REPORT

Executive Summary

This project aims to develop a new convolutional neural network designed specifically for computer vision tasks. The dataset comprises a subset of the "Dog-vs-Cats" dataset available on Kaggle. The limitation of available data poses a significant challenge in creating an effective model."

"Convolutional neural networks, commonly known as 'convnets', represent a well-established deep learning model with a strong track record of success in computer vision tasks. Their remarkable feature lies in their capacity to learn and identify spatial patterns present in images. This characteristic renders them particularly suitable for tasks like image recognition, object detection, and segmentation. Despite the constraints of limited available data, it is firmly believed that the 'convnet' model can still deliver satisfactory results. This confidence stems from 'convnets' ability to acquire knowledge and generalize effectively from compact datasets by detecting and recognizing relevant image features. The process involves training the model with the limited dataset, refining it through transfer learning techniques, and assessing its performance through suitable evaluation metrics.

Overall, the goal is to develop an accurate and efficient convolutional neural network that can effectively classify images from the "Dog-vs-Cats" dataset with a limited amount of data.

Problem

The Cats-vs-Dogs dataset is a binary classification problem where the goal is to predict whether an image belongs to the dog class or a cat class.

Techniques

Dataset:

The Cats-vs-Dogs dataset comprises 25,000 images, evenly split between dogs and cats, with a compressed size of 543MB. Upon downloading and decompressing the dataset, I created a new dataset with three subsets:

- A training set consisting of 1,000 samples from each class.
- A validation set with 500 samples from each class.
- A test set containing 500 samples from each class.

Due to the larger image size and the complexity of the problem at hand, it is necessary to expand the capacity of the neural network**.**

To do this, I extended the Conv2D + MaxPooling2D design by one more level. In addition to increasing network capacity, this will make the feature maps smaller so that they won't be too big when they get to the Flatten layer. As I advanced through the network layers, the feature maps gradually decreased in size until they were 7x7 just before the Flatten layer. The input photos are initially 150x150 in size. Although this choice of input size is somewhat random, it is suitable for the given problem.

Preprocessing:

Image File Processing**:**

- Read the image files.
- Decode the JPEG content, converting it into grids of RGB pixels.
- Transform these grids into floating-point tensors.
- Normalize the pixel values, scaling them from the original [0, 255] range to the [0, 1] interval. This is done to ensure that the input values are within a suitable range for neural networks.

Data Augmentation:

To enhance the model's accuracy, data augmentation techniques can be used. Data augmentation is a method that allows to achieve reliable results even when working with limited datasets. It involves generating additional data from the available training samples by applying random transformations. This approach ensures that the model is exposed to a wide variety of images during training, enhancing its ability to generalize effectively.

For this specific task, I intended to randomly apply transformations such as rotating, flipping and zooming to the training dataset's images. This approach generates variations of the existing images, increasing the dataset's diversity and enhancing the model's robustness.

Pre-trained model:

If the original dataset is both extensive and varied, a pre-trained network can be utilized as a generic model, and its features can be applied to many different computer vision tasks. This capability to transfer learned features across different tasks is one of the key strengths of deep learning when compared to other machine learning techniques.

As an example, a large convolutional neural network that has been trained on the ImageNet dataset, which contains 1.4 million labeled images and 1,000 different classes can be considered. This dataset includes several animal classes, such as various breeds of cats and dogs. The architecture of this network is called VGG16, which is a simple and widely used convnet architecture for ImageNet. There are two primary methods for utilizing a pre-trained network: fine-tuning and feature extraction. In this, I used feature extraction, first without data augmentation, and then with data augmentation to achieve better results.

Results:

The table below shows the accuracy and validation loss for each approach.

TABLE FOR MODEL CREATED FROM SCRATCH

TRAIN SIZE	TEST SIZE	VALIDATION SIZE	DATA AUGMENTATION	TRAIN ACCURACY (%)	VALIDATION ACCURACY (%)
1000	500	500	NO	77.2	67.7
1000	500	500	YES	68.5	71.8
1500	500	500	NO	83.7	70.9
1500	500	500	YES	53.7	57.9

1500	1000	500	YES	83.2	75.3
1500	1000	500	NO	68.03	69.4

TABLE FOR PRE-TRAINED MODEL

Data Augmentation	Train Accuracy (%)	Validation Accuracy (%)
NO	99.7	97.5
YES	95.9	97.3

The sample sizes for the train, test, and validation sets are shown in the tables above along with the model configurations. Results with and without data augmentation, for models trained with an increase in train size, or with various train and validation sizes, are included for the model created entirely. For the pre-trained model, I compared the accuracy and validation accuracy with and without data augmentation.

The results show that the models that were regularly trained with data augmentation were unable to perform better than those that were trained without it. The accuracy of the model is additionally enhanced by expanding the training set or modifying the size of the validation set. When I compared the pre-trained model with and without data augmentation, I found that neither the accuracy nor the validation accuracy of the model was improved. Pre-trained models typically perform better overall than models that are created from scratch, especially when handling limited training data.