MaskSim: Detection of synthetic images by masked spectrum similarity analysis

Supplementary Material

1. Detailed visualization

We further visualize the masks and spectrum references of our proposed method trained on different classes of synthetic images in Fig. 6. The detection of different classes of synthetic images depends on different combinations of frequencies which meanwhile share certain informative frequencies. For instance, the masks and spectrum references for Stable Diffusion family, Midjourney and Firefly images having the similar grid of peak values show the importance of the 8- and 16-period frequency components for their detection, while the detection of DALL·E images rely on certain peak frequency components at both axes. In addition to the common peak frequencies, each class of images requires its own distinct subset of frequencies for detection.

We also visualize the average spectra of pristine images of Raise dataset [15] in Fig. 7 and the average spectra of different classes of synthetic images in Fig. 8. The images for each spectrum have undergone different post-compressions by JPEG, and have been preprocessed by the DnCNN denoiser of the model trained for detecting a specific class of synthetic images. Except for DALL·E 2, each synthetic average spectrum shows a similar regular grid of peak values, while the same grid is also present in the pristine average spectra. When zooming in, the peaks at the regular grid are clearer when the JPEG compression is stronger, due to the fact that JPEG compression is processed on 8x8 and 16x16 blocks and leave the similar artifacts at the 8- and 16-period frequency components.

2. Comparison with ResNet-50

A further comparison was performed between our proposed architecture and ResNet-50 [36] which is one of the most popular architectures for synthetic image classification. We trained the ResNet-50 classifier in the same data scheme using the pre-trained weights for classification task on ImageNet [16]. Similarly to what was done for our method, we trained one ResNet-50 detector for each class of synthetic images with the same data augmentation, and evaluated the performance of the SD-2 detector, its generalization ability of merged detector and its generic detection ability at different compression quality factors. The average performances presented by AUC over all the tested classes of images are shown in Tab. 5.

As can be seen, our proposed architecture is generally more effective than ResNet-50 detector except for the generalized performance for JPEG-compressed images compressed at quality factor 70. This can be attributed to the fact that the ResNet-50 overfits easily on the used training set.

post-JPEG	method	SD-2	generalized	generic
None	ResNet-50	88.3	82.9	95.9
	ours	**90.9**	**89.5**	**98.3**
Q=90	ResNet-50	87.1	83.1	95.6
	ours	**90.3**	**87.5**	**97.9**
Q=80	ResNet-50	86.2	82.4	95.2
	ours	**87.8**	**84.0**	**96.6**
Q=70	ResNet-50	86.5	**82.4**	95.3
	ours	**86.6**	81.7	**95.5**

Table 5. The average AUC (%) over all the classes of synthetic images for our detection method and for the classifier based on ResNet-50. Both detection methods are trained, validated and tested in the same data scheme. Different post-JPEG compressions at quality factors Q=90, 80 and 70 have been applied to the tested images.

Also, our proposed method is able to explicitly discover the peak frequencies that contribute to the generalization ability for detecting different classes of synthetic images, while ResNet-50 composed of small convolution kernels can be less sensitive to these informative peak frequencies.

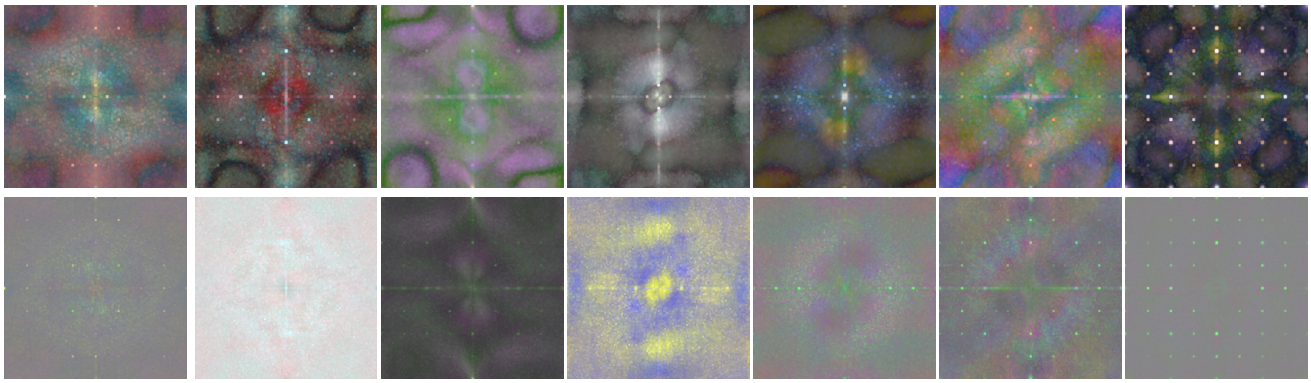

Figure 6. The masks (top) and spectrum references (bottom) of the proposed detection models trained on different classes of images compressed by JPEG at quality factors between 65 and 100.

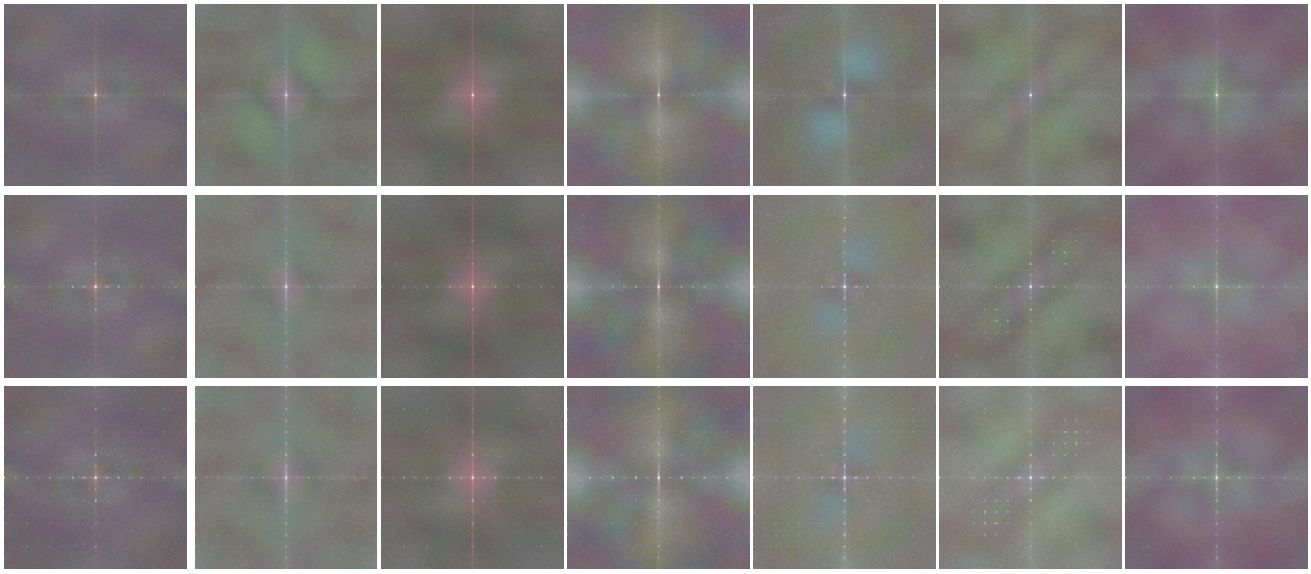

Figure 7. The average spectra of pristine images from Raise dataset. Each column shows the average pristine spectra after the preprocessing of the model trained on the corresponding class of synthetic images. The three rows show the average pristine spectra of uncompressed images (top) and images compressed by JPEG respectively at quality factors 90 (middle) and 70 (bottom).

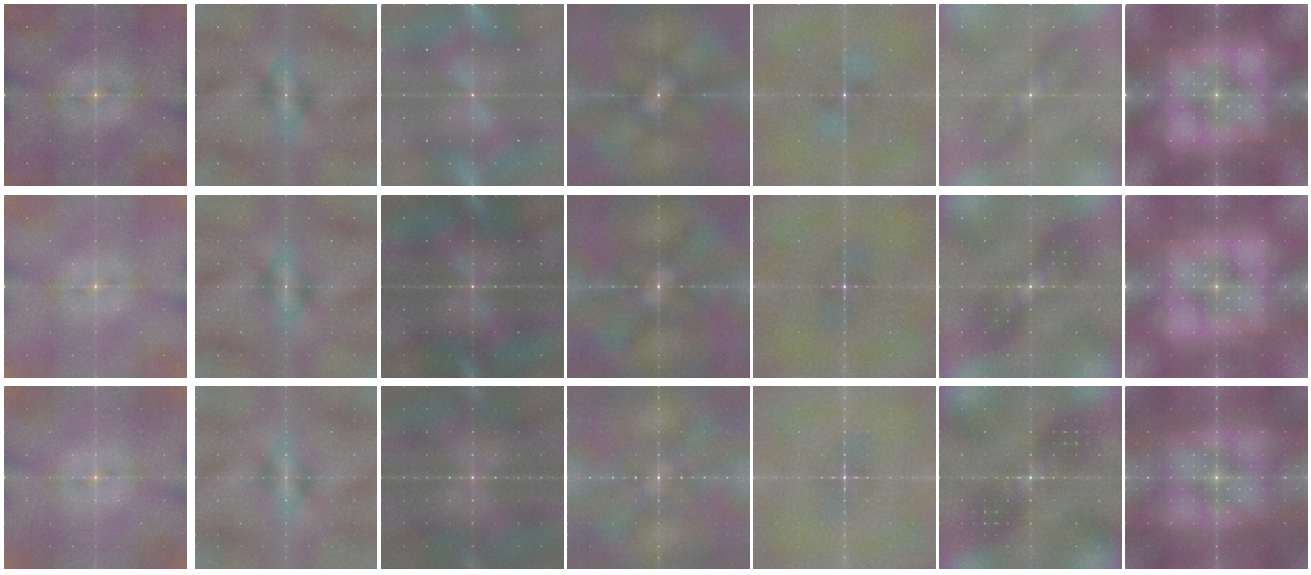

Figure 8. The average spectra of synthetic images of different classes. Each column shows the average spectra of a class of synthetic images after the preprocessing of the model trained on the same class of synthetic images. The three rows show the average spectra of unprocessed synthetic images (top) and synthetic images compressed by JPEG respectively at quality factors 90 (middle) and 70 (bottom).