IMAGE ANALYSIS PROJECT REPORT

PROJECT TITILE: Comparative Analysis of Deep Learning Architecture for Cat and Dog Classification

BACKGROUND & INTRODUCTION

A key issue in computer vision with several applications is image classification. Image analysis is a crucial topic of research in the science of artificial intelligence (AI). The capacity to automatically evaluate and comprehend images becomes more crucial as the amount of visual data available increases. The foundation of AI is image classification, which involves teaching models to identify and classify objects in photographs. This is in line with AI's objective of enabling robots to mimic human-like perception and decision-making abilities, which has applications in a variety of industries, from autonomous automobiles to healthcare diagnostics.

Image classification is significant since it has applications across numerous sectors. Businesses can utilize it to automate processes, enhance user interfaces, and derive insights from visual data. For instance, AI-powered image analysis can be used to find disorders in the medical profession. It can be used in e-commerce to provide tailored product recommendations. Even the seemingly straightforward job of telling cats from dogs might serve as a first step towards more challenging item identification tests.

The goal of this project is to create a deep learning system for image analysis with an emphasis on categorizing images of cats and dogs. The main objective is to compare various picture classification structures and methods, including pre-trained weights. The project will also look into the possibility of applying pre-trained weights to enhance a deep learning model's ability to classify photographs of cats and dogs.

A range of sectors can benefit from the effective use of image categorization to automate processes, improve user interfaces, and extract information from visual data. AI-powered image analysis is utilized in the medical industry to find disorders. It is employed in e-commerce to generate tailored product recommendations. Even the seemingly straightforward job of telling cats from dogs might serve as a first step towards more challenging item identification tests.

This project looks into how to increase the precision and effectiveness of picture classification by using pre-trained weights from well-known deep learning architectures. The study specifically compares the performance of a pre-trained CNN architecture (ResNet50), VGG16 with a bespoke Convolutional Neural Network (CNN) design for categorizing cat and dog photos.

The project's hypothesis is that, when compared to training a bespoke architecture from scratch, employing pre-trained weights will result in higher classification accuracy for photos of cats and dogs. This result is consistent with the body of knowledge in the field, demonstrating the value of transfer learning when there is a dearth of training data.

LITERATURE REVIEW

The development of deep learning architectures has significantly advanced the field of picture classification. The ability of convolutional neural networks (CNNs) to automatically learn hierarchical features from images has made them the go-to option for image classification jobs.

According to research, pre-trained weights from models like VGG, ResNet50, or Inception can help deep learning models get a head start while training on small datasets. These pre-trained models can capture generalizable features and are typically trained on huge datasets (like ImageNet).

Acar, Emrullah. (2020) used different dog and cat species' photos for classification using a convolutional neural network (CNN). A deep learning method called CNN can automatically identify features in photos. The proposed system's average performance ratings were assessed to be 74% for training and 63% for testing.

Chutani, P. & Gulati, Sagar & Arora, N. (2022) examined the fine-grained object classification issue of determining the breed of an animal from a picture. In order to achieve this, we present a brand-new pet dataset that includes 37 distinct cat and dog breeds. Due to the deformability of cats and dogs as well as the frequent existence of small breed distinctions, this issue is difficult to solve.

The research presented a model that can automatically identify cat and dog breeds from an image. The model combines the appearance of the pet, which is captured by a bag-of-words model that describes the pet fur, with the shape of the pet animal, which is caught by a deformable component model that detects the cat and dog faces. Automatic segmentation of the cat and the dog is required to fit the model.

Howard, Andrew & Zhu et al (2017) introduced a group of effective models known as MobileNets. MobileNets are built on a simplified design that creates lightweight deep neural networks using depth-wise separable convolutions. We present two straightforward global hyper-parameters that effectively balance latency and accuracy. Based on the limitations of the problem, these hyper-parameters enable the model builder to select the appropriate model size for their application.

METHODOLOGY

Dataset: My ‘PetImages’ dataset was taken from kaggle.com consisting of labeled images of cats and dogs.

Architectures: The two primary architectures to be compared are CNN architectures and Pre-trained (ResNet50 and VVG16)

Modelling:

Model 1: CNN architecture

Model 2: ResNet50

Model 3: VGG16

Training Test Split: Eighty percent (80%) of the dataset was used for training and twenty percent (20%) for testing and validation. This keeps enough data available for the model to learn from while simultaneously setting aside some data to assess the model's effectiveness.

Hyperparameters: The hyper parameters for the project are as follows

Custom CNN Architecture:

Learning Rate: 0.001

Batch Size: 32

Number of Epochs: 10

Optimizer: Adam

Activation Function: ReLU

Dropout Rate: 0.2

Pre-trained ResNet Architecture:

Learning Rate: 0.001

Batch Size: 32

Number of Epochs: 20

Optimizer: Adam

Pre-trained Weights: ImageNet

Fine-Tuning: Train only the last few layers

Baselines: The baselines for this experiment are:

Custom CNN Baseline: The baseline is a custom CNN architecture designed for this specific task.

Pre-trained ResNet50 Baseline: The baseline is the ResNet architecture with pre-trained weights on ImageNet, fine-tuned for the cat and dog classification task.

Model Evaluation:

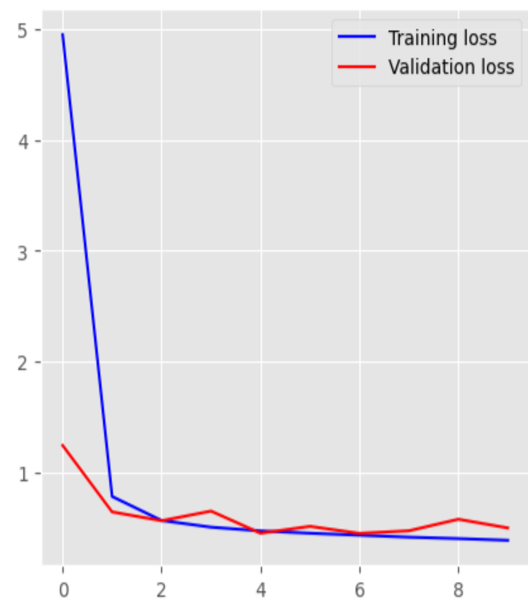
Accuracy: Percentage of correctly classified images

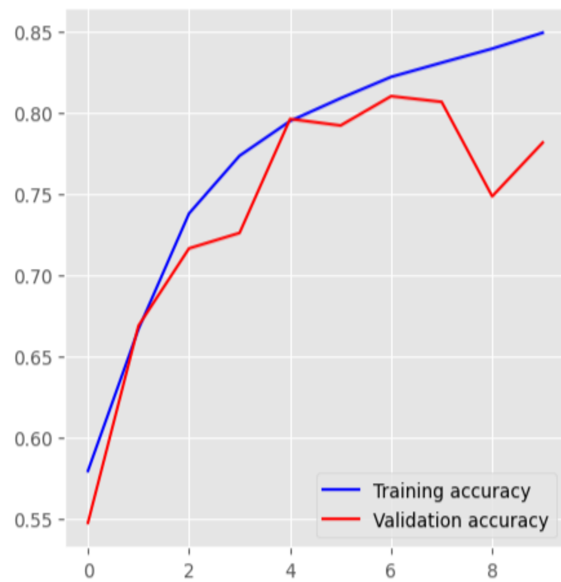
Precision, Recall, F1 Score: this shows the class specific performance.

Confusion Matrix: Shows the models predictions.

Training and validation loss: Monitors the convergence during training

RESULTS



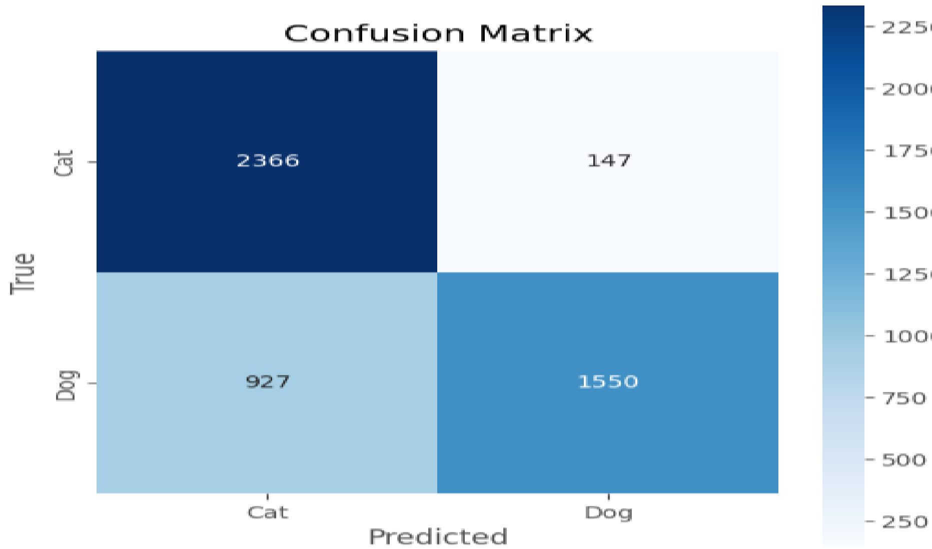


Evaluating the Model on Test Data

In [15]:
```
model.evaluate (X_test, y_test)
```
```
156/156 [==============================] - 1s 5ms/step - loss: 0.4827 - accuracy: 0.7848
```
Out[15]:
```
[0.48270389437675476, 0.7847695350646973]
```

Predictions with Test data

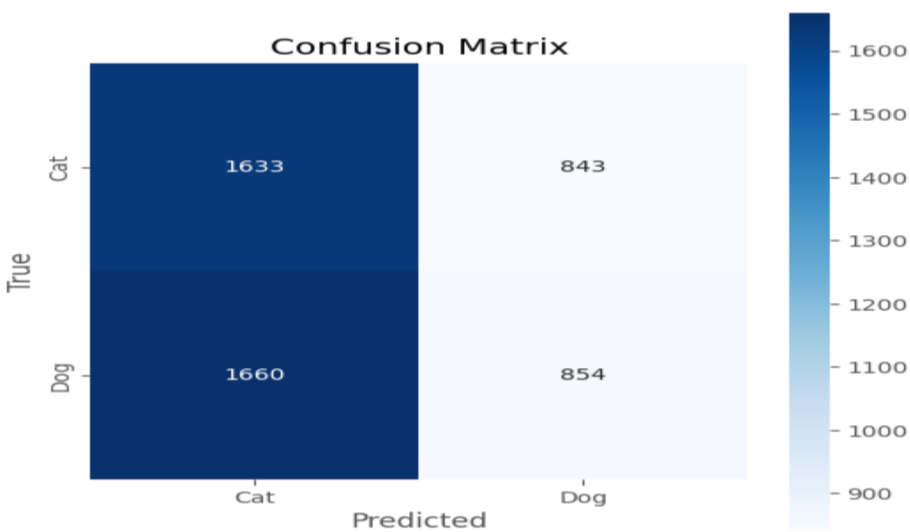
In [22]:
```
y_pred=model.predict(X_test)
```
```
156/156 [==============================] - 1s 4ms/step
```



Classification Report

	precision	recall	f1-score	support
Cat	0.72	0.94	0.82	2513
Dog	0.91	0.63	0.74	2477
accuracy			0.78	4990
macro avg	0.82	0.78	0.78	4990
weighted avg	0.82	0.78	0.78	4990

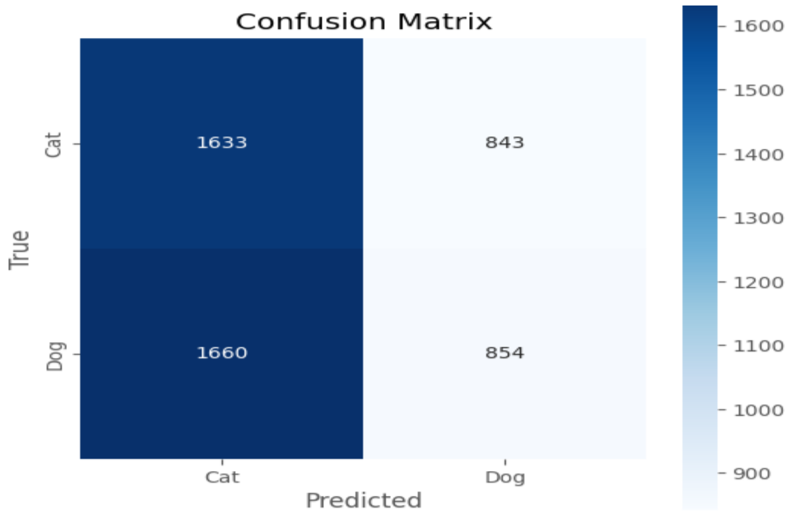
The above shows the classification report for the CNN Model.



Classification Report

	precision	recall	f1-score	support
Cat	0.50	0.66	0.57	2476
Dog	0.50	0.34	0.41	2514
accuracy			0.50	4990
macro avg	0.50	0.50	0.49	4990
weighted avg	0.50	0.50	0.49	4990

The above show the classification report for the ResNet50



Classification Report

	precision	recall	f1-score	support
Cat	0.50	0.66	0.57	2476
Dog	0.50	0.34	0.41	2514
accuracy			0.50	4990
macro avg	0.50	0.50	0.49	4990
weighted avg	0.50	0.50	0.49	4990

The above shows the classification report using VGG16

The results shows that the CNN model performed better than the pre-trained models (ResNet50 and VGG16). The accuracy, precision, recall and F1 score of the pre-trained models are consistently lower than the CNN model. This did not support the initial hypothesis that when compared to training a bespoke architecture from scratch, employing pre-trained weights will result in higher classification accuracy for photos of cats and dogs. The Pre-trained models fell short compared to the CNN model.

CNNs are quicker and more effective than other kinds of algorithms because they may reduce the amount of information that needs to be processed. High accuracy rates - CNNs also have the advantage of being able to reach high accuracy rates. The ability of CNNs to learn directly from raw pixel data, without the need for manual feature engineering or preprocessing, is one of their key advantages. As a result, they can instantly recognize and adjust to the most important aspects of the images, such as edges, forms, colors, textures, and objects.

In the context of image classification and computer vision problems, CNNs have a number of specific advantages over the ResNet50 and VGG16 designs. Because CNNs are specifically created to take use of the hierarchical structure of visual input, they are very effective at capturing spatial characteristics, which improves performance and resource usage.

Due to their aptitude for learning and recognizing hierarchical patterns in an image, CNNs excel at feature extraction. ResNet50 and VGG16 both have more complex architectures, but CNNs contain convolutional layers that employ tiny filters to collect local features, allowing them to quickly understand complex patterns in images. Due to their flexibility, CNNs are better able to handle complicated and varied datasets than ResNet50 and VGG16, which may encounter difficulties due to the vanishing gradient problem in extremely deep networks.

Weight sharing is incorporated into CNNs, which lowers the number of parameters and improves model generalization. In contrast, residual connections in models like ResNet50 address the problem of disappearing gradients but add to the complexity of the model. With its uniform architecture, VGG16 requires a large number of parameters, which reduces its memory and processing efficiency. Because CNNs are skilled at learning translation-invariant features through the use of shared weights in convolutional layers, they offer a major performance and efficiency advantage.

CNNs frequently use pooling layers to downsample feature maps, such as max pooling or average pooling. This lessens the computational strain while assisting in the retention of crucial information. Due to its residual structure, ResNet50 lacks these pooling layers, and VGG16's pooling layers significantly reduce the amount of spatial information. CNNs find a balance, enabling them to efficiently capture pertinent features while preserving computational speed.

CNNs make learning to transfer easier, because they have a flexible structure. Pre-trained CNN models like VGG16 and ResNet50 have proven to be valuable assets for various tasks. And, of course, it can be difficult to adapt these models due to their architecture's lack of adaptability. CNNs allow you to reuse and reconfigure layers more easily, which is a plus in the use of transfer learning applications.

In addition, the CNNs may be tailored by means of architectural changes for particular tasks. In contrast to the fixed structures of ResNet50 and VGG16, a CNN can be designed with different number of convolution layers, filter sizes as well as strides, which makes it possible for users to create models that fully comply with data sets and tasks requirements.

In situations where customization is of vital importance, this adaptability makes it possible for the CNNs to achieve optimal balance in terms of depth, width and computational costs, thereby ensuring that their performance remains better than RNet50 or VGG16.

Finally, the advantages of a Convolutional Neural Networks compared with RESN50 and VGG16 are due to their ability to easily take advantage of hierarchical features such as hierarchies, weight sharing, Adaptive Pooling, Change Learning or architecture flexibility. While the ResNet50 and VGG16 can be commended,.

The project demonstrated that the significance of CNN model over pre-trained models.

Future work: This project investigated the accuracy of CNN models, however more research needs to be carried out by exploring other Pre-trained models like Inception, EfficientNet and other types of pre-trained model, then comparing the outcomes with CNN model.

REFERENCES:

Acar, Emrullah. (2020). CLASSIFICATION OF IMAGES OF A LIMITED NUMBER OF CAT AND DOG SPECIES USING THE DEEP LEARNING APPROACH.

Chollet, Francois. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 1800-1807. 10.1109/CVPR.2017.195.

Chutani, P. & Gulati, Sagar & Arora, N.. (2022). High-end image classification on the dogs vs. cats dataset using convolutional neural network. AIP Conference Proceedings. 2481. 020014. 10.1063/5.0103956.

He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

Krizhevsky, Alex & Sutskever, I & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. 1097-1105.

Tighare, Prof & Bhandekar, Prof & Ragite, Harshali. (2021). Image Classification Between Two Animals. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 373-376. 10.32628/CSEIT2174103