



# Single View to Stereo Image Synthesis

Progress Report 3  
December 15, 2025,  
Tarik Eker, Niklas Munkes, Kaan Yücel, Julian Jurcevic, Ronny Göttler



## New Dataset (1/4)



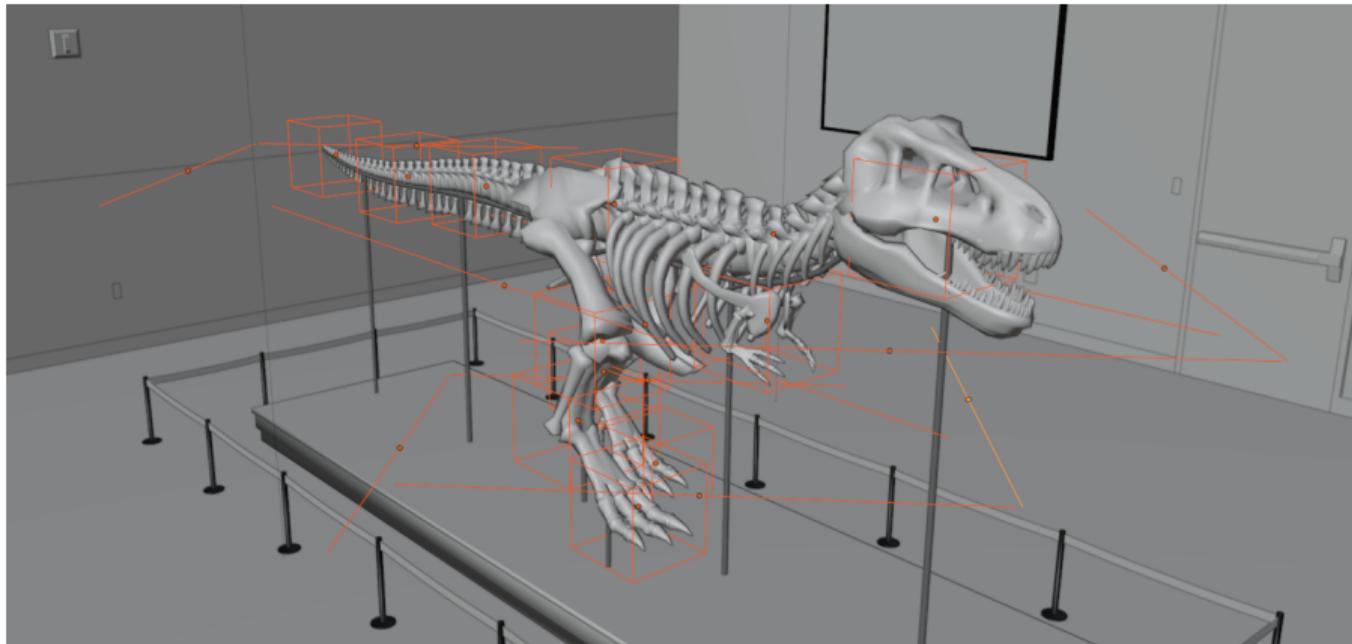


## New Dataset (2/4)



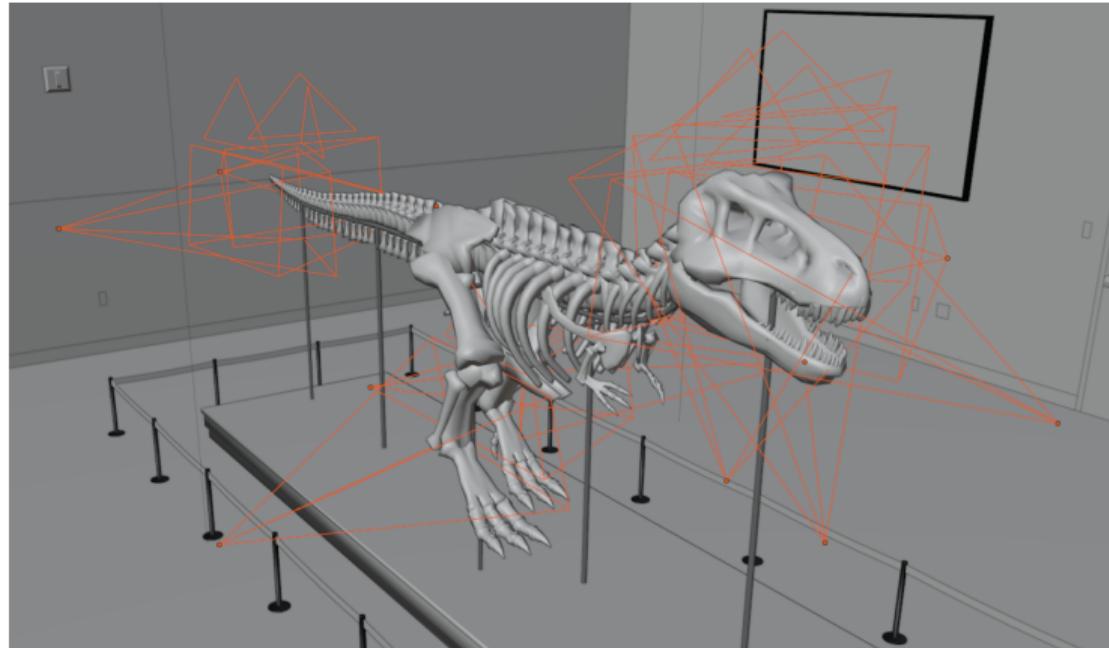


## New Dataset (3/4)





## New Dataset (4/4)



# New Architecture: StereoDiffusion [4]

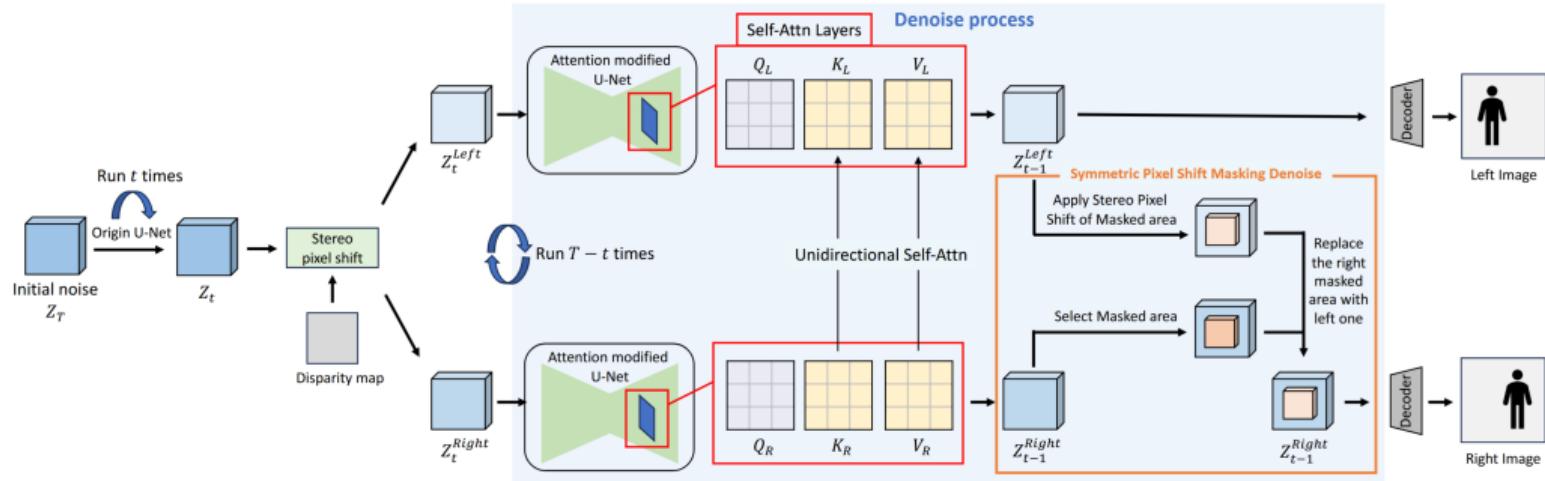


Figure: From [4]

---

## StereoDiffusion: General Characteristics

- Training-free: utilizes pretrained DPT [1] and StableDiffusion [2] without fine-tuning
- TODO

---

# StereoDiffusion: Stereo Pixels Shift

**Theirs**

TODO

**Ours**

TODO



---

## Qualitative Analysis: Input and Gold Output

TODO



## Qualitative Analysis: Their Result



**Figure:** Stereo images generated with StereoDiffusion + disparity map from DPT



## Qualitative Analysis: Our Result



**Figure:** Stereo images generated with StereoDiffusion + depth map from DPT + sensor data



---

## Qualitative Analysis: Our Result with Modified Baseline Distance

TODO

---

## Qualitative Analysis: Their Result Deblurred

TODO

---

## Possible Improvements

- switch to a better/state-of-the-art monocular depth estimation model (see for example [3, 5])
- TODO



## New Dataset: Credits

- "Tyrannosaurus Rex Fossil Museum" (<https://skfb.ly/pBEzK>) by Chenchanchong is licensed under Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>).
- "Frazer Nash Le Mans" (<https://skfb.ly/pDP6y>) by Kryox Shade is licensed under Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>).
- ground texture from "The Mardou museum" (<https://skfb.ly/pqxlz>) by OuterspaceSoftware is licensed under Creative Commons Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/4.0/>).
- "Discovery Hall" (<https://skfb.ly/6UQZ8>) by nickcramer is licensed under CC Attribution-NonCommercial-NoDerivs (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).
- "Grand Piano" (<https://skfb.ly/6UUUnK>) by Amatsukast is licensed under CC Attribution-NonCommercial-ShareAlike (<http://creativecommons.org/licenses/by-nc-sa/4.0/>).
- "Steam Train (Valley Railroad #97)" (<https://skfb.ly/6TPKL>) by Kefla is licensed under CC Attribution-NonCommercial-NoDerivs (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

- [1] Ranftl, René, Alexey Bochkovskiy & Vladlen Koltun (2021). *Vision Transformers for Dense Prediction*. <https://arxiv.org/abs/2103.13413>.
- [2] Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser & Björn Ommer (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*.  
<https://arxiv.org/abs/2112.10752>.
- [3] Sun, Boyuan, Modi Jin, Bowen Yin & Qibin Hou (2025). *Depth Anything at Any Condition*.  
<https://arxiv.org/abs/2507.01634>.
- [4] Wang, Lezhong, Jeppe Revall Frisvad, Mark Bo Jensen & Siavash Arjomand Bigdeli. 2024. StereoDiffusion: Training-Free Stereo Image Generation Using Latent Diffusion Models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 7416–7425. DOI: [10.1109/cvprw63382.2024.00737](https://doi.org/10.1109/cvprw63382.2024.00737).
- [5] Zhang, Jiuling (2025). *Survey on Monocular Metric Depth Estimation*.  
<https://arxiv.org/abs/2501.11841>.