Improvement of Robustness to Noise for Medical Image Segmentation by using Self-Supervised Learning Approach*

Yuta Konishi[0000−0003−0565−564X] and Takio Kurita[0000−0003−3982−6750]

Hiroshima University,
1-7-1 Kagamiyama, Higashi Hiroshima, 739-8521, Japan

Abstract. It is crucial to make the trained model robust to the distortions such as pixel noises in medical image segmentation. Recently it has been pointed out that self-supervised learning (SSL) methods such as SimCLR, VICReg, and Barlow Twins are closely related to spectral methods such as Laplacian Eigenmaps, Multidimensional Scaling, etc. This means that SSL can construct features invariant to the perturbations introduced by data augmentations. Since invariant feature extraction is also fundamental in medical image segmentation, in this paper, we proposed introducing SSL loss as a regularizer in U-Net for medical image segmentation. Pixel noise is applied to the training samples, and invariant features to such distortions are extracted in the hidden layer of U-Net. The effectiveness of the proposed approach is experimentally confirmed using the subset of Sunnybrook Cardiac Data (SCD) and Abdominal organs segmentation dataset by Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) Challenge.

Keywords: Invariant feature extraction · Self-supervised learning · Medical image segmentation · U-Net · SimCLR.

1 Introduction

Medical image segmentation is used to identify the pixels of organs or lesions from medical images such as CT or MRI images and is regarded as one of the most important tasks in medical image analysis [12]. Deep learning is now recognized as one of the best approaches for medical image segmentation [30]. Many network architectures, such as the fully convolutional neural network (FCN) [17] or U-Net [24], have been used to segment medical images.

U-Net is one of the most well-known architectures for medical image segmentation. The encoder-decoder architecture is utilized, and skip connections between different stages of the network are introduced, as shown in Fig.1. Many researchers applied the U-Net base model for medical image segmentation[6, 8].

Invariant feature extraction is one of the central topics in machine learning and pattern recognition, and it is also important in deep learning. The standard

* Supported by KEKEN 21K12049.

approach to making robust to unnecessary variations is to train a deep learning model by using a large number of training samples that include all possible variations. There are some researches in which invariant features are extracted by using deep learning. For example, pose-invariant features are extracted using Convolutional Neural Networks (CNN) for pose-invariant face recognition [1]. Metric learning has also often been used for invariant feature extraction [13, 16]. Ueda et al. proposed an invariant feature extraction method using Gradient Reversal Layer (GRL) [27].

Self-Supervised Learning (SSL) is one of the most promising methods to learn data representations that generalize across downstream tasks [3]. Labels for the training samples are not required, but the knowledge of what makes some samples semantically close to others is trained. Usually, semantic similarity is constructed by augmenting the training samples through data augmentations.

One of the basic SSL methods is SimCLR (a simple framework for contrastive learning of visual representations) [5]. SimCLR learns representations by maximizing agreement between differently augmented views of the same sample via a contrastive loss in the latent space. Recently Balestriero et al. [3] demonstrated that SSL methods such as SimCLR [5], VICReg [4], and Barlow Twins [29] are closely related with the spectral methods such as Laplacian Eigenmaps, Multidimensional Scaling, etc. This means that SSL is extracting features (embeddings) that are invariant to the perturbations introduced by data augmentations.

The invariant feature extraction is also fundamental in supervised learning. Ramyaa et al. proposed to combine Barlow Twins loss with the standard cross entropy loss for the supervised learning with CNN [23].

In this paper, we propose to use SSL loss as a regularizer in U-Net-based medical image segmentation. Pixel noise is applied to the training samples as the distortions to the medical images, and the invariant features to such distortions are extracted by introducing SSL loss in the hidden layers of the U-Net. To show the effectiveness of the proposed approach, we have performed experiments using the subset of Sunnybrook Cardiac Data (SCD) [22] and Abdominal organs segmentation dataset by Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) Challenge[14].

The contributions of this paper are summarized as follows:

(1) SSL loss in the hidden layers of the U-Net is introduced to make the trained model for medical image segmentation robust to the pixel noises.
(2) The effectiveness of the proposed approach is experimentally confirmed using the subset of Sunnybrook Cardiac Data (SCD) [22] and Abdominal organs segmentation dataset by Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) Challenge[14].

2 Related Work

2.1 Medical Image Segmentation

Image segmentation is a computer vision technique that divides a region in an image into several objects. It has been applied in a wide range of fields,

such as medical image analysis, scene understanding, robotic perception, video surveillance, augmented reality, image compression, automatic driving, and so on [19].

Image segmentation plays a crucial role in many medical image analyses in which the pixels of organs or lesions are identified from medical images such as CT or MRI images [21, 12]. Deep learning is now recognized as one of the best approaches for medical image segmentation [30].

Many network structures have been used for medical image segmentation. One of the basic deep learning models is Convolutional Neural Networks (CNNs). A CNN consists of a stack of layers such as convolution, pooling, and fully connected layers [26]. In the fully convolutional neural network (FCN) proposed by Long et al. [17], the fully convolutional layer is used at the last layer instead of the fully-connected layers in the standard CNN. With this replacement, the network makes a dense pixel-wise prediction easy.

One of the most well-known structures for medical image segmentation is U-Net, proposed by Ronneberger et al. [24]. By introducing deconvolution, the encoder-decoder architecture is realized in U-Net. Also, U-Net introduces skip connections between different stages of the network. These connections bypass the information between the layers of equal resolution in the encoding path to the decoding path. This is the most important property of U-Net. Many researchers applied the U-Net base model for medical image segmentation [6, 8].

Deep learning-based models have achieved good segmentation accuracy. However, to train the network, a large number of annotated training samples are required [17]. Collecting such huge training samples is often very tough and expensive in medical image analysis. The most common approach to increase the size of the training samples is data augmentation in which a set of perturbations are applied to the images in the training samples [18, 7, 20].

Another solution to this problem is transfer learning. Transfer learning employs the knowledge learned in a different source domain to a target task [25]. Transfer learning has been proven to have better performance when the tasks of the source and target network are more similar.

It is common in medical images that the anatomy of interest only occupies a very small portion of the image. Namely, most pixels belong to the background area, while these small organs (anomalies) are more important for medical diagnosis. Training a network with such data often leads to the trained network being biased toward the background. A popular solution to this issue is sample re-weighting, where a higher weight is applied to the foreground patches. Dice loss is often used for automatic re-weighting [15, 24, 31, 28].

Another approach is introducing the prior knowledge into the loss function as a regularizer. For example, Euler characteristics (EC) from topology are used to calculate the number of isolated objects on segmented vessel regions in the fundus image. It is used as the regularizer for training [10]. It is also useful to utilize information on the neighboring pixel relationship. Hakim et al. [11] proposed introducing a regularization term defined based on the differences of neighboring

pixels. The regularization term can be represented as Graph Laplacian computed from the output of the network and the ground-truth image.

2.2 Self-Supervised Learning and Invariant Feature Extraction

Recently it has been shown that Self-Supervised Learning (SSL) can extract features with the same level as supervised learning with large training samples [3]. SSL can build representations of data without labels and give significant advances in various applications such as natural language processing, speech processing, and computer vision [2].

In SSL for computer vision applications, the distortions or perturbations are added to the original image. The features extracted from the distorted images are trained so that they are close to each other. This is achieved by maximizing the similarity of representations obtained with different distortions using a variant of Siamese networks [9]. Thus SSL can learn invariant representations (embeddings) to the added distortions of the input image.

SimCLR SimCLR (a simple framework for contrastive learning of visual representations) is one of the basic methods for contrastive self-supervised learning [5]. SimCLR learns representations by maximizing agreement between differently augmented views of the same sample via a contrastive loss in the latent space.

Let $\{\boldsymbol{x}_k|k=1,\ldots,N\}$ be the training samples in a mini-batch. At first, for each training sample in the mini-batch, a stochastic data augmentation is applied to randomly generated two views of the same sample, denoted $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}_j$, which are considered as a positive pair. Then we obtain $2N$ pairs of the augmented samples derived from the samples in the mini-batch. These augmented samples include a positive pair $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}_j$, which are generated from the same training sample $\boldsymbol{x}_i$. The pairs of the augmented samples are fed into the neural network encoder to get the hidden representation $\boldsymbol{h}_i = f(\tilde{\boldsymbol{x}}_i)$. The contrastive loss is applied after the hidden representation is mapped by a small neural network projection head as $\boldsymbol{z}_i = g(\boldsymbol{h}_i)$.

The loss function for a positive pair of examples (i, j) is defined as

$$l_{i,j} = -log\frac{\exp(sim(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k\neq i]} \exp(sim(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)} \tag{1}$$

where $sim(\boldsymbol{u}, \boldsymbol{v}) = \dfrac{\boldsymbol{u}^T\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$ is the cosine similarity between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ and $\mathbb{1}_{[k\neq i]}$ is an indicator function evaluating to 1 if $k \neq i$. τ denotes a temperature parameter that controls the scale. This loss function can be used to learn to keep positive pairs in a mini-batch close together and other pairs apart.

Other SSL methods For SSL, it is important to prevent a collapse in which the encoders produce constant or non-informative representations. Bardes et al.

proposed VICReg (Variance-Invariance-Covariance Regularization), which explicitly avoids the collapse problem with two regularization terms [4]. One term maintains the variance of each embedding dimension above a threshold, and the other decorrelates each pair of variables.

Another method is Barlow Twins, which applies H. Barlow's redundancy-reduction principle [29]. The objective function of Barlow Twins measures the cross-correlation matrix between the embeddings of two identical networks fed with distorted versions of a batch of samples and tries to make this matrix close to the identity matrix. This makes the embedding vectors of distorted versions similar while minimizing the redundancy between the components of these vectors. It is reported that Barlow Twins is competitive with state-of-the-art methods for SSL.

Balestriero et al. [3] demonstrated that SSL methods such as SimCLR, VICReg, and Barlow Towns are closely related to spectral methods such as Laplacian Eigenmaps, Multidimensional Scaling, etc. This shows that invariant feature extraction is fundamental in SSL.

Since it is obvious that the invariant feature extraction is also important in supervised learning, Barlow Twins loss is combined with the standard cross-entropy loss as a regularizer in the supervised learning with CNN [23]. This paper proposes to use SSL loss as a regularizer in U-Net for medical image segmentation.

3 Proposed Method

3.1 Overview of the network architecture

As discussed in 2.2, SSL can learn representations that are invariant to the distortions applied to images. Taking advantage of this property, we designed a mechanism to promote learning that is robust to pixel noises in the medical image segmentation tasks.

A branch of the linear layer is connected to the middle layer of the segmentation model (U-Net) as shown in Figure 1, and SSL is performed with the output vectors of the branch. Each part of U-Net is named as shown in Figure 1, and a branch for SSL is connected to an arbitrary location.

3.2 Training flow

We follow the learning method used in SimCLR's paired data learning. Gaussian noise are applied to the original images, and they are paired with the original images.

The original and the distorted images are fed to the mainstream (U-Net), and the outputs of the U-Net are used to compute the segmentation loss function for each image. The SSL branch of the sub-stream outputs the feature vectors of the original and the distorted images, and these vectors are used to compute the SSL loss function. This allows us to capture representations that are invariant

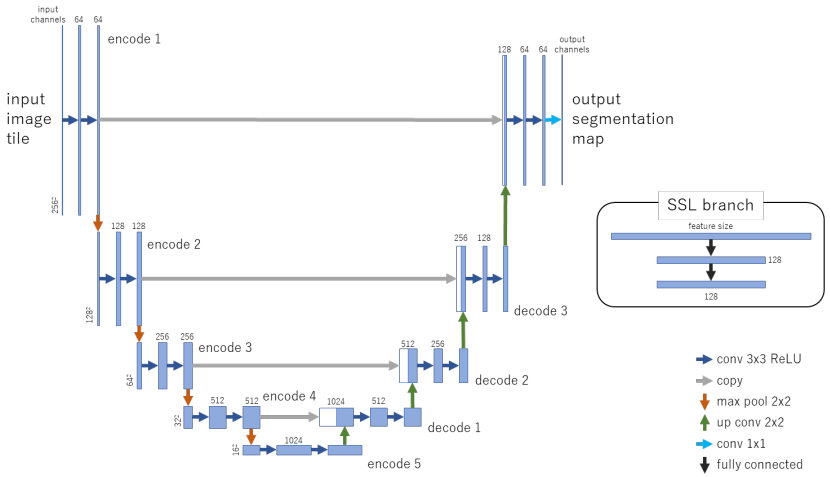

Fig. 1. Overview of U-Net and the SSL branch. The SSL branch is connected to the location encode1, encode2, encode3, encode4, encode5, decode1, decode2 or decode3.

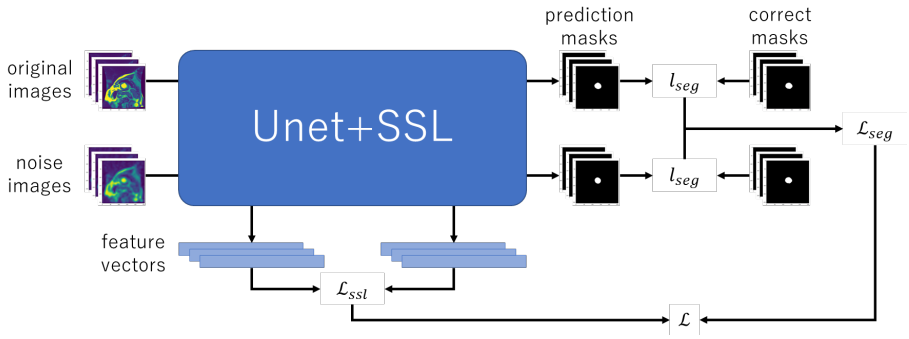

Fig. 2. Training flow of the proposed method. The best accuracy is obtained when SSL branch is connected to encode2 or decode2.

to this distortion (noise) during training for the segmentation task and to make the trained model robust to such distortions. The overview of the training flow of the proposed method is shown in Figure 2.

In the proposed learning flow, Segmentation learning and SSL are performed simultaneously. This means that the loss functions must be computed and fused. In this study, the loss function is defined as the weighted sum of the loss functions of segmentation and SSL as

$$\mathcal{L} = \lambda \mathcal{L}_{seg} + (1 - \lambda)\mathcal{L}_{ssl} \tag{2}$$

where $\mathcal{L}_{seg}$ and $\mathcal{L}_{ssl}$ are the loss functions of segmentation and SSL and λ is a hyper-parameter to control the ratio of the two loss functions. Since the segmentation loss function is computed for each of the original and the distorted images, the segmentation loss $\mathcal{L}_{seg}$ is defined by their respective averages as

$$\mathcal{L}_{seg} = \frac{l_{seg}(Y_{original}) + l_{seg}(Y_{noise})}{2} \tag{3}$$

where l_{seg} is the loss function of segmentation and $Y_{original}$ and Y_{noise} are the outputs of U-Net for the original and the distorted images. In this study, Cross-entropy Loss is used for segmentation loss and InfoNCE Loss (eq (1))

$$l_{i,j} = -log \frac{\exp(sim(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(sim(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

is used for SSL loss.

4 Experiments

4.1 Datasets

To evaluate the effectiveness of the proposed approach, we have performed experiments using two datasets. They are the subset of Sunnybrook Cardiac Data (SCD) [22] and Abdominal organs segmentation dataset by Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) Challenge[14]. Images of the datasets are resized (bilinear) to 256×256 pixels.

Subset of SCD. The SCD also called the 2009 Cardiac MR Left Ventricular Segmentation Challenge data, consists of 45 cine MRI images of various patients and conditions. The SCD subset used in this study consists of gray-scale cardiac MRI images (short-axis images) and expert-masked data of the left ventricular region. The masked data is a binary image with 1 inside the region of the left ventricle and 0 in other regions. The training data set consists of 234 image pairs, and the validation data set consists of 26 image pairs. They do not overlap each other.

Abdominal organs segmentation dataset by CHAOS challenge. The CHAOS Challenge is aimed at segmenting organs (liver, kidneys, spleen) from abdominal CT and MRI data. CT and MRI are provided in DICOM image data, each with masked images of abdominal organs. The CT dataset is data acquired for the pre-evaluation of living liver transplant donors and is intended for the segmentation of the liver. The MRI data set consists of data from two different sequences (T1-DUAL and T2-SPIR) and is intended for the segmentation of the four abdominal organs (liver, right and left kidneys, and spleen). The MRI T2-SPIR data set was used in this experiment. As mentioned earlier, this data set is DICOM image data, so it was converted to JPEG image data for easier handling. Of the total MRI images, 531 were used as training data and 92 as validation data. The classes to be classified are the four abdominal organs (liver, right and left kidneys, and spleen) as described above.

4.2 Experimental details

Distortions The distortion used in this study is Gaussian noise. Gaussian noise is statistical noise that has the same probability density function as the Gaussian distribution. The noise image was generated by adding 0.3 times the Gaussian noise (standard normal distribution) to the original image.

Learning parameters The batch size was set to 9, and Adam was used as the optimizer. For the subset of SCD, the number of epochs was set to 100, and the learning rate was set to 0.001, which was multiplied by 0.5 every 25 epochs. The weight decay was set to 0.001. For the Abdominal organs segmentation dataset, the number of epochs was set to 250, and the learning rate was set to 0.0001, which was multiplied by 0.5 every 40 epochs. The weight decay was set to 0.01.

Evaluation Multi-class IoU and pixel-wise accuracy, which are common metrics for segmentation tasks, were used for evaluation. After training the model, prediction using the trained model is performed on the original images and the distorted images, and each is evaluated.

4.3 Ablation study using SCD dataset

To confirm the usefulness of our proposed method, we conducted a preliminary experiment using the subset of SCD. The U-Net was trained with only the original images of the subset of SCD. This is denoted as baseline1. Also, the U-Net was trained with the samples in which 50% of the samples are replaced with the distorted samples. This is denoted as baseline2. Then the performance of the proposed method is compared with these baselines.

In these experiments, the parameter λ in the loss function was set to 0.8.

Optimum Location of SSL branch To find the best location of the SSL branch, we connected the SSL branch in different layers in the U-Net and evaluated the test accuracy of the trained models for the original test samples and the distorted images of the test samples. In this experiment, SSL branch was connected to 8 locations of U-Net encode1-5, decode1-3, and each of them was trained to compare their performance. The results of the experiment are shown in Table 1.

Table 1. Comparison of accuracy for the subset of SCD. To find the optimal location of the SSL branch of the proposed method, the accuracy was evaluated by changing the location of the SSL branch in the U-Net.

	multi-class IoU (%)		pixel-wise accuracy (%)	
model	original images	distorted images	original images	distorted images
baseline 1	94.63	49.12	99.81	98.23
baseline 2	94.28	93.53	99.80	99.77
encode1	94.54	93.07	99.81	99.75
encode2	95.32	**94.06**	99.83	**99.79**
encode3	94.47	90.44	99.80	99.66
encode4	94.16	92.50	99.78	99.72
encode5	94.87	92.48	99.82	99.72
decode1	94.28	92.83	99.79	99.75
decode2	**95.56**	93.92	**99.84**	99.78
decode3	94.71	93.05	99.81	99.75

From Table 1, it is noticed that the accuracy of the proposed method is better than the baselines, especially for the distorted test images. This means that the proposed approach can make the trained model robust to distortion such as pixel noise.

The best accuracy was achieved at decode2 for the original images and encode2 for the distorted images. The results suggest that visually relevant features, such as the contours of objects in the image, are more effective in making the learned model robust to variations such as pixel noise. In contrast, the extraction of class information is more important for the original images. Thus the reason why these results are obtained is probably that the distortions (pixel noises) used in this experiment are local and the deeper layers are probably effective for more global distortions.

In subsequent experiments, the SSL branch is connected to encode2, which had the best accuracy for the distorted images, and decode2, which had the best accuracy on the decoder side.

Optimum value of λ. Next, we performed experiments to find the best value of the parameter λ which controls the valance between the segmentation loss and SSL loss. The accuracy for the original test samples and the distorted samples was evaluated by changing the parameter λ from 0.1 to 0.9 in 0.1 increments.

Table 2. Comparison of accuracy by λ (%)

	multi-class IoU		pixel-wise accuracy	
λ	original images	distorted images	original images	distorted images
0.1	93.92	90.31	99.78	99.66
0.2	94.39	91.57	99.80	99.70
0.3	93.67	85.92	99.78	99.51
0.4	94.17	93.29	99.78	99.76
0.5	94.34	92.63	99.79	99.73
0.6	94.73	92.54	99.80	99.74
0.7	95.08	93.18	99.82	99.76
0.8	**95.32**	**94.06**	**99.83**	**99.79**
0.9	94.93	93.84	99.82	99.78

Table 2 shows the accuracy obtained for each parameter λ. The best accuracy is achieved for the distorted images when the value of λ is 0.8. It is also noticed that the proposed method is robust to the small changes of this parameter λ.

In the following experiments, the value of λ is set to 0.8.

4.4 Experiments with Abdominal organs segmentation dataset

We have also conducted experiments on the Abdominal organs segmentation dataset, which consists of color images and the task is multi-class segmentation. The results are summarized in Table 3.

Table 3. Comparison of accuracy with Abdominal organs segmentation dataset.

	multi-class IoU (%)		pixel-wise accuracy (%)	
model	original images	distorted images	original images	distorted images
baseline 1	**87.45**	19.11	**99.43**	95.52
baseline 2	85.47	81.63	99.23	98.86
encode2	83.92	79.24	99.10	98.89
decode2	83.89	**82.64**	99.00	**98.91**

From Table 3, it can be confirmed that for distorted images, the best accuracy for both multi-class IoU and per-pixel accuracy is obtained when the SSL branch is connected to decode2. This suggests that in the case of multi-class segmentation, the accuracy can be improved by acquiring features from the U-Net decoder side using the SSL branch.

Figure 3 shows the segmentation results for the distorted training image of the Abdominal organs segmentation dataset. It can be seen that the proposed method is able to segment organ contours more clearly than baseline2. The results show that the proposed method is robust to distortions such as pixel noise by introducing SSL loss in the hidden layer of U-Net.

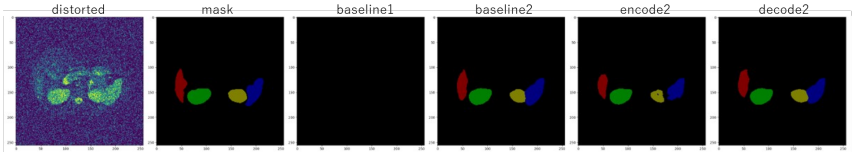

Fig. 3. Comparison of the segmentation results of the baselines and the proposed methods for the distorted testing images. (red: liver, green: right kidney, yellow: left kidney, blue: spleen)

5 Conclusion

We proposed a learning method for U-Net with SSL to make the trained model robust against image distortions such as pixel noise. The proposed method (U-Net with SSL) can construct the segmentation model by extracting features that are invariant to distortions in the paired data. The effectiveness of the proposed approach was experimentally confirmed by using the subset of Sunnybrook Cardiac Data (SCD) and Abdominal organs segmentation dataset.

In this paper, we used only pixel noise as image distortion. We think the approach proposed in this paper can apply to the other types of image distortions. Experiments for such distortions will be our future works.

References

1. Ahmed, S.B., Ali, S.F., Ahmad, J., Adnan, M., Fraz, M.M.: On the frontiers of pose invariant face recognition: a review. Artificial Intelligence Review **53**(4), 2571–2634 (2020)
2. Baevski, A., Hsu, W.N., Xu, Q., Babu, A., Gu, J., Auli, M.: Data2vec: A general framework for self-supervised learning in speech, vision and language. arXiv preprint arXiv:2202.03555 (2022)
3. Balestriero, R., LeCun, Y.: Contrastive and non-contrastive self-supervised learning recover global and local spectral embedding methods. arXiv preprint arXiv:2205.11508 (2022)
4. Bardes, A., Ponce, J., LeCun, Y.: Vicreg: Variance-invariance-covariance regularization for self-supervised learning. arXiv preprint arXiv:2105.04906 (2021)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
7. Golan, R., Jacob, C., Denzinger, J.: Lung nodule detection in ct images using deep convolutional neural networks. In: 2016 international joint conference on neural networks (IJCNN). pp. 243–250. IEEE (2016)
8. Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., Rokovyi, 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. pp. 638–647. Springer (2018)
9. 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 (CVPR'06). vol. 2, pp. 1735–1742. IEEE (2006)
10. Hakim, L., Kavitha, M.S., Yudistira, N., Kurita, T.: Regularizer based on euler characteristic for retinal blood vessel segmentation. Pattern Recognition Letters **149**, 83–90 (2021)
11. Hakim, L., Zheng, H., Kurita, T.: Improvement for single image super-resolution and image segmentation by graph laplacian regularizer based on differences of neighboring pixels. Manuscript submitted for publication (2021)
12. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging **32**(4), 582–596 (2019)
13. Hu, J., Lu, J., Tan, Y.P.: Discriminative deep metric learning for face verification in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1875–1882 (2014)
14. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nürnberger, A., Maier-Hein, K.H., Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (Apr 2021). https://doi.org/https://doi.org/10.1016/j.media.2020.101950, http://www.sciencedirect.com/science/article/pii/S1361841520303145
15. Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A.: Deep mri brain extraction: A 3d convolutional neural network for skull stripping. NeuroImage **129**, 460–469 (2016)
16. Liu, Y., Gong, X., Chen, J., Chen, S., Yang, Y.: Rotation-invariant siamese network for low-altitude remote-sensing image registration. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **13**, 5746–5758 (2020)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
18. Milletari, F., Ahmadi, S.A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., Bötzel, K., et al.: Hough-cnn: deep learning for segmentation of deep brain regions in mri and ultrasound. Computer Vision and Image Understanding **164**, 92–102 (2017)
19. Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D.: Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2021). https://doi.org/10.1109/TPAMI.2021.3059968
20. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621 (2017)
21. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annual Review of Biomedical Engineering **2**(1), 315–337 (2000). https://doi.org/10.1146/annurev.bioeng.2.1.315, https://doi.org/10.1146/annurev.bioeng.2.1.315, pMID: 11701515

22. Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G.: Evaluation framework for algorithms segmenting short axis cardiac mri. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge (07 2009). https://doi.org/10.54294/g80ruo
23. Ramyaa, M., Jonathan, M., Kurita, T.: Supervised learning for convolutional neural network with barlow twins. In: ICANN2022 (submitted) (2022)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
25. Shie, C.K., Chuang, C.H., Chou, C.N., Wu, M.H., Chang, E.Y.: Transfer representation learning for medical image analysis. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 711–714 (2015). https://doi.org/10.1109/EMBC.2015.7318461
26. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis **63**, 101693 (2020). https://doi.org/https://doi.org/10.1016/j.media.2020.101693, https://www.sciencedirect.com/science/article/pii/S136184152030058X
27. Ueda, M., Kanda, K., Miyao, J., Miyamoto, S., Nakano, Y., Kurita, T.: Invariant feature extraction for cnn classifier by using gradient reversal layer. In: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 851–856. IEEE (2021)
28. Yudistira, N., Kavitha, M., Itabashi, T., Iwane, A.H., Kurita, T.: Prediction of sequential organelles localization under imbalance using a balanced deep u-net. Scientific reports **10**(1), 1–11 (2020)
29. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow twins: Self-supervised learning via redundancy reduction. In: International Conference on Machine Learning. pp. 12310–12320. PMLR (2021)
30. Zhou, T., Ruan, S., Canu, S.: A review: Deep learning for medical image segmentation using multi-modality fusion. Array **3**, 100004 (2019)
31. Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L.: A fixed-point model for pancreas segmentation in abdominal ct scans. In: International conference on medical image computing and computer-assisted intervention. pp. 693–701. Springer (2017)