Image and Video Super-resolution with GANs

Allan Raventos, Jason Liu, and Max Spero
Stanford University

Introduction

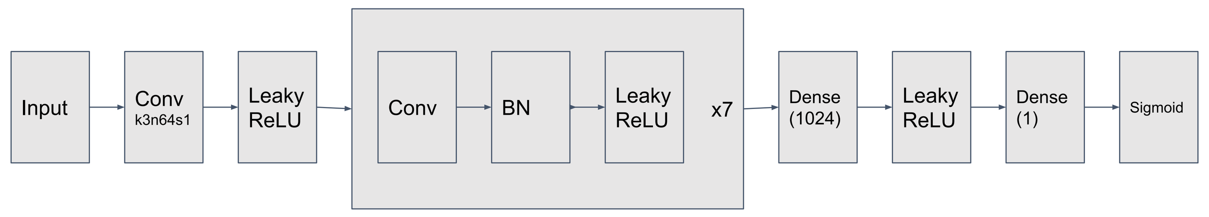
Image downscaling is an innately lossy process. No matter how good an upscaling algorithm is, there will always be some amount of high-frequency data lost from a downscale-upscale function performed on an image. Ultimately, even the best upscaling algorithms cannot effectively reconstruct data that does not exist. We propose a fix to certain situations where that problem appears, by using GANs to "hallucinate" high-frequency data in a super-resolved image that does not exist in the smaller image. Using this method, we claim no accurate reconstruction of lost data, but rather a plausible guess at what this lost data might be.

Objectives

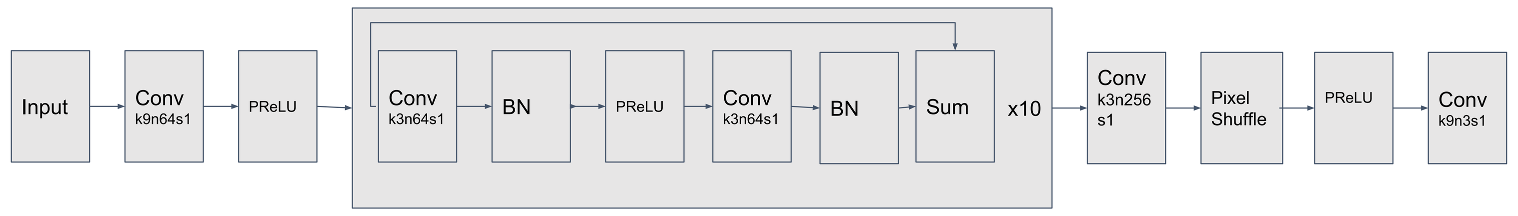
Our goal for this project is to first train a GAN that would perform super-resolution on images. Then, we wanted to extend our model with 3D convolutions, and compare the new and old models' performance on super-resolution for video. A 3-D convolution allows us to access not only spatial, but also temporal information, and we were interested in seeing what improvements this would yield.

Architecture

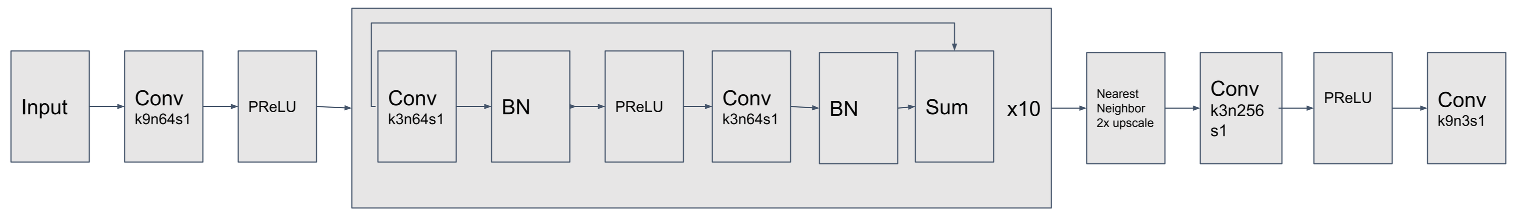
Discriminator: (every 2 conv layers, stride=2, double the number of filters)



Generator: (Original)



Generator: (Nearest neighbor/Conv layer)



Methods

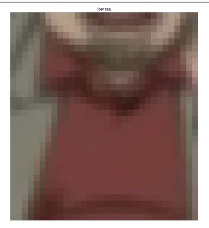
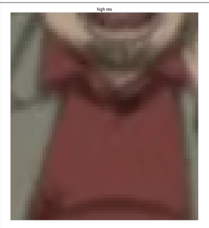
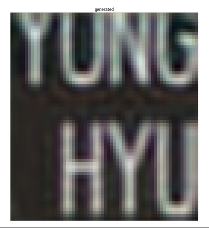
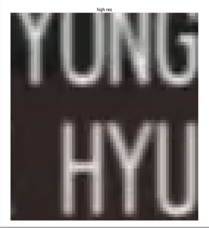
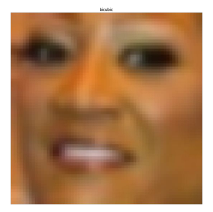
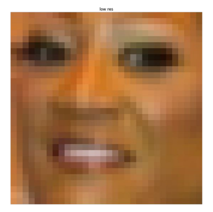
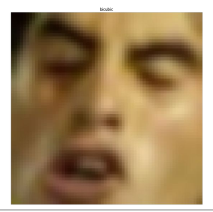
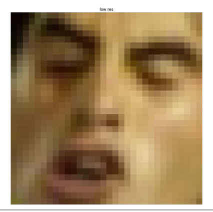
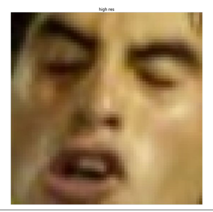
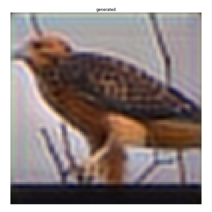
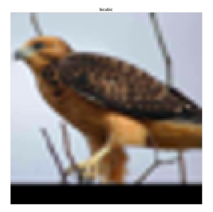
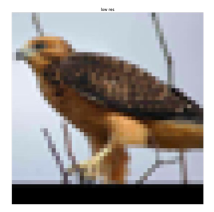
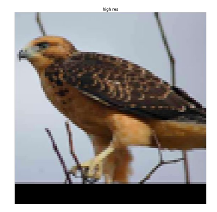
We implemented a variety of methods when training our GAN

1. Content Loss
2. Pixel-Shuffle Upscaling
3. Resize-Convolution Upscaling
4. Generator Pre-training

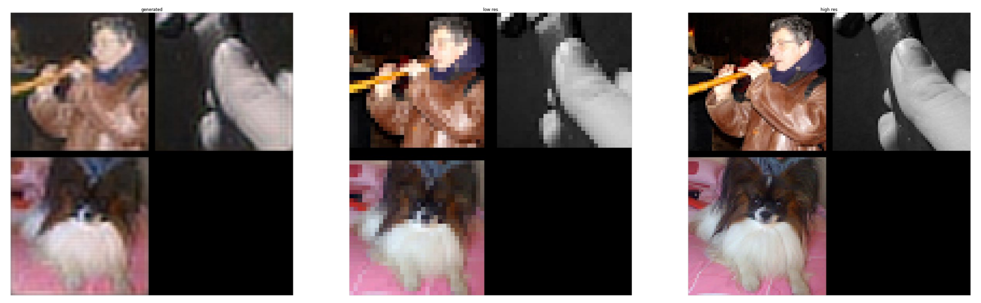
We also tested our GAN on a various datasets, yielding the results shown below.

1. Imagenet
2. STL-10
3. CelebA
4. Center crop of an anime episode

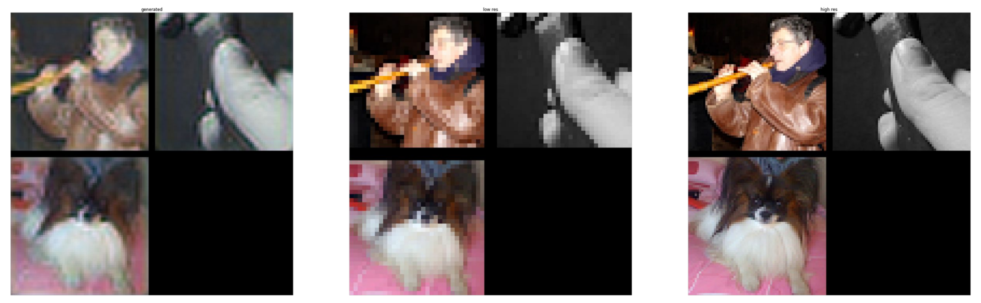
Results

	Upscaled with Generator	Upscaled with Bicubic	Original Low-res	True High-res
Lines are bolder in generated image				
Edges seem sharper but there is some color noise				
Still some blur but able to capture more of the hair texture				
Hallucinates eye shadow, which may be in other pictures				
Very close to original				
Has better wing texture than bicubic does				
Has trouble resolving background				

1. Pixel Shuffle (See grid artifacts)



2. Resize + Conv



Conclusions and Future Work

- Use of resize convolution mitigates "gridding" effect observed when pixel shuffling
- Performance of GAN seems to improve with pretraining of generator network (D has better fake examples to work with)
- Sharp edges are more difficult to produce when the dataset contains samples from a large number of classes

Future work

- Train GAN on fewer output classes, iteratively increasing to full ImageNet
- HR and SR accuracy-based learning schedules
- Compute content loss using other VGG-19 layers
- Extend networks to handle video by adding a time dimension to inputs and convolutional filters
- Evaluate performance of video GAN as a function of frame spacing

References

- Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." *arXiv preprint arXiv:1609.04802* (2016).
- Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.
- Odena, Augustus, Vincent Dumoulin, and Chris Olah. "Deconvolution and checkerboard artifacts." *Distill* 1.10 (2016): e3.
- Adam Coates, Honglak Lee, Andrew Y. Ng An Analysis of Single Layer Networks in Unsupervised Feature Learning AISTATS, 2011
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.
- Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. **Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.** *University of Massachusetts, Amherst, Technical Report 07-49*, October, 2007.
- Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- https://www.tensorflow.org
- Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." *arXiv preprint arXiv:1412.0767* (2014).