Comparative Analysis of Transfer Learning in Dog Breed Identification

R Sahithi Reddy
Artificial Intelligence and Machine Learning
R N S Institute of Technology
Bengaluru, India
sahithireddyreddim@gmail.com

Nishitha R Lokande
Artificial Intelligence and Machine Learning
R N S Institute of Technology
Bengaluru, India
lokandenish3@gmail.com

Navya M
Artificial Intelligence and Machine Learning
R N S Institute of Technology
Bengaluru, India
navyam352@gmail.com

***Abstract*—This paper presents a comparative study of transfer learning techniques in the context of dog breed identification. Two distinct datasets are employed to assess model performance, with and without the integration of a center loss function. Our approach focuses on transfer learning and explores data augmentation techniques to expand dataset size, enhancing model robustness. This research provides valuable insights into the effectiveness of transfer learning strategies across diverse datasets, offering a comprehensive understanding of their applicability in the domain of dog breed identification.**

Keywords—Dog Breed Identification, Transfer Learning, Center Loss Function

I. Introduction

In the era of Artificial intelligence and deep learning, image classification tasks have revolutionized as well as witnessed substantial advancements over the course of time. One such intriguing challenge is the identification of dog breeds from images, a task that not only fascinates dog enthusiasts but also serves as a benchmark problem in computer vision. This research paper delves into the domain of dog breed identification, leveraging the power of transfer learning with a focus on the InceptionNet V3 model.

Our investigation goes beyond the mere application of a single model. We introduce a comparative analysis by testing the InceptionNet V3 model with and without the incorporation of a center loss function. To ensure the comprehensiveness of our study, we employ two distinct datasets: the Kaggle dataset[1] and the Stanford dataset[2]. These datasets represent different sources, and by comparing their accuracy, we aim to gain deeper insights into the model's performance across diverse data domains.

The remainder of this document is structured as follows: Section II provides an in-depth exploration of the datasets and preprocessing techniques, while Section III elaborates on the InceptionNet V3 model. Section IV delves into our experimental setup, detailing the performance of the model with and without the center loss function on the two datasets. Section V showcases the results and their implications. Section VI presents conclusions drawn from our research and outlines future directions. Finally, Section VII lists the references that have guided our study.

II. Dataset and Preprocessing

The foundation of our research lies in the utilization of two essential datasets: the Stanford Dogs dataset[2] and the Kaggle dataset[1]. Comprising 120 distinct dog breeds, these datasets are a subset of the expansive ImageNet[3] collection. To facilitate our model's training and evaluation, we partitioned each dataset into dedicated training and validation subsets. Our preprocessing pipeline involves data augmentation techniques to enhance dataset diversity and promote model generalization. Furthermore, images from the Kaggle dataset[1] are resized to 224x224 pixels, while those from the Stanford dataset[2] are resized to 299x299 pixels, ensuring uniformity in input dimensions for our Inception V3 model.

III. Inception V3

Inception V3, a convolutional neural network, is part of Google's Inception architecture series, designed for image classification and object recognition. Introduced in 2015, it features innovative "Inception modules" that use parallel convolutions of varying sizes to capture a wide range of image features efficiently. Filter sizes in these modules typically include 1x1, 3x3, and 5x5, allowing the network to simultaneously process fine and coarse-grained features. Inception V3 is renowned for its high accuracy and computational efficiency, incorporating techniques like batch normalization and factorized convolutions to reduce model size and enhance training speed. It's widely utilized in various computer vision applications.

IV. Experiments

After applying data augmentation and dividing the datasets into training and validation subsets, our experimentation focuses on fine-tuning the model with critical hyperparameters. This includes adjusting attention parameters, optimizing weight decay, and implementing dropout mechanisms for enhanced model performance.

A. Data Augmentation

To mitigate the risk of overfitting in our dog breed identification research, we employ a prevalent strategy: data augmentation. This process involves several key transformations on the dataset. Firstly, we rescale the images to a 1.0/255.0 scale, standardizing their intensity. Secondly, we introduce a 20-degree rotation to diversify the dataset further. Additionally, we apply width and height shift ranges of 0.2, enabling variations in image positioning. To augment the dataset's variability, we horizontally flip images. Finally, to ensure robust model evaluation, we allocate 20% of the dataset for validation, allowing us to assess model performance effectively. Data augmentation plays a pivotal role in enhancing model generalization and reducing overfitting, thereby contributing to the accuracy of our dog breed identification system.

B. Learning and Hyperparameters

In our endeavor to construct an effective dog breed identification system through transfer learning, we recognize the paramount significance of meticulous model parameterization. Commencing this process, we embark on the delicate task of configuring various parameters that are pivotal in determining the model's performance. One such parameter is weight decay, meticulously tuned to 0.0001. This serves as a regulatory mechanism, crucial for preventing the model from over-adapting to the training data, a condition known as overfitting, which could result in suboptimal performance on new, unseen data. In a bid to enhance the model's adaptability and its ability to generalize patterns, we apply a dropout rate of 0.3. During training, this entails randomly deactivating certain connections within the model, promoting its robustness.

Our approach to handling the datasets involves tailored strategies. For the Stanford dataset[2], we prescribe a center loss value of 0.5, whereas for the Kaggle dataset[1], the center loss is computed differently, yielding a value of 0.0068. Our training regimen commences with an initial weight decay and learning rate set at 1e-3. Training unfolds over three epochs, striking a delicate equilibrium between achieving high accuracy and optimizing computational resource utilization.

To fine-tune the model for each dataset, we unfreeze and adjust specific layers, tailoring their configurations to harmonize with the dataset characteristics. We opt for the ReLU optimizer, and the Inception V3 model forms the backbone for both datasets. Initially, we abstain from integrating attention mechanisms, although we retain the flexibility to explore their potential benefits in future iterations. These judiciously calibrated parameters collectively underpin our objective of optimizing model performance while mitigating the risk of overfitting.

V. Results

The evaluation of our dog breed identification model using transfer learning primarily relies on the accuracy metric, which represents the mean percentage of correctly classified classes within a given dataset. This section presents the results of our experiments on the Kaggle and Stanford datasets[1-2], comparing the performance of our model with and without the center loss function.

For the Kaggle dataset[1], the training accuracy without the center loss function was observed to be 0.6949, accompanied by a loss of 1.3426. In contrast, the validation accuracy for the same dataset reached 0.7881, with a corresponding validation loss of 0.8932. However, when we introduced the center loss function into the training process, we noticed a performance improvement. The training accuracy increased to 0.7501, with a reduced training loss of 0.8909. The validation accuracy further improved to 0.8180, accompanied by a reduced validation loss of 0.6237. These results indicate that the inclusion of the center loss function enhanced the model's ability to correctly classify dog breeds in the Kaggle dataset[1].

Turning to the Stanford dataset[2], our model achieved a sparse categorical accuracy of 0.6181 without the center loss function, with a corresponding loss of 1.7206. However, with the center loss function incorporated into the training process, the training accuracy substantially improved to 0.7252, with a reduced training loss of 1.2065. The validation accuracy for the Stanford dataset[2] also demonstrated improvement, reaching 0.8089, while the validation loss decreased to 0.8474.

Overall, the results suggest that our model, based on Inception V3 architecture and equipped with hyperparameters such as weight decay, dropout, and the center loss function, performed admirably in identifying dog breeds, particularly excelling with the Kaggle dataset[1]. These findings underscore the effectiveness of transfer learning and the potential for further optimization in breed recognition tasks.

TABLE I. Accuracy

Experiment Type	**Comparative Dataset Accuracy**		
	Validation Accuracy	***Kaggle dataset[1]***	***Stanford dataset[2]***
Results with center loss	Epoch 9 Accuracy	81.80%	80.89%
Results without center loss	Epoch 3 Accuracy	78.81%	61.81%

VI. Conclusion and Future Work

In this study, we employed the Inception V3 model for dog breed identification, evaluating its performance on both Kaggle and Stanford datasets[1-2]. Our focus centered on the efficacy of the center loss function. While the Kaggle dataset[1] demonstrated slightly superior accuracy compared to Stanford, the adoption of the center loss function showcased its prowess in mitigating overfitting. This was evident in the reduction of loss and concurrent improvement in model accuracy.

For future investigations, we propose exploring alternative optimizers such as Adam or Adafactor and fine-tuning hyperparameters like weight decay and dropout rates. Additionally, the incorporation of attention parameters holds promise for further enhancing the model's capabilities.

References

[1] Kaggle, "Dog Breed Identification," Kaggle, [Online]. Available: https://www.kaggle.com/c/dog-breed-identification.
[Accessed: August 14, 2023].

[2] Stanford University, “Stanford Dogs Dataset,” [Online]. Available: http://vision.stanford.edu/aditya86/ImageNetDogs/. [Accessed: August 14, 2023].

[3] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255.

[4] Y. Patel, “Dog-Breed-Prediction” GitHub, Available: https://github.com/yashpatel-py/Dog-Breed-Prediction, [Accessed: August 14, 2023].

[5] M. Tiwari, "Dog_Breed_Classifier_CNN," GitHub, [Online]. Available:https://github.com/Tiwarim386/Dog_Breed_Classifier_CNN. [Accessed: August 14, 2023].

[6] J. A. Nievarghese, "Dog Breed Identification: Keras NASNetLarge (93.63%)," Kaggle, [Online]. Available:https://www.kaggle.com/code/jithinanievarghese/dog-breed-identification-keras-nasnetlarge-93-63#Data-preprocessing. [Accessed: August 14, 2023].

[7] A. Nesterov, "Dog Breed Identification: ViT CNNs - Loss 0.12," Kaggle,[Online].Available:https://www.kaggle.com/code/andreynesterov/dog-breed-identification-vit-cnns-loss-0-12. [Accessed: August 14, 2023].