Bag of Deep Features for Classification of Gigapixel Histological Images

Nadia Brancati
Institute for High Performance Computing and Networking,
National Research Council of Italy
Naples, Italy
nadia.brancati@icar.cnr.it

Crispino Cicala
Computer Scientist, Milan, Italy
cr.cicala@gmail.com

Daniel Riccio
University of Naples Federico II
Naples, Italy
daniel.riccio@unina.it

Maria Frucci
Institute for High Performance Computing and Networking,
National Research Council of Italy
Naples, Italy
maria.frucci@icar.cnr.it

Abstract. **Convolutional Neural Networks (CNNs) have proven to be one of the most powerful tools for solving complex problems in the field of pattern recognition and image analysis, even if serious challenges remain. Indeed, one of the main drawbacks of CNNs is their inability to cope with very high-resolution images. In areas other than digital pathology, image resizing is often the simplest and most effective solution. However, histopathological images not only show a very high resolution, but also contain a lot of information at the detail level, making this strategy completely ineffective. Other approaches partition the image into small patches and analyze them independently, losing the context information that is fundamental in digital pathology. In this paper, we present a method based on a compressed representation of the Whole Slide Image (WSI), by building a 3D tensor, that preserves the topological and morphological information relating to the proximity relationships between the patches of the WSI. Tensors are used to train a CNN to solve a binary classification task. This technique has been evaluated for the analysis of gigapixel Hematoxylin and Eosin (H&E) histological images with the aim of supporting the diagnosis of breast cancer. Several experiments have been performed on the Camelyon16 dataset by generating different types of 3D tensors. The results of the proposed approach on the breast cancer classification task have been compared with some state-of-the-art approaches.**

Keywords: **histological images, deep learning, clustering**

I. INTRODUCTION

In the field of Computer Aided Diagnosis (CAD), one of the main challenges concerns the analysis of WSIs obtained by scanning tumor tissues stained with H&E and commonly used for the diagnosis of tumor pathologies. Unfortunately, deep learning approaches cannot be applied directly to WSI because of their very high resolution. WSI are generally made up of trillions of pixels that cannot be managed by current deep learning systems. Over the years, different approaches have been proposed trying to meet computational needs but preserving the information needed to perform different tasks of analysis, including the classification of the disease. Most classification approaches are primarily based on partitioning the entire WSI image into patches small enough to be processed independently by a deep network. The class of the entire WSI is usually inferred by combining the decisions obtained for the individual patches [1–3]. Unfortunately, all these approaches neglect the information provided by relationships between patterns presented by individual patches, making the prediction of the CNN an isolated result. Recent methods [4–6] map the WSI into a new compressed and dense feature space by rearranging patch-wise feature vectors in a grid-based representation aiming to preserve spatial correlations of different patches. Although these methods save most of the discriminatory information and the grid representation can be used to train a CNN to classify the entire WSI, the contextual analysis of each point of the grid (i.e the feature vector of a patch) is limited to its 3x3 neighboring in the grid. Indeed, the analysis of relationships between patterns present in the single patches is performed through 3D convolutional operations.

This study proposes a solution for applying CNNs to histopathological images that works on the entire image, but preserving both detail and contextual information by widely extending the contextual analysis of each patch. A set of reference patches is mapped into a high dimensional deep features space, so that bags of deep features words are constructed using a clustering algorithm. A whole slide is partitioned into patches, which are projected into the same high dimensional deep features space and the co-occurrences of the deep features words are considered to build a 3D tensor that represent the entire image in a more compact way.

The experiments were conducted on the Camelyon16 dataset for the binary classification task of breast cancer. Comparisons with the state of the art confirm that the proposed method opens up to the future possibility of further extending this method aiming to further reduce the amount of data to be processed, while still obtaining good results for classification tasks.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

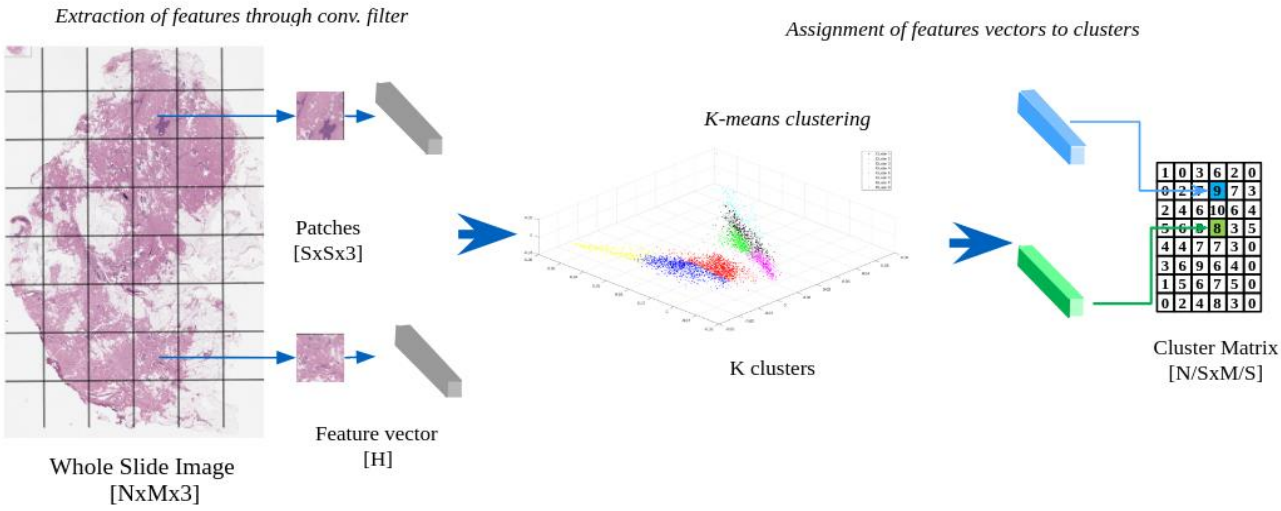


Fig. 1. Matrix cluster feature extraction. A WSI is divided into a set of patches and each of them is mapped to a feature vector using a pre-trained ResNet-18. To each feature vector is assigned a label cluster obtained computing its minimum distance from the feature vectors representing the centroids of the previously computed clusters. The set of label clusters is rearranged in a matrix according to the original spatial arrangement of the patches

II. METHOD

The system takes a WSI of any size as input and its pipeline is as follows. The image is partitioned into non-overlapping patches, which are projected into a high dimensional deep feature space by means of suitably fine-tuned CNN. Our strategy requires pre-processing to divide in clusters the feature vectors of the patches in which any WSI of the training set has been partitioned. The information about the clusters is used to map later each WSI to be analyzed in a dense feature space that is fed to the image-level classifier.

After the preprocessing phase, the overall framework is functionally divided into two main stages, namely Tensor-based Feature Extractor and Tensor-based Classifier.

A. Pre-processing

A pre-trained ResNet-18 [7] is fine-tuned for a binary classification, i.e. in order to distinguish between malignant and benign tissue. The fine-tuning is performed by considering patches extracted from a reference set of WSI that have been manually annotated by pathologists and are provided with the dataset adopted for the experiments. The trained network is then used to extract a feature vector of length H for each considered patch, that is then projected into a H-dimensional deep features space. The set of deep features vectors extracted from reference patches undergo a clustering process that is performed by K-means to form a set of K bags of deep features words.

Clusters might include irrelevant or redundant information, so that a post-process is applied aiming to balance data in each cluster and to remove data associated with patches including no tissue. Finally, for each cluster i, the corresponding centroid V is stored in the i-th row of a matrix MV of $K \times H$ size. The matrix MV will be used to perform the assigned patches extracted from a WSI to the corresponding cluster. In other words, MV allows assigning a patch to the corresponding bag of deep features words.

B. Tensor-based Feature Extractor and Classifier

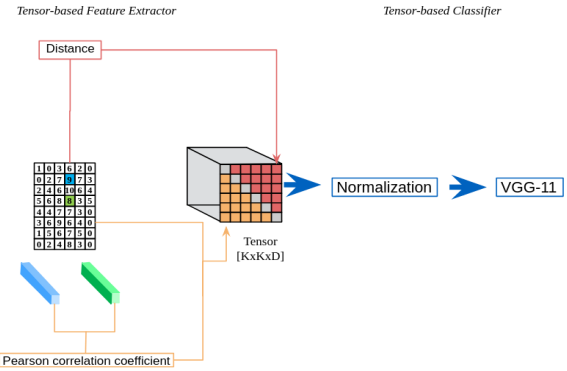
This step is devoted to the generation of a 3D tensor which stores information on the relationships between each couple of different patches lying at distance less or equal to D in the WSI input. The distance D is the value of the proximity radius determining the contextual area considered for each patch, i.e. D is the maximum distance between two different patches of a WSI for which the relationship between the corresponding features can be taken into account.

In the following, two different patches of a WSI at distance less or equal to D will be indicated as adjacent patches. Each analyzed patch has size $S \times S \times 3$.

Given an input WSI, namely W, with size $N \times M \times 3$, it is partitioned in non overlapping patches of size $S \times S \times 3$. Each patch $p_{i,j}$ is projected in the H-dimensional deep features space by computing its deep feature vector and assigned with the cluster $k \in K$ with the minimum Euclidean distance. Distances are computed between the patch feature vector and the cluster centroids. A cluster matrix CM of size $N/S \times M/S$ is constructed, where each (i, j) corresponds to a patch in W, and stores the index k of the cluster the patch has been assigned with. The index k can be also thought of as the cluster label of the patch $p_{i,j}$ (see Fig. 1).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The cluster matrix CM is then used to build a 3D tensor storing two different types of information about the relationship between feature vectors of adjacent patches of *W*. In more details, let *T* be a tensor, with size $K \times K \times D$, whose elements are initially set to 0. The tensor *T* is dealt as the union of two equivalent prisms P_{dis} *and* P_{cor} each of them of height *D* and with bases formed by orthogonal triangles with legs of length *K*. The prism P_{dis} contains the information related to the distribution of clusters among the adjacent patches of W, while P_{cor} includes information on the correlations between the feature vectors of the adjacent patches of *W*.



a.

Fig. 2. Our tensor-based network. Two different sets of features are extracted independently on the basis of the information of the cluster matrix and feature vectors of the patches. These sets are stored in two different symmetric volumes of the tensor *T*. For each couple of different patches with a given distance $d \leq D$ and with feature vectors belonging to two defined clusters, the first volume (red part) specifies the occurrence in *W* of the selected pattern, while the second volume of *T* (orange part) includes the sum of correlation indexes between the feature vectors associated to the patches of the selected pattern in *W*. Then, the tensor *T* is normalized and is fed a deep network for the classification tasks

In more details, let be $p_{i,j}$ and $p_{i',j'}$ two patches in W lying at the distance $d \leq D$, which have been assigned to clusters *k* and *k'* respectively, that is $CM(i, j) = k$ and $CM(i', j') = k'$. The set $\{k, k', d\}$ (or also $\{k', k, d\}$) identifies a pattern *SP* in *W* represented by any couple of patches of *W* at distance *d* in *W* and with feature vectors belonging to *k-th* and *k'-th* clusters. The occurrence of *SP* in *W* is stored into the tensor P_{dis}. In particular, the point (m, n, d) of *T*, with $m = min(k, k')$ and $n = max(k, k')$, is incremented by 1 every time the pattern *SP* is detected in *W*. Concomitantly, the sum of correlation indexes between the feature vectors associated to the patches in W characterizing SP is stored in P_{cor}. In particular, for each detected SP in *W*, the point (n, m, d) of *T* is incremented by the value of the Pearson coefficient [8] computed between the feature vectors of the patches belonging to the current *SP*. See Fig. 2.

A normalization process is applied on each slice of *T* to obtain values between -1 and 1. This operation is performed adopting the mapping function [9] that is a quasi sigmoid normalization. Finally, the normalized tensor *T* is fed to a VGG-11 network [10].

III. Experiments and Results

The performance of the proposed method has been evaluated on the publicly available histopathology image Camelyon16 dataset [11].

Different experiments have been performed to select: a) the deep networks both for feature vector extraction and for classification; b) the patches set for the clustering process; c) the normalization function and finally, d) the values of *K* and *D*. Moreover, different strategies have been considered for training of our network, considering either single parts or the whole tensor, aimed at assessing the potential contribution of different kinds of information in T. For the sake of brevity, only some of these experiments will be presented in this paper. Comparisons with recent state-of-the-art techniques are provided on the same task with respect to the same testing protocols.

A. Dataset and cluster data preparation

The Camelyon16 dataset contains 400 H&E WSIs of sentinel lymph nodes of breast cancer obtained from two independent sets collected in Radboud University Medical Center (Nijmegen, the Netherlands) and in the University Medical Center Utrecht (Utrecht, the Netherlands). The dataset is originally split into 270 WSIs (160 of normal tissue and 110 containing metastasis) for the training phase and 130 WSIs (80 of normal tissue and 50 containing metastasis) for the test phase; this original splitting was preserved in our experiments. All WSIs of the training set containing metastases are accompanied by manual annotations that have been used for both the training of the ResNet-18 and the selection of patches used for the clustering processes. In particular, 120156 patches have been extracted from the WSI training set to fine-tune the ResNet18 and 15000 patches coming from the WSI test set were used for clustering. The involved patches were appropriately selected from many different images, equally distributing them according to their type, normal or tumor tissue. On the basis of different experiments, the number of clusters *K* was set to 256.

B. Experimental Setup and Results

In this study, each analyzed patch has size $S \times S \times 3$, with *S* equal to 224 and ResNet-18 has been adopted as feature extractor for both the clustering process and the generation of the tensor *T*. The extracted features are one-dimensional vectors of length H=512 elements.

We propose three different scenarios for the classification, depending on whether only one part of *T* (P_{dis} or P_{cor}) or the whole tensor *T* is involved in the analysis.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

For each strategy, the results have been evaluated with tensors at different depths, in particular for D = 4, 8, 16 and 32.

The performance of different approaches has been compared in terms of standard metrics, namely Accuracy, F-Measure, Specificity and Sensitivity. The performance has been also measured in terms of the Area under the ROC Curve (AUC). The numerical results of these experiments are reported in Table I.

TABLE I. RESULTS

	D	*AUC*	*Acc.*	*F-score*	*Spec.*	*Sens.*
T	4	0,61	0,56	0,56	0,45	0,73
	8	0,72	**0,64**	0,66	**0,51**	0,79
	16	0,61	0,60	0,66	0,48	0,70
	32	0,61	0,62	0,69	0,50	0,70
P_{dis}	4	0,51	0,62	**0,77**	–	0,62
	8	0,55	0,62	**0,77**	–	0,62
	16	0,50	0,62	0,73	0,50	0,66
	32	0,49	0,44	0,51	0,32	0,56
P_{cor}	4	0,71	0,63	0,65	**0,51**	0,79
	8	**0,75**	0,63	0,65	**0,51**	0,78
	16	0,70	0,58	0,56	0,47	**0,80**
	32	0,72	0,59	0,60	0,47	0,75

In Table I, the best value for each measure is written in bold, while the best result for each type of tensor is written on a gray background. Considering the values for each strategy as a whole, setting D = 8 represents the best choice for the maximum distance between two different patches of a WSI for whose relationship between the relative features can be taken into account. The highest values of accuracy (0,68), specificity (0,51) and sensitivity (0,79) are obtained when the whole tensor T is considered. The highest value of F1 score (0,77) is provided by P_{dis} (0,66 for T and 0,65 for P_{cor}). The best performance in terms of AUC (0,75) is obtained considering P_{cor} (0,72 for T and 0,55 for P_{dis}). The remaining measures for P_{cor} show values similar to those obtained for T. Thus, the best strategy can be considered the one based on P_{cor} and for D = 8. For this configuration, Fig. 3 shows the confusion matrix and the ROC curve.

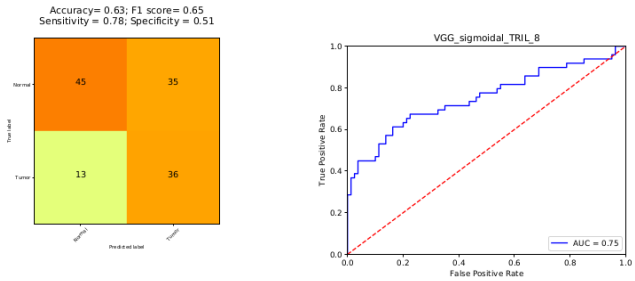


Fig. 3. Confusion matrix and ROC curve of the network, taking into account only P_{cor} with D = 8

The classification result in terms of AUC is comparable with those of the studies in [4] and [5]. The capacity of these methods to reduce the whole-slide images into a compact format was tested in [4] by using three different networks: the Bidirectional Generative Adversarial Network (BiGAN), a Variational AutoEncoder (VAE) and a discriminative model based on contrast training, while the method [5] is based on two Attention networks (AN). The AUC of the BiGAN, VAE and contrastive networks are respectively 0.70, 0.67 and 0.65, while AN provides an AUC equal to 0.71. The results are in line with many of those obtained from the method presented in this study, in which the level of abstraction of whole-slide image representation has increased. However, the result obtained by our model is quite relevant, since for depth levels 4, 8, 16 and 32, it has obtained an AUC of 0.71, 0.75, 0.70 and 0.72 respectively, which is equal to or higher than that obtained from [4] and [5] methods. Thus, the proposed method can represent giga-pixel images in an alternative way, while preserving the ability to discriminate images by classes even at a higher level of abstraction.

IV. CONCLUSIONS

In this paper, a methodology for the analysis of histological images has been proposed which extracts 3D tensors by constructing a grid of clustered deep features and extracting information related to the proximity of the patches. These tensors allow a compact representation of WSIs that can be analyzed by deep learning techniques. Results have shown that a tensor constructed by considering a proximity radius of 8 patches and the correlation measures between the different patches provides the best performance. With this type of image synthesis, the results obtained by the network exceed those obtained by recent studies proposed in the literature, opening up to the future possibility of extending this approach to further reduce the amount of data to be processed, while still obtaining good results in classification tasks.

ACKNOWLEDGMENTS

This work was supported by the project “Campania Oncotherapy - Fighting tumor resistance: an integrated

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

multidisciplinary platform for an innovative technological approach to oncotherapies – Regione Campania".

REFERENCES

[1] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. Shih, ... and F. González, "High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection". PloS one, vol. 13, no. 5, 2018.

[2] K. Das, S. Conjeti, J. Chatterjee, and D. Sheet, "Detection of breast cancer from whole slide histopathological images using deep multiple instance cnn," IEEE Access, vol. 8, pp. 213 502–213 511, 2020.

[3] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in International Conference Image Analysis and Recognition. Springer, 2018, pp. 914–922.

[4] D. Tellez, G. Litjens, J. van der Laak, and F. Ciompi, "Neural image compression for gigapixel histopathology image analysis," IEEE transactions on pattern analysis and machine intelligence, 2019.

[5] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata, and S. Hassanpour, "Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides," JAMA network open, vol. 2, no. 11, pp. e1 914 645–e1 914 645, 2019.

[6] N. Brancati, G. De Pietro, D. Riccio, and M. Frucci "Gigapixel Histopathological Image Analysis using Attention-based Neural Networks". arXiv preprint arXiv:2101.09992, 2021.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[8] S. Stigler, "Francis Galton's Account of the Invention of Correlation". Statistical Science, vol. 4, no. 2, pp. 73–79, 1989.

[9] M. De Marsico and D. Riccio "A new data normalization function for multibiometric contexts: A case study", International Conference Image Analysis and Recognition. Springer. 2008, pp. 1033–1040.

[10] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv: 1409.1556, 2014.

[11] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," Jama, vol. 318, no. 22, pp. 2199–2210, 2017.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.