**Project Topic:** Crowd Counting

**Team Members:**

Chandan Kumar Reddy Mannem

Monesa Thoguluva Janardhanan

Eliseo Garcia

Sai Krishna Uddagiri

Aparajitha Sriram

This project is more “business focussed” than “academically focussed”.

Questions our project addresses:

1. What problem will your Computer Vision solution solve, and for whom?

Crowding counting helps to gauge an estimate of how many people or objects are present in a particular place. Crowd counting is of importance in computer vision mainly due to its wide range of applications in marketing, video surveillance, public safety management, disaster management, etc.

1. What value will it provide them? What are their pain points?

Knowing the estimate of the number of people present in an area is of vital importance in the field of advertisements as it helps to get a count of an average number of eyeballs on each ad, like in a football match with tons of ads from sponsors. Another use case is in shops which will help them estimate their flow of customers and help in product placement.

It also plays a major role in disaster management and maintaining public safety like preventing overcrowding, stampedes and planning evacuation routes in case of emergencies.

Some of the main challenges with crowd counting are occlusion (blockage or obstruction of view), clutter, irregular object distribution and varying illumination.

1. How big is the potential market?

The people counting market is projected to touch USD 1100 million by 2022 according to a brand-new report by Market Research Future (MRFR) (source, [People Counting Systems Market to Expand Owing to Smart Crowd Management Solutions](https://www.globenewswire.com/news-release/2018/09/17/1571980/0/en/People-Counting-Systems-Market-to-Expand-Owing-to-Smart-Crowd-Management-Solutions-People-Counting-System-Market-Revenue-Surpass-USD-1100-Mn-by-2022.html)). It seems the market can expand at a CAGR (Compound annual growth rate) close to 14% based on their application in retail and transportation.

The Americas market is projected to grow well due to the increase in the number of malls in the US and Canada and more airports adopting people-counting systems recently.

The Asia Pacific region is expected to grow at a 16% CAGR by 2022 mainly due to the enormous increase in retail stores, supermarkets and shopping malls that form the potential customer base for the people counting market.

1. Do other similar solutions exist?

Many companies like ShopperTrak, RetailNext, V-Count, etc, provide solutions to count people using the live feed from surveillance video. However, our solution is to count people from static images. There are apps like countingthings and DotDotGoose that provide solutions for counting objects from static images.

1. Would your business have any competitors? Who are they? How are they doing?

Current key players in the people counting domain include companies like ShopperTrak, RetailNext who operate specifically in the retail domain for shopping malls, super markets and retail stores; V-Count, DILAX, Irisys, CountWise and FLIR’s include public spaces like transportation hubs, museums and domains like defence, security, marine in addition to retail analytics in shopping spaces. IRIS products specialize specifically in passenger counting. Hikvision and Axis Communications specialize in security and video surveillance.

Apps like countingthings and DotDotGoose provide solutions for counting objects from static images.

1. How are potential customers dealing with these issues now?

There are businesses who use a ticketing system for events which would take care of tallying the counts. Some businesses still employ manual methods which involve an employee to manually jot down counts as people enter a venue for an event.

1. How much would a customer be willing to pay for your product?

Products that come embedded with cameras and sensors for object detection come in prices ranging from $250 to $900 based on their sophisticated tech. Most of these products deal with real time feed from surveillance cameras unlike our idea to use static images to count the number of people. There are apps that count objects from images which are priced at around $100 per month licenses. So, our product could have a potential value of about the same range.

1. Are your customers individuals or businesses? Be sure to cite the sources you reference in your research.

Customers for people counting products mainly seem to be businesses as its application lies in tracking and estimating the count of people in office or public spaces, in events and in modes of transportation. The companies like V-Count, ShopperTrak, RetailNext, etc, are some of the key players in this field and most of their customer base involves businesses who want to keep track of their customers.

1. What academic work is relevant to your project topic? Pick 3 papers, ask us for help if you need it.

We found 3 papers which are

“An evaluation of crowd counting methods, features and regression models” - <https://www.sciencedirect.com/science/article/abs/pii/S1077314214001611>,

“Deep Spatial Regression Model for Image Crowd Counting” - <https://arxiv.org/ftp/arxiv/papers/1710/1710.09757.pdf>,

“Counting Crowd with Fully Convolutional Networks” - <https://ieeexplore.ieee.org/document/8221101>.

The following 3 were suggested by Aravind, the TA:

“You Only Look Once: Unified, Real-Time Object Detection” - <https://arxiv.org/abs/1506.02640>,

“SSD: Single Shot MultiBox Detector” - <https://arxiv.org/abs/1512.02325>

1. What makes these papers important/relevant?

Each of these papers explore the methods to count people effectively. The paper on the evaluation of different crowd counting methods gives a broad perspective of various methods that are usually used for people counting and the other papers explore the possible ideas of using deep learning and neural networks for solving this problem.

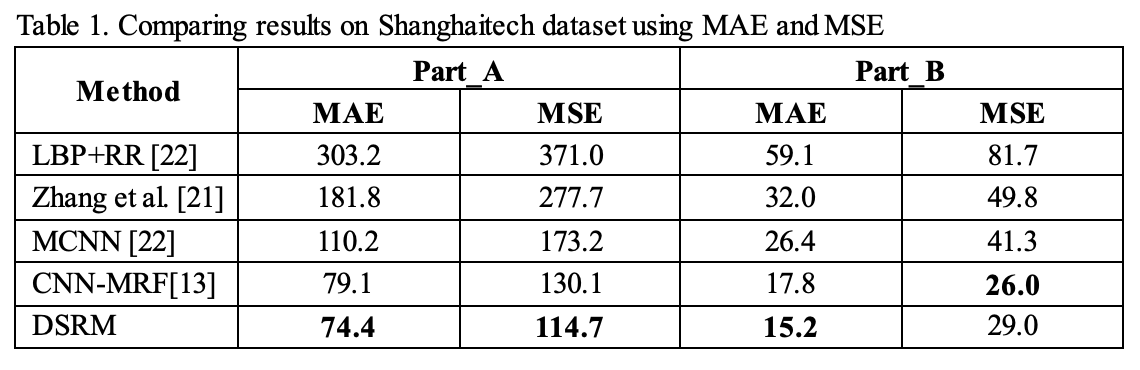
**Literature review:**

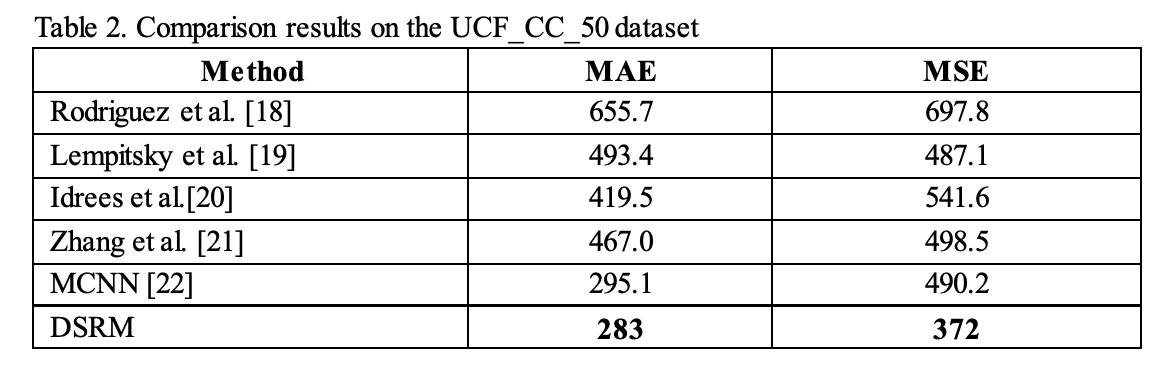
Reference [6] **Link:** <https://arxiv.org/pdf/1710.09757.pdf>

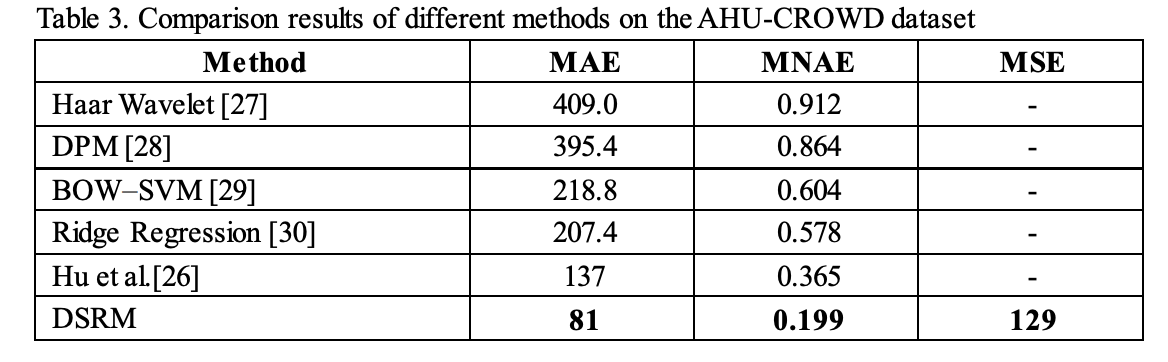
**Deep Spatial Regression Model for Image Crowd Counting**

**Datasets** used for this model Shanghaitech dataset, UCF\_CC\_50 dataset, AHU-CROWD dataset.

The proposed model in this paper is based on Convolutional Neural Network (CNN) and long short term memory (LSTM). First, they put the images into a pretrained CNN to extract a set of high-level features. Then the features in adjacent regions are used to regress the local counts with a LSTM structure which takes the spatial information into consideration. The final global count is obtained by a sum of the local patches







The above result shows the DSRM model that they used is predicting results with more accuracy when compared to other traditional models.

Reference [7] **Link:** <https://ieeexplore.ieee.org/document/8221101>

**Crowd Counting with Fully Convolutional Networks**

**Datasets: WorldExpo’10 crowd counting dataset**

This paper proposed a fully Convolutional Neural Network model (FCN). This paper talks about the different algorithms in crowd counting such as detection based counting, global regression based methods and density estimation that counts by integrating density map. This paper demonstrates about the obtaining the head count by integrating the density map of the crowd. They are obtaining the density ground truth by creating a sum of Guassian kernel and the center of the head. Coming to the Crowd FCN Model, it is composed of two parts one is the down sampling network and the other the up sampling network. The activation function that have used for this Neural Network Is ReLU. Compared to the above discussed methods, this paper indicates that the results are better than the state-of-the-art crowd counting methods.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Method* | *Scene 1* | *Scene 2* | *Scene 3* | *Scene 4* | *Scene 5* | *Average* |
| LBP + RR [9] | 13.6 | 59.8 | 37.1 | 21.8 | 23.4 | 31.0 |
| Crowd CNN [9] | 10.0 | 15.4 | 15.3 | 25.6 | 4.1 | 14.1 |
| Fine-tuned Crowd CNN[9] | 9.8 | 14.1 | 14.3 | 22.2 | 3.7 | 12.9 |
| MCNN [10] | 3.4 | 20.6 | 12.9 | 13.0 | 8.1 | 11.6 |
| Our crowd FCN model | 3.5 | 14.3 | 16.0 | 15.2 | 3.9 | 10.6 |

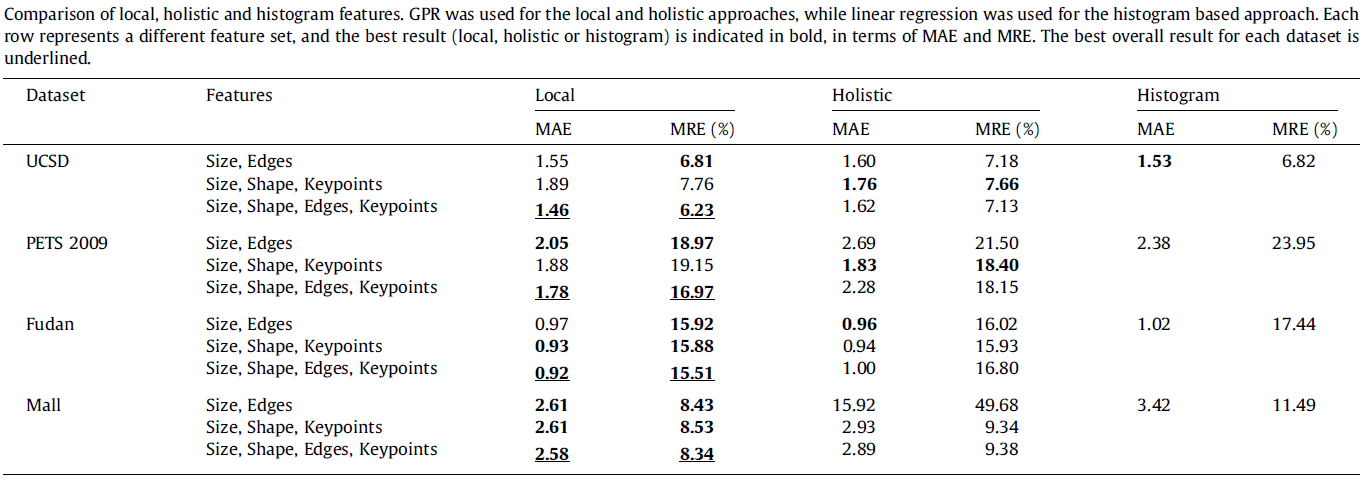
Reference [8] **Link:**

<https://www.sciencedirect.com/science/article/abs/pii/S1077314214001611>

**An evaluation of crowd counting methods, features and regression models**

**Datasets: UCSD, PETS 2009, Fudan, Mall and Grand Central datasets**

This research paper presents an evaluation of different approaches to crowd counting. This paper compares holistic, local and histogram based approaches across five public datasets and contrasts it with regression models. The regression models evaluated were Gaussian Process Regression (GPR), linear regression, K Nearest Neighbour (KNN) and Neural Network (NN). The authors found that local methods outperformed holistic and histogram based approaches and GPR outperformed KNN, NN and linear regression.



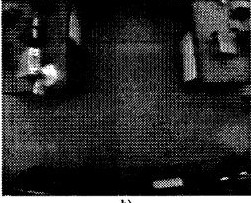
**Article - A system for counting people in video images using neural networks to identify the background scene**

[URL - https://www.sciencedirect.com/science/article/pii/0031320395001638](http://www.sciencedirect.com/science/article/pii/0031320395001638)

The article describes a system of counting the amount of visible people from images. The environment shown in the image contained objects such as doors and environmental features such as a reflective floor. The article describes a system that requires inexpensive hardware and tools and uses images as the main input. This summary explains the pre-processing method mentioned in this article.

The system used images as the primary source of input for processing. It is described that images were captured using a CCD video camera and then analysed to determine the number of people present” (Schofield 1421). Figure 1 shows an example of the images used. The background of the images used were meant to be varied in order to assess performance in different environments. This was done by “including variations in lighting levels and patterns, and the occasional movement of objects, such as doors, that might appear in the scene” (Schofield 1421). One of the key problems in dealing with crowd detection mentioned is that people look different from different angles and have different physical features, but objects are most likely to be consistent in shape (Schofield 1422). This would prove important as an image could be processed to remove a large portion of the background and simplify the crowd counting process. The process used was split into three separate stages.

Figure 1 - “Crowd Counting” (Schofield 1423).



The first stage, or pre-processing stage, was mostly concerned with getting the right thresholds and filters onto an image. While high-resolution images might provide more detail, low-resolution images were found to be sufficient for crowd counting. The experiment used (362x268) medium-resolution images for developing an algorithm, but later discovered that it performed well with (100x100) or lower-resolution images as well (Schofield 1423). In order to increase processing speed, sub-sampling was used to decrease the size and pixel count of images during this stage. It was described that “thresholding is a process that requires some care because it removes much information from the image… A simple fixed threshold at (say) grey level 128 was not considered appropriate since this method takes on account of variations in ambient lighting levels” (Schofield 1423). A threshold that was adaptive to light changes in the form of a “function of the mean gray level” of the greyscale image was described as not being able to deal with small unique lighting changes, but did do well with global light changes (Schofield 1423).

In order to deal with these non-uniform changes in light intensity, an “adaptive local threshold” was used. This was accomplished with “the grey level of each pixel… compared with the average grey level of the pixels in its immediate neighbourhood” (Schofield 1424). This can be compared to basic edge detection where a small patch of an image is taken and its local values are compared to each other. A problem with this approach is that it can create a lot of noise. The major problem found with this approach was that it did not remove object edges, which would cause the gaps between two people to be counted as one person due to it interfering with finding the center point of people (Schofield 1424).

The second stage was the background identification stage. In order to process images more, a background removal method was used. It used RAM-based neural networks due to having advantages over neural network architectures. (Schofield 1424).They can be “trained using examples of a single training class and they are fully trained after a single pass through the training set” (Schofield 1424). The neural network was only trained on images with no people and only the background. The reason for this was so that it could detect similar backgrounds and remove them from the images it was given. Using this approach, “the system was able to discriminate between parts of the background scene and non-background objects. This is important since it would not have been practi- cally possible to train a neural network on a represen- tative set of all non-background scenes” (Schofield 1424).

This can be useful for our Lowe’s crowd counting system as we will be able to detect customers in an environment where people may be partially blocked by store features or lighting. Since the store layout is kept consistent in terms of inventory placement, we can try to subtract the background of an image in order to remove some of it. This can improve processing time, and leads to less image pixels having to be processed. It can also help reduce errors from reflective surfaces or objects that have similar physical characteristics, such as size, to people.

References:

1. Convolutional-Neural Network-Based Image Crowd Counting: Review, Categorization, Analysis, and Performance Evaluation, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6983207/>
2. People Counting Systems Market to Expand Owing to Smart Crowd Management Solutions, <https://www.globenewswire.com/news-release/2018/09/17/1571980/0/en/People-Counting-Systems-Market-to-Expand-Owing-to-Smart-Crowd-Management-Solutions-People-Counting-System-Market-Revenue-Surpass-USD-1100-Mn-by-2022.html>
3. People Counting System Market, <https://www.marketsandmarkets.com/Market-Reports/people-counting-system-market-206905443.html>
4. People Counting System Market Size By Technology, <https://www.gminsights.com/industry-analysis/people-counting-system-market>
5. People Counting System Market Share, Size 2020 Industry Overview, Key Players Analysis, Emerging Opportunities, Comprehensive Research Study, Competitive Landscape and Potential of Industry from 2020-2026, <https://www.marketwatch.com/press-release/people-counting-system-market-share-size-2020-industry-overview-key-players-analysis-emerging-opportunities-comprehensive-research-study-competitive-landscape-and-potential-of-industry-from-2020-2026-2020-09-15>
6. Deep Spatial Regression Model for Image Crowd Counting, <https://arxiv.org/ftp/arxiv/papers/1710/1710.09757.pdf>
7. Counting Crowd with Fully Convolutional Networks, <https://ieeexplore.ieee.org/document/8221101>
8. An evaluation of crowd counting methods, features and regression models, <https://www.sciencedirect.com/science/article/abs/pii/S1077314214001611>
9. You Only Look Once: Unified, Real-Time Object Detection, <https://arxiv.org/abs/1506.02640>
10. SSD: Single Shot MultiBox Detector, <https://arxiv.org/abs/1512.02325>
11. Schofield, A.J., et al. *A System for Counting People in Video Images Using Neural Networks to Identify the Background Scene*. 7 June 2001, [www.sciencedirect.com/science/article/pii/0031320395001638.](http://www.sciencedirect.com/science/article/pii/0031320395001638.)