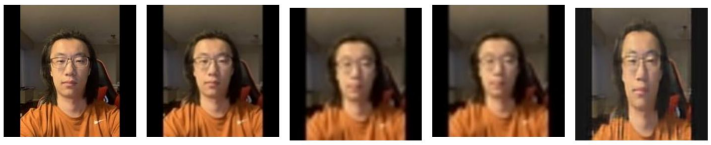
Smooth compression and reconstruction of human portrait image using convolutional auto-encoder

Link to Presentation: https://youtu.be/HBlx5rpkG_8

Sheng Tan, shengtan@stanford.edu

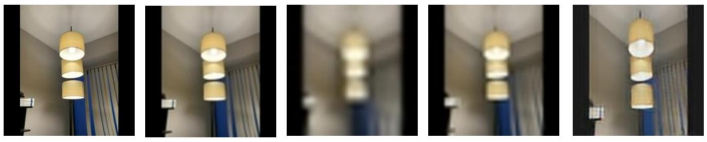
This project explores the possibility of using a convolutional auto-encoder to encode higher resolution images to data stream and reconstruct the image. A 192 x 192 pixels image is passed into the neural network, encoded to lower latent space and then decoded to its original size. The reconstructed images can not out perform the results produced by GANs, but the structure provides a lossy compression algorithm to human portrait images that could allow fast data transferring as well as a feature extraction network for other trainnings.

Performance on training images from left to right: original, 2x down sampling, 4x, 8x, auto-encoder result



Performance on test images(photo of author) from left to right: original, 2x down sampling, 4x, 8x, auto-encoder result

The result of the encoding and decoding process certainly outperforms the naive max-pooling/bicubic interpolation mechanism with equivalent or even lower compression rate. The mean squared error between the original image and the reconstructed image is 3.1e-4, which is ±4.5 per pixel on average on [0,255] scale, whereas the mean squared error of 2x, 4x, and 8x down sampling are respectively 1.9e-4, 4.5e-4, 9.3e-4.

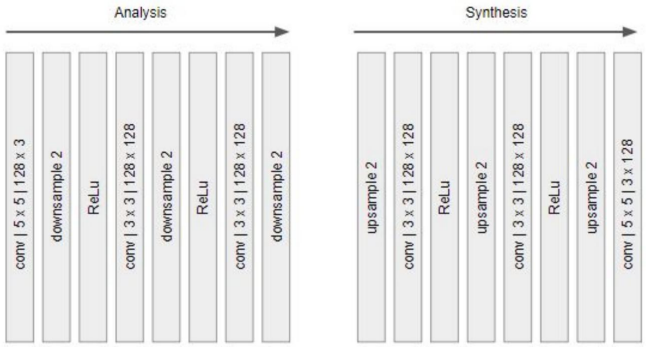


Performance on test images(photo of lamp) from left to right: original, 2x down sampling, 4x, 8x, auto-encoder result

Surprisingly, even though the network was never trained on any non human portrait images, it still performs decently enough. The filters could be caputering more general features than only the ones specific to human beings, which could be studied further

Sample training images from CelebA data set, 178 x 218, RGB images. It is a open source image set that contains 200,000+ human portrait images

Since this project focuses on unsupervised learning, we did not need to label any data or perform any feature extration: images are simply resized to 192x192 and normalized to [0,1]



An autoencoder is built for this project, with ReLu instead of GND/IGDN as its activation function and using 5x5 and 3x3 filters instead of 9x9 and 5x5 filters for higher resolution images [1]

Future work on this project could be trainning a even deeper autoencoder and further compress the input data. The inner most layer could also be used as inputs to GANs for image reconstruction, which would be a promising next step to take as well

[1] Ballé, Johannes, Valero Laparra, and Eero P. Simoncelli. "End-to-end optimized image compression." arXiv preprint arXiv:1611.01704 (2016).