Robust Estimation of Upscaling Factor on Double JPEG Compressed Images

Wei Lu, *Member, IEEE*, Qin Zhang, Shangjun Luo, Yicong Zhou, *Senior Member, IEEE*, Jiwu Huang, *Fellow, IEEE*, and Yun-Qing Shi, *Life Fellow, IEEE*

***Abstract*—As one of the most important topics in image forensics, resampling detection has developed rapidly in recent years. However, the robustness to JPEG compression is still challenging for most classical spectrum-based methods, since JPEG compression severely degrades the image contents and introduces block artifacts in the boundary of the compression grid. In this article, we propose a method to estimate the upscaling factors on double JPEG compressed images in the presence of image upscaling between the two compressions. We first analyze the spectrum of scaled images and give an overall formulation of how the scaling factors along with the parameters of JPEG compression and image contents influence the appearance of tampering artifacts. The expected positions of five kinds of characteristic peaks are analytically derived. Then, we analyze the features of double JPEG compressed images in the block discrete cosine transform (BDCT) domain and present an inverse scaling strategy for the upscaling factor estimation with a detailed proof. Finally, a fusion method is proposed that through frequency-domain analysis, a candidate set of upscaling factors is given, and through analysis in the BDCT domain, the optimal estimation from all candidates is determined. The experimental results demonstrate that the proposed method outperforms other state-of-the-art methods.**

***Index Terms*—Image forensics, image resampling detection, JPEG block artifacts, scaling factor estimation.**

Manuscript received December 24, 2020; accepted March 25, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62072480, Grant U2001202, Grant U19B2022, and Grant U1736118; in part by the National Key Research and Development Program of China under Grant 2019QY2202 and Grant 2019QY(Y)0207; in part by the Key Areas Research and Development Program of Guangdong under Grant 2019B010136002 and Grant 2019B010139003; in part by the Key Scientific Research Program of Guangzhou under Grant 201804020068; and in part by Shenzhen Research and Development Program under Grant GJHZ20180928155814437. This article was recommended by Associate Editor Q. M. J. Wu. *(Corresponding author: Wei Lu.)*

Wei Lu, Qin Zhang, and Shangjun Luo are with the School of Computer Science and Engineering, Guangdong Province Key Laboratory of Information Security Technology, Ministry of Education Key Laboratory of Machine Intelligence and Advanced Computing, Sun Yat-sen University, Guangzhou 510006, China (e-mail: luwei3@mail.sysu.edu.cn; zhangq359@mail2.sysu.edu.cn; luoshj9@mail2.sysu.edu.cn).

Yicong Zhou is with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: yicongzhou@um.edu.mo).

Jiwu Huang is with the Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China, also with the Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060, China, and also with the Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518055, China (e-mail: jwhuang@szu.edu.cn).

Yun-Qing Shi is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: shi@njit.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2021.3069999.

Digital Object Identifier 10.1109/TCYB.2021.3069999

I. Introduction

In THE past two decades, the widespread use of digital cameras, along with image editing software [1], has given rise to the demand of an automatic detector for tampered images, which promotes the development of the forensics field [2]–[4]. As a kind of passive image tampering detection method, resampling forensics refers to the detection of image geometric transformations (e.g., scaling, rotation, and shearing), which are actually achieved by numerical interpolation and resampling. Resampling detection is one of the most important topics in forensics, because malicious tampering is usually performed by geometrically adapting some new image elements to the original scene [5], and such adaption may require the employment of geometric transformations.

Despite the diversity of the proposed methods, most of the available detectors share a common processing structure. In the first step, a residual signal from the observed image is extracted as the feature of detecting resampling artifacts. Depending on the foundation of each method, this signal can be obtained in different ways. In the observation of the specific periodic correlations between the pixels of resampling images, Popescu and Farid [6] proposed a global predictor to extract the residual signal. With an attempt at reducing the computational complexity, their method was later improved by Kirchner [7] with a fast and local linear predictor, where an automatic detector based on the maximum gradient of the p-maps spectrum was proposed. Gallagher [8] and Mahdian and Saic [9] proved that interpolated signals and their derivatives exhibit periodicity in their second-order statistics, where resampling artifacts are observable. Besides, the residual signal including resampling traces can also be obtained by computing the difference of predictor coefficients stemming from adjacent rows/columns of the image [10].

In the second step, given the residual signal extracted as mentioned above, a decision on whether the observed image has been resampled can be rendered according to the diverse criterion of each method. For example, a postprocessing step is applied in the frequency domain to detect the presence of spectral peaks, which is related to the periodicities introduced by the resampling process [8], [9]. In addition to applying postprocessing in the frequency domain, some approaches (e.g., [11], [12]) directly check if a group of candidate resampling factors satisfies the underlying linear relationship induced by resampling and interpolation operation. Apart from the spectrum-based methods, other detectors

avoid the frequency-based analysis by making use of a support vector machine (SVM) to take the final decision. For instance, in [13] a set of features is gathered from the normalized energy density for varying window size of the image. While in [14], Vázquez-Padín *et al.* derived a detector capable of discriminating between upscaled images and genuine images on the basis of SVD as well. This method was further improved in [15]. Finally, most of these detection approaches can also be oriented toward the parameter estimation of the tampering operation, based on the fact that the frequency of spectral peaks of the residual signal is directly related to the resampling factor. One of the best estimators of the resampling factor is proposed by Liu and Kirchner [16], in which an end-to-end CNN framework is trained to estimate the factor of a 64×64 image patch. Such CNN-based methods [17]–[19] have attracted a lot of interest in recent years. However, it is still challenging for downscaling factor estimation because the characteristics of the downscaling scenario are much weaker [20].

The remaining problem of resampling forensics is the robustness to lossy compression, for example, JPEG compression. Regardless of whether JPEG compression is implemented before or after resampling operation, the performance of the resampling detector will be seriously influenced, only if the quality factor (QF) of JPEG compression is low enough. On the one hand, JPEG compression is acting as the low pass filter in the frequency domain. On the other hand, periodic artifacts are introduced in the boundary of JPEG blocks, which have similar statistics as the resampling signals. The case of JPEG compressed images in the presence of resampling before compression has been mentioned from the very beginning [6], [8], [21]. Gallagher [8] proposed to ignore the influence of post-JPEG compression on resampling artifacts and regard JPEG compression as nearest neighbor interpolation, which introduces prominent peaks to the frequency domain in the location of $k/8$ (since the location of such JPEG peaks is fixed, there is little effect on the performance of the forensics detector as long as the characteristic peaks of resampling do not overlap with JPEG peaks). Most subsequent research works in this regard adopted this standpoint. However, our previous research [22] proposed that there are nonlinear coupling effects between upscaling and JPEG compression, shown as frequency mixing peaks. The case of JPEG compression prior to resampling was first studied by Kirchner and Gloe [23], which showed the spectral peaks that refer to pre-JPEG compression would be shifted by the following resampling operation. By detecting the shifted JPEG peaks, the resampling factor can be estimated. This theory was inherited by several other researchers [24]–[26]. In our previous research [26], we constructed a statistic based on the interval distribution of adjacent extremum points, which only respond to the JPEG blocking artifacts despite resampling operation. These statistics can help to estimate the resampling factor in the case of downscaling. As the combination of the previous two cases, the forensics of the resampling image with both pre-JPEG and post-JPEG compression are much more difficult than the uncompressed case. In our previous work [27], by denoising the investigated image to eliminate the JPEG blocking artifacts, we proposed a double-JPEG upscaled factor estimation scheme to make the resampling peaks more easily detected. Besides, the method of Bianchi and Piva [24] can reverse the operation chain of double-JPEG compression with the help of prior information on the resampling factor, given that the pre-JPEG compression is stronger than the post-JPEG compression. Nguyen and Katzenbeisser [28] found that the main difficulty of this case is the presence of many suspicious peaks in the spectrum. Yet, no further analyses and discussions could be found in these research works.

In this article, we concentrate on the upscaling factor estimation of double-JPEG compressed images in the presence of image upscaling between the two compressions (abbreviated as upscaled double-JPEG images in the following contents). The tampering operation chain pre-JPEG-upscaling-post-JPEG is decomposed into pre-JPEG-upscaling and upscaling-post-JPEG, which are integrated into the scaling model in our previous research [22], [26], respectively. For the case of upscaling-post-JPEG, we surpass our previous research [22] by giving a closed-form proof of the frequency mixing effect under the simplification of the quantization effect in the block discrete cosine transform (BDCT) domain. Then, the theories for pre-JPEG-upscaling model and upscaling-post-JPEG model are combined, and we present that there are more than 20 characteristic peaks in the spectrum of upscaled double-JPEG images, which can be classified into five groups. A numerical simulation of the compression process based on the 2-D autoregressive model of order 1 [AR(1) model] [15] of untampered images shows how the parameters of two JPEG compressions influence the relative amplitude of different characteristic peaks, along with various image contents. Since such influence is indeterminate for a given tampering case, it is hard to directly obtain the estimation of an upscaling factor from the frequency domain without any prior information. Through the analysis of nonaligned double JPEG compression, we prove that the unique estimation of the upscaling factor can be obtained in the BDCT domain with a proposed inverse scaling strategy. The main contributions of this work are as follows.

1) We present an in-depth analysis in the frequency domain of the second-order statistics of upscaled double JPEG images, which gives an exact formulation for the location of all characteristic peaks. Based on this formulation, a spectrum method is proposed to obtain some candidates of the scaling factor.
2) We propose an inverse scaling strategy to select the optimal estimation of the scaling factor between the candidates, which is based on the statistical model in the BDCT domain. The validity of the proposed strategy is proved mathematically, along with the robustness analysis under different scaling factors.

The remainder of this article is organized as follows. The analysis of upscaled double JPEG images in frequency domain and BDCT domain is shown in Sections II and III, respectively. Section IV gives the whole process of the proposed joint-domain fusion estimation method. Section V shows the experimental results of the proposed method and discusses

the robustness issues. Eventually, the concluding remarks and future works are given in Section VI.

II. Frequency-Domain Analysis of Scaled Images

In this section, a comprehensive analysis of scaling on JPEG images is presented in the frequency domain. There are four situations for scaled images when considering the operation chain decompression-scaling-compression:

1) scaling on genuine images, where both the untampered images and tampered images are stored in lossless-compression format;
2) scaling on pre-JPEG images, where the untampered images are stored in JPEG format and the tampered ones are in lossless-compression format;
3) scaling on post-JPEG images, where the untampered images are stored in lossless-compression format and the tampered ones are in JPEG format;
4) scaling on double-JPEG images, where both the untampered and tampered images are stored in JPEG format.

The first three situations have been deeply analyzed by the previous works and the fourth situation, that is, the frequency analysis of double-JPEG images, is the most important part of this article.

The following parts of this section are organized as follows. First, Section II-A briefly reviews the previous frequency analysis of the first three situations. Second, Section II-B combines the previous research to explain the complicated frequency structure of upscaled double-JPEG images. Using a simulation on the AR(1) model, we also show how different tampering parameters and image contents influence the frequency structure. The notation used hereafter is summarized in Table I.

TABLE I
Notation

$\mathbf{X}$	two-dimensional signal
x	one-dimensional signal
λ	resampling factor
$h(\bullet)$	interpolation function
$\mathrm{E}\{\bullet\}$	expectation operator
$\mathrm{Cov}\{\bullet\}$	covariance operator
ω	frequency of spectrum peak
$\mathrm{frac}(\bullet)$	function to return the fractional part of real number
$\lfloor\bullet\rfloor$	round-down function
e_x	JPEG quantization error of x
$\mathbf{r}_x$	auto-correlation function of x
$\mathbf{R}_x$	the Fourier transformation of $\mathbf{r}_x$
T	period
$\mathbf{N}$	white noise matrix
$\mathbf{U}$	AR correlation matrix
ρ	one-step correlation coefficients
σ_x	standard deviation of x
q	quantization step
QF	quantization factor of JPEG
$\mathbf{x}$	vector form of signal x
$\mathbf{y}$	vector form of signal in BDCT domain
$\mathcal{Q}(\bullet)$	quantization operator
$\mathcal{D}(\bullet)$	dequantization operator
$\mathbf{D}$	matrix of BDCT transform
$\mathbf{D}^{-1}$	matrix of IBDCT transform
$\Pr\{\bullet\}$	probability measure
$\mathbf{\Lambda}$	resampling matrix
$f(\bullet)$	probability distribution function
$r(\bullet)$	row dominant ratio
H	Integer Periodicity Maps

A. Review of the Prior Works

Usually, the digital image is stored as a 2-D digital matrix with finite size and finite set of values, for example, $\mathbf{X} = [X_{i,j}] \in \mathbb{Z}_l^{M\times N}$, where M and N are the size of matrix and l is the number of the gray level. Since both scaling operation and BDCT transform are linear and separable in each space dimension, to make the frequency domain analysis more compact, we model the digital image as a 1-D discrete signal with infinite length, denoted as $x_0(n) : \mathbb{Z} \to \mathbb{R}$. For genuine images, this discrete signal $x_0(n)$ is a sample of analog signal $x_0(t) : \mathbb{R} \to \mathbb{R}$.

According to Gallagher [8], when performing the scaling operation with a factor of λ on $x_0(n)$, the scaled signal $x_1(n)$ could be expressed as

$$x_1(n) = \sum_{i\in\mathbb{Z}} x_0(i)h\Big(\frac{n}{\lambda} - i\Big) \tag{1}$$

where $h(\bullet)$ is the low-pass filter for interpolation, also known as "interpolation kernel." This function is symmetric around 0 and has finite support [26], for example, $h(x) \neq 0 \iff x \in [-d, d]$.

The genuine images are the original output of the digital camera, in which prominent random components can be found even if the input optical field is a smooth function [29]. Without loss of generality, we only consider the random components of genuine images and assume that the genuine signal $x_0(n)$ is wide-sense stationarity

$$m_{x_0}(n) = E\{x_0(n)\} = \mu \tag{2}$$

$$c_{x_0}(n,\tau) = \mathrm{Cov}\{x_0(n), x_0(n+\tau)\} = c_0(\tau) \tag{3}$$

where $E\{\bullet\}$ and $\mathrm{Cov}\{\bullet\}$ are the expectation operator and covariance operator for the random process, respectively. Then, the scaled signal $x_1(n)$ should be a two-ordered cyclostationary process [30], which means the autocorrelation function of it varies periodically with the variable n, and prominent spectral lines can be found in the DFT spectrum of the autocorrelation function with the frequencies of

$$\omega_{\mathrm{rs}}^{(i)} \triangleq \mathrm{frac}\left(\frac{i}{\lambda}\right) \quad \text{and} \quad \omega_{\mathrm{rs,sy}}^{(i)} \triangleq 1 - \omega_{\mathrm{rs}}^{(i)} \tag{4}$$

where

$$\mathrm{frac}(x) \triangleq x - \lfloor x \rfloor \tag{5}$$

where $\lfloor\bullet\rfloor$ is the round-down function, i is the order of harmonic waves, and $\omega_{\mathrm{rs,sy}}^{(i)}$ are symmetric peaks of $\omega_{\mathrm{rs}}^{(i)}$. To see the whole proof of (4), refer to our previous research works [22], [26].

The higher the order of harmonic, the weaker the corresponding peak is. Usually, only the first harmonic peaks $\omega_{\mathrm{rs}}^{(1)}$ are prominent in $\mathbf{R}_{x_1}(\omega,\tau)$. For the case of upscaling, the general factor estimation method is to take the highest peak between frequencies 0 and 0.5 as $\omega_{\mathrm{rs}}^{(1)}$, and with (4), the estimation would be

$$\tilde{\lambda}_1 = \frac{1}{\omega_{\mathrm{rs}}^{(1)}} \quad \text{and} \quad \tilde{\lambda}_2 = 1 - \tilde{\lambda}_1 \tag{6}$$

where $\tilde{\lambda}_1 > 2$ and $1 < \tilde{\lambda}_2 < 2$, as a consequence of aliasing. In most previous research works, the aliasing result was not distinguished anymore.

The model of image scaling on pre-JPEG images is a simple generalization of the genuine one mentioned above. It is well known that the quantization in BDCT domain introduces block artifacts in the boundary of each block [31]–[33]. Robertson and Stevenson [31] proved that the JPEG quantization error $e_{x_0}(n)$ in the space domain is a two-ordered cyclostationary process if the genuine signal is wide-sense stationary

$$\mathbf{r}_{e_{x_0}}(n,\tau) = \mathbf{r}_{e_{x_0}}(n+8k,\tau) \quad \forall k \in \mathbb{Z} \tag{7}$$

where $\mathbf{r}_{e_{x_0}}(n,\tau)$ is the autocorrelation function of e_{x_0}.

Denoting the scaled version of $e_{x_0}(n)$ as $e^{\lambda}_{x_0}(n)$ and recalling (1), we have

$$e^{\lambda}_{x_0}(n) = \sum_{i\in\mathbb{Z}} e_{x_0}(i) h\Big(\frac{n}{\lambda} - i\Big). \tag{8}$$

Liu *et al.* [26] formulated the relationship between the correlation function of $e^{\lambda}_{x_0}(n)$, denoted as $\mathbf{r}_{e^{\lambda}_{x_0}}$, and λ in the case of scaling an image after JPEG compression

$$\mathbf{r}_{e^{\lambda}_{x_0}}(n+8\lambda k,\tau) = \mathbf{r}_{e^{\lambda}_{x_0}}(n,\tau) \quad \forall k \in \mathbb{Z} \tag{9}$$

which means $e^{\lambda}_{x_0}$ is two-order cyclostationary in a period of $T = 8\lambda$. Based on (9), Liu *et al.* [26] found the connection between the spectral peaks of scaled pre-JPEG images, namely, shifted JPEG peaks ω_{sfjp}, and the scaling factor λ

$$\omega^{(i)}_{\text{sfjp}} = \frac{i}{8\lambda}. \tag{10}$$

In our previous work [22], apart from JPEG peaks ω_{jp} and scaling peaks ω_{rs}, we find a third kind of peak in the frequency spectrum of upscaled post-JPEG images. We define these peaks as JPEG-scaling-mixing peaks

$$\omega^{(i)}_{\text{jp-rs-mix}} \triangleq \omega^{(1)}_{\text{rs}} + \omega^{(i)}_{\text{jp}} \triangleq \text{frac}\Big(\frac{1}{\lambda} + \frac{i}{8}\Big). \tag{11}$$

In this previous work [22], we also ignore the quantization effect in low BDCT channels and propose a BDCT high channels vanish model based on the assumption that the coefficients in high channels are discarded by the JPEG compression. This model is facilitated to capture the essence of the frequency mixing effect. Based on the BDCT high channels vanish model, the operation of JPEG compression is simplified to a linear operator. Equation (11) can be proved by a classical analysis in the DFT domain. More details can be found in [22].

B. Frequency-Domain Analysis of Scaling on Double-JPEG Images

Combining the conclusion in the previous works together, we obtain two qualitative rules for JPEG quantization and scaling of two-order cyclostationary signals.

1) *Frequency Translation Rule:* When a two-order cyclostationary signal with period T is scaled by interpolation in a factor of λ, the period of the output signal is scaled to $T\lambda$

$$x_1(n) = I_{\lambda}\Big\{x_0^{[T]}(n)\Big\} = x_0^{[T\lambda]}(n) \tag{12}$$

where $I_{\lambda}\{\bullet\}$ is the operator of scaling, and the superscript $[T]$ denotes the cyclostationary period. Especially, the discrete stationary signal can be considered as cyclostationary with $T = 1$

2) *Frequency Mixing Rule*: When a two-order cyclostationary signal with period T is compressed by any blockwise linear transformation and quantized in the transform domain and then decompressed, the introduced compression error is also cyclostationary in a period of frequency mixing

$$x_1(n) = J_{T_1}\Big\{x_0^{[T_2]}(n)\Big\} = x_0^{[T_2]}(n) + e^{[\text{mix}(T_1,T_2)]}_{x_0^{[T_2]}}(n) \tag{13}$$

where $J_{T_1}\{\bullet\}$ is the operator of quantization in the BDCT domain, the subscript $[T_1]$ is the size of the DCT block, and the frequency mixing term is the summation of different periods

$$e^{[\text{mix}(T_1,T_2)]}_{x_0^{[T_2]}}(n) = \sum_k e^{\left[\frac{T_1T_2}{T_1+kT_2}\right]}_{x_0^{[T_2]}}(n). \tag{14}$$

Especially, the period T_1 of blockwise compression is also included in $\text{mix}(T_1, T_2)$.

Applying the two rules on the inspected operation chain pre-JPEG-upscaling-post-JPEG with a stationary signal $x_0(n)$ as input, we have

$$\begin{aligned} x_2(n) &= J_8\{I_{\lambda}\{J_8\{x_0(n)\}\}\} \\ &= J_8\Big\{I_{\lambda}\Big\{x_0(n) + e^{[8]}_{x_0}\Big\}\Big\} \\ &= J_8\Big\{x_0^{[\lambda]}(n) + e^{[8\lambda]}_{x_0}(n)\Big\} \\ &= x_0^{[\lambda]}(n) + e^{[8\lambda]}_{x_0}(n) + e^{[\text{mix}(8,\lambda)]}_{x_0^{[\lambda]}}(n) + e^{[\text{mix}(8,8\lambda)]}_{e^{[8\lambda]}_{x_0}}(n) \end{aligned} \tag{15}$$

which means there are five kinds of intrinsic cyclostationary components in the output signal $x_2(n)$, shown as five kinds of peaks in the variance spectrum.

1) *Scaling peaks*, denoted as $\omega^{(i)}_{\text{rs}} \triangleq \text{frac}(i/\lambda)$, and in most time only the first harmonic $\omega^{(1)}_{\text{rs}}$ is prominent. These peaks are caused by the term $x_0^{[\lambda]}(n)$.
2) *JPEG peaks*, denoted as $\omega^{(i)}_{\text{jp}} \triangleq \text{frac}(i/8)$, and the amplitude for different i is almost the same. These peaks are caused by the post-JPEG compression noise, which come from the last two terms $e^{[\text{mix}(8,\lambda)]}_{x_0^{[\lambda]}}(n)$ and $e^{[\text{mix}(8,8\lambda)]}_{e^{[8\lambda]}_{x_0}}(n)$.
3) *JPEG-scaling-mixing peaks*, denoted as $\omega^{(i)}_{\text{jp-rs-mix}} \triangleq \omega^{(1)}_{\text{rs}} + \omega^{(i)}_{\text{jp}}$, and the amplitude for different i seems indeterminate. These peaks are caused by the frequency mixing between scaling signal and post-JPEG compression noise, shown as the third term $e^{[\text{mix}(8,\lambda)]}_{x_0^{[\lambda]}}(n)$.
4) *Shifted JPEG peaks*, denoted as $\omega^{(i)}_{\text{sfjp}} \triangleq \text{frac}(i/8\lambda)$. These peaks are caused by the rescaled pre-JPEG noise, shown as the second term $e^{[8\lambda]}_{x_0}(n)$.

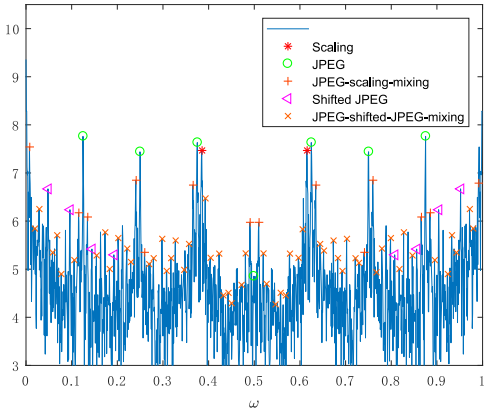


Fig. 1. Spectrum of Lena with double JPEG compression and upscaling. $QF1 = QF2 = 70$, $\lambda = 2.6$. The feature peaks have been marked. y axis is in logarithmic scale.

5) *JPEG-shifted-JPEG-mixing peaks*, denoted as $\omega^{(i,j)}_{\text{jp-sfjp-mix}} \triangleq \text{frac}(\omega^{(i)}_{\text{sfjp}} + \omega^{(j)}_{\text{jp}})$. These peaks are caused by the frequency mixing between scaled pre-JPEG noise and post-JPEG compression noise, shown as the last term $e^{[\text{mix}(8,8\lambda)]}_{e^{[8\lambda]}_{x_0}}(n)$.

One example of the test image Lena, which undergoes double JPEG compression and scaling, is shown in Fig. 1, with five different peaks marked. Obviously, almost all of the prominent peaks have been captured by our proposed model.

Intuitively, one would consider that there should be an analytical expression for the amplitude of different peaks with parameters of tampering as variables, since the analyses in the previous section are quantitative. Unfortunately, this idea does not work, because the error of JPEG compression has been simplified in our analysis, which is the BDCT high channels vanish model. In fact, there are more than five influencing factors in the operation chain pre-JPEG-upscaling-post-JPEG, including the QF of pre-JPEG $QF1$, QF of post-JPEG $QF2$, scaling factor λ, the kind of scaling kernel, and the contents of genuine image. The last one within them, the image contents, is the hardest one to grasp and is also the most important one. One example is that if the genuine signal $x_0(n)$ is white noise, the out signal $x_2(n)$ will only show the periodicity of scaling λ, discarding the blocking artifacts of JPEG compression.

In order to thoroughly analyze the influences on the spectrum of double-JPEG upscaled images affected by different parameters, inspired by the recent research [15] of signal scaling, we choose an AR(1) model as the genuine signal $x_0(n)$. The AR(1) model is a mathematical tool to simulate the texture and luminance characteristics of nature images. The content of the simulated image is subjected to its parameters. The inherent complexity of this forensics problem can be shown in the spectrum of the tampered AR(1) signal when we change the parameters of tampering operation and image contents. The 2-D AR(1) model is, in fact, a finite impulse response of 2-D white noise

$$\mathbf{X} = \mathbf{U}\mathbf{N}\mathbf{U}^T \tag{16}$$

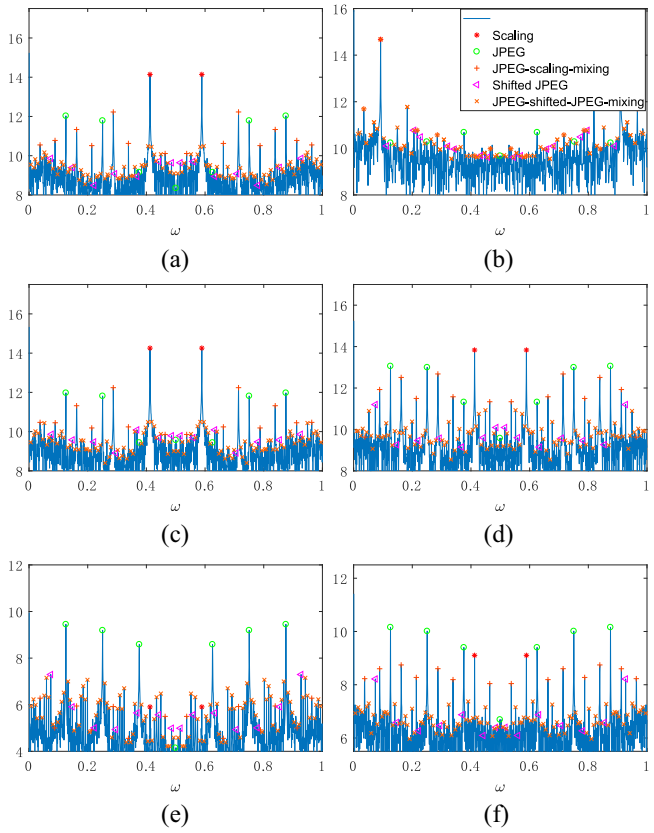


Fig. 2. Simulation of AR(1) model. $N = 1000$, $Q = 100$, and the variance σ^2_X is normalized by 255. (a) is the baseline with $QF1 = QF2 = 70$, $\lambda = 1.7$, $\rho = 0.1$, and $\sigma^2_X = 0.9$. (b)–(f) only change one parameter of (a), respectively. (b) $\lambda = 1.1$. (c) $QF1 = 40$. (d) $QF2 = 40$. (e) $\sigma^2_X = 0.1$. (f) $\rho = 0.9$. Y axis is in logarithmic scale.

where $\mathbf{N}$ is an $(M + Q - 1) \times (M + Q - 1)$ random matrix with i.i.d random variable elements, for example, $\mathbf{N}(i, j) \sim \mathcal{N}(0, \sigma^2_N)$, and $\mathbf{U}$ is a Toeplitz matrix of size $M \times (M+Q-1)$

$$\mathbf{U}(i,j) = \begin{cases} \rho^{Q-1-(j-i)}, & \text{if } (j-1) \in [0, Q-1] \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where ρ is the one-step correlation coefficients, and Q is the length of the truncation window. There are five parameters of this simulation, for example, $QF1$, $QF2$, λ, ρ, and σ^2_X, where σ^2_X is the variance of $\mathbf{X}$ and is controlled by σ^2_N as $\sigma^2_X = \sigma^2_N/(1 - \rho^2)$.

The process of using the AR(1) model for parameter estimation on double-JPEG upscaled images can be summarized as follows. First, we use the AR(1) model to generate the texture image as mentioned above. Later, the operation chain is performed on the generated image, which is JPEG compression followed by upscaling and JPEG compression again. In this way, we obtain a double-JPEG upscaled image. Finally, the spectrum of this tampered image is extracted for frequency analysis. In each combination of five parameters, several samples of $\mathbf{X}$ are simulated to undergo tampering, and the estimation of variance spectrum is the average of all samples. The cases for six different combinations are shown in Fig. 2, where Fig. 2(a) is the control group and is different from each other with only one parameter.

Some conclusions can be drawn from this simulation. First, the decrease of λ would weaken the relative amplitude of suspicious peaks, as shown in Fig. 2(b), according to the experimental results on natural images. In fact, these suspicious peaks are almost invisible when $\lambda < 1$. This means the classical estimation method should work well when λ is not too high. Second, the image contents have great influence on the relative amplitude of all characteristic peaks, shown in Fig. 2(e) and (f). In some cases, especially when the image is too dim or smooth, it is hard to distinguish $\omega_{\text{rs}}^{(i)}$ from other characteristic peaks, which is the main difficulty of classical spectrum-based methods. However, enough information for parameter estimation is preserved in the low BDCT channels, which will be shown in the next section.

III. BDCT Domain Analysis of Scaled Double JPEG Images

From the previous analysis, we know that the power spectrum of upscaled double JPEG images has a complicated structure, shown as many suspicious characteristic peaks with different amplitude. The main reason is that the process of post-JPEG compression has strong nonlinear effect, resulting in a frequency mixing phenomenon for a periodic input signal. Although the relative amplitude of these characteristic peaks is undeterminable, it is possible to obtain some candidates of the upscaling factor with the help of the aforementioned formulations. The concrete process is described in Section IV.

What we are concerned about is how to select the optimal estimation of the upscaling factor from the candidates. Inspired by the classical research of nonaligned double JPEG compression (NA-DJPG) forensics [32], [33], we propose an inverse scaling strategy, which zooms the inspected signal by a factor of $1/\lambda$ and then investigates the distribution of DCT coefficients. This idea is straightforward because in the classical research of NA-DJPG, the gridding shift of pre-JPEG compression is estimated by enumerating all shift parameters, where the BDCT coefficients will indicate the quantization effect of pre-JPEG compression only if the BDCT transform is aligned with the pre-JPEG compression. The double JPEG compression with upscaling can be seen as a variant of NA-DJPG, where the gridding of pre-JPEG compression is not shifted but zoomed.

In the rest of this section, we will prove the validity of the inverse scaling strategy and discuss its robustness with respect to different scaling factors.

A. Model of BDCT Domain Quantization and Inverse Scaling

Without loss of generality, the following analysis is implemented on 1-D signals and 1-D BDCT transform. Assume that an original uncompressed signal $\mathbf{x}$ is quantized in the BDCT domain

$$\mathbf{y}^{(1)} = \mathcal{D}(\mathcal{Q}(\mathbf{Dx})) \quad \text{and} \quad \mathbf{x}^{(1)} = \mathbf{D}^{-1}\mathbf{y}^{(1)} \tag{18}$$

where $\mathbf{x}$ is an infinite vector of random variables

$$\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots]^T. \tag{19}$$

$\mathbf{D}$ and $\mathbf{D}^{-1}$ are the matrix of the BDCT transform and inverse BDCT transform, which are block diagonal with each submatrix in the diagonal as the standard 8×8 DCT transform matrix, and $\mathcal{Q}(\bullet)$ and $\mathcal{D}(\bullet)$ are the model quantization and dequantization processes. Considering that $\mathbf{x}$ has finite moment and infinite support, then the BDCT coefficients $\mathbf{y}^{(1)}$ are discrete random variables

$$\Pr\left\{\mathbf{y}_i^{(1)} = a\right\} > 0 \Longleftrightarrow a = mq_i, m \in \mathbb{Z} \tag{20}$$

where q_i is the quantization step of $\mathcal{Q}(\bullet)$. Then, we consider the process of upscaling with factor of λ_1

$$\mathbf{x}^{(2)} = \mathbf{\Lambda}^{\lambda_1}\mathbf{x}^{(1)} \tag{21}$$

$$\mathbf{\Lambda}^{\lambda_1}(i,j) = h(j/\lambda_1 - i). \tag{22}$$

This equation is the matrix form of (1), where $\mathbf{\Lambda}^{\lambda_1}$ is the upscaling matrix of a specific interpolate kernel, and is a sparse matrix for most common kernels.

Based on our inverse scaling strategy, the scaled signal $\mathbf{x}^{(2)}$ is zoomed by the second factor λ_2 and then BDCT transformed

$$\mathbf{x}^{(3)} = \mathbf{\Lambda}^{\lambda_2}\mathbf{x}^{(2)} \quad \text{and} \quad \mathbf{y}^{(3)} = \mathbf{D}\mathbf{x}^{(3)}. \tag{23}$$

Combining (23) with (22), we have

$$\begin{aligned}\mathbf{y}^{(3)} &= \mathbf{D}\mathbf{\Lambda}^{\lambda_2}\mathbf{\Lambda}^{\lambda_1}\mathbf{D}^{-1}\mathbf{y}^{(1)} \\ &= \mathbf{D}\mathbf{\Lambda}^{\lambda_2,\lambda_1}\mathbf{D}^{-1}\mathbf{y}^{(1)} \\ &= \mathbf{D}\mathbf{\Lambda}^{\lambda_2,\lambda_1}\mathbf{D}^{T}\mathbf{y}^{(1)} \\ &= \mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}\mathbf{y}^{(1)}\end{aligned} \tag{24}$$

where $\mathbf{\Lambda}^{\lambda_2,\lambda_1}$ is the transform matrix of a successive zooming operation, where λ_1 and λ_2 are the factor of first and second zooming, respectively. Consider that both BDCT transform and its matrix $\mathbf{D}$ are orthogonal; hence, the inverse BDCT transform matrix $\mathbf{D}^{-1}$ is equal to the transposed version $\mathbf{D}^T$. The matrix product $\mathbf{D}\mathbf{\Lambda}^{\lambda_2,\lambda_1}\mathbf{D}^T$ is the 2-D BDCT transform of the matrix $\mathbf{\Lambda}^{\lambda_2,\lambda_1}$, denoted as $\mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}$.

It is obvious that when $\lambda_2 = \lambda_1 = 1$, the transform matrix $\mathbf{\Lambda}^{\lambda_2,\lambda_1}$ is the identity matrix, hence $\mathbf{y}^{(3)} = \mathbf{y}^{(1)}$, the same as the situation of $\lambda_1 \in \mathbb{Z}, \lambda_2 = 1/\lambda_1$. For other situations, the transform matrix $\mathbf{\Lambda}^{\lambda_2,\lambda_1}$ is not an identity matrix if one of the scaling factors is not integer. A detailed discussion and proof of the successive geometric transformations can be found in the research of Chen *et al.* [30]. In the next section, we will discuss in which case the inverse scaling images have the most prominent characteristics in the BDCT domain.

B. Probability Distribution of BDCT Domain Quantization and Inverse Scaling

Once $\mathbf{\Lambda}^{\lambda_2,\lambda_1}$ is not an identity matrix, $\mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}$ should not be sparse, and the variable $\mathbf{y}^{(3)}$ that we finally inspect is a linear combination of $\mathbf{y}^{(1)}$

$$\mathbf{y}_i^{(3)} = \sum_j \mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}(i,j)\mathbf{y}_j^{(1)}. \tag{25}$$

The probability distribution function (PDF) of $\mathbf{y}_i^{(3)}$ can be derived by a multiple convolution of $\mathbf{y}_j^{(1)}$

$$f_{\mathbf{y}_i^{(3)}}(x) = f_{d(i,1)\mathbf{y}_1^{(1)}}(x) * f_{d(i,2)\mathbf{y}_2^{(1)}}(x) * \cdots * f_{d(i,j)\mathbf{y}_j^{(1)}}(x) * \cdots \tag{26}$$

where

$$f_{d(i,j)\mathbf{y}_j^{(1)}}(x) = \frac{1}{d(i,j)} f_{\mathbf{y}_j^{(1)}}\left(\frac{x}{d(i,j)}\right) \tag{27}$$

$$d(i,j) = \mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}(i,j). \tag{28}$$

Although this multiple convolution is intractable, an approximate analysis can be applied. According to the centralized energy property of BDCT, most nonzero elements in each row of $\mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}$ should be close to zero. We permutate the row elements of $\mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}$ in descending order

$$\{d(i,1), d(i,2), \ldots, \} = \{d(i,[1]), d(i,[2]), \ldots\}$$
$$d(i,[1]) \geq d(i,[2]) \geq d(i,[3]) \cdots \tag{29}$$

Without loss of generality, we assume that one element is dominant in the corresponding row, inspired by the concept of "diagonally dominant"

$$d(i,[1]) > \sum_{j>1} \left| d\left(i,[j]\right) \right|. \tag{30}$$

Hence, $\mathbf{y}_i^{(3)}$ in (26) can be considered as the summation between the prominent element and other elements

$$\mathbf{y}_i^{(3)} = \mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}(i,[1])\mathbf{y}^{(1)}_{[1]} + \overline{\mathbf{y}}^{(1)}_{[1]} \tag{31}$$

where

$$\overline{\mathbf{y}}^{(1)}_{[1]} = \sum_{j>1} \mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}\left(i,[j]\right)\mathbf{y}^{(1)}_{[j]}. \tag{32}$$

Meanwhile, (26) can be rewritten as follows:

$$\begin{aligned} f_{\mathbf{y}_i^{(3)}}(x) &= f_{d(i,1)\mathbf{y}_1^{(1)}}(x) * f_{d(i,2)\mathbf{y}_2^{(1)}}(x) * \cdots * f_{d(i,j)\mathbf{y}_j^{(1)}}(x) * \cdots \\ &= f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) * f_{d(i,[2])\mathbf{y}^{(1)}_{[2]}}(x) \\ &\quad \times \cdots * f_{d(i,[j])\mathbf{y}^{(1)}_{[j]}}(x) * \cdots \end{aligned} \tag{33}$$

Here, we just rearrange the order of $f_{d(i,[j])\mathbf{y}^{(1)}_{[j]}}(x)$ in (26). Combined with (31) and (32), (33) can be replaced with the convolution between the PDF of the prominent element and other elements

$$\begin{aligned} f_{\mathbf{y}_i^{(3)}}(x) &= f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) \\ &\quad \times \left(f_{d(i,[2])\mathbf{y}^{(1)}_{[2]}}(x) * \cdots * f_{d(i,[j])\mathbf{y}^{(1)}_{[j]}}(x) * \cdots \right) \\ &= f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) * f_{d(i,[2])\mathbf{y}^{(1)}_{[2]} + \cdots + d\left(i,[j]\right)\mathbf{y}^{(1)}_{[j]} + \cdots}(x) \\ &= f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) * f_{\sum_{j>1}\mathbf{D}^{\Lambda}_{\lambda_2,\lambda_1}(i,[j])\mathbf{y}^{(1)}_{[j]}}(x) \\ &= f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) * f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x). \end{aligned} \tag{34}$$

Although each variable $\mathbf{y}^{(1)}_{[j]}$ is discrete random variable, the linear combination $\overline{\mathbf{y}}^{(1)}_{[1]}$ has degraded into continuous random variable in some sense, which converges uniformly to the Gaussian variable, according to CLT (central limit theorem). Equation (25) is then simplified as a summation of a discrete random variable with a continuous random variable, and the corresponding PDF (34) should be a single convolution

$$\begin{aligned} f_{\mathbf{y}_i^{(3)}}(x) &= f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) * f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x) \\ &= \sum_n f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(n) f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x-n) \\ &= \sum_n f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}\left(nd(i,[1])q_{[1]}\right) f_{\overline{\mathbf{y}}^{(1)}_{[1]}}\left(x - nd(i,[1])q_{[1]}\right) \\ &= \sum_n f_{[1]}\left(k_{i,[1]}n\right) f_{\overline{\mathbf{y}}^{(1)}_{[1]}}\left(x - k_{i,[1]}n\right) \end{aligned} \tag{35}$$

where

$$f_{[1]}(x) = f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x) \tag{36}$$

$$k_{i,[1]} = d(i,[1])q_{[1]} \tag{37}$$

where $k_{i,[1]}$ is the resized version of the quantization step $q_{[i]}$. The image of $f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x)$, which is the PMF of $d(i,[1])\mathbf{y}^{(1)}_{[1]}$, should be discrete spectral line with equal interval $k_{i,[1]}$. Because $f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x)$ is an unimodal distribution, once the width of $f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x)$ is smaller than $k_{i,[1]}$, the image of $f_{\mathbf{y}_i^{(3)}}(x)$ should be approximated to $f_{d(i,[1])\mathbf{y}^{(1)}_{[1]}}(x)$ with each discrete spectral line extended into a peak. Such image shows a trace of periodicity. In this situation, the information of the quantization step can be restored from the PDF. Besides, if the width of $f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x)$ is close or bigger than $k_{i,[1]}$, the shifted term $f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x - k_{i,[1]}n)$ will overlap with $f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x - k_{i,[1]}(n+1))$, which will eliminate the periodicity of $f_{\mathbf{y}_i^{(3)}}(x)$. Since the width of $f_{\overline{\mathbf{y}}^{(1)}_{[1]}}(x)$ is bounded by the variance of $\overline{\mathbf{y}}^{(1)}_{[1]}$, we propose a rough condition of which the distribution of $\mathbf{y}_i^{(3)}$ preserves the periodicity of first quantization

$$2\sigma_{\overline{\mathbf{y}}^{(1)}_{[1]}} < d(i,[1])q_{[1]} \tag{38}$$

$$\sigma^2_{\overline{\mathbf{y}}^{(1)}_{[1]}} = \sum_{j>1} d\left(i,[j]\right)^2 \sigma^2_{\mathbf{y}^{(1)}_{[j]}}. \tag{39}$$

Under the i.i.d assumption of $\mathbf{y}^{(1)}$, (38) can be further interpreted as

$$2\sigma_{\mathbf{y}^{(1)}} < r(i;\lambda_1,\lambda_2)q \tag{40}$$

$$r(i;\lambda_1,\lambda_2) = d(i,[1]) \Big/ \left(\sum_{j>1} d\left(i,[j]\right)^2 \right)^{1/2} \tag{41}$$

where $r(i;\lambda_1,\lambda_2)$ is defined as the "row-dominant ratio." Note that in (40), the variance term $\sigma_{\mathbf{y}^{(1)}}$ is determined by the image content, and the quantization step q is also fixed before investigated by the forensics estimator. Hence, the bigger the row dominant ratio $r(i;\lambda_1,\lambda_2)$ is, the more prominent the periodicity of $f_{\mathbf{y}_i^{(3)}}(x)$ is.

C. Numerical Simulation of BDCT Domain Quantization and Inverse Scaling

Since $r(i;\lambda_1,\lambda_2)$ is controlled by the two scaling factors λ_1 and λ_2, by enumerating different values of λ_2 and investigating the periodicity of $f_{\mathbf{y}_i^{(3)}}(x)$, we can estimate the value of λ_1. To further illustrate the relationship between $r(i;\lambda_1,\lambda_2)$ and λ_1,λ_2, numerical simulation is applied for different kernels and different scaling factors, and the result is shown as

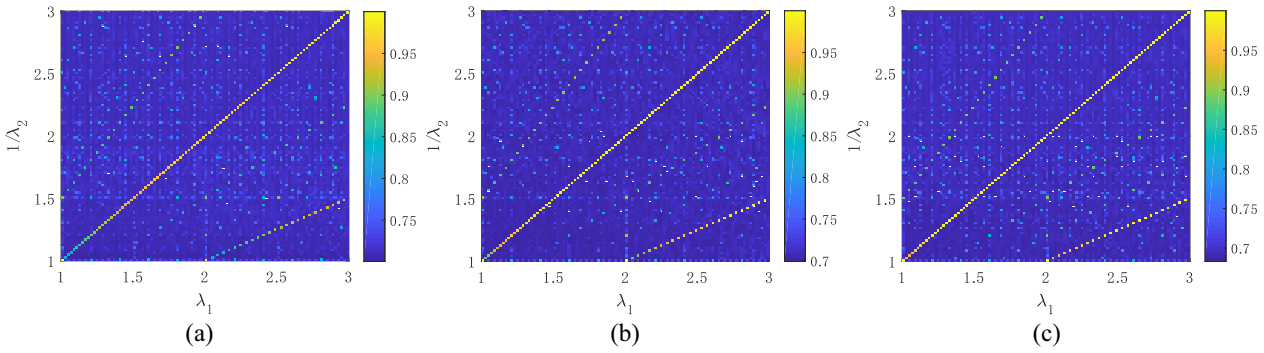


Fig. 3. Simulation of $\bar{r}_{\min}(\lambda_1, \lambda_2)$ for different kernels in form of pseudocolor maps. (a) Linear. (b) Cubic. (c) 1-D lanczos3.

a pseudocolor map in Fig. 3. The scaling matrix $\mathbf{\Lambda}^{\lambda_2,\lambda_1}$ with finite size is used to calculate $r(i; \lambda_1, \lambda_2)$. Since $r(i; \lambda_1, \lambda_2)$ is a vector for fixed λ_1, λ_2, the minimal of this vector is selected. To make the "row-dominant ratio" more visually pleasant, we transform it into a "normalized row-dominant ratio"

$$\bar{r}(i; \lambda_1, \lambda_2) = d(i, [1])/\left(\sum_{j\in\mathbb{Z}^+} d\big(i, [j]\big)^2\right)^{1/2}. \quad (42)$$

Note that compared to (41), the index of summation is changed. The minimal element of vector $\bar{r}(i; \lambda_1, \lambda_2)$ is selected

$$\bar{r}_{\min}(\lambda_1, \lambda_2) = \min_{i\in\mathbb{Z}^+} \bar{r}(i; \lambda_1, \lambda_2). \quad (43)$$

The result of numerical simulation shows that for fixed λ_1, the value of $r(i; \lambda_1, \lambda_2)$ achieves the maximum only when $\lambda_2 = 1/\lambda_1$, according with intuition. The diagonal element in the left bottom is smaller than these in the right upper, which implies a bad performance of the inverse scaling strategy when $\lambda_1 \approx 1$. However, this can be conquered if $\mathbf{\Lambda}^{\lambda_2}$ is selected as the pseudoinverse of $\mathbf{\Lambda}^{\lambda_1}$, which is applied in the proposed method. It is worth noting that in the right bottom of the figure, some points have similar values to the diagonal elements, which happens when $\lambda_2 = 2/\lambda_1$, namely, "half-inverse scaling." However, in experiments we found this have little effect on the final estimation.

IV. Proposed Method

In this section, a blind estimation method for the upscaling factor of the double JPEG image is proposed, which is based on the analysis of the spectrum structure and BDCT domain feature in previous sections. Note that the proposed method is a pure estimation method, because there has been many methods [13], [28], [34] for the classification and detection of image upscaling, and some of these methods are also very robust to pre-JPEG compression and post-JPEG compression. For a image under investigation, all of the information about post-JPEG compression can be acquired from the file header. The main problem is the estimation of the upscaling factor and parameters of pre-JPEG compression. To simplify our discussion, we assume that the image inspected has already been correctly classified as having undergone double-JPEG compression and upscaling between two compression. A block diagram of the proposed method is shown in Fig. 4.

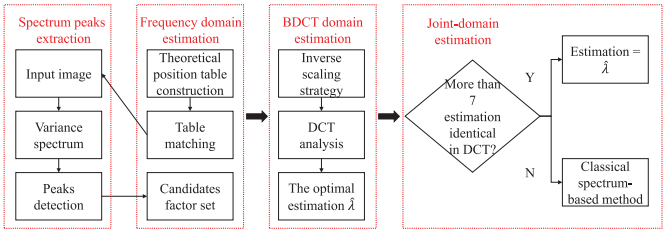


Fig. 4. Block diagram of the proposed method.

The proposed method has three main steps.

1) *Frequency-Domain Estimation:* The power spectrum of the investigated image is calculated, and several candidates of the upscaling factor are obtained.
2) *BDCT-Domain Estimation:* The inverse scaling strategy is applied to the investigated image, and the optimal estimation is selected from these candidates.
3) *Joint-Domain Estimation:* The confidence of the optimal estimation is examined to decide whether this estimation should be modified by the frequency-domain estimation.

A. Frequency-Domain Estimation

The situation of upscaled double JPEG image is far more complicated than upscaled genuine image. As shown in Fig. 1, more than 20 feature peaks exist in the spectrum of inspected image, and there is no clear winner among those peaks except for the JPEG peaks $\omega_{\text{jp}}^{(i)}$, which provide no useful information about scaling and pre-JPEG compression. Hence, we have to ignore the amplitude information of each feature peak and classification for each peak, and the spectrum is then reduced to a peaks-location-vector

$$v_f = [\omega_1, \omega_2, \ldots], v_f \in \mathbb{R}^n, \text{ s.t. } \omega_1 < \omega_2 < \cdots \quad (44)$$

The mapping from the upscaling factor λ to v_f is an injection, but not bijection. However, for a given vector v_f, the number of its preimages is finite. For a given double-compressed image, several candidates of upscaling factor λ can be selected by inversing this mapping. Inspired by Popescu and Farid's work [6], a searching-matching approach is used to solve this problem as follows.

1) A table is constructed that contains the theoretical positions of all the feature peaks discussed in Section III for

all possible value of the upscaling factor. Considering the fact that an over-upscaled image will be too blurred, we will restrict ourself to $1 < \lambda < 2.5$, with a step size $\Delta_\lambda = 0.01$. Not all the feature peaks are considered, since some of them are too weak. For the scaling peaks, only the first harmonic $\omega_{\text{rs}}^{(1)}$ is included. All of the JPEG peaks are ignored for their fixed location. All of the JPEG-scaling-mixing peaks $\omega_{\text{jp-rs-mix}}^{(i)}$ are included. For the shifted JPEG peaks $\omega_{\text{sfjp}}^{(i)}$, the first two are included, that is, $i = 1, 2$. For the JPEG-shift-JPEG-mixing peaks $\omega_{\text{jp-sfjp-mix}}^{(i,j)}$, only the ones related to the first shifted JPEG peaks are included, that is, $i = 1, j = 1, 2, \ldots, 7$. Finally, for each value of λ, 17 feature peaks are considered. To be brief, the table construction can be summarized in three steps as follows. First, for all the possible value of the upscaling factor, we generate a double-JPEG upscaled image with the corresponding factor and extract its variance spectrum by [8]. Second, we gather the location of the spectral peaks as mentioned above as the theoretical position of the corresponding factor. Finally, all theoretical positions of all the possible factors are added together to construct the table.

2) Calculate the spectrum of the image under investigation using the method in [8], denoted as $f_s(\omega)$. To be specific, the spectrum extraction step can be briefly summarized as two steps. The first step is to compute the second derivative of each row of the input image. The second step is to average over rows and compute the corresponding DFT. After the spectrum is extracted, the local maximum points in the spectrum are selected with a fixed interval of $\Delta_\omega = 0.01$. Both the location and amplitude of these maximum points are preserved, with the one corresponding to JPEG peaks excluded, resulting in a tuple sequence $v_s = [[\omega_1, f_s(\omega_1)], [\omega_2, f_s(\omega_2)], \ldots,]$. Then, the amplitude elements $f_s(\omega_i)$ in this sequence are normalized by the maximum one of them. The final result is denoted as v_s^n, which should include all of the feature peaks.
3) For each value of λ in the matching table, we count how many candidates in the tuple sequence v_s^n are located in the corresponding 17 theoretical positions. Since the spectrum of digital image is discrete and the location of one feature peak is a rational number, we consider a match if a candidate ω_i in v_s^n is around the two neighbors of the theoretical position. The counting result is also weighted by the normalized amplitude of each candidate $\log(f_s^n(\omega_i))$. Finally, the λ values with the top N highest count are taken as the candidates of optimal estimation and sent to the next step of estimation in the BDCT domain. Here, we set $N = 6$ and the result candidates set is denoted as $\Lambda^c = \{\lambda_1^c, \lambda_2^c, \ldots, \lambda_6^c\}$.

B. BDCT-Domain Estimation

Once the candidate set of the upscaling factor Λ^c is obtained, the inverse scaling strategy is applied to the inspected image $\mathbf{X}^{(0)}$, which is then transformed into BDCT coefficients

$$\mathbf{X}_i^{(1)} = \mathbf{\Lambda}^{\lambda_i^c}\mathbf{X}^{(0)}, \lambda_i^c \in \Lambda^c \tag{45}$$

$$\mathbf{Y}_{i,dx,dy}^{(1)} = \mathbf{D}_{dx,dy}\mathbf{X}_i^{(1)}, dx, dy \in \{0, 1, \ldots, 7\} \tag{46}$$

where $\mathbf{D}_{dx,dy}$ is the grid shifted version of the BDCT transform matrix, with dx and dy as the latent variable of grid shift in each dimension. Then, the empirical distribution function for each channel of BDCT coefficients is calculated

$$h(n; i, dx, dy, j) = \sum_{k,l} \chi_{\left\{n-0.5 \le \mathbf{Y}_{i,dx,dy,j}^{(1)}(k,l) \le n+0.5\right\}} \quad j \in 1, 2, \ldots, 9 \tag{47}$$

where $\chi_{\{\bullet\}}$ is the indicative function of the probability event, and $\mathbf{Y}_{i,dx,dy,j}^{(1)}$ is the jth channel of $\mathbf{Y}_{i,dx,dy}^{(1)}$, subscript j in a zigzag scanning order. Because the energy of natural image is concentrated on the low frequencies, only the first nine channels of $\mathbf{Y}_{i,dx,dy,j}^{(1)}$ are considered. To measure the quantization effects of the empirical distribution, we adopt the "integer periodicity maps" in [32], since it is unsupervised and robust. The quantization step q_i is an integer in most situations, and only the frequencies that are reciprocal of an integer are considered

$$H(m; i, dx, dy, j) = \Big| \sum_n h(n; i, dx, dy, j) e^{-j\frac{2\pi n}{m}} \Big|, m \in \mathbb{Z}. \tag{48}$$

It seems that the maximum element in $H(m; i, dx, dy, j)$ is the one corresponding to the optimal estimation. However, this assumption would be wrong in most situations, because $H(m; i, dx, dy, j)$ has not been normalized with respect to i and j, where $H(m; i, dx, dy, j)$ corresponding to the minimal λ_i^c and maximal m is prone to be the maximum element. Since the changes of grid shift dx, dy have little influence on the total energy, the spectrum of coefficients is normalized over these two dimensions

$$\overline{H}(m; i, dx, dy, j) = \frac{H(m; i, dx, dy, j)}{\sum_{dx', dy'} H(m; i, dx', dy', j)}. \tag{49}$$

Based on the normalized version, the optimal estimations $\hat{\lambda}_j$ of the upscaling factor are obtained for all channels j

$$\hat{\lambda}_j = \lambda_{\hat{i}_j}^c \quad \text{where} \quad \hat{i}_j = \arg\max_i \max_{dx,dy,m} \overline{H}(m; i, dx, dy, j). \tag{50}$$

C. Joint-Domain Estimation

Since the first nine channels in the BDCT domain are used for the optimal estimation, there are nine individual estimations obtained in the previous section, that is, $\hat{\Lambda} = \{\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_9\}$. In most cases, these nine estimations would not be completely identical. There are many possible reasons for this inconsistency.

First, the DC channel may perform the behavior of pre-JPEG compression no matter which factor of the inverse scaling is considered. This is caused by the pure color regions in the image, in which the neighboring $m \times m, m > 8$ pixels have the same color. For an 8×8 block in such a region, only the DC channel has nonzero value. No matter what the value of λ is, the grid of post-JEPG compression is aligned with pre-JPEG compression in such region. Hence, the distribution of DC channels in pure color regions is similar to the one of

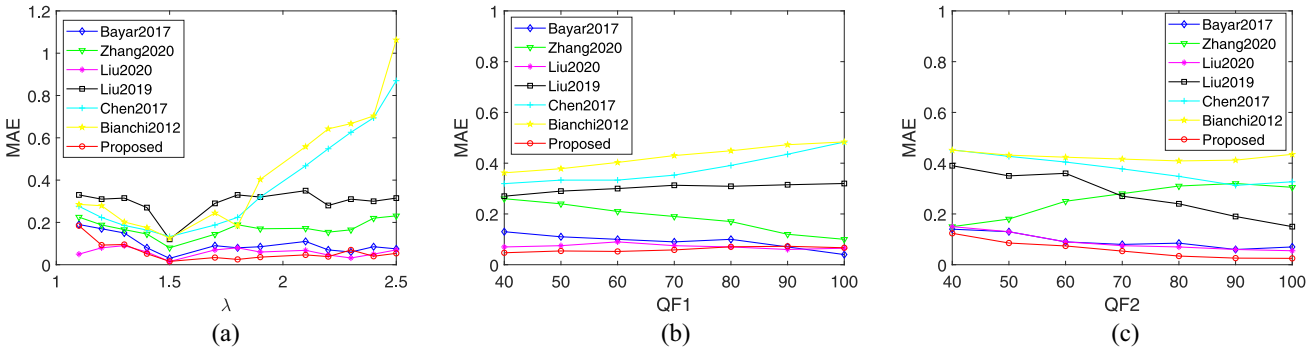


Fig. 5. Performance in terms of the MAE for upscaling factor estimation with different variables. (a) λ, averaged over all $QF1$ and $QF2$. (b) $QF1$, averaged over all $QF2$ and λ. (c) $QF2$, averaged over all $QF1$ and λ.

aligned double-JPEG images. For an image with much regions of pure color, the proposed BDCT domain estimation method would select random answer from the DC channel. A solution to this problem is directly ignoring the answer estimated by the DC channel.

Second, the noise introduced by post-JPEG compression is ignored in the inverse scaling strategy, which would destroy the quantization trace of pre-JPEG compression, especially when $QF2 < QF1$. This phenomenon has been verified in the research of nonaligned double-JPEG compression [32], [33]. When the trace of pre-JPEG compression is completely destroyed by post-JPEG compression, the BDCT domain method should lose its efficiency, which means that the estimation selected by each channel should be randomly picked from the candidates set Λ^c. Although we do not have any prior information about pre-JPEG compression, just by inspecting the consistency of nine estimation $\hat{\lambda}_j$, we can know whether these estimations are valid or not. In particular, if more than seven of $\hat{\lambda}_j$ are identical, the mode of them is adopted as the final estimation. Otherwise, these estimations are excluded. In this case, since $QF1 > QF2$, the trace of pre-JPEG compression is also very weak in the frequency spectrum, and the scaling peaks $\omega_{\text{rs}}^{(i)} = i/\lambda$ should be with the highest amplitude in the spectrum. Hence, the final estimation is acquired by the classical frequency-domain estimation method, for example, the one from [8].

V. Experiments

To evaluate the performance of the proposed method in a quantitative way, experiments are conducted over a large group of test sets, which differs in the parameters of tampering. These test sets originate from 500 different uncompressed images captured by Nikon cameras, which belong to the Dresden Image Database [35], in consistent with the previous research.

We pretreated the genuine images before tampering by graying and downsampling to avoid the influence of CFA interpolation [6], [23], [26], [30]. Then, the 500 grayscale images (1504×1000 pixels) are tampered by an operation chain of pre-JPEG-upscaling-post-JPEG. The corresponding parameters of the operation chain are the upscaling factor λ, the QF of pre-JPEG compression $QF1$, and the QF of post-JPEG compression $QF2$. Because the JPEG peak is in the same location as the first harmonic of $\lambda = \{1.6, 2.0\}$, all of the considered algorithms will fail in the situation, so the corresponding situations are ignored. The parameter set of the upscaling scale is $\Lambda = \{1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 1.8, 1.9, 2.1, 2.2, 2.3, 2.4, 2.5\}$, with the uniform distribution as priori assumption. Since the bicubic kernel is the most popular one on the Internet, only the bicubic kernel is considered. To avoid the influence of different image sizes, after the upscaling step, the tampered image is cropped with only the centered 1024×1024 pixels of upscaled version preserved. The possible values of $QF1$ and $QF2$ are taken from the same set $\{100, 90, \ldots, 40\}$. The resulted testing image sets have many subsets, denoted as $G\{i, j, k\}$. The variables i, j, and k refer to different parameters of $QF1$, λ, and $QF2$, respectively. For example, $G\{1, 13, 7\}$ have 500 tampered images (1024×1024 pixels), which suffer from upscaling with a factor of 2.5 and kernel of bicubic, and double-JPEG with $QF1$ of 100 and $QF2$ of 40. Finally, the test image sets include 318 500 images.

The proposed method is compared with the algorithm from [17] [22], and [24]–[27], with the default parameters mentioned in referring research. The one from [24] has three versions as "estimated," "prior," and "oracle." Since the proposed method is a blind estimation method of upscaling factor without any prior information of λ, only the "estimated" one of [24] is considered. To present the performance comprehensively, we take the performance in terms of the mean absolute error (MAE) and accuracy as criterion [26], [27]. The values of MAE and accuracy are calculated over each subsets $G\{i, j, k\}$, resulting in a 3-D performance table $E\{i, j, k\}$ including 637 elements.

This performance table is too big to be showed in the text; hence, it is processed by two kinds of data dimensionality reduction method. The first method is the "drill-up" operation, where $E\{i, j, k\}$ is averaged over any two dimensions, resulting in three 1-D tables, shown in Figs. 5 and 6. The second method is the "slicing" operation, which takes into account only one fixed value of λ, generating a 2-D table. Here, due to the limitation of this article, we present the situation of $\lambda = \{1.2, 1.5, 2.3\}$ with two algorithms [24], [25], which are most relevant to the research scenario, that is, JPEG-resampling-JPEG operation chain, shown in Tables II–IV,

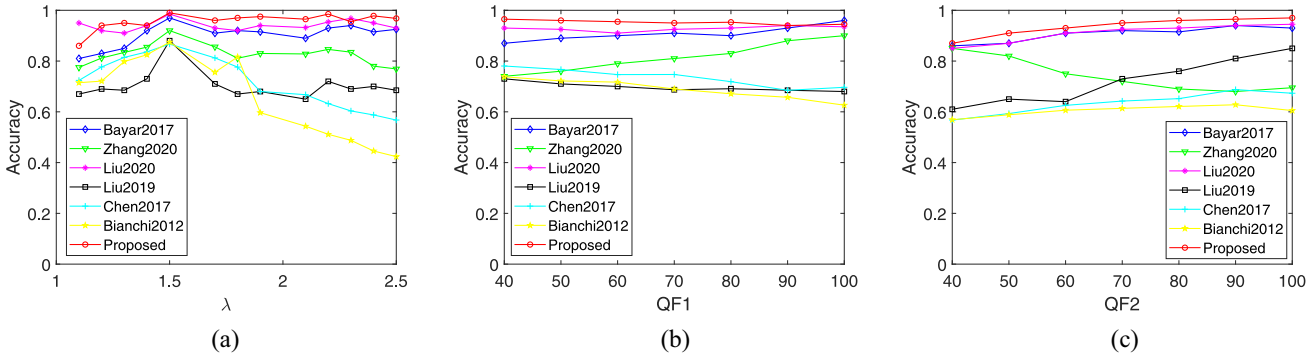


Fig. 6. Performance in terms of the accuracy for upscaling factor estimation with different variables. (a) λ, averaged over all $QF1$ and $QF2$. (b) $QF1$, averaged over all $QF2$ and λ. (c) $QF2$, averaged over all $QF1$ and λ.

TABLE II
PERFORMANCE IN TERMS OF THE MAE FOR UPSCALING FACTOR ESTIMATION. THE UPSCALING FACTOR IS 1.2

$QF1$	methods	$QF2$ 100	90	80	70	60	50	40
100	Ours	**0.0002**	**0.0108**	**0.0396**	**0.0859**	**0.0978**	**0.1596**	**0.2525**
	Bianchi [24]	0.3240	0.3212	0.3183	0.3178	0.3251	0.3121	0.3164
	Chen [25]	0.3004	0.3175	0.3055	0.2989	0.3024	0.3247	0.2932
90	Ours	**0.0160**	**0.0192**	**0.0179**	**0.0784**	**0.1106**	**0.1559**	**0.2395**
	Bianchi [24]	0.2719	0.3203	0.3215	0.3284	0.3037	0.3154	0.3264
	Chen [25]	0.1045	0.2732	0.3182	0.3212	0.3146	0.3123	0.2930
80	Ours	**0.0246**	**0.0170**	**0.0278**	**0.0399**	**0.0952**	**0.1446**	**0.2548**
	Bianchi [24]	0.2221	0.3014	0.3334	0.3475	0.3412	0.3230	0.3191
	Chen [25]	0.0500	0.1871	0.3014	0.3015	0.2894	0.3016	0.3058
70	Ours	**0.0183**	**0.0167**	**0.0533**	**0.0757**	**0.0830**	**0.1124**	**0.2179**
	Bianchi [24]	0.1806	0.2352	0.2954	0.3376	0.3303	0.3275	0.3205
	Chen [25]	0.0206	0.1297	0.2209	0.3048	0.2950	0.3147	0.3024
60	Ours	**0.0125**	**0.0213**	**0.0568**	**0.1068**	**0.0990**	**0.1125**	**0.1956**
	Bianchi [24]	0.1332	0.1734	0.2565	0.2992	0.3269	0.3208	0.3182
	Chen [25]	0.0165	0.0687	0.1668	0.2507	0.2700	0.2833	0.2948
50	Ours	**0.0072**	0.0444	**0.0929**	**0.1315**	**0.1354**	**0.1333**	**0.1861**
	Bianchi [24]	0.1164	0.1555	0.2156	0.2600	0.3029	0.3246	0.3265
	Chen [25]	0.0191	**0.0432**	0.1283	0.1916	0.2365	0.2594	0.3235
40	Ours	**0.0000**	0.0299	**0.0637**	**0.1338**	**0.1511**	**0.1778**	**0.1891**
	Bianchi [24]	0.1003	0.1295	0.1699	0.2330	0.2728	0.2982	0.2966
	Chen [25]	0.0202	**0.0256**	0.0800	0.1470	0.1868	0.2315	0.2832

TABLE III
PERFORMANCE IN TERMS OF THE MAE FOR UPSCALING FACTOR ESTIMATION. THE UPSCALING FACTOR IS 1.5

$QF1$	methods	$QF2$ 100	90	80	70	60	50	40
100	Ours	**0.0007**	**0.0196**	**0.0283**	**0.0356**	**0.0592**	**0.0636**	**0.0698**
	Bianchi [24]	0.2714	0.2210	0.1921	0.1966	0.1828	0.1732	0.1837
	Chen [25]	0.2739	0.2354	0.2594	0.2780	0.2675	0.2613	0.2769
90	Ours	**0.0003**	**0.0135**	**0.0175**	**0.0314**	**0.0402**	**0.0500**	**0.0608**
	Bianchi [24]	0.2344	0.2056	0.1669	0.1502	0.1808	0.1777	0.1949
	Chen [25]	0.0756	0.0754	0.2054	0.2666	0.2754	0.2841	0.2871
80	Ours	**0.0003**	**0.0048**	**0.0071**	**0.0237**	**0.0388**	**0.0246**	**0.0316**
	Bianchi [24]	0.1801	0.1698	0.1493	0.1120	0.1068	0.1247	0.1529
	Chen [25]	0.0413	0.0347	0.0764	0.1216	0.2176	0.2463	0.2732
70	Ours	**0.0024**	**0.0000**	**0.0034**	**0.0097**	**0.0245**	**0.0167**	**0.0112**
	Bianchi [24]	0.1389	0.1262	0.1223	0.1034	0.0847	0.0765	0.1065
	Chen [25]	0.0286	0.0343	0.0377	0.0796	0.1192	0.1746	0.2345
60	Ours	**0.0034**	**0.0009**	**0.0012**	**0.0032**	**0.0118**	**0.0072**	**0.0075**
	Bianchi [24]	0.1188	0.1073	0.1026	0.0890	0.0887	0.0769	0.0668
	Chen [25]	0.0274	0.0288	0.0460	0.0594	0.0979	0.1229	0.1722
50	Ours	**0.0121**	**0.0000**	**0.0000**	**0.0015**	**0.0070**	**0.0041**	**0.0072**
	Bianchi [24]	0.0899	0.0826	0.0901	0.0859	0.0761	0.0792	0.0651
	Chen [25]	0.0272	0.0321	0.0391	0.0523	0.0810	0.1008	0.1578
40	Ours	**0.0033**	**0.0000**	**0.0000**	**0.0010**	**0.0036**	**0.0012**	**0.0029**
	Bianchi [24]	0.0699	0.0640	0.0674	0.0690	0.0540	0.0549	0.0629
	Chen [25]	0.0304	0.0335	0.0386	0.0447	0.0616	0.0868	0.1277

TABLE IV
PERFORMANCE IN TERMS OF THE MAE FOR UPSCALING FACTOR ESTIMATION. THE UPSCALING FACTOR IS 2.3

$QF1$	methods	$QF2$ 100	90	80	70	60	50	40
100	Ours	**0.0005**	**0.0045**	**0.0133**	**0.0491**	**0.0605**	**0.0485**	**0.0818**
	Bianchi [24]	0.7171	0.5961	0.5677	0.5621	0.5841	0.6066	0.6033
	Chen [25]	0.6887	0.6365	0.6553	0.6684	0.6681	0.6879	0.7032
90	Ours	**0.0918**	**0.0302**	**0.0687**	**0.0572**	**0.0599**	**0.0458**	**0.0657**
	Bianchi [24]	0.7301	0.6067	0.5800	0.5754	0.6004	0.6135	0.6006
	Chen [25]	0.6470	0.5274	0.6274	0.6362	0.6873	0.6825	0.7201
80	Ours	**0.1006**	**0.0605**	**0.0463**	**0.0689**	**0.0700**	**0.0457**	**0.0866**
	Bianchi [24]	0.8327	0.6924	0.5408	0.6132	0.5764	0.5834	0.5981
	Chen [25]	0.6936	0.5326	0.5548	0.6333	0.6397	0.6678	0.6788
70	Ours	**0.0407**	**0.0586**	**0.0558**	**0.0717**	**0.0792**	**0.0586**	**0.0769**
	Bianchi [24]	0.8780	0.7683	0.6509	0.6128	0.5691	0.6181	0.6213
	Chen [25]	0.7143	0.5162	0.5308	0.5735	0.5539	0.5981	0.6403
60	Ours	**0.0434**	**0.0573**	**0.0533**	**0.1003**	**0.1081**	**0.0706**	**0.1232**
	Bianchi [24]	0.9513	0.8232	0.7172	0.6501	0.5986	0.5208	0.5787
	Chen [25]	0.7284	0.5741	0.5543	0.4951	0.5793	0.5926	0.6271
50	Ours	**0.0460**	**0.0517**	**0.0602**	**0.1079**	**0.1224**	**0.0820**	**0.1453**
	Bianchi [24]	0.9318	0.8398	0.8043	0.7049	0.6186	0.5391	0.5301
	Chen [25]	0.7734	0.6587	0.6258	0.5749	0.5989	0.6253	0.6062
40	Ours	**0.0126**	**0.0621**	**0.0903**	**0.1000**	**0.1156**	**0.1039**	**0.1596**
	Bianchi [24]	1.0281	0.8935	0.8219	0.7384	0.6255	0.5369	0.5325
	Chen [25]	0.7210	0.6241	0.5679	0.5670	0.6076	0.6078	0.6059

respectively. In each subset $G\{i, j, k\}$, the best result between three methods is marked in boldface.

Figs. 5 and 6 show the MAE and accuracy performance of several resampling factor estimation methods for varying parameters. As is shown in Fig. 5, the proposed method has better performance than other six methods in almost all situations and dimensions. First, the performance of the proposed method is robust with respect to different tampering parameters, which implies that the scaling peaks $\omega_{\text{rs}}^{(1)}$ are the most robust feature when the inspector has no more prior information. Second, the influence on the proposed method caused by $QF1$ and $QF2$ resembles the way $QF1$ and $QF2$ influence [24]–[26]. To be specific, when $QF1$ decreases, the performance is getting better. When $QF2$ decreases, the performance is getting worse. These methods rely on the feature of shifted JPEG peaks in the spectrum, which is caused by the pre-JPEG compression and would be weakened by the post-JPEG compression. While [22] is on the contrary, the performance of [22] increases as $QF1$ increases, and his performance degrades as $QF2$ increases. This is because his method bases on the location of $\omega_{\text{jp-rs-mix}}^{(i)}$. When $QF1$ decreases, the amplitude of shifted JPEG peaks increases, which means there are more noise peaks occurring in the spectrum, interfering the detection of resampling-JPEG mixture peaks. In the meanwhile, when $QF2$ increases, the JPEG block artifacts degrade, which weakens the amplitude of JPEG peaks and resampling-JPEG peaks. In this case, the performance of [22] gets worse. Besides, the performances of [27] and [17] are relatively stable whatever $QF1$ and $QF2$ are. This is because [17] is a CNN-based method, which is robust to JPEG compression, and [27] first denoises the investigated image to eliminate JPEG peaks while maintaining the resampling peaks only. Later, he divides the denoised image into several patches to estimate individually and produces final estimation in terms of each estimation in patch, which decreases the influence caused by noise and JPEG peaks, and hence, produces excellent performance.

Finally, the performance of [24] and [25] has a nonlinear relationship with λ. When λ is slightly bigger than 1, the increase of λ will contribute to the performance. But when λ is close to or bigger than 2, its increase has a negative influence on MAE. This puzzling phenomenon can be explained in two aspects. On the one hand, these two methods rely on the first shifted JPEG peak $\omega_{\text{sfjp}}^{(1)} = 1/8\lambda$, whose amplitude has a positive relationship with λ. The increase of λ would strengthen the significance of $\omega_{\text{sfjp}}^{(1)}$, hence helping for estimation. On the other hand, the amplitude of JPEG-scaling-mixing peaks $\omega_{\text{jp-rs-mix}}^{(i)}$ also has a positive relationship with λ. One of $\omega_{\text{jp-rs-mix}}^{(i)}$ is located in the frequency interval $[0, 1/8]$, which is easily confused with $\omega_{\text{sfjp}}^{(1)}$. When λ is slightly bigger than 1, the amplitude of $\omega_{\text{jp-rs-mix}}^{(i)}$ is quite small and the confusion is negligible. With the increase of λ, $\omega_{\text{jp-rs-mix}}^{(i)}$ is more prominent; hence, the confusion between it and $\omega_{\text{sfjp}}^{(1)}$ will become stronger and make the estimation error increase. Since the proposed method has considered all kinds of peaks in the spectrum, its performance is more robust to λ.

VI. Conclusion

In this article, we addressed the upscaling factor estimation of double-JPEG compressed images in the presence of image upscaling between the two compressions. We first highlighted the complicated spectrum structure of upscaled double-JPEG images, which is caused by the coupling effect between upscaling and double JPEG compressions. Specifically, we presented that there are five kinds of characteristic peaks in the spectrum, along with the exact formulation of frequencies derived from a simplified model of BDCT domain quantization. The proposed theory is verified by the simulation based on the AR(1) model of untampered images, which also shows the nondeterminacy for the relative amplitude of different peaks, influenced by image contents and tampering parameters. The confusion among different characteristic peaks makes it hard to estimate the factor only by the peaks location information. Inspired by the research of nonaligned double JPEG compression, the problem was then analyzed in the BDCT domain. We proved that the unique estimation of the upscaling factor can be obtained in the BDCT domain with a proposed joint-domain fusion estimation method. The experimental results have demonstrated that the proposed method outperforms the state-of-the-art methods reported in the current literature. However, in the case of strong post-JPEG compression and slight upscaling, the performance of the proposed method could be improved further. Therefore, our further research will try to integrate the supervised learning method and develop a more robust estimator for upscaled double-JPEG images.

References

[1] J. Shen, D. Wang, and X. Li, "Depth-aware image seam carving," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1453–1461, Oct. 2013.

[2] H. Zeng, J. Liu, J. Yu, X. Kang, Y. Q. Shi, and Z. J. Wang, "A framework of camera source identification Bayesian game," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1757–1768, Jul. 2017.

[3] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

[4] Z. Wang *et al.*, "Person reidentification via discrepancy matrix and matrix metric," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 3006–3020, Oct. 2018.

[5] A. Ferreira *et al.*, "Behavior knowledge space-based fusion for copy-move forgery detection," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4729–4742, Oct. 2016.

[6] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758–767, Feb. 2005.

[7] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in *Proc. ACM Workshop Multimedia Security*, Oxford, U.K., 2008, pp. 11–20.

[8] A. C. Gallagher, "Detection of linear and cubic interpolation in JPEG compressed images," in *Proc. Can. Conf. Comput. Robot Vis.*, Victoria, BC, Canada, May 2005, pp. 65–72.

[9] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 3, pp. 529–538, Sep. 2008.

[10] M. Kirchner, "Linear row and column predictors for the analysis of resized images," in *Proc. ACM Workshop Multimedia Security*, Roma, Italy, 2010, pp. 13–18.

[11] Y. T. Kao, H. J. Lin, C. W. Wang, and Y. C. Pai, "Effective detection for linear up-sampling by a factor of fraction," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3443–3453, Aug. 2012.

[12] D. Vázquez-Padín, P. Comesaña, and F. Pérez-González, "Set-membership identification of resampled signals," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, Guangzhou, China, Nov. 2013, pp. 150–155.

[13] X. Feng, I. J. Cox, and G. Doerr, "Normalized energy density-based forensic detection of resampled images," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 536–545, Jun. 2012.

[14] D. Vázquez-Padín, P. Comesaña, and F. Pérez-González, "An SVD approach to forensic image resampling detection," in *Proc. Eur. Signal Process. Conf.*, Nice, France, 2015, pp. 2067–2071.

[15] D. Vázquez-Padín, F. Pérez-González, and P. Comesaña-Alfaro, "A random matrix approach to the forensic analysis of upscaled images," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 9, pp. 2115–2130, Sep. 2017.

[16] C. Liu and M. Kirchner, "CNN-based rescaling factor estimation," in *Proc. ACM Workshop Inf. Hiding Multimedia Security*, Paris, France, 2019, pp. 119–124.

[17] B. Bayar and M. C. Stamm, "A generic approach towards image manipulation parameter estimation using convolutional neural networks," in *Proc. ACM Workshop Inf. Hiding Multimedia Security*, Philadelphia, PA, USA, 2017, pp. 147–157.

[18] J. Han, H. Chen, N. Liu, C. Yan, and X. Li, "CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3171–3183, Nov. 2018.

[19] W. Wu, Y. Yin, X. Wang, and D. Xu, "Face detection with different scales based on faster R-CNN," *IEEE Trans. Cybern.*, vol. 49, no. 11, pp. 4017–4028, Nov. 2019.

[20] C. Pasquini and R. B$^{\text{'}}$ohme, "Information-theoretic bounds for the forensic detection of downscaled signals," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1928–1943, Jul. 2019.

[21] X. Liu, W. Lu, T. Huang, H. Liu, Y. Xue, and Y. Yeung, "Scaling factor estimation on JPEG compressed images by cyclostationarity analysis," *Multimedia Tools Appl.*, vol. 78, no. 7, pp. 7947–7964, 2019.

[22] Q. Zhang, W. Lu, T. Huang, S. Luo, Z. Xu, and Y. Mao, "On the robustness of JPEG post-compression to resampling factor estimation," *Signal Process.*, vol. 168, Mar. 2020, Art. no. 107371.

[23] M. Kirchner and T. Gloe, "On resampling detection in re-compressed images," in *Proc. Int. Workshop Inf. Forensics Security*, London, U.K., Dec. 2009, pp. 21–25.

[24] T. Bianchi and A. Piva, "Reverse engineering of double JPEG compression in the presence of image resizing," in *Proc. Int. Workshop Inf. Forensics Security*, Tenerife, Spain, 2012, pp. 127–132.

[25] Z. Chen, Y. Zhao, and R. Ni, "Detection of operation chain: JPEG-resampling-JPEG," *Signal Process. Image Commun.*, vol. 57, pp. 8–20, Sep. 2017.

[26] X. Liu, W. Lu, Q. Zhang, J. Huang, and Y.-Q. Shi, "Downscaling factor estimation on pre-JPEG compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 618–631, Mar. 2020.

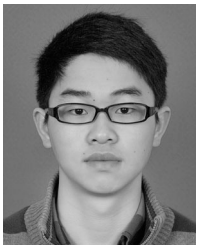
[27] X. Liu, W. Lu, Y. Xue, and Y. Yeung, "Upscaling factor estimation on double JPEG compressed images," *Multimedia Tools Appl.*, vol. 79, nos. 19–20, pp. 12891–12914, 2020.

[28] H. C. Nguyen and S. Katzenbeisser, "Detecting resized double JPEG compressed images-using support vector machine," in *Proc. Int. Conf. Commun. Multimedia Security*, Magdeburg, Germany, 2013, pp. 113–122.
[29] T. H. Thai, R. Cogranne, and F. Retraint, "Camera model identification based on the heteroscedastic noise model," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 250–263, Jan. 2014.
[30] C. Chen, J. Ni, Z. Shen, and Y. Q. Shi, "Blind forensics of successive geometric transformations in digital images using spectral method: Theory and applications," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2811–2824, Jun. 2017.
[31] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 27–38, Jan. 2005.
[32] T. Bianchi and A. Piva, "Detection of nonaligned double JPEG compression based on integer periodicity maps," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 842–848, Apr. 2012.
[33] W. Wang, J. Dong, and T. Tan, "Exploring DCT coefficient quantization effects for local tampering detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 10, pp. 1653–1666, Oct. 2014.
[34] L. Li, J. Xue, and Z. Tian, "Moment feature based forensic detection of resampled digital images," in *Proc. ACM Int. Conf. Multimedia*, Barcelona, Spain, 2013, pp. 569–572.
[35] T. Gloe and R. Bohme, "The 'Dresden Image Database' for benchmarking digital image forensics," in *Proc. ACM Symp. Appl. Comput.*, vol. 2, 2010, pp. 1584–1590.

Wei Lu (Member, IEEE) received the B.S. degree in automation from Northeast University, Shenyang, China, in 2002, and the M.S. degree and the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2005 and 2007, respectively.

He was a Research Assistant with Hong Kong Polytechnic University, Hong Kong, from 2006 to 2007. He is currently a Professor with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include multimedia forensics and security, data hiding and watermarking, and privacy protection.

Prof. Lu is an Associate Editor for the *Signal Processing* and the *Journal of Visual Communication and Image Representation*.



Qin Zhang received the B.S. degree from the School of Physics, Sun Yat-sen University, Guangzhou, China, in 2016, and the M.S. degree in cyber security from Sun Yat-sen University in 2020.

His research interests include multimedia security and forensics.

Shangjun Luo received the B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018. He is currently pursuing the M.S. degree with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China.

His research interests include multimedia security and forensics.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA, in 2008 and 2010, respectively.

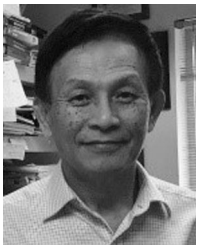
He is an Associate Professor and the Director of the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou received the Third Price of Macao Natural Science Award as a Sole Winner in 2020 and a co-recipient in 2014 and was a recipient of the Best Editor Award for his contributions to *Journal of Visual Communication and Image Representation* in 2020. He has been a leading Co-Chair of Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society since 2015. He serves as an Associate Editor for IEEE TRANSACTIONS ON NEUTRAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and four other journals. He is a Fellow of Society of Photo-Optical Instrumentation Engineers and was recognized as the "Highly Cited Researcher" in Web of Science in 2020.

Jiwu Huang (Fellow, IEEE) received the B.S. degree from Xidian University, Xi'an, China, in 1982, the M.S. degree from Tsinghua University, Beijing, China, in 1987, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 1998.

He is currently a Professor with the College of Information Engineering, Shenzhen University, Shenzhen, China. His current research interests include multimedia forensics and security.

Prof. Huang was the General Co-Chair of IEEE Workshop on Information Forensics and Security in 2013 and a TPC Co-Chair of IEEE Workshop on Information Forensics and Security in 2018. He is an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is a member of IEEE Signal Processing Society Information Forensics and Security Technical Committee.



Yun-Qing Shi (Life Fellow, IEEE) received the M.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1980, and the Ph.D. degree from the University of Pittsburgh, Pittsburgh, PA, USA, in 1987.

He joined the New Jersey Institute of Technology, Newark, NJ, USA, in 1987. He has authored/coauthored more than 400 papers, one book, five book chapters, an editor of ten books, and holds 30 U.S. patents. His research interests include data hiding, forensics, information assurance, visual signal processing, and communications.

Dr. Shi has served as an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART I: REGULAR PAPERS. He has been a member of a few IEEE technical committees since 2005. He has become a Fellow of National Academy of Inventors in the end of 2017.