

# Encoding histopathological WSIs using GNN for scalable diagnostically relevant regions retrieval

Yushan Zheng<sup>1,2</sup>, Bonan Jiang<sup>3</sup>, Jun Shi<sup>4</sup>, Haopeng Zhang<sup>2,1</sup>, and Fengying Xie<sup>2,1</sup>

<sup>1</sup> Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China {yszheng@buaa.edu.cn}

<sup>2</sup> Image Processing Center, SA, Beihang University, Beijing 100191, China

<sup>3</sup> Beijing-Doblin International College, Beijing University of Technology, Beijing 100124, China

<sup>4</sup> School of Software, Hefei University of Technology, Hefei 230601, China

**Abstract.** The research on content-based histopathological image retrieval (CBHIR) has become popular in recent years. CBHIR systems provide auxiliary diagnosis information for pathologists by searching for and returning regions that are contently similar to the region of interest (ROI) from a pre-established database. To retrieve diagnostically relevant regions from a database that consists of histopathological whole slide images (WSIs) for query ROIs is challenging and yet significant for clinical applications. In this paper, we propose a novel CBHIR framework for regions retrieval from WSI-database based on hierarchical graph neural networks (GNNs). Compared to the present CBHIR framework, the structural information of WSI is preserved by the proposed model, which makes the retrieval framework more sensitive to regions that are similar in tissue distribution. Moreover, benefited from the hierarchical GNN structures, the proposed framework is scalable for both the size and shape variation of ROIs. It allows the pathologist defining the query region using free curves. Thirdly, the retrieval is achieved by binary codes and hashing methods, which makes it very efficient and thereby adequate for practical large-scale WSI-database. The proposed method is validated on a lung cancer dataset and compared to the state-of-the-art methods. The proposed method achieved precisions above 82.4% in the irregular region retrieval task, which are superior to the state-of-the-art methods. The average time of retrieval is 0.514 ms.

**Keywords:** Digital pathology · Histopathological image analysis · GNN · CBIR · ACDC-LungHP.

## 1 Introduction

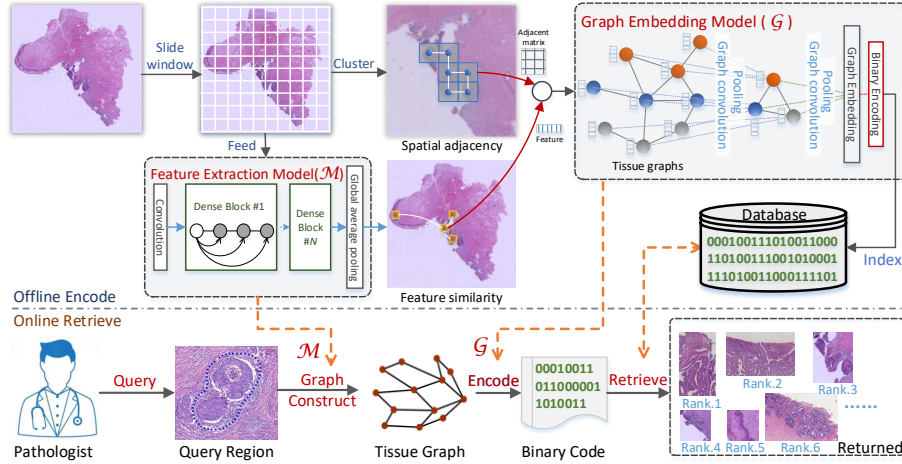
With the development of whole slide imaging techniques for digital pathology, the computer aided cancer diagnosis methods based on histopathological image analysis (HIA) have been widely studied. Content-based histopathological image retrieval (CBHIR) is an emerging approach in this domain [1]. CBHIR searches

a pre-established WSI database for the regions the pathologist concerned and provides contently similar regions to the pathologists for reference. Compared to the typical HIA methods based on image segmentation and classification [2], CBHIR methods can provide more valuable information including the diagnostically similar regions, the meta-information, and the diagnosis reports of experts stored along with the cases in the digital pathology system.

The present studies for CBHIR are generally on databases consist of image blocks or patches in a fixed size. However, the practical histopathological databases generally consist of digital whole slide images and the query regions created by the pathologists are in different sizes and shapes. It is challenging to efficiently retrieve regions from large-scale database containing WSIs in very high pixel-resolutions and accurately return the similar cases the pathologists needed. To meet the efficiency requirement for large-scale histopathological database retrieval, the binary encoding and hashing techniques have been successfully introduced to accelerate the process of retrieval [1, 3]. To index WSIs for region-level retrieval, the WSIs are commonly divided into small patches following the sliding window paradigm. However, the diagnosis of cancer with tissue sections not only depends on the local nuclei features but also the contextual information from a broad region surrounding the nuclei. Several retrieval strategies have been developed to improve the scalability of the retrieval framework to size variation of query regions [4–6]. These methods mainly applied feature vector quantification approach, e.g. pooling operations, to embed the features of local patches, which is a convenient to generate uniform representation for irregular tissue regions. However, the adjacent relationship of different type of biopsy objects are lost during the feature quantification, for which the structure similarity between tumor regions are difficult to recognize in the procedure of retrieval.

In this paper, we propose a novel framework for histopathological image retrieval from large-scale WSI-database based on graph neural network (GNN) [8] and hashing method. The proposed framework for CBHIR is illustrated in Fig. 1. The WSIs are first divided into patches and converted into image features using a pre-trained convolutional neural networks (CNN). Then, graphs of tissue are established based on spatial relationships and feature distances of patches. Finally, the tissue-graphs are fed into the designed GNN-Hash model to obtain the indexes for retrieval. The retrieval strategy proposed in this paper is scalable for both the size and shape of query ROIs, which makes the proposed method more applicable than the present methods.

The novelty and contribution of this paper to the problem is two-fold. Firstly, we designed a novel graph-based WSI encoding approach based on GNN. Besides the local features of tissue regions, the distribution of tissue is considered and preserved in the process of WSI encoding. The accuracy and scalability of the CBHIR framework has been improved. Secondly, we combined the GNN structure with binary encoding and designed a GNN-Hash model. The GNN-Hash model is trained end-to-end from graphs with variable number of nodes to the binary-like codes, which makes the retrieval framework both structural-



**Fig. 1.** The proposed CBHIR framework. In the offline stage, tissue graphs are first constructed based on the spatial adjacency and feature similarity of patches. Then, the graphs are embedded into binary codes and stored to the retrieval database. When retrieving, the region the pathologist queries are converted into a binary code and then the diagnostically relevant regions are retrieved and returned to pathologists.

preserving and time-saving. It determines that the proposed method is applicable for practical large-scale WSI database.

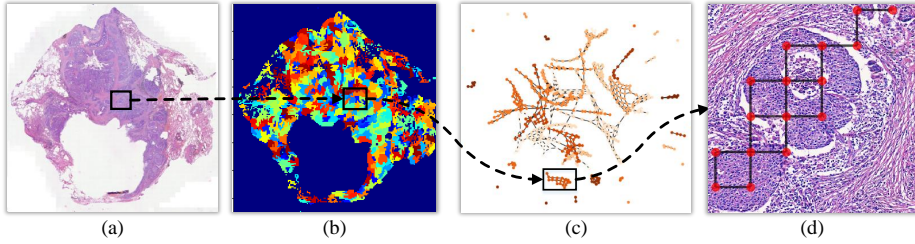
The remainder of this paper is organized as follows. Section 2 introduces the methodology of the proposed method. The experiment is presented in Section 3. Section 4 summarizes the contributions.

## 2 Methods

Motivated by the development of graph information analysis (e.g. protein structure recognition and social network analysis), we propose to establish graph structures to depict the adjacent relationship of local regions in WSIs. The details of the proposed methods are introduced in this section.

### 2.1 Tissue graph construction

A graph  $G$  can be represented as  $(\mathbf{A}, \mathbf{X})$ .  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is an adjacent matrix that defines the connectivity of the  $n$  nodes in  $G$ , where  $a_{ij} = 1$ ,  $a_{ij} \in \mathbf{A}$  represents the  $i$ -th and the  $j$ -th node in the graph are connected, and  $a_{ij} = 0$ , otherwise.  $\mathbf{X} \in \mathbb{R}^{n \times d}$  denotes the node feature matrix assuming each node is represented as a  $d$ -dimensional column vector. In this paper, the sub-regions in the WSI are described by graphs. The flowchart to construct graphs for a WSI is presented in Fig. 2. First, the WSI is divided into non-overlapping patches using a sliding window. Then, the patches are fed into a pre-trained convolutional neural network (CNN) to extract patch features  $\mathbf{X}$ . Considering that the tumor regions



**Fig. 2.** The flowchart of tissue graphs, where (a) is a digital WSI, (b) illustrates the sub-regions, (c) shows the graphs established on sub-regions, and (d) jointly presents a graph and its corresponding regions where the nodes are drawn on the centroids of patches.

in a WSI varies in shape and size, we propose mosaicking the adjacent patches into irregular sub-regions. Specifically, the agglomerative hierarchical clustering algorithm [7] is employed to merge the patches based on the similarities between patch features. To ensure the sub-regions are spatially connected, only the 4-connected patches are allowed to be mosaicked. After the clustering, a set of graphs  $\{G_i\}$  can be established (Fig. 2 (c)). The set  $\{G_i\}$  can cover the entire content of the WSI and thereby can be used to index the WSI for retrieval.

## 2.2 GNN-Hash for graph encoding

To establish the retrieval indexes, the graphs are needed to be encoded into vectors in equal dimensions. It is challenging to simultaneously embed the node attributes and edge information into an uniform representation. Graph neural network (GNN) [8] is an emerging techniques for graph information embedding, which has been proven effective in histopathological image analysis [9]. Generally, GNNs can be represented following message-passing architecture  $\mathbf{H}^{(k)} = M(\mathbf{A}, \mathbf{H}^{(k-1)}; \theta^{(k)})$ , where  $H^{(k)} \in \mathbb{R}^{n \times d}$  denotes the embeddings on the  $k$ -th step of passing,  $M$  is the message propagation function [10–12] that depends on the adjacent matrix  $\mathbf{A}$ , the output of the previous step  $\mathbf{H}^{(k-1)}$ , and the set of trainable parameters  $\theta^{(k)}$ .  $\mathbf{H}^{(0)}$  is the original attributes of the nodes, i.e., the CNN features of patches  $\mathbf{X}$  in our method. For simplicity, a GNN module with multiple steps of embeddings can be represented as  $\mathbf{Z} = GNN(\mathbf{A}, \mathbf{X})$ . Multiple GNNs can be stacked to learn deep representations of graphs. Recently, Ying et al. [13] proposed a differentiable graph pooling module (DiffPool), which enables the hierarchical GNNs to be trained in end-to-end fashion. Specifically, an additional GNN with a softmax output layer is designed to learn a matrix  $\mathbf{S} \in \mathbb{R}^{n_l \times n_{l+1}}$ , which is used to assign the output representations of the  $l$ -th GNN to  $n_{l+1}$  clusters. Then, the input  $\mathbf{X}^{(l+1)}$  and the adjacent matrix  $\mathbf{A}^{(l+1)}$  for the next GNN are obtained by equations:

$$\begin{aligned} \mathbf{X}^{(l+1)} &= \mathbf{S}^{(l)\top} \mathbf{Z}^{(l)} \in \mathbb{R}^{n_{l+1} \times d} \\ \mathbf{A}^{(l+1)} &= \mathbf{S}^{(l)\top} \mathbf{A}^{(l)} \mathbf{S}^{(l)} \in \mathbb{R}^{n_{l+1} \times n_{l+1}} \end{aligned}$$

In our method, the network is used to learn representations that are effective for data retrieval. To ensure the framework is applicable to practical large-scale pathological database, we modified the output of the hierarchical GNNs to generate a GNN-Hash structure. Specifically, a binary encoding layer is defined based on the last graph embeddings  $\mathbf{Z}^{(L)} \in \mathbb{R}^{N \times d}$ :  $\mathbf{Y} = \tanh(\mathbf{Z}^{(L)}\mathbf{W} + \mathbf{b})$ , where  $\mathbf{W}$  and  $\mathbf{b}$  are the weights and bias for a linear projection.  $\mathbf{Y} \in (-1, 1)^{N \times d_h}$  is the network outputs that can be simply converted into binary codes by equation  $\mathbf{B} = \text{sign}(\mathbf{Y}) \in \{-1, 1\}^{N \times d_h}$ , where  $N$  is the number of graphs and  $d_h$  is the dimension of binary codes. The loss function for training the GNN-Hash is defined as

$$J = \frac{1}{N} \left\| \frac{1}{d_h} \mathbf{Y} \mathbf{Y}^T - \mathbf{C} \right\|_F^2 + \lambda \left\| \mathbf{W}^T \mathbf{W} - \mathbf{I} \right\|_F^2$$

where  $\mathbf{C} \in \{-1, 1\}^{N \times N}$  is the pair-wise label matrix in which  $c_{ij} = 1$  represents the  $i$ -th and  $j$ -th tissue graphs are relevant and  $c_{ij} = -1$  otherwise, and  $\lambda$  is a weight coefficient. Finally, the proposed GNN-Hash structure is trained end-to-end from the input graph  $G$  with CNN-features to the output  $\mathbf{Y}$ .

### 2.3 Retrieval using binary codes

For each WSI in the retrieval database, a set of binary codes ( $\mathbf{B}$ ) that represent the graphs in the WSI can be obtained using the trained model. When retrieving, the region the pathologist queries is divided into patches and converted into binary code using the same model. Then, the similarities between the query code and those in the database are measure using Hamming distance. After ranking the similarities, the top-ranked regions are retrieved and finally returned to the pathologist.

## 3 Experiments

### 3.1 Experimental setting

The experiments were conducted on the ACDC-LungHP dataset<sup>5</sup> [14], which contains 150 WSIs within lung cancer regions annotated by pathologists. In the evaluation, 30 WSIs were randomly selected as the testing dataset and the remainders were used to train the retrieval models and establish the retrieval database. To evaluate the scalability to size and shape variations of query ROIs, the WSIs were divided into irregular sub-regions under  $40\times$  lenses using the clustering approach introduced in section 2.1. Consequently, 2894 query regions are generated and 16261 regions are created to establish the retrieval database. The number of patches in each region ranges from 1 to 136 with a median number of 37.

<sup>5</sup> The dataset is accessible at <https://acdc-lunghp.grand-challenge.org/>. Since the annotations of testing part of the data set are not yet published, only the 150 training WSIs of the data were used in this paper.

**Table 1.** Comparison of retrieval performance with the state-of-the-art methods.

Methods	P@50	P@200	MAP@50	MAP@200	Complexity	Time
Shi et al. [3] (2018)	0.726*	0.723*	0.747*	0.733*	$\mathcal{O}(n)$	180 ms
Zheng et al. [16] (2018)	0.759	0.745	0.781	0.761	$\mathcal{O}(m \log n)$	86.1 ms
Ma et al. [4] (2018)	0.735	0.726	0.756	0.736	$\mathcal{O}(n)$	0.507 ms
Jimenez et al. [6] (2017)	0.750	0.745	0.770	0.753	$\mathcal{O}(mn)$	4.97 s
GNN-LR**	0.797	0.794	0.812	0.802	$\mathcal{O}(n)$	1.20 ms
GNN-Hash	<b>0.826</b>	<b>0.824</b>	<b>0.838</b>	<b>0.833</b>	$\mathcal{O}(n)$	0.514 ms

\* This is the result for fixed-size-patch ( $224 \times 224$  pixels) retrieval task. Because there are 489,330 patches in the retrieval database, the average running time of this method is much longer than other methods with  $\mathcal{O}(n)$  complexity.

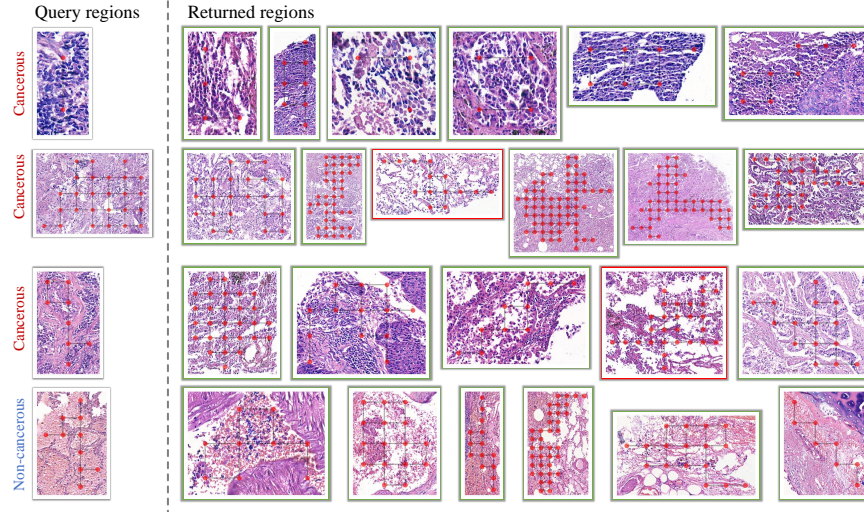
\*\* This is the retrieval result based on the last representation (LR) of the designed hierarchical GNNs structure but trained using the classification loss proposed in [13].

The DenseNet-121 structure [15] pre-trained for classification task within the training WSIs was used as the feature extractor of patches. Two GNN modules within three embedding steps and two DiffPool modules were stacked to generate the GNN-Hash structure. The dimension of node embeddings was tuned from 60 to 120 with a step of 20 on the training set and determined as 100. The bit number of the output binary code was set  $d_h = 32$ . The proposed method was implemented in python with pytorch. All the experiments were conducted on a computer with an Intel Core i7-7700k CPU of 4.2 GHz and a GPU of Nvidia GTX 1080Ti.

### 3.2 Results and discussion

The proposed method is compared with 4 state-of-the-art-methods [3, 4, 6, 16] proposed for histological image retrieval. For fair comparison, the backbones of the compared methods are the same DenseNet-121 structure used in our model. In the evaluation, the graphs that contain more than 10% cancerous pixels referring to the pathologists' annotation are defined as *cancerous graph*, and the remainders are defined as *non-cancerous graph*. The retrieval precision (P@R) and the mean average precision (MAP@R) of the top R returned graphs are used as the evaluation metrics. The experimental results are summarized in Table 1.

Overall, the proposed method achieved the best retrieval performance. The query regions and the database regions in [3] are limited to square patches in a fixed size. Hence, we regarded the patches in graphs as the retrieval instances and reported the patch-level retrieval performance for this method. Besides, the other methods are scalable to the size variation of query regions. Ma et al. [4] proposed quantifying the patch features in a region using max-pooling operation and calculating the similarities between regions based on the pooled representations. The percentage of patches of different tissue types is ignored by the pooling operation, which has reduced the retrieval accuracy. Zheng et al. [16] and Jimenez et al. [6] proposed measuring the similarity of two regions by an ensemble of feature distances of all the patch pairs across the two regions. The local similarities between the regions are considered in the two methods, which



**Fig. 3.** Visualization of the retrieval performance of the proposed method, where the first column provides 4 query regions, the top-returned regions are ranked on the right, the relevant return regions (has the same label with the query region) are framed in green and the irrelevant regions are in red. The locations of the graph are meanwhile displayed. A high-quality copy of this picture is submitted as supplemental material.

has improved the MAP@200 to 0.761 and 0.753, respectively. The main drawback of the two methods is that the adjacent relationship of tissue objects cannot be effectively described, which has severely affected the performance of retrieval. In contrast, the proposed method constructed graphs within tissue regions. The information of tissue allocation has been sufficiently described and well preserved via the hierarchical embeddings of the proposed GNN-Hash model. It contributes to a significant improvement of MAP@200 from 0.761 to 0.833, compared to that obtained by Zheng et al. [16]. The retrieval results obtained by the proposed framework are visualized in Fig. 3. It presents that the proposed model can adapt to the variation of size and shape of query regions and the diagnostically relevant regions to the query regions are successfully retrieved.

The efficiency is also important for CBHIR system. The computational complexity (by  $\mathcal{O}$  notation) relevant to the pixel size of query region  $m$  and the scale of the database  $n$  for the compared methods are given in Table 1. Correspondingly, the average time consumption for retrieval are also compared (the feature extraction time is not involved). In our method, the computation for retrieval is irrelevant to the size of query region after the encoding and thereby the complexity is  $\mathcal{O}(n)$ . Moreover, benefiting from the binary encoding, the similarity measurement is time-saving than those (e.g. GNN-LR) based on float-type high-dimensional features. When the order of magnitudes of WSI in database increases and the content in the database is abundant, a hashing table will be

pre-established. Then, the retrieval can be easily achieved by a table-lookup operation, for which the complexity of retrieval can be further reduced to  $\mathcal{O}(1)$ .

## 4 Conclusion

In this paper, we proposed a novel histopathological image retrieval framework for large-scale database consisting of WSIs. The instances in the database are defined based on graphs and are converted into binary codes by the designed GNN-Hash model. The experimental results have demonstrated that the proposed model achieves the state-of-the-art retrieval performance and is scalable to size and shape variations of query regions and can effectively retrieve relevant regions that contain similar content and structure of tissue. It allows pathologists to create query regions by free-curves on the digital pathology platform. Benefiting from hashing structure, the retrieval process is completed based on hamming distance, which is very time-saving. Overall, the proposed method is effective, efficient and practical for large-scale WSI database retrieval.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61771031, 61371134, 61471016, and 61501009), China Postdoctoral Science Foundation (No. 2019M650446) and Motic-BUAA Image Technology Research Center.

## References

1. Li Z, Zhang X, Mller H, et al.: Large-scale retrieval for medical image analytics: A comprehensive review. *Medical image analysis* **43**, 66–84 (2018)
2. Bejnordi B. E., Veta M., Van Diest P. J., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **318**(22), 2199–2210 (2017)
3. Shi X., Sapkota M., Xing F., et al.: Pairwise based deep ranking hashing for histopathology image classification and retrieval. *Pattern Recognition* **81**, 14–22 (2018)
4. Ma Y., Jiang Z., Zhang H., et al.: Generating Region Proposals for Histopathological Whole Slide Image Retrieval. *Computer Methods and Programs in Biomedicine* **159**, 1–10 (2018)
5. Zheng Y., Jiang Z., Zhang H., et al.: Histopathological whole slide image analysis using context-based CBIR. *IEEE Transactions on Medical Imaging* **37**(7), 1641–1652 (2018)
6. Jimenez-del-Toro O., Otálora S., Atzori M., et al.: Deep multimodal casebased retrieval for large histopathology datasets. *International Workshop on Patch-based Techniques in Medical Imaging*, pp. 149–157 (2017)
7. Day W H E, Edelsbrunner H.: Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification* **1**(1): 7–24 (1984)



8. Wu Z., Pan S., Chen F., et al.: A comprehensive survey on graph neural networks. arXiv:1901.00596 (2019)
9. Li R., Yao J., Zhu X., et al.: Graph CNN for Survival Analysis on Whole Slide Pathological Images. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 174–182. Springer, Heidelberg (2018)
10. Gilmer J., Schoenholz S. S., Riley P. F., et al.: Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning (ICML), pp. 1263–1272 (2017)
11. Hamilton W., Ying Z., Leskovec J.: Inductive representation learning on large graphs. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), pp. 1024–1034 (2017)
12. Kipf T. N., Welling M.: Semi-supervised classification with graph convolutional networks. In: Proceedings of International Conference on Learning Representations (ICLR), 2017
13. Ying Z., You J., Morris C., et al. Hierarchical graph representation learning with differentiable pooling. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), pp. 4805–4815 (2018)
14. Li Z., Hu Z., et al.: Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study, arXiv:1803.05471v1 (2018)
15. Huang G., Liu Z., Van Der Maaten L., et al.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp. 4700–4708 (2017)
16. Zheng Y., Jiang Z., Zhang H., et al.: Size-scalable content-based histopathological image retrieval from database that consists of WSIs. *IEEE Journal of Biomedical and Health Informatics* **22**(4), 1278–1287 (2018)