Copy-Move Forgery Detection Based on Pyramid Correlation Network

Peng Liang [1], Yuting Wu [2]*, Huimin Zhao [1], Wa He [1], Gang Hao[1], Shaofa Li [3]

[1]*School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, 510665, China*

[2]*Guangdong Vocational College of Post and Telecom, Guangzhou, 510630, China*

[3]*School of Computer Science, South China University of Technology, Guangzhou, 510642, China*

** Corresponding author.*

Abstract

Block-based image copy-move detection algorithms disregard the spatial layout of the features, leading to the poor detection performance under small-region tampering samples. Therefore, we propose a pyramid correlation network (PCNet) for copy-move forgery detection, whose goal is to obtain rich and detailed image representation via a pyramid cascaded correlation architecture. Experimental results show that PCNet outperforms the comparison algorithm on USCISI, CASIA and CoMoFoD data sets. Compared to the benchmark model BusterNet, F1 scores of PCNet has increased by 33.84% and 30.62% on CASIA CMFD dataset and CoMoFoD dataset respectively.

Keywords:*copy-move; image forgery detection; feature extraction; deep learning*

I. Introduction

Nowadays, digital image has remarkable role in various areas such as journalism, legal service and military. However, with the widespread of photo-editing tools, image forgery has seriously threatened the authenticity and integrity of images. Image copy-move forgery, where one part of image is copied and moved to the other position in the same image, is one of common image forgery. In order to remove the traces left by copy-move forgery, various transformations such as noise adding, image blurring and part deformation are often applied to the tampered image. Generally, there are two common ways to detect copy-move forgery with various transformations. One is to use transformation-invariant feature to learn robust representations, which have many well-known feature descriptors, such as SIFT and SURF. The other is to obtain a robust representation learn from large amount of data. Both of above methods have drawbacks. The hand crafted design of invariant features are infeasible for composited transformations. Learning a robust representation from large amount of data usually costing expensive training. What's more, transformation parameters are assumed fixed and used to build copy-move forgery detection model. The assumption decreases model generalization ability while processing new task with unknown transformation parameters.

In this paper, a pyramid correlation network (PCNet) is proposed for image copy-move forgery detection. The main contribution of our work is that the pyramid correlation feature extraction network enrich spatial information which reinforce the ability of detecting small forgery regions.

II. Related Work

Recently, a great deal of research have been carried out in the field of copy-move forgery detection. The algorithms proposed so far may be categorized depending on the schemes of feature extraction. Based on this criterion, copy-move forgery detection methods may be divided as keypoint-based methods[1-6] and block-based methods[7-18].

In keypoint-based methods, SIFT and SURF were usually adopted as the feature descriptor. Li et al. [1] tried to

divide the image into blocks by using SLIC algorithm, and performed keypoint matching under the constraints of the generated blocks to detect copy-move regions. In [2-4], SIFT was selected as the feature descriptor, while in [5-6], SURF was adopted. Silva E et al. [6] used SURF to detect key points and performed feature matching via Nearest Neighbor Distance Ratio. Their method can work under various challenging conditions, but peforom poorly when forgery region is small or homogeneous.

In block-based methods, multiple features were employed to describe overlapping blocks such as DCT[7], PCA[8], DWT and SVD[9], Zernike moments[10], LBP[11]. Pun et al. [12] proposed an forgery detection algorithm by using block-based feature to locate candidate regions and then use keypoint feature to detect forgery region. Their method is robust to various transformations such as JPEG compression, geometric transformation and downsampling. The limitation conventional methods suffered is that such methods often fail due to the insufficient key points when forgery region is too small or texture smoothly.

In recent years, there are some block-based methods implement copy-move forgery detection via deep neural architecture[13-18]. Rao Y et al. [13] proposed an end to end DNN forgery detection solution which extractd image hierarchical block features from input images. presented a new image forgery detection method based on deep learning technique, which utilizes a convolutional neural network (CNN) to automatically learn hierarchical representations from the input RGB color images. Barni M et al. [14] designed a 4-Twins Net to distinguish the source and target regions of tampered images, where the proposed CNN architecture is capable of capturing interpolation artefacts and boundary inconsistencies of forgery regions. Wu Y et al. [15-18] developed a deep neural network which applied convolution and deconvolution modules to directly generate the forgery detection mask from the input image. Wu Y et al. [18] proposed a manipulation tracing network which can handles several types of image forgery such splicing, copy-move, removal, enhancement. Nevertheless, DL-based methods still have two limitations. One is requiring large amount of data to train a robust representation for various transformations. The other is the convolutional network extract feature map at a fixed kernel which is inflexible for part transformation.

III. Methodology

3.1 Overview

The flowchart of proposed PCNet is presented in Figure 1. It takes an initial image $I \in R^{256\times256\times3}$ as input, and outputs a binary copy-move forgery mask $Y \in R^{256\times256\times1}$. The network architectures of proposed PCNet is shown in Fig. 1.

The key of copy-move forgery detection problem is calculating the similarity between two image patches. To achieve this goal, we define a pyramid correlation network (PCNet), which is composed of two main steps: (1) pyramid correlation pyramid feature extraction, which compute the correlations between two feature pixel and concatenate them as feature pyramid; (2) forgery pixel prediction, which predict a pixel is forged or not by using the correlation pyramid features.

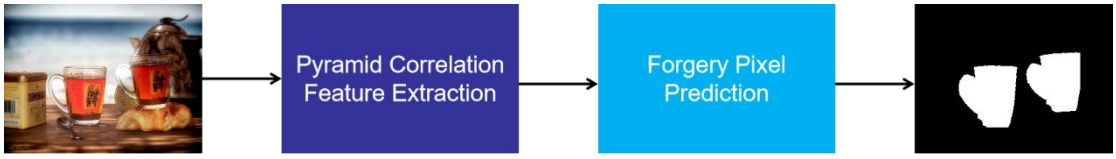

Fig. 1 The Flowchart of proposed PCNet method

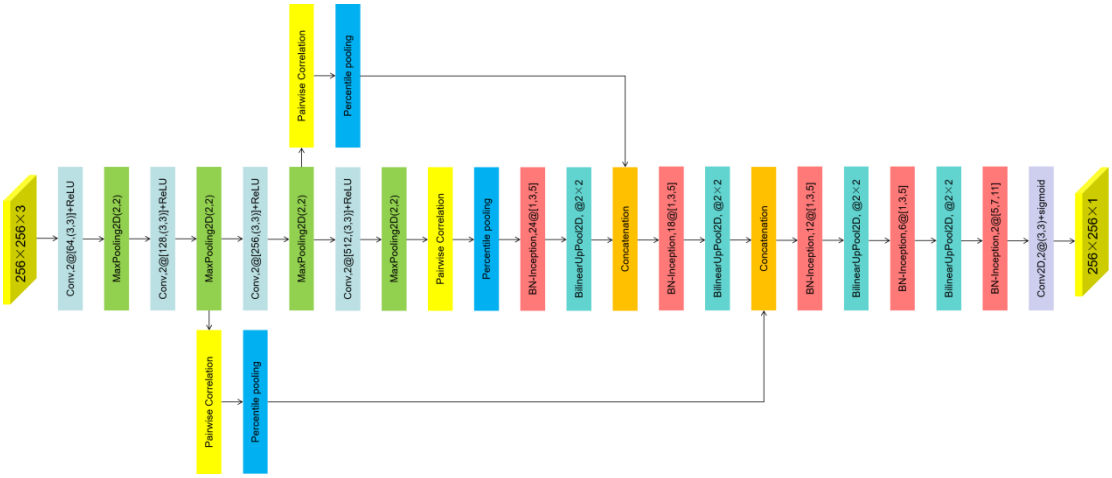

Fig. 2 The network architecture of PCNet solution

3.2 Pyramid Correlation Feature Extraction

Compared to the feature extracted from CNN model which designs for other image classification tasks, the resulting features are more robust to image forgery detection with different transformations. Since the convolution feature representation unit is design for image-level classification task, it aims to obtain the semantic of regions, and it is not appropriated for pixel-level forgery detection task. We find that a simple but effectively fine-to-coarse feature representation is more suitable and better alternative for pixel-level forgery detection.

Given an input image, and produces feature maps at four scales with a scaling step of convolution and max pooling layers. These feature maps extracted from the first to four blocks will be fed to a correlation feature pyramid structure.Without loss of generality, feature extraction takes an manipulation image, extracts feature from feature extractor, compute and sort feature correlation similarity via correlation similarity layer, and applies concatenate layer to concate all features as a whole. More precisely, we apply three layers of Conv with parameter sets 2@[64,(3,3)], 2@[128,(3,3)], and 2@[256,(3,3)] respectively, which followed by the relu activation and MaxPooling. As the result, the feature extraction part outputs feature tensers of size 256×256×64, 128×128×128, 64×64×256, and 32×32×512 at different convolution layers. Fig. 2 shows thenetwork architectures of PCNet.

We compute feature correlation score by computing pairwise similarity and select meaningful features via Percentile Pooling. Let $f \in R^{32\times32\times512}$ be a feature tensor extracted from feature layer, i.e. $f = \{f[i_r, i_c]\}_{i_r, i_c \in [0,\cdots,31]}$. For a feature tensor *f*, given two patch-like feature $f[i](i = (i_r, i_c))$ and $f[j](j = (j_r, j_c))$, the Cosine Similarity is used to quantify the feature correlation by:

$$\cos(\theta_{i,j}) = \frac{\tilde{f}[i] \cdot \tilde{f}[j]}{\|\tilde{f}[i]\| \times \|\tilde{f}[j]\|} \tag{1}$$

$$\tilde{f}[i] = (f[i] - \mu[i])/\sigma[i] \tag{2}$$

Where $\mu[i]$represents the mean of $f[i]$, and $\sigma[i]$denotes its standard deviation. $\tilde{f}[i]$ is the normalized version of $f[i]$.

For a given $f[i]$, a score vector $s[i]$is obtained by computing cosine similarity between $f[i]$and a total of

1024 possible $f[j]$, and sort it in the descending order, namely

$$s[i] = sort([\cos(\theta_{i,0}), \cdots, \cos(\theta_{i,j}), \cdots, \cos(\theta_{i,1023})]) \tag{3}$$

Percentile pooling is used to select the top *K* scores as the preliminary matching result, and the pooled percentile score vector is denoted in Eq(4)

$$P[i] = s[i][k] \tag{4}$$

We append a image convolution with 3×3 kernel on the preliminary matching result. The core intuition behind this step is to eliminate false matches from the preliminary matching result by comparing the neighborhoods of the matched features. The image convolution enhances the data reliability, but does not change the resolution of feature.

Finally, all correlation score vectors are concatenated into a cascaded pyramid vectors using Concatenation, and all the correlation vector can be considered as a whole instead of considering disjointedly. Specifically, the high resolution correlation score vector contains features from low resolution ones. In order to suppress the feature repetitiveness from high resolution correlation score vectors, we assign smaller weights to the high resolution vectors in Concatenation function. Given a series of correlation score vectors $S_k, k = 0,...,n-1$, where S_0 denotes the highest resolution vector and S_n denotes the highest resolution vector, the output of cascade correlation feature P_n can be described as follow:

$$P_n = \left[\frac{1}{2^n} S_0, \cdots \frac{1}{2^{n-k}} S_k, \cdots \frac{1}{2} S_{n-1} \right] \tag{4}$$

3.3 Forgery Pixel Prediction

In ths section, a copy-move forgery mask is generated for forgery pixel prediction.The resolution of similarity score vector is different from forgery image. We thus need to restore the original resolution and thereby generate a copy-move mask consistent with the size of the input image. To achieve this goal, we define a Mask Decoder, which doubles the size of feature map by convolution and upsampling. Specifically, the input feature map will take into three branches (namely, 1×1 convolution, 3×3 convolution, and 5×5 convolution) to reduce channels, and then the three convolved features will be merged. Finally, Bilinear Interpolation will be appended on the merged map to generate the final feature map. More preciously, we restore image using BN-Inception with parameters [64,64,18], [128,128,12], and [256,256,6] respectively, and each BN-Inception is followed by a BilinearUpPool2D[33]. pixel-level forgery is predicted via Binary Classifier with the sigmoid activation.

IV. Experiment

The experiments aim to evalute the efficiency, generalizability and robust to various transformations of proposed PCNet. The PCNet is chosen for comparison with three benchmark methods-- a dense field-based CMFD [6], a deep matching and validation network (DMVN) [17] and a CNN-based method (BusterNet) proposed in [18]. Firstly, we evaluate the efficiency of PCNet with different numbers of training samples. Secondly, followed [18], we evalute the generalizability and robust of PCNet on the two benchmark data sets. Experiments are conducted on Ubuntu 18.04 platform with CPU i7 8700, memory 16GB, graphics card GTX 1080Ti X2, hard disk 2T. Tensorflow framework is applied for network training and testing.

4.1 Dataset

USCISI: The dataset consists of 10^5 images which are taken from SUN2012 and Microsoft COCO dataset, and the copy-move tampered images are obtained by means of geometric transformation.

CASIA TIDE v2.0: The dataset consists of 1313 tampered samples that are of copy-move forgery.

CoMoFoD: CoMoFoD provides 5000 forged images belonging to 25 categories. There are 200 base forged images in the base category, and the remaining 24 categories are created through various transformations(JPEG compression, noise adding, image blurring, brightness change, color reduction, contrast adjustments) on the base category images to hide forgery clues.

4.2 Parameters Setting

Parameters setting of proposed PCNet method is listed in Table.

Table 1. Parameters setting of proposed PCNet method

Parameter	Value
Epochs	50
Optimizer	Adam
Learning rate	0.0001
Loss function	binary_crossentropy
Train data set	90000 images from USCISI dataset
Test data set	10000 images from USCISI dataset
Evaluation data set	CASIA CMFD dataset with 1313 images, CoMoFoD dataset with 200 images
Activation	Relu
Stride of pooling	2
Padding of pooling	Valid
Stride of convolution	1
Padding of convolution	Same

4.3 Evaluation Metrics

The performance of copy-move forgery detection is evaluated by precision, recall, F_1 scores (PRF) and ROC curve. The precision, equation (1), depends on the variables: TP and FP while the recall, equation (2), depends on the variables: TP and FN. The F_1 scores, equation (3), combines precision and recall for better evaluation results.

$$Presicion = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Presicion + Recall} \tag{7}$$

TP is the pixel number of the forgery regions that are correctly detected. FP is the pixel number of the authentic region that are mistakenly detected as forged regions. FN is the pixel number of the forgery regions that are undetected. We use two protocols for pixel-level evaluation: Micro Average and Macro Average. Micro Average is to evaluate the overall performance including non-forged images through calculating the overall PRF on the entire data set,, while Macro Average is only applicable to a subset of forged images through calculating the average PRF

for each sample, but better quantifies the localization performance.

4.4 Results and Analysis

1) Evluation with different numbers of training images on USCISI dataset: In this section, since the compared methods [6] and [17] are key-point based method, we only evaluate the efficiency of PCNet with benchmark end-to-end solution BusterNet[18] on different numbers of training images which are choosen fromUSCISI dataset. Two methods are tested on CASIA CMFD dataset, and both evaluated with Micro Average pixel-level evaluation. The results are shown in Table2. It can be seen that PCNet is as good as BusterNet when trained with a great number of training samples(100k images). Compared with proposed PCNet, the BusterNet perform poorly when trained with a small number of training samples(10k images). One possible explanation is that the PCNet leverages hierarchical features by enriching the spatial information.

Table2. Performance comparison with BusterNet on CASIA CMFD dataset

Number of training samples	Index	BusterNet[18]	PCNet
10k	Precision	16.77	53.14
	Recall	9.51	41.53
	F_1	12.14	**46.62**
50k	Precision	49.23	65.21
	Recall	27.74	52.19
	F_1	35.48	**57.98**
100k	Precision	79.32	73.56
	Recall	50.12	57.42
	F_1	61.43	**64.49**

2) Comparison with others on two benchmark datasets: The comparison results on two benchmark datasets are shown in Table3 and Table4. These experiments are trained on the USCISI dataset and evaluated on the CASIA CMFD dataset. It can be seen from Table3 thatPCNet outperforms others on CASIA CMFD dataset. On the one hand, PCNet obtains more detailed features by constructing a cascaded pairwise correlation feature thereby improving CMFD performance. On the other hand, more detailed feature representation also causes some authentic regions are mistakenly detected as forged regions, this is why the micro average Precision of PCNet is lower than that of BusterNet. PCNet's F_1 scores outperforms others' on all three evaluation protocols. Furthermore, by comparing the performance of PCNet and BusterNet, Fig. 3 shows that PCNet improves AUC by 4% and 7% at both image-level and pixel-level, respectively.

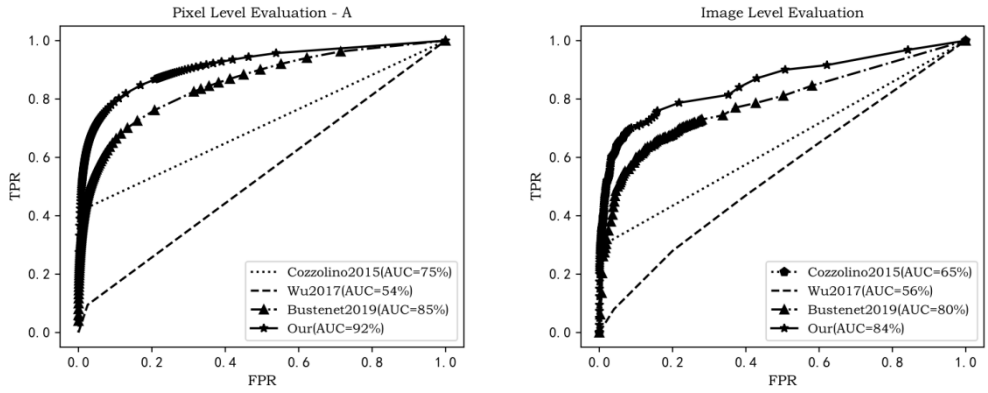

Fig. 3 AUC performance comparison on CASIA CMFD dataset

3) Impact of the duplicated region scale: To evaluate the impact of the scale of the duplicated region to the performance, we divide the CASIA CMFD dataset into three kinds of subset based on the proportion of duplicated region in a whole image: small duplicated region samples (proportion: 0-0.04), medium duplicated region samples (proportion: 0.04-0.13) and large duplicated region samples (proportion: 0.13-0.8), and make comparison between PCNet and BusterNet on three subsets. As shown in Table4, compared with BusterNet, enhanced CMFD performance is observed in PCNet on all three subsets, especially on the small and medium duplicated region samples, which illustrates PCNet not only achieves good CMFD performance under large-region tampering samples, but also outstanding under small-region or medium-region tampering samples. This is because PCNet obtains richer local features by using cascaded pairwise feature extraction, thereby capturing a set of features capable to detect the small copy-moved regions. It further validates our idea that cascaded pairwise feature can effectively improve CMFD performance.

4)

Table3 Performance comparison with others on CASIA CMFD dataset

Evaluation Protocol	Index	Cozzolino[6]	Wu[17]	BusterNet[18]	PCNet
Micro Average	Precision	83.12	17.06	79.32	73.56
	Recall	51.28	10.60	50.12	57.42
	F_1	63.43	13.08	61.43	**64.49**
Macro Average	Precision	24.92	23.97	47.93	62.78
	Recall	26.81	13.79	36.27	51.31
	F_1	25.43	14.64	37.86	**50.67**
Image Level Evaluation Protocol	Precision	99.51	66.37	78.22	79.31
	Recall	30.61	73.59	73.89	75.43
	F_1	46.82	69.80	75.98	**77.61**

Table4 Performance analysis of different copy-moved region scale on CASIA CMFD

Dataset Partition (number of images)	Macro Average PRF	BusterNet[18]	PCNet	Performance Improvement
Small duplicated region (474)	Precision	0.23	**0.45**	**0.96**
	Recall	0.17	**0.41**	**1.41**
	F_1	0.17	**0.37**	**1.18**
Medium duplicated region (521)	Precision	0.55	**0.68**	0.24
	Recall	0.40	**0.56**	0.4
	F_1	0.42	**0.57**	0.36
Large duplicated region (318)	Precision	0.73	**0.79**	0.08
	Recall	0.57	**0.59**	0.04
	F_1	0.61	**0.63**	0.03

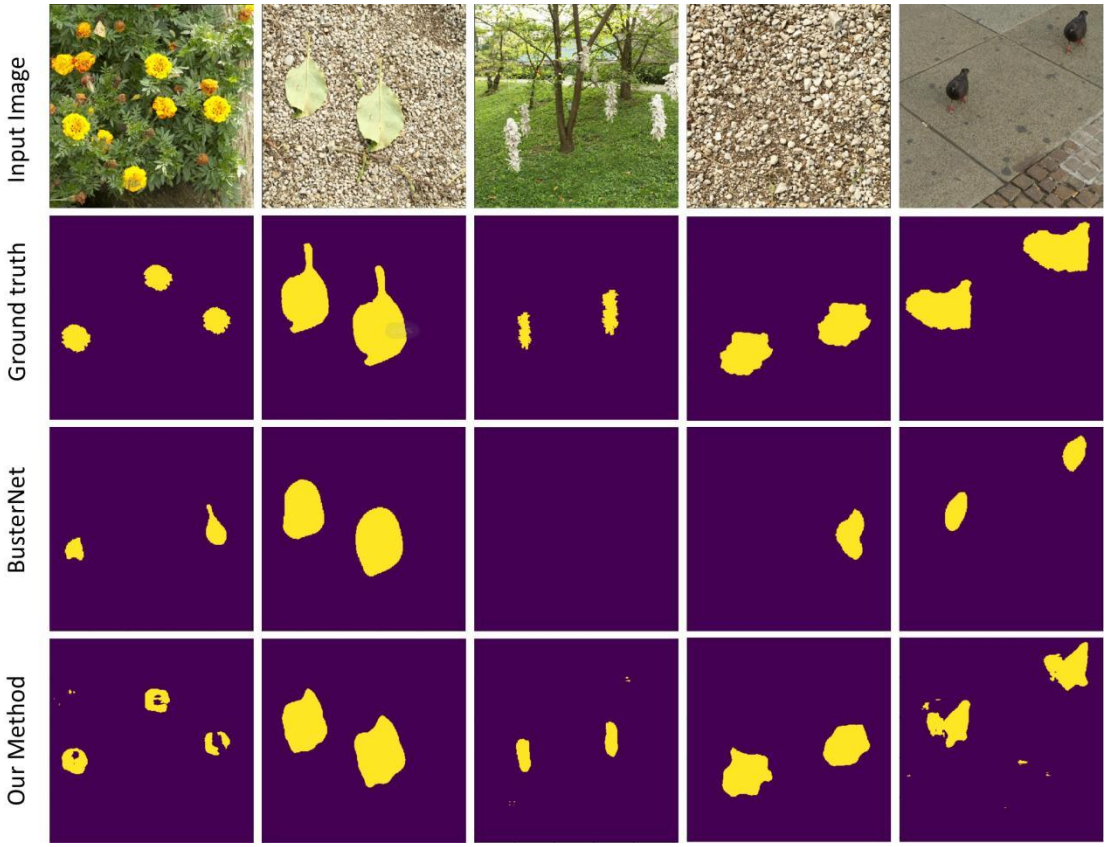

Fig. 4. Sample image copy-move forgery detection results on CoMoFoD dataset. Rows from top to bottom are: input image, ground truth, BusterNet[18], and our PCNet results.

4) Robustness against various transformations:In this section, the robustness of PCNet against various transformations is evaluated on the CoMoFoD dataset. According to the detection criteria defined in [18], Table5 shows the number of images with F_1 scores > 0.5 under each transformation. It can be clearly seen that PCNet outperforms others on all transformations. We also conduct performance analysis on the entire dataset. Fig. 5 shows the pixel-level F_1 scores with Macro Average protocal under different transformations. As can be seen from the chart, except for serious JPEG compression transformation, PCNet is robust against various transformations. Followed [18], the experimental results on the base category of CoMoFoD dataset is shown in Table6, and the proposed PCNet achieves better detection performance than others on the COMOFD dataset with no transformation. Compared with BusterNet, its Precision increased by 0.27%, Recall increased by 24.04%, F_1 scores increased by 11.58%, and the number of images with F_1 scores > 0.5 increased by 49.35%.

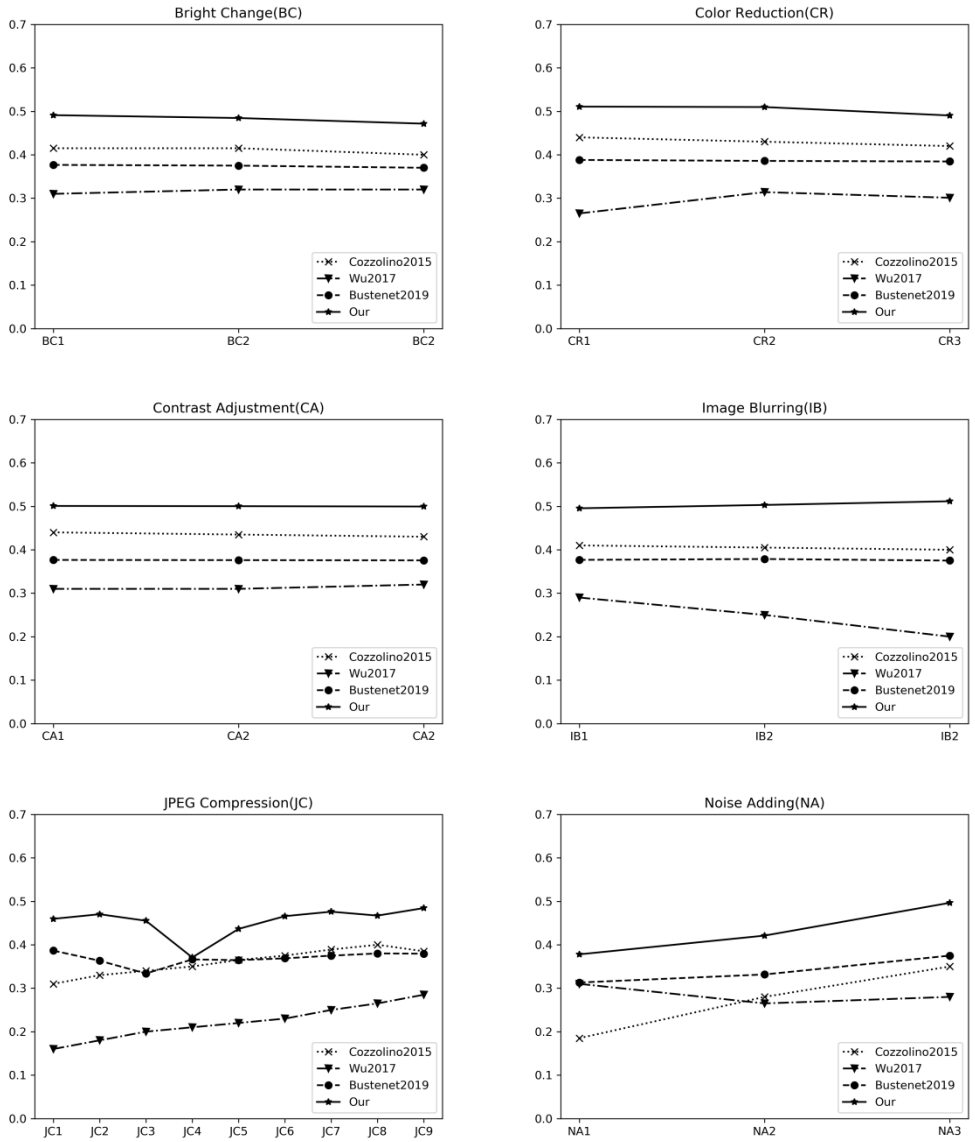

Fig. 5 Pixel-level F_1 scores (y-axis) on CoMoFoD under transformations (x-axis)

Table5 Comparison performance on CoMoFoD dataset

transformations	Cozzolino[6]	Wu[17]	BusterNet[18]	PCNet
Brightness change 1	94	53	77	**113**
Brightness change 2	94	50	76	**106**
Brightness change 3	88	53	77	**108**
Contrast adjustments 1	98	48	77	**114**
Contrast adjustments 2	96	50	77	**113**
Contrast adjustments 3	96	50	77	**112**
Color reduction 1	97	48	77	**116**
Color reduction 2	95	51	77	**120**
Color reduction 3	92	50	75	**122**
Image blurring 1	91	54	81	**96**

Image blurring 2	88	53	82	**99**
Image blurring 3	84	32	72	**99**
JPEG compression 1	69	18	69	**78**
JPEG compression 2	73	21	67	**93**
JPEG compression 3	75	26	76	**105**
JPEG compression 4	77	29	76	**109**
JPEG compression 5	81	38	75	**110**
JPEG compression 6	83	33	76	**113**
JPEG compression 7	87	42	78	**117**
JPEG compression 8	92	42	78	**116**
JPEG compression 9	87	36	75	**114**
Noise adding 1	41	38	71	**93**
Noise adding 2	66	39	74	**119**
Noise adding 3	68	41	78	**120**

Table6 CMFD performance comparisons on CoMoFoD dataset

Evaluation Protocol	Index	Cozzolino[6]	Wu[17]	BusterNet[18]	PCNet
Macro Average	Precision	39.92	36.29	50.35	50.62
	Recall	47.61	40.41	37.49	61.53
	F_1	41.83	31.13	37.82	**49.40**
F_1 scores > 0.5	Number	93	53	77	**115**
	Precision	84.22	61.11	76.37	77.46
	Recall	93.58	71.48	73.98	79.46
	F_1	**87.82**	63.13	73.08	78.45

V. Conclusion

In this paper, an end-to-end copy-move forgery detection method based on pyramid correlation network(PCNet). Several experiments are conducted to demonstrate that the efficiency, generalizability and robust to various transformations of PCNet. Thanks to the richer spatial information of the features and appropriate post-processing strategy, PCNet not only achieves good CMFD performance under large-region tampering samples, but also outstanding under small-region or medium-region tampering samples. Compared with BusterNet, F_1 scores of PCNet has increased by 33.84% and 30.62% on CASIA CMFD dataset and CoMoFoD dataset respectively. In future work, we will try to efficiently reduce the computing complexity.

Acknowledgements

This work was supported by National Natural Science Foundation of China (62072123), State Key Program of Guangdong College (2020ZDZX3059), Science and Technology Plan Project of Guangdong Province (2017A040405058), Natural Science Foundation of Guangdong Province (2018A0303130187).

Reference

[1] Li J, Li X, Yang B, et al. Segmentation-Based Image Copy-Move Forgery Detection Scheme. IEEE Transactions on Information Forensics & Security, 2017, 10(3):507-518.

[2] Pan X, Lyu S. Region duplication detection using image feature matching. IEEE Transactions on Information Forensics and Security, 2010, 5(4):857-867.

[3] Amerini I, Ballan L, Caldelli R, et al. A SIFT-Based Forensic Method for Copy–Move transformation Detection and Transformation Recovery. IEEE Transactions on Information Forensics & Security, 2011, 6(3):1099-1110.

[4] Wang C, Zhang Z, Li Q, et al. An Image Copy-Move Forgery Detection Method Based on SURF and PCET. IEEE Access, 2019, 7:170032-170047.

[5] Cozzolino D, Poggi G, Verdoliva L. Efficient Dense-Field Copy–Move Forgery Detection. IEEE Transactions on Information Forensics & Security, 2015, 10(11):2284-2297.

[6] Silva E, Carvalho T, Ferreira A, et al. Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes. Journal of Visual Communication & Image Representation, 2015, 29.

[7] A Jessica Fridrich, B David Soukal, A Jan Lukáš. Detection of copy-move forgery in digital images. proceedings of digital forensic research workshop, 2003.

[8] Mahdian B, Saic S. Detection of copy–move forgery using a method based on blur moment invariants. Forensic Science International, 2007, 171(2-3):180-189.

[9] Li G, Wu Q, Tu D, et al. A Sorted Neighborhood Approach for Detecting Duplicated Regions in Image Forgeries Based on DWT and SVD[C]// Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, ICME 2007, July 2-5, 2007, Beijing, China. IEEE, 2007.

[10] Ryu S J, Kirchner M, Lee M J, et al. Rotation Invariant Localization of Duplicated Image Regions Based on Zernike Moments. IEEE Transactions on Information Forensics & Security, 2013, 8(8):1355-1370.

[11] Li L, Li S, Zhu H, et al. An efficient scheme for detecting copy-move forged images by local binary patterns. Journal of Information Hiding & Multimedia Signal Processing, 2013, 4(1):46-56.

[12] Pun C M, Yuan X C, Bi X L. Image Forgery Detection Using Adaptive Oversegmentation and Feature Point Matching. Information Forensics and Security, IEEE Transactions on, 2015, 10(8):1705-1716.

[13] Rao Y, Ni J. A deep learning approach to detection of splicing and copy-move forgeries in images[C]// 2016 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2016.

[14] Barni M,Phan Q T, Tondi B. Copy Move Source-Target Disambiguation through Multi-Branch CNNs. 2019.

[15] Wu Y, Abd-Almageed W, Natarajan P. Image Copy-Move Forgery Detection via an End-to-End Deep Neural Network[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.

[16] Wu Y, Abd-Almageed W, Natarajan P, "BusterNet: Detecting copy-move image forgery with source/target localization," in Proc. of ECCV 2018, 2018, pp. 170–186.

[17] Wu Y, AbdalmageedW, Natarajan P. Deep Matching and Validation Network -- An End-to-End Solution to Constrained Image Splicing Localization and Detection. 2017.

[18] Wu Y, Abdalmageed W, Natarajan P. ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries With Anomalous Features[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.