StyleTex: Style Image-Guided Texture Generation for 3D Models – *Supplementary Material* –

ZHIYU XIE*, State Key Lab of CAD&CG, Zhejiang University, China
YUQING ZHANG*, State Key Lab of CAD&CG, Zhejiang University, China
XIANGJUN TANG, State Key Lab of CAD&CG, Zhejiang University, China
YIQIAN WU, State Key Lab of CAD&CG, Zhejiang University, China
DEHAN CHEN, State Key Lab of CAD&CG, Zhejiang University, China
GONGSHENG LI, Zhejiang University, China
XIAOGANG JIN†, State Key Lab of CAD&CG, Zhejiang University, China

ACM Reference Format:
Zhiyu Xie, Yuqing Zhang, Xiangjun Tang, Yiqian Wu, Dehan Chen, Gongsheng Li, and Xiaogang Jin. 2024. StyleTex: Style Image-Guided Texture Generation for 3D Models – *Supplementary Material* –. 1, 1 (September 2024), 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

In this supplement, we first provide the attributes of 3D models and style images we used in this paper in Sec. A. Following that, in Sec. B, we present the implementation details of our method, including training details (Sec. B.1), the method of texture map extraction (Sec. B.2), details of quantitative evaluation matrix (Sec. B.3), and baseline implementation details (Sec. B.4). Finally, we present the effect of the CFG scale λ_{cfg} and the style guidance scale λ_{style} in Sec. C.

A 3D MODEL / STYLE IMAGE ATTRIBUTION

In this paper, we use 3D models sourced from the Objaverse [Deitke et al. 2023] and Sketchfab [Sketchfab [n. d.]] under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. The models are utilized without their original textures to focus solely on the impact of our stylized texture generation method.

Each model used from Sketchfab is attributed as follows:

- "Baby Animals Statuettes" by Andrei Alexandrescu.
- "Dragon Fruit" by Andrei Alexandrescu.
- "Durian The King of Fruits" by Laithai.
- "Backpack" by mickeymoose1204.
- "Treasure chest" by DailyArt.
- "Molino De Viento _ Windmill" by BC-X.
- "MedivalHouse | house for living | MedivalVilage" by JFredchill.
- "Bouddha Statue Photoscanned" by amcgi.
- "Bunny" by vivienne0716.
- "bird" by rudolfs.
- "Vase .::RAWscan::." by Andrea Spognetta (Spogna).
- "Leather Wooden Chest - 3D scan Quixel Megascans" by Guay0.
- "Stone" by mesropash97.
- "Stone" by Xephira.
- "Stone Entrance" by DJMaesen.
- "Chinese Bridge Ornament" by artfletch.
- "Chinese Hall" by LSDWaterPipe.
- "TeaScroll Clubhouse Scene" by Anaïs Faure.
- "Chinese House" by GloomyGN.
- "Chinese Dragon Fan" by Mrs. Chief.
- "Chinese Lacquer Shanxi Console Table" by Arts and Materials Lab.
- "chinese cup" by Konstantin Morozov.
- "Teaset" by nuts.
- "victorian Cabinet" by lagesnpiet.
- "Sakura Cherry Blossom " by ffish.asia / floraZia.com.
- "Madrona Invasives" by dipietron.
- "Chinese storehouse" by LSDWaterPipe.
- "Oriental Building" by N01516.
- "Chinese style tea table" by DailyArt.
- "Porcelain China Vase" by rz.
- "Chinese lamp" by Coffeek.
- "Carrot Cake" by Greg Zaal.
- "Painted Wooden Chair 02" by Kirill Sannikov.
- "Dutch Ship Large 01" by James Ray Cock.
- "Marble Bust 01" by Rico Cilliers.
- "Brass Vase 03" by Rico Cilliers.
- "Jug 01" by Kuutti Siitonen.
- "Arm Chair 01" by Kirill Sannikov.
- "Wooden Candlestick" by Josh Dean.
- "Carved Wooden Elephant" by Greg Zaal.
- "Pot Enamel 01" by Kuutti Siitonen.
- "Rock Face 02" by Dario Barresi.

*Equal contribution
†Corresponding author.

Authors' addresses: Zhiyu Xie, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China, xiezhiyu@zju.edu.cn; Yuqing Zhang, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China, 3180102110@zju.edu.cn; Xiangjun Tang, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China, xiangjun.tang@outlook.com; Yiqian Wu, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China, onethousand1250@gmail.com; Dehan Chen, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China, cdh573885@outlook.com; Gongsheng Li, Zhejiang University, Hangzhou, Zhejiang, China, ligongshengzju@foxmail.com; Xiaogang Jin, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang, China, jin@cad.zju.edu.cn.

ACM XXXX-XXXX/2024/9-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

- "Boulder 01" by Rico Cilliers.
- "ancient television on a table" by ricksticky.
- "Lowpoly viking helmet" by Dmytro Rohovyi.
- "ballon" by Nyilonelycompany.
- "Woman's shirt" by Jacen Chio.
- "Ratus" by dringoth.
- "Octopus Clay Model [Re-upload]" by abot86.
- "Underwood 4-Bank Typewriter (Portable)" by Ed Swinbourne.
- "MOUNTAIN_ BAKPOD©" by CIMORO.
- "Turtle Project" by Scott Teel.
- "Fishy" by steamsoldier.
- "The Megaphone" by ezgi bakim.
- "sled pig" by maksimpetrik.
- "Low poly army boots" by tipicultbiomassa.
- "Stanford Bunny PBR" by hackmans.
- "Minotaur Statue" by plasmaernst.
- "Stagecoach" by Tuuttipingu.
- "Cartoon Penguin WiP v1" by Drakahn Finlay.
- "Headphones Sony low" by danok98.
- "Box" by KlGrimm.
- "Aged Traffic Cone" by Eydeet.
- "Rep. 17" by SGMADEO_1.
- "#powertool" by Digital Dressmaker.
- "Vintage Gold Pocket Watch" by Daz.
- "Post apocalyptic style retro telephone" by Sousinho.
- "Fan" by Escoly.
- "vaza" by protva2011.
- "Piano" by DarksProducer.
- "Biker" by KulerRuler.
- "Seashell 4K Photogrammetry | Game Ready asset" by Photogrammetry Guy.
- "Damaged Leather Recliner" by Gravity Jack.
- "Dirtbike" by Thunder.
- "Hand Painted SeaHorse" by stormk90.
- "Garbage can - Stylized" by Uricaro97.

Our style reference images are sourced from Civitai or directly generated using SD XL. Style Images sourced from Civitai are attributed as follows:

- BohoAI - konyconi
- Glass Sculptures
- IvoryGoldAI - konyconi
- Glass mouse
- Woodenmade
- Doctor Diffusion's Abstractor
- style-of-marc-allante
- Ice cream
- Pixel Particles Style [SDXL]
- Ink woman
- Opal Style [LoRA 1.5+SDXL]
- Moss Beasts
- Anime Lineart / Manga-like Style
- GENTLECAT style
- XL Realistic gold carving art style
- Glacial Ice Style [SD1.5]
- style ceshi
- style of Milton Glaser [SDXL] 368
- style of Raymond Duchamp-Villon [SDXL] 133
- Style-Darkestdungeon
- Ice Style XL
- zyd232's Ink Style
- Pastel color
- XL Realistic silver carving art style
- Necronomicon Pages
- (SDXL)chinese style illustration
- Gelato Style
- niji - geometric_shapes
- Schematics
- InkPunk XL

B IMPLEMENTATION DETAILS

B.1 Training Details

Our texture generation pipeline is developed in Threestudio [Guo et al. 2023] with Stable Diffusion 1.5 [Rombach et al. 2022]. Our evaluation dataset includes 100 3D models from Objaverse [Deitke et al. 2023] and Sketchfab [Sketchfab [n. d.]] (see details in Sec. A). The stylistic images for our experiments are derived from the internet or generated by diffusion models (see details in Sec. A). The content text prompts y_{ref} for these style images are obtained via GPT-4 [Achiam et al. 2023].

We optimize the texture field for 2500 iterations using an Adam optimizer with a leaning rate of 0.005. During the optimization phase, we employ the pre-trained depth and normal ControlNet [Zhang et al. 2023] to ensure the alignment of the texture details with the geometry of the input mesh. Both the depth map and normal map are rendered in camera space and subsequently normalized to adhere to ScanNet's standards [Dai et al. 2017]. In the main paper, the hyperparameters λ_{cfg} in Eq. 7 and λ_{style} in Eq. 8 are both set as 7.5.

B.2 Texture Map Extraction

After obtaining the optimized texture field, we employ a post-processing procedure to ensure the storability, editability, and applicability of the textures across various rendering platforms by transforming the texture field into a texture map with a resolution of 1024^2. Specifically, similar to [Chen et al. 2023; Munkberg et al. 2022], we sample the texture field using either the model's inherent UV map or one automatically generated by xatlas [Young 2021]. Furthermore, we apply the UV edge padding technique to fill in the empty regions between UV islands, effectively eliminating unwanted seams.

B.3 Quantitative Evaluation Matrix

The quantitative metrics used in our paper are derived from two aspects: alignment with the style of the reference image, and alignment with the text prompts.

Gram Matrix Distance. Drawing from traditional 2D style transfer methods [Gatys et al. 2015, 2016; Johnson et al. 2016], the squared Frobenius norm of the difference between the Gram matrices of the reference image and the rendered views of the generated textures can be employed to quantify the stylistic divergence:

$$D_{GM}^{j} = ||G_j^{\phi}(I_{ref}) - G_j^{\phi}(I_{render})||_F^2, \tag{1}$$

$$G_j^{\phi}(I)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(I)_{h,w,c} \phi_j(I)_{h,w,c'}, \quad (2)$$

where $\phi_j(I)$ is the activations at the jth layer of the VGG network ϕ for the input image I, which is a feature map of shape $C_j \times H_j \times W_j$.

CLIP Score. CLIP Score [Hessel et al. 2021] is a metric that quantifies the semantic similarity between images and texts. For a rendered view with visual CLIP embedding $\mathbf{v}$ and a given text prompt with textual CLIP embedding $\mathbf{c}$, we set $w = 2.5$ and compute CLIP Score as:

$$CLIP_s(\mathbf{c}, \mathbf{v}) = w * max(cos(\mathbf{c}, \mathbf{v}), 0). \quad (3)$$

B.4 Baseline Implementation Details

In our implementation of TEXTure [Richardson et al. 2023], we adhere to its texture-from-image methodology. As TextureDreamer's [Yeh et al. 2024] source code is not publicly available, we reproduce their method using threestudio [Guo et al. 2023]. Due to the absence of specific training details for DreamBooth [Ruiz et al. 2023] in their publication, we utilize the code and default parameters from the Diffusers library to train DreamBooth with LoRA using a single reference image. IPDreamer [Zeng et al. 2023] is a two-stage 3D generation method, with the first stage optimizing geometry and the second stage optimizing appearance. We skip the first stage and feed the input mesh directly to the second stage to optimize the surface color. SyncDreamer [Liu et al. 2023] is a method that synthesizes multi-view consistent images based on a given mesh, making it compatible with any 2D image-guided method during the denoising process. Consequently, we employ Instant Style [Wang et al. 2024] to infuse the style of the reference image.

C EFFECT OF GUIDANCE SCALE

In this section, we conduct an investigation into the impact of two hyperparameters: the CFG scale λ_{cfg} and the style guidance scale λ_{style}. As illustrated in Fig. 1, we visualize the influence of both λ_{cfg} and λ_{style}. Our observations indicate that an increase in λ_{style} can effectively enhance detail and style guidance. However, if λ_{style} becomes excessively high and λ_{cfg} is unable to match it, the content text prompt, which serves as a negative prompt in the CFG term, may fail to perform its role adequately, leading to content leakage issues. It is worth noting that our optimal values for λ_{cfg} and λ_{style} are suitable for all objects and require no further modification during inference.

REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

Rui Chen, Yongwei Chen, Ningxin Jiao, and Kui Jia. 2023. Fantasia3D: Disentangling Geometry and Appearance for High-quality Text-to-3D Content Creation. In *2023 IEEE/CVF International Conference on Computer Vision, ICCV 2023*. IEEE, 22189–22199.

Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas A. Funkhouser, and Matthias Nießner. 2017. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE Computer Society, 2432–2443.

Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. 2023. Objaverse: A Universe of Annotated 3D Objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023*. IEEE, 13142–13153.

LA Gatys, AS Ecker, and M Bethge. 2015. Texture Synthesis Using Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*. 262–270.

Leon Gatys, Alexander Ecker, and Matthias Bethge. 2016. A Neural Algorithm of Artistic Style. *Journal of Vision* 16, 12 (2016), 326–326.

Yuan-Chen Guo, Ying-Tian Liu, Ruizhi Shao, Christian Laforte, Vikram Voleti, Guan Luo, Chia-Hao Chen, Zi-Xin Zou, Chen Wang, Yan-Pei Cao, and Song-Hai Zhang. 2023. threestudio: A unified framework for 3D content generation. https://github.com/threestudio-project/threestudio.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. 2021. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021*. 7514–7528.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision - ECCV 2016 - 14th European Conference (Lecture Notes in Computer Science, Vol. 9906)*. 694–711.

Yuan Liu, Cheng Lin, Zijiao Zeng, Xiaoxiao Long, Lingjie Liu, Taku Komura, and Wenping Wang. 2023. SyncDreamer: Learning to Generate Multiview-consistent Images from a Single-view Image. *arXiv preprint arXiv:2309.03453* (2023).

Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*. 8270–8280.

Elad Richardson, Gal Metzer, Yuval Alaluf, Raja Giryes, and Daniel Cohen-Or. 2023. Texture: Text-guided texturing of 3d shapes. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 54, 11 pages. https://doi.org/10.1145/3588432.3591503

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*. IEEE, 10674–10685.

Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023*. 22500–22510.

Sketchfab. [n. d.]. Sketchfab - The best 3D viewer on the web. https://www.sketchfab.com

Haofan Wang, Qixun Wang, Xu Bai, Zekui Qin, and Anthony Chen. 2024. InstantStyle: Free Lunch towards Style-Preserving in Text-to-Image Generation. *arXiv preprint arXiv:2404.02733* (2024).

Yu-Ying Yeh, Jia-Bin Huang, Changil Kim, Lei Xiao, Thu Nguyen-Phuoc, Numair Khan, Cheng Zhang, Manmohan Chandraker, Carl S Marshall, Zhao Dong, et al. 2024. TextureDreamer: Image-guided Texture Synthesis through Geometry-aware Diffusion. *arXiv preprint arXiv:2401.09416* (2024).

Jonathan Young. 2021. Jpcy/Xatlas. https://github.com/jpcy/xatlas.git

Bohan Zeng, Shanglin Li, Yutang Feng, Hong Li, Sicheng Gao, Jiaming Liu, Huaxia Li, Xu Tang, Jianzhuang Liu, and Baochang Zhang. 2023. Ipdreamer: Appearance-controllable 3d object generation with image prompts. *arXiv preprint arXiv:2310.05375* (2023).

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023*. IEEE, 3813–3824.

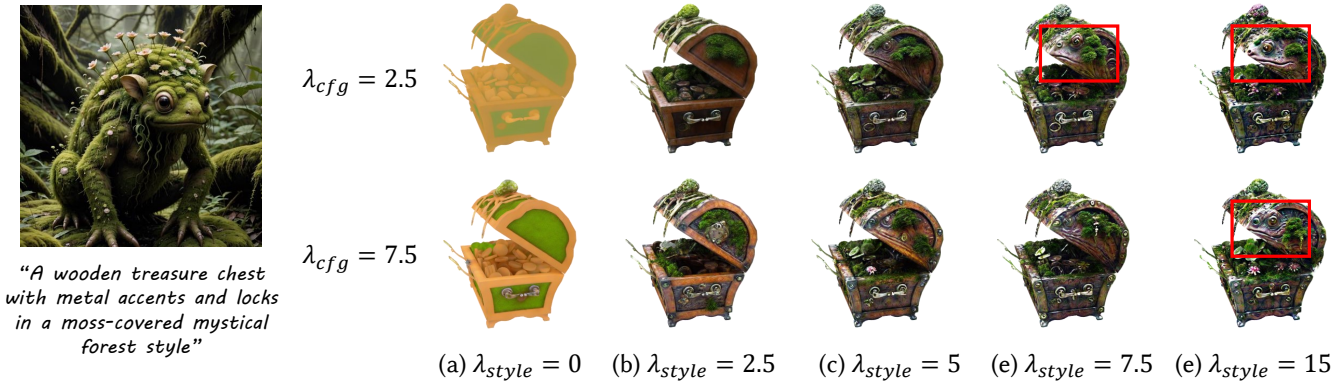

Fig. 1. Stylized texture generation with different λ_{cfg} and λ_{style}.