Programming Assignment 5

Report submitted to

Indian Institute of Technology Mandi for the course

of

CS671: Deep Learning and Applications

by

Piyush Verma (B20163)

Yash Bhotmagey (B20173)

INDIAN INSTITUTE OF TECHNOLOGY MANDI

April 2023

Abstract

In this assignment we trained CNN classification models on a subset of the Caltech-101 dataset. We compared different architectures and studied how the CNN models learn by visualizing the feature maps, gradient maps and localization maps as well as the image patches that maximally fire a neuron.

Contents

Convolutional Neural Networks

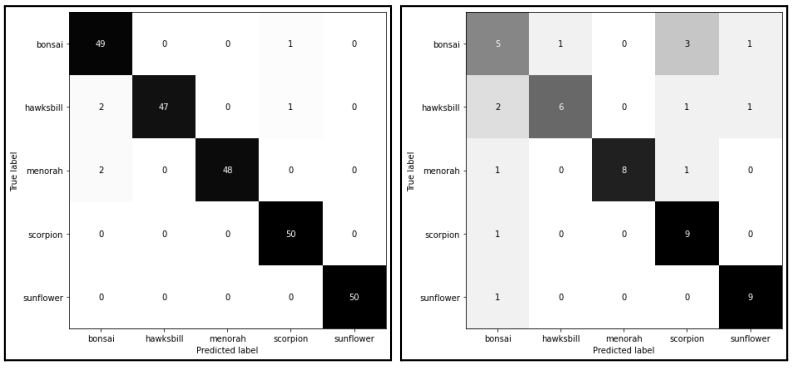
Task 1

- Different architectures of CNN were trained on a subset of the Caltech-101 dataset for image classification.
- The classification accuracy of each of the models on the validation dataset was used to determine the best architectures for our task.
- The feature maps for all the convolutional layers of the best architecture were plotted for an image to visualize the features computed in each layer.
- The neuron that was maximally activated in the output layer was traced back to determine the patch in the original image that caused it to activate maximally.

Architecture 1 (4 layer CNN)

Layer Num	Layer (Type)	Output Shape	# Parameters
1	Conv2D	(None, 56, 56, 8)	2912
	MaxPooling 2D	(None, 27, 27, 8)	0
2	Conv2D	(None, 27, 27, 16)	3216
	MaxPooling 2D	(None, 13, 13, 16)	0
	Flatten	(None, 2704)	0
3	Dense	(None, 128)	346240
4	Dense	(None, 5)	645

Table 1: Layer description for architecture 1



(a) **(b)**

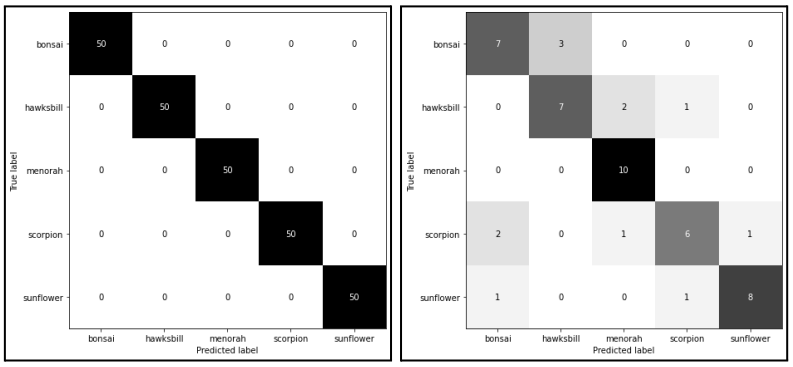
Figure 1: Confusion matrices for (a) Training Set and (b) Validation Set

- The classification accuracy on the training set was 97.60%.
- The classification accuracy on the validation set was 74.00%.

Architecture 2 (5 layer CNN)

Layer Num	Layer	Output Shape	# Parameters
1	Conv2D	(None, 56, 56, 8)	2912
	MaxPooling 2D	(None, 27, 27, 8)	0
2	Conv2D	(None, 27, 27, 16)	3216
	MaxPooling 2D	(None, 13, 13, 16)	0
3	Conv2D	(None, 13, 13, 32)	4640
	MaxPooling 2D	(None, 6, 6, 32)	0
	Flatten	(None, 1152)	0
4	Dense	(None, 128)	147584
5	Dense	(None, 5)	645

Table 2: Layer description for architecture 2



(a) **(b)**

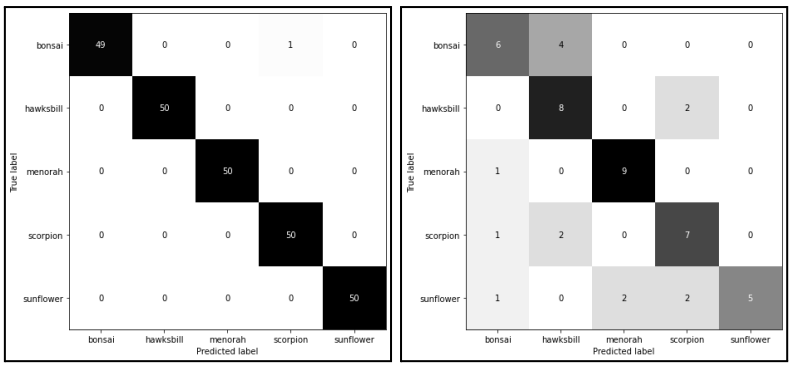
Figure 2: Confusion matrices for (a) Training Set and (b) Validation Set

- The classification accuracy on the training set was 100%.
- The classification accuracy on the validation set was 74.00%.

Architecture 3 (6 layer CNN)

Layer Num	Layer (Type)	Output Shape	# Parameters
1	Conv2D	(None, 56, 56, 8)	2912
	MaxPooling 2D	(None, 27, 27, 8)	0
2	Conv2D	(None, 27, 27, 16)	3216
	MaxPooling 2D	(None, 13, 13, 16)	0
3	Conv2D	(None, 13, 13, 32)	4640
4	Conv2D	(None, 13, 13, 64)	18496
	MaxPooling 2D	(None, 13, 13, 64)	0
	Flatten	(None, 2304)	0
5	Dense	(None, 128)	295040
6	Dense	(None, 5)	645

Table 3: Layer description for architecture 3



(a) **(b)**

Figure 3: Confusion matrices for (a) Training Set and (b) Validation Set

- The classification accuracy on the training set was 99.60%.
- The classification accuracy on the validation set was 70.00%.

Best Model Architecture

- Both architectures 1 and 2 have the same classification accuracies on the validation set. Hence architecture 2 is chosen as the best architecture for having higher training set accuracy.
- The chosen model has an accuracy of 76.00% on the test set.

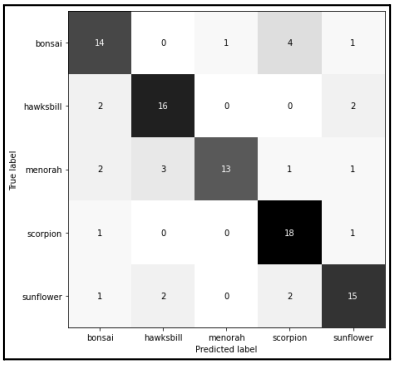


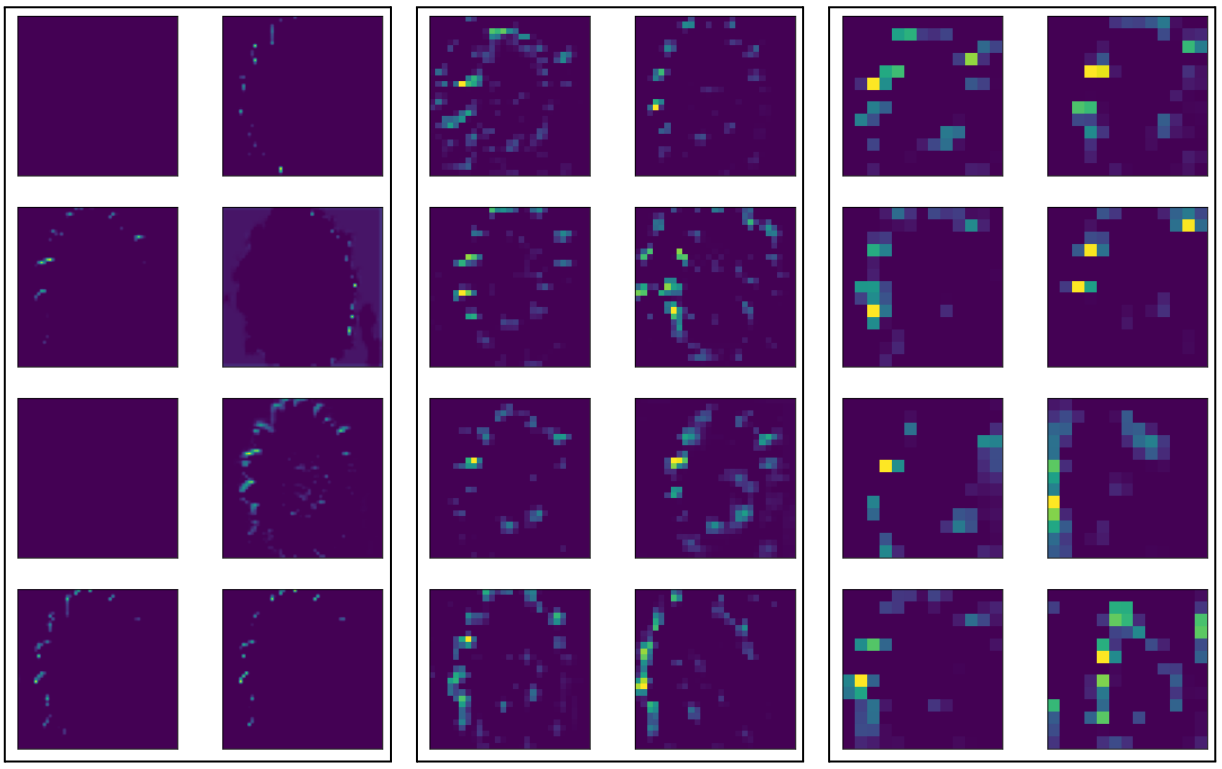
Figure 4: Confusion matrix for best model for Test Set

- Furthermore, it can be seen that although the training set accuracy for all the architectures is high, they perform poorly on the test and validation dataset. This means that the models have overfit the training dataset which is probably a consequence of the small dataset.

Feature Maps of Different Convolutional Layers (Architecture 2)

- A sample image from the sunflower class is passed through the best model architecture (5 layer CNN) and the feature maps for all the convolutional layers are plotted.

Figure 5: Sample image from the sunflower class



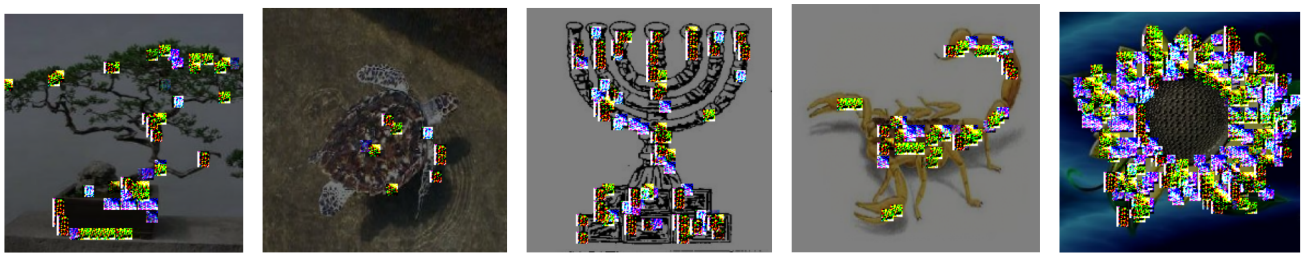
(a) **(b)** **(c)**

Figure 6: Feature maps for (a) Conv layer 1, (b) Conv layer 2 and (c) Conv layer 3

- A lot of similarities can be seen between the original image and the feature maps of that image.
- The feature maps for the first convolutional layer contain thin strokes resembling the boundary of the sunflower. The layer has learned to detect very basic features such as edges in the image.

- The feature maps for the second and third convolutional layers contain patches resembling the petals of the sunflower. These layers have learned more complex features about the image.
- It can also be seen that the dimensions of the feature maps reduce with every layer which matches with the layer description in Table 2.

Maximally Activating Patches



(a) Bonsai **(b)** Hawksbill **(c)** Menorah **(d)** Scorpion **(e)** Sunflower

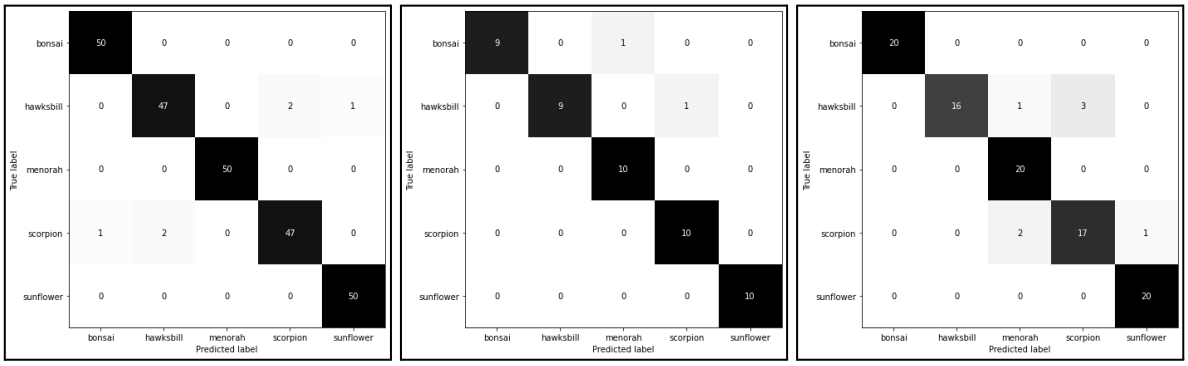
Figure 7: Maximally activating for sample images from all the classes

- Figure 7 represents the patches in the original image that maximally activate the classifying neuron in the output layer i.e these patches are most responsible for determining the class of that image by the model.
- It can be seen that the patches for each of the images contain some information unique to their classes such as,
    - Green leaves on Bonsai trees
    - The head of Hawksbill
    - Lamp on Menorah
    - Claws of Scorpion
    - Yellow petals of Sunflower

Task 2

- The VGG19 model pretrained on the ImageNet dataset modified for the given dataset by retraining its classification layer and its performance was compared to the best model from task 1.
- The neuron that was maximally activated in the output layer was traced back to determine the patch in the original image that caused it to activate maximally.
- Guided-Backpropagation algorithm was used to visualize the influence of input pixels on the neurons in the last convolutional layer of pretrained VGG19 network.
- GradCAM was used to visualize the localization map (heat map) which highlights the important regions in an image for predicting its class and other classes.

Model Performance



(a) **(b)** **(c)**

Figure 8: Confusion matrices for (a) Training set, (b) Validation set and (c) Test set

- The classification accuracy on the training set was 97.60%.
- The classification accuracy on the validation set was 96.00%.
- The classification accuracy on the test set was 93.00%.
- Compared to the best model from task 1 the classification accuracy of the modified VGG19 model is lower. However the model performs far better than any of the previous models on the validation and test set.

Maximally Activating Patches

(a) Bonsai **(b)** Hawksbill **(c)** Menorah **(d)** Scorpion **(e)** Sunflower

Figure 9: Maximally activating for sample images from all the classes

- Figure 9 represents the patches in the original image that maximally activate the classifying neuron in the output layer i.e these patches are most responsible for determining the class of that image by the model.
- Compared to Figure 7 these patches encompass a larger portion of the image and contain more information about its class. For example in 9(b) and 9(d) the patches crop out the bodies of the Hawksbill and Scorpion from their surroundings.

GradCAM

(a) Bonsai

(b) Hawksbill

(c) Menorah

(d) Scorpion

(e) Sunflower

Figure 11: Localization map using GradCAM

- The heat maps overlaid on the images depict the important regions in an image for predicting its class.