🔍 Task A: Gender Classification

✅ **Objective:** Build a robust deep learning model to classify face images as male or female and Ensure high generalization and fairness under data imbalance and real-world variability.

⚙️ Core Approach :

Model: Swin V2 Transformer (Swin_V2_S) — a state-of-the-art Vision Transformer with strong inductive bias and global feature extraction.

Classifier Head: Modified with Dropout + Linear and ReLU for stability and regularization.

Input Size: 224×224, compatible with pretrained weights.

🌟 Key Innovations & Uniqueness :

✅ **Transformer-based architecture:** Swin V2 Outperforms CNNs in handling global facial structure and subtle gender cues.

✅ **Dynamic Data Augmentation Pipeline:** Strong yet controlled augmentations (e.g., random affine, jitter, grayscale) for domain generalization and robustness to unseen test conditions.

✅ **Fairness-aware Training:** Used **WeightedRandomSampler** to balance class representation during training without oversampling or generating synthetic data and Applied CrossEntropyLoss with **class weights and label smoothing** for soft decision boundaries.

✅ **Automatic Mixed Precision (AMP):** Boosted performance on GPU using torch.cuda.amp for faster and memory-efficient training.

✅ **Adaptive LR Scheduling:** ReduceLROnPlateau to automatically lower learning rate on validation stagnation — preventing overfitting and training collapse.

✅ **Early Stopping with Checkpointing:** Prevents overfitting by monitoring validation loss and restoring best weights.

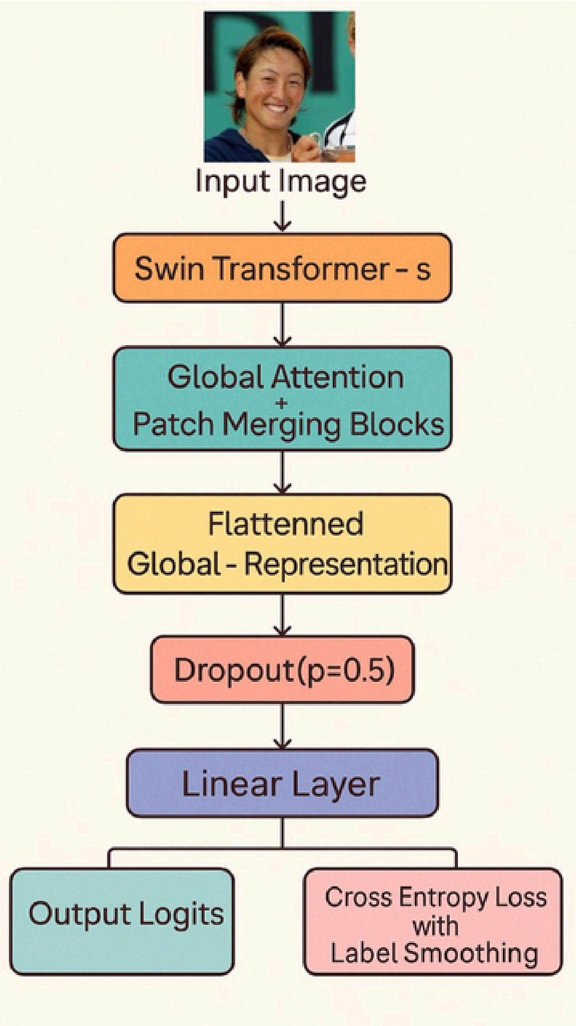

Figure 1 : diagram of the model, used in Task A

🔍 Task B: Face Verification & Matching

✅ **Goal:** Determine whether a test image (possibly distorted) belongs to the same identity as a given reference image — face verification, not classification.

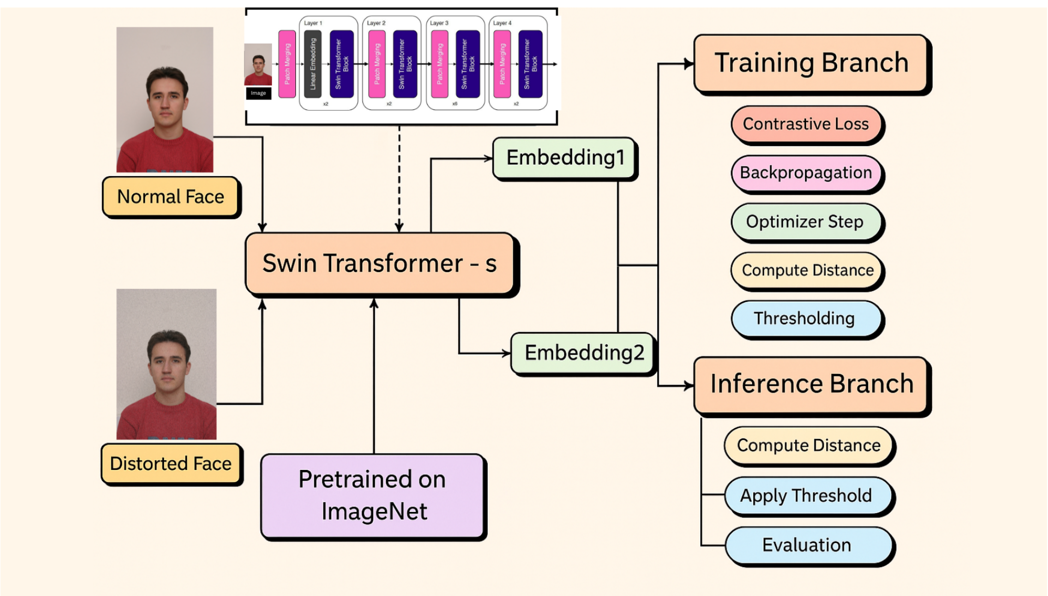

Figure 2 : workflow diagram of the methodology, used for Task B

⚙️ Key Design & Innovation :

✅**Input:** Image pairs (clean & distorted, same or different identity).

✅**Siamese Network:** Employs a dual-stream architecture based on the Swin V2 transformer backbone (shared weights) for robust embedding extraction.

✅**Embedding:** Generates L2-normalized embeddings to improve generalization and representation quality for scale-invariant similarity.

✅**Similarity Metrics:**
- Euclidean distance for contrastive separation.
- Cosine similarity for directional closeness.

✅**Loss Function:** Utilizes a hybrid loss combining Contrastive Loss (measuring embedding distance) and Cosine BCEWithLogits Loss (capturing embedding direction), which combines metric learning and binary similarity classification.

$$\textit{Hybrid Loss} = \alpha \cdot \textit{Contrastive} + (1-\alpha) \cdot \textit{BCEWithLogits (Cosine\_Similarity)}.$$

This **dual-perspective approach** enhances discrimination capability and model robustness.

✅**Training:**
- Implements mixed-precision training to optimize computational efficiency.
- Includes gradient clipping, early stopping, and learning rate scheduling to ensure stable and effective training.

✅**Inference:**
- Uses **cosine similarity** as the primary metric instead of Euclidean distance and **τ threshold for binary match** decision.
- This scale-invariant similarity metric aligns with normalized embeddings and facilitates clear thresholding and ROC curve analysis.

✅**Generalization:** Supports unseen identities and real-world distortions, due to robust design and data augmentation.

🧪 Why This Design Works

🔄 Contrastive Loss pushes apart dissimilar pairs, pulls together similar ones — using Euclidean distance.

➕ Cosine BCEWithLogits Loss treats similarity as a binary classification task over cosine similarity, ideal for normalized embeddings.

🔄 Combining both helps the model generalize to unseen identities and be robust to intra-class variations (like distortions).

💡 Using cosine similarity at inference makes thresholding easier and more consistent across conditions.