A Survey of
Deep Learning-Based Source Image Forensics

Introduction
Source Camera Identification

CONTENTS

01 Introduction

02 Source Camera Identification

2.1 Traditional Convolutional Neural Networks(T.CNN)

2.2 Data Enhancement (D.E.)

2.3 Fusion and Ensemble (F./E.)

2.4 Patch Selection (P.S.)

2.5 Classifier (C.)

01

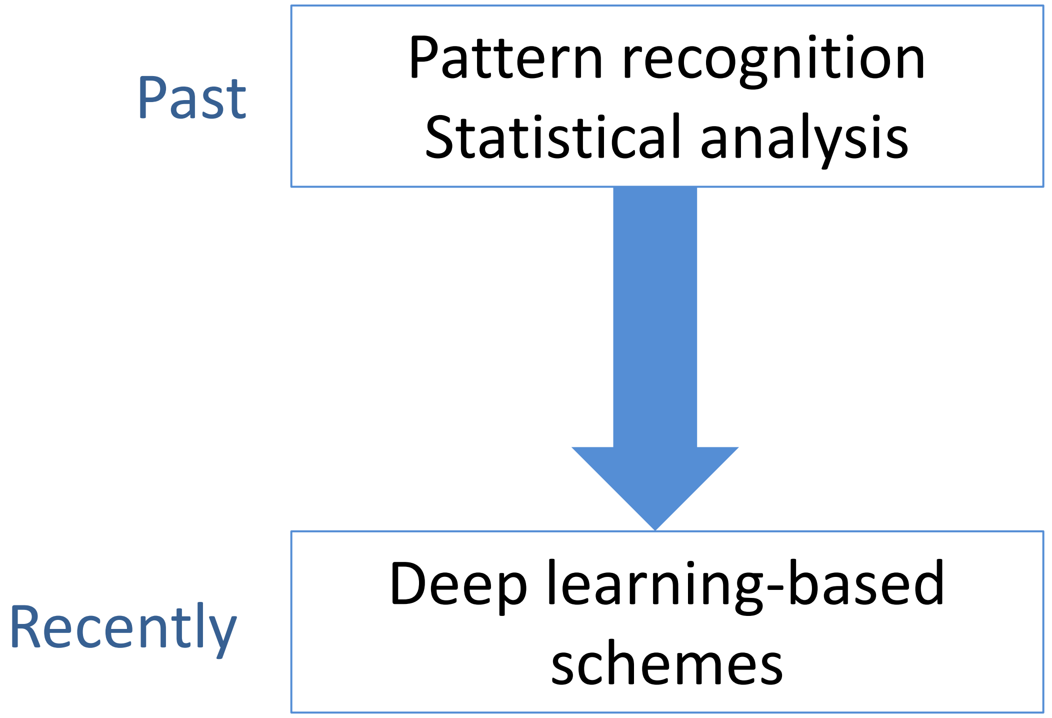
Introduction

01 Introduction

Purpose

With the advent of techniques based on artificial intelligence(AI) that can be exploited by malicious actors to spread “fake news”

In order to verify the authenticity and integrity of a digital image, a number of techniques, known collectively as “digital image forensics”



01 Introduction

Preprocessing by a spatial filter to improve their signal-to-noise(SNR)

Classifier result for the original image can be obtained by voting among the pixel patches

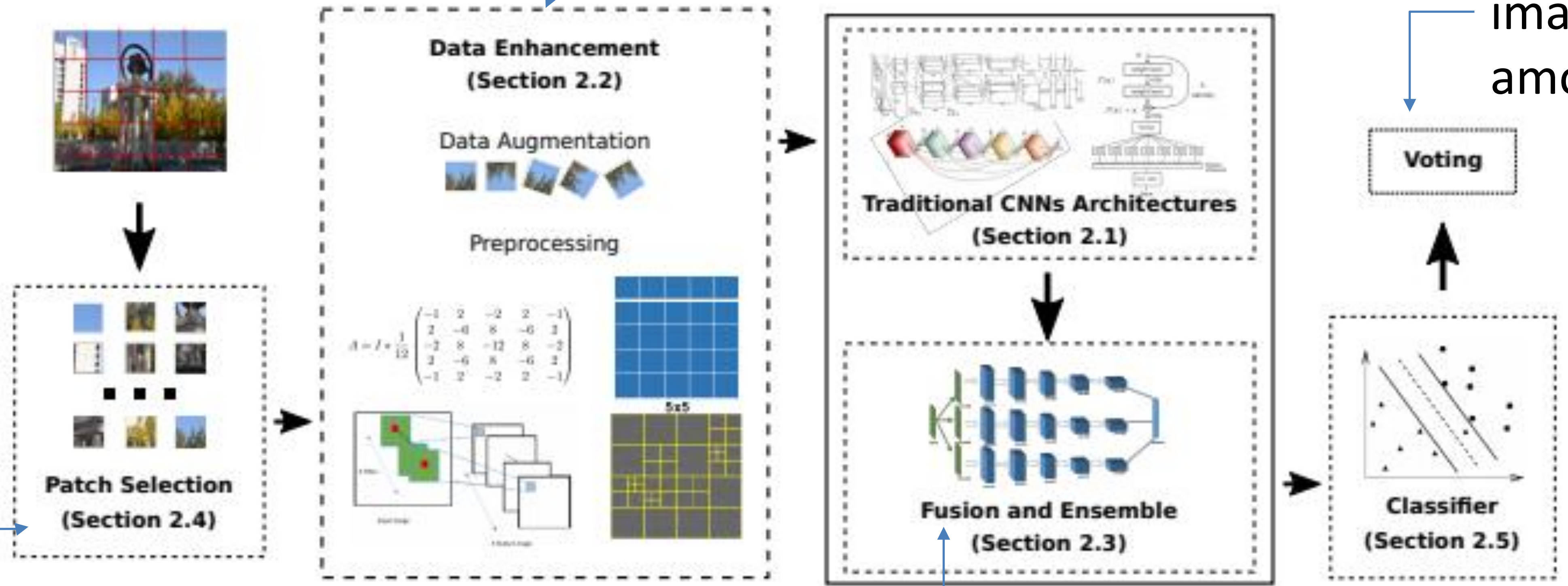


Figure 2. The framework of the deep learning-based algorithms for source forensics. (Section X.X) indicates the subsection where the related technique is described in detail.

Cropped into small and fixed-size pixel patches
Patches are processed
Choose the patches by selection strategy

- By having a softmax layer at the end of the network
- By training a separate classifier on the features extracted by the last layer of the CNN

02 Source Camera Identification

2.1 Traditional Convolutional Neural Networks(T.CNN)
2.2 Data Enhancement (D.E.)
2.3 Fusion and Ensemble (F./E.)
2.4 Patch Selection (P.S.)
2.5 Classifier (C.)

02 Source Camera Identification

2.1 Traditional Convolutional Neural Networks(T.CNN)

Purpose : which is to trace where an image is from

Bondi[23]

- Deep learning based schemes Source camera identification Firstly introduced by Bondi
- Path-breaking method used a simple architecture with five layers including three conv layers, two FC layers

On full resolution image

	Arch.	Input Size	D.A.	F./E.	P.S.	C.	Train : Test	Dataset	Perf. (Patch)		Perf. (Voting)	
									Model	Sensor	Model	Sensor
[23]	A1	48 × 48 × 3	-	-	-	Softmax	7:3	Dresden [62]	72.9% (27)	29.8% (74)	94.1% (27)	-

Freire-Obregon[24]

- Proposed six layer CNN, including two conv layers, one max pooling layer, three fully connected layers
- Used in L-ReLU, led to slightly better performance than ReLU

[24]	A2	32 × 32 × 3	-	-	-	Softmax	MICHE-I [63]	98.1% (3)	91.1% (5)	-	-

02 Source Camera Identification

2.1 Traditional Convolutional Neural Networks(T.CNN)

Huang[25]

- Presented an architecture similar to the on proposed by Bondi
- Improve over the accuracy obtained by Freire-Obregon
- Using Batch Normalization and more conv layers

[25]	A3	36 × 36 × 3	-	-	-	SVM	8:2	Dresden [62]	-	-	-	99.9% (10)

Yao[26]

- Put forward a 13 layers CNN
- Robust against JPEG compression and noise adding
- Not resistant to re-scaling operation

[26]	A4	64 × 64 × 3	-	-	✓	Softmax	3:2	Dresden [62]	93% (25)	-	>98% (25)	-

Chen[26]

- Using ResNet with 26 layers
- Accuracy 99.12%(brand-), 94.73%(model-), 45.81%(devide-level)

[27]	A5	256 × 256 × 3	-	-	-	Softmax	7:3	Dresden [62]	94.7% (27)	45.8% (74)	-	-

02 Source Camera Identification

2.1 Traditional Convolutional Neural Networks(T.CNN)

Ding[28]

Multi task learning

- Combining ResNet with multi-task learning strategy
- Three task(brand-level, model-level, sensor-level) are integrated into one framework

[28]	ResNet-modified	48 × 48 × 3	✓	-	-	Softmax		Dresden [62]	-	-	79.71% (27)	53.4% (74)

Marra[28]

- Used XceptionNet, voting strategy

[29]	A6	64 × 64 × 3	-	-	-	Softmax	8:2	VISION [64]	-	80.77% (35)	-	97.47% (35)
	DenseNet-40	32 × 32 × 3							-	87.96% (35)	-	95.06% (35)
	DenseNet-121	224 × 224 × 3							-	93.88% (35)	-	99.10% (35)
	XceptionNet	299 × 299 × 3							-	95.15% (35)	-	99.31% (35)

02 Source Camera Identification

2.1 Traditional Convolutional Neural Networks(T.CNN)

Realistic case is the open-set scenario, where information about query image is not completely known
A number of deep learning methods for this more challenging scenario have been proposed.

Bayar and Stamm[34]

- Aims to judge whether the device that captured the query image is known or unknown

1. Uses a confidence score mapping with a thresholding strategy to evaluate whether the true source camera model is known or unknown
2. Uses different classifier on features extracted by a CNN

Mayer and Stamm[35]

1. Features are extracted form the last layer of a CNN and fed into a Siamese network to learn a measurement of source similarity
2. Verifying if two query images are captured by same device or not.

Source Camera Identification tasks closely follow the one of architectures proposed for computer vision tasks

02 Source Camera Identification

2.2 Data Enhancement (D.E.)

Including data augmentation and pre-processing → way to improve the performance

Bondi[36,37]

- Normalized the images by subtracting the pixel-wise mean value

Kamal[31]

- Random crops, random rotations, image manipulations(JPEG compression, gamma correlation, and resizing), image addition, empirical mode decomposition

Result

Adding more images has a great impact to performance
Using manipulated images enhances the robustness of the CNN model

02 Source Camera Identification

2.2 Data Enhancement (D.E.)

Computer vision task are seriously dependent on the image contents

Whereas the opposite is true when dealing with source camera identification

The correct class to be attributed to an example is heavily dependent on the noise component introduced by camera acquisition

To reduce the interference of the image contents

G: spatial filter F: denoising filter I : input N: noise

$N = I - F(I).$ $N = I * G$

Tuama

chose wavelet-based denoiser (have been widely used in model-based schemes based on PRNU

HP filters yield better results than wavelet-based denoiser when used in CNN-based scheme

Bayar and Stamm

evaluated the effect of median filter(3x3)

Ding

Evaluated the case of gaussian filter residuals and verified their effectiveness

02 Source Camera Identification

2.2 Data Enhancement (D.E.)

Yang
Presented self-learning filters as a way to further improve the SNR

Bayar and Stamm
Proposed a novel constrained convolution which ensures that learned high-pass filters are within a given bound

Wang
Instead of designing the filter, used LBP to code the image

1. Image → LBP coding operation
2. Fed into CNN

- self-learning filters, constrained convolutions, and LBP coding → outperform HP filters

Zuo
Indicate that the CNN model without pre-procession provides better performances

Most DE technique aims to reduce the influence of image contents by filtering out information deemed not useful

02 Source Camera Identification

2.3 Fusion and Ensemble (F./E.)

Fusion and Ensemble Aim to enhance performance by fusing multiple models and features together

Yang[42]

- Merging three models → increasing accuracy

[42]	A11	64 × 64 × 3	√	√	-	Softmax	4:1	Dresden [62]	-	94.14% (9)	-	-

Bayar and Stamm[38]

- Combined constrained convolutions and MFR at the first layer of CNN → slightly increasing

[38]	A8	256 × 256 × 2	√	√	-	ET	4:1	Dresden [62]	98.58% (26)	-	-	-

Kamal[31]

- Ensemble feature of DenseNet201 trained using three image scales (64x64, 128x128, 256x256) → beneficial

[31]	DenseNet-201 + SE-Block	256 × 256 × 1	√	√	√	SE-block	3.2:1	SPC2018 [7]	98.37% (10, weighted)	-	-	-

Ferreira[33]

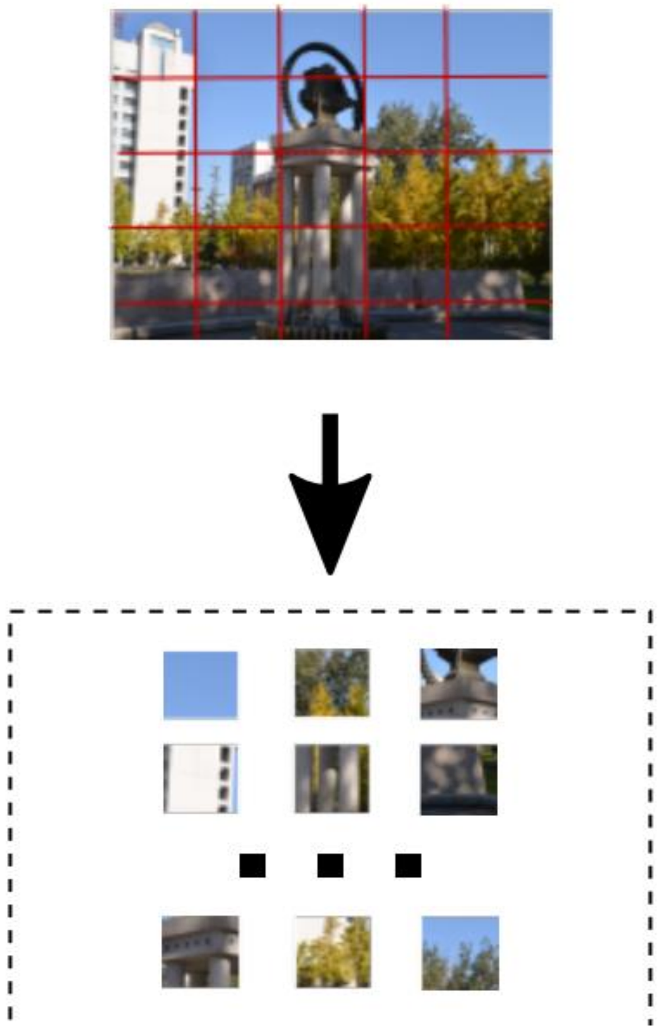
- Integrate InceptionNet and XceptionNet → boost increase

[33]	Inception-Xception	299 × 299	-	√	√	Softmax		SPC2018 [7]	93.29% (10, weighted)	-	-	-

02 Source Camera Identification

2.4 Patch Select (P.S.)

A good strategy for choosing the best pixel patches to be used for CNN training can be essential to obtain higher performance



Bondi[37] , Kamal[31]

- only select for training the pixel patches whose average values are close to half of the image dynamic range.
- Pixel patches with higher measure value are used to train the CNN model

Guera[46]

- proposed a CNN-based solution to estimate

Yang[43]

1. Pixel patches were separated into three subsets according to their mean and Variance
2. Different CNN model would be trained on each subset
3. Query pixel patches would be classified using the model corresponding to their characteristics.

dynamic range of a patch is considered to be the best descriptor for its usefulness for the task at hand.

02 Source Camera Identification

2.5 Classifier (C.)

Some image forensics researchers have recently explored if the adoption of different classifiers can improve the performance

Classifiers can achieve better performance with respect to simple softmax layers

SVM, ERT, cosine similarity measure, nearest mean score, squeeze and excitation block