IDENTIFYING DIGITAL FORGERY IN IMAGES

Soujanya Manasa (Team lead) - 1RVU22CSE159
Gowri Galgali - 1RVU22CSE060
Lohith R Gowda - 1RVU22CSE093
Sudhanva M S- 1RVU22CSE166

Abstract

The widespread use of image forgery techniques has made it hard to differentiate between authentic and forged images today. In this report, we are showcasing the efficiency of the Convolutional Neural Network (CNN) in the process of identification of image forgery, by utilising the technique of transfer learning, we are comparing the performances of two pre-trained and widely known models, ResNet50 and DenseNet121, to classify the images into two categories, authentic or forged. One of the prominent changes is the fine-tuning of the last neural network layer such that it behaves like a classifier. We have created a custom dataset from the pre-existing datasets, which are **CASIA.V2, Columbia uncompressed image splicing detection, Image Manipulation dataset, and COMOFO dataset**. These datasets are labeled and include forgery images of methods of copy-move and splicing. These datasets have been trained on a Python platform where frameworks like TensorFlow and Keras are utilised. These models are trained using the GPU provided by KAGGLE (NVidia K80 GPUs). We have used a classification report as the evaluation metric for our results, in addition to this detailed classification report we have also provided the analysis of AUC-ROC. Our experiments

have shown an f1 score of **83%** and **81%** with the ResNet50 and DenseNet121 respectively (The dataset being used is slightly imbalanced). Further, we have also incorporated the localisation of the image forgery region, which includes a ResNet-based feature extractor and decoder network, which generates masks, which would help in localisation. For this task, we used the dataset **CASIA.V1 (with ground truths)**. This project aims to use CNN to solve the problem of image forgery.

Introduction

Have you ever browsed through social media on your devices and came across a picture or video that seemed perfect to be real? In this era, manipulating images has become more accessible leading to a blurry distinction, between what is real and what is forged.

Detection of forgery in images helps us to know if a given image is altered or not to do so we need a wide number of features as there are multiple forgery techniques such as copy-move forgery, splicing, inpainting, and deepfake techniques. Features that are

based on CNN, i.e., convolution network network, help us to classify the images. Convolution neural networks have multiple layers composed of neurons to process the input features and extract them, the layers get more complex with the depth of the model of the network, hence making the extraction of features more accurate.

The main objective of this project is to classify the images and form a comparison between two CNN models for the same task. Here in our approach, we have utilized ResNet50 and DenseNet121 for the comparison. Further after developing these models, we took up another challenge of localising the forged region in the images. We used a ResNet feature extractor and decoder network for image segmentation and masking the forged region.

Related Work

Most previous studies of forgery detection in images use different methods to identify manipulated photos. Earlier these were visual inspection and analysis of the metadata. But now in more recent times with the increase in the number of tools more advanced computational methods for forgery detection have been utilised. One of the most efficient methods involves using CNN, i.e. convolution neural network.

Doegara et al. ["CNN-based Image Forgery Detection using pre-trained AlexNet Model"] Demonstrated a similar approach by using the pre-trained model Alex net for forgery detection and localization they also used an SVM classifier which helped them achieve an accuracy of 93.94% on the MICC-F220 dataset. [1]

Nalluri et al. [Image Forgery Detection using ResNet50] presented a method that combined both CNNs with a ResNet50 architecture and Error Level Analysis(ELA) on the CASIA 2.0 Image dataset Which contains annotated images that indicate whether an image is authentic or tampered with. [2]

Zanardelli et al. [Image forgery detection: a survey of recent deep-learning approaches] did a detailed survey of recent deep-learning approaches to image forgery detection, that focused on copy move and splicing attacks their analysis pointed out that deep-learning techniques were the most relevant approach for image forgery detection. [3]

Ronneberger et al. [U-Net: Convolutional Networks for Biomedical
Image Segmentation] Introduced the U-Net architecture, which was designed for the segmentation of biomedical images. This architecture comprised a contracting path(encoder) to capture context and an expanding path(decoder) for precise localization and showed higher performance on different biomedical image segmentation tasks. [4]

Nimsuk and Paewboontra[Compact CNN Model for Classifying Rose Apple Species and Detecting Their Skin Defects] Came up with a compact CNN model to classify the different species of rose apple and compared it with traditional models like ResNet50 and Densenet121. This model significantly reduced the number of parameters and also the prediction time without compromising on the accuracy. [5]

Khalil et al. [Enhancing Digital Image Forgery Detection Using Transfer Learning]

showcased an approach to enhance the detection of forgery in digital images by the use of transfer learning. This method involved the detection of the differences in the compression quality between forged and original image areas. They compared eight pre-trained models suited for binary classification, with the MobileNetV2 model showing the highest accuracy in detection (around 95%) with fewer parameters and lesser training time. [6]

All the above-mentioned studies demonstrate the effectiveness of using Convolutional Neural Networks and the use of pre-trained models for image forgery detection, localisation, and other similar related tasks. Several techniques, architectures, and datasets have been explored, with the prime objective being to improve accuracy and deployment capabilities.

Our project tries to compare the performance metrics and the results of the same approach by other state-of-the-art methods and well-known models such as ResNet50 and DenseNet121 on a combination of datasets, It showcases how effective DenseNet121 and ResNet50 models perform when it comes to image forgery detection.
We decided to use these two models as they are said to be a few of the best models trained over a similar problem, specifically RestNet50 and DenseNet121 because of parameter efficiency that is they achieve stronger performance compared to models with deeper architecture such as ResNet101 and DenseNet121.
The other reason is that they are comparatively easier to implement using the preexisting Keras and TensorFlow framework.

We have also tried to incorporate the localisation of the forgery area by using a Resnet extractor and decoder.

Methodology

Our approach towards the problem statement was to build the two models according to our requirements and further fine-tune them to get a good result.
The methodology includes loading the datasets, preprocessing them using OpenCV, and using the method**. resize(),** the images have been resized to a dimension of 224x224, and further, the authentic images have been labeled as 0 and forged as 1**.** Models are further developed using transfer learning.

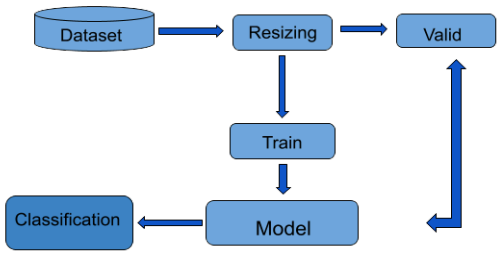


Fig 1: The process of model training and validation

ResNet50

ResNet [2] is A deep neural network architecture that was designed to address the vanishing gradient problem. It contains skip connections which are also known as residual connections. These connections allow the gradients to flow to the initial layers during backpropagation. This helps in better training of the deeper networks.

ResNet50, with 50 layers, is a variant of ResNet has achieved high performance on the well-known ImageNet dataset at the time of its introduction and is one of the best-trained models for Image classification tasks. Using Keras and TensorFlow, we have built a model where only a few layers are being trained and not frozen. We have used average pooling layers to reduce the dimension ReLU activation function and L2 regularisation. A learning rate schedular has also been used to reduce the learning rate exponentially as the epochs increase. After this training process, the model is tested using the validation set to check the score for the same.

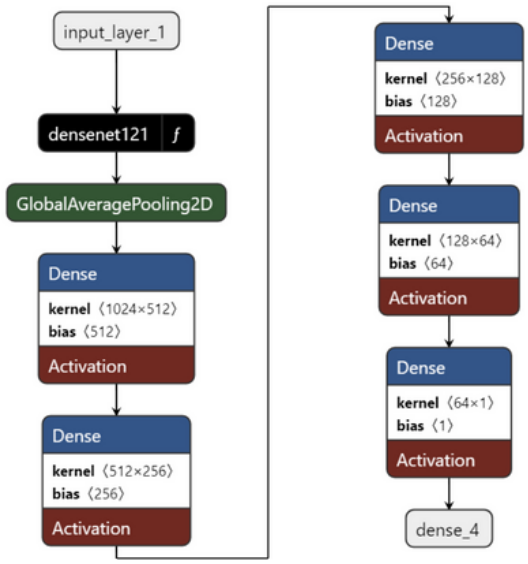


Fig 2: **ResNet50 model**

DenseNet121

DenseNet, [7] introduced in 2017, builds upon the success of residual networks (ResNet) by introducing the concept of dense connections. In DenseNet, each neural network layer is connected to all the other layers in a forward manner, which results in better feature reuse and gradient propagation. These patterns allow the flow of information and the reuse of features that were learned by the previous layers. DenseNet121 variant has 121 layers and has demonstrated better parameter efficiency and performance compared to most other models known until then, these architectures are widely adopted in various computer vision applications due to their ability to learn feature representations well for image classification and other related tasks. We have incorporated a DenseNet121 model with freezing most of the layers from the base model and similar to the ResNet50 model used a global pooling layer to reduce the spatial dimension of the images and further activation function sigmoid is used at the last fully-connected layer to be a classifier.

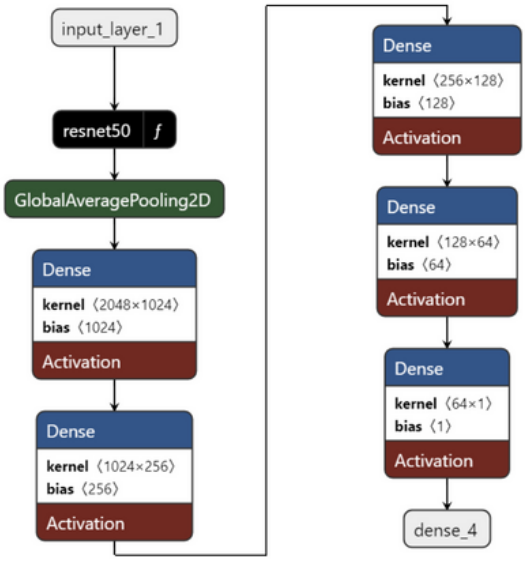


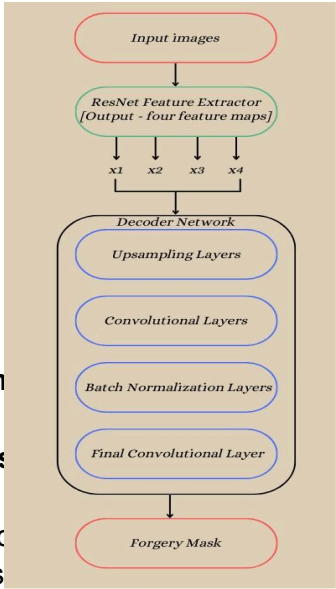
Fig 3: DenseNet121 model

Localisation Model - ResNet Extractor

Localization in image forgery detection involves finding regions within an image that

have been forged. Localization can be achieved by using techniques that analyze inconsistency in color and texture. For instance, these algorithms look for identical patterns in the image in the case of copy-move forgery detection. Methods that often use CNNs are commonly used for localization where these trained networks learn to understand and recognize patterns of forgery and generate a binary mask to highlight areas where there are high chances of forgery within the image. The value of each pixel in the mask showcases the likelihood of forgery.

A ResNet feature extractor is used to extract features from various layers of the Resnet model, further, the forward method obtains feature maps at different resolutions. The decoder network works for upsampling and combining features of different resolutions and further generates a binary mask.



Experim

Datasets

To train ed a large dataset c he forgery methods Splicing. This increased the diversity of the data, hence giving the model a wider range to be trained upon. The dataset used is a combination of CASIA.V2(12615), Columbia uncompressed image splicing detection(363), image manipulation dataset(96), and COMOFO dataset(10003). The resolution of the images belonged to a wide range, hence we went ahead with only resizing the images. The image dataset is labeled, which is classified into two categories, forged(10351) and authentic(12726).

For the development of models, we split this customised dataset into the ratio of 70:30 (training: validation).

Setup

The algorithm is developed in Python, using sklearn and Tensorflow framework to build the model for training and inference of deep neural networks. The library Keras minimizes the coding and reduces the complexity of building the model. Keras also provided us with these pre-trained models, ResNet50 and DenseNet121 to be directly imported and used. We are also importing cv2, for image processing. We trained our model on the GPU provided by KAGGLE, which is NVidia K80 GPUs, this reduced the computational time significantly.

Implementation

ResNet50

The model's training was a full pass of an epoch, the model was updated accordingly to reduce the difference between the expected and actual output. The number of epochs used for training this model was 7.

Increasing the number of epochs started overfitting the model and hence reduced the validationscore.

The top layer of the model has not been considered such that our own customized classification layer can be used. Further first few layers of the model have been frozen, such that those weights do not get updated

and the weights are only finetuned.' GlobalAveragePooling2D()' is used to reduce the spatial dimension of the feature maps, we have included dense layers with ReLU activation function and L2 regularisation.
The final fully-connected layer is a single layer that has a sigmoid function, suitable for binary classification.
Adam optimizer is used and the Cross_Entropy loss function is tackled in this model.

DenseNet121

Similar to ResNet50 this model was also trained for 7 epochs, for the same reason as stated above. Expect for the last 70 layers rest of them have been frozen only for fine-tuning. A sequential model uses the pre-trained base model while using GlobalAveragePooling2D(), ReLU activation function, and L2 regularisation. Similar to the above model, the final fully-connected layer, the single layer has a sigmoid function, suitable for binary classification. The model is compiled using the Adam classifier and Cross_Entropy loss function.
We have also used a method from Keras known as early stopping, which monitors the validation accuracy and stops the process if there isn't a good change.

Localisation Model- ResNet Extractor

We have used the nn.module from which the ResNet extractor inherits the feature extractions. The ResNet's base model is initialised for using the initial input weights, further, the layers of the model layers are frozen for only fine-tuning. The forward method processes the input from the model to obtain the feature maps at different layers.

A decoder network uses nn.ConvTranspose2d to defined upsampling layers and nn.Conv2d for convolutional layers. Now these extracted features are used to generate the forgery masks.
The model is trained to generate the forgery region and detect the same.

Implementation
https://github.com/LohithR22/Image-Forgery-Detection

Evaluations

As a forgery detection application, [3] it is a binary classifier hence we are following the evaluation by creating a classification of the report, the key metrics of which are confusion matrix, accuracy, precision, recall, f1 score, and support.
The confusion matrix here provided us with the details of TP (True Positive), FP(False Positive), FN(False Negative), and TN(True Negative); this evaluation provides us with the performance ground truth of the model.
In our case, Accuracy is the score given to the model based on the training, that is Precision is the accuracy of the model to predict and identify the forged images, and Recall is the ability of the model to correctly identify the actual forged images from the entirety of the forged images, F1 score indicates the model has good Precision Recall and Support finally giving us the information regarding the instances of each class the model has made predictions on.
AUC curve and ROC.

Results
These are the results of our model training, considering the forged class(1):

ResNet50

False Positive =718
False Negative = 422
F1 Score= 83%
AUC=92%

DenseNet121
False Positive =623
False Negative = 552
F1 Score=81%
AUC=92%

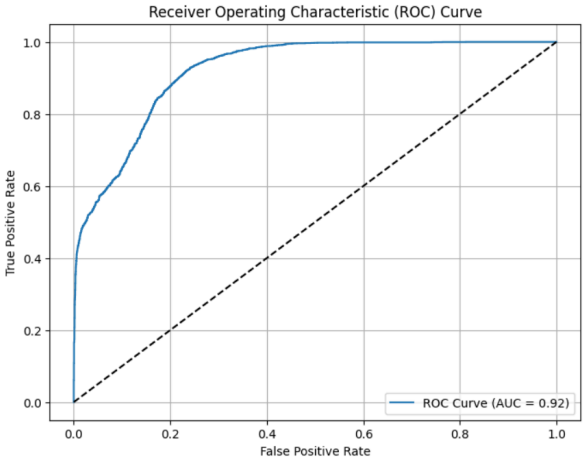


Fig4: ROC for ResNet50 model

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 7, 7, 2048)	23,587,712
global_average_pooling2d_2 (GlobalAveragePooling2D)	(None, 2048)	0
dense_8 (Dense)	(None, 1024)	2,098,176
dense_9 (Dense)	(None, 256)	262,400
dense_10 (Dense)	(None, 128)	32,896
dense_11 (Dense)	(None, 64)	8,256
dense_12 (Dense)	(None, 1)	65

Total params: 77,862,277 (297.02 MB)

Trainable params: 25,936,385 (98.94 MB)

Non-trainable params: 53,120 (207.50 KB)

Optimizer params: 51,872,772 (197.88 MB)

Fig5: Training process and Parameter count of ResNet50

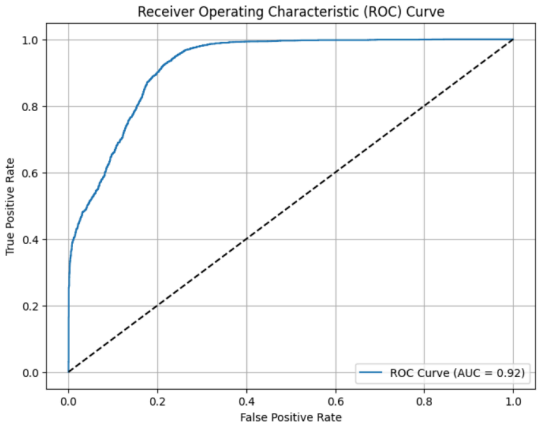


Fig6: ROC for DenseNet121 model

Layer (type)	Output Shape	Param #
densenet121 (Functional)	(None, 7, 7, 1024)	7,037,504
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1024)	0
dense (Dense)	(None, 512)	524,800
dense_1 (Dense)	(None, 256)	131,328
dense_2 (Dense)	(None, 128)	32,896
dense_3 (Dense)	(None, 64)	8,256
dense_4 (Dense)	(None, 1)	65

Total params: 23,037,253 (87.88 MB)

Trainable params: 7,651,201 (29.19 MB)

Non-trainable params: 83,648 (326.75 KB)

Optimizer params: 15,302,404 (58.37 MB)

Fig5: Training process and Parameter count of ResNet121

	ResNet50	DenseNet121
F1 Score	0.83	0.81
AUC	0.92	0.92

Accuracy	0.84	0.83

(For 7 epochs)

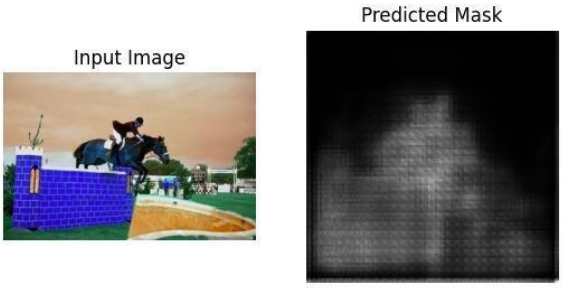


Fig6: The output mask result of the localisation model.

Conclusion

Inference

This project shows the potential of convolution neural networks in the field of image forensics, image forgery detection in particular. This paper compares ResNet50 and DenseNet121 architectures in identifying forged images. Transfer learning and fine-tuning the models ensured that both the models achieved good results with F1 scores of 83% and 81% respectively, on a dataset that comprised images forged using various forgery methods. Based on our above comparison, we can infer that on valid data ResNet50 has a slight advantage with a better f1 score.
Comparison based on parameters, f1 score ratio, we see that the DenseNet121 model shows a better ratio result.
Localisation model still needs to be trained further to reach our desired goal.

Limitations

This project does have limitations, which span multiple areas.

- The dataset used for this project includes only splicing and copy move and hence would not give accurate results on other forgeries.
- Localisation model, which uses RestNet as an extractor and a decoder block, didn't put out the best mask result, the code could be further enhanced for better performance.
- The models have not been tested with real data.

Future Scope

- Further fine-tuning of the models can be done, by exploring different loss functions, freezing and unfreezing the layers can give a much better result.
- In the case of localisation, the extractor and decoder can further be more diligently trained over a larger dataset of ground truths for better results of the masks.

References

[1] A. a. D. M. a. G. K. Doegar, "CNN Based Image Forgery Detection Using Pre-trained AlexNet Model," *International Journal of Computational Intelligence & IoT,* vol. 02, no. 01, 2019.

[2] T. K. T. R. T. P. S. S. S. S. Nalluri Brahma Naidu, "Image Forgery Detection using ResNet50," *Ijraset Journal For Research in Applied Science and Engineering Technology,* vol. 12, no. 3, 2024.

[3] G. F. Zanardelli Marcello, "Image forgery detection: a survey of recent deep-learning approaches," *Springer,* vol. 82, no. 12, 2023.

[4] O. F. P. B. T. Ronneberger, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Springer,* vol. 9351, 2015.

[5] W. P. N. Nimsuk, "Compact CNN Model for Classifying Rose Apple Species and Detecting Their Skin Defects," *IEEE Xplore,* pp. 136-139, 2021.

[6] A. a. G. A. a. E. H. a. S. G. a. G. H. Khalil, "Enhancing Digital Image Forgery Detection Using Transfer Learning," *IEEE Access,* p. 1, 2023.

[7] Z. L. L. v. d. M. K. Q. W. Gao Huang, "Densely Connected Convolutional Networks," 2018.

[8] X. Z. S. R. J. S. Kaiming He, "Deep Residual Learning for Image Recognition," *arXiv.*