

深層学習を用いた動物分類に関する既存研究

2.1 赤外線画像に対する既存研究

気候変動や人口増加が生態系に与える影響を把握し、野生動物と人間の持続可能な共存を実現することは、重要な課題であるこの課題を解決するため、生態系モニタリングの重要性が世界的に高まっている [1, 2]。生態系モニタリングの手法として、監視カメラなどを用いた観測が広く採用されており、特定地域における動物種の個体数推定だけでなく、環境に対する各動物種の生態の観察や研究が行われている [3]。中でも、カメラトラップは、観察者による直接的な介入を最小限に抑えることが可能であり、観察者の存在が個体の行動に与える影響を軽減することができるため、野生動物の監視ツールとして広く活用されている [4, 5]。カメラトラップは、赤外線センサなどを用いた自動撮影により人的労力を削減することができ、近年のデジタルカメラの高性能化に伴い長期間にわたる連続的なモニタリングが可能である。一方で、カメラトラップを使用した生態系モニタリングでは、複数箇所にカメラトラップを設置するため膨大な画像枚数を取得することも多く [6, 7]、記録された画像・動画中から人手による動物の有無や種の推定は多大なアノテーションコストを要する [8]。加えて、種の分類には専門的な知識が必要であることも作業員確保のコスト面での課題である。さらに、技術革新は今後も進むと予想されるのに対して、アノテーションコストを著しく下げることは困難であるため、このギャップは今後一層拡大していくと予想される [9]。したがって、これらの課題を解決するため、カメラトラップによって撮影された画像・動画中から自動で動物を検出・識別する手法の実現が望まれている。

近年では、画像処理技術と機械学習を用いた野生動物の自動識別手法が研究されている。2012 年の画像処理タスクに関連する深層学習技術の登場以降、画像処理分野の様々な分野において CNN に基づく手

法が盛んに研究されている。また、その多くの分野において CNN は高い性能を実現しており、カメラトラップによる動物識別に関する研究においても CNN を用いた手法がいくつか提案されている。Tan ら [10] は、2014 年から 2020 年にかけて撮影された約 25,000 枚の自作データセットを用いて、YOLOv5, FCOS(Fully Convolutional One-Stage Object Detection), Cascade R-CNN の 3 つの検出ネットワークでの比較検証をおよび映像に適用した動物認識の性能を評価している。Tabak ら [11] は、全米 5 箇所で撮影された約 300 万枚のカメラトラップ画像を用いて、独自の CNN により動物の画像分類を行っている。

しかしながら、上記のような既存研究の分類モデルを学習するために用いられた大規模なデータセットは、多くの撮影場所や長期間に渡った撮影によって蓄積された画像で構成されている。したがって、これらの既存研究は、個人的な利用での撮影や狭い範囲の地域での撮影など、多くの画像を集められるとは限らない状況を考慮していないため実用的では無い。また、夜間に行動する動物の撮影には赤外線カメラを用いることが有効である。しかし、赤外線カメラで撮影された画像は色情報を含まないなど、可視光カメラで撮影された画像とは映り方が異なる。したがって、真に実用的な夜間の動物モニタリングの実現に向けて、少数の赤外線で撮影された画像のみでも学習可能な深層学習モデルについての検討が急務である。

このような既存研究の課題解決に向けた研究として、少数の赤外線画像を用いた動物分類が挙げられる。Kishi ら [12] は、米国南西部の 140 箇所で撮影された画像 3000 枚を用いて、CNN による少数の赤外線画像を用いた動物分類を行っている。また、少数のデータを用いて深層学習モデルを効果的に学習させることを目的とする Few-Shot Learning (FSL) の分野において有効な手法である転移学習とデータ拡張の赤外線画像に対する有効性を検証した。転移学習は、事前に大量のデータを用いて学習したモデルを新しいタスクに用いる手法であり、本研究では、画像認識タスクに一般的に用いられるデータセットである ImageNet と ImageNet を擬似赤外線化した画像、さらに数式から生成されたフラクタル画像による転移学習の有効性をそれぞれ検証した。一方、データ拡張については、一般的な画像変換によるデータ拡張と画像の一部をマスクし隠すことでモデルの汎化性能を向上させる Random Erasing、画像処理後の画像を組み合わせることで新しい画像を生成しモデルの頑健性を向上させる Augmix などの有効性を比較検証した。結果として、転移学習では疑似赤外線画像、フラクタル画像、ImageNet の順に効果があることが示された。また、データ拡張においては、最新の手法である AugMix が特に有効である

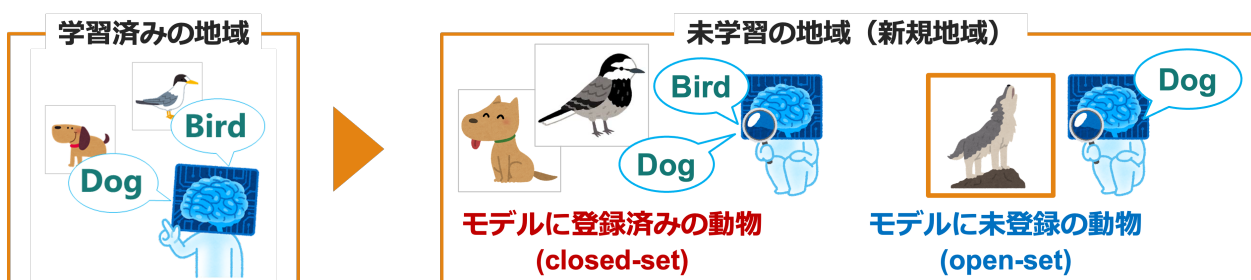


図 2.1: 一般的な分類モデルが未見データを識別できないことを示す例

と結論づけられている。しかし、Kishi らの研究の課題として、転移学習では通常大量のデータでモデルを学習する手法であるが、本研究では 1 クラスあたり 100 枚という限られたデータのみを使用している。そのため、厳密には転移学習ではなく、別ドメイン画像を用いたファインチューニングによる、少数の赤外線画像分類への有効性の検証を行った研究だと解釈できる。よって、転移学習を赤外線画像タスクに用いた際の有効性は検証できていない。また、データが少ない状況を想定し、学習に使用する画像は 600 枚としているが、実運用を想定した場合では 600 枚程度の収集すら困難な場合も考えられる。さらに、モデルの評価実験におけるテスト用データセットには学習時と同じ動物のクラスが使用されていることから、運用する地域ごとに 600 枚程度の学習データを集めることが前提の研究となっている。つまり、一度学習したモデルを別地域にて即時に適用することが困難だと考えられる。そして、赤外線画像に対する事前学習やデータ拡張の評価を行っているものの、一般的なカラー画像との比較が行われていない。したがって、各手法が赤外線画像のみに対して効果的であるか、それとも一般的な画像分類においての精度向上に寄与しているのかを判断することができない。

特に重要な課題として、モデルの学習に使用されるデータに含まれる動物種が少ない場合、実用時には学習データに含まれなかった動物が出現することが挙げられる。一般的な分類モデルは、そういった未見の動物を正しく識別できず、学習済みのクラスに強制的に分類してしまう。図 2.1 は、一般的な分類モデルが未見動物を識別できない例を示している。

図 2.1 で示される通り、一般的な分類モデルは分類したい画像を、画像から抽出した特徴量が類似している既存クラスに分類するため、未見データであっても等しく既存クラスに分類されてしまう。そのため、もし学習時データに含まれない危険な動物が実用時に現れた場合、モデルは判別することができず、重大な事故を引き起こす可能性がある。

以上より、未見の動物を識別できるモデルの開発は急務である。なお、未見の識別を行う能力をモデ

ルに付与する問題設定をオープンセット認識 (Open Set Recognition, OSR) と呼ぶ。

2.1.1 Few-Shot Open-Set Recognition に関する既存研究

少数のデータを用いた分類 (FSL) と未見の識別 (OSR) を両立することは非常に困難な課題であるが、近年この問題を解決するための新たな分野 Few-Shot Open-Set Recognition (FSOSR) が注目されている。FSOSR は、少数の画像を用いた学習をもとに正確な分類を行うことと、学習に存在しないクラスがテスト時に現れた際に未見データであると識別・拒絶できるモデルの構築を目的とした分野である。FSOSR において未見識別が可能となることを説明した図を図 2.2.1 に示す。図 2.2.1 では、既存の分類モデルが抱えていた未見データでも既存クラスに分類してしまうという課題を、分類したい画像の特徴量と既存クラスの類似度がある閾値以上でなければ未見と識別することで解決できるようになったことを示している。しかし、FSOSR において、FSL と OSR の手法をただ組み合わせただけでは、画像の分類と未見識別の両方を高水準で実現することが困難なことが示されている [13]。その原因として、OSR の問題設定では、学習データが非常に多いのに対し、FSL では学習データが非常に少ないという違いにより、既存の OSR 手法が FSL への適用が困難であることが挙げられる。

上記の問題の解決策として、メタ学習を FSOSR に応用した手法 [13] が提案された。メタ学習とは、FSL の分野において有効な手法の一種であり「学習する方法を学習する」という概念に基づいた手法である。FSL に用いられるメタ学習では、モデルは少数のデータを用いた分類に適応する必要があるため、少数データの分類を適切に行うことができる特徴抽出を学習する。本研究では、いくつかあるメタ学習の手法の中でも Prototypical Networks (ProtoNet) [14] を用いたメタ学習に着目する。ProtoNet は、一般的な深層学習モデルが画像をヘッドを用いてクラス分類確率まで畳み込み分類を行うのに対して、畳み込みを行う直前の特徴ベクトルを利用することにより画像の分類を行う手法である。クラス分類確率と特徴ベクトルの関係について、一般的な深層学習モデル構造の例として ViT のモデル構造を図 2.2.2 に示す。一般的な分類問題では、クラス分類確率が最も高いクラスに画像が分類されるが、FSL においては、少量のデータを使用するため画像の特徴が豊富に表現されている特徴ベクトル同士を比較することにより効果的に分類を行うことが可能となる。

FSL の問題設定では、いくつかの見本となるデータが渡された際、分類を行いたい対象データが見本データのどのクラスに属するかを正確に分類することが求められる。この時、正確な分類を行うために

は、分類モデルは対象画像から特徴を抽出する際に、見本データに類似する特徴ベクトルを抽出できる必要がある。メタ学習では、追加データを用いて、対象データと見本データの特徴ベクトル間の類似度を高めるような特徴抽出を繰り返し学習することで、新しいデータに対しても正確な分類ができるようにモデルの訓練を行う。

FSOSR に拡張されたメタ学習においては、未見データを識別するために、未見データを他の見本クラスとは類似度が低くなるような損失を追加することで、未見の識別と分類の両方を学習することを可能とした。

本研究では、2.1 章で述べた少数の赤外線画像における動物分類の問題解決のために新しい問題設定を提案するとともに、提案する問題設定におけるメタ学習の有効性について検証を行う。

参考文献

- [1] Joeri Zwerts, PJ Stephenson, Fiona Maisels, Marcus Rowcliffe, Christos Astaras, Patrick Jansen, Jaap van der Waarde, Liesbeth Sterck, P.A. Verweij, Tom Bruce, Stephanie Brittain, and Marijke Kuijk, “Methods for wildlife monitoring in tropical forests: Comparing human observations, camera traps, and passive acoustic sensors,” *Conservation Science and Practice*, Vol. 3, 2021.
- [2] Jabili Bandaru, Nikitha Basa, P Raghavendra, and A. Sirisha, “Review on various techniques for wildlife monitoring and alerting systems,” In *Proceedings of the International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, Vol. 1, pp. 1–5, 2024.
- [3] Franck Trollet, Marie-Claude Huynen, Cédric Vermeulen, and Alain Hambuckers, “Use of camera traps for wildlife studies. a review,” *Biology Agriculture Science Environnement*, Vol. 18, pp. 446–454, 2014.
- [4] 本郷 峻, “霊長類学におけるカメラトラップ研究,” *霊長類研究*, Vol. 34, No. 1, pp. 53–64, 2018.
- [5] Baydaa Sh. Z. Abood, Manjula B. M, Zainab abed Almoussawi, N Shilpa, and Aboothar mah-mood Shakir, “Revolutionizing wildlife monitoring: A novel approach to camera trap image analysis with yolov5,” In *Proceedings of the 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, pp. 1–6, 2023.
- [6] Roland Kays, Brian S. Arbogast, Megan Baker-Whatton, Chris Beirne, Hailey M. Boone, Mark Bowler, Santiago F. Burneo, Michael V. Cove, Ping Ding, Santiago Espinosa, André Luis Sousa Gonçalves, Christopher P. Hansen, Patrick A. Jansen, Joseph M. Kolowski, Travis W. Knowles, Marcela Guimarães Moreira Lima, Joshua Millspaugh, William J. McShea, Krishna Pacifici, Arielle W. Parsons, Brent S. Pease, Francesco Rovero, Fernanda Santos, Stephanie G. Schuttler,

- Douglas Sheil, Xingfeng Si, Matt Snider, and Wilson R. Spironello, “An empirical evaluation of camera trap study design: How many, how long and when?,” *Methods in Ecology and Evolution*, Vol. 11, No. 6, pp. 700–713, 2020.
- [7] Xingfeng Si, Roland Kays, and Ping Ding, “How long is enough to detect terrestrial animals? estimating the minimum trapping effort on camera traps,” *PeerJ*, Vol. 2, p. e374, 2014.
- [8] Rajasekaran Thangaraj, Sivaramakrishnan Rajendar, Sanjith M, Rithick Saran K, Sudev Sasikumar, and Chandhru L, “Automated recognition of wild animal species in camera trap images using deep learning models,” In *Proceedings of the Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pp. 1–5, 2023.
- [9] 安藤 正規, 中塚 俊介, 相澤 宏旭, 中森 さつき, 池田 敬, 森部 絢嗣, 寺田 和憲, and 加藤 邦人, “深層学習 (deep learning) によるカメラトラップ画像の判別,” *哺乳類科学*, Vol. 59, No. 1, pp. 49–60, 2019.
- [10] Mengyu Tan, Wentao Chao, Jo-Ku Cheng, Mo Zhou, Yiwen Ma, Xinyi Jiang, Jianping Ge, Lian Yu, and Limin Feng, “Animal detection and classification from camera trap images using different mainstream object detection architectures,” *Animals*, Vol. 12, No. 15, 2022.
- [11] Michael A. Tabak, Mohammad S. Norouzzadeh, David W. Wolfson, Steven J. Sweeney, Kurt C. Vercauteren, Nathan P. Snow, Joseph M. Halseth, Paul A. Di Salvo, Jesse S. Lewis, Michael D. White, Ben Teton, James C. Beasley, Peter E. Schlichting, Raoul K. Boughton, Bethany Wight, Eric S. Newkirk, Jacob S. Ivan, Eric A. Odell, Ryan K. Brook, Paul M. Lukacs, Anna K. Moeller, Elizabeth G. Mandeville, Jeff Clune, and Ryan S. Miller, “Machine learning to classify animal species in camera trap images: Applications in ecology,” *Methods in Ecology and Evolution*, Vol. 10, No. 4, pp. 585–590, 2019.
- [12] Koki Kishi, Masako Kishimoto, Sulfayanti Situju, Hironori Takimoto, and Akihiro Kanagawa, “Few-shot learning for CNN-based animal classification in camera traps using an infrared camera,” In *Proceedings of the 10th IIAE International Conference on Intelligent Systems and Image Processing (ICISIP)*, pp. 11–17, 2023.

- [13] Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos, “Few-shot open-set recognition using meta-learning,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8795–8804, 2020.
- [14] Jake Snell, Kevin Swersky, and Richard S. Zemel, “Prototypical networks for few-shot learning,” *CoRR*, Vol. abs/1703.05175, 2017.