Improving SRGAN for Object Detection

Ivan Abraham, Wyatt McAllister, Justin Wasserman
Paris' Information Processing Systems (PIPS) Conference
University of Illinois at Urbana-Champaign

Abstract

In [1] the authors implemented a convolutional neural network which, when trained on sample images, is able to perform science-fiction like image *enhancement* as commonly shown in television. This network is termed *Generative Adversarial Network for Super Resolution (SRGAN)*. A very similar model in [2] has a more effective performance metric of object detection. This work shows that, for this metric, the SRGAN does better when multiple downsampling methods are used to generate the training images.

Background

SRGAN Architecture The SRGAN network architecture presented in [1], and adapted from the implementation in [3], is as follows..

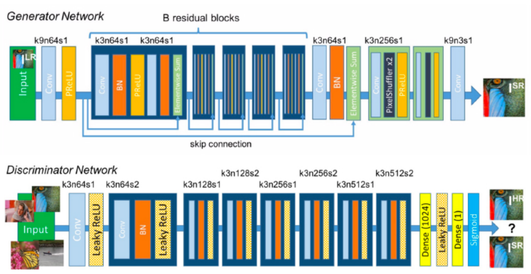

Figure 1: Architecture of Generator and Discriminator Networks with corresponding filter size (k), feature number (n), and filter time step (s) indicated for each convolutional layer.

Object Detection Metric For the object detection metric [4], the original SRGAN Network fails to detect several objects on many images, such as the one below.

Figure 2: Object Detection (OD): Original Image

If the network trains with the use of multiple downsampling schemes, it will succeed for images like this one.

Figure 3: OD (Multi-Blur)

Figure 4: OD (SRGAN)

Examples of Network Performance

As in the image below, our network training scheme, which we call Multi - Blur, reduces image artifacts in homogeneous image regions, at the expense of smoothing textures within objects.

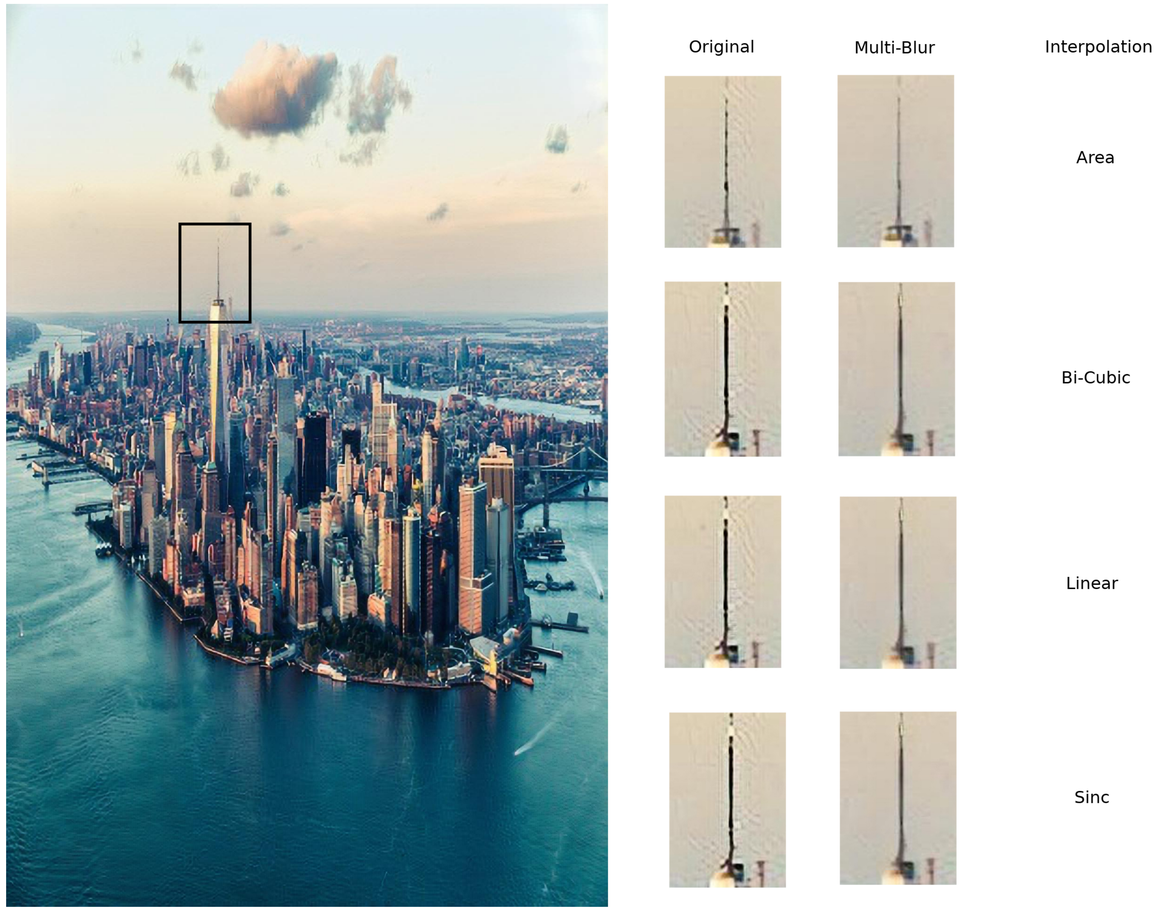

Figure 5: SRGAN Network Architecture

It also does well with fine grained features such as text. As shown below, the Multi-Blur training scheme produces readable text for this excerpt of Moby-Dick, while the original network fails to do so.

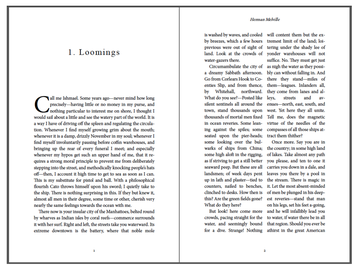

Figure 6: Original Image

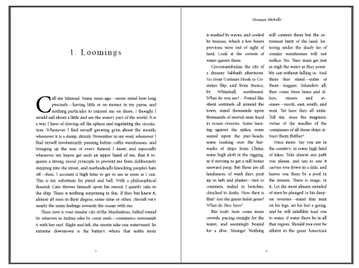

Figure 7: SR Image (Multi-Blur)

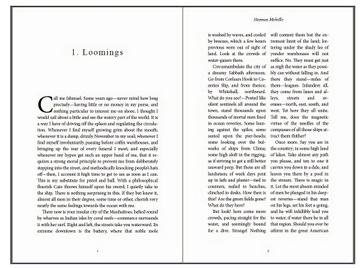

Figure 8: SR Image (SRGAN)

Results

This network training scheme, Multi-Blur, was tested on the VOC 2012 data set, and achieved superior performance in terms of object detection error, for all interpolation schemes tested.

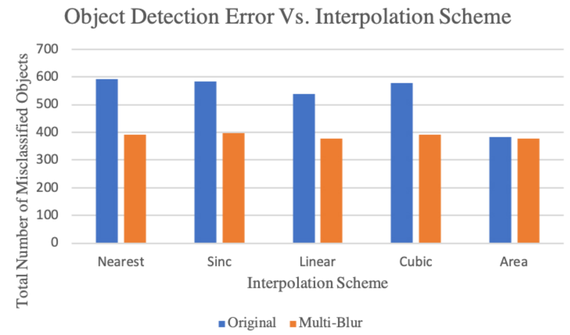

Figure 9: Object Detection Error vs. Interpolation Scheme

The reason for this performance improvement is hypothesized to be the fact that the learned weights for the generative network for the super-resolved images reflect all of the interpolation schemes, resulting in a smoothing out of the image textures. The hypothesis is that this smoothing results in a lower variation in the texture information within objects, improving object detection.

Conclusion

- + This training scheme improves the SRGAN network performance for object detection.
- + This comes at the expense of smoothing out image textures within each object.
- + While less photorealistic, the generated images have less image artifacts, and do well with images with regular features like text.

References

[1] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, page 4, 2017.

[2] Mehdi SM Sajjadi, Bernhard Schölkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. *arXiv preprint arXiv:1612.07919*, 2016.

[3] leftthomas on GitHub. Implementation of SRGAN. `https://github.com/leftthomas/SRGAN`, 2018. [Online; accessed 12-Nov-2018].

[4] ryoppippi and ayooshkathuria on GitHub. Implementation of the YOLO-V3 Object Detection Method. `https://github.com/ayooshkathuria/pytorch-yolo-v3`, 2018. [Online; accessed 12-Nov-2018].

Acknowledgement

Supported by PIPS.