Technical Overview: Combining ConvNeXt and ArcFace to Classify Gender and Verify Faces

1. Overview of the Approach

To address two tasks at once, I created a dual-head deep learning architecture:
Task A: Binary gender classification
Task B: Verification of face (identity matching)

With two specialized heads—one for gender classification and one for face verification using ArcFace loss—the model uses a ConvNeXt backbone for reliable feature extraction. To optimize performance for both tasks, training is divided into three stages.

2. Pipeline for Preprocessing

Resizing:
To standardize input dimensions and adhere to pretrained backbone requirements, all images are resized to 224 x 224 pixels.

Augmentation:
I use a variety of augmentations during training, including noise, color jitter, perspective changes, rotations, and horizontal flips.
This improves robustness to variations in the real world, decreases overfitting, and diversifies the data.

Normalization:
ImageNet mean and standard deviation are used to normalize images.
Why: Promotes stability and convergence by guaranteeing that pixel values are on a uniform scale.

3. Architecture of Models

ConvNeXt Backbone Convolutional Layers:
Spatial inductive bias, hierarchical feature extraction.
Empirically, GELU Activation is a smooth, non-linear transformation that outperforms ReLU in deep networks.
Enhance productivity and capture channel-specific features with depthwise convolutions.

ConvNeXt:
Combines the advantages of ViT-inspired designs with contemporary ConvNets.

Pretrained weights enhance generalization and speed up convergence.

Phase 1: Categorization by Gender Head MLP + BatchNorm + ReLU + Dropout:
Complex patterns are captured by deep MLP.
Training is stabilized by BatchNorm.
Non-linearity is introduced by ReLU.
Dropout stops overfitting.

Gender Logits (2 classes):
The output layer generates logits for both genders.
Logits are converted to probabilities using Softmax.
CrossEntropyLoss with Label Smoothing: Enhances calibration and generalization while penalizing inaccurate predictions.

Phase 2: Learns discriminative face embeddings using ArcFace Verification Head MLP + BatchNorm + ReLU + Dropout.
512-dim Embedding:
Each face is represented in a condensed, expressive manner.
An essential component of angular margin learning is L2 normalization, which projects embeddings onto a unit hypersphere.

In angular space, ArcFace Loss (Additive Angular Margin) enforces both inter-class separation and intra-class compactness.
ArcFace:
cutting edge for facial recognition that is resistant to intra-class variance.

Phase 3: Fine-Tuning the Joint
Joint Optimization (Multi-task Loss): A weighted sum of gender and ArcFace losses is used to train both heads simultaneously.
Why: Makes use of shared features, enhances the overall usefulness of the model, and guarantees the performance of both tasks.

4. Adjusting Hyperparameters

Important Hyperparameters:
Epochs:-
Gender: 15
Confirmation: 40
Joint: 8
Justification: Selected on the basis of convergence analysis and validation performance.

Learning Rates:

Different rates for ArcFace loss, gender head, verification head, and backbone.
ExponentialLR and CosineAnnealing are schedulers for quick adaptation and smooth decay.

Batch Size:
32 (tradeoff between GPU memory and convergence)

Dropout:
0.4–0.5 (the empirically ideal range for regularization)

Label Smoothing:
0.1 (improves calibration)

ArcFace Scale/Margin:
0.5 / 64

Weight Decay:
1e-3 (to avoid overfitting)

Enhancement of Data:
Intensity and probability are adjusted for optimal validation F1/accuracy.

Tuning Procedure:
Grid search and validation set monitoring were used to adjust the hyperparameters, giving gender accuracy and F1 verification top priority. To avoid overfitting, schedulers and early stopping were employed.

5. Innovations and Justifications Staged Training:

Prior to joint optimization, each head can achieve nearly optimal performance.
Multi-task learning increases model efficiency and feature sharing.
For verification, ArcFace Loss is better than softmax or triplet loss.
For SOTA results, the modern ConvNet Backbone (ConvNeXt) combines CNNs' inductive bias with a transformer-like design.
Strongness to real-world deployment is ensured by extensive augmentation and normalization.
Using a well-designed pipeline and hyperparameter strategy, this architecture combines the most recent developments in deep learning for face analysis to provide reliable, high-accuracy performance on both gender classification and face verification.