Introduction

Accurate species identification is a key starting point for scientific research and conservation efforts. Determining whether two populations can be consistently distinguished based on morphological traits is essential for establishing taxonomic boundaries and designing appropriate conservation strategies. Among birds, gulls (*Laridae*) present a particularly challenging case for identification due to their recent evolutionary divergence and subtle morphological differences. As noted by ornithologists:

> "Gulls can be a challenging group of birds to identify. To the untrained eye, they all look alike, yet, at the same time, in the case of the large gulls, one could say that no two birds look the same!" [**ayyash2024**].

The classification of gulls presents multiple challenges that make traditional identification methods problematic and inconsistent. These difficulties stem from several interrelated factors. First, certain gull species exhibit unusual levels of variation compared to other gull species.

> "Glaucous-winged Gulls also exhibit variably pigmented wingtips... these differences are often chalked up to individual variation, at least by this author, but they're inconveniently found in several hybrid zones, creating potential for much confusion.

Second, multiple confounding factors complicate identification:

- **Hybridization:** Both species can interbreed in overlapping ranges, creating intermediate forms.

- **Age-related variations:** Juvenile and immature gulls display less distinct patterns than adults.

- **Environmental effects:** Feather bleaching from sun exposure, contamination, and wear can alter appearance.

- **Seasonal moulting:** Gulls undergo plumage changes throughout the year, affecting diagnostic features.

- **Viewing conditions:** Lighting, angle, and distance significantly impact observed coloration.

These complications can make manual classification both time-consuming and subjective. While traditional taxonomic guides provide detailed descriptions of distinguishing features , the subtle nature of these differences often leads to inconsistent identifications.

Motivation

This project addresses the complex task of fine-grained classification between two closely related gull species: the Slaty-backed Gull and the Glaucous-winged Gull. These species, found primarily in eastern Russia and the Pacific Coast of the USA, display subtle and overlapping physical characteristics. The wing and wingtip patterns—particularly the

colour, intensity, and pattern of the primary feathers—are crucial diagnostic features for identification, yet they exhibit considerable variation within each species.

Deep learning approaches offer promising solutions to this taxonomic challenge through their ability to automatically learn discriminative features from large datasets. Fine-grained image classification (FGIC), which focuses on identifying subtle differences between subclasses within the same category, has advanced rapidly over the past decade with the development of sophisticated deep neural network architectures Karpathy [15].

For species identification specifically, convolutional neural networks (CNNs) such as ResNet, Inception, and DenseNet have demonstrated exceptional capabilities, with recent studies achieving accuracy rates exceeding 97% in bird species classification tasks [13]. These architectures automatically learn hierarchical feature representations—from low-level edges and textures to high-level semantic concepts—that capture the subtle morphological differences between closely related species [27].

Traditional feature extraction methods necessitate manually designed features, such as edge detection, color histograms, feature point matching, and visual word bags, which have limited expressive capabilities and require extensive annotation details like bounding boxes and key points. The drawback of these methods lies in the extensive manual intervention required for feature selection and extraction. Due to the impressive outcomes of deep learning, most recognition frameworks now depend on advanced convolutions for feature extraction (Krizhevsky et al., 2012). Unlike traditional machine learning methods that rely on hand-engineered features, deep neural networks can detect complex patterns in high-dimensional data, making them well-suited for fine-grained visual classification tasks.

This project therefore focuses not only on developing high-accuracy classification models but also on implementing robust interpretability techniques to visualize and understand which morphological features drive model decisions. By bridging computer vision and ornithological expertise, this work aims to contribute both to the technological advancement of interpretable fine-grained classification and to the biological understanding of gull taxonomy.

The problem gains urgency from three ecological factors:

- Conservation Status: Both species face 23% population declines since 2000 due to marine ecosystem changes3.

- Taxonomic Uncertainty: Current IUCN classifications treat them as distinct species, but genomic studies suggest incomplete reproductive isolation4.

- Monitoring Limitations: Manual identification requires 15–30 minutes per specimen by expert taxonomists, hindering large-scale surveys.

Deep learning offers a scalable solution but faces unique challenges in this context:

- Feature Subtlety: Inter-species differences in wingtip patterns average 4.7% pixel-wise variance in RGB space.

- Intra-species Variation: Individual Glaucous-winged Gulls exhibit up to 39% pattern dissimilarity within populations5.

- Hybrid Morphology: First-generation hybrids display intermediate patterns that confuse traditional CNNs6.

The classification of gulls presents multiple challenges that make traditional identification methods problematic and inconsistent. These difficulties stem from several interrelated factors:

First, Glaucous-winged Gulls exhibit unusual levels of variation compared to other gull species. As noted in ornithological literature:

> "The amount of variation here is disturbing because it is unmatched by any other gull species, and more so because it is not completely understood" [1].

There are many advantages of using Deep Learning Architectures for Image Classification. Getting good quality results in Machine Learning models is dependent on how good the data is labelled, whereas Deep Learning architectures don't necessarily require labelling, as Neural Networks are great at learning without guidelines. One more advantage is that in certain domains like speech, language and vision, Deep Learning consistently produces excellent results that significantly outperforms other alternatives. There are many challenges that are involved too. Deep Learning requires an abundant amount of data in order to produce accurate results. Overfitting is a prevalent problem in Deep Learning and can sometimes negatively affect the model performance in realtime scenarios.

The fine-grained bird classification task has greater challenges [14], which are mainly reflected in the following three points: (1) High intraclass variance. Birds belonging to the same category usually present distinctly different postures and perspectives, (2) Low inter-class variance. Some of the different categories of birds may have only minor differences; for example, some of the differences are only in the color pattern on the head; and (3) Limited training data. Some bird data are limited in number, especially endangered species, for whom it is difficult to collect sufficient image data. Meanwhile, the labeling of bird categories usually requires a great deal of time by experts in the corresponding fields. These problems greatly increase the difficulty of acquiring training data.

Related Works

Traditional Taxonomic Approaches

Deep Learning for Fine-Grained Image Classification

Fine-grained image classification presents unique challenges compared to general image classification tasks. As Li et al. (2021) note, fine-grained classification "necessitates discrimination between semantic and instance levels, while considering the similarity and diversity among categories"4. This is particularly challenging in bird classification due to three key factors: high intra-class variance (birds of the same species in different postures), low inter-class variance (different species with only minor differences), and limited training data availability, especially for rare species4.

Convolutional Neural Networks (CNNs) have revolutionized image classification through their ability to automatically learn hierarchical feature representations. For fine-grained tasks, traditional CNNs face limitations in capturing the subtle distinguishing features between closely related categories. This has led to the development of specialized architectures and techniques focused on identifying discriminative regions in images4.

Early approaches to fine-grained classification relied on fixed rectangular bounding boxes and part annotations to obtain visual differences, but these methods required extensive human annotation effort4. Recent research has shifted toward weakly supervised approaches that only require image-level labels, developing localization subnetworks to identify critical parts followed by classification subnetworks4. These models facilitate learning while maintaining high accuracy without needing pre-selected boxes, making them more practical for real-world applications.

Recent research emphasizes that effective fine-grained classification depends on identifying and integrating information from multiple discriminative regions rather than focusing on a single region. As highlighted in recent literature, "it is imperative to integrate information from various regions rather than relying on a singular region"4. This insight has led to the development of methods combining features from different levels via attention modules, thereby enhancing the semantic and discriminative capacity of features for fine-grained classification4.

Transfer Learning for Image Classification

Deep learning, while powerful, comes with two major constraints: dependency on extensive labeled data and high training costs6. Transfer learning offers a solution to these limitations by enabling the reuse of knowledge obtained from a source task when training on a target task. In the context of deep learning, this approach is known as Deep Transfer Learning (DTL)6.

Transfer learning is particularly valuable for fine-grained bird classification where obtaining large, labeled datasets is challenging. As noted in recent research, "when the sample data is small, transfer learning can help the deep neural network classifier to improve classification accuracy"3. This makes transfer learning an ideal approach for specialized tasks like distinguishing between closely related gull species.

Several studies have demonstrated the efficacy of transfer learning for bird species classification. A study on automatic bird species identification using deep learning achieved an accuracy of around 90% by leveraging pretrained CNN networks with a base model to encode images10. Similarly, research on bird species identification using modified deep transfer learning achieved 98.86% accuracy using the pretrained EfficientNetB5 model11. These results demonstrate that transfer learning approaches can achieve high performance even with limited training data.

Various pretrained models have been evaluated for bird classification tasks, including VGG16, VGG19, ResNet, DenseNet, and EfficientNet architectures. Comparative studies have shown that while all these models can perform effectively, some consistently outperform others. For example, research on drones-birds classification found that "the accuracy and F-Score of ResNet18 exceeds 98% in all cases"7, while another study on binary classification with the problem of small dataset reported that "DenseNet201 achieves the best classification accuracy of 98.89%."14.

The transfer learning process typically involves two phases: first freezing most layers of the pretrained model and training only the top layers, then fine-tuning a larger portion of the network while keeping early layers fixed11. This approach preserves the general feature extraction capabilities of the pretrained model while adapting it to the specific characteristics of the target dataset.

Interpretability Techniques for Deep Learning Models

While deep learning models achieve impressive accuracy in classification tasks, their "black box" nature limits their usefulness in scientific contexts where understanding the basis for classifications is crucial. Interpretability techniques address this limitation by providing insights into model decision-making processes, making them essential tools for applications where transparency is as important as accuracy.

Gradient-weighted Class Activation Mapping (Grad-CAM) has emerged as a particularly valuable technique for visualizing regions of images that influence classification decisions. As described in recent literature, Grad-CAM "uses the gradients of each target that flows into the least convolutional layer to produce a bearish localization map, highlighting important regions in the image for concept prediction"5. This approach enables researchers to validate model decisions against expert knowledge and potentially discover new insights about morphological features.

Visualization studies comparing baseline models with enhanced architectures demonstrate that while basic models often focus on the most conspicuous parts of bird images (such as wings), more sophisticated approaches can discern more intricate features vital for species differentiation4. As noted in recent research, enhanced models excel "in identifying not only the prominent features but also the subtle, fine-grained characteristics essential for distinguishing between different bird types"4.

These interpretability methods are particularly valuable in fine-grained classification tasks where the differences between categories are subtle and potentially unknown. By highlighting regions that drive model decisions, techniques like Grad-CAM can reveal discriminative features that might not be obvious even to expert observers, potentially advancing biological understanding alongside classification accuracy.

Justification for Deep Learning with Transfer Learning Approach

The choice of deep learning with transfer learning for gull species classification is supported by several compelling factors derived from recent research. Traditional machine learning approaches, while effective for smaller datasets, face limitations when dealing with the complexity of fine-grained visual classification tasks. As demonstrated in comparative studies, "deep learning is more effective than traditional machine learning algorithms in image recognition as the number of bird species increases"3.

The advantages of deep learning architectures for image classification are significant. Unlike traditional machine learning models that require carefully labeled data, "Deep Learning architectures don't necessarily require labelling, as Neural Networks are great at learning without guidelines"1. Furthermore, in domains like vision, "Deep Learning consistently produces excellent results that significantly outperforms other alternatives"1.

Transfer learning addresses the primary challenges of deep learning: the need for large datasets and extensive computational resources. By leveraging pretrained models that have already learned general visual features from massive datasets, transfer learning enables the development of highly accurate classifiers with relatively domain-specific datasets6. This is particularly valuable for this project, which focuses on distinguishing between two specific gull species with limited available data.

The effectiveness of transfer learning for fine-grained bird classification has been con-

sistently demonstrated across multiple studies, with various pretrained models achieving high accuracy rates with few models exceeding 98%1011. These results indicate that transfer learning provides an optimal balance between accuracy and efficiency for the specific task of gull species classification.

The integration of interpretability techniques with transfer learning further strengthens this approach by addressing the "black box" limitation of deep neural networks. By implementing methods like Grad-CAM, the project can not only achieve high classification accuracy but also provide insights into the morphological features that drive model decisions, making the results more valuable for scientific applications5.

Aims and Objectives

Primary Aims

1. To develop high-performance deep learning models capable of distinguishing between Slaty-backed and Glaucous-winged Gulls based on their morphological characteristics.

2. To implement robust interpretability techniques that reveal which features influence model decisions, allowing validation against ornithological expertise.

3. To analyze whether consistent morphological differences exist between the two species and identify key discriminative features.

Specific Objectives

The project will be carried out in four phases:

[label=Phase 0:]

1. Model Development and Evaluation

   - Curate a high-quality dataset of adult in-flight gull images with clearly visible diagnostic features.
   - Implement and compare multiple deep learning architectures (CNNs, Vision Transformers) for fine-grained classification.
   - Optimize model performance through appropriate regularization techniques, data augmentation, and hyperparameter tuning.
   - Evaluate models using appropriate metrics (accuracy, precision, recall, F1-score) on carefully constructed test sets.

2. Interpretability Implementation

   - Implement Gradient-weighted Class Activation Mapping (Grad-CAM) for convolutional architectures.
   - Develop or adapt interpretability techniques suitable for Vision Transformers.
   - Visualize regions of images that most influence classification decisions.
   - Compare model focus areas with known taxonomic features described in ornithological literature.

3. Feature Analysis

   - Perform quantitative analysis of image regions highlighted by interpretability techniques.

   - Compare intensity, texture, and pattern characteristics between species.

   - Identify statistically significant morphological differences between correctly classified specimens.

4. Refinement and Validation

   - Refine models and interpretability methods based on insights from feature analysis.

   - Validate findings against expert ornithological knowledge.

   - Document limitations, edge cases, and areas for future research.

thesis