Supplementary Material

This supplementary document provides a detailed ablation study justifying the choice of hyperparameters and configurations for the artifact detection methods proposed in the paper *JPEG AI Image Compression Visual Artifacts: Detection Methods and Dataset*. It includes evaluations of the methods' performance across various parameter configurations. All comparisons presented in this document were conducted on a test dataset specific to each artifact type, consisting of 100 images with and without artifacts.

The following sections describe the specific evaluations and results for each artifact detection method. Each section highlights the key hyperparameters and configurations affecting the performance of the respective method.

A. Texture-Distortion Detection

We analyzed the impact of the internal similarity metric and kernel size used for averaging. As shown in Table I, the best AUC score of 0.90 was achieved with the MS-SSIM metric and a kernel size of 32, outperforming other combinations of metrics and kernel sizes.

TABLE I
AUC results for texture-distortion detection method using different internal metrics and kernel sizes.

Internal Metric / Kernel Size	**16**	**32**	**64**	**128**	**256**
MS-SSIM [1]	0.88	**0.90**	0.89	0.88	0.85
SSIM [2]	0.89	0.88	0.86	0.86	0.84
PSNR	0.81	0.81	0.81	0.82	0.79

B. Boundary-Texture-Distortion Detection

We evaluated the influence of the kernel size used in the average pooling operation. The results in Table II demonstrate that the best AUC score of 0.87 was achieved with a kernel size of 128. Based on these results, a kernel size of 128 was selected for the boundary-texture-distortion detection method.

TABLE II
AUC results for boundary-texture-distortion detection method using different kernel sizes.

Kernel Size	**AUC**
64	0.86
128	**0.87**
256	0.86

C. Large-Color-Distortion Detection

We evaluated the effects of kernel size and the hyperparameters t_1 and t_2. Table III compares AUC scores for different kernel sizes, and Table IV provides details for t_1 and t_2 with the best-performing kernel size of 64.

D. Small-Color-Distortion Detection

We evaluated the metric with different kernel widths and calculated AUC scores on the test set of images to determine which kernel width is more suitable for detecting small color distortions. Table V shows the AUC scores for different kernel widths.

TABLE III
AUC results for large-color-distortion detection method using different kernel sizes.

Kernel Size	**AUC**
32	0.84
64	**0.85**
128	0.84

TABLE IV
AUC results for large-color-distortion detection method using different t_1 and t_2 values for kernel size 64.

t_2 \ t_1	**2**	**3**	**4**
6	0.85	0.84	0.68
7	0.83	**0.85**	0.76
8	0.81	0.84	0.79
9	0.79	0.81	0.83

TABLE V
AUC results for small-color-distortion detection method using different kernel widths

Kernel width	**AUC**
3	0.64
33	0.77
59	**0.78**
85	0.77

E. Text-Distortion Detection

We evaluated internal similarity metrics and thresholds for crop area and detector confidence. Table VI shows the AUC scores for different metrics, while Table VII highlights the effect of varying thresholds. The best AUC score of 0.90 was achieved with FSIM, a crop area threshold of 300, and a confidence threshold of 0.7.

TABLE VI
AUC results for text-distortion detection method using different internal metrics.

Internal Metric	**AUC**
FSIM [3]	**0.90**
PSNR	0.86
SSIM	0.85
VIF(P) [4]	0.60
MS-SSIM	0.52
IW-SSIM [5]	0.51
NLPD [6]	0.52

TABLE VII
AUC results for text-distortion detection method using different detector confidence thresholds and crop area thresholds.

Crop Area \ **Confidence**	**0.6**	**0.7**	**0.8**
100	0.85	0.87	0.87
200	0.86	0.87	0.87
300	0.88	**0.90**	0.86
400	0.88	0.88	0.83
600	0.80	0.80	0.76

References

[1] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003.

[2] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[3] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, pp. 2378–2386, 2011.

[4] Hamid R. Sheikh and Alan C. Bovik, "Image information and visual quality," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004.

[5] Zhou Wang and Qiang Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, 2011.

[6] Valero Laparra, Johannes Ballé, Alexander Berardino, and Eero P Simoncelli, "Perceptual image quality assessment using a normalized laplacian pyramid," *Electronic Imaging*, 2016.