Dog Breed Classifier

**A. Data**

Source - <https://www.kaggle.com/c/dog-breed-identification/data>

You are provided with a training set and a test set of images of dogs. Each image has a filename that is its unique id. The dataset comprises 120 breeds of dogs. The goal is to create a classifier capable of determining a dog's breed from a photo. The list of breeds is as follows:

affenpinscher

afghan\_hound

african\_hunting\_dog

airedale

american\_staffordshire\_terrier

appenzeller

australian\_terrier

basenji

basset

beagle

bedlington\_terrier

bernese\_mountain\_dog

black-and-tan\_coonhound

blenheim\_spaniel

bloodhound

bluetick

border\_collie

border\_terrier

borzoi

boston\_bull

bouvier\_des\_flandres

boxer

brabancon\_griffon

briard

brittany\_spaniel

bull\_mastiff

cairn

cardigan

chesapeake\_bay\_retriever

chihuahua

chow

clumber

cocker\_spaniel

collie

curly-coated\_retriever

dandie\_dinmont

dhole

dingo

doberman

english\_foxhound

english\_setter

english\_springer

entlebucher

eskimo\_dog

flat-coated\_retriever

french\_bulldog

german\_shepherd

german\_short-haired\_pointer

giant\_schnauzer

golden\_retriever

gordon\_setter

great\_dane

great\_pyrenees

greater\_swiss\_mountain\_dog

groenendael

ibizan\_hound

irish\_setter

irish\_terrier

irish\_water\_spaniel

irish\_wolfhound

italian\_greyhound

japanese\_spaniel

keeshond

kelpie

kerry\_blue\_terrier

komondor

kuvasz

labrador\_retriever

lakeland\_terrier

leonberg

lhasa

malamute

malinois

maltese\_dog

mexican\_hairless

miniature\_pinscher

miniature\_poodle

miniature\_schnauzer

newfoundland

norfolk\_terrier

norwegian\_elkhound

norwich\_terrier

old\_english\_sheepdog

otterhound

papillon

pekinese

pembroke

pomeranian

pug

redbone

rhodesian\_ridgeback

rottweiler

saint\_bernard

saluki

samoyed

schipperke

scotch\_terrier

scottish\_deerhound

sealyham\_terrier

shetland\_sheepdog

shih-tzu

siberian\_husky

silky\_terrier

soft-coated\_wheaten\_terrier

staffordshire\_bullterrier

standard\_poodle

standard\_schnauzer

sussex\_spaniel

tibetan\_mastiff

tibetan\_terrier

toy\_poodle

toy\_terrier

vizsla

walker\_hound

weimaraner

welsh\_springer\_spaniel

west\_highland\_white\_terrier

whippet

wire-haired\_fox\_terrier

yorkshire\_terrier

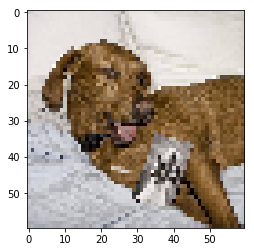
B. Preprocessing

**1 Resize and normalize data**

We define the new image size applied for all images

All the images do not have the same shape. For our model, we need to resize them to the same shape. We use the common practice to reshape them as a square. We also need to normalize our dataset by dividing by 255 all the pixel values. The new pixels values will be in the range [0,1].

Let’s check one image from the training dataset:



**2 Extract the most represented breeds**

We will reduce the database so that we can reduce the complexity of our model. In addition, it will help for the calculation as there will be only N breeds to classify.

**2 One-Hot Labels**

We did one-hot encoding for our labels data.

c. The Model:

**CNN with TensorFlow — Defining Layers**

The architecture will be like this:

* 1st Convolutional Layer with 32 filters
* Max pooling
* Relu
* 2nd Convolutional Layer with 64 filters
* Max pooling
* Relu
* 3rd Convolutional Layer with 128 filters
* Max pooling
* Relu
* DropOut
* Flatten Layer
* Fully Connected Layer with 500 nodes
* Relu
* DropOut
* Fully Connected Layer with n nodes (n = number of breeds)

**Google each of them and write a short piece of about three sentences each to explain**

From a bigger picture, a CNN architecture accomplishes 2 major tasks: feature extraction (convolution + pooling layers) and classification (fully-connected layers).

Model Hosting

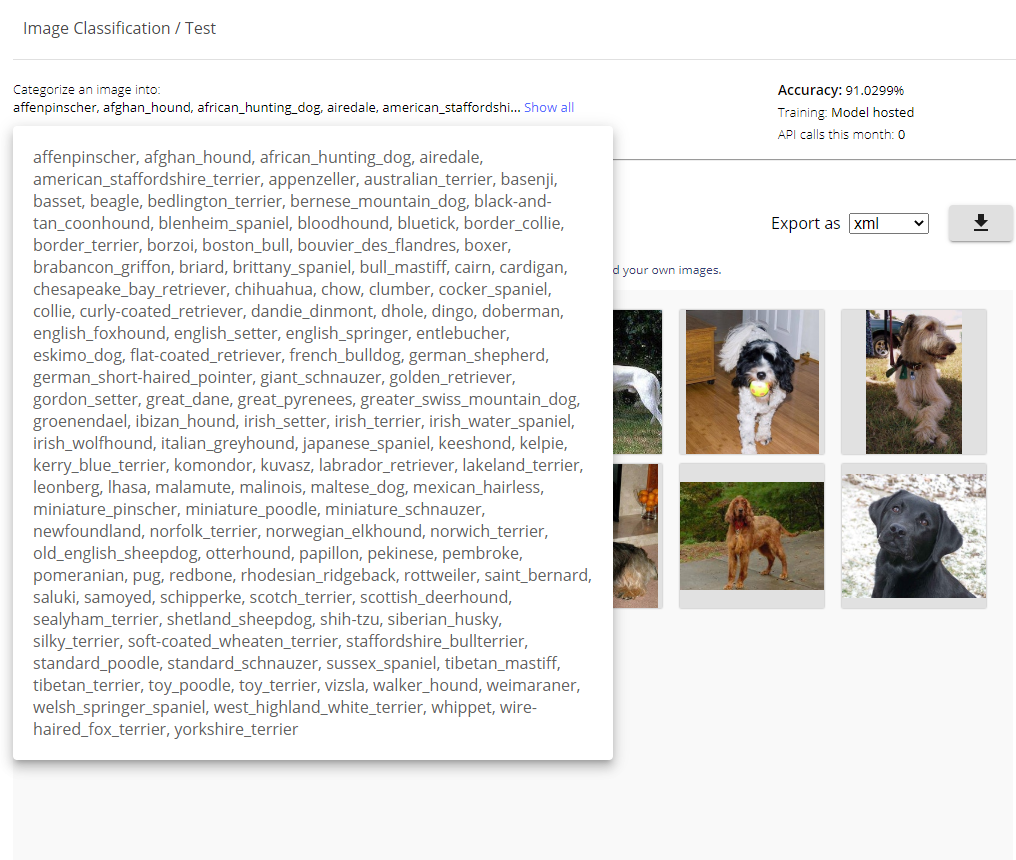
Model is hosted with Nanonets

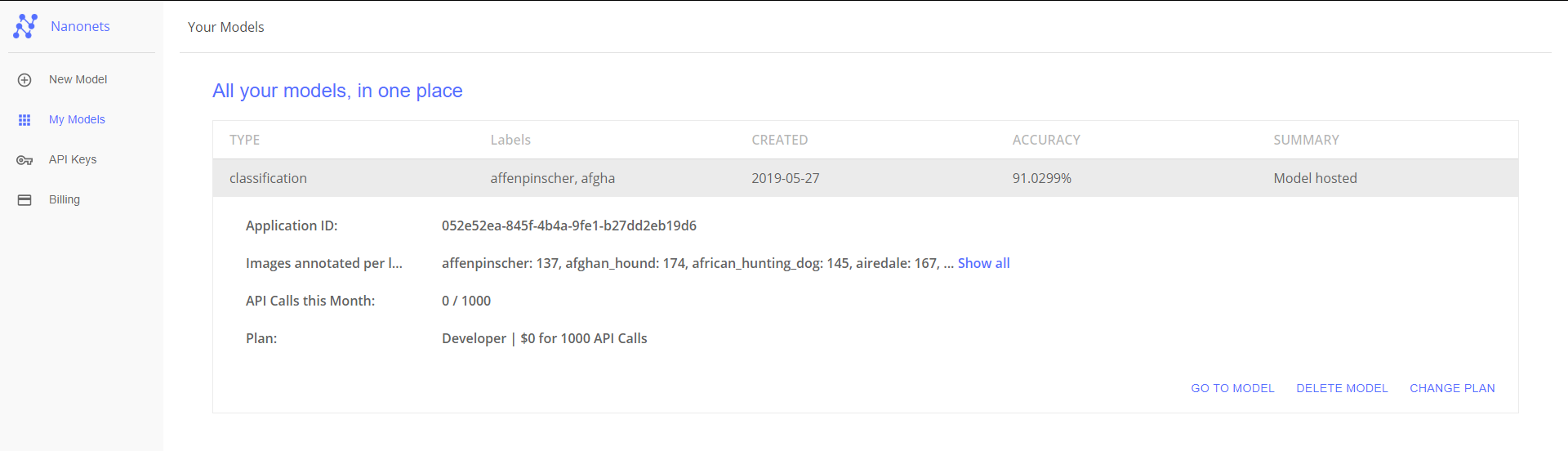
Because the NanoNets models are heavily pre-trained, We made a few modifications to the data From this model, we got 99.3% test accuracy. This is 5% more than the original model in spite of adding just 1/60th of the data! The reason that NanoNets model performs better is:

large amount of pre-training,

optimal hyper-parameter selection,

and data augmentation.

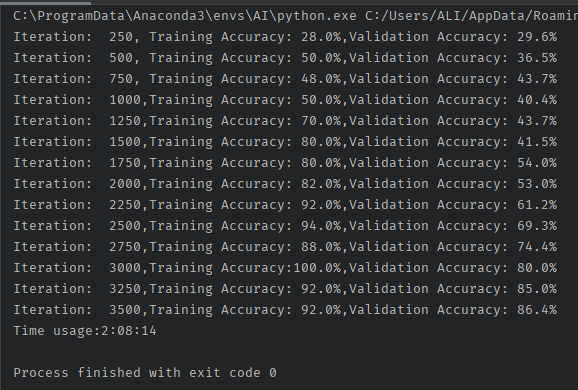




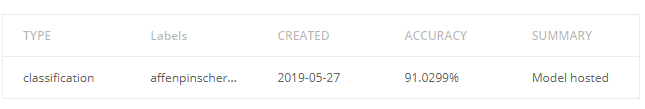
D. Accuracy/Evaluation

We split our train data in two: a training set and a validation set. Therefore, we can check the accuracy of the model train made from the ‘training set’, on the validation set.

Original CNN Model



Hosted CNN Model on Nanonets



4. Implementation

