RRU-Net: The Ringed Residual U-Net for Image Splicing Forgery Detection

Xiuli Bi†, Yang Wei†, Bin Xiao*, Weisheng Li
School of Computer Science & Technology, Chongqing University of Posts and Telecommunication
{bixl}@cqupt.edu.cn, {yalesaleng}@outlook.com, {xiaobin, liws}@cqupt.edu.cn

Abstract

Detecting a splicing forgery image and then locating the forgery regions is a challenging task. Some traditional feature extraction methods and convolutional neural network (CNN)-based detection methods have been proposed to finish this task by exploring the differences of image attributes between the un-tampered and tampered regions in an image. However, the performance of the existing detection methods is unsatisfactory. In this paper, we propose a ringed residual U-Net (RRU-Net) for image splicing forgery detection. The proposed RRU-Net is an end-to-end image essence attribute segmentation network, which is independent of human visual system, it can accomplish the forgery detection without any preprocessing and post-processing. The core idea of the RRU-Net is to strengthen the learning way of CNN, which is inspired by the recall and the consolidation mechanism of the human brain and implemented by the propagation and the feedback process of the residual in CNN. The residual propagation recalls the input feature information to solve the gradient degradation problem in the deeper network; the residual feedback consolidates the input feature information to make the differences of image attributes between the un-tampered and tampered regions be more obvious. Experimental results show that the proposed detection method can achieve a promising result compared with the state-of-the-art splicing forgery detection methods.

1. Introduction

Recently, the widespread availability of image editing software makes it extremely easy to edit or even change the digital image content, which is becoming a fearful problem. Struggling to the public trust in photographs, in this paper, our research is specifically focused on the image splicing forgery detection. The splicing forgery copies parts of one image and then pastes into another image to merge a new image as shown in Fig. 1.(a). Because the tampered regions come from other images, the differences of image attributes between the un-tampered and tampered regions exist, such as lighting, shadow, sensor noise, camera reflection and so on, which can be utilized to identify an image suspected of being tampered with and to locate the tampered regions in the forgery image. The existing splicing forgery detection methods have tried to make use of some feature extraction methods for exploring the differences of image attributes. According to the feature extraction methods used in the existing splicing forgery detection methods, they can be mainly classified into two classes: traditional feature extraction-based detection methods and convolutional neural network (CNN)-based detection methods.

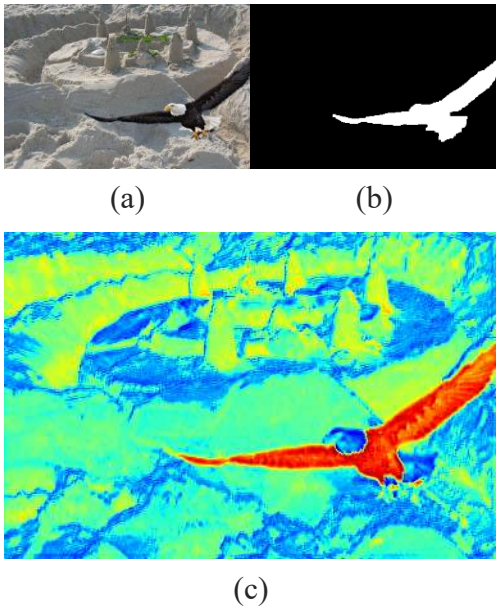


Figure 1. An enhanced input features by the residual feedback in the proposed RRU-Net. (a) The splicing forgery image; (b) The ground-truth image; (c) The global response of the enhanced input of the first building block in the proposed RRU-Net.

For traditional feature extraction-based methods, they can be broadly categorized into four types: detection methods based on the image essence attribute [3, 25, 34], detection methods based on the imaging device attribute [6, 8, 13,

† denotes equal contribution
* denotes corresponding author

17], detection methods based on the image compression attribute [12,14,30], and detection methods based on the hash techniques [23, 26, 29, 35]. These detection methods generally focus on one specified image attribute, thus they have the following limitations on real-world tasks: a) The detection methods based on the image essence attribute may fail if some hidden processes (such as the overall fuzzy operation) are performed after the process of splicing forgery; b) If the device noise intensity of an image is weak, the detection methods based on the imaging device attribute may fail; c) The detection methods based on the image compression attribute can only detect the image saved by JPEG format; d) The detection methods based on the hash techniques rely on the hash of the original un-tampered image, so they cannot be strictly regarded as the blind type of forgery detection method.

In recent years, convolutional neural network (CNN) has achieved a great success in the research area of computer vision. The feature extraction and mapping of CNN give some researchers an insight that CNN can also be adapted to accomplish the image splicing forgery detection, CNN is originally used to judge whether the image has been tampered in [20], but it cannot locate the tampered regions. In [33], the authors try to locate the tampered regions by CNNs, but the detected regions just can be shown by the inaccurate rough areas that are made up of some square white blocks. These two detection methods based on CNN are just the preliminary attempts, their results are not ideal. For improving the detected tampered regions, the detection methods [1, 27] use the non-overlapping image patch as the input of CNNs. However, when an image patch totally comes from the tampered regions, this image patch will be judged un-tampered label. In [15], the authors utilize the bigger image patch to reveal the image attributes of the tampered regions, however, the detection method may fail if the forgery image is small. For the existing CNN-based detection methods, since they use the image patch as the input of the network, the contextual spatial information is lost, which easily causes incorrect prediction. Moreover, when the network architecture is deeper, the gradient degradation problem will appear and the discrimination of features will become weaker, which will lead to the splicing forgery detection more difficult or even fail.

For overcoming the drawbacks of traditional feature extraction-based methods, meanwhile, further solving the problems of current CNN-based detection methods, a ringed residual U-Net (RRU-Net) is proposed in this paper. RRU-Net is an end-to-end image essence attribute segmentation network, which is independent of human visual system, it can directly locate the forgery regions without any preprocessing and post-processing. Furthermore, RRU-Net can effectively decrease incorrect prediction since it makes better use of the contextual spatial information in a image.

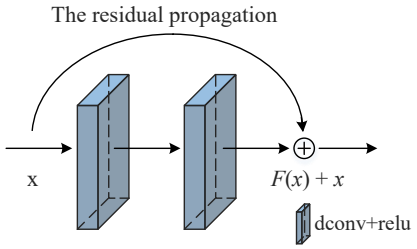


Figure 2. Residual propagation in a building block. The 'dconv' represents a dilated convolution operation, and the 'relu' represents a nonlinear operation.

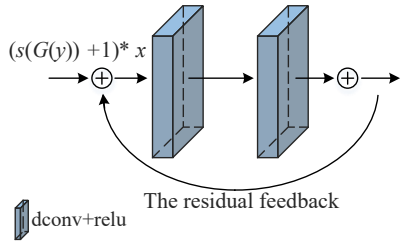


Figure 3. Residual feedback in a building block.

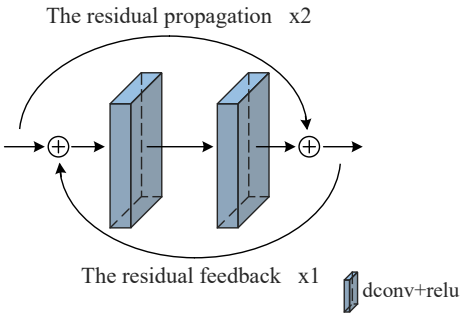


Figure 4. Ringed residual structure. The 'x1' represents this operation is performed only once. The 'x2' represents this operation is performed twice.

And most of all, the ringed residual structure in RRU-Net can strengthen the learning way of CNN and simultaneously prevent the gradient degradation problem of deeper network, which ensure the discrimination of image essence attribute features be more obvious while the features are extracted among layers of network.

2. Related Work

U-Net: U-Net has been proposed by Olaf Ronneberger *et al*. [21] in 2015. U-Net is a great success in the neuronal structure segmentation, its framework is pathbreaking since features are propagated among layers. In U-Net, the context information is captured by a contracting path (successive layers), the output feature is upsampled and then combined with the high-resolution features propagated by a symmetric expanding path, which reduces the loss of detail information and enables precise location. Therefore, some image segmentation methods [4,10] based on U-Net have been

subsequently proposed. In fact, the image splicing forgery detection can be viewed as a complicated image segmentation task independent of the human visual system. We need to segment out the tampered regions in a image that cannot be distinguished by the human eyes at most time. The only way to locate the tampered regions depends on the differences of image essence attributes, which can be discovered by extracting the discriminative features. Though the U-Net can extract some relatively shallow discriminative features among layers of network, only the two sides of the U-Net structure are interacting, which is not enough for confirming the tampered regions. Besides, the gradient degradation problem [7] will appear when network architecture is deeper.

ResNet: ResNet has been proposed by Kaiming He *et al*. [7] in 2015, it has won the championship in the classification task of ImageNet match. In ResNet, the residual mapping is defined as Eq.(1).

$$y = F(x) + x \tag{1}$$

For a few stacked layers, x represents the input, y is the output, the operation $F(x) + x$ is performed by a shortcut connection and element-wise addition. The residual mapping is proposed to solve the gradient degradation problem in deeper networks. For the splicing forgery detection, the gradient degradation problem will cause another extra serious problem. The discrimination of image essence attribute features will be weaker through the direct multilayer structure, which makes the differences of image essence attributes become hard to discover. For solving the gradient degradation problem and simultaneously strengthening the learning way of CNN, the residual mapping should be utilized more efficiently.

3. The Ringed Residual U-Net (RRU-Net)

3.1. Residual Propagation

According to the discussion above, the differences of image essence attributes are the significant basis for detecting image splicing forgery, however, the gradient degradation problem will destroy the basis when the network architecture gets deeper. For solving the gradient degradation problem, we add the residual propagation to each stacked layers. A building block is shown in Fig. 2, which consists of two convolutional (dilated convolution [31], dconv) layers and residual propagation. The output of the building block is defined as:

$$y_f = F(x, \{W_i\}) + W_s * x, \tag{2}$$

where, x and y_f are the input and output of the building block, W_i represents the weights of layer i, the function $F(x, \{W_i\})$ represents the residual mapping to be learned. For the example in Fig. 2 that has two convolutional layers, $F = W_2\sigma(W_1 * x)$ in which σ denotes ReLU [19] and the biases are omitted for simplifying notations. The linear projection W_s is used to change the dimension of x to match the dimension of $F(x, \{W_i\})$. The operation $F + W_s * x$ is performed by a shortcut connection and element-wise addition.

The residual propagation looks like the recall mechanism of the human brain. We may forget the previous knowledge when we learn several more new knowledge, so we need the recall mechanism to help us arouse those previous fuzzy memories.

3.2. Residual Feedback

It is obvious that, in splicing forgery detection, if the differences of image essence attributes between the un-tampered and tampered regions can be further strengthened, the performance of the detection can be further improved. In [36], the proposed method superposes the additional difference of noise attribute by passing the forgery image through an SRM filter layer to enhance detection results. The SRM filter layer has a certain effect, however, it is a manual choosing method and can only for the RGB image forgery detection. Moreover, when the un-tampered and tampered regions come from the cameras with the same brand and model, the SRM filter layer will reduce effectiveness sharply, since they have same noise attribute. For further strengthening the differences of image essence attributes, the residual feedback is proposed, which is an automatic learning method and not just focus on one or several specific image attributes. Furthermore, we design a simple and effective attention mechanism, which take advantage of ideas of Hu *et al*. [9], and then we add it on the residual feedback to pay more attention to the discriminative features of input information. In this attention mechanism, we opt to employ a simple gating mechanism with a sigmoid activation function to learn a nonlinear interaction between discriminative feature channels and avoid diffusion of feature information, and then we superpose the response values obtained by sigmoid activation on input information to amplify differences of image essence attributes between the un-tampered and tampered regions. The residual feedback in a building block is shown Fig. 3 and is defined as Eq.(3),

$$y_b = (s(G(y_f)) + 1) * x \tag{3}$$

where, x is the input, y_f is the output of residual propagation defined in Eq.(2), y_b is the enhanced input. The function G is a linear projection, which is used to change the dimensions of y_f. The function s is a sigmoid activation function.

In contrast to the recall mechanism imitated by the residual propagation, the residual feedback seems to act as the consolidation mechanism of the human brain, we need to

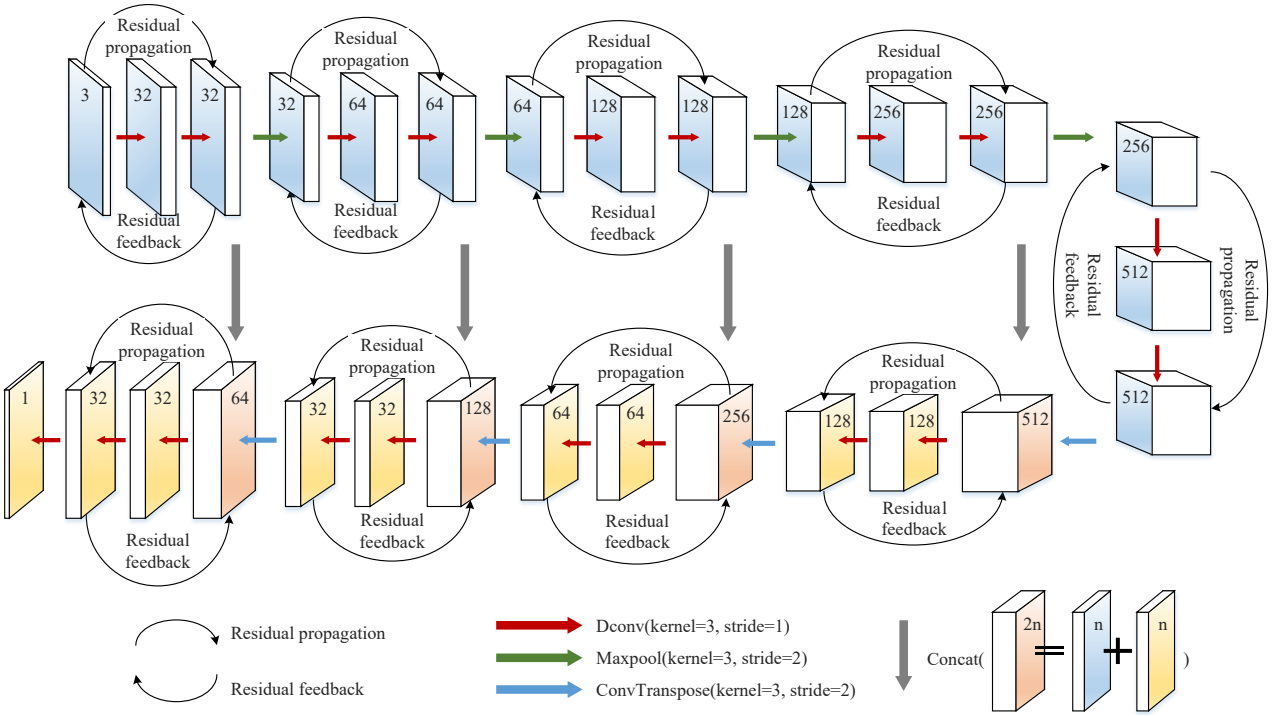


Figure 5. The network architecture of RRU-Net. The number on the box represents the number of features.

consolidate the knowledge already learned by us to obtain the new feature comprehensionp. The residual feedback can amplify the differences of image essence attributes between the un-tampered and tampered regions in the input, as shown in Fig. 1.(c), the tampered region 'eagle' is amplified to global maximal response values by the residual feedback. Furthermore, it also has two far-reaching effects: (1) the strengthening of the discriminative features can simultaneously be viewed as the repression of the negative label features; (2) the convergence rate of network in the training process is more fast.

3.3. Ringed Residual Structure and Network Architectures

The proposed ringed residual structure that combines the residual propagation and the residual feedback is shown in Fig. 4. The residual propagation is just like the recall mechanism of the human brain, which recalls the input feature information to solve the degradation problem in the deeper network; the residual feedback consolidates the input feature information to make the differences of image essence attributes between the un-tampered and tampered regions be amplified. To sum up, the ringed residual structure guarantees the discrimination of image essence attribute features be more obvious while the features are extracted among layers of network, which can achieve better and stable detection performance than traditional feature extraction-based detection methods and existing CNN-based detection methods. The network architecture of RRU-Net is shown in Fig. 5, it is an end-to-end image essence attribute segmentation network, and can directly detect the splicing forgery without any preprocessing and post-processing.

4. Evaluation Experiment and Comparative Analysis

For evaluating the performance of the proposed RRU-Net, we carry out various experiments in terms of effectiveness and robustness. Meanwhile, the proposed method is compared with some other image splicing forgery detection methods under different cases.

Experimental Datasets: We chose two public datasets for evaluation, i.e., CASIA [24] and COLUMB [8]. On CASIA, the splicing forgery regions are objects, which are small and elaborate. On COLUMB, the splicing forgery regions are some simple, large, and meaningless regions. For training RRU-Net better, we resize the size of images in the training and validation sets to 384×256, and then we perform data augmentation with random Gaussian noise, JPEG compression and random overturn, which quadruples the capacity of the two datasets. All of experimental datas is listed in Tab. 1, on CASIA, we randomly select 715 sets of images that contain the original images and forgery images as the training set, 35 sets of images as the validation set, and then 100 sets of images as the testing set. Similarly, on COLUMB, 125 sets of images are chosen as the training

Sets	Cases	Parameters	Range	Step	**CASIA** [24]	**COLUMB** [8]
Training Set	Augmented Splicing	—	—	—	3575	625
	Original Image	—	—	—	715	125
Validation Set	Plain Splicing	—	—	—	175	50
	Original Image	—	—	—	35	10
Testing Set	Plain Splicing	—	—	—	100	44
	Original Image	—	—	—	100	44
	JPEG Compression	Quality Factor	$50 \sim 90$	10	500	220
	Noise Corruption	Variance	$0.002 \sim 0.01$	0.002	500	220

Table 1. The generation of training, validation and testing sets based on CASIA [24] and COLUMB [8].

set, 10 sets of images as the validation set, and then 44 sets of images as the testing set. The Augmented Splicing represents the combination of augmented datasets (2860 images) and the Plain Splicing datasets (715 images), moreover, for comparing and analyzing the robustness of the image splicing forgery detection method, JPEG compression and noise corruption are applied to the forgery datasets to create various attack cases.

- JPEG Compression: after a splicing forgery image is created, the splicing forgery image will be saved in JPEG format with different compression quality factor.
- Noise Corruption: after a splicing forgery image is created, white Gaussian noise with mean value 0 and different variances will be added to the splicing forgery image.

As listed in Tab. 1, 7038 images are used in total in the following experiments. For the fair comparison, we convert all experimental images from TIFF format to JPEG format with quality factor 100%, since the detection methods based on the compression property only can detect the image in JPEG format.

Evaluation Metrics: For image splicing forgery detection, at the pixel level, the significant evaluation is the performance of locating the tampered regions. The evaluation metrics are the number of correctly detected tampered pixels (*TP*), the number of incorrectly detected tampered pixels (*FP*), and the number of incorrectly detected un-tampered pixels (*FN*). In the following experiments, we use *Precision*, *Recall*, and F-measure to evaluate the performance of the proposed splicing forgery detection methods in pixel level. *Precision* is defined in Eq.(4), which denotes the probability that the detected regions are the truly tampered regions in the ground-truth image. *Recall* is the probability that the tampered regions in the ground-truth image are correctly detected, which is defined in Eq.(5). F-measure combines *Precision* and *Recall* to one measure to synthetically evaluate the performance of detection method, it is formulated in Eq.(6). In the experiments, the *Precision*, *Recall*, and F-measure are the mean values of the testing set. Moreover, the performance of distinguishing the un-tampered image and tampered image is another significant evaluation, which means the un-tampered image should not be detected as the tampered image and vise versa. For further demonstrating the detection effects of the proposed RRU-Net, at the image level, we evaluate the detection results by using the accuracy rate.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

Compared Detection Methods: For comparing the performance of the proposed RRU-Net, we choose three traditional feature extraction-based detection methods, two CNN-based detection methods and two semantic segmentation methods, they are DCT [30], CFA [5], NOI [18], DF-Net [15] C2R-Net [27], FCN [16] and DeepLab v3 [2]. The DCT is an inconsistent detection method for JPEG DCT coefficient histogram. In CFA, the interference in the color filter array(CFA) interpolation pattern is modeled as a mixture of Gaussian distributions to detect the tampered regions. The NOI detects the splicing regions by using the wavelet Filtering to extract the local image noise variance modeling. Compared detection methods (DCT, CFA, and NOI) have been mainly implemented by Zampoglou, Papadopoulos and Kompatsiaris [32]. We choose this version algorithms to estimate in our experiments. The C2R-Net and DF-Net detect the tampered regions by using CNN, they use image patch as the input of CNN, the algorithm codes are provided by authors. The DF-Net is inefficient on CASIA, since this method uses $64 * 64$ image patch as the input and the images on CASIA are small, we do not present its result on this dataset. The FCN and DeepLab v3 are classic and

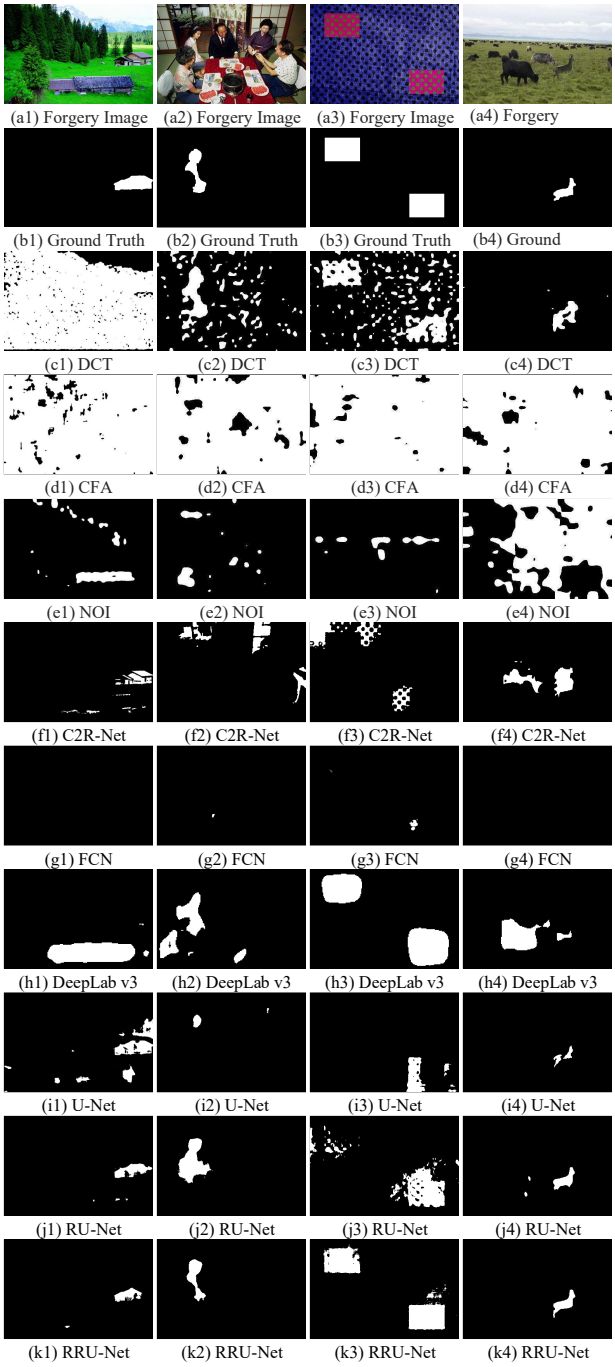


Figure 6. The splicing forgery detection result by RRU-Net and other eight comparative detection methods.

effective detection methods for image semantic segmentation task, both of them have achieved better detection performance. Moreover, we utilize two detection methods the U-Net and the residual U-Net (RU-Net) to further evaluate validity of the residual residual structure in RRU-Net. Implementation of U-Net refers to its original structure in [21], and the structure of RU-Net gets rid of the residual feedback in RRU-Net.

Implementation Detail: The RRU-Net and the compared detection methods are run on a computer with Intel Xeon E5-2603 v4 CPU and NVIDIA GTX TITAN X GPU. The parameters of all compared detection methods are set according to their best performances. The U-Net, RU-Net, and RRU-Net are implemented by PyTorch. In the training process of RRU-Net, we utilize random value to initial parameters and use stochastic gradient descent with a batch size of 10 samples, the momentum is 0.9, the weight decay is 0.0005 and the initial learning rate is 0.1. The group normalization (GN) [28] is used to normalize scattered data in high dimensional space since a batch size of 10 samples is insufficient to support for batch normalization (BN) [11]. The cross-entropy introduced in [22] is used as a loss function.

4.1. Detection at Pixel Level

4.1.1 Detection Results under Plain Splicing Forgery

In this subsection, the proposed RRU-Net and other compared detection methods are estimated under the case of plain splicing forgery, the detected results of four examples are shown in Fig. 6. From a subjective perspective, it is clear that the performance of the RRU-Net is better than the other eight detection methods. For more objective and fair comparisons, we calculate the averages of *Precision*, *Recall*, and F-measure of detection results on both of the two datasets, which is listed in Tab. 2. It can be seen that RRU-Net is better than other nine detection methods in *Precision*, *Recall* and F-measure. Although the *Recall* of RRU-Net is a little worse than the DCT [30] and DeepLab v3 [2], from the subjective perspective, we can find that the DCT almost loses effectiveness and the detection effect of the DeepLab v3 is far worse than the RRU-Net.

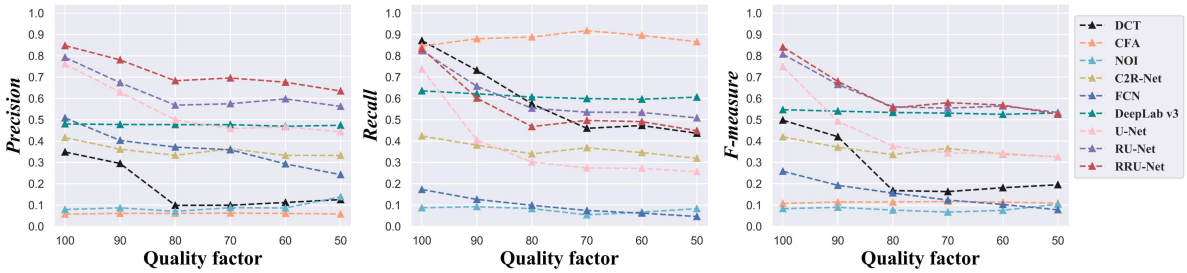
4.1.2 Detection Results under Various Attacks

For further verifying the effectiveness and robustness of the proposed detection method, we also evaluate the performance of the detection methods under various attacks, including JPEG compression and noise corruption.

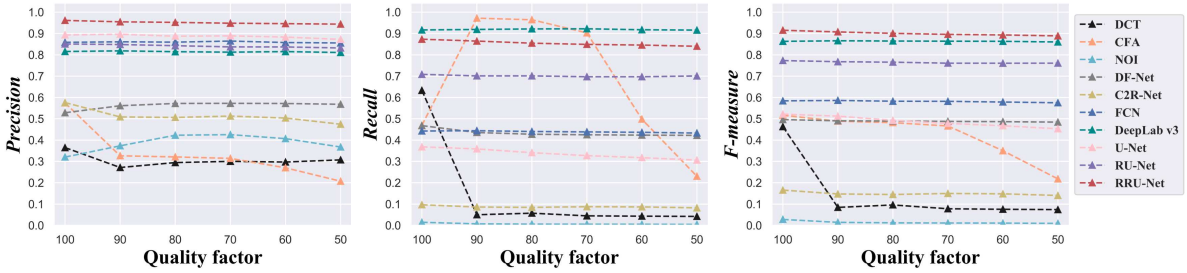
Experimental Results under JPEG Compression Attack. The comparative experiment results under JPEG compression attack are shown in Fig. 7. In Fig. 7, three columns indicate *Precision*, *Recall*, and F-measure of the comparative experiment results respectively. The first and second rows are the experiment results under different quality factor of JPEG compression on CASIA and COLUMB respectively. From this Fig. 7, it can be easily observed that the RRU-Net is better than the other eight detection methods in *Precision* and F-measure on CASIA, and it is slightly lower than the DCT, the CFA, the DeepLab v3 and the

Methods	CASIA [24]			COLUMB [8]		
	Precision	*Recall*	*F*-measure	*Precision*	*Recall*	*F*-measure
DCT [30]	0.349	**0.871**	0.498	0.365	0.633	0.463
CFA [5]	0.057	0.846	0.108	0.574	0.469	0.517
NOI [18]	0.079	0.088	0.083	0.321	0.015	0.028
DF-Net [15]	-	-	-	0.528	0.468	0.496
C2R-Net [27]	0.417	0.424	0.420	0.576	0.097	0.166
FCN [16]	0.509	0.173	0.259	0.859	0.443	0.584
DeepLab v3 [2]	0.481	0.636	0.547	0.815	**0.917**	0.863
U-Net [21]	0.761	0.737	0.749	0.893	0.369	0.522
RU-Net	0.783	0.814	0.798	0.851	0.708	0.773
RRU-Net	***0.848***	*0.834*	***0.841***	***0.961***	*0.873*	***0.915***

Table 2. Detection results under the plain splicing forgery. The sign '-' denotes that the result is not available in the Table.



(a1) JPEG-compression (a2) JPEG-compression (a3) JPEG-compression

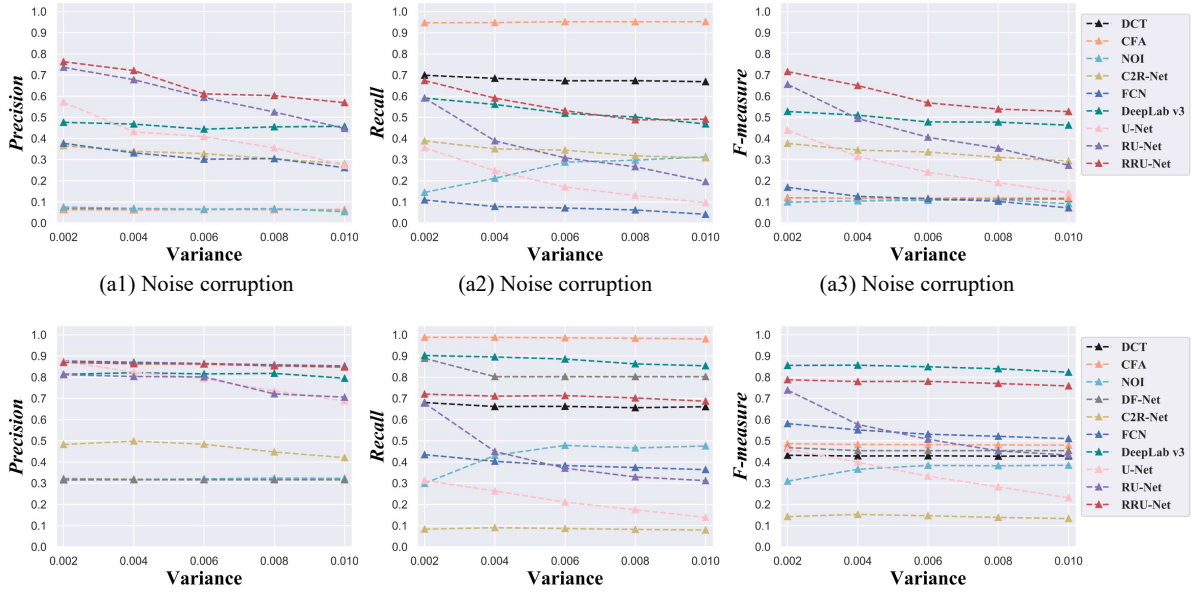


(b1) JPEG-compression (b2) JPEG-compression (b3) JPEG-compression

Figure 7. Comparison results under JPEG compression attacks. The three columns represent the *Precision*, *Recall*, and *F*-measure. (a1) - (a3) represent the experiment results on CASIA; (b1) - (b3) represent the experiment results on COLUMB.

RU-Net in *Recall*. Similarly, the RRU-Net is better than the other nine detection methods in *Precision* and *F*-measure on COLUMB, and it is only slightly lower than the DeepLab v3 in *Recall*. The *Recall* of the RRU-Net is worse than the CFA and the DCT, the reason is both of them detect almost the whole image as the tampered regions. The performance and robustness of the U-Net without the residual propagation and the residual feedback are far lower than the RU-Net and the RRU-Net. It can be found through the experiments that the detection results of the RRU-Net are preferable than the other detection methods and have high robustness under JPEG compression attack on the two datasets.

Experimental Results under Noise Corruption Attack. The comparative experiment results under noise (Gaussian-distributed additive noise) corruption attack are shown in Fig. 8. Fig. 8.(a1-a3) and Fig. 8.(b1-b3) represent the experimental results of noise corruption with different variances (mean of the random distribution is 0) on CASIA and COLUMB respectively. On CASIA, the *Precision* and *F*-measure of the RRU-Net are better than the other eight detection methods. On COLUMB, the *Precision* of the RRU-Net is better than the other nine detection methods, and the *F*-measure of the RRU-Net is slightly lower than the DeepLab v3. The robustness of the RU-Net without the residual feedback is weak under noise corruption attack



(a1) Noise corruption (a2) Noise corruption (a3) Noise corruption

(b1) Noise corruption (b2) Noise corruption (b3) Noise corruption

Figure 8. Comparison results under noise corruption attack. The three columns represent the *Precision*, *Recall*, and F-measure. (a1) - (a3) represent the experiment results on CASIA; (b1) - (b3) represent the experiment results on COLUMB.

Methods	**Accuracy**
DCT [30]	52.78%
CFA [5]	59.63%
NOI [18]	63.89%
DF-Net [15]	15.28%
C2R-Net [27]	46.53%
FCN [15]	68.4%
DeepLab v3 [27]	69.4%
U-Net [21]	67.2%
RU-Net	72.6%
RRU-Net	***76%***

Table 3. The detection results of the RRU-Net and the other nine detection methods at image level.

on both of the two datasets. From the above analysis, the RRU-Net shows better and stable performance under noise corruption attack on the two datasets.

4.2. Detection at Image Level

For comparing the performance of the RRU-Net and other detection methods at the image level, we carry out an experiment to identify the un-tampered image and the tampered image. 144 plain splicing forgery images and 144 original images are selected as testing images from CASIA and COLUMB. The accuracy of the detection methods is listed in Tab. 3. It is clear that the accuracy of RRU-Net is better than other nine detection methods, which proves RRU-Net not only can locate the tampered regions in splicing forgery images but also can judge whether an image has or has not been tampered.

5. Conclusion

In this paper, we propose a ringed residual U-Net (RRU-Net) for image splicing forgery detection, which is an end-to-end image essence property segmentation network and can achieve the forgery detection without any preprocessing and post-processing. Inspiring by the recall and consolidation mechanisms of the human brain, the proposed RRU-Net strengthens the learning way of CNN by the propagation and feedback process of the residual. Simultaneously, we also prove the validity of the ringed residual structure in RRU-Net from theoretical analysis and experimental comparison. We will further explore and visualize the latent discriminative feature between tampered and un-tampered regions to explain the key issues of image splicing forgery detection in our future works.

Acknowledgment

This work was partly supported by the National Natural Science Foundation of China (61572092, U1713213 and 61806032), the National Science & Technology Major Project (2016YFC1000307-3), the Chongqing Research Program of Application Foundation & Advanced Technology (cstc2018jcyjAX0117) and the Scientific & Technological Key Research Program of Chongqing Municipal Education Commission (KJZD-K201800601).

References

[1] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. Manjunath. Exploiting spatial structure for localizing manipulated image regions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4970–4979, 2017.

[2] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[3] W. Chen, Y. Q. Shi, and W. Su. Image splicing detection using 2-d phase congruency and statistical moments of characteristic function. In *Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, page 65050R. International Society for Optics and Photonics, 2007.

[4] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.

[5] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva. Image forgery localization via fine-grained analysis of cfa artifacts. *IEEE Transactions on Information Forensics and Security*, 7(5):1566–1577, 2012.

[6] H. Gou, A. Swaminathan, and M. Wu. Noise features for image tampering detection and steganalysis. In *ICIP (6)*, pages 97–100. Citeseer, 2007.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[8] Y.-F. Hsu and S.-F. Chang. Detecting image splicing using geometry invariants and camera characteristics consistency. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 549–552. IEEE, 2006.

[9] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[10] V. Iglovikov and A. Shvets. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*, 2018.

[11] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[12] M. K. Johnson and H. Farid. Exposing digital forgeries in complex lighting environments. *IEEE Transactions on Information Forensics and Security*, 2(3):450–461, 2007.

[13] M. K. Johnson and H. Farid. Exposing digital forgeries through specular highlights on the eye. In *International Workshop on Information Hiding*, pages 311–325. Springer, 2007.

[14] Z. Lin, J. He, X. Tang, and C.-K. Tang. Fast, automatic and fine-grained tampered jpeg image detection via dct coefficient analysis. *Pattern Recognition*, 42(11):2492–2501, 2009.

[15] B. Liu and C.-M. Pun. Deep fusion network for splicing forgery localization. In *ECCV workshop*, page 15, 2018.

[16] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[17] B. Mahdian and S. Saic. Detection of resampling supplemented with noise inconsistencies analysis for image forensics. In *Computational Sciences and Its Applications, 2008. ICCSA'08. International Conference on*, pages 546–556. IEEE, 2008.

[18] B. Mahdian and S. Saic. Using noise inconsistencies for blind image forensics. *Image and Vision Computing*, 27(10):1497–1503, 2009.

[19] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

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

[21] O. Ronneberger, P. Fischer, and T. 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.

[22] J. Shore and R. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on information theory*, 26(1):26–37, 1980.

[23] Z. Tang, X. Zhang, X. Li, and S. Zhang. Robust image hashing with ring partition and invariant vector distance. *IEEE Trans. Information Forensics and Security*, 11(1):200–214, 2016.

[24] C. v2.0. http://forensics.idealtest.org/casiav2/. 2009.

[25] W. Wang, J. Dong, and T. Tan. Effective image splicing detection based on image chroma. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1257–1260. IEEE, 2009.

[26] X. Wang, K. Pang, X. Zhou, Y. Zhou, L. Li, and J. Xue. A visual model-based perceptual image hash for content authentication. *IEEE Transactions on Information Forensics and Security*, 10(7):1336–1349, 2015.

[27] Y. Wei, X. Bi, and B. Xiao. C2r net: The coarse to refined network for image forgery detection. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 1656–1659. IEEE, 2018.

[28] Y. Wu and K. He. Group normalization. *arXiv preprint arXiv:1803.08494*, 2018.

[29] C.-P. Yan, C.-M. Pun, and X.-C. Yuan. Quaternion-based image hashing for adaptive tampering localization. *IEEE Transactions on Information Forensics and Security*, 11(12):2664–2677, 2016.

[30] S. Ye, Q. Sun, and E.-C. Chang. Detecting digital image forgeries by measuring inconsistencies of blocking artifact. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 12–15. IEEE, 2007.

[31] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[32] M. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris. Large-scale evaluation of splicing localization algorithms for web images. *Multimedia Tools and Applications*, 76(4):4801–4834, 2017.

[33] Y. Zhang, J. Goh, L. L. Win, and V. L. Thing. Image region forgery detection: A deep learning approach. In *SG-CRC*, pages 1–11, 2016.

[34] X. Zhao, J. Li, S. Li, and S. Wang. Detecting digital image splicing in chroma spaces. In *International Workshop on Digital Watermarking*, pages 12–22. Springer, 2010.

[35] Y. Zhao, S. Wang, X. Zhang, and H. Yao. Robust hashing for image authentication using zernike moments and local features. *IEEE transactions on information forensics and security*, 8(1):55–63, 2013.

[36] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis. Learning rich features for image manipulation detection. *arXiv preprint arXiv:1805.04953*, 2018.