AU-Blendshape for Fine-grained Stylized 3D Facial Expression Manipulation

Hao Li[1,2], Ju Dai[2,*], Feng Zhou[3], Kaida Ning[2], Lei Li[4,5], Junjun Pan[1,2,*]
[1]Beihang University, [2]Peng Cheng Laboratory, [3]North China University of Technology,
[4]University of Washington, [5]University of Copenhagen
{lih09, daij, ningkd}@pcl.ac.cn, zhoufeng@ncut.edu.cn, lilei@di.ku.dk, pan_junjun@buaa.edu.cn

[*] Corresponding authors

AU-Blendshape for Fine-grained Stylized 3D Facial Expression Manipulation

Supplementary Material

1. Rationale

In this paper, we contribute a facial action unit blendshape guided dataset, AUBlendSet, for fine-grained stylized facial expression manipulation. Meanwhile, we propose AUBlendNet for generating stylized AU-Blendshape control bases. Extensive experiments demonstrate the potential and importance of AUBlendSet and AUBlendNet in 3D facial animation tasks.

This supplemental document contains six sections:

- **Section A** describes the selected facial AUs in detail;
- **Section B** gives the FACS coding table to build the mapping relations between common expressions and AUs;
- **Section C** illustrates the fine-grained AU-guided facial expression manipulation interface;
- **Section D** provides the specific contents of the user study;
- **Section E** conducts ablation analysis for the key component designing and parameter setting of AUBlendNet;
- **Section F** describes the supplemental video.

Section A

The facial AU describes facial muscle movement and provides a reasonable interactive control basis for facial expression manipulation. Considering that some eye AU movements only involve varying degrees of movement, to achieve maximum fine-grained expression manipulation while ensuring good interaction, we simplify the description of eye AU to reduce redundancy while removing the AU descriptions of tongue and head posture. Thus, we ultimately retain 32 AU for AU-Blendshape representations. Table 1 gives the detailed information of selected facial AUs.

Table 1. AU selection and description.

AU1	AU2	AU4	AU5
Inner brow raiser	Out brow raiser	Brow lowerer	Upper lip raiser
AU6	AU7	AU9	AU10
Cheek raiser	Lid tightener	Nose Wrinkler	Upper lip raiser
AU11	AU12	AU14	AU15
Nasolabial deepener	Lip corner depressor	Dimpler	Lip corner depressor
AU16	AU17	AU18	AU20
lower lip depressor	Chin raiser	Lip Pucker	Lip stretcher
AU22	AU23	AU24	AU25
Lip Funneler	Lip Tightener	Lip pressor	Lip part
AU26	AU27	AU28	AU29
Jaw Drop	Mouth stretch	Lip Suck	Jaw thrust
AU30L	AU30R	AU33	AU45
Jaw sideways (left)	Jaw sideways (right)	Cheek blow	Blink
AU61	AU62	AU63	AU64
Eyes turn left	Eyes turn right	Eyes up	Eyes down

Section B

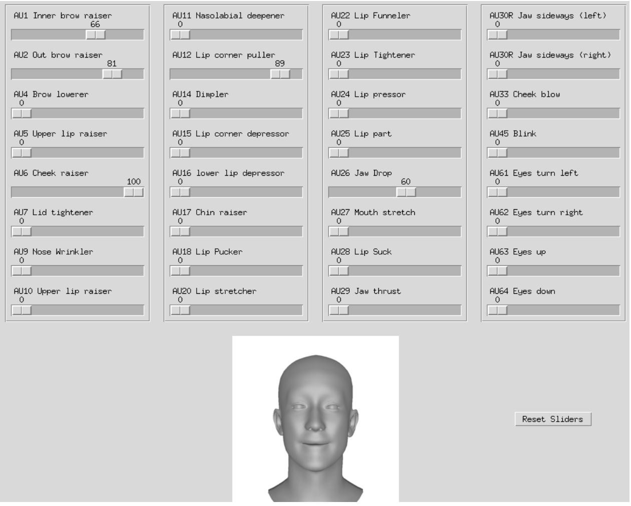
The FACS provides the mapping rules of facial AUs and expressions, enabling users to manipulate and generate corresponding target expressions with our AUBlendNet quickly [1]. Table 2 illustrates the mapping relationships for seven common expressions and action units.

Table 2. FACS mapping table for common expressions.

Expresions	Action Units
Happy	AU6 +AU12
Sad	AU1+AU4+AU15
Surprise	AU1+AU2+AU5+AU27
Fear	AU1+AU2+AU4+AU5+AU7+AU20+AU26
Anger	AU4+AU5+AU7+AU9+AU23
Disgust	AU9+AU5+AU16
Contempt	AU4+AU14+AU64

Section C

With AUBlendSet and AUBlendNet, we can achieve fine-grained facial expression manipulation with diversified emotions. We designed a control interface where users can freely combine different AUs and control their activation strength, as illustrated in Figure 1.



AU1 Inner brow raiser 66
AU2 Out brow raiser 81
AU4 Brow lowerer 0
AU5 Upper lip raiser 0
AU6 Cheek raiser 100
AU7 Lid tightener 0
AU9 Nose Wrinkler 0
AU10 Upper lip raiser 0

AU11 Nasolabial deepener 0
AU12 Lip corner puller 89
AU14 Dimpler 0
AU15 Lip corner depressor 0
AU16 lower lip depressor 0
AU17 Chin raiser 0
AU18 Lip Pucker 0
AU20 Lip stretcher 0

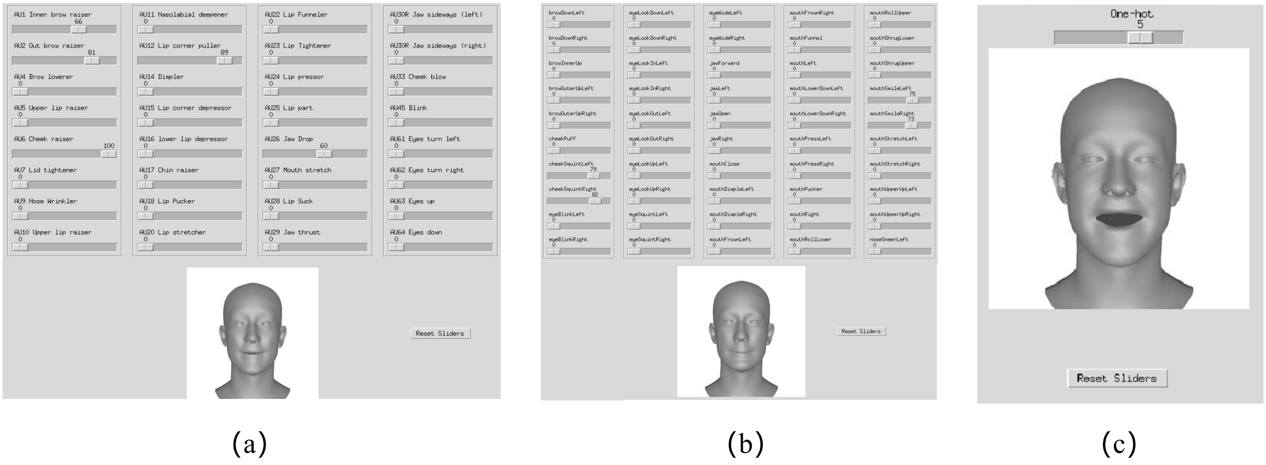
AU22 Lip Funneler 0
AU23 Lip Tightener 0
AU24 Lip pressor 0
AU25 Lip part 0
AU26 Jaw Drop 60
AU27 Mouth stretch 0
AU28 Lip Suck 0
AU29 Jaw thrust 0

AU30R Jaw sideways (left) 0
AU30R Jaw sideways (right) 0
AU33 Cheek blow 0
AU45 Blink 0
AU61 Eyes turn left 0
AU62 Eyes turn right 0
AU63 Eyes up 0
AU64 Eyes down 0

Figure 1. Illustration of AU control interface.

Section D

We compare our AUBlendNet with ARkit-based [2] and one-hot-based emotion control based on LGLDM [3] for facial expression manipulation. All the methods are developed to create corresponding facial programs, as shown in Figure 2. We recruited 32 participants to evaluate the emo-



(a)

AU1 Inner brow raiser 66
AU2 Out brow raiser 81
AU4 Brow lowerer 0
AU5 Upper lip raiser 0
AU6 Cheek raiser 100
AU7 Lid tightener 0
AU9 Nose Wrinkler 0
AU10 Upper lip raiser 0
AU11 Nasolabial deepener 0
AU12 Lip corner puller 89
AU14 Dimpler 0
AU15 Lip corner depressor 0
AU16 lower lip depressor 0
AU17 Chin raiser 0
AU18 Lip Pucker 0
AU20 Lip stretcher 0
AU22 Lip Funneler 0
AU23 Lip Tightener 0
AU24 Lip pressor 0
AU25 Lip part 0
AU26 Jaw Drop 60
AU27 Mouth stretch 0
AU28 Lip Suck 0
AU29 Jaw thrust 0
AU30R Jaw sideways (left) 0
AU30R Jaw sideways (right) 0
AU33 Cheek blow 0
AU45 Blink 0
AU61 Eyes turn left 0
AU62 Eyes turn right 0
AU63 Eyes up 0
AU64 Eyes down 0

Reset Sliders

(b)

browDownLeft 0
browDownRight 0
browInnerUp 0
browOuterUpLeft 0
browOuterUpRight 0
cheekPuff 0
cheekSquintLeft 79
cheekSquintRight 82
eyeBlinkLeft 0
eyeBlinkRight 0
eyeLookDownLeft 0
eyeLookDownRight 0
eyeLookInLeft 0
eyeLookInRight 0
eyeLookOutLeft 0
eyeLookOutRight 0
eyeLookUpLeft 0
eyeLookUpRight 0
eyeSquintLeft 0
eyeSquintRight 0
eyeWideLeft 0
eyeWideRight 0
jawForward 0
jawLeft 0
jawOpen 0
jawRight 0
mouthClose 0
mouthDimpleLeft 0
mouthDimpleRight 0
mouthFrownLeft 0
mouthFrownRight 0
mouthFunnel 0
mouthLeft 0
mouthLowerDownLeft 0
mouthLowerDownRight 0
mouthPressLeft 0
mouthPressRight 0
mouthPucker 0
mouthRight 0
mouthRollLower 0
mouthRollUpper 0
mouthShrugLower 0
mouthShrugUpper 0
mouthSmileLeft 75
mouthSmileRight 73
mouthStretchLeft 0
mouthStretchRight 0
mouthUpperUpLeft 0
mouthUpperUpRight 0
noseSneerLeft 0

Reset Sliders

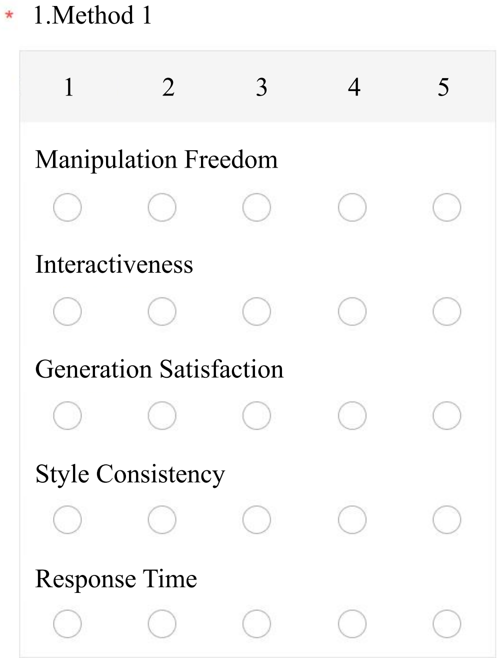
(c)

One-hot 5

Reset Sliders

Figure 2. Visualization diagrams of user study manipulation program. (a) AUBlendNet, (b) Arkit-based, (c) One-hot based.

tional face generation results of our AUBlendNet, Arkit-based, and one-hot-based methods. Participants are asked to manipulate seven different facial expressions for six characters and evaluate the degree of manipulation freedom, interactiveness, generation satisfaction, style consistency, and response time during the editing process using five scales. The detailed evaluation contents of the questionnaire are shown in Figure 3. We report the statistical results in Figure 7 in the paper. Our model consistently obtains the best or equivalent performance.



* 1.Method 1

1	2	3	4	5

Manipulation Freedom

○ ○ ○ ○ ○

Interactiveness

○ ○ ○ ○ ○

Generation Satisfaction

○ ○ ○ ○ ○

Style Consistency

○ ○ ○ ○ ○

Response Time

○ ○ ○ ○ ○

Figure 3. User study evaluation contents.

Table 3. Ablation experiments. **NoL** denotes the number of layers.

Model	$\text{MSE}_S \downarrow$ $(\times 10^{-9})$	$\text{MES}_M \downarrow$ $(\times 10^{-9})$
AUCodeBook NoL = 4	8.71	13.86
AUCodeBook NoL = 6	5.36	9.51
AUCodeBook NoL = 8	**4.55**	**7.62**
AdaLN-Zeros NoL =4	7.84	12.26
AdaLN-Zeros NoL =6	4.91	8.33
AdaLN-Zeros NoL =8	**4.55**	**7.62**
Cross-Attention	5.54	8.96
AdaLN-Zeros	**4.55**	**7.62**

Section E

We conduct a series of ablation experiments to validate the selection of key designs of AUBlendNet for stylized facial expression manipulation. The results are shown in Table 3.

Firstly, we compare the performance variations of the AUCodeBook with different layer numbers and set the layer number of AdaLN-Zeros to 8. It can be found that as the layer number increases, the 3D emotional face error of different metrics is also reduced. Secondly, we compare the AdaLN-Zeros with different layer numbers and set the layer number of AUCodeBook to 8. We observe that when the layer number increases, the model also performs better. Finally, we compare different conditional injection strategies, that is, the Cross-Attention and AdaLN-Zeros. The results in Table 3 show that condition injection through AdaLN-Zeros performs better than the Cross-Attention manner.

Section F

The demonstration video first introduces our stylized AUBlendSet dataset, consisting of many character themes. We then give several examples of the corresponding AU-

Blendshape control bases. Subsequently, we present the visualization of stylized facial expression manipulation results for seven emotions. Finally, we demonstrate the performance of combining AUBlendNet with traditional speech-driven methods for fine-grained stylized emotional facial animation.

References

[1] Paul Ekman and Wallace V Friesen. Facial action coding system. *EPNB*, 1978. 1

[2] Timo Menzel, Mario Botsch, and Marc Erich Latoschik. Automated blendshape personalization for faithful face animations using commodity smartphones. In *VRST*, pages 22:1–22:9, 2022. 1

[3] Wenfeng Song, Xuan Wang, Yiming Jiang, Shuai Li, Aimin Hao, Xia Hou, and Hong Qin. Expressive 3d facial animation generation based on local-to-global latent diffusion. *TVCG*, 2024. 1