1M-03

Image Inpainting における 顔属性ラベルを考慮した Landmark 推定手法

大道 博文† 黒澤 義明† 目良 和也† 竹澤 寿幸†

†広島市立大学大学院 情報科学研究科

1. はじめに

昨今の新型コロナウィルスの影響により，季節を問わずにマスクを着用する機会が増えた．これに伴い，対面のコミュニケーションにおいてはマスクをした状態で接することが標準になりつつある．しかし，マスクの着用によって表情の大部分が隠されてしまうため，相手に意図しない情報が伝わってしまう恐れがある．本研究ではマスクで表情が隠された領域を破損とみなし，それを修復する課題 Image Inpainting に焦点を当てる．

近年の Image Inpainting では Generative Adversarial Networks（GAN）[1]を用いた手法が数多く提案されている[2][3][4]．その中でも Landmark を参考にした手法[5]では顔の向きや構造に基づいて画像修復が可能であることを示している．しかし，破損顔画像から推定された Landmark は真値とは異なる場合がある．これはコミュニケーションシステムへの応用を考えると誤った情報を伝達することになる．

そこで本研究では図 1 のように破損顔画像にラベル情報を付与した Landmark 推定手法を提案する．具体的には入力が破損顔画像とラベル情報，出力が Landmark としてモデルを設計する．また，口の開閉における Loss 関数も新たに定義する．評価として，従来手法と提案手法の修復画像の表情に着目して比較，議論する．

2. 提案手法

本章では[5]の Landmark Prediction Module に改良を加えた手法を提案する（図 1）．

2.1 入力情報

従来研究は破損顔画像だけを入力としていたが，本研究では CelebA[6]の顔属性ラベルを加えて学習を行う．CelebA は巨大な顔画像データセットであり，各画像に対して 40 種類のラベルがバイナリ値として付与されている．本研究では Smiling ラベルを用いて入力情報を構成する．

まず，Smiling ラベルのバイナリ値に Embedding 処理を行い，16 次元のベクトル表現にする．次に，破損顔画像のサイズ（256×256×3）に合わせるため，複製処理を加える．そして，それらのチャンネル次元軸を基準にして連結する．最終的な入力のサイズは（256×256×（3+1））となり，これを入力情報とする．

2.2 ネットワーク

本研究では従来研究同様，MobileNet-V2[7]をベースにしてネットワークを構築した．従来手法と異なる点は Landmark を出力する全結合層のほかに，ラベル情報を出力する全結合層を新たに追加したことである．

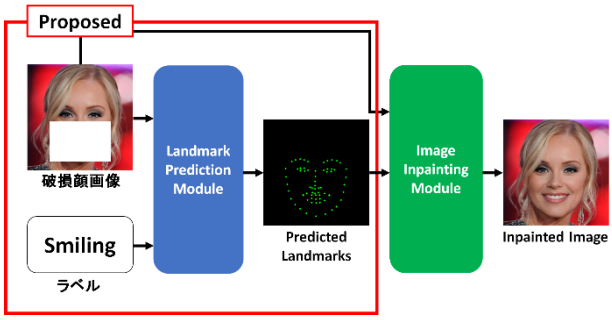

図 1: 本研究の概要図

2.3 Loss 関数

従来研究では単純に実測値と予測値の Landmark の差分をユークリッド距離として計算していた．しかし，未知のデータを入力した時，真値とは異なる結果を出力する場合がある．

そこで本研究では従来手法に加えて，Smiling ラベルを考慮した口の開閉における Loss 関数を新たに定義する．定義した Loss 関数を以下に示す．なお，太字で示した 3 つが本論文で提案する Loss 関数である．

- Landmark Loss

$$L_{lmk} := \left\| \hat{L} - L_{gt} \right\|_2^2 \quad (1)$$

- **Mouth Corner Distance Loss**

$$\boldsymbol{L_{MCD} := \frac{1}{2}\sum_{i=1}^{2}\left(\hat{D}_{MC}{}^{(i)} - D_{MC}{}^{(i)}\right)^2} \quad (2)$$

- **Lip Distance Loss**

$$\boldsymbol{L_{LD} := \frac{1}{2}\sum_{i=1}^{2}\left(\hat{D}_{Lip}{}^{(i)} - D_{Lip}{}^{(i)}\right)^2} \quad (3)$$

- **State Loss**

$$\boldsymbol{L_{state}(x, class) := -\log\left(\frac{\exp(x^{(class)})}{\sum_j \exp(x^{(j)})}\right)} \quad (4)$$

Landmark Loss は従来手法と同様の Loss 関数であり，$\hat{L}$ は予測値の Landmark，L_{gt} は実測値の Landmark を示している．

Mouth Corner Distance Loss は実測値と予測値の口角の距離の差分を定義した Loss 関数である．$\hat{D}_{MC}$ は予測値の口角の距離，D_{MC} は実測値の口角の距離，$D_{MC}{}^{(i)}$ の i は座標の添え字を示している．

Lip Distance Loss は実測値と予測値の唇の距離の差分を定義した Loss 関数である．$\hat{D}_{Lip}$ は予測値の唇の距離，D_{Lip} は実測値の唇の距離，$D_{Lip}{}^{(i)}$ の i は座標の添え字を示している．

State Loss はラベル情報の状態を定義した Loss 関数である．この関数は Auxiliary Classifier GAN[8]から着想を得て

Facial Landmark Prediction Method Considering Face Attribute Label in Image Inpainting: Hirofumi Omichi, Yoshiaki Kurosawa, Kazuya Mera, Toshiyuki Takezawa, Graduate School of Information Sciences, Hiroshima City University

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

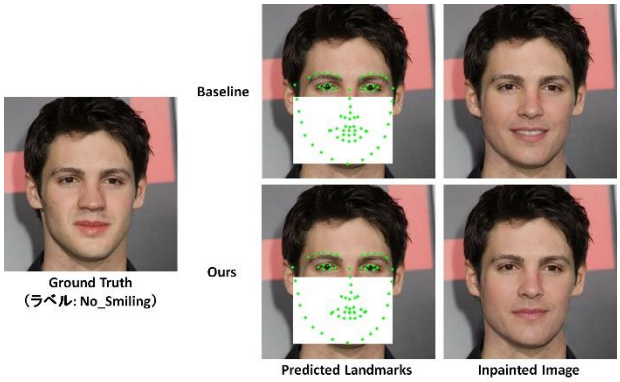

図 2: 従来手法と提案手法の比較（No_Smiling）

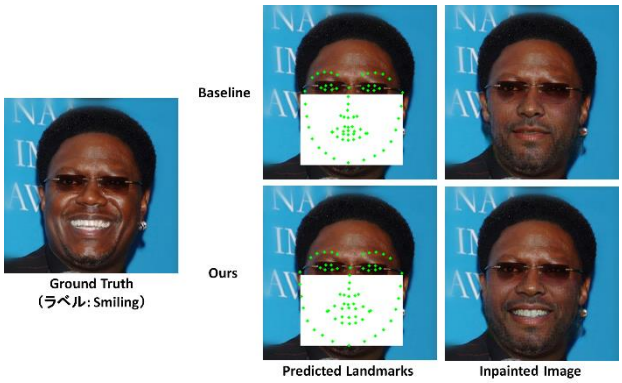

図 3: 従来手法と提案手法の比較（Smiling）

いる．xは予測値のラベルベクトル，$class$ は正解ラベル，$x^{(j)}$のjはラベルベクトルの添え字を示している．なお，本研究では Smiling ラベルのバイナリ値を使用するため，出力されるベクトルの次元数は 2 である．また，バイナリ値の 1 を Smiling，-1 を No_Smiling と定義する．

新たに 3 つの Loss 関数を定義することで，Smiling であれば口角と唇の距離が長くなり，No_Smiling であれば口角と唇の距離が短くなるように学習が進むと期待される．

3. 実験

本章では従来手法と提案手法との比較実験を行い，修復画像の表情について議論する．

3.1 Dataset

本研究では CelebA-HQ[9]を使用する．Landmark については FAN[10]を用いて 68 個の顔座標点を抽出する．学習時の破損領域は抽出した Landmark を参考に顔の下半分が隠れるように設定した．

3.2 比較実験

本研究の提案手法は Landmark Prediction Module の改良であるため，実験もその部分に限定する．また，出力結果が Landmark であることから，視覚的に理解しやすいようにその Landmark と破損顔画像から修復された画像を比較対象とする．その時の Image Inpainting Module については CelebA-HQ の Pre-Train モデルを使用する．

実験に使用した Train データは 27,998 枚，validation データは 2,000 枚，Iteration は 1,000,000，画像サイズは 256×256，Batch Size は 16 に設定した．実験結果を図 2 および図 3 に示す．各図の左側には元の顔画像（Ground Truth）と付与されているラベル情報を表示している．対して，右側には 1 行目に従来手法（Baseline），2 行目に提案手法（Ours）を表示し，1 列目に推定された Landmark（Predicted Landmarks），2 列目に修復された顔画像（Inpainted Image）を表示している．

図 2 および図 3 より，提案手法の方が従来手法に比べてより元の表情に近く修復されている．これは提案手法が口角と唇の距離を No_Smiling ラベルでは短く推論し，Smiling ラベルでは長く推論した結果が反映されていると考えられる（図中の Predicted Landmarks を参照）．このことより，提案手法の有効性を確認できた．

4. おわりに

本研究では破損顔画像にラベル情報を付与した Landmark 推定手法を提案した．[5]の Landmark Prediction Module の入力情報，ネットワーク，Loss 関数それぞれに改良を行った．比較実験の結果，提案手法の方が従来手法に比べてより元の表情に近い顔画像修復ができることを確認した．

今後の課題としては Loss 関数の改良や定量的評価を行うことが挙げられる．また，発話音声から推定した話者感情を顔属性ラベルと対応付けることによって，クロスモーダルな顔画像修復実験を行う予定である．

謝辞

本研究の一部は国立研究開発法人科学技術振興機構（JST）の研究成果展開事業「センター・オブ・イノベーション（COI）プログラム」グラント番号 JPMJCE1311 の支援を受けたものである．

参考文献

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: Generative Adversarial Nets. In Advances in Neural Information Processing Systems (NIPS), pp.2672–2680, 2014.

[2] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang: Generative Image Inpainting with Contextual Attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.5505–5514, 2018.

[3] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang: Free-Form Image Inpainting with Gated Convolution. In IEEE International Conference on Computer Vision (ICCV), pp.4471–4480, 2019.

[4] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi: EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. In IEEE International Conference on Computer Vision Workshop (ICCVW), 2019.

[5] Y. Yang, X. Guo, J. Ma, L. Ma, and H. Ling: LaFIn: Generative Landmark Guided Face Inpainting. arXiv:1911.11394v1, 2019.

[6] Z. Liu, P. Luo, X. Wang, and X. Tang: Deep Learning Face Attributes in the Wild. In IEEE International Conference on Computer Vision (ICCV), pp.3730–3738, 2015.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen: MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.4510–4520, 2018.

[8] A. Odena, C. Olah, and J. Shlens: Conditional Image Synthesis with Auxiliary Classifier GANs, In Proceedings of the 34th International Conference on Machine Learning, PMLR 70:2642-2651, 2017.

[9] T. Karras, T. Aila, S. Laine, and J. Lehtinen: Progressive Growing of GANs for Improved Quality, Stability, and Variation. In International Conference for Learning Representations (ICLR), 2018.

[10] A. Bulat and G. Tzimiropoulos: How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In IEEE International Conference on Computer Vision (ICCV), pp.1021–1030, 2017.

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.