International Journal of Pattern Recognition and Artificial Intelligence

Altered Handwritten Text Detection in Document Images Using Deep Learning

--Manuscript Draft--

Manuscript Number:	IJPRAI-D-23-01569R1
Full Title:	Altered Handwritten Text Detection in Document Images Using Deep Learning
Article Type:	Research Paper
Keywords:	Altered Handwritten Text; Document Forgery; Feature Extraction; Deep Transfer Learning; Multiclass Classification
Corresponding Author:	Palaiahnakote Shivakumara, Ph.D University of Malaya Faculty of Computer Science and Information Technology: Universiti Malaya Faculty of Computer Science and Information Technology Kuala Lumpur, Kuala Lumpur MALAYSIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Malaya Faculty of Computer Science and Information Technology: Universiti Malaya Faculty of Computer Science and Information Technology
Corresponding Author's Secondary Institution:	
First Author:	Gayatri Patil
First Author Secondary Information:	
Order of Authors:	Gayatri Patil
	Palaiahnakote Shivakumara, Ph.D
	Shivanand S Gornale
	Daniel P Lopresti
Order of Authors Secondary Information:	
Abstract:	Handwritten documents possess immense significance in domains such as law, history, and administration. However, they are vulnerable to forgery, which can undermine their credibility and reliability. This paper aims to establish a dependable technique for identifying altered text in handwritten document images, even in scenarios with high levels of noise and blur. Our study investigates ten distinct categories of handwritten text that have been altered through various forgery operations. The suggested approach employs the deep neural architectures VGG16 and Resnet50 as feature extractors. The architecture comprises three parts: feature extraction using individual models, a feature fusion layer, and a classification layer. Initially, we optimize the training process and feature extraction using VGG16 and ResNet50. The feature vectors obtained from both models are then fused together in the feature fusion layer and input into the classification layer for the classification task. Experiments are conducted on a custom-created dataset as well as benchmark datasets including ICPR FDC, IMEI Forged Number, and Kundu to demonstrate that the proposed method is superior to existing approaches.
Response to Reviewers:	Point to Point Responses to Reviewer’s Comments Title: Altered Handwritten Text Detection in Document Images Using Deep Learning Paper ID: No. IJPRAI-D-23-01569 Reviewer#1 Comment 1.1: Authors have established a dependable technique for identifying altered text in handwritten document images, even in scenarios with high levels of noise and blur. I think this paper is quite interesting with algorithms improvement. Experiments have proved its effectiveness. However, I have several suggestions for authors to improve.

Response 1.1: Thank you very much for your encouraging feedback. We have considered your comments and suggestions and modified the manuscript to improve the quality and clarify of the paper. The changes we have made to the Related Work, the Proposed Method, and the Experimental Results sections are marked in yellow.

Comment 1.2: 1. English writing could be improved. 2.I have several citation suggestions for you.
https://doi.org/10.1016/j.patrec.2023.05.025, https://doi.org/10.1145/3587038, https://doi.org/10.1016/j.patrec.2022.05.004

Response 1.2: Thank you very much for your suggestions. The revised manuscript has been reviewed and edited by a native English speaker to improve the quality of the writing. We have added the citations you suggested as [3, 4 and 27] in the revised manuscript.

Reviewer#2

Comment 2.1: The problem addressed in this paper is very interesting and appropriate in the context of handwriting recognition. Here are few minor suggestions:

Response 2.1: Thank you for your comments and suggestions. We have incorporated these in the revised manuscript. The changes we have made are marked in yellow.

Comment 2.2: The reason of choosing VGG19 and ResNet as a feature extractor need to be mentioned.

Response 2.2: Thank you. We have the added more motivation in the Introduction Section in the revised manuscript. In our study, the selection of VGG16 and ResNet50 as feature extractors for altered handwritten document images in blurry and noisy environments is motivated by their unique capabilities in addressing the challenges presented by such situations. VGG16's simple yet efficient design, utilizing small convolutional filters, is well-suited for capturing intricate spatial features. This enables it to effectively identify subtle patterns in handwritten content, even when obscured by blur or noise. On the other hand, ResNet50's innovative residual learning framework excels in handling noisy environments by mitigating the vanishing gradient problem, ensuring that the model can proficiently learn and extract meaningful features from handwritten text amidst the presence of noise. By integrating the architectural features of VGG16 and ResNet50, we leverage the synergistic advantages to establish a feature extraction framework that effectively tackles the challenges posed by blur and noise in altered handwritten documents. Consequently, this approach significantly improves the performance of our model in demanding real-world scenarios.

Comment 2.3: Why other popular networks are not tried?

Response 2.3: Thank you for raising this point. We agree that there are other popular models for feature extraction and classification in the literature. For example, ImageNet is powerful method for extracting visual features and hence is well suited for object classification and recognition in contrast to the combination of VGG and ResNet proposed in this work for altered handwritten text classification [11]. We included this observation in motivating our approach in the revised manuscript. We believe it would be interesting future research to perform experiments combining ImageNet and ResNet for feature extraction and classification.

Comment 2.4: Please improve the figure 2 for better readability.

Response 2.4: We apologize for the poor quality. We re-drew the figure so that the text in the architecture is now clear and readable, as shown in Figure 2 in the revised manuscript.

Comment 2.5: Show some error samples in the discussion and also mention about the reason for such error.

Response 2.5: Thank you for this suggestion. We have added more samples of failure cases in Figure 14. In addition, we have provided possible reasons for such misclassifications and discuss future work to address these challenges in Section 4.4 in the revised manuscript.

Reviewer#3

Comment 3.1: This paper has to be revised and improved significantly as follows:

Response 3.1: We have carefully considered all of the comments and suggestions made by the reviewers to improve the quality and clarity of the revised manuscript. These can be seen throughout the introduction, related work, proposed methodology, and experimental results sections in the paper.
Comment 3.2: The novelty or fresh new ideas is not very clear, does not have enough comparisons with other methods in the literature.

Response 3.2: Thank you for this feedback. Detecting altered handwritten text in noisy and blurred environments is a complex and challenging problem compared to the normal case of forgery detection in document images. We believe that exploring the combination and optimization of existing models for addressing this problem is a useful contribution to the field. In our paper, we adapt VGG16 and the ResNet50 rather than using the baseline models as shown in Figure 4 and Figure 6. Furthermore, the way we fuse the two adapted models is new as shown in Figure 2. To show that the proposed work improves on state-of-the-art methods, we implemented and tested a number of existing methods [1, 2, 14, 15, 18 and 22]. We also added some additional references to recent work [24, 25, 26] which are cited in the Related Work section in the revised manuscript. Our motivation and a list of key contributions are presented in the Introduction section in the revised manuscript. An existing method [22] has been added as a comparative study in the Experimental Section.

Comment 3.3: What is its relation to PR & AI? (Pattern Recognition & Artificial Intelligence) or related area(s)? Should explain it more clearly, i.e., What exactly is it talking about? AI? PR? or Image Processing? Computer Vision? or? A good, complete and thorough bibliography is part of a well written paper.

Response 3.3: We have attempted to clarify these points in the revised manuscript. We note that when forgery operations are performed on images, they introduce distortions, altering the content. This results in unique patterns. We note that when forgery operations are performed on such images, it introduces distortions, altering the content. This results in unique patterns that are detectable. To extract such patterns, the method we present explores the combination of deep learning models to extract these unique patterns. Overall, our approach combines techniques from pattern recognition, image processing, computer vision, and artificial intelligence.

The above explanations is added to the Introduction section in the revised manuscript.
Comment 3.4: Need to add more recent related publications in the reference list including, for examples 2015, 2016, 2017PR & AI subfield,

Response 3.4: Thank you for this suggestion. Yes, we cited the existing methods [22, 24, 25, 26] in the Related Work section and the same were updated in the Reference section in the revised manuscript. The newly cited methods are marked in yellow.

Comment 3.5: Compare your work with others in the literature, advantages vs disadvantages, in depth and width,

Response 3.5: Thank you for raising this point. As discussed in our response to Comment 3.4, existing methods are discussed in the paper and compared and contrasted to our proposed approach. This critical analysis includes discussing advantages and disadvantages of the existing methods. Overall, the related work section has been updated in the revised manuscript to reflect this.

Comment 3.6: Should use real images with noisy and testing samples set size should be large enough to show your results are indeed convincing, reliable, and meaningful.

Response 3.6: Thank you – we agree it would be helpful to include more samples of

	images and have done so. Our dataset includes sample forged images created artificially as well as images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially in Figure 8. Our dataset includes sample forged images created artificially and images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially from Figure 8. We believe the dataset we have constructed is sufficiently diverse for evaluating our proposed method. We have made suitable changes to Section 4.1 in the revised manuscript.

Altered Handwritten Text Detection in Document Images Using Deep Learning

Gayatri Patil[1*], Palaiahnakote Shivakumara[2], Shivanand S. Gornale[1] and Daniel P. Lopresti[3]
[1]Department of Computer Science, Rani Channamma University, Belagavi, India

Email: gayatripatil865@gmail.com, Email: shivanand1971@rcub.ac.in

[2]School of Science, Engineering and Environment (SEE), University of Salford, Manchester, UK

Email: S.Palaiahnakote@salford.ac.uk

[3]Lehigh University, USA, lopresti@cse.lehigh.edu

Abstract

Handwritten documents possess immense significance in domains such as law, history, and administration. However, they are vulnerable to forgery, which can undermine their credibility and reliability. This paper aims to establish a dependable technique for identifying altered text in handwritten document images, even in scenarios with high levels of noise and blur. Our study investigates ten distinct categories of handwritten text that have been altered through various forgery operations. The suggested approach employs the deep neural architectures VGG16 and Resnet50 as feature extractors. The architecture comprises three parts: feature extraction using individual models, a feature fusion layer, and a classification layer. Initially, we optimize the training process and feature extraction using VGG16 and ResNet50. The feature vectors obtained from both models are then fused together in the feature fusion layer and input into the classification layer for the classification task. Experiments are conducted on a custom-created dataset as well as benchmark datasets including ICPR FDC, IMEI Forged Number, and Kundu to demonstrate that the proposed method is superior to existing approaches.

Keywords: Altered Handwritten Text, Document Forgery, Feature Extraction, Deep Transfer Learning, Multiclass Classification

1. Introduction

Handwritten document forgery is the act of imitating, copying, or tampering with a portion of a handwritten document such as a signature, handwritten text, or a hand drawn image. of the goal is to alter an existing document or create a new document with the intent of deceiving others. With the increased availability of image editing tools and software, the prevalence of handwriting forgeries has increased, giving rise to a number of sensitive applications such as generating fake suicide notes, forging certificates, and tampering with legal documents [1,2]. The results can be very convincing for an untrained eye.

Handwriting analysis and text detection [3, 4] has been used for decades in forensic investigations to judge whether a document is genuine or a forgery. While many techniques have been developed to detect forged handwriting, these approaches are not as effective for forged handwriting affected by noise and blur [5,6]. As a result, noise or distortion resulting from degradations such as blur, low contrast and low resolution are not accounted for. Moreover, aging of paper and ink and differences in writing styles and pens can also introduce challenges [7,8]. Detecting forged or altered handwriting can be a complex problem. Hence, there is need for new methods that are robust to noise and blur.

This study categorizes forged handwriting detection as a ten-class problem, which includes as variations: Blur, Noise, copy paste, Insertion, copy paste + insertion, copy paste+ Blur, Copy paste + Noise, Insertion + Blur, Insertion + Noise and Normal. A detailed explanation of these ten different classes is presented in the Dataset Section of this paper. Sample images for each class are shown in Fig. 1 which illustrate the complexity of the problem.

We note that when forgery operations are performed on such images, it introduces distortions, altering the content. This results in unique patterns that are detectable. To extract such patterns, the method we present explores the combination of deep learning models to extract these unique patterns. Overall, our approach combines techniques from pattern recognition, image processing, computer vision, and artificial intelligence.

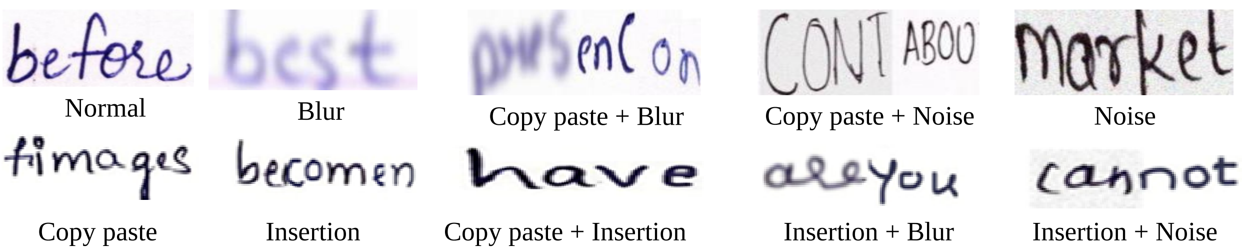


Figure. 1. Sample Images of ten different forgey types.

Thus, the objective of this study is to develop a new approach for detecting altered handwritten text in noisy and blurred environments, including situations combining multiple effects. It is observed that when images are affected by noise and blur, specific patterns appear. As the level of noise and blur increases, these patterns become more noticeable [1]. On the other hand, forgery operations do not typically produce any discernible distortion patterns. Additionally, when text is affected by multiple operations, the level of distortion uncertainty increases. Inspired by the success of deep learning models for solving complex classification problems, we explore the combination of VGG16 and ResNet50 for feature extraction and identification of these 10 classes. The reason for choosing the VGG16 and ResNet50 as feature extractors for our problem is that these models exhibit special capabilities for tackling the inherent challenges. VGG16's simple yet efficient design, utilizing small convolutional filters, is well-suited for capturing intricate spatial features [9]. This enables it to effectively identify subtle patterns in handwritten content, even when obscured by blur or noise. On the other hand, ResNet50's innovative residual learning framework excels in handling noisy environments by mitigating the vanishing gradient problem, ensuring that the model can proficiently learn and extract meaningful features from handwritten text amidst the presence of noise [10]. By integrating the architectural features of VGG16 and ResNet50, we leverage the synergistic advantages to establish a feature extraction framework that effectively tackles the challenges that arise in this problem domain. Consequently, this significantly improves the performance of our approach in demanding real-world scenarios. There are, of course, other popular models for feature extraction and classification available in the literature. For example, ImageNet, is a powerful method for extracting visual features and hence is well-suited for object classification and recognition [11] in contrast to the combination of VGG16 and ResNet50 we employ in our work on altered handwritten text classification.

The key contributions of this paper are as follows:

- Exploring the combination of VGG16 and ResNet50 for forged handwritten text detection in complex environments.
- Fusing the features extracted by VGG16 and ResNet50 in a novel way to optimize the performance of our system.
- Demonstrating that integrating the strengths of VGG16 and ResNet50 produces results that improve on the state of the art.

The paper is divided into sections that start with a discussion of deep learning algorithms for detecting forged words in handwritten documents. It also highlights important contributions in this field and provides details of our proposed approach. We then describe the datasets used in our algorithm design and system assessment, and present experimental results and our analysis. The paper concludes by discussing possibilities for future research.

2. Related Work

This section presents a comprehensive literature review of prior research conducted on detecting forgeries in handwritten document images.

Kundu et al [2] conducted a study with the aim of classifying text into four distinct categories: noisy, blurred text, normal text, and forged text. Their objective was to categorize the categorization of texts into these four groups. In order to achieve this, their methodology employs spectral density and variations as distinguishing features to identify forged text. However, it is crucial to acknowledge that this method is only effective for images that have been affected by a single forgery operation, and not for multiple operations.

Amr Megahed et al [12] suggested a method for identifying handwriting forgery by analyzing ink color characteristics. Instead of traditional approaches, their method employs image processing to identify different inks used in the forged text. This involves scanning handwritten documents as images and segmenting them into individual objects. From each object, nine features are extracted based on the red, green, and blue channels using measures such as mean, standard deviation, and skewness. It is important to note that this method is specifically designed to detect copy-paste or insertion types of forgery and may not be effective for other types of forgery, especially in noisy or blurry environments.

Priyanka Roy and Soumen Bag [13] developed a method to detect forgery in handwritten documents by analyzing the writer's handwriting. They used a sliding window technique that moves across the document from top-left to bottom-right, similar to a scanner. They tested the method on a IAM dataset of 10,000-word images from 50 people and found that it successfully detects forgery. However, it may not be able to identify different types of forgeries. Lokesh Nandanwar et al [14] have suggested a technique for identifying forged handwritten words by utilizing Chebyshev-Harmonic-Fourier-Moments (CHFM) and deep Convolutional Neural Networks. Their approach concentrates on detecting incongruities and abnormal modifications caused by forgery actions such as noise, blur, and copy-pasting on handwritten papers. Nonetheless, it should be noted that this method is specifically intended for detecting individual forgery actions and may not be efficient in cases where multiple forgery actions are involved.

Ying Chen and Shuhui Gao [15] studied the detection of forged numeral handwriting using a convolution neural network. Their approach targets the identification of forged numbers in handwritten documents, rather than detecting forged text. Priyanka Roy and Soumen Bag [16] have presented a novel concept for detecting fraudulent handwriting, specifically through word alteration. Their innovative approach allows for the detection of forged documents by analyzing visually similar ink from various pens, such as blue and black. The proposed technique entails scrutinizing handwritten words for any modifications, whereby added letters significantly alter the intended meaning. This method is framed as a binary classification problem, offering a reliable solution to identifying fraudulent handwriting.

Shivakumara et al [17] developed a technique to detect forged IMEI numbers using a fusion process and color space approach. Their utilization of connected component analysis demonstrates a meticulous and refined technique for detecting forged IMEI numbers. However, it is worth noting that their experimentation solely focused on images of IMEI numbers, without considering handwritten text images. As a result, it is uncertain if the method is suitable for identifying forged handwritten text images. The study by Nandanwar et al [18] suggests a technique to identify fake text in different types of document images using the phase spectrum derived from a combination of DCT and FFT. The method extracts phase statistics as features and uses a Support Vector Machine for classifying original and forged images. However, this approach is only effective for clean images and has limitations in other scenarios.

Humayun et al. [19] presents a method that uses unsupervised learning to automatically detect the various inks in a hyper-spectral document image. This approach can be beneficial in the initial phases of identifying forgery in hyper-spectral document images but not as a complete method for detecting forgery in handwritten document images. The aim of the method discussed in Patil et.al. [1] proposed a method to tackle the difficulties of identifying altered handwritten text in environments that are noisy and blurred. This goal is achieved by combining statistical, gradient, and texture features with a Bayesian classifier. The study emphasizes the novelty of the proposed approach, the development of a distinctive dataset that encompasses ten categories of altered handwritten text, and the superior performance of the proposed method in comparison to existing methods.

The method explained in Jaiswal [20] presents an approach for detecting ink mismatches in hyperspectral document images using unsupervised deep learning. It utilizes a convolutional auto encoder (CAE) to extract deep features from the images, and then applies logistic regression (LR) for classification. However, this method is specifically designed for detecting ink mismatches in blue and black ink only. It may not be able to effectively detect multiple forgery operations that have been performed on the document images. Priyanka Roy [21] has developed a technique for identifying forged handwritten legal documents by analyzing the writing style. Their method focuses on legal business contracts, cheques, and invoices, and can determine whether a document is genuine or fake based on the writers handwriting. However, in situations where multiple forgery operations have occurred in a noisy and blurry environment, the method may not be effective in detecting the forgery.

Qu et al. [22] introduced a framework known as Document Tampering Detector (DTD) to identify tampered text in document images. This framework combines a Frequency Perception Head (FPH) and a Multi-View Iterative Decoder (MID) to detect subtle clues in challenging situations. The method focuses on printed document images that have been altered using three forgery techniques: copy-paste, splicing, and generation. However, a limitation of this method is that it only works well with printed document images that have a small number of forgery operations. It may not be effective for identifying tampering in handwritten document images that have multiple forgery operations in noisy or blurry environments. Fadl, S. et al. [23] proposed a method that aims to automatically detect altered handwritten documents and identify the forged part using DE-Net. To locate the forged parts in the forged document images, a localization schema is applied, where each forged document is segmented into objects. A fused feature vector is generated for each object using color histograms of the R, G, and B channels. The structural similarity index (SSIM) is then used to identify the lower similarity parts as forged. However, the main limitation of this method is that it only works for addition and alteration operations on document images based on ink mismatch and may not be effective for multiple forgery operations in handwritten document images with noisy and blurry environments.

Cha et al. [24] proposed a novel method to detect counterfeit handwriting created by inexperienced individuals using an automated system. Their study reveals that skilled forgeries exhibit a slower writing speed, leading to a more “wrinkled” appearance compared to genuine writing. The authors explain the process of obtaining authentic and forged handwriting samples, as well as extracting various handwriting features, including the wrinkleless characteristic. Experimental results show that the proposed system achieves an accuracy rate of 89% in identifying forged handwriting by utilizing a neural network trained with eight handwriting distance features, including the wrinkleless feature. This proposed method may fail to work on detecting forgeries when text is affected by multiple forgery operations. Chen et al. [25] introduced a novel approach to identify instances of website tampering through text comparison. The proposed method entails extracting the homepage, JS, CSS, and image files from the website for subsequent analysis. Four comparison algorithms, namely string comparison, hash-based (MD5, SHA-1), and compression-based, are assessed. The experimental findings reveal variations in the performance of these algorithms, with string comparison demonstrating notable advantages but the method may not work when the website contains text with noisy and/or blurry environments. Fahn et al. [26] proposed a forgery detection system that uses branchlet features and Gaussian mixture models (GMMs). The system analyzes input handwriting images and extracts character stroke skeletons. From these skeletons, branchlet points are extracted to obtain handwriting features. The extracted feature data is grouped, and GMMs are created for each group. The similarity between each group and the input images is measured. Input images with higher similarity values are considered indicators of genuineness. A new GMM is created using these genuine images to measure similarity with all input images. The sample mean and standard deviation of the measured similarity values are calculated. Input handwriting images with similarity values below a certain threshold are predicted to be forgeries.

After reviewing the past methods, it can be seen that detecting multiple forgeries in handwritten documents in noisy or blurred environments remains an open problem. Our goal is to develop a method that can identify multiple forgeries in handwritten document images. A summary of the existing methods discussed earlier is listed in Table 1. None of the methods achieve high accuracy for handwritten document images affected by multiple forgery operations.

Table 1. Details of few existing methods for forgery detection in document images

Author	**Methodology**	**Dataset**	**Classifier**	**Results**	**Objective**	**Drawback**
G. Patil et.al. [1]	Statistical, Histogram Oriented Gradients (HOG), Local Binary Pattern (LBP)	Own Dataset, , ICPR 2018 FCD, ACPR 2019, and IME number	Naïve Bayes	96.80%	is able to classify multiple forgery operations.	It still misclassifies the images.
Kundu et al [2]	spectral density	Own dataset, Bharadwaj et at. [27], Elkasrawi et al. [28]	Neural Network	77.5%	To identify forged handwritten words	Effective for images with one alteration, but less effective for images with multiple forgery operations.
Amr Megahed et al [12]	R,G,B color components	Own created English handwriting dataset	Outlier detection method	85%	For identifying handwriting forgery by analyzing ink color characteristics	The method does not work in noisy and blurry environments
Priyanka Roy and Soumen Bag [13]	contour related sliding window based features	IAM dataset and IDRBT check image dataset	REPTree classifier	89.64%	To detect forgery in handwritten documents by analyzing the writer's handwriting.	Is able to identify different types of forgeries
Nandanwar et al[14]	combination of Chebyshev Harmonic Fourier-Moments and deep Convolutional Neural Networks	Own dataset, ICPR 2018 FCD, ACPR 2019, and IME number	High pass Deep CNN	78.6%	To detect forged handwritten words.	May fail with obscured forged text due to natural handwriting variations.
Chen et al. [15]	convolutional neural networks,	handwritten forged numeral samples database	Neural Network	95.35%	To identify forged numeral handwriting	May not be effective for complex background forgery operations.
Priyanka Roy and Soumen Bag [16]	statistical features	Own dataset	MLP classifier	83.71%	To detect fraudulent handwriting through word alteration	May not be effective for words affected by blur and noise
Shivakumar a et al [17]	Connected components	Forged IMEI Numbers. Roy et al [31, Bharadwaj et al. [27]	Neural Network	82%	to detect forged IMEI numbers	The method may not work on handwritten images.
Nandanwar et al [18].	phase spectrum using the Discrete Cosine Transform (DCT)	forged IMEI numbers and air ticket images	Support vector machine	80%	To identify fake IMEI numbers and modified tickets for security purposes.	Does not work on blurry and noisy images.
Humayun et al. [19]	K-means clustering	iVision HHID dataset	KNN	96%	To automatically detect the various	Is not a complete method for

					inks in a hyper-spectral document image	detecting forgery in handwritten document images
Priyanka Roy [21]	SIFT Features, SURF (Speeded Up Robust)Features	IAM dataset and IDRBT cheques dataset	Ensemble	90%	for identifying forged handwritten legal documents by analyzing the writing style.	It does not work on blurry and noisy images
Qu et al. [22]	Frequency Perception Head and Multi-View Iterative Decoder	DocTamper	Based on printer classification techniques	92%	To identify tampered text in document images	Limited to printed document images with few forgery operations and does not work in noisy and blurry environments
Cha et al. [24]	Based on Wrinkleless of text	Own dataset	Feed Forward Neural Network	89%	to detect forged handwriting created by inexperienced individuals using an automated system	May fail to work on detecting forgery when text is affected by multiple forgery operations
Fahn et al. [26]	Branchlet features and Gaussian mixture models(GMMs)	IAM Handwriting Database	Gaussian mixture models classifier	95%	To detect the forgery in handwritings	Does not work on blurry and noisy images

3. Proposed Methodology

As noted from the previous section, detecting multiple forgery in the handwritten document images is a complex problem. Our key insight is that forgery operations on images introduces distinctive distortion. This observation has been used to classify forged and original images [2]. However, to classify the images affected by forgery + noise and forgery + blur from other images, it is observed that the pixel distribution in forgery + noise and forgery + blur is arbitrary, while the pixel distribution in noisy and blurred images that do not involve forgeries follow a more regular pattern [2]. The reason is that Gaussian noise and blur usually generates regular distributions whilst forgeries do not. To exploit the above observation, inspired by deep learning models, namely VGG16 and ResNet50 that are successful in extracting dominant and contextual features for classification [9,10, 29], we explore the combination of VGG16 and ResNet50 to create a unified model for forgery detection. We believe that contextual feature has the ability to differentiate original, blurred, noise, forged images, forgery + noise and forgery + blur. A block diagram is shown in Fig. 2, which depicts the steps, including the flow and the various classes of forged text to be classified.

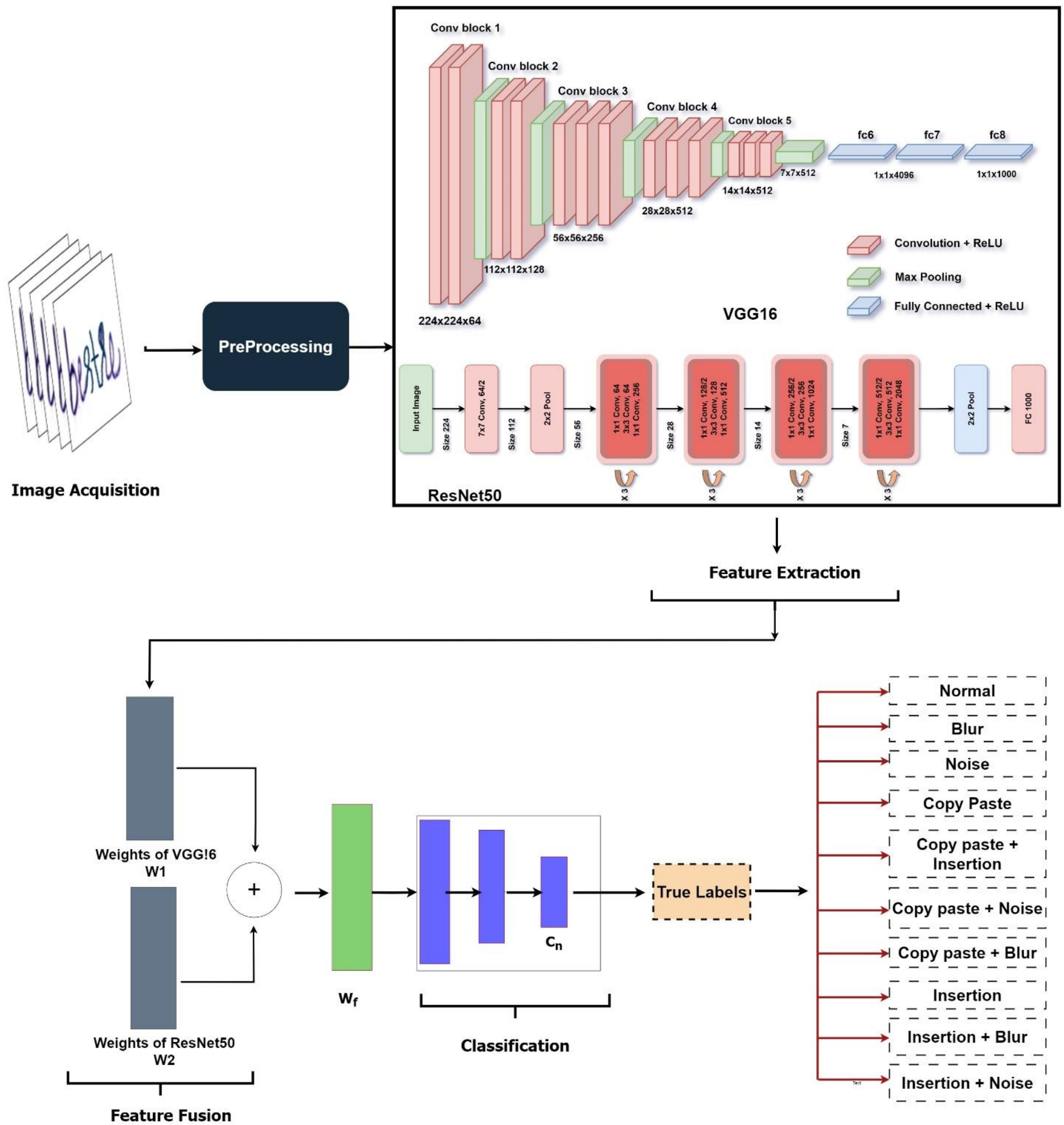


Figure 2. An architectural diagram that outlines the proposed approach.

3.1. VGG16 and ResNet50 for Feature Extraction

Visual Geometry Group (VGG16); Andrew Zisserman and Karen Simonyan, members of Oxford's Visual Geometry Group [9], first proposed the VGG model in 2013 and created a prototype for the 2014 ImageNet Challenge. The model, named after its characteristics, consists of a neural network layer that is 16-deep and contains 138 million parameters. By today's standards, this is a large network. Despite its complexity, the VGG-net architecture is straightforward and incorporates essential features of convolutional neural networks. The benefit of using VGG16 is capable of extracting various low, mid, and high-level features from the input image as it processes through the layers. The early layer's capture simple features, and as the information flows through the network, the features become more abstract and complex. This hierarchical representation allows VGG16 to perform well on a wide range of visual recognition tasks, especially image classification. The details description of the VGG16 model is represented in Figure 3.

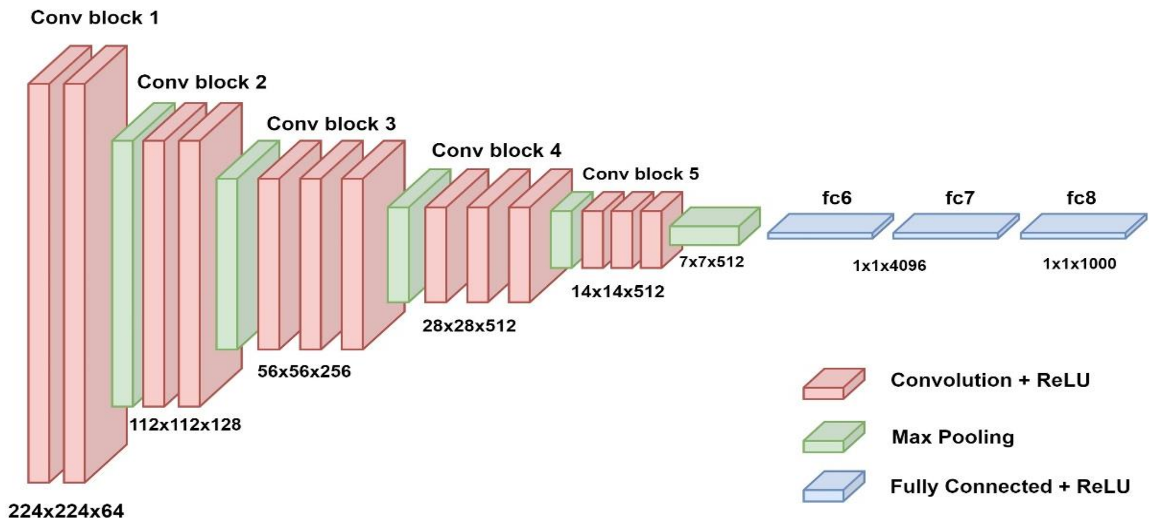


Figure 3. Architecture of VGG16 Deep Neural Network

Figure 3 represents the architecture of VGG16 Deep Neural Network. The network is made up of 13 convolutional layers and 3 fully connected layers, which are composed of small convolution filters. The input of a fixed-size 224×224 RGB image is received at our ConvNets during the training phase. The preprocessing we perform involves subtracting the average RGB value, which is calculated based on the training set, from each pixel. The image then goes through several convolutional layers, where we use 3×3 filters to capture the concepts of left/right, up/down, and center. In some cases, we also use 1×1 convolution filters, which act as a linear transformation of the input channels followed by a non-linear step. The convolution stride remains at 1 pixel, and the spatial padding of the convolutional layer input is set to preserve the spatial resolution after convolution. For 3×3 conv. layers, the padding is 1 pixel. We perform spatial pooling using five max-pooling layers, applied after certain conv. layers. Max-pooling is done over a 2×2-pixel window with a stride of 2. After the stack of convolutional layers, we have three Fully-Connected (FC) layers. The first two FC layers have 4096 channels each, while the third FC layer handles the 1000-way ILSVRC classification with 1000 channels (one for each class) [32, 33]. The final layer is the SoftMax layer. Figure 4(a) represents the visualization of intermediatary layers of VGG16 on sample image of our custom created dataset.

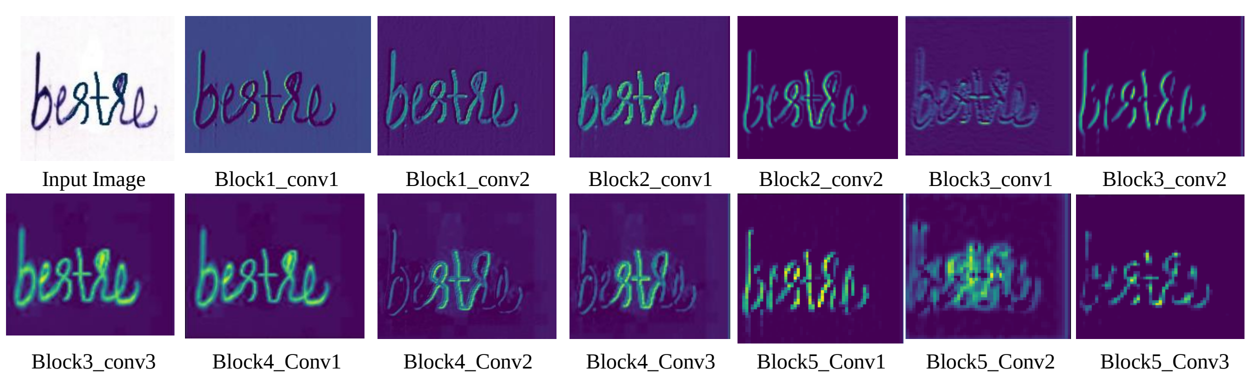


Figure 4(a). Visualization of the learning of the intermediate layers of VGG16 on sample image of our custom created Dataset

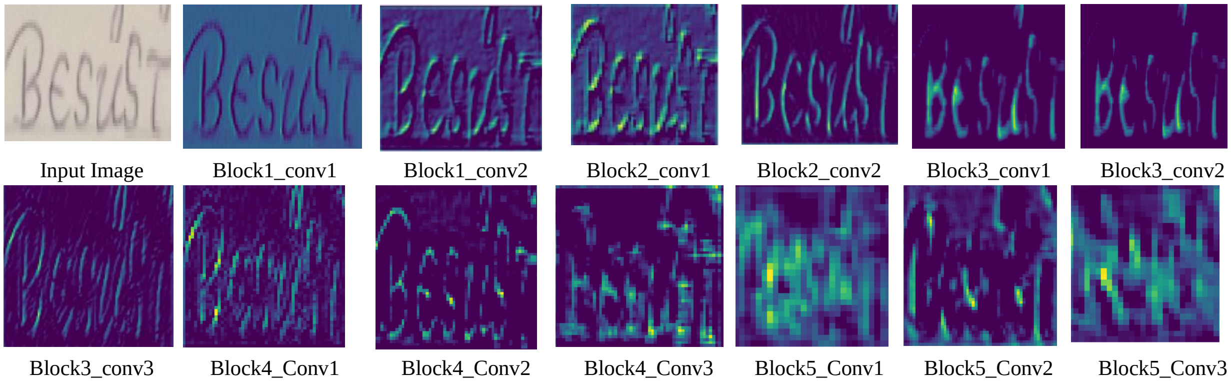


Figure 4(b). Visualization of the learning of the intermediate layers of VGG16 on a sample naturally blurred image

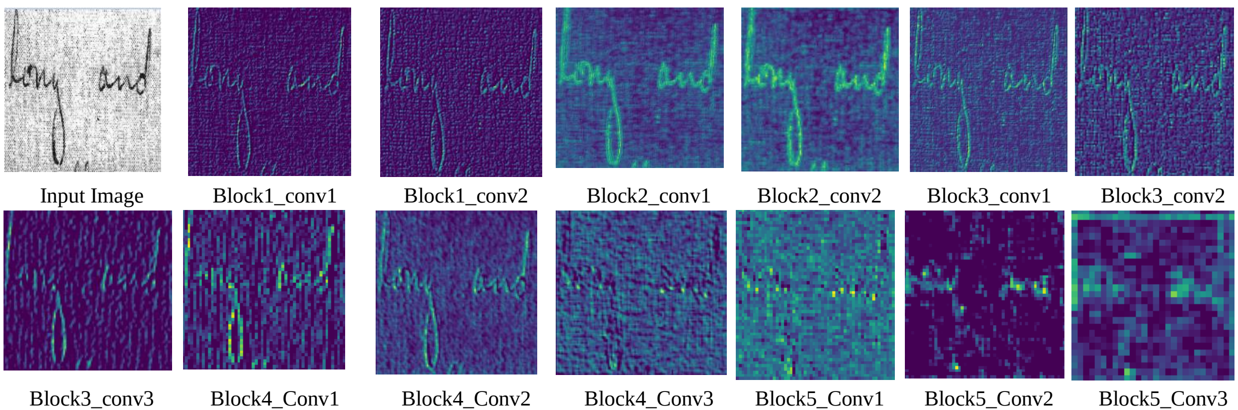


Figure 4(c). Visualization of the learning of the intermediate layers of VGG16 on a sample naturally noisy image

A visualization of intermediate layers in the VGG16 and ResNet50 models when applied to naturally blurred and noisy images as is shown in Figure 4(b), (c) for VGG16 and Figure 6(b), (c). This research aims to uncover the mechanisms by which these architectures can effectively handle the challenges presented by both blur and noise. In doing so, valuable insights can be gained, leading to the development of forgery detection approaches that exhibit enhanced performance in complex real-world scenarios.

Resnet50 for Feature Extraction: ResNet-50 is a noteworthy deep neural network architecture, comprising a stack of 50 convolutional layers, each integrated with skip connections, also referred to as residual connections. These ingenious residual connections play a pivotal role in addressing the widespread vanishing gradient problem often encountered in the training of deep neural networks. In conventional deep networks, the increase in layer depth typically compounds the challenges of training, resulting in higher error rates and diminished accuracy. However, ResNet-50's well-crafted design adeptly overcomes these hurdles, allowing the model to proficiently capture intricate features while efficiently managing a controlled number of trainable parameters. Consequently, ResNet-50 emerges as a streamlined yet remarkably potent model, suitable for a diverse range of computer vision tasks.

Utilizing ResNet-50 for feature extraction in the analysis of handwritten forged text offers distinct advantages in fraud detection. Its deep architecture excels at capturing intricate features, enabling the detection of subtle fraudulent patterns. The detailed architectural representation of Resnet50 is represented in Figure 5. The model's capacity for transfer learning reduces the need for extensive labeled data, while its generalization ability ensures adaptability to various forgery scenarios, resulting in improved detection accuracy and reduced false positives. ResNet-50's speed and robustness make it suitable for real-time processing, and the hierarchical features it extracts provide interpretability, aiding investigators in understanding flagged document anomalies, ultimately enhancing fraud detection capabilities. Its effectiveness in dealing with forged text lies in its ability to capture intricate features crucial for fraud detection. When applied to such images, the model processes them through its convolutional layers, extracting patterns and details that may reveal inconsistencies or irregularities in handwritten text, or document layouts.

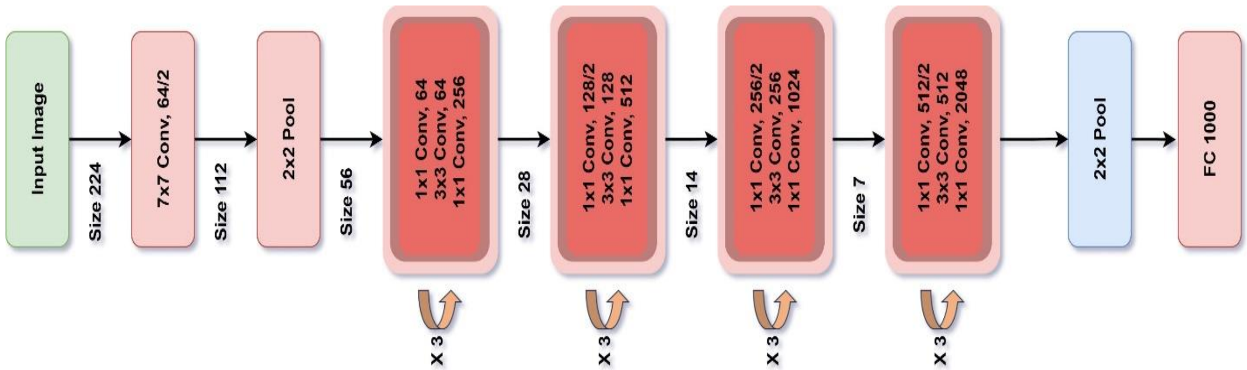


Figure 5. Architecture of ResNet50 Deep Neural Network

It takes input images of a fixed size, typically 224 × 224 pixels. One of its key innovations is the use of residual blocks, which contain convolutional layers with 3×3 kernels, allowing it to capture intricate image features. ResNet-50 includes several residual blocks, and these residual connections enable the training of very deep networks by mitigating the vanishing gradient problem [10]. The use of residual connections in ResNet-50 proves particularly advantageous, as it allows the network to recognize deviations between authentic and forged elements by focusing on the residual information, thereby aiding in the detection of subtle discrepancies. Furthermore, ResNet-50's deep architecture and transfer learning capabilities enable it to generalize across various types of forgery attempts, making it adept at identifying fraudulent documents. After the convolutional layers, ResNet-50 typically concludes with fully connected layers, though the specific details of the fully connected layers can vary depending on the application, with the last layer often having as many units as there are classes in the classification task. Figure 6(a)-(c) represents the visualization of intermediatary layers of ResNet50 on sample image of our custom created dataset.

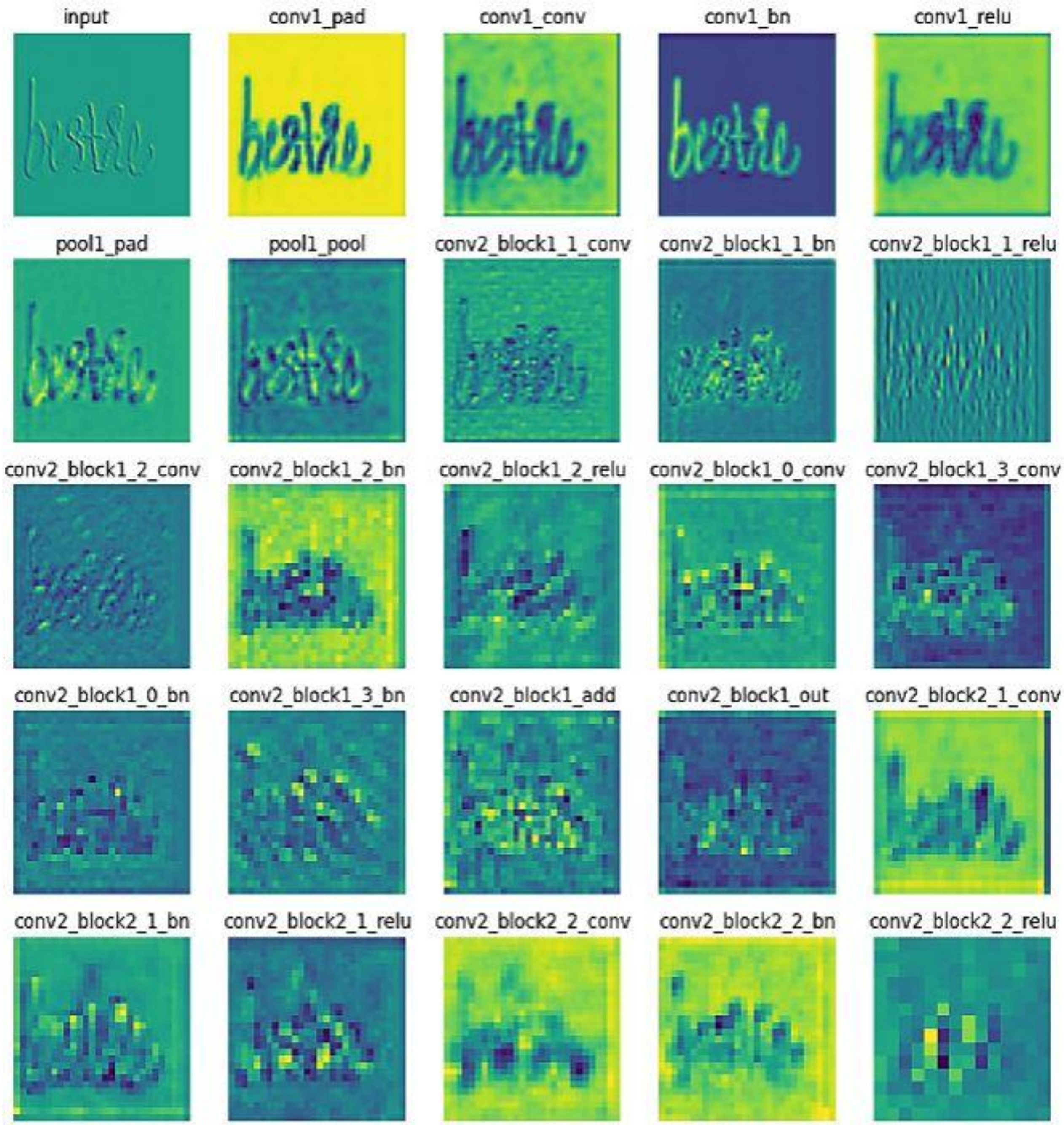


Figure 6(a). Represents the visualization of few intermediatary layers of ResNet50 on sample image of our custom created dataset.

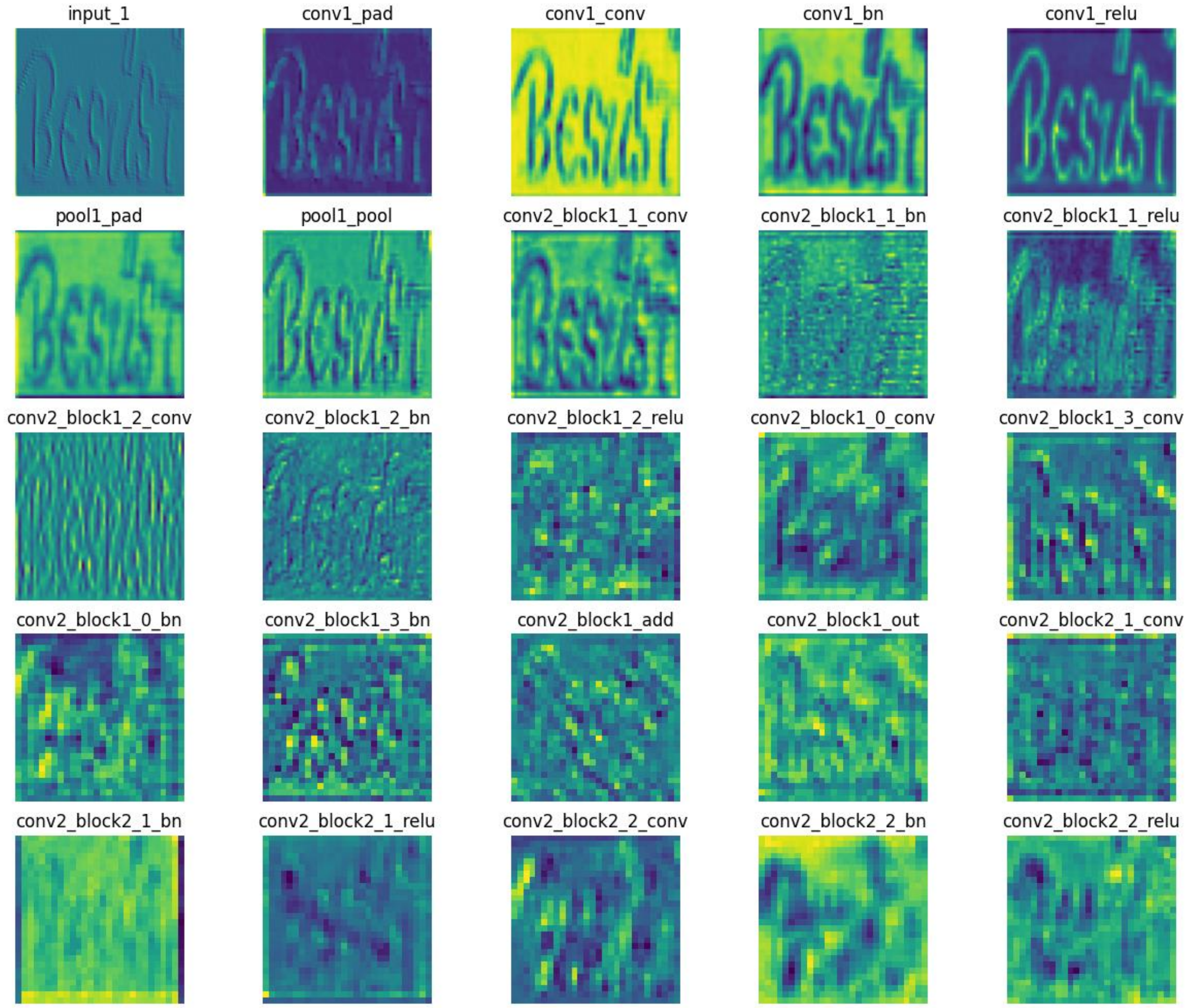


Figure 6(b). Visualizations of a few intermediary layers of ResNet50 on a naturally blurred image

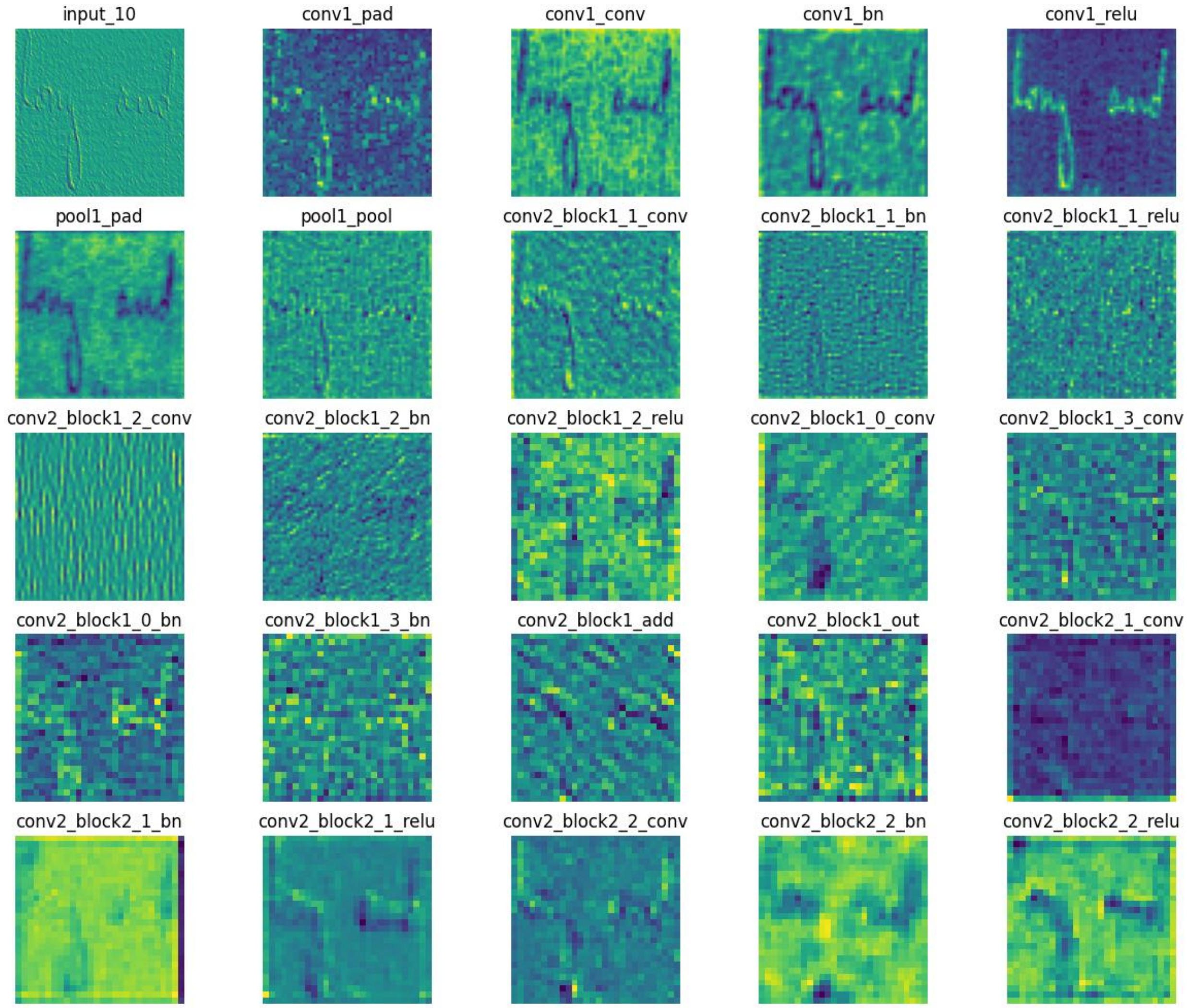


Figure 6(c). Visualizations of a few intermediary layers of ResNet50 on a naturally noisy image

Heat map: In the field of deep learning, a heat map of an image commonly denotes a visualization technique that emphasizes particular regions or areas of interest in an image. Heat maps help gain understanding of where a neural network is concentrating its attention or making significant decisions while processing the image data. **Grad CAM (Gradient-weighted Class Activation Mapping)** is a visualization technique that highlights the important regions of an input image that contribute to the prediction made by a deep learning model. It calculates the gradients of the target class score with respect to the feature maps of the last convolutional layer of the model. These gradients are then used to obtain the importance weights for each channel of the feature maps. By taking a weighted sum of the feature maps using these importance weights, Grad CAM generates the final heat map. It provides a heat map that indicates the regions of the image that are most relevant for the model's decision-making process [32].

To create the class-specific localization map Grad-CAM $L^{C}_{Grad\ CAM} \in R^{u,v}$ with dimensions u x v (width x height respectively) for a specific class c, we start by calculating the gradients of the score for class C, y^{C}, concerning the feature map activations A^{k} in a convolutional layer, denoted as $\frac{\partial y^{c}}{\partial A^{k}}$. These gradients are then globally averaged across the spatial dimensions, typically indexed by i for width and j for height. This pooling process yields neuron importance weights α_k^c, indicating the relevance of each feature map A^k to the classification of class c. Finally, these importance weights are used to construct the desired localization map $L_{Grad\ CAM}^C$ [33].

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \tag{1}$$

Where , $\frac{1}{Z} \sum_i \sum_j$ represents the global average pooling and $\frac{\partial y^c}{\partial A_{i,j}^k}$ represents the gradients via backpropagation.

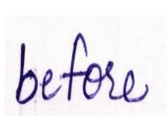
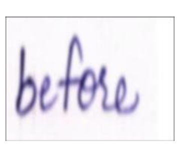
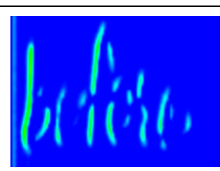
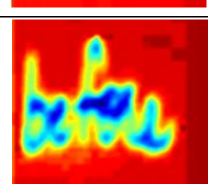
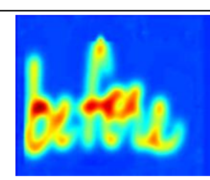
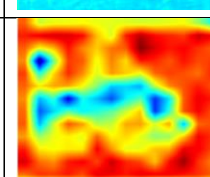
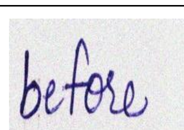
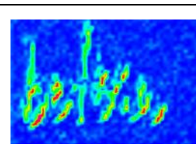
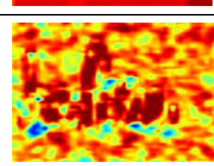
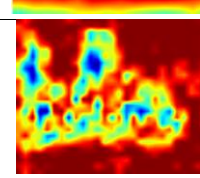
During the process of backpropagation gradients for activations, it is necessary to perform a precise calculation to determine α_k^c. This calculation involves iteratively multiplying the weight matrices with the gradients related to activation functions until reaching the final convolution layer where the gradients are propagated. The weight α_k^c represents a partial linear representation of the deep network downstream from A and reflects the importance of feature map k for a specific target class C. To calculate this weight, we combine forward activation maps and then apply a ReLU function.

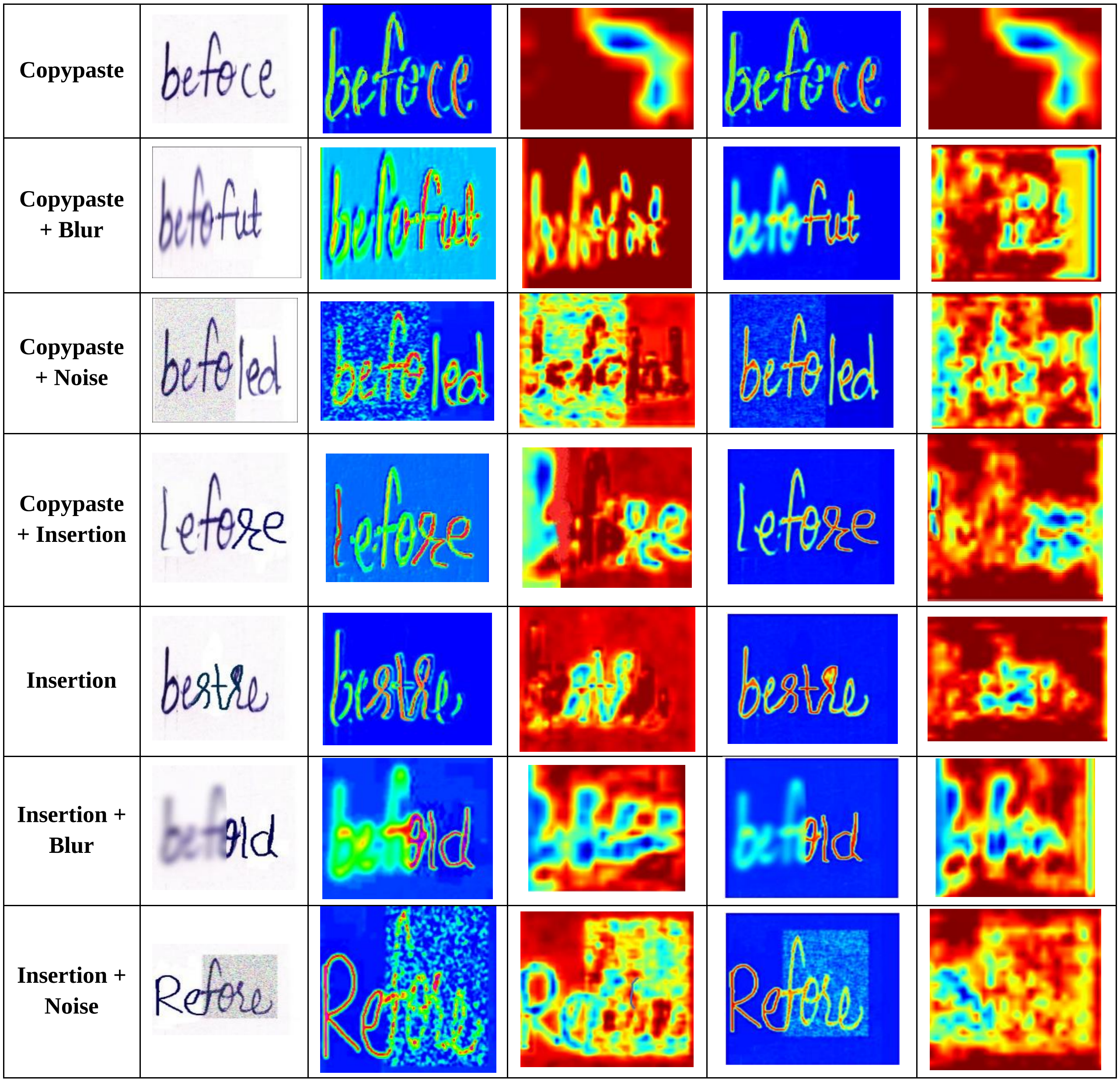
$$L_{Grad\ CAM}^C = \boldsymbol{ReLU}\left(\sum_{\boldsymbol{k}} \alpha_k^c A^k\right) \tag{2}$$

Where, $\left(\sum_{\boldsymbol{k}} \alpha_k^c A^k\right)$ is a linear combination of maps

It should be noted that this will produce a coarse heat map that is the same size as the convolutional feature maps. We use a ReLU activation function on the combination of maps because we are only concerned with the features that positively affect the desired class, meaning the pixels that need to have their intensity increased to increase y^C. On the other hand, negative pixels are probably associated with other categories in the image. Table 2 represents the Heat maps and Class Activation maps for ten different forgery operations including the original image generated by using the VGG16 and ResNet50 deep neural networks.

Table 2. Visualization of Heat maps and Class Activation maps for ten different forgery operations including the original image generated by using the VGG16 and ResNet50 deep neural networks.

Class	Input Images	Heat Map using VGG16	Class Activation Map using VGG16	Heat Map using ResNet50	Class Activation Map using ResNet50
Normal					
Blur					
Noise					



By observing the table, it can be seen that GRAD CAM highlights the regions on an original image that have the most influence on the network's decisions regarding the image. In general, it emphasizes objects or characteristics that are important for the classification task or the specific features learned by the network. When dealing with a blurred image, it could face difficulty in recognizing particular areas of interest due to the absence of clearly defined edges or features to concentrate on. The heat map generated by Grad-CAM might seem less accurate and spread out. For noisy images, Grad-CAM has the ability to identify and emphasize regions that the model has learned to focus on. It specifically highlights areas where the noise is prominent. For the copy-paste forgery operation, the Grad-CAM highlights the pasted area as it differs from the surrounding context Attention based on gradients reveals a clear distinction between the original image and the pasted portion.

When dealing with the operations of copy-past and blurring for forgery, this particular combination may not be very effective in producing the grad-CAM highlights. This is because the blur and copied content may compete for attention. Looking at the copy-paste + Noise operation, Grad-CAM is unlikely to emphasize the pasted region, but it will emphasize regions with distinct noise patterns. For Copy paste and inserted images CAM would probably emphasize both the boundary of the region that has been copied and pasted, as well as the area where new content has been inserted. These particular areas would stand out noticeably due to their distinction from the surrounding background.

For an insertion forgery, Grad-CAM can highlight the inserted object because the model treats it as an additional feature. Additionally, it may indicate some attention towards the surrounding area if the inserted object is attempting to blend into the scene. Looking at Insertion + Blur image forgery CAM has the ability to emphasize the boundary of the inserted region, as well as the areas where blurring was implemented, as they stand out noticeably from the surrounding background. For insertion + Noise operation Grad-CAM will concentrate on the regions where insertion and noise have been applied, accentuating the areas that have been altered using these two methods.

3.2. Feature Fusion for Classification

The motivation behind combining features from different convolutional neural networks (CNNs) such as VGG16 and ResNet50 in deep learning is to take advantage of the strengths of each network. VGG16 is known for its simple and consistent structure, which allows it to capture detailed features effectively. On the other hand, ResNet50 is recognized for its ability to handle the problem of vanishing gradients in deep networks through deep residual learning. By merging features from both networks, we aim to create a more comprehensive representation of the input data. This fusion enables the model to learn intricate details from VGG16 and capture abstract features from ResNet50. By combining these diverse features, the resulting model can potentially achieve higher accuracy and adaptability in a variety of tasks. This is because it benefits from the unique perspectives of both networks, ultimately enhancing the overall performance of the deep learning model.

In the proposed method, feature fusion refers to the fusion of feature vectors of training images extracted from the VGG16 and ResNet50 model. The extracted feature vectors are combined, and a new model is constructed for multiclass classification with a dropout layer, batch normalization, and a SoftMax output layer. This model is trained using the combined features and evaluated on a test dataset, reporting accuracy and a confusion matrix. feature vector extracted from the images using VGG16 and ResNet50 are $F_V = (F_{V1}, F_{V2}, F_{V3,} \dots \dots \dots \dots ., F_{Vn}) \in R^n$ and $F_R = (F_{R1}, F_{R2}, F_{R3,} \dots \dots \dots \dots ., F_{Rn}) \in R^m$, where Rn represents the n-dimensional Feature vector and Rm represents the m-dimensional Feature vector respectively as shown in figure 2. The feature fusion is realized by the concatenation of F_V and F_R, and result is represented by F_f that is an (m + n)-dimensional vector as represented in equation 3. The feature fusion is realized by the following formula.

$$F_f = F_V \oplus F_R = (F_{V1}, F_{V2}, F_{V3,} \dots \dots \dots \dots ., F_{Vn}, F_{R1}, F_{R2}, F_{R3,} \dots \dots \dots \dots ., F_{Rn}), F_f \in R^{n+m} \quad (3)$$

where the elements $(F_{V1}, F_{V2}, F_{V3,} \dots \dots \dots \dots ., F_{Vn})$ of F_V and the elements $(F_{R1}, F_{R2}, F_{R3,} \dots \dots \dots \dots ., F_{Rn})$

of F_R construct a new vector $(F_{V1}, F_{V2}, F_{V3,} \dots \dots \dots \dots ., F_{Vn}, F_{R1}, F_{R2}, F_{R3,} \dots \dots \dots \dots ., F_{Rn})$ m) to express the fused feature vector F_f.

Classification: The classification task can be accomplished by merging the feature vectors of the VGG16 and ResNet50 models. Multiple fully connected layers can be employed for classification. The ReLU function is typically used as the activation function for each neuron in the fully connected layer to enhance network performance. The output layer, which is the last fully connected layer, typically utilizes the SoftMax function as the activation function. The final classification is implemented by the output layer.

The classification layer takes in input F_f and produces an output $C_n = (C_{n1}, C_{n2}, C_{n3,} \ldots \ldots \ldots \ldots, C_{nn},$ which is a C-dimensional feature vector where the dimension is equal to the total number of classes. To improve the model's classification ability, the categorical cross-entropy loss function is utilized in this study.

The overall algorithm is as follows.

Algorithm: **Classification of forgery in handwritten document images using deep features.**

Input: Forged handwritten document images.

Output: Classification of Forgery into ten class classifications.

Step 1: Create an image Data Store for reading images, labeling them based on their respective folder names, and including all subfolders within the directory.

Step 2: Rescale the input images to ensure their size matches that of the input layer in VGG-16.

Step 3: The data is divided into training and test sets, with 70% of images per category used for training and 30% for testing the network in each folder.

Step 4: Employ methods for extracting features using VGG16 and ResNet50 and Compute Grad-CAM.

Step 5: Feature fusion and classification.

Step 6: Analysis of performance metric evaluation to test the dataset.

Step 7: Tabularize the results.

Algorithm ends.

4. Experiments and Results

4.1. Dataset Creation and Evaluation

This study aims to classify different types of forgery that appear in handwritten document images, particularly in a noisy and blurry environment. To accomplish this classification task, we propose a novel method. The effectiveness of any algorithm is determined by how the model is trained. During the training process, the entire dataset is divided into training, testing, and validation sections with a split ratio of 70:20:10.

A literature review reveals that existing datasets only contain images with specific isolated forgery actions such as noise, blur, copy-paste, and insertions. However, they do not include handwritten documents with multiple forgery actions. In order to tackle this issue, we have created our own dataset consisting of 1300 forged handwritten words, divided into 10 different classes. Each class includes 130 forged images for further examination as shown in Figure 7 pie chart. The description of ten altered classes are as follows:

The Normal class (Original Class) includes handwritten document images that are genuine and have not undergone any manipulation. The Blur and Noise class contains the words affected by Gaussian blur and Gaussian Noise respectively. Copy paste class includes the words created copy paste operation and Insertion class contains the words created by insertion operation where target part of the word is erased, and different characters are inserted using insertion operation. The class copy paste + insertion contains the forged words affected by both copy paste and insertion operation. Copy paste + Noise, copy paste + Blur, Insertion + Noise and insertion + Blur class contains the forged words where the half of the word is affected by copy paste and Insertion operation and another half part of the word is affected by noise and blur operations respectively. Figure 8 represents some sample images for multiple forgery operations of ten different classes from our custom created dataset where, the forged text is underlined by different colors. Blue color indicates Blur operation, red color indicates Noise, green color indicates Copy paste and pink color indicates the Insertion operation.

Our dataset includes sample forged images created artificially and images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially from Figure 8. We believe the dataset we have constructed is sufficiently diverse for evaluating our proposed method.

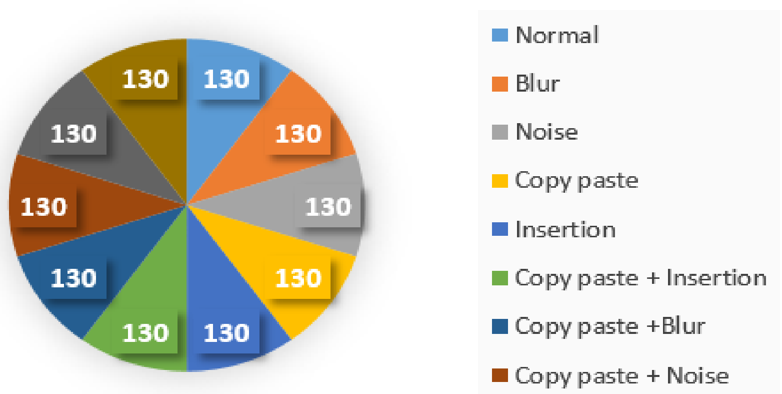


Figure 7. Pie chart representation of forged image dataset classification

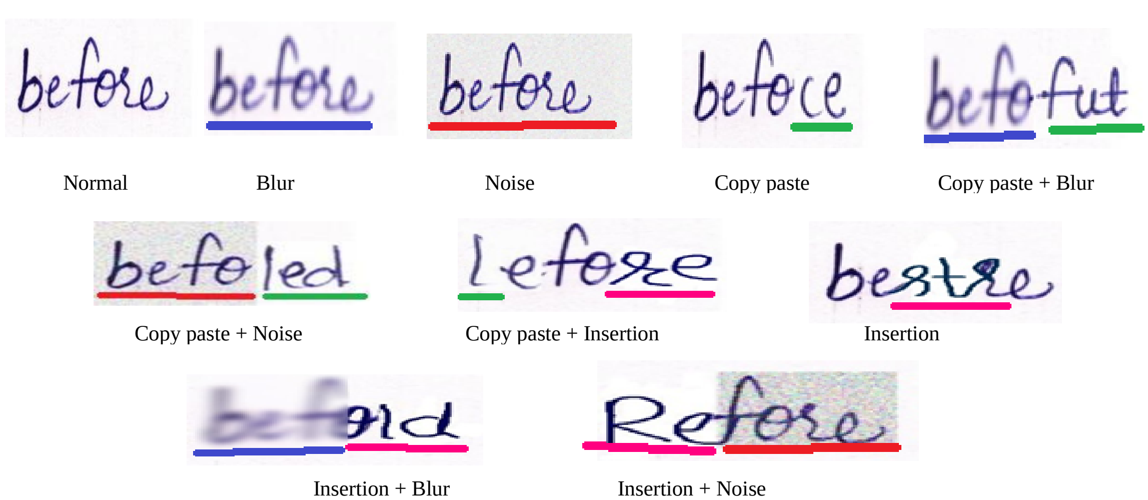
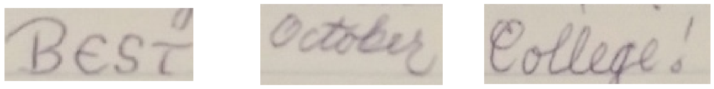
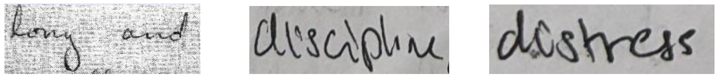


Figure 8. Sample images for multiple forgery operation from our custom created dataset



(a) Images with real blur



(b) Images with real noise

Figure 9. Samples of real noisy and blurred images

Kundu Dataset [2]: The dataset consists of 800 images of handwritten words, with 200 images in each of the four classes: Blur, Noise, Forged, and Original. However, images with multiple forgery operations are not included in the dataset. **IMEI Forged Number dataset [17]:** Forgery images are created using IMEI numbers found on mobile phone cases and packaging. This dataset is more complicated for classifying forged images compared to our own handwritten dataset, Kundu et al., and the ICPR FDC 2018 datasets. Each class consists of 500 samples, resulting in a total of 1000 samples for experimentation. **ICPR 2018 Fraud Detection Dataset (FDC) [34]:** Most of the modified images in this scenario have numerals that have been changed at the character level. These images have a plain background similar to the document images in our custom dataset. However, in the Kundu et al. dataset, the images have both numerals and printed characters, but not handwritten text. This dataset presents a two-class classification challenge as it includes both original and altered prices. The altered samples make up the forgery class, with 301 samples, while the unaltered samples from the original class with 527 samples. Sample images

We employed a variety of data augmentation techniques, including image rotation, scaling, and reflection in the x and y axis, to increase the complexity of the dataset. This resulted in a greater number of images in each dataset, with our custom dataset containing 650 images in each category, totaling 6500 images. In comparison, Kundu et al.'s dataset has 1000 images in each category and 4000 images in total. The ICPR 2018 fraud detection contest dataset has 1505 original images and 2635 forged images, while the Forged IMEI number dataset has 2500 images in each category. Overall, our method used 19,640 images for experimentation, and we compared its effectiveness with existing g techniques.

To demonstrate the efficacy of the proposed method, we implemented five recently developed forgery detection methods. These methods were chosen because they exhibit the same goal as the proposed method. Kundu et al. [2] used spectral density and variations as features to develop a method for detecting forged handwritten text images at the word level. Nandanwar et al. [18] used a phase spectrum derived from DCT and phase statistical features to detect forgeries in IMEI number images and air ticket images. Chen et al. [15] used convolutional neural networks to develop a method for detecting forged numeral handwritten text.

Nandanwar et al. [14] implemented Chebyshev-Harmonic-Fourier-Moments and a Deep Convolutional Neural Network to detect forgery in handwritten words. Patil et.al [1] implemented an approach for altered handwritten text detection using combination of statistical, gradient, and texture features with a Bayesian classifier which classifies 10 class classification problem of forged handwritten document images. Qu et al.

[22] propose the Document Tampering Detector (DTD) framework to identify tampered text in document images.

These six approaches were selected for comparison because they all seek to accomplish the same objective as the suggested strategy. Furthermore, Patil et.al [1], Kundu et al. [2], Nandanwar et al. [14] and Chen et al. [15] concentrate on handwritten text, whereas Nandanwar et al. [18] concentrate on images and text in PDF images and Qu et al. [22] concentrates on printed document images. This comparative study aims to show that methods developed for detecting forged text created by a single operation may not be effective for detecting forged text affected by multiple operations.

In order to evaluate the robustness of the suggested method, standard datasets are utilized for testing purposes. These datasets consist of various examples of handwritten text with different types of forgery, modified prices in receipts, and forged IMEI number images on mobile phone cases.

To assess the effectiveness of both the proposed and existing methods, standard measures are taken into account. These measures include generating the confusion matrix and calculating the average classification rate. The average classification rate is determined by obtaining the mean of the diagonal elements of the confusion matrix, as defined in Equation (4).

$$Accuracy = \frac{correct\ predictions}{total\ no.of\ \ predictions} X100 \tag{4}$$

Implementation Details

The anaconda Navigator is utilized for executing the complete model, which serves as a graphical user inters supporting different implementation platforms (https://docs.anaconda.com/anaconda/navigator/install/). Jupyter Notebook (version 6.4.12) is a platform for executing interactive computing notebooks on the web. Python is used to write the code (https://www.python.org/). The algorithm utilized the import statement to incorporate libraries like TensorFlow, Keras, sklearn, pandas, NumPy, and matplotlib.

4.2. Ablation Study

In this study, a pre-trained deep neural architectural model was used to evaluate proposed work. The experiments were conducted on our custom created dataset of forged text samples from ten classes. The samples were preprocessed and normalized before being divided into training and testing sets. The key steps to achieve the required results in the proposed work involve utilizing the pretrained models VGG16 and ResNet50 along with various classifiers such as VGG16 and ResNet50. As a result, in order to evaluate the efficiency and impact of each of the important stages, we conducted the following experiments using our custom dataset: (i) Using VGG16 deep learning model as a feature extractor with VGG16, Classifier and (ii) Using ResNet50 deep learning model as a feature extractor with ResNet50 Classifier and (iii) Fusion of extracted features from VGG16 and ResNet50 using CNN model SoftMax layer for classification and the results are reported in Table 3**.**

Table 3. Classification rate on custom created dataset (in %)

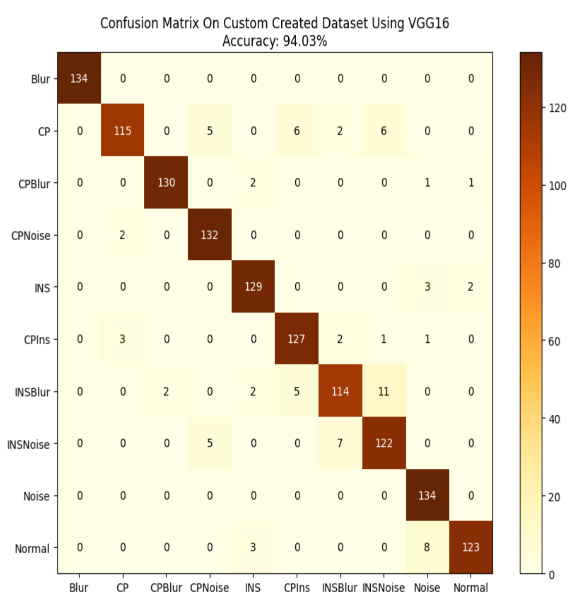
Dataset	VGG16	ResNet50

		Train	Val	Train	Val
Own custom Created Dataset	Loss	0.2986	0.1597	0.3553	0.2107
	Accuracy	0.9041	0.9406	0.8668	0.9001
	Test Accuracy	94.03		91.94	
	Feature Fusion of VGG16 + ResNet50				
		Train		Val	
	Loss	0.2347		0.1165	
	Accuracy	0.9169		0.9599	
	Test Accuracy	0.9694			

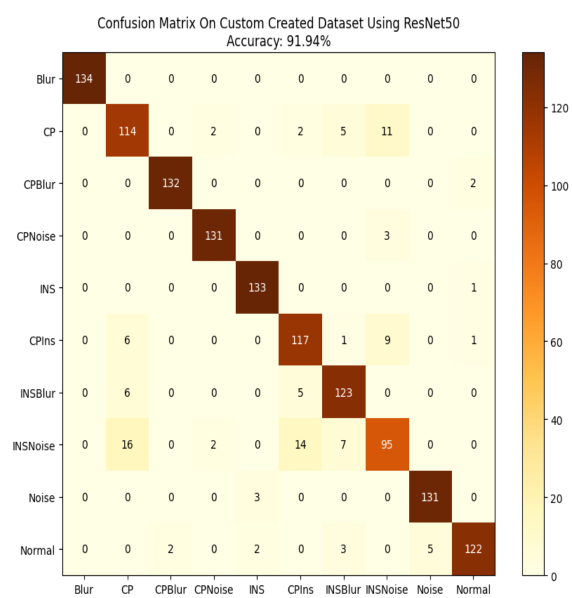
It can be seen, based on the classification rate of individual and fusion of features as shown in Table 3, that every feature scheme contributes to a successful classification outcome. The VGG16 model achieved a better classification rate compared to the ResNet50 model. It can also be observed that the feature fusion of VGG16 and ResNet50 model outperforms well on our dataset compared to the results reported for the individual features extraction and classification schemes.

4.3. Experimental results on our custom created dataset

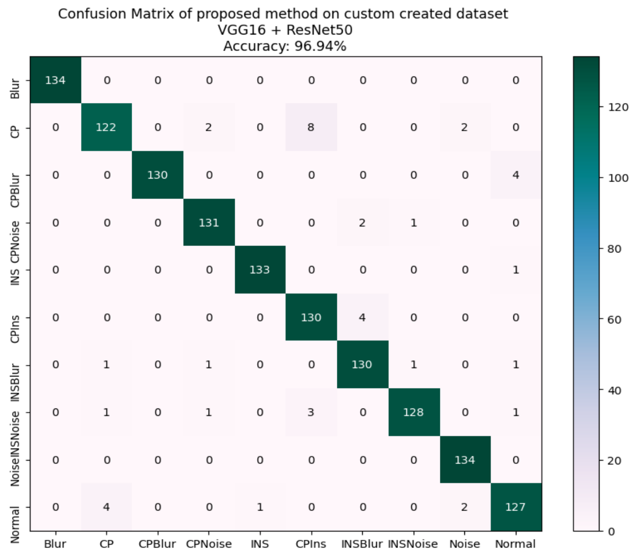
Our first experiment was conducted on our own dataset which consists of 6500 images of the ten different classes described earlier. The dataset was evaluated using two different pre-trained models, ResNet50 and VGG16. These results can be seen in Table 3. Here, for each model the hyper-parameters are fine-tuned using Adam optimizer with learning rate = 0.001, Epochs = 20 and Categorical-Cross Entropy as the loss function. VGG16 achieved 94.03% accuracy while ResNet50 obtained 91.94%. The feature fusion of these two models achieved 96.94%. Figure 10(a) and 10(b). illustrate the confusion matrix of VGG16 and ResNet50, respectively, while 10(c) presents the confusion matrix for the feature fusion of VGG16 and ResNet50. The class abbrevations used in Figure 10(a),(b) and (c) are as follows: Blur-Blur. CP-Copy Paste, CPBlur-Copy Paste+Blur, CPNoise-Copy Paste+Noise, INS-Insertion, INSblur-Insertion+Blur, INSNoise-Insertion+Noise, CPIns-Copy Paste+Insertion, Noise-Noise and Normal-Normal.



(a) Confusion Matrix of VGG16 on custom created dataset



(b) Confusion Matrix of ResNet50 on custom created dataset



(c) Confusion matrix for fusion of two models on custom created dataset

Figure 10. Confusion matrix of proposed method on custom created dataset.

Table 4 displays the average Classification Rate (CR) for both existing methods and our new proposed approach on our custom dataset. The table reveals that our method outperforms the others, obtaining the highest CR. This is attributed to the fact that the existing methods [2, 14, 15, 18, 22] were originally intended for two-class and four-class classification, whereas the proposed method and G. Patil et.al [1] was specifically designed for ten-class classification. Additionally, the images in our dataset have undergone multiple operations, resulting in noise and blur, which the existing methods struggle to handle.

Table 4. Implementation results of existing and proposed method on custom created dataset (in %)

Dataset	Methods	Results
Custom Created dataset	G. Patil et.al [1]	94.89
	Kundu et al[2]	74.41
	Nandanwar et al. [14]	77.83
	Chen et al. [15]	80.25
	Nandanwar et al. [18]	79.57
	Qu et al. [22]	78.67
	Proposed	**96.94**

Based on the data from Table 4, it can be observed that the technique of Patil et.al [1] outperforms Chen et al. [15], Nandanwar et al. [14], Qu, C et al [22], Nandanwar et al. [18], and Kundu et al. [2] among the six existing methods. However, it is not as good as our proposed method in terms of CR. This is because G. Patil et.al [1] employs a combination of statistical, Histogram Oriented Gradients (HOG) and local binary patterns to tackle the difficulties in identifying altered handwritten text that is noisy and blurred, whereas Kundu et al. [2], Chen et al. [15], Nandanwar et al. [14], Nandanwar et al. [18] and Qu et al. [22] use transforms that are less robust to these effects. As demonstrated in this experiment, however, the combination of feature fusion and sequential classifiers proposed in our study is capable of effectively handling various scenarios of altered handwritten text making it superior to the existing methods in terms of CR.

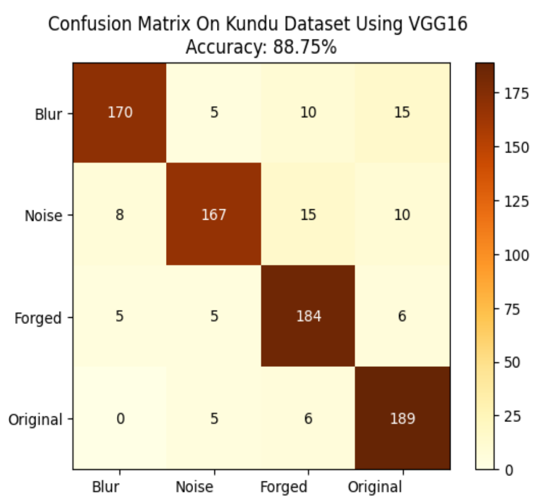
4.4. Experimental Results on Benchmark Datasets

Table 5-Table 7 shows classification rates of the existing and proposed methods on the IMEI [17] dataset, the Kundu [2] dataset and the ICPR FDC [34] datasets, respectively**.** From Table 5, it can be observed that the individual models VGG16 and Resnet50 achieve 88.75% and 85.38% respectively. If we compare the individual models with G. Patil et.al [1], the results are inferior. This is due to the combination of statistical, gradient, and texture features that have the potential to be more effective than VGG16 and ResNet50. Handcrafted features have the ability to capture specific patterns, which makes them suitable for small datasets and tasks that require domain knowledge and interpretability. On the other hand, deep learning models like VGG16 and

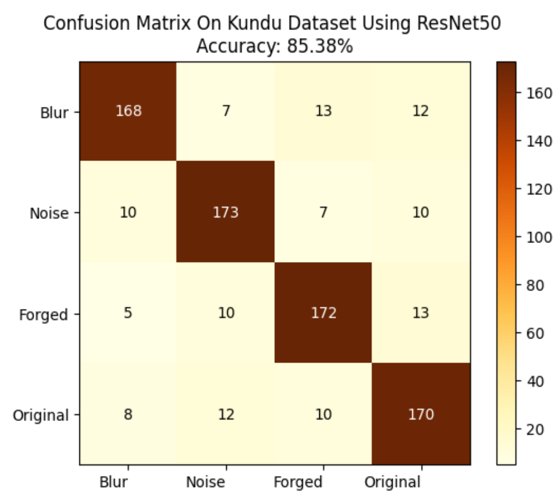
ResNet50 needs large amounts of data and may not perform well in these situations due to overfitting. The same conclusions can be drawn for the IMEI [17] and ICPR FDC [34] datasets. When it comes to the feature fusion of VGG16 and ResNet50, it achieves 90.10% CR, which is superior to Patil et.al [1] and the other existing methods.

Table 5. Implementation results of existing methods on Kundu Dataset (in %)

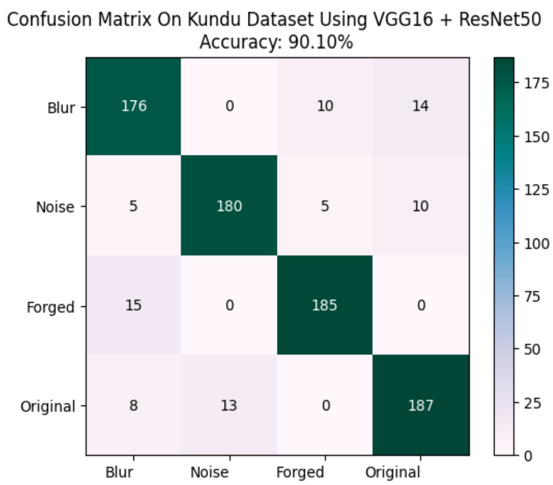
Dataset	Methods	Results
Kundu Dataset	G. Patil et.al [1]	90.02
	Kundu et al[2]	76.15
	Nandanwar et al. [14]	70.07
	Chen et al. [15]	78.00
	Nandanwar et al. [18]	86.46
	Qu et al. [22]	71.01
	VGG16 Model ResNet50 Model **Proposed (VGG16+ResNet50)**	88.75 85.38 **90.10**



(a) Confusion Matrix on Kundu Dataset Using VGG16



(b) Confusion Matrix on Kundu Dataset Using ResNet50



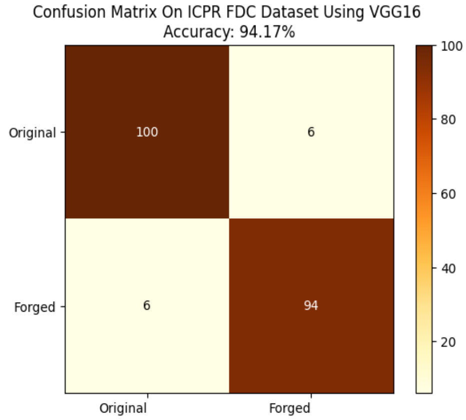
(c) Confusion Matrix of proposed method on Kundu Dataset

Figure 11. Confusion Matrix of individual models and Fused Features on Kundu dataset

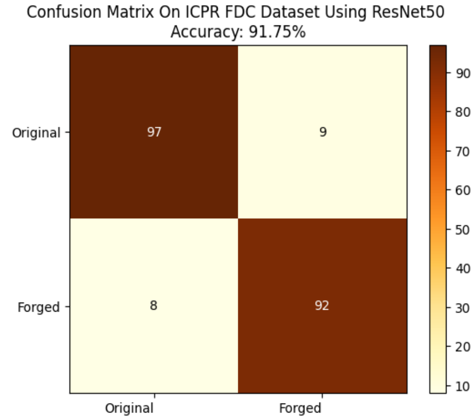
In comparison to the ICPR 2018 FDC and IMEI number datasets, our method performs poorly on the Kundu dataset in terms of classification rate. This is because the Kundu et al. dataset consists of four-class problems, while the other two datasets only have two-class problems. It is evident that complexity rises with the number of classes. Similarly, when comparing the CR on the IMEI number and ICPR 2018 FDC datasets using the suggested method, the CR on the IMEI number dataset is higher than on the ICPR 2018 FDC dataset. This is because the ICPR 2018 FDC number dataset includes images with complex backgrounds. The method suggested by G. Patil et.al [1] performs well compared to the methods of Kundu et al [2], Chen et al. [15], Nandanwar et al. [14], Nandanwar et al. [18] and Qu et al. [22]. The method [1] was developed for addressing the challenges of altered handwritten text detection in noisy and blurred environments. The confusion matrices of the individual VGG16 and ResNet50 models as well as the proposed fusion method on the Kundu dataset, the ICPR FDC dataset, and the IMEI Number dataset are provided in Figure 11(a), (b), (c), Figure12(a), (b), (c)] and Figure 13(a), (b), (c)], respectively.

Table 6. Implementation results of existing methods on ICPR FDC Dataset (in %)

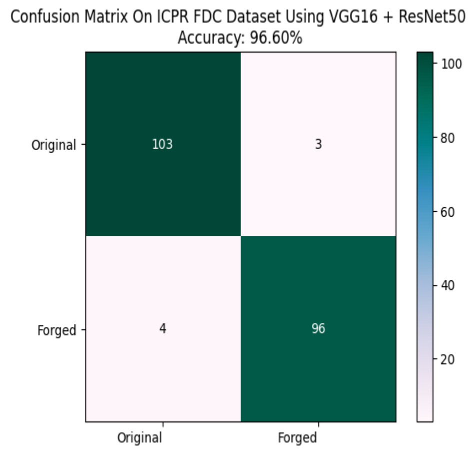
Dataset	Methods	Results
ICPR FDC Dataset	G. Patil et.al [1]	94.35
	Kundu et al[2]	82.1
	Nandanwar et al. [14]	79.5
	Chen et al. [15]	91.8
	Nandanwar et al. [18]	84.4
	Qu et al. [22]	90.27
	VGG16 Model	94.17
	ResNet50 Model	91.75
	Proposed (VGG16+ResNet50)	**96.60**



(a). Confusion Matrix on ICPR FDC Dataset Using VGG16



(b). Confusion Matrix on ICPR FDC Dataset Using ResNet50



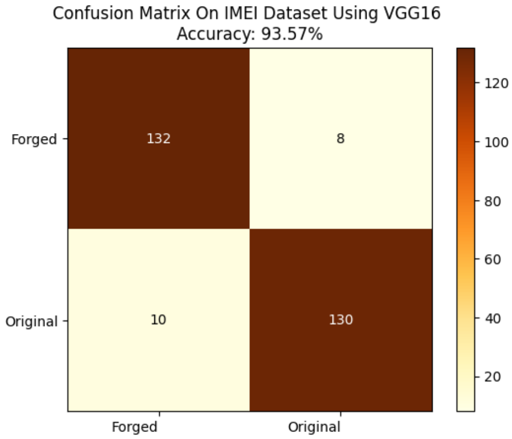
(c). Confusion Matrix of proposed method on ICPR FDC Dataset

Figure 12. Confusion Matrix of individual models and Fused Features on ICPR FDC dataset

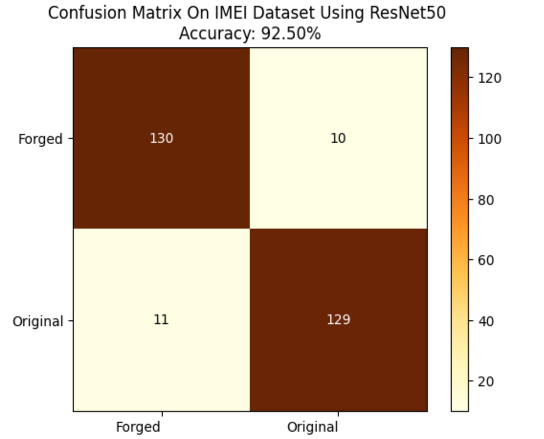
From Table 6, it can be noted that the classification rates of VGG16 and Resnet50 on the ICPR FDC dataset are 94.17% and 91.75%, respectively. Also, from Table 7 it can be observed that the classification rates of VGG16 and ResNet50 on the IMEI Number dataset are 93.57% and 92.50%, respectively. The Classification Rate of the Kundu et al. method falls from 82.1% to 67.56% according to Tables 6 and 7. Similarly, the CR of the Nandanwar et al. [18] technique decreases from 84.4% to 75.87%, and the Classification Rate of the Chen et al. [15] approach decreases from 91.8% to 78.09%. The classification rate of Qu et al [22] is unchanged. Finally, the CR of Patil et.al [1] increases from 94.35% to 99.90%. The reason is that the method [1] is robust to altered handwritten text affected by multiple forgery operations.

Table 7. Implementation results of existing methods on IMEI Dataset (in %)

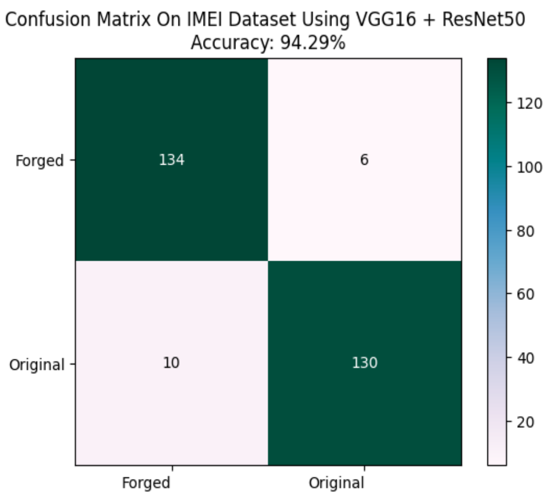
Dataset	Methods	Results
IMEI Dataset	Patil et.al [1]	99.90
	Kundu et al[2]	67.56
	Nandanwar et al. [14]	82.45
	Chen et al. [15]	78.09
	Nandanwar et al. [18]	75.87
	Qu et al. [22]	89.80
	VGG16 Model	93.57
	ResNet50 Model	92.50
	Proposed (VGG16+ResNet50)	**94.29**



(a) Confusion Matrix on IMEI Number Dataset Using VGG16



(b) Confusion Matrix on IMEI Number Dataset Using ResNet50



(c) Confusion Matrix of proposed method on IMEI Number Dataset

Figure 13. Confusion Matrix of individual models and Fused Features on IMEI Number dataset

Table 6 and 7 show that the proposed method achieves the best classification rate on the ICPR FDC and IMEI number datasets. The method yields 96.60% on the ICPR FDC dataset and 94.29% on the IMEI Number dataset. The reason for this is that the proposed dataset contains 10 classes, including noisy and blurred images, whereas the other two datasets only offer two classes—original and forged text. Furthermore, it is observed from Table 5-Table 7 that the proposed method is the best compared to all the existing methods on our and ICPR FDC datasets. However, for the IMEI dataset, the method [1] achieves the best results compared to the proposed method. The method [1] does not explore deep learning models while the proposed method explores deep learning models. Sometimes, due to the small number of samples for training, there are chances of causing overfitting which may lead to misclassification in the case of our method. In addition, IMEI is not complex for classification compared to the proposed dataset because the IMEI dataset has only two classes while our dataset has 10 classes. Since the proposed method is developed for complex classification problems, sometimes for simple classification problem, the deep learning based method may not work well.

The method we have presented for detecting forgeries in different handwritten images is generally effective but has some limitations. Despite being designed specifically to handle multiple alterations; images are sometimes misclassified as shown in Figure 14. This is because these manipulations can cause the images to lose important information, leading to misclassifications. The misclassification of handwritten forged images often occurs due to the intricate variability in individual writing styles, as well as challenges in extracting consistent features for accurate pattern recognition. Moreover, the presence of noise, distortions, and subtle nuances in the falsified content adds complexity to the classification process. Copy-paste operations with noise may be misidentified as just noise, while insertions with noise might be erroneously labeled as just noise. Copy-paste with blur may be classified as blur, and the presence of noise in a copy-paste context could lead to misclassification as noise. Such misclassifications highlight the complexity of accurately categorizing altered handwritten text images. These issues indicate that there is room for improvement.

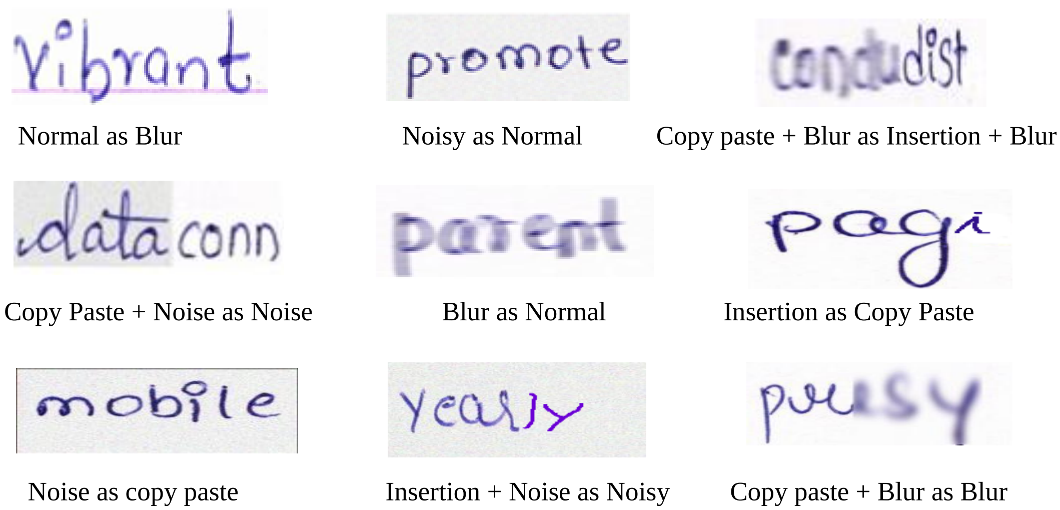


Figure 14. Misclassification samples from the proposed method

5. Conclusion and Future Work

This study examines the application of deep neural networks (specifically transfer learning) to address the problem of classifying multiple altered handwritten text that are distorted by noise and blur. The use of deep learning models, namely, the feature fusion of VGG16 and ResNet50, are employed by utilizing feature learning to identify similarities between images of forged and genuine handwriting. We presented experiments on a custom dataset we developed ourselves as well several standard datasets. Based on these experimental results, our method shows superior performance compared to existing methods in all cases. Nonetheless, the method still makes some errors on challenging input as discussed in experimental section. To tackle this problem, our plan is to explore natural language processing concepts to detect altered text affected by severe distortion and writing style.

References

1. Patil, G., Shivakumara, P., Gornale, S. S., Pal, U., & Blumenstein, M. (2022). A new robust approach for altered handwritten text detection. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-022-14242-8

2. Kundu S, Shivakumara P, Grouver A, Pal U, Lu T, Blumenstein M (2019) A new forged handwriting detection method based on Fourier spectral density and variation. In: Proc. Asian Conference on Pattern Recognition (ACPR), pp 136–150. https://doi.org/10.1007/978-3-030-41404-7_10.

3. Wu, Y., Kong, Q., Yong, L., Narducci, F., & Wan, S. (2023). CDText: Scene text detector based on context-aware deformable transformer. Pattern Recognition Letters, 172, 8-14.

4. Wu, Y., Zhang, W., & Wan, S. (2022). CE-text: A context-Aware and embedded text detector in natural scene images. Pattern Recognition Letters, 159, 77-83

5. Khan RA, Lone SA (2020) A comprehensive study of document security system, open issues and challenges. Multimed Tools Appl 80:7039–7061. https://doi.org/10.1007/s11042-020-10061-x

6. Mehrjardi, F. Z., Latif, A. M., & Zarchi, M. S. (2023). Copy-move forgery detection and localization using deep-learning. International Journal of Pattern Recognition and Artificial Intelligence.

7. Gornale, S. S., Patil, G., & Benne, R. (2022). Document Image Forgery Detection Using RGB Color Channel. Transactions on Engineering and Computing Sciences, 10(5), 1–14. https://doi.org/10.14738/tmlai.105.13126

8. Cha, S. H., Tappert, C. C., Gibbons, M., & Chee, Y. M. (2004). Automatic detection of handwriting forgery using a fractal number estimate of wrinkliness. International Journal of Pattern Recognition and Artificial Intelligence, 18(07), 1361-1371.

9. K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.

10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vol. 2016-December, pp. 770– 778). IEEE Computer Society. https://doi.org/10.1109/CVPR.2016.90.

11. Lin, Y., Lu., Y., Li, Z., Liu, Z and Liu, W. (2023), “Fusion and inpainting: A novel salient object detection network”, In Proc. ICIVC, pp 189-193. https://doi.org/10.1109/ICIVC58118.2023.10270136.

12. Megahed, A., Fadl, S. M., Han, Q., & Li, Q. (2018). Handwriting forgery detection based on ink colour features. In Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS (Vol. 2017-November, pp. 141–144). IEEE Computer Society. https://doi.org/10.1109/ICSESS.2017.8342883

13. Roy, P., & Bag, S. (2019). Detection of Handwritten Document Forgery by Analyzing Writers’ Handwritings. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 11941 LNCS, pp. 596–605). Springer. https://doi.org/10.1007/978-3-030-34869-4_65

14. Nandanwar, L., Shivakumara, P., Kundu, S., Pal, U., Lu, T., & Lopresti, D. (2020). Chebyshev-Harmonic-Fourier-Moments and deep CNNs for detecting Forged handwriting. In Proceedings - International Conference on Pattern Recognition (pp. 6562–6569). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICPR48806.2021.9412179

15. Chen, Y., & Gao, S. (2020). Forgery Numeral Handwriting Detection based on Convolutional Neural Network. In Proceedings of 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference, ITOEC 2020 (pp. 201–205). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ITOEC49072.2020.9141830

16. Roy, P., & Bag, S. (2019). Forensic Performance on Handwriting to Identify Forgery Owing to Word Alteration. In ISBA 2019 - 5th IEEE International Conference on Identity, Security and Behavior Analysis. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ISBA.2019.8778490

17. Shivakumara P, Basavaraja V, Gowda HS, Guru DS, Pal U, Lu T (2018) A new RGB based fusion for forged IMEI number detection in Mobile images. 2018 16th international conference on Frontiers in handwriting recognition (ICFHR), pp 386–391. https://doi.org/10.1109/icfhr-2018.2018.00074

18. Nandanwar L, Shivakumara P, Kanchan S, Basavaraja V, Guru DS, Pal U, Blumenstein M (2020) DCTphase statistics for forged IMEI numbers and air ticket detection. Expert Syst Appl 164:114014. https://doi. org/10.1016/j.eswa.2020.114014

19. Humayun, M. F., Malik, H. W., & Alvi, A. A. (2022). A Simplified Un-Supervised Learning Based Approach for Ink Mismat S. Elkasrawi and F. Shafait, "Printer Identification Using Supervised Learning for Document Forgery Detection," 2014 11th IAPR International Workshop on Document Analysis Systems, Tours, France, 2014, pp. 146-150, doi: 10.1109/DAS.2014.48.

20. Jaiswal, G., Sharma, A., & Yadav, S. K. (2022). Deep feature extraction for document forgery detection with convolutional autoencoders. Computers and Electrical Engineering, 99, 107770.

21. Roy, Priyanka and Bag, Soumen, Investigation of Forgery in Handwritten Legal Documents Based on Writing Style Analysis. Available at

SSRN: https://ssrn.com/abstract=4485590 or http://dx.doi.org/10.2139/ssrn.4485590

22. Qu, C., Liu, C., Liu, Y., Chen, X., Peng, D., Guo, F., & Jin, L. (2023). Towards Robust Tampered Text Detection in Document Image: New Dataset and New Solution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5937-5946).
23. Fadl, S., Hosny, K. M., & Hammad, M. (2023). Automatic fake document identification and localization using DE-Net and color-based features of foreign inks. Journal of Visual Communication and Image Representation, 92, 103801.
24. Cha, S. H., Tappert, C. C., Gibbons, M., & Chee, Y. M. (2004). Automatic detection of handwriting forgery using a fractal number estimate of wrinkliness. International Journal of Pattern Recognition and Artificial Intelligence, 18(07), 1361-1371
25. CHEN, S. Y., MA, H., & CHEN, Q. J. (2017). Tamper Detection of Batch Websites Based on Text Comparison. In Computer Science and Technology: Proceedings of the International Conference (CST2016) (pp. 573-579).
26. Fahn, C.-S., Lee, C.-P., & Chen, H.-I. (2016). A text independent handwriting forgery detection system based on branchlet features and Gaussian mixture models. 2016 14th Annual Conference on Privacy, Security and Trust (PST). doi:10.1109/pst.2016.7906952
27. Bhardwaj, D., & Pankajakshan, V. (2016). Image Overlay Text Detection Based on JPEG Truncation Error Analysis. IEEE Signal Processing Letters, 23(8), 1027–1031. https://doi.org/10.1109/LSP.2016.2581311
28. S. Elkasrawi and F. Shafait, "Printer Identification Using Supervised Learning for Document Forgery Detection," 2014 11th IAPR International Workshop on Document Analysis Systems, Tours, France, 2014, pp. 146-150, doi: 10.1109/DAS.2014.48.
29. Wu, Y., Zhang, L., Gu, Z., Lu, H., & Wan, S. (2023). Edge-AI-driven framework with efficient mobile network design for facial expression recognition. ACM Transactions on Embedded Computing Systems, 22(3), 1-17
30. Kaur, T., & Gandhi, T. K. (2019). Automated brain image classification based on VGG-16 and transfer learning. In Proceedings - 2019 International Conference on Information Technology, ICIT 2019 (pp. 94–98). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICIT48102.2019.00023.
31. Roy, S., Shivakumara, P., Pal, U., Lu, T., & Tan, C. L. (2016). New tampered features for scene and caption text classification in video frame. In Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR (Vol. 0, pp. 36–41). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICFHR.2016.0020
32. S. Poppi, M. Cornia, L. Baraldi and R. Cucchiara, "Revisiting The Evaluation of Class Activation Mapping for Explainability: A Novel Metric and Experimental Analysis," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 2299-2304, doi: 10.1109/CVPRW53098.2021.00260.
33. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
34. Artaud C, Sidère N, Doucet A, Ogier J, Yooz VPD (2018) Find it! Fraud detection contest report. 2018 24th International Conference on Pattern Recognition (ICPR), pp 13–18. https://doi.org/10.1109/ICPR.2018.8545428

Point to Point Responses to Reviewer’s Comments

Title: Altered Handwritten Text Detection in Document Images Using Deep Learning

Paper ID: No. IJPRAI-D-23-01569

Reviewer#1

Comment 1.1: Authors have established a dependable technique for identifying altered text in handwritten document images, even in scenarios with high levels of noise and blur. I think this paper is quite interesting with algorithms improvement. Experiments have proved its effectiveness. However, I have several suggestions for authors to improve.

Response 1.1: Thank you very much for your encouraging feedback. We have considered your comments and suggestions and modified the manuscript to improve the quality and clarify of the paper. The changes we have made to the Related Work, the Proposed Method, and the Experimental Results sections are marked in yellow.

Comment 1.2: 1. English writing could be improved. 2.I have several citation suggestions for you.

https://doi.org/10.1016/j.patrec.2023.05.025, https://doi.org/10.1145/3587038, https://doi.org/10.1016/j.patrec.2022.05.004

Response 1.2: Thank you very much for your suggestions. The revised manuscript has been reviewed and edited by a native English speaker to improve the quality of the writing. We have added the citations you suggested as [3, 4 and 27] in the revised manuscript.

Reviewer#2

Comment 2.1: The problem addressed in this paper is very interesting and appropriate in the context of handwriting recognition. Here are few minor suggestions:

Response 2.1: Thank you for your comments and suggestions. We have incorporated these in the revised manuscript. The changes we have made are marked in yellow.

Comment 2.2: The reason of choosing VGG19 and ResNet as a feature extractor need to be mentioned.

Response 2.2: Thank you. We have the added more motivation in the Introduction Section in the revised manuscript. In our study, the selection of VGG16 and ResNet50 as feature extractors for altered handwritten document images in blurry and noisy environments is motivated by their unique capabilities in addressing the challenges presented by such situations. VGG16's simple yet efficient design, utilizing small convolutional filters, is well-suited for capturing intricate spatial features. This enables it to effectively identify subtle patterns in handwritten content, even when obscured by blur or noise. On the other hand, ResNet50's innovative residual learning framework excels in handling noisy environments by mitigating the vanishing gradient problem, ensuring that the model can proficiently learn and extract meaningful features from handwritten text amidst the presence of noise. By integrating the architectural features of VGG16 and ResNet50, we leverage the synergistic advantages to establish a feature extraction framework that effectively tackles the challenges posed by blur and noise in altered handwritten documents. Consequently, this approach significantly improves the performance of our model in demanding real-world scenarios.

Comment 2.3: Why other popular networks are not tried?

Response 2.3: Thank you for raising this point. We agree that there are other popular models for feature extraction and classification in the literature. For example, ImageNet is powerful method for extracting visual features and hence is well suited for object classification and recognition in contrast to the combination of VGG and ResNet proposed in this work for altered handwritten text classification [11]. We included this observation in motivating our approach in the revised manuscript. We believe it would be interesting future research to perform experiments combining ImageNet and ResNet for feature extraction and classification.

Comment 2.4: Please improve the figure 2 for better readability.

Response 2.4: We apologize for the poor quality. We re-drew the figure so that the text in the architecture is now clear and readable, as shown in Figure 2 in the revised manuscript.

Comment 2.5: Show some error samples in the discussion and also mention about the reason for such error.

Response 2.5: Thank you for this suggestion. We have added more samples of failure cases in Figure 14. In addition, we have provided possible reasons for such misclassifications and discuss future work to address these challenges in Section 4.4 in the revised manuscript.

Reviewer#3

Comment 3.1: This paper has to be revised and improved significantly as follows:

Response 3.1: We have carefully considered all of the comments and suggestions made by the reviewers to improve the quality and clarity of the revised manuscript. These can be seen throughout the introduction, related work, proposed methodology, and experimental results sections in the paper.

Comment 3.2: The novelty or fresh new ideas is not very clear, does not have enough comparisons with other methods in the literature.

Response 3.2: Thank you for this feedback. Detecting altered handwritten text in noisy and blurred environments is a complex and challenging problem compared to the normal case of forgery detection in document images. We believe that exploring the combination and optimization of existing models for addressing this problem is a useful contribution to the field. In our paper, we adapt VGG16 and the ResNet50 rather than using the baseline models as shown in Figure 4 and Figure 6. Furthermore, the way we fuse the two adapted models is new as shown in Figure 2. To show that the proposed work improves on state-of-the-art methods, we implemented and tested a number of existing methods [1, 2, 14, 15, 18 and 22]. We also added some additional references to recent work [24, 25, 26] which are cited in the Related Work section in the revised manuscript. Our motivation and a list of key contributions are presented in the Introduction section in the revised manuscript. An existing method [22] has been added as a comparative study in the Experimental Section.

Comment 3.3: What is its relation to PR & AI? (Pattern Recognition & Artificial Intelligence) or related area(s)? Should explain it more clearly, i.e., What exactly is it talking about? AI? PR? or Image Processing? Computer Vision? or? A good, complete and thorough bibliography is part of a well written paper.

Response 3.3: We have attempted to clarify these points in the revised manuscript. We note that when forgery operations are performed on images, they introduce distortions, altering the content. This results in unique patterns. We note that when forgery operations are performed on such images, it introduces distortions, altering the content. This results in unique patterns that are detectable. To extract such patterns, the method we present explores the combination of deep learning models to extract these unique patterns. Overall, our approach combines techniques from pattern recognition, image processing, computer vision, and artificial intelligence.

The above explanations is added to the Introduction section in the revised manuscript.

Comment 3.4: Need to add more recent related publications in the reference list including, for examples 2015, 2016, 2017PR & AI subfield,

Response 3.4: Thank you for this suggestion. Yes, we cited the existing methods [22, 24, 25, 26] in the Related Work section and the same were updated in the Reference section in the revised manuscript. The newly cited methods are marked in yellow.

Comment 3.5: Compare your work with others in the literature, advantages vs disadvantages, in depth and width,

Response 3.5: Thank you for raising this point. As discussed in our response to Comment 3.4, existing methods are discussed in the paper and compared and contrasted to our proposed approach. This critical analysis includes discussing advantages and disadvantages of the existing methods. Overall, the related work section has been updated in the revised manuscript to reflect this.

Comment 3.6: Should use real images with noisy and testing samples set size should be large enough to show your results are indeed convincing, reliable, and meaningful.

Response 3.6: Thank you – we agree it would be helpful to include more samples of images and have done so. Our dataset includes sample forged images created artificially as well as images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially in Figure 8. Our dataset includes sample forged images created artificially and images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially from Figure 8. We believe the dataset we have constructed is sufficiently diverse for evaluating our proposed method. We have made suitable changes to Section 4.1 in the revised manuscript.

Point to Point Responses to Reviewer's Comments

Title: Altered Handwritten Text Detection in Document Images Using Deep Learning

Paper ID: No. IJPRAI-D-23-01569

Reviewer#1

Comment 1.1: Authors have established a dependable technique for identifying altered text in handwritten document images, even in scenarios with high levels of noise and blur. I think this paper is quite interesting with algorithms improvement. Experiments have proved its effectiveness. However, I have several suggestions for authors to improve.

Response 1.1: Thank you very much for your encouraging feedback. We have considered your comments and suggestions and modified the manuscript to improve the quality and clarify of the paper. The changes we have made to the Related Work, the Proposed Method, and the Experimental Results sections are marked in yellow.

Comment 1.2: 1. English writing could be improved. 2.I have several citation suggestions for you.

https://doi.org/10.1016/j.patrec.2023.05.025, https://doi.org/10.1145/3587038, https://doi.org/10.1016/j.patrec.2022.05.004

Response 1.2: Thank you very much for your suggestions. The revised manuscript has been reviewed and edited by a native English speaker to improve the quality of the writing. We have added the citations you suggested as [3, 4 and 27] in the revised manuscript.

Reviewer#2

Comment 2.1: The problem addressed in this paper is very interesting and appropriate in the context of handwriting recognition. Here are few minor suggestions:

Response 2.1: Thank you for your comments and suggestions. We have incorporated these in the revised manuscript. The changes we have made are marked in yellow.

Comment 2.2: The reason of choosing VGG19 and ResNet as a feature extractor need to be mentioned.

Response 2.2: Thank you. We have the added more motivation in the Introduction Section in the revised manuscript. In our study, the selection of VGG16 and ResNet50 as feature extractors for altered handwritten document images in blurry and noisy environments is motivated by their unique capabilities in addressing the challenges presented by such situations. VGG16's simple yet efficient design, utilizing small convolutional filters, is well-suited for capturing intricate spatial features. This enables it to effectively identify subtle patterns in handwritten content, even when obscured by blur or noise. On the other hand, ResNet50's innovative residual learning framework excels in handling noisy environments by mitigating the vanishing gradient problem, ensuring that the model can proficiently learn and extract meaningful features from handwritten text amidst the presence of noise. By integrating the architectural features of VGG16 and ResNet50, we leverage the synergistic advantages to establish a feature extraction framework that effectively tackles the challenges posed by blur and noise in altered handwritten documents. Consequently, this approach significantly improves the performance of our model in demanding real-world scenarios.

Comment 2.3: Why other popular networks are not tried?

Response 2.3: Thank you for raising this point. We agree that there are other popular models for feature extraction and classification in the literature. For example, ImageNet is powerful method for extracting visual features and hence is well suited for object classification and recognition in contrast to the combination of VGG and ResNet proposed in this work for altered handwritten text classification [11]. We included this observation in motivating our approach in the revised manuscript. We believe it would be interesting future research to perform experiments combining ImageNet and ResNet for feature extraction and classification.

Comment 2.4: Please improve the figure 2 for better readability.

Response 2.4: We apologize for the poor quality. We re-drew the figure so that the text in the architecture is now clear and readable, as shown in Figure 2 in the revised manuscript.

Comment 2.5: Show some error samples in the discussion and also mention about the reason for such error.

Response 2.5: Thank you for this suggestion. We have added more samples of failure cases in Figure 14. In addition, we have provided possible reasons for such misclassifications and discuss future work to address these challenges in Section 4.4 in the revised manuscript.

Reviewer#3

Comment 3.1: This paper has to be revised and improved significantly as follows:

Response 3.1: We have carefully considered all of the comments and suggestions made by the reviewers to improve the quality and clarity of the revised manuscript. These can be seen throughout the introduction, related work, proposed methodology, and experimental results sections in the paper.

Comment 3.2: The novelty or fresh new ideas is not very clear, does not have enough comparisons with other methods in the literature.

Response 3.2: Thank you for this feedback. Detecting altered handwritten text in noisy and blurred environments is a complex and challenging problem compared to the normal case of forgery detection in document images. We believe that exploring the combination and optimization of existing models for addressing this problem is a useful contribution to the field. In our paper, we adapt VGG16 and the ResNet50 rather than using the baseline models as shown in Figure 4 and Figure 6. Furthermore, the way we fuse the two adapted models is new as shown in Figure 2. To show that the proposed work improves on state-of-the-art methods, we implemented and tested a number of existing methods [1, 2, 14, 15, 18 and 22]. We also added some additional references to recent work [24, 25, 26] which are cited in the Related Work section in the revised manuscript. Our motivation and a list of key contributions are presented in the Introduction section in the revised manuscript. An existing method [22] has been added as a comparative study in the Experimental Section.

Comment 3.3: What is its relation to PR & AI? (Pattern Recognition & Artificial Intelligence) or related area(s)? Should explain it more clearly, i.e., What exactly is it talking about? AI? PR? or Image Processing? Computer Vision? or? A good, complete and thorough bibliography is part of a well written paper.

Response 3.3: We have attempted to clarify these points in the revised manuscript. We note that when forgery operations are performed on images, they introduce distortions, altering the content. This results in unique patterns. We note that when forgery operations are performed on such images, it introduces distortions, altering the content. This results in unique patterns that are detectable. To extract such patterns, the method we present explores the combination of deep learning models to extract these unique patterns. Overall, our approach combines techniques from pattern recognition, image processing, computer vision, and artificial intelligence.

The above explanations is added to the Introduction section in the revised manuscript.

Comment 3.4: Need to add more recent related publications in the reference list including, for examples 2015, 2016, 2017PR & AI subfield,

Response 3.4: Thank you for this suggestion. Yes, we cited the existing methods [22, 24, 25, 26] in the Related Work section and the same were updated in the Reference section in the revised manuscript. The newly cited methods are marked in yellow.

Comment 3.5: Compare your work with others in the literature, advantages vs disadvantages, in depth and width,

Response 3.5: Thank you for raising this point. As discussed in our response to Comment 3.4, existing methods are discussed in the paper and compared and contrasted to our proposed approach. This critical analysis includes discussing advantages and disadvantages of the existing methods. Overall, the related work section has been updated in the revised manuscript to reflect this.

Comment 3.6: Should use real images with noisy and testing samples set size should be large enough to show your results are indeed convincing, reliable, and meaningful.

Response 3.6: Thank you – we agree it would be helpful to include more samples of images and have done so. Our dataset includes sample forged images created artificially as well as images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially in Figure 8. Our dataset includes sample forged images created artificially and images with real noise and blur. Sample images with real noise and blur are shown in Figure 9, where one can see the images look like the images created artificially from Figure 8. We believe the dataset we have constructed is sufficiently diverse for evaluating our proposed method. We have made suitable changes to Section 4.1 in the revised manuscript.