Image2Minecraft

Andrew Teeter Gregory Smelkov Matthew Clinton

March 26, 2021

Project description. The goal of Image2Minecraft is to create a tool to facilitate the conversion of 2D images to a 3D representation within the voxel based sandbox game. To achieve this we will implement existing architectures to first generate a 3D model from an image and then create a world in Minecraft featuring the model constructed out of the available voxels in the game.

Due to the massive computational resource and time requirement for training and the number of good solutions to the problem of generating 3D models from a single 2D image, we plan to use use an existing trained architecture that allows us to focus our efforts on recreating that model in Minecraft. We will train a convolutional neural network to choose better blocks for the in-game model to ensure Image2Minecraft can more accurately recreate the image.

Related work. We plan on drawing heavily from the implementation of Zubi'c and Lio [4] because the implementation presented in the paper has promising results in generating 3D models from images without rendering. We also may draw from Chen et. al. [2] for an alternative approach to generating 3D models from images. If we have enough time, we plan to compare our results from using these different architectures.

Datasets and evaluation. We plan on using the ShapeNet [1] dataset for original tests and then expand our project to include the MSR-Object3D-300[3] dataset both of which can be created as point clouds.

To measure how accurately Image2Minecraft recreates the image in Minecraft, we plan to render both the original synthesized object and its voxel representation from the same view - then use the difference between their renders as a means to optimize our voxel representation with respect to the original

model. This is preferable, as opposed to mapping blocks to the color space, due to the intricate nature of different block materials.

Split of the work.

Matthew will focus on implementing the image to 3D object.

Andrew will focus on implementing the 3D object to voxel network.

Greg will focus on implementing the voxel to Minecraft.

We will work together on fine-tuning our tool.

References

[1] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

[2] W. Chen, Jun Gao, Huan Ling, Edward Smith, J. Lehtinen, A. Jacobson, and S. Fidler. Learning to predict 3d objects with an interpolation-based differentiable renderer. In *NeurIPS*, 2019.

[3] Qiang Hao, Rui Cai, Z. Li, Lei Zhang, Yanwei Pang, Feng Wu, and Y. Rui. Efficient 2d-to-3d correspondence filtering for scalable 3d object recognition. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 899–906, 2013.

[4] Nikola Zubi'c and P. Lió. An effective loss function for generating 3d models from single 2d image without rendering. *ArXiv*, abs/2103.03390, 2021.