Residual Dense Network Super-Resolution Artifact Elimination Through Incremental Rotation

Anonymous Authors[1]

1. Abstract

Image super-resolution is a rapidly growing area of research in machine learning and computer vision. The core goal of this is to take a low-resolution image and through an algorithm, upscale it to a higher resolution. This can be done trivially by resampling, but this leads to a lot of blurriness in the final image and a loss of detail. The focus of modern research is to upscale an input image while both extrapolating more details and preserving the original structure of the image. There are a variety of approaches to this. Discussed in this paper are SinGAN, Residual Dense Network, and Second-order Attention Network super-resolution methods. Each of these implementations takes a unique approach to the problem with varying results. Residual Dense Network upscaling results in a noticeable number of artifacts in the output image. The expansion of this network that is discussed in this paper performs the RDN upscaling on multiple variations of an image and combines them to help eliminate the artifacts. The most obvious of the artifacts are horizontal aberrations in the image and outlines in areas of high contrast. This method successfully eliminates almost all of the horizontal artifacts and severely diminishes the level of outlining while still performing better than baseline methods. In addition, the lack of mention of these weaknesses in the original implementation paper allows for these issues to go unnoticed and possibly affect the applications that the RDN is used for.

2. Paper Summaries and Reviews

2.1. Paper 1: SinGAN: Learning a Generative Model from a Single Natural Image

2.1.1. SUMMARY:

In *SinGAN: Learning a Generative Model from a Single Natural Image*, Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli discuss a new system that can be trained by a single training image and can perform a vast number of image manipulations. The paper begins by discussing the concept of Generative Adversarial Networks (GANs) and their unique ability to process and create images. GANs are highly successful at generating realistic images given a prompt when trained on large sets of classified data.

SinGAN (Rott Shaham et al., 2019) takes this a step further, generating images from a single training image, forgoing the need for vast amounts of labeled images. This greatly simplifies the ease of use since there is no need to compile a large training dataset. A key component that leads to SinGAN's success is its use of a collection of GANs operating at different scales on a single image. Small-scale GANs capture and recreate fine details like texture and image quality. At the same time, medium-scale GANs are generating objects and key features in the training image in the same style as it due to the small-scale GANs. Finally, large-scale GANs recreate overarching structures and arrangements of features in the training image. Using only a small number of scales leads to the generation of images with the same look and texture as the training image, but without any of the major features and arrangement of objects. As the scale increases, larger structures and patterns are generated.

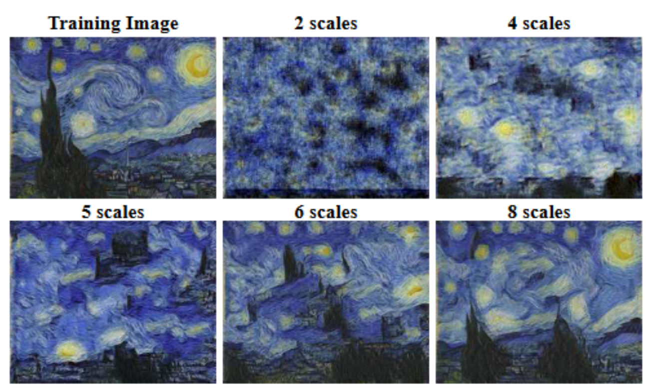

Figure 1. Effects of differing numbers of scales on the generated image.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

2.1.2. Review:

The strength of this technology is in its wide variety of applications. This has vast potential for image manipulation beyond just creating like images from a training data point. Five examples given are paint-to-image, editing, harmonization, super-resolution, and animation. (Rott Shaham et al., 2019) In paint to image, SinGAN is fed a training image, then given a poorly painted scene which is then recreated within the style of the training image. Editing with SinGAN allows a user to rearrange and scale objects in an image which are then modified by the system to mesh with the original unmodified image. Harmonization takes a training image style and input image with clashing styles and transforms them to work in the context of the trained style. Super-resolution allows an image to be up-scaled to a higher resolution and is a unique case where the training and input images are the same. Animation allows for a single image to be animated into a series of images from context. The ability for all of this to be done from a single algorithm and all using just one training image makes SinGAN incredibly versatile and valuable.

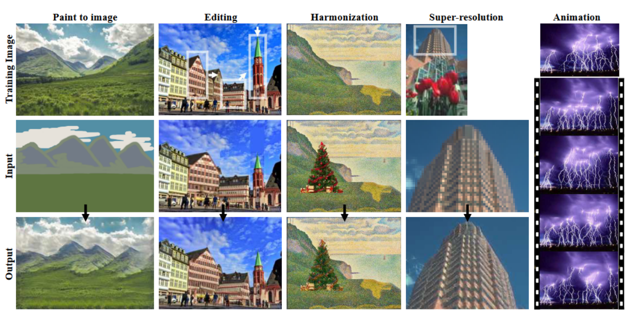

Figure 2. Example of the wide variety of use cases for SinGAN.

As far as weaknesses go, it can only create variations on what it sees, as it only has an understanding of the patterns it is trained on. The example the authors use is that if the model is trained on a photograph of a dog, it will be unable to generate different breeds of dogs. It will create variations of that specific dog, but since it is not also a classifier, it will not classify the subject as a dog and be able to reference other potential representations or breeds of dog. I would like to ask the authors if it would be possible to combine this technology of scaled GANs with established image recognition and classification in order to create a system capable of doing both. A system that could understand the component parts of a training image, then use scaled GANs to generate a new image in the same style and layout as the original while also altering the subjects themselves would be incredibly powerful.

This paper covers a breadth of image processing applications, but the one that links best to the other papers covered is the application to image upscaling. Granted, that is a small component of SinGAN's ability and doesn't technically require scaled GANs to do. However, there are many advantages to the method used here. SinGAN up-scales images by performing basic upsampling on the training/input image, adding noise to it, and then passing it through a final generator. This process is repeated and produces an up-scaled image that preserves fine details and repeated structures. Many other algorithms fall short in this and produce alterations of repeated patterns that differ from the original low-resolution input.

2.2. Paper 2: Residual Dense Network for Image Super-Resolution

2.2.1. Summary:

Residual Dense Network for Image Super-Resolution by Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu describes methods for up-sampling images using a deep convolutional neural network (CNN). The technology unique to this approach that sets it apart from others in the field of image super-resolution is the ability to utilize larger structures from within the image in order to result in a better up-scaled final product. Whereas many upscaling technologies simply look at the information in the immediate vicinity of each pixel, this model considers overarching structures. Feature recognition allows the model to learn about important global aspects of the image and ensure they are preserved throughout the upscaling process.

A common tactic among many image super-resolution models is to use the original low-resolution image in the up-scaling process to preserve major repeated structures and features. The downside of this is a quadratic increase in computational complexity and thus a large increase in the processing requirements and runtime. Dense block processing improves the performance of the model but makes training more difficult. The residual dense network (RDN) configuration proposed by Dai et al. (2019) allows the model to use the original low-resolution images and the benefits gained with that while lowering the training costs typically associated with it, resulting in an overall better-performing image super-resolution algorithm.

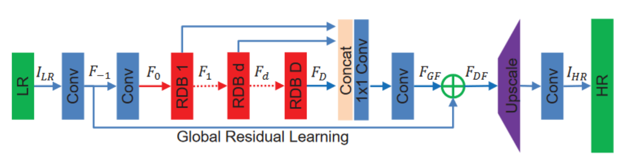

Figure 3. Structure of a Residual Dense Network.

2.2.2. REVIEW:

RDN's strength lies in its combination of residual dense blocks (RDBs) on a local level with global features extracted from the original low-resolution image. Utilizing large-scale features from the training dataset throughout the up-scaling process has noticeable results in decreasing blurring and maintaining structure in the final result relative to other methods of super-resolution. This method is highly robust and performs well across a variety of training and testing images.

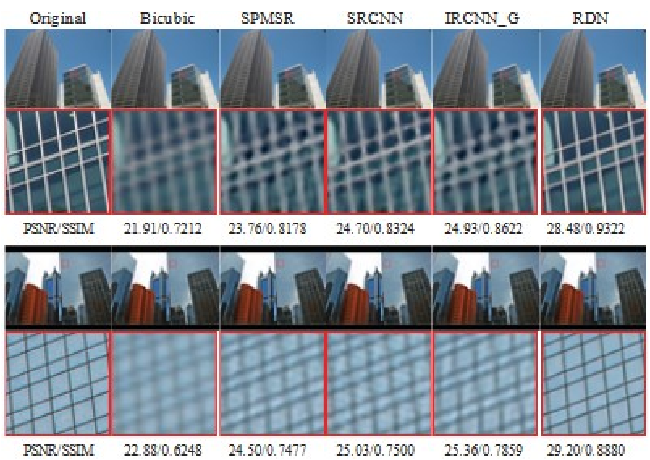

Figure 4. Comparison of results of RDN with other common approaches.

Zhang et al. (2018b) does not list any major drawbacks other than the asymptotic behavior of final image quality over training time. While RDNs are a more efficient method of super-resolution and require less computation for superior results relative to other neural network approaches, they are still fairly intensive due to multiple layers of CNNs being required. I would like to ask the authors what other drawbacks there are to this implementation, and what types or arrangements of images it tends to perform worse on relative to other similar approaches. This paper uses a similar approach to super-resolution as Dai et al. (2019), referencing the original low-resolution image to preserve global structures. RDN falls short of the vast range of image manipulation applications available with SinGAN (Rott Shaham et al., 2019) in that it is only applicable to image upscaling.

2.3. Paper 3: Second-order Attention Network for Single Image Super-Resolution

2.3.1. SUMMARY:

In *Second-order Attention Network for Single Image Super-Resolution* by Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xai, and Lai Zhang, the authors implement a second-order attention network (SAN) to enable better feature preservation and expression in the final high-resolution output image. In addition, non-locally enhanced residual group (NLRG) learning captures global contexts to improve structure representation throughout the up-scaling process. Many other CNN-based super-resolution algorithms are recursive and gradually upsample their previous outputs, discarding the inputs after each iteration. This model makes use of the original training image and intermediate images to increase performance. Using second-order channel attention (SOCA), the network can adaptively focus on feature sets that contribute the most information to the overall image. These key developments work in tandem to produce superior pattern preservation when up-scaling an image.

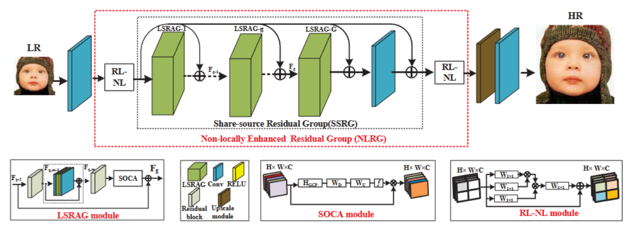

Figure 5. Structure of a Second-order Channel Attention Network.

This paper proposes another method for image super-resolution which works to maintain larger-scale structures within the original image. Its specific use of NLRG allows it to utilize information within the image to enhance the quality of the result globally, as it is not bound to only working with local patterns in the super-resolution process. Finally, SOCA allows it to focus on the most information-dense parts of an image where repeated textures which are up-scaled poorly would be most noticeable by the eye. This allows it to work more efficiently by focusing on key parts of the image and spending less time on unimportant parts.

2.3.2. REVIEW:

The strengths of a SOCA network lay mainly in its accurate results when up-scaling textures. Textures are relatively dense in information and are thus focused on by the network and preserved very well. When compared to other image super-resolution algorithms, SOCA networks perform very strongly overall and even better when it comes to texture quality specifically. Their main weakness is images with many repeated straight edges. Edges are relatively simple repeated patterns and thus are focused on less by the network, giving results that are often second to others like residual channel attention networks (RCAN) (Zhang et al., 2018a). Similar to Zhang et al. (2018b), SOCA networks are only designed for image super-resolution and lacks the broad range of image processing applications of SinGAN (Rott Shaham et al., 2019).

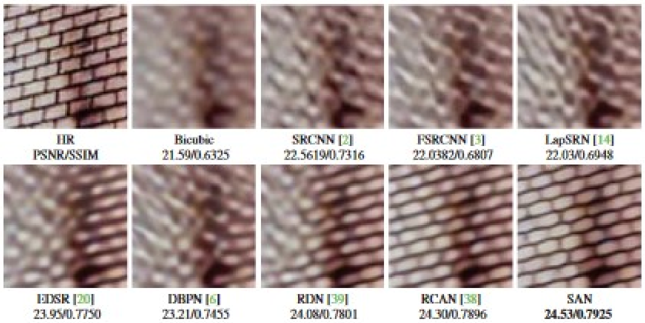

Figure 6. Comparison of results of SOCA networks with other common approaches.

Given the strengths and weaknesses of SOCA networks, I would ask the authors if there would be a way to have the network increase focus on some lower information aspects of the image to preserve repeated simple patterns like straight edges. Coupled with the impressive results it gives with repeated textures, this might make the network more well-rounded overall and make up for its existing shortcomings. Another downside of this implementation is its lack of application range, similar to Zhang et al. (2018b), when compared to a GAN-based approach like SinGAN (Rott Shaham et al., 2019).

3. Implementation, Results, and Discussion

3.1. Implementation

3.1.1. Use of preexisting code

My implementation goal was to create a working setup of a residual dense network (RDN) to generate super-resolution versions of a given image then improve on it. This relates to the paper Zhang et al. (2018b) in which the authors discuss the advantages and results of such a network. There is a code repository listed in the paper that also provides pre-trained models to eliminate the need to perform the computationally intensive training phase and get the RDN working quickly. This code repository can be found at `https://github.com/yulunzhang/RDN`. The original paper was implemented in Lua and MATLAB, but I was able to find a code repository where it had been converted to python at `https://github.com/idealo/image-super-resolution`. My implementation expands on this repository specifically.

Initially, I just implemented the Residual Dense Network from the paper and was able to successfully perform the up-scaling on any provided test image. Below is a sample of the network running on an image of a tree. It is clear that the quality is greatly improved compared to the original low-resolution image and is also superior to a bicubic upscaling.

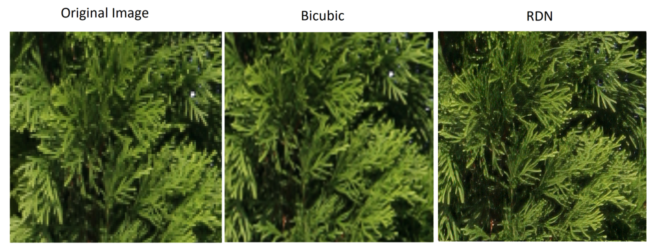

Figure 7. Comparison of RDN versus bicubic upscaling.

These results are very promising, as it shows the good performance of the RDN in image super-resolution. The next step in the project was to begin making and gathering test images to start to identify weaknesses in the algorithm. Once these were identified, I would be able to work on counteracting these shortcomings.

3.1.2. Expanding on the design

One place that I found the original paper lacking was in the discussion of weaknesses. The authors don't go into depth on any of the places where RDNs can fall short of other methods of image super-resolution. To look for weaknesses, I started by first running some basic test images through the RDN. Once I started to notice patterns, I made some custom test images which were able to make the issues more obvious. I was able to find two distinct issues with the output images. Firstly, I noticed that there were some horizontal aberrations in the image. These are present through almost the entire image but are more pronounced over some parts than others depending on the content of the image. The other artifact is actually additive and will create outlines when presented with high-contrast image content.

Figure 8. Closeup on the RDN output image, with visible horizontal aritfacting. It is clearest on the eyebrows and fur.

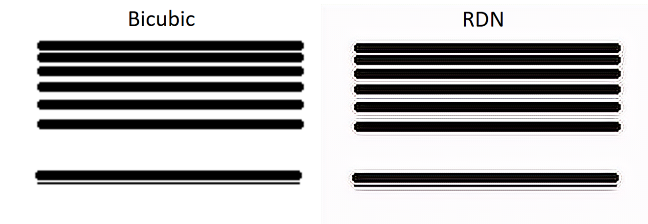

Figure 9. Comparison of bicubic versus RDN upscaling showing outlining artifact.

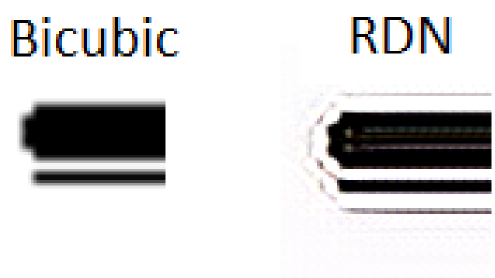

Figure 10. Closeup on the comparison of bicubic versus RDN upscaling, with the outlining artifact clearly visible.

To fix the issue of the horizontal lines, my approach was to rotate the image by a various number of degrees and perform the upscaling on these new images. Running the RDN on these rotated images and then averaging the outputs provides an image free of horizontal artifacting. The issue I ran into here is that the RDN takes a rectangular image as input, and could not simply be given a skewed image. In order to solve this, I used the imgutils python library to rotate the image. This, however, cuts off any part of the image rotated outside of the shape of the original image. To combat this, I added a black border to the outer edge of the image so that it could be rotated without loss of content. The size of this border depends on the size of the image, as a square image will not need a lot of a border, whereas a long rectangular image would need substantial space on the top to rotate into. Once the RDN upscales the rotated image, it needs to be rotated back to its original orientation in order for the averaging to work properly. To do this, the output image is rotated back the same number of degrees that it was rotated before, but in the opposite direction. To remove the border and restore the output image to the same size as the original, it is cropped using array indexing. The resulting image is then appended to an array along with images from the other rotations. Finally, the images are all averaged together and displayed as a final image. Not only does this remove the horizontal artifacts to a level at which they are essentially invisible, but also has a big impact on the outline artifacts.

3.2. Evaluation Method and Results

3.2.1. Results

The evaluation of results when those results are images is usually somewhat difficult, as it is often a qualitative process rather than a quantitative one. There are some methods of quantifying the quality of an image, but that falls outside the scope of this project. Simple visual analysis is enough in this case to see the clear benefit that rotational averaging can provide. As seen in the figures, the horizontal artifacts are largely gone and the outlines are severely diminished. All of this is done while maintaining an image quality still greater than that of the baseline bicubic upscaling.

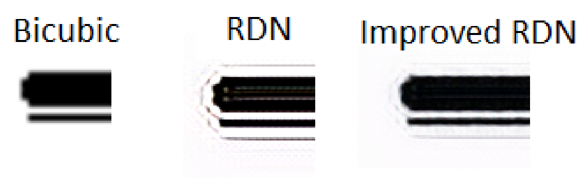

Figure 11. Closeup on the comparison of bicubic versus RDN and the improved RDN upscaling implemented.

Figure 12. Closeup on the improved RDN output image without horizontal aritfacting.

3.2.2. Audience testing

To ensure real results and remove the possibility for bias in the comparison of the images, I performed some audience testing on both the baboon test image and the black and white line image. For the baboon image, the main responses were that the bicubic was the most blurry and the original RDN image looked "pained." The horizontal lines were noticeable in the original RDN image and were said to be absent in the improved RDN test image. For the black and white line test image, audiences found the improved RDN image still clearer than the bicubic image, with distinctly less outline artifacting than the original RDN implementation. In these cases, the process outlined in this paper was able to produce an ideal result by creating a more detailed and higher-resolution image than bicubic while minimizing the amount of visible artifacting.

3.3. Discussion of Results

These results show that when it comes to image super-resolution, there is usually a trade-off that has to be made between image sharpness and artifacting. While the RDN output is very sharp, it is visibly not an original image. Audiences are also able to pick up that while the image is certainly clearer, the image is not original. The rotational averaging method discussed in this paper is able to eliminate these horizontal artifacts and help remove some of the "painted" look of the picture. This made it more likely to be perceived as an unaltered image. As more and more detail is drawn from an image, it seems that there becomes a higher likelihood that things such as pixels and minor upscaling distortions are amplified and lead to a result that no longer represents something that is realistically what a camera would produce. These results also show the benefit of drawing a final image from a large variation of upscaled images in an effort to combat this artifacting. While this may end up leading to a slight loss of quality, it produces something that could reasonably have been produced by a camera or other means.

Larger insights can be made from the original description of using Residual Dense Networks for image super-resolution. As discussed, the paper regarding the implementation of the RDN I used does not discuss any weaknesses of the algorithm, when the outline artifacts represent an issue for potential uses of these images. For example, say some researchers want to train a machine learning algorithm meant to perform object recognition on images captured by a high-resolution camera. Their training dataset, however, may consist of images taken at a lower quality for a number of reasons. The hardware could be not completed yet, or the customer cannot afford to lend them the use of the high-resolution camera for long enough to capture training data, etc. A reasonable idea would be to use an image super-resolution method like RDN to upscale the training images to the same resolution that the final hardware will produce. As shown in this paper, weaknesses in the algorithm may create artifacting that the neural network begins to use as a feature to help classify input images. This could yield very good results in testing, yet could lead to very poor performance in the final application. Because the paper that originally described this RDN made no mention of these weaknesses, there is no way that a researcher would know about this without taking the time to test the RDN extensively. It is for this reason that I believe it is crucial to discuss the shortcomings of a system when writing about it.

References

Dai, T., Cai, J., Zhang, Y., Xia, S.-T., and Zhang, L. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. URL https://openaccess.thecvf.com/content_ICCV_2019/html/Shaham_SinGAN_Learning_a_Generative_Model_From_a_Single_Natural_Image_ICCV_2019_paper.html.

Rott Shaham, T., Dekel, T., and Michaeli, T. Singan: Learning a generative model from a single natural image. In *Computer Vision (ICCV), IEEE International Conference on*, 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Dai_Second-Order_Attention_Network_for_Single_Image_Super-Resolution_CVPR_2019_paper.html.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. Image super-resolution using very deep residual channel attention networks, 2018a. URL https://arxiv.org/abs/1807.02758.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018b. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Residual_Dense_Network_CVPR_2018_paper.html.