Comprehensive Analysis of CFSM

Shreyansh Pathak
MTech PhD AI , IIT Jodhpur

October 21, 2024

Introduction to CFSM - Detailed Analysis

Core Concept:

- **Semi-constrained Dataset (X):**
  - Contains n face images with identity labels
  - Well-controlled imaging conditions
  - Consistent lighting, pose, and quality
  - Mathematically represented as $X = \{X\}_{i=1}^{n}$
- **Target Dataset (Y):**
  - Contains m images without labels: $Y = \{Y\}_{i=1}^{m}$
  - Variable factors: Lighting conditions, Motion blur, Resolution variations, Environmental effects, Pose changes

Technical Objectives:

- Learn style transfer between domains
- Maintain identity preservation
- Control synthesis attributes
- Generate realistic unconstrained variations

Model Overview - Architecture Deep Dive

Encoder (E):

- **Purpose:** Feature extraction
- **Input:** RGB image $X \in \mathbb{R}^{W \times H \times 3}$
- **Output:** Content features $C = E(X)$
- **Architecture:**
  - Convolutional layers
  - Instance normalization
  - Feature maps at multiple scales

Decoder (G):

- **Function:** Image generation
- **Inputs:** Content features C, Style code z
- **Process:** Feature upsampling, Style injection via AdaIN, Final image reconstruction

Multimodal Image Translation Network - Technical Details

Network Architecture:

- Input normalization: $[-1, 1]$ range
- Progressive downsampling with skip connections to decoder

AdaIN Integration:

$$AdaIN(x, y) = y_s \cdot \frac{x - \mu(x)}{\sigma(x)} + y_b$$

where x is the content feature and y_s, y_b are the style parameters.

Generation Process:

- Progressive upsampling
- Style-modulated convolutions
- Final tanh activation

Adversarial Learning Framework - Comprehensive Analysis

Discriminator Architecture:

- PatchGAN structure with a 70x70 receptive field
- Markovian discrimination ensuring local and global consistency

Loss Components:

- Real Sample Processing: $L_{\text{real}} = -\mathbb{E}_{Y \sim Y}[\log(D(Y))]$
- Fake Sample Processing: $L_{\text{fake}} = -\mathbb{E}_{X \sim X, z \sim Z}[\log(1 - D(\hat{X}))]$
- Generator Adversarial Loss: $L_{\text{adv}} = -\mathbb{E}_{X \sim X, z \sim Z}[\log(D(\hat{X}))]$

Domain-Aware Linear Subspace Model - Mathematical Foundation

Subspace Construction:

- Basis Matrix U with dimensions $d \times q$, where $U^T U = I$
- Style Coefficient $o \sim \mathcal{N}_q(0, I)$ controls attribute strength
- Mean Style μ, learned domain center

Orthogonality Enforcement:

$$L_{\text{ort}} = \|U^T U - I\|_1$$

Style Control Mechanism - Detailed Operation

Direction Control System:

- Basis Vector Roles: u_1 for Lighting direction, u_2 for Blur amount, u_3 for Pose variation

Control Process:

$$z = Uo + \mu \quad \hat{X} = G(C, MLP(z))$$

Magnitude Control:

- Style Strength: $a = \|o\|$, Range: $[l_a, u_a]$
- Control Mechanisms: Linear interpolation, Adaptive scaling, Boundary conditions

Identity Preservation - Technical Implementation

Feature Extraction:

- ArcFace Network, Pre-trained weights, Fixed during training
- 512-D feature vectors

Similarity Computation:

$$SC(f_1, f_2) = \frac{f_1 \cdot f_2}{\|f_1\|\|f_2\|}$$

where f_1 and f_2 are features of original and generated images.

Magnitude Function:

$$g(a) = \frac{(a - l_a)(u_m - l_m)}{u_a - l_a} + l_m$$

Complete Loss Functions - Detailed Analysis

Loss Components Breakdown:

- Adversarial Loss: $\lambda_{\text{adv}} = 1$
- Orthogonality Loss: $\lambda_{\text{ort}} = 1$
- Identity Loss: $\lambda_{\text{id}} = 8$

Optimization Strategy:

- Two-phase training: Warm-up phase and Full optimization

Guided Face Synthesis - Methodology

Purpose:

- Training Data Enhancement: Diverse variations, Controlled difficulty, Identity preservation
- FR Model Improvement: Robustness, Generalization, Domain adaptation

Implementation:

- Synthesis Process: Style sampling, Guided perturbation, Quality assessment
- Feedback Loop: FR model performance, Style adjustment, Iterative refinement

Adversarial Guidance - Technical Process

Perturbation Optimization:

$$\delta^* = \arg\max L_{\text{cla}}(F(\hat{X}), I) \quad \text{s.t.} \|\delta\|_\infty < \epsilon$$

FGSM Implementation:

- Gradient Computation: $\nabla_z = \frac{\partial L_{\text{cla}}}{\partial z}$
- Update Rule: $\delta^* = \epsilon \cdot \text{sgn}(\nabla_z)$
- Constraints: $\epsilon = 0.314$, Clip bounds, Normalization

FR Model Integration - System Design

Training Pipeline:

- Batch processing: Original images and synthetic variations
- Combined forward pass for loss computation

Loss Computation:

$$L_{\text{total}} = L_{\text{cla}} + L_{\text{reg}}$$

- Classification loss L_{cla}
- Regularization terms L_{reg}

Update Procedure:

- Gradient computation and parameter update
- Synthesis guidance and model refinement

Dataset Distribution Measure - Mathematical Analysis

Similarity Computation:

$$S(A, B) = \frac{1}{q} \sum_{i} SC(u_i^A + \mu_A, u_i^B + \mu_B)$$

- Basis comparison: Direction alignment and magnitude correlation
- Mean offset: Domain center distance and style distribution overlap

Applications:

- Dataset alignment, Transfer learning, Domain gap measurement

Implementation Details - Technical Specifications

Image Processing:

- Resolution: 112x112
- Preprocessing: Face alignment, Normalization, Augmentation

Network Parameters:

- Architecture: $q = 10$ basis vectors, $d = 128$ style dimension, AdaIN layers
- Training: Batch size = 32, Learning rate = 2×10^{-4}, Adam optimizer

Results and Evaluation - Performance Analysis

Dataset Statistics:

- MS-Celeb-1M: 1M+ images, 100K+ identities, controlled conditions
- WiderFace: 70K images, unconstrained scenarios, various challenges

Benchmark Performance:

- IJB-B: 1,845 subjects, 21.8K images, 55K video frames
- IJB-C: 3,500 subjects, 31,334 images, 117,542 video frames

Evaluation Metrics:

- TAR@FAR, Verification accuracy, Identity preservation, Style transfer quality