ArCSEM: Artistic Colorization of SEM Images via Gaussian Splatting

Takuma Nishimura[1] Andreea Dogaru[1] Martin Oeggerli[2] Bernhard Egger[1]
[1]Friedrich-Alexander-Universität Erlangen-Nürnberg [2]Micronaut
{takuma.nishimura, andreea.dogaru, bernhard.egger}@fau.de info@micronaut.ch

Abstract

Scanning Electron Microscopes (SEMs) are widely renowned for their ability to analyze the surface structures of microscopic objects, offering the capability to capture highly detailed, yet only grayscale, images. To create more expressive and realistic illustrations, these images are typically manually colorized by an artist with the support of image editing software. This task becomes highly laborious when multiple images of a scanned object require colorization. We propose facilitating this process by using the underlying 3D structure of the microscopic scene to propagate the color information to all the captured images, from as little as one colorized view. We explore several scene representation techniques and achieve high-quality colorized novel view synthesis of a SEM scene. In contrast to prior work, there is no manual intervention or labelling involved in obtaining the 3D representation. This enables an artist to color a single or few views of a sequence and automatically retrieve a fully colored scene or video. Project page: https://ronly2460.github.io/ArCSEM

1. Introduction

We use images captured by a Scanning Electron Microscope (SEM). SEMs are used for the examination and analysis of nanoscale structures. An electron gun generates a beam that thoroughly scans the surface. As the beam interacts with the surface, it emits signals. The microscope's detectors capture these emitted electrons. The quantity of electrons detected from each point is then converted into corresponding pixel values, resulting in a high-resolution grayscale image that reveals the intricate surface structure. SEM images share similarities with optical images, exhibiting diffuse and specular reflectance and effects similar to optical shadowing. The fundamental distinction lies in the particle flow: in SEM imaging, the particles travel in the opposite direction compared to optical imaging.

We experiment with multi-view grayscale images of a pollen granule captured by tilting the sample while keeping the microscope fixed. However, tilting alters the incident angle between the electron beam and the surface, which causes the emitted electrons to vary across regions of the sample. This induces view-dependent variations in electron emission and scattering, which are perceived as illumination changes in the final SEM images.

Leveraging our captured grayscale dataset, we achieve novel view synthesis (NVS) via a precise 3D representation of the pollen modeled with 2D Gaussian Splatting (2DGS) [11]. To address the aforementioned illumination variations, we apply an image specific affine color transformation (ACT) to the Gaussians, as proposed by [4]. Based on the 3D representation, we introduce colors into the scene, guided by artistic intuition. In addition to our grayscale dataset, we incorporate up to five color images created by a professional artist, Martin Oeggerli. These color images are then used for the colorization of the scene, by adapting ideas from [41] to propagate the color information via pseudo-colors and semantic correspondences across views. We showcase the capabilities of our method, ArCSEM, by generating expressive colorized novel views of an SEM scene with artistic guidance.

The key contributions of our work are as follows:

- We obtain a precise and intricate 3D representation of a pollen captured by SEM, enabling novel view synthesis.
- We achieve 3D colorization of the grayscale 3D scene using a limited number of manually colored images, enabling colored novel view synthesis.

This is a short version of the AI for Visual Arts Workshop paper to be presented at the non-archival ECCV Beyond workshop in ECCV 2024. The full paper includes a comparison to other methods, ablation studies, a statement regarding the limitations of the method, as well as a more detailed related work section. In this short paper, we focus on the proposed method.

2. Related work

Colorization techniques can be broadly categorized into two main approaches: statistical methods [29] and semantic methods [2, 10, 18, 39]. Beyond 2D images, 3D colorization methods have also been developed for meshes [38], point clouds [6, 20], and voxels [35]. The few works [8, 31]

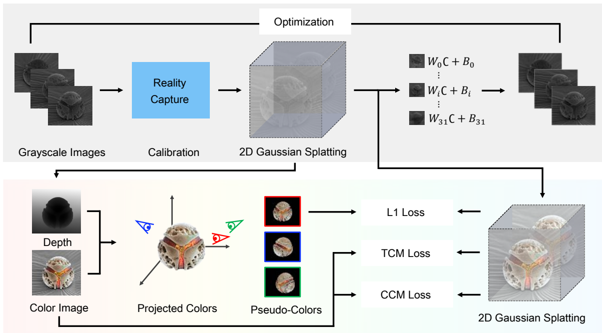


Figure 1. Overview of our two-stage approach: (a) **Grayscale training**: We fit 2DGS [11] with an image-specific affine color transformation to the grayscale images calibrated with RealityCapture [28]. (b) **Colorization**: 2DGS depth maps are used to project colors from limited manually colorized images into 3D space as pseudo-colors. Together with the input color views, the pseudo-colors guide the colorization of the grayscale model via L1, TCM, and CCM loss functions.

that consider the colorization of SEM images are limited to the 2D domain and work without specific color guidance. In contrast, we focus on artistic 3D colorization of SEM images, using specific color inputs, which allows artists to guide and control the process.

Only one line of works [43, 44] has focused on 3D shape reconstruction of complex objects from SEM images, particularly of a cat flea employing traditional photogrammetry and complex computer graphics techniques, which demand extensive mathematical calculations and laborious work. In contrast, our pipeline provides a much simpler method for representing 3D scenes, it is straightforward and requires no customization to model grayscale scenes. NVS techniques, such as Neural Radiance Fields (NeRF) [25], have enabled the photorealistic rendering of arbitrary views. The versatility of these 3D representations enabled the development of various methods for appearance editing with diverse control modalities. Image-guided approaches [3, 12, 21, 26, 37, 41] leverage visual references to control the editing process. Text-based methods [5, 9, 15, 30, 32, 33, 42] employ natural language descriptions to manipulate scene appearances. Other methods [7, 17, 19, 24, 27] allow for manual color specification or tool-based editing. Recently, Gaussian Splatting [11, 14] has emerged as a breakthrough technique, representing scenes explicitly with numerous 3D Gaussians primitives. Providing real-time rendering and competitive quality, the framework has already enabled several editing techniques [1, 13, 22, 36]. Our proposed method belongs to the latter line of works, in contrast to existing approaches that mainly focus on stylization or color replacement using extracted color palettes, we build on Ref-NPR [41] which is more suitable for our grayscale-to-color setting.

3. Method

Our proposed method, illustrated in Fig. 1, employs a two-stage training process: grayscale 3D scene optimization and colorization. We start with a grayscale scene representation by fitting 2DGS on the SEM image dataset calibrated with RealityCapture [28]. To handle varying illumination, we apply an affine color transformation to the decoded color, using image-specific weights and biases. In the second stage, we use the grayscale model to generate depth maps and project the artist-provided colors into 3D space. For views without color data, we use a nearest-neighbor search to obtain pseudo-colors. Finally, we fine-tune the initial grayscale model using the color images and the computed pseudo-colors by keeping the geometry fixed and optimizing the spherical harmonics coefficients for all degrees using losses inspired from Ref-NPR [41].

3.1. Grayscale Training

Our grayscale dataset is acquired using SEM. Although SEM utilizes parallel electron beams, resulting in orthographic projection, we approximate it using perspective projection to ensure compatibility with existing NVS methods. Under this setting, an exceptionally large focal length of approximately 50,000 pixels is estimated, with the cameras positioned very far from the scene content. We obtained satisfying results using RealityCapture [28] with shared intrinsic parameters across all images, while allowing for image-specific distortion coefficients.

We build our method based on 2DGS [11], which uses 2D oriented Gaussian disks to represent the scene. 2DGS succeeded in modeling the SEM images, rendering accurate depth maps and grayscale images. Alternatively, we also considered 3D Gaussian Splatting [14] which represents the scene using 3D Gaussians, but encountered issues especially with visible floaters caused by poor geometry fitting.

To accommodate varying illumination conditions, we employ an affine color transformation (ACT) as proposed by [4]. Each Gaussian holds spherical harmonics coefficients, which are subsequently converted to output intensity values. Prior to rasterization, we apply an image-dependent transformation on the decoded illumination, $\mathbf{L}$. During the subsequent colorization stage, we found that applying this affine transformation led to a degradation in output quality, so we omit it in the second stage. The transformation uses three weights $\mathbf{W} = \{w_1, w_2, w_3\}$ and biases $\mathbf{b} = \{b_1, b_2, b_3\}$ and is defined as $\mathbf{L}' = \mathbf{W} \cdot \mathbf{L} + \mathbf{b}$. When rendering novel views, we average the weights and biases of the training views, resulting in plausible and consistent illumination.

3.2. Colorization

Our colorization method is based on Ref-NPR [41], which was introduced for 3D stylization. We made several key modifications and enhancements to adapt it to our dataset and colorization needs. We rely on three components: pseudo-color supervision for views lacking color information, a Template-based Correspondence Module for propagating colors via the grayscale feature space, and a Coarse Color-Matching Loss to ensure global color consistency.

3.2.1 Pseudo-color supervision.

For color transfer, we first utilize the depth information of the grayscale model to unproject the input pixels of the colored views into the 3D space. Then, for each pixel in the other views we compute a pseudo-color as the color of the closest colored point to its unprojected location. This color is then used as a supervision signal via the loss defined in Eq. 1, where $\hat{\mathbf{C}}_{\text{pc}}$ denotes the pseudo-color, and $\hat{\mathbf{C}}_{\hat{x}}$ refers to the color rendered by the model. If there is no colored point within a given radius, we exclude the respective pixel from the loss calculation.

$$\mathcal{L}_{\text{pseudo-color}} = \frac{1}{N_{\text{pc}}} \left\| \hat{\mathbf{C}}_{\text{pc}} - \hat{\mathbf{C}}_{\hat{x}} \right\|_1 . \tag{1}$$

Ref-NPR employs Reference Ray Registration with a grid system, where colors are assigned to grids and a single color is selected for each pixel. However, this approach does not consider the distance between points during the selection phase; it only filters out pixels that exceed a threshold in the final image, making it difficult to create precise pseudo-colors and potentially resulting in visual artifacts.

Our approach is designed to accommodate high-resolution datasets without encountering memory constraints. The grid-based method in Ref-NPR becomes computationally infeasible for our data, as the higher number of required grids to match our cinematic resolution would exhaust available memory resources.

Ref-NPR also considers the cosine similarity of ray directions in the final image, we found that this factor had minimal impact on quality on our data. Therefore, we simplified our approach by concentrating on spatial proximity.

Pseudo-colors for the background regions often introduce noise and inconsistencies across images, as the number of background pixels varies significantly between images, potentially skewing the loss calculation in Eq. 1. To address these issues, we employ Segment Anything Model [16] to generate precise masks of the pollen granule. This segmentation allows us to effectively extract only the pollen and eliminate the interference of the background elements.

3.2.2 Template-based Correspondence Module.

We employ TCM proposed by Ref-NPR as a loss function to propagate colors to the areas that do no have a ground truth color assigned in the colorized input views, by using matches in the feature space of the grayscale images. This loss minimizes the cosine distance between the features $F_{\hat{I}_g}$ of the rendered color image and a constructed guidance feature $\hat{F}_{I_g}$ of the view I_g. The grayscale image I_g is fed into a VGG network [23] to extract the feature map F_{I_g}. The feature maps of the reference colorized images S_k and their grayscale version I_k are extracted as F_{S_k} and F_{I_k} respectively. For each location (i, j) in the guidance feature map $\hat{F}_{I_g}^{(i,j)}$, we consider the grayscale feature $F_{I_g}^{(i,j)}$ and search for the nearest grayscale feature across reference views $F_{I_k}^{(i^*,j^*)}$ and take the corresponding feature of the colorized image $F_{S_k}^{(i^*,j^*)}$, as defined in Eq. 2.

$$\begin{gathered} \hat{F}_{I_g}^{(i,j)} = F_{S_k}^{(i^*,j^*)}, \\ \text{where}(i^*, j^*), k = \arg \min_{(i',j'),k'} \text{dist} \left(F_{I_g}^{(i,j)}, F_{I_{k'}}^{(i',j')} \right), \end{gathered} \tag{2}$$

$$\mathcal{L}_{\text{TCM}} = \text{dist}(F_{\hat{I}_g}, \hat{F}_{I_g}) \tag{3}$$

3.2.3 Coarse Color-Matching Loss.

Although TCM helps estimate color for occluded regions, it can result in global color inconsistencies and mismatches. To address this limitation, we also consider a coarse color-matching loss [41] that operates at the patch level to minimize color differences, as defined in Eq. 4. Using the index (i^*, j^*) obtained in Eq. 2, let $\bar{C}$ denote the average color of a patch, and C_{S_k} and C_{I_g} refer to the patches in the input color image and rendered image respectively.

$$\mathcal{L}_{\text{coarse-color}} = \frac{1}{N} \sum_{i,j} \| \bar{C}_{I_g}^{(i,j)} - \bar{C}_{S_k}^{(i^*,j^*)} \|_2^2 . \tag{4}$$

3.2.4 Optimization.

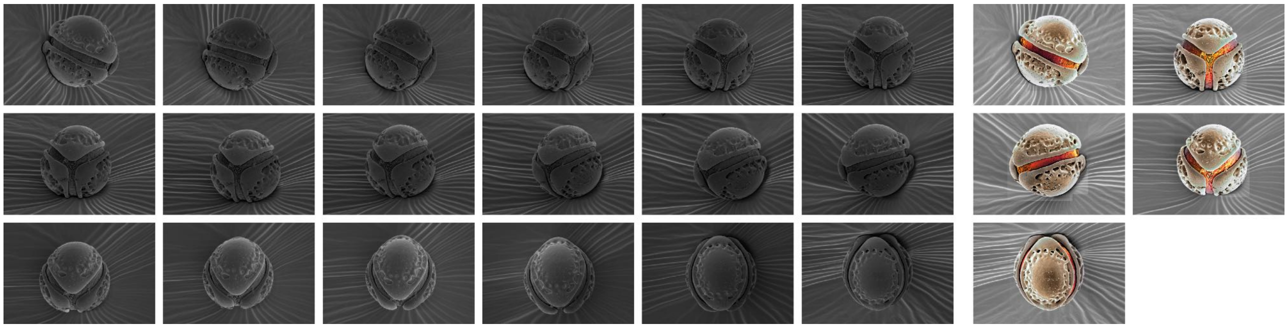
For views with available colorized image, we directly optimize the L1 loss between rendered and input images. For the other views, our final loss function is defined in Eq. 5. λs are respective weights for each loss term.

$$\mathcal{L} = \lambda_{\text{pc}} \mathcal{L}_{\text{pseudo-color}} + \lambda_{\text{TCM}} \mathcal{L}_{\text{TCM}} + \lambda_{\text{cc}} \mathcal{L}_{\text{coarse-color}} \tag{5}$$

4. Experiments

4.1. Dataset

Our dataset comprises 32 high-resolution (3072×2048) SEM images of a pollen, captured along two primary axes. The horizontal axis consists of 20 images spanning from left



(a) Grayscale (b) Color

Figure 2. Our dataset. (a) A subset of 18 out of 32 grayscale images, arranged left to right in the first two rows, and front to top in the bottom row. (b) All manually colored images shown in the following order: leftmost, center, rightmost, angled, and top view

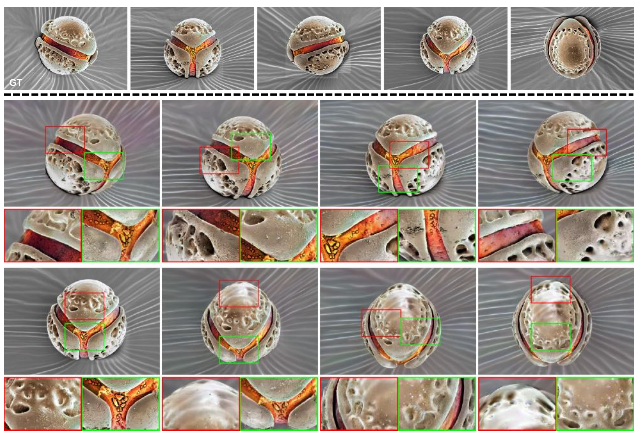


Figure 3. Novel views and closeups generated by our method. Top: All input color images. Bottom: Synthesized novel views at 3072×2048 resolution, with corresponding closeups.

to right, providing a comprehensive lateral view, while the vertical axis includes 12 images, with the camera moving in an arc from the frontal view to the top view of the pollen. We illustrate 18 of these grayscale images in Fig. 2(a), and use the entire set of 32 images as our training dataset. Fig. 2(b) shows colorized versions of 5 images, manually colorized by an artist, Martin Oeggerli.

4.2. Implementation details

The grayscale 2DGS and 3DGS models are trained for 60,000 epochs, with 20,000 additional epochs for colorization. We initialize the ACT by adding small random perturbations to the identity transformation (weights $w_i = 1$ and biases $b_i = 0$, for $i = 1, 2, 3$) and optimize the parameters using a learning rate of 0.0001.

4.3. NVS and Colorization

We qualitatively compare the backbones on grayscale novel view synthesis and quantitatively on rendering the training views by evaluating common image quality metrics: PSNR, SSIM [34], and LPIPS [40].

Model	PSNR (↑)	SSIM (↑)	LPIPS (↓)
3DGS [14]	**37.50**	**0.902**	**0.461**
2DGS [11]	35.25	0.867	0.511
2DGS + ACT (Ours)	36.32	0.890	0.489

Table 1. Evaluation of rendered images. 3DGS and our method achieve the best quantitative scores for reproducing the training views, demonstrating the ability of modelling SEM images.

All models were trained at the full 3072×2048 resolution. Although the quantitative evaluation presented in Tab. 1 indicates that and 3D Gaussian Splatting (3DGS) [14] outperforms in terms of rendering quality of the training views, 2DGS produces superior results in terms of perceived qualitative novel view quality. Moreover, the integration of ACT improves the results of 2DGS across all considered metrics. To obtain the final result, we trained our model with five color images at the original resolution (3072×2048). The generated novel views and corresponding close-ups are shown in Fig. 3. Our method accurately captures fine details such as black outlines of small circular protrusions at the center and the gradual transition of red hues from center to periphery. Overall, our method achieved uniform and high-quality colorization across the entire scene.

5. Conclusion

We achieved cinematic colorization of pollen images captured by a Scanning Electron Microscope. Our approach, which incorporates color projection onto 3D space, affine color transformation, a Template-based Correspondence Module, and a Coarse Color-Matching loss, demonstrated high quality performance on our dataset. From an artistic perspective, introducing color to a monochrome realm offers a visually arresting and mesmerizing experience. Moreover, our method particularly enables us to reduce the number of viewpoints artists need to color manually. By eliminating the manual annotations through our novel view synthesis process, our approach not only enhances efficiency but also opens new creative possibilities for artists.

Acknowledgments

We would like to thank Maximilian Weiherer for valuable discussions and support with the camera calibration. Andreea Dogaru was funded by the German Federal Ministry of Education and Research (BMBF), FKZ: 01IS22082 (IRRW). The authors are responsible for the content of this publication. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project b112dc IRRW. NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) – 440719683.

References

[1] Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, and Guosheng Lin. Gaussianeditor: Swift and controllable 3d editing with gaussian splatting, 2023. 2

[2] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *Proceedings of the IEEE international conference on computer vision*, pages 415–423, 2015. 1

[3] Pei-Ze Chiang, Meng-Shiun Tsai, Hung-Yu Tseng, Wei-Sheng Lai, and Wei-Chen Chiu. Stylizing 3d scene via implicit representation and hypernetwork. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1475–1484, 2022. 2

[4] François Darmon, Lorenzo Porzi, Samuel Rota-Bulò, and Peter Kontschieder. Robust gaussian splatting. *arXiv preprint arXiv:2404.04211*, 2024. 1, 2

[5] Jiahua Dong and Yu-Xiong Wang. Vica-nerf: View-consistency-aware 3d editing of neural radiance fields. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[6] Rongrong Gao, Tian-Zhu Xiang, Chenyang Lei, Jaesik Park, and Qifeng Chen. Scene-level point cloud colorization with semantics-and-geometry-aware networks. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2818–2824. IEEE, 2023. 1

[7] Bingchen Gong, Yuehao Wang, Xiaoguang Han, and Qi Dou. Recolornerf: Layer decomposed radiance fields for efficient color editing of 3d scenes. *arXiv preprint arXiv:2301.07958*, 2023. 2

[8] Israel Goytom, Qin Wang, Tianxiang Yu, Kunjie Dai, Kris Sankaran, Xinfei Zhou, and Dongdong Lin. Nanoscale microscopy images colorization using neural networks. *arXiv preprint arXiv:1912.07964*, 2019. 1

[9] Ayaan Haque, Matthew Tancik, Alexei A Efros, Aleksander Holynski, and Angjoo Kanazawa. Instruct-nerf2nerf: Editing 3d scenes with instructions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19740–19750, 2023. 2

[10] Mingming He, Dongdong Chen, Jing Liao, Pedro V Sander, and Lu Yuan. Deep exemplar-based colorization. *ACM Transactions on Graphics (TOG)*, 37(4): 1–16, 2018. 1

[11] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 1, 2, 4

[12] Yi-Hua Huang, Yue He, Yu-Jie Yuan, Yu-Kun Lai, and Lin Gao. Stylizednerf: consistent 3d scene stylization as stylized nerf via 2d-3d mutual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18342–18352, 2022. 2

[13] Vishnu Jaganathan, Hannah Hanyun Huang, Muhammad Zubair Irshad, Varun Jampani, Amit Raj, and Zsolt Kira. Ice-g: Image conditional editing of 3d gaussian splats, 2024. 2

[14] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 4

[15] Hyunsu Kim, Gayoung Lee, Yunjey Choi, Jin-Hwa Kim, and Jun-Yan Zhu. 3d-aware blending with generative nerfs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22906–22918, 2023. 2

[16] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 3

[17] Zhengfei Kuang, Fujun Luan, Sai Bi, Zhixin Shu, Gordon Wetzstein, and Kalyan Sunkavalli. Palettenerf: Palette-based appearance editing of neural radiance fields. *arXiv preprint arXiv:2212.10699*, 2022. 2

[18] Manoj Kumar, Dirk Weissenborn, and Nal Kalchbrenner. Colorization transformer. In *International Conference on Learning Representations*, 2021. 1

[19] Jae-Hyeok Lee and Dae-Shik Kim. Ice-nerf: Interactive color editing of nerfs via decomposition-aware weight optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3491–3501, 2023. 2

[20] Jitao Liu, Songmin Dai, and Xiaoqiang Li. Pccn: Point cloud colorization network. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3716–3720. IEEE, 2019. 1

[21] Kunhao Liu, Fangneng Zhan, Yiwen Chen, Jiahui Zhang, Yingchen Yu, Abdulmotaleb El Saddik, Shijian Lu, and Eric P Xing. Stylerf: Zero-shot 3d style transfer of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8338–8348, 2023. 2

[22] Kunhao Liu, Fangneng Zhan, Muyu Xu, Christian Theobalt, Ling Shao, and Shijian Lu. Stylegaussian: Instant 3d style transfer with gaussian splatting. *arXiv preprint arXiv:2403.07807*, 2024. 2

[23] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734, 2015. 3

[24] Alessio Mazzucchelli, Adrian Garcia-Garcia, Elena Garces, Fernando Rivas-Manzaneque, Francesc Moreno-Noguer, and Adrian Penate-Sanchez. Irene: Instant recoloring of neural radiance fields, 2024. 2

[25] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106, 2021. 2

[26] Thu Nguyen-Phuoc, Feng Liu, and Lei Xiao. Snerf: stylized neural implicit representations for 3d scenes. *arXiv preprint arXiv:2207.02363*, 2022. 2

[27] Lukas Radl, Michael Steiner, Andreas Kurz, and Markus Steinberger. LAENeRF: Local Appearance Editing of Neural Radiance Fields. In *CVPR*, 2024. 2

[28] Realitycapture. Realitycapture. `http://www.capturingreality.com`. 2

[29] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001. 1

[30] Liangchen Song, Liangliang Cao, Jiatao Gu, Yifan Jiang, Junsong Yuan, and Hao Tang. Efficient-nerf2nerf: Streamlining text-driven 3d editing with multiview correspondence-enhanced diffusion models. *arXiv preprint arXiv:2312.08563*, 2023. 2

[31] Emiel Venema. Colorizing scanning electron microscopy images with diffusion models. Master's thesis, The University of Bergen, 2023. 1

[32] Can Wang, Ruixiang Jiang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. Nerf-art: Text-driven neural radiance fields stylization. *arXiv preprint arXiv:2212.08070*, 2022. 2

[33] Yi Wang, Jing-Song Cheng, Qiao Feng, Wen-Yuan Tao, Yu-Kun Lai, and Kun Li. Tsnerf: Text-driven stylized neural radiance fields via semantic contrastive learning. *Computers & Graphics*, 116:102–114, 2023. 2

[34] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 2004. 4

[35] Zhenpei Yang, Lihang Liu, and Qixing Huang. Learning generative neural networks for 3d colorization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. 1

[36] Dingxi Zhang, Zhuoxun Chen, Yu-Jie Yuan, Fang-Lue Zhang, Zhenliang He, Shiguang Shan, and Lin Gao. Stylizedgs: Controllable stylization for 3d gaussian splatting, 2024. 2

[37] Kai Zhang, Nick Kolkin, Sai Bi, Fujun Luan, Zexiang Xu, Eli Shechtman, and Noah Snavely. Arf: Artistic radiance fields. In *European Conference on Computer Vision*, pages 717–733. Springer, 2022. 2

[38] Mohan Zhang, Jing Liao, and Jinhui Yu. Deep exemplar-based color transfer for 3d model. *IEEE transactions on visualization and computer graphics*, 28(8):2926–2937, 2020. 1

[39] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 649–666. Springer, 2016. 1

[40] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 4

[41] Yuechen Zhang, Zexin He, Jinbo Xing, Xufeng Yao, and Jiaya Jia. Ref-npr: Reference-based non-photorealistic radiance fields for controllable scene stylization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4242–4251, 2023. 1, 2, 3

[42] Jingyu Zhuang, Chen Wang, Lingjie Liu, Liang Lin, and Guanbin Li. Dreameditor: Text-driven 3d scene editing with neural fields. *arXiv preprint arXiv:2306.13455*, 2023. 2

[43] Jasenko Zivanov. *Reconstruction of intricate surfaces from scanning electron microscopy*. PhD thesis, University_of_Basel, 2017. 2

[44] Jasenko Zivanov and Thomas Vetter. Multiview recon-

struction of complex organic shapes. In *BMVC*, pages 157–1, 2015. 2