

## 進捗報告

### 1 今週やったこと

ViT における多義図形識別はタッチ識別か否かの確認, TransGAN の動作確認, Attention map を用いた判断根拠に対する考察

### 2 タッチ識別に関して

タッチ識別か否かを確認するために元画像を 9 分割し, resize したものを入力画像とすることで識別率および混同行列による分布を確認した. 入力画像は全  $216 \times 9 = 1944$  枚とし, ラベルは分割する前の画像と同じものを付与した. 表 2 に縦軸を真値, 横軸を ViT による予測値とした混同行列を示す. また, 表 1 に元の 216 枚を ViT を用いて識別した混同行列結果を示す. 単に各値を 9 倍したものはなかった. 全体の識別率は元画像のほうが 93.1 単純に識別率が下がっているため全てをタッチ識別しているわけではないことがわかる.

表 1: 元画像の混同行列

真値	多義図形	61	5	6
	風景画	0	72	0
	肖像画	1	3	68
		多義図形	肖像画	風景画
		ViT による予測値		

表 2: 9 分割画像の混同行列

真値	多義図形	414	135	99
	風景画	58	545	45
	肖像画	59	195	394
		多義図形	肖像画	風景画
		ViT による予測値		

9 分割した画像の識別結果を attention map を用いて示したものを図 1 に示す. 明らかに顔が写っている画像に関しては元画像が多義図形画像であったとしても肖像画と判断しており, また背景部分に対しては正しく風景画と判断しているものもある. 元画像が風景画であっても分割して resize すると顔が映っている部分に対しては

肖像画と判断しているため, 画像から判断しているとわかった. 一方, 明らかに顔が映っていないにも関わらず多義図形や肖像画と判断してる画像もあるため, タッチ識別をしていないというわけではなくタッチ識別できるものはして, 画像から判断できるものも識別していると考えられる.

### 3 TransGAN の動作確認

TransGAN の動作確認が図 2 のエラーによりうまくいかなかった. 解決策は模索中. ViT の時と同様, 先に CNN ベースの GAN である DCGAN を予備実験として実装したい.

### 4 Attention map について

濃淡のかかり方が微妙であるため, どの部分を判断根拠としているかが判断しづらくなっている. 今回行ったタッチ識別か否かに対する実験から, タッチ識別もしているが画像から判断もしているという結果になった.

### 5 今後の方針

attention map のより明瞭な表現の実装, 予備実験として DCGAN の実装, TransGAN の実装



図 1: 左 : 分割した元画像 右 : attention map を用いた画像

```
(base) ito@merciserv:~/TransGAN-master$ sh exps/stl10_test.sh
Files already downloaded and verified
=> resuming from ./pretrained_weight/stl_checkpoint.pth
Traceback (most recent call last):
  File "test.py", line 171, in <module>
    main()
  File "test.py", line 122, in main
    assert os.path.exists(args.load_path)
AssertionError
```

図 2: TransGAN の動作確認中のエラー