令和 6 年度　修士論文

深層学習による複数種類の動物に対する顔個体識別の特性比較

宮崎大学大学院　工学研究科　工学専攻　機械・情報系コース
T2303394
長友　祐磨
指導教員　椋木　雅之教授

令和 7 年 1 月 23 日

概要

本論文では、複数種類の動物のデータセットを用いて、それらの様々な組み合わせで、転移学習や Joint Training を行い、顔個体識別の有効性を検証した。対象としては、人、チンパンジー、牛、ヤギ、馬の 5 種類を扱った。顔個体識別の学習モデルには人の顔個体識別の代表的な手法である ArcFace を利用した。転移学習、Joint Training を様々な組み合わせで行い、識別精度を比較した。また、特徴活性化マッピングの 1 つである Grad-CAM を用いて、学習モデルが顔画像のどこに注目しているかを可視化することで、転移学習や Joint Training による学習モデルの変化を比較した。

実験結果から、転移学習に関しては、チンパンジー、ヤギ、馬の小規模データセットでは、大規模データセットである人や牛からの転移学習による識別精度の向上が見られた。Grad-CAM によるヒートマップを比較しても、大規模データセットからの転移学習により、顔を捉えられるようになり、注目する場所が安定した。また、似た顔同士の動物での転移学習が有効な傾向にあると分かった。

Joint Training に関しては、5 種類の動物を合わせて学習した場合は、チンパンジーを除く 4 種類の動物で、テストデータに対する識別精度が低下した。データセットの規模にばらつきがあったため、すべての動物での Joint Training は有効な手法ではなかった。

ヤギと馬、人とチンパンジーの似た顔の小規模データセット同士の Joint Training については、それぞれ識別精度が向上した。Grad-CAM による特性比較でも、安定して顔やその部位を捉えている組み合わせもあり、識別精度には表れない性能の向上を確認した。

目次

第 1 章

はじめに

顔個体識別とは、入力された顔画像が特定の個体と一致するかどうか判別したり、その顔がどの個体であるかを特定したりする技術である。人の顔個体識別の精度は非常に高く、スマートフォン端末のユーザ認証などに用いられている。顔個体識別の手法では、深層学習モデルが主流になっていて、Triplet loss を用いた FaceNet[1] や、Softmax loss[2] をベースにした ArcFace[3] や CosFace[4] などの様々な手法が提案されている。また、新しい損失関数の提案以外にも、深層学習モデルのパフォーマンスを向上させるための手法として、ある分野で学習されたモデルを別の分野の学習に適用する転移学習 [5] や、複数のタスクやモデルを同時に学習する Joint Training[6] なども提案されている。

人の顔個体識別の性能向上を受けて、人以外の動物に関しても、顔個体識別の活用が期待されている。例えば、動物保護のために野生動物の個体や行動範囲を把握したり、生態系の分析を行うために動物を個体識別して、個体間や群れの中での関係性を観察するなどの方法がとられている。従来は GPS やマイクロチップなどの装置を動物に装着していたが、動物に対する福祉の面から非接触、非侵襲で個体情報が得られる顔個体識別は良い方法だと考えられている。近年の動物の顔個体識別の研究では、人の顔個体識別に使われる深層学習手法を用いて、識別対象となる動物のみのデータセットを学習したものが多い。しかし、特定の動物のみのデータセットを用いた学習では、データ数が不足しがちで、あまり良い精度を実現できていないものも多い。

本論文では、複数の動物のデータセットを用いて、それらの様々な組み合わせで転移学習や Joint Training を行い、複数種類の動物に対する顔個体識別の有効性を検証する。対象となる動物としては、人、チンパンジー、牛、ヤギ、馬の 5 種類を扱う。顔個体識別の学習モデルには ArcFace を利用する。また、特徴活性化マッピングの 1 つである Grad-CAM[7] を用いて、学習モデルが顔画像のどこに注目しているかを可視化する。こ

れにより、先述した転移学習や Joint Training による学習モデルの変化を比較する。

具体的には、2 つの観点から学習による精度向上を検証した。1 つ目の観点は、画像の枚数が多く、バリエーションが豊富な大規模データセットの動物からの転移学習である。大規模データセットで学習したモデルを利用すれば、画像の枚数やバリエーションが少ない小規模データセットの動物も高い精度で識別できると考えられる。2 つ目の観点は、似た顔の動物同士での転移学習や Joint Training である。転移学習するモデルのデータセットが小規模でも、似た顔の動物同士であれば、タスクがより近いものになるので、識別精度の向上が期待できる。

以下、2 章では顔個体識別の従来研究について述べる。3 章では本研究で使用する ArcFace を用いた動物の顔個体識別手法と、Grad-CAM を用いた特性比較について述べる。4 章では識別精度が向上する転移学習の組み合わせを検証し、その結果について述べる。5 章では本研究の結論と今後の課題について述べる。

第 2 章

深層学習による顔個体識別

2.1　顔個体識別と深層学習

顔個体識別とは、入力された顔画像が特定の個体と一致するかどうか判別したり、その顔がどの個体であるかを特定したりする技術である。近年、この分野では深層学習による手法が主流になっている。

深層学習とは、多層構造のニューラルネットワークを用いて、データから特徴を学習する人工知能技術である。人間が計算法を作成する必要のあった特徴量を、大量の学習データを用いることで、モデルが自動で学習し、高次元な特徴を得ることができる。学習モデルの出力は損失関数によって評価される。損失関数とは、機械学習モデルが予測した結果と、実際の正解データとの間の誤差を定量的に測るための指標である。モデルは損失関数の値が最小になるようにパラメータを更新する。選択した損失関数によってモデルが学習する方向性や特性が決まるため、タスクに応じて適切なものを選ぶ必要がある。

顔個体識別においては、深層学習モデルが入力された顔画像を、埋め込み空間と呼ばれる、モデルが個体識別しやすくなる空間で、特徴ベクトルとして表現する。特徴ベクトルは顔の形状、色、部位の配置など、顔に固有の情報を数値的に表現したものである。この表現を用いることで、異なる顔画像間の類似性を評価することができる。

2.2　人の顔個体識別の従来手法

人の顔個体識別は、深層学習と多量のデータセットによって高い識別精度を実現することができていて、スマートフォンの顔認証、監視カメラの自動追跡、医療分野での患者識別など、多様な応用が実現している。その中で重要な役割を果たしている技術の一つが深

層距離学習（Deep Metric Learning）である。深層距離学習は、画像間の類似性を測定するための特徴ベクトルを学習する手法であり、同一人物の顔画像が埋め込み空間で近い位置に、異なる人物の顔画像が離れた位置になるようにニューラルネットワークを訓練する。代表的なアルゴリズムとしては、Triplet Loss を用いた FaceNet[1] や Softmax loss[2] をベースにした ArcFace[3] や CosFace[4] などの手法がある。

2.2.1 FaceNet

FaceNet[1] は、Triplet Loss を用いて、埋め込み空間における特徴ベクトル間のユークリッド距離で個体識別を行う手法である。

学習は、基準となる Anchor 画像、Anchor と同じクラスに属する Positive 画像、Anchor と異なるクラスに属する Negative 画像の 3 つを 1 組 (Triplet) として行われる。各入力画像は、畳み込みニューラルネットワーク (CNN) によって埋め込み空間のベクトルとして表現される。埋め込み空間上で Anchor と Positive の距離 d_p と Anchor と Negatve の距離 d_n を計測する。クラス分布を安定させるために、固定値であるマージン α を d_p に加え、$d_p + \alpha \leq d_n$ となるように学習を行う。ただし、実際には $d_p - d_n + \alpha$ が負の値になる場合は 0 に切り上げる。最終的には特徴ベクトルの距離を最小化し、アンカーとネガティブの特徴ベクトルの距離を最大化する (図 2.1) 。

損失関数は以下のとおりである。

$$L_{triplet} = [d_p - d_n + \alpha]_+ \tag{2.1}$$

$$\text{ただし } [x]_+ = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

これにより、顔特徴を高次元のベクトルとして表現できるようになり、識別性能を高めることができる。

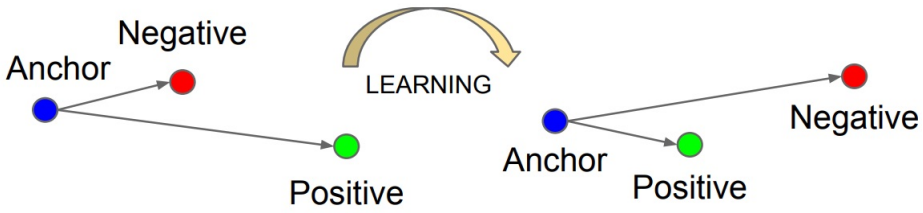

図 2.1　Triplet における学習工程 [1]

2.2.2 Softmax loss ベースの手法

Softmax loss[2] とは、多クラス分類の損失関数として利用されるもので、入力されたものが、どのクラスに属するかを確率で表現する。式は以下の通りになる。

$$L_{softmax} = -\frac{1}{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{N} e^{W_j^T x_i + b_j}} \tag{2.2}$$

- クラス数は N
- x_i は i 番目の学習サンプルの特徴ベクトル
- y_i は i 番目の学習サンプルの正解クラス
- W_j は j 番目のクラスの重み行列で各クラスの代表ベクトルとみなせる
- b_j は j 番目のクラスのバイアス項 (固定値)

ここで、$W_j^T x_i$ に注目すると、ベクトルの内積の定義より、

$$\begin{aligned} W_j^T x_i &= \|W_j^T\|\|x_i\|\cos\theta_j \\ \cos\theta_j &= \frac{W_j^T x_i}{\|W_j^T\|\|x_i\|} \end{aligned} \tag{2.3}$$

となる。バイアス項を 0、θ_j について $0 \leq \theta_j \leq \pi$ とすると、$W_{y_i}^T x_i$ が大きくなるように学習するということは、x_i と正解クラスの代表ベクトル $W_{y_i}^T$ の角度を 0 に近づけるように学習するということになる。しかし、Softmax loss をそのまま利用すると、異なるクラス同士の類似度を低くしたりするような効果はないため、クラス分布が安定しないという問題があった。

この問題を解決するために ArcFace[3] や CosFace[4] などの手法が提案されている。ここでは、代表的な手法である ArcFace について説明する。

ArcFace は、入力画像に対して畳み込みニューラルネットワークが抽出する特徴量ベクトルを、同じクラス同士の特徴量ベクトルの角度は小さく、異なるクラス同士の特徴量ベクトルの角度は大きくなるようにモデルを学習させる (図 2.2)。

損失関数は、以下の通りである。

$$L_{AMP} = -\frac{1}{N}\log\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1, j\neq y_i}^{N} e^{s\cos\theta_j}} \tag{2.4}$$

代表ベクトルと入力画像からの特徴ベクトルをそれぞれ正規化することで、$W_j^T x_i$ を $\cos\theta_j$ と置くことが出来る。ただし、正解クラスの y_i に対しては、マージン m を加える。

これにより、正解クラスとの角度が大きくなるため、マージンを加えない場合よりも、モデルは同じクラス内の特徴をより近づけ、異なるクラス間の特徴をより遠ざける。図 2.3 は 8 クラスのクラス分布を 2 次元で表現したものだが、マージンを加えない softmax では、クラス同士が繋がり円状に分布している。一方で、ArcFace はクラス同士が分離しているため、識別性が高く安定した分布になっている。また、正規化によりベクトルの大きさを 1 に固定した後、内積が角度情報だけに依存する形になるが、数値が小さすぎると、内積の範囲が小さくなり、softmax の機能が低下する可能性がある。これを防ぐために、スケールパラメータ s を用いて内積を拡大し、勾配のスケールを調整する。

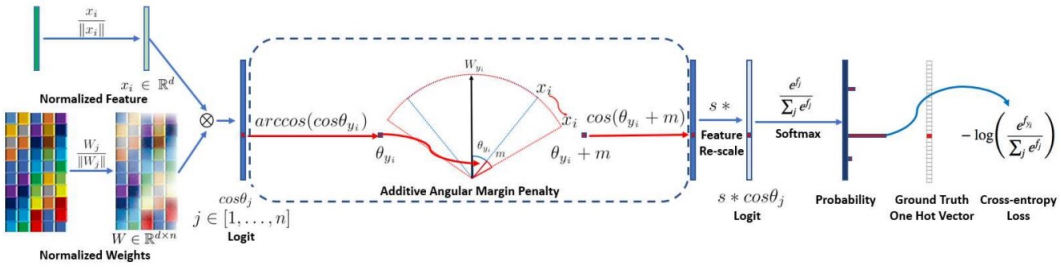

図 2.2　Arcface の処理の流れ [3]

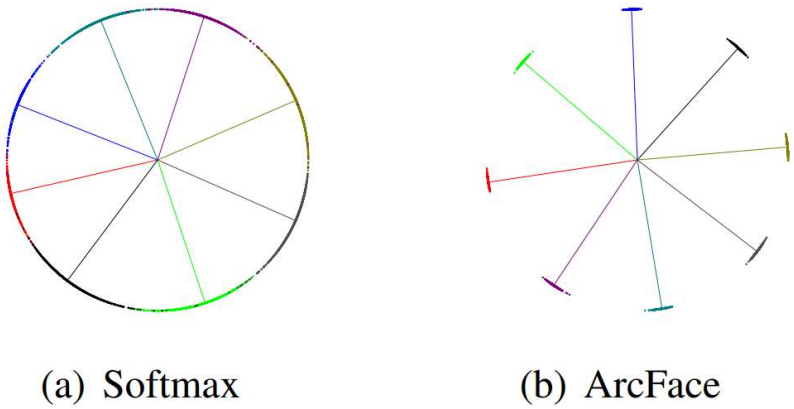

図 2.3　クラス分布の比較 [3]

さらに、マージンやスケールなどのハイパーパラメータを自動で調整する AdaCos[8] や、質の良い画像をクラス分布の中心に寄せる MagFace[9] などが提案されている。Triplet Loss と識別精度に大きな差はないが、最適な Triplet の組み合わせを導き出す必要がなく、学習データを流し込むだけで良いため、近年では角度による個体識別手法が多く提案されている (図 2.4)。

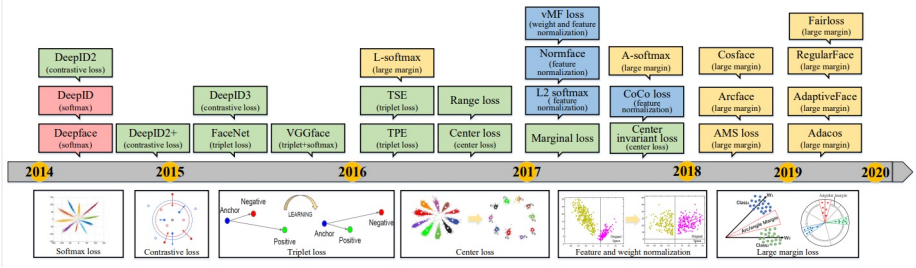

図 2.4　Deep Face Recognition: A Survey[10]

2.3　複数のデータセットやタスクを用いた学習法

深層学習モデルのパフォーマンスを高めるのは、先述したような損失関数の改善だけでなく、複数のデータセットや、関連するタスクを用いた学習法も有効である。大規模なデータセットで学習済みの優れたモデルを利用して、新しいタスクの学習を行う転移学習 (Transfer Learning)[5] や、関連するタスクを同時に学習することで各タスクの精度を向上させる Joint Training[6] などがある。

2.3.1　転移学習

転移学習 (Transfer Learning)[5] とは、あるタスクのために事前に学習したモデルを、類似した別のタスクを学習するモデルの開始点として使用するディープラーニングの手法である。ランダムな初期値からネットワークを学習させる場合、通常は膨大なデータを用いることで、モデルが高い精度を実現できるようになる。大規模なデータセットで学習済みのモデルを転移学習に利用することで、画像処理では、画像のエッジや色などの一般的な特徴を既に学習できているため、転移先のタスクを学習するために必要な学習時間が短縮される。さらに、元のタスクと転移先のタスクに関連性が高いと、転移なしの学習よりもモデルの精度が高まる可能性がある (図 2.5)。

また、新しいタスクで使用できるデータが少ない場合にも、事前学習済みのモデルを用いた転移学習を行うと、データ不足にもかかわらず、精度が高く安定したモデルを生成できる。

ただし、転移学習にはいくつかの注意点がある。転移先の学習データがあまりにも少ない場合は、モデルが過学習を起こす可能性もあり、データ拡張が必要になることある。また、逆に転移先の学習データが膨大になると、事前学習モデルの影響が小さくなり、転移学習の効果が表れないことがある。

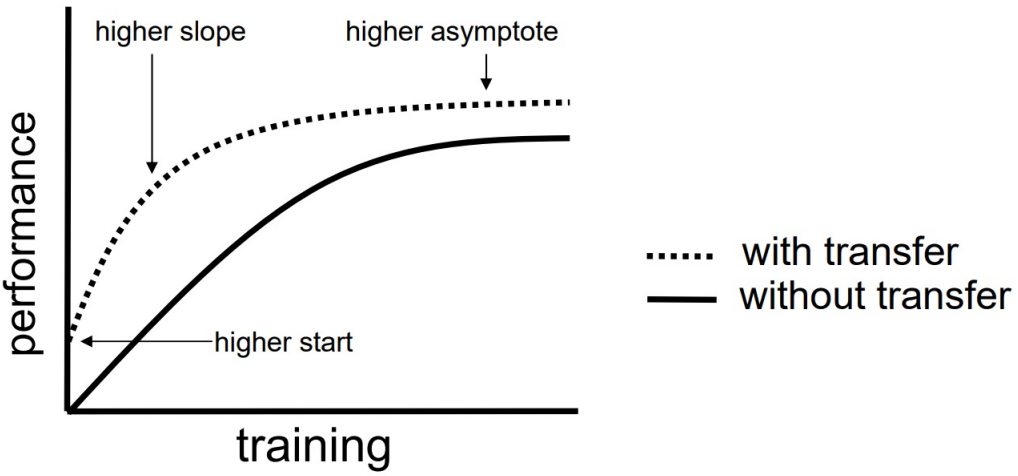

図 2.5　Transfer Learning[5]

2.3.2　Joint training

Joint Training[6] は、複数の関連するタスクを同時に学習する手法である。タスク間の相互関係を利用し、単一のモデル内でタスクを共有するパラメータを学習することで、全体の性能を向上させることを目的としている。これにより、各タスクが独立して学習する場合に比べて、効率的かつ高い精度を達成できる場合がある。複数のタスクが同じモデルで学習されるため、一方のタスクでの学習が他方のタスクの性能向上に寄与する場合がある。また、複数タスクのデータセットを統合して学習するため、各タスクのためのデータセットが十分でなくとも補完できる。さらに、モデルがタスク間で共通する特徴を学習することで、汎用性が高まり、新しいデータや未学習タスクにも適応する。ただし、タスク間で学習データの差が大きくなってしまうと、片方のタスクが優先されることがある。また、転移学習と比べると、タスク間の相互関係の重要度が高く、タスク同士の関係性が弱

いと、あまり性能が向上しない。

2.4 動物の顔個体識別と従来研究

動物の顔個体識別は、人と同様に顔の特徴を基に個体を識別する技術である。この技術は、生態調査や動物保護、農業、ペット管理など、多岐にわたる分野での活用が期待されている。家畜やペットなどの動物の行動や健康状態を追跡するために個体識別は重要である。従来はタグや GPS 装置などを対象となる動物個体に装着させる手法が用いられてきた。しかし、これらの手法には手間がかかるという課題がある。顔個体識別は、非接触で比較的簡単な方法として注目されている。また、動物に対する福祉の面から非接触、非侵襲で個体情報が得られる顔個体識別は良い方法だと考えられている。顔個体識別では、自動撮影カメラで撮影された画像から、タグや装置を使わずに個体識別が可能になる。

実際に、動物の顔個体識別を行った研究例も多く存在する。Andreas ら [11] では、ガボールフィルタとスパース表現を用いることで抽出した特徴量と、SURF[12] による特徴点をもとにした特徴量を組み合わせて顔個体識別を行っている。兒玉 [13] は、モバイル端末から撮影した牛の顔画像から、VGG16 を用いて特徴量を抽出し、事前に登録済みの個体の中で最も類似している個体識別番号を取得することで牛の管理を支援する CowFindAR を提案した。Yu ら [14] は、ヤギの顔について顔個体識別を行っている。顔の部位をランドマークとして (図 2.6)、畳み込みニューラルネットワークで特徴を抽出し、softmax を用いてクラス識別を行っている。また、この研究では、人の顔画像のデータセットである VGGFace で学習したモデルから転移学習を行い、高精度の識別を実現している。また、篠田ら [15] は、13 種類の動物で合計個体数が 25 万、合計画像数およそ 100 万枚の大規模データセット Petface を提案している (図 2.7)。この論文では、Triplet や Softmax、ArcFace など人の顔識別で用いられる手法を用いて実験を行っている。また、各動物ごとに学習した結果と、全ての動物で Joint Training した結果を比較しており、一部の動物は Joint Training を行うことで、識別精度を向上させることができている。

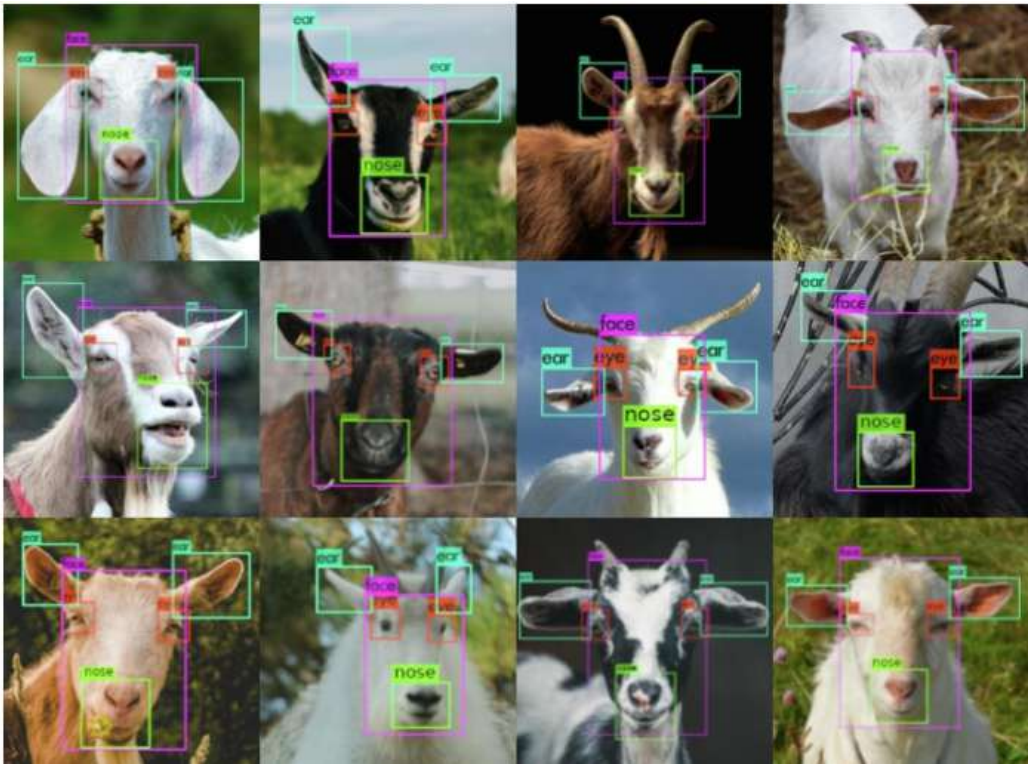

図 2.6　ヤギのランドマーク [14]

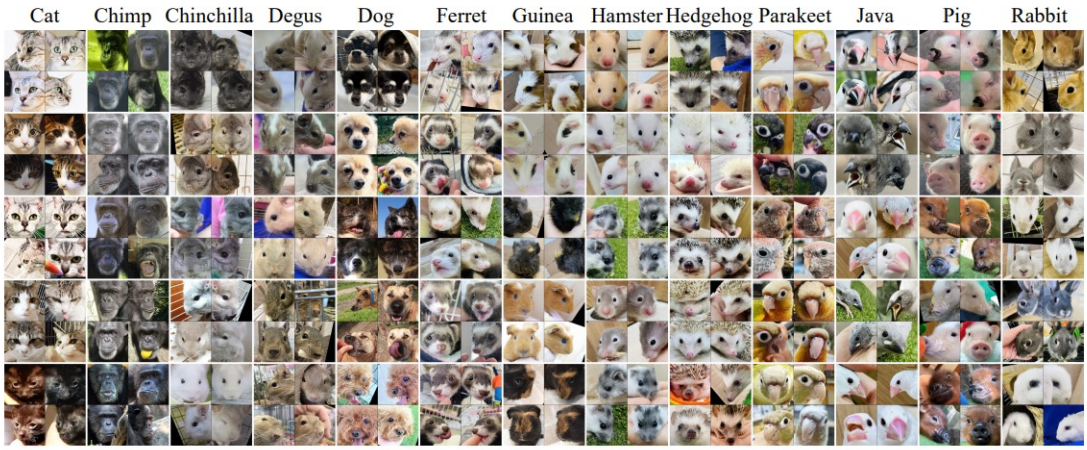

図 2.7　Petface[15]

第 3 章

ArcFace を用いた動物の顔個体識別手法

3.1　動物の顔識別個体の課題

2.4 節では、動物の顔個体識別が行われた例について説明したが、人の顔個体識別と比べると、十分な識別精度に達していない動物が多く、改善の余地がある。特定の動物のみのデータセットを用いた学習では、データ数が不足しがちで、あまり良い精度を実現できていないものも多い。[11] ではチンパンジーの顔個体識別について識別精度が 90% 未満である。また、[15] でも、13 種類の動物で Joint Training を行っているが、6 種類の動物で識別精度が 90% を越えていない。これは、各動物に対する適切な学習方法が確立されていないためだと考えられる。

本論文では、人、チンパンジー、牛、ヤギ、馬のデータセットを用いた様々な組み合わせで、転移学習や Joint Training を行い、複数種類の動物に対する顔個体識別の有効性を検証する。顔個体識別の学習モデルには ArcFace を利用する。また、特徴活性化マッピングの 1 つである Grad-CAM を用いて、学習モデルが顔画像のどこに注目しているかを可視化する。これにより、先述した転移学習や Joint Training による学習モデルの変化を比較し、顔個体識別における各動物の関係性や、学習モデルが改善される学習方法について考察する。

3.2　データセット

本論文では、人、チンパンジー、牛、ヤギ、馬のデータセットを用いて、これらの動物の顔個体識別の特性比較を行う。以下では、各動物のデータセットについて述べる。

3.2.1　人

人のデータセットには VGGFace2[16] を使用する (図 3.1)。VGGFace2 は、個体数 9131、画像枚数 330 万枚を超える大規模のデータセットである。様々な年齢や性別、人種、角度、表情、照明条件を含むため、大規模データセットとして、[3] や [9] などでも使用されている。

図 3.1　VGGFace2 の一部

3.2.2　チンパンジー

チンパンジーのデータセットには、Alexander ら [17] が作成した CZoo を使用する (図 3.2)。このデータセットは、動物園で飼育されたチンパンジーを撮影したもので、個体数は 24、画像枚数は約 2000 枚である。角度や表情にバリエーションがあり、植物などで顔の一部が隠れている画像や、角度により顔の部位がほとんど見えない画像なども含まれている小規模データセットである。

図 3.2　CZoo の一部

3.2.3　牛

牛のデータセットには、実際の農場で撮影したものを使用する (図 3.3)。牛の顔を撮影した動画から切り出した画像になっていて、個体数 475、画像枚数約 49 万枚である。人のデータセットほど大規模なデータセットではないが、チンパンジーや後述する動物と比べると比較的個体数や画像枚数が多く、本論文では大規模データセットとして扱う。

図 3.3　牛データセットの一部

3.2.4　ヤギ

ヤギのデータセットには [14] で提案されたデータセットを使用する (図 3.4)。牛のデータセットと同様に、ヤギを撮影した動画から画像を切り出している。個体数が 10 で画像枚数約 1200 枚と個体数と画像枚数共に少ない、小規模データセットである。

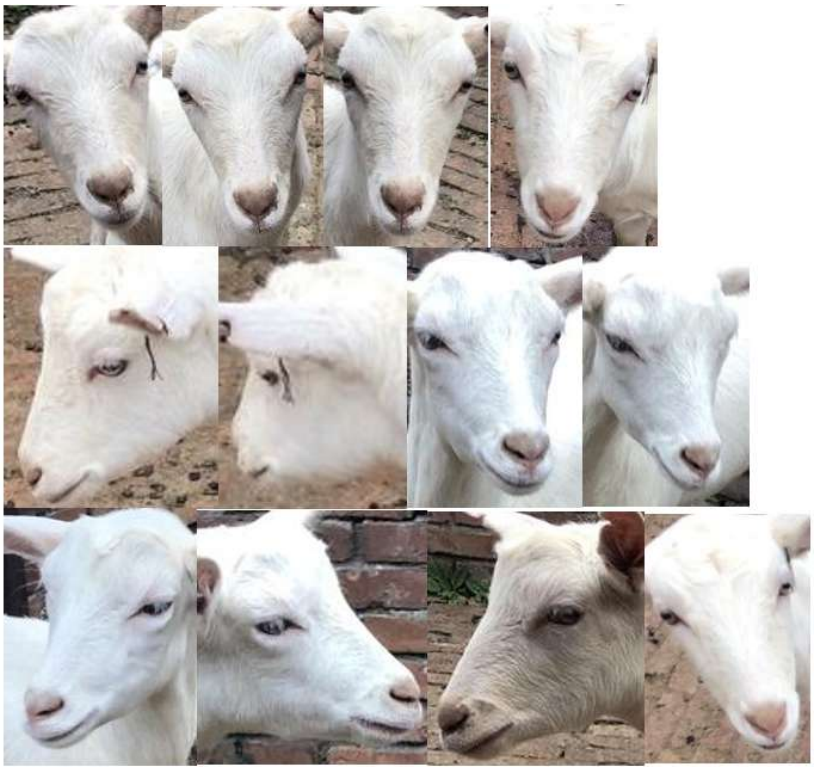

図 3.4　ヤギデータセットの一部

3.2.5　馬

馬のデータセットには、THoDBRL2015 Database[18] を使用する (図 3.5)。馬のデータセットは、個体数 47、画像枚数約 1400 枚の小規模データセットである。各個体には、正面、左、右の写真が含まれている。馬の毛色は茶、白、黒が含まれ、模様のある馬も存在する。

図 3.5　THoDBRL2015 Database の一部

3.3　ArcFace による識別精度比較

[14] や [15] から、人以外の動物にも人の顔識別のために提案された手法は有効だと考えられるため、ArcFace を用いる。ArcFace の構造を図 3.6 に示す。ArcFace は、特徴ベクトルを生成する back bone と学習時に個体への分類を行う head からなる。本研究では、back bone に Resnet18[19] を使用する。head には前述した ArcFace の損失関数を使用する。転移学習では、事前学習済みモデルの back bone の重みを利用する。また、推論時には head を除去して、backbone で出力した特徴ベクトル同士を比較することで個体識別を行う。識別精度の測定は訓練用とは異なるテスト用のデータを用いて行う。測定方法は、まず、テスト用のデータを半分に分割し、一方のデータから特徴ベクトルを抽出して個体を登録する。その後、もう一方のデータから特徴ベクトルを抽出し、登録済みの特徴ベクトルと比較し、最も類似する特徴ベクトルをもつ個体に識別する。この正答率を識別精度として扱う。

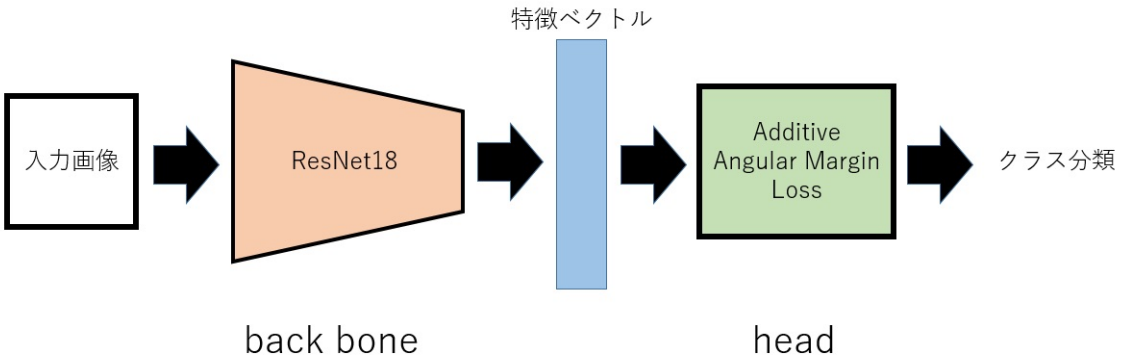

図 3.6　ArcFace の構造

3.4　Grad-CAM

Grad-CAM(Gradient-weighted Class Activation Mapping)[7] は、クラス活性化マッピング (Class activation mapping) 手法の 1 つである。クラス活性化マッピングは、画像分類モデルが特定のクラスを予測する際に、どの領域が重要だったかを視覚化する手法で、Grad-CAM では、特定のクラスに対する損失関数の勾配を用いて、最終畳み込み層の特徴マップの重要性を計算する。具体的には、特徴マップごとに損失の勾配を計算し、

その平均を重みとして使用する。特定のクラスを c、c に対する出力を y^c、最終畳み込み層の特徴マップの k 番目のチャネルを A^k、特徴マップのサイズを $Z(i \times j)$、重みを α_k^c とすると、

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{3.1}$$

と表すことが出来る。これにより、各特徴マップが、y^c として判別するためにどれだけ寄与しているかを計算したことになる。

この重み α_k^c を特徴マップにかけ合わせて線形結合し、ヒートマップとして視覚化することで、モデルが特定のクラスを予測する際に注目した領域を示す (式 3.2)。ReLU を適用するのは、重要度の低い領域を除外し、重要度の高い領域のみを強調するためである。

$$L^c = ReLU(\sum_k \alpha_k^c A^k) \tag{3.2}$$

また、Grad-CAM はヒートマップを入力画像の上に表示することで、元の解像度での視覚化を可能にしており、モデルの判断根拠を理解しやすくなる。

本研究では、Grad-CAM を用いて、各学習法によって顔画像に対する学習モデルのふるまいの変化を比較し、各動物の組み合わせの相性を、識別精度の変化を踏まえて考察する。ArcFace の Head を除去して Grad-CAM を適用する。head を除去しない場合、特定クラスへの出力スコアが対象のため、ヒートマップはそのクラスに特化した領域を強調するが、head を除去した場合は、特徴ベクトルが対象となる。この際、特定クラス c の代わりに、特徴ベクトルの中で最大の要素に着目して視覚化する。ヒートマップはモデルが顔画像のどの部分を特徴抽出の際に重視したかを示す。

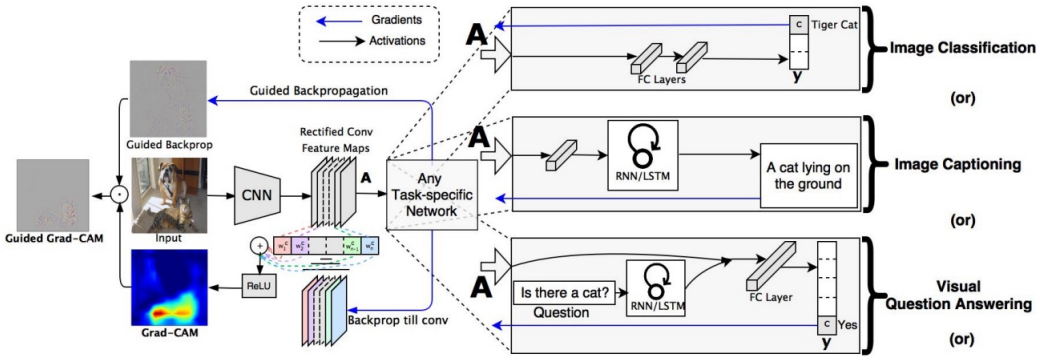

図 3.7　Grad-CAM [7] より

第 4 章

実験

4.1　実験設定

本研究では、複数の動物のデータセットを用いて、それらの様々な組み合わせで転移学習や Joint Training を行い、複数種類の動物に対する顔個体識別の有効性を検証する。対象とする動物は、人、チンパンジー、牛、ヤギ、馬の 5 種類である。各画像はサイズを 224 × 224 に調整した。

それぞれの動物のデータセットは、訓練検証用とテスト用に分割し、訓練検証用とテスト用で異なる個体を使用した。データセットの詳細を表 4.1 に示す。

表 4.1　データセット内訳

動物	訓練検証用個体	訓練検証用画像枚数	テスト用個体	テスト用画像枚数
人	7,000	2,541,526	2,131	766,514
チンパンジー	14	1,267	10	842
牛	400	414,881	75	76,700
ヤギ	7	906	3	385
馬	31	930	16	480

4.1.1　データ拡張

データ拡張として、以下の手法を適用した：

- アフィン変換
- 中央切り抜き
- 左右反転
- ぼかし

ただし、小規模データセットで転移学習を行わない場合、一部のデータ拡張を削減することで収束性を向上させた。

4.1.2　学習条件

学習条件は以下のように設定した。

- **バッチサイズ**: 300
- **学習回数**: すべての手法で 100 エポック
- **学習率**: 0.01 から開始し、10 エポックごとに 0.5 倍に減少

4.2　転移学習による識別精度比較

ここでは、5 種類の動物の様々な組み合わせで転移学習を行った結果について述べる。各動物について識別を行い、訓練 (train) と検証 (valid) とテスト (test) における精度を表に記している。訓練と検証は学習時の head が含まれた状態でのクラス分類精度になる。テストは 3.3 節で述べた、head を除去した状態でテストデータの半分を登録して残りのデータで識別を行った際の精度になる。

4.2.1　人

表 4.2 に人に対する識別精度を示す。大きな差ではないが、転移学習なしで学習したモデルが最も良い精度になっている。これは人のデータセットが大規模で、転移学習の影響が小さくなっているからだと考えられる。

表 4.2　人に対する各転移学習の識別精度

事前学習した動物	train	valid	test
なし	**0.85**	**0.89**	**0.93**
チンパンジー	0.83	0.88	0.92
牛	0.84	0.88	0.92
ヤギ	0.83	0.87	0.90
馬	0.84	0.87	0.92

4.2.2　チンパンジー

チンパンジーについては、転移学習なしでの学習で収束しなかったため、アフィン変換と中央切り抜きのみをデータ拡張に使用した。

表 4.3 にチンパンジーに対する識別精度を示す。訓練検証では、人と牛からの転移学習が最も良い精度になっていて、テストについては、人からの転移学習が最も良い精度になっている。

表 4.3　チンパンジーに対する各転移学習の識別精度

事前学習した動物	train	valid	test
なし	0.93	0.87	0.69
人	**0.99**	**0.95**	**0.85**
牛	**0.99**	**0.95**	0.83
ヤギ	0.93	0.85	0.69
馬	0.89	0.82	0.70

4.2.3　牛

表 4.4 に牛に対する識別精度を示す。訓練検証では、転移学習なしが、テストについては、チンパンジーからの転移学習が最も良い精度になっている。ただし、各手法に大きな精度の差はない。これは、人と同様に牛は大規模データセットなため転移学習の影響が小さくなっているからだと考えられる。

表 4.4　牛に対する各転移学習の識別精度

事前学習した動物	train	valid	test
なし	**0.99**	**0.96**	0.93
人	**0.99**	0.94	0.92
チンパンジー	**0.99**	0.94	**0.94**
ヤギ	0.99	0.95	0.91
馬	0.99	0.94	0.92

4.2.4　ヤギ

ヤギについては、転移学習なしでの学習が収束しなかったため、アフィン変換と中央切り抜きのみをデータ拡張に使用した。

表 4.5 にヤギに対する識別精度を示す。検証とテストのどちらも人での転移学習が良い精度であった。ただし、チンパンジーを除くどの動物からの転移学習でもテストデータに対する識別精度が向上しており、小規模データセットである馬からの転移学習でも効果があると分かる。

表 4.5　ヤギに対する各転移学習の識別精度

事前学習した動物	train	valid	test
なし	**1.00**	0.86	0.82
人	0.97	**0.96**	**0.94**
チンパンジー	0.96	0.93	0.81
牛	0.92	0.92	0.90
馬	0.95	0.90	0.83

4.2.5　馬

表 4.6 に馬に対する識別精度を示す。訓練検証では、人からの転移学習が最も良い精度になっているが、テストについては、牛からの転移学習が最も良い精度になっている。また、チンパンジーを除くどの動物からの転移学習でもテストデータでの識別精度が向上し

ている。

表 4.6　馬に対する各転移学習の識別精度

事前学習した動物	train	valid	test
なし	0.95	0.89	0.86
人	**1.00**	**0.99**	0.91
チンパンジー	0.99	0.95	0.86
牛	0.99	0.95	**0.93**
ヤギ	0.97	0.96	0.88

4.2.6　転移学習の有効性についての考察

転移学習に関して、まず、人と牛の大規模データセットでは、各手法であまり識別精度が変化しなかった。このことから、学習するデータセットが十分な大規模データセットの場合には、転移学習を行っても、あまり精度向上に寄与しないと考えられる。チンパンジー、ヤギ、馬の小規模データセットでは、共通して人や牛からの転移学習による識別精度の向上が見られた。このことから、小規模データセットに対しては、大規模データセットからの転移学習が有効であると考えられる。

また、チンパンジーの識別では人からの転移学習が最も精度が高く、ヤギと馬のそれぞれの転移学習でもやや精度が向上していた。このことから、似た顔同士の動物での転移学習が有効な傾向にあると言える。

4.3　Joint Training による識別精度の比較

ここでは、Joint Training を行った結果について述べる。学習時には、異なる種類の動物のデータを同列に扱い、各個体を 1 クラスとする。

実験では、まず、5 種類すべての動物を合わせて Joint Training を行い、Joint Training 全般の有効性について評価する。

次に小規模データセット同士で類似した動物同士の Joint Training の有効性について評価する。Joint Training の組み合わせにはヤギと馬及び、データ数を減らした人とチンパンジーを使用する。また、これらについては、転移学習との組み合わせについても評価する。

4.3.1　5 種類の動物での Joint Training

人、チンパンジー、牛、ヤギ、馬の 5 種類の動物のデータセットを合わせて Joint Training を行い、各動物のテストデータでの識別結果を評価する。比較のために、Joint Training なしで各動物単独での学習も行った。結果を表 4.7 に示す。

識別精度を比較すると、チンパンジーのみ識別精度が向上し、他の動物は精度が変わらないか低下している。

表 4.7　5 種類の動物での Joint Training

動物	Joint test	No Joint test
人	0.90	0.93
チンパンジー	0.69	0.61
牛	0.78	0.88
ヤギ	0.74	0.74
馬	0.86	0.88

4.3.2　ヤギと馬での Joint Training

データセットの規模と、タスクの類似性を考慮して、小規模データセットで同じ草食動物のヤギと馬による Joint Training を行った。また、大規模データセットの人や牛で事前に学習したモデルでの転移学習と組み合わせた Joint Training も行っている。ヤギの識別結果を表 4.8 に、馬の識別結果を表 4.9 に示す。

Joint Training なしの結果と比較すると、ヤギと馬の両方で識別精度が向上している。一方で、転移学習と Joint Training を組み合わせても、馬の識別に対する人の転移学習以外では精度は向上していない。

表 4.8　Joint Training によるヤギの識別結果

事前学習した動物	Joint test	No Joint test
なし	0.84	0.82
人	0.94	0.94
牛	0.89	0.90

表 4.9　Joint Training による馬の識別結果

事前学習した動物	Joint test	No Joint test
なし	0.90	0.86
人	0.96	0.91
牛	0.92	0.99

4.3.3　チンパンジーと人での Joint Training

チンパンジーと類似している小規模タスクを用意するために、VGGFace2 の一部を取り出して、小規模データセットとして Joint Training を行った。個体数は 15、画像枚数は 750 枚を取り出した。識別結果を表 4.10 に示す。人のデータセットは Joint Training に使用しているため、転移学習には牛の事前学習モデルのみ使用している。

Joint Training なしの結果と比較すると、識別精度は向上している。しかし、ヤギや馬と同様に、転移学習と組み合わせても精度は向上していない。

表 4.10　Joint Training によるチンパンジーの識別結果

事前学習した動物	Joint test	No Joint test
なし	0.72	0.69
牛	0.79	0.83

4.3.4　Joint Training の有効性に関する考察

Joint Training に関しては、5 種類の動物すべてを合わせて学習した場合はチンパンジーを除く 4 種類の動物で、テストデータに対する識別精度が低下していた。これはデー

タセットの割合の不均衡が原因と考えられる。この Joint Training では、人の学習データがほとんどを占めるため、他の動物の個体についてよりも人の分類を重視した方が損失が低くなる。そのため、他の動物については重視されなくなるため識別精度が低下したと考えられる。

ヤギと馬、人とチンパンジーの似た顔で小規模データセット同士の Joint Training については、それぞれ識別精度が向上した。この結果から、類似している動物での Joint Training は有効だと考えられる。

一方で、転移学習と比較すると、すべての場合で、転移学習を行った学習モデルの方が識別精度が高かった。Joint Training と転移学習を組み合わせた場合でも、あまり精度が改善しないため、転移学習の方が有効であると考えられる。

4.4 Grad-CAM による比較

Grad-CAM による転移学習や Joint Training による学習モデルの変化を比較する。まず、精度に大きな変化がなかった人と牛の学習モデルの出力を確認する。次に、小規模データセットのチンパンジー、ヤギ、馬については、精度が向上した学習モデルからの出力を転移学習と Joint Training に分けて比較する。

4.4.1 大規模データセットのヒートマップ

人について、転移学習を行わずに学習したモデルのヒートマップを図 4.1 に示す。Grad-CAM の出力するヒートマップは、赤い領域ほど注目度が高く、青い領域はあまり注目していないことを示す。図 4.1 では、鼻付近に注目していることが確認できる。一方、牛から転移学習を行ったモデル（図 4.2）でも、鼻を中心に注目している様子はほとんど変化していない。牛についても同様の傾向が見られる。転移学習を行わずに学習したモデル（図 4.3）と、人から転移学習を行ったモデル（図 4.4）を比較しても、注目領域にほとんど差異が確認されなかった。

図 4.1　人：転移学習なし

図 4.2　人：牛からの転移学習

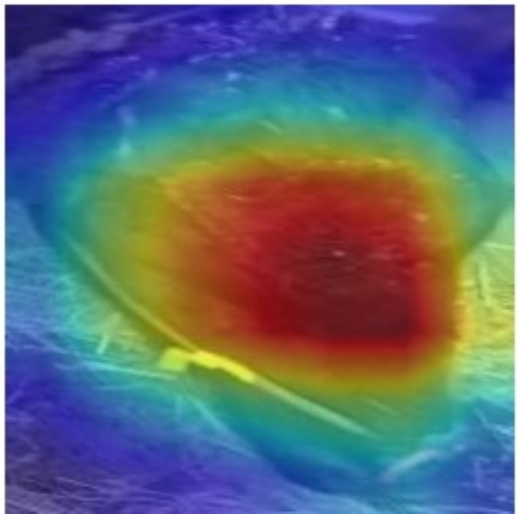

図 4.3　牛：転移学習なし

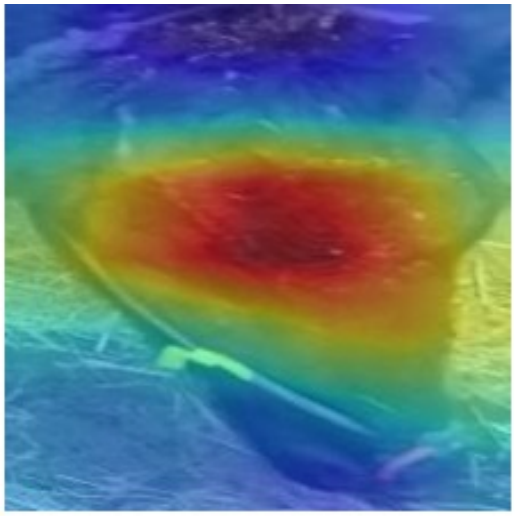

図 4.4　牛：人からの転移学習

4.4.2　小規模データセットの転移学習でのヒートマップ

チンパンジーについて、転移学習を行わずに学習したモデル（図 4.5）では、識別精度が低く、注目する場所が不安定で顔を正確に捉えることができていない。一方、人から転移学習したモデル（図 4.6）では、顔の領域に注目している。注目する場所が安定することで識別精度が向上したと考えられる。同様に、ヤギにおいても転移学習や Joint Training を行わない場合（図 4.7)、鼻付近に注目する傾向は見られるが、注目範囲が不安定で背景の影響を受けている。一方、人から転移学習したモデル（図 4.8）では、目と鼻の間に注目が集中しており、顔を捉えることができている。馬の場合も、転移学習や Joint Training を行わないモデル（図 4.9）では、顔の領域をある程度捉えてはいるが注目が不安定である。これに対し、牛から転移学習を行ったモデル（図 4.10）では、注目が鼻に集中している。

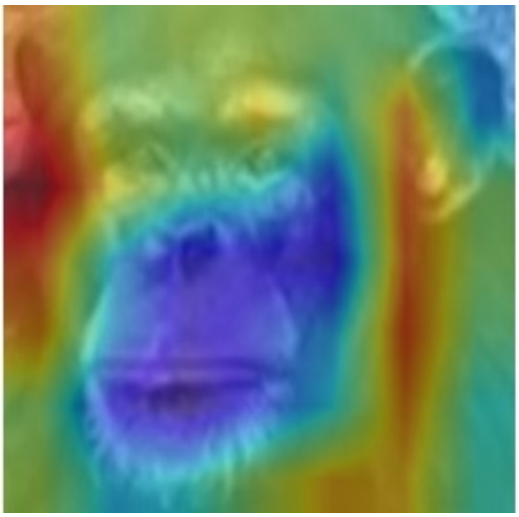

図 4.5　チンパンジー：転移学習なし

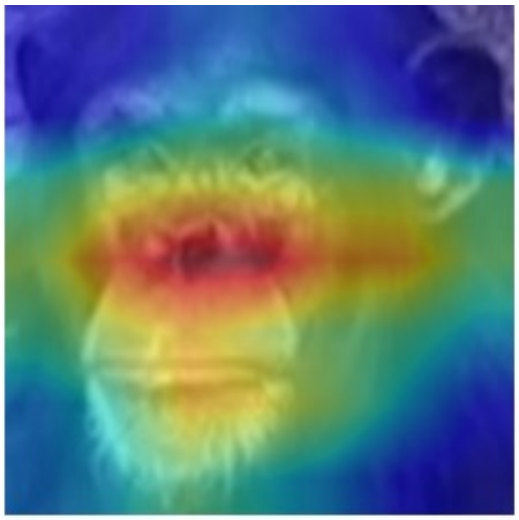

図 4.6　チンパンジー：人からの転移学習

図 4.7　ヤギ：転移学習なし

図 4.8　ヤギ：人からの転移学習

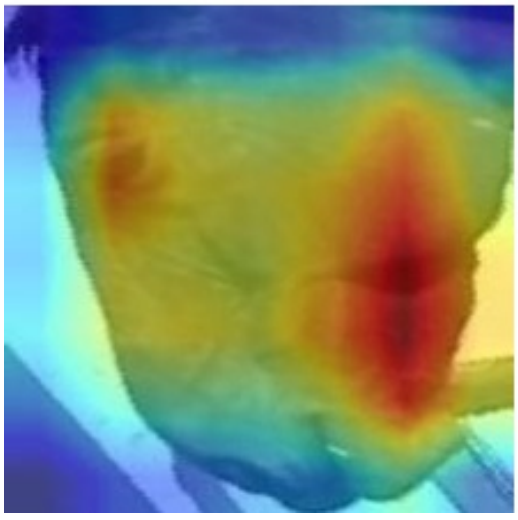

図 4.9　馬：転移学習なし

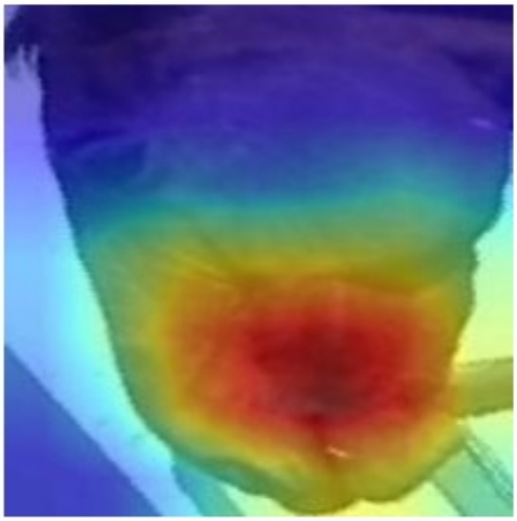

図 4.10　馬：牛からの転移学習

4.4.3　小規模データセットの Joint Training でのヒートマップ

チンパンジーについて、Joint Training を行わずに学習したモデル (図 4.11) は顔を捉えられていない。一方、人と Joint Training を行ったモデル（図 4.12）は、識別精度の面では大きな向上は見られなかったが、転移学習を行った学習モデル (図 4.6) と同様に顔に注目していることが確認できる。ヤギについても、Joint Training を行わずに学習したモデル（図 4.13）では顔を捉えられていないが、馬と Joint Training を行ったモデル（図 4.14）では顔を的確に捉えるようになっていることが分かる。馬については、Joint Training を行わずに学習したモデル（図 4.15）では顔の注目点がやや不安定である。一方、ヤギと Joint Training を行ったモデル（図 4.16）では、やや牛で転移学習した図 4.10 に似た出力で、注目点が鼻に近くなっている。

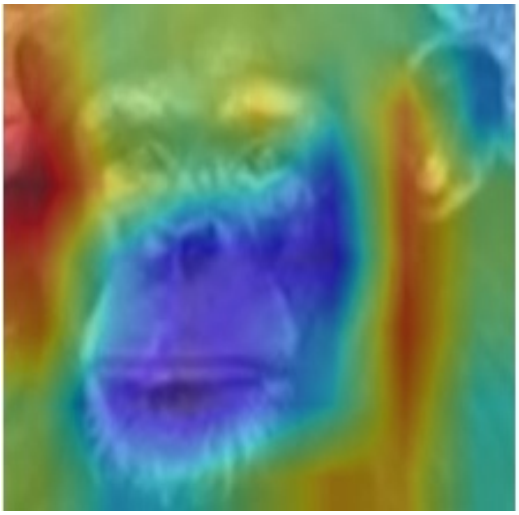

図 4.11　チンパンジー：Joint Training なし

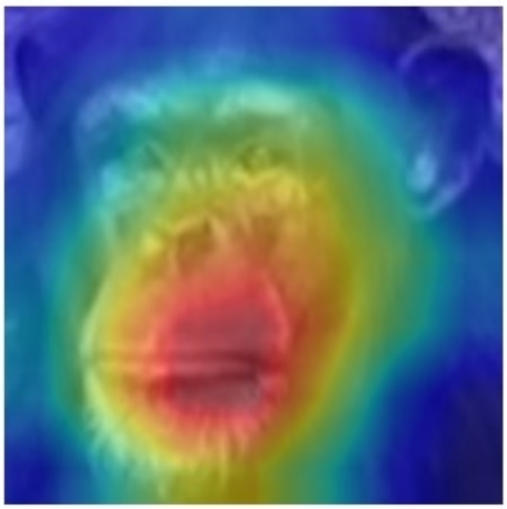

図 4.12　チンパンジー：人との Joint Training

図 4.13　ヤギ：Joint Training なし

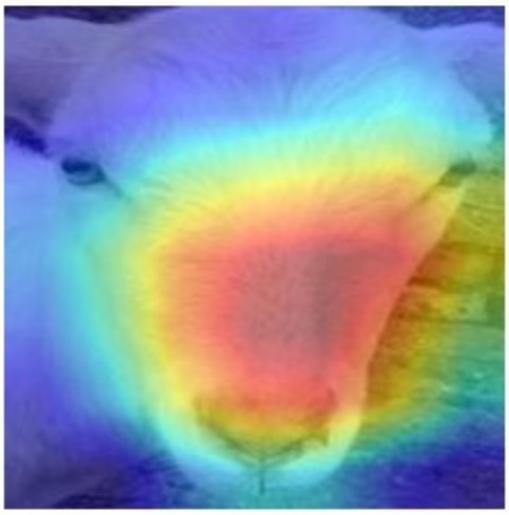

図 4.14　ヤギ：馬との Joint Training

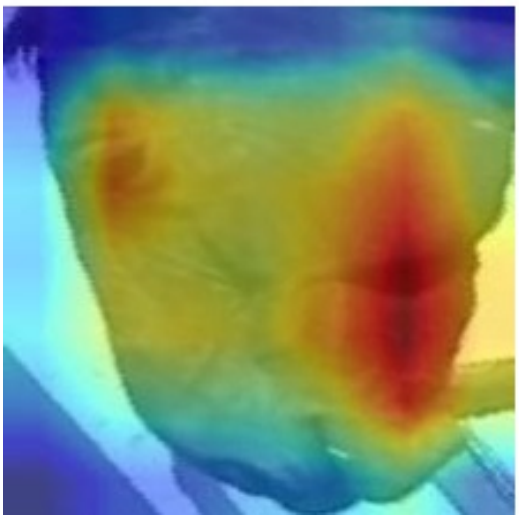

図 4.15　馬：Joint Training なし

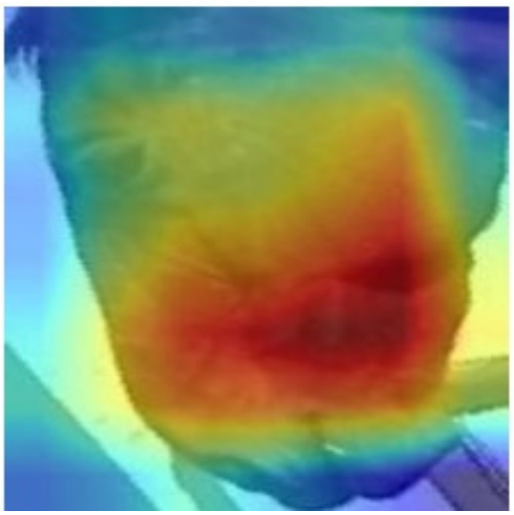

図 4.16　馬：ヤギとの Joint Training

4.4.4　Grad-CAM による特性比較の考察

転移学習については、Grad-CAM によるヒートマップを比較すると、各小規模データセットのみで学習したモデルは注目する場所が不安定であった。一方で、大規模データセットからの転移学習を行ったモデルでは、顔を捉えられるようになり、注目する場所が安定した。このことが、識別精度の向上に繋がったと考えられる。

Joint Training については、ヤギやチンパンジーでは転移学習を行った学習モデルと同等に顔を捉えていた。識別精度には表れていないが、学習モデルの性能は向上しているのではないかと考えられる。

第 5 章

おわりに

本論文では、人、チンパンジー、牛、ヤギ、馬の 5 種類の動物の顔個体識別について、人の顔個体識別の代表的な手法である ArcFace による識別精度比較と Grad-CAM による特性比較を行い、各動物の様々な組み合わせでの転移学習や Joint Training の有効性を検証した。

転移学習に関しては、まず、人と牛の大規模データセットでは、各手法であまり識別精度が変化しなかった。このことから、学習するデータセットが十分である場合、転移学習を行っても識別精度の向上にあまり寄与しないと考えられる。

チンパンジー、ヤギ、馬の小規模データセットでは、共通して人や牛からの転移学習による識別精度の向上が見られた。Grad-CAM によるヒートマップを比較しても、各小規模データセットのみで学習したモデルは注目する場所が不安定であったが、大規模データセットからの転移学習により、注目する場所が安定した。また、チンパンジーの識別では人からの転移学習が最も精度が高く、ヤギと馬のそれぞれの転移学習についても、やや精度が向上していた。この結果から、似た顔同士の動物での転移学習が有効な傾向にあると言える。

Joint Training については、ヤギと馬、人とチンパンジーの似た顔の小規模データセットで識別精度が向上した。Grad-CAM による特性比較では、転移学習を行った学習モデルと同等に、安定して顔やその部位を捉えている組み合わせもあり、識別精度には表れない性能の向上を確認した。

今後の課題としては、さらなる動物種の追加や動物の組み合わせによる識別性能の分析を進めるとともに、識別精度を向上させるための学習手法の確立を目指すことが挙げられる。また、実際の使用を想定し、野外環境や異なる撮影条件下での識別性能の評価を行い、より実用的な動物個体識別モデルへ改善する必要がある。

謝辞

本論文の作成にあたり、丁寧な対応とご指導をして下さった椋木雅之教授に深く感謝申し上げます。実験やデータセットの調査などでも助力して頂きありがとうございました。

工学部教育研究支援技術センターの皆様には、修士と技術職員の兼業で多大な配慮を頂きました。おかげさまで、兼業の中でも本論文を作成できました。ありがとうございました。

お互いの研究を進めるにあたって助言し合うことで切磋琢磨した椋木研究室の皆様に感謝申し上げます。これからのご活躍をお祈り申し上げます。

本研究は科研基盤 (C)23K11151 の助成を受けて実施した。

参考文献

[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[2] Ludwig Boltzmann. Studien uber das gleichgewicht der lebendigen kraft zwischen bewegten materiellen punkten. In *Wiener Berichte 58*, pp. 517–560, 1868.

[3] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 10, p. 5962–5979, October 2022.

[4] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5265–5274, 2018.

[5] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264. IGI global, 2010.

[6] Jonathan Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'14, p. 1799–1807, Cambridge, MA, USA, 2014. MIT Press.

[7] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE inter-*

national conference on computer vision, pp. 618–626, 2017.

[8] Xiao Zhang, Rui Zhao, Yu Qiao, Xiaogang Wang, and Hongsheng Li. Adacos: Adaptively scaling cosine logits for effectively learning deep face representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10823–10832, 2019.

[9] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14225–14234, 2021.

[10] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, Vol. 429, pp. 215–244, 2021.

[11] Alexander Loos and Andreas Ernst. An automated chimpanzee identification system using face detection and recognition. *EURASIP Journal on Image and Video Processing*, Vol. 2013, pp. 1–17, 2013.

[12] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, Vol. 110, No. 3, pp. 346–359, 2008.

[13] 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, 宮崎大学大学院修士論文, 2021.

[14] Masum Billah, Xihong Wang, Jiantao Yu, and Yu Jiang. Real-time goat face recognition using convolutional neural network. *Computers and Electronics in Agriculture*, Vol. 194, p. 106730, 2022.

[15] Risa Shinoda and Kaede Shiohara. Petface: A large-scale dataset and benchmark for animal identification, 2024.

[16] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pp. 67–74. IEEE, 2018.

[17] Alexander Freytag, Erik Rodner, Marcel Simon, Alexander Loos, Hjalmar S Kühl, and Joachim Denzler. Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates. In *Pattern Recognition: 38th German Conference, GCPR 2016, Hannover, Germany, September 12-15, 2016, Proceedings 38*, pp. 51–63. Springer, 2016.

[18] Islem Jarraya, Wael Ouarda, and Adel M. Alimi. Thodbrl2015 database, 2021.

[19] Andreas Veit, Michael Wilber, and Serge Belongie. Residual networks behave like ensembles of relatively shallow networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, p. 550–558, Red Hook, NY, USA, 2016. Curran Associates Inc.