

# 第11回定期ミーティング

2024/12/24

早稲田大学 基幹理工学研究科  
電子物理システム学専攻 史研究室  
石黒将太郎・野口颯汰

# 1. 先行研究紹介

---

A) NPHM

## 2. 実装状況

---

## 3. 今後の研究計画

---

# Learning Neural Parametric Head Models

Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Runz ,  
Lourdes Agapito, Matthias Nießner  
Technical University of Munich, Synthesia, University College London

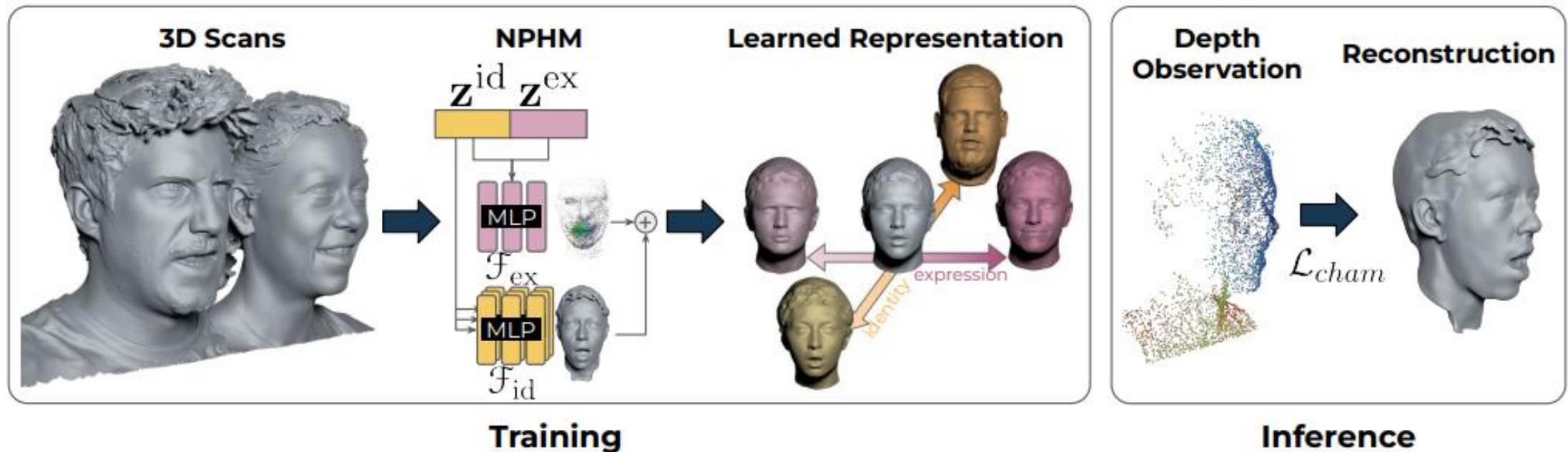


図1：NPHMのモデルアーキテクチャ

## 取り組んだ問題と背景

- ✓ 3D Faceを表現するためのパラメトリック空間として主流であるのは、DECAにも使用されているFLAMEであり、複数の3D Faceデータセットをテンプレートフィッティング後に、PCA分析をベースとした手法で組み立てられる
- ✓ PCAベースの手法は、入力データに対する剛性に優れるという特徴を持つが、局所的な表面の微細な表現に弱いことや、固定されたテンプレートメッシュを基にするため髪形や歯の生成ができない

## このモデルの特徴

- ✓ ニューラルパラメトリックヘッドモデル(NPHM)は、顔のキーポイントを中心とする局所座標が対象とする小さなMLPを複数導入することにより、局所的な表面の表現が得意
- ✓ SDFによってcanonical spaceでの頭部ジオメトリを表現してから、ポーズ空間で形状変化を学ぶため、アイデンティティと表情を表現する潜在空間の分離度が高い
- ✓ 255人を対象とした平均3.5Mサイズの頭部スキャンを5200以上用意し、これらをもとにトレーニングすることでトレーニング時の正確なフィッティング誤差を得る

## 3Dデータセット

Num. Subjects	255 (188m/67f)
Total num. Scans	5200
Num. Vertices/Scan	$\approx 1.5\text{M}$

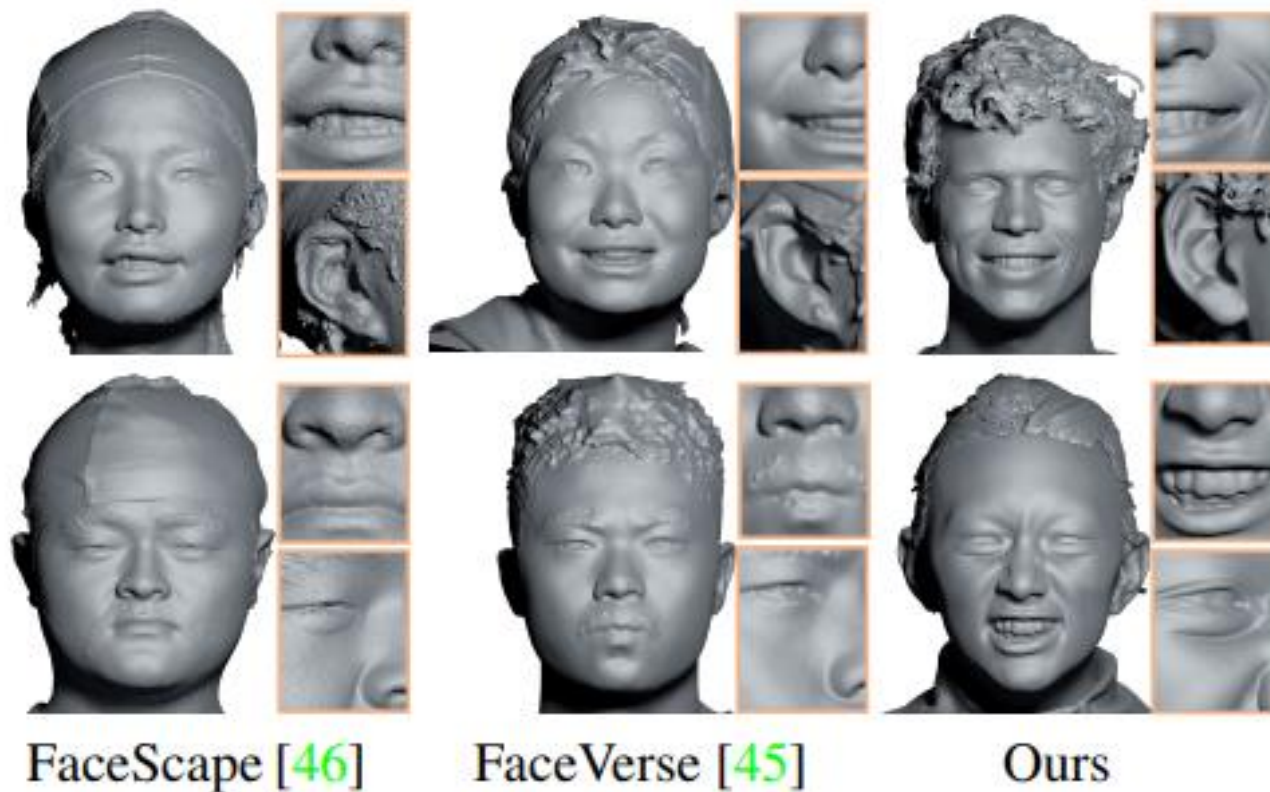


表2：NPHMの3Dデータセットについて

## モデルアーキテクチャ

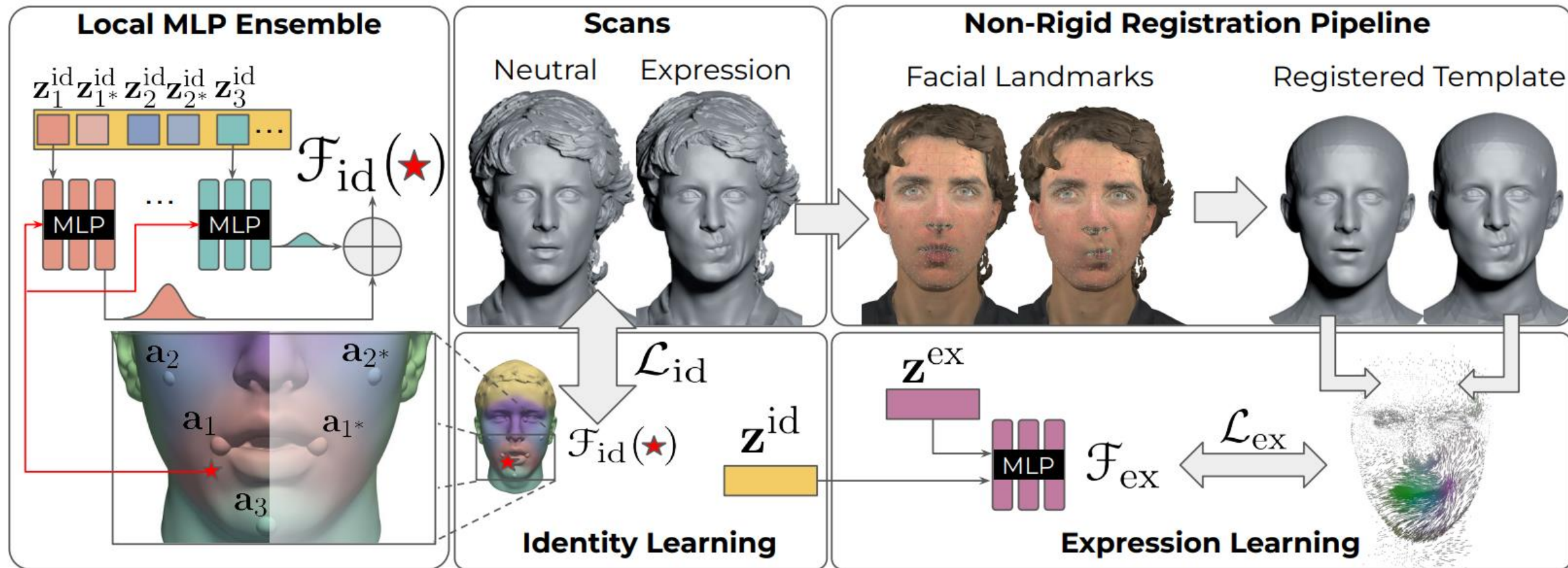


図2：NPHMのトレーニング過程



## 損失関数

$$\mathcal{L}_{\text{id}} = \sum_{j \in J} \mathcal{L}_{\text{IGR}} + \lambda_a \|\hat{\mathbf{a}}_j - \mathbf{a}_j\|_2^2 + \lambda_{\text{sy}} \mathcal{L}_{\text{sy}} + \lambda_{\text{reg}}^{\text{id}} \|\mathbf{Z}_j^{\text{id}}\|_2^2,$$

$$\mathcal{L}_{\text{ex}} = \sum_{\substack{i, j \in J, L \\ x \in X_{j, l}}} \|\mathcal{F}_{\text{ex}}(x, \mathbf{z}_{j, l}^{\text{ex}}, \hat{\mathbf{z}}_j^{\text{id}}) - \delta(x)_{j, l}\|_2^2 + \lambda_{\text{reg}}^{\text{ex}} \|\mathbf{z}_{j, l}^{\text{ex}}\|_2^2$$

表2：NPHMの3Dデータセットについて

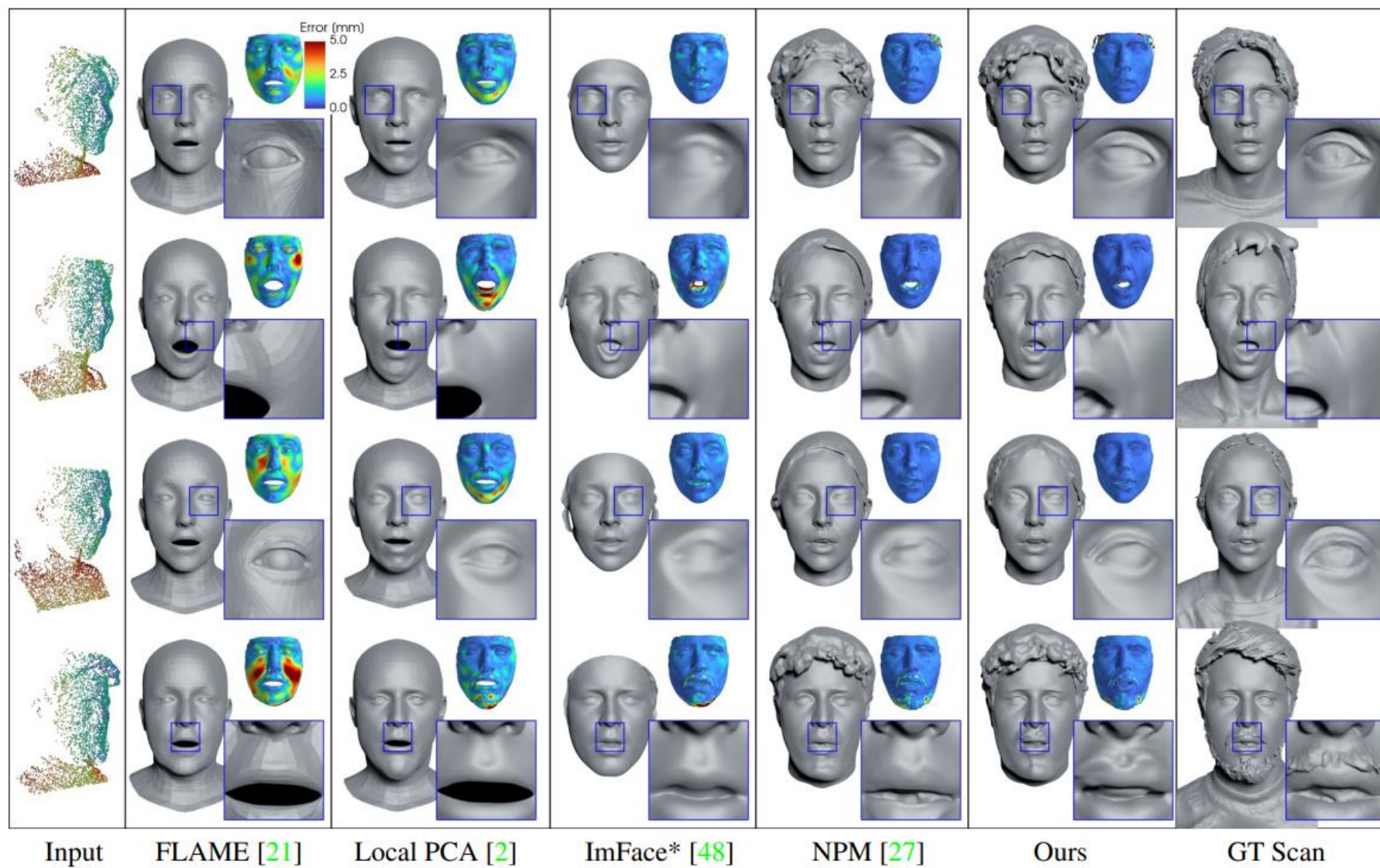
## 定量評価(identity)

Method	$L_1$ -Chamfer ↓	N. C. ↑	F-Score@1.5 ↑
BFM [32]	1.341e−2	0.936	0.319
FLAME [21]	0.640e−2	0.931	0.530
Global PCA [2]	0.563e−2	0.954	0.571
Local PCA [2]	0.416e−2	0.960	0.756
ImFace [48]	0.404e−2	0.954	0.832
ImFace* [48]	0.312e−2	0.971	0.883
NPM [27]	0.200e−2	0.975	0.947
Ours	<b>0.182e−2</b>	<b>0.978</b>	<b>0.954</b>

\* trained on our data



## 定性評価(identity)

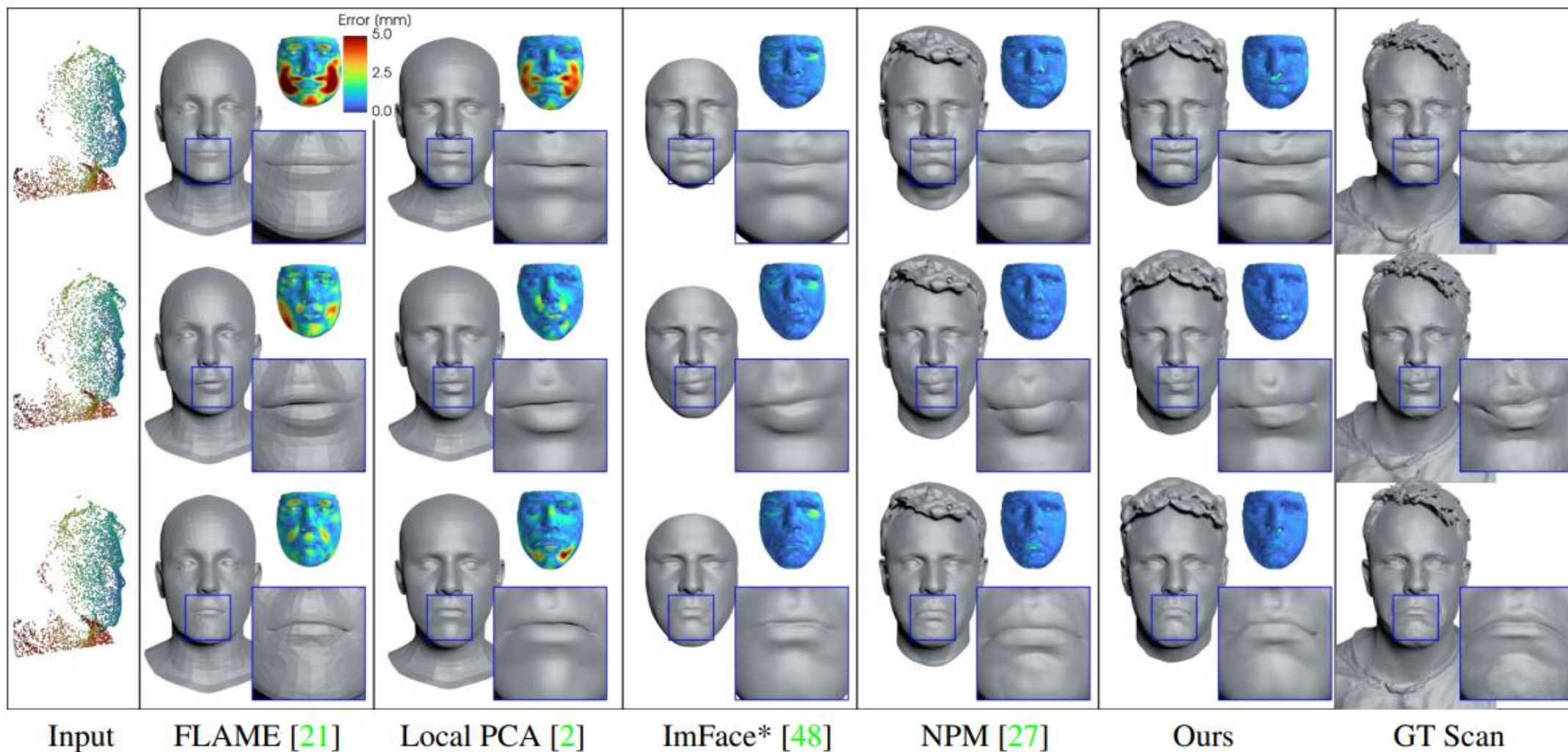


## 定量評価(expression)

Method	$L_1$ -Chamfer ↓	N. C. ↑	F-Score@1.5 ↑
BFM [32]	1.271e−2	0.937	0.508
FLAME [21]	0.679e−2	0.924	0.351
Global PCA [2]	0.515e−2	0.956	0.606
Local PCA [2]	0.535e−2	0.950	0.641
ImFace [48]	0.369e−2	0.959	0.824
ImFace* [48]	0.321e−2	<b>0.971</b>	0.879
NPM [27]	0.299e−2	0.962	0.891
Ours	<b>0.272e−2</b>	0.969	<b>0.913</b>

\* trained on our data

## 定性評価(expression)



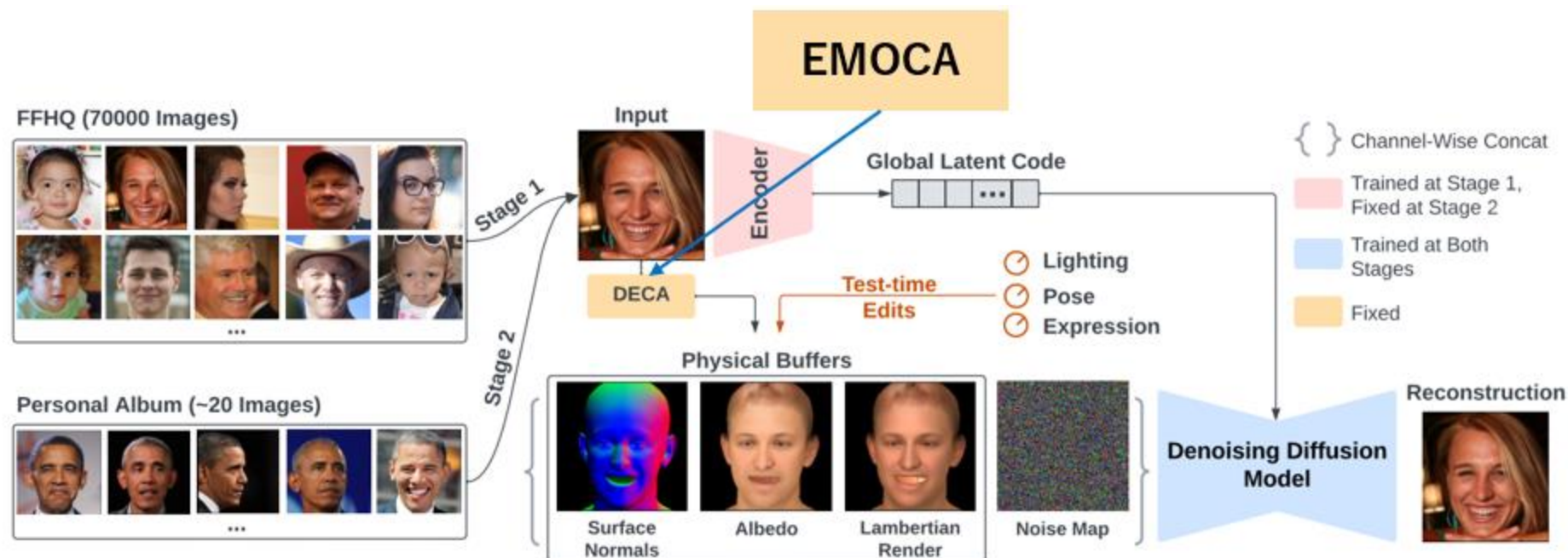


## FFHQのEMOCA由来の物理条件を用いたStage1完了

使用GPU : RTX 3080ti Laptop  
Image size :  $256 \times 256$   
Batch size : 8  
Max step : 5000  
Global Encoder : ResNet18

学習時間 : 約8日間

- AffectNetデータセット追加
- ResNet50 verもやりたい



## 無表情への変換

Source



Target



Stage1 : DECA  
S : DECA、 T : DECA



Stage1 : EMOCA  
S : DECA、 T : EMOCA



## 笑顔への変換

Source



Target



Stage1 : DECA  
S : DECA、 T : DECA



Stage1 : EMOCA  
S : DECA、 T : EMOCA





## 笑顔への変換

Source



Target



Stage1 : DECA  
S : DECA、 T : DECA



Stage1 : EMOCA  
S : DECA、 T : EMOCA



## しかめっ面への変換

Source

Target

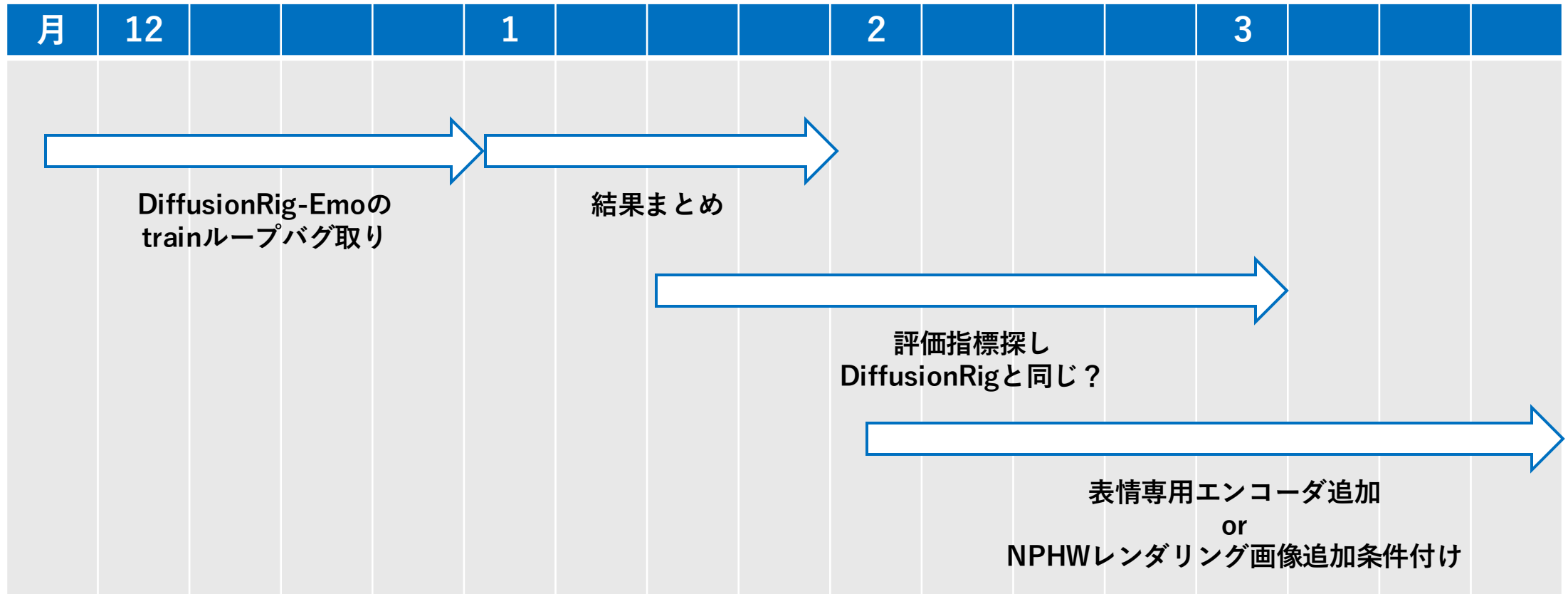
Stage1 : DECA  
S : DECA、 T : DECA

Stage1 : EMOCA  
S : DECA、 T : EMOCA



## 2月までには現在のモデル性能について結論づけたい

Step1



## 1. 研究テーマ

---

## 2. 実装状況

---

## 3. 今後の研究計画

---

## Step1：表情編集に特化したDDIMを訓練

## Step2：変換前後で $\beta$ 変化しないように訓練

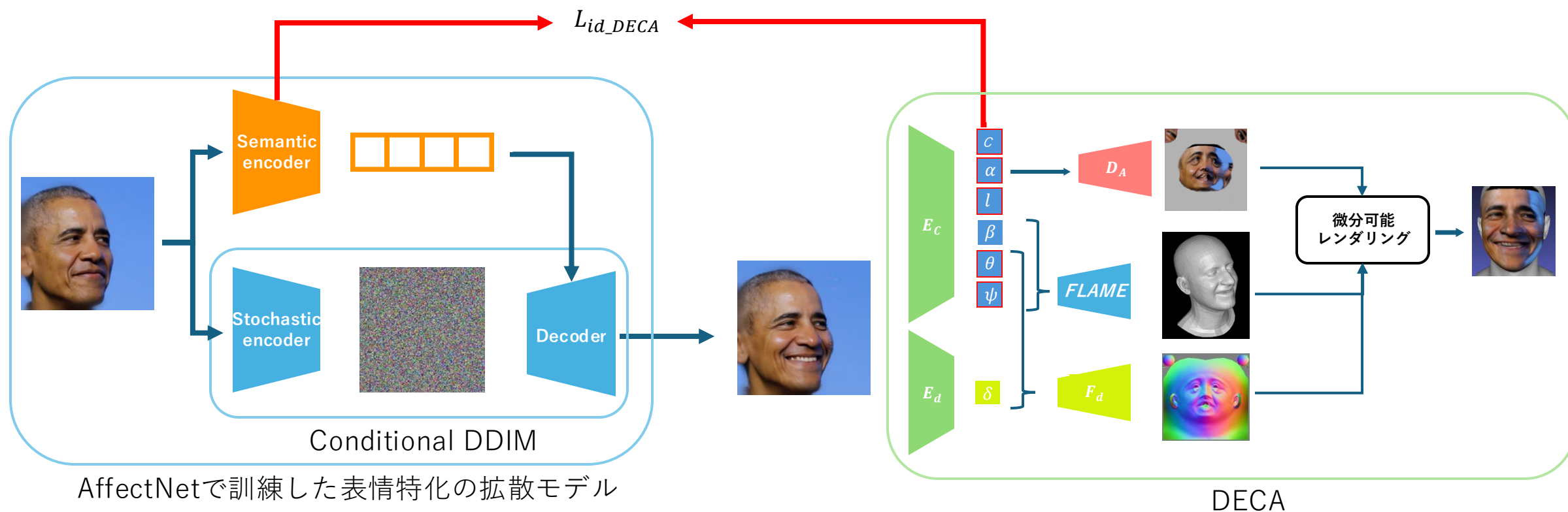


図. DECA[1]とDiffusionAutoencoders[2]を元にした提案モデル

[1] YAO FENG et al. Learning an Animatable Detailed 3D Face Model from In-The-Wild Images

[2] Konpat Preechakul et al. Diffusion Autoencoders: Toward a Meaningful and Decodable Representation

### 提供モデル

- ✓ CelebA データセットで分類機を訓練  
(分類機のみいけそう)
- ✓ 40種類の属性
- ✓ アノテーションテキストの中身(30000)

```
Brown_Hair Male Mouth_Slightly_Open Smiling ...  
0.jpg -1 1 1 -1 ...  
:  
:
```

- ✓ 実際のトレーニングの流れ
  1. LMDB形式に変更
  2. 画像サイズを128×128もしくは256×256に変換
  3. Pytorch\_lightningで訓練

### 自作モデル

- ✓ AffectNetで分類機を訓練
- ✓ 8種類の感情
- ✓ アノテーション中身
  - aro.npy : Arousalの値
  - exp.npy : 表情ラベル(インデックス)
  - lnd.npy : ランドマークの座標
  - val.npy : Valenceの値



表. AffectNetの表情カテゴリー



Neutral	75374
Happy	134915
Sad	25959
Surprise	14590
Fear	6878
Disgust	4303
Anger	25382
Contempt	4250
None	33588
Uncertain	12145
Non-Face	82915
Total	420299

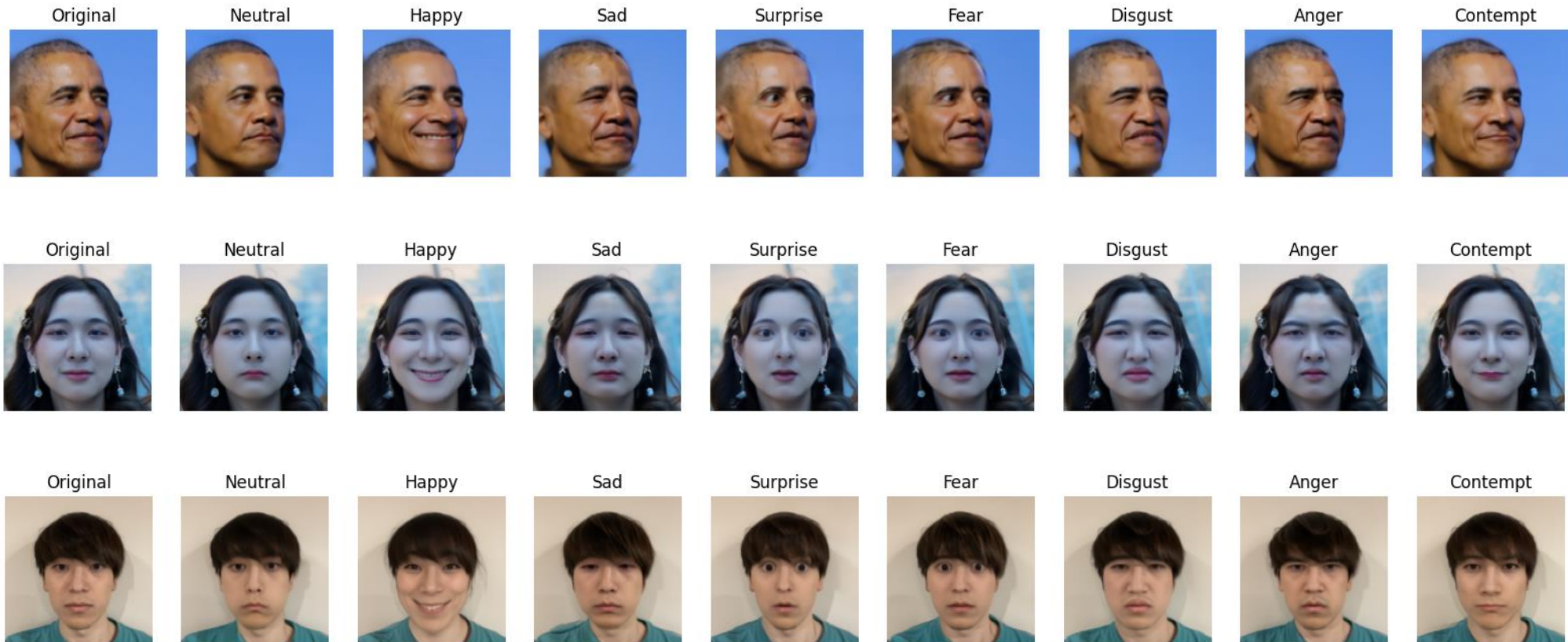


287651  
Label\_0 Label\_1 Label\_2 Label\_3 Label\_4 Label\_5 Label\_6 Label\_7  
0.jpg -1 1 -1 -1 -1 -1 -1 -1  
1.jpg 1 -1 -1 -1 -1 -1 -1 -1  
2.jpg 1 -1 -1 -1 -1 -1 -1 -1  
3.jpg -1 1 -1 -1 -1 -1 -1 -1  
5.jpg -1 -1 -1 -1 -1 -1 1 -1  
7.jpg -1 -1 -1 -1 -1 -1 1 -1  
10.jpg -1 1 -1 -1 -1 -1 -1 -1  
13.jpg -1 1 -1 -1 -1 -1 -1 -1  
15.jpg 1 -1 -1 -1 -1 -1 -1 -1  
16.jpg 1 -1 -1 -1 -1 -1 -1 -1  
18.jpg -1 -1 -1 -1 -1 -1 1 -1  
21.jpg -1 -1 1 -1 -1 -1 -1 -1  
22.jpg -1 1 -1 -1 -1 -1 -1 -1  
23.jpg -1 1 -1 -1 -1 -1 -1 -1  
27.jpg 1 -1 -1 -1 -1 -1 -1 -1

図. ラベル付した結果

LMDB形式に変換

 data.mdb	10.83 GB	MDB ファイル	2024-12-16 23:02:38
 lock.mdb	8 KB	MDB ファイル	2024-12-16 23:02:38



8種類の表情変換が可能に & Contempt・Fearなど不自然

```
(diffae) ls@labshi@ls@labshi-SYS-5049A-TR:~/workspace-cloud/hayata.noguchi/diffae_affectnet$ python run_ffhq128_cls.py
conf: ffhq128_autoenc_cls
Global seed set to 0
loading pretrain ... 130M
step: 1019986
loading latent stats ...
/home/ls@labshi/anaconda3/envs/diffae/lib/python3.8/site-packages/pytorch_lightning/callbacks/model_checkpoint.py:446: UserWarning: Checkpoint directory checkpoints/ffhq128_autoenc_cls exists and is not empty.
  rank_zero_warn(f"Checkpoint directory {dirpath} exists and is not empty.")
/home/ls@labshi/anaconda3/envs/diffae/lib/python3.8/site-packages/pytorch_lightning/callbacks/model_checkpoint.py:432: UserWarning: ModelCheckpoint(save_last=True, save_top_k=None, monitor=None) is a redundant configuration. You can save the last checkpoint with ModelCheckpoint(save_top_k=None, monitor=None).
  rank_zero_warn(
Using native 16bit precision.
GPU available: True, used: True
TPU available: False, using: 0 TPU cores
IPU available: False, using: 0 IPUs
local seed: 0
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0,1,2,3]

| Name          | Type          | Params
-----|-----|-----
0 | model          | BeatGANsAutoencModel | 128 M
1 | ema_model      | BeatGANsAutoencModel | 128 M
2 | classifier     | Linear        | 4.1 K
3 | ema_classifier | Linear        | 4.1 K
-----|-----|-----
8.2 K   Trainable params
257 M   Non-trainable params
257 M   Total params
1,028,394 Total estimated model params size (MB)
Epoch 1:  4%|███████| 386/8989 [00:27<10:02, 14.27it/s, loss=0.19, v_num=]
```

- 途中で訓練が終了した原因調査
- 最適なエポック等  
ハイパーパラメーターの調整

評価指標	目的	計算手法/特徴	使用目的
PSNR	ピクセルレベルの類似度	平均二乗誤差 (MSE) を基に計算	再構築品質評価
SSIM	構造的類似性	輝度・コントラスト・構造の3要素	再構築品質評価
LPIPS	知覚的類似性	学習済みネットワークの特徴空間での距離	再構築品質評価
感情分類精度	感情転送性能	HSEmotionでターゲット感情との一致率を計算	感情操作の正確性評価
CSIM	被写体のアイデンティティ保持	CosFaceモデルでの特徴ベクトル間のコサイン類似度	被写体特徴の保持性能評価
ユーザースタディ	リアリズムと感情表現の主観的評価	ペア比較法・感情識別タスク	視覚的品質と感情表現の検証

表情変換の精度を評価する  
コードを作成

# 今年度中に計画してる部分の実装を目指す

