

害獣自動認識のためのアライグマ画像データベースの構築と 深層学習による認識の試行

Construction of a Raccoon Image Database and a Trial of Deep Learning for Vermin Recognition

初田慎弥¹・大野真史¹・泉知諭²・孟林³

Shinya Hatsuda, Masashi Ono, Tomonori Izumi and Lin Meng

¹立命館大学大学院 理工学研究科 (〒525-8577 滋賀県草津市野路東1-1-1)

Graduate Student, Ritsumeikan University, Graduate School of Science and Engineering

²立命館大学教授 理工学部電子情報工学科 (〒525-8577 滋賀県草津市野路東1-1-1)

Professor, Ritsumeikan University, Dept. of Electronic and Computer Engineering

³立命館大学助教 理工学部電子情報工学科 (〒525-8577 滋賀県草津市野路東1-1-1)

Assistant Professor, Ritsumeikan University, Dept. of Electronic and Computer Engineering

This report gives construction of an image database of racoons and a trial to apply the deep learning technique to vermin recognition. Photographs and movies of racoons and some other similar animals are taken in a zoological facility and then clipped and labeled with the aid of GUI tool developed by ourselves to construct an image database including 11,600 racoons, 2,700 racoon dogs, and 21,500 backgrounds. A trial of deep learning with the database is conducted utilizing an open-source software framework "Caffe". Our experiment shows that the system achieves about 90% of recognition rate.

Keywords: image recognition, machine learning, deep learning, convolutional neural network, vermin, racoon

1. はじめに

アライグマは北米原産の生物であり日本において過去ペットとして人気があったが、現在それらが野生化して様々な問題を引き起こしている。その中の一つに文化財の破壊がある。屋根裏に侵入する、糞尿で汚染するといった被害が報告されている¹⁾。文化財などへの被害を防ぐため最近では不審者を検知するとサイレン等を発報する防犯システムがある。防犯対策としては人間と動物を誤判別することで不要な誤報が多発する。しかし、現在ではこのようなシステムは動物を人間と見誤っての誤報が多く管理者にシステムを停止されてしまうことも少なくない。さらに害獣対策としては建造物に害を為す動物（特にアライグマ）とそれ以外の実害の無い動物の区別をつけることが求められる。本研究は害獣自動認識により監視カメラシステムを高度化し文化財を害獣被害から防ぐことを目的としている。また、アライグマの繁殖状況や分布などの分析にも、フィールドカメラと害獣自動認識技術の応用が期待される。

画像認識にはパターンマッチ、SVM(Support Vector Machine)やニューラルネットワークなどの機械学習が提案されている。パターンマッチはある程度形が決まっている文字など幾何学図形に適している²⁾。機械学習はコンピュータが何かしらの入力を受けてプログラムの内部状態を変更し、適応していく過程のことを言う。機械学習では人間が見つけた特徴や法則性ではなく、コンピュータが見つけた特徴や法則性を主として処理する。そのため特徴、法則性を人間がプログラム化しづらい場合や人間では判別できない場合などの時に機械学習は人間を超える効果を発揮することがある。ILSVRC2014³⁾ という物体認識のコンテストで畳み

込みニューラルネットワーク(Convolutional Neural Network:CNN)という手法⁴⁾が提案されそれまでにない高い性能を示したことで注目を集めている。畳み込みニューラルネットワークを中核とした機械学習は深層学習(deep learning)と呼ばれる。本稿は、深層学習の害獣認識への適用の試行を報告する。

機械学習は大量の学習用サンプルを必要とし、自動認識システムの構築にはまず画像を収集し、画像とラベル(属性、正解)の組からなるデータベースを構築しなければならない。ペットや家畜として一般的な動物については公開された機械学習用のデータベースが見られるが、アライグマやそれに似たタヌキ、アナグマなどの動物については報告者らが知る範囲では大規模なデータベースは見受けられない。そこで、害獣を含む野生動物のデータベースの構築も併せて進めていく。本稿ではその初期試行を報告する。まず鳥獣飼育施設にて動画・写真を撮影した。画像切り出しおよびラベル付けを支援するソフトウェアツールを開発し、手作業によりアライグマ11,600枚、タヌキ2,700枚、背景21,500枚のデータベースを構築した。

次に、畳み込みニューラルネットワークの処理を行うプログラムが必要となる。本稿ではディープラーニングのフレームワークであるCaffe⁵⁾を用いる。CaffeはBerkeley Vision and Learning Center (BVLC)で開発されたオープンソースソフトウェアであり、畳み込みニューラルネットワークに基づく深層学習を行うフレームワークである。使用するデータベース、ネットワークの構造、学習スケジュールなどの設定を与えることで、深層学習およびその評価を実行する。今回は、畳み込み層とプーリング層それぞれ3層、全結合層2層の畳み込みニューラルネットワークにバッチサイズ100、イタレーション5000で学習させた。構築したデータベースの計35,800枚のうち、学習に32,500枚、テストに3,300枚を使用して評価したところ93%の認識率を達成した。

計算負荷の高い学習部分を事前に行い、その際の畳み込みニューラルネットワークの構造と学習後のパラメータを利用することで将来的に文化財を害獣被害から防ぐシステムへの活用を見込める。

2. データベースの構築

機械学習には大量の学習データセットが必要である。学習データセットは多ければ多いほど対象の特徴を汎化することができ、未知のデータにも対応することができるようになる。学習度合いはネットワークの構造や識別対象などによって左右され、適切な量を定めることは難しい。公開され広く参照されている事例を見ると、0から9の手書き数字の認識用に公開されているMNIST⁶⁾では合計60,000枚、動物や飛行機など10のカテゴリを持つCIFAR-10⁷⁾では合計50,000枚用意されている。また、学習が十分に汎化されているか、過学習を起こしていないかを確認するためのテストデータセットも大量に必要であり、こちらはMNIST、CIFAR-10ともに10,000枚用意されている。どちらもカテゴリ数は10であり、1カテゴリごとに数千単位でデータセットが用意されている。

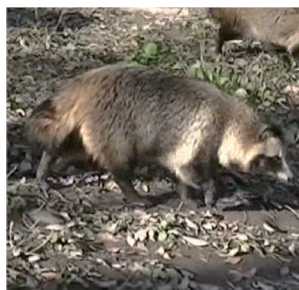
(1) 飼育施設での撮影

データセットの元となる画像を得るために害獣を動画で撮影する。今回はアライグマを認識することが主たる目的であるので、アライグマとアライグマに似ているタヌキ、アナグマ、またアライグマと同じく外来生物であるハクビシン(図1左)を撮影の対象とした。撮影にあたっては、前後左右斜め、俯瞰、仰望など様々な方向からの、さらに文化財に侵入する状況を考慮して柱を登る、穴に入るなど様々な体勢の画像を撮ることとした。これを実現するために角材とペット用の玩具(図1右)(製品名:DCMわくわくツリーハウス5段)、誘導用の餌を用いた。関西野生生物研究所の支援を得て弘前市農業政策課の協力の下、(財)弘前市みどりの協会 弥生いこいの広場にて2016年11月14日13:00~16:00に撮影を行った。天候は晴れであった。関西野生生物研究所のアライグマ研究者によると、本来アライグマは夜行性であるが昼間も行動し、昼夜の行動に違いはないように思われるとのことである。

撮影には2台の家庭用ビデオカメラ(JVC GZ-HM670, JVC GZ-E565)を用いた。2台とも動画形式はAVCHD、画素数は1920×1080、フレームレートは30fpsである。アライグマは、撮影者2名の手持ちにより、飼育場の柵内および柵外(上からの俯瞰)から撮影した。施設の許可を得て、ペット用玩具や誘引用の餌を用いた。また、タヌキほかの動物を撮影した。これらの動物については飼育場に立ち入ることができず、柵外からの撮影となった。



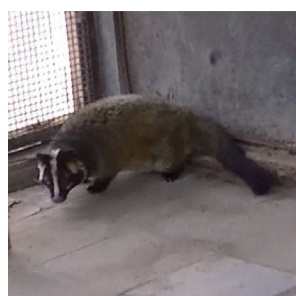
アライグマ



タヌキ



アナグマ



ハクビシン



図1 撮影した動物とペット用玩具によるアライグマ撮影の様子

(2) 画像の切り出しとラベル付け

まず動画から静止画を抽出する。ここでは既存のツールを用いて以下の手順で行った。Any Video Converter フリー版⁸⁾を用いて、AVCHD形式からXvid⁹⁾コーデックによるAVI形式に変換する。そしてNiVE2¹⁰⁾という画像ツールを用いてPNG形式の静止画を抽出する。

次に画像から対象物の画像を切り出し、ラベルを付けてデータセットを作成する。その作業を支援するために、GUI画像アノテーションツールを開発した。Microsoft Visual Studio 2015を用いC#で開発した。このツールでは、フォルダ内の大量の静止画から必要なものを選び、ラベルを設定して対象物の領域を矩形でマークすることができる。開発したツールによる切り出しラベル付け作業の様子を図2に示す。GUIツールによる指定後、OpenCVを利用したPythonのスクリプトで画像を切り出し、64×64サイズに正規化してPNG形式で保存し、また画像ファイル名とラベルの組のリストを保存する。

撮影条件と作業時間の都合により、今回は撮影画像の中からアライグマ、アライグマとよく似ており目撃頻度も高いタヌキ、そして背景の画像に注力した。作成した画像はアライグマ11,600枚、タヌキ2,700枚、背景21,500枚の3種類で、これを学習用32,500枚とテスト用3,300枚に振り分けた。画像の順番が時系列のままでは似た画像が続いてしまうため、剰余を用いてシャッフルした。作成した画像の例を図3に示す。また、作成したデータセットの枚数を表1にまとめる。

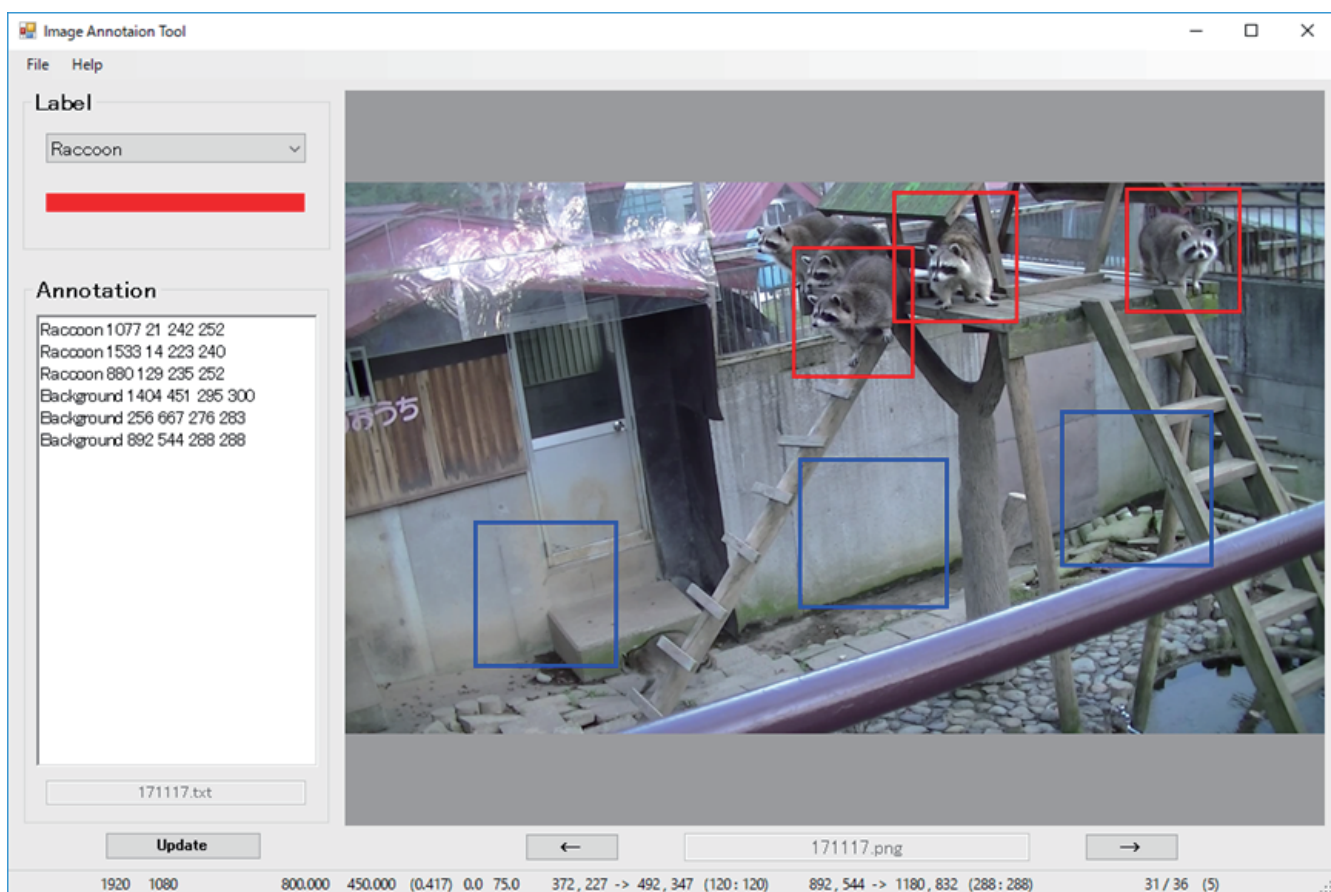


図2 作成したGUI画像アノテーションツール



図3 切り出した画像の例（左からアライグマ、タヌキ、背景）

表1 作成したデータセットの内訳

	アライグマ	タヌキ	背景	合計
学習用	10000 枚	2500 枚	20000 枚	32500 枚
テスト用	1600 枚	200 枚	1500 枚	3300 枚

3. 害獣分類器の構築と評価

ここではオープンソースの深層学習フレームワークであるCaffeを用いて害獣分類器を構築し評価する。画像認識の応用としては、生物種のグループ（例えば害獣が否か）の分類、生物種の分類、個体の分類などが考えられる。最も詳細な個体の分類は学習用データベースの構築が困難であり、少なくとも生物種が分類

できればグループの分類も可能となる。害獣認識の第一歩として、ここでは一般的な生物種毎の分類を試みる。

(1) ネットワーク構成

害獣分類器の畳み込みニューラルネットワークの構成はCaffeに参照モデルとして用意されているCIFAR-10_quickを基にする。これは自動車、猫など十種類の対象に対する画像分類器である。害獣分類器のネットワーク構造を図5に示す。CIFAR-10_quickに倣い、畳み込み層とプーリング層はそれぞれ3層とする。畳み込み層のフィルタサイズはすべて 5×5 とし、ストライドは1、パディングは2とする。プーリング層は1層目だけ最大値を算出し、2、3層目は平均値を出力する。畳み込み層の2、3層目とプーリング層の1層目の出力にはReLU関数を適用する。全結合層の1層目のユニット数は64とする。今回はアライグマ、タヌキ、背景の3種類に分類するため、出力層となる全結合層の2層目はユニット数を3に変更する。

(2) アライグマデータセットを用いた学習

学習およびテストの各種パラメタも基本的にCIFAR-10_quickに倣う。学習時およびテスト時のバッチサイズはともに100とする。テスト時のバッチ処理の回数はデータ数にあわせて33(=テスト画像3,300枚/バッチサイズ)に変更する。学習画像が32,500枚あるので、バッチの325イタレーションで学習データが一巡する。

(3) 分類器の評価

分類器の学習の様子を図4に示す。横軸(iteration) はバッチのイタレーション回数、縦軸(accuracy)は正答率である。正答率はアライグマ、タヌキ、背景をまとめた値であり、破線 train accuracy は学習データ（各バッチ無作為均等分配100枚）に対する正答率、実線 test accuracy はテストデータに対する正答率である。また、表2に正答率の抜粋を示す。5000イタレーションの学習により最終的にtrain accuracyは97%に、test accuracyは約93%となった。test accuracyの向上は2,000イタレーション程度（32,500枚のテスト画像およそ六巡程度）で概ね収束していることが分かる。その後もtrain accuracyは多少上下するが、これはテストが常に同じ3,300枚に対する正答率なのに対し、学習は順に変わるバッチ100枚に対する正答率であるためと思われる。

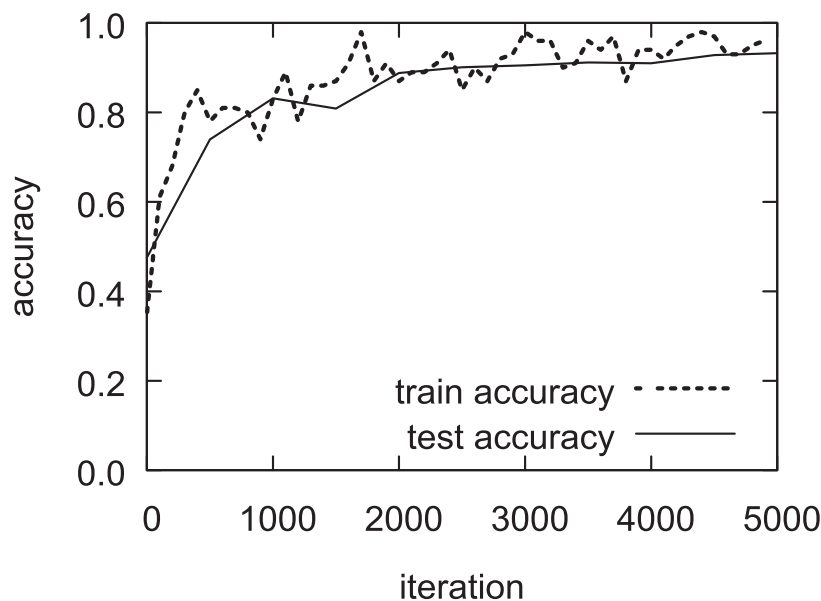


図4 構築した害獣分類器の正答率の変化

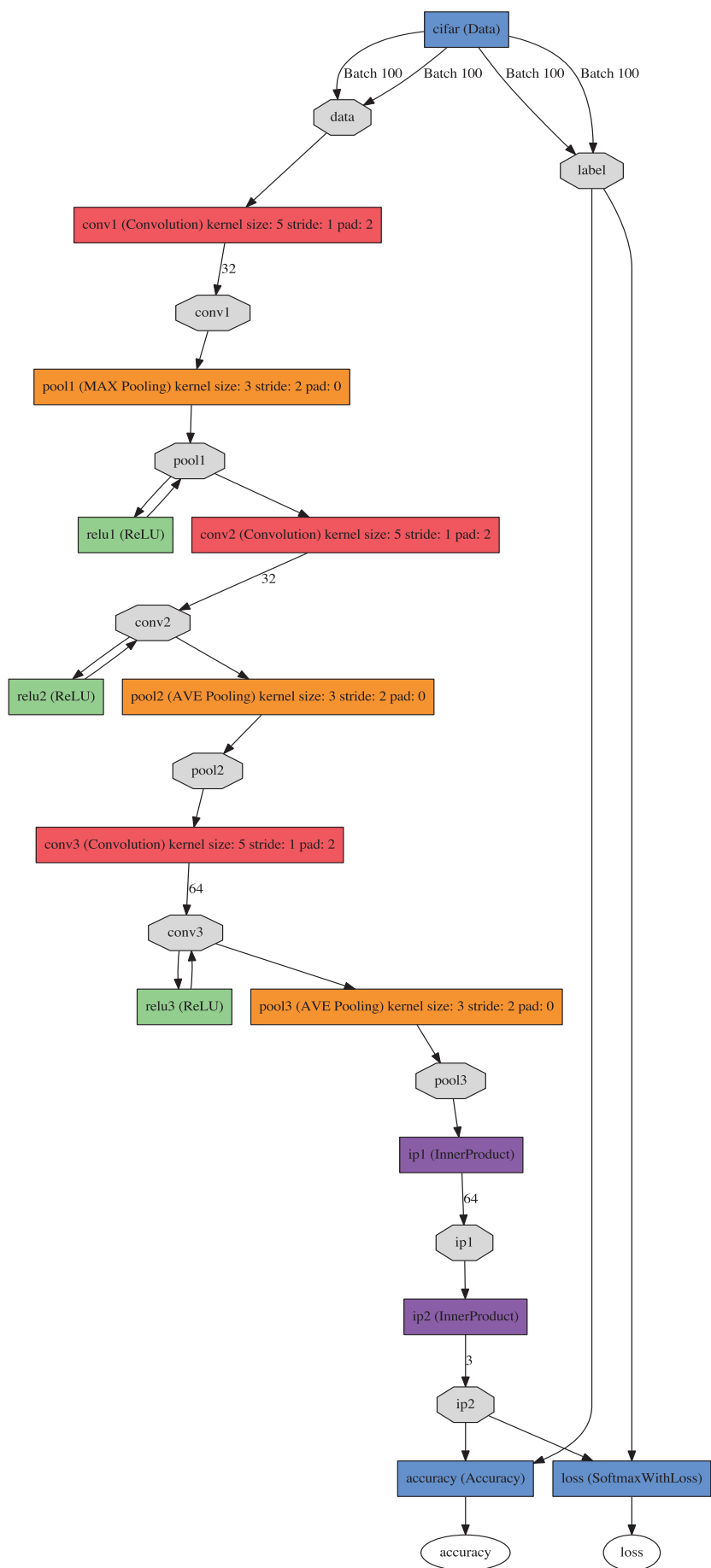


図5 害獣分類向け畳み込みニューラルネットワークの構造

表2 構築した害獣分類器の正答率の変化の表（抜粋）

iterations	train accuracy	test accuracy
0	0.35	0.48
500	0.78	0.74
1000	0.83	0.83
1500	0.87	0.81
2000	0.87	0.89
2500	0.85	0.90
3000	0.98	0.91
3500	0.96	0.91
4000	0.94	0.91
4500	0.97	0.93
5000		0.93

最終的なtest accuracyの内訳を表3に示す。2～4行目は入力画像の種類を示し、2～4列目はそれらに対する最終的な分類器の回答を示す。右下がりの対角線が正答数を、それ以外が誤答数となる。5列目は本人拒否率(False Reject Rate: FRR)を示し、5行目は他人受入率(False Accept Rate: FAR)を示す。

表3 テスト用画像に対する正誤回答数の内訳

回答 入力画像	アライグマ	タヌキ	背景	FRR
アライグマ	1465枚	16枚	119枚	0.084
タヌキ	16枚	177枚	7枚	0.115
背景	58枚	7枚	1435枚	0.043
FAR	0.048	0.115	0.081	

4. おわりに

本稿では、アライグマ、タヌキ、背景を含む機械学習用データベースを構築し、畳み込みニューラルネットワークを用いた害獣分類器を試行した。この害獣分類器は認識率93%という十分な値を達成した。画像の深層学習のリファレンスとして広く用いられているCIFER-10のサンプルテストセットでの認識率が80～90%程度であり、初期試行としては93%は良好な認識率と判断できる。今回使用したのは特定の動物飼育施設の特定の個体（アライグマ6頭）について、特定の一日昼間（11月の冬毛）のみで撮影したものであり、今後、他の個体・環境の画像も収集し、未学習の個体・環境で撮影した画像による評価も行っていかなければならない。また、今回は撮影のみのアナグマやハクビシンについてもデータベース化をすすめていきたい。さらに、監視カメラで実現するための組込みシステムとしての実装と性能向上、アライグマの繁殖状況の分析や歴史文化財の保護システムへの応用に向けて開発を続けていきたい。なお、本研究で構築しているデータベースは一般に公開していく予定である。

謝辞：本研究に際してご支援・ご協力をいただいた一般財団法人 弘前市みどりの協会 弥生いこいの広場、弘前市農業政策課、関西野生生物研究所の関係各位に感謝の意を示す。本研究は、歴史都市防災研究所 研究拠点形成支援プログラムの支援を受けている。

参考文献

- 1) 環境省, “アライグマ防除の手引き (計画的な防除の進め方)” . https://www.env.go.jp/nature/intro/3control/files/manual_racoon.pdf, (2017-02-13閲覧)
- 2) 長尾知晴, 安居院猛, 長橋宏. “遺伝的手法を用いた2値図形のパターンマッチング” . 電子情報通信学会論文誌 D-II . 電子情報通信学会, pp.557-565, 1993
- 3) Stanford Vision Lab. “Large Scale Visual Recognition Competition 2014” , ImageNet, <http://image-net.org/challenges/LSVRC/2014/index>, (2017-02-13閲覧)
- 4) Karen Simonyan, Andrew Zisserman. “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION”. ICLR2015
- 5) Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor. “Caffe: Convolutional Architecture for Fast Feature Embedding”. 2014. arXiv preprint arXiv:1408.5093, <http://ucb-icsi-vision-group.github.io/caffe-paper/caffe.pdf>, (2017-02-11閲覧)
- 6) Yann LeCun, Corinna Cortes, Christopher J.C. Burges. “THE MNIST DATABASE” . MNIST handwritten digit database. <http://yann.lecun.com/exdb/mnist/>, (2017-02-02閲覧)
- 7) Alex Krizhevsky . “The CIFAR-10 dataset” . CIFAR-10 and CIFAR-100 database. <https://www.cs.toronto.edu/~kriz/cifar.html>, (2017-02-02閲覧)
- 8) Anvsoft Inc. “Any Video Converter フリー版” . フリー動画変換ソフト. <http://jp.any-video-converter.com/any-video-converter-free.php>, (2017-02-02閲覧)
- 9) “Xvid” . <https://www.xvid.com/>, (2017-02-02閲覧)
- 10) mes. “NiVE2” . NiVE2 Wiki, <http://nive.jp/NiVE2/>, (2017-02-02 閲覧)