JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

SelfME: Self-Supervised Motion Learning for Micro-Expression Recognition

Xinqi Fan[1], Xueli Chen[1], Mingjie Jiang[1], Ali Raza Shahid[1,2], Hong Yan[1]

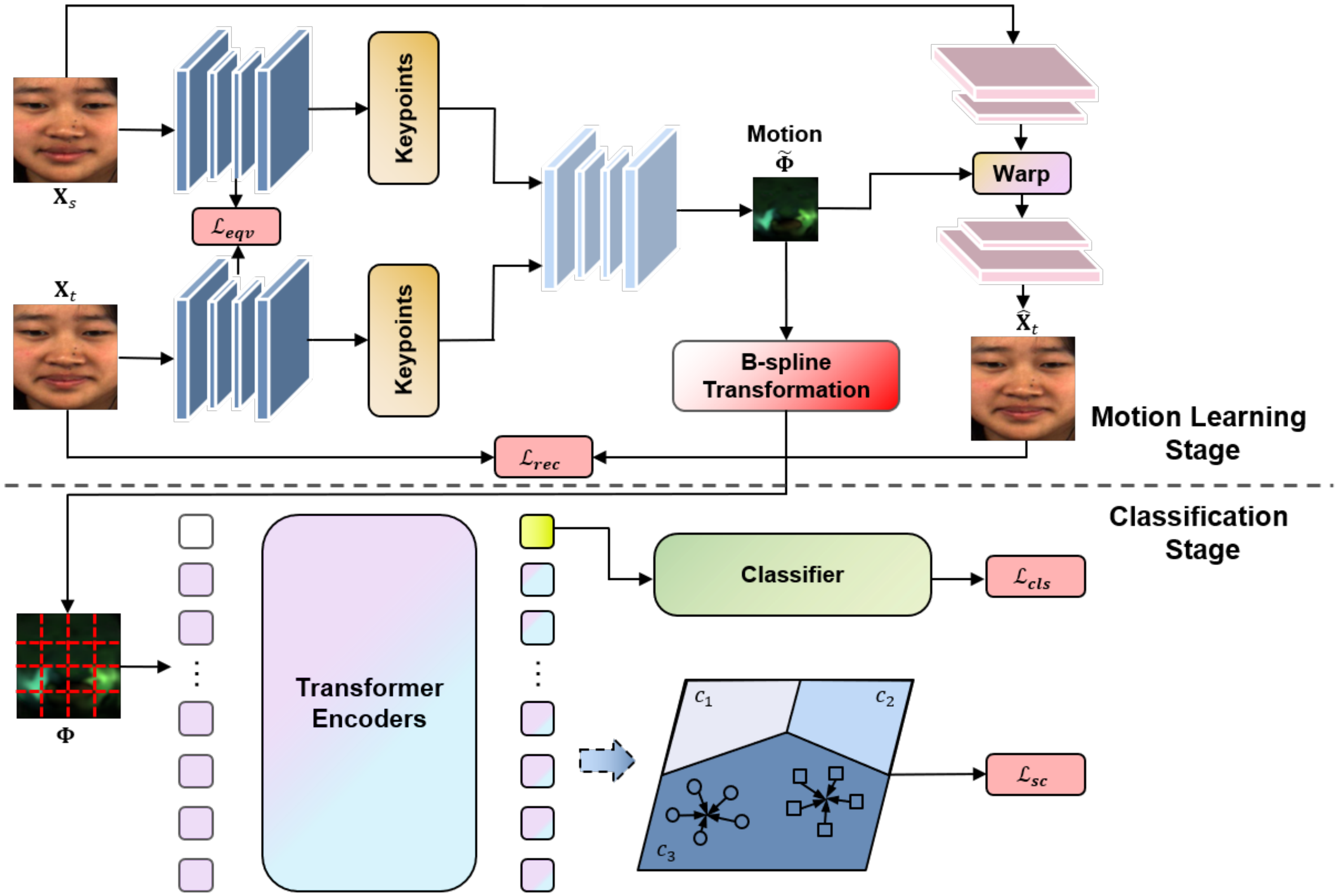
1 City University of Hong Kong, 2 COMSATS University Islamabad

xinqi.fan@my.cityu.edu.hk

WED-PM-141

SelfME Overview

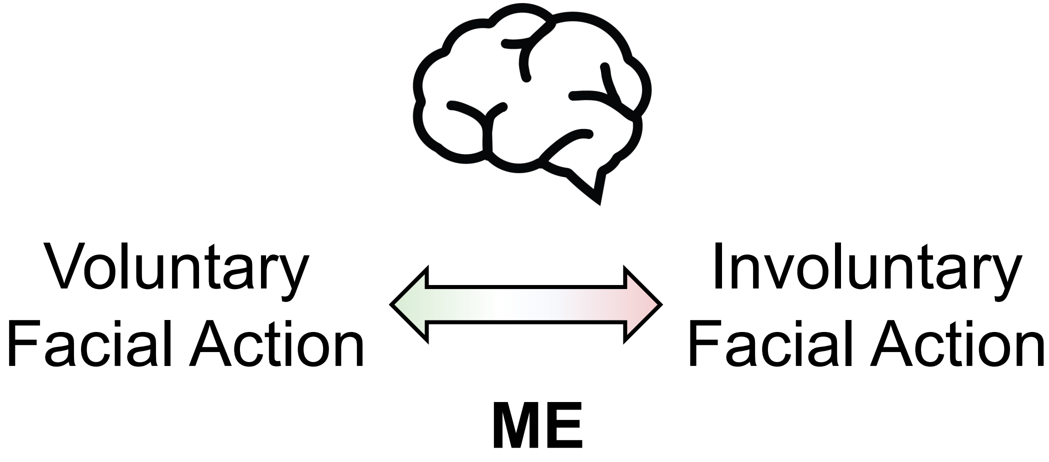
- Facial micro-expression (ME)
- Imperceptible to the naked eye
- Self-supervised motion representation
- Symmetric contrastive vision transformer (SCViT)



Introduction

Facial Micro-Expression (ME)

- Brief spontaneous facial movement
- Genuine emotion
- Characteristics
  - Subtle in intensities
  - Brief in duration (<0.5 seconds)
  - Affect small areas
- Applications
  - National security
  - Political psychology
  - Medical care

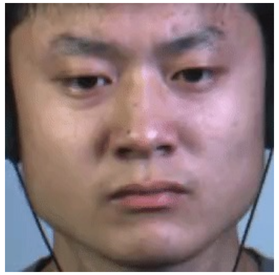


Voluntary Facial Action

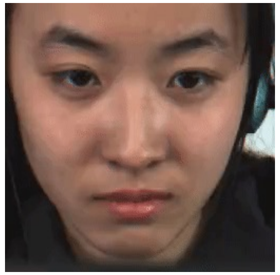
Involuntary Facial Action

ME

Positive



Negative

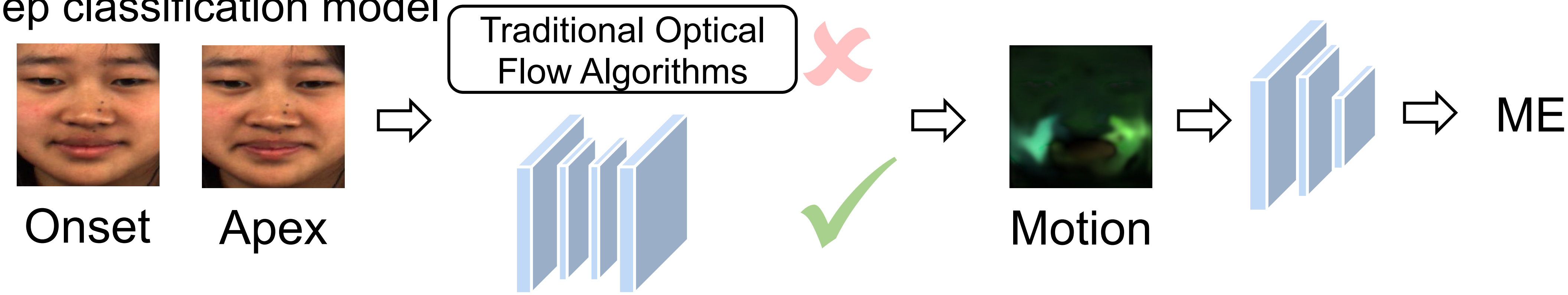


Surprise

Introduction

Motivation: Innovation of ME Pipeline

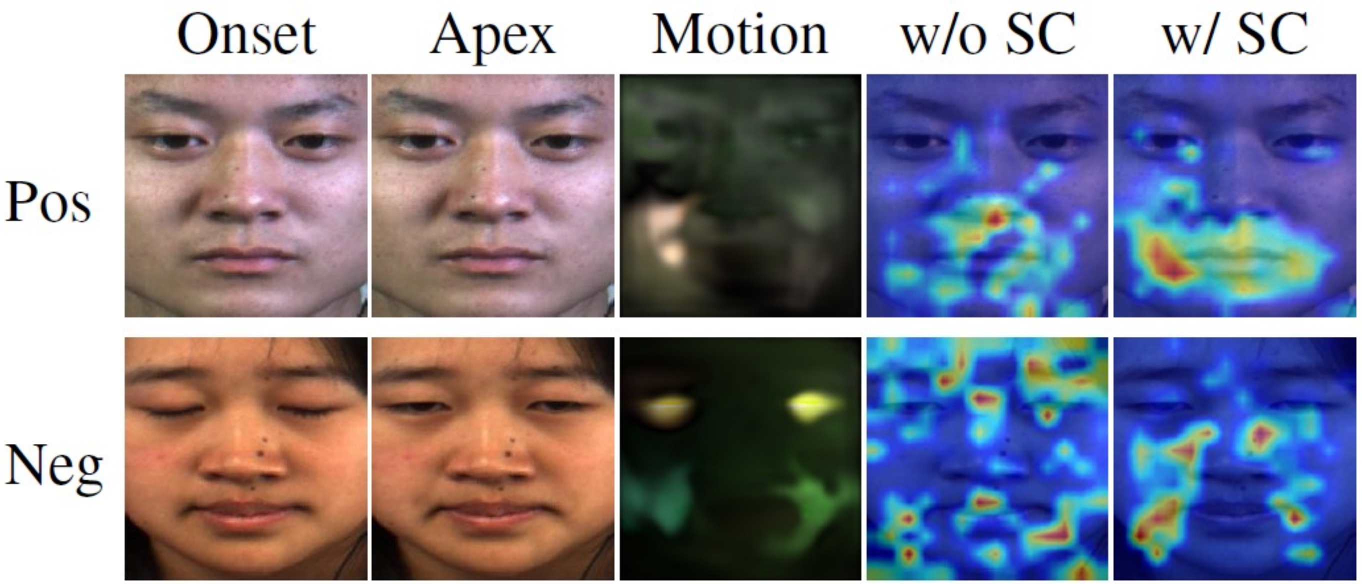
- Motion representation
- Current pipeline
  - Motion extracted by traditional optical flow algorithms
  - Deep classification model
- Proposed pipeline
  - Motion learned by deep self-supervised learning
  - Deep classification model



Introduction

Motivation: Symmetry of Facial Actions in ME

- Symmetric facial actions
- Neglibible intensity differences



- Problem

  Symmetry ignored by learned motions

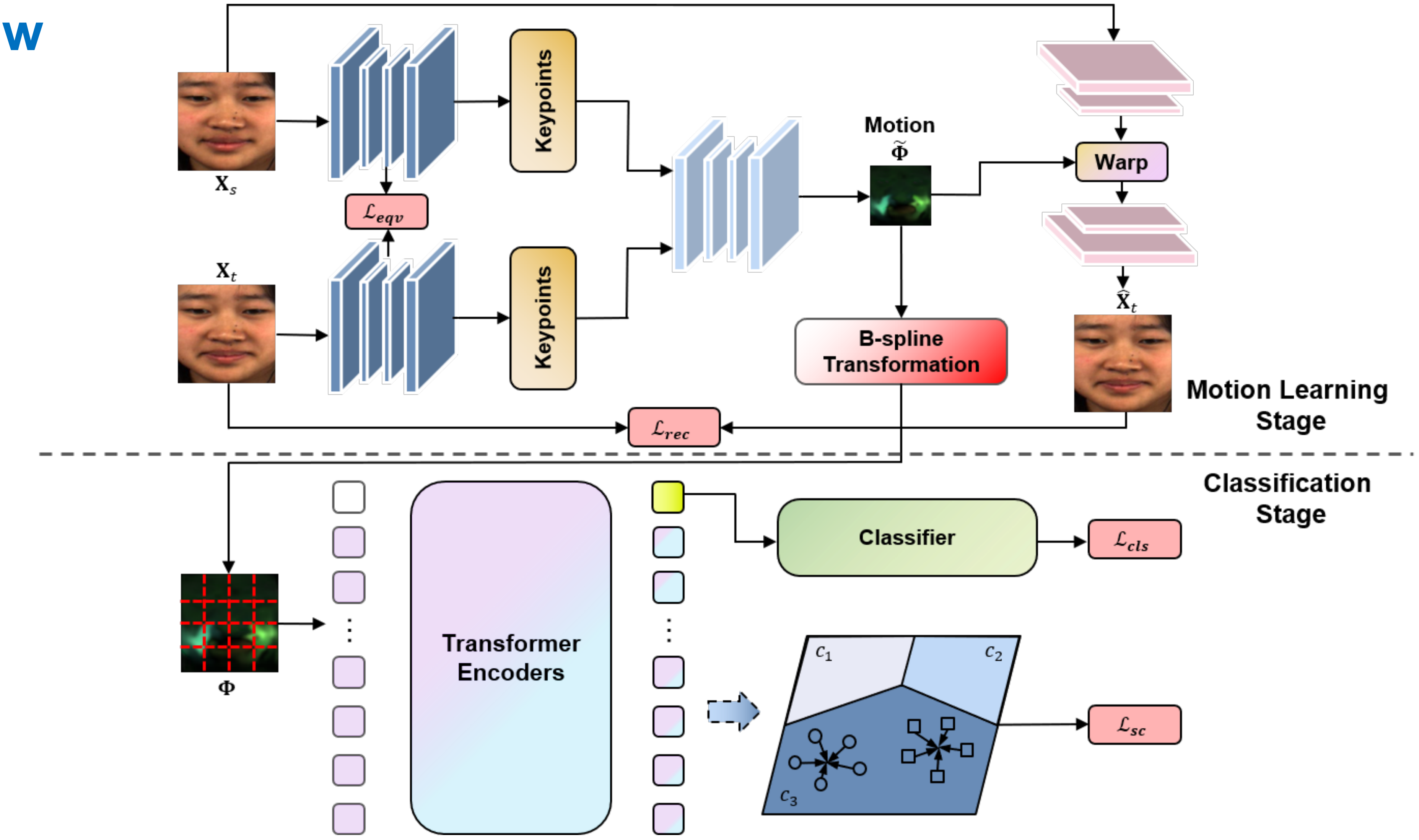
- Solution

  Symmetric contrastive constraint

Methodology

Method Overview

- Motion Learning
- Classification

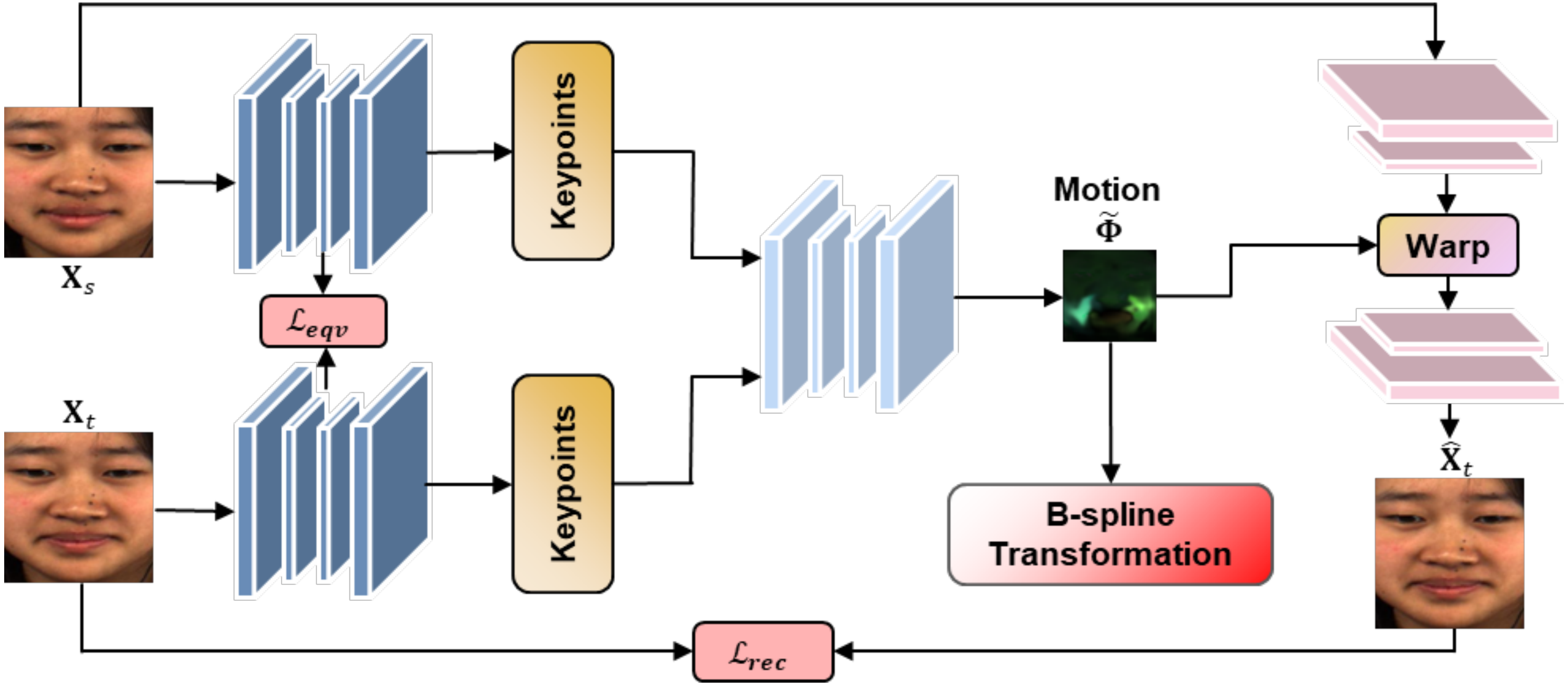


Methodology

Motion Learning Stage

- Self-supervised motion learning
- Reconstruction
- Warp

- Keypoints
- Sparse motion
- Dense motion
- B-spline

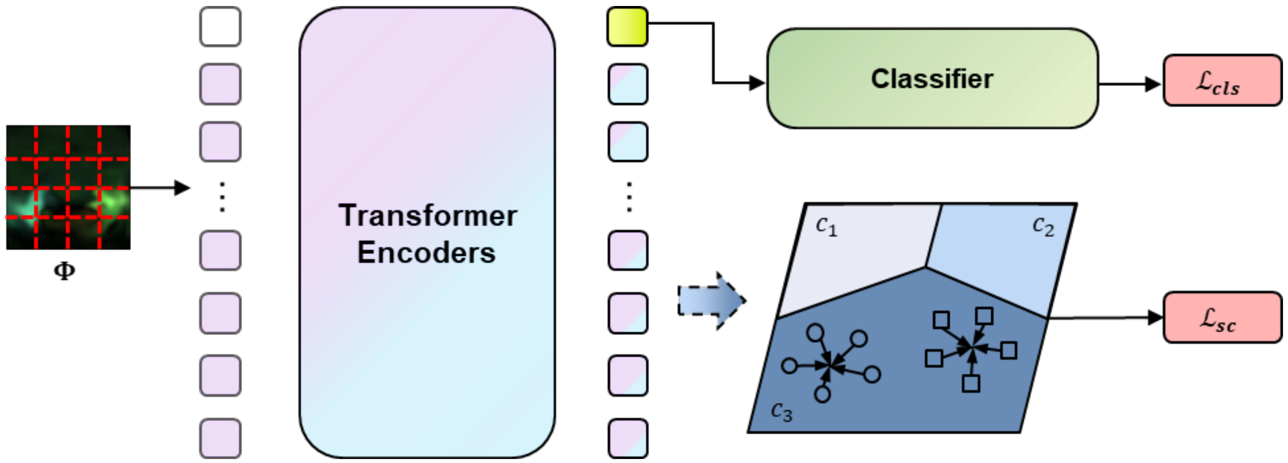


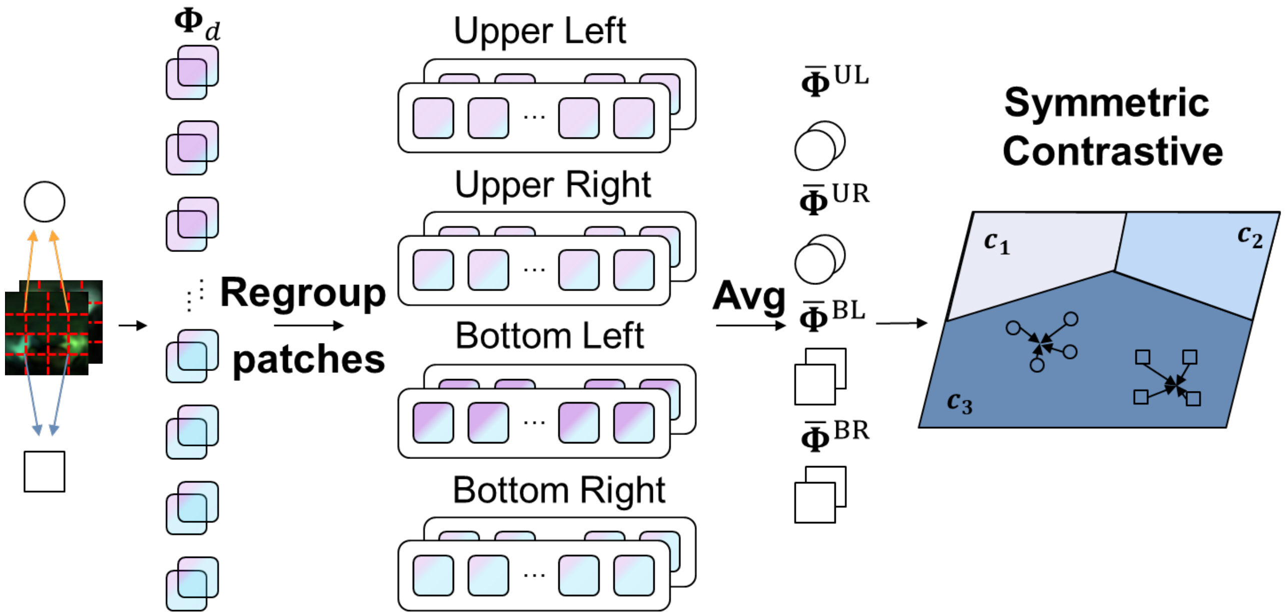
Methodology

Classification Stage

- Vision transformer (ViT)
  - Patches
  - Small and subtle features
  - Violate geometry and symmetry
- Symmetric contrastive
  - Left →← Right
  - Regroup patches
  - 4 regions
  - $\mathcal{L}_{sc}$

$$\mathcal{L}_{sc} = \sum_{i \in I} \frac{-1}{|P(i,\alpha,\beta)|} \sum_{\alpha,\beta \in R} \sum_{p \in P(i,\alpha,\beta)} \log \frac{\exp\left(\bar{\boldsymbol{\Phi}}_i^{\alpha} \cdot \bar{\boldsymbol{\Phi}}_p^{\beta} / \tau\right)}{\sum_{a \in A(i,\alpha,\beta)} \exp\left(\bar{\boldsymbol{\Phi}}_i^{\alpha} \cdot \bar{\boldsymbol{\Phi}}_a^{\beta} / \tau\right)}$$

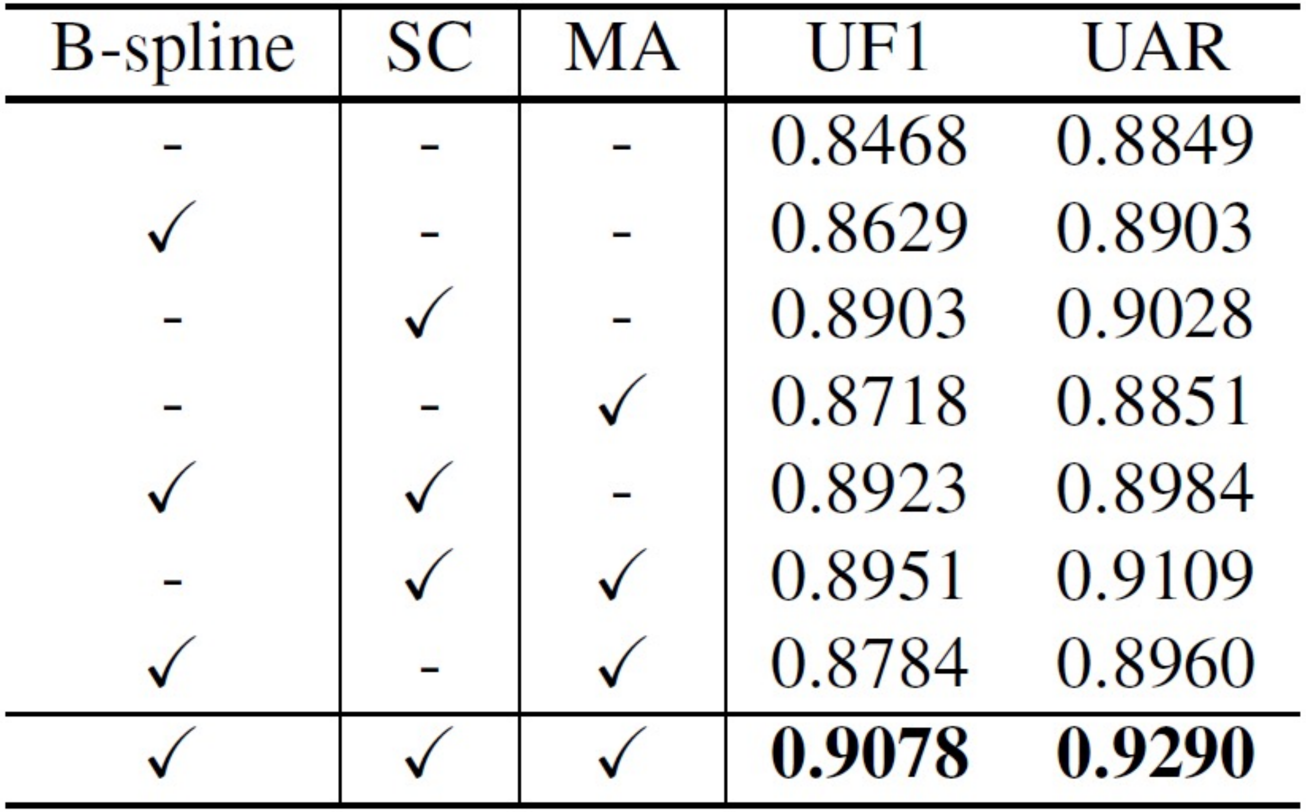




Experiment and Result

Ablation Study

- Symmetric contrastive (SC)
- B-spline transformation
- Motion amplification (MA)

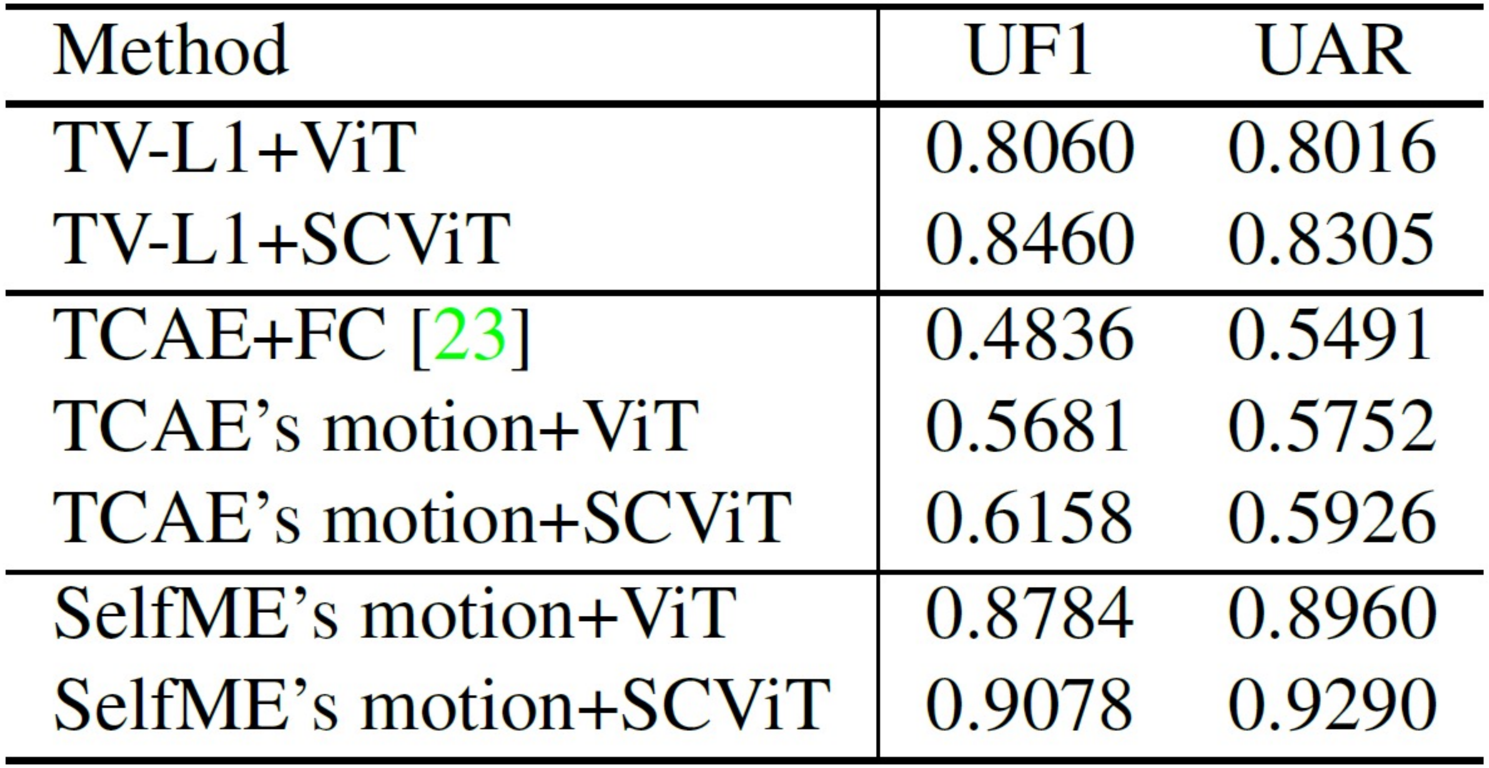


B-spline	SC	MA	UF1	UAR
-	-	-	0.8468	0.8849
✓	-	-	0.8629	0.8903
-	✓	-	0.8903	0.9028
-	-	✓	0.8718	0.8851
✓	✓	-	0.8923	0.8984
-	✓	✓	0.8951	0.9109
✓	-	✓	0.8784	0.8960
✓	✓	✓	**0.9078**	**0.9290**

Experiment and Result

Impact of the Learned Motion

- SelfME’s motion is better than
  - TV-L1
  - TCAE’s motion



Method	UF1	UAR
TV-L1+ViT	0.8060	0.8016
TV-L1+SCViT	0.8460	0.8305
TCAE+FC [23]	0.4836	0.5491
TCAE’s motion+ViT	0.5681	0.5752
TCAE’s motion+SCViT	0.6158	0.5926
SelfME’s motion+ViT	0.8784	0.8960
SelfME’s motion+SCViT	0.9078	0.9290

Twin-cycle autoencoder (TCAE) was proposed for AU detection [CVPR’19, TPAMI’20].

Experiment and Result

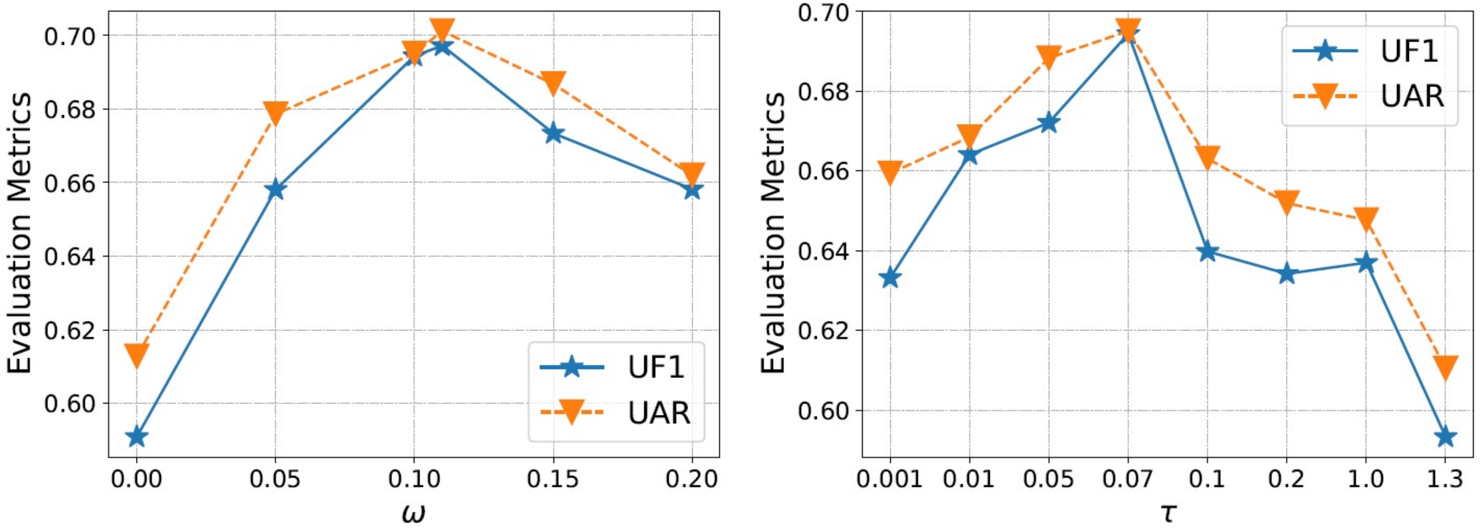
Hyperparameter Analysis

- Best trade-off weight

  $\omega = 0.11$

- Best sharpen temperature

  $\tau = 0.07$

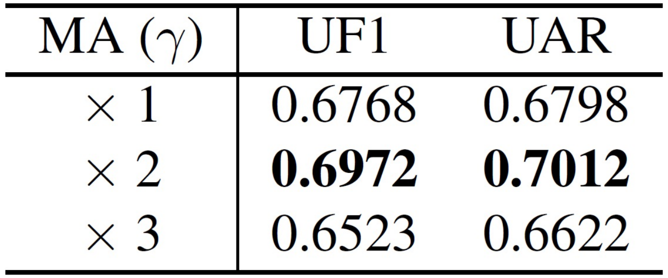


(a) Hyperparameter ω.

(b) Hyperparameter τ.

- Best motion amplification

  $\gamma = 2$

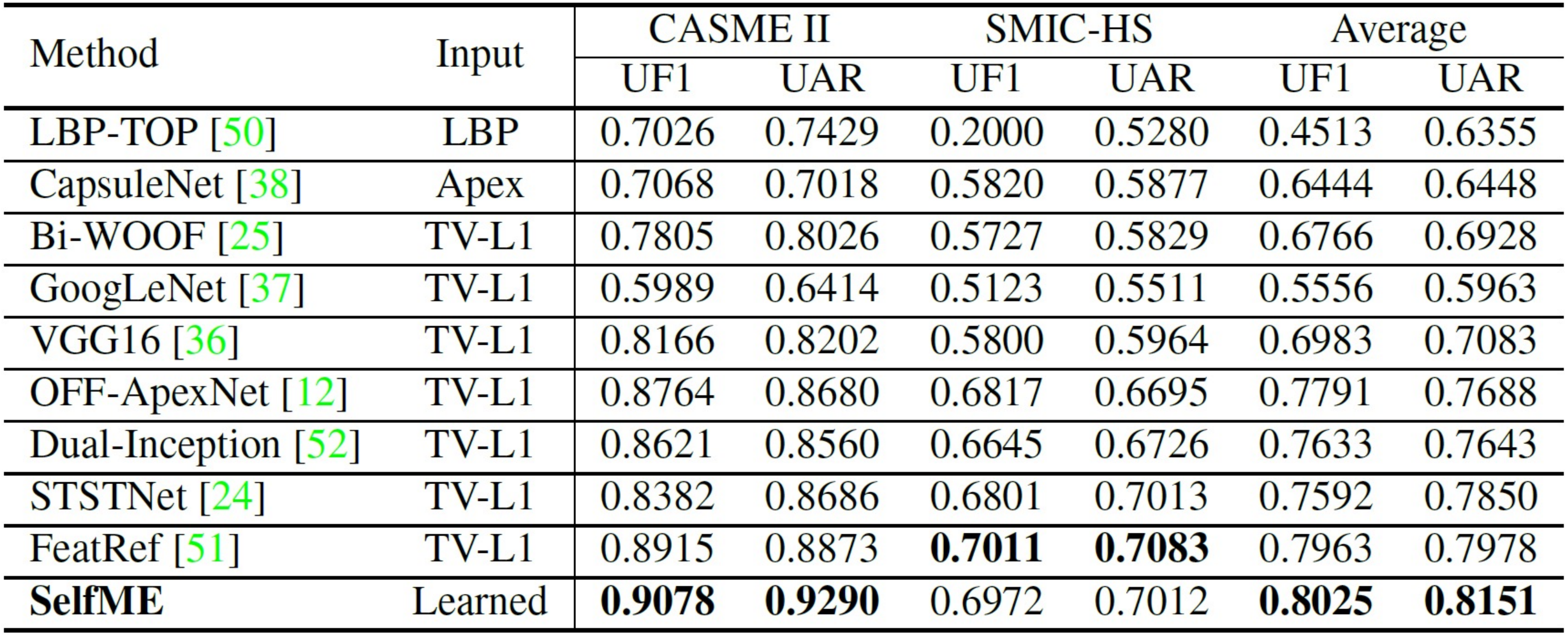


MA (γ)	UF1	UAR
× 1	0.6768	0.6798
× 2	**0.6972**	**0.7012**
× 3	0.6523	0.6622

Experiment and Result

Comparison with the State-of-the-Art

- The 1st self-learned motion representation for MER



Method	Input	CASME II		SMIC-HS		Average	
		UF1	UAR	UF1	UAR	UF1	UAR
LBP-TOP [50]	LBP	0.7026	0.7429	0.2000	0.5280	0.4513	0.6355
CapsuleNet [38]	Apex	0.7068	0.7018	0.5820	0.5877	0.6444	0.6448
Bi-WOOF [25]	TV-L1	0.7805	0.8026	0.5727	0.5829	0.6766	0.6928
GoogLeNet [37]	TV-L1	0.5989	0.6414	0.5123	0.5511	0.5556	0.5963
VGG16 [36]	TV-L1	0.8166	0.8202	0.5800	0.5964	0.6983	0.7083
OFF-ApexNet [12]	TV-L1	0.8764	0.8680	0.6817	0.6695	0.7791	0.7688
Dual-Inception [52]	TV-L1	0.8621	0.8560	0.6645	0.6726	0.7633	0.7643
STSTNet [24]	TV-L1	0.8382	0.8686	0.6801	0.7013	0.7592	0.7850
FeatRef [51]	TV-L1	0.8915	0.8873	**0.7011**	**0.7083**	0.7963	0.7978
SelfME	Learned	**0.9078**	**0.9290**	0.6972	0.7012	**0.8025**	**0.8151**

Thank you !

Please feel free to discuss and ask questions.