

Image Inpainting Detection Based on High-Pass Filter Attention Network

Can Xiao^{1,2}, Feng Li^{1,2,*}, Dengyong Zhang^{1,2}, Pu Huang^{1,2}, Xiangling Ding³ and Victor S. Sheng⁴

¹Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, Changsha University of Science and Technology, Changsha, 410114, China

²School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, 410114, China

³School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan, 411004, China

⁴Department of Computer Science, Texas Tech University, Lubbock, 79409, USA

*Corresponding Author: Feng Li. Email: lif@csust.edu.cn

Received: 13 January 2022; Accepted: 11 March 2022

Abstract: Image inpainting based on deep learning has been greatly improved. The original purpose of image inpainting was to repair some broken photos, such as inpainting artifacts. However, it may also be used for malicious operations, such as destroying evidence. Therefore, detection and localization of image inpainting operations are essential. Recent research shows that high-pass filtering full convolutional network (HPFCN) is applied to image inpainting detection and achieves good results. However, those methods did not consider the spatial location and channel information of the feature map. To solve these shortcomings, we introduce the squeezed excitation blocks (SE) and propose a high-pass filter attention full convolutional network (HPACN). In feature extraction, we apply concurrent spatial and channel attention (scSE) to enhance feature extraction and obtain more information. Channel attention (cSE) is introduced in upsampling to enhance detection and localization. The experimental results show that the proposed method can achieve improvement on ImageNet.

Keywords: Image inpainting detection; spatial attention; channel attention; full convolutional network; high-pass filter

1 Introduction

Image inpainting is to complete the damaged or lost parts according to the existing data information of the image or remove the target object to ensure that the inpainted region under different scenarios and background perfectly, so that the inpainted image remains naturally, It is hard for the observer to directly observe the damaged or modified marks with the naked eye. As shown in Fig. 1.

With the advancement of science and technology and the development of modern society, digital image information has become an inseparable part of people's lives, so tampered images such as inpainting will flood in people's life and study. Image inpainting plays an important role in many ways. For example, in the photography and film industries, it can achieve certain special effects; in the restoration of cultural relics, it can repair damaged frescoes; in medical imaging, it can also provide the basis for medical



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

diagnosis. It is precise because image inpainting has penetrated all aspects of people's lives that the use of image inpainting should be legalized, and inpainting detection is particularly important.

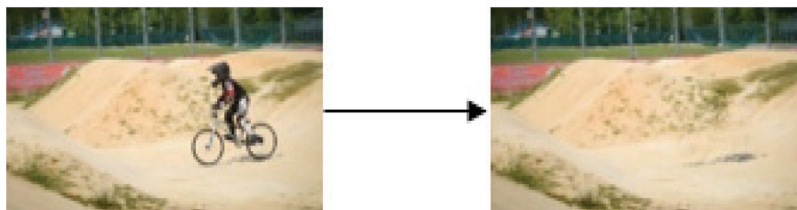


Figure 1: Image inpainting. The left is the original image, and the right is the inpainted image

In recent years, image inpainting has made significant development [1–13]. For traditional inpainting methods based on diffusion or patch [1–5], either the area of the inpaintable hole is very limited, or it cannot inpaint complex scenes or missing objects that are not in the image. For the inpainting method based on deep learning [6–13], the shortcomings of the traditional methods that are perfectly solved, which not only the inpaintable area is enlarged, but also the details of the image can be well predicted and the target that does not exist in the missing image can be automatically generated.

Methods based on deep learning have achieved relatively good inpainting results in image inpainting tasks, and people use image editing software or technology to easily manipulate images. These make the forensics process more difficult, so inpainting detection remains to be further developed. If the processed images are maliciously used by criminals, such as deleting objects in evidence and removing watermarks visible in copyright. Because the manipulation operation distorts the real information of the image, it will cause harm to people's daily life and normal political, economic and social order. Forensics for image inpainting and tampering can determine whether the image to be forensic has been tampered with and the specific inpainted and tampered area of the image to be forensic, which helps people to better identify the real information expressed by the image. Therefore, the study of image inpainting forensics has great social significance and value.

Although satisfactory achievements have been made in the field of image inpainting, there is still a lot of work worth studying about image tampering detection, especially deep image inpainting detection. Many detection methods either only detect for traditional inpainting, or the detection accuracy is not high enough [14–16]. With the continuous development and improvement of neural networks, it is also used in many studies. Li et al. [17] used Convolutional Neural Networks (CNN) combined with high-pass filtering architecture for image inpainting detection and achieved satisfactory performance, but the model did not consider the spatial and channel correlations of feature maps. On the other hand, attention mechanism has been widely used in many fields, which greatly encourages researchers to explore attention mechanism to further improve the performance of network models. Attention mechanism can excite more information from feature maps and can be well integrated into other studies [18–21]. Inspired by this, we combine high-pass filtered fully convolutional networks with spatial and channel attention for image inpainting detection.

The main innovations of this paper are roughly reflected in these aspects. First, we propose a deep learning-based high-pass filtered attention fully convolutional network to detect and localize image regions for deep inpainting operations. Second, the method uses scSE in the feature extraction stage to enhance feature extraction. During upsampling, more attention is paid to channel information to enhance detection and localization. Finally, our model achieves promising results on the public dataset ImageNet.

We arrange this in the subsequent chapters of this article. Section 2 introduces the research content of image inpainting, image inpainting detection methods and attention mechanism. Section 3 presents the

proposed network framework and implementation process. Section 4 introduces the experimental environment, discusses and analyzes the experimental results. Section 5 makes a summary of existing work and discusses future research.

2 Related Work

2.1 Image Inpainting

Many inpainting methods have been developed for image inpainting. These include traditional methods based on diffusion or patches, as well as those based on deep learning. Based on the diffusion method, Li et al. [1] analyzed the diffusion process and found the variation of the Laplacian image, and determined the inpainting area according to the channel information, but the detailed texture of the image could not be completed. Sridevi et al. [2] proposed an image inpainting algorithm, which used Fourier transform to remove noise and blur and processed image boundaries well. But it cannot deal with areas of the curved structure. Based on the patch method, Barnes et al. [3] proposed a random algorithm using incremental update calculations to fill in missing regions by iteratively searching for similar patches. Ružić et al. [4] proposed contextual textures based on Markov random fields to speed up candidate patch search. Although the inpainting area has increased and can be completed by similar patches in the surrounding region, it cannot repair complex scenes or objects that are not in the missing image. Zeng et al. [5] proposed a method to determine the priority of patch based on significance mapping and gray entropy, but the inpainting effect was unclear for larger areas.

In order to solve the shortcomings of traditional methods, learning-based methods have been developed. Pathak et al. [6] were the first to introduce the use of neural networks for image inpainting. There are also some methods based on Generative Adversarial Networks (GAN) [7–9]. Partial convolution [10] only deals with the information of valid regions, and Yu et al. [11] introduced gated convolution on this basis. In a relatively new study, Wang et al. [12] performed image inpainting by adaptively selecting features and normalization. Wang et al. [13] introduced a parallel multi-resolution network for image inpainting.

2.2 Image Inpainting Detection

Some inpainting detection methods have been developed. For example, in some traditional methods, Chang et al. [22] search for inpainting regions by computing similar blocks between regions. Liang et al. [23] also search for inpainted regions by computing similarity hashes, but this approach is limited to detecting simple operations without post-processing. With the improvement of neural networks, Zhu et al. [24] used CNN for detection patch-based image inpainting. Li et al. [17] combined high-pass filtering on the basis of CNN for image inpainting detection, which has the highest correlation with the research content of this paper. As the structure of the network model continues to develop and improve, some networks in other fields are used for inpainting detection. Zhang et al. [25] used feature pyramid network for forensics, but the shortcomings of this method are also obvious, only the detection area is small and the inpainting is based on diffusion. Wang et al. [26] used Mask R-CNN which was originally used for object detection for inpainting detection, and the types of data that can be detected have increased. Although current methods can accomplish certain forensic tasks, there is still a long way to go to improve the generality of detection methods to datasets.

2.3 Attention Mechanism

Recent research shows that the attention mechanism in deep learning has been widely used in many fields [20–21,27,28]. However, Neural Networks is still a frequently used method in computer vision [29–35]. Self-attention is just beginning to slowly seep into the body of research, either complementing existing structures or replacing them entirely. But attention mechanisms have always been a popular technique.

The SE block is the simplest kind of attention mechanism. Hu et al. [18] did not use strict attention and recalibrated the weights in the cropping of feature maps. By using self-attention to model the interdependence between convolutional feature channels, we studied re-weighting the channel response in a certain layer of CNN. Roy et al. [19] introduced spatial excitation on the basis of channel excitation, and finally proposed parallel spatial and channel SE block (scSE), and recalibrated feature maps based on the spatial and channel to obtain the final result. As a result, more information of the feature maps is stimulated, and satisfactory results are achieved, which can be easily applied to other fields. Inspired by this process, we introduce the SE block into a high-pass filtered fully convolutional network for image inpainting detection.

3 Proposed Method

Recent studies have shown that there has been a lot of image forensics work based on convolutional neural networks. These methods are constantly improving and perfecting. The trained models have achieved very good detection results and can be well generalized to other fields. Since the outstanding performance of the HPFCN [17] in image inpainting detection, we consider introducing its main structure into our research and introduce an attention mechanism to propose a deep learning-based image inpainting detection network.

3.1 The Proposed Framework

The proposed model utilizes channel and spatial information in the feature extraction stage, and scSE is introduced to enhance feature extraction and obtain more information. We call it HPACN-sc. Another model pays more attention to the connection between channels in upsampling. We introduce channel attention (cSE) to enhance the effect of detection and location. We call it HPACN-c. The third model uses the method described above to add both scSE and cSE. We call it HPACN-scc. The network structure of our proposed inpainting detection method is shown in Fig. 2.

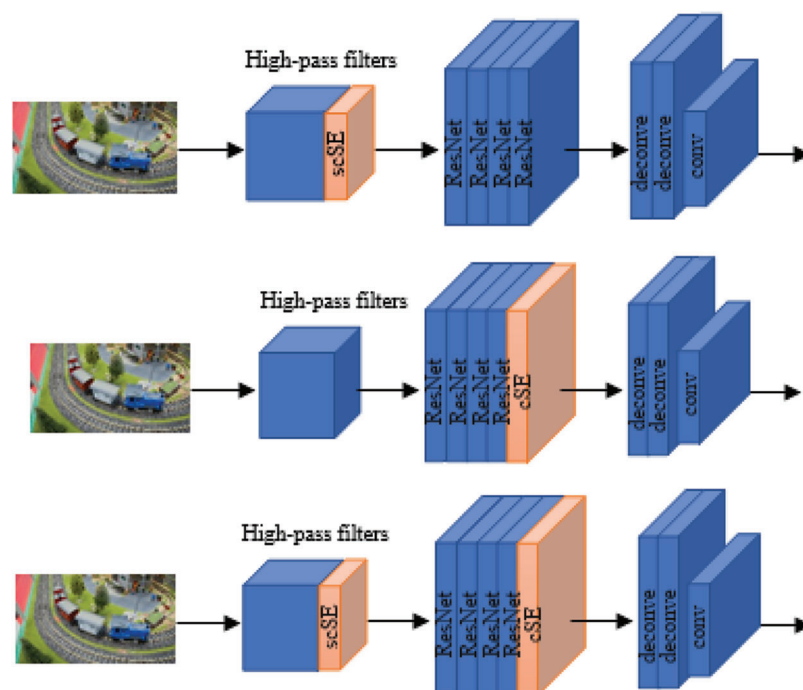


Figure 2: Network framework of the proposed method. The three pictures from top to bottom represent HPACN-sc, HPACN-c, and HPACN-scc

3.2 Three Kinds of SE Blocks

For the feature map X , a new feature map U is obtained by transforming $F_{tr}, F_{tr}: X \rightarrow U, X \in \mathbb{R}^{H \times W \times C'}$, $U \in \mathbb{R}^{H \times W \times C}$. where $X = [x_1, x_2, \dots, x_{C'}]$, $U = [u_1, u_2, \dots, u_C]$, $u_i \in \mathbb{R}^{H \times W}$, C' and C represent the corresponding channel number, and F_{tr} is the convolution operator. Use SE block to process U to obtain a new U' and use it in subsequent layers.

The feature map is globally average pooled to obtain the statistic $Z \in \mathbb{R}^{1 \times 1 \times C}$, the k -th value of Z is shown in Eq. (1).

$$Z_k = \frac{1}{H \times W} \sum_i^H \sum_j^W u_k(i, j) \quad (1)$$

The vector Z passes through two fully-connected layers and a ReLU, and finally performs sigmoid activation to obtain $Z' = W_1(\sigma(W_2 Z))$. where $W_1 \in \mathbb{R}^{C \times C/2}$ and $W_2 \in \mathbb{R}^{C/2 \times C}$. Then squeeze along the spatial and excite the characteristic map U in the channel to obtain the recalibrated and excited U_{cSE} . See Eq. (2).

$$U_{cSE} = F_{cSE}(U) = [\sigma(Z'_1)u_1, \sigma(Z'_2)u_2, \dots, \sigma(Z'_c)u_c] \quad (2)$$

where $\sigma(Z'_i)$ represents the importance of the i -th channel. This recalibration pays more attention to the more important channels while ignoring the unimportant ones.

Consider again, for the feature map $U = [u^{1,1}, \dots, u^{i,j}, \dots, u^{H,W}]$, $u^{i,j}$ represents the corresponding spatial position. The spatial projection is obtained by convolution, and then the sigmoid is mapped back to the original position. That is $q = W_{sq} * U$, where $W_{sq} \in \mathbb{R}^{1 \times 1 \times C \times 1}$, $q \in \mathbb{R}^{H \times W}$. Then squeeze along the channel and excite the feature map U in spatial to obtain the recalibrated and excited U_{sSE} . As shown in Eq. (3).

$$U_{sSE} = F_{sSE}(U) = [\sigma(q_{1,1})u^{1,1}, \dots, \sigma(q_{i,j})u^{i,j}, \dots, \sigma(q_{H,W})u^{H,W}] \quad (3)$$

where $\sigma(q_{i,j})$ represents the importance of the spatial position (i, j) . This recalibration pays more attention to the more important spatial positions while ignoring the unimportant ones.

The scSE block is to get the squeeze and excitation of parallel spatial and channel by adding U_{sSE} and U_{cSE} . That is $U_{scSE} = U_{sSE} + U_{cSE}$. In theory, this kind of recalibration will focus on important parts of the channel and spatial position at the same time, and it will achieve better results.

For a more detailed introduction, please refer to the content in [18] and [19]. For these three types of SE blocks, we introduce scSE block and cSE block among them for the corresponding image forensic research.

3.3 Implementation Process

3.3.1 Preprocessing Module

The image of the dataset are used as input of the preprocessing module, and the traces left by the image tampering operation are enhanced by high-pass filtering, and the 9-channel image residual of the focused spatial information and channel information is obtained by using scSE. The output terminal of the preprocessing module is connected to the input terminal of the feature extraction module; here the high-pass filtering consists of three deep convolutions with a stride of 1. The size of the filter kernel is 3×3 and is learnable. More detailed content can be found in [17].

3.3.2 Feature Extraction Module

ResNet v2 [36] made changes on the basis of ResNet v1 [37], first batch normalization and ReLU then convolution, and the feature extraction module is constructed according to this structure. This module is used to collect distinguishable features from the above image residuals to obtain a feature map, which consists of

four identical ResNet blocks, each of which contains two bottleneck units, and each unit includes three convolution layers with 1×1 , 3×3 , 1×1 convolutional kernels respectively. The specific settings can be found in [17]. After the previous scSE processing, feature extraction will pay attention to both spatial and channel information, and finally, 1024 feature maps are obtained. These obtained feature maps, after being processed by the cSE block, will be used as the input of the subsequent upsampling module.

3.3.3 Upsampling Location Module

After the processing of the previous module, the upsampling location module pays more attention to the dependence between channels. Use transposed convolution to improve the spatial resolution to obtain the category label of each pixel, and finally realize the output of tampering and location. Where the transpose convolution, including the first transpose convolution, the kernel size is 8×8 , the output channel is 64 and the stride is 4; Second transpose convolution, kernel size is 8×8 , the output channel is 4 and the stride is 4. The transposed convolution uses a learnable bilinear kernel to perform four times upsampling twice to keep the number of elements in the feature map before and after upsampling the same. Finally, a convolution with a kernel size of 5×5 and a stride of 1 is performed to further weaken the checkerboard artifacts, and finally realize the output of tampering and location.

4 Experiments

The framework used in the experiments in this paper is TensorFlow. The configuration and environment of the experiment are as follows: Ubuntu 16.04, GPU: NVIDIA GeForce RTX 2080Ti, CUDA version is 10.0.

4.1 Dataset and Evaluation Criteria

The images in this paper are from ImageNet [38], the quality factor of these images is mostly 75, and the inpainting method in [7] is used to process the images. The dataset has a total of 60,000 images, in which training data and validation data are configured in a ratio of 5:1. The configuration of the dataset refers to HPFCN, because the model itself is not perfect for detecting complex situations, so the mask corresponding to the dataset is a 10% rectangular area in the center of the image.

As shown in Tab. 1. We compare HPACN with DCC [14], SCO [15], MFCN [16] and HPFCN [17]. The settings and data of some of the methods are from [17], and use Recall, Precision, F1-score, and IoU to evaluate the effect of the model's image inpainting detection.

Table 1: Comparison of detection results between our method and other different methods

Detector	Recall	Precision	F1-score	IoU
DCC	63.13	37.51	38.88	26.62
SCO	83.08	24.57	32.92	21.07
MFCN	57.15	53.04	45.01	33.73
HPFCN	89.84	95.22	91.53	85.93
HPACN (ours)	91.93	94.38	92.46	87.09

4.2 Ablation Study

We compare FCN, ACN-c, HPFCN, HPSE-scc, HPSE-sc and HPSE-c. FCN in this paper represents a fully convolutional network. ACN-c represents an attention convolutional network with channel attention added in upsampling. HPFCN represents the high-pass filtered fully convolutional network in [17].

HPACN-scc, HPACN-sc, and HPACN-c represent high-pass filtering attention fully convolutional networks that add both scSE and cSE, only scSE, and only cSE, respectively. Their comparison results are shown in [Tab. 2](#).

Table 2: Localization results (%) of ablation study. The model is trained on the imagenet dataset

Detector	Recall	Precision	F1-score	IoU
FCN	82.22	92.16	84.54	76.63
ACN-c	74.86	94.64	80.54	72.13
HPFCN	89.84	95.22	91.53	85.93
HPACN-scc (ours)	90.09	95.45	91.86	86.38
HPACN-sc (ours)	89.48	95.86	91.66	86.16
HPACN-c (ours)	91.93	94.38	92.46	87.09

We found that the processing of high-pass filtering is one of the important factors for the subsequent channel attention to play a role. The three models introduced have different degrees of improvement, especially HPACN-c, which had the most significant effect, the corresponding F1-score improved by 0.9%, IoU improved by 1.1%. It can be seen from the experimental results that without the preprocessing of high-pass filtering, the direct use of channel attention leads to performance degradation. The possible reason is that the high-pass residual after high-pass filtering preprocessing makes the difference between the inpainted area and the original area more obvious, and the feature difference between the corresponding channels may also be amplified. Exciting more important channels will significantly improve the performance.

Because the result of the forensic process in this paper is the generation of pixel-by-pixel localization maps, and spatial information is very important to the entire pixel-level forensics process. For the feature extraction module, both spatial location information and channel information are very important. One possible explanation is that in the upsampling module, where the forensic results of the class labels of each pixel are about to be generated, each spatial location information is irreplaceable, which makes the contribution of some channels less important and squeezing unimportant channels will significantly improve model performance.

The proposed method can effectively detect and locate the inpainting region in our dataset, and examples of detection is shown in [Fig. 3](#). In addition, we also found that if channel attention is used at all stages, or used arbitrarily, it will lead to the decline of model performance. In future work, we will continue to modify the network structure to improve the generalization of the model to the dataset and further improve the performance of the model.

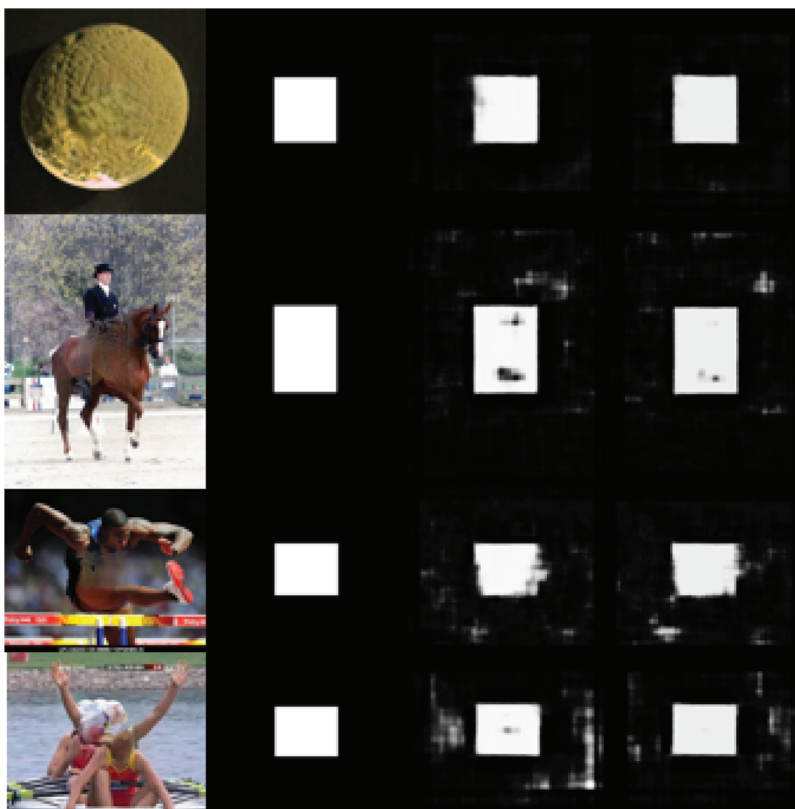


Figure 3: Examples of the localization effect maps of the deep inpainting. The four columns from left to right represent inpainted images, groundtruth, localization maps of HPFCN [17] and the localization maps of the proposed method

5 Conclusion

This paper propose a high-pass filtered attention full convolutional network to detect and locate the image inpainting. Taking into account the high-pass filter full convolutional network, the spatial location and channel dependence are not considered in the feature extraction and the upsampling stage. We combine spatial and channel information to enhance feature extraction, so as to obtain more information, and use channel attention to improve the localization of image inpainting detection. The experimental show that our method has achieved satisfactory results. In the following research, we will consider using more advanced inpainting methods to process the dataset, improve the generalization of the model to detect complex inpainting images.

Funding Statement: This project is supported by the National Natural Science Foundation of China under Grant 62172059, 61972057 and 62072055, Hunan Provincial Natural Science Foundations of China under Grant 2020JJ4626, Scientific Research Fund of Hunan Provincial Education Department of China under Grant 19B004, Postgraduate Scientific Research Innovation Project of Hunan Province under Grant CX20210811.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] H. Li, W. Luo and J. Huang, "Localization of diffusion-based inpainting in digital images," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 3050–3064, 2017.
- [2] G. Sridevi and S. S. Kumar, "Image inpainting based on fractional-order nonlinear diffusion for image reconstruction," *Circuits Systems and Signal Processing*, vol. 38, no. 8, pp. 3802–3817, 2019.
- [3] C. Barnes, E. Shechtman, A. Finkelstein and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. on Graphics*, vol. 28, no. 3, pp. 1–24, 2009.
- [4] T. Ružić and A. Pižurica, "Context-aware patch-based image inpainting using markov random field modeling," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 444–456, 2014.
- [5] J. Zeng, X. Fu, L. Leng and C. Wang, "Image inpainting algorithm based on saliency map and gray entropy," *Arabian Journal for Science & Engineering*, vol. 44, no. 4, pp. 3549–3558, 2019.
- [6] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 2536–2544, 2016.
- [7] S. Iizuka, E. Simo-Serra and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.
- [8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu *et al.*, "Generative image inpainting with contextual attention," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 5505–5514, 2018.
- [9] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi and M. Ebrahimi, "Edgeconnect: Generative Image Inpainting with Adversarial Edge Learning," arXiv preprint arXiv:1901.00212, 2019.
- [10] G. Liu, F. A. Reda, K. J. Shih, T. Wang, A. Tao *et al.* "Image inpainting for irregular holes using partial convolutions," in *Proc. of the European Conf. on Computer Vision*, Munich, Germany, pp. 85–100, 2018.
- [11] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu *et al.* "Free-form image inpainting with gated convolution," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, Korea, pp. 4471–4480, 2019.
- [12] N. Wang, Y. Zhang and L. Zhang, "Dynamic selection network for image inpainting," *IEEE Transactions on Image Processing*, vol. 30, pp. 1784–1798, 2021.
- [13] W. Wang, J. Zhang, L. Niu, H. Ling, X. Yang *et al.*, "Parallel multi-resolution fusion network for image inpainting," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, Korea, pp. 14559–14568, 2021.
- [14] H. Li, B. Li, S. Tan and J. Huang, "Detection of Deep Network Generated Images Using Disparities in Color Components," arXiv preprint arXiv:1808.07276, 2018.
- [15] M. Huh, A. Liu, A. Owens and A. A. Efros, "Fighting fake news: Image splice detection via learned self-consistency," in *Proc. of the European Conf. on Computer Vision*, Munich, Germany, pp. 101–117, 2018.
- [16] R. Salloum, Y. Ren and C. C. Kuo, "Image splicing localization using a multi-task fully convolutional network (MFCN)," *Journal of Visual Communication and Image Representation*, vol. 51, pp. 201–209, 2018.
- [17] H. Li and J. Huang, "Localization of deep inpainting using high-pass fully convolutional network," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, Korea, pp. 8301–8310, 2019.
- [18] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
- [19] A. G. Roy, N. Navab and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, Granada, Spain, pp. 421–429, 2018.
- [20] G. Hou, J. Qin, X. Xiang, Y. Tan and N. N. Xiong, "Af-net: A medical image segmentation network based on attention mechanism and feature fusion," *Computers, Materials & Continua*, vol. 69, no. 2, pp. 1877–1891, 2021.
- [21] M. Choi, H. Kim, B. Han, N. Xu and K. M. Lee, "Channel attention is all you need for video frame interpolation," in *Proc. of the AAAI Conf. on Artificial Intelligence*, Palo Alto, California, USA, pp. 10663–10671, 2020.
- [22] I. C. Chang, J. C. Yu and C. C. Chang, "A forgery detection algorithm for exemplar-based inpainting images using multi-region relation," *Image and Vision Computing*, vol. 31, no. 1, pp. 57–71, 2013.

- [23] Z. Liang, G. Yang, X. Ding and L. Li, "An efficient forgery detection algorithm for object removal by exemplar-based image inpainting," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 75–85, 2015.
- [24] X. Zhu, Y. Qian, X. Zhao, B. Sun and Y. Sun, "A deep learning approach to patch-based image inpainting forensics," *Signal Processing: Image Communication*, vol. 67, pp. 90–99, 2018.
- [25] Y. Zhang, F. Ding, S. Kwong and G. Zhu, "Feature pyramid network for diffusion-based image inpainting detection," *Information Sciences*, vol. 572, no. 9, pp. 29–42, 2021.
- [26] X. Wang, S. Niu and H. Wang, "Image inpainting detection based on multi-task deep learning network," *IETE Technical Review*, vol. 38, no. 1, pp. 149–157, 2021.
- [27] P. Han, M. Zhang, J. Shi, J. Yang and X. Li, "Chinese q&a community medical entity recognition with character-level features and self-attention mechanism," *Intelligent Automation & Soft Computing*, vol. 29, no. 1, pp. 55–72, 2021.
- [28] K. Prabhu, S. SathishKumar, M. Sivachitra, S. Dineshkumar and P. Sathiyabama, "Facial expression recognition using enhanced convolution neural network with attention mechanism," *Computer Systems Science and Engineering*, vol. 41, no. 1, pp. 415–426, 2022.
- [29] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [30] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 3431–3440, 2015.
- [31] R. Girshick, "Fast R-CNN," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Boston, MA, USA, pp. 140–1448, 2015.
- [32] F. Zhang, N. Cai, J. Wu, G. Cen, H. Wang *et al.* "Image denoising method based on a deep convolution neural network," *IET Image Processing*, vol. 12, no. 4, pp. 485–493, 2018.
- [33] H. P. Wu, Y. L. Liu and J. W. Wang, "Review of text classification methods on deep learning," *Computers, Materials & Continua*, vol. 63, no. 3, pp. 1309–1321, 2020.
- [34] C. L. Wang, Y. L. Liu, Y. J. Tong and J. W. Wang, "GAN-GLS: Generative lyric steganography based on generative adversarial networks," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 1375–1390, 2021.
- [35] D. Y. Zhang, J. W. Hu, F. Li, X. L. Ding, A. K. Sangaiah *et al.* "Small object detection via precise region-based fully convolutional networks," *Computers, Materials & Continua*, vol. 69, no. 2, pp. 1503–1517, 2021.
- [36] K. He, X. Zhang, S. Ren and J. Sun, "Identity mappings in deep residual networks," in *Proc. of the European Conf. on Computer Vision*, Springer, Cham, Amsterdam, The Netherlands, pp. 630–645, 2016.
- [37] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.
- [38] J. Deng, W. Dong, R. Socher, L. Li, K. Li *et al.*, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248–255, 2009.