CS766 – Project Proposal
Super Resolution Image Enhancement

Asher Elmquist (`amelmquist@wisc.edu`)
Eric Brandt (`elbrandt@wisc.edu`)

Due: February 14, 2020

1   Project Overview

The purpose of this project is to explore image super-resolution and understand how creating higher resolution or higher quality pictures can assist in downstream tasks such as object recognition or image segmentation. To understand the effect of super-resolution, we propose implementing a convolutional neural network for super-resolution based on the current state of the art. Beyond recreation of a current algorithm, we will study the general nature of the trained model, and explore the application of super-resolution in object detection accuracy and precision. This document covers the relevance of the project, a brief outline of the state of the art, a detailed description of the project plan, and a timeline that we plan to follow in order to accomplish the outlined tasks.

2   Project Relevance

Generating super-resolution images from low resolution images has been used in medical imaging [1], [2], astronomy imaging [3] and security imaging [4]. Where small or blurry objects need to be identified, a higher resolution image increase the performance of existing object recognition algorithms. If super-resolution techniques deliver on their promises efficiently, we can transparently substitute lower resolution or more highly compressed images where previously costly high resolution images were required. Image storage, network transmission and video encoding/decoding are several such examples. Super-resolution is an interesting task in and of itself, but we propose this project as a way to better understand and explore the potential for super-resolution networks to occupy one step in a pipeline for object detection or semantic segmentation tasks. The ability to locate objects potentially with sub-pixel precision in an image has interesting future applications for photogrammetry and metrology. To this end, our proposal centers on the implementation and application of super-resolution.

3   State of the Art

A recent and comprehensive overview of state of the art super-resolution algorithms and network structures can be found in [4]. A short discussion of key points as well as important notes will be given here as they relate to the project proposal. This discussion will cover generative adversarial networks (GANs) versus supervised learning for super resolution as well as a well-studied network structure that has been shown to work well in image enhancement architectures: ResNet [5].

The problem of super-resolution (SR) is in the non-uniqueness of a high-resolution (HR) image generated from a lower-resolution (LR) image. For any LR there are multiple plausible HR that would be faithfully represented by the LR. One way to generate data for SR is to down sample an HR image. Unfortunately, generating training samples from down sampling can lead to small artifacts in the network when trained to directly undo the down sampling algorithm [4]. One way of circumventing this is to use an unsupervised learning approach so as not to unwittingly compute a mapping from input-output samples, but instead to compute a mapping from a distribution of inputs to a distribution of outputs. Generative adversarial

networks accomplish exactly this and have been shown to work well for SR [6]. While GANs would improve SR on real-world low-resolution images, we can explore SR in a more efficient manner when we have more control of the datasets and training as is the case with the supervised approach. State of the art results are shown from a supervised method in [7].

In super-resolution, the network architecture plays an important role in the accuracy and efficiency of the network. Many image processing network architectures (including SR) are built on a well-studied convolutional network called ResNet. This network uses a recursive structure to learn small changes in the image. The network is made up of residual blocks with a fraction of the input added directly to the output of a later block. These connections are known as skip connections and are a fundamental component of the ResNet architecture [5] to reduce the complexity of the loss surface and reduce convergence to local minima. These residual blocks, along with up-scaling via deconvolution form the basis of many SR network architectures [4] and are detailed in the implementation given in [7].

4 Project Plan

The project will be split into four sections that will allow exploration beyond the state of the art. The first part of the project will be to recreate state of the art super-resolution results using machine learning. In itself this is a difficult task, but there is significant room to explore beyond what has been discussed in the previous section. To go further, we propose studying the generalization of the trained network by training and testing across different image domains. That is, we will train on a set of images, say buildings, and then evaluate the network on an animal image data set. In addition, we propose applying state-of-the-art, off-the-shelf object detection networks such as YOLO-V3 [8] and SSD [9] to understand if object recognition can be improved using super-resolution. Time-permitting, we will then extend to understand if other image improvement could help with object detection such as denoising or sharpening. Since similar networks for super-resolution can be modified for general image enhancement, this is a natural continuation of the project. Finally, we hope to pay attention to the time required to perform SR on images of particular sizes to evaluate the efficacy of real-time SR for live-acquisition applications on either general-purpose or dedicated hardware. Further details and data plans are laid out in the following sections.

4.1 Implement and Train Super-Resolution Network

Following the state of the art implementations, we will use a residual CNN as it has been shown to reduce error propagated across layers when given sufficient skip connections. To limit training time, we will avoid using generative adversarial networks. This will result in requiring the network to be trained through supervised learning, meaning we will require input and output image pairs. We do not foresee this being an issue as we can generate a dataset from a single set of high resolution images. This will also give us flexibility in the resolutions we choose to understand the extent to which resolution can be increased without unintended artifacts. For verification, our super-resolution can be compared against ground truth (original image) and bicubic or bilinear interpolation of the downsampled/corrupted image. We can also qualitatively compare our results to state-of-the-art results from the papers discussed previously.

4.2 Run Domain Transfer Study

We propose studying the general nature of our network by having three distinct datasets. These may be, for example, buildings, cars/roadways, and animals, each of which have a large corpora of readily available data. We will then hold out one dataset during training to use during testing to see if there are domain specific artifacts that appear on images with significant content differences. We will apply this holdout to each subset in turn to understand the full effects.

4.3 Apply Object Detection Networks

To understand if super-resolution improves object detection, YOLO-V3 and SSD will be used with pretrained weights. These networks have been shown to give state-of-the-art object detection accuracy. The accuracy of the pretrained networks will be run on the lower resolution images to generate a baseline accuracy for

comparison. The networks will then be run on the higher resolution (cropped when necessary to create one-to-one comparison) to understand if super-resolution improves detection rate, position-precision, or class confidence. Multiple data sets will also allow us to understand if there is significant difference when detecting difference object classes.

4.4 Further Extend To General Image Enhancements

If time permits, we can extend this project to look at other image enhancements for improving object detection. Since the residual CNNs have the same architecture as general image processing networks, the network proposed can be adjusted and trained to denoise, or sharpen images at the lower resolution. We would then perform the same studies discussed above on the resulting denoised or sharpened images.

5 Timeline

In order to meet the class deadlines and adhere to the project plan, the following timeline is proposed.

- **February 14th:** Project Proposal Due

- **March 13th:** Complete Super-Resolution Implementation

- **March 20th:** Complete Domain Transfer Study

- **March 25th:** Project Mid-Term Report due *Possible re-alignment of goals based on progress thus far*

- **April 3rd:** Complete Object Detection Study

- **April 17th:** Complete Additional Image Enhancement Study

- **April 27-May 1st:** Project Presentations

- **May 4th:** Project Webpage Due

References

[1] C.-H. Pham, C. Tor-Díez, H. Meunier, N. Bednarek, R. Fablet, N. Passat, F. Rousseau, and H. ene Meunier, "Multiscale brain MRI super-resolution using deep 3D convolutional networks Multiscale brain MRI super-resolution using deep 3D convolutional," *Computerized Medical Imaging and Graphics*, vol. 77, 2019. DOI: 10.1016/j.compmedimag.2019.101647. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01635455v2.

[2] M.-I. Georgescu, R. T. Ionescu, and N. Verga, "Convolutional neural networks with intermediate loss for 3d super-resolution of ct and mri scans," *ArXiv*, vol. abs/2001.01330, 2020.

[3] H. Zhang, P. Wang, C. Zhang, and Z. Jiang, "A comparable study of cnn-based single image super-resolution for space-based imaging sensors," in *Sensors (Basel)*, vol. 19, MDPI, pp. 32–34.

[4] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[6] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[7] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, Springer, 2016, pp. 21–37.