

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することは、野生動物と人間の共存において重要な課題である。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。具体的には、監視カメラなどを用いることにより、特定地域の動物種の個体数の推定だけでなく、その環境における各種の行動や生態について観察や研究が行われている。特に、野生鳥獣の監視ツールとして、人間が直接関与せず、また、生態系への影響が極力少ない遠隔での自動撮影を行うカメラトラップが広く使用されている。カメラトラップは、モーションセンサや赤外線センサによって自動での撮影を可能にしているため、人間の労力削減と 24 時間のモニタリングが可能である。また、野生鳥獣が現れそうな場所の樹木や地面にカメラを固定するだけの単純な設置方法であることも利点である。カメラトラップを使用した撮影では、数年間にわたった撮影期間や、多くの撮影箇所の設定などによって、膨大な画像枚数を取得していることも少なくない。しかしながら、記録された画像・映像中から鳥獣の有無や種の推定を人手で行うことは多大なコストを要する。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。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鳥獣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

[1] 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 ObjectDetection), 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 [2] は、全米 5 か所で撮影された約 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を用いることが有効である。しかし、図 2.1.1 で示すように赤外線撮影された画像は色情報を含まないなど、可視光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 [1] は、米国南西部の 140 か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 2.1.2 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 2.1.2 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。

このような課題を解決するため、少数の赤外線画像を用いた動物分類の研究が行われている [3]。しかし、既存研究では、評価時に分類対象となる動物種は全て学習済みであると仮定しているが、実運用において、深層学習モデルを特定の地域に適用する際、モデルが対象地域に生息する全ての動物種を学習しているとは限らない。このような状況において、未学習の動物種は学習済みの動物種に強制的に誤分類され、モデルの性能は著しく低下することが知られている。この問題はオープンセット問題 (Open-Set Problem) と呼ばれ、この問題に対処するため、学習済みクラスの分類を行いつつ、未学習クラスを検出するオープンセット認識 (Open-Set Recognition, OSR) 手法が提案されている [4, 5]。

さらに近年では、少数データでもオープンセット認識を可能にする Few-Shot Open-Set Recognition

(FSOSR) [6] が注目を集めている。代表的な FSOSR 手法として、少数データ学習 (Few-Shot Learning, FSL) 分野で有効な手法とされているメタ学習を OSR に拡張することにより FSL と OSR を同時に実現した PEELER [6] や、変換の一貫性に基づき未学習クラスを検出することによって、擬似的な未学習クラスサンプルを必要としない SnaTCHer [7] が挙げられる。

しかし、これらの手法は可視光画像を対象としており、赤外線画像に対して性能評価がなされていない。また、FSOSR では未学習クラスを単一種として扱っているが、未学習クラスのアノテーションや追加学習を考慮すると、実用的には未学習クラスも複数種に分類できることが望ましい。

本論文では、夜間の野生動物モニタリングの実現を目的とした、より実用的な問題設定である「Infrared Few-shot Open-set Recognition (IFOR)」を提案する。IFOR では少量の赤外線画像データのみを用いて、特定地域に生息するモデルに学習済みの動物種を正確に分類し、かつ、未学習の動物種の検出を可能にすることを目指す。

加えて、IFOR ではドメインシフトに対する頑健性の評価も行う。ドメインシフトとは、学習データと評価データが異なる地域で収集された場合に生じる課題であり、背景や撮影環境の違い、同じ動物種の地域差による外見の違いによってモデルの性能が低下する現象を指す。ドメインシフトを考慮することにより、地理的条件に依存せず、様々な場所に適用可能な汎用性の高いシステムの実現が期待される。

IFOR の実現に向けて、赤外線画像に有効な既存手法の特定に加え、既存の FSOSR 手法の 1 つであるメタ学習フレームワークが IFOR に対して効果的であるか検証を行う。まず、赤外線画像に有効な特徴抽出器を特定するため、テクスチャ特徴に焦点を当てている CNN や、形状特徴 [8] を重視することで知られている Vision Transformer (ViT) [9] などの代表的な特徴抽出器の有効性を赤外線画像に対して評価する。次に、IFOR フレームワーク内の FSL タスクに有効なアプローチの 1 つである転移学習について検証する。転移学習では、事前学習のタスクと本番環境でのタスクの類似度が重要だと考えられている。そこで、一般的な ImageNet データセットを用いた事前学習と並行して、事前学習に色情報を持たないフラクタル画像を用いる Formula-Driven Supervised Learning (FDSL) [10] の有効性を探る。最後に、赤外線画像を分類する際、小規模データセットから汎用的な特徴抽出を行うための学習戦略であるメタ学習の IFOR における有効性を、ドメインシフトの条件下で評価する。特に、IFOR においては学習済みクラスの正確な画像分類と未学習クラスの検出が不可欠であるため、メタ学習による有効性を従来の学習方法であるミニバッチ学習と比較する。

さらに, IFOR を発展させ, 未学習クラスに対する多クラス分類の精度向上にも取り組む. 特徴空間上で各クラス分布がコンパクトに表現されることによって多クラス分類が容易になると仮定し, クラスタリングに基づく損失関数を用いてクラス内分散の最小化・クラス間分散の最大化を図る. クラス内分散の最小化では, 異常検知タスクで用いられている k-means 損失 [11] を導入する. クラス間分散の最大化では, k-means クラスタリングによって得られる各クラスタ中心を利用した損失関数, Between-Class 損失 (BC 損失) を提案し, 性能評価を行う.

以下, 第 2 章では深層学習を用いた動物分類に関する既存研究について述べる. 第 3 章では夜間の野生動物モニタリングの実現に向けてより実用的な問題設定を提案し, 様々な手法の有用性について述べる. 第 4 章では評価実験を行い, その結果及び考察を多面的な方向性から述べる. 最後に第 5 章では結論を述べる.

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

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng, “Conditional gaussian distribution learning for open set recognition,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13477–13486, 2020.
- [5] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman, “Open-set recognition: A good closed-set classifier is all you need,” In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.
- [6] Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos, “Few-shot open-set recog-

- nition using meta-learning,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8795–8804, 2020.
- [7] Minki Jeong, Seokeon Choi, and Changick Kim, “Few-shot open-set recognition by transformation consistency,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12561–12570, 2021.
- [8] Shikhar Tuli, Ishita Dasgupta, Erin Grant, and Thomas L. Griffiths, “Are convolutional neural networks or transformers more like human vision?,” *arXiv preprint arXiv:2105.07197*, 2021.
- [9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
- [10] Hirokatsu Kataoka, Kazushige Okayasu, Asato Matsumoto, Eisuke Yamagata, Ryosuke Yamada, Nakamasa Inoue, Akio Nakamura, and Yutaka Satoh, “Pre-training without natural images,” *International Journal of Computer Vision (IJCV)*, 2022.
- [11] Chin-Chia Tsai, Tsung-Hsuan Wu, and Shang-Hong Lai, “Multi-scale patch-based representation learning for image anomaly detection and segmentation,” In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3992–4000, 2022.