Optimizing Face Recognition with Lightweight Architectures and Noise-Tolerant Algorithms

Adeer Khan

December 9, 2024

1 Introduction

Face recognition has become a pivotal technology in a wide array of applications. The rapid advancements in deep learning techniques have substantially enhanced both the accuracy and efficiency of face recognition models, with recent efforts focusing on lightweight architectures that are easy to deploy while effectively addressing challenges such as noisy label issues.

Our primary objective is to train a face recognition model that generalizes effectively across diverse demographic groups. To achieve this, we implement a two-step training strategy: initially training on a large-scale Caucasian dataset, followed by fine-tuning on a Korean dataset. This training process involves managing substantial label noise within the Korean dataset, which is mitigated using the TURN algorithm—a novel approach for fine-tuning pre-trained models in the presence of noisy labels [1]. The TURN algorithm employs a two-step refinement process that first performs linear probing to train a classifier while freezing the feature extractor, thereby preventing distortion from noisy labels. It then utilizes a Gaussian Mixture Model (GMM) to identify and cleanse noisy samples, followed by fine-tuning on the cleansed dataset. Notably, during the TURN process, the Generalized Cross Entropy (GCE) loss function proved more effective for both linear probing and fine-tuning, as it robustly mitigates the impact of noisy labels. This methodology ensures that the model maintains high performance despite demographic shifts and label inconsistencies.

We selected the GhostFaceNetV2 [2] model due to its superior efficiency and reduced computational complexity compared to state-of-the-art convolutional neural networks (CNNs) like ResNet18. GhostFaceNetV2 leverages Ghost modules, which utilize inexpensive linear transformations to generate additional feature maps from intrinsic features, effectively reducing feature map redundancy. Furthermore, GhostFaceNetV2 incorporates an attention mechanism to capture long-range dependencies, enhancing its ability to represent complex facial features. During the initial training phase on the Caucasian dataset, the ArcFace loss function [3] was employed due to its proven effectiveness in maximizing inter-class variance and minimizing intra-class variance in embedding spaces. Additionally, the adoption of the ADOPT [4] optimizer during the initial training phase offered superior convergence properties without the need for problem-dependent hyperparameter tuning, addressing the non-convergence issues associated with optimizers like Adam. By integrating GhostFaceNetV2 with the ADOPT optimizer and the TURN algorithm, we enhance the model's robustness and generalization capabilities, achieving high-performance face recognition under resource constraints and demographic variability.

2 Experiments

The datasets used in this study were structured as follows: the train1 dataset, consisting of 981,007 images from 2,882 identities (Caucasian males), was used for the primary training phase. The train2 dataset, containing 83,298 images of 1,230 identities (Korean mixed gender) with label noise, was prepared for fine-tuning experiments. The validation set included 3,185 images from 50 identities (mixed-gender Korean), providing a benchmark for evaluating model generalization. Additionally, the test set comprised 3,105 images, used to further assess model performance under unseen scenarios.

2.1 Experimental Setup

Train1: Initial Training Phase

The experiments for Train1 were conducted using the GhostFaceNetsV2 architecture. The model utilized the ArcFace loss function to maximize inter-class variance and minimize intra-class variance in facial embeddings. Training was performed with a batch size of 150, a learning rate of 0.001 decayed by a factor of 0.9 per epoch using the StepLR scheduler, and over 30 epochs. The ADOPT optimizer ensured robust convergence. The setup included 2,882 classes (Caucasian male identities), with a margin of 0.1, a scale of 30, and no dropout (0.0). The feature dimension of the model was set to 1,024. Data loading was optimized with 5 threads, and validation was performed every 5 epochs on a set of 50 Korean mixed-gender identities.

Train2: TURN Training Phase

For Train2, the TURN algorithm was applied to adapt the model to the noisy dataset. Since, the code for the algorithm wasn't available online, it was developed from scratch. The training process consisted of two phases: linear probing and fine-tuning.

During linear probing, the pre-trained model from Train1 was frozen, and only the classifier was trained. This phase used the Adam optimizer, a batch size of 100, a learning rate of 0.0005 decayed by a factor of 0.65 per epoch (StepLR scheduler), and lasted 30 epochs. The Generalized Cross Entropy (GCE) loss function ($q = 0.7$) was employed to address noisy labels. A threshold of 0.85 was used to select clean samples, resulting in 53,058 samples for fine-tuning.

Fine-tuning updated the entire model over 50 epochs, with Adam optimizer, which provided better convergence results. The setup included a batch size of 100, a weight decay of 0.0001, no dropout (0.0), a margin of 0.5, a scale of 30.0, and a feature dimension of 1,024.

All experiments were conducted using an NVIDIA GeForce RTX 4070 GPU, with 4–5 threads across the experiments, ensuring streamlined execution for high-cardinality datasets and noisy label scenarios.

Table 1: Design Choices Across Runs for Train1 and Train2

Parameter	**Value**
Model	ResNet18, GhostFaceNetsV2
Loss Function	ArcFace, Softmax, Triplet
Optimizer	Adam, AdamW, Adopt, SGD
Scheduler	StepLR, CosineLR
Learning Rate	0.001 (initial)
Batch Size	150, 200, 220, 330
Epochs	20, 30, 50, 100
Margin	0.1, 0.2
Scale	15, 30
Dropout	0.0, 0.2

Parameter	**Value**
Model	GhostFaceNetsV2
Initial Model	Pre-trained on Train1
Algorithm	TURN
Loss Function	Generalized Cross Entropy
Optimizer	Adam
Scheduler	StepLR
Learning Rate	0.0005, 0.001
Learning Rate Decay	0.65, 0.85
Weight Decay	0.0001, 0.0005
Batch Size	100, 150
Epochs (Probing)	10, 30
Epochs (Fine-Tuning)	20, 50
Threshold	0.85
Dropout	0.0
Scale	30.0
Margin	0.5
Selected Clean Samples	53,058 – 59,809

Table 2: Validation Equal Error Rates (EER) for Different Models and Batch Sizes

Model	**Batch Size**	**Validation EER (%)**		**Parameters (M)**
		Train 1	**Train 2**	
ResNet18	150	21.19	16.24	11.26
ResNet18	330	22.79	17.85	11.50
GhostFaceNetsV2	150	20.16	9.60	6.66
GhostFaceNetsV2	**330**	**18.33**	**8.91**	**6.11**

3 Ablation Study and Conclusion

A comprehensive ablation study highlights the following key insights:

- **Loss Functions:** ArcFace in Train1 and GCE in Train2 were crucial for achieving high inter-class variance and robustness against noisy labels.
- **Model Architectures:** GhostFaceNetsV2 outperformed ResNet18 with half parameters.
- **Training Strategy:** The TURN algorithm effectively reduced EER in Train2 by isolating clean samples through linear probing and enhancing generalization via fine-tuning with Generalized Cross Entropy loss.
- **Optimization:** The new Adopt converged well early on Train1 but overfitted on Train 2, and therefore was replaced with Adam optimizer. StepLR scheduler ensured robust convergence with both optimizers.

Conclusion: The integration of GhostFaceNetsV2 with TURN and well-optimized hyperparameters significantly improved performance. These findings underscore the importance of combining lightweight architectures with robust training strategies to achieve effective noise mitigation and superior face recognition under challenging conditions.

References

[1] Sumyeong Ahn, Sihyeon Kim, Jongwoo Ko, and Se-Young Yun. Fine tuning pre trained models for robustness under noisy labels. *arXiv preprint arXiv:2310.17668*, 2023.

[2] Mohamad Alansari, Oussama Abdul Hay, Sajid Javed, Abdulhadi Shoufan, Yahya Zweiri, and Naoufel Werghi. Ghostfacenets: Lightweight face recognition model from cheap operations. *IEEE Access*, 11:35429–35446, 2023.

[3] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5962–5979, October 2022.

[4] Shohei Taniguchi, Keno Harada, Gouki Minegishi, Yuta Oshima, Seong Cheol Jeong, Go Nagahara, Tomoshi Iiyama, Masahiro Suzuki, Yusuke Iwasawa, and Yutaka Matsuo. Adopt: Modified adam can converge with any β_2 with the optimal rate, 2024.