AHAN: Asymmetric Hierarchical Attention Network for Identical Twin Face Verification

Hoang-Nhat Nguyen
Faculty of Mathematics and Informatics, Hanoi University of Science and Technology

Problem Statement

The Twin Verification Challenge:
State-of-the-art face recognition achieves **99.8%** accuracy on standard benchmarks, but drops dramatically to **88.9%** for identical twins.

Why is this hard?
Twins share nearly identical facial structure. Standard models focus on *global features* while discriminative info lies in *subtle, non-genetic* variations

AHAN: Multi-stream network analyzing faces at global, local, and asymmetric levels simultaneously. **92.3%** Twin Verification | **88.5%** Hard Twin

Proposed Method

Key Insight: Successful twin discrimination requires simultaneous analysis at three complementary levels.

1. Hierarchical Cross-Attention

Motivation: Different facial regions need different analysis strategies—eyes contain rich texture, jawline captures shape.

Multi-Scale Cross-Attention:

$$\mathbf{A}_k^{(s)} = \text{softmax}\left(\frac{\mathbf{Q}_k^{(s)}(\mathbf{K}^{(s)})^T}{\sqrt{d}}\right)\mathbf{V}^{(s)}$$

Scale-Adaptive Aggregation:

$$\mathbf{f}_k = \sum_s \alpha_{k,s} \cdot \text{Pool}(\mathbf{A}_k^{(s)})$$

Enables region-specific analysis at optimal scales.

2. Facial Asymmetry Attention Module

Motivation: Asymmetry differs even between identical twins.

Left-Right Decomposition:

$$\mathbf{X}_L, \mathbf{X}_R^{flip} = \text{Split \& Flip}(\mathbf{X})$$

Bidirectional Cross-Attention:

$$\mathbf{A}_{L\to R} = \text{softmax}\left(\frac{\mathbf{Q}_L\mathbf{K}_R^T}{\sqrt{d}}\right)\mathbf{V}_R$$

Asymmetry Signature:

$$\mathbf{f}_{asym} = \text{Pool}(|\mathbf{A}_{L\to R} - \mathbf{A}_{R\to L}|)$$

Captures unique left/right differences as identifiers.

3. Twin-Aware Pair-Wise Cross-Attention

Motivation: Standard training focuses on obvious features, but for twins these are shared genetic traits.

Twin Distraction Cross-Attention:

$$\mathbf{K}_{comb} = [\mathbf{K}_a; \mathbf{K}_t]$$

$$\mathbf{A}_{TA} = \text{softmax}\left(\frac{\mathbf{Q}_a\mathbf{K}_{comb}^T}{\sqrt{d}}\right)\mathbf{V}_{comb}$$

Training only (p=0.5, layers 6-9): **Zero inference overhead**

Uses each subject's twin as the hardest possible distractor.

Forces learning of truly individuating patterns.

AHAN Architecture Overview

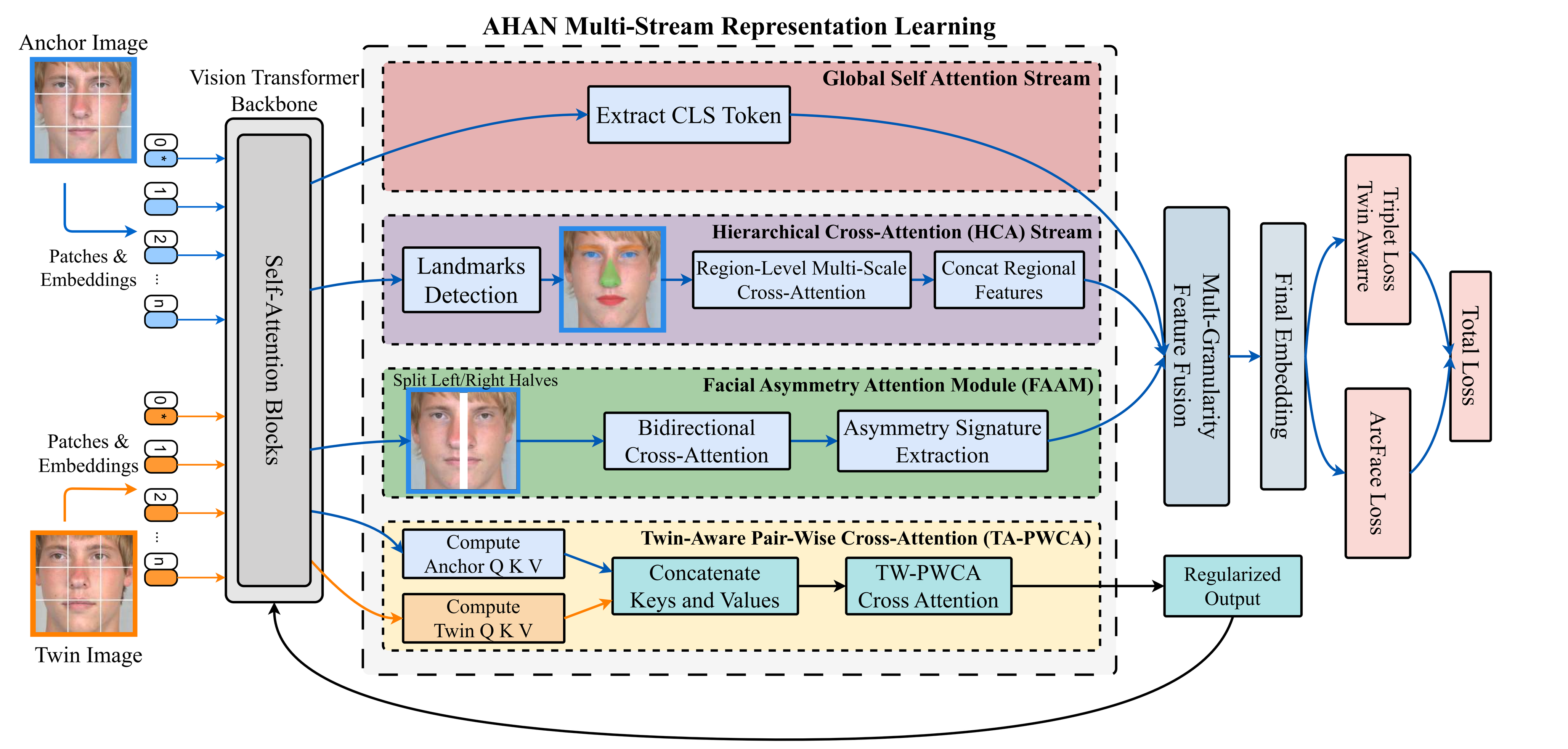

Figure: AHAN employs a ViT backbone with three parallel streams: (1) **Global Self-Attention** for holistic features, (2) **HCA** for multi-scale part-based analysis, and (3) **FAAM** for asymmetry signatures. **TA-PWCA** uses twin images as distractors during training.

Attention Visualization

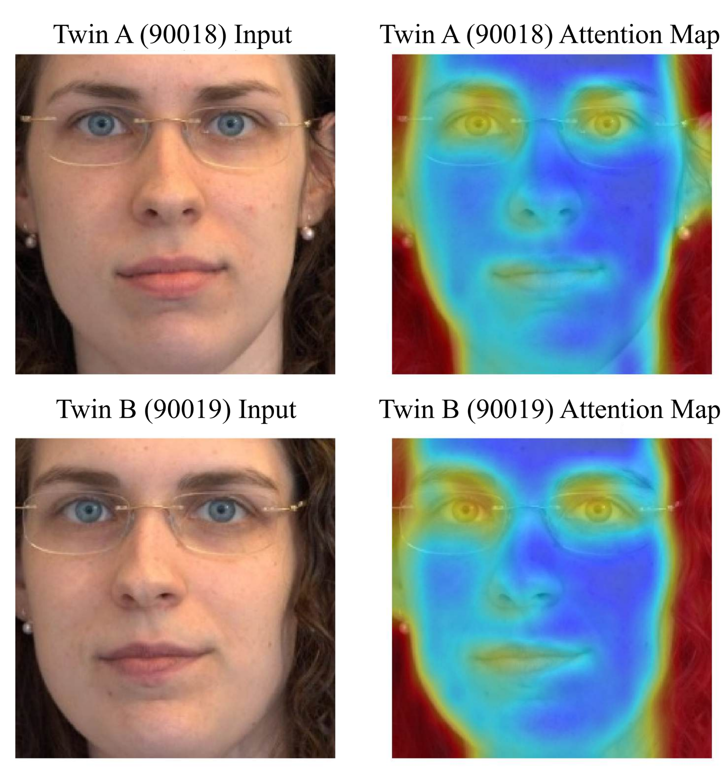

Figure: Attention maps for an identical twin pair.
AHAN learns to focus on discriminative regions (warm colors) that differentiate twins, including subtle textures around eyes and unique asymmetric facial patterns.

Experimental Results

Dataset: ND-TWIN - 24,050 images from 435 people (identical twin pairs)

Main Results on ND_TWIN Dataset

Method	General Verification			Twin Verification			Hard Twin Verification		
	Acc	AUC	TAR@1%	Acc	AUC	TAR@1%	Acc	AUC	TAR@1%
Non Transformer-based Face Recognition									
ArcFace	98.5	99.4	95.6	88.9	93.8	82.4	85.3	90.6	78.4
MagFace	98.4	99.3	95.4	88.5	93.4	81.9	85.0	90.3	78.0
AdaFace	98.2	99.2	95.1	88.2	93.1	81.5	84.7	90.0	77.6
CosFace	98.0	99.1	94.8	87.5	92.5	80.6	84.1	89.4	76.8
CurricularFace	98.3	99.3	95.2	87.1	92.1	80.1	83.6	88.9	76.2
UniFace	97.9	99.0	94.6	86.7	91.8	79.6	83.2	88.5	75.7
Transformer-based Methods									
TransFace	97.5	98.7	94.0	85.2	90.4	77.8	81.8	87.2	73.9
TransFG	97.3	98.5	93.6	84.8	90.0	77.3	81.4	86.8	73.4
AHAN (Ours)	**99.1**	**99.8**	**97.2**	**92.3**	**96.4**	**87.6**	**88.5**	**93.5**	**82.8**

Ablation Study

Configuration	TAR@0.1%	EER	TVA
Baseline (ViT-B)	52.1	24.7%	81.2%
+ HCA	63.4	19.3%	87.1%
+ FAAM	58.9	21.8%	84.7%
+ TA-PWCA	67.8	17.1%	89.9%
+ HCA + FAAM	69.2	16.4%	90.4%
+ HCA + TA-PWCA	74.6	14.2%	91.8%
+ FAAM + TA-PWCA	71.3	15.7%	90.7%
Full AHAN	**78.4**	**12.1%**	**92.3%**

Conclusion: AHAN achieves **92.3%** twin verification (**+3.4%** over ArcFace) through multi-granularity analysis + asymmetry modeling + twin-aware training. This establishes a new paradigm for extreme fine-grained biometric verification.

References: [1] Phillips et al. "Distinguishing identical twins by face recognition." *FG*, 2011. [2] Zhu et al. "Dual Cross-Attention Learning for Fine-Grained Visual Categorization." *CVPR*, 2022.