Dissecting Convolutional Neural Networks using Activation Maps in Rice Image Classfication

Mentor : Dr. Tushar Sandhan
Indian Institue of Technology Kanpur

ABSTRACT

Explainable Artificial Intelligence (X-AI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterize model accuracy, fairness, transparency and outcomes in AI-powered decision making.
In our case we were tasked to apply the same over a VGG16 classification model.

INTRODUCTION

The aim of the project was explainable - AI, to explain how the model works and which features it uses to classify the image.

We were tasked to apply the same over a VGG16 classification model.

Initially we trained our VGG16 model on the original dataset.

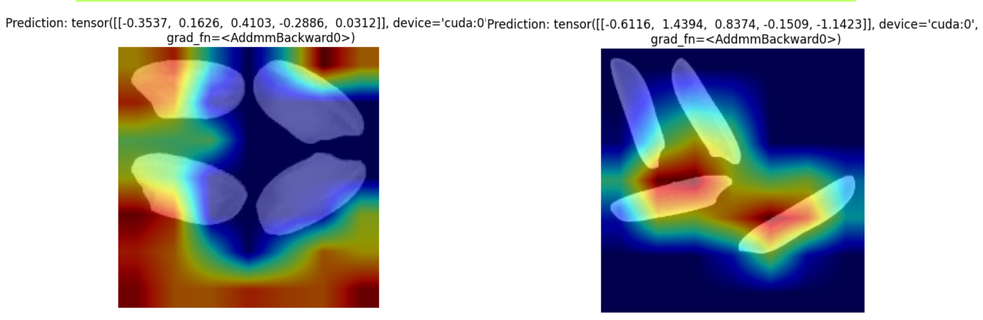
We generated new dataset using transformations and tested our model on it.

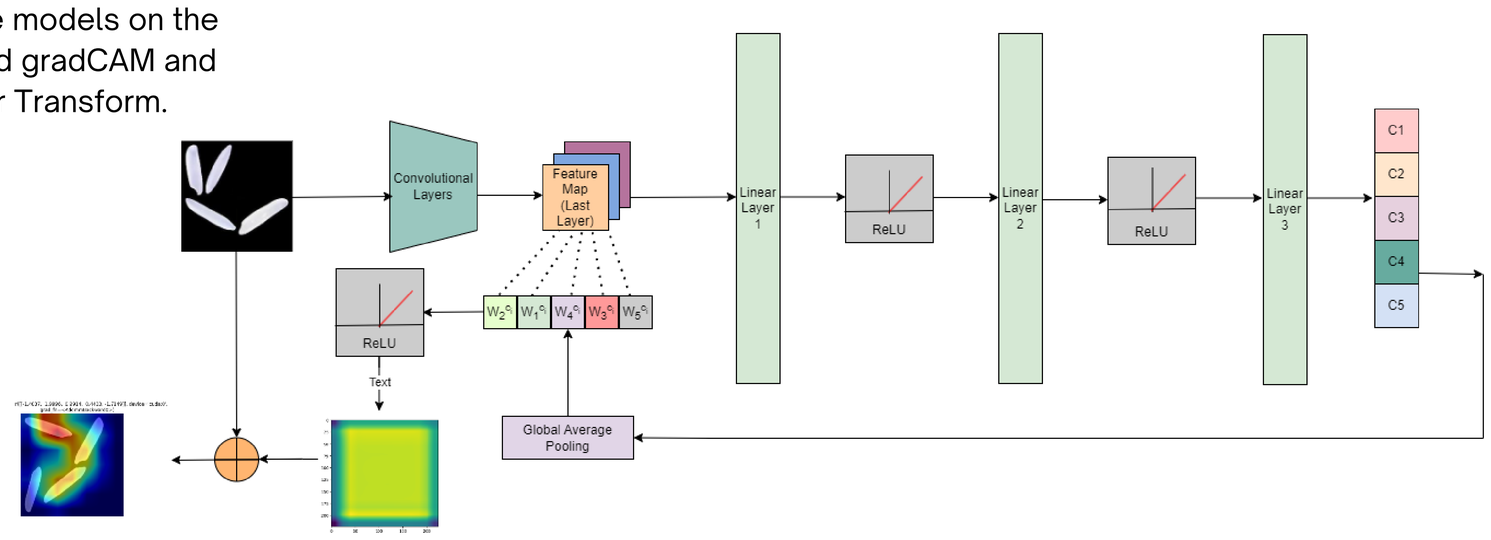
We then tested the models on the dataset,and applied gradCAM and Elliptical Fourier Transform.

METHODOLOGY AND MODELS USED

- We were given a rice image dataset of 5 different classes.
- Generated new datasets using **OpenCV**, **PyTorch** and various transformations(**Affine**).
- We used **VGG-16** model for our classification purpose.
- VGG16 is a **convolutional neural network model** that's used for image recognition. It's unique in that it has only **16 layers** that have weights, as opposed to relying on a large number of hyper-parameters. It's considered one of the best vision model architectures.
- Then we applied **GradCAM** over the model and obtained the heatmaps which were used in analysing the model.
- Then we applied **Elliptical Fourier Transfrom** over the image to obtain the Fourier Coefficents to generate the contours.

gradCAM





MATHEMATICS BEHIND.....

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$

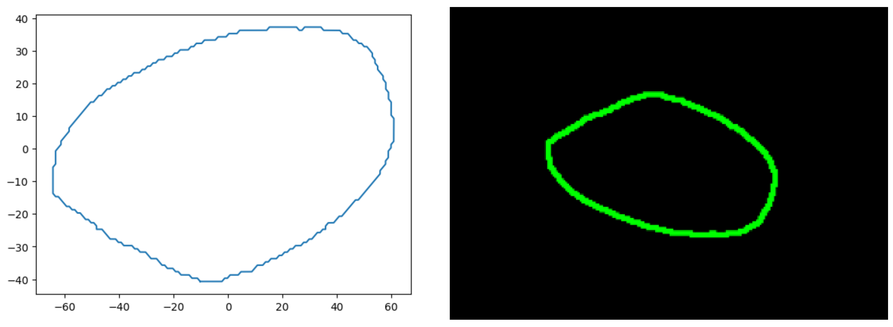
$$L_{\text{Grad-CAM}}^c = ReLU\underbrace{\left(\sum_k \alpha_k^c A^k\right)}_{\text{linear combination}}$$

The above images are **heatmaps generated using GradCAM** using the prediction tensor obtained by passing a test image to the trained model combined with the images . **The colour variation shows the degree of the importance of features as used by the model to classify.**

DATASET

- We used a dataset consisting of **5 different classes of rice** .
- It consisted of **15000 images of each class**, totaling upto **75000 images over the 5 classes**.
- Using **OpenCV** and **PyTorch** Library functions we were able to **generate** ample dataset so that we could train the model over more realistic images.
- This was done so that we could include in our model more **randomising features of real life images of the grains.**

ELLIPTICAL FOURIER TRANSFORM



MATHEMATICS BEHIND.....

$$x(t) = \sum_{n=1}^{N}\left[A_n \cos\left(\frac{2\pi nt}{T}\right) + B_n \sin\left(\frac{2\pi nt}{T}\right)\right]$$

$$y(t) = \sum_{n=1}^{N}\left[C_n \cos\left(\frac{2\pi nt}{T}\right) + D_n \sin\left(\frac{2\pi nt}{T}\right)\right]$$

The second image represents the contour of a single rice grain generated using **cv2 Library** and used this to generate **Elliptical Fourier Coefficents**, then this coefficents were fed to **reconstruct the contour** again as represented in the first image.

CONCLUSION

- The **gradCAM heatmaps** show us the regions and the features of the image that the model took into account with **varying degrees of importance** while training.
- By observing the gradCAM heatmaps we were also able to infer that the **model does not take into account** the **individual aspects of grains** and in some instances even the **spatial orientation** of the grains while classifying.
- The **Elliptical Fourier Transform** generated contours will prove out to be instrumental for training **SVMs** (Support Vector Machines), which will be useful when deploying the classification model over **systems with less computational power such as over drones** as **VGG16** is a **very dense and heavy model** to deploy over such systems.

REFERENCES

[1] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015
[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626

Application No.: 2330401
Name : Porav Rohilla
Mentor : Dr. Tushar Sandhan
Mentor Dept.: Electrical Engineering