Proceedings of Machine Learning Research 189, 2022 ACML 2022

FF-Net: An End-to-end Feature-Fusion Network for Double JPEG Detection and Localization

Bo Liu boliu@cqupt.edu.cn
Ranglei Wu s190201095@stu.cqupt.edu.cn
Xiuli Bi bixl@cqupt.edu.cn
Bin Xiao xiaobin@cqupt.edu.cn
Chongqing University of Posts and Telecommunications, Nan'an District, Chongqing 400060, China

Editors: Emtiyaz Khan and Mehmet Gönen

Abstract

In the real-world, most images are saved in JPEG format, so many forged images are partially or totally composed of JPEG images and then saved in JPEG format again. In this case, exposing forged images can be accomplished by the detection of double JPEG compressions. Although the detection methods of double JPEG compressions have greatly improved, they rely on handcrafted features of image patches and cannot locate forgery at pixel-level. To break this limitation, we propose an end-to-end feature-fusion network (FF-Net) for double compression detection and forgery localization. We find that JPEG compression fingerprint primarily exists on the high-frequency component of an image, and the singly and doubly compression yield different fingerprints. Therefore, we design two encoders cooperatively to learn the compression fingerprint directly from the whole image. A decoder is deployed to locate the regions with different compression fingerprints at pixel-level based on the learned compression fingerprint. The experiment results verify that the proposed FF-Net can detect and locate the forged regions more accurately than these existing detection methods. Besides, it has a good generalization ability that the network trained on one compression case can work in numerous compression cases. Moreover, it can detect different local forgeries, including copy-move, splicing, and object-removal.

Keywords: Double JPEG compressions; JPEG fingerprint; end-to-end; pexel-level; forgery detection.

1. Introduction

JPEG compression is a lossy compression scheme proposed by the Joint Photographic Experts Group, and it is the most widely-used image format. Many forged images are partially or totally generated by JPEG images. Meanwhile, JPEG compression is usually used as a post-processing operation because it can significantly reduce tampering traces. Considering the JPEG scheme compresses an image in each 8×8 blocks, if the local image forgery, such as splicing, copy-move, and object-removal, is accomplished in JPEG images as shown in Figure 1, the tampered regions will cause some 8×8 unaligned blocks covering the former compression blocks. Since the blocks are not aligned, they override the former compression traces in these blocks. When the final forged image is saved in JPEG format, the tampered regions seem to be singly JPEG compressed while the un-tampered regions are doubly JPEG compressed. Therefore, detecting single/double JPEG compressions can help expose these local image forgeries.

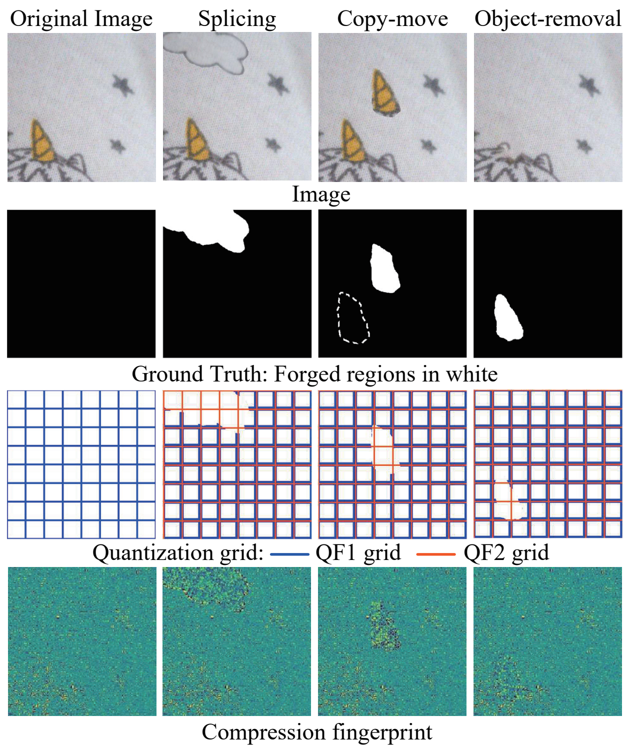


Figure 1: Three local forgeries in JPEG images. In the third row, QF1 grid represents the 8*8 blocks in the original JPEG images; QF2 grid represents the 8×8 blocks in the final saved images in JPEG format. The dashed line in ground truth indicates the original region in copy-move forgery.

For double JPEG compressions detection, researchers have proposed various traditional methods. The image-level authentication methods perform forgery detection based on whether the image owns a consistent JPEG compression fingerprint. Lukáš and Fridrich (2003) found that an image with double JPEG compressions will yield double quantization (DQ) effect and then presents an algorithm to detect such double compressions. Fu et al. (2007) found that Benford's rule occurs in JPEG coefficients, so they utilized Benford's rule to discriminate doubly-compressed JPEG images from singly-compressed JPEG images. Pixel-level detection methods can locate the regions with different compressions in a forged image. In pixel-level detection, Bianchi and Piva (2012) used the Bayesian method to calculate the probability of each 8×8 DCT block being doubly JPEG compressed according to the DQ effect. Besides, Farid (2009) proposes a method to detect and locate the regions with different compressions via extraction of JPEG ghost.

In recent years, with the development of convolutional neural networks (CNNs), many CNN-based methods to detect double JPEG compressions have been proposed. However, these methods are performed at image-level. Wang and Zhang (2016) designed a 1-D CNN to automatically distinguish the singly compressed regions and doubly compressed regions by classifying DCT histograms. Barni et al. (2017) utilized CNNs in the pixel domain, the noise domain, and the DCT domain to perform detection tasks, respectively. This method can obtain comprehensive information from a JPEG image but is quite complicated. Amerini et al. (2017) proposes a multi-domain CNN to improve the performance of frequency domain-based CNN. Park et al. (2018) inserted the quantization tables into a CNN to detect double JPEG compressions with mixed JPEG quality factors. Furthermore, they designed a 3D CNN Ahn et al. (2019) to deal with the task and got a better performance.

Although these CNN-based methods have made great progress, they still rely on the statistics of DCT coefficients, in particular DCT coefficient histograms, which causes two issues. First, these methods learn DCT coefficient histograms from image patches by a complex handcraft extraction process; Second, these methods obtain each image patch's classification result, which essentially is not a pixel-level detection. We proposed an end-to-end Feature-Fusion Network (FF-Net) to detect double JPEG compressions and localize forged areas at pixel-level to break these limitations. The main contributions of our work in this paper can be summarized as follows:

- We proposed JPEG compression fingerprint, which is based on the study of JPEG compression's impact on the high-frequency of images and the analysis of the differences caused by various quantization processes.
- We proposed FF-Net, an end-to-end network for double JPEG compressions detection and pixel-level forgery localization by direct learning the compression fingerprint.
- To the best of our knowledge, the proposed FF-Net is the first network that can actually detect various local image forgeries in JPEG images.

2. JPEG Compression Fingerprint

For an image saved in JPEG format, JPEG compression reduces the storage space mainly by choosing a quality factor. This quality factor corresponds to a quantization table that is used to quantize the DCT coefficients of the image in the quantization process in the JPEG compression pipeline. Meanwhile, the quantization process is the main reason causing the deterioration of image quality. Therefore, the quality deterioration caused by JPEG compression has its unique pattern determined by the quantization table. Since the response of the quantization process in the spatial domain is the suppression of high-frequency in the frequency domain, the change of high-frequency in the frequency domain of the JPEG image also can represent the change of JPEG compression. Therefore, we used some common high-frequency extraction methods to explore the traces left by JPEG compressions.

Spatial Rich Model(SRM) Fridrich and Kodovsky (2012), which contains 30 high-pass filters, has been proposed for image steganographic analysis. To extract the high-frequency of images under different compressions, one SRM filter is applied to the images with single compression, double compressions, and combined compressions separately, as shown in Figure 2. However, from the response maps in Figure 2-(c1-c3), we cannot observe the differences in the three scenarios. By analyzing the histograms of these response maps, we find that most values in response maps are distributed in [-10 10]. Therefore, we map the values in [-10 10] to a color bar, and the processed response maps are shown in the fourth row of Figure 2. As shown in Figure 2-(d1-d3), the response of the double compression image is weaker than that of the single compression image. Furthermore, a clear transitional zone can be seen at the edges between singly and doubly compressed regions, proving that the quantization process causes image quality deterioration. The deterioration can be verified on a larger scale, and the traces between aligned and unaligned compression can be easily discriminated. Figure 3 shows the average amplitude in the vertical and horizontal direction in the DFT domain of image blocks randomly sampled from 360 SRM filtered images. Each image block was undergone aligned and unaligned compression. In each case, although the compression qualities are the same, the traces of aligned and unaligned compression are far different and can be discriminated easily.

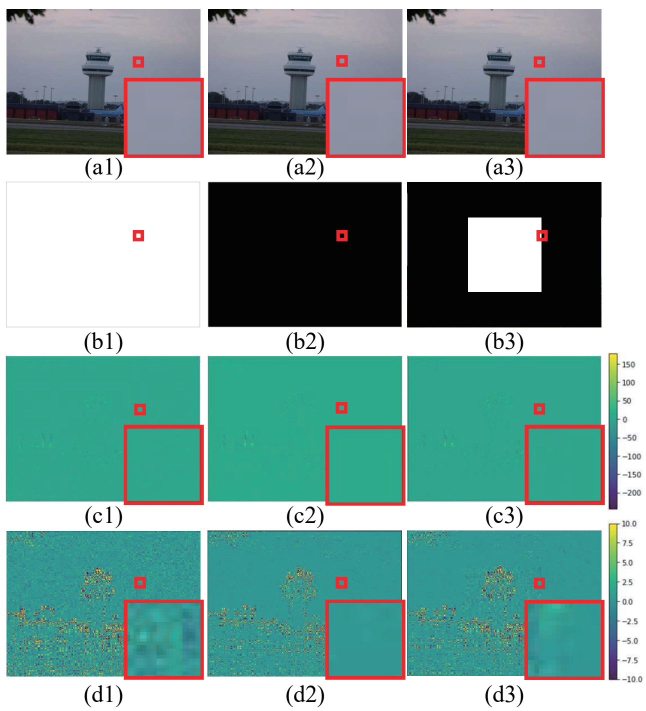


Figure 2: The filtered results of a SRM filter on an image under different compression cases (R-channel). (a1) the image is compressed with QF1_S=90, (a2) the image is first compressed with QF1_D=50 and then recompressed with QF2_D=90. (a3) the image is composited by (a1) and (a2). (b1-b3) shows the single compression regions by white color and the twice compression regions by black color. (c1-c3) shows the filtered results of (a1-a3). (d1-d3) shows these values between [-10 10] in (c1-c3). The red box enlarges the part of the edge regions between single and twice compression.

To verify this insight again, Discrete Wavelet Transform (DWT), which can decompose the texture information of images and amplify the details of images effectively by extracting

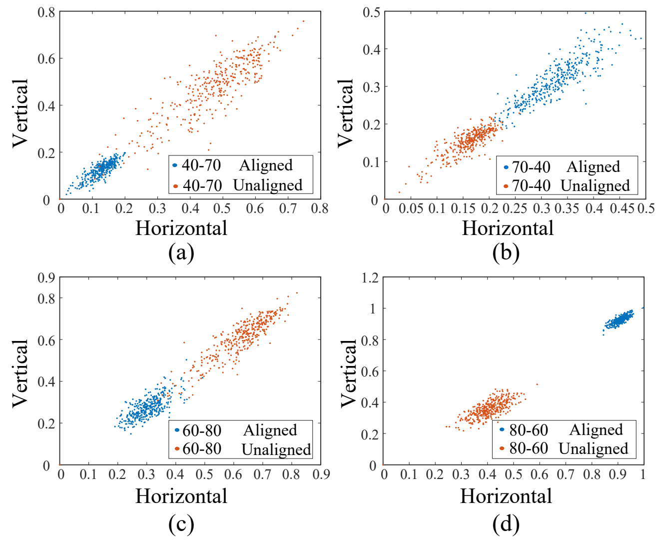


Figure 3: The aligned and unaligned compression yield different traces. The numbers in the legends indicate the first and second compression quality applied in JPEG compressions.

the gradient of pixels, can also be utilized like SRM. We did the same analysis on the filtered results processed by DWT. A three-level Haar DWT was applied to process Figure 2-(a3), and the results are shown in Figure 4. It can be seen that there is a clear transitional zone at the edges between the singly and the doubly compressed regions. Moreover, this transitional zone, which shows the difference between the singly and the doubly compressed regions, is more obvious with higher-level DWT.

3. Feature-Fusion Network(FF-Net)

Based on the insight in the above section, a locally forged image, which had been partially or wholely composited by images in JPEG format and then saved as JPEG format again, is fed into a network. If the network can focus on learning JPEG compression fingerprint rather than the image content, it can authenticate images and locate the regions with different JPEG compression fingerprints based on the learned JPEG compression fingerprint.

Following this idea, we chose three semantic segmentation networks: SegNet Badrinarayanan et al. (2017), U-Net Ronneberger et al. (2015) and DenseU-Net Li et al. (2018), which are popular for the image segmentation tasks. We utilized them to learn the compression fingerprint directly from the input image. The four examples of training and test images are shown in Figure 5-(a), and they were composited by single compression and double compressions. As shown in Figure 5-(b), white color denotes the regions compressed only one time ($QF1_S$=70), and black color regions are the regions compressed twice ($QF1_D$-$QF2_D$=50-70).

It can be seen from the experimental results in Figure 5 that three semantic segmentation networks tend to segment the objects. In the test images where the singly compressed

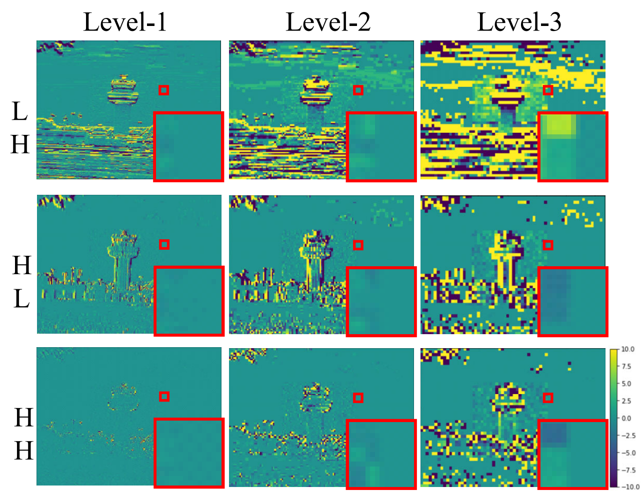


Figure 4: The filtered results by DWT on figure 2-(a3)(R-channel). From left to right are the results of level-1, level-2, and level-3 DWT. From top to bottom are the results of LH, HL and HH subbands.

regions have no semantic information in the first and second columns of Figure 5, three semantic segmentation networks failed to detect the regions. In the test images with salient objects in the third and fourth columns of Figure 5, they detected a part of objects. However, the detected regions are not the regions with the different compressed fingerprints. This experiment proves that the three semantic segmentation networks can hardly directly learn the JPEG compression fingerprint from the input image.

Since the compression fingerprint of a JPEG image can be explored in its high-frequency component while the semantic segmentation networks cannot learn the compression fingerprint from the input image directly, we considered how to expose the high-frequency of the image to help the network learn the compression fingerprint. Therefore, we designed a spatial rich models (SRM) encoder and a discrete wavelet transform (DWT) encoder based on the study of JPEG compression fingerprint. By combining the two encoders, we proposed an end-to-end Feature-Fusion Network (FF-Net). The framework of FF-Net is shown in Figure 6.

We firstly designed SRM convolutional layers and embedded them into an encoder of SegNet to form an SRM encoder. The SRM convolutional layers consist of five SRM filters(as shown in Figure 7, and our experiment have verified that this combination achieved the best performance than other combinations). Fixing the parameters of the first convolutional layer of the network to the value of SRM can suppress the extraction of semantic information of the image, forcing the network to pay attention to the high-frequency component of the image and accurately learn the compression fingerprint. The SegNet(+SRM) was trained and tested under the same condition as before. Compared to the experimental results of SegNet, SegNet(+SRM) can accurately detect and locate the regions with different compression fingerprints(as shown in Figure5). The results prove that the learned

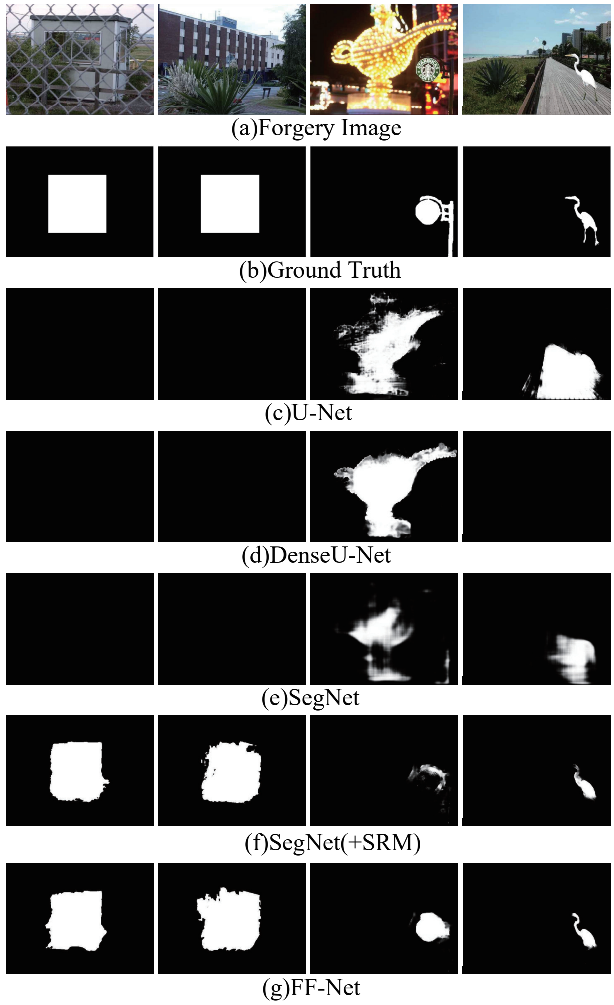


Figure 5: The detection results under (50-70, 70) by FF-Net and other comparative detection methods.

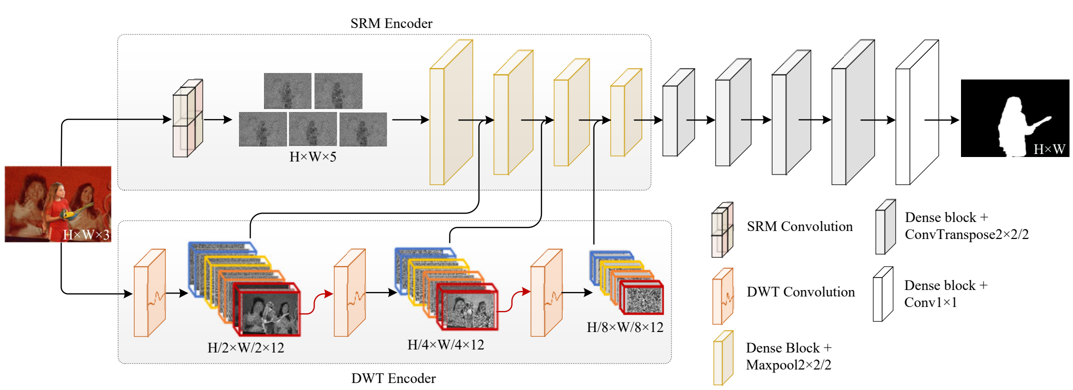


Figure 6: The structure of Feature-Fusion Network.

$$\frac{1}{12}\times\begin{bmatrix}-1&2&-2&2&-1\\2&-6&8&-6&2\\-2&8&-12&8&-2\\2&-6&8&-6&2\\-1&2&-2&2&-1\end{bmatrix}\quad\frac{1}{12}\times\begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\-2&8&-12&8&-2\\2&-6&8&-6&2\\-1&2&-2&2&-1\end{bmatrix}\quad\frac{1}{12}\times\begin{bmatrix}-1&2&-2&0&0\\2&-6&8&0&0\\-2&8&-12&0&0\\2&-6&8&0&0\\-1&2&-2&0&0\end{bmatrix}\quad\frac{1}{12}\times\begin{bmatrix}-1&2&-2&2&-1\\2&-6&8&-6&2\\-2&8&-12&8&-2\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}\quad\frac{1}{12}\times\begin{bmatrix}0&0&-2&2&-1\\0&0&8&-6&2\\0&0&-12&8&-2\\0&0&8&-6&2\\0&0&-2&2&-1\end{bmatrix}$$

Figure 7: The five SRM filters in SRM encoder.

features by the SRM encoder can be consequently utilized in the decoder of Seg-Net to discriminate the regions with different compression fingerprints.

Furthermore, we designed a DWT encoder (as shown in Figure 8) and added the DWT encoder to the SegNet(+SRM) for enhancing the capability of the network to learn the compression fingerprint. Because the filtering process of DWT is actually a sum-pooling operation, the features obtained from each level of the DWT encoder and the features obtained from each pooling layer of the SRM encoder can be well aligned and fused. In addition, because DWT does not introduce additional parameters, the network will not occupy more memories with the DWT encoder.

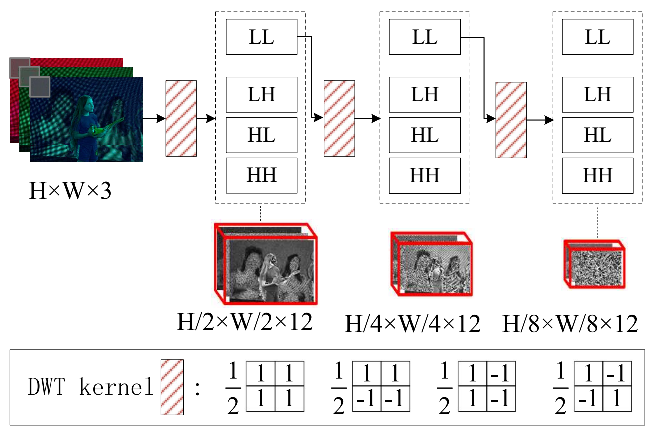


Figure 8: The structure of DWT encoder.

Our task focuses on an image's high-frequency information, which is a low-level feature rather than a high-level representation, thus the features generated by the initial layers are essential. We replaced the original convolution layers with dense connections Huang et al. (2017) to encourage the utilization of low-level features in the whole network. In this paper, each dense block consists of an input layer and three convolutional layers, and the input of each convolutional layer in a dense block is defined in Eq. (1):

$$input_{conv(i)} = inputs + \sum_{j}^{i-1} output_{conv(j)}, \tag{1}$$

Where $inputs$ represents the input of dense block, $input_{conv(i)}$ represents the input of i^{th} convolution in this block, and $output_{conv(j)}$ represents the output of j^{th} convolution in this block.

Finally, the proposed FF-Net was trained and tested under the same experimental condition as before. The test results of FF-Net(shown in Figure 5-(g)) are much more accurate than SegNet(+SRM), owing to the DWT encoder, which provides the network with more subtle features, and dense connection increases the utilization of features in the network.

4. Experiments and Analysis

To evaluate the proposed FF-Net's performance, the ablation experiments and the comparative experiments were carried out in this section. Meanwhile, to explore the generalization capability of FF-Net, we generated the test sets under various compression cases.

4.1. Datasets

Since the previous double JPEG compressions detection methods classify image blocks, there is no dataset specialized for double JPEG compressions detection. For evaluating the proposed method, we generated a set of training and test datasets of singly and doubly compressed images with different quality factors (QF). All datasets were built from images in UCID Schaefer and Stich (2003) and MIT-Adobe FiveK Bychkovsky et al. (2011). The UCID consists of 1338 TIFF images with the resolution of 512×384 (or 384×512). MIT-Adobe FiveK consists of 5000 raw images. The UCID was divided into two parts without the same images. One part contains 300 images used to generate the training sets, and the other contains the remaining images used to generate the test sets.

To generate the training sets, 200 images in the first part were selected as background images, and 20 images were selected as object providers. Having the background images compressed with $\mathrm{QF1}_D \in \{30, 35, \ldots, 90, 95\}$, they were duplicated into 20 copies. Then some regions in the first object provider image were spliced to all images in the first copy set, and the pasted position was random. After that, the same manipulation was applied to the remaining object provider images and duplicated background images. Finally, all manipulated images were recompressed with $\mathrm{QF2}_D \in \{50, 55, \ldots, 90, 95\}$, namely $\mathrm{QF1}_S$.

To generate the test sets, 80 images were selected as background images and another 80 images as object provider images in the second part. The manipulation was the same as the one used in training sets, except that each object provider image and background image are one-to-one corresponded. In this way, each compression case contains 80 test images. In addition, to make a subjective comparison with the previous double JPEG compressions detection method, we generated several large test images with MIT-Adobe FiveK.

4.2. Evaluation Metrics

The methods' performances for detecting double JPEG were evaluated by the P, R, and F. In the ground truth, white represents single JPEG compression, while black represents double JPEG compression. P, defined by Eq. (2), is the ratio of the pixels correctly classified to single JPEG compression to all pixels classified to single JPEG compression. R, defined by Eq. (3), is the ratio of the pixels correctly classified to single JPEG compression to the ground truth. TP and FP denote the numbers of correctly classified and erroneously classified pixels, respectively, and FN is the number of falsely missed pixels. The F is the weighted harmonic mean of P and R and is given by Eq. (4). The P,R,F in experiments is the average of test sets.

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{4}$$

1pt

Table 1: Performance comparisons of different methods under four compression cases.

2*Method		(40-60,60)			(50-70,70)			(60-80,80)			(70-90,90)		
		P	R	F	P	R	F	P	R	F	P	R	F
4*Ablation methods	SegNet	0.254	0.157	0.194	0.268	0.314	0.289	0.383	0.369	0.376	0.686	0.382	0.491
	SegNet(+SRM)	0.805	0.634	0.709	0.848	0.695	0.764	**0.891**	0.683	0.774	0.948	0.822	0.881
	SegNet(+SRM+DWT)	0.885	0.615	0.726	**0.887**	0.692	0.777	0.887	0.723	0.797	0.945	0.822	0.879
	DenseSegNet(+SRM)	0.819	0.655	0.728	0.876	0.682	0.767	0.87	0.734	0.796	0.944	0.834	0.886
6*Comparison methods	U-Net	0.381	0.162	0.227	0.329	0.155	0.211	0.477	0.284	0.356	0.537	0.326	0.406
	DenseU-Net	0.151	0.077	0.102	0.153	0.079	0.104	0.207	0.096	0.131	0.439	0.233	0.304
	RRU-Net	0.167	0.608	0.262	0.168	0.603	0.263	0.163	0.692	0.264	0.204	0.643	0.31
	FusionNet	0.487	0.371	0.421	0.488	0.439	0.462	0.528	0.5	0.513	0.532	0.557	0.544
	Park et al.	0.038	0.061	0.047	0.02	0.035	0.025	0.02	0.035	0.025	0.027	0.056	0.036
	FF-Net	**0.891**	**0.687**	**0.776**	0.877	**0.744**	**0.805**	0.888	**0.759**	**0.818**	**0.951**	**0.857**	**0.902**

4.3. Implementation Details

We implemented the model by PyTorch. A Batch Normalization (BN) Ioffe and Szegedy (2015) is added between the convolutional layer and the rectified linear unit (ReLU) to normalize the output features of each convolutional layer. FF-Net and the compared detection methods ran on the GPU of NVIDIA Tesla V100 of 16GB memory size with CUDA version 10.1 and CUDNN version 6.0. The model was trained using Adam optimizer with a batch size of 10. The model parameters were optimized using the binary cross-entropy loss function to accelerate model convergence. The learning rate is initialized to 0.0001. The parameters of all compared methods were set according to their best performances.

4.4. Experimental Comparison and Analysis

4.4.1. Ablaction Experiments

The effectiveness of the SRM encoder, DWT encoder, and the dense connection was evaluated subjectively. For the fairness of the comparison, we trained all ablation methods under four compression cases, namely: (40-60,60), (50-70,70), (60-80,80), and (70-90,90). The P, R and F of detection results are listed in Table 1. It can be seen from the results that FF-Net is superior to other ablation methods in every compression case. Moreover, we find that when the QF gets smaller, the performance of all detection methods will decline, and the advantage of FF-Net becomes more obvious. $QF1_S$=90 is a slight compression, while $QF1_D$=70 is relatively a strong compression. For the single compression with $QF1_S$=90 and the double compressions with $QF1_D$-$QF2_D$=70-90, the networks can easily localize the

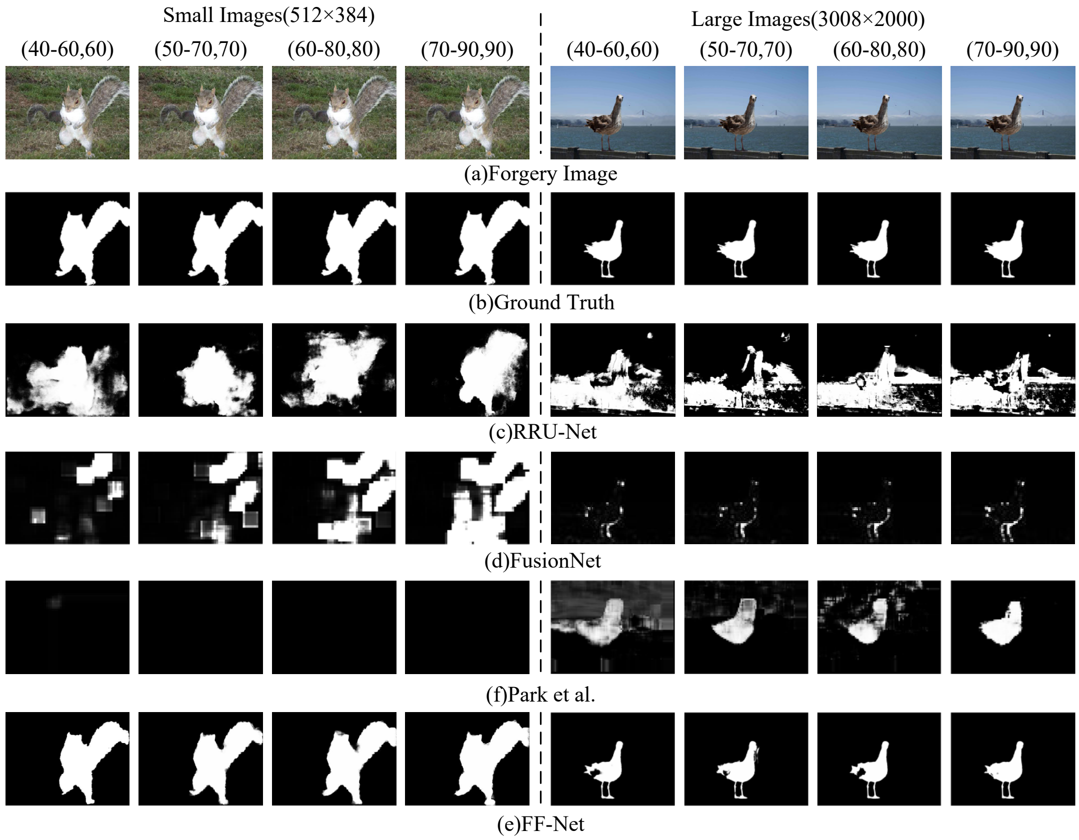


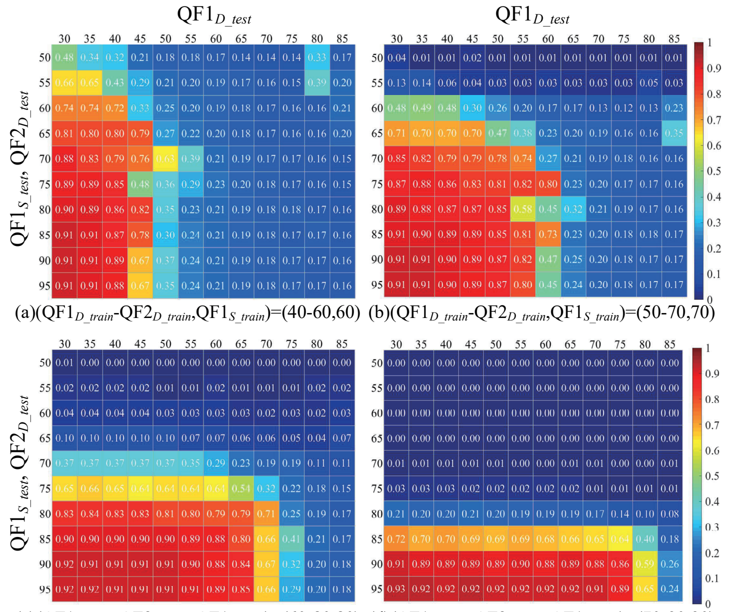
Figure 9: The detection results by FF-Net and other three comparison methods. Detection results of small images are to the left of the dotted line, and detection results of large images are to the right of the dashed line.

tampered regions because the differences of compression fingerprints are obvious. With the decreasing of QF, the differences in compression fingerprints become indistinguishable, decreasing the detection performance. Since the dense connection and the DWT encoder increase the ability of FF-Net to distinguish the differences of compression fingerprints, FF-Net keeps better detection performance than the other methods.

4.4.2. Comparative Experiments

To compare the proposed FF-Net's performance, we chose two CNN-based detection methods proposed for splicing forgery detection and one detection method for double JPEG compressions. Ringed residual U-Net (RRU-Net) Bi et al. (2019) combines the residual propagation with the residual feedback modules to form a ringed residual structure, which improves the utilization of feature maps.

Deep fusion network (FusionNet) Liu and Pun (2020) concentrates on learning the low-level forensic features to detect splicing forgery. Since the existing double JPEG compressions detection methods all rely on DCT coefficient histograms to analyze image patches



(a)(QF1$_{D_train}$-QF2$_{D_train}$,QF1$_{S_train}$)=(40-60,60) (b)(QF1$_{D_train}$-QF2$_{D_train}$,QF1$_{S_train}$)=(50-70,70)

(c)(QF1$_{D_train}$-QF2$_{D_train}$,QF1$_{S_train}$)=(60-80,80) (d)(QF1$_{D_train}$-QF2$_{D_train}$,QF1$_{S_train}$)=(70-90,90)

Figure 10: The generalization capability of FF-Net. The number in each grid represents the metric F of the test result in the compression case.

and then use sliding window for localization, we only chose Park et al. (2018) as a comparison method. The P, R and F of detection results are listed in Table 1, and Figure 9 shows several detection results by different methods. It can be seen from Figure 9 and Table 1, FF-Net outperforms other comparison methods both subjectively and objectively. RRU-Net was ineffective for double JPEG compressions detection because it cannot learn fingerprints from the input images. FusionNet only detected the edges, and the detection results were quite inaccurate. Park et al. (2018) was invalid in detecting small images, and when it detected large images, the detection of details was so weak that the feet of the seabirds were not detected in Figure 9-(h) while FF-Net presented promising results.

4.5. Generalization Capability

For exploring the generalization capability of FF-Net, we used FF-Net, which was trained under one compression case to test various compression cases. The detection results are shown in Figure 10. It can be seen from the results that FF-Net has advanced generalization capability. It can be observed, when QF1$_{D_test}\leq$QF1$_{D_train}$ and QF2$_{D_test}\geq$QF2$_{D_train}$, the detection results are promising. When QF1$_{D_test}>$QF1$_{D_train}$ or QF2$_{D_test}<$QF2$_{D_train}$, the detection capability of the network begins to decline drastically, and fails soon. We guess that when QF1$_{D_test}>$QF1$_{D_train}$, the network may regard the doubly-compressed regions with two slight compressions as a single compression region with single strong compression. On the contrary, when QF2$_{D_test}<$QF2$_{D_train}$, the network may regard the single compression regions with a strong compression as doubly-compressed regions with two slight compressions.

The proposed FF-Net locates the regions with double JPEG compressions based on the compression fingerprint to detect any local forgery that produces different compression fingerprints. To verify the detection capability of FF-Net, we manipulated images by object-removal, copy-move, color-changing, blurring, and object-resizing using Adobe Photoshop. The detection results of various image manipulations are shown in Figure 11. From the results, we can see FF-Net accurately detected and located the tampered regions of each image manipulation.

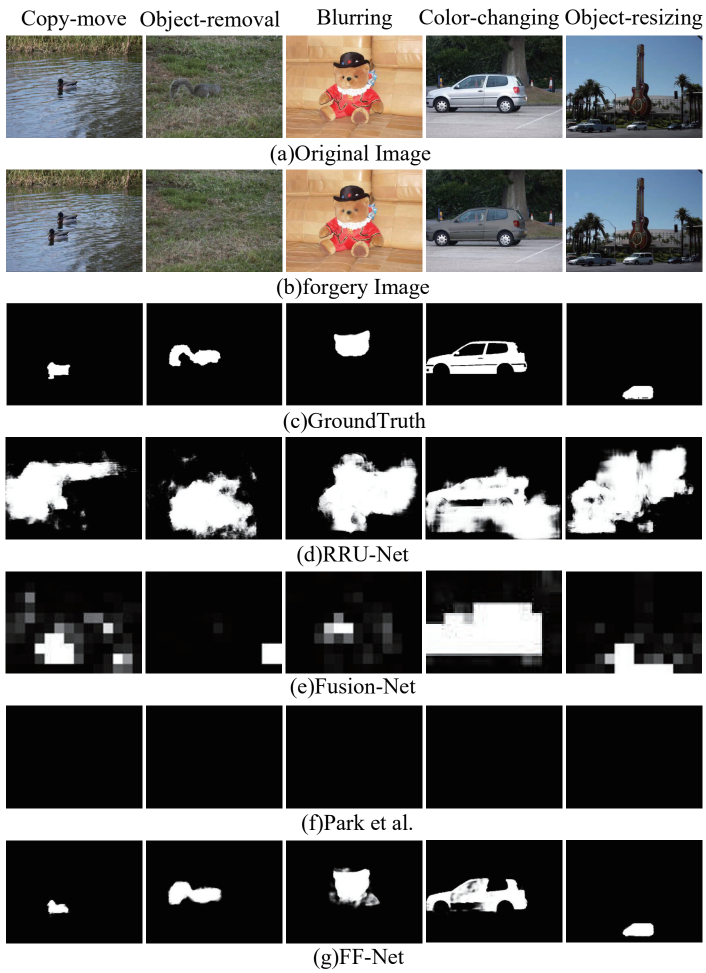


Figure 11: The comparison of detection results in various local forgeries.

5. Conclusion

In this paper, we proposed an end-to-end Feature-Fusion Network (FF-Net) to detect and localize local image forgery by analyzing JPEG compressions. We first studied the differences in compression fingerprints caused by different quantization processes. According to the study of JPEG compression fingerprint, we designed an SRM encoder and a DWT encoder to learn the compression fingerprint. Based on the learned compression fingerprint, the proposed FF-Net can locate the regions with different compression fingerprints at pixel-level accurately. Moreover, the proposed FF-Net has good generalization capability and can detect various image manipulations. In our future works, we will further explore the key issues of double JPEG compressions detection.

Acknowledgments

This work was supported in part by the National Key Research and Development Project under Grant 2019YFE0110800 and Grant 2016YFC1000307-3, in part by the National Natural Science Foundation of China under Grant 62172067 and Grant 61976031, in part by the National Major Scientific Research Instrument Development Project of China under Grant 62027827.

References

W Ahn, S-H Nam, M Son, H-K Lee, and S Choi. End-to-end double jpeg detection with a 3d convolutional network in the dct domain. *Electronics Letters*, 56(2):82–85, 2019.

Irene Amerini, Tiberio Uricchio, Lamberto Ballan, and Roberto Caldelli. Localization of jpeg double compression through multi-domain convolutional neural networks. In *2017 IEEE Conference on computer vision and pattern recognition workshops (CVPRW)*, pages 1865–1871. IEEE, 2017.

Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

Mauro Barni, Luca Bondi, Nicolò Bonettini, Paolo Bestagini, Andrea Costanzo, Marco Maggini, Benedetta Tondi, and Stefano Tubaro. Aligned and non-aligned double jpeg detection using convolutional neural networks. *Journal of Visual Communication and Image Representation*, 49:153–163, 2017.

Xiuli Bi, Yang Wei, Bin Xiao, and Weisheng Li. Rru-net: The ringed residual u-net for image splicing forgery detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

Tiziano Bianchi and Alessandro Piva. Image forgery localization via block-grained analysis of jpeg artifacts. *IEEE Transactions on Information Forensics and Security*, 7(3):1003–1017, 2012.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011*, pages 97–104. IEEE, 2011.

Hany Farid. Exposing digital forgeries from jpeg ghosts. *IEEE transactions on information forensics and security*, 4(1):154–160, 2009.

Jessica Fridrich and Jan Kodovsky. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, 2012.

Dongdong Fu, Yun Q Shi, and Wei Su. A generalized benford's law for jpeg coefficients and its applications in image forensics. In *Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, page 65051L. International Society for Optics and Photonics, 2007.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018.

Bo Liu and Chi-Man Pun. Exposing splicing forgery in realistic scenes using deep fusion network. *Information Sciences*, 526:133–150, 2020.

Jan Lukáš and Jessica Fridrich. Estimation of primary quantization matrix in double compressed jpeg images. In *Proc. Digital forensic research workshop*, pages 5–8, 2003.

Jinseok Park, Donghyeon Cho, Wonhyuk Ahn, and Heung-Kyu Lee. Double jpeg detection in mixed jpeg quality factors using deep convolutional neural network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 636–652, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Gerald Schaefer and Michal Stich. Ucid: An uncompressed color image database. In *Storage and Retrieval Methods and Applications for Multimedia 2004*, volume 5307, pages 472–480. International Society for Optics and Photonics, 2003.

Qing Wang and Rong Zhang. Double jpeg compression forensics based on a convolutional neural network. *EURASIP Journal on Information Security*, 2016(1):1–12, 2016.