Multi-Cell Multi-Task Convolutional Neural Networks for Diabetic Retinopathy Grading

Kang Zhou[1,2], Zaiwang Gu[2,3], Wen Liu[1], Weixin Luo[1], Jun Cheng[2], Shenghua Gao[1], Jiang Liu[1]

[1] ShanghaiTech University, Shanghai, China
[2] Chinese Academy of Sciences, Ningbo, China
[3] Shanghai University, Shanghai, China

Introduction

Diabetic Retinopathy (DR) is a non-negligible eye disease among patients with Diabetes Mellitus, and automatic retinal image analysis algorithm for the DR screening is in high demand.

- **Problem:**
  Label: 0, 1, 2, 3, 4 (Larger number means the severity of the disease becomes more significant)
- **Task:**
  Input: image / ***Output***: it's grade
- **Challenge** (DR grading ≠ general image classification):
  - The classes in DR grading are ***relational*** while in general image classification are not
  - The image resolution of DR images is ***significantly higher*** than that of general images

Contribution

- We propose a Multi-Task Learning strategy to simultaneously improves the classification accuracy and discrepancy between ground-truth and predicted label.
- We propose a Multi-Cell CNN architecture which not only accelerates the training procedure, but also improves the classification accuracy.

Proposed Method

- **Multi-Task Learning:** Softmax loss doesn't consider the ***relationships*** of DR images with different stages:

$$L_1 = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\}\log(\mathrm{Prob}_{ij})\right]$$

Mean Square Error (MSE) loss is ***not robust*** for classification task:

$$L_2 = \frac{1}{m}\sum_{i=1}^{m}(y - y^{(i)})^2$$

Multi-task loss:

$$L = L_1 + L_2$$

- **Multi-Cell Architecture:** ***Small resolution*** image often leads to information loss especially when the lesion is small. ***Large resolution*** image will introduce more computational costs and lead to the gradient vanishing/exploding problem in optimization.

SPATIAL RESOLUTION OF INPUT IMAGE AND SOME FEATURE MAP

input image	224×224	256×256	448×448	720×720
before switch	5×5	8×8	12×12	21×21
after multi-cell	5×5	8×8	5×5	4×4

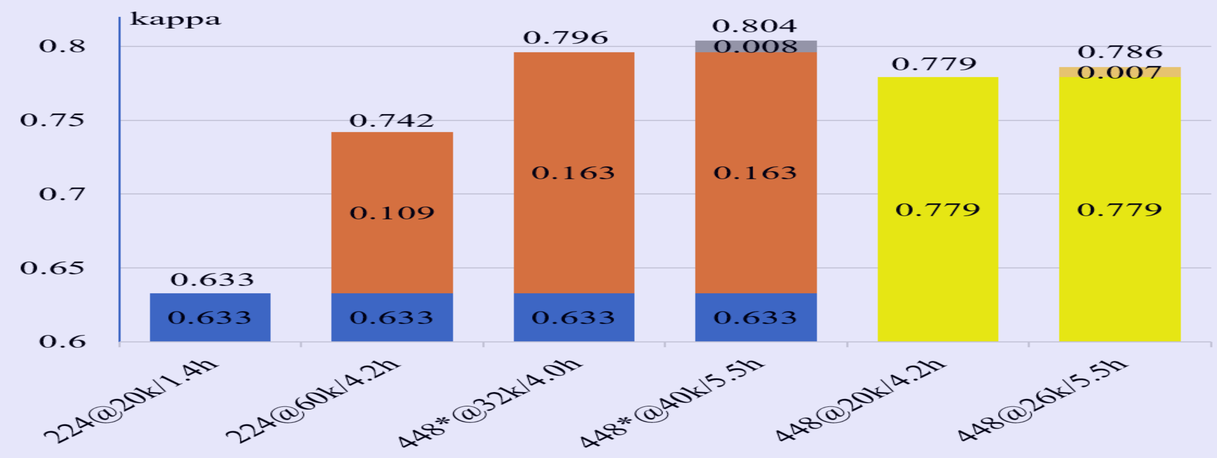
Multi-Cell Architecture ***gradually increase the depth of network architecture and the resolution of images.***

Experiment

- To evaluate each module of M²CNN, we conduct ablation experiments.

RESULTS OF EACH MODULE

Train	MSE	CE	Multi-Task		M²CNN	
Test	scores	prob.	scores	prob.	scores	prob.
224×224	0.720	0.725	0.742	0.718	-	-
448×448	0.790	0.772	0.812	0.782	**0.830**	0.812
720×720	0.835	0.751	**0.841**	0.826	**0.844**	0.842



- Compare our M²CNN method with the former methods achieving the best performance on Kaggle challenge and the state-of-the-art method (Zoom-in-Net).

COMPARISON WITH OTHER ALGORITHMS

Algorithm	val set	test set
Min-pooling	0.860	0.849
Zoom-in-Net	0.857	0.849
o_O	0.854	0.844
Reformed Gamblers	0.851	0.839
M-Net+A-Net	0.837	0.832
BaseNet	0.835	0.828
BaseNet+MT	0.841	0.838
M²CNN	0.844	0.841