

1. はじめに

畳み込みニューラルネットワークは、出力層をタスク毎に増やすことで、複数タスクを同時に学習および推定するマルチタスクラーニングの実現が可能である。従来のマルチタスクラーニングでは、学習するタスクに共通する低次特徴を抽出し、その特徴から各タスク固有の高次特徴を抽出するため、各タスク固有の低次特徴を抽出および利用することができないという問題がある。本研究では、マルチタスクラーニングで獲得できるタスク共通の特徴と各タスク固有の特徴を低次から高次まで同時に抽出する新たなマルチタスクラーニングのネットワーク構造として、Separation Multi-task Networks を提案する。提案手法は、これら2種類の特徴を同時に抽出することで、それぞれの特徴を考慮した学習および推定を可能とする。評価実験では、顔画像を用いて顔器官点検出と顔属性推定を行い、提案手法の有効性を示す。

2. マルチタスクラーニング

マルチタスクラーニング [1] は、複数タスクを同時に学習及び推定する手法である。この手法の利点として、学習する全てのタスクに共通する特徴表現を獲得できることが挙げられる。複数タスクを同時に学習することで、各タスク固有の特徴に加え、タスクに依存しない特徴を獲得できる。マルチタスクラーニングは、図1のように学習する全てのタスクに共通する特徴を抽出する Shared Layers と、各タスク固有の特徴を抽出して最終的な推定結果を出力する Task-specific Layers から構成される。このマルチタスクラーニングは、タスクに共通する低次特徴を抽出し、その特徴から各タスク固有の高次特徴を抽出しているため、各タスク固有の低次特徴を抽出および利用することができないという問題点がある。

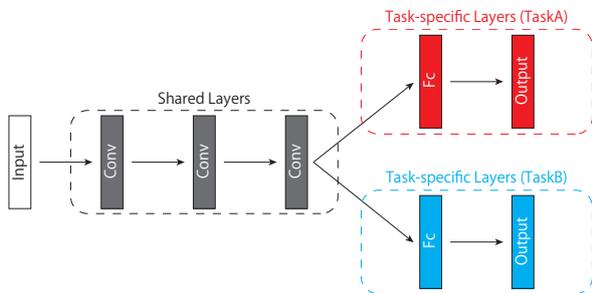


図 1：マルチタスクラーニング

3. 提案手法

提案する Separation Multi-task Networks は、マルチタスクラーニングで獲得できるタスク共通の特徴と各タスク固有の特徴を低次から高次まで同時に抽出する。提案手法は、2種類の特徴を同時に抽出するため、図2のように Shared Layers と Task-specific Layers を並列に構築する。提案手法では、入力画像を各 Layer の1層目に入力し、Shared Layers ではタスク共通の特徴、Task-specific Layers では各タスク固有の特徴をそれぞれ獲得する。そして、獲得した特徴マップを次層に入力する際、Task-specific Layers は前層で獲得したタスク固有の特徴と Shared Layers で獲得したタスク共通の特徴を連結して入力する。これにより、提案手法では、タスク共通の特徴とタスク固有の特徴を考慮した学習および推定が可能となる。

3.1 学習方法

タスク共通の特徴とタスク固有の特徴をそれぞれ獲得するため、本研究では、次の2段階学習によって提案手法の学習を行う。これにより、Shared Layers はタスク共通の特徴、Task-specific Layers はタスク固有の特徴のみを捉えるように学習することができる。

Stage 1 図1に示すマルチタスクラーニングを用いて複数タスクを同時に学習する。これにより、タスク共通の特徴

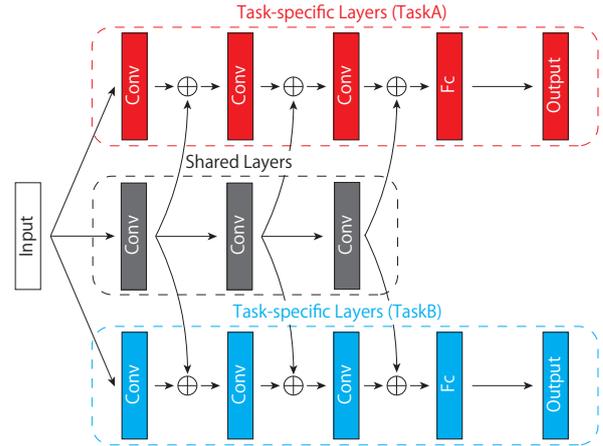


図 2：Separation Multi-task Networks

を抽出するための Shared Layers のパラメータを事前に獲得できる。Stage 1 では、各タスクの学習誤差をもとに誤差逆伝播法を用いてパラメータを更新する。学習するタスクを TaskA, TaskB とし、それぞれの学習誤差を E_{TaskA} , E_{TaskB} としたとき、ネットワーク全体の学習誤差 E_{all} は式 (1) で計算される。

$$E_{all} = E_{TaskA} + E_{TaskB} \quad (1)$$

Stage 2 図2に示すタスク固有の低次特徴を抽出するための層となる Task-specific Layers を追加して学習する。このとき、Shared Layers には Stage 1 で獲得したパラメータを固定して用いる。これにより、各 Task-specific Layers はタスク固有の特徴のみを抽出するように学習できる。Stage 2 の学習においても、Stage 1 と同様に各タスクの学習誤差をもとに誤差逆伝播法を用いてパラメータを更新する。

3.2 Channel-wise Convolution の導入

提案手法では、タスク共通の特徴とタスク固有の特徴を連結するため、Shared Layers とは異なり、各 Task-specific Layers に入力される特徴マップのチャンネル数は増加する。そのため、Task-specific Layers を学習させる際に、Stage 1 で獲得したパラメータを Task-specific layers の初期パラメータとして利用することができない。そこで、本研究では、2種類の特徴を連結させた後にカーネルサイズ 1×1 の畳み込み処理 (Channel-wise Convolution) を行い、各 Task-specific Layers に入力する特徴マップのチャンネル数を調整する。これにより、各 Task-specific Layers に、Stage 1 で獲得したパラメータを初期パラメータとして利用し、ファインチューニングすることが可能となる。

4. 評価実験

評価実験として、CelebA Dataset[4] を用いて両目、鼻、口角の5点の顔器官点検出と帽子や黒髪、笑顔などの40種類の顔属性推定を行い、提案手法の有効性を調査する。顔器官点検出の比較手法に、提案手法の Stage 1 で使用した従来のマルチタスクラーニングをベースラインとして用いる。また、顔属性推定ではベースラインに加え、従来の顔属性推定手法として、FaceTracer[2], PANDA-w[3], PANFA-l[3], LNet+ANet[4] との推定精度を比較する。また、提案手法は、Channel-wise Convolution を導入したモデル、さらに、Task-specific Layers の学習時に、Stage 1 で獲得したパラメータを Task-specific Layers の初期パラメータとして利用したモデルとの推定精度を比較し、Channel-wise Convolution の有効性についても調査する。

本実験では、162,770 枚を学習サンプル、19,962 枚を評価サンプルとして使用する。また、入力画像サイズは 128×128 ピクセル、バッチサイズは 32、学習の更新回数は 100 epoch とする。最適化手法には、MomentumSGD を用いる。このとき、学習率を 0.001、Momentum 係数を 0.9 とする。

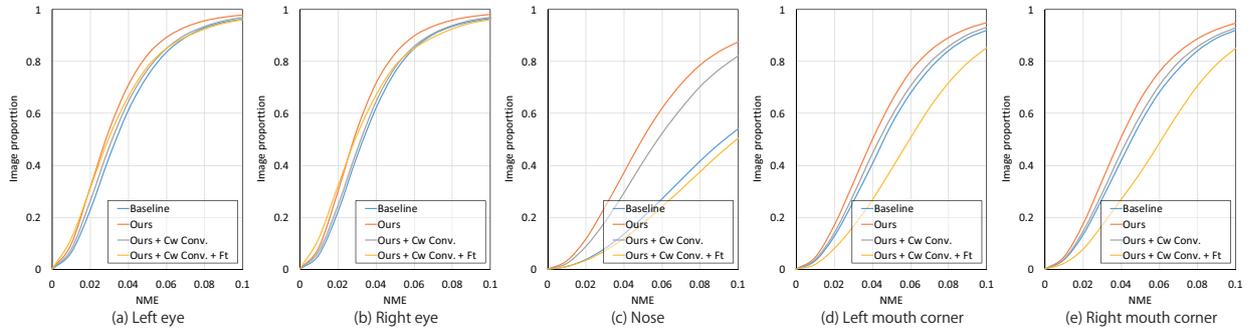


図 3 : 各顔器官点の検出精度の変化

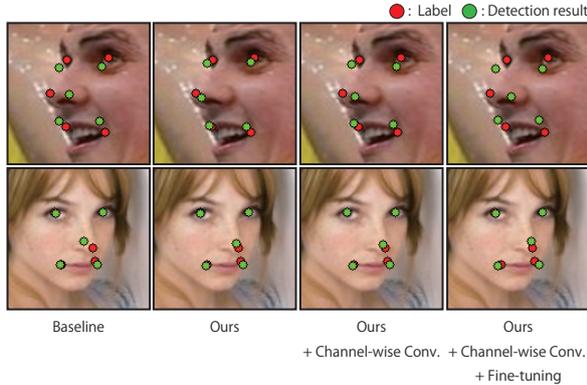


図 4 : 顔器官点の検出結果例

顔器官点検出の評価には、式 (2) に示す Normalized Mean Error (NME) を用いて比較する。ここで、 t_i と y_i はそれぞれ顔器官点 i における正解座標と出力座標、 t_{le} と t_{re} はそれぞれ左目と右目の正解座標を示す。

$$NME_i = \frac{\|t_i - y_i\|_2}{\|t_{le} - t_{re}\|_2} \quad (2)$$

4.1 顔器官点検出の比較

ベースラインと提案手法による各顔器官点の検出精度の変化を図 3 に示す。図 3 から、提案手法は、全ての顔器官点においてベースラインを上回る検出精度を達成した。特に、鼻の検出精度を比較した場合、提案手法は大幅に検出精度が向上した。このことから、タスク共通の特徴とタスク固有の特徴を考慮した学習および推定を行う提案手法は、顔器官点検出において有効であることがわかる。

ベースラインと提案手法による顔器官点の検出結果例を図 4 に示す。ここで、赤丸は教師信号、緑丸は検出結果を示す。図 4 から、顔が横を向いている場合、ベースラインでは、全ての顔器官点にズレが発生している。一方、提案手法はこれらのズレを改善していることがわかる。

4.2 顔属性推定の比較

従来手法と提案手法による顔属性推定の推定精度を表 1 に示す。表 1 から、Channel-wise Convolution とファインチューニングを導入したモデルの場合、平均推定精度が約 90.0% と最も高い推定精度を達成した。このことから、顔器官点検出と同様に、提案手法は、顔属性推定においても有効であることがわかる。そして、マルチタスクラーニングにおいて、タスク固有の低次特徴を利用することは、推定精度向上に有効であることがわかる。

4.3 特徴マップの可視化

提案手法で抽出するタスク共通の特徴とタスク固有の特徴の違いを調査するために、特徴マップの可視化を行う。提案手法における Shared Layers と各 Task-specific Layers で出力される特徴マップの可視化結果例を図 5 に示す。図 5 から、顔器官点検出では Shared Layers で捉えていないエッジなどの特徴を抽出できている。同様に、顔属性推定では、顔のシワなどのコンテキスト情報を抽出できている。このことから、提案手法はタスク共通の特徴と各タスク固有の特徴を分けて抽出していることがわかる。

表 1 : 顔属性推定の平均推定精度 [%]

	Average
FaceTracer [2]	81.1
PANDA-w [3]	79.8
PANDA-l [3]	85.4
LNets + ANet [4]	87.3
Baseline	89.9
Ours	89.6
Ours + Channel-wise Conv.	89.4
Ours + Channel-wise Conv. + Fine-tuning	90.0

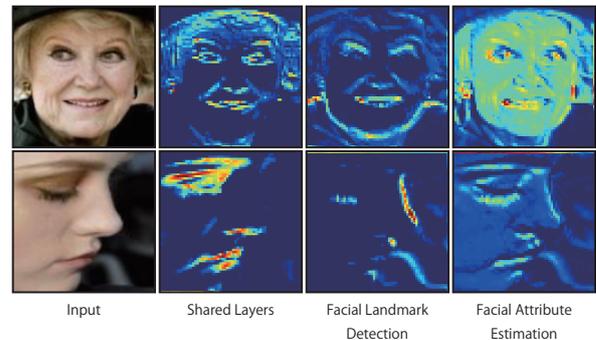


図 5 : 特徴マップの可視化結果例

5. おわりに

本研究では、タスク共通の特徴とタスク固有の特徴を同時に抽出する新たなマルチタスクラーニングのネットワーク構造として、Separation Multi-task Networks を提案した。評価実験では、顔器官点検出および顔属性推定において、提案手法は従来手法を上回る検出精度を達成した。また、タスク固有の低次特徴を用いることで推定精度を向上できることを示した。今後は、顔画像以外のタスクへの応用と提案手法の 2 段階学習から一貫学習への改善を検討する。

参考文献

- [1] R. Caruana, "Multitask Learning", Machine Learning Journal, Vol.28, No.1, pp.41-75, 1997.
- [2] N. Kumar, *et al.*, "FaceTracer: A Search Engine for Large Collections of Images with Faces", ECCV, 2008.
- [3] N. Zhang, *et al.*, "PANDA: Pose Aligned Networks for Deep Attribute Modeling", CVPR, 2014.
- [4] Z. Liu, *et al.*, "Deep Learning Face Attributes in the Wild", ICCV, 2015.

研究業績

- [1] 松井涼 等, "Separation Multi-task Networks による顔器官点と顔属性の同時推定", PRMU 研究会, 2018.
- [2] R. Matsui, *et al.*, "Simultaneous Estimation of Facial Landmark and Attributes with Separation Multi-task Networks", VISAPP, 2019.

(他 学会発表 2 件)