Evolving Systems (2023) 14:545–556
https://doi.org/10.1007/s12530-022-09446-0

ORIGINAL PAPER

MiniNet: a concise CNN for image forgery detection

Shobhit Tyagi[1] · **Divakar Yadav**[1]

Received: 14 February 2022 / Accepted: 6 June 2022 / Published online: 24 June 2022

Abstract
The exponential growth of technology has made images and videos popular digital objects. The increase in the number of visual imagery, crimes such as Identity theft, privacy invasion, fake news, etc. has also increased. The paper proposes a simple, easy-to-train, fully Convolutional Neural network, named MiniNet to detect forged images with high accuracy. The model is evaluated on existing image forgery datasets which consist of Authentic and tampered images. The proposed model achieved an accuracy of more than 95% for the 140 K Real and Fake Faces and 93% for CASIA datasets. Multiple Ablation studies are conducted on various state-of-the-art (SOTA) CNN models to check their performance on the given dataset. The objective is to assess the ability of CNN in detecting Image tampering. The experiments are done based on different aspects such as self-attention, positional encoding, and depth of the model. The minimal architecture used for image forgery detection is presented along with the performance achieved on different well-known datasets.

Keywords Image forgery detection · Deep learning · CNN

1 Introduction

Images have become an integral part of everyone's daily lives. From mobiles to computer screens, Images are present everywhere. Nowadays, Images are used in schools, and colleges, for presentations in offices, and as pictures on social media (Ghai et al. 2021). This increase in digital content also attracts malicious activities such as Identity theft, insurance fraud, fake news, and even forgery of documents with malicious intent. It is imperative to develop new approaches which are capable of detecting forged images from the real ones. Commonly used manipulations include Splicing, copy-move, and Object Removal. Figure 1 shows examples of Images manipulations.

(1) **Splicing** This technique copies parts from multiple images and pastes them onto another image. The purpose is to add information or element to an image. Splicing is the most common type of Image forgery technique that can be found on the internet. The sign on the building in Fig. 1b is cut from another image and then added to the original image. Because some regions of the tampered images come from other images, spliced images are easier to detect. There are visual differences between the Original and tampered images in Image attributes such as Shadow, lighting, sensor noise, and camera reflection.
(2) **Copy-Move** This manipulation is used for adding or hiding information from an image (Krishnaraj et al. 2022). For example: adding a region from the same image to hide or misrepresent the information provided by the original image. The person shown in the Fig. 1c is added from the same image.
(3) **Object Removal** It can be used to remove or replace unwanted objects/regions from an image. For example, filling missing parts of an image using Image inpainting. The person in Fig. 1d is missing as compared to the original image.

Some manipulation techniques depend upon the type of software used and the type of forgery one is wanted. Some of them are hard to detect even for an expert user. Moreover, with the development of GANs (Goodfellow et al. 2020 Oct 22) and Vision Transformers (Alexey et al. 2021), one is capable of generating perfect artificial images. These

✉ Shobhit Tyagi
shobhit.tya@gmail.com

Divakar Yadav
dsy99@rediffmail.com

[1] Department of Computer Science and Engineering, NIT, Hamirpur, HP, India

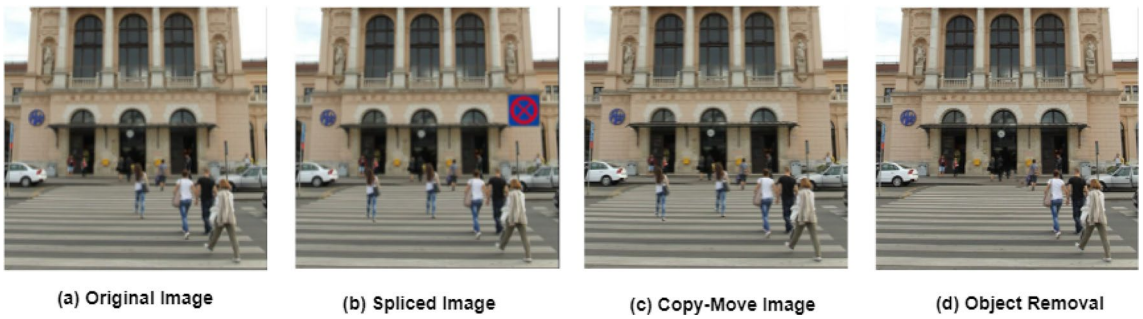

Fig. 1 Image Forgery Examples : From left to right **a** Original Image, **b** Spliced Image (Adding a sign), **c** Copy-Move Image (Duplicating a person), **d** Removal (Removing a person)

manipulated images can be used to deliver false information such as creating fake social media accounts in someone's name, tampering with evidence in a criminal matter, fake news, and mass manipulation, etc. In the last decade, the number of Cyber-crime is increased at a rapid speed. It is imperative to take necessary actions to avoid such crimes. Image forensics is the discipline that deals with the authenticity of an image or image content. It is also known as Forensic Image analysis. The forensics Community is trying to design or develop new approaches which are capable to counter Cyber-crimes. In recent years, there exist several methods for detecting and accessing the image content and authenticity (Stamm Matthew et al. 2013; Kharrazi et al. 2005). The paper addresses the problems caused by Image forgeries and proposes a Concise CNN model for detecting Image forgeries. The proposed architecture addressed two limitations: a) the model can detect forgeries irrespective of Image size and tampering attack, and b) the model is end-to-end trainable on new forgery datasets which help in generalization for new images.

In this work, the main contributions are as follows. First, a concise CNN called MiniNet is proposed. The model is inspired by the high image representational power of CNN and can detect image forgeries such as copy-move, splicing, and others. The network consists of optimal layers making it the most miniature image forensics model to detect forgeries with high accuracy. Second, because of the small size, the proposed architecture does not require Residual/Skip connections to avoid deep convolutional network problems (such as Vanishing gradient, overfitting). Furthermore, the network size also reduces the overall computational overhead and helps training faster. Finally, Experiments show that the proposed model is easier to generalize on other datasets and outperforms SOTA methods by a large margin. The proposed work mainly focuses on more common forgeries such as Splicing and Copy-move. The Proposed network can be generalized for different forgeries depending upon the dataset used for training.

The paper is structured as follows. Section 2 studies the existing methods for Image forgery detection, their advantages, and disadvantages. Section 3 introduces a Concise Convolutional Neural network (CNN) called MiniNet for generalized Image forgery detection. Section 4 discusses the training details and experimental results of MiniNet on benchmark models. Finally, Sect. 5 gives the conclusion of this paper.

2 Related work

This section focused on some of the earlier works on Image forgery detection based on Convolutional Neural networks (CNN's). In the 1980s, LeCun et al. (1989) developed the first CNN architecture to train handwritten digits. The same architecture is used in MNIST[1] handwritten number dataset in 1988. But due to lack of technological advancements in hardware and development of methods such as Bayesian Networks (Heckerman and Wellman 1995) and Support vector machines (Cortes and Vapnik 1995). In recent years, CNN's made a comeback in 2012 after winning the ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al. 2015) (ILSVRC). The ILSRVC won by AlexNet (Krizhevsky et al. 2012) created a new pathway for CNN's. Later on, big conglomerate such as Google, IBM, Microsoft, etc, developed their own CNN-based Deep learning models named GoogleNet (Szegedy et al. 2015), ResNet (He et al. 2016) which are used for feature extraction and classification tasks.

CNN's can also address forgery detection tasks such as Image forgery detection (Bi et al. 2019), Deepfake detection (Afchar et al. 2018), visual forgery localization (Wu et al. 2019) etc. Bayar and Stamm (2016) proposed a new Convolutional layer to detect image features that have been altered by manipulations such as blurring, noising, resizing, etc.

[1] http://yann.lecun.com/exdb/mnist/.

Their method is trained on a large number of authentic and manipulated images. In 2015, Chen et al. (2015) introduced a CNN based forgery detection approach using media filters. The proposed model can detect tampering of median filtering and copy paste type. Later, the authors also published a study (Bayar and Stamm 2017) which shows the effect of different CNN architectures on image forensics. Bunk et al. (2017) use a combination of CNN and Long Short term memory (LSTM) networks to detect tampered regions in an image. Szegedy et al. (2015) proposed a CNN-based architecture named RRU-Net capable of detecting image splicing forgery without any preprocessing and post-processing. The model uses residual propagation to recall input features to avoid vanishing gradient problems in the deeper networks. Their experiments show promising results on datasets such as CASIA (Dong et al. 2013) and COLUMB (Hsu et al. 2006). Later, He et al. (2016) also proposed an end-to-end network that performs both detection and localization without any extra pre and post-processing. Their network is capable to detect forgeries like removal, enhancement, and unknown types. Other Image Forensics techniques are based on particular issues such as JPEG artifacts removal e.g. Jiang et al. (2021) proposed a flexible blind CNN (FBCNN) capable of predicting adjustable quality factors for JPEG image artifacts removal. Visual Chirality is a notion that objects are different from their mirror images e.g. Lin et al. (2020) uses this notion to show how statistics of Visual data are changed by the reflection. Their approach has application in the field of data augmentation, self-supervised learning, and image forensics. Huh et al. (2018) proposed a self-consistency learning algorithm capable of detecting and localizing image manipulations. The model was trained on a large dataset of real photographs. Their experiments show SOTA results on several image forensics benchmarks without seeing any manipulated images in training.

In recent papers, CNN-based models (Katiyar and Bhavsar 2022; Kumar and Meenpal 2022; Koul et al. 2022) have achieved a *state-of-the-art* performance for Image forgery detection (Tyagi and Yadav 2022). However, the major limitation of these CNN models is their generalization and performance issues on New unseen images. The performance of CNN models is based on three factors which are (a) feature extraction, (b) classifier used, (c) size of the dataset, and (d) data augmentation (optional).

2.1 Convolutional neural networks (ConvNets or CNNs)

A ConvNet is a class of deep neural networks that typically has one or more convolutional layers. A convolution is an linear operation of sliding a filter over the input. The operation involves the multiplication of a set of weights with the input data. The first Convolutional layer applies convolution operation over the entire input data, and passes the result onto the next convolutional layer. ConvNets are mainly used for Computer vision, image processing, object recognition (Uijlings et al. 2013), classification (Rani and Jain 2022), segmentation (Girshick et al. 2014), and various other tasks.

2D Convolution Layer (Conv2D) It is the most common type of convolution in which a filter of size (aXb) "slides" over 2D input data (typically an image) performing element-wise multiplication. The multiplication is performed between the array of input data and a 2D array of weights called Filter (or Kernel). The Kernel applies the same operation over the entire input data array and transforms the 2D input array into a feature map that summarizes the presence of detected features of the input. A filter and the input must always have the same number of channels, also known as Depth. If the image has 3 channels (e.g. a depth of 3), then a filter applied to that image must also have 3 channels (e.g. depth of 3).

The proposed method uses Conv2D for Feature extraction as the CNN learns important features without human supervision, the proposed architecture can detect Image forgeries irrespective of Image size and tampering type.

3 Proposed method

3.1 Architecture of the model

The proposed CNN architecture started with complex models and a large number of layers. The aim is to design a minimal CNN model capable of detecting image manipulation efficiently as compared to the larger architectures (such as DenseNet, EfficientNet, etc.). The source code of MiniNet is available online[2]

The Proposed model contains four convolutional layers from C1 to C4. Max pool layer of size (2×2) is used after each Convolutional layer (CL) to reduce the dimensions of feature maps and select the maximum element present in a region of the feature map generated by Convolutional Layers. A constant stride of 1 for all convolutional filters. The input image of size ($h \times w \times c$) where h and w represent the height and width of the image and c represents the number of channels values. CNNs uses filters (also known as Kernels) to detect features, such as edges throughout the input image. The Kernel of the size ($k \times k \times c$) is used for convolution, where k represents the filter size.

A. the first Convolutional layer (C1) takes an input of size $128 \times 128 \times 3$ and has a kernel size of 3×3. C1 produces a tensor with 32 feature maps as output.

[2] https://github.com/shobhittya/MiniNet.

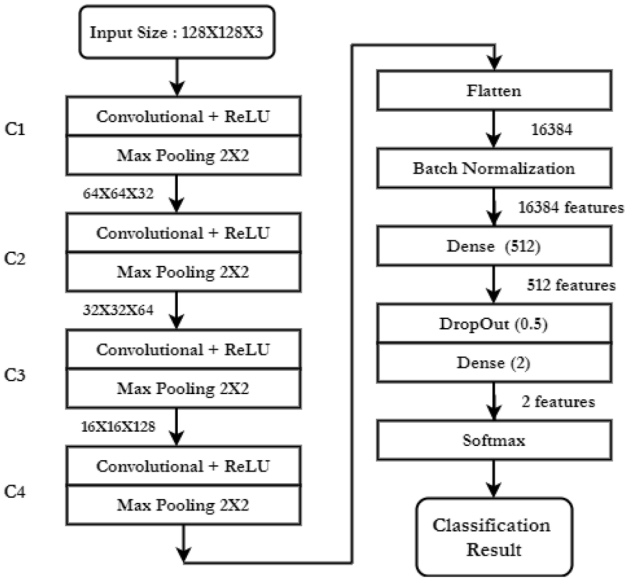

Fig. 2 The network architecture of MiniNet. Layers and parameters are displayed in the boxes. The Outputs are shown next to arrows

B. the C2 contains 64 filters of size 32×32 and produces a tensor with 64 feature maps as output.
C. the C3 contains 128 filters of size 16×16. It produces 128 feature maps as output.
D. the last Convolutional layer C4 contains 256 filters of size 8×8 and produces a tensor of 256 feature maps as output.

The most popular Rectified linear unit (ReLU) (Nair and Hinton 2010; Agarap 2018) activation function is used alongside the Convolutional layers to improve performance and generalization by adding non-linearities to the network. By rectifying the values of inputs less than zero, ReLU is able to eliminates the vanishing gradient problem which can be observed in other types of activation functions. The main advantage of ReLU is faster execution which reduces the computation time . The mathematical representation of ReLU is shown in the Eq. 1, where (x) is the weight of the input element.

$$f(x) = \max(0, x) = \begin{cases} x_{a,} & \text{if } x_a \geq 0 \\ 0, & \text{if } x_a < 0 \end{cases} \quad (1)$$

The successive Convolutional and pooling layers are followed by flatten (to convert the data into 1-D tensor). Later, a batch normalization layer is added to standardize the distribution of the inputs so that the network can represent the identity transform (Ioffe and Szegedy 2015). Finally, the Dense or fully-connected layers are added with Dropout (Srivastava et al. 2014) to improve the robustness of the model. The Dropout layer randomly drops the input to zero during the training time to avoid overfitting. The Network architecture is shown in the Fig. 2.

3.2 MiniNet

The proposed method automatically detects forgeries in images and mainly focuses on two common image forgeries types: Splicing and Copy-move. Complexity and Quantity are the two main factors traditional Image forensics methods [55, 56] are unable to detect tampered images. The Complexity problem occurs due to poor image quality and different formats used to store different types of images. Another problem is Quantity, which refers to the number of forged images found on the images and can be easily produced using online tools and software. Other factors include Image Compression that degrades the data, multiple forgeries used in a single image, low-level formatting, etc. The paper uses deep learning techniques to overcome these problems and presents a minimal CNN named MiniNet to detect fake images. MiniNet has three main advantages: the network uses a low number of layers (a total of 8 layers) to focus on image-level identification, and the small size increases the training speed and avoids overfitting. Finally, the lightweight model can generalize well on different forgeries types. The test results on the existing datasets show that the proposed method can detect tampered images with high speed and accuracy and can also be used on slower devices.

Once the training is completed, the trained model can be used to classify new image input. The authenticity of the new image can be classified using CNN. The Softmax in the output layer is an activation function that can predict the class of an image (whether authentic or fake) with the help of a multinomial probability distribution.

4 Experiments

This section shows the results of the proposed method to detect the image forgeries. The model is evaluated on two datasets named 140k Real and Fake Faces (RFF) and CASIA. Both the datasets are publicly available and easy to download. This work also experimented the performance of other well-known state-of-the-art (SOTA) models using different architectures (i.e. DenseNet, Vision Transformers) as backbone.

4.1 Datasets

The proposed model is trained on two Open-source Image forgery detection datasets. For training, the datasets are divided into 60%, 20%, and 20% for Training, Validation, and Testing. The model is trained on a training set so that the model can generalize better on unseen images. The Dataset

Fig. 3 Sample Images from 140k RFF dataset (Tyagi and Yadav 2021). The original image **A** is on the left and its corresponding spliced image **B** is on the right

description along with other details is depicted in Table 1. For Image source attribution datasets such as DRESDEN (Gloe and Böhme 2010) is publicly available. The Dataset contains more than 13k images of 18 different camera models. For Video Forgery in faces, datasets such as FaceForensics (Rössler et al. 2018) or DeeperForensics (Jiang et al. 2020) are available (Fig. 3). To access these datasets, one has to simply request the authors by filling a google form.

(1) **140K Real and Fake Faces (RFF)** The 140k RFF[3] is an open-source image forgery detection dataset available of Kaggle. The dataset contains a total of 140k images from which 70k are Real face images (from Flickr[4]) collected by Nvidia and 70 k is Fake faces (generated by Style-GAN (Tyagi and Yadav 2021)) sampled from 1 million Fake face[5] The images are resized into 256px, and split into validation (20 k) and test set (20 k). The size of the dataset is around 4 GB. Images are in JPG format and some examples are shown in Fig. 5.

(2) **CASIA** (Dong et al. 2013) This datasets are developed to provide researchers with realistic open-source image tampering datasets[6] The v1.0 dataset contains a total of 1721 images with cut-paste tampered images while the v2.0 contains a total of 12,614 images with both cut-paste and copy-move tampered images. The images are available in JPEG format. Images from CASIA dataset are shown in Fig. 4.

Fig. 4 Images from CASIA dataset (Dong et al. 2013). The original image **A** is on the left and its corresponding spliced image **B** is on the right

Fig. 5 Sample Images from COVER dataset (Wen et al. 2016). The original image **A** is on the left and its corresponding spliced image **B** is on the right

(3) **COVERAGE** (Wen et al. 2016) It contains copy-move forged (CMFD) images with their originals with similar but genuine objects (SGOs). It contains a total of 100 forged images along with their Ground truth masks. In COVERAGE[7], the forged-original image pairs are annotated with (a) the duplicated and forged region masks, and (b) the tampering factor/similarity metric. Images from COVER dataset are shown in Fig. 5.

4.2 Loss function and training settings

The Model is implemented using Python 3.5 with The help of multiple libraries such as Tensorflow (Abadi et al. 2016) 2.6.0, Keras (Chollet et al. 2015) 2.1.5, Pandas (McKinney 2010) 1.3.4, and, Numpy (Harris et al. 2020) 1.14.2, and etc. The Network is trained with 16 GB of RAM and 160 GB of space on the Google Colab[8] Pro platform. The Accuracy

[3] https://www.kaggle.com/xhlulu/140k-real-and-fake-faces.

[4] https://www.kaggle.com/c/deepfake-detection-challenge/discussion/122786.

[5] https://www.kaggle.com/c/deepfake-detection-challenge/discussion/121173.

[6] https://github.com/namtpham/casia1groundtruth.

[7] https://github.com/wenbihan/coverage.

[8] https://colab.research.google.com.

Table 1 Datasets description used in results

Dataset	#Authentic/ #Forged Images	Image Size	Tampering Type
140k RFF	70 k/70 k	256x256	Generated using StyleGAN
CASIA [5]	7491/5123	240x160	Copy-Move
		900x600	Copy-Paste
COVERAGE [33]	100/100	Multiple	Copy-Move

is measured on a single GPU Tesla P100 with a batch size of 64 images of size $128 \times 128 \times 3$ using ADAM (Kingma and Ba 2014) optimizer. ADAM is a commonly used first-order gradient-based Stochastic optimizer that is computationally efficient and requires little memory. As two classes are involved: 'real' or 'fake', the proposed work uses binary cross-entropy as the loss function. The Binary Cross-entropy (or Log loss) is the negative average of the log of corrected predicted probabilities. It is shown in the Eq. 2, where (y_i) is the i_{th} label and p_i is the i_{th} predicted probability of the class. Binary classification is a problem where inputs are divided into two labels based on their features.

$$\text{Log loss} = \frac{1}{N}\sum_{i=1}^{N} -\Big(y_i{}^{\star} \log\left(p_i\right) + \left(1-y_i\right)^{\star} \log\left(1-p_i\right)\Big) \tag{2}$$

The initial learning rate of 10^{-3} is used with the decay of 0.001 after every 50 epochs down to 5×10^{-6}. The Network is trained for about 10 h for each dataset i.e., about 150 epochs for 140 k RFF and about 200 epochs for CASIA.

4.3 Image classification results

Table 2 represents the comparison of popular Image forgery Detection methods on studied datasets. The proposed architecture is evaluated on the studied datasets. The proposed work evaluate the model at image level with benchmark metrics: Accuracy and F_1-Scores. The performance is better for the higher value. The proposed model reached an accuracy of more than 95% for the 140K Real and Fake Faces and 93% for CASIA datasets. In this work, Data augmentation techniques are not used on the input data. Depending on the dataset, one can use such techniques to further improve the network's performance.

Baseline Methods The proposed model is compared with popular methods such as CFA1 (Ferrara et al. 2012), ELA (Neal 2007), SPAN (Xuefeng and Zhang 2020) etc. based on F_1-Scores. The work also compares the accuracies with other Deep Neural Networks (DNN) models achieved on the respective datasets. Some of the methods are described below :

- **CFA1:** The paper presents a Color Filter Array (CFA) to discriminate between original and tampered images. The drawback of this approach is the assumption that the image is acquired using a CFA and due to demosaicing algorithm, the tampering removes the artifacts.
- **NOI1:** The paper proposes a novel segmentation approach capable of partitioning an image with homogeneous noise levels. The authors used noise inconsistencies for blind image forensics.
- **SPAN:** The paper uses a spatial pyramid attention network for detection and localization of forged images from DRESDEN Image database (Gloe and Böhme 2010). The approach uses a Deeper VGG network as a backbone along with Bayar (Bayar and Stamm 2016) and SRM (Fridrich and Kodovsky 2012) Convolutional layers to extract rich features.

Table 2 F1-Score comparison of popular Image forgery Detection methods. "-" represents that the results are not available in the literature. Bold represents the best F1-Score

Method	**F1 Score**	
	CASIA	**COVER**
CFA1 [45]	0.207	0.190
ELA [46]	0.214	0.222
NOI1 [47]	0.263	0.269
SPAN [48]	0.382	0.558
RGB-N [49]	0.408	0.437
MSCRF [54]	0.617	0.729
HFSRNet [59]	0.467	0.624
MiniNet	0.910	0.867
Precision	0.903	0.852
Recall	0.936	0.882

Table 3 Comparison analysis with DNN-based Image forgery methods on CASIA dataset

Authors	Model	#Layers	Accuracy
Zhang et al. [43]	Stacked Autoencoders	–	91.09%
Bi et al. [20]	Ring Residual U-Net	23	76%
Wu et al. [21]	BusterNet	16	77.49%
Chen et al. [41]	CNN using median filters	9	85.14%
Wu et al. [67]	ManTraNet (Self- supervised learning)	–	56.14%
Ali et al. [62]	CNN with double image compression	–	92.23%
MiniNet	Based on Vanilla CNN	8	93.75%

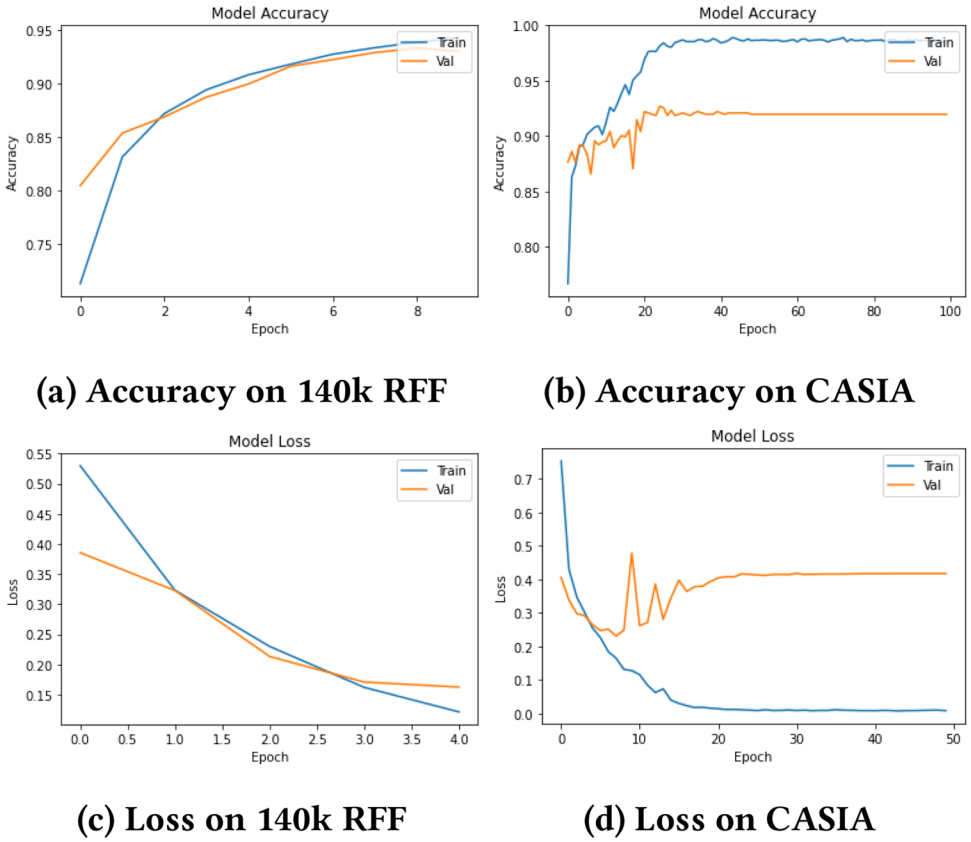

Fig. 6 Accuracy and loss analysis of MiniNet

- **RGB-N** The paper proposes a faster end-to-end trainable R-CNN network capable of detecting tampered regions in a forged image. They used two streams: RGB stream to extract features and, Noise stream that leverages the noise filters and to discover the noise inconsistency between original and tampered regions.
- **Multi-Semantic CRF** The paper proposed a CRF-based attention model to characterize the local correlations between the neighboring pixels. The model uses a fully connected CNN as feature extractor along with a attention residual learning to include local pattern correlation.
- **HFSRNet** The authors proposes a Hybrid features and Semantic reinforcement network (HFSRNet) for image forgery detection. The model uses Long-Short term memory (LSTM) network to capture traces of manipulating artifacts from image patches. HFSRNet mainly uses encoding and decoding network and has four main aspects: hybrid g=features, encoding network, semantic reinforcement, and decoding network.

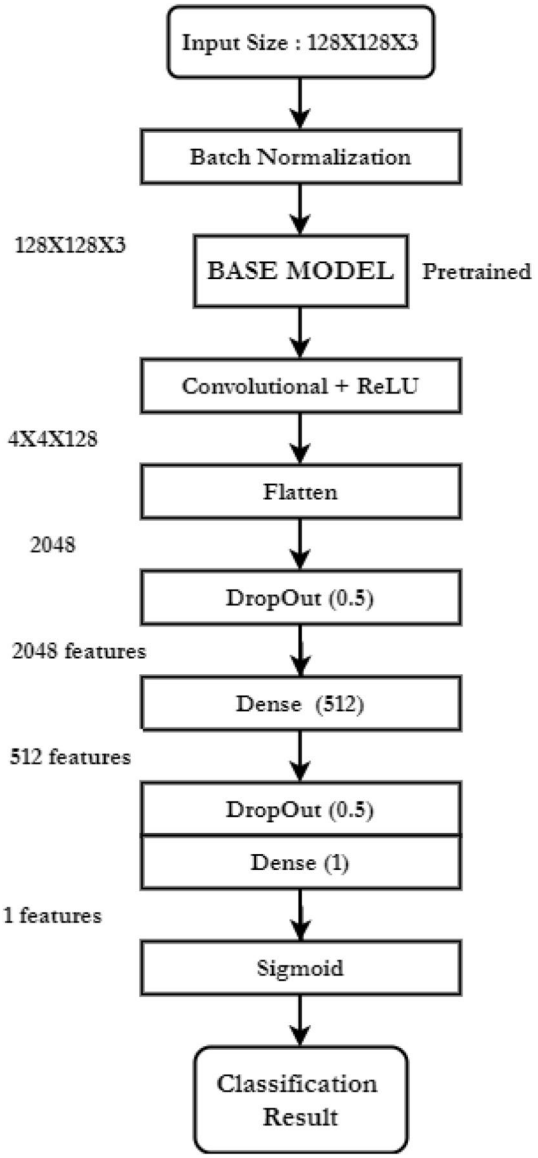

Fig. 7 The network architecture of MiniNet with benchmark models. Layers and parameters are displayed in the boxes. The Outputs are shown next to arrows

4.4 Comparison analysis with existing deep neural network based image forgery methods

The performance analysis of MiniNet and other existing CNN-based forgery detection methods on the CASIA dataset is shown in Table 3. The accuracy of the model is deteriorated with the resolution i.e., with low resolution is reduced, the accuracy also decreases. As the network is concise and small (only 8.5 million parameters). With the training of a few hours on Standard GPU, good performance accuracy can be obtained. The MiniNet model loss is shown in Fig. 6. It is worth noticing that the performance is achieved without any data augmentation techniques such as MixUp, CutMix, RandAugment, etc.

4.5 Ablation study

To better understand the network's behavior, ablation studies are performed. The proposed work also checks which input features are more relevant to the network's output. The ablation studies by Girshick et al. (2014) conclude that the representational power of the CNNs comes from their convolutional layers rather than Dense/fully connected layers. For a better understanding of the model, earlier convolutional layers are frozen and replaced with state-of-the-art (SOTA) models.

Several SOTA models such as DenseNet-121 (Huang et al. 2017), EfficientNet-b0 (Mingxing and Le 2019), Efficient-V2 (Mingxing and Le 2021), etc are used for ablations. The Convolutional layers of MiniNet are replaced with benchmark models pre-trained on the 'ImageNet' (Deng et al. 2009) dataset. The rest of the Network classifier remains the same. The improved Network architecture of MiniNet is shown in Fig. 7.

Table 4 Accuracy performance in Ablations

Base network	Image size	# Params (in million)	FLOPs (in billion)	Peak Memory (in MB)	Accuracy	
					140 k RFF	CASIA
DenseNet-121 [11]	224^2	8.8	3	32	98.22	94.57
EfficientNet-b0 [12]	224^2	5.1	0.4	21	97.66	92.91
EfficientNet-v2 [13]	224^2	24	8.8	92	98.96	92.0
InceptionResnNet-v2 [14]	128^2	56	17	214	97.53	93.11
ViT-b32 [15]	128^2	87	55.4	346	95.71	95.39
MiniNet	128^2	8.5	0.75	35	95.18	93.75

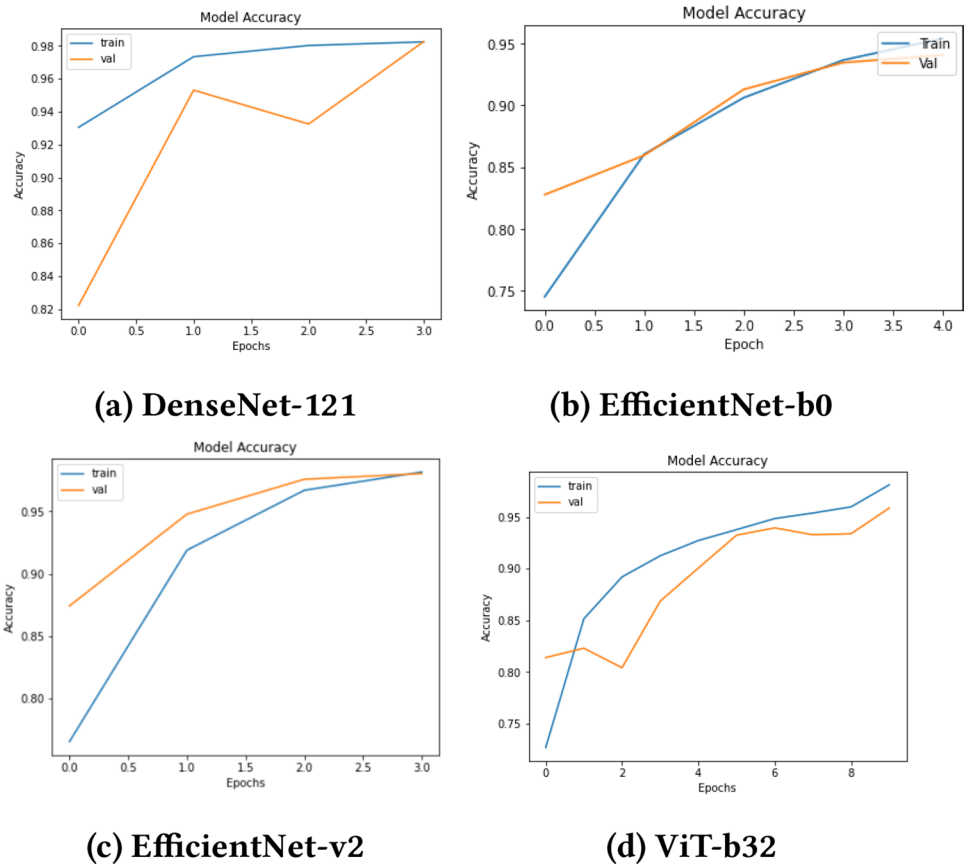

Fig. 8 Accuracy performance of the SOTA models on 140k RFF Dataset

(a) DenseNet-121 **(b) EfficientNet-b0**

(c) EfficientNet-v2 **(d) ViT-b32**

4.6 Performance analysis with SOTA models

The proposed model is trained several benchmark networks on the studied datasets. The Accuracy Performance of different State-of-the-Art (SOTA) Models on 140k RFF and CASIA Dataset is shown in Table 4. The accuracy performance of the SOTA models on the 140 k RFF dataset is shown in Fig. 8 and on CASIA dataset is in Fig. 9.

5 Conclusion

Recently society faces several threats from image and video tampering. In this paper, a CNN named MiniNet is introduced that is capable of efficiently detecting fake images without any pre-processing. The results show a high accuracy on popular datasets with low computational cost. This paper also studies the popular architectures to understand how convolutional layers influence the network's performance.

The proposed model achieved promising results on well-known datasets. The Convolutional layers are the building blocks of a CNN model. These layers generate feature maps that help in tasks such as image classification, segmentation, etc. The proposed method MiniNet uses CNNs and believes that if provided with enough data, it has the potential to improve and can also be used in visual forgery detection. In future, the proposed model can be extended to be used for video forgery, multi-class classification of different tampering types, image and video segmentation.

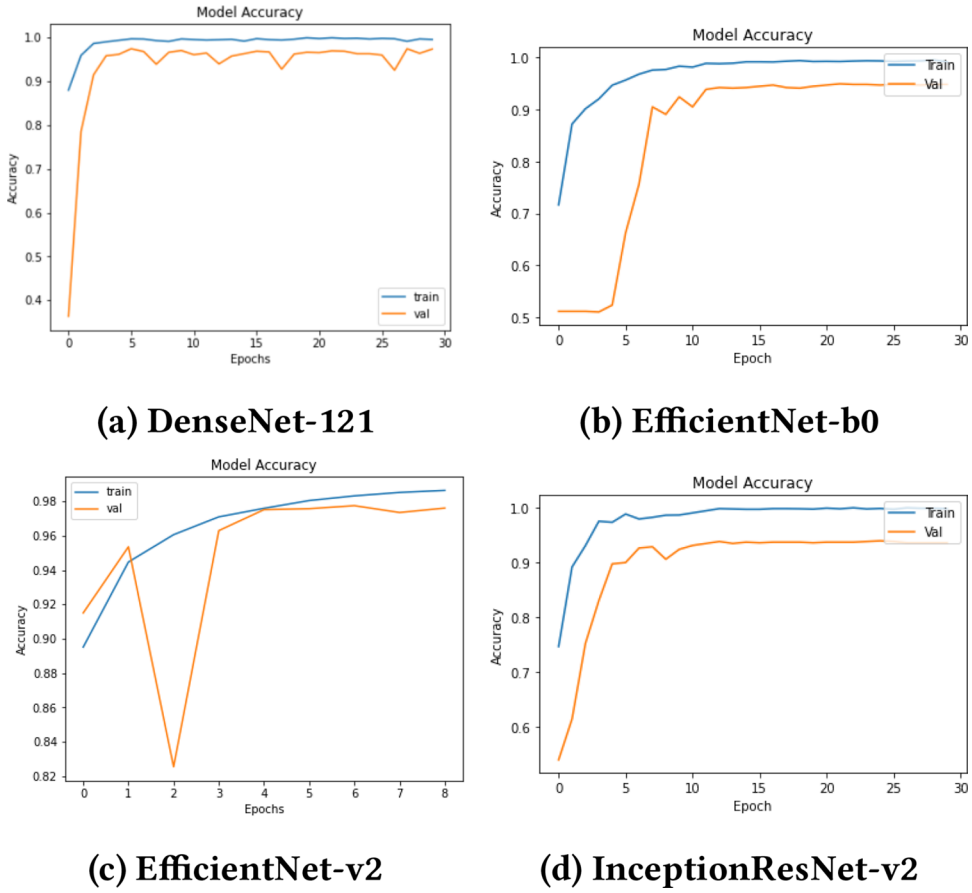

Fig. 9 Accuracy and loss analysis in ablations

Declarations

Conflict of interest The authors have no competing interests to declare.

References

Abadi M et al (2016) Tensorflow: a system for large-scale machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16)

Afchar D et al (2018) Mesonet: a compact facial video forgery detection network. In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE

Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv:1803.08375

Ahmad M, Khursheed F (2021) Digital Image Forgery Detection Approaches: A Review. Applications of Artificial Intelligence in Engineering, Springer, Singapore

Alexey D, Lucas B, Alexander K, Dirk W, Xiaohua Z, Thomas U, Mostafa D, Matthias M, Georg H, Sylvain G et al (2021) An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations

Ali SS et al (2022) Image forgery detection using deep learning by recompressing images. Electronics 11.3:403

Bayar B, Matthew CS (2016) A deep learning approach to universal image manipulation detection using a new convolutional layer. In: Proceedings of the 4th ACM workshop on information hiding and multimedia security

Bayar B, Stamm MC (2017) Design principles of convolutional neural networks for multimedia forensics. Electron Image 2017(7):77–86

Bi X et al (2019) RRU-Net: the ringed residual U-Net for image splicing forgery detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops

Bunk J et al (2017) Detection and localization of image forgeries using resampling features and deep learning. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, 2017

Chen H et al (2021) Hybrid features and semantic reinforcement network for image forgery detection. Multimed Syst pp 1–12

Chen J, Kang X, Liu Y, Wang ZJ (2015) Median filtering forensics based on convolutional neural networks. IEEE Signal Process Lett 22(11):1849–1853

Chollet F, et al. Keras. https://keras.io, 2015

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Deng J et al (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE

Dong J, Wang W, Tan T (2013) Casia image tampering detection evaluation database. In: 2013 IEEE China summit and international conference on signal and information processing. IEEE

Ferrara P, Bianchi T, De Rosa A, Piva A (2012) Image forgery localization via fine-grained analysis of cfa artifacts. IEEE Trans Inf Forensics Secur 7(5):1566–1577

Fridrich J, Kodovsky J (2012) Rich models for steganalysis of digital images. IEEE Trans Inf Forensics Secur 7(3):868–882
Ghai A, Pradeep K, Samrat G (2021) A deep-learning-based image forgery detection framework for controlling the spread of misinformation. Information Technology and People
Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580–587)
Gloe T, Rainer B (2010) The 'Dresden Image Database' for benchmarking digital image forensics. In: Proceedings of the 2010 ACM Symposium on Applied Computing
Gloe T, Böhme R (2010) The dresden image database for benchmarking digital image forensics. J Digit Forensic Pract 3(2–4):150–159
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–44
Harris CR, Millman KJ, van der Walt SJ et al (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2
He K et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition
Heckerman D, Wellman MP (1995) Bayesian networks. Commun ACM 38(3):27–31
Hsu, Yu-Feng, and Shih-Fu Chang. "Detecting image splicing using geometry invariants and camera characteristics consistency." 2006 IEEE International Conference on Multimedia and Expo. IEEE, 2006
Huang G et al (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
Huh M et al (2018) Fighting fake news: Image splice detection via learned self-consistency." Proceedings of the European Conference on Computer Vision (ECCV)
Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning, PMLR
Jiang, Jiaxi, Kai Zhang, and Radu Timofte. "Towards flexible blind JPEG artifacts removal." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021
Jiang, Liming, et al. "Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020
Kadam, Kalyani, Swati Ahirrao, and Ketan Kotecha. "AHP validated literature review of forgery type dependent passive image forgery detection with explainable AI." International Journal of Electrical and Computer Engineering (2088-8708) 11.5 (2021)
Katiyar, Ankit, and Arnav Bhavsar. "Image Forgery Detection with Interpretability." arXiv preprint arXiv:2202.00908 (2022)
Kharrazi M, Sencar H, Memon N (2005) Blind source camera identification. In: IEEE International Conference on Image Processing
Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014)
Koul, Saboor, et al. "An efficient approach for copy-move image forgery detection using convolution neural network." Multimedia Tools and Applications (2022): 1-19
Krishnaraj, N., et al. "Design of Automated Deep Learning-Based Fusion Model for Copy-Move Image Forgery Detection." Computational Intelligence and Neuroscience 2022 (2022)
Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
Kumar, Nitish, and Toshanlal Meenpal. "Salient keypoint-based copy-move image forgery detection." Australian Journal of Forensic Sciences (2022): 1-24
LeCun Y et al (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1.4:541–551
Lin, Zhiqiu, et al. "Visual chirality." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020
Mahdian B, Saic S (2009) Using noise inconsistencies for blind image forensics. Image Vis Comput 27(10):1497–1503
McKinney, Wes. "Data structures for statistical computing in python." Proceedings of the 9th Python in Science Conference. Vol. 445. No. 1. 2010
Mingxing T, Quoc VL (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946
Mingxing T, Quoc VL (2021) Efficientnetv2: smaller models and faster training. In: International conference on machine learning
Nair V, Geoffrey EH (2010) Rectified linear units improve restricted boltzmann machines. ICML
Neal Krawetz and Hacker Factor Solutions (2007) A picture's worth. Hacker Factor. Solutions 6(2):2
Rani A, Jain A (2022) Digital image forgery detection under complex lighting using Phong reflection model. J Electron Imaging 31(5):051402
Rao Y, Ni J, Xie H (2021) Multi-semantic CRF-based attention model for image forgery detection and localization. Signal Process 183:108051
Rössler, Andreas, et al. "Faceforensics: A large-scale video dataset for forgery detection in human faces." arXiv preprint arXiv:1803.09179 (2018)
Russakovsky O et al (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115.3:211–252
Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014)
Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
Stamm Matthew C, Min W, Ray LKJ (2013) Information forensics: an overview of the first decade. IEEE Access 1:167–200
Szegedy C et al (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence
Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015
Tyagi, Shobhit, and Divakar Yadav. "A Comprehensive Review on Image Synthesis with Adversarial Networks: Theory, Literature, and Applications." Archives of Computational Methods in Engineering (2021): 1-21
Tyagi, Shobhit, and Divakar Yadav. "A detailed analysis of image and video forgery detection techniques." The Visual Computer (2022): 1-21
Uijlings JR, Van De Sande KE, Gevers T, Smeulders AW (2013) Selective search for object recognition. Int J Comput Vision 104(2):154–171
Wen B, Zhu Y, Subramanian R, Ng T, Shen X, Winkler S (2016) COVERAGE - A Novel Database for Copy-move Forgery Detection, in Proc. IEEE Int. Conf, Image Processing (ICIP)
Wu, Yue, Wael AbdAlmageed, and Premkumar Natarajan. "Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019

Xuefeng Hu and Zhihan Zhang. Span: Spatial pyramid attention network for image manipulation localization. In ECCV, 2020

Zhang, Ying, et al. “Image Region Forgery Detection: A Deep Learning Approach.” SG-CRC 2016 (2016): 1-11

Zhou Peng , Xintong Han, Vlad I Morariu, and Larry S Davis. Learning rich features for image manipulation detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1053-1061, 2018