This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI

10.1109/ACCESS.2020.3011752, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

**Local-Source Enhanced Residual Network for Steganalysis of Digital Images**

**WONHYUK AHN1, HANEOL JANG2, SEUNG-HUN NAM1, IN-JAE YU1, AND HEUNG-KYU LEE1**

* School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea

2 The Affiliated Institute of ETRI, Daejeon, South Korea

Corresponding author: Heung-Kyu Lee (heunglee@kaist.ac.kr).

 **ABSTRACT** Steganalysis refers to the study of identifying hidden messages in images inserted by steganography. Although detection performance is greatly improved when adopting convolutional neural networks (CNNs), they require sophisticated tricks, such as preprocessing for suppression of image content, using absolute and truncated activation functions, and utilizing domain knowledge. These tricks help networks train stably and mitigate the convergence problem of early stages in training, but they also restrict the flexibility of CNNs, which limits their performance. In this paper, we propose a local-source enhanced residual network (LSER) with end-to-end learning. The LSER is simple in its architecture but has two distinct characteristics from previous methods. First, the LSER uses residual blocks without any normalization. We find batch normalization is an unnecessary module in our framework. Second, a local-source skip connection is added to bypass features of different levels, which allows more abundant feature representation. Moreover, the LSER exhibits state-of-the-art results compared with the existing work in both spatial and JPEG domain steganalysis. Furthermore, a simple self-ensemble method further improves its performance without any side information.

 **INDEX TERMS** Steganalysis, low-level signal classification, convolutional neural networks, inconsisten-cies in noise pattern

**I. INTRODUCTION**

TEGANOGRAPHY is a secret communication method Sby hiding messages in images that are sent to the intended recipient. The most important requirement of steganography is that it should be impossible for steganalysis, which is a counterpart method to steganography, to distin-guish between ordinary images (cover) and images contain-ing secret messages (stego) [1]. Steganalysis is essentially re-quired to capture small disturbance, which cannot be found in cover images. For example, Fig. 1 illustrates example cover, stego, and residual (the pixel or quantized DCT coefficients difference between them) images. It is difficult to distinguish between the cover and stego images with the human eye because steganography inserts messages by flipping the LSB on the textual regions rather than the flat regions. As it deals with the subtle signal in digital content, the advances in steganalysis have been adopted in other applications dealing with subtle signals such as camera model identification [2], digital image forensics [3]–[5], content authentication, and

image-processing history.

The most powerful steganalysis was built using machine learning, starting from the work by Avcibacs et al. [8] and Lyu et al. [9]. This trend continues, and the state-of-the-art steganalysis methods have benefited from recent advances in deep learning, especially the works on ImageNet classifica-tion tasks [10]–[17]. However, most of the convolutional neu-ral network (CNN)-based steganalysis methods suffer from a convergence problem with randomly initialized parameters and require sophisticated tricks such as preprocessing the images to suppress the image content [10], [11], [13], [14], [17], using absolute and truncated activation functions [11], [14], [17], and utilizing the domain knowledge [12], [15], [18].

This is because the classification task in the computer vision field and steganalysis have different characteristics. At first, the original CNNs were usually developed for tack-ling high-level vision tasks such as ImageNet or CIFAR-10 classification. These CNNs usually have a feature reduction

|  |  |
| --- | --- |
| VOLUME xx, 2020 | 1 |

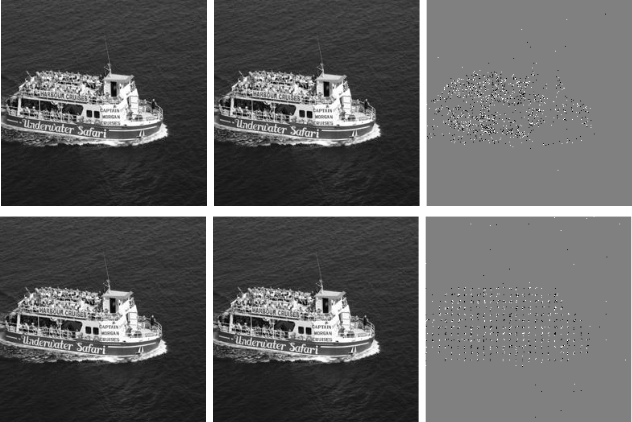
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

(a) (b) (c)



**FIGURE 1:** An example of (a) cover, (b) stego (secret messageinserted), and (c) residual images (cover-stego). Top row are the case of non-compressed images, where the messages are embedded into the pixels using S-UNIWARD [6]. Bottom row are the case of JPEG images with a quality factor of 95, where the messages are embedded into the quantized DCT coefficients using UED-JC [7]. We note that residual images are represented in the pixel and the DCT domain, respectively. Steganalysis aims to classify the cover and stego images, where the difference between them is very subtle.

module with a pooling layer or convolution with stride. However, these components remove the noise-like signal in-serted by steganography and make CNNs challenging to learn from scratch. In this sense, steganalysis can be categorized as a low-level vision task and requires a special network architecture for exploring low-level features.

Meanwhile, one of the representative low-level vision task is image super-resolution (SR). Although SR focuses on image restoration considering the local and global charac-teristics and does not classify subtle signals, various studies based on deep learning have been presented for dealing with low-level signals. Some of the methodologies in SR can inspire better CNN architecture for steganalysis but have not yet been explored in steganalysis.

In this paper, we propose a local-source enhanced residual network (LSER) for identifying spatial and JPEG steganogra-phy without any constraints. The LSER is inspired by recent advances in SR tasks that deal with low-level signals, but it is a non classification task. Overall, the LSER consists of three local-source residual groups, and each of them uses an enhanced residual, which is a normalization-free residual, as a basic building unit. In the end, the LSER replaces global average pooling with second-order pooling, which succes-sively captures discriminative features between channels. The experimental results reveal that the LSER converges stably and quickly with randomly initialized parameters and outperforms the existing state-of-the-art methods for both spatial and JPEG domain steganalysis.

In summary, our main contributions are three-fold:

We propose an LSER, which is a clean end-to-end framework for both spatial and JPEG steganalysis with state-of-the-art performance.

We find batch normalization (BN) harms the detection

of hidden messages in our framework, and use enhanced residual (ER) instead, which is the normalization-free residual block.

We propose a local-source skip connection, where fea-tures at different levels can be bypassed, and it allows more abundant feature representation.

1. **RELATED WORK**

**A. STEGANALYSIS**

Steganography can be classified into spatial steganography and JPEG steganography depending on the embedding do-main. Spatial steganography inserts messages into images by modifying the pixel values (top of Fig. 1) and JPEG steganography inserts messages by modifying the quantized discrete cosine transform (DCT) coefficients in the DCT domain (bottom of Fig. 1). On the other hand, steganalysis aims to identify inconsistencies between the cover and stego images. Steganalysis has also been built with two categories: spatial and JPEG steganalysis, depending on the targeting steganography.

With the development of CNN frameworks, CNN-based steganalysis methods have worked better than previous handcrafted-feature based methods. However, many of them have convergence problems with random initialized variables as the footprints of hidden messages are very subtle. Thus, they usually employed some tricks to help the network to converge. For instance, preprocessing images with high-pass filters is one of the typical techniques to identify the stego image embedded by spatial steganography. It suppresses the image content and reveals the noise stego signals. Fridrich et al. [19] proposed steganalysis rich model (SRM) and de-fined 30 linear and non-linear high-pass filters. The methods in [10], [11] use KV kernel of SRM, and the methods in [13], [14] utilize all 30 basic filters of SRM for preprocessing images. Zhu-Net [17] improved the performance by making 30 basic filters learnable rather than fixing the weights. In addition, domain knowledge is often incorporated in CNNs, especially for JPEG steganalysis. Xu et al. [12] used 4 4 DCT basis functions to transform images from the pixel domain into the DCT domain. Chen et al. [15] proposed a JPEG-phase CNN employing the PhaseSplit module in the middle, which splits and relocates the features into 64 groups by the JPEG phase.

SRNet [16] is a breakthrough in CNN-based steganalysis in that it can be trained with randomly initialized parameters without any tricks. The authors found that it is crucial to adopt unpooled layers, which maintains the input dimen-sion by disabling pooling and using stride 1, in the front part of the detector. SRNet consists of seven layers with-out downsampling among 12 layers, which leads to extract noise features regardless of the steganography algorithms and embedding domain. It outperforms previous CNN-based steganalysis methods in both the spatial and JPEG domain with first end-to-end architecture. However, Zhu-Net, which has a preprocessing layer that is initialized with high-pass filters, shows better performance against spatial steganog-

2 VOLUME xx, 2020

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

raphy, which indicates that SRNet may not be the optimal architecture for identifying subtle signals.

Recently, considerable attention has been paid to mak-ing steganalysis more practical. It ranges from identifying diversified stego images inserted by various steganography algorithms [20], [21] to detecting hidden messages on color images of an arbitrary size with various post-processing tech-niques, such as the demosaicing algorithm, resizing factor, denoising, sharpening, and enhancements tools [20]–[22]. To date, SRNet has been adopted as a baseline network to extend the practicality of steganalysis for these tasks because of its training ability and performance. In this sense, it is beneficial for detectors to have other baseline architectures that are better than SRNet to perform a more practical, “real-life” steganalysis.

**B. IMAGE SUPER-RESOLUTION**

SR refers to the process of recovering high-resolution images from low-resolution images, which is one of the important image processing techniques in computer vision. A variety of SR methods have been proposed for a few decades, and deep learning-based SR models have achieved state-of-the-art performance in recent years. In this section, we note the recent advances commonly used in SR models, that we take for our proposed LSER.

1) Removing Batch Normalization

Batch normalization (BN) [23] is proposed to reduce the covariate shift of networks and helps the training process by smoothing the optimization landscape [24]. It allows a higher learning rate and less careful initialization. It is known as an essential module in both image recognition and steganalysis [16], [17]. However in SR, Lim et al. [25] first argued that BN loses the scale information of each image and gets rid of range flexibility, which lowers SR performance. It also saves the memory cost up to 40%. From this work, state-of-the-art SR models often disable BN in their architectures [26]–[28].

2) Global and Local-source Residual Learning

In SR, many works [26]–[30] employ global-source residual learning, which refers to letting networks learn only the resid-uals between low-resolution and high-resolution images. It is valid because SR is an image-to-image translation task where the input image is highly correlated with the target images. With global-source residual learning, SR models only focus on the residual map to restore the missing high-frequency details. Local-source residual learning [26]–[28] is similar to residual learning in ResNet [31] to alleviate the degradation problem caused by increasing the network depth. It usually bypasses between layers and is added to make multiple shortcuts with different depths inside the network.

Inspired by the aforementioned modules in SR, we propose an LSER, which is a clean end-to-end network for steganaly-sis of digital images. We detail our LSER in the next section.

**A. NETWORK FRAMEWORK**

As shown in Fig. 2, our LSER consists of four parts: shallow feature extraction, unpooled feature extraction, deep feature reduction, and classification part. Let’s denote I and y as the input and its corresponding label of the LSER. We only adopt one convolutional (Conv) layer to extract the shallow feature F0 from the suspicious input

|  |  |
| --- | --- |
| F0 = HSF (I); | (1) |

where HSF ( ) stands for the convolution operation. Then the extracted shallow feature F0 is used for unpooled feature extraction with the local-source group (LSG) module, which produces the deep feature as follows:

|  |  |
| --- | --- |
| FUNPL = HLSG(F0); | (2) |

where HLSG( ) represents the LSG based unpooled feature extraction module, which consists of M ERs with a local-source skip connection. Then, the extracted unpooled feature FUNP L is used for deep feature reduction with two local-source downsample group (LSDG) modules. So we can further have

|  |  |
| --- | --- |
| FDF = HLSDGs(FUNP L); | (3) |

where HLSDGs( ) is two stacked LSDG, where LSDG is the same as LSG except for the downsampling. It induces a net-work to extract high-level features from low-level features. Lastly, the reduced deep feature FDF is convoluted with one 3 3 kernel and is pooled using global second-order pooling

|  |  |
| --- | --- |
| FP L = HGSoP (H3 3(FDF )); | (4) |

where H3 3 and HGSoP ( ) denote the Conv layer and global second order pooling with iterative matrix square root nor-malization (iSQRT-COV) [32], respectively. It explores the feature distribution and captures the feature statistics that are higher than the first-order for more discriminative represen-tations. It replaces global average pooling and steganalysis using iSQRT-COV also proposed [33]. The probability y^ of the existence of a hidden message is predicted via one fully connected layer and sigmoid function

|  |  |
| --- | --- |
| y^ = Sigmoid(HF C(FP L)); | (5) |

where Sigmoid( ) and HF C( ) denote the sigmoid function and fully connected layer, respectively.

Given a training set of images and their corresponding labels denoted by fIi; yi gNi , the goal of training the LSER is to minimize the binary cross entropy (BCE) loss

N

1 X

BCE(y; y^) = (yi log y^i+(1 yi) log(1 y^i)); (6)

N

i=1

where N is the batch size. The loss function is optimized by AdamW [34].

**III. PROPOSED METHOD**

|  |  |
| --- | --- |
| VOLUME xx, 2020 | 3 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

4

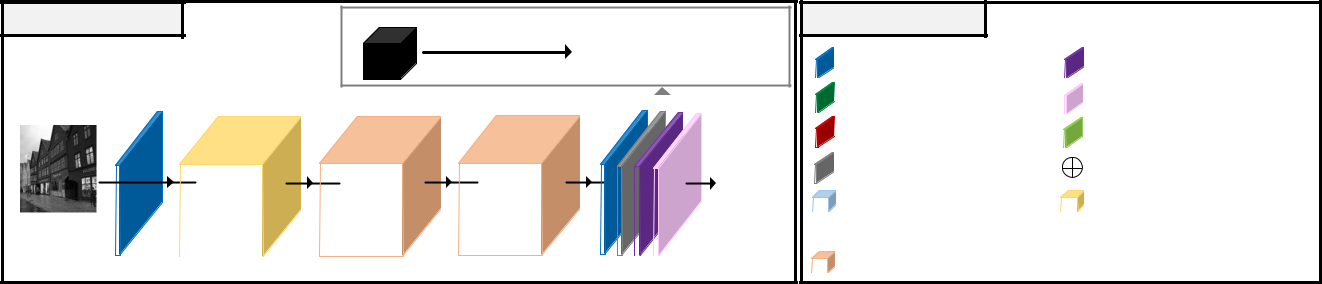
where Hg;3 3( ) and Hg;skip( ) denote the function of the 3 3 Conv layer at the tail and the function of the skip connection of the group, respectively. Hg;skip( ) is an iden-tity mapping function and the stride of Hg;3 3 is set to 1 for LSG. For LSDG, we use the downsample module (DM) as an operator Hg;skip for a concise downsampling. DM splits the Fg;0 into two parts (see blue elements of DM in Fig. 2), and each is convoluted on different locations with the 1 1 Conv layer with half number of channels. The outputs of both are concatenated to match the number of

where Hg;m( ) denotes the function of the m-th ER. Fg;m1 and Fg;m are the corresponding input and output. We observe that simply stacking the ER often results in divergence during training. It usually happens when training the steganalysis detector, so many researchers have proposed sophisticated tricks to alleviate it, as mentioned in Section II-A. The local-source skip connection is introduced to stabilize the training of the detector and also improves the performance by bypassing features

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images



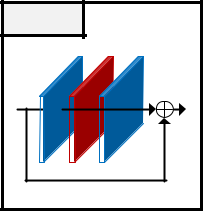
**Proposed LSER**

*C*: 64

Input

( × × )

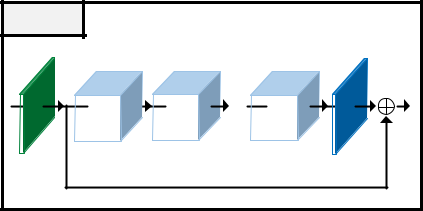
*s*: 1



*d*

*w*

*h*



* = 1, 2,⋯, = ℎ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  | 1 |  |
| 1 |  |  |  | 2 |  |
| − | | − |  |  |
|  |  |
|  |  |  |
|  |  | |  |  |  |
|  | |  |  |  |
|  |  |  |  |
|  |  | |  |  |  |
|  | *C*: 64 |  |  |  |  |
|  |  |  |  |  |  |

Cover

/ Stego

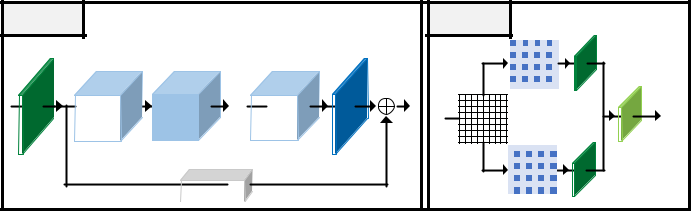
*s*: 1

**CNN components**

|  |  |
| --- | --- |
| 3 × 3 Conv layer | Fully connected layer |
| 1 × 1 Conv layer | Sigmoid |
| ReLU | Concatenation |
| iSQRT-COV | Element-wise addition |
| Enhanced residual (ER) | Local-source group (LSG) |

 Downsample module (DM)

Local-source downsample group (LSDG)



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ER** |  | **LSG** |  | **LSDG** | **DM** |
| *C*: 64 | *C*: 64 | *C*: 64 | *C*: 64 | *C*: 64 | *C*: 64 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | ... |  |  |  |  | ... |  |  |
| *s*: 1 | *s*: 1 | *s*: 1 | ER-1 | ER-2 | ER-*M* | *s*: 1 | *s*: 1 | ER-1 | ER-2 | ER-*M* | *s*: 2 |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**FIGURE 2:** Proposed local-source enhanced residual network (LSER) for spatial and JPEG steganalysis, whereCandsdenote the number ofchannels and stride for the convolutional layer, respectively. W and H represent the width and height for the input images.



**B. LOCAL-SOURCE RESIDUAL LEARNING**

We now give more details about our LSG and LSDG, which have a local-source skip connection. They contain a 1 1 Conv layer, M ERs, a 3 3 Conv layer, and a skip function. The input of the g-th LSG or LSDG first goes through 1 1 the Conv layer to learn the channel correlation via

|  |  |
| --- | --- |
| Fg;0 = Hg;1 1(Fg1 ) | (7) |

where Hg;1 1( ) denotes the function of the 1 1 Conv layer in the g-th LSG or LSDG, and Fg1 and Fg;0 are the corresponding input and output. Then, Fg;0 is used as the input of ERs, and the m-th ER (Fig. 2) can be formulated as

|  |  |
| --- | --- |
| Fg;m = Hg;m(Fg;m1 ) | (8) |

channels of Hg;3 3(Fg;M ). Comparing the proposed DM with other downsampling methods, it utilizes half the feature maps rather than ignoring three-quarters of them [16], [31] and is computationally efficient. It is inspired by performance gain with utilizing more feature maps in SR [35].

Such a local-source skip connection allows more abundant features to be bypassed during training. The LSG extracts low-level information by keeping the feature dimension at shallow layers. On the other hand, LSDG reduces the feature dimensions by downsampling, and makes the network to extract high-level features from low-level signals.

**C. ENHANCED RESIDUAL**

Residual networks exhibit excellent performance in various applications. BN layer has become an essential building unit in the residual block for image classification tasks. Most of the existing state-of-the-art steganalysis detectors have also adopted BN in their residual blocks. However, BN gets rid of the range flexibility of features by normalizing them, it is better to remove the BN layers when dealing with low-level signals, as shown in recent SR works [25]–[28].

Inspired by the above observations, we use the residual block without BN, called ER, in our network. In Fig. 3, we

Fg = Hg;3 3(Fg;M ) + Hg;skip(Fg;0) (9) compare the residual block of SRNet and the LSER. It is

a first attempt to use a normalization-free residual module in the steganalysis detector. We experimentally demonstrate that the model with ER achieves better performance than that with the residual module with BN. With the proposed ER, we can build a deeper and wider model under limited computational resources because removing the BN layers saves GPU memory usage of the model by up to 40%. Moreover, our LSER can work on mini-batches of size 2, which is the case in which CNNs with BN usually fail.

VOLUME xx, 2020

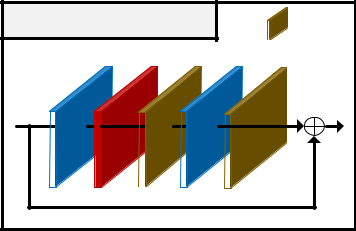
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access

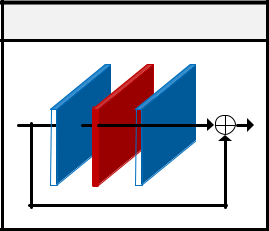


Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

|  |  |  |
| --- | --- | --- |
| **Type 2 in SRNet** | | : BN |
| *s*: 1 | *s*: 1 |  |



**Enhanced residual (ER)**



*s*: 1 *s*: 1

2) Evaluation Metric

The performance of detectors is measured by error rate on the test set under equal priors PE = minPFA 12 (PFA + PMD), where PF A and PMD are the false-alarm and missed-detection probabilities. We also provide the area under the curve (AUC) and plot the receiver operating characteristic (ROC) curve for the selected cases. One random splitting is used for evaluating the experimental results as in [16]

**FIGURE 3:** Comparison of the residual block in SRNet (type 2) andthe enhanced residual (ER) block

**D. IMPLEMENTATION DETAILS**

Now we specify the implementation details of our proposed LSER. We put one LSG and two LSDGs after shallow feature extraction. In each LSG and LSDG, we set the number of ERs to M = 5. The Conv layers in all modules have C = 64 filters. We would like to note that bias of all Conv layers is set to be trainable to compensate for the absence of BN layers, and the performance is degraded without bias learning. We set 3 3 as the size of all Conv layers except the first Conv layer in LSG and LSDG and DM for downsampling. Lastly, we set the number of Newton-Schulz iterations to seven for global second order pooling. Our LSER have 1.28 million parameters and its computational complexity is 58.57 MACs (multiply–accumulate operation).

**IV. EXPERIMENTS**

To demonstrate the effectiveness of the proposed LSER, two state-of-the-art steganalysises, SRNet and Zhu-Net, are compared with the LSER.

**A. SETTINGS**

We clarify the experimental settings regarding datasets, eval-uation metrics, training settings, and curriculum learning.

1) Datasets

Following [14], [16], [17], we use the union of BOSSbase 1.01 [36] and BOWS2 [37], each containing 10,000 greyscale images, reducing them from their original size of 512 512 to 256 256 using imresize with the default setting in MAT-LAB. For JPEG source images, this source is additionally compressed with quality factors 75 and 95 and default setting in MATLAB. The entire BOWS2 dataset is used for training. We randomly divided the images from BOSSBase into train-ing, validation, and test sets in the ratio of 4:1:5. We consider two spatial domain steganographic methods: S-UNIWARD

1. and WOW [38] and two JPEG domain steganographic algorithms: J-UNIWARD [6] and UED-JC [7], using MAT-LAB implementations with a random embedding key. In summary, 2 14; 000 cover and stego images were used for training, 2 1; 000 for validation, and 2 5; 000 for testing each steganographic method. In addition, the JPEG images are decompressed without rounding to an integer.

because it is not computationally feasible to train all the networks on multiple splits.

3) Training Settings

Data augmentation is performed on the training set, which is randomly rotated by 90 , 180 , and 270 and is flipped horizontally. In each training batch, 16 cover images and their paired 16 stego images are constructed, and each of the cover-stego pairs is augmented the same way among eight cases. Different augmentations for each cover and stego image result in divergence for all steganalysis methods. All models are trained up to 200 epochs by AdamW [34] with B1 = 0:9, B2 = 0:999, = 108 , and weight decay = 105 . The learning rate is initialized as 104 and then is reduced to 105 after 100 epochs. Our proposed LSER was implemented on the PyTorch [39] framework on an NVIDIA TITAN RTX GPU.

4) Curriculum Learning

The above training strategy is applied for all embedding algorithms for payload of 0.4 bpp (bit per pixel) / bpnzac (bits per non-zero AC DCT coefficients). For the remaining payloads (0.1, 0.2, and 0.3), we build the detector via curricu-lum learning [40] following previous work [16], described symbolically as 0:4 ! 0:3 ! 0:2 ! 0:1. In other words, the detector trained for a payload of 0.4 is fine-tuned on a payload of 0.3 for up to 100 epochs. The learning rate is set to 105 and then decreases to 106 after 50 epochs. Then, the detector for a payload of 0.2 is fine-tuned with the pretrained model for a payload of 0.3, and so on.

**B. GEOMETRIC SELF-ENSEMBLE**

We propose a simple self-ensemble strategy inspired by the SR work [25]. During testing time, we rotated and flipped the images to generate seven augmented inputs Ii = Ti(I) for each sample, where Ti the represents geometric transfor-mation including the identity. With a trained network, we can acquire the predicted probabilities fy^1; : : : ; y^8 g. Finally, we average the predicted probabilities together to determine the self-ensemble predicted probability, as follows:

y^se = 1 8 y^i: (10)

X

8

i=1

This simple self-ensemble method does not require addi-tional training of the separate model. It is useful especially when the model size and training time matter. We would like to note that it also does not require additional information

|  |  |
| --- | --- |
| VOLUME xx, 2020 | 5 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

**TABLE 1:** Effects of different modules. We report the best error ratePEon the validation set of J-UNIWARD at 0.4 bpnzac for QF 75

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Base | Ra | Rb | Rc |
|  |  |  |  |  |
| Enhanced Residual (ER) |  | ! |  | ! |
|  |  |  |  |  |
| Local-source Residual Learning |  |  | ! | ! |
|  |  |  |  |  |
| PE | 0.0640 | 0.0615 | 0.0600 | 0.0558 |

**TABLE 2:** Detection error ratePEand AUC for the SRNet, Zhu-Net, LSER, and LSER+ for four payloads in bpp and two spatial domainembedding algorithms

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Domain | Steganography | Detector |  |  | bpp |  |  |
|  |  |  |  |  |
| 0.1 | 0.2 | 0.3 | 0.4 |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  | SRNet | 0.3602/0.7150 | 0.2624/0.8369 | 0.1928/0.9098 | 0.1527/0.9400 |  |
|  | S-UNIWARD | Zhu-Net | 0.3339/0.7611 | 0.2466/0.8593 | 0.2082/0.8980 | 0.1492/0.9408 |  |
|  | LSER | 0.2846/0.8201 | 0.1860/0.9176 | 0.1263/0.9598 | 0.0957/0.9762 |  |
|  |  |  |
| Spatial |  | LSER+ | 0.2782/0.8299 | 0.1828/0.9220 | 0.1220/0.9627 | 0.0904/0.9782 |  |
|  |  |  |  |  |  |  |
|  | SRNet | 0.3047/0.7939 | 0.2100/0.8963 | 0.1532/0.9415 | 0.1139/0.9669 |  |
|  |  |  |
|  | WOW | Zhu-Net | 0.2818/0.8167 | 0.2092/0.8953 | 0.1571/0.9372 | 0.1109/0.9666 |  |
|  | LSER | 0.2375/0.8678 | 0.1554/0.9412 | 0.1084/0.9698 | 0.0815/0.9826 |  |
|  |  |  |
|  |  | LSER+ | 0.2290/0.8774 | 0.1453/0.9467 | 0.1037/0.9733 | 0.0756/0.9849 |  |
|  |  |  |  |  |  |  |  |

on embedding algorithms as opposed to selection-channel-aware steganalysis [41]–[44]. We find that it provides ad-ditional performance gain against both spatial and JPEG steganography in terms of the error rate PE and AUC. We denote the method with self-ensemble by adding the postfix ‘+’ to the method name (i.e., LSER+).

**C. ABLATION STUDY**

As discussed in Section III, our LSER contains two main components including ER and local-source residual learn-ing. To verify the effectiveness of different modules, we compared ER with its variant trained and tested it on J-UNIWARD at 0.4 bpnzac for a quality factor (QF) 75. The specific performance is listed in Table 1. Base refers to the basic baseline following the network framework of the LSER described in Section III-A, but employing a residual block with BN (see the left one in Fig. 3) and not employing a local-source skip connection described in Equation (9). From Table 1, we can see that Base reaches an error rate of 6.4%. The results from Ra to Rc verify the effectiveness of the proposed modules, because they show the performance gain compared to the Base model.

Specifically, Ra replaces the residual module to ER by removing BN, which obtains slightly better performance than the Base with saving GPU memory usage of up to 40%. This indicates that the BN layers are unnecessary modules in our LSER framework. When a local-source skip connection is added alone (i.e., Rb), the performance can be improved from a 6.4% error rate to a 6.0% error rate. The main reason for the performance gain from the local-source skip connection is that it allows a different level of features to be bypassed, where low-level features for LSG and more high-level features for LSDG. When both Ra and Rb are used (i.e.,

Rc), the performance can be further improved, achieving 5.58%. It can be seen that two components discriminately capture the traces of message embedding by steganography, and contribute more performance gain together. These com-parisons firmly demonstrate the effectiveness of the proposed components.

**D. RESULTS ON THE SPATIAL DOMAIN**

To evaluate our LSER on spatial domain steganography, we compared our method with the state-of-the-art methods: SR-Net [16] and Zhu-Net [17]. As mentioned in Section IV-B, we adopted a self-ensemble method to further improve our LSER denoted as the LSER+. We considered four payloads: 0.1 to 0.4 bpp (bits per pixel) for WOW [38] and S-UNIWARD [6]. The detection error rate PE and AUC are shown in Table 2. Depending on the algorithm and payload, the LSER improves upon Zhu-Net by up to 5% for the PE. Moreover, the pro-posed self-ensemble LSER+ further improves the LSER for all payloads and steganography in terms of PE and AUC. The biggest improvement is typically observed in smaller payloads. This indicates that the LSER can capture smaller changes in images than other methods.

Fig. 4 illustrates an example of the progression of the validation error and loss when training the LSER, SRNet, and Zhu-Net on WOW at 0.4 bpp. The LSER (red line) shows clearly better convergence speed and performance than SRNet and Zhu-Net in terms of the error rate and loss. We would like to note that SRNet is similar to the LSER in that its goal is to design a clean end-to-end architecture for steganalysis in both the spatial and JPEG domain. However, SRNet is less effective than Zhu-Net, which initializes the preprocessing layer with heuristic high-pass filters. So we can judge that SRNet can be further improved with more

6 VOLUME xx, 2020

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access

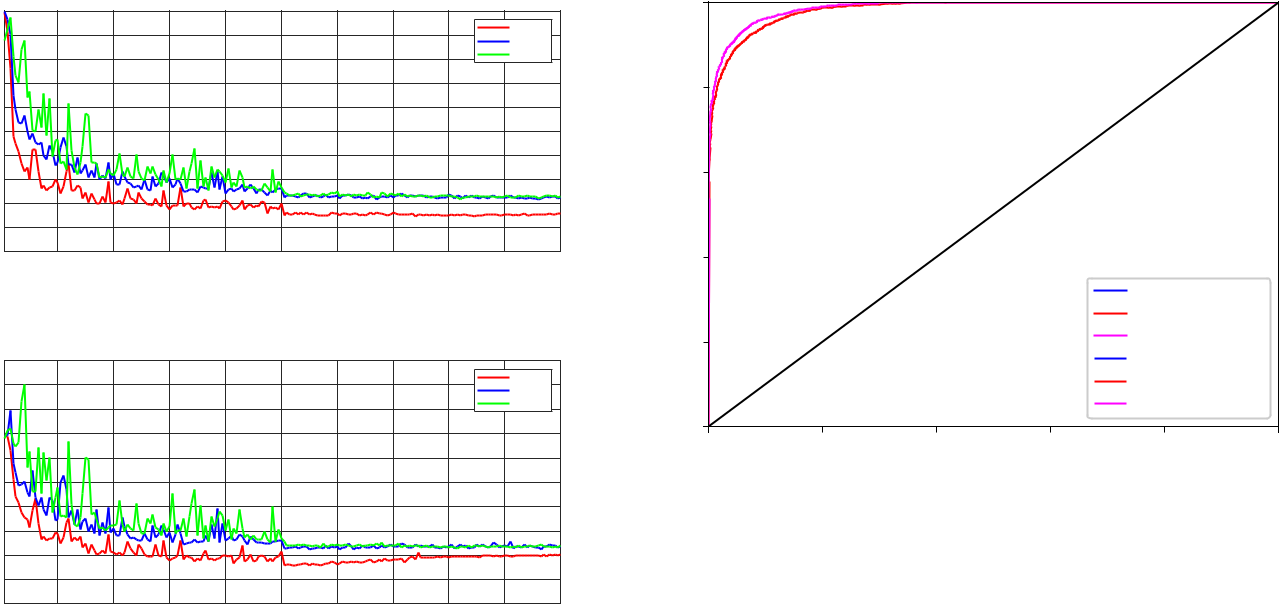


Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

**TABLE 3:** Detection error ratePEand AUC for the SRNet, LSER, and LSER+ for four payloads in bpnzac and two JPEG domain embeddingalgorithms

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Domain | Steganography | | Detector |  | bpnzac | |  |  |
|  |  |  |  |  |
| 0.1 | 0.2 | 0.3 | 0.4 |  |
|  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  | SRNet | 0.3750/0.6843 | 0.2444/0.8375 | 0.1606/0.9249 | 0.0999/0.9659 |  |
|  |  | J-UNIWARD | LSER | 0.3115/0.7582 | 0.1787/0.9037 | 0.1031/0.9656 | 0.0577/0.9875 |  |
|  | QF 75 |  | LSER+ | 0.2987/0.7791 | 0.1679/0.9175 | 0.0919/0.9724 | 0.0500/0.9905 |  |
|  |  |  |  |  |  |  |  |
|  |  | SRNet | 0.2154/0.8761 | 0.1063/0.9672 | 0.0528/0.9902 | 0.0324/0.9969 |  |
|  |  |  |  |
|  |  | UED-JC | LSER | 0.1176/0.9589 | 0.0501/0.9917 | 0.0272/0.9977 | 0.0225/0.9980 |  |
| JPEG |  |  | LSER+ | 0.1054/0.9653 | 0.0440/0.9935 | 0.0246/0.9982 | 0.0187/0.9988 |  |
|  |  |  |  |  |  |  |  |
|  |  | SRNet | 0.4528/0.5728 | 0.3928/0.6639 | 0.3206/0.7584 | 0.2507/0.8378 |  |
|  |  |  |  |
|  |  | J-UNIWARD | LSER | 0.4226/0.6170 | 0.3284/0.7396 | 0.2510/0.8451 | 0.1844/0.9083 |  |
|  | QF 95 |  | LSER+ | 0.4079/0.6403 | 0.3092/0.7692 | 0.2276/0.8680 | 0.1651/0.9262 |  |
|  |  |  |  |  |  |  |  |
|  |  | SRNet | 0.3729/0.6920 | 0.2680/0.8193 | 0.1976/0.9021 | 0.1314/0.9544 |  |
|  |  |  |  |
|  |  | UED-JC | LSER | 0.2904/0.7966 | 0.2098/0.8920 | 0.1409/0.9522 | 0.0746/0.9834 |  |
|  |  |  | LSER+ | 0.2728/0.8203 | 0.1949/0.9085 | 0.1274/0.9618 | 0.0650/0.9874 |  |
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 50 |  |  |  |  |  |  |  |  |  |  | 1.0 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 45 |  |  |  |  |  |  |  | LSER |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  | SRNet |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 40 |  |  |  |  |  |  |  | Zhu-Net |  |  |  |  |  |  |  |  |  |
| (%) |  |  |  |  |  |  |  |  |  |  | 0.8 |  |  |  |  |  |  |
| 35 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| error | 30 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Validation | 25 |  |  |  |  |  |  |  |  |  | PositiveRate |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 20 |  |  |  |  |  |  |  |  |  |  | 0.6 |  |  |  |  |  |  |
|  | 15 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 10 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 0 |  |  |  |  |  |  |  |  |  | True | 0.4 |  |  |  |  |  |  |
|  | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 |  |  |  |  |  |  |
|  |  |  |  |  | Epoch |  |  |  |  |  |  |  |  |  |  | SRNet, 0.1 bpnzac |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | (a) Validation error for LSER, SRNet, and Zhu-Net | | | | | | | |  |  |  |  |  |  |  | LSER, 0.1 bpnzac |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | 0.2 |  |  |  | LSER+, 0.1 bpnzac |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | SRNet, 0.4 bpnzac |  |  |
|  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 0.9 |  |  |  |  |  |  |  | LSER |  |  |  |  |  |  | LSER, 0.4 bpnzac |  |  |
|  | 0.8 |  |  |  |  |  |  |  | SRNet |  |  |  |  |  |  | LSER+, 0.4 bpnzac |  |  |
|  |  |  |  |  |  |  |  | Zhu-Net |  |  |  |  |  |  |  |  |
| loss | 0.7 |  |  |  |  |  |  |  |  |  |  | 0.00.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |  |
| 0.6 |  |  |  |  |  |  |  |  |  |  |  |  | False Positive Rate | |  |  |  |
| Validation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0.5 |  |  |  |  |  |  |  |  |  | **FIGURE 5:** ROC curves of the SRNet, LSER, and LSER+ for UED-JC | | | | | | |  |
|  | 0.4 |  |  |  |  |  |  |  |  |  |  |
|  | 0.3 |  |  |  |  |  |  |  |  |  | at 0.1 and 0.4 bpnzac for QF 95 | | | |  |  |  |  |
|  | 0.2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 0.1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 0 |  |  |  |  |  |  |  |  |  | for payloads of 0.1 to 0.4 bpnzac (bits per non-zero AC DCT | | | | | | |  |
|  | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 |  |
|  |  |  |  |  | Epoch |  |  |  |  |  | coefficients) are tested for QF 75 and 95. The results of the | | | | | | |  |
|  | (b) Validation loss for LSER, SRNet, and Zhu-Net | | | | | | | |  |  |  |
|  |  |  | experiments are shown in Table 3. | | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



|  |  |  |
| --- | --- | --- |
| **FIGURE 4:** Validation error and loss for WOW at 0.4 bpp | Our LSER exhibits a significant improvement compared |  |
|  | to SRNet for all payloads, steganography, and QF in terms |  |
| adequate architecture for universal steganalysis. | of PE and AUC. Again, the LSER+ further improves perfor- |  |
| mance. We would like to note the gap between LSER and |  |
|  | LSER+ is larger in the JPEG domain steganalysis than in |  |
| **E. RESULTS ON JPEG DOMAIN** | spatial domain steganalysis. We can confirm that proposed |  |
| For the JPEG domain, we only considered SRNet to evaluate | self-ensemble works regardless of the embedding domain |  |
| the proposed LSER and LSER+ because SRNet outperforms | and steganography. Fig. 5 illustrates six ROC curves for the |  |
| the existing methods by a large margin [16]. Many works | SRNet, LSER, and LSER+ for UED-JC for two payloads. It |  |
| use SRNet as a baseline to make steganalysis robust to real- | reveals that the LSER and LSER+ are more reliable detectors |  |
| world images. It includes steganalysis for color JPEG images | than SRNet regardless of payloads. |  |
| with high-resolution [21], and high QF with small payloads | We trained Zhu-Net, which is proposed for spatial domain |  |
| [45]. In this experiment, J-UNIWARD [6] and UED-JC [7] | steganalysis, for J-UNIWARD and UED-JC. It achieves a |  |
| VOLUME xx, 2020 | 7 |  |

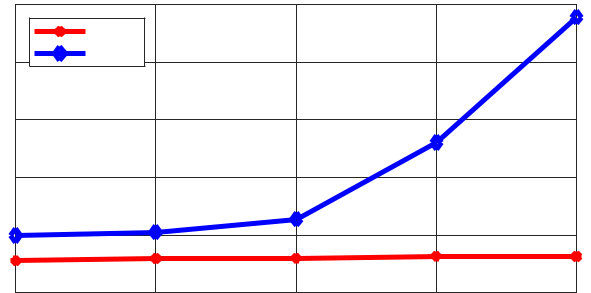
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 50 |  |  |  |  |  |
|  |  | LSER |  |  |  |  |
|  | 40 | SRNet |  |  |  |  |
|  |  |  |  |  |  |
| (%) | 30 |  |  |  |  |  |
| rate |  |  |  |  |  |
|  |  |  |  |  |  |
| Error | 20 |  |  |  |  |  |
|  |  |  |  |  |  |
|  | 10 |  |  |  |  |  |
|  | 0 |  |  |  |  |  |
|  | 32 | 16 | 8 | 4 | 2 |  |



is 47.8%, which is similar to the random guessing detector. The red line indicates the error rate of the LSER, which has very similar performance across a wide range of batch sizes from 32 to 2 (5.58% for 32 and 6.35% for 2). The LSER is clearly robust to small batch sizes, which indicates a deeper and wider model can be trained under limited memory. This is especially important when only small batch sizes are available such as targeting images of “real life.” where the images are color images with high-resolution (e.g., higher than 1024 1024).

Batch size

**FIGURE 6:** Classification errorvs.batch size for the LSER and SRNetfor J-UNIWARD at 0.4 bpnzac

15.36 % error rate for J-UNIWARD and 5.51 % error rate for UED-JC, which is much worse than SRNet. Because the initialization of Zhu-Net is done for capturing the noise in the spatial region, useful information in the DCT domain is removed by those kernels.

**F. CLASSIFICATION ERROR RATE** VS. **BATCH SIZE**

Researchers have found that it is important to keep the input size in the shallow layers to extract noise features by prepro-cessing with high-pass filters [13], [14] or adopting unpooled modules [16], [17] for deep-learning based steganalysis. These designs require considerable memory usage alone. Moreover, most steganalysis models adopt the BN layer in their architecture following great success in computer vision tasks. It requires additional memory use and a sufficiently large batch size for BN to work properly (e.g., 32 per worker). A small-batch leads to an inaccurate estimation of batch statistics, and reducing the batch size increases the model error dramatically. Shared normalization [46], which uses a larger batch to obtain the initial statistics, may alleviate the small-batch problems but it cannot remove the root cause that features are extracted based on batch statistics. These two factors make CNN-based steganalysis to be narrow and shallow architecture and restrict to detecting images of a higher resolution than 256 256. Although there are some alternative normalization methods that are irrelevant to the batch size proposed in the computer vision field, such as instance normalization [47], layer normalization [48], and group normalization [49], steganalysis with these normaliza-tion methods has not been proposed so far and may harm the noise features left in the images.

On the other hand, the LSER uses ER, which is a normalization-free residual block, as a basic unit of opera-tion. In Section IV-C, we demonstrate that BN is an unneces-sary module in our LSER framework, and simply removing it provides a performance gain. Fig. 6 presents the validation error rate for the LSER and SRNet in terms of the batch size per worker. Although SRNet for batch size of 32 achieves 9.98%, its performance is degraded severely with smaller batch sizes. At a batch size of 2, the validation error rate

**V. FUTURE WORK**

The LSER can be an alternative baseline architecture design for detecting inconsistencies in the noise patterns of images. Its design is simple but exhibits effective performance with-out any constraints or heuristics. For future work, we will search for hyperparameters such as the number of channels, ER, LSG, and LSDG and their combinations when limited resources are given. To improve the architecture, attention mechanisms can be introduced into ER, LSG, and LSDG following the great successes in image recognition [50] and image restoration tasks [27], [28]. However, adding these modules require additional computational resources thus an efficient architecture should be considered.

**VI. CONCLUSION**

Although the ultimate goal of steganalysis by itself is to detect inconsistencies in noise patterns of images, the current steganalysis designs typically incorporate heuristic tricks that help the convergence but constrain the flexibility of models. SRNet is the first clean end-to-end CNN-based steganalysis and the first detector identifying hidden messages in the spa-tial and JPEG domains. Our goal is to design an alternative baseline architecture that is more adequate for capturing uni-versal low-level signals. In this paper, we proposed an LSER for both spatial and JPEG domain steganalysis. Inspired by recent advances in SR works, we used ER, which removes unnecessary normalization layers in the residual block. We also added a local-source skip connection between groups of ERs to allow the features to be bypassed at the different levels. With those two insights, the LSER exhibits state-of-the-art performance in both spatial and JPEG domain steganalysis. Furthermore, a simple geometric self-ensemble method was introduced, which further improves the LSER performance without any side information.

**ACKNOWLEDGMENT**

This work was partly supported by the National Research Foundation of Korea (NRF) grant (NRF-2019R1A2C2084569) and the Institute of Information & communications Technology Planning & Evaluation (IITP) grant (No. 2017-0-01671, Development of high reliability image and video authentication service for smart media environment) funded by the Korea Government (MSIT).

8 VOLUME xx, 2020

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

**REFERENCES**

1. J. Fridrich, Steganography in digital media: principles, algorithms, and applications. Cambridge University Press, 2009.
2. A. Tuama, F. Comby, and M. Chaumont, “Camera model identification with the use of deep convolutional neural networks,” in 2016 IEEE Inter-national workshop on information forensics and security (WIFS). IEEE, 2016, pp. 1–6.
3. Y. Rao and J. Ni, “A deep learning approach to detection of splicing and copy-move forgeries in images,” in 2016 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2016, pp. 1–6.
4. P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, “Learning rich features for image manipulation detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1053–1061.
5. S.-H. Nam, W. Ahn, S.-M. Mun, J. Park, D. Kim, I.-J. Yu, and H.-K. Lee, “Content-aware image resizing detection using deep neural network,” in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 106–110.
6. V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” EURASIP Journal on Information Security, vol. 2014, no. 1, p. 1, 2014.
7. L. Guo, J. Ni, and Y. Q. Shi, “Uniform embedding for efficient jpeg steganography,” IEEE transactions on Information Forensics and Security, vol. 9, no. 5, pp. 814–825, 2014.

˙

1. I. Avcıba¸s, M. Kharrazi, N. Memon, and B. Sankur, “Image steganalysis with binary similarity measures,” EURASIP Journal on Advances in Signal Processing, vol. 2005, no. 17, p. 679350, 2005.
2. S. Lyu and H. Farid, “Detecting hidden messages using higher-order statistics and support vector machines,” in International Workshop on information hiding. Springer, 2002, pp. 340–354.
3. Y. Qian, J. Dong, W. Wang, and T. Tan, “Deep learning for steganalysis via convolutional neural networks,” in Media Watermarking, Security, and Forensics 2015, vol. 9409. International Society for Optics and Photonics, 2015, p. 94090J.
4. G. Xu, H.-Z. Wu, and Y.-Q. Shi, “Structural design of convolutional neural networks for steganalysis,” IEEE Signal Processing Letters, vol. 23, no. 5, pp. 708–712, 2016.
5. G. Xu, “Deep convolutional neural network to detect j-uniward,” in Pro-ceedings of the 5th ACM Workshop on Information Hiding and Multime-dia Security, 2017, pp. 67–73.
6. M. Yedroudj, F. Comby, and M. Chaumont, “Yedroudj-net: An efficient cnn for spatial steganalysis,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 2092–2096.
7. J. Ye, J. Ni, and Y. Yi, “Deep learning hierarchical representations for image steganalysis,” IEEE Transactions on Information Forensics and Security, vol. 12, no. 11, pp. 2545–2557, 2017.
8. M. Chen, V. Sedighi, M. Boroumand, and J. Fridrich, “Jpeg-phase-aware convolutional neural network for steganalysis of jpeg images,” in Proceed-ings of the 5th ACM Workshop on Information Hiding and Multimedia Security, 2017, pp. 75–84.
9. M. Boroumand, M. Chen, and J. Fridrich, “Deep residual network for ste-ganalysis of digital images,” IEEE Transactions on Information Forensics and Security, vol. 14, no. 5, pp. 1181–1193, 2018.
10. R. Zhang, F. Zhu, J. Liu, and G. Liu, “Depth-wise separable convolutions and multi-level pooling for an efficient spatial cnn-based steganalysis,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1138–1150, 2019.
11. J. Huang, J. Ni, L. Wan, and J. Yan, “A customized convolutional neural network with low model complexity for jpeg steganalysis,” in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2019, pp. 198–203.
12. J. Fridrich and J. Kodovsky, “Rich models for steganalysis of digital images,” IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, pp. 868–882, 2012.
13. J. Butora and J. Fridrich, “Detection of diversified stego sources with cnns,” Electronic Imaging, vol. 2019, no. 5, pp. 534–1, 2019.
14. Y. Yousfi, J. Butora, J. Fridrich, and Q. Giboulot, “Breaking alaska: Color separation for steganalysis in jpeg domain,” in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2019, pp. 138– 149.
15. R. Cogranne, Q. Giboulot, and P. Bas, “The alaska steganalysis challenge: A first step towards steganalysis,” in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2019, pp. 125–137.
16. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, 2015.
17. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, “How does batch normalization help optimization?” in Advances in Neural Information Processing Systems, 2018, pp. 2483–2493.
18. B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017,
    1. 136–144.
19. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2472–2481.
20. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Pro-ceedings of the European Conference on Computer Vision (ECCV), 2018,
    1. 286–301.
21. T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, “Second-order atten-tion network for single image super-resolution,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 11 065–11 074.
22. J. Kim, J. Kwon Lee, and K. Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1646– 1654.
23. Y. Tai, J. Yang, X. Liu, and C. Xu, “Memnet: A persistent memory network for image restoration,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 4539–4547.
24. 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, 2016, pp. 770–778.
25. P. Li, J. Xie, Q. Wang, and Z. Gao, “Towards faster training of global co-variance pooling networks by iterative matrix square root normalization,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 947–955.
26. X. Deng, B. Chen, W. Luo, and D. Luo, “Fast and effective global covariance pooling network for image steganalysis,” in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2019,
    1. 230–234.
27. I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in International Conference on Learning Representations, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7
28. F. Toutounchi and E. Izquierdo, “Advanced super-resolution using lossless pooling convolutional networks,” in 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019, pp. 1562–1568.
29. P. Bas, T. Filler, and T. Pevny,` “” break our steganographic system”: the ins and outs of organizing boss,” in International workshop on information hiding. Springer, 2011, pp. 59–70.
30. P. Bas and T. Furon, “Bows-2,” 2007.
31. V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” in 2012 IEEE International workshop on information forensics and security (WIFS). IEEE, 2012, pp. 234–239.
32. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” 2017.
33. Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in Proceedings of the 26th annual international conference on machine learning, 2009, pp. 41–48.
34. T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, “Selection-channel-aware rich model for steganalysis of digital images,” in 2014 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2014, pp. 48–53.
35. W. Tang, H. Li, W. Luo, and J. Huang, “Adaptive steganalysis against wow embedding algorithm,” in Proceedings of the 2nd ACM workshop on Information hiding and multimedia security, 2014, pp. 91–96.
36. ——, “Adaptive steganalysis based on embedding probabilities of pixels,” IEEE Transactions on Information Forensics and Security, vol. 11, no. 4,
    1. 734–745, 2015.
37. T. D. Denemark, M. Boroumand, and J. Fridrich, “Steganalysis features for content-adaptive jpeg steganography,” IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1736–1746, 2016.
38. J. Butora and J. Fridrich, “Reverse jpeg compatibility attack,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1444– 1454, 2019.

|  |  |
| --- | --- |
| VOLUME xx, 2020 | 9 |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011752, IEEE Access



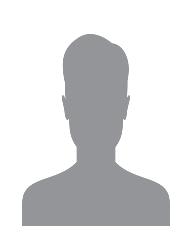
Ahn et al.: Local-source Enhanced Residual Network for Steganalysis of Digital Images

1. S. Wu, S.-h. Zhong, and Y. Liu, “A novel convolutional neural network for image steganalysis with shared normalization,” IEEE Transactions on Multimedia, vol. 22, no. 1, pp. 256–270, 2019.
2. D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance normalization: The missing ingredient for fast stylization,” arXiv preprint arXiv:1607.08022, 2016.
3. J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” arXiv preprint arXiv:1607.06450, 2016.
4. Y. Wu and K. He, “Group normalization,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
5. J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Pro-ceedings of the IEEE conference on computer vision and pattern recogni-tion, 2018, pp. 7132–7141.

**WONHYUK AHN** received the B.S. degree insoftware and computer engineering from Ajou University, South Korea, in 2016. He is currently pursuing the Ph.D. degree with the School of Computing, Korea Advanced Institute of Science and Technology (KAIST). His research interests include multimedia forensics, computer vision, and machine learning.



**HANEOL JANG** received his BE degree in infor-mation computer engineering from Ajou Univer-sity, in 2012, and the MS degree in computer sci-ence from KAIST, in 2014 and the Ph.D. degree in school of computing from KAIST, Korea, in 2018. Since 2018, he has been a senior researcher in the Affiliated Institute of ETRI. His research interests include multimedia security, machine learning, and computer vision.

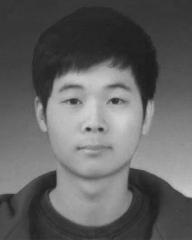


**SEUNG-HUN NAM** received the B.S. degreein Information Communication Engineering from Dongguk University, Seoul, South Korea, in 2013, and the M.S. degree in School of Computing from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2015. He is currently pursuing the Ph.D. degree in Multimedia Computing Lab., School of Com-puting, KAIST. His research interests include var-ious aspects of digital watermarking, multimedia



forensics, and deep learning.

**IN-JAE YU** received the B.S. degree in De-partment of Mathematical Sciences from Korea Advanced Institute of Science and Technology (KAIST), Korea, in 2015. He received his M.S. degree in School of Computing from KAIST, in 2017. He is currently working toward his Ph.D. degree in Multimedia Computing Lab., School of Computing, KAIST. His research interests include digital multimedia forensics and deep learning.



**HEUNG-KYU LEE** received the B.S. degree inelectronics engineering from Seoul National Uni-versity, Seoul, South Korea, in 1978, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Tech-nology (KAIST), South Korea, in 1981 and 1984, respectively. Since 1986, he has been a Professor with the School of Computing, KAIST. He has authored/co-authored over 200 international jour-nal and conference papers. His major interests are



digital watermarking, digital fingerprinting, and digital rights management. He has been a Reviewer of many international journals, including the Journal of Electronic Imaging, Real-Time Imaging, and the IEEE Transactions on Circuits and Systems for Video Technology. His major interests are information hiding, digital watermarking, and multimedia forensics.



10 VOLUME xx, 2020

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.