Bird Species Detection using Deep Learning on the CUB-200-2011 Dataset

Megh Patel (Enrollment No.: AU2544020)

Konark Karia (Enrollment No.: AU2544011)

September 15, 2025

Project Proposal for AI-Lab Project

---

1. Project Title

Bird Species Detection using Deep Learning

2. Project Overview / Introduction

Fine-grained image recognition aims to distinguish categories with subtle visual differences, such as bird species. The Caltech–UCSD Birds 200–2011 (CUB-200-2011) dataset provides 11,788 images of 200 species with annotations for parts, bounding boxes and attributes. Traditional CNN classifiers fail to focus on discriminative regions (e.g., beak color, wing tips), making fine-grained recognition challenging.

Recent research has proposed attention-based models:

- Recurrent Attention CNN (RA-CNN) learns to zoom progressively into discriminative regions without part labels.
- Multi-Attention CNN (MA-CNN) discovers multiple part attentions via channel grouping and jointly optimizes localization and classification.
- Part-based R-CNN leverages region proposals and geometric constraints to detect semantic parts for pose-normalized features.

This project will build a bird-species detection system by exploring these methods, comparing their performance, and deploying the best-performing model via a lightweight web interface.

3. Objectives

- Study fine-grained recognition techniques for bird classification.
- Preprocess the CUB-200-2011 dataset (resizing, normalization, augmentation).
- Implement and train RA-CNN, MA-CNN, and Part-based R-CNN on the dataset.
- Compare models on accuracy, F1-score, and part-localization quality.
- Deploy the best model as a web app allowing image upload and species prediction.

4. Literature Review / Background

Dataset: CUB-200-2011 introduces rich annotations and is a benchmark for subordinate categorization.

RA-CNN: Fu et al. propose a recurrent attention network that iteratively crops and enlarges discriminative regions, trained with intra-scale classification and inter-scale ranking losses, achieving state-of-the-art accuracy without part annotations.

MA-CNN: Zheng et al. design a multi-attention mechanism that groups convolutional channels into parts, learning both localization and fine-grained features with a channel-grouping loss and classification loss.

Part-based R-CNN: Zhang et al. extend R-CNN to localize object parts via region proposals and geometric priors, enabling pose-normalized fine-grained classification.

5. Tools and Technologies

- Python 3, Jupyter Notebook / Google Colab
- Frameworks: PyTorch (preferred) or TensorFlow/Keras
- Libraries: OpenCV, NumPy, Matplotlib, scikit-learn
- Deployment: Flask or Streamlit for web inference

6. Dataset

Dataset: Caltech–UCSD Birds 200–2011

- 200 bird species, 11,788 images
- Annotations: 15 part locations, 312 attributes, bounding boxes

Preprocessing:

- Resize images to 224×224 or model-specific input size
- Normalize pixel values; augment with flips, crops, color jitter
- Split: 70% train, 15% validation, 15% test

7. Timeline

Week	Task	Deliverable
1	Literature survey of RA-CNN, MA-CNN, Part-RCNN	Summary report
2	Dataset exploration, preprocessing	Cleaned dataset
3	Implement baseline CNN	Baseline accuracy
4	Implement RA-CNN	RA-CNN trained model
5	Implement MA-CNN and Part-RCNN	Trained models
6	Comparative evaluation, hyperparameter tuning	Metrics table
7	Web interface and final documentation	Working demo + report

8. References

1. Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2011). *The Caltech–UCSD Birds-200-2011 Dataset.*
2. Fu, J., Zheng, H., Mei, T. (2017). *Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition.* CVPR.
3. Zheng, H., Fu, J., Mei, T., Luo, J. (2017). *Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition.* ICCV.

4. Zhang, N., Donahue, J., Girshick, R., Darrell, T. (2014). *Part-based R-CNNs for Fine-Grained Category Detection*. ECCV.