社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

Separation Multi-task Networksによる顔器官点と顔属性の同時推定

松井　　涼†　山下　隆義†　藤吉　弘亘†

† 中部大学　〒 487–8501 愛知県春日井市松本町 1200
E-mail: †ryo@mprg.cs.chubu.ac.jp, ††{takayoshi,fujiyoshi}@isc.chubu.ac.jp

あらまし　マルチタスクラーニングは，単一の DCNN で複数タスクを同時に学習および推定する手法である．しかし，従来のマルチタスクラーニングでは，タスク間で共通する低次の特徴のみを抽出し，その特徴から高次のタスク固有の特徴を抽出しているため，低次のタスク固有の特徴を利用できない問題がある．本研究では，タスク間で共通する特徴と各タスク固有の特徴を同時に抽出する Separation Multi-task Networks を提案する．提案手法は，2 つの特徴を同時に抽出することで，2 つの特徴を考慮した学習および推定を可能とする．CelebA を用いた評価実験では，提案手法は従来手法を上回る推定精度を達成した．

キーワード　マルチタスクラーニング，Deep Convolutional Neural Network，Separation Multi-task Networks

1. はじめに

人の顔には，年齢，性別，人種など様々な属性情報が含まれている．これらの属性情報は顔照合や顔向き推定，美肌加工などのアプリケーションの実現に必要であり，顔画像を解析することで属性情報を取得することができる [1].

コンピュータビジョンの分野では，Deep Convolutional Neural Network(DCNN) が画像分類問題において高い認識精度を達成して以降，顔器官点検出や顔属性推定も同様に，DCNN を用いることで高精度な推定が可能となった [1]〜[4]．しかし，1 つのタスクに対して 1 つの DCNN を構築すると，属性情報毎にネットワークが必要となるため，計算コストが膨大になる．

マルチタスクラーニング [4]〜[11] は，単一の DCNN で複数タスクを同時に学習および推定する手法である．この手法は，タスク数に比例して増加する学習および推定時間を大幅に削減することができる．これにより，複数のタスクから得られる情報が必要な顔画像解析においてマルチタスクラーニングは有効的な手法である．

マルチタスクラーニングは，学習する全てのタスクで共通する特徴表現とタスク固有の特徴表現を階層的に学習する．すなわち，低層では全タスクに共通する特徴，上位層はタスク固有の特徴を捉えるようになる．複数タスクを同時に学習することで，タスク固有の特徴に加えて，全タスク共通の特徴を獲得できるため，より一般的な特徴表現を獲得することができる．

顔画像解析の手法として，顔器官点検出に加え，顔向きや性別などの属性を同時に学習する Task-Constrained Deep Convolutional Network(TCDCN) [4] が提案されている．顔画像解析以外にも，法線推定とセマンティックセグメンテーションを同時に行う Cross-stitch Networks [6]，物体検出と物体追跡を同時に行う手法 [7] などにマルチタスクラーニングが導入されている．また，マルチタスクラーニングはコンピュータビジョン分野のみならず，自然言語処理 [8] などの様々な分野においても用いられている．従来のマルチタスクラーニング手法の多くは，学習する全てのタスクで共通する特徴マップを抽出して共有する Shared Layers と Shared Layers から抽出された特徴マップをもとに各タスク固有の特徴を抽出し，各タスクの推定結果を出力する Task-specific Layers から構成される．しかし，従来のマルチタスクラーニングを用いる場合，タスク間で共通する低次の特徴を抽出し，その特徴から高次のタスク固有の特徴を抽出しているため，低次のタスク固有の特徴を抽出および利用できない問題点がある．

本研究では，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，全てのタスクで共通する特徴と各タスク固有の特徴を同時に抽出する Separation Multi-task Networks を提案する．提案手法は，Shared Layers は全てのタスクで共通の特徴を抽出する一方，Task-specific Layers は前層のタスク固有の特徴と全タスク共通の特徴を連結して入力する．これにより，提案手法は 2 つの特徴を考慮した学習および推定が可能となる．提案手法の特長を以下に示す．

- タスク共通の特徴と各タスク固有の特徴を同時に抽出することで 2 つの特徴を考慮した学習および推定が可能
- 2 つの特徴を連結したあとに Channel-wise Convolution を導入することで各 Task-specific Layers に入力される特徴マップのチャンネル数を調整可能
- 各 Task-specific Layers が Shared Layers で捉えていない特徴を抽出するように学習可能

2. マルチタスクラーニング

マルチタスクラーニング [5] は，DCNN の手法の 1 つで，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は，単一の DCNN とは異なり，学習する全てのタスクで特徴表現を共有しながら複数タスクを同時に最適化するように学

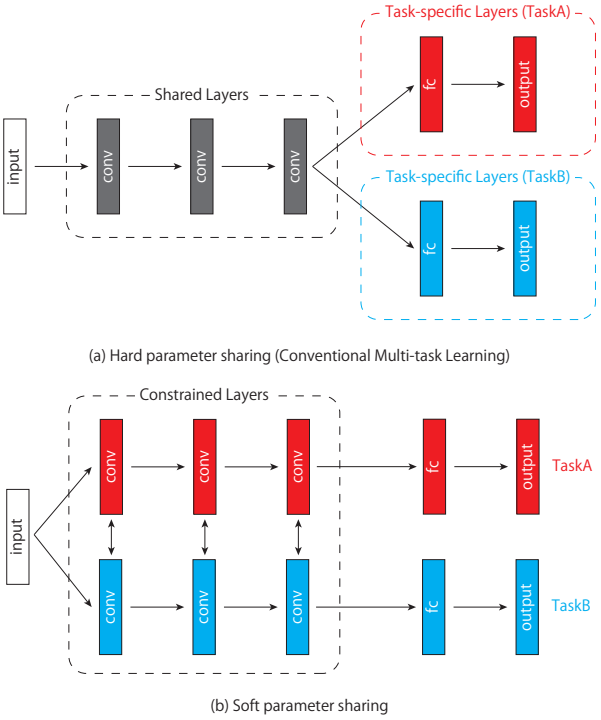

図 1　一般的に用いられているマルチタスクラーニングのネットワーク構造．(a) は Hard parameter sharing，(b) は Soft parameter sharing を示す．

習する．この手法の利点として，より一般的な特徴表現を獲得することが可能であることが挙げられる．これは，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可能であることから言える．また，マルチタスクラーニングはタスクに依存しない特徴を用いて特定のタスクの精度を向上させることも可能である．DCNN におけるマルチタスクラーニングには，Hard parameter sharing と Soft parameter sharing の 2 種類がある [12].

2.1　Hard parameter sharing

Hard parameter sharing [5] は，複数タスクで中間層のパラメータを共有するネットワークモデルである．Hard parameter sharing は，図 1(a) のように学習する全てのタスクで共通する特徴マップを抽出する Shared Layers と Shared Layers から抽出された特徴マップをもとに各タスク固有の特徴を抽出し，各タスクの推定結果を出力する Task-specific Layers から構成される．また，Hard parameter sharing には，Shared Layers と Task-specific Layers を分ける位置 (branch point) がある．この branch point はあらかじめ決定しておく必要がある．決定方法は，あらかじめ適当に決めるか，全てのパターンを試して最も精度が良い場所を利用する．Shared Layers は，学習する全タスクに有効な特徴表現を学習するため，単一の DCNN と比べて過学習を抑制することができる．Hard parameter sharing はコンピュータビジョン分野をはじめ，様々な分野で，一般的に用いられているマルチタスクラーニングのネットワークモデルである [4] [6] [9].

法線推定とセマンティックセグメンテーションを同時に行う Cross-stitch Networks [6] では，学習するタスクごとに DCNN を構築し，Cross-stitch units を用いてタスク間で共通する特徴表現を学習する．Cross-stitch units は，各タスクで獲得した特徴マップをチャンネルごとに全結合により統合し，タスク間で共通する特徴マップを獲得する．獲得したタスク共通の特徴マップは各タスクの DCNN に入力される．この手法は，単一の DCNN と比べて法線推定とセマンティックセグメンテーションの精度が向上している．特に，データ数が少ないカテゴリに対して，著しく認識精度が向上している．他にも，カスケード状にタスクを設置し，インスタンスセグメンテーションを行う Multi-task Network Cascades [9] が提案されている．この手法は，タスク間で共通する特徴を抽出した後に，物体検出，マスク推定，カテゴリ推定の順にタスクを処理してインスタンスセグメンテーションを行う．このとき，各タスクへの入力はタスク間で共有する特徴マップに加え，1 つ前のタスクの出力結果を入力する．このようなカスケード構造にすることで，高精度なインスタンスセグメンテーションを実現している．

Task-Constrained Deep Convolutional Network(TCDCN) [4] では，メインタスクを顔器官点検出，サブタスクを顔向きや性別，笑顔などの属性推定として同時に学習する．そして，メインタスクである顔器官点検出の学習をサブタスクが補助して顔器官点検出の検出精度向上を実現している．すなわち，属性情報が顔器官位置を検出する際の限定させる情報となっている．また，TCDCN は Task-wise early stopping により，サブタスクの学習を途中で停止させ，サブタスクの過学習によるメインタスクへ与える悪影響を抑制している．

2.2　Soft parameter sharing

Soft parameter sharing は，各タスクのパラメータ間の距離を均一化するネットワークモデルである [10] [11]．Soft parameter sharing は各タスクのパラメータ間の距離を均一化する Constrained layers から構成される．また，Hard parameter sharing と異なり，Soft parameter sharing は，タスクごとに個別のネットワークを構築する．Soft parameter sharing のネットワークモデル例を図 1(b) に示す．各タスクのパラメータ間の距離の均一化には，L2 ノルムを用いた手法 [10] やトレースノルムを用いた手法 [11] がある ．しかし，Soft parameter sharing は，1 つのタスクに対して 1 つの DCNN を構築する必要があるため，推定するタスク数に比例して計算コストが膨大になる．

3.　提 案 手 法

Hard parameter sharing を用いる場合，タスク間で共通する低次の特徴を抽出し，その特徴から高次のタスク固有の特徴を抽出する．そのため，低次のタスク固有の特徴を抽出および利用することができない．本研究では，タスク共通の特徴とタスク固有の特徴を低次から高次まで網羅して抽出する Separation Multi-task Networks を提案する．本章では，3.1 節で提案手法のネットワーク構造，3.2 節で提案手法の学習方法，そして，3.3 節で Channel-wise Convolution について述べる．

3.1　ネットワーク構造

提案手法のネットワーク構造は，タスク共通の特徴とタスク固有の特徴を低次から高次まで同時に抽出するために Hard

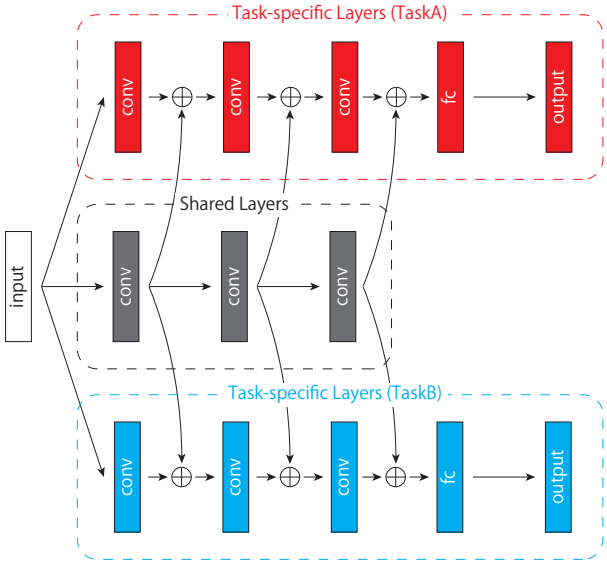

図 2　Separation Multi-task Networks

parameter sharing と異なり，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，入力画像を各タスク毎の Layer に入力し，Shared Layers ではタスク共通の特徴マップ，Task-specific Layers では各タスク固有の特徴マップをそれぞれ獲得する．このとき，前層で獲得したタスク固有の特徴マップと Shared Layers で獲得したタスク共通の特徴マップを連結し，Task-specific Layers に入力される．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．そして，Hard parameter sharing と同様に，最終的な推定結果は各 Task-specific Layers から出力される．

3.2 学習方法

全タスクで共通する特徴とタスク固有の特徴をそれぞれ獲得するために，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. まず，図 1(a) の Hard parameter sharing をベースとした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，学習する全タスク間で共通する特徴を抽出するための Shared Layers のパラメータを事前に獲得することができる．Stage1. では，各タスクの学習誤差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) で定義される．このとき，学習するタスクが回帰推定タスクの場合は平均二乗誤差，認識タスクの場合はソフトマックスクロスエントロピーとなる．

$$E_{all} = E_{TaskA} + E_{TaskB} \tag{1}$$

Stage 2. 次に，図 2 のように Shared Layers と並列処理を行う Task-specific Layers を学習する．このとき，図 2 の Shared Layers には Stage 1. で獲得した Shared Layers のパラメータを固定して用いる．すなわち，Task-specific layers のパラメータのみ学習する．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される．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からパラメータを更新する．

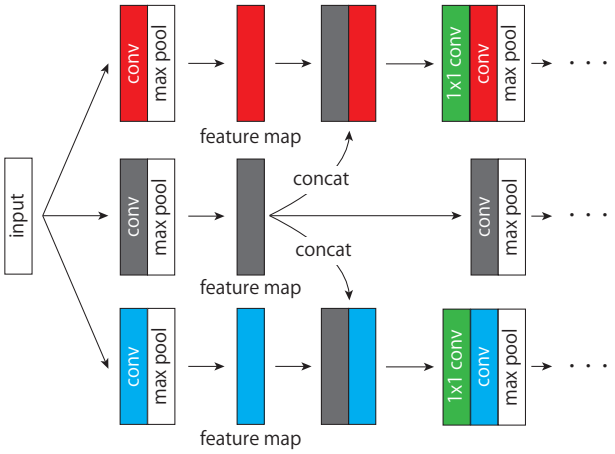

図 3　Channel-wise Convolution

表 1　本実験で使用する提案手法のネットワーク構造の詳細

	TaskA	Shared	TaskB
Input	size = 128 × 128, channels = 3		
Conv1-1	ksize = 3 × 3, channels = 64, pad = 1		
Conv1-2			
Max pooling	ksize = 2 × 2		
Conv2-1	ksize = 3 × 3, channels = 128, pad = 1		
Conv2-2			
Max pooling	ksize = 2 × 2		
Conv3-1	ksize = 3 × 3, channels = 256, pad = 1		
Conv3-2			
Conv3-3			
Max pooling	ksize = 2 × 2		
Conv4-1	ksize = 3 × 3, channels = 512, pad = 1		
Conv4-2			
Conv4-3			
Max pooling	ksize = 2 × 2		
Fc	2048		2048
Output	10		80

3.3 Channel-wise Convolution

提案手法では，タスク共通の特徴マップを各タスク固有の特徴マップに連結するため，各 Task-specific Layers に入力される特徴マップのチャンネル数は Shared Layers とは異なり増加する．そのため，ImageNet で学習した VGGNet [13] や ResNet [14] などの公開されている学習済みモデルを Task-specific Layers に導入してファインチューニングすることができない．そこで，提案手法では，図 3 のように特徴マップの連結後に 1 × 1 の畳み込み (Channel-wise Convolution) を導入する．これにより，各 Task-specific Layers に入力される特徴マップの次元数を調整することができ，事前学習モデルの利用が可能となる．

4. 評価実験

提案手法の有効性を調査するために評価実験を行う．評価実

表 2　ベースラインと提案手法による顔器官点の検出精度 [%]

	Channel-wise Convolutioin	Fine-tuning	Left eye	Right eye	Nose	Left mouth corner	Right mouth corner	Average
Baseline (教師信号との平均誤差)			96.3 (0.013)	96.5 (0.013)	54.0 (0.038)	92.0 (0.017)	91.9 (0.017)	86.1 (0.020)
Ours (教師信号との平均誤差)			**97.7 (0.011)**	**98.0 (0.011)**	**87.5 (0.019)**	**94.9 (0.015)**	**94.7 (0.015)**	**94.6 (0.014)**
	✓		96.9 (0.012)	97.0 (0.012)	82.2 (0.022)	93.1 (0.016)	93.0 (0.016)	92.4 (0.016)
	✓	✓	96.1 (0.012)	96.0 (0.012)	50.6 (0.037)	85.4 (0.021)	85.1 (0.021)	82.6 (0.021)

験では，CelebA Dataset [3] を用いて顔器官点検出と顔属性推定を行う．比較対象として，顔器官点検出手法では，3.2 節の Stage1. で使用した Hard parameter sharing をベースとした従来のマルチタスクラーニングによる手法をベースラインとして用いる．顔属性推定では，顔器官点検出の比較に用いるベースラインに加え，FaceTracer [15]，PANDA-w と PANFA-l [16]，LNets+ANet [3] と推定精度を比較する．また，提案手法は，Separation Multi-task Networks とこれに Channel-wise Convolution を導入したモデル，さらに Task-specific Layers に事前学習モデルをファインチューニングしたモデルの 3 種類を実装し，Channel-wise Convolution の有効性についても調査する．

さらに，提案手法を解析するために，branch point の変更による推定精度の変化の調査と Shared Layers と各 Task-specific Layers の特徴マップの可視化を行う．

4.1　実 験 概 要

本実験で使用する CelebA Dataset [3] は約 20 万枚の顔画像から構成されているデータセットで，両目，鼻，口角の 5 点の顔器官点と帽子や黒髪，笑顔などの 40 種類の顔属性が付与されている．学習サンプルは 162,770 枚，評価サンプルは 19,962 枚である．

本実験で使用する提案手法のネットワークモデルの詳細を表 1 に示す．ここで、TaskA を顔器官点検出，TaskB を顔属性推定とする．また，ネットワークの活性化関数には ReLU を使用する．各手法の学習方法には，Stochastic Gradient Descent(SGD) に Momentum 項を追加して学習する．このとき，学習率を 0.001，Momentun 係数を 0.9 とする．学習の更新回数は 100 エポック，バッチサイズは 32 とする．入力する画像サイズは，128 × 128 である．

本実験での顔器官点検出の評価方法は，各顔器官点において式 (2) を満たした場合を検出成功とする．ここで，(x_i, y_i) と (x'_i, y'_i) はそれぞれ教師信号と出力結果の座標，L は両目間の距離，α はしきい値を示す．本実験では，しきい値 α を 0.1 とし，検出結果と教師信号との誤差が両目間の距離の 10%以内の場合，検出成功とする．

$$\frac{\sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2}}{L} \leqq \alpha \qquad (2)$$

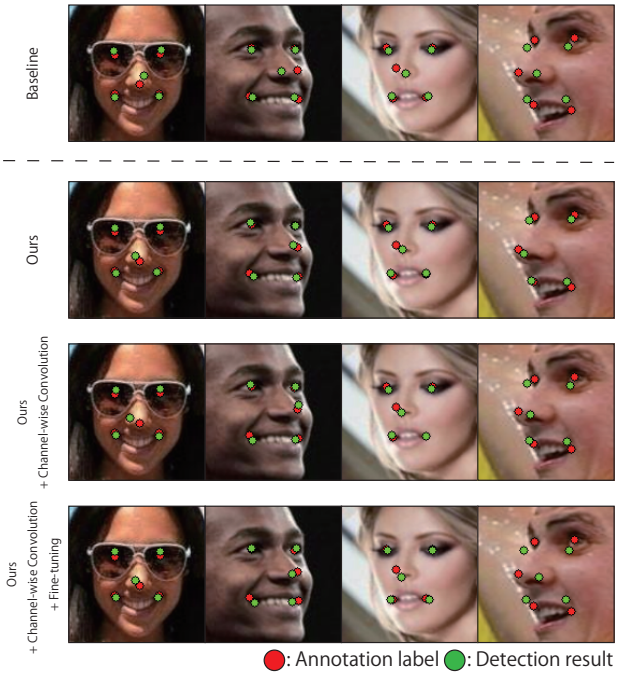

図 4　ベースラインと提案手法による顔器官点の検出結果例

4.2　従来手法との推定精度の比較

4.2.1　顔器官点検出

ベースラインと提案手法による顔器官点検出の検出精度と出力結果と教師信号の平均誤差を表 2 に示す．表 2 から，提案手法はベースラインと比較して平均検出精度が約 8.5%向上している．また，各顔器官点に対する検出精度を比較した場合，提案手法は全ての顔器官点においてベースラインより高精度な検出精度を達成している．特に，鼻の検出においてベースラインの検出精度が約 54.0%に対して，提案手法は約 87.5%と約 30.0%向上させることができた．このことから，全タスクで共通する特徴と各タスク固有の特徴を同時に抽出する提案手法は，顔器官点検出において有効であることがわかる．一方，提案手法に Channel-wise Convolution を導入したモデルは，ベースラインの精度を上回るが，導入しないモデルを下回る結果となった．また，Channel-wise Convolution を導入して各 Task-specific Layers を事前学習したモデルを用いてファインチューニングした場合，ベースラインを下回る結果となった．

ベースラインと提案手法による顔器官点検出結果例を図 4 に

表 3　各手法による各顔属性の推定精度 [%]

	Channel-wise Convolution	Fine-tuning	5 o Clock Shadow	Arched Eyebrows	Attractive	Bags Under Eyes	Bald	Bangs	Big Lips	Big Nose	Black Hair	Blond Hair	Blurry	Brown Hair	Bushy Eyebrows	Chubby	Double Chin	Eyeglasses	Goatee	Gray Har	Heavy Makeup	High Cheekbones	Male
FaceTracer [15]			85.0	76.0	78.0	76.0	89.0	88.0	64.0	74.0	70.0	80.0	81.0	60.0	80.0	86.0	88.0	98.0	93.0	90.0	85.0	84.0	91.0
PANDA-w [16]			82.0	73.0	77.0	71.0	92.0	89.0	61.0	70.0	74.0	81.0	77.0	69.0	76.0	82.0	85.0	94.0	86.0	88.0	84.0	80.0	93.0
PANDA-l [16]			88.0	78.0	**81.0**	79.0	96.0	92.0	67.0	75.0	85.0	93.0	86.0	77.0	86.0	86.0	88.0	98.0	93.0	94.0	90.0	86.0	97.0
LNets+ANet [3]			91.0	79.0	**81.0**	79.0	98.0	95.0	68.0	78.0	**88.0**	**95.0**	84.0	80.0	90.0	91.0	92.0	99.0	95.0	97.0	90.0	**87.0**	**98.0**
Baseline			93.6	**81.8**	80.7	**83.8**	98.1	95.0	69.8	82.7	85.9	94.7	95.7	85.6	**91.8**	**95.5**	**96.1**	**99.4**	**97.2**	97.8	89.9	86.0	97.5
Ours			93.3	81.2	80.0	83.1	98.0	95.0	69.5	82.3	85.4	94.4	95.7	84.9	91.6	95.0	96.0	**99.4**	97.0	97.8	89.7	85.4	97.3
	✓		93.3	80.9	79.7	82.8	98.0	95.0	69.3	81.8	85.0	94.2	95.6	85.3	91.4	95.0	95.9	99.3	96.9	97.8	89.5	85.3	97.1
	✓	✓	**93.7**	81.6	80.5	83.7	**98.2**	**95.1**	**70.0**	**83.0**	85.9	94.9	**95.8**	**86.3**	91.7	95.4	96.0	**99.4**	97.0	**98.0**	**90.1**	85.9	97.5

	Mouth Slightly Open	Mustache	Narrow Eyes	No Beard	Oval Face	Pale Skin	Pointy Nose	Receding Hairline	Rosy Cheeks	Sideburns	Smiling	Straight Hair	Wavy Hair	Wearing Earrings	Wearing Hat	Wearing Lipstick	Wearing Necklace	Wearing Necktie	Young		Average
FaceTracer [15]	87.0	91.0	82.0	90.0	64.0	83.0	68.0	76.0	84.0	94.0	89.0	63.0	73.0	73.0	89.0	89.0	68.0	86.0	80.0		81.1
PANDA-w [16]	82 .0	83.0	79.0	87.0	62.0	84.0	65.0	82.0	81.0	90.0	89.0	67.0	76.0	72.0	91.0	88.0	67.0	88.0	77.0		79.8
PANDA-l [16]	93.0	93.0	84.0	93.0	65.0	91.0	71.0	85.0	87.0	93.0	**92.0**	69.0	77.0	78.0	96.0	93.0	67.0	91.0	84.0		85.4
LNets+ANet [3]	92.0	95.0	81.0	95.0	66.0	91.0	72.0	89.0	90.0	96.0	**92.0**	73.0	**80.0**	82.0	**99.0**	93.0	71.0	93.0	**87.0**		87.3
Baseline	**93.2**	96.8	86.6	95.4	71.9	96.4	74.9	92.1	94.4	**97.5**	91.5	**80.1**	79.6	**88.1**	98.5	93.3	85.4	95.4	86.8		89.9
Ours	92.9	**96.9**	86.4	95.3	71.3	**96.8**	74.6	**92.3**	94.4	97.3	91.4	79.4	78.9	87.8	98.6	93.0	84.8	95.3	86.2		89.6
	92.8	96.6	86.5	95.4	70.6	96.7	74.4	91.9	94.3	97.3	91.0	79.0	78.3	87.5	98.6	92.7	84.4	95.1	85.8		89.4
	93.0	96.8	**86.7**	**95.6**	**72.6**	96.6	**75.5**	**92.3**	**94.5**	97.4	91.7	**80.1**	79.6	**88.1**	98.5	**93.4**	**85.6**	**95.5**	86.6		**90.0**

示す．ここで，赤点が教師信号，緑点が検出結果を示す．図 4 からも，提案手法はベースラインに比べて高い検出精度を達成していることがわかる．また，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おいてズレが発生している．しかし，提案手法を用いることで改善されていることがわかる．

4.2.2　顔属性推定

ベースラインと提案手法，そして，従来手法の各顔属性の推定結果を表 3 に示す．表 3 から，提案手法は平均推定精度において，従来手法を上回る推定精度を達成していることがわかる．また，各顔属性ごとに比較した場合，提案手法は"Eyeglasses"や"Pale Skin"，"Receding Hairline"など特定のタスクにおいて最も高い推定精度を達成している．提案手法に Channel-wise Convolution を導入したモデルは，導入しないモデルと同等の推定精度となった．しかし，Channel-wise Convolution を導入して各 Task-specific Layers に事前学習したモデルを用いてファインチューニングした場合，平均推定精度が約 90%と最も高い推定精度を達成した．以上のことから，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．

4.3　branch point の変更

提案手法は，Shared Layers，Task-specific Layers ともに同じネットワーク構造を用いている．そして，入力画像を各レイヤーに同時に入力している．そこで，本実験では，図 5 のように，各 Task-specific Layers にタスク共通の特徴マップを入力する branch point について検討する．ネットワークの構造

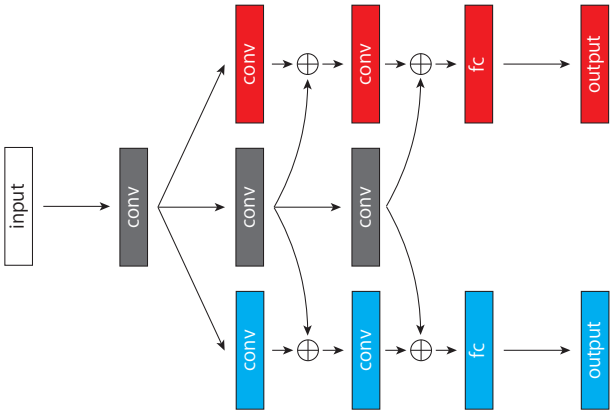

図 5　提案手法における branch point の変更例

は 4.1 節と同様のパラメータを使用する．また，本実験では branch point を表 1 の Conv1-2，Conv2-2，Conv3-3 とした場合の 3 種類のネットワークモデルを実装する．比較対象は，同じ branch point の Hard parameter sharing をベースとしたマルチタスクラーニングをベースラインと変更前の提案手法とする．

branch point の変更による推定精度の変化を表 4 に示す．表 4 から，提案手法は 2 つのタスクにおいて従来手法を下回る推定精度となった．また，提案手法は変更前の方が推定精度が高いことがわかる．このことから，マルチタスクラーニングにおいて，タスク共通の特徴は提案手法のように各タスクに対してより補助的に用いる方が有効であることがわかる．

表 4　branch point の変更による推定精度の変化

	Branch point		
	Conv1-2	Conv2-2	Conv3-3
Facial Landmark Detection (Baseline)	**92.90**	**95.58**	**95.21**
Facial Landmark Detection (Changed Ours)	80.46	79.51	78.85
Facial Attribute Estimation (Baseline)	**89.93**	**89.88**	**89.92**
Facial Attribute Estimation (Changed Ours)	88.84	88.85	88.83

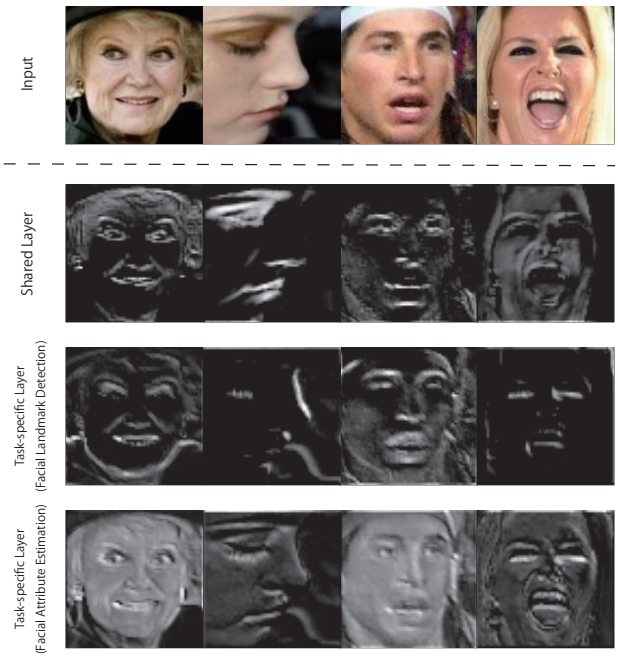

図 6　Separation Multi-task Networks の特徴マップの可視化結果例

4.4　特徴マップの可視化

Separation Multi-task Networks の解析として特徴マップの可視化を行う．Separation Multi-task Networks の Shared Layers と各 Task-specificLayers における表 1 の Conv1-2 で出力される特徴マップの可視化結果例を図 6 に示す．図 6 から，Task-specific Layers(顔器官点検出) は Shared Layers で捉えていないエッジなどの特徴を抽出していることがわかる．同様に，Task-specific Layers(顔属性推定) は顔のシワなどのコンテキストを抽出していることがわかる．このことから，Separation Multi-task Networks はタスク間で共通する特徴と各タスク固有の特徴を分けて抽出しているといえる．

5.　おわりに

本研究では，タスク間で共通する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手法として Separation Multi-task Networks を提案した．提案手法を用いることで，タスク共通の特徴と各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った．さらに，Channel-wise Convolution を導入することで，入力する特徴マップのチャンネル数を調整できるようにした．評価実験では，CelebA による顔器官点検出と顔属性推定を行い，提案手法がどちらのタスクにおいても従来手法を上回る推定精度を達成したことで有効性を示した．

今後の課題としては，顔画像以外への応用が挙げられる．また，提案手法は 2 段階学習によってタスク共通の特徴とタスク固有の特徴を学習するため，この 2 つの特徴を分けながら End-to-End で学習可能なネットワークモデルへの改善も検討する．

文　献

[1] Y. Zhao, F. Tang, W. Dong, F. Huang, and X. Zhang, “Joint face alignment and segmentation via deep multi-task learning,” Multimedia Tools and Applications, pp.1–18, 2018.

[2] J.-J. Lv, X. Shao, J. Xing, C. Cheng, X. Zhou, et al., “A deep regression architecture with two-stage re-initialization for high performance facial landmark detection,” Computer Vision and Pattern Recognition, p.4, 2017.

[3] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” International Conference on Computer Vision, pp.3730–3738, 2015.

[4] Z. Zhang, P. Luo, C.C. Loy, and X. Tang, “Facial landmark detection by deep multi-task learning,” European Conference on Computer Vision, pp.94–108, 2014.

[5] R. Caruana, “Multitask learning,” Learning to learn, vol.1, pp.95–133, Springer, 1998.

[6] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, “Cross-stitch networks for multi-task learning,” Conference on Computer Vision and Pattern Recognition, pp.3994–4003, June 2016.

[7] C. Feichtenhofer, A. Pinz, and A. Zisserman, “Detect to track and track to detect,” International Conference on Computer Vision, pp.3038–3046, 2017.

[8] P. Liu, X. Qiu, and X. Huang, “Adversarial multi-task learning for text classification,” Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp.1–10, 2017.

[9] J. Dai, K. He, and J. Sun, “Instance-aware semantic segmentation via multi-task network cascades,” Conference on Computer Vision and Pattern Recognition, pp.3150–3158, 2016.

[10] L. Duong, T. Cohn, S. Bird, and P. Cook, “Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser,” Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol.2, pp.845–850, 2015.

[11] Y. Yang and T.M. Hospedales, “Trace norm regularised deep multi-task learning,” arXiv preprint arXiv:1606.04038, pp.1–4, 2016.

[12] S. Ruder, “An overview of multi-task learning in deep neural networks,” arXiv preprint arXiv:1706.05098, pp.1–14, 2017.

[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” International Conference on Learning Representations, pp.1–14, 2015.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Computer Vision and Pattern Recognition, pp.770–778, 2016.

[15] N. Kumar, P.N. Belhumeur, and S.K. Nayar, “Facetracer: A search engine for large collections of images with faces,” Oct. 2008.

[16] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev, “Panda: Pose aligned networks for deep attribute modeling,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp.1637–1644, 2014.