PAniC-3D: Stylized Single-view 3D Reconstruction from Portraits of Anime Characters

Shuhong Chen*, Kevin Zhang*, Yichun Shi^, Heng Wang^, Yiheng Zhu^, Guoxian Song^, Sizhe An^, Janus Kristjansson*, Xiao Yang^, Matthias Zwicker*
*University of Maryland - College Park, ^Bytedance/TikTok

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

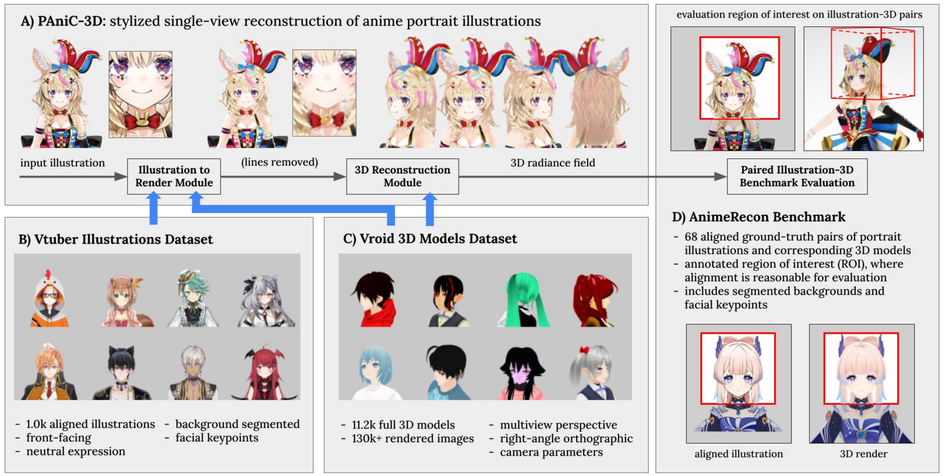
Challenges

(A1) *Character portrait illustrations have hair/accessories with complex geometry, and* **(A2)** *are shaded with NPR contour lines*

(B+C) *Lack of available 3D models and 2D portrait illustrations in this domain*

(D) *Stylized reconstruction is an ambiguous problem that is difficult to evaluate*

Contributions



Abstract

We propose **PAniC-3D**, a system to reconstruct stylized 3D character heads directly from illustrated **(p)ortraits of (ani)me (c)haracters**. Our anime-style domain poses unique challenges to single-view reconstruction compared to natural images of human heads, and there is a lack of data suitable to train and evaluate this ambiguous stylized reconstruction task. Our proposed **(A)** PAniC-3D architecture crosses the illustration-to-3D domain gap with a line-filling model, and represents sophisticated geometries with a volumetric radiance field. We train our system with **(B+C)** two large new 2D and 3D datasets, and evaluate on **(D)** a novel AnimeRecon benchmark of illustration-to-3D pairs. PAniC-3D **(E)** significantly outperforms baseline methods, and provides data to establish the task of stylized reconstruction from portrait illustrations.

Code + data available at:
`github.com/ShuhongChen/panic3d-anime-reconstruction`

Contributions A
- Omaru Polka, Hololive, https://virtualyoutuber.fandom.com/wiki/Omaru_Polka

Contributions B, left-to-right, top-to-bottom
- Kazami Kuku, 774 inc., https://virtualyoutuber.fandom.com/wiki/Kazami_Kuku
- Ayunda Risu, Hololive, https://virtualyoutuber.fandom.com/wiki/Ayunda_Risu
- Kagami Kira, Hololive, https://virtualyoutuber.fandom.com/wiki/Kagami_Kira
- Vestia Zeta, Hololive, https://virtualyoutuber.fandom.com/wiki/Vestia_Zeta
- Shibuya HAL, Neo-Porte, https://virtualyoutuber.fandom.com/wiki/Shibuya_HAL
- Era, Nijisanji, https://virtualyoutuber.fandom.com/wiki/Era
- Ibrahim, Nijisanji, https://virtualyoutuber.fandom.com/wiki/Ibrahim
- Araka Luto, PRISM Project, https://virtualyoutuber.fandom.com/wiki/Araka_Luto

Contributions C, left-to-right, top-to-bottom
- Red Hoodie, by Ice Guardian, https://hub.vroid.com/en/characters/8651763155335600083/models/2050558999678266090
- 1号, by せいじ, https://hub.vroid.com/en/characters/5514349705731891231/models/4248196668590370400
- Hatsune Miku Magical Mirai 2016, by Liza-chan, https://hub.vroid.com/en/characters/5938496229132058825/models/5213375051399188660
- Fyre Skar, by Fyre Skar, https://hub.vroid.com/en/characters/8075507551952010391/models/8919846134426001840
- PG-chan (PAC Global-chan), by a certain member, https://hub.vroid.com/en/characters/3125779151067201179/models/2022267982343604441
- はやと, by ユニ, https://hub.vroid.com/en/characters/1877189575628873831/models/6797828888939495741
- Test 8, by KseKeshka, https://hub.vroid.com/en/characters/5079994044987193276/models/6534563580070628532
- sia17-1, by J_Sergeant, https://hub.vroid.com/en/characters/4019600350590104152/models/4665808090417218631

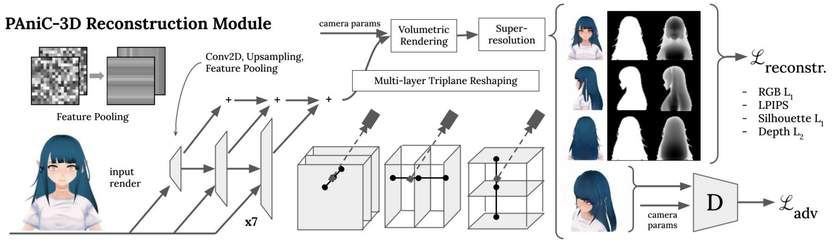
Contributions D
- illustration: Omaru Polka, Hololive, https://virtualyoutuber.fandom.com/wiki/Omaru_Polka
- 3d model: Omaru Polka, Hololive, https://3d.nicovideo.jp/works/td78507
- illustration: Sangonomiya Kokomi, Genshin Impact, https://genshin-impact.fandom.com/wiki/Sangonomiya_Kokomi
- 3d model: Sangonomiya Kokomi, Genshin Impact, https://activity.hdslb.com/blackboard/static/20220823/c84ef0977c17fb1198f6887261fea35f/SBvSX0RxRme.zip

Figure A1
- kawaii, by Atmada Art, https://hub.vroid.com/en/characters/5784086077974039059/models/2298748470490203300

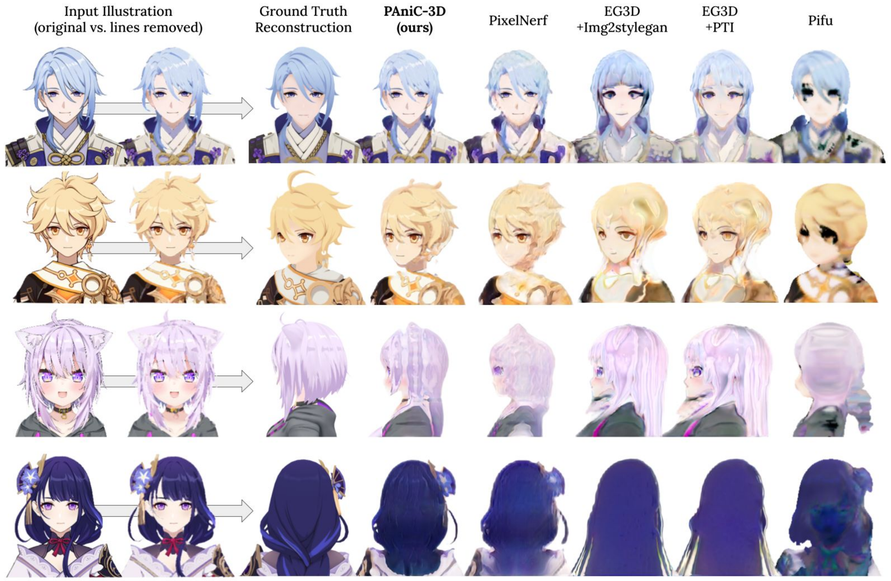
Figure A2
- Yozora Mel, Hololive, https://virtualyoutuber.fandom.com/wiki/Yozora_Mel

Figure E, left-to-right, top-to-bottom
- illustration: Kamisato Ayato, Genshin Impact, https://genshin-impact.fandom.com/wiki/Kamisato_Ayato
- 3d model: Kamisato Ayato, Genshin Impact, https://activity.hdslb.com/blackboard/static/20220328/279f2a710d12c63fc8bc0aeb4e1a3556/nAguYew4kB.zip
- illustration: Aether, Genshin Impact, https://genshin-impact.fandom.com/wiki/Traveler
- 3d model: Aether, Genshin Impact, http://static.biligame.com/caster_custom_assets/ys/gczj/kong.rar
- illustration: Nekomata Okayu, Hololive, https://virtualyoutuber.fandom.com/wiki/Nekomata_Okayu
- 3d model: Nekomata Okayu, Hololive, https://3d.nicovideo.jp/works/td63648
- illustration: Raiden Shogun, Genshin Impact, https://genshin-impact.fandom.com/wiki/Raiden_Shogun
- 3d model: Raiden Shogun, Genshin Impact, https://activity.hdslb.com/blackboard/static/20210826/279f2a710d12c63fc8bc0aeb4e1a3556/GA2FmFGUE0.zip

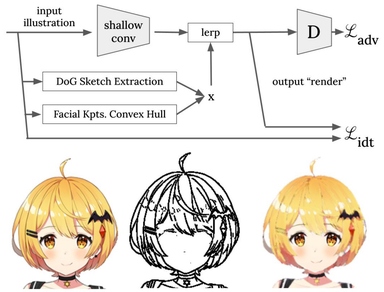
PAniC-3D Reconstruction Module



(A1) 3D reconstruction module. The front-orthographic input rendering is fed to a series of upsampling convolutions, with intermediate feature pooling to help distribute information along common triplane axes. The final feature stack is reshaped into a multi-layer triplane, which is volumetrically rendered and super-resolutioned to the final output. Reconstruction losses are applied to front, left, right, and back views (right view omitted in figure), and adversarial loss is applied to a random perspective view.



PAniC-3D Illustration-to-Render Module



(A2) Illustration-to-render module. We design a simple yet effective network to cross the domain gap by removing illustration contour lines absent in a diffuse 3D render, while retaining lines present around facial features like the eyes and mouth.

(E) Comparison to baselines. Illustrations with lines removed by our illustration-to-render module are fed to various reconstruction frameworks; our PAniC-3D system delivers plausible reconstructions, while other methods struggle to preserve identity and predict reasonable geometry. Note that metrics between the displayed ground-truth and prediction are restricted to an ROI bounding box/rectangular prism (unshown) during evaluation. Please refer to our paper for quantitative comparison to other methods, ablations of model components, and additional results.

Yu, A., Ye, V., Tancik, M., & Kanazawa, A. (2020). pixelnerf: Neural radiance fields from one or few images. In 2021 IEEE. In CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4576-4585).
Chan, E. R., Lin, C. Z., Chan, M. A., Nagano, K., Pan, B., De Mello, S., ... & Wetzstein, G. (2022). Efficient geometry-aware 3D generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16123-16133).
Abdal, R., Qin, Y., & Wonka, P. (2019). Image2stylegan: How to embed images into the stylegan latent space?. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4432-4441).
Roich, D., Mokady, R., Bermano, A. H., & Cohen-Or, D. (2022). Pivotal tuning for latent-based editing of real images. ACM Transactions on Graphics (TOG), 42(1), 1-13.
Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2304-2314).