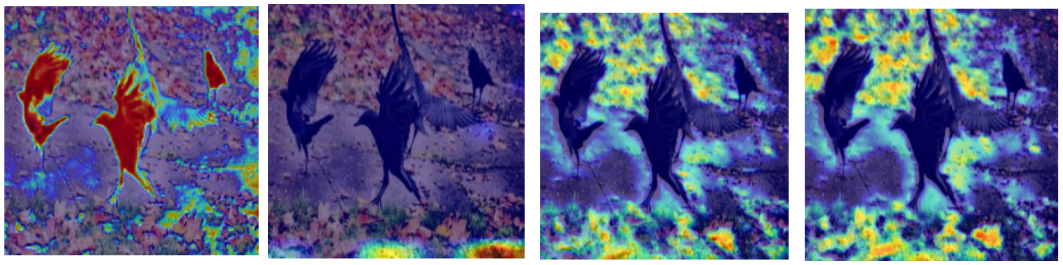
Savudam Sai Sathvik 2024JRB2022 Nattibayannagari Parthasaradhi Reddy 2024JRB2028

We got to this final architecture after trying more than 20 other architectures. Here is the thought process we went through before finalizing our architecture:

- First we thought resizing images is a bad idea because we lose a lot of information, so we tried training on 1200, 720, 480, 224 size pictures. Interestingly we found resizing to 224 is good, as the transition from convolutions layers and MLP layers is smoother for neural networks with depth 8. High pixel size pictures require much deeper networks.
- We noticed the key area to focus was the transition from convolution to MLP layers. We tried different kernels sizes and maxpool combinations, good decent results.
- Later we read research papers on popular image classification softwares like Alexnet, and Yolo research papers and our model is inspired from alexnet.
- We tried different kernel sizes on the first layer 3x3, 5x5, 7x7, 11x11. We found that kernel size 11 and 7 are capturing important big features, while kernel size 3 is trying to capture small features like curves and edges. As you can see below(kernel 3,5,7,11 Gradcam cnn1). We choose kernel size 7.



Final Architecture:

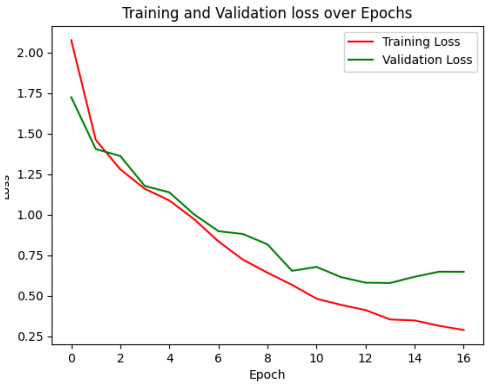
We are using 5 convolution layers, and 4 fully connected layers.

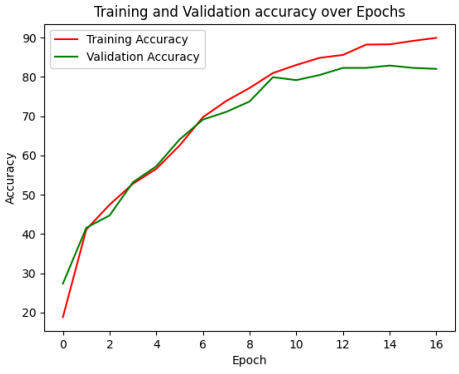
Layer No	Type of Layer	channel size/ No of Perceptrons	Activation Function	Kernel size	Stride
1	CNN	48	ReLu	7 X 7	4
	Max Pool			5 X 5	2
2	CNN	128	ReLu	3x3	1
	Max Pool			3x3	2
3	CNN	198	ReLu	3x3	1
4	CNN	198	ReLu	3x3	1
5	CNN	128	ReLu	3x3	1
6	ANN	2048	ReLu		
7	ANN	2048	ReLu		
8	ANN	1000	ReLu		
9	ANN	10			

Optimizer: Adam
Epochs: 20 (Max).Early stopping is applied when, model starts to overfit.
Batch size: 32
Learning rate: 0.0005
No of Parameters: 19675458 (**20M** approx).

Data Agumentation after resizing(224)	Methods	Early Stopping	Validation Accuracy
Random Horizontal Flip + RandomRotate (5 degrees)	Normal	No	89.91
Random Horizontal Flip + Color Jitter (0.7,1.3)	Normal	Yes	82.28
Random Horizontal Flip + Color Jitter (0.7,1.3)	Dropout	Yes	80.37
Random Horizontal Flip + Color Jitter (0.7,1.3)	Batch Normalization	Yes	81.12

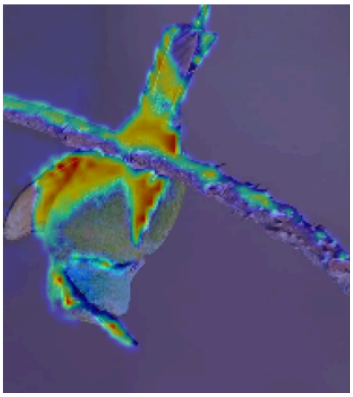
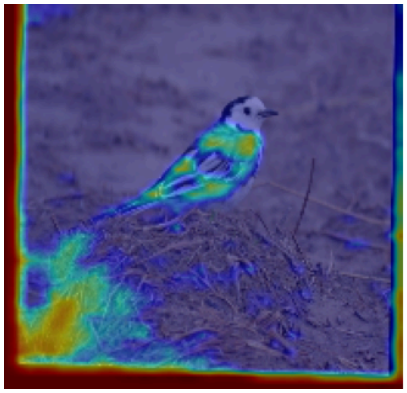
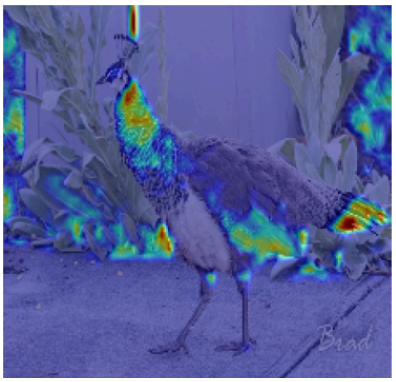
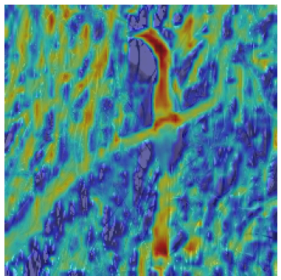
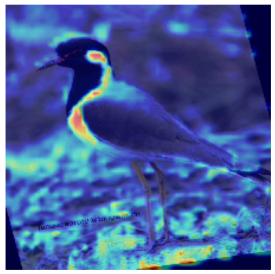
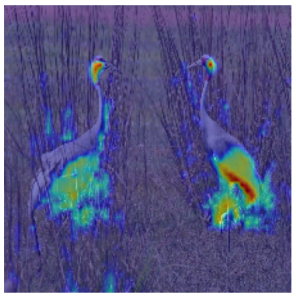
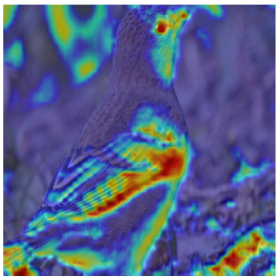
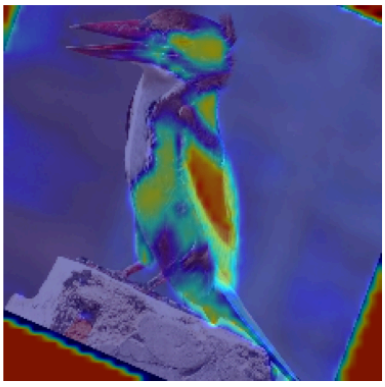
Performance:
Training Loss: 0.3549 at 14th epoch
Validation Loss: 0.5795 at 14th epoch
Training accuracy: 88.22
Validation accuracy: 82.28

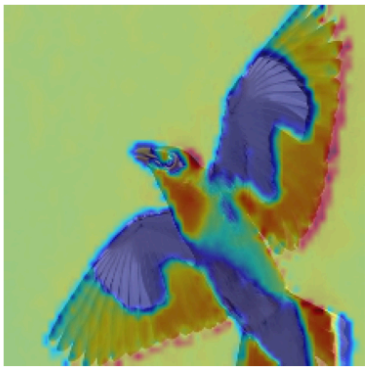




We are using early stop to stop training the model to prevent overfitting(patience=3) and saving the best model with the lowest loss.

Gradcam CNN1 features:



For most of the birds the features extracted are beak, wings(wings edges), feathers of the birds, bird tail and legs, and it also extracting unique color features of birds like in class 0,1.