

## Applying AI Techniques for Building Vision Products

### Abstract

Deep learning based solutions are gaining more importance to solve the problems in computer vision such as segmentation, object detection, object tracking, object classification, object recognition, etc. due to high accuracy of the solutions and availability of computing resources. The primary challenge of deep learning is to prepare huge datasets, training deep neural networks on such datasets and deploying on low power embedded devices such as smartphones. This workshop provides

- A glimpse of deep learning, specifically applied to development of computer vision solutions
- Various challenges involved in training and fine-tuning of a deep neural networks
- Optimization of deep learning solutions to enable real-time execution on embedded systems

### Session 1: Introduction to Deep Learning Speaker: Dr. Sandeep Palakkal

This session covers basic concepts in Machine Learning and an overall picture of Machine Learning as practiced in the industry. It also includes an introduction to theory and concepts in deep learning & Deep Neural Networks (DNNs), evolution of Convolutional Neural Networks (CNN), 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.

#### Hands-on:

The audience are introduced to basic building blocks in a CNN such as convolution, max-pooling, and ReLU operations. The audience will implement some of these operations in python and appreciate the nuts and bolts of a deep neural network. The session will also focus on training and testing of CNNs. The hands-on experiments will include the following.

1. Test a CNN (e.g., GoogleNet, Mobilenet, etc.), pre-trained on ImageNet
2. Train a light-weight CNN on CIFAR-10 dataset for 10-way image classification
3. Learn the impact of hyper-parameters such as learning rate on training & convergence of a CNN

### Session 2: Application of DL to Computer Vision Speaker: Dr. Venkat Peddigari

Semantic segmentation is one of the key problems in the field of computer vision as it helps in complete scene understanding, which is quite useful in many applications such as autonomous navigation, human computer interaction and augmented reality. Semantic segmentation provides detailed pixel level classification of images, which clusters parts of images belonging to the same class.

#### Hands-on:

The audience will learn to build a simple semantic segmentation neural network on a sample dataset, which includes training and testing to provide end-to-end experience of building semantic segmentation solution.

1. Build a sample semantic segmentation network for a given training and test dataset
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: Mr. Pankaj Kumar Bajpai**

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 briefly on stereo camera calibration concepts and its necessity for stereo depth estimation. The later part of session, introduces the classical computer vision based correspondence matching problem for depth estimation and then further discusses on the latest advancement and approaches of deep learning based stereo vision methods. Finally, we give a glimpse of various applications of depth in mobile phone cameras and autonomous driving.

**Hands-on:**

Audience will learn inference from pre-trained CNN model for stereo depth estimation on publicly available Sceneflow dataset. The hands on session is designed mainly to understand the advantages of various variations in network architecture. The hands-on experiments will include the testing on three different variations.

1. Vanilla Dispnet network
2. Dispnet with refinement network
3. Dispnet with refinement network + smoothness constraint

**Session 4: On-Device Deep Learning Speaker: Dr. Narasinga Rao Miniskar**

Many of the Deep Learning based solutions for Computer Vision, Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) tasks are realized on cloud due to huge memory and computational requirements. It is a huge challenge to achieve real time performance of these solutions on low power embedded devices. Many researchers are focusing on solving the challenges in bringing the Deep Learning solutions to embedded devices (On-Device DL) from the cloud. The primary focus of the talk is about how to reduce the computational complexity of the DNNs, by applying techniques such as pruning, lower precision arithmetic and layer optimizations, without altering the accuracy.

**Hands-on:**

Hands-on experiments include the following,

1. Realization of face detection application on embedded device
2. Exploration of convolution operation and performance analysis through Android APK