-proliferative DR	

The task was simplified to a binary classification by merging 0 and 1 labels to class 0 (non referable DR, NRDR) and 2,3, and 4 labels to class 1 (referable DR, RDR).

2. Input Pipeline

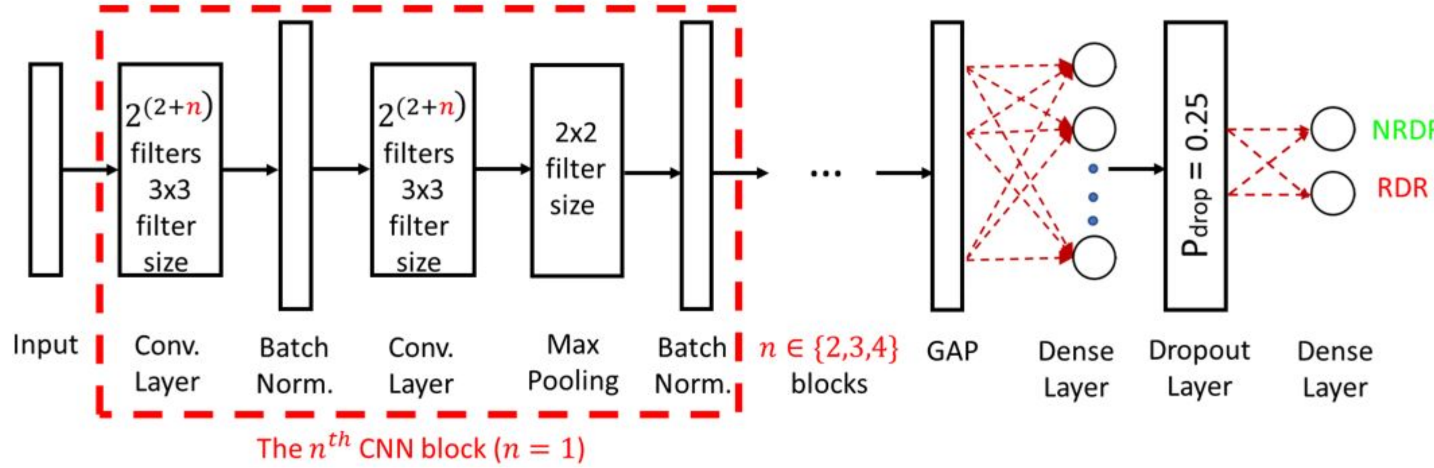
For efficient data loading the dataset was saved as a custom tensorflow dataset. Then the following processing and augmentation steps were applied to the images:



Through offline augmentation the images from the less prevalent class were augmented more, to ideally reach an equal amount of 0 and 1 labeled samples. The augmentation percentage was chosen through hyperparameter optimization.

3. Model architecture

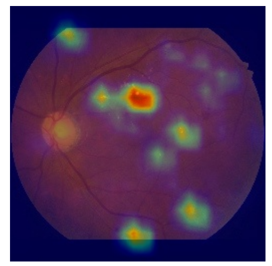
The model architecture was created by adding batch normalization layers to a VGG-based model, leading to a total number of 83146 model parameters. The hyperparameters of the model were determined by hyperparameter optimization.



4. Results

Despite our assumption, image processing did not show to improve the model performance. The best results were even achieved when using no image processing at all. This might be due to the small size of the dataset and rather high quality of the dataset images, which proves further processing steps to be unnecessary. Improvements could however be achieved through balancing and augmenting the dataset, leading to a $+7\%$ increase for validation and test accuracy.

	Avg 10 runs		Best overall	
	Val accuracy	Test accuracy	Val accuracy	Test accuracy
Plain dataset	82,96	77,88	84,27	81,31
+ balancing	88,99	78,04	90,04	85,22
+ augmentation	90,62	84,74	93,27	89,62
Ben Graham	89,82	82,30	93,27	88,06
CLAHE	90,11	82,45	91,66	87,78



Guided GradCam

Guided Grad-Cam was used to visualize the image regions with the greatest influence on the label decision. Overall the results on the IDRID dataset are satifsfactory, however, the accuracy on a subset of 5481 images from the Kaggle DR dataset was only 62,08%. This shows the limits of a small dataset. For further model improvements we recommend the usage of additional training data.

Deep Learning Lab 20/21, February 9, Stuttgart, Germany