

令和4年度卒業論文

FaceNetを用いた牛の個体識別

宮崎大学工学部情報システム工学科

時任 大輝

指導教員 椋木 雅之 教授

概要

本論文では、深層学習による顔認識技術としてFaceNetを用い、FaceNetの牛への適用可能性について調査する。FaceNetとは、Googleが2015年に開発した顔認識手法である。顔画像から特徴量を抽出し、その特徴量をもとに顔の識別を行う。FaceNetは人間の顔認識タスクにおいて高い精度を実現している。また、人間の顔認識手法を動物に適用した研究もある。例えばBearIDと呼ばれる、FaceNetを用いたヒグマの顔認識手法では、ヒグマの個体識別において高い精度を実現している。しかし、人間やヒグマと比較して牛の顔は角度による形状変化が大きく、形状の性質が異なるため、人間の顔認識手法を牛の個体識別に適用した研究事例は少ない。

本研究では、牛の顔画像のデータセットを用意し、FaceNetによる個体識別のモデルを構築し、識別率の評価実験を行う。その結果、異なる牛を正しく異なると判定することはできたが、同じ牛を同じと判定する性能はやや低かった。また、先行研究である高宗の研究と同様の評価方法と比較した結果、本研究で作成したモデルは人の顔を学習したFaceNetのモデルと比較して高い精度が得られた。

目次

1. はじめに	1
2. 牛の個体識別	3
2.1 牛個体識別の必要性	3
2.2 牛個体識別の従来手法	3
2.3 顔画像による個人識別	4
3. FaceNetを用いた牛画像からの個体識別	6
3.1 牛顔画像データセットの準備	6
3.2 個体識別のモデル構築	10
4. 実験	12
4.1 実験1 モデル評価(正解率・適合率・再現率・特異度)	12
4.1.1 実験データ	12
4.1.2 評価方法	13
4.1.3 閾値の決定	14
4.1.4 各モデルの評価結果	16
4.1.5 考察	19
4.2 実験2 モデル評価(CMC曲線)	20
4.2.1 CMC曲線による評価	20
4.2.2 考察	21
5. 結論	22
謝辞	23
参考文献	24

1. はじめに

近年、農業においても自動化技術が普及してきている。その中で畜舎内の個体識別は重要な課題の1つである。個体識別は牛の健康管理や繁殖管理など様々な場面で行われている。従来は耳標やRFIDタグなどが用いられることが多かった。しかし耳標は汚れや牛の動きで読み取りにくかったり、破損・欠落で利用できない場合がある。また、RFIDタグでは、タグの埋め込みで牛にストレスを与える上、タグ自体や読み取り装置のコストが問題となる。

このような問題を解決するために、画像認識技術が用いられることが増えてきている。近年、深層学習による画像認識技術は高い精度を実現している。画像認識技術のタスクの中でも人間の顔認識は重要な課題の1つである。そのため顔認識に関する研究は多数展開されている。

人間の顔認識手法の1つにFaceNet[1]がある。FaceNetとはGoogleが2015年に開発した顔認識手法である。顔画像から特徴量を抽出し、その特徴量をもとに顔の識別を行う。FaceNetは人間の顔認識において高い性能を示している。

人間の顔認識手法を動物に適用した研究もある。例えばBearIDと呼ばれる、FaceNetを用いたヒグマの顔認識手法[2]では、ヒグマの個体識別において高い精度を実現している。しかし、人間やヒグマと比較して牛の顔は角度による形状変化が大きく、形状の性質が異なるため、人間の顔認識手法を牛の個体識別に適用した研究事例は少ない。

そこで、本研究では深層学習による顔認識技術としてFaceNetを用い、FaceNetの牛への適用可能性について調査する。FaceNetは人間の顔認識タスクにおいて高い精度を実現しており、牛の顔認識にも適用可能であると考えられる。

本研究の目的は、FaceNetの牛顔認識への適用可能性の調査である。牛の顔画像のデータセットを用意し、個体識別のモデルを構築し、識別率の評価実験を行う。

本論文の構成は次のとおりである。第2章では、牛の個体識別の必要性と従来手法について述べると共に、FaceNetの牛顔認識への適用可能性について考察する。第3章では、FaceNetを用いた牛の個体識別手法について具体的に述べる。4章では、生成したFaceNetモデルの性能を評価する。最後に第5章では、結論と今後の課題を述べる。

2. 牛の個体識別

2.1 牛個体識別の必要性

牛の個体識別は農業において重要な課題の一つであり、牛の健康管理や繁殖管理などに欠かせないものである。牛の一生を通じて健康状態や生産性を正確に追跡することができるとともに、繁殖計画の策定や管理のための基礎データを提供することができる。このような情報を利用することで農業経営者は最適な繁殖管理策を採用でき、生産性の向上やコストの削減に寄与することができると考えられる。

また、牛トレーサビリティ制度により、牛管理者に対して国内で飼われている牛の個体識別情報の届け出が義務化されている。牛トレーサビリティ制度とは、牛や牛肉に関する問題が発生した際に、それらの追跡を可能にし、国産牛肉の安全・安心を確保することを目的に制定された制度である。届け出が必要な個体識別情報には10桁の個体識別番号、性別や種別(例:黒毛和種)、出生から飼養地までの情報などが含まれる。そのため、牛管理者には正確な牛の個体管理が求められる。

2.2 牛個体識別の従来手法

牛個体識別の従来手法には耳標の読み取り、RFIDタグの埋め込み、画像認識による手法などがある。

耳標から個体識別情報を読み取る手法では1次元バーコード、QRコード、RFIDタグが使われている。1次元バーコードとは、幅と長さ限定された長い一列の線とスペースからなる識別システムである。これらの線とスペースをスキャンすることにより、情報を読み取る事ができる。QRコードとは、二次元グリッド上に描かれた

複雑な模様からなる識別システムである。QRコードはカメラを使用して読み取ることができる。RFIDタグとは、電波によってマイクロチップに記述された電子データを非接触で読み書きする手法である。1次元バーコードやQRコードではタグを1枚ずつ読み取る必要があるのに対して、RFIDタグは電波によってタグを複数同時に読み取ることが可能である。また、RFIDタグのマイクロチップを牛の皮下に直接埋め込むという手法もある。

高宗による研究[3]では、様々な特徴抽出器などを用いて牛顔画像から個体識別を行い、特徴抽出器の有効性を評価する実験を行っている。実験では人の顔を学習したモデルであるFaceNetとVGGFace、一般物体認識用のデータセットで学習したVGG16、VGG19、ResNet50をそれぞれ用いて牛顔画像から特徴抽出を行い、精度を比較した。結果として、識別の精度が最も高いVGG16では、27頭の肉牛に対して92.8%、24頭の乳牛に対して97.5%の高い精度を示し、牛顔画像を用いた個体識別の可能性が示された。

2.3 顔画像による個人識別

画像認識による手法の1つに、FaceNetがある。FaceNetとはGoogleが2015年に開発した顔認識手法である。FaceNetは顔画像から特徴量を抽出し、その特徴量をもとに顔の識別を行う。

FaceNetは3つのステージから構成されている。1つ目のステージでは、顔画像から特徴量を抽出するための畳み込み層(Convolutional Neural Network; CNN)を用いる。2つ目のステージでは、特徴量を正規化し、顔の識別に適した特徴量に変換するための全結合層を用いる。3つ目のステージでは、特徴量をもとに顔の識別を行うためのTriplet loss層を用いる。

Triplet loss層では、類似した画像間の距離を近づけ、類似していない画像間の距離を離すことを目的とする。Triplet lossはアンカー画像、正解画像、不正解画像の3つの画像から構成される。アンカー画像は、正解画像と同じクラスに属する画像であり、不正解画像は、正解画像とは異なるクラスに属する画像である。学習の際には、アンカー画像から正解画像への距離がアンカー画像から不正解画像への距離よりも小さくなるように学習を行う。これによりアンカー画像と正解画像の類似度が高く、アンカー画像と不正解画像の類似度が低くなる。Triplet lossは距離関数としてユークリッド距離を用いることが一般的である。これにより類似した画像間の距離が小さく、類似していない画像間の距離が大きくなる。このように、FaceNetは距離の算出方法を学習しているとみなせる。識別時には、既知の牛の顔画像を最低一枚用意し、未知の牛の顔画像との距離を求め、距離があらかじめ定めた閾値以内であれば同じ牛個体の牛顔画像と判定する。これにより、一度FaceNetの学習を行えば、識別したい各牛個体について、顔画像を最低一枚用意すれば、識別が行える。

本研究では深層学習による顔認識技術としてFaceNetを用い、FaceNetの牛への適用可能性について調査する。高宗は人で学習済みのFaceNetを使って評価を行っていた。しかし、人と牛の顔の構造は大きく異なるため、正当な評価とは言えない。本論文では牛の顔画像でFaceNetの学習を行った上で、牛の顔認識に適用可能かを調査する。

3. FaceNetを用いた牛画像からの個体識別

3.1 牛顔画像データセットの準備

牛顔画像のデータセットを用いて学習と評価を行う。データセットは畜産農場から収集した牛顔画像から構成される。

データセットの準備にあたっては、以下のステップを踏む。

1. 畜産農場で牛顔を含む動画を収集する。
2. 牛顔検出によって動画から牛顔画像を抽出する。
3. 収集した画像を人手でラベル付けする。
4. 牛顔画像を学習用データと評価用データに分割する。

まず、畜産農場で牛顔を含む動画を収集する。動画は、牛が牛舎内で隔柵から顔を出して餌槽内の餌を食べている状態で撮影した(図1)。また、撮影時には個体識別の正解が分かるよう、最初に耳標を撮影した後、上下左右にカメラを動かしながら撮影した。牛自体も比較的自由に動くため、様々な角度から牛の顔を撮影できている。一方で、暗い牛舎内での撮影のため、一部にはボケやブレが見られた。動画内では主に1頭の牛を撮影しているが、左右の牛の顔も映り込んでいる。

次に、牛顔検出によって動画から牛顔画像を抽出する。牛顔検出には、兒玉による研究[4]で用いられた、牛顔検出のための事前学習済みYOLOv5[5]のモデルを使用する。YOLO[6]は高速かつ正確な物体検出アルゴリズムの一つである。YOLOv5はフレームワークとしてPython向けオープンソース深層学習ライブラリであるPyTorchを使っているYOLOの改良版である。兒玉の作成したモデルは、「牛全体」と「牛顔」の2クラスでラベル付けした5921枚の画像を含むデー

タセットと事前学習済みモデルを用いて作成されている。本研究では、このモデルを用いて、収集した各動画に対して牛顔検出を実行し、「牛顔」の信頼度が0.9以上のフレームのみを牛顔画像として抽出する。

さらに、得られた牛顔画像を人手でラベル付けする。ラベル付けは、図2に示すようなディレクトリ構造に従う。datasetsディレクトリ下には、対応する動画ファイル名がラベルとして付与されたディレクトリがある。これらのディレクトリ下には各ラベルに対応する牛顔画像が格納されている。各牛顔画像は“(ラベル名)_(4桁の連番).jpg”のような命名規則に従う。4桁の連番は牛顔検出順に付与されている。この時、ディレクトリ内に異なる牛顔画像が混入していないかどうかを目視で確認する。もしディレクトリ内に異なる牛顔画像が混入していた場合は、その画像を削除する。



図1 撮影した動画の例

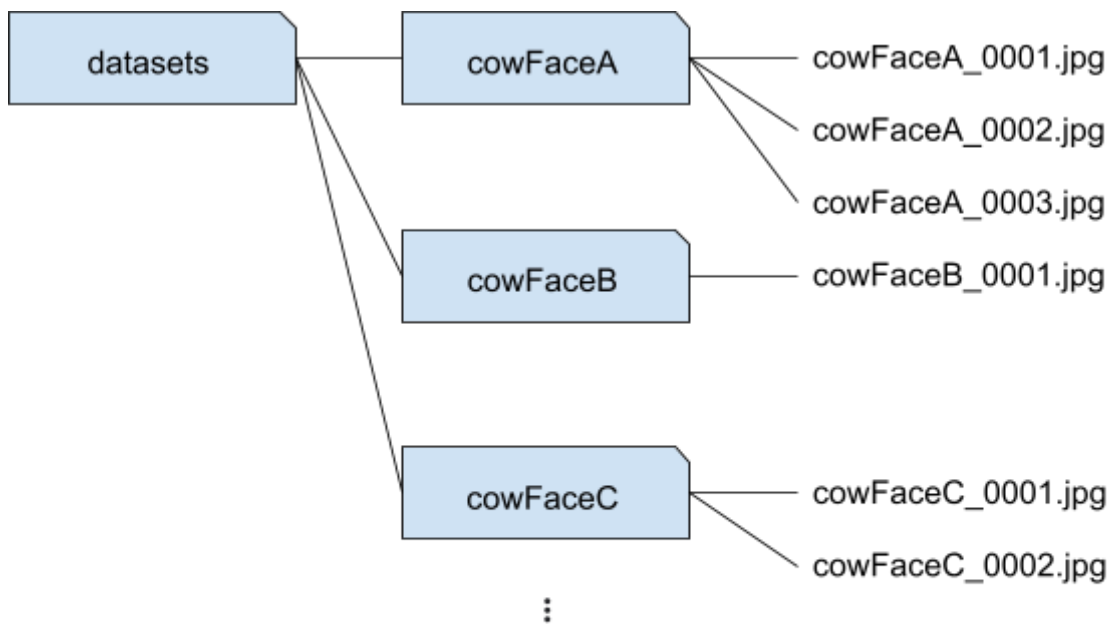


図2 データセットのディレクトリ構造

以上の手順によって牛顔画像データセットを準備する。このデータセットは、図3に示すような594頭、57097枚の画像から構成される。このデータセットの内訳を図4のヒストグラムに示す。このヒストグラムは横軸が1クラスあたりの画像枚数、縦軸が各画像枚数におけるクラスの出現頻度を表す。1クラスあたりの画像枚数は平均112枚、最小2枚、最大523枚である。このデータセットを学習用データと評価用データに分割する。約9割にあたる531頭を学習用データとして、残りの63頭を評価用データとする。

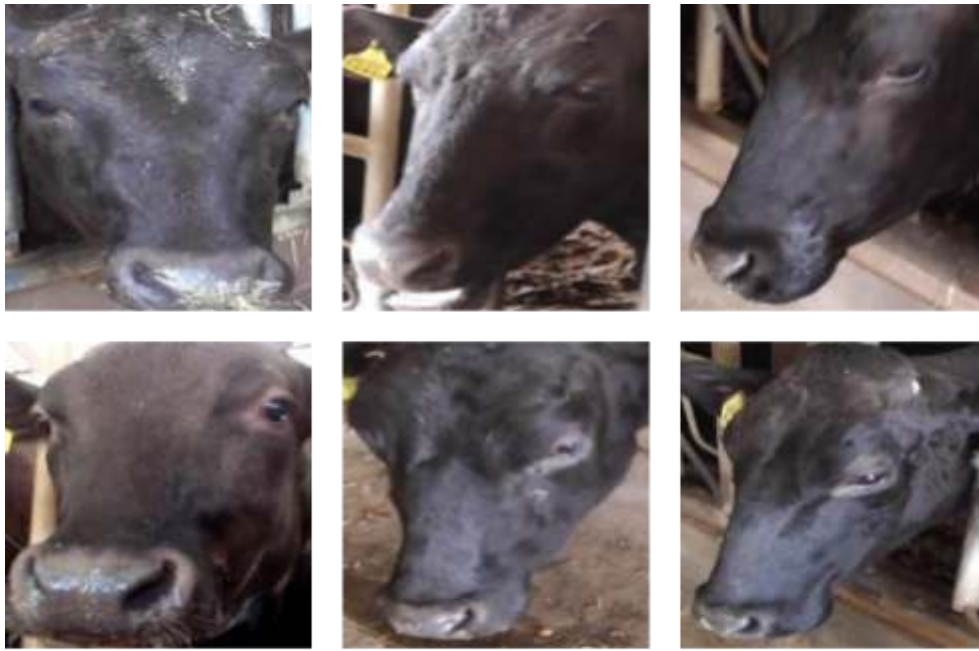


図3 牛顔画像データセットの画像例

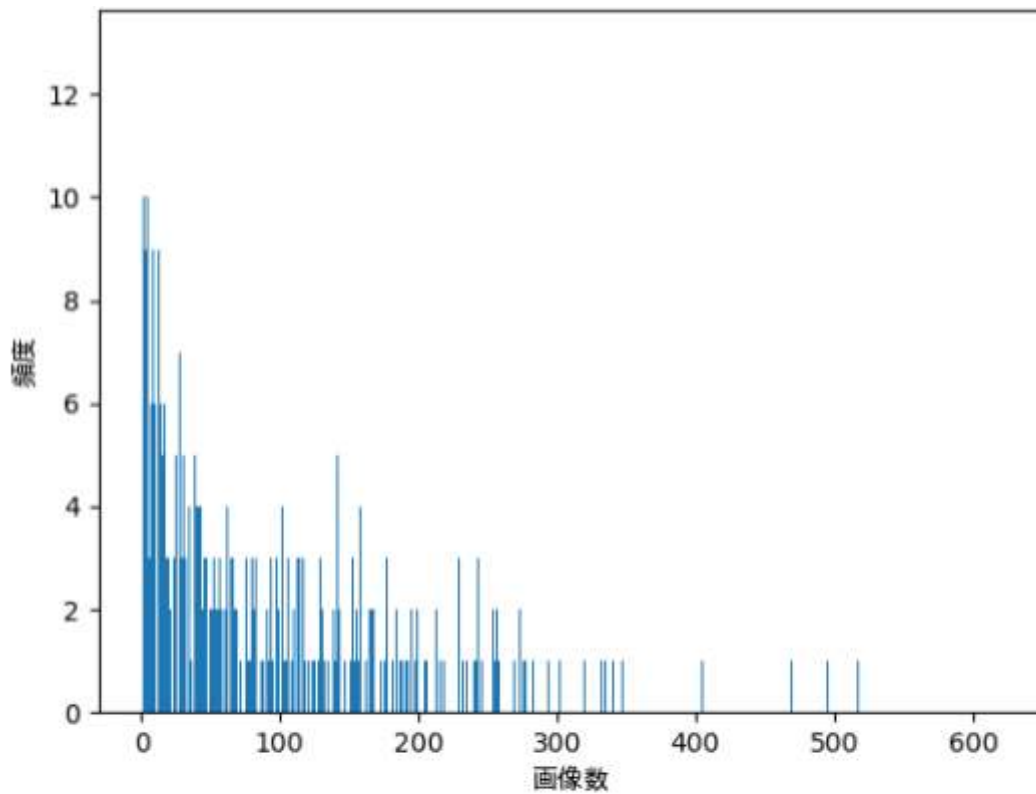


図4 牛顔画像データセットの個体毎の画像枚数分布

3.2 個体識別のモデル構築

TensorFlow版のFaceNetの実装[7]を用いて牛の個体識別のモデルを構築する。学習用データは531頭、50856枚の牛顔画像で構成される。学習用データの例を図5に示す。その後、学習によって得られたモデルを用いて、評価用データの牛顔画像の特徴量を抽出し、その特徴量を元に類似度を計算することで、個体識別を行う。特徴量は128次元を出力する。特徴量間の類似度は、ユークリッド距離によって計算する。



図5 学習用データの例

本研究では、学習用データの各種パラメータの異なる3つのモデルを構築する。各モデルの仕様は表1の通りである。

表1 学習用データのパラメータ

	画像サイズ	背景情報
model1	160 × 160画素	あり
model2	90 × 90画素	あり
model3	90 × 90画素	なし(center crop)

これら学習用データの異なる3つのモデルを構築することで、学習用データのパラメータが牛顔の識別率に与える影響を検証する。model1は図5の顔画像を160×160画素にリサイズして学習を行う。図5に示したように、画像には牛の顔だけでなく、背景も映り込んでいる。model2は、入力画像の大きさを90×90画素に縮めて学習を行う。解像度が学習に与える影響を評価するために用いる。model3は図5の画像を160×160画素にリサイズした上で、中央の90×90画素を切り出すことで背景を除去して、学習を行う。背景の影響を評価するために用いる。次節で、評価実験について述べる。

各モデルを約500エポック、約3日間かけて学習する。

4. 実験

構築した個体識別モデルを用いて、牛顔画像データセットの評価用データを用いた精度評価を行う。

4.1 実験1 モデル評価(正解率・適合率・再現率・特異度)

4.1.1 実験データ

評価用データについては学習に使用しなかった、63頭、6241枚の牛顔画像を使用する。評価用データの例を図6に示す。これらの画像は学習用データとは異なる牛顔画像であり、モデルが未知のデータに対してどの程度の性能を発揮するかを評価するために使用される。また、モデルの評価の際には、モデルに合わせて牛顔画像をリサイズをする。



図6 評価用データの例

4.1.2 評価方法

本研究ではモデルの性能を評価するために、正解率、適合率、再現率、特異度、という4つの評価指標を用いる。各指標の定義は以下の通りである。

TruePositive(TP) = (本当は同じ牛顔に対して同じ牛顔と予測したケース)

TrueNegative(TN) = (本当は異なる牛顔に対して同じ牛顔と予測したケース)

FalsePositive(FP) = (本当は異なる牛顔に対して同じ牛顔と予測したケース)

FalseNegative(FN) = (本当は同じ牛顔に対して異なる牛顔と予測したケース)

$$\text{(正解率)} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{(適合率)} = \frac{TP}{TP + FP}$$

$$\text{(再現率)} = \frac{TP}{TP + FN}$$

$$\text{(特異度)} = \frac{TN}{TN + FP}$$

正解率はモデルが牛顔を正しく認識する割合を示す。正解率が高いほど、モデルは牛顔を正確に認識することができる。

適合率はモデルが同じ牛顔として予測する画像のうち、実際に同じ牛顔である割合を示す。適合率が高いほど、モデルは間違った予測をすることが少ない。

再現率は実際に同じ牛顔である画像のうち、同じ牛顔として予測する割合を示す。再現率が高いほど、モデルは実際に同じ牛顔である画像を異なる牛顔として見逃してしまうことが少ない。

特異度は実際に異なる牛顔である画像のうち、正しく異なる牛顔だと予測できた割合を示す。特異度が高いほど、モデルは実際に異なる牛顔を同じ牛顔として予測することが少なく、同じ牛顔以外の画像を正確に認識することができる。

これらの評価指標を組み合わせることで、顔認識モデルの全体的な性能を評価する。

4.1.3 閾値の決定

実験では二枚の牛顔画像をそれぞれ学習済みのFaceNetに入力して特徴量を算出し、それらの間のユークリッド距離が与えられた閾値以下であれば同じ牛個体、閾値より大きければ異なる牛個体として判別する。閾値は、学習データに対するF1スコアが最大となる値を実験により定める。F1スコアの定義は以下の通りである。

$$(F1スコア) = \frac{2 \times (\text{適合率}) \times (\text{再現率})}{(\text{適合率}) + (\text{再現率})}$$

F1スコアは、適合率と再現率の調和平均で、適合率と再現率のバランスを考慮した評価指標である。F1スコアが最大となる値を閾値として設定することで、評価実験を行う際に、適合率と再現率のバランスを取りながら、個体識別の精度を最大化する。

学習データの任意の2枚の画像間で閾値を変えながら個体識別を行った際のF1スコアの変化を図7に示す。閾値を0.0から1.0まで0.01ごとに変えた結果をプロットしている。それぞれのモデルに対して、図中に示した最適な閾値を採用した。

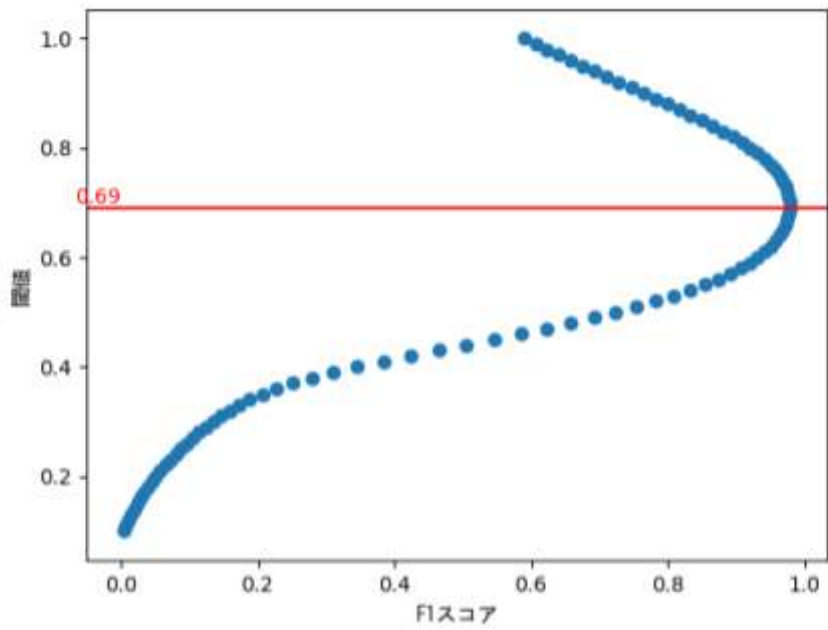


図7(a) F1スコアによる閾値の決定(model1)

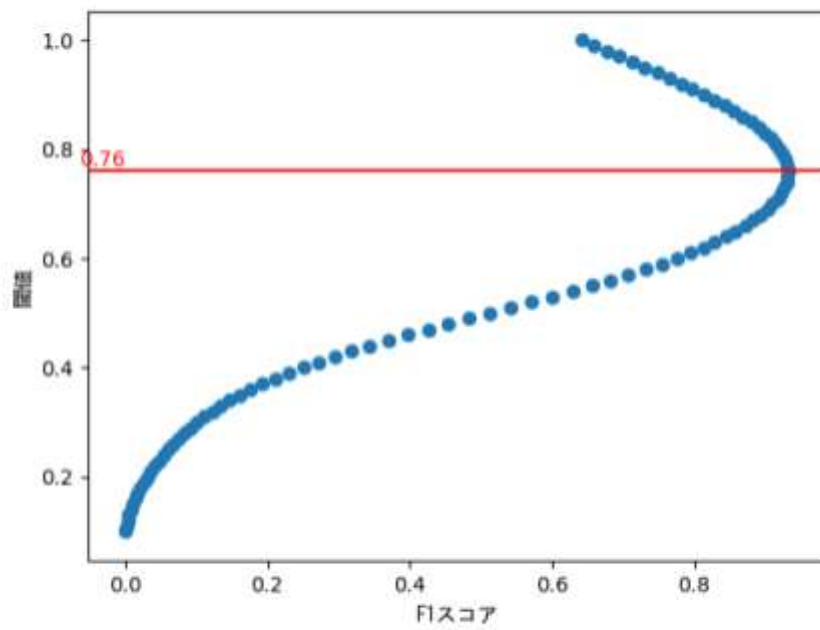


図7(b) F1スコアによる閾値の決定(model2)

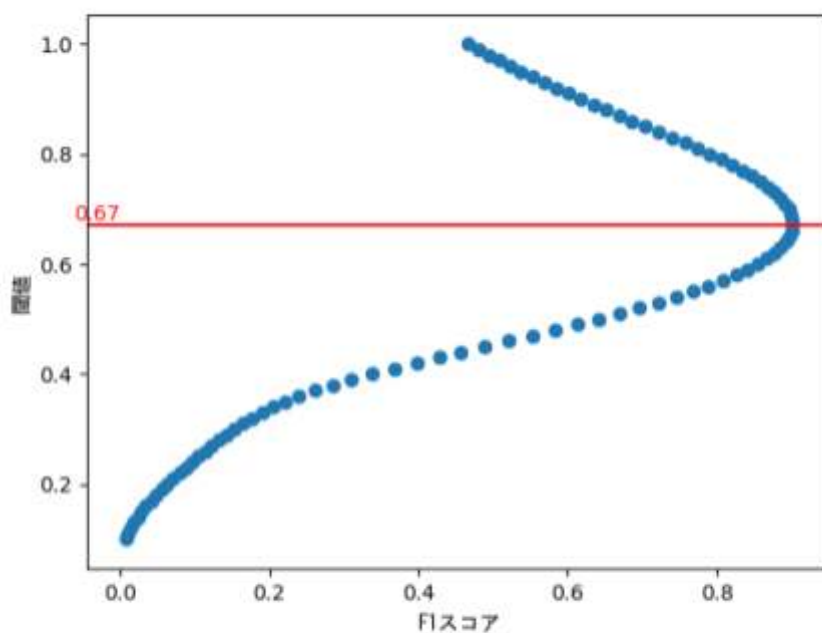


図7(c) F1スコアによる閾値の決定(model3)

4.1.4 各モデルの評価結果

先に述べた各評価指標を使って、各モデルの性能を比較する。評価用データの6241枚の画像から全ての二枚の組み合わせを取り出し、各モデルに与えて特徴量を算出する。得られた特徴量間のユークリッド距離が前節で設定した閾値以下であれば同じ牛、閾値より大きければ異なる牛顔として判定する。この判定結果について、4.1.2節で述べた4つの評価指標で評価する。各モデルの評価結果を表2に示す。また、各モデルの識別結果を図8に示す。

表2 各モデルの評価結果

	model1	model2	model3
正解率(%)	94.8	94.6	93.6
適合率(%)	87.6	87.9	72.3
再現率(%)	42.7	38.7	34.6
特異度(%)	99.4	99.5	98.8

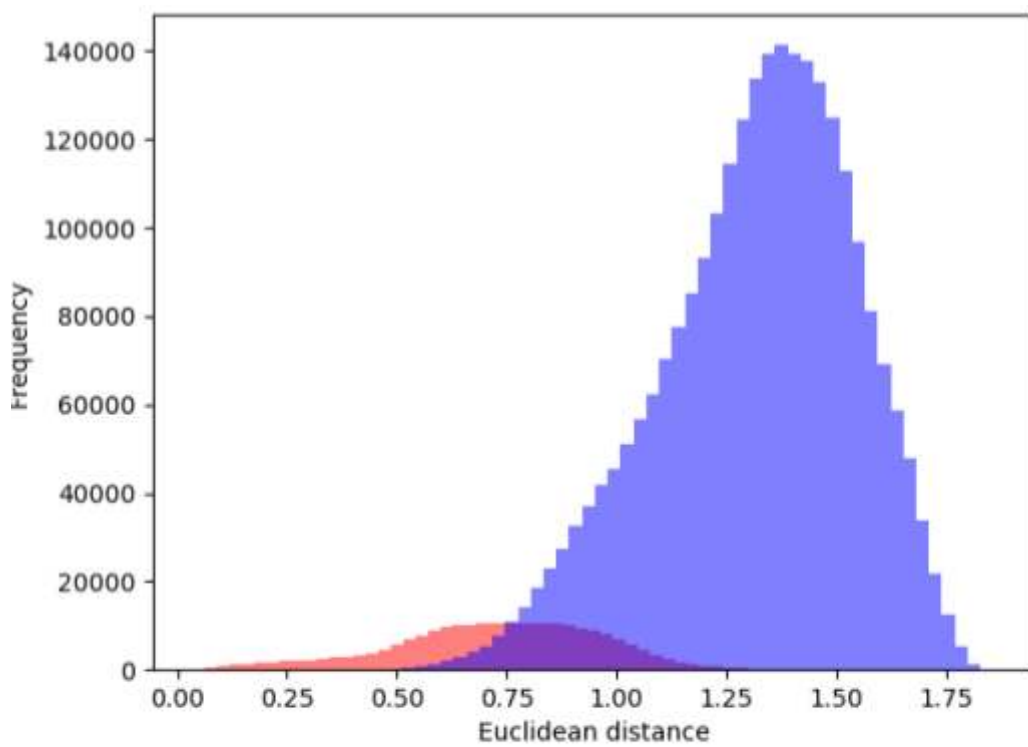


図8(a) 識別結果(model1)

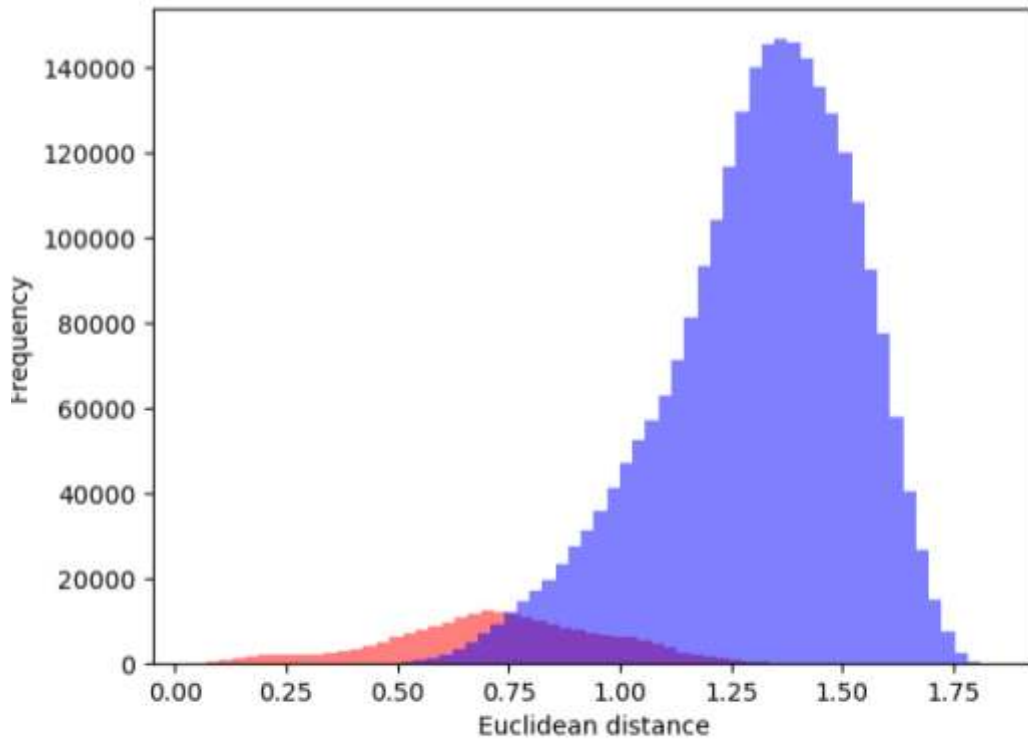


図8(b) 識別結果(model2)

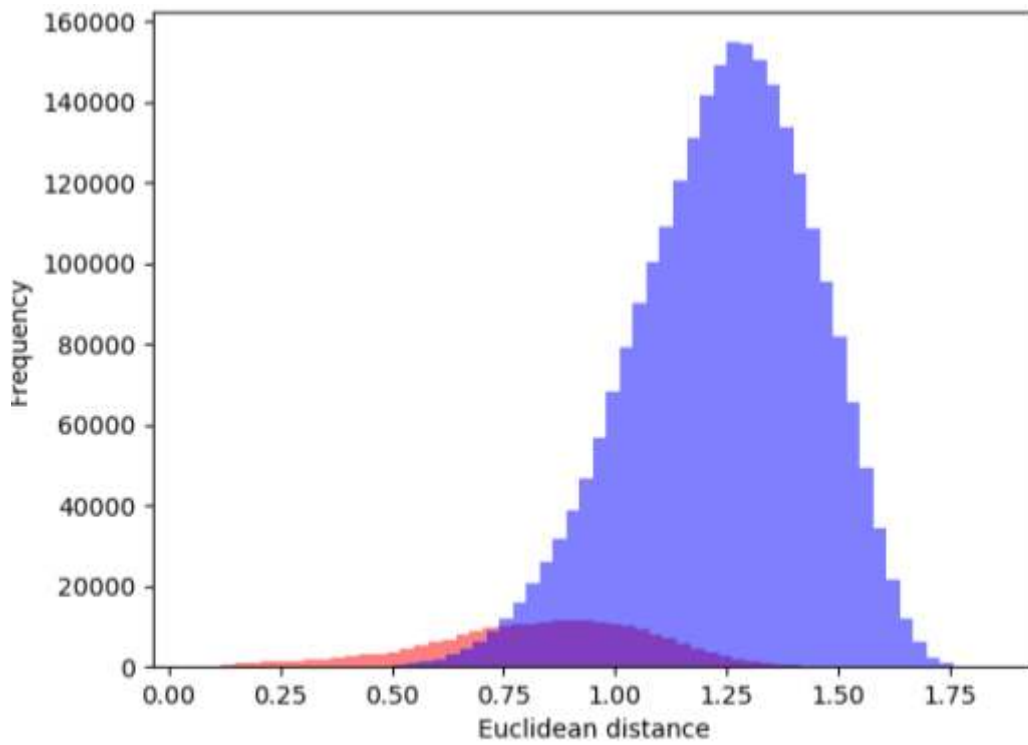


図8(c) 識別結果(model3)

4.1.5 考察

まず、表2の評価結果より、各モデルの性能を評価する。

全体の傾向としては、正解率と特異度はいずれも90%以上と高い。一方、再現率は低く、同じ牛を異なる牛として見逃してしまう場合がある。適合率もやや低く、異なる牛を同じ牛と誤判定している場合がある。

model1は160×160pxの画像を学習用データとして使用している。正解率は94.8%だが、適合率は87.6%と低い。再現率も42.7%と、低い傾向にある。特異度は99.4%と高い。

model2は90×90pxの画像を学習用データとして使用している。正解率は94.6%だが、適合率は87.9%と低い。再現率も38.7%と、低い傾向にある。特異度は99.5%と高い。

model3は90×90pxで中央部分のみを切り取っている画像を学習用データとして使用している。正解率は93.6%だが、適合率は72.3%とかなり低い。再現率も34.6%と、より低い傾向にある。特異度は98.8%とやや低い。

全体的にmodel1とmodel 2の結果は似ているが、model1の方が再現率は高い。model2よりmodel1の方が再現率が高い理由としては、大きな画像サイズを使用するとより多くの特徴が捕捉され、正確性が向上するからであると考えられる。また、model3は各指標ともに低い傾向にあり、背景の写り込みの影響はあまりないと考えられる。

次に、図8の各モデルの識別結果について考察する。

このヒストグラムでは、同じ牛顔間のユークリッド距離の分布(赤色)と異なる牛顔間のユークリッド距離の分布(青色)を比較している。2つの分布が乖離しているほど、モデルが未知のデータに対して正しい識別をしていることを示す。結果は、

各モデルとも両分布の重なりが大きく、正しい識別ができていない場合があることが確認される。

4.2 実験2 モデル評価(CMC曲線)

4.2.1 CMC曲線による評価

高宗の研究[3]では、CMC曲線を使用した評価が行われていた。CMC曲線は、特徴量間の照合によって得られた類似度をもとに順位付けを行い、正解クラスが各順位までに含まれる割合をグラフに表したものである。本研究での牛の顔画像で学習したFaceNetを高宗の研究と比較するために、同様の設定で実験を行った。また、この実験では高宗による研究[3]で用いられた人の顔画像で学習したFaceNet(FaceNet + SVM)との比較も行う。

実験1と同様に、評価用データについては学習に使用しなかった、63頭、6241枚の牛顔画像を使用する。

評価は、5分割交差検証により行う。各牛の顔画像を5分割し、4つを既知の牛顔画像、残り1つを未知の牛顔画像とする。未知の牛の顔画像それぞれについて、全ての既知の牛の顔画像と特徴量のユークリッド距離を求める。距離が近いものから順にRank1からRank4とし、正解率を求める。これを未知の牛の顔画像とする分割を5通りに入れ替えて行う。その識別精度の平均値をCMC曲線で示す。各モデルのCMC曲線を図9に示す。

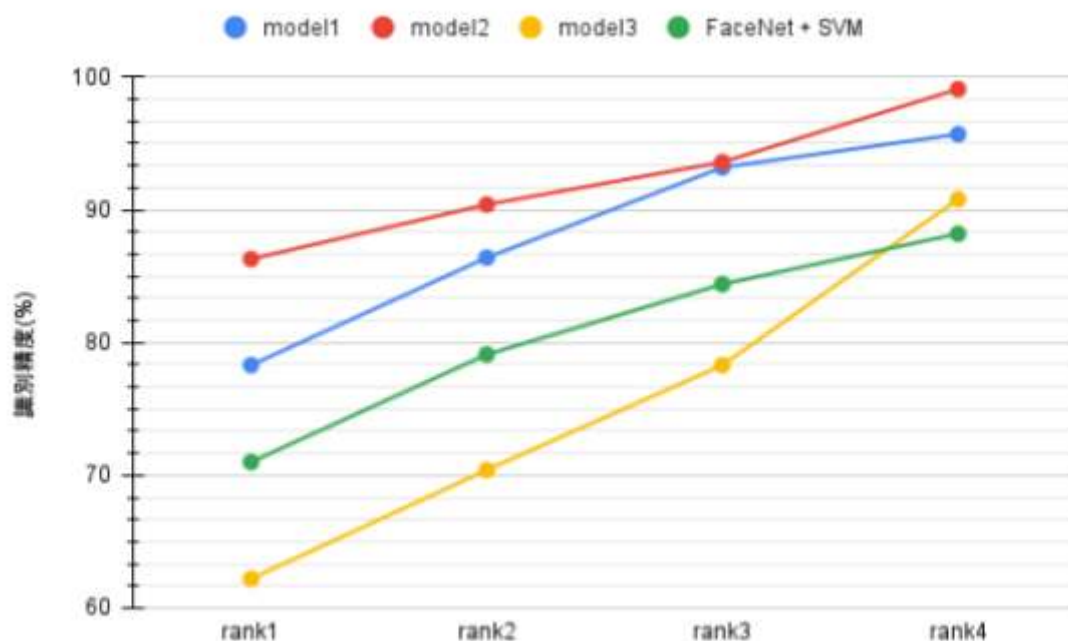


図9 CMC曲線による評価結果

4.2.2 考察

全体的にmodel2が最も高い精度を示している。model1とFaceNet+SVMは比較的高い精度を示しているが、model3は精度が低い。一方、model1とmodel2はFaceNet+SVMと比較して高い精度を示している。高宗の研究では、FaceNet + SVMは肉牛27頭に対してrank1で60%程度の識別精度を示していたが、本研究で用いた評価用データでは70%を超える識別精度を示している。これは、高宗の研究では既知の牛1頭あたりの画像枚数が6枚と少量であったことが影響しているとみられる。同じ高宗の研究では、同条件でVGG16という学習済みの深層学習ネットワークを用いると90%以上の識別精度を示している。本研究の条件下で、VGG16を使った手法と本研究で行った牛顔画像を学習したFaceNetによる個体識別との比較を行うことは、今後の課題である。

5. 結論

本研究では深層学習による顔認識技術としてFaceNetを用い、FaceNetの牛への適用可能性について調査した。牛の顔画像のデータセットを用意し、個体識別のモデルを構築し、識別性能の評価実験を行った。その結果、異なる牛を正しく異なると判定することはできたが、同じ牛を同じと判定する性能はやや低かった。また、先行研究である高宗の研究と同様の評価方法で比較した結果、本研究で作成したモデルは人の顔を学習したFaceNetのモデルと比較して高い精度が得られた。

今後の課題としては、まず、本研究の条件下で、VGG16を使った手法と本研究で行った牛顔画像を学習したFaceNetによる個体識別を比較することが挙げられる。また、評価に同一の動画内の画像を使っていたため、違う環境下で撮影した画像を使った評価が必要であると考えられる。

謝辞

本研究を行うにあたり、多大な助言と支援をいただいた椋木雅之教授に感謝いたします。そして、研究の相談や議論を通して、多くの知識や示唆をいただいた椋木研究室の皆様にお礼申し上げます。

参考文献

- [1] Schroff Florian, et al. “FaceNet: A Unified Embedding for Face Recognition and Clustering” *Computer Vision and Pattern Recognition*, 2015, pp 815-823.
- [2] Melanie Clapham, et al. “Automated facial recognition for wildlife that lack unique markings: A deep learning approach for brown bears” *Ecology and Evolution*, 10, 2020, pp. 12883-12892.
- [3] 高宗 伸幸, “牛の正面顔画像による個体識別における転移学習の有効性評価”, 宮崎大学工学部情報システム工学科 卒業論文, 2020
- [4] 兒玉 光平, “CowFindAR: 牛顔個体識別を用いたモバイル端末向け管理情報提示システム”, 宮崎大学大学院工学研究科 修士論文, 2021
- [5] Glenn Jocher, “YOLOv5”, <https://github.com/ultralytics/yolov5>
- [6] Joseph Redmon, et al. “You Only Look Once: Unified, Real-Time Object Detection”, arXiv:1506.02640, 2015
- [7] David Sandberg, “facenet”, <https://github.com/davidsandberg/facenet>