

Leveraging U-Transformer Outpainting for Image Interpolation

Traditionally, image interpolation or inpainting tries to fill in a removed section of an image. There are a variety of methodologies for image interpolation, but our hope is to attempt a new approach where we can interpolate a scene between two disparate images by using image outpainting or image extrapolation. Outpainting or image extrapolation is the process of taking an image and generating an extension of the image beyond the original border. In general, outpainting is a difficult task because it requires the model to expand beyond a known region to generate scenes that look real. Our idea is to take two images and see if we can train a GAN-like architecture to outpaint these two images towards one another to complete a scene. We will be basing this project on the work done in [Generalised Image Outpainting with U-Transformer by Gao et. al \(2022\)](#), which leveraged a transformer-based GAN they call a U-transformer. In the past, image outpainting only focused in the horizontal direction. However, in this paper, Gao et al, were able to outpaint horizontally and vertically. We hope to use this architecture to train a new U-Transformer that can interpolate between images. This is different from interpolation in the past which was able to use the context of the surrounding image to fill in a removed section. Instead, our focus is on seeing if the model can understand how to connect different scenes in a realistic way. If successful, it means that the model understands relationships between disparate images and can model the visual world in a more precise way.

The original data used by Gao et al. in training the U-transformer is the NS-Outpainting data set which is available on [GitHub](#). We can leverage this data set and other scene-based datasets like [ADE20K](#) or [SEL](#) and crop images to our specific task. The original paper's implementation of the U-Transformer was in PyTorch and is available on [GitHub](#). We will also implement our new model using PyTorch and leverage the U-Transformer. However, we will alter the original paper's architecture significantly to attempt image interpolation using outpainting instead of just outpainting images outright. To evaluate the results, we can compare the generated interpolations to actual scenes and use an image similarity metric like the structural similarity index or other image analysis metrics like peak signal-to-noise ratio. We can also look specifically at the smoothness of where the outpaintings meet. In addition, because this is a GAN based architecture, we should use metrics like inception score and Fréchet inception distance to measure the performance of the GANs. All of these measures were used in Gao et al. (2022). Lastly, there will probably need to be some manual inspection of model quality. Hopefully a good discriminator will be able to assess the realistic-ness of an image.

