Name: Rahul Bhagat

Module 4 – Project: Module 4 (Deep Learning - train a network CNN for image classification on mini-ImageNet datasets)

Part A: CNN Architecture Experiments on Mini-ImageNet

Abstract

This report presents a systematic study of convolutional neural network (CNN) architecture and training parameters for image classification on the mini-ImageNet dataset. By varying network depth, filter capacity, pooling strategy, stride, and training duration, we analyze their effect on classification performance and identify a high-performing model.

Overview

Convolutional neural networks are widely used for visual recognition tasks, where architectural choices play a crucial role in determining performance. In this work, a configurable CNN was trained from scratch on mini-ImageNet. A small number of controlled experiments were designed to study the influence of key parameters without performing exhaustive hyperparameter search.

Experiment

Three CNN configurations were evaluated.

Each experiment differs in convolutional depth, number of filters, fully connected layers, pooling strategy, stride, and training epochs. All models were evaluated on a held-out test set.

The results are shown in the below table Table-1.

Experiment	Conv Filters	Conv Layers	FC Layers	Max Pooling	Stride	Epochs	Test Accuracy
1	[32, 64]	2	[128]	Yes	1	10	41.66%
2	[32, 64, 128]	3	[256, 128]	Yes	1	15	43.2%
3	[32, 64]	2	[128]	No	2	20	Lower than Experiment #2

Table 1: Results CNN configurations

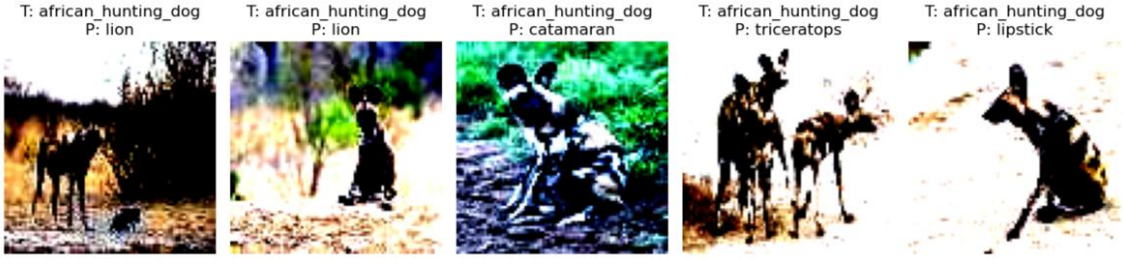
Observation

The experimental results show that increasing the depth of the CNN and the number of filters leads to a noticeable improvement in classification accuracy. The deeper model trained for more epochs achieved the best performance. In contrast, replacing max pooling with stride-based down- sampling has actually reduced accuracy, which indicates loss of fine-grained spatial information.

Misclassification

A subset of misclassified test images produced by the best-performing model was visually inspected. Common error patterns include confusion between visually similar classes, influence of background context, and partial visibility of objects.

Below figure shows some misclassification image T: True Label, P: Prediction



Figer1 : Misclassification

Summary

This study confirms that deeper CNN architectures with sufficient filter capacity and training time are better suited for the mini-ImageNet classification task. Max pooling proved more effective than stride-based down-sampling in preserving discriminative features. The best-performing model identified in this study was subsequently used for interpretability analysis in Part B.

Part B: Occlusion Sensitivity Analysis on Mini-ImageNet

Abstract

This part of the study focuses on understanding the decision-making behavior of the best-performing convolutional neural network (CNN) obtained in Part A. Using occlusion sensitivity analysis, we investigate whether the model bases its predictions on object-centric features or relies on surrounding contextual cues. The analysis is performed on selected test images by systematically occluding local regions and observing changes in classification confidence.

Overview

While quantitative accuracy provides a measure of model performance, it does not explain how or why a model arrives at a particular prediction. Occlusion sensitivity is an interpretability technique that helps identify image regions that are most influential for a model's decision. By masking small regions of the input image and monitoring confidence changes, insights into spatial feature dependence can be obtained.

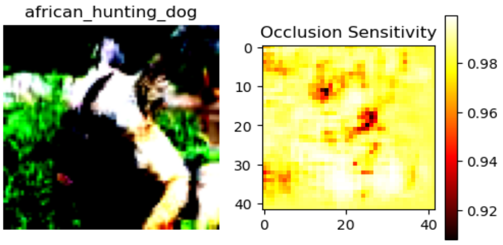
Experiment

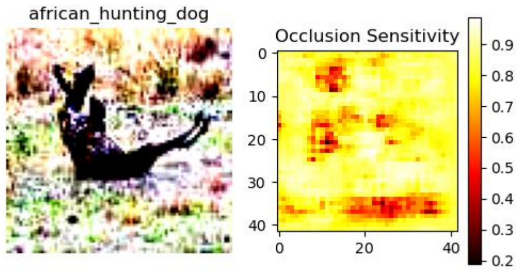
The CNN model with the highest test accuracy from Part A was selected for this analysis. Approximately ten correctly classified images from the test set were chosen. For each image, an N × N occlusion window was slid across the image in both horizontal and vertical directions. At each spatial location, the pixels within the window were replaced with gray values, and the modified image was passed through the trained model.

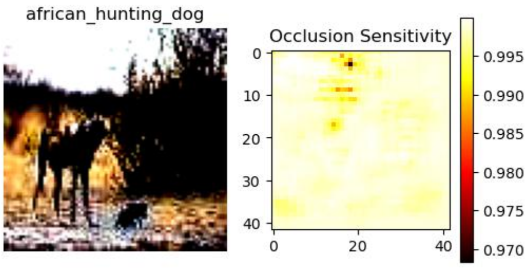
The *SoftMax* probability corresponding to the true class was recorded as confidence (i, j). The resulting confidence matrix was visualized as a heatmap to highlight regions critical to the model's prediction.

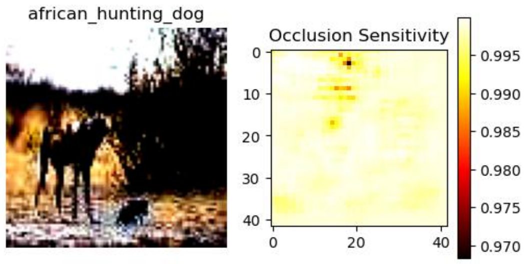
Below is the table that summarizes these results and images with occlusion sensitivity

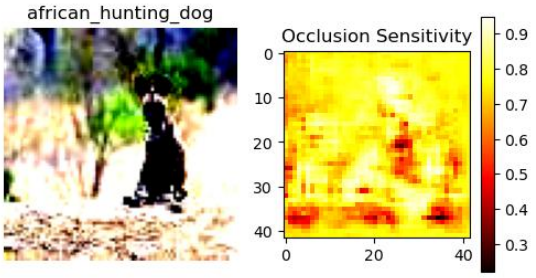
Parameter	Value	Reason for Choice	Effect Observed
Occlusion Window Size (N)	8 × 8	Balances locality and computational cost	Clearly highlights discriminative regions
Stride	2	Reduces computation while preserving spatial trends	Smooth confidence maps
Replacement Value	Gray pixels	Neutral input after normalization	Suppresses local features effectively
Images Analyzed	~10 test images	Representative qualitative analysis	Consistent behavioral patterns observed

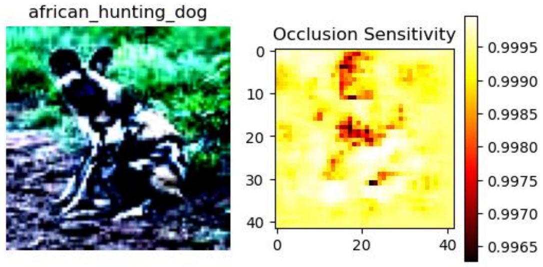

african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.98
0.96
0.94
0.92

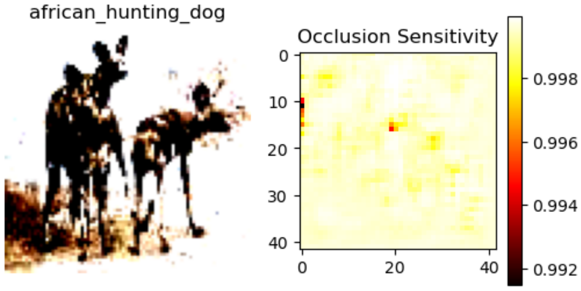

african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

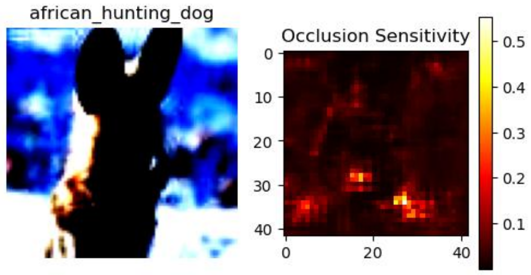

african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.995
0.990
0.985
0.980
0.975
0.970

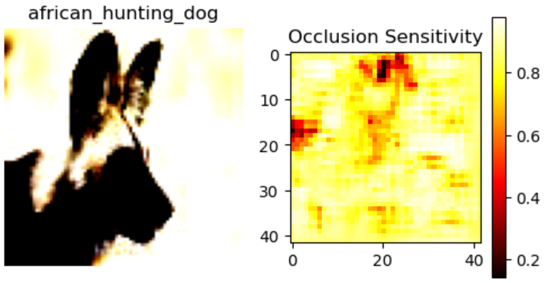

african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.995
0.990
0.985
0.980
0.975
0.970

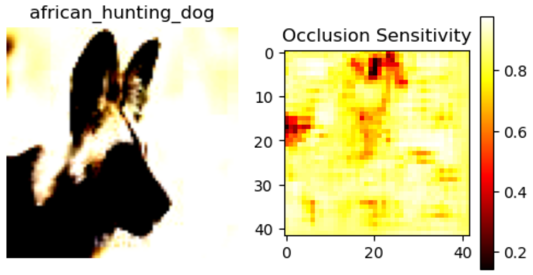

african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.9
0.8
0.7
0.6
0.5
0.4
0.3


african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.9995
0.9990
0.9985
0.9980
0.9975
0.9970
0.9965


african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.998
0.996
0.994
0.992


african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.5
0.4
0.3
0.2
0.1


african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.8
0.6
0.4
0.2


african_hunting_dog
Occlusion Sensitivity
0
10
20
30
40
0
20
40
0.8
0.6
0.4
0.2

Summary

This occlusion sensitivity analysis provides qualitative evidence that the best-performing CNN model primarily relies on object-relevant regions for classification decisions.

The results indicate successful spatial feature learning, while also highlighting limitations related to background dependence in certain cases. These insights complement the quantitative results from Part A and contribute to a more comprehensive understanding of the model’s behavior.