ICCV HONOLULU HAWAII
OCT 19-23, 2025

SHANGHAI JIAO TONG UNIVERSITY

FixTalk: Taming Identity Leakage for High-Quality Talking Head Generation in Extreme Cases

Shuai Tan, Bill Gong, Bin Ji, Ye Pan

tanshuai0219@outlook.com

What is Talking Head Generation?

Source Image

+

Driving Audio

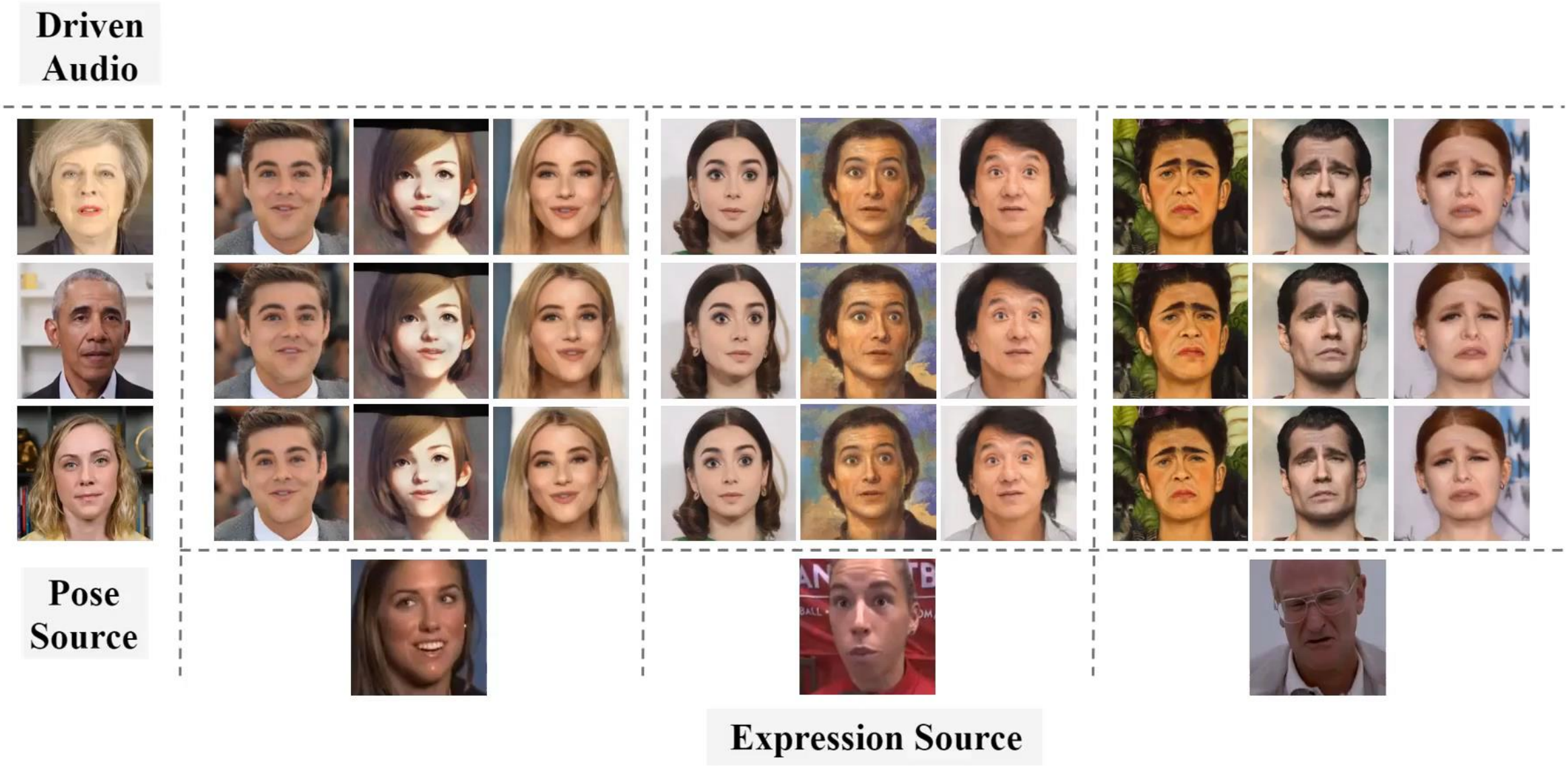
OR

Driving Video

=

Generated Talking Head Videos

The Results of Previous Works: EDTalk [ECCV'24 Oral]

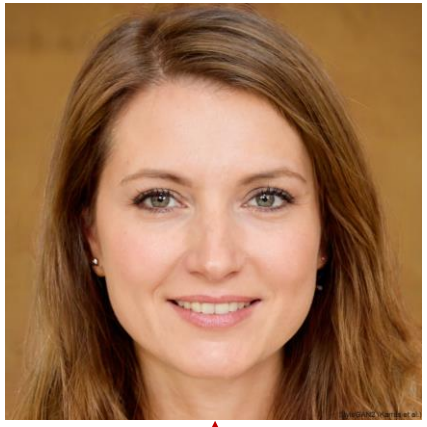


The Results of Previous Works: EDTalk [ECCV'24 Oral]

Vital Observations and Motivations

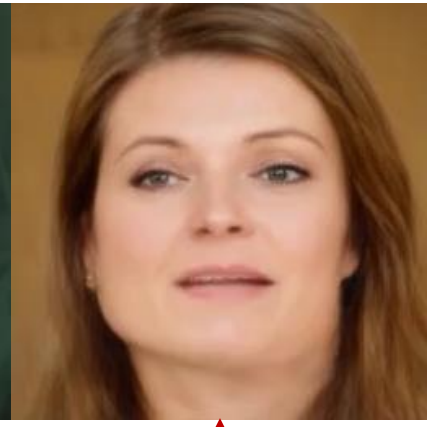
Identity Leakage

Rendering Artifacts



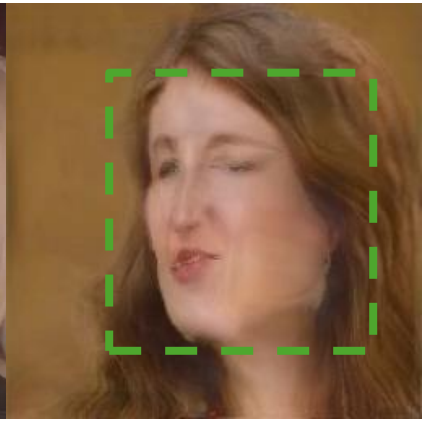
Source Image

Driven video #1



Result #1

Driven video #2

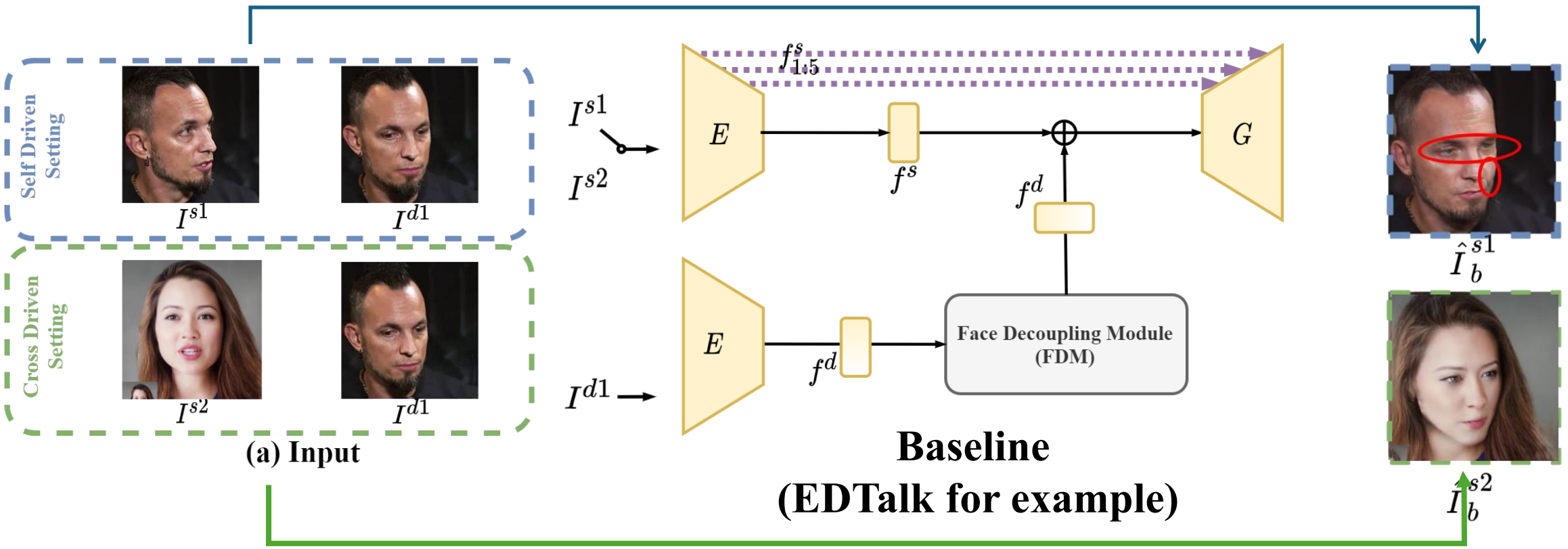


Result #2

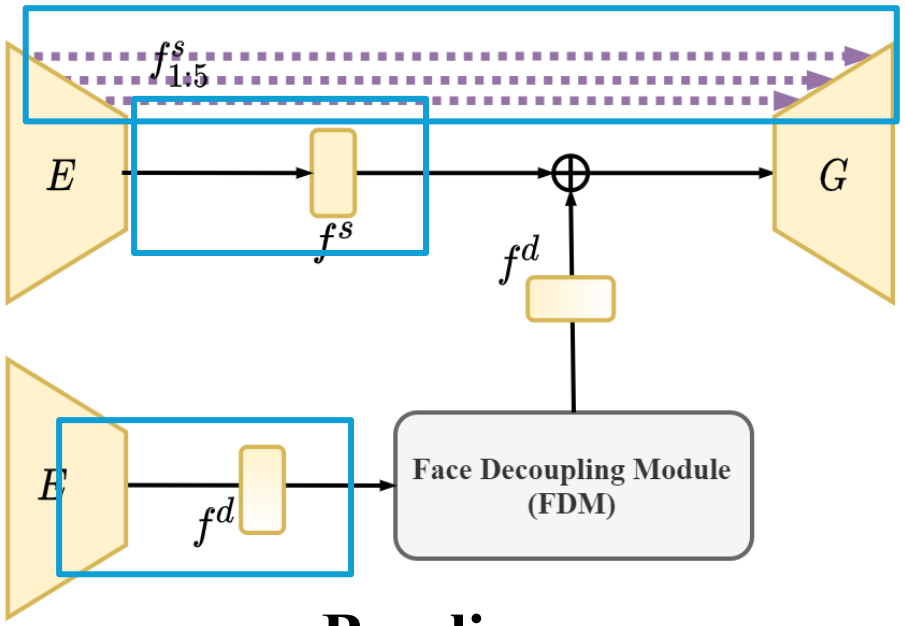
Low CSIM caused by
Identity Leakage

FixTalk: Taming Identity Leakage for High-Quality Talking Head Generation in Extreme Cases

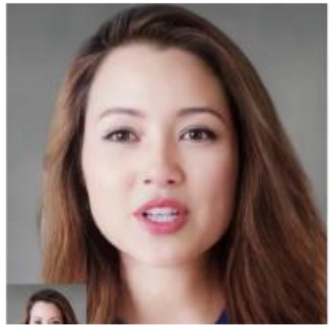
Exploration in Baseline



Exploration in Baseline



Baseline
(EDTalk for example)



Source Image

Cross Driven Results

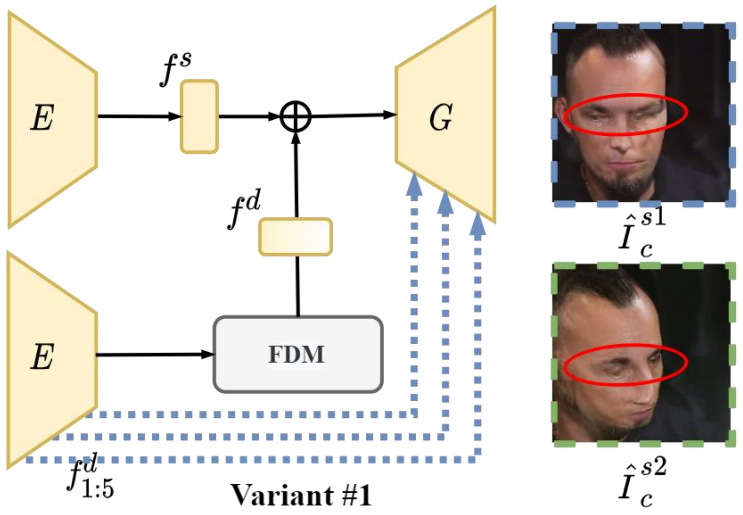
Driven Image

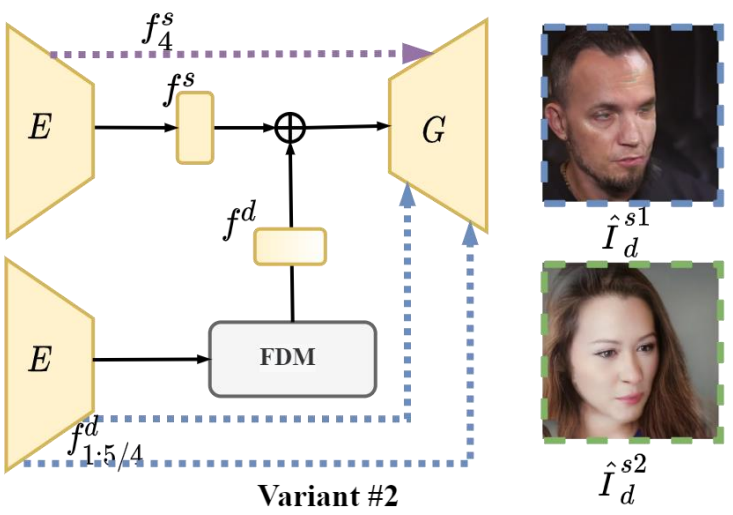
Generated results

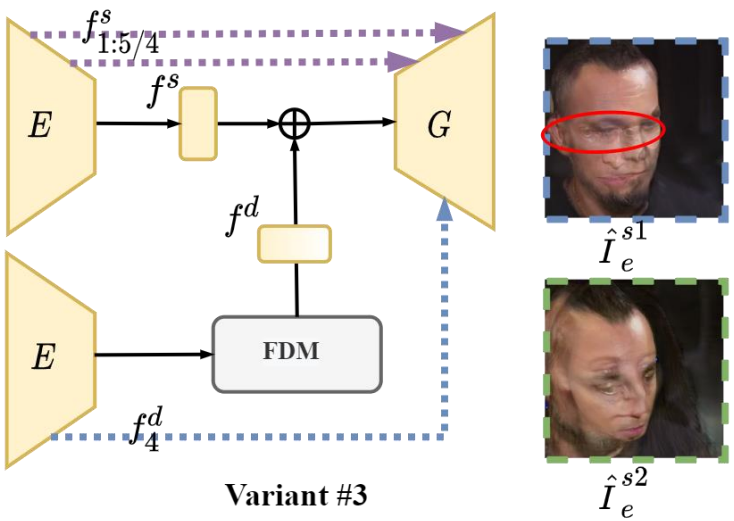
Similar face shape

Identity Leakage

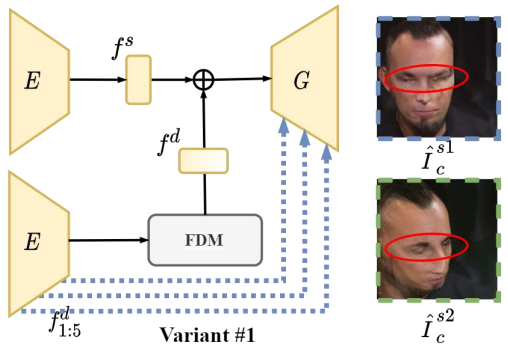
Exploration in Baseline

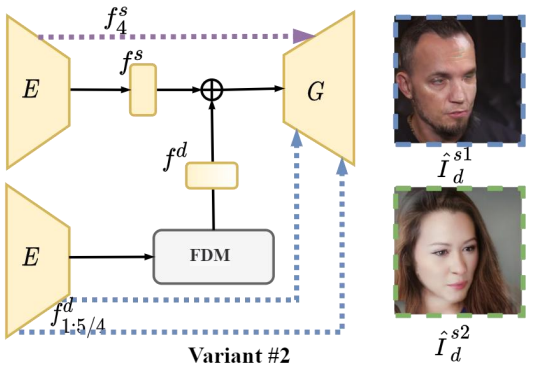


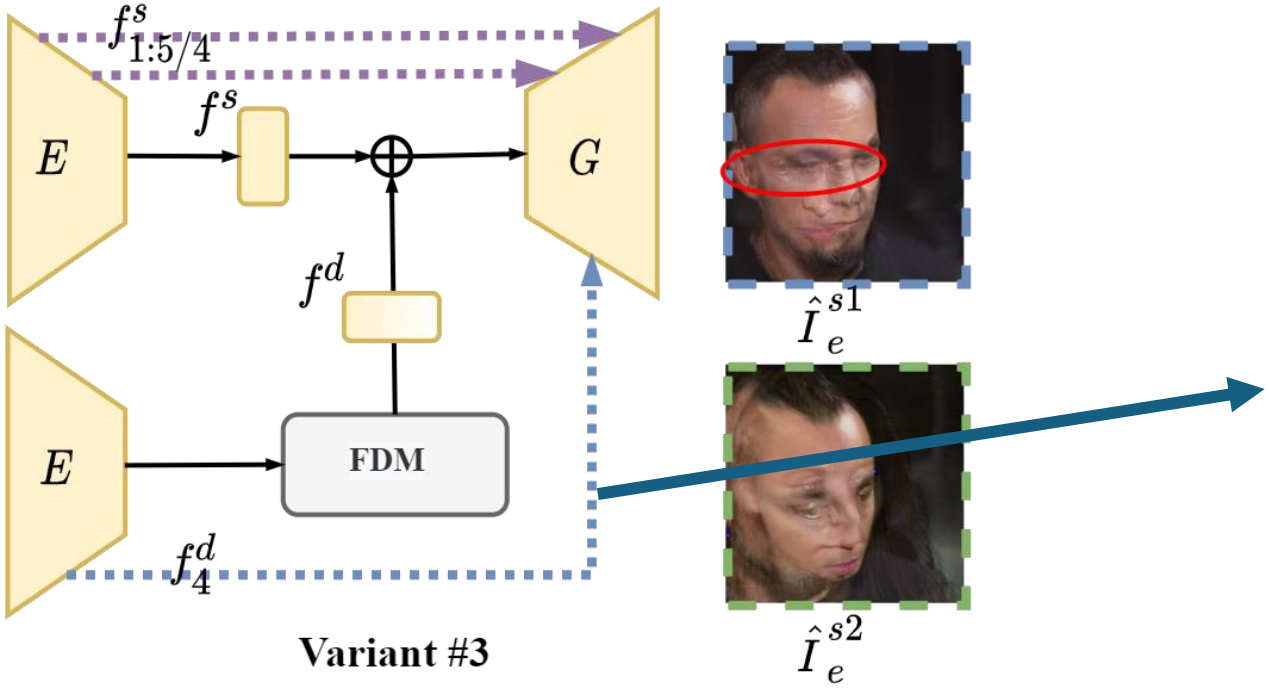




Exploration in Baseline

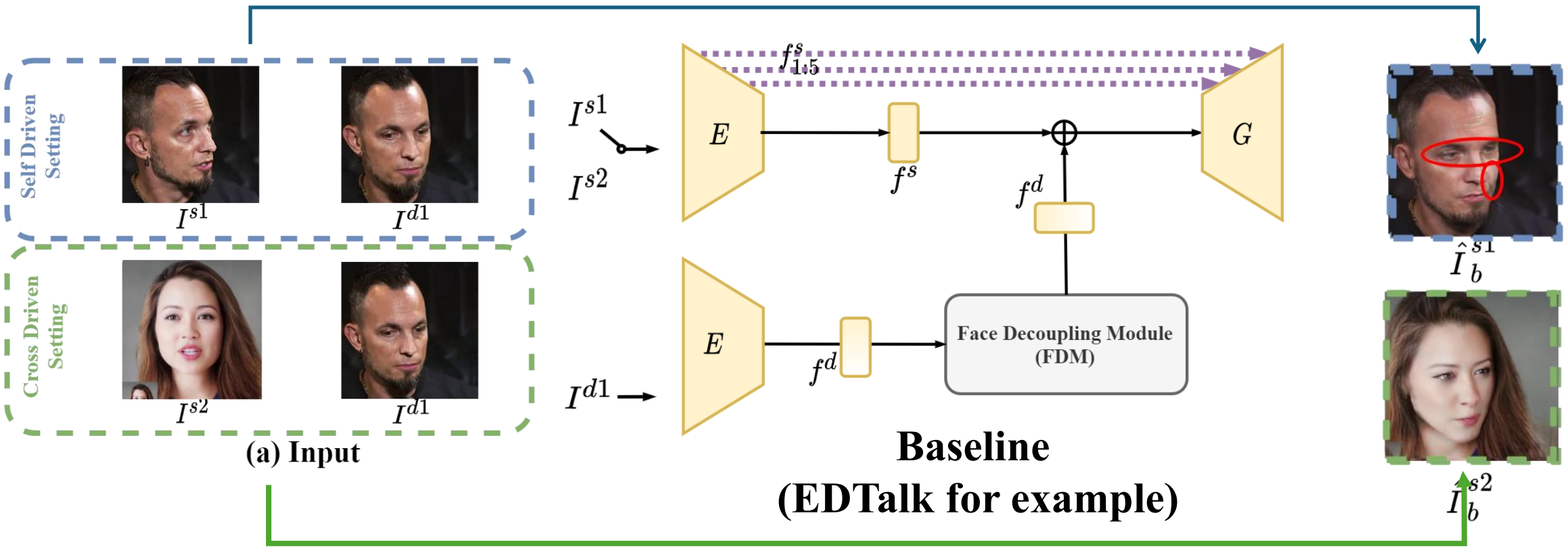




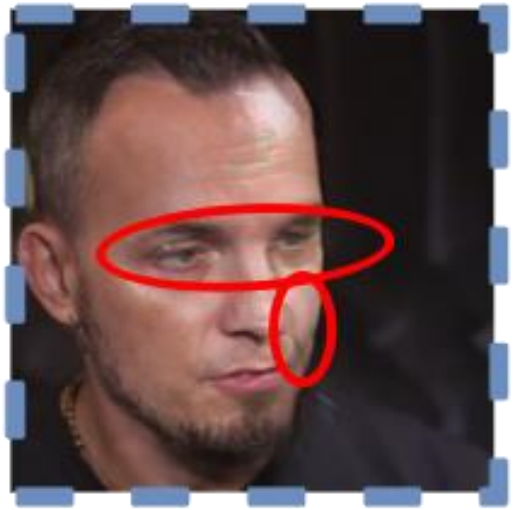


Carries identity information

Exploration in Baseline



Exploration in Baseline

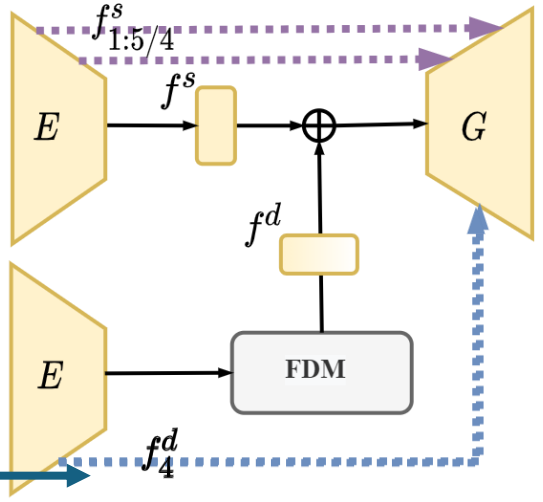


Self-Driven Result

>

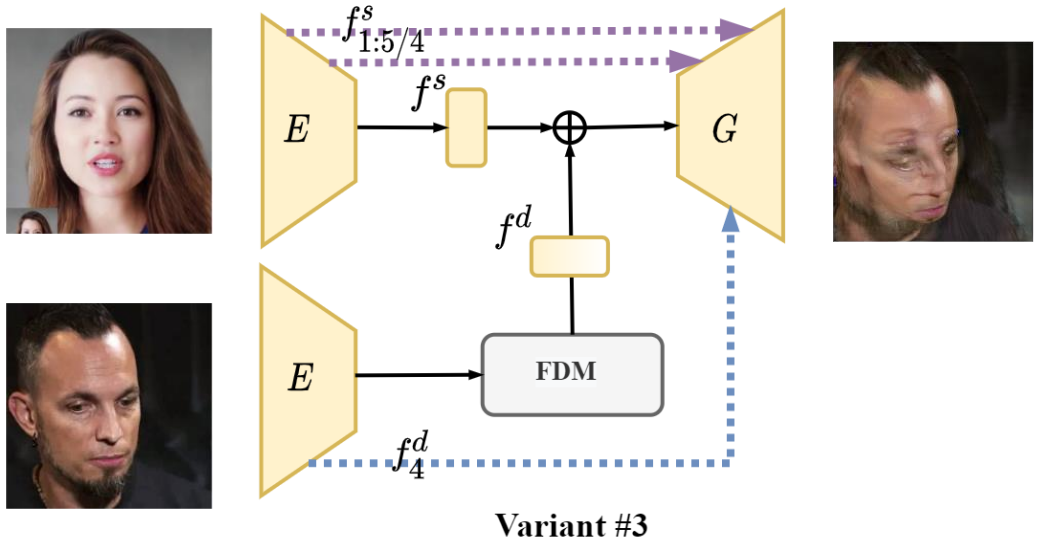
Cross-Driven Result

- Identity leakage from same identity can be beneficial for animation in self-driven scenarios
- We aim to extend this benefit to cross-driven animation
- Rely on

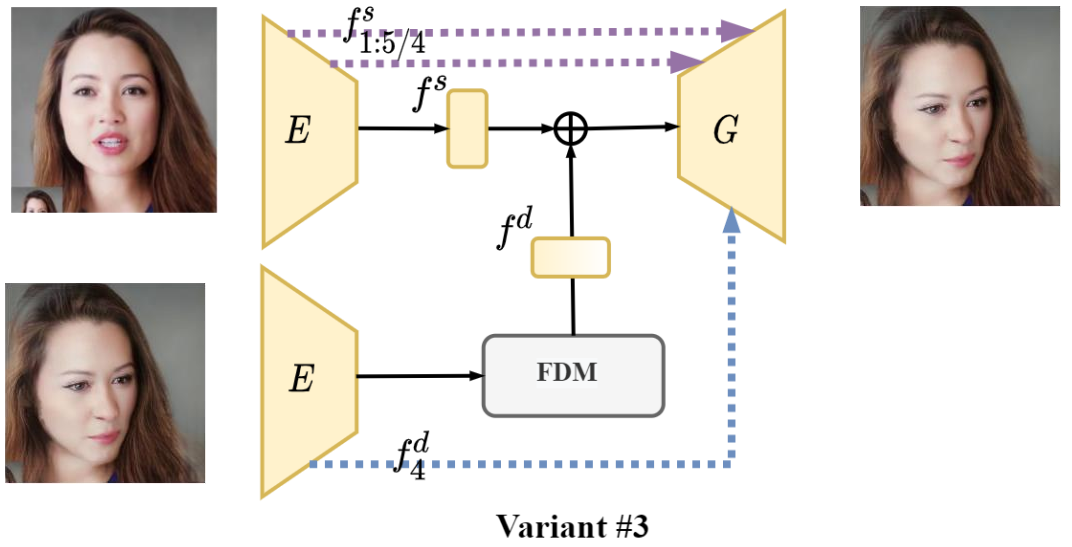


Variant #3

Exploration in Baseline

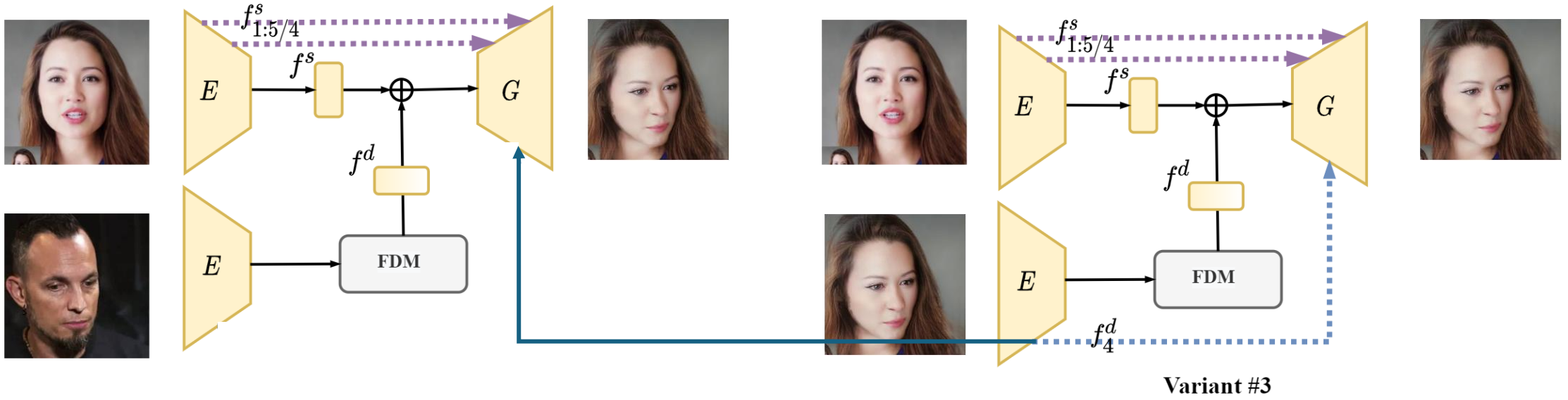


Cross-driven setting



Self-driven setting

Exploration in Baseline

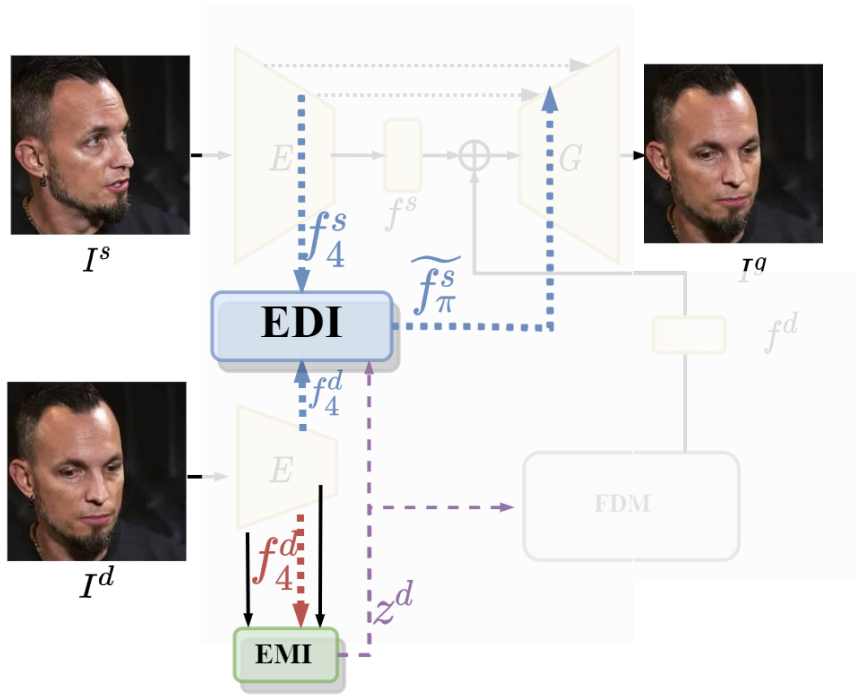


Cross-driven setting

Self-driven setting

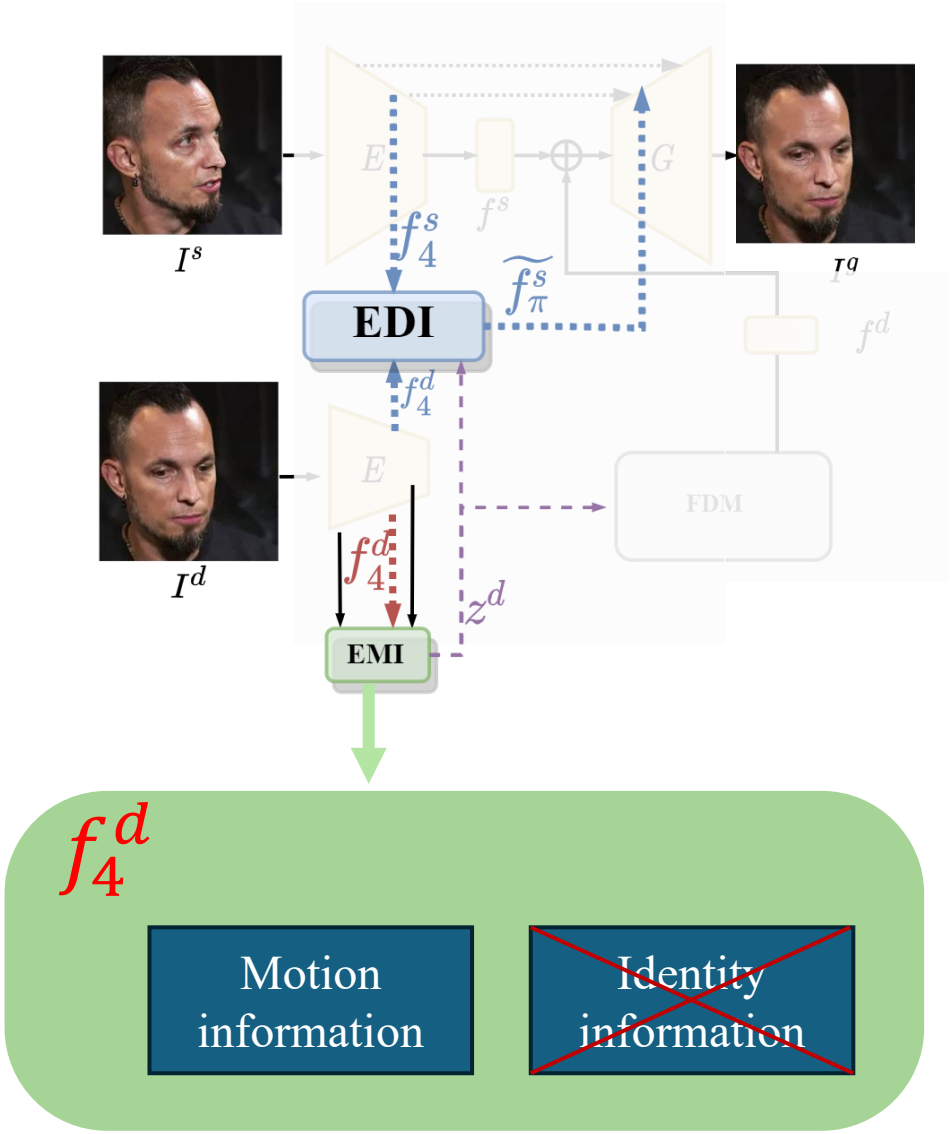
But how get the feature?

Framework of FixTalk



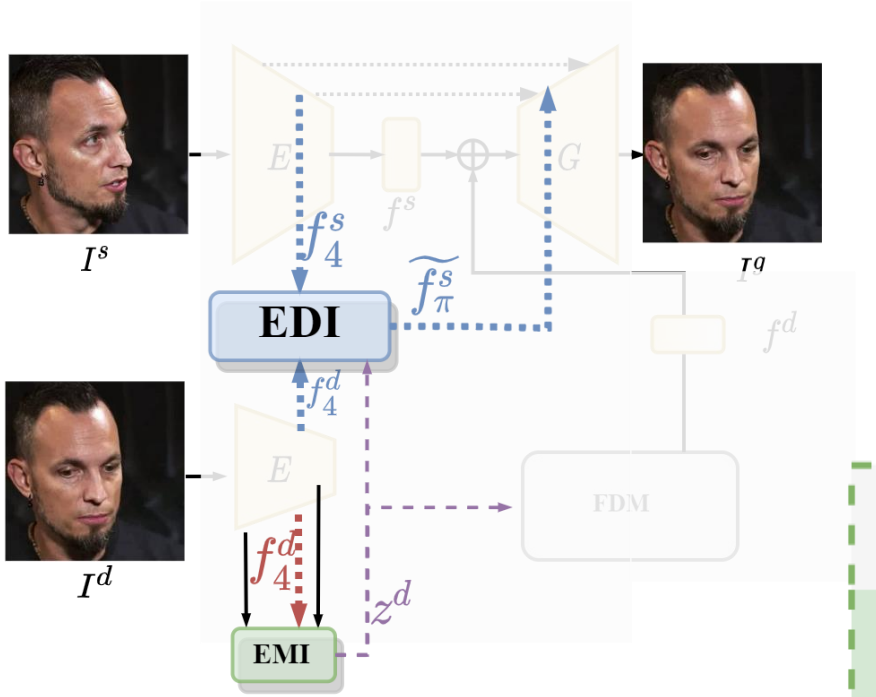
- **EMI: Enhanced Motion Indicator**
- **EDI: Enhanced Detail Indicator**

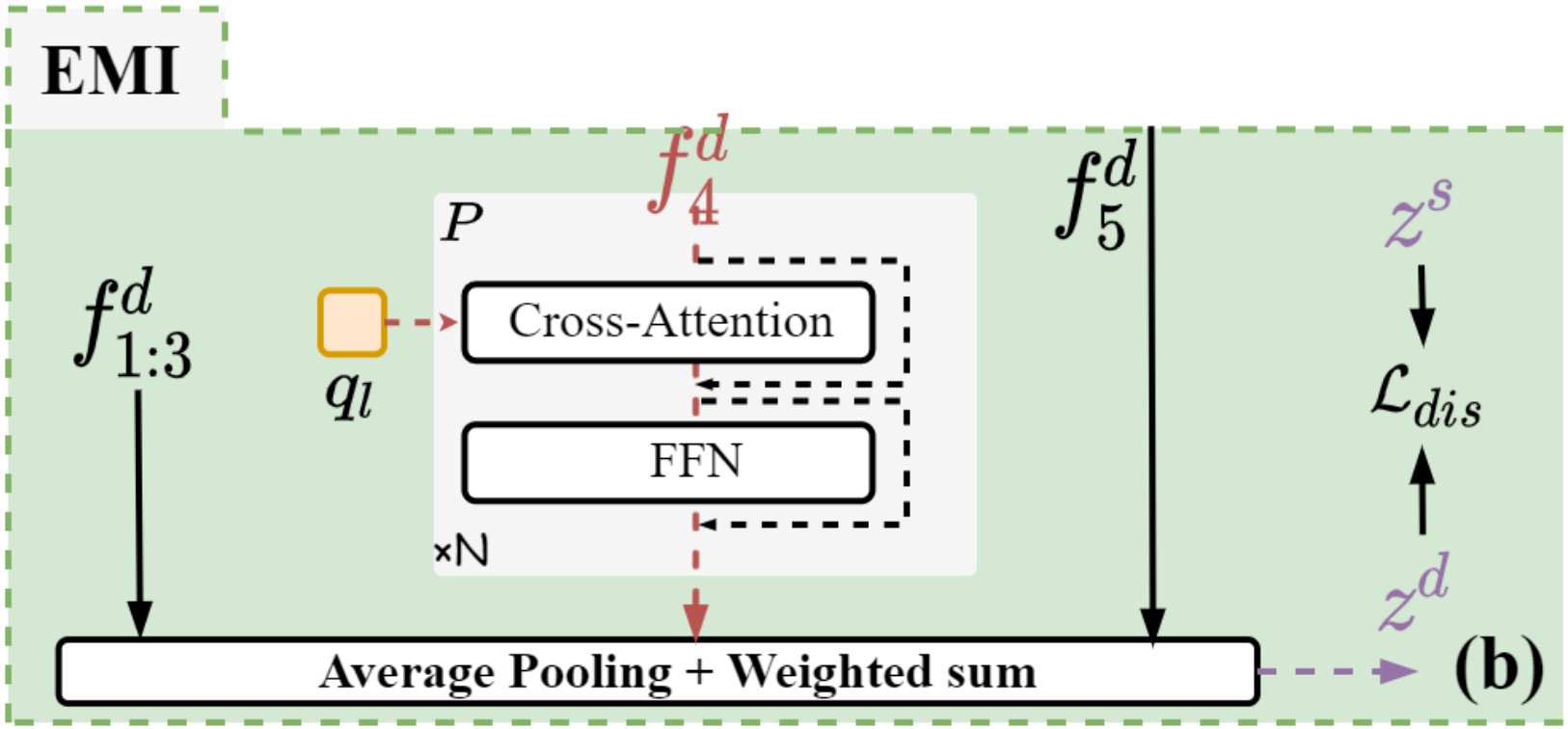
Framework of FixTalk



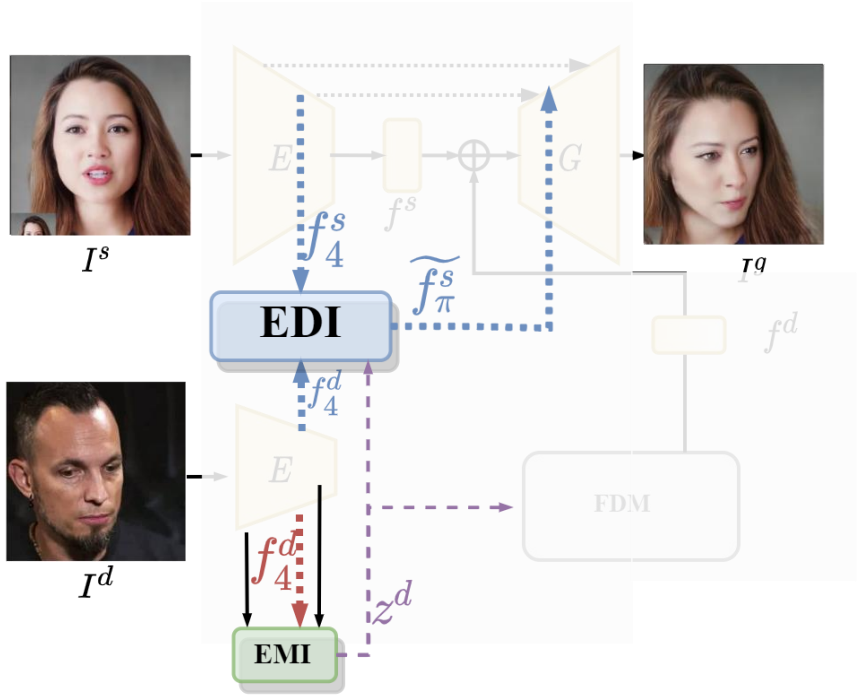
- Isolate motion-related features from the identity-containing fourth-layer feature
- → Address the limitations of identity leakage

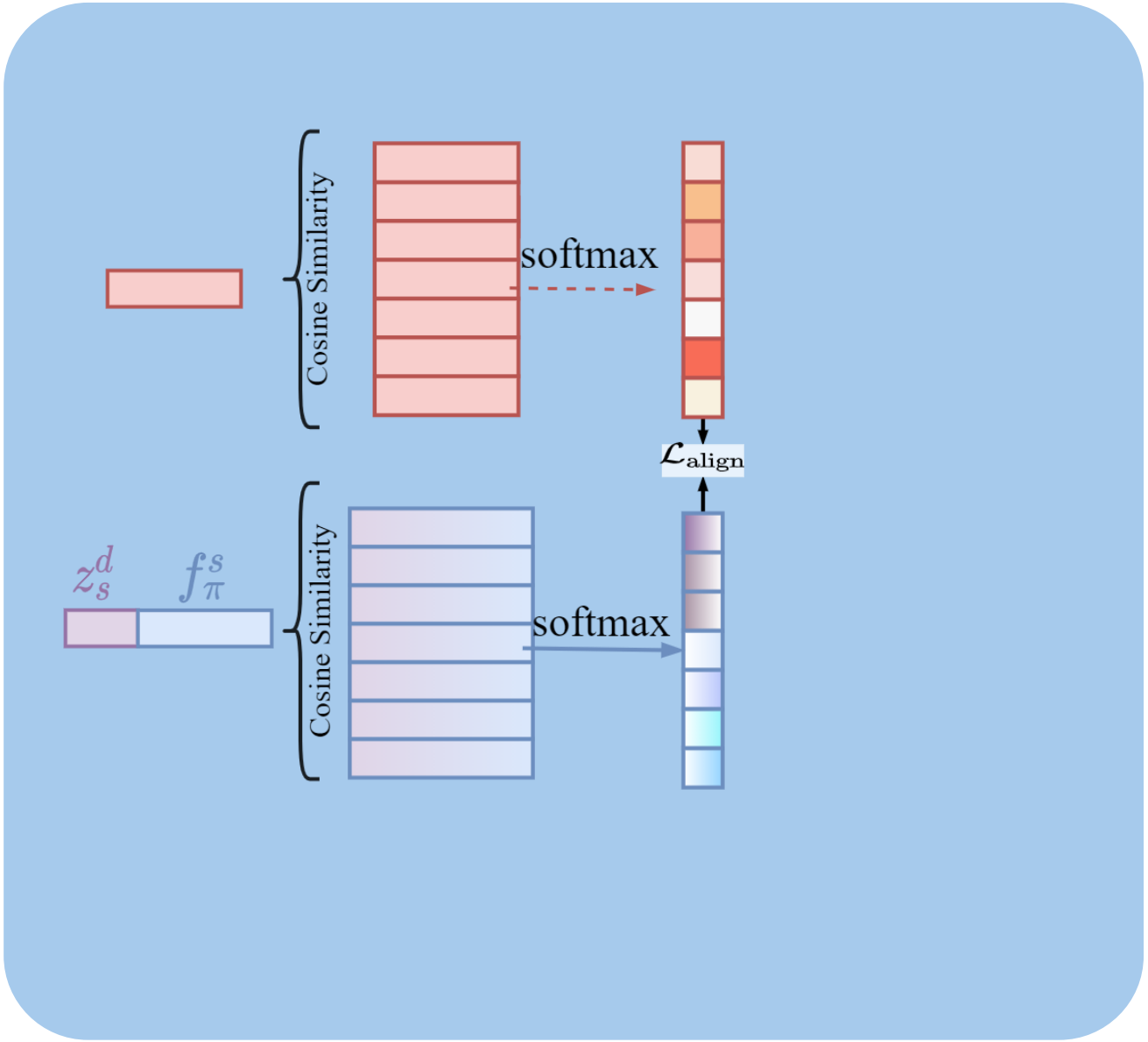
Framework of FixTalk





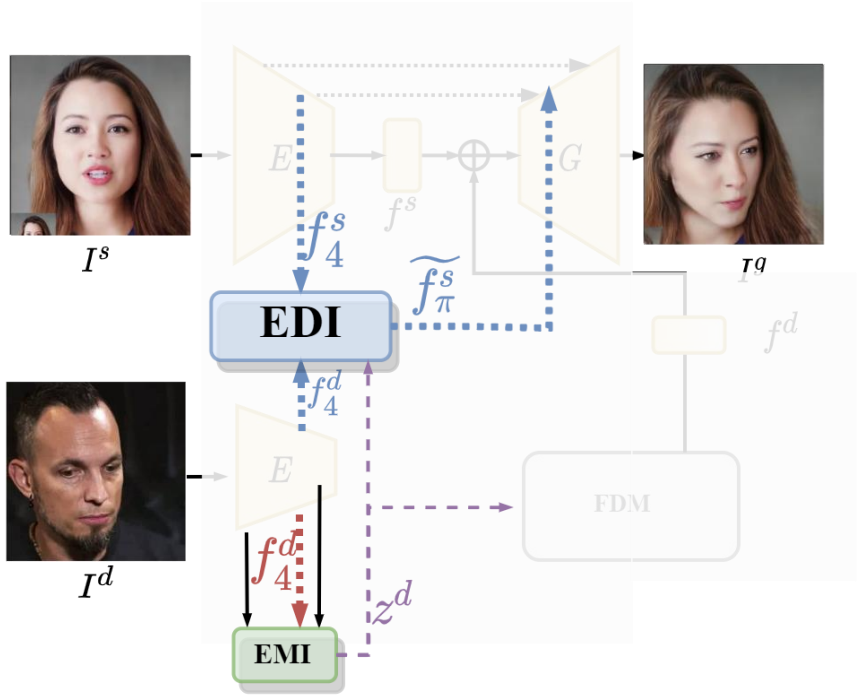
Framework of FixTalk

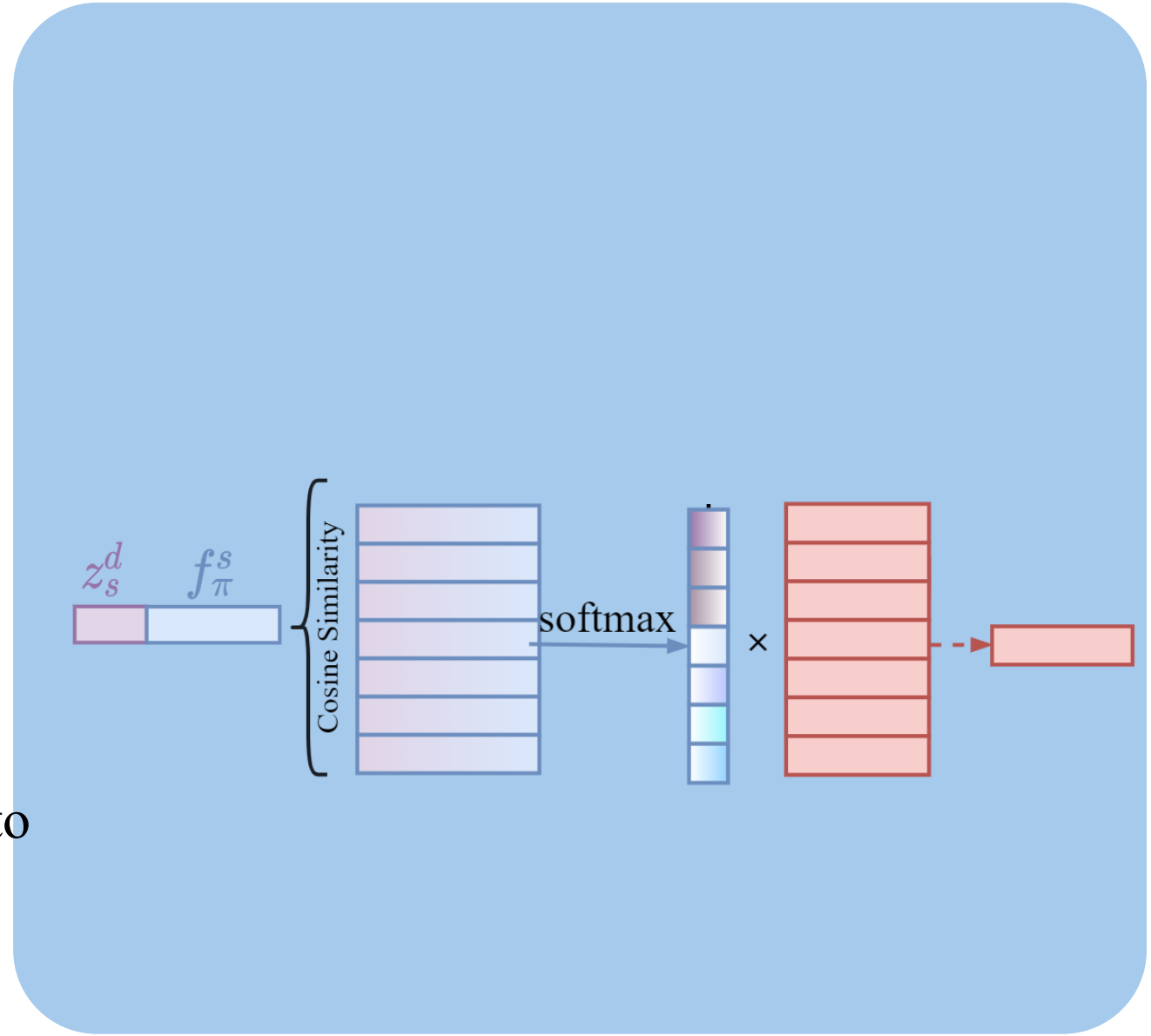




- Utilize extra storage space to retain features that contribute to identity leakage during training

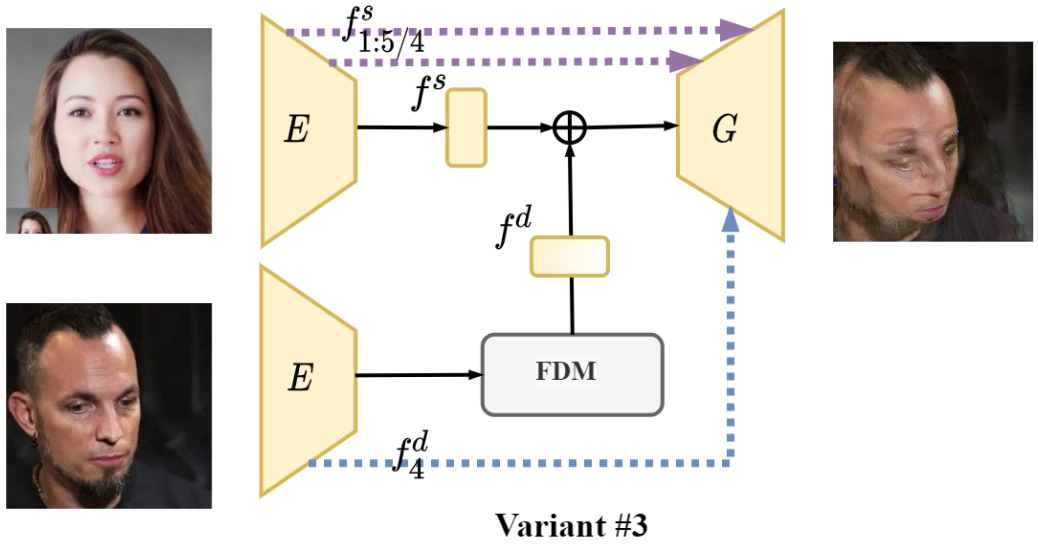
Framework of FixTalk



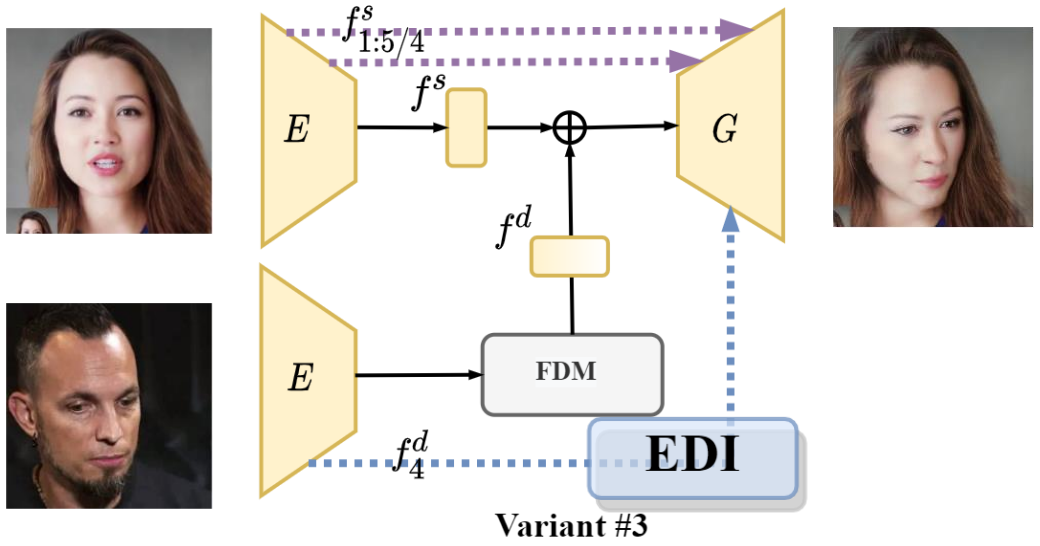


- Utilize extra storage space to retain features that contribute to identity leakage during training
- Design appropriate queries during inference to recall the most relevant identity-leaked features
- → Eliminate artifacts

Exploration in Baseline

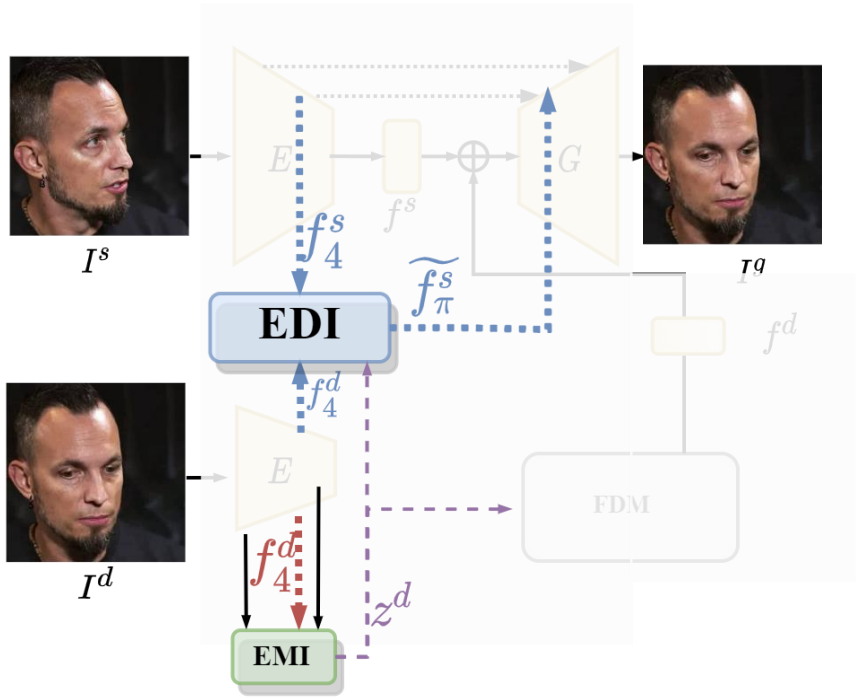


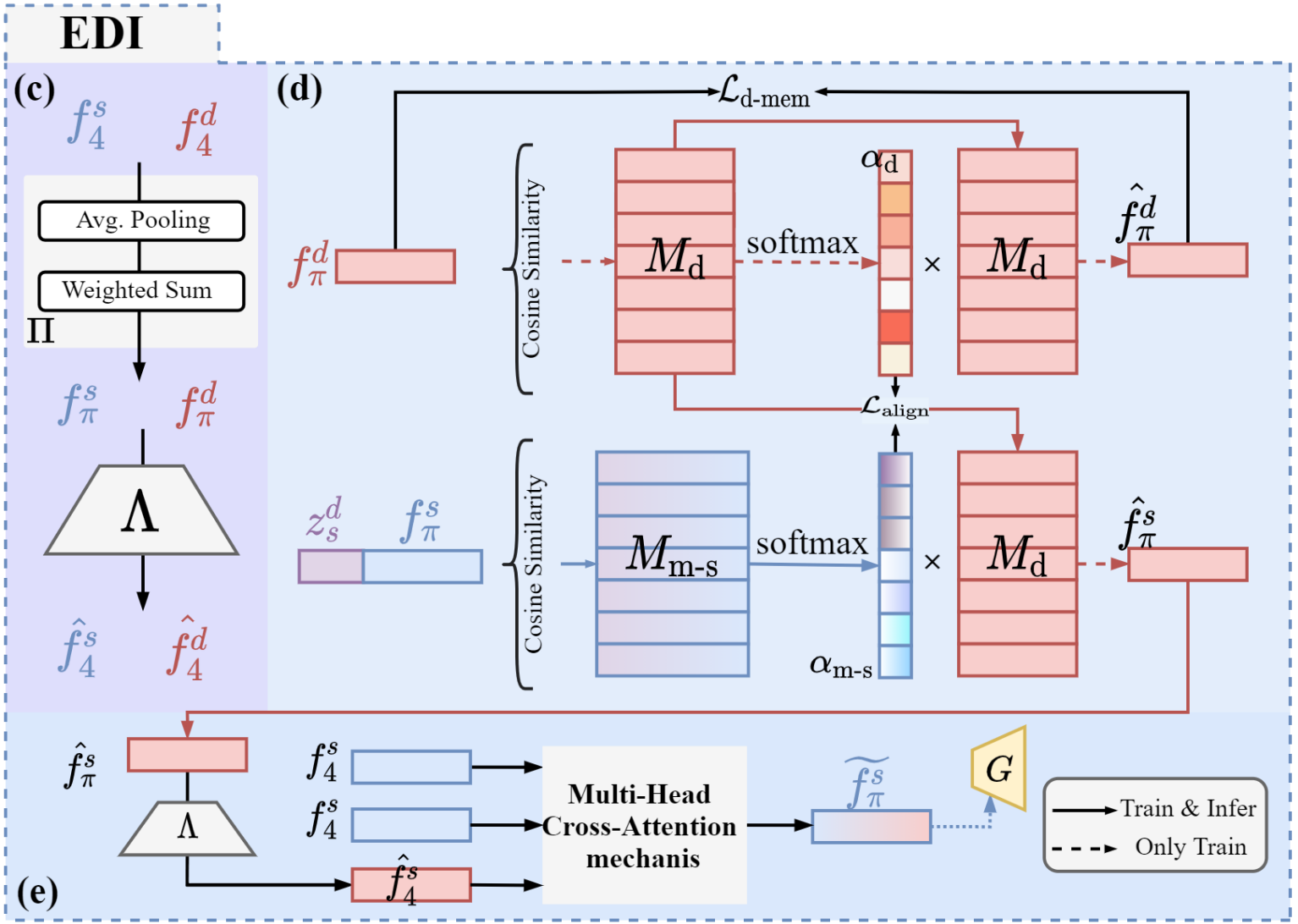
Without EDI



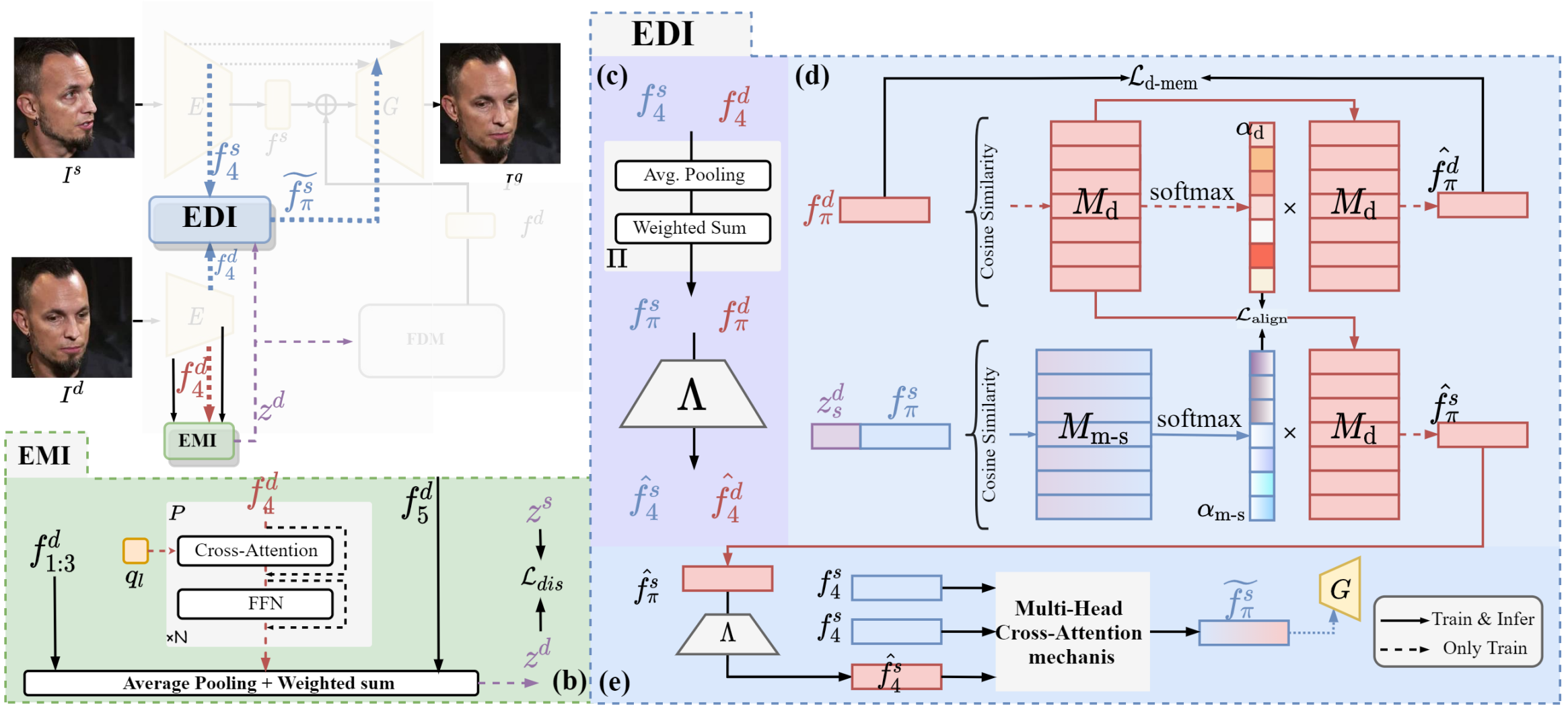
With EDI

Framework of FixTalk





Framework of FixTalk



Comparison with SOTA Video-Driven Talking Head Generation Methods

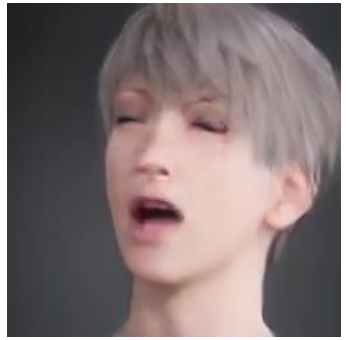
- All test data are unseen in the training set.

Comparison with SOTA Video-Dirven Talking Face Generation Methods

Source Image

Driven Video

FOMM [NeurIPS' 19]



Face-Vid2Vid [CVPR' 21]

LIA [ICLR' 22]

DaGAN [CVPR' 22]

Other methods suffer from Rendering Artifact.

DPE [CVPR' 23]

MCNET [ICCV' 23]

EDTalk [ECCV' 24]

Only Mouth

Only Pose

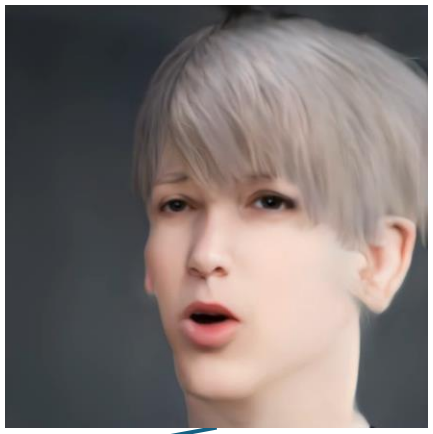
All

FixTalk

Comparison with SOTA Video-Dirven Talking Face Generation Methods

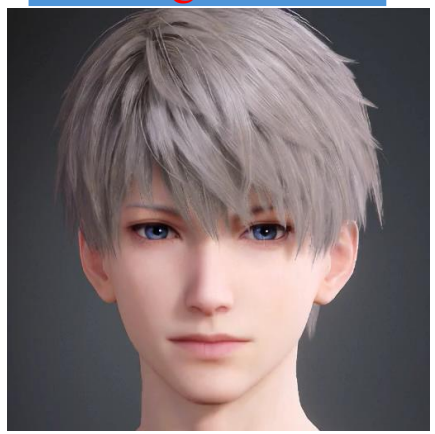
Source Image

Driven Video



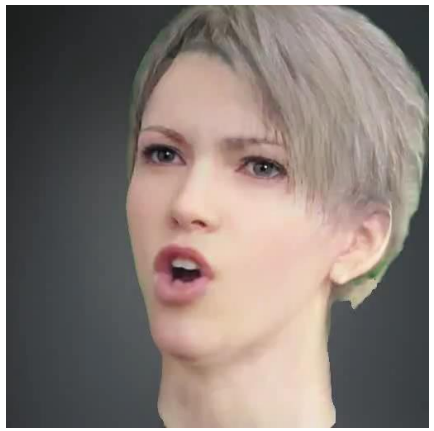
StyleHEAT [ECCV' 22]

Wrong motion



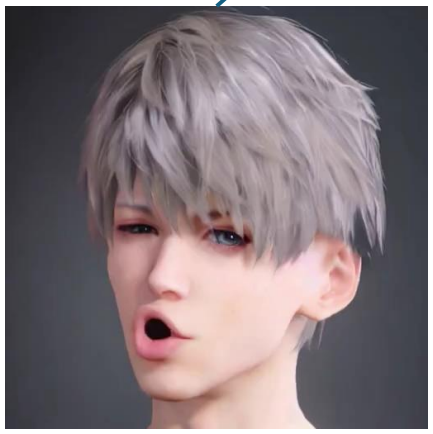
LivePortrait [arXiv' 24]

Identity Leakage

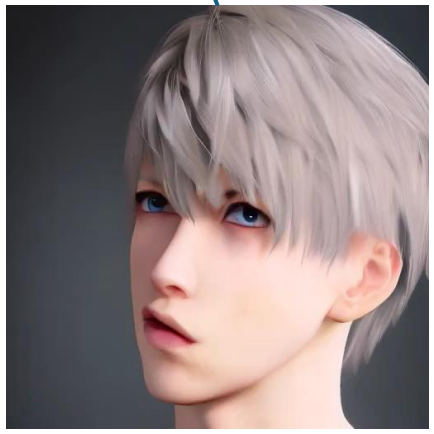


EmoPortrait [CVPR' 24]

Facial Deformation



X-Portrait [SIGGRAPH' 24]



FollowYourEmoji [SIGGRAPH Asian' 24]

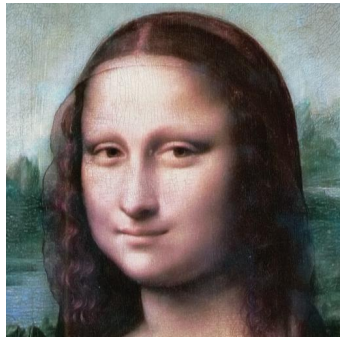
Only Mouth

Only Pose

All

FixTalk

Comparison with SOTA Video-Dirven Talking Face Generation Methods



Source Image

Driven Video

FOMM [NeurIPS' 19]

Face-Vid2Vid [CVPR' 21]

LIA [ICLR' 22]

DaGAN [CVPR' 22]

DPE [CVPR' 23]

MCNET [ICCV' 23]

EDTalk [ECCV' 24]

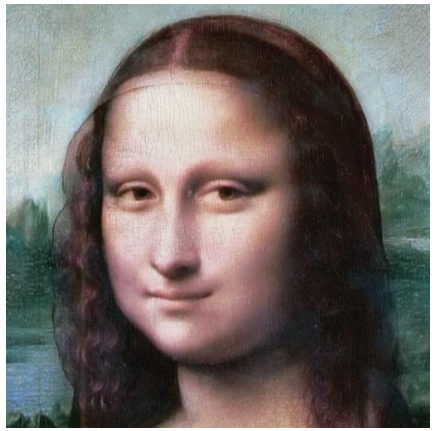
Only Mouth

Only Pose

All

FixTalk

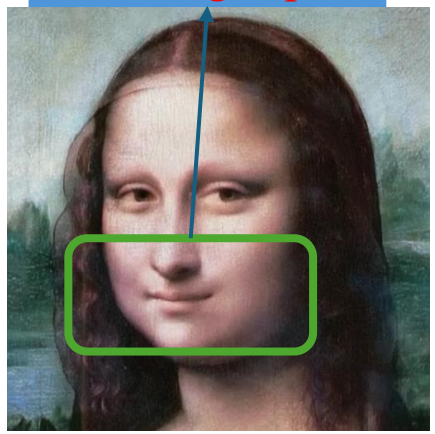
Comparison with SOTA Video-Dirven Talking Face Generation Methods



Source Image

Driven Video

StyleHEAT [ECCV' 22]



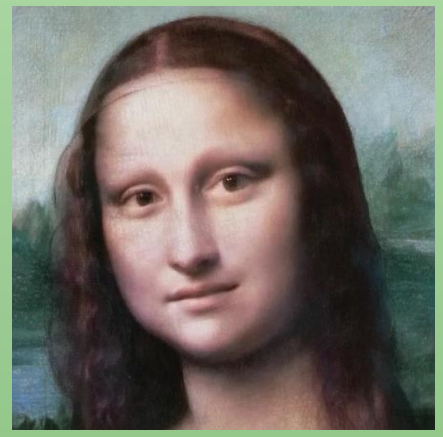
LivePortrait [arXiv' 24]

EmoPortrait [CVPR' 24]

X-Portrait [SIGGRAPH' 24]

FollowYourEmoji [SIGGRAPH Asian' 24]

Only Mouth



Only Pose

All

FixTalk

Comparison with SOTA Audio-Driven Talking Head Generation Methods

- All test data are unseen in the training set.

Comparison with SOTA Audio-Dirven Talking Face Generation Methods

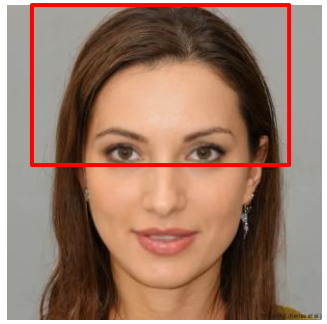
Source Image

Extreme Pose Source

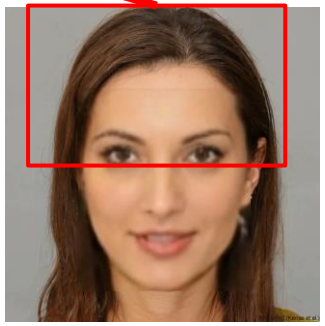
Expressive Expression Source

Fixed pose

Wav2Lip [MM' 20]



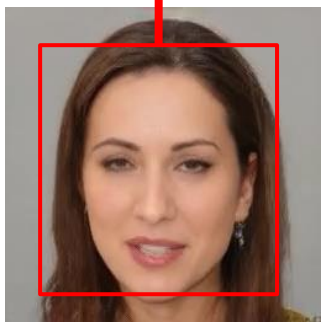
IP-LAP [CVPR' 23]



TalkLip [CVPR' 23]

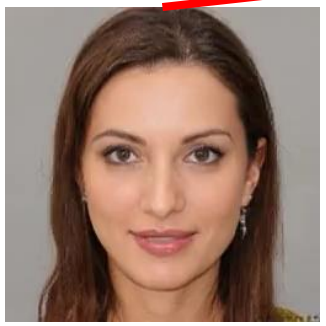
MuseTalk [arXiv' 24]

Identity Loss



Audio2Head [IJCAI' 21]

Frame Jitter

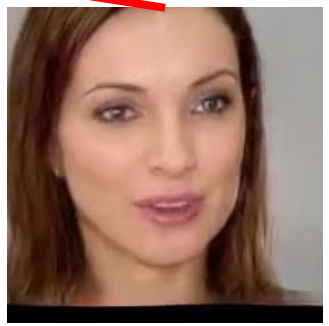


AVCT [AAAI' 22]

SadTalker [CVPR' 23]

Wrong pose

AniTalker [MM' 24]



PC-AVS [CVPR' 21]

FixTalk

Pose-Controllable Talking Head Methods

Comparison with SOTA Diffusion-based Talking Face Generation Methods

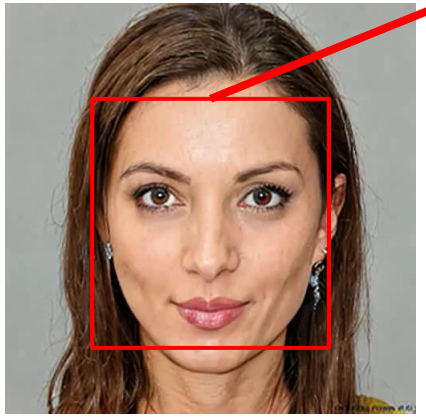
Source
Image

Expressive
Expression Source

Extreme
Pose Source

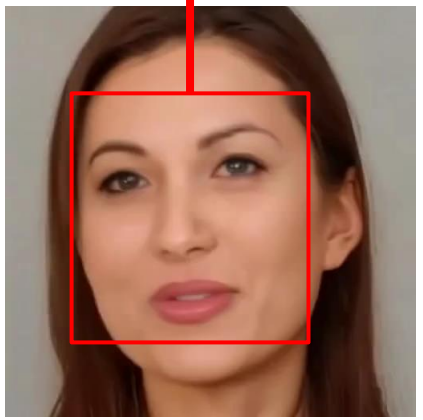
Artifacts

Identity Loss

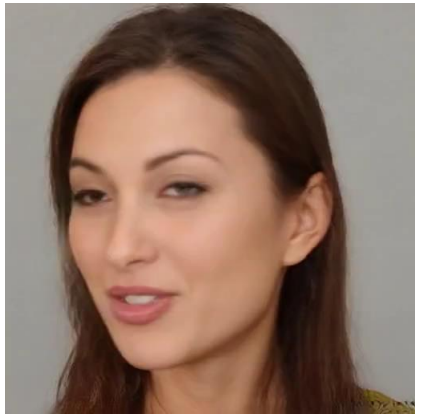


V-Express
[arXiv' 24]

Hallo
[arXiv' 24]



EchoMimic
[AAAI' 25]



FixTalk

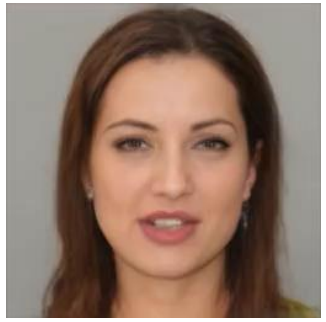
Diffusion-Based model: **1. cannot** achieve **expression/pose control**; **2. Long** inference time; 3. More training data

Comparison with SOTA Emotional Talking Face Generation Methods

Source Image

Extreme Pose Source

Expressive Expression Source



EAMM [SIGGRAPH' 22]

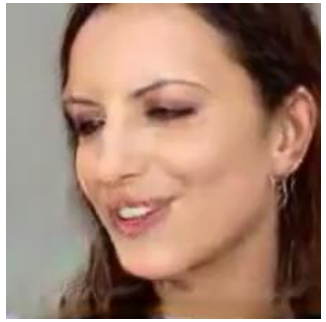
EmoGen [McGE'23]

StyleTalk [AAAI' 23]

SAAS [AAAI' 24]

DreamTalk [arXiv' 23]

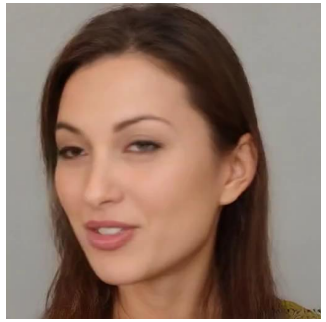
FlowVQTalker [CVPR' 24]



PD-FGC [CVPR' 23]

EAT [ICCV' 23]

EDTalk [ECCV' 24]



FixTalk

Pose-Controllable and expression-control Talking Head Methods

Comparison with SOTA Audio-Driven Talking Face Generation Methods

Source Image

Extreme Pose Source

Expressive Expression Source

Wav2Lip [MM' 20]

IP-LAP [CVPR' 23]

TalkLip [CVPR' 23]

MuseTalk [arXiv' 24]

Audio2Head [IJCAI' 21]

AVCT [AAAI' 22]

SadTalker [CVPR' 23]

AniTalker [MM' 24]

PC-AVS [CVPR' 21]

FixTalk

Pose-Controllable Talking Head Methods

Comparison with SOTA Emotional Talking Face Generation Methods

Source Image

Extreme Pose Source

Expressive Expression Source

EAMM [SIGGRAPH' 20]

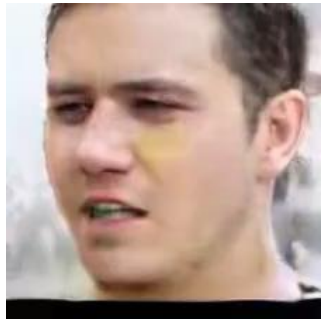
EmoGen [McGE'23]

StyleTalk [AAAI' 23]

SAAS [AAAI' 24]

DreamTalk [arXiv' 23]

FlowVQTalker [CVPR' 24]



PD-FGC [CVPR' 23]

EAT [ICCV' 23]

EDTalk [ECCV' 24]

FixTalk

Pose-Controllable and expression-control Talking Head Methods

Thanks for your watching