COMP 6721 (Applied Artificial Intelligence) – Progress Report

Breed Identification System (Group B)

1. Problem Statement

Breed Identification System is designed to classify the breed-class of an animal. With scientific advancement, curiosity has also grown prompting humans to learn more about different species as within the same species there can be numerous breeds. This is difficult to distinguish with the naked eye considering the number of breeds that exist for any animal. Deep learning can be used to solve this issue. Computer Vision models can be built to aid this, we plan on using a combination of scratch training and transfer learning using state-of-the-art CNN models to build a solution of the same.

Challenges: Differences in images used in the dataset, featuring animals of the same breed in a variety of lightings and positions, which is a problem we encountered. The other issue is the lack of balanced datasets which can have hit on overall performance.

Goal of the project is to explore and provide detailed analysis (based on our evaluation metrics) of how different CNN architectures fares against the chosen datasets of animals. We will do a thorough comparison and point out appropriate model through which high accuracy can be achieved. We believe any improvements we make in this project can be ported to other animal breed identification without too much trouble through transfer learning.

2. Proposed Methodologies

Dataset and Preprocessing: The table below shows the different datasets we are using.

Animal	Number of Images	Image sizes	No. of breeds
Dog [1]	20.6 k	{400 - 500} X {310 - 345}	120
Cat [2]	127 k	{300 - 330} X {250 - 270}	67
Fish [3]	9 k	590 X 445	9

The fish dataset was evenly balanced with 1000 images per class (9k total). Whereas, in the dog dataset some classes had less than 75 images, so we eliminated them and reduced the dataset to 50 breeds with a total dropping to 9719 images. The Cat dataset was initially pruned, reducing the number of images to 68k from 127 k and the number of classes to 27 from 67 in the original dataset. Images in the removed classes range from 500 images to 68 images and one class had around 52k images. CNN Models: In terms of CNN architectures [6], we are using ResNet18, ResNet50 and MobileNetV2 and have created 9 instances as train from scratch (considering 3 datasets on each of the three CNN models). We also have 3 instances of transfer-learning the dog dataset. So, there are a total of 12 model instances whose results are discussed below. The reason for choosing these three models is based on our time-taken vs performance analysis of the popular CNN architectures trained for object identification/classification. The ResNet18 model has 72 layers with 18 deep layers and ResNet50 is a CNN with 50 deep layers. However, ResNet50 outperformed the transfer learning method because this pretrained network can classify images into 1000 object categories, including animals, which is relevant to our problem. MobileNetV2, which is a convolutional neural network with 53 deep layers. Being the lightest of 3, naturally it gave mixed results in scratch training but was able to achieve good performance via transfer learning. We have maintained the following hyperparameters across all models.

Optimizer $\rightarrow$ Adam , Loss $\rightarrow$ Cross Entropy Loss, Learning rate $\rightarrow$ 0.001 , Batch size $\rightarrow$ 64

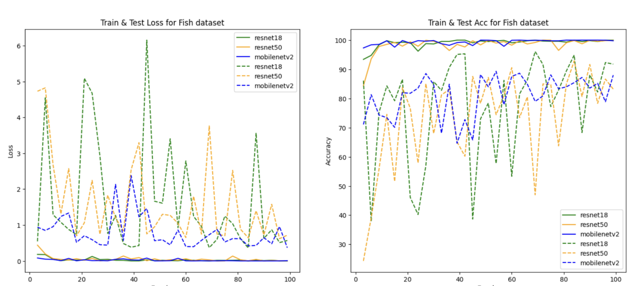
3. Attempts of Solving Problems

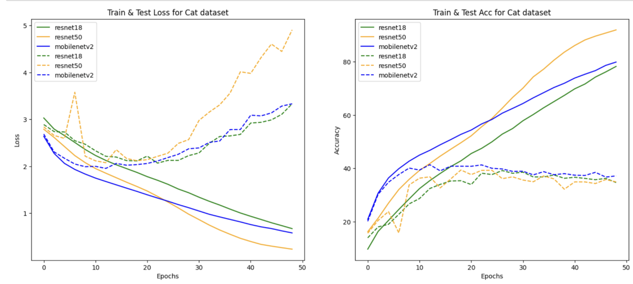
So far, we have 12 model-dataset instances (9 scratch and 3 transfer learning). We have achieved good results across all evaluation metrics on the Fish dataset. Unfortunately for the dog dataset, we obtained very low numbers in evaluation metrics while training from scratch during our first attempt due to improper image distribution among breeds and a smaller number of training samples for each class. Later, the dataset was pruned, and an attempt was made to use standard normalized values (ImageNet) but only achieved 30% accuracy. Finally, during the third iteration, the normalized values were generated for the dataset and then we merged the validation train images instead of our standard 70/10/20 split to increase the train sample size. Accuracies ranging from 58-64% was observed which is a huge leap compared to the previous two trials. For the cat dataset, similar pruning/dropping methods were followed where the number of classes were reduced from 67− > 55− > 27. In our last cut any class below 500 samples were dropped to improve performance. For our latest iteration we used normalized values generated for the dataset. Accuracies started at 15% and through our trails have reached a meteoric 59For evaluation, we went ahead with the built functions of scikit-learn library (classification [5] report). We also calculated class-wise accuracies to understand the discrepancies in evaluation between classes, which helped us understand are generally underfitting when compared to others. (In Cat dataset American Bobtail: 13.23%, American Shorthair: 21.65% these two classes were failing to achieve good acc even after 100s of epochs of training) which raises the question about the quality of images in a few breeds (future improvements). The hyper-parameters [4] like learning rate, batch-size, or optimizer function for transfer learning on cat dataset will be modified to get better results (Ablation Study).

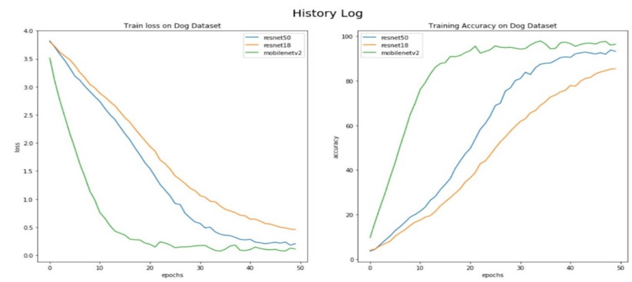
Results Obtained While Training from Scratch:

D $\rightarrow$ Dog ,F $\rightarrow$ Fish , C $\rightarrow$ Cat

	ResNet - 18			ResNet - 50			MobileNetV2		
	D	F	C	D	F	C	D	F	C
Accuracy (%)	69	99	58	69	91	58	64	81	63
Precision (%)	48	99	50	60	92	54	65	85	69
Recall (%)	41	99	49	56	91	51	62	81	51
F-1score (%)	41	99	49	56	91	55	63	80	69
Time Taken per epoch	100	100	500	120	148	784	90	120	477

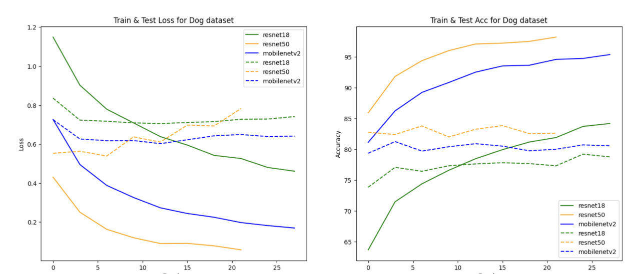






Transfer Learning Results for Dog Dataset:

Animal	Number of Images	Image sizes	No. of breeds
Accuracy (%)	77	83	80
Precision (%)	78	84	81
Recall (%)	77	83	80
F1-score (%)	76	82	80
Time taken per echos (sec)(%)	128	146	122



4. Future Improvements

We have curated the train vs test plots of all 12 instances for loss and accuracy. We plan to visualize the dataset using the TSNE tool. To improve the accuracy of cat resnet-18 model, we are planning to do a detailed ablation study on one of the hyperparameters of our choice.

References

[1] https://www.kaggle.com/competitions/dog-breed-identification/data. 1

[2] https://www.kaggle.com/datasets/ma7555/cat-breeds-dataset. 1

[3] https://www.kaggle.com/datasets/crowww/a-large-scale-fish-dataset. 1

[4] Margherita et a Grandini. "metrics for multi-class classification: an overview." arxiv abs/2008.05756 (2020): n. pag. 2

[5] Xiangyu AU Ren Shaoqing AU Sun Jian PY He, Kaiming AU Zhang. - 2016/06/01 sp - 770 ep - 778 t1 - deep residual learning for image recognition do - 10.1109/cvpr.2016.90 er. 2

[6] Menglong AU Chen Bo AU Kalenichenko Dmitry AU Wang Weijun AU Weyand Tobias AU Andreetto Marco AU Adam Hartwig PY Howard, Andrew AU Zhu. 2017/04/16 sp - t1 mobilenets: Efficient convolutional neural networks for mobile vision applications er. 1

A. Appendix

Dog Sample images:

Fish Sample images:

Cat Sample images: