**DEEP LEARNING FOR SEMANTIC SEGMENTATION OF UAV VIDEOS**

The traditional approaches use image segmentation deep learning methods and extend it to frames of video. This does not take into account the temporal features of subsequent frames. This paper states video segmentation method using a combination of FCN (Full Convolutional Network) and Conv-LSTM (Convolutional Long Short Term Memory). The architecture stated in this paper’s methodology consist of FCN instead of CNN (Convolutional Neural Network) because FCN can be used for arbitrary sized inputs and it generates pixel level segmentation output. The input videos are captured using UAV (Unmanned Aerial Vehicle). The input video frames are fed into FCN and the output segmentations of FCN are grouped into blocks of 4 frames with 3 overlapping frames in consecutive blocks. These blocks are then fed to Conv-LSTM. Conv-LSTM has a loop in its network which allows previous state features to be used in the next frames. The results of this paper show that using a combination of FCN and Conv-LSTM increases the accuracy of almost all classes.

**EFFECTIVE OBJECT DETECTION FROM TRAFFIC CAMERA VIDEOS**

This paper states the work done by a group of researchers for the NVIDEA AI Challenge. This challenge consisted of video recording of a traffic intersection. The objective was to classify different kinds of four wheelers, 2 wheelers, pedestrians, etc. This paper compared 2 deep learning methods namely, R-FCN (Region-based Full Convolutional Network) and Faster RCNN (Region-based Convolutional Neural Network). This paper used the dataset provided by NVIDEA which consisted of 8 hour length video with annotated frames for training and testing. The transformations this paper used was flipping of frames. The architecture used in this paper consists of a RPN (Region Proposal Network) which gives score to regions in the frames. These region scores are then passed to the above 2 deep learning networks. The results show that R-FCN performed better than Faster RCNN. Also, R-FCN has pixel level classification design which can be easily applied to any general detection task. The results were not able to differentiate similar appearing classes (like variants of four wheelers). This results can be improved by using temporal features and using more dataset and variance.

**NEURAL NETWORK-BASED VEHICLE AND PEDESTRIAN DETECTION FOR VIDEO ANALYSIS SYSTEM**

This paper compares 3 neural network architectures for vehicle and pedestrian detection. The three neural network architectures compared are R-CNN (Region-based Convolutional Neural Network), YOLO (You Only Look Once), and SSD (Single Shot multibox Detector). The parameters used for comparision were AUC (Area Under precision-recall Curve), mAP (mean Average Precision), and processing time. Different variants/networks of these 3 architectures were used testing (Faster R-CNN + Resnet-101, Faster R-CNN + InceptionResnet-2, YOLOv3, SSD + MobileNet-1, SSD + MobileNet-2). Results show that Faster-RCNN performs better in terms of AUC and mAP. The drawback of this architecture is that RCNN uses grids of region whereas other 2 methods process the entire image at once. Thus the processing time of Faster-RCNN method is higher than other 2 methods. This processing time is crucial for real time processing where the accuracy requirements cannot always be met.

**OBJECT DETECTION AND TRCACKING BASED ON CONVOLUTIONAL NEURAL NETWORKS FOR HIGH-RESOLUTION OPTICAL REMOTE SENSING VIDEO**

This paper states a method for object detection in high resolution remote sensing videos using neural networks. The considered architectures were YOLO (You Only Look Once) and Faster R-CNN (Region-based Convolutional Neural Network). Faster R-CNN was selected for this paper because it produces better result for smaller sized objects which are present in remote sensing video. The dataset used in this paper is sattelite videos capturing aircraft and other similar objects. This paper uses sub sampling strategy which uses sliding window mechanism. This sub samping strategy allows for adapting to different resolution images. The proposed architecture uses temporal features of previous frame. The RoI (Region of Interest) generated for the previous frame is fused with current frame’s RoI. In this way, the current frame’s RoI is passed on to the next frame. This temporal feature has improved the results as compared to Faster R-CNN.

**OBJECT DETECTION ON VIDEO IMAGES BASED ON RFCN AND GROWCUT ALGORITHM**

This paper focuses on automation of picking objects from inventory in the logistics industry due to decrease in the productive age population in Japan. As there are a lot of different objects in logistics industry, the object detection algorithm is required in the robot vision. The dataset used by this paper is PASCAL Visual Object Classes 2007 dataset consisting of large number of classes. The algorithm this paper uses is R-FCN (Region-based Full Convolutional Network) for object detection. After object detection, the paper’s methodolgy uses Growcut algorithm for segmentation. This segmentation refines the bounding box produced by R-FCN to actual shape of the object using background and foreground seeding method. This paper uses Kineet for Windows v2 (Kinect v2) for capturing RGB images and depth images. This RGBD channels acts as the input. And thus, by using refined segmentation and depth information, robot vision can measure the depth of the object for picking. The mean accuracy precision obtained by this paper is 67%.

**OBJECT DETECTION, CLASSIFICATION AND TRACKING METHODS FOR VIDEO SURVEILLANCE: A REVIEW**

This paper states various methods of object detection, classification and tracking of objects in videos. The detection methods are Motion-based Object Detection (including Background Subtraction, Frame Differencing, Optical Flow Method) and Appearance-based Object Detection (including Template Matching and Feature Extraction). The classification of the detected objects can be done based on 4 attributes: shape and size, color, texture, and motion. This paper also states methods of tracking of the classified objects. This paper compares the different methods and sub-methods mentioned above with their advantages and disadvantages.

**PANORAMA STITCHING, MOVING OBJECT DETECTION AND TRACKING IN UAV VIDEOS**

This paper states methodology for panaroma stitching of frames obtained by UAV (Unmanned Aerial Vehicle). The subsequent frames are compared and merged to form a continuous panaromic image. The paper also states object detection methodology using lq-estimator which is the normalization function comparing features of the image. This method suffers from outliers and hence tuning is required for normalizing function. The detected object is then tracked in the subsequent frames using thresholding. The results show that the methods mentioned in this paper works well for both smooth textured and rough textured images.

**VIDEO OBJECT DETECTION FOR TRACTABILITY WITH DEEP LEARNING METHOD**

This paper states object detection methods in video dataset using deep learning methods for traceability. Annotated database is constructed first marking the object of interest and training the model off line. The architecture used in this paper is CNN (Convolutional Neural Network). Once trained, the model is deployed online and objects are detected and traced using Guassian background modelling and other non parametric modelling. The results show that this deep learning method is suitable for downloaded videos as well as real time video analysis.

**VIDEO PREDICTIVE OBJECT DETECTOR**

This paper states a methodology for Predictive Object Detection (POD) using deep learning techniques. The architectures used in this paper uses YOLOv2 (You Only Look Once version 2) and LSTM (Long Short Term Memory). The YOLO architecture has time benefits as it processes the image only once. The LSTM architecture is used to incorporate temporal features of previous frames in the video. The paper suggests 2 different architectures PODv1 and PODv2 which differs in the feature mapping stage. The PODv2 gives better results.

**VIDEO STREAM ANALYSIS IN CLOUDS AN OBJECT DETECTION AND CLASSIFICATION FRAMEWORK FOR HIGH PERFORMANCE VIDEO ANALYTICS**

This paper proposes a method for video stream analysis using cloud computing for object detection and classification. The traditional methods are time consuming even after using GPU cores as the resolution of today’s cameras are increasing. This paper states that saving the input video data on the cloud, decoding the videos, and moving operator function code to the computing nodes on the cloud can give scalability and time benefits. The nodes on the cloud have GPU cores thereby collectively improving the time required for analysis of videos.