A ROTATION-INVARIANT CONVOLUTIONAL NEURAL NETWORK FOR IMAGE ENHANCEMENT FORENSICS

Yifang Chen[1], Zixian Lyu[1], Xiangui Kang[1], and Z. Jane Wang[2]

1. School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China, 510006
2. Department of ECE, University of British Colombia, Vancouver, Canada.

ABSTRACT

Many proposed complex convolutional neural network (CNN) models in image forensics are with a large number of parameters, requiring a huge number of training data and having the risk of being overfitting. Considering the desired rotation invariance in the detection of some specific image manipulations, i.e., image enhancement, we propose employing convolutional filters with an isotropic architecture in the CNN model which can significantly reduce the required number of CNN parameters. With the same weights in symmetric positions, the proposed filter can extract rotation-invariant features for image enhancement forensics. Experimental results show that the proposed rotation-invariant CNN models with much less parameters can achieve much better performance, e.g., yielding more than 13% improvement in terms of detection accuracy in Gamma correction forensics. It also achieves significantly better generalization performances on different databases and better robustness against JPEG compression when compared with the popular BayarNet in [16].

Index Terms—Image enhancement forensics, convolutional neural networks, rotation invariant, constrained isotropic filter

1. INTRODUCTION

With the rapid development of digital media editing technology, digital image manipulation becomes very easy and convenient even for an inexperienced forger with the aid of user-friendly photo-editing software, e.g., Adobe Photoshop. During the past decades, multimedia forensics has been an active research area, and many blind forensic techniques were proposed by utilizing statistical fingerprints to verify the authenticity of digital multimedia data. Previous studies mainly focus on detecting different types of alterations, which can be broadly divided into two categories:

1) Non-content-changing operations, including resampling [1], compression [2] and image enhancement operations which includes sharpening filtering [3], contrast enhancement (e.g., Gamma correction [4]–[6] and S mapping [7]) and median filtering [8][9].

2) Content-changing operations, e.g., splicing and composition [10]–[12].

Image enhancement operations are commonly used as a retouching tool. While these operations alter the perceptual quality without changing the content of a digital image, their detections are still forensically significant. Since image enhancement may be used as a part of an operation chain to hide the forgery of an image, its detection can serve as a warning sign for possible image forgery. Much research has focused on identifying histogram peak/gap artifacts in images and then developing algorithms to detect these traces [4]–[7]. These forensic algorithms work well under the assumption that the gray-level histogram of an unaltered image exhibits a smooth contour. However, digital images are often post-processed in real applications, such as a heavy compression with a middle/low quality factor (QF). Post-processing might weaken or even remove these aberrant features in gray-level histogram. In such a scenario, the existing approaches [4]–[7] may fail to detect the enhancement operations.

In recent years, there has been a growing number of deep learning approaches used in blind image forensics [13]–[17]. These CNN models with a huge number of parameters require a large-scale training data to boost the performance. However, it is difficult to collect a large number of forensic images practically on demand, whereas training CNNs with limited data increases the risk of being overfitting. Therefore, designing a novel CNN architecture to learn robust and general forensic features with fewer parameters is vital for the effectiveness of detection.

Rotation invariance is an important factor to be considered in the detection of image manipulations. Take contrast enhancement as an example, the enhanced image is created via linear or nonlinear mapping for each pixel to adjust global brightness and contrast of the original image. This operation of mapping is identical for the image no matter in rotation of a multiple of 90 degrees or in mirror symmetry. Rotation invariance is essential and a general feature for most of enhancement operations.

In this work, we propose a novel rotation-invariant CNN and focus on six common enhancement operations including unsharp masking sharpening (UMS), Gamma

This work was supported by NSFC (Grant Nos. U1536204, 61772571, 61702429), and the special funding for basic scientific research of Sun Yat-sen University (6177060230).

correction, S mapping, histogram equalization, median filtering and Gaussian filtering. Based on the state-of-the-art CNN architecture [16], all convolutional filters with the kernel size above one are replaced by the proposed isotropic convolutional layers. The constrained isotropic filter (CIF) layer of the proposed CNN serves as the pre-processing module to efficiently suppress the effect of image contents. The following convolutional layers with isotropic architecture can adaptively learn statistical features related to image enhancement detection. Comparing with the original CNN [16], our proposed CNN model with much less parameters has shown better capability on image enhancement detections. Through a series of experiments, we also document the better generalization performance for different databases and the robustness against JPEG compression with a low QF.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the proposed rotation-invariant CNN structure. Experimental results, comparisons and analysis are included in Section 3. Finally, the concluding remarks are drawn in Section 4.

2. THE PROPOSED ROTATION-INVARIANT CNN

A rotation-invariant CNN is proposed in this paper. Compared with the CNN architecture used for forensics applications [16] which is referred as BayarNet, all convolutional filter is replaced with an isotropic convolutional filter, which will be elaborated in Section 2.3. In [16], the authors have investigated the detection of certain image manipulations, but rarely focused on image enhancement detection, so we have redesigned the CNN with isotropic convolutional filters and successfully applied it to the detection of six common enhancement operations. Some design consideration are examined and justified through the experiments.

2.1. Overall Network Architecture

Fig. 1 illustrates the architecture of the proposed rotation-invariant CNN, which contains 6 groups: one preprocessing layer (Group1), other four layer groups (Group 2 ~ 5) and one classification module (Group 6). The feature preprocessing layer is produced to feed to the first convolutional layer, while the last layer of the convolutional module outputs the features to a fully-connected layer followed by an n-way Softmax layer, which produces a distribution of n-class labels.

In the original BayarNet architecture [16], a 5×5 convolutional layer with the constrained filter (CF) serves as the preprocessing layer to adaptively learn pixel values dependencies. Extracting such dependency features may be effective for detecting operations based on adjacent pixels, such as MF, but it is not appropriate to detect histogram-based operations. Therefore, we propose a constrained isotropic filter (CIF) layer to suppress the correlated components (image content) while largely capture useful statistical features related to the enhancement operations.

Group	Output size	Process
Group 1	256×256	Constrained Isotropic Conv 3×(5×5) ,stride=1
Group 2	64×64	Isotropic Conv 96×(7×7), stride=2
		BN+PReLU
		Max pooling 3×3, stride=2
Group 3	32×32	Isotropic Conv 64×(5×5), stride=1
		BN+PReLU
		Max pooling 3×3, stride=2
Group 4	16×16	Isotropic Conv 64×(5×5), stride=1
		BN+PReLU
		Max pooling 3×3, stride=2
Group 5	8×8	Conv 128×(1×1), stride=1
		BN+PReLU
		Average pooling 3×3, stride=2
Group 6	1×1	fully-connected (200 neuros) PReLU
		fully-connected (200 neuros) PReLU
		fully-connected(classes neuros) softmax

Fig. 1. Illustration of the proposed rotation-invariant CNN architecture.

In our investigation, the first five groups of BayarNet (including the preprocessing layer) are viewed as "feature extractors", which are afterwards combined and compacted into a low-dimensional feature vector used for detection. Given the directional invariance of enhancement operations, we modify the convolutional layers of the first four layers into the isotropic architecture to force the statistical modeling to take into account the rotation invariance in enhancement-related features.

In this way of design, the whole CNN can adaptively

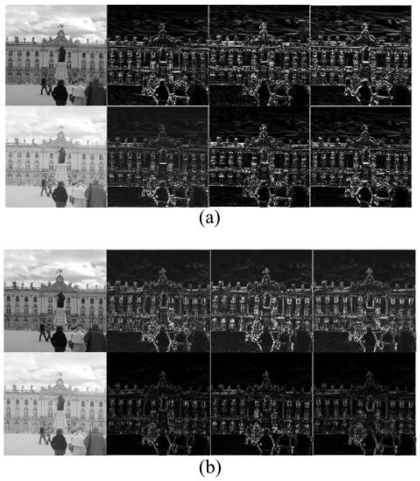

Fig. 2. The images (1st column) and their corresponding feature maps (2nd ~4th column) of the preprocessing layer obtained by (a) CF and (b) CIF respectively, on a raw image (top) and its enhanced (γ = 0.5) version (bottom).

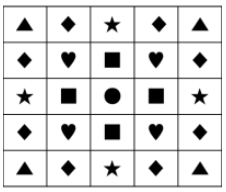

Fig. 3. The weights of a 5×5 isotropic filter.

learns the rotation-invariant features for image enhancement detection and thereby prevents the statistical modeling from grasping local content information. A series of experiments have verified that the proposed rotation-invariant CNN can extract more robust and general enhancement-related features when compared with BayarNet.

2.2. Constrained Isotropic Filter Layer

A convolutional layer with constrained filter (CF) is used in [17] to suppress the image's content and adaptively learn manipulation detection features. To accomplish this, the filter in the preprocessing layer of CNN has the following constraints:

$$\begin{cases} w(0,0) = -1 \\ \sum_{l,m\neq 0} w(l,m) = 1 \end{cases} \quad (1)$$

where, $w(l, m)$ is the filter weight at the (l, m) position and $w(0, 0)$ is the filter weight at the center of the filter window. Each filter in this layer is initialized by randomly choosing each filter weight, then the constraints in (1) are enforced.

Here we enforce the aforementioned isotropic constraints, as illustrated in Fig. 3, to this preprocessing constrained filters and obtain the preprocessing constrained isotropic filters (CIF). The feature maps of the preprocessing layer for a raw image and its enhanced image (γ = 0.5) extracted via CF and CIF are visualized in Fig. 2. The raw image and its enhanced image are shown in the 1st column in Fig. 2, the 1st-3rd feature maps of the preprocessing layer are shown in the 2nd ~4th column in Fig. 2. For the display purpose, the values are normalized to [0, 1]. We can see more obvious contrast difference between CIF feature maps when compared with that of CF, suggesting that CIF is a better choice to extract enhancement-related features automatically.

2.3. The Proposed Isotropic Convolutional Layer

It is noteworthy that image manipulation might have the attribute of directional invariance, which can be used to learn robust and general features for detection. Filters of the convolutional layer allowing feature evolution freely into any form tend to extract some features irrelevant to the enhancement detection, (e.g. the content-dependent features). Therefore, we propose the convolutional layer with the following isotropic constraints. $w(i, j)$ is the weight of a $N \times N$ ($N > 1$) filter at i^{th} row and j^{th} column. With the center $w(N, N)$, all the other $w(i, j)$ are both center symmetrical and mirror symmetrical.

The weights of a 5 × 5 isotropic filter are shown in Fig. 3, the objects with the same shape indicate the weights with the same value. In implementation, during each iteration, each weight is updated based on the stochastic gradient in back-propagation, then the average value of the weights in symmetric positions, *i.e.*, the weights with the same shape as shown in Fig. 3, are calculated. The isotropic constraints are enforced via assigning the average value to the corresponding weights with the same shape.

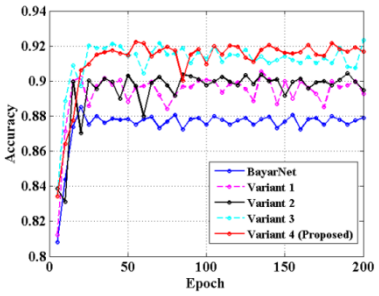

Fig. 4. Comparison of the performance vs. epoch when employing different variants and BayarNet.

No matter rotated by a multiple of 90 degrees, the filters performs same operations on images, thereby serve as the extractor that adaptively learns the properties of rotation invariance. Furthermore, the amount of CNN parameters can be significantly reduced. Take a 5 × 5 filter for instance, there are only 6 parameters in the isotropic filter which is around a quarter of the original one. With more filters, there would be more obvious advantages on reducing the number of parameters. As a result, the proposed CNN model with low complexity can be learned to extract robust and general enhancement-related features.

3. EXPERIMENTAL RESULTS

The primary database used in this study is BOSSbase v1.01 containing 10,000 uncompressed images, which were initially taken by seven cameras in the RAW format and transformed to 8-bit gray-scale images. All images are cropped from the center to size of 256 × 256. The enhanced image versions are generated via the following 6 types of enhancement operations: unsharp masking sharpening (UMS) with different settings of σ and λ: σ = 1, λ = 1.5; σ = 1.3, λ = 1; σ =0.7, λ = 1, Gamma correction with γ = 0.5 and 2, S mapping, histogram equalization, median filtering with 5 × 5 filter and Gaussian filtering with 5 × 5 filter. Therefore, for each classification problem, the dataset contains 10,000 pairs of images. 8,000 pairs are randomly selected for training; the remaining 2,000 pairs are used for testing. In the training phase, each CNN submodel is trained on 6,000 pairs and validated on 2,000 pairs. Only the training pairs contribute to updating the weights, and the validation pairs are used to determine when to stop the training of CNN. All experiments employing the CNN reported in this study are performed on a modified version of the Caffe toolbox on Nvidia Tesla K80 GPUs. The training parameters of the stochastic gradient descent approach are set according to paper [16]. A mini-batch of 64 images is used as the input for each training iteration.

3.1. Comparison with Different Variants

In order to investigate the influence of the numbers of the layers modified to the proposed isotropic architecture, we propose different variants based on BayarNet and test their accuracy for the detection of UMS with σ = 1, λ = 1.5. Considering the obvious of performance gap between different variant, here we use smaller images which are

TABLE I. BOSSBASE: DETECTION ACCURACY (%) FOR ENHANCEMENT.

Method / Operation type		Proposed	BayarNet
UMS	$\sigma=1, \lambda=1.5$	98.54	95.93
	$\sigma=1.3, \lambda=1$	97.72	92.90
	$\sigma=0.7, \lambda=1$	97.39	89.19
Gamma Correction	$\gamma=0.5$	95.96	86.14
	$\gamma=2$	94.98	81.29
S Mapping		96.21	94.25
Histogram Equalization		99.22	97.83
Median Filtering		99.98	98.91
Gaussian Filtering		99.95	98.93

TABLE II. BOW: DETECTION ACCURACY (%) FOR ENHANCEMENT.

Method / Operation type		Proposed	BayarNet
UMS	$\sigma=1, \lambda=1.5$	97.39	94.89
	$\sigma=1.3, \lambda=1$	97.02	90.60
	$\sigma=0.7, \lambda=1$	97.25	87.17
Gamma Correction	$\gamma=0.5$	85.23	78.25
	$\gamma=2$	83.37	70.86
S Mapping		89.99	87.68
Histogram Equalization		98.25	95.50
Median Filtering		99.05	98.97
Gaussian Filtering		99.28	98.01

cropped from center to size 64 × 64. There are four variants considered: Variant 1 to Variant 4. The index of the variant indicates the number of isotropic convolutional layers. Variant 1 indicates that only the convolutional filters in Group 1 as shown in Fig.1 is isotropic, Variant 4 indicates that all convolutional filters in Group 1 ~ Group 4 are isotropic. In Fig. 4, the training accuracy of different variants and BayarNet are shown. It can be seen that all the variants outperform the BayarNet. Variant 4 is obviously better than Variant 1 and Variant 2. There is a slight improvement on accuracy when compared with Variant 3, but Variant 4 has the lowest model complexity in terms of the number of model parameters. Therefore, we choose Variant 4 as our proposed CNN architecture for image enhancement detection.

3.2. Performance of the Proposed Scheme

We compare the detection performance of the proposed CNN with BayarNet when tested on 256 × 256 images. Since the performance reaches saturation after 180 epochs as shown in Fig. 4, for fair comparison, both are trained under the same conditions, e.g., both through 200 epochs. Table I summarizes the test performances of their optimized models. It is observed that the proposed approach significantly outperforms BayarNet for all enhancement detections, especially for the challenging detection (e.g., GC). Compared with BayarNet, the proposed CNN approach improves the accuracies of GC detections by more than 9% for $\gamma = 0.5$ and 13% for $\gamma = 2$ respectively, and the average accuracy results for all the detections cases are improved by around 5%.

3.3. Generalization Performance and Robustness

To illustrate the better generalization performance of our proposed isotropic CNN, after both CNN models have been trained on BOSSbase database as mentioned in 3.2, we test both trained models on the benchmark BOW database, which consists of a total of 10,000 256 × 256 gray-scale images. Table II shows the accuracy results in detecting six enhancement operations. It is obvious that our proposed CNN model still achieves similar much higher accuracy than that of BayarNet despite the discrepancy of the database.

To evaluate the robustness against JPEG compressiong, our optimized models mentioned in 3.2 are used to detect image enhancement under JPEG compression (QF = 40) via transferring leaning. For each kind of operation detection, the dataset used in section 3.2 are compressed with QF = 40 to establish a new database with 10,000 pairs of images. 8,000 pairs of the new database are randomly selected for training and fine-tuning; the remaining, 2,000 pairs, are used for testing. The parameters of the first five groups (Group 1 ~ 5) are transferred from aforementioned optimized models, as mentioned in 3.2, and fixed in the training. Those parameters of the classification module in Group 6 are trained with fine-tuning. The test results for assessing the robustness of our proposed CNN are reported in Table III, which demonstrates its robustness against post-processing even for low-quality JPEG compression.

TABLE III. BOSSBASE: DETECTION ACCURACY (%) FOR ENHANCEMENT UNDER JPEG POST-PROCESSING.

Method / Operation type		Proposed	BayarNet
UMS	$\sigma=1, \lambda=1.5$	94.46	90.91
	$\sigma=1.3, \lambda=1$	94.71	89.96
	$\sigma=0.7, \lambda=1$	95.92	85.89
Gamma Correction	$\gamma=0.5$	93.88	86.45
	$\gamma=2$	95.02	80.92
S Mapping		91.47	82.58
Histogram Equalization		94.73	93.66
Median Filtering		99.66	97.47
Gaussian Filtering		99.05	98.95

4. CONCLUSION

Convolutional neural networks (CNNs) recently were shown promising in the field of digital image forensics, meanwhile how to learn robust and general features of forensics with fewer parameters remains an important and difficult problem to improve detection effectiveness. Given the directional invariance of image enhancement operations, in this paper a rotation-invariant CNN was proposed and successfully applied to six common image enhancement operations including unsharp masking sharpening (UMS), Gamma correction, S mapping, histogram equalization, median filtering and Gaussian filtering. The state-of-the-art CNN architecture for forensics applications is used for comparison. Our proposed isotropic CNN architecture can reduce a large amount of parameters and achieve higher accuracy on BOSSbase database for all six image enhancement operation detections. It also shows better generalization performances on the BOW database and improved robustness against JPEG compression. In the future, we plan to extend the proposed methodological framework and the algorithmic platform for the detections of other types of image manipulations, especially for the operations with directional invariance.

5. REFERENCES

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

[2] T. Bianchi and A. Piva, "Detection of non-aligned double JPEG com-pression based on integer periodicity maps," *IEEE Transactions on Information Forensic and Security*, vol. 7, no. 2, pp. 842–848, 2012.

[3] G. Cao, Y. Zhao, R. Ni, and A. C. Kot, "Unsharp masking sharpening detection via overshoot artifacts analysis," *IEEE Signal Processing Letters*, vol. 18, no. 10, pp. 603–606, 2011.

[4] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *IEEE Transactions on Information Forensic and Security*, vol. 5, no. 3, pp. 492–506, 2010.

[5] M. C. Stamm and K. J. R. Liu, "Forensic estimation and reconstruction of a contrast enhancement mapping," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2010, pp. 1698–1701.

[6] G. Cao, Y. Zhao, and R. Ni, "Forensic estimation of gamma correctionin digital images," in *Proc. of 17th IEEE International Conference Image Processing* (ICIP), 2010, pp. 2097–2100.

[7] G. Cao, Y. Zhao, R. Ni, and X. Li, "Contrast Enhancement-Based Forensics in Digital Images," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 515–525, 2014.

[8] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Proc. of SPIE, Electronic Imaging, Media Forensics and Security II*, 2010, pp. 1–12.

[9] C. Chen, J. Ni, and J. Huang, "Blind detection of median filtering in digital images: A difference domain based approach," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4699–4710, 2013.

[10] D. Mahajan, R. Ramamoorthi, and B. Curless, "A theory of frequency domain invariants: Spherical harmonic identities for BRDF/lighting transfer and image consistency," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 30, no. 2, pp. 197–213, 2008.

[11] I. Yerushalmy and H. Hel-Or, "Digital image forgery detection based on lens and sensor aberration," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 71–91, 2011.

[12] J. O'Brien and H. Farid, "Exposingphoto manipulation with inconsistent reflections," *ACM Transaction on Graphics*, vol. 31, pp. 1–11, 2012.

[13] J. Chen, X. Kang, Y. Liu and Z. J. Wang, "Median Filtering Forensics Based on Convolutional Neural Networks," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1849-1853, Nov. 2015.

[14] Rao Y, Ni J. "A deep learning approach to detection of splicing and copy-move forgeries in images," in *Proc. of 8th IEEE International Workshop on Information Forensics and Security* (WIFS), 2016, pp. 1-6.

[15] Bondi, Luca, et al. "First Steps Toward Camera Model Identification With Convolutional Neural Networks." *IEEE Signal Processing Letters*, vol.24, no.3, pp. 259-263, 2017.

[16] Bayar B, Stamm MC, "Design principles of convolutional neural networks for multimedia forensics," in *Proc. of 2017 IS&T International Symposium on Electronic Imaging*, 2017, pp. 77-86.

[17] Bayar B, Stamm MC, "On the robustness of constrained convolutional neural networks to jpeg post-compression for image resampling detection," in *Proc. of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2017, pp. 2152-2156.