Article

Improving Breast Tumor Multi-Classification from High-Resolution Histological Images with the Integration of Feature Space Data Augmentation

Nadia Brancati [*] and Maria Frucci

Article

Improving Breast Tumor Multi-Classification from High-Resolution Histological Images with the Integration of Feature Space Data Augmentation

Nadia Brancati *[ID] and **Maria Frucci** [ID]

Institute for High Performance Computing and Networking, National Research Council of Italy (ICAR-CNR), Via P. Castellino 111, Naples 80131, Italy
* Correspondence: nadia.brancati@cnr.it

Abstract: To support pathologists in breast tumor diagnosis, deep learning plays a crucial role in the development of histological whole slide image (WSI) classification methods. However, automatic classification is challenging due to the high-resolution data and the scarcity of representative training data. To tackle these limitations, we propose a deep learning-based breast tumor gigapixel histological image multi-classifier integrated with a high-resolution data augmentation model to process the entire slide by exploring its local and global information and generating its different synthetic versions. The key idea is to perform classification and augmentation in feature latent space, reducing the computational cost while preserving the class label of the input. We adopt a deep learning-based multi-classification method and evaluate the contribution given by a Conditional Generative Adversarial Network-based data augmentation model on the classifier's performance for three tumor classes in the BRIGHT Challenge dataset. The proposed method has allowed us to achieve an average F1 equal to 69.5, considering only the WSI dataset of the Challenge. The results are comparable to those obtained by the Challenge winning method (71.6), also trained on the annotated tumor region dataset of the Challenge.

Keywords: deep learning; whole slide image; augmentation; classification; feature space

---

1. Introduction

In recent years, deep learning has played a key role in developing methods for the automatic classification of histological images to support pathologists in tumor diagnosis [1,2]. One of the first challenges of any research in the field of deep learning for computational pathology is the collection of representative datasets. The greater the diversity and amount of data in the training set, the more significantly does the model's performance improve. However, creating a large volume of data with corresponding annotations requires complex and time-consuming work by pathologists. Moreover, in many cases, the distribution of disease subclasses in the histological image datasets collected can be highly skewed in the presence of rare tumors. To mitigate these issues, reliable data augmentation methods can generate synthetic data for the classification training model. Data augmentation not only expands the image dataset but also helps to reduce any data imbalance, thereby preventing overfitting and enhancing the model's capacity to generalize.

Classic image data augmentation techniques consist in geometric transformations and photometric shiftings. They include image processing techniques based on image transformations (flipping, reflection, zooming, scaling etc.) and color transformations (histogram matching, change in brightness, contrast etc.) [3]. Such methods produce new samples that are similar to the original ones, an increase which can slightly improve the performance of the deep learning method. Other augmentation methods are based on image mixing and deleting techniques to compose real images, either by pasting object masks in backgrounds or by mixing two images to hide or highlight image characteristics [4–6]. However, the samples obtained with this specific type of augmentation on medical images may not meet the requirements of a clinical characterization of the histological images. Unlike augmentation

techniques mainly based on image processing methods, other algorithms generate synthetic data using deep learning models. In particular, after the initial development of Generative Adversarial Networks (GANs), models generating completely new images have become a popular method for data augmentation [7,8].

However, in the context of histological image processing, using standard deep learning techniques, regardless of the purpose (classification or data augmentation), is not feasible in terms of managing high-dimensional histological images, i.e., Whole Slide Images (WSIs). It is impossible to feed all the pixels from such images into a neural network simultaneously. Thus, WSI processing is generally inferred by combining the decisions obtained for the single patches in which the WSI input has been split [9,10]. The drawback of such a strategy is the reduced analysis of the spatial relations between these tiles, which produces a potentially debilitating analysis of the weak patterns and an inappropriate generation of synthetic data. These issues are even more evident when the absence of annotation for regions of interest (ROIs) within WSIs does not allow the addressing of significant patches.

Specifically, for the classification task, a tumor area in a WSI cannot be arranged into a category by analyzing the individual parts without preserving the spatial correlations between the patches and the interaction between the entire tumor region and the neighborhood tissue microenvironment information. For the augmentation task, generating individual synthetic patches might introduce excessive noise, potentially disrupting weak patterns or applying too few variations, failing to introduce sufficient diversity to enhance the robustness of the classification models. Therefore, there is a need for innovative data generation approaches customized for high-dimensional WSIs which strike the right balance between incorporating diversity and preserving essential patterns within these images.

With the aim of preserving the spatial correlations in the WSI, some recent methods for the classification task only [11–13] and for the data augmentation improvement classification [14,15] propose mapping images from a low-level pixel space to a higher-level latent space using neural networks. Features are extracted patch-wise by a network and rearranged to form a compressed image which saves most of the discriminative information and can be used to classify the entire WSI. Based on this approach, we introduce a model to perform data augmentation in feature space instead of in image space by relinquishing visual evidence but reducing computational costs and promoting global image analysis. In particular, we introduce a data augmentation model that does not generate images but rather "synthetic" feature maps. Using a residual network dedicated to extracting features from each patch into which the image has been partitioned, the WSI input is transformed into a grid-based feature map (GFM). Next, the proposed data augmentation model, based on traditional methodologies (rotation, flipping and translation) and a trained generative network, operates on the GFM to generate synthetic GFMs. This model has been integrated into a histological breast tumor image multi-classifier, which analyzes both the GFMs of the input WSI and the generated data.

In detail, together with classic augmentation methodologies, we have used a Conditional Generative Adversarial Network (cGAN) [16], a variant of a standard GAN, to generate the synthetic features of a WSI. Integrating such a network with the deep learning-based multi-classification architecture presented in [11] and adopted as the baseline of the BRIGHT Challenge[1], we have evaluated the contribution given by the proposed augmentation approach to the performance of the multi-classifier for three breast tumor classes. The adopted model uses only the label of the WSI (weak label annotation) for the classification, not requiring any further pixel-level or patch-level annotations. To achieve this, the experimentation has been conducted considering only the annotated WSIs of the reference dataset. The results show that the performance of the proposed model is comparable with that obtained by the winning team which secured first place in the BRIGHT Challenge[2], whose method

---

[1]   https://research.ibm.com/haifa/Workshops/BRIGHT/
[2]   https://www.synapse.org/#!Synapse:syn26480664/wiki/616847

is based on the combination of strongly-annotated data (from ROIs) and weakly-annotated data (from WSIs) [17], proving better than other recent methods [18,19].

Specifically, the highlights of the proposed work are:

- performing the classification and augmentation in feature latent space allows the model to deal with the local and global information of gigapixel WSIs;
- working in feature latent space reduces the computational cost while preserving the class label of the input; and
- combining a two-stage augmentation process into an all-in-one model enables the network to generate and classify the WSIs during the training without a separate classifier, increasing the classification accuracy on the testing dataset.

2. Method

The proposed approach to the classification-augmentation integration is illustrated in Figure 1. A deep neural network is employed as a backbone to extract a more compact and meaningful representation of the original data, reducing its dimensionality. In this context, the features extracted in this unsupervised manner can be seen as a representation of the latent feature space, which is an abstract representation of data or information. The input image is split into patches, each of which is represented by using an embedding structure. These feature vectors are generated using a Convolutional Neural Network (CNN) and are aggregated into a GFM according to the spatial location of the corresponding patches in the WSI input. Online data augmentation techniques are applied to the GFM during the training, reducing the storage resources required to implement the data augmentation. A cGAN network generates new "synthetic" GFM representations for the specific type of class c of the WSI input. At the same time, classic affine transformations produce different placements of the feature vectors into the initial GFM. Next, an Attention-Based Neural Network is adopted for the WSI multi-classification [11].
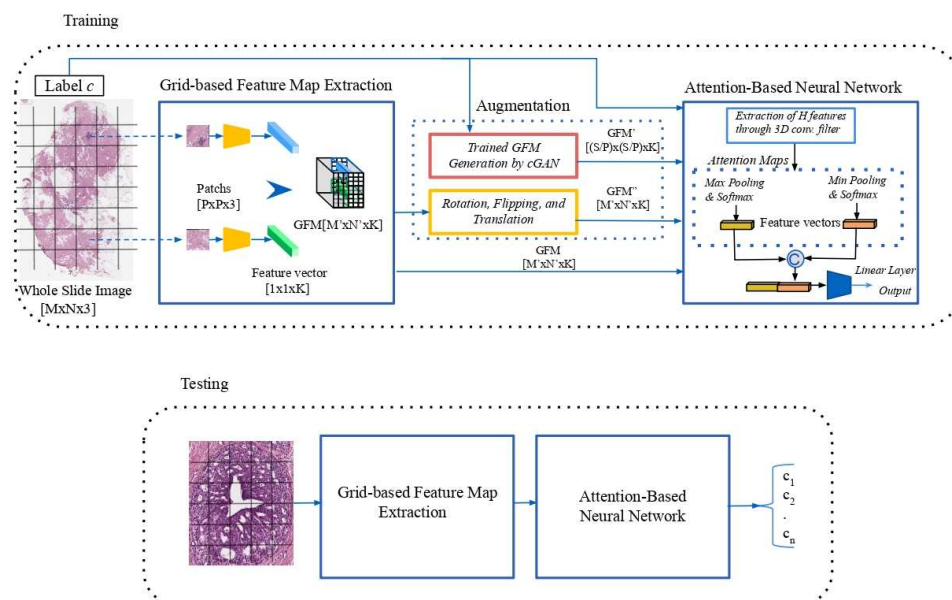

Figure 1. The diagram's higher and lower sections illustrate the training and testing phases, respectively. For the training, after transforming the WSI input into a GFM, new GFM representations are generated by an augmentation module. An Attention-Based Neural Network processes the GFM input of one of the generated GFMs.

2.1. Grid-based Features Map Extraction

The purpose of the process of extraction of the GFM is to produce a more compact representation of the WSI input that could be managed in its entirety by the following processing pipeline. The input is a gigapixel WSI, with a dimension of $M \times N \times 3$, and it is partitioned into a set of non-overlapping patches of a size equal to $P \times P \times 3$, mapped into a 1D feature vector with length K by applying a CNN. According to the patch-wise spatial order, a GFM of size $(M/P) \times (N/P) \times K$ is used to store the extracted feature vectors.

2.2. GFM Augmentation

In the following sections, two different processes producing new versions of the GFM are described.

2.2.1. Standard Data Augmentation

The training GFM undergoes standard affine transformations, which include horizontal and vertical flips, shifts and rotations. These transformations do not alter the content of the GFM input. They are applied dynamically during batch generation throughout the training process, utilizing standard image processing functions in the PyTorch torch-vision library.

2.2.2. Conditional Generative Adversarial Network (cGAN)

Generative adversarial networks are deep learning models that use two competing neural networks, a Generator and a Discriminator, to create realistic data. The generator expects to fake sufficiently realistic samples and the discriminator seeks to differentiate real data from those produced by the generator. Specifically, the generator takes random noise as the input and generates synthetic samples intended to be similar to the real training data. The discriminator receives in input the candidate synthesized by the generator and a real sample of the dataset by performing accurate real/fake binary classification, i.e., indicating the probability that the synthetic input is real. The training of these modules is called adversarial because both compete against each other regarding the training objective. The two modules are trained iteratively and the performance increases alternately.

cGANs use additional information to control the generation process, providing the ability to create samples belonging to a specific class. This process can be useful in applications that aim to generate data with specific properties or characteristics. The architecture of a cGAN is similar to that of a traditional GAN, where a specific label c is concatenated with the input noise to feed the generator. The generator then produces a sample that is intended to match label c. In the same way, the discriminator not only determines whether the sample is realistic but also whether the sample matches the label c. In this work, the label c identifies a specific tumor class of the dataset used for the experiments.

A conditional component, introduced into the generator and discriminator losses, ensures that the generated data are consistent with the provided label c. The generator loss evaluates how well the discriminator identifies the generated data as real data. In this work, the Mean Squared Error (MSE) between the fake output and real label is used to minimize the probability that the discriminator can distinguish between generated and real data. The discriminator loss measures how well the discriminator correctly performs the binary classification. It is split into two components:

- Discriminator Loss Real: this quantifies how well the discriminator correctly identifies real data as real.
- Discriminator Loss Fake: Quantifies how well the discriminator correctly identifies generated data as fake.

In this work, the first (second) loss component has been computed as the MSE between the real (fake) output and the real (fake) label. The adopted cGAN structure [3] requires inputs of the same size and with a square shape. For this reason, only for the training of the cGAN, each WSI input is resized at the same dimension of $S \times S \times 3$, where S has been selected on the basis of the medium size of the WSI training dataset. See Figure 2 .
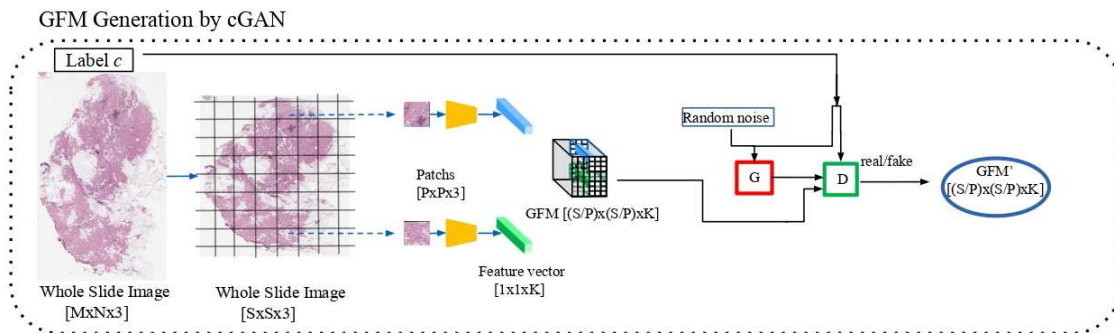

Figure 2. Synthetic GFM generation using a cGAN: each WSI input is resized at the dimension of $S \times S \times 3$. The GFM representation and the label class of the WSI are given as input to a cGAN, which produces a new synthetic GFM of the same dimension.

2.3. Convolutional Neural Network (CNN) Classifier

The learning path of the multi-classifier is an Attention-Based Neural Network consisting of a 3D Convolutional layer followed by two attention modules. The Convolutional Layer merges the information provided by the features of the neighboring patches, producing a new compact volume fed independently to two attention modules, each producing a 3D attention map. The attention modules, based separately on min- and max-pooling mechanisms, are employed to lead the classifier to focus on features that are considered more expressive for the class learned. The two different sets of attention maps contain complementary information. They are inserted into the network collaboratively to highlight discriminative feature channels while suppressing irrelevant information with respect to the actual classification task. Each attention map is used to assign weights to features in the examined GFM and produces a feature vector. The concatenation of the two feature vectors is fed to a linear layer that produces an image-level label.

3. Experiments and results

Different experiments have been performed to assess the performance of the proposed method on the histological image dataset provided by the BRIGHT Challenge. In particular, different strategies have been considered for the training of the classification and cGAN networks to assess the potential contribution of cGAN augmentation. The performance of the proposed model has been compared with that provided by state-of-the-art techniques on the same task for the same testing protocols of the BRIGHT Challenge.

3.1. Dataset

The dataset used is employed in the BRIGHT Challenge, which includes all the images of the BRACS dataset [20]. This dataset was built through the collaboration of the National Cancer Institute - Scientific Institute for Research, Hospitalization and Healthcare (IRCCS) "G. Pascale Foundation", the Institute for High Performance Computing and Networking (ICAR) of the National Research Council

---

[3]   https://github.com/eriklindernoren/PyTorch-GAN/blob/master/implementations/cgan/cgan.py

(CNR) and IBM Research–Zurich. It includes both annotated WSIs and ROIs. However, in this work, only the WSIs have been considered for the experiments. The WSIs of the BRIGHT testing set are made publicly accessible without annotations since the organizers of the BRIGHT Challenge perform the performance evaluation of the classification algorithms according to a blind protocol. The authors of the current work are among the organizers of the Challenge, and they can easily access the annotations of the testing set. The BRIGHT dataset includes three breast lesion types, namely Non-cancerous, Pre-cancerous, and Cancerous, which are further subdivided into six specified tumor subtypes, namely Pathological Benign (PB), Usual Ductal Hyperplasia (UDH), Flat Epithelia Atypia (FEA), Atypical Ductal Hyperplasia (ADH), Ductal Carcinoma in Situ (DCIS) and Invasive Carcinoma (IC). Table 1 shows the distribution in the training, validation and testing sets and the origin of the WSIs for the Challenge according to the lesion type and tumor subtype. The WSIs have been obtained by scanning slides using an Aperio AT2 scanner at a resolution of 0.25 $\mu m / pixel$ with a 40× magnification factor.

Table 1. Number of WSIs from the BRIGHT dataset for the three breast tumor classes and the six tumor subtypes, divided into training, validation and testing sets.

	Non-cancerous		Pre-Cancerous		Cancerous		Total
	PB	UDH	FEA	ADH	DCIS	IC	
Training	131	65	30	36	49	112	423
Validation	16	9	11	12	12	20	80
Testing	34	33	33	33	33	34	200

In the BRIGHT Challenge, two WSI classification tasks have been established:

- Task 1: a 3-class WSI classification is required by grouping the original six tumor subtypes into three lesion types: Non-cancerous (PB+UDH), Pre-cancerous (ADH+FEA), and Cancerous (DCIS+IC).
- Task 2: a 6-class WSI classification is required to perform a fine-grained subtyping of the tumors within the WSIs. Six tumor subtypes have been considered: PB, UDH, ADH, FEA, DCIS, and IC.

In this work, different experiments have been performed and discussed for Task 1, while only the final results are shown for Task 2.

3.2. Training Protocol

All the WSIs have been normalized using the method proposed in [21]. Morphological operations have been employed to eliminate the peripheral white areas of the WSI inputs, which represent the slide's background outside the tissue section. The experiments have mainly been focused on the 3-class task for which a WSI magnification of 2.5× has produced the best results. In general, a low magnification analysis is useful to obtain an overview of the sample, allowing the pathologist to identify basic tissue structures and architecture. Thus, images with a lower magnification help to solve the first task of the classification into three classes. Conversely, the pathologist can examine cellular details at higher magnification levels, such as the cell morphology, mitosis and other features that distinguish various lesions [22,23]. Therefore, by analyzing higher-resolution images (e.g., at a 10× magnification), the 6-class task might be better resolved. As the GFM extraction module, the WSI input is split into a set of non-overlapping patches with a size equal to $P \times P \times 3$, with $P = 224$, and a residual neural network ResNet-34 pre-trained on natural images from ImageNet [24], adopted to generate feature vectors with length $K = 512$ for each patch.

3.2.1. cGAN Training

As specified in Section 2.2.2, the cGAN network receives as input a GFM obtained by resizing the WSI input at the $S \times S \times 3$ dimension. The value $S = 7168$ has been selected on the basis of the

evaluation of the medium size of the WSIs of the training set. Thus, all the generated GFMs have dimensions equal to $32 \times 32 \times 512$. The cGAN has been trained for 1000 epochs with a batch size of 64. The MSE Loss has been considered for the backpropagation and an Adam optimizer [25] has been adopted with a learning rate equal to $1e - 6$. Figure 3 shows the Generator and Discriminator losses for the training.
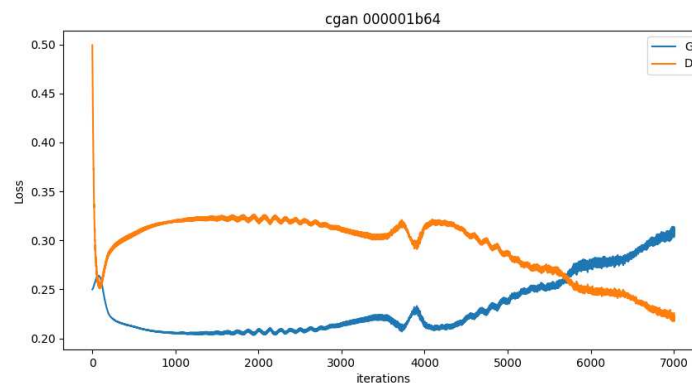

Figure 3. The Generator and Discriminator loss curves (blue and orange lines, respectively) during the training: after approximately 750 epochs, the generator begins to improve, while the performance of the discriminator deteriorates.

3.2.2. Classifier training

The multi-classifier analyzes the GFMs obtained by the WSI input, affine transformations and trained cGAN without limitations about the shape and size of the WSI input. The online augmentation includes nine affine transformations of the GFM input (three rotations at $90°$, $180°$ and $270°$, two flips, horizontally and vertically, and four shifts, right, left, up and down) and the trained cGAN, which generates nine synthetic versions consistently with the class of the WSI input. Additionally, to augment further the amount of the training data, the non-normalized WSI input has been considered. Thus, a set of 20 different feature map representations may be generated for each WSI input. Indeed, to obtain a better balance of the training data set, the number of samples generated for the classes consisting of a high number of data must be lower than that of the minority classes. For this reason, in correspondence with inputs belonging to the majority classes, not all 20 feature map representations are produced. An augmentation selection is randomly provided through the type specification of the affine transformations to apply and the number of synthetic versions to generate. Table 2 shows the balancing of the initial training dataset after the controlled augmentation process. In the original dataset, the number of WSIs belonging to the pre-cancerous class proves to be less than that of the non-cancerous and cancerous classes. Thus, all 20 different feature map representations have been produced only for the Pre-cancerous lesions. The attention-based network has been trained for 100 epochs with a batch size equal to 8. Cross Entropy Loss has been considered for the backpropagation and an AdamW optimizer [26] has been adopted with an initial learning rate equal to $1e - 4$ and using the "learning rate annealing with restarts" technique. The combination of these two learning methods is aimed at enhancing the efficiency and effectiveness of the neural network training and limit overfitting. In detail, AdamW is a stochastic optimization technique based on the separation of weight decay from the gradient update. While the original implementation of Adam applies weight decay directly to the weights during the update, AdamW applies weight decay after each weight update step, preventing excessive weight shrinkage during the training. The weight decay has been set equal to $1e - 5$. Different learning rates are set in the "learning rate annealing with restarts" technique and "restarts" are scheduled during the training. During each cycle, the learning rate gradually decreases over time but is periodically reset to a higher value ("restart period"). The restart period has been set equal to 6, i.e., every six epochs, and the learning rate is equal to the initial value $1e - 4$.

Table 2. Number of WSIs in the original training dataset and the balanced augmented dataset.

	Non-cancerous	Pre-Cancerous	Cancerous	Total
Original Dataset	196	66	161	**423**
Augmented Dataset	1960 (196x10)	1320 (66x20)	1932(161x12)	**5212**

3.3. Result and discussion

Considering the strong imbalance of the training dataset, the F1-score has been chosen as a metric to evaluate the proposed method. Table 3 shows a comparison with three recent state-of-the-art gigapixel WSI classification methods. The first method [19] is a recent weakly supervised framework which processes the entire slide by exploring its local and global information. The deep network is split into numerous gradient-isolated modules, each trained independently with local supervision. Moreover, a Random Feature Reconstruction model is introduced to enhance the performance and optimize the GPU utilization. The publicly available online code[4] has been used to obtain the results of the method on the WSI BRIGHT dataset. The second method, namely WINM, has been presented to the BRIGHT Challenge through two submissions with different settings of the learning parameters, ranking in first and second place. WINM is a semi-supervised multiple-instance learning method based on cross-slide contrastive learning in which patch-level annotations are used to regularize the attention mechanism. For this reason, the annotated tumor region training dataset is also used in the learning phase. The team submitted only a draft paper on WINM to the challenge organizers and presented in [17] a modified version of WINM, tested on different datasets, which do not require the involvement of the annotated ROIs. The performance of the method in [17] is not evaluable on the WSI BRIGHT dataset due to the fact that the code is not publicly available. The third method [18], ranked in third place in the BRIGHT Challenge, is based on a Multiple Instance Learning CNN, allowing the combination of strongly annotated data (from ROIs) and weakly annotated data (from WSIs) via the optimization of a multi-task loss function. The data in Table 3 on the performance in terms of the F1-score of the five different classifiers highlight how the proposed method provides a good performance even if a simple network architecture is adopted and no additional annotations in the training phase other than those related to the labeling of WSIs are used. Except for [18], all the compared methods provide the best (worst) performance for the Cancerous (Pre-cancerous) class. However, the proposed method obtains the highest value for the Non-cancerous class. The performance of the proposed method is significantly higher than that of [18,19] in terms of the F1-score for both the average and the single classes, although [18] involves the ROIs in the training phase.With respect to WINM, the results are comparable, although WINM uses the annotated training ROIs. Indeed, only for the Cancerous class is the evaluation gap higher. In contrast, for the Pre-cancerous class, the proposed method has a slightly lower (higher) performance than the first-ranking (second-ranking) WINM. These results highlight the importance of label-conditional data generation on the performance of the classifier. This is even more evident from the results of some ablative experiments, shown in Table 4 and described below.

---

[4]   https://github.com/cvlab-stonybrook/local_learning_wsi/blob/main

Table 3. Comparison of the proposed method for the 3-classes task, in terms of the F1-score, with recent state-of-the-art methods.

Experiment	F1-average	F1-Non-Cancerous	F1-Pre-Cancerous	F1-Cancerous
[19]	65.3	68.0	54.0	74.2
WINM 1° rank	71.6	72.5	62.3	80.0
WINM 2° rank	69.6	70.8	52.7	85.3
[18]	65.0	71.8	51.6	71.7
Proposed Method	69.5	74.0	59.5	75.2

In the first experiment (namely AT_NcGAN), only affine transformations have been considered in the augmentation module; in the second experiment (namely NAT_cGAN), only the contribution of the cGAN is evaluated; and in the last experiment (namely AT_3GAN), the cGAN has been substituted by three traditional GANs, one for every single label. When only one augmentation method is adopted (AT_NcGAN or NAT_cGAN), the experiments have always been conducted with a controlled balancing while avoiding excessive increases in the number of samples to prevent overfitting. Therefore, for these two types of experiments, the amount of the generated data is lower than that obtained with AT_3GAN and the proposed method. The greatest increase in performance is achieved for the Pre-cancerous class when generative adversarial networks (conditional or not) are adopted for the data augmentation. Pre-cancerous tumors are very similar to cancerous ones [20]; thus, the affine transformations, generating samples very similar to the original data, provide only a slight improvement in the classifier performance in terms of discriminating between pre-cancerous and cancerous tumors. The use of only the cGAN clearly allows for a significant improvement in performance for the pre-cancerous cases without introducing substantial changes for the other two classes. Although the combination of the affine transformations with a generative adversarial network allows the obtainment of a greater number of different samples, preventing overfitting, the performance in terms of the F1-average of AT_3GAN is worse than that obtained with NAT_cGAN. Indeed, each GAN is trained with fewer samples for a given tumor class. For these reasons, the proposed augmentation model based on the combination of the affine transformations and a conditional GAN provides the best performance in terms of the F1-score for both the average and the single classes.

Table 4. Ablative experiments adopting different combinations of augmentation processes: AT_NcGAN for only affine transformations, NAT_cGAN for only cGAN and AT_3GAN for a fusion of the affine transformations with three GANs, one for each label.

Experiment	F1-average	F1-Non-Cancerous	F1-Pre-Cancerous	F1-Cancerous
AT_NcGAN	61.6	71.0	44.2	69.4
NAT_cGAN	62.4	70.5	47.8	68.9
AT_3GAN	62.1	66.7	48.2	71.3
Proposed Method	69.5	74.0	59.5	75.2

We have also evaluated the proposed method for Task 2 of the Challenge (i.e., for six classes), ranking sixth in the challenge leaderboard, preceded by WINM and [18] with different parameter settings. This task is more challenging as it involves a greater number of classes and a higher potential for inter-class ambiguities. Thus, the involvement of annotated ROIs in the network training phase has

significantly benefited WINM and [18] in terms of performance. The comparison with [19], shown in Table 5, highlights a better performance in terms of F1 score for both the average and the four classes.

Table 5. Comparison of the proposed method for the 6-classes task, in terms of F1-score, with a recent weakly supervised method.

Experiment	F1-average	F1-PB	F1-UDH	F1-FEA	F1-ADH	F1-DCIS	F1-IC
[19]	40.3	46.1	41.5	31.8	13.0	45.3	64.1
Proposed Method	41.5	49.6	28.0	40.0	20.0	40.8	71.2

4. Conclusions

In this work, a WSI multi-classification using weakly supervised learning integrated with a high-resolution data augmentation model has been applied to the imbalanced dataset of breast gigapixel WSIs of the BRIGHT Challenge. Rather than using the conventional approach based on the analysis of the single patches, the notion of an intrinsic feature relationship across patches has been used to realize a compact representation of the WSI input as a GFM. Both classification and data augmentation operate at the compact representation level, allowing an efficient analysis of the gigapixel WSIs. Activating the augmentation module during the classifier training brings two advantages: high-resolution synthetic data generation and reduced storage resources required to implement the data augmentation. The augmentation module is based on affine transformations and a cGAN operating on the GFM input. Experiments have been performed to evaluate the contribution of the augmentation module for the multi-classification. This study shows that a well-designed and adequately trained deep learning model can achieve, considering only a weak label dataset, an average F1 comparable to that of the winning method of the BRIGHT Challenge, also trained on the annotated tumor region dataset, and a result better than that obtained by other recent methods.

Author Contributions: Conceptualization, N.B., and M.F.; methodology, N.B., and M.F.; software, N.B. ; validation, N.B. and M.F. ; formal analysis, M.F. ; data curation, N.B. and M.F.; supervision, M.F.; original draft preparation, N.B., and M.F.. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Huang, Y.; Zhao, W.; Wang, S.; Fu, Y.; Jiang, Y.; Yu, L. ConSlide: Asynchronous Hierarchical Interaction Transformer with Breakup-Reorganize Rehearsal for Continual Whole Slide Image Analysis. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 21349–21360.
2. Wetstein, S.C.; de Jong, V.M.; Stathonikos, N.; Opdam, M.; Dackus, G.M.; Pluim, J.P.; van Diest, P.J.; Veta, M. Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. *Scientific reports* **2022**, *12*, 15102.
3. Khalifa, N.E.; Loey, M.; Mirjalili, S. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review* **2022**, pp. 1–27.
4. Naveed, H.; Anwar, S.; Hayat, M.; Javed, K.; Mian, A. Survey: Image mixing and deleting for data augmentation. *arXiv preprint arXiv:2106.07085* **2021**.
5. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple copy-paste is a strong data augmentation method for instance segmentation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 2918–2928.
6. Wang, F.; Wang, H.; Wei, C.; Yuille, A.; Shen, W. CP 2: Copy-Paste Contrastive Pretraining for Semantic Segmentation. European Conference on Computer Vision. Springer, 2022, pp. 499–515.

7. Chen, Y.; Yang, X.H.; Wei, Z.; Heidari, A.A.; Zheng, N.; Li, Z.; Chen, H.; Hu, H.; Zhou, Q.; Guan, Q. Generative adversarial networks in medical image augmentation: A review. *Computers in Biology and Medicine* **2022**, *144*, 105382.

8. AlAmir, M.; AlGhamdi, M. The Role of generative adversarial network in medical image analysis: An in-depth survey. *ACM Computing Surveys* **2022**, *55*, 1–36.

9. Gecer, B.; Aksoy, S.; Mercan, E.; Shapiro, L.G.; Weaver, D.L.; Elmore, J.G. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. *Pattern recognition* **2018**, *84*, 345–356.

10. Zhang, X.; Liu, C.; Li, T.; Zhou, Y. The whole slide breast histopathology image detection based on a fused model and heatmaps. *Biomedical Signal Processing and Control* **2023**, *82*, 104532.

11. Brancati, N.; De Pietro, G.; Riccio, D.; Frucci, M. Gigapixel histopathological image analysis using attention-based neural networks. *IEEE Access* **2021**, *9*, 87552–87562.

12. Tomita, N.; Abdollahi, B.; Wei, J.; Ren, B.; Suriawinata, A.; Hassanpour, S. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. *JAMA network open* **2019**, *2*, e1914645–e1914645.

13. Tellez, D.; Litjens, G.; van der Laak, J.; Ciompi, F. Neural image compression for gigapixel histopathology image analysis. *IEEE transactions on pattern analysis and machine intelligence* **2019**, *43*, 567–578.

14. Ma, P.; Lu, H.; Yang, B.; Ran, W. GAN-MVAE: A discriminative latent feature generation framework for generalized zero-shot learning. *Pattern Recognition Letters* **2022**, *155*, 77–83.

15. Wu, H.; Gao, R.; Sheng, Y.P.; Chen, B.; Li, S. SDAE-GAN: enable high-dimensional pathological images in liver cancer survival prediction with a policy gradient based data augmentation method. *Medical image analysis* **2020**, *62*, 101640.

16. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* **2014**.

17. Wang, X.; Xiang, J.; Zhang, J.; Yang, S.; Yang, Z.; Wang, M.H.; Zhang, J.; Yang, W.; Huang, J.; Han, X. SCL-WC: Cross-slide contrastive learning for weakly-supervised whole-slide image classification. *Advances in neural information processing systems* **2022**, *35*, 18009–18021.

18. Marini, N.; Wodzinski, M.; Atzori, M.; Müller, H. A Multi-Task Multiple Instance Learning Algorithm to Analyze Large Whole Slide Images from Bright Challenge 2022. 2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC). IEEE, 2022, pp. 1–4.

19. Zhang, J.; Zhang, X.; Ma, K.; Gupta, R.; Saltz, J.; Vakalopoulou, M.; Samaras, D. Gigapixel whole-slide images classification using locally supervised learning. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022, pp. 192–201.

20. Brancati, N.; Anniciello, A.M.; Pati, P.; Riccio, D.; Scognamiglio, G.; Jaume, G.; De Pietro, G.; Di Bonito, M.; Foncubierta, A.; Botti, G.; others. BRACS: A dataset for breast carcinoma subtyping in h&e histology images. *Database* **2022**, *2022*, baac093.

21. Macenko, M.; Niethammer, M.; Marron, J.S.; Borland, D.; Woosley, J.T.; Guan, X.; Schmitt, C.; Thomas, N.E. A method for normalizing histology slides for quantitative analysis. 2009 IEEE international symposium on biomedical imaging: from nano to macro. IEEE, 2009, pp. 1107–1110.

22. Neary-Zajiczek, L.; Beresna, L.; Razavi, B.; Pawar, V.; Shaw, M.; Stoyanov, D. Minimum resolution requirements of digital pathology images for accurate classification. *Medical Image Analysis* **2023**, *89*, 102891.

23. Mayouf, M.S.; Dupin de Saint-Cyr, F. Curriculum Incremental Deep Learning on BreakHis DataSet. Proceedings of the 2022 8th International Conference on Computer Technology Applications, 2022, pp. 35–41.

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

25. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.

26. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* **2017**.