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

Master's degree in Artificial Intelligence and Robotics

SAPIENZA
UNIVERSITÀ DI ROMA

Authors:
Denise Landini
Alessandro Lambertini

Course: Vision and Perception
Professor: Irene Amerini

Academic Year: 2020-2021

Introduction
Why RRU-Net ?

Goal: image splicing forgery detection.

RRU-Net is an end-to-end image essence attribute segmentation network.

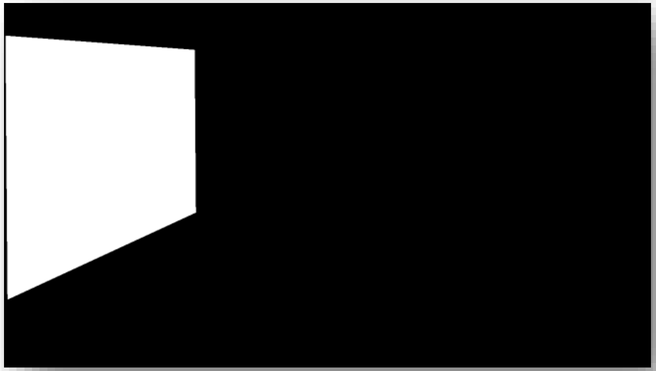
Characteristics:

- It is indipendent of human visual system
- It can locate the forger regions without preprocessing and post-processing
- It decreases incorrect prediction becuase it uses in a better way the contextual spatial information in an image
- It guarantess that the discrimination of image essence attribute features be more evident while the features are extracted among the layers' network. This can obtain better and stable performance than the traditional feature extraction-based detection methods and existing CNN-based detection method.

Dataset
Structure

The **dataset** is composed by 184 images.
In each image we have to identify what is the forgery object/person/detail.

Input image



Output image

In this example, we can see that the image on the television is the forgery object.

Architecture

Residual Propagation and Residual Feedback

The **RRU structure** combines:

RESIDUAL PROPAGATION

It recalls the input feature information, to solve the gradient degradation problem.

Architecture:

- It is added in each stacked layers
- It has 2 convolutional layers
- It uses ReLU as activation function.

RESIDUAL FEEDBACK

It is an automatic learning method that consolidate the input feature information to make the differences of image essence attributes between the untampered and tampered regions be amplified.

Architecture:

- It uses a sigmoid activation function in order to learn a nonlinear interaction between discriminate feature channels and avoid diffusion of feature information.

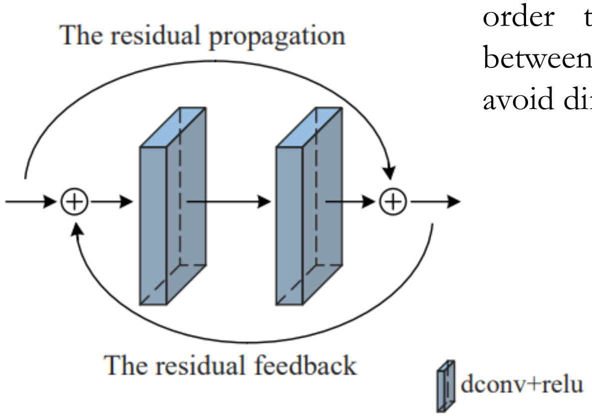


Figure: Representation of residual propagation and residual feedback.

Architecture

Complete structure of RRU-Net

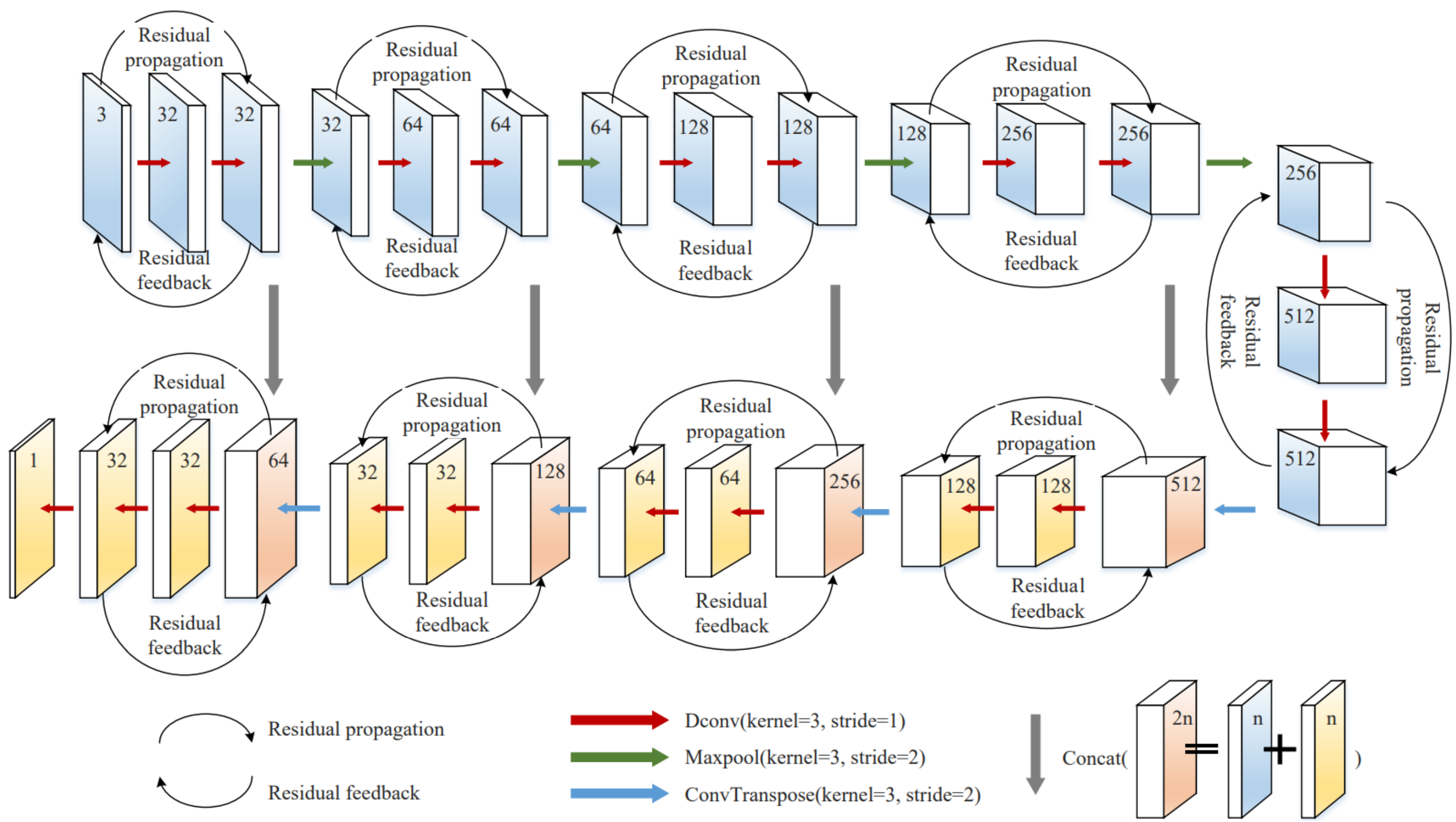


Figure: RRU-Net architecture. The numbers on the box represent the numbers of features.

Evaluation of RRU-Net

Metrics used

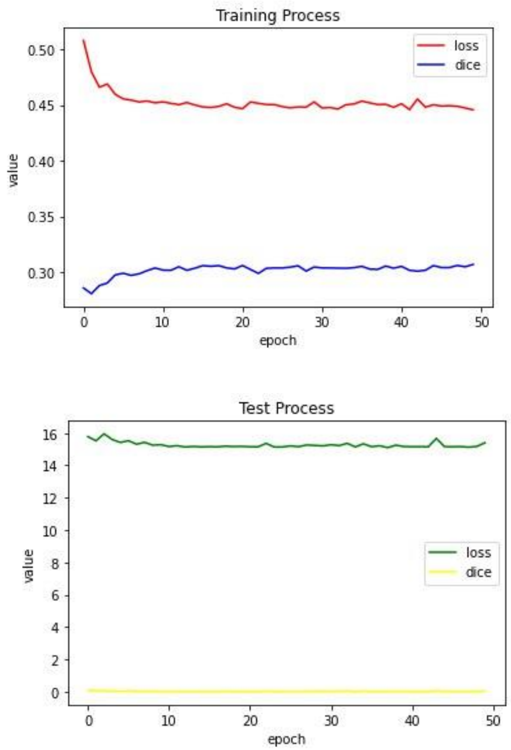


Figure 1: Values of dice and loss function for training and test processes after 50 epochs.

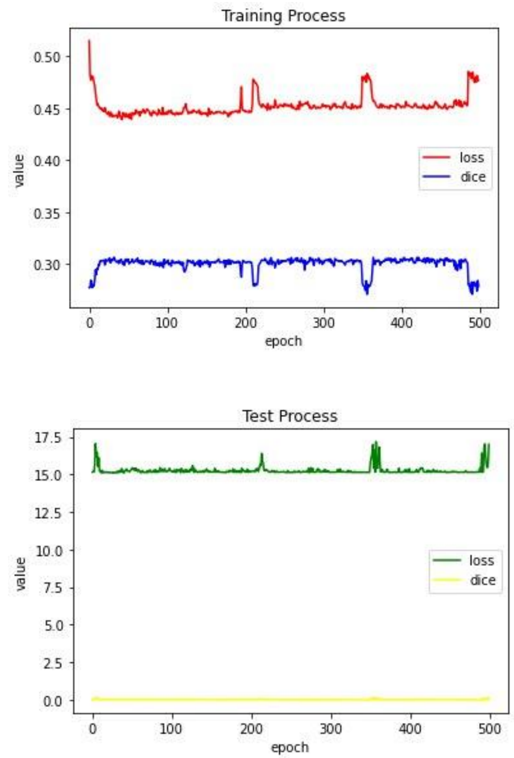


Figure 2: Values of dice and loss function for training and test processes after 500 epochs.

- **Dice score:**
$$\frac{2 \cdot Area\ of\ Overlap}{total\ pixels\ combined}$$
- **Loss function:**
$$BinaryCrossEntropy$$

Discussion:
First, we trained the network 50 times as we read in the paper but we though that it can improve the performance, so we trained a lot the network to see 2 things:

1) As far as the network can improve
2) At which point we can see the phenomenon of overfitting.

Results obtained

Comparison with true output and our output

In **Figure 1**, we can see that in our output the forgery “person” is noticed, but we have not the result as clear as in the second output that is the true output.

In **Figure 2**, we can see that in our output the forgery person is not noticed so much, the only better forgery points that are noticed are the face and one hand.

Reason: we suppose that the reasons is due to the different dataset that we have used. In fact, our dataset is small, instead those of the paper is more complex.

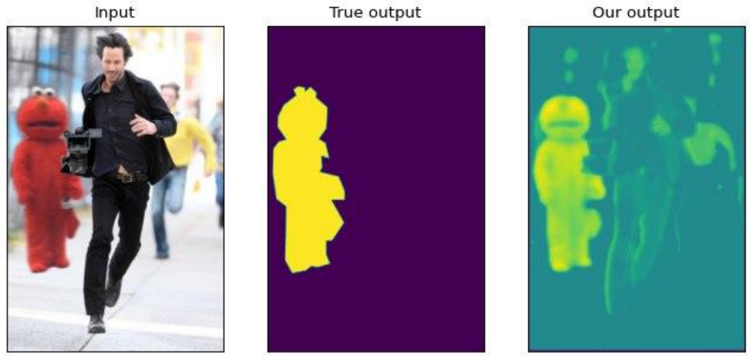


Figure 1: One of the better output that we have obtained.

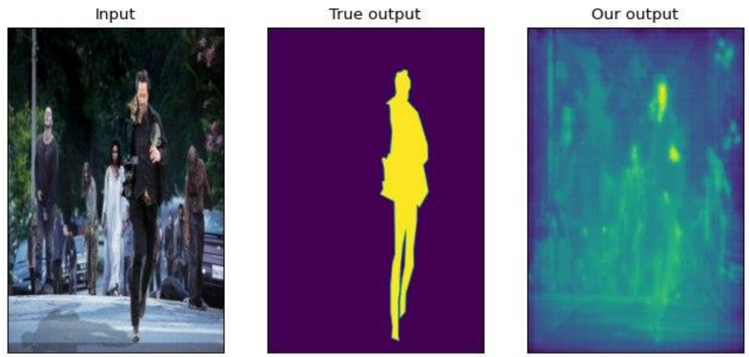


Figure 2: One of the worse output that we have obtained.

Personal Research

FSM block

We have decided to add the FSM block between the encoder and the decoder in order to see if the performance are better. FSM is used with U-Net to improve the performance and to capture long-range spatial information.

- Input feature map: $X_0 \in R^{H \times W \times C_0}$.
- Convolution 3×3 over I → B1
  Aim: used to filter out the irrelevant feature from the input I, because of this we obtain a feature map with depth $C < C_0$
- Three convolutions:
  - 1×1 → B2
  - 1×1 → B3
  - 1×1 → B4 indicates the representation of the input signal.
- B2 ⊗ B3 with softmax function → B5.
  Aim: relation map $f(x_i, x_j)$ which is the combination of dot product and softmax.
- B4 ⊗ B5 → B6
- Convolution 1×1 over B6, B1 ⊕ B6
- Convolution 3×3 over B6 → B7.
  Aim: obtain the initial dimension of the feature map.
- I ⊕ B7 → output of FSM (RIT)
  Aim: avoid overfitting.

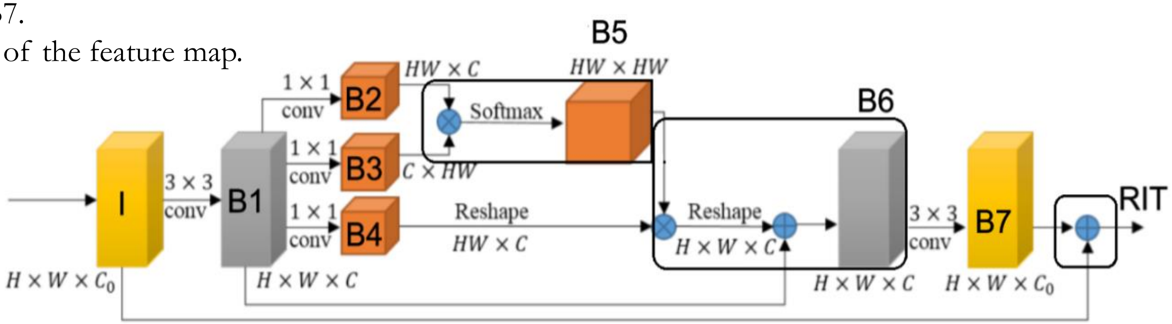


Figure: FSM architecture.

Result obtained with FSM block

In the **Figures,** we can see that by adding the FSM block our network are able to detect the forgery and a kind of halo is created around it.
During the training we have faster better performance, so we think that the FSM block will help to improve the performance of RRU-Net.

Figure 1: Example 1 by using FSM block.

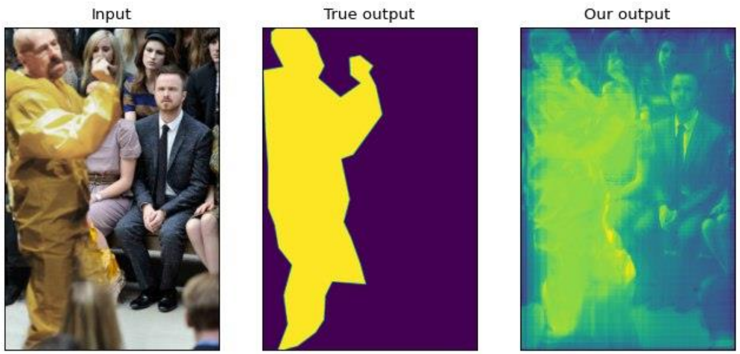


Figure 2: Example 2 by using FSM block.

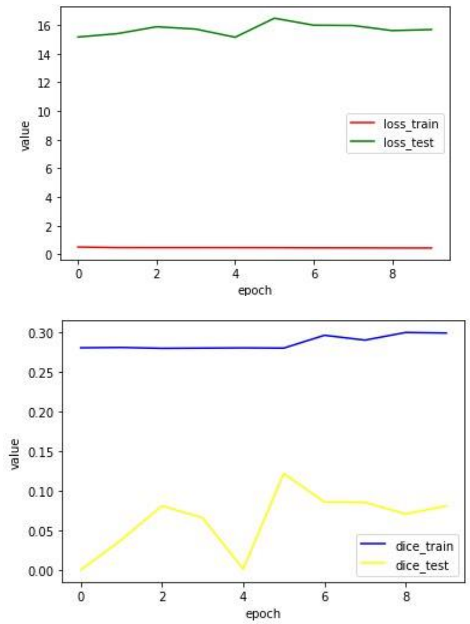


Figure 3: Values of loss function and dice for training and test processes after 10 epochs.

Comparisons of results obtained

Conclusion

Model	Dice	#Parameters
RRU-Net (paper)	0.7	4.09M
RRU-Net (our)	0.1	4.09M
RRU-Net with FSM	0.1	4.25M

Conclusions:

- The dataset that we have used is different form which in the paper. In fact, our dataset is smaller than that used in the paper, so the performance are different.
- We think that if we use a better dataset and we train our network with FSM block, we will obtain better performance than using only RRU-Net
- RRU-Net with FSM allow us to have faster better performance.

Thanks for the attention!

SAPIENZA
UNIVERSITÀ DI ROMA

Denise Landini
Alessandro Lambertini