Automatic Feature Learning to Grade Nuclear Cataracts Based on Deep Learning

Xinting Gao[1], Stephen Lin[2], and Tien Yin Wong[3]

[1] Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
[2] Microsoft Research, P. R. China
[3] Singapore Eye Research Institute, Singapore

Abstract. Cataracts are a clouding of the lens and the leading cause of blindness worldwide. Assessing the presence and severity of cataracts is essential for diagnosis and progression monitoring, as well as to facilitate clinical research and management of the disease. Existing automatic methods for cataract grading utilize a predefined set of image features that may provide an incomplete, redundant, or even noisy representation. In this work, we propose a system to automatically learn features for grading the severity of nuclear cataracts from slit-lamp images. Local filters learned from image patches are fed into a convolutional neural network, followed by a set of recursive neural networks to further extract higher-order features. With these features, support vector regression is applied to determine the cataract grade. The proposed system is validated on a large population-based dataset of 5378 images, where it outperforms the state-of-the-art by yielding with respect to clinical grading a mean absolute error (ε) of 0.322, a 68.6% exact integral agreement ratio (R_0), a 86.5% decimal grading error ≤ 0.5 ($R_{e0.5}$), and a 99.1% decimal grading error ≤ 1.0 ($R_{e1.0}$).

1 Introduction

The lens of a human eye is optically transparent, consisting mostly of water and protein. Due to its shape, clarity and refractive index, the lens is able to focus light onto the retina, where the visual stimuli is transmitted through the optic nerve to the brain. Any clouding or loss of clarity in the lens is called a cataract, and the blockage of light results in impaired vision or even blindness [1]. Cataracts are the leading cause of visual impairment worldwide, accounting for more than 50% of blindness in developing countries. As most cataracts are age-related, the global trend of aging populations is expected to increase the prevalence of cataracts, with the number of blind people projected to reach 75 million by 2020 [2]. Mass screening and timely treatment of cataracts in the elderly is thus essential to improve quality of life and reduce health care costs.

The lens can be anatomically divided into three layers: an outer layer called the capsule, a central compacted core called the nucleus, and the cortex which surrounds the nucleus. Cataracts that occur in the nucleus are the most common

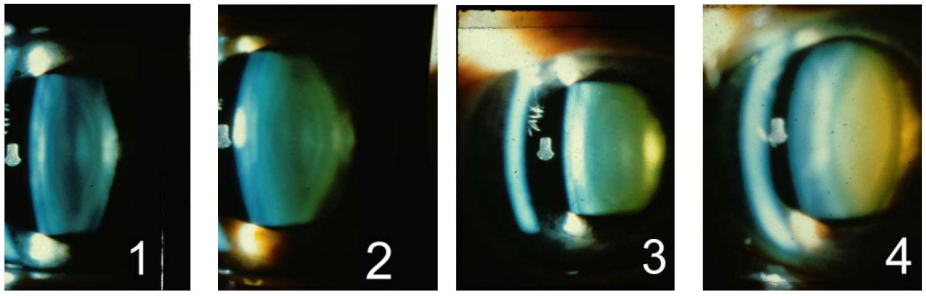

Fig. 1. Standard photographs of the Wisconsin grading system. The severity of the nuclear cataracts increases in the images from left to right, with greater brightness and lower contrast between anatomical landmarks. In addition, the color of the nucleus and posterior cortex exhibits more of a yellow tint due to brunescence.

type and will be the focus of this work. Since they appear as a homogeneous increase of opacification and coloration of the nucleus, they can be clearly seen in cross-sectional views of the lens in slit-lamp images [3].

For practical reasons, automatic methods for screening cataracts are needed, since manual examination, either directly through a slit-lamp microscope or indirectly through photographic comparisons to the Wisconsin grading protocol [4] (see Fig. 1), is time-consuming, expensive and subjective [5]. Though visually distinguishing grade 1 from grade 3 or 4 may be easy, it is difficult to determine precise grades on a continuous scale, which is critical for monitoring the progression of cataracts. In fact, human intra-grader agreement is only 70-80%, and inter-grader agreement is about 65% [4].

Existing techniques for automatic grading of nuclear cataracts utilize features designed according to the grading protocol [6–9]. These feature sets have the advantage of being low-dimensional; however, they may be incomplete, redundant, or even contain irrelevant (noisy) elements. For example, some earlier methods extract features from the whole lens [7, 8], though it was later shown that the anterior cortex provides no information for nuclear cataract grading [9]. In [9], bag-of-features (BOF) descriptors are extracted from different parts of the lens, and group sparsity regression (GSR) is used to select the features, parameters and models simultaneously. Although the BOF model automatically learns a codebook and represents each image as a histogram of visual words, the local feature descriptors must be defined in advance. Furthermore, as a global representation of local features, the BOF model has limited ability to encode geometric information.

Different from previous techniques, we propose in this work to automatically learn features for nuclear cataract grading in slit-lamp images. Toward this end, we adopt the deep learning framework of convolutional-recursive neural networks (CRNNs) [10], which are able to extract discriminative higher-order features because of its hierarchical structure. With these features, we apply support vector regression (SVR) to obtain grading estimates. Experiments demonstrate that

the learned features have greater discriminative power and the proposed system attains higher overall performance than previous methods.

Deep learning has been used in medical image processing for registration, segmentation and classification [11–16]. For our problem, we adopt the CRNN deep learning framework because of its ability to extract high-order semantic information from images of a realistic size (e.g., images of 1536×2048 resolution in our case). Although there exist many deep learning methods for learning features, most of them can handle only small images or local patches in practice [11–15], due to the considerable number of parameters that need to be learned for larger images. In the medical imaging domain, it is difficult if not infeasible to obtain a sufficiently large amount of data to effectively learn so many parameters (e.g., over one million parameters for supervised deep learning networks which typically consist of about seven layers). By contrast, the design of CRNNs allows for unsupervised learning within a hierarchical structure, which enables scaling up to realistic image sizes without requiring substantial training data [17, 10]. In this paper, we leverage the unsupervised learning method of the CRNN framework to learn features automatically. This work represents the first application of deep learning for diagnosing large medical images, and it is shown to achieve better overall performance than the existing cataract grading techniques. With respect to clinical grading, our method yields a mean absolute error (ε) of 0.322, a 68.6% exact integral agreement ratio (R_0), a 86.5% decimal grading error ≤ 0.5 ($R_{e0.5}$), and a 99.1% decimal grading error ≤ 1.0 ($R_{e1.0}$) on a large population-based dataset of 5378 images.

2 Method

In this section, we first introduce the CRNN deep learning framework and then the proposed automatic system which consists of three components: region of interest (ROI) and structure detection, feature learning and image representation, and grading.

2.1 Convolutional-Recursive Neural Networks

The unsupervised Convolutional-Recursive Neural Network (CRNN) method was proposed by Socher et al. [10]. It consists of three steps: pre-training CNN filters from randomly selected patches, generating local representations of each image by feeding the filters into a convolutional neural network (CNN) layer, and learning hierarchical feature representations using multiple recursive neural networks (RNNs) with random weights.

Pre-training CNN Filters The CNN filters are learned from randomly selected image patches that have been normalized and whitened. There exist several methods for learning the filters, such as sparse auto-encoder, sparse restricted Boltzmann machine, k-means clustering, and Gaussian mixtures. Among them,

k-means clustering has been shown to achieve the best performance [18]. K-means clustering aims to minimize the sum of squared Euclidean distances between patches, represented as a vector x, and their nearest cluster centers m_k. The standard 1-of-K, hard-assignment coding scheme is as follows:

$$f_k(x) = \begin{cases} 1 & \text{if } k = argmin_j ||m_j - x|| \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The learned K filters, $\{f_k, k = 1, 2...K\}$, will be used in the convolutional layer of the CNN.

Convolutional Neural Network A convolutional neural network consists of a convolution layer and a pooling layer. In the convolution layer, the set of learned filters, f_k, is convolved with the entire image to yield K corresponding feature maps. In the pooling layer, each feature map is sub-sampled by average or max pooling, which makes the resulting features invariant to translation and small deformations.

Recursive Neural Networks Recursive neural networks learn hierarchical feature representations by applying the same neural network recursively in a tree structure. The output of each neural network in an RNN is a parent vector computed from a set of child vectors, where the children at the bottom of the tree represent features generated by the CNN. Through this hierarchy, features of local image regions are merged into a higher-order, image-level feature representation. The RNN model can be trained through back-propagation [17]. In [10], Socher et al. demonstrated that a fixed tree structure can achieve good performance with the CNN as its preceding layer. Furthermore, multiple RNNs with random weights produce high quality features. As the learning is unsupervised, it is feasible to explore a large set of RNNs efficiently. It is particularly suitable for medical image processing applications where large amounts of labeled data are difficult or expensive to acquire.

2.2 Automatic Feature Learning to Grade Nuclear Cataracts

In applying the CRNN feature learning method to nuclear cataract grading, the lens structure is first detected and anatomical sections of the lens are segmented. Then CRNNs are applied to each section to learn a representation for that part of the lens. Finally, SVR is applied to the concatenated features to estimate the cataract grade. This procedure is illustrated in Fig. 2.

Lens Structure Detection To detect the lens structure, we employ the method in [7], which uses an active shape model learned from a training set with manually annotated landmark points. We then extract the central part of the lens around the visual axis as done in [9], but remove the anterior cortex section

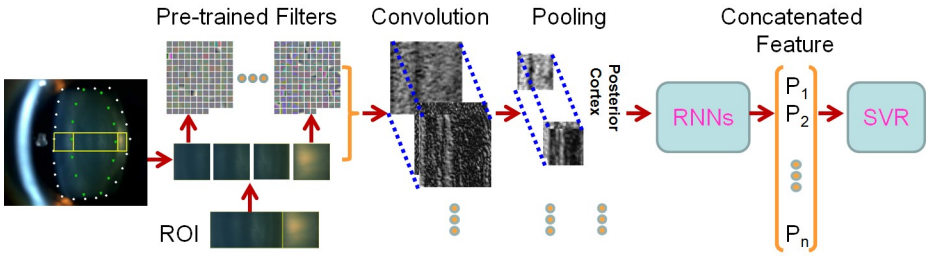

Fig. 2. Overview of deep learning based nuclear cataract grading. Regions of interest (ROIs) within detected lens structures are convolved with learned local filters, and then pooled for extracting higher-order features from recursive neural networks (RNNs). With the resulting feature vectors, final grading results are obtained using support vector regression (SVR).

since it contains no information for nuclear cataract grading according to both the grading protocol [4] and the state-of-the-art grading technique [9]. Therefore, our region of interest contains only the posterior cortex and nucleus. After extracting the ROI, the posterior cortex section is resized to $n_s \times n_s$ and the nucleus is resized to $n_s \times (n_s \times 2)$ as done in [9]. The resized nucleus is further divided into three half-overlapping sections: the anterior nucleus, central nucleus and posterior nucleus. Features are learned and extracted from each of the $n_s \times n_s$ sections. By aligning these sections to specific anatomical structures that are geometrically similar among individuals, discriminative features can be more effectively extracted even with a relatively small amount of data.

Feature Learning The features are learned for each section independently. First, we randomly extract local patches from a specific section of the training images for each grading category. Each patch has a spatial dimension of $n_p \times n_p$ and three color channels (R, G *and* B). Then, k-means clustering is used to generate the local filters from the randomly selected samples. Fig. 2 shows the filters learned for the posterior cortex section and the anterior nucleus section, which capture standard edge and color features.

The local filters are used in the convolutional layer of the CNN, followed by rectification, local contrast normalization, and average pooling. The invariance of the obtained feature to translation and small deformations helps to compensate for inaccuracies in structure detection, bringing greater robustness to the system. Each section of size $n_s \times n_s \times 3$ is convolved with the K square filters of size $n_p \times n_p \times 3$. This results in K feature maps of size $(n_s - n_p + 1) \times (n_s - n_p + 1)$. In the pooling layer, each feature map is processed by average pooling over $n_a \times n_a$ regions with a stride size of n_l. The size of the final pooled feature map for each section is $n_c = ((n_s - n_p + 1) - n_a)/n_l + 1$. The CNN layer thus produces a $K \times n_c \times n_c$ dimensional 3D feature map, with each feature vector $c_i \in \mathbb{R}^K$.

To extract higher-order features from the low-level feature map $C \in \mathbb{R}^{K \times n_c \times n_c}$, multiple random RNNs are applied. For each RNN, the basic element is a 3D

random matrix $W \in \mathbb{R}^{K \times b^2 \times K}$, where b is the block size which determines a set of local windows to merge into a parent vector $p \in \mathbb{R}^K$. The neural network is as follows:

$$p = g\left(W \begin{bmatrix} c_1 \\ \vdots \\ c_{b^2} \end{bmatrix}\right), \tag{2}$$

where g is a nonlinear function such as $tanh$ and $c_i \in \mathbb{R}^K$ are the feature vectors obtained in the CNN layer. Eq. 2 is recursively applied to the whole feature map without overlapping blocks to obtain a new layer R_1. Then Eq. 2 is applied again with the same weights W to the vectors in R_1, resulting in a second RNN layer R_2. The same procedure is repeated until only one parent vector is left. As the weight W is randomly generated without any supervised learning, multiple RNNs are needed to extract the higher-order features. Finally, each section is represented by the $N \times K$-dimensional vectors obtained through the N RNNs. We concatenate the feature vectors from all four of the sections to represent the image. The learned features are fed into an RBF ϵ-SVR [19] to obtain the final grading result.

3 Experiments

In this section, we first evaluate the proposed method by comparing it with the state-of-the-art nuclear cataract grading technique [9] using the dataset employed in their paper, ACHIKO-NC. Then, we compare the proposed learned features with the handcrafted features presented in [7] using the same learning method. The ACHIKO-NC dataset is comprised of 5378 images with decimal grading scores that range from 0.1 to 5.0. The scores are determined by professional graders based on the Wisconsin protocol [4], with higher decimal scores indicating greater severity of the cataract. The protocol takes the ceiling of each decimal grading score as the integral grading score, *i.e.*, a cataract with a decimal grading score of 1.2 has an integral grading score of 2. ACHIKO-NC contains 94 images of integral grade 1, 1874 images of integral grade 2, 2476 images of integral grade 3, 897 images of integral grade 4, and 37 images of integral grade 5. Since the unbalanced data distribution of ACHIKO-NC may skew a learned prediction model towards middle-grade estimates, we set the training sample size of each grade to 20 as done in [9].

3.1 Evaluation Criteria

In this work, we use the same four evaluation criteria as in [9] to measure grading accuracy, namely the exact integral agreement ratio (R_0), the percentage of decimal grading errors ≤ 0.5 ($R_{e0.5}$), the percentage of decimal grading errors

≤ 1.0 ($R_{e1.0}$), and the mean absolute error (ε), which are defined as

$$\begin{aligned} R_0 &= \frac{|\lceil G_{gt} \rceil = \lceil G_{pr} \rceil|_0}{N}, & R_{e0.5} &= \frac{\left||G_{gt} - G_{pr}| \leq 0.5\right|_0}{N}, \\ R_{e1.0} &= \frac{\left||G_{gt} - G_{pr}| \leq 1.0\right|_0}{N}, & \varepsilon &= \frac{\sum |G_{gt} - G_{pr}|}{N}, \end{aligned} \tag{3}$$

where G_{gt} denotes the ground-truth clinical grade, G_{pr} denotes the predicted grade, $\lceil \cdot \rceil$ is the ceiling function, $|\cdot|$ denotes the absolute value, $|\cdot|_0$ is a function that counts the number of non-zero values, and N is the number of testing images ($N = |G_{gt}|_0 = |G_{pr}|_0$).

The first metric (R_0) is based on grading protocol and has large quantization error, while the third metric ($R_{e1.0}$) also provides a relatively weak measure. On the other hand, the second ($R_{e0.5}$) and fourth (ε) metrics are more significant, since they measure performance at a finer scale and provide a better reflection of a method's utility in monitoring the progression of this disease. $R_{e0.5}$ has the most narrow tolerance among the four evaluation criteria, which makes it the most significant in evaluating the accuracy of grading.

3.2 Comparisons

We compare our method to the state-of-the-art technique, GSR [9], using the same dataset, experimental setting and reporting methods that they used. We also evaluate our method in relation to the method proposed in [7], which uses handcrafted features and RBF ϵ-SVR. Testing is conducted over twenty rounds. For each round, twenty images of each grade are selected randomly as the training data, and the remaining 5278 images are used for testing, which follows the training/testing sample ratio in [9, 7]. In training, optimal parameters for SVR and GSR were selected for each method by cross-validation. For the proposed method, $n_s = 148$, $n_p = 9$, $n_a = 10$, $n_l = 5$, $K = 128$, $N = 64$, and $b = 3$, which results in a feature dimension for the four sections of $N \times K \times 4 = 64 \times 128 \times 4 = 32768$. The results are listed in Table 1 in terms of mean value and standard deviation of R_0, $R_{e0.5}$, $R_{e1.0}$ and ε over the twenty rounds. The evaluations of $R_{e0.5}$, $R_{e1.0}$ and ε were found to be statistically significant, with associated p-values of [0.0920, 2.0383e-10, 3.5271e-14, 9.8171e-11] for the four metrics.

As mentioned previously, the $R_{e0.5}$ and ε metrics measure performance at a finer scale, and thus offer a better indication of a method's utility in disease progression monitoring. For these important metrics, our method achieves an improvement of 3.7% in $R_{e0.5}$ and 8.3% in ε on the large population-based database (5378 images). These represent meaningful improvements in light of the impact of accurate diagnoses on cataract patients.

From the results, we also have the following observations. First, since our method and [7] both use RBF ϵ-SVR for regression, the better performance of our method indicates that our learned features obtained via the CRNN deep learning framework provide a better representation than the handcrafted features of [7]. Second, although GSR is able to reduce the noise and increase the

Table 1. Performance comparisons for nuclear cataract grading methods

Method	R_0	$R_{e0.5}$	$R_{e1.0}$	ε
Proposed	**0.686±0.009**	**0.865±0.010**	**0.991±0.001**	**0.322±0.009**
BOF + GSR [9]	0.682±0.004	0.834±0.005	0.985±0.001	0.351±0.004
RBF ϵ-SVR [7]	0.658±0.014	0.824±0.016	0.981±0.004	0.354±0.014
our improvement over [9]	0.6%	3.7%	0.6%	8.3%

accuracy of structured BOF group features, its performance is still limited by the representation power of the BOF group features. The proposed learned features characterize the image well and furthermore encode high-level semantic information, which leads our method to better performance.

3.3 Computation Time

The proposed approach provides objective assessments at speeds comparable to state-of-the-art methods, making it useful for assisting and improving clinical management of the disease in the context of large-population screening. On a four-core 2.4GHz PC with 24GB RAM, the total training time using 100 images is about 1899 seconds, and it takes 17 seconds for prediction of one image. By comparison, the techniques of [9] and [7] run on the same computing platform at a speed of 20.45 seconds and 25.00 seconds per image, respectively.

3.4 Analysis and Discussion

The features learned in the CRNN framework depend on the sections from which they are extracted in a lens image. We empirically studied the effect of different lens sectioning on our method by extracting features under the following settings: 2-sections (posterior cortex and full nucleus), 3-sections (posterior cortex, full nucleus, and anterior cortex), 4-sections (the current implementation with posterior cortex, posterior nucleus, central nucleus, and anterior nucleus), and 5-sections (same as 4-sections plus anterior cortex). We also examined the BOF + GSR method [9] under its original 3-section setting and with 5-sections. The results, listed in Table 2, show that for our CRNNs, including the anterior cortex always leads to lower performance, as seen by comparing 3-section CRNNs to 2-section CRNNs, and 5-section CRNNs to 4-section CRNNs, with either regression method. These results support the findings in [9] that the anterior cortex introduces noise into the classification. However, an examination of group weights when using CRNN features with GSR shows that GSR does not eliminate the CRNN features extracted from the anterior cortex, which suggests that GSR may not fully remove noisy elements from a feature set. The analysis of Table 2 additionally indicates that a finer partition of the nucleus leads to more discriminative CRNN and BOF features (comparing 4-sections vs. 2-sections, and 5-sections vs. 3-sections). In fact, it is seen that 5-section BOF outperforms the original 3-section BOF in [9].

For both CRNN and BOF features, k-means clustering is employed to learn local filters or a codebook. Besides being able to capture more global geometric and semantic information, CRNN differs from BOF in the representation of color, as BOF applies k-means to local patches in each color channel separately, while CRNN applies it to full color patches in a way that the learned filters characterize both standard edge features and color features. The ability to model correlated color information provides CRNN features with greater discriminative power.

Table 2. Analysis of different lens sectioning

Method	Feature	R_0	$R_{e0.5}$	$R_{e1.0}$	ε
RBF ϵ-SVR	*2-section CRNNs*	0.645±0.013	0.819±0.014	0.985±0.003	0.358±0.012
	3-section CRNNs	0.631±0.013	0.801±0.016	0.981±0.003	0.372±0.012
	4-section CRNNs	**0.686±0.009**	**0.865±0.010**	**0.991±0.001**	**0.322±0.009**
	5-section CRNNs	0.677±0.008	0.857±0.010	0.990±0.002	0.329±0.008
	3-section BOF [9]	0.615±0.013	0.799±0.012	0.980±0.002	0.375±0.011
	5-section BOF	0.654±0.018	0.820±0.019	0.976±0.007	0.360±0.017
GSR	*2-section CRNNs*	0.654±0.021	0.823±0.027	0.985±0.005	0.355±0.022
	3-section CRNNs	0.643±0.024	0.808±0.030	0.982±0.005	0.366±0.024
	4-section CRNNs	0.679±0.011	**0.850±0.015**	**0.989±0.002**	**0.335±0.012**
	5-section CRNNs	0.672±0.009	0.843±0.013	0.987±0.002	0.341±0.011
	3-section BOF [9]	0.682±0.004	0.834±0.005	0.985±0.001	0.351±0.004
	5-section BOF	**0.687±0.009**	0.838±0.011	0.987±0.002	0.345±0.008

4 Conclusion and Future Work

We have proposed a new method for nuclear cataract grading based on automatic feature learning. Difficulty in finding the right features has been a limiting factor in research on automatic cataract grading, and this work brings a new approach that directly addresses this issue in a systematic and general manner, in contrast to resorting to heuristic handpicked features. Through deep learning, discriminative features that characterize high-level semantic information are effectively extracted. In tests on the *ACHIKO-NC* dataset comprised of 5378 images, our system achieves a 68.6% exact agreement ratio (R_0) against clinical integral grading, a 86.5% decimal grading error ≤ 0.5 ($R_{e0.5}$), a 99.1% decimal grading error ≤ 1.0 ($R_{e1.0}$), and 0.322 mean absolute error, which represents significant improvements over the state-of-the-art method.

This approach has the potential to be applied to other eye diseases. For example, different handcrafted features are used in optic cup/disc segmentation to assess the progression of glaucoma and to detect drusen for assessment of age-related macular degeneration (AMD). Features extracted through this type of deep learning approach may potentially lead to improved performance in these cases.

References

1. Kanski, J.: Clinical Ophthalmology: A Systematic Approach. Elsevier (2007)
2. : Vision 2020: International Agency for the Prevention of Blindness 2010 report (2010)
3. Wong, T., Loon, S., Saw, S.: The epidemiology of age related eye diseases in asia. Br. J. Ophthalmol. **90** (2006) 506–511
4. Klein, B., Klein, R., Linton, K., Magli, Y., Neider, M.: Assessment of cataracts from photographs in the beaver dam eye study. Ophthalmology **97** (1990) 1428–33
5. Thylefors, B., Chylack, J.L., Konyamia, K., Sasaki, K., Sperduto, R., Taylor, H., S., W.: A simplied cataract grading system – the WHO cataract grading group. Ophthalmic Epidemiology **9** (2002) 83–95
6. Fan, S., Dyer, C., Hubbard, L., Klein, B.: An automatic system for classification of nuclear sclerosis from slit-lamp photographs. In: MICCAI. (2003) 592–601
7. Li, H., Lim, J., Liu, J., Mitchell, P., Tan, A., Wang, J., Wong, T.: A computer-aided diagnosis system of nuclear cataract. IEEE Trans. Biomedical Engineering **57** (2010) 1690–1698
8. Huang, W., Chan, K., Li, H., Lim, J., Liu, J., , Wong, T.: A computer assisted method for nuclear cataract grading from slit-lamp images using ranking. IEEE Trans. Medical Imaging **30** (2011) 94–107
9. Xu, Y., Gao, X., Lin, S., Wong, D., Liu, J., Xu, D., Cheng, C., Cheung, C., Wong, T.: Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. In: MICCAI. (2013) II:468–475
10. Socher, R., Huval, B., Bhat, B., Manning, C., Ng, A.: Convolutional-recursive deep learning for 3d object classification. In: Adv. Neural Information Processing Systems. (2012)
11. Cireşan, D., Giusti, A., Gambardella, L., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: MICCAI. (2013) II:411–418
12. Habibzadeh, M., Krzyżak, A., Fevens, T.: White blood cell differential counts using convolutional neural networks for low resolution images. In: MICCAI. (2013) II:263–274
13. Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., Shen, D.: Unsupervised deep feature learning for deformable registration of mr brain images. In: MICCAI. (2013) II:649–656
14. Liao, S., Gao, Y., Oto, A., Shen, D.: Representation learning: A unified deep learning framework for automatic prostate mr segmentation. In: MICCAI. (2013) II:254–261
15. Prasoon, A., Petersen, K., Igel, D., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: MICCAI. (2013) II:246–253
16. Brosch, T., Tam, R.: Manifold learning of brain mris by deep learning. In: MICCAI. (2013) II:633–640
17. Socher, R., Lin, C., Ng, A., Manning, C.: Parsing natural scenes and natural language with recursive neural networks. In: Int. Conf. Machine Learning. (2011)
18. Coates, A., Lee, H., Ng, A.Y.: An analysis of single-layer networks in unsupervised feature learning. In: Int. Conf. Artificial Intelligence and Statistics (AISTATS). (2011)
19. Chang, C., Lin, C.: Libsvm: a library for support vector machines. ACM Trans. Intel. Sys. and Tech. **2** (2011) 1–27