

Image quality assessment for determining efficacy and limitations of Super-Resolution Convolutional Neural Network (SRCNN)

Chris M. Ward, Josh Harguess, Brendan Crabb, Shibin Parameswaran
 Space and Naval Warfare Systems Center Pacific
 53560 Hull Street, San Diego, CA 92152-5001
 cward@spawar.navy.mil, joshua.harguess@navy.mil, brendancrabb8388@pointloma.edu,
 shibin.parameswaran@navy.mil

ABSTRACT

Traditional metrics for evaluating the efficacy of image processing techniques do not lend themselves to understanding the capabilities and limitations of modern image processing methods - particularly those enabled by deep learning. When applying image processing in engineering solutions, a scientist or engineer has a need to justify their design decisions with clear metrics. By applying blind/referenceless image spatial quality (BRISQUE), Structural SIMilarity (SSIM) index scores, and Peak signal-to-noise ratio (PSNR) to images before and after image processing, we can quantify quality improvements in a meaningful way and determine the lowest recoverable image quality for a given method.

1. INTRODUCTION

Since the inception of digital imagery, there has been a demand for higher and higher resolution images and videos. Image resolution describes the details contained in an image. Higher resolution, in general, means that more information can be extrapolated from the imagery. Two principal areas motivate the desire for higher resolution imagery: improvement of information for human analysis, and performance improvement of machine perception. While digital image resolution can be classified in different ways: spatial resolution, spectral resolution, temporal resolution, etcetera, this study focuses on methodology for improving spatial resolution.

In recent years, deep learning algorithms and convolutional neural network (CNN) architectures have gained momentum in solving many problems in machine learning (ML) and have, in particular, benefited computer vision research areas. One architecture of interest is the Super-Resolution Convolutional Neural Network (SRCNN), which has demonstrated the application of deep learning to image enhancement.

Additionally, work in no-reference image quality metrics, a growing area of research within computer vision, has produced a relatively new metric model: blind/referenceless image spatial quality evaluator (BRISQUE), which has gained popularity in the evaluation of scene statistics and quantification of loss of naturalness.¹

In this paper we will examine SRCNN and its ability to reconstruct imagery degraded by various means. We will quantify our results by applying the BRISQUE metric, and others, and compare the results with qualitative observations. The organization of the paper is as follows: In Section 2.1 we briefly describe SRCNN and why it was chosen as a vehicle for experimentation in this study. Section 2.2 and 2.3 introduce the datasets and metrics used for the study, respectively. Section 2.4 describes our methodology, and gives the results of the study. We conclude the paper in Section 3.

2. EXPERIMENTATION

In this section, we introduce the datasets used for this paper, briefly explain the methodology for measuring the efficacy of super resolution reconstruction, and present our results.

2.1 Super-Resolution Convolutional Neural Network (SRCNN)

SRCNN² was used as an experimentation vehicle due to its success in recent studies and a growing interest in super-resolution applications. SRCNN was initialized with weights learned from training on the ILSVRC 2013 ImageNet dataset, as described in Dong, Loy, He, and Tong.²

2.2 Data

For this evaluation we utilized familiar imagery from the Set 5 [3] and Set 14 [4] datasets.

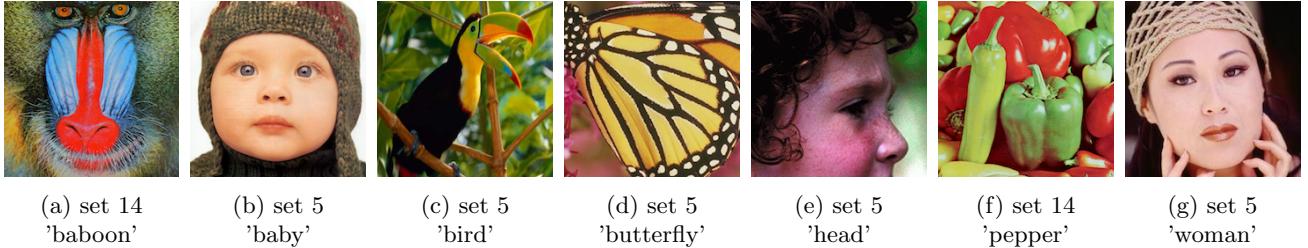


Figure 1: Test images sourced from set 5 and set 14

2.3 Metrics

The performance of SRCNN in different scenarios was measured using two full-reference metrics and one no-reference metric. The two full reference metrics used were the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index. A MATLAB implementation of the algorithm for calculating the SSIM index from Wang et al.⁵ was used throughout testing.

The no-reference metric utilized was the Blind/Referenceless Image Spatial Quality Evaluator(BRISQUE).¹ A MATLAB implementation of this algorithm from Mittal et al.⁶ was used to calculate this metric. PSNR was calculated using the native MATLAB function psnr(A,ref).

2.3.1 BRISQUE: No-reference Image Quality

BRISQUE is a no-reference metric for evaluating Natural Scene Statistics. No-reference metrics assume that no pristine example of an image is present at the time of assigning a metric its quality.⁷ However, within the scope of this experiment, we apply this metric to a reference image in order to observe the impact of image reconstruction on the statistical regularity of the imagery under test. Smaller BRISQUE values indicate low distortion and larger values indicate high levels of image distortion. That is to say, the BRISQUE score is inversely proportional to image quality.

BRISQUE was selected as an evaluation metric in this experiment because it was designed for strong correlation to human judgment of quality across different types of distortion.¹ Because we aim to evaluate an image-correction method intended to enhance imagery in a subjectively human way, BRISQUE is a particularly appealing metric for this experiment.

Given the commonality of subjective observation between SRCNN and BRISQUE, we hypothesize that we will observe high BRISQUE scores commensurate with imagery degradation, and improved/lower BRISQUE scores for reconstructed imagery.

2.3.2 SSIM: Structural SIMilarity Index

SSIM is a full-reference metric that describes the statistical similarity of two images. In our experimentation we take a sample image and declare it as a pristine reference, regardless of any preexisting degradation, noise, or anomalies. Processed imagery is then compared to this reference in the computation of the SSIM metric.

As described in Wang, Bovik, Sheikh, and Simoncelli,⁵ we use a mean SSIM (MSSIM) index to evaluate the image quality holistically, given that the local application of SSIM provides a more accurate representation of statistical features in an image. Using the SSIM Index, we aim to quantify the similarity between images before and after successive rounds of processing.

The expected behavior of this metric is a gradually decreasing Index value that is consistent with the amount of applied processing.

2.3.3 PSNR: Peak Signal-to-Noise Ratio

The Peak Signal-to-Noise Ratio represents the ratio of the reference image pixels to noise pixels introduced during processing. In the scope of this experiment, PSNR is used to evaluate how well processing and correction methods restore a degraded image to its baseline state. Reference images are assumed perfect with zero noise giving them a PSNR score of ∞ . Processed (reconstructed) images are referenced to their respective baseline images.

In examination of the reconstructed imagery, a high PSNR indicates an effectively reconstructed image, whereas a low score indicates persistence of distortion after reconstruction.

2.4 Methodology

For all experiments, SRCNN basic network settings of $f_1 = 9$, $f_2 = 1$, $f_3 = 5$, $n_1 = 64$, and $n_2 = 32$ were used.² This network was trained with an upscaling factor of 3 on over five million sub-images from the ILSVRC 2013 ImageNet detection training partition. To handle color images in the RGB color space, the network design was adapted to deal with three channels simultaneously by setting the input channels to $c = 3$. This network was trained on the luminance channel in the YCbCr color space as a single-channel ($c = 1$) network; however, previous work² has showed that training on the Y channel only produced good performance, as measured by average PSNR (dB) scores on color images in the RGB color space because of the high cross-correlation among RGB channels.

2.4.1 Image Compression

The performance of SRCNN on single-image super-resolution was tested on images with JPEG compression artifacts to gauge its performance at reconstructing poor quality images of different types. Using a JPEG compression artifact generator, images in our dataset were given compression artifacts of varying degrees. These images were then reconstructed using the SRCNN and evaluated using PSNR, SSIM, and BRISQUE.

2.4.2 Successive Image Correction

Using the quality metrics PSNR, SSIM, and BRISQUE, the effect of consecutive rounds of SRCNN correction was observed. To begin, each image in our dataset was shrunk and upsampled by a scaling factor of 3 using bicubic interpolation to produce a low-resolution image that was the same size as the original. This image was then reconstructed using SRCNN and evaluated using our three image quality metrics. Without rescaling, this image was processed with SRCNN three additional times, observing our key metrics after each consecutive iteration.

2.4.3 Scaling Factor

The performance of SRCNN at single-image super resolution was tested on images resized by different scaling factors. Each image in our dataset was shrunk and upsampled by a scaling factor of 2, 3, and 4 using bicubic interpolation to produce a low resolution image that was the same size as the original. This image was then reconstructed using the SRCNN and evaluated using PSNR, SSIM, and BRISQUE. Using the reconstructed image as the input, this method of resizing and reconstructing was repeated an additional three times to observe the effects of repeated resizing and single-image super resolution.

2.4.4 Incremental Scaling vs Large Single-Shot Scaling

In order to evaluate the ability of SRCNN to correct progressive amounts of interpolation distortion, we processed images scaled by the same factor, but with a varying number of upsampling stages. Once again, we used bicubic interpolation the method of image scaling.

We performed two consecutive upsampling operations of 2x scaling factor and then evaluated the performance of SRCNN reconstruction on the upsampled imagery. We then upsampled the same images in one operation, but by a scaling factor of 4x. The images were then processed with SRCNN reconstruction, and the results were compared.

2.5 Results

This section describes the observations made during our experimentation. In general our results were consistent with our initial assumptions. Each experiment yielded its own interesting data points and anomalies of note.

2.5.1 Effect of Image Compression

The results of applying SRCNN to imagery degraded with JPEG compression artifacts were not as predicted. As expected, insertion of JPEG artifacts degraded each of our metrics - figure 3b shows a consistent reduction in structural similarity, and raised levels of noise are apparent in figure 3c. Post-interpolation BRISQUE scores also degraded in every image under test [Figure 3a]. These results are qualitatively observable, noting the visible artifacts present in figures 2b and 2e.

Because SRCNN was not trained to correct compression artifacts, we were uncertain how well it would reconstruct imagery with this type of degradation. Qualitatively, there was little improvement after reconstruction. Note the persistence of compression artifacts in figures 2c and 2f. Interestingly, however, SRCNN reconstruction **improved** the BRISQUE scores for nearly every sample tested, besting the score of our reference images in several cases.

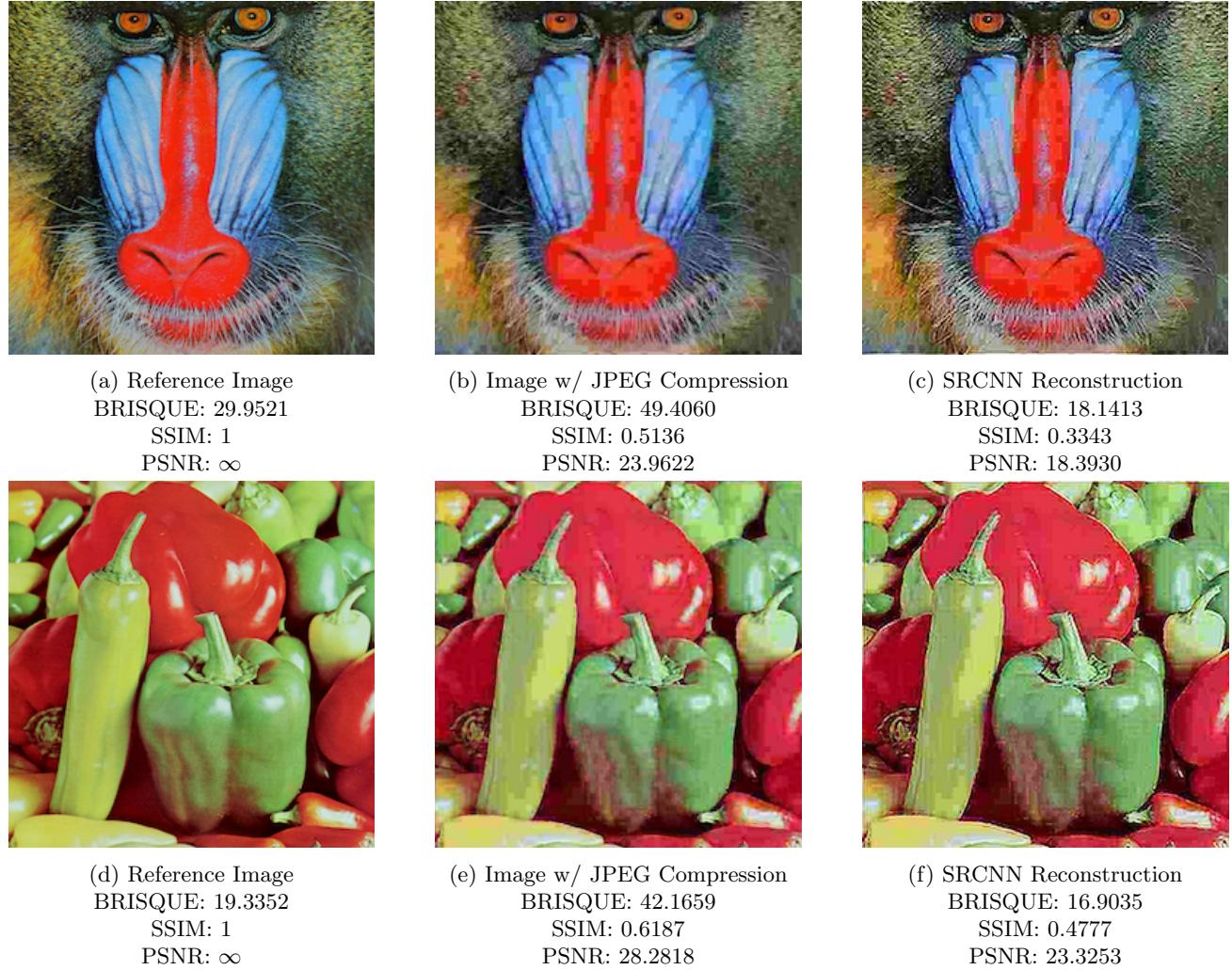


Figure 2: Effect of compression artifacts

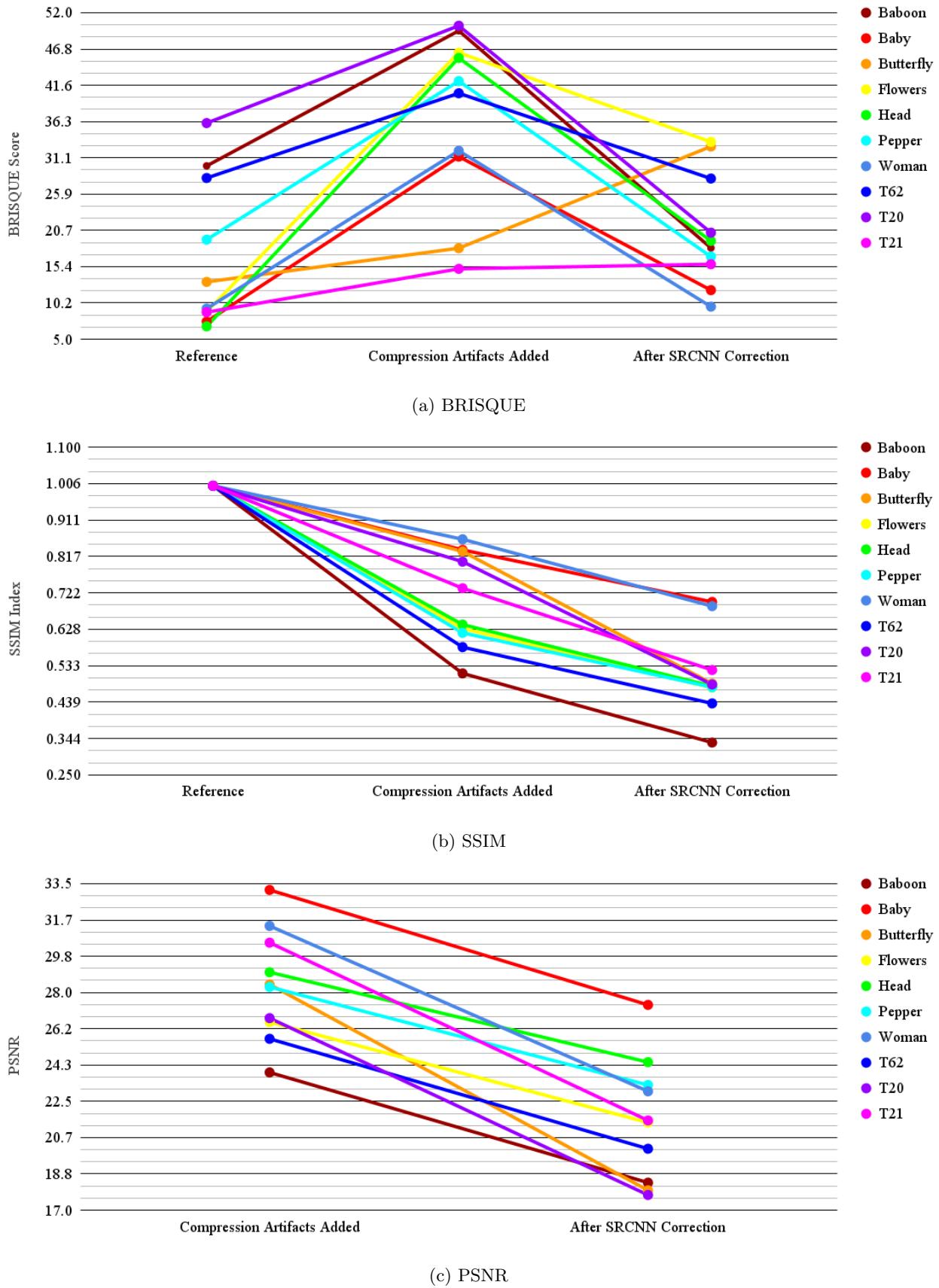


Figure 3: Effect of compression artifacts

2.5.2 Effect of Successive Image Correction

Both BRISQUE and SSIM metrics were degraded by interpolation and, as expected, SRCNN reconstruction improved BRISQUE and PSNR measurements. While each image tested saw improved BRISQUE scores after a single reconstruction pass, subsequent passes yielded nominal improvement and ultimately had a negative impact (Figure 8). Single-pass reconstruction had no significant impact on structure similarity, but as seen in figure 7 there was significant structure change with each subsequent iteration. In section 2.5.3 we discuss how scaling factor influences these trends.

Qualitative observations show that 'sharpness' increases with each reconstruction pass. At 2x interpolation factor we begin to see 'ringing' about edges after three passes through SRCNN. This edge-ringing can be clearly observed after four passes in figures 4f, 4l, and 4r. Based on the Gibbs phenomena, we attribute this effect to the progressive approximation of a discontinuous function by SRCNN. Increasing BRISQUE scores confirm a loss of 'naturalness' in our test imagery.

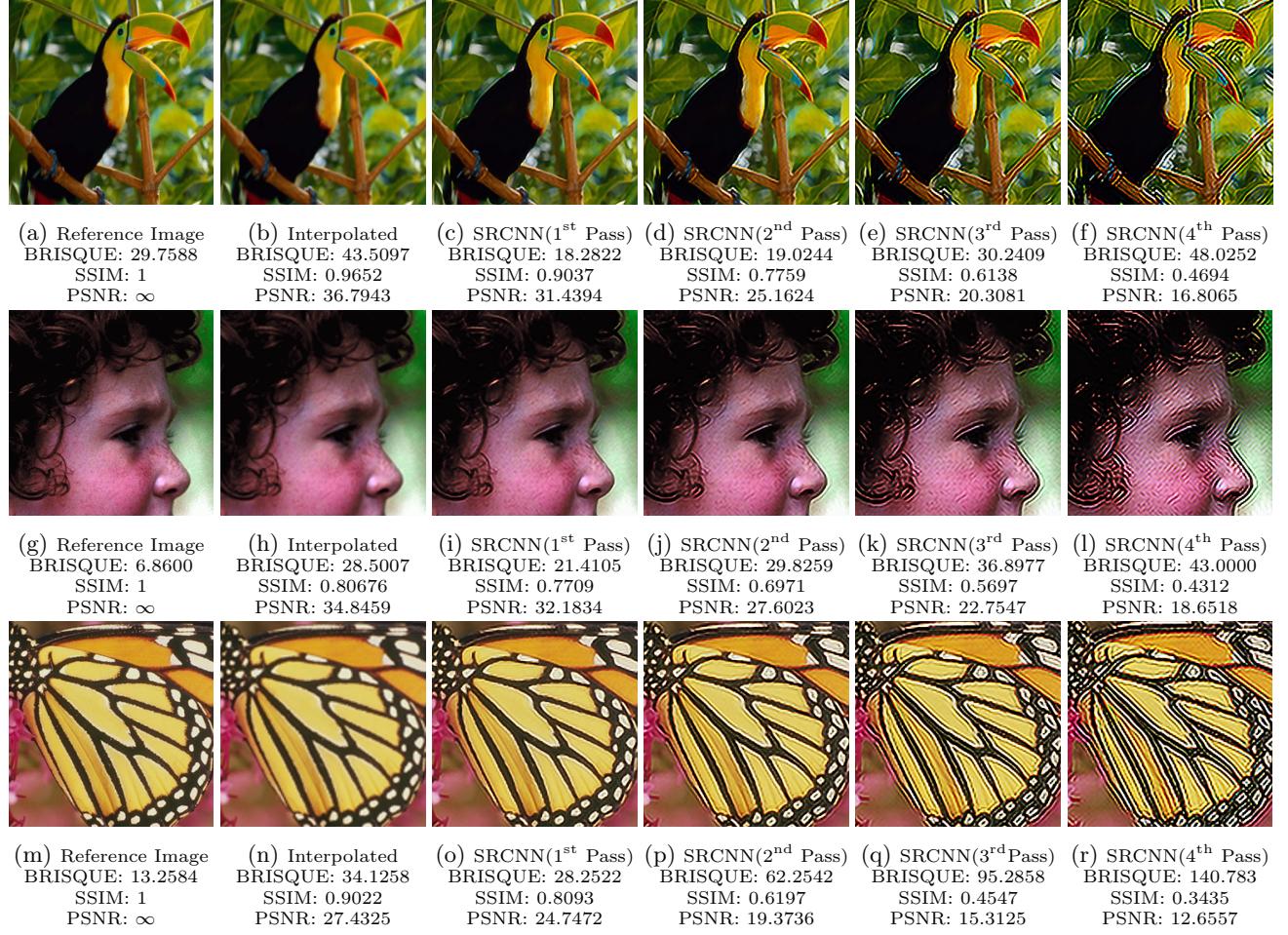


Figure 4: Effect of successive SRCNN correction at 2x Interpolation factor

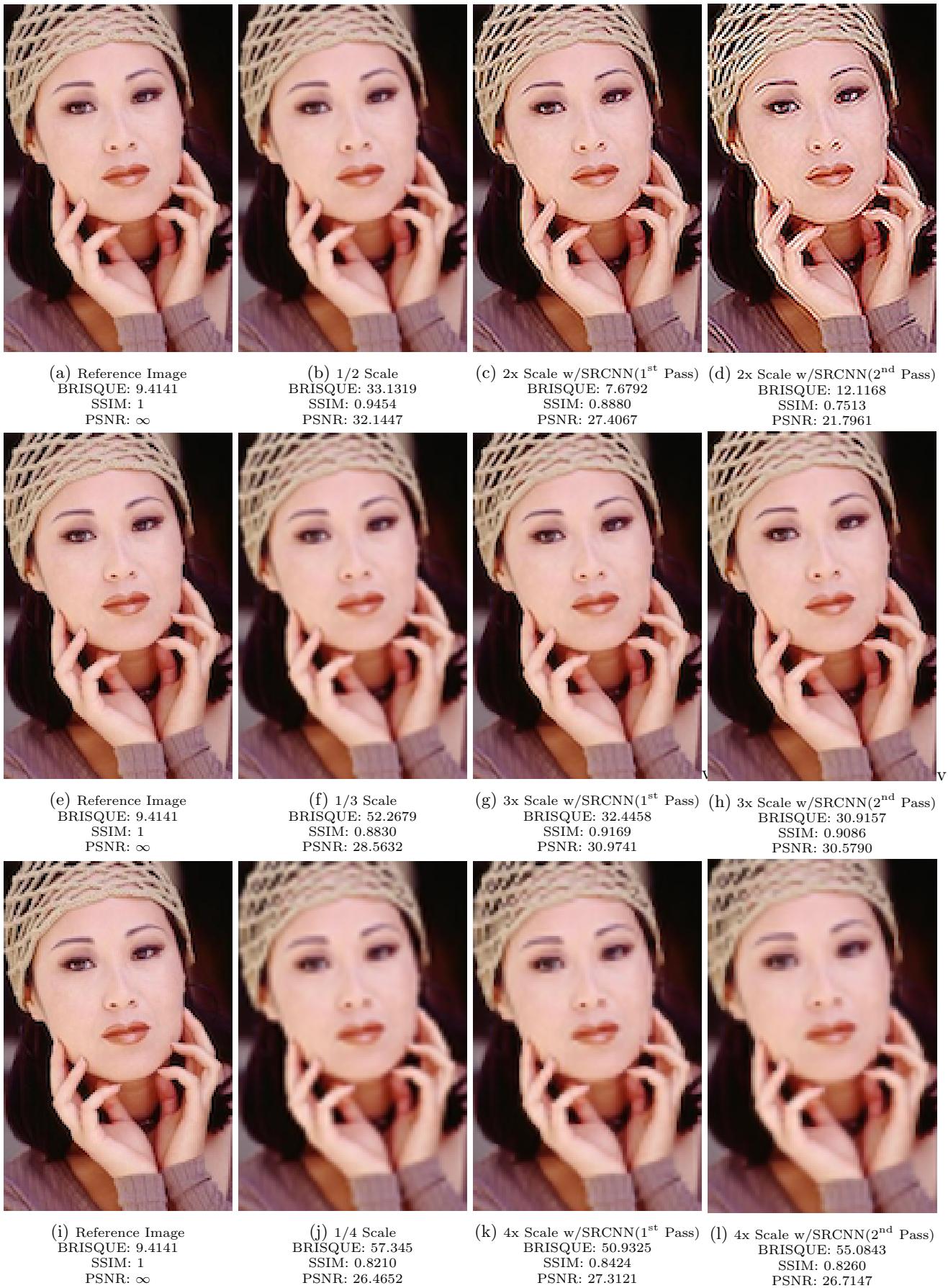
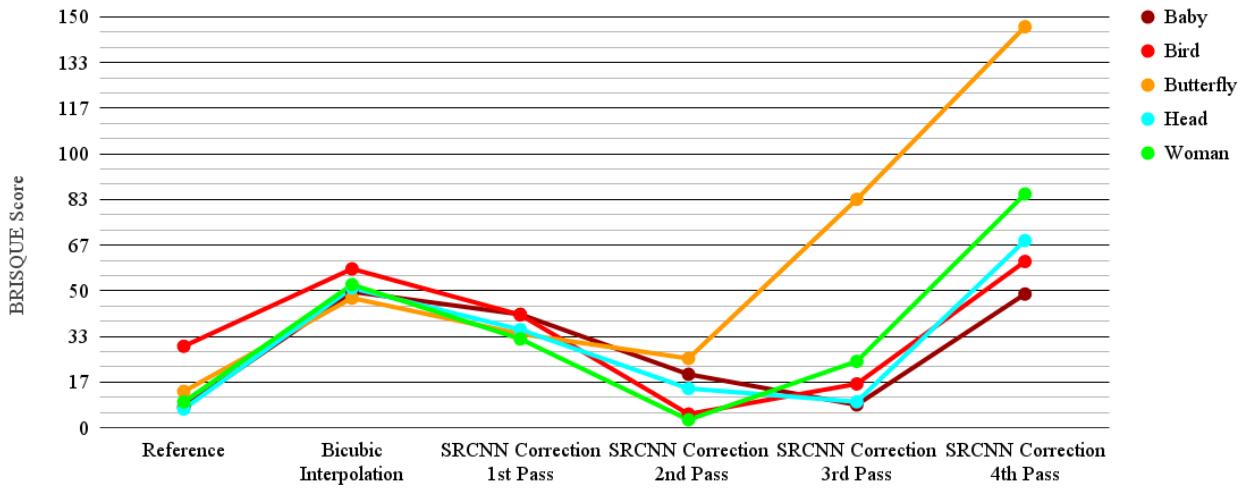
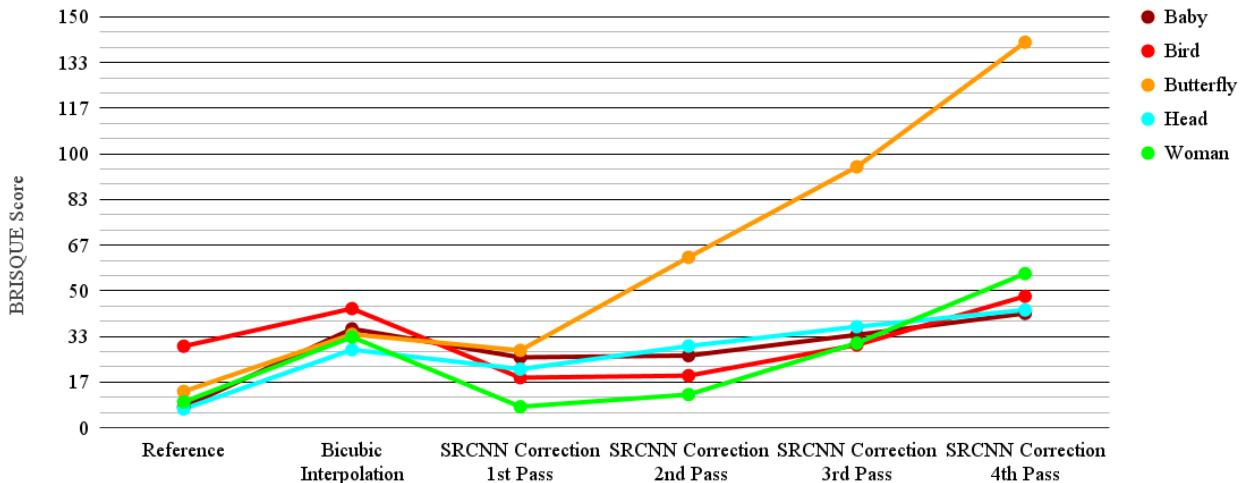


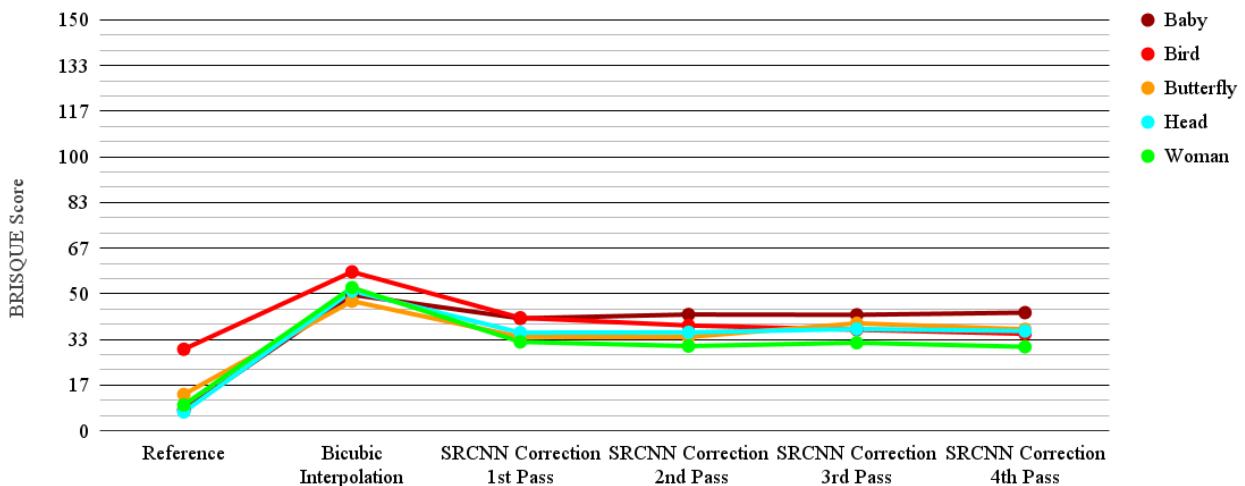
Figure 5: Effect of Scaling Factor



(a) 1x Bicubic Interpolation Factor

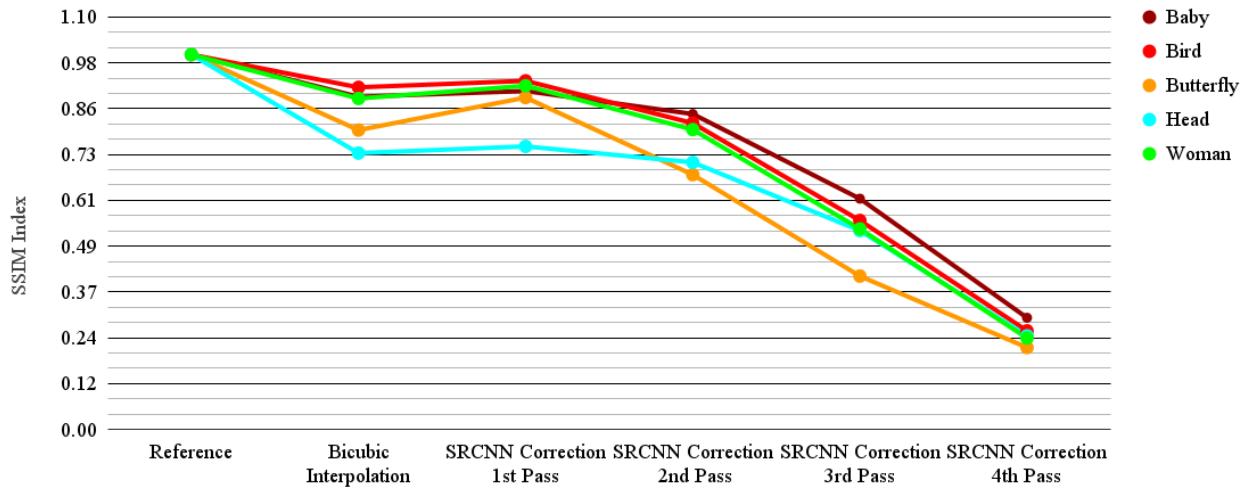


(b) 2x Bicubic Interpolation Factor

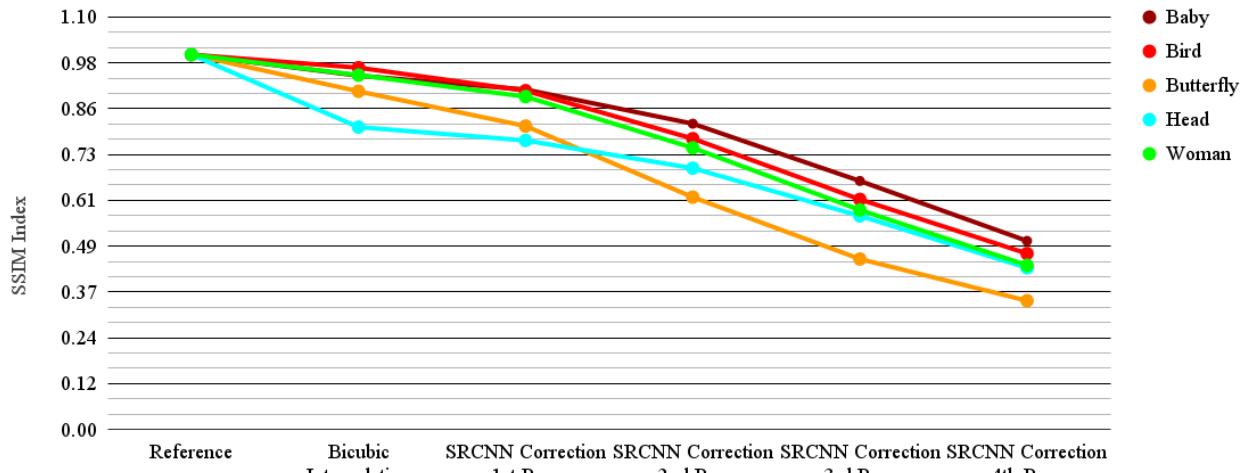


(c) 3x Bicubic Interpolation Factor

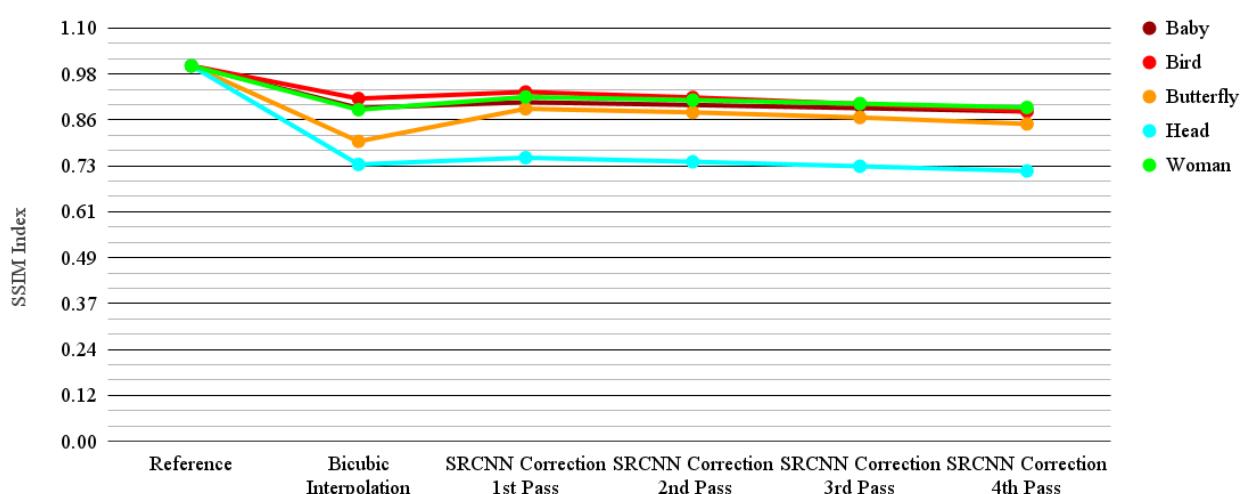
Figure 6: Effects of SRCNN on BRISQUE Score



(a) 1x Bicubic Interpolation Factor

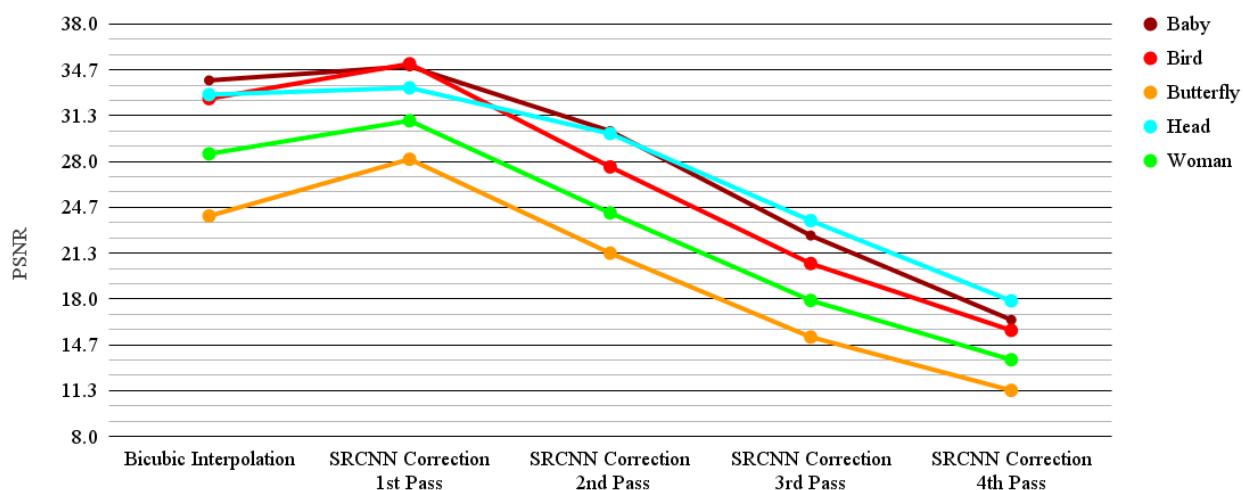


(b) 2x Bicubic Interpolation Factor

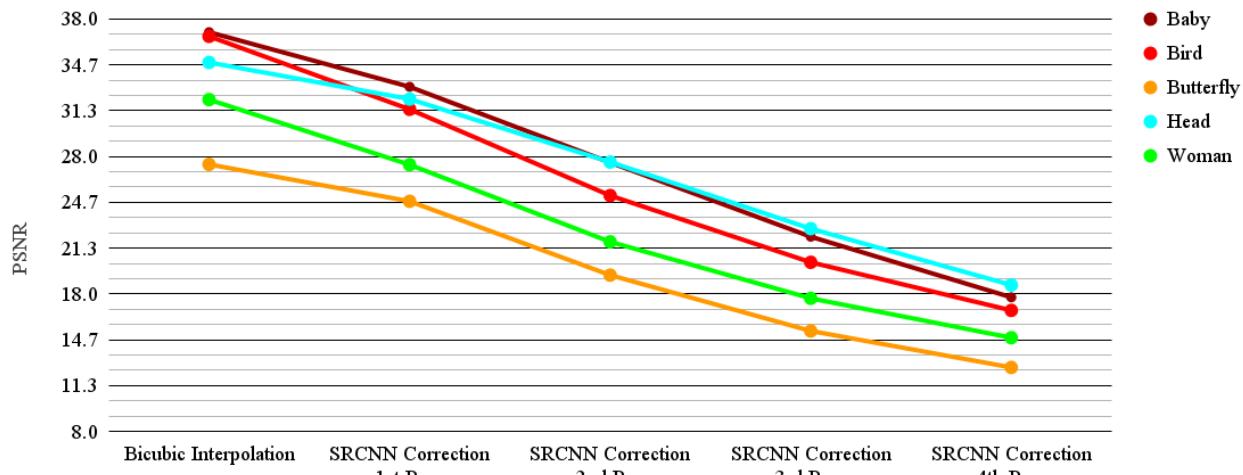


(c) 3x Bicubic Interpolation Factor

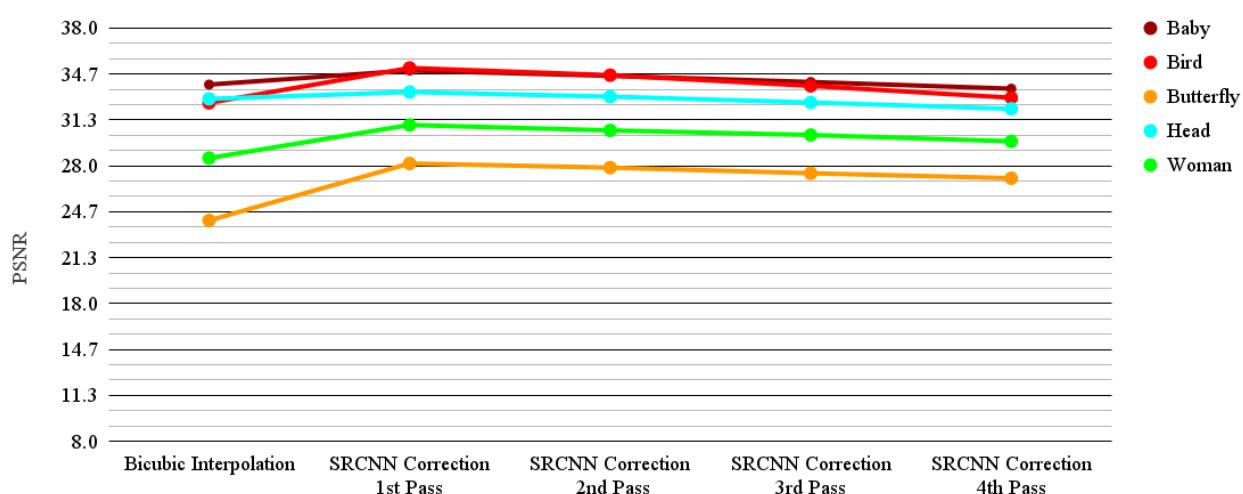
Figure 7: Effects of SRCNN on SSIM Index



(a) 1x Bicubic Interpolation Factor



(b) 2x Bicubic Interpolation Factor



(c) 3x Bicubic Interpolation Factor

Figure 8: Effects of SRCNN on PSNR

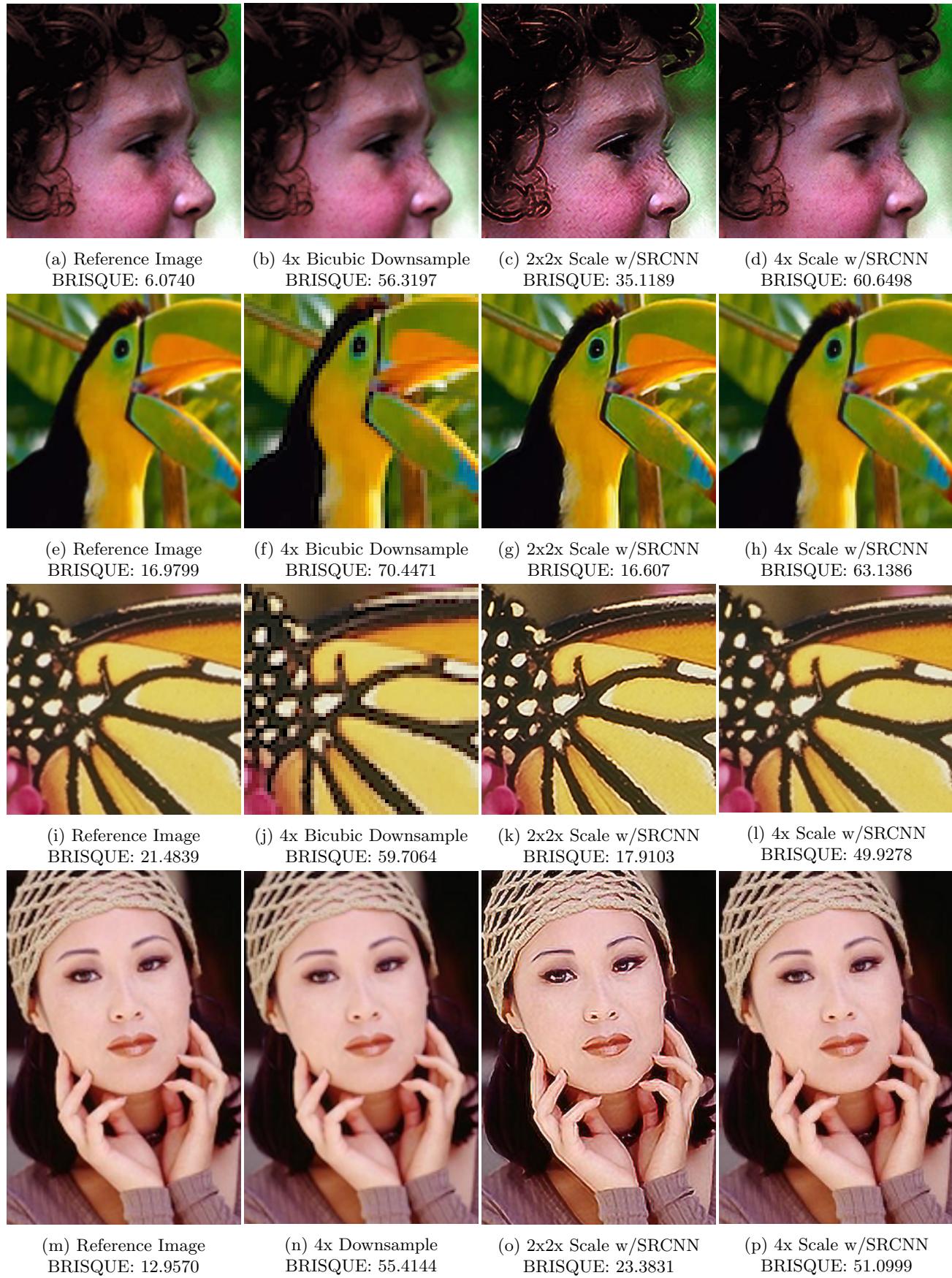


Figure 9: Efficacy of correction on incremental upsampling vs large single-shot upsampling

2.5.3 Effect of Scaling Factor

Examination of data in Figures 6, 7, and 8 show the effect of scaling factor on our test imagery. As plotted, we note that increased scaling factor mitigates the efficacy of SRCNN and it effects on BRISQUE, SSIM, and PSNR. This is particularly true in the 3rd and 4th reconstruction passes. Moreover, this observation holds true qualitatively. In figure 5 we see that images scaled by a factor of two, have a much more visual response to SRCNN. When compared to imagery scaled by a 3x factor, we see that the sharpening/enhancement is much more prominent in the 2x test case. Imagery scaled at 4x is even less responsive to SRCNN reconstruction passes.

2.5.4 Efficacy of Reconstruction on Incremental Upsampling vs Large Single-Shot Upsampling

The efficacy of SRCNN reconstruction on incrementally scaled imagery (2x2x) outperformed 4x single-shot scaling in every test case. Examination of Figure 9 shows much sharper details present in all 2x2x cases. Quantitatively, the BRISQUE measurements of 2x2x test cases agree with visual inspection, yielding better scores than 4x in each tested image. In several particularly noteworthy cases (Figures 9g and 9k), SRCNN reconstruction yielded BRISQUE scores that surpass those of their corresponding reference images (Figures 9e and 9i).

3. CONCLUSION AND FUTURE WORK

In this paper we have applied three different metrics (BRISQUE, SSIM, and PSNR) to imagery that has been modified by varying means, and then reconstructed using the Super-Resolution Convolutional Neural Network (SRCNN). SRCNN reconstructed images as expected, sharpening imagery affected by bicubic interpolation. The approach of using the BRISQUE algorithm to evaluate SRCNN revealed that SRCNN successfully restores the 'naturalness' of imagery, but is not without limitation. Additionally, we observed the role that image scaling factor plays on the efficacy of SRCNN.

We also observed that other types of distortion, such as JPEG compression artifacts, are not only resistant to SRCNN reconstruction, but produce erratic BRISQUE scores. One area of future work is to study the 'gaussianess' of these images to better understand why the BRISQUE metric improved despite the persistence of compression artifacts after SRCNN reconstruction.

We are interested in using BRISQUE and SRCNN to further study future image processing neural networks, and advance work in adversarial imagery⁷ by attempting to correct adversarial features with SRCNN. In general, we hope to work toward building more robust metrics for analyzing deep-learning architectures for image processing, and use them in the development of high performing deep learning architectures.

REFERENCES

- [1] Mittal, A., Moorthy, A. K., and Bovik, A. C., "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing* 21(12), 4695–4708 (2012).
- [2] Dong, C., Loy, C. C., He, K., and Tang, X., "Learning a deep convolutional network for image super-resolution," in [European Conference on Computer Vision], 184–199, Springer (2014).
- [3] Bevilacqua, M., Roumy, A., Guillemot, C., and Alberi-Morel, M. L., "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," (2012).
- [4] Zeyde, R., Elad, M., and Protter, M., "On single image scale-up using sparse-representations," in [International conference on curves and surfaces], 711–730, Springer (2010).
- [5] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* 13(4), 600–612 (2004).
- [6] Mittal, A., Moorthy, A., and Bovik, A., "Brisque software release," (2011).
- [7] Harguess, J., Miclat, J., and Raheema, J., "Using image quality metrics to identify adversarial imagery for deep learning networks," in [SPIE Defense+ Security], 1019907–1019907, International Society for Optics and Photonics (2017).