Intelligent Automation & Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.041181

ARTICLE

Deep Learning-Based Digital Image Forgery Detection Using Transfer Learning

Emad Ul Haq Qazi[1,*], Tanveer Zia[1], Muhammad Imran[2] and Muhammad Hamza Faheem[1]

[1]Center of Excellence in Cybercrimes and Digital Forensics (CoECDF), Naif Arab University for Security Sciences (NAUSS), Riyadh, 14812, Saudi Arabia

[2]School of Engineering, Information Technology and Physical Sciences, Federation University, Brisbane, QLD 4000, Australia

*Corresponding Author: Emad Ul Haq Qazi. Email: qabdulrab@nauss.edu.sa

Received: 13 April 2023 Accepted: 26 June 2023 Published: 27 February 2024

ABSTRACT

Deep learning is considered one of the most efficient and reliable methods through which the legitimacy of a digital image can be verified. In the current cyber world where deepfakes have shaken the global community, confirming the legitimacy of a digital image is of great importance. With the advancements made in deep learning techniques, now we can efficiently train and develop state-of-the-art digital image forensic models. The most traditional and widely used method by researchers is convolution neural networks (CNN) for verification of image authenticity but it consumes a considerable number of resources and requires a large dataset for training. Therefore, in this study, a transfer learning based deep learning technique for image forgery detection is proposed. The proposed methodology consists of three modules namely; preprocessing module, convolutional module, and the classification module. By using our proposed technique, the training time is drastically reduced by utilizing the pre-trained weights. The performance of the proposed technique is evaluated by using benchmark datasets, i.e., BOW and BOSSBase that detect five forensic types which include JPEG compression, contrast enhancement (CE), median filtering (MF), additive Gaussian noise, and resampling. We evaluated the performance of our proposed technique by conducting various experiments and case scenarios and achieved an accuracy of 99.92%. The results show the superiority of the proposed system.

KEYWORDS

Image forgery; transfer learning; deep learning; BOW dataset; BOSSBase dataset

1 Introduction

Digital imaging is becoming an essential component of our daily lives with the advancement of modern technologies. Images are incorporated into our daily lives, and detecting whether these images are forged or not is a challenging problem. Different type of techniques are employed to alter the quality of images in terms of dimensions or pixel density. Researchers are actively involved in developing new methodologies to detect image manipulations and restoring techniques. Some researchers have used statistical fingerprints for the authenticity of digital data without access to the

source, but it may be destroyed by different manipulations. The techniques widely used by researchers are as follows:

i) Resampling
ii) JPEG compression
iii) Contrast Enhancement (CE)
iv) Median Filtering (MF)
v) Additive Gaussian Noise (AGN)

Resampling is a technique that depends on empirical analysis of data, instead of parametric theory. Resampling has the same goal as a parametric statistical test to make an inferential decision [1]. Resampling differs from parametric theory in how the same goal is achieved using two different approaches. There are three ways to achieve resampling: jackknife, bootstrap, and permutation test. Bootstrap is considered an effective and commonly used technique.

JPEG is an image compression standard that was accepted as an international compression standard in 1992. Discrete Cosine Transformation (DCT) is used by JPEG compression for coding transformations. It is a compression technique that provides a high degree of compression and reduces the size of content, but some video frames, sound waves, and pixels could be removed.

The contrast enhancement (CE) technique is widely used for improving image quality. In [2], the authors proposed a technique that computes the local mean of the 3×3 sub-image and a global mean of the complete image. Then, the smoothing of the image is done using a local mean filter using the mean values of neighbor images.

Median filtering (MF) is a digital filtering technique that is used to remove the noise from a signal or image. It is often utilized as a preprocessing step for later steps. This technique runs through each pixel one by one and replaces that entry with the median of neighboring members. This approach is known as the sliding window approach.

Additive Gaussian noise is a type of noise that is intrinsic to the information system and Adding White Gaussian Noise (AWGN) is a model used in information technology to copy the behavior that occurs in nature.

The above techniques can be used to verify the authenticity of the images [3–6]. With the development of modern techniques such as CNN [7], which is used as a core architecture for many methods. Most of the generic methods lose potential information such as pixel dependency. Many models use the preprocessing layer as an input to the CNN model despite the original images. The algorithms used for preprocessing fail to remove all the image contents. Features extracted are data-dependent leading to less generalizability when the same model is applied to different databases. There is an overhead of the size of the dataset as well as a considerable amount of data is required to train the model.

To train a deep neural network, we need highly efficient computational resources that require a huge amount of data to learn for a considerable amount of time. The human brain also uses the knowledge between the tasks to identify the characteristics of an object and past learning experiences are applied to encountering new tasks. Similarly, transfer learning is a mechanism to transfer knowledge in one or more tasks and utilize it for the identification of related target tasks.

In this paper, we propose a novel deep learning-based approach using transfer learning, which benefits from the weights that are learned from previously trained datasets without having to

explicitly run these huge models and saving tremendous re-courses, and identifying multiple image manipulations.

The objective of this paper is to demonstrate an approach that is based on deep learning for image forgery that detects JPEG compression, contrast enhancement, median filtering, additive Gaussian noise, and resampling.

The model proposed in this study trains for steganalysis on BOSSBase. After training the models, we apply transfer learning to the forensic model from the steganalysis model. With this approach, our proposed model trains a forensic model by employing a transfer learning-based approach which benefits in terms of increasing performance and reducing the training time. We also propose a feature transfer strategy among different data sources, a well-trained model on BOSSBase serves as a source for training a model on new data sources which essentially uses a small amount of data and requires minimal iterations. Our main contributions are given below:

- Proposed deep learning based efficient and robust image forgery detection system.
- Detection of JPEG compression, contrast enhancement, median filtering, additive Gaussian noise, and resampling using proposed approach.
- Presented a comprehensive analysis of existing state-of-the-art techniques.
- Validation and comparison of the proposed model with existing techniques.

The rest of the paper is organized as follows: Section 2 presents the literature review; Section 3 describes the proposed technique based on CNN and transfer learning. Section 4 presents the experimental results and discussion. Section 5 concludes the study.

2 Literature Review

Machine learning is taking over the current era and is transforming the current era into a revolutionary practical world where intelligent machines perform tasks that were done by humans in the past. The different multimedia forensics models have been proposed in previous research studies. In the following paragraphs, we review the state-of-the-art techniques that have been proposed for image forgery detection.

Liu et al. [8] recognized content-based image copy detection using CNN. The authors detected unauthorized digital images by transferring the image dataset by several image processing manipulations. The proposed scheme works efficiently for recognizing images with scaling, rotation, and other manipulations.

The second approach that most of the researchers used is to extract the features from images using hand-crafted algorithms and pass extracted features into CNN. Generally, the CNN model does not cover statistical properties, it considers tempered one and original as the same. The preprocessing layer at the start of the CNN model has been added by many researchers in the area of digital image forensics. Chen et al. [6] used the median filtering-based CNN model to detect images that have been manipulated by using the first layer of the CNN model which accepts the image as an input and outputs the median filtering residual (MFR). In this study, they worked on a new type of CNN called constraint convolutional layer which can jointly suppress and adapt image content and adaptably learn the manipulation detection features [9]. They developed and trained their model on the Wild Testing Dataset which consisted of 50,000 grayscale patches and 8,350 unedited patches.

In another study, Younis et al. [10] proposed an approach for training image forgery detection models. The proposed model applies prior knowledge which is transferred to the new model from

previous steganalysis models. The proposed approach obtained an accuracy of 94.8% and successfully accelerated conversions of the CNN model but the only limitation is it does not improve the image quality.

Yerushalmy et al. [11] proposed that no digital water marketing and comparison are required for the verification of image authenticity. The authors claimed that photo characteristics collected during the acquisition process provide evidence of legitimacy in and of themselves. The image artifacts are used as markers to evaluate the authenticity of an image. This technique achieved a significantly good result.

In another study, the transfer-based methodology was proposed by Zhan et al. [12]. The authors used the steganalysis model to gain prior knowledge. They achieved an accuracy of 97.36% on BOW and BOSSBase datasets. Barad et al. [13] presented a survey based on different deep learning models to evaluate the effectiveness of different machine learning models in different scenarios. They used different publicly available datasets for this purpose.

Wu et al. [14] detected image splicing forgery detection using ringed residual U-net (RRU-Net). They used an end-to-end segmentation network for image forgery detection. The main purpose of this study was to introduce a technique that was based on the human brain mechanism for developing RRU-Nets and can detect images without pre-processing and post-processing. The goal of this technique was to optimize the learning capacity of the proposed algorithm using a recall and consolidation mechanism. The experimental results show that this technique gives better results as compared to the traditional techniques.

Bayar et al. [4] presented a deep learning-based image forgery detection technique. The idea behind this technique was to identify different ways to identify and learn how image modifications are done. They proposed that modification of the image affects the neighbor pixels and focused on the local operational association between pixels, then detect the forgery in an image. Unsurprisingly, the cutting-edge solutions are deep learning-based, with a special focus on pixel-level modification detection [15–18]. The researchers however consider only two classes legitimate and manipulated making the task appear to be of semantic segmentation based on Images.

Dirik et al. [19] proposed a technique that uses a color filter for the detection of image tampering. It computes a simple threshold-based classifier and tested their approach over authentic and benchmark computer-generated datasets. The experimental analysis shows that the proposed technique performs better as compared to traditional image tampering techniques.

In similar research studies [20–22], the authors observed the changes in the image acquisition phase by saving the images in compressed format. When the image is being processed, there are some traces left on the image which can be used for verification via digital authenticity.

Zhuang et al. [23] proposed an image tampering localization technique using a dense fully connected convolutional network. The authors focused on detecting editing tools using Photoshop in order to address the issue of localization of tampered images. Furthermore, the authors develop a data generation methodology for generating training da-ta using Photoshop scripting, which may mimic human actions and create large-scale training examples.

The overview of the state-of-the-art methods given above indicates that most of the existing methods do not give good performance for image forgery. They are using techniques that do not extract the discriminative information and their performance depends on the tuning of various parameters. In view of the decisive victory of DL over the other methods, the DL-based method can be employed to improve the generalization and accuracy of an image authentication system.

3 Proposed Technique Based on Deep Learning and Transfer Learning

Deep learning and transfer learning are powerful techniques in the field of detecting cyber-attacks, particularly in the context of image forgery. Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) play a pivotal role in these approaches. Here, we propose a technique that combines deep learning, transfer learning, and CNNs to detect image forgery.

3.1 Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) is an Artificial Neural Network (ANN) that has been used for analyzing and detecting images in various fields [24–26]. CNN is widely used for image analysis but there are some other datasets for which CNN can be used for classification and other problems as well. CNN is an artificial neural network that has the specialization to detect and select image patterns and gives output containing information from them. This feature of CNN makes it more useful for image analysis. A CNN has convolutional layers called hidden layers that make it well-performing for image diagnosis. With these hidden layers, now CNN has multiple layers, but the major layers of CNN are convolutional layers that take the convolution of image matrices.

The convolution layers take an input image and transform the output. Then the out-put is fed into the next convolutional layer. This whole process is based on the convolution operation. A CNN can detect patterns more precisely with the help of filters. The patterns may include shape textures, objects, multiple edges, etc. These are the filters that can detect the edges and that filter is named edge detector. Some filters may detect circles; some detect corners or patches, etc. Hence, based on the difference between neighbor pixel values, the CNN model detects tampering in images. The schematic diagram of the proposed deep learning-based system is shown in Fig. 1.

CNN is a multi-layer deep learning model which has linear stacked residual connections called local receptive field used for better efficiency and its performance can be improved by weight sharing. This architecture helps in learning many complex features of input that simple machine learning struggles to do. The main feature of CNN architecture is that it can get local features from input i.e., an image at higher layers, and combine them at lower layers for complex features.

Since CNN is a deep neural network containing multiple layers, each layer connects every neuron from the previous layer to every neuron in the next layer to identify the particular class correctly. The activation function takes the output from the last connected layer and outputs the score of each class.

3.2 Proposed Architecture

The proposed deep learning based digital image forgery detection system is based on transfer learning and a convolution module. Fig. 1 shows the proposed CNN architecture. We denote an image as a Xi and pass it to the preprocessing layer P. The preprocessed image P is then forwarded to the convolution module $conv_m$ which interin is then prosessed through convolution groups $conv_{m1} \ldots conv_8$ after which the classification module clf_m makes the classification.

Our proposed deep learning-based approach is an extension of the CNN architecture that is presented in [27] and shown in Fig. 1. The proposed architecture has the following modules:

i) Preprocessing module
ii) Convolutional module
iii) Classification module

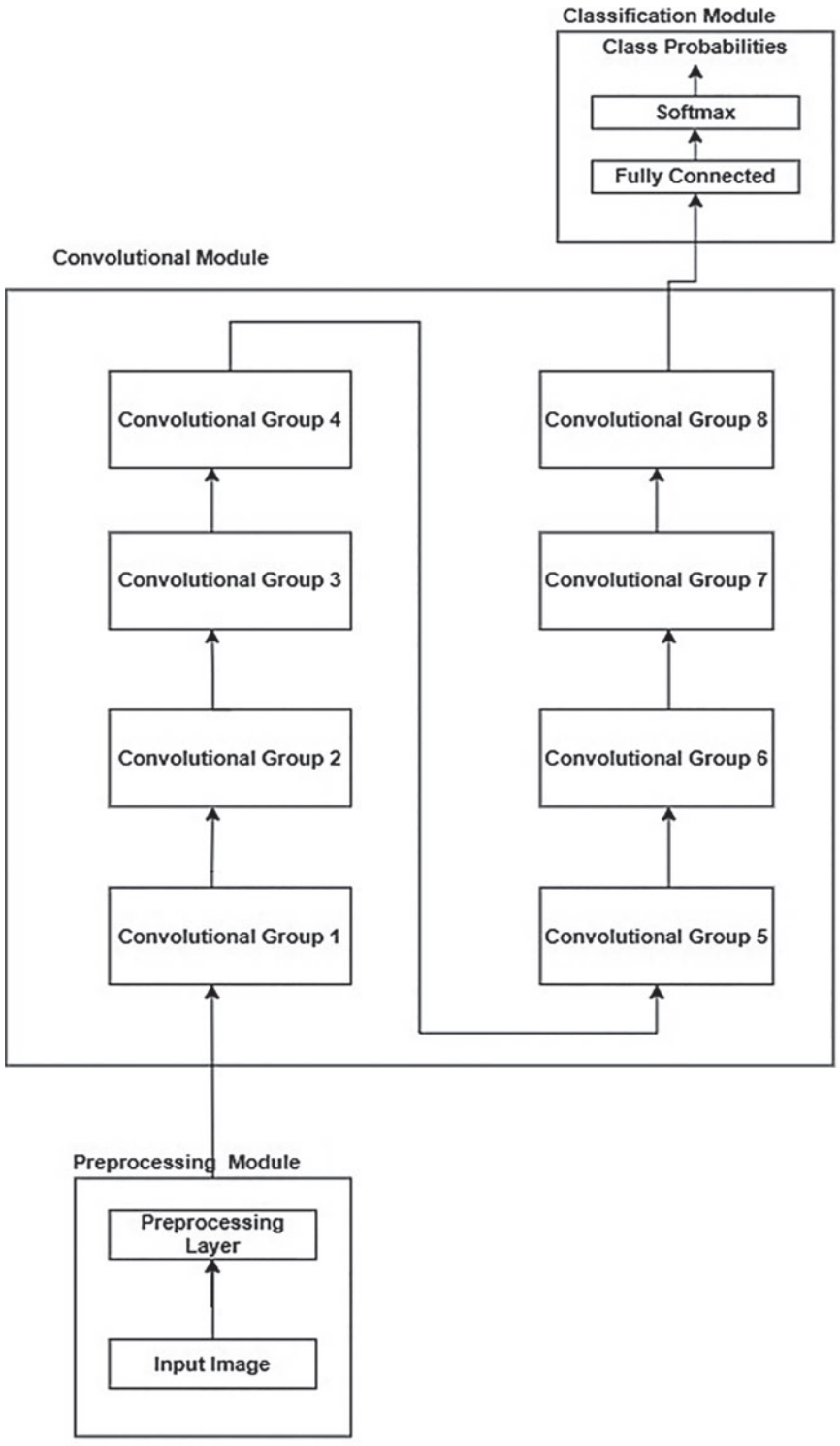

Figure 1: Proposed CNN architecture

The preprocessing module takes an input image that is passed to preprocessing layer which will enhance the stego noise. The convolutional module will receive the input from preprocessing layer on which it applies six different convolution layer blocks as shown in Fig. 1. The convolutional module outputs a 256-D (dimension) to a fully connected layer of the classification module with an n-way SoftMax layer (the network decides based on the output) and produces n-class labels.

3.3 Architecture Discussion

Transfer learning is a technique in machine learning where a previously trained model is used as the foundation for a model on a new task. In transfer learning, a model that has been trained for one job is reused for a different, similar task as an optimization to facilitate quick progress while modeling the new task [28,29]. It is the concept of trans-fer-ring knowledge in terms of weights learned from one model to another model, which are learned on millions of images containing thousands of objects. The advantage of employing transfer learning is that we utilize learned weights on large datasets efficiently such as ImageNet, to our specific models without having to train them. Furthermore, we also reduce the complexity, and cost of our model by utilizing the pre-trained weights which reduce the cost and initialize the model with meaningful weights.

3.3.1 Case 1: Transfer of Parameters between Tasks

Considering the transfer between tasks, we essentially transfer the parameters to the forensic model from steganalysis. By using this approach, we come up with two scenarios, first, which part of the steganalysis model should be transferred, and second; the number of layers to be transferred. We employ the standard approach for transfer learning by training the base network and then transferring the first n layers to the end network, the rest of the layers of the end network are initialized randomly. Generally, shallow layers are transferred between tasks as they are more general compared to the deep layers, hence increasing the transferability.

Fig. 2 shows the proposed architecture based on deep learning. In Fig. 2a, we propose a steganalysis model which is based on the standard backpropagation that takes the input X, the steganalysis model is based on the BOSSBase dataset. The rectangular boxes are labeled as parameters that are used for preprocessing which are kept the same during the entire experiment. We are training the models on different datasets which are shown in Fig. 2 in different colors. The white line/bar represents the activation layer, batch normalization, and pooling layers in between the convolution layers.

Fig. 2b shows the proposed forensic model which has almost similar architecture to the above-discussed steganalysis model apart from the fully connected layer. In the proposed model, the fully connected layer of model Y consists of six neurons in the multi-class classification task and two neurons in the binary-class classification task. The BOSSBase dataset is used for creating the input Y for the image forensics task. The parameter transfer strategy from the steganalysis model to the forensic model is shown in Fig. 2c. First n layers of weights of the forensic model are used from the steganalysis model whereas the final 8-n layers are being initialized randomly. We have trained the entire network on the input Y keeping in mind that the parameters are kept as same in the pre-processing layer in the target model.

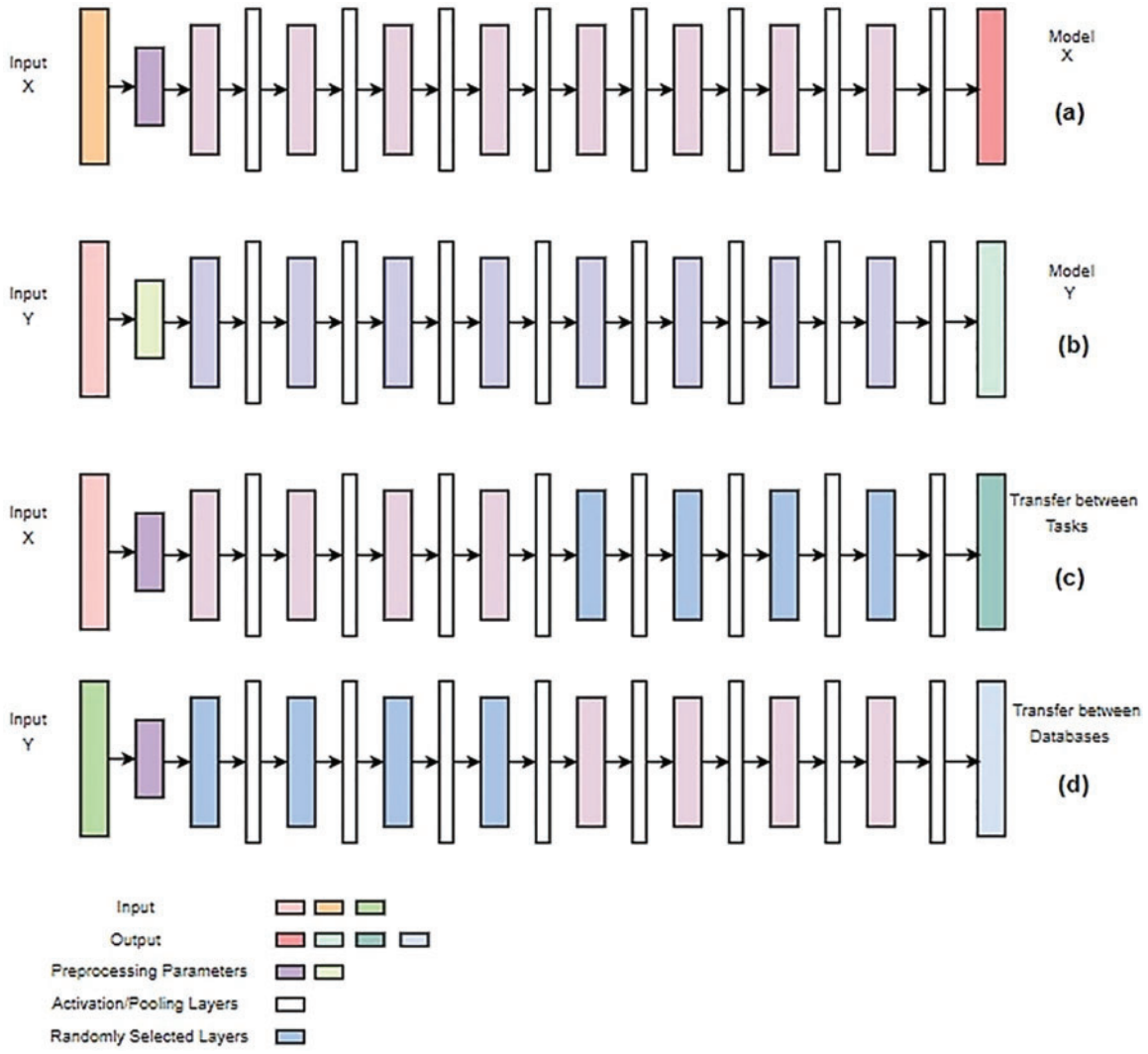

Figure 2: Proposed setting. (a) Trained model (b) untrained model (c) feature transfer among tasks (d) feature Transfer between databases

3.3.2 Case 2: Transfer of Parameters between Databases

The second part of our work focuses on the transfer of parameters among the two forensic models learned from separate datasets. Transfer learning is employed in forensic tasks; the training costs have been reduced as the tasks are similar to one another while using multiple databases.

Labeling the end data is very expensive. The data needs to be labeled when the targets are achieved, so the network is fine-tuned using labeled training data. For the domain adaptation, we have used the technique discussed in [30] to reduce the similarity between domains. Apart from this we also need to train an optimized model to obtain reasonable accuracy. Since we aim to meet both conditions, we retained the shallow layers as they have a larger proportion concerning deep layers. Then we use a shallow layer to train the deep layer classifier. As shown in Fig. 2d, we are considering the use of the first and the last 8-n layers from the network given model X, whereas the rest of the layers have been

randomly initialized. After initializing, the network is trained on the dataset Y which has been created from the BOW dataset.

4 Experimental Results and Discussion

In this study, we used a benchmark BOSSBase dataset for training and validation purposes. The experiments are conducted with transfer learning between a task that contains 10,000 grayscale images. We tested the transfer learning approach between datasets on the benchmark BOW dataset, which also contains 10,000 grayscale images. Various image manipulations are used to produce the tampered files.

i) JPEG files are generated using JPEG compression.
ii) Images are filtered using 3×3 and 5×5 kernels.
iii) Bilinear interpolation with scaling factors (SF) is used for resampled images resampling (resizing) ranging between 1.1 and 1.5.
iv) Gamma correction is used to generate Contrast Enhancement images with $\gamma = 0.4$ and $\gamma = 2.0$.
v) Datasets of Gaussian noise are generated by the process of Adding White Gaussian Noise (AWGN) while setting the value of standard deviation to 1.0 and 2.0, respectively.

We used parameter transfer for the steganalysis model to detect S-UNIWARD (universal distortion function that hides messages in pixel values) with 0.4 bpp on the BOSSBase dataset. Data embedding is then used to generate stego images. Since the dataset contains 10,000 images, we divided the dataset equally into training and testing samples, i.e., 5000 images each for training and testing purposes.

For implementation, we used an Intel i7 processor with 64 GB of RAM and an Nvidia GTX 1060 GPU to conduct the experiments. The training parameters are set as follows:

Learning rate lr = 0.0001, decay = 0.004, momentum = 0.9 and batch set = 50 images. Overfitting is minimized by utilizing regularization.

4.1 Case 1: Transfer of Parameters between Tasks

The performance of the transfer learning-based model against the number of layers transferred is shown in Fig. 3a. We have calculated the average testing accuracy for all the binary classification models after 100 iterations. Here, n is indicating the convolution layers which were transferred. The learning rate is 0.0001 was adopted during the training phase.

After conducting a set of experiments, eleven models are considered for the binary classification task which has achieved the highest accuracy while testing the aforementioned image manipulation. A random selection of 3000 images for training purposes, whereas 5000 images were used in the testing phase. By the experimental analysis, we conclude that our proposed approach can be trained by using small amounts of data along with the parameters which are transferred. We have mentioned the accuracy scores in Table 1. Except for the contrast enhancement (CE) detection. When CNN is employed with transfer learning, the model converges quite faster when compared to training the model from scratch as the transfer of knowledge takes place. Fig. 3c represents the learning curve that is used for the JPEG detection to observe as an example. We can observe from Fig. 3c that the model converges faster. The loss curves are represented in Fig. 3d. The loss is minimal when we use transfer learning by using the pre-trained weights.

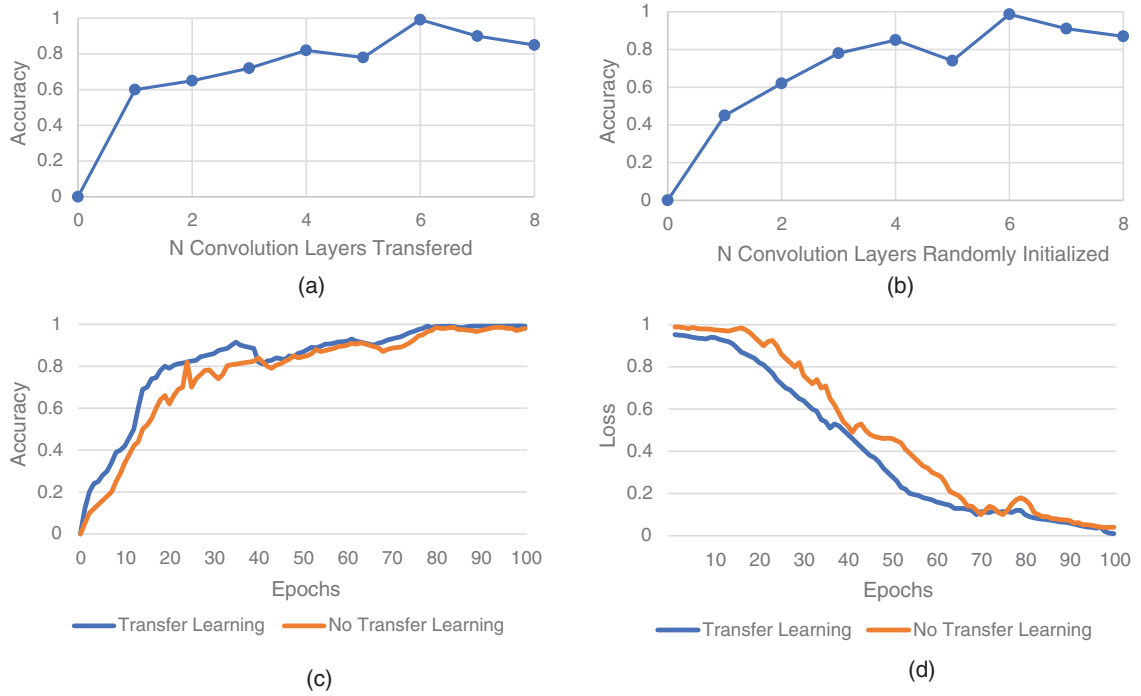

Figure 3: Performance evaluation using transfer learning (a). Transfer among tasks, (b). Transfer among databases, (c). Accuracy comparison between with and without transfer learning, (d). Loss with and without transfer learning

Table 1: Testing accuracy on BOSSBase for binary classifier

Modification	JPEG (70)	JPEG (80)	JPEG (90)	Median filtering (3×3)	Median filtering (5×5)	Contrast enhancement ($\gamma = 0.4$)
Accuracy	99.92%	99.92%	99.72%	99.92	99.92	85.7%
Modification	Contrast enhancement ($\gamma = 2.0$)	Resampling (SF = 1.1)	Resampling (SF = 1.1)	AWGN ($\sigma = 1$))	AWGN ($\sigma = 2$)	
Accuracy	95.70	98.9%	99.50%	99.92%	99.92%	

A slight change in the CNN architecture was made for multi-class classification where we changed the number of output neurons to 6. A total set of 70,000 images were used for training and testing where 20% of the images were used as training data and the rest as testing data. We have mentioned the accuracy scores in Table 3 and achieved an average accuracy of 97.79%.

4.2 Case 2: Transfer of Parameters between Databases

In Fig. 3b, we observed the accuracy peak at n = 5. The performance of the BOW dataset is shown in Tables 2 and 4. using the same experimental settings. Table 1 represents the variation in accuracy after modifications in the environment and Table 3 similarly depicts the performance of our proposed approach on the BOSSBase dataset.

Table 2: Testing accuracy on BOW for binary classifier

Modification	JPEG (70)	JPEG (80)	JPEG (90)	Median filtering (3×3)	Median filtering (5×5)	Contrast enhancement ($\gamma = 0.4$)
Accuracy	99.85%	99.85%	99.42%	99.92	99.92	85.40%
Modification	Contrast enhancement ($\gamma = 2.0$)	Resampling (SF = 1.1)	Resampling (SF = 1.1)	AWGN ($\sigma = 1$))	AWGN ($\sigma = 2$)	
Accuracy	95.90	98.32%	99.80%	99.92%	99.92%	

Table 3: Testing accuracy on BOSSBase for multi classifier

Test/Prediction	Original	JPEG (70)	Median filtering (3×3)	Contrast enhancement ($\gamma = 0.4$)	Resampling (SF = 1.1)	AWGN ($\sigma = 2$)
Original	99.10%	0.72%	0.21%	0.02%	0.05%	0.13%
JPEG (70)	0.025%	99.92%	0.02%	0.07%	0.02%	0.03%
Median filtering (3×3)	0.06%	0.01%	99.92%	0.03%	0.05%	0.02%
Contrast enhancement ($\gamma = 0.4$)	16%	0.2%	0.3%	88%	1.5%	0.2%
Resampling (SF=1.1)	0.2%	0.1%	0.01%	0.18%	99.87%	0.15%
AWGN (σ =2)	0.00%	0.05%	0.04%	0.02%	0.08%	99.95%

Table 4: Testing accuracy on BOW for multi classifier

Test/Prediction	Original	JPEG (70)	Median filtering (3×3)	Contrast enhancement ($\gamma = 0.4$)	Resampling (SF = 1.1)	AWGN ($\sigma = 2$)
Original	99%	0.60%	0.52%	0.35%	0.00%	0.1%
JPEG (70)	0.04%	99.92%	0.00%	0.07%	0.05%	0.05%
Median filtering (3×3)	0.23%	0.10%	99.60%	0.05%	0.35%	0.08%
Contrast enhancement ($\gamma = 0.4$)	16%	0.75%	0.02%	85%	1.30%	0.04%
Resampling (SF = 1.1)	0.05%	0.06%	0.30%	0.20%	99.70%	0.07%
AWGN ($\sigma = 2$)	0.00%	0.05%	0.01%	0.02%	0.1%	99.90%

4.3 Comparison

We compare the results of our deep learning based proposed technique with existing techniques that are being evaluated on the CASIA v2, DVMM data sets. Table 5 shows the accuracy comparison of our deep learning based proposed technique. Researchers have used various techniques to perform image forgery detection in order to detect images which have been tampered. Most studies only

proposed an approach to detect a single type of image tampering detection mainly image splicing and copy paste/ cut move. Previous studies have employed approaches namely: CNN, RNN, MFCN, edge probability map, RRU-Net and wavelet decomposition to name a few.

Table 5: Comparison with existing techniques

Ref.	Targeted tampering	Method	Dataset	Accuracy (%)
[12]	AWGN, Gaussian blurring, median filtering	CNN, error predictions	Images from 12 different cameras	99.10
[31]	Image splicing	DCT coefficient analysis pattern recognition	CASIA v2	90.1
[32]	Image splicing	Steerable pyramid transform and local binary pattern	CASIA v1	94.89
[33]	Image splicing	Markov features	DVMM	93.55
[34]	Copy move, cut-paste	CNN	Columbia DVMM, CASIA v1 and CASIA v2	98.04
[35]	Copy move, cut-paste	Mask R-CNN, ResNet-101	Cover, Columbia	For Cover Avg Precision 93 and for cover 97
[36]	Image splicing	MFCN, edge probability map	CASIA v1	For CASIA v1 0.52 MCC score
[37]	Cut paste	CNN	Dresden database	Localization accuracy 81 & Detection accuracy 82
[38]	JPEG double compression, cut-paste	Features through multi-domain CNN	UCID	95
[39]	Cut-paste	Landscapes with noise and Autoencoder	Images which are taken from 7 electronic devices	0.41 F-measure
[40]	Cut-paste	Image residuals and RRU-Net	CASIA, Columbia	93.94
[41]	Cut-paste, copy-move	Daubechies wavelet decomposition and SAE	CASIA v1, Columbia and CASIA v2	90.09
[42]	Attacks inform of the combination of transformations	CNN and AlexNet	MICC-F220	93.94
[6]	Median filtering and cut paste	Median filtering residuals and CNN	Boss base, BOSS RAW, NRCS Gallery UCID, Dresden	85.14
Proposed work	Resampling, JPEG compression, contrast enhancement, median filtering, and additive Gaussian noise	CNN	BOSSBase [43]	99.92

Our proposed model is successfully able to detect five most widely used tampering techniques namely; Resampling, JPEG compression, Contrast enhancement, Median filtering, and Additive gaussian noise with the highest accuracy of 99.92% using CNN based approach.

5 Conclusion and Future Work

Digital image forgery detection using deep learning is considered one of the most efficient and reliable mechanisms for the verification of the legitimacy of an image. In this study, a deep learning-based approach is proposed for forensic analysis where we have shown the transfer of knowledge between steganalysis and forensic models. We also discuss our proposed strategy of transfer learning when applying the forensic model to some other databases. For the task of parameter transferring, our proposed approach rectifies the problem of degradation by fine-tuning the model with some labeled data that is taken from a new database. We also reduced the complexity of our model by utilizing the pre-trained weights. We evaluate the performance of our technique based on benchmark datasets, i.e., BOW and BOSSBase that detect five forensic types which include JPEG compression, contrast enhancement, median filtering, additive Gaussian noise, and resampling. By conducting various experiments, we evaluate our proposed method and achieve the highest accuracy of 99.92%. The experimental model shows good results but also boosts the training process. The proposed technique could be further extended to transfer knowledge between multiple multimedia forensic tasks. The comparison with existing approaches demonstrates the superiority of the suggested system. The suggested approach will aid in the identification of image alterations and open the way for future studies into identifying numerous forms of image forgery manipulations.

In the future, we hope to conduct additional research on the reduction of input characteristics utilizing techniques such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Autoencoders, and so on. Future research will also investigate the usage of RNN, LSTM, and GRU-based architectures.

Acknowledgement: The authors would like to express their deep thanks to the Vice Presidency for Scientific Research at Naif Arab University for Security Sciences for their kind encouragement of this work.

Funding Statement: This work was supported by Security Research Center at Naif Arab University for Security Sciences (Project No. SRC-PR2-01).

Author Contributions: Conceptualization, Emad Ul Haq Qazi, and Tanveer Zia; methodology, Emad Ul Haq Qazi; investigation, Emad Ul Haq Qazi, Tanveer Zia, Muhammad Imran, and Muhammad Hamza Faheem; resources, Tanveer Zia; software, Emad Ul Haq Qazi; validation, Emad Ul Haq Qazi, Tanveer Zia, Muhammad Imran, and Muhammad Hamza Faheem; visualization, Emad Ul Haq Qazi, Tanveer Zia, Muhammad Imran, and Muhammad Hamza Faheem; writing–original draft preparation, Emad Ul Haq Qazi; writing–review and editing, Emad Ul Haq Qazi, Tanveer Zia, Muhammad Imran, and Muhammad Hamza Faheem; supervision, Tanveer Zia, and Muhammad Imran; project administration, Tanveer Zia; funding acquisition, Tanveer Zia. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Materials: Not applicable.

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

References

[1] W. H. Beasley and J. Rodgers, "Resampling methods," *The Sage Handbook of Quantitative Methods in Psychology*, vol. 1, no. 3, pp. 362–386, 2013.

[2] A. Singh, S. Yadav and N. Singh, "Contrast enhancement and brightness preservation using global-local image enhancement techniques," in *Fourth Int. Conf. on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, Solan, Himachal Pradesh, India, pp. 291–294, 2016.

[3] L. Baroffio, L. Bondi, P. Bestagini and S. Tubaro, "Camera identification with deep convolutional networks," arXiv preprint arXiv:1603.01068, 2016.

[4] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proc. of the 4th ACM Workshop on Information Hiding and Multimedia Security*, Galicia, Vigo, Spain, pp. 5–10, 2016.

[5] C. Chen, Y. Q. Shi and W. Su, "A machine learning based scheme for double JPEG compression detection," in *2008 19th Int. Conf. on Pattern Recognition*, Tampa, Florida, USA, pp. 1–4, 2008.

[6] J. Chen, X. Kang, Y. Liu and Z. J. Wang, "Median filtering forensics based on convolutional neural networks," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1849–1853, 2015.

[7] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," in *Int. Conf. on Engineering and Technology (ICET)*, Peshawar, Pakistan, pp. 1–6, 2017.

[8] X. Liu, J. Liang, Z. Y. Wang, Y. T. Tsai, C. C. Lin *et al.,* "Content based image copy recognition using CNN," *Electronics*, vol. 12, no. 9, pp. 2029, 2020.

[9] B. B. Bayar and M. C. Stamm, "Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2691–2706, 2018.

[10] A. Younis, I. Tariq and S. M. Shehata, "Image forgery detection based on deep transfer learning," *European Journal of Electrical Engineering and Computer Science*, vol. 3, no. 5, pp. 125, 2019. https://doi.org/10.24018/ejece.2019.3.5.125

[11] I. Yerushalmy and H. Hel-Or, "Digital image forgery detection based on lens and sensor aberration," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 71–91, 2011.

[12] Y. Zhan, Y. Chen, O. Zhang and X. Kang, "Image forensics based on transfer learning and convolutional neural network," in *Proc. of the 5th ACM Workshop on Information Hiding and Multimedia Security*, Philadelphia, Pennsylvania, USA, pp. 165–170, 2017.

[13] Z. J. Barad and M. M. Goswami, "Image forgery detection using deep learning: A survey," in *6th In. Conf. on Advanced Computing and Communication Systems*, Coimbatore, India, pp. 571–576, 2020.

[14] Y. Wu, W. A. Almageed and P. Natarajan, "BusterNet: Detecting copy-move image forgery with source/-target localization," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 168–184, 2018.

[15] X. Hu, Z. Zhang, Z. Jiang, S. Chaudhuri, Z. Yang *et al.,* "SPAN: Spatial pyramid attention network for image manipulation localization," in *ECCV*, vol. 2, pp. 312–328, 2020.

[16] G. Mahfoudi, B. Tajini, F. Retraint, F. Morain-Nicolier and M. Pic, "DEFACTO: Image and face manipulation dataset," in *EUSIPCO*, Coruna, Spain, pp. 1–5, 2019.

[17] Y. Wu, W. AbdAlmageed and P. Natarajan, "ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features," in *CVPR*, Long Beach, CA, USA, pp. 9543–9552, 2019.

[18] P. Zhou, B. Chen, X. Han, M. Najibi and L. Davis, "Generate, segment, and refine: Towards generic manipulation segmentation," in *Proc. of the AAAI Conf. on Artificial Intelligence*, New York, USA, vol. 34, no. 7, pp. 13058–13065, 2020.

[19] A. E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," in *Proc. of the Int. Conf. on Image Processing (ICIP)*, Cairo, Egypt, pp. 1497–1500, 2009.
[20] H. Farid, "Image forgery detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, 2009.
[21] T. Van Lanh, K. S. Chong, S. Emmanuel and M. S. Kankanhalli, "A survey on digital camera image forensic methods," in *Proc. of the IEEE Int. Conf. on Multimedia and Expo*, Venice, Italy, pp. 16–19, 2007.
[22] B. Mahdian and S. Saic, "A bibliography on blind methods for identifying image forgery," *Signal Processing: Image Communication*, vol. 25, no. 6, pp. 389–399, 2010.
[23] P. Zhuang, H. Li, S. Tan, B. Li and J. Huang, "Image tampering localization using a dense fully convolutional network," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2986–2999, 2021.
[24] T. Sercu, C. Puhrsch, B. Kingsbury and Y. LeCun, "Very deep multilingual convolutional neural networks for LVCSR," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, pp. 4955–4959, 2016.
[25] X. Zhang, J. Zhao and Y. LeCun, "Character-level convolutional networks for text classification," *Advances in Neural Information Processing Systems*, vol. 28, pp. 649–657, 2015.
[26] J. Tompson, R. Goroshin, A. Jain, Y. LeCun and C. Bregler, "Efficient object localization using convolutional networks," in *Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 648–656, 2015.
[27] G. Xu, H. Z. Wu and Y. Q. Shi, "Ensemble of CNNs for steganalysis: An empirical study," in *Proc. of the 4th ACM Workshop on Information Hiding and Multimedia Security*, Vigo Galicia, Spain, ACM, pp. 103–107, 2016.
[28] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu *et al.,* "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
[29] R. Liu, Y. Shi, C. Ji and M. Jia, "A survey of sentiment analysis based on transfer learning," *IEEE Access*, vol. 7, pp. 85401–85412, 2019.
[30] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," arXiv preprint arXiv:1412.3474, 2014.
[31] Z. Lin, J. He, X. Tang and C. K. Tang, "Automatic and fine-grained tampered jpeg image detection," *Via DCT Coefficient Analysis Pattern Recognition*, vol. 42, no. 11, pp. 2492–2501, 2009.
[32] G. Muhammad, M. Al-Hammadi, M. Hussain and G. Bebis, "Image forgery detection using steerable pyramid transform and local binary pattern," *Machine Vision and Applications*, vol. 25, pp. 1–11, 2013.
[33] Q. Zhang, W. Lu, R. Wang and G. Li, "Digital image splicing detection based on Markov features in DCT and DWT domain," *Multimedia Tools and Applications*, vol. 45, no. 12, pp. 4292–4299, 2012.
[34] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *2016 IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Abu Dhabi, The United Arab Emirates, pp. 1–6, 2016.
[35] X. Wang, H. Wang, S. Niu and J. Zhang, "Detection and localization of image forgeries using improved mask regional convolutional neural network," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 4581–4593, 2019.
[36] R. Salloum, Y. Ren and C. C. J. Kuo, "Image splicing localization using a multi-task fully convolutional network (MFCN)," *Journal of Visual Communication and Image Representation*, vol. 51, pp. 201–209, 2018.
[37] L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J. Delp *et al.,* "Tampering detection and localization through clustering of camera based CNN features," in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, Hawaii, USA, vol. 2, pp. 1855–1864, 2017.
[38] I. Amerinia, T. Uricchioa, L. Ballana and R. Caldellia, "Localization of JPEG double compression through multi-domain convolutional neural networks," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, Hawaii, USA, pp. 1865–1871, 2017.
[39] D. Cozzolino and L. Verdoliva, "Single-image splicing localization through autoencoder-based anomaly detection," in *IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Abu Dhabi, The United Arab Emirates, pp. 1–6, 2016.

[40] Y. Wei Bi, B. Xiao and W. Li, "RRU-Net: The ringed residual U-Net for image splicing forgery detection," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, 2019.

[41] Y. Zhang, J. Goh, L. Win and V. Thing, "Image region forgery detection: A deep learning approach," in *Proc. of the Singapore Cyber-Security Conf. (SG-CRC)*, Singapore, pp. 1–11, 2016.

[42] A. Doegara, M. Dutta and G. Kumar, "CNN based image forgery detection using pre-trained AlexNet model," in *Proc. of Int. Conf. on Computational Intelligence & IoT (ICCIIoT)*, NERIST and NIT Mizoram, India, no. 1, 2018.

[43] T. Penvy, T. Filler and T. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Information Hiding: 12th Int. Conf., IH 2010*, Calgary, Canada, vol. 12, pp. 161–177, 2010.