Supplementary Material — Median Filtered Image Quality Enhancement and Anti-Forensics via Variational Deconvolution

Wei Fan, Kai Wang, François Cayre, and Zhang Xiong

This Supplementary Material contains some experimental results and method descriptions which are not included in the manuscript due to the page limit and for the sake of brevity. Here, we may employ some notations or references from the paper without explanation. Please refer to the manuscript for details.

1 ROC Curves of Different MF Forgeries Against Scalar-based Median Filtering Detectors

Figure 1 shows the ROC curves of different MF forgeries against the four scalar-based detectors [3]-[5]. Among all the median filtering anti-forensic methods in comparison, ours is able to bring the ROC curves the closest to the random guess.

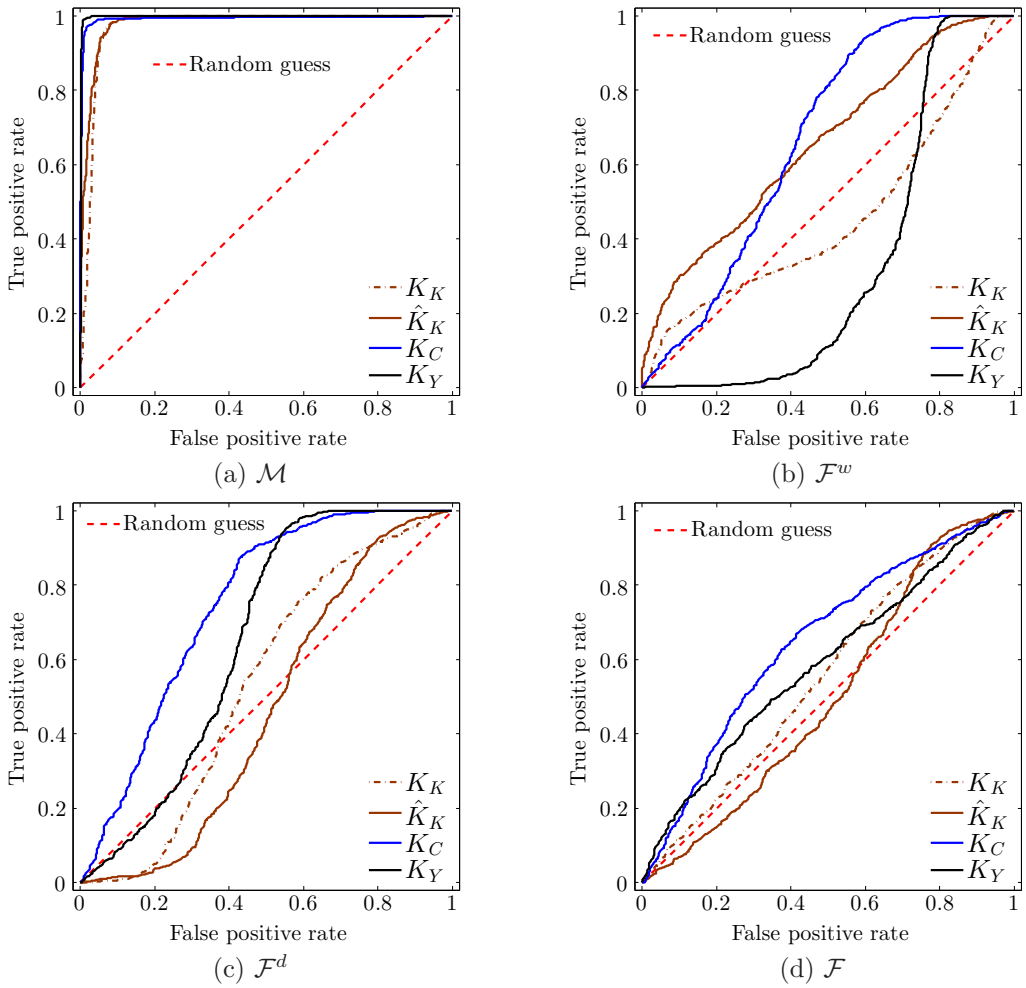


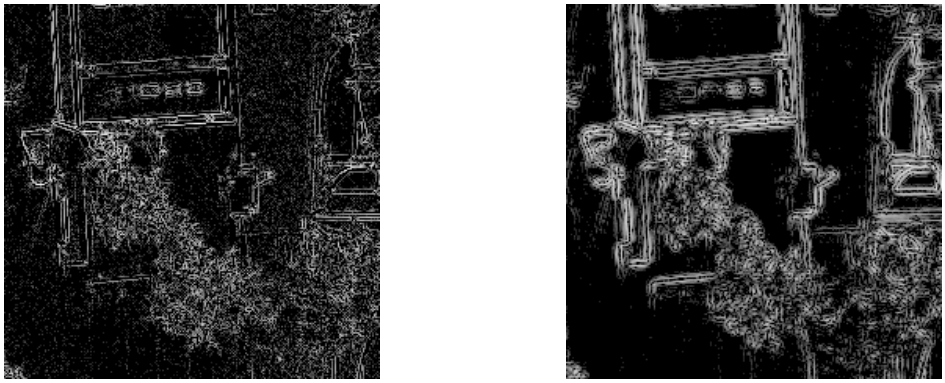
Figure 1: ROC curves of the MF image $\mathcal{M}$, Wu *et al.*'s MF forgery $\mathcal{F}^w$ [14], Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15], and our MF forgery $\mathcal{F}$, against the four scalar-based median filtering detectors K_K [3], $\hat{K}_K$ [3], K_C [4], and K_Y [5]. Results are obtained on MFTE dataset.

2 Example Results of the Proposed Median Filtering Anti-Forensic Method

Figure 2-(c) is an example MF forgery $\mathcal{F}$ created using the median filtering anti-forensic method proposed in the paper. With comparison to the MF image shown in Figure 2-(b), we can see that the image quality has been improved and the image blurring has been mitigated. Besides, Figure 2-(d) illustrates the ground-truth difference image between the original image and the MF image. Figure 2-(e) shows the difference image of our MF forgery with respect to the MF image. The two difference images -(d) and -(e) are rather visually similar. It can also be seen that the proposed method, to some extent, is able to reproduce the noise-like pattern of the median filtering difference, especially in the textured areas.

Even without the pixel value perturbation, another version of our MF forgery $\mathcal{F}'$ is also capable of creating similar noise-like pattern in the difference image with respect to the MF image. We refrain from showing the corresponding result due to the fact that it is hard to check the difference between $|\mathcal{F}' - \mathcal{M}|$ and $|\mathcal{F} - \mathcal{M}|$ by human naked eyes. This can also be inferred from the very small image quality metric value differences between $\mathcal{F}'$ and $\mathcal{F}$ (see the Table III of the manuscript). Moreover, the merit of the proposed pixel value perturbation can be found in the further improvement of median filtering artifact hiding ability and pixel value difference histogram restoration capability (see the Table III of the manuscript).

(a) $\mathcal{I}$ (b) $\mathcal{M}$ (c) $\mathcal{F}$



(d) $|\mathcal{I} - \mathcal{M}|$ (e) $|\mathcal{F} - \mathcal{M}|$

Figure 2: Example results (close-up images) of an MFTE image. For a better visibility of the difference images shown in (d) and (e), we have taken logarithm of the pixel value differences and afterwards carried out a normalization.

3 De-autocorrelation

As pointed out in the manuscript, in an optimistic testing scenario (the SVM-based detectors are trained on original and MF images) for median filtering anti-forensics, different kinds of MF forgeries have good forensic undetectability against the SVM-based detectors K_{SPAM}^{S686}, K_{MFF}^{S44}, K_{GLF}^{S56}, and K_{LTP}^{S220}. However, $\mathcal{F}^w$, $\mathcal{F}^d$ and $\mathcal{F}$ all fail to fool K_{AR}^{S10} [11] when the replacement rate is high (see the Fig. 7 of the manuscript). In this section, we propose to further perform a so-called de-autocorrelation operation to $\mathcal{F}'$, in order to defeat K_{AR}^{S10} [11] in this testing scenario.

The autoregressive model adopted by K_{AR}^{S10} [11] exploits the relations between median filter residual neighbors, which in general are more correlated to each other after the image is median filtered. It is known that the autoregressive coefficients are related to the autocorrelation values via the well-known Yule-Walker equations. We conducted experimental study and find that the autocorrelation values of the median filter residual increase after the image is median filtered. It is hard to directly regularize the autoregressive coefficients, however relatively easy to minimize the autocorrelation values. Hence, we propose to minimize the autocorrelation values of the median filter residual of $\mathcal{F}'$ in order to fool the detector K_{AR}^{S10}. Given an image to be processed, the median filter residual is computed and each column and row of which is processed using the proposed de-autocorrelation operation. At last, the MF forgery is generated by subtracting the median filtered version of $\mathcal{F}'$ by the processed median filter residual (see the Eq. (6) of the manuscript).

More specifically, for an $n \times 1$ median filter residual vector $\mathbf{g}$ extracted from a column or a row of the median filter residual matrix of $\mathcal{F}'$, we hope to minimize its first N autocorrelation values by:

$$\tilde{\mathbf{g}} = \arg\min_{\mathbf{g}} \sum_{j=1}^{N} w_j \left| \sum_{i=1}^{n-j} \mathbf{g}_i \mathbf{g}_{i+j} \right|, \tag{1}$$

where w_j is the weight balancing different autocorrelation values. The above optimization problem can be solved using the widely used subgradient method. In practice, we set $N = 6$ as we observe that the first 6 entries of the MFRAR feature [11] have relatively big differences between the original and the MF images. Moreover, we empirically set the weights $\{w_j\}$ for $j = 1, 2, \cdots, 6$ as 0.5, 1, 0.1, 0.5, 0.1, and 0.1, respectively. This gives us satisfying MF forgeries which can achieve a good forensic undetectability while slightly decreasing the image quality of the resulting image $\mathcal{F}''$ (see the last paragraph of Sec. IV-D and Fig. 7 of the manuscript for the relevant results). We leave the thorough analysis and validation of this preliminary de-autocorrelation algorithm as a future effort.

4 Some Experimental Results of Fontani and Barni's Method

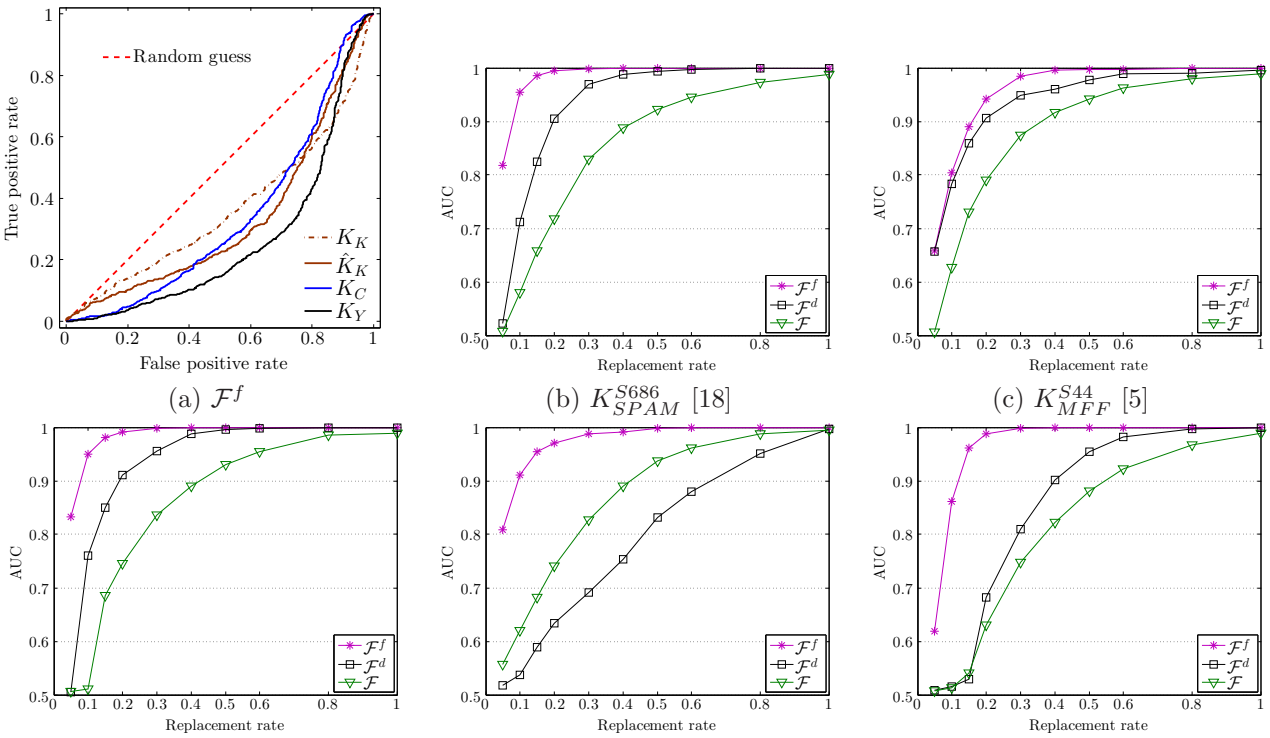
In [15], Dang-Nguyen *et al.* have demonstrated that their median filtering anti-forensic method outperforms Fontani and Barni's pioneer method [13]. Hence, we refrain from the comparison with Fontani and Barni's method in the manuscript. However, for completeness consideration, in this Supplementary Material we still present some experimental results of Fontani and Barni's median filtering anti-forensic method obtained on MFTE dataset. During our experimental simulation, we follow the algorithm description in [13] for implementing the anti-forensic method. Yet, please note that there might exist certain difference between our implementation and Fontani and Barni's [13], possibly leading to (slightly) different experimental results.

For the sake of brevity, let $\mathcal{F}^f$ denote the MF anti-forensic image created from the MF image $\mathcal{M}$ using Fontani and Barni's method. Table 1 reports the performance comparison of Fontani and Barni's MF forgery $\mathcal{F}^f$, Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$, and our MF forgery $\mathcal{F}$. The evaluation metrics include image quality, anti-forensic performance against four scalar-based detectors, as well as KL divergence of pixel value difference histograms between the original images and the MF anti-forensic images. Figure 3-(a) shows the ROC curves achieved by $\mathcal{F}^f$ when tested against the scalar-based detectors. From Figure 3-(b) to -(g), the AUC values achieved by $\mathcal{F}^f$ at different image replacement rates when tested against the SVM-based detectors are plotted.

From Table 1, it can be seen that both $\mathcal{F}^d$ [15] and $\mathcal{F}$ outperform Fontani and Barni's MF forgery $\mathcal{F}^f$ [13] in terms of average PSNR and SSIM values on MFTE dataset. Comparing Figure 1-(d) and Figure 3-(a), in general the proposed median filtering anti-forensic method is able to move the ROC curves of the scalar-based detectors closer to the random guess line than Fontani and Barni's method [13]. The experimental results shown from Figure 3-(b) to -(g) indicate that, $\mathcal{F}^f$ performs the worst when tested against the SVM-based detectors. Though the last four columns in Table 1 show that $\mathcal{F}^f$ is able to achieve the lowest KL divergence values (which however do not necessarily imply a good forensic undetectability against existing detectors), it is of lower image quality and can be more easily detected by both scalar-based and SVM-based detectors compared with $\mathcal{F}^d$ and $\mathcal{F}$.

Table 1: From the 2nd to the 5th columns, the average PSNR and SSIM values are reported. The following 4 columns show the AUC of different kinds of images against the 4 scalar-based detectors [3]-[5]. The average KL divergence values of the pixel value difference histograms between the original images and the MF anti-forensic images are listed in the last 4 columns. Results are obtained on MFTE dataset.

	Image quality				Anti-forensic performance				KL divergence			
	oPSNR	oSSIM	mPSNR	mSSIM	K_K [3]	$\hat{K}_K$ [3]	K_C [4]	K_Y [5]	$\mathbf{f}^1$	$\mathbf{f}^2$	$\mathbf{f}^3$	$\mathbf{f}^4$
$\mathcal{F}^f$ [13]	30.1129	0.9638	32.3215	0.9776	0.3626	0.3310	0.3389	0.2609	0.0287	0.0319	0.0158	0.0165
$\mathcal{F}^d$ [15]	33.4272	0.9714	36.4076	0.9871	0.5347	0.4635	0.7479	0.6518	0.0547	0.0563	0.0383	0.0389
$\mathcal{F}$	37.5184	0.9901	38.8653	0.9898	0.5595	0.5061	0.6490	0.5886	0.0484	0.0449	0.0272	0.0238



(a) $\mathcal{F}^f$ (b) K^{S686}_{SPAM} [18] (c) K^{S44}_{MFF} [5]

(d) K^{S56}_{GLF} [8] (e) K^{S10}_{AR} [11] (f) K^{S220}_{LTP} [12]

Figure 3: (a) shows the ROC curves of Fontani and Barni's MF forgery $\mathcal{F}^f$ [13] against the four scalar-based median filtering detectors K_K [3], $\hat{K}_K$ [3], K_C [4], and K_Y [5]. (b)-(f) are results obtained by Fontani and Barni's MF forgery $\mathcal{F}^f$ [13], Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15] and our MF forgery $\mathcal{F}$ against the SVM-based detectors. Here, the AUC values are plotted as a function of the image replacement rate under the worst-case scenario for anti-forensics. Results are achieved on MFTE dataset. The SVM-based detectors are trained on MFTR dataset.

5 Some Experimental Results Obtained on BOSSBase Dataset

BOSSBase contains 10000 high-resolution images stored in RAW format. It is suggested to use the *UFRaw* utility to firstly transfer them into the PPM format, before they are further converted to 8-bit grayscale PGM images. For each original high-resolution grayscale BOSSBase image, a 512×512 sub-image is cropped from its center for forensic testing. Similar to the strategy in the manuscript (see the first paragraph of Sec. III), over-smoothed images which cannot yield valid outputs for K_K, $\hat{K}_K$, K_C, or K_Y are excluded. At last, we have 9466 512×512 8-bit grayscale images, among which 1499 images are randomly selected and put into BBMFPE dataset for parameter estimation of the generalized Gaussian distribution for the image prior. The rest 8000 images constitute the BBMFTest dataset for forensic testing. In order to test the SVM-based median filtering forensic detectors, the BBMFTest dataset is further randomly split into two subsets: BBMFTR and BBMFTE, each of which contains 4000 images. BBMFTR is used to train the SVM-based detectors, while BBMFTE is for forensic testing of these SVM-based detectors.

The creation of Wu *et al.*'s MF forgery $\mathcal{F}^w$ [14] on BOSSBase dataset requires more computation time than we can afford. In [15], Dang-Nguyen *et al.* have shown that their method outperforms Fontani and Barni's [13]. It is again demonstrated by experimental results achieved on MFTE dataset in Section 4 in this Supplementary Material. Therefore, we refrain from comparing the performance of our MF forgery with theirs,

whereas the relevant results achieved on MFTE dataset can be found in the manuscript and in Section 4 of this Supplementary Material. In this section, we use the MF image $\mathcal{M}$ and Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15] for experimental comparison on BOSSBase dataset.

5.1 Scalar-based Detectors

In order to evaluate the performance of Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15] on BOSSBase dataset, we also need to properly tune the parameter T so that $\mathcal{F}^d$ has a good tradeoff between the forensic undetectability and the image quality. This parameter search is carried out on 100 randomly picked images from BBMFTest, varying $T \in \{1, 2, \cdots, 6\}$. In the end, we use $T = 4$ for generating Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15]. As to our quality enhanced MF image $\mathcal{M}^p$ and MF anti-forensic image $\mathcal{F}$, the parameter selection is performed in a similar way to that presented in the Sec. IV of the manuscript. In the end, for $\mathcal{M}^p$, we set $\omega = 0.7$, $\lambda = 5000$, $\gamma = 800$ and use the AVE kernel. Regarding $\mathcal{F}$, the AVEE kernel is adopted with $\omega = 0.1$, $\lambda = 1500$, and $\gamma = 500$.

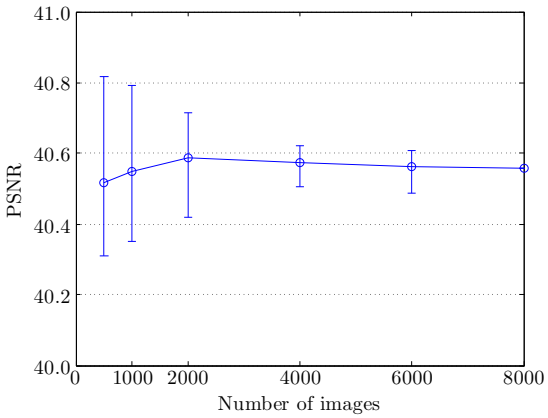
Please see Table 2 for the performance evaluation of different kinds of (quality enhanced) MF (anti-forensic) images, in terms of image quality, AUC values, and the average KL divergence values. It can be seen that, compared with the MF image $\mathcal{M}$, $\mathcal{M}^p$ is able to achieve 0.4286 dB of oPSNR gain and 0.0025 of oSSIM gain on average. It again proves the efficacy of the proposed framework for MF image quality enhancement. For our MF anti-forensic forgery $\mathcal{F}$, though its average oPSNR value is slightly lower than that of the MF image, the average oSSIM value still sees some improvement. Nevertheless, the average oPSNR value (40.5575 dB) of our forgery $\mathcal{F}$ is quite high (> 40 dB), implying a high visual resemblance between the original image and the anti-forensic image, and it is also noticeably higher than the oPSNR value of $\mathcal{F}^d$ [15]. Moreover, we can see that our forgery $\mathcal{F}$ has a better tradeoff between the forensic undetectability and the image quality than Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15].

Table 2: From the 2nd to the 5th columns, the average PSNR and SSIM values are reported. The following 4 columns show the AUC of different kinds of images against the 4 scalar-based detectors [3]-[5]. The average KL divergence values of the pixel value difference histograms between the original images and the (quality enhanced) MF (anti-forensic) images are listed in the last 4 columns. Results are obtained on BBMFTest dataset.

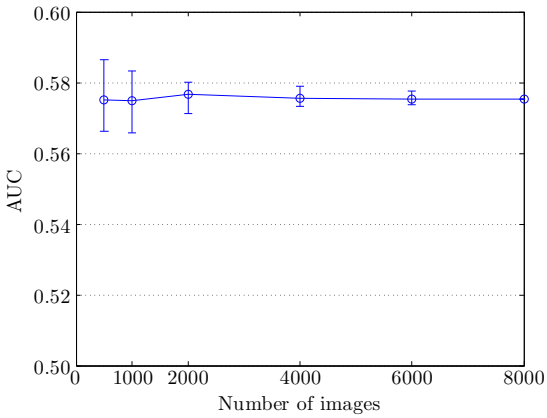
	Image quality				Anti-forensic performance				KL divergence			
	oPSNR	oSSIM	mPSNR	mSSIM	K_K [3]	$\hat{K}_K$ [3]	K_C [4]	K_Y [5]	$\mathbf{f}^1$	$\mathbf{f}^2$	$\mathbf{f}^3$	$\mathbf{f}^4$
$\mathcal{M}$	41.0530	0.9863	–	–	0.9820	0.9937	0.9417	0.9971	0.1957	0.1931	0.1225	0.1149
$\mathcal{M}^p$	41.4816	0.9888	51.2180	0.9985	0.9090	0.9715	0.7770	0.9338	0.1623	0.1584	0.1008	0.0945
$\mathcal{F}^d$ [15]	37.7061	0.9795	40.9854	0.9921	0.6297	0.4701	0.5203	0.6820	0.0961	0.0968	0.0774	0.0744
$\mathcal{F}$	40.5575	0.9891	43.3182	0.9934	0.5754	0.4624	0.4539	0.6543	0.0648	0.0628	0.0385	0.0357

It can be seen that it is relatively hard to improve the image quality of the MF image on BBMFTest dataset compared with MFTE dataset used in the manuscript. An underlying reason is that most images from BOSSBase are very smooth, which can also be reflected by the extremely high average oPSNR value (up to 41.0530 dB) of the MF image. One possible reason might be that the *UFRaw* utility for generating the PGM images from RAW files does not implement the sharpening operation in the camera processing pipeline, as mentioned by its authors (see `http://ufraw.sourceforge.net/` for details). The median filtering can be taken as an image smoothing operation, and here on BOSSBase dataset has relatively small impact to smooth images. Too much similarity between the MF image and the original image may weaken the performance of the proposed image variational deconvolution method.

In order to study the statistical significance of the number of images used in forensic testing, for each $K \in \{500, 1000, 2000, 4000, 6000, 8000\}$, K images are randomly picked from BBMFTest dataset and the average image quality and AUC values achieved by the scalar-based median filtering detectors are computed. For each K, this evaluation is repeated for 20 times. Hence, the mean, maximum, minimum, and standard deviation of different metrics over the 20 evaluations can be calculated. Figure 4-(a) shows the mean, maximum and minimum average oPSNR values, while -(b) plots the mean, maximum and minimum AUC values achieved by the detector K_K [3]. Tables 3-4 also list the mean and standard deviation pairs for different numbers of images and different metrics. We can see that the standard deviation slightly decreases as the number of images increases. It can also be observed that even when the number of testing images is relatively small, say $K = 1000$, the results are quite stable and reliable compared to those obtained from all the 8000 images of BBMFTest dataset. This can be more precisely reflected by the small standard deviations listed in Tables 3-4, and the small differences between the maximum and minimum values in Figure 4. Note that, although Figure 4 only shows the results for oPSNR and the AUC against the detector K_K [3], we actually have the same observation for other metrics, *i.e.*, the maximum and the minimum values are always close to each other.



(a) oPSNR



(b) K_K

Figure 4: (a) shows the mean, maximum and minimum average oPSNR values achieved for different numbers of images in consideration; whereas (b) plots the mean, maximum and minimum of the AUC value achieved by detector K_K. Note the scale of the y-axis in the figures, it can be seen that the maximum or the minimum values are in fact close to the mean.

Table 3: The mean average PSNR/SSIM value and its standard deviation over the 20 evaluations, for different numbers of images K. Results are obtained on BBMFTest dataset.

	oPSNR	oSSIM	mPSNR	mSSIM
$K = 500$	$(40.5177, 0.1699)$	$(0.9889, 0.0005)$	$(43.2835, 0.2273)$	$(0.9933, 0.0003)$
$K = 1000$	$(40.5481, 0.1176)$	$(0.9892, 0.0004)$	$(43.2975, 0.1279)$	$(0.9934, 0.0001)$
$K = 2000$	$(40.5874, 0.0879)$	$(0.9891, 0.0002)$	$(43.3578, 0.0978)$	$(0.9934, 0.0001)$
$K = 4000$	$(40.5737, 0.0305)$	$(0.9891, 0.0001)$	$(43.3305, 0.0485)$	$(0.9934, 0.0000)$
$K = 6000$	$(40.5631, 0.0311)$	$(0.9891, 0.0001)$	$(43.3182, 0.0325)$	$(0.9934, 0.0000)$
$K = 8000$	$(40.5575, \quad - \quad)$	$(0.9891, \quad - \quad)$	$(43.3182, \quad - \quad)$	$(0.9934, \quad - \quad)$

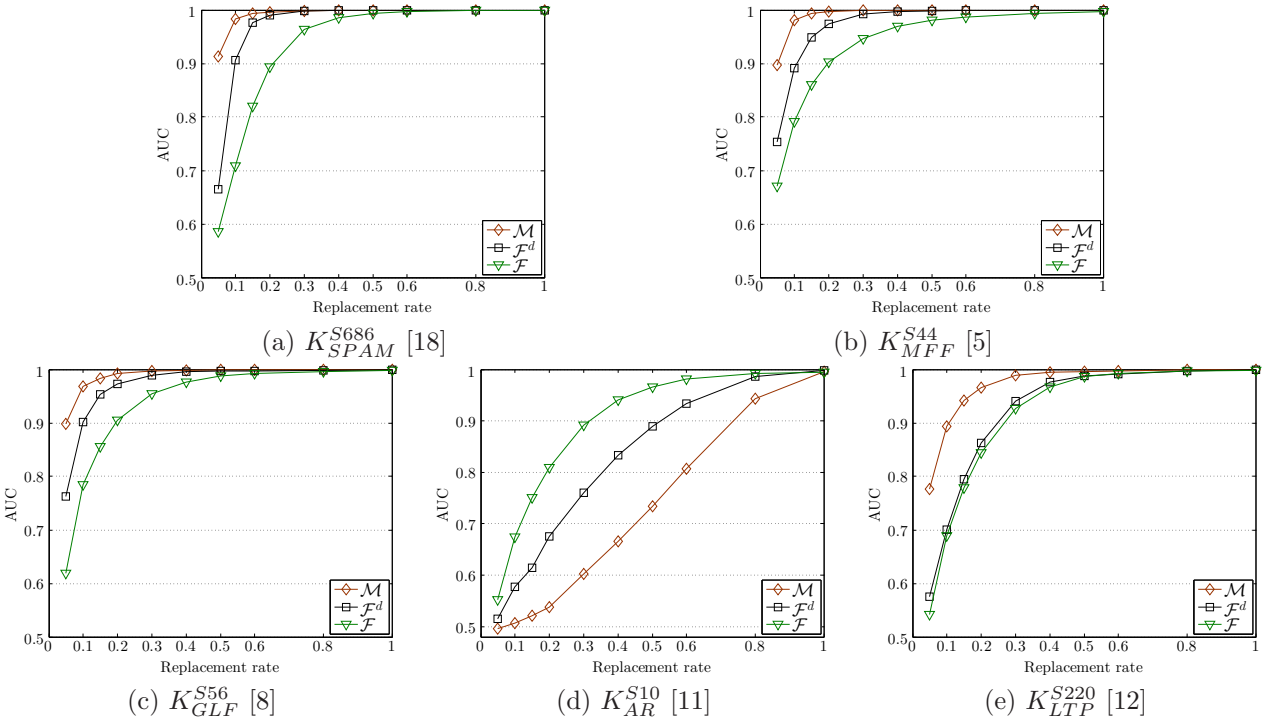
Table 4: The mean AUC value achieved by the scalar-based median filtering detectors [3]-[5] and its standard deviation over the 20 evaluations, for different numbers of images K and forensic detectors. Results are obtained on BBMFTest dataset.

	K_K [3]	$\hat{K}_K$ [3]	K_C [4]	K_Y [5]
$K = 500$	$(0.5752, 0.0052)$	$(0.4629, 0.0066)$	$(0.4541, 0.0096)$	$(0.6573, 0.0105)$
$K = 1000$	$(0.5749, 0.0048)$	$(0.4614, 0.0072)$	$(0.4541, 0.0050)$	$(0.6536, 0.0075)$
$K = 2000$	$(0.5767, 0.0024)$	$(0.4627, 0.0027)$	$(0.4539, 0.0026)$	$(0.6553, 0.0040)$
$K = 4000$	$(0.5755, 0.0016)$	$(0.4629, 0.0026)$	$(0.4536, 0.0023)$	$(0.6530, 0.0021)$
$K = 6000$	$(0.5752, 0.0009)$	$(0.4620, 0.0015)$	$(0.4542, 0.0013)$	$(0.6539, 0.0013)$
$K = 8000$	$(0.5754, \quad - \quad)$	$(0.4624, \quad - \quad)$	$(0.4539, \quad - \quad)$	$(0.6543, \quad - \quad)$

5.2 SVM-based Detectors

As mentioned earlier, in order to test the SVM-based median filtering forensic detectors, the 8000 images in BBMFTest dataset is randomly split into two subsets: BBMFTR and BBMFTE, each of which contains 4000 images. Similar to the worst-case testing scenario in the manuscript, Figure 5 shows the AUC values achieved by the MF image $\mathcal{M}$, Dang-Nguyen *et al.*'s MF forgery $\mathcal{F}^d$ [15], and our MF forgery $\mathcal{F}$ at different image replacement rates against different SVM-based detectors. The results shown in Figure 5 are to some extent similar to those shown in the Fig. 6 of the manuscript, which are obtained on MFTE dataset. Except for detector K_{AR}^{S10} [11], our MF forgery $\mathcal{F}$ outperforms $\mathcal{M}$ and $\mathcal{F}^d$ [15] against all the SVM-based detectors in consideration. The AUC values achieved by our MF forgery $\mathcal{F}$ are almost at the same level for the same replacement rate across different SVM-based detectors, indicating a stable performance of the proposed MF image anti-forensic method.

In order to study the statistical significance of the number of training images in the SVM-based forensic testing, we adopt a similar experimental setting to Section 5.1 of this Supplementary Material. For each $K \in \{500, 1000, 2000, 3000, 4000\}$, K images are randomly picked from BBMFTR dataset for training the SVM-based detectors, meanwhile K images are randomly picked from BBMFTE dataset for forensic testing. For each kind of SVM-based detector, each kind of image and each K, forensic tests are conducted, and



(a) K^{S686}_{SPAM} [18] (b) K^{S44}_{MFF} [5]

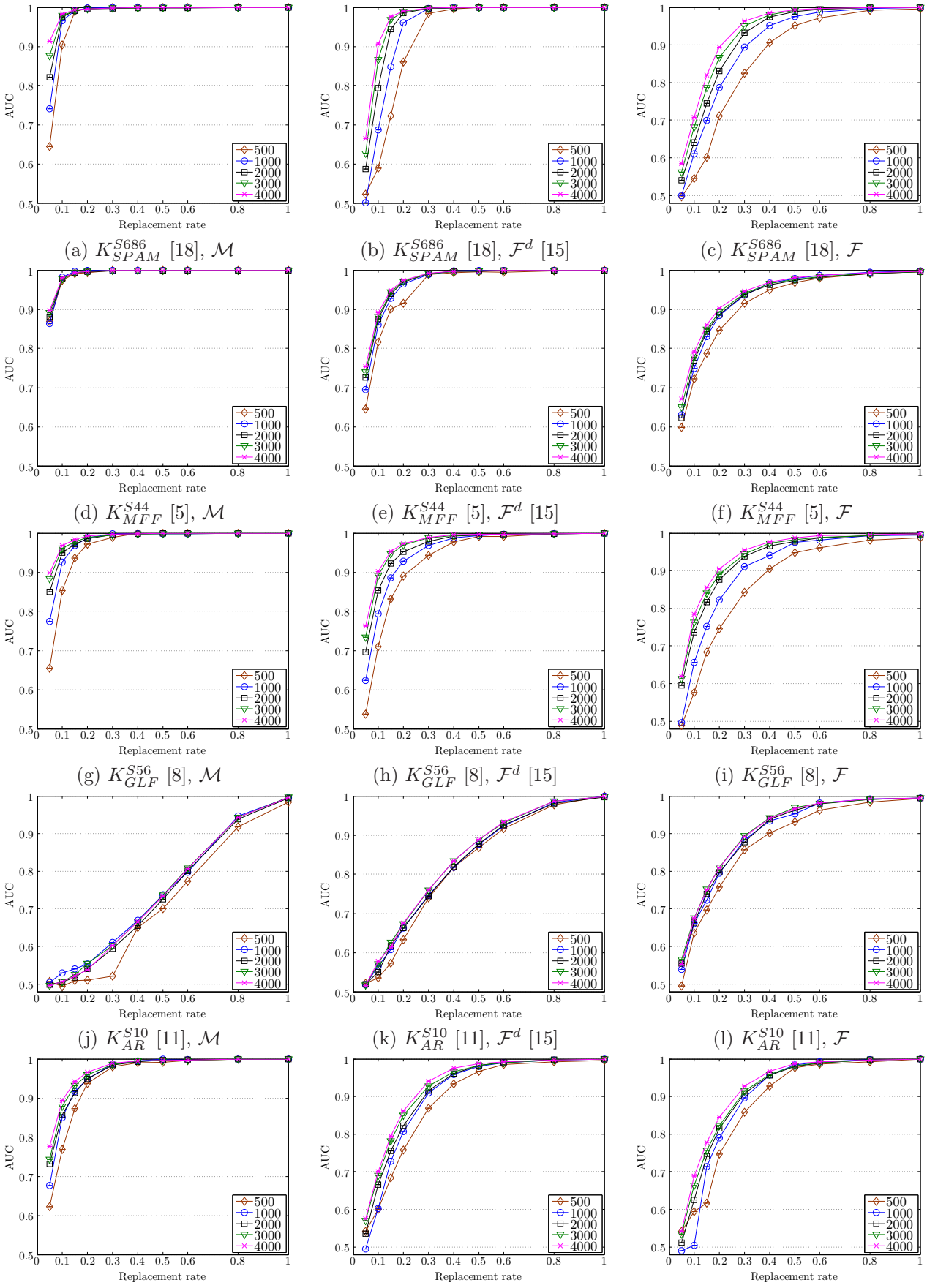
(c) K^{S56}_{GLF} [8] (d) K^{S10}_{AR} [11] (e) K^{S220}_{LTP} [12]

Figure 5: The AUC value as a function of image replacement rate for different kinds of images $\mathcal{M}$, $\mathcal{F}^d$ [15] and $\mathcal{F}$, when tested using the SVM-based detectors, under the worst-case scenario. Results are achieved on BBMFTE dataset, with detectors trained on BBMFTR dataset.

thereafter an AUC curve (in function of the image replacement rate) is plotted. In Figure 6, for each kind of image and each kind of SVM-based detector, five AUC curves are plotted in a sub-figure when varying the value of K. We can see that generally the forensic performance of the SVM-based detector is improved with increasing number of training images. Yet, there is no dramatic performance improvement when more training images are used, especially when the replacement rate is high. This can be partly explained by the fact that the SVM is by concept a classifier less sensitive to the problem of curse of dimensionality as pointed out in the classical textbook of Vapnik "The Nature of Statistical Learning Theory" on page 298.

The forensic performance variations of the SVM-based detectors are also different for different forensic features. Indeed, for K^{S686}_{SPAM} [18] and K^{S56}_{GLF} [8], the performance of the detector is improved by using more images for training the detector, especially when the replacement rate is low. By contrast, the performance of K^{S44}_{MFF} [5], K^{S10}_{AR} [11] and K^{S220}_{LTP} [12] barely changes with different numbers of images in use. Furthermore, no matter the number of images used, the comparison result between $\mathcal{F}^d$ and $\mathcal{F}$ remains the same, that is, $\mathcal{F}$ achieves a better forensic undetectability than $\mathcal{F}^d$ against all the considered SVM-based median filtering forensic detectors except K^{S10}_{AR}.

In all, we consider that there is a complex relation between SVM performance and various factors including not only the feature, the test image and the replacement rate, but also factors not considered here such as the type of kernel and the grid used for parameter search. The relevant theoretical analysis and experimental study, at least in the realm of image forensics, still remain an open research problem that deserves considerable future efforts.



(a) K^{S686}_{SPAM} [18], $\mathcal{M}$ (b) K^{S686}_{SPAM} [18], $\mathcal{F}^d$ [15] (c) K^{S686}_{SPAM} [18], $\mathcal{F}$

(d) K^{S44}_{MFF} [5], $\mathcal{M}$ (e) K^{S44}_{MFF} [5], $\mathcal{F}^d$ [15] (f) K^{S44}_{MFF} [5], $\mathcal{F}$

(g) K^{S56}_{GLF} [8], $\mathcal{M}$ (h) K^{S56}_{GLF} [8], $\mathcal{F}^d$ [15] (i) K^{S56}_{GLF} [8], $\mathcal{F}$

(j) K^{S10}_{AR} [11], $\mathcal{M}$ (k) K^{S10}_{AR} [11], $\mathcal{F}^d$ [15] (l) K^{S10}_{AR} [11], $\mathcal{F}$

(m) K^{S220}_{LTP} [12], $\mathcal{M}$ (n) K^{S220}_{LTP} [12], $\mathcal{F}^d$ [15] (o) K^{S220}_{LTP} [12], $\mathcal{F}$

Figure 6: Each sub-figure plots the AUC values as a function of image replacement rate for a certain kind of image against a certain SVM-based median filtering forensic detector, under the worst-case scenario. Different lines in a sub-figure correspond to different numbers of images in use in the BBMFTR dataset for training or in the BBMFTE dataset for testing.