4-2022

Understanding the Mechanism of Deep Learning Frameworks in Lesion Detection for Pathological Images with Breast Cancer

Wei-Wen Hsu
Old Dominion University

Chung-Hao Chen
Old Dominion University

Chang Hao

Yu-Ling Hou

Xiang Gao

See next page for additional authors

Original Publication Citation

Hsu, W. W., Chen, C. H., Hao, C., Hou, Y. L., Gao, X., Shao, Y., Zhang, X., Wang, J., He, T. & Tai, Y. (2022). Understanding the mechanism of deep learning frameworks in lesion detection for pathological images with breast cancer. *Sensors and Materials*, *34*(4), 1337-1349. https://doi.org/10.18494/SAM3629

Sensors and Materials, Vol. 34, No. 4 (2022) 1337–1349
MYU Tokyo

S & M 2895

Understanding the Mechanism of Deep Learning Frameworks in Lesion Detection for Pathological Images with Breast Cancer

Wei-Wen Hsu,[1,2] Chung-Hao Chen,[1] Chang Hao,[2] Yu-Ling Hou,[2] Xiang Gao,[2] Yun Shao,[2] Xueli Zhang,[3] Jingjing Wang,[3] Tao He,[2] and Yanhong Tai[3*]

[1]Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529, USA
[2]Institute of Big Data Technology, Bright Oceans Corporation, Beijing 100093, China
[3]Department of Pathology, The Fifth Medical Center of PLA General Hospital, Beijing 100093, China

(Received September 15, 2021; accepted January 13, 2022)

Keywords: CADe system, lesion detection, deep features, visual interpretability

With the advances of scanning sensors and deep learning algorithms, computational pathology has drawn much attention in recent years and started to play an important role in the clinical workflow. Computer-aided detection (CADe) systems have been developed to assist pathologists in slide assessment, increasing diagnosis efficiency and reducing misdetections. In this study, we conducted four experiments to demonstrate that the features learned by deep learning models are interpretable from a pathological perspective. In addition, classifiers such as the support vector machine (SVM) and random forests (RF) were used in experiments to replace the fully connected layers and decompose the end-to-end framework, verifying the validity of feature extraction in the convolutional layers. The experimental results reveal that the features learned from the convolutional layers work as morphological descriptors for specific cells or tissues, in agreement with the diagnostic rules in practice. Most of the properties learned by the deep learning models summarized detection rules that agree with those of experienced pathologists. The interpretability of deep features from a clinical viewpoint not only enhances the reliability of AI systems, enabling them to gain acceptance from medical experts, but also facilitates the development of deep learning frameworks for different tasks in pathological analytics.

1. Introduction

In recent years, computer-aided technologies have been widely adopted to achieve automated inspection for health care delivery, having a wide range of applications from scalp inspection(1) to lung nodule detection(2) in radiology and lesion classification(3) in histopathology. However, biomedical image analysis is a complex task that relies on highly skilled domain experts, such as radiologists and pathologists. In pathology, the manual process of slide assessment is laborious and time-consuming, and incorrect interpretations due to specialist fatigue or stress may occur. Moreover, there has been an insufficient number of registered pathologists, especially in developing countries. As a result, the workload of pathologists has increased, which is becoming

*Corresponding author: e-mail: yanhongtai@hotmail.com
https://doi.org/10.18494/SAM3629

ISSN 0914-4935
https://myukk.org/

a severe problem in pathology. Recently, the techniques of image processing and machine learning have significantly advanced, and computer-aided detection/diagnosis (CADe/CADx) systems[4–7] have been developed to assist pathologists in slide assessment. Working as second opinion systems, they are designed to alleviate the workload of pathologists and avoid missing inspections.

Many early studies on machine learning focused on the development of classifiers. However, data scientists found feature extraction for data representation to be the bottleneck of performance in classification and detection tasks. Therefore, feature engineering, which focuses on methods to extract features and make machine learning algorithms work effectively, is becoming increasingly critical for overall performance. In representation learning, scientists aim to develop techniques that allow a system to automatically discover the representations needed for classification or detection from raw data. Since 2012,[8] the framework of deep convolutional neural networks (DCNNs) has achieved outstanding performance in many applications of computer vision. Many studies have shown that the classification results obtained using features extracted from deep convolutional networks, known as deep features, outperform those obtained with conventional approaches using hand-crafted features.[4–7] Accordingly, the deep learning framework has been widely adopted for pathological image analysis. Nonetheless, it has been difficult for medical specialists to accept such CADe/CADx systems with deep learning approaches since the deep learning framework operates in an end-to-end manner, taking raw images as the input and deriving the outcomes directly, and there is a lack of theoretical explanation of the mechanism for such systems with deep learning approaches. Most developers have solely focused on the efficacy of outcomes, without explaining why their proposed frameworks are effective.[9] Consequently, many medical specialists treat the deep learning framework as a "black box" and have doubts about the feasibility of such systems in clinical practice.

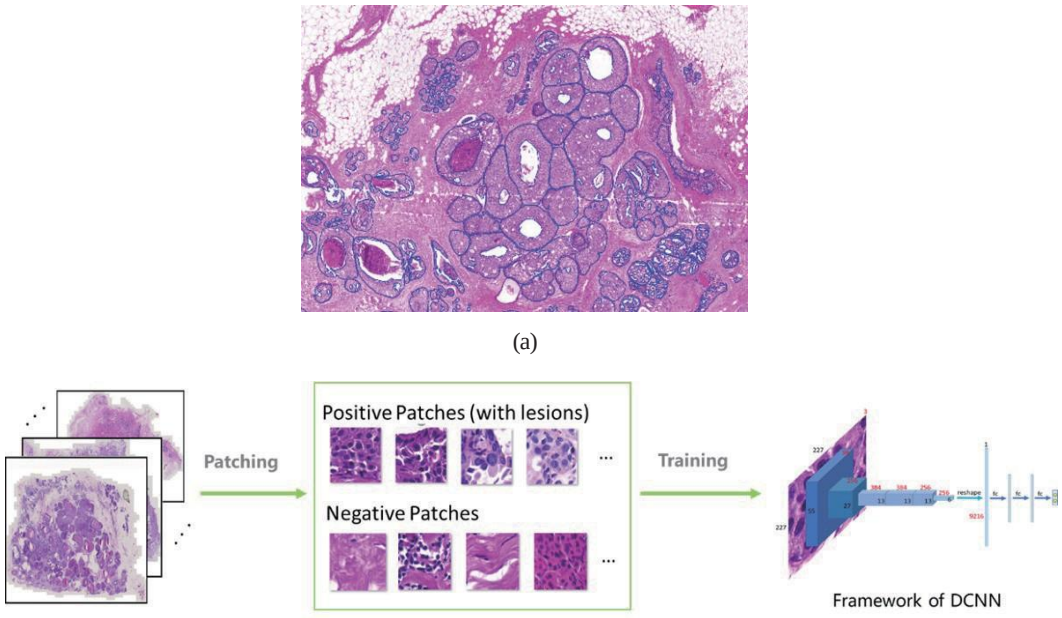
A DCNN comprises convolutional layers and fully connected layers that perform feature extraction and classification, respectively, during the optimization process. In convolutional layers, local features such as colors, end points, corners, and oriented edges are collected in the shallow layers. These local features in the shallow layers are integrated into larger structural features such as circles, ellipses, and specific shapes or patterns with increasing layer depth. Afterward, these structural patterns or shapes constitute high-level semantic representations that describe the feature abstraction for each category.[10] On the other hand, fully connected layers take the features extracted from the convolutional layers as the inputs and act as a classifier. These fully connected layers can encode the spatial correspondences of these semantic features and convey the co-occurrence properties between patterns or objects.[11]

Many studies have focused on the visual interpretability of deep learning models on the datasets of natural images[10,12–15] and showed that the mechanism of deep learning frameworks follows the prior knowledge for each category in the classification. The process of the classification system is concordant with human intuition in tasks of image classification.[16] However, for pathological image analysis, there has been insufficient research on explanations about the mechanism of systems with deep learning approaches, and the feasibility of such systems continues to be questioned by medical specialists.

The purpose of this study is to provide the visual interpretability of models to explain the mechanism of the deep learning framework in tasks of lesion detection for histology images. We studied the properties of the deep features extracted by deep learning models for lesion detection using digital pathology images at high magnification (×40). Four serial experiments were conducted to verify the properties of the following extracted DCNN features: (1) transferability to other classifiers, (2) meaningfulness in classification, (3) interpretability in terms of the domain knowledge of pathology, and (4) inspiration for exploring new cues in pathological image analysis. To focus on the properties of feature extraction in the convolutional layers, classifiers such as the support vector machine (SVM) and random forests (RF) were used in the experiment to replace the fully connected layers and decompose the end-to-end framework. However, we do not compare the performance of classifiers or discuss whether the substitution of fully connected layers can improve the performance because it is not the objective of this study.

2. Materials and Methods

In this study, 27 H&E stained samples of breast tissues with ductal carcinoma in situ (DCIS) were collected and digitized to the format of whole-slide images (WSIs). All lesions of DCIS were delineated precisely in blue by a registered pathologist, as shown in Fig. 1(a), and confirmed by a second registered pathologist. The original dataset was split into two sets: 15 cases for training and the remaining 12 cases for testing. To perform lesion detection through the WSIs, many small patches were sampled under high magnification (×40), a process called patching.[5,17]



(a)

(b)

Fig. 1. (Color online) Annotations of lesions and training of DCNN model for lesion detection. (a) Precise delineations of DCIS lesions on a WSI. (b) Training procedure in the deep learning framework for lesion detection.

There were two kinds of sampling sets: positive and negative sets. The positive set collected the patches with tumorous cells by sampling inside the annotated regions. On the other hand, the patches with normal cells or normal tissues were sampled outside the annotated regions and comprised the negative set. There were about 200k patches (positive samples: 98268, negative samples: 97997, total samples: 196265) sampled from the training dataset and about 85k patches (positive samples: 42115, negative samples: 41998, total samples: 84113) sampled from the testing dataset with a balanced class distribution in the initial stage of data acquisition. The training procedure in the deep learning framework for lesion detection is shown in Fig. 1(b).

In our experiments, the ImageNet pre-trained model of AlexNet[8] was used to perform transfer learning.[18] Within the pre-trained model of AlexNet, the feature size for each patch was 9216 × 1 in the classification. Since the classifiers of SVM and RF were used to replace the fully connected layers to decompose the end-to-end DCNN framework, the dataset would have been too large for SVM and RF if all 200k sampling patches were used in training. Therefore, to reduce the size of the dataset to make training feasible, 20k patches (positive:negative = 1:1) were randomly selected from the original training dataset, which were used as the real training dataset to fine-tune the deep learning model.[19] For performance evaluation, 10k patches (positive:negative = 1:1) were also randomly collected from the original testing dataset as the real testing dataset to compute the out-sample accuracy in patch classification.

To observe the influential patterns used in patch classification, the size of the field-of-views (FOVs) was computed to derive the mappings between the units (neurons) and their corresponding FOVs in the input image, as shown in Fig. 2. In DCNN models, the number of channels in the assigned convolutional layer indicates the number of learnable filters that represent particular features (the number of channels is 256 in the experiments). The units (neurons) in each channel represent the spatial orientation with respect to its corresponding FOV in the input image. A unit with high activation means that the learned pattern has a strong response to the corresponding region (FOV) in the image, reflecting the matching level between them. For visualization,[20] the activations of units in the assigned convolutional layer were recorded for all patches, and all patches were ranked by the units' activations for each channel.

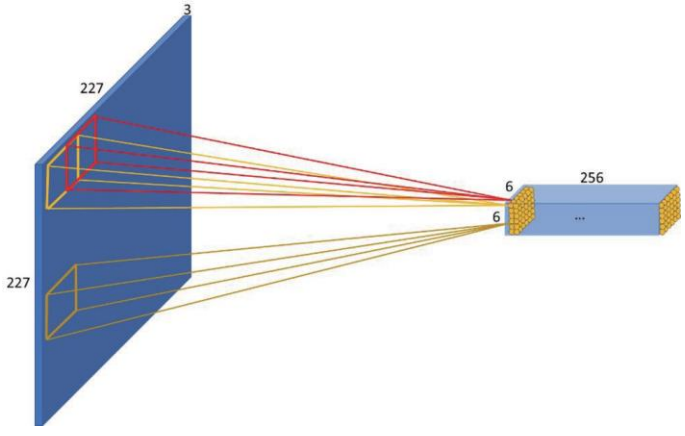


Fig. 2. (Color online) Mappings between units and their corresponding FOVs.

Afterward, the patches with the units having the top 100 activations for each channel were collected with the corresponding high-response region highlighted in a yellow bounding box, as shown in Fig. 3. Moreover, the corresponding activation maps were resized to the same size as the input image for better observation of the learned pattern and its spatial distribution. Figure 3 shows one of the examples where the learned pattern reflects the distribution of lymphocytes.

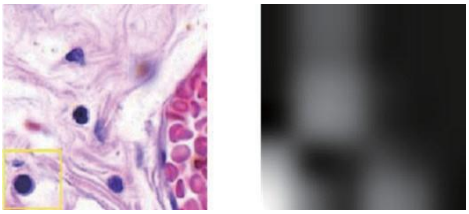
3. Experimental Results and Discussion

3.1 Exp #1: feature extraction in DCNN

Motivation: Even though the deep learning model is an end-to-end structure, it can be decomposed into two parts: convolutional layers for feature extraction and fully connected layers for classification. The goal of this experiment is to verify that the features extracted by the deep learning models are meaningful in classification and can be incorporated in other classifiers, rather than being exclusive to neural networks.

Hypothesis: Features extracted from the convolutional layers are meaningful in classification and can also be used with other classifiers.

Model: The end-to-end pre-trained AlexNet model was used in training and testing, and its structure is shown in Fig. 4. For the control group, the fully connected layers in AlexNet were replaced by other classifiers, including logistic regression (LR), SVM, and RF, as shown in Fig. 5.



(a) (b)

Fig. 3. (Color online) Patch (a) with the highest activation unit in channel No. 49 and (b) its corresponding activation map.

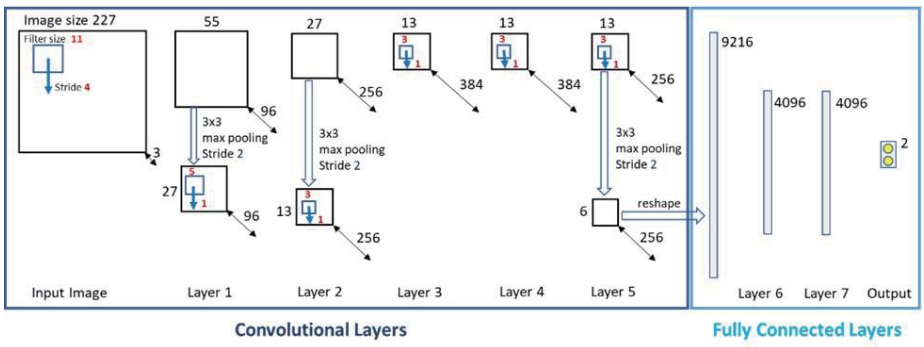


Fig. 4. (Color online) Structure of the end-to-end AlexNet model.

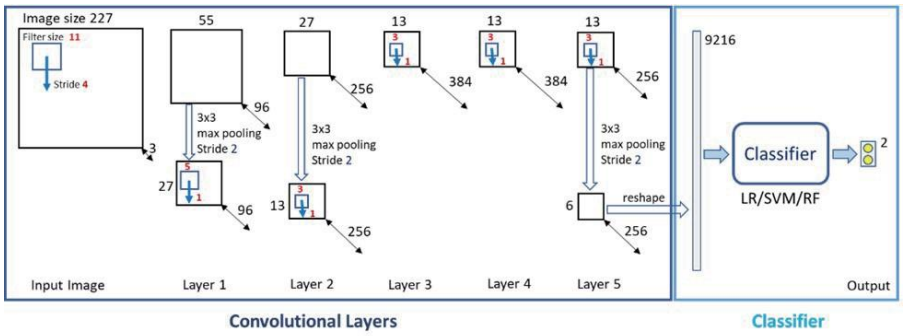


Fig. 5. (Color online) Fully connected layers in AlexNet replaced by other classifiers (LR/SVM/RF) for the control group.

Table 1
Comparison among four different classifiers.

Model	In-sample accuracy	Out-sample accuracy
AlexNet (9216)	0.999	0.978
CNN + LR (9216)	1	0.980
CNN + SVM (9216)	1	0.974
CNN + RF (9216)	1	0.966

Results and Discussion: The performances using different classifiers in training and testing are listed in the columns of in-sample accuracy and out-sample accuracy, respectively, in Table 1. The testing results show tiny differences in accuracy rates among the classifiers. This means that the features extracted from the convolutional layers are not restricted to end-to-end neural networks. These features are meaningful in classification and can also be incorporated in other classifiers. From Table 1, it is noteworthy that overfitting seemed to occur on the model trained with RF, whereas the highest out-sample accuracy rate was achieved for the model trained with LR. This may imply that a simpler model can lead to better performance on the out-sample dataset due to its better generalization property.

3.2 Exp #2: visualization of deep features

Motivation: In the previous experiment, the deep learning model demonstrated its capability to distinguish patches with and without lesions, and the deep features learned by the DCNN models are meaningful in classification. In this experiment, the patterns that contribute to the decision making of the classifier are revealed to clarify the mechanism of deep learning models from a pathological perspective.

Hypothesis: Most deep features learned by the DCNN models agree with the pathological rules used in classification.

Model: The fine-tuned AlexNet model from Exp #1 was used for visualization, and forward propagation was performed through the convolutional layers for the input patch to derive its corresponding activation map in each channel, as shown in Fig. 6.

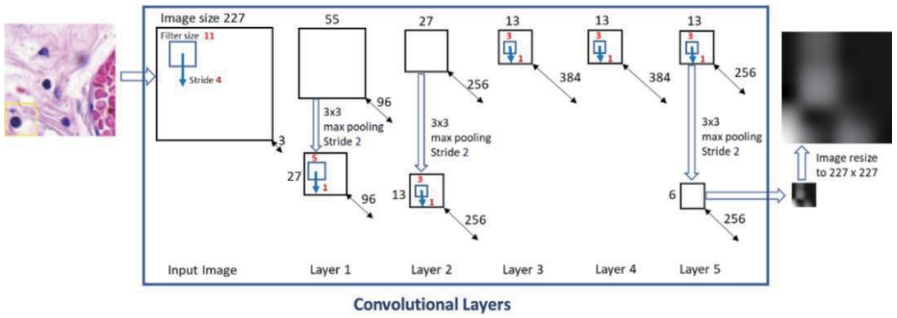


Fig. 6. (Color online) Activation map generated from the results of forward propagation through the convolutional layers in one of the channels.

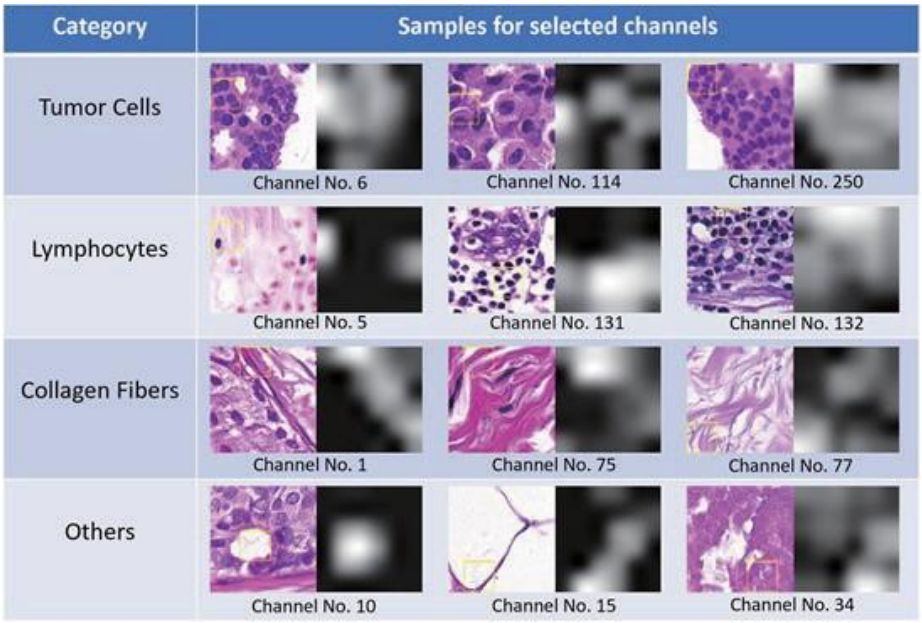


Fig. 7. (Color online) Activation maps reflecting the high-response regions. Most of the learned filters in the DCNN can work as morphological descriptors to detect specific cells and tissues.

Results and Discussion: The sampling patches and the corresponding activation maps for the selected channels are presented and classified by the pathological categories in Fig. 7. From the observations, most of the learned filters in the DCNN work as morphological descriptors to detect specific cells and tissues. Also, the activation maps reflect the spatial distribution of the patterns learned from the input patches. Interestingly, in this experiment, only the regions with lesions were manually labeled by the pathologists; however, we found that the deep learning models are able to analyze the main components in the patches and categorize them by their characteristics. That is, in lesion detection, the deep learning models not only detect the distribution of tumor cells but also recognize lymphocytes, collagen fibers, and some other non-cell structural tissues such as luminal spaces, necrosis, and secretions. The results show that the deep features learned by the DCNN model agree with the clinical insights in pathology, and our hypothesis holds.

3.3 Exp #3: feature reduction

Motivation: In image classification tasks on natural images, the spatial arrangement of object parts is an essential characteristic for the deep learning models in object recognition. For example, eyes are expected to be detected above a nose or mouth for a human face in an image. However, for pathological images, since patches were sampled at high magnification (×40), cells and tissues were arbitrarily distributed in the small sampling patches, as shown in Fig. 8. The information of spatial positions among objects becomes meaningless and irrelevant in the task of patch classification here.

Hypothesis: Characteristics of the spatial orientations of objects can be ignored within the small patches of view at a high magnification, and feature reduction can be applied to speed up the system.

Model: The previous experiment showed that the deep learning models can recognize tumor cells, lymphocytes, and collagen fibers. Most of the learned DCNN features can be regarded as detectors for these categories. Since we assume that the information of spatial orientations for these elements can be ignored within the small sampling patches, the tasks of patch classification can be accomplished by checking whether any lesion exists in the patch without knowing its exact orientation. Accordingly, a 13 × 13 average pooling layer was adopted to replace the original 6 × 6 max pooling layer in AlexNet-Layer 5. The modified model is shown in Fig. 9. As a result, the total number of features in classification is reduced from 6 × 6 × 256 (9216) to 1 × 1 × 256 (256). The size of the features was 1/36 that in the original structure.

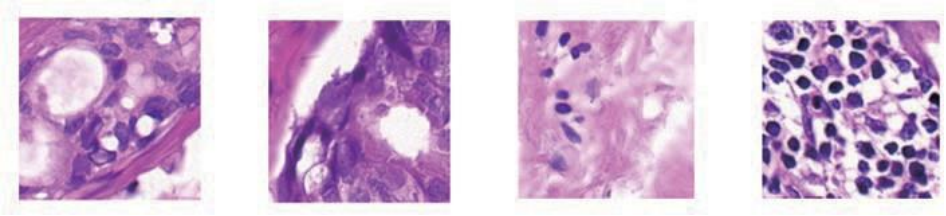


Fig. 8. (Color online) Cells and tissues arbitrarily distributed in the sampling patches.

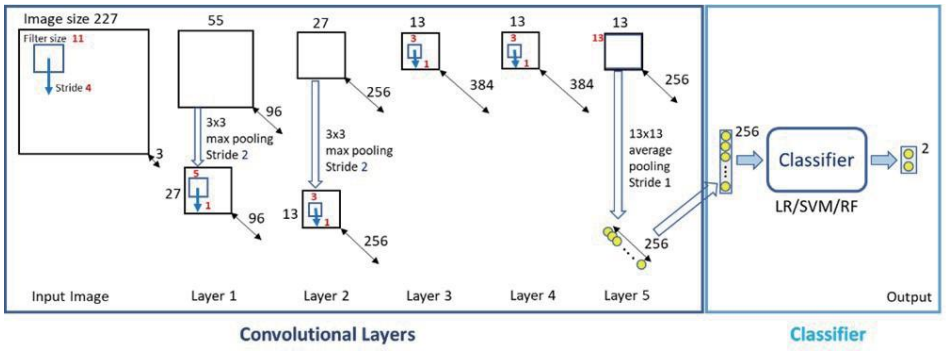


Fig. 9. (Color online) Modified model using 13 × 13 average pooling layer to discard spatial information.

Table 2
Comparisons among four different classifiers.

Model	In-sample accuracy	Out-sample accuracy
AlexNet (9216)	0.999	0.978
CNN + LR (9216)	1	0.980
CNN + LR (256)	0.985	0.979
CNN + SVM (9216)	1	0.974
CNN + SVM (256)	0.990	0.976
CNN + RF (9216)	1	0.966
CNN + RF (256)	1	0.978

Results and Discussion: For comparison, the performances before and after feature reduction are listed in Table 2. Despite having a feature size 36 times smaller than the original one, the out-sample accuracy remains at the same level or even slightly better. That means that the characteristic of the spatial orientations of objects is redundant and can be discarded within the small-size sampling patches, which proves the hypothesis. Since the model's complexity drops after feature reduction, the results suggest that constraining the complexity of the model can give the model better generalization property to prevent it from overfitting and achieve better out-sample accuracy. Moreover, after reducing the number of features from 9216 to 256, the system of lesion detection became 23% faster in execution. The performance was improved in terms of both efficacy and efficiency using the modified model.

3.4 Exp #4: feature selection

Motivation: After feature reduction, the same method of visualization as in Exp #2 was used to observe the patterns learned from the modified model in Exp #3. The visualization results are summarized in Fig. 10. The original activation maps from the modified model were of size 13 × 13 before resizing, and the corresponding size of the FOV for each unit was about the same as a cancerous nucleus in the patch. Therefore, the high-response regions in the activation maps very

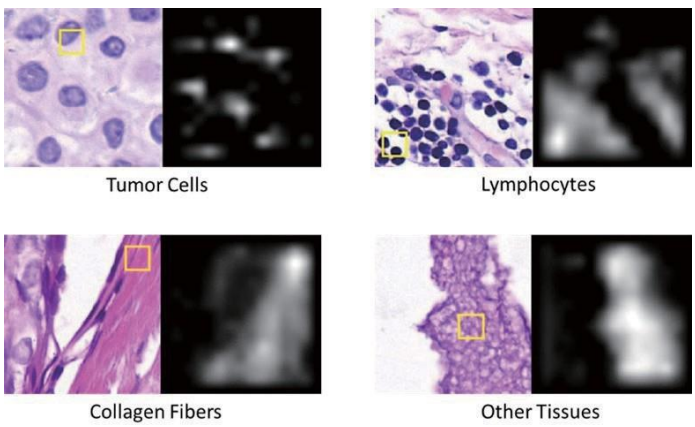


Fig. 10. (Color online) Visualization of the deep features using the modified model in Exp #3. The high-response regions reflect the distribution of specific cells or tissues.

closely reflect the distribution of tumor cells (more precisely than the results obtained using the original AlexNet model in Exp #2). We also found that the deep learning models can reveal the co-occurrence properties of patterns by exploring the data. Figure 11 shows that the deep learning models not only focused on the characteristics of cancerous nuclei but also noticed the effect of cytoplasmic clearing around those cancerous nuclei. In this experiment, the purpose of feature selection was to better understand how the trained model utilizes these 256 deep features from Exp #3.

Method: All 256 features from Exp #3 were partitioned into two groups. One group was used to collect the features that can convey clinical insights, which means that the features can work as detectors for specific cells or tissues, similarly to the features collected in Figs. 7 and 10, referred to as "recognizable features" here. On the other hand, the rest of the features that cannot be assigned to a specific category in pathology belonged to the other group of "unrecognizable features." Figure 12 shows examples of unrecognizable features. From observations, 151 out of the 256 features were categorized into the group of recognizable features, 43 of which were cell-structure features, such as tumor cells or lymphocytes. These 43 cell-structure features were

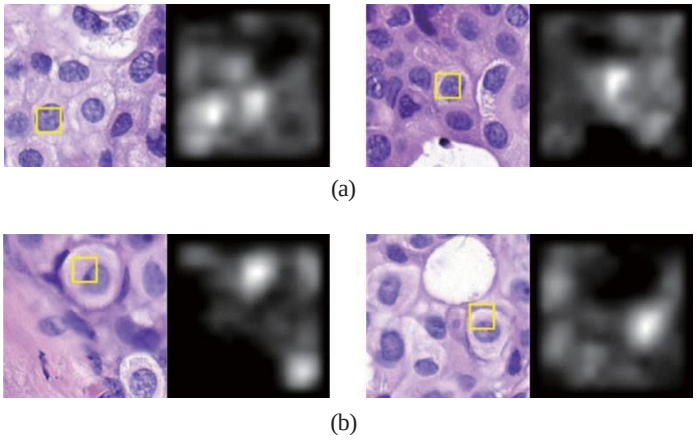


Fig. 11. (Color online) Co-occurrence properties of patterns learned from the training dataset. (a) Learned filter targets on cancerous nuclei. (b) Units with high activations on the regions of cytoplasmic clearing around cancerous nuclei.

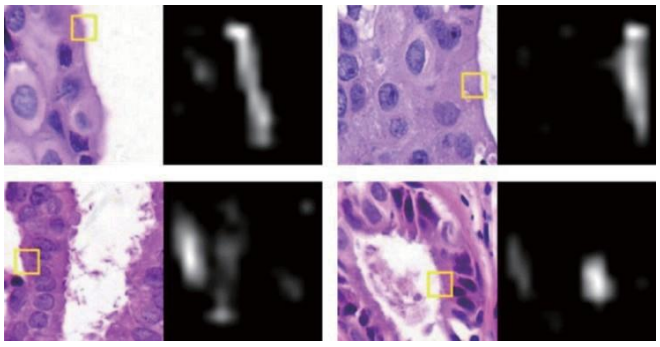


Fig. 12. (Color online) Examples of unrecognizable features.

collected manually in this experiment to further reduce the feature size (from 256 to 43). Another 43 features were randomly selected from the group of unrecognizable features as the control group for comparison.

Hypothesis: In manual lesion inspection, pathologists usually focus on different types of cells and then determine whether the cells are cancerous from their morphological properties. Similarly, we argue that if we further reduce the feature size by only selecting the cell-structure features, lesion detection should also be achieved. The model trained with the cell-structure features was expected to outperform the model trained with the unrecognizable features under the same feature size since the cell-structure features are more useful and important from a pathological perspective.

Results and Discussion: In this experiment, the RF classifier was used to perform consistent comparisons of performance among all scenarios starting from the first experiment. Table 3 lists the results of comparisons with the original model after feature reduction and after feature selection. In Table 3, the set of 43 cell-structure features from the group of recognized features is denoted as (43), and the other set of 43 features randomly selected from the group of unrecognized features is denoted as (<u>43</u>). After feature selection, the results show that the performance decreased for both models compared with the model trained with all 256 features. Also, the model trained with the selected 43 cell-structure features outperformed the model trained with the <u>43</u> unrecognizable features. Surprisingly, the model trained with the <u>43</u> features randomly selected from the group of unrecognizable features still maintained an out-sample accuracy of over 94%. This implies that the features not intuitive to specialists may still be useful for machines and statistically discriminative in classification. Accordingly, the top 43 important features ranked by the importance property from RF out of all 256 features were collected, and the feature set is denoted as (*43) in Table 3. The model trained with the top 43 important features outperformed the model trained with the 43 cell-structure features. In the feature set of (*43), 33 features belonged to the group of recognizable features, among which 14 features were related to the cell structure and 19 features were related to collagen fibers or other tissues. The remaining 10 features were from the group of unrecognizable features. Figure 12 shows examples of unrecognizable features that were discriminative in patch classification. The activation maps in Fig. 12 show that the learned filter drives a high response to the cytoplasmic parts of the tumor cells near interstitial spaces. These discriminative but unrecognizable features extracted by the deep learning models merit further study to find reasonable correlations with pathological knowledge and may facilitate the research of new characteristics in diagnosis.

Table 3
Performance before and after feature selection.

Model	In-sample accuracy	Out-sample accuracy
AlexNet (9216)	0.999	0.978
CNN + RF (9216)	1	0.966
CNN + RF (256)	1	0.978
CNN + RF (43)	1	0.961
CNN + RF (<u>43</u>)	1	0.947
CNN + RF (*43)	1	0.974

4. Conclusion

In this study, four experiments were conducted to study the properties of the deep features learned by DCNN models. In the first experiment, we verified that these DCNN features are transferable and meaningful in the classification for histology images. By visualizing the deep features in the second experiment, we found that most of the learned filters in the DCNN can work as morphological descriptors to detect specific cells and tissues, in accordance with the categories in pathology. The results revealed the insights of the deep learning model in lesion detection to demonstrate its validity. The learned filters can also be exploited for quantitative assessment in the assessment tasks of tumor-infiltrating lymphocytes and tumor-stroma ratio. In the third experiment, we modified the model on the basis of prior knowledge to obtain better efficacy and efficiency. Furthermore, we ranked all features by importance to compare the viewpoints of humans and machines in the fourth experiment. We found that more than half of the extracted features were interpretable by the domain knowledge of pathology, although the other unrecognizable features also seemed discriminative in the classification. The deep learning frameworks are useful for summarizing rules in classification. These rules learned from big data should be further studied to facilitate the development of AI technology and research in the medical field.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 81972485).

References

1 S.-Y. Jhong, P.-Y. Yang, and C.-H. Hsia: Sens. Mater. **34** (2022) 1259.
2 C.-C. Lee, S.-T. Tsai, and C.-H. Yang: Sens. Mater. **30** (2018) 1859. https://doi.org/10.18494/SAM.2018.1899
3 C.-J. Lin, S.-Y. Jeng, and C.-L. Lee: Sens. Mater. **33** (2021) 315. https://doi.org/10.18494/SAM.2021.3015
4 A. Janowczyk and A. Madabhushi: J. Pathol. Inf. **7** (2016) 29. https://doi.org/10.4103/2153-3539.186902
5 D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck: arXiv print (2016). https://arxiv.org/abs/1606.05718
6 B. E. Bejnordi, M. Balkenhol, G. Litjens, R. Holland, P. Bult, N. Karssemeijer, and J. Laak: IEEE Trans. Med. Imaging **35** (2016) 2141. https://doi.org/10.1109/TMI.2016.2550620
7 A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi: Proc. SPIE Int. Soc. Opt. Eng. (2014) 9041. https://doi.org/10.1117/12.2043872
8 A. Krizhevsky, I. Sutskever, and G. E. Hinton: Int. Conf. Neural Information Processing Systems **1** (2012). https://doi.org/10.1145/3065386
9 B. Korbar, A. M. Olofson, A. P. Miraflor, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, and S. Hassanpour: IEEE Conf. Computer Vision and Pattern Recognition Workshops (2017) 821–827. https://doi.org/10.1109/CVPRW.2017.114
10 M. D. Zeiler and R. Fergus: European Conf. Computer Vision (2014). https://doi.org/10.1007/978-3-319-10590-1_53
11 Q.-S. Zhang and S.-C. Zhu: Front. Inf. Technol. Electron. Eng. **19** (2018) 27. https://doi.org/10.1631/FITEE.1700808
12 K. Simonyan, A. Vedaldi, and A. Zisserman: arXiv print (2013). https://arxiv.org/abs/1312.6034
13 A. Mahendran and A. Vedaldi: IEEE Conf. Computer Vision and Pattern Recognition (2015) 5188–5196. https://doi.org/10.1109/CVPR.2015.7299155
14 B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba: IEEE Conf. Computer Vision and Pattern Recognition (2016) 2921–2929. https://doi.org/10.1109/CVPR.2016.319

15 R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra: Int. J. Comput. Vis. **128** (2016) 336. https://doi.org/10.1007/s11263-019-01228-7
16 Q. Zhang, R. Cao, F. Shi, Y.-N. Wu, and S.-C. Zhu: arXiv print (2017). https://arxiv.org/abs/1708.01785
17 Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe: arXiv print (2017). https://arxiv.org/abs/1703.02442
18 J. Yosinski, J. Clune, Y. Bengio, and H. Lipson: Int. Conf. Neural Information Processing Systems (2014) 3320–3328. https://arxiv.org/abs/1411.1792
19 F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte: IEEE Int. Conf. Systems, Man and Cybernetics (2017) 1868–1873. https://doi.org/10.1109/SMC.2017.8122889
20 Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, and E. Chang: BMC Bioinf. **18** (2017) 281. https://doi.org/10.1186/s12859-017-1685-x