

# VAEを用いた笑顔認識が困難な顔画像の検出にむけて

羽路 悠斗<sup>†</sup> 下西 慶<sup>‡</sup> 近藤 一晃<sup>‡</sup> 上田 博唯<sup>‡</sup> 中村 裕一<sup>‡</sup>

<sup>†</sup> 京都大学 工学部情報学科

<sup>‡</sup> 京都大学 学術情報メディアセンター

## 1. はじめに

経時的な表情認識は多岐にわたる応用分野があり、広く普及している。従来手法は、正面を向いてはっきりと顔が映る理想的な状況で撮影された顔を主な対象としているが、横を向いたり顔の一部が隠れているなど、表情認識が困難な要素を含む場面も現実には多く存在する。事前にこれらの困難な顔画像を検出して弾ければ、従来手法が想定する入力データを担保でき、表情認識の信頼性を高められる。顔向き<sup>[1]</sup>やオクルージョン<sup>[2]</sup>といった個別の要素に対応する手法は存在するが、そもそも表情認識が困難なあらゆる要素を事前に想定することは困難である。本研究では、表情認識が困難な顔画像は認識器の学習時のデータとして基本的に対象とされていない点に着目し、学習データに含まれていない画像を検出することでこの問題に対処する。具体的には、認識器の一部を移植したVAE (Variational Auto Encoder) を用いて、再構成誤差を手掛かりに学習データと異なる分布の画像を検出する。本稿では、表情の中でも特に笑顔に着目し、笑顔認識の予測精度とVAEの再構成誤差の関連を調査した結果について述べる。

## 2. 提案手法

1章で述べたように本研究では、学習データに含まれていない画像をVAEで検出することで、表情認識が困難な画像を検出することを目指す。VAEはエンコーダーとデコーダーからなる、入力を復元して出力する再構成型の生成機であり、学習データに対する尤度最大化を目的として学習するため、学習データに含まれない画像の検出に適していると考え

Toward detecting difficult samples for smile recognition using Variational Auto Encoder Y. Haneji, K. Shimonishi, K. Kondo, H. Ueda, Y. Nakamura (Kyoto Univ.)

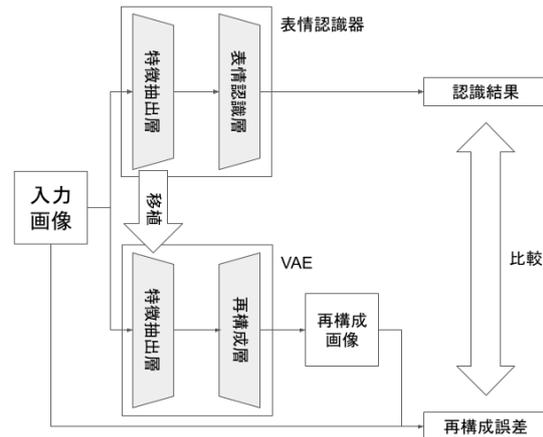


図1 表情認識器の一部を移植したVAEの再構成誤差と表情認識精度の比較

えられる。

図1に示すように、表情認識器の特徴抽出層のアーキテクチャと重みをVAEのエンコーダーに移植し、表情認識器と同等の特徴抽出力を与える。表情認識の学習データと同じ顔画像でVAEを学習する。これらにより、表情認識の学習データ分布に含まれていない顔画像は、学習後のVAEでの再構成が上手くいかず、再構成画像が元画像と大きく異なることが期待される。

以上のように学習したVAEで顔画像を再構成し、再構成の精度を表す任意の再構成誤差を手掛かりに、表情認識が困難な顔画像を検出する。

## 3. 評価

### 3.1 方法

提案手法を用いて、笑顔認識器の予測精度とVAEの再構成誤差の関連を評価した。

笑顔認識には、我々のグループがこれまでに提案してきた、2枚の顔画像を比較することで笑顔を認識する深層学習ネットワーク<sup>[3]</sup>を利用した。VAEは層が深いと上手く学習できないことが知られているため、笑顔認識器の特徴抽出層 (VGG16) の下位

7層のみをエンコーダーに移植した。

飲食を伴うビデオ通話中の被験者を正面から撮影した 30fps の動画から切り出した 41146 枚の顔画像の集合（以下、顔画像データセットと呼ぶ）から、学習用と評価用の顔画像ペアのデータセットを作成した。動画の音声情報や周辺フレーム情報も参考にし、どちらがより笑顔かアノテーションした。学習用のデータセットは、顔画像データセットから正面向き同士で笑顔度の差がある 313 ペアを取り出して作成した。

学習後の VAE を用いて、顔画像データセットの全ての画像の再構成誤差を算出して、再構成誤差が大きい顔画像を検出した。再構成誤差の算出には、VAE の損失関数の再構成誤差項である、入力画像と出力画像のクロスエントロピー誤差を用いた。

定性的評価として、再構成誤差が大きい顔画像が、経験的に表情認識が困難と思われる要素を含んでいるかどうかを評価した。定量的評価として、再構成誤差が大きい値の顔画像と通常値の顔画像の笑顔認識精度の違いを評価した。時間的に隣接した顔画像は再構成誤差も近い値を示すことが多いため、ウィンドウサイズが片側 250 フレームのピーク検出で局所的に再構成誤差が大きい顔画像を選んだ。さらに、大域的に再構成誤差が大きい顔画像に絞るために、99 パーセント以上の再構成誤差の顔画像に絞った。再構成誤差が通常値の顔画像を大きい値の顔画像と同じ枚数ランダムに取り出し、別途 10 枚の顔画像をランダムに取り出し笑顔比較の相手としてペアを組み、評価用のデータセットを作成した。笑顔比較相手の顔画像は、比較元に関わらず同一のものを用いた。どちらがより笑顔かアノテーションできなかったペアは、評価用データセットから除外した。再構成誤差が通常値の顔画像は 34 ペア、大きい値の顔画像は 28 ペア得られた。

### 3.2 結果と考察

再構成誤差が大きい顔画像は、顔向きが正面向きから大きく外れていたり、咀嚼中であったり、飲料を飲んでいたり、経験的に表情認識が困難と思われる要素が見られた。また、図 2 に示すように、再

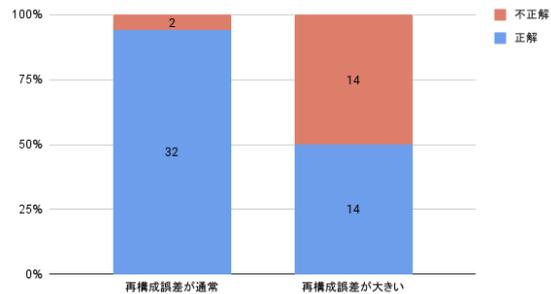


図 2 再構成誤差による笑顔認識精度の違い

構成誤差が通常値の顔画像の笑顔認識の正解数が 32 ペア、不正解数が 2 ペアで正解率が約 95% であったのに対し、再構成誤差が大きい顔画像の笑顔認識の正解数が 14 ペア、不正解数が 14 ペアで正解率は 50% と実際に笑顔認識の予測精度が落ちた。再構成誤差が大きい顔画像の認識精度はチャンスレベルだが、顔向きやオクルージョンにより、前節で述べた他の情報を参考にしないと、どちらがより笑顔か判断できないためと考えられる。

従って、想定する入力データを担保し、表情認識の信頼性を高めるためには、再構成誤差が大きい顔画像は表情認識の対象から弾くべきであると考えられる。

## 4. まとめ

表情認識が困難な顔画像を、VAE の再構成誤差を手掛かりに検出する手法を提案した。再構成誤差が大きい顔画像には表情認識が困難な要素が見られ、実際に笑顔認識の予測精度が落ちることがわかった。今後は、再構成誤差のより良い指標の探索や、白飛び、ピンぼけ、雑多な背景を含むデータセットでの検証に取り組みたい。

## 参考文献

- [1] Erik Murphy-Chutorian and Mohan Manubhai Trivedi. Head Pose Estimation in Computer Vision: A Survey. *IEEE Trans. on PAMI*, Vol.31, No.4, pages 607–626, 2009.
- [2] Dahua Lin and Xiaoou Tang. Quality-Driven Face Occlusion Detection and Recovery. *IEE Conf. on CVPR*, pages 1–7, 2007.
- [3] Kazuaki Kondo et al. Siamese-structure Deep Neural Network Recognizing Changes in Facial Expression According to the Degree of Smiling. In *Proc. of ICPR*, pages 4605–4612, 2021.