Comparative Analysis of Deep Learning Architectures for Image Classification: VGG vs. ResNet101

Parnian Taheri[a]

[a]*Learning in Brain and Machine,*

Abstract

In this study, we compare the efficacy of two popular feature extraction techniques, namely VGG and ResNet101, for image classification tasks. Leveraging pre-trained deep learning models, we extract high-level features from images of dogs and cats and evaluate their performance using a Support Vector Machine (SVM) classifier. Our dataset consists of 1000 images of each class for training and 500 images for testing. By harnessing the discriminative power of VGG and ResNet101, we aim to develop a robust image classification system capable of accurately distinguishing between dogs and cats.

1. Introduction

Image classification is a fundamental task in computer vision with applications ranging from object recognition to medical image analysis. Deep learning models, with their ability to learn hierarchical representations directly from raw data, have emerged as powerful tools for feature extraction in image classification tasks. Among these models, VGG and ResNet101 are widely used for their effectiveness in capturing intricate patterns and features from images.

In this study, we investigate the performance of VGG and ResNet101 as feature extractors for classifying images of dogs and cats. The choice of dogs and cats as target classes allows us to evaluate the models' ability to distinguish between visually similar categories, thereby providing insights into their discriminative power.

Our approach involves leveraging pre-trained VGG and ResNet101 models, which have been trained on large-scale image datasets such as ImageNet. These models have learned rich feature representations through millions of iterations of supervised learning, making them well-suited for transfer learning to new tasks.

Using our dataset comprising 1000 images of each class for training and 500 images for testing, we extract high-level features from the images using both VGG and ResNet101. These features are then fed into a Support Vector Machine (SVM) classifier, which learns to differentiate between dogs and cats based on the extracted features.

2. Dataset

Our Dataset consists of 3000 images of cats and dogs, 1000 images of each for training and 500 for testing. These images are of all varieties, with both low and high quality. Moreover, cat images and dog images are assumed to be 1 and 0 respectively.

3. Feature Extraction

The feature extraction process plays a crucial role in image classification tasks as it involves capturing meaningful representations from raw input images. In this study, we explore two prominent deep learning architectures, VGG and ResNet101, for feature extraction.

3.1. VGG

VGG (Visual Geometry Group) is a convolutional neural network (CNN) architecture known for its simplicity and effectiveness. It consists of multiple convolutional layers followed by max-pooling layers, with progressively increasing depth. By stacking convolutional layers, VGG can learn hierarchical features of increasing complexity, making it well-suited for a wide range of image classification tasks.

The key characteristic of the VGG architecture is its deep convolutional structure, comprising multiple layers of convolutional and max-pooling operations. VGG variants typically consist of 16 or 19 layers, with small 3x3 convolutional filters and max-pooling layers of 2x2 size. The use of small filters allows the network to capture fine-grained spatial information efficiently, while the max-pooling layers reduce the spatial dimensions of the feature maps, leading to hierarchical feature representations.

One of the notable aspects of VGG is its uniform architecture, where convolutional layers are stacked on top of each other with the same filter size and stride, followed by max-pooling layers. This uniformity simplifies the design and implementation of the network and enables straightforward model configurations.

3.2. ResNet101

ResNet101 is a deep convolutional neural network architecture that has garnered widespread acclaim in the field of computer vision. Developed by Kaiming He et al. in 2015, ResNet101 represents a significant advancement in the design

of deep neural networks, particularly in addressing the challenge of training very deep networks effectively.

The architecture of ResNet101 consists of a series of convolutional layers organized into blocks, with each block containing a set of convolutional, batch normalization, and activation layers. The key innovation of ResNet101 lies in the introduction of residual blocks, where the input to a block is added to its output, allowing the network to learn residual functions instead of directly mapping inputs to outputs. This residual learning approach enables the network to effectively learn the desired mapping, even for very deep architectures.

One of the notable attributes of ResNet101 is its transferability, meaning that the features learned by the network on one dataset can be effectively transferred and fine-tuned for related tasks or domains. This transfer learning capability makes ResNet101 particularly valuable for researchers and practitioners working on image-related tasks, as it allows them to leverage pre-trained models and adapt them to their specific applications with relatively little effort.

4. Classification

In the classification stage, we employ a Support Vector Machine (SVM) to learn a decision boundary based on the high-level features extracted by the ResNet101 model. SVM is a powerful machine learning algorithm known for its ability to effectively separate data points belonging to different classes by finding the optimal hyperplane that maximally separates them in the feature space.

4.1. Support Vector Machine

Support Vector Machines (SVMs) are a powerful class of supervised learning algorithms used for classification, regression, and outlier detection tasks. Developed by Vladimir Vapnik and his colleagues in the 1990s, SVMs have become widely popular due to their ability to handle high-dimensional data and their effectiveness in finding complex decision boundaries.

At the heart of SVMs is the concept of finding the hyperplane that best separates data points belonging to different classes in the feature space. In the case of binary classification, this hyperplane aims to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors. By maximizing the margin, SVMs aim to find the decision boundary with the greatest separation between classes, leading to better generalization performance on unseen data.

One of the key advantages of SVMs is their ability to handle datasets with many features (i.e., high-dimensional data) without overfitting. SVMs achieve this by focusing only on the support vectors, which are the data points closest to the decision boundary and crucial for defining the decision boundary. This sparsity property makes SVMs memory-efficient and computationally efficient, especially for large-scale datasets.

5. Results

In the following, the result of each is attached.

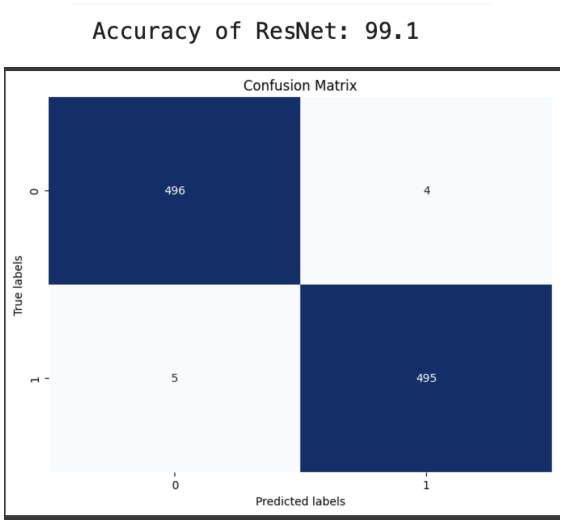

Figure 1: Accuracy & Confusion Matrix of ResNet101

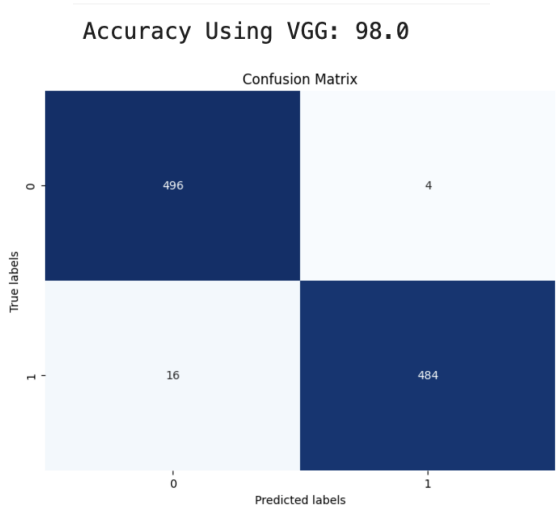

Figure 2: Accuracy & Confusion Matrix of VGG

As is shown in the results, the accuracy of ResNet is better for 1.1 %. Moreover, most of the errors for both methods are for False Negative Rate, this means the number of dogs, which are assumed to be cats are more. However, the running time for ResNet was about 6 hours unlike the VGG, which was about an hour.

6. Concluaion

Overally speaking, this project was to compare the results of two commonly used Deep Learning method, ResNet101 and

VGG on dataset consists of images of dogs and cats. According to the results we can say that the accuracy and performance of ResNet architecture is quite better than VGG, however, due to the really long time that was needed for its training, It's not worth using it.