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!" [?].

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 [?]. Current approaches to FGIC can be broadly categorized into two frameworks: intensely supervised learning, which requires detailed annotations of discriminative parts, and weakly supervised learning, which achieves comparable or superior results using only image-level labels [?].

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 [?]. 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 [?]. Modern approaches further enhance classification performance through attention mechanisms that help models focus on discriminative regions without explicit part annotations and multi-network learning strategies that capture complementary visual information at different scales [?].

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. However, these models often function as "black boxes," providing little insight into their decision-making processes—a critical limitation in scientific applications where understanding the reasoning behind classifications is as important as the classifications themselves.

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.

Motivation

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" [?].

This variability extends to wingtip patterns, which are critical for species identification but inconsistent within populations.

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.

Deep learning offers a powerful alternative by automatically learning distinctive features from large datasets. However, the "black box" nature of most neural networks limits their usefulness in scientific contexts where understanding the basis for classifications is crucial. This project therefore emphasizes interpretability alongside accuracy, ensuring that model decisions can be validated against expert knowledge and potentially yield new insights into the morphological basis of species differentiation.

By focusing specifically on adult in-flight images where wingtip patterns are most visible, this project addresses the core taxonomic question while minimizing confounding variables. The resulting interpretable classification system aims to provide both a practical identification tool and a scientific instrument for exploring morphological variation within and between these closely related species.

Chapter 1

Background and Related Work

Explaining what your project does that is new or is better than existing work in the same field.

Test citation [?].

Chapter 2

Design

2.1 Methodology: Model Development & Interpretability Implementation

2.1.1 Architectural Design Philosophy

The model development process followed three core principles derived from recent advances in fine-grained classification research:

- **Hierarchical Feature Learning:** Prioritize architectures capable of capturing multi-scale features (local textures and global patterns) [?, ?].
- **Attention-Aware Processing:** Incorporate mechanisms to focus on discriminative regions without manual part annotations [?, ?].
- **Interpretability by Design:** Ensure architectural compatibility with gradient-based visualization techniques [?, ?].

This framework guided the selection and modification of baseline architectures to address the specific challenges of gull wingtip analysis.

2.1.2 Model Implementations

CNN Architectures

ResNet-50 Adaptation

- **Rationale:** Residual connections enable stable training on limited data by preserving gradient flow [?, ?].
- **Implementation:**
  1. Load ImageNet-pre-trained ResNet-50.
  2. Replace final FC layer with:
     - Dropout(0.5)
     - Linear(in_features=2048, out_features=2)
  3. Freeze initial blocks, progressively unfreeze during fine-tuning.

Inception-v3 Enhancement

- **Rationale:** Multi-scale feature extraction crucial for analyzing variable wingtip sizes [?, ?].
- **Implementation:** Add batch normalization before final FC layer and modify auxiliary classifiers.

VGG16 Optimization Shallow architecture enables precise gradient flow for interpretability. Added spatial dropout (p=0.2) before final pooling and implemented gradient clipping (max_norm=2.0).

...

2.1.3 Vision Transformer Adaptations

Base ViT Implementation

- **Rationale:** Global attention captures wingtip-body relationships [?, ?].
- **Architectural Details:** Replace the head with:
  1. LayerNorm
  2. Dropout(0.3)
  3. Linear($768 \rightarrow 2$)

...

2.1.4 Data Augmentation Strategy

The augmentation pipeline included:

- Geometric Transformations: Random rotation ($\pm 10°$), affine translation (5%), and random resized crop (95–100% scale).
- Photometric Adjustments: Color jitter (brightness/contrast/saturation = 0.1) and sharpness adjustment (factor = 1.2).
- Normalization: Using ImageNet statistics.

Research showed that controlled geometric augmentations improved bird classification accuracy by 18% and subtle color variations reduced overfitting to lighting conditions by 42% ([?, ?]).

...

Chapter 3

Implementation

Chapter 4

Evaluation

Explaining how your software was tested (using different datasets or in different environments), statistical evaluation of performance, results of user evaluation questionnaires, etc.

Chapter 5

Summary and Reflections

Including a discussion of results in a wider context (considering other work).

5.1 Project management

Covering the tasks as a part of your work plan and progress as well as how time and resources are managed.

5.2 Contributions and reflections

Providing the details of your achievements and contributions including innovation, creativity and novelty (if there is any) as well as a personal reflection on the plan and your experience of the project (a critical appraisal of how the project went).

Chapter 6

User Manuals

Chapter 7

User Evaluation Questionnaire