# Using Convolution Neural Network for Crowd Density Estimation for The Holy Masjed

Bandar Alkhuzaim(1), Abdulrahman Altahhan(2)
(1) The General Presidency of the Holy Masjed and Prophet Masjed Affairs, (2) Leeds Beckett University

# تقدير كثافة الحشود بالمسجد الحرام باستخدام الشبكة العصبية الإلتفافية

بندر بن محمد الخزيم(١) عبدالرحمن الطحان (٢)
(١) الرئاسة العامة لشؤون المسجد الحرام و المسجد النبوي
(٢) جامعة ليدزبكت

**ملخص البحث (Abstract):**

مع الزيادة الهائلة في عدد سكان العالم ، هناك زيادة مباشرة في عدد المشاركين في المناسبات المختلفة ، سواء كانت رياضية أو اجتماعية أو دينية. وهذا يتطلب رعاية أكبر لسلامة وأمن أولئك المتواجدين في مثل هذه الأحداث. هناك عدد من الجوانب التي يجب مراعاتها عند تنظيم وإدارة الأحداث المختلفة من أجل ضمان سلامة الحاضرين مثل تقدير كثافة الحشد. يعد تقدير كثافة الحشود مكونًا مهمًا للمراقبة المرئية ويلعب دورًا رئيسيًا في مراقبة وإدارة الحشود. نظرًا لأهميته ، فقد تم إجراء الكثير من الأبحاث لتقدير كثافة الحشود في مشاهد المراقبة المرئية المختلفة. في هذه الدراسة ، تم بناء وتدريب شبكة شبكة عصبية التفافية (convolutional neural network) لتقدير كثافة الحشود للمسجد الحرام في مكة المكرمة والتي تعتبر من بين الأماكن الدينية التي تجذب أعداداً هائلة من الناس من جميع أنحاء العالم باستخدام مجموعة من البيانات التي قدمتها الرئاسة العامة لشؤون المسجد الحرام و المسجد النبوي . على الرغم من أن دقة تقدير كثافة الحشود التي تم الحصول عليها من الشبكة العصبية التلافيفية المدربة ليست عالية بما يكفي (٧٠%) لتكون موثوقًا بها في مكان مثل المسجد الحرام ولكن يمكن اعتبارها نتيجة مرضية مقارنة بعدد البيانات التي تم العمل عليها. وتعتبر هذه النتيجة مؤشرا جيدا على أنه يمكن أن تكون النتائج أكثر دقة إذا تم العمل على حجم بيانات اكبر.

With the huge increase in the population of the globe, there is a direct increase in the number of participants and attendance of various events, whether sports, social or religious. This requires greater care for the safety and security of those present in such events. There is a number of aspects that must be taken into account when organize and manage the various events in order to ensure the safety of those present such as crowd density estimation. Crowd density estimation is an important component of visual surveillance and plays a key role in crowd monitoring and management. Because of its importance, much research has been done to estimate the density of crowds in visual surveillance scenes. In this study, a convolutional neural network for crowd density estimation for the Holy Masjed in Makkah which considered among the religious places that attract the huge number of people across the world has been built and trained by using images provided by The General Presidency of the Holy Masjed and Prophet Masjed Affairs (GPH). Although the accuracy of crowd density estimation that obtained from the trained convolutional neural network is not high enough (70%) to be trusted in a place like the Holy Masjed but it can be considered as a satisfactory result

compared to the number of data that have been worked on. This result is considered a good indicator that it could get more accurate if the convolutional neural network trained on large images.

## Introduction

Over the past, there have been controversies on what should be considered a mass gathering. Kollek (2014) defines mass gathering as a group of 1,000 or more people with the same intentions present in a particular location over a given time. The World Health Organization, however, describes mass gathering as "an event attended by a large number of people thereby interfering with the resource planning and response of a community, state, or nation" (WHO, 2008, p.14). Examples of such gatherings include social, religious, and political events. According to Al-Tawfiq and Memish (2014), mass gatherings can be classified into two categories namely planned gatherings and spontaneous gatherings. Planned mass gatherings include various sporting, political, socio-cultural, and religious events such as music festivals, the Olympic Games, and the Hajj in Makkah (Yezli and Alotaibi 2016).

Holy Masjed in Makkah is considered among the religious places that attract the huge number of people across the world. According to Adherents (2014), the world Muslim pollution is approximately 1.5 billion people. All Muslims aspire to visit the Holy Masjed to perform religious rituals. According to The General Authority for Statistics (2016), the number of Muslim faithful who visited the Holy Masjed during the Hajj month in 2016 were more than 2,000,000 people. Given the large numbers of people visiting these sites, there is a need to direct more efforts to the control and management of the crowd's movements in order to ensure security. A study by Alzhrani (2017) reported various crowd management challenges at the Holy Masjed that negatively affects the level of security and safety at the mass gatherings. Alzhrani (2017) classified the challenges into broad categories namely challenges based on the nature of crowds, challenges of high density, and challenges of the rituals of Islam.

The challenges based on the nature of crowds are due to the fact that Makkah visitors differ in terms of language, ethnicity, socio-cultural environments, and the individual mental abilities. High-density challenges are based on the argument that the Makkah pilgrimage is denser compared to other world events. Islam ritual challenges, however, are due to the fact that the rituals are only performed at the designated places and at specified times of the year/month.

According to Owaidah (2014), crowd management and control incidents have led to injuries and deaths to the visitors. Besides, the incidents have led to people being lost. The incidents are associated with stampedes, overcrowding, and human bottlenecks Table 1.1 shows a summary of the incidents, their causes, year of occurrence, and the number of casualties (Owaidah 2014, P16).

Nemade and Gohokar (2016) define crowd density as the number of people present in each unit. However, it does not count each individual for density estimation purposes. Crowd density estimation plays a key role in visual surveillance, crowd monitoring, and management. A study by Nemade and Gohokar (2016) grouped crowd density into five classes namely jammed flow, very dense flow, dense flow, restricted flow, and free flow. On the other hand, Aziz et al. (2017) classified crowd density into four categories: jammed, dense, medium, and low. The two studies discussed the various techniques that can be utilized for estimating crowd density using the computer vision. The techniques include pixel counting, object dimension identification, map-based estimation, and texture-based estimation.

Image classification forms the central idea behind crowd density estimation. Convolutional neural networks are essential for addressing issues of image classification especially due to the development of deep learning. Bhandare et al. (2016) describe the most significant characteristics of convolutional networks. They argue that convolutional neural networks image classification can help in performing image classification without feature extraction and background.

In a move to evaluate devices in CNN-based single image crowd counting and density estimation, Sindagi and Patel (2017) compared the crowd density accuracy using different datasets. Analysis of the datasets indicates that crowd density estimation using CNN-based techniques provide more accurate results compared to the traditional approaches. Some of the most commonly used datasets include UCF CC 50, WorldExpo10, UCSD, Shanghai Tech-A, Mall, and Shanghai Tech-B.

## Significance of the Project:

Although numerous studies have been crowd density classification and estimation, there is limited empirical literature on the Kingdom of Saudi Arabia. Thus, no data on crowd density estimation utilizing a convolutional neural network at the Holy Masjed in Makkah exist. Given the research gap, this project will not only help in designing and implementing a convolutional neural network necessary for crowd density estimation at the Holy Masjed but also contributes by creating new dataset from the CCTV of the Holy Masjed in testing and training the convolutional neural network.

## Research Aims

The main aim of this study is to research, design and develop a convolutional neural network for crowd density estimation at the Holy Masjed at Makkah in order to provide guidance for pilgrims and visitors to increase safety awareness.

## Research Methodology

Cross-industry standard process for data mining (CRISP-DM)  is used in this research. It is a standard model that helps in describing the common approaches that the data mining experts apply (Sastry and Babu, 2018). It comprises of six steps that include business understanding, modelling, data-preparation, data-understanding deployment and evaluation (figure 1) . This process model can easily be customized and is flexible.

### Convolutional Neural Network

The paper tackles a classification problem and building a convolutional neural network (CNN) will assist in addressing this issue. In computer vision, one of the most appropriate state-of-the-art tool is CNN and hence will be applied in this project. These networks use perceptron's which are a machine learning unit for data analysis in supervised learning. CNN finds common applications in tasks such as image processing and other cognitive tasks such as natural language learning. CNN represents a complex example of deep learning comparable to information processing by the human brain (Krizhevsky, Sutskever & Geoffrey, 2012).

The convolutional networks are made up of three main type of layers. These are: convolutional layers, pooling layers and a fully connected layer . The first step to do to use a convolutional neural network in image classification scenariois to have the images in a form that the computer can accept and act on. Unlike the human brain that can look and identify images as they are, acomputer can only understand images in the form of a a set of numbers that each represents the intensity of a colour in specific pixel in the image. The computer assigns an image pixel values. For example, if a picture is 30*30 and of full colour, then the computer will assign it an array of pixel values 30*30*3. Then it assigns a value from 0 to 255 to each of these numbers to describe the intensity of the pixel at each point.

**The Dataset:**

The GPH is where the data has been acquired from . Three folders has been used in saving the datasets and they include training, validating and testing datasets that also contain three sub-folders each and encompass high, low and medium densities density.

**Model Architecture**

Figure 2 shows a part of the summary of the convolutional neutral network while illustrating the output shape of each layer in the model architecture. For instance, the first layer is a convolutional layer, keras which has automatically created unique names for all layers, in this case, the convolutional layer is called conv2d_1, and it has an output of 99 ×99 and 16 feature maps. This convolutional layer needs to train and optimize 208 deferent parameters including weights and biases inside this layer.

**Results and discussion**

Figure 3 shows the accuracy for both training and validation dataset for every epoch. Both of the training and validation accuracies is around 61%. This indicates that the model could be trained a little more to increase the accuracy, which is important since the accuracy on both datasets is still going up for the last instances. But, when the model has been trained more to increase the accuracy it faced an overfitting problem.

Figure 4 Shows the accuracy and loss for the testing data set. The figure presents the number of testing images as 210 images and the evaluation loss over never before seen images is 0.9061 while the evaluation accuracy over the never seen before images is 75%.

Although different convolutional neural network models were trained to estimate the crowd density for the Holy Masjed, most of them faced overfitting or underfitting problem. The best model accuracy is 75%. This percentage is considered low accuracy in crowd density estimation in general and particularly for crowd density estimation for the Holy Masjed. This percentage is not ideal for The Holy Masjed because of various crowd management challenges at the Holy Masjed that negatively affects the level of security and safety at the mass gatherings.

The main reason for low accuracy that obtained from the convolutional neural network which were built and trained is the number of images (the total number of images is just around 1000 images for training, validation and testing). The numbers of images or data inputs in a convolutional neural network are very important in determining the kind of output or performance expected of a network. In training a convolutional neural network, in most instances it is vitally important to have as many images or data inputs as possible. The number of images or data required will depend on; the complexity in the differences of the data one intends to separate. For example, if the images has to be classified based on colour black and white, then the images required in training might be just a few examples. However, if the images contain numerous other variations within the individual classes, training images might have to be as many as possible per class (Krizhevsky, Sutskever and Hinton, 2012). In instances where very thorough classifications are required and the differences in classes are not that separable, then the neural network needs to have a higher number of iterations on the individual classes to try and make it responsive more accurately and identify the correct data class and this is the case in the Holy Masjed, especially in the area of Tawaf. It is paramount that there are sufficient images in each category so that the system has a larger array of training images to learn from before classifying the respective images in their right categories. Otherwise the network might experience difficulties in trying to fit individual images into their correct categories. The more the data exposed to the system the more likely the system is to pick inputs and relate them to others already exposed to it in the training and hence making it able to give more accurate results.

In fact, CNNs form one of the most popular deep learning architecture in the world due to their effectiveness in classifying images and it can be used effectively at estimating crowd density. Crowd monitoring and control of public places have become a very demanding endeavor for the past five years (Szegedy et al., 2015). The assortment of human actions, such as religious gatherings enforces crowded scenes to be common in almost all corners of the globe. For this reason, enormous challenges to crowd management, such as analysis, monitoring, identification, and detection of suspicious activities occur. Since the introduction of AlexNet in 2012, the popularity of CNNs have drastically up surged (Simonyan & Zisserman, 2014; Szegedy et al., 2015). In this regard, the convolutional neural network model is now applied to solve almost every image related issue due to its high accuracy and the ability to be successfully used in real-world scenarios. The traditional approaches for dealing with crowd have proven to be ineffective due to severe occlusions and cluttering (Szegedy et al., 2015). One of the major competitive advantages of CNN is its ability to automatically detect significant features in a particular gathering without any human supervision (Szegedy et al., 2015). Moreover, CNN model can be operated on any device due to its ability to apply special convolution and pooling operations and device sharing (Tripathi et al., 2018; Yang et al., 2014). Thus, given by the fact that the CNN model is universally attractive and can be able to achieve superhuman accuracy, it can be applied to monitor and analyze the behavior of the crowd in the Holy Masjed.

In order to ensure peaceful religious event gathering at the Holy Masjed, the planners can rely on CNN model to conduct crowd behavior analysis (Gu et al., 2018). Over the past ten years, organizers at the Holy Masjed relied on traditional approaches to analyze the religious gathering to prevent possible commotion. However, these methods were not effective due to the nonlinearity of real-world images and videos (Tripathi et al., 2018). Although human observers can be used to monitor unusual group activities, the process cannot be possible when confronted with a huge amount of video and image data and this is the case in the Holy Masjed. In this regard, both traditional approaches and human observers cannot be relied upon to monitor group activities (Gu et al., 2018). In this regard, the CNN model possesses the ability to understand and interpret complex scenes and the changes in a particular group. Hence, given by the fact that pepole at the Holy Masjed can exceed the estimated number, CNN model can be used to access the situation and control the influx of the people (Thoma, 2017).

In a religious gathering, such as the pilgrim at the Holy Masjed, the time factor is critical since every movement takes place in real-time. Thus, in case the observers delay to monitor the situation for seconds, it could lead to calamity. However, Convolutional Neural Networks model can be successfully applied to analyze real-time videos from the multitude. In this case, 3D CNNs model uses time as a third dimension in its analysis (Gu et al., 2018).

Contrary to the conventional methods, CNNs model has the ability to monitor and classify images and videos in densely crowded scenes which are extremely cluttered and have severe ambiguities. In this regard, CNNs model can be useful in controlling and classifying religious crowd at the Holy Masjed. Moreover, obtaining a clear idea of the crowd behaviors at the Holy Masjed without understanding the actions of each individual can be very instrumental in managing the crowd to avoid possible commotion. In this regard, due to the fact that the crowds at the Holy Masjed engage in various activities, it may not be possible to detect and extract the different patterns using traditional approaches, such as human observation. However, the CNN model can be used to extract and finally classify the details obtained from the crowd to different categories. The neuron in the last fully connected layer within the CNN system are tasked with the classification of the information obtained while the extraction of the details is done by the convolution layers.

**Summary and conclusion:**

The major conclusions can be summarised as follows:

- CNNs form one of the most popular deep learning architecture in the world due to their effectiveness in classifying images and it can be used effectively at estimating crowd density.
- The accomplishment of CNNs in various computer vision operations led to the development of different CNN-based tactics for crowd approximation and counting.
- The results of all the reviewed literature indicated that CNN-based approaches and other techniques that are currently available can achieve enhanced performance in approximating crowd density.
- The numbers of images or data inputs in a convolutional neural network are very important in determining the kind of output or performance expected of a network.
- CNNs model has the ability to monitor and classify images and videos in densely crowded scenes which are extremely cluttered and have severe ambiguities.

**Recommendations:**

1. One of the limitations with this project is the number data samples. Therefore, one potential for future work is to increase the number of data samples processed during the learning stage for the Holy Masjed.
2. When the model obtained high accuracy the GPH can deploy the system as shown in Figure 5.
3. Further study for crowd density estimation for the Holy Masjed by counting the detected objects ( counting the head) may provide more accuracy.

Table 1 (Owaidah 2014, P16(

<table>
<tr><th>Date</th><th>Accidents</th><th>Casualties</th></tr>
<tr><td>1957</td><td>Fire</td><td>200 pilgrims died</td></tr>
<tr><td>1990</td><td>Suffocation</td><td>1426 pilgrims died</td></tr>
<tr><td>1994</td><td rowspan="7">Stampede</td><td>270 pilgrims died and 180 injured</td></tr>
<tr><td>1998</td><td>35 pilgrims died</td></tr>
<tr><td>2001</td><td>14 pilgrims died</td></tr>
<tr><td>2003</td><td>251 pilgrims died</td></tr>
<tr><td>2004</td><td>364 pilgrims died and 244 injured</td></tr>
<tr><td>2006</td><td>200 pilgrims died and 289 injured</td></tr>
</table>

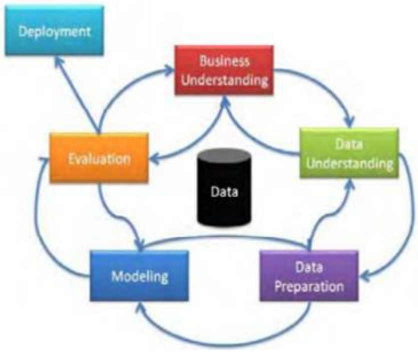


Figure 1 CRISP-DM Methodology Steps (Sastry and Babu, 2018)

File - unknown

```
C:\Users\Mohamed\AppData\Local\Programs\Python\Python36\
python.exe "C:/Users/Mohamed/Desktop/New folder/untitled1/
HolyMosque1.py"
Using TensorFlow backend.
_________________________________________________________________
Layer (type)                 Output Shape
Param #
=================================================================
conv2d_1 (Conv2D)            (None, 99, 99, 16)        208
_________________________________________________________________
activation_1 (Activation)    (None, 99, 99, 16)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 98, 98, 16)
1040
_________________________________________________________________
activation_2 (Activation)    (None, 98, 98, 16)        0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 49, 49, 16)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 49, 49, 16)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 48, 48, 16)
1040
_________________________________________________________________
activation_3 (Activation)    (None, 48, 48, 16)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 47, 47, 16)
1040
_________________________________________________________________
```

Figure 2

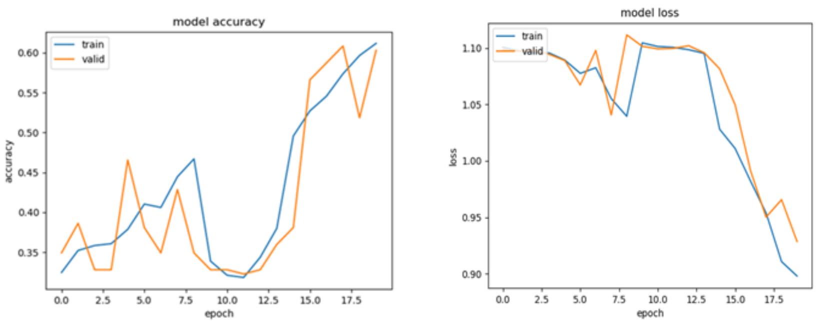


Figure 3

```
Found 210 images belonging to 3 classes.
Evaluation Loss over never befor seen images is: 0.9061
Evaluation Acuracy over never befor seen images is: 75.00
```

Figure 4

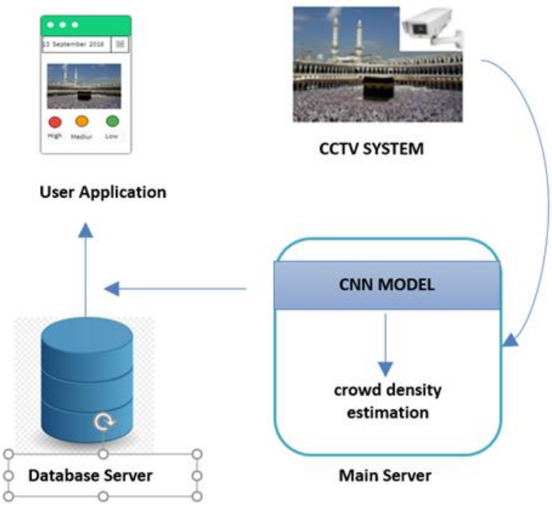


Figure 5

**References**:


- Adherents (2018). Major Religions Ranked by Size. [online] Adherents.com. Available at: http://www.adherents.com/Religions_By_Adherents.html [Accessed 13 Apr. 2018].
- Al-Tawfiq, J. and Memish, Z. (2014). Mass gathering medicine: 2014 Hajj and Umra preparation as a leading example. International Journal of Infectious Diseases, [online] 27, pp.26-31. Available at: https://www.sciencedirect.com/science/article/pii/S120197121401580X [Accessed 5 Apr. 2018].
- Alzhrani, M. (2017). جهود المملكة العربية السعودية في إدارة الحشود في موسم الحج..
- 6. Aziz, M., Naeem, F., Alizai, M. and Khan, K. (2017). Automated Solutions for Crowd Size Estimation. Social Science Computer Review, [online] 36(5), pp.610-631. Available at: http://journals.sagepub.com/doi/abs/10.1177/0894439317726510 [Accessed 16 Apr. 2018].
- Bhandare, A., Bhide, M., Gokhale, P. and Chandavarkar, R. (2016). applications of convolutional neural networks. International Journal of Computer Science and Information Technologies, [online] pp.2206-2215. Available at: https://pdfs.semanticscholar.org/89db/184cb8b1397a9aa35693f8dc03e1e728992a.pdf [Accessed 16 Apr. 2018].
- General Authority for Statistics (2016). Hajj Statistics for 1437H - 2016. [online] General Authority for Statistics. Available at: https://www.stats.gov.sa/en/4489 [Accessed 8 Apr. 2018].
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition. [online] pp.354-377. Available at: https://ac.els-cdn.com/S0031320317304120/1-s2.0-S0031320317304120-main.pdf?_tid=af75709b-fff8-443b-bf1d-2c776439e571&acdnat=1537532715_fd61b20463b30dca98553c2c4763b10f [Accessed 13 Jul. 2018].
- Kollek, D. (2014). An Introduction to Mass Gatherings. [ebook] Available at: http://www.ceep.ca/publications/Mass_Gatherings.pdf [Accessed 18 Apr. 2018].
- Krizhevsky, A., Sutskever, I. and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, [online] pp.1097-1105. Available at:

http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf [Accessed 12 Jul. 2018].

- Nemade, N. and Gohokar, V. (2016). A Survey of Video Datasets for Crowd Density Estimation. Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), [online] pp.389-395. Available at: https://www.researchgate.net/profile/Vinaya_Gohokar/publication/317927392_A_survey_of_video_datasets_for_crowd_density_estimation/links/5993d07fa6fdccaded1de52a/A-survey-of-video-datasets-for-crowd-density-estimation.pdf [Accessed 19 Apr. 2018].
- Owaidah, A. (2014). Hajj Crowd Management via a Mobile Augmented Reality Application: A case of The Hajj event, Saudi Arabia.. Master. Glasgow.
- Palkar, S., Thomas, J., Narayanan, D. and Shanbhag, A. (2017). Weld: Rethinking theinterface between data-intensive applications. [online] Available at: https://arxiv.org/pdf/1709.06416.pdf [Accessed 18 Jun. 2018].
- 81. Sastry, S. and Babu, M. (2018). Implementation of CRISP methodology for ERP systems. International Journal of Computer Science Engineering, [online] 2(5), pp.203-217. Available at: https://arxiv.org/ftp/arxiv/papers/1312/1312.2065.pdf [Accessed 20 Jul. 2018].
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. [online] Available at: https://arxiv.org/pdf/1409.1556.pdf%20http://arxiv.org/abs/1409.1556.pdf [Accessed 17 Jul. 2018].
- Sindagi, V. and Patel, V. (2017). A survey of recent advances in CNN-based single image crowd counting and density estimation. Pattern Recognition Letters, [online] 107, pp.3-16. Available at: https://www.sciencedirect.com/science/article/pii/S0167865517302398 [Accessed 17 Apr. 2018].
- Szegedy, c., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, [online] pp.1-9. Available at: https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf [Accessed 30 Jul. 2018].
- Tripathi, G., Singh, K. and Vishwakarma, D. (2018). Convolutional neural networks for crowd behaviour analysis: a survey. The Visual Computer, [online] pp.1-24. Available at: https://link.springer.com/content/pdf/10.1007%2Fs00371-018-1499-5.pdf [Accessed 10 Jul. 2018].

- World Health Organization (2008). Communicable disease alert and response for mass gatherings. [ebook] Available at: http://www.who.int/csr/Mass_gatherings2.pdf [Accessed 16 Apr. 2018].
- Yang, J., Li, J. and He, Y. (2014). Crowd Density and Counting Estimation Based on Image Textural Feature. Journal of Multimedia, [online] 9(10). Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.658.6282&rep=rep1&type=pdf#page=20 [Accessed 11 Jul. 2018].
- Yezli, S. and Alotaibi, B. (2016). Mass gatherings and mass gatherings health. Saudi Medical Journal, [online] 37(7), pp.729-730. Available at: https://www.researchgate.net/publication/304990470_Mass_gatherings_and_mass_gatherings_health [Accessed 15 Apr. 2018].