Journal of Universal Computer Science, vol. 25, no. 9 (2019), 1151-1173
submitted: 12/8/18, accepted: 9/7/19, appeared: 28/9/19 © J.UCS

Designing Statistical Model-based Discriminator for Identifying Computer-generated Graphics from Natural Images

Mingying Huang
(School of Computer Science and Technology, Hangzhou Dianzi University
Hangzhou, China
hmycug@126.com)

Ming Xu
(School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China
mxu@hdu.edu.cn)

Tong Qiao[1]
(School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China
Zhengzhou Science and Technology Institute, Zhengzhou, China
tong.qiao@hdu.edu.cn)

Ting Wu, Ning Zheng
(School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China
wuting@hdu.edu.cn, nzheng@hdu.edu.cn)

Abstract: The purpose of this paper is to differentiate between natural images (NI) acquired by digital cameras and computer-generated graphics (CG) created by computer graphics rendering software. The main contributions of this paper are threefold. First, we propose to utilize two different denoising filters for acquiring the first-order and second-order noise of the inspected image, and analyze its characteristics with assuming that residual noise follows the proposed statistical model. Second, under the framework of the hypothesis testing theory, the problem of identifying between NI and CG is smoothly transferred to the design of the likelihood ratio test (LRT) with knowing all the nuisance parameters, and meanwhile the performance of the LRT is theoretically investigated. Third, in the practical classification, using the estimated model parameters, we propose to establish a generalized likelihood ratio test (GLRT). A large scale of experimental results on simulated and real data directly verify that our proposed test has the ability of identifying CG from NI with high detection performance, and show the comparable effectiveness with some prior arts. Besides, the robustness of the proposed classifier is verified with considering the attacks generated by some post-processing techniques.

Key Words: natural image, computer-generated graphic, digital image forensics, statistical noise model, hypothesis testing

Category: D.4.6

[1] Corresponding author

1 Introduction and contributions

In past few decades, the industry of rendering software has remarkably developed, for instance, Adobe Photoshop and Autodesk Maya with the ability of yielding stunningly computer-generated graphics (CG) very similar to the real object or scene. On the one hand, the technique of CG indeed enriches the human-being's daily life; on the other hand, the faked object or scene most possibly interferes our judgment for differentiating between natural images (NI) acquired by an imaging device and CG. Besides, it also results into both legal and scientific issues since the highly realistic CG might be used in the scenarios such as journalism, academic community, and even judicial trials. For example, a malicious attacker might generate a large-scale unrealistic images using a rendering software. The computer-generated graphics are possibly spread on the Internet, that serve as the faked evidence in the court, leading to misguided judicial judgment. In fact, this type of cybercrime to some extent threatens the reliability and authenticity of cyberspace. Hence, the classification between NI and CG remains one of the primary tasks in the forensic community. Thus, developing reliable methods with high accuracy for identifying CGs from actual photographs generated from digital cameras are necessary.

1.1 State of the art

Fortunately, digital image forensics is a possible technique to solve the proposed problem of classification. Digital image forensics is the technique of identifying the source of the obtained image (image origin identification) (see [Caldelli et al., 2017, Qiao et al., 2015a, Qiao et al., 2017, Yao et al., 2018, Qiao and Retraint, 2018]) or authenticating if the inspected image has been tampered (image content integrity) (see [Swaminathan et al., 2008, Birajdar and Mankar, 2013, Zhou et al., 2017, Qiao et al., 2018, Qiao et al., 2019, Zhao et al., 2019]). In the digital forensic community, steganography in fact is a type of tampering technique, in which the image content integrity is damaged by hidden secret messages. Some current studies have been done to consider steganalysis for detecting secret information (see [Luo et al., 2016, Ma et al., 2018, Zhang et al., 2018] for instance). In this context, it is proposed to focus on discrimination between NI and CG, belonging to the field of image origin identification.

In general, let us classify forensic methodologies into two categories: active forensics and passive forensics. Active forensics involves forensic techniques which authenticate a digital image by using prior-embedded relevant information after image acquisition, referring to as a digital watermark or signature [Potdar et al., 2005]. Unfortunately, the main drawback of active approaches requires strict coordination that any tampering may break the built-in information. On the contrary, passive techniques need no prior embedded information from an image,

instead focusing on the intrinsic features such as content textures, the statistical model of residual noise, and the algorithm of the color filter array (CFA) (see [Qiao et al., 2013, Gallagher, 2005]). Relying on the characteristics of the image acquisition pipeline, many passive methodologies have been proposed to deal with the forensic problem, which also inspires us to distinguish between NI and CG in this context.

In [Lyu and Farid, 2005] Lyu et al. extracted the first four order statistics involving mean, variance, skewness, and kurtosis, which are computed from the wavelet coefficients after discrete wavelet transform (DWT), where samples are trained as features, together with the machine-learning mechanism such as support vector machine (SVM). Lyu et al. opened the way of designing a typical DWT statistical discrimination model consisting of first and higher order wavelet statistics for classifying between NI and CG. On the basis of the scheme [Lyu and Farid, 2005], In [Ozparlak and Avcibas, 2011] Mader et al. extracted the statistics from ridgelet and contourlet wavelet transform (CWT), which capture more useful information for classifying between NI and CG. The method proposed in [Wang et al., 2016] overcame the drawbacks of DWT and CWT, and improved the identification accuracy. Accordingly, wavelet coefficients always play an important role of designing an effective classifier, which also inspires us, in this context, to utilize the wavelet-related algorithm for dealing with the problem of feature extraction.

In particular, the traces left by the procedure of demosaicing can also be used to design the discriminator. For instance, for detecting traces of demosaicing within NI, in [Gallagher and Chen, 2008] Gallagher et al. proposed to apply Fourier analysis to the image after high pass filtering, for capturing the presence of periodicity in the variance of interpolated coefficients, and then designed the classifier by using the peak value presented in the transformed domain. However, in the practical classification, since the peak value is possibly interfered with the texture of the inspected image, the robustness of the algorithm proposed in [Gallagher and Chen, 2008] cannot be guaranteed. In [Qiao et al., 2013] Qiao et al. proposed to use the property of the residual noise characterized by its statistical parameters, referring to as expectation and variance. With the help of the hypothesis testing theory, the designed mechanism of classification indeed improves the detection accuracy. However, the statistical performance of the discriminator is not theoretically analyzed, resulting in that we cannot obtain the theoretical upper bound of detection rate. In [Peng et al., 2017] Peng et al. found that multi-fractal spectrum can not only represent the overall texture feature of an inspected image, but also describe the local texture feature. In general, the value of the multi-fractal spectrum curve corresponds to the dimension of fractal sets surrounded by different precision value. The larger the precision value is, the more complex the texture of the image is. In virtue of the typical procedure

of acquiring CG, referring to as modeling, rendering, and illumination, Peng et al. assumed that the CG cannot thoroughly imitate the complicated contour (or the texture feature) of natural scenes. Thus its surface is generally smoother than that of NI. Besides, in current research, a very intriguing algorithm has been proposed in [Mader et al., 2017]. Discarding any design of discriminator, the authors found that human-being observer performance on differentiating CG from NI can be significantly improved with the proper training, feedback, and incentives.

Generally, extracted features are usually trained for generating a typical discriminator, referring to as supervised algorithms. Alternatively, those discriminative features can also be used for directly classifying, or designing some statistical models, and then the model-based statistical detectors are designed, referring to as un-supervised algorithms. For clarity, in the following discussion, let us primarily generalize both advantages and disadvantages of two categories:

- **Supervised algorithms**: In the digital forensic community, almost all the supervised algorithms are designed based on the machine learning mechanism, specifically SVM [Lyu and Farid, 2005, Ozparlak and Avcibas, 2011, Wang et al., 2016, Peng et al., 2017, Long et al., 2017] or current-hot CNN [De Rezende et al., 2018]. The discrimination between extracted features, obtained from a large scale of training samples, can be expressed in the form of classifiers. These classifiers are then used to distinguish between NI and CG. The effectiveness of the characteristics describing the corresponding type of images might determine the overall performance of the established discriminator. To our knowledge, the supervised algorithms always dominate the study of designing the classifier. However, the challenging problems of the supervised algorithms are that high dimensions of features might not be efficient during the stage of training the model, especially dealing with a large scale of samples. In addition, the performance based on supervised algorithms can only be empirically investigated mainly relying on a given validated dataset, not be analytically studied using a statistical model.

- **Un-supervised algorithms**: The methods in this category refer to the differentiation between NI and CG using statistical features or statistical model (see [Qiao et al., 2013, Mader et al., 2017, Gallagher and Chen, 2008, Dirik and Memon, 2009]), and it mainly focuses on the intrinsic features generated from the image acquirement procedure. For instance, the traces generated by CFA interpolation serving as the typical features can be used for representing the unique characteristic of NI in the frequency domain (see [Gallagher and Chen, 2008]) or describing the statistical model to differentiate between NI and CG (see [Qiao et al., 2013]). In general, the unique characteristic of NI can be directly used to design

an effective discriminator without the training stage. Thus, the efficiency (or computation cost) of the algorithm can be improved compared to supervised algorithms. It should be noted that the main advantage of the designed detectors in this category lies in the theoretical explanation for the principals of the algorithms whose validation is not only evaluated using empirical results. In addition, few current algorithms focus on investigating un-supervised algorithms based on the statistical model, or theoretically studying the performance of the designed discriminator. To fill the gap, let us establish a typical statistical model-based discriminator for distinguishing between NI and CG.

1.2 Contributions of the paper

In this paper, we extract the residual noise of an image to establish the statistical model, and then design the Likelihood Ratio Test (LRT) and the Generalized Likelihood Ratio Test (GLRT) based on the hypothesis testing theory. The main contributions of this paper can be summarized below:

1. By removing the disturbance caused by pixels' heterogeneity (of the property of image texture), it is proposed to devise a typical filter to extract the first-order residual noise in the spatial domain. Then by using a regression parametric model, the second-order noise, empirically following the Gaussian distribution, can be successfully expressed in the frequency domain.

2. In an ideal scenario, where all the nuisance model parameters are perfectly known, the optimal LRT is designed, and we mainly analyze its statistical performance. The advantage of the LRT is that it can easily serve as an upper bound of the detection power for discriminating between NI and CG images.

3. In a practical scenario, in which parameters of the proposed model remain unknown, we first develop the algorithm to predict the concerning parameters. Then in the case of adopting the estimated model parameters, the practical GLRT is established. Also, its statistical performance can be analyzed and applied to our practical classification between NI and CG.

4. Solid experimental results show the sharpness of the theoretically established LRT and the good performance of the practically designed GLRT. Besides, in comparison with current detectors, our proposed detector performs the comparable relevance. In addition, the robustness of the proposed classifier can be verified with considering the attacks generated by some post-processing techniques.

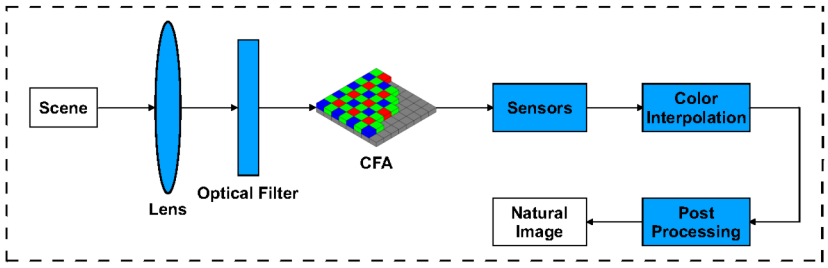

Figure 1 : Illustration of an imaging process in a digital still camera.

1.3 Organization of the paper

The rest of the paper is organized as follows. The image features extraction is investigated in Section 2, meanwhile, the first-order and second-order noise are extracted. In virtue of the proposed statistical model studying from the residual noise, the optimal detection algorithm for classification is established in Section 3. Then in Section 4, the practical detection scheme for discrimination is presented. Subsequently, the numerical experimental results are provided in Section 5. Finally, both conclusion and discussion are given in Section 6.

2 Image features extraction

2.1 Description of CFA interpolation

Current methodologies mainly rely on the extracted fingerprints generated from the stage of the image acquisition pipeline. Hence, let us first generally illustrate the typical image acquisition pipeline in a digital still camera (see Fig. 1). The light spectrum from the scene permeates through the optical system involving for instance, the lens and the optical filter. Portions (or useful parts) of the original light, which is controlled by the CFA pattern, are finally recorded by the camera sensors. Most digital cameras use a CFA pattern to filter the light ray from the real-world scene. The CFA pattern consisting of three color channels samples only one of three primary color components, referring to as red, green, and blue, at a corresponding channel photosite. These sparsely sampled color values are named as mosaic (or non-interpolated) images. To re-construct a full-color image from a mosaic image, the algorithm of CFA interpolation is required to pad two missing color intensities for each pixel photosite. Typically, CFA interpolation, also defined as demosaicing algorithm, is usually applied to the mosaic image by estimating the unrecorded pixel values. Then, after other post-processing stages (e.g. white balancing and gamma correction), a natural image can be obtained. In our establishment of the discriminator, the first step is to extract the residual noise in virtue of the unique characteristic of the CFA interpolation. Therefore,

in the following investigation, let us first extract the first-order noise from the image based on the characteristic of the CFA interpolation.

2.2 Extraction of first-order noise

In general, to deal with NI, the changes of the interpolation algorithm unavoidably arise some forensic traces that can be reliably detected. Besides, the interpolated pixels of NI characterize the unique features, which cannot be carried on by CG. Let us first design a typical high-pass filter to obtain the first-order noise characterizing the CFA pattern features. In this context, we propose to use Bayer model (one type of the most adopted CFA pattern) as our CFA pattern. Without the loss of generality, our proposed discriminator can be smoothly extended to other pattern.

Specifically, the first-order noise extraction can be processed as follows. First, we select only the green color channel of the given image [2], then the gray-level image $\mathbf{I}$ is convolved with a high-pass filter in order to extract the first-order noise representing the details of image. Because the detailed information can describe the CFA feature better than that of the original NI. In addition, the periodicity of the green channel of NI can be exposed. However, by proceeding the same operation, CG do not carry that typical feature, such as periodicity. In this context, we propose to design three different high-pass filters for first-order noise extraction.

Paradigm one: The image $\mathbf{I}$ is convolved with a designed typical high-pass operator, which is formulated as:

$$\mathrm{H} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

where $\mathbf{I}$ denotes the pixel intensity of the green color channel. Using that high-pass filter, the differences between the central element (or pixel intensity in this context) and its four neighboring elements are enlarged. The residual image primarily representing the noise can characterize the more NI features caused by CFA interpolation than the original NI without filtering.

Paradigm two: Also, we can design the second denoising filter by using a directional filter bank used in [Holub and Fridrich, 2013]. We utilize a set of three linear shift-invariant filters represented by the kernels $\mathcal{D} = \{\mathbf{K}^{(e)}\}, e \in \{1, 2, 3\}$. They can be used to evaluate the smoothness of a given image $\mathbf{I}$ along the horizontal, vertical, and diagonal directions by computing the so-called directional residual noise $\mathbf{W}^{(e)} = \mathbf{K}^{(e)} \star \mathbf{I}$, where the symbol "$\star$" denotes

[2] In this context, we use the word "image" to denote a natural image acquired by a camera device or a graphic generated by rendering software.

a mirror-padded convolution guaranteeing $\mathbf{W}^{(k)}$ has the same size with original image $\mathbf{I}$. In this context, it is proposed to exclusively use kernels $\mathcal{D}$ built from one-dimensional 16-tap Daubechies wavelet decomposition filters $\mathbf{l}$ and $\mathbf{h}$:

$$\mathbf{K}^{(1)} = \mathbf{l} \cdot \mathbf{h}^T, \;\; \mathbf{K}^{(2)} = \mathbf{h} \cdot \mathbf{l}^T, \;\; \mathbf{K}^{(3)} = \mathbf{h} \cdot \mathbf{h}^T. \tag{1}$$

In this case, the filters correspond to two-dimensional vertical LH, horizontal HL, and diagonal HH wavelet directional high-pass filters respectively. Then the residual images $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{W}^{(3)}$ of $\mathbf{I}$ coincide with the wavelet vertical LH, horizontal HL, and diagonal HH directional decomposition respectively.

Paradigm three: Still, it is proposed to utilize the wavelet denoising filter. The noise-free image and noise have different statistical characteristics after wavelet transform. Specifically, the main energy of the image itself corresponds to the large wavelet coefficients, but the remaining energy (or noise) corresponds to the small wavelet coefficients. Based on the assumption, we can set an appropriate threshold of wavelet coefficients for distinguishing between the main and the remaining energy. The value of wavelet coefficients larger than the threshold is considered to be a useful signal while that of wavelet coefficients smaller than the threshold refers to noise.

In the first scale of wavelet decomposition, the denoising operation for 3 high-pass subbands can not extract the noise completely. Thus, still, the low-pass subband (LL) in the first scale needs to be processed using the wavelet decomposition. It is proposed to employ Daubechies 8-tap wavelet decomposition with the whole 4 scales, which have been empirically effective in our noise extraction. Unlike Paradigm two, Paradigm three utilizes wavelet decomposition extracting multiple high-pass subbands in different scales, resulting in the more decomposed noise.

Due to that the high frequential component of NI involves the traces of demosaicing introduced by CFA interpolation, the pixels of image with the typical interpolation algorithm can exhibit high correlation. Although the extracted first-order residuals are capable of exposing the differences between NI and CG, those discriminations cannot help us design an effective classifier. Because the residual noise still contains some remnants, dependent of the edges and complex texture regions. Therefore, in the next subsection, we further conduct the second-order filtering.

2.3 Extraction of second-order noise

Generally, to design the statistical model-based discriminator, it is proposed to transform the first-order noise from the spatial domain to the frequency domain. To this end, the mean of the first-order noise along each diagonal is estimated. Then its frequential representations can be obtained by using Fast

Fourier Transform (FFT). Finally, the one-dimensional noise as a vector can be obtained (see [Qiao et al., 2013] for details). In the transformed domain, the second-order noise can be extracted by using a regression parametric model. It is proposed to use the least square algorithm to deal with the one-dimensional noise. In the frequency domain, non-overlapping channel $\mathbf{z}_k = \{z_{k,j}\}$ of the random variables (or noise) were partitioned as detailed in Section 5.2, the channel index $k \in \{1, ..., K\}$. The following polynomial representation predicts the estimation of element $z_{k,j}$, $\forall j \in \{1, ..., m\}$, of vector $\mathbf{z}_k$,

$$\widehat{z}_{k,j} = a_{k,0} + a_{k,1} \cdot x_{k,1} + a_{k,2} \cdot x_{k,2}^2 + \ldots + a_{k,n} \cdot x_{k,m}^n, \tag{2}$$

where $\{a_{k,0}, a_{k,1}, \ldots, a_{k,n}\}$ expresses the parameter vector of the regression model, $\{1, x_{k,1}, \ldots, x_{k,m}^n\}$ denotes the in-order variable vector, and j represents the index of random variables. It should be noted that the $z_{k,j}$ represents one sample of $\mathbf{z}_k$.

In practice, Eq. (2) could be expressed by using the following formulation:

$$\mathbf{z}_k = \mathbf{X} \cdot \begin{pmatrix} a_{k,0} \\ a_{k,1} \\ \vdots \\ a_{k,n-1} \\ a_{k,n} \end{pmatrix}, \quad \mathbf{z}_k = \begin{pmatrix} z_{k,1} \\ z_{k,2} \\ \vdots \\ z_{k,m-1} \\ z_{k,m} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{k,1} & x_{k,1}^2 & \ldots\ldots & x_{k,1}^{n-1} & x_{k,1}^n \\ 1 & x_{k,2} & x_{k,2}^2 & \ldots\ldots & x_{k,2}^{n-1} & x_{k,2}^n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{k,m} & x_{k,m}^2 & \ldots\ldots & x_{k,m}^{n-1} & x_{k,m}^n \end{pmatrix} \tag{3}$$

Then, in virtue of least square algorithm, we can estimate the parameters of polynomial fitting model by:

$$\begin{pmatrix} a_{k,0} \\ a_{k,1} \\ \vdots \\ a_{k,n-1} \\ a_{k,n} \end{pmatrix} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{z}_k. \tag{4}$$

Let us define the second-order noise, for each non-overlapping channel $\mathbf{z}_k$, $n_{k,j}$ can be formulated as:

$$n_{k,j} = z_{k,j} - \widehat{z}_{k,j}, \tag{5}$$

where $\widehat{z}_{k,j}$ denotes the estimation of $z_{k,j}$. With the proposed regression model Eq. (2), one can obtain the second-order noise of the image. Assuming that the random variable $n_{k,j}$ is independent and identically distributed (IID), and can be modelled by the Gaussian distribution written as:

$$n_{k,j} \sim \mathcal{N}(\mu_k, \sigma_k^2), \tag{6}$$

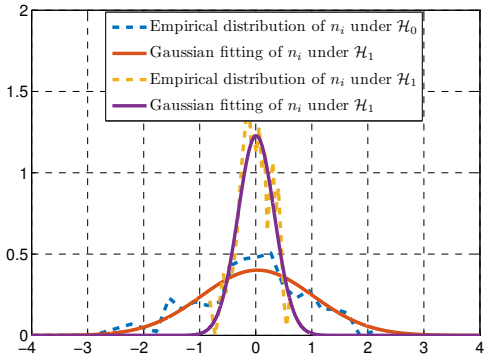

Figure 2 : Empirical distribution of the estimated second-order noise n_i, $\mathcal{H}_0$ and $\mathcal{H}_1$ see Section 3.1.

where μ_k is the expectation of the model and σ_k represents the standard deviation of the distribution.

For clarity and simplicity, let us extend the assumption for one channel of the second-order noise to all the channels. An alternative representation of those second-order noise is usually adopted by gathering the second-order noise. Let us assume that all the random variables with different channels of the image follow the Gaussian model with the same expectation and standard deviation. Due to that μ_k and σ_k within different channels are very approximate, it is reasonable that we can formulate the statistical distribution of the second-order residuals by $\mathbf{n} = \{n_i\}$, and $n_i \sim \mathcal{N}(\mu, \sigma^2)$, where, $i \in \{1, \ldots, l\}$, with $l = K \cdot m$, while μ and σ denotes the expectation and the standard deviation respectively.

It is difficult to establish the statistical distribution of an image $\mathbf{I}$. However, the second-order residuals of an image approximately follow the Gaussian model with sharing similar expectation and variance among different channels. This assumption has been empirically evaluated on a dataset. As Fig. 2 illustrates, we show the comparison between the empirical result of second-order noise n_i and its the Gaussian fitting.

3 Optimal detector for classification: designing the LRT

3.1 Classification problem formulation

This section aims at presenting the optimal LRT and studying its statistical performance. The statistical test is designed based on the second-order noise n_i. Each type image can be characterized by its Gaussian parametric model with parameters $\boldsymbol{\theta}_t = (\mu_t, \sigma_t^2)$, $t \in \{0, 1\}$. Hence, within the framework of the hypothesis testing theory, the classification problem can be cast into the two

following binary detection:

$$\begin{cases} \mathcal{H}_0 : \ \{n_i \sim \mathcal{N}(\mu_0, \sigma_0^2), \forall i = (1, ..., l)\} \\ \mathcal{H}_1 : \ \{n_i \sim \mathcal{N}(\mu_1, \sigma_1^2), \forall i = (1, ..., l)\} \end{cases} \tag{7}$$

where (μ_0, σ_0^2) represents the expectation and variance under hypothesis $\mathcal{H}_0 = \{\text{the image is NI}\}$, and (μ_1, σ_1^2) for hypothesis $\mathcal{H}_1 = \{\text{the image is CG}\}$. For solving statistical detection problem above, Lehmann [Lehmann and Romano, 2006, Theorem 3.2.1] states that the most powerful test for classifying between $\mathcal{H}_0$ and $\mathcal{H}_1$ at a given false positive rate (FPR) α_0 in the class $\mathcal{K}_{\alpha_0}$ can be described below. Let

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{\boldsymbol{\theta}_t} \mathbb{P}_0[\delta(\mathbf{n}) = \mathcal{H}_1] \leq \alpha_0 \right\} \tag{8}$$

be the class of test, with an upper-bounded FPR α_0. Here, $\mathbb{P}_t[\cdot]$ stands for the probability under $\mathcal{H}_t, t \in \{0, 1\}$, and the supremum over $\boldsymbol{\theta}_t$ has to be understood as whatever the distribution parameters might be, in order to ensure that the false alarm probability α_0 can not be exceeded. It is aimed at finding a test δ maximizing the power function, defined by the true positive rate (TPR). Among all the tests in $\mathcal{K}_{\alpha_0}$, the LRT is the most powerful test, which maximizes the detection power:

$$\beta_\delta = \mathbb{P}_1[\delta(\mathbf{n}) = \mathcal{H}_1], \tag{9}$$

equals to minimize the false negative rate $\alpha_1 = \mathbb{P}_1[\delta(\mathbf{n}) = \mathcal{H}_0] = 1 - \beta_\delta$.

In the following subsection, the LRT is first described in details and then its statistical performance is analytically established.

3.2 Design of likelihood ratio test

We assume that the statistical discriminator parameters $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0)$, $\boldsymbol{\theta}_1 = (\mu_1, \sigma_1)$ are all known, the classification problem can be transformed to a statistical test between two simple hypotheses. Based on the assumption that the random variables n_i is IID, the LRT can be represented by the following decision rule:

$$\delta(\mathbf{n}) = \begin{cases} \mathcal{H}_0 \ \text{ if } \ \Lambda(\mathbf{n}) = \sum_{i=1}^{l} \Lambda(n_i) \leq \tau \\ \mathcal{H}_1 \ \text{ if } \ \Lambda(\mathbf{n}) = \sum_{i=1}^{l} \Lambda(n_i) > \tau \end{cases} \tag{10}$$

where the solution of $\mathbb{P}_0 [\Lambda(\mathbf{n}) > \tau] = \alpha_0$ is denoted as the decision threshold τ, to guarantee that the FPR equals α_0. Based on the Gaussian model Eq. (7), the probability mass function (PMF) under two hypotheses can be respectively written as: $P_{\boldsymbol{\theta}_0}$ and $P_{\boldsymbol{\theta}_1}$, then one can describe the log Likelihood Ratio (LR) for one observation as:

$$\Lambda(n_i) = \log \frac{P_{\boldsymbol{\theta}_1}[n_i]}{P_{\boldsymbol{\theta}_0}[n_i]}. \tag{11}$$

In practical detection, since parameter μ_t are too small, resulting into that it can be prescribed as 0. Therefore, the LR is easy to be represented as:

$$\Lambda(n_i) = \log\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2} n_i^2. \tag{12}$$

3.3 Statistical performance of LRT

Based on the assumption that the second-order noise n_i is IID, we propose to adopt some asymptotic theorems, due to in this context the number of observations of the second-order noise is large enough. Then under hypothesis $\mathcal{H}_t,\ t \in \{0,1\}$, $E_{\mathcal{H}_t}(\Lambda(n_i))$ and $V_{\mathcal{H}_t}(\Lambda(n_i))$ denotes the expectation and the variance of the LR $\Lambda(n_i)$ respectively. Note that we assume $\mu_t = 0$, and for each LR $\Lambda(n_i)$ in $\Lambda(\mathbf{n})$, the expectation and the variance can be expressed by:

$$E_{\mathcal{H}_t}(\Lambda(n_i)) = \log(\frac{\sigma_0}{\sigma_1}) + \frac{\sigma_1^2 - \sigma_0^2}{2\sigma_0^2\sigma_1^2}\sigma_t^2, \tag{13}$$

$$V_{\mathcal{H}_t}(\Lambda(n_i)) = \frac{\left(\sigma_1^2 - \sigma_0^2\right)^2}{4\sigma_0^4\sigma_1^4} 2\sigma_t^4,\ \ t \in \{0,1\}. \tag{14}$$

Note that those moments can be calculated analytically, Lindeberg's central limit theorem(CLT) [Lehmann and Romano, 2006, Theorem 11.2.1] states that as the number of samples l of the second-order noise tends to infinity it holds true that:

$$\frac{\sum_{i=1}^{l}\left(\Lambda(n_i) - E_{\mathcal{H}_t}(\Lambda(n_i))\right)}{\left(\sum_{i=1}^{l} V_{\mathcal{H}_t}(\Lambda(n_i))\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1),\ \ t \in \{0,1\}, \tag{15}$$

where $\xrightarrow{d}$ represents the convergence in distribution; $\mathcal{N}(0,1)$ denotes the standard normal distribution with zero expectation and unit variance. This takes crucial interest to establish the statistical properties of the LRT (see [Qiao et al., 2015b, Qiao et al., 2014]). In fact, once the moments of the LR have been calculated analytically under hypothesis $\mathcal{H}_0$, one can normalize the LR $\Lambda(\mathbf{n})$ under hypothesis $\mathcal{H}_0$ as follows:

$$\widetilde{\Lambda}(\mathbf{n}) = \frac{\Lambda(\mathbf{n}) - \sum_{i=1}^{l} E_{\mathcal{H}_0}(\Lambda(n_i))}{(\sum_{i=1}^{l} V_{\mathcal{H}_0}(\Lambda(n_i)))^{1/2}}. \tag{16}$$

Then let us define the normalized LRT with $\widetilde{\Lambda}(\mathbf{n})$ by:

$$\widetilde{\delta} = \begin{cases} \mathcal{H}_0 \ \text{ if } \ \widetilde{\Lambda}(\mathbf{n}) < \widetilde{\tau} \\ \mathcal{H}_1 \ \text{ if } \ \widetilde{\Lambda}(\mathbf{n}) \geq \widetilde{\tau} \end{cases} \tag{17}$$

thus, it is straightforward to establish the statistical properties of the LRT Eq. (17).

Proposition 1. *Assuming that for the image classification as case within the two simple hypotheses Eq.* (7)*, in which both parameters* μ_t *and* σ_t *are all known, for clarity,* Φ *and* Φ^{-1} *respectively represent the cumulative distribution function (cdf) of the standard normal distribution and its inverse, then for any* $\alpha_0 \in (0,1)$*, the decision threshold can be defined by:*

$$\widetilde{\tau} = \Phi^{-1}(1-\alpha_0). \tag{18}$$

Proposition 2. *Assuming that for the image classification problem as case within the two simple hypotheses Eq.* (7)*, in which both parameters* μ_t *and* σ_t *are all known, for any decision threshold* $\widetilde{\tau}$*, the power function associated with the proposed test* $\widetilde{\delta}$ *Eq.* (17) *is given by:*

$$\beta_{\widetilde{\delta}} = 1 - \Phi\left(\sqrt{\frac{v_0}{v_1}} \cdot \Phi^{-1}(1-\alpha_0) + \frac{e_0 - e_1}{\sqrt{v_1}}\right), \tag{19}$$

where $e_t = \sum_{i=1}^{l} E_{\mathcal{H}_t}(\Lambda(n_{i,t}))$, $v_t = \sum_{i=1}^{l} V_{\mathcal{H}_t}(\Lambda(n_{i,t}))$, $\quad t \in \{0,1\}$.

Eqs. (18) and (19) emphasize the main advantage of the normalized LR as described in Eq. (17): it allows to set any threshold that guarantees a FPR independently from any distribution parameter.

4 Practical detector for classification: designing the GLRT

4.1 Model parameter estimation

In the practical scenario, it is much more realistic to assume that parameters of Gaussian model-based discriminator are unknown. Our proposed classification aims at identifying the given image **I** acquired by either digital camera or computer rendering software.

In this section, let us devise the GLRT for dealing with the problem described in Eq. (7). Hence, first we have to estimate expectation of the Gaussian model parameters by the Maximum Likelihood (ML) algorithm as follows:

$$\bar{\mu}_t = \frac{1}{M \cdot l} \sum_{m=1}^{M} \sum_{i=1}^{l} n_i^m, \tag{20}$$

and the mean of variance is denoted as:

$$\bar{\sigma}_t^2 = \frac{1}{M \cdot (l-1)} \sum_{m=1}^{M} \sum_{i=1}^{l} (n_i^m - \frac{1}{l} \sum_{i=1}^{l} n_i^m)^2, \tag{21}$$

where M denotes the total number of images.

4.2 Design of generalized likelihood ratio test

Generally, the solution of the Generalized Likelihood Ratio (GLR) essentially consists in replacing the unknown parameters by its ML estimation. Immediately, let us define the GLRT $\widehat{\delta}_1(\mathbf{n})$ as follows:

$$\widehat{\delta}_1(\mathbf{n}) = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}_1(\mathbf{n}) = \sum_{i=1}^{l} \widehat{\Lambda}_1(n_i) < \widehat{\tau}_1 \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}_1(\mathbf{n}) = \sum_{i=1}^{l} \widehat{\Lambda}_1(n_i) \geq \widehat{\tau}_1 \end{cases} \tag{22}$$

where $\widehat{\tau}_1$ refers to the solution of following:

$$\mathbb{P}_{\mathcal{H}_0}\left[\widehat{\Lambda}_1(\mathbf{n}) \geq \widehat{\tau}_1\right] = \alpha_0. \tag{23}$$

For each sample, the log GLR (refers to the log value of generalized likelihood ratio), $\widehat{\Lambda}_1(n_i)$ can be expressed by:

$$\widehat{\Lambda}_1(n_i) = \log \frac{P_{\widehat{\boldsymbol{\theta}}_1}[n_i]}{P_{\widehat{\boldsymbol{\theta}}_0}[n_i]}. \tag{24}$$

Based on the assumption that $\bar{\mu}_t$ is infinitely close to 0, refers to Eq. (20), here, $\widehat{\boldsymbol{\theta}}_0 = (0, \bar{\sigma}_0^2)$ and $\widehat{\boldsymbol{\theta}}_1 = (0, \bar{\sigma}_1^2)$ denote the estimates of statistical parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, with the specific elements shown in Eq. (7), one can rewrite the GLR Eq. (24) as:

$$\widehat{\Lambda}_1(n_i) = \log\left(\frac{\bar{\sigma}_0}{\bar{\sigma}_1}\right) + \frac{1}{2}\left(\frac{n_i^2}{\bar{\sigma}_0^2} - \frac{n_i^2}{\bar{\sigma}_1^2}\right), \tag{25}$$

where $\bar{\sigma}_t, t \in \{0, 1\}$, denotes the averaged estimation of the second-order noise's standard variance of the training dataset, as given in Eqs. (20) and (21). In order to normalize the GLR $\widehat{\Lambda}_1(\mathbf{n})$ of the second-order noise, Eq. (25) can be formulated as:

$$\widehat{\Lambda}_2(\mathbf{n}) = \frac{\widehat{\Lambda}_1(\mathbf{n}) - e_t^\star}{\sqrt{v_t^\star}}. \tag{26}$$

where $e_t^\star = \sum_{i=1}^{l} E_{\mathcal{H}_t}\left(\widehat{\Lambda}_1(n_{i,t})\right)$ and $v_t^\star = \sum_{i=1}^{l} V_{\mathcal{H}_t}\left(\widehat{\Lambda}_1(n_{i,t})\right)$, indicates the expectation and variance of GLR $\widehat{\Lambda}_1(\mathbf{n})$, under hypothesis $\mathcal{H}_t$, $t \in \{0, 1\}$ respectively. Hence, for the second-order noise $\mathbf{n}$ of an inspected image, the classification problem can be easily formulated by the normalized GLRT:

$$\widehat{\delta}_2(\mathbf{n}) = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}_2(\mathbf{n}) < \widehat{\tau}_2 \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}_2(\mathbf{n}) \geq \widehat{\tau}_2 \end{cases} \tag{27}$$

Thus, let us establish the statistical properties of the GLRT.

Proposition 3. *Assuming that the second-order noise of an image is modeled by the proposed Gaussian parametric model, and both linear model statistically*

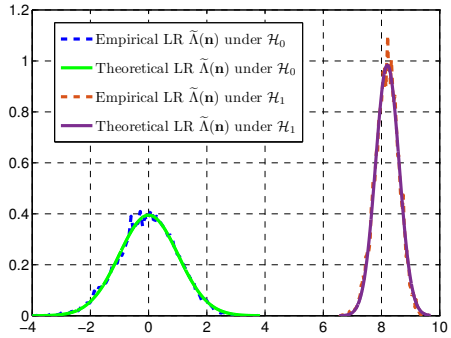

Figure 3 : LR Comparison between empirical and theoretical distribution of $\widetilde{\Lambda}(\mathbf{n})$ under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively.

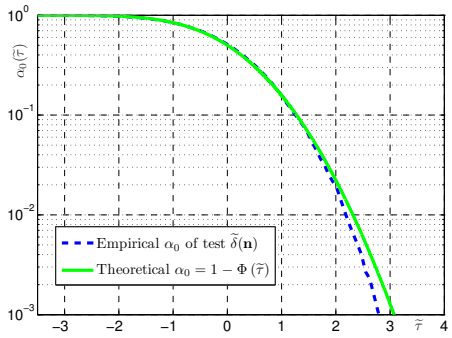

Figure 4 : Comparison between the theoretical FPR α_0 and the empirical results plotted as a function of the $\widetilde{\tau}$.

unknown parameters $\widehat{\boldsymbol{\theta}}_0 = (0, \bar{\sigma}_0^2), \widehat{\boldsymbol{\theta}}_1 = (0, \bar{\sigma}_1^2)$ *are estimated as Eqs.* (20) *and* (21)*, then for any* $\alpha \in (0, 1)$ *the decision threshold of the proposed GLRT* $\widehat{\delta}_2$ *can be calculated by:*

$$\widehat{\tau}_2 = \Phi^{-1}(1 - \alpha_0). \tag{28}$$

Proposition 4. *Assuming that the second-order noise of an image are modeled by the proposed Gaussian parametric model, and both linear model unknown statistical parameters* $\widehat{\boldsymbol{\theta}}_0 = (0, \bar{\sigma}_0^2), \widehat{\boldsymbol{\theta}}_1 = (0, \bar{\sigma}_1^2)$ *are estimated as in Eqs.* (20) *and* (21)*, for any decision threshold* $\widehat{\tau}_2$*, the detection power function associated with test* $\widehat{\delta}_2$ *Eq.* (27) *can be calculated by:*

$$\beta_{\widehat{\delta}_2} = 1 - \Phi\left(\sqrt{\frac{v_0^\star}{v_1^\star}} \cdot \Phi^{-1}(1 - \alpha_0) + \frac{e_0^\star - e_1^\star}{\sqrt{v_1^\star}}\right). \tag{29}$$

5 Experimental results and analysis

This section illustrates theoretical and empirical experimental results of our proposed statistical discriminator, comparative analysis with other prior arts [Qiao et al., 2013, Lyu and Farid, 2005, Peng et al., 2017, Gallagher and Chen, 2008], and robustness investigation of our discriminator.

5.1 Results on simulated images

To evaluate our theoretically established results, we use the simulated noise which contains the images' characteristics. Specifically, we first generate two groups of random variables, simulating the second-order noise, where the mean and variance are pre-set based on the CG and NI samples. In this work, NI

is characterized by $\boldsymbol{\theta}_0 = (0, 0.4747)$, referring to as the model parameters of n_i under hypothesis $\mathcal{H}_0$; meanwhile, CG is characterized by $\boldsymbol{\theta}_1 = (0, 0.1968)$ under hypothesis $\mathcal{H}_1$. It is necessary to note that pixels of the NI exhibit high correlation caused by the CFA interpolation. In the stage of noise extraction, the selected threshold of wavelet coefficients is larger than the value of the optimal threshold of NI proposed in [Donoho, 1995], which leads to the extracted first-order residuals of NI with a larger value, and meanwhile that the variance of the first-order residuals of NI is larger than CG.

In fact, those parameters are estimated parameters for two different types of images from the Dresden dataset [Gloe and Böhme, 2010] and our collected CG dataset. Thus, one can construct two sets of random variables by repeating 10000 simulations respectively. Then, the first set of 10000 images representing NI, consisting of 383 realizations [3] of random variables for each image; the second set of 10000 images representing CG, consisting of the same number of random variables.

The Fig. 3 illustrates the comparison between the theoretical and empirical distribution of the optimal LR $\widetilde{\Lambda}(\mathbf{n})$. Under hypothesis $\mathcal{H}_0$, the empirical result of the LR $\widetilde{\Lambda}(\mathbf{n})$ approximately follows the standard Gaussian distribution, which directly verifies the correctness of the theoretically established statistical property for the proposed LRT. Similarly, under hypothesis $\mathcal{H}_1$, the empirical and theoretical distribution of the optimal LR $\widetilde{\Lambda}(\mathbf{n})$ are nearly superposed. In this scenario, the Gaussian distribution is characterized by expectation $\frac{e_1-e_0}{\sqrt{v_0}}$ and variance $\frac{v_1}{v_0}$ (see Section 3.3), which also validates the effectiveness of the established statistical classifier. In addition, our statistical test can warrant the prescribed FPR. Thus, let us compare the empirical and theoretical FPR α_0 of the optimal LR $\widetilde{\Lambda}(\mathbf{n})$ as a function of the decision threshold $\widetilde{\tau}$, see Fig. 4. This figure emphasizes that the proposed LRT has the ability to guarantee the prescribed FPR in practice. In some cases ($\widetilde{\tau} \geq 2$), we emphasize that the slight discrimination between two curves might be caused by the inaccuracy of the CLT which can hardly model the tails of the distribution.

5.2 Results on authentic dataset

In the following experiments, we select 400 NIs of various indoor and outdoor scenes from Dresden dataset [Gloe and Böhme, 2010], and 400 CGs from our own collected dataset which are created by using rendering software or downloaded from the Internet including computer portraits, various indoor and outdoor scenes; then let us crop the central 256×256 region of each sample which carries sufficient information to characterize the corresponding type of images. Besides, the CGs downloaded from the Internet cannot own the same size as the NIs. In

[3] The value 383 is calculated from NI dataset.

our proposed algorithm, we have to guarantee that the size of an inquiry image remains unchanged.

In Section 2.2, three paradigms of the first-order noise extraction have been studied. Then let us compare the detection power using different filters. It is proposed to use the accuracy (ACC) and Area Under Curve (AUC) as metric to evaluate the performance of the detector with three different filters. The ACC of the detector using the first, the second and the third filter is 70.40%, 75.63%, and 90.08%; meanwhile the AUC is 0.732, 0.774, and 0.920. It is clear that the detector using the third filter can achieve the optimal result. Because the third one utilizes the wavelet decomposition of multiple scales which could obtain more information including the first-order noise. Next, we will discuss the settings of another regression model filter for extracting second-order noise in the frequency domain.

The regression model-based filter is formulated by a polynomial order n equal to 4, and the size of the vector $\mathbf{z}_k$ is set with $m = 64$ observations. Note that to avoid dealing with the outlier (or abnormal variables), possibly resulting into the unstable results in the stage of parameter estimation, the first channel $\mathbf{z}_1$ and the last $\mathbf{z}_K$ set of noise are both excluded from our calculation.

The detection power (or TPR) of the LRT and GLRT are both illustrated in Fig. 5. The ROC that is the β_δ as a function of FPR α_0, of empirical established performance Eq. (29) is compared. It should be noted that the ROC curve of the LRT is generated from the 10000 simulations, while the GLRT obtained from an authentic dataset. In fact, the loss of power indeed exists between discriminators, caused by the mismatch of the model fitting and parameter estimation. In practical discrimination, because of pixel inhomogeneity, the proposed Gaussian model cannot fit all the inspected images. Besides, another limitation is that our proposed algorithm of estimation is not optimal, which can be improved by using other better-performed algorithms. Nevertheless, our proposed GLRT indeed provides a general framework for designing a un-supervised model-based detector, which has not been widely investigated in the forensic community.

5.2.1 Comparative analysis

We carry out the experiments by comparing our proposed discriminator with some prior arts [Qiao et al., 2013,Lyu and Farid, 2005,Peng et al., 2017,Gallagher and Chen, 2008] in Fig. 6. Because the algorithms in [Lyu and Farid, 2005, Peng et al., 2017] require a large scale of labeled images to train the designed discriminators. We propose to randomly segment 9 portions of the size 256×256 from each image of NI and CG, and obtain 3600 NIs and 3600 CGs that are used for comparative analysis with other classification schemes. We use the

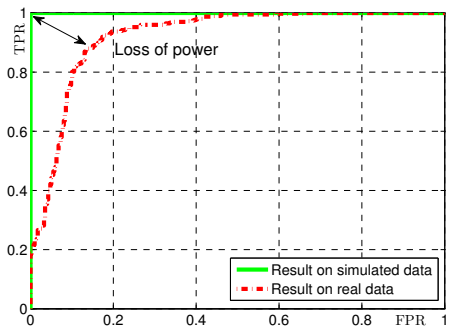

Figure 5 : Different data comparison.

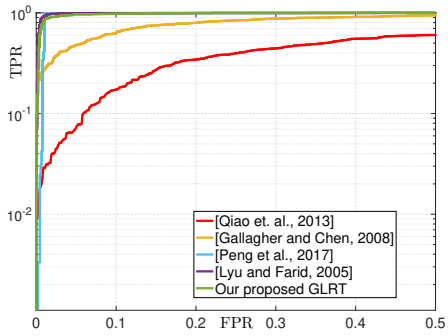

Figure 6 Different motheds comparison.

Table 1 : Comparison of detection power using different methods.

Metric \ Method	Proposed	[Qiao et al., 2013]	[Lyu and Farid, 2005]	[Peng et al., 2017]	[Gallagher and Chen, 2008]
P_{NI}	95.33%	56.90%	98.78%	99.11%	80.33%
P_{CG}	96.44%	57.35%	96.78%	97.89%	73.44%

identification accuracy for NI and CG as metrics, and which is formulated by:

$$P_{NI} = \frac{TP}{P}, \quad \text{and} \quad P_{CG} = \frac{TN}{N} \tag{30}$$

where TP and TN denote the number of correctly classified NI and CG, P and N is the number of real positive cases and real negative cases respectively in the dataset. Table 1 illustrates the identification accuracy comparison of methods [Qiao et al., 2013, Lyu and Farid, 2005, Peng et al., 2017, Gallagher and Chen, 2008]. For LIBSVM [Chang and Lin, 2011] discriminator with the linear kernel proposed in [Lyu and Farid, 2005, Peng et al., 2017], we use the hold-out to estimate the identification accuracy. To guarantee the reliability of experiments, first, we randomly select 300 NIs and 300 CGs from the image dataset as training dataset, and set remaining 100 NIs and 100 CGs as testing dataset. Then by cropping 9 non-overlapped patches for each image, the selected training image dataset is extended to a dataset with 2700 NIs and 2700 CGs, and the remaining testing dataset is extended to a dataset with 900 NIs and 900 CGs.

Compared to the supervised detectors in [Lyu and Farid, 2005, Peng et al., 2017], the performance of our discriminator is very slightly worse than that of the prior art. Because the proposed scheme only utilizes the second-order noise as features for image authentication. However, the supervised scheme requires a large scale of training data (at least 500 training images in [Peng et al., 2017]; at least 4800 training images in [Lyu and Farid, 2005]), we only use 20 images to estimate the parameters, as shown in Table 2. As the number of trained images degrades, the detection power is relevant stable, meaning that our statistical classifier can remain its detection power with the limited given samples while the supervised algorithms [Lyu and Farid, 2005,Peng et al., 2017] cannot perform

Table 2 : The detection power β_δ comparison using different number of training samples under different FPR α_0.

n \ α_0	0.1	0.2	0.3	0.4	0.5
10	97.00%	99.00%	99.56%	99.67%	99.89%
20	97.11%	99.00%	99.56%	99.67%	99.89%
30	97.22%	99.00%	99.56%	99.67%	99.89%
40	97.22%	99.00%	99.56%	99.67%	99.89%
50	97.22%	99.00%	99.56%	99.67%	99.89%

well. Our discriminator outperforms both two other un-supervised detectors [Gallagher and Chen, 2008, Qiao et al., 2013]. Due to the considerable small size 256×256 of tested images, the detector of [Qiao et al., 2013] requiring the large size of the inspected image cannot collect enough observations for establishing the concerning detector, resulting in unsatisfying performance. However, our proposed discriminator cannot easily be interfered with the image size (see Fig. 7d). In addition, the detection performance of [Gallagher and Chen, 2008] largely relies on the unique feature characterized by the peak value in the frequency spectrum, which is not very robust to the images with variable textures.

5.2.2 Analysis of robustness

In recent studies, some researchers focus on the images in the social networks, which have been compressed or resized. In that scenario, we have to consider the proposed algorithm can resist against the attacks caused by post-processing operations. In order to analyze the robustness of the proposed statistical model, different post-processing operations to the images. The operation includes: compressing images saved as JPEG format, resizing images, adding Gaussian white noise to images and cropping images.

The settings of the quality factors for JPEG compression are 98, 96, 94, and 92. We use 20 NIs and 20 CGs to estimate the model parameters that will be used to image classification. Since high-resolution images with large quality factors are prevalent anywhere, it makes sense that we evaluate the robustness of our proposed algorithm in the case of large quality factors. As shown in Fig. 7a, it presents ROC curves of the proposed discrimination for uncompressed and compressed images. With decreasing the quality factor, the performance of the proposed discriminator is degraded. In fact, when the quality factor arrives at 80, the detection performance is nearly close to randomly guess. However, it still preserves a high detection performance with comparatively large quality factors.

Then, it is proposed to investigate the detection performance of resized images. First, we resize the image with scaling factors of 0.8, 1.2 and 1.5, and then use 20 NIs and 20 CGs to estimate the model parameters. As Fig. 7b illustrates, the detection performance is degraded severely when the scaling

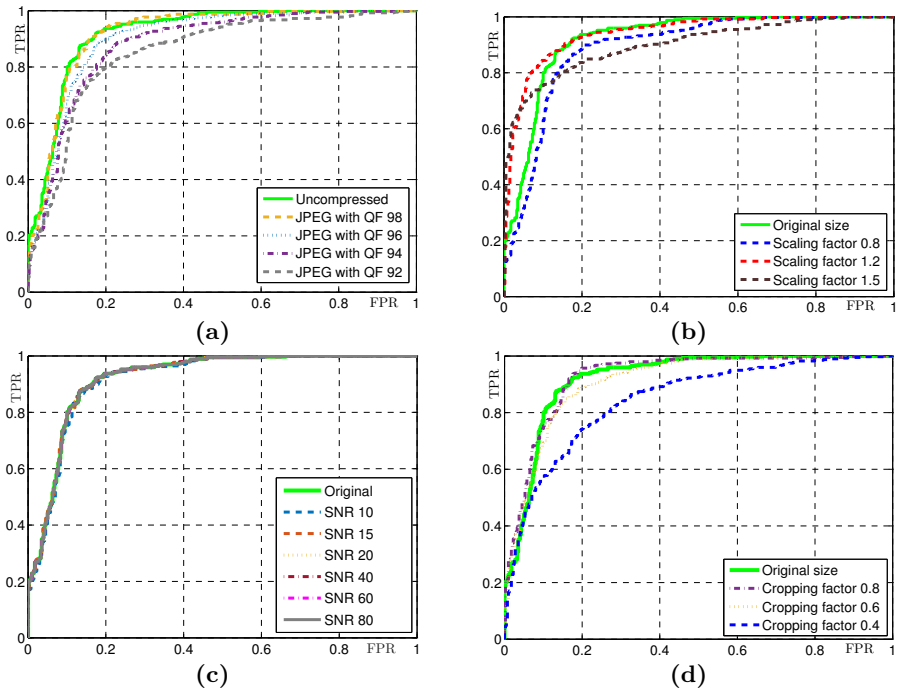

Figure 7 : The ROC curves comparison with (a) different compression quality factors; (b) different scaling factors; (c) adding different Gaussian noise; (d) different cropping factors.

factor is 1.5. The main reason is that the image size enlargement can be regarded as a linear interpolation of local pixels. Based on the principle that pixels of the NI with the same interpolation model or interpolation algorithm exhibit high correlation, the pixels of CG introduced high correction with the linear interpolation of enlargement, which can reduce the difference between NI and CG, and meanwhile will eventually influence the classification accuracy.

Thirdly, we add Gaussian white noise with the SNR (dB) of 80, 60, 40 and 20, and use 20 NIs and 20 CGs to estimate the model parameters. As shown in Fig. 7c, with increasing noise intensity (decreasing the value of SNR), the discrimination performance remains stable, testifying the proposed statistical classifier could resist the attack of adding noise effectively.

Last, we acquire images with different cropping factors of 0.8, 0.6 and 0.4, then we choose 20 NIs and 20 CGs for estimating the model parameters which will be used to image classification. As Fig. 7d illustrates, our proposed detector still perform very well when cropping factor equals to 0.6 or 0.8. While the cropping factor equals to 0.4, the detection power of our proposed classifier degrades. Because of the reduction of the size of the test images, we cannot

obtain enough pixels, resulting in the lack of cumulative difference between NI and CG for our image classification.

6 Conclusion and discussion

In this paper, we propose an approach to differentiate between natural images and computer-generated graphics. We not only focus on the problem of real images classification by designing the GLRT, but also theoretically establish a statistical discriminator, referring to as LRT, for classifying the simulated data. Each type of image is characterized by its Gaussian distribution of second-order noise with parameters $(\mu_t, \sigma_t), t \in \{0, 1\}$, which is extracted using a regression model-based filter in the frequency domain. Then the problem of image classification is cast into the framework of the hypothesis testing theory. Assuming that all of the parameters are known, the statistical performance of the LRT is analytically established, meanwhile the statistical property is studied. In the practical scenario, based on the estimated parameters, the proposed GLRT shows efficient classification performance under the prescribed FPR. Moreover, numerical results verify that the proposed scheme can achieve fairly good performance with a good robustness against to some post-processing techniques.

In fact, in the community of image forensics for classifying between NI and CG. The supervised methodologies dominate the research. Again, we have to admit that our proposed discriminator cannot outperform (but very close to) that of the supervised algorithms such as [Lyu and Farid, 2005,Peng et al., 2017]. However, our algorithm indeed provides an alternative solution to deal with the problem of classification between NI and CG. Supposing that the forensic analyzer cannot acquire a large scale of labeled images for training, the accuracy of the supervised will not be guaranteed.

Acknowledgements

This work is funded by the cyberspace security major program in National Key Research and Development Plan of China under grant No. 2016YFB0800201, the Natural Science Foundation of China under grant No. 61702150 and and No. 61572165, the Public Research Project of Zhejiang Province under grant No. LGG19F020015, the Key research and development plan project of Zhejiang Province under grant No. 2017C01062 and No. 2017C01065.

References

[Birajdar and Mankar, 2013] Birajdar, G. K. and Mankar, V. H. (2013). Digital image forgery detection using passive techniques: A survey. *Digital Investigation*, 10(3):226–245.

[Caldelli et al., 2017] Caldelli, R., Becarelli, R., and Amerini, I. (2017). Image origin classification based on social network provenance. *IEEE Transactions on Information Forensics and Security*, 12(6):1299–1308.

[Chang and Lin, 2011] Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27.

[De Rezende et al., 2018] De Rezende, E. R., Ruppert, G. C., Theophilo, A., Tokuda, E. K., and Carvalho, T. (2018). Exposing computer generated images by using deep convolutional neural networks. *Signal Processing: Image Communication*, 66:113–126.

[Dirik and Memon, 2009] Dirik, A. E. and Memon, N. (2009). Image tamper detection based on demosaicing artifacts. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1497–1500. IEEE.

[Donoho, 1995] Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627.

[Gallagher and Chen, 2008] Gallagher, A. and Chen, T. (2008). Image authentication by detecting traces of demosaicing. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE.

[Gallagher, 2005] Gallagher, A. C. (2005). Detection of linear and cubic interpolation in jpeg compressed images. In *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*, pages 65–72. IEEE.

[Gloe and Böhme, 2010] Gloe, T. and Böhme, R. (2010). The Dresden image database for benchmarking digital image forensics. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 1584–1590. Acm.

[Holub and Fridrich, 2013] Holub, V. and Fridrich, J. (2013). Digital image steganography using universal distortion. In *Proceedings of the first ACM workshop on Information hiding and multimedia security*, pages 59–68. ACM.

[Lehmann and Romano, 2006] Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

[Long et al., 2017] Long, M., Peng, F., and Zhu, Y. (2017). Identifying natural images and computer generated graphics based on binary similarity measures of prnu. *Multimedia Tools and Applications*, pages 1–18.

[Luo et al., 2016] Luo, X., Song, X., Li, X., Zhang, W., Lu, J., Yang, C., and Liu, F. (2016). Steganalysis of hugo steganography based on parameter recognition of syndrome-trellis-codes. *Multimedia Tools and Applications*, 75(21):13557–13583.

[Lyu and Farid, 2005] Lyu, S. and Farid, H. (2005). How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 53(2):845–850.

[Ma et al., 2018] Ma, Y., Luo, X., Li, X., Bao, Z., and Zhang, Y. (2018). Selection of rich model steganalysis features based on decision rough set α-positive region reduction. *IEEE Transactions on Circuits and Systems for Video Technology*.

[Mader et al., 2017] Mader, B., Banks, M. S., and Farid, H. (2017). Identifying computer-generated portraits: The importance of training and incentives. *Perception*, 46(9):1062–1076.

[Ozparlak and Avcibas, 2011] Ozparlak, L. and Avcibas, I. (2011). Differentiating between images using wavelet-based transforms: a comparative study. *IEEE Transactions on Information Forensics and Security*, 6(4):1418–1431.

[Peng et al., 2017] Peng, F., Zhou, D.-l., Long, M., and Sun, X.-m. (2017). Discrimination of natural images and computer generated graphics based on multi-fractal and regression analysis. *AEU-International Journal of Electronics and Communications*, 71:72–81.

[Potdar et al., 2005] Potdar, V. M., Han, S., and Chang, E. (2005). A survey of digital image watermarking techniques. In *Industrial Informatics, 2005. INDIN'05. 2005 3rd IEEE International Conference on*, pages 709–716. IEEE.

[Qiao and Retraint, 2018] Qiao, T. and Retraint, F. (2018). Identifying individual camera device from raw images. *IEEE Access*, 6:78038–78054.

[Qiao et al., 2013] Qiao, T., Retraint, F., and Cogranne, R. (2013). Image authentication by statistical analysis. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5. IEEE.

[Qiao et al., 2015a] Qiao, T., Retraint, F., Cogranne, R., and Thai, T. H. (2015a). Source camera device identification based on raw images. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 3812–3816. IEEE.

[Qiao et al., 2017] Qiao, T., Retraint, F., Cogranne, R., and Thai, T. H. (2017). Individual camera device identification from jpeg images. *Signal Processing: Image Communication*, 52:74–86.

[Qiao et al., 2015b] Qiao, T., Retraint, F., Cogranne, R., and Zitzmann, C. (2015b). Steganalysis of jsteg algorithm using hypothesis testing theory. *EURASIP Journal on Information Security*, 2015(1):2.

[Qiao et al., 2019] Qiao, T., Shi, R., Luo, X., Xu, M., Zheng, N., and Wu, Y. (2019). Statistical model-based detector via texture weight map: Application in re-sampling authentication. *IEEE Transactions on Multimedia*, 21(5):1077–1092.

[Qiao et al., 2018] Qiao, T., Zhu, A., and Retraint, F. (2018). Exposing image resampling forgery by using linear parametric model. *Multimedia Tools and Applications*, 77:1501–1523.

[Qiao et al., 2014] Qiao, T., Zitzmann, C., Retraint, F., and Cogranne, R. (2014). Statistical detection of jsteg steganography using hypothesis testing theory. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 5517–5521. IEEE.

[Swaminathan et al., 2008] Swaminathan, A., Wu, M., and Liu, K. R. (2008). Digital image forensics via intrinsic fingerprints. *Information Forensics and Security, IEEE Transactions on*, 3(1):101–117.

[Wang et al., 2016] Wang, J., Li, T., Shi, Y.-Q., Lian, S., and Ye, J. (2016). Forensics feature analysis in quaternion wavelet domain for distinguishing photographic images and computer graphics. *Multimedia Tools and Applications*, pages 1–17.

[Yao et al., 2018] Yao, H., Qiao, T., Xu, M., and Zheng, N. (2018). Robust multi-classifier for camera model identification based on convolution neural network. *IEEE Access*, 6:24973–24982.

[Zhang et al., 2018] Zhang, Y., Qin, C., Zhang, W., Liu, F., and Luo, X. (2018). On the fault-tolerant performance for a class of robust image steganography. *Signal Processing*, 146:99–111.

[Zhao et al., 2019] Zhao, Y., Zheng, N., Qiao, T., and Xu, M. (2019). Source camera identification via low dimensional prnu features. *Multimedia Tools and Applications*, 78(7):8247–8269.

[Zhou et al., 2017] Zhou, Z., Wang, Y., Wu, Q. J., Yang, C.-N., and Sun, X. (2017). Effective and efficient global context verification for image copy detection. *IEEE Transactions on Information Forensics and Security*, 12(1):48–63.