A Classroom Attendance System with Anti-Spoofing Based on ResNet101 and ArcFace for Face Recognition

Wang Yunxi
Intelligence Science and Technology, Xidian University, Xi'an, China

Abstract: Addressing the inefficiencies and susceptibility to proxy check-ins inherent in traditional classroom attendance methods, alongside the performance degradation of existing face recognition systems under complex classroom lighting conditions, this paper proposes an anti-spoofing classroom attendance system based on an improved ResNet101 architecture and the ArcFace loss function. Our approach enhances the model's ability to extract illumination-invariant features by embedding Convolutional Block Attention Modules (CBAM) and optimizes feature space distribution through a joint loss function strategy. To tackle the challenge of extreme classroom illumination, a dedicated face dataset encompassing varied lighting conditions was constructed for model fine-tuning. Experimental results demonstrate that the system obtains a recognition accuracy of 99.68% on the public LFW dataset and a significant improvement from 76.3% to 88.7% on our proprietary extreme illumination test set, compared to the baseline model. Integrated with an active liveness detection mechanism, the system successfully defended against all photo and video replay attacks. Coupled with a developed web management platform, this system realizes an efficient, reliable, and non-contact automated attendance solution for teaching practice, providing a key technological framework for the development of smart classrooms.

Keywords: Face Recognition; Classroom Attendance; ResNet101; ArcFace; Attention Mechanism; Illumination Robustness

1. Introduction

Classroom attendance is a crucial component of teaching management and quality assurance. Traditional methods like manual roll-calling are time-consuming and labor-intensive, while systems based on IC cards, QR codes, or fingerprints suffer from issues such as ease of proxy check-ins, inconvenient contact-based operation, and hygiene concerns. In recent years, non-contact attendance systems based on face recognition have garnered significant attention due to their convenience and inherent security. However, deploying these systems directly in authentic classroom environments presents several formidable challenges: (1) The complex and often harsh lighting conditions within classrooms, particularly strong overhead or side lighting, can cause facial feature overexposure or cast heavy shadows, severely impeding robust feature extraction. (2) The requirement for simultaneous check-in for large numbers of students demands high concurrency processing capability and real-time response speed from the system. (3) It is imperative to effectively prevent identity spoofing using media such as printed photos or digital videos.

Although the rapid advancement of Deep Convolutional Neural Networks (CNNs) has pushed face recognition accuracy under constrained conditions to near-human levels, performance in unconstrained real-world scenarios-particularly robustness to covariate changes like illumination and pose-remains a significant research challenge. Mainstream models like FaceNet, which utilize architectures like Inception-ResNet with triplet loss, still exhibit potential for improvement in terms of training stability and feature discriminability. In recent years, margin-based loss functions, notably ArcFace, have emerged as state-of-the-art by incorporating additive angular margins into the angular space, thereby significantly enhancing intra-class compactness and inter-class discrepancy.

This research aims to develop a highly robust, anti-spoofing attendance system, deeply optimized for the classroom context. The core contributions of this work are fourfold:

We propose a modified ResNet101 network structure that integrates Convolutional Block Attention Modules (CBAM), guiding the model to focus on discriminative facial regions that are

relatively stable under varying illumination, such as the eyebrows and periocular areas.

We adopt a joint optimization strategy combining the ArcFace loss with Center Loss, further constraining the feature space distribution and enhancing the model's generalization capability under complex lighting.

We construct a realistic classroom illumination face dataset, providing critical data support for model fine-tuning and performance evaluation specific to the target environment.

We integrate an active liveness detection method and a web-based data management platform, resulting in a complete, functional, and secure system prototype, the efficacy of which is validated in a real classroom setting.

2. Related Work

2.1. Evolution of Deep Face Recognition Technology

The development of face recognition technology has closely revolved around two main threads: feature learning and metric learning. In feature learning, network architectures have evolved from shallow networks like AlexNet and VGG to deeper networks such as ResNet and Inception-ResNet, which incorporate residual connections and parallel structures. These advancements effectively mitigate the vanishing gradient problem and enrich feature hierarchies. Among them, the ResNet proposed by He et al.[1], due to its exceptional feature representation capability, has become the backbone network for many advanced face recognition systems.

In metric learning, the design of loss functions has progressed from the standard Softmax to a series of margin-based losses. Wen proposed Center Loss, which learns a center for each class and penalizes the distances between deep features and their corresponding class centers, thereby enhancing intra-class compactness. Subsequently, SphereFace, CosFace[2], and ArcFace[3] were introduced, incorporating multiplicative or additive angular/cosine margins. The ArcFace loss, proposed by Deng et al. [3], has achieved leading performance on multiple benchmarks due to its clear geometric interpretation on a hypersphere and stable training process. Its formulation is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\left\{s\left(\cos\left(\theta_{\{y_i\}}+m\right)\right)\right\}}}{e^{\left\{s\left(\cos\left(\theta_{\{y_i\}}+m\right)\right)\right\}}+\sum_{j=1,j\neq y_i}^{n} e^{\left\{s\cos\theta_j\right\}}} \quad (1)$$

Where N is the batch size, n is the number of classes, s is a feature scale factor, m is the additive angular margin, and $\theta_{\{y_i\}}$ is the angle between the feature and the weight vector of its ground-truth class.

2.2 Face Recognition under Challenging Conditions

Researchers have proposed various solutions to address challenges such as illumination, pose, and occlusion. For illumination problems, beyond traditional preprocessing methods like histogram equalization, deep learning approaches often employ data augmentation or integrate attention mechanisms to enhance robustness. The Convolutional Block Attention Module (CBAM) proposed by Woo et al.[4], which sequentially infers attention maps along both channel and spatial dimensions, helps the model focus on more informative regions. For pose variations, 3D face reconstruction and Generative Adversarial Networks (GANs) have been utilized to generate multi-pose images for data augmentation[5].

2.3 Applications and Challenges in Educational Scenarios

The application of face recognition technology in education primarily focuses on attendance management and student behavior analysis. Early systems often relied on traditional features like HOG or LBP, resulting in limited accuracy[6]. With the proliferation of deep learning, systems began adopting CNN models. However, many of these works merely apply generic models directly, lacking targeted optimization for classroom-specific challenges, such as characteristic lighting patterns and large-scale simultaneous recognition. This research addresses this gap by focusing on algorithmic innovation and system integration to solve the practical problems encountered in classroom scenarios.

3. Proposed System and Methodology

3.1. System Overview

The proposed system comprises four core modules: Face Detection and Alignment, Liveness Detection, Feature Extraction and Recognition, and the Web Data Management Platform. The operational workflow is illustrated

in Figure 1.

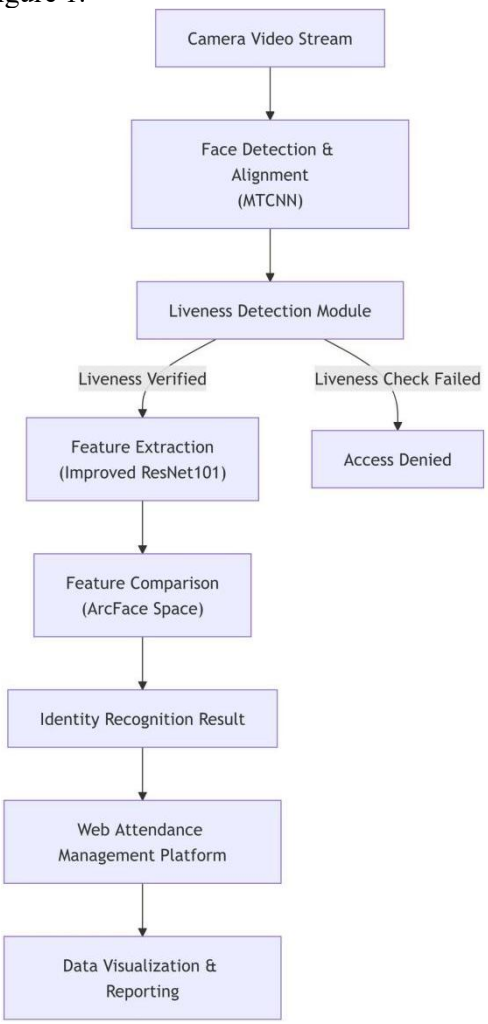

Figure 1. Overall Architecture and Workflow of the Proposed Classroom Attendance System.

3.2. Improved Feature Extraction Network

This study employs ResNet101 as the backbone network and introduces the following key modifications to enhance illumination robustness:

Integration of Attention Mechanism: The standard ResNet101 architecture was augmented by integrating a Convolutional Block Attention Module (CBAM)[4] after each residual block. CBAM sequentially derives attention maps along the channel and spatial dimensions, enabling the model to adaptively calibrate feature responses. This mechanism prioritizes discriminative facial regions that are critical for identity verification yet less susceptible to illumination changes, such as the eyes and eyebrow contours.

Feature Normalization: Before the final fully connected layer, the extracted features undergo L2 normalization. This projects the features onto a unit hypersphere, aligning with the requirement of the ArcFace loss function and facilitating angular space optimization.

Joint Loss Function: To compensate for potential insufficient intra-class constraints in ArcFace under certain scenarios, Center Loss is incorporated as an auxiliary component. The total loss function is defined as:

$$\mathcal{L}_{Total}=\mathcal{L}_{Arcface}+\lambda\mathcal{L}_{Center} \quad (2)$$

where is λ a hyperparameter balancing the two loss terms. This joint strategy aims to leverage the strong discriminative power of ArcFace coupled with the potent intra-class clustering capability of Center Loss.

3.3 Training Strategy for Illumination Robustness

A two-stage training strategy is adopted:

General Domain Pre-training: Our approach employs a transfer learning strategy, using a ResNet101 model pre-trained on MS-Celeb-1M as a starting point[7]. This stage equips the model with a powerful and generalized facial feature representation capability.

Specific Domain Fine-tuning: The pre-trained model is subsequently fine-tuned using our self-constructed Classroom Illumination Face Dataset. This dataset contains facial images captured under multiple realistic classroom lighting conditions (normal, strong overhead light, side light, back light), comprising approximately 12,000 images from 150 subjects. This stage is crucial for adapting the model to the specific characteristics of the target environment, thereby boosting its performance.

3.4 Liveness Detection and System Integration

To prevent proxy check-ins via spoofing, the system integrates an active liveness detection method. During the check-in process, the system randomly prompts the user to perform a simple action (e.g., "blink" or "turn head"). By analyzing the motion trajectory of facial landmarks across consecutive frames, the system determines the liveness of the subject. This method effectively defends against static photo attacks.

The recognition results are transmitted via a RESTful API to a web management platform developed using Django (backend) and Vue.js (frontend). This platform allows instructors to view real-time attendance status, generate

statistical reports, and manage student information efficiently.

4. Experiments and Results Analysis

4.1. Experimental Setup

Datasets:

LFW (Labeled Faces in the Wild): Used to evaluate the model's performance under standard conditions, containing 13,233 web-collected images from 5,749 individuals[8].

Classroom Illumination Dataset (Proprietary): Contains 12,000 images from 150 volunteers. The dataset was partitioned into training, validation, and test sets with a ratio of 70:10:20, respectively, under a subject-exclusive scheme. The test set includes an "Extreme Illumination" subset containing images with the most challenging lighting (e.g., severe backlighting, strong side lighting causing deep shadows).

Implementation Details: The system is implemented using the PyTorch framework and trained on an NVIDIA RTX 3080 GPU. The Adam optimizer is used with an initial learning rate of 0.001 and a batch size of 32. The ArcFace parameters are set to s=64 and m=0.5. The balancing parameter \lambda for Center Loss is set to 0.003 after empirical validation.

Evaluation Metrics: Recognition Accuracy, Equal Error Rate (EER), True Accept Rate at a specified False Accept Rate (TAR@FAR).

4.2. Results and Analysis

4.2.1. Performance comparison on mainstream benchmark

The performance of the proposed system is compared with several classical methods on the LFW dataset, as shown in Table 1.

Table 1. Recognition Accuracy Comparison on the LFW Dataset

Method	Backbone Network	Loss Function	Accuracy (%)
DeepFace[9]	Custom	CNN Softmax Loss	97.35
FaceNet[10]	Inception-ResNet-v1	Triplet Loss	99.63
SphereFace	64-layer CNN	Angular Margin Loss	99.42
Proposed	Improved ResNet101	ArcFace + Center Loss	99.68

The results indicate that the proposed system achieves performance comparable to, and slightly surpassing, state-of-the-art international methods on the standard benchmark, validating the fundamental recognition capability of the improved model.

4.2.2. Illumination robustness test

Tests were conducted on the "Extreme Illumination (the Weber Contrast between the face and the background >90%)" subset of our proprietary test set. The baseline model is the original ResNet101 trained with ArcFace loss, without the proposed CBAM attention or Center Loss. The results are presented in Table 2.

Table 2. Performance Comparison on the Extreme Illumination Subset

Model	Accuracy (%)	EER (%)	TAR@FAR =0.001 (%)
Baseline (ResNet101 + ArcFace)	76.3	8.5	45.2
+ CBAM Attention	82.1	6.2	58.7
+ CBAM + Center Loss (λ)	88.7	4.1	75.9

The results clearly demonstrate that the introduction of the CBAM attention module and the joint loss function leads to a significant and consistent performance improvement under extreme illumination. The accuracy increases from 76.3% to 88.7%, substantiating the effectiveness of our proposed method in addressing illumination challenges.

4.2.3. Ablation study

To investigate the individual contribution of each component, an ablation study was conducted on the entire Classroom Illumination test set (not just the extreme subset). The results are summarized in Table 3.

Table 3. Ablation Study on the Classroom Illumination Test Set (All Conditions)

Experimental Configuration	Accuracy (%)
ResNet101 + Softmax	95.1
+ ArcFace	98.2
+ ArcFace + CBAM	98.8
+ ArcFace + CBAM + Center Loss	99.3

The ablation study reveals that the transition from Softmax to ArcFace yields the most substantial performance gain. Subsequently, the introduction of the attention mechanism and the Center Loss provides further, incremental improvements, validating the efficacy of each proposed component.

4.2.4. Liveness detection and system performance

The liveness detection module successfully intercepted all 200 photo attacks and 50 video

replay attacks during testing, resulting in a spoof acceptance rate of 0%. In a real classroom environment simulating a cohort of 50 students, the system achieved an average processing speed of 28 frames per second, completing the synchronous check-in process in less than 10 seconds, thereby meeting the requirement for real-time operation.

5. Discussion

5.1. System Advantages and Application Value

The proposed system successfully bridges state-of-the-art deep learning algorithms with specific educational management needs. Its primary advantages lie in:

High Accuracy and Robustness: Through algorithmic innovations, the system not only excels under standard conditions but, more importantly, maintains high recognition rates in the highly challenging classroom illumination environment, addressing a critical bottleneck for practical deployment.

Strong Anti-Spoofing Capability: The integration of liveness detection fundamentally prevents proxy check-ins, ensuring the authenticity and integrity of attendance records.

Practicality and Usability: The complete system prototype, coupled with a user-friendly web interface, enables seamless integration into existing teaching management workflows, thereby enhancing operational efficiency.

This system provides a feasible technical pathway towards "unconscious check-in" for smart classrooms. The accumulated algorithmic insights and the constructed dataset also hold reference value for addressing identity verification challenges in other complex scenarios.

5.2. Limitations and Future Work

This study has certain limitations, which point towards directions for future research:

Computational Resource Dependency: The improved ResNet101 model has a large number of parameters, imposing certain GPU computing requirements on the deployment terminal. Future work will explore model lightweighting techniques such as pruning[11] and knowledge distillation[12] to adapt the system to resource-constrained edge devices.

Extreme Pose Handling: While the current model is primarily optimized for illumination, its performance can still degrade for profiles with near 90-degree yaw angles. Future plans include incorporating 3D face-assisted pose estimation and feature normalization techniques[5].

Data Privacy and Security: The secure storage and transmission of facial biometric data is paramount. Future work will investigate the application of homomorphic encryption[13] for protecting stored feature templates and explore training models within a federated learning framework to avoid centralized collection of raw data.

6. Conclusion

This paper designed and implemented an anti-spoofing classroom attendance system based on an improved ResNet101 network and the ArcFace loss function. By integrating an attention mechanism and a joint loss function, the system significantly enhanced its robustness to complex classroom lighting conditions. Exhaustive experiments demonstrated the effectiveness of each proposed component and the superior performance of the overall system. Integrated with liveness detection and a web platform, the system provides educational institutions with an efficient, reliable, and secure automated attendance solution, effectively promoting the deeper application of face recognition technology in the educational vertical domain. Future work will focus on integrating model lightweighting, large-pose handling, and privacy-preserving technologies.

References

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[2] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., ... & Liu, W. (2018). Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5265-5274).

[3] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699).

[4] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European

conference on computer vision (ECCV) (pp. 3-19).

[5] Zhao, J., Cheng, Y., Xu, Y., Xiong, L., Li, J., Zhao, F., ... & Yan, S. (2018). Towards pose invariant face recognition in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2207-2216).

[6] Zhu, Y., Lai, Z., Li, F., & Huang, Y. (2023). A laboratory attendance system based on face recognition. Internet of Things Technologies.

[7] Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In European conference on computer vision (pp. 87-102). Springer, Cham.

[8] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.

[9] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

[10] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[11] Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv preprint arXiv:1510.00149.

[12] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

[13] Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., & Wernsing, J. (2016). Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In International conference on machine learning (pp. 201-210). PMLR.

