A Systematic Approach to Detect Spliced and Forged Images using Deep Learning Technique

N. Brahma Naidu[1], Assistant Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

E. Harish[2], **G. Vamsi Kumar**[3], **B. Sai Ravi Teja**[4], **B.Naveen**[5]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1,2,3,4,5]nbnaidu1208@gmail.com,
enugantiharish15@gmail.com,vamsigamidi1725@gmail.com,
ravitejabhimavarapu99@gmail.com,naveenbollimuntha7801@gmail.com

DOI;10.48047/IJFANS/V11/I12/191

Abstract

The widespread accessibility of image editing tools has made it simpler to alter the contents of digital figures as multimedia technology has advanced. Moreover, photographs are more susceptible to counterfeiting when they are distributed through an open channel using information and communication technology (ICT). Due to the flaws in the telecommunications infrastructure, it is possible for hackers to make subtle but deceptive alterations to picture databases. If edited with a malicious intent, the phoney photographs might create serious social and legal problems. The use of sophisticated tools built through deep learning techniques can accurately detect changes to the mathematical image is required for the discovery of image forgeries. The sensitivity in photos is typically concealed through splicing forgeries. Extreme contrast is introduced by splicing in the edges, smooth areas, and corners. In-depth Learning

Keywords: Deep Learning, splicing, image editing, forgeries, concealed.

Introduction

Forgery and spliced images are two common types of digital image manipulation that are often used to deceive viewers or to create false evidence. Forgery refers to the act of creating a fake image from scratch or altering an existing image to misrepresent or falsify information. For example, a person might forge a document or photo to change its contents or create a false impression. Forgery can be done using a variety of tools, such as photo editing software, and can range from simple alterations to highly sophisticated manipulations that are difficult to detect. Spliced images, on the other hand, are created by combining two or more images to create a composite that appears to be a single, authentic photograph. For example, a spliced image might combine a person's face from one photo with a different background from another photo. Spliced images can be used to create misleading or deceptive images that appear to be real.Both forgery and spliced images can be difficult to detect, especially if they are done using sophisticated techniques.

The below image Fig.1 is an example of a tampered image followed byFig.2which is an example of an authentic image.

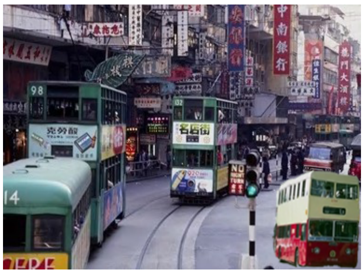

Fig1: Tampered Image

Fig2: Authentic Image

However, there are several methods for detecting these types of image manipulation, including digital forensics, image analysis, and machine learning algorithms. These techniques can help to identify inconsistencies, artifacts, or anomalies in images that may indicate manipulation, and can be used to determine the authenticity and integrity of digital images.

Image forgery detection refers to the process of identifying whether a digital image has been manipulated or tampered with in some way. Image forgery can take many forms, such as adding, deleting, or modifying objects or details in the image. It can be done for a variety of reasons, such as to create misleading or deceptive images, to hide information, or to create false evidence.Image forgery detection is important in a variety of fields, including law enforcement, journalism, and digital forensics. In these fields, it is often necessary to determine the authenticity and integrity of digital images, especially in cases where the images are used as evidence.

There are several techniques for image forgery detection, including visual inspection, metadata analysis, digital forensics, and image analysis. Visual inspection involves looking at the image to identify inconsistencies or anomalies that may indicate manipulation. Metadata analysis involves examining the metadata associated with the image to determine if it has been altered. Digital forensics involves analyzing the digital data associated with the image to determine its origin and history. Image analysis involves using computational methods to identify patterns and features in the image that may indicate manipulation.

Recent advances in deep learning and computer vision have led to the development of more sophisticated techniques for image forgery detection, such as using Convolutional Neural Networks (CNNs) and Error Level Analysis (ELA) together. These techniques can help to identify even subtle forms of image forgery, and are becoming increasingly important in the fight against digital image manipulation.

Error Level Analysis (ELA) is a technique for digital image analysis that can be used to detect areas of an image that have been manipulated or edited. It works by comparing the error levels in different areas of the image, which can reveal areas where the image has been altered.ELA involves taking an image and re-saving it at a different quality level, typically at a lower compression rate. This creates a new version of the image that has a different level of error or noise. By comparing the original image to the newly generated version, the areas that have been edited or manipulated will appear as regions of higher error levels.

CNNs are made up of many layers of linked neurons that can recognize patterns and features in the images they are trained on.A picture is supplied into the network's first layer as the input to a CNN. To find patterns in the image, such as edges or corners, the first layer uses a number of filters, called convolutions.The output of the first layer is then passed to the next layer, where additional convolutions are applied to identify more complex patterns.

CNNs have revolutionized the field of computer vision, and are widely used in applications such as image recognition, object detection, and facial recognition. They have also been applied to other domains, such as natural language processing and speech recognition. The ability of CNNs to automatically learn and identify features in images makes them a powerful tool for a wide range of tasks in machine learning and artificial intelligence.

Literature Survey

In order to detect spliced images, Zhang et al. [1] devised a model that focuses on moment characteristics taken from the multi-size block discrete cosine transform (MBDCT) and image quality metrics (IQMs). This approach compares the statistical differences between an original image and a modified version. The accuracy of the experiments the authors conducted using the Columbia Dataset [2] was 89.16%.

The discrete cosine transform (DCT) domain is used to extract Markov features, and Pham et al[3].'s fast technique for detecting picture splicing generates a feature vector by merging two types of Markov features, coefficient-wise and block-wise. A support vector machine (SVM) is then used to determine whether a query image is genuine or a forgery using the feature vector. They obtained a 96.90% accuracy using the CASIA v2.0 dataset.

Wu et al. [4] offered an addition to the image splicing detection problem where they have worked with two independent photographs and evaluated the risk of one image being tampered with the other. Also, they developed a deep neural network called the deep matching and validation network for locating and detecting picture splicing (DMVN). They employed CASIA v2.0 and the Nimble 2017 datasets in their tests. 94.15% and 79.08%, respectively, are the precision and recall results of this system.

In order to identify the most significant evidence of a forgery, straight from the training data, Pomari et al. [5] introduced a novel method for detecting splicing in digital images by combining the high representational capacity of illuminate maps and convolutional neural networks. Their research suggests a technique that does away with the laborious feature engineering procedure, enables the identification of forged regions in a picture, and is demonstrated to produce a classification accuracy of more than 96%.

To demonstrate their method for detecting image falsification, Muhammad et al. [6] looked into the steerable pyramid transform (SPT) and local binary pattern (LBP). To create a feature vector, they essentially applied the STT to the Cb and Cr channels of the YCbCr picture colour space. Next, they used LBP histograms to describe the texture in each SPT sub band. In order to detect image forgeries, they use support vector machine (SVM), a classifier that is based on the feature vector. For the CASIA v2.0 dataset, the obtained accuracy is 97.33% [7].

Problem Identification

The increase of fake and spliced images on friendly media has become a main issue in recent age. With the ease of access to powerful concept editing software and the skill to quickly disseminate figures online, it has become more and more easy for individuals to devise and share manipulated concepts. This can have serious consequences, in the way that spreading misinformation and publicity, damaging reputations, and sabotaging the integrity of visual television. As social news continues to play an increasingly main role in shaping common belief and discourse, it is more important than ever to cultivate effective techniques for detecting and barring image guidance.To address these challenges, researchers are developing advanced algorithms and techniques for image analysis, such as deep learning and computer vision methods. These methods can identify subtle patterns and features in images that may indicate manipulation, even in cases where the manipulation is highly sophisticated. Error level analysis and Convolutional Neural Networks are some of the Algorithms used to detect the spliced and fake images. [8-16]

Methodology

Error level analysis is a technique used to detect image manipulation by comparing the error level of different regions in an image. The basic idea is that when an image is saved and compressed multiple times, the error level in the compressed regions will be higher than the error level in the original regions. By comparing the error levels in different regions of an image, it is possible to detect areas that have been manipulated or added to the image.

This Error level analysis is applied on every image present in the dataset named CASIA2 and the error levels are examined. These images are labelled as 0 and 1 representing spliced or tampered and authentic respectively.Now the conversion to categorical format takes place where each label is represented as a one-hot encoded vector of length 2, where the index corresponding to the label's value is set to 1 and all other indices are set to 0. For example, a label of 0 would be converted to [1, 0] and a label of 1 would be converted to [0, 1].

This is useful for training a neural network model with multiple output classes, as it provides a way to represent categorical data that is compatible with the model's architecture. The categorical format allows the model to predict a probability distribution over all possible output classes for each input image, and the predicted class can be determined by taking the index of the highest probability value in the output vector.

A deep learning model called a convolutional neural network (CNN) is utilised largely for image and video categorization tasks. Convolutional, pooling, and fully linked layers are among the layers that make up a CNN's architecture. The input image is convolved over by the convolutional layers using a collection of trainable filters, creating feature maps that capture various facets of the image. While the fully connected layers carry out classification based on the features extracted by the convolutional layers, the pooling layers down sample the feature maps to lessen the computational complexity of the model. Now a CNN model is built using several input layers which are fully connected followed by final dense layer with 2 units and softmax activation function, which outputs a probability distribution over the two classes (authentic or manipulated).

Implementation

Using Keras, a high-level deep learning API created by Google for the implementation of neural networks, the suggested method is carried out. It is used to make the implementation of neural networks simple and is developed in Python.Kaggle, google colab,jupyter notebook platforms can be used to implement the proposed solution.

A dataset is a collection of data that is organized and stored together, often in a structured format.Datasets are commonly used in machine learning and other data-driven applications to train and test models or algorithms. A dataset typically includes both input data, such as features or attributes, and output data, such as labels or target values, that are used to train or evaluate a model.

In order to be useful for machine learning, a dataset must be large and diverse enough to capture the full range of variation and complexity in the real-world data that the model is intended to process. Additionally, the data must be carefully pre-processed and cleaned to ensure that it is suitable for analysis and modelling.

Dataset used in this solution named as CASIA contains 7941 images which are authentic and 5124 images which are tampered.

Now Error level analysis is performed on all the images present in the dataset. A forensic technique called error level analysis can be used to spot areas of an image that have varied levels of compression. The method could be used to identify whether a photo has undergone digital editing. It's important to learn more about JPEG compression in order to comprehend the strategies better. JPEG (Joint Photographic Experts Group) is a technique for digital image lossy compression. This approach for data encoding compresses data by removing or losing some of it. The degree of compression could be determined as a suitable trade-off between image quality and picture size. The standard JPEG compression ratio is 10:1. The output obtained for every image is recorded with the values 0, 1 where 1 indicates an authentic image and 0 indicates a tampered image.

Following that, the CNN model is built with several fully connected input layers and with relu activation, which introduces the property of non-linearity to a deep learning model and solves the vanishing gradients issue, and followed by a final dense layer with two units and a softmax activation function, which outputs a probability distribution over the two classes. Now the CNN model is trained using the pre-processed data i.e. The Ela performed data. For tasks involving image processing, recognition, and classification, convolutional neural networks (CNNs), a type of deep learning algorithm, are frequently utilised. They are made to recognise and understand the characteristics of images since their structure is modelled after that of the visual cortex in the brain. In order to recognise patterns and features in the images they are trained on, CNNs are made up of numerous layers of interconnected nodes, or neurons. A picture is supplied into the network's first layer as the input to a CNN. In order to detect patterns like edges or corners, the first layer applies a number of filters, or convolutions, to the image. Afterwards, the second layer receives the output from the first layer and applies more convolutions to it.

CNNs have transformed computer vision and are now widely utilised in tasks including object detection, facial recognition, and image recognition. They have also been applied to other domains, such as natural language processing and speech recognition. A valuable tool for a variety of machine learning and artificial intelligence tasks, CNNs have the capacity to automatically learn and recognise elements in images.

A batch length of 32 is utilised for training the batch. For model training and evaluation, the pre-processed dataset is divided into training and validation sets. When the test size option is set to 0.2, it signifies that 80 percent of the data will be used for training and only 20 percent for validation. During training, the model iteratively modifies the values of its parameters to minimise the loss function using the optimization technique defined in the model's compilation stage. After the completion of training, the model is ready to predict the label of the given image i.e., authentic or tampered.

Results & Conclusion

This proposed solution to detect spliced and forged images performed on CASIA dataset resultsthe following outputwhich is depicted in the below Table 1.

S.No	Metric	Value
1.	Accuray	0.9412
2.	Loss	0.2285

Table1: Output values

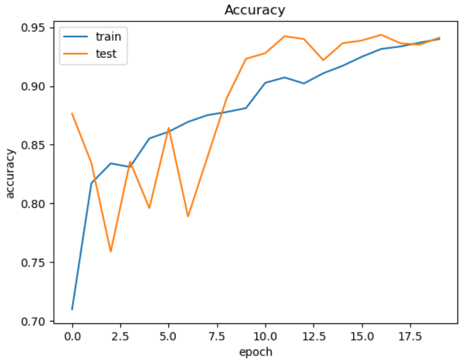

Fig 3: accuracy epoch graph

In Fig.3 the graph, with epoch on the x-axis and accuracy on the y-axis, is commonly used to visualise the performance of a machine learning model over the training period. The x-axis represents the training process's iterations or epochs, and the y-axis represents the model's performance on the training set.

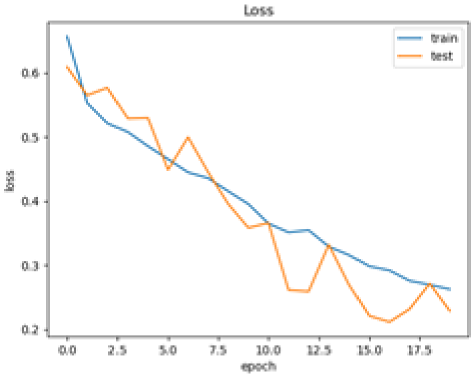

Fig 4: loss epoch graph

In Fig4, the graph, with epoch on the x-axis and loss on the y-axis, is commonly used to visualise the performance of a machine learning model during training. The x-axis represents the number of epochs or iterations of the training process, while the y-axis represents the model's loss on the training data.

A table called a confusion matrix is used to assess how well a categorization model is working. It displays the amount of accurate and inaccurate predictions the model made on a set of test data that were broken down into all conceivable classes or categories.

The four main elements of a confusion matrix are:

True Positive (TP): A positive (or present) class was successfully predicted by the model.

False Positive (FP): The model predicted the wrong class, thinking it was positive (or present), when it was actually negative (or absent).

True Negative (TN): A negative (or nonexistent) class that the model accurately anticipated.

False Negative (FN): The model predicted a positive class when the actual class was positive, which was erroneous (or present).

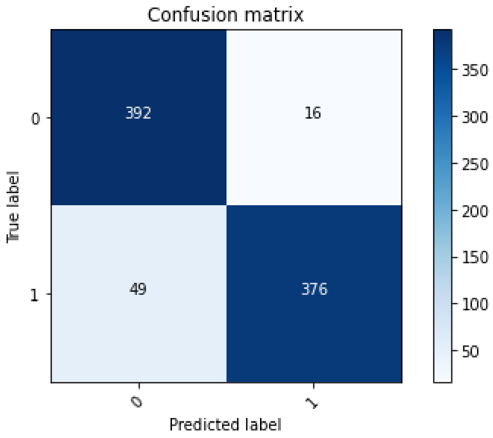

Fig 5: confusion matrix

The confusion matrix from Fig 5, we get the results of TP, FP, TN, FN as 392, 16, 49, 376 respectively.

Future Scope

The above model performs well in detecting the genuine and tampered images.The method Error level analysis used in this paper is not foolproof and can be tricked by certain types of manipulations i.e. if an image is saved and compressed multiple times at different levels of compression, the error level analysis may not be able to accurately detect manipulation.This authorizes thebasic inspiration of our future research in this direction.Also, it hopeful appealing to investigate either the tampereddomains of the image maybe rebuilt from the information ofposition of counterfeiting, in a computationally feasible habit.

References

[1] Zhang, Z., Kang, J., Ren, Y.: An effective algorithm of image splicing detection. IEEE Int. Conf. Comput. Sci. Softw. Eng. 1, 1035–1039 (2008)

[2] Hsu, Y.-F., Chang, S.-F.: Detecting image splicing using geometry invariants and camera characteristics consistency. In: IEEE International Conference on Multimedia and Expo, pp. 549-552 (2006)

[3] Pham, N.T., Jong-Weon, L., Goo-Rak, K., Chun-Su, P.: Efficient image splicing detection algorithm based on markov features. Multimedia Tools

[4] Wu, Y., Abd-Almageed, W., Natarajan, P.: Deep matching and validation network: An end-to-end solution to constrained image splicing localization and detection. In: Proceedings of the 25th ACM

[5] Pomari, T., Ruppert, G., Rezende, E., Rocha, A., Carvalho, T.: Image splicing detection through illumination inconsistencies and deep learning. In: 25th IEEE International Conference on Image Processing (ICIP), pp. 3788-3792 (2018)

[6] Muhammad, G., Al-Hammadi, M.H., Hussain, M., Bebis, G.: Image forgery detection using steerable pyramid transform and local binary pattern. Mach. Vis. Appl. 25(4), 985–995 (2014)

[7] Dong, J., Wang, W., Tan, T.: Casia image tampering detection evaluation database. In: IEEE China Summit and International Conference on Signal and Information Processing, pp. 422-426 (2013)

[8] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[9] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[10] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[11] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[12] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[13] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

[14] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[15] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[16] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.