Self-supervised Visual Representation Learning for Histopathological Images

Pengshuai Yang[1], Zhiwei Hong[2], Xiaoxu Yin[1], Chengzhan Zhu[3], and Rui Jiang[1(✉)]

[1] Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division, Beijing National Research Center for Information Science and Technology, Department of Automation, Tsinghua University, Beijing 100084, China
ruijiang@tsinghua.edu.cn

[2] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

[3] Department of Hepatobiliary and Pancreatic Surgery, The Affiliated Hospital of Qingdao University, Qingdao 266000, Shandong, China

Abstract. Self-supervised learning provides a possible solution to extract effective visual representations from unlabeled histopathological images. However, existing methods either fail to make good use of domain-specific knowledge, or rely on side information like spatial proximity and magnification. In this paper, we propose CS-CO, a hybrid self-supervised visual representation learning method tailored for histopathological images, which integrates advantages of both generative and discriminative models. The proposed method consists of two self-supervised learning stages: cross-stain prediction (CS) and contrastive learning (CO), both of which are designed based on domain-specific knowledge and do not require side information. A novel data augmentation approach, stain vector perturbation, is specifically proposed to serve contrastive learning. Experimental results on the public dataset NCT-CRC-HE-100K demonstrate the superiority of the proposed method for histopathological image visual representation. Under the common linear evaluation protocol, our method achieves 0.915 eight-class classification accuracy with only 1,000 labeled data, which is about 1.3% higher than the fully-supervised ResNet18 classifier trained with the whole 89,434 labeled training data. Our code is available at https://github.com/easonyang1996/CS-CO.

Keywords: Self-supervised learning · Stain separation · Contrastive representation learning · Histopathological images

1 Introduction

Extracting effective visual representations from histopathological images is the cornerstone of many computational pathology tasks, such as image retrieval [26,29], disease prognosis [1,30], and molecular signature prediction [7,9,18].

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12902, pp. 47–57, 2021.
https://doi.org/10.1007/978-3-030-87196-3_5

Due to the powerful representation ability, deep learning-based methods gradually replace the traditional handcraft-feature extraction methods and become the mainstream. Deep learning-based methods usually rely on a large amount of labeled data to learn good visual representations, while preparing large-scale labeled datasets is expensive and time-consuming, especially for medical image data. Therefore, to avoid this tedious data collection and annotation procedure, some researchers take a compromise and utilize the ImageNet-pretrained convolutional neural network (CNN) to extract visual representations from medical images [9,30]. However, this compromise ignores not only the data distribution difference [21] between medical and natural images, but also the domain-specific information.

Considering the aforementioned dilemma, self-supervised learning is one of the feasible solutions, which attracts the attention of a growing number of researchers in recent years. Self-supervised learning aims to learn representations from large-scale unlabeled data by solving pretext tasks. In the past few years, research on self-supervised visual representation learning has made great progress. The existing self-supervised learning methods in computer vision field can be categorized into generative model-based approaches and discriminative model-based approaches in the light of the type of associated pretext tasks [16,22]. In earlier times, generative pretext tasks like image inpainting [24] and image colorization [31,32] are proposed to train an autoencoder for feature extraction; discriminative self-supervised pretext tasks such as rotation prediction [10], Jigsaw solving [23], and relative patch location prediction [8], are designed to learn high-level semantic features.

Recently, contrastive learning [13], which also belongs to discriminative approaches, achieves great success in self-supervised visual representation learning. The core idea of contrastive learning is to attract different augmented views of the same image (positive pairs) and repulse augmented views of different images (negative pairs). Based on this core idea, MoCo [14] and SimCLR [4] are proposed for self-supervised visual representation learning, which greatly shrink the gap between self-supervised learning and fully-supervised learning. The success of MoCo and SimCLR shows the superiority of contrastive learning. Furthermore, the following related work BYOL [12] and SimSiam [5] suggest that negative pairs are not necessary for contrastive learning, and they have become the new state-of-the-art self-supervised visual representation learning methods.

Above studies are about natural images. As for medical images, Chen et al. [3] developed a self-supervised learning model based on image context restoration and proved the effectiveness on several tasks. Specific to histopathological images, Gildenblat [11] and Abbet [1] utilized the unique spatial proximity information of whole slide images (WSIs) to establish self-supervised learning methods, relying on the plausible assumption that adjacent patches share similar content while distant patches are distinct. Xie [28] and Sahasrabudhe [25] also proposed self-supervised learning approaches based on magnification information, specially for histopathological image segmentation. However, using such

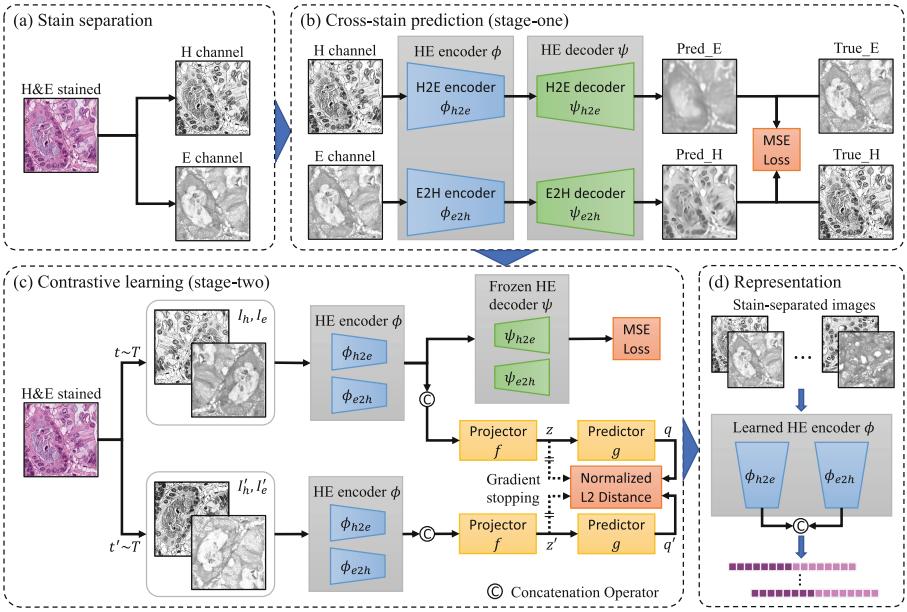


Fig. 1. The framework of the proposed CS-CO method.

side information limits the applicability of these methods. As far as we know, there is still a lack of universal and effective self-supervised learning methods for extracting visual representations from histopathological images.

In this paper, we present CS-CO, a novel hybrid self-supervised histopathological image visual representation learning method, which consists of **C**ross-**S**tain prediction (generative) and **CO**ntrastive learning (discriminative). The proposed method takes advantage of domain specific knowledge and does not require side information like image magnification and spatial proximity, resulting in better applicability. Our major contributions are summarized as follows.

- We design a new pretext task, i.e. cross-stain prediction, for self-supervised learning, aiming to make good use of the domain specific knowledge of histopathological images.
- We propose a new data augmentation approach, i.e. stain vector perturbation, to serve histopathological image contrastive learning.
- We integrate the advantages of generative and discriminative approaches and build a hybrid self-supervised visual representation learning framework for histopathological images.

2 Methodology

2.1 Overview of CS-CO

As illustrated in Fig. 1, the CS-CO method is composed of two self-supervised learning stages, namely cross-stain prediction and contrastive learning, both of which are specially designed for histopathological images. Before self-supervised learning, stain separation is firstly applied to original H&E-stained images to generate single-dye staining results, which are called H channel and E channel images respectively. These stain-separated images are used at the first self-supervised learning stage to train a two-branch autoencoder by solving the novel generative pretext task of cross-stain prediction. Then, at the second stage, the learned HE encoder is trained again in a discriminative contrastive learning manner with the proposed stain vector perturbation augmentation approach. The HE decoder learned at the first stage is also retained as a regulator at the second stage to prevent model collapse. After the two-stage self-supervised learning, the learned HE encoder can be used to extract effective visual representations from stain-separated histopathological images.

2.2 Stain Separation

In histopathology, different dyes are used to enhance different types of tissue components, which can be regarded as domain-specific knowledge implicit in histopathological images. For the commonly used H&E stain, cell nuclei will be stained blue-purple by hematoxylin, and extracellular matrix and cytoplasm will be stained pink by eosin [2]. The stain results of hematoxylin and eosin are denoted as H channel and E channel respectively. To restore single-dye staining results from H&E stain images and reduce the stain variance to some extent, we utilize the Vahadane method [27] for stain separation.

To be specific, for an H&E stained image, let $I \in \mathbb{R}^{m \times n}$ be the matrix of RGB intensity, $V \in \mathbb{R}^{m \times n}$ be the relative optical density, $W \in \mathbb{R}^{m \times r}$ be the stain color matrix, and $H \in \mathbb{R}^{r \times n}$ be the stain concentration matrix, where $m = 3$ for RGB images, r is the number of stains, and n is number of pixels. According to the Beer-Lambert law, the relation between V and H, W can be formulated as Eq. (1), where $I_0 = 255$ for 8-bit RGB images.

$$V = \log \frac{I_0}{I} = WH \tag{1}$$

Then, W and H can be estimated by solving the sparse non-negative matrix factorization problem as Eq. (2) proposed by [27].

$$\begin{aligned} \min_{W,H} \quad & \frac{1}{2}||V - WH||_F^2 + \lambda \sum_{j=1}^{r} ||H(j,:)||_1, \\ \text{s.t.} \quad & W, H \geq 0, \quad ||W(:,j)||_2^2 = 1 \end{aligned} \tag{2}$$

From the estimated stain concentration matrix H, the H channel and E channel images I_h and I_e can be restored as Eq. (3).

$$I_h = I_0 \exp(-H[0,:]), \quad I_e = I_0 \exp(-H[1,:]) \tag{3}$$

2.3 Cross-stain Prediction

At the first stage of the proposed self-supervised learning scheme, a deep neural network is trained to learn visual representations by solving the novel pretext task of cross-stain prediction. The deep model is composed of two independent autoencoders: one is for predicting E channel images from corresponding H channel images (H2E), and the other is the inverse (E2H). We denote the encoder and decoder of H2E branch as ϕ_{h2e} and ψ_{h2e}. The E2H branch is denotated similarly. For the sake of simplicity, we also denote the combination of ϕ_{h2e} and ϕ_{e2h} as HE encoder ϕ, and the combination of ψ_{h2e} and ψ_{e2h} as HE decoder ψ.

As shown in Fig. 1(b), restored H channel and E channel images are input into the two autoencoders separately, and the mean square error (MSE) losses are computed between the predicted and true images in both two branches.

$$I_{pred\_e} = \psi_{h2e}(\phi_{h2e}(I_h)), \quad I_{pred\_h} = \psi_{e2h}(\phi_{e2h}(I_e)) \tag{4}$$

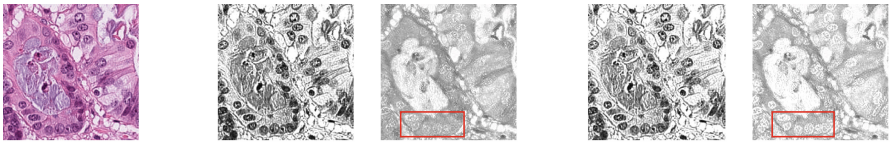
$$\mathcal{L}_{cs} = MSELoss(I_{pred\_e}, I_e) + MSELoss(I_{pred\_h}, I_h) \tag{5}$$

By solving this proposed generative pretext task, the HE encoder can capture low-level general features from histopathological images. In addition, based on the characteristics of H&E stain mentioned in Sect. 2.2, the HE encoder is also expected to be sensitive to details which imply the correlation between nuclei and cytoplasm.

2.4 Contrastive Learning

Based on the two-branch autoencoder learned at the first stage, we adopt contrastive learning at the second stage to learn discriminative high-level features. Inspired by [5], we reorganize our model into the Siamese architecture, which is composed of the HE encoder ϕ, a projector f, and a predictor g. All parameters are shared across two branches. The HE decoder ψ is also kept in one branch as an untrainable regulator to prevent model collapse. The weights of ϕ and ψ learned at the first stage are loaded as initialization.

During contrastive learning, a pair of H channel and E channel images of the same H&E stained image is regarded as one data sample. As shown in Fig. 1(c), for each data sample, two transformations t and t' are sampled from the transformation family T for data augmentation. The transformation family includes stain vector perturbation, RandomFlip, RandomResizedCrop, and GaussianBlur. After transformation, the derived two randomly augmented views (I_h, I_e) and (I'_h, I'_e) are input into the Siamese network separately. For each augmented view, the contained H channel and E channel images are firstly encoded by ϕ_{h2e} and ϕ_{e2h} respectively. The outputs are pooled and concatenated together



(a) H&E stain (b) Stain separation (c) Stain vector perturbation

Fig. 2. Stain vector perturbation. (a) The original H&E stain image. (b) Stain separation results using Vahadane [27] method. (c) Stain separation results with proposed stain vector perturbation ($\sigma = 0.05$). The red box outlines the visible differences. (Color figure online)

as one vector. Subsequently, the projector f and predictor g are applied to the vector sequentially.

For two augmented views, denoting the outputs of the projector f as $z \triangleq f(\phi(I_h, I_e))$ and $z' \triangleq f(\phi(I'_h, I'_e))$ and the outputs of predictor g as $q \triangleq g(z)$ and $q' \triangleq g(z')$, we force q to be similar to z' and q' to be similar to z by minimizing the symmetrized loss:

$$\mathcal{L}_{co} = \frac{1}{2}||\tilde{q} - \tilde{z'}||_2^2 + \frac{1}{2}||\tilde{q'} - \tilde{z}||_2^2 \tag{6}$$

where $\tilde{x} \triangleq \frac{x}{||x||_2}$ and $||\cdot||_2$ is ℓ_2-norm. z and z' are detached from the computational graph before calculating the loss.

As for the frozen pretrained HE decoder ψ, it constrains the generalization of features extracted by the HE encoder ϕ by minimizing Eq. (5), so as to ensure no collapse occurs on the HE encoder ϕ. The total loss is formulated as Eq. (7), where α is the weight coefficient (in our implementation, $\alpha = 1.0$).

$$\mathcal{L}_{tot} = \mathcal{L}_{co} + \alpha\mathcal{L}_{cs} \tag{7}$$

Stain Vector Perturbation. Since the input images are gray, many transformations of colorful image cannot be used for contrastive learning. To guarantee the strength of transformation, we customize a new data augmentation approach called stain vector perturbation for histopathological images. Inspired by the error of stain vector estimation in stain separation, we disturb elements of the estimated W with $\epsilon \sim N(0, \sigma^2)$ to obtain the perturbed stain vector matrix W'. With W', another stain concentration matrix H' can be derived, and the corresponding H channel and E channel images can be restored from H'. The results of stain vector perturbation are shown in Fig. 2.

2.5 Representation Extraction

After two-stage self-supervised learning, the learned HE encoder ϕ can be used for visual representation extraction. As shown in Fig. 1(d), for an H&E stained

image, the corresponding H and E channel images are firstly restored via stain separation and then input into the learned HE encoder ϕ. The outputs are pooled and concatenated together as the extracted visual representation.

3 Experimental Results

Dataset. We evaluate our proposed CS-CO method on the public dataset NCT-CRC-HE-100K [17]. The dataset contains nine classes of histopathological images of human colorectal cancer and healthy tissue. The predefined training set contains 100,000 images and the test set contains 7180 images. The overall nine-class classification accuracy on test set is 0.943 as reported in [19], which is achieved by fully-supervised learning with VGG19. It is worth noting that we exclude images belonging to the background (BACK) class for training and testing when we evaluate visual representation learning methods. The reason is that background is always non-informative and can be easily distinguished by simple threshold-based methods. The final sizes of training and test set are 89,434 and 6333 respectively, and the eight-class classification accuracy on the test set is reported as the evaluation metric.

Implementation Details. For the proposed CS-CO method, we use ResNet18 [15] as the backbone of the encoders ϕ_{h2e} and ϕ_{e2h}. The decoders ψ_{h2e} and ψ_{e2h} are composed of a set of upsampling layers and convolutional layers. The projector f and predictor g are both instantiated by the multi-layer perceptron (MLP), which consists of a linear layer with output size 4096 followed by batch normalization, rectified linear units (ReLU), and a final linear layer with output dimension 256. At the first training stage, we use SGD to train our model on the whole training set. The batch size is 64 and the learning rate is 0.01. At the second stage, for fast implementation, we train the model with Adam on 10,000 randomly selected training data. The batch size is 96, the learning rate is 0.001, and the weight decay is 1×10^{-6}. Early stopping is used at both stages for avoiding over-fitting.

According to the common evaluation protocol [20], a linear classifier is trained for evaluating the capacity of representation. The linear classifier is implemented by a single linear layer and trained with SGD. The batch size is 32 and the learning rate is 0.001. Early stopping is also used for avoiding over-fitting.

Method Comparison. We firstly train a ResNet18 model with the whole eight-class training data to establish the fully-supervised baseline. Then, we choose three types of methods to compared with our proposed CS-CO method. The first type contains two fixed ResNet18 models, one is random initialized, and the other is pretrained on ImageNet [6]. The second type contains two state-of-the-art contrastive learning methods: BYOL [12] and Simsiam [5]. The last type also contains two methods specifically proposed for medical images by Chen et al. [3] and Xie et al. [28] respectively. Except that the two ResNet18 models of

Table 1. Linear evaluation results (5-fold cross-validation) with different size (n) of training data.

Fully-supervised ResNet18		0.903 ± 0.015	
Methods	$n = 100$	$n = 1000$	$n = 10000$
Random initialized ResNet18	0.134 ± 0.050	0.181 ± 0.070	0.427 ± 0.003
ImageNet pretrained ResNet18	0.628 ± 0.040	0.802 ± 0.012	0.844 ± 0.002
BYOL [12]	0.811 ± 0.011	0.898 ± 0.007	0.891 ± 0.005
SimSiam [5]	0.797 ± 0.029	0.897 ± 0.004	0.890 ± 0.005
Chen's method [3]	0.215 ± 0.067	0.661 ± 0.014	0.711 ± 0.003
Xie's method [28]	0.109 ± 0.042	0.507 ± 0.007	0.586 ± 0.009
CS-CO	$\mathbf{0.834 \pm 0.018}$	$\mathbf{0.915 \pm 0.004}$	$\mathbf{0.914 \pm 0.002}$

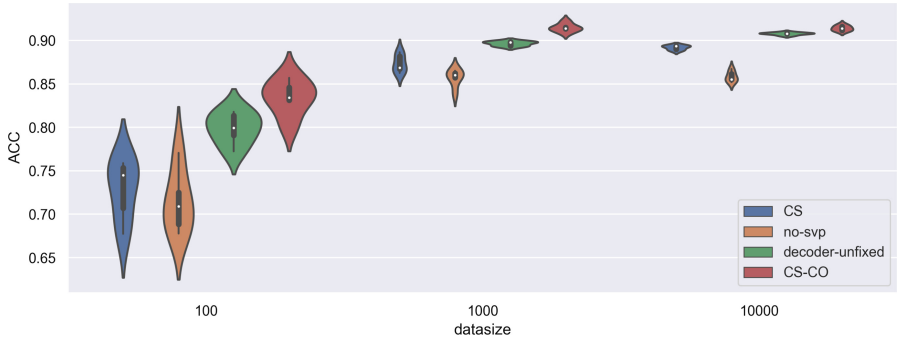


Fig. 3. Ablation study results (5-fold cross-validation).

the first type don't need to be trained, both our CS-CO method and the latter two types of self-supervised learning methods are firstly trained using the whole training data as unlabeled data.

Rather than using the whole training set, we randomly sample 100, 1000, and 10000 data from the training set and extract their visual representations with each method for the following linear classifier training. In this way, the impact of large data size can be stripped, and the classification accuracies on the test set can more purely reflect the representation capacity of each method. As shown in Table 1, our proposed CS-CO method demonstrates superior representation capacity compared to other methods. Furthermore, with only 1,000 labeled data and the linear classifier, our CS-CO method even outperforms the fully-supervised ResNet18 which is trained on the whole training set.

Ablation Study. We conduct ablation studies to explore the role of the following three key components of the proposed CS-CO method. 1) Contrastive learning: To verify whether the contrastive learning enhances the visual representation capacity, we do linear evaluation on the CS model, which is only trained by solving the

cross-stain prediction task. In the cases of different amount of training data, the average test accuracies of CS model are 0.782, 0.873, and 0.892, which shows obvious gaps from the original CS-CO model. 2) Stain-vector perturbation: To demonstrate the effectiveness of stain-vector perturbation, we remove it from the transformation family of contrastive learning, and train another CS-CO model which is denoted as *no-SVP*. As shown in Fig. 3, the performance of *no-SVP* model is even worse than CS model, which suggests that stain-vector perturbation is crucial for contrastive learning. 3) Frozen-decoder: We also make the HE decoder trainable at the second training stage to train the CS-CO model, which is denoted as *decoder-unfixed*. As Fig. 3 shows, the *decoder-unfixed* model doesn't collapse but performs slightly worse than the original CS-CO model.

4 Conclusion

In this paper, we have proposed a novel hybrid self-supervised visual representation learning method specifically for histopathological images. Our method draws advantages from both generative and discriminative models by solving the proposed cross-stain prediction pretext task and doing contrastive learning with the proposed stain-vector perturbation augmentation approach. The proposed method makes good use of domain-specific knowledge and has good versatility. Linear evaluation results on dataset NCT-CRC-HE-100K suggest that our method outperforms current state-of-the-art self-supervised visual representation learning approaches. In future work, we intend to use the representations extracted by the proposed CS-CO method to solve downstream tasks such as disease prognosis and molecular signature prediction, so as to further prove the effectiveness of the proposed method in practice.

Acknowledgements. This work was partially supported by the National Key Research and Development Program of China (No. 2018YFC0910404), the National Natural Science Foundation of China (Nos. 61873141, 61721003), the Shanghai Municipal Science and Technology Major Project (No. 2017SHZDZX01), the Tsinghua-Fuzhou Institute for Data Technology, the Taishan Scholars Program of Shandong Province (No. 2019010668), and the Shandong Higher Education Young Science and Technology Support Program (No. 2020KJL005).

References

1. Abbet, C., Zlobec, I., Bozorgtabar, B., Thiran, J.-P.: Divide-and-rule: self-supervised learning for survival analysis in colorectal cancer. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 480–489. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_46
2. Chan, J.K.: The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology. Int. J. Surg. Pathol. **22**(1), 12–32 (2014)
3. Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., Rueckert, D.: Self-supervised learning for medical image analysis using image context restoration. Med. Image Anal. **58**, 101539 (2019)

4. 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)
5. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15750–15758 (2021)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
7. Ding, K., Liu, Q., Lee, E., Zhou, M., Lu, A., Zhang, S.: Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 294–304. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_29
8. Doersch, C., Gupta, A., Efros, A.A.: Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1422–1430 (2015)
9. Fu, Y., et al.: Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat. Cancer **1**(8), 800–810 (2020)
10. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. In: International Conference on Learning Representations (2018)
11. Gildenblat, J., Klaiman, E.: Self-supervised similarity learning for digital pathology. arXiv preprint arXiv:1905.08139 (2019)
12. Grill, J.B., et al.: Bootstrap your own latent: a new approach to self-supervised learning. arXiv preprint arXiv:2006.07733 (2020)
13. 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 2006), vol. 2, pp. 1735–1742. IEEE (2006)
14. 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)
15. 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, pp. 770–778 (2016)
16. Jing, L., Tian, Y.: Self-supervised visual feature learning with deep neural networks: a survey. IEEE Trans. Pattern Anal. Mach. Intell. (2020)
17. Kather, J.N., Halama, N., Marx, A.: 100,000 histological images of human colorectal cancer and healthy tissue. https://doi.org/10.5281/zenodo.1214456
18. Kather, J.N., et al.: Pan-cancer image-based detection of clinically actionable genetic alterations. Nat. Cancer **1**(8), 789–799 (2020)
19. Kather, J.N., et al.: Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med. **16**(1), e1002730 (2019)
20. Kolesnikov, A., Zhai, X., Beyer, L.: Revisiting self-supervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1920–1929 (2019)
21. Liu, Q., Xu, J., Jiang, R., Wong, W.H.: Density estimation using deep generative neural networks. Proc. Natl. Acad. Sci. **118**(15), e2101344118 (2021)
22. Liu, X., et al.: Self-supervised learning: generative or contrastive **1**(2). arXiv preprint arXiv:2006.08218 (2020)

23. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 69–84. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_5
24. 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)
25. Sahasrabudhe, M., et al.: Self-supervised nuclei segmentation in histopathological images using attention. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 393–402. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_38
26. Shi, X., Sapkota, M., Xing, F., Liu, F., Cui, L., Yang, L.: Pairwise based deep ranking hashing for histopathology image classification and retrieval. Pattern Recogn. **81**, 14–22 (2018)
27. Vahadane, A., et al.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging **35**(8), 1962–1971 (2016)
28. Xie, X., Chen, J., Li, Y., Shen, L., Ma, K., Zheng, Y.: Instance-aware self-supervised learning for nuclei segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 341–350. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_33
29. Yang, P., et al.: A deep metric learning approach for histopathological image retrieval. Methods **179**, 14–25 (2020)
30. Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med. Image Anal. **65**, 101789 (2020)
31. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 649–666. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_40
32. Zhang, R., Isola, P., Efros, A.A.: Split-brain autoencoders: unsupervised learning by cross-channel prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1058–1067 (2017)