IJCARS manuscript No.
(will be inserted by the editor)

Robust Endocytoscopic Image Classification based on Higher-Order Symmetric Tensor Analysis and Multi-Scale Topological Statistics

Hayato Itoh · Yukitaka Nimura · Yuichi Mori · Masashi Misawa · Shin-Ei Kudo · Kinichi Hotta · Kazuo Ohtsuka · Shoichi Saito · Yutaka Saito · Hiroaki Ikematsu · Yuichiro Hayashi · Masahiro Oda · Kensaku Mori

Received: 22 January 2020 / Accepted: 2 September 2020

H. Itoh
Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
Tel.: +81-52-789-5688
Fax: +81-52-789-3815
E-mail: hitoh@mori.m.is.nagoya-u.ac.jp

Y. Nimura
Information Strategy Office, Information and Communications, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

Y. Mori, M. Misawa, and S.-E. Kudo
Digestive Disease Center, Showa University Northern Yokohama Hospital, Chigasaki-chuo 35-1, Tsuduki-ku, Yokohama, 224-8503, Japan

K. Hotta
Division of Endoscopy, Shizuoka Cancer Center, Shimonagakubo 1007, Nagaizumi-cho, Sunto-gun, Shizuoka, 411-8777, Jpan

K. Ohtsuka
Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Yushima 1-5-45, Bunkyo-ku, Tokyo 113-8510, Japan

S. Saito
Department of Gastroenterology, Cancer Institute Hospital of Japanese Foundation for Cancer Research, Ariake 3-8-31, Koto-ku, Tokyo 135-8550, Japan

Y. Saito
Endoscopy Division, National Cancer Center Hospital, Tsukiji 5-1-1 , Chuo-ku, Tokyo, 104-0045, Japan

H. Ikematsu
Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwanoha 6-5-1, Kashiwa, 277-8577, Japan

Y. Hayashi, M. Oda, and K. Mori
Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

Abstract *Purpose* : An endocytoscope is a new type of endoscope that enables users to perform conventional endoscopic observation and ultramagnified observation at the cell level. Although endocytoscopy is expected to improve the cost-effectiveness of colonoscopy, endocytoscopic image diagnosis requires much knowledge and high-level experience for physicians. To circumvent this difficulty, we developed a robust endocytoscopic (EC) image classification method for the construction of a computer-aided diagnosis (CAD) system, since real-time CAD can resolve accuracy issues and reduce interobserver variability.
Method: We propose a novel feature extraction method by introducing higher-order symmetric tensor analysis to the computation of multi-scale topological statistics on an image, and we integrate this feature extraction with EC image classification. We experimentally evaluate the classification accuracy of our proposed method by comparing it with three deep-learning methods. We conducted this comparison by using our large-scale multi-hospital dataset of about 55,000 images of over 3,800 patients.
Results: Our proposed method achieved an average 90% classification accuracy for all the images in four hospitals even though the best deep-learning method achieved 95% classification accuracy for images in only one hospital. In the case with a rejection option, the proposed method achieved expert-level accurate classification. These results demonstrate the robustness of our proposed method against pit-pattern variations, including differences of colours, contrasts, shapes, and hospitals.
Conclusions: We developed a robust EC image classification method with novel feature extraction. This method is useful for the construction of a practical CAD system, since it has sufficient generalisation ability.

Keywords Endocytoscopy · CAD · pathological pattern classification · machine learning · texture analysis

1 Introduction

The early detection of colorectal cancer is a critical problem because the survival rate for the cancer particularly depends on the stage of the cancer at diagnosis. Optical diagnosis of diminutive colorectal polyps is a promising approach for the improvement of the cost-effectiveness of colonoscopy and the reduction of polypectomy-related complications, according to the European Society of Gastrointestinal Endoscopy and the American Society of Gastrointestinal Endoscopy (ASGE). A new optical device, an endocytoscope has recently been developed for minimally-invasive diagnosis. Endocytoscopy enables users to perform direct observation of cells and their nuclei on the colon wall, that is, pit patterns at a maximum of 500-times ultramagnification, as shown in Fig. 1. Endocytoscopy can become an alternative method to biopsy as a tool for real-time diagnosis [1], since pit patterns on endocytoscopic (EC) images, as shown in Fig. 2, capture similar strictures to histopathological patterns. However, much pathological knowledge and clinical experience are necessary for

accurate endocytoscopy diagnosis. For example, Mori et al. [2] report that diagnostic performances of non-experts and experts for small ($\leq$10 mm) lesions are and % and 89-93% accuracy, respectively, in the recognition of neoplastic lesions.

Image classification techniques have the potential to resolve accuracy issues and reduce interobserver variability in optical diagnosis for endoscopic images [2–11]. Therefore, an accurate and robust image classifier is essential to the construction of a real-time computer-aided diagnosis (CAD) system. For this classification, determining how to represent image patterns is the most important factor. Jachin et al. proposed texture-feature extraction with selection of a region of interest [3]. Mesejo et al. proposed feature extraction based on texture, colour, and shape information [4]. Häfner et al. proposed colour-texture feature extraction [5]. Tamaki et al. proposed transformation-invariant local-feature-based feature extraction [6] and various kinds of wavelet-based feature extraction [7]. Mori et al. proposed the combination of local binary features and nuclei features [2]. In addition to these handcraft features, deep-learning-based feature extraction and classification have been proposed recently [8–10]. These image classification techniques are useful for the construction of a CAD system. Kominami et al. reported that non-expert endoscopists with a image-classification-based CAD system may more easily achieve sufficient accuracy to meet the criterion of the preservation and incorporation of valuable endoscopic innovation initiatives of ASGE [11].

We developed a robust EC image classifier as an essential function in several requirements for the construction of a practical CAD system. Since a practical CAD system might be used in many hospitals, a classifier should have a robust response to variants of pit patterns and generalisation ability for new (unseen) data. In the process of developing a robust EC image classifier, we have proposed a novel feature extraction method by introducing higher-order symmetric tensor analysis (HOSTA) [12,13] to the computation of topological statistics on an image. Previous works [5–10] suggest that topological information is essential for representing pit patterns on the surfaces of polyps. As the natural extension of texture features [14] used in previous works [3–5], HOSTA enable us to compute detailed topological statistics by decomposing the local distribution of gradients into several principal directions. By computing topological statistics with local contrast normalisation for results of HOSTA, we obtain robust features against variations of pit patterns with differences of colour, contrast, and shape. This robustness results in the high generalisation ability in EC image classification, since the staining styles in different hospitals generate colour and contrast variations in addition to variations of the shapes of pit patterns. Using this novel feature extraction and a linear classifier, we constructed an EC image classifier. We experimentally evaluated the classification performance of our classifier by comparing it with three deep-learning methods to demonstrate the validity of the proposed feature extraction. We conducted this comparison by using our large-scale multi-hospital dataset of about 55,000 images of over 3800 patients, which is the largest EC image dataset, to the best of our knowledge. In the evaluation, we trained our clas-

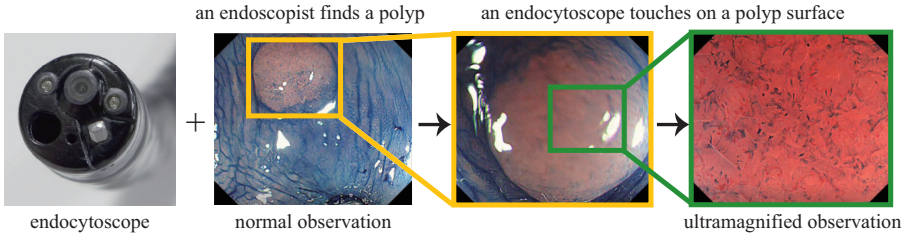


Fig. 1 Endocytoscope and ultramagnified observation. The first figure shows the CF-H290ECI endocytoscope (Olympus, Tokyo, Japan). As shown from the second to fourth images, endocytoscopy offers an ultramagnified view in vivo, where the tip of the endocytoscope touches the surface of a polyp.

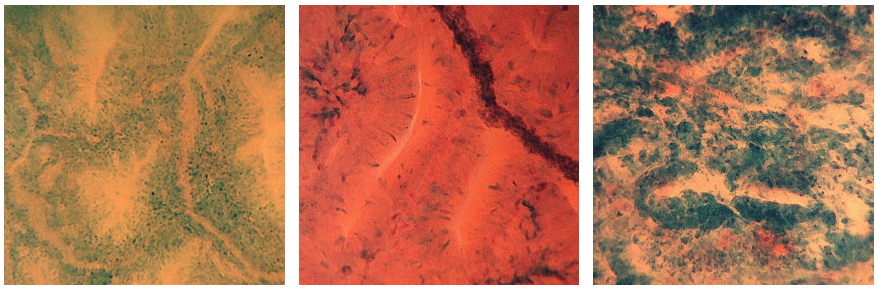


Fig. 2 Typical examples in endocytoscopy. (a) non-neoplasia (b) adenoma (c) invasive cancer. Adenoma and invasive cancer are categorised as neoplasia.

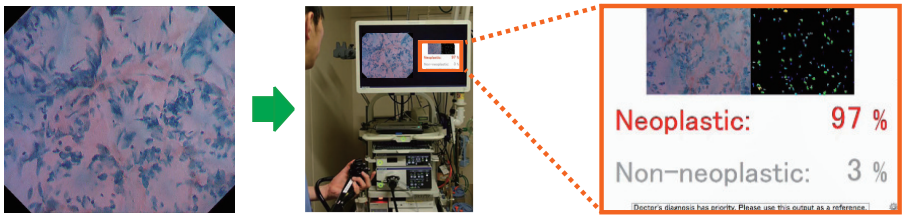


Fig. 3 CAD system based on EC image classifier for endocytoscopy. Left: input endocytoscopic image. Right: indication of classification result.

sifier with only one hospital's data and tested it with three other hospitals' data to demonstrate its generalisation ability against unseen data.

2 Methods

2.1 Neoplasia classification by endocytoscopic images

For binary classification between non-neoplastic and neoplastic EC images, we set categorical labels of non-neoplasia and neoplasia to be 0 and 1, respectively. For a label $\mathcal{L} \in \{0, 1\}$ and an input image $\mathcal{X}$, we assume that $\mathcal{X}$ belongs to a category of $\mathcal{L}$ with a probability $P(\mathcal{L}|\mathcal{X})$. To construct a precise decision function

$$g(\mathcal{X}) = \begin{cases} 1, & P(\mathcal{L} = 1|\mathcal{X}) > \tau, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where τ is a criterion that gives the minimum classification error, we have to construct an appropriate probabilistic model $P(\mathcal{L}|\mathcal{X})$. After we decide the value of τ, we adopt a rejection option to reject a classification of low output probability due to its low confidence [15]. For a query $\mathcal{X}$, the classifier outputs

$$\begin{cases} g(\mathcal{X}), & \max(\{P(\mathcal{L}|\mathcal{X})|\mathcal{L} \in \{0, 1\}\}) > \kappa, \\ \text{reject a classification}, & \text{otherwise}, \end{cases} \tag{2}$$

where κ is a rejection criterion. The constructed classifier is implemented in a CAD system to show the classification results with estimated probability, as shown in Fig. 3.

2.2 Proposed method

We constructed the pipeline shown in Fig. 4(a) for endocytoscopic image classification. For this pipeline, we proposed texture feature extraction based on HOSTA and multi-scale topological statistics.

Let $X : \Omega \in \mathbb{R}^2 \to \mathbb{R}$ be a grayscale image obtained from an RGB image $\mathcal{X}$. We apply a median filter of 3×3 kernel to X for denoising in pre-processing. Furthermore, we apply the two-dimensional isotropic Gaussian filter G_σ of a standard deviation σ to a denoised image. A grayscale value of $G_\sigma(X)$ is a scalar, that is, a 0th-order tensor. $\nabla G_\sigma(X)$ is a vector, that is, a 1st-order tensor. For a smoothed image $G_\sigma(X)$, a Hessian is defined by $\boldsymbol{H} = (\nabla \otimes \nabla)G_\sigma(X)$. $\boldsymbol{H}$ is a matrix, that is a 2nd-order tensor. Using an l-fold outer product, we have a lth-order symmetric tensor as an extension of a Hessian matrix for $G_\sigma(X)$ by

$$\mathcal{T} = (\nabla^{\otimes l})G_\sigma(X) = (\nabla \otimes \nabla \otimes \nabla \cdots)\nabla G_\sigma(X), \tag{3}$$

where l is the number of nabla ∇. Note that a 2nd-order symmetric tensor is a Hessian matrix.

In an analogy of a rank-1 matrix, a symmetric rank-1 lth-order tensor is defined by using the lth-fold outer product. Therefore, any symmetric lth-order

tensor $\mathcal{T}$ can be expressed by a linear combination of rank-1 tensors

$$\mathcal{T} = \sum_{i=1}^{l} \lambda_i^{(l)} \boldsymbol{v}_i^{(l)\otimes l}, \quad \boldsymbol{v}_i^{(l)} = \begin{pmatrix} \cos(-\theta + (i-1)\pi/l) \\ \sin(-\theta + (i-1)\pi/l) \end{pmatrix}, \tag{4}$$

where $\boldsymbol{v}_i^{(l)}$ corresponds to $\lambda_i^{(l)}$ with the condition $\lambda_i^{(l)} \geq \lambda_{i+1}^{(l)}$ in descending order. In Eq. (4), a relaxation of orthogonality is introduced for l two-dimensional vectors in a fixed spacing of π/l between neighbours as principal directions. HOSTA decomposes a local distribution of gradients at each point into its principal directions. For $l = 2, 3, 4$, a decomposition method has been proposed with proofs of unique solutions [13].

By selecting a scale from $\sigma = 1, 2, 4, 8$, we obtained a best smoothed image. For the selected scale, we have distributions of $G_\sigma(X)$, $|\nabla G_\sigma(X)|$, $\lambda_{\max}^{(l)}$, and $\lambda_{\min}^{(l)}$ with $l = 2, 3, 4$ as eight images. For $l = 2, 4$, we set $\lambda_{\max}^{(l)} = \max(\{\lambda_i^{(l)}\}_{i=1}^{l})$ and $\lambda_{\min}^{(l)} = \min(\{\lambda_i^{(l)}\}_{i=1}^{l})$. For $l = 3$, we set $\lambda_{\max}^{(l)} = \max(\{|\lambda_i^{(l)}|\}_{i=1}^{l})$ and $\lambda_{\min}^{(l)} = \min(\{|\lambda_i^{(l)}|\}_{i=1}^{l})$ due to the odd-function property of homogeneous form of odd-order symmetric tensors [13]. Figure 4(b) shows an example of eight images obtained by HOSTA. By using eigenvalues of principal axes for higher-order derivatives, we can express detail local topology of a grayscale-value distribution on an image as handcraft feature maps.

We then extract multi-scale topological statistics from eight images obtained from HOSTA. For one of these eight images, we compute a discritised 16-level grayscale image $I(k, l) \in \{1, 2, 3, \ldots 16\}^{H \times W}$, where H and W express height and width, respectively, by applying local constant normalisation with a kernel of 32×32 at each point. By using $I(k, l)$, we compute gray-level co-occurrence

$$\begin{aligned} p(i, j; d, \theta) = \sharp\{(k, l) \in (H \times W) \mid \\ I(k, l) = i, I(k + \lfloor d \sin\theta + 0.5 \rfloor, l + \lfloor d \cos\theta + 0.5 \rfloor) = j\}, \end{aligned} \tag{5}$$

where we use cardinality operator $\sharp$, and set indexes of grayscale level $i, j \in \{1, 2, \ldots, 16\}$, a distance to neighbours $d \in \{2, 5, 8\}$, and a connection direction $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. From one image, we obtain a gray-level co-occurrence matrix $\boldsymbol{C}_{d,\theta} = (p(i, j; d, \theta))$, and its 13 topological statistics defined in Haralick feature, except for the max correlation coefficient [14]. For each d, we average each of the 13 statistics among four directions. This averaging results in rotation-invariant feature extraction. We concatenate these statistics for eight HOSTA results with three scales of $d \in \{2, 5, 8\}$, and obtain a 312-dimensional feature vector for an input image. Note that we select $\sigma = 2$ as the best hyperparameter with respect to classification performance.

We finally classify extracted feature vector $\boldsymbol{x} \in \mathbb{R}^{312}$ using a linear support vector machine (SVM) [16]. If we have ideal feature extraction, a linear classifier can theoretically achieve robust classification [10]. Furthermore, a linear SVM is apt to overfit to a given pattern distribution less than nonlinear methods such as nonlinear SVM and deep-learning methods. A SVM gives a function $f(\cdot)$, which returns the score between an input feature vector and the

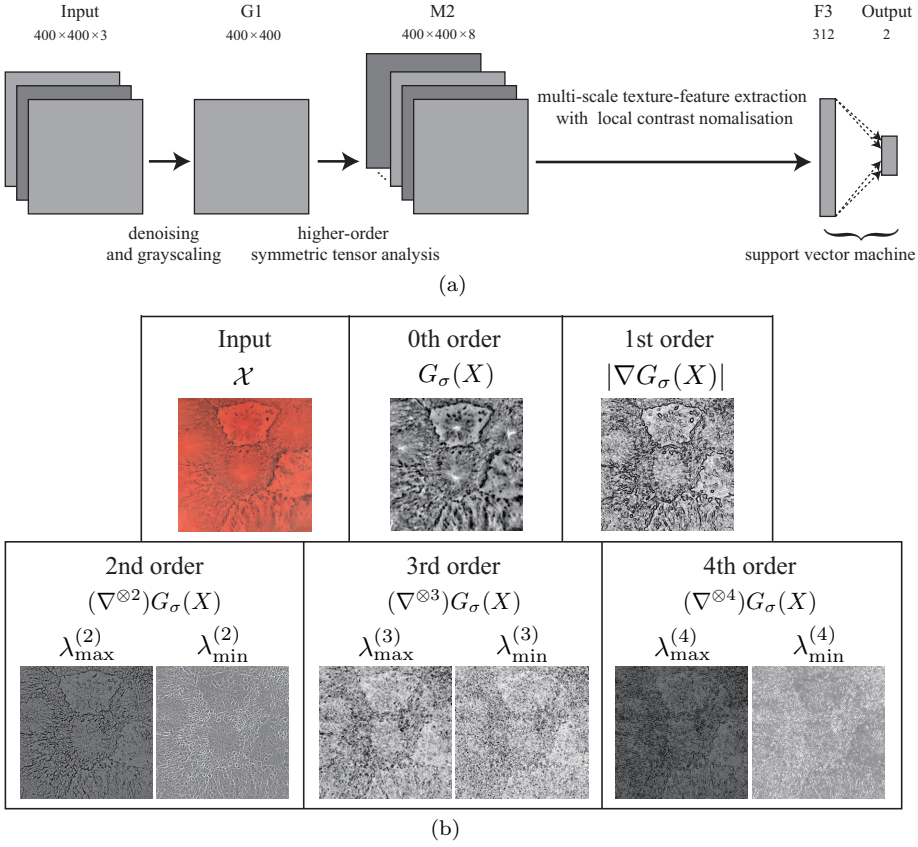


Fig. 4 Proposed method. (a) pipeline of proposed feature extraction and classification. (b) example of lth-order tensor analysis. We generate eight images from an input colour image for the computation of topological statistics. Here, we set $l = 0, 1, 2, 3, 4$ and $\sigma = 2$.

hyperplane that differentiates the categories. For the two categories with label $\mathcal{L} \in \{0, 1\}$, the output of the pipeline is given by the approximated probability

$$P(\mathcal{L} = 1|\boldsymbol{x}) \approx \frac{1}{1 + \mathrm{e}^{(af(\boldsymbol{x})+b)}}, \tag{6}$$

where a and b are the parameters in Platt's method [17]. We can obtain the decision function $f(\cdot)$ and parameters a, b by training of a SVM and maximum likelihood estimation, respectively.

2.3 Deep-learning methods

We constructed three deep-learning architectures for evaluations of our proposed method. For all these architectures, we used a softmax function to obtain

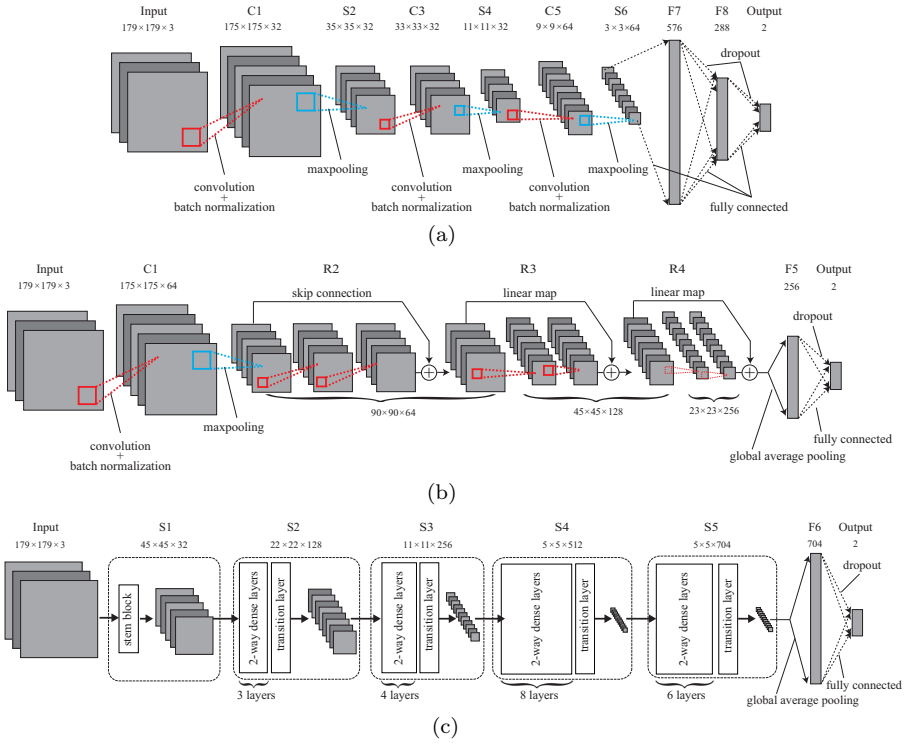


Fig. 5 Architectures of deep-learning methods. (a) three-convolution-layer CNN. (b) seven-convolution-layer ResNet. (c) 21-dense-layer PeleeNet. In (a)-(c), the number beside the squares shows the size of the array in processing. Note that an input image is a three-channel RGB image.

the function in Eq. (1). In our previous works, we found that a deep structure is inappropriate for texture-based EC image classification [9,10], since the fixed-scale texture can be expressed as a simple combination of low-level geometrical features. For the extraction of a simple combination of low-level geometrical features, a shallow architecture is suitable [8,18]. Therefore, we constructed our first architecture as a three-block convolutional neural network (CNN), as shown in Fig. 5(a). We used leaky ReLU [19] of negative slope 0.30 for activation functions in each layer. We applied batch normalisation [20] and dropout [21] with a ratio of 0.50 for convolutional and fully-connected layers, respectively. For this architecture, we assumed three cases: no weight regularisation, and L_1 and L_2 regularisation (weight decay) [15] for each convolution layer. We set the coefficient of regularisation term to be 1.0×10^{-2} for these weight decays. The total number of parameters in the CNN was 529,794.

For the the second architecture, we constructed a three-residual-unit ResNet by replacing three blocks of the first architecture with residual units [22]. Figure 5(b) illustrates our ResNet. Note that one convolution layer was added

before the three residual units in this ResNet. The learning of ResNet is a kind of ensemble learning, which consists of shallow convolutional layers in ResNet [23]. For the building a residual unit, we adopted the same structure of the original building block [22]. For the weight decay in each residual unit, we set a coefficient of 1.0×10^{-5}. We used dropout with a ratio of 0.20 for the outputs of the global average pooling. The total number of parameters in the ResNet was 1,228,034.

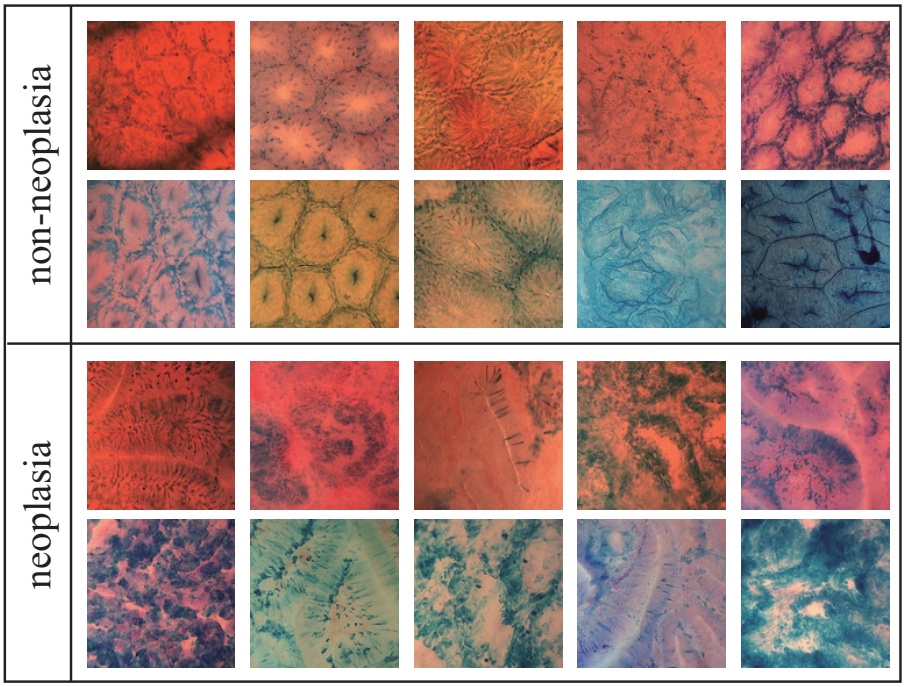
For the third architecture, we used PeleeNet [24] by changing the size of the input and output layers as shown in Fig 5(c). This deep architecture is a refined variant of MobileNet, which was designed for mobile applications with less spatial complexity, with 21 dense layers of 105 convolutional layers. We could use a large minibatch size for training of the PeleeNet, since it had only 2,114,250 parameters for 105 convolutional layers.

3 Image acquisition and dataset construction

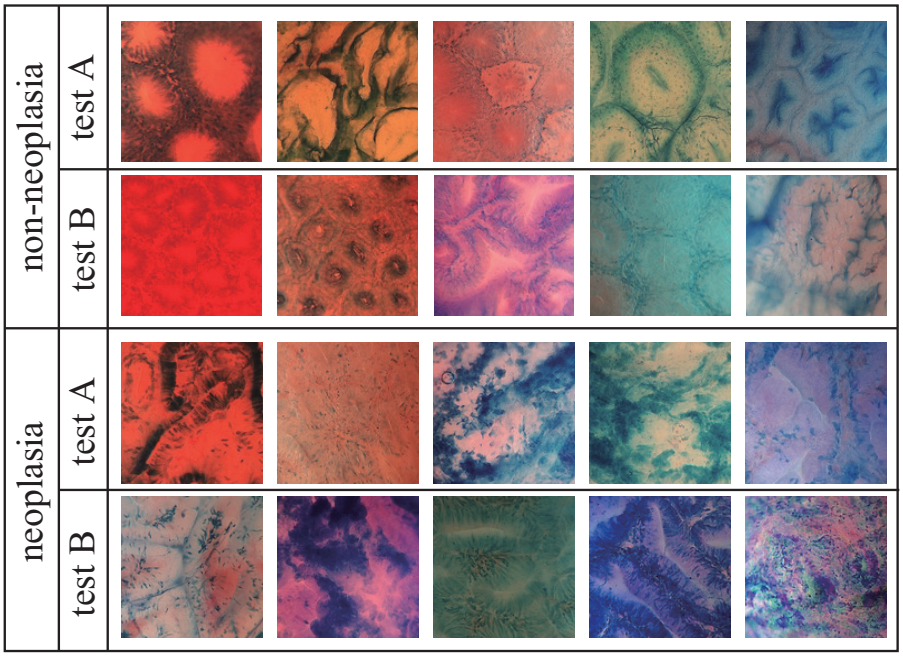
We built an EC image dataset for the development of a robust EC image classification method. First, we collected about 50,000 EC images of 3,522 patients of conventional colonoscopy at a main hospital, where endocytoscopy had been developed by using a prototype of an endocytoscope over many years, with 47 endoscopists. For this collection, an endoscopist took from 10 to 200 images of each polyp by changing the target position at its stained surface. From rare pit patterns, an endoscopist collected many images at several positions to collect a wide variety of patterns. Second, we divided these images into three types of data: training, validation, and test-A data, without the duplication of patients. Finally, we collected about 4,300 EC images from normal colonoscopy at three other hospitals, for use as test-B data. These were collected from January 2016 to July 2019 with IRB approvals of each facility. Based on pathological diagnosis via biopsy, we gave a categorical label to each image. Table 1 summarises the amount of the four kinds of data. Figures 6(a) and (b) show examples of images in the training data, and test-A and -B data, respectively. The comparison between Figs. 2 and 6 summarises the large variety of EC image patterns due to the difference of colour, contrast, and shape of the pit patterns. The part of the difference of colour and contrast comes from the difference of staining styles with two dyes: crystal violet and methylene blue.

4 Experimental results

We conducted four experiments to evaluate our dataset and proposed classification method. First, we evaluated the scales of our dataset as to whether it had enough pit patterns to obtain generalisation ability against the validation and test-A data. Second, we compared the performance on the training and validation data among the CNN, the CNN with weight decay, the ResNet, and the PeleeNet. This showed the baseline characteristics of each architecture.



(a)



(b)

Fig. 6 Examples of images. (a) training data. (b) test-A and test-B data.

Table 1 Summary of the number of EC images in our dataset, which consists of training, validation, test-A, and test-B data. Test-A and -B data consist of images captured in a single main hospital, and in three other hospitals, respectively. Training data contains images of about 2,800 cases (patients). Note that there is no duplication of patients among these divided training, validation, and test-A and test-B data.

	training	validation	test A	test B	total
♯ of non-neoplasia	13,180	1,592	1,638	662	17,599
♯ of neoplasia	26,462	3,213	3,262	3,965	36,902
♯ of both categories	39,642	4,805	4,900	4,267	54,501

Third, we evaluated the classification performance of our proposed method by comparing it with the deep-learning methods. Finally, we evaluated the classification accuracy of our proposed method with the rejection option defined in Eq. (2). For the evaluations, we defined accuracy, sensitivity, and specificity by the ratios of correctly classified images for all, neoplastic, and non-neoplastic images, respectively. In all of the experiments with the deep-learning methods, we used a single GPU V100 of 32 GB (NVIDIA) with Keras of Tensorflow backends. For training of the models, we used binary-cross entropy with weight balancing for all of the deep-learning methods.

For the first experiment, we trained the three-block CNN by using one eighth, one quarter, one half, and all of the training data. For training of the CNN, we set the mini-batch size to be 1,024 with He's initialisation [25] and an Adam optimiser [26] of base learning rate lr= 1.0×10^{-4} for 300 epochs. At 100, 150, 200, and 250 epochs, the learning rate was multiplied by 0.10. In this training, we used data argumentation with a rotation transform of $0, 30, 60, \ldots, 330$ degrees with random mirroring with respect to the horizontal and vertical axes. From the four-type training, we selected four trained models that had the highest classification accuracy in each training for the validation data with $\tau = 0.50$. Using these selected trained models, we confirmed the accuracy in the training, validation, and test-A data. Note that fine-tuning with ImageNet [28] did not contribute to the increase of the classification accuracy in preliminary experiments. This was caused by the differences of colour contrasts around edges, corners, and blobs between EC images and natural images. Figures 7(a)-(d) illustrate the learning curves of four-type training. Figure 7(e) summarises the relationship between the size of the training dataset and the accuracy.

For the second experiment, we trained the CNN with weight decay of L_1 and L_2, the ResNet, and the PeleeNet. For the training of the CNN with weight decay, we used the same procedure as in the first experiment. For the training of the ResNet and the PeleeNet, we set the mini-batch size to be 512 and 256, respectively. Training of the ResNet and the PeleeNet were performed with an Adam optimiser of base learning rate $lr = 1.0 \times 10^{-4}$ for 250 epochs. At 100, 150, and 200 epochs, the base learning rate lr is multiplied by 0.10. We used the same data argumentation procedure as for the CNN. We selected four

trained models that gave the highest classification accuracy in each training for the validation data with $\tau = 0.50$. Figures 8(a)-(d) illustrate the learning curves of these trained models.

For the third experiment, we trained a linear SVM with the proposed HOSTA-based texture features described in Sec 2.2. For the training of the SVM, we performed five-fold cross-validation for the mixture of the training and validation data to find the best hyperparameter in the SVM. We then had the six classification methods, and we evaluated the accuracy of these methods for test-A and test-B data. Figures 9(a) and (b) show the ROC curves for test-A and test-B data, respectively.

For the fourth experiment, we evaluated the classification accuracy of our proposed method with the rejection option. Figure 10 shows the accuracy for the rejection rate, which is the ratio of rejected queries in test-A and -B data.

5 Discussion

In Figs. 7(a) to (d), the gap between the training and validation curves became smaller as the number of cases increased. As summarised in Fig. 7(e), when there were over 1400 cases of training data, the accuracy gap between the validation and test data was almost zero. Furthermore, the accuracy gap among the training, validation, and test data was close to zero for the trained CNN with 2800 cases of training data. These results show the validity of our data generation, since the training data is enough to obtain generalisation ability for the validation and test-A data.

The comparison among Fig. 7(d) and Figs. 8(a) and (b) shows that the weight decay reduces the gap between the training and validation data. Furthermore, Fig. 8(c) illustrates that residual learning leads to better learning than a simple shallow CNN for the training and validation data. As shown in Fig. 8(d), the PeleeNet gives the highest accuracy for the training data, even though the accuracy for the validation data is almost the same as for the ResNet. Moreover, in Fig. 9(a), the ResNet, the PeleeNet, the CNN with weight decay, the CNN, and the proposed method give higher performance in descending order for the test-A data. These results imply that the PeleeNet is too deep of an architecture for EC image classification.

In Fig. 9(b), the proposed method is more accurate than all of the deep-learning methods for the test-B data. The PeleeNet and the ResNet had the worst and second worst performances, respectively. The comparison between Figs. 9(a) and (b) clarified that deep-learning methods overfit the pattern distribution of the main hospital. In particular, deeper architectures are apt to overfit to the pattern distributions on the training data. The proposed handcraft feature is mathematically well defined, and it does not depend on the pattern distribution of the training data. The proposed method achieves an average classification accuracy of 0.93 with rejection criterion $\kappa = 0.70$ for all hospitals, as shown in Fig. 10. Two expert endoscopists commented that about 10-15% of test images, which were rejected by the rejection option,

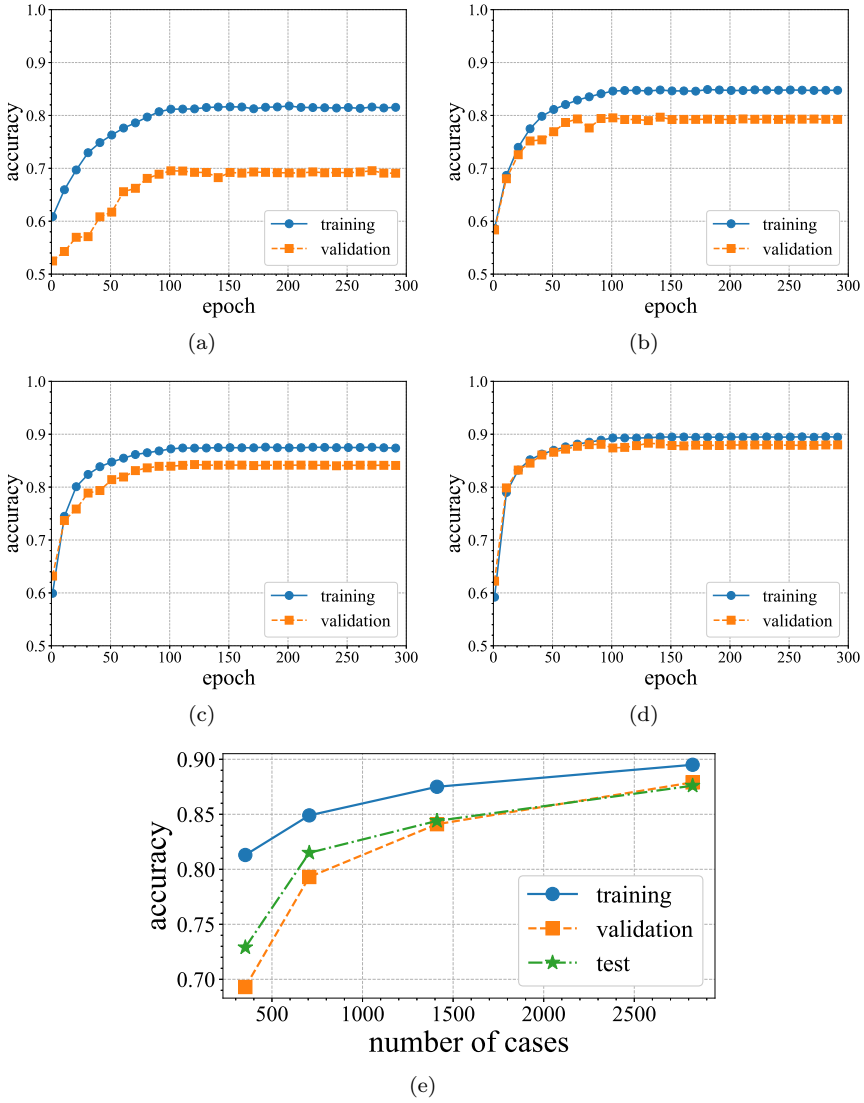


Fig. 7 Learning curves with glowing of training data for the CNN. (a)-(d) show the learning curves for learning with one eighth, one quarter, one half, and all of the training data, respectively. (e) Glowing of the classification accuracy with respect to the number of cases in the training data. In (a)-(d), the horizontal and vertical axes show the epoch and the accuracy for the training data. In (e), the horizontal and vertical axes show the number of cases in images for training and accuracy, respectively.

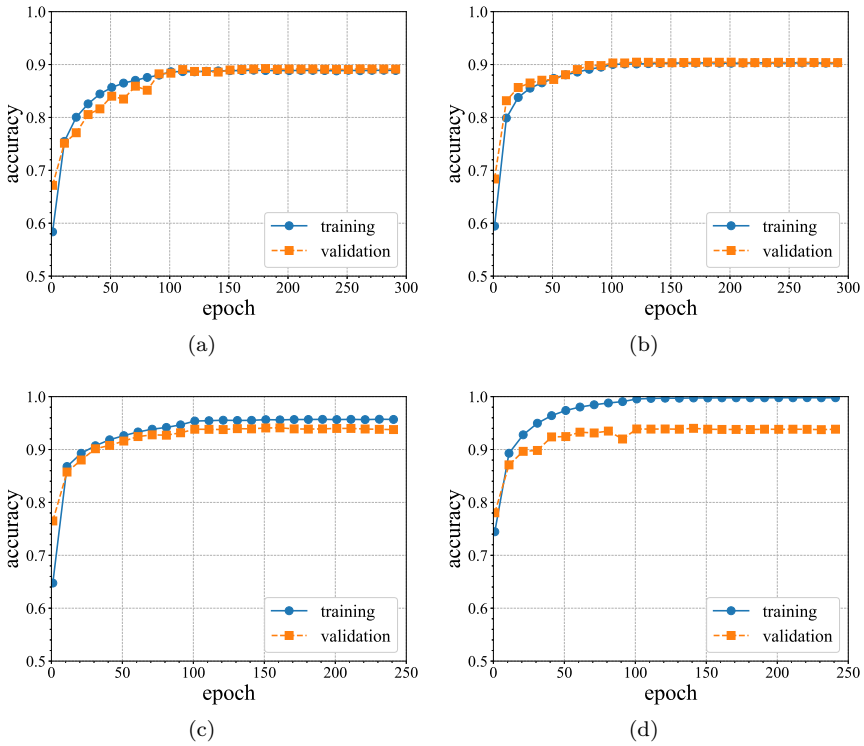


Fig. 8 Learning curves for learning with weight decay and residual units. (a) L_1-regularised CNN, (b) L_2-regularised CNN, and (c) ResNet. (d) PeleeNet. The horizontal and vertical axes show the epoch and accuracy for the training and validation data.

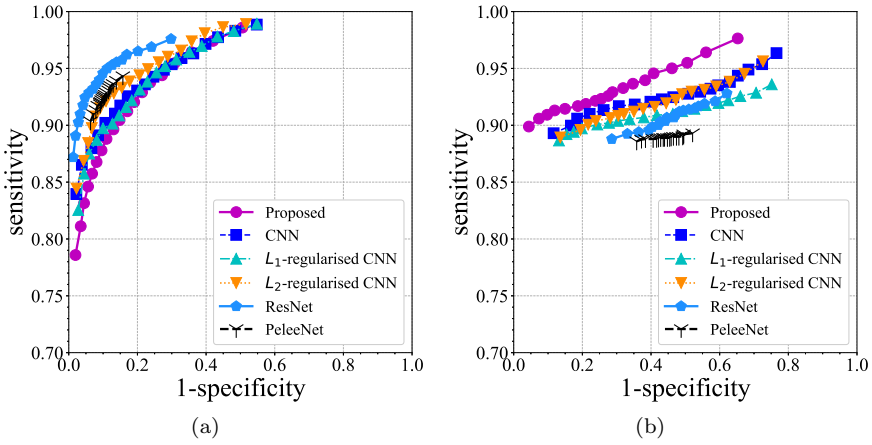


Fig. 9 ROC curves. (a) Test A (b) Test B. In (a) and (b), the horizontal and vertical axes express 1−specificity and sensitivity, respectively. For plotting, we changed the decision criteria τ in Eq. (1).

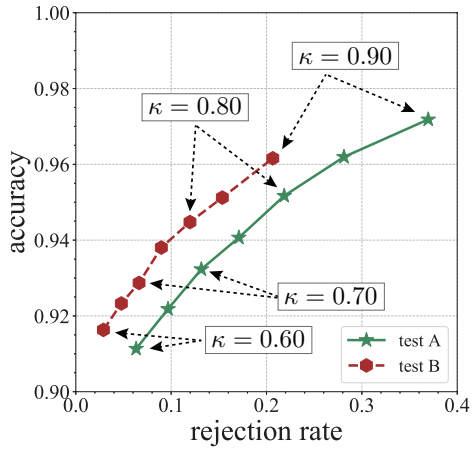


Fig. 10 Classification performance of proposed method with the rejection option for test-A and test-B data. The horizontal and vertical axes show the rejection rate and accuracy, respectively. We set $\kappa \in \{0.60, 0.65, \ldots, 0.90\}$ in Eq. (2) for plotting these curves.

were difficult queries even for endoscopists. This performance is approximately equal to that of an expert endoscopist [2,27].

6 Conclusions

We developed a robust EC image classifier by introducing HOSTA to the computation of topological statistics and integrating it into a linear SVM. This classification method with the rejection option had almost the same performance as expert endoscopists did in the evaluation with a large-scale multi-hospital dataset. Furthermore, comparison of the classification performances among the proposed methods and three deep-learning methods validated the high generalisation ability of our classification method, even though the deep-learning methods overfitted a specific hospital's distribution. A large-scale multi-hospital comparison revealed the difficulty of constructing a deep learning model with generalisation ability.

Acknowledgements This study was funded by grants from AMED (19hs0110006h0003), JSPS MEXT KAKENHI (26108006, 17H00867, 17K20099, 19K08403), and the JSPS Bilateral Joint Research Project.

Conflicts of interest

Kudo SE, Misawa M, and Mori Y received lecture fees from Olympus. Ohtsuka K reports personal fees and nonfinancial support from Olympus outside of this work. Mori K is supported by Cybernet Systems and Olympus (Research

grant) in this work, and by NTT outside of the submitted work. The other authors have no conflicts of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical committee of Nagoya University (No. 351, 357) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained via an opt-out procedure from all individual participants included in the study.

References

1. Mori Y, Kudo SE, Ikehara N, Wada Y, Kutsukawa M, Misawa M, Kudo T, Kobayashi Y, Miyachi H, Yamamura F, Ohtsuka K, Inoue H, Hamatani S (2013) Comprehensive diagnostic ability of endocytoscopy compared with biopsy for colorectal neoplasms: a prospective randomized noninferiority trial, Endoscopy, 45(2): 98-105
2. Mori Y, Kudo SE, Chiu PW, Singh R, Misawa M, Wakamura K, Kudo T, Hayashi T, katagiri A, Miyachi H, Ishida F, Maeda Y, Inoue H, Nimura Y, Oda M, Mori K (2016) Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study. Endoscopy 48(12): 1110-1118
3. Fu JJ, Yu YW, Lin HM, Chai JW, Chen CC (2014) Feature Extraction and Pattern Classification of Colorectal Polyps in Colonoscopic Imaging. Computerized Medical Imaging and Graphics 38(4): 267-275
4. Mesejo P, Pizarro D, Abergel A, Rouquette O, Beorchia S, Poincloux L, Bartoli A (2016) Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy. IEEE Transactions on Medical Imaging 35(9): 2051-2063
5. Häfner M, Liedlgruber M, Uhl A, Vécsei A, Wrba, F (2012) Color treatment in endoscopic image classification using multi-scale local color vector patterns. Medical Image Analysis 16(1): 75-86
6. Tamaki T, Yoshimura J, Kawakami M, Raytchev B, Kaneda K, Yoshida S, Takemura Y, Onji K, Miyaki R, Tanaka S(2013) Computer-aided colorectal tumor classification in NBI endoscopy using local features. Medical Image Analysis 17(1): 78–100
7. Wimmer G, Tamaki T, Tischendorf JJ, Häfner M, Yoshida S, Tanaka S, Uhl A (2016) Directional wavelet based features for colonic polyp classification. Medical Image Analysis, **31**: 16-36
8. Tamaki T, Sonoyama S, Hirakawa T, Raytchev B, Kaneda K, Koide T, Yoshida S, Mieno Hiroshi, Tanaka S (2016) Computer-Aided Colorectal Tumor Classification in NBI Endoscopy using CNN Features. Proc. Korea-Japan joint workshop on Frontiers of Computer Vision:61-65
9. Itoh H, Mori Y, Misawa M, Oda M, Kudo SE, Mori K (2018) Cascade classification of endocytoscopic images of colorectal lesions for automated pathological diagnosis. Proc. SPIE Medical Imaging 10575:269-274
10. Itoh H, Lu Z, Mori Y, Misawa M, Oda M, Kudo SE, Mori K (2018) Discriminative Feature Selection by Optimal Manifold Search for Neoplastic Image Recognition. Proc. ECCV workshops 4: 534-549
11. Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, Tamaki T, Koide T, Kaneda K, Chayama K (2016) Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointestinal Endoscopy 83: 643-649
12. Schultz T, Weickert J, Seidel HP (2009) A Higher-Order Structure Tensor. In: Visualization and Processing of Tensor Fields. *Springer*

13. Schultz T (2011) Topological Features in 2D Symmetric Higher-Order Tensor Fields. Computer Graphics Forum 30(3):841-850
14. Haralick RM, Shanmugam K, Dinstein I (1973) Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3(6): 610-621
15. Bishop CM (2011) Pattern Recognition and Machine Learning 2nd Edition, Springer
16. Vapnik VN (1998) Statistical Learning Theory. *Wiley*
17. Platt JC (1999) Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. Advances in large margin classifier: 61-74
18. Zeiler MD, Fergus R (2014) Visualizing and Understanding Convolutional Networks. Proc. ECCV: 818-833
19. Maas AL, Hannun AY, Ng AY (2013) Rectifier Nonlinearities Improve Neural Network Acoustic Models. Proc. International Conference on Machine Learning 30(1)
20. Ioffe S, Szegedy C (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proc. International Conference on Machine Learning 37: 448–456
21. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15: 1929-1958
22. He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. Proc. CVPR: 770-778
23. Veit A, Wilber M, Belongie S. (2016) Residual Networks Behave Like Ensembles of Relatively Shallow Networks. Proc. NIPS: 550–558
24. Wang J, Li X, Ling CX (2018) Pelee: A Real-Time Object Detection System on Mobile Devices, Proc. NIPS: 1963-1972
25. He K, Zhang X, Ren S, Sun J (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, Proc. ICCV: 1026-1034
26. Kingma DP, Ba J (2015) Adam: A Method for Stochastic Optimization. Proc. ICLR:1-13
27. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, Ohtsuka K, Urushibara F, Kataoka S, Ogawa Y, Maeda Y, Takeda K, Nakamura H, Ichimasa K, Kudo T, Hayashi T, Wakamura K, Ishida F, Inoue H, Itoh H, Oda M, Mori K (2018) Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Annals of Internal Medicine 169(6): 357-366
28. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC and Li FF. (2015) ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 115(3): 211-252