Evaluating methods for the detection of defocused images

Comparing AWS sharpness measurement with four eye-region sharpness metrics; Laplacian Variance, Perceptual Blur Metric, Tenengrad Variance and Wavelet Coefficients Variance. A statistical analysis of the ability of each metric to differentiate between images of known sharpness was performed, along with assessments of speed.

1. Introduction

The goal of this research is to improve the reliability and accuracy of the automatic focus measurement currently employed for the detection and exclusion of out of focus images.

The current method relies on the face detection sharpness response returned by AWS, which has been shown to be imperfect. Although it's response has good linearity within each image, i.e. a series of progressively less focused versions of the same image yield steadily decreasing sharpness values, it seems highly affected by image content, so the overall values change across different images. This inhibits our ability to set a sharpness threshold, because, depending on the value set, we will always exclude some in-focus images or include some out-of-focus images.

One way to compensate for the bias that differing image content introduces in sharpness measurement would be to run the measurement over a region of the image which does not vary significantly between images. For this reason, the proposed methods all measure sharpness over the eye region of the face in each image. The eye region was chosen because an in-focus image of the eye is guaranteed to have regions of fine detail, so the absence of these details can be considered a definite sign of blur. Additionally the eye is the region of the image that we care about being in-focus; a photograph with narrow depth of field, where most of the image is out-of-focus but the eyes are in focus would still be considered acceptably sharp. A final benefit of running the measurement over a small region of the image is that it takes substantially less time to compute.

2. Methodology

An analysis of prior approaches to sharpness detection [1] provided a starting point and a selection of potential algorithms. Three of the highest performing algorithms were chosen from the paper along with one from a Python image processing library [2]. Under ideal conditions we would test against a series of photographs that were taken both in-focus and out-of-focus, as this replicates the conditions under which the metrics will be used. Finding enough out-of-focus images to run the experiment on however would be difficult. A large number of images would be required for the findings to be considered statistically significant. For this reason a series of 855 in-focus images were used, which were synthetically defocused by applying five Gaussian blurs of increasing standard deviation. Gaussian blur is visibly different from the defocus caused by a lens, however I couldn't easily replicate true lens blur programmatically. It seems plausible that it wouldn't make much difference, however future experiments should test this theory.

The data collection process consisted of opening each image, sending to AWS to collect sharpness data and eye coordinates, running each metric on a 96px x 96px box around each eye, saving the results to a CSV file, then blurring the image and repeating for all of the 5 blur amounts.

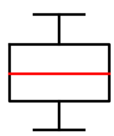
The results contain the blur amount, AWS sharpness score, alongside the return values from each metric for the left eye, right eye, mean value of both eyes and the time taken.

3. Results

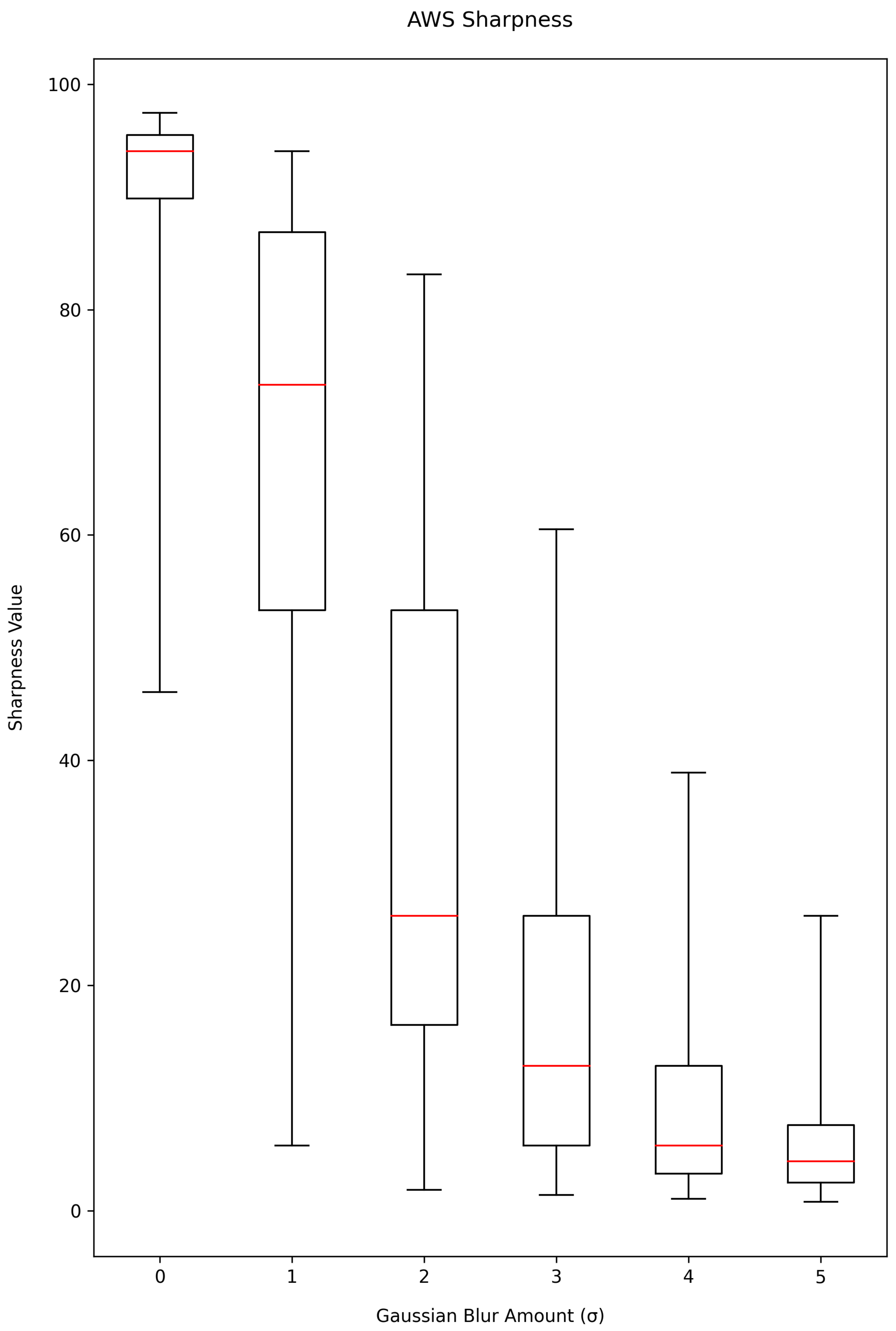
The mean times per image, taken across all images for each metric are as follows:

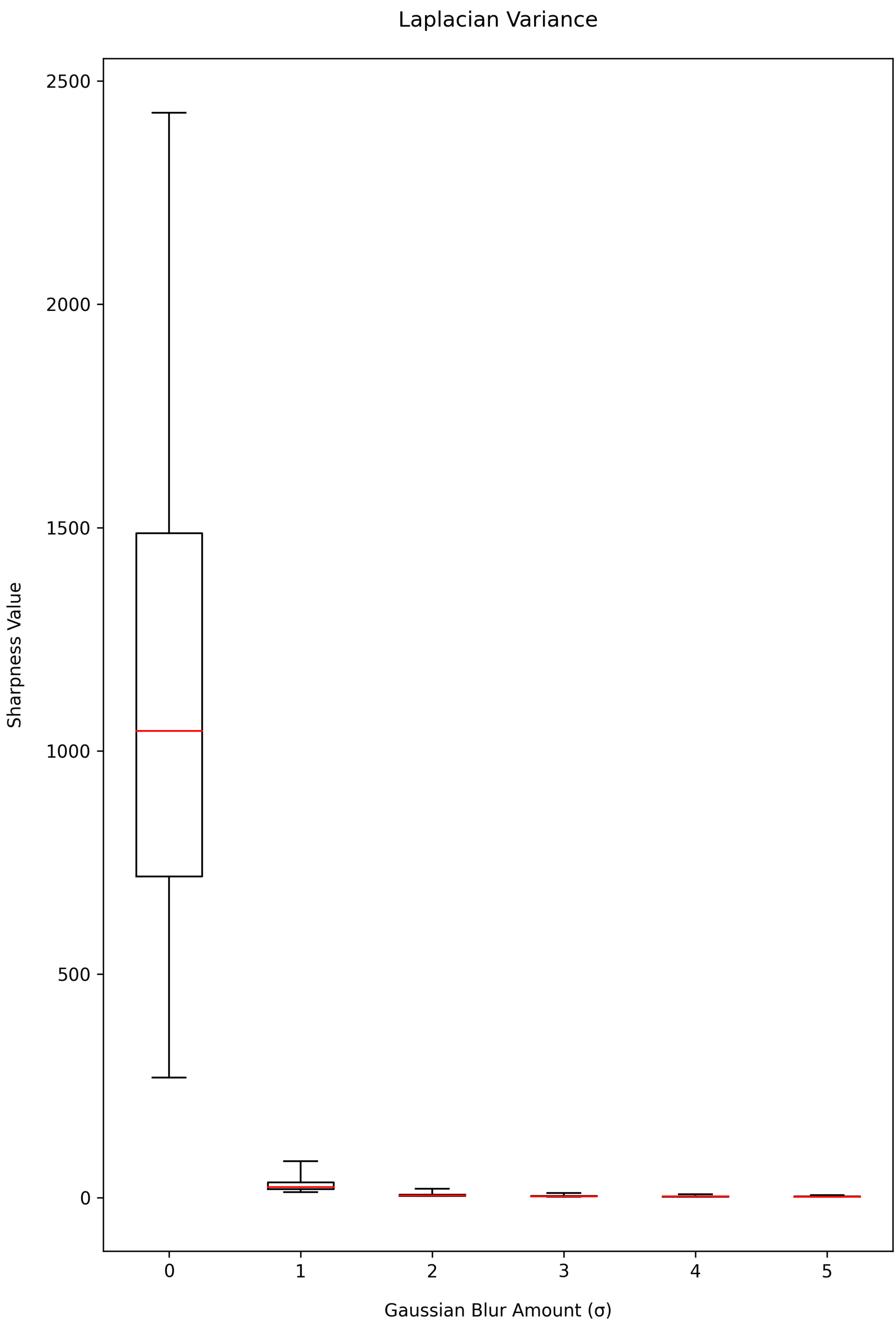
Laplacian Variance	1.87 ms
Perceptual Blur Metric	13.49 ms
Tenengrad Variance	2.17 ms
Wavelet Coefficients Variance	3.72 ms

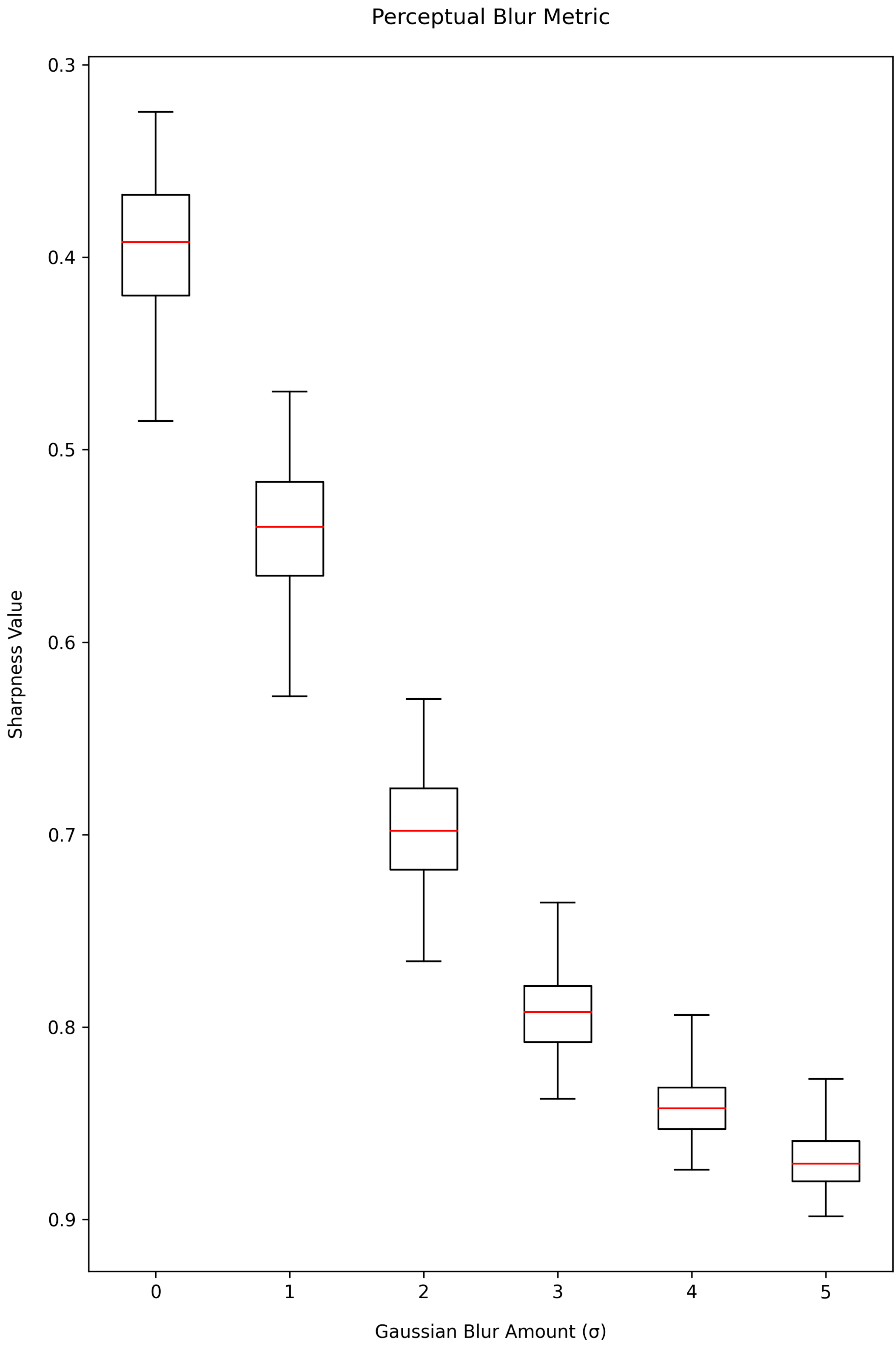
The box plots, illustrating the distribution of response for each metric, can be found on the following pages. Plots which show greater separation between each box demonstrate a greater ability of that metric to objectively characterise blur. The features of the box are explained below. Outliers were omitted from the plot because their inclusion impeded legibility and their exclusion doesn't materially affect interpretation.

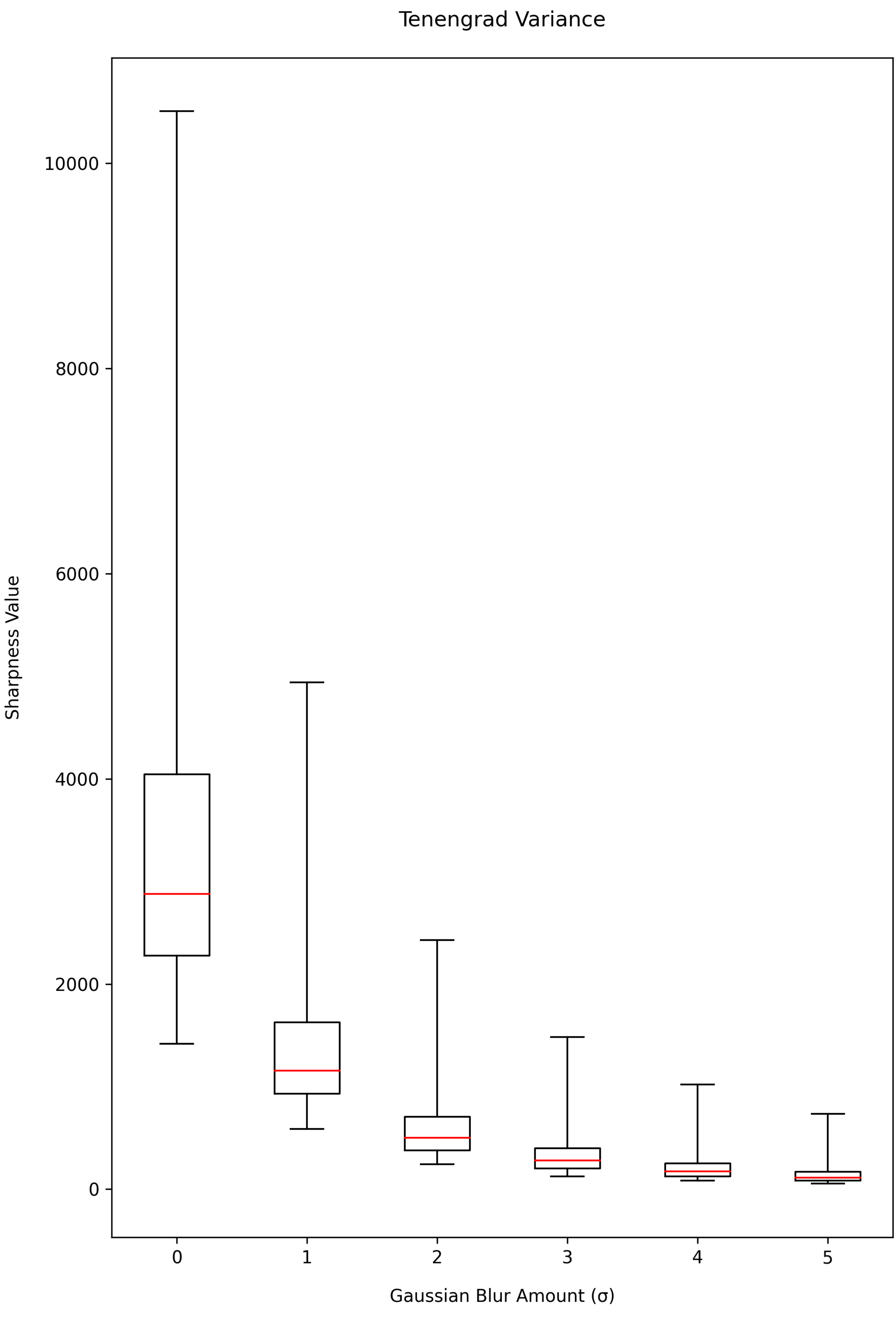


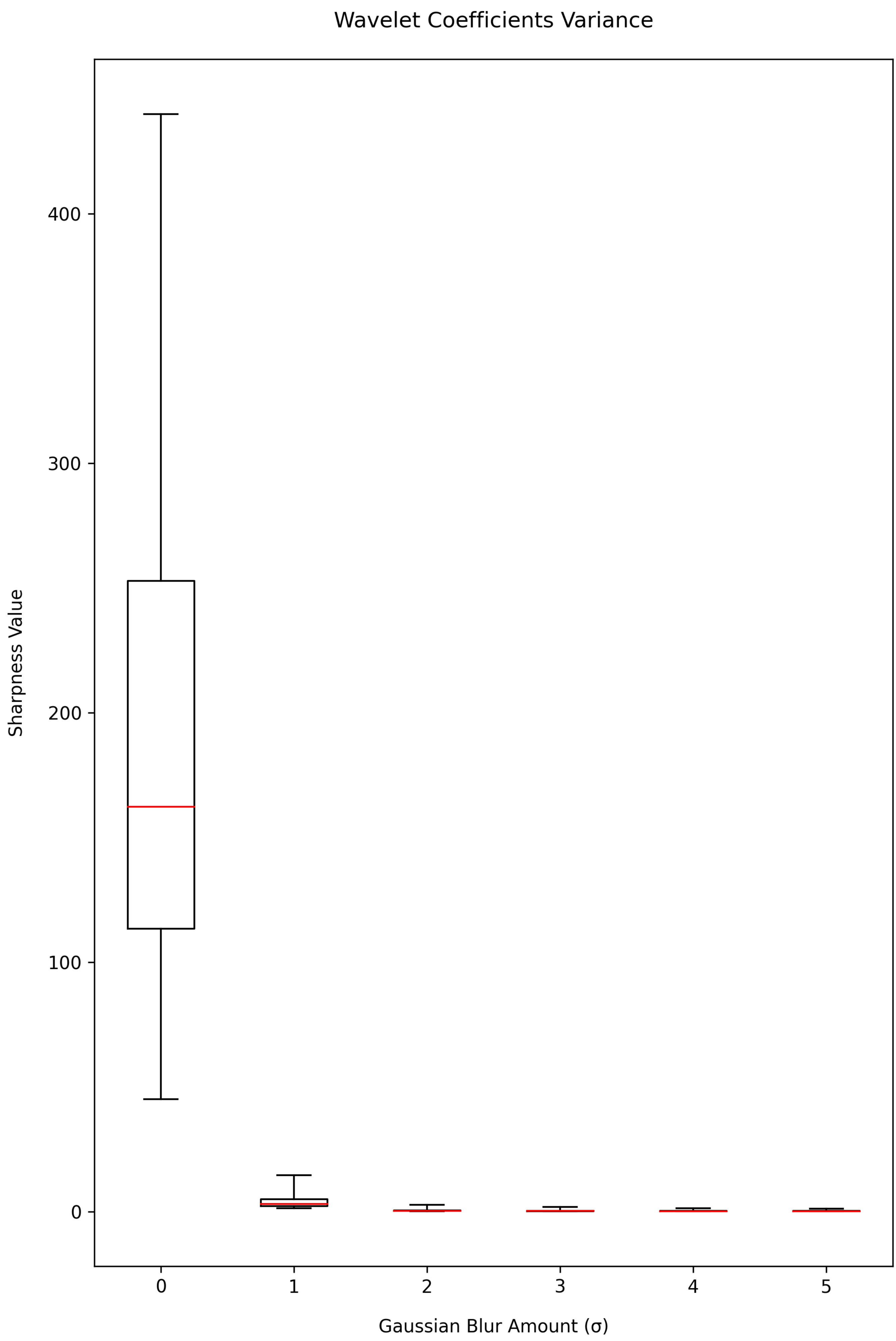
98th percentile
3rd quartile
median
1st quartile
2nd percentile


AWS Sharpness
Sharpness Value
Gaussian Blur Amount (σ)
0
20
40
60
80
100
0
1
2
3
4
5


Laplacian Variance
Sharpness Value
2500
2000
1500
1000
500
0
0
1
2
3
4
5
Gaussian Blur Amount (σ)


Perceptual Blur Metric
Sharpness Value
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
1
2
3
4
5
Gaussian Blur Amount (σ)


Tenengrad Variance
Sharpness Value
0
2000
4000
6000
8000
10000
0
1
2
3
4
5
Gaussian Blur Amount (σ)


Wavelet Coefficients Variance
Sharpness Value
400
300
200
100
0
0
1
2
3
4
5
Gaussian Blur Amount (σ)

4. Conclusion

Laplacian Variance and Wavelet Coefficients Variance are the clear winners out of the four metrics tested. Their response across all images is proportional to the amount of blur in each image and so have a high ability to differentiate between in-focus and out-of-focus images. Given that Laplacian Variance shows the fastest computation time of the two, it should be the preferred option.

5. References

[1] Said Pertuz, Domenec Puig, Miguel Angel Garcia, Analysis of focus measure operators for shape-from-focus in: Pattern Recognition, 7 November 2012

[2] Frédérique Crété-Roffet, Thierry Dolmiere, Patricia Ladret, Marina Nicolas. The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric. SPIE Electronic Imaging, Symposium Conf Human Vision and Electronic Imaging, Jan 2007, San Jose, United States. pp.EI, 6492-16. ffhal-00232709