3D AttnGAN for Text to Voxel Generation

Yoon, Seungwoo (School: Chung Nam Samsung Academy)

Text to image generation, that is about understanding the text descriptions and generating images from them, has been one of the most important machine learning research fields due to its high utility on diverse applications in industrial fields and academic research. This research aims to expand text to image generation field to text to voxel generation field, which is about generating 3D objects from text descriptions, since it has higher potential in VR, metaverse, 3D printing, real-world industries, etc. To implement text to voxel generating artificial intelligence, this research proposes 3D Attentional Generative Adversarial Networks (3D AttnGAN), which is a modification of Attentional Generative Adversarial Networks (AttnGAN) originally used in text to image generation. In 3D AttnGAN, 2D Convolutional Neural Network (CNN) of AttnGAN is replaced with 3D CNN. This change causes the memory size problem that means enlarged memory size occupied by the model. To solve the memory size problem, 3D AttnGAN uses stage freeze technique, training each stages sequentially. Bottleneck layers were installed in front and end of the CNN in the residual blocks of stage 2 and stage 3, decreasing the memory size while the performance doesn't decrease largely. The experiment conducted on captioned ShapeNet dataset qualitatively showed that 3D AttnGAN can create proper 3D objects following the given text descriptions and quantitatively proved the performance enhancement caused by attention with DAMSM loss 28.7% decreased. With 3D AttnGAN, this research established the foundation of Text to Voxel Generation research field.

Awards Won: