IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

Securing Net Banking Transaction with Facial Recognition-Based Verification Systems

Mohanasundram A[1], Senthilmurugan R[2], Sathish Kumar K[3], Harikrishnan P[4], Mooventhan L[5]
Assistant Professor, Computer Science and Engineering[1]
Students, Computer Science and Engineering[2,3,4,5]
Mahendra Institute of Engineering and Technology, Namakkal, India

Abstract: *Face recognition systems are increasingly used in biometric security for convenience and effectiveness. However, they remain vulnerable to spoofing attacks, where attackers use photos, videos, or masks to impersonate legitimate users. This research addresses these vulnerabilities by exploring the Vision Transformer (VIT) architecture, fine-tuned with the DINO framework utilizing Celeb A-Spoof, CASIA SURF, and a proprietary dataset. The DINO framework facilitates self-supervised learning, enabling the model to learn distinguishing features from un label data. We compared the performance of the proposed fine-tuned VIT model using the DINO framework against traditional models, including CNN Model Efficient Net b2, Efficient Net b2 (Noisy Student), and Mobile VIT on the face anti-spoofing task. Numerous tests on standard datasets show that the VIT model performs better than other models in terms of accuracy and resistance to different spoofing methods. Our model's superior performance, particularly in APCER (1.6%), the most critical metric in this domain, underscores its improved ability to detect spoofing relative to other models. Additionally, we collected our own dataset from a biometric application to validate our findings further. This study highlights the superior performance of transformer-based architecture in identifying complex spoofing cues, leading to significant advancements in biometric security*

Keywords: Face recognition

I. INTRODUCTION

A. Overview

Face recognition systems (FRS) are vital to modern security, offering efficient biometric authentication for applications like smartphone unlocking and access control. These systems are particularly effective in sensitive areas, where they can restrict unauthorized access and enhance reliability Smartphone-based FRS is also being explored, focusing on feature extraction algorithms and security challenges However, they are vulnerable to spoofing attacks, where impostors use photos, videos, or masks to mimic legitimate users and deceive the system . Even simple identity spoofing methods, such as using mobile

The associate editor coordinating the review of this manuscript and camera shots or social media photos, can compromise the security of these systems . This vulnerability requires developing strong anti-spoofing techniques to accurately differentiate between genuine and spoofed faces. .

Several recent studies have demonstrated the potential of vision transformers in face anti-spoofing . Current research addresses this problem using the Vision Transformer (VIT) architecture, fine-tuned with the DINO (Emerging Properties in Self-Supervised Vision Transformers) framework. The DINO framework facilitates self-supervised learning, enabling the model to learn distinguishing features from un label data. However, existing self-supervised approaches often struggle with generalization across diverse and unseen spoofing techniques, and their limited. We hypothesize that a transformer-based model, trained on a large and diverse dataset, can effectively capture the nuanced features indicative of spoofing. As a result, it can outperform traditional CNN models.

Face anti-spoofing poses unique challenges, especially due to the lack of label spoofing data and the constantly evolving techniques to bypass security systems. Self supervised learning, like DINO, provides a significant advantage

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

by allowing the model to learn from large amounts of un label data, reducing the need for expensive and time consuming label data. This is particularly valuable for face antispoofing, as collecting and label diverse spoofing samples poses a challenge. By using self-supervised learning, our model can better generalize across a wider range of spoofing attacks and adapt to unseen threats.

In this study, we utilized multiple benchmark datasets to evaluate the performance of our proposed Vision Transformer (VIT) model, fine-tuned using the DINO framework. Besides these established datasets, we also gathered a unique dataset from a biometric application. The contributions of this study are as follows:

- Introducing the Vision Transformer (ViT) architecture fine-tuned with the DINO (Emerging Properties in SelfSupervised Vision Transformers) framework for face anti-spoofing. While ViTs have been used in face antispoofing, integrating the DINO framework in this area has not been extensively investigated.
- Comparative Analysis of the proposed model with traditional models, including CNN Model EfficientNet b2 and EfficientNet b2 (Noisy Student), Mobile ViT.
- Improvement in anti-spoofing performance, reflected in the APCER. Our comparative analysis shows that our DINO-based ViT model significantly outperforms other models, demonstrating the ability to identify spoofing attacks better.

One of our model's main distinctions is its integration with the DINO framework, which employs self-supervised learning. This allows our model to learn from unlabeled data and generalize better across various spoofing attacks.

Although Vision Transformers have been previously applied to face anti-spoofing, the integration of the DINO framework remains underexplored in this context. Our work addresses this gap by introducing a novel approach and utilizing DINO's self-supervised learning capabilities to enhance

Model robustness against spoofing attacks .To our knowledge, this is the first application of the DINO framework in the context of face anti-spoofing. It fills a critical gap in the existing literature and offers new insights into the potential of self-supervised ViTs in biometric security.

The paper is structured as follows: Section I is this introduction. Section II presents an overview of the works related to face anti-spoofing. Next, Section III describes the methods employed in this study, including data collection, vision transformers, and the DINO framework. Experimental results are presented in Section IV, followed by a Discussion in Section V and Future Works in Section VI. Finally, the concluding remarks are drawn in Section

II. RELATED WORK

This section reviews the existing methods for face antispoofing, including traditional and deep learning-based approaches. The vulnerability of face recognition systems to spoofing attacks has been extensively studied.

Initial methods for face anti-spoofing mainly used handcrafted features and traditional machine learning techniques. For instance, some researchers used SURF - speededup robust features as a patented local feature detector and descriptor, and Fisher vector encoding is an image feature encoding and quantization technique to enhance face spoof detection. Still, these methods struggled with generalizing to new and unseen spoofing attacks. Similarly, researchers focused on smartphone-based face unlock systems, emphasizing the limitations of these traditional methods in dynamic and varied attack scenarios .

A range of other methods has been proposed for face antispoofing, including Haralick texture features , image quality assessment , patch and depth-based CNNs , and multi-feature video let aggregation . These methods have shown promising results in distinguishing between genuine and spoofed face appearances. Other approaches include general image quality assessment , color texture analysis and pulse detection from face videos , all of which have demonstrated effectiveness in detecting various types of spoofing attacks. Combining FRS with other security systems, such as RFID, has also been suggested to strengthen security .

Since the emergence of deep learning, Convolutional Neural Networks (CNNs) have become popular in face antispoofing research. Several studies have demonstrated the effectiveness of CNNs in learning features directly from data leading to improved liveness detection performance. However, these models require large, diverse datasets and often struggle with generalization to novel spoofing techniques due to their reliance on local feature extraction.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

Several recent studies have explored the use of transformer architectures in face anti-spoofing, with promising results. Studies and both achieved competitive performance using ViT transformers, with the latter introducing a relation aware mechanism. The performance was further improved by deepening the transformer network loop depth and introducing adaptive transformers for robust cross-domain face anti-spoofing, respectively . Other similar studies focused on generalizability, with the former proposing a domain-invariant vision transformer and the latter demonstrating the effectiveness of vision transformers for zero-shot face anti-spoofing . The other work presents UDG-FAS, the first Unsupervised Domain Generalization framework for Face Anti-Spoofing . This framework uses large volumes of unlabeled data to learn generalizable features, thereby improving performance in low-data scenarios for face anti-spoofing. Another study introduces FM-ViT, a transformer-based framework that outperforms existing single-modal frameworks . Adaptive vision transformers for robust few-shot cross domain face anti-Spoofing was proposed in the other recent study .The generalizability of vision transformers was further improved with the Domain-invariant Vision Transformer (DiVT) . Next, the study developed a convolutional vision transformer-based framework for robust performance against unseen domain data.

As we see, recent advancements in Vision Transformers (ViTs) offer a promising alternative. Unlike CNNs, ViTs capture global dependencies via self-attention mechanisms, potentially enhancing their ability to identify subtle, global spoofing cues. Studies have explored the application of ViTs for unseen face anti-spoofing, showcasing their potential in handling unseen attacks. Further research emphasized the effectiveness of transformers in incorporating relation-aware mechanisms for improved spoof detection.

Recent studies illustrate the relevance of handling masked face detection in real-time scenarios, which can be extended to an anti-spoofing approach, enhancing the robustness of face detection systems. A CAFFE-modified MobileNetV2 (CMNV2) model for masked face age and gender identification was proposed , achieving 96.54% accuracy by focusing on key facial areas like the eyes, forehead, and ears. Similarly, authors developed a Caffe-MobileNetV2 model for detecting masked and unmasked faces in both photos and real-time video , with an impressive accuracy of 99.64%. These studies highlight the importance of feature extraction from the periocular region and above, which aligns with challenges in facedetection and antispoofingunder occluded conditions.

Specific challenges frequently arise in face anti-spoofing research, including difficulties in generalizing across different domains and datasets, the constraints imposed by limited data, and technical obstacles related to methodologies such as anomaly detection and black-box discriminators. Crossdomain face anti-spoofing, such as the domain gap and limited data, can lead to poor generalization of models to new domains. Furthermore, the generalization capabilities of classifiers, particularly when applied to diverse databases, are often questioned, as they may not consistently perform well across different datasets.

DINO FRAMEWORK

Recent research has explored the DINO framework for visual transformers, demonstrating its effectiveness in various computer vision tasks. DINO-based models have shown remarkable performance in object detection and segmentation . The framework has been extended to improve few-shot keypoint detection. The original DINO paper highlighted the method's ability to learn rich visual representations without labels, achieving state-of-the-art results on ImageNet.

Many studies demonstrate the effectiveness of DINO in object detection and masked autoencoder domains. focuses on learning patch-level representations, which are crucial for accurate object detection. DINO's self-supervised vision transformers enable the model to learn detailed representations of image patches, improving performance in detecting and recognizing objects. Lastly, and both demonstrate how DINO's features can be effectively utilized in masked autoencoders, enabling these models to reconstruct masked image regions more efficiently. These studies demonstrate DINO's versatility and effectiveness across various computer vision applications.

The DINO framework has been explored in the context of security, particularly in adversarial attack scenarios . For example, studies have analyzed the robustness of self-supervised Vision Transformers trained with DINO against adversarial attacks, showing that these models can be more resilient than those trained through supervised learning These works have focused on evaluating the robustness of DINO in adversarial contexts and exploring defense strategies to enhance model security. However, despite these advancements, no previous studies have applied DINO

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

specifically to face anti-spoofing. Our research addresses this gap by employing the DINO framework to enhance the performance of Vision Transformers in detecting spoofing attacks, thereby contributing a novel application of DINO in the domain of biometric security. By doing so, we demonstrate the potential of self-supervised learning frameworks like DINO to improve real-world security applications, particularly in face anti-spoofing significantly.

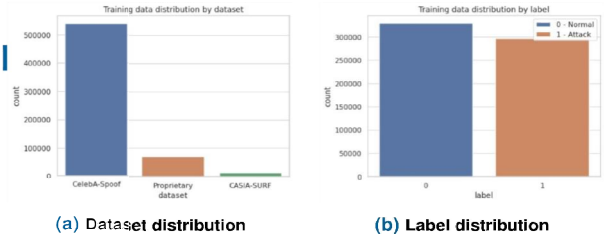
So, unlike traditional supervised approaches that rely heavily on labeled datasets, DINO excels in tasks like face anti-spoofing by focusing on its ability to capture global dependencies and learn discriminative features from large amounts of unlabeled data. This leads to improved generalization to diverse spoofing attacks that may not be present in traditional training datasets. By leveraging the ViT architecture, DINO allows the model to detect subtle details indicative of spoofing, making it particularly wellsuited for this task.

III. METHODOLOGY

A. DATA

In this research, we employed several benchmark datasets to assess how well our proposed Vision Transformer (ViT) model fine-tuned with the DINO framework. These datasets are selected for their diversity and coverage of various spoofing techniques, ensuring a thorough evaluation of the model's capabilities.

The CelebA-Spoof [41] dataset is an extensive dataset created especially for face anti-spoofing tasks. It contains over 625,000 images of 10,000 subjects, incorporating



(a) Dataset distribution (b) Label distribution

FIGURE 1. Trainig data distribution by dataset and label.

various spoofing attacks, including printed photos, replayed videos, and 3D masks. The dataset's extensive range of spoofing techniques and high subject diversity make it an excellent resource for training and evaluating anti-spoofing models, ensuring they can generalize well to different types of attacks.

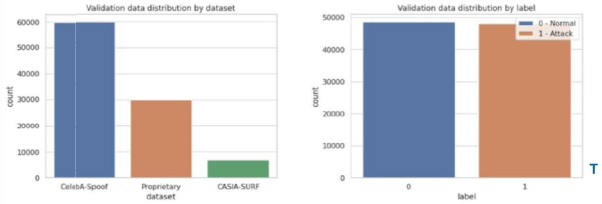
The CASIA-SURF [42], [43] dataset includes 21,000 images captured in multiple modalities: RGB, Depth, and Infrared. This multi-modal approach provides rich information that deep learning models can leverage to improve spoof detection accuracy. The dataset is particularly useful for evaluating the effectiveness of models in scenarios where different types of image data are available, improving the robustness of anti-spoofing systems.

In addition to these well-known public datasets, we used a proprietary dataset, which we collected from a biometrics application; it is owned and controlled by a company. This dataset was created during sessions flagged as suspicious and non-suspicious. During biometric authentication, subjects were often asked to turn their heads or move closer, resulting in a dataset of 100,000 images. Each subject underwent multiple biometric sessions, providing diverse images under various conditions. These images are unlabeled. Due to privacy concerns and the sensitive nature of the biometric data, this dataset cannot be publicly disclosed. We aim to train a Vision Transformer (ViT) on this unlabeled data using a self-supervised learning approach.

So, the dataset used in this study consists of images from three sources: CelebA-Spoof, a proprietary dataset, and CASIA-SURF. The training data distribution, as depicted in the first set of plots (see Fig 1), shows that the majority of the data comes from the CelebA-Spoof dataset with 543,424 images, followed by the proprietary dataset with 69,234 images, and CASIA-SURF with 14,879 images (Table 1). For the validation data, the distribution is similar, with 59,762 images from CelebA-Spoof, 29,856 images from the proprietary dataset, and 6,892 images from CASIASURF. (see Fig 2) These distributions highlight the reliance on the CelebA-Spoof dataset for training and validation, supplemented by the proprietary and CASIA-SURF datasets to provide diverse images for evaluating the model's

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

performance across different sources. This diverse dataset composition ensures the robustness and generalizability of the face anti-spoofing models developed in this study.



(a) Dataset distribution (b) Label distribution

FIGURE 2. Validation data distribution by dataset and label.

TABLE 1. Distribution of data in train and validation sets.

Dataset	Split	Total
CelebA-Spoof	Train	543424
CelebA-Spoof	Validation	59762
Proprietary	Train	69234
Proprietary	Validation	29856
CASIA-SURF	Train	14879
CASIA-SURF	Validation	6892

TABLE 2. Distribution of labels in train and validation sets.

Split	Label 0 (Normal)	Label 1 (Attack)
Train	329850	297687
Validation	48520	47990

The label distribution (Table 2) also indicates a balanced representation of normal and attack labels in both training and validation sets, which is essential for accurate model training and evaluation.

Fig. 3 shows a sample of images from the dataset used in this study. The dataset includes a wide range of face images, both genuine and spoofed, to train and evaluate the face antispoofing models. The images in the sample illustrate various spoofing techniques, such as printed photos (images 5-7), screen displays (images 3, 9-10), and genuine face images. Each image is labeled as ‘‘live’’ or ‘‘spoof,’’ highlighting the ground truth for training and validation purposes.

B. VISION TRANSFORMER (VIT)

Vision Transformers significantly impacted the field of computer vision [44]. ViT architecture treats an image as a sequence of patches, similar to how words are treated in text processing using Transformers [45]. Each image is split into a grid of non-overlapping patches, then linearly embedded and provided with positional embeddings. These embeddings go through a standard Transformer encoder, which uses multi-head self-attention mechanisms to understand the connections between different patches (see Fig. 4).

The self-attention mechanism in Transformers can be defined as:

$$\text{Attention}(Q,K,V) = \text{soft max} \sqrt{\frac{QKT}{dk}} V \quad (1)$$

where Q (queries), K (keys), and V (values) are derived from the input embeddings, and dkis the dimension of the keys.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

FIGURE 3. Sample images from the dataset illustrating various genuine (‘‘live’’) and fake (‘‘fake’’) examples. The dataset includes various facial images, including spoofing techniques such as printed and screen images.

Mobile ViT [46] is an effective neural network architecture, merging the capabilities of Vision Transformers (ViTs) with Convolutional Neural Networks (CNNs). Mobile ViT's hybrid design enables it to capture both global and local image features that are important for the face antispoofing domain.

C. DINO (DISTILLATION WITH NO LABELS)

DINO is a self-supervised learning approach that trains the model to generate similar embeddings for different views of the same image [35]. This is done using a student-teacher training setup, where the student network learns to imitate the output of the teacher network. Architecture is shown in Fig. 5.

• Teacher Network: A fixed pre-trained network that provides stable target representations.

• Student Network: A trainable network that learns to predict the teacher's representations.

The DINO framework helps the ViT model learn discriminative features from large amounts of unlabeled data. This is particularly useful for tasks like face antispoofing, where labeled data may be limited. It will help the model train on our data without labels.

D. EFFICIENTNET B2

EfficientNet b2 is a CNN model optimized for both efficiency and performance [47]. It uses a compound scaling method that proportionally increases the network's width, depth, and resolution, resulting in improved accuracy with fewer parameters and reduced computational cost. To enhance its performance further, we employed the noisy student [48] training approach, which iteratively trains the model on our custom unlabeled dataset, leveraging self training with noise to improve robustness and accuracy. For training, we utilized the CelebA-Spoof and CASIA-SURF datasets. Additionally, our proprietary dataset, consisting of 100,000 images, was incorporated into the training process using the noisy student approach, enhancing the model's ability to generalize across different spoofing scenarios.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

E. PROPOSED APPROACH

1) EXPERIMENTAL SETTINGS

To tackle the issue of face anti-spoofing, we fine-tuned a Vision Transformer (ViT) model using the DINO framework. Our approach leverages ViTs' ability to capture global dependencies in the input data via self-attention mechanisms, which enhances their ability to detect subtle, global spoofing cues. Though Vision Transformers have been applied to face anti-spoofing in prior research, incorporating the DINO framework within this context has received limited attention. We compared how well the ViT model performed against traditional models, including CNN Model Efficien tNet b2, Efficient Net b2 (Noisy Student), and Mobile ViT, to see how effective transformer-based methods are in this field. Our models were trained on two NVIDIA A100 40 GB GPUs.

The detailed training procedure is outlined in Algorithm 1.

We selected the Adam optimizer [49] due to its ability to adapt the learning rate for each parameter automatically, and it is effective because, in deep learning problems, the loss function landscape can be extremely non-convex. It is particularly suitable for deep models such as Vision Transformers.

Focal Loss [50] was used to handle the issue of class imbalance, a frequent challenge in face anti-spoofing tasks. It reduces the impact of easy-to-classify examples, enabling the model to concentrate more effectively on complex cases, such as identifying spoofed faces.

Using fp16 half precision enables faster training and reduces memory usage, especially when dealing with large models or datasets. This approach also allows for larger batch sizes, speeding up the training process on GPUs with limited memory. It helps us to speed up the training process and handle limitations on our GPUs.

The One Cycle LR scheduler [51] modifies the learning rate throughout the training process by initially setting it low, gradually increasing it to a peak, and then reducing it. his helps the model to converge faster and perform better by enabling it to explore a range of learning rates during training. It showed better convergence compared to other we use daugmentations such as Channel Shuffle, schedulers.

2) DISTINGUISHED FEATURES

During the training process, various data augmentation techniques were used to enhance the robustness and generalizability Of the face anti-Spoofing models. The Visulazation

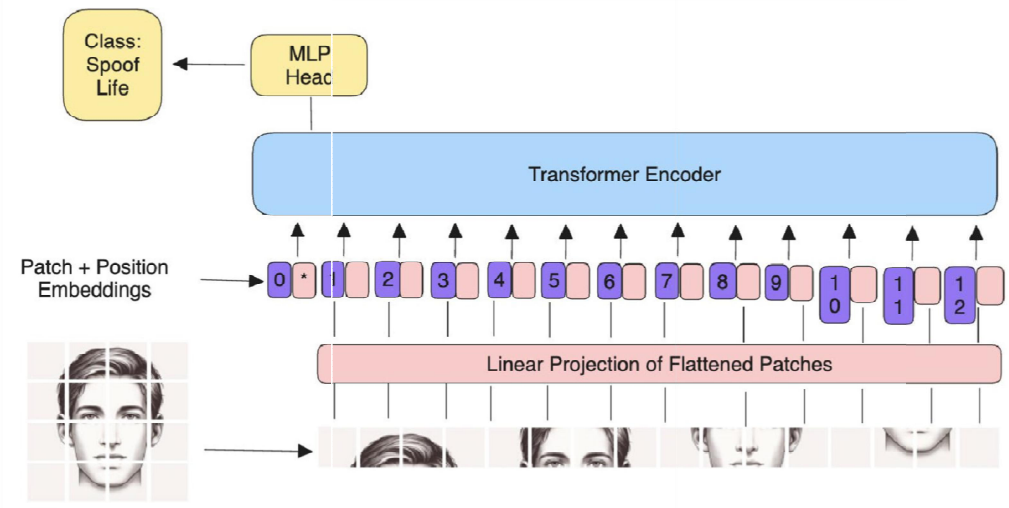


FIGURE 4. The input face image is split into patches, which are then projected linearly and embedded with positional information. These embeddings go into the Transformer encoder, which processes the sequence of patches. Next, the encoder's output is passed through a multilayer perceptron (MLP) head to classify the image as either ''spoof'' or ''live.'' .

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

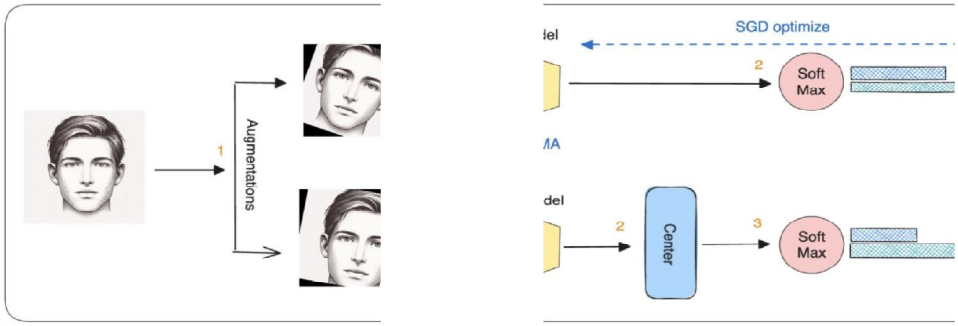


FIGURE 5. This figure illustrates the DINO (Distillation with No Labels) model training process. It starts with image augmentations (1), where two augmented views of the same image are generated. The student model processes one view, while the teacher model processes the other (2). The teacher model's outputs are centered and passed through a softmax layer (3). The student's outputs are optimized using Stochastic Gradient Descent (SGD) to match the teacher's outputs via an exponential moving average (EMA) update(4), minimizing the cross-entropy loss between the student's and teacher's redictions.

of augmentations can be seen in Fig. 6. These augmentations were categorized into four main groups:

1) Color Transformations. To provide color variations and simulate different lighting conditions, scenarios, we used augmentations such as Image Compression and a combination of blurring techniques such as Blur with a blur limit of 3 to 7, Motion Blur with a blur limit of 7 to 21, and Gauss Noise for variability in noise levels.
2) Affine Transformations .We used augmentations Such as Rotate And Flip to Provide Geometric Variations and Enhance the Model ability to generative across different orientation and perspective
3) Quality Degradations. To simulate various image visualization quality issues that might be encountered in real-world LR with One Cycle LR scheduler; end end return Trained model M
4) Cropping and Padding. To alter the spatial composition of the images, we used Crop And Pad with a percentage range of -10% to +23% which randomly crops and pads the images, ensuring the model can handle partial occlusions and varying framing conditions.

The steps of the training Algorithm are as follows:

1) Data Preparation.Split images into patches and create patch embeddings with positional encodings.

Algorithm 1 Training DINOv2 for Liveness and
AntiSpoofing Classification
Input: Dataset D (CelebA-Spoof, CASIA-SURF,
Proprietary), Image Size 224 × 224, Patch Size
14 × 14
Initialization: DINOv2 model M with pre-trained weights, with Batch Size B = 4, Learning Rate LR
0.001;
Output: ViT model DINOv2 for epoch = 1 to 300 do
for each batch B in D do Resize images in B to 224 × 224; Apply augmentations to images;
Forward pass through M with half-precision (fp16);

2) Self-Supervised Pre-training. Use the DINO framework to pre-train the ViT model on a large dataset of unlabeled facial images.

3) Fine-tuning. Replace the decoder with a binary classification layer and fine-tune the model on labeled face anti-spoofing datasets.

4) Evaluation. Compare the performance of the ViT model with EfficientNet b2 using standard metrics.

See Fig. 7 and Algorithm 1 for the detailed training algorithm steps. The algorithm initializes the DINOv2 model with pre-trained weights and a batch size of 4, using 224×224 image inputs with a 14 × 14 patch size. Over 300 epochs, images are augmented, passed through the model in halfprecision, and trained using FocalLoss and the Adam optimizer with a OneCycleLR learning rate scheduler.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the models, we used standard metrics in face anti-spoofing, including APCER, BPCER, ACER [52], and accuracy. We express APCER and BPCER in terms of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN)

APCER (Attack Presentation Classification Error Rate): This is the rate of attack presentations (spoof attempts) incorrectly classified as bona fide (genuine) presentations.

$$APCER = FP/FP + TN \quad (2)$$

BPCER (Bona Fide Presentation Classification Error Rate): This is the rate of bona fide presentations incorrectly classified as attack presentations.

$$BPCER = FN/FN + TP \quad (3)$$

ACER (Average Classification Error Rate): This is the mean of APCER and BPCER, providing a single metric to evaluate the model's overall performance.

$$ACER = APCER + BPCER/2 \quad (4)$$

In face anti-spoofing systems, APCER and BPCER present a trade-off Fig. 8. Minimizing APCER (reducing false acceptance of spoofs) can increase BPCER (false rejection of genuine attempts) and vice versa. Balancing these rates is crucial for effective performance.

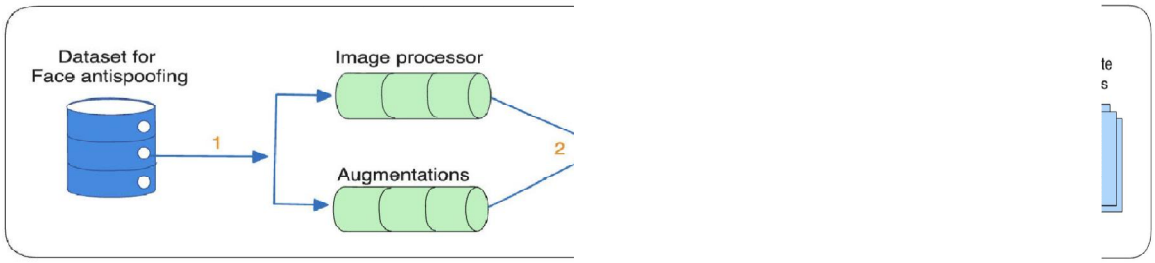
For our comparison experiment, we used four models: EfficientNet b2 and the same model, but enhanced with the Noisy Student technique, MobileViT v2, and ViT (DINO). EfficientNet b2 was selected for its strong baseline performance and efficiency. The Noisy Student version of EfficientNet b2 was included to explore the impact of semisupervised learning on model robustness. MobileViT v2 was chosen for its balance of efficiency and performance. Finally, ViT (DINO) was included as the primary model in our research, focusing on its ability to leverage self-supervised learning through a transformer-based architecture.

The performance metrics for considered models are summarized in Table 3. The results demonstrate that the ViT (DINO) model significantly outperforms the other models performance on different datasets.

TABLE 3. Comparison of EfficientNet and ViT (DINO) Models (all datasets combined). across all evaluation metrics. Table 4 compares the model's

The results, summarized in Tables 3, 4, demonstrate that the ViT (DINO) model consistently outperforms the other models across all evaluation metrics. For instance, ViT (DINO) achieves the lowest APCER (1.6%) compared to 22.5% for EfficientNet b2 and 5.5% for MobileViT v2. Similarly, BPCER for ViT (DINO) is minimal at 0.1%, outperforming the other models. The ACER metric further confirms ViT (DINO)'s superior balance in handling both attack and bona fide presentations, with a score of 0.8% compared to 11.75% for EfficientNet b2 and 2.98% for MobileViT v2. Moreover, ViT (DINO) consistently delivers the highest overall accuracy, reaching 99.8%, underscoring its excellent capability in distinguishing genuine faces from spoofed ones across various datasets.

The enhancement in anti-spoofing efficacy is evident in the APCER metric. Our comparative analysis reveals that our DINO-based ViT model greatly surpasses the EfficientNet B2 model in performance. Notably, it achieves an APCER of 1.6%, markedly lower than the 22.5% by the EfficientNet model. This substantial improvement underscores our model's enhanced capability to detect spoofing attacks.



DOI: 10.48175/IJARSCT-26387

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

Metric	EfficientNet b2	EfficientN
APCER	**22.5**	
BPCER	0.95	
ACER	11.75	
Accuracy (%)	98.2	

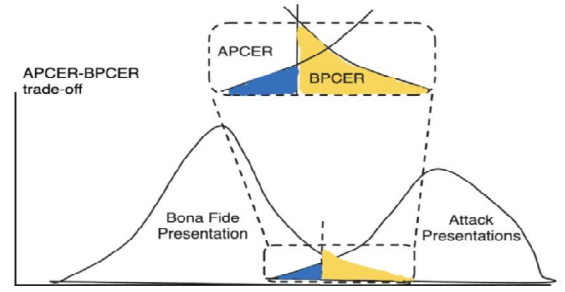


FIGURE 8. Balancing the Attack Presentation Classification Error Rate (APCER) and the Bona Fide Presentation Classification Error Rate (BPCER) is important. The overlapping areas show misclassifications: APCER (blue) represents attack presentations wrongly classified as real, and BPCER (yellow) shows real presentations wrongly classified as attacks. Finding the right balance between these two metrics is key to improving the performance of face anti-spoofing systems.

Our model's strong performance in APCER demonstrates its superior ability to detect spoofing over other models, as it is the most critical metric, even if others perform slightly better. APCER and BPCER are more crucial than overall accuracy in face anti-spoofing because they directly measure how well a system handles security threats. APCER shows how often the system mistakenly accepts spoof attacks as genuine, which is critical since even a few errors can lead to serious security breaches. BPCER, on the other hand, indicates how often genuine users are wrongly denied access, which can cause significant frustration. Since datasets in this field are often imbalanced, accuracy alone can be misleading–it might appear high even if the model fails to detect spoofing effectively. This is why standards like ISO/IEC 30107-3 focus on APCER and BPCER, as these metrics more accurately reflect the system's performance in real-world security scenarios.

Fig. 10 illustrates the trends for APCER, BPCER, ACER, and accuracy over 50 training epochs for all models. The plot demonstrates a significant decrease in APCER for all models, with the ViT (DINO) model consistently maintaining a lower APCER throughout the training process. The BPCER plot highlights the reduction in BPCER, where the ViT (DINO) model shows superior performance by achieving a lower BPCER than other models. The ACER plot indicates the overall classification error rates, significantly improving the ViT (DINO) model's ability to balance APCER and BPCER. The accuracy plot illustrates the higher overall accuracy of the ViT (DINO) model, indicating better general performance in distinguishing genuine and spoofed faces.

Fig. 9 presents the confusion matrices for all models. The ViT (DINO) model demonstrates superior classification performance with the lowest APCER and BPCER values, resulting in fewer false positives and false negatives. The confusion matrix for ViT (DINO) highlights its ability to accurately distinguish between genuine and spoofed faces, leading to high accuracy. MobileViT also shows strong performance with low error rates, while both EfficientNet b2 models, though achieving high accuracy, exhibit higher APCER and BPCER, reflecting a relatively higher rate of misclassification when compared to MobileViT and ViT (DINO).

V. DISCUSSION

A. WHY APCER IS SIGNIFICANTLY DECREASED?

As our experimental observations demonstrated, APCER significantly decreased after we trained the ViT model, with even greater improvements when fine-tuned using the DINO framework. The decrease in APCER reflects the model's

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

ability to more accurately distinguish between real and spoofed faces, reducing the risk of security breaches in face recognition systems. This improvement is critical because APCER directly measures the model's effectiveness in identifying spoof attacks, a key concern in biometric security applications.

The superior performance of ViT-based models can be attributed to their ability to capture global patterns and dependencies across the entire image, rather than focusing only on localized features, as is common with traditional CNN models. ViTs are particularly well-suited for face antispoofing tasks because they can detect subtle inconsistencies, such as unnatural lighting or distortions in spoofed faces. However, the DINO framework's selfsupervised pre-training further enhances the model's capability to learn discriminative features from large amounts of unlabeled data. By using this data, the DINO framework enables the ViT model to generalize better to diverse spoofing techniques that may not be present in traditional training datasets. This results in a model that is more robust against novel and complex spoofing attacks.

The attention visualizations for spoofandliveclassimages, as shown in the figures Fig. 11, reveal how the Vision Transformer (ViT) model, fine-tuned with DINO, selectively focuses on different regions of the images when making classifications. In the case of spoof class images Fig. 11b, the attention maps demonstrate that the model concentrates on areas that often exhibit unnatural artifacts or inconsistencies, such as reflections, edges, or distortions typically found in spoofing attacks. In contrast, for the live class images Fig. 11a, the attention maps show a more evenly distributed focus on natural, coherent facial features, such as skin texture, smoothness, and uniform lighting patterns. This distinction between how the model handles real and spoofed images illustrates the model's effectiveness in focusing on relevant features for classification.

In contrast, the EfficientNet B2 model, although optimized for efficiency and performance, relies on local feature extraction through convolutional layers. This localized focus may limit its ability to generalize to novel and sophisticated spoofing attacks that require a detailed understanding of the face's overall structure. Additionally, the traditional supervised learning approach used for training EfficientNet B2 may not fully exploit the potential of the available data, leading to suboptimal generalization. This limitation led us to experiment with training EfficientNet B2 using the Noisy Student method, a semi-supervised approach that uses both labeled and unlabeled data. This approach improved performance metrics, including APCER, but the results were still not as good as the self-supervised ViT model fine-tuned with DINO.

The findings of this study suggest that adopting transformer-based architectures, such as ViT, fine-tuned with self-supervised learning frameworks like DINO, or even CNN-based models enhanced with semi-supervised learning frameworks like Noisy Student, can significantly improve face anti-spoofing systems. These advancements have practical implications for improving the security and reliability of biometric authentication systems, which are vision transformers by incorporating relation-aware transformers to zero-shot anti-spoofing and data mechanisms and adaptive-avg-pooling-based attention. augmentation, respectively, achieving Next, [29] and [55] extend the application of vision

B. COMPARISON WITH RECENT STUDIES

Let's review how the current study's results compare to previous studies. Many studies have explored using vision transformers in face anti-spoofing, with promising results.Many studies demonstrate the effectiveness of these models in detecting anomalies and achieving robust performance across different domains [11], [13], [28], [53]. Studies [27] and [54] further enhance the capabilities of state-of-the-art performance. Lastly, [56] reports significant improvements in accuracy and reduced equal error rates using transformer-based models. These studies collectively highlight the potential of vision transformers in enhancing the security of face recognition systems. Our findings back up these prior research works.

VI. CONCLUSION

In this study, we presented a novel application of the DINO framework within Vision Transformers for face antispoofing. This approach addresses the limited exploration of DINO's self-supervised learning capabilities in this context. Several benchmark datasets were used to assess the effectiveness of the model.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, May 2025 Impact Factor: 7.67

The ViT (DINO) model consistently outperformed other models across all key metrics (especially in APCER), indicating its superior ability to distinguish between genuine and spoofed faces. Our comparative experiments demonstrated that the ViT (DINO) model consistently outperformed other state-of-the-art models, including EfficientNet B2, EfficientNet B2 with Noisy Student, and MobileViT, particularly in key metrics like APCER. This improvement is crucial as it addresses the growing threat of spoofing attacks in various applications, from personal device security to access control in high-security environments. The findings underscore the importance of adopting cutting-edge AI technologies to safeguard biometric systems against increasingly sophisticated spoofing techniques.

In general, study findings suggest that incorporating DINO into ViTs enhances their robustness against spoofing attacks, offering valuable insights into the potential of selfsupervised learning in biometric security. The results indicate that integrating DINO into ViTs can enhance their performance in biometric security applications. This contributes to a broader understanding of how selfsupervised learning techniques can be effectively applied in this domain.

REFERENCES

[1] E. Vazquez-Fernandez and D. Gonzalez-Jimenez, ‘‘Face recognition for authentication on mobile devices,’’ Image Vis. Comput., vol. 55, pp. 31–33, Nov. 2016, doi: 10.1016/j.imavis.2016.03.018.

[2] R. V. Petrescu, ‘‘Face recognition as a biometric application,’’ SSRN Electron. J., vol. 3, pp. 237–257, Apr. 2019, doi: 10.2139/ssrn.3417325.

[3] M. P.Nagesh,‘‘Face recognition systems,’’ Int.J.Res.Appl.Sci. Eng. Technol., vol. 11, no. 3, pp. 962–964, Mar. 2023, doi: 10.22214/ijraset.2023.49567.

[4] T. I. Dhamecha, S. Ghosh, M. Vatsa, and R. Singh, ‘‘Kernelized heterogeneity-aware cross-view face recognition,’’ Frontiers Artif. Intell., vol. 4, Jul. 2021, Art. no. 670538, doi: 10.3389/frai.2021.670538.

[5] D. A. Chowdhry, A. Hussain, M. Z. Ur Rehman, F. Ahmad, A. Ahmad, and M. Pervaiz, ‘‘Smart security system for sensitive area using face recognition,’’ in Proc. IEEE Conf. Sustain. Utilization Develop. Eng.
Technol. (CSUDET), May 2013, pp. 11–14, doi:
10.1109/CSUDET.2013.6670976.

[6] A. AbdElaziz, ‘‘A survey of smartphone-based face recognition systems for security purposes,’’ Kafrelsheikh J. Inf. Sci., vol. 2, no. 1, pp. 1–7, Aug. 2021, doi: 10.21608/kjis.2021.5484.1006.

[7] N. Erdogmus and S. Marcel, ‘‘Spoofing face recognition with 3D masks,’’ IEEE Trans. Inf. Forensics Security, vol. 9, no. 7, pp. 1084–1097, Jul. 2014.

[8] B. Hamdan and K. Mokhtar, ‘‘The detection of spoofing by 3D mask in a 2D identity recognition system,’’ Egyptian Informat. J., vol. 19, no. 2, pp. 75–82, Jul. 2018.

[9] L. Omar and I. Ivrissimtzis, ‘‘Evaluating the resilience of face recognition systems against malicious attacks,’’ in Proc. 7th U.K. Brit. Mach. Vis. Workshop, 2015, pp. 5.1–5.9.

[10] L. Omar and I. Ivrissimtzis, ‘‘Designing a facial spoofing database for processed image attacks,’’ in Proc. 7th Int. Conf. Imag. Crime Detection Prevention (ICDP), 2016, pp. 1–6.

[11] L. Abduh, L. Omar, and I. Ivrissimtzis, ‘‘Anomaly detection with transformer in face anti-spoofing,’’ J. WSCG, vol. 31, nos. 1–2, pp. 91–98, Jul. 2023.

[12] A. Liu, Z. Tan, Z. Yu, C. Zhao, J. Wan, Y. Liang, Z. Lei, D. Zhang, S. Z. Li, and G. Guo, ‘‘FM-ViT: Flexible modal vision transformers for face antispoofing,’’ IEEE Trans. Inf. Forensics Security, vol. 18, pp. 4775–4786, 2023, doi: 10.1109/TIFS.2023.3296330.

[13] C.-H. Liao, W.-C. Chen, H.-T. Liu, Y.-R. Yeh, M.-C. Hu, and C.-S. Chen, ‘‘Domain invariant vision transformer learning for face anti-spoofing,’’ in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), Jan. 2023, pp. 6087–6096.

[14] Y. Lee, Y. Kwak, and J. Shin, ‘‘Robust face anti-spoofing framework with convolutional vision transformer,’’ in Proc. IEEE Int. Conf. Image Process. (ICIP), Oct. 2023, pp. 1015–1019.

[15] Z. Boulkenafet, J. Komulainen, and A. Hadid, ‘‘Face antispoofing using speeded-up robust features and Fisher vector encoding,’’ IEEE Signal Process. Lett., vol. 24, no. 2, pp. 141–145, Feb. 2017.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

[16] K. Patel, H. Han, and A. K. Jain, ‘‘Secure face unlock: Spoof detection on smartphones,’’ IEEE Trans. Inf. Forensics Security, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.
[17] A. Agarwal, R. Singh, and M. Vatsa, ‘‘Face anti-spoofing using Haralick features,’’ in Proc. IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst.
(BTAS), Sep. 2016, pp. 1–6.
[18] E. Fourati, W. Elloumi, and A. Chetouani, ‘‘Face anti-spoofing with image quality assessment,’’ in Proc. 2nd Int. Conf. Bio-eng. Smart Technol. (BioSMART), Aug. 2017, pp. 1–4.
[19] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, ‘‘Face anti-spoofing using patch and depth-based CNNs,’’ in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), Oct. 2017, pp. 319–328.T. A. Siddiqui, S. Bharadwaj, T. I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha, ‘‘Face anti-spoofing with multifeature videolet
[20] T. A. Siddiqui, S. Bharadwaj, T. I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha, ‘‘Face anti-spoofing with multifeature videolet aggregation,’’ in Proc. 23rd Int. Conf. Pattern Recognit. (ICPR), 2016, pp. 1035–1040.
[21] J. Galbally and S. Marcel, ‘‘Face anti-spoofing based on general image quality assessment,’’ in Proc. 22nd Int. Conf. Pattern Recognit., Aug. 2014, pp. 1173–1178.
[22] Z. Boulkenafet, J. Komulainen, and A. Hadid, ‘‘Face anti-spoofing based on color texture analysis,’’ in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2015, pp. 2636–2640.
[23] X. Li, J. Komulainen, G. Zhao, P.-C. Yuen, and M. Pietikäinen,
‘‘Generalized face anti-spoofing by detecting pulse from face videos,’’ in Proc. 23rd Int. Conf. Pattern Recognit. (ICPR), Dec. 2016, pp. 4244–4249, doi: 10.1109/ICPR.2016.7900300.
[24] A. Aff, M. Awedh, and M. H. A. Alghamdi, ‘‘RFID and face recognition based security and access control system,’’ Int. J. Innov. Res. Sci., Eng.
Technol., vol. 2, no. 11, pp. 5955–5964, Jan. 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:13542387
[25] S. Garg, S. Mittal, P. Kumar, and V. Anant Athavale, ‘‘DeBNet: Multilayer deep network for liveness detection in face recognition system,’’ in Proc. 7th Int. Conf. Signal Process. Integr. Netw. (SPIN), Feb. 2020, pp. 1136–1141.
[26] S. Jafri, S. Chawan, and A. Khan, ‘‘Face recognition using deep neural network with ‘LivenessNet’,’’ in Proc. Int. Conf. Inventive Comput. Technol. (ICICT), 2020, pp. 145–148.
[27] Z. Wang, Q. Wang, W. Deng, and G. Guo, ‘‘Face anti-spoofing using transformers with relation-aware mechanism,’’ IEEE Trans. Biometrics, Behav., Identity Sci., vol. 4, no. 3, pp. 439–450, Jul. 2022.
[28] H.-P. Huang, D. Sun, Y. Liu, W.-S. Chu, T. Xiao, J. Yuan, H. Adam, and M.-H. Yang, ‘‘Adaptive transformers for robust few-shot crossdomain face anti-spoofing,’’ in Proc. Eur. Conf. Comput. Vis., Jan. 2022, pp. 37–54.
[29] A. George and S. Marcel, ‘‘On the effectiveness of vision transformers for zero-shot face anti-spoofing,’’ in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), Aug. 2021, pp. 1–8.
[30] Y. Liu, Y. Chen, M. Gou, C.-T. Huang, Y. Wang, W. Dai, and H. Xiong, ‘‘Towards unsupervised domain generalization for face antispoofing,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2023, pp. 20597–20607.
[31] B. A. Kumar and M. Bansal, ‘‘Face mask detection on photo and realtime video images using caffe-MobileNetV2 transfer learning,’’ Appl. Sci., vol. 13, no. 2, p. 935, Jan. 2023, doi: 10.3390/app13020935.
[32] B. A. Kumar and N. K. Misra, ‘‘Masked face age and gender identification using caffe-modified MobileNetV2 on photo and real-time video images by transfer learning and deep learning techniques,’’ Expert Syst. Appl., vol. 246, Jul. 2024, Art. no. 123179. [Online]. Available:
https://www.sciencedirect.com/science/article/pii/S0957417424000447
[33] F. Li, H. Zhang, H. Xu, S. Liu, L. Zhang, L. M. Ni, and H.-Y. Shum, ‘‘Mask DINO: Towards a unified transformer-based framework for object detection and segmentation,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 3041–3050.
[34] C. Lu, H. Zhu, and P. Koniusz, ‘‘From saliency to DINO: Saliencyguided vision transformer for few-shot keypoint detection,’’ 2023, arXiv:2304.03140.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, May 2025 Impact Factor: 7.67

[35] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, and A. Joulin, ''Emerging properties in self-supervised vision transformers,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 9630–9640.
[36] S. Yun, H. Lee, J. Kim, and J. Shin, ''Patch-level representation learning for self-supervised vision transformers,'' 2022, arXiv:2206.07990.
[37] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, ''Masked autoencoders are scalable vision learners,'' 2021, arXiv:2111.06377. [38] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. So Kweon, and S. Xie, ''ConvNeXt v2: Co-designing and scaling ConvNets with maskedautoencoders,'' 2023, arXiv:2301.00808.
[39] N. Inkawhich, G. McDonald, and R. Luley, ''Adversarial attacks on foundational vision models,'' 2023, arXiv:2308.14597.
[40] N. Inkawhich, G. McDonald, and R. Luley, ''Adversarial attacks on foundational vision models,'' 2023, arXiv:2308.14597.
[41] J. Rando, N. Naimi, T. Baumann, and M. Mathys, ''Exploring adversarial attacks and defenses in vision transformers trained with DINO,'' 2022, arXiv:2206.06761.
[42] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, ''Celebaspoof: Large-scale face anti-spoofing dataset with rich annotations,'' in Proc. Eur. Conf. Comput. Vis., pp. 70–85, 2020.
[43] S. Zhang, A. Liu, J. Wan, Y. Liang, G. Guo, S. Escalera, H. J. Escalante, and S. Z. Li, ''CASIA-SURF: A large-scale multi-modal benchmark for face anti-spoofing,'' IEEE Trans. Biometrics, Behav., Identity Sci., vol. 2, no. 2, pp. 182–193, Apr. 2020.
[44] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. Z. Li, ''A dataset and benchmark for large-scale multi-modal face anti-spoofing,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 919–928.
[45] N. Ilinykh and S. Dobnik, ''What does a language-and-vision transformer see: The impact of semantic information on visual representations,'' Frontiers Artif. Intell., vol. 4, Dec. 2021, Art. no. 767971, doi:10.3389/frai.2021.767971.
[46] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly,J. Uszkoreit, and N. Houlsby, ''An image is worth 16×16 words:
Transformers for image recognition at scale,'' 2020, arXiv:2010.11929.
[47] S. Mehta and M. Rastegari, ''Separable self-attention for mobile vision transformers,'' 2022, arXiv:2206.02680.
[48] M. Tan and Q. V. Le, ''Efficientnet: Rethinking model scaling for convolutional neural networks,'' in Proc. 36th Int. Conf. Mach. Learn., vol. 97, 2020, pp. 6105–6114.
[49] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, ''Self-training with noisy student improves imagenet classification,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 10687–10698.
[50] D. P. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, arXiv:1412.6980.
[51] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, ''Focal loss for dense object detection,'' 2017, arXiv:1708.02002.
[52] L. N. Smith and N. Topin, ''Super-convergence: Very fast training of neural networks using large learning rates,'' 2017, arXiv:1708.07120.
[53] Information Technology—Biometric Presentation Attack Detection Part 3: Testing and Reporting, Standard ISO/IEC 30107-3:2023, Int. Org. for Standardization, 2023. [Online]. Available: https://www.iso.org/standard/79520.html
[54] M. Marais, D. Brown, J. Connan, and A. Boby, ''Facial liveness and antispoofing detection using vision transformers,'' in Proc. Southern Afr. Telecommun. Netw. Appl. Conf. (SATNAC), Aug. 2023, pp. 1–6.
[55] J. Yang, F. Chen, R. K. Das, Z. Zhu, and S. Zhang, ''Adaptive-avgpooling based attention vision transformer for face anti-spoofing,'' in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2024, pp. 3875–3879, doi: 10.1109/ICASSP48485.2024.10446940.
[56] J. Orfao and D. van der Haar, ''Keyframe and GAN-based data augmentation for face anti-spoofing,'' in Proc. 12th Int. Conf. Pattern Recognit. Appl. Methods,2023,pp.629–640doi:10.5220/0011648400003411.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

[57] K. Watanabe, K. Ito, and T. Aoki, ‘‘Spoofing attack detection in face recognition system using vision transformer with patch-wise data augmentation,’’ in Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu.Summit Conf. (APSIPA ASC), Nov.2022,pp.1561–1565,doi: 10.23919/APSIPAASC55919.2022.9979996.
[58] Q. Fan, Q. You, X. Han, Y. Liu, Y. Tao, H. Huang, R. He, and H. Yang,‘‘Vitar: Vision transformer with any resolution,’’ 2024, arXiv:2403.18361.
[59] P. Kozlov, A. Akram, and P. Shamoi, ‘‘Fuzzy approach for audiovideo emotion recognition in computer games for children,’’ Proc. Comput. Sci.,vol. 231, pp. 771–778, Jan. 2024, doi: 10.1016/J.PROCS. 2023.12.139.