AVIRAL SINGH – 2022CS11290 **PARTH VERMA – 2022CS11936**

Report on Bird Classification Using CNN

Introduction

In this report, we present a convolutional neural network (CNN) model developed for bird species classification. The model is trained on a dataset of bird images and utilizes data augmentation and regularization techniques to enhance performance.

Model Architecture

The CNN model consists of five convolutional blocks followed by fully connected layers. Each convolutional block includes a convolutional layer, batch normalization, ReLU activation, and max pooling. The architecture is designed to handle input images of size **300x200** pixels.

Total no of parameters in the model = 1,704,458

Layer Type	Output Size	Filter Size / Stride	Activation
Input Layer	3 x 300 x 200	-	-
Convolutional Block 1			
- Conv2D (3 → 32)	32 x 300 x 200	3 x 3 / 1	ReLU
- BatchNorm2D	32 x 300 x 200	-	-
- MaxPool2D	32 x 150 x 100	2 x 2 / 2	-
Convolutional Block 2			
- Conv2D (32 → 64)	64 x 150 x 100	3 x 3 / 1	ReLU
- BatchNorm2D	64 x 150 x 100	-	-
- MaxPool2D	64 x 75 x 50	2 x 2 / 2	-
Convolutional Block 3			
- Conv2D (64 → 128)	128 x 75 x 50	3 x 3 / 1	ReLU
- BatchNorm2D	128 x 75 x 50	-	-
- MaxPool2D	128 x 37 x 25	2 x 2 / 2	-
Convolutional Block 4			
- Conv2D (128 → 256)	256 x 37 x 25	3 x 3 / 1	ReLU
- BatchNorm2D	256 x 37 x 25	-	-
- MaxPool2D	256 x 18 x 12	2 x 2 / 2	-
Convolutional Block 5			
- Conv2D (256 → 512)	512 x 18 x 12	3 x 3 / 1	ReLU
- BatchNorm2D	512 x 18 x 12	-	-
- MaxPool2D	512 x 9 x 6	2 x 2 / 2	-
Global Average Pooling	512 x 1 x 1	AdaptiveAvgPool2D	-
Fully Connected Layers			
- Flatten	512	-	-
- Linear (512 → 256)	256	-	ReLU
- Dropout (p=0.5)	256	-	-
- Linear (256 → 10)	10	-	-

Training and Validation Loss and Accuracies vs. Epochs

Observation:

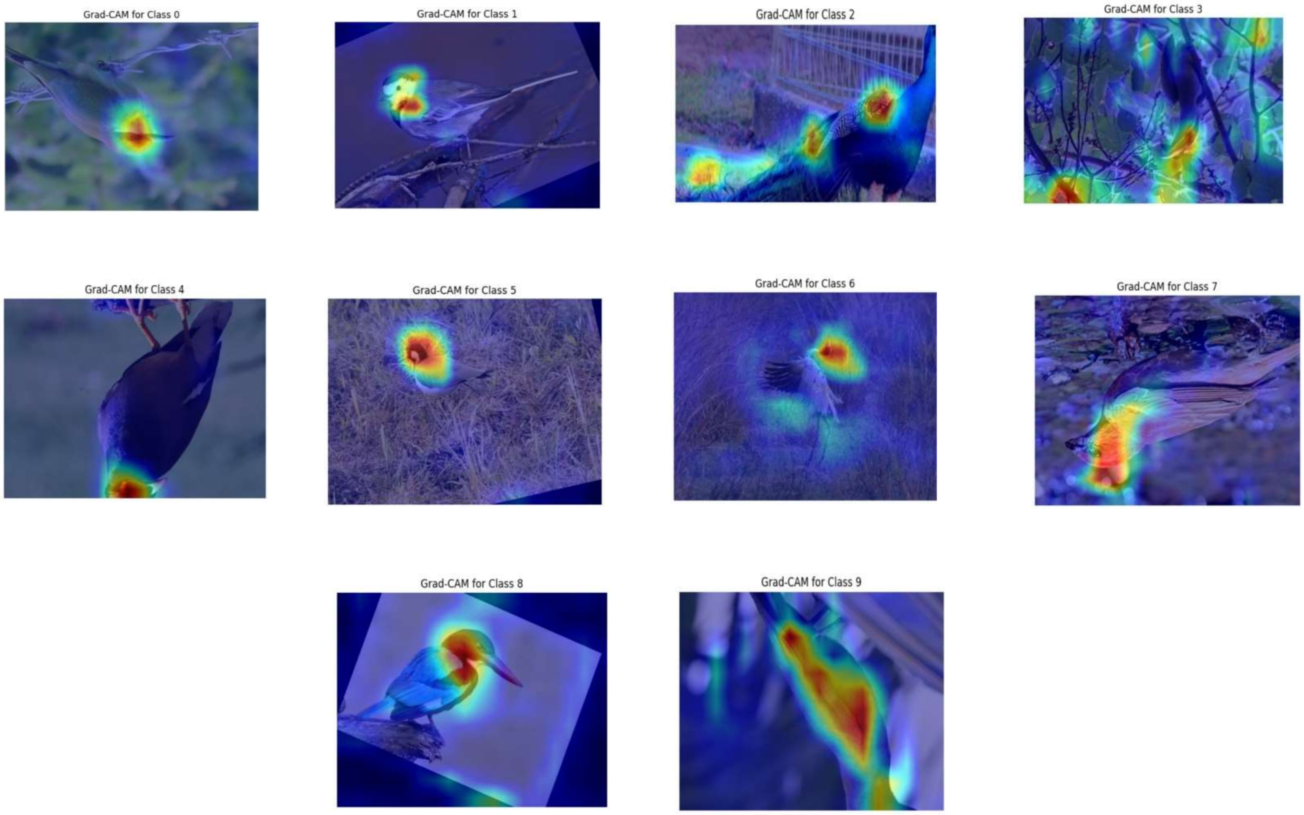
- Training loss has smooth decrement, while validation loss has minute variations.
- Training accuracy increases, reaches almost 99.98%, validation accuracy reaches 97.29%.
- Ass the no of epochs increases model better fits to data and due to regularization introduced also generalizes well.

Effect of Model Optimization Techniques

Technique Used	Validation Accuracy (%)	Observation
Baseline (No Augmentation, No Regularization)	93.57	Initial model performance without optimizations.
Data Augmentation (Horizontal Flip, Vertical Flip, Rotation)	95.89	Improvement due to increased data diversity.
Regularization (Dropout p=0.5 and Batch Normalization)	96.19	Reduction of overfitting through dropout layers.
Combined Regularization with Data Augmentation	96.65	Combining Regularization further increases accuracy
Combined Techniques (adding a learning rate custom scheduler)	97.29 Macro F1 Score: 0.9711 Micro F1 Score: 0.9729	Best performance achieved with combined techniques.

Early stopping and L2 Regularization were not much effective in improving accuracy, since our model requires less no of parameters (approx. 1.7M), therefore does not highly overfits.

Class Activation Maps (CAM)



Analysis:

- **Class 0**: Model focuses on the head, likely using shape and colour for identification.
- **Class 1**: Head and neck highlighted, suggesting reliance on head patterns.
- **Class 2**: Broader body and feather patterns focused, indicating body features are key.
- **Class 3**: Distributed focus across the bird, implying use of general colour or posture.
- **Class 4**: Head and neck region emphasized; head features likely distinct for this class.
- **Class 5**: Strong focus on eyes and beak, suggesting unique head traits are essential.
- **Class 6**: Head and wing areas highlighted, likely due to specific patterns there.
- **Class 7**: Head and upper body focused, combining chest details for identification.
- **Class 8**: Head area alone is key, implying sufficient distinctive features.
- **Class 9**: Distributed across head and body, using multiple regions for classification.

In summary, the Grad-CAM visualizations reveal that the model predominantly relies on the head and upper body regions of the birds for classification across classes. This is expected, as these areas often have distinct features that differentiate bird species. In cases where the focus is more distributed, the model might be considering general colour or shape patterns when specific features are less pronounced.