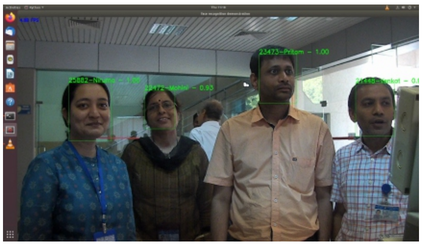
Deep Learning

8

Deep Learning based Face Recognition System

Pritam Prakash Shete* and Dinesh Madhukar Sarode

Computer Division, Bhabha Atomic Research Centre, Mumbai 400085, INDIA



Face recognition system developed in-house

ABSTRACT

Today, face recognition systems are able to achieve the state of the art performance using machine learning. Computer Division, BARC has designed and developed in-house face recognition system using deep learning. It recognizes more than 20,000 registered persons in an unconstrained environment. In this article, we discuss a life cycle of machine learning project with the help of the developed system. We discuss selection of performance metrics, datasets, and neural networks for achieving satisfactory performance. We talk about model deployment using edge computing. We emphasize the importance of data-centric AI development and model monitoring for maintaining an acceptable level of performance.

KEYWORDS: *Face recognition, Deep learning, Data-centric AI development, Model monitoring*

Introduction

Bhabha Atomic Research Centre (BARC) is an organization of strategic importance. Security cameras have been installed in the organization premises for video surveillance. Online intelligent face recognition systems support as well as enhance the power of manual video surveillance. Computer Division, BARC has designed and developed a face recognition system using in-house machine learning framework. Numbers of machine learning design patterns[1] are utilized during the development. It has achieved the state of the art (SOTA) accuracy using the large composite data sets. It contains 20,000 persons from open dataset and representative persons from Computer Division, BARC. In-house field trials have shown ~0% false negative and ~3% false positive rates in an unconstrained environment. A life cycle of a machine learning project starts with project scoping. It involves identifying candidate problems and defining the scope of the project with performance metrics and required resources. Then appropriate open dataset is selected or in-house dataset is collected. Then neural network models are selected, trained, and evaluated in modeling stage. These trained models are continuously monitored for acceptable performance in monitoring stage. Model retraining and dataset curation are essential to address performance degradation due to data drift or concept drift. In this article, we discuss a life cycle of machine learning project with the help of the developed face recognition system.

System Architecture

In-house face recognition system uses a full HD camera connected to a workstation with a high end graphics card for edge computing. The process is divided into three stages as shown in Fig.1. Initially, human faces are detected within input image stream. Then facial features are extracted for these detected faces. Finally these features are utilized for face classification or recognition. Intersection over Union (IoU) and Mean Average Precision(mAP) are used to track the progress of face detection. Accuracy metric is utilized to evaluate face classification.

Dataset

Deep learning models require large datasets to achieve the state of the art performance. Open dataset such as Image Net dataset is driving force for research, where as in-house datasets are essential for local adaptations and continuous model monitoring and improvement.

Open datasets

WIDER FACE dataset[2] is a benchmark dataset for face detection. It is used for training a deep neural network. It contains 32,203 images with 3,93,703 annotated faces for bounding boxes and attributes such as scale, pose, occlusion and so on.

Microsoft dataset [4] is utilized for training of deep neural networks for facial feature extraction and face classification. The base dataset contains approximately 12 lakh aligned facial images for 20,000 celebrities. The novel dataset contains 1,000images for 1,000 celebrities.

In-house dataset

A standard operating procedure is established for in-house dataset collection. Representative employees from Computer Division, BARC have been registered. Pre-processing is performed to align facial images. In-house dataset is combined with open dataset for further processing. Rebalancing design pattern[1] is employed to reduce dataset imbalance.

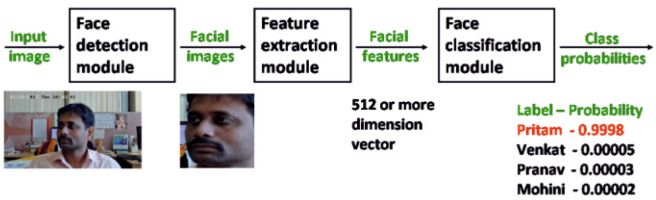


Fig.1: Overview of face recognition system.

**Author for Correspondence: Pritam Prakash Shete*
E-mail: ppshete@barc.gov.in

Modeling

Numbers of neural network models are studied for face detection and face or image classification in the modeling stage.

Face detection

In-house Multi Task Cascaded Convolutional Neural Networks and You Only Look Once (YOLO)[5] are trained using WIDER FACE dataset for face detection. 94.5% accuracy is achieved on FDDB dataset[3]. YOLO provides better performance for full HD image stream.

Model centric AI development

Inception model is selected as the baseline model for performance evaluation. Inception, Inception-ResNet, and ResNet[6] models are selected for model centric AI development.

Transfer learning

Transfer learning design pattern[1] is utilized to fine-tune models trained using Image Net dataset. As shown in Table 1, comparable test dataset accuracy is observed for all models.

Table 1: Model centric AI development using transfer learning.

Models	Inception	Inception- ResNet	ResNet50-v2	ResNet152 - v2
Accuracy	98.5%	98.6%	98.6%	98.65%

Without transfer learning

Models are trained from scratch using Microsoft dataset. As shown in Table 2, the accuracy is marginally better than respective models trained using transfer learning. It is mainly due to the use of Microsoft dataset as it matches better than Image Net dataset for face classification.

Table 2: Model centric AI development without using transfer learning.

Models	Inception	Inception- ResNet	ResNet50-v2	ResNet152 - v2
Accuracy	99.1%	99.2%	99.2%	99.3%

Data centric AI development

Data centric AI development is performed using in-house ResNet18-v2 model. As shown in Table 3, just 75% accuracy is observed without using any data augmentation. Random crop and resize is the most dominant data augmentation for face or image classification.

Table 3: Data centric AI development using data augmentations.

Data augmentation	No data augmentation	Random crop & random resize	Colour augmentation	Horizontal flip augmentation
Accuracy	75.0%	97.5%	98.2%	98.3%

Data curation

Rebalancing design pattern[1] is employed to reduce dataset imbalance. Microsoft dataset is pre-processed to match the deployment images. As shown in Table 4, maintaining aspect ratio and native image resolution provides the best performance.

Table 4: Data centric AI development using data curation.

Data augmentation	Native dataset	Aspect ratio	Aspect ratio plus native image resolution
Accuracy	98.75%	99.6%	99.8%

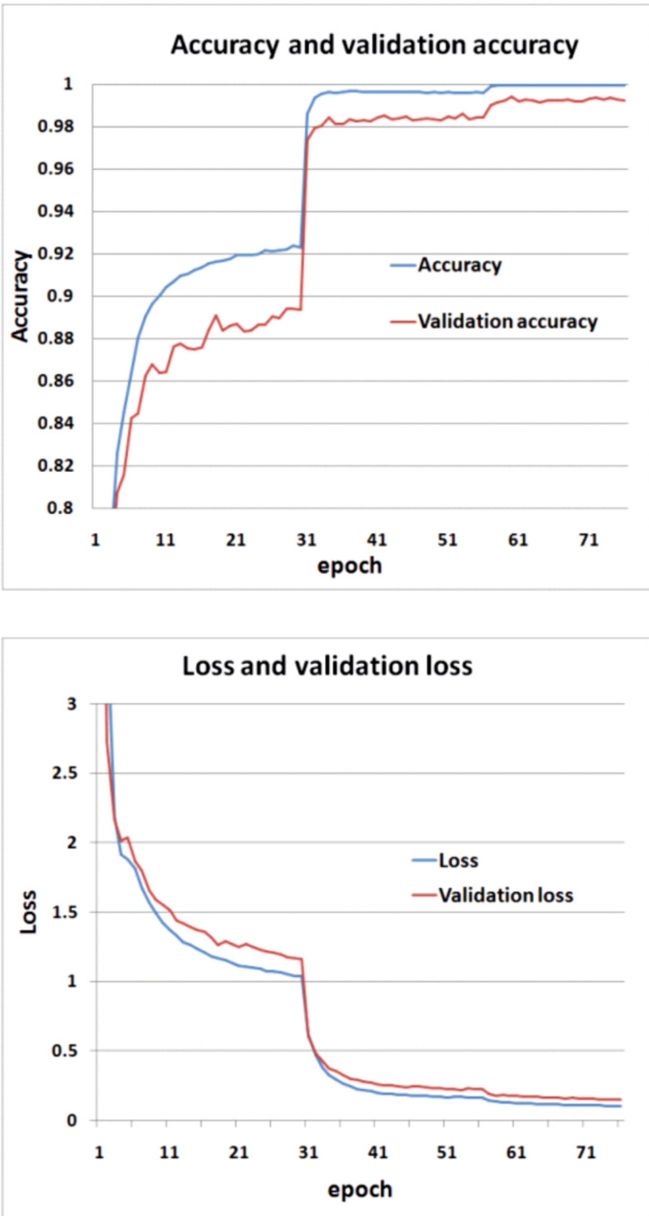


Fig.2: Model performance on training and validation datasets.

As shown in Fig.2, the low model bias and the competency of in-house ResNet18-v2 model is indicated by nearly 100% training accuracy. Data augmentations and model regularizations have reduced the gap between training and validation accuracy as well as the model variance. Training and validation losses are initially reduced sharply and then saturated with epochs.

Field Trials and Monitoring

Field trials conducted at Computer Division, BARC helped us to get more insight about the working environment and arriving at important considerations about camera, optimization of the in-house model and identifying failure cases and thereby improving the overall system.

Camera Selection

Various types of cameras are assessed for quality of acquired image stream. Image quality is improved and motion blur is reduced using appropriate camera placement and parameters. Camera is installed to cover the shortest and the tallest person without operational constrains.

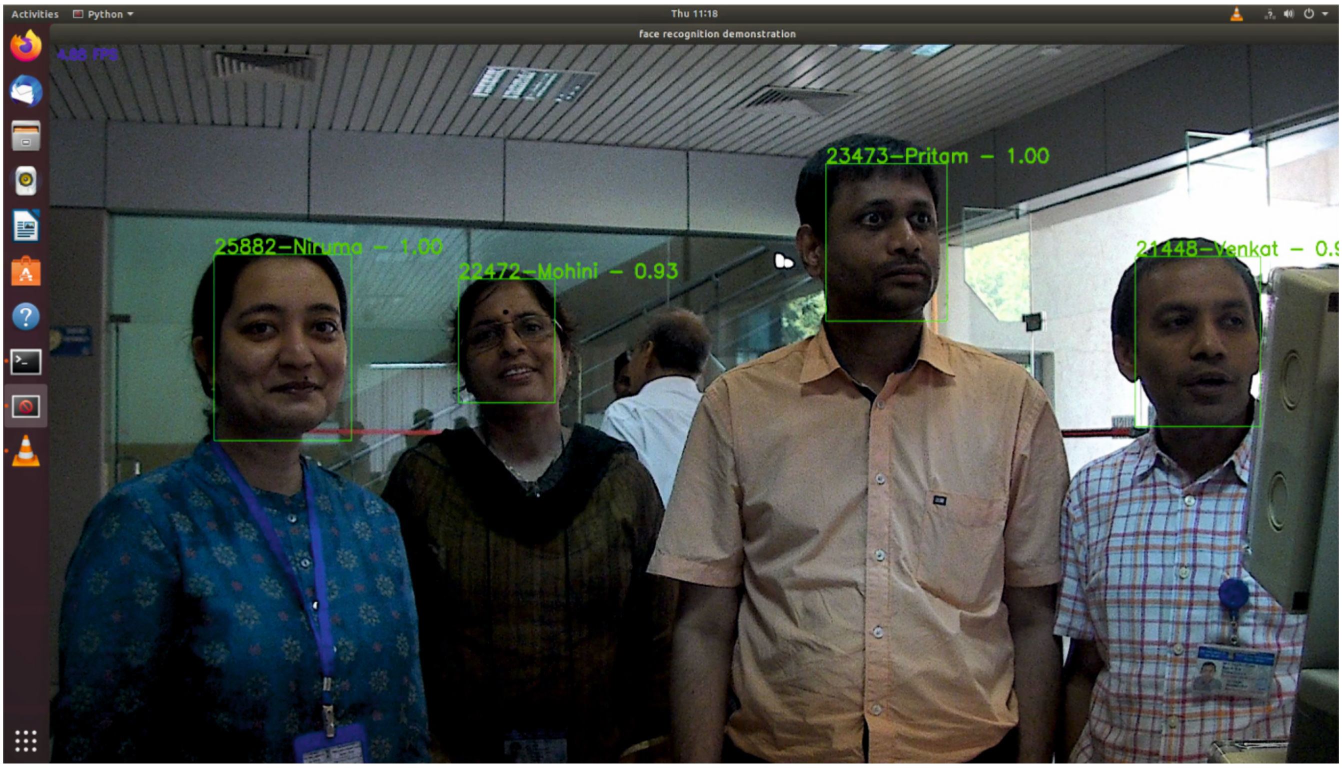


Fig.3: In-house face recognition system in action at Computer Division, BARC.

Post training optimization

Software level optimization is performed using Open CV-DNN and Tensor RT based compiler. Model quantization is utilized for algorithm level optimization for faster inference.

Edge computing

Edge computing is employed as compared to prediction server for live processing and low prediction latency. It has 100 ms latency and runs live with 10 to 15 frames per second.

Model monitoring

System performance is continuously monitored for more than 9 months. Failure cases due to eye blinking, face occlusion, facial poses, and lighting conditions are identified. These are corrected using the in-house dataset and exhaustive data augmentations.

Results

The developed system recognizes persons with eye-glasses, turban, cap and so on. It also identifies persons with different ages and different facial poses. As shown in Fig. 3, it is able to recognize multiple persons using images acquired from full HD camera. The field trials conducted in an unconstrained environment achieved ~0% false negative and ~3% false positive rate.

Conclusion

In this article, we have discussed a workstation based face recognition system developed in-house based on deep learning techniques and its full life cycle. During the field trials we have demonstrated the competency of system even in an unconstrained environment. The system achieved ~99.8% accuracy on open datasets containing more than 20,000 persons.

References

[1] Valliappa Lakshmanan and et. al., Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps, O'Reilly Media, October 2020.

[2] Yang and et. al., "WIDER FACE - A Face Detection Benchmark", in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[3] Vidit Jainand et. al., "FDDB - A Benchmark for Face Detection in Unconstrained Settings", Technical Report UM-CS-2010-2009, University of Massachusetts, 2010.

[4] Yandong Guo and et. al., "MS-Celeb-1M - A Dataset and Benchmark for Large-Scale Face Recognition", in Proc. of European Conference on Computer Vision, ECCV 2016.

[5] Alexey Bochkovskiy and et. al., "YOLOv4 - Optimal Speed and Accuracy of Object Detection", arXiv, 2020.

[6] Kaiming He and et. al., "Deep Residual Learning for Image Recognition", in Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.