Contrastive Learning of Single-Cell Phenotypic Representations for Treatment Classification

Alexis Perakis[1]([✉]), Ali Gorji[2], Samriddhi Jain[2], Krishna Chaitanya[3], Simone Rizza[2], and Ender Konukoglu[3]

[1] KOF Swiss Economic Institute, ETH Zurich, Zurich, Switzerland
perakis@kof.ethz.ch
[2] ETH Zurich, Zurich, Switzerland
[3] Computer Vision Lab, ETH Zurich, Zurich, Switzerland

Abstract. Learning robust representations to discriminate cell phenotypes on microscopy images is important for drug discovery. Drug development efforts typically analyse thousands of cell images to screen for potential treatments. Early works focus on creating hand-engineered features from these images or learn such features with deep neural networks in a fully or weakly-supervised framework. Both require prior knowledge or labelled datasets. Therefore, subsequent works propose unsupervised approaches based on generative models to learn these representations. Recently, representations learned with self-supervised contrastive loss-based methods have yielded state-of-the-art results on various imaging tasks compared to earlier unsupervised approaches. In this work, we leverage a contrastive learning framework to learn appropriate representations from single-cell fluorescent microscopy images for the task of Mechanism-of-Action classification. The proposed work is evaluated on the annotated BBBC021 dataset, and we obtain state-of-the-art results in NSC, NCSB and drop metrics for an unsupervised approach. We observe an improvement of 10% in NCSB accuracy and 11% in NSC-NSCB drop over the previously best unsupervised method. Moreover, the performance of our unsupervised approach ties with the best supervised approach. Additionally, we observe that our framework performs well even without post-processing, unlike earlier methods. With this, we conclude that one can learn robust cell representations with contrastive learning. We make the code available on GitHub (https://github.com/SamriddhiJain/SimCLR-for-cell-profiling).

Keywords: Fluorescent microscopy · Phenotypes · Profiling · Cell images · Cell representations · Unsupervised learning · Contrastive learning

1   Introduction

An effective approach in the field of drug discovery is to relate treatments under development to existing ones. Comparing an unknown to a known treatment enables finding desired similarities and avoiding unwanted effects. Cells of interest are exposed to chemical compounds and then imaged using various microscopy techniques. A cell's response mechanism upon treatment is called Mechanism-of-Action (MOA). Such response mechanisms modify a cell's phenotype (morphology) to various degrees. To compare different treatments we can classify them into different MOAs. To perform this classification, we require cell representations that accurately capture the cell morphology. Morphological cell profiling uses image recognition techniques to construct these cell representations. Typically, cell profiling-based MOA classification is done on thousands of images in a transductive learning setting in the literature, where the whole data are included in the training and evaluation process.

In the literature, many works have focused on automating the tedious task of learning meaningful single-cell representations and utilize them downstream for MOA assignment. Initial works like [24,27] rely on creating expert engineered features. Later, deep neural networks are used to learn these features. Early works making use of neural networks rely on fully supervised approaches [11,21] to learn these features, where [11] implements multi-scale convolutional neural networks to extract cell morphology directly from images. Alternatively, [3] proposes a weakly supervised learning method. Other works as in [12,28] use the metadata information to devise pseudo-labels for the network supervision. All the fully or weakly supervised methods above suffer from the requirement of labeled sets, and annotating such images by experts leads to high costs and time-consuming efforts. Even in [12,28], the labels acquired from metadata can be prone to imprecise labeling due to the nature of the data and treatment analysis techniques. Some other works use transfer learning techniques [1,26]. With these approaches, it may not be feasible to acquire appropriate labeled datasets for pre-training.

These reasons encourage the usage of unsupervised learning approaches that provide the following advantages: large unlabeled datasets can be directly used in training, and no pre-training labelled datasets are required. In approaches as above [5,13,19,22], the aim is to learn robust single-cell representations. In [19] the authors perform clustering on the learned cell representations and use the cluster assignments as labels, where the implementation is similar to [5]. Alternatively, some works use generative models like variational autoencoders [22] or generative adversarial networks [13] to learn such representations.

For representation learning, many recent works propose self-supervised learning methods using unlabeled data. Recently, self-supervised approaches based on a contrastive loss [14] have yielded state of the art performance for imaging tasks such as classification, object detection, segmentation [7,15,17,33] on benchmark natural image datasets as well as for medical imaging tasks on MR [6], CT [34,35], X-ray [2,29,32,36], dermatology [2], electron microscopy [18] image datasets and ECG signals [20]. These contrastive loss based approaches outper-

form traditional self-supervision based pretext tasks (e.g., rotation [10], inpainting [25]) and generative models [8,9].

In cell profiling, the datasets contain thousands or millions of images. Unsupervised methods are promising and provide more viable solutions for such applications than supervised methods. Hence, in the proposed work, we leverage the popular contrastive learning framework [7] for learning robust cell representations using only unlabeled data. In contrastive loss-based learning [14], representations are learned by contrasting positive examples to negative examples. As in [7], we train the network to pull the representations of positive examples to be close in the latent space and push the negative examples representations to be far away from positive examples.

Our contributions are:

– We are the first to learn single-cell representations using a contrastive learning framework in an unsupervised setting.
– For a cell profiling dataset, we evaluate and find the most important components and hyper-parameters used in the contrastive framework such as: (i) encoder network size, (ii) data augmentation strategies, (iii) projection head size, (iv) batch size value and (v) temperature parameter.
– We achieve state-of-the-art NSCB accuracy and NSC-NSCB drop for an unsupervised method. Our unsupervised results in NSC, NSCB, and drop metrics match the state-of-the-art for a supervised method.
– The learned single-cell representations perform well even without any postprocessing, unlike earlier works.

2 Methods

We divide this section into two parts: (a) representation learning and (b) MOA classification. In the first part, we learn representations in a contrastive learning framework as proposed in [7]. In the second part, we use these representations for the downstream task of Mechanism-of-Action classification of the treatment profiles.

(a) Representation Learning: In the contrastive framework, we learn a global representation for each input image as illustrated in Fig. 1. We follow [7] and sample a mini-batch of images of size N from the whole dataset X. Then, for each sampled image x we apply two random transformations t_i and t_j (sampled from a set of transformations T). The transformed images are denoted by $\tilde{x}_i = t_i(x)$ and $\tilde{x}_j = t_j(x)$ to obtain $2N$ images in the batch, as shown in Fig. 1. For the encoder network, we use a ResNet [16] denoted by f to get the representation h where $h_i = f(x_i)$, which is followed by a projection head g. The output latent representation is given by z where $z_i = g(f(x_i))$. The two transformed images arising from a given image x are denoted as the positive pair. The remaining $2(N-1)$ images in the batch act as the negative pairs and form the negative images set Ω^-. The contrastive loss applied on the positive pairs of output latent

representations is defined as follows:

$$l(\tilde{x}_i, \tilde{x}_j) = -\log \frac{e^{\text{sim}(z_i,z_j)/\tau}}{e^{\text{sim}(z_i,z_j)/\tau} + \sum_{x_n \in \Omega^-} e^{\text{sim}(z_i,g(f(x_n)))/\tau}} \tag{1}$$

where z_i, z_j are the output latent representations of the positive pair (x_i, x_j) and x_n are the corresponding negative images from the set Ω^-. τ denotes the temperature parameter. The similarity between two representations is computed using cosine similarity, which is defined as $\text{sim}(z_i, z_j) = z_i^T z_j / \|z_i\| \|z_j\|$.
The net contrastive loss across all positive pairs in the batch is given below:

$$L_{net} = \frac{1}{2N} \sum_{k=1}^{N} [l(t_i(x_k), t_j(x_k)) + l(t_j(x_k), t_i(x_k))] \tag{2}$$

By optimizing the loss L_{net}, we enable the network to learn representations of a positive pair for a given image to be similar under different transformations such as crop, rotation, color jitter, etc. Also, they should be dissimilar to representations of the remaining images in the batch that constitute the negative set. With this optimization, we aim to learn robust and meaningful global representations $h = f(.)$ that can be used for downstream tasks.
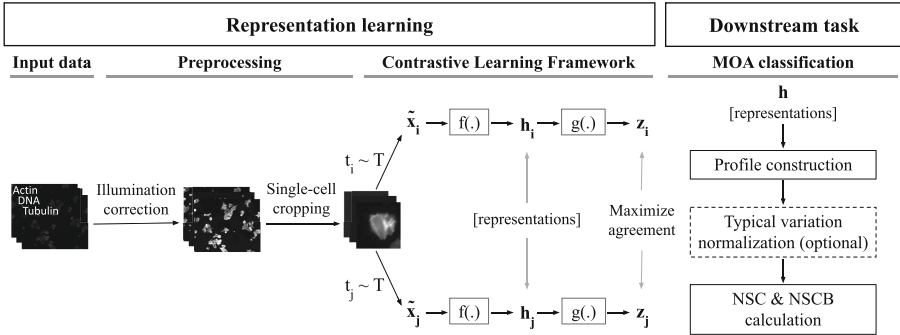
Fig. 1. Our experimental pipeline can be split into two parts: representation learning and downstream task. In representation learning, we feed single-cell images into the contrastive learning framework [7], where single-cell representations are extracted. Subsequently, these representations are aggregated into treatment profiles and postprocessed (optional step) in the downstream task. We use the resulting profiles for MOA classification and calculate NSC and NSCB accuracies.

(b) MOA Classification: Our proposed method is evaluated on a downstream MOA classification task. The final latent representations z_i and projection head $g(.)$ are only used for the representation learning stage and are discarded for the classification step. The Mechanism-of-Action classification is performed on the learned cell representations h_i that are output from the encoder network.

We use two widely applied procedures for evaluating how MOA are assigned to treatments: not-same-compound (NSC) matching from [24] and not-same-compound-and-batch (NSCB) matching from [1]. In a first step, single-cell representations are aggregated into *treatment profiles* as described in [24]. This aggregation results in a treatment profile vector for each treatment in the dataset.

In the second step, one MOA is assigned to each treatment profile. Both NSC and NSCB scores are 1st nearest neighbour MOA classification accuracies where the distance measure used is cosine distance. Treatments are assigned an MOA label one at a time. In NSC matching, for a given treatment, we search the representation space for the nearest neighbour not containing the same compound. Recall that a treatment is a compound-concentration pair. In NSCB matching, the search further excludes all treatments from the same experimental batch.

3   Datasets and Network Details

Dataset: Our work uses the annotated part of the BBBC021 dataset [4] available from the Broad Bioimage Benchmark Collection [23]. BBBC021 consists of multi-channel images captured from human MCF-7 breast cancer cells exposed to chemical compounds for 24 h. Cells are imaged by fluorescent microscopy. The 3 grey-scale channels represent DNA, B-tubulin and F-actin. The proposed method is evaluated on the subset of BBBC021 that has previously been labelled for MOAs by [24]. There are 12 distinct MOA present in this subset for a total of 103 treatments and 38 compounds. Each treatment corresponds to a compound-concentration combination. The control cells are treated with DMSO (dimethyl sulfoxide). We solely use the ground truth annotated part of BBBC021 for our study, in line with the evaluation strategy applied in all the works in the literature [1,3,11–13,19,21,22,24,26,27,31].

Data Pre-processing: We apply the illumination correction algorithm [27] on the original images as performed in [3]. Then, we crop all the single-cell instances from these images to fixed image dimensions of 96×96, based on the cell locations introduced by [24].

Post-processing: We use typical variation normalization (TVN) [3,19,22,31] to reduce batch effects and improve profiles. It comprises of a *whitening* step and a *correlation alignment* (CORAL) [30] step. We also evaluate the effects of using only whitening [19,22,31] as well as no post-processing.

Network Details: We choose ResNet [16] for the encoder network $f(.)$ and MLP layers for the projection head network $g(.)$.

Training Details: We train the network with the following parameter settings. We choose Adam as the optimizer with a learning rate of $3e^{-4}$ and a weight

decay of $1e^{-5}$. We run the initial hyper-parameter evaluation for 150 epochs on a subset of 15% of the annotated images. We run the final experiments with the optimal set of hyper-parameters for up to 600 epochs on the complete set of annotated data. All our experiments are performed on a Nvidia Titan Xp GPU.

4   Experiments

Experimental Setup: The part of BBBC021 annotated for MOAs contains 2'526 original images which translates into 454'793 single-cells. First, we tune our hyperparameters on a subset of the annotated BBBC021 dataset containing 15% of the images as performed by [21]. This significantly reduces our training time and compute resources. In a second step, we train our framework on the whole annotated BBBC021 dataset and report our final results. Note that we perform training and evaluation in a transductive learning setting on the whole annotated data in line with the evaluation strategy used in the earlier works.

Evaluation: We use NSC, NSCB scores and the NSC-NSCB drop to measure the performance.

Ablation Study of Hyper-parameters: For the ablation study of hyper-parameters, we only use 15% of the annotated images. We report all results for these experiments with TVN post-processing. (Refer to the Supplementary for results using only whitening or without post-processing.) We evaluate the following hyper-parameters to analyze their effect on downstream performance.

 (i) **Encoder Network Size** $f(.)$: For the encoder, the following ResNet sizes are evaluated: ResNet18, ResNet50, ResNet101. The remaining hyper-parameters are investigated for a ResNet50 encoder as it yielded the best results.
 (ii) **Data Augmentation strategy**: We explore different data augmentations (T) such as crop, flip, rotations by 90°, color jitter, grey-distortion, and Gaussian blur. In the default setting, we have all augmentations. We experiment by removing one augmentation at a time from this main set of augmentations to analyze the importance of each.
(iii) **Projection head** $g(.)$: We explore three types of projection heads: the identity, a linear projection and a non-linear projection (two layer MLP).
(iv) **Batch Size**: We evaluate the following batch sizes used for each training iteration: 64, 128, 256.
 (v) **Temperature coefficient**: As done in [7], we evaluate the effect of temperature for the following values: $\tau = 0.05, 0.1, 0.5, 1$ for different combinations of other hyper-parameters.

Final Experiments and Comparison: We choose the hyper-parameters for the final set of experiments from the ablation results. They are a ResNet50 as the encoder network, a two-layer MLP as the projection head, an augmentation strategy without the grey distortion, a batch size of 256 and a temperature parameter of $\tau = 0.5$. Here, the evaluation is performed in a transductive learning setting on the whole annotated dataset as described earlier.

TVN Post-processing : We also evaluate if there is any difference in performance with and without TVN post-processing applied to the learned representations.

5   Results

Ablation Study of Hyper-parameters
(i) Encoder size: In Table 1, we present the results for different ResNet architectures evaluated as the encoder network. We observe that ResNet18 slightly outperforms both ResNet50 and ResNet101. However, when we compare ResNet50 to ResNet18, we observe smaller values for NSC-NSCB drop with and without post-processing for ResNet50 (Refer to Table 4 in Supplementary). Hence, we choose ResNet50 for the remaining hyper-parameter evaluation.
(ii) Data augmentation strategies: In Table 1, we present our results for the different augmentation strategies adopted. We drop one augmentation at a time to analyze its importance. We observe that dropping the grey color distortion significantly improves NSC and NSCB performance. This could be due to grey-scale averaging over the three channels. It may not be meaningful for such datasets as they contain relatively independent pixel intensities in each channel, with different channels capturing different parts of a cell. Typical RGB channels contain dependent information. The most important augmentation was found to be cropping followed by color jitter. This is also observed in earlier works [7] on natural images.

Other hyper-parameter evaluations are presented in the Supplementary Material in Tables 5, 6 and 7. We summarize our findings to be the following: (iii) a two-layer MLP (non-linear projection) and linear projection yield similar NSC and NSCB scores. They both yield higher scores than the identity. (iv) A batch size of 256 yields higher NSC and NSCB scores on average over batch sizes of 64 and 128. (v) The temperature coefficients of 0.1 and 0.5 yield higher scores compared to a large value of 1 and very small value of 0.05.

Table 1. Model performance is shown for different encoder network sizes ((a) f size) and for one augmentation removed at a time ((b) Augmentation removal) under TVN post-processing conditions. Jitter, Grey, and Blur refer to color-jitter, Grey color distortion, Gaussian Blur, respectively.

Metric	(a) f size			(b) Augmentation removal						
	ResNet18	ResNet50	ResNet101	None	Crop	Flip	Rotation	Jitter	Grey	Blur
NSC	95%	93%	92%	94%	83%	93%	94%	91%	96%	92%
NSCB	91%	91%	90%	91%	70%	88%	87%	88%	94%	88%

Final Experiments and Comparison: In Table 2, we present our results from the final set of experiments on the whole annotated dataset in a transductive setting as evaluated in earlier works. We observe that the proposed unsupervised contrastive learning framework yields better results than earlier unsupervised works with an improvement over the best method [19] of 10% and 11% in NSCB scores and drop respectively. We also observe that the proposed framework yields similar scores to supervised counterparts [1] where large number of annotations or suitable pre-training datasets are required to achieve such high performance.

Table 2. MOA classification accuracy. * indicates results without post-processing.

Supervised				Unsupervised			
Method	NSC	NSCB	Drop	Method	NSC	NSCB	Drop
Ljosa et al. [24]	94%	77%	17%	Janssens et al. [19]	**97%**	**85%**	12%
Singh et al. [27]	90%	85%	5%	Lafarge et al. [22]	93%	82%	**11%**
Ando et al. [1]	**96%**	**95%**	**1%**	Lafarge et al. [22] *	92%	72%	20%
Pawlowski et al. [26]	91%	NA	NA	Our work *	**97%**	94%	3%
Caicedo et al. [3]	95%	89%	6%	Our work + Whitening	**96%**	**95%**	**1%**
				Our work + TVN	**97%**	92%	5%

TVN Post-processing: We observe that our method performs well even without applying post-processing such as TVN or whitening. Earlier works are sensitive to the post-processing step where a 10% decrease in NSCB is observed for [22]. Our work has only a 1% decrease in NSCB when removing the whitening step, as shown in Table 2.

t-SNE Plot: Figure 2 shows a t-SNE visualization of the treatment profiles obtained with our final experiment. Here, the treatments are classified into the 12 MOAs available. In Fig. 2, we can observe that only 4 out of 103 and 5 out of 92 treatments are classified incorrectly during NSC and NSCB MOA assignment respectively. NSC and NSCB mis-classifications are marked with black squares and black diamonds respectively.
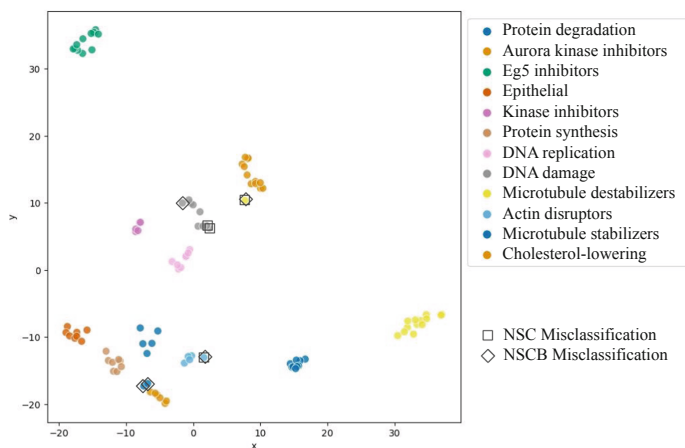
Fig. 2. t-SNE plot for treatment profiles subject to whitening post-processing.

6   Conclusion

In this work, we conclude that using an unsupervised approach with contrastive learning on single-cell images leads to excellent representations in morphological cell profiling. With this framework, we demonstrate state of the art results in an unsupervised learning setting for the downstream task of MOA classification. In the unsupervised setting, our NSCB score of 95% for MOA matching is the highest classification accuracy reported and similarly the drop of 1% between our NSC and NSCB scores is the best reported. Furthermore, our unsupervised results are identical to the state of the art transfer learning approach [1] and higher than the supervised approach [3], both relying on labels. We perform an ablation study of hyper-parameters and conclude that encoder size as well as data augmentations are the most crucial hyper-parameters for obtaining maximum improvements on cell profiling datasets. Finally, we also show that the performance of the resulting representations does not deteriorate even when removing post-processing techniques.

References

1. Ando, D.M., McLean, C.Y., Berndl, M.: Improving phenotypic measurements in high-content imaging screens. BioRxiv, p. 161422 (2017)
2. Azizi, S., et al.: Big self-supervised models advance medical image classification. arXiv preprint arXiv:2101.05224 (2021)
3. Caicedo, J.C., McQuin, C., Goodman, A., Singh, S., Carpenter, A.E.: Weakly supervised learning of single-cell feature embeddings. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9309–9318 (2018)
4. Caie, P.D., et al.: High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol. Cancer Ther. **9**(6), 1913–1926 (2010)

5. 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 (ECCV), pp. 132–149 (2018)

6. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: Advances in Neural Information Processing Systems, vol. 33 (2020)

7. 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)

8. Donahue, J., Krähenbühl, P., Darrell, T.: Adversarial feature learning. ICLR (2017)

9. Donahue, J., Simonyan, K.: Large scale adversarial representation learning. In: Advances in Neural Information Processing Systems, vol. 32 (2019)

10. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728 (2018)

11. Godinez, W.J., Hossain, I., Lazic, S.E., Davies, J.W., Zhang, X.: A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics **33**(13), 2010–2019 (2017)

12. Godinez, W.J., Hossain, I., Zhang, X.: Unsupervised phenotypic analysis of cellular images with multi-scale convolutional neural networks. BioRxiv, p. 361410 (2018)

13. Goldsborough, P., Pawlowski, N., Caicedo, J.C., Singh, S., Carpenter, A.E.: CytoGAN: generative modeling of cell images. BioRxiv, p. 227645 (2017)

14. 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)

15. 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)

16. 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)

17. Hjelm, R.D., et al.: Learning deep representations by mutual information estimation and maximization. ICLR (2019)

18. Huang, G.B., Yang, H.F., Takemura, S.y., Rivlin, P., Plaza, S.M.: Latent feature representation via unsupervised learning for pattern discovery in massive electron microscopy image volumes. arXiv preprint arXiv:2012.12175 (2020)

19. Janssens, R., Zhang, X., Kauffmann, A., de Weck, A., Durand, E.Y.: Fully unsupervised deep mode of action learning for phenotyping high-content cellular images. BioRxiv, p. 215459 (2020)

20. Kiyasseh, D., Zhu, T., Clifton, D.A.: CLOCS: contrastive learning of cardiac signals. arXiv preprint arXiv:2005.13249 (2020)

21. Kraus, O.Z., Ba, J.L., Frey, B.J.: Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics **32**(12), 52–59 (2016). https://doi.org/10.1093/bioinformatics/btw252

22. Lafarge, M.W., Caicedo, J.C., Carpenter, A.E., Pluim, J.P.W., Singh, S., Veta, M.: Capturing single-cell phenotypic variation via unsupervised representation learning. In: International Conference on Medical Imaging with Deep Learning, pp. 315–325. PMLR (2019)

23. Ljosa, V., Sokolnicki, K., Carpenter, A.E.: Annotated high-throughput microscopy image sets for validation. Nat. Meth. **9**, 637 (2012). https://doi.org/10.1038/nmeth.2083

24. Ljosa, V., et al.: Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J. Biomol. Screen. **18**(10), 1321–1329 (2013)
25. 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)
26. Pawlowski, N., Caicedo, J.C., Singh, S., Carpenter, A.E., Storkey, A.: Automating morphological profiling with generic deep convolutional networks. BioRxiv, p. 085118 (2016)
27. Singh, S., Bray, M.A., Jones, T., Carpenter, A.: Pipeline for illumination correction of images for high-throughput microscopy. J. Microsc. **256**(3), 231–236 (2014)
28. Spiegel, S., Hossain, I., Ball, C., Zhang, X.: Metadata-guided visual representation learning for biomedical images. BioRxiv, p. 725754 (2019)
29. Sriram, A., et al.: COVID-19 prognosis via self-supervised representation learning and multi-image prediction. arXiv preprint arXiv:2101.04909 (2021)
30. Sun, B., Feng, J., Saenko, K.: Correlation Alignment for Unsupervised Domain Adaptation. In: Csurka, G. (ed.) Domain Adaptation in Computer Vision Applications. ACVPR, pp. 153–171. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58347-1_8
31. Tabak, G., Fan, M., Yang, S., Hoyer, S., Davis, G.: Correcting nuisance variation using Wasserstein distance. PeerJ **8**, e8594 (2020)
32. Vu, Y.N.T., Wang, R., Balachandar, N., Liu, C., Ng, A.Y., Rajpurkar, P.: Contrastive learning leveraging patient metadata improves representations for chest x-ray interpretation. arXiv preprint arXiv:2102.10663 (2021)
33. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3733–3742 (2018)
34. Xie, Y., Zhang, J., Liao, Z., Xia, Y., Shen, C.: PGL: prior-guided local self-supervised learning for 3D medical image segmentation. arXiv preprint arXiv:2011.12640 (2020)
35. Yan, K., et al.: Self-supervised learning of pixel-wise anatomical embeddings in radiological images. arXiv preprint arXiv:2012.02383 (2020)
36. Zhang, Y., Jiang, H., Miura, Y., Manning, C.D., Langlotz, C.P.: Contrastive learning of medical visual representations from paired images and text. arXiv preprint arXiv:2010.00747 (2020)