EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

Ig3D: Integrating 3D Face Representations in Facial Expression Inference

Lu Dong*, Xiao Wang*, Srirangaraj Setlur, Venu Govindaraju, and Ifeoma Nwogu

* Equal Contribution

University at Buffalo The State University of New York

NATIONAL AI INSTITUTE for Exceptional Education

ABAW

Motivation

- Reconstructing 3D faces with facial geometry from single images has allowed for major advances in animation, generative models, and virtual reality. However, this ability to represent faces with their 3D features is not as fully explored by the facial expression inference (FEI) community. Our contributions in this work are threefold:
  - We provide insights into the key parameters of 3D face representations within the context of facial emotion inference.
  - We introduce two architectures for integrating 3D representations: intermediate fusion and late fusion.
  - Extensive experiments demonstrate the efficiency of our method, surpassing the state-of-the-art in both AffectNet Valence-Arousal (VA) estimation and RAF-DB classification.

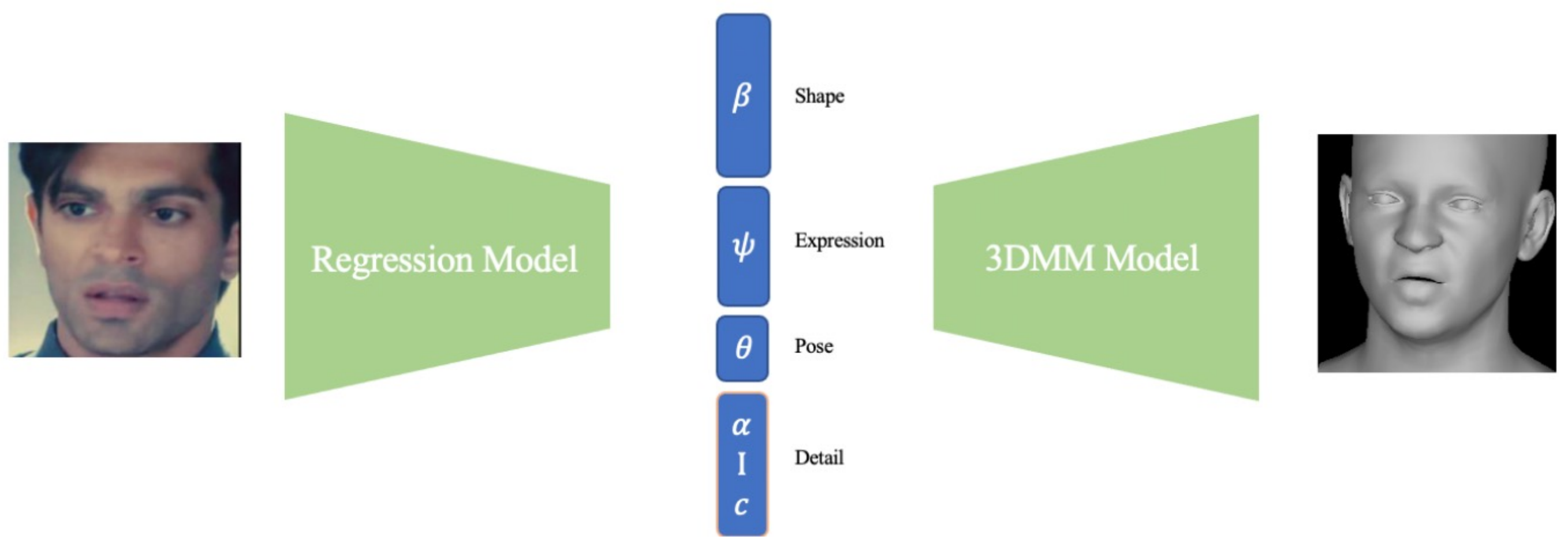

Fig. 1: A standard pipeline for 3D facial geometry reconstruction from an image.

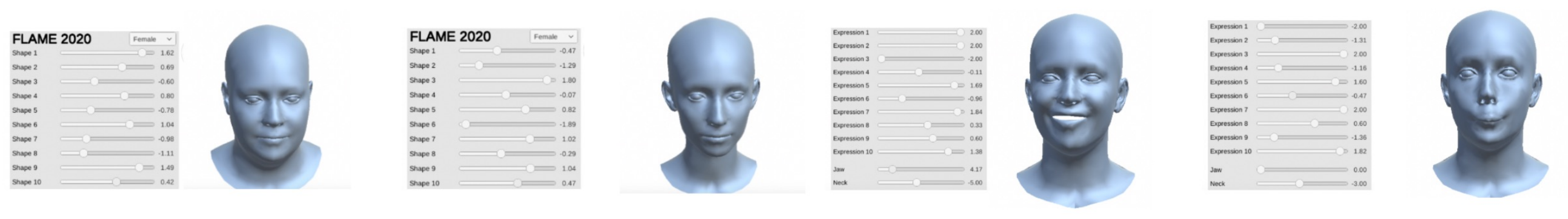

Fig. 2: 3D Representation Visualization.

Ig3D Overview

Comparison of 3D Face Representations

- The EMOCA model regresses a total of 334 parameters: 100 for shape, 50 for emotional expressions, 6 for pose, 100 for detail, 50 for texture, and others including pose-dependent and articulated components.
- The SMIRK model regresses to 358 standard parameters of which 300 are shape, 50 are expression and 6 are pose. Other additional parameters include camera parameters and those specific to the neural rendering process used in SMIRK.

Loss Function:

Discrete Expression Inference

$$Loss = L_{CE} + \frac{\alpha}{\alpha+\beta+\gamma} \times L_{MSE} + \frac{\beta}{\alpha+\beta+\gamma} \times (1 - L_{CCC}) + \frac{\gamma}{\alpha+\beta+\gamma} \times (1 - L_{PCC}) \quad (1)$$

Continuous Expression Inference

$$Loss_{combined} = L_{weightedCE} + w_1 \cdot L_{MSE} \quad (2)$$

$$Loss_{va} = L_{CCC} + w_2 \cdot L_{MSE} \quad (3)$$

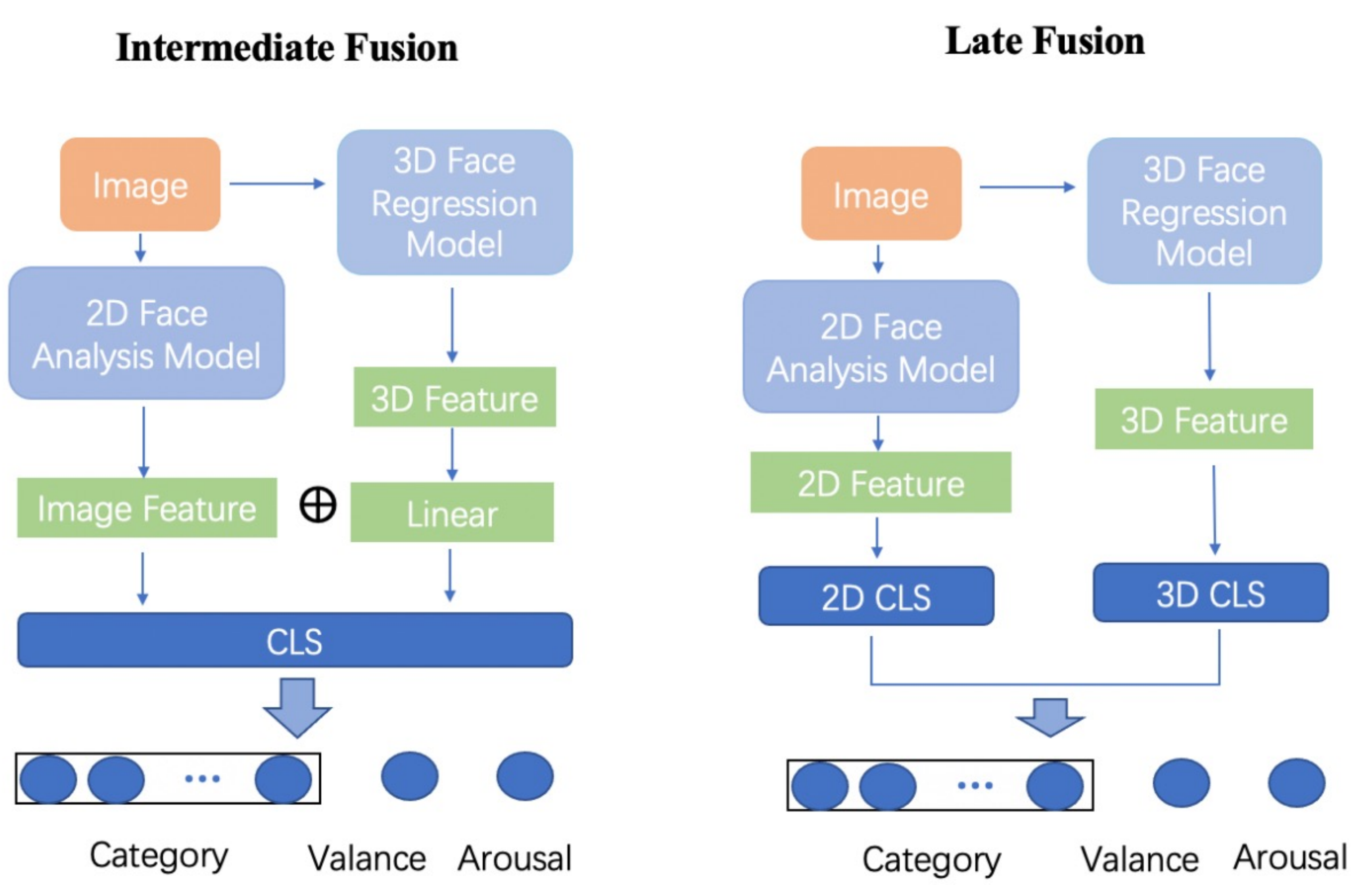

Fig. 3: Overview of the 3D Representation Fusion Architecture

Experiments

Evaluation Metrics:

- **Discrete Expression Inference**
  - Accuracy; F1 score; Precision; Recall.
- **Continuous Expression Inference**
  - Mean Squared Error (MSE);
  - Mean Absolute Error (MAE);
  - Root Mean Squared Error (RMSE);
  - Concordance Correlation Coefficient (CCC);

Quantitative Results

Table 2: Classification Comparison of EMOCA and SMRIK 3D Representations only (no fusion) on AffectNet Dataset.

3D Classifier	Accuracy ↑	F1 ↑	Precision ↑	Recall ↑
$CLS_{Smirk3D-short}$	0.5461	0.5459	0.5477	0.5461
$CLS_{Smirk3D-full}$	0.5546	0.5547	0.5569	0.5546
$CLS_{Emoca3D-short}$	**0.5723**	**0.5726**	**0.5758**	**0.5723**
$CLS_{Emoca3D-full}$	0.5703	0.5704	0.5768	0.5703

Table 3: Classification Comparison of EMOCA and SMRIK 3D Representations only (no fusion) on RAF-DB Dataset. Due to the unbalanced test dataset, we report both weighted and macro average metrics for a comprehensive evaluation. **Acc** stands for Accuracy, **F1** for F1 score, **P** for Precision, and **R** for Recall.

3D Classifier	Acc ↑	Weighted Avg			Macro Avg		
		F1 ↑	P ↑	R ↑	F1 ↑	P ↑	R ↑
$CLS_{Smirk3D-short}$	0.7378	0.7418	0.7475	0.7418	0.6421	0.6386	0.6482
$CLS_{Smirk3D-full}$	0.7557	0.7584	0.7631	0.7557	0.6585	0.6568	0.6627
$CLS_{Emoca3D-short}$	0.7862	0.7873	0.7895	0.7862	0.6965	0.6908	0.7037
$CLS_{Emoca3D-full}$	**0.7927**	**0.7946**	**0.7985**	**0.7927**	**0.7073**	**0.7043**	**0.7118**

Table 4: Classification Comparison of Different Fusion Architectures on AffectNet Dataset.

Framework	Accuracy ↑	F1 ↑	Precision ↑	Recall ↑
DDAMFN (our reproduction)	0.6324	0.6323	0.6353	0.6324
Intermediate Fusion				
$F_{2D} + F_{Smirk3D}$	0.6117	0.6098	0.6128	0.6117
$F_{2D} + F_{Emoca3D}$	0.6234	0.6232	0.6276	0.6234
Late Fusion				
Max with $CLS_{Smirk3D}$	0.6267	0.6260	0.6273	0.6267
Max with $CLS_{Emoca3D}$	0.6294	0.6292	0.6306	0.6294
Mean with $CLS_{Smirk3D}$	0.6262	0.6266	0.6315	0.6262
Mean with $CLS_{Emoca3D}$	0.6289	0.6295	0.6338	0.6289
Weighted with $CLS_{Smirk3D}$	0.6364	0.6367	0.6408	0.6364
Weighted with $CLS_{Emoca3D}$	**0.6379**	**0.6381**	**0.6379**	**0.6379**

Table 5: Classification Comparison of Different Fusion Architectures on RAF-DB Dataset. Due to the unbalanced test dataset, we report both weighted and macro average metrics for a comprehensive evaluation. **Acc** stands for Accuracy, **F1** for F1 score, **P** for Precision, and **R** for Recall.

Framework	Acc ↑	Weighted Avg			Macro Avg		
		F1 ↑	P ↑	R ↑	F1 ↑	P ↑	R ↑
DDAMFN (our reproduction)	0.9016	0.9013	0.9022	0.9016	0.8554	0.8686	0.8451
Intermediate Fusion							
$F_{2D} + F_{Smirk3D}$	0.9006	0.9007	0.9018	0.9006	0.8489	0.8561	0.8435
$F_{2D} + F_{Emoca3D}$	0.8996	0.8990	0.8989	0.8996	0.8501	0.8559	0.8453
Late Fusion							
Max with $CLS_{Smirk3D}$	0.8989	0.8984	0.8989	0.8989	0.8527	0.8656	0.8426
Max with $CLS_{Emoca3D}$	0.8941	0.8944	0.9021	0.8941	0.8462	0.8643	0.8485
Mean with $CLS_{Smirk3D}$	0.9030	0.9024	0.9041	0.9030	0.8561	0.8829	0.8361
Mean with $CLS_{Emoca3D}$	0.9130	0.9135	0.9178	0.9130	0.8413	0.8414	0.8521
Weighted with $CLS_{Smirk3D}$	0.9106	0.9099	0.9110	0.9106	0.8689	0.8914	0.8516
Weighted with $CLS_{Emoca3D}$	**0.9400**	**0.9393**	**0.9397**	**0.9400**	**0.8958**	**0.9090**	**0.8860**

Table 6: Comparison with Previous SOTA models for Discrete FEI on RAF-DB Dataset.

Method	Accuracy [%]	Date [mm-yy]
FMAE [50]	93.09	07-2024
S2D [8]	92.57	12-2023
BTN [18]	92.54	07-2024
ARBEx [64]	92.37	05-2023
DDAMFN [71]	92.34	07-2023
Ours	**94.00**	07-2024

Table 7: Continuous VA Results from Different Fusion Architectures on AffectNet Dataset.

Framework	MSE ↓	MAE ↓	RMSE ↓	CCC ↑
$CAGE_{va}$ (Our reproduction)	0.1044	0.2377	0.3230	0.7814
3D Representation				
$Regresser_{Emoca3D}$	0.1061	0.2483	0.3257	0.7637
Feature Fusion				
$F_{2D} + F_{Emoca3D}$	0.1061	0.2398	0.3257	0.7749
Late Fusion				
Max with $Regressor_{Emoca3D}$	0.1052	0.2419	0.3243	0.7727
Min with $Regressor_{Emoca3D}$	0.1053	0.2441	0.3245	0.7726
Mean with $Regresser_{Emoca3D}$	**0.0956**	0.2325	**0.3092**	0.7891
Weighted with $Regresser_{Emoca3D}$	0.0958	**0.2316**	0.3095	**0.7901**

Table 8: Benchmark Comparison for VA Inference on AffectNet Dataset.

Framework	$RMSE_{val}$↓	$RMSE_{aro}$↓	CCC_{val} ↑	CCC_{aro} ↑	Date[mm-yy]
VGG-G [5]	0.356	0.326	0.710	0.629	03-2021
CAGE [62]	0.331	0.305	0.716	0.642	04-2024
Ours	**0.323**	**0.294**	**0.724**	**0.650**	07-2024

Project **Paper**