Deepfake-Proof eKYC: A Unified System for Identity Verification and Forgery Detection

Diana Mary, Raghav Agarwal, and Armugam Ajay
T-REX

***Abstract*—The proliferation of sophisticated image editing and deepfake generation tools poses a significant threat to the security of digital identity verification systems. To address this, we propose a unified eKYC system that integrates two specialized deep learning models: a hybrid spatial-frequency detector for robust forgery identification and a lightweight metric learning model for efficient face identification. Our forgery detection module combines an EfficientNet-B3 backbone with a custom frequency-domain head, achieving a validation accuracy of 99.65% and an AUC of 0.9999. For identity verification, we employ a MobileNetV3-Small backbone trained with an ArcFace loss function to produce highly discriminative facial embeddings, achieving a validation AUC of 0.880 on unseen identities. This paper details the architecture, training, and performance of this dual-component system, demonstrating its effectiveness in creating a secure, real-time eKYC solution.**

***Index Terms*—Deepfake Detection, eKYC, Identity Verification, Face Identification, Forgery Detection, Deep Metric Learning, ArcFace, MobileNetV3, EfficientNet.**

I. Introduction

Automated identity verification from digital images is a fundamental challenge in computer vision with widespread applications. The core task is to create systems that are not only accurate but also robust to real-world variations and computationally efficient for deployment on edge devices. However, the ease with which images can be manipulated has led to a surge in digital forgeries, from simple edits to complex deepfakes, which threaten to erode trust in digital media.

To combat this, we have developed a unified eKYC system with two core capabilities. The first is forgery detection. Forgery detection methods typically operate in either the spatial domain (analyzing visual content like textures and edges) or the frequency domain (analyzing periodic structures altered by operations like compression or splicing). While models in a single domain have shown success, a hybrid approach combining both has the potential for greater robustness.

The second capability is face identification, for which deep metric learning has become a leading paradigm. These methods learn an embedding function that maps face images to a high-dimensional vector space where images of the same person are clustered together. Loss functions like ArcFace [1] have shown exceptional performance by incorporating an additive angular margin penalty, encouraging more discriminative features.

This paper presents a unified system featuring a hybrid spatial-frequency model for forgery detection and a lightweight face embedder using MobileNetV3-Small with an ArcFace head for identification. We demonstrate that this combined approach provides a comprehensive solution for secure eKYC.

II. Proposed Methodology

Our proposed solution is a multi-stage pipeline designed for accuracy and efficiency. The system integrates two specialized deep learning models to handle the distinct tasks of forgery detection and identity verification. The overall architecture is depicted in Figure 1.

A. A Hybrid Approach for Forgery Detection

To identify manipulated images, we developed the **Hybrid-Detector**, a model that extracts and fuses features from both the spatial and frequency domains. This synergy allows the model to identify a wider range of manipulations, from high-level semantic inconsistencies to subtle, low-level compression artifacts. The architecture is shown in Figure 2.

1) Spatial Path: A pre-trained EfficientNet-B3 model serves as the backbone. The feature map is processed by an adaptive average pooling layer and a feed-forward network to generate a 128-dimensional spatial feature vector.

2) Frequency Path: This path captures artifacts in the frequency domain. The input RGB image is converted to grayscale, and a 2D Fast Fourier Transform (FFT) is computed. The log-magnitude of the FFT is then flattened and passed through an MLP to extract a 128-dimensional frequency feature vector.

3) Final Classification: The feature vectors from both paths are concatenated and fed into a final classification head, which uses a Sigmoid activation function to output a forgery probability score.

B. A Lightweight Model for Face Identification

For verifying identity, we designed a lightweight model focused on generating discriminative embeddings for face images. The model, shown in Figure 3, is composed of a feature extractor and a specialized training head.

1) Embedder: A pre-trained MobileNetV3-Small model acts as the feature extractor. Its classification layer is replaced with a head that projects the feature vector into a 512-dimensional embedding space. The final embedding is L2-normalized.

2) ArcMarginProduct (ArcFace Head): This head is used only during training. It computes the cosine similarity between embeddings and learnable class-weights, applies an additive angular margin (m=0.5), and scales the logits (s=30.0) before passing them to the Cross-Entropy Loss function.

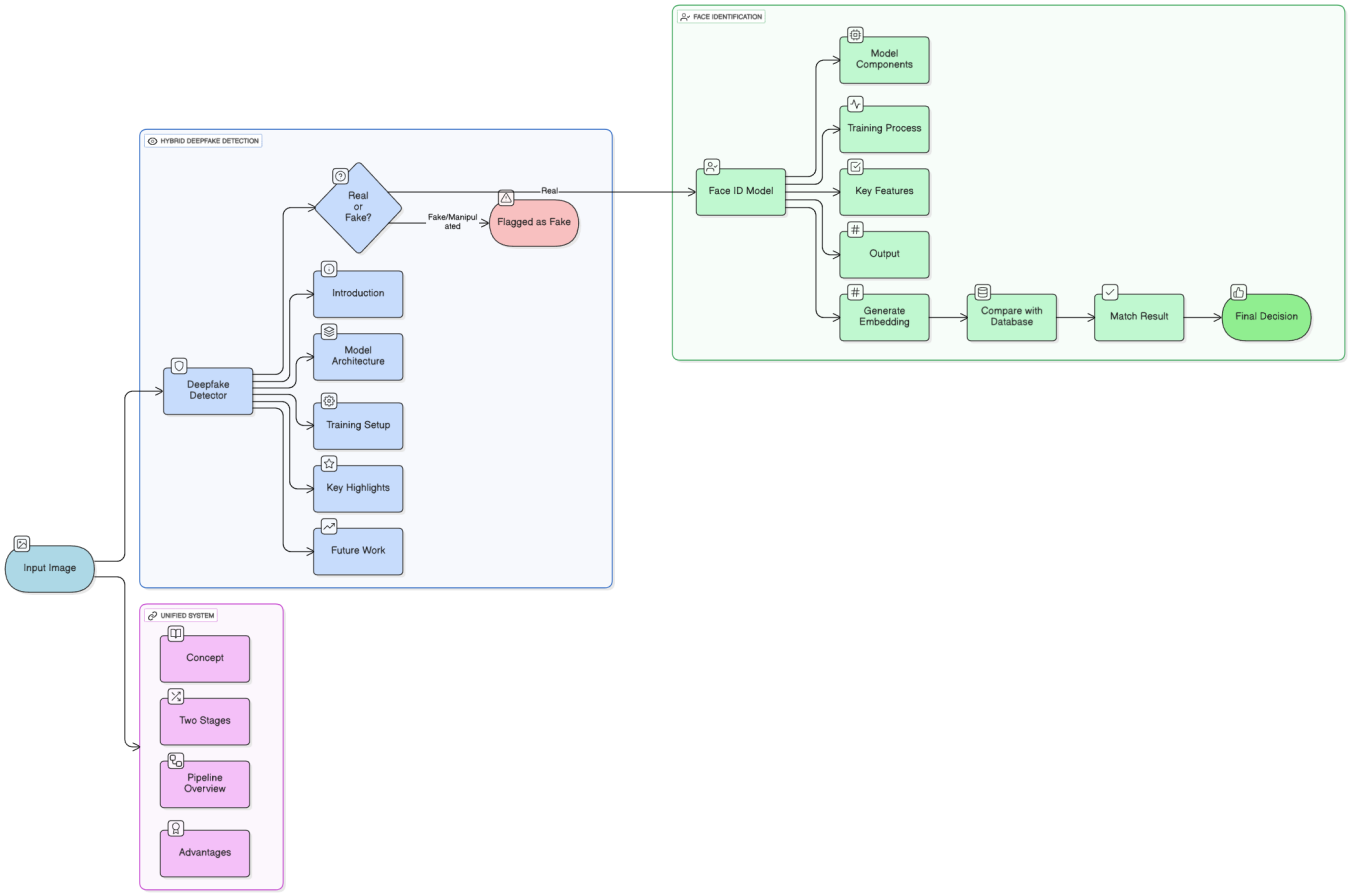

Fig. 1. High-level architecture of our unified eKYC system. The system takes an input image and processes it through two parallel modules to ascertain both its authenticity and the identity of the subject.

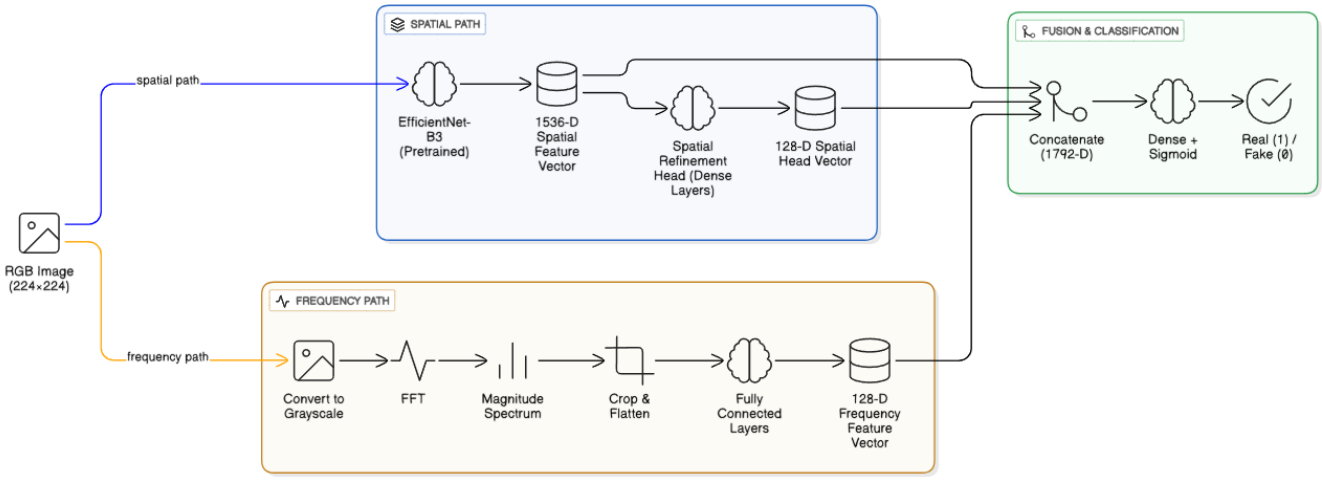

Fig. 2. Architecture of the HybridDetector model for forgery detection. It consists of a spatial path using an EfficientNet-B3 backbone and a frequency path that analyzes the FFT of the image.

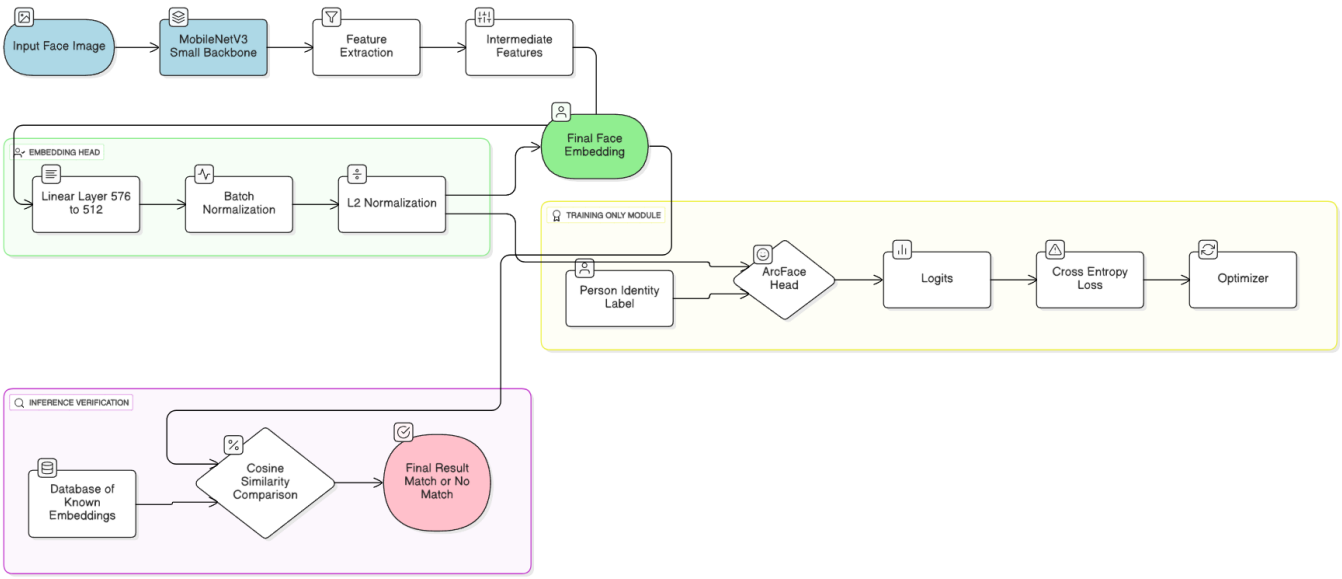

Fig. 3. Architecture of the Face Identification model. A MobileNetV3-Small backbone extracts features, which are projected into a 512-D embedding. The ArcFace head is used during training to enforce a discriminative margin.

III. Datasets and Preprocessing

A. Dataset for Forgery Detection

The forgery model was trained on a dataset of real and forged images. A CSV file mapped each image to its 'real' or 'fake' label. A custom PyTorch Dataset class handled loading, label encoding, and transformations.

B. Dataset for Identity Verification

The identification model was trained on a dataset where each person represents a unique class. The dataset was split by identity: 80% for training (80 people) and 20% for validation (20 people), ensuring generalization testing on unseen individuals. Augmentations included random resized cropping, color jitter, and horizontal flipping.

IV. Experimental Setup

All models were implemented using PyTorch and trained on NVIDIA T4 GPUs. The specific training parameters for each model are detailed below.

A. Forgery Detection Setup

The HybridDetector model was trained for 8 epochs using the parameters summarized in Table I.

B. Identity Verification Setup

The face identification model was trained for 20 epochs. The training protocol and hyperparameters are detailed in Table II.

TABLE I
Training Parameters for Forgery Detection

Parameter	Value
Model Backbone	EfficientNet-B3
Input Image Size	224x224 pixels
Batch Size	16
Epochs	8
Loss Function	Binary Cross-Entropy (BCELoss)
Optimizer	Adam
Learning Rate	2e-4
Weight Decay	1e-5
LR Scheduler	StepLR (factor 0.6 every 3 epochs)

TABLE II
Training Parameters for Identity Verification

Parameter	Value
Model Backbone	MobileNetV3-Small
Input Image Size	160x160 pixels
Embedding Dimension	512
Batch Size	128
Epochs	20
Loss Function	Cross-Entropy Loss
Optimizer	AdamW
Learning Rate	3e-4
Weight Decay	1e-4
LR Scheduler	Cosine Annealing
ArcFace Margin (m)	0.5
ArcFace Scale (s)	30.0

V. Results and Analysis

A. Forgery Detection Results

The HybridDetector converged quickly to a low error rate. The model weights yielding the highest validation AUC were

saved. As shown in Table III, the final model achieved a validation accuracy of 99.65%, a validation AUC of 0.9999, and an Equal Error Rate (EER) of 0.0040.

TABLE III
FORGERY DETECTION PERFORMANCE ON THE VALIDATION SET

Epoch	Val Loss	Val Acc.	Val AUC	EER
1	0.0454	0.9858	0.9985	0.0144
2	0.0284	0.9898	0.9996	0.0095
4	0.0160	0.9948	0.9998	0.0050
7	0.0123	0.9952	0.9999	0.0045
8	**0.0139**	**0.9965**	**0.9999**	**0.0040**

B. Identity Verification Results

The face identification model was evaluated on the unseen validation set after each epoch. The best performance was achieved at Epoch 8, as detailed in Table IV. The model achieved a validation AUC of 0.880, EER of 0.190, and a best F1-score of 0.813, demonstrating its ability to generalize to new identities.

TABLE IV
IDENTITY VERIFICATION PERFORMANCE ON THE VALIDATION SET

Epoch	Val AUC	Val EER	F1@best	Threshold
1	0.7261	0.3237	0.6990	0.489
4	0.8578	0.2200	0.7939	0.228
7	0.8766	0.1913	0.8099	0.178
8	**0.8796**	**0.1900**	**0.8130**	**0.178**
9	0.8745	0.1825	0.8182	0.153
12	0.8707	0.2062	0.8043	0.108
20	0.8623	0.2037	0.7985	0.093

VI. CONCLUSION AND FUTURE WORK

In this work, we presented a unified system for eKYC that effectively combines two specialized models: a hybrid spatial-frequency model for image forgery detection and a lightweight face identification model using a MobileNetV3-Small backbone with ArcFace loss. Our results show that this dual-path approach achieves excellent performance on both tasks. The forgery detector is highly accurate at identifying manipulated images, while the face embedder demonstrates a strong ability to generalize to unseen identities.

Future work could involve training on larger, more diverse datasets to further improve generalization, exploring different efficient backbones, and deploying the unified system in a real-world application to test its performance and latency.

REFERENCES

[1] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.

[2] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," University of Massachusetts, Amherst, Technical Report 07-49, 2007.

[3] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1–11.