Fully Deep Learning for Slit-lamp Photo based Nuclear Cataract Grading

Chaoxi Xu[*1,2,3], Xiangjia Zhu[*4], Wenwen He[*4], Yi Lu[*4], Xixi He[3], Zongjiang Shang[5], Jun Wu[5], Keke Zhang[4], Yinglei Zhang[4], Xianfang Rong[4], Zhennan Zhao[4], Lei Cai[4], Dayong Ding[3], and Xirong Li[**1,2,3]

[1]MOE Key Lab of DEKE, Renmin University of China
[2]AI & Media Computing Lab, School of Information, Renmin University of China
[3]Vistel AI Lab, Visionary Intelligence Ltd.
[4]Eye and ENT Hospital of Fudan University
[5]School of Electronics and Information, Northwestern Polytechnical University

Abstract. Age-related cataract is a priority eye disease, with nuclear cataract as its most common type. This paper aims for automated nuclear cataract grading based on slit-lamp photos. Different from previous efforts which rely on traditional feature extraction and grade modeling techniques, we propose in this paper a fully deep learning based solution. Given a slit-lamp photo, we localize its nuclear region by Faster R-CNN, followed by a ResNet-101 based grading model. In order to alleviate the issue of imbalanced data, a simple batch balancing strategy is introduced for improving the training of the grading network. Tested on a clinical dataset of 157 slit-lamp photos from 39 female and 31 male patients, the proposed solution outperforms the state-of-the-art, reducing the mean absolute error from 0.357 to 0.313. In addition, our solution processes a slit-lamp photo in approximately 0.1 second, which is two order faster than the state-of-the-art. With its effectiveness and efficiency, the new solution is promising for automated nuclear cataract grading.

Keywords: Nuclear cataract grading · Slit-lamp photos · Deep learning

1 Introduction

This paper studies automated *nuclear cataract grading* based on slit-lamp photos. A cataract is clouding of the lens in the eye. As the lens is to focus light rays onto the retina, such clouding leads to decrease in vision including blurry vision, faded colors halos around light, trouble seeing at night, *etc.* The most common factor of cataract is ageing [13]. Age-related cataract is reported to be responsible for over 50% of world blindness, and thus considered as a priority eye disease by the World Health Organization[1]. Nuclear cataract, which involves

---

[*] Equal contributions
[**] *Corresponding author: Xirong Li (xirong@ruc.edu.cn)*
[1] https://www.who.int/blindness/causes/priority/en/index1.html
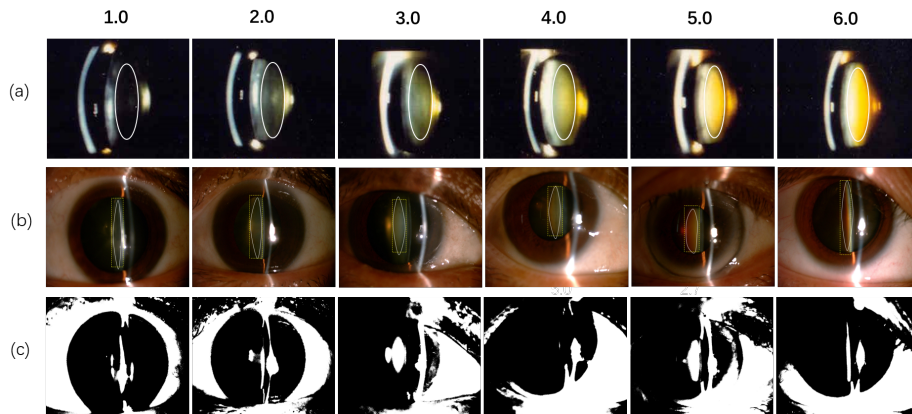
Fig. 1. Slit-lamp photos showing nuclear cataract of six scales, with scale 1.0 as the mildest and scale 6.0 as the severest. (a) Reference photos specified by the LOCS III [3], with the nuclear area in each photo marked out by a white ellipse. (b) Slit-lamp photos collected from a clinical scenario, with their nuclear areas (white ellipses) labeled by ophthalmologists. (c) Intensity-based binarization on (b), required by previous works [5, 7, 10, 15] as a prerequisite step to localize a region of interest. Note that the white of the eye and specular highlights in the background make this step ineffective. By contrast, we train Faster R-CNN [12], a state-of-the-art object detection network, to directly localize the nuclear zone, see the yellow bounding boxes with dotted borders in (b). Best viewed in digital format.

the central or *nuclear* area of the lens, is the most common type of age-related cataracts [1].

In clinical practice, nuclear cataract is diagnosed by an eye examination, where an ophthalmologist uses a slit-lamp to obtain a magnified view of the eye structures including the nuclear area in detail. Based on the opacification of the nuclear area, the cataract is further graded on a scale from 1.0 to 6.0. In particular, given a slit-lamp photo, a grade is manually estimated by comparing the photo against exemplars provided by the lens opacities classification system LOCS III [3], see Fig. 1(a). Apparently, the manual grading process not only requires well-trained eyes and is also time-consuming.

While automating the grading process is much in demand, challenges exist. The first challenge is how to reliably localize the nuclear region in a relatively complex background. Slit-lamp photos collected from a clinical scenario contain the eyelids, the sclera and the cornea with spectral highlights, as exemplified in Fig. 1(b). Previous works [5, 7, 10, 15] share an initial step, where a region of interest (ROI) is roughly localized by thresholding 20% to 30% of the brightest pixels in a gray-scale photo. However, such an intensity-based binarization cannot cope with the background noise, as Fig. 1(c) shows. One might argue the necessity of nuclear region localization. Indeed, we notice an early study by Fan *et al.* [4] where the whole image is used. Considering that the nuclear region

Table 1. An overview of recent methods for nuclear catatract grading. Different from the previous works, our method is based on fully deep learning.

Method	Region Detection	ROI	ROI Representation	Grading
Li *et al.* [10]		Lens	21-d intensity and color feature	SVM regression
Huang *et al.* [7]	Intensity-based binarlization + Active shape model [9]	Lens	6-d intensity and color feature	Nearest neighbor
Xu *et al.* [15]		Central part of lens	12,600-d bag-of-quantized color moment feature	Group Sparsity Regression
Gao *et al.* [5]		Central partof lens	32,768-d CRNN feature	SVM regression
This paper	Faster R-CNN	Nuclear region	ResNet-101	

contributes less than 3% of the pixels in a slit-lamp photo, grading based on the whole image is suboptimal.

Given the ROI successfully localized, the second challenge is how to derive a vectorized representation of the ROI, upon which a grading (or regression) model will be built. Due to variance in slit-lamp photography including the lighting condition, the skill of the technician, and the eye condition of the subject, both intra-grade divergence and inter-grade similarity exists in the photometric appearance of the nuclear area. Good features are thus needed. Earlier works rely on handcrafted features. To describe the intensity and color statistics in the lens, Li *et al.* [10] extract a 21-dim feature, while a shorter 6-dim feature is used in Huang *et al.* [7]. Xu *et al.* [15] represent the central part of the lens by a 12,600-dim bag-of-quantized color moment feature. The latest work by Gao *et al.* [5] employ convolutional-recursive neural networks (CRNN) [14] to extract a 32,768-dim deep feature. The authors then train an RBF-kernel SVM regression model for nuclear cataract grading. In all the above works (Table 1), feature extraction and the grading are separated and thus cannot be optimized jointly.

For answering the two challenges in nuclear cataract grading, this paper makes the following contributions:

1. We propose a fully deep learning based solution, which is the first work of its kind. In particular, we localize the nuclear region by Faster R-CNN [12]. With the nuclear region as input, a ResNet-101 [6] based grading model is trained. In contrast to the previous works, the use of ResNet-101 allows us to naturally achieve feature extraction and grading in a unified framework. Hence, the proposed solution is not only more effective and also computationally more efficient.

2. Tested on a clinical dataset of 157 slit-lamp photos from 39 female and 31 male patients, our solution outperforms the state-of-the-art [5], reducing the mean absolute error (MAE) from 0.357 to 0.313.
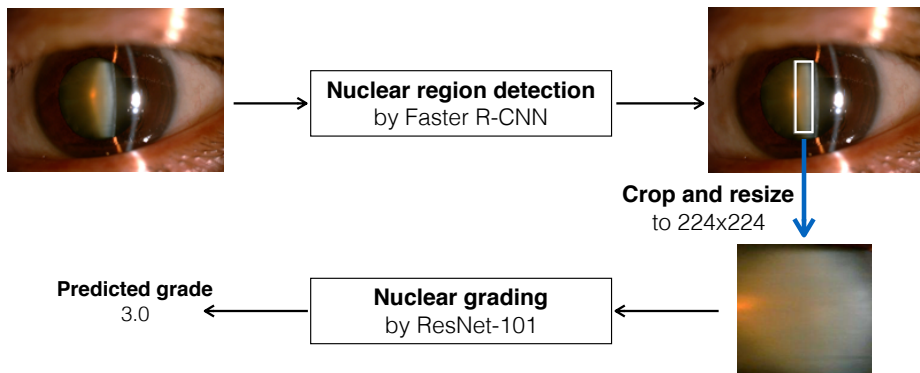
Fig. 2. An illustration of the proposed fully deep learning based solution for slit-lamp based nuclear cataract grading.

2 Proposed Solution

Given a slit-lamp photo from a specific patient having nuclear cataract, our goal is to automatically grade the severity level of the cataract in terms of the photometric appearance of the nuclear area. We propose a two-step solution. As shown in Fig. 2, the nuclear region is localized by Faster R-CNN, a state-of-the-art CNN for object detection. The second step is grading, where the detected region is fed to a grading model with ResNet-101, a state-of-the-art CNN for image classification, as its backbone. We use x to indicate a given photo, x_{nuc} for the sub-image corresponding to the nuclear area, and $y \in [1.0, 6.0]$ as the predicted grade. Accordingly, the above solution is expressed as follows,

$$
\begin{cases}
x_{nuc} \leftarrow \text{Faster R-CNN}(x), \\
y \leftarrow \text{ResNet-101}(x_{nuc}).
\end{cases}
\tag{1}
$$

2.1 Nuclear Region Localization by Faster R-CNN

Faster R-CNN by Ren *et al.* [12] consists of two subnetworks, *i.e.,* a region proposal network (RPN) followed by a detection network. The RPN proposes a number of bounding boxes that have the highest probability of containing ROIs. The detection network is responsible for discriminating foreground from background and refining the predicted location and size of the ROIs. Since a slit-lamp photo has one nuclear region, we select the top-ranked ROI as x_{nuc}.

To train Faster R-CNN for nuclear region localization, we adopt an SGD [2] optimizer with its default hyper-parameters setup except for the learning rate which is empirically set to 0.001. The mini-batch size is set to 4. The iteration is set to 100K batches and the learning rate is decayed by 0.1 at the 60K and the 80K batch. We set three anchor scales, *i.e.,* 3, 8 and 16 and two aspect ratios,

i.e., 1:1 and 1:4. The input image is sized to 500×500. Nuclear regions are well localized, with an IoU of 0.763 on average.

Note that Faster R-CNN cannot be directly used to grade a detected nuclear region. We therefore develop a grading model as follows.

2.2   ResNet-101 based Grading

According to the LOCS III [3], the estimated grade y shall have a precision of one decimal place. In other words, the value of y can be 1.3, 2.8, 5.3 *etc.* Therefore, we formulate the grading task as a regression problem. We depart from ResNet-101 pretrained on ImageNet [6], which requires an input size of 224×224. So we resize x_{nuc} accordingly. We substitute a regression layer for the original classification layer. We use the mean square error (MSE), a common loss function for regression. We minimize the MSE loss by the Adam [8] optimizer with its default hyper-parameters setup except for the learning rate, which is initially set to 0.0001, and decayed by 0.1 every 30 epochs. An early stop occurs if the performance on a validation set does not increase in 15 consecutive epochs. We perform data augmentation on the detected regions, changing brightness, saturation and contrast by a factor uniformly sampled from $[0, 2]$ respectively. The three operations are executed in a randomized order.

Batch balancing. The grades, ranging from 1.0 to 6.0, are not uniformly distributed. In our experimental data, the amount of photos graded between 3.0 and 4.0 is the largest. When constructing a mini-batch at random, the grading model will learn from unbalanced instances and thus become biased. To alleviate the issue, we introduce a simple batch balancing strategy. In particular, we quantize the grades into five groups, *i.e.,* $[1, 2)$, $[2, 3)$, $[3, 4)$, $[4, 5)$ and $[5, 6]$. When constructing a mini-batch, we randomly select the same amount of instances, which is 25, from each group, making the batch fully balanced.

3   Experiments

3.1   Experimental Setup

Dataset. As we have noted, there is no public dataset available for training and evaluating models for nuclear cataract grading. Our experiment data, provided by our hospital partner, consists of 847 slit-lamp photos from 214 female and 141 male patients with nuclear cataract. As shown in the second row of Fig. 1, these photos have complex background showing pupil, sclera, eyelid *etc.* , which are irrelevant with respect to the task and thus noisy. The photos have been graded collectively by six experienced ophthalmologists based on the LOCS III criteria [3]. Besides, the experts mark out the nuclear area in each photo.

We make a patient-based data partition, dividing randomly the dataset into three disjoint subsets, *i.e.,* training, validation and test, at a ratio of 7:1:2. A profile of the three subsets is shown in Table 2.

Performance metrics. Nuclear cataract grading is essentially a regression problem, so we report the mean absolute error (MAE), a common regression

Table 2. Profile of slit-lamp photos used in this work. In order to avoid over-fitting, a specific patient is exclusively assigned to the training, validation or test sets, so images from the same patient appear only in one dataset. For the purpose of approximately showing the grade distribution, we quantize the grades into five groups.

Dataset	Patients	Ages	Photos	Photos per group				
				$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6]$
Training	155 females + 94 males	17–95	587	73	152	209	108	45
Validation	20 females + 16 males	50–94	103	13	22	41	20	7
Test	39 females + 31 males	39–87	157	24	22	67	37	7

Table 3. Evaluating the choice of input and the influence of batch balancing for the grading model. Lower MAE and higher Accuracy are better.

Input	Batch balancing	MAE	Accuracy(%)
Whole image	✗	0.482	67.5
	✓	0.362	79.6
Nuclear region	✗	0.357	81.5
	✓	**0.313**	**84.7**

Table 4. Group-based evaluation. Lower MAE is better.

Input	Batch balancing	MAE per group				
		$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6]$
Whole image	✗	**0.325**	0.459	0.476	0.422	1.457
	✓	0.513	0.423	0.279	**0.186**	1.386
Nuclear region	✗	0.654	0.423	**0.260**	0.305	**0.343**
	✓	0.454	**0.309**	0.313	0.197	0.443

error metric. As an absolute error smaller than 0.5 is considered acceptable in the clinical practice, we report Accuracy, defined as the percentage of test photos that meet this criterion.

3.2 Experiment 1. Ablation Study

Choice of the input for grading. As Table 3 shows, using the detected nuclear region as input outperforms the whole image. Class activation maps are shown at Fig. 3, where red colors indicate highly activated regions. For the grading model with the whole image as input, the nuclear region is activated, suggesting that the model can indeed focus on the correct region. However, the relatively small ROI makes the model less effective. These results justify the necessity of nuclear region localization.

The influence of batch balancing. As Table 3 and 4 show, batch balancing is beneficial, regardless of the input.

(a) Whole image

(b) CAM of whole image

(c) Nuclear region detected by Faster R-CNN

(d) CAM of nuclear region
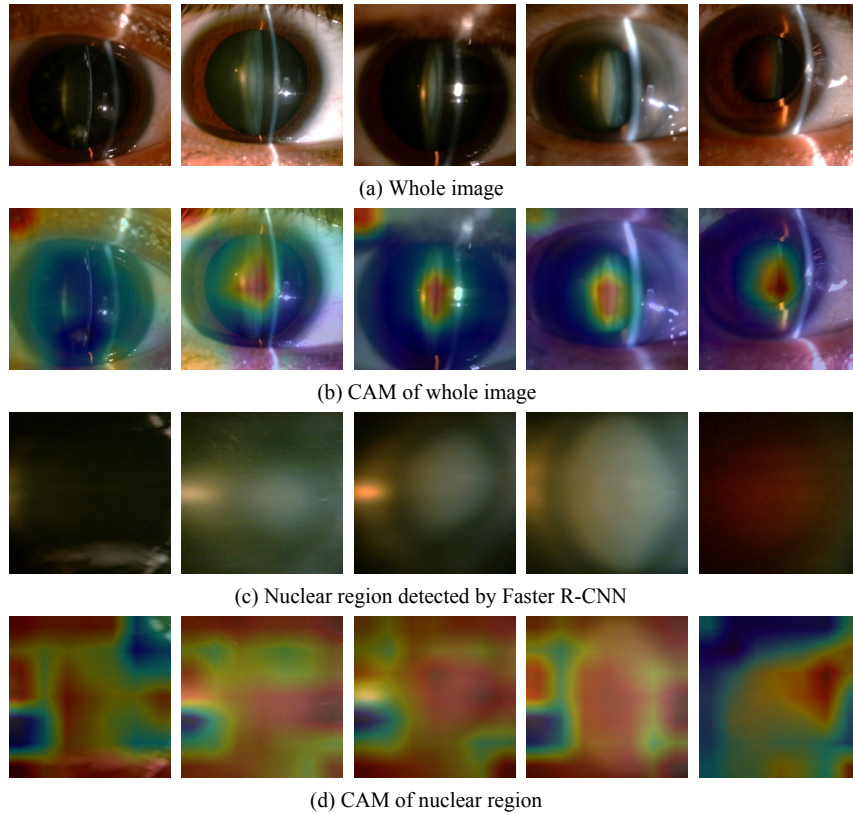
Ground truth:	1.3	2.2	3.0	4.0	5.6
Whole image:	2.1	2.5	**3.1**	**4.0**	4.4
Nuclear region:	**1.3**	**2.1**	3.2	3.8	**5.3**

Fig. 3. Visualizing the discriminative locations of an input image for nuclear cataract grading by class activation mapping (CAM) [16]. Each column indicates a specific test image. The second row and the last row are CAMs obtained when using the whole image and the detected nuclear as the input to the grading model, respectively. Red regions show high activations. Decimals in **bold font** means they are more close to the ground truth.

3.3 Experiment 2. Comparison with the State-of-the-Art

We compare with Gao *et al.* [5], the current state-of-the-art. As aforementioned, their ROI detection method mostly fails on the new dataset. So for a more fair comparison, our implementation of [5] uses the ROI found by Faster R-CNN as a candidate region and crops it as specified in [5].

As Table 5 shows, the proposed method surpasses Gao *et al.* [5] in terms of both MAE and Accuracy. Moreover, our method is computationally more efficient. On a normal computer with a 3.6GHz six-core CPU, 64GB RAM and a

Table 5. Comparison with SOTA. Lower MAE and higher Accuracy are better.

Method	MAE	Accuracy(%)
Gao *et al.* [5]	0.357	82.2
This paper	**0.313**	**84.7**

GTX 1080ti GPU, grading an image costs approximately 0.1 second. By contrast, the grading process of [5] takes 17 seconds per image, tested on a PC with a four-core 2.4GHz CPU and 24GB RAM. While our machine is more powerful, the efficiency is largely due to the fully deep learning property of our method.

The proposed method is general and can, in principle, be applied to recognizing other types of cataracts, *e.g.,* pediatric cataract [11].

4 Conclusions

A fully deep learning based solution for slit-lamp photo based nuclear cataract grading is developed. A test set of 157 slit-lamp photos from 70 patients verifies the effectiveness of the proposed solution. Concerning the choice of the input of the grading model, using the nuclear region localized by Faster R-CNN is better than using the whole image, with the mean absolute error (MAE) from 0.482 to 0.357. The proposed batch balancing strategy is also helpful, with MAE reduced to 0.313. Consequently, the new solution surpasses the state-of-the-art which has MAE of 0.357.

References

1. Asbell, P.A., Dualan, I., Mindel, J., Brocks, D., Ahmad, M., Epstein, S.: Age-related cataract. The Lancet **365**(9459), 599–609 (2005)
2. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: COMPSTAT, pp. 177–186. Springer (2010)
3. Chylack, L.T., Wolfe, J.K., Singer, D.M., Leske, M.C., Bullimore, M.A., Bailey, I.L., Friend, J., McCarthy, D., Wu, S.Y.: The lens opacities classification system III. Arch. Ophthalmology **111**(6), 831–836 (1993)
4. Fan, S., Dyer, C.R., Hubbard, L., Klein, B.: An automatic system for classification of nuclear sclerosis from slit-lamp photographs. In: MICCAI (2003)
5. Gao, X., Lin, S., Wong, T.Y.: Automatic feature learning to grade nuclear cataracts based on deep learning. IEEE Trans. Biomedical Engineering **62**(11), 2693–2701 (2015)

6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
7. Huang, W., Chan, K.L., Li, H., Lim, J.H., Liu, J., Wong, T.Y.: A computer assisted method for nuclear cataract grading from slit-lamp images using ranking. IEEE Trans. Medical Imaging **30**(1), 94–107 (2011)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014)
9. Li, H., Chutatape, O.: Boundary detection of optic disk by a modified asm method. Pattern Recognition **36**(9), 2093–2104 (2003)
10. Li, H., Lim, J.H., Liu, J., Mitchell, P., Tan, A.G., Wang, J.J., Wong, T.Y.: A computer-aided diagnosis system of nuclear cataract. IEEE Trans. Biomedical Engineering **57**(7), 1690–1698 (2010)
11. Liu, X., Jiang, J., Zhang, K., Long, E., Cui, J., Zhu, M., An, Y., Zhang, J., Liu, Z., Lin, Z., et al.: Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. PloS one **12**(3), e0168606 (2017)
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS (2015)
13. Robman, L., Taylor, H.: External factors in the development of cataract. Eye **19**(10), 1074 (2005)
14. Socher, R., Huval, B., Bath, B., Manning, C.D., Ng, A.Y.: Convolutional-recursive deep learning for 3d object classification. In: NIPS (2012)
15. Xu, Y., Gao, X., Lin, S., Wong, D.W.K., Liu, J., Xu, D., Cheng, C.Y., Cheung, C.Y., Wong, T.Y.: Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. In: MICCAI (2013)
16. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR (2016)