# COL331 Principles of Artificial Intelligence

## Assignment 3 - Part 1 - Report

Sanyam Garg (2022CS11078) & Aneeket Yadav (2022CS11116)

**Introduction**

This report illustrates the architecture, results and various optimisations involved in a designing a CNN image classification model for 10-class supervised classification of bird images. We achieved validation accuracy of 94.93% subject to the following conditions: `torch.manual_seed(0),torch.cuda.manual_seed(0),torch.cuda.manual_seed_all(0), torch.backends.cudnn.deterministic = True.` Train and validation sets were randomly generated in an 80-20 ratio from the full train dataset provided without explicitly maintaining the same distribution of classes in train and validation sets to create a tougher validation environment.

**Auxilliary Features**

Utilises the following special modules-

Other Model features:

- Batch size- 64
- Loss function- Label Smoothing Cross Entropy Loss, smoothing = 0.1
- Optimizer- AdamW optimizer with initial Learning rate 0.001
- Scheduler- Cosine Annealing Scheduler with T_max = 10
- Weights initialization:- None
- Total parameters:- Approx 10.6 million
- Input image size:- 224x224

**Model Architecture**

1. **Input**: 224 × 224 × 3, RGB Image Input
2. **Initial Feature Extraction**:
    a. Conv2d: 224 × 224 × 64, 3×3 kernel, stride=1, padding=1
    b. BatchNorm2d + GELU: 224 × 224 × 64
    c. Conv2d: 112 × 112 × 64, 3×3 kernel, stride=2, padding=1
    d. BatchNorm2d + GELU: 112 × 112 × 64
    e. MaxPool2d: 56 × 56 × 64, 2×2 kernel, stride=2
3. **Block 1**:
    a. ResidualBlock: 56 × 56 × 128, with SEBlock attention

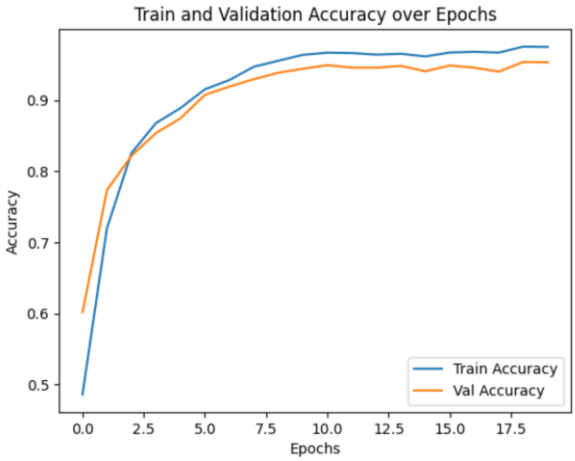
    b. MaxPool2d: 28 × 28 × 128, 2×2 kernel, stride=2
4. **Double Residual Block**:
    a. ResidualBlock 1: 28 × 28 × 256, with SEBlock attention
    b. ResidualBlock 2: 28 × 28 × 256, with SEBlock attention
    c. MaxPool2d: 14 × 14 × 256, 2×2 kernel, stride=2
5. **Inception Module**:
    a. InceptionBlock: 14 × 14 × 512, 4 paths (1×1 conv, 1×1→3×3 conv, 1×1→5×5 conv, pool→1×1 conv)
6. **SEBlock**: 14 × 14 × 512, channel attention
7. **MaxPool2d**: 7 × 7 × 512, 2×2 kernel, stride=2
8. **Spatial Attention**: 7 × 7 × 512, 7×7 kernel spatial attention
9. **Deep Features**:
    a. Conv2d: 7 × 7 × 1024, 3×3 kernel, dilation=2, padding=2
    b. BatchNorm2d + GELU: 7 × 7 × 1024
    c. DropBlock2d: 7 × 7 × 1024, block_size=7, drop_prob=0.1
    d. MaxPool2d: 3 × 3 × 1024, 2×2 kernel, stride=2
10. **Pooling Branches**:
    a. AdaptiveAvgPool2d: 1 × 1 × 1024, global average pooling
    b. AdaptiveMaxPool2d: 1 × 1 × 1024, global max pooling
    c. Concatenate: 2048, flatten and concatenate
11. **Classifier**:
    a. Dropout: 2048, p=0.3
    b. Linear + GELU + BN: 1024
    c. Dropout: 1024, p=0.3
    d. Linear + GELU + BN: 512
    e. Linear: num_classes, final classification layer

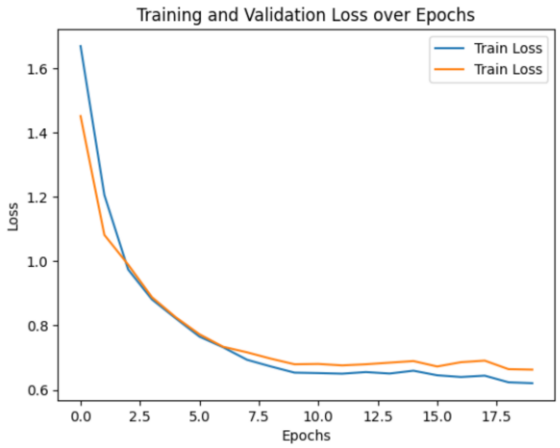
**Optimisations**

We have added batch normalisation, dropouts. We replaced ReLU with GeLU for better gradient flow. We also tried focal loss and Cauchy loss but found label smoothing cross entropy even without weights to perform better in terms of convergence. Adding the spacial attention also improved performance. In general, a deeper model allowed convergence in fewer epochs. For determination of optimal hyperparameters, we did not use k-fold cross entropy as that was too time taking and simply observed the results from the first n epochs and treating them as a heuristic. Tried OneCycleLR scheduler but found its convergence rate to be quite slow. In general, the interpretability of the model is low.

**Variation of Loss and Accuracy over time**

The graphs show that the transition is quite smooth. In fact, model reached peaked accuracy in just 11 epochs, with each epoch taking around 3min 40s to train and 50s to validate the data.



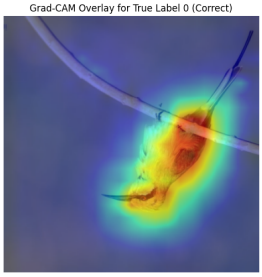




**GRAD-CAM Interpretation**

The CAMs show that the model generally identifies a bird from its face and beak. However, the CAM for the crow shows that it is being identified from a portion of its feathers, which is not very intuitive as multiple birds can have black feathers. Sometimes, the model can get overfitted to the boundaries of the bird, due to which misclassifications can occur. Please find the GRAD cams, one for each training example on the next page.

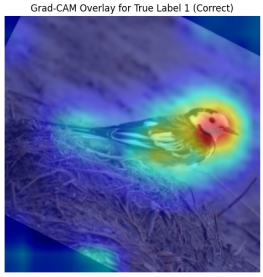
Original Image

Grad-CAM Overlay for True Label 0 (Correct)



Original Image

Grad-CAM Overlay for True Label 1 (Correct)

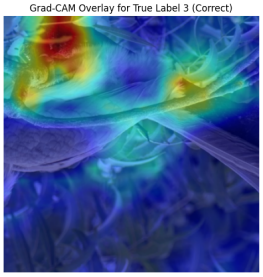


Original Image

Grad-CAM Overlay for True Label 2 (Correct)

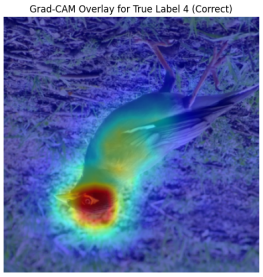
Original Image

Grad-CAM Overlay for True Label 3 (Correct)

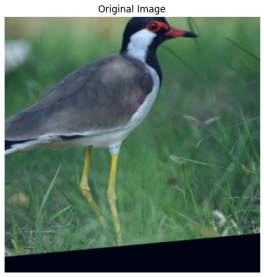


Original Image

Grad-CAM Overlay for True Label 4 (Correct)



Original Image



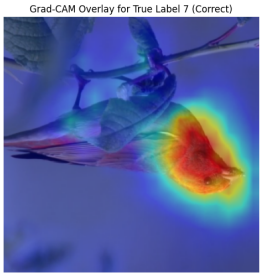
Grad-CAM Overlay for True Label 5 (Correct)

Original Image

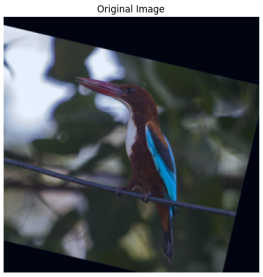
Grad-CAM Overlay for True Label 6 (Correct)

Original Image

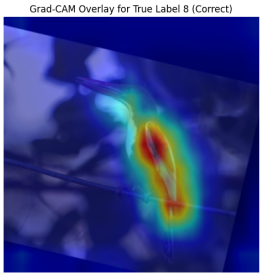
Grad-CAM Overlay for True Label 7 (Correct)



Original Image



Grad-CAM Overlay for True Label 8 (Correct)

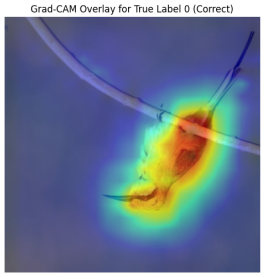


Original Image

Grad-CAM Overlay for True Label 9 (Correct)

Original Image

Grad-CAM Overlay for True Label 0 (Correct)



Original Image

Grad-CAM Overlay for True Label 1 (Correct)