

# 第12回定期ミーティング

2025/01/14

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. 研究テーマ

---

## 2. 実装状況

---

## 3. 今後の研究計画

---

表情変換顔生成モデル(Diffusion)と2Dto3D顔形状を生成するモデルを組み合わせることで、3Dの表情変換を軽量化かつ自然に行う

## 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

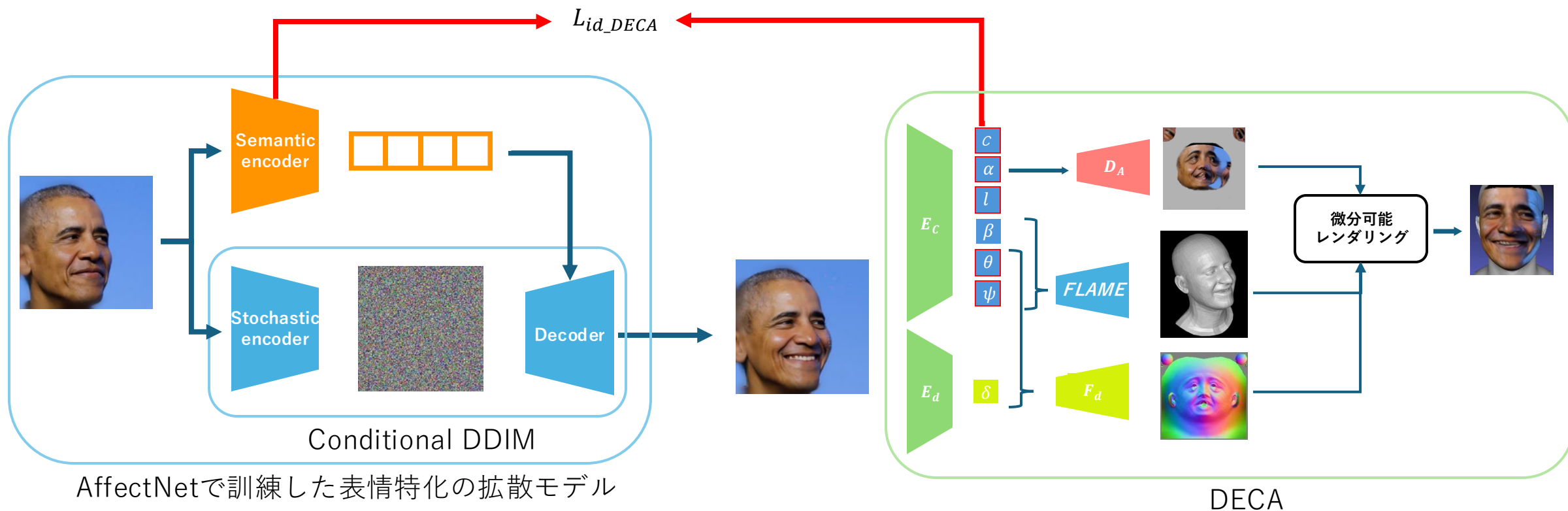


図. DECA[1]とDiffusionAutoencoders[2]を元にした提案モデル

[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

今後、すべての表情変換に対応できるように枚数を均一する必要あり

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.jpg	1	-1	-1	-1	-1	-1	-1
2.jpg	1	-1	-1	-1	-1	-1	-1
3.jpg	-1	1	-1	-1	-1	-1	-1
5.jpg	-1	-1	-1	-1	-1	1	-1
7.jpg	-1	-1	-1	-1	-1	1	-1
10.jpg	-1	1	-1	-1	-1	-1	-1
13.jpg	-1	1	-1	-1	-1	-1	-1
15.jpg	1	-1	-1	-1	-1	-1	-1
16.jpg	1	-1	-1	-1	-1	-1	-1
18.jpg	-1	-1	-1	-1	-1	-1	-1
21.jpg	-1	-1	1	-1	-1	-1	-1
22.jpg	-1	1	-1	-1	-1	-1	-1
23.jpg	-1	1	-1	-1	-1	-1	-1
27.jpg	1	-1	-1	-1	-1	-1	-1

図. ラベル付した結果



表. 訓練に使ったデータセット

表情	枚数
Neutral	74874
Happy	134415
Anger	25382
Surprise	14090
Fear	6378
Contempt	3750
Sad	25459
Disgust	3803

## CelebデータセットとAffectNetデータセットの総枚数の差によるエポック数の変更

### 【前回のエラー】

```
(diffae) islabshi@islabshi-5Y5-5049A-TR:~/workspace-cloud/hayata.noguchi/diffae_affectnet$ python run_ffhq128_cls.py
conf: ffhq128_autoenc_cls
Global seed set to 0
loading pretrained ... 130M
step: 1019986
loading latent stats ...
/home/islabshi/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/islabshi/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

Celebデータセット

	枚数
データセットサイズ	30000
バッチサイズ	32
総サンプル数	300000
エポック数	10

AffectNet変更前

	枚数
データセットサイズ	287651
バッチサイズ	32
総サンプル数	300000
エポック数	1.04

AffectNet変更後

	枚数
データセットサイズ	287651
バッチサイズ	32
総サンプル数	2876510
エポック数	10

多少の変化が得られたが、主観評価では、優劣判断不可

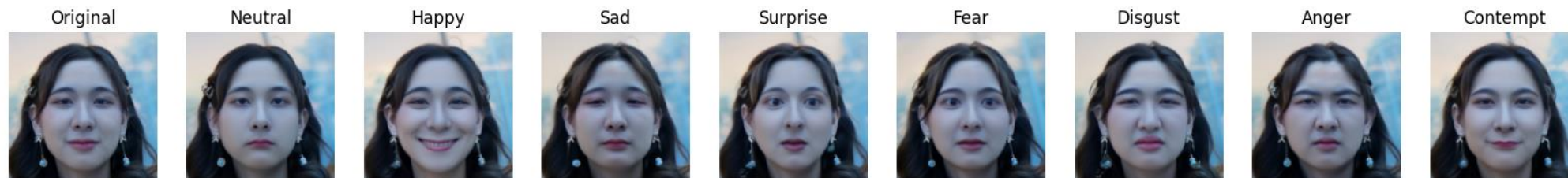
Epoch:1.04



Epoch:10



Epoch:1.04



Epoch:10

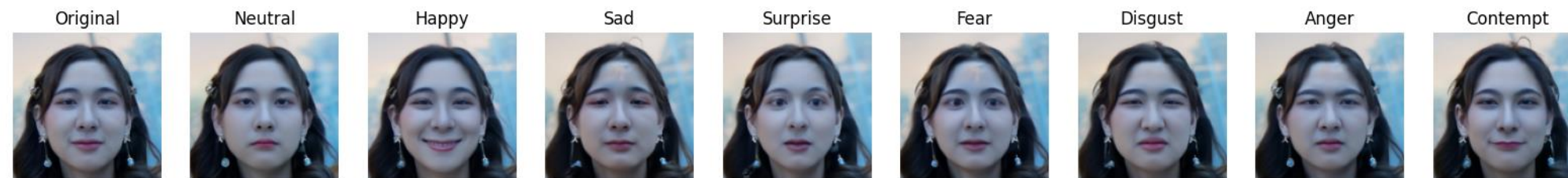


表. 表情変換拡散モデルでの評価指標

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

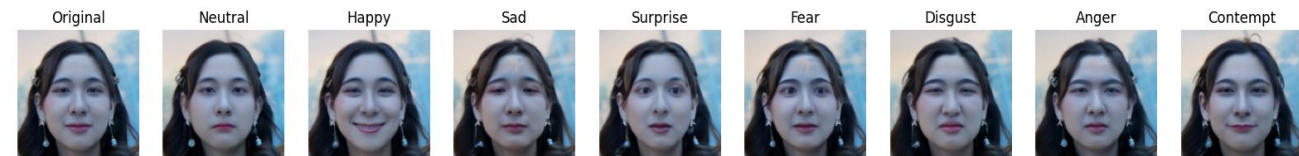


表. CSIMの結果

Emotion	CSIM Score
Happy	0.9991289551530050
Neutral	0.9992253874641800
Sad	0.999353341531626
Disgust	0.9993047995447710
Surprise	0.9991803598517230
Anger	0.9992440512510850
Fear	0.9992428167300590
Contempt	0.9991535004929470



Emotion	CSIM Score
Happy	0.9991500822195200
Neutral	0.99937070177147
Sad	0.9991631364899060
Disgust	0.9991973576733690
Surprise	0.9990699802366610
Anger	0.9991830557048840
Fear	0.9991767291212830
Contempt	0.9990630171746850







## 1. 実装状況

---

## 2. 研究計画

---

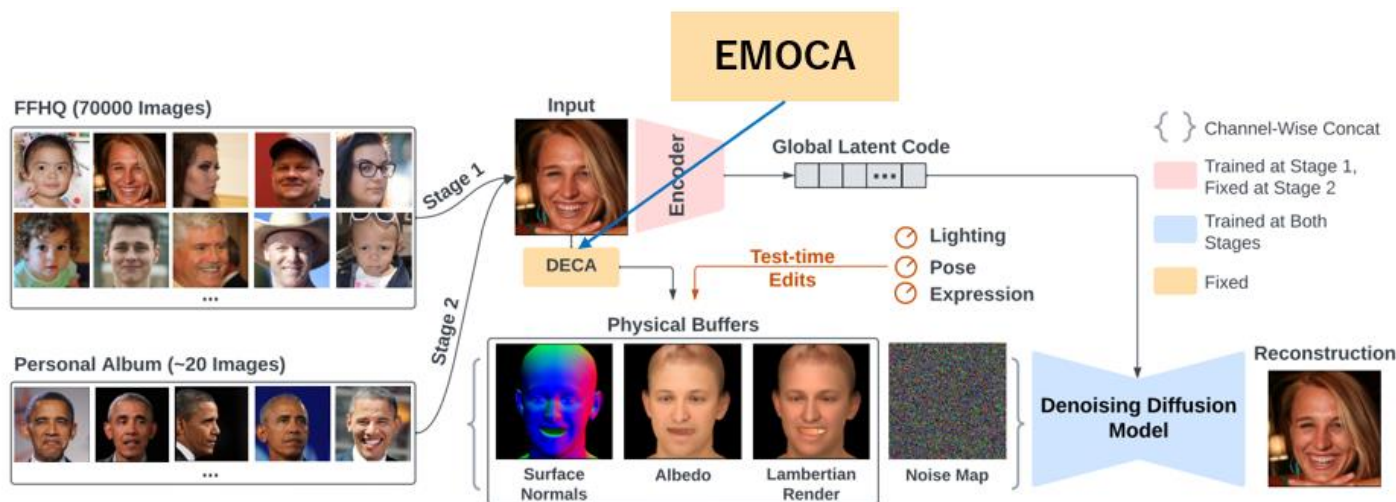
## 分散学習の実装完了

学習時間：8日間

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

学習時間：約8時間

使用GPU：RTX 4090(×2)  
Image size：256×256  
Batch size：8  
Max step：5000  
Global Encoder：ResNet50



## 2月末までにやりたいことリスト

- ~~ResNet50 verもやりたい~~
- バッチ数増加
- 5000step⇒10000step
- AffectNetデータセット追加 (70000枚⇒420000枚)
- AffectNetにアノテーションされている valence・arousalを用いて評価の実装 (**EmoNet**)

## 無表情への変換

Source



Target



Stage1 : EMOCA  
S : DECA、 T : EMOCA  
ResNet18



Stage1 : EMOCA  
S : DECA、 T : EMOCA  
ResNet50



## 笑顔への変換

Source



Target



Stage1 : EMOCA  
S : DECA、 T : EMOCA  
ResNet18



Stage1 : EMOCA  
S : DECA、 T : EMOCA  
ResNet50





## 笑顔への変換

Source



Target



Stage1 : EMOCA  
S : DECA、T : EMOCA  
ResNet18



Stage1 : EMOCA  
S : DECA、T : EMOCA  
ResNet50



## しかめっ面への変換

Source

Target

Stage1 : EMOCA  
S : DECA、T : EMOCA  
ResNet18

Stage1 : EMOCA  
S : DECA、T : EMOCA  
ResNet50





# Stage1      5000step $\Rightarrow$ 10000step

5000step



7000step



10000step



