Capstone Project Proposal

This document is a Capstone Project Proposal, presented to Udacity for completion of Machine Learning Engineer Nanodegree.

Domain Background

Computer vision has greatly evolved in the past few years. Datasets like ImageNet are used to train complex models that perform with an accuracy that would be unimaginable a few years ago. In 2013 a competition was held (https://sites.google.com/site/fgcomp2013/) targeting fine-grained image classification, which is image classification of objects that are visually very similar. The dataset used on this project was taken from this competition.

My interest in airplanes come from an early age. I always loved to go to the airport to observe them. Later I had the opportunity to work in an aircraft manufacturer, which only increased this interest. Then I began trying to identify airplane models every time I would be close to them, and it became a kind of game I would plane with myself. Airplane model identification is an interesting challenge for computer vision, since they are sometimes so similar that even humans have issues with it.

Problem Statement

Different aircraft models can be very similar, even for human vision. Identifying aircraft with high accuracy is an interesting problem, which I propose to solve in this project.

Datasets and Inputs

The dataset used on this project is collection of 10200 images of aircraft, with 100 images for each of 102 classes, or aircraft models. The dataset is available on https://www.robots.ox.ac.uk/~vgg/data/fgvc-aircraft/.

The following is a description of the dataset labels as described on the dataset page:

"Aircraft models are organized in a four-levels hierarchy. The four levels, from finer to coarser, are:

- ***Model**, e.g. Boeing 737-76J. Since certain models are nearly visually indistinguishable, this level is not used in the evaluation.*
- ***Variant**, e.g. Boeing 737-700. A variant collapses all the models that are visually indistinguishable into one class. The dataset comprises 102 different variants.*
- ***Family**, e.g. Boeing 737. The dataset comprises 70 different families.*
- ***Manufacturer**, e.g. Boeing. The dataset comprises 41 different manufacturers."*

Solution statement

The solution proposed uses transfer learning technique, or the use of a pre-trained Convolutional Neural Network, like VGG-16 or ResNet, to extract the features of the images, and train some fully-connected layers to identify the images according to the classes. My proposal is

to develop 3 models, to classify the aircraft according to 3 criteria: manufacturer, family and variant. This system is going to be used by a web application in which users will be able to input an aircraft image and see the prediction made by the model on that image.

Benchmark model

I propose to use the results of the 2013 competition as a benchmark. The results are available at https://sites.google.com/site/fgcomp2013/results, and I reproduce them below:

Team	Additional Details	Aircraft
Inria-Xerox		**81.4581**
CafeNet*	Results of all 5 verticals using bounding boxes at test time.	78.8479
Inria-Xerox		75.8776
VisionMetric*	Distance metric learning with combo features	75.4875
Symbiotic		75.8476
Inria-Xerox		80.5881
CognitiveVision*	Using ILSVRC-2012 training set (~1.3M images) to pre-training the DNN.	67.4167
DPD_Berkeley*	Track 1 results on test data using the method described above.	68.4668
VisionMetric	Distance metric learning with LLC features	73.9274
CognitiveVision	Results by the method described in the abstract.	58.8059
MPG	SIFT, RGB-SIFT, Opponent-SIFT, C-SIFT. Fisher Vector with 256 Gaussians, 8 regions. Logistic regression classifiers.	9.45095
MPG	Different configurations for dogs and cars data (single scale)	9.45095
Infor_FG*	DCNN	30.393
InterfAIce	IGBA v1.2, 4 cycles, 4h20min	5.79058

It is not explicitly declared on the website, but it is presumed that this accuracy results refer to the accuracy of models on the "Variant" classes.

Evaluation metric

The proposed evaluation metric is simple average accuracy across the classes.

Project design

The project will be developed in Python using Pytorch on Jupyter notebooks. Data will be pre-processed to comply with input format of the pre-trained CNNs used. Also data augmentation is going to be used. Pre-trained models that are being considered are AlexNet, VGG and ResNet. The convolutional layers weights of these models will be kept as pre-trained for feature extraction, and a final classification layer (or layers) will be trained to image classification. A website is going to be developed for images input and evaluation. The system will be deployed using AWS platform.