Project pre-proposal

Project title: Real-time object detection in videos
Team number: 8 / Dig Vijay Kumar Yarlagadda (Class ID: 26)

Project goal and Objectives: -

Motivation:
There are many excellent systems proposed for object recognition in images but very few of them are adaptable for use in videos. Due to recent advances in deep learning, accuracy has been greatly improved in detecting objects, facilitating use of these systems in real-time systems like autonomous vehicles, patient care etc.,

Significance:
Object detection in images is a fundamental and a long researched task to be performed before performing further analysis of content for applications like scene recognition, action recognition and prediction, face detection, translation etc., The challenge of detection in real-time is not being thoroughly researched, designing a real-time system which can achieve state-of-art accuracy in object detection, classification and segmentation can find usage in wide range of fields. It can act as a fundamental framework on which various applications can be built upon.

Uniqueness:
Previous works in object recognition in videos [1] [2]focused on using computer vision techniques and generic machine learning models for accomplishing state-of-art results. Deep learning based method are increasing being developed [3] [4] but most of them rely on vertical scaling of expensive GPUs. In our system, we plan to use an efficient system which can be leverage horizontal scaling capabilities (training still using GPUs and recognition using more efficient usage of resources) provided by frameworks such as Apache Spark, while still providing state-of-art performance.

Objectives:
- Adapt deep learning based object recognition frameworks such as Tensorflow Inception v4 [4] [7], Facebook DeepMask+SharpMask [5] [6] to videos.
- Designing a system that can perform the object recognition using above mentioned frameworks in real-time using Apache Spark, Storm and Kafka.

System Features:
- Input video fed to the system from rear camera of android device.
- The system will be pre-trained on training datasets and a model is generated. Based on available computing resources the video is divided into 30/60 frames/second and objects will be classified/detected and segmented.

Related work:

Clarifai[8] provides a visual recognition API which can detect objects and present them with confidence scores but it doesn't display the segmentation (separation) between objects. Further, it is a commercial API which cannot be built upon for application specific implementations. The system proposed in [3] uses a Deep Neural Network with five convolutional layers and R-CNN for labeling, but the framework we plan to use, Inception V4[4] uses RESNET[9], which is deeper and could provide better accuracy and performance results. Other systems such as [1], [10] are not generic, instead they are tailored towards specific applications.

Backup project:

Deep learning based real-time action recognition system:

A system that can detect actions such as jogging, boxing etc., in videos using Deep Networks. [11] [12]

Bibliography:

[1]. Shao, Mang, et al. "A comparative study of video-based object recognition from an egocentric viewpoint." *Neurocomputing* 171 (2016): 982-990.

[2]. Toshev, Alexander, Ameesh Makadia, and Kostas Daniilidis. "Shape-based object recognition in videos using 3D synthetic object models." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.

[3]. Wang, Li, and Dennis Sng. "Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey." *arXiv preprint arXiv:1512.03131*(2015).

[4]. Szegedy, Christian, Sergey Ioffe, and Vincent Vanhoucke. "Inception-v4, inception-resnet and the impact of residual connections on learning." *arXiv preprint arXiv:1602.07261* (2016).

[5]. Zagoruyko, Sergey, et al. "A MultiPath Network for Object Detection." *arXiv preprint arXiv:1604.02135* (2016).

[6]. https://code.facebook.com/posts/561187904071636

[7]. https://research.googleblog.com/

[8]. https://www.clarifai.com/

[9]. He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385* (2015).

[10]. Rahman, Junaedur. "Motion detection for video surveillance." (2008).

[11]. Wang, Pichao, et al. "Deep convolutional neural networks for action recognition using depth map sequences." *arXiv preprint arXiv:1501.04686*(2015).

[12]. Baccouche, Moez, et al. "Sequential deep learning for human action recognition." *International Workshop on Human Behavior Understanding*. Springer Berlin Heidelberg, 2011.

[13]. Karpathy, Andrej, et al. "Large-scale video classification with convolutional neural networks." *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.