



Forensic discrimination between traditional and compressive imaging by blurring kernel investigation

Ali Taimori¹ · Hadi Zayyani² · Farokh Marvasti¹

Received: 14 September 2022 / Revised: 24 April 2023 / Accepted: 8 August 2023 /
Published online: 12 October 2023
© Crown 2023

Abstract

Image forensics encompasses a set of scientific tests to investigation of a suspected event via intrinsic clues of imaging pipeline. Traditional image sensing at the Nyquist-Shannon rate as well as the new modality of compressive imaging below the rate are two main types of sensing in photography and imaging applications. Hence, for forensic investigators, it would importantly necessitate the ability to discriminate among images captured by them. However, due to the complex nonlinear nature of imaging processes, investigating imagers' traces is a difficult task. To this intent, we first systematically model the imaging pipelines as an encoder-decoder pair. For exploring distinguishable traces, we mathematically simplify and linearize the pair for compressive imaging and two main forms of traditional image sensing with or without compression. Our theoretical analyses on the approximate linear models reveal blurring kernels of different imagers have discriminability. To validate it in real-world scenarios, we considered the whole imaging process as an inverse problem and estimated the blurring kernel based on a deconvolution approach, where the discriminability is also justified by information visualization. Then, we designed a pipeline classification system, where a deep convolutional neural network is trained by the estimated blurring kernels to be able to classify the three imaging systems. Our results in compressive imaging identification show an accuracy improvement about 3.7 % in comparison to the best result among compared methods. Implementation codes are available for research and development.

Keywords Compressive sensing · Convolutional neural network · Data compression · Deconvolution · Image forensics · Imaging · Systems modeling

✉ Ali Taimori
alitaimori@yahoo.com

Hadi Zayyani
zayyani@qut.ac.ir

Farokh Marvasti
marvasti@sharif.edu

¹ Department of Electrical Engineering, Sharif University of Technology, Tehran 14588-89694, Iran

² Qom University of Technology (QUT), Qom, Iran

1 Introduction

Image forensics encompasses a set of scientific tests to investigation of a suspected event/crime via left image clues. The science enables governments, the police, and courts to discover the origin of events/crimes [18, 29, 39, 49]. To do the task, forensic experts require to know imaging pipelines which lead to the acquisition of digital images, from scene sensing to storage/transmission. This realizes by understanding of sorts of imaging and their differences. Two general types of image sensing exist: (1) traditional image acquisition, and (2) Compressive Sensing (CS)-based imaging. Conventional imagers such as surveillance cameras and those on portable devices first acquire an image at the Nyquist-Shannon rate, and then, if required for storage or transmission purposes, discard redundant data by compression like the widely used JPEG-family compression. However, the technology of Compressive Sensing (CS) which operates below the Nyquist-Shannon rate, has recently led a new generation of image acquisition modality, called compressive imaging [43]. This, as a type of computational imaging, provides advantages in comparison to conventional devices such as considerable reduction in hardware components, power consumption optimization, and fast image acquisition time. In a compressive imaging device, the sampling and compression processes are done simultaneously in one step [50]. In other words, the compression can be directly performed during sensing. Random sampling-based imagers such as single-pixel camera [16], imaging beyond the visible spectrum like hyperspectral and infrared cameras [10], dynamic [21] and intelligent CS-based sensing systems [50], radar imaging [2], wireless CS networks [56], constrained medical imaging systems such as CS-equipped magnetic resonance imaging [57] and lens-less imagers [46, 54], and scientific imaging devices [41] are paradigms of imaging based on the CS technology.

1.1 Literature review

In the conventional imaging, forensic analyses developed in the literature are categorized into two general categories, including: 1) approaches that classify the source of a questionable image by assigning the image to a device make and model [33, 55], and 2) methods that detect or localize forged regions in digital images [3, 4, 51]. In both categories, discovering intrinsic signatures left from recorded images plays a key role for forensic investigators. The forensic footprints themselves can be extracted by three general techniques [53]. We call them as: (a) feature engineering-based methods, in which forensic examiners extract handcrafted features for classification [3, 52]; (b) deep learning-based methods which try to extract forensic features directly from data in an automatic manner [4, 6]; and, (c) forensics-informed methods which represent hybridized variants of methods of (a) and (b). This paper falls into the category of forensics-informed methods, where we combine forensic clues with deep learning for incorporating the knowledge of forensic examiner in machine learning processes.

Although the above-mentioned valuable studies have been carried out in the context of traditional image forensics, a lack of theoretical and applied researches is seen for a deep understanding of parts and functions of the compressive imaging systems from the forensic perspective. This issue causes another important problem about the forensic cognition of differences between the kinds of imaging. For example, consider the case in which the task of a forensic examiner is identifying whether the source of a query image is from a traditional or a compressive imaging manner. Solving this problem necessitates extracting signatures that can discriminate well among the diverse systems.

Let us state that an image with an uncompressed file format such as Bitmap (BMP) be under an investigation. The image may come from a CS imaging or not. If the digital photo has been acquired from a conventional imager, it may have been already compressed via a source coder such as the well-known JPEG2000 (JP2) compression standard or acquired in an uncompressed file format, i.e., raw imaging. Hence, in the considered scenario, three main types of imaging systems are introduced to be identified, including compressive imaging, conventional sensing plus compression, and conventional raw imaging. It is notable that both JP2 and most of compressive sensing algorithms employ Discrete Wavelet Transform (DWT) [45, 47]. This points out the fact that the confusion between these types of imaging systems is more probable. Therefore, the main complexity is on JPEG2000 standard among other image compression techniques. In [11, 12], Chu et al. found empirical probability mass functions of wavelet coefficients for different imaging systems follow Laplace-like distributions with different parameters of location and diversity. Based on the observation, they employed a 2-step thresholding-based decision making process to separate traditional sensing from compressive one. Its first step contains a distribution detector based on maximum likelihood estimator to discriminate uncompressed images from JP2 and compressively sensed images. The second distribution-based detector gains Expectation Maximization (EM) algorithm to classify the remainder, namely traditionally sensed JP2 images from compressively sensed ones.

1.2 Our approach and contributions

Due to the complex and nonlinear nature of imaging processes, investigating imagers' traces is a difficult task. To address this issue, we first systematically model an imaging pipeline as an encoder-decoder pair to bring more insight into analytical investigations. To be able to explore distinguishable traces, we mathematically simplify and linearize the compressive imaging as well as the two main forms of traditional image sensing. Our theoretical analyses on the approximate linear models revealed imaging blurring kernels make discriminability, where we followed them for further analysis. Here, the discriminability means the degree to which an intelligent machine can separate an imaging source from another. To validate the discriminability of the proposed footprints in real-world nonlinear scenarios, we considered the whole imaging process as an inverse problem and estimated the blurring kernel numerically based on a deconvolution approach [30], where the discriminability is also justified by information visualization. Then, we schematized a deep learning based pipeline to automatically classify blurring kernels from diverse imaging sources. To this end, we trained a Convolutional Neural Network (CNN) by the estimated blurring kernels from an image set including images of the three imaging systems and evaluated the effectiveness of blurring kernel signatures. Main advantages of the proposed method over the approaches in categories of (a) and (b) are higher performance, due to incorporating the knowledge of forensic analyzer, and simpler deep-network architecture, because of learning from blurring kernel images which are very small in size and sparse in comparison to the images (See Table 1.). The novelties of this paper are:

- modeling compressive and conventional imagers and approximating them by linear models for simplifying forensic investigations,
- revealing blurring kernel clues left by different imaging systems as discriminative features and proving their discriminability both theoretically and statistically,
- proposing a deep learning framework for the application of imaging source identification based on the engineered blurring kernel features.

Table 1 Simplified encoder-decoder models of compressive and conventional imagers

Imaging system	Encoder/decoder model	Blurring kernel	Baboon's example	Application areas
Compressive sensing		$H_C \approx \hat{A}^{-1}A$		<ul style="list-style-type: none"> – Multi- or hyper-spectral devices – Infrared cameras – Biomedical equipment – Radar imaging – Scientific imagers
Raw format camera		$H_U = \mathbf{I}^{-1}\mathbf{I}$		<ul style="list-style-type: none"> – Professional photography – Consumer electronics
Compressed format		$H_J \approx \mathbf{T}^{-1}\hat{\mathbf{T}}$		<ul style="list-style-type: none"> – Consumer electronics – Surveillance cameras – Robotics – Industrial imagers

The rest of this paper is organized as follows. In Section 2, we derive three general models for the imaging systems. Section 3 proposes a CNN classifier for identifying the source of imaging based on learning from blurring kernels. Section 4 designs a series of experiments to show the effectiveness of the proposed approach in different situations. Finally, we conclude the paper in Section 5.

2 Modeling and linearization

2.1 Encoder-decoder modeling of compressive imaging

Different setups may be considered for designing application-specific compressive imagers, which share similar behaviors [16, 50]. For instance, in single-pixel camera [16], the scene image $\mathbf{X} \in \mathbb{R}^{h \times w}$ is first concentrated by a primary lens on a light modulator such as Digital Micro-mirror Device (DMD) consisting of a configurable array of mirrors, LCD, and coded modulation. The dimensions h and w represent the height and the width of the image \mathbf{X} , respectively. For each measurement cycle from a total of m measurements, a fraction of the analogy image projected onto DMD is selected by a predefined random pattern, modeled as the column vector $\mathbf{a}_i \in \mathbb{R}^n$, in which $n = h \times w$. Then, the randomly sampled image is collected on a single photo-diode sensor by another secondary lens. The sensor signal corresponding to the modulated light is converted to a single discrete measurement $y_i \in \mathbb{R}, \forall i = 1, 2, \dots, m$. The cycle is repeated m times, which result in the measurement vector $\mathbf{y} \triangleq \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{R}^m, \mathbf{A} \in \mathbb{R}^{m \times n}$ represents a wide sensing matrix arranged by $\mathbf{A} = [\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_m^T]^T$, and $\mathbf{n} \in \mathbb{R}^m$ models the additive noise of measurements. There is usually $m \ll n$. The sampling rate or compressive ratio is defined as:

$$R_s \triangleq \frac{m}{n}. \tag{1}$$

The vector \mathbf{x} represents a vectorized version of the input compressible image \mathbf{X} . The sensing matrix may be defined as $\mathbf{A} \triangleq \mathbf{M}\mathbf{F}$, in which the matrices \mathbf{F} and \mathbf{M} denote the sparsifying transform and sampler, respectively. Therefore, rewriting the measurement vector \mathbf{y} in terms of the sparse-signal vector \mathbf{s} yields:

$$\mathbf{y} = \underbrace{\mathbf{A}}_{\triangleq \mathbf{M}\mathbf{F}} \mathbf{x} + \mathbf{n} = \mathbf{M} \underbrace{\mathbf{F}\mathbf{x}}_{\triangleq \mathbf{s}} + \mathbf{n} = \mathbf{M}\mathbf{s} + \mathbf{n}. \tag{2}$$

In practice, the light modulator requires calibration one or more times during the image acquisition process. A decoder algorithm gets both the measurement vector \mathbf{y} and the sensing matrix \mathbf{A} , and then reconstructs the image scene. Single-pixel camera recovers the sparse signal $\hat{\mathbf{s}}$ using the ℓ_1 -norm convex minimization algorithm, also called Basis Pursuit (BP) in the literature [15]. This optimizer estimates $\hat{\mathbf{s}}$ in a search space as:

$$\hat{\mathbf{s}} = \min_{\mathbf{s}} \|\mathbf{s}\|_{\ell_1} \text{ s.t. } \|\mathbf{y} - \mathbf{M}\mathbf{s}\|_{\ell_2}^2 \leq \epsilon, \tag{3}$$

where $\|\mathbf{n}\|_{\ell_2}^2 = \|\mathbf{y} - \mathbf{M}\mathbf{s}\|_{\ell_2}^2$ denotes the energy of measurements noise. It is set to be less than or equal to the small value ϵ in sparse recovery algorithms. The symbol $\|\cdot\|_{\ell_p}$ represents the ℓ_p -norm. Finally, the original vectorized signal is recovered by:

$$\hat{\mathbf{x}} = \mathbf{F}^{-1}\hat{\mathbf{s}}. \tag{4}$$

Generally, the process of imaging and representation of a digital image in the CS technique can be modeled by two basic blocks, including: the encoder \mathcal{G} and the decoder \mathcal{G}^{-1} . Figure 1

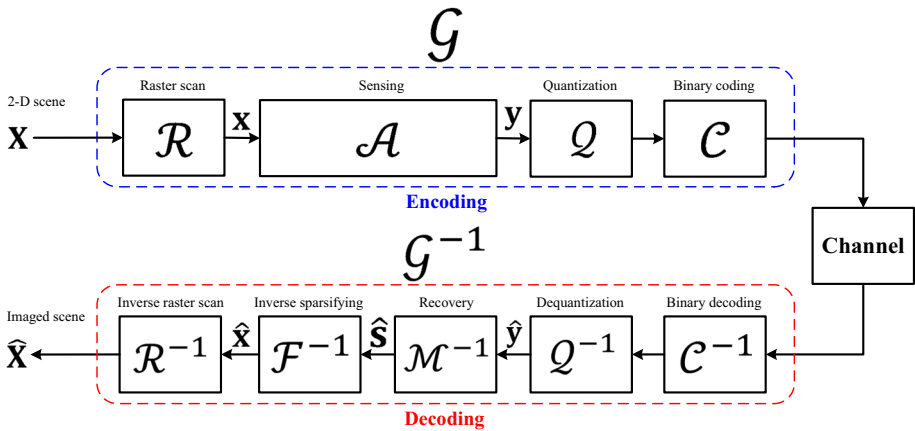


Fig. 1 Modeling a compressive imaging system by an encoder-decoder pair for imaging a 2-D scene, which includes the linear and nonlinear mathematical operators

shows the processes of encoding and decoding in a compressive imaging system. The encoding process is usually performed inside a compressive imager. The decoding process may be accomplished by using computers or servers with powerful hardware resources. The need to powerful hardware is due to the computational complexity of nonlinear decoding algorithms. Because compressive sensing performs a many-to-one transform in the encoding phase, the decoder is not exactly the inverse system of the encoder, i.e., the function \mathcal{G} is not invertible and $\mathcal{G}^{-1}\{\mathcal{G}(\mathbf{X})\} \neq \mathbf{X}$. Therefore, we define $\hat{\mathbf{X}} \triangleq \mathcal{G}^{-1}\{\mathcal{G}(\mathbf{X})\}$ and probe the information loss as intrinsic artifacts for compressive imaging forensics.

In the encoding phase shown in Fig. 1, the operators \mathcal{R} , \mathcal{A} , \mathcal{Q} , and \mathcal{C} represent the parallel raster scanner, sensing operator, quantizer, and coder, respectively. In the decoding phase, the blocks \mathcal{R}^{-1} , \mathcal{F}^{-1} , \mathcal{M}^{-1} , \mathcal{Q}^{-1} , and \mathcal{C}^{-1} denote the inverse parallel raster scanner, inverse of the sparsifying transform (\mathcal{F}), reconstructor, dequantizer, and decoder, respectively. The task of the block \mathcal{R} is to scan parallel-wise in the vertical direction of signal samples. This block converts a matrix to a vector, i.e., $\mathbf{x} = \mathcal{R}\{\mathbf{X}\}$, and its inverse system, \mathcal{R}^{-1} , is exactly available, i.e., $\hat{\mathbf{X}} = \mathcal{R}^{-1}\{\hat{\mathbf{x}}\}$. This also forensically means the scan operation is loss-less and does not has any traces. The operator \mathcal{F} maps the signal onto a transform domain in which the signal is sparse. For instance, the transformations of Fourier, discrete cosine transform, wavelet, and Haar have such a property. The inverse of these transformations are exactly at hand without information loss. The sampling operator \mathcal{A} can be modeled by the matrix \mathbf{A} . To do this, two general CS approaches exist. One is using random Gaussian matrices, where each measurement may be consisted of a weighted linear combination of all signal samples [14]. Another is random sampling techniques [8], in which each measurement consists of a combination of some random samples of the signal. The latter mechanism, also used in single-pixel camera, exhibits simpler structure and yet higher performance in comparison to random Gaussian matrices for a wide range of applications [1, 17, 50, 58]. In the block \mathcal{A} , the information loss exists due to the many-to-one conversion. This means the operator \mathcal{M}^{-1} recovers the signal by an approximate nonlinear method. The reconstruction of the signal is performed by convex optimization or iterative procedures [50], such as the algorithms of BP [15], IBA [59], IMAT [36], IMATI [58], ISTA [24], IHT [5], ADMM [7], BCS [27], and ISP [44]. The quantizer \mathcal{Q} is a lossy system and leaves quantization errors in practice. The coder \mathcal{C} is generally a loss-less operation such as Huffman or arithmetic codes and the

related inverse system is completely available. Here, we assume the communication channel or storage space is ideal, i.e., error-, noise-, and distortion-less. The recovered image from a scene in compressive imaging can be formulated as:

$$\hat{X} = \mathcal{R}^{-1} \{ \mathcal{F}^{-1} [\mathcal{M}^{-1} (\mathcal{Q}^{-1} (\mathcal{C}^{-1} (\mathcal{C} (\mathcal{Q} (\mathcal{A} (\mathcal{R} (X))))))))] \}. \tag{5}$$

The above equation describes both the graphically shown steps of scene encoding and decoding in Fig. 1.

2.2 Simplification of compressive and conventional imagers

2.2.1 Linearized model of compressive imaging

As seen in (5), the model of a compressive imaging system is complex and nonlinear. This hardens its forensic investigations. For simplicity, we concentrate our focus on the measurement block \mathcal{M} and ignore other negligible sources of errors such as the quantization noise. This consideration is due to the fact that the processes of quantization and binary coding shown in Fig. 1 may be arbitrary for a compressive imager. For instance, in single-pixel camera, the measurement vector y is directly used in the decoding phase to restore an image from its unquantized measurements. However, the measurement process is always present, and because of combining data (i.e., the dimensionality reduction), introduces more significant traces than the quantizer. As mentioned earlier, in the decoding phase, the operator \mathcal{M}^{-1} for signal reconstruction uses convex optimization methods such as the ℓ_1 -norm [15] or nonlinear iterative algorithms [5, 7, 24, 27, 36, 44, 50, 58, 59]. Applying such an operation may leave some artifacts such as blurring and high frequency oscillations. We will linearize the CS decoder to reveal artifacts as forensic footprints for compressive imaging identification purposes. The simplified model of compressive imaging is considered as:

$$\hat{X} = \mathcal{R}^{-1} \{ \mathcal{F}^{-1} [\mathcal{M}^{-1} (\mathcal{A} (\mathcal{R} (X)))] \}. \tag{6}$$

For simplicity, we rewrite (6) in the following vector form:

$$\hat{x} = \mathcal{F}^{-1} \{ \mathcal{M}^{-1} [\mathcal{A} (x)] \}. \tag{7}$$

The details of all simplified imaging models including the above compressive sensing is summarized in Table 1. The input signal x is sampled by using the measurement operator \mathcal{M} .

Example 1 (Random sampling as a case of CS) In [8], Candès et al. have used random sampling as a special case of CS for Fourier matrices. Here, for simplicity of the theoretical studies, we utilize this efficient mechanism in sampler. Consider a random sampling mask represented by the vector $\mathbf{m} = [m_1, m_2, \dots, m_7]^T = [0 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1]^T$. In this model, each row of the measurement matrix \mathbf{M} consists of only a single 1 and all of the other entries are 0, so that:

$$\mathbf{M} = \begin{pmatrix} 0 & m_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & m_7 \end{pmatrix}. \tag{8}$$

Lemma 1 (Blurring kernel for compressive imager) Consider the compressive sensing model in Table 1, where the recovered signal \hat{x}_C is determined by:

$$\hat{x}_C = \mathcal{G}_C^{-1} \mathcal{G}_C (x). \tag{9}$$

The encoder of compressive sensing, \mathcal{G}_C , is modeled as a linear under-determined system of equations. However, its decoder, \mathcal{G}_C^{-1} , is a nonlinear system, because the signal is generally recovered by nonlinear iterative algorithms. Figure 2 models the decoder. To be able to analyze the system behavior, we model the decoding function by an approximate linear system around the operating point. Appendix A provides details of the derivation of the linear model of compressive imaging decoder. In such a situation, the reconstructed signal $\hat{x}_C(n)$ can be modeled by the convolution of the signal $x(n)$ with the unknown blurring kernel $h_C(n)$, i.e., $\hat{x}_C(n) = h_C(n) * x(n)$. Based on Toeplitz and circulant matrices [23], the convolution operation can be rewritten in an algebraic vector-matrix form as:

$$\hat{x}_C \approx H_C x, \tag{10}$$

where the matrix H_C represents the blurring kernel of compressive imaging system. If the matrix $\hat{\mathbf{M}}^{-1}$ is the linearized approximate model of the system \mathcal{M}^{-1} , then:

$$H_C \approx (\hat{\mathbf{M}}\mathbf{F})^{-1} (\mathbf{M}\mathbf{F}). \tag{11}$$

By defining the matrix $\hat{\mathbf{A}} \triangleq \hat{\mathbf{M}}\mathbf{F}$, we have:

$$H_C \approx \hat{\mathbf{A}}^{-1} \mathbf{A}. \tag{12}$$

2.2.2 Linearized model of conventional imaging without compression

In cameras which deliver raw images, pixels are directly measured from CCD or CMOS sensors without any post-processing or compression on the sensed pixels. This type of imaging is usually utilized for professional digital photography. To formulate the problem in this case, we extend the simplified CS model shown in Table 1 to this type of conventional imaging. In the CS representation, the matrix \mathbf{M} may be considered as a diagonal matrix, where only a given percent of its main diagonal elements with randomly distributed locations, are 1 and in other entries are 0. The ratio of the number of main diagonal elements with the value 1 to the length of signal gives the sampling rate as $R_s = \frac{\|\text{diag}(\mathbf{M})\|_{\ell_0}}{n} \times 100$ in percent. The function $\text{diag}(\mathbf{M})$ gets diagonal entries of the matrix \mathbf{M} . By this interpretation, a conventional sensing model can be interpreted as a complete measurement of the vector containing the signal \mathbf{x} . Table 1 shows the imaging system with uncompressed file format (hereafter RAW for simplicity). In this scheme, the measurement matrix is equal to the identity matrix, i.e., $\mathcal{T}=\mathbf{I}$. In the decoding phase, just the main signal is usable.

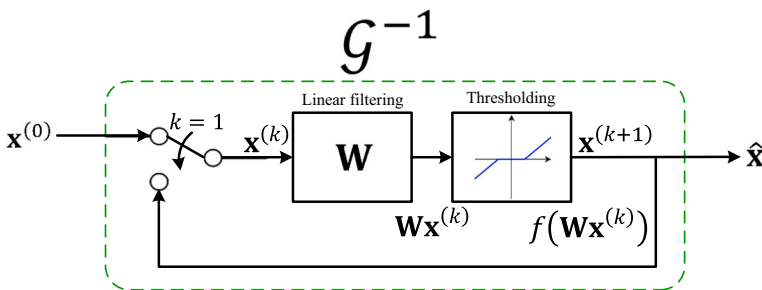


Fig. 2 Closed-loop modeling of general iterative sparse recovery approaches

Example 2 (Sampling matrix for RAW imaging) As mentioned above, the measurement matrix (8) can be rewritten in an equivalent form as:

$$\mathbf{M} = \text{diag}(\mathbf{m}). \tag{13}$$

The function $\text{diag}(\mathbf{m})$ creates the square diagonal matrix \mathbf{M} with the elements of vector \mathbf{m} on its main diagonal. In this case, the sampling rate is $R_s = 42.86\%$. However, for conventional sensing, the matrix shown in (13) converts to an identity matrix, namely $\mathbf{M} = \mathbf{I}_{7 \times 7}$ with $R_s = 100\%$.

Lemma 2 (Blurring kernel for RAW imager) Consider the traditional imaging model in Table 1 with raw image format. The recovered signal \hat{x}_U is determined by:

$$\hat{x}_U = \mathcal{G}_U^{-1} \mathcal{G}_U(x), \tag{14}$$

which the operators \mathcal{G}_U and \mathcal{G}_U^{-1} denote conventional encoder and decoder, respectively. Based on the vector-matrix representation, the raw imaging model can be described as:

$$\hat{x}_U = H_U x, \tag{15}$$

where the matrix $H_U = \mathbf{I}^{-1} \mathbf{I}$ represents the blurring kernel of the raw imaging system. This results in:

$$H_U = I, \tag{16}$$

which also means the identity matrix \mathbf{I} is the blurring kernel of raw imaging. Based on this model, the original signal is perfectly reconstructible in the decoder. Therefore, there exists $\hat{x}_U = x$.

2.2.3 Linearized model of conventional imaging with compression

Most of ordinary cameras such as those embedded in today smart cell phones, at first, grab an image and then compress its content. Similar to uncompressed raw imaging, the total number of pixels are first measured. Then, after some optional post-processing operations, the 2-D signal is coded using a built-in source coder such as JP2 in order to reduce the bit rate. In Table 1, we have planned this kind of imaging, where the total compression operation has modeled by the operator \mathcal{T} . In JP2 compression standard, the main encoding steps include the transform to wavelet domain, quantization, and binary coding, which are performed sequentially (See Fig. 3 for details.). The compression ratio in the quantizer controls the required bit rate, defined as:

$$R_c \triangleq \frac{\text{The input image size}}{\text{The output compressed size}}. \tag{17}$$

This parameter is a real number more than or equal to 1. In the decoder \mathcal{T}^{-1} , the reverse corresponding operations are applied, i.e., the binary decoding, dequantization, and inverse wavelet transform, respectively. Based on this model, the only intrinsic fingerprint of JP2 is the nonlinear quantizer system in the encoder and other operations are linear and losslessly decodable. For simplicity of analytical studies, the quantizer may be approximated by a linear model such as the linearized model of Σ - Δ modulator [35, 42] for describing noise shaping phenomenon.

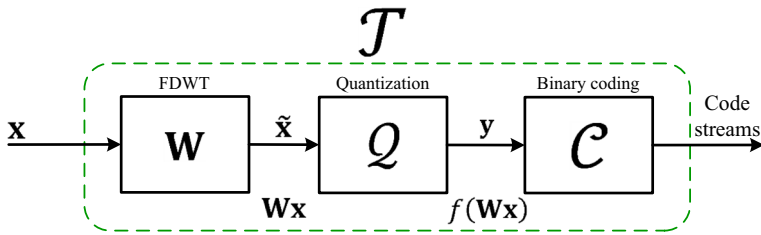


Fig. 3 Basic building blocks of JPEG2000 compressor

Lemma 3 (Blurring kernel for imaging with compression) Consider the traditional imaging with compression shown in Table 1. The recovered signal \hat{x}_J is determined by:

$$\hat{x}_J = \mathcal{G}_J^{-1} \mathcal{G}_J(x), \tag{18}$$

in which the operators \mathcal{G}_J and \mathcal{G}_J^{-1} denote the encoder and the decoder of conventional imager with JP2 compression, respectively. To find an approximate blurring kernel, we linearize the model of JP2 compressor in Fig. 3. A derivation of the linearization of JP2 quantizer is provided in Appendix B. In this case, the reconstructed signal in the form of vector-matrix representation can be formulated as:

$$\hat{x}_J \approx H_J x. \tag{19}$$

The matrix $H_J \approx (\mathbf{T}\mathbf{I})^{-1} (\hat{\mathbf{T}}\mathbf{I})$ represents the blurring kernel of conventional imaging with compression, in which the matrices $\hat{\mathbf{T}}$ and \mathbf{T}^{-1} denote the linearized approximate model of the compressor \mathcal{T} and the linear decompressor \mathcal{T}^{-1} , respectively. The final blurring matrix is simplified as:

$$H_J \approx \mathbf{T}^{-1} \hat{\mathbf{T}}. \tag{20}$$

3 Imaging identification

In Section 2, we provided a rough estimate of the blurring kernel status for different imaging systems. Due to the differences among blurring matrices, they can potentially be exploited as features for the identification of imaging history. Here, we consider the whole imaging process as an inverse problem and calculate the blurring kernel matrix more accurately from each individual input image of a provided dataset based on a blind deconvolution approach. By leveraging Lemmas 1, 2, and 3, we justify the discriminability of matrices for different imagers theoretically; and, at the same time, we verify the discriminability of the estimated kernels from real data statistically. At the end, we provide a learning and evaluation framework for the task of imaging source identification based on our proposed blurring kernel features. Figure 4 shows the architecture of our proposed imaging classification scheme, where its parts are introduced below.

3.1 Dataset preparation

Designing any classifier requires an appropriate dataset of training and test. To prepare our experimental data, we employed standard databases of raw images, which are widely used in image processing and forensics tasks [51]. We used the images set of Never-compressed Color

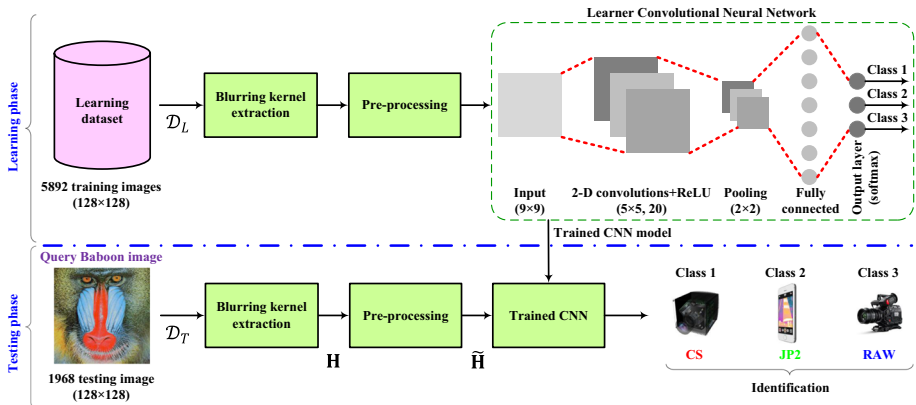


Fig. 4 A pipeline for the forensic identification of traditional and compressive imaging based on our proposed blurring kernel signatures, including the two phases of learning and testing

Image Database (NCID) [32], including 5150 raw images with the dimensions 256×256 and BMP file format. To manage hardware resources, we provided our experimental images set as follows. We first divided NCID images into $w \times h = 128 \times 128$ non-overlapping blocks. Then, we randomly selected a subset of them, including $N_r = 655$ raw image patches. This yields the initial data set $\mathcal{D}_I = \{\mathbf{x}^i\}_{i=1}^{N_r}$.

We performed experiments under different settings of imaging systems. The sampling rate in compressive sensing technology is usually set at the rates less than 50% of the signal length [12]. To this intent, we drew the experimental sampling rates from the set $\mathcal{S} = \{25\%, 40\%, 50\%, 67\%\}$ to generate different compressively sensed images based on the encoder/decoder designed in single-pixel camera [16]. This means that we used Gaussian measurements and BP algorithm for CS sampling and recovery, respectively. According to the above sampling rates, the total number of CS images is $N_r \cdot |\mathcal{S}| = 2620$, where CS dataset is defined as $\mathcal{D}_C \triangleq \{\hat{\mathbf{x}}_C^i\}_{i=1}^{N_r \cdot |\mathcal{S}|}$. The symbol $|\cdot|$ denotes the cardinality of a set.

To create JP2 dataset, we adjusted the parameter of compression ratio, R_c , in MATLAB JPEG2000 coder similar to the mechanism utilized in [12]. In this case, for a fair comparison, we set the compression ratio of JPEG2000 coder in a manner that Peak Signal to Noise Ratio (PSNR) of conventional imager with compression is equal to PSNR of the compressive imaging system. Assume the elements of the set $\{C, J, U\}$ refer the compressive imager, conventional imaging with JP2 compression, and conventional imaging system with uncompressed file format, respectively. The expectation in the regime satisfies $\mathbb{E}\{\text{PSNR}|J\} \approx \mathbb{E}\{\text{PSNR}|C\}, \forall R_s \in \mathcal{S}$. For each R_s , there exists:

$$\mathbb{E}\{\text{PSNR}|k\} \triangleq \frac{10}{N_r} \sum_{i=1}^{N_r} \log_{10}\left(\frac{M^2}{\frac{1}{n} \|\mathbf{x}^i - \hat{\mathbf{x}}_k^i\|_{\ell_2}^2}\right), \forall k \in \{C, J\}. \tag{21}$$

The parameter M is the maximum value of a pixel, which is 255 for an 8-bit unsigned integer image. In our experiments, the resultant set of compression ratio corresponding to the set \mathcal{S} is equal to $\mathcal{P} = \{107.7899, 72, 54, 33.8\}$. This process results in $N_r \cdot |\mathcal{P}| = 2620$ JP2 images of different compression ratios, where JP2 dataset is defined as $\mathcal{D}_J \triangleq \{\hat{\mathbf{x}}_J^i\}_{i=1}^{N_r \cdot |\mathcal{P}|}$. The wavelet basis employed for creating both CS and JP2 images was biorthogonal 4.4 with 4 levels of decomposition.

To prepare a balanced uncompressed images set \mathcal{D}_U , in addition to the database \mathcal{D}_I , we randomly chose other 1965 images from the 128×128 patches without replacement, placed in the set \mathcal{D}_O . Then, we collected all of them into the balanced database $\mathcal{D}_U \triangleq \{\mathbf{x}^i\}_{i=1}^{|\mathcal{D}_I|+|\mathcal{D}_O|}$, including $|\mathcal{D}_I| + |\mathcal{D}_O| = 2620$ images. In this case, the number of data for each class in the set $\{C, J, U\}$ is the same. The reason behind this is that learning from balanced classes generally leads to an appropriate performance [26].

Our final database consists of the union of all the above datasets as:

$$\mathcal{D} \triangleq \bigcup_{i \in \{C, J, U\}} \mathcal{D}_i, \quad (22)$$

in which the total number of images is $|\mathcal{D}| = 7860$. We exploited the hold-out cross validation mechanism [20]. So, the first 75% of data of each sub-class selected as the learning set and the remainder for testing, where their sizes are respectively equal to $\mathcal{D}_L = \frac{\lfloor 0.75N_i \rfloor \cdot |\mathcal{D}|}{N_i} = 5892$ and $\mathcal{D}_T = 1968$ images. The symbol $\lfloor \cdot \rfloor$ denotes the nearest integer function. Our data separation strategy guarantees to avoid any information leakage of images with the same scene/content from the learning phase to the test phase. For providing more training data, we have also used McGill CCID dataset [38]. We selected the central 128×128 region from each of the 1096 TIFF images available in it [51].

For examining further the generalization capability, in addition to the 1968 test images, we provided two different test sets including 48 images chosen from Microsoft Research Cambridge Object Recognition Image Database and well-known image processing test-set images such as Baboon, Cameraman, Lena, etc. We utilized 27 randomly selected, cropped, and resized images from Microsoft database. Also, 21 images were chosen from well-known test images of the size 128×128 [50].

3.2 Blurring kernel estimation

We model an imaging system as an inverse problem to be able to numerically estimate the blurring kernel matrix $\mathbf{H}_i \in \mathbb{R}^{a \times b}$, $\forall i \in \{C, J, U\}$ by deconvolution. We are given an image from an imaging system, the problem is to reconstruct the blurring kernel of the original scene. Deconvolution is known as a paradigm of inverse problems [40]. Various direct and indirect deconvolution methods are available in order to estimate a degradation kernel and a deblurred version of an image such as Wiener filter and Richardson-Lucy algorithm [31, 48]. One of the best algorithms is the modified maximum-a-posterior-based deconvolution algorithm proposed in [30], which has used here. By using EM optimization, this algorithm iterates alternatively between two steps: one for solving a latent sharp image, and another for finding a blurring kernel. In order to blindly deconvolve the blurring kernel, it should be taken into consideration that the kernel dimensions are generally much less than the image dimensions [3]. Also, the shape of the degradation kernel is usually considered as a square with the length a . Table 1 illustrates the estimated blurring kernels of the length $a = 9$ pertaining to Baboon image with the size 128×128 for the three different imaging systems. In this example, the simulated compressive imager is based on Gaussian measurements and BP recovery algorithm with the sampling rate $R_s \approx 34\%$. The difference among blurring kernels can be clearly seen with the naked eye.

Theorem 1 (*Blurring kernels discriminability of imagers*) *Let different imaging systems be modeled by $\widehat{x}_i \approx H_i x$, $\forall i \in \{C, J, U\}$, in which the vectors x and \widehat{x}_i , and the matrix \mathbf{H}_i denote the latent original sharp signal, the output signal of the i^{th} imaging system, and the blurring*

kernel of the i^{th} imager, respectively. The elements of the set $\{C, J, U\}$ represent compressive imager, conventional imager with JPEG2000 compression, and conventional imaging system with uncompressed file format, respectively. Then, the blurring kernels of compressive \mathbf{H}_C , conventional with uncompressed format \mathbf{H}_U , and conventional with compression \mathbf{H}_J imaging systems are discriminable.

Proof Distinct Lemmas 1, 2, and 3 demonstrate that the blurring kernels of compressive imager, conventional imager with uncompressed format, and conventional imager with compression are determined by inconsistent matrices $\mathbf{H}_C \approx \hat{\mathbf{A}}^{-1}\mathbf{A}$, $\mathbf{H}_U = \mathbf{I}$, and $\mathbf{H}_J \approx \mathbf{T}^{-1}\hat{\mathbf{T}}$, respectively, where the matrix \mathbf{I} is an identity matrix and the approximate kernels \mathbf{H}_C and \mathbf{H}_J are derived from affine models with different slope and y -intercept parameters. Accordingly, the blurring kernels of various imaging models are discriminable, i.e., $\mathbf{H}_i \neq \mathbf{H}_j, \forall i, j \in \{C, J, U\}, i \neq j$, except for the following two worst cases:

- Case I: If the compressive ratio $R_s \rightarrow 100\%$, then CS imaging tends to conventional sensing. To demonstrate this, substituting $\mathbf{M} = \mathbf{I}$ in (11) gives $\mathbf{H}_C = (\mathbf{IF})^{-1}(\mathbf{IF}) \Rightarrow \mathbf{H}_C = \mathbf{F}^{-1} \underbrace{\mathbf{I}^{-1}\mathbf{I}}_{=\mathbf{I}} \mathbf{F}$, where results in $\mathbf{H}_C = \mathbf{F}^{-1}\mathbf{F}$, and ultimately $\mathbf{H}_C = \mathbf{I}$. This situation means the lack of discriminability between CS imager and RAW imaging as the compressive ratio approaches the maximum rate 100%.
- Case II: If the compression ratio $R_c \rightarrow 1$, then $\hat{\mathbf{T}} = \mathbf{T} = \mathbf{I}$, which yields $\mathbf{H}_J = \mathbf{I}$. This means that as the compression ratio approaches the minimum rate 1, the conventional imaging with compression tends to RAW imaging. This leads to the lack of discriminability between conventional imaging systems with compression and without compression.

Also, to statistically validate the discriminability of the proposed signatures on real data as supportive evidence of our proof, we visualized the situation of estimated blurring kernels in a two-dimensional cluster representation by using the information visualization tool presented in [52]. This visualizer is a bi-level dimensionality reduction technique. At the first level, the kernel features are sparsely coded via the technology of compressive sensing and group least absolute shrinkage and selection operator algorithm. Using the unsupervised non-parametric t-SNE dimensionality reduction algorithm [34], sparse vectors are mapped onto a 2-D space with the components v_1 and v_2 at the second level of the visualizer. For visualization, we randomly selected with replacement 50% of the dataset \mathcal{D} for training and remainder for evaluation. In the visualizer, we adjusted the parameters of the number of measurements as $d = 10$ and the regularizer as $\lambda = 0.5$. Figure 5 illustrates the 2-D visualized results at two random runs. It is noticeable that due to the stochastic nature of the visualization algorithm [52], clustering results at various random runs are naturally different. The visualization results also justify the blurring kernels discriminability of imagers. Specifically, visualized information reveals that:

- point clouds pertaining to the cluster JP2 are separated well,
- special groups of CS and RAW samples have overlap, and
- distinct islands convey species information belonging to individual classes. These species may be generated by different ratios of coders and other latent variables of imaging systems (In experiments, we have analyzed the behavior of these sub-classes in Section 4.2.2.).

□

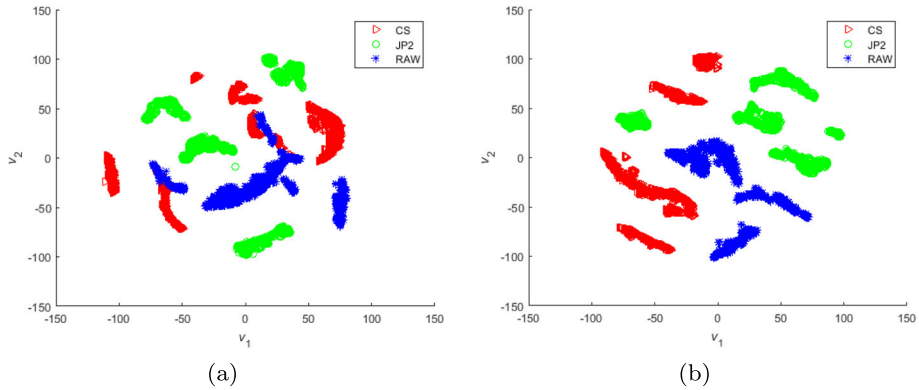


Fig. 5 2-D visualization of the discriminability of blurring kernel features for various imaging systems at two different random runs (a) and (b)

3.3 Convolutional neural network for classification of imaging systems

CNN is one of the well-known models in deep machine learning, which is originally used for detection and identification of objects in images and videos [22]. Unlike feature engineering methods such as SVM which features are manually extracted by domain-specific experts, CNN determines characteristics in an automatic manner directly from image pixels. The structure of a CNN consists of two main building blocks. The first block contains feature detection layers and the second block includes classification layers. In the feature detection block, useful features are extracted by filtering, thresholding, and nonlinear down-sampling. These operations are applied by using a successive set of processing steps on the input image pixels including convolution, rectified linear unit (ReLU), and pooling, respectively. Convolution parameters can be obtained through training. In the classification stage, a vectorized version of the features obtained from the first block are fed to a Fully Connected (FC) network such as multi layer perceptron in order to train weights and biases of the network from learning data. CNN employs softmax activation function to predict output probabilities of classes.

Different from the conventional usage of deep CNNs, here we train a network in a hybridized manner aware from the discriminability of blurring kernels. The scheme leads to a forensics-informed CNN for the application of source identification of the imaging systems. To do this, we convert the estimated 2-D blurring kernels \mathbf{H} to corresponding gray-level images $\tilde{\mathbf{H}}$ via a pre-processing step, and then feed them to a CNN as seen in Fig. 4. In the pre-processing of CNN learning/testing, we first normalize kernel values in the range $[0, 1]$. Then, the normalized values are converted into 8-bit unsigned integers in the range $[0, 255]$. We examined different topologies for learning CNN and finally selected the structure having maximum training accuracy. Based on our experiments on different architectures, a CNN network with a small number of convolutional and fully-connected layers is sufficient to learn the small, informative gray-level blurring kernel features. The employed optimal architecture consists of an $a \times b = 9 \times 9$ input blurring kernel image (Section 4.1 explains how we selected this optimal size.), a convolutional layer including 20 filters of the dimensions 5×5 , a ReLU layer, a non-overlapping max pooling layer with the pool size 2×2 , a fully-connected layer with 3 neurons, and an output layer with softmax activation function which generates the probabilities of classes C_1 to C_3 . In the 3-class classification problem, we have

defined the classes C_1 , C_2 , and C_3 as the representatives of CS, JP2, and RAW, respectively. The optimization procedure used for training was stochastic gradient descent algorithm with momentum. The settings of the maximum number of epochs and the initial learning rate equal to 30 and 0.0003, respectively.

4 Experiments

In this section, we design a series of numerical experiments to show the effectiveness of the proposed footprints as well support our theoretical findings. We have compared the proposed forensics-informed CNN approach to techniques of k -Nearest Neighbor (k -NN) [13], Multi-Layer Perceptron (MLP) neural network [25], SVM [9, 25], conventional CNN [22], and the method in [12]. We have performed two different experiments: 1) using only the BP algorithm [15], and 2) investigating the generalization capability under various unseen CS algorithms including BCS [27], IMAT [36], IMATI [58], and ISP [44]. It is important to note that the IMAT and IMATI algorithms utilize random sampling in sensing, and BCS and ISP approaches employ Gaussian measurements of all samples of signal.

4.1 Evaluation on different dimensions of blurring kernel

The dimensions of estimated blurring kernel are the parameters that may affect a restoration process. Here, the goal is to evaluate the accuracy (See (25).) of our proposed pipeline identification system vs the blurring kernel length a , ranging from 3 to 13. Among the compared approaches, the classifiers based on k -NN, MLP, and SVM are trained by our blurring kernel features for a fair comparison. Hence, we also sought their optimal sizes. Figure 6 depicts the accuracy vs blurring kernel dimensions. A growth-decline pattern is seen among methods. There is almost a congruency of the best accuracy on the size 9×9 . This size notifies that features information is enough for an appropriate classification without

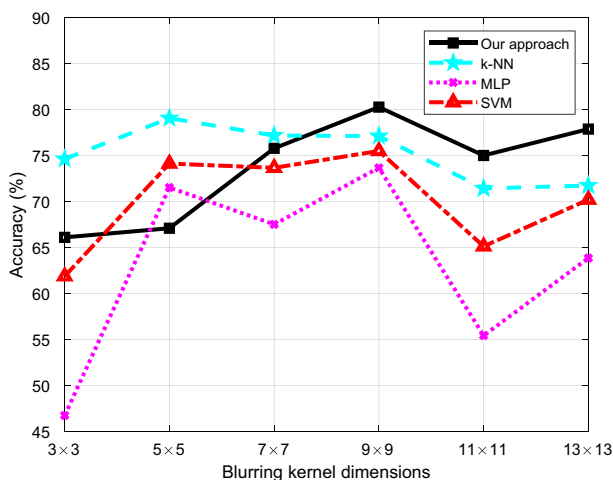


Fig. 6 The performance under variations of blurring kernel dimensions for those methods that use our proposed blurring kernel traces as forensic features

needing to redundancies due to increasing dimensions. Therefore, for next experiments of the paper, we set $a^* = 9$ for all the methods.

4.2 Results and comparison

4.2.1 Implementation information of compared approaches

Simulations have done in MATLAB and run on an Intel Core i7 2.2GHz machine. In the k -NN classifier, we used the number of nearest neighbors $k = 3$. The architecture of MLP includes: 81 input features, 2 hidden layers with 70 and 3 neurons respectively at the first and second layers, 3 neurons in the output layer with softmax activation function. The maximum number of epochs and the initial learning rate are the same with our method.

For the SVM, we utilized an one vs one division strategy with linear kernel. Its pre-processing step includes normalization of features in the range $[0, 1]$. To do this, we first considered an individual Gaussian distribution for each feature, in such a way that $\mathcal{N}(\mu_i, \sigma_i), \forall i \in [1, a \times b]$. Then, we obtained the sample mean and standard deviation parameters for each feature from the training data. At the end, the normalization for a training/testing sample is performed as:

$$\tilde{\mathbf{h}} = (\mathbf{h} - \boldsymbol{\mu}) \boldsymbol{\sigma}, \quad (23)$$

where the vectors \mathbf{h} and $\tilde{\mathbf{h}}$ represent vectorized versions of matrices \mathbf{H} and $\tilde{\mathbf{H}}$, respectively. So, we have $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_{a \times b}]^T$ and $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_{a \times b}]^T$. The symbol $\boldsymbol{\sigma}$ denotes the entry-by-entry division. For implementing the SVM classifier, we used LIBSVM toolbox [9].

For the conventional CNN, the images of learning dataset are directly utilized to automatically extract features. In its training phase, we used an optimal structure including an input gray-level image set of the size $w \times h = 128 \times 128$, 30 filters of the dimensions 3×3 , the pool size 2×2 without overlapping, and a fully-connected network with 3 outputs. The maximum number of epochs and the initial learning rate were 30 and 0.0001, respectively.

We reproduced the method [12] and approximated the thresholding parameters τ_1 and τ_2 of the first and second detectors from the learning set \mathcal{D}_L . In a grid search, we varied the threshold τ_1 from 0.001 to 0.002 and the threshold τ_2 from 0.002 to 0.003, and found the optimal $\tau_1^* = 0.0015$ and $\tau_2^* = 0.002$ values, so that the accuracy constraint on the learning data is maximized. In this method, we also utilized 5 levels of wavelet decomposition which lead to the best performance and set the number of iterations in EM algorithm equal to 100.

4.2.2 Performance under various compressive and compression ratios

It is important to obtain the performance of classification for sub-classes of each sensing system. The compressive ratio of compressive imaging in (1) and the compression ratio of JPEG2000 encoder in (17) are main parameters for controlling the required bandwidth. Settings of these ratios directly affect the quality of images in their decoding phase. Hence, the estimated kernels vary based on the amount of compressive/compression ratios. Figure 7(a) and (b) compare the accuracy of different methods for detecting CS and JP2 classes under various compressive and compression ratios, respectively.

As seen from Fig. 7(a), by increasing the compressive ratio, the accuracy for detecting the class CS decreases. It is due to the fact that the higher compressive ratio results in the lower compressive sensing artifacts. This increases the probability of error for misclassification of

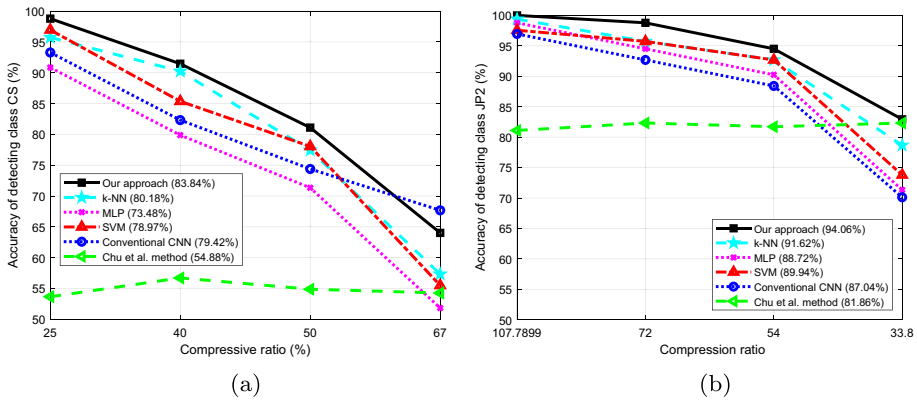


Fig. 7 (a) The performance comparison of identifying compressive sensing under various compressive ratios, and (b) the performance comparison of detecting JPEG2000 under various compression ratios. The values in parentheses in the legend of figures show averaging on their own x-axis values

the class CS as the class RAW and vice versa, that justifies Case I from Theorem 1 (See also Table 2 for more details.). Inversely, the performance of detecting the class JP2 decreases by reducing the compression ratio, as seen in Fig. 7(b). The lower compression ratio leads to the lower compression artifacts, which increases the error probability of misclassification between the JP2 and RAW classes. This also justifies Case II from Theorem 1. The values in parentheses in the legend of Fig. 7(a) and (b) show the average accuracy on all corresponding ratios. The proposed approach has the first accuracy rank among different competing methods. The method in [12] exhibits almost the same accuracy for various ratios. However, its performance is unbalanced among the three classes and lower than other methods, especially for detecting the class CS depicted in Fig. 7(a). The overall performance for detecting the class JP2 is better than the class CS, as justified by the discriminability visualization in Section 3.2.

4.2.3 Confusion matrix of imaging identification

Confusion matrix is a standard means of quantifying performance of a classifier in detail [19]. In this paper, the confusion matrix C for an imaging classifier is defined as:

$$C = \begin{matrix} & \begin{matrix} CS & JP2 & RAW \end{matrix} \\ \begin{matrix} CS \\ JP2 \\ RAW \end{matrix} & \begin{bmatrix} P(CS|CS) & P(CS|JP2) & P(CS|RAW) \\ P(JP2|CS) & P(JP2|JP2) & P(JP2|RAW) \\ P(RAW|CS) & P(RAW|JP2) & P(RAW|RAW) \end{bmatrix} \end{matrix} \quad (24)$$

The conditional probabilities in the main diagonal elements of C represent the correct classification and the remainder off-diagonal entries show the misclassification probabilities. For simplicity of comparison, we have collected confusion matrices of the compared approaches in Table 2. Overall accuracy in terms of percent is calculated by:

$$Accuracy = \frac{100}{3} [P(CS|CS) + P(JP2|JP2) + P(RAW|RAW)]. \quad (25)$$

From Table 2, the accuracy is 80.28, 77.13, 73.68, 75.51, 82.01, and 55.64 for our purposed, k -NN, MLP, SVM, conventional deep CNN, and Chu et al. [12] approaches,

Table 2 Confusion table for imaging identification of different approaches in %

Confusion table	Ground truth												
	CS	JP2		Our		[12]		Our		RAW			
	Our	k-NN	MLP	SVM	CNN	CNN	[12]	Our	k-NN	MLP	SVM	CNN	[12]
Predicted													
CS	83.84	80.18	73.48	78.97	79.42	54.88	1.07	2.28	3.96	1.68	1.68	2.13	32.01
JP2	0.24	0	0.61	0.61	3.05	19.66	94.06	91.62	88.72	89.94	87.04	81.86	5.03
RAW	16	19.82	25.91	20.43	17.53	25.46	4.88	6.1	7.32	8.38	11.28	16.01	62.96
													59.6
													58.84
													57.62
													79.57
													30.18

respectively. This means that the methods of conventional CNN and our approach achieve the first and second ranks in terms of the overall accuracy, respectively. The first position of the conventional CNN is because of the performance difference in detecting the class C_3 . However, the bold values in Table 2 reveal merits of our approach. Except for the class RAW, our method is ranked number one for detecting the classes CS and JP2 among competing approaches. The confusion matrices also show, for all compared methods, the misclassification errors between the classes CS and RAW are more than the other errors. The correct classification of the class JP2 is also more than the other imaging systems, because the blurring kernels of JPEG2000 images are more discriminative. Therefore, the best accuracy is obtained for predicting the class JP2 and the worst one for the class RAW. Specifically for the class RAW, comparing our detection result to that of conventional CNN shows that the automatically-extracted features by means of the convolution plus pooling represent better discriminability than the blurring kernels estimated by the deconvolver. However, in terms of the input image size, the speed of our scheme is more than the competing CNN with less memory requirement.

4.3 Generalization under unseen CS methods and datasets

This experiment examines the generalization capability of the pipeline identification system under different unseen compressive sensing approaches as well as unseen data. To this intent, we first trained our identification method on all 7860 data extracted from NCID database. It is noticeable that the CS recovery algorithm used for generating images of the compressive imaging class is BP algorithm. Then, for generalization test, we applied unseen CS algorithms of BCS, IMAT, IMATI, ISP, and BP (as a baseline) on unseen databases from the Microsoft and well-known test-set images. For a fair comparison, the sampling rate of all recovery algorithms set about 32% for Microsoft Object Recognition Database and 34% for well-known images test-set.

Figure 8 depicts the results of correct identification of the class CS, i.e., the probability $P(C_1|C_1)$ in %, for different CS algorithms on the databases of Microsoft and well-known test-set images. The reported performance metrics are average values for 10 training-testing runs. As shown in the chart, the identification system has promisingly predicted blurring kernels labels for both the unseen CS approaches and the unseen data. The best performance is belonging to the baseline BP, due to training on NCID kernels of the same recovery algorithm. The worst identification result is for IMATI method, where its interpolation mechanism during reconstructing iterations causes imaging artifacts to alleviate somewhat. Except for the method BP, the performance on Microsoft dataset is more than well-known images test-set in other sparse recovery algorithms.

However, it seems that learning blurring kernels of various recovery algorithms in the training phase can potentially improve the performance of compressive imaging identification. For instance, in the worst performance for IMATI, we curiously trained our approach with blurring kernels containing this type of CS to monitor how our model acts. For this purpose, at first, our network was trained with the datasets of NCID and CCID. Then, we evaluated its performance on the datasets of Microsoft and well-known images. Detailed confusion matrix in % is presented as:

$$\mathbf{C} = \begin{matrix} & \begin{matrix} \text{CS} & \text{JP2} & \text{RAW} \end{matrix} \\ \begin{matrix} \text{CS} \\ \text{JP2} \\ \text{RAW} \end{matrix} & \begin{bmatrix} 86.04 & 0 & 1.67 \\ 0 & 93.13 & 25 \\ 13.96 & 6.87 & 73.33 \end{bmatrix} \end{matrix} \tag{26}$$

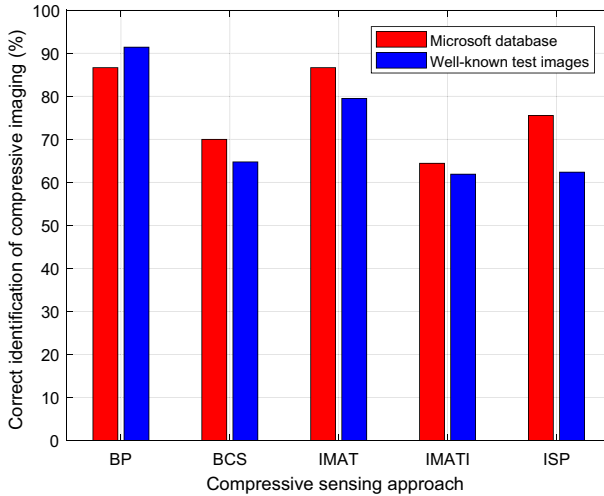


Fig. 8 Correct identification of compressive imaging in % for different sparse recovery algorithms and two test-set databases

The result $P(C_1|C_1) = 86.04\%$ shows a considerable 22.86% improvement in IMATI CS identification than that one obtained in Fig. 8 with 63.18%. Also, for the best result on the baseline BP, the confusion matrix would be:

$$C = \begin{matrix} & \begin{matrix} CS & JP2 & RAW \end{matrix} \\ \begin{matrix} CS \\ JP2 \\ RAW \end{matrix} & \begin{bmatrix} 91.46 & 0 & 3.33 \\ 0 & 94.38 & 26.67 \\ 8.54 & 5.62 & 70 \end{bmatrix} \end{matrix}, \tag{27}$$

which shows 2.41% improvement in BP CS identification.

4.4 Effect of quantization noise on CS identification

In compressive imaging, CS measurements may be quantized by means of a quantizer. Such quantized measurements introduce quantization noise in CS encoding phase. So, the aim of this experiment is to investigate the effect of the quantization noise on the performance of compressive imaging identification under various compressive ratios and compare the obtained results with those of unquantized measurements. In case of quantized measurements, the entries of the measurement vector y are rounded to the nearest integer values. For evaluation, we employed a set of unseen data chosen from the well-known test-set images. Figure 9 reports the performance of the pipeline compressive imaging identification system vs compressive ratio for both scenarios of unquantized and quantized CS measurements. The performance metrics are average values for 10 training-testing runs. The results demonstrate a perfect CS identification at the compressive rate 20% for both unquantized and quantized measured data. This also means that for low compressive ratios, the quantization noise has no any effect on the performance. Although the identification performance in case of noisy quantized measurements is less than unquantized ones for the mid and high compressive rates, the results of unquantized measurements reveal that the dominant fall in performance is mainly related to the measurement process in CS encoding but not quantization.

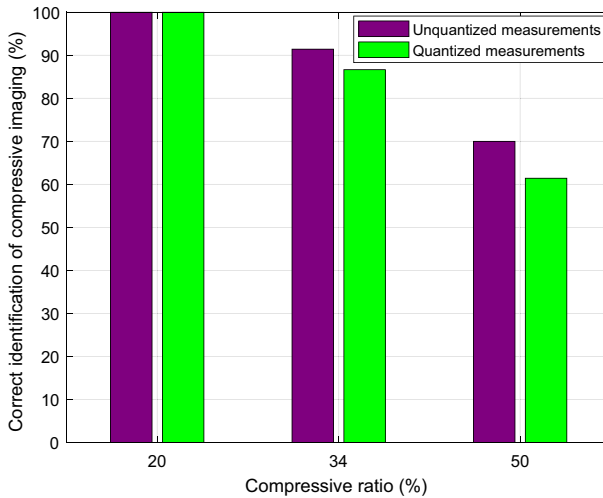


Fig. 9 The performance of compressive imaging identification vs compressive ratio for both unquantized and quantized CS measurements

5 Conclusion

This paper modeled both compressive and traditional image sensing systems as an encoder-decoder pair. To analyze behavior of the complex imaging systems for forensic investigation purposes, simplification and linearization of the imagers were provided. Thanks to insights into the modeling of different imagers, we found blurring kernels of imaging systems make a discrimination capability for identifying the source imaging device that has been captured a suspect image. Therefore, we considered the process of image acquisition as a blind deconvolution problem and estimated the blurring kernel numerically. Theoretical and statistical tests were performed for analyzing the discriminability of blurring kernels of different imaging systems. We also fed our revealed traces to a pipeline deep learning-based system for the application of imaging source identification. We reached a promising correct classification of compressive imaging about 84% on 1968 test samples. To manage hardware resources in simulations, we trained the network with a relatively small-scale training set including 5892 images for showing the concept of our method. As shown in experiments, the performance of our forensics-informed deep network can improve by increasing training data.

As future studies, being aware of noise characteristics or sharpness properties of edges can provide complementary footprints for imaging source identification purposes. Our modeling also facilitates developmental approaches for forensic investigations in connection with the next-generation, state-of-the-art compressive imaging systems. For instance, the forensic identification among a set of compressive sensing-based imaging systems can reveal the history of image acquisition such as make and model of the source compressive imaging device that has been generated a questionable image as well as settings under which the image has been captured.

Acknowledgements This research was jointly sponsored by Iran National Science Foundation and ACRI of Sharif University of Technology under agreement numbers 95/SAD/47585 and 7000/6642, respectively. The authors also thank Prof A. Amini and other researchers in Multimedia and Signal Processing Lab for their valuable comments.

Data Availability All data generated or analyzed during this study, simulation codes, and their supplementary information files are included in the published article as supplementary materials.

Declaration

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Derivation of the compressive decoder linear model

Consider the decoder of a compressive imaging system as:

$$\hat{\mathbf{x}} = \mathcal{G}^{-1}(\mathbf{y}) = \mathbf{F}^{-1}\mathcal{M}^{-1}(\mathbf{y}), \tag{28}$$

where the vector $\mathbf{y} \triangleq \mathbf{A}\mathbf{x}$ contains measurements. Iterative algorithms are known as a general class of approximate sparse decoding approaches to solve (28). They can be modeled by successive linear filtering and point-wise thresholding [28], such as the algorithms of ISTA, IHT, IMAT, and ADMM. In Fig. 2, we have modeled an iterative compressive imaging decoder as a closed-loop system, in which the recovered signal \mathbf{x} at the $(k + 1)$ th iteration with the initial guess $\mathbf{x}^{(0)}$ is formulated as:

$$\mathbf{x}^{(k+1)} = f(\mathbf{W}\mathbf{x}^{(k)}), \tag{29}$$

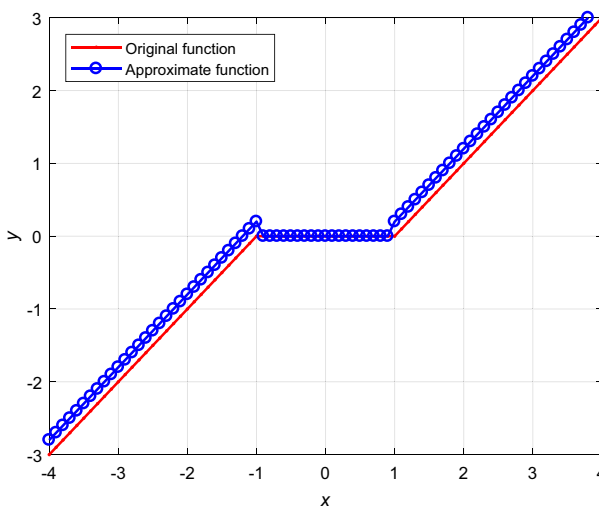


Fig. 10 The function of soft-thresholding and its approximation

which the matrix \mathbf{W} represents an algorithm-dependent linear filter, the nonlinear operator $f(\cdot)$ is a thresholding function, and $\widehat{\mathbf{x}} \triangleq \mathbf{x}^{(N)}$ after the last iteration N . One of the popular thresholding functions utilized in sparse recovery is soft-thresholding as:

$$y = f(x) = \begin{cases} \text{sgn}(x)(|x| - \Delta), & |x| \geq \Delta \\ 0, & |x| < \Delta \end{cases}, \tag{30}$$

where the parameter Δ is a threshold value. Based on Taylor series expansion [37], a linear approximation of soft-thresholding function around the known point x_0 can be represented by:

$$y = f(x) \approx f(x_0) + f'(x_0)(x - x_0), \tag{31}$$

for which the derivative of the function $f(x)$ can be rewritten by the unit step function $u(x)$ terms as:

$$f'(x) = u(-x - \Delta) + u(x - \Delta). \tag{32}$$

This function is one for $|x| \geq \Delta$ and zero for $|x| < \Delta$. Therefore, there exist $y \approx x - \Delta$ and $y \approx x + \Delta$ for $x \geq \Delta$ and $x \leq -\Delta$, respectively. Figure 10 simulates both soft-thresholding function and its linear approximation for $\Delta = 1$ and $\Delta x \triangleq x - x_0 = 0.2$, which shows small errors. Without loss of generality for the linear approximation of an arbitrary thresholding function, (28) can be rewritten as $\widehat{\mathbf{x}} \approx \mathbf{F}^{-1}\widehat{\mathbf{M}}^{-1}\mathbf{y}$, in which the matrix $\widehat{\mathbf{M}}^{-1}$ represents the linearized approximate model of the system \mathcal{M}^{-1} . By defining the matrix $\widehat{\mathbf{A}} \triangleq \widehat{\mathbf{M}}\mathbf{F}$, we have $\widehat{\mathbf{x}} \approx \widehat{\mathbf{A}}^{-1}\mathbf{Ax}$. □

Appendix B: Derivation of the linearization of JPEG2000 quantizer

As shown in Fig. 3, the basic operations of JPEG2000 compressor are the forward DWT, quantization, and arithmetic coding. Since the binary decoder \mathcal{C}^{-1} is exactly the inverse system of \mathcal{C} , the forensic model of JPEG2000 compressor can be formulated by:

$$\mathbf{y} = \mathcal{T}(\mathbf{x}) = f(\mathbf{W}\mathbf{x}). \tag{33}$$

Here, the vectors \mathbf{x} and \mathbf{y} , and the matrix \mathbf{W} denote the input signal, quantized coefficients, and the wavelet transform, respectively. The nonlinear quantization function $f(\cdot)$ for a given wavelet coefficient in JPEG2000 standard is formulated as:

$$y = f(x) = \text{sgn}(x) \lfloor \frac{|x|}{\Delta_b} \rfloor, \tag{34}$$

where the function Δ_b determines the quantization step size. For small values of Δ_b , the function in (34) approaches the line $y \approx \frac{1}{\Delta_b}x$. Hence, the quantizer can be linearized by the first-order approximation $y = f(x) \approx f(x_0) + f'(x_0)(x - x_0)$. The derivative of the function $f(x)$ is as:

$$f'(x) = 2\delta(x) \lfloor \frac{|x|}{\Delta_b} \rfloor + \text{sgn}(x) \frac{d}{dx} \lfloor \frac{|x|}{\Delta_b} \rfloor, \tag{35}$$

which the function $\delta(x)$ denotes the unit impulse function. Using the change of variable $\lambda \triangleq \frac{|x|}{\Delta_b}$, the derivative $\frac{d}{dx} \lfloor \lambda \rfloor$ is equal to:

$$\frac{d}{d\lambda} \frac{d\lambda}{dx} \lfloor \lambda \rfloor = \frac{x}{\Delta_b|x|} \{ \quad + \delta(\lambda + 1) + \delta(\lambda) + \delta(\lambda - 1) + \quad \}, \tag{36}$$

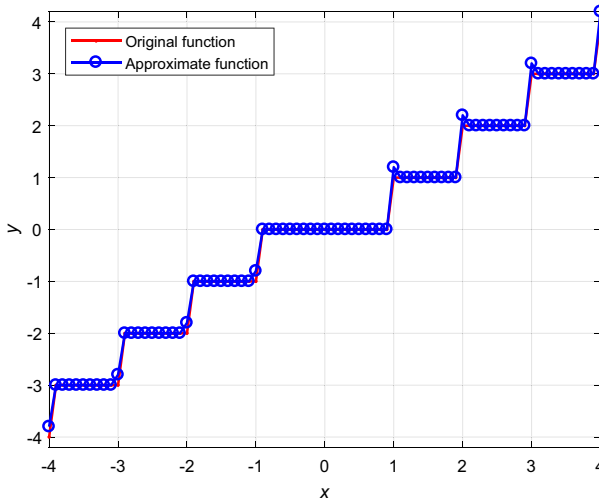


Fig. 11 The function of JPEG2000 quantizer and its approximation

where for $\lambda \in \mathbb{R}^+$ yields:

$$\frac{d}{dx} \lfloor \frac{|x|}{\Delta_b} \rfloor = \frac{1}{\Delta_b} \{ -\delta(x + 2\Delta_b) - \delta(x + \Delta_b) + \delta(x - \Delta_b) + \delta(x - 2\Delta_b) + \dots \}. \tag{37}$$

By inserting (37) into (35) and utilizing the sampling property of the impulse function as $g(x)\delta(x - x_0) = g(x_0)\delta(x - x_0)$, we have ultimately:

$$f'(x) = \frac{1}{\Delta_b} \left\{ \sum_{n=-\infty}^{\infty} [\delta(x - n\Delta_b)] - \delta(x) \right\}. \tag{38}$$

The function $f'(x)$ is only non-zero for $x = n\Delta_b, \forall n = \pm 1, \pm 2, \dots$. Therefore, for both $n < 0$ and $n > 0$, we have $y \approx \frac{1}{\Delta_b}x$, which make sense. Figure 11 depicts JPEG2000 quantizer and its linear approximation for $\Delta_b = 1$ and $\Delta x \triangleq x - x_0 = 0.2$. The differences are only seen for the discontinues points. For the above linearized model, (33) is approximated by $\mathbf{y} \approx \hat{\mathbf{T}}\mathbf{x}$, where the matrix $\hat{\mathbf{T}}$ corresponds to the linearized model of the compressor \mathcal{T} . JPEG2000 decompressor \mathcal{T}^{-1} , as the inverse of the system presented in Fig. 3, can be described by $\hat{\mathbf{x}} \approx \mathbf{W}^{-1}\delta \odot \mathbf{y} \triangleq \mathbf{T}^{-1}\hat{\mathbf{T}}\mathbf{x}$, in which the matrices \mathbf{W}^{-1} and \mathbf{T}^{-1} , the vector δ , and the symbol \odot represent the inverse DWT, decompression matrix, quantization step entries, and point-wise multiplication, respectively. \square

References

1. Abtahi A, Azghani M, Marvasti F (2019) An adaptive iterative thresholding algorithm for distributed MIMO radars. *IEEE Trans Aerosp Electron Syst* 55(2):523–533
2. Alonso MT, López-Dekker P, Mallorquí JJ (2010) A novel strategy for radar imaging based on compressive sensing. *IEEE Trans Geosci Remote Sens* 48(12):4285–4295
3. Bahrami K, Kot AC, Li L, Li H (2015) Blurred image splicing localization by exposing blur type inconsistency. *IEEE Trans Inf Forensics Security* 10(5):999–1009

4. Bayar B, Stamm MC (2018) Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection. *IEEE Trans Inf Forensics Security* 13(11):2691–2706
5. Blumensath T, Davies ME (2010) Normalized iterative hard thresholding: guaranteed stability and performance. *IEEE J Sel Topics Signal Process* 4(2):298–309
6. Boroumand M, Chen M, Fridrich J (2019) Deep residual network for steganalysis of digital images. *IEEE Trans Inf Forensics Security* 14(5):1181–1193
7. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J et al (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Learn* 3(1):1–122
8. Candès EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory* 52(2):489–509
9. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2(3), article 27
10. Chen H, Xi N, Song B, Chen L, Zhao J, Lai KWC, Yang R (2013) Infrared camera using a single nano-photodetector. *IEEE Sensors J* 13(3):949–958
11. Chu X, Stamm MC, Lin WS, Liu KR (2012) Forensic identification of compressively sensed images. In: *IEEE ICASSP*, pp 1837–1840
12. Chu X, Stamm MC, Liu KR (2015) Compressive sensing forensics. *IEEE Trans Inf Forensics Security* 10(7):1416–1431
13. Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
14. Crespo Marques E, Maciel N, Naviner L, Cai H, Yang J (2019) A review of sparse recovery algorithms. *IEEE Access* 7:1300–1322
15. Donoho DL, Elad M, Temlyakov VN (2006) Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans Inf Theory* 52(1):6–18
16. Duarte MF, Davenport MA, Takbar D, Laska JN, Sun T, Kelly KF, Baraniuk RG (2008) Single-pixel imaging via compressive sampling. *IEEE Signal Process Mag* 25(2):83–91
17. Esmaili A, Kangarshahi EA, Marvasti F (2018) Iterative null space projection method with adaptive thresholding in sparse signal recovery. *IET Signal Process* 12(5):605–612
18. Farid H (2016) *Photo forensics*. MIT Press
19. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27(8):861–874
20. Fukunaga K (1990) *Introduction to statistical pattern recognition*, 2nd edn. Morgan Kaufmann
21. Godaliyadda GDP, Ye DH, Uchic MD, Groeber MA, Buzzard GT, Bouman CA (2018) A framework for dynamic image sampling based on supervised learning. *IEEE Trans Comput Imag* 4(1):1–16
22. Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) *Deep learning*, vol 1. MIT Press Cambridge
23. Gray R (2006) Toeplitz and circulant matrices: a review. *FnT Commun Inf Theory* 2(3):155–239
24. Gregor K, LeCun Y (2010) Learning fast approximations of sparse coding. In: *ICML*, pp 399–406
25. Haykin S (2009) *Neural networks and learning machines*, 3rd edn. Pearson Education India
26. He H, Garcia EA (2009) Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 21(9):1263–1284
27. Ji S, Xue Y, Carin L (2008) Bayesian compressive sensing. *IEEE Trans Signal Process* 56(6):2346–2356
28. Jin KH, McCann MT, Froustey E, Unser M (2017) Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process* 26(9):4509–4522
29. Kåvrestad J (2017) *Guide to digital forensics: a concise and practical introduction*
30. Levin A, Weiss Y, Durand F, Freeman WT (2011) Efficient marginal likelihood optimization in blind deconvolution. In: *IEEE CVPR*, pp 2657–2664
31. Levin A, Weiss Y, Durand F, Freeman WT (2011) Understanding blind deconvolution algorithms. *IEEE Trans Pattern Anal Mach Intell* 33(12):2354–2367
32. Liu Q, Sung AH, Qiao M (2011) Neighboring joint density-based JPEG steganalysis. *ACM Trans Intell Syst Technol* 2(2), article 16
33. Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. *IEEE Trans Inf Forensics Security* 1(2):205–214
34. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605
35. Marvasti F (2012) *Nonuniform sampling: theory and practice*. Springer Science & Business Media
36. Marvasti F, Azghani M, Imani P, Pakrouh P, Heydari SJ, Golmohammadi A, Kazerouni A, Khalili M (2012) Sparse signal processing using iterative method with adaptive thresholding (IMAT). In: *IEEE ICT*, pp 1–6
37. Ogata K (2004) *System dynamics*, 4th edn. Pearson Prentice Hall
38. Olmos A, Kingdom FAA (2004) A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33(12):1463–1473
39. Piva A (2013) An overview on image forensics. *ISRN Signal Process Article ID* 496701
40. Proakis JG, Manolakis DG (2006) *Digital signal processing: Principles, algorithms, and applications*, 4th edn. Pearson Prentice Hall

41. Radwell N, Mitchell KJ, Gibson GM, Edgar MP, Bowman R, Padgett MJ (2014) Single-pixel infrared and visible microscope. *Optica* 1(5):285–289
42. Razavi B (2016) The Delta-Sigma Modulator [A Circuit for All Seasons]. *IEEE Solid-State Circuits Mag* 8(2):10–15
43. Romberg J (2008) Imaging via compressive sampling. *IEEE Signal Process Mag* 25(2):14–20
44. Sadeghi M, Babaie-Zadeh M (2016) Iterative sparsification-projection: Fast and robust sparse signal approximation. *IEEE Trans Signal Process* 64(21):5536–5548
45. Sadeghigol Z, Kahaei MH, Haddadi F (2016) Model based variational Bayesian compressive sensing using heavy tailed sparse prior. *Signal Process: Image* 41:158–167
46. Satat G, Tancik M, Raskar R (2017) Lensless imaging with compressive ultrafast sensing. *IEEE Trans Comput Imag* 3(3):398–407
47. Sayood K (2012) Introduction to data compression, 4th edn. Morgan Kaufmann
48. Schuler C, Hirsch M, Harmeling S, Scholkopf B (2016) Learning to deblur. *IEEE Trans Pattern Anal Mach Intell* 38(7):1439–1451
49. Stamm MC, Wu M, Liu KR (2013) Information forensics: An overview of the first decade. *IEEE Access* 1:167–200
50. Taimori A, Marvasti F (2018) Adaptive sparse image sampling and recovery. *IEEE Trans Comput Imag* 4(3):311–325
51. Taimori A, Razzazi F, Behrad A, Ahmadi A, Babaie-Zadeh M (2016) Quantization-unaware double JPEG compression detection. *J Math Imaging Vis* 54(3):269–286
52. Taimori A, Razzazi F, Behrad A, Ahmadi A, Babaie-Zadeh M (2017) A novel forensic image analysis tool for discovering double JPEG compression clues. *Multimed Tools Appl* 76(6):7749–7783
53. Taimori A, Razzazi F, Behrad A, Ahmadi A, Babaie-Zadeh M (2021) A part-level learning strategy for JPEG image recompression detection. *Multimed Tools Appl* 80(8):12235–12247
54. Tan J, Niu L, Adams JK, Boominathan V, Robinson JT, Baraniuk RG, Veeraraghavan A (2018) Face detection and verification using lensless cameras. *IEEE Trans Comput Imag* 5(2):180–194
55. Tuama A, Comby F, Chaumont M (2016) Camera model identification with the use of deep convolutional neural networks. In: *IEEE WIFS*, pp 1–6
56. Wang Y, Wang D, Zhang X, Chen J, Li Y (2016) Energy-efficient image compressive transmission for wireless camera networks. *IEEE Sensors J* 16(10):3875–3886
57. Ye, JC (2019) Compressed sensing MRI: A review from signal processing perspective. *BMC Biomed Eng* 1–17
58. Zayed AI, Schmeisser G (2014) *New perspectives on approximation and sampling theory*. Springer
59. Zayyani H, Babaie-Zadeh M, Jutten C (2009) An iterative Bayesian algorithm for sparse component analysis in presence of noise. *IEEE Trans Signal Process* 57(11):4378–4390

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.