

序論

生物多様性が生み出す生態的機能は、温室効果ガス、エネルギー、水などの物質・資源循環など、地球に不可欠な役割を多数持っており、人間社会の基盤となっている [1, 2]。生態的機能のうち、特に人間社会の便益につながるものを生態系サービスと呼ぶ。しかしながら、昨今、人間による土地開発や気候変動によって、陸域生態系の劣化と生物多様性の損失は進行しており [3, 4]、生態系サービスの享受が困難になっている。したがって、生態系の劣化を低減し、生物多様性の損失を抑制することは、現代社会における根本的な社会課題であり、その解決が急務である。これらの課題は持続可能な開発目標（SDG15: 陸の豊かさを守ろう）などの国際目標としても掲げられ、日本でも生物多様性国家戦略 [5] をはじめとする様々な政策や企業の取り組みが開始されている。

これらの社会課題に対して、継続的な野生動物との共存を実現するため、生態系モニタリングが注目を集めている。生態系モニタリングは、自然環境の空間的・時間的変化の把握に有効な方法であり、自然保護や環境保全に重要なデータを提供する。モニタリング実施の効果として生態系に生じた異常の早期発見が可能であり、迅速な対策により生態系の回復期間の短縮やコスト削減へと繋がることが期待される。

費用対効果の高い生態系モニタリングを行う上で、カメラトラップの使用は極めて重要である [6]。カメラトラップは、赤外線センサなどを用いてカメラの前を通り過ぎる動物を感知し、自動的に撮影するため、動物にストレスを与えることなく、撮影者によるバイアスを排除したデータ収集が可能である [7, 8]。これらのカメラは、比較的低価格であり、限られた電力資源で効率的に動作するため、広範囲に長期間の撮影が可能である [9, 10]。また、赤外線カメラの使用により、夜間の撮影も可能である。近年、カメラトラップによって膨大な画像や動画データが低コストで収集できるようになり、深層学習モデルを用いた野生動物の正確な検出と分類に期待が高まっている [11]。

野生動物の分類に関して、畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) を用いた手法がいくつか提案されている。しかし、自然環境下における野生動物に対する分類タスク特

画像

[~]
文庫



(a) 可視光画像



(b) 赤外線画像

図 1: 撮影方法の違いによる物体の写り方

有の課題として、特定地域における十分な量の学習用画像の収集コストが高いことや [12], 赤外線カメラによって撮影された画像は色情報が欠落していること [13] などが挙げられる。図 1 に撮影方法の違いによる物体の写り方の違いを示す。既存の動物分類手法のほとんどは図 1(a) のような可視光画像に焦点を当てており、図 1(b) のような赤外線画像に対して取り組んだ研究は少ない。また、赤外線画像と可視光画像では物体の写り方が大きく異なるため、既存の可視光画像に対する手法を赤外線画像に対して適用すると精度が大きく低下する。

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 [13]。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) と呼ばれる。この問題に対処するため、学習済みクラスのカテゴリを行いつつ、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 [14, 15]。

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) [16] 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 [16] 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 [17] 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 [18] を重視することで知られている Vision Transformer (ViT) [19] 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) [20] 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、クラスター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 [21] を導入する。クラス間分散の最大化で

学習済み!

未学習データ?
クラス?

である。

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。

参考文献

- [1] Millennium Assessment and Albert Van Jaarsveld, *Ecosystems and Human Well-Being: Biodiversity Synthesis*. 2005.
- [2] Bradley Cardinale, Kristin Matulich, David Hooper, Jarrett Byrnes, J. Duffy, Lars Gamfeldt, Patricia Balvanera, Mary O'Connor, and Andrew Gonzalez, "The functional role of producer diversity in ecosystems," *American journal of botany*, Vol. 98, pp. 572–92, 2011.
- [3] Tim Newbold, Lawrence Hudson, Samantha Hill, Sara Contu, Igor Lysenko, Rebecca Senior, Luca Börger, Dominic Bennett, Argyrios Choimes, Ben Collen, Julie Day, Adriana De Palma, Sandra Diaz, Susy Echeverria-Londono, Melanie Edgar, Anat Feldman, Morgan Garon, Michelle Harrison, Tamera Alhusseini, and Andy Purvis, "Global effects of land use on local terrestrial biodiversity," *Nature*, Vol. 520, pp. 45–50, 2015.
- [4] Forest Isbell, Andrew Gonzalez, Michel Loreau, Jane Cowles, Sandra Diaz, Andy Hector, David Wardle, Mary O'Connor, J. Duffy, Lindsay Turnbull, Patrick Thompson, and Anne Larigauderie, "Linking the influence and dependence of people on biodiversity across scales," *Nature*, Vol. 546, pp. 65–72, 2017.
- [5] 環境省, "生物多様性国家戦略 2023-2030～ネイチャーポジティブ実現に向けたロードマップ～," tech. rep., 環境省 自然環境局, 東京, 2023. https://www.biodic.go.jp/biodiversity/about/initiatives6/files/1_2023-2030text.pdf (2024 年 12 月 7 日確認) .
- [6] Liang Jia, Ye Tian, and Junguo Zhang, "Domain-aware neural architecture search for classifying animals in camera trap images," *Animals*, Vol. 12, p. 437, 2022.
- [7] Scott Newey, Paul Davidson G, Sajid Nazir, Gorrry Fairhurst, Fabio Verdicchio, Robert Irvine, and Rene van der Wal, "Limitations of recreational camera traps for wildlife management and

- conservation research: A practitioner' s perspective," *Ambio*, Vol. 44, pp. 624–635, 2015.
- [8] Chunbiao Zhu, Thomas H. Li, and Ge Li, "Towards automatic wild animal detection in low quality camera-trap images using two-channelled perceiving residual pyramid networks," In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 2860–2864, 2017.
- [9] Stefan Schneider, Graham W. Taylor, and Stefan Kremer, "Deep learning object detection methods for ecological camera trap data," In *Proceedings of the 15th Conference on Computer and Robot Vision (CRV)*, pp. 321–328, 2018.
- [10] Christin Carl, Fiona Schönfeld, Ingolf Profft, Alisa Klammer, and Dirk Landgraf, "Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model," *European Journal of Wildlife Research*, Vol. 66, No. 62, 2020.
- [11] 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.
- [12] Stefan Schneider, Saul Greenberg, Graham Taylor, and Stefan Kremer, "Three critical factors affecting automated image species recognition performance for camera traps," *Ecology and Evolution*, Vol. 10, No. 7, pp. 3503–3517, 2020.
- [13] 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.
- [14] 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.
- [15] 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.

- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.

