mathematics

MDPI

Article

Improving Cancer Metastasis Detection via Effective Contrastive Learning

Haixia Zheng, Yu Zhou and Xin Huang *

College of Data Science, Taiyuan University of Technology, Taiyuan 030024, China; zhenghaixia@tyut.edu.cn (H.Z.); zhouyu@tyut.edu.cn (Y.Z.)
* Correspondence: huangxin@tyut.edu.cn

Abstract: The metastasis detection in lymph nodes via microscopic examination of H&E stained histopathological images is one of the most crucial diagnostic procedures for breast cancer staging. The manual analysis is extremely labor-intensive and time-consuming because of complexities and diversities of histopathological images. Deep learning has been utilized in automatic cancer metastasis detection in recent years. The success of supervised deep learning is credited to a large labeled dataset, which is hard to obtain in medical image analysis. Contrastive learning, a branch of self-supervised learning, can help in this aspect through introducing an advanced strategy to learn discriminative feature representations from unlabeled images. In this paper, we propose to improve breast cancer metastasis detection through self-supervised contrastive learning, which is used as an accessional task in the detection pipeline, allowing a feature extractor to learn more valuable representations, even if there are fewer annotation images. Furthermore, we extend the proposed approach to exploit unlabeled images in a semi-supervised manner, as self-supervision does not need labeled data at all. Extensive experiments on the benchmark Camelyon2016 Grand Challenge dataset demonstrate that self-supervision can improve cancer metastasis detection performance leading to state-of-the-art results.

Keywords: convolutional neural network; contrastive learning; self-supervision; deep learning; breast cancer detection

MSC: 68-11

Citation: Zheng, H.; Zhou, Y.; Huang, X. Improving Cancer Metastasis Detection via Effective Contrastive Learning. *Mathematics* **2022**, *10*, 2404. https://doi.org/10.3390/math10142404

Academic Editors: Radu Tudor Ionescu and Ion Necoara

Received: 27 May 2022
Accepted: 7 July 2022
Published: 8 July 2022

1. Introduction

Cancer is currently one of the major causes of death for people all over the world. It is estimated that 14.5 million people have died of cancer, and by 2030 this figure is expected to exceed 28 million. The most common cancer for women is breast cancer. Every year, 2.1 million people around the world are diagnosed with breast cancer, according to World Health Organization (WHO) [1]. Due to the high rate of mortality, considerable efforts were made in the past decade to detect breast cancer from histological images so as to improve survival through early breast tissue diagnosis.

Since lymph node is the first position of breast cancer metastasis, metastasis identification of lymph node is one of the most essential criteria for early detection. In order to analyze the characteristics of tissues, pathologists examine tissue slices under the microscope [2]. The tissue slices are traditionally directly observed with a histopathologist's naked eyes and visual data are assessed manually based on prior medical knowledge. The manual analysis is highly time consuming and labor expensive due to the intricacies and diversities of histopathological images. At the same time, highly depending on histopathologist's expertise, workload, and current mood, the manual diagnostic procedure is subjective and limited repeatability. In addition, in the face of escalating demands for diagnostics with increased cancer incidence, there is a serious shortage of pathologists [3].

Hundreds of biopsies must be diagnosed daily by pathologists, thus it is almost impossible to thoroughly examine the entire slides. However, if only regions of interest are investigated, the chance of incorrect diagnosis may increase. To this end, in order to increase the efficiency and reliability of pathological examination, it is required to develop automatic detection techniques.

However, automated metastasis identification in sentinel lymph node from whole-slide image (WSI) is extremely challenging for the following reasons: first, the hard imitations in normal tissues usually look similar in morphology to metastatic areas, which leads to many false positives; second, the great varieties in biological structures and textures of metastatic and background areas; third, the varied circumstances of histological image processing, such as staining, cutting, sampling and digitization, enhance the variations of the appearance of image. This usually happens while tissue samples are taken at different time points or from different patients. Last but not least, WSI is incredibly huge, around 100,000 pixels $\times$ 200,000 pixels, and may not be directly input into any emerging method for cancer identification. Therefore, one of the major issues for automatic detection algorithms is how to analyze such a large pixel image effectively.

Artificial Intelligence (AI) technologies have developed rapidly in recent years. Especially in computer vision, image processing, and analysis, they have achieved outstanding breakthroughs. In histopathological diagnosis, AI has also exhibited potential advantages. With the help of AI-assisted diagnostic approaches, valuable information about diagnostics may be speedily extracted from big data, alleviating the workload of pathologists. At the same time, AI-aided diagnostics have more objective analysis capabilities and can avoid subjective discrepancies of manual analysis. To a certain extent, the use of artificial intelligence can not only improve work efficiency, but also reduce the rate of misdiagnosis by pathologists.

In the past few decades, a lot of works for breast histology image recognition have been developed. Early research used hand-made features to capture tissue properties in a specific area for automatic detection [4–6]. However, hand-made features are not sufficiently discriminative to describe a wide variety of shapes and textures. Recently, a deep Convolutional Neural Network (CNN) has been utilized to detect cancer metastases that can learn more effective feature representation and obtain higher detection accuracy in a data-driven approach [7–9]. The primary factor that may degrade the performance of CNN-based detection methods is the insufficiency of training samples, which may cause overfitting during the training process. In most medical circumstances, it is unrealistic to require understaffed radiologists to spend time creating such huge annotation sets for every new application. Therefore, in order to address the problem of lack of sufficient annotated data samples, it is critical to build less data-hungry algorithms capable of producing excellent performance with minimal annotations.

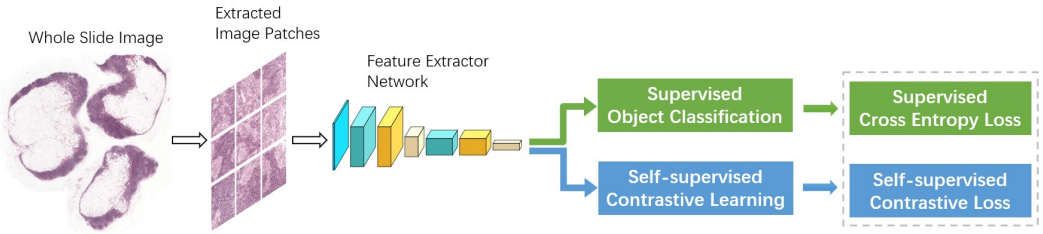

Figure 1. Overview of the proposed architecture.

Self-supervised learning is a new unsupervised learning paradigm that does not require data annotations. In this paper, we propose a multi-task setting where we train the backbone model through joint supervision from the supervised detection target-task and an additional self-supervised contrastive learning task, as shown in Figure 1. Unlike most multi-task cases, where the goal is to achieve desired performance on all tasks at the same time, our aim is to enhance the performance of the backbone model through exploiting the

supervision from the additional contrastive learning task. More specifically, we extend the initial training loss with an extra self-supervised contrastive loss. As a result, the artificially augmented training task contributes to learning a more diverse set of features. Furthermore, we can incorporate unlabeled data to the training process, since self-supervision does not need labeled data. Through increasing the number and diversity of training data in this semi-supervised manner, one may expect to acquire stronger image features and achieve further performance improvement.

The remainder of this work is organized as follows. In Section 2, we briefly review the related works. Section 3 describes our proposed approach in detail. Section 4 demonstrates the experimental results and comparisons. Finally, this work is summarized in Section 5.

2. Related Work

In this section, we provide an overview of the relevant literature on breast cancer detection and self-supervised contrastive learning.

2.1. Breast Cancer Detection

The primary and most widely diagnosed cause of women's cancer is breast cancer. Early breast cancer diagnosis and identification can significantly decrease the mortality rate. The microscopic inspection of lymph nodes close to the breast is a core component in breast cancer staging [10]. However, this procedure requires extremely qualified pathologists, and it takes quite a lot of time, especially for tiny lymph nodes. Therefore, Computer Aided Diagnosis has thus been established to improve the consistency, efficiency, and sensitivity of metastasis identification [11].

In earlier years, most designed approaches employed hand-crafted features. Spanhol et al. demonstrate classification performance based on several hand-made textural features for distinguishing malignant from benign [4]. Some works merged two or more hand-made features to enhance the accuracy of detection. In [5], graph, haralick, Local Binary Patterns (LBP), and intensity features were used for cancer identification in H&E stained histopathological images. The histopathological images were represented via fusing color histograms, LBP, SIFT, and some efficient kernel features, and the significance of these pattern features was also studied in [6]. However, it needs considerable efforts to design and validate these hand-made features. In addition, the properties of tissues with great variations in morphologies and textures cannot properly be represented, and consequently their detection performance is poor.

With the emergence of powerful computers, deep learning technology has made remarkable progress in a variety of domains, including natural language understanding, speech recognition, computer vision and image processing [12]. These methods have also been successfully employed in various modalities of medical images for detection, classification, and segmentation tasks [13]. Bejnordi et al. built a deep learning system to determine the stromal features of breast tissues associated with tumor for classifying Whole Slide Images (WSIs) [14]. Spanhol et al. utilized AlexNet to categorize breast cancer in histopathological images to be malignant and benign [7]. Bayramoglu et al. developed two distinct CNN architectures to classify breast cancer of pathology images [8]. Single-task CNN was applied to identify malignant tumors. Multi-task CNN has been used for analyzing the properties of benign and malignant tumors. The hybrid CNN unit designed by Guo et al. could fully exploit the global and local features of image, and thus obtain superior prediction performance [9]. Lin et al. proposed a dense and fast screening architecture (ScanNet) to identify metastatic breast cancer in WSIs [15,16]. In order to fully capture the spatial structure information between adjacent patches, Zanjani et al. [17,18] applied the conditional random field (CRF), whereas Kong et al. [19] employed 2D Long Short-Term Memory (LSTM) on patch features, respectively, which are first obtained from a CNN classifier. As the limited number of training samples in medical applications may be insufficient to learn a powerful model, some methods [20–22] transferred deep and rich

feature hierarchies learned from a large number of cross-domain images, for which training data could be easily acquired.

2.2. Self-Supervised Contrastive Learning

Self-supervised learning is a new unsupervised learning paradigm. Recent research has shown that, by minimizing a suitable unsupervised loss during training, self-supervised learning can obtain valuable representations from unlabeled data [23–26]. The resulting network is a valid initialization for subsequent tasks.

The current revival of self-supervised learning started with intentionally devised annotation-free pretext tasks, such as colorization [27], jigsaw puzzle solving [25], relative patch prediction [23], and rotation prediction [26,28]. Although more complex networks and longer training time can yield good results [29], these pretext tasks more or less depend on ad-hoc heuristics, limiting the generality of learnt representations.

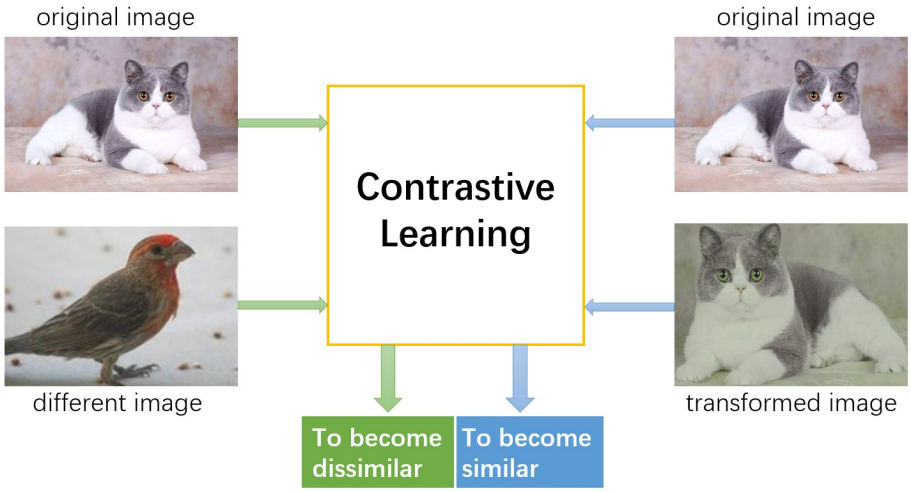

Figure 2. The core idea of contrastive learning: pushing the representations of original and transformed images closer together while separating the representations of original and different images far apart from each other.

Contrastive learning is a discriminative technique that uses contrastive loss [30] to group similar instances closer together and dissimilar instances far apart from each other [31–35], as indicated in Figure 2. Similarity is defined in an unsupervised manner. It is usually considered that various transformations of an image are similar [36]. Ref. [37] employed domain-specific knowledge of videos to model the global contrastive loss. Authors in [38–40] maximized mutual information (MI) between global and local features from different layers of an encoder network, which is comparable to contrastive loss in implementation [41]. Some works utilized memory bank [31] or momentum contrast [32,33] to obtain more negative samples in each batch. MoCo [32], SimCLR [35], and SwAV [42] with modified algorithms generated similar performance with the state-of-the-art supervised method on the ImageNet dataset [43].

We propose in this work to incorporate self-supervision into the detection target in a multi-task setting. Our aim is to enhance the main task's performance through the supervision signal of the accessional task. Self-supervision as an accessional task will contribute to performance improvement.

3. Methodology

Our proposed method is introduced in this section. Figure 1 displays the overall framework of this method. The details of each component are presented in the following subsections.

3.1. Patch Representation with CNN

As the size of Whole-Slide Image (WSI) is extremely huge, smaller image patches (for example, 256 pixels × 256 pixels) are first extracted from WSIs, then deep Convolutional Neural Network (CNN) is utilized to learn effective feature representation of these image patches. Convolutional neural network $F(\bullet)$ generally consists of convolution, spatial pooling, and nonlinear activation layers, which may map an image patch x to a vector with a given length $h \in \mathbb{R}^d$, i.e.,

$$h = F(\Phi; x) \tag{1}$$

where Φ is the parameter of network $F(\bullet)$.

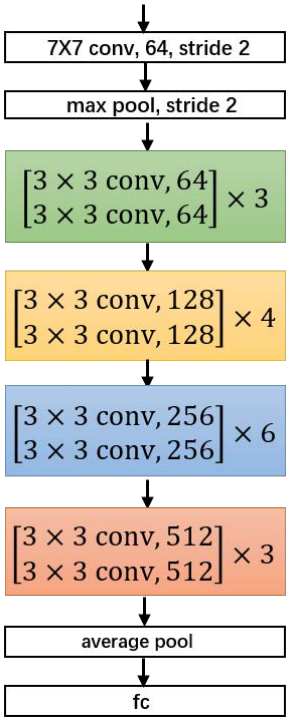

Figure 3. Flow chart of ResNet-34.

Our architecture is compatible with a wide range of network structures with no restrictions. We choose the widely used ResNet [44] as a feature extractor due to its balance between learning capability and network scale. The activation after the average pooling layer is employed as the feature representation of image patch. Figure 3 displays the flow chart of ResNet-34.

3.2. Explored Contrastive Learning Model

Our work is based on SimCLR [35], which has been shown to achieve state-of-the-art performance. SimCLR not only outperforms previous work, but it is also simpler, as it does not require specialized architectures [38,39] or a memory bank [31–34].

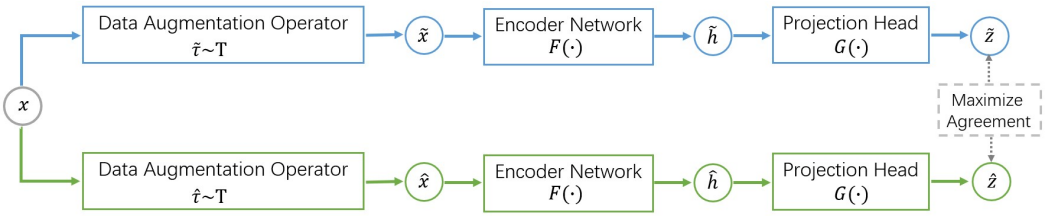

Figure 4. Contrastive learning pipeline for self-supervised training.

SimCLR learns representations in the latent space by maximizing agreement over different data augmentations of the same sample using contrastive loss. As seen in Figure 4, the SimCLR model is composed of four primary components:

(1) A stochastic data augmentation module transforms each given sample randomly to produce two associated versions of the sample, named $\hat{x}$ and $\widetilde{x}$, which are regarded as a similar pair. In this work, we employ three augmentations serially: random cropping followed by rescaling to the original size, random color distortions, and random Gaussian blur.
(2) A encoder network $F(\bullet)$ generates representation vectors of the transformed samples.
(3) A projection head $G(\bullet)$, typically a shallow Multilayer Perceptron (MLP) with one hidden layer, projects representation vectors to a latent space, in which a contrastive loss function is designed.
(4) A contrastive loss function is constructed for a contrastive prediction problem.

For a given similar pair of images $(\hat{x}, \widetilde{x})$, the contrastive loss is defined as:

$$l(\hat{x}, \widetilde{x}) = -log\frac{exp(sim(\hat{z}, \widetilde{z})/\epsilon)}{exp(sim(\hat{z}, \widetilde{z})/\epsilon) + \sum_{\bar{x}\in\Psi^-} exp(sim(\hat{z}, G(F(\bar{x})))/\epsilon).} \quad (2)$$

Here, $\hat{z} = G(F(\hat{x}))$, $\widetilde{z} = G(F(\widetilde{x}))$, $\hat{x}$, and $\widetilde{x}$ are two different transformed versions of x, i.e., $\hat{x} = \hat{\tau}(x)$ and $\widetilde{x} = \widetilde{\tau}(x)$, where $\hat{\tau}, \widetilde{\tau} \in T$ are two stochastic data augmentation operators from transformation set T [30,34–36,38]. ϵ is a temperature scaling parameter. The set Ψ^- contains all images that are dissimilar to x, as well as their transformations. Similarity in the representation space is defined by dot product between ℓ_2 normalized vectors, i.e., $sim(a, b) = a^T b/||a||||b||$.

Minimizing the loss $l(\hat{x}, \widetilde{x})$ in Equation (2) improves the consistency of the representations of $\hat{x}$ and $\widetilde{x}$, known as $\hat{z}$ and $\widetilde{z}$, as well as the inconsistency of the representation of x and those dissimilar images.

Based on the defined loss for a similar pair of images in Equation (2), the overall contrastive loss can be defined as:

$$L_{contrastive} = \frac{1}{|\Psi^+|} \sum_{\forall(\hat{x},\widetilde{x})\in\Psi^+} [l(\hat{x}, \widetilde{x}) + l(\widetilde{x}, \hat{x})] \quad (3)$$

where Ψ^+ contains all similar pairs of images built from a given image set X.

3.3. Boosting Detection via Self-Supervision

In order to reduce overfitting problems brought by the limited number of training samples and learn valuable and discriminative image features, we propose to employ the latest advances in self-supervised feature learning to further enhance the existing detection methods. As shown in Figure 1, we propose to boost feature extractor training through an extra self-supervised task in addition to the primary detection task.

We consider two approaches to incorporate self-supervision into optimization objectives: (1) exploiting an accessional loss function based on contrastive learning; and (2) using unlabeled data during training in a semi-supervised manner. These two techniques are described below in more detail.

3.3.1. Optimize with Accessional Contrastive Loss

We incorporate self-supervision into the target task by adding an accessional contrastive loss. More specifically, for the whole training dataset, the optimization objective of the proposed method is defined as:

$$min_{\Phi,\Theta} L_{CE}(\Phi; D_{training}) + \gamma L_{contrastive}(\Phi, \Theta; X_{training}) \quad (4)$$

where $L_{contrastive}(\Phi, \Theta; X_{training})$ denotes self-supervised contrastive loss on training data $X_{training} = \{x|(x, y) \in D_{training}\}$ without considering their class labels.

$L_{contrastive}(\Phi, \Theta; X_{training})$ depends on a feature extractor's parameters Φ and the parameters Θ of a contrastive learning network. L_{CE} stands for the supervised cross-entropy loss. The importance of the self-supervised contrastive loss is controlled by hyperparameter γ.

3.3.2. Optimize with Semi-Supervised Contrastive Loss

The contrastive learning term $L_{contrastive}$ in Equation (4) does not need class labels, so it can be easily extended to learn from additional unlabeled samples as well. Obviously, if unlabeled images set $X_{unlabeled}$ is available in addition to $D_{training}$, we can take full advantage of them in the contrastive learning task by redefining the optimization objective as:

$$min_{\Phi,\Theta} L_{CE}(\Phi; D_{training}) + \gamma L_{contrastive}(\Phi, \Theta; X_{training} \cup X_{unlabeled}) \quad (5)$$

During the training feature extractor, the contrastive loss on additional unlabeled data is also minimized at the same time. Thus, the feature extractor's visual scope is opened up in the hope that this will improve its capacity to handle new classes with limited data. This can be thought of as a semi-supervised training method.

In order to optimize Equations (4) and (5), we employ Stochastic Gradient Descent (SGD) with a mini-batch to train the deep neural network. To be specific, the parameters of network are updated by a back propagation (BP) strategy.

After generating the probability map of WSI at patch level, we apply a non-maxima suppression algorithm [45,46] to obtain the coordinates of cancer metastases, which repeat two steps until, in the heatmap, there exists no value larger than a certain threshold: (1) search the maximum and its corresponding coordinate; (2) all values in the range of a given radius around the maximum are set to 0.

4. Experiments

4.1. Dataset

We made extensive experiments on the Camelyon16 challenge dataset (https://camelyon17.grand-challenge.org/Data/, (accessed on 15 July 2021)), which was obtained from two different institutes: the University Medical Centre Utrecht (Utrecht UMC) and Radboud University Medical Centre (Radboud UMC). These two Medical Centers utilize different digital slide scanners to produce the TIF WSIs. The TIF WSIs from Utrecht UMC were created with a 40× objective lens (level-0 pixel size, 0.226 × 0.226 µm) through a digital slide scanner (NanoZoomerXR Digital slide scanner C12000-01; Hamamatsu Photonics). The TIF WSIs from Radboud UMC were created with a 20× objective lens (level-0 pixel size, 0.243 × 0.243 µm) through a digital slide scanner (Pannoramic 250 Flash II; 3D Histech). The Camelyon16 challenge dataset contains 400 TIF WSIs in total, including 110 tumorous and 160 normal WSIs for training, 50 tumorous and 80 normal WSIs for test. Pathologists have carefully annotated the cancer metastasis locations and regions in the format of binary mask, with a few exceptions reported in [46]. The statistics of this dataset can be seen in Table 1.

Table 1. The statistics of the Camelyon16 dataset.

Sources	Training		Test	
	Tumor	Normal	Tumor	Normal
Utrecht UMC	40	60	50	80
Radbound UMC	70	100		
Total	110	160	50	80

Most WSIs are larger than 60,000 pixels × 100,000 pixels. The size of the whole Camelyon16 dataset is about 700 gigabytes. The WSIs are usually saved in a pyramid structure of multi-resolution, with several down-sampled versions of the source image [47].

As shown in Figure 5, the source image with the highest resolution is labeled as level-0, while other down-sampled images are labeled as level-1 to level-n.

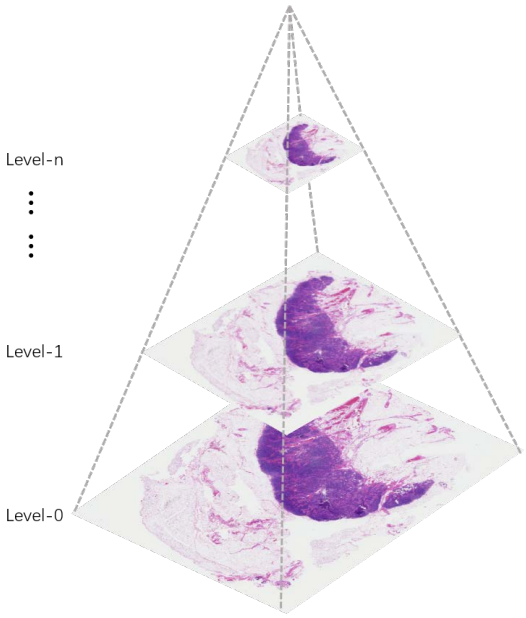

Figure 5. The multi-resolution pyramid structure of the whole-slide image.

4.2. Implementation Details

Given a test WSI, our purpose is to detect whether this slide includes tumors and locate the areas of these tumors. The model is trained with smaller image patches extracted from the slides, due to a huge size and limited number of slides. A patch is labeled to be tumorous if there is at least one pixel marked as a tumor within the patch area. In general, just a tiny part of the slide includes biological tissue of interest, whereas the remaining majority is background and fat. To reduce computation, the Otsu algorithm [48] is applied to remove the background regions of every training slide.

The training model was challenging because of the tumorous class imbalance among the large number of patches. There are 10,000 to 400,000 patches (median 90,000) in each slide. However, only 20 to 150,000 tumorous patches (median 2000) are found in each tumor slide. The proportion of tumorous patches is between 0.01% and 70% (median 2%). We adopt an effective sampling strategy to avoid bias towards slides that include more patches (both normal and tumorous). First, the "normal" or "tumorous" class is selected with the same probability. Second, we select a slide containing this class of patches uniformly at random, and then sample patches from the slide. In comparison, several existing approaches pre-sample a number of patches from each slide [49], which restrict the diversities of training patches.

In the training stage, several strategies of data augmentation were applied. Since pathological slides have no canonical directions, the eight directions are all valid. The input patches are rotated by four multiples of 90°, the left-right flip and rotations are repeated. Color jitter was added using Torchvision transforms with parameters used in [18,46]: hue with a maximum delta of 0.04, saturation with a maximum delta of 0.25, contrast with a maximum delta of 0.75, and brightness with a maximum delta of 64/255. By subtracting 128 and dividing 128, the pixel values of patches were normalized. In order to optimize the whole architecture, we employed a stochastic gradient descent with a 0.001 learning rate and 0.9 momentum. We set $\gamma = 1.0$, $\epsilon = 0.5$ in our experiments. Our approach is implemented using PyTorch-1.6.0 [50] and trained using NVIDIA GeForce GTX 2080 Ti GPU.

In the test stage, a tumorous probability heatmap is generated by performing prediction over patches in a sliding window with a stride of 64 across the whole slide. For every patch, rotations and left-right flip are applied to generate the prediction results in the

eight directions, and the patch-level tumor prediction is finally obtained by averaging the prediction results in the eight directions. The maximum value in the tumorous probability heatmap is taken as a prediction result of each slide.

4.3. Evaluation Metrics

In order to assess the performance of tumor region localization and WSI classification, we employ two Camelyon16 evaluation metrics. The area under receiver operating characteristic (Area Under ROC, AUC) [51] is utilized for the evaluation of slide-level classification performance. The Free Response Operating Characteristic (FROC) curve [52] is utilized for performance evaluation in tumor detection and localization.

The FROC curve is defined as a sensitivity plot versus the average number of false positives per image under various probability truncation values. To be specific, for each heatmap, a list of coordinates and related predictions will be first generated. The maximum prediction value is recorded among all the coordinates within each annotated tumor area. FPs are the number of coordinates that fall outside tumor areas. The FROC score is defined as the average sensitivity at six predefined false positive rates: 1/4, 1/2, 1, 2, 4, and 8 false positives. Higher average FROC score means better detection performance.

In order to produce the points for FROC calculation, the Camelyon16 challenge winner thresholded the heatmap to generate a bitmask, and recorded a single prediction value for every connected component in the bitmask. Instead, based on a certain probability map, we employ the non-maxima suppression algorithm [45,46] to obtain cancer metastasis coordinates.

4.4. Results

4.4.1. Self-Supervision as Accessional Contrastive Loss

We first evaluate the effectiveness of the additional self-supervised contrastive loss and report results in Table 2. Compared with traditional ResNet, the trained network in the method "ResNet with accessional contrastive loss" not only includes original ResNet but also contains an extra contrastive learning task. Contrastive learning is a paradigm of unsupervised learning and does not need class labels. In the contrastive learning task, our employed SimCLR model is trained by maximizing agreement between different data augmentations of the same data sample through contrastive loss in the latent space. The self-supervised contrastive learning task as an auxiliary task provides a surrogate supervision signal for feature learning and enables feature extractors to learn more discriminative visual representations. It is demonstrated in Table 2 that adding self-supervision improves the detection performance, and that the gain in performance is especially significant for a larger-scale network structure, such as ResNet-34.

Table 2. Performance evaluation of accessional contrastive loss.

Approaches	FROC Score
ResNet-18	0.7814
ResNet-18 with accessional contrastive loss	**0.7986**
ResNet-34	0.7463
ResNet-34 with accessional contrastive loss	**0.7721**

4.4.2. Self-Supervision as Semi-Supervised Contrastive Loss

Then, the proposed semi-supervised training strategy is evaluated using unlabeled WSIs in the Camelyon16 test dataset. We make comparisons "ResNet with semi-supervised contrastive loss" method with the following two methods: "original ResNet backbone without self-supervision" and "ResNet with accessional contrastive loss", which is ResNet backbone with self-supervision but no access to unlabeled images. Compared with the "ResNet with accessional contrastive loss" method, in the "ResNet with semi-supervised contrastive loss" method, the auxiliary contrastive learning task is trained using both training data and

extra unlabeled test data in the Camelyon 2016 Grand Challenge dataset. Through increasing the number and diversity of training data in this semi-supervised manner, the "ResNet with semi-supervised contrastive loss" method can learn richer and stronger image features. The results in Table 3 indicate that that our proposed semi-supervised strategy does actually enhance detection performance by exploiting unlabeled images.

Table 3. Performance evaluation of semi-supervised contrastive loss.

Approaches	FROC Score
ResNet-18	0.7814
ResNet-18 with accessional contrastive loss	0.7986
ResNet-18 with semi-supervised contrastive loss	**0.8124**
ResNet-34	0.7463
ResNet-34 with accessional contrastive loss	0.7721
ResNet-34 with semi-supervised contrastive loss	**0.8013**

4.4.3. Comparison with Prior Works

In Table 4, we compare our approach with several prior works, which were submitted by different top institutions and organizations to the Camelyon16 challenge. For our approach, we use ResNet-18 with semi-supervised contrastive loss, which previously produces the best results. We observe from Table 4 that our approach is superior to other methods in both tumor localization and WSI classification tasks. It should be noted that detection performance of our approach outperforms that from pathologists by about 8%, which highlights our approach's capability to detect metastasis in lymph node biopsies. It is also indicated that our approach not only produces an objective solution, but also achieves superior localization results. In the training process, our method is composed of two tasks: the supervised task and additional contrastive learning task. The supervised task utilizes all the training data of the Camelyon16 challenge dataset (110 tumorous and 160 normal WSIs), which is the same as other compared studies. In the semi-supervised learning manner, the contrastive learning task employs the whole Camelyon16 challenge dataset (including all the training and test data). Since contrastive learning is a paradigm of unsupervised learning, it does not need class labels of the data. In a word, during the training stage, our method utilizes the whole Camelyon16 challenge dataset, including 270 labeled training data and 130 unlabeled test data, whereas other compared studies only use 270 labeled data from training data of the Camelyon16 challenge dataset. With the same number of labeled data, our method achieves further performance improvement via additional unlabeled data in the contrastive learning task, in comparison with other studies. By increasing the size and variety of training data, this task augmentation enables the model to learn richer and more discriminative image features. In addition, the self-supervised task attempts to reduce deep learning methods' dependence on large amounts of labeled data.

Table 4. Performance comparison on the Camelyon16 dataset.

Approaches	FROC Score	AUC Score
Human performance	0.7325	0.9660
Radboud Uni.(DIAG)	0.5748	0.7786
Middle East Tech. Uni.	0.3889	0.8642
HMS, Gordan Center, MGH	0.7600	0.9763
NLP LOGIX Co. USA	0.3859	0.8298
EXB Research Co.	0.5111	0.9156
DeepCare Inc.	0.2439	0.8833
University of Toronto	0.3822	0.8149
Ours	**0.8124**	**0.9857**

We also display the detection probability result for test_071 in order to emphasize the excellent performance of our approach. It can be observed from Figure 6 that our approach generates a detection probability map with better visual quality, and that metastasis detection result in the third column is in great agreement with the ground truth annotations generated by expert pathologists in the second column. It is obvious that our approach can detect metastases very well within the whole-slide images.

(**a**) (**b**) (**c**)

Figure 6. Predicted probability map for Test_071 by our approach. (**a**) original image test_071. (**b**) ground truth annotation. (**c**) predicted probability map.

4.5. Discussion

The comparison of various self-supervised methods is difficult because it requires another base task that utilizes the learned representations [23,29]. Considering that our framework enables simple combination of the detection approach and any kind of self-supervised learning algorithm, we propose that our framework can be used as an alternative method to assess the performance of various self-supervised algorithms. To achieve that, the only change we need to make for our framework is to use the specific self-supervised loss. The ultimate detection performance could then be utilized to evaluate the self-supervised method. We leave a more comprehensive and detailed comparison of self-learned representations for future work.

5. Conclusions

The need for a large number of annotation images to achieve superior performance using deep learning methods is still a major challenge in medical image analysis. In order to reduce the requirement for a large annotation dataset, we propose to include an extra self-supervised contrastive loss during detection model training. In addition, we extend the proposed method to a semi-supervised setting and obtain further performance gain because self-supervised loss does not need labeled data at all. We also investigate the effectiveness of the additional self-supervised loss through conducting extensive quantitative experiments on the Camelyon2016 Grand Challenge dataset. The experimental results indicate that including self-supervision improves detection performance significantly. Finally, we expect that our approach will be applied to other kinds of medical image analysis tasks, especially in less-annotation clinical conditions.

Author Contributions: H.Z. investigated the ideas, designed the method, and wrote the manuscript; Y.Z. provided the suggestions on the experimental setup and analytical results; X.H. wrote the manuscript and provided funding supports. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China under Grant No. 62002255, Shanxi Scholarship Council of China No. 2021-038, and the Applied Basic Research Project of Shanxi Province No. 20210302123130.

Institutional Review Board Statement: The study did not involve humans or animals.

Informed Consent Statement: The study did not involve humans.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: (https://camelyon17.grand-challenge.org/Data/) (accessed on 15 July 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Ramos-Vara, J.A. Principles and methods of immunohistochemistry. *Methods Mol. Biol.* **2011**, *691*, 83–96. [PubMed]
3. Humphreys, G.; Ghent, A. World laments loss of pathology service. *Bull. World Health Organ.* **2010**, *88*, 564–565. [PubMed]
4. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 1455–1462. [CrossRef] [PubMed]
5. Cruz-Roa, A.A.; Ovalle, J.; Madabhushi, A.; Osorio, F. A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. In Proceedings of the 16th International Conference on Medical Image Computing and Computer Assisted Intervention, Nagoya, Japan, 22–26 September 2013; pp. 403–410.
6. Kandemir, M.; Hamprecht, F.A. Computer-aided diagnosis from weak supervision: A benchmarking study. *Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc.* **2015**, *42*, 44–50. [CrossRef] [PubMed]
7. Spanhol, F.; Oliveira, L.S.; Cavalin, P.R.; Petitjean, C.; Heutte, L. Deep features for breast cancer histopathological image classification. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Banff, AB, Canada, 5–8 October 2017; pp. 1868–1873.
8. Bayramoglu, N.; Kannala, J.; Heikkilä, J. Deep learning for magnification independent breast cancer histopathology image classification. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2440–2445.
9. Guo, Y.; Dong, H.; Song, F.; Zhu, C.; Liu, J. Breast Cancer Histology Image Classification Based on Deep Neural Networks. In *International Conference Image Analysis and Recognition*; Springer: Cham, Switzerland, 2018; Volume 10882, pp. 827–836.
10. Apple, S.K. Sentinel Lymph Node in Breast Cancer: Review Article from a Pathologist's Point of View. *J. Pathol. Transl. Med.* **2016**, *50*, 83–95. [CrossRef]
11. Litjens, G.; Sánchez, C.; Timofeeva, N.; Hermsen, M.; Nagtegaal, I.; Kovacs, I.; Kaa, H.; Bult, P.; Ginnneken, B.V.; Laak, J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **2016**, *6*, 26286. [CrossRef]
12. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Essen, B.C.V.; Awwal, A.A.S.; Asari, V.K. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *arXiv* **2018**, arXiv:1803.01164.
13. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; der Laak, J.A.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
14. Ehteshami Bejnordi, B.; Linz, J.; Glass, B.; Mullooly, M.; Gierach, G.; Sherman, M.; Karssemeijer, N.; van der Laak, J.; Beck, A. Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images. In Proceedings of the IEEE 14th International Symposium on Biomedical Imaging, Melbourne, VIC, Australia, 18–21 April 2017; pp. 929–932.
15. Lin, H.; Chen, H.; Dou, Q.; Wang, L.; Qin, J.; Heng, P.A. ScanNet: A Fast and Dense Scanning Framework for Metastatic Breast Cancer Detection from Whole-Slide Images. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 539–546.
16. Lin, H.; Chen, H.; Graham, S.; Dou, Q.; Rajpoot, N.; Heng, P.A. Fast ScanNet: Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection. *IEEE Trans. Med. Imaging* **2019**, *38*, 1948–1958. [CrossRef]
17. Zanjani, F.G.; Zinger, S.; With, P. Cancer detection in histopathology whole-slide images using conditional random fields on deep embedded spaces. In Proceedings of the Digital Pathology, Houston, TX, USA, 6 March 2018.
18. Yi, L.; Wei, P. Cancer Metastasis Detection with Neural Conditional Random Field. *arXiv* **2018**, arXiv:1806.07064.
19. Kong, B.; Xin, W.; Li, Z.; Qi, S.; Zhang, S. Cancer Metastasis Detection via Spatially Structured Deep Network. In *International Conference Image Analysis and Recognition*; Springer: Cham, Switzerland, 2017; pp. 236–248.
20. Xie, J.; Liu, R.; Luttrell, J.; Zhang, C. Deep Learning Based Analysis of Histopathological Images of Breast Cancer. *Front. Genet.* **2019**, *10*, 80. [CrossRef] [PubMed]
21. de Matos, J.; de Souza Britto, A.; Oliveira, L.; Koerich, A.L. Double Transfer Learning for Breast Cancer Histopathologic Image Classification. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
22. Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Breast Cancer Diagnosis with Transfer Learning and Global Pooling. In Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 16–18 October 2019; pp. 519–524.
23. Doersch, C.; Gupta, A.; Efros, A.A. Unsupervised Visual Representation Learning by Context Prediction. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1422–1430.
24. Pathak, D.; Krähenbühl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
25. Noroozi, M.; Favaro, P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. In Proceedings of the ECCV, Amsterdam, The Netherlands, 11–14 October 2016.
26. Gidaris, S.; Singh, P.; Komodakis, N. Unsupervised Representation Learning by Predicting Image Rotations. *arXiv* **2018**, arXiv:1803.07728.
27. Zhang, R.; Isola, P.; Efros, A.A. Colorful Image Colorization. In Proceedings of the ECCV, Amsterdam, The Netherlands, 11–14 October 2016.

28. Chen, T.; Zhai, X.; Ritter, M.; Lucic, M.; Houlsby, N. Self-Supervised GANs via Auxiliary Rotation Loss. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 12146–12155.
29. Kolesnikov, A.; Zhai, X.; Beyer, L. Revisiting Self-Supervised Visual Representation Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 1920–1929.
30. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1735–1742. [CrossRef]
31. Wu, Z.; Xiong, Y.; Yu, S.X.; Lin, D. Unsupervised Feature Learning via Non-parametric Instance Discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3733–3742.
32. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R.B. Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9726–9735.
33. Misra, I.; van der Maaten, L. Self-Supervised Learning of Pretext-Invariant Representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 6706–6716.
34. Tian, Y.; Krishnan, D.; Isola, P. Contrastive Multiview Coding. In Proceedings of the ECCV, Glasgow, UK, 23–28 August 2020.
35. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G.E. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv* **2020**, arXiv:2002.05709.
36. Dosovitskiy, A.; Springenberg, J.T.; Riedmiller, M.; Brox, T. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS), Montreal, QC, Canada, 8–13 December 2014.
37. Tschannen, M.; Djolonga, J.; Ritter, M.; Mahendran, A.; Houlsby, N.; Gelly, S.; Lucic, M. Self-Supervised Learning of Video-Induced Visual Invariances. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 13803–13812.
38. Bachman, P.; Hjelm, R.D.; Buchwalter, W. Learning Representations by Maximizing Mutual Information Across Views. In Proceedings of the NeurIPS, Vancouver, BC, Canada, 8–14 December 2019.
39. Hénaff, O.J.; Srinivas, A.; Fauw, J.D.; Razavi, A.; Doersch, C.; Eslami, S.M.A.; van den Oord, A. Data-Efficient Image Recognition with Contrastive Predictive Coding. *arXiv* **2020**, arXiv:1905.09272.
40. Hjelm, R.D.; Fedorov, A.; Lavoie-Marchildon, S.; Grewal, K.; Trischler, A.; Bengio, Y. Learning deep representations by mutual information estimation and maximization. *arXiv* **2019**, arXiv:1808.06670.
41. Tschannen, M.; Djolonga, J.; Rubenstein, P.K.; Gelly, S.; Lucic, M. On Mutual Information Maximization for Representation Learning. *arXiv* **2019**, arXiv:1907.13625.
42. Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; Joulin, A. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. *arXiv* **2020**, arXiv:2006.09882.
43. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, F.F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NA, USA, 27–30 June 2016.
45. Ciresan, D.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* **2013**, *16 Pt 2*, 411–418.
46. Liu, Y.; Gadepalli, K.; Norouzi, M.; Dahl, G.E.; Kohlberger, T.; Boyko, A.; Venugopalan, S.; Timofeev, A.; Nelson, P.Q.; Corrado, G.S.; et al. Detecting Cancer Metastases on Gigapixel Pathology Images. *arXiv* **2017**, arXiv:1703.02442.
47. Goode, A.; Gilbert, B.; Harkes, J.; Jukie, D.; Satyanarayanan, M. Openslide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Informatics* **2013**, *4*, 27. [CrossRef] [PubMed]
48. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]
49. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A.H. Deep Learning for Identifying Metastatic Breast Cancer. *arXiv* **2016**, arXiv:1606.05718.
50. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; Devito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic Differentiation in PyTorch. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017) Autodiff Workshop, Long Beach, CA, USA, 9 December 2017.
51. Hanley, J.A.; Mcneil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [CrossRef] [PubMed]
52. Chakraborty, D.P. Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. *Med. Phys.* **1989**, *16*, 561–568. [CrossRef]