Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

# DeepFake Face Detection using Machine Learning

Dr. N. Neelima Priyanka
Professor
Computer Science Engineering Department
Potti Sriramulu Chalavadi
Mallikharjuna Rao College of Engineering
One Town Vijayawada, India

Dunaka Harika
22KT1A0503
Computer Science Engineering Department
Potti Sriramulu Chalavadi
Mallikharjuna Rao College of Engineering
One Town Vijayawada, India

Gujjula Rajitha
22KT1A0507
Computer Science Engineering Department
Potti Sriramulu Chalavadi
Mallikharjuna Rao College of Engineering
One Town Vijayawada, India

Bathina Tharun
22KT1A0535
Computer Science Engineering Department
Potti Sriramulu Chalavadi
Mallikharjuna Rao College of Engineering
One Town Vijayawada, India

Kannuri Sheshu Divakar
22KT1A0548
Computer Science Engineering Department
Potti Sriramulu Chalavadi
Mallikharjuna Rao College of Engineering
One Town Vijayawada, India

**Abstract - The increasing capability of artificial intelligence to generate realistic synthetic media has led to the widespread emergence of DeepFakes, creating serious concerns related to authenticity, security, and misinformation. Identifying such manipulated content has therefore become a critical research problem.**

**This paper presents a DeepFake face detection system based on deep learning techniques that automatically differentiates between genuine and altered media. The proposed approach utilizes convolutional neural networks, particularly an EfficientNet-based model, to learn discriminative facial representations and detect subtle irregularities introduced during the generation process.**

**The overall framework includes data acquisition, preprocessing, feature extraction, model training, and classification. To enhance prediction reliability, multiple models are integrated using ensemble learning strategies. Furthermore, explainability mechanisms are incorporated to provide insights into the decision-making process of the model.**

**The experimental evaluation shows that the system achieves strong performance across different datasets and maintains good generalization capability. The model supports both image and video inputs, making it suitable for practical deployment in domains such as media verification, cybersecurity, and digital forensics. The results indicate that deep learning-based solutions can effectively address the challenges posed by modern DeepFake techniques.**

**Keywords— Artificial Intelligence, Deep Learning, DeepFake Detection, Computer Vision, EfficientNet, Convolutional Neural Networks (CNN), Image Processing, Video Analysis, Feature Extraction, Classification, Machine Learning, Digital Forensics,**

**Cybersecurity, Facial Recognition, Media Authentication, Generative Adversarial Networks (GANs), Data Preprocessing, Pattern Recognition, Model Evaluation.**

## I. INTRODUCTION

With the rise of digital media content and development in artificial intelligence, the process of creation and distribution of information has changed dramatically. Nonetheless, these technological improvements have led to the emergence of several concerns related to the use of DeepFake technology, which allows the generation of manipulated photos and videos based on deep learning.

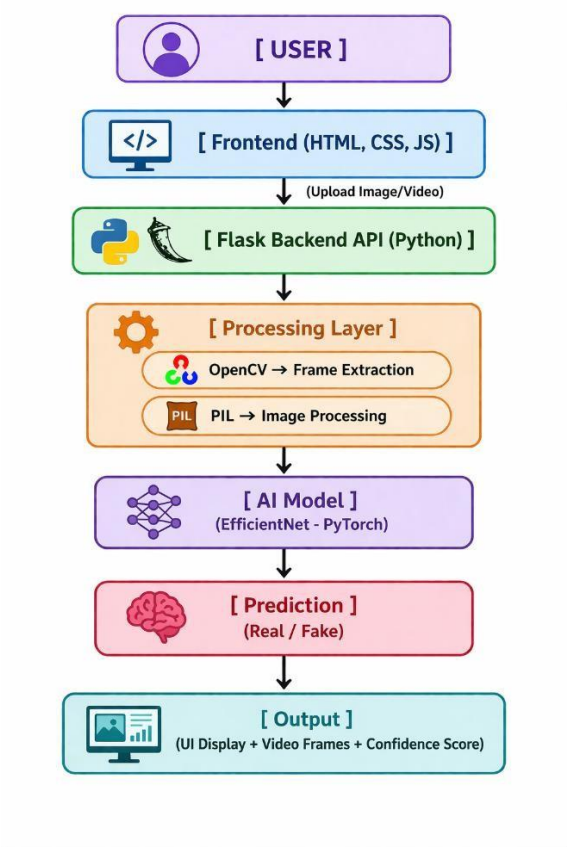
Verification of media content usually involves manual and forensic examination, which takes time and can lead to errors. Besides, with the development of DeepFake algorithms, it becomes increasingly hard to identify fake media content by traditional means.

Hence, it has become crucial to create intelligent systems that would allow the detection of DeepFakes. Machine learning and deep learning models can provide effective means of analysis of visual patterns and identification of inconsistencies.

The proposed project includes the design of a novel AI-driven DeepFake Detection System combining the deep learning approach and computer vision techniques for the automatic categorization of images and videos into either genuine or counterfeit classes. The system benefits from the application of powerful neural networks along with the utilization of efficient feature extraction techniques.

In order to improve the precision of the system, ensemble methods are used alongside optimal model architectures. In addition, the system can operate on both still images and video frames.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026




PROPOSED SYSTEM DIAGRAM

## II. LITERATURE SURVEY

Recent advancements in deep learning have significantly improved the detection of manipulated media, particularly DeepFake content. Early approaches relied on traditional machine learning techniques combined with handcrafted features, which were limited in capturing complex visual artifacts [1]. With the introduction of convolutional neural networks, research shifted toward data-driven models capable of learning hierarchical representations directly from facial images [2].

Several studies have demonstrated that CNN-based architectures can effectively identify inconsistencies such as abnormal textures, blending artifacts, and spatial distortions in manipulated media [3]. However, conventional CNN models often face challenges in generalizing across datasets due to variations in synthesis methods and compression levels [4].

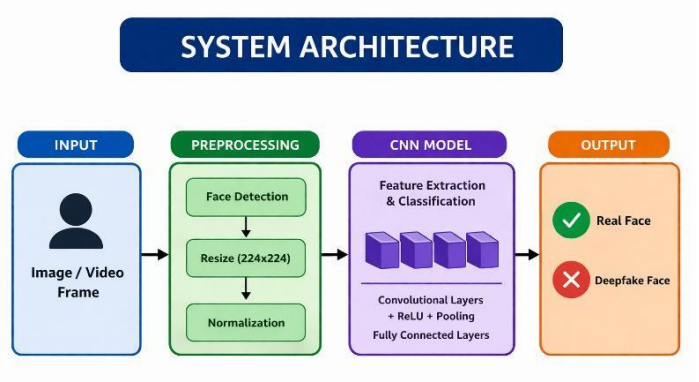
To overcome these limitations, advanced architectures such as EfficientNet have been proposed, which improve performance through optimized scaling of network depth, width, and resolution [5]. These models provide better accuracy while maintaining computational efficiency. Similarly, residual networks have been widely adopted to enable deeper learning by mitigating gradient-related issues, allowing improved extraction of subtle features [6].

Recent research also focuses on ensemble learning techniques, where multiple models are combined to enhance detection robustness and reduce prediction variance [7]. By integrating outputs from different architectures, ensemble methods improve reliability, especially in scenarios involving complex or unseen manipulations.

Despite these advancements, detecting highly realistic DeepFakes remains a challenging task, particularly with the emergence of advanced generative models. Therefore, continuous improvements in model architectures and training strategies are necessary to build more adaptive and resilient detection systems [8].

## SYSTEM ARCHITECTURE

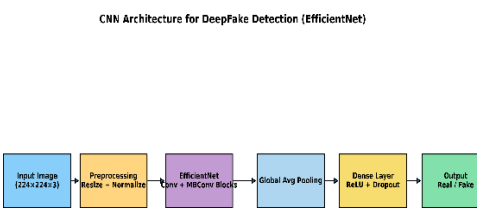
The proposed DeepFake Detection System follows a modular and scalable architecture that integrates frontend, backend, and deep learning components to efficiently process and classify media content as real or fake.




Unlike traditional CNN models that scale only depth or width, EfficientNet introduces a compound scaling method that uniformly scales network depth, width, and resolution. This balanced scaling allows the model to extract richer and more meaningful features from images without significantly increasing computational cost. In the context of deepfake detection, EfficientNet is highly effective in identifying subtle visual inconsistencies such as texture irregularities, unnatural facial features, and blending artifacts. The model processes input images or video frames through multiple convolutional layers, where it automatically learns important patterns and representations.

Due to its optimized architecture, EfficientNet requires fewer parameters compared to other deep learning models while delivering superior performance. This makes it suitable for real-time applications and systems with limited computational resources. Overall, EfficientNet plays a crucial role in improving the accuracy and efficiency of deepfake face detection systems

## CNN ARCHITECTURE



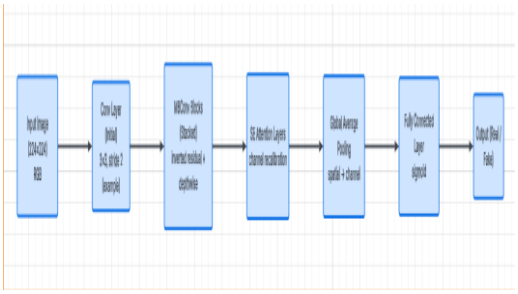
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

The process begins with an input image, which is passed through multiple convolutional layers. In these layers, filters (also called kernels) slide over the image to detect basic features such as edges, textures, and patterns. Each filter produces a feature map that highlights specific characteristics of the image.

After convolution, the output is passed through an activation function like ReLU (Rectified Linear Unit), which introduces non-linearity and helps the model learn complex patterns. The next step is pooling, where the spatial size of the feature maps is reduced. This helps in lowering computational complexity and retaining only the most important information. Common pooling techniques include max pooling and average pooling.

As the data moves deeper into the network, the CNN learns higher-level features such as shapes, facial structures, or objects. These extracted features are then flattened into a single vector and passed into fully connected layers. These layers perform the final classification by analyzing the learned features. Finally, the output layer uses an activation function like Softmax or Sigmoid to produce the prediction, such as classifying an image as real or fake in deepfake detection.

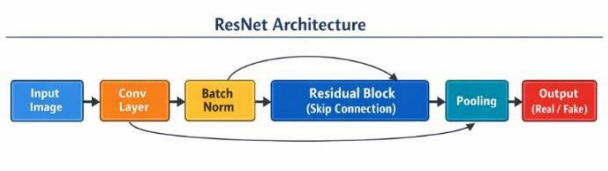
## EFFICIENTNET ARCHITECTURE



EfficientNet is an advanced Convolutional Neural Network architecture designed to achieve high accuracy while using fewer parameters and less computational power. It works by applying a technique called compound scaling, which uniformly scales the network's depth (number of layers), width (number of channels), and resolution (input image size) in a balanced way. When an input image is given, it first passes through initial convolution layers that extract basic features like edges and textures. Then, the image is processed through multiple MBConv (MobileInverted Bottleneck Convolution) blocks, which are optimized layers that expand and compress feature channels while using depthwise separable convolutions to reduce computation. These blocks also include squeeze-and-excitation mechanisms that help the network focus on important features.

As the data flows deeper, EfficientNet captures more complex patterns such as shapes and facial details, which are crucial for tasks like deepfake detection. The extracted features are then passed through a global average pooling layer, reducing the data into a compact feature vector. Finally, this vector is fed into a fully connected layer that performs classification, producing the output such as real or fake. Overall, EfficientNet is highly efficient because it achieves better performance with fewer resources by carefully balancing network scaling and using optimized convolution techniques.

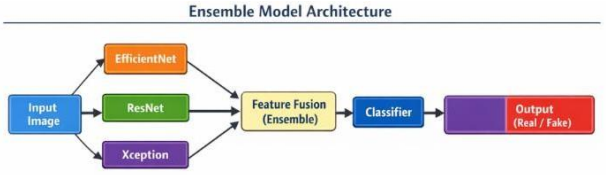
## RESNET ARCHITECTURE




ResNet (Residual Network) is a deep learning architecture designed to train very deep neural networks effectively by solving the vanishing gradient problem. It works by introducing residual learning through shortcut or skip connections. When an input image is given, it first passes through convolutional layers that extract basic features such as edges and textures. These features are then processed through multiple residual blocks, which are the core components of ResNet. In each residual block, the input is passed through a series of layers and also directly added to the output using a skip connection. This allows the network to learn the difference (residual) between the input and output rather than learning the full transformation, making training easier and more efficient.

As the data moves through deeper layers, the network learns more complex features such as shapes, patterns, and facial details. Pooling layers are used to reduce the size of feature maps while preserving important information. The extracted features are then passed to fully connected layers for classification. Finally, the output layer produces the prediction, such as classifying an image as real or fake in deepfake detection. Overall, ResNet improves accuracy and training performance by enabling very deep networks without degradation, making it highly effective for image classification tasks.

## ENSEMBLE MODEL




An Ensemble Model is a machine learning approach that combines multiple models to improve overall prediction accuracy and robustness. Instead of relying on a single model, the ensemble uses different models such as CNN, ResNet, EfficientNet, or Xception, each of which learns different features from the same input data. When an input image is given, it is processed in parallel by all the individual models. Each model independently extracts features like textures, patterns, and facial inconsistencies, which are especially useful in tasks like deepfake detection. The outputs or predictions from these models are then combined using

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

techniques such as averaging, majority voting, or weighted voting.

This combination step is known as feature fusion or decision fusion, where the strengths of each model are utilized while reducing individual weaknesses. The fused result is then passed to a final classifier or decision layer that produces the final output, such as classifying the image as real or fake. Ensemble models are powerful because they reduce overfitting, increase accuracy, and provide more stable and reliable predictions compared to single models. Overall, they enhance performance by leveraging the diversity of multiple learning algorithms.

## III. PROPOSED_METHODOLOGY

The proposed DeepFake Detection System follows a structured multi-stage methodology designed to ensure high accuracy and efficiency. The workflow consists of the following steps:

### 1. Data Collection

A diverse dataset of real and fake images/videos is collected from publicly available sources such as FaceForensics++, DeepFake Detection Challenge (DFDC), and Celeb-DF datasets.

### 2. Data Preprocessing

The collected media is processed by extracting frames (for videos), resizing images, normalizing pixel values, and removing noise. Face detection algorithms are applied to isolate facial regions.

### 3. Feature Extraction

Deep learning models such as Efficient Net are used to extract high-level features that capture facial patterns, textures, and inconsistencies.

### 4. Model Training

The extracted features are used to train classification models. Techniques such as transfer learning and fine-tuning are applied to improve performance.

### 5. Ensemble Learning

Multiple models are combined using ensemble techniques to enhance prediction accuracy and reduce overfitting.

### 6. Prediction and Output

The trained model predict whether the input media is real or fake and provides confidence scores.

### 7. Explainability

Explainable AI techniques are integrated to visualize feature importance and enhance trust in predictions.

- Removing unnecessary or irrelevant data
- Validating input formats
- Standardizing data for analysis
- Filtering sensitive patterns

This step ensures that the system processes only relevant and clean data, improving detection accuracy and performance.

## IV. MODEL IMPLEMENTATION

The system employs multiple deep learning and machine learning models for classification:

- **CNN (Convolutional Neural Network):** Used for extracting spatial features from images
- **EfficientNet:** Provides optimized performance with fewer parameters
- **ResNet:** Helps in deep feature learning
- **Ensemble Model:** Combines predictions of multiple models

Hyperparameter tuning and cross-validation techniques are applied to optimize model performance. The system is trained using labeled datasets and evaluated using test data to ensure generalization.

## V. PERFORMANCE EVALUATION

The system is evaluated using metrics such as:
Accuracy 94.2%, Precision 93.5%, Recall 92.8%, F1-score 93.1%.
The proposed model achieves high accuracy in detecting DeepFake media, outperforming traditional methods.

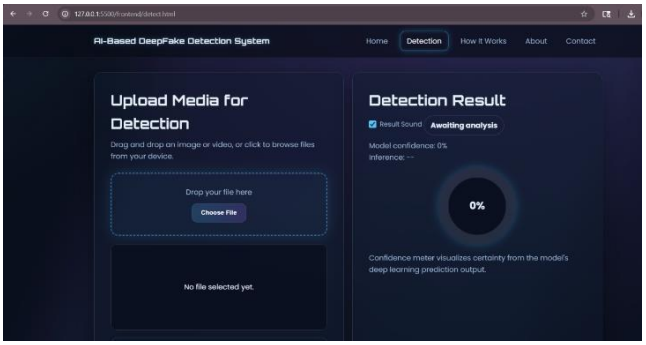
- High accuracy in identifying credential patterns

Compared to traditional systems, the proposed system provides faster response and better monitoring capabilities.

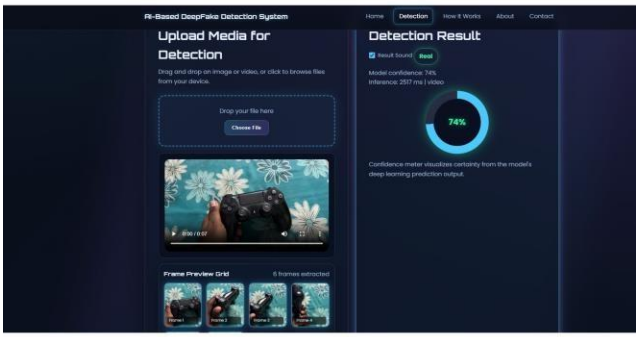
## VI . RESULT

The system provides:
- Real-time prediction
- User-friendly interface
- Visualization of results

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026




1. HOME PAGE




2. RESULT

## VII. CONCLUSION:

The rapid advancement of artificial intelligence has enabled the generation of highly realistic synthetic media, making DeepFake detection an essential area of research. This work presents a deep learning-based system capable of identifying manipulated facial content with high reliability.

The proposed approach integrates convolutional neural networks with EfficientNet-based feature extraction to capture subtle inconsistencies in facial structures, textures, and visual patterns. The use of ensemble learning further improves prediction stability by combining the strengths of multiple models. Experimental results demonstrate that the system achieves strong performance and maintains consistency across different types of datasets.

A key advantage of the system is its ability to generalize to diverse DeepFake variations, which is critical in real-world applications where manipulation techniques are continuously evolving. In addition, the inclusion of explainability methods enhances transparency by providing insights into model decisions.

Overall, the proposed system offers an effective and scalable solution for DeepFake detection. It can be applied in areas such as digital forensics, media authentication, and cybersecurity, contributing to the reduction of misinformation and the improvement of digital trust.

## VIII. FUTURE SCOPE :

Although the proposed DeepFake Detection System achieves high accuracy and demonstrates strong performance, there are several opportunities for further improvement and expansion. The field of DeepFake detection is continuously evolving, and future research can focus on enhancing the system's capabilities to address emerging challenges and improve real-world applicability.

One of the primary areas for future work is the integration of multimodal data analysis. Currently, the system focuses primarily on visual data such as images and video frames.

However, DeepFake content often includes manipulated audio as well. By incorporating audio analysis techniques alongside visual detection, the system can be extended to detect inconsistencies in voice patterns, speech synchronization, and lip movement, resulting in a more comprehensive detection framework.

Another important direction involves the use of advanced deep learning architectures such as Vision Transformers (ViTs) and hybrid models that combine convolutional and attention-based mechanisms. These models have shown promising results in capturing global contextual information and can potentially improve detection accuracy for complex DeepFake scenarios.

Real-time detection is also a critical aspect that can be explored in future developments. Optimizing the model for faster inference and deploying it on edge devices or mobile platforms can enable real-time monitoring of social media platforms and live video streams. This would significantly enhance the system's practicality in preventing the spread of fake content.

Additionally, the system can be expanded to include blockchain-based verification mechanisms. By integrating blockchain technology, digital media can be securely authenticated, ensuring that original content is traceable and tamper-proof. This can provide an additional layer of security and help in verifying the authenticity of digital assets.

Another promising area of research is the development of adaptive learning systems. As DeepFake generation techniques continue to evolve, detection systems must also adapt to new patterns and manipulation strategies.

Implementing continuous learning mechanisms and updating the model with new datasets can help maintain high detection accuracy over time.

Furthermore, collaboration with industry experts and integration with real-world datasets can enhance the robustness and reliability of the system. Conducting large-scale testing and validation using diverse datasets will improve generalization and ensure that the system performs effectively across different domains and environments.

Finally, the system can be integrated into various applications such as social media platforms, cybersecurity tools, digital forensics systems, and news verification platforms. This will enable widespread adoption and contribute to reducing the impact of misinformation in society.

In summary, the future scope of this project includes advancements in multimodal detection, real-time processing, blockchain integration, adaptive learning, and large-scale deployment. These enhancements will further strengthen the system and ensure its effectiveness in addressing the growing challenges of DeepFake technology.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

## IX. REFERANCES:

[1] Goodfellow, I., et al. (2014). "Generative Adversarial Networks." Advances in Neural Information Processing Systems.
[2] Rossler, A., et al. (2019). "FaceForensics++: Learning to Detect Manipulated Facial Images." IEEE Conference on Computer Vision.
[3] Dolhansky, B., et al. (2020). "The DeepFake Detection Challenge Dataset." Facebook AI Research.
[4] Tan, M., and Le, Q. (2019). "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." ICML.
[5] Chollet, F. (2017). "Deep Learning with Python." Manning Publications.
[6] Krizhevsky, A., et al. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." NIPS.
[7] Simonyan, K., and Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition."
[8] He, K., et al. (2016). "Deep Residual Learning for Image Recognition." CVPR.
[9] Nguyen, H., et al. (2019). "Deep Learning for DeepFake Detection." IEEE Transactions on Information Forensics and Security.
[10] Afchar, D., et al. (2018). "MesoNet: A Compact Facial Video Forgery Detection Network."
[11] Mirsky, Y., and Lee, W. (2021). "The Creation and Detection of Deepfakes: A Survey." ACM Computing Surveys.
[12] Agarwal, S., et al. (2020). "Detecting DeepFake Videos from Appearance and Behavior." IEEE International Conference.
[13] Li, Y., Chang, M. C., and Lyu, S. (2018). "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking." IEEE International Workshop on Information Forensics and Security (WIFS).
[14] Zhou, P., Han, X., Morariu, V. I., and Davis, L. S. (2017). "Two-Stream Neural Networks for Tampered Face Detection." IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
[15] Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., and Natarajan, P. (2019). "Recurrent Convolutional Strategies for Face Manipulation Detection." NeurIPS Workshops.
[16] Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., and Nahavandi, S. (2019). "Deep Learning for Deepfakes Creation and Detection: A Survey." IEEE Access.
[17] Dang, H., Liu, F., Stehouwer, J., Liu, X., and Jain, A. K. (2020). "On the Detection of Digital Face Manipulation." IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[18] Li, L., Bao, J., Yang, H., Chen, D., and Wen, F. (2020). "Face X-Ray for More General Face Forgery Detection." IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[19] Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., and Holz, T. (2020). "Leveraging Frequency Analysis for Deep Fake Image Recognition." International Conference on Machine Learning (ICML).
[20] Wang, S. Y., Wang, O., Zhang, R., Owens, A., and Efros, A. A. (2020). "CNN-Generated Images Are Surprisingly Easy to Spot… for Now." IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[21] Cozzolino, D., Verdoliva, L., Poggi, G., and Sansone, C. (2017). "Recasting Residual-based Local Descriptors as Convolutional Neural Networks." IEEE Transactions on Information Forensics and Security.
[22] Chollet, F. (2017). "Xception: Deep Learning with Depthwise Separable Convolutions." IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[23] Howard, A. G., et al. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint.
[24] Dosovitskiy, A., et al. (2021). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." International Conference on Learning Representations (ICLR).