

様々な観測角度の歩行映像に対する マルチタスク学習を用いた人物年齢推定

山野 広大¹ 村松 大吾¹ 八木 康史¹

概要: 機械学習やパターン認識, 画像処理の分野の技術発展に伴い, 生体特徴を用いた人物の性質を推定する様々な手法が提案されている. 人物の性質の推定の中でも年齢推定はマーケティングの観点から非常に注目を集めている. また, 生体特徴に中でも歩行行動に基づく特徴は映像中小さく映った人物であっても解析できる可能性があり, 防犯カメラ映像を対象とした解析に利用可能であると考えられる. しかし, 防犯カメラ映像では, 人々は様々な方向に歩いているため, 観測方向の違いに頑健な手法が求められる. その手法の一つとして, 様々な観測角度の歩行映像を用いて一つのモデルを作成する方法がある. この方法により, 観測角度に頑健な手法を構築することが可能であるが, この方法は, 観測角度に関する情報を一切利用しない方法である. 著者らは観測角度の情報を有効に利用することで精度改善が実現できるのではないかと考えた. そこで, 本研究では様々な角度から撮影された歩行データに対して, 観測角度の情報を利用して年齢推定を行う手法を二つ検討した. 一つ目は, マルチタスク学習により, 観測角度の分類と年齢推定を同時に行う手法, 二つ目は, 敵対的な学習により, 抽出する特徴から観測角度の情報を取り除く手法である. 実験では2~87歳の10,307人からなる公開データベースを用いて評価を行なった. その結果, 大幅な精度改善には至らなかったものの, 精度改善の可能性を確認でき, また精度改善のための今後の課題が明らかとなった.

1. はじめに

人間の歩行行動に基づく特徴は一般的に歩容と呼ばれる. 歩容は歩幅や腕の振り, 関節の曲がり具合など様々な情報を含んでいることが特長である. また, 指紋や虹彩に基づく特徴と比べ, 対象人物とデータ取得用センサの距離が離れていても, 個人認証や人物属性に有効な特徴が得られることから多方面の分野での利用が検討されている.

本稿では, 歩容を用いた人物年齢推定を対象とする. 人物の性質の推定の中でも年齢推定はマーケティングの観点からも非常に注目されており, 歩容に基づく年齢推定に関する研究は多岐にわたる [1], [2], [3], [4], [5]. これらの年齢推定の既存手法は, 人物を横から撮影した歩行映像のみを対象としているものが多い. しかし, 歩容解析で対象となる映像は横向きだけとは限らない. 特に防犯カメラ映像などを解析する場合には, 様々な方向の歩行映像が解析対象となりうるが, 様々な方向から撮影された歩行映像を対象とした研究は, 限られる.

阪田らは, 様々な方向から撮影された歩行映像を対象とした年齢推定手法を検討している [6]. この研究では 14 方

向で撮影された歩行データに対し, 角度ごとにそれぞれ個別のモデルを作成する手法と, 角度に依らない統一モデルを作成する手法を用いて年齢推定精度を評価している. この評価結果においては, 角度毎に個別のモデルを構築するよりも, 統一モデルを構築したほうが, 高い精度が得られる, と報告されている. 歩容は観測角度によりその見えが大きく変化するため, 人間による歩行理解では, 角度の違いを考慮しつつ処理が行われていると考えられる. その点を考慮すると, 角度の違いを認識しつつ処理する手法のほうが有効と考えられる. そこで, 本研究では様々な観測角度で撮影された歩容画像に対する年齢推定において, その観測角度の情報を利用する手法を検討する. 具体的には, マルチタスク学習 [7] の主タスクとして年齢回帰タスク, 補助タスクとして角度分類タスクを同時に行う角度情報受容手法を検討する. また, 敵対的学習 [8] の枠組みを用いて歩容画像から特徴を抽出する段階で角度の情報を取り除く角度情報排除手法も検討した. 本稿では, 10,307人からなる公開歩容データベースを利用した評価結果を報告する.

2. 角度情報受容手法

この章ではマルチタスク学習の主タスクとして年齢の回帰問題, 補助タスクとして撮影角度の分類問題を与え, 年

¹ 大阪大学
Osaka University

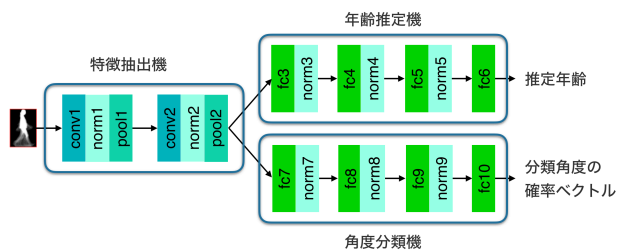


図 1 角度情報受容手法のネットワーク構造

年齢推定と撮影角度の分類を同時に達成しようとするアプローチについて述べる。

2.1 概要

マルチタスクでは補助タスクを設定することで主タスクの性能向上が期待できる場合がある。また、人間は歩行映像から年齢推定を行うときに角度について認識しているという事実が存在する。以上の点を踏まえ、補助タスクとして角度分類を行うことで主タスクである年齢推定の性能が向上するのではないかと仮説のもと研究を行なった。本手法では被験者の年齢推定と観測角度の分類を同時に行うことで、年齢推定と角度分類に共通して有用な特徴抽出が期待できる。

2.2 ネットワーク構造

本手法で用いるネットワーク構造は白神らの GEINet[9]をもとに作成した。GEI を入力とし、畳み込み層 (conv)、最大プーリング層 (pool)、正規化層 (norm) の組み合わせを 2 組、全結合層 (fc)、正規化層の組み合わせを 3 組からなるネットワークである。その構造を図 1 に示す。このネットワークは大きく分けると特徴抽出器、年齢推定器、角度分類器の 3 つから成り立っている。特徴抽出器は畳み込み層、最大プーリング層、正規化層で構成されており、入力となる GEI から有効な情報を抽出することを目的としている。年齢推定器、角度分類器は全結合層、正規化層から構成されており、特徴抽出器が生成した特徴に対してそれぞれ目的となる年齢の推定や角度の分類をすることを目的としている。ネットワークの各層のパラメータの設定を表 1 に示す。fc6, fc10 層は出力層であるため年齢推定の出力である fc6 のユニット数は 1、観測角度分類のユニット数は観測角度の分類数である 14 とした。

2.3 損失関数

マルチタスク学習では各タスクに損失関数を設け、それぞれの損失関数の重み付き和をネットワーク全体の損失関数として設定することが一般的である。ここでは年齢推定で用いた損失関数、観測角度分類で用いた損失関数、ネットワーク全体の損失関数について説明する。

N 個の学習データがあるとき、学習データセットの各

表 1 各層の設定

層名	カーネル数	カーネルサイズ	ストライド	ユニット数
conv1	81	5 × 5	1	-
pool1	-	3 × 3	2	-
conv2	45	7 × 7	1	-
pool2	-	2 × 2	2	-
fc3,7	-	-	-	1024
fc4,5,8,9	-	-	-	256
fc6	-	-	-	1
fc10	-	-	-	14

データを $(x_n, t_n, \mathbf{v}_n), n = 1, 2, \dots, N$ とする。ここで x_n は n 番目の入力 GEI, t_n はその年齢ラベル, \mathbf{v}_n は、角度に関する M 次元の One-hot ベクトルある。年齢推定においては被験者の実年齢 t_n と推定年齢 y_n との平均絶対誤差 (MAE: Mean Absolute Score) を損失関数として設定した。ここで、入力を x とし、特徴抽出器の関数を $f(x; \omega_e)$, 年齢推定のための FC 群の関数を $g(\cdot; \omega_a)$, 角度分類のための関数を $h(\cdot; \omega_v)$ とすると推定年齢は

$$y_n = g(f(x; \omega_e); \omega_a) \quad (1)$$

と計算され、この推定年齢と、正解ラベルの差の絶対値が MAE であり、次の式で計算される。

$$MAE = \frac{1}{N} \sum_{n=1}^N |t_n - y_n| \quad (2)$$

撮影角度分類においては正解の撮影角度 v_n と推定角度 \mathbf{Z} との交差エントロピー (cross entropy) を損失関数として設定する。ここで、 $\mathbf{v}_n = [v_{n1}, v_{n2}, \dots, v_{nM}]^T$ は One-hot ベクトルであり、 $\mathbf{Z}_n = [Z_{n1}, Z_{n2}, \dots, Z_{nM}]$ はネットワークの観測角度に関する M 次元の出力ベクトルである。

$$L_v = -\frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M v_{nm} \log Z_{nm} \quad (3)$$

となる。ここで、 $Z_n = h(f(x_n; \omega_e); \omega_v)$ である。

ネットワーク全体の損失関数は、重みの係数 λ を用いて

$$L(\omega_e, \omega_a, \omega_v) = \frac{1}{N} \sum_{n=1}^N (|t_n - y_n| - \lambda \sum_{m=1}^M v_{nm} \log Z_{nm}) \quad (4)$$

と表現できる。

2.4 学習方法

学習ではこの損失関数を最小にするパラメータを求める。学習により求められるパラメータセット

$$[\omega_e^*, \omega_a^*, \omega_v^*] = \underset{\omega_e, \omega_a, \omega_v}{\operatorname{argmin}} L(\omega_e, \omega_a, \omega_v) \quad (5)$$

となり、学習データを利用してこのパラメータを設定する。学習が終了した後、未知の入力データ x に対する推定年齢 y は、

$$y = g(f(x; \omega_e^*); \omega_a^*) \quad (6)$$

で求める。

3. 角度情報排除手法

この章ではマルチタスク学習の主タスクとして年齢の回帰推定問題、補タスクとして撮影角度の分類問題を与え、敵対的学習の枠組みを利用し、抽出特徴に観測角度の情報が含まれないように、積極的に取り除く手法について述べる。

3.1 概要

本手法では観測角度の分類を行う分類器と、年齢推定を行いながら観測角度の分類を行っていく生成器を交互に学習することによって観測角度情報を取り除いた特徴を抽出し、その特徴を利用して年齢推定を行うことで観測角度に依存しない年齢推移の達成を目指す。角度情報を取り除くことは、角度不変の特徴を抽出することに繋がると考えられ、角度不変の特徴が抽出できれば、特徴空間において複雑な推定モデルが不要となり、汎化性の高いモデルの構築が期待できると考えた。

2章で議論した手法では年齢を推定するタスクと角度を分類するタスクを組み合わせることで、年齢情報も角度情報も保持する特徴の抽出を目指したが、本章で検討する手法は年齢情報を保持しているが、角度情報は保持しないという点が大きく異なる。この角度情報を保持しないように実装するには、少し工夫が必要になる。単純に式(3)で示す角度に関する Loss を最大化するように損失関数を設定し学習してもうまくいかない。それは、角度情報を保持しつつ、Loss を高くするように意図的に間違えるような角度識別器 $h(\cdot; \omega_v)$ を作ればよいことになり、これでは角度情報をなくす、という目的を達成できない。一方で、損失関数を工夫し、エントロピーが最大になるようにマルチタスクで学習する方法も考えられるが、この場合も、角度識別器 $h(\cdot; \omega_v)$ のみを工夫することで、それが実現できる可能性があり、抽出特徴 $f(x; \omega_e)$ から角度の情報を抜く、という目的を同じく達成できない。そこで、敵対的に学習を行う方法を検討する。本手法では、2章で検討した角度情報受容手法により構築されたネットワークを初期ネットワークとして学習を開始する。この時点で、抽出される特徴は、年齢情報も角度情報も保有している。そして角度分類器は、その特徴から角度分類ができるように学習されている。次のステップでは、特徴抽出器と年齢推定器の学習を行う。なお、以降では特徴抽出器と年齢推定器を合わせたものを推定器と呼ぶ。この学習では観測角度を分類する層の角度分類器のパラメータは固定し、この角度分類器において、角度分類が困難かつ年齢推定が可能となるような特徴空間の生成・特徴空間からの年齢推定を行えるようにする。その次のステップでは、角度分類器のみの学習を行う。この学習では特徴空間を生成する特徴抽出器のパラメータを固定し、特徴抽出器が作成した特徴空間に対して角度の

分類の精度が上がるように学習する。推定器と角度分類器を交互に学習することで観測角度情報が排除された年齢推定に有効な特徴の抽出を狙う。なお、推定器と角度分類器の繰り返しを何回行うかはハイパーパラメータとする。

3.2 ネットワーク構造

本手法では図1と同じネットワークを利用した。なお、ネットワークの各層のパラメータの設定も角度情報受容手法と同じものを用いた。

3.3 損失関数

本手法は推定器と角度分類器に交互に異なる学習を行わせる。そのため推定器学習時と角度分類器学習時で異なる損失関数を設定した。そこで以下の3つの損失関数を用いた。

- (1) 年齢推定を行うための損失関数
- (2) 撮影角度を正しく分類するための損失関数
- (3) 撮影角度をわからなくするための損失関数

以上の損失関数のうち1, 3は推定器学習時に用い、2は角度分類器学習時に用いる。また、1は式(2)、2は式(3)を用いる。

3.3.1 推定器学習

特徴抽出器は固定された角度分類精度を下げつつ年齢推定精度が上がるような特徴を抽出することが望まれ、年齢推定器はその抽出された特徴から精度よく年齢を推定できることが望まれる。そのため主タスクとして年齢の回帰推定問題、補助タスクとして角度推定精度を下げるような分類問題のマルチタスク学習を行う。主タスクである年齢の回帰推定問題の目的関数は式(2)で示したMAEを用いた。

補助タスクである角度分類精度を下げるような分類問題に対しては以下の損失関数を使用した。

$$L_{sub} = \frac{1}{N} \sum_{n=1}^N \left| \log \frac{1}{C} - \sum_{m=1}^M v_{nm} \log Z_{nm} \right| \quad (7)$$

ここで、 C は分類クラス数を表す。分類器では角度分類の精度を下げるのが目的であるため、推定結果の精度が $\frac{1}{C}$ となることが望ましい。今回の問題設定では、角度分類タスクは14角度のクラス識別であるため、 $C = 14$ を交差エントロピーの式に代入すると2.639となる。つまり、交差エントロピーが2.639に近づくほど分類しにくくなっていると言える。推定器と角度分類器を組み合わせたネットワーク全体の損失関数は、重みの係数 λ を用いて

$$L = \frac{1}{N} \sum_{n=1}^N (|t_n - y_n| + \lambda \left| \log \frac{1}{C} - \sum_{m=1}^M v_{nm} \log Z_{nm} \right|) \quad (8)$$

とする。

3.3.2 角度分類器学習

分類器は与えられた特徴空間に対して撮影角度の分類精

度を上げることが目的なので損失関数として式 (3) で示した交差エントロピーを用いて学習を行う。ここで重要なことは、この学習は、直前の推定器学習により構築された特徴抽出器から抽出される、角度分類が困難になった特徴に対して、角度分類をするように学習することである。この枠組みが敵対的であり、この繰り返しにより、角度情報を排除した特徴抽出の実現を目指す。

3.4 学習方法

角度情報排除手法の学習の流れを図 2 に示す。まず、角度情報受容手法で構築されたネットワークを読み込む。この時点で、抽出される特徴は年齢情報も角度情報も保有している。そして、年齢推定器、角度分類器は抽出された特徴に対してそれぞれ年齢推定、角度分類ができるように学習される。次に、推定器の学習を行う。この学習では角度分類器のパラメータは固定し、この角度分類器において、角度分類が困難かつ年齢推定が可能となるような特徴空間の生成・特徴空間からの年齢推定を行えるように学習する。その次のステップでは、角度分類器のみの学習を行う。この学習では特徴空間を生成する特徴抽出器のパラメータを固定し、特徴抽出器が作成した特徴空間に対して角度分類の精度が向上するように学習する。そして推定器と角度分類器の学習は、繰り返しステップ数 S 回繰り返させる。

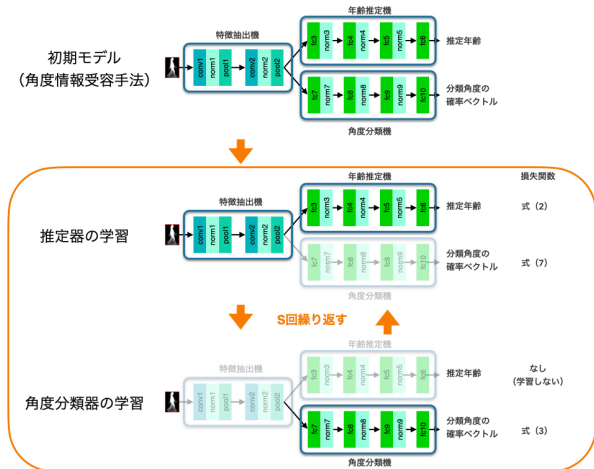


図 2 角度情報排除手法の学習の流れ

4. 実験

4.1 データセット

本実験は OU-ISIR Multi-View Large-Population (MVLPL) [10] のデータセットを用いて行った。被験者の歩行コースは、コースの中心と一致する円の 4 分の 1 に沿うように 15 度間隔で 7 台のカメラが設置されている。このコースを被験者が往復で歩くことにより 0 度から 90 度、180 度から 270 度の 14 の観測角度からの歩行映像が撮影され、この映像を処理することで、歩行者のシルエット画

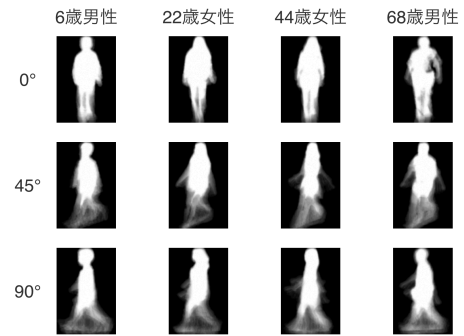


図 3 データセット内の GEI サンプル

表 2 角度情報受容手法のハイパーパラメータ

角度情報受容手法のハイパーパラメータ	
ミニバッチサイズ	128
エポック数	100
初期学習率	0.001
活性化関数 (最終層を除く)	ReLU
活性化関数 (fc6)	恒等写像
活性化関数 (fc10)	Softmax
最適化手法	Adam
Dropout(pool2 - fc3, pool2 - fc7)	0.8
Dropout(norm3 - fc4, norm4 - fc5, norm5 - fc6, norm7 - fc8, norm8 - fc9, norm9 - fc10)	0.5

像列を得た。被験者は 2~87 歳の 10,307 人 (男性 5,114 人・女性 5,193 人) で構成されている。データセット内の GEI サンプルを図 3 に示す。提案した手法評価は同一の方法でデータセットを訓練データ、バリデーションデータ、テストデータに分割して用いている。

4.2 実験設定

4.2.1 ネットワーク学習設定

角度情報受容手法の学習のハイパーパラメータを表 2 に、角度情報排除手法の学習のハイパーパラメータを表 3 に示す。これらの設定は阪田らの様々な方向から撮影された歩行映像を対象とした年齢推定手法 [6] の設定に従った。

4.2.2 ネットワークパラメーター設定

一般的に、学習回数を増やすと損失関数は減少していくが、ある一定以上学習を続けると学習データでは正解率が高い一方でバリデーションデータでは正解率が低くなってしまふ、つまり、学習データだけに最適化されてしまふ汎用性が低い状態に陥る過学習を防ぐために学習ではドロップアウトを採用した。なお、テストで用いるネットワークの重みは両手法とも全エポックの中でバリデーションの損失関数の値が最小となったモデルの重みを採用した。

4.3 評価方法

いずれの提案手法も前述の MAE を用いて評価を行う。また、検証方法としては 8-分割交差検証を用いた。さらに、

表 3 角度情報排除手法のハイパーパラメータ

ミニバッチサイズ	128
生成器エポック数	20
分類器エポック数	10
初期ネットワーク	角度情報受容手法の $\lambda = 5$ で作成した ネットワーク
初期学習率	0.001
活性化関数 (最終層を除く)	ReLU
活性化関数 (fc6)	恒等写像
活性化関数 (fc10)	Softmax
最適化手法	Adam
Dropout(pool2 - fc3, pool2 - fc7)	0.8
Dropout(norm3 - fc4, norm4 - fc5, norm5 - fc6, norm7 - fc8, norm8 - fc9, norm9 - fc10)	0.5

表 4 各モデルの年齢推定における MAE の比較

手法名	MAE[歳]
シングルタスク	7.034
角度情報受容手法 ($\lambda = 1$)	6.915
角度情報受容手法 ($\lambda = 5$)	6.980
角度情報受容手法 ($\lambda = 10$)	7.006
角度情報排除手法 ($\lambda = 3, S = 1$)	7.010
角度情報排除手法 ($\lambda = 3, S = 2$)	6.949
角度情報排除手法 ($\lambda = 3, S = 3$)	6.928
角度情報排除手法 ($\lambda = 5, S = 1$)	7.083
角度情報排除手法 ($\lambda = 5, S = 2$)	7.211
角度情報排除手法 ($\lambda = 5, S = 3$)	7.230

学習データは全学習データの 9 割をパラメータ学習のための訓練データに, 1 割をバリデーションデータに分割した。

4.4 結果

まず, 各手法で得たモデルごとの MAE を表 4 に示す。

4.4.1 角度情報受容手法の結果

まず, 角度情報受容手法の式 (5) 内にあるパラメータ λ を変更して学習を行った場合の学習曲線を図 4, 5, 6, に示す。この学習曲線は, パラメータ λ 毎で, 同一の訓練データ, バリデーションデータを用いた結果を示している。縦軸に損失関数の値, 横軸にエポック数をとっている。訓練データに対する損失関数の値はエポックの増加とともに減少し続けているが, バリデーションデータに対する損失関数の値は 20 エポック前後を境に減少が止まっていることが確認できる。

次に, 角度情報受容手法で得たモデルごとの撮影角度の分類精度を表 5 に示す。表 4 より, MAE に関しては $\lambda = 1$ の場合が最も良く, λ の値を大きくするに従って MAE は悪くなっていることが読み取れる。表 5 より, 角度分類精度に関しては λ の値を変化させても精度に大きな違いは表れていないことがわかる。なお, シングルタスクの年齢推

表 5 各モデルの年齢推定における角度の分類精度

手法名	角度分類精度
角度情報受容手法 ($\lambda = 1$)	0.996
角度情報受容手法 ($\lambda = 5$)	0.996
角度情報受容手法 ($\lambda = 10$)	0.996

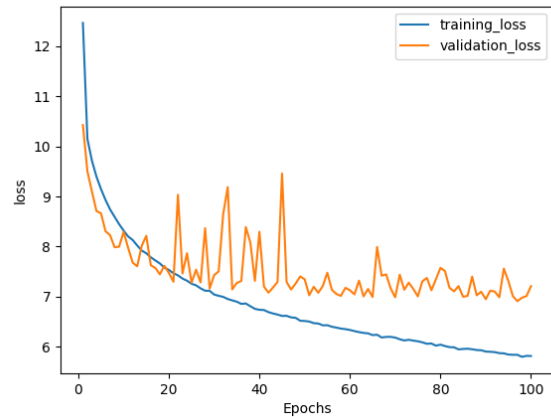


図 4 角度情報受容手法における $\lambda = 1$ での損失関数学習曲線

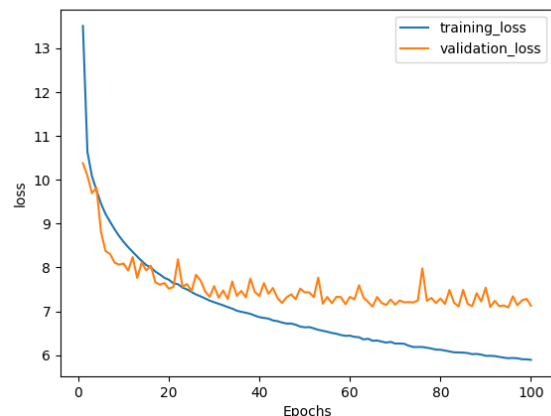


図 5 角度情報受容手法における $\lambda = 5$ での損失関数学習曲線

定はマルチタスクと同じ条件で行なっている。

次に各モデルの撮影角度ごとの MAE を図 7 に示す。図 7 からは, 全ての観測角度においてシングルタスクよりも角度情報受容手法がわずかに性能が上回っていることが確認できる。

4.4.2 角度情報排除手法の結果

角度情報排除手法において, 推定器と角度分類器の繰り返し回数は非常に重要なハイパーパラメーターとなる。ここでは繰り返し回数を繰り返しステップ数 S を用いて表現する。

$\lambda = 3$ の場合の MAE 学習曲線を図 8 に, $\lambda = 3$ の場合の角度分類精度学習曲線を図 9 に, $\lambda = 5$ の場合の MAE 学習曲線を図 10 に, $\lambda = 5$ の場合の角度分類精度学習曲線を図 11 に示す。図 8, 10 では, 縦軸に MAE の値, 横

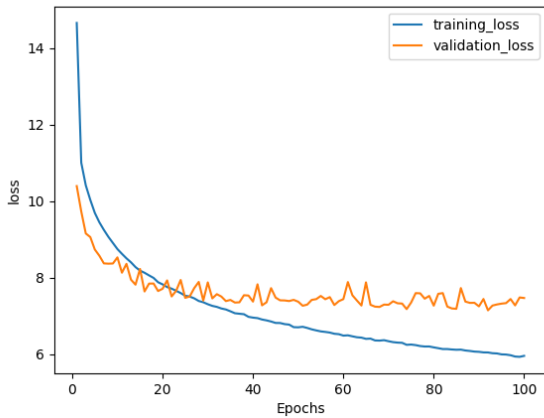


図 6 角度情報受容手法における $\lambda = 10$ での損失関数学習曲線

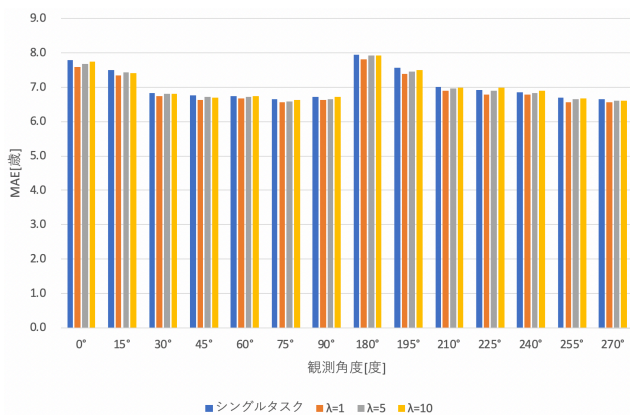


図 7 角度情報受容手法の各モデルの撮影角度ごとの MAE

軸にエポック数をとっている。図 9, 11 では、縦軸に角度分類精度の値、横軸にエポック数をとっている。これらのエポック数とは、推定器、角度分類器のエポック数は区別せず表記している。つまり、0~20 エポックは推定器のステップ 1, 20~30 エポックは角度分類器のステップ 1, 30~50 エポックは推定器のステップ 2 という表記である。図 8 から、推定器のパラメータ学習データに対する MAE はステップを増加させるごとに減少しているのに対して、バリデーションデータに対する MAE はステップ 2, ステップ 3 で振動していることが確認できる。図 10 から、推定器の MAE はステップ 2, ステップ 3 の初めの 1~5 エポックほどで MAE が上昇し、その後減少する傾向があることが読み取れる。これは、式 (8) で示した推定器全体の損失関数の値を下げるために、MAE が下がった時の全体の損失関数の減少度合いと比較して、角度分類精度が下がった時の全体の損失関数の減少度合いの方が大きいためであると考えられる。図 9, 11 から、推定器の学習では期待通り角度分類精度が下がり、角度分類器の学習では角度分類精度が上がっていることが確認できる。角度情報排除手法で得たモデルごとの MAE を表 4 に示す。なお、シングルタスクの年齢推定はマルチタスクと同じ条件で行なってい

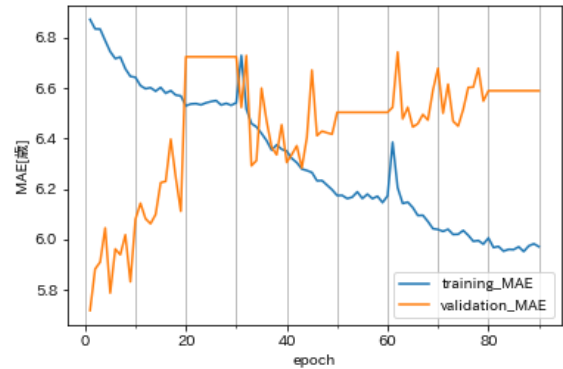


図 8 角度情報排除手法における $\lambda = 3$ での MAE 学習曲線

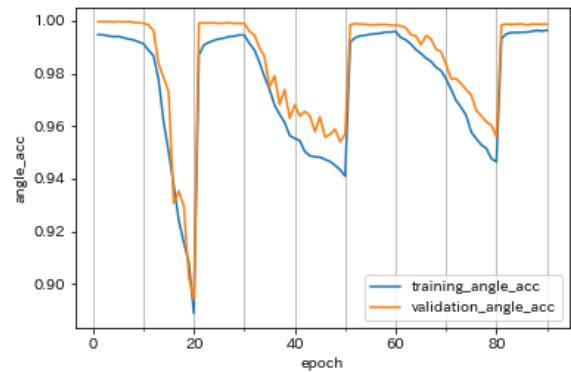


図 9 角度情報排除手法における $\lambda = 3$ での角度分類精度学習曲線

る。表 4 から、 $\lambda = 3$ の場合はステップ数を大きくするにつれて MAE がわずかに向上しており、 $\lambda = 5$ の場合はステップ数を大きくするにつれて MAE がわずかに悪化していることがわかる。また、 $\lambda = 3$ ではすべてのモデルにおいてシングルタスクよりもわずかに推定精度が良いのに対して、 $\lambda = 5$ ではすべてのモデルにおいてシングルタスクよりもわずかに推定精度が悪いことが読み取れる。次に角度情報排除手法の一部のモデルの撮影角度ごとの MAE を図 12 に示す。図 12 からは、 $\lambda = 3$ の場合は、ほとんどの観測角度においてシングルタスクよりも角度情報排除手法がわずかに性能が上回っており、 $\lambda = 5$ の場合は、ほとんどの観測角度においてシングルタスクよりも角度情報排除手法がわずかに性能が下回っていることが確認できる。

4.5 各手法の比較

表 4 から、角度情報受容手法の λ を 1 に設定した場合が最も精度が高くなったことがわかる。また、両手法とも最も精度が高い設定ではシングルタスクの精度をわずかに上回っている。

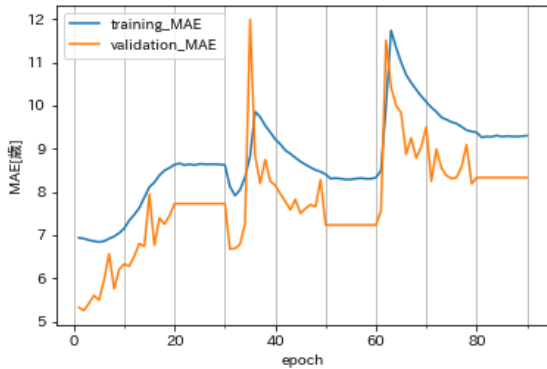


図 10 角度情報排除手法における $\lambda = 5$ での MAE 学習曲線

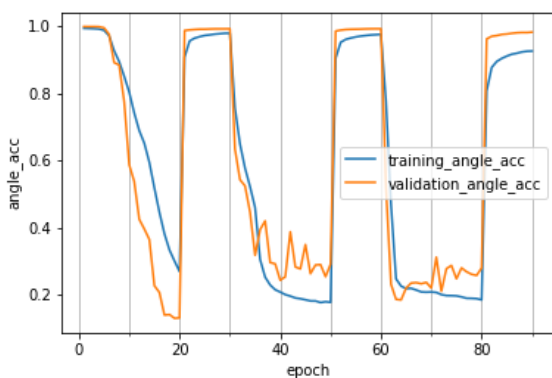


図 11 角度情報排除手法における $\lambda = 5$ での角度分類精度学習曲線

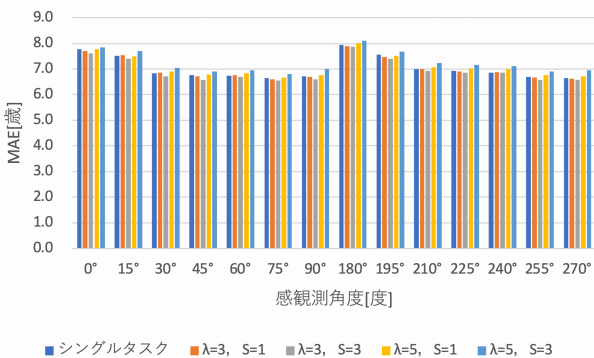


図 12 角度情報排除手法の一部のモデルの撮影角度ごとの MAE

5. 考察

5.1 各手法の学習過程に対する考察

図 4, 6 で示した角度情報受容手法の学習曲線から, 角度分類に関する損失関数の重みを増加させるに従ってバリデーションデータでの損失関数の値のばらつきが小さくなっていることがわかる. これは, 年齢推定と比較して角度分類が容易であり, 角度分類に関する損失関数が安定して小さいことが要因として考えられる. つまり, 全体の損失関数のうち, 安定して小さい角度分類に関する損失関数

の割合が大きくなることでばらつきが小さくなっていると考えた.

5.2 各手法の結果に対する考察

図 7 から, 観測角度情報を用いないシングルタスクの手法と比較して, 補助タスクとして撮影角度の分類問題を与えた角度情報受容手法は全ての撮影角度においてわずかに年齢推定の精度が向上していることがわかる. これは, 補助タスクとして角度分類を行うことで主タスクである年齢推定の性能が向上するという仮説が正しかったためであると考えられる. また, 表 4 から, 角度分類に関する損失関数の重みを増加させるに従って年齢推定の精度がわずかに低下していることがわかる. これは, 補助タスクの損失関数の重みを増加させるに連れて, 学習時のパラメータ更新において補助タスクの損失関数を下げようとする動きが大きくなるためであると考えられる.

角度情報排除手法は推定器, 角度分類器の各エポック数や, 繰り返すステップ数 S , 式 (8) 中の重み係数 λ , 初期ネットワークで用いるモデルなど, 角度情報受容手法と比較してハイパーパラメータの種類が多い. シングルタスクの手法と比較して, 制度の大幅な改善がみられないことから, それぞれのハイパーパラメータを変更しての比較が必要であり, 最適なパラメータを設定した時に推定精度が向上することは十分考えられる.

5.3 各手法の比較に対する考察

表 4 から, 各手法の MAE を比較して, 角度情報受容手法の λ を 1 に設定した場合が最も精度が高くなったことが読み取れる. しかし, シングルタスクの MAE と比較して性能が大幅に改善されたとは言えない. また, 角度情報排除手法はシングルタスク, 角度情報受容手法と比較して大幅な性能改善は達成できていない. これは, 推定器は固定された角度分類精度を下げつつ年齢推定精度が上がるような特徴を抽出することでできていないことが要因の一つであると考えた. 期待している特徴が抽出できていない原因としては, 式 (7) で示した, 角度分類を下げるような損失関数が期待した効果を実現できていないことが考えられる. 式 (7) の損失関数では, 入力された GEI に対して, 角度分類器が出力した推定結果の精度が $\frac{1}{C}$ となるようにしているが, 出力された推定結果以外の $C-1$ つの角度に対しても同様に推定結果が $\frac{1}{C}$ に近づくことが望ましい.

6. おわりに

本稿では既存手法の入力である歩容画像の観測角度が歩行方向に対して真横方向で撮影されているものが多く用いられており, 汎化性が低いという問題があったため, より汎用性の高い手法を構築すべく, 様々な観測角度からの歩容を対象とした年齢推定課題に着目した. この課題におい

ては、角度の違いをうまく考慮することで、年齢推定精度の改善が期待されるため、角度と年齢を同時に推定するアプローチ、特徴空間から角度情報を取り除く二つの異なるアプローチを試みた。

その結果、両手法ともに撮影角度ごとの MAE は年齢のみを推定する手法と比較して、現時点では大幅な改善を実現できていない。その要因としては角度情報受容手法に関しては、式 (4) 中の最適な λ の値を見つけていないといったことや、補助タスクによる主タスクの性能向上を活かしきれていないといったことが考えられる。角度情報排除手法に関しては、式 (7) で示した角度分類精度を下げるような損失関数が期待した効果を実現できていないといったことや、角度分類器の各エポック数や、繰り返しステップ数 S 、式 (8) 中の重み係数 λ 、初期ネットワークで用いるモデルなどのハイパーパラメータが最適でないことが考えられる。これらの要因に対して角度情報受容手法については、バリデーションデータを用いて最適な λ の値を見つけることや、性別などの角度情報以外の情報を補助タスクとして与えるというアプローチ、角度情報排除手法については、角度分類精度を下げるような損失関数に変更することや、最適なハイパーパラメータをバリデーションデータを用いて設定するアプローチを行うことを実施していきたい。

謝辞 この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものである。

参考文献

- [1] 万波秀年, 榎原靖, 八木康史. 歩容における性別・年齢の分類と特徴解析. 電子情報通信学会論文誌 D, Vol. 92, No. 8, pp. 1373–1382, 2009.
- [2] Jiwen Lu and Yap-Peng Tan. Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Transactions on Human-Machine Systems*, Vol. 43, No. 2, pp. 249–258, 2012.
- [3] 奥村麻由, 榎原靖, 八木康史ほか. 大規模歩容データベースを用いたガウス過程回帰による年齢推定の評価. コンピュータビジョンとイメージメディア (CVIM), Vol. 2011, No. 33, pp. 1–8, 2011.
- [4] Haiping Zhu, Yuheng Zhang, Guohao Li, Junping Zhang, and Hongming Shan. Ordinal distribution regression for gait-based age estimation. *Science China Information Sciences*, Vol. 63, No. 2, p. 120102, 2020.
- [5] Atsuya Sakata, Yasushi Makihara, Noriko Takemura, Daigo Muramatsu, and Yasushi Yagi. Gait-based age estimation using a densenet. In *Asian Conference on Computer Vision*, pp. 55–63. Springer, 2018.
- [6] 阪田篤哉, ホサインシャヘラ, 村松大吾, 榎原靖, 八木康史. 様々な角度の歩行映像を用いた人物属性推定. *IEICE-BioX2019-60*, Vol. 119, No. 214, pp. 27–29, 2019.
- [7] Abrar H Abdalnabi, Gang Wang, Jiwen Lu, and Kui Jia. Multi-task cnn model for attribute prediction. *IEEE Transactions on Multimedia*, Vol. 17, No. 11, pp. 1949–1959, 2015.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [9] Kohei Shiraga, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *2016 international conference on biometrics (ICB)*, pp. 1–8. IEEE, 2016.
- [10] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, Vol. 10, No. 1, p. 4, 2018.