

| | |
|------|---|
| タイトル | 深層学習によるネットワークカメラ画像を対象としたシマフクロウの検出と評価 |
| 著者 | 長谷川, 大; HASEGAWA, Dai; 早矢仕, 有子; HAYASHI, Yuko |
| 引用 | 北海学園大学工学部研究報告(50): 65-76 |
| 発行日 | 2023-01-13 |

深層学習によるネットワークカメラ画像を対象とした シマフクロウの検出と評価

長谷川 大*・早矢仕 有子*

Evaluation of Detecting Blakiston's Fish Owls from Network Camera Images Using Deep Learning

Dai HASEGAWA* and Yuko HAYASHI*

概要

シマフクロウの生態調査・保護活動の一環として、生息域に設置されたネットワークカメラの画像を市民に公開する取組を進めている。このネットワークは赤外線センサによる動体検出を行って撮影する他に、15分に1度の間隔で定期的に撮影する仕組みになっている。しかしながら、日々増え続ける画像のうち、シマフクロウが撮影されている画像はごくわずかであり、公開する画像の取捨選択に手間を要している。そこで本研究は、深層学習による画像認識技術を適用して、ネットワークカメラの画像を対象にシマフクロウが撮影されている画像を判別することを試みる。本研究では2019年1月から2020年7月の期間に撮影された画像をフクロウあり画像（12,938枚）とフクロウなし画像（138,539枚）に分類したデータセットを用いた。本データセットに対して、一般的な画像分類タスクにおいて高性能な事前学習済みのVGG19モデルを用いて学習および分類を行った。実験の結果、システムがフクロウありと判定した画像のうち94.6%が実際にフクロウが写っている写真であったことから、高い適合率であることが確認できた。またテストしたフクロウあり画像のうち89.2%を正しくフクロウありと判定することができており、適合率、再現率ともに高性能な判定モデルを構築することができた。

1 はじめに

かつて北海道全域に広く分布していたシマフクロウは、生息環境消失に伴い分布域縮小と個体数減少が続いたが、1984年からの国の保護事業により次第に回復の兆しが現れ、現在は東部を中心に約165羽の生息が確認されている¹。シマフクロウを保全するために、国は生息地を特

*北海道大学工学部生命工学科

*Department of Life Science and technology, Faculty of Engineering, Hokkai-Gakuen University

定できる情報を非公開とし、生息地への人の立入を防ごうとしてきた。しかし、インターネットの普及により生息情報の拡散が加速したことでカメラマンやバードウォッチャーの生息地への入り込みが急増し、シマフクロウへの悪影響が懸念されている⁸⁾。そこで、シマフクロウ営巣地に遠隔監視システムを構築することで繁殖巣や採餌場所への人の接近を抑止するとともに、画像を公開配信することで市民の目でシマフクロウの子育てを見守る取り組みを進めてきた(図1)⁶⁾。見守る目の数が増えることで監視効果が高まるとともに、シマフクロウへの関心や知識が高まる効果も期待できる。

画像は、シマフクロウが繁殖に利用している巣箱に2台、採餌場所に1台を設置したネットワークカメラから配信した。巣箱に設置したカメラの1台は巣の天井から巣底を撮影し、もう1台は巣の天井から巣の入口を撮影した。どちらも赤外線暗視型カメラであり、昼夜を通しての画像閲覧が可能であった。このネットワークカメラは赤外線センサーによる動体検出を行って撮影する他に、15分に1度の間隔で定期的に撮影する仕組みになっている。日々増え続ける画像のうち、シマフクロウが撮影されている画像はごくわずかであり、公開する画像の取捨選択に手間を要している。そのため、シマフクロウが写っている画像を自動的に分類することができれば労力の削減になる。

一方、コンピュータを用いたパターン認識・画像認識の分野では、2012年に発表されたニューラルネットワークを用いた深層学習技術が、一般物体認識タスクの精度を飛躍的に向上させた³⁾。このブレイクスルー以降、深層学習技術は大きく注目され、これまでにさらなる発展を見せている。そこで本研究は、深層学習による画像認識技術を利用して、ネットワークカメラの画像を対象にシマフクロウが撮影されている画像を判別することを試みる。本研究では2019年1月から2020年7月の期間に撮影された画像をフクロウあり画像(12,938枚)とフクロウなし画像(138,539枚)に分類したデータセットを用いた。本データセットに対して、一般的な画像分類タスクにおいて高性能な事前学習済みのVGG19モデル⁴⁾を用いて学習および分類を行う。

2 先行研究

本研究が分類対象とする画像はアングルが固定された画像であるため、シマフクロウが写っている部分を除けばどの画像も変化は少ない。そのため、深層学習を使わずに比較的シンプルなルールベースの画像処理によってシマフクロウが写っている画像を特定できる可能性があった。先行研究⁷⁾では、シマフクロウが写っていない画像と分類対象画像の差分をみることでシマフクロウが写っている画像を検出する方法の検討が行われている。先行研究では、以下の判

⁴⁾環境省北海道地方環境事務所：シマフクロウ保護増殖事業「シマフクロウの概要」, https://hokkaido.env.go.jp/post_9.html, 2022年1月。

定アルゴリズムが提案されている。

まずフクロウの写っていない画像40枚（以降，基本画像）を選択した。このとき，積雪がある季節や，昼夜によって大きく画像に変化が見られることから，夏の昼，夏の夜，冬の昼，冬の夜の4つの条件から各10枚ずつ選んだ。これらの画像はグレースケールに変換した。



図 1：シマフクロウ一家の見守り日記（Webサイト）

次に，判定を行う画像もグレースケール化し，基本画像との画素値差分の総和を計算する。1枚の判定対象画像に対する40枚の基本画像との画素差分40組から最小値を得る。これは条件の異なる基本画像のなかから最も類似の場面を選択するためである。この画素差分が閾値以上である場合にシマフクロウが写っていると判定し，閾値を下回っている場合はシマフクロウが写っていない画像と判定する。

閾値は，次のように求めた。40枚の基本画像に対して1枚のランダムに選択したシマフクロウが写っていない画像との差分を計算し，40組の画像差分から最小値を得る。これをシマフクロウが写っていない画像を変更して，40回繰り返す，それぞれ最小値を得る。40個の最小値の中の最大値を閾値とした。この方法で得られた閾値はシマフクロウが写っていない画像同士を比較した際でも生じる画素差分の最大値に近似されると考えられる。何か異物が写っていた場合には，これよりも大きな画素差分が得られると期待される。

シマフクロウが写っている写真100枚とシマフクロウが写っていない写真100枚の合計200枚をテスト画像として選定し，上記のアルゴリズムに従って判定を行った結果，システムがフクロウありと判定した画像は91枚，フクロウなしと判定した画像は109枚であった。フクロウあ

表1：データセットの概要

| | 採餌場所 | 巣の中からみた入口 | 天井からみた巣底 | 計 |
|----------|---------|-----------|----------|----------|
| フクロウあり画像 | 3,328枚 | 1,467枚 | 8,143枚 | 12,938枚 |
| フクロウなし画像 | 51,327枚 | 51,028枚 | 36,184枚 | 138,539枚 |
| 計 | 54,655枚 | 52,495枚 | 44,327枚 | 151,477枚 |

りと判定された画像の中で実際にフクロウが写っている正しい画像は60枚であり、またフクロウなしに判定された画像の中で実際にフクロウが写っていなかった画像は69枚であった。

この結果は、大量の画像からシマフクロウが写っている画像を選定するには不十分であると言える。特に、フクロウありと判定した画像のうち実際にフクロウが写っていた写真は65%であり、35%はシマフクロウが写っていない画像であるため、やはり人手による選定が必要になってしまう。また、実際にフクロウが写っていた写真のうち、40%を見逃してしまっていることも問題と言える。このように判定精度が向上しなかったのは、同一アングルの写真であっても光の当たり方、季節の変化などによって予想以上に画像毎の変化は多様であり、ルールベースによる場面分けをすることが難しかったことが要因と考えられる。本研究では深層学習によって、人手では設定できない量のパラメータを持つ判定モデルを作成することにより、より高精度な判定システムの構築を目標とする。

3 方法

3.1 データセット

2015年～2020年7月までにシマフクロウの生息域で3つのアングル（採餌場所、巣の中からみた入口、天井から見た巣底）から撮影された約44万枚の画像のうち、2019年1月から2020年7月までに撮影された3つのアングル（採餌場所、巣の中からみた入口、天井から見た巣底）の画像、151,477万枚を使用した。

先行研究⁷⁾により、北海学園大学工学部生命工学科の学生1名がこれらの画像に対してシマフクロウが写っている画像（以後、フクロウあり画像）とシマフクロウが写っていない画像（以降、フクロウなし画像）に分類を行っている。分類の結果、フクロウあり画像が12,938枚、フクロウなし画像が138,539枚であった。各アングル毎の分類結果を表1、および画像例を図2に示す。

3.2 ネットワークアーキテクチャ

図3に、本研究で用いたネットワーク構成を示す。本研究ではImageNetデータベース¹⁾により事前学習されたVGG19ネットワーク⁴⁾を転移学習することで分類モデルを作成する。VGGネットワークは、深い畳み込み層をもったネットワーク構成になっており、高い画像認識能力を持つことが確認されている。またImageNetは1,400万枚以上の画像からなる一般画像認識用



図 2：データセット内の画像例

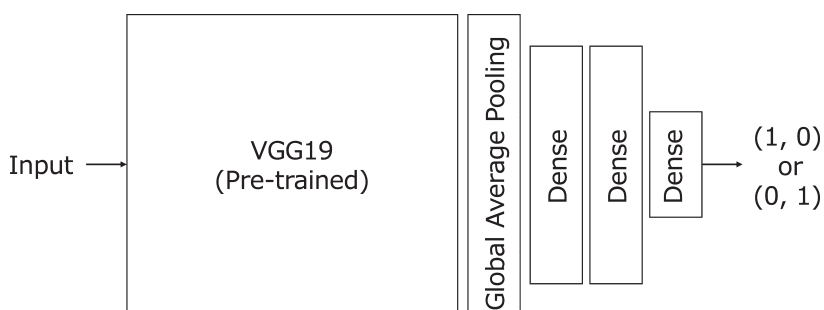


図 3：Network Architecture

のデータセットで、20,000カテゴリの画像がある．このImageNetで画像分類を学習済みのVGG19ネットワークの出力層側に、Global Average Pooling 2D層、および全結合層が3層の計4層を追加した．また学習にあたってはVGG19ネットワークの重みパラメータは固定として、追加した4層のレイヤのパラメータを学習によって変更する．追加した4層には、過学習を避けるためBatch Normalization²⁾と5%Dropout⁵⁾を適用している．また追加層の活性化関数は、中間層ではReLU関数、最後の出力層ではsoft-max関数を用いた．

表 2：データスプリット

| | 学習用データ | 検証用データ | テスト用データ |
|----------|----------|---------|---------|
| フクロウあり画像 | 10,352枚 | 1,293枚 | 1,293枚 |
| フクロウなし画像 | 110,833枚 | 13,853枚 | 13,853枚 |
| 合計 | 121,185枚 | 15,146枚 | 15,146枚 |

3.3 学習

対象としたデータセットのデータをすべてシャッフルし、トレーニングデータ：検証データ：テストデータを8：1：1になるよう分割した．表 2 に分割後の画像数を示す．データ分

割の結果，学習に用いるデータは121,185枚，検証およびテストに用いるデータがそれぞれ15,146枚となった．ここで，検証用データとは学習中に学習の進行を確認するために使用するデータであり，テスト用データとは学習後に分類精度を確認するために使用するデータとなる．

3.3.1 データ拡張

学習用データを使用して学習する際に，データジェネレータによるデータ拡張を行った．このデータジェネレータは学習用のデータから画像を取り出す際に，RGBカラー画像を各色0から1の範囲にリスケールし，また画像の拡大縮小，回転，平行移動，反転を行う．学習用のデータにはシマフクロウが写っている画像が10,352枚あるが，画像に収められなかったシマフクロウの姿や写り方はより多様であったことは想像に難くない．このデータ拡張は，シマフクロウが写っている画像に限りがあっても，擬似的に様々なバリエーションのシマフクロウの画像を生成して学習させることを意図している．

3.3.2 データバランシング

本研究で使用する画像データは，フクロウあり画像が12,938枚に対してフクロウなし画像が138,539枚と10倍以上の枚数がある．一般的に，正例と負例のトレーニングデータ量に偏りがある場合は，適切な学習結果が得られない可能性があることが知られている．そのため，フクロウなし画像を間引き，データ量の偏りを是正した条件での実験も行うこととする．

具体的には，トレーニングデータのフクロウあり画像とフクロウなし画像の割合が，「1：10.7（全て）」，「1：8」，「1：5」，「1：2」の条件で実験を行う．データ量を変更した条件での画像数を表3に示す．割合を変更していない「1：10.7」条件では，フクロウなし画像が138,539枚であるのに対して，「1：8」条件では82,816枚，「1：5」条件では，51,760

表3：データバランシング

| 条件 | 1：10.7 | 1：8 | 1：5 | 1：2 |
|----------|----------|---------|---------|---------|
| フクロウあり画像 | 10,352枚 | 10,352枚 | 10,352枚 | 10,352枚 |
| フクロウなし画像 | 110,833枚 | 82,816枚 | 51,760枚 | 20,704枚 |

表4：実験結果

| | VGG19 | VGG19 (1：8) | VGG19 (1：5) | VGG19 (1：2) |
|---------|-------|-------------|-------------|-------------|
| 適合率（あり） | 0.989 | 0.984 | 0.960 | 0.946 |
| 適合率（なし） | 0.984 | 0.985 | 0.987 | 0.990 |
| 再現率（あり） | 0.829 | 0.834 | 0.864 | 0.892 |
| 再現率（なし） | 0.999 | 0.999 | 0.997 | 0.995 |
| F値（あり） | 0.902 | 0.902 | 0.910 | 0.918 |
| F値（なし） | 0.992 | 0.992 | 0.992 | 0.993 |

枚, 「1 : 2」条件では20,704枚となった. フクロウなし画像を間引く際にはランダムに画像を削除し, 特定の角度や特定の時期の画像ばかりが削除されないよう配慮した.

3.3.3 学習の停止

上述のネットワークおよび学習条件で学習を行い, 学習の進行状況として学習データと検証データにおける損失および正答率を確認した. 「1 : 2」条件における, エポック毎の学習の様子を図4に示す. データジェネレータを用いていることでデータは無制限に拡張可能であるため, ここではランダムに選ばれた1,024枚の画像による学習を1エポックとした. また検証データも各エポックで1,024枚をランダムに用いて損失と正答率を計算している. 図4から, エポックが進む毎に損失が減少し, 学習が進行していることがわかる. しかしながら, 正答率は20エポックを過ぎたあたりから高い水準で横ばいに推移し, これ以上の学習は過学習となることが予想されたことから, 全ての条件で100エポックで学習を打ち切った.

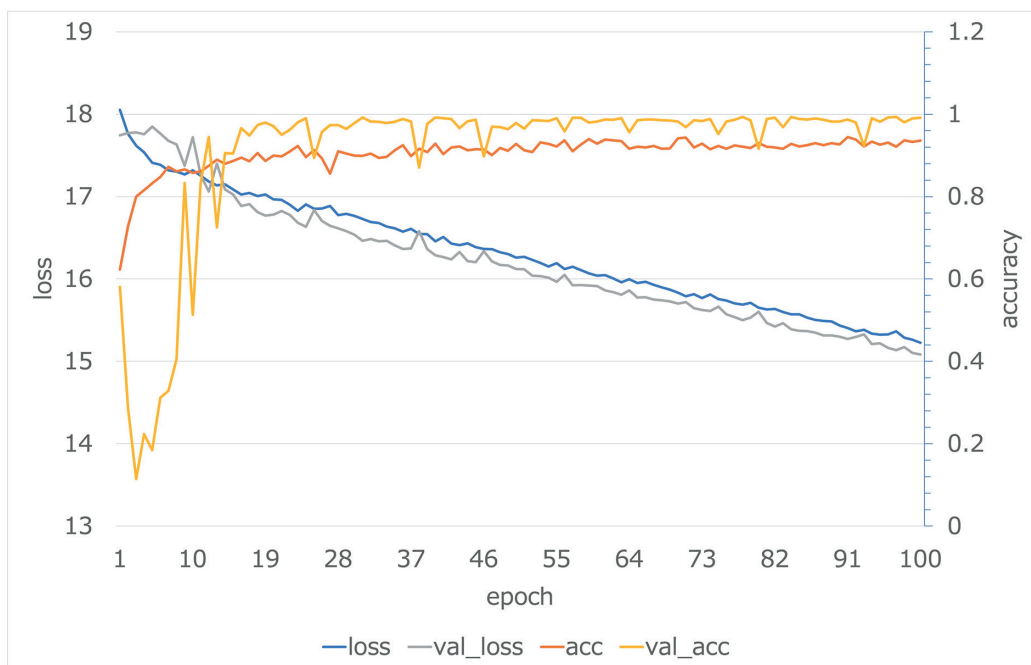


図4：学習過程

4 結果

実験の結果を表4に示す. ここで正例(「フクロウあり」クラス)に対する適合率 (precision), 再現率 (recall), F値 (f-value) は以下の式で計算される.

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$f\text{-value} = \frac{2 \times precision \times recall}{precision + recall}$$

ただし、TP, FP, FN, TNは次の通りとする。

- True Positive (TP) : 真の値が正例のものに対して、正例と予測したもの
- False Positive (FP) …真の値が負例のものに対して、正例と予測したもの
- False Negative (FN) …真の値が正例のものに対して、負例と予測したもの
- True Negative (TN) …真の値が負例のものに対して、負例と予測したもの

一般に、適合率と再現性はトレードオフの関係にあるため、総合的な性能はF値で評価することになる。また、本研究の課題では、シマフクロウあり画像を網羅的かつ誤りなく検出することが目標になるため、とくに正例に対するF値が重要となる。表4より、フクロウあり画像となし画像の割合が「1 : 2」条件でF値(あり)が0.918と最大を示しており、適合率と再現率のバランスが最も良かった条件であった。またデータ量を調整していない条件では適合率(あり)が0.989と最大であったが、再現率が0.829と低下した。

5 考察

フクロウあり画像となし画像の割合が「1 : 2」条件でF値(あり)が0.918と最大となった。この条件では適合率(あり)が0.946であり、フクロウありの判定をしたときには94.6%の精度で正しい判定を行っていることがわかる。一方で、再現率(あり)は0.892であり、フクロウあり画像が入力されたときに10%程度見逃してしまうという結果になった。本研究課題では大量の画像からフクロウあり画像を選定して一般に公開することを目標としているが、適合率が高いほどフクロウなし画像を誤って提示することが少なくなく、また再現率が高いほどフクロウあり画像を見逃してしまう割合が少ない。本研究で得られた判別モデルは十分に実用可能な性能を有していると考えられる。

また、フクロウなし画像を間引いた実験結果から、やはりフクロウあり画像となし画像のデータ量を均等に近づけることで適合率と再現率のバランスが良くなりF値が向上することが確認できた。しかしながら、フクロウなし画像を間引くことで適合率は若干低下することから、より適合率が重要な場面やタスクでは間引く前のモデルを利用するほうが良いと言える。

表5に、テスト用データ(フクロウあり画像1,293枚、フクロウなし画像13,853枚)の判定結果の詳細を示す。表5に示されているように、フクロウあり画像1,293枚のうち1,154枚を正

しく判定できており，そのうち採餌場所の画像が234枚，巣の中の画像が137枚，巣の天井の画像が783枚であった．**図5**，**図6**，**図7**に採餌場所，巣の中，巣の天井の角度において正しく判定できたフクロウあり画像の例を示す．いずれの角度においてもシマフクロウの位置や姿勢の違い，昼夜などの光量の違いなどに関わらず，正しい判定ができていることが確認できる．

表5：VGG19（1：2）の実験結果

| | 判定あり | 判定なし |
|----------|--------|---------|
| フクロウあり画像 | 1,154枚 | 139枚 |
| フクロウなし画像 | 66枚 | 13,787枚 |



図5：シマフクロウの検出成功例（採餌場所）



図6：シマフクロウの検出成功例（巣の中）

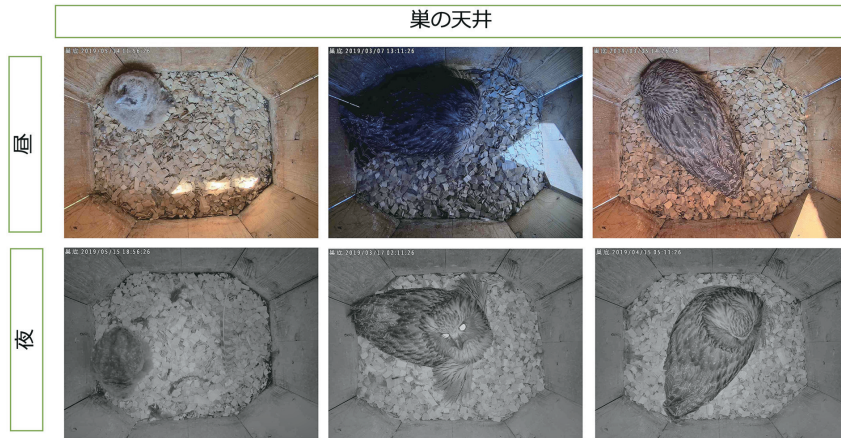


図7：シマフクロウの検出成功例（巢の天井）



図8：シマフクロウの検出失敗例（正解あり判定なし）

一方で、テストデータのフクロウあり画像139枚に対して誤った判定が行われた。図8に、正解ありの画像に対して判定をなしとしてしまった画像の例を示す。判定を失敗した139枚のうち、採餌場所の画像が117枚で最も多く、巢の中の画像が5枚、巢の天井の画像が17枚であった。図8に見られるように、小さくシマフクロウが写っている画像やシマフクロウの体の一部しか写っていない画像などで見落としがあり、フクロウなしと判定してしまうケースが存在することが確認できた。特に採餌場所アングルだけに限定してみるとフクロウあり画像を見逃す傾向が高いと言える。

また、フクロウなし画像のうち66枚に対して、誤った判定を行なっている。判定を失敗した66枚のうち、採餌場所の画像が42枚、巢の中の画像が3枚、巢の天井の写真が21枚であった。図9に正解なしの画像に対して判定ありとした画像例を示す。図9から、葉や降雪、局所的な光が写っている画像に判定ミスが見られた。また少数ではあったが、巢の中の画像には元デー

タの分類ミスで実際にシマフクロウの体の一部が写っている画像が混入していたため、判定ミスに数えられていることがあった。

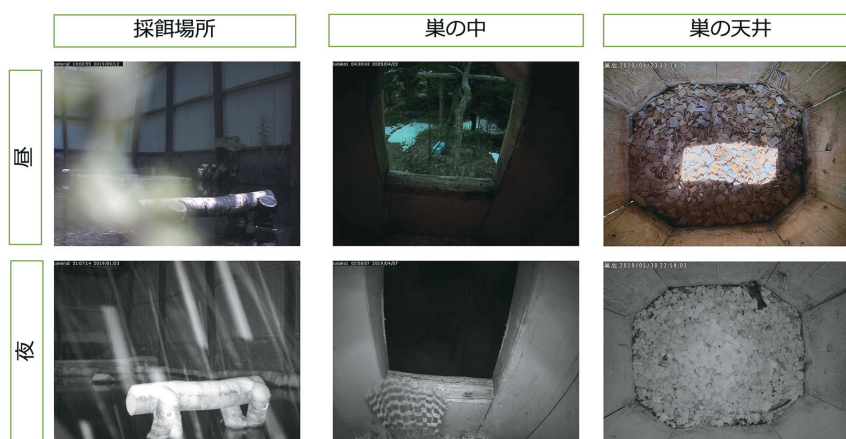


図9：シマフクロウの検出失敗例（正解なし判定あり）

6 おわりに

本研究では、シマフクロウの生態調査・保護活動の一環として生息域に設置されたネットワークカメラで撮影された多量の画像からシマフクロウが撮影されている画像を判別することを目的とし、深層学習による画像認識技術を利用した判定モデルを作成した。フクロウあり画像とフクロウなし画像に分類された151,477枚の画像データのうち、121,185枚を学習用データとし、事前学習されたVGG19ネットワークを転移学習した。また、フクロウあり画像に対してフクロウなし画像のデータ量が極端に多かったことから、フクロウなし画像を間引いたデータによる実験も行った。実験の結果、フクロウあり画像となし画像の割合が「1：2」のとき、適合率（あり）が0.946、再現率（あり）が0.892、F値（あり）が0.918ともっとも精度が向上した。今後は現在開発が進められているWebシステムに判別モデルを組み込み、自動判別機能の実装を行っていきたい。

謝辞

本研究は令和2年度北海学園大学共同学術助成、三井物産環境基金2012年度研究助成、同2015年度研究助成、および2016年度旭硝子財団研究助成を受けた。また、(株)構研エンジニアリング、シスコン(株)、(株)土谷製作所、(株)NTTドコモ北海道支社の技術協力を受けた。

参考文献

- 1) Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet : A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp.248–255. Ieee, 2009.
- 2) Sergey Ioffe and Christian Szegedy. Batch normalization : Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp.448–456, 2015.
- 3) Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol.25. Curran Associates, Inc., 2012.
- 4) Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv : 1409.1556*, 2014.
- 5) Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol.15, No.1, pp.1929–1958, 2014.
- 6) 小島望 (編) 高橋満彦 (編) 畠山武道 (監). 野生動物の餌付け問題, 第12章シマフクロウへの給餌と餌付け (早矢仕有子), pp.191–206. 地人書館, 2016.
- 7) 望月透. 自動撮影カメラによる野生動物画像データセットの整備と分析. 北海学園大学工学部生命工学科 2020年度卒業論文, 2021.
- 8) 早矢仕有子. 絶滅危惧種シマフクロウを対象とした写真撮影者の特性および観光利用における問題点. 保全生態学研究, p.2036, 2022.