Ranking Loss: A Ranking-Based Deep Neural Network for Colorectal Cancer Grading in Pathology Images

Trinh Thi Le Vuong[1], Kyungeun Kim[2], Boram Song[2], and Jin Tae Kwak[1(✉)]

[1] School of Electrical Engineering, Korea University, Seoul 02841, Korea
jkwak@korea.ac.kr
[2] Department of Pathology, Sungkyunkwan University of Medicine, Seoul 06355, Korea

Abstract. In digital pathology, cancer grading has been widely studied by utilizing hand-crafted features and advanced machine learning and deep learning methods. In most of such studies, cancer grading has been formulated as a multi-class categorical classification problem, likely overlooking the relationship among different cancer grades. Herein, we propose a ranking-based deep neural network for cancer grading in pathology images. Utilizing deep neural networks, pathology images are mapped into a latent space. Built based upon a triplet loss, a ranking loss is devised to maximize the inter-class distance among cancer grades in the latent space with respect to the aggressiveness of cancer, leading to the correct ordering or rank of pathology images. To evaluate the proposed method, a number of colorectal pathology images have been employed. The experimental results demonstrate that the proposed approach is capable of predicting cancer grades with high accuracy, outperforming the deep neural networks without the ranking loss.

Keywords: Colorectal cancer grading · Multi-class classification · Ranking loss · Triplet loss · Convolutional neural network

1 Introduction

Cancer is a primary disease worldwide [1]. Once diagnosed with cancer, a cancer grade is assigned upon the histologic assessment of tissue specimens by pathologists under a microscope, describing how abnormal cancer cells are in comparison to healthy (or normal) cells, i.e., the aggressiveness of the cancer. Although it is dependent on the types of cancer and grading system, cancer grading mainly adopts 1–3 (or –4) grading scale such as well, moderately, and poorly differentiated (or undifferentiated) cancer, in other words, the low, intermediate and high grade cancer. In clinics, the cancer grade is used as a primary determinant in treatment planning and patient care [2]; thus, accurate cancer grading not only leads to an improved patient outcome but also a substantial reduction in the cost and complexity of cancer treatment. However, there are several factors that limit the accuracy and efficiency of the current practice of cancer grading: 1) *low-throughput*: cancer grading is, by and large, manually conducted, i.e., time-consuming; 2) *subjective*:

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12908, pp. 540–549, 2021.
https://doi.org/10.1007/978-3-030-87237-3_52

cancer grading is mainly based upon qualitative measures, prone to inter- and intra-observer variability [3]; 3) *workload*: pathology services are under increasing pressure due to the upsurge of the volume and complexity of workloads per pathologist [4], leading to over-fatigue and burnout of pathologists as well as decrease in the quality of the service. Therefore, an alternative digital pathology tool that can permit automated, prompt, and objective decision-making for tissue specimens could aid in improving diagnostic yield and accuracy of cancer grading.

Many digital pathology tools have been developed to improve the accuracy, efficiency and robustness of cancer pathology [5]. Earlier works, in general, conducts cancer grading in two steps – feature extraction and classification. The feature extraction step includes a set of statistical, morphological and/or texture features, so called hand-crafted features. The classification step mainly utilizes supervised machine learning algorithms such as support vector machine (SVM) [6], decision tree [7], Bayes [8] and boosting [9]. This approach has been applied to cancer grading in prostate [7], brain [10], and colon [11]. Nowadays, deep convolutional neural networks (CNNs) are extensively utilized for digital pathology; for instance, tissue segmentation [12], nuclei detection [13] and segmentation [14], cancer detection [15], and pathology image synthesis [16]. Several efforts have been also made to develop CNN models for cancer grading in prostate [17], breast [18], and brain [19]. Such models have been built based upon the state-of-the-art network architectures, including inception module [20], residual layout [21], densely-connected layers [22], and depth-wise separable convolutions [23]. In most of the previous works, cancer grading has been studied as a multi-class categorical problem. These methods have been sought to either classify tissue specimens into multiple classes at once or classify tissue specimens in a cascaded manner. However, cancer grades are not independent to each other. The higher the grade is, the more aggressive the cancer is, i.e., it is an ordinal classification problem. Distinguishing high grade cancer from intermediate grade cancer is more difficult than from low grade cancer or healthy tissue. The hierarchy of cancer grades has not been fully utilized in developing digital pathology tools for cancer grading.

In this paper, we propose a ranking-based deep convolutional neural network (RankCNN) for colorectal cancer grading in pathology images. The proposed network is built based upon the following principles: 1) there exists a function f that maps tissues into data points (or embeddings) in an Euclidean latent (or embedding) space; 2) the similarity/dissimilarity between tissues (of the same or differing cancer grades) is measurable using distance metrics in the latent space; 3) the more (semantically) similar tissues are, the (metrically) closer the corresponding data points are; 4) the difference in the cancer grades of the tissues is directly related to the distance between the corresponding embeddings in the latent space. The mapping function f can be obtained via the-state-of-the-art deep convolutional neural networks. The distance between the embeddings can be measured using distance metrics, e.g., a Euclidean distance. To enforces the constraints of the distance between tissues in the latent space, we, in particular, introduce a ranking loss that could aid in obtaining the correct order of tissues. Built based upon the triplet loss [24], the ranking loss aims to maximize the relative distance between the tissues of differing grades in regard to the order of pathology grades.

2 Methodology

The overview of the proposed approach is illustrated in Fig. 1. It utilizes a deep convolutional neural network to conduct cancer grading. To exploit the nature ordering of cancer grades, we devise a loss function, so called ranking loss, and utilize to optimize the network. The ranking loss extends and generalizes triplet loss. We first introduce the triplet loss and then present the proposed ranking loss.

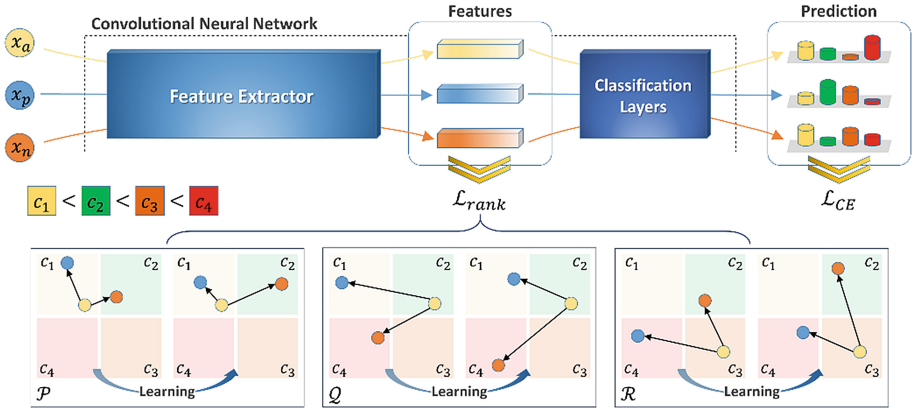

Fig. 1. The overview of the proposed method. x_a, x_p, and x_n denote an anchor, positive, and negative sample, respectively. c_1, c_2, c_3, and c_4 are the four classes that are ordered as $c_1 < c_2 < c_3 < c_4$. $\mathcal{P}$ denotes triples where x_a and x_p belong to the same class. $\mathcal{Q}$ and $\mathcal{R}$ include triplets from three different classes. In $\mathcal{Q}$, the class label of x_a is closer to the class label of x_p than that of x_n. In $\mathcal{R}$, the class label of x_a is equidistant from that of x_p and x_n, but x_p is ranked higher than x_a. $\mathcal{L}_{rank}$ and $\mathcal{L}_{CE}$ are ranking loss and cross entropy loss, respectively

2.1 Loss Functions

Two loss functions are employed to optimize the neural network as follows:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{rank} \tag{1}$$

where $\mathcal{L}_{CE}$ and $\mathcal{L}_{rank}$ denote cross entropy loss and ranking loss, respectively. λ is a weight factor that is associated with $\mathcal{L}_{rank}$ ($\lambda = 2$). $\mathcal{L}_{CE}$ has been widely accepted for the (categorical) classification task. $\mathcal{L}_{rank}$ is the loss function that is designed to capture the natural ordering of class labels, i.e., the ordinal classification task.

Let $\{(x_i, y_i)\}_{i=1}^{N}$ be a set of pathology images and ground truth labels where $x_i \in \mathbb{R}^{w \times h \times d}$ is the i^{th} pathology image, $y_i \in \mathcal{C} = \{c_1, c_2, \ldots, c_K\}$ is the ground truth label and N is the number of images. w, h and d denote the width, height and the number of channels, respectively. K is the cardinality of the class label. For the categorical classification, $\mathcal{L}_{CE}$ is used to measure the total entropy between the prediction and ground truth as follows:

$$\mathcal{L}_{CE} = -\sum_{i}^{N} \sum_{j}^{K} I_k(y_i) \log\left(o_{i,k}\right) \tag{2}$$

where $I_k(y)$ is a binary indicator where $I_k(y)$ is 1 if $k \in y$ and 0 otherwise. $o_{i,k}$ is the prediction for the image x_i and class label k.

$\mathcal{L}_{rank}$ utilizes the relative distance among the images with different class labels. It is defined on an embedding or feature representation of an image x_i, not the final prediction on the image. In the feature space $\mathbb{R}^D$, the relative distance among the images may be associated with the class labels. In other words, the distance between the images that belong to the same class or are closer to each other with respect to the class label becomes smaller than the distance between the images that are farther away from the perspective of the class labels. $\mathcal{L}_{rank}$ is built based upon the triplet loss. Thus, we first introduce the triplet loss and then provide the detailed description of the ranking loss.

Triplet Loss. The triplet loss is defined on a series of triplets $\{x_a, x_p, x_n\}, y_a = y_p \neq y_n$. For an (anchor) image x_a, the triplet loss is designed to have x_a is closer to all other (positive) images x_p of the same class than it is to any (negative) image x_n with a different class label from x_a. The triplet loss is formulated as:

$$\mathcal{L}_{tri} = \sum_{\substack{x_a, x_p, x_n \\ y_a = y_p \neq y_n}} \left[D(x_a, x_p) - D(x_a, x_n) + \alpha\right]_+ \tag{3}$$

where $[z]_+ = max(z, 0)$ and α is a margin. $D(x_i, x_j) = \|f(x_i) - f(x_j)\|_2^2$ where $f(x_i)$ is the feature representation of an image x_i. It should be noted that the number of all possible triplets grows exponentially. It is not necessary to consider all the triplets since most of the trivial triplets are easily and quickly separated in the feature space. Hence it is common to utilize only the *hard* (or *moderately hard*) triplets to compute $\mathcal{L}_{tri}$ [25]. For an anchor image x_a, the *hard* positive image is the one that is furthest away from x_a and *hard* negative image is the one that is closest to x_a. Thus, Eq. (3) can be rewritten as:

$$\mathcal{L}_{tri} = \sum\nolimits_{(i,j,k)\in\mathcal{P}} \sum\nolimits_{a}^{N_t} \mathcal{H}\left(x_a^{c_i}, x_p^{c_j}, x_n^{c_k}, \alpha\right), \tag{4}$$

$$\mathcal{H}\left(x_a^{c_i}, x_p^{c_j}, x_n^{c_k}, \alpha\right) = \left[\max_{p=1,\ldots,N_j} D\left(x_a^{c_i}, x_p^{c_j}\right) - \min_{n=1,\ldots,N_k} D\left(x_a^{c_i}, x_n^{c_k}\right) + \alpha\right]_+ \tag{5}$$

where $\mathcal{P} = \{(i, j, k) : c_i = c_j \neq c_k\}$, $x_a^{c_i}$ is an image with a class label c_i and N_i is the number of images in the class c_i.

Ranking Loss. In cancer pathology, there is a natural ordering of cancer grades $y \in \{c_1, c_2, \ldots, c_K\}$. Without loss of generality, let $c_1, c_2, \ldots, c_K$ be the values that are related on an ordinal scale; for example, $c_1 = 1, c_2 = 2, \ldots, c_K = K$. In the ordinal classification, the goal of a ranking function h is to learn the relative rank among the images, i.e., for any pair of images $\{x_i, x_k\}$ it holds that

$$h(x_i) > h(x_k), y_i > y_k. \tag{6}$$

This can be easily extended to any triplets $\{x_i, x_j, x_k\}$:

$$h(x_i, x_j) > h(x_i, x_k), y_i = y_j > y_k. \tag{7}$$

Replacing h with D, it is trivial to rewrite the formula as follows:

$$D\big(x_i, x_j\big) + \alpha > D(x_i, x_k), y_i = y_j > y_k. \tag{8}$$

This indicates that the rank between the images can be expressed with respect to the triplets, i.e., the triplet loss. However, the triplet loss can partially handle the rank among the images of multiple classes since the loss only considers a pair of class labels at a time. To extend the triplet loss and to fully exploit the ordering of the class labels, the ranking loss takes the triplets from three different classes into account. The ranking loss is formulated as:

$$\mathcal{L}_{rank} = \sum\nolimits_{(i,j,k)\in\mathcal{P}} \sum\nolimits_{a}^{N_i} \mathcal{H}\Big(x_a^{c_i}, x_p^{c_j}, x_n^{c_k}, \alpha_{\mathcal{P}}\Big) + \sum\nolimits_{(i,j,k)\in\mathcal{Q}} \sum\nolimits_{a}^{N_i} \mathcal{H}\Big(x_a^{c_i}, x_p^{c_j}, x_n^{c_k}, \alpha_{\mathcal{Q}}\Big) + \sum\nolimits_{(i,j,k)\in\mathcal{R}} \sum\nolimits_{a}^{N_i} \mathcal{H}\Big(x_a^{c_i}, x_p^{c_j}, x_n^{c_k}, \alpha_{\mathcal{R}}\Big) \tag{9}$$

where $\mathcal{Q} = \big\{(i,j,k) : \big|c_i - c_j\big| < |c_i - c_k|, c_i \neq c_j \neq c_k\big\}$, $\mathcal{R} = \big\{(i,j,k) : \big|c_i - c_j\big| = |c_i - c_k|, c_i \neq c_j \neq c_k, c_j > c_k\big\}$, and $\alpha_{\mathcal{P}}$, $\alpha_{\mathcal{Q}}$, and $\alpha_{\mathcal{R}}$ are the margins ($\alpha_{\mathcal{P}} = 1.0$, $\alpha_{\mathcal{Q}} = 1.0$, and $\alpha_{\mathcal{R}} = -0.5$). The first term is the same as the triple loss. It focuses on minimizing the relative distance among the images in the same class with respect to the images of different classes. The second and third term include the triplets from three different classes. The second term is to consider the cases where the anchor images are closer to positive images than to negative images with respect to class labels. The third term is the constraint on the cases where positive and negative images are equidistant from anchor images with respect to the class labels, but positive images are ranked higher than negative images. These two terms aim to push negative images further away, forcing the minimum distance between anchor and negative images to be larger than the maximum distance between anchor and positive images, following the ordering of the class labels.

2.2 Network Architecture

We employed three different types of backbone CNNs, which are widely used and built based upon different principles, to evaluate the effectiveness of the proposed ranking loss. The three types include 1) DenseNet (DenseNet121): uses all the subsequent layers via concatenation and scales the network by its depth, [22], MoblieNet (MobileNetV2) [23]: is a light-weight network using depthwise separable convolutions, and 3) EfficientNet (EfficientNet-B0) [26]: adopts a compound scaling method that balances the width, depth, and image resolution. Given a pathology image x_i, all three types of architectures generate two distinct outputs: 1) the high dimensional features $f(x_i)$: the output of the last convolutional block and 2) the prediction result o_i: the output of the classification layers given $f(x_i)$. Due to the difference in the input size from the original implementation, the average-pooling layer (AvgPool) in the classification layers of the three architectures is replaced by the gobal AvgPool. The cross entropy loss is utilized to assess the prediction results and the ranking loss is adopted to evaluate the high dimensional features.

2.3 Evaluation Metrics

We quantitatively evaluate the performance of the proposed method using five evaluation metrics: 1) accuracy (*ACC*), 2) recall (*REC*), 3) Cohen's kappa (κ), 4) Matthews correlation coefficient (*MCC*) [27], and 5) macro-average F1-score ($F1_{mac}$).

3 Experiments and Results

3.1 Datasets

Colorectal tissue samples were stained with hematoxylin and eosin (H&E) and scanned at x20 optical magnification. The tissue samples were collected from six colorectal tissue microarrays (TMAs) and three whole slide images (WSIs). An experienced pathologist reviewed the tissue samples and identified and delineated distinct histological regions, including benign (BN) and well-differentiated (WD), moderately-differentiated (MD), and poorly-differentiated (PD) tumors. From the WSIs and TMAs, tissue patches of size 1024×1024 pixels were extracted and resized to 512×512. ~10,000 image patches, including 1600 BN, 2322 WD, 4105 MD, and 1830 PD patches, were generated and further divided into a training (898 BN, 1615 WD, 2623 MD, and 1245 PD), validation (444 BN, 374 WD, 810 MD, and 238 PD), and test (262 BN, 334 WD, 672 MD, and 347 PD) dataset. We note that the data partition was conducted at WSI- and TAM-level.

3.2 Implementation Details

All the networks are optimized using Adam optimizer with parameter values ($\beta_1 = 0.5$, $\beta_2 = 0.999$, $\varepsilon = 1e-8$). The learning rate is set to $1e-4$. After 30 epochs, it is decayed by 10. λ is set to 2 after cross-validation experiments within the training set only. We train the networks for 60 epochs. The best model on the validation dataset is chosen for the evaluation on the test dataset. During training, the following data augmentation techniques are applied: 1) a random rotation in a range of [0, 60] degree, 2) a random scaling in a range of [0.8, 1.2], 3) a random translation in a range of [0.02, 0.02] with respect to the width and height of the input image, and 4) a random change in the brightness, contrast, saturation, and hue with a factor in a range of [0.95, 1.05]. For the loss calculation, We set the weighting factor for the ranking loss (λ) to 2.0 and the margins $\alpha_{\mathcal{P}}$, $\alpha_{\mathcal{Q}}$, and $\alpha_{\mathcal{R}}$ to 1.0, 1.0, and -0.5, respectively. Setting $\alpha_{\mathcal{R}} = -0.5$, we relax the constraint for the equidistant cases.

3.3 Results and Discussions

The experimental results of colorectal cancer grading are available in Table 1. Equipped with the proposed $\mathcal{L}_{rank}$, RankCNNs with different backbone networks, including DenseNet, MobileNet, and EfficientNet, obtained $\geq$86.91% *ACC*, $\geq$0.8693 *REC*, $\geq$0.8254 κ, $\geq$0.8255 *MCC*, and $\geq$0.8693 $F1_{mac}$. Among the three backbone networks, EfficientNet consistently outperformed other two backbone networks and MobileNet showed the lowest performance over the five evaluation metrics. To assess the effectiveness of the proposed $\mathcal{L}_{rank}$, we repeated the same experiments without $\mathcal{L}_{rank}$. For each

Table 1. Results of cancer grading.

Network	Loss	*ACC* (%)	*REC*	κ	*MCC*	$F1_{mac}$
DenseNet	$\mathcal{L}_{CE}$	81.50	0.8153	0.7533	0.7577	0.8191
	$\mathcal{L}_{CE}$ & $\mathcal{L}_{tri}$	85.20	0.8522	0.8027	0.8038	0.8554
	$\mathcal{L}_{CE}$ & $\mathcal{L}_{rank}$	**87.00**	**0.8703**	**0.8267**	**0.8275**	**0.8727**
MobileNet	$\mathcal{L}_{CE}$	85.20	0.8523	0.8027	0.8053	0.8579
	$\mathcal{L}_{CE}$ & $\mathcal{L}_{tri}$	86.53	0.8655	0.8204	0.8207	0.8676
	$\mathcal{L}_{CE}$ & $\mathcal{L}_{rank}$	**86.91**	**0.8693**	**0.8254**	**0.8255**	**0.8693**
EfficientNet	$\mathcal{L}_{CE}$	81.88	0.8191	0.7584	0.7668	0.8293
	$\mathcal{L}_{CE}$ & $\mathcal{L}_{tri}$	85.39	0.8541	0.8052	0.8055	0.8547
	$\mathcal{L}_{CE}$ & $\mathcal{L}_{rank}$	**88.80**	**0.8882**	**0.8507**	**0.8508**	**0.8881**

type of backbone networks, the networks were optimized using $\mathcal{L}_{CE}$ only (PlainCNNs) and a combination of $\mathcal{L}_{CE}$ and $\mathcal{L}_{tri}$ (TripletCNNs). Regardless of the type of backbone networks, RankCNNs were substantially superior to both PlainCNNs and TripletCNNs, demonstrating the additive value of the proposed $\mathcal{L}_{rank}$. PlainCNNs, which perform the categorical classification task only, showed the worst performance. TripletCNNs, conducting both categorical ($\mathcal{L}_{CE}$) and (partial) ordinal classification ($\mathcal{L}_{tri}$) tasks, achieved $\leq$86.53% *ACC*, $\leq$0.8655 *REC*, $\leq$0.8204 κ, $\leq$0.8207 *MCC*, and $\leq$0.8676 $F1_{mac}$. These results demonstrate that the ordinal classification could aid in improving the performance of cancer grading and the ranking loss could better exploit the hierarchy of cancer grades. However, the additive value of $\mathcal{L}_{rank}$ was different among the backbone networks. Using $\mathcal{L}_{CE}$ only, MobileNet was better than other two networks, but the performance gain by $\mathcal{L}_{rank}$ and $\mathcal{L}_{tri}$ was lesser for MobileNet. As a result, using $\mathcal{L}_{rank}$, MobileNet was poorer than other two networks. This indicates that the effect of $\mathcal{L}_{rank}$ could vary with the backbone networks.

Moreover, we visualized and quantitatively evaluated the classification results of the proposed approach (Fig. 2 Visual assessment of colorectal cancer grading. Blue, yellow, green, and red colors represent benign, WD, MD, and PD, respectively.). For each tissue image, we slid a rectangular window of size 1024 $\times$ 1024 pixels with a step size of 256 pixels, generating a set of image patches. The image patches were resized by half and used to produce the probabilities for the four classes. Averaging the probabilities over the overlapping patches, we assigned a class label per pixel in the tissue image. For various pathology images, the prediction results are corresponding to the ground truth maps provided by pathologists.

Table 2 inter-class distances among different cancer grades. shows the inter-class distances among different cancer grades where the distance was computed using the features generated by EfficientNet. Numbers represent the average distances and standard deviations in parenthesis. RankCNNs showed substantially higher inter-class distances than other networks (p-value $\ll$ 0.01). In comparison to PlainCNNs and TripletCNNs, RankCNNs, on average, exhibited ~3 to 7-fold and ~1 to 3-fold larger distances among

different cancer grades, respectively. Moreover, the larger distances we obtained, the more aggressive cancer grades are. For instance, the average distance of BN to WD, MD, and PD was 22.02, 24.51, and 30.80, respectively, i.e., linearly increasing as cancer progresses.

Table 2. Inter-class distances among different cancer grades.

Loss	BN-WD	BN-MD	BN-PD	WD-MD	WD-PD	MD-PD
$\mathcal{L}_{CE}$	4.19	4.22	4.25	1.33	2.98	2.26
	(2.05)	(2.05)	(2.05)	(0.69)	(0.64)	(0.85)
$\mathcal{L}_{CE}$ & $\mathcal{L}_{tri}$	11.30	11.03	12.71	1.79	6.83	5.50
	(4.05)	(4.14)	(4.23)	(1.83)	(1.42)	(2.00)
$\mathcal{L}_{CE}$ & $\mathcal{L}_{rank}$	**22.02**	**24.51**	**30.80**	**5.27**	**10.82**	**6.80**
	(2.64)	(3.03)	(3.13)	(3.12)	(2.98)	(2.49)

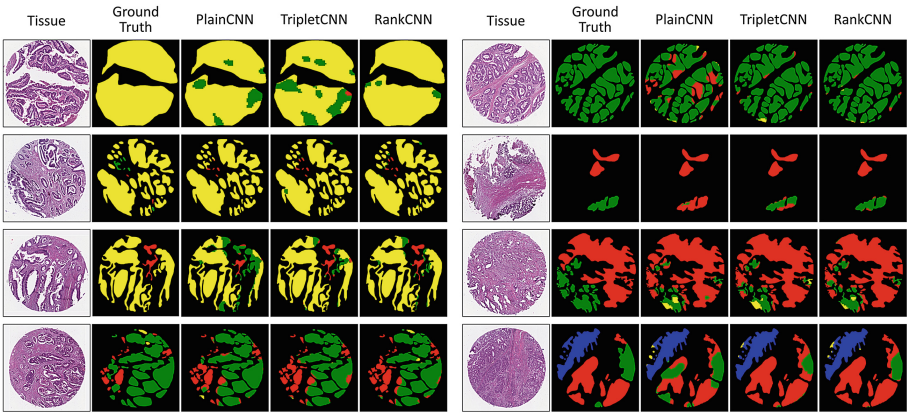

Fig. 2. Visual assessment of colorectal cancer grading. Blue, yellow, green, and red colors represent benign, WD, MD, and PD, respectively.

4 Conclusions

We present a ranking-based deep convolutional neural network that is equipped with the state-of-the-art deep neural network and the ranking loss. The ranking loss attempts to achieve a larger inter-class distance among different pathology classes with respect to the aggressiveness of cancer, leading to the correct ordering of tissue samples with differing cancer grades. The experimental results demonstrate that a substantial performance gain can be obtained without any major modifications of the network architecture, highlighting the effectiveness of the proposed ranking loss.

Acknowledgements. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.NRF-2021R1A4A1031864).

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**, 394–424 (2018)
2. Berglund, R.K., Jones, J.S.: A practical guide to prostate cancer diagnosis and management. Clevel. Clin. J. Med. **78**, 321 (2011)
3. Elmore, J.G., et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA **313**, 1122–1132 (2015)
4. Williams, B.J., Bottoms, D., Treanor, D.: Future-proofing pathology: the case for clinical adoption of digital pathology. J. Clin. Pathol. **70**, 1010–1018 (2017)
5. Niazi, M.K.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. Lancet Oncol. **20**, e253–e261 (2019)
6. Kwak, J.T., Hewitt, S.M., Sinha, S., Bhargava, R.: Multimodal microscopy for automated histologic analysis of prostate cancer. BMC Cancer **11**, 62 (2011)
7. Doyle, S., Feldman, M.D., Shih, N., Tomaszewski, J., Madabhushi, A.: Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. BMC Bioinformatics **13**, 282 (2012)
8. Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A.: A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Trans. Biomed. Eng. **59**, 1205–1218 (2010)
9. Kwak, J.T., Hewitt, S.M.: Multiview boosting digital pathology analysis of prostate cancer. Comput. Methods Prog. Biomed. **142**, 91–99 (2017)
10. Wang, X., et al.: Machine learning models for multiparametric glioma grading with quantitative result interpretations. Front. Neurosci. **12**, 1046 (2019)
11. Xu, Y., et al.: Multi-label classification for colon cancer using histopathological images. Microsc. Res. Techniq. **76**, 1266–1277 (2013)
12. Vu, Q.D., Kwak, J.T.: A dense multi-path decoder for tissue segmentation in histopathology images. Comput. Methods Prog. Biomed. **173**, 119–129 (2019)
13. Hou, L., et al.: Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. Pattern Recogn. **86**, 188–200 (2019)
14. Graham, S., et al.: Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. **58**, 101563 (2019)
15. Litjens, G., et al.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Rep. **6**, 26286 (2016)
16. Hou, L., Agarwal, A., Samaras, D., Kurc, T.M., Gupta, R.R., Saltz, J.H.: Robust histopathology image analysis: to label or to synthesize? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8533–8542 (2019)
17. Arvaniti, E., et al.: Automated Gleason grading of prostate cancer tissue microarrays via deep learning. Sci. Rep. **8**, 1–11 (2018)
18. Aresta, G., et al.: Bach: Grand challenge on breast cancer histology images. Med. Image Anal. **56**, 122–139 (2019)
19. Ertosun, M.G., Rubin, D.L.: Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. In: AMIA Annual Symposium Proceedings, p. 1899. American Medical Informatics Association

20. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778
22. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708
23. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520
24. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823
25. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification (2017). https://arxiv.org/abs/1703.07737
26. Tan, M., Le, Q.V.: Efficientnet: rethinking model scaling for convolutional neural networks (2019). https://arxiv.org/abs/1905.11946
27. Jurman, G., Riccadonna, S., Furlanello, C.: A comparison of MCC and CEN error measures in multi-class prediction. PLoS One **7**, e41882 (2012)