CS6780: Digital Video Processing
Course Project

Divya K Raman
EE15B085

16th May 2019

1 Introduction

In our Deep Learning course project, we explored the problem of style transfer and experimented on cycleGAN in detail. We proposed and implemented a pipeline based on object detection to solve the object transfiguration failure cases(horse zebra case in particular). But what if the input is a video? The hue of the background changes when a horse is transformed to a zebra or vice versa and to combat this, the horse needs to be detected in every frame. Training a neural network, even for object detection, is a costly process. The number of frames per video can be really large and the number of distinct objects per video can also vary. We propose a method based on video segmentation for object style transfer in videos with diverse backgrounds.

2 Literature survey

Literature survey for style transfer, cycleGAN, semGAN and object detection was presented in the DL course project report. We present the literature survey for video segmentation in this report.

In [2], which is the most popular paper for background subtraction, the authors model each pixel as a mixture of Gaussians and update parameters as frames arrive using a simple update rule based on standard multivariate gaussian distribution equations. The learning rate needs to be tuned carefully for good results. We studied this paper in detail in assignment 1.

Conditional Random fields have been explored intensely for video segmentation. [3] proposes a new approach to learning a discriminative model of object classes, incorporating appearance, shape and context information efficiently. Unary classication and feature selection is achieved using shared boosting to give an efficient classifier which can be applied to a large number of classes.

The use of a Conditional Random Field allows them to incorporate shape, texture, color, location and edge cues in a single unified model and they incorporate all these three terms in the conditional probability of the class labels c given an image x. [4] uses a heirarchial CRF model for the same problem.

In [5], the authors propose an algorithm for spatio temporal segmentation. The authors describe a general framework for segmentation that identifies the spatiotemporal coherences of video data. In this framework, coherent motion regions are identied iteratively by generating hypotheses of motion then classifying each location of the image to one of the hypotheses. The paper details an implementation of the segmentation algorithm based on the affine motion model.

In [6], the authors propose an efficient algorithm that considers video segmentation and optical flow estimation simultaneously. For optical flow estimation, particularly at object boundaries, they compute the ow independently in the segmented regions and recompose the results. The process is called object ow and the authors demonstrate the effectiveness of jointly optimizing optical flow and video segmentation using an iterative scheme.

In [1], the authors describe a probabilistic approach to video segmentation using Gaussian mixture models. We describe the paper in detail in the following section. We have implemented this paper and tested its performance on various datasets.

3 Paper: Image Segmentation in Video Sequences - A Probabilistic Approach

We implement the paper 'Image Segmentation in Video Sequences: A Probabilistic Approach'. In this paper, the authors propose a method to obtain a mixture of Gaussians classification model for each pixel using a modified version of the incremental EM algorithm. The number of gaussians in the mixture of gaussians equals the number of labels we wish to segment the video into. Each gaussian predicts the probability of a pixel belonging to a particular label. To resolve storage constraints, the authors use an incremental EM method, the parameters of the Gaussian model are updated as frames(of the video) arrive. Initialisation is done using the first frame. As frames arrive, 'sufficient statistics' are updated and using these updated values, the parameters of the mixture of Gaussian model are updated. The equations are as in Page 4 and Page 5 of the paper.

implementation.ipynb has the basic implementation of the paper. We need to tune the initialisation, and set appropriate input and output files to generate results on various videos.

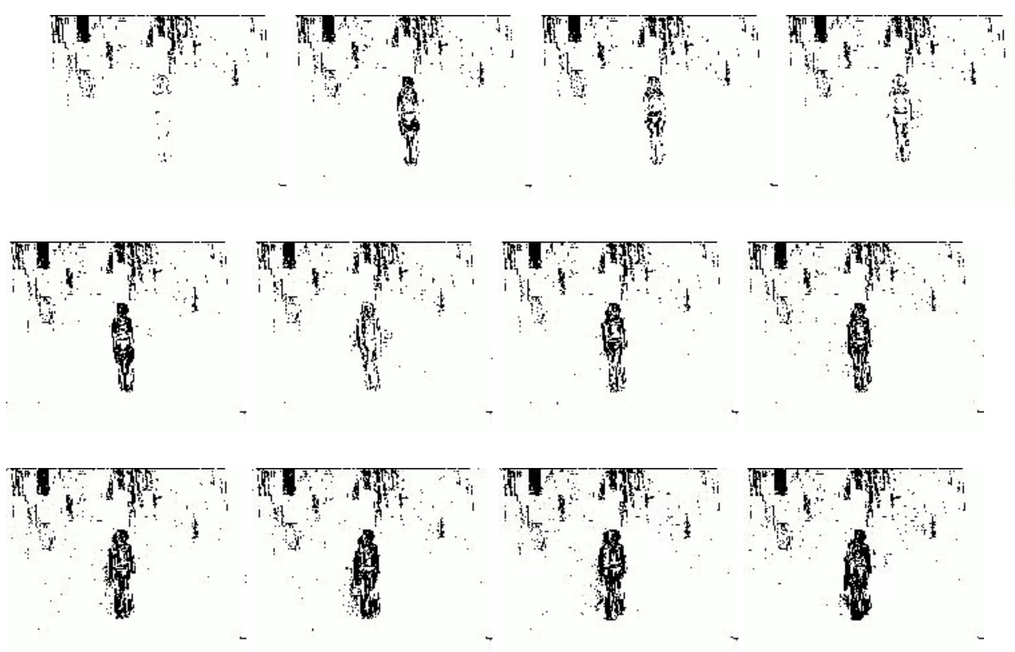
This method works much faster than the method described in [2]. Also, it is highly accurate. Good initialisation of parameters is important for best results.

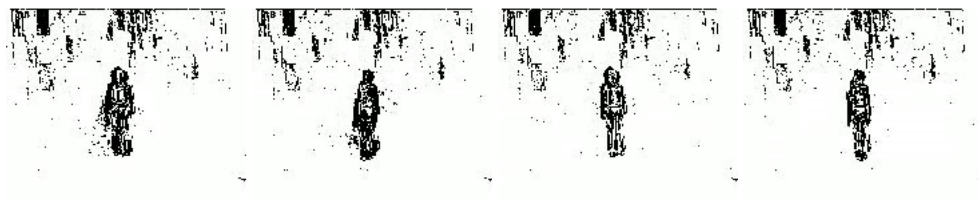
3.1 Why is this superior than the GMM algorithm?

Firstly, this algorithm is much faster. Secondly, the GMM algorithm works well when it is enough to detect moving objects in a video and segmentation is not necessary. However, for video cycleGAN, we need to segment out different objects in the video.

3.2 Results

jumpprobabilistic .ipynb runs the algorithm on the jump video. The background subtraction results are stored in jumpoutput.avi. A few sample frames are shown below.





runprobabilistic .ipynb runs the algorithm on the run video. The background subtraction results are stored in runoutput.avi. A few sample frames are shown below.

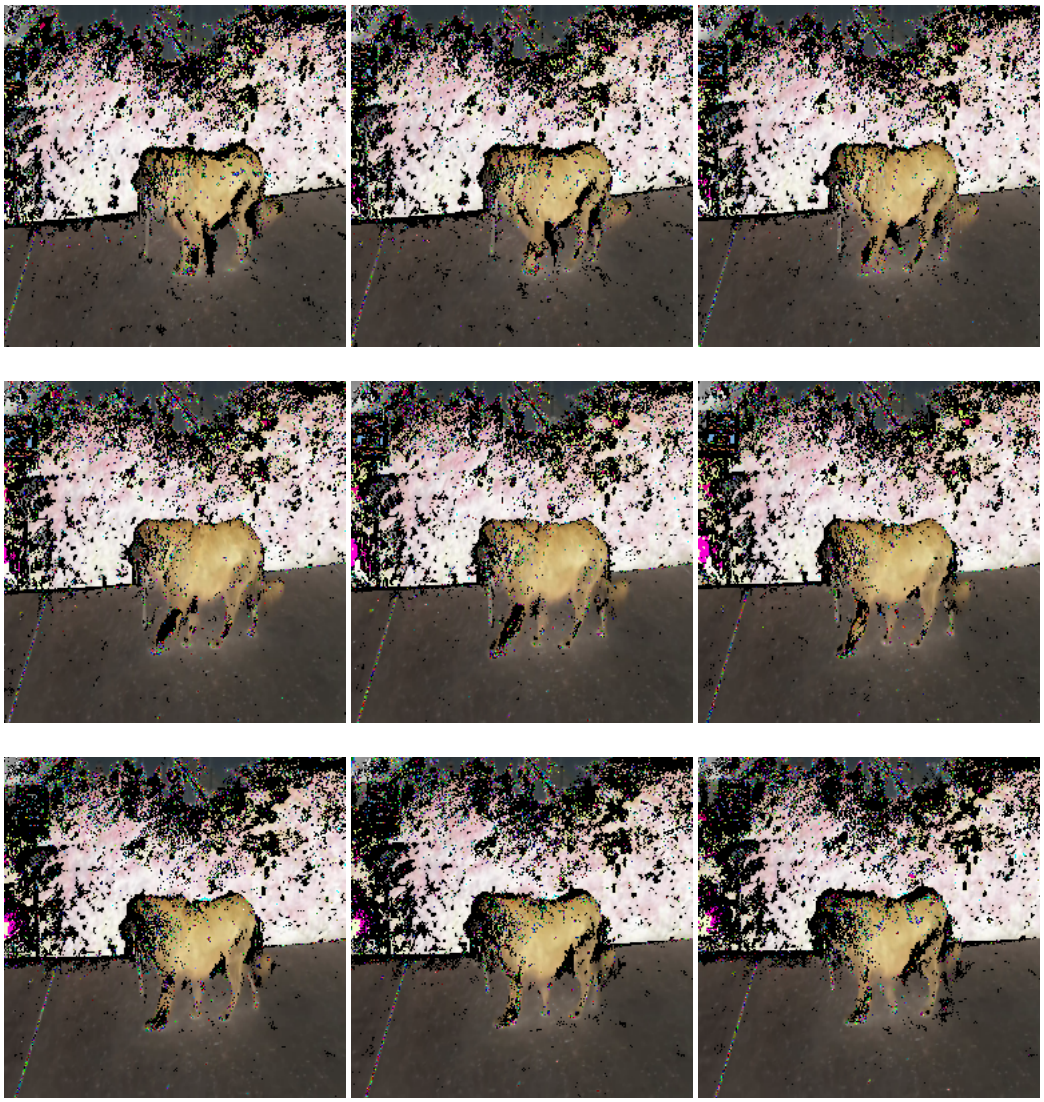
4 Horsifying videos

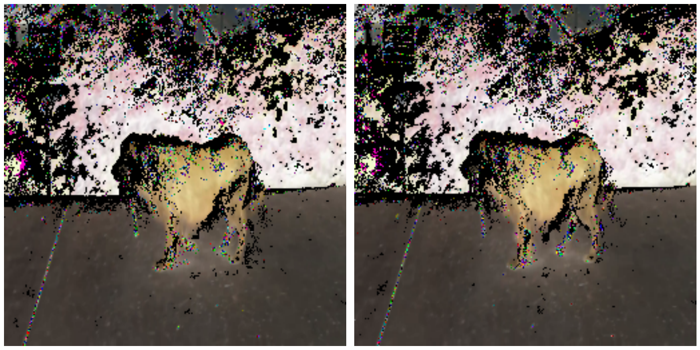
Steps:

1. Train the cycleGAN model on the horse-zebra image dataset.
2. Perform video segmentation on the input video to be horsified and obtain the foreground mask of the object to be horsified. For this, set the object pixels to 1 and all other pixels to 0.
3. Now, create a video using the foreground mask video such that all pixels other the object's pixels is set to 0.
4. Run cycleGAN on each frame of the video and stitch the transformed foreground with its background.

This gives us a horsified video whilst preserving the hue of the background as well as transforming only the object of interest.

4.1 Results





5 References

1. Friedman, N., Russell, S. (1997, August). Image segmentation in video sequences: A probabilistic approach. In Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence (pp. 175-181). Morgan Kaufmann Publishers Inc.

2. Stauffer, C., Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) (Vol. 2, pp. 246-252). IEEE.

3. Shotton, J., Winn, J., Rother, C., Criminisi, A. (2006, May). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In European conference on computer vision (pp. 1-15). Springer, Berlin, Heidelberg.

4. Russell, C., Kohli, P., Torr, P. H. (2009, September). Associative hierarchical crfs for object class image segmentation. In 2009 IEEE 12th International Conference on Computer Vision (pp. 739-746). IEEE.

5. Wang, J., Adelson, E. H. (1994, March). Spatio-temporal segmentation of video data. In Image and Video Processing II (Vol. 2182, pp. 120-132). International Society for Optics and Photonics.

6. Tsai, Y. H., Yang, M. H., Black, M. J. (2016). Video segmentation via object flow. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3899-3908).