

3. Cao, T., Huang, C., Hui, D.Y.T., Cohen, J.P.: A benchmark of medical out of distribution detection. In: Uncertainty and Robustness in Deep Learning Workshop at ICML (2020), <http://arxiv.org/abs/2007.04250>
4. Dikici, E., Bigelow, M., Prevedello, L.M., White, R.D., Erdal, B.S.: Integrating AI into radiology workflow: levels of research, production, and feedback maturity. *J. Med. Imag.* **7**(1), 016502 (2020)
5. Dodge, Y.: Kolmogorov-Smirnov Test, pp. 283–287. Springer, New York (2008). [https://doi.org/10.1007/978-0-387-32833-1\\_214](https://doi.org/10.1007/978-0-387-32833-1_214)
6. Eche, T., Schwartz, L.H., Mokrane, F.Z., Dercle, L.: Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome underspecification. *Radiol. Artif. Intell.* **3**(6), e210097 (2021)
7. Finlayson, S.G., et al.: The clinician and dataset shift in artificial intelligence. *N. Engl. J. Med.* **385**(3), 283–286 (2021)
8. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Computer Vision and Pattern Recognition (2017), <https://arxiv.org/abs/1608.06993>
9. Irvin, J., et al.: CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. CoRR abs/1901.07031 (2019), <http://arxiv.org/abs/1901.07031>
10. Kermany, D., Zhang, K., Goldbaum, M., et al.: Labeled optical coherence tomography (OCT) and chest x-ray images for classification. *Mendeley data* **2**(2) (2018)
11. Klaise, J., Van Looveren, A., Cox, C., Vacanti, G., Coca, A.: Monitoring and explainability of models in production. arXiv preprint [arXiv:2007.06299](https://arxiv.org/abs/2007.06299) (2020)
12. van Leeuwen, K.G., Schalekamp, S., Rutten, M.J.C.M., van Ginneken, B., de Rooij, M.: Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur. Radiol.* **31**(6), 3797–3804 (2021). <https://doi.org/10.1007/s00330-021-07892-z>
13. Mahajan, V., Venugopal, V.K., Murugavel, M., Mahajan, H.: The algorithmic audit: working with vendors to validate radiology-AI algorithms—how we do it. *Acad. Radiol.* **27**(1), 132–135 (2020)
14. Mehri, M.H.R., van Ooijen, P., Homan, M.: Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. *Eur. Radiol.* **31**(4), 1805–1811 (2021)
15. Mildener, P., Eichelberg, M., Martin, E.: Introduction to the DICOM standard. *Eur. Radiol.* **12**(4), 920–927 (2002)
16. Pearson, K.: X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **50**(302), 157–175 (1900). <https://doi.org/10.1080/14786440009463897>
17. Sculley, D., et al.: Machine learning: the high interest credit card of technical debt. In: NIPS Workshop 2014 (2014)
18. Shafaei, A., Schmidt, M., Little, J.J.: Does your model know the digit 6 is not a cat? a less biased evaluation of “outlier” detectors. CoRR abs/1809.04729 (2018), <http://arxiv.org/abs/1809.04729>
19. Tadavarthi, Y., et al.: The state of radiology AI: considerations for purchase decisions and current market offerings. *Radiol. Artif. Intell.* **2**(6), e200004 (2020)
20. Tariq, A., et al.: Current clinical applications of artificial intelligence in radiology and their best supporting evidence. *J. Am. Coll. Radiol.* **17**(11), 1371–1381 (2020)
21. West, E., Mutasa, S., Zhu, Z., Ha, R.: Global trend in artificial intelligence-based publications in radiology from 2000 to 2018. *Am. J. Roentgenol.* **213**(6), 1204–1206 (2019)

22. Wiggins, W.F., et al.: Imaging AI in practice: a demonstration of future workflow using integration standards. *Radiol. Artif. Intell.* **3**(6), e210152 (2021)
23. Zenati, H., Foo, C.S., Lecouat, B., Manek, G., Chandrasekhar, V.R.: Efficient GAN-based anomaly detection. arXiv preprint [arXiv:1802.06222](https://arxiv.org/abs/1802.06222) (2018)



# DiffMix: Diffusion Model-Based Data Synthesis for Nuclei Segmentation and Classification in Imbalanced Pathology Image Datasets

Hyun-Jic Oh and Won-Ki Jeong<sup>(✉)</sup>

College of Informatics, Department of Computer Science and Engineering, Korea University, Seoul, South Korea

wkjeong@korea.ac.kr

**Abstract.** Nuclei segmentation and classification is an important process in pathological image analysis. Deep learning-based approaches contribute significantly to the enhanced accuracy of this task. However, these approaches suffer from an imbalanced nuclei data composition, which results in lower classification performance for rare nuclei class. In this study, we proposed a realistic data synthesis method using a diffusion model. We generated two types of virtual patches to enlarge the training data distribution, which balanced the nuclei class variance and increased the chance of investigating various nuclei. Subsequently, we used a semantic label-conditioned diffusion model to generate realistic and high-quality image samples. We demonstrated the efficacy of our method based on experimental results on two imbalanced nuclei datasets, improving the state-of-the-art networks. The experimental results suggest that the proposed method improves the classification performance of the rare type nuclei classification, while showing superior segmentation and classification performance in imbalanced pathology nuclei datasets.

**Keywords:** Diffusion models · Data augmentation · Nuclei segmentation and classification

## 1 Introduction

In digital pathology, nuclear segmentation and classification are crucial tasks in disease diagnosis. Because of the diverse nature (e.g., shape, size, and color) and large numbers of nuclei, nuclei analysis in whole-slide images (WSIs) is a challenging task for which computerized processing has become a de facto standard [11]. With the advent of deep learning, challenges in nuclei analysis, such as color inconsistency, overlapping nuclei, and clustered nuclei, have been effectively addressed through data-driven approaches [6, 8, 15, 23]. Recent studies have addressed nuclear segmentation and classification simultaneously.

---

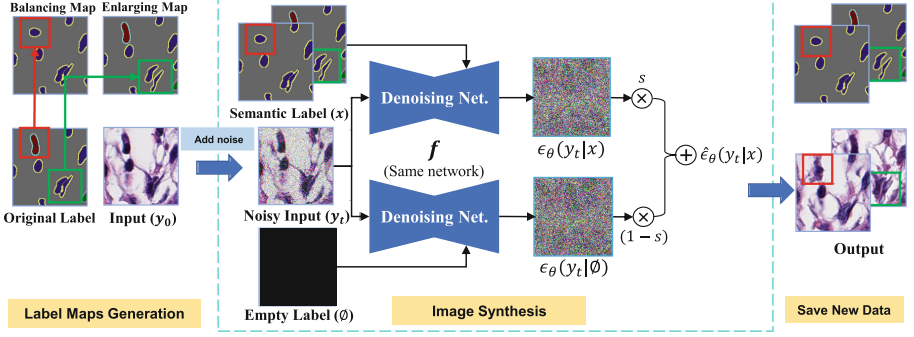
**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-43898-1\\_33](https://doi.org/10.1007/978-3-031-43898-1_33).

For example, HoVer-Net [8] and SONNET [6] performed nuclei segmentation using a distance map to identify nucleus instances and then assigned an appropriate class to each of them. Although such deep-learning-based algorithms have shown promising performance and have overcome various challenges in nuclei analysis, data imbalance among nuclei types in training data has become a major performance bottleneck [5].

Data augmentation [2, 16] can be an effective solution for compensating data imbalance and generalizing DNN by expanding the learnable training distribution using virtual training data. Several studies have been conducted on image classification tasks. Mixup [22] interpolates pairs of images and labels to generate virtual training data. CutOut [3] randomly masks square regions of the input during training. CutMix [21] cuts patches from the original images and pastes them onto other training images. Recently, a generative adversarial network(GAN) [7, 10, 12, 18] has been actively studied for pathology data augmentation. However, training a GAN is challenging owing to its instability and the need for hyper parameter tuning [4]. Moreover, most previous studies mainly focused on nuclei segmentation without considering nuclei classification. Recently, Doan *et al.* [5] proposed a data regularization scheme that addresses the data imbalance problem in pathological images. The main concept is to cut the nuclei from a scarce class image and paste them onto the nuclei from an abundant class image. Because the source and target nuclei differ in size and shape, a distance-based blending scheme is proposed. This method slightly reduces the data-imbalance problem; however, it only considers pixel values for blending and some unrealistic blending, artifacts can be observed, which is the main limitation of this method.

The main motivation for this study stems from the recent advances in generative models. Recently, the denoising diffusion probabilistic model(DDPM) [9] has gained considerable attention owing to its superior performance, which surpasses that of conventional GANs, and it has been successfully adopted in conditional environments [4, 13, 20]. Among these, we were inspired by Wang *et al.* [19], semantic diffusion model(SDM) which can synthesize a semantic image conditioned on the semantic label map. Because data augmentation for nuclei segmentation and classification requires accurate semantic images and label map pairs, we believe that SDM effectively fits the data augmentation scenario of our unbalanced nuclei data while allowing more realistic pathology image generation compared to pixel-blending or GAN-based prior work.

In this study, we proposed a novel data augmentation technique using a conditioned diffusion model, DiffMix, for imbalanced nuclear pathology datasets. DiffMix consists of the following steps. First, we trained the SDM with semantic map guidance, which consists of instance and class-type maps. Next, we built custom label maps by modifying the existing imbalanced label maps. We changed the nuclei labels and randomly shifted the locations of the nuclei mask to ensure that the number of each class label was balanced and the data distribution expanded. Finally, we synthesized more diverse, semantically realistic, and



**Fig. 1.** Framework of DiffMix. First, we generate custom semantic label maps( $x$ ) and noisy images( $y_t$ ). Second, we synthesize image samples with a pretrained semantic diffusion model conditioned on the custom masks. Semantic label  $x$  is custom mask to enlarge the data distribution. Finally, we utilize the synthesized image and label pairs in the training DNN for nuclear segmentation and classification.

well-balanced new pathology nuclei images using SDM conditioned on custom label maps. The main contributions of this study are summarized as follows:

- We introduce a data augmentation framework for imbalanced pathology image datasets that can generate realistic samples using semantic diffusion model conditioned on two custom label maps, which can enlarge the data distribution.
- We demonstrate the efficacy and generalization ability of our scheme with experimental results on two imbalanced pathology nuclei datasets, GLySAC [6] and CoNSeP [8], improving the performance of state-of-the-art networks.
- Our experiments demonstrated that the optimal approach for data augmentation depends on the level of imbalance, balancing sample numbers, and enlarging the training data distribution, which are critical factors for consideration.

## 2 Method

In this section, the proposed method is described in detail. DiffMix operates through several steps. First, we trained the SDM first on the training data. Balancing label maps comprise several rare class labels, and enlarging label maps are composed of randomly shifted nuclei. Finally, using the pre-trained SDM and custom label maps, we synthesized realistic data to train on imbalanced datasets. Before discussing DiffMix, a brief introduction of SDM is presented. An overview of the proposed method is presented in Fig. 1.

### 2.1 Preliminaries

The SDM is a conditional denoising diffusion probabilistic model (CDPM) conditioned on semantic label maps. Based on the CDPM, SDM follows two fundamental diffusion processes *i.e.*, forward and reverse process. The reverse process was

a Markov chain with Gaussian transitions. When the added noise is sufficiently large, the reverse process is approximated by a random variable  $\mathbf{y}_T \sim \mathcal{N}(0, \mathbf{I})$ , defined as follows:

$$p_\theta(\mathbf{y}_{0:T}|\mathbf{x}) = p(\mathbf{y}_T) \prod_{t=1}^T p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}) \quad (1)$$

$$p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}) = \mathcal{N}(\mathbf{y}_{t-1}; \mu_\theta(\mathbf{y}_t, \mathbf{x}, t), \Sigma_\theta(\mathbf{y}_t, \mathbf{x}, t)) \quad (2)$$

The forward process implements Gaussian noise addition for  $T$  timesteps based on variance schedule  $\{\beta_1, \dots, \beta_T\}$  as follows:

$$q(\mathbf{y}_t|\mathbf{y}_{t-1}) = \mathcal{N}(\mathbf{y}_t; \sqrt{1 - \beta_t}\mathbf{y}_{t-1}, \beta_t\mathbf{I}) \quad (3)$$

For  $\alpha_t := 1 - \beta_t$  and  $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$ , we can write the marginal distribution as follows,

$$q(\mathbf{y}_t|\mathbf{y}_0) = \mathcal{N}(\mathbf{y}_t; \sqrt{\bar{\alpha}_t}\mathbf{y}_0, (1 - \bar{\alpha}_t)\mathbf{I}) \quad (4)$$

The conditional DDPM was optimized to minimize the negative log-likelihood of the data for the particular input and condition information. If noise in the data follows Gaussian distribution with a diagonal covariance matrix  $\Sigma_\theta(y_t, x, t) = \sigma_t\mathbf{I}$ , denoising can be the optimization target by removing the noise assumed to be present in data as follows,

$$\mathcal{L}_{t-1} = \mathbb{E}_{y_0, \epsilon} [\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t}\mathbf{y}_0 + \sqrt{1 - \alpha_t}\epsilon, \mathbf{x}, t)\|_2] \quad (5)$$

## 2.2 Semantic Diffusion Model (SDM)

The SDM is a U-Net-based network that estimates noise from a noisy input image. Contrary to other conditional DDPMs, the denoising network of the SDM processes the semantic label map  $x$  and noisy input  $y_t$  independently. When  $y_t$  is fed into the encoder,  $x$  is injected into the decoder to fully leverage semantic information [19]. For training, the SDM is trained similarly to the improved DDPM [14] to ensure that it predicts the noise involved in reconstructing the input image as well as variances to enhance the log-likelihood of the generated images. To improve the sample quality, SDM utilizes classifier-free guidance for inference.

SDM replaces the semantic label map  $x$  with an empty (null) map  $\emptyset$  to separate the noise estimated under the label map guidance by  $\epsilon_\theta(y_t|x)$ , from the noise estimated in an unconditioned case  $\epsilon_\theta(y_t|\emptyset)$ . This strategy allows the inference of the gradient of log probability, expressed as follows:

$$\epsilon_\theta(y_t|x) - \epsilon_\theta(y_t|\emptyset) \propto \nabla_{y_t} \log p(y_t|x) - \nabla_{y_t} \log p(y_t) \propto \nabla_{y_t} \log p(x|y_t) \quad (6)$$

During the sampling process, the disentangled component  $s$  is increased to improve the samples from the conditional diffusion models, formulated as follows:

$$\hat{\epsilon}(y_t|x) = \epsilon_\theta(y_t|x) + s \cdot (\epsilon_\theta(y_t|x) - \epsilon_\theta(y_t|\emptyset)) \quad (7)$$

### 2.3 Custom Semantic Label Maps Generation

Figure 1 illustrates the process of creating custom label maps to condition the semantic diffusion model for synthesizing the desired data based on the original input image label  $y_0$ . We prepared custom semantic label maps to condition the SDM and used it to synthesize data for our imbalanced datasets. Therefore, we considered two types of semantic label maps to improve imbalanced datasets. First, balancing maps to balance the number of nuclei among different nuclei types. GradMix [5] increased the number of fewest type nuclei in datasets by cutting, pasting, and smoothing both images and labels. However, we used only mixed labels in our experiments. Second, enlarging maps perturbs the positions of the nuclei instances to diversify the datasets. We randomly moved nuclei positions on semantic maps to synthesize diverse image patches with SDM by conditioning them with unfamiliar semantic maps to allow the diffusion model to generate significant patches.

### 2.4 Image Synthesis

We synthesized the virtual data with a pretrained SDM conditioned on custom semantic label maps  $x$ . Figure 1 depicts the data sampling process in the SDM. Before inputting the original image  $y_0$  into diffusion net  $f$ , we added noise to  $y_0$ . The semantic label and noisy image  $y_t$  were used simultaneously. To synthesize virtual images, we built two label maps and trained a semantic diffusion model. Before data synthesis, we added noise to the input image  $y_0$ . Subsequently, we input  $y_t$  and  $x$  to the pretrained denoising network  $f$ ,  $y_t$  for the encoder, and  $x$  for the decoder. As SDM generates samples, it uses an empty label  $\emptyset$  to generate the unconditioned output. The image was sampled from an existing noisy patch, depending on the predefined time steps. Therefore, we generated patches conditioned on custom semantic label maps. In this process, we added noise to the input image  $y_0$ ; thus, we input the noisy input  $y_t$  into the pretrained denoising network  $f$ , and following [19], we input the custom semantic label maps to the decoder parts. Thereafter, the semantic label maps condition the SDM to synthesize image data that satisfy the label maps. We sampled the new data using the DDIM [17] process, which reduces the number of sampling steps, but with high-quality data.

## 3 Experiments

### 3.1 Datasets

In this study, we used two imbalanced nuclear segmentation and classification datasets. First, GLySAC [6], comprising 59 H&E images of size  $1000 \times 1000$  pixels, was split into 34 training images and 25 test images. GLySAC comprised 30875 nuclei and was grouped into three nuclei types: 12081 lymphocytes, 12287 epithelial, and 6507 miscellaneous. Second, CoNSEP [8], comprising 41 H&E images of size  $1000 \times 1000$  pixels, was divided into 27 training images and 14 test images. CoNSEP comprised 24319 nuclei, and four nuclei classes: 5537 epithelial nuclei, 3941 inflammatory, 5700 spindle, and 371 miscellaneous nuclei.

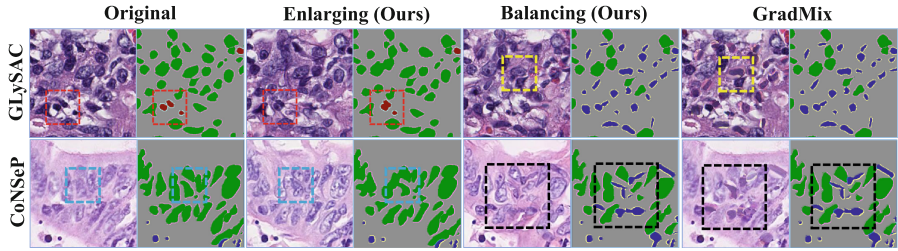
**Table 1.** Quantitative results on GLySAC and CoNSeP. We implemented our scheme on two state-of-the-art networks and compared with GradMix. From Dice to PQ metrics state segmentation performance, the metrics from Acc to  $F^S$  indicate classification performance. The highest scores for the same network and dataset are highlighted in **bold**. **Red** indicates cases in which the performance was at least 3% higher than that of the other methods.

Dataset	Method	Segmentation					Classification				
		Dice	AJI	DQ	SQ	PQ	Acc	$F^E$	$F^L$	$F^M$	$F^S$
GLySAC	HoVer-Net	0.839	0.670	0.807	0.787	0.637	0.713	0.565	0.556	0.315	—
	GradMix	0.839	0.672	0.809	0.789	0.640	0.703	0.551	0.551	0.320	—
	DiffMix-E	0.838	0.669	0.806	0.789	0.640	<b>0.719</b>	0.572	<b>0.560</b>	0.321	—
	DiffMix-B	<b>0.840</b>	<b>0.673</b>	<b>0.811</b>	<b>0.791</b>	<b>0.642</b>	0.697	0.573	0.519	0.304	—
	DiffMix	0.837	0.669	0.806	0.790	0.639	0.716	<b>0.582</b>	0.541	<b>0.324</b>	—
	SONNET	0.835	0.660	0.789	0.792	0.627	0.679	0.511	0.511	0.305	—
	GradMix	0.835	0.658	0.787	0.790	0.625	0.680	0.506	0.509	0.312	—
	DiffMix-E	0.837	0.662	<b>0.793</b>	<b>0.793</b>	<b>0.631</b>	<b>0.700</b>	0.533	<b>0.524</b>	<b>0.334</b>	—
	DiffMix-B	<b>0.839</b>	0.661	0.788	0.792	0.627	0.687	0.530	0.507	0.300	—
	DiffMix	0.837	<b>0.663</b>	<b>0.793</b>	0.791	0.630	0.694	<b>0.538</b>	0.513	0.312	—
CoNSeP	HoVer-Net	0.835	0.545	0.636	0.758	0.483	0.799	0.588	0.490	0.204	0.478
	GradMix	<b>0.836</b>	0.562	<b>0.658</b>	0.765	0.504	0.802	0.598	<b>0.519</b>	0.144	0.494
	DiffMix-E	0.832	0.550	0.645	0.760	0.492	0.804	0.602	0.486	0.223	0.493
	DiffMix-B	0.835	0.558	0.653	0.762	0.499	0.809	0.595	0.496	0.324	0.498
	DiffMix	<b>0.836</b>	<b>0.563</b>	<b>0.658</b>	<b>0.766</b>	<b>0.505</b>	<b>0.818</b>	<b>0.604</b>	0.501	<b>0.363</b>	<b>0.508</b>
	SONNET	0.841	0.564	0.646	0.766	0.496	0.863	0.610	0.618	0.367	0.560
	GradMix	0.840	0.561	0.639	0.764	0.489	0.861	0.600	<b>0.639</b>	0.348	0.555
	DiffMix-E	0.842	0.567	0.648	<b>0.767</b>	0.498	0.860	0.604	0.600	0.374	0.557
	DiffMix-B	0.842	0.562	0.636	0.765	0.488	0.857	0.600	0.606	0.335	0.557
	DiffMix	<b>0.844</b>	<b>0.570</b>	<b>0.649</b>	0.766	<b>0.499</b>	<b>0.873</b>	<b>0.622</b>	0.627	<b>0.463</b>	<b>0.575</b>

### 3.2 Implementation Details

We used an NVIDIA RTX A6000 to train the SDM for 10000epochs. For data synthesis, we implemented a DDIM-based diffusion process from 1000 to 100 and added noise to the input image to the SDM, setting  $T$  as 55. In our scheme,  $\emptyset$  is defined as the all-zero vector, as in [19] and  $s = 1.5$  when sampling both datasets. We conducted experiments on two baseline networks, SONNET [6] and HoVer-Net [8]. SONNET was implemented using Tensorflow version 1.15 [1] as the software framework with two NVIDIA GeForce 2080 Ti GPUs. HoVer-Net was trained using PyTorch 1.11.0, as the software framework, with one NVIDIA GeForce 3090 Ti GPU. We implemented 4-fold cross validation for SONNET and 5-fold cross validation for HoVer-Net. For a fair comparison, we trained process with the same number of iterations and changed only the epoch numbers depending on the size of training set. DiffMix and GradMix used all the original patches and the same numbers of synthesized patches In case of DiffMix-B, it comprised original data and balancing map-based patches. Similarly, the DiffMix-E training set comprised enlarging map-based patches with the original training set. Therefore, we trained each baseline network for 100 epochs, 75





**Fig. 2.** Qualitative comparison of synthesized patches. From left to right: Original data, enlarging map(ours), balancing map(ours), and GradMix, respectively. Top to bottom: GLySAC and CoNSEP datasets, respectively. Each image and its corresponding semantic label were paired together. In this work, we utilized the same labels for balancing maps and GradMix. Comparing the result of balancing and GradMix patches, our method generates semantically harmonized patches.

epochs for DiffMix-B and DiffMix-E, and trained 50 epochs for GradMix and DiffMix.

### 3.3 Results

Figure 2 shows a qualitative comparison of the synthesized patches. The original patch is shown on the left, followed by the enlarging, balancing, and GradMix patches. The two types of patches were well harmonized with the surrounding structure compared to GradMix. Moreover, using our scheme, many patches can be synthesized using a semantic diffusion model.

For a quantitative evaluation, we implemented two state-of-the-art networks, HoVer-Net and SONNET, on two publicly imbalanced nuclei-type datasets, GLySAC and CoNSEP. Table 1 lists the results of the five experiments per network for each dataset. Furthermore, we conducted ablation studies on balanced (DiffMix-B) and enlarged (DiffMix-E) patch datasets. Before analyzing the experimental results, we computed the proportion of the least common nuclei type in each dataset. We found that miscellaneous nuclei accounted for approximately 19% of the nuclei in GLySAC, but only 2.4% in CoNSEP. This indicates that GLySAC is more balanced in terms of nuclei type than CoNSEP. Considering this information, we analyzed the experimental results. DiffMix-E exhibited the highest classification performance on the GLySAC dataset. This result indicates that enlarging semantic map-based data synthesis, such as DiffMix-E, provides sufficient opportunities to enlarge the learning distribution for GLySAC. However, for the CoNSEP dataset, DiffMix-E performed lower than the other methods, suggesting that the dataset is slightly balanced; thus, it is important to expand the available data distribution. DiffMix showed the highest performance in most metrics, with 4% and 9% margins from the second-highest result in classifying miscellaneous data, successfully diminishing the classification performance variability among class types. Furthermore, DiffMix improved the segmentation

and classification performances of the two state-of-the-art networks, even when compared with GradMix.

## 4 Conclusion

In this study, we introduced **DiffMix**, a semantic diffusion model-based data augmentation framework for imbalanced pathology nuclei datasets. We experimentally demonstrated that our method can synthesize virtual data that can balance and enlarge imbalanced nuclear pathology datasets. Our method also outperformed the state-of-the-art GradMix in terms of qualitative and quantitative comparisons. Moreover, **DiffMix** enhances the segmentation and classification performance of two state-of-the-art networks, HoVer-Net and SONNET, even in imbalanced datasets, such as CoNSeP. Our results suggest that **DiffMix** can be used to improve the performance of medical image processing tasks in various applications. In the future, we plan to improve the performance of the diffusion model to generate various pathological tissue types.

**Acknowledgements.** This work was partially supported by the National Research Foundation of Korea (NRF-2019M3E5D2A01063819, NRF-2021R1A6A1A 13044830), the Institute for Information & Communications Technology Planning & Evaluation (IITP-2023-2020-0-01819), the Korea Health Industry Development Institute (HI18C0316), the Korea Institute of Science and Technology (KIST) Institutional Program (2E32210 and 2E32211), and a Korea University Grant.

## References

1. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. In: Osdi, Savannah, GA, USA, vol. 16, pp. 265–283 (2016)
2. Chapelle, O., Weston, J., Bottou, L., Vapnik, V.: Vicinal risk minimization. *Adv. Neural Inf. Process. Syst.* **13**, 1–7 (2000)
3. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. *arXiv preprint [arXiv:1708.04552](https://arxiv.org/abs/1708.04552)* (2017)
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. *Adv. Neural Inf. Process. Syst.* **34**, 8780–8794 (2021)
5. Doan, T.N.N., Kim, K., Song, B., Kwak, J.T.: Gradmix for nuclei segmentation and classification in imbalanced pathology image datasets. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022*. LNCS, pp. 171–180. Springer, Heidelberg (2022). [https://doi.org/10.1007/978-3-031-16434-7\\_17](https://doi.org/10.1007/978-3-031-16434-7_17)
6. Doan, T.N., Song, B., Vuong, T.T., Kim, K., Kwak, J.T.: Sonnet: a self-guided ordinal regression neural network for segmentation and classification of nuclei in large-scale multi-tissue histology images. *IEEE J. Biomed. Health Inf.* **26**(7), 3218–3228 (2022)
7. Gong, X., Chen, S., Zhang, B., Doermann, D.: Style consistent image generation for nuclei instance segmentation. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3994–4003 (2021)
8. Graham, S., et al.: Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019)

9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020). <https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf>
10. Hou, L., Agarwal, A., Samaras, D., Kurc, T.M., Gupta, R.R., Saltz, J.H.: Robust histopathology image analysis: to label or to synthesize? In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8533–8542 (2019)
11. Li, X., et al.: A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. *Artif. Intell. Rev.* **55**(6), 4809–4878 (2022)
12. Lin, Y., Wang, Z., Cheng, K.T., Chen, H.: InsMix: towards realistic generative data augmentation for nuclei instance segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022*. LNCS, vol. 13432, pp. 140–149. Springer, Heidelberg (2022). [https://doi.org/10.1007/978-3-031-16434-7\\_14](https://doi.org/10.1007/978-3-031-16434-7_14)
13. Nichol, A., et al.: Glide: towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741* (2021)
14. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: *International Conference on Machine Learning*, pp. 8162–8171. PMLR (2021)
15. Raza, S.E.A., et al.: Micro-net: a unified model for segmentation of various objects in microscopy images. *Med. Image Anal.* **52**, 160–173 (2019)
16. Simard, P.Y., LeCun, Y.A., Denker, J.S., Victorri, B.: Transformation invariance in pattern recognition—tangent distance and tangent propagation. In: Orr, G.B., Müller, K.-R. (eds.) *Neural Networks: Tricks of the Trade*. LNCS, vol. 1524, pp. 239–274. Springer, Heidelberg (1998). [https://doi.org/10.1007/3-540-49430-8\\_13](https://doi.org/10.1007/3-540-49430-8_13)
17. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020)
18. Wang, H., Xian, M., Vakanski, A., Shareef, B.: Sian: style-guided instance-adaptive normalization for multi-organ histopathology image synthesis. *arXiv preprint arXiv:2209.02412* (2022)
19. Wang, W., et al.: Semantic image synthesis via diffusion models. *arXiv preprint arXiv:2207.00050* (2022)
20. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022*. LNCS, vol. 13438, pp. 35–45. Springer, Heidelberg (2022). [https://doi.org/10.1007/978-3-031-16452-1\\_4](https://doi.org/10.1007/978-3-031-16452-1_4)
21. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: regularization strategy to train strong classifiers with localizable features. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2019)
22. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412* (2017)
23. Zhao, B., et al.: Triple u-net: hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. *Med. Image Anal.* **65**, 101786 (2020)