CAMERA MODEL IDENTIFICATION WITH RESIDUAL NEURAL NETWORK

*Yunshu Chen, Yue Huang, Xinghao Ding**

School of Information Science and Engineering, Xiamen University, Xiamen, China

ABSTRACT

With the development of multimedia, camera model identification from given images has attract large attentions in cyber-forensic area recently. The task has achieved a great improvement due to some deep learning methods, where the features are extracted with the stacked architectures. However, it should be considered that both low-level and high-level features have contributions to the recognition. In this paper, we investigate the task with another deep learning model, residual neural network (ResNet). Proposed framework has been evaluated on the experiments of brand-attribution, model-attribution and device-attribution. Besides, we also include cell phone model identification in the brand-attribution experiment for the first time. The classification results have demonstrated that the proposed work has the ability of enhancing the identification performances compared with existing methods in each specific task. The proposed work can be considered as an effective approach on image forensics.

Index Terms— camera model identification, image forensics, deep learning, ResNet

1. INTRODUCTION

With the development of science and information, digital images have penetrated into our social life. The extensive use of digital images has also promoted the development and application of digital images editing software such as Adobe Photoshop, Instagram and so on. However, as downloading and copying digital material are becoming easier over the years, verification of image primitiveness and authenticity has become urgent need. Image forensics [1] technology is proposed in this context, aiming to authenticate the primitive and authenticity of the images by blind analysis. It is useful for copyright infringement cases, ownership attribution, as well as for detecting the disseminators of illicit material.

*Corresponding author: dxh@xmu.edu.cn. The work is supported in part by National Natural Science Foundation of China under Grants 61571382, 81671766, 61571005, 81671674, U1605252, 61671309 and 81301278, Guangdong Natural Science Foundation under Grant 2015A030313007, Fundamental Research Funds for the Central Universities under Grant 20720160075, 20720150169, CCF-Tencent research fund, Natural Science Foundation of Fujian Province of China (No.2017J01126), and the Science and Technology funds from the Fujian Provincial Administration of Surveying, Mapping, and Geoinformation.

The starting point for digital forensics is to analyze and understand the operational history of digital images by extracting the inherent traces left in the digital image cycle. There are three parts in the complete cycle of digital image processing: image acquisition, image coding and image editing. Each part of the full cycle described above leaves different operating traces (fingerprint characteristics). In this paper, we focus on image acquisition fingerprint.

In the research of image acquisition fingerprint, the digital image is analyzed by low-level such as the characteristics of the lens, the characteristics of the sensor and the CFA mode [2, 3]. Traditional camera model identification methods require to compute a model, Photo Response Non-Uniformity(PRNU) [4, 5, 6], to identify a camera and evaluate a statistical proximity(correlation) between the model and the tested image. Lukas et.al [5] proposed the methods which used sensor pattern noise as a fingerprint for uniquely identifying sensors to identify a camera device. Choi et.al [7] used the lens radial distortion to identify a camera source. Each camera has a unique radial distortion pattern, so it can be used as a camera fingerprint to identify camera model. Dirik et.al [8] used the sensor dust patterns in digital single lens reflex cameras as a means of device identification. Different from the traditional method with hand-crafted feature, Baroffio [9] and Tuama [10] referred to use convolutional neural networks(CNN) to identify the camera source, which can extract high vision-level features from images shot with different cameras automatically. The work in [9, 10] can be recommended as the state-of-the-art in the task.

In the proposed work, we want to re-consider the task in the view of real-world application. Firstly, in previous works, it has been discussed that low-level features such as the lens characteristics also contribute to the recognition [2, 3]. Secondly, traditional CNN is a still shallow network, since the stacked architecture has the problem of vanishing/exploding gradients [11]. Thus, CNN is still limited to be a real deep neural network. Finally, with the rapid development of mobile Internet, the number of images captured by cell phones is dramatically increasing. But very few discussions have been reported to cell phone camera identification.

ResNet is a recent developed deep learning method, and it can be considered as the state-of-the-art CNN model [11]. Compared with traditional CNN models and the stacked architectures, a typical unit in ResNet contains a residual map-

ping function and a shortcut, where both low-level and high-level features are extracted simultaneously. Thus the feature extraction can be improved by the boosting effect. It is much easier for ResNet to increase the depth of the network for higher vision level features [12].

In this paper, we first propose to use a deep learning method ResNet to identify the camera source, the architecture of which makes it possible to jointly utilize the features from high-level vision and low-level vision. The proposed work has been evaluated on four different tasks in image forensics, including brand-attribution, model-attribution, device-attribution from cameras and brand-attribution between cell phones and cameras. The experiment results via proposed model have classification accuracy higher than 99% in 13 camera brands identification, 94% in discriminating 27 camera models, 45% in discriminating 74 camera devices and 97% in discriminating 13 camera brands and 6 cell phone brands. The contributions can be summarized as: 1) consider the ability of extracting features from both low-level and high-level, a ResNet based camera model identification method is proposed to enhance the performances; 2) the proposed method is able to extract higher level features with the real deep neural network model; 3) to the best of our knowledge, it is the first report that also considers cell phone identification, which may receive increasing attentions in recent forensics areas of mobile multimedia.

2. RESNET FOR CAMERA IDENTIFICATION

Deep learning methods show amazingly good performance in several computer vision applications like image classification and object recognition [13]. Identifying which camera model shot a given image within a set of possible candidates is quite similar to image classification. The input data is an image and the output is a label to the camera that shot the input picture.

Depth of deep learning has great influence in classifying and recognition. The same as many visual recognition tasks which have greatly benefited from very deep models, camera model identification also works well in deeper networks. However learning better networks is not as easy as stacking more layers. The problem of vanishing/exploding gradients comes as the increasing depth [11]. The degradation of training accuracy also indicates that stacked CNN is not easy to optimize. But the recognition accuracy of shallower network such as traditional CNN is worse than deeper networks. ResNet solves these problems cleverly. Compared with traditional CNN, ResNet adds a shortcut connection [11] to every few stacked layers. Instead of direct fitting a desired underlying mapping, ResNet lets these layers fit a residual mapping, which makes the data flow between networks more smoothly, and solves the problem of gradient disappeared and degradation of training accuracy.

CNN use the characteristic of last layer to classify and identify. From the experiments results of [9, 10], we can see that characteristic of high level is effective for camera model identification. In the same time, the traditional method with hand-crafted feature [2, 3] maily focus on the characteristics of the lens, the characteristics of the sensor and the CFA mode, all of which are low level features. Therefore, we consider use ResNet model which can boost a set of discriminative features from multiple layers together and send these features to classification to judge which camera shot the given image. ResNet integrate low/mid/high/ level features and classifiers in an end-to-end multilayer fashion, and the levels of features can be enriched by the number of staked layers(depth). It has led to a series of breakthroughs for image classification.

Table 1. Characteristics and meta-parameters of ResNet.

Layer_name	Parameter	Value
Conv1	$Kernel\_size$ Num_out $Stride$ Pad	7×7 64 2 1
Conv2_x	$\begin{bmatrix} Kernel\_size & Num\_out \\ Kernel\_size & Num\_out \end{bmatrix}$ *repeat	$\begin{bmatrix} 3\times3 & 64 \\ 3\times3 & 64 \end{bmatrix}$ *3
Conv3_x	$\begin{bmatrix} Kernel\_size & Num\_out \\ Kernel\_size & Num\_out \end{bmatrix}$ *repeat	$\begin{bmatrix} 3\times3 & 128 \\ 3\times3 & 128 \end{bmatrix}$ *3
Conv4_x	$\begin{bmatrix} Kernel\_size & Num\_out \\ Kernel\_size & Num\_out \end{bmatrix}$ *repeat	$\begin{bmatrix} 3\times3 & 256 \\ 3\times3 & 256 \end{bmatrix}$ *3
Conv5_x	$\begin{bmatrix} Kernel\_size & Num\_out \\ Kernel\_size & Num\_out \end{bmatrix}$ *repeat	$\begin{bmatrix} 3\times3 & 512 \\ 3\times3 & 512 \end{bmatrix}$ *3
Full Connect	$GlobalPooling$ Num_out	Average 13/27/74/19
Softmax Loss	-	-
Accuracy	-	-

From above explanation, we show that:1) ResNet is easy to optimize and can easily enjoy accuracy gains from greatly increased depth, but the simply stack layers (stacked CNN) exhibit higher training error when the depth increases; 2) CNN only use the characteristic of last layer to classify and identify and ignore the low level features. However, ResNet can integrate multiple features together to classify and recognize, which works well in camera model identification.

To train a camera model identification system with ResNet, we need:

1. to define the meta-parameters of the ResNet, such as the number of layers, the number and the size of the filters in convolutional layers. Details have been listed in Table 1.

2. to define a proper cost function to be minimized during training process. In this paper, we choose a soft-max function to compute the loss.

3. to prepare a suitable dataset of training and testing process. This will be mentioned in Section 3. In order to fit the ResNet conditions, we resize all of the images into 256×256. Minimum error rate is recorded after convergence

3. EXPERIMENTS

In order to validate the effectiveness of the proposed approach, four experiments including brand-attribution, model-

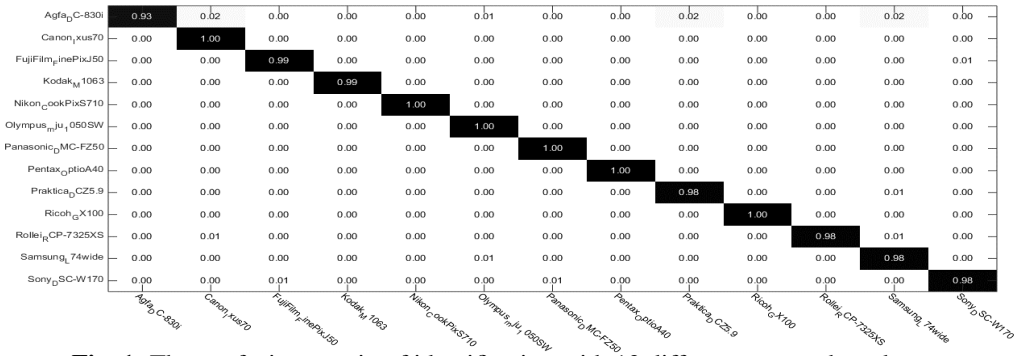

Fig. 1. The confusion matrix of identification with 13 different camera brands.

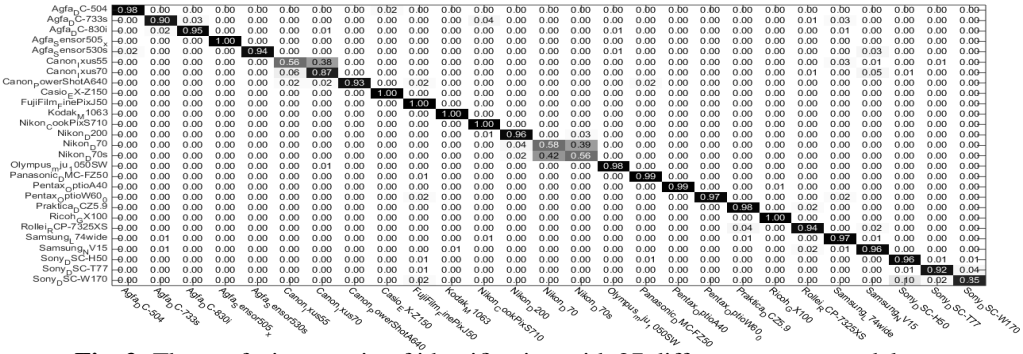

Fig. 2. The confusion matrix of identification with 27 different camera models.

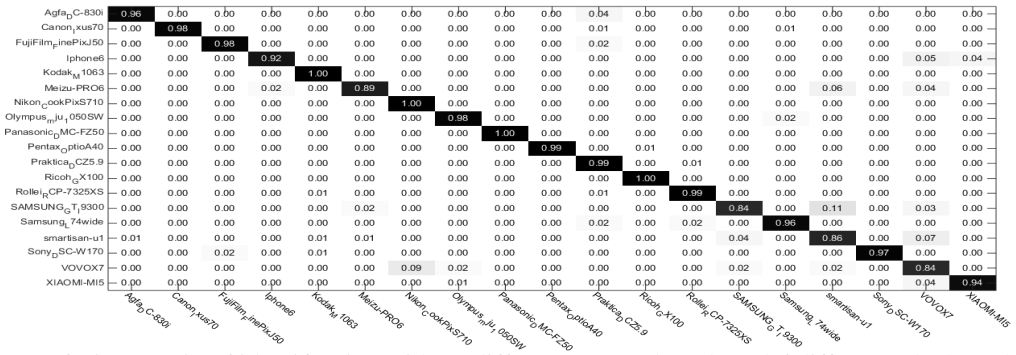

Fig. 3. The confusion matrix of identification with 13 different camera brands and 6 different telephone brands.

attribution and device-attribution have been applied. There are two public datasets in the validation. The first one contains 16960 natural images from 74 devices at 13 camera brands in Dresden database [14]. And the second set includes 6 personal cell phone brands which has 1444 images and are described in Table 2. In order to protect the privacy, we only choose the cell phone photos based on scenery, plants or food. Each brand has different photos but all of them have similar contents. The experiments are done in a server with NIVDIA GeForce 1080 GPU, 64G RAM and a56 Intel(R)Xeon(R) CPU E5-2683 V3@2.00GHz. Before any further manipulation, the dataset is divided into training and testing sets randomly, where 70% of the data is chosen for training and the rest 30% is for the testing data. There is no overlapped between training and testing data.

Table 2. Another 6 different cell phone brands added in brand-attribution experiment to identify cell phone source.

Seq.	Brand	Model	Original Resolution	No. Images 256×256
1	Samsung	GT-I9300	3264×2448	210
2	IPhone	6	3264×2448	319
3	VIVO	X7	3120×4160	202
4	Smartisam	U1	3096×4128	256
5	Xiaomi	MI5	2448×3264	268
6	Meizu	Pro6	5312×3984	189

3.1. Brand-attribution camera identification

The first experiment uses 13 different camera brands in order to check if the proposed model can identify different camera brands. ResNet model is trained on the 13 camera brands, in totally about 10038 images which are resized into 256×256.

Then we use the trained ResNet model to detect what brand of the camera each image in the test set comes from, and to count the accuracy of the recognition. Experiment result shows that the overall accuracy is equal to 99.12%. As shown by the confusion matrix (Fig. 1), the proposed model is quite good at discriminating different brands. Almost each brand can achieve 100% accuracy.

3.2. Model-attribution camera identification

This experiment is tested on 27 camera models in Dresden database [14], some of which come from the same brand. There are 16960 images in total. In the experiment, we can achieve a total identification accuracy as high as 94.73%. Fig. 2 shows the confusion matrix of the model-level camera identification task.

From Fig. 2 we can see that, the best identification accuracy is recorded for the camera models Agfa_sensor_505x, Canon_EXZ150, FujiFilm_FinePixJ50 etc., all of which gain 100% accuracy. That is because there are not other camera models coming from the same manufacturer as them. Similarly, the identification accuracy decreases due to the fact that the captured images from camera models of the same manufacturer are sometimes harder to separate, such as Canon_xus70 and Canon_xus55, Nikon_D70 and Nikon_D70s. Canon_xus55 has 38% to be mistaken as Canon_xus70, while Nikon_D70 has about 40% possibility to be mistaken as Nikon_D70s. That is because different camera models from the same brand have the strong feature similarity [15]. Because of the use of global pooling, we can only extract one 256×256 patch in test which can reach the same 94% accuracy as the existing CNN [9] that should extract 25 patches in test.

3.3. Device-attribution camera identification

In this experiment, we input images from 74 different camera devices, some of them may have same brand or same model. The total accuracy on the test set is only 45.81%. It is a known that the accuracy decreases as the number of classes increases. From the confusion matrix, we can also discover that, different devices from same brand or same model are hard to separate.

Although these devices have lower accuracy, their correct rate is still higher than the error probability. For example, we test three different FujiFilm_FinePixJ50 devices, as shown in Table 3, they still have more than 40% probability to classify correctly. But we have to recognize that ResNet model can identify different types of camera brands well, and the performance of a single instance of the same camera model is also quite good, whereas when there are multiple cameras belonging to the same model, the performance of our system degrades. However, compared with CNN model [9], which only gained 29.8% accuracy (shown in Table 4), the proposed model has received great progress.

Table 3. The parts of the confusion matrix of 74 different camera devices (0, 1, 2 are device index).

Different devices of FujiFilm_FinePixJ50	0	1	2
0	**0.5645**	0.1452	0.2742
1	0.2941	**0.4706**	0.2353
2	0.3030	0.2879	**0.4091**

Table 4. Identification accuracies for all the experiments compared to CNN [9], AlexNet and GoogLeNet, bold denotes the best results.

Method	EXP1 13 labels	EXP2 27 labels	EXP3 74 labels	EXP4 19 labels
CNN [9]	-	72.9%(1patch) 94.1%(25patches	29.8%	-
AlexNet [16]	62.21%	45.41%	16.19%	61.06%
GoogLeNet [17]	96.31%	87.25%	35.09%	92.81%
ResNet	**99.12**%	**94.73**%	**45.81**%	**97.73**%

3.4. Brand-attribution from telephone and camera

Cell phones become more and more popular in our daily life. It is very convenient for people to take pictures from cell phone. Identifying cell phone camera models also have practical significance, but it is more challenged. So we add another 6 cell phone models into 13 camera models mentioned in brand-attribution experiment. The precision for this task is as high as 97.73%.

An image shot by camera or cell phone is determined by lens, sensor, image processor and camera program (algorithm). From the confusion matrix (Fig. 3), it can be observed that the proposed model also identifies cell phone models with satisfied accuracy. Among the 6 cell phone models, XIAOMI-MI5 has the highest accuracy of 94%. Samsung-GT-I9300 has 11% probability of being mistaken as Smartisan-U1. It can be explained that the camera manufacturer of Smartisan-U1 and Samsung-GT-I9300 is the same (Samsung).But they have different imaging algorithm. Therefore, the proposed algorithm still receive a high accuracy. A conclusion of results from all the experiments have been shown in Table 4. It can be observed that proposed model outperform existing CNN [9] based model in device-attribution identification task. We also compare our result with the most popular classifer models, AlexNet [16] and GoogleNet [17]. The experiment results show that the proposed model has good performance in all tasks.

4. CONCLUSIONS

In this paper, a camera model identification method is proposed based on ResNet. Compared with existing deep learning methods in the task, the proposed method has stronger learning power, and can extract features from different levels simultaneously. Both of them bring benefit to the proposed application. The method is evaluated with four experiments. The higher accuracy demonstrate that proposed method outperform the one with CNN. Besides, we also extend to test on telephone model identification in brand-attribution experiment for the first time, which can be considered as a great progress on image forensics field.

5. REFERENCES

[1] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Computing Surveys*, vol. 43, no. 4, pp. 26, 2011.

[2] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on cfa interpolation," in *IEEE International Conference on Image Processing (ICIP)*, 2005.

[3] S. Milani, P. Bestagini, M. Tagliasacchi, and S. Tubaro, "Demosaicing strategy identification via eigenalgorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[4] J. Lukas, J. Fridrich, and M. Goljan, "Determining digital image origin using sensor imperfections," in *Proc. SPIE, Image and Video Communications and Processing*, 2005.

[5] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[6] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," in *Proc. SPIE, Media Forensics and Security*, 2009.

[7] Kai San Choi, Edmund Y Lam, and Kenneth KY Wong, "Source camera identification using footprints from lens aberration," in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 60690J–60690J.

[8] A. E. Dirik, H. T. Sencar, and N. Memon, "Source camera identification based on sensor dust characteristics," in *IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, 2007.

[9] Luca Baroffio, Luca Bondi, Paolo Bestagini, and Stefano Tubaro, "Camera identification with deep convolutional networks," *arXiv preprint arXiv:1603.01068*, 2016.

[10] A. Tuama, F. Comby, and M. Chaumont, "Camera model identification with the use of deep convolutional neural networks," in *IEEE International Workshop on Information Forensics and Security*, 2016.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[12] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, pp. 1–21, 2015.

[13] Y. LeCun and Y. Bengio, "The handbook of brain theory and neural networks," *chapter Convolutional Networks for Images, Speech, and Time Series*, pp. 255–258, 1998.

[14] T. Gloe and R. B-ohme, "The dresden image database for benchmarking digital image forensics," *Journal of Digital Forensic Practice*, vol. 3, no. 2-4, pp. 150–159, 2010.

[15] M. Kirchner and T. Gloe, "Forensic camera model identification," *Handbook of Digital Forensics of Multimedia Data and Devices*, pp. 329–374, 2015.

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.