Deepfake detection and classification using local surface geometrical features

1st Sivabalamurugan M
Dept. of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidhyapeetham
India

2nd Swapna T.R
Dept. of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidhyapeetham
India

***Abstract*—The rise of Deepfake technology presents a critical challenge to the integrity of digital media, prompting the need for advanced forgery detection techniques. This paper proposes a novel approach for detecting deepfake image forgery by integrating local surface geometry analysis into the detection process. Leveraging the inherent alterations in local surface geometry caused by deepfake generation, our method extracts and merges this information with the original image, enhancing the discriminative features essential for accurate identification of manipulated content. Using EfficientNet-B0 as the classifier, our method achieves a remarkable accuracy rate of 98% in correctly identifying deepfake images. This high accuracy underscores the effectiveness of our approach in detecting sophisticated deepfake content. By combining local surface geometry analysis with deep learning-based classification, our method represents a significant advancement in the field of deepfake detection, offering a promising solution to combat the proliferation of manipulated media and preserve the authenticity of digital imagery.**

***Index Terms*—EfficientNet, local surface geometry, deepfake, UprightNet**

I. Introduction

The rise of deepfake technology has ushered in an era where the authenticity of visual content is increasingly called into question . Deepfake algorithms, empowered by machine learning techniques, can seamlessly alter images and videos, often with disturbingly convincing results as said in the review. As a consequence, the potential for misinformation, identity theft, and privacy breaches has become a pressing concern for society. Traditional methods for detecting forged images primarily rely on pixel-level analysis, which can be easily circumvented by sophisticated manipulation techniques.Figure-1 illustrate the example of an deepfake with the both original and real image where the expression of a person is changed artificially

To address these challenges, this paper proposes a novel approach that leverages the local surface geometry of images for deepfake detection. By focusing on geometric features, which encapsulate intrinsic properties of the image structure, our method aims to create a more robust and tamper-resistant detection framework. Specifically, we explore the utilization of advanced geometric analysis techniques to extract discriminative features from different regions of the image. These features are then utilized to train a deep neural network classifier, leveraging the efficiency and effectiveness of the state-of-the-art EfficientNet architecture.

In this study, we present our methodology in detail, highlighting its key components and theoretical foundations. Furthermore, we conduct comprehensive experiments to evaluate the performance of our approach on diverse datasets containing both real and deepfake images. Through our experiments, we aim to demonstrate the efficacy and reliability of our method in accurately distinguishing between genuine and manipulated visual content. Overall, this research contributes to the ongoing efforts to develop robust solutions for combating the proliferation of deepfake content and safeguarding the integrity of digital media platforms.

Fig. 1. Here the left side image is original and the right side image is fake

II. Related Works

Deepfake detection is a recent problem where the people cannot identify whether the image is fake or not using their eyes .Mainly the manipulations are focused on face tampering and this are threating the people like said in [1] even the AFIS cannot find the manipulated human faces.Deeraj et al. [2] developed a method to extract the image features in pixel level using ELA (error level analysis) and giving the image to a Dense CNN for further feature extraction and classification. But the technique worked for only lossy images . Some authors where tried to find the surface level descriptors [1], [3] to extract the important feature from the image like texture, camera orientation, illumination.Andrea et al [3] developed a

novel technique whether they used a method called surface which is pipeline which is used to generate a image with import feature It made using global surface descriptor which is generated using uprightnet [4],and they are giving the surface images to the available pretrained models such as ResNet50, MobileNetV2, EfficientNet-B0.Some other local descriptors where also introduced by authors like Arnab et al [5] where he introduced the weber local descriptor which was introduced by chen et al [6] to find the local feature using intensity of the pixels and they where given to CNN models for further feature extraction and classification. Zahid et al [1] used local image descriptors which is a pixel level analysis to find the pixels which are different from their neighbour .some authors like Atmik et al [7] employed only the CNN model to detect the forgery in image but this were giving only 80 percent accuracy. so the deepl earning models where giving good results then the other methods like machine learning method and statistical method said by RANA et al [8] in their review paperVedant et al [9] developed a layered approach where they followed normal CNN for feature detection then passing the feature to the LSTM layer, finally they used softmax function for classification. Rineesh et al [10] introduced a multipath convolutional neural network (CNN) with three modules is used, each of which is stacked with a convolutional block attention mechanism. The first two modules in the dual-path paradigm are a Resnet module and a Densenet module. The Resnet component enables for feature reuse while Densenet allows for the investigation of new features. The parallel Inception Resnet module contains a one-dimensional feature reduction module with residual connections. Vidhyasagar et al [12] where using a YOLO weights and ResNet50v2 architecture to find the authenticity of the image. And authors like Bhuvaneswari et al [13] where using SqueezeNet to find the image forgery and also some authors like Kalpana et al [14] also using filters in the preprocessing steps to highlight the forged place in the image and training it using the CNN model. The literature survey identified a gap that though Global surface descriptors gave an accuracy of 75% in identifying forgery , the local surface geometry were not used for deepfake detection. But we can use local surface geometry to get the features of the image and this can be used for the image forgery detection. Based on the survey , we have identified the following objectives.

(i) Converting the image into a local surface geometry
(ii) Combining the local surface geometry and the original image
(iii) Classifying the image with the EfficientNet model

A. Dataset

For our study, we utilize a dataset sourced from Kaggle [11], comprisin 75000 images each of both real and fake images, in that 60000 images are used for training purposes. In that 15000 images each where used for validation purposes. The dataset consists of a diverse range of visual content, encompassing various images of faces of the persons, subjects, and backgrounds, thereby ensuring its representativeness and suitability for training deep learning models. Each image is labeled as either ”real” or ”fake,” indicating its authenticity status

III. Proposed Methodology

Our methodology is structured into three key steps, each strategically designed to leverage specific aspects of the image data to enhance the accuracy of detection. They are performed in the form of a pipeline first the image is converted into local surface geometry then it is combined with the original image and the it Is trained to do the classification using EfficientNet.In Fig-2 we could see our architectural diagram of the methodology

A. Local surface geometry

The features of surfaces within a localized area or proximity are referred to as local surface geometry [4]; these surfaces are usually examined at individual points or pixels. It entails comprehending the surface's orientation, curvature, and other geometric characteristics within this constrained area. For a variety of applications, including computer vision, robotics, and graphics, where exact information of the surrounding environment is necessary for tasks like object detection, navigation, or generating realistic scenes, this analysis is frequently crucial.

Local surface geometry in computer vision problems can be represented at each location in an image by descriptors like surface normals, curvature, or tangent vectors. These descriptors offer important details regarding the composition and form of surfaces or objects in the picture. Surface normals, for instance, show the direction perpendicular to a surface at a certain place, whereas tangent vectors depict the surface's local orientation or direction.In Fig-3 we can see the example of an local surface generated images with the original images

It is especially important to comprehend local surface geometry in situations when surfaces display intricate shapes or varying curvatures. For example, precise estimation of local surface geometry aids in the reconstruction of realistic and detailed three-dimensional representations of objects or situations when reconstructing three dimensions from pictures. Similarly, understanding local surface geometry helps with accurate action planning and execution in robotics applications such as grasping or manipulation.

Overall, local surface geometry plays a fundamental role in various fields, serving as a basis for understanding and analyzing the structure of surfaces within a localized region. Its accurate estimation and representation are essential for numerous computer vision, robotics,image forgery detection and graphics tasks, enabling machines to perceive and interact with their environment effectively.

B. UprightNet

UprightNet [3] is a novel approach for estimating the orientation of a camera in an indoor scene using only a single RGB image. Unlike previous methods that rely solely on black-box regression techniques through deep

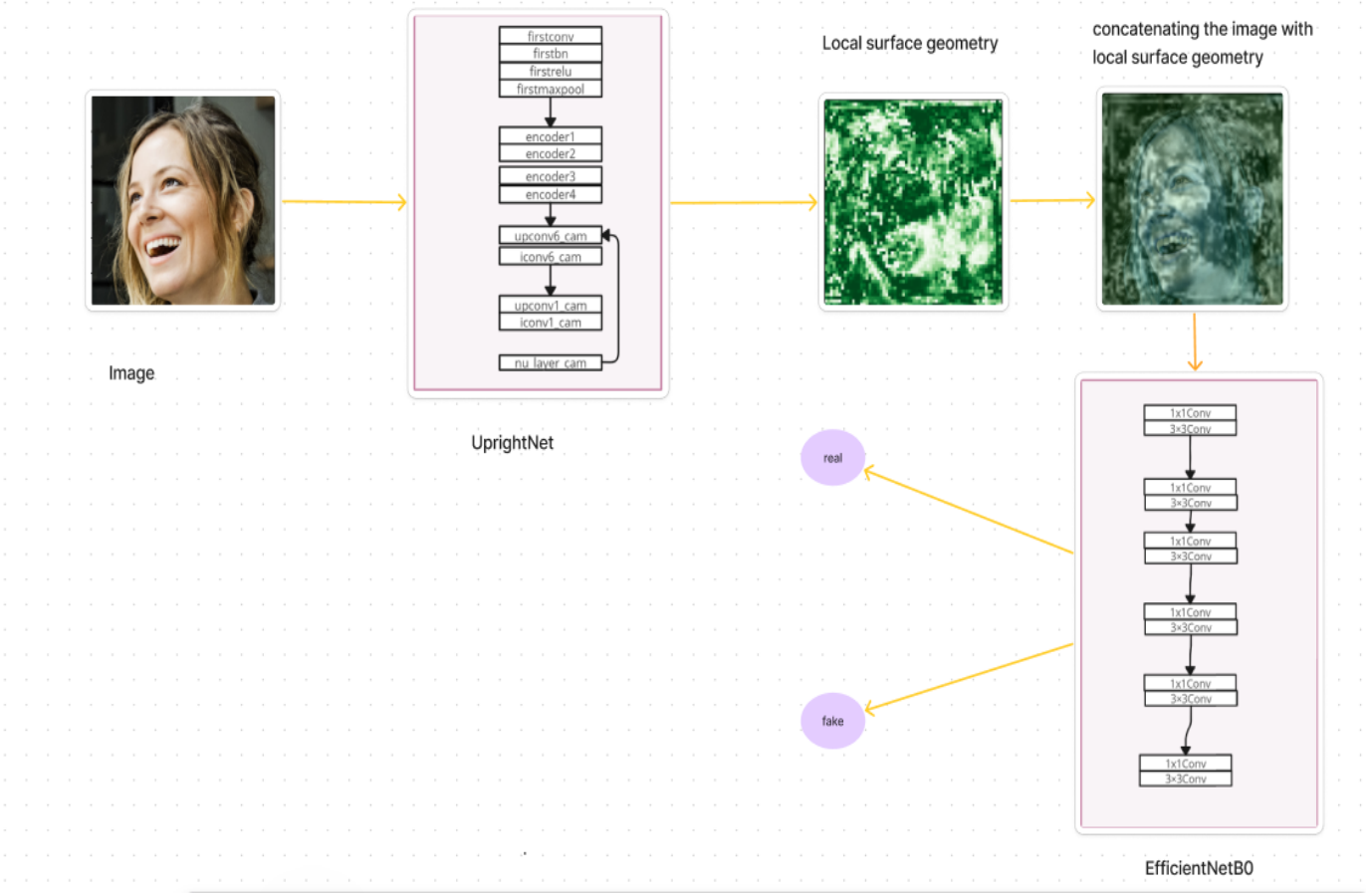

Fig. 2. The above figure shows the architecture of the LSG and EfficientNet

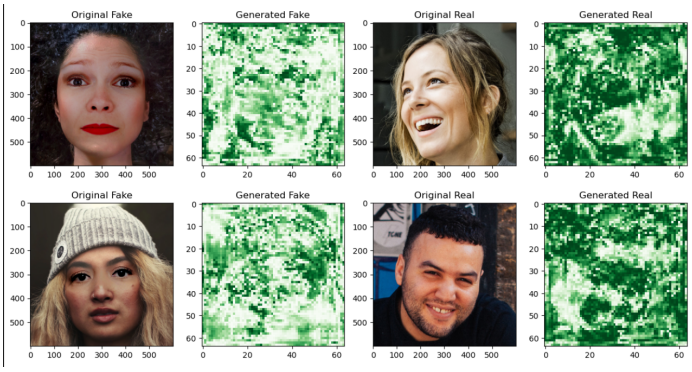

Fig. 3. we illustrate an example of the original face crop alongside manipulated versions generated by DFDC (first column). Subsequently, we present the corresponding local surface geometry (LSG) feature extracted by UpRightNet after processing the input face images (second column).

learning, UprightNet integrates explicit geometric reasoning into its framework. This is achieved by designing a network that predicts two representations of the scene's geometry: one in the local camera coordinate system and another in a global reference coordinate system.In Fig-4 we can see the architectural diagram of the UprightNet which used for generating the local surface geometry(LSG) The key innovation of UprightNet lies in its ability to solve for the camera orientation by aligning these two predicted geometries using a differentiable least squares module. This alignment process effectively determines the rotation needed to best match the local and global geometry representations, providing a precise estimation of the camera's orientation.

C. *Combining the original image with generated image:*

Alpha blending: Alpha blending is a technique used to combine two images by linearly interpolating their pixel values based on an alpha mask. The alpha mask specifies the transparency of each pixel in the overlaying image, allowing for smooth transitions between the two images. This technique is widely used in image processing and computer graphics due to its ability to create visually appealing composite images with seamless integration of different layers.In Fig-5 we can see the example of the orginal image and the combined image side by side

D. *EfficientNet*

In our proposed pipeline, we leverage a deep convolutional neural network (CNN) as the classification model to discern

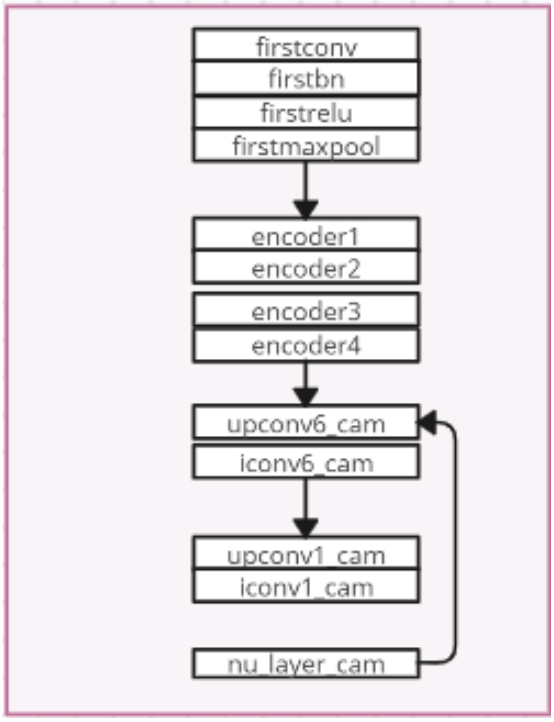

Fig. 4. Above image illustrate the architecture of the UprightNet which is used to find the local surface geometry.

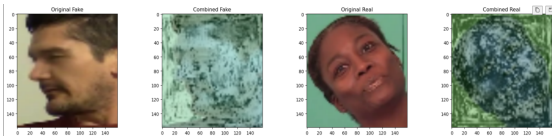

Fig. 5. Here the first image is the original fake image and second one is combined third one is original real image and fourth one is combined real image (alpha mark=0.2)

between authentic and manipulated images. Specifically, we adopt the EfficientNet architecture renowned for its efficacy in handling diverse image data with exceptional efficiency. EfficientNet models are characterized by a compound scaling method that optimizes network depth, width, and resolution simultaneously, ensuring superior performance while minimizing computational resources. These models excel in parameter efficiency, computation efficiency, scalability, and generalization, making them ideal for resource-constrained environments like mobile and edge devices. During training, the EfficientNet model learns to extract pertinent features from input images and make accurate predictions based on these features, with the possibility of fine-tuning to further enhance performance. Evaluation on a separate validation dataset allows us to measure the model's effectiveness in classifying real and fake images, assessing metrics such as accuracy, precision, recall, and F1 score to validate its robustness. By incorporating the EfficientNet architecture into our pipeline, we aim to achieve efficient and reliable detection of deepfake manipulations in visual content

IV. Result and Analysis

TABLE I
Training Procedure

Model	'efficientnet-b0' vairent
Optimization	Adam optimizer
learning rate	0.001
Loss function	cross-entropy loss function
Epochs	10
Batch size	32

A. Analysis of Local Geometric Representations

Upon visual inspection, it is observed that the LSG features may exhibit similarities across different faces, appearing uniform regardless of the facial content. This uniformity is attributed to the frontal framing of faces in the global upright coordinate system, akin to the consistent representation observed in indoor environments. Light green pixels in the LSG feature correspond to surfaces whose normal are perpendicular to the upward vector, resembling walls in a room and most facial pixels. Conversely, regions colored with shades dark green encode surfaces parallel to the ground, typically located at the top and bottom of the image

Although the LSG feature may initially appear minimally informative, subtle details become more discernible to a neural network trained to detect synthetic patterns and anomalies in geometric estimations of the face. The presence of white spaces in generated images and black spaces in original images suggests a potential correlation with the magnitude of local surface geometry features, indicating their significance in discriminating between authentic and manipulated images.

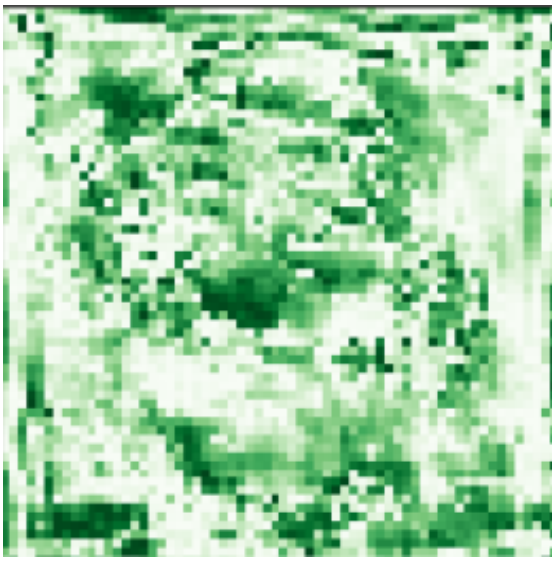

Fig. 6. Above image illustrate the local surface geometry of a image .This is a sample image to illustrate the results

Black Regions (refer Fig.6): Black regions in the surface Normals typically represent areas where there is no variation in the surface orientation. These areas can correspond to flat

or uniform surfaces in the scene. In an image that has been manipulated or generated synthetically, black regions might indicate areas where the algorithm has failed to generate realistic surface normal, potentially revealing inconsistencies or artifacts.

White Regions(refer Fig.6): White regions, on the other hand, represent regions where there is a significant variation in surface orientation. These areas can correspond to edges, corners, or other complex surface structures. In a deepfake image, white regions might indicate areas where the manipulation has introduced irregularities or distortions that affect the surface normal

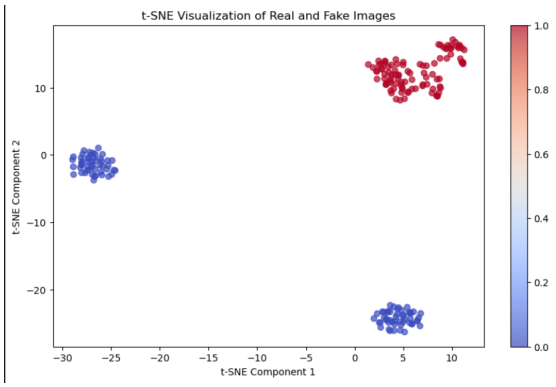

Fig. 7. The tsne graph illustrate the difference between the fake and real images

Two Distinct Groups: The presence of two distinct groups of points, one in blue and the other in red, suggests that the t-SNE algorithm has successfully separated the real and fake images into two clusters based on their feature representations. Each point in the plot represents an image, and the position of the points is determined by the similarity of their features.

Blue Cluster (Real Images): The blue cluster likely represents the real images. Since these images are genuine and come from the real dataset, they are expected to share common features and thus form a cohesive cluster. The fact that the blue cluster is separated from the red cluster indicates that the features of real images are different from those of fake images.

Red Cluster (Fake Images): The red cluster likely represents the fake images generated by the model. These images are synthetically generated and may exhibit different characteristics compared to real images. The separation of the red cluster from the blue cluster suggests that the features extracted from fake images are distinct from those of real images.

Large Separation: The significant separation between the blue and red clusters indicates that there are substantial differences in the features of real and fake images. This suggests that the model has successfully learned to generate fake images that are visually distinct from real images, at least in terms of the extracted features used for t-SNE visualization.

The table-2 mentions about our investigation into image forgery detection within the realm of deepfake images, the accuracies garnered by ResNet, DenseNet, and EfficientNet provide profound insights into the efficacy of these neural

TABLE II
ARCHITECTURES PERFORMANCE

Architecture	Accuracy
ResNet	99.98
DesNet	99.97
EefficientNet	98.20

network architectures. ResNet, exhibiting the highest accuracy at 99.98%, showcases its remarkable proficiency in discerning between authentic and manipulated regions within images. This exceptional performance underscores ResNet's ability to effectively learn and generalize intricate local surface geometry features crucial for identifying forged content in deepfake images. DenseNet closely follows with an accuracy of 99.97%, highlighting its robustness and proficiency comparable to ResNet. Despite minor disparities, both ResNet and DenseNet demonstrate strong capabilities in detecting image forgeries, owing to their deep and densely connected architectures, which enable them to extract and integrate features across various scales with precision.

In contrast, EfficientNet, while achieving a respectable accuracy of 98.20%, presents potential limitations in capturing fine-grained details and complex spatial relationships essential for accurate forgery detection. This discrepancy may stem from EfficientNet's reliance on compound scaling, which optimizes model size and computational efficiency across multiple dimensions, potentially compromising its ability to discern subtle manipulations compared to ResNet and DenseNet. Nonetheless, the achieved accuracy reaffirms EfficientNet's viability as a candidate for forgery detection tasks, particularly in scenarios where computational resources are constrained or where a balance between performance and efficiency is crucial.

The notable accuracies attained by all three models underscore the promising role of deep learning techniques in combating image forgery within the context of deepfake images. These results emphasize the significance of architectural considerations and optimization strategies in designing effective forgery detection systems. Further exploration, including the analysis of misclassified samples and the exploration of alternative hyperparameters and augmentation techniques, holds promise for enhancing the accuracy and robustness of forgery detection systems in addressing the growing threat posed by the proliferation of deepfake content across digital platforms.

CONCLUSION

The investigation underscores the critical role of local surface geometry (LSG) data extracted by UpRightNet in deepfake detection tasks. Integration of LSG features with neural network models presents a promising avenue to bolster forensic analysis techniques, thereby enhancing the accuracy of detecting synthetic manipulations and fortifying the resilience against digital forgeries. However, further exploration and experimentation are warranted to fully harness the poten-

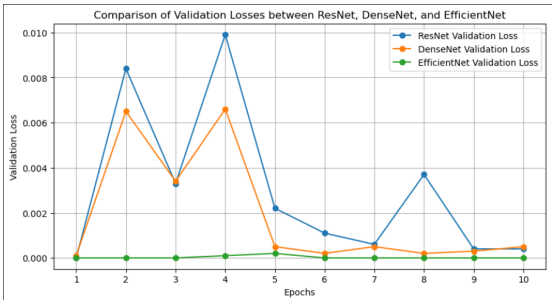

Fig. 8. The graph illustrates the comparison of training loss between the architectures.

tial of LSG features in combating the proliferation of false content in digital media.

Noteworthy results were attained across multiple models, with ResNet and DenseNet demonstrating exceptional accuracy in discerning genuine from manipulated images. ResNet, achieving a remarkable accuracy of 99.98%, and DenseNet, closely following with an accuracy of 99.97%, showcase the robustness and efficacy of deep convolutional architectures in identifying subtle manipulations within deepfake images. These results affirm the capability of ResNet and DenseNet to effectively learn and generalize intricate features, contributing significantly to the advancement of forgery detection techniques.

Furthermore, the EfficientNet model delivered outstanding performance, achieving an impressive accuracy of 98%. This accomplishment signifies the model's efficacy in discriminating between authentic and forged photos, with profound implications for reinforcing the reliability and authenticity of visual content. The notable achievements across ResNet, DenseNet, and EfficientNet mark a substantial milestone in the ongoing pursuit of developing robust methods for identifying deepfake manipulations, underscoring their pivotal role in safeguarding the trustworthiness of visual content in the digital landscape.

REFERENCES

[1] Zahid Akhtar, Dipankar Dasgupta "A Comparative Evaluation of Local Feature Descriptors for DeepFakes Detection" Proceedings of the Third International Conference on Inventive Research in Computing Applications (ICIRCA-2021) IEEE Xplore Part Number: CFP21N67-ART

[2] Dheeraj J C, Krutant Nandakumar "Detecting Deepfakes Using Deep Learning" 2021 6th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), August 27th & 28th 2021

[3] Andrea Ciamarra, Roberto Caldelli "Deepfake detection by exploiting surface anomalies: the SurFake approach" IEEE Xplore.

[4] Wenqi Xian, Zhengqi Li "UprightNet: Geometry-Aware Camera Orientation Estimation from Single Images" IEEE Xplore

[5] Arnab Banerjee , Nibaran Das "Weber local descriptor for image analysis and recognition" Published online: 25 November 2020

[6] Jie Chen , Shiguang Shan "WLD: A Robust Local Image Descriptor" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2009, TPAMI-2008-09-0620

[7] Atmik Ajoy, Chethan U Mahindrakar "DeepFake Detection using a frame based approach involving CNN"

[8] MD SHOHEL RANA, MOHAMMAD NUR NOBI3 "Deepfake Detection: A Systematic Literature Review" IEEE Access Received January 25, 2022, accepted February 16, 2022, date of publication February 24, 2022, date of current version March 10, 2022.

[9] Vedant Jolly, Mayur Telrandhe "CNN based Deep Learning model for Deepfake Detection" 2022 2nd Asian Conference on Innovation in Technology (ASIANCON) Pune, India. Aug 26-28, 2022

[10] Rineesh Babu P, Madhu S. Nair "Deepfake Detection using Multi-path CNN and Convolutional Attention Mechanism" 2022 IEEE 2 nd Mysore Sub Section International Conference (MysuruCon)

[11] Link to the dataset:https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images

[12] Vidhyasagar B.S ,Suresh , Shanmughanathan, Senthilkumar "Deep learning-based image forgery detection system" International Journal of Electronic Security and Digital Forensics

[13] Bhuvaneswari R., Enaganti, Karun Kumar "Robust Image Forgery Classification using SqueezeNet Network" 2023 1st International Conference on Advances in Electrical, Electronics and Computational Intelligence, ICAEECI

[14] Kalpana K, Amritha P.P. "Image manipulation detection using Deep Learning in tensor flow" International Journal of Control Theory and Applications

[15] Dhanishtha Patil, Vaibhav Narawade. "A Novel Approach to Image Forgery Detection Techniques in Real World Applications" International Journal of Control Theory and Applications

[16] Bahar Uddin Mahmud, Afsana Sharmin. "Deep Insights of Deepfake Technology : A Review" In book: Applications of Artificial Intelligence and Machine Learning

[17] Sahu, Aditya Kumar,Biradar, Vaishali D,Sri Vigna Hema V."A Study on Content Tampering in Multimedia Watermarking",ISSN,2662995X

[18] Sandotra, Neha,Arora, Bhavna."A comprehensive evaluation of feature-based AI techniques for deepfake detection"Neural Computing and Applications

[19] Zhu, Yizhe,Zhang, Chunhui, Gao, Jialin,Rui, Zihan."High-compressed deepfake video detection with contrastive spatiotemporal distillation"Neurocomputing