Applying AI Techniques for Building Vision Products

Abstract

Traditionally Image Processing and Computer Vision solution design and development, have been using tried and tested set of methods and algorithms, for solving many basic problems related to segmentation, object detection, object tracking, object classification, object recognition. These basic blocks are required in multitude of end-user applications involving scene understanding, scene augmentation and context awareness such as e-commerce solutions, autonomous navigation, human robot interaction etc... However, even the best of these solutions have failed to successfully handle challenging scenarios involving reflections, occlusions, shadows, partial objects, perspective depth information, texture handling, foreground background matching etc..

These solutions have seen a revolutionary new trend with the onset of AI techniques and paradigms related to machine learning and deep learning, and there has been dramatic progress during the past 4 to 5 years. In these newer approaches, researchers have proposed and adapted such methods successfully to obtain classification and recognition accuracies very near to 100%. This has led to differentiated possibilities, wherein, there have been instances, involving training, testing and validating huge datasets encompassing the variety of scenarios required to solve challenging problems such as those mentioned above, which was unthinkable earlier. This is been made possible also due to recent trends in GPU farm infra, cloud infra and advances in AI based embedded heterogeneous computing. Specifically many of these CNN, DNN, RNN models are amenable for adaptation to new use cases without requiring extensive / computation heavy retraining. Many packages also exist which offer flexible Python APIs to compose custom pipelines which have features related to dataset handling (synthesis, augmentation, etc..), transfer learning and fine tuning pre-trained models for Object classification, Object detection and Image matching, evaluating trained models against standard baseline models, and also for deploying such trained models.

Given the above context, this workshop provides a glimpse of machine learning and deep learning techniques and algorithms, applied to designing and developing computer vision solutions. Such solutions have been the backbone of many productized features in embedded platforms such as smartphones. The workshop gives a peek into sample technology challenges in few real-life user scenarios, and demonstrates how these can be attempted to solve by designing computer vision solutions based on AI principles.

Session 1: Introduction to ML / DL Basics

Speaker : Dr Sandeep Palakkal

Presentation Content :

This session starts by giving a very brief introduction to basic concepts in Machine Learning and an overall picture of Machine Learning as practiced in the industry. This session also introduces theory and concepts in deep learning & deep neural networks. Evolution of convolutional neural networks (CNN) are reviewed, followed by a discussion on the training and testing procedure of CNN. The overall objective of the session is to provide quick glimpse of the practical technical challenges in training and testing CNNs, which are further explained in the hands-on session that follows

Hands-on Content:

In the hands-on session, the audience are introduced to training and testing of machine learning models and CNNs. The hands-on experiments will include the following.

1. Linear logistic regression: a simple classification problem using linear logistic regression using a simple dataset (e.g., https://www.kaggle.com/crawford/gene-expression)
2. Test a CNN (e.g., GoogleNet, Mobilenet, etc.), which is pretrained on ImageNet for 1000-way image classification problem
3. Train a light-weight CNN on CIFAR-10 dataset for 10-way image classification
4. Train a CNN for emotion Classification using an open source dataset (e.g., https://mmifacedb.eu/)

Session 2: Application of DL to Computer Vision

Speaker : Dr Venkat Peddigari

Presentation Content :

Semantic segmentation is one of the key problems in the field of computer vision as it aids us in complete scene understanding which is quite useful in many applications such as autonomous vehicles, human computer interaction, augmented reality etc.,. Semantic segmentation provides detailed pixel level classification of images, which essentially clusters parts of the image belonging to the same class. This is particularly suited for Advanced Driver Assistance Systems (ADAS) as these applications require accurate segmentation of the obstacles and road boundaries, which can run in real-time with low memory requirements. With the advent of deep architectures such as Convolutional Neural Nets (CNNs) and various optimization techniques such as convolution factorization, network pruning and weight quantization, many state-of-the-art techniques adopted Deep Neural Networks for performing semantic image segmentation.

A semantic segmentation architecture can be broadly classified into encoder and decoder networks. The encoder network is generally a pre-trained classification network to learn

discriminative features at a lower resolution. The major task of the decoder is to perform dense classification at pixel level by semantically projecting the discriminative features learnt by encoder at lower resolution onto the high resolution pixel space

Hands-on Content:
In the hands-on session, the audience will learn fine tuning and inferring of pre-trained CNN based networks (e.g., ENet, SkipNet, etc.,) to perform semantic segmentation on sample datasets (e.g., CityScapes, Kitti etc.,). The hands-on experiments will include the following.
1. Comparative study of pre-trained model with various input resolutions
2. Fine tuning of the network to focus only on limited set of classes relevant to ADAS use case

Session 3 : Application of DL to Computational Intelligence
Speaker : Dr Rituparna Sarkar
Presentation Content :
In this session, we introduce concepts of stereo vision. The stereo vision is necessary to estimate the distance of objects from the imaging device. In the first part of the talk, we will discuss stereo camera calibration concepts and why it is necessary in stereo depth estimation. We first introduce the classical computer vision based correspondence matching problem for depth estimation and then move on to discuss the latest advancement and approaches of deep learning based stereo vision methods. We further demonstrate how DL based solutions have extended the stereo concept to solve single image based depth estimation problem. Finally, we give a glimpse of various applications of depth in mobile phone cameras and autonomous driving.

Hands-on Content:
In the hands-on session, the audience is introduced to testing of CNN models for stereo depth estimation for different image sets. The hands on session are designed mailny to understand the advantages and drawbacks of the approaches in real-life scenarios. The hands-on experiments will include the following.
1. Test patch based CNN model for the following criteria
a. Calibrated and un-calibrated images
b. Low light scenario
c. Transparent/reflective objects and others.
2. Test end to end CNN network for similar test image criteria as above

Session 4 : Enabling AI on Embedded Platforms
Speaker : Sharan Allur & Venkappa Mala
Presentation Content :
Deep Learning based solutions have been producing better than State of the Art results for many Computer Vision, Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) tasks. Currently many of these use cases are realized on cloud due to huge memory and

computational requirement, which makes it difficult to achieve real time performance on low power embedded devices. Many researches around the world started focusing on brining the AI capabilities to embedded devices (On-Device AI) from the cloud due to security, privacy and latency concerns. The primary focus of the talk is about how to reduce the computational complexity of the DNNs (pruning, lower precision arithmetic etc.) without altering the accuracy and also acceleration on underlying heterogeneous compute resources

Hands-on Content:

Hands-on experiments include the following,

- Introduction to Build environment
- Cross compilation
- Deploy and Run a model using Caffe or TensorFlow on Android

Timeline:

Topic - Applying AI Techniques for Building Advanced Vision Products"					
Sl No.	**From**	**To**	**Topic**	**Speaker**	**Remarks**
1	9-00 AM	9-15 AM	Welcome & Context Setting	Dr Lokesh Boregowda	
2	9-15 AM	10-45 AM	Introduction to ML / DL Basics	Dr Sandeep Palakkal	30 min Talk + 1 hour hands-on
3	10-45 AM	11-00 AM	TEA / COFFEE BREAK		
4	11-00 AM	12-30 PM	Application of DL to Computer Vision	Dr Venkat Peddigari	30 min Talk + 1 hour hands-on
5	12-30 PM	1-30 PM	LUNCH BREAK		
6	1-30 PM	3-30 PM	Application of DL to Computational Intelligence	Dr Rituparna Sarkar	30 min Talk + 1 hour hands-on
7	3-30 PM	3-45 PM	TEA / COFFEE BREAK		
8	3-45 PM	5-15 PM	Enabling AI on Embedded Platforms	Sharan Allur	30 min Talk + 1 hour hands-on
9	5-15 PM	5-30 PM	Wrap-up & Closure Session	Dr Lokesh Boregowda	

Techfest Office
Students' Activity Centre
IIT Bombay, Powai, Mumbai
Maharashtra - 400076