This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

URCDM: Ultra-Resolution Image Synthesis in Histopathology

Sarah Cechnicka[1], James Ball[1], Matthew Baugh[1], Hadrien Reynaud[1], Naomi Simmonds[2], Andrew P.T. Smith[2], Catherine Horsfield[2], Candice Roufosse[2], and Bernhard Kainz[1,3]

[1] Departament of Computing, Imperial College London, UK
[2] NHS Trust, London, UK
[3] Friedrich–Alexander University Erlangen–Nürnberg, DE
sc7718@imperial.ac.uk

Abstract. Diagnosing medical conditions from histopathology data requires a thorough analysis across the various resolutions of Whole Slide Images (WSI). However, existing generative methods fail to consistently represent the hierarchical structure of WSIs due to a focus on high-fidelity patches. To tackle this, we propose Ultra-Resolution Cascaded Diffusion Models (URCDMs) which are capable of synthesising entire histopathology images at high resolutions whilst authentically capturing the details of both the underlying anatomy and pathology at all magnification levels. We evaluate our method on three separate datasets, consisting of brain, breast and kidney tissue, and surpass existing state-of-the-art multi-resolution models. Furthermore, an expert evaluation study was conducted, demonstrating that URCDMs consistently generate outputs across various resolutions that trained evaluators cannot distinguish from real images. All code and additional examples can be found on GitHub.

Keywords: Generative · Diffusion · Histopathology · Gigapixel Images.

1 Introduction

Professional pathologists rely on both low and high-magnification views to make critical decisions. At lower magnifications, they assess overall tissue architecture to identify areas of interest, such as tumour distribution. Medium magnifications allow for closer examination of smaller tissue structures and cell groupings, providing details on cellular changes. At the highest magnifications, individual cell morphology is scrutinized for specific abnormalities crucial for accurate diagnosis. Each level of magnification contributes essential information, enabling a comprehensive pathological assessment.

Machine learning techniques [29,13] could partially automate such diagnosis, but, unfortunately, access to Whole Slide Image (WSI) data from real patient populations is often restricted due to privacy concerns, making it difficult to develop robust models. Diffusion models have demonstrated significant value in

this area, facilitating the creation of synthetic data for medical education and the development of machine learning models for automated analysis, all while ensuring the protection of patient privacy.

In histopathology, where staining differences hinder generalisability, diffusion models achieve a more dependable performance across diverse data sources [1]. Most research focuses on the use of generative approaches at small-scale patch levels, where these techniques excel at generating local, high-fidelity details [9,15,25]. However, significant challenges arise when transitioning from patch-level analysis to WSI, such as memory constraints, long sampling times, and a lack of training data. Additionally, these methods risk data curation biases, as training and validation datasets for patch-level problems are often compiled by the same experts or under identical guidelines, potentially skewing task outcomes [4]. Many existing methods are not applicable for large context images, as they were originally designed for much lower resolutions [23,17]. This transition introduces complexities related to preserving contextual information and ensuring seamless integration between different magnification levels. The development of tailored algorithms capable of handling such scale is therefore crucial. This is particularly important as accurate generation and analysis at the WSI level increases the number and variety of available data samples. This also enables the use of more complex downstream algorithms, which operate on the entire image at different scales, *e.g.*, You Only Look Twice [5].

Contributions. We introduce a method leveraging Ultra-Resolution Cascaded Diffusion Models (URCDM)s to produce high-fidelity, photo-realistic histopathology images at the WSI scale, marking a first in the field. Our approach uniquely captures detailed features at multiple magnifications and supports long-range contextual understanding, overcoming the memory limitations of attention -based models. It achieves this at a significantly reduced computational cost, facilitating efficient image generation even in data-intensive WSI learning contexts.

Related Work. The strategy of enhancing datasets with artificially created images has been a long-standing idea, yet its realisation evolved slowly [26,14,7]. Techniques based on auto-encoders [11] or Generative Adversarial Networks (GAN)s [6] often exhibited significant limitations, mainly due to a noticeable domain gap between real and synthetic images. This makes machine learning on enriched data difficult since it is often easier for downstream application models, like pathology detection, to discriminate data sources over desired training signals, which introduces significant confounders. This is further exasperated when dealing with gigapixel images like WSI, when more information is handled and a bottom-up or top-down approach through patching is required.

Diffusion models have been shown to produce synthetic images that better match the distribution of target datasets [18,21,22]. Recent applications of Cascaded Diffusion Model (CDM)s [22] have improved multi-scale information integration in medical imaging, pushing the boundaries of diagnostic accuracy and patient care [19]. These diffusion models are chained together, first generating a low-resolution image, which gets repeatedly upsampled by each model in the cascade. Compared to attempting to train one model directly at the target

resolution, this process allows for smaller models and parallel training, making these large-scale experiments much more tractable in time,compute resources and allows for seamless edges.

2 Method

Diffusion models have emerged as powerful generative models, capable of being parameterized through various approaches to accurately model complex data distributions. In this work, we follow the Stochastic Differential Equation definition [24]. We define our forward (Eq. 1) and backward (Eq. 2) processes as:

$$\mathrm{d}\mathbf{x}_+ = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}, \tag{1}$$

$$\mathrm{d}\mathbf{x}_- = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}, \tag{2}$$

where $\mathrm{d}\mathbf{x}_+$ is the noise to be added to the sample, while $\mathrm{d}\mathbf{x}_-$ is the noise to be removed from the sample, following the epsilon (noise prediction) objective. $\mathbf{w}$ and $\bar{\mathbf{w}}$ are forward and backward Wiener processes, respectively introducing stochasticity. $\mathbf{f}(\cdot, t)$ is the drift coefficient that models the deterministic part of the state's change in the diffusion process over time. $g(\cdot)$ is the diffusion coefficient of $\mathbf{x}(t)$. $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is the score function we learn with a neural network. We use this parametrisation for all our diffusion models.

CDMs work by first generating a low-resolution image I_0 with a base model C_{ϕ_0} from Gaussian noise. I_0 is subsequently used as a conditional input for a second diffusion model C_{ϕ_1}[16], which generates an image of higher resolution I_1, that is based on I_0. Generally, the nth super-resolution stage will condition on the lower-resolution image I_{n-1} generated in the previous stage. We can write

$$C_{\phi_n}\left(\sigma, I_{n-1}\right) = F_{\phi_n}\left(c_{\text{noise}}\left(\sigma\right), I_{n-1}\right), C_{\phi_0}\left(\sigma\right) = F_{\phi_0}\left(c_{\text{noise}}\left(\sigma\right)\right). \tag{3}$$

URCDMs extend this principle further. The overall pipeline consists of three stages which form a cascade, with each outputting a WSI at higher and higher resolutions, as shown in Figure 1. Within each stage there is a CDM, which aims to upsample a specific region of interest. Every CDM starts by generating an image of dimension 64×64 and then upscales it twice, first to 256×256 and then to 1024×1024.

The first stage directly uses a standard CDM to output an entire WSI of size 1024×1024. The second stage then uses a separate CDM to upsample patches of this image using a combination of conditioning and inpainting. We adapt the CDM so that each stage is also conditioned on a 1024×1024 patch of the image from the previous stage, centered on the patch being upsampled. This allows it to see both the region for upsampling and its surrounding context, ensuring that the generated patch is consistent with the high-level structure of the image. Our second adaptation starts with how we use the CDM to generate a grid of overlapping patches. When generating a new patch, using inpainting, any pixels in regions that have already been generated as part of neighbouring patches are kept constant throughout the reverse diffusion process. Giving the diffusion model

this information enables it to produce patches that seamlessly blend together without any artifacts between patch boundaries. At the end of stage 2 all the upsampled patches are combined together to make a WSI of size 6400×6400.

Stage 3 then follows the same structure as stage 2, except now the images used for conditioning are taken from the output of stage 2, resulting in the final WSI having a resolution of $41\,344 \times 41\,344$.

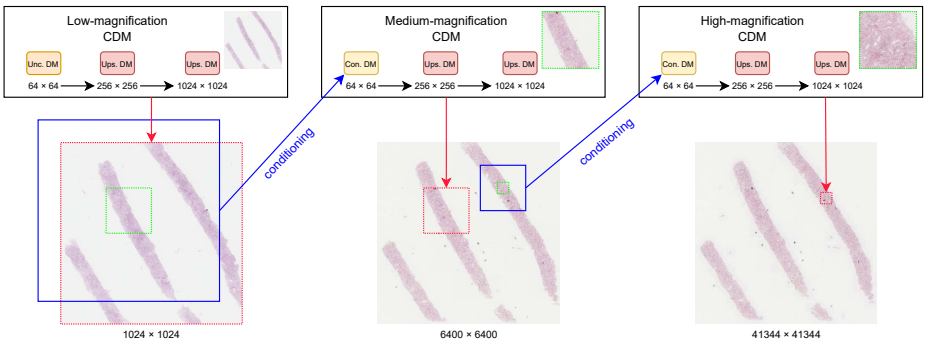

Fig. 1: Detailed overview of the URCDM image generation process. The medium and high-magnification CDMs are sampled many times, and the patches generated are stitched together; one sample is shown as an example. A blue outline indicates the lower-magnification conditioning image, to teach the context for the new generation process. A green outline indicates the resultant patch that will be 'zoomed in' on and generated by a baseline CDM. Red lines indicate the output of each magnified image. Not to scale.

In our proposed URCDM, there are nine diffusion models in total that form three separate CDMs. All models are trained independently and in parallel on nine separate Nvidia A100 GPUs, which takes 24 hours for 200,000 steps. During inference a low-quality WSI is generated with the use of the low-resolution CDM with size 1024×1024. Overlapping patches of this generated low-resolution image are then used to condition the second CDM, by giving it the spatial context needed for generation. The second model generates images of size 1024×1024 for the centre or each conditioning patch. After stitching them together, and accounting for overlaps, the resolution of the WSI increases to 6400×6400. This process is repeated for the final high-resolution model, yielding a final synthetic WSI of $41\,344 \times 41\,344$ pixels, which takes approximately 24h on nine separate Nvidia A100 GPUs to generate. The size of the medium-magnification CDM was chosen to be approximately halfway between low-resolution and high-resolution magnifications. The URCDM is not restricted to these resolutions, which can trivially be changed to suit other datasets.

Implementation details: We base on Imagen [22] using the `imagen-pytorch` [27] library. Each CDM targets a different magnification of the overall image. Gradi-

ent clipping was implemented (set to 1) as well as v-parametrisation, to avoid lower quality fine details, blurring or heavy distortion of lower magnification images when 'zooming in'. Inpainting [3] is used to smoothly merge generated patches together with minimal seams. The resulting dependency between patches is especially important when batch-processing images or sampling on multiple GPUs, as it determines when patches can be generated in parallel. This ensures reasonable sampling speeds. 12.5% of overlap (both in the vertical and horizontal directions) was chosen for our experiments to minimize patching artefacts while reasonably decreasing the total number of patches generated. High-magnification patches of the whole slide image that are mostly white are ignored in both training and sampling. Instead, all white patches are replaced with an upscaled version of the medium-magnification image. All images were cropped or padded to $41\,300 \times 41\,300$ pixels.

3 Evaluation and Results

Datasets and Preprocessing: We explore our approach on three datasets for evaluation. The first dataset is the public low-grade glioma dataset from The Cancer Genome Atlas (TCGA) archive [20] containing 344 WSIs, hereafter referred to as GLIOMA. The second is a public The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) dataset [28]. The dataset comprises 1978 high-resolution WSIs of various tissues that have been stained using a variety of protocols, designated as BREAST. The final dataset is a proprietary collection of 428 Kidney Transplant Pathology WSI slides with an average resolution of $40\,000 \times 40\,000$ per slide, named KIDNEY. Ultra-resolution images must consistently appear realistic at multiple scales both when 'zoomed-out' and when 'zoomed-in'. Lower-resolution patches generated by a Lower Resolution Diffusion Model (LRDM) are evaluated against StyleGAN3 [10], and the Morphology Focused Diffusion Probabilistic Model (MorphDiff) ([17]. To accommodate the size limitations of the models, which handle images up to 1024×1024, all WSIs were segmented into patches. These patches were then resized to 64×64 for the initial diffusion model training, augmented with basic transformations, and further scaled to 256×256 for the first super-resolution model. An LRDM has been chosen as the backbone for outpainting for fair comparison 1.

Secondly, long-distance spatial coherency is evaluated. The realism of high-resolution images is compared against baseline unconditional diffusion models generating high-resolution images using outpainting. All datasets used for experimenting on WSI level are first cropped to patches size $40\,000 \times 40\,000$ pixels. These images are then downsampled to 1024×1024 for the low magnification models and undergo the same transformations and resizing to 256×256 and 64×64 as in the low-resolution models for training. The $40\,000 \times 40\,000$ patches are then separately resized to 6400×6400 and cropped to 1024×1024 for the mid-magnification model and undergo the same training preprocessing as the LRDM. Finally, for the high-magnification model the $40\,000 \times 40\,000$ patches are

cropped to 1024×1024 patches that then follow the same training preprocessing as all other models.

FID and pFID: The Fréchet Inception Distance (FID) [8] is a metric to evaluate the image quality and diversity of generated images by comparing two Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$, and $\mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, where r and g represent real and generated images respectively. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are found by fitting to latent-space feature vectors of real images and generated images [8]. FID becomes impractical for very high-resolution images due to computational constraints and the need to capture local image quality across vast areas. Patch-FID (pFID), introduced by [2] is a variation of FID [8] for very high-resolution images and is used for the evaluation of gigapixel performance. It works by taking random crops at random scales from the image and computing the FID between the real and generated patches by sampling different patches at matching scales and positions [2].

IP and IR: Improved Precision (IP) [12] quantifies the proportion of generated samples that align with the real data manifold.Conversely, Improved Recall (IR) measures the extent to which the real data manifold is encapsulated within the synthetic data manifold, reflecting the model's ability to capture data diversity. These metrics are derived by computing the k-Nearest Neighbors distances for each sample, facilitating a non-parametric approximation of the data manifolds. While FID is a robust and popular metric for comparing the overall similarity of real and generated data distributions, IR and IP offer a more detailed, balanced assessment of a generative model's ability to produce diverse, high-quality samples that closely match the characteristics of the real data.

Metric Implementation: To accurately model URCDM's capabilities various post-processing steps were meticulously followed. For pFID-50k we utilized various crops at different magnifications, all resized to 1024×1024 pixels. These were kept consistent with the training dataset for each cascading level, facilitating an accurate measure of the generated images' fidelity compared to the real images. Given the computational intensity of our method, especially at higher magnifications, we use IP and IR metrics on approximately 10 000 patches for each magnification level, except for magnification 0, which was limited to the real samples available. For magnifications 1 and 2, the analysis was based on roughly 10 000 patches randomly sampled from about 800 generated images. Although this limited the performance on both magnifications, this decision was necessary as generating a single full-size WSI at magnification 2 requires approximately 24 hours on 8 NVIDIA A100 GPUs. The values in Table 1 are then uniformly averaged over these three performance measurements.

Results: The quantitative results of the experiments can be seen in Tables 1, 3. Table 1 provides a general comparison with other state-of-the-art models whereas Table 3 shows detailed generative results for our model across 3 magnification scales. Figure 2 shows randomly selected example images generated by the URCDM and comparison methods. URCDMs consistently outperform all state-of-the-art methods on all three datasets. Generally, URCDMs exhibit considerable plausible diversity across different regions at high magnification levels while

Table 1: Results on the unconditional generative model and baselines trained on all datasets report FID-10k, IP and IR on 10k image of 1024×1024 patches (MorphDiff [17] uses 128×128). Results of pFID-50k, IP and IR on different methods of generating ultra-resolution images using diffusion models on 1024×1024 patches.

	Model	KIDNEY				GLICOMA				BREAST			
		Fidelity ↓		IP ↑	IR ↑	Fidelity ↓		IP ↑	IR ↑	Fidelity ↓		IP ↑	IR ↑
		FID_{10k}	$pFID_{50k}$			FID_{10k}	$pFID_{50k}$			FID_{10k}	$pFID_{50k}$		
patch	StyleGAN3 [10]	38.62	n.a	0.74	0.08	**48.59**	n.a	**0.72**	0.08	-	n.a	-	-
patch	Morph-Diff [17]	NO PUBLIC CODE				20.11	n.a	0.26	0.85	NO PUBLIC CODE			
patch	LRDM	**10.35**	n.a	**0.82**	**0.57**	34.31	n.a	0.70	**0.85**	66.17	n.a	0.91	0.30
WSI	Outpainting [3]	n.a	150.15	0.63	0.13	n.a	192.89	0.60	0.23	n.a	166.24	0.56	0.20
WSI	URCDM (ours)	n.a	**39.52**	**0.70**	**0.18**	n.a	**52.38**	**0.76**	**0.51**	n.a	**67.66**	**0.69**	**0.31**

Table 2: Human expert evaluation. TP = true positives, FP = false positives. p = proportion of incorrectly classified samples. The final averaged total values in bold for the expectation corrected proportion $|p - 0.5|$.

	User	TP	FP	p	$\lvert p-0.5\rvert$		TP	FP	p	$\lvert p-0.5\rvert$
LRDM	Pathol. 1	250	179	0.4172	0.0823	URCDM	61	66	0.5197	0.0197
LRDM	Pathol. 2	106	145	0.5777	0.0777	URCDM	152	6	0.0380	0.4620
LRDM	Pathol. 3	29	99	0.7734	0.2734	URCDM	28	33	0.5410	0.0410
LRDM	Pathol. 4	-	-	-	-	URCDM	47	10	0.1754	0.3246
	Non-expert	110	162	0.5956	0.0956		29	21	0.4200	0.0800
	Total	495	585	0.5417	**0.1074**		317	136	0.3002	**0.2219**

Table 3: Detailed average IP and IR metrics for 1024×1024 patches across three magnification levels for all three datasets.

Model	mag. 0		mag. 1		mag. 2	
	IP↑	IR↑	IP↑	IR↑	IP↑	IR ↑
KIDNEY	0.87	0.42	0.60	0.10	0.62	0.03
BREAST	0.54	0.45	0.70	0.20	0.83	0.27
GLICOMA	0.76	0.63	0.61	0.40	0.91	0.50

preserving a high quality. At lower magnification levels, the images accurately represent the fundamental structures of the WSI.

Human perception study: To assess the perceived realism of synthetic images, a blinded comparison platform was developed, presenting expert pathologists with pairs of real and synthetic images from the KIDNEY dataset, and tasking them with identifying the authentic one. Results are shown in Table 2. A p-value close to 0.5 means that the users are essentially guessing. Pathologists 2 and 4 were familiar with our approach and found a shortcut during high magnification evaluation. They looked for rare sample preparation artefacts, like tissue folding, which are not present in URCDM since these parts are discarded during regular pathology assessment. In Table 2 the proportion of incorrectly classified samples p varies considerably between users, with some having a strong preference for real images, and others having a strong preference for fake images. This balances out with the total proportion being close to 0.5 for LRDM.

Discussion URCDM presents a shift from outpainting-based approaches. At a low resolution LRDM outperform both StyleGans and MorphDiff [17]. This is likely due to GANs often causing unnatural symmetries to appear in the images, including ringing artefacts (see Figure 2 KIDNEY/StyleGAN). The differential

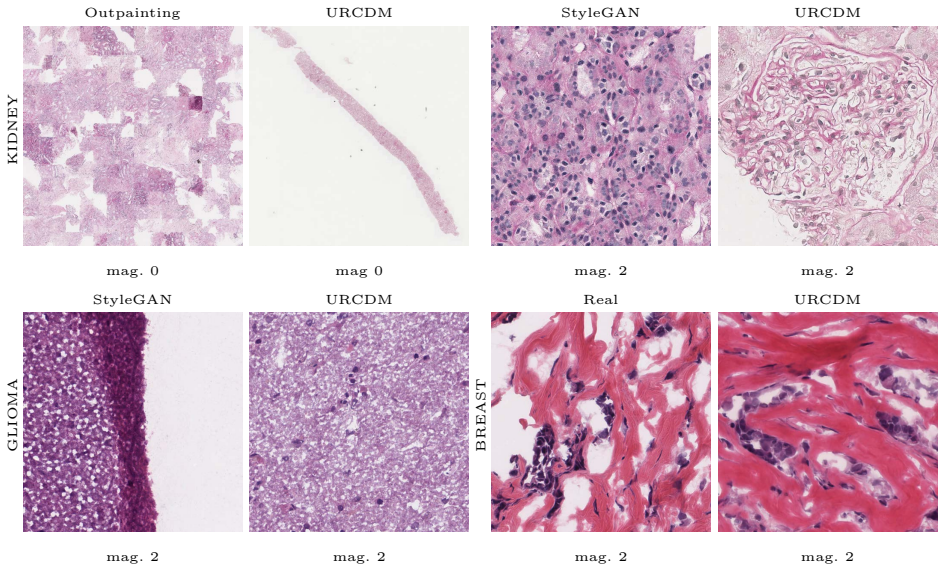

Fig. 2: Random URCDM samples vs. Outpainting and StyleGAN. More qualitative examples can be found in the Appendix.

learning strategy employed by CDMs, where distinct networks incrementally learn at each stage of the cascade, offers a tangible advantage over MorphDiff [17].

Occasionally, at high magnification factors, isolated details of the images generated using outpainting are subjectively of higher quality, however, the lack of spatial coherency leads to poor pFID, IP and IR scores. The texture of the tissue for the KIDNEY dataset in Table 2 is realistically wispy in regions of the kidney further from the cortex, like in a real WSI, the BREAST and GLIOMA datasets also follow similar spacial alignments. Additionally, structures that are only visible at high magnification are of high quality, like the glomerulus and tubules in the KINDEY or the smooth tissue surrounded by nuclei in BREAST in magnification 2 of Figure 2. For trained pathologists, some fine details in the high-resolution image of Figure 2 can be poorer than those in the real WSIs, which is likely due to URCDMs having to maintain consistency through all nine stages. Base U-Nets are required to consistently and accurately 'zoom in' the lower magnification image, and failure to do so will result in visually diverging neighbouring patches. Moreover, the performance analysis in Table 3 illustrates a disparity in image quality across varying magnifications, underscoring a specific challenge in generating mid-magnification images. This disparity is primarily attributed to the constrained sampling process for magnification level 1, where the next stage relies on fully generated images, limiting the diversity and potentially the realism of the generated images. The unique challenge presented by the BREAST dataset, where larger image sizes mean that the $40\,000 \times 40\,000$ centre-crop may frequently capture non-informative white space, highlights the impact

of dataset characteristics on the model's performance; this also lead to StyleGAN failing to learn the distribution.

4 Conclusion

URCDMs are a novel way of generating images with more than 10^9 pixels. Images generated by URCDMs are spatially coherent over long distances and are plausible at different scales. Fine details remain clear and are more coherent than images from outpainting, which is crucial for ultra-resolution imagery applications that use various image scales [5]. Future work will focus on computational efficiency, the use of the same CDM for all magnifications, and multi-modal learning.

Acknowledgments: S. Cechnicka is supported by the UKRI Centre for Doctoral Training AI4Health (EP / S023283/1). Support was also received from the ERC project MIA-NORMAL 101083647, the State of Bavaria (HTA) and DFG 512819079. HPC resources were provided by NHR@FAU of FAU Erlangen-Nürnberg under the NHR project b180dc. NHR@FAU hardware is partially funded by the DFG – 440719683. Dr. Roufosse is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London (ICL). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Dr Roufosse's research activity is made possible with generous support from Sidharth and Indira Burman. Human samples used in this research project were obtained from the Imperial College Healthcare Tissue & Biobank (ICHTB). ICHTB is supported by NIHR Biomedical Research Centre based at Imperial College Healthcare NHS Trust and ICL. ICHTB is approved by Wales REC3 to release human material for research (22/WA/2836)

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

References

1. Cechnicka, S., Ball, J., Reynaud, H., Arthurs, C., Roufosse, C., Kainz, B.: Realistic data enrichment for robust image segmentation in histopathology. In: Koch, L., Cardoso, M.J., Ferrante, E., Kamnitsas, K., Islam, M., Jiang, M., Rieke, N., Tsaftaris, S.A., Yang, D. (eds.) Domain Adaptation and Representation Transfer. pp. 63–72. Springer Nature Switzerland, Cham (2024)
2. Chai, L., Gharbi, M., Shechtman, E., Isola, P., Zhang, R.: Any-resolution training for high-resolution image synthesis. In: European Conference on Computer Vision (2022)
3. Cheng, Y.C., Lin, C.H., Lee, H.Y., Ren, J., Tulyakov, S., Yang, M.H.: Inout: Diverse image outpainting via gan inversion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11431–11440 (2022)
4. Ciga, O., Xu, T., Nofech-Mozes, S., Noy, S., Lu, F.I., Martel, A.L.: Overcoming the limitations of patch-based learning to detect cancer in whole slide images. Scientific Reports | **11**, 8894 (123). https://doi.org/10.1038/s41598-021-88494-z

5. Etten, A.V.: You only look twice: Rapid multi-scale object detection in satellite imagery (2018)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
7. Gupta, L., Klinkhammer, B.M., Boor, P., Merhof, D., Gadermayr, M.: Gan-based image enrichment in digital pathology boosts segmentation accuracy. In: MICCAI 2019, Part I 22. pp. 631–639. Springer (2019)
8. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium (2017). https://doi.org/10.48550/ARXIV.1706.08500, https://arxiv.org/abs/1706.08500
9. Jose, L., Liu, S., Russo, C., Nadort, A., Di Ieva, A.: Generative adversarial networks in digital pathology and histopathological image processing: A review. Journal of Pathology Informatics **12**(1), 43 (2021). https://doi.org/https://doi.org/10.4103/jpi.jpi_103_20, https://www.sciencedirect.com/science/article/pii/S2153353922001651
10. Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-free generative adversarial networks. In: Proc. NeurIPS (2021)
11. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
12. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., Aila, T.: Improved precision and recall metric for assessing generative models. CoRR **abs/1904.06991** (2019)
13. Laak, J., Litjens, G., Ciompi, F.: Deep learning in histopathology: the path to the clinic. Nature Medicine (2021). https://doi.org/10.1038/s41591-021-01343-4, https://doi.org/10.1038/s41591-021-01343-4
14. Lin, Y., Wang, Z., Cheng, K.T., Chen, H.: Insmix: Towards realistic generative data augmentation for nuclei instance segmentation. In: MICCAI 2022, Part II. pp. 140–149. Springer (2022)
15. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, June 28 - July 1, 2009. pp. 1107–1110. IEEE (2009). https://doi.org/10.1109/ISBI.2009.5193250, https://drive.google.com/file/d/1eZGi1wUdyxVOYADXUbxZiVtajlztSnGL
16. Meng, C., Rombach, R., Gao, R., Kingma, D.P., Ermon, S., Ho, J., Salimans, T.: On distillation of guided diffusion models (2023)
17. Moghadam, P.A., Dalen, S.V., Martin, K.C., Lennerz, J., Yip, S., Farahani, H., Bashashati, A.: A morphology focused diffusion probabilistic model for synthesis of histopathology images. In: 2023 IEEE conference on computer vision and pattern recognition (2023)
18. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, Alec, e.a.: Zero-Shot Text-to-Image Generation (February 2021), arXiv:2102.12092
19. Reynaud, H., Qiao, M., Dombrowski, M., Day, T., Razavi, R., Gomez, A., Leeson, P., Kainz, B.: Feature-conditioned cascaded video diffusion models for precise echocardiogram synthesis. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 142–152. Springer Nature Switzerland, Cham (2023)
20. RL, G., AP, H., V, F., HE, V., DR, L., WA, K., LM., S.: Toward a shared vision for cancer genomic data. N Engl J Med. (2016). https://doi.org/10.1056/NEJMp1607591

21. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models (April 2022), arXiv:2112.10752
22. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S.K.S., Ayan, B.K., Mahdavi, S.S., Lopes, R.G., Salimans, T., Ho, J., Fleet, D.J., Norouzi, M.: Photorealistic text-to-image diffusion models with deep language understanding (2022). https://doi.org/10.48550/ARXIV.2205.11487, https://arxiv.org/abs/2205.11487
23. Shrivastava, A., Fletcher, P.T.: Nasdm: Nuclei-aware semantic histopathology image generation using diffusion models. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 786–796. Springer Nature Switzerland, Cham (2023)
24. stanford Song, Y.S., Sohl-Dickstein, J., Brain, G., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. ICLR (2021)
25. Wagner, N., Fuchs, M., Tolkach, Y., Mukhopadhyay, A.: Federated stain normalization for computational pathology. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 14–23. Springer Nature Switzerland, Cham (2022)
26. Wang, J., Perez, L., et al.: The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks Vis. Recognit **11**(2017), 1–8 (2017)
27. Wang, P.: lucidrains/imagen-pytorch: Implementation of Imagen, Google's Text-to-Image Neural Network, in Pytorch — github.com. https://github.com/lucidrains/imagen-pytorch (2022), [Accessed 12-Nov-2022]
28. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature Publishing Group (2013). https://doi.org/10.1038/ng.2764, http://www.cancergenome.nih.gov/.
29. Wu, B., Moeckel, G.: Application of digital pathology and machine learning in the liver, kidney and lung diseases. Journal of Pathology Informatics **14**, 100184 (2023). https://doi.org/10.1016/j.jpi.2022.100184