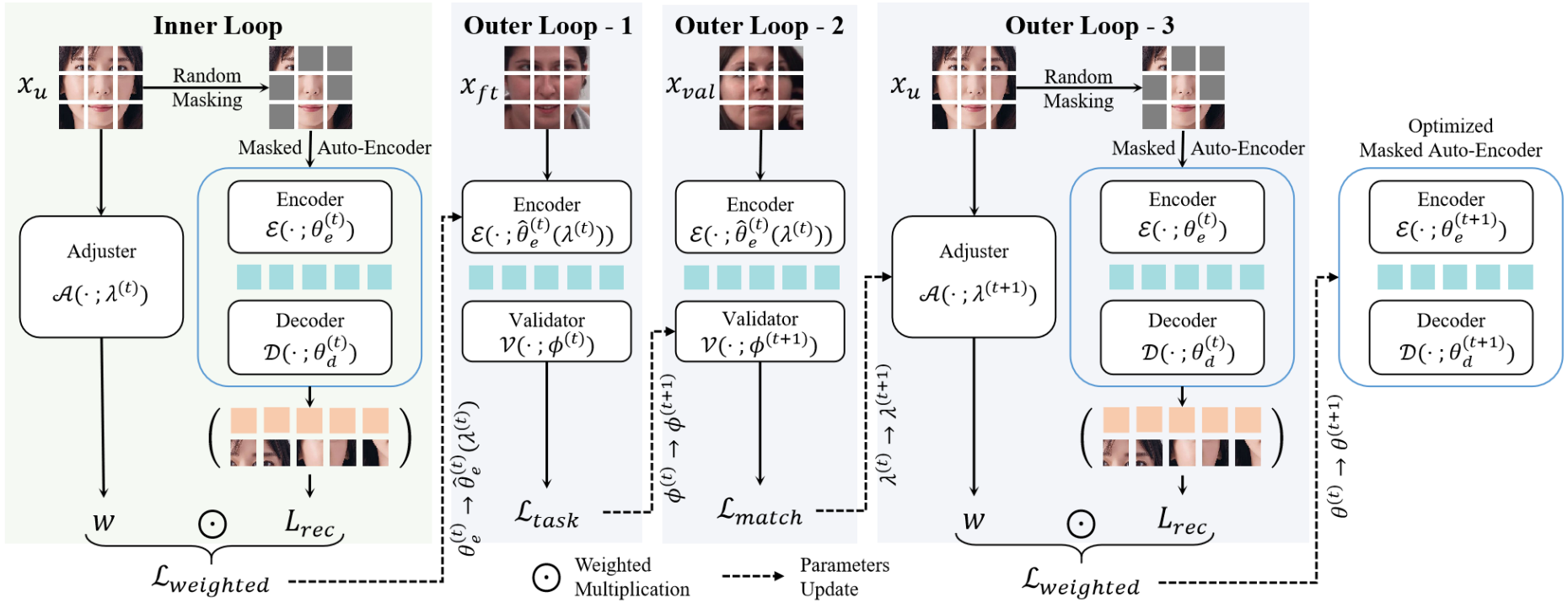
Task-aware Pre-training for Facial Action Analysis

Xuran SUN, Jiabei ZENG, Shiguang SHAN

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50286-3

Problems & Ideas

- **Challenges in Facial Action Analysis (FAA)**
  - The need for expert annotators in FAA tasks makes it difficult to obtain large-scale labeled datasets.
  - General-purpose pre-training methods struggle to capture the fine-grained facial motion features required for FAA tasks.
- **Our Key Idea – TAME Framework**
  - Task-Aware Masked Auto-Encoder (TAME) dynamically customizes its pre-training objective in a task-aware manner, guided by feedback from downstream FAA performance.



Main Contributions

- **Meta-learning-inspired Pre-training Strategy**
  - We propose a novel pre-training framework inspired by meta-learning, featuring a dual-loop optimization process.
- **State-of-the-art Performance**
  - TAME achieves new SOTA results on three core FAA tasks: Action Unit (AU) Detection, AU Intensity Estimation and Pain Estimation

Categories	Methods	Backbones	Pre-training Datasets	AU Detection BP4D F1 score ↑	AU Intensity Est. DISFA ICC ↑	Pain Est. UNBC-McMaster Acc. ↑
SOTA Methods	FAT	CNN-MHSA	ImageNet	0.64	-	-
	RA-UWML	CNN-ROI-MHSA	FEC	-	0.62	-
	Wang et al.	CNN-BiLSTM	VGGFace	-	-	0.90
SSL Methods	MAE	ViT-B	AffectNet	0.66	0.62	0.89
	MoCov3	ViT-B	AffectNet	0.55	0.49	0.86
Foundation Models	CLIP	ViT-B	LAION-2B	0.49	0.32	0.85
	CLIP	ViT-B	AffectNet	0.60	0.53	0.87
Ours	TAME	ViT-B	AffectNet	**0.69**	**0.68**	**0.92**

Comparison of model performance on AU detection, AU intensity estimation and pain estimation tasks. Best results in bold.