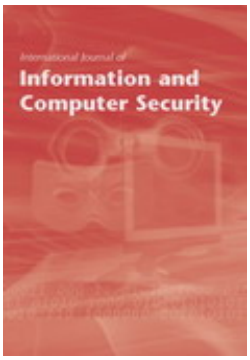


International Journal of Information and Computer Security

ISSN online: 1744-1773 - ISSN print: 1744-1765
https://www.inderscience.com/ijics

Image tampering detection based on feature consistency attention

Junlin Gu, Yihan Xu, Juan Sun, Weiwei Liu

DOI: 10.1504/IJICS.2023.10053595

Article History:

Received:	05 September 2022
Last revised:	22 November 2022
Accepted:	22 December 2022
Published online:	19 February 2024

Image tampering detection based on feature consistency attention

Junlin Gu*, Yihan Xu, Juan Sun and Weiwei Liu

College of Computer,
Jiangsu Vocational College of Electronics and Information,
Huai'an 223003, China
Email: 21968118@qq.com
Email: 36266709@qq.com
Email: 327652626@qq.com
Email: 371927989@qq.com
*Corresponding author

Abstract: Recently, the development of computer image technology makes image tampering more and more convenient, causing a large number of image tampering accidents. The existing scheme uses manual features and depth features to analyse the forgery traces, and has achieved good results. However, the existing schemes lack the analysis of essential traces and have defects in generalisation performance. In this paper, an image tampering detection scheme based on feature consistency attention is proposed. The inconsistency between the real region and the background region is used to improve the detection ability of the algorithm for unknown images. The scheme uses the feature extraction module to extract the deep semantic features of the image, and then calculates the feature correlation between the tampered region and the background region to maximise the correlation within the region and minimise the correlation between the background region and the tampered region. The scheme can learn the common traces of tampering process, which is expected to achieve better generalisation effect. Experimental results show that the proposed scheme is superior to several existing schemes in detecting tampered images.

Keywords: image tampering detection; deep neural network; feature consistency.

Reference to this paper should be made as follows: Gu, J., Xu, Y., Sun, J. and Liu, W. (2024) 'Image tampering detection based on feature consistency attention', *Int. J. Information and Computer Security*, Vol. 23, No. 1, pp.1–15.

Biographical notes: Junlin Gu received his Master's degree from the Guilin University of Technology. He is currently working at the Jiangsu Vocational College of Electronics and Information. His research interests include data analysis.

Yihan Xu is currently working at the Jiangsu Vocational College of Electronics and Information. His research interests include data analysis and programming.

Juan Sun is currently working at the Jiangsu Vocational College of Electronics and Information. His research interests include data analysis and digital forensics.

Weiwei Liu is currently working at the Jiangsu Vocational College of Electronics and Information. His research interests include data analysis and programming.

1 Introduction

The development of digital image processing technology has greatly reduced the threshold of digital image editing. Even users who do not have professional image processing knowledge can use various image processing tools to modify the content of digital images at will (Trump's Phone Call to a January 6 Witness Could Be a Federal Crime, 2022). True, this brings considerable convenience and fun, but inevitably leads to some security problems. The attacker deliberately tampered with the image content, and uploaded the tampered image to the network and social media, trying to spread false messages to disturb or manipulate public opinion, resulting in a very bad impact. In order to deal with the security risks caused by tampering with images, it is urgent to study the technology that can identify the authenticity of digital images.

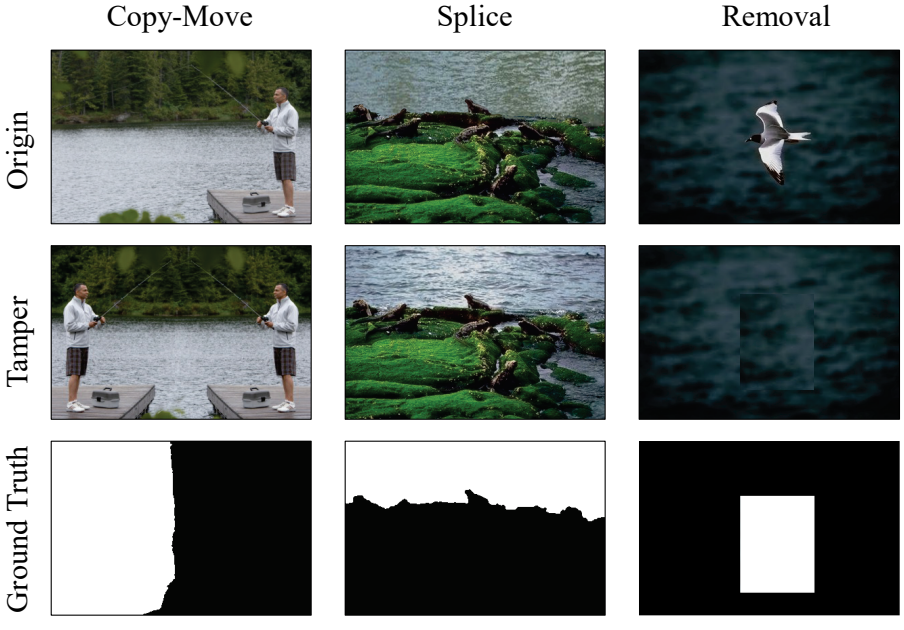
In recent years, the problem of image tampering localisation has attracted more and more attention of researchers, and new image tampering localisation methods have emerged. With the remarkable excellent performance of deep learning in computer vision and other applications, various deep learning models are introduced into image forensics. The existing schemes based on deep learning have been able to achieve tamper classification and tamper localisation tasks. However, there is no targeted design for tamper detection tasks. The detection ability of weak tampering is weak, and the effect of cross-dataset detection is not satisfactory. As shown in Figure 1, the existing schemes can show good results on the known dataset, but it is difficult to accurately detect tampering on the unknown dataset. How to achieve accurate and practical tamper detection has become an urgent problem to be solved.

In this paper, we aim to design a high generalisation tamper detection scheme. There are many inconsistencies between the image tampered region and the background region in texture continuity and noise mode. We hope to learn such inconsistency to assist the forgery localisation task and achieve accurate detection. The contributions of this paper are summarised as follows:

- We proposed a tampering detection scheme based on feature consistency analysis, which pays more attention to the inconsistency between tampered region and background region. The scheme uses the feature extraction module to extract the deep semantic features of the image, and then calculates the feature correlation between the tampered region and the background region to maximise the correlation within the region and minimise the correlation between the background region and the tampered region. The scheme can learn the common traces of tampering process, which is expected to achieve better generalisation effect.

- A tampering detection method based on multi-task learning is proposed. The forgery localisation module is introduced to segment the tampered region and provide guidance for the forgery classification tasks. At the same time, the classification module could perform forgery detection but lacks explanation for its results. The localisation module could provide enough explanation in the practical scenes.
- Experimental results show that the proposed scheme is superior to several existing schemes in detecting tampered images. Both intra-dataset and cross-dataset experiments demonstrate our effectiveness in detecting tampered images.

Figure 1 Images generated by different tampering methods (see online version for colours)



Note: There are three main tampering methods, including copy-move, splice and removal. The three lines are original image, tampered image and ground truth, respectively.

The rest of this paper is arranged as follows. In Section 2, we introduce the research on tampering detection based on deep learning. The proposed image tampering detection scheme based on feature consistency attention is described in Section 3. In Section 4, we analyse the experimental results to prove the effectiveness of the scheme. Finally, we summarise the proposed scheme in Section 5.

2 Related work

In this section, we present the research on image tampering, tampering detection, and feature consistency.

2.1 Image tampering

Digital image content change tampering is one of the most important ways of image tampering, which can arbitrarily reshape the image content and significantly change the image semantic information. There are various tampering methods for image content changes, such as splicing, copy-move, and removal.

As shown in Figure 1, splicing tampering refers to the purpose of hiding an important target by copying a region or target from another image and pasting it to the image (Nirkin et al., 2019; Li et al., 2020). Copy-move tampering is to copy any region or target in the image and paste it to other locations in the same image. The copied region is processed through scaling, rotation, translation and other operations. Since the copy part comes from the same image, the image texture attributes (such as colour, noise and texture) have no obvious visible changes, bringing more difficulties for the tampering detection task (Christlein et al., 2012; Tralic et al., 2013). Removal tampering is to delete an area or target in the image for some purpose, and then filling it with the surrounding area (Criminisi et al., 2004). Image embellishment is a common image repair operation, which is regarded as an acceptable method for tampering with images. It does not lead to significant changes in images, but emphasises (or reduces) some ideal (or unsatisfactory) features of images, which are commonly used in magazine photos and film production.

2.2 Tampering detection

Image tamper localisation is to identify and mark the tampered region in the image. In recent years, deep learning technology has been gradually introduced into the field of image tampering localisation. Inspired by the manual rich model feature proposed by Fridrich and Kodovsky (2012) for steganalysis, Rao and Ni (2016) initialised the first layer of CNN network through all the fixed filters involved in the rich model, and used CNN features and SVM for tamper localisation. Bayar and Stamm (2018) developed a new form of convolution layer, and used the constructed CNN architecture to detect three types of tampering operations: image splicing, copy-move and removal. Bappy et al. (2017a) proposed two methods based on resampling features and deep learning to detect and locate image operations. They used Gaussian conditional random field model and long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) network to optimise the location, and finally determined the tampered area. All the methods described above allow small image blocks to be detected and tampered regions can be located by sliding window analysis of the entire image.

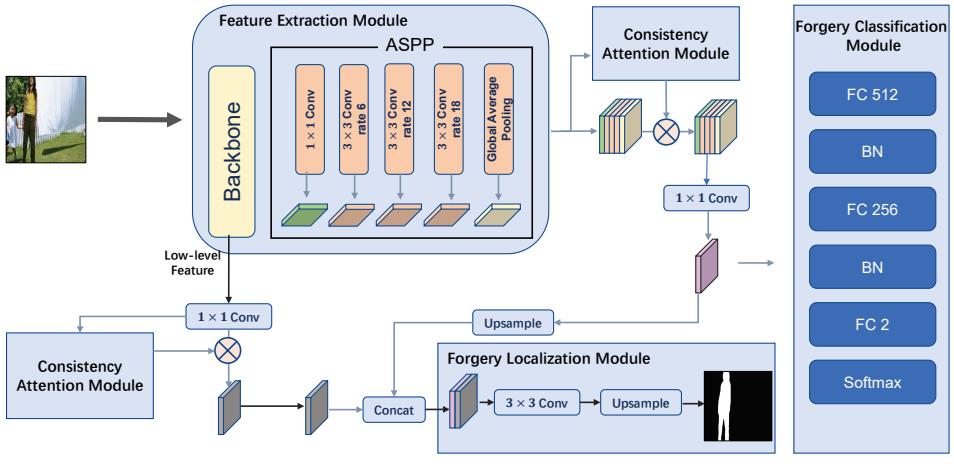
These solutions are mainly dependent on low-level features, provided that high-level features do not help detect possible operations. However, imperfect image editing (such as poor stitching quality images) may leave traces, such as unnatural tampering with boundaries under strong contrasts. Therefore, Li and Huang (2019) used a fixed high-pass filter at the first layer of the proposed depth network, while Wu et al. (2018) combined it with the standard filter. However, these two papers proposed a fully connected network (FCN) to obtain the binary positioning map as the output, and also considered the image-level detection. Bappy et al. (2017b) proposed another general architecture for tamper localisation with high reliability, which uses resampling feature, LSTM unit and encoder-decoder network to separate tampered regions from non-tampered regions. Due to the laborious image labelling process, the current technical

level of detecting these tampered images suffers from the lack of training data. To solve this problem, Zhou et al. (2020) proposed a tampering image generation process, which uses the currently available dataset to create real sample results.

2.3 Feature consistency

The inconsistency in tampered images has been studied in recent works (Mayer and Stamm, 2020). The inconsistency measurement is mainly performed by calculating the similarity score between image blocks (Mayer and Stamm, 2019; Bondi et al., 2016; Huh et al., 2018). Zhou et al. (2017) proposed the double-stream network based on faster RCNN, attempting to find symptoms from tampered faces and low-level instability. However, the training process requires steganalysis feature traction. Nirkin et al. (2021) used the proposed facial recognition and context recognition network for detecting deep forgery. At present, most deep learning frameworks rely on training data. It is difficult to effectively detect abnormality when the images are never encountered by the models. Self-learning and self-consistency detection are effective methods to solve this problem. Through the internal self-consistency judgement of the image, the abnormal regions in the image can be detected. This method can solve the problem of insufficient tampered image dataset, and conduct consistency training with a large number of different source image blocks, so that the model can learn the differences between images.

Figure 2 The overview of the proposed scheme which is composed of five modules: feature extraction module, hash generation module, consistency constraint module, tampering detection and classification modules (see online version for colours)



3 The proposed scheme

In this section, we elaborate on the motivation and details of the proposed scheme.

3.1 Design concept

The inconsistency between the tampered region and the background region is a common phenomenon in tampered images. This scheme constrains the consistency of the two at the feature level, and reduces the feature correlation between the tampered region and the background region, so as to carry out the tampering detection task.

The process of this scheme is shown in Figure 2. The scheme is mainly composed of feature extraction module, consistent attention module, forgery location module and forgery classification module. Among them, the feature extraction module extracts rich semantic features from the image. The consistency analysis module analyses the extracted features and selects important features related to consistency. The forgery location module is used to segment the forgery area. The forgery classification module is used to identify the authenticity of the image.

3.2 Feature extraction module

The feature module is used for tamper semantic feature extraction. Feature extraction is the most important part of the tamper detection task, and good features can provide direct help for tamper classification. Firstly, the pre-trained model on ImageNet dataset is used as the basic extractor for feature extraction to extract effective image semantic features. Then, the feature pyramid model is used to extract the rich features of multi-scale and multi-level. The feature pyramid consists of five small modules, including 1×1 convolution, 3×3 convolution (sampling rate is 8), 3×3 convolution (sampling rate is 12), 3×3 convolution (sampling rate is 18) and global average pooling layer. Each module extracts a feature map and forms the final feature map after splicing. After the feature pyramid model, the features extracted in the previous step can be effectively extended to provide more effective and targeted information for subsequent modules.

3.3 Consistency attention module

The consistency attention module is used to extract inconsistent features from the background and tampered regions. Because the image fusion operation must be involved in the tampering process, there will be inherent inconsistency between the tampered region and the background region. This scheme uses consistency attention module to analyse and extract inconsistency. Firstly, the multiscale features extracted from the feature extraction module are input into the convolution layer to calculate the attention weight, focusing on the important features related to inconsistency. It should be noted that the generated attention weight has the same dimension as the original feature. Then the attention weight is multiplied by the original feature matrix to obtain the weighted feature matrix. This feature matrix is the feature matrix given by attention.

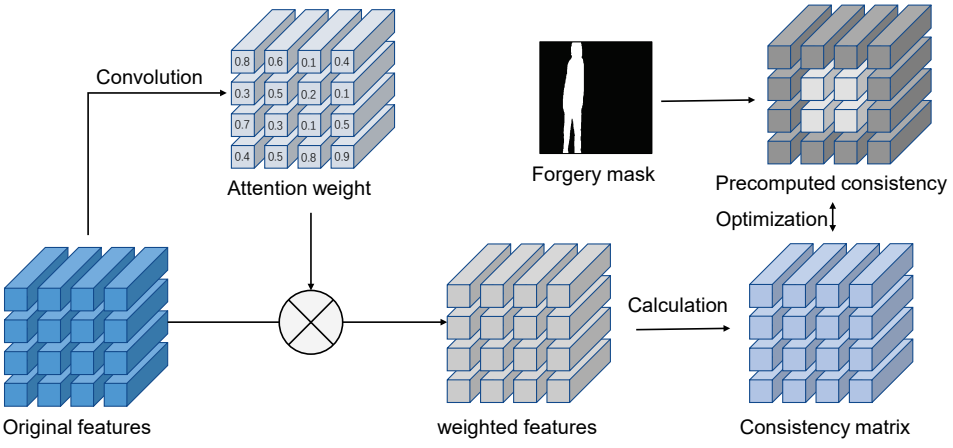
The initial generated attention-weighted features are not directional, so it is necessary to continuously optimise the modules to learn consistency-related features. The correlation between tampered region and tampered region should be strong, while the correlation between tampered region and background region should be weak. In this

scheme, we calculate the consistency metric matrix by Pearson correlation coefficient calculation. The formula is as follows,

$$\rho_{xy} = \frac{\sum \frac{x_i - \mu_x}{\sigma_x} \frac{y_i - \mu_y}{\sigma_y}}{n}, \tag{1}$$

where x denotes the source vector, the y denotes the target vector, μ_x represents the average value of x, σ_x represents the standard deviation of x, and the n is the value number of vector.

Figure 3 The optimisation process of consistency attention module (see online version for colours)



Note: The attention weight is applied to find important features for consistency analysis.

For the forgery mask with H $*$ W size, the feature correlation between each pixel and other pixels is calculated. On this basis, we obtain the precomputed consistency matrix. After that, we use the weighted features to calculate the feature consistency matrix. For the H $*$ W $*$ C matrix, the correlation between the eigenvectors corresponding to each point on the H $*$ W surface is calculated, and the consistency matrix is obtained. By optimising the similarity between these two matrices, we make the attention module pay more and more attention to consistency-related features. The loss function is shown below:

$$L_{ca} = BCELoss(T', T), \tag{2}$$

where T' is calculated correlation matrices and the T is the pre-computed correlation matrices.

3.4 Forgery localisation module

The forgery localisation task can accurately style the forgery area, thus providing guidance for forgery classification. In this scheme, the underlying texture features

extracted from the feature extraction module and the high-level semantic features after attention assignment are combined to perform the forgery and localisation task. The attention-given feature matrix is expanded by the size of the upper sampling layer, and then superimposed with the underlying texture features for forgery localisation. Through multiple convolutions, the final positioning result is obtained. The optimisation of this process is realised by the combination of DiceLoss and BCELoss. Previous work has shown that the combination of the two can achieve more accurate positioning, thus providing clearer guidance for forgery classification tasks. The loss function is as follows.

$$L_{loc} = 1 - \frac{2|P \cap \hat{P}|}{|P| + |\hat{P}|}, \tag{3}$$

where $\hat{P}$ denotes the predicted tampering mask and P represents the .

3.5 Forgery classification modules

The forgery classification module is used to identify the authenticity of the image. Features weighted by attention are used for forgery classification, and multi-layer full connection layer is used to analyse the difference of feature distribution between true and false images, so as to realise accurate forgery classification. We use BCE loss function to optimise the module. The loss function is as follows.

$$L_{cls} = -(c \times \log(\hat{c}) + (1 - c) \times \log(1 - \hat{c})). \tag{4}$$

where $\hat{c}$ is the predicted category and the c is the corresponding category.

Finally, the total loss is composed of three parts: classification loss, localisation loss and consistency attention loss. The total loss is as follows,

$$Loss = L_{cls} + \lambda_1 L_{loc} + \lambda_2 L_{ca}, \tag{5}$$

where λ_1, λ_2 are balanced parameters.

4 Experiments

In this section, we verify the effectiveness of the scheme. Firstly, the dataset and experimental settings are introduced, and then the detection performance of the scheme is verified. Finally, the ablation experiments are carried out to verify the effectiveness of each module.

4.1 Experimental settings

4.1.1 Dataset setting

To verify the effectiveness of the scheme, we used four datasets, including CASIA V2 (Dong et al., 2013), Realistic Tampering Dataset (RTD) (Korus and Huang, 2016), COVERAGE (Wen et al., 2016) and Faceswap dataset (Rössler et al., 2019). Among

them, the CASIA V2 dataset contains 5,123 tampered images and corresponding original images, and the tampered images contain 3,295 copy-move images and 1,828 spliced images.

The Faceswap dataset is composed of 2,000 face tampering images, swapping another face to the current image. RTD is a manually tampered dataset containing 220 1,920×1,080 uncompressed images. COVERAGE is a copy-move tampering dataset. 100 pairs of tampered images and original images. In the experiment, the CASIA V2 dataset is divided into training set, verification machine and test set according to 8:1:1, and then the model is trained and tested, and then the generalisation experiment is carried out in several other datasets.

4.1.2 Experimental details

In the experiment, we used MIOU and MPA indicators to evaluate the performance of the forgery positioning task. Among them, MIOU is used to characterise the average ratio of intersection and union of true and predicted values of each category, which can describe the prediction accuracy. MPA directly calculates the average prediction accuracy of each pixel. For forgery classification tasks, we directly use AUC and ACC for performance evaluation. AUC represents the area under the ROC curve, the closer to 1, the better performance. ACC represents the probability of accurate prediction. We use Adam optimiser to optimise the network training, the learning rate is set to 1e-3, decay set to 1e-4. The weight of loss function is set to 1.

4.2 Evaluation of detection performance

In this section, we first evaluate the detection performance of the scheme on the known dataset, and then verify the generalisation performance of the scheme on the unknown dataset.

4.2.1 Intra-dataset evaluation

After the proposed scheme is trained on the training dataset, we evaluate the effect of the trained model in the test dataset. We first perform evaluation on the non-noise dataset. As shown in Table 1, our scheme obtains the best detection performance on the CASIAV2 and FaceSwap dataset. It can be seen that the simple network-based methods fails to detect the manipulation in the images. They obtained high AUC values but low ACC values. This can be attributed to that some images in the CASIA V2 dataset are manipulated in a small degree, which increase the difficulty of accurate detection. Such schemes tend to classify all images as original images, and thus get high precision when classifying original images. Our scheme employs the consistency attention module to find the feature inconsistency, so as to explore the inconsistency in the forged images. The experimental results on two datasets show that our scheme could classify both original images and forged images correctly.

To prove the robustness of our scheme, we also perform the evaluation on post-processing datasets. A variety of post-processing operations (Gaussian noise, JPEG compression, blur, etc.) are involved to process images in the test dataset of CASIA V2. The experimental results are shown in Table 2. For images without post-processing, our

scheme outperforms several existing schemes. After image post-processing, our scheme maintains higher performance when comparing with the existing schemes. It can be seen that our scheme has a good performance in the detection of known data.

Table 1 The intra-dataset detection performance of our scheme

Methods	*CASIA V2*		*FaceSwap*	
	AUC	*ACC*	*AUC*	*ACC*
S-Resnet50 (He et al., 2016)	94.99	85.29	99.36	98.66
S-Resnet101 (He et al., 2016)	95.38	85.94	99.79	98.82
S-Densenet121 (Huang et al., 2017)	95.00	86.13	99.75	98.54
S-Inception_v3 (Szegedy et al., 2016)	92.03	81.71	98.83	96.57
Zhang et al. (2016)	96.94	91.09	98.69	96.74
Ours	**98.86**	**95.57**	**99.57**	**98.95**

Table 2 The detection performance when the images are processed by post-processing operations

Methods	*Post-processing operations*			
	Gauss noise	*JPEG compression*	*Blur*	*Resize*
S-Resnet101 (He et al., 2016)	94.65	95.12	94.98	93.23
S-Densenet121 (Huang et al., 2017)	95.34	95.86	94.13	93.71
Zhang et al. (2016)	96.98	95.84	93.5	93.64
Ours	**98.03**	**98.92**	**97.23**	**97.78**

4.2.2 Cross-dataset evaluation

This section evaluates the generalisation performance of the scheme. After training on CASIAv2 dataset, the model is tested on other unknown datasets. As shown in Table 3, we perform the generalisation evaluation on the Coverage and RTD datasets. These two datasets are all made manually, and have realistic visual performance. The accuracy of existing schemes is around 50%, which can be regarded as random guess. Although our performance is not good as the intra-dataset detection, our AUC and ACC values are higher than existing schemes.

Table 3 The cross-dataset detection performance of our scheme

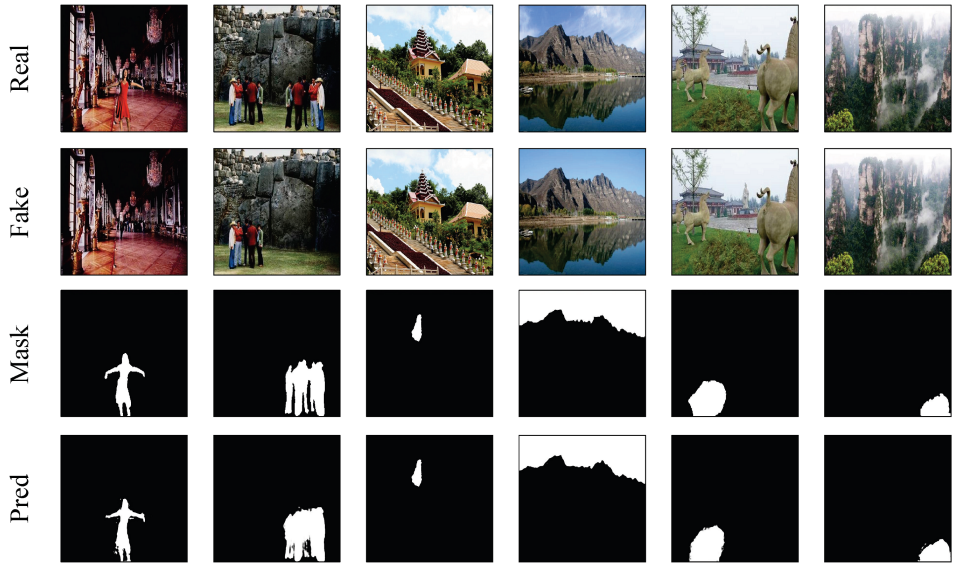
Methods	*Coverage*		*RTD*	
	AUC	*ACC*	*AUC*	*ACC*
S-Resnet50 (He et al., 2016)	52.62	52.68	57.02	54.24
S-Resnet101 (He et al., 2016)	51.90	49.11	54.77	53.87
S-Densenet121 (Huang et al., 2017)	51.80	51.79	56.11	53.12
S-Inception_v3 (Szegedy et al., 2016)	51.05	50.89	56.07	54.39
Zhang et al. (2016)	52.05	50.12	57.64	54.92
Ours	**61.97**	**57.25**	**59.65**	**56.54**

Note: The S-* means network with classification abilities. The model is trained on the CASIAV2 dataset and evaluated on the Coverage and RTD.

4.2.3 The performance of forgery localisation

In the scheme, we use forgery localisation task to assist the training of detection model. This section evaluates the effect of forgery localisation module. As shown in Figure 4, the scheme can accurately detect the image and accurately locate the forgery area. The pixel accuracy of our scheme is 99.45% and the MIOU is 0.9288. The localisation task could provide enough guidance for classification tasks.

Figure 4 The localisation performance on CASIA V2 dataset (see online version for colours)



Note: The original images, corresponding tampered images, tampering masks and the predicted outputs are shown on each row, respectively.

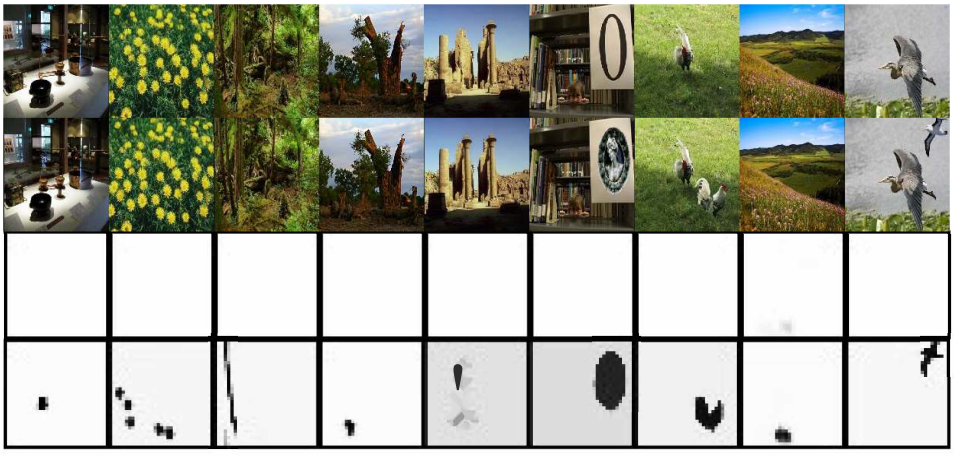
4.2.4 Consistency attention analysis

This section evaluates the performance of the consistency attention module. After the model is trained on the CASIAV2 dataset, the consistency attention module can learn the consistency-related features. We calculate the average value of consistency matrix to demonstrate the distinctiveness between original consistency and tampered consistency. As shown in Figure 5, the consistency attention module could distinguish the difference between the original images and the tampered images effectively.

4.3 Ablation experiment

This section evaluates the effectiveness of each module designed in the scheme, including tamper localisation module and consistency attention module.

Figure 5 The consistency matrix visualisation on CASIA V2 dataset (see online version for colours)



Note: The original images, corresponding tampered images, consistency matrices of original images and consistency matrices of tampered images are shown on each row, respectively.

4.3.1 *Tampering localisation module*

This module is used to assist tamper classification tasks. In the experiment, we removed this module and carried out related experiments on the CASIA V2 dataset. As Table 4 shows, the performance of the detection model on the dataset increases with the increase of the weight of the localisation loss. It can be seen that when the weight of tamper localisation loss is increased to 5, the detection model has obtained good performance. After that, increasing the localisation loss weight only brings less improvement, even affect the classification task of the model.

Table 4 Detection performance under different localisation loss weight

Weight	0.1	1	2	5	10	100
AUC	96.45	97.96	98.45	98.86	98.89	98.85

Table 5 The ablation experiment of localisation module and consistency attention module (CAM)

	Dataset	
	CASIA V2	*FaceSwap*
w/o localisation	97.34	98.52
w/o CAM	97.54	98.29
Ours	**98.86**	**99.57**

Note: The performance is evaluated by AUC metric.

4.3.2 Tampering detection task

In the experiment, we performed ablation experiments on the consistency attention module. As shown in Table 5, the addition of localisation module can significantly improve the convergence speed of the model. At the same time, we perform the ablation of consistency attention module. It can be seen that the CAM is beneficial for our scheme.

5 Conclusions

In this paper, we propose a detection scheme based on consistent attention to detect inherent differences in tampered images. Our scheme makes the detection model pay more attention to the real/tamper inconsistency in the image. We design a consistent attention module, which uses the attention mechanism to assign importance to the weights extracted from the model and pay more attention to the features related to inconsistency. Multi-task learning also enhances the detection performance of the model. The experimental results show that the consistent attention mechanism can significantly enhance the detection ability of the model.

Acknowledgements

This work is supported in part by the Jiangsu Province Department of Industry and Information Technology Key Technology Innovation Project Orientation Program under grant numbers 141-62-65, in part by the Jiangsu Provincial Science and Technology Department Digital Public Service Platform Project under grant numbers 93208000931.

References

Bappy, J., Mohammed, T.M., Nataraj, L., Flenner, A., Chandrasekaran, S., Roy-Chowdhury, A., Peterson, J.H.L. et al. (2017a) 'Detection and localization of image forgeries using resampling features and deep learning', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.69–77.

Bappy, J.H., Roy-Chowdhury, A.K., Bunk, J., Nataraj, L. and Manjunath, B. (2017b) 'Exploiting spatial structure for localizing manipulated image regions', *Proceedings of the IEEE International Conference on Computer Vision*, pp.4970–4979.

Bayar, B. and Stamm, M.C. (2018) 'Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection', *IEEE Transactions on Information Forensics and Security*, Vol. 13, No. 11, pp.2691–2706.

Bondi, L., Baroffio, L., Güera, D., Bestagini, P., Delp, E.J. and Tubaro, S. (2016) 'First steps toward camera model identification with convolutional neural networks', *IEEE Signal Processing Letters*, Vol. 24, No. 3, pp.259–263.

Christlein, V., Riess, C., Jordan, J., Riess, C. and Angelopoulou, E. (2012) 'An evaluation of popular copy-move forgery detection approaches', *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 6, pp.1841–1854.

Criminisi, A., Pérez, P. and Toyama, K. (2004) 'Region filling and object removal by exemplar-based image inpainting', *IEEE Transactions on image processing*, Vol. 13, No. 9, pp.1200–1212.

Dong, J., Wang, W. and Tan, T. (2013) 'Casia image tampering detection evaluation database', *2013 IEEE China Summit and International Conference on Signal and Information Processing*, pp.422–426.

Fridrich, J. and Kodovsky, J. (2012) 'Rich models for steganalysis of digital images', *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 3, pp.868–882.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) 'Deep residual learning for image recognition', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation*, Vol. 9, No. 8, pp.1735–1780.

Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K.Q. (2017) 'Densely connected convolutional networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.4700–4708.

Huh, M., Liu, A., Owens, A. and Efros, A.A. (2018) 'Fighting fake news: image splice detection via learned self-consistency', *Proceedings of the European Conference on Computer Vision (ECCV)*, pp.101–117.

Korus, P. and Huang, J. (2016) 'Evaluation of random field models in multi-modal unsupervised tampering localization', *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, pp.1–6.

Li, H. and Huang, J. (2019) 'Localization of deep inpainting using high-pass fully convolutional network', *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.8301–8310.

Li, L., Bao, J., Yang, H., Chen, D. and Wen, F. (2020) 'Advancing high fidelity identity swapping for forgery detection', *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.5074–5083.

Mayer, O. and Stamm, M.C. (2019) 'Forensic similarity for digital images', *IEEE Transactions on Information Forensics and Security*, Vol. 15, No. 2020, pp.1331–1346.

Mayer, O. and Stamm, M.C. (2020) 'Exposing fake images with forensic similarity graphs', *IEEE Journal of Selected Topics in Signal Processing*, Vol. 14, No. 5, pp.1049–1064.

Nirkin, Y., Keller, Y. and Hassner, T. (2019) 'FSGAN: subject agnostic face swapping and reenactment', *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.7184–7193.

Nirkin, Y., Wolf, L., Keller, Y. and Hassner, T. (2021) 'Deepfake detection based on discrepancies between faces and their context', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 10, pp.6111–6121.

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

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M. (2019) 'FaceForensics++: learning to detect manipulated facial images', *International Conference on Computer Vision (ICCV)*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) 'Rethinking the inception architecture for computer vision', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2818–2826.

Tralic, D., Zupancic, I., Grgic, S. and Grgic, M. (2013) 'CoMoFoD – new database for copy-move forgery detection', *Proceedings ELMAR-2013*, IEEE, pp.49–54.

Trump's Phone Call to a January 6 Witness Could Be a Federal Crime (2022) [online] https://www.businessinsider.com/january-6-trump-phone-call-could-be-witness-tampering-2022-7 (accessed 7 November 2022).

Wen, B., Zhu, Y., Subramanian, R., Ng, T-T., Shen, X. and Winkler, S. (2016) 'Coverage – a novel database for copy-move forgery detection', *IEEE International Conference on Image processing (ICIP)*, pp.161–165.

Wu, Y., Abd-Almageed, W. and Natarajan, P. (2018) 'Image copy-move forgery detection via an end-to-end deep neural network', *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp.1907–1915.

Zhang, Y., Goh, J., Win, L.L. and Thing, V.L. (2016) 'Image region forgery detection: a deep learning approach', *SG-CRC*, Vol. 2016, pp.1–11.

Zhou, P., Chen, B-C., Han, X., Najibi, M., Shrivastava, A., Lim, S-N. and Davis, L. (2020) 'Generate, segment, and refine: towards generic manipulation segmentation', *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, pp.13058–13065.

Zhou, P., Han, X., Morariu, V.I. and Davis, L.S. (2017) 'Two-stream neural networks for tampered face detection', *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp.1831–1839.