Deployment Solution Architecture

Shengping Jiang, October 03, 2020

1 The goals of the project

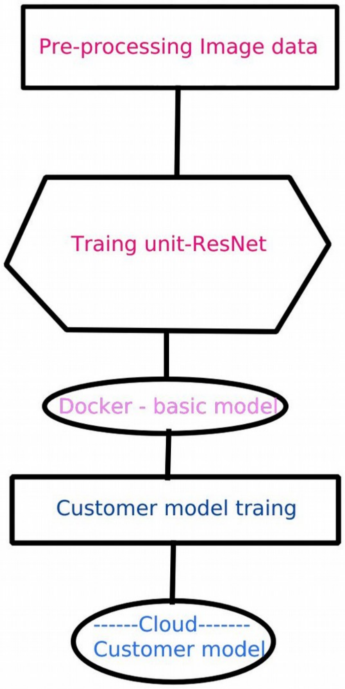
This project will build a ML application for recognizing people with masked face. It is a research project. Below are goals of the project:
a) Able to recognize a person as same person when he/she is with or without a mask, from a webcam or IP camera
b) It will be deploymented as a web application or a off-line application (Windows version or/and Linux version)
c) It can be used in a small or middle size company for general entry management

2 Project description

a) The major components of the system? What are the inputs and
outputs?
Training unit, customer UI (management, monitor camera, test result output), system management (training image input, trained model deployment)
b) Where and how will the data be stored?
Training data are saved in a Linux sever
c) How will data get from one component of the system to another?
The basic face recognition model is trained from a ResNet network. It will be used to predict 128D vector from a face (with or without a mask). Pre-processed training imaged will be inputed to the ResNet through mini-batches
d) What is the lifecycle of the ML/DL model?
○ How frequently do you need to retrain your model? Is it at fixed intervals
The first training will use a big number of face images. The images may belong to 500 or more groups. Each group has at least 5 images
The model may be retrained by each half year
○ What kind of data do you need for retraining? How will you store and
manage it?
We should add new images to retrain the model. It means we add new groups or replace some of existing groups. All images are stored in our server
○ How do you know if the retrained model is good enough to deploy?
Model will be tested in different ways before deployment
○ How will the retrained model be deployed?
The model will be deployed through a docker image (using a cloud PC is an option)
○ How will the retrained model be stored as an artifact?
github
e) How will the system be monitored? How will you debug it if there are problems?
We test the model monthly and also through customer’s feedback to monitor the model

3 **Project Submission Steps**

1) diagram of deployment architecture
Use docker to deploy basic model
Use cloud PC to deploy customer model



2) Project design

The major goal of the project is to recognize masked face. The existing face recognition model such as dlib model was trained by using un-masked face images. Therefore, we need to retrain a basic face prediction model to produce 128D face vector.

Development approach:

1 Collect images of people with mask and without mask<br>
2 Use Dlib ResNet training program to retrain a basic model
3 Deploy basic model and a training unit of customer as a docker image
4 Collect customer’s images
5 Use Dlib CNN face detector to detect face from images. Use retrained basic model to generate 128D vector(face) per image
6 Use customer image encoding data to train a KNN model.
7 Deploy trained KNN model and customer UI to cloud

3) Check list

Data collection
Data Preprocessing (corping, resizing, clustering, etc.)
Training
Algorithm Tuning
Testing
Customer feedback
Improving