Vol.2019-CVIM-217 No.26
2019/5/30

セマンティックセグメンテーション手法を用いたシルエットに基づく歩容認証の性能評価

小川 景矢[†1,a)]　村松 大吾[†1,b)]　槇原 靖[†1,c)]　八木 康史[†1,d)]

概要：防犯カメラなどから取得された低解像度の映像や，顔が確認できない映像からでも個人認証が可能となる歩容認証は，犯罪捜査やマーケティングなどの分野での利用が期待されているものの，現時点においてその利用はあまり進んでいない．それは，主要な歩容認証手法は高精度なシルエット画像の利用を前提としているものの，高精度シルエット画像の自動生成が容易でないためである．一方で，近年画像認識分野においては，深層学習を用いた様々な手法が提案されている．その一つにセマンティックセグメンテーションと呼ばれる，画像内に移っているものに対するカテゴリ分類と領域分割を同時に行う手法があり，これにより人物領域の自動抽出も可能である．そこで，本研究では，代表的なセマンティックセグメンテーション手法である Mask-RCNN と Refinenet-Net を対象とし，これらを用いた歩容認証の大規模歩容データベースに対する性能評価結果を報告する．

1. はじめに

近年我々の周りの防犯カメラの数は増加の一途をたどっており，我々にとって身近なものとなっている．防犯カメラの撮影画像はマーケティング調査や，犯罪捜査などでの利用が期待されている．マーケティング調査では，顧客の買い物時の動作を解析することで販売促進につながると考えられ，犯罪捜査では犯行現場で犯人の撮影や逃走経路の確認などが行えると考えられる．特に犯罪捜査においては，防犯カメラから得られる人物や車などの映像から犯人の逮捕に至った事件もあり，イギリスでは容疑者に対する歩容認証の結果が裁判において証拠として認められた実例もある [1]．しかし，複数の人物が映っている防犯カメラ映像から特定人物を見つける場合や，対象となる一対の人物映像が同一人物か判断する場合において捜査員の目視確認では限界があり効率的ではないため，防犯カメラの映像から特定人物を自動的に検出できるシステムは重要だと考えられる．また，見守り等では，今まさに現場で起きている事象の把握が望まれるため，オンラインで解析できることが望ましい．しかし，防犯カメラの映像に対してオンラインで特定人物の検出をするためには，防犯カメラの映像から人物の検出や，検出された人物の認証を高精度かつ実時間で実施する必要がある．映像から人物の認識を実現する技術の一つに歩容認証 [2] がある．歩容認証は，人の歩き方の特徴を比較することで人物を認証する手法であり，歩容認証は理想的な環境下にて取得された 3,000 人を超える歩行者映像データを用いた精度評価では，94 %の 1 位認証率を達成したと報告されている [3]．日本において 2009 年に犯罪捜査の有力情報として初めて活用され，2013 年には鑑定ソフトウェア [4] が開発されるなど，新たな犯罪捜査の手法として注目されている．歩容認証に関する研究の多くは認証の部分に焦点を当てているものが多く，様々な歩容特徴を用いた手法が提案されている．一般的に歩容認証で用いる歩容特徴は，モデルに基づく歩容特徴 [5] と，見えに基づく歩容特徴 [6] の二つに分類可能である．モデルに基づく歩容認証は人物の頭部や関節などの骨格のモデル化を行い，そのモデルを入力画像に当てはめて照合を行う．しかし画像の解像度が低い場合のモデル化は難しいため，映像自体が低解像度映像であったり，また，対象人物が映像中小さく映っていたりすることが多い防犯カメラの映像ではモデルに基づく人物認証は難しい．そのため，解像度の影響が少なく計算コストも比較的小さい，見えに基づく歩容特徴を用いる認証が現在歩容認証の主流となっている．見えに基づく歩容認証ではシルエット画像などを用いて画像を直接解析し，人物モデルを用いずに特徴を抽出する．実際歩容認証を想定した公開データベースはこのシルエットの形式で公開されているものが多い [3]．しかし，

†1 現在，大阪大学
Presently with Osaka Uniersity
a) k-ogawa@am.sanken.osaka-u.ac.jp
b) muramatsu@am.sanken.osaka-u.ac.jp
c) makihara@am.sanken.osaka-u.ac.jp
d) yagi@am.sanken.osaka-u.ac.jp

見えに基づく歩容認証では高精度のシルエット画像が得られることを想定している．

見えに基づく歩容認証の実現に必要な手続きは，映像からの人物の検出，追跡，シルエット抽出，認証に大別される．人物の検出は映像中の一枚一枚の画像内の人物の位置を推定し，シルエット抽出では，画像の背景領域と人物領域を分離し，人物領域を取得する．追跡では，画像一枚一枚独立に検出された人物の対応付けを行う．認証では，検出，追跡，シルエット抽出の結果から，歩容の特徴を抽出し，抽出した特徴の類似性を評価することで，人物の同一性を判定する．防犯カメラの映像解析に歩容認証が適用可能となれば，捜査員による目視での調査よりはるかに効率的に映像解析が可能になる．そのためには，認証手法の要素である，人物の検出，追跡，シルエット抽出，認証を実行時間を考慮したうえで高精度化する必要がある．本研究では，シルエット抽出について注目することにした．近年，画像認識の分野において成果を上げているニューラルネットワークを用いたセマンティックセグメンテーションの手法を用いることで，人物のシルエット画像を作成し，それを利用して個人認証を行う．本研究では，セマンティックセグメンテーションを利用した場合の，認証に利用するシルエット画像の特性，実行時間，個人認証の認証率について評価していく．歩容認証の手法の中でセマンティックセグメンテーションによって実装されるのは，人物の検出とシルエット抽出である．本研究で利用する個人認証の手法としては，近年画像認識の分野で大きな成功を収めている深層学習に注目した手法で画像データの学習に適している畳み込みニューラルネットワーク (CNN：Convolutional Neural Network) を用いた GEINet [7] であり，画像特有の空間的近接性が考慮され，高精度の認識が可能となる．CNN は生体認証に関しても静脈認証 [8] や歩容認証 [9] で実際に利用されている．本研究では，精度評価には，大規模歩容データベースである OU-MVLP [10] を用いて実験を行った．このデータベースには各被験者に対して観測視点の異なる歩行データが含まれているため，様々な角度における精度評価が可能である．今回はこのデータベース内のデータの中から，歩行者の進行方向から 0°， 45°， 90°， 180°の 4 方向から撮影したデータを用いて評価実験を行い，セグメンテーションの手法を利用した歩容認証の性能評価を行った．

2. 関連研究

2.1 歩容の特徴表現

見えに基づく表現では，服装の色や模様に影響を受けない，シルエットに基づく表現が多く利用されている．例えば平均シルエット [11] として知られている，歩容エネルギー画像 (GEI: Gait Energy Image) [6] は取得した映像から抽出された歩行者のシルエット画像列を 1 歩行周期分で平均することで得られる．他にも 1 周期の歩行画像列から時間的なエントロピーを計算した Gait Entropy Image (GEnI) [12] や，シルエット画像間のオプティカルフローを利用した Gait Flow Image (GFI) [13]，歩行周期に基づく周波数領域特徴を利用した Frequency domain feature (FDF) [14] などがある．

2.2 セグメンテーションの手法

GEI を作成するためにはシルエット画像が必要となる．歩行画像列から歩行者のシルエットを抽出するためにはセグメンテーションにより，歩行者の領域を取得することが考えられる．セグメンテーションの手法は数多く提案されている [15]．グラフカット [16] や時空間において局所的な領域に対して色変換を行った背景差分に基づく手法 [17] などがある．これらに対し，近年はニューラルネットワークを利用したセマンティックセグメンテーションの手法も数多く提案されている．セマンティックセグメンテーションとは，画素レベルでその画素がどの物体かを認識するタスクを行う手法のことである．シルエット画像を作成することは画素レベルで人の領域を取得することと同じであるため，セマンティックセグメンテーションにより人物領域を取得しシルエット画像を作成する．セマンティックセグメンテーションを行うニューラルネットワークの中で，全層畳み込みネットワークの一つである U-Net [18] は局所的特徴と全体的位置情報を同時に利用することで領域抽出を効率よく行うことが可能である．また，Vijay Badrinarayanan らによって提案された SegNet [19] は道路や標識など車を運転する際に必要となる情報を処理することに焦点を置いたネットワークである．ほかにも SegNet を改良した ENet [20]，ERFNet [21] は精度は下がるが実行時間を短縮させることを目的として提案されたネットワークである．

2.3 歩容認証の手法

認証の手法は，歩容特徴間で類似度や相違度を計算することが基本であり，例えば L2 ノルムを用いる手法をはじめ，数多く提案されており，Martin と Xiang が提案したサポートベクターマシンによるランク学習による認証 [22] や Mansur らが提案した異なる視点からの歩行特徴画像に対して視点に依存する射影行列を用いる手法がある．深層学習を歩容認証に利用した事例として，Wu らの研究 [23] や Shiraga [7] らの研究等が挙げられる．その中でも Shiraga らの研究 [7] により提案された GEINet は，同一被験者の複数方向から撮影した画像を訓練データとして学習させることで，観測視点の情報は不要でかつ，視点変化に頑健な歩容認証を可能とする特徴がある．

3. セマンティックセグメンテーションを利用した歩容認証の手法

歩容認証を実現するためにはまず，歩行画像列からシルエット画像群を作成することが必要である．その上で作成したシルエット画像を位置合わせし，アスペクト比を保持したままサイズの正規化をすることで，正規化シルエット画像列 (GSV: Gait Silhouette Volume) を作成し，GSV を時間軸方向に正規化自己相関を計算することで歩行周期を求め，GEI を作成する．作成した GEI を歩容の特徴とし認識を行う．本研究では，シルエット画像を作成するための手法としてセマンティックセグメンテーション手法の訓練済みのネットワークを利用する．今回使用するセマンティックセグメンテーションの手法は Mask R-CNN [24] と Refine-Net [25] を選択した．

3.1 Mask R-CNN

映像からシルエット画像を作成する手法の一つとして Mask R-CNN [24] を利用する．本研究では訓練済みの，facebookresearch [26] で配布されているネットワークを利用する．また，配布されているネットワークは CoCo Dataset で学習したものである．Mask R-CNN は入力された画像を CNN に入力して，対象物体である可能性が高い領域を予想する．ピクセルごとに対象の物体かどうかの判定を行い，同じ物体と判定されたピクセルをまとめることでセグメンテーションを行う．その上で，人物と判定された領域を選択することで，人物シルエットを抽出する．[26] で配布されているプログラムでは，物体領域の候補数をパラメータにより指定することが可能である．本研究では既定値である 1000 と，その半分の 500 をパラメータとして与えた場合の 2 通りで実験を行う．候補数が大きいほど実行時間は増加するが，シルエット抽出の精度が向上すると考えられる．この 2 通りで歩容認証精度に関しても比較を行う．本稿では説明上，パラメータが 1000 のものを scale=1000，パラメータが 500 のものを scale=500 と述べていく．

3.2 Refine-Net

本研究で使用するセマンティックセグメンテーションの手法の 2 つ目が Refine-Net [25] であり，Gousheng Lin らが公開しているネットワーク [25] を利用する．また，利用するネットワークは PASCAL VOC を利用して学習したものである．Refine-Net は，高解像度なセマンティックセグメンテーションのためにマルチパス入力を実現する手法であり，マルチパスの中で低解像の特徴をアップサンプリングし複数の解像度の特徴を融合させることで，高解像度の予測を行える特性を有している．入力画像から ResNet を利用して抽出した複数のサイズの特徴マップをネットワークに入力することで，セグメンテーションを行い結果画像を出力する．配布されているネットワーク [25] では，入力画像を複数の大きさにリサイズし，リサイズした画像毎にセグメンテーションを行い，それを融合することで精度を向上させている．このリサイズの大きさを 3 種類と指定した場合と 5 種類と指定した場合の 2 通りで実験を行う．リサイズの種類が多いほど実行時間は増加するが，シルエット抽出の精度が向上すると考えられる．この 2 通りで歩容認証精度に関しても比較を行う．本稿では説明上，5 種類のものを 5scales，3 種類のものを 3scales と呼ぶ．

3.3 歩容特徴抽出

本研究で利用する GEI の作成には，まず歩行者を撮影した映像から抽出されたシルエット画像に対して外接矩形で人物領域を切り出し，位置合わせをしたうえでアスペクトを保持しながらリサイズし，GSV を作成する．GSV 画像列から自己相関を用いて歩行周期を計算し，歩行 1 周期分の画像列を取り出す．そして取り出した一歩行周期分の GSV について式 (1) で各画素に対して時間方向の平均をとことで，GEI を作成する．画像位置 (x, y) における GEI の値 G(x,y) は

$$G(x,y) = \frac{1}{T}\sum_{t=1}^{T} B_t(x,y) \tag{1}$$

と計算する．ここで，T は一歩行周期中のフレーム数，$B_t(x,y)$ は t フレーム目の GSV における位置 (x,y) の値である．

3.4 CNN による認証

GEI に基づく個人認証手法として Shiraga らが提案した CNN を用いた GEINet [7] を利用する．

3.4.1 GEINet のネットワーク構造

今回利用する GEINet の構造を図 1 に示す．

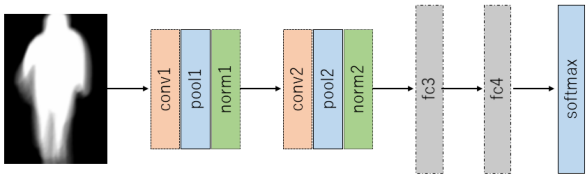


図 1: GEINet のネットワーク構造

GEINet は GEI を入力として受け取り，訓練データの各被験者との類似度を出力する．GEINet は全部で 8 層のネットワークで，図 1 の conv は畳み込み層を，pooling はプーリング層を，norm は正規化層を，fc は全結合層を，softmax はソフトマックス層をそれぞれ示している．畳み込み層は，畳み込み演算により入力データの形状を維持し

たままデータを出力する層であり，画像の空間的特徴を保持することが可能となる．プーリング層は畳み込み層から受け取ったデータに対して空間サイズを小さくし，扱いやすい形に変形するために情報を圧縮しダウンサンプリングする．正規化層ではプーリング層の結果の値を正規化する．この正規化層では，Local response normalization (LRN) [27] を利用している．次の層の fc3 はユニット数 1024 の全結合層であり，活性化関数は ReLU [28]，ドロップアウト [29] を利用している．全結合層 fc4 は訓練に使用する入力データの被験者の数と同じ N 個のユニットを持つ．学習の際にはソフトマックス関数を適用し，各被験者 ID に対する類似度を出力する．したがって，fc4 層の i 番目のユニットは i 番目の被験者との類似度を表すため，認証の際には fc4 層の N 次元ベクトルを特徴量とし認証に利用する，GEINet で設定されている各層の値を表 1 に示す．

表 1: GEINet の各層の値の設定

Layer	Kernels	Size/stride	Act.	Pooling
conv1	18	7 × 7 × 1/1	ReLU	
pool1		2 × 2/2		Max pooling
conv2	45	5 × 5 × 18/1	ReLU	
pool2		3 × 3/2		Max pooling

ここで，Kernels は畳み込み層の出力のチャンネル数，Size は畳み込みフィルタのサイズ，stride はフィルタのストライド数，Act. は活性化関数を表している．

3.4.2 GEINet の学習

GEINet の学習では CNN の重みパラメータ w を M 個の訓練用 GEI データ $\{I_1,...,I_M\}$ を用いて最適化していく．訓練用データには被験者 N 人分の歩容特徴が含まれているとする．入力される GEI を I としたときの GEINet の出力ベクトルの n 次元目の出力は，n 番目の fc4 層の出力を $v_{fc4}(I;w)_n$ とすると，出力ベクトルの n 番目の要素のソフトマックスの出力 $v^{'}(I:\omega)_n$ は

$$v'(I;w)_n = \frac{exp(v_{fc4}(I;w)_n)}{\sum_{j=1}^{N} exp(v_{fc4}(I;w)_j)} \tag{2}$$

で計算した値となる．パラメータの学習の際には以下に示す交差エントロピーの値を損失関数として考慮し，その値を小さくするパラメータを学習する．

$$L_1(w) = -\sum_{m=1}^{M}\sum_{n=1}^{N} \delta_{y_m n} log v'(I_m;w)_n \tag{3}$$

ここで $\delta_{y_m n}$ はクロネッカーのデルタ，y_m は M 番目のデータの被験者 ID とする．したがって，式 (3) を最適化するように学習することで，被験者 ID と対応するラベルに高い値を出力するようになる．

3.4.3 認証

認証時には学習した GEINet を特徴抽出に利用する．一般的に物体認識と人物認証では学習時とテスト時のラベルの持つ意味が異なる．物体認識などでは学習時と認識時で正解ラベルは同じものを利用するが，人物認証では同じとなるとは限らない．CNN では大量の学習データを必要とするが，認証対象人物のデータを大量に用意することは現実的ではないため，人物認証では，大量に準備可能な対象人物以外のデータを学習データとして利用する．その為，出力ラベルには，対象人物のラベルが存在しない．したがって，学習したネットワークの出力ラベルをそのまま認証結果として利用するのではなく，ネットワークを特徴抽出に利用し，ネットワークの中間出力を特徴量とし，比較を行うことで認証とする．本研究では特徴比較に L2 ノルムを計算することで特徴間の相違度を定義する．プローブとギャラリーそれぞれの GEI を I_p, I_g とし，それらを GEINet に入力したときの fc4 層の出力をそれぞれ $v_{fc4}(I_p;w)$，$v_{fc4}(I_g;w)$ とすると，相違度は次のように表される．

$$dist(I_p, I_g; w) = ||v_{fc4}(I_p;w) - v_{fc4}(I_g;w)||_2 \tag{4}$$

式 (4) で計算される相違度は，他人同士の GEI から抽出された特徴間では大きな値となり，本人同士の GEI から抽出された特徴間では小さな値となると考えられる．そこで、この相違度を用いて認証を行う．一対一認証の場合には，計算される相違度としきい値を比較することで本人か他人かの判定を行う．一対 N 認証の場合には，プローブの特徴に対する，ギャラリーの相違度を小さい順に並べ，判定を行う．

3.5 利用したデータセット

手法の評価には，世界最大の歩容データベースである OU-ISIR Multi-View Large Population Dataset(OU-MVLP) [30] を利用する．OU-MVLP では 1 万人を超える被験者の複数回の歩行を 14 方向から撮影している．本実験ではこのうち，進行方向に対し 0°， 45°， 90°， 180° の角度から撮影されたそれぞれに歩行分の歩行映像を利用した．OU-MVLP ではクロマキー処理によって作成された GEI も公開されているため，クロマキーにより作成された GEI とセマンティックセグメンテーション手法を用いて作成した GEI の間でも精度評価を行っていく．

3.6 実験設定

各撮影角度データに，Mask R-CNN の手法と，Refine-Net の手法を各々利用し，歩行者のシルエット画像列を作成する．その後シルエット画像列から，GEI を作成し，GEINet による認証を行う．この実験を通して，セマンティックセグメンテーションの性能を処理時間，シルエット画像のク

表 2: 各角度における処理時間と失敗系列数

手法	角度	処理時間 [sec/frame]	失敗系列 [シーケンス]
Mask R-CNN (scale = 1000)	0°	0.29	34
MAsk R-CNN (scale = 500)		0.27	35
Refine-Net (3scales)		2.7	61
Refine-Net (5scales)		4.3	51
Mask R-CNN (scale = 1000)	45°	0.27	18
MAsk R-CNN (scale = 500)		0.26	19
Refine-Net (3scales)		2.6	292
Refine-Net (5scales)		4.3	291
Mask R-CNN (scale = 1000)	90°	0.27	32
MAsk R-CNN (scale = 500)		0.26	34
Refine-Net (3scales)		2.6	48
Refine-Net (5scales)		4.2	40
Mask R-CNN (scale = 1000)	180°	0.29	160
MAsk R-CNN (scale = 500)		0.28	160
Refine-Net (3scales)		2.7	427
Refine-Net (5scales)		4.4	428

表 3: 歩容認証精度比較 (EER, Rank-1)

手法	角度	EER[%]	Rank-1[%]
Mask R-CNN (scale = 1000)	0°	7.63	65.34
MAsk R-CNN (scale = 500)		8.09	63.79
Refine-Net (3scales)		9.15	60.90
Refine-Net (5scales)		9.00	64.46
クロマキー		5.72	75.67
Mask R-CNN (scale = 1000)	45°	6.31	76.63
MAsk R-CNN (scale = 500)		6.34	75.89
Refine-Net (3scales)		5.86	78.91
Refine-Net (5scales)		6.09	78.47
クロマキー		5.81	82.63
Mask R-CNN (scale = 1000)	90°	4.90	78.55
MAsk R-CNN (scale = 500)		4.89	79.08
Refine-Net (3scales)		5.19	77.30
Refine-Net (5scales)		4.79	78.44
クロマキー		4.59	83.95
Mask R-CNN (scale = 1000)	180°	6.57	71.03
MAsk R-CNN (scale = 500)		6.57	70.16
Refine-Net (3scales)		5.95	74.36
Refine-Net (5scales)		6.29	75.54
クロマキー		4.41	82.66

オリティ，歩容認証の認証精度の観点から評価する．

4. 結果

4.1 セマンティックセグメンテーションによるシルエット抽出

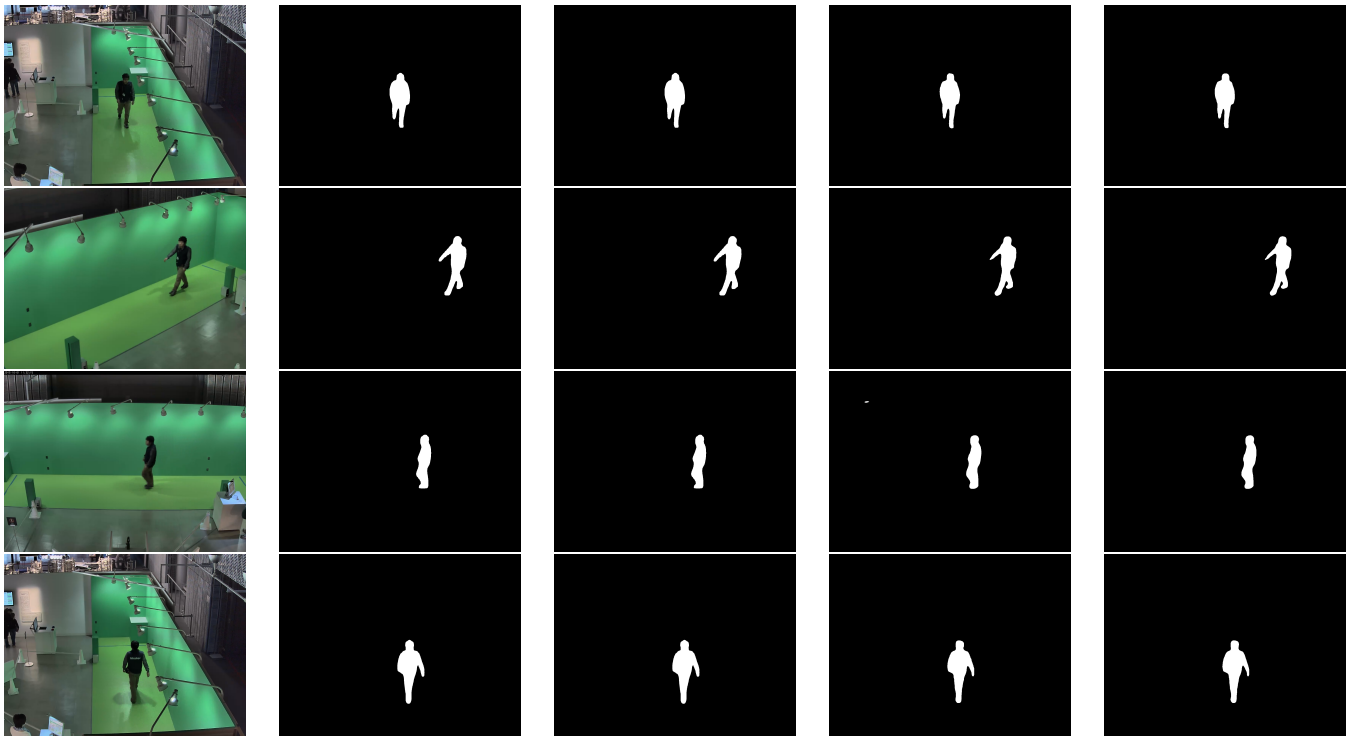
図 2 に，角度の異なる画像を入力したときのそれぞれの手法の出力結果を示す．出力結果のシルエットは大きな誤りがないことがわかる．表 2 に，各手法によるセマンティックセグメンテーションの処理時間と，失敗系列数を示す．ここで，失敗系列とは，人物を検出できなかった画像が存在する画像系列の数である．

セマンティックセグメンテーションの処理時間計測に使用したマシンのスペックは，GeForce GTX 1080 を 1 台，システムプロセッサに Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz, 実装メモリ (RAM)16.0GB，システムの種類は 64 ビット　オペレーティングシステム，x64 ベースプロセッサである．また，処理時間の計測方法は 10 フレームの画像セットを入力として 50 画像セット実行したときの 1 フレームあたりの平均処理時間を計算している．

表 2 より，処理時間と失敗系列数ともに Mask R-CNN のほうが良い結果となっている．Refine-Net のセグメンテーションが失敗した例を図 4 に示す．図 4 では Mask R-CNN では人物領域が正しく抽出できているのに対し，Refine-Net に基づく手法では，人物領域が検出されず，人物領域に対して人物以外のラベリングが行われている．ここでは，上半身を犬，下半身を馬と判断してしまっている．Refine-Net はこのような人物領域を人物以外として判断する誤りが MASK-RCNN に比べて多く発生した．

4.2 歩容認証の精度評価

前述したとおり，セマンティックセグメンテーション手法により，正しくシルエットを検出できるフレーム数が異なるため，本評価では，OU-MVLP で用いられている評価プロトコルを利用するのではなく，検討した 4 つの手法全てにおいて，全てのフレームにおいてシルエットを検出できた被験者データを用いる．利用できるデータのうち，6000 人のデータを学習用，残りのデータをテスト用として利用した．本実験では，0°のデータセットでは　2027 人，45°のデータセットでは　3248 人，90°のデータセットでは 3609 人，180°のデータセットでは　2192 人の歩行映像をテストデータとして利用した．認証精度は一対一認証及び一対多認証それぞれにおいて評価した．一対一認証の精度評価値として，本人拒否率 (FRR: False Rejection Rate) と他人受入率 (FAR: False Acceptance Rate) を計算し，プロットすることで作成できる受信者操作時特性曲線 (ROC 曲線: Receiver Operating Characteristic curve) をプロットするとともに，等誤り率（EER: Equal Error Rate）を示す．また，一対多認証の精度評価として，本人同士の類似度が全体の R 番目以内に入る割合を示した累積識別精度特性曲線 (CMC 曲線：Cumulative Match Caracteristic curve) ををプロットするとともに，1 位認証率（Rank-1）を求める．表 3 には各角度における EER と Rank-1 値をまとめた．また図 4 は，角度ごとの ROC 曲線及び CMC 曲線を示す．



(a) (b) (c) (d) (e)

図 2: 入力画像と処理結果．各行は異なる角度の結果であり，上から 0°， 45°， 90°， 180°である．また各列は， (a) 入力画像, (b) Mask R-CNN scale=500 の実行結果, (c) Mask R-CNN scale=1000 の実行結果, (d) Refine-Net 3scales の実行結果, (e) Refine-Net 5scales の実行結果である．

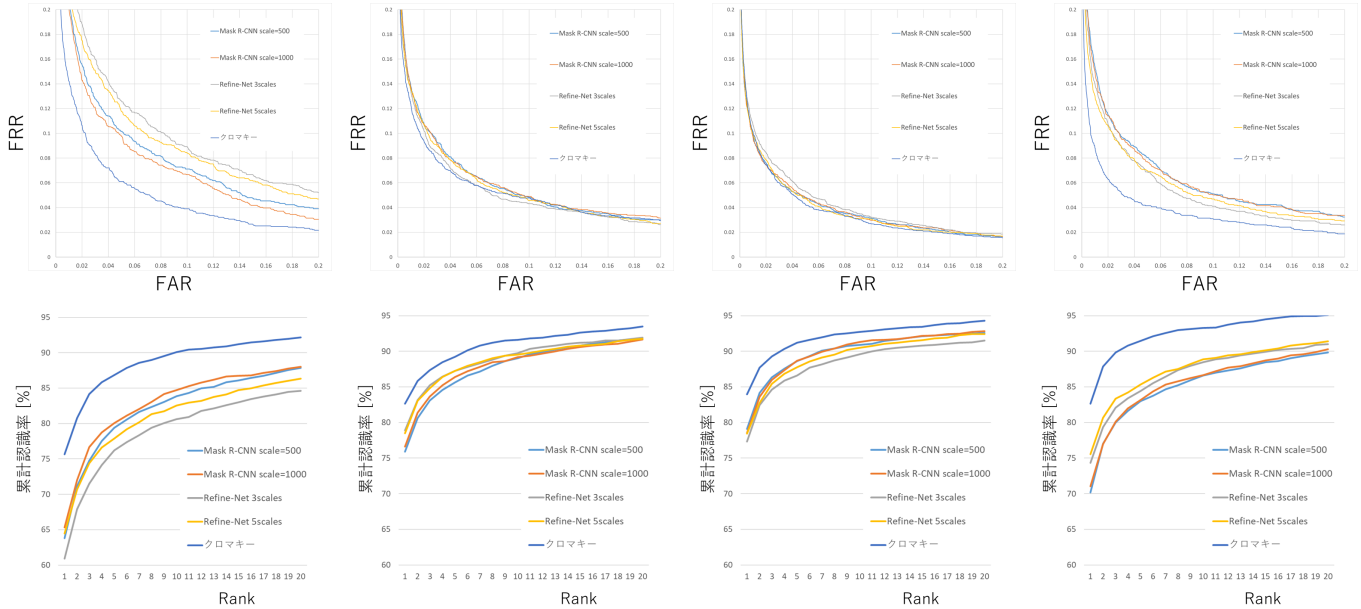
(a) (b) (c) (d) (e)

図 3: Refine-Net の失敗例の入力画像と各手法の出力結果．(a) 入力画像, (b) Mask R-CNN scale=500 の出力結果，(c) Mask R-CNN scale=1000 の実行結果, (d) Refine-Net 3scales の実行結果, (e) Refine-Net 5scales の実行結果 (Refine-Net の実行結果の中で薄い桃色は人物，濃い桃色は馬，紫は犬とラベル付けされている．)

5. 考察

5.1 シルエット抽出に関する考察

本研究で使用している Mask R-CNN は COCO DATASET を利用し，Refine-Net は 3.2 章で述べたように PASCAL VOC 2012 を利用して学習されている．これらのデータセットは一般的な物体検出を考えたデータセットであるためどちらも人物領域に特化していないが，Refine-Net のほうが人物領域に対して犬や牛などのラベリングを行ってしまう例が多くみられた．この人物への異なるラベル付けの多い理由はネットワークのアルゴリズム，学習に利用したデータセットが関係していると考えられるが，今後検証が必要である． 表 2 から，処理時間に大きな差があることがわかる．これは Refine-Net のネットワーク構造によるものと考えられる．Refine-Net は低解像度の情報をアップサンプリングした結果を次の実行で利用する．また最終的に，入力画像のそれぞれのサイズの出力結果を融合する必要がある．したがって，アップサンプリングの結果待ち，すべての大きさの入力画像に対する結果待ちの時間が必要となってくる．それに対して，Mask R-CNN のネットワーク構造は基本的な畳み込みニューラルネットワークによるものであるため，Refine-Net と比べて実行時間は短くなっている．また表 2 から，Mask R-CNN も Refine-Net もパラメータの数を減らすことで失敗系列数が多くはなっているが処理時間は減少していることや角度によって失敗系列数が大きく変化していることがわかる．特に 0°や 180°のデータにおいて失敗系列数が多いのは照明によって被験者が遮蔽されることが多かったことが原



(a) (b) (c) (d)

図 4: ROC 曲線と CMC 曲線．上段が ROC 曲線であり，下段が CMC 曲線である．各列は (a) 0°のデータセット，(b) 45°のデータセット，(c) 90°のデータセット，(d) 180°のデータセットをそれぞれ利用した場合の結果である．

因だと考えられる．また，Refine-Net は人物領域に対して人物以外のラベリングを行うことがあり，誤検出が Mask R-CNN に比べて多くなる結果となった．

5.2 歩容認証に関する考察

OU-MVLP は，グリーンバックの環境下で撮影されているため，クロマキーを使用することで，高精度のシルエットが抽出可能である．そこで，本論文では，このクロマキーにより抽出したシルエットに基づく歩容認証精度と比較を行うことで，シルエット精度について考察を行う．表 3 と図 4 の結果を見ると 90°においてはセマンティックセグメンテーションの精度とクロマキーの精度一対一認証ではそれほど大きな差となっていないが，正面や背面では認証精度に大きな差が確認できる．これより，セマンティックセグメンテーションのシルエット抽出精度には改善の余地があることがわかる．また角度によって，認証精度の高いセマンティックセグメンテーション手法が異なる，という結果となった．

6. おわりに

本研究では，セマンティックセグメンテーション手法を用いた歩容認証に注目し，セマンティックセグメンテーションの代表的な手法である Mask R-CNN と Refine-Net をシルエット抽出に用いた歩容認証においてそれぞれの手法ごとにおけるシルエット抽出の精度，またそのシルエット画像を利用した歩容認証の認証精度の比較を行った．防犯カメラから取得される映像に対してセマンティックセグメンテーションを行い歩容認証を行うためには，シルエット抽出の精度のさらなる改善が重要であると考えられる．また，セマンティックセグメンテーションを行うニューラルネットワークには，処理時間に注目したネットワークも存在する．そのようなネットワークを利用した場合の処理時間，歩容認証の認証精度の差についても確認していきたい．

実際に防犯カメラから取得した映像を利用することを考えたときに，今回利用したデータセットよりも低解像度の動画を利用することもあると思われる．そのため，解像度の変化によって歩容認証の認証精度にどの程度影響があるかも確認していきたいと考えている．また，今回使用したデータセットはグリーンバックという管理された環境下で取得されたデータであるため，今後は実環境で撮影されたデータを利用した実験も行っていきたいと考えている．

参考文献

[1] Bouchrika, I., Goffredo, M., Carter, J. and Nixon, M.: On Using Gait in Forensic Biometrics, *Journal of Forensic Sciences*, Vol. 56, No. 4, pp. 882–889 (online), DOI: 10.1111/j.1556-4029.2011.01793.x (2011).

[2] Nixon, M. S., Tan, T. N. and Chellappa, R.: *Human Identification Based on Gait (The Kluwer International Series on Biometrics)*, Springer-Verlag, Berlin, Heidelberg (2005).

[3] Iwama, H., Okumura, M., Makihara, Y. and Yagi, Y.: The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 5, pp. 1511–1521 (online), DOI: 10.1109/TIFS.2012.2204253 (2012).

[4] Iwama, H., Muramatsu, D., Makihara, Y. and Yagi, Y.: Gait Verification System for Criminal Investigation,

IPSJ Trans. on Computer Vision and Applications, Vol. 5, pp. 163–175 (2013).

[5] Bobick, A. F. and Johnson, A. Y.: Gait recognition using static, activity-specific parameters, *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1, pp. I–423–I–430 vol.1 (2001).

[6] Man, J. and Bhanu, B.: Individual recognition using gait energy image, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322 (online), DOI: 10.1109/TPAMI.2006.38 (2006).

[7] Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T. and Yagi, Y.: GEINet: View-invariant gait recognition using a convolutional neural network, *2016 International Conference on Biometrics (ICB)*, pp. 1–8 (online), DOI: 10.1109/ICB.2016.7550060 (2016).

[8] Shioji, R., Ito, S.-i., Ito, M. and Fukumi, M.: Personal Authentication Based on Wrist EMG Analysis by a Convolutional Neural Network, pp. 12–18 (online), DOI: 10.12792/icisip2017.006 (2017).

[9] Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T. and Yagi, Y.: On Input/Output Architectures for Convolutional Neural Network-Based Cross-View Gait Recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1 (online), DOI: 10.1109/TCSVT.2017.2760835 (2018).

[10] Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T. and Yagi, Y.: Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition, *IPSJ Trans. on Computer Vision and Applications*, Vol. 10, No. 4, pp. 1–14 (2018).

[11] Liu, Z. and Sarkar, S.: Simplest representation yet for gait recognition: averaged silhouette, *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 4, pp. 211–214 Vol.4 (online), DOI: 10.1109/ICPR.2004.1333741 (2004).

[12] Lee, S., Liu, Y. and Collins, R.: Shape Variation-Based Frieze Pattern for Robust Gait Recognition, *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (online), DOI: 10.1109/CVPR.2007.383138 (2007).

[13] Lam, T. H., Cheung, K. and Liu, J. N.: Gait flow image: A silhouette-based gait representation for human identification, *Pattern Recognition*, Vol. 44, No. 4, pp. 973 – 987 (online), DOI: https://doi.org/10.1016/j.patcog.2010.10.011 (2011).

[14] Sagawa, R., Makihara, Y., Echigo, T. and Yagi, Y.: Matching Gait Image Sequences in the Frequency Domain for Tracking People at a Distance, Vol. 3852, pp. 141–150 (online), DOI: 10.1007/11612704_15 (2006).

[15] Fu, K. and Mui, J.: A survey on image segmentation, *Pattern Recognition*, Vol. 13, No. 1, pp. 3 – 16 (online), DOI: https://doi.org/10.1016/0031-3203(81)90028-5 (1981).

[16] Wang, J., Makihara, Y. and Yagi, Y.: Human tracking and segmentation supported by silhouette-based gait recognition, *2008 IEEE International Conference on Robotics and Automation*, pp. 1698–1703 (online), DOI: 10.1109/ROBOT.2008.4543445 (2008).

[17] Makihara, Y. and Yagi, Y.: Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation, *2008 19th International Conference on Pattern Recognition*, pp. 1–4 (online), DOI: 10.1109/ICPR.2008.4761121 (2008).

[18] Ronneberger, O., P.Fischer and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, Vol. 9351, Springer, pp. 234–241 (2015).

[19] Badrinarayanan, V., Kendall, A. and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, *CoRR*, Vol. abs/1511.00561 (online), available from ⟨http://arxiv.org/abs/1511.00561⟩ (2015).

[20] Paszke, A., Chaurasia, A., Kim, S. and Culurciello, E.: ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation, *CoRR*, Vol. abs/1606.02147 (online), available from ⟨http://arxiv.org/abs/1606.02147⟩ (2016).

[21] Romera, E., lvarez, J. M., Bergasa, L. M. and Arroyo, R.: Efficient ConvNet for real-time semantic segmentation, *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1789–1794 (online), DOI: 10.1109/IVS.2017.7995966 (2017).

[22] Chen, X. and Xu, J.: Uncooperative Gait Recognition, *Pattern Recogn.*, Vol. 53, No. C, pp. 116–129 (online), DOI: 10.1016/j.patcog.2015.11.016 (2016).

[23] Wu, Z., Huang, Y. and Wang, L.: Learning Representative Deep Features for Image Set Analysis, *IEEE Transactions on Multimedia*, Vol. 17, No. 11, pp. 1960–1968 (online), DOI: 10.1109/TMM.2015.2477681 (2015).

[24] He, K., Gkioxari, G., Dollár, P. and Girshick, R. B.: Mask R-CNN, *CoRR*, Vol. abs/1703.06870 (online), available from ⟨http://arxiv.org/abs/1703.06870⟩ (2017).

[25] Lin, G., Milan, A., Shen, C. and Reid, I.: RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation, *CVPR* (2017).

[26] Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P. and He, K.: Detectron, https://github.com/facebookresearch/detectron (2018).

[27] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems 25* (Pereira, F., Burges, C. J. C., Bottou, L. and Weinberger, K. Q., eds.), Curran Associates, Inc., pp. 1097–1105 (online), available from ⟨http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf⟩ (2012).

[28] Nair, V. and Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines, *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, USA, Omnipress, pp. 807–814 (online), available from ⟨http://dl.acm.org/citation.cfm?id=3104322.3104425⟩ (2010).

[29] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, Vol. 15, pp. 1929–1958 (online), available from ⟨http://jmlr.org/papers/v15/srivastava14a.html⟩ (2014).

[30] Iwama, H., Okumura, M., Makihara, Y. and Yagi, Y.: The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 5, pp. 1511–1521 (online), DOI: 10.1109/TIFS.2012.2204253 (2012).