



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

www.iith.ac.in

AI5100 hackathon presentation

Multi class classification

Srivatsan T
cs20btech11062@iith.ac.in

M Dinesh Kumar
ep20btech11011@iith.ac.in

The Dataset



Background

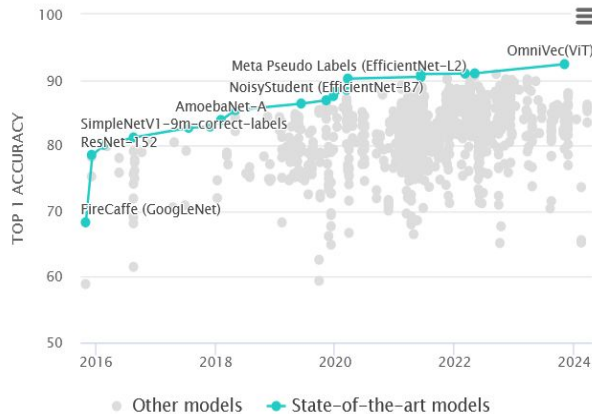
- Task at hand is multiclass classification.
- The dataset contains 50 classes.
- Found it is a subset of **ILSVRC** dataset.

Research

- Looked into the best performing models of ILSVRC.
- Narrowed down options to **ResNet** & **ViT**.
- Researched through the codes of ViT & ResNet.

Strategies

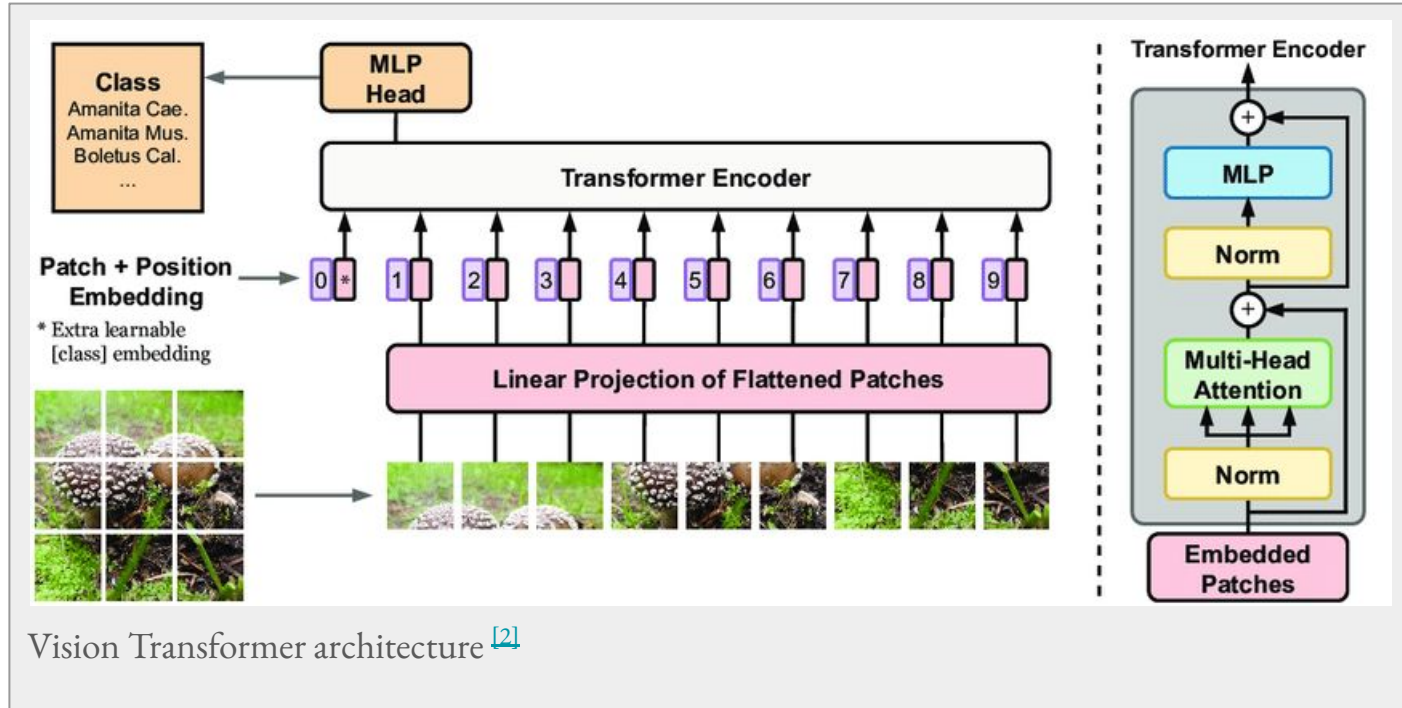
- Started with ViT since it is the **best performing** state of the art model.
- If it doesn't work out choose the more reliable ResNet model (went with **ResNet50**).



Benchmark of image classification task from 2016 to 2024 [\[1\]](#)



Attempts : Vision Transformers



Vision Transformer architecture [\[2\]](#)



Attempts : Vision Transformers

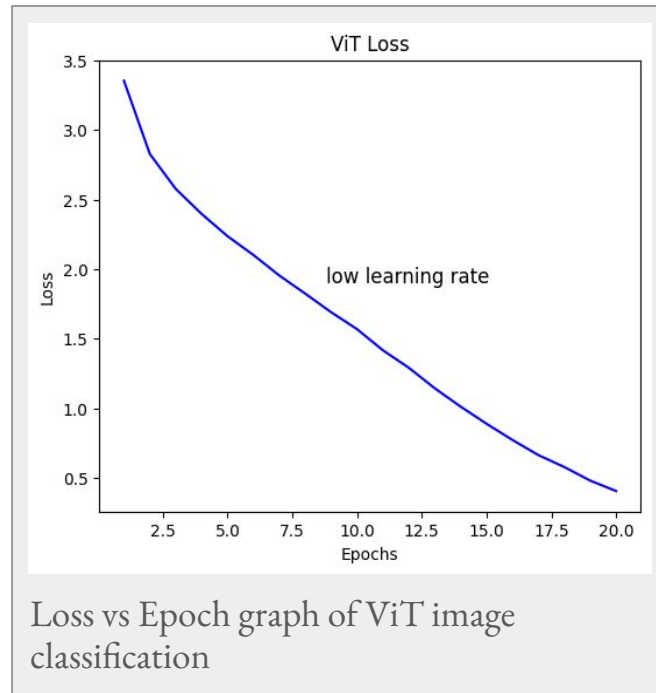


Optimal hyper-parameters not achieved (too many)

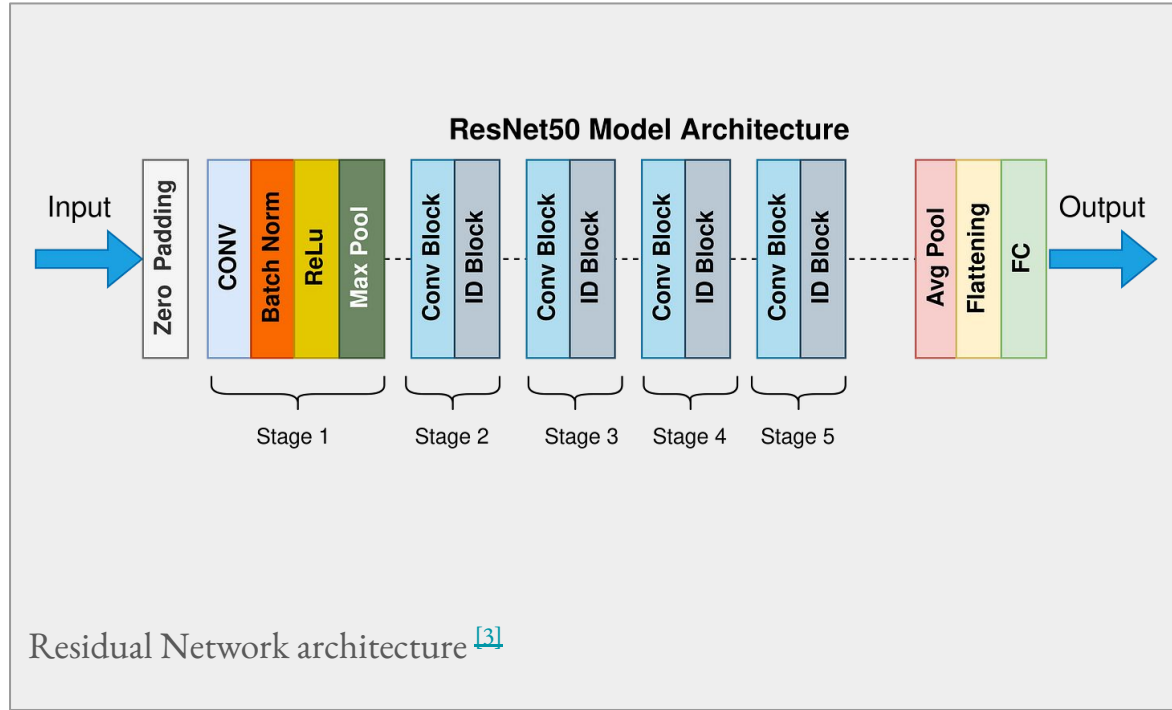
- patch_size = 28
- dim = 1024
- depth = 8
- heads = 8
- mlp_dim = 2048
- dropout = 0.1
- emb_dropout = 0.1
- dim_head = 64
- batch_size = 400
- learning_rate = 0.0001
- num_epochs = 20

Signs of overfitting validation loss kept increasing

- Validation Loss - Epoch [1/20],
Loss: 3.040864667892456
... (decreasing ↓)
Validation Loss - Epoch [10/20],
Loss: 2.431074371337891
... (increasing ↑)
Validation Loss - Epoch [15/20],
Loss: 2.802235622406006
... (increasing ↑)
Validation Loss - Epoch [20/20],
Loss: 3.342058563232422



Attempts : Residual Network



Attempts : Residual Network (ResNet50)



Less hyperparameters less worry

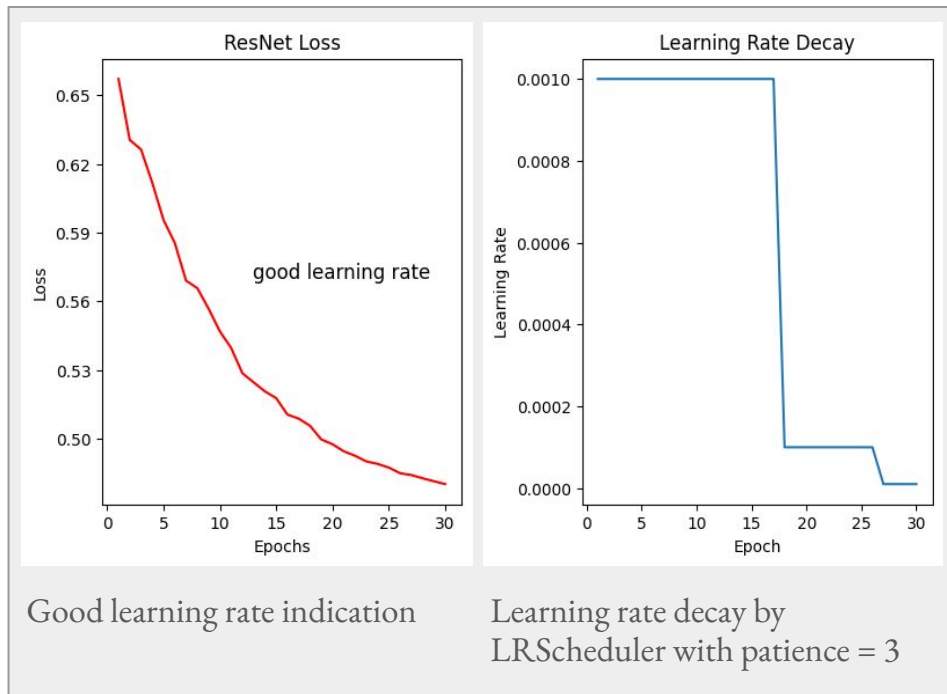
- batch_size = 128
- learning_rate = 0.01
- num_epochs = 60
- patience = 3
- weight_decay = 0.0001
- momentum = 0.9

Improved performance

Validation Loss - Epoch [1/30], Loss:
0.5851225259396937, Accuracy: 0.7966153846153846

... (decreasing ↓)

Validation Loss - Epoch [30/30], Loss:
0.5289132474304793, Accuracy: 0.8417948717948718



Future Works



- ResNet50 could have been **pre-trained** on the ImageNet dataset to obtain better generalisation and, hence, better results.
- Could have tried out **variants** of ResNet which have known to improve the performance of the model on **ImageNet dataset**.
- Efforts could have been taken to explore other architectures like **VGG** which might have **improve** the performance





Thank You



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad
