CellStyle: Improved Zero-Shot Cell Segmentation via Style Transfer

Rüveyda Yilmaz (✉)[0009−0007−7351−698X], Zhu Chen[0009−0009−9847−7686], Yuli Wu[0000−0002−6216−4911], and Johannes Stegmaier[0000−0003−4072−3759]

Institute of Imaging and Computer Vision, RWTH Aachen University, Germany
Email: rueveyda.yilmaz@lfb.rwth-aachen.de

Abstract. Cell microscopy data are abundant; however, corresponding segmentation annotations remain scarce. Moreover, variations in cell types, imaging devices, and staining techniques introduce significant domain gaps between datasets. As a result, even large, pretrained segmentation models trained on diverse datasets (source datasets) struggle to generalize to unseen datasets (target datasets). To overcome this generalization problem, we propose CellStyle, which improves the segmentation quality of such models without requiring labels for the target dataset, thereby enabling zero-shot adaptation. CellStyle transfers the attributes of an unannotated target dataset, such as texture, color, and noise, to the annotated source dataset. This transfer is performed while preserving the cell shapes of the source images, ensuring that the existing source annotations can still be used while maintaining the visual characteristics of the target dataset. The styled synthetic images with the existing annotations enable the finetuning of a generalist segmentation model for application to the unannotated target data. We demonstrate that CellStyle significantly improves the cell segmentation performance across diverse datasets by finetuning multiple segmentation models on the style-transferred data. The source code for CellStyle is publicly available at https://github.com/ruveydayilmaz0/cellStyle.

Keywords: Cell Segmentation · Diffusion Models · Domain Adaptation

1 Introduction

Automated cell segmentation is essential for biomedical research, facilitating the extraction and analysis of cellular morphology and spatial organization [6]. However, achieving accurate instance segmentation remains challenging due to the substantial variability in imaging modalities, cell types, and staining protocols [10,23]. This is caused by domain shifts, leading to failed generalization when applied to unseen microscopy datasets with different characteristics [15]. To address this, previous studies [11,16,24] have proposed generalist models trained on diverse datasets to enhance applicability across different datasets. Often, these generalist models are finetuned on the target dataset by a human-in-the-loop approach [24] or through pre-generated labels [16, 21, 24]. Alternatively, other

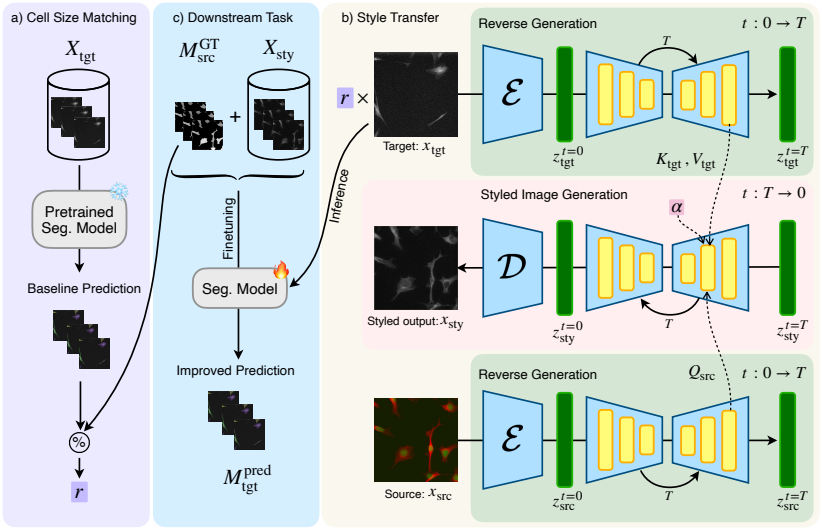

Fig. 1: Overview of the CellStyle pipeline: (a) Cell Size Matching: the average cell length in X_{tgt} is estimated using a pretrained segmentation model and compared to X_{src}, to compute the cell size ratio r which is used to scale x_{tgt}; (b) Style Transfer: the pretrained diffusion model generates x_{sty} based on x_{tgt} and x_{src}; (c) Downstream task: X_{sty} and ground truth labels $M_{\text{src}}^{\text{GT}}$ from X_{tgt} are used to finetune segmentation models.

studies [1,8,25,26,28,29] have addressed the challenge of cell instance segmentation using classical methods, diffusion models or generative adversarial networks (GANs) by generating annotated synthetic images derived from a limited set of labeled real data or by domain transfer, which adapts the visual features from one dataset to another one [2, 12, 15, 17, 27]. The generated images are then utilized as training datasets for instance segmentation models. Although there are improvements in the segmentation quality, these models require supervised [1,2,8,15,25,28,29] or unsupervised [12,17,27] training of a generative model or manual adjustments to the simulated image models [26].

In this work, we introduce CellStyle, a method designed to enhance the instance segmentation performance for cell microscopy images without requiring annotation labels for the target dataset. CellStyle leverages existing annotated datasets to address the limitations of generalist zero-shot segmentation methods. It achieves this by adapting key visual attributes of microscopy images—such as texture, color, and noise—from an unannotated target dataset to an annotated source dataset using a pretrained diffusion backbone. This transformation preserves the original cell shapes, allowing the generated images to retain the annotations of the source dataset while ensuring they reflect the visual characteristics of the target dataset. By finetuning pretrained segmentation models on style-transferred images paired with source annotations—without using any

labels from the target dataset—we enable a form of zero-shot adaptation, where the model can be applied directly to the unannotated target data. Our main contributions are threefold: (1) To the best of our knowledge, CellStyle is the first work proposing zero-shot domain transfer for cell microscopy imaging without requiring the training of a generative model. (2) We generate a diverse set of synthetic datasets that can be used to improve the downstream task performance on target data. (3) Experimental results demonstrate that CellStyle significantly enhances the zero-shot performance of cell segmentation models across various datasets.

2 Method

Diffusion models are generative models designed to synthesize data from pure noise by learning a data distribution [14]. Training involves progressively adding noise to a clean image sample and learning to predict this noise using the following loss function:

$$L(\theta) = \mathbb{E}_{x,\epsilon,t}\left[\|\epsilon - \epsilon_\theta\left(x_t, t\right)\|^2\right], \tag{1}$$

where ϵ is the noise added to a clean image x_0 and $\epsilon_\theta(x_t, t)$ is the prediction for this noise based on x_t, t and the learned parameters θ.
The inference begins with a pure noise sample x_T and a clean image is iteratively generated over T steps. Denoising Diffusion Implicit Models (DDIMs) follow the same training procedure but achieve significantly faster inference [22]. This is accomplished by formulating the diffusion process as a non-Markovian model, in contrast to the Markovian process used in [14]. As a result, during inference, DDIMs can generate images of comparable quality in approximately 50 iterations, rather than requiring $T \sim 10^3$ iterations. Building on DDIMs, Latent Diffusion Models further improve inference speed by performing diffusion in a spatially compact latent space [20]. An autoencoder [9] is used to encode the images into latent representations before the diffusion process and reconstruct them back into the image space after the diffusion process. To enable style transfer on cell microscopy images without requiring additional training of a generative model on labeled data, we adopt Stable Diffusion (SD) as our diffusion backbone, which leverages this latent diffusion architecture.

To adapt an annotated source image x_{src} to a target image x_{tgt} with no annotations, we incorporate the finding outlined in [4, 13]. The finding shows that queries (Q) in the SD UNet attention blocks govern the shapes and spatial layouts of the generated objects, while keys (K) and values (V) control the other visual attributes such as texture, color, brightness, etc. Specifically, after encoding the images x_{src} and x_{tgt} into the lower-dimensional representations $z_{\text{src}}^{t=0}$ and $z_{\text{tgt}}^{t=0}$ using the autoencoder from SD, we predict their corresponding noisy latent representations $z_{\text{src}}^{t=T}$ and $z_{\text{tgt}}^{t=T}$ by simulating a reverse generation process from $t = 0$ to $t = T$ [22]. During this reverse process, we cache the queries (Q_{src}) corresponding to z_{src}^{t} and keys and values corresponding to z_{tgt}^{t} (K_{tgt} and V_{tgt})

Table 1: The selected pairs (X_{src}, X_{tgt}) for the experiments, the calculated average cell size ratios r, and the computed adaptive score scaling ratios α.

Pair	X_{src}	X_{tgt}	r	α	Pair	X_{src}	X_{tgt}	r	α
1	MP6843	Fluo-MSC	1.0	1.5	4	Fluo-HeLa	Fluo-GOWT1	2.5	1.1
2	Huh7	DIC-HeLa	3.0	1.5	5	SHY5Y	MP6843	3.5	1.2
3	BV-2	Fluo-HeLa	2.1	1.2	6	NuI Kidney	NuI Cardia	1.0	1.0

at multiple levels of the self-attention blocks in the decoder of the SD UNet. Next, starting with $z_{\text{sty}}^{t=T} = z_{\text{src}}^{t=T}$, we generate the styled latent $z_{\text{sty}}^{t=0}$ corresponding to x_{sty} using Q_{src}, K_{tgt} and V_{tgt} (see Fig. 1). However, directly applying this approach to pairs of cell microscopy images can result in weak style transfer, depending on the disparity in average cell sizes between the datasets X_{src} and X_{tgt}. This issue arises because when computing *Attention(Q_{src}, K_{tgt}, V_{tgt})*, image patches from different images may lack strong feature correspondences when object sizes differ significantly. To address this, we first compute the average cell lengths for both X_{src} and X_{tgt} and use the resulting cell size ratio r to resize the images in X_{tgt} before performing style transfer. The cell lengths for X_{src} are derived from the ground truth (GT) annotations $M_{\text{src}}^{\text{GT}}$, while a pretrained segmentation model [16] is used to predict the annotations $M_{\text{tgt}}^{\text{pred}}$ for approximating the average cell lengths in X_{tgt} (see Fig. 1a).

Another important consideration is the decrease in the attention map values obtained with (Q_{src}, K_{tgt}) compared to the original self-attention maps. This is because the correspondence between Q and K is higher when they are derived from the same image. To account for this, [4] computes the average standard deviation ratio between the original self-attention scores and those generated using Q and K from different images. The attention scores are then scaled based on this ratio to maintain consistency. However, when generating data across different dataset pairs with varying degrees of similarity, using a fixed scaling ratio for attention scores is not optimal (the corresponding experimental results are given in Section 3). Instead, we propose computing this ratio separately for each dataset pair, which we term as *adaptive score scaling ratio* α. Prior to performing style transfer, the standard deviations of the attention scores are computed across a small set of randomly selected samples from X_{src} and X_{tgt}. These values are then averaged over all diffusion timesteps T to calculate α. Subsequently, during style transfer, the attention maps computed between Q_{src} and K_{tgt} within the self-attention blocks of SD are scaled accordingly.

3 Experiments

Datasets: We conduct experiments using publicly available datasets, including MP6843 from the Cell Image Library [30]; BV-2, Huh7, and SHSY5Y from LiveCell [7]; DIC-C2DH-HeLa (DIC-HeLa), Fluo-N2DL-HeLa (Fluo-HeLa), Fluo-

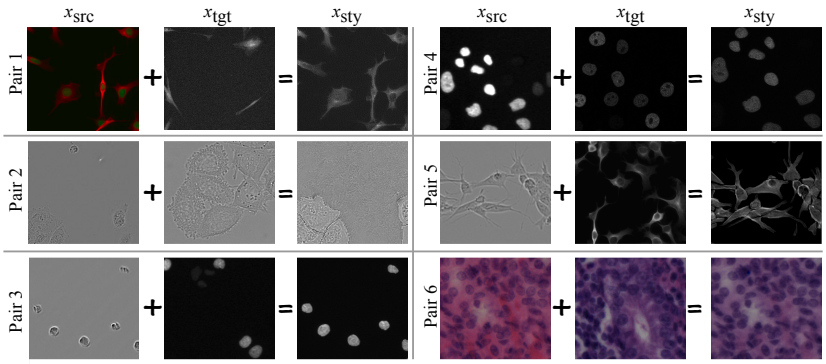

Fig. 2: Sample qualitative results (x_{sty}) for each pair along with the corresponding source (x_{src}) and the target (x_{tgt}) images.

C2DL-MSC (Fluo-MSC), and Fluo-N2DH-GOWT1 (Fluo-GOWT1) from the Cell Tracking Challenge (CTC) [19]; and human kidney and cardia datasets from NuInsSeg [18].

Experimental Setup: We conduct our experiments using the pretrained SD v1.5 model with 50 diffusion timesteps. The experiments also include the use of different numbers of timesteps; however, due to space limitations, we cannot include those results in this paper. For each dataset pair, we pick 4,000 combinations $\{(x^i_{\text{src}}, x^i_{\text{tgt}})\}_{i=1}^{4000}$ from (X_{src}, X_{tgt}) and generate the corresponding styled images $\{x^i_{\text{sty}}\}_{i=1}^{4000}$ using an Nvidia L40S GPU. During the generation process, we extract Q_{src}, K_{tgt}, and V_{tgt} from the last six attention layers of the model to be used when generating X_{sty}. To demonstrate the capabilities of CellStyle, we present experimental results on six dataset pairs (see Table 1). The pairings are made based on the morphological characteristics of the cells in the images, with nuclei images paired with other nuclei images and cytoplasm images paired with other cytoplasm images. Additionally, Table 1 provides the predicted average cell size ratios (r) and the adaptive score scaling ratios (α) computed for each dataset pair.

Evaluation Methods: We evaluate CellStyle on the downstream task of instance segmentation using Cellpose [24], Stardist [21], and Mediar [16]. First, we assess the performance of the pretrained models on X_{tgt} and then finetune them separately on X_{src} and X_{sty} using the models' default parameter configurations. The training data for the base Cellpose model originally contains images from the MP6843 dataset. For proper experimentation on pairs 1 and 5, we manually remove those images from the training dataset and train Cellpose from scratch. This modified version of the base Cellpose model is used exclusively for experiments on these pairs, while the original base model is applied to all other pairs. Stardist was originally trained on images featuring star-convex shapes, limiting its applicability to other cell shapes. Therefore, we train it from scratch on the Cellpose training data for better generalization. The training data for

Table 2: The quantitative results for Cellpose [24], Stardist [21], and Mediar [16] for each pair. For each segmentation model, the performance of the base model, the finetuned model on X_{src}, and X_{sty} are given.

Method	Pair 1		Pair 2		Pair 3		Pair 4		Pair 5		Pair 6	
	SEG	DET	SEG	DET	SEG	DET	SEG	DET	SEG	DET	SEG	DET
Cellpose	0.20	0.30	0.79	0.55	0.76	0.92	**0.90**	0.0	0.62	0.66	0.31	0.54
Cellpose+X_{src}	0.48	0.72	0.69	0.55	0.76	0.92	0.45	0.22	0.0	0.0	0.52	0.82
Cellpose+X_{sty}	**0.62**	**0.83**	**0.81**	**0.85**	**0.81**	**0.95**	0.79	**0.93**	**0.76**	**0.89**	**0.55**	**0.83**
Stardist	0.0	0.08	0.26	0.21	0.75	**0.95**	0.69	0.94	**0.36**	0.46	0.10	0.18
Stardist+X_{src}	0.02	0.10	0.28	0.22	0.64	0.84	0.41	0.82	0.01	0.0	0.52	0.80
Stardist+X_{sty}	**0.28**	**0.56**	**0.60**	**0.77**	**0.77**	0.94	**0.87**	**0.98**	0.32	**0.47**	**0.56**	**0.81**
Mediar	0.26	0.58	0.83	0.96	0.66	0.95	0.85	0.94	0.67	0.85	0.48	0.71
Mediar+X_{src}	0.24	**0.59**	0.83	0.96	0.68	0.96	0.77	0.95	0.66	0.84	0.57	0.87
Mediar+X_{sty}	**0.31**	0.54	**0.85**	**0.97**	**0.77**	**0.97**	**0.87**	**0.98**	**0.69**	**0.90**	**0.58**	**0.89**

the pretrained Mediar contains a collection of datasets, including Cellpose [24], DataScienceBowl 2018 [3], LiveCell [7] and Omnipose [5]. Again, to ensure proper experimentation, we exclude the images used in our experiments from the Mediar training set and retrain the model from scratch. This process is carried out separately for each pair, letting images from one pair be included in the training process for another pair. We further compare our results to [1, 8, 26, 28, 29] that also generate annotated synthetic cell microscopy data. However, they rely on labeled target data during generative model training, which presents a disadvantage for CellStyle. To compensate for this in the comparisons to those models, we use a mixture of our synthetic data and a set of held-out target data with labels when finetuning the segmentation models.

For the final quantitative evaluations, we use the segmentation accuracy measure (SEG) and the detection accuracy measure (DET) [19]. SEG measures how accurately the cell structures are spatially segmented while DET assesses whether the cells are correctly detected. For false detections, DET introduces a penalty term that reduces the final score.

Experimental Results: We present our experimental results in a zero-shot setting for the selected segmentation models in Table 2. The first row for each model reports the SEG and DET scores for the generalizable pretrained base models. Next, we provide the scores obtained when the base model is finetuned on X_{src} and evaluated on X_{tgt}. Finally, the results when the base model is finetuned using $\{X_{\text{sty}}, M_{\text{src}}^{\text{GT}}\}$ are presented. In general, the base Stardist model tends to underperform compared to other methods. This limitation is inherent to its design, as it was specifically developed for segmenting star-convex shapes, such as cell nuclei. As a result, its performance declines on non-convex cytoplasm images, particularly on pairs 1 and 5. As shown in Table 2, CellStyle

Table 3: Segmentation results for (a) Fluo-GOWT1 and (b) Fluo-HeLa on Cellpose [24], Stardist [21], and Mediar [16]. For fairness, the CellStyle outputs are combined with real target images to finetune the segmentation models, as other methods use labeled target data when training the generative models. The best-performing results are highlighted in bold, while the second-best are underlined.

(a) Comparative segmentation results for Fluo-GOWT1

Method	[8]		[29]		Ours+Real	
	SEG	DET	SEG	DET	SEG	DET
Cellpose	<u>0.88</u>	0.91	0.87	<u>0.96</u>	**0.91**	**0.98**
Stardist	0.44	0.79	<u>0.84</u>	<u>0.86</u>	**0.87**	**0.98**
Mediar	0.79	<u>0.78</u>	<u>0.91</u>	**0.97**	**0.92**	**0.97**

(b) Comparative segmentation results for Fluo-HeLa

Method	[26]		[8]		[29]		[28]		[1]		Ours+Real	
	SEG	DET	SEG	DET	SEG	DET	SEG	DET	SEG	DET	SEG	DET
Cellpose	0.70	0.88	0.75	<u>0.94</u>	<u>0.80</u>	0.93	0.71	0.83	0.76	0.81	**0.83**	**0.96**
Stardist	0.70	**0.98**	0.62	0.87	0.72	0.91	<u>0.75</u>	0.85	<u>0.75</u>	<u>0.95</u>	**0.77**	<u>0.95</u>
Mediar	0.83	<u>0.97</u>	0.68	0.88	0.84	<u>0.97</u>	0.81	**0.98**	<u>0.85</u>	<u>0.97</u>	**0.88**	**0.98**

significantly improves the zero-shot segmentation quality compared to the base models in terms of both SEG and DET. It is important to note that the overall segmentation performance should be assessed by considering the average of SEG and DET scores [19]. This is mainly because the GT segmentation labels are not present for all the cells in the CTC dataset, while cell centers are fully annotated. Since SEG is computed based on these GT masks, only the regions containing annotations are evaluated, meaning that false positives in the predicted masks are not penalized. However, this limitation can be overcome by considering the arithmetic mean of SEG and DET, referred to as the overall performance measure (OP_{CSB}) [19]. For example, in Table 2, the base Cellpose model achieves a higher SEG score for Pair 4, while upon inspecting the predicted masks, we observed numerous background regions that were falsely segmented as cells. This cannot be captured directly by SEG, hindering proper comparisons of the predictions. Alternatively, when OP_{CSB} is considered, the performance of our approach (with $0.5\times(\text{SEG} + \text{DET}) = 0.86$) significantly surpasses the base model (with $0.5\times(\text{SEG} + \text{DET}) = 0.45$) even for this particular pair. Additionally, in Table 3, we compare our method to other generative models on the downstream segmentation task in a non-zero-shot setting since they use labeled data during the training of generative models. The comparisons are conducted using the datasets that were also generated by these methods, namely Fluo-HeLa and Fluo-GOWT1. The compared methods were specifically designed for these

Table 4: The results for the ablation experiments using Cellpose [24], Stardist [21], and Mediar [16]. For the whole CellStyle pipeline denoted as $+X_{\text{sty}}$, we calculated $\alpha = 1.5$ or $r = 1.0$ for some pairs (see Table 1). This causes the results for the configurations $+X_{\text{sty}} - \alpha$ or $+X_{\text{sty}} - r$ to coincide with the setting from $+X_{\text{sty}}$ for those pairs. These results are replaced by '-' to avoid redundancy.

Method	Pair 1		Pair 2		Pair 3		Pair 4		Pair 5		Pair 6	
	SEG	DET	SEG	DET	SEG	DET	SEG	DET	SEG	DET	SEG	DET
Cellp+r	0.48	0.72	0.61	0.80	0.65	0.84	0.78	0.0	0.60	0.73	0.52	0.82
Cellp+$X_{\text{sty}} - \alpha$	-	-	-	-	0.80	0.93	**0.79**	0.90	0.72	0.81	0.54	0.81
Cellp+$X_{\text{sty}} - r$	-	-	0.43	0.0	0.59	0.82	0.39	0.63	0.04	0.0	-	-
Cellp+X_{sty}	**0.62**	**0.83**	**0.81**	**0.85**	**0.81**	**0.95**	**0.79**	**0.93**	**0.76**	**0.89**	**0.55**	**0.83**
Strd+r	0.02	0.10	0.03	0.05	0.72	0.90	0.85	0.96	0.03	0.0	0.52	0.80
Strd+$X_{\text{sty}} - \alpha$	-	-	-	-	0.76	**0.94**	0.85	**0.98**	0.31	0.46	0.54	0.80
Strd+$X_{\text{sty}} - r$	-	-	0.52	0.73	0.18	0.30	0.39	0.67	0.03	0.0	-	-
Strd+X_{sty}	**0.28**	**0.56**	**0.60**	**0.77**	**0.77**	**0.94**	**0.87**	**0.98**	**0.32**	**0.47**	**0.56**	**0.81**
Mdr+r	0.24	0.50	0.79	0.96	0.75	0.95	0.85	0.95	0.63	0.84	0.57	0.87
Mdr+$X_{\text{sty}} - \alpha$	-	-	-	-	0.68	0.96	**0.87**	**0.98**	0.68	0.89	0.56	0.88
Mdr+$X_{\text{sty}} - r$	-	-	0.77	0.86	0.65	0.92	0.80	0.96	0.60	0.77	-	-
Mdr+X_{sty}	**0.31**	**0.54**	**0.85**	**0.97**	**0.77**	**0.97**	**0.87**	**0.98**	**0.69**	**0.90**	**0.58**	**0.89**

datasets, and adapting them to all the datasets used in this work was not feasible without major modifications. While our approach demonstrates competitive performance to those models even for the zero-shot scenario (see Table 2), when combined with real data, it surpasses them at least by 1% in terms of OP_{CSB}, the average of SEG and DET (see Table 3).

Ablation Experiments: To assess the significance of individual components in CellStyle, we conduct ablation experiments. Specifically, we test the effects of (i) performing only cell size matching and finetuning the segmentation models on X_{src}, (ii) using a constant attention score scaling ratio of $\alpha = 1.5$ during style transfer, (iii) performing style transfer without cell size matching, i.e., with $r = 1.0$, and (iv) using the whole CellStyle pipeline (see Table 4). When only cell size matching is used ($+r$), the performance tends to drop, which is more pronounced for the pairs that are significantly different in terms of structure or color. Similarly, when a fixed attention score scaling ratio $\alpha = 1.5$ is used for all the pairs ($+X_{\text{sty}} - \alpha$), the overall segmentation performance decreases compared to the adaptive approach, where α is specifically calculated for each pair of datasets. Intuitively, the value of this parameter varies based on the similarity of cell characteristics between the paired datasets, with higher values of α for lower similarity, and vice versa. Additionally, performing style transfer without cell size matching ($+X_{\text{sty}} - r$) results in a reduction in segmentation quality primarily due to structural size differences between X_{src} and X_{tgt} causing blurry or indistinct representations in X_{sty}.

4 Conclusion

We introduced CellStyle to improve the zero-shot performance of pretrained cell instance segmentation models. By leveraging style transfer, CellStyle transforms a labeled source dataset to match the visual characteristics of an unlabeled target dataset while preserving cell morphology. The generated styled images, combined with the source annotations, enable the finetuning of segmentation models without requiring additional target dataset labels. Our experimental results demonstrate that CellStyle significantly enhances segmentation performance compared to baseline methods and alternative generative models.
Acknowledgments: This work was partially funded by the German Research Foundation DFG (STE2802/5-1).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bähr, D., Eschweiler, D., Bhattacharyya, A., Moreno-Andrés, D., Antonin, W., Stegmaier, J.: Cellcyclegan: Spatiotemporal microscopy image synthesis of cell populations using statistical shape models and conditional gans. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 15–19 (2021)
2. Bouteldja, N., Hölscher, D.L., Bülow, R.D., Roberts, I.S., Coppo, R., Boor, P.: Tackling stain variability using cyclegan-based stain augmentation. Journal of Pathology Informatics **13**, 100140 (2022)
3. Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al.: Nucleus segmentation across imaging experiments: the 2018 data science bowl. Nature Methods **16**(12), 1247–1253 (2019)
4. Chung, J., Hyun, S., Heo, J.P.: Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8795–8805 (2024)
5. Cutler, K.J., Stringer, C., Lo, T.W., Rappez, L., Stroustrup, N., Brook Peterson, S., Wiggins, P.A., Mougous, J.D.: Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. Nature Methods **19**(11), 1438–1448 (2022)
6. Durkee, M.S., Abraham, R., Clark, M.R., Giger, M.L.: Artificial intelligence and cellular segmentation in tissue microscopy images. The American Journal of Pathology **191**(10), 1693–1701 (2021)
7. Edlund, C., Jackson, T.R., Khalid, N., Bevan, N., Dale, T., Dengel, A., Ahmed, S., Trygg, J., Sjögren, R.: Livecell—a large-scale dataset for label-free live cell segmentation. Nature Methods **18**(9), 1038–1045 (2021)
8. Eschweiler, D., Yilmaz, R., Baumann, M., Laube, I., Roy, R., Jose, A., Brückner, D., Stegmaier, J.: Denoising diffusion probabilistic models for generation of realistic fully-annotated microscopy image datasets. PLOS Computational Biology **20**(2), e1011890 (2024)

9. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12873–12883 (2021)
10. Gogoberidze, N., Cimini, B.A.: Defining the boundaries: challenges and advances in identifying cells in microscopy images. Current Opinion in Biotechnology **85**, 103055 (2024)
11. Greenwald, N.F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C.C., McIntosh, B.J., Leow, K.X., Schwartz, M.S., et al.: Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nature Biotechnology **40**(4), 555–565 (2022)
12. Haq, M.M., Huang, J.: Adversarial domain adaptation for cell segmentation. In: Medical Imaging with Deep Learning. pp. 277–287 (2020)
13. Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., Cohen-or, D.: Prompt-to-prompt image editing with cross-attention control. In: International Conference on Learning Representations (2023)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
15. Keaton, M.R., Zaveri, R.J., Doretto, G.: Celltranspose: Few-shot domain adaptation for cellular instance segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 455–466 (2023)
16. Lee, G., Kim, S., Kim, J., Yun, S.Y.: Mediar: Harmony of data-centric and model-centric for multi-modality microscopy. In: Proceedings of The Cell Segmentation Challenge in Multi-modality High-Resolution Microscopy Images. Proceedings of Machine Learning Research, vol. 212, pp. 1–16 (2023)
17. Liu, D., Zhang, D., Song, Y., Zhang, F., O'Donnell, L., Huang, H., Chen, M., Cai, W.: Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4243–4252 (2020)
18. Mahbod, A., Polak, C., Feldmann, K., Khan, R., Gelles, K., Dorffner, G., Woitek, R., Hatamikia, S., Ellinger, I.: Nuinsseg: A fully annotated dataset for nuclei instance segmentation in h&e-stained histological images. Scientific Data **11**(1), 295 (2024)
19. Maška, M., Ulman, V., Delgado-Rodriguez, P., Gómez-de Mariscal, E., Nečasová, T., Guerrero Peña, F.A., Ren, T.I., Meyerowitz, E.M., Scherr, T., Löffler, K., et al.: The cell tracking challenge: 10 years of objective benchmarking. Nature Methods **20**(7), 1010–1020 (2023)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
21. Schmidt, U., Weigert, M., Broaddus, C., Myers, G.: Cell detection with star-convex polygons. In: Medical Image Computing and Computer Assisted Intervention. pp. 265–273. Springer (2018)
22. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021)
23. Stacke, K., Eilertsen, G., Unger, J., Lundström, C.: Measuring domain shift for deep learning in histopathology. IEEE Journal of Biomedical and Health Informatics **25**(2), 325–336 (2020)
24. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature Methods **18**(1), 100–106 (2021)

25. Sturm, M., Cerrone, L., Hamprecht, F.A.: Syncellfactory: Generative data augmentation for cell tracking. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 304–313. Springer (2024)
26. Svoboda, D., Ulman, V.: Mitogen: a framework for generating 3d synthetic time-lapse sequences of cell populations in fluorescence microscopy. IEEE Transactions on Medical Imaging **36**(1), 310–321 (2016)
27. Yang, S., Zhang, J., Huang, J., Lovell, B.C., Han, X.: Minimizing labeling cost for nuclei instance segmentation and classification with cross-domain images and weak labels. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 697–705 (2021)
28. Yilmaz, R., Eschweiler, D., Stegmaier, J.: Annotated biomedical video generation using denoising diffusion probabilistic models and flow fields. In: International Workshop on Simulation and Synthesis in Medical Imaging. pp. 197–207. Springer (2024)
29. Yilmaz, R., Keven, K., Wu, Y., Stegmaier, J.: Cascaded diffusion models for 2D and 3D microscopy image synthesis to enhance cell segmentation. In: 2025 IEEE International Symposium on Biomedical Imaging (2025)
30. Yu, W., Lee, H.K., Hariharan, S., Bu, W.Y., Ahmed, S.: Ccdb:6843, mus musculus, neuroblastoma. Dataset. https://doi.org/10.7295/W9CCDB6843