

# MediaPipe で生成した顔点群に対する Conditional VAE を用いた表情操作

間世田 秀<sup>†</sup>赤松 茂<sup>††</sup>

† 法政大学大学院理工学研究科

†† 法政大学理工学部

## 1. はじめに

表情は人のコミュニケーションにとって重要なものであり、多くの研究が行われてきた。近年、深層学習を用いた顔画像生成に関する研究が多く行われているが、それらは 2 次元画像がほとんどであり、3 次元顔画像の生成については、データセットの少なさや処理の複雑さなどから十分に検討されていない。そこで本研究では、MediaPipe[1]を用いて 2 次元顔画像を点群化、点群をテーブルデータとして扱い、Conditional VAE(以降 CVAE)[2]で学習・表情操作する手法を提案する。

## 2. 提案手法

MediaPipeとは Google が提供している Machine Learning ソリューションであり、本研究ではその 1 つである Face Mesh を用いる。Face Mesh は、2 次元顔画像のランドマーク 478 点の 3 次元座標を推定できる機能である。この機能を用いて、1 人の表情を 478×3 次元のテーブルデータとして扱い、CVAE で表情ラベルとともに学習させる。学習後、対象データと対象表情ラベルでエンコードを行い、得られた「潜在変数」と「変化先表情ラベル」を用いてデコードを行うことで表情を操作する。本研究では真顔と笑顔の 2 表情を対象とし、表情操作は、真顔→笑顔、笑顔→真顔の 2 パターンである。

## 3. 実験方法

本研究では CelebA データセット[3]の真顔約 4 万枚、笑顔約 5 万枚を用いて学習や評価を行った。データを MediaPipe で点群化し、傾き・大きさ・位置を正規化したものを用いて CVAE で学習、操作を行った。CVAE のネットワークを表 1 に示す。全てのネットワークを全結合層で構成した。データの 1 万枚を test データとして、残りの 9 割を train データ、1 割を validation データとして学習した。

また、操作された表情を評価するため、正規化データを用いて 3 層全結合ニューラルネットワーク識別器を作成し、test データ 1 万枚を表情操作、評価した。識別器の accuracy は 0.92 で評価器として十分な精度である。

表1. CVAE のネットワーク

|                 |            |          |          |
|-----------------|------------|----------|----------|
| Encoder層        | 2層         | 潜在変数の次元  | 2        |
| Decoder層        | 2層         | バッチサイズ   | 256      |
| ノード数 (Encoder)  | 256→128    | Dropout率 | 0.2      |
| ノード数 (Decoder)  | 128→256    | 最適化手法    | Adam     |
| 活性化関数 (Encoder) | Leaky ReLu | 学習率      | 1.00E-06 |
| 活性化関数 (Decoder) | ReLu       |          |          |

## 4. 実験結果と考察

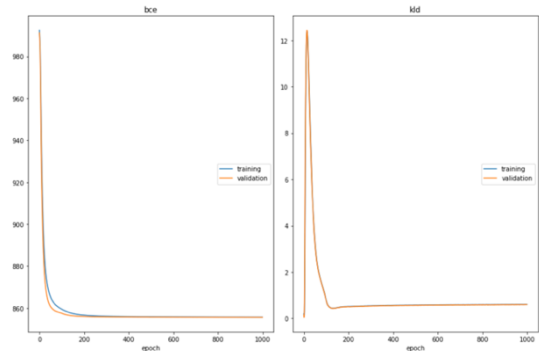


図1. CVAE の loss 推移

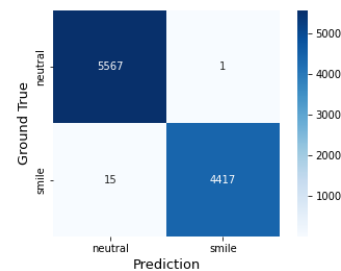


図 2. 混同行列

CVAE の損失関数である再構成誤差 (bce)と KL divergence (kld)の推移を図 1 に示す。どちらも収束しており、学習終了時 validation の loss 値はそれぞれ 856(bce)、0.595(kld)であった。test データの表情操作後、評価器で識別した混同行列を図 2 に示す。Ground Truth が表情操作後ラベルで、Prediction が評価器による識別結果である。操作失敗したものは、真顔→笑顔の 15 例、笑顔→真顔の 1 例であった。CVAE の潜在空間はクラスラベルと独立の特徴を表しており、今回失敗した人物顔は、train データの中に変化先表情の類似顔がなかったため、操作が失敗したと考えられる。

## 5. むすび

本研究では、Conditional VAE を用いて真顔→笑顔の表情点群を相互操作する手法を提案した。結果として、多くの人物について表情を操作することができた。今後は他表情でも提案手法が有効か検討する。

## 参考文献

- [1] C. Lugaresi, et al., IEEE. CVPR, 2019.
- [2] D. P. Kingma, et al., Advances in neural information processing systems, 27, 2014.
- [3] Z. Liu, et al., In: Proceedings of the IEEE international conference on computer vision, p. 3730–3738, 2015.