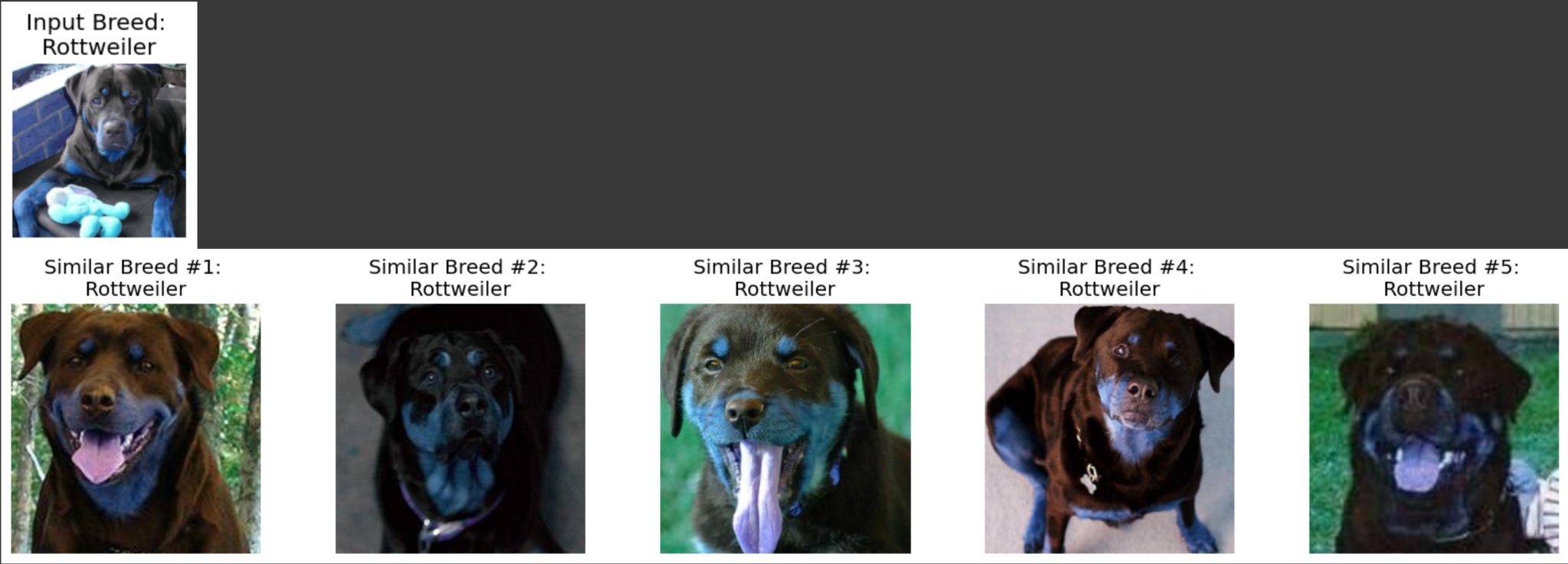
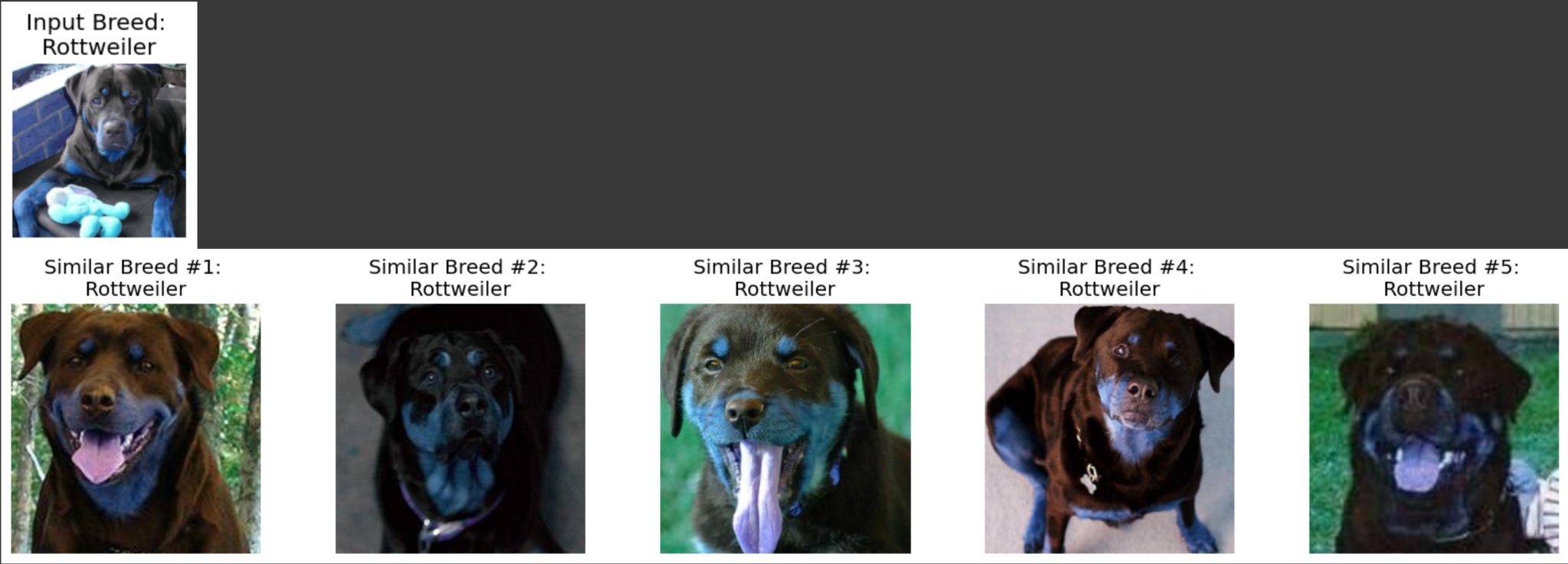
****

****

**Introduction**

Our project delves into the fascinating field of image recognition, specifically focusing on identifying and finding images of dogs that closely resemble a given input image. This is similar to the feature on popular online retail platforms such as Amazon, where customers are shown products similar to the one they are viewing. Our approach harnesses the power of the k-nearest neighbors (KNN) algorithm, a well-established method in machine learning, on the Stanford Dogs dataset. Our system can accurately pinpoint and return images of dogs that bear the most resemblance to the input image.

This project isn't only about matching images; it addresses a real-world need in various sectors. For veterinarians, it offers a tool to correctly identify dog breeds, which is crucial in diagnosing breed-specific ailments and prescribing appropriate treatments. Similarly, potential dog owners exploring different breeds for their future pets can benefit immensely from this technology. It aids in understanding the physical traits of various breeds, helping them make an informed decision.

In the broader context, this project contributes to the advancement of image recognition technology and its application in everyday life. By focusing on a specific domain like dog breeds, it showcases how machine learning can be tailored to meet specialized needs. Furthermore, it can serve as a framework for more advanced applications, such as animal welfare and conservation efforts, where accurate identification of animals plays a critical role.

**Literature Review**

In the realm of dog breed identification using artificial intelligence, there has been a distinct focus on employing transfer learning with pre-trained Convolutional Neural Networks (CNN). This approach is particularly prominent in several studies that utilized the Stanford Dogs Dataset, a comprehensive collection of dog images by breed. This dataset serves as a benchmark for assessing the performance of various AI models in breed classification. Our project, exploring the k-nearest neighbors (KNN) technique, sits within this context but diverges in its methodological approach.

Varshney et al. (2019) exemplify this CNN-centric trend, applying transfer learning from the Inception V3 and VGG16 models. Their findings highlighted Inception V3’s superiority, achieving an 85% accuracy rate. Raduly et al. (2020) also pursued this path but focused on optimizing transfer learning from NASNet-A mobile and Inception Resnet V2. Their research showed that Inception Resnet V2 outperformed NASNet-A, reaching an impressive 90.69% accuracy. Another significant contribution came from Borwarnginn et al. (2021), who retrained pre-trained CNNs like MobileNetV2, InceptionV3, and NASNet. Their most successful model was NASNet combined with data augmentation, achieving an accuracy of 89.92%.

However, not all studies using this approach yielded similar high accuracy. Mulligan and Rivas (2022) adopted a different strategy, utilizing a pre-trained Xception model coupled with a dense layer and a Softmax output for 120 dog breeds. Their model, while innovative, resulted in a more modest balanced accuracy of 54.80% and a log-loss of 9.5954. In contrast, Ayanzadeh and Vahidnia (2018) achieved better results with their exploration of DenseNet-121, ResNet50, DenseNet169, and GoogleNet. Their best performance was with ResNet50, which, after data augmentation and fine-tuning, reached a high accuracy of 89.66%.

These varied outcomes in using transfer learning and CNN architectures for dog breed identification highlight the complexity and challenges inherent in this task. Our project's use of KNN presents an alternative approach, offering a different perspective in this diverse and evolving research landscape.

**Dataset**

For our project, we employed the Stanford Dog Dataset, a comprehensive collection comprising 20,580 images, representing a diverse array of 120 different dog breeds. This dataset, sourced from ImageNet, is particularly notable for its extensive range of breeds and the high quality of images and annotations. Each image in the dataset is accompanied by an XML annotation file, which includes crucial details such as image dimensions, class labels, and bounding boxes. These bounding boxes are essential as they highlight the presence of one or more dogs in each image.

To prepare this dataset for our analysis, we undertook a meticulous preprocessing routine. Each bounding box from the annotation files was carefully cropped and resized to create a uniform image size of 256x256 pixels. This resizing was achieved using OpenCV’s robust functions for image reading, writing, interpolation, and resizing, ensuring the consistency and quality of the images were maintained.

In our quest to optimize image preprocessing, we experimented with different cropping techniques. One such technique involved cropping along the longer dimension of the bounding box and then centering the image against a black background of 256x256 pixels. This method aimed to preserve as much information about the dog and its features as possible. However, upon evaluation, we found that this approach did not significantly enhance the outcome of our analysis. Consequently, we decided to continue with our initial strategy of cropping along the smaller dimension of the bounding box.

This preprocessing phase resulted in a total of 22,125 uniformly sized and high-quality dog images, providing a solid foundation for our analysis. By standardizing the image size and format, we ensured that our k-nearest neighbors model could efficiently and effectively process the dataset, enabling accurate breed identification and comparison across the diverse range of dog breeds in the Stanford Dog Dataset.

**Baseline**

In our project, we established a baseline approach to provide a reference point for evaluating the effectiveness of our k-nearest neighbors model in identifying dog breeds. This baseline is a simple random guessing strategy, which serves to contextualize the performance of more sophisticated algorithms.

Top-1 Accuracy Metric:

Our baseline uses the Top-1 accuracy metric. Here, for a given input dog image, the approach involves randomly selecting one breed from the 120 breeds available in our dataset. Then, it returns an image of a dog from the randomly chosen breed. Top-1 accuracy is calculated as the proportion of instances where the breed of the dog in the randomly selected image matches the breed of the input image. This is measured over the total number of instances where the baseline approach was applied. Given the random nature of this method and the dataset comprising 120 breeds, we anticipate a Top-1 accuracy of or 0.83%.

Top-5 Accuracy Metric

Alongside Top-1, we also employ Top-5 accuracy as a metric. In this case, the process involves making five random guesses for each input image, each time randomly selecting from the 120 breeds (with replacement) and returning an image for each chosen breed. The Top-5 accuracy is determined by the ratio of instances where at least one of the five randomly returned images matches the breed of the input dog image, out of the total instances evaluated. For Top-5 accuracy, the expected value is calculated as or 4.1%.

This baseline approach, though simplistic, is crucial in providing a fundamental benchmark for our project. It allows us to assess the effectiveness of our k-nearest neighbors model by comparing its accuracy against these basic random guessing accuracies. Such comparison helps in quantifying the improvement our model offers over a rudimentary method, thereby highlighting the value added by employing a more sophisticated AI-based approach in dog breed identification.

**Main approach**

In our project, we implemented an advanced approach combining the capabilities of pre-trained Keras application models with the k-nearest neighbors algorithm, tailored for identifying dog breeds through image analysis. This method effectively marries deep learning's feature extraction prowess with the straightforward comparative logic of traditional machine learning.

We began by utilizing pre-trained models from Keras Applications, specifically designed for their proficiency in image feature extraction, having been trained on the expansive ImageNet dataset. We tested with different pre-trained models including NASNetLarge, InceptionResNetV2, Xception, ResNet50V2, DenseNet201, VGG16, ConvNeXtBase, EfficientNetV2M, and detailed information about these models is shown in **Figure 1**.

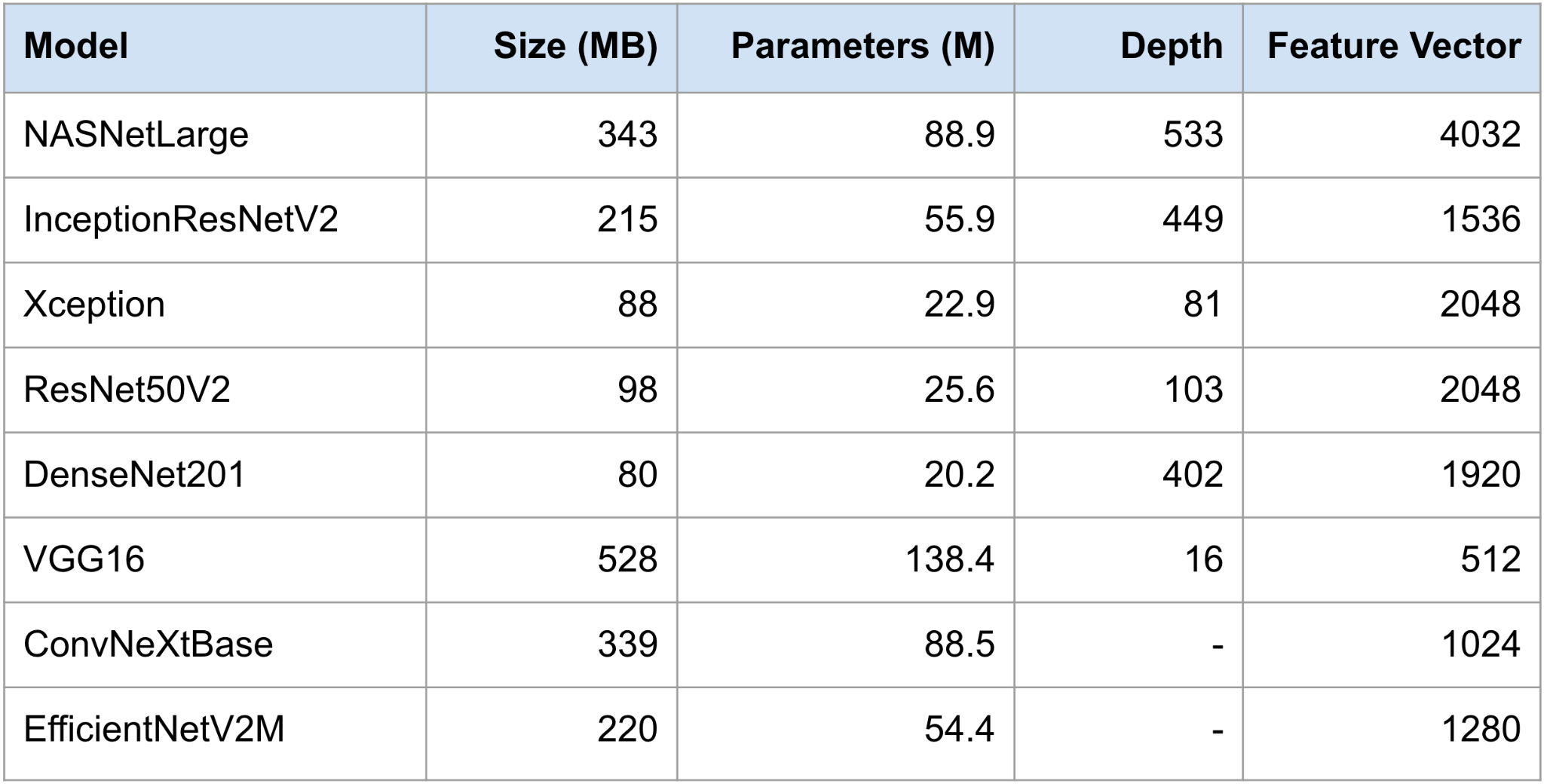


Figure 1: This shows the size, parameters, depth, and dimension of the feature vector produced for each of the pre-trained models tested with.

We checked each model’s Top-1 and Top-5 accuracy to evaluate its performance, which is described in further detail in the Evaluation metric section. The k-nearest neighbors algorithm plays a crucial role following the feature extraction process. Using the KNeighborsClassifier with n=6, we input the feature vector of a given dog image to identify the six most similar feature vectors. This number is chosen because the first feature vector returned is always the image itself, so we select vectors 1-6 to find the five most visually similar dogs. These images and their corresponding feature vectors are stored in a DataFrame, allowing for quick retrieval of the images once the nearest neighbors are identified. It is important to note that by using a KNN, we are not using the labels (dog breed information) at all in this process. Instead, we run the chosen pre-trained model on all images to generate a feature vector for each one, and then use KNN to find the similar feature vectors.

Our development environment leverages the computational power of Google Colab, using a T4 GPU which is significantly faster than traditional CPU processing. The project is developed with Python and employs Keras and TensorFlow for implementing the model and facilitating transfer learning. We assess our model's performance using Top-1 and Top-5 accuracy metrics, testing it on 1,000 random images to identify five similar dogs.

After evaluating Top-1 and Top-5 accuracies for each model, we chose the best performing model to use for fine-tuning, to further improve our results. For the fine-tuning process, the cropped images (22,125) were divided into 3 datasets with the ratio of 80:10:10 (17,701 in training, 2,212 in validation, 2,212 in test). First, we checked accuracy with no training on the test set. Then, we did some initial training by freezing all layers and adding a Dense fully connected layer, followed by a softmax layer for classification of the breed into 120 different categories, as specified in the dataset. We followed this up with the final fine-tuning phase where we unfroze the last 10-20% of layers in the model but kept the first 90% frozen. We tested which number in the range of 10-20% performed the best, and recompiled the model. We chose to do the initial training for 5 epochs and fine-tuning for 2 epochs, as anything above these values would lead to overfitting, which can be seen by a drop in validation accuracy and an increase in validation loss.

**Evaluation Metric**

In assessing the performance of our model for dog breed identification, we employ both Top-1 and Top-5 accuracy metrics, which serve as quantitative measures of the model's accuracy.

Top-1 Accuracy is a measure of precision, indicating the model's ability to correctly identify the breed of the input dog image on its first attempt. To calculate Top-1 accuracy, we first examine the breed of the input image which was stored in the DataFrame. Then, we compare it to the breed of the most similar feature vector found by the KNN. We check how many times the breeds match out of the total number of predictions ran:

Top-1 Accuracy = ​

Top-5 Accuracy expands the criteria for a correct prediction by considering a prediction successful if the actual breed of the dog is among the model's top five guesses. This metric provides insight into the model's ability to rank the correct breed highly, even if it is not the first prediction. The equation for Top-5 accuracy is:

Top-5 Accuracy =

These metrics allow us to quantitatively evaluate the model's effectiveness. Top-1 accuracy indicates the model's precision, while Top-5 accuracy shows its recall ability — how well it can retrieve correct results from its set of predictions.

The model's performance was initially tested on a subset of 1000 random images to assess NASNet's performance as a pre-trained model. Following fine-tuning, the accuracies were re-evaluated on a larger test dataset comprising 2212 images. This evaluation process helps to ensure that the model not only generalizes well beyond the data it was trained on but also maintains high accuracy when confronted with new, unseen images.

**Results & Analysis**

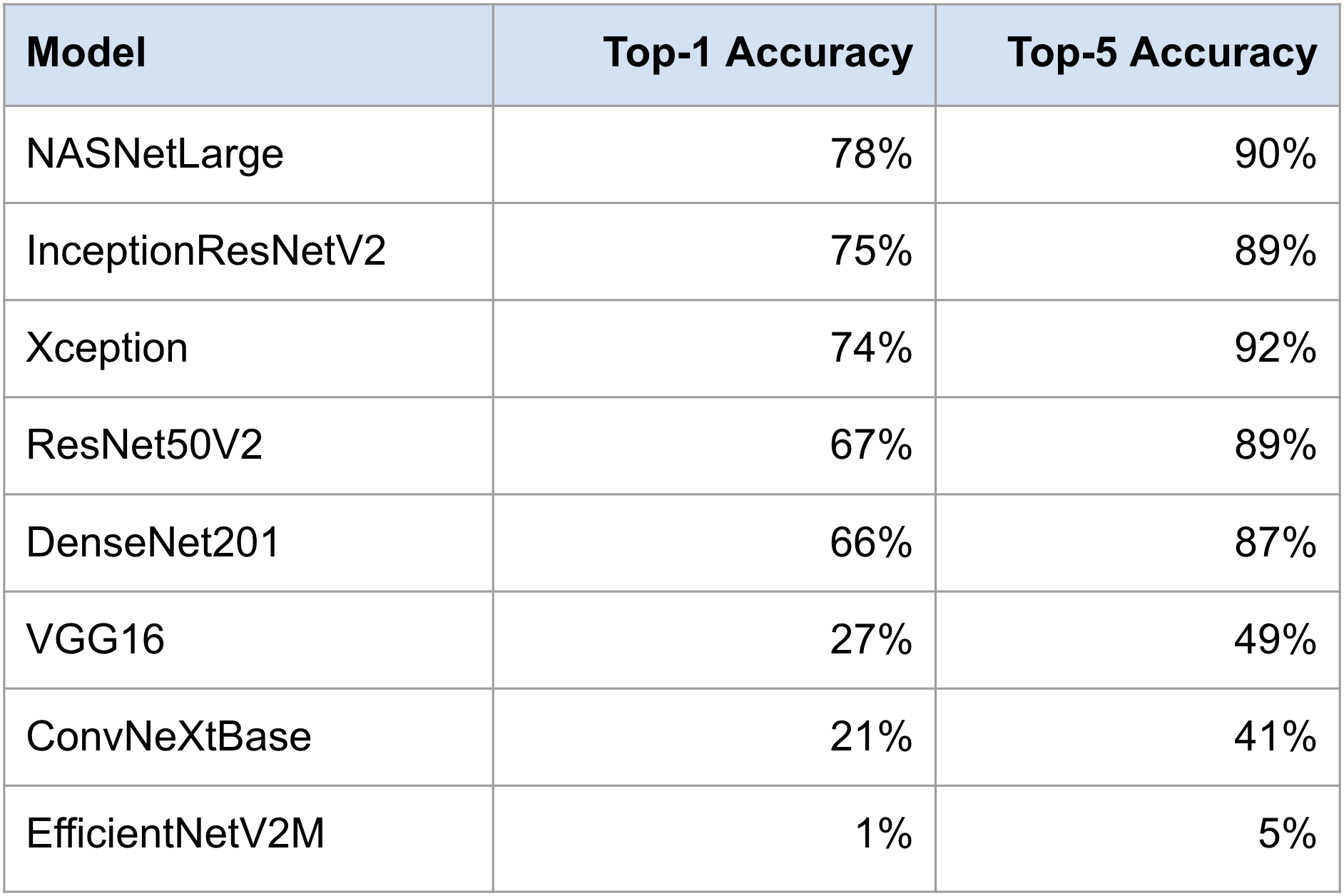


Figure 2: This shows the Top-1 and Top-5 accuracies on a random subset of 1000 images when testing with different pre-trained models shown in the left-most column.

**Figure 2** shows the performance of each pre-trained model when tested on a random subset of 1000 input dog images. Since NASNetLarge was the best performing model with the highest Top-1 Accuracy and a high Top-5 accuracy, we decided to move forward and complete the fine-tuning process on NASNetLarge.

Fine-Tuning

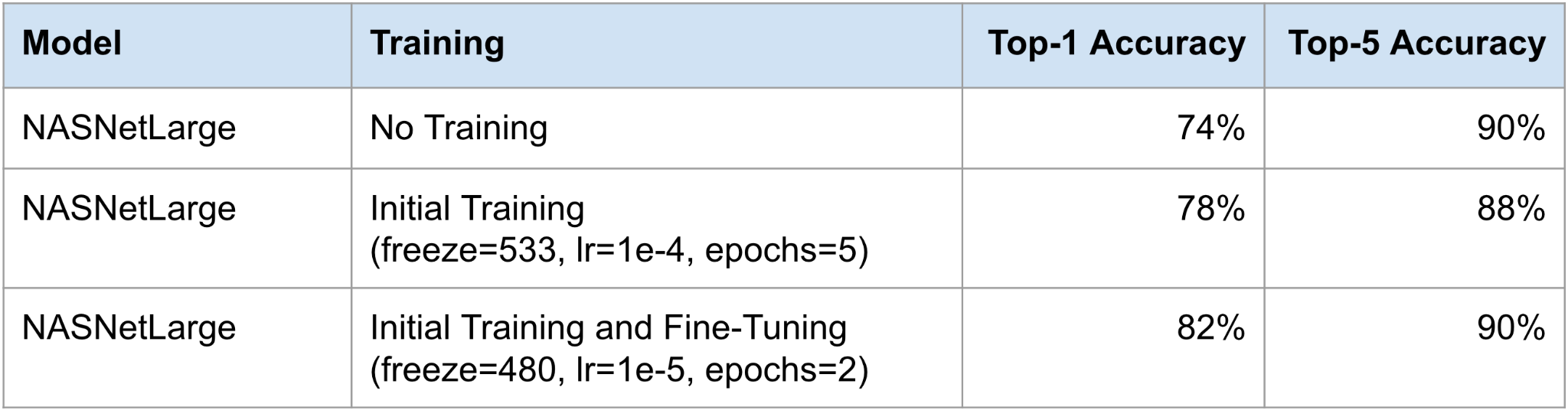
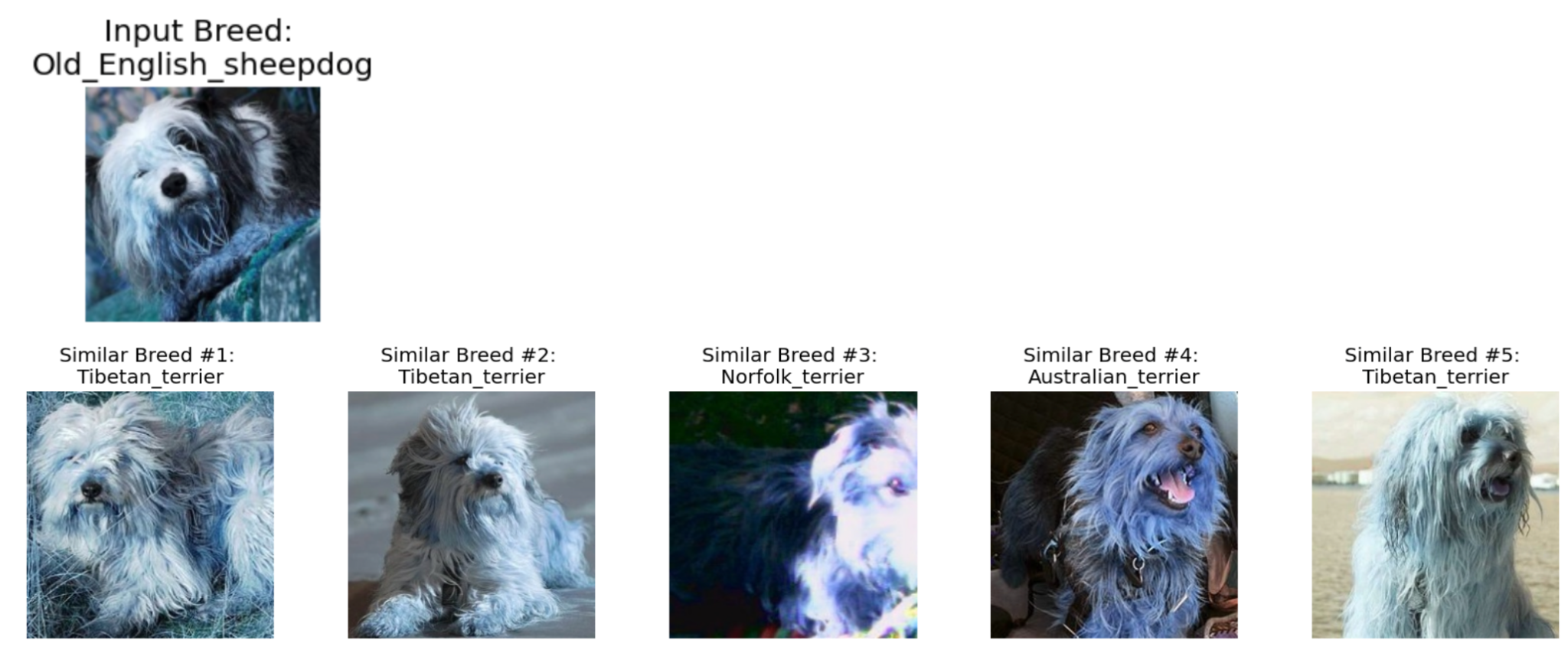


Figure 3: This shows the effect that ‘No Training’, ‘Initial Training’ and ‘Initial Training and Fine-Tuning’ has on the Top-1 and Top-5 accuracy for the NASNetLarge model.

**Figure 3** shows the effect that fine-tuning had on the NASNetLarge model. Initially, with no training NASNetLarge achieved Top-1 accuracy of 74% and Top-5 accuracy of 90% on the test set of 2212 images. After initial training the accuracy for Top-1 went up to 78% and Top-5 decreased slightly to 88%. After fine-tuning was complete, Top-1 accuracy went as high as 82% and Top-5 accuracy stabilized at 90%. Based on these results, it seems that fine-tuning has a large effect on increasing the Top-1 accuracy, but not much effect on Top-5 accuracy.

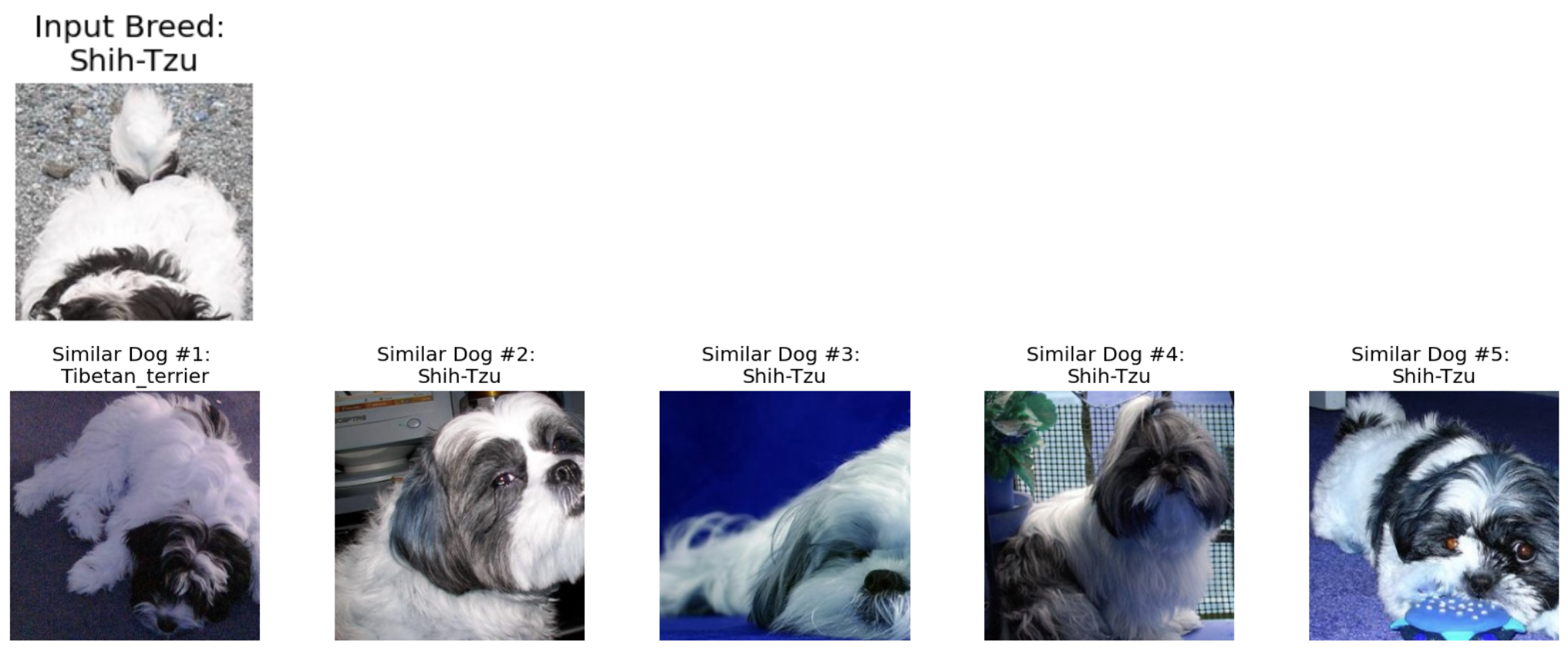
**Error Analysis**

We noticed some specific cases where our model is struggling:

* When the input dog breed looks very similar to another breed:

Here, we see that the Old English Sheepdog and the Norfolk terrier, Australian terrier and Tibetan terrier breeds look very similar.

* When the cropping algorithm is unable to picture the entire dog.



Here, we see that the cropped input Shih-Tzu image is missing most of the dog’s face and is mostly the torso. The most similar image here is a Tibetan terrier in similar colors. However, in the Top-5 we still see that Shih-Tzu was found in 4 out of the 5 similar images. This consistency indicates that the breed's distinctive features are being effectively captured and recognized by the model. Such results also imply that the cropping algorithm employed in our preprocessing stage is performing adequately, as it does not seem to significantly impede the model's ability to discern and match breed characteristics.

**Future Work**

Moving forward we plan to explore vision transformers, which are considered the oracle method right now for finding similar dog images. The best performing ViT currently is ViT-NeT with a Top-1 accuracy of 93.6%. It would be an interesting challenge to try or reproduce these results and experiment with fine-tuning on this pre-trained model. We would also love to work on developing an app for veterinarians which could help them in identifying dog breeds and prescribing the correct medications best on the breed.

**Code**

Link to Jupyter Notebook: <https://colab.research.google.com/drive/1nRNKIYg8_TwfrHyVY5lHreX9xjFs68S-?usp=sharing>

**References**

1. Borwarnginn, P.; Kusakunniran, W.; Karnjanapreechakorn, S.; Thongkanchorn, K. Knowing Your Dog Breed: Identifying a Dog Breed with Deep Learning. Int. J. Autom. Comput. 2021, 18, 45–54.
2. Raduly, Z.; Sulyok, C.; Vadaszi, Z.; Zolde, A. Dog Breed Identification Using Deep Learning. In Proceedings of the IEEE 16th International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, 13–15 September 2018.
3. Varshney, A.; Katiyar, A.; Singh, A.K.; Chauhan, S.S. Dog Breed Classification Using Deep Learning. In Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 25–27 June 2021.
4. Mulligan, K.; Rivas, P. Dog Breed Identification with a Neural Network over Learned Representations from The Xception CNN Architecture. In Proceedings of the 21st International Conference on Artificial Intelligence (ICAI 2019), Xuzhou, China, 22–23 August 2019.
5. Ayanzadeh, A.; Vahidnia, S. Modified deep neural networks for dog breeds identification. Preprints.org 2018, 2018120232.
6. Papers With Code. (n.d.). Stanford Dogs Benchmark (Fine-Grained Image Classification). <https://paperswithcode.com/sota/fine-grained-image-classification-on-stanford-1>