

令和 6 年度  
修 士 論 文

メタ学習に基づく Infrared Few-shot  
Open-set Recognition を考慮した動物分類

岡山県立大学大学院 情報系工学研究科  
システム工学専攻

学籍番号 : K623009    氏名 : 岸 孝樹

令和 7 年 1 月 30 日 提出

# Animal Classification Considering Infrared Few-shot Open-set Recognition Based on Meta-learning

K623009      Koki Kishi

## ABSTRACT

Camera traps are crucial for cost-effective ecosystem monitoring. In particular, there are high expectations for using infrared cameras to capture night images. However, there are challenges with classifying infrared images, such as the lack of color information and insufficient training data.

This paper proposes the novel Infrared Few-shot Open-set Recognition (IFOR) framework for wildlife monitoring using camera traps. The framework considers three key challenges: infrared image classification, limited training data, and unknown species detection. We evaluated the effectiveness of meta-learning with different feature extractors (CNN and ViT) and transfer learning approaches within the IFOR framework. Furthermore, to improve the multi-class classification accuracy for unknown animal species, k-means loss and Between-Class loss were introduced to minimize intra-class variance and maximize inter-class variance. The experiments utilized data collected from different regions for training and evaluation and assessed the model's performance under domain shift conditions. Experimental results showed that ViT with ImageNet pre-training and meta-learning was effective for IFOR, and the proposed loss functions effectively enhanced the classification accuracy of both known and unknown species.

The results of this research provide important insights for realizing effective wildlife monitoring systems under limited data conditions. Moreover, they are expected to be a foundation for developing practical machine-learning models in ecological research.

# 目次

<b>第 1 章</b>	<b>序論</b>	1
<b>第 2 章</b>	<b>深層学習を用いた動物分類に関する既存研究</b>	4
2.1	赤外線画像に対する動物分類に関する既存研究 . . . . .	4
2.2	Few-Shot Open-Set Recognition に関する既存研究 . . . . .	6
<b>第 3 章</b>	<b>メタ学習に基づく Infrared Few-shot Open-set Recognition を考慮した動物分類</b>	8
3.1	Infrared Few-shot Open-set Recognition (IFOR) . . . . .	8
3.2	IFOR に対して有効な手法の提案 . . . . .	9
3.2.1	特徴抽出器 . . . . .	9
3.2.2	転移学習 . . . . .	10
3.2.3	メタ学習 . . . . .	11
3.3	未登録クラスに対する多クラス分類の高精度化に向けたクラスタリングに基づく損失関数 . . . . .	17
3.3.1	メタ学習にクラスタリングを導入する目的 . . . . .	17
3.3.2	損失関数 . . . . .	18
<b>第 4 章</b>	<b>評価実験</b>	20
4.1	データセット . . . . .	20
4.2	未登録クラスの検出に対する提案手法の評価 . . . . .	22
4.2.1	実験条件 . . . . .	22
4.2.2	実験結果及び考察 . . . . .	23
4.3	未登録クラスの多クラス分類に対する損失関数の評価 . . . . .	26
4.3.1	実験条件 . . . . .	26
4.3.2	実験結果及び考察 . . . . .	27
<b>第 5 章</b>	<b>結論</b>	30
<b>謝辞</b>		32
<b>研究業績</b>		33
<b>参考文献</b>		35

# 第1章

## 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（SDG15：陸の豊かさも守ろう）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・時間的变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]。図 1.1 にカメラトラップの例を示す。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



図 1.1. カメラトラップの例



(a) 可視光画像



(b) 赤外線画像

図 1.2. 撮影方法（カメラ）の違いによる物体の写り方

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成本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].

野生動物画像の分類に関して、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 [12, 13, 14, 15]. 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 [16]、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 [17] などが挙げられる。図 1.2 に撮影方法（カメラ）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 1.2(a) のような可視光画像に焦点を当てており、図 1.2(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した場合、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深層学習モデルの学習に用いた動物分類に関する研究が行われている [17]. 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おり、この問題はオープンセット問題 (Open-Set Problem) [18] と呼ばれる。このオープンセット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 [19, 20].

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) [21] 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 [21] 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 [22] 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、赤外線画像に対する性能評価が行われていない。また、FSOSR では未学習クラスを單一種として扱っているが、新しい動物種に対するアノテーションや追加学習に要するコストを考慮すると、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必要である。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など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。

本論文では、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 [23] を重視することで知られている Vision Transformer (ViT) [24] 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) [25] 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。特徴空間上で各学習済みクラスの分布がコンパクトに表現されることにより、未学習データに対しても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 [26] 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である Between-Class 損失 (BC 損失) を提案する。

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から述べる。最後に第 5 章では結論を述べる。

## 第 2 章

# 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動物分類に関する既存研究

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 [27, 28]。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、固有の環境に対する各動物種の生態観察や研究が行われている [29]。特に、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 [30, 31]。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 [32, 33]、記録された画像・動画中に対する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 [34]。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による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、アノテーションコストを著しく下げる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 [35]。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 [12, 13]。国際的な画像認識アルゴリズム性能コンペティションとして知られる ImageNet Large Scale Visual Recognition Challenge (ILSVRC) の 2012 年大会において、AlexNet [36] が画期的な成功を収めて以降、画像処理分野の様々なタスクにおいて CNN に基づく手法が盛んに研究されている [37, 38]。また、その多くのタスクで 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 [14, 15, 31, 34]。Tan ら [11] 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、YOLO-V5 [39]、FCOS (Fully Convolutional One-Stage Object Detection) [40]、Cascade R-CNN [41] の 3 つの検出ネットワークでの比較検証、及び映像に適用した動物認識の性能を評価している。Tabak ら [42] は、全米 5 箇所で撮影された約 300 万枚のカメラトラッ



図 2.1. 従来の分類モデルによるオープンセット問題の例

普画像を用いて、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、多くの撮影場所や数年間に及ぶ長期間の撮影によって蓄積された画像で構成されている。そのため、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、十分な画像データを収集できない状況には適しておらず、実用面での課題が残る。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が、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、真に実用的な生態系モニタリングシステムの実現に向けて、赤外線カメラによって撮影された少數の学習用データから、効率的に学習可能な深層学習モデルの開発が急務である。

このような既存研究の課題解決に向けた研究として、少數の赤外線画像を用いた動物分類が検討されている。Kishi ら [17] は、米国南西部の 140 箇所で撮影された画像 3,000 枚を用いて、CNN による少數の赤外線画像を用いた動物分類を行っている。この先行研究では、少數データを用いた効率的な深層学習モデルの学習を目的とする 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が検証された。まず、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適用する手法である。この先行研究では、画像認識タスクで一般的に用いられる ImageNet データセット、ImageNet データセット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が検証された。一方、データ拡張については、一般的な幾何学変換や色変換などの画像変換に加え、画像の一部をマスクし隠すことによってモデルの汎化性能を向上させる Random Erasing や、複数の画像処理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 などの有効性が検証された。実験の結果、転移学習では疑似赤外線画像、フラクタル画像、ImageNet の順に効果が高いことが示され、データ拡張については AugMix が特に有効であることが明らかになった。

Kishi らの研究では、新規地域に対する深層学習モデルの適用開始時の状況を想定しており、学習に使用する画像は 1 クラスあたり 50 枚としている。しかし、実運用を想定すると 1 クラスあたり 50 枚程度の画像収集すら困難な場合も考えられる。また、評価実験における評価用データセットでは学習用データセットと同じ動物種のみが使用されており、モデルの適用地域に生息する全ての動物種がモデルに登録されている閉集合が仮定されている。しかし、モデルの実運用開始時において、対象地域に生息する全ての動物種のデータを網羅的に収集することは現実的ではない。従来の分類モデルでは、学習データに含まれていないモデルに未登録の動物を正しく識別できず、登録済みのクラスに強制的に分類されてしまうオープンセット問題が存在する。図 2.1 は、従来の分類モデルが未登録の動物を識別できない例を示している。図 2.1 に示される通り、犬と鳥のみを学習したモデルを新規地域に適用した場合、モデルに登録済みの犬や鳥は分類できるが、未登録の動物であるオオカミは登録済みの動物種に強制的に分類されてしまう。このような誤分類

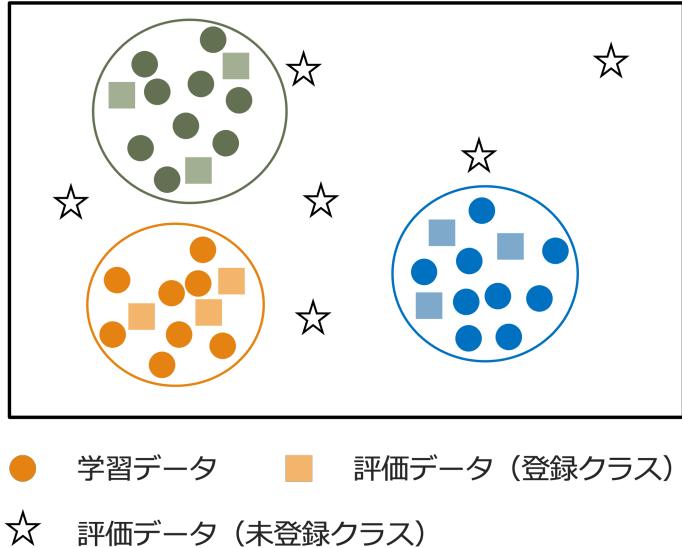


図 2.2. OSR における未登録データの検出プロセスの一例

はモデルの精度低下へと繋がるため、未登録の動物種を適切に検出できる OSR モデルの開発が急務となっている。なお、本論文では、モデルに登録されているクラスセットをクローズドセット (closed-set)，モデルに未登録のクラスセットをオープンセット (open-set) と呼ぶ。

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

昨今の第 3 次 AI ブームにおける機械学習システムの高い性能や実用的な成果は、主にクローズドセットタスクにおいて達成されてきた。これらのシステムでは、学習用データセットと評価用データセットに同一のクラスが含まれることを前提としており、システムの評価は学習時に登録されたオブジェクトクラスのみを対象として実施される。しかしながら、より実用的なシステムの実現を目指す場合、現実的な問題設定としてオープンセット問題への対応が不可欠である [43]。機械学習ベースのシステムの利用が進むにつれ、幅広いアプリケーションにおいて高い頑健性を備えた手法が要求されており、OSR 技術に注目が集まっている [44]。OSR は、学習時に想定していないクラスのデータが入力された場合でも、システムを頑健に機能させるためのアプローチの 1 つとして位置付けられている。図 2.2 に OSR における未登録データの検出プロセスの一例を示す。既存の分類モデルでは未登録データを登録クラスに分類してしまう課題があったが、分類対象の画像の特徴量と登録クラスの類似度が特定の閾値を下回る場合に未登録クラスとして検出することで、この課題を解決することが可能となった。

実世界における動的な環境下での認識・分類タスクに対して、より実用的なモデルを構築するためには未登録クラスへの対応が不可欠である。しかし、学習時に想定されていないクラスは無数に存在する可能性があり、それら全てを事前に予測して学習データに含めることは現実的ではない [45]。特に、深層学習モデルはデータ駆動型のアプローチのため、適切な帰納バイアスを獲得するために大量の学習データを必要とする。同様に、既存の OSR 手法においても大量の学習データの利用を前提としており、その適用範囲は局所的な場面に限定されている [46]。

この課題に対し、近年では Few-Shot Open-Set Recognition (FSOSR) という新たな研究分野が注目を集めている [21, 47]。FSOSR は、少数の画像データからの効率的な学習による正確な分

類を行うことと、学習データに存在しないデータを未登録クラスとして検出・拒絶できるモデルの構築を目的とした研究分野である。

Jeong ら [22] は、変換一貫性 (transformation consistency) という概念に基づき、未登録クラスを検出する SnaTCHer を提案した。これは、類似した入力データは特徴空間での変換後も近い位置関係を保つという性質を利用している。未登録クラスの入力データは、登録クラスとは異なる特徴空間を形成する傾向があるため、変換後の特微量は登録クラスのプロトタイプから大きく離れることが期待される。プロトタイプとは、登録クラスを代表する特徴ベクトルのことであり、登録クラスから得られた特微量の平均値を求めるこによって生成される。この手法の最大の利点は、疑似的な未登録クラスサンプルを必要としない点である。OSR などにおける従来手法が未登録クラスの分布を直接推定するのに対し、SnaTCHer は特微量の相対的な変換問題として扱うことで、より効率的な学習を実現した。また、様々な特微量変換手法との組み合わせ実験により、分類性能を低下させることなく、未登録クラスサンプルの検出性能を向上させることが確認された。

一方で、Huang ら [48] は、閾値の設定に依存しない新しい手法として Task-Adaptive Negative Envision (TANE) を提案した。TANE は、登録クラスのプロトタイプから負例のプロトタイプを生成し、これを用いてタスクに応じた動的な拒絶境界を学習する。具体的に、登録クラスの代表点に注意機構を適用して負例のプロトタイプを生成し、入力データとの類似度が計算される。もし全ての登録クラスに対する予測スコアが負例プロトタイプに対するスコアよりも低い場合、その入力データは未登録クラスとして拒絶される。

本論文では、限られたデータに対する赤外線動物分類の実現に向けて新しい問題設定を提案するとともに、今後の赤外線動物分類タスクの発展に向けて様々な手法の有効性を検証する。

## 第3章

# メタ学習に基づく Infrared Few-shot Open-set Recognition を考慮した動物分類

### 3.1 Infrared Few-shot Open-set Recognition (IFOR)

夜間に活動する動物を撮影するためには赤外線カメラを用いる必要があり、その結果として得られる画像は赤外線画像に限定される。しかし、既存の分類モデルのほとんどは可視光画像を対象としており、色情報を持たない赤外線画像への適用可能性については未だに検証の余地が残されている。本論文では、夜間における生態系モニタリングの実現に向けて、より実用的なモデルの構築支援を目的とし、少数の赤外線画像を用いた動物分類と未登録の動物識別という新たな問題設定 Infrared Few-shot Open-set Recognition を提案する。IFOR では、特定の地域に生息する野生動物の画像を大量に収集することが困難である現状を考慮し、モデルの学習に使用できるデータが限られている環境下でのシステムの運用を想定している。そのため本論文では、より実用的な夜間の生態系モニタリングに向けて、各クラス当たり 1 枚から 30 枚程度の少数画像を学習に用いた分類を行う。

また、収集されるデータが限られているという制約により、モデルの学習データは特定の地域に生息する動物を網羅的に含んでいない可能性が高い。したがって、実運用の際には学習データに含まれない動物が出現する可能性が考えられる。このような状況下では、モデルが未登録の動物を認識できずに登録済みのクラスに誤って分類してしまうため、未登録の動物種を正しく検出するシステムの実現が望まれている。

加えて、IFOR では、モデルの性能評価において、学習に用いたデータセットとは異なる地域で収集されたデータセットを評価時に使用する。これにより、地域間における環境や動物種の差異など、新規地域に対するモデルの適応能力を評価することが可能となる。このように、学習用データセットと評価用データセットのデータ分布が異なる状況はドメインシフトと呼ばれる。IFOR では、この地域間のドメインシフトを意図的に導入することにより、モデルの頑健性と汎用性を定量的に評価し、より広範な地域に対して適用可能なモデルの開発を推進する。

## 3.2 IFOR に対して有効な一手法の提案

本節では 3.1 節で提案した IFOR フレームワークに対する効果的な手法について論じる。提案手法は、IFOR を構成する「赤外線画像」、「少数データ学習」、「未登録クラスの検出」という三つの要素に着目し、各要素に対して効果的なアプローチを組み合わせることで、より高度な動物分類の実現を目指すものである。第一に、赤外線画像における特徴抽出に関して、CNN と ViT [24] という異なる特性を持つ二種類の特徴抽出器の有効性を検証する。第二に、少数データの問題に対しては、ImageNet や FDSL などの大規模データセットを用いた転移学習により、モデルの汎化性能の向上を図る。第三に、未登録データへの対応として、メタ学習アルゴリズムを導入し、少数データにおける分類精度と未登録クラスの検出能力の向上を実現する。

### 3.2.1 特徴抽出器

画像処理分野において、深層学習技術の登場以降、様々な分野において CNN を用いた手法が盛んに研究されている。これらの手法は様々なタスクにおいて高い精度を実現しており、従来のハンドクラフト特徴量に基づいた識別手法によるアプローチから、大規模な画像データセットを用いて学習された CNN の使用へと顕著なパラダイムシフトをもたらした。

CNN の一種である Residual Networks (ResNet) [49] は深い畳み込みニューラルネットワークを効率的に訓練するために開発された深層学習モデルである。本研究では、畳み込みニューラルネットワークの層が 18 層である ResNet18 を使用する。ResNet18 は層が比較的浅く、小規模なデータセットにおいて有効性が示されているため、本研究の特徴抽出器として採用する。

一方、近年 CNN に対する新たなアプローチとして、畳み込みを使用しない ViT が注目を集めている。ViT は、自然言語処理タスクで成功を収めた Transformer を画像処理タスクに応用したものであり、画像認識の多くの分野において CNN よりも高い性能を発揮している。ただし、ViT が最大限のモデルパフォーマンスを発揮するためには大規模なデータセットで学習を行う必要があり、転移学習を行わない場合には CNN と比較して性能が低下することが知られている [24]。

先行研究において、CNN はテクスチャ特徴の抽出に優れているのに対し、ViT は形状特徴の抽出に重点を置いていることが示されている [23]。本研究で対象とする赤外線画像は色情報を欠いているため、分類時には形状特徴がより重要な手がかりとなる。このことから、形状特徴の抽出に優れた ViT は赤外線画像の分類により適していると考えられる。そこで、異なる特性を持つ 2 つの特徴抽出器として ResNet18 及び ViT を用いて、赤外線画像に対する有効性を比較検証する。

さらに本論文では、赤外線画像に対して基盤モデルを用いた意味的な特徴抽出にも取り組む。上述したような CNN や ViT は画像のみを用いた学習を行うため、獲得できる表現能力に制限があった。しかし近年の研究では、テキストと画像のペアを用いた学習によって、より効果的な画像表現の学習が可能であることが示された。特に、Alec らはウェブから得た膨大なテキストと画像のペアを学習することにより、様々な下流タスクに転用可能な基盤モデル Contrastive Language-Image Pre-training (CLIP) を提案した [50]。図 3.1 に CLIP の事前学習の概要を示す。CLIP は対照学習 (Contrastive Learning) と呼ばれる学習手法を採用しており、類似した特徴量間（正例）の類似度は大きく、類似していない特徴量間（負例）の類似度は小さくなるように学習が行われる。具体的には、 $N$  個のテキストと画像のペアが入力され、画像エンコーダ (Image Encoder) やテキストエンコーダ (Text Encoder) によって、それぞれの特徴量が埋め込

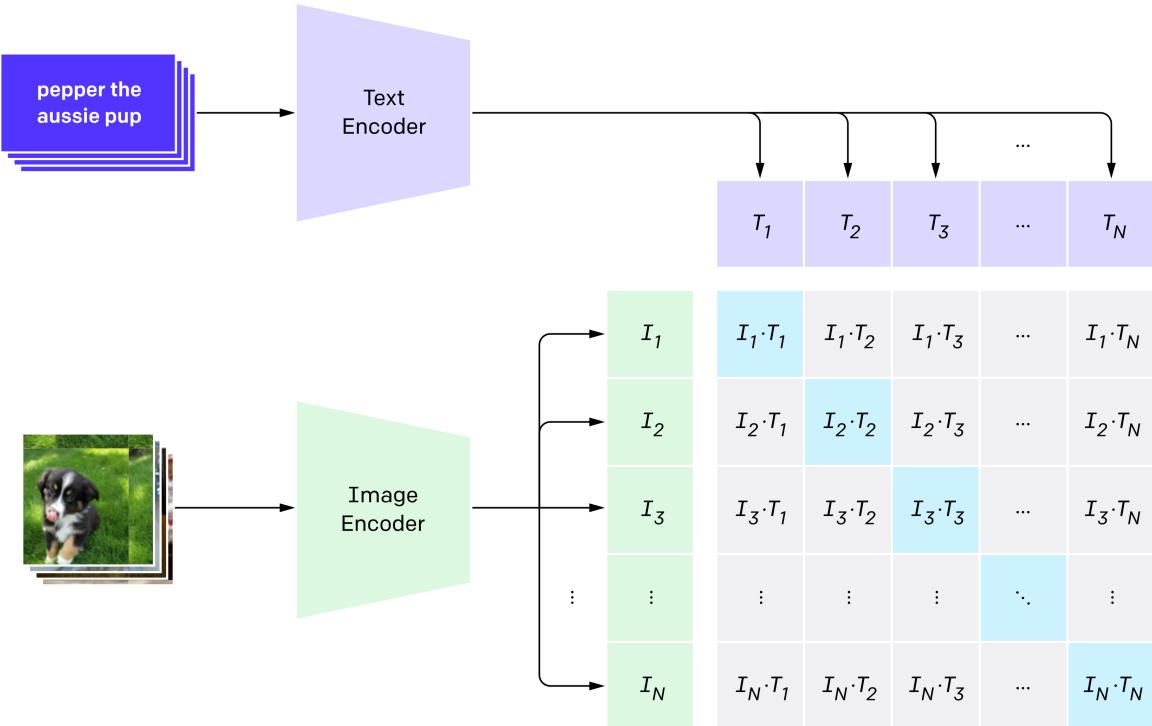


図 3.1. CLIP の事前学習 [50]

まれる。 $N$  枚の画像から得られる特徴量は  $I_1, I_2, \dots, I_N$  と、 $N$  個のテキストから得られる特徴量は  $T_1, T_2, \dots, T_N$  と表される。次に、埋め込まれた特徴量間のコサイン類似度が計算され、正例のコサイン類似度  $I_1 \cdot T_1, I_2 \cdot T_2, \dots, I_N \cdot T_N$  を最大化し、負例のコサイン類似度を最小化するよう画像エンコーダとテキストエンコーダのパラメータを更新する。この対照学習により、CLIP は画像のみの学習では表現が困難であった、意味的な特徴抽出が可能となる。また、CLIP は膨大な量の画像とテキストのペアを学習したことにより、タスク固有のデータセットによる追加学習を必要としない Zero-Shot タスクに対して高いパフォーマンスを示すことが確認されている。

本研究では、テキストを用いた対照学習によって意味的な特徴表現が可能である CLIP の IFOR に対する有効性を検証する。ただし、本研究では大規模なテキストと画像による対照学習が行われた CLIP モデルについて、画像エンコーダのみを特徴抽出器として採用する。

### 3.2.2 転移学習

一般的に、深層学習モデルは大規模なデータセットを用いた学習により高い汎化能力を獲得することが知られている。一方で、学習用データを十分に確保できない場合、過学習が起こる可能性が高く、モデルが十分な汎化性能を得ることは極めて困難である。実世界のタスクにおいては、多くの場合、大規模な学習データセットの構築が現実的に困難である。

これらの問題に対して、小規模なデータセットを用いて効率的にモデルの学習を行う FSL タスクでは、転移学習が重要な解決策として知られている。図 3.2 に転移学習の概要を示す。転移学習とは、事前に別のタスクから得られた知識を活用し、関連する新しいタスクに対して深層学習モデルの汎化性能を向上させる手法である。この手法により、広範な学習データから得られた汎用的なモデルの知識を転移させることで、少量の学習データしかない場合においても深層学習モデルは高精度な分類が可能となる。

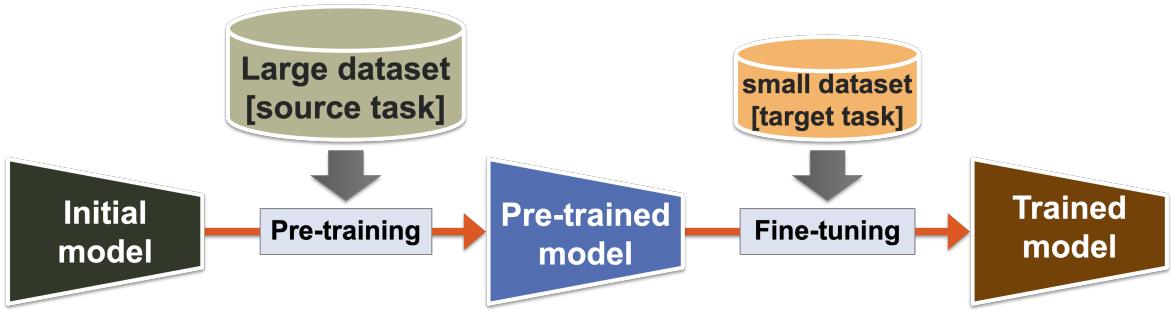
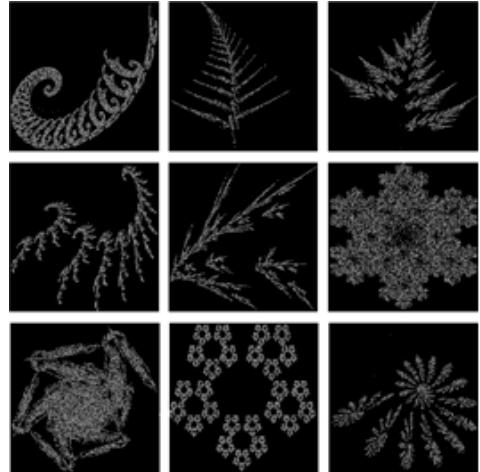


図 3.2. 転移学習の概要



(a) ImageNet の画像例



(b) FDSL の画像例

図 3.3. ImageNet と FDSL の画像例

本研究では IFOR フレームワークにおける転移学習の有効性について検証を行う。転移学習による性能向上において、事前学習のタスク（ソースタスク）と本番環境でのタスク（ターゲットタスク）の類似性は重要な要素である。そこで、IFOR のターゲットタスクが赤外線画像であることを考慮し、色情報を含まないフラクタル画像を事前学習に用いる Formula-Driven Supervised Learning (FDSL) [25] の適用可能性について評価を行う。また、事前学習データセットとして様々な画像認識タスクで標準的に用いられており、包括的な画像を含んだ大規模データセット ImageNet を用いた事前学習の有効性についても検証を行う。図 3.3 に、それぞれのデータセットにおける画像の例を示す。FDSL では、数式によって生成されたフラクタル幾何画像を用いており、色情報が含まれないため、形状特徴を強調した特徴抽出器の学習が期待できる。

### 3.2.3 メタ学習

メタ学習は、学習方法の学習として知られており、FSL における効率的なアプローチとして広く認識されている。図 3.4 にメタ学習の概要図を示す。メタ学習では、様々なタスクによって構成される学習単位をエピソードと呼び、深層学習モデルは複数のエピソードを通じて学習アルゴリズムを改善し、限られたデータに対する汎化性能を強化する。各タスクは、それぞれ  $K$  個のデータを持つ  $N$  個のクラスで構成されており、このタスク設定は “ $N$ -Way,  $K$ -Shot 分類” と呼ばれる。

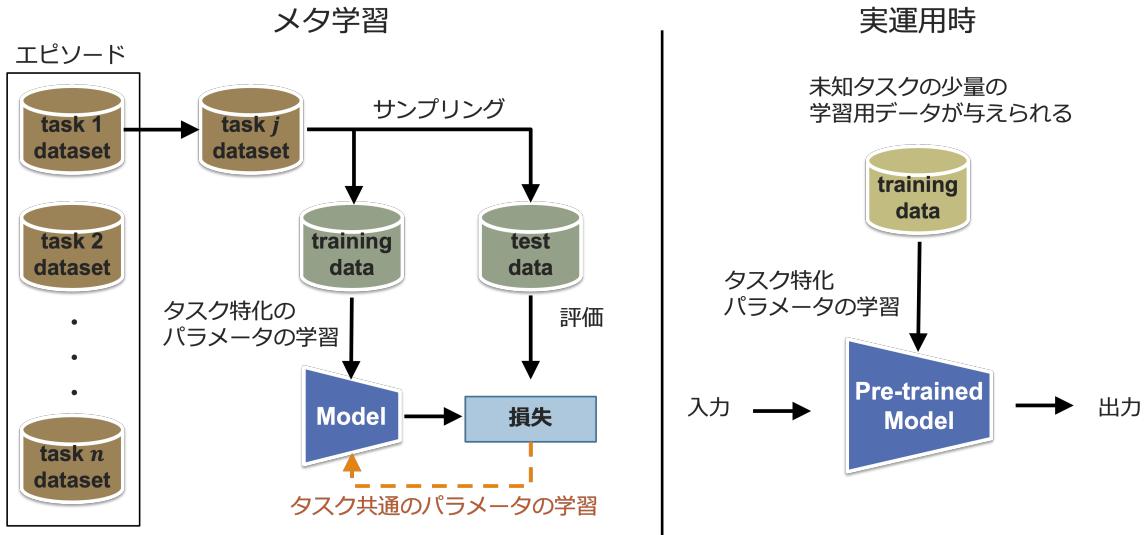


図 3.4. メタ学習の概要

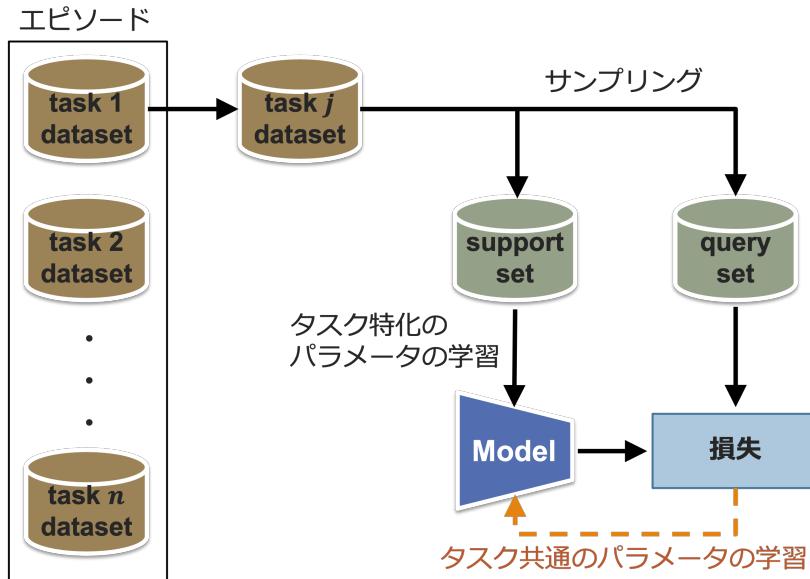


図 3.5. ProtoNet の概要

メタ学習の各エピソードでは、ランダムに選択された学習タスクに基づいてモデルパラメータが更新される。このプロセスにより、モデルは各エピソードで異なるタスクへの対応を求められ、特定のサブセットではなくより一般的な特徴表現の獲得が期待される。

Snell らは、FSL の代表的なメタ学習手法である Prototypical Networks (ProtoNet) を提案した [51]。ProtoNet の概要を図 3.5 に示す。ProtoNet では、モデルに登録するクラスセットであるサポートセット (support-set) と、サポートセットを評価するためのクエリセット (query-set) を用いて学習を行う。ProtoNet は、入力データと各クラスのプロトタイプとの距離に基づいて分類のための特徴空間を学習し、少数データにおける分類を実現する。プロトタイプはサポートセットの埋め込みベクトルの平均として定義される。具体的に、サポートデータは各クラスのプロトタイプを中心としたクラスタを形成するような空間に埋め込まれ、分類時には、クエリデータの埋め込みベクトルに最も近いプロトタイプを持つクラスが予測クラスとして分類される。こ

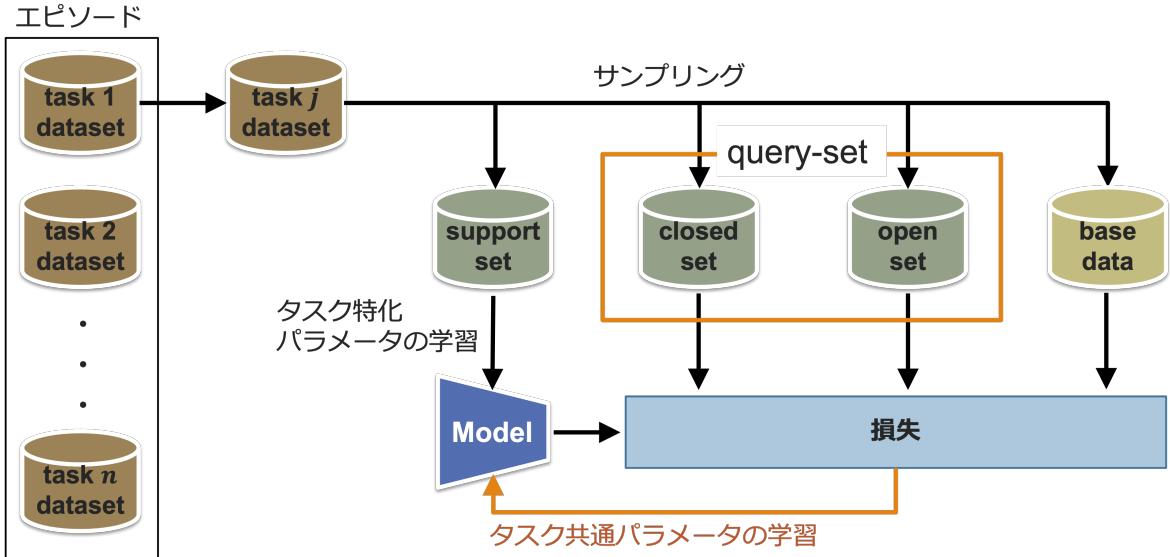


図 3.6. PEELER の概要

のような距離に基づく分類手法により、FSLにおいて課題となる過学習に対処している。

近年、メタ学習アルゴリズムが Few-Shot Open-Set Recognition (FSOSR) の分野に拡張され、登録クラスの分類と未登録クラスの検出の両方を同時に高い精度で実現する手法が提案された。Liu らは、モデルの学習過程で登録クラスの分類と未登録クラスの検出に取り組む PEELER アルゴリズムを提案した [21]。従来のソフトマックス分類器は、学習クラスを過剰に適合させる傾向があるため、未登録クラスの検出が困難であった。PEELER はこの課題に対し、エピソードごとに新規クラスをランダムに選択し、これらのクラスの事後エントロピーを最大化することにより未登録クラスの検出能力の向上を図っている。さらに、メタ学習をオープンセット認識に拡張したことにより、より一般化された特徴抽出における表現力を獲得し、認識タスクの様々なスケールや複雑さに対して効果的な学習フレームワークを提供する。

PEELER の概要を図 3.6 に示す。各タスクではサポートセット (support-set) と呼ばれる登録用データとクエリセット (query-set) と呼ばれる評価用データを使用する。 $N$ -Way,  $K$ -Shot 分類におけるサポートセットは  $\mathcal{D}^S = \{\mathbf{x}_i^S, y_i^S\}_{i=1}^{NK}$  と表される。ここで、 $\mathbf{x}_i^S \in \mathcal{X}^S$  はサポートセットの入力画像空間  $\mathcal{X}^S$  における入力画像であり、 $y_i^S \in \mathcal{Y}^S$  は登録クラス空間  $\mathcal{Y}^S$  における教師ラベルを示す。また、 $N$  はサポートセットのクラス数、 $K$  は各クラスのサンプル数を表す。さらに、クエリセットはサポートセットと同じクラスから構成されるクローズドクエリセット (closed-query set) と、サポートセットと異なるクラスから構築されるオープンクエリセット (open-query set) の 2 つに分けられる。クローズドクエリセットは  $\mathcal{D}^C = \{\mathbf{x}_i^C \in \mathcal{X}^C, y_i^C \in \mathcal{Y}^S\}_{i=1}^{NQ}$  と表される。ここで、 $\mathbf{x}_i^C$  はクローズドクエリセットの入力画像空間  $\mathcal{X}^C$  における入力画像であり、 $y_i^C$  は登録クラス空間  $\mathcal{Y}^S$  における教師ラベルを示す。また、 $N$  はクローズドクエリセットのクラス数、 $Q$  は各クラスのサンプル数を表す。一方で、オープンクエリセットは  $\mathcal{D}^O = \{\mathbf{x}_i^O \in \mathcal{X}^O, y_i^O \in \mathcal{Y}^O\}_{i=1}^{NU}$  と表される。ここで、 $\mathbf{x}_i^O$  はオープンクエリセットの入力画像空間  $\mathcal{X}^O$  における入力画像であり、 $y_i^O$  は未登録クラス空間  $\mathcal{Y}^O$  における教師ラベルを示す。また、オープンクエリセットはモデルに未登録のデータ集合であるため  $\mathcal{Y}^S \cap \mathcal{Y}^O = \emptyset$  が成り立つ。サポートセットやクローズドクエリセットがクラス数を明示的に定義するのに対し、オープンクエリセットは未登録クラスの検出を目的とするため、クラス数を定義せず総サンプル数  $N^U$  のみで表される。PEELER ではこの未登

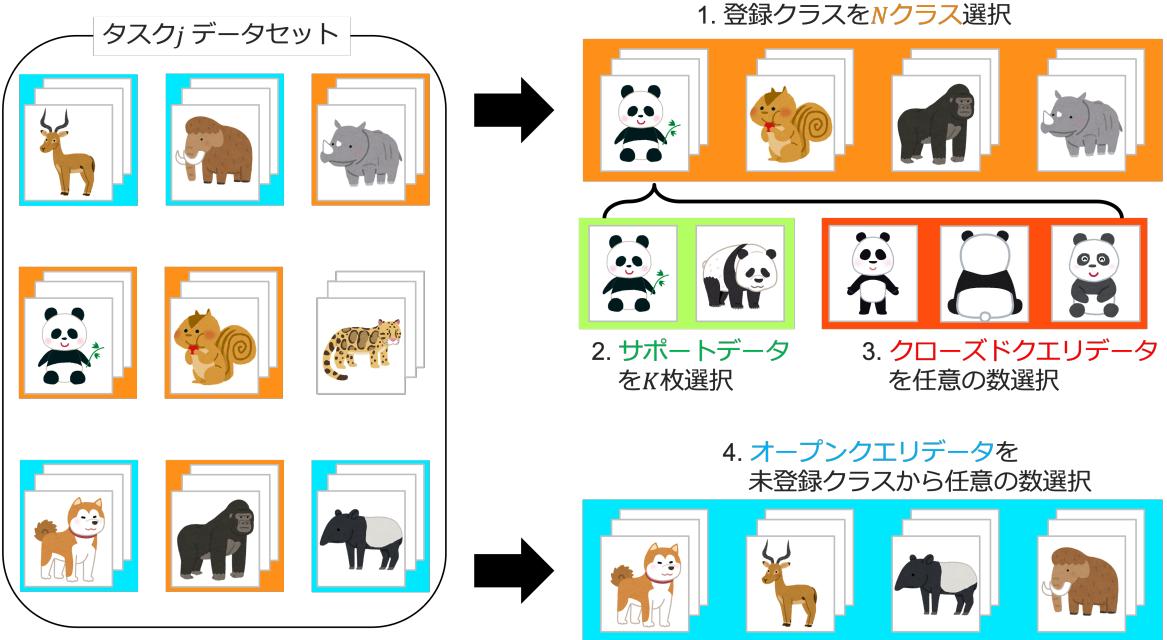


図 3.7. 登録クラス・未登録クラスの選択方法

録クラスを单一クラスとして定義することにより、未登録クラスの検出に焦点を当てている。サポートデータ、クローズドクエリデータ、オープンクエリデータの選択方法の概要を図 3.7 に示している。PEELER は、プロトタイプと入力データとの距離に基づいて分類を行う。具体的には、クエリデータが各プロトタイプの閾値より大きい場合は未登録クラス、閾値よりも小さい場合は最も近いプロトタイプのクラスに分類される。PEELER はモデルに登録されたクラスの分類と未登録クラスの検出において高い精度を達成するため、FSL 損失、OSR 損失、分類損失の 3 つの異なる損失関数を採用している。FSL 損失は、プロトタイプとクローズドクエリセットを近づけることで、少數データにおける登録クラスの分類性能を向上させる。OSR 損失はプロトタイプとオープンクエリセットを離すことで、未登録クラスの検出精度を向上させる。最後に、分類損失は、ランダムな画像から構成されるベースデータから適切な特徴を抽出し、モデルの分類能力を最適化するように設計されている。エピソードを通してタスク間で異なるクラスセットを学習することにより、モデルは特定のタスクではなく、タスク間の共通性を学習することが期待される。

FSL 損失の導出には、学習用データセットから選択されるサポートセットとクローズドクエリセットが用いられる。メタ学習の過程において、クローズドクエリデータから抽出された特徴ベクトルをサポートデータの正解クラスに近づける学習を行うことで、モデルは正確な分類を実現する特徴マッピングが習得可能となる。図 3.8 では、FSL 損失で学習することにより効果的に登録クラスの分類が可能になる例を示している。以下に具体的な FSL 損失の導出過程を述べる。

まず、プロトタイプとクローズドクエリデータの特徴ベクトル間のユークリッド距離を計算する。

$$dist(f_\phi(\mathbf{x}^C), \mu_k) = (f_\phi(\mathbf{x}^C) - \mu_k)^\top (f_\phi(\mathbf{x}^C) - \mu_k) \quad (3.1)$$

ここで、 $f_\phi(\mathbf{x}^C) \in \mathcal{F}$  はクローズドクエリセットにおける入力画像  $\mathbf{x}^C \in \mathcal{X}^C$  の特徴ベクトルを表す。ただし、 $\mathcal{X} \subseteq \mathbb{R}^D$  は  $D$  次元の入力画像空間、 $\mathcal{F} \subseteq \mathbb{R}^V$  は  $V$  次元の特徴空間を表す。また、 $f_\phi : \mathcal{X} \rightarrow \mathcal{F}$  はニューラルネットワークによる特徴抽出器であり、 $\phi$  は学習可能なパラメータ

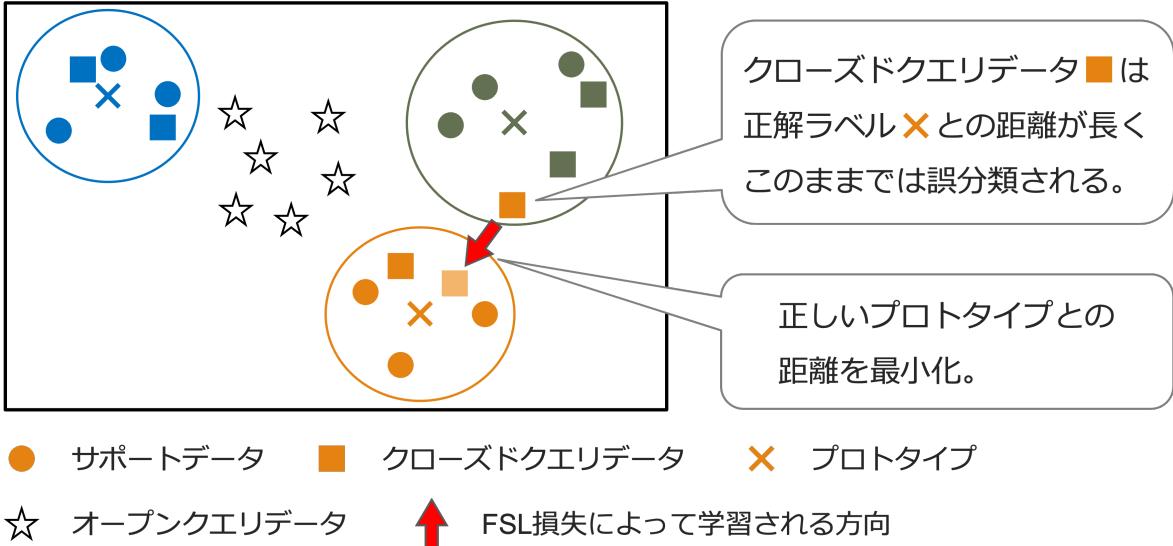


図 3.8. FSL 損失での学習が登録クラスの分類精度を向上させる例

である.  $\mu_k$  は  $k$  番目のクラスのプロトタイプを表し, 以下の式で算出される.

$$\mu_k = \frac{1}{K} \sum_{\mathbf{x}_i^S \in \mathcal{X}_k^S} f_\phi(\mathbf{x}_i^S) \quad (3.2)$$

ここで,  $\mathcal{X}_k^S$  はクラス  $k$  におけるサポートデータ集合である.

次に, ユークリッド距離に負の符号を付けてソフトマックス関数に適用する.

$$p_\phi(y^C = k | \mathbf{x}^C, M) = \frac{\exp(-dist(f_\phi(\mathbf{x}^C), \mu_k))}{\sum_{i \in \mathcal{Y}^S} \exp(-dist(f_\phi(\mathbf{x}^C), \mu_i))} \quad (3.3)$$

ここで,  $p_\phi(\cdot | \cdot, \cdot)$  は分類確率を表し,  $M = \{\mu_0, \mu_1, \dots, \mu_{N-1}\}$  はプロトタイプ集合を示す. 式 3.3 より, サポートデータとクローズドクエリデータの特徴ベクトル間のユークリッド距離が短いほど, 正しく分類できる確率が高くなることが分かる. よって, 深層学習モデルの学習において, プロトタイプとクローズドクエリデータ間の距離が長い場合に大きな損失を与えることが望ましい. 最終的に, FSL 損失はクロスエントロピー損失を用いて以下のように計算される.

$$\mathcal{L}_{\text{FSL}}[y^C, \mathbf{x}^C] = \sum_{(\mathbf{x}_i^C, y_i^C) \in \mathcal{D}^C} -\log p_\phi(y_i^C | \mathbf{x}_i^C, M) \quad (3.4)$$

次に, OSR 損失の計算の際は, 学習用データセットから選択されるサポートセットとオープンクエリセットが用いられる. ここで用いるサポートセットとは, 前述した FSL 損失の計算時に利用したものと同一のサンプルセットである. 一方, オープンクエリセットは, 未登録クラスの検出能力を評価するためのサンプルセットであり, 学習用データセットからサポートセットと異なるクラスの画像がランダムに選択される.

メタ学習の過程において, オープンクエリデータから抽出された特徴ベクトルを全てのプロトタイプから遠ざけることにより, 深層学習モデルは正確な未登録の検出能力を習得することが期待される. 図 3.9 に, OSR 損失で学習することにより効果的に未登録クラスの検出が可能になる例を示す. 以下に具体的な OSR 損失の導出過程を述べる.

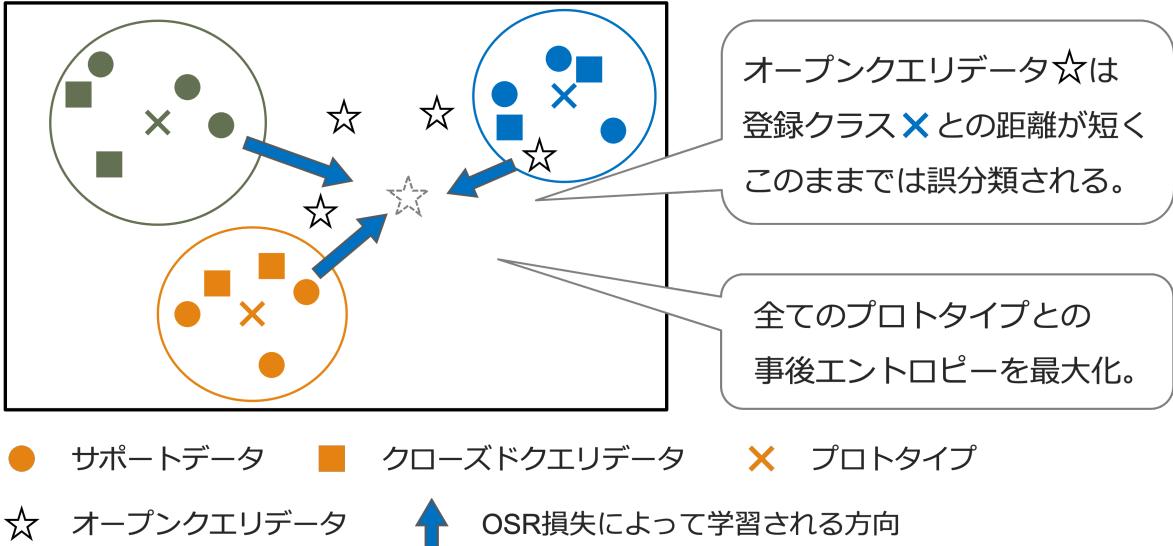


図 3.9. OSR 損失での学習が未登録クラスの検出精度を向上させる例

モデルは、正しく未登録の検出を行うために、未登録クラスからのサンプルに遭遇した際、サポートデータのどのクラスにおいても大きな確率を割り当てるべきではない。この場合、プロトタイプ集合に対するオープンクエリセットの最大のクラス確率  $\max_{k \in \mathcal{Y}^S} p_\phi(k | \mathbf{x}^O, M)$  が小さければ、未登録クラスのサンプルを適切に棄却することができる。この目的を達成するため、PEELER アルゴリズムでは、オープンクエリセットのサンプルに対して、登録済みクラスへの分類確率の最小化を図る。これは、オープンクエリデータの事後エントロピーを最大化すること、すなわち負のエントロピーを用いることで実現可能である。この最適化を実現するための損失関数として、OSR 損失は以下のように計算される。

$$\mathcal{L}_{\text{OSR}}[\mathbf{x}^O] = \sum_{k \in \mathcal{Y}^S} p_\phi(k | \mathbf{x}^O, M) \log p_\phi(k | \mathbf{x}^O, M) \quad (3.5)$$

最後に、分類損失の導出では、タスク  $j$  データセットから任意の数の画像枚数がベースデータとして選択される。ベースデータは  $\mathcal{D}^B = \{\mathbf{x}_i^B \in \mathcal{X}^B, y_i^B \in \mathcal{Y}^B\}_{i=1}^{IJ}$  と表される。ここで、 $\mathbf{x}_i^B$  はベースデータの入力画像空間  $\mathcal{X}^B$  における入力画像であり、 $y_i^B$  は教師ラベル空間  $\mathcal{Y}^B$  における教師ラベルを示す。また、 $I$  はベースデータのクラス数、 $J$  は各クラスのサンプル数を表す。この分類損失は、モデルが新しいドメインにおける分類タスクに対して、一般的かつ有用な特徴抽出を学習するために使用される。モデルは、ベースデータから特徴抽出を行う際は特徴空間上での距離による分類ではなく、分類ヘッドを用いて学習を進める。これは、学習用データセットに含まれる全てのクラスの分類問題を解くことと同義である。分類確率の計算は、以下のソフトマックス関数を用いて行われる。

$$p(y^B = k | \mathbf{x}^B; \phi, \mathbf{w}_k) = \frac{\exp(\mathbf{w}_k^\top f_\phi(\mathbf{x}^B))}{\sum_{i \in \mathcal{Y}^B} \exp(\mathbf{w}_i^\top f_\phi(\mathbf{x}^B))} \quad (3.6)$$

ここで、 $\mathbf{w}_k$  は特徴抽出器の重みベクトルを表す。よって、分類損失はクロスエントロピー損失を用いて以下のように表される。

$$\mathcal{L}_{\text{base}}[y^B, \mathbf{x}^B] = \sum_{(\mathbf{x}_i^B, y_i^B) \in \mathcal{D}^B} -\log p(y_i^B | \mathbf{x}_i^B) \quad (3.7)$$

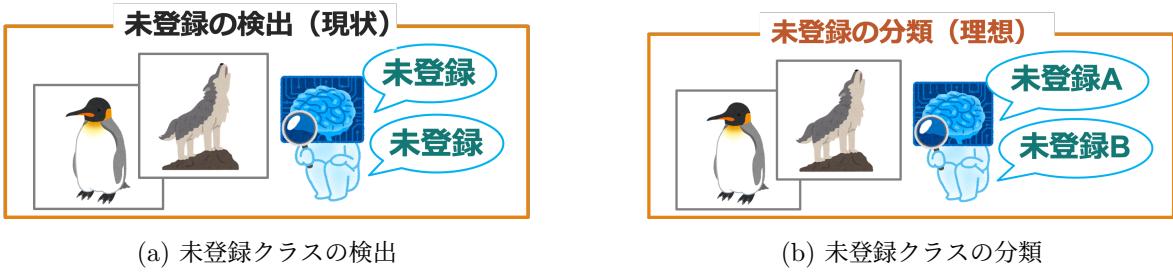


図 3.10. 未登録クラスの多クラス分類

最終的に, FSL 損失, OSR 損失及び分類損失を線型結合し, バックプロパゲーションすることによりモデルのパラメータを更新する. この学習手法により, 深層学習モデルは登録クラスの正確な分類と未登録クラスの検出の両方を効果的に実現することが可能となる. 具体的には以下の最適化問題を解くことにより,  $e \in \{1, 2, \dots, N_e\}$  エピソードにおけるモデル  $h^*$  を更新する.

$$h^* = \arg \min_h \left\{ \sum_{(x_i, y_i) \in \mathcal{D}^C | y_i \in \mathcal{Y}^S} \mathcal{L}_{\text{FSL}}[y_i, h'(x_i)] + \lambda \sum_{(x_i, y_i) \in \mathcal{D}^O | y_i \in \mathcal{Y}^O} \mathcal{L}_{\text{OSR}}[h'(x_i)] + \sigma \sum_{(x_i, y_i) \in \mathcal{D}^B | y_i \in \mathcal{Y}^B} \mathcal{L}_{\text{base}}[y_i^B, x_i^B] \right\} \quad (3.8)$$

ここで,  $h'$  は  $e$  エピソードにおいてサポートセットが登録されたモデルであり, 学習アルゴリズム  $\mathcal{M}(\cdot)$ ,  $e - 1$  エピソードにおけるモデル  $h$  を用いて以下のように表される.

$$h' = \mathcal{M}(h, D^S) \quad (3.9)$$

本研究では, FSOSR に用いられるメタ学習手法の 1 つである PEELER を 3.1 節で提案した IFOR に適用し, その有効性を検証する. FSOSR と比較して, IFOR はターゲットタスクが赤外線画像であることや, 学習時と評価時に異なるデータセットを使用しているためドメインシフトが存在するなど, より厳しい問題設定となっている. したがって, これらの本質的に異なる問題設定に対して, メタ学習アプローチの汎用性と頑健性を実証的に評価する.

### 3.3 未登録クラスに対する多クラス分類の高精度化に向けたクラスタリングに基づく損失関数

#### 3.3.1 メタ学習にクラスタリングを導入する目的

既存の OSR や FSOSR は, 未登録クラスの検出に取り組んでおり, 多クラス分類に適した特徴空間の構築が十分に行われていなかった. 図 3.10 に未登録クラスの多クラス分類の概要を示す. 3.2 節では, 既存手法と同様に図 3.10(a) のような未登録クラスの検出に取り組んでいた. これに対し本説では IFOR を発展させ, 図 3.10(b) のような未登録クラスに対する多クラス分類精度の向上に焦点を当てている. 図 3.11 は, 学習時における特徴空間上の各クラス分布を示す 2 つの例である. 四角形の枠線は特徴空間, 点は特徴ベクトル, 点を含む丸い枠線はクラス分布領域の境界を表している. 図 3.11(a) 及び図 3.11(b) に示された特徴空間は, いずれも登録済み

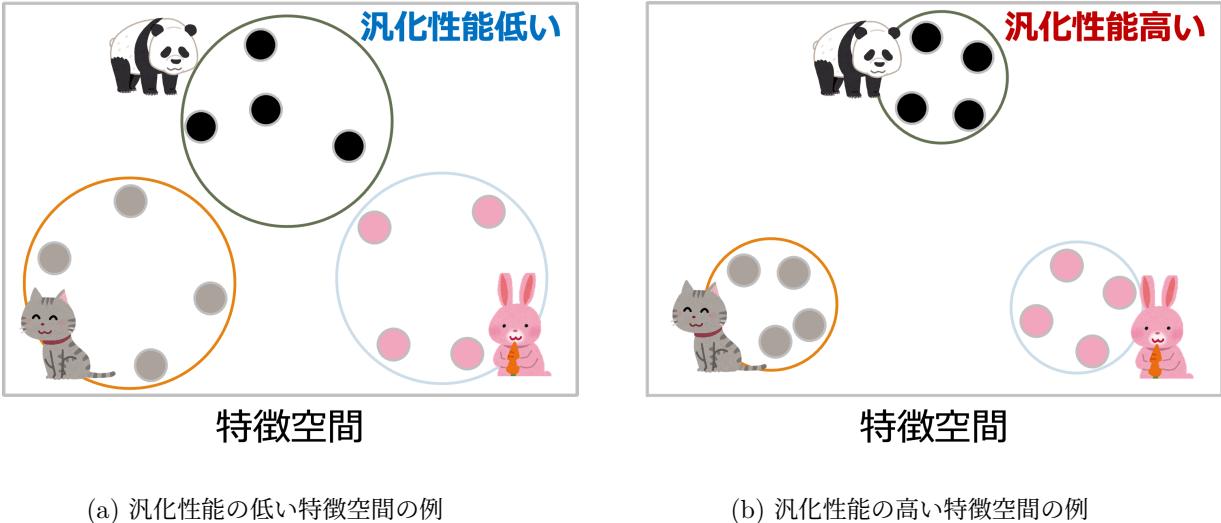


図 3.11. 学習時における特徴空間上の各クラス分布の例

のクラスに対して同程度の分類精度であるが、評価時のドメインシフトなどを考慮した場合、図 3.11(b) の特徴空間の方が、新規地域に対する高い汎化性能を有していると考えられる。これは、図 3.11(b) の方がクラス内分散が小さく、かつクラス間分散が大きいため、新規地域においても分類が容易となるような特徴空間の構築が期待できるからである。

したがって、本研究では、メタ学習にクラスタリングに基づく損失関数を導入することにより、クエリセットに対するクラス内分散の最小化・クラス間分散の最大化を実現する。これにより、IFOR における未登録クラスの多クラス化に向けて分類精度の向上を図る。

### 3.3.2 損失関数

本研究では、FSL 損失、OSR 損失、分類損失に加え、3.3.1 で述べたクラスタリングに基づく損失関数として k-means 損失と Between-Class 損失（BC 損失）を導入する。k-means 損失は、Chin ら [26] が異常検知タスクにおいて提案した損失関数であり、k-means クラスタリングによって集約される類似した性質を持つ特徴量から、より識別的な特徴表現の学習を可能にする。IFOR フレームワークにおいて、k-means 損失は以下のように定義される。

$$\mathcal{L}_{\text{k-means}} = \sum_i \min_k \|f(x_i) - c_k\|_2 \quad (3.10)$$

ここで、 $f(x_i)$  は  $i$  番目の入力画像を特徴抽出器  $f(\cdot)$  に入力した際の特徴量、 $c_k$  は  $k$  番目のクラスタ中心を表す。

k-means 損失による学習の例を図 3.12 に示す。この損失関数の最小化により、各クラスタ中心とそのクラスタに属する特徴ベクトルとの距離が最小となることが期待される。本研究では、類似した性質を持つ特徴量のクラスタリングにより、特徴空間上の各クラスのクラス内分散を最小化することを目指し、k-means 損失の有効性を検証する。

一方、BC 損失は、クラス分布のコンパクトな表現に加え、各クラスの分布が可能な限り離れているような特徴空間の構築が、多クラス分類性能の向上に寄与するという考えに基づいている。IFOR フレームワークにおいて、BC 損失は以下のように定義される。

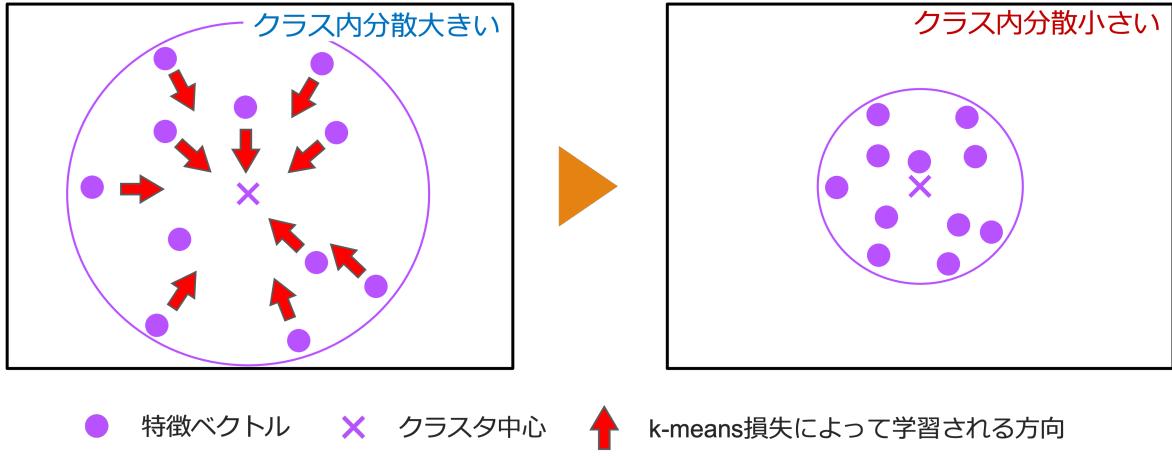


図 3.12. k-means 損失によってクラス内分散が小さくなる例

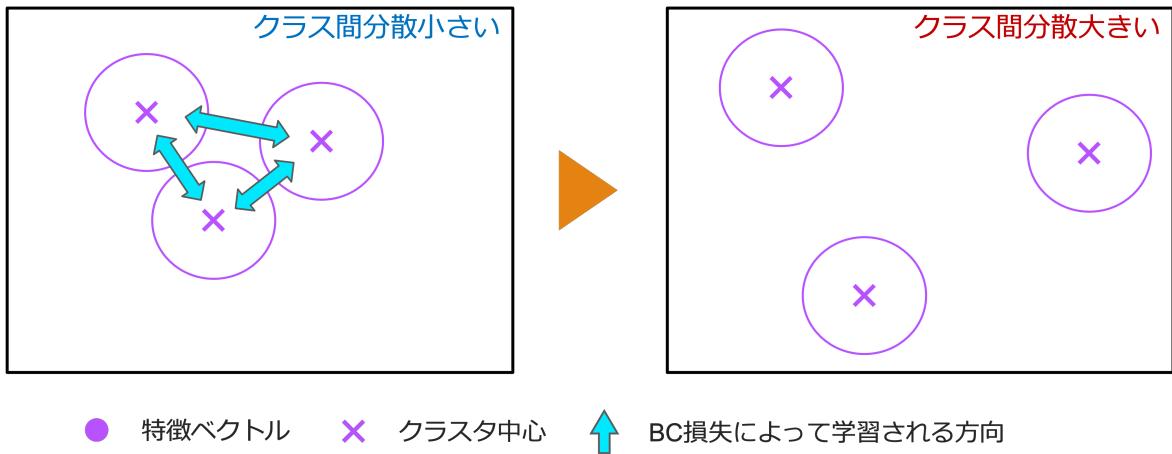


図 3.13. Between-Class 損失によってクラス間分散が大きくなる例

$$\mathcal{L}_{\text{Between-Class}} = -\log \sum_{k_1} \sum_{k_2} \|c_{k_1} - c_{k_2}\|_2 \quad (3.11)$$

ここで,  $c_{k_1}$  と  $c_{k_2}$  は  $k_1$  番目,  $k_2$  番目のクラスタ中心を表す. 負の符号を付与することにより, クラス間分散の最大化問題を損失関数の最小化問題として扱っている.

BC 損失による学習の例を図 3.13 に示す. この損失関数では, k-means クラスタリングによって得られる各クラスタ中心間の距離を最大化することにより, クラス間分散の最大化を図っている.

## 第 4 章

# 評価実験

### 4.1 データセット

IFOR 手法の開発において、地域間のドメインシフトを考慮したモデルの性能評価は重要である。このドメインシフトを実現するため、学習用と評価用のデータセットを異なる地域から選定した。具体的には、学習用データセットとして南米の野生動物画像を集めた WCS Camera Traps (WCS) [52] を、評価用データセットとして北米の野生動物画像から構成される Caltech Camera Traps (CCT) [53] をそれぞれ採用した。

本実験では、提案手法の有効性を多角的に検証するため、各データセットにおいて赤外線画像と可視光画像それぞれから構成される 2 種類のデータセットを作成した。このアプローチにより、提案手法が赤外線画像特有の性質に対して有効であるのか、あるいは、赤外線画像や可視光画像を問わず広義の画像分類に対して有効であるのかを検証することが可能となる。

データセットの前処理では、各画像を赤外線画像と可視光画像に分類し、アノテーションとして提供されているバウンディングボックスに基づいてクロッピングを行い、動物が存在している領域を切り出した。バウンディングボックスに基づき動物領域が切り出された例を図 4.1 に示す。その後、実験の公平性を確保するため、赤外線画像と可視光画像による学習用データセット間でクラス数と画像枚数を同数に統一した。同様に、評価用データセットについてもクラス数と画像

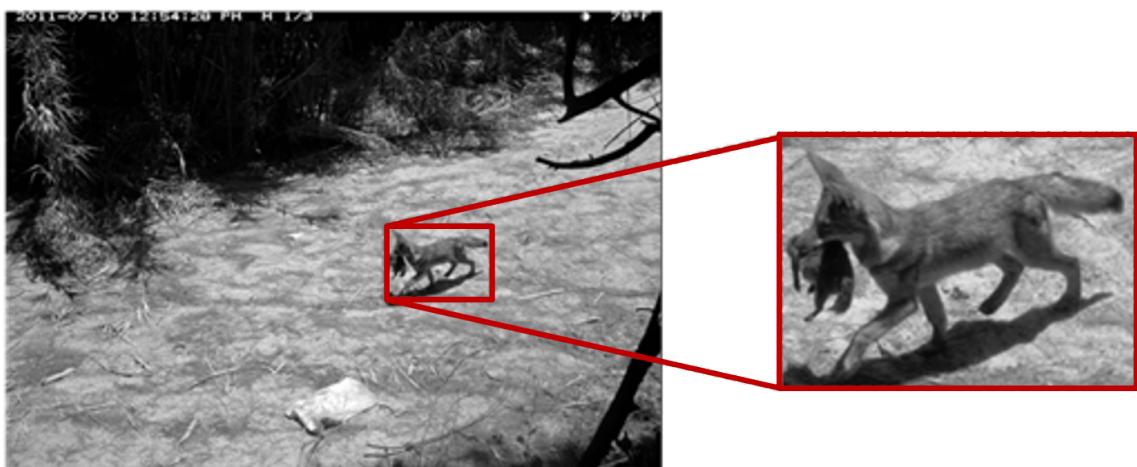


図 4.1. バウンディングボックスによって動物領域が切り出されている例

表 4.1. 学習用 (WCS Camera Traps) データセットの詳細

動物種	赤外線画像 (枚)	可視光画像 (枚)
<i>acryllium vulturinum</i>	500	500
<i>aepycceros melampus</i>	500	500
<i>bos taurus</i>	500	500
<i>capra aegagrus</i>	<b>337</b>	500
<i>cephalophus nigrifrons</i>	500	500
<i>crax rubra</i>	500	500
<i>cuniculus paca</i>	500	<b>495</b>
<i>dasyprocta punctata</i>	500	500
<i>equus quagga</i>	500	500
<i>giraffa camelopardalis</i>	500	500
<i>leopardus pardalis</i>	500	500
<i>loxodonta africana</i>	500	500
<i>madoqua guentheri</i>	500	500
<i>mazama americana</i>	500	500
<i>mazama temama</i>	500	500
<i>meleagris ocellata</i>	500	500
<i>mitu tuberosum</i>	500	500
<i>nasua narica</i>	<b>350</b>	500
<i>panthera onca</i>	500	500
<i>papio anubis</i>	500	500
<i>pecari tajacu</i>	500	500
<i>phacochoerus africanus</i>	500	<b>317</b>
<i>psophia leucoptera</i>	500	500
<i>puma concolor</i>	500	500
<i>syncerus caffer</i>	500	500
<i>tapirus terrestris</i>	500	500
<i>tayassu pecari</i>	500	500
<i>urocyon cinereoargenteus</i>	500	500
合計	13,687	13,812

枚数を同一に設定した。具体的には、学習用データセットは 28 クラス、各クラス約 500 枚ずつの合計 14,000 枚の画像で構成され、評価用データセットは 11 クラス、各クラス 100 枚ずつの合計 1,100 枚の画像により構成される。これらの処理により、学習用データセットとして赤外線画像及び可視光画像による 2 種類の WCS データセット、評価用データセットとして赤外線画像及び可視光画像による 2 種類の CCT データセットの計 4 種類のデータセットを作成した。ただし、画像の前処理過程において、学習用データセットでは赤外線画像が 13,687 枚、可視光画像が 13,812 枚と若干の不均衡が生じたが、この僅かな差異は実験結果に重大な影響を及ぼさないと考えられる。これらの前処理によって作成された各データセットの詳細を表 4.1 及び表 4.2 に示す。

表 4.2. 評価用 (Caltech Camera Traps) データセットの詳細

動物種	赤外線画像（枚）	可視光画像（枚）
bird	100	100
bobcat	100	100
cat	100	100
coyote	100	100
deer	100	100
dog	100	100
fox	100	100
opossum	100	100
rabbit	100	100
skunk	100	100
squirrel	100	100
合計	1,100	1,100

## 4.2 未登録クラスの検出に対する提案手法の評価

### 4.2.1 実験条件

本実験では、IFOR フレームワークにおける特徴抽出器の有効性を評価するため、異なる性質を持つ CNN ベース及び ViT ベースの特徴抽出器について検証を行う。CNN ベースのモデルには、18 層の深さと約 1,170 万のパラメータを有する ResNet18 [49] を採用した。ResNet18 は、小規模から中規模のデータセットに対して、十分な分類性能を維持しながら効率的な処理が可能なアーキテクチャとして知られている。一方、ViT ベースのモデルとして、Data-efficient image Transformers (DeiT) [54] の最軽量モデルである Deit-Ti を利用した。このアーキテクチャは、約 500 万のパラメータを有し、モデルのスケールと計算効率の両立を実現している。DeiT は、入力画像を  $16 \times 16$  pix. のパッチに分割し、自己注意機構を適用することによって特徴抽出を行う。これらのモデルは ImageNet データセットまたは FDSL を用いた事前学習が適用されており、それぞれの転移学習手法が IFOR フレームワークのモデル性能に与える影響について検証することが可能である。

IFOR では、特定の地域に生息する野生動物の画像を大量に収集することが困難な状況を想定している。このような実世界での制約を考慮し、本研究で用いるデータ設定は極めて厳しい条件である 5-Way, 1-Shot 問題として定義した。本実験設定は、新規地域において、わずか 1 クラスあたり 1 枚の画像を収集するだけでモデルが利用できる状況を仮定しており、システムの初期導入時に想定される最小限のデータ条件である。

学習の際には、各エピソードにつき学習用データセットから 10 クラスがランダムに選択される。このうち 5 クラスは登録クラス、残りの 5 クラスは未登録クラスとして設定される。サポートセットとクローズドクエリセットには登録クラスからのデータのみが使用され、オープンクエリセットには未登録クラスからのデータが割り当てられる。

メタ学習の各学習エピソードでは、学習用データセットからサポートセットとして 1 クラスに

つき 1 枚, クローズドクエリセットとして 1 クラスにつき 15 枚, オープンクエリセットとして 1 クラスにつき 15 枚の画像を用いる. これに加えて, 75 枚の画像がベースデータとしてランダムに選択される. 最適化関数には Adam を採用し, 入力画像は  $224 \times 224$  pix. にリサイズを行った. 学習エピソードの総数は 30,000 エピソードとし, 学習過程でモデルが局所最適解に陥るのを防ぐため, 10,000 エピソード時と 20,000 エピソード時に学習率を 0.1 倍ずつ減少させた. 本実験では, 様々な転移学習手法や特徴抽出器の組み合わせを検証するため, 各実験設定における最適な初期学習率が異なる. そこで, 初期学習率を  $10^{-6}$  から  $10^{-2}$  まで 10 倍ずつ増加させて最適な学習率の探索を行い, 最も高い精度が得られた重みを採用した.

評価実験では, 新規地域で収集された限られたデータによる実運用を想定し, 評価用データセットを用いて実用的な条件下でのモデルの性能評価を行った. 本実験は 5-Way, 1-Shot 問題として定義され, 5 つのクラスにそれぞれ 1 枚ずつ, 合計 5 枚のサポートセットによる評価を行った. クエリセットも学習時と同様に, サポートセットと同一の 5 クラスを登録クラス, サポートセットと異なる 5 クラスを未登録クラスとして設定した. 具体的な評価手順は次の通りである.

まず, 評価用データセットからランダムに選択された 5 つのクラスから, それぞれ 1 枚の画像をプロトタイプとしてモデルに登録する. 次に, 特徴抽出器によって得られたクローズドクエリセットの特徴ベクトルを距離が最も近いプロトタイプのクラスへと分類し, 正しく分類できた割合に基づいて分類精度を測定する. オープンクエリデータを未登録クラスとして検出するモデルの能力は, AUROC (Area Under the Receiver Operating Characteristic Curve) 指標により評価される. AUROC は, クエリデータとプロトタイプ間の距離スコアに対する閾値を変化させた際の, 偽陽性率 (False Positive Rate, FPR) と真陽性率 (True Positive Rate, TPR) の関係をプロットしたグラフの曲線下の面積として定義されている. この指標は値が 100% に近づくほど, 未登録クラスに対する検出精度が高いことを示す. ここで, TPR はオープンクエリデータを正しく未登録クラスとして検出できたサンプルの割合を表し, FPR は登録クラスを誤って未登録クラスとして検出したサンプルの割合を示す. 評価用データセットからのクラス選択が特定のクラスに集中することで生じるバイアスなどを排除し, 実験結果の妥当性を担保するため, この評価手順を 10,000 エピソードにわたり実施した.

#### 4.2.2 実験結果及び考察

本実験では, 異なる特徴抽出器である ResNet18 と ViT について, ImageNet 転移学習, フラクタル転移学習, 転移学習なしの 3 つの実験条件で比較を行った. なお, 転移学習なしの条件では, モデルの重みをランダムに初期化し, WCS データセットを用いたメタ学習によりモデルのパラメータをスクラッチから更新した. これらの条件下における IFOR に対するそれぞれの実験結果を表 4.3 に示す. 表 4.3 に示す全ての実験において, モデルの学習にメタ学習を適用した. 各特徴抽出器に対する実験結果から, 赤外線画像の分類において ViT が ResNet18 よりも高い精度を示すことが明らかとなった. ただし, 転移学習なしの条件下では, ViT は ResNet18 よりも低い精度を示した. これは, 転移学習を適用しない場合, ViT よりも CNN ベースのモデルの方が優れた性能を示すという従来の知見 [24] と一致する結果である.

表 4.3 より, ResNet18 と ViT の両モデルにおいて, ImageNet を用いた転移学習が赤外線画像と可視光画像の双方のデータセットに対して最も高い精度を示した. FDSL による転移学習は, 転移学習なしの場合と比較して, 赤外線画像と可視光画像ともに精度をわずかに改善しただけであった. このような傾向が見られた背景として, フラクタル画像のドメインが自然画像とは異なる

表 4.3. IFOR に対する各特徴抽出器と転移学習の組み合わせによる実験結果

学習方法		メタ学習 (PEELER)					
特徴抽出器		ResNet18			ViT		
転移学習		ImageNet	FDSL	なし	ImageNet	FDSL	なし
赤外線画像	分類精度 (%)	45.8	38.8	38.6	<b>51.0</b>	36.5	36.2
	AUROC (%)	58.4	54.3	56.3	<b>61.0</b>	54.4	54.5
可視光画像	分類精度 (%)	53.0	32.8	33.2	<b>60.2</b>	32.7	31.4
	AUROC (%)	60.8	55.2	54.3	<b>67.8</b>	54.6	53.9

表 4.4. ImageNet 転移学習を用いた場合の ViT による各学習方法の IFOR に対する実験結果

特徴抽出器		ViT		
転移学習		ImageNet		
学習方法		メタ学習 (PEELER)	従来法 (ミニバッチ学習)	なし
赤外線画像	分類精度 (%)	<b>51.0</b>	39.9	39.6
	AUROC (%)	<b>61.0</b>	53.3	55.0

ることから、赤外線画像や可視光画像の動物分類タスクにおいて効果が限定的となった可能性が考えられる。一方で、FDSL を用いた先行研究では、自然画像で構成される CIFAR10 データセットにおいて、FDSL によって転移学習されたモデルが ImageNet を用いて転移学習したモデルを凌駕する性能を示している [25]。この知見から、FDSL を赤外線画像分類に効果的に適用するためには、自然画像で構成されたデータセットによるファインチューニングが必要である可能性を示唆している。

本研究では、IFOR フレームワークにおけるメタ学習の影響も検証しており、その結果を表 4.4 に示す。本実験では、ImageNet で事前学習された ViT モデルを特徴抽出器として使用し、WCS データセットを用いて動物分類タスクのためにファインチューニングを行った。なお、「学習方法」によってファインチューニングのアプローチが異なることに注意が必要である。メタ学習の場合、分類ヘッドを使用せずに各エピソードで特徴抽出器をファインチューニングしている。具体的には、特徴空間におけるサポートセットとクローズドクエリセット間、並びに、サポートセットとオープンクエリセット間の距離に基づいて特徴抽出器のパラメータを更新した。一方、従来の学習手法であるミニバッチ学習の場合、28 個のクラスノードを持つ分類ヘッドを用いて、特徴抽出器を含む全ての層を更新した。この手法は、ImageNet を用いた事前学習と、ターゲットタスクデータセットを用いたミニバッチ学習を組み合わせた一般的な学習アプローチであり、本研究におけるベースラインとして位置付けされる。また、学習なしの場合、ImageNet で事前学習された重みを WCS データセットでの追加学習を行わずに使用した。本研究の評価方法は特徴空間上の距離に基づいて分類を行うため、一般的な転移学習で行われる重みの凍結や分類ヘッドの再学習は不要であった。メタ学習との公平な評価条件を確保するため、学習に使用する画像枚数を統一し、学習エポック数を 690 エポックに設定した。初期学習率は  $10^{-6}$  から  $10^{-2}$  まで 10 倍ずつ増加させて最適値の探索を行い、学習過程において、学習率は 230 エポックごとに 0.1 倍ずつ減少させた。IFOR フレームワークにおけるモデルの性能評価は、評価用データセットを用いて 10 エポックごとに検証を行い、得られた最も高い精度を最終的な実験結果として採用した。

表 4.5. IFOR における基盤モデルの追加学習の有無による分類性能の実験結果

特徴抽出器	CLIP		ViT-Base	
転移学習	WIT		ImageNet-21k	
学習方法	メタ学習 (PEELER)	なし	メタ学習 (PEELER)	なし
赤外線画像	分類精度 (%)	58.6	36.7	<b>61.1</b>
	AUROC (%)	64.8	54.9	<b>68.6</b>
				39.8
				58.0

表 4.4 は、IFOR フレームワークにおいて、ミニバッチ学習や学習なしの条件と比較して、メタ学習が分類精度と AUROC の両方を顕著に改善したことを示している。ミニバッチ学習は、学習なしの条件と比較して分類精度をわずかに向上させたものの、AUROC は低下する結果となつた。この結果は、従来の学習手法であるミニバッチ学習が 28 クラスの識別に特化した特徴抽出器の学習を行うため、特徴空間上の各クラスの分布が未登録クラスの検出に適さない形で最適化されたことを示唆している。一方、メタ学習では特徴空間における距離関係を直接的に学習することにより、登録クラスの分類と未登録クラスの検出を同時に考慮した特徴表現の獲得が可能となり、結果として AUROC の向上につながったと考えられる。

本実験を通じて、赤外線画像による動物分類タスクの複雑性が明らかとなり、特に色情報の欠如に起因する課題の重要性が示された。さらに、可視光画像と赤外線画像の分類性能の比較分析により、赤外線画像特有の課題の顕著さが確認された。

また、本実験では IFOR における基盤モデルを用いた意味的な特徴抽出の有効性についても検証を行った。本実験で用いる特徴抽出器のモデル構造には、ViT の標準的なモデルサイズである ViT-Base を採用した。基盤モデルについては、ViT ベースの CLIP モデルとして ViT-B/16 を利用し、以降、これを CLIP と表記する。一方、比較対象であるテキストを用いた学習が行われていない ViT は ViT-Base と表記する。これらのアーキテクチャは、約 8,600 万のパラメータをしており、入力画像を  $16 \times 16$  pix. のパッチに分割し特徴抽出が行われる。CLIP はインターネットを介して収集された約 4 億組の画像とテキストのペアから構成される WebImageText (WIT) という大規模データセットを使用して事前学習されている。一方で、ViT-Base は ImageNet-1k の拡張版データセットであり、21,000 クラスを含む ImageNet-21k を用いて事前学習が行われている。

表 4.5 に CLIP の追加学習の有無が赤外線画像の分類性能に与える影響を示している。実験結果から、追加学習の有無に関わらず IFOR における CLIP の性能は ViT-Base に劣ることが明らかとなった。本実験では、特徴空間上の距離に基づいた評価を行ったが、先行研究 [50] ではテキストエンコーダを用いた分類を行っている。CLIP の学習ではテキストの特徴ベクトルと画像の特徴ベクトルとの類似度に基づき分類を行っているため、画像の特徴ベクトルのみを用いた分類には不向きである可能性がある。この事実から、CLIP を IFOR フレームワークに効果的に適用するためには、テキストエンコーダを用いた類似度計算が必要である可能性を示唆している。

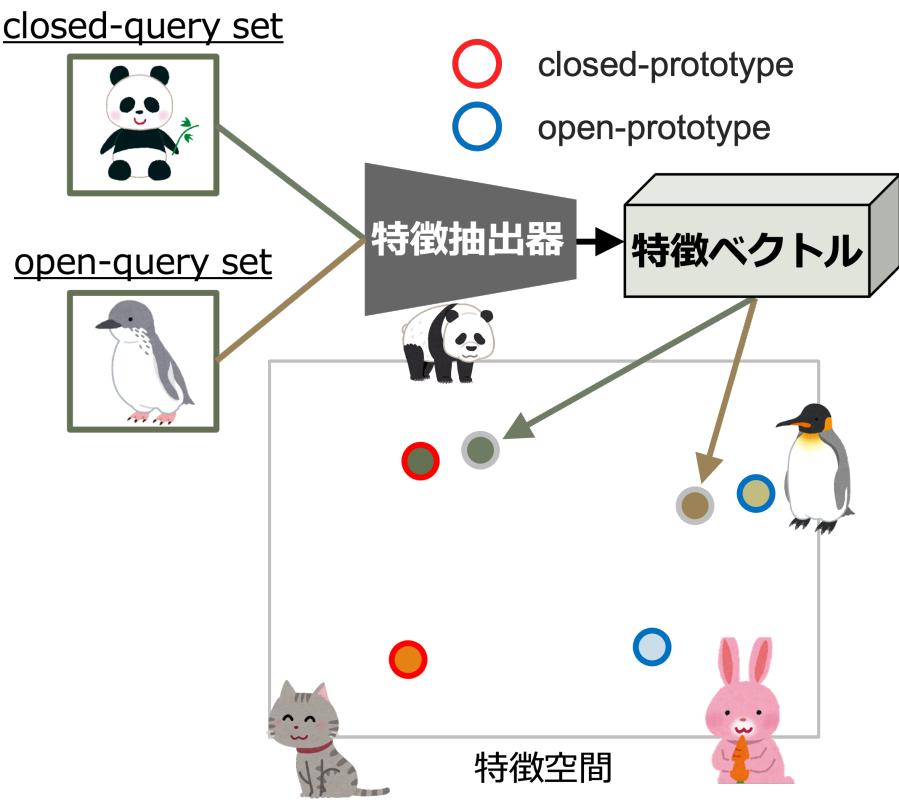


図 4.2. クローズドプロトタイプとオープンプロトタイプを用いた評価方法

### 4.3 未登録クラスの多クラス分類に対する損失関数の評価

#### 4.3.1 実験条件

未登録の動物種に対する分類精度の評価に際し、本実験では新たな評価手法を導入した。4.2 節の実験の評価時には、登録クラスから得られるプロトタイプのみを用いて分類精度を求めていたが、本実験では登録クラスのプロトタイプに加えて、未登録クラスから算出されるプロトタイプを用いて未登録クラスに対する分類精度も測る。ここで、登録クラスから構成されたサポートセットをクローズドサポートセット (closed-support set) と呼び、それから得られるプロトタイプをクローズドプロトタイプ (closed-prototype) と呼ぶ。同様に、未登録クラスから構成されたサポートセットをオープンサポートセット (open-support set) と呼び、それから得られるプロトタイプをオープンプロトタイプ (open-prototype) と呼ぶ。図 4.2 にクローズドプロトタイプとオープンプロトタイプを用いた評価方法の概要を示している。本評価方法においても、特徴抽出器に入力されたクエリセット中の画像は特徴空間上にプロットされ、距離が最も近いプロトタイプのクラスへと分類される。分類精度は、このようにしてプロトタイプベースの分類が正しく行われた割合に基づき測定される。ただし、4.2 節とは異なり、クエリデータの特徴点がオープンプロトタイプに近ければそのプロトタイプが属する未登録クラスに分類され、クローズドプロトタイプに近ければそのプロトタイプが属する登録クラスに分類される。この評価において、クローズドクエリデータのみの分類精度は Closed Accuracy、オープンクエリデータのみの分類精度は Open Accuracy、クローズドクエリデータとオープンクエリデータの分類精度を平均した精度は

表 4.6. IFOR における k-means 損失と BC 損失のアブレーション結果

学習フレームワーク	k-means	BC	Closed Accuracy (%)	Open Accuracy (%)	All Accuracy (%)	AUROC (%)
PEELER			38.3	37.8	38.1	<b>49.7</b>
	✓		38.3	37.8	38.1	<b>49.7</b>
		✓	38.3	37.8	38.1	<b>49.7</b>
	✓	✓	39.0	38.3	38.7	49.6
PEELER (w/o 分類損失)	✓		39.3	<b>39.2</b>	39.2	49.6
		✓	<b>39.5</b>	<b>39.2</b>	<b>39.3</b>	49.6
	✓	✓	39.3	<b>39.2</b>	39.2	49.6

All Accuracy として表される。

本実験の評価設定としても 5-Way, 1-Shot 問題を採用し, 評価用データセットから各エピソードにおいて 5 つの登録クラスと 5 つの未登録クラスを選択した。このうち, クローズドサポートセットとオープンサポートセットには 1 クラスにつき 1 枚の画像を選択し, クローズドクエリセットとオープンクエリセットには 1 クラスにつき 15 枚の画像を用いて評価を行った。

### 4.3.2 実験結果及び考察

表 4.6 に未登録の動物種の分類結果を示す。本実験では, 距離学習に基づく FSL 損失と OSR 損失に加え, 分類損失を組み合わせた既存の PEELER フレームワークをベースラインとし, k-means 損失, BC 損失の有効性について検証を行った。また, 全ての実験において, PyTorch Image Models (timm) ライブラリ [55] によって提供される, ImageNet で事前学習済みの ViT モデル `deit_tiny_patch16_224` を特徴抽出器として用いた。

結果より, 既存手法の PEELER に k-means 損失, または, BC 損失を個別に組み合わせた場合, モデルの分類精度はベースラインとほぼ同等の性能であった。一方, k-means 損失と BC 損失の両方をベースラインの PEELER に組み合わせることによって, ベースラインより高い性能を達成することが確認された。次に, 分類損失を用いない PEELER に k-means 損失または BC 損失を組み合わせた結果を確認すると, k-means 損失のみ, BC 損失のみ, k-means 損失及び BC 損失を組み合わせた場合のいずれにおいてもベースラインの精度より高くなることが確認された。この結果は, IFOR の 5-Way, 1-Shot 問題において, 分類損失を用いずに k-means 損失や BC 損失を単体で用いることの有効性を示している。

4.2 節の表 4.3, 4.4 並びに 4.5, 本節の表 4.6 に示された結果は, 新規地域においてサポートセットとしてわずか 1 枚の画像のみを使用するという, 極めて厳しい条件下で評価されている。しかし, システムの継続的な運用に伴い, サポートセットとして利用できる画像の数は自然に増加することが期待される。このような実運用シナリオを考慮し, 図 4.3 に示すように, 評価時の Shot 数  $K$  が分類精度と AUROC に与える影響について検証を行った。この評価では, 特徴抽出器として ViT, 転移学習手法として ImageNet, 学習方法としてメタ学習, 損失関数として FSL 損失, OSR 損失, BC 損失を採用したモデルを使用した。実験の結果, 図 4.3(a) に示すように, 1 クラスあたりのサポートデータを 30 枚, すなわち, 30-Shot まで増加させることによって分類精度が約 70% に達することを確認した。また, 表 4.7 に各カテゴリーにおける分類精度を示す。

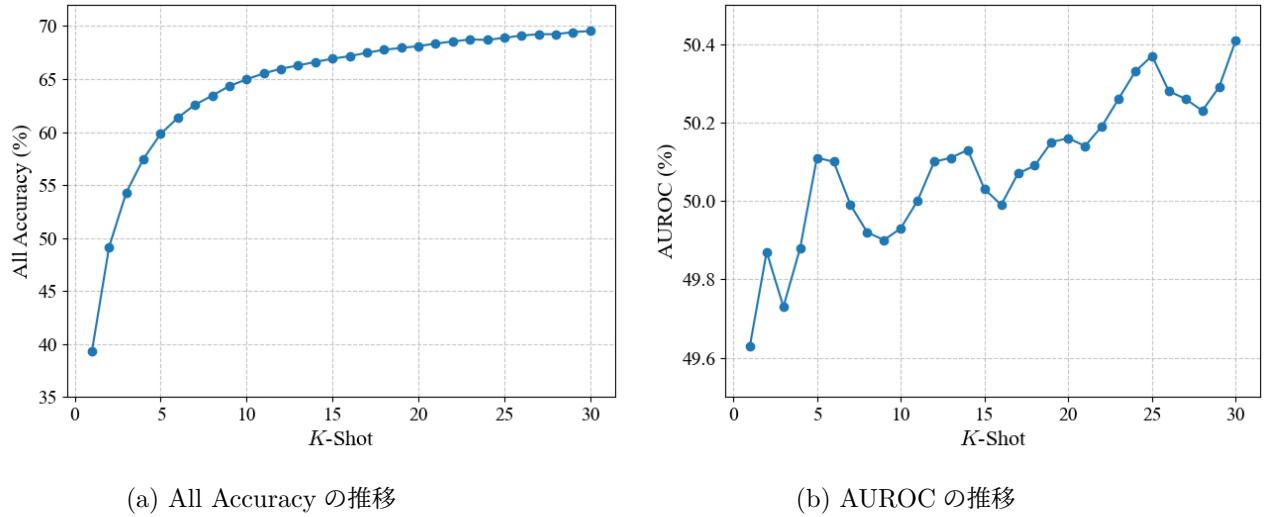


図 4.3. 評価時におけるサポートデータ数  $K$  が変化した際の All Accuracy と AUROC

表 4.7. サポートデータの Shot 数が変化した際の各カテゴリの分類精度

動物種	1-shot	5-shot	10-shot	20-shot	30-shot
bird	26.1	49.2	55.4	58.9	60.1
bobcat	36.7	55.7	59.8	61.6	64.7
cat	19.4	34.5	40.9	46.6	50.5
coyote	25.8	41.0	47.9	53.1	54.5
deer	55.4	81.9	86.1	88.0	88.1
dog	29.6	55.1	62.7	67.0	67.5
fox	34.3	47.7	54.3	58.9	60.7
opossum	47.2	70.1	75.5	79.1	80.6
rabbit	36.2	59.0	64.5	66.7	68.2
skunk	64.3	83.5	85.6	86.3	86.9
squirrel	56.7	79.1	80.9	81.6	81.9
All Accuracy	39.2	59.7	64.9	68.0	69.4

各実験結果の詳細な分析により、本研究における主要な課題として動物の姿勢による誤分類が明らかとなった。特に、動物の背面のみが撮影された画像や複雑な姿勢をとる個体の画像など、出現頻度の低い姿勢パターンに対して分類精度の低下が観察された。図 4.4 は、このような姿勢の影響により誤分類された画像の具体的な事例を示している。この課題に対する実用的な解決策としては、カメラトラップにより連続的に撮影された画像群をシーケンスデータとして扱うアプローチが有効である。複数の連続画像における時間的文脈を考慮することにより、単一の画像からでは判断が困難な姿勢に対しても高精度な分類が期待できる。この改善案は、モデルの頑健性及び実環境における適用可能性の向上に寄与しつつ、IFOR フレームワークの本質的な利点を維持することが可能である。



(a) coyote の画像例



(b) bobcat の画像例

図 4.4. 姿勢が原因で誤分類された画像例（正解カテゴリ： coyote, 予測カテゴリ： bobcat）

# 第 5 章

## 結論

本論文では、限られた赤外線画像を用いた効果的な野生動物モニタリングのための新たな概念として、Infrared Few-shot Open-set Recognition (IFOR) を提案した。本研究では、最小限の赤外線画像データを用いて、モデルに登録された動物の分類と未登録の動物の識別を実現する IFOR フレームワークを開発した。

包括的な分析により得られた主要な知見として、まず特徴抽出器について、ViT は赤外線画像の分類において CNN (ResNet18) と比較して優れた性能を示し、分類精度が 5.2% 向上した。これは、ViT の大域的な注意機構が赤外線野生動物画像に対する特徴抽出において高い有効性を持つことを示唆している。一方で、基盤モデルである CLIP は ViT-Base と比較して有効性を示せなかった。この結果を踏まえ、今後の研究課題として、テキストエンコーダやプロンプトを用いた学習フレームワークの検証が必要である。これにより、IFOR におけるテキストと画像のペアを用いた対照学習の有効性を理論的かつ実験的に明らかにすることが可能となる。転移学習については、ImageNet による事前学習が、フラクタル画像ベースの事前学習及び事前学習なしの場合と比較して、一貫して優れた性能を示した。このことから、自然画像から学習した特徴抽出器が赤外線野生動物分類タスクに対して高い汎化性能を持つことを示している。メタ学習に関しては、提案したメタ学習アプローチが従来の学習手法であるミニバッチ学習と比較して分類精度を 11.1% 向上させ、未登録クラスに対する検出性能である AUROC を 7.7% 改善した。この結果は、データ数が限られた条件下において、モデルを新規クラスに適応させる際のメタ学習の有効性を実証している。未登録クラスの多クラス分類においては、k-means 損失と Between-Class 損失の導入により既存の PEELER モデルと比較して、登録クラスと未登録クラスを含む全体の分類精度 (All Accuracy) を改善した。これらの実験により、クラスタリングに基づく損失関数が未登録クラスの多クラス分類に対して有効であることが実証された。

また、異なる地域から収集された学習用データセットと評価用データセットを用いることにより、ドメインシフト下におけるモデルの適応性に関する知見を得た。さらに、学習に用いるサポートデータの Shot 数の増加がモデルの分類性能に及ぼす影響を検証し、Shot 数を増やすことで分類精度が大幅に向上することを明らかにした。これらの知見は、夜間環境という困難な条件下における野生動物モニタリング技術の進展に貢献するとともに、生態学の研究において直面する限られたデータ条件下でも適用可能な、効率的かつ汎用性の高い機械学習モデルの開発に向けた基盤となる成果を提供している。

野生動物モニタリングの実用化に向けた今後の課題として、本研究で取り組んだ画像分類に加えて、物体検出手法の開発が挙げられる。特に、少數の赤外線画像データという制約下において

も、画像内の動物領域を高精度に検出可能な手法の開発が急務である。また、将来の研究における有望な方向性として、IFOR フレームワークの性能を潜在的に向上させる多段階転移学習アプローチが挙げられる。具体的には、ImageNet のような大規模データセットにより事前学習されたモデルを基盤とし、Snapshot Serengeti のような大規模なカメラトラップ画像データセットを用いた中間的なファインチューニングを経て、最終段階で、ターゲットタスクである CCT や WCS のような小規模な赤外線画像データセットを用いたメタ学習を行うアプローチである。

# 謝辞

本研究を進めるにあたり、全過程を通じて御助言、御指導を頂きました、滝本 裕則 教授に深謝の意を表します。

また、本研究及び本論文の執筆において、様々な面で御協力を頂きました、金川 明弘 岡山県立大学 名誉教授に感謝申し上げます。

そして、本研究の遂行にあたり日頃から御討論、御協力を頂いた数理情報メディア工学研究室の皆様に厚く御礼申し上げます。特に、本研究の基盤となる研究を進められた伊藤 嵐丸 氏、岸本 昌子 氏に感謝の意を表します。

最後に、研究以外の面でも、日常的な交流を通じて精神的な支えとなってくれた友人たちにお礼申し上げます。日々の何気ない会話や共に過ごした時間は、研修生活を豊かなものにしてくれました。

# 研究業績

## 学術雑誌への掲載論文

1. 北山晃生, 岸孝樹, 古賀荘翠, 滝本裕則, 金川明弘, “室内での見守り実現に向けた全方位画像に対する複数人物追跡,” 日本福祉工学会誌, Vol. 26, No. 2, pp. 2-8 (7 ページ) , 2024 年 11 月.
2. Koki Kishi, Ranmaru Ito, Sulfayanti Faharuddin Situju, Hironori Takimoto, and Akihiro Kanagawa, “Animal Classification Considering Infrared Few-shot Open-set Recognition,” *Cogent Engineering*, (Accepted), 2025 年 1 月.

## 国際会議での研究発表

1. ◎ Koki Kishi, Masako Kishimoto, Sulfayanti Faharuddin 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 2023 (ICISIP2023)*, pp. 11-17 (7 ページ) , 2023 年 9 月.
2. ◎ Koki Kishi, Ranmaru Ito, Sulfayanti Faharuddin Situju, Hironori Takimoto and Akihiro Kanagawa, “Few-Shot Learning for Animal Classification in Camera Traps using an Infrared Camera,” In *Proceedings of the SICE Festival 2024 with Annual Conference (SICE FES 2024)*, WeBT7.7, pp. 420-423 (4 ページ) , 2024 年 8 月.

## 国内学会（全国レベル）での研究発表

1. ◎古賀荘翠, 北山晃生, 岸孝樹, 滝本裕則, 金川明弘, “複数物体追跡における ID スイッチ抑制のための Motion SORT の提案,” 第 28 回 パターン計測シンポジウム, PM108\_04 (4 ページ) , 2023 年 11 月.
2. ◎岸孝樹, 伊藤嵐丸, 植田諒大, Sulfayanti Faharuddin Situju, 滝本裕則, “Infrared Few-shot Open-set Recognition を考慮したクラスタリングとメタ学習による動物分類,” 第 29 回 パターン計測シンポジウム, PM109\_02 (8 ページ) , 2024 年 11 月.

## 国内学会（支部・県レベル）での研究発表

1. ◎岸孝樹, 伊藤嵐丸, 滝本裕則, 金川明弘, “赤外線カメラトラップにおける動物分類のための少数データ学習,” 第 26 回 IEEE 広島支部学生シンポジウム (IEEE HISS 26th) 論文集, TP-A-25, pp. 91-94 (4 ページ) , 2024 年 11 月.

## 受賞

1. The 10th IIAE International Conference on Intelligent Systems and Image Processing  
2023 Best Paper Award, 2023 年 9 月（対象は上記の国際会議での研究発表 (1)) .
2. 計測自動制御学会 計測部門 パターン計測部会 令和 6 年度優秀論文賞, 2024 年 11 月（対象  
は上記の国内学会（全国レベル）での研究発表 (2)）.
3. IEEE 広島支部学生シンポジウム (HISS) 優秀研究賞, 2024 年 11 月（対象は上記の国内学会  
(支部・県レベル) での研究発表 (1)）.

# 参考文献

- [1] Anantha Kumar Duraiappah, Shahid Naeem, Tundi Agardy, Neville J. Ash, H. David Cooper, Sandra Diaz, Daniel P. Faith, Georgina Mace, Jeffrey A. McNeely, Harold A. Mooney, Alfred A. Oteng-Yeboah, Henrique Miguel Pereira, Stephen Polasky, Christian Prip, Walter V. Reid, Cristian Samper, Peter Johan Schei, Robert Scholes, Frederik Schutyser, and Albert van Jaarsveld, *Ecosystems and Human Well-Being: Biodiversity Synthesis*. Millennium Ecosystem Assessment Series, Washington, DC: World Resources Institute (WRI), 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～,” 環境省 自然環境局, 東京, 2023. [https://www.biodic.go.jp/biodiversity/about/initiatives6/files/1\\_2023-2030text.pdf](https://www.biodic.go.jp/biodiversity/about/initiatives6/files/1_2023-2030text.pdf), 生物多様性センター (2024 年 12 月 7 日閲覧).
- [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, Gorry 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-channeled perceiving residual pyramid networks,” In *Proceedings of the IEEE International Conference on Computer Vision Work-*

- shops (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 Klamm, 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] Asmita Manna, Nilam Upasani, Shubham Jadhav, Ruturaj Mane, Rituja Chaudhari, and Vishal Chatre, “Bird image classification using convolutional neural network transfer learning architectures,” *International Journal of Advanced Computer Science and Applications*, Vol. 14, No. 3, 2023.
  - [13] Anishka Mohanty, Guillermo Goldsztein, and Raphaël Pellegrin, “Fish species image classification using convolutional neural networks,” *Journal of Student Research*, Vol. 11, 2022.
  - [14] Nikita Agarwal, Tina Kalita, and Ashwani Kumar Dubey, “Classification of insect pest species using cnn based models,” In *Proceedings of the International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, pp. 862–866, 2023.
  - [15] Subash Neeli, Chandra Sekhar Reddy Guruguri, Adithya Ram Achari Kammara, Visalakshi Annepu, Kalapraveen Bagadi, and Venkata Rami Reddy Chirra, “Bird species detection using cnn and efficientnet-b0,” In *Proceedings of the International Conference on Next Generation Electronics (NEleX)*, pp. 1–6, 2023.
  - [16] 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.
  - [17] Koki Kishi, Masako Kishimoto, Sulfayanti Faharuddin 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.
  - [18] Walter Scheirer, Anderson Rocha, Archana Sapkota, and Terrance Boult, “Toward open set recognition,” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, pp. 1757–72, 2013.
  - [19] 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.

- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] Joeri Zwerts, PJ Stephenson, Fiona Maisels, Marcus Rowcliffe, Christos Astaras, Patrick Jansen, Jaap van der Waarde, Liesbeth Sterck, P.A. Verweij, Tom Bruce, Stephanie Britain, 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.
- [28] 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.
- [29] Franck Trolliet, Marie-Claude Huynen, Cédric Vermeulen, and Alain Hambuckers, “Use of camera traps for wildlife studies. a review,” *Biology Agriculture Science Environment*, Vol. 18, pp. 446–454, 2014.
- [30] 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.
- [31] Baydaa Sh. Z. Aboot, Manjula B. M, Zainab abed Almoussawi, N Shilpa, and Aboothar mahmood 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.

- [32] 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.
- [33] 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.
- [34] 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.
- [35] 安藤 正規, 中塚 俊介, 相澤 宏旭, 中森 さつき, 池田 敬, 森部 純嗣, 寺田 和憲, 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.
- [36] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” In *Proceedings of the Advances in Neural Information Processing Systems* (F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, eds.), Vol. 25, Curran Associates, Inc., 2012.
- [37] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science*, Vol. 7, 2016.
- [38] Sue Han Lee, Hervé Goëau, Pierre Bonnet, and Alexis Joly, “New perspectives on plant disease characterization based on deep learning,” *Computers and Electronics in Agriculture*, Vol. 170, p. 105220, 2020.
- [39] Yiming Fang, Xianxin Guo, Kun Chen, Zhu Zhou, and Qing Ye, “Accurate and automated detection of surface knots on sawn timbers using yolo-v5 model,” *BioResources*, Vol. 16, pp. 5390–5406, 2021.
- [40] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He, “Fcos: Fully convolutional one-stage object detection,” In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [41] Zhaowei Cai and Nuno Vasconcelos, “Cascade r-cnn: Delving into high quality object detection,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6154–6162, 2018.
- [42] 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.
- [43] Chuanxing Geng, Sheng-Jun Huang, and Songcan Chen, “Recent advances in open set recognition: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 10, pp. 3614–3631, 2021.
- [44] Jiayin Sun and Qiulei Dong, “A survey on open-set image recognition,” *arXiv preprint arXiv:2312.15571*, 2023.
- [45] Atefeh Mahdavi and Marco Carvalho, “A survey on open set recognition,” In *Proceedings of the IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pp. 37–44, 2021.
- [46] Haoyu Wang, Guansong Pang, Peng Wang, Lei Zhang, Wei Wei, and Yanning Zhang, “Glocal energy-based learning for few-shot open-set recognition,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7507–7516, 2023.
- [47] Yongjuan Che, Yuexuan An, and Hui Xue, “Boosting few-shot open-set recognition with multi-relation margin loss,” In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23* (Edith Elkind, ed.), pp. 3505–3513, International Joint Conferences on Artificial Intelligence Organization, 2023. Main Track.
- [48] Shiyuan Huang, Jiawei Ma, Guangxing Han, and Shih-Fu Chang, “Task-adaptive negative envision for few-shot open-set recognition,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7171–7180, 2022.
- [49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [50] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever, “Learning transferable visual models from natural language supervision,” *arXiv preprint arXiv:2103.00020*, 2021.
- [51] Jake Snell, Kevin Swersky, and Richard S. Zemel, “Prototypical networks for few-shot learning,” *arXiv preprint arXiv:1703.05175*, 2017.
- [52] Wildlife Conservation Society, “WCS Camera Traps,” <https://lila.science/datasets/wcscameratraps>, Labeled Information Library of Alexandria: Biology and Conservation (LILA BC) (2025年1月8日閲覧).
- [53] Sara Beery, Grant Van Horn, and Pietro Perona, “Recognition in terra incognita,” In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [54] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Herve Jegou, “Training data-efficient image transformers & distillation

- through attention,” In *Proceedings of the 38th International Conference on Machine Learning (ICML)* (Marina Meila and Tong Zhang, eds.), Vol. 139, pp. 10347–10357, Proceedings of Machine Learning Research (PMLR), 2021.
- [55] Ross Wightman, “Pytorch image models,” 2019. <https://github.com/rwightman/pytorch-image-models>, GitHub repository (2025年1月22日閲覧).