Research Proposal of Dog Breed Identification using Deep Learning

Karolayne Isabel Gaona Chong
Earlham College
Richmond, Indiana, USA
kigaona20@earlham.edu

ABSTRACT

Pet identification is important for veterinary care, pet ownership, and animal welfare and control. This proposal presents a solution for identifying dog breeds using dog images. The proposed method applies a deep learning approach to identify the dog breeds. The starting point for this method is transfer learning by retraining existing pre-trained convolutional neural networks on the Stanford Dog database. Three classification architectures will be used. These classifiers will take images as input and generate feature matrices based on their architecture. The stages these classifiers will undergo to create feature vectors are 1) Convolution to generate feature maps and 2) Max Pooling: highlight features are extracted from the feature maps. Data augmentation is applied to the database to improve classification performance.

ACM Reference Format:
Karolayne Isabel Gaona Chong. 2018. Research Proposal of Dog Breed Identification using Deep Learning. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The field of Machine Learning in Computer Vision is growing. Technological developments have impacted the field to gain increased momentum [8]. Machine learning improves system performances by learning from experience via computational methods [14]. It allows data analysis to be performed by a computer in place of humans [8]. The main task of machine learning is to develop learning algorithms that build models from data by feeding the learning algorithm with experience data that can make predictions on new observations [14]. There are multiple applications for Machine Learning, and classification problems are one of them [6]. The process that the system uses to learn from data is iterative. This means as the model continues to be exposed to more data, it will learn from the new data [8]. The motivation for this paper is to learn how to use classification tools to identify dogs from different breeds through digital photographs. Currently, human face identification is successfully used for authentication and security in many applications. There are multiple efforts in researching the animal recognition field [2]. Dogs are known for being man's best friend. There are over 180 dog breeds. According to Borwarnginn et al., dog breed recognition can be important in providing proper training and health training. Usually, dog breed recognition is done by humans; however, some dog breeds may be difficult to recognize due to their physical similarities. The current papers use the methodology and results of fine-tuning Convolutional Neural Networks (CNN) for two different architectures using the Stanford Dog Database (see Figure 2.) [10]. CNNs are becoming more popular for image classification [8]. CNN is a feed-forward Artificial Neural Networks and deep learning class, and it can take images as input. CNNs have at least one fully connected layer followed by the desired number of fully convolutional layers as a standard multilayer network [1].

Computer vision is a field of Artificial Intelligence that uses computers to get information from images, videos, etc. Some technologies used in Computer Vision include image processing, Machine Learning, and pattern recognition. Deep learning, a subset of Machine Learning, can create different features during training from the original images [2]. In this project, I am aiming to use a fine-tuned CNN, specifically ResNet, NASNet, and InceptionV3, on the Stanford Dogs Database to enhance the accuracy and efficiency of dog breed identification from digital photographs and evaluate the incorporation of flipped images from the same database as data augmentation technique that may contribute to the robustness of the model.

2 BACKGROUND AND RELATED WORK

Several traditional techniques were used before machine learning in image identification problems became more popular. These methods, which relied on handcrafted features and heuristics to recognize patterns in images, Feature Extraction, Image Segmentation, Handcrafted Neural Networks, etc. The research of Borwarnginn et al. addresses some of the challenges of some of the traditional approaches that attempted to solve the problem of dog breed classification using methods such as Coarse to fine classification [4][2]. However, for image classification problems, it seems like machine learning outperforms traditional techniques.

3 METHODOLOGY

This project attempts to be a deep-learning architecture to improve dog breed identification using the Stanford Dog Database with 20,580 images. This database contains 120 different dog breeds and is a subset of ImageNet. It is separated into a training data set with 12,000 images and a test database with 8,580 images. Both sets have images of different sizes, and every image has a label representing the dog's breed [10]. Using a similar approach as Borwarnginn et

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

al., the project will require three main phases: data preparation, training, and testing. The research of Rhodes and Borwarngin et al. decided to identify the dogs' faces rather than a broader picture frame, excluding the dogs' bodies and backgrounds. Borwarnginn et al. focused on dog face recognition, using a CNN-based model for dog breed identification and applying transfer learning and data augmentation techniques. Additionally, Raduly et al. proposed a system that takes as input an image of a dog provided by the user and returns information about the dog's breed. I am attempting to use the whole image from the raw data with the dog's images. The data will be divided into two categories for the training and testing process [2].

- *Data Preparation*: The Stanford Dog Database is one of the public databases for dog breed classification. According to [5], images of 120 breeds of dogs worldwide. This dataset has been built using images and annotation from ImageNet for the task of fine-grained image categorization. My goal is to use the images raw from the database. Half of the training set will be augmented using flipping. Horizontal axis flipping is much more common than flipping the vertical axis [11]. TensorFlow will be used to do the augmentation work.

- *The dog breed identification model*: This model will undergo training on an extensive database comprising dog images to get the ability to identify various dog breeds. The initial phase of training involves the understanding of traditional attributes [2]. As the models go deeper into the database, they attain a holistic understanding involving the identification of entire objects, such as dogs, and parts of them such as tails or ears, and the relationships between these elements. A natural characteristic of these models lies in their autonomous learning capability, requiring no explicit instructions for feature identification. The necessity for manual specification of these parts is obviated, as the model autonomously understands relevant features through a process denoted as feature learning. This methodology emulates the model's training to adeptly recognize dog breeds, alleviating the need for exhaustive manual intervention.

- *Model Training*: a CNN undergoes a crucial iterative process to learn and associate features from a vast labeled database of dog images with their respective breeds. Each dog image serves as input to the model, and the corresponding breed label acts as the ground truth against which the model refines its predictions. The training process involves adjusting the internal parameters of the model through optimization techniques. This process occurs over multiple epochs, with each epoch representing a complete iteration through the entire training database. The training database is typically processed in smaller batches, and the model's performance is periodically evaluated on a separate validation set to ensure effective generalization. Metrics such as accuracy guide the assessment of the model's learning progress. Optional practices, including early stopping and fine-tuning, may be employed to prevent over-fitting and enhance overall model performance. The success of model training is fundamental to the subsequent testing and deployment phases, determining the model's ability to accurately classify dog breeds in unseen data.

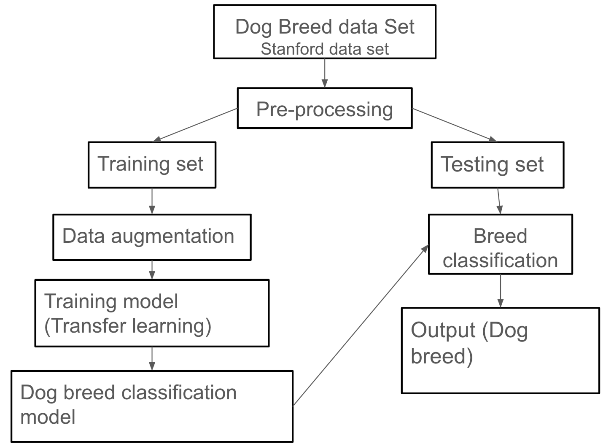

Figure 1: Overview of the proposed framework. This figure is adapted from the graph in [2].

3.1 Transfer Learning

According to Raduly et al., transfer learning helps improve the performance of models without starting from scratch. Instead of training models completely new, it uses ones that have already learned common features, often from the ImageNet database. To make these models more useful for specific tasks, researchers fine-tune their knowledge on a focused dataset. This way, they keep the general knowledge the models gained earlier but tailor it to be really good at a particular job. This approach not only makes learning more efficient but also lets models quickly adapt to new tasks, making transfer learning a useful way to boost performance. The learning of these models means fine-tuning the training data with learned weights and biases. This research will be composed of three public CNNs that will perform feature extraction [1].

- *ResNet*: Residual Neural Network is a deep learning model in which the weight layers learn residual functions with reference to the layer inputs. A Residual Network[1] is a network with skip connections that perform identity mappings, merged with the layer outputs by addition [3].
- *NASNet*: is an architectural building block on a small database and then transfer the block to a larger database. The NASNet model achieves state-of-the-art results with smaller model size and lower complexity (FLOPs) [13].
- *InceptionV3*: is a convolutional neural network for working with image analysis and object detection. Inception helps with classification of objects [12].

3.2 Feature Extraction

The features are extracted using pre-trained models, which are trained on ImageNet to a fully connected layer [1], and the final layers for identification are fixed.

Figure 2: Example of Stanford dog Public Database

3.3 Data Preparation

Deep learning has outperformed in computer vision tasks. However, it requires a lot of data training to avoid over-fitting [2]. The most common method to avoid over-fitting on training data is to use different transformations before the feed-forward passes during the training [10]. The over-fitting behavior happens when the model gives accurate predictions with training data but not new data. There can be many reasons why this undesired behavior happens, but one of them is when the training data is too small and does not have enough data samples. Although many researchers apply data augmentation, there are many ways in which this problem can be approached[2].

3.4 Data Augmentation

Once the transfer learning phase is done, data augmentation will be applied to the training data set using TensorFlow methods to flip images (see Figure 3.) Data flipping is a technique that directly augments the existing images by performing a geometric transformation of the image. The augmented images keep the same label as the original image. In this project, data augmentation is used to increase the database [2]. According to [9], all the transformations of data augmentation are affine transformations of the original image that take the form:

$$y = Wx + b$$

The data will be transform within some set of parameters. This process will contribute to two things 1) it will increase the data set that is use to work with and it helps ensure that the model can recognize the breed of the dog even if the image is slightly moved [7].

3.5 Evaluation

To evaluate the proposed method of this project, two scenarios will be evaluated: 1)applying data augmentation and 2) not applying data augmentation. Similarly to Sulyok, I may use accuracy as a metric for evaluation and precision to have more results. The accuracy will be captured on the training and test databases. This metric represents the mean percentage of correctly classified classes on a database.

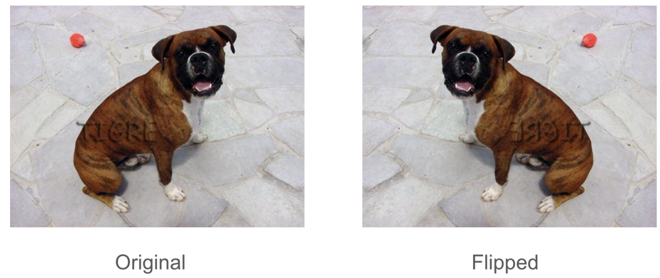

Figure 3: Example of an image from the Stanford Dog Database being flipped horizontally

3.6 Timeline

January.

- **Week 1-2:** Project definition, goal setting and literature and proposal review. Go back to all used literature and read it again and search for their code. Simultaneously, check my database and create a rough architecture model of my implementations.
- **Week 3-4:** Reviewing and database processing. I will start very roughly to do changes to my database if needed to fit my model. As well as starting to set Linux so I can run my project as well as writing the first lines of code.

February.

- **Week 1-2:** Data Reprocessing and putting together models. Keep working on writing code and running program. Implementation of TensorFlow and CNN architectures. Simultaneously, be working on the writing final paper first draft.
- **Week 3-4:** Manipulating, and debugging . Working heavily in solving all the code and implementation issues.

March.

- **Week 1-2:** Model implementation. Working heavily in trying to put all the architecture together.
- **Week 3-4:** Have a decent final draft for final paper and keep debugging.

April.

- **Week 1-2:** Model training and evaluation. Have some results. If needed, fix errors.
- **Week 3-4:** Results analysis and discussion. Generate data visualizations to use on paper and final presentation poster. Present project at Epic Expo at Earlham College.

May.

- **Week 1-2:** Finalize final paper.

4 CONCLUSION

The motivation of this model is to learn how to use a machine learning classification tool in order to classify images and dog breeds. I tried the CNN method to see how well they can predict dog breeds. I am also curious about how these compare with other such as Xception.

REFERENCES

[1] Durga Bhavani, Mir Habeebullah Shah Quadri, and Y Ram Reddy. 2019. Dog Breed Identification Using Convolutional Neural Networks on Android. *CVR Journal of Science and Technology* 17, 1 (2019), 62–66.

[2] Punyanuch Borwarnginn, Worapan Kusakunniran, Sarattha Karnjanapreechakorn, and Kittikhun Thongkanchorn. 2021. Knowing your dog breed: Identifying a dog breed with deep learning. *International Journal of Automation and Computing* 18 (2021), 45–54.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[4] Yuqi Huo, Yao Lu, Yulei Niu, Zhiwu Lu, and Ji-Rong Wen. 2019. Coarse-to-fine grained classification. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1033–1036.

[5] Aditya Khosla. [n. d.]. ImageNet Dogs Dataset. http://vision.stanford.edu/aditya86/ImageNetDogs/. Accessed: <insert date here>.

[6] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* 160, 1 (2007), 3–24.

[7] Aditya Mehta. 2020. *Using Deep Learning to identify dog breeds*. https://towardsdatascience.com/computer-vision-part-2-8e07029955ee Decmeber,2023.

[8] Kaitlyn Mulligan and Pablo Rivas. 2019. Dog breed identification with a neural network over learned representations from the xception cnn architecture. In *21st International conference on artificial intelligence (ICAI 2019)*.

[9] Luis Perez and Jason Wang. 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621* (2017).

[10] Zalán Ráduly, Csaba Sulyok, Zsolt Vadászi, and Attila Zölde. 2018. Dog breed identification using deep learning. In *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, 000271–000276.

[11] Connor Shorten and Taghi M Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of big data* 6, 1 (2019), 1–48.

[12] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[13] Sik-Ho Tsang. 2019. *Review: NASNet — Neural Architecture Search Network (Image Classification)*. https://sh-tsang.medium.com/review-nasnet-neural-architecture-search-network-image-classification-23139ea0425d December,2023.

[14] Zhi-Hua Zhou. 2021. *Machine learning*. Springer Nature.