COS30082 – Apply Machine Learning

Assignment -1

Student Name: Nguyen Dinh Dung

Student ID: 104772138

SWINBURNE UNIVERSITY OF TECHNOLOGY

Contents

Nguyen Dinh Dung
104772138

Abstract

This paper presents a comparative study on multi-class bird species classification using the Caltech-UCSD Birds 200 (CUB-200) dataset. Three models — ResNet-50, EfficientNet-B3, and Vision Transformer (ViT-8) — are evaluated on this fine-grained classification task. Performance is assessed using Top-1 accuracy and Average Accuracy per Class, highlighting each model's strengths and weaknesses. The results offer insights into the trade-offs between convolutional networks and transformer-based architectures for species classification with limited training data.

Resnet 50

Selecting and Configuring the Base Model:

- **ResNet-50 Backbone:** ResNet-50 was chosen for its deep residual learning capabilities and strong feature extraction, pre-trained on ImageNet. Its rich feature hierarchy helps handle the fine-grained classification challenges of the CUB-200 dataset.
- **Transfer Learning:** By leveraging pre-trained features, transfer learning reduces data requirements and accelerates training, improving accuracy — particularly helpful given the dataset's quality issues.

Fine-Tuning Strategy:

- **Selective Layer Training:** Layers before the fifth are frozen to retain general features, while later layers are fine-tuned to capture bird-specific traits.
- **Full Adaptability:** Optionally, setting *fine_tune_start* to a negative value makes all layers trainable for deeper adaptation to the dataset.

Output Layer Reconfiguration:
The original ResNet-50 output layer is replaced with a custom sequence for classifying 200 bird species. It includes:

- Linear layer reducing features to 512
- ReLU activation for non-linearity
- Dropout (rate = 0.5) to reduce overfitting
- Final linear layer mapping to 200 classes

Regularization and Overfitting Mitigation:

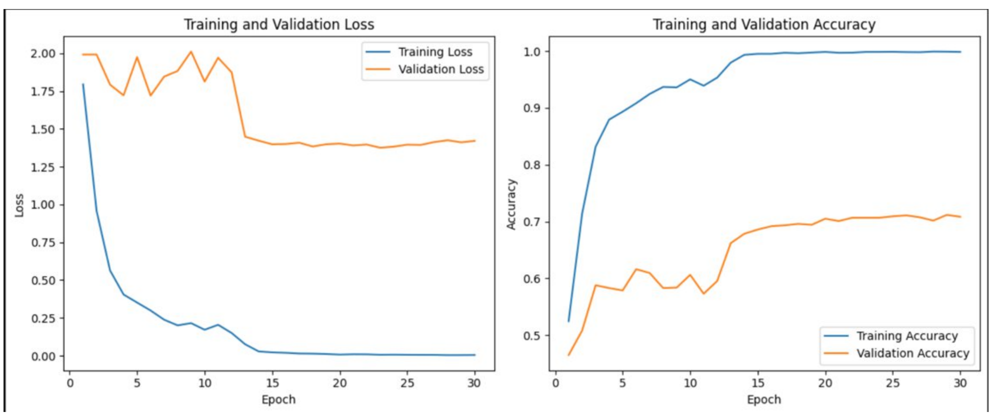
- **Dropout:** Applied (rate = 0.5) to promote robustness by preventing reliance on specific neurons.
- **Layer Freezing:** Freezing early layers preserves general features and reduces overfitting.

Training and Validation Process:

- **Loss & Optimizer:** Cross Entropy Loss with Adam optimizer (initial LR = 0.001).
- **Early Stopping:** Stops training if validation loss stagnates.
- **LR Scheduler:** ReduceLROnPlateau lowers learning rate by 0.1 when validation loss plateaus.

Data Handling and Augmentation:
Training images are resized, center-cropped, converted to tensors, and normalized to standardize inputs and improve generalization.

Results and Discussion:



- **Training Performance:** Training loss dropped from 1.794 to 0.0047, with accuracy rising from 52.47% to 99.86%, showing effective learning.
- **Validation Performance:** Validation loss reached its lowest at epoch 14 (1.4218), while accuracy improved from 46.51% to 71.18% by epoch 29, demonstrating good generalization.

Analysis:

- **Early Training:** Initial low accuracy and high loss were expected as the model began learning.
- **Mid Training:** Validation performance stabilized and improved as parameters adjusted and learning rate decreased.

- **Late Training:** Training accuracy neared perfection, raising overfitting concerns, but validation accuracy continued improving, showing reasonable generalization.

Efficient Net B3

Model Overview

- **EfficientNet-B3**: A high-capacity CNN with strong performance on image classification tasks.
- **Improved Autoencoder (U-Net Style)**: Used for denoising and preprocessing the images before feeding them into the classifier.

The design balances **strong feature extraction (EfficientNet)** with **noise reduction (Autoencoder)**, aiming for **robust generalization**.

Autoencoder Denoising Phase

Design Choices:

- **U-Net style architecture** with **skip connections** to preserve spatial details.
- **Batch normalization layers** to stabilize training.
- **ConvTranspose layers** for upsampling (decoder) to recover image resolution.
- **Pretraining phase** for denoising using **MSE loss**, focusing on reconstructing noise-free images.

Data Handling and Augmentation

Data Pipeline:

- Custom CUB200Dataset class properly loads **image paths and labels** from annotation files.
- Augmentations during training include:
  - **Horizontal flips** and **random rotations** (mild geometric changes)
  - **Color jitter** (color variability)
  - **Random resized crops** (spatial variability)

Observations:

- **384×384 resolution** is appropriate for EfficientNet-B3, ensuring images are not overly shrunk. (EfficientNet-B3's optimal size)
- **Normalization using ImageNet means/std** ensures compatibility with the EfficientNet pretrained weights.
- Including **random resized crops** is particularly useful for small datasets like CUB-200.

EfficientNet-B3 Classifier

Architecture Modification:

- **Classifier head replaced** with a new nn.Linear layer to match 200 bird species.
- Pretrained weights (weights='DEFAULT') speed up convergence.

Gradual Unfreezing Strategy:

- **4 Training Stages**:
  - Stage 1: Train only classifier (all backbone frozen).
  - Stage 2: Unfreeze last feature block (high-level features).
  - Stage 3: Unfreeze another block.
  - Stage 4: Unfreeze a third block.

```
[ ]  # Training stages for gradual unfreezing
     training_stages = [
         {'epochs': 10, 'unfreeze': None, 'lr': 0.001},            # Train only classifier
         {'epochs': 10, 'unfreeze': 'features.7', 'lr': 0.005},   # Unfreeze last feature block
         {'epochs': 10, 'unfreeze': 'features.6', 'lr': 0.0001},   # Unfreeze one more block
         {'epochs': max_epochs, 'unfreeze': 'features.5', 'lr': 0.0005}  # Unfreeze one more block
     ]
```

- Each stage **has its own learning rate and schedule (decreasing as more layers are unfrozen) (Cosine Annealing)**.

Training Strategy

Key Elements:

- **Denoising pipeline** applied at both training and testing stages ensures consistency.

- **Early stopping with patience=15** ensures training halts if no improvement occurs.
- **Separate pretraining for autoencoder** prevents unnecessary coupling between denoising and classification training.

Evaluation Metrics

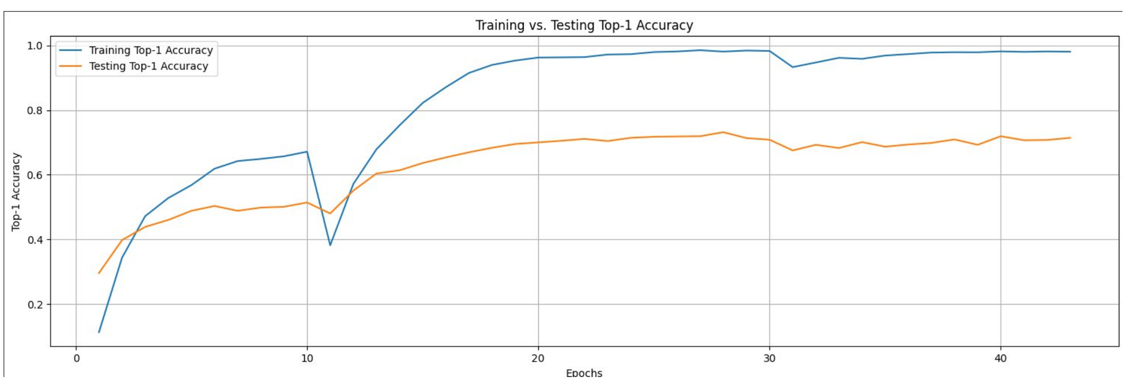
Top-1 Accuracy

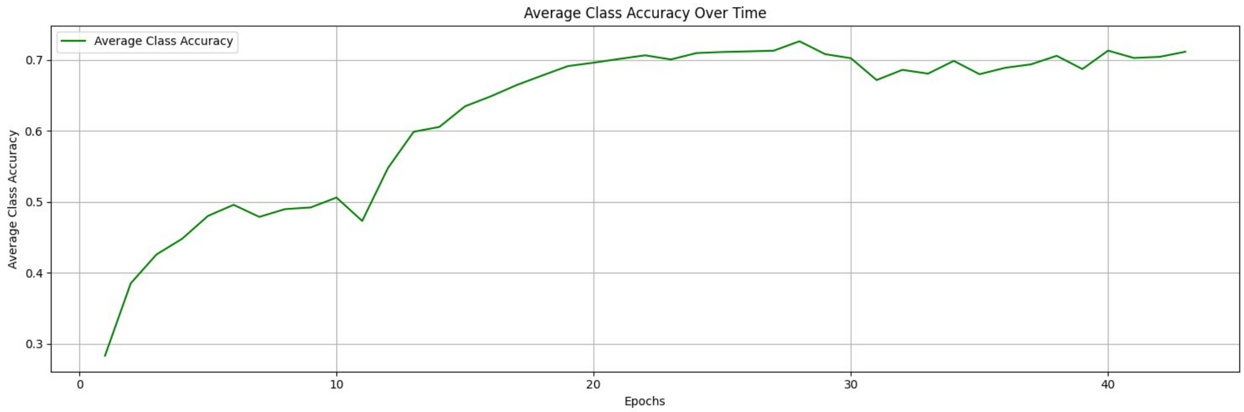
- Tracks overall correctness across all predictions.
- **Final Top-1 Accuracy: 73.17%** – respectable for a 200-class fine-grained classification task.

Average Class Accuracy

- Tracks per-class accuracy, avoiding the dominance of common species.
- **Final Average Class Accuracy: 72.60%** – this is very close to Top-1, which suggests relatively balanced performance across species.

Visualization and Monitoring





ViT: vit_base_patch8_224

Model Selection & Configuration

- **ViT-Base-Patch8-224** was chosen for its ability to capture both local and global patterns using 8x8 patches — ideal for fine-grained bird species classification.
- **Transfer Learning** with **pre-trained ImageNet-21k weights** helped accelerate training and improve generalization.

Fine-Tuning & Regularization

- **Full Model Fine-Tuning:** All layers were trainable to adapt to the CUB-200 dataset.
- **Output Layer:** Replaced with a linear classifier for 200 species.
- **Augmentations:** Resize, random horizontal flip, color jitter, random rotation, and normalization.
- **Regularization:**
    - **Label Smoothing (0.1)** improved robustness.
    - **Dropout within ViT blocks** reduced overfitting.

Training Process

- **Optimizer & Loss:** AdamW with **CrossEntropyLoss**.
- **LR Scheduling:** Cosine Annealing with **3-epoch warmup**.

- **Early Stopping:** Monitored **Top-1 accuracy**, stopping after **7 epochs without improvement**.

Results

- **Training Loss:** Reduced from **3.41** to **0.87**.
- **Train Accuracy:** Reached **99.94%**.
- **Top-1 Test Accuracy:** Peaked at **89.87%**.
- **Average Per-Class Accuracy:** Reached **89.60%**.

Analysis

- **Strong Early Gains:** Augmentations, warmup, and transfer learning helped the model adapt quickly.
- **Overfitting Mitigation:** Early stopping and label smoothing controlled overfitting.
- **Generalization:** The model performed well across species, though challenging fine-grained differences still limited accuracy.