I.J. Modern Education and Computer Science, **2019, 6, 1-4**
Published Online June 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2019.06.01

A Review of Self-supervised Learning Methods in the Field of Medical Image Analysis

Jiashu Xu
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, 03056, Ukraine
Email: jiashu.xu.1@gmail.com

Received: 06 April 2021; Accepted: 20 May 2021; Published: 08 August 2021

Abstract: In the field of medical image analysis, supervised deep learning strategies have achieved significant development, while these methods rely on large labeled datasets. Self-Supervised learning (SSL) provides a new strategy to pre-train a neural network with unlabeled data. This is a new unsupervised learning paradigm that has achieved significant breakthroughs in recent years. So, more and more researchers are trying to utilize SSL methods for medical image analysis, to meet the challenge of assembling large medical datasets. To our knowledge, so far there still a shortage of reviews of self-supervised learning methods in the field of medical image analysis, our work of this article aims to fill this gap and comprehensively review the application of self-supervised learning in the medical field. This article provides the latest and most detailed overview of self-supervised learning in the medical field and promotes the development of unsupervised learning in the field of medical imaging. These methods are divided into three categories: context-based, generation-based, and contrast-based, and then show the pros and cons of each category and evaluates their performance in downstream tasks. Finally, we conclude with the limitations of the current methods and discussed the future direction.

Index Terms: Medical image analysis, Self-Supervised learning, Unsupervised learning, Visual feature learning, Contrastive Learning.

1. Introduction

Self-supervised learning was first introduced in robotics, in which the training data was automatically labeled by utilizing the relationship between different sensor signals. Then, Deep learning borrowed this idea, Unsupervised Natural Language Processing tasks make significant development also benefited from this idea. Recent years, in Computer Vision tasks, new SSL frameworks have seen a blossoming phenomenon. The paradigm of SSL algorithms is conducive to medical image analysis tasks, there is a well-known phenomenon, labeled medical datasets are difficult to assemble, and the time cost and labor cost of manual labeling are enormous, while unlabeled medical data is numerous. SSL builds proxy tasks to perform representation learning from large-scale unlabeled data, through this process can improve downstream task performance. Therefore, the application of self-supervised learning methods in the field of medical image analysis is of great significance.

However, there is a big difference between medical images and natural images. How to reasonably use the existing SSL framework to solve the task of medical image analysis is the main research problem. This article reviews the self-supervised learning methods applied in the field of medical image analysis, and aims to evaluate existing methods, find new research directions, and provide help for subsequent researchers. Through the systematic analysis of the most cutting-edge articles, found that the application of self-supervised learning in the medical field still has great application potential. According to the characteristics of the proxy task of these self-supervised learning methods, we mainly divide them into three categories: context-based, generation-based, and contrast-based. These methods will be specifically introduced in Sections 2, 3, and 4. In Section 5, Integrate the experimental results of downstream tasks performed on medical image datasets. Comprehensive evaluation of these existing SSL methods.

2. Context-based Self-supervised Learning

In the self-supervised learning framework, the proxy tasks based on contextual semantics are usually geometric transformation prediction, flipping, rotation angle prediction [1], Jigsaw Puzzles [2], etc. Some related research demonstrated that these methods are more effective than transfer learning (from nature images domain) in specific medical image analysis tasks [30].

2.1. Predicting Rotation

The pretext-task of predicting the rotation angle is to construct a pre-training dataset by randomly rotating the image of the raw data, through this way pre-train the network. This process enables the network to understand the relationship between spatial features within the image. A. Hatamizadeh.et al. used this method as a pretext-task to pre-train the network architecture based on V-net [5] and Inception-ResNet V2 [4]. Then the pre-trained model is used for Lung lobe segmentation tasks and DR classification tasks [3]. Their experimental results suggest that their method is improved compared to the pre-trained model based on the ImageNet dataset. Imran, A. A. Z et al. proposed an SSL-based multi-task learning model (S^4 MTL) [6], use geometric transformation function t(x) as a proxy task in the framework to generate pseudo-label as supervision signals. In general, the predicting rotation is a simple pretext-task, compared with the model trained from scratch, this pre-training model converges faster, but the model performance improvement is limited.

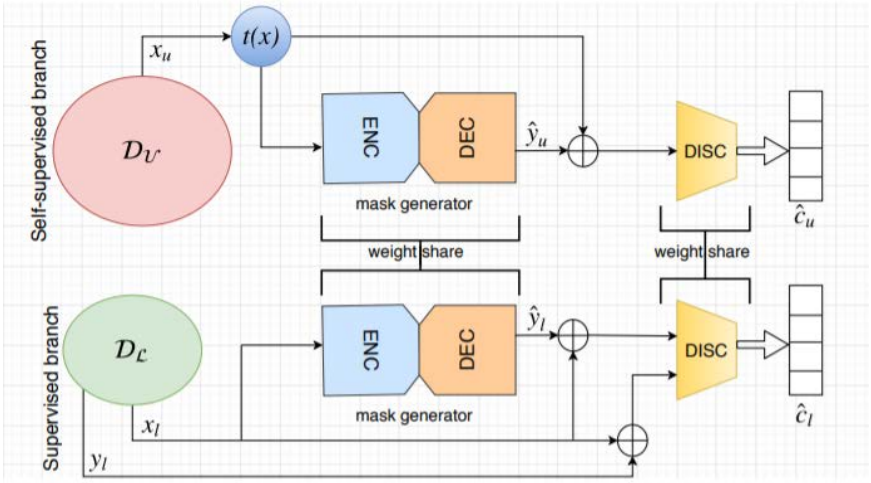

Fig. 1. multi-task learning model (S^4 MTL) [6].

Therefore, some researchers will combine multiple pretest-tasks to improve SSL framework performance. For instance, a new SSL framework consists of the prediction of geometric transformations and the reconstruction of CT scans [7]. Through reconstruction, the network can identify abnormal information at the pixel level. By predicting the geometric transformation of the CT scans, the network can capture the semantic information of the global context. This new framework has made a significant improvement for anomaly detection in brain CT scans. Since the rotation prediction is performed on 2D images, some researchers try the 3D rotation prediction [9]. In the 3D rotation prediction pretext-task, the input 3D image is randomly rotated by different degree {0°, 90°, 180°, 270°} in the 3D coordinate system (x, y, z). In Fig.2. Each axis has four rotation angles, so there are a total of 12 possible rotations. Whatever which coordinate axis is rotated, when the rotation angle is zero, it is the same as the original image, so this is a 10-way classification problem.

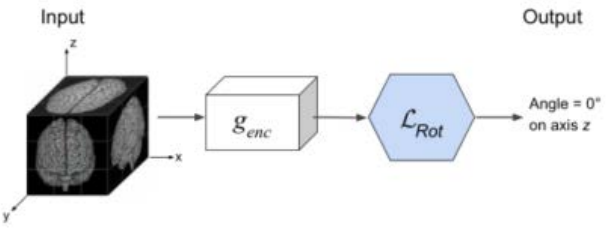

Fig. 2. Predict the 3D image rotation degree [9].

2.2 Jigsaw Puzzles

The jigsaw puzzle is another pretext-task based on context spatial semantic. Usually, this method to recognize the order of the shuffled sequence of patches from the same image. Inspired by Jigsaw puzzles Y. Li, et al. [8] increased the complexity of the pretext-task, on the basis of shuffled the order of the patches, also randomly rotated each patch by {0°, 90°, 180°, 270°}, as shown in Fig.3. These shuffled patches as the input of the network then trained to recognize the right order, through the process the network can capture high-level semantic information.

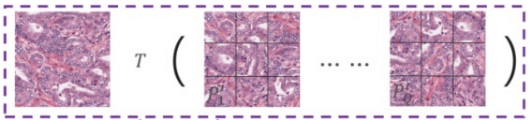

Fig. 3. Jigsaw puzzles for histopathological images

In inspired by the 2D Jigsaw puzzle, Aiham et al. proposed 3D Jigsaw puzzle method as SSL proxy task [9].

Similar to the 2D method, it is an n-way classification task, aims to find the right sequence index of the shuffled 3D patches.

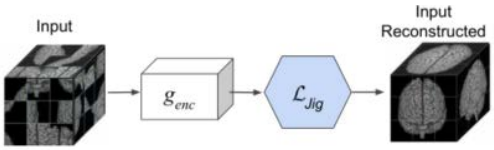

Fig. 4. 3D Jigsaw puzzles [9]

2.3. Rubik's Cube

Rubik's Cube can be regarded as a derivative of the 3D Jigsaw puzzle. X. Zhuang et al. first proposed playing Rubik's cube as a proxy task for volumetric medical data [10]. Divide the 3D medical image into a set (2×2×2) of cubes, and then randomly rotate cubes, the proxy task aims to recover the original 3D image, cubes rearrangement, and cubes rotation. Compared with 3D Jigsaw puzzles, this method adds Cube's rotation operations. Basis on this idea, Rubik's Cube + [12] and Rubik's Cube ++ [11] methods have been developed. The pipeline of the Rubik's Cube is illustrated in Fig.5. This method is more conducive to the contextual feature extraction of 3D unlabeled medical images.

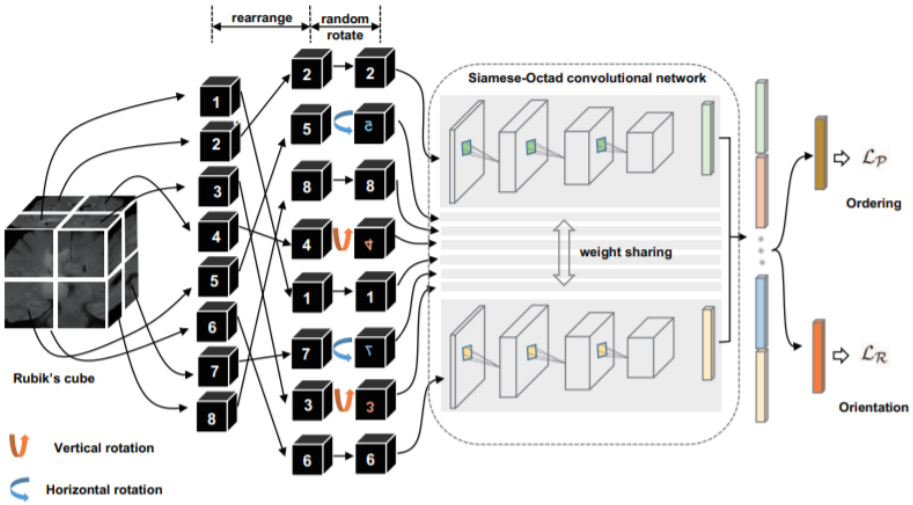

Fig. 5. 3D Rubik's Cube [10]

2.4 Specific Pretext-Tasks Based On The Context Of The Medical Image Domain

Different from the previous context-based methods, Li Sun et al. [13] focuses on the context information of the anatomical region of medical images and proposed a two-level feature learning method, first learning local textural features from the regional anatomical level, and then learning global contextual features from the patient level. Furthermore, in Bai's work use the characteristics of different Cardiac MR view planes to build anatomical position prediction pretext-task [14]. These ideas got rid of generic SSL methods and developed proxy tasks utilize the specific context information of medical images.

3. Generated-Based Self-supervised Learning

In this section, we will introduce generation-based SSL methods in the medical image domain, including autoencoder models [19] and GANs [20].

3.1 Restore Based On The Autoencoder Model

The purpose of this SSL pretext-task is to recover the original sub-volume after its transformation (sub-volume, cropped from unlabeled CT images by random size and random location). The original sub-volume x_i through the transformation function $f(\cdot)$ obtain $\tilde{x}_i = f(x_i)$. Take $\tilde{x}_i$ as the input of the autoencoder to restore the original sub-volume. The restoration pretext-tasks optimizer function is expressed as:

$$\min_{\theta_E \theta_G} L_{res} = \sum_{x_i \in D} \| G\left(E\left(\tilde{x}_i\right)\right) - x_i \|_2 \tag{1}$$

Where G and E are the encoder and decoder. θ_E *and* θ_G are the parameters of the autoencoder. Here the transformations can be a non-linear transformation, local-shuffling, out-painting, in-painting. In the Models Genesis [15,18] SSL framework integrates a variety of transformations and shared the same autoencoder to restore the x_i. This allows the model to learn latent representations from multiple perspectives and then transfer them to specific downstream tasks. Inspired by Models Genesis, subsequent researchers proposed Semantic Genesis [16] and Universal Model [17]. Semantic Genesis further extends the framework of Models Genesis to add self-classification branch, using less computational overhead, actually a classification head is added at the end of the encoder in the autoencoder. Semantic Genesis framework extracts semantics representation from the consistent and recurrent anatomical patterns. The pipeline of the Models Genesis and Semantic Genesis are illustrated in Fig.6, 7.

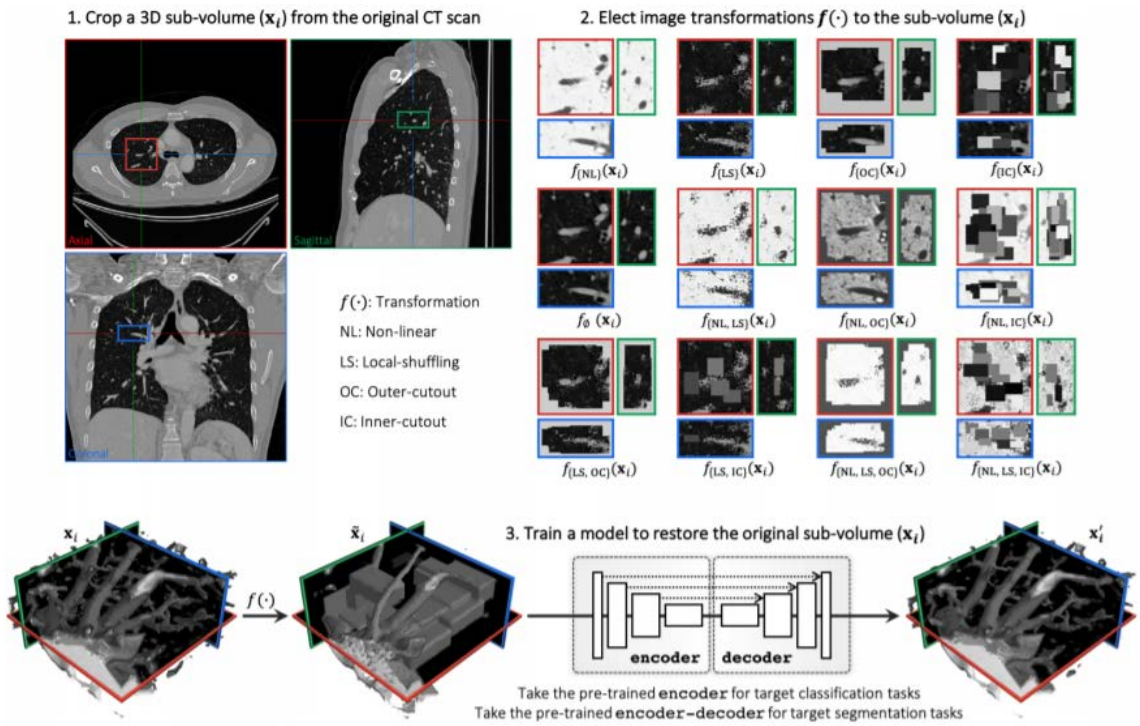

Fig. 6. The pipeline of the Models Genesis[15].

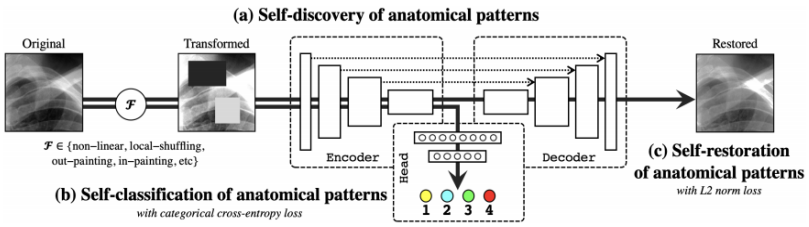

Fig. 7. The pipeline of the Semantic Genesis[16].

Universal Model adds Modality Invariant Representation Learning tasks and Multi-level Feature Learning classification tasks compare to the pipeline of the Models Genesis. Therefore, this method is dedicated to solving the generalization problem of multi-task and multi-modality in the medical image analysis domain. The pipeline of the Universal Model is illustrated in Fig.8.

3.2 Based on GANs

As the most successful generative model in recent years, GAN is widely employed in computer vision tasks. Ross et al. proposed a re-colorization proxy task that utilizes GAN to recolor medical images (endoscopic video data) [21]. First, the raw data is transformed into the CIELAB Color space, in CIELAB Color space the L-channel as input, train the GAN to generate the a-channels and b-channels. Then transfer the generator network, which has the ability to extract the low-level semantic information, to the target segmentation task. According to the variants of GAN, many proxy tasks can be constructed, for instance the reconstruction of patches using Wasserstein GAN [22], use conditional GAN [23] for image colorization [3], and self-supervised CycleGAN framework for ultrasound image super-resolution [24]. Compared with previous pretext-tasks, the framework of GANs-based is more complicated, the self-supervised CycleGAN framework is illustrated in Fig.9. it also means longer training time costs[58].

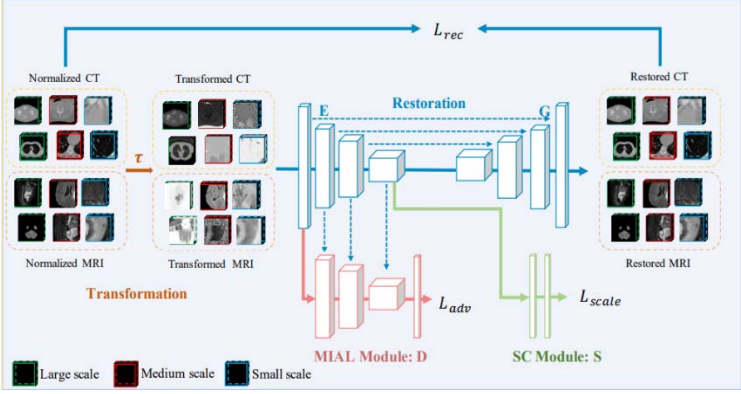

Fig. 8. The pipeline of the Universal Model [17].

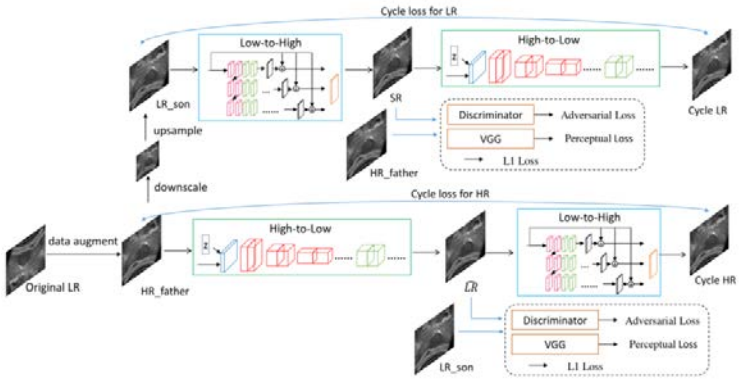

Fig. 9. The self-supervised CycleGAN framework

4. Contrast-Based Self-supervised Learning

Contrast learning as a form of self-supervised learning has been the front-runner in the unsupervised natural image domain, and these frameworks based on comparative learning constantly narrow the gap between unsupervised and supervised learning, for instance, CPC [32], SimCLR [25], MoCo [26], BYOL [27], SimSiam [28], Twins [29]. The basic idea of comparative learning is that different transformations of a sample image have similar representations and these representations should be different from the different sample images, then, use unlabeled data to train the neural network by minimizing the contrast loss [31]. However, its application in the medical imaging domain is still in infancy. This section will introduce the application in the field of medical image analysis based on different contrastive learning methods, here we divide those CL methods into 2 types: global-local contrast and context-context contrast.

4.1 Global-local Contrast

Global-local Contrast is aims to modeling the relationship between local semantic representation and global semantic representation of an instance. This method focuses on the prediction of the encoder-decoder architectures at the pixel level and learns the global and local level representations while considering the local representation. For medical images, the data of different patients have inherent consistency, because the structure of human organs is consistent, such as MRI (magnetic resonance imaging) and CT (computed tomography). Therefore, encoding the same anatomical region of the medical images will get the similar embedding, leverage this characteristic to construct the global contrastive loss [33]. In contrast, Local representation aims to distinguishing pixel-level differences in neighboring regions. In [33,35], by constructing InfoNCE loss [34], the network determines whether the output comes

from the similar distribution or the dissimilar distribution to define the global contrastive loss and the local contrastive loss. InfoNCE loss is defined as:

$$L_N = -E_x\left[\log\frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right)+\sum_{J=1}^{N-1}\exp\left(f(x)^T f(x_j)\right)}\right] \tag{2}$$

Where x is from a set $X = \{x_1 \dots x_n\}$, x^+ are transformed sample. Minimizing the loss narrows the distance between the representations of x and x^+, while increasing the distance between the representation of x and other dissimilar images.

The Global-local Contrast learning framework is illustrated in Fig.10. The input volumetric images divided into four partitions, leverage the similarity of the slices in different volumetric images defining positive samples and negative samples. Representations z are extracted by encoder $e(\cdot)$ and projection head $g(\cdot)$.

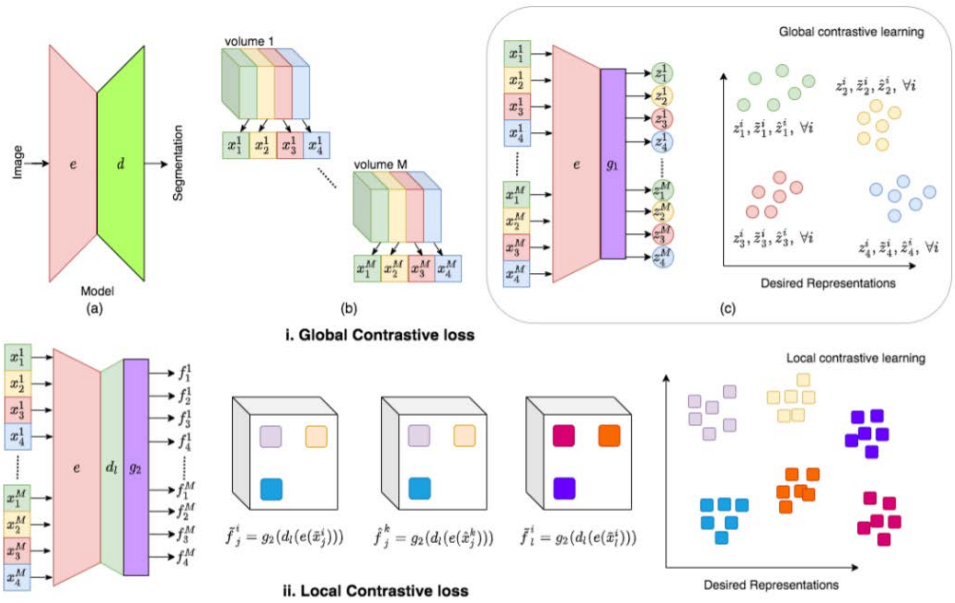

Fig. 10. Global-local Contrast learning framework

Compared with [33], SAM [35] utilize global embedding and local embedding to construct InfoNCE loss in the embedding space. This framework is more general, suitable for various tasks and has a relatively simple structure. The pipeline of SAM is illustrated in Fig.11. First, the volumetric image x and random transform sample x' as the input. Then, generate global and local embedding information through the network. Finally, the global and the local InfoNCE loss to encourage positive pairs to have similar embeddings, while pushing negative samples apart.

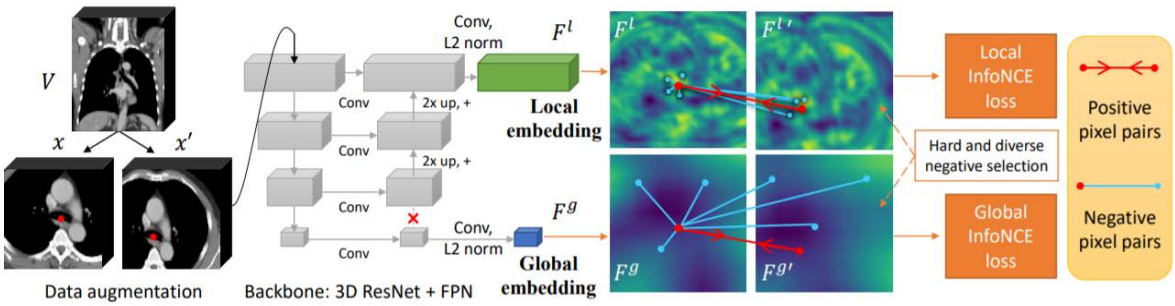

Fig. 11. The pipeline of SAM

4.2 Context-Context Contrast

Context-context contrastive learning directly utilizes the relationships between the global representations of different samples as what metric learning does, essentially, leveraging instance discrimination as a pretext task. Recently, CPC[32], SimCLR[25], MoCo[26], BYOL[27] based on Context-context contrast methods achieved great performance in the nature image domain,. These methods have inspired researchers in the medical domain, next we will introduce the practical applications of these methods in the field of medical domain.

4.2.1 CPC

Contrastive Predictive Coding (CPC) is a contrastive method that can be applied to any form of data such as text, voice, video, image, etc. For image data, a sample can be seen as a sequence of pixels or image blocks, then CPC learns the feature representation of spatial information. Inspired by CPC, TCPC [36] was proposed, which first uses Simple Linear Iterative Clustering (SLIC) [37] to locate the potential lesion area, and then modified CPC to learn 3D feature representation from the sub-volumes containing the lesion areas. TCPC framework as shown in the Fig.12.

TCPC adopts a u-shaped path and there are fewer differences in the content of medical images compared with natural images, so it uses a larger cube size than 2D CPC and 3D CPC.

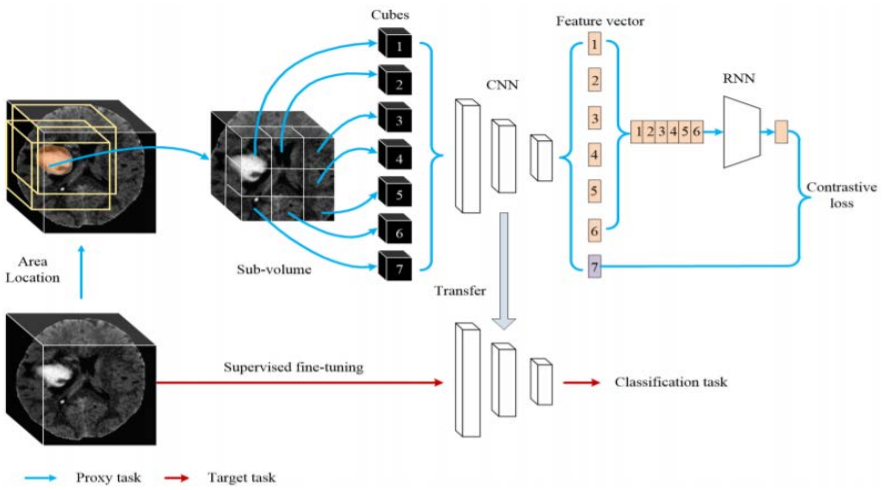

Fig. 12. TCPC framework [36]

4.2.2 MOCO

In MoCo [26], the idea of using instance discrimination through momentum contrast was further developed, which greatly increased the number of negative samples. Benefiting from the successful performance of MoCo, then researchers proposed MoCo-CXR [38] framework on unlabeled chest X-ray's dataset, and [39] further research was conducted on the basis of MoCo-CXR.

The medical image has the characteristics of small pixel area where the abnormality's part is located, gray-scale image, a large amount of unlabeled data [53], etc. So, some general data augmentations algorithms in MOCO are not suitable for medical images. MoCo-CXR modified the data augmentation strategy used to generate views suitable for the chest X-ray, such as horizontal flipping and random rotation. Experimental results show that MoCo-CXR-pretrained model outperformance than ImageNet-pretrained model, further verify the availability of MoCo in chest X-rays. following MoCo-CXR, MedAug [40] was proposed, which uses multiple images as a way to increase the number of positive pair choices, in this process, patient metadata is introduced to assist in constructing positive pairs. The process illustrated in Fig13. Their experiments proved that the performance of the model can be improved by using positive pairs selected by patient metadata.

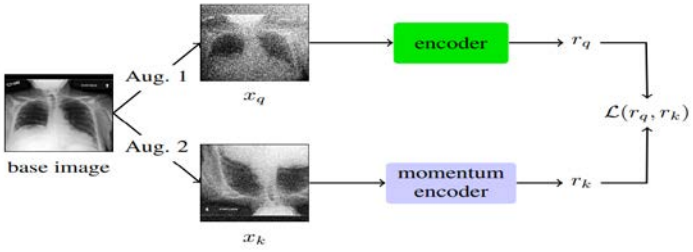

Fig. 13. Self-Supervised Pretraining using Momentum Contrast Learning

The medical image has the characteristics of small pixel area where the abnormality's part is located, gray-scale

image, a large amount of unlabeled data, etc. So, some general data augmentations algorithms in MOCO are not suitable for medical images. MoCo-CXR modified the data augmentation strategy used to generate views suitable for the chest X-ray, such as horizontal flipping and random rotation. Experimental results show that MoCo-CXR-pretrained model outperformance than ImageNet-pretrained model, further verify the availability of MoCo in chest X-rays. following MoCo-CXR, MedAug [40] was proposed, which uses multiple images as a way to increase the number of positive pair choices, in this process, patient metadata is introduced to assist in constructing positive pairs. The process illustrated in Fig13. Their experiments proved that the performance of the model can be improved by using positive pairs selected by patient metadata.

In [39], a self-supervised learning algorithm for COVID-19 prediction is proposed, which actually borrows from the overall framework of MoCo and adds some data augmentation strategy, such as, random Gaussian noise, random cropping, interpolation. Then, this self-supervised pretraining achieved the highest AUC scores on Single Image Prediction tasks of interest.

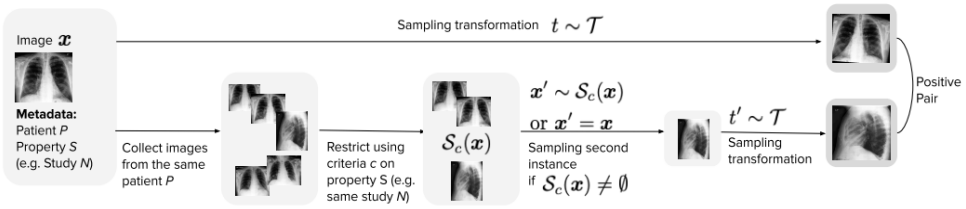

Fig. 14. The process selecting positive pairs used patient metadata for contrastive learning

4.2.3 SimCLR

SimCLR follows the end-to-end framework and is different from MoCo, which uses momentum contrast, and in order to deal with the large-scale negative sample problem, SimCLR chooses a larger batch. In [41] work, employ SimCLR framework applies to histopathological image analysis. The author conducted different downstream tasks to demonstrate that the feature extraction method based on comparative learning is significantly better than the baseline based on ImageNet. Furthermore, Inspired by SimCLR, Multi-Instance Contrastive Learning (MICLe) [42] was proposed. This method utilizes multiple images per patient to construct positive pairs, not only different augmentations of the same image but also different images of the same medical pathology, the construct process are shown in the Fig.15.

In MICLe, only use N pair of minibatch and lightweight data augmentation. Experiments show that MICLe significantly improves the accuracy and achieve the state-of-the-art result on the dermatology condition classification task.

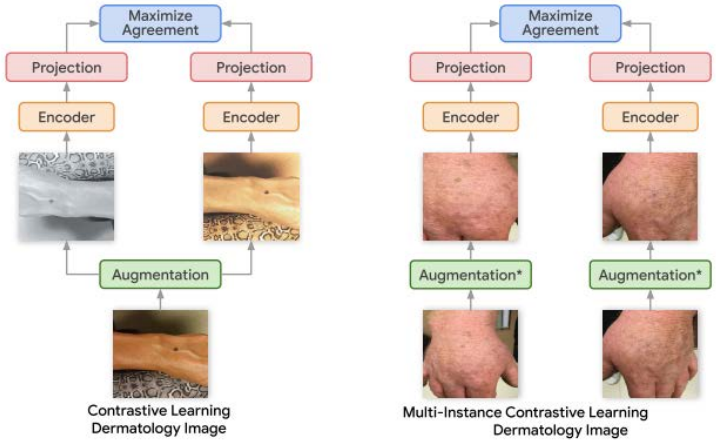

Fig.15. Multi-Instance Contrastive Learning (MICLe) framework.

4.2.4 BYOL

BYOL is a more radical approach, first discards negative sampling in contrastive learning but achieves an even better result, the model learns image representations using the online-target framework. The online network on an augmented view of an image to predict the representation of another augmented view of the same image produced by the target network. Based on the framework of BYOL, Prior-Guided Local (PGL) [43] focuses on Local consistency

loss and minimizes Local consistency loss based on the spatial and regional location relationship of the two augmented views. Fig.16 illustrated the difference between global consistency loss and local consistency loss.

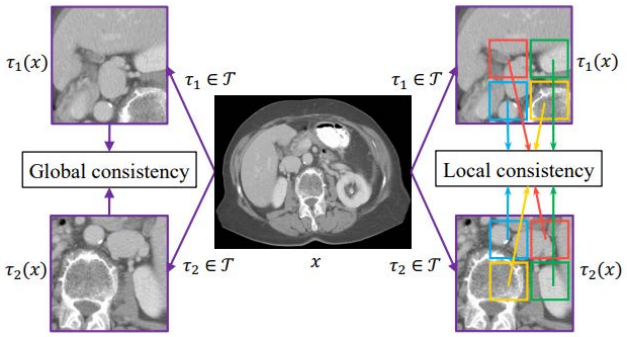

Fig. 16. Global consistency loss (BYOL) and local consistency loss (PGL)

PGL model framework including a data augmentation module τ and a prior dual-path module, as the Fig.17 shows. First, the unlabeled training data is augmented by module τ to obtain x_1, x_2, and then fed to the prior dual-path for feature extraction and alignment. The goal of the prior-guided aligner is to construct the spatial relation between f_1 and f_2 on the prior information of augmentation transformation and align the features.

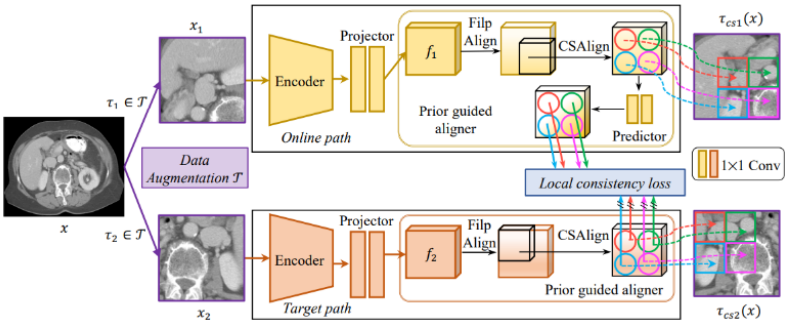

Fig. 17. PGL model framework.

5. Performance Comparison

For specific downstream tasks or target tasks in the field of medical imaging, they are mainly divided into two categories: classification and segmentation. This section integrates the performance of SSL methods on different datasets.

5.1 LUNA 2016

LUNA 2016 [46] dataset comes from the 2016 LUNA Nodule Analysis competition. Table 1. Shows the classification task results of the self-supervised learning methods on the LUNA 2016 dataset, all of the results, including the mean and standard deviation across ten trials. From the results, the self-supervised methods are conducive to the improvement of the performance of the target task, even performance is better than supervised learning, as illustrated in fig.18.

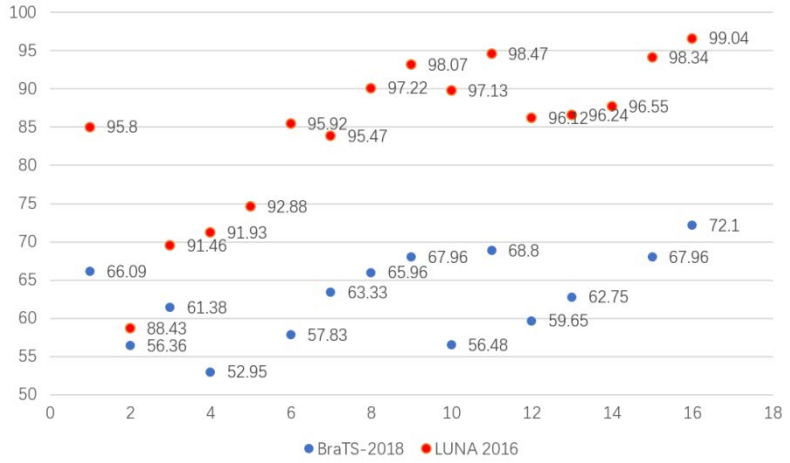

Fig. 18. The result of SSL methods performed on the LUNA 2016 and BraTS-2018, the first line of result data is the supervised method.

Table 1. Performance on LUNA 2016

Method	**Supervised**	**AUC%**(mean±s.d.)
MedicalNet [45]	✓	95.80±0.49
Autoencoder	×	88.43±10.25
In-painting [48]	×	91.46±2.97
Patch shuffling	×	91.93±2.32
De-noising [49]	×	95.92±1.83
Jigsaw	×	95.47±1.24
Rubik's Cube	×	96.24±1.27
DeepCluster [50]	×	97.22±0.55
Self-restoration	×	98.07±0.59
3D Rotation	×	97.13±0.81
Semantic Genesis 3D	×	98.47±0.32
3D Jigsaw	×	96.12±0.63
3D CPC	×	92.88±1.56
TCPC	×	96.55±0.97
Model Genesis	×	98.34±0.44
Universal Model	×	99.04±0.23

5.2 BraTS-2018

BraTS 2018 [47] dataset comes from Multimodal Brain Tumor Segmentation Challenge. The results of these experiments on the BraTS 2018 dataset are shown in Table 2.

Through these experimental results, it seems that whether segmentation tasks or classification tasks, the self-supervised learning methods redesigned based on medical image characteristics has higher performance.

Table 2. Performance on BraTS 2018

Method	**Supervised**	**IoU**%(mean±s.d.)
MedicalNet	✓	66.09±1.35
Autoencoder	×	56.36±5.32
In-painting	×	61.38±3.84
Patch shuffling	×	52.95±6.92
De-noising	×	57.83±1.57
Jigsaw	×	63.33±1.11
Rubik's Cube	×	62.75±1.93
DeepCluster	×	65.96±0.85
Self-restoration	×	67.96±1.29
3D Rotation	×	56.48±1.78
Semantic Genesis 3D	×	68.80±0.30
3D Jigsaw	×	59.65±0.81
Model Genesis	×	67.96±1.29
Universal Model	×	72.10±0.67

5.3 Performance of Contrastive Learning

Contrastive learning methods achieve The-state-of-the-art in self-supervised learning methods of natural images. The experimental results in Table 3 show that in the medical image analysis domain, the contrast-based methods also achieved The-state-of-the-art. The performance of the modified comparative learning method is more prominent and those methods more suitable for medical image analysis.

Table 3. Performance of Contrastive Learning methods

Method	Dataset	Dice %	Iou %
Random Init	KiTS [51]	81.57	74.63
Models Genesis	KiTS	82.32	75.51
BYOL	KiTS	84.06	77.56
PGL	KiTS	84.29	78.27
Random Init	Liver dataset[52]	73.97	66.79
Models Genesis	Liver dataset	74.74	67.68
BYOL	Liver dataset	74.82	68.09
PGL	Liver dataset	76.05	69.06
Random Init	Spleen dataset [52]	93.23	88.08
Models Genesis	Spleen dataset	94.20	89.44
BYOL	Spleen dataset	94.56	89.83
PGL	Spleen dataset	95.60	91.61

6. Discussions and Future Directions

The application of self-supervised learning methods in the medical field has achieved great success, and obtaining good performance that exceeds the supervised pre-training model on certain medical image tasks. However, directly applying existing self-supervised learning methods is not necessarily suitable for medical image tasks [57].

6.1 Selection of Data Augmentation and Pretext Tasks

In medical images, human organs are consistent, and only a small part of the lesions are different [55]. So, some data augmentation methods and some pretext-tasks will destroy the main information of the raw data. Therefore, it is necessary to construct suitable pretext-tasks and data augmentation strategies according to the characteristics of medical image data. By designing effective pretexts, using the information other than the part to be predicted to predict a certain subset of information, mining the potential laws of large-scale unsupervised data. Learn data representations and inductive biases that help improve downstream task performance. These data representations cover good semantics or structural meanings.

6.2 Extract features with Multiple Pretext Tasks

Most existing self-supervised learning methods in the medical domain learn data features by training a pretext-task. Different pretext-tasks mean different supervision signals, which can help the network learn more data features. The existing methods construct pre-tasks from both local and global perspectives [33], so that the model not only learns global features, but also focuses on fine-grained features, and has achieved good experimental results on medical image segmentation tasks. Whether it is through the experimental results of the literature or the perspective of algorithm proof, multiple Pretext Tasks are helpful to extract the potential law of data distribution.

6.3 Dataset Imbalance

The bias of the dataset is a normal phenomenon, but this phenomenon is more prominent in medical images. In a large amount of unlabeled medical data, the data of the disease is actually much smaller than the data of the normal person. Use self-supervision to overcome the inherent "data bias" and learn better initialization feature information that is not related to labels from unbalanced datasets [44]. The problem of data imbalance needs to be considered in the construction of the self-supervised framework in the medical domain.

6.4 Redesign Contrastive Learning

In the field of medical imaging, directly applying the existing Contrastive Learning framework has limited improvement in experimental results. According to the characteristics of medical images, suitable negative samples can be constructed, then extract more useful data features [56,58]. In order to improve the performance of self-supervised learning methods in the field of medical image analysis, we need to further explore how to construct negative examples and how to better adapt self-supervised learning methods to downstream tasks.

7. Conclusions

This article has extensively reviewed the latest applications of self-supervised methods in the field of medical image analysis. Self-supervised learning can use unlabeled data or even unbalanced data to extract latent features. This unsupervised learning paradigm naturally adapts to the problems of medical image data. We have clearly categorized recent methods and introduced the pipelines of these methods separately, introduces background knowledge and important frameworks. Finally, discussed the future research directions, and puts forward the issues that need to be paid attention to when designing new methods and paradigms. In short, our work fills the gap of review papers in the medical image analysis based on self-supervised learning and researchers can easily grasp the cutting-edge ideas of this

domain.

Acknowledgment

This research has been partially supported by China Scholarship Council (CSC), and Special thanks should go to my supervisor professor Sergii Stirenko, for his instructive advice and useful suggestions on my paper.

References

[1] S. Gidaris, P. Singh, N. Komodakis.” Unsupervised Representation Learning by Predicting Image Rotations,” Proceedings of the International Conference on Learning Representations, 2018.
[2] M. Noroozi, P. Favaro. “Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles,” In European conference on computer vision, 2016, pp: 69-84.
[3] T. Nima, Y. Hu, J. Cao, X. Yan, Y. Xiao, Y. Lu, J. Liang, et al. “Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data.” In 2019 IEEE 16th International Symposium on Biomedical Imaging,2019, pp. 1251-1255.
[4] C. Szegedy, S. Ioffe, V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1. 2017.
[5] A. Hatamizadeh, P. Shilpa, X. Ding, D. Terzopoulos, and N. Tajbakhsh. "Automatic segmentation of pulmonary lobes using a progressive dense V-network." In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, pp. 282-290.
[6] Imran, A. A. Z., Huang, C., Tang, H., Fan, W., Xiao, Y., Hao, D., et al. “Partly Supervised Multitask Learning.” arXiv preprint arXiv:2005.02523,2020.
[7] A. R. Venkatakrishnan, S. T. Kim, R. Eisawy, F. Pfister, & Navab, N. “Self-Supervised Out-of-Distribution Detection in Brain CT Scans.” arXiv e-prints, arXiv-2011,2020.
[8] Y. Li, J. Chen, X. Xie, K. Ma, & Y. Zheng. “Self-Loop Uncertainty: A Novel Pseudo-Label for Semi-supervised Medical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention.” pp. 614-623, Oct. 2020
[9] A. Taleb,W. Loetzsch, N. Danz, J. Severin, T. Gaertner, B. Bergner, & C. Lippert. “3d self-supervised methods for medical imaging.” arXiv preprint arXiv:2006.03829, 2020.
[10] X. Zhuang, Y. Li, Y. Hu, K. Ma, Y. Yang, and Y. Zheng, “Self-supervised Feature Learning for 3D Medical Images by Playing a Rubik’s Cube.” arXiv preprint arXiv:1910.02241, 2019.
[11] X. Tao, Y. Li, W. Zhou, K. Ma, & Y. Zheng, Revisiting Rubik’s cube: self-supervised learning with volume-wise transformation for 3D medical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, October. pp. 238-248.
[12] J. Zhu, Y. Li, Y. Hu, K. Ma, S. K. Zhou, & Y. Zheng. Rubik’s Cube+: A self-supervised feature learning framework for 3D medical image analysis. Medical Image Analysis, 2020, 64, 101746.
[13] S. Li, K. Yu, and K. Batmanghelich. "Context Matters: Graph-based Self-supervised Representation Learning for Medical Images." arXiv preprint arXiv:2012.06457 ,2020.
[14] Bai, W., Chen, C., Tarroni, G., Duan, J., Guitton, F., Petersen, S. E., ... & Rueckert, D. "Self-supervised learning for cardiac mr image segmentation by anatomical position prediction." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 541-549. 2019.
[15] Zhou, Z., Sodha, V., Pang, J., Gotway, M. B., & Liang, J. and Jianming Liang. "Model’s genesis." Medical image analysis 67: 101840, 2021.
[16] Haghighi, F., Taher, M. R. H., Zhou, Z., Gotway, M. B., & Liang, J. "Learning Semantics-enriched Representation via Self-discovery, Self-classification, and Self-restoration." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 137-147. 2020.
[17] Zhang, X., Zhang, Y., Zhang, X., & Wang, Y. Universal Model for 3D Medical Image Analysis. arXiv preprint arXiv:2010.06107, 2020.
[18] Zhou, Z., Sodha, V., Siddiquee, M. M. R., Feng, R., Tajbakhsh, N., Gotway, M. B., & Liang, J. "Model’s genesis: Generic autodidactic models for 3d medical image analysis." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 384-393.2019.
[19] Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. Self-supervised learning for medical image analysis using image context restoration. Medical image analysis, 58, 101539. 2019.
[20] Yu, L., Zhang, W., Wang, J., & Yu, Y. Seqgan: Sequence generative adversarial nets with policy gradient. In Proceedings of the AAAI conference on artificial intelligence. 2017, February,Vol. 31, No. 1.
[21] Ross, T., Zimmerer, D., Vemuri, A., Isensee, F., Wiesenfarth, M., Bodenstedt, S., ... & Maier-Hein, L. Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. International journal of computer assisted radiology and surgery, 13(6), 925-933. 2018.
[22] M. Arjovsky, S. Chintala, L. Bottou, “Wasserstein gan,” arXiv:1701.07875,2017.
[23] Larsson, G., Maire, M., & Shakhnarovich, G. Colorization as a proxy task for visual understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6874-6883. 2017.
[24] Liu, H., Liu, J., Hou, S., Tao, T., & Han, J. Perception consistency ultrasound image super-resolution via self-supervised CycleGAN. Neural Computing and Applications, 1-11. 2021.
[25] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. A simple framework for contrastive learning of visual representations. In International conference on machine learning. 2020, November, pp. 1597-1607.

[26] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9729-9738. 2020.
[27] Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., ... & Valko, M. Bootstrap your own latent: A new approach to self-supervised learning. arXiv preprint arXiv:2006.07733. 2020.
[28] Chen, X., & He, K. Exploring Simple Siamese Representation Learning. arXiv preprint arXiv:2011.10566.2020.
[29] Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. arXiv preprint arXiv:2103.03230. 2021.
[30] Zhang, P., Wang, F., & Zheng, Y. Self-supervised deep representation learning for fine-grained body part recognition. In 2017 IEEE 14th International Symposium on Biomedical Imaging. April ,2017, pp. 578-582. 2017.
[31] Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 1735–1742. 2006.
[32] Henaff, O.J. Data-efficient image recognition with contrastive predictive coding. In IEEE Conference on Computer Vision and Pattern Recognition,pp. 4182-4192, 2019.
[33] Chaitanya, K., Erdil, E., Karani, N., & Konukoglu, E. Contrastive learning of global and local features for medical image segmentation with limited annotations. arXiv preprint arXiv:2006.10511. 2020.
[34] Oord, A. V. D., Li, Y., & Vinyals, O. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748. 2018.
[35] Yan, K., Cai, J., Jin, D., Miao, S., Harrison, A. P., Guo, D., et al. Self-supervised Learning of Pixel-wise Anatomical Embeddings in Radiological Images. arXiv preprint arXiv:2012.02383.2020.
[36] Zhu, J., Li, Y., Hu, Y., & Zhou, S. K. Embedding Task Knowledge into 3D Neural Networks via Self-supervised Learning. arXiv preprint arXiv:2006.05798.2020
[37] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. vol. 34, pp. 2274–2282,2012.
[38] Sowrirajan, H., Yang, J., Ng, A. Y., & Rajpurkar, P. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. E-print arXiv:2010.05352,2020
[39] Sriram, A., Muckley, M., Sinha, K., Shamout, F., Pineau, J., Geras, K. J., ... & Moore, W. (2021). COVID-19 Prognosis via Self-Supervised Representation Learning and Multi-Image Prediction. E-print arXiv:2101.04909,2021.
[40] Vu, Y. N. T., Wang, R., Balachandar, N., Liu, C., Ng, A. Y., & Rajpurkar, P. MedAug: Contrastive learning leveraging patient metadata improves representations for chest X-ray interpretation. arXiv preprint arXiv:2102.10663.2021.
[41] Ciga, O., Martel, A. L., & Xu, T. Self-supervised contrastive learning for digital histopathology. arXiv preprint arXiv:2011.13971. 2020.
[42] Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., et al. Big Self-Supervised Models Advance Medical Image Classification. arXiv preprint arXiv:2101.05224. 2021.
[43] Xie, Y., Zhang, J., Liao, Z., Xia, Y., & Shen, C. (2020). PGL: Prior-Guided Local Self-Supervised Learning for 3D Medical Image Segmentation. arXiv preprint arXiv:2011.12640. 2020.
[44] Yang, Y., & Xu, Z. Rethinking the value of labels for improving class-imbalanced learning. arXiv preprint arXiv:2006.07529. 2020.
[45] Chen, S., Ma, K., & Zheng, Y. Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625. 2019.
[46] Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., van den Bogaard, C., Cerello, P., ... & Jacobs, C. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Medical image analysis, 42, 1-13. 2017.
[47] Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629. 2018.
[48] Pathak, D., Krahenbuhl, 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. pp. 2536-2544. 2016.
[49] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A., & Bottou, L. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research, 11(12). 2010.
[50] Caron, M., Bojanowski, P., Joulin, A., & Douze, M. Deep clustering for unsupervised learning of visual features. In Proceedings of the European Conference on Computer Vision. pp.132-149. 2018.
[51] Heller, N., Isensee, F., Maier-Hein, K. H., Hou, X., Xie, C., Li, F., et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. Medical Image Analysis, 67, 101821. 2021.
[52] Simpson, A. L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063. 2019.
[53] Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., ... & Stirenko, S. Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer. In International Conference on Computer Science, Engineering and Education Applications. Springer, Cham, pp. 638-647, 2018.
[54] Yang, Y. H., Xu, J. S., Gordienko, Y., & Stirenko, S. Abnormal Interference Recognition Based on Rolling Prediction Average Algorithm. In International Conference on Computer Science, Engineering and Education Applications. Springer, Cham, pp. 306-316. 2020.
[55] Sulema, Y., Kerre, E. and Shkurat, O. Vector Image Retrieval Methods Based on Fuzzy Patterns. International Journal of Modern Education and Computer Science, 12(3), 2020.
[56] Md. Rahat Khan, A. S. M. Shafi, " Statistical Texture Features Based Automatic Detection and Classification of Diabetic Retinopathy", International Journal of Image, Graphics and Signal Processing, Vol.13, No.2, pp. 53-61, 2021.
[57] Farzaneh Nikroorezaei, Somayeh Saraf Esmaili, " Application of Models based on Human Vision in Medical Image Processing: A Review Article", International Journal of Image, Graphics and Signal Processing, Vol.11, No.12, pp. 23-28, 2019.

[58] Kama, R., Chinegaram, K., Tummala, R.B. and Ganta, R.R. Segmentation of Soft Tissues and Tumors from Biomedical Images using Optimized K-Means Clustering via Level Set formulation. International Journal of Intelligent Systems and Applications, 11(9), p.18, 2019.

Authors' Profiles

Jiashu Xu received a master's degree from the Department of computing engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". Now is a Ph.D. student from the same university. His research interests include self-supervised learning, unsupervised learning, computer vision, GAN, and their applications in the medical image domain.

How to cite this paper: Jiashu Xu, " =A Review of Self-supervised Learning Methods in the Field of Medical Image Analysis", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.13, No.4, pp. 33-46, 2021.DOI: 10.5815/ijigsp.2021.04.03