Effective Splicing Localization Based on Image Local Statistics

P. N. R. L. Chandra Sekhar and T. N. Shankar

Abstract In the digital era, people freely share pictures with their loved ones and others using smartphones or social networking sites. The news industry and the court of law use the pictures as evidence for their investigation. Simultaneously, user-friendly photo editing tools make the validity of pictures on the internet are questionable to trust. Intense research work is going on in image forensics over the last two decades to bring out such a picture's trustworthiness. In this paper, an efficient statistical method based on Block Artificial Grids in double compressed JPEG images is proposed to identify areas attacked by image manipulation. In contrast to existing approaches, the proposed approach extracts the local characteristics from individual objects of the manipulated image instead of the entire image, and pair-wise dissimilarity is obtained between those objects and exploits the manipulated region, which has the highest variance among other objects. The experimental results reveal the proposed method's superiority over other current methods.

Keywords Splicing localization · Object segmentation · Block artificial grids · Cosine dissimilarity

1 Introduction

Nowadays, in digitization, people strongly connect with social networking sites and freely share their ideas, pictures, and comments. In present-day society, images are used extensively in many ways, such as evidence in the court of law, journalism, science, and forensics discovery [1]. The Government is also taking positive steps towards digitizing all fields to reach the public. Simultaneously, the rapid growth of technology in developing powerful image editing tools induced an interest to make

P. N. R. L. C. Sekhar (✉) · T. N. Shankar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India
e-mail: cpnrl@gitam.edu

T. N. Shankar
e-mail: tnshankar2004@kluniversity.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

H. Jahankhani et al. (eds.), *Cybersecurity, Privacy and Freedom Protection in the Connected World*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-68534-8_14

the images or videos manipulate with ease and cannot be traceable with human vision. Copy-move, splicing, re-sampling, cloning few manipulation attacks which are frequently to tamper digital images. If an image has undergone these attacks and uses for evidence, it significantly impacts the trustworthiness of such evidence [2]. It is a challenge to distinguish the original with manipulated and establish the integrity and authenticity of digital images [3].

Image Forensics, a branch of Multimedia Forensics, aims at developing powerful techniques and tools towards detecting manipulation attacks on images [4]. In traditional methods like watermarking, authentication is considered an active method where authentic code is embedded in the original image to verify authenticity. Whereas blind or passive methods do not require any external clue to assess the authenticity of the image. Many image tampering techniques work on the assumption that pictures taken from different cameras or different processing operations introduce different inherent patterns into tampered image [5]. Furthermore, these underlying patterns consistent throughout the original image, and when any manipulation attacks it, there will be inconsistency in those patterns of tampered image. These intrinsic inconsistency statistics can thus be used as forensic features to identify image tampering [6].

In image splicing, a part of the source image copied and pasted into the donor image. The post-processing techniques applied to the tampered image made human vision challenging to find such attacks [7]. This challenge attracted many researchers to find methods for image splicing detection. These techniques extract image features and use classification techniques to reveal the forgery, and achieve even high success rates [8]. However, it is worth locating the tampered region in many real-time purposes to gain confidence [9]. Splicing localization brings many more challenges as it requires pixel-level analysis rather than image-level analysis [4, 10].

The images captured by digital cameras are stored in Joint Photographic Experts Group (JPEG) format. The JPEG format uses lossy compression and responsible for the proliferation of images on the internet and social networking sites. In JPEG compression, the digital image divides into 8×8 non-overlapping blocks, and for every block, the discrete cosine transform (DCT) is evaluated and then quantized using a standard quantization matrix. When a splicing attack manipulates the image, it introduces discontinuities, and these statistical traces, such as JPEG quantization artefacts and JPEG grid alignment discontinuities, are used to exploit tampering attacks [11, 12].

The tampered region blocks will undergo single compression in a splicing attack while the remaining blocks will have double compression. For double compression (DQ) artefacts, a model of periodic DCT patterns is created in [13] and evaluated each block of the image concerning its conformance of the model. Any block whose probability distribution distinguishes from the original classifies as blocks manipulated by a tampering attack. A similar approach found in [14] where the authors assume that the distribution of JPEG coefficients changes with the number of recompressions and proposes a training a set of support vector machines (SVM) for the first digit artefacts and estimated the probability distribution of each block as single or double compressed thereby exposed the splicing attack.

In [15] proposed an alternative method to exploit DQ artefacts. They compare the discontinuities using the quality factor adopted in the tampered region with the principle that a JPEG ghosts—a local spatial minimum- will correspond to the tampering attack. The limitation of the method is; it works only if the tampered region has a lower quality factor than the rest of the image. An alternative to the DQ discontinuities, in [16], the authors created a model on the entire image DCT coefficient distributions using the degree of quantization. The inconsistencies become indicative of the tampering attack. The difference between this method and DCT-based is that the output is not probabilistic, making the technique relatively difficult to interpret although efficient.

Other techniques use JPEG grid discontinuities as they occur during compression by placing spliced objects that misaligned the 8×8 block grid. The 8×8 block creates the grid even when the image compress with high quality and these discontinuities are invisible to human eyes but can be exposed using filtering. The absence or misalignment of the 8×8 grid with the rest of the image can become a fingerprint to exploit a tampering attack in [17]. In [18], the authors extracted local features from the intensity of the blocking pattern. Any variations in those features indicate the block grid's absence or misalignment to detect splicing attacks. In continuation the authors of [12] expose tampering detection and localization by the probability distribution of its DCT coefficients. They used three features that can truly distinguish tampered regions from original ones and obtain accurate localization results. The drawback of their method is the refining of the probability map in post-processing, and it is a remediate strategy that influences localization results. To eliminate the drawback, [19] used a mixture model based on normalized grey level co-occurrence matrix (NGLCM) and obtained more accurate localization with the prior knowledge of both tampered and original regions. To get this, they used conditional probabilities of tampered regions and original regions of DCT blocks in first, second, and third-order statistics. Still, their method is time consuming and the rate of false alarm is high.

In this paper, we move towards proposing a forensic technique that can localize the tampered region from a single JPEG image with double compression. Unlike other techniques that produce probability maps from 8×8 DCT coefficients, we proposed an efficient statistical model that uses block artificial grids (BAG) [18] and expose localization of spliced object.

1.1 Our Contribution

Over the years, various splicing localization techniques have proposed that there is still scope for robustness and effectiveness, as splicing is complex in nature. In this regard, we are offering the following contributions to our proposed work. (i) Instead of extracting features on the image level, we segment the image into individual objects and obtain features from each object. (ii) For each object, we estimate the variance

of the BAG noise (iii) Instead of probability maps, we used pair-wise dissimilarity to classify the suspicious objects from original ones to expose tampered object.

The paper is organized as follows: Sect. 1 describes JPEG fingerprints from block artificial grids to speed up computation time. In Sect. 2 the proposed statistical method to expose splicing attack in JPEG image is described. The experimental and evaluation results present in Sect. 3, and finally, the paper concluded in Sect. 4.

2 Proposed Method

In this work, our primary goal is to localize the tampered region of the spliced JPEG image. The proposed work is framed into three levels: object-level image segmentation to extract individual objects in the spliced image, estimate the variance of each object using block artificial grids, and pair-wise dissimilarity among objects to localize tampered region as shown in Fig. 1.

2.1 Object Segmentation

Object detection is a challenging computer vision problem that solves object detection and classification. Among several object detection techniques, Mask R-CNN [20] is a widely used framework for object detection developed by Facebook research. It outperforms COCO suite challenge consisting of instance segmentation, bounding-box object detection, and person key point detection. It is a simple extension of Faster R-CNN with predicting the object's mask and easy to estimate human poses.

Using the Mask R-CNN framework, as shown in Fig. 2, we performed object detection and segmentation [21] for the given spliced image and extracted individual masks of all objects. Then for each mask, find its object from the input image along with the bounding box area. We split each object into foreground object consisting of the object mask region and the background object consisting of the remaining part in the bounding box from each bounding box object.

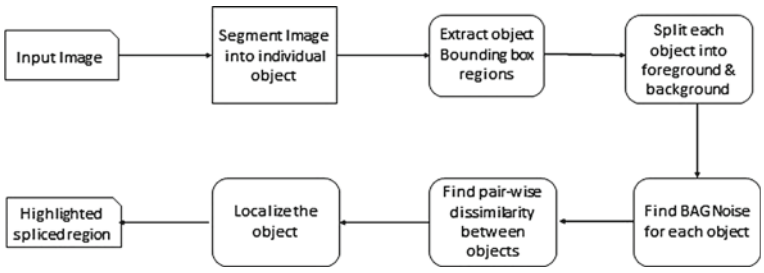

Fig. 1 The Proposed Frame Work

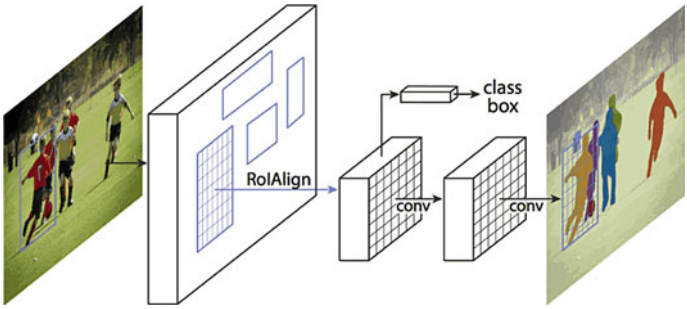

Fig. 2 Mask R-CNN Frame Work adopted from [20]

2.2 Block Artificial Grids

When the image is compressed with lossy JPEG, it leaves horizontal and vertical breaks in the image and is commonly refers as Block Artificial Grids (BAG). The BAGs of the entire image are roughly at the border an 8 × 8 block with a periodicity of 8 at both horizontal and vertical edges. When any attack alters the image, then the BAGs appear within the block instead of at borders. Thus this JPEG fingerprint is used in image forensics [12]. While compress the image using a digital camera, it introduces noise such as natural noise, BAG noise due to the JPEG compression factor. The artificial grid lines in an 8× 8 block are feeble than the border edges. In [18], the authors extracted weak horizontal and vertical lines with a periodicity of 8 separately to enhance these weak lines, and then combined them is referred to as BAGs.

In this paper, we focus on extracting BAGs in colour images. Since the luminance component in the JPEG standard is 8 × 8 block, we used only the luminance component rather than Cb and Cr of components of YCbCr image. The second-order difference of an image regards as weak horizontal edges of an image. For the given image $I(m, n)$, the absolute second-order difference $d(m, n)$ is obtained by

$$d(m,n) = |2I(m,n) - I(m+1,n) - I(m-1,n)| \tag{1}$$

To enhance the weak edges and remove the interference coming from strong image edges, a median filter is applied. To further reduce the edge influence as in [18] ignored differentials greater than an experimental threshold. Then the enlarged horizontal edges are accumulated for every two subsequent blocks as:

$$e(m,n) = \sum_{i=n-16}^{16} d(m,i) \tag{2}$$

Then to equalize the amplitudes throughout the resultant image, a local median reduces from each element.

$$e_r(m,n) = e(m,n) - median[\{e(i,n)|m-16 \leq i \leq m+16\}] \quad (3)$$

Thus, the weak horizontal edge image $w_h$is obtained by applying periodical median filter as:

$$w_h(m,n) = median[\{e_r(i,n)|i = m-16, m-8, m, m+8, m+16\}] \quad (4)$$

where $w_h(m, n)$ are elements of extracted horizontal BAG lines. The five elements in Eq. 4 with spacing 8 used in the median filter, makes the strong BAGs and weak BAGs smooth, and others can be removed. As more elements used in the median filter, BAGs can be extracted in a better way.

The vertical BAGs $w_v$are extracted similarly.

$$w_v(m,n) = median[\{e_r(m,i)|i = n-16, n-8, n, n+8, n+16\}] \quad (5)$$

The final BAG is obtained by combining Eqs. 4 and 5 as

$$w_b(m,n) = w_h(m,n) + w_v(m,n) \quad (6)$$

Equation 6 gives BAGs for the original image. When an image is attacked by tampering, the BAGs appear at some abnormal position such as centre of the block. So, for a fixed 8×8 block w_{mn} these abnormal BAGs can be obtained as [5]:

$$\begin{aligned} w_{mn} = & \text{Max}\{\sum_{i=2}^{7} w_b(i,n)|2 \leq n \leq 7\} \\ & - \text{Min}\{\sum_{i=2}^{7} w_b(i,n)|n = 1,8\} \\ & + \text{Max}\{\sum_{i=2}^{7} w_b(m,i)|2 \leq m \leq 7\} \\ & + Min\{\sum_{i=2}^{7} w_b(m,i)|m = 1,8\} \end{aligned} \quad (7)$$

2.3 *Localization of Splicing Region*

Using mark R-CNN object detection framework [21] as discussed in Sect. 2.1 the individual objects split into foreground object and background object-which is assumed to have similar characteristics of the whole image. Then for object extract the BAGs as discussed in Sect. 2.2. To expose discrepancies in BAGs of individual objects, we

find BAG noise as:

$$\mu = \frac{1}{R} \sum w_{mn}(i, j) \tag{8}$$

$$\sigma = \frac{1}{R} \sum (w_{mn}(i, j) - \mu)^2 \tag{9}$$

μ is mean, σ is variance, and R represents the no of BAG features in w_{mn}.

After that, we used pair-wise dissimilarity between foreground and background objects to detect the tampered object. For each pair of distinct objects, let the BAG noise is estimated be S_1 and S_2. Then the cosine dissimilarity between the objects defined as:

$$L_D = 1 - \frac{C(S_1, S_2) + 1.0}{2} \tag{10}$$

where

$$C(S_1, S_2) = \frac{{S_1}^T . S_2}{\|S_1\| . \|S_2\|.} \tag{11}$$

$C(S_1, S_2)$ is the cosine angle between two BAG noises. This metric $L_D$gives values in the range [0, 1]. Where the values near to 0 represent similar BAG noise levels of both objects, and near to 1 represents different levels.

From the dissimilarity matrix, we find the pair that gives maximum dissimilarity, and for each object in the pair, identify the object which has maximum dissimilarity with other objects and expose as tampered object.

In summary, the method described above includes three aspects. (i) We extract features from individual segments rather than whole image. (ii) For each individual object, extract the BAG noise. (iii) Using pair-wise dissimilarity between objects localize the tampered object which has maximum dissimilarity. As a result, localizing accuracy, as well as computational complexity are improved.

3 Experimental and Performance Analysis

This section evaluates the proposed method on two datasets and compares its performance with other recent technique.

Typically, CASIA dataset [22] is a widely used evaluation dataset for JPEG image splicing forgery detection, and it consists of 7491 authentic and 5123 spliced images with JPEG, TIFF, and BMP types of images. We used the Mask R-CNN framework for object segmentation, so we randomly selected 1000 tampered images of animals, persons, birds, vehicles with the size 384 × 256. The proposed method test on those chosen tampered images of the CASIA dataset for localizing spliced regions.

The qualitative evaluation of splicing images on the CASIA dataset shows in Fig. 3. The first row consisting of randomly chosen four images, and the ground truth masks are in the second row. The proposed method results are in the third row, except the spliced region masked as white. From the results, our method's superiority is very clearly evident to localize the spliced region. The advantage of object segmentation is clear evidence in our results.

To increase the robustness of the proposed method, we evaluated our approach on the Image Manipulation Dataset (IMD) [23]. The dataset contains 48 high resolution JPEG compressed images with size 3264 × 2448 and different quality factors ranging from 20 to 100%. The images were cropped to 2048 × 1536 to reduce the computational complexity and spliced each other and obtained 600 spliced images. The proposed method applies to those images.

The evaluation results on the customized IMD spliced dataset from [23] are shown in Fig. 4. The first row contains four sample images from the dataset. The ground truth masks are in the second row, and the proposed method results are in the third row. From the results, the proposed method works well on the high-resolution images.

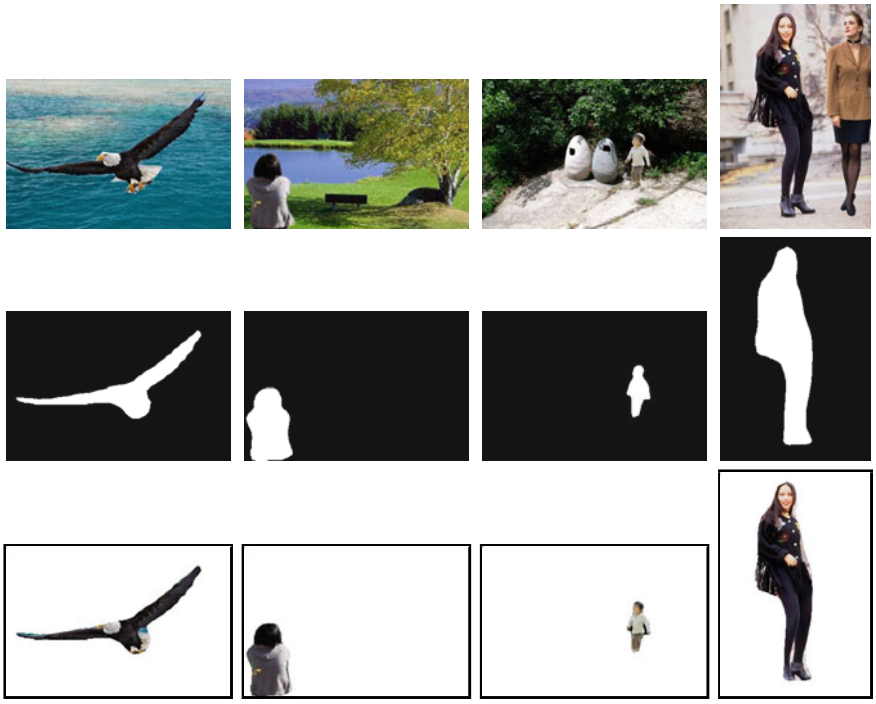

Fig. 3 Visual evaluation of proposed method on CASIA dataset

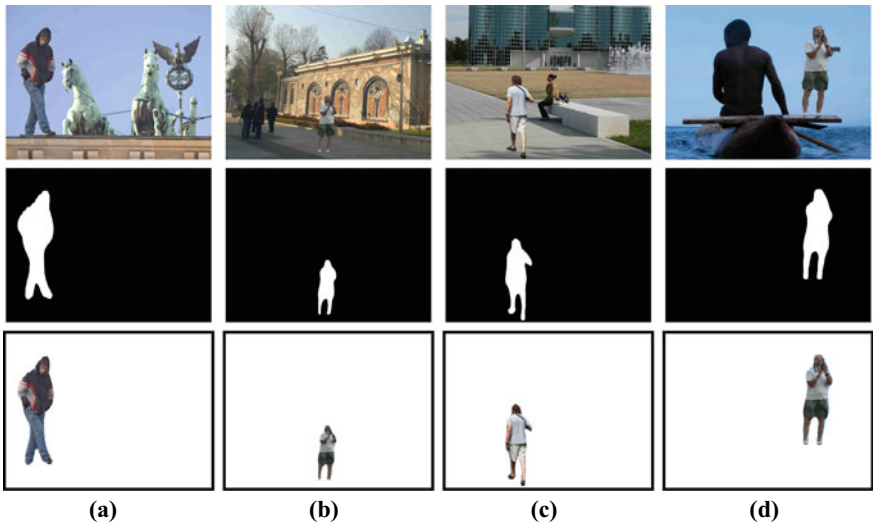

Fig. 4 Visual evaluation of proposed method on the high-resolution images from [23]

3.1 Localization Accuracy

The accuracy of splicing localization evaluates based on pixel-level F-measure. To evaluate, we used two metrics True Positive Rate (TPR), a measure of the rate of pixels that are truly detected as spliced, and False Positive Rate (FPR), a measure of the rate of pixels that are falsely detected as spliced.

$$TPR = \frac{TP}{TP + FN} * 100 \tag{12}$$

$$FPR = \frac{FP}{FP + N} * 100 \tag{13}$$

where TP is True Positive, FP is False Positive, TN is True Negative, and FN is False Negative. It expects to have high TPR and low FPR in the results. From these metrics, the F-measure defines as follows:

$$F = 2 * \frac{TPR * FPR}{TPR + FPR} \tag{14}$$

We evaluated average TPR and FPR and F-measure for all the selected images from the CASIA dataset and compared them with a recent method to analyse the proposed method. [19]. The method of [19] is based on a normalized grey level co-occurrence matrix on 8×8 DCT coefficients and using the Bayesian posterior

Table 1 Comparative results on CASIA and IMD datasets using average F-measure

Method	CASIA 2.0	IMD
NGLCM	0.6524	0.5572
Proposed	0.7852	0.0692

probability map, localized the tampering objects. To evaluate the superiority of the proposed method, we compared our results with recent methods.

Table 1 contains the Comparative results of the proposed method with [19] method on both datasets based on average F-measure. From the results, it is evident that BAG noise on individual objects in the proposed method enables us to have much superior performance than [19].

The method is robust when it has stable performance even after applying some post-processing operations on the spliced image. To evaluate the proposed method robustness, we applied JPEG compression with different quality factors, Gaussian blur, and added Gaussian noise to all the spliced images and tested.

For JPEG compression, 8 different quality factors ranging from 20 to 90 are considered. For Gaussian blur, Gaussian smoothing kernel with standard deviation $\sigma = 1.0$ is considered and for Gaussian noise, variance of 0.03 and 0.05 are considered.

The evaluation results on IM Dataset are shows in Table 2. As the quality factor (QF) in JPEG compression decreases and additional post-processing operations included, the NGLCM method decrease in its average F-measure values. In contrast, the proposed method has superior as well as stable performance even in such situations.

The IM dataset images are very high-resolution, and we try to downscale the quality factor to the lowest level 20. Figure 5 is a graph showing the performance of the proposed method with other existing method. NGLCM method decreases its average F-measure as the JPEG compression quality factory is reduced to 20. The proposed method outperforms and gives stable performance even when the quality factor reduces because the BAGs are affected only in those objects than the rest of the image.

Table 2 Comparative results for robustness on IM dataset using average F-measure

Method	(JPEG compression)		(Gaussian blur)	(Gaussian noise)	
	QF = 50	QF = 70	$\sigma = 1.0$	Variance = 0.03	Variance = 0.05
NGLCM	0.3934	0.4323	0.5412	0.5389	0.5395
Proposed	0.6418	0.6596	0.7520	0.7514	0.7520

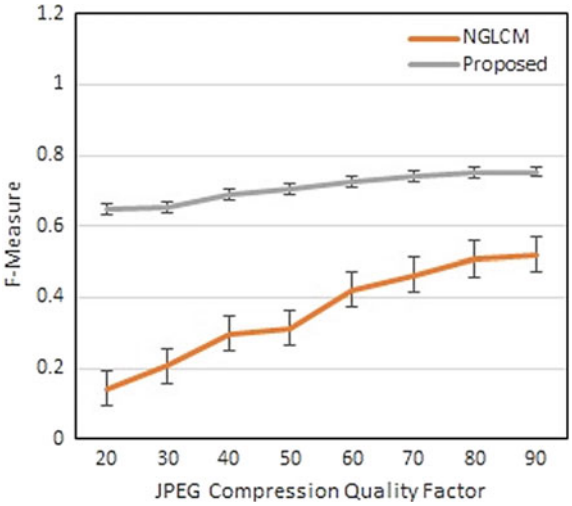

Fig. 5 Comparative results of JPEG quality factory with F-measure

3.2 Computational Complexity

The effectiveness of any method depends on its average computation time spent is minimal to get the desired result. In the proposed method, after segmenting the individual objects, we obtain BAG features from each object instead of the whole image by saving a lot of computation time. For localization, also we used a simple statistical method instead of unsupervised learning techniques. Table 3 gives the average running time spent by each method. Among the two methods, the proposed method takes less time than other existing methods.

Table 3 Average running time

Method	Proposed	NGLCM
(Average running time in secs)	16.8	78.9

4 Conclusion

This paper proposed an effective splicing localization method using local statistics of the image. When the JPEG image is splicing with another image's object, the block artificial grids in the tampered area move from 8×8 grid lines to its centre. Taking this clue as a feature descriptor, we exposed splicing forgery through object segmentation. The method is straightforward, effective than other conventional methods that use JPEG fingerprints. The proposed method also robust even when the quality factor is low in high-resolution JPEG compression. The method fails on low-resolution images, and we considered it as our future work.

References

1. Ali Qureshi M, Deriche M (2014) A review on copy move image forgery detection techniques. In: IEEE 11th international multi-conference on systems, pp 1–5
2. Redi JA, Taktak W, Dugelay J (2011) Digital image forensics: a booklet for beginners. Multimed Tools Appl 51:133–162
3. Farid H (2009) Image forgery detection a survey. IEEE Signal Process Mag 26(2):16–25
4. Birajdar GK, Mankar VH (2013) Digital image forgery detection using passive techniques. Digit Invest Int J Digit Forensics Incident Response 10(3):226–245
5. Liu B, Pun CM, Yuan XC (2014) Digital image forgery detection using JPEG features and local noise discrepancies. Sci World J 1–12
6. Chandra Sekhar PNRL, Shankar TN (2016) Review on image splicing forgery detection. Int J Comput Sci Inf Secur 14(11):471–475
7. Bahrami K, Kot AC, Li L (2015) Blurred image splicing localization by exposing blur type inconsistency. IEEE Trans Inf Forensics Secur 10(5):999–1009
8. Zhang Y, Zhao C, Pi Y, Li S (2012) Revealing Image splicing forgery using local binary patterns of DCT coefficients. In: Liang Q et al (eds) Communications, signal processing, and systems. Lecture notes in electrical engineering, vol 202, pp 181–189
9. Chandra Sekhar PNRL, Shankar TN (2019) Splicing localization based on noise level inconsistencies in residuals of color channel differences. IJRTE 8(3):764–769
10. He Z, Lu W, Sun W, Huang J (2012) Digital image splicing detection based on Markov features in DCT and DWT domain. IEEE Trans Pattern Recognit 45(12):4292–4299
11. Zampoglou M, Papadopoulos S, Kompatsiaris Y (2017) Large-scale evaluation of splicing localization algorithms for web images. Multimed Tools Appl 76(4):4801–4834
12. Bianchi T, Piva A (2012) Image forgery localization via block-grained analysis of JPEG artifacts. IEEE Trans Inf Forensics Secur 7(3):1003–1017
13. Lin Z, He J, Tang X, Tang CK (2009) Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. Pattern Recognit 2492–2501
14. Amerini I, Becarelli R, Caldelli R, Del Mastio A (2014) Splicing forgeries localization through the use of first digit features. In: IEEE international workshop on information forensics and security (WIFS), pp 143–148
15. Farid H (2009) Exposing digital forgeries from JPEG ghosts. IEEE Trans Inf Forensics Secur 4(1):154–160
16. Bianchi T, De Rosa A, Piva A (2011) Improved DCT coefficient analysis for forgery localization in JPEG images. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2444–2447
17. Luo W, Qu Z, Huang J, Qiu G (2007) A novel method for detecting cropped and recompressed image block. In: International conference on acoustics speech and signal processing, vol 2, pp 117–220

18. Li W, Yuan Y, Yu N (2009) Passive detection of doctored JPEG image via block artifact grid extraction. Signal Process 89:1821–1829
19. Xue F, Wei Lu, Ye Z, Liu H (2019) JPEG image tampering localization based on normalized gray level co-occurrence matrix. Multimed Tools Appl 78:9895–9918
20. He K, Gkioxari G, Dollar P, Girshick R (2017) 'Mask R-CNN'. In: EEE international conference on computer vision (ICCV), pp 2980–2988
21. Abdulla W (2017) Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask RCNN
22. Dong J, Wang W, Tan T (2013) CASIA image tampering detection evaluation database. In: IEEE China summit and international conference on signal and information processing, Beijing, pp 422–426
23. Christlein V, Riess C, Jordan J, Riess C, Angelopoulou E (2012) An evaluation of popular copy move forgery detection approaches. IEEE Trans Inf Forensics Secur 7(6):1841–1854