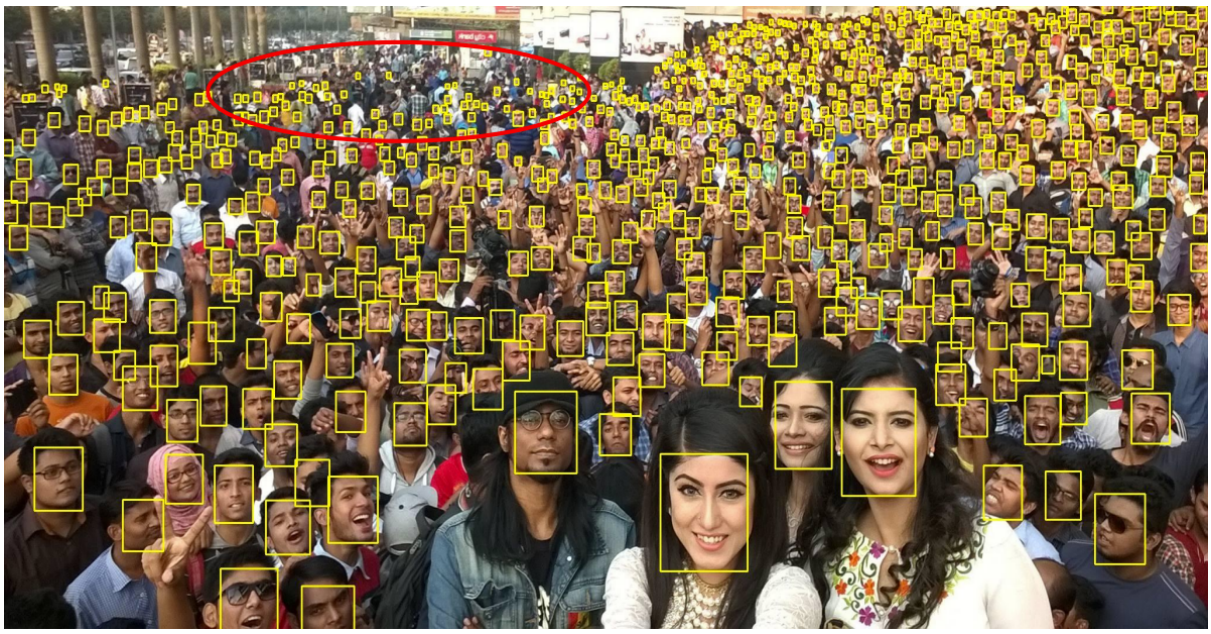
Enhancing Dense Crowd counting CNN's Report

Name	ID	Email
Rebecca Azoulay	342531837	rebecca.a@campus.technion.ac.il
Christian Shakkour	208157826	christian.s@campus.technion.ac.il



Contents

1 Introduction

With the rising demand for counting crowds among protests in ISRAEL and around the world, we aim to Enhance the crowd counting of CNN's by adding an input layer that will tile the raw image into NxN small images that will be enlarged and sharpened and feed to the network a bit like the inverse YOLO concept where each tile is feed forward. the crowd count in the end is the sum of all NxN counts. The datasets available are broad, we will be focusing on the most popular datasets: jhu-crowd-1, shanghaitech, ucf-cc-50-1. The two main CNN networks are CAT-CNN (Crowd Attention Convolutional Neural Network), and MT-CNN (Multi Column Convolutional Neural Network)

2 Related work

2.1 Detection-based crowd counting

Traditional detection-based methods often employ motion and appearance cues in video surveillance to detect each individual in a crowd [43, 5, 29]. They are suffered from heavy occlusions among people. Recent methods in the deep fashion learn person detectors relying on exhaustive bounding box annotations in the training images. For instance, the bounding boxes were manually annotated on partial of SHB and trained a Faster R-CNN for crowd counting. The annotation cost can be very expensive and sometimes impractical in very dense crowds. Our work instead uses only the point-level annotations to learn the detection model. There are some other works particularly focusing on small object detection, e.g. faces. A face detection method based on the proposal network helped detect and localize faces in a single stage detector like SSD. The face crowds tackled in these works are however way less denser than those in crowd counting works; moreover, these works are typically trained with bounding box annotations.

2.2 Regression-based crowd counting

Since the accuracy of detection-based method is not very high in the highly congested scene, researchers attempt to use regression-based methods to handle this problem. The regression-based methods learn a mapping between highlevel features and crowd counts. The high-level features are extracted from low-level information such as edge information, texture information, and segmentation information, then the crowd count is regressed according to high-level features. The Gaussian regression algorithm was proposed to learn a mapping between feature maps and crowd counts. The Poisson regression algorithm was employed to model the crowd count as the Poisson random variable. To decrease crowd counting errors. Some outstanding features were utilized such as people's heads to regress the crowd count. Although the regression-based methods can regress the crowd counts directly, the location information of each person is omitted.