

令和元年度卒業論文

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

高宗 伸幸

指導教員 椋木 雅之 教授

目次

1. はじめに	2
2. 転移学習による牛の個体識別.....	4
2.1 転移学習	4
2.2 識別処理の流れ	4
2.3 特徴量の抽出.....	5
2.3 識別器による判定.....	7
3. 実験.....	8
3.1 実験 1 精度評価	8
3.1.1 実験データ	8
3.1.2 評価方法.....	9
3.1.3 実験結果.....	10
3.2 実験 2 ヒートマップによる確認.....	13
3.2.1 評価方法.....	13
3.2.2 実験結果.....	13
3.3 実験 3 人の顔画像を与えた場合の検証.....	19
3.3.1 実験データ	19
3.3.2 評価方法.....	19
3.3.3 実験結果.....	19
4. おわりに	21
謝辞.....	22
参考文献	23

1. はじめに

畜産農業において家畜として飼われている牛には、伝染病を持っているなどの要注意な個体が存在する。これらの個体を識別することが重要である。一般に牛の識別には耳のタグが用いられるが、この方法では顔の向きやタグの汚れによって即座の判別が難しい場合があり、より簡便な識別手法が必要とされている。

そこで、牛の生体情報を利用して個体の識別を行う研究が行われてきた。先行研究^[1]では古典的な顔認識手法である主成分分析や線形判別分析、独立成分分析、およびそれらの発展手法を用いて個体の識別を行った。これによって、牛の顔の情報による個体識別の可能性が示された。しかしこの手法は精度的に十分とは言えず、より高精度な手法が必要と言える。

一方、近年、深層学習を用いた手法がコンピュータビジョンの多くの分野において大きな成果を上げており、人の顔認識においても高い精度を示している^[2]。これは、深層学習が人の顔の特徴を上手く捉えられている証拠である。牛の場合でも深層学習により同様に顔の特徴を捉えられると考えられる。一方で、一般に深層学習を行う場合には大量の学習データが必要となる。しかし、大規模な学習データの準備には多くの労力を必要とする。そこで、学習済みの深層学習モデルを利用した転移学習という手法がある。この転移学習を用いることで、少ない学習データで高い精度を実現することが可能である。

本研究では、転移学習を用いて牛の個体識別を行う。転移学習には2種類の学習済み深層学習モデルを用いる。1つは人間の顔を学習したモデルである。顔画像による個体識別という共通のタスクであるため、このモデルによる手法は高い精度が期待される。もう1つは、一般物体認識用に学

習されたモデルである。これは様々な物体認識の転移学習に用いられるモデルであり、1 つ目のモデルとの比較のために用いる。これらのモデルを特徴抽出器として用いて、牛の正面顔画像による個体識別に対する転移学習の有効性の評価を行う。

2. 転移学習による牛の個体識別

2.1 転移学習

転移学習とは、あるタスクについて学習済みのモデルがある場合に、関連する新規のタスクに対して学習済みのモデルを利用する手法の総称である。大規模なデータセットを用いて学習された畳み込みニューラルネットワーク(CNN)の出力は汎用性が高く、非常に強力な特徴抽出器として機能することが知られており、少ない学習データで高い性能を実現可能にする^[3]。

本研究では、学習済みのモデルを特徴量抽出器として利用し、識別器への転移学習を行う。学習済みモデルから得られる高品質な特徴量を識別器の学習に利用することで、少ない学習データで高精度な識別器の学習が期待される。

2.2 識別処理の流れ

本研究では、画像中から牛の正面顔画像を切り出したものを学習済みモデルの入力に用いて顔の特徴量を抽出する。抽出した特徴量を正規化し、識別器の学習データとして与えることで転移学習を行う。学習された識別器に対して未知の正面顔画像から得られた特徴量を与え、どのクラスに属するか判定をすることで、個体識別を行う。処理の流れを図1に示す。

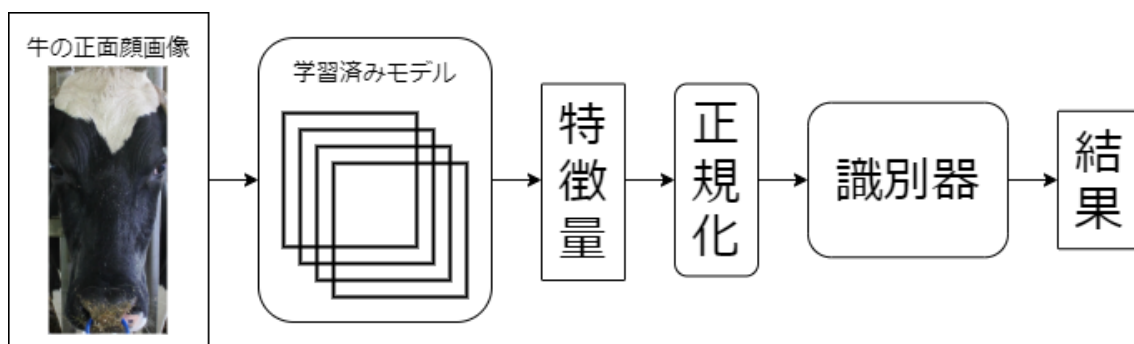


図 1 識別処理の流れ

2.3 特徴量の抽出

本研究では特徴抽出器として、FaceNet^[2]、VGGFace^[4]、VGG16^[5]、VGG19^[5]、ResNet50^[6]の 5 つの学習済みモデルを用いる。FaceNet と VGGFace は人の顔を学習したモデルである。牛の場合でも顔画像による個体識別という共通のタスクであるため、高い精度が期待できる。VGG16 と VGG19、ResNet50 は一般物体認識用に ImageNet^[7]という大規模データセットによって学習されたモデルである。これら 3 つのモデルは様々な物体認識の転移学習に用いられており、人の顔用モデルとの比較のために用いる。これらのモデルから、FaceNet は特徴抽出用に設計された層の出力、他の 4 つは全結合層の直前のプーリング層の出力を特徴量として用いる。各特徴量の次元数は、FaceNet と VGG16、VGG19 が 512 次元、VGGFace と ResNet50 が 2048 次元である。これらのモデルに対して、図 2 のような牛画像から図 3 のように牛の正面顔を切り取ったものを入力として与え、上記の各層での出力を特徴量として抽出する。得られた特徴量は平均が 0、分散が 1 になるように正規化を行う。



図 2 牛画像

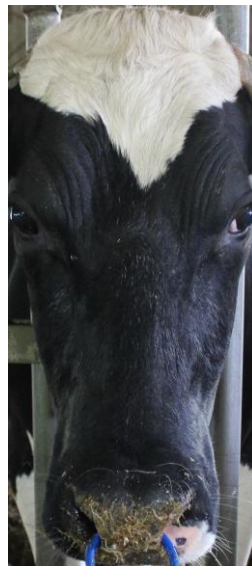


図 3 切り出した牛の正面顔画像

2.3 識別器による判定

得られた特徴量を用いて個体識別を行う。識別器には、サポートベクターマシン(SVM)を用いる。SVMは、未知の入力に対して高い識別性能を得ることが出来る優れた識別器の1つである。一般にSVMは2値分類を行うが、多クラス分類への拡張を行うために、一対他分類^[8]を導入する。これはクラス数がC個ある場合に、正解クラスとその他のクラスの2値分類を行う識別器をC個学習する。推論時にはすべての識別器に予測を行わせ、正解であるという予測確率が最も高いクラスを最終的な予測結果とすることで、多クラス分類を実現する。学習データの特徴量でこの識別器の学習を行い、テストデータの特徴量がどのクラスに属するかを分類し、個体の識別を行う。

3. 実験

3.1 実験 1 精度評価

それぞれの学習済みモデルによる転移学習の性能を評価するために、まず各モデルから得られた特徴量を用いた顔認識の精度評価を行う。

3.1.1 実験データ

実験に用いるデータは、乳牛 24 頭、肉牛 27 頭の 51 頭に各 10 枚の計 510 枚の顔画像を用いる。この顔画像は図 3 のように事前に手作業で切り出したものであり、個体ごとに撮影日時や撮影を行った機材が違う画像が含まれている。切り出した画像の例を図 4, 5 に示す。図 4, 5 から、乳牛には模様があるものが多く、肉牛には模様がなく、個体ごとの特徴が少ないことが分かる。



図 4 乳牛の正面顔画像



図 5 肉牛の正面顔画像

3.1.2 評価方法

各個体 10 枚のうち 6 枚を学習データ, 残りの 4 枚をテストデータとしてランダムに選択する交差実験を 5 回行い, その識別精度の平均値を Rank4 までの累積識別精度特性(CMC)曲線で示す. CMC 曲線は, ある未知の入力が識別器に与えられたとき, 予測確率の高い順に並べられた中に正解クラスが各順位までに含まれる割合を示したグラフである. Rank4 であれば, 4 番目までに正解が含まれる割合を示す. このとき, 乳牛だけの場合と肉牛だけの場合, すべての個体が含まれる場合の 3 つの結果を示す.

3.1.3 実験結果

表 1, 2, 3, 図 6, 7, 8 に乳牛, 肉牛, 全個体に各学習済みモデルを適用して得られた特徴量を用いた識別精度を示す. 表 1 と図 6 が乳牛, 表 2 と図 7 が肉牛, 表 3 と図 8 が全個体の結果である.

表 1 乳牛における各モデルの精度(%)

Rank	人の顔用モデル		一般物体認識用モデル		
	FaceNet	VGG Face	VGG16	VGG19	ResNet50
1	82.7	78.8	97.5	96.3	96.0
2	93.8	91.7	99.2	98.8	99.0
3	96.7	96.0	99.6	99.6	99.2
4	97.7	96.7	99.6	99.8	99.2

表 2 肉牛における各モデルの精度(%)

Rank	人の顔用モデル		一般物体認識用モデル		
	FaceNet	VGG Face	VGG16	VGG19	ResNet50
1	59.4	57.4	92.8	91.9	91.7
2	71.1	72.4	96.7	95.9	96.9
3	78.5	80.2	98.1	97.6	98.3
4	83.0	85.7	98.5	98.1	99.1

表 3 全個体における各モデルの精度(%)

Rank	人の顔用モデル		一般物体認識用モデル		
	FaceNet	VGG Face	VGG16	VGG19	ResNet50
1	71.0	67.9	96.0	95.6	95.1
2	82.8	81.6	98.1	97.9	97.8
3	87.4	87.5	98.5	98.5	98.6
4	89.7	91.0	99.0	98.9	98.9

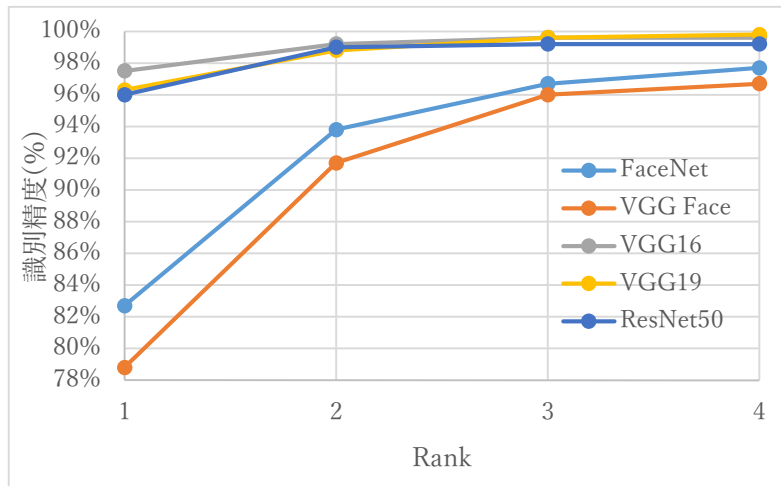


図 6 乳牛における各モデルの CMC 曲線

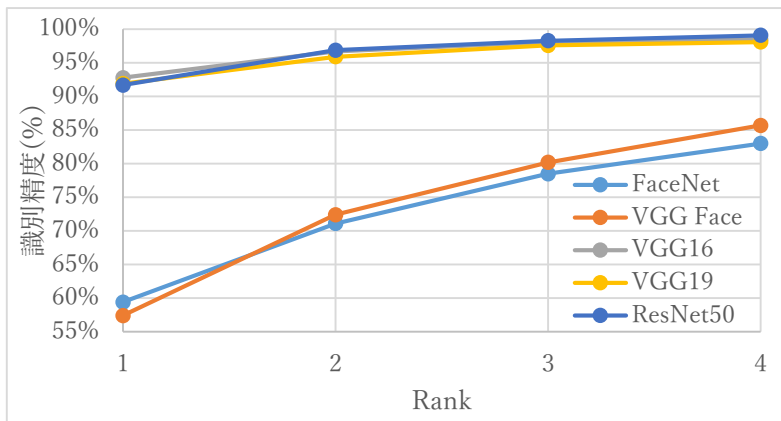


図 7 肉牛における各モデルの CMC 曲線

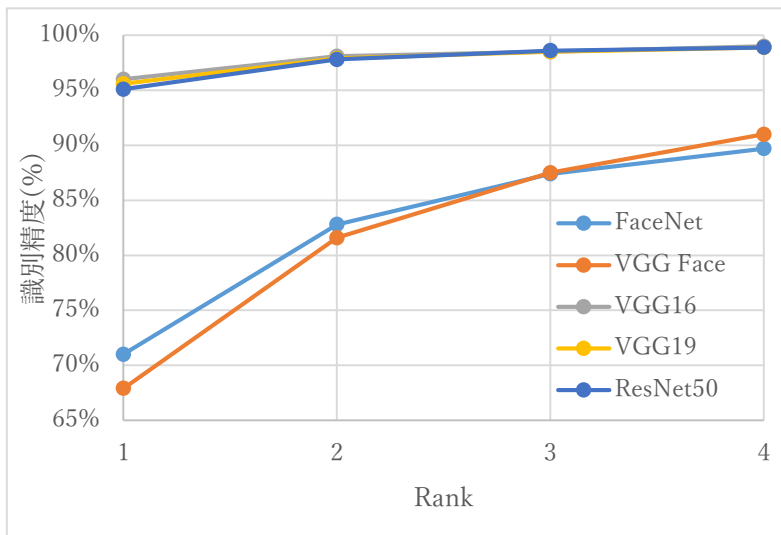


図 8 全個体における各モデルの CMC 曲線

表 1, 2, 3 と図 6, 7, 8 より, 各モデルの結果を比較する. まず, 人の顔用モデルを見ると, いずれの場合でも明らかに精度が低くなっていることが分かる. 乳牛のみの場合ではある程度の精度が出ているが, 肉牛に関しては FaceNet, VGGFace のどちらのモデルも Rank1 識別精度が 60%を下回っており, 全個体で識別を行った場合の結果に大きく影響している. これは, 人の顔と牛の顔の形状が大きく異なることによって, モデルが牛の顔の特徴を正しく捉えられていないためと考えられる. 乳牛であれば模様がある個体が多いのである程度の特徴量の差が出ているが, 模様の全くない肉牛ではその差がほとんど出ず, 精度に差が出たと考えられる. これに対して, 一般物体認識用のモデルはどれも高い精度を示している. 乳牛の場合では識別精度が 96%を上回り, 最も高いものでは 97.5%を示している. 肉牛の場合でも 92%程度の高い精度を示している. 最終的な全個体での結果もすべて 95%を上回っており, これは人の顔用モデルの Rank4 識別精度を上回る結果となっている.

この結果より, 牛の顔認識のタスクにおいては, 人の顔用モデルを転移学習するよりも, 一般物体認識用モデルを転移学習した方が高い精度を見込めることが分かった.

本研究と先行研究^[1]は実験に用いた画像や頭数, 実験方法などの違いから直接的に比較するのは難しいが, 先行研究での識別精度は最大 82.48%であり, 性能的にも本研究の方が優れていると考えられる.

3.2 実験2 ヒートマップによる確認

実験1より各モデルの精度については分かったが、これらのモデルが実際に牛の顔の特徴を捉えているかどうかは分からない。そこで各モデルが画像中のどの部分に注目しているか調べる。人の顔用モデルと一般物体認識用モデルを比較するために、VGGFace と VGG16, VGG19, ResNet50 の中間層の出力を用いてヒートマップを作成する。FaceNet はモデルの構造が非常に複雑であったため、ヒートマップの作成が難しく、本実験では使用しない。

3.2.1 評価方法

牛の顔画像をそれぞれのモデルに入力として与え、各モデルの中間層の出力から得られたヒートマップを比べて評価を行う。このヒートマップは赤に近い部分ほど強く反応していることを表しており、モデルが画像中のどの部分に注目しているかを調べることができる。CNN はチャンネル数が多く、すべてを表示するのは難しいため、ヒートマップは平均強度が強い順に左上から右下にかけて6つ表示する。

3.2.2 実験結果

図9~16に乳牛、肉牛を各モデルに入力して得られたヒートマップを示す。図9, 10はVGGFace, 図11, 12はVGG16, 図13, 14はVGG19, 図15, 16はResNet50の結果である。

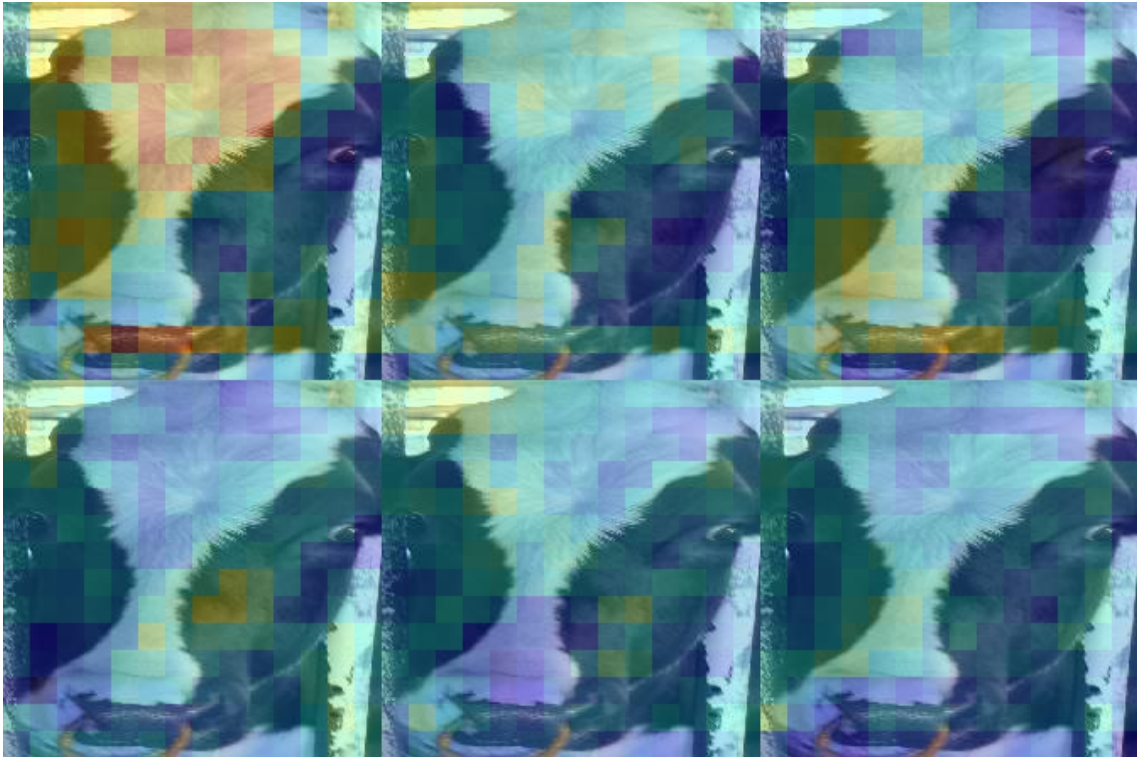


図 9 VGGFace による乳牛のヒートマップ

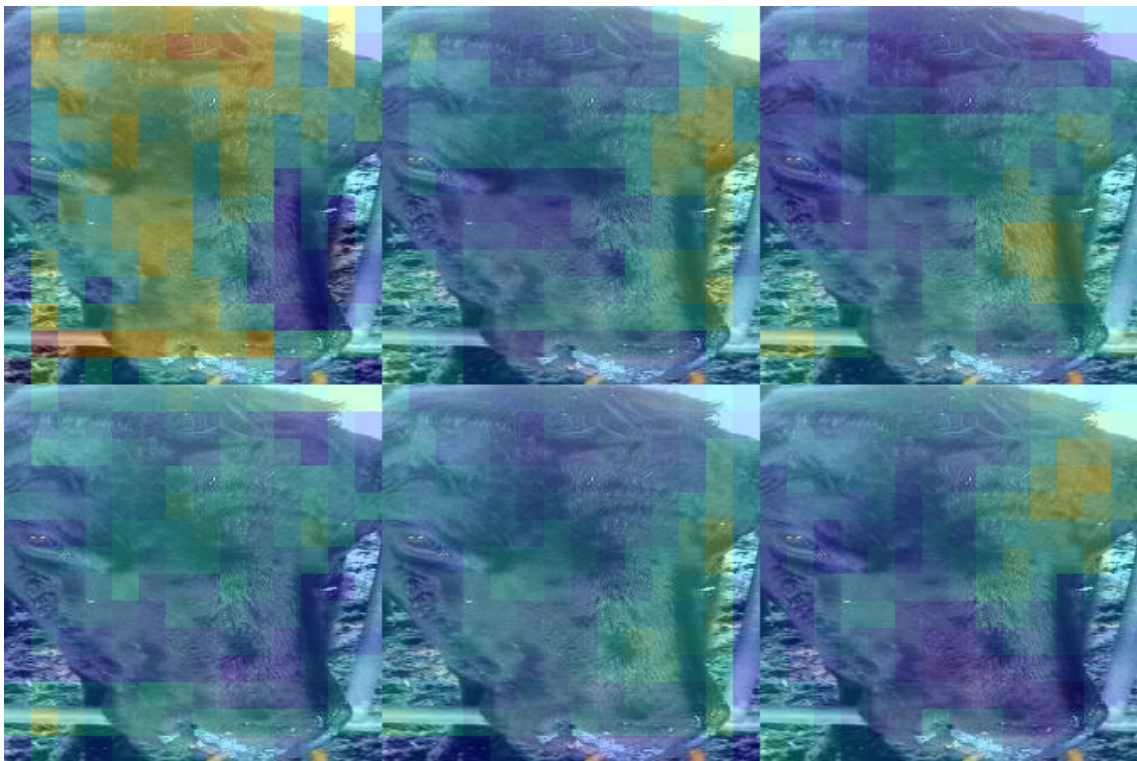


図 10 VGGFace による肉牛のヒートマップ

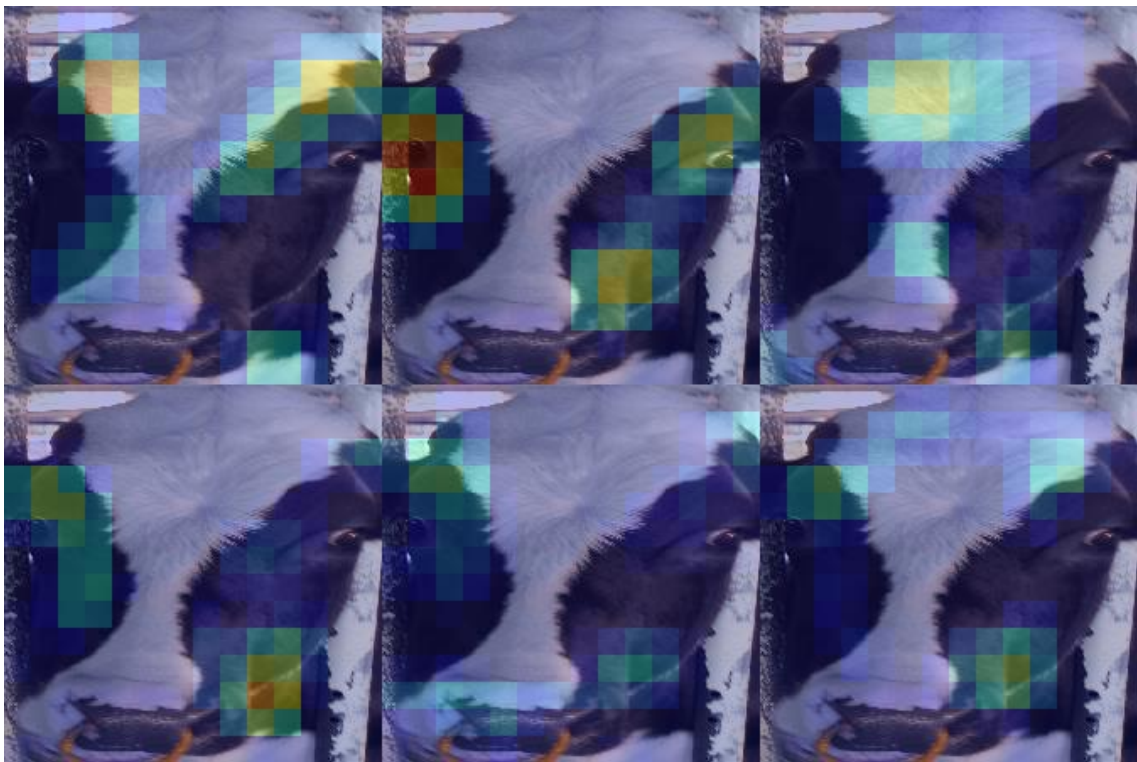


図 11 VGG16 による乳牛のヒートマップ

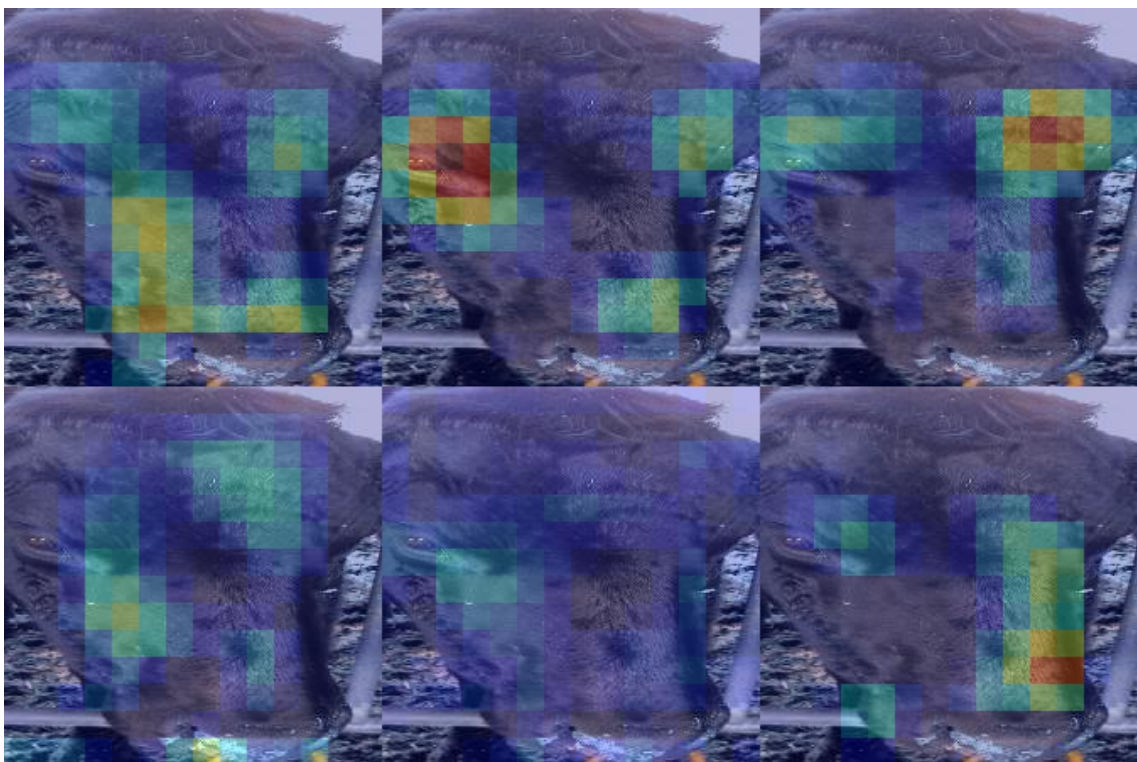


図 12 VGG16 による肉牛のヒートマップ

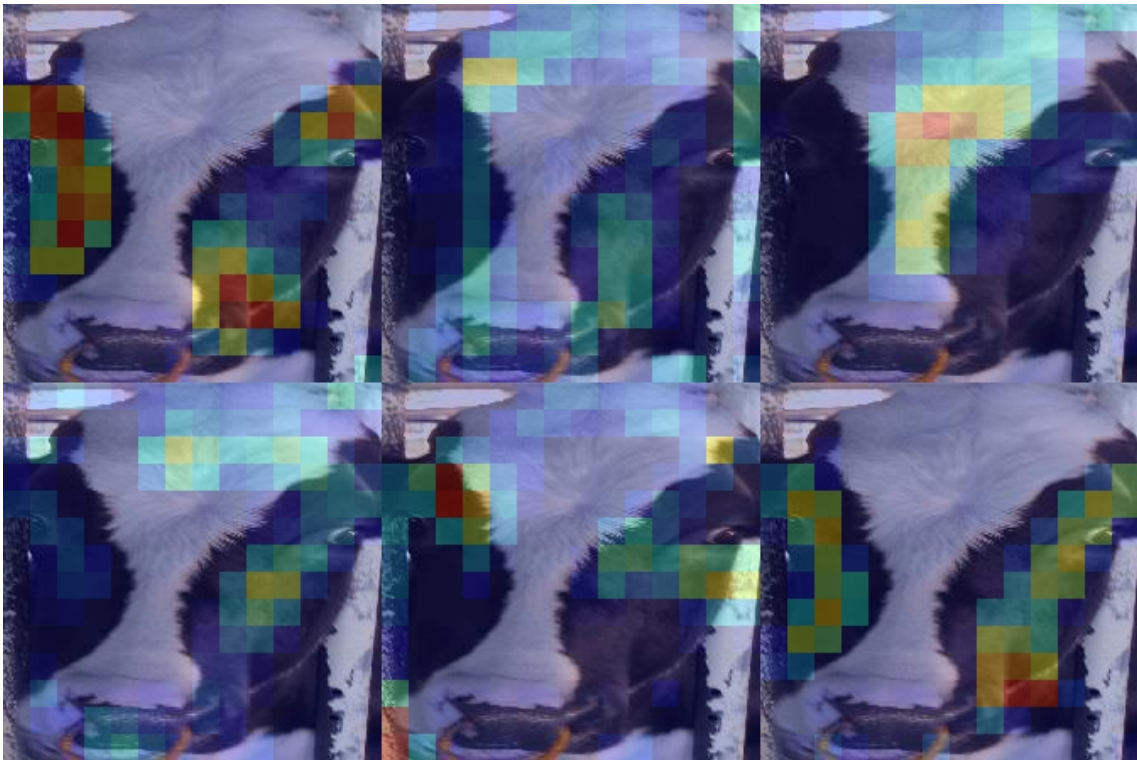


図 13 VGG19 による乳牛のヒートマップ

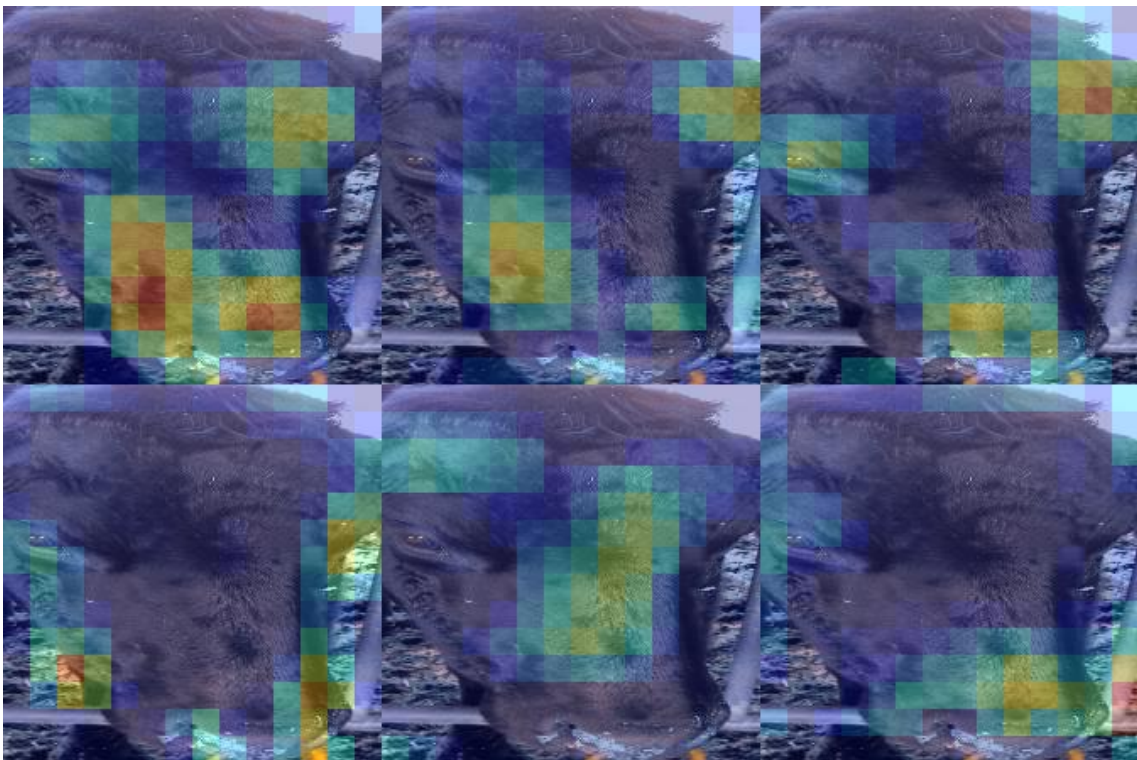


図 14 VGG19 による肉牛のヒートマップ

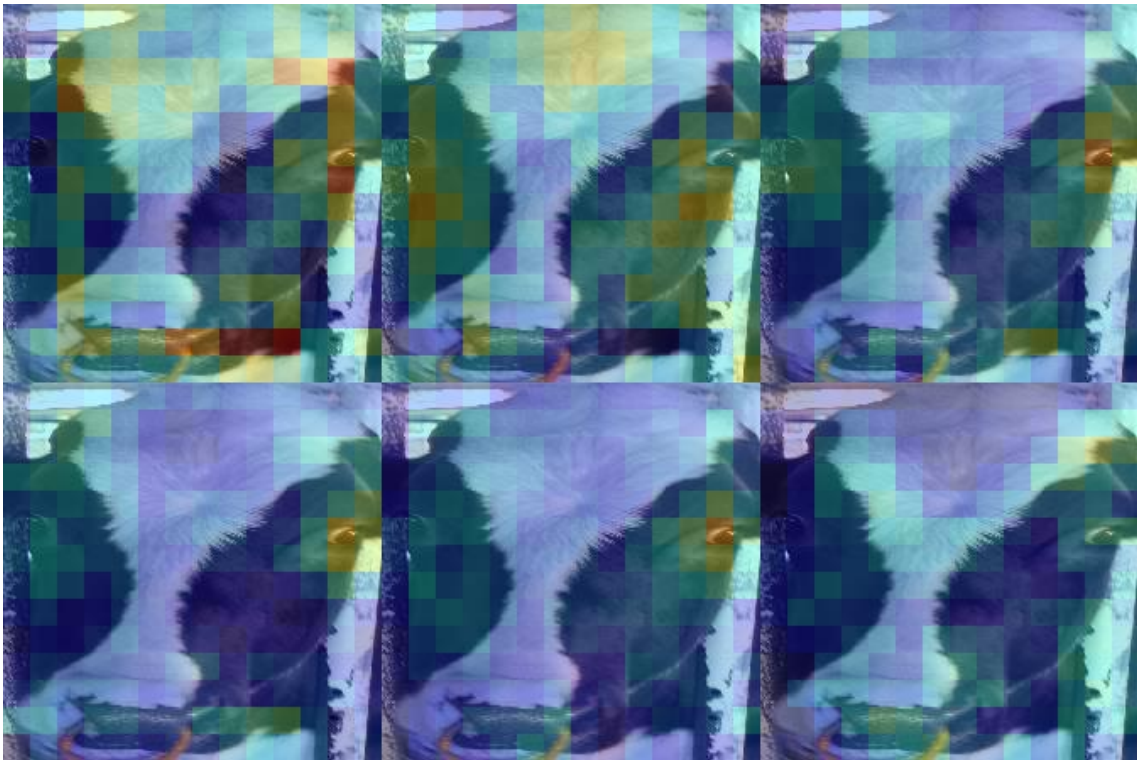


図 15 ResNet50 による乳牛のヒートマップ

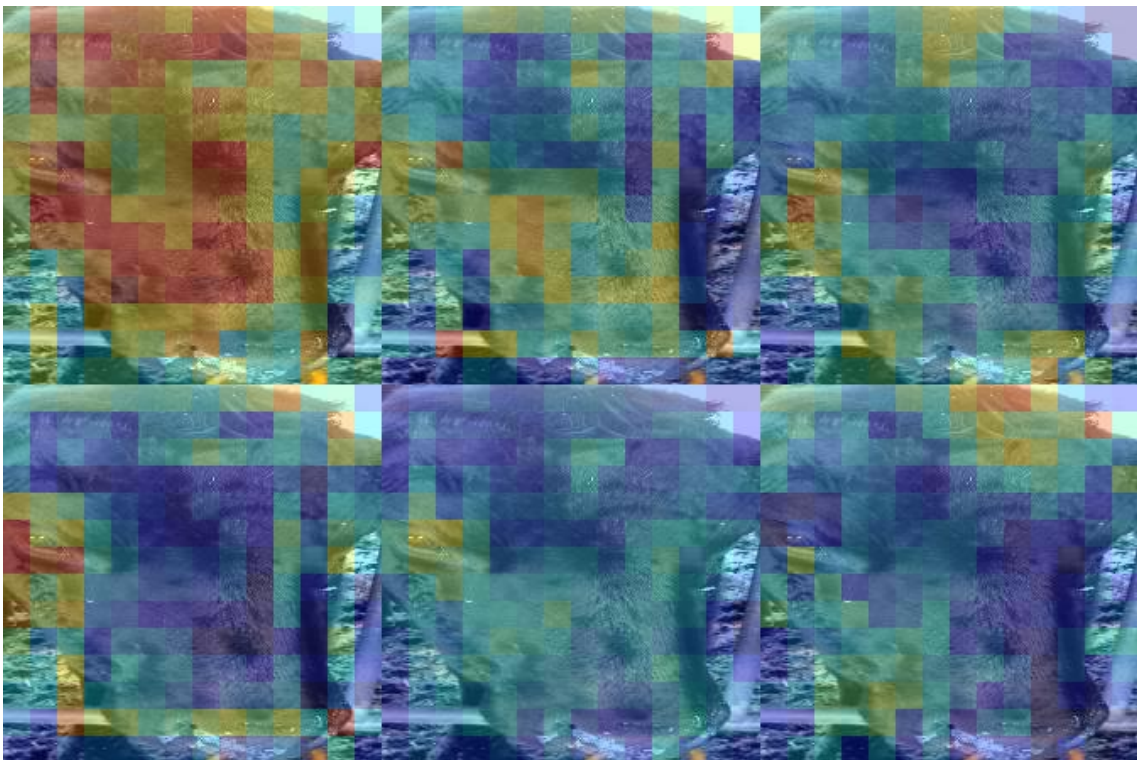


図 16 ResNet50 による肉牛のヒートマップ

図 9, 10 の結果を見ると、背景部分を含んだ全体に反応しており、モデルが牛の顔の中で、どの部分を注目しているかほとんど分からない。この結果から、実験 1 で述べたとおり、牛の顔が人の顔と比べて形状が大きく異なることで、牛の顔の特徴を正しく捉えられていないと考えられる。それに対して、図 11, 12, 13, 14 の結果を見ると、乳牛であれば目や頭部の毛並み、顔の模様といった部分を、肉牛では目や毛並み、目から鼻にかけての形状など、特徴となりそうな部分に注目していることが分かる。図 15, 16 の結果では、前の 4 つに比べて全体的に反応しているものの、目や顔の模様、毛並みなどにより強く反応している。また、人の顔用モデルと比べて背景よりも牛の顔の領域に強く注目している。これらの結果から、一般物体認識用モデルは牛の特徴をうまく捉えられており、その結果、高い精度に繋がったと考えられる。また、肉牛と乳牛のどちらにおいても目に強く反応しているものがあり、目が個体識別を行う際の特徴として重要である可能性が考えられる。

この結果より、人間の顔用モデルでは牛の顔には適用が難しく、一般物体認識用モデルは牛の顔から特徴をうまく捉えることができていたことが分かった。

3.3 実験3 人の顔画像を与えた場合の検証

人の顔用モデルが牛の顔の特徴を捉えることが難しいことが実験2で分かった。このとき、人の顔用モデルに人の顔画像を用いた場合、どの部分に注目するかを調べることで、本当にこのモデルが人以外には不適切なのかを検証する。比較のため、人の顔画像を VGGFace に入力して得られたヒートマップを示す。

3.3.1 実験データ

The Japanese Female Facial Expression (JAFFE) データセット^[9]の中から1枚を選出した。

3.3.2 評価方法

人の顔画像を入力として VGGFace に与え、実験2と同様にヒートマップを作成し、牛の顔から作成したヒートマップと比較する。

3.3.3 実験結果

図17に人の顔を VGGFace に入力して作成したヒートマップを示す。

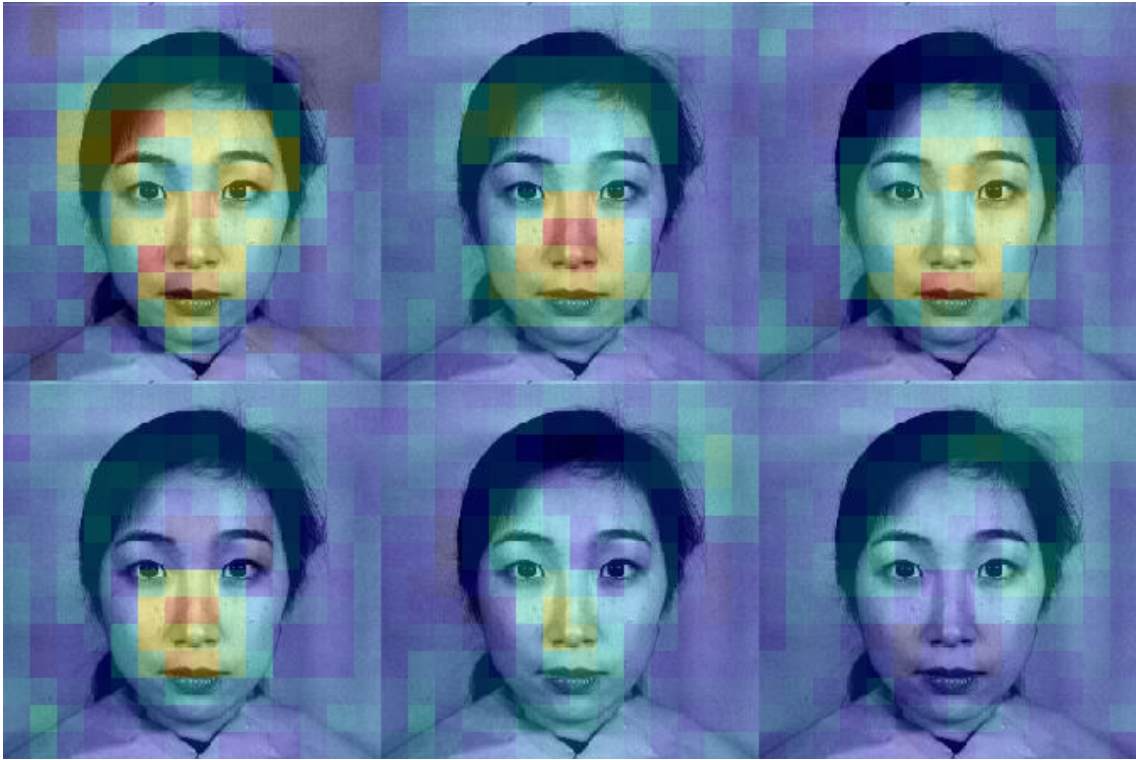


図 17 人の顔から作成したヒートマップ

図 17 の結果を見ると、人の顔から口や鼻、頬の周りや顔全体など、特徴となりそうな部分に注目していることが分かる。実験 2 の図 9, 10 の結果と比較すると、人の顔の場合では顔の特徴を正しく捉えられていることが分かる。これらの結果から、人の顔用モデルは人の顔の場合には顔の特徴を正しく捉えられているといえる。また同時に、人の顔とは形状が大きく異なる場合には、その特徴をうまく捉えられないことも分かった。

4. おわりに

本研究では、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の検証を行った。当初の想定では、人の顔用モデルは顔画像による個体識別という共通のタスクを行うので、高い精度が出ると予想していた。しかし、実験の結果、人間とは形状の大きく異なる牛の顔に対しては人の顔用モデルはうまくその特徴を捉えられず、精度を出すことができないことが分かった。それに対し、一般物体認識用モデルは、肉牛乳牛問わず、非常に高い精度を示した。全個体を使用した場合の精度は 95%以上であり、これは人の顔用モデルの識別精度を大きく上回る結果となった。結論として、牛の正面顔画像による個体識別に転移学習を用いる場合、使用する学習済みモデルは一般物体認識用モデルを利用した方が良いことが分かった。

今後の課題としては、まず、実験用の個体数を増やす必要があると考えている。先行研究^[1]の実験では計 300 頭の牛に対して個体識別を行っている。本研究では、精度測定の実験に計 51 頭の牛の正面顔画像を用いたが、先行研究と比べ個体数が十分でなく、個体数をより増やした場合に精度にどの程度影響が出るか実験を行う必要がある。また、本研究で人の顔用モデルをそのまま牛の顔に適用することは難しいことが分かったが、FaceNet のようなネットワーク構造を持つモデルに対して、牛の顔を大量に学習させることで、牛の顔用モデルを作成することができれば、より高い精度が得られる可能性がある。また、本研究では、画像中の牛の顔の切り取りを人の手で行っているが、最終的に自動で切り取りを行う必要があるため、切り取り用のシステムを別途作成する必要がある。

謝辞

本研究を作成するにあたり，丁寧で熱心なご指導を頂いた指導教員の
椋木雅之教授に感謝いたします。また，本研究のきっかけを与えてくださ
った農学部獣医学科の関口敏准教授，牛の顔画像を撮影するためにご協
力頂いた農学部附属フィールド科学教育研究センターの小林郁雄准教授
と工学部教育研究支援センター技術専門職員の高塚佳代子さん，同研究
室の兒玉光平さんに感謝いたします。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] Santosh Kumar, Shrikant Tiwari, Sanjay Kumar Singh, “Face Recognition of Cattle: Can it be Done?”, National Academy of Sciences, India Section A: Physical Sciences, 2016, pp 137-148.
- [2] Florian Schroff, Dmitry Kalenichenko, James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, Computer Vision and Pattern Recognition, 2015, pp 815-823.
- [3] 中山 英樹, “深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習”, 電子情報通信学会技術報告 SP, 115(146), 2015, pp.55-59.
- [4] O. M. Parkhi, A. Vedaldi, A. Zisserman, “Deep Face Recognition”, British Machine Vision Conference, 2015.
- [5] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, International Conference on Learning Representations, 2015.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Microsoft Research, “Deep Residual Learning for Image Recognition”, Computer Vision and Pattern Recognition, 2016, pp770-778.
- [7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei, “ImageNet: A large-scale hierarchical image database”, Computer Vision and Pattern Recognition, 2009, pp248-255.
- [8] W. Wei, G. Xiaorong, G. Shangkai, ”One-Versus-the-Rest(OVR) algorithm: An extension of common spatial patterns(CSP) algorithm to multi-class case” Engineering in Medicine and Biology Society, 2005, pp. 2387-2390.
- [9] <https://zenodo.org/record/3451524>, 2020 年 1 月 28 日にアクセス.