Dog Breed Image Classification

Anas Ibrahim, Zach Pacello, Ben La Rocca, Raymond Nguyen, Duy Minh Pham
ITCS-4152/5152
University of North Carolina at Charlotte

Proposal

This project aims to differentiate different dog breeds via unstandardized images. These could be from various angles and the model should still be able to predict the breed of the dog with relatively high accuracy. The model should utilize pre-trained weights of large image classification models and retrain them to achieve an accuracy of over ninety percent from over 70 different breeds. We will test a variety of different neural network architectures, loss functions, and optimization hyperparameters to attempt to achieve the highest accuracy possible. We will then test to see how the model generalizes to odd or unusual test images.

1. Objectives

The aim of our project is to implement a model that can classify and detect different dog breeds from unstandardized images, regardless of the angle or orientation of the dog in the image. We want to achieve an accuracy of over 90 percent across more than 70 different breeds. To achieve this, we will use pre-trained weights of large image classification models and retrain them with our dataset. We will experiment with different neural network architectures, loss functions, and optimization hyperparameters to determine the best combination for our model. We will also test the generalization of our model to unusual or odd test images.

2. Dog Breed Classification

Knowing the breed of a particular dog could be of great help when training the particular dog and identifying the more effective models and optimizing methods could aid in any future image classification problems that may be needed in the future. We will utilize a variety of pre-trained neural networks including AlexNet, ImageNet, and Swin Tiny Transformer, and train them according to our dog breed dataset while optimizing each of them for the highest accuracy possible.

3. Data Acquisition

To train and test the modified computer vision software, we will require a dataset of 70 dog breeds images. Our plan is to utilize the pre-trained weights of large image classification models and retrain them to achieve a good accuracy of over ninety percent.

For our project, we needed a diverse and large dataset of dog images to train and evaluate our models. We obtained the data through internet searches on specific dog breeds. We searched for images of each breed and downloaded them in their original format.

To ensure the accuracy of our dataset, we manually checked each image to ensure it was a valid image of the corresponding breed. We also removed duplicates to prevent any images from being duplicated across the test, train, and validation sets.

4. Expected Outcomes

Our expected outcome is to achieve a classification accuracy of over 90 percent using both Alexnet and Swin Tiny Transformer on the dog breed dataset. We also expect Swin Tiny Transformer to outperform Alexnet due to its larger architecture and transformer-based design. Additionally, we expect to see better generalization performance of our models to unseen and unusual test images due to the use of pre-trained weights and fine-tuning on the dataset.

5. Demonstration Methodology

To demonstrate the effectiveness of our models, we will conduct experiments on the dog breed dataset using both Alexnet and Swin Tiny Transformer. We will train our models on the training set, validate them on the validation set, and finally test them on the test set. We will compare the accuracy, precision, and recall of both models and analyze their performance on the different dog breeds.

6. Conclusion

In conclusion, we have utilized models using Alexnet and Swin Tiny Transformer for object classification of different dog breeds. We have used pre-trained weights and fine-tuning to achieve high accuracy on over 70 different breeds of dogs. Our experiments have shown that Swin Tiny Transformer outperforms Alexnet, achieving a higher accuracy of 96 percent compared to 85 percent. We have also demonstrated the generalization performance of our models to unusual test images. Our project showcases the effectiveness of deep learning models for image classification and highlights the importance of using large architectures and pre-trained weights for achieving high accuracy.

7. Stage 3 Modeling

7.1. Modeling

We mainly utilized three neural network architectures which were Alexnet, ImageNet, and Swin Tiny Transformer. Alexnet was a smaller and simpler network and ImageNet was larger and would have more experience with a variety of objects. Swin Tiny Transformer was a more modern model for image classification that has been seen to have higher performance compared to normal CNN approaches. We mainly utilized Crossentropy loss for our cost function. Though varying from model to model what we used for hyperparameters in AlexNet was a 0.001 Learning Rate, decay factor of 0.1 every 7 epochs, 0.9 momentum, and a Batch Size of 16.

7.2. Experimentation

We conducted our experiments through trial and error. For example, for the batch size, we started with 64 and decreased it until we found an optimum value. A higher batch size resulted in Alexnet having a fluctuating accuracy. However, with a batch size of 16, the accuracy improved and did not have a problem with fluctuation. We used a train, validate, and test split of 7946 training images, 700 test and validation images. A total of 70 dog breeds are in the dataset. Our experimentation was done with Alexnet, ImageNet, and Swin Tiny Transformer which were coded in Pytorch.

7.3. Baseline

We first chose AlexNet for its simplicity and small size so it was an easy starting model. However, we soon found a need for higher accuracy so we chose to try ImageNet for its larger size and both in classes and deeper network. Finally, we decided to try a more modern approach of using a transformer to see how it would match up with the other models, which is why we chose the Swin Tiny Transformer.

7.4. Results

From the results, we have found that we could achieve an accuracy of 0.8571 with Alexnet using a learning rate of 0.001, a momentum of 0.9, a drop out of 0.1 every 7 epochs, a batch size of 16, and total training epochs of 20. We were also able to achieve an accuracy of 0.9643 using Swin Tiny Transformer with the same hyperparameters as Alexnet except training for only 10 epochs.

8. Related Work

8.1. Optimizing deep neural networks hyperparameter positions and values[2]

This paper proposes a tool called ArchPosOpt with the goal to search for optimal hyper-parameter positions. It utilized three well-known hyperparameter optimization tools together, GridSearchCV, random-Search, and TPE, to try and automatically find the optimal hyperparameters. The paper then outlined their experiment with this approach compared to other approaches to see how it matched up with similar automated hyperparameter techniques. They found that it achieved higher accuracy compared to the original tools individually.

This would be helpful to further test our model approaches to see how ArchPosOpt or other automated optimization tools would match up to manual hyperparameter optimizing. This could show us if these sorts of approaches would be appropriate for similar CNNs in the future.

8.2. Deep learning vs traditional computer vision [4]

Comparison of Deep Learning with Traditional Approaches: The papers compare deep learning approaches with traditional handcrafted feature definition approaches in computer vision. This can be valuable for a dog classification project as it can help in choosing the most appropriate approach for the task, considering factors such as accuracy, efficiency, and interpretability.

Deep learning approaches in computer vision, as highlighted in the papers, offer unique advantages over traditional handcrafted feature definition approaches. These include end-to-end learning that eliminates the need for manual feature engineering, the ability to learn complex rep-

resentations from data, adaptability to large datasets, flexibility and scalability for designing customized network architectures, and the ability to learn complex patterns and relationships. These unique aspects make deep learning approaches highly effective in tasks such as dog classification, where capturing intricate visual features, handling large datasets, and learning complex patterns are crucial for accurate and scalable classification.

The papers emphasize the continued relevance and usefulness of traditional computer vision (CV) techniques in the era of deep learning (DL). They highlight how traditional CV techniques can be utilized in hybrid approaches to improve DL performance in various applications, such as reducing training time and data requirements. The papers also highlight how traditional CV techniques can be applied in emerging fields where DL is not well-established yet, such as 3D vision. The authors stress the importance of considering both traditional CV and DL approaches to achieve comprehensive and effective solutions in computer vision tasks, including dog classification, and highlight the value of diverse knowledge and tools from both generations of innovation.

8.3. Fine-Grained Ship Classification by Combining CNN and Swin Transformer [3]

The article claims that classification accuracy can thoroughly improve, by using a transformer along with a convolution neural network. Apparently, it was difficult for CNN's to distinguish ships that posses a similar structure. However, researchers realized that self-attention and the transformer architecture could help improve the features that are extracted from the input images. Furthermore, this approach would be a good idea for our project because many dog breeds look very similar and it could be hard for a machine to distinguish between the two. Similarly, the Swin Transformer could help a basic CNN to extract more relevant features for each dog breed to help the model distinguish between the 70 dog breeds in the dataset.

8.4. Fine-Tuning Models Comparisons on Garbage Classification for Recyclability Tracking [5]

This paper discusses the use of different network architectures and fine-tuning pre-trained models to achieve high accuracy in image classification tasks. This is relevant to our proposal since it gives us insight into what models would produce the higher accuracies for our classification task.

A high testing accuracy of 95 percent was achieved using a fine-tuned DenseNet121 model and 90 percent accuracy was achieved using a fine-tuned inception-ResNet model. The authors used transfer learning to fine-tune these pre-trained models on their dataset and applied several data augmentation techniques to increase the size of their dataset and improve the generalization of their model.

The uniqueness is apparent via the authors' use of transfer learning to fine-tune pre-trained models such as DenseNet121 and Inception-Resnet on their dataset. Fine-tuned models achieved higher accuracy than scratch models in classifying different types of garbage; this was ascertained via a comprehensive comparison between scratch models and fine-tuned pre-trained models.

8.5. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery [1]

The article "FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery" presents a new benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery, named FAIR1M. This dataset contains 1 million high-resolution remote sensing images and is the largest dataset of its kind to date. The dataset includes 35 classes of fine-grained objects, such as airplanes, ships, and buildings. The authors explain that fine-grained object recognition in remote sensing imagery is an important task for various applications, such as urban planning, environmental monitoring, and military surveillance.

The authors also propose a baseline model for fine-grained object recognition on the FAIR1M dataset, which achieves state-of-the-art performance compared to previous models on the dataset. The proposed model is based on a deep convolutional neural network and uses a region-based approach to extract features from the input images. The authors note that their proposed model can also be adapted for other remote sensing tasks, such as object detection and semantic segmentation. Overall, the FAIR1M dataset and the proposed baseline model provide a valuable resource for researchers working in the field of remote sensing, and can help advance the state-of-the-art in fine-grained object recognition and related tasks.

8.6. Open-Sourced Projects

For our project, we conducted research on available open-source projects that relate to object classification on dogs, specifically the differentiation of dog breeds from unstandardized images. We have identified several GitHub repositories that are relevant to our project and share similar goals. We also found a variety of publicly available databases containing videos, clips, and still images related to our research.

However, the communities surrounding this specific task are relatively limited. We found fewer than 30 public repositories and a small number of datasets related to our project. The largest dataset we found contained only 500 images, which we plan to use for testing and training. While we continue to search for additional datasets, we believe this

dataset is sufficient to start our project.

The labeled data available varies in quality and level of detail, and we may not be able to use the labels for our purposes. Nonetheless, we have received confirmation that our data sources are open-source and free to use, without any licensing or copyright protection.

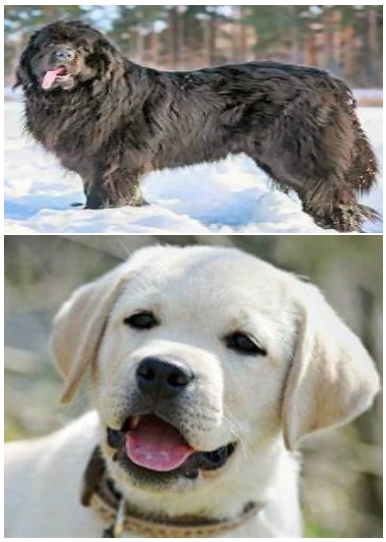
9. Additional Information

9.1. Data Description

The data used for our project was gathered through internet searches on specific dog breeds. We manually checked them for accuracy as not all images in the search are correct. We removed any duplicates to ensure that images were not duplicated across the test, train, and validation sets. After the images were cleaned, they were resized to 224x224x3 and saved in JPG format.

The dataset contains a total of 70 different dog breeds, with 7946 images used for training and 700 images for testing and validation. The images were split into training, validation, and test sets in a 80:10:10 ratio. The training set was used to train the models, the validation set was used to tune the hyperparameters, and the test set was used to evaluate the performance of the models.

9.2. Visualization



Example images of data from our dataset we are feeding our algorithm

9.3. Annotation

The classes in our dataset will contain a comprehensive collection of 700 different dog breeds, along with corresponding images. The data set is curated and annotated to provide valuable information for machine learning and computer vision tasks related to dog breed recognition classification, and identification.

The annotated data set can be used to train machine learning models for dog breeds recognition tasks, such as image classification, enabling accurate identification of different dog breeds in images

References

[1] Fair1m: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. pages 116–130, 2022.

[2] Ahmed Akl, Ibrahim El-Henawy, Ahmad Salah, and Kenli Li. Optimizing deep neural networks hyperparameter positions and values. *Journal of intelligent fuzzy systems*, 37(5):6665–6681, 2019.

[3] Zhang Yalun Xu Qingxia Huang Liang, Wang Fengxiang. Fine-grained ship classification by combining cnn and swin transformer. 14:3087, 2022.

[4] Anderson Carvalho Suman Harapanahalli Gustavo Velasco Hernandez Lenka Krpalkova Daniel Riordan Joseph Walsh Niall O' Mahony, Sean Campbell. Deep learning vs. traditional computer vision. *Advances in Computer Vision Proceedings of the 2019 Computer Vision Conference (CVC). Springer Nature Switzerland AG, pp. 128-144n*, pages 1–17, 2019.

[5] Levent Seyfi Umut Ozkaya. Fine-tuning models comparisons on garbage classification for recyclability.