G-FAN: Graph-Based Feature Aggregation Network for Video Face Recognition

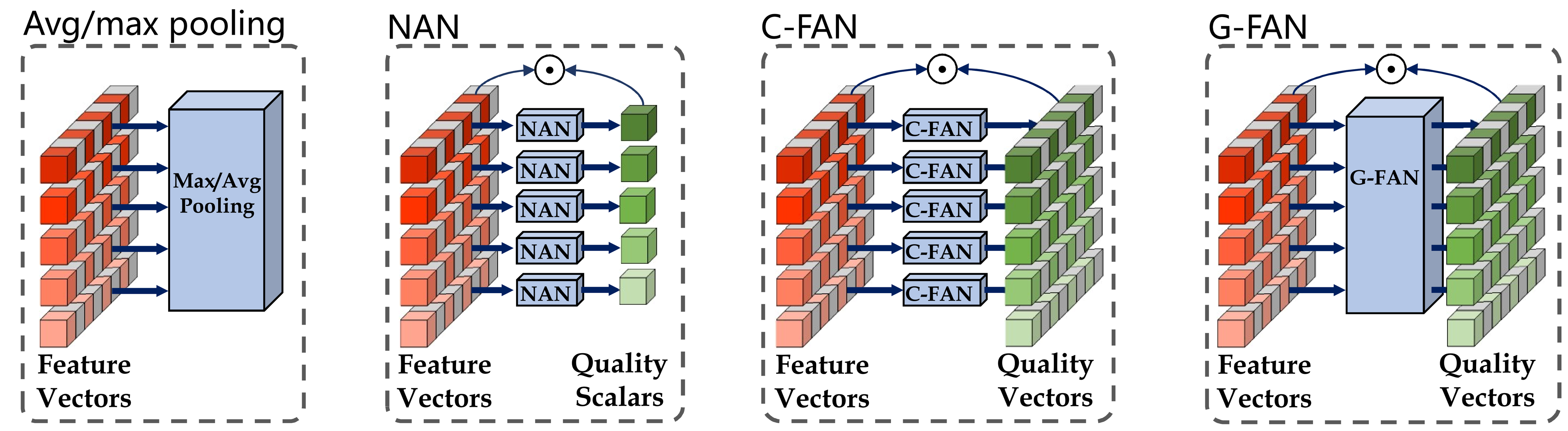
He Zhao, Yongjie Shi, Xin Tong, Jingsi Wen, Xianghua Ying*, and Hongbin Zha
Key Laboratory of Machine Perception (MOE), School of EECS, Peking University

1. Motivation and Contribution

Motivation:

★ Video face recognition exhibits great challenges due to huge intra-class variability and high inter–class ambiguity.

★ On the other hand, a video clip could provide useful temporal and multi-view information which is profitable for recognition.



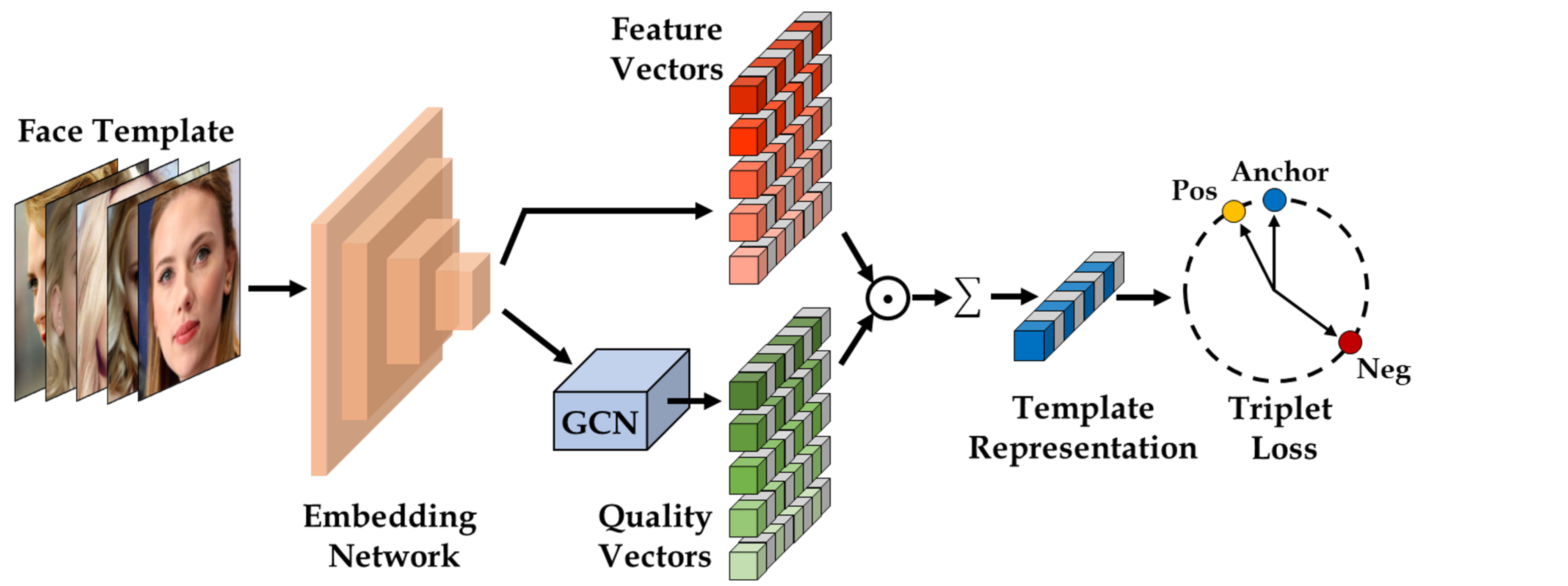
Contribution:

★ We propose the graph-based feature aggregation network, which firstly utilizes correlation information between face frames to aggregate frame features.

★ G-FAN achieves SOTA performance on three public benchmarks, including YTF, IJB-A, and IJB-C.

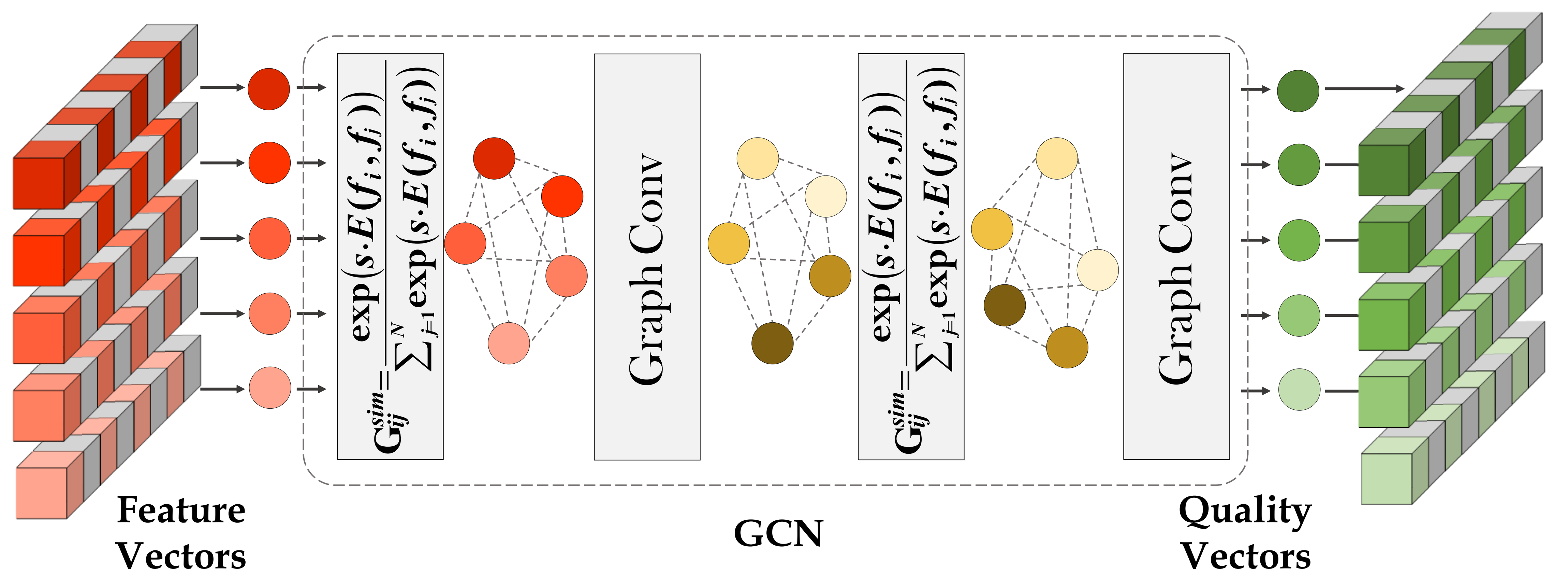
★ We qualitative analysis the output quality vectors of G-FAN which provides an explanation for the effectiveness.

2. Overview



★ G-FAN incorporates an Embedding Network for feature extraction and a GCN for feature aggregation.

3. GCN Module



★ For a template of features $F = \{f_1, f_2, ..., f_N\}$, we build a fully connected graph with N nodes, where each node represents an image feature. The pairwise similarity is calculated as $E(f_i, f_j) = cos(f_i, f_j)$

★ We apply a two-layer GCN to perform reasoning on the graph and outputs quality vectors. $F^{(l+1)} = \sigma\left((G^{sim})^{(l)} F^{(l)} W^{(l)}\right)$,

★ The quality vectors are normalized by Softmax: $w_{i,j} = \frac{\exp(q_{i,j})}{\sum_{k=1}^{N} \exp(q_{kj})}$. The aggregated face representation of a template is obtained by pooling the features with the normalized quality vectors: $r = \sum_{i=1}^{N} f_i \odot w_i$,

4. Experiment Results

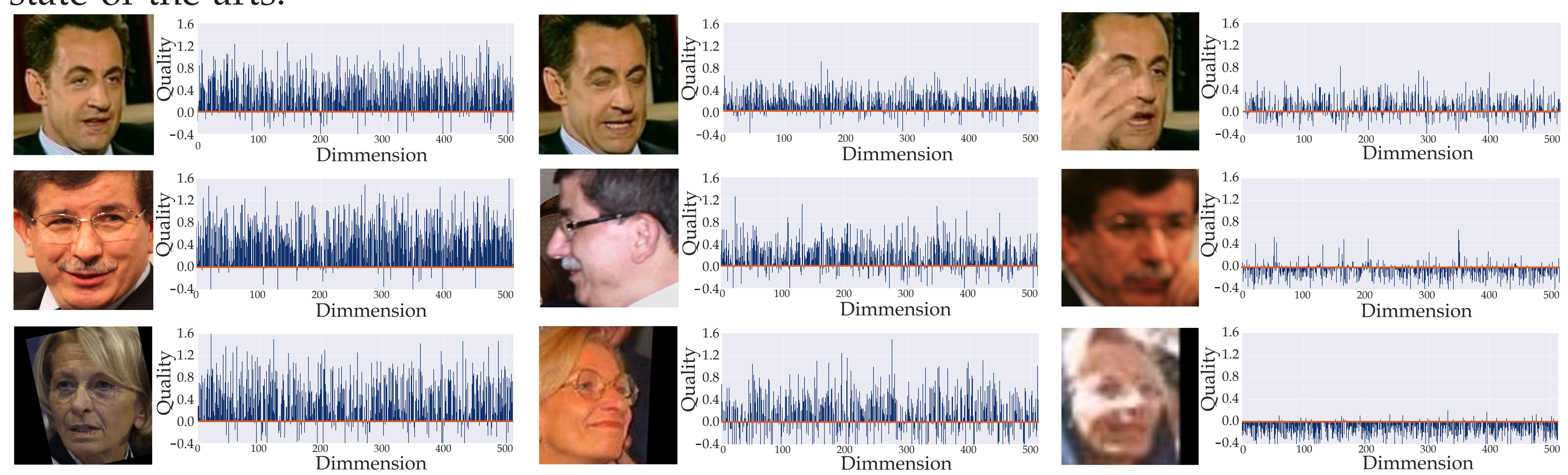
TABLE III
VERIFICATION ACCURACY (%) ON IJB-A BENCHMARK COMPARED WITH BASELINE METHODS AND OTHER STATE-OF-THE-ART METHODS. THE TRUE ACCEPT RATES (TAR) VS FALSE ACCEPT RATES (FAR) ARE REPORTED.

Method	1:1 Verification TAR		
	FAR=0.001	FAR=0.01	FAR=0.1
Crystal Loss [31]	94.80	97.10	98.50
NAN [6]	88.10	94.10	97.80
QAN [30]	89.31	94.20	98.02
M-FAN []	94.44	96.56	98.00
C-FAN [7]	91.59	93.97	-
FAN [12]	93.61	97.28	98.94
GhostVLAD [17]	93.50	97.20	99.00
AFRN [29]	94.90	98.50	**99.80**
Average	92.85	97.36	99.10
*C-FAN**	94.74	97.85	99.27
G-FAN	**95.97**	**98.64**	**99.55**

TABLE IV
VERIFICATION ACCURACY (%) ON IJB-C BENCHMARK COMPARED WITH BASELINE METHODS AND OTHER STATE-OF-THE-ART METHODS. THE TRUE ACCEPT RATES (TAR) VS FALSE ACCEPT RATES (FAR) ARE REPORTED.

Method	1:1 Verification TAR		
	FAR=1e-5	FAR=1e-4	FAR=1e-3
Crystal Loss [31]	87.35	92.29	95.63
CosFace [24]	86.94	91.82	95.37
ArcFace [8]	87.28	92.13	95.55
AdaCos [32]	88.03	92.40	95.65
APA [33]	85.5	92.06	96.23
AFRN [29]	88.30	93.00	96.30
Average	87.16	91.89	95.37
*C-FAN**	87.90	92.57	95.74
G-FAN	**89.42**	**93.83**	**96.38**

★ Experiments on three public video face benchmarks show that G-FAN outperforms other state-of-the-arts.



★ The qualitative analysis demonstrates that G-FAN can maintain the discriminative features while discarding the noisy features.