PROPOSAL SKRIPSI NON KELAS

Sistem Perhitungan Kerumunan Orang Berbasis Aplikasi Web dengan *Convolutional Neural Network* untuk Video dari Kamera Pengintai

Crowd Counting System on Web Application Based on Convolutional Neural Network for Surveillance Video

Topik: Artificial Intelligence

2001567156 / Audrey Chrysler / Computer Science Global Class / 087777102050
2001545180 / Rivaldi Gunarso / Computer Science Alam Sutera / Global Class / 08170089200
2001598451 / Triladias Puten / Computer Science Global Class / 081291367575

BINUS University
2019

Diperiksa oleh

D5544 - Tjeng Wawan Cenggoro, S.Kom., M.TI.

Table of Contents

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

1.1 Introduction

The work on crowds has gained much attention in recent years for various types of applications and systems that focuses on public safety and traffic control such as the video surveillance. There are numerous research working on various aspects of the analysis of the crowded scenes such as detecting anomalies and crowd counting (Faisal, Sleit, & Alsayyed, 2018). The crowding evaluation used in various environments are addressed in the research field of the surveillance system where the visual surveillance such as the video surveillance are employed for that matter. Automatic decision is also limited to the selected scenes or spaces acquired with different cameras. The different areas that are monitored by video surveillance are presented to the operator of the set of monitors in the control room. Where the monitored surveillance consist of all the footage of live video monitored by a human operator (Regazzoni & Tesei, 1995). Since there are new generation of automatic and semi-automatic visual surveillance system where they are characterized by their capabilities of image processing such as the ability to classify whether it is a human or animals as well as crowd counting (Regazzoni & Tesei, 1995; Faisal, Sleit, & Alsayyed, 2018). As technology advancements continue to grow, new systems include robust algorithms are introduced to aim at detecting motion, validating and making estimation in a specific environment.

The proliferation usage of Closed-Circuit Television (CCTV) coverage have increased the past decade due to the rapid growth of the world's population (Dailey, 2013). Since there are more likely high-density crowd in an area where there are also more likely chances for it to be out of control which could lead to casualties (Luo et al., 2018). Therefore, it is important to use utilize CCTV for safety and surveillance purposes. The implementation of the visual surveillances system of the human flow is used for the purpose of monitoring areas to detect unusual events happening, threat detections, managing the high traffic roads, video surveillance, as well as knowing the demographic of people in that certain area and various industries including streets, banks, train stations, airports, theme parks, shopping malls and other public places (Nakatsuka, Iwatani, & Katto, 2008; Luo et al., 2018; Roqueiro & Petrushin, 2007).

From the recent studies, the implementation today that the crowd counting done with density estimation still faces some challenges. The satisfactions are not met by the system due to some limitations that are present such as the effect of perspective of the camera, most of all the lack of accuracy of the estimation of the crowd density due to the perspective distortion. As for the system implemented on the streets, it is difficult to locate the pedestrians accurately because of the scales and distances as well as the hardware. Where in high density places, it is unavoidable that people will be obscured and the crowd density distribution will be uneven due to the perspective changes and the changes of the density.

In the research of Cross- scene Crowd Counting via Deep Convolutional Neural Networks the challenges faced in the current system is the perspective distortion is due to the result of the lack of additional training in different spaces or places. Most models are trained in one specific scene therefore the system has difficulty when it is used in other places. Crowd segmentation is also a challenging problem where it cannot be accurately obtained when in most crowded scene particularly if there are lack of data training (Zhang, Li, Wang, & Yang, 2015). Where even for images that are also in the same dataset, the distribution or the pattern of the crowd density will not be in uniform. Which results to the final crowd counting estimation to be under-estimation or overestimation in crowd images (Zou et al., 2019). Most problem surfaces when implementing the system is the inaccuracy that occurs when the system

is to do the crowd counting. On some research, people used the face detection program to determine the crowd count unfortunately this is not an effective method. Since, the objective is to be able to count everyone that is in the frame of the camera, not only the ones that the camera can detect within the frame. With this method, it is tremendously affected by the angle of the camera and the program should be able to count the crowds within the frame not only the people's faces that are visible which will lead to a poor result in the estimation (Roqueiro & Petrushin, 2007).

The system of counting people is an important part of the automatic surveillance system where the task of crowd counting can be the primary focus of the system (Roqueiro & Petrushin, 2007). The crowd evaluation from crowd counting data can be conducted in real time are commonly used to get accurate metrics on the demographic of the population. In this paper, several methods will be proposed in order to achieve a result that will be reliable and is according to the main objective which is to be able to create a platform that will be able to provide accurate data in crowd counting. Where the crowd counting system based on Convolutional Neural Network using web application as the main instrument to mark photos or real time videos from surveillance videos to get a more accurate result. Where the web application will enable an effective presentation to the users (Nakatsuka, Iwatani, & Katto, 2008). With the new combined methods that will be conducted that will be able to provide a more accurate result, will be able to better the current state of the system today.

1.2 Problem Formulation

Would the combined methods of previous research applied to the web application revise and raise the level of accuracy of the crowd counting system?

CHAPTER 2. LITERATURE REVIEW

2.1 Related Works & Proposed Idea

With the advancements of the video surveillance have revolutionized the industry the developments made are rising with the growth of the cyber security solutions. However, building an automatic detection system with a high level of accuracy is still a challenge.

Brostow and Cipolla worked out a system where they were able to **detect specific people in crowds.** However, there is a fault in the system where they encountered when there are noises or other objects that exist such as stores and kiosk (Brostow & Cipolla, 2006).

As for Pathan et al. they worked on the system where it **counts by the erroneous movements** in a public place. The system has encountered a difficulty when it comes to the accuracy of the result due to the methods used which is the subtraction process in detecting people (Pathan, Al-Hamadi, & Michaelis, 2010).

Krausz and Bauckage presented the idea of a system that will automatically identify the critical situation during a congestion by **using an alarm system**. When the system was implemented, they found the error in the detection even at the normal state (Krausz & Bauckhage, 2011).

In the research of Zhao and Nevatia, they used the articulated ellipsoids to be able to model the human form as well as the colour histograms to model different appearance of people. They also use an augmented Gaussian distribution for the model of the background. As

the moving head pixels are detected, the MCMC approach is used to help better the probability of the multi person configurations.

Counting by detection. This method of crowd counting particularly utilizes sliding window-based detection algorithms in order to count the number of instances inside of an image (Toptaka et al, 2014). By using these methods, it will inflict the background clutter and the presence of high density crowd. To overcome this, researchers used another method to count the crowd, which is to count by global regression.

Counting by global regression. There are some works that have been proposed to count the pedestrians by detection (Wang & Wang, 2011) or trajectory-clustering (Rabaud & Belongie, 2006). But for the crowd counting problem, these methods are limited by severe obstructions between people. Some of the existing methods (Chan, Liang & Vasconcelos, 2008; Kong, Gray & Tao, 2006) tried to predict global regression counts by using regressors trained with low-level features. These are more suitable for crowded environments and more efficient in terms of computational process. (Change Loy et al, 2013) acquainted semi-supervised regression and data transferring methods to reduce the needed amount of training data, but it still needs some labels from the target crowd scene. (Idrees et al, 2013) estimated the number of individuals in dense crowds based on multi-source information from images, but no from surveillance videos.

Counting by density estimation. Counting by global regression ignores spatial information of pedestrians. (Lempitskyet et al, 2010) introduced an object counting method through pixel-level density map regression. Besides considering spatial information, another advantage of density regression based methods is that they are able to estimate object counts in any region of an image. Taking this advantage, an interactive object counting system was introduced in (Arteta et al, 2014), which visualized region counts to help users determine the relevance feedback efficiently.

In this paper, the proposed idea is to implement the crowd counting using CNN based on the web application as the instrument to gather the data from either the pictures uploaded or the live footage of the video surveillance that will be able to retrieve more accurate data. The working crowd counting approaches could be divided into several categories (Loy et al, 2013), which are: counting by detection, counting by global regression, and counting by density estimation.

2.2 Crowd Counting System

Crowd counting is a technique to estimate numbers of people in an image or a video (Kurama, 2019). This method considers the estimation problem into regression problem, and some of them try to create a heatmap of the possible location of people's heads using Convolutional Neural Network ("Algorithm Spotlight: Crowd Counter", 2018). According to Loy, Chen, Gong & Xiang (2013), the taxonomy of crowd counting algorithm can be grouped into three paradigms:

a. Counting by Detection

Monolithic detection: Detection on a full-body appearance of a set of training images. Whole body monolithic detector can generates reasonable detection in sparse scene.

Part-based detection: different with the monolithic detection method, this method uses the part-based method. For instance, instead of a full-body, one can construct boosted classifiers for specific body part such as the head or shoulder to count the people. Similar to monolithic detection, partial detection relaxes the rigid visibility assumption of the entire body, making it more stable in crowded scenes.

Shape matching detection: the detection that is revealing not only the count and location but also the pose of each person in the scenes.

b. Counting by Clustering

Clustering method counting is based on the assumption that individual motion field or visual features are relatively uniform, thereby allowing for the grouping of coherent object trajectories to represent independently moving entities. First, the method generates a set of crowd-based person hypotheses. The hypotheses are then iteratively refined by assigning small patches of the crowd to the hypotheses based on the constancy of motion fields and color in the garment.

c. Counting by Regression

Counting by regression intentionally prohibits actual individual segregation or monitoring of features, but estimates crowd density based on a systematic and collective explanation of crowd patterns.

2.3 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a Deep Learning algorithm that can take in an input image, assign it to various aspects of image and be able to differentiate one from another (Saha, 2018). CNN is specialized kind of neural network for processing data that has a known, grid-like topology (Rukmanda, Sugeng & Murfi, 2018). According to Goodfellow, Bengio & Courville (2016), the name "Convolutional Neural Network" indicates that the network has a mathematical operation called **convolution**. The network uses a special architecture which is particularly well-adapted to classify images, those special architecture makes the network fast to train because it trains deep and has many layer networks which is good to classify the images (Ahn, 2016).

According to Yamashita, Nishio, Do & Togashi (2018), CNN architecture includes several building blocks, Convolution Layers, Pooling Layers, and Fully Connected Layers. These are the three layers (Xu et al., 2017):

a. Convolution Layer

This layer is the core layer in the architecture of CNN. Kernels as a filters are convolved with the overlapped subsets of the input across the width and height. The output feature maps is corresponding to the convolution layer are generated by adding a bias and applying a non-linear activation function.

b. Pooling Layer

Also known as subsampling layer, which is generally inserted between the convolution layer. There are three commonly used pooling methods: max-pooling, mean pooling,

and stochastic pooling. However, the most efficient method is **max-pooling** method. The benefit of this are; decreasing the size lead to less computational overhead and it works against over-fitting.

c. Fully Connected Layer

The fully-connected layer is similar to the way that neurons are arranged in a traditional neural network. It takes the high-level filtered images and translate it into votes. This layer are the primary building block of traditional neural network.

It consists of a stack of several convolution layers and a pooling layers, followed by one or more fully connected layers. Forward propagation is the step where input data are transformed into output through these layers.

2.4 Web Application

Web application is a set of web pages that are generated in response to user request. According to Capka (2019), web application works mainly through the client requesting specific documents from the server. Web applications can be designed for many variety of uses and can be used by anyone; from an organization to an individual for numerous reasons ("What is Web Application (Web Apps) and its Benefits", 2019).

There are six different types of web applications:

a. Static Web Application

The term “static” comes from the web application’s lack of flexibility. Static web applications have their own pages generated by a server and offer a little interactivity. Static web application is often difficult to maintain and this are not well-suited for a mobile environment ("Types of Web Applications: From a Static Web Page to a Progressive Web App - DZone Web Dev", 2019).

b. Dynamic Web Application

Dynamic web application is more complex in technical level. This web application is using databases for loading the data and the content will change and updated each time the user access it. Dynamic website application have an administration panel where the panel itself can help to correct or modify the application’s content (text and images). The dynamic web application can use several languages such as PHP and ASP.

c. Online Store or E-Commerce

This kind of web application is more complex than dynamic web application because it must enable and include electronic payments with many payment methods. The administration panel is also needed for this website for input, update, and delete the product and price module.

d. Portal Web Apps

The portal website means the application which can access several categories of homepage. The website itself can be consisted plenty of things such as forums, chat, email, and browser.

e. Animated Web Application

The animated web application approach allows to present content with animated effects. An animation itself is usually associated with Flash technology.

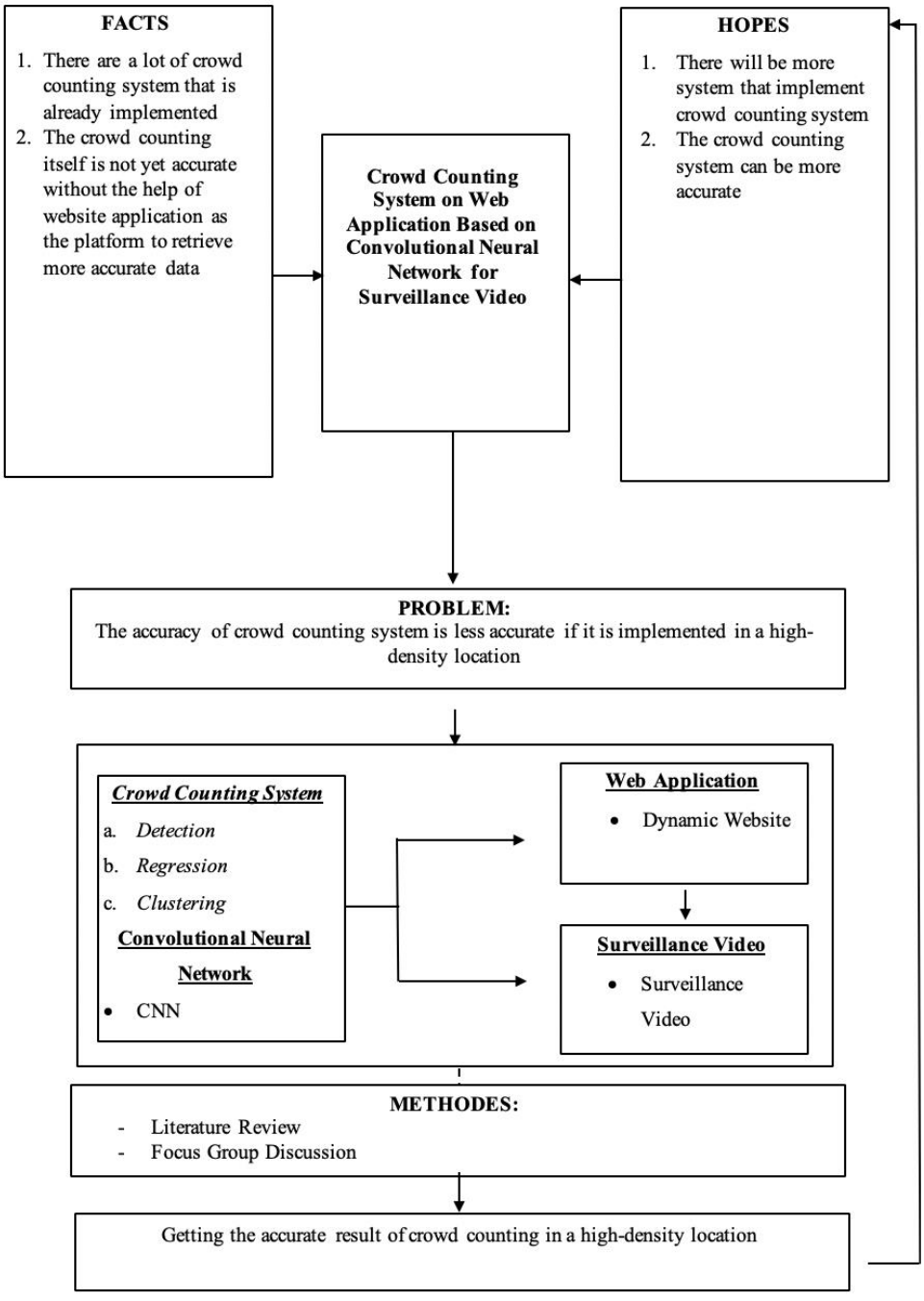
f. Content Management System

It is an administrator panel for the internal user to make an update or changes to the website. The management system is very beneficial to have because the contents of the website must be continually updated when it comes to web development.

CHAPTER 3. RESEARCH METHODOLOGY

3.1 Methodology

3.1.1 Research Framework:



3.1.2 Data Collection Methods:

Several data collection methods will be obtained to get data and information that were needed to fulfill the requirements of developing this research. The data collected will help improve the understanding of the process and the analysis of each data. The data collection methods that will be used such as:

a. Literature Review

This data collection is the most common one to gather information, which obtaining data from journals, research papers, books and official documents. Each data that we obtain from a literature will refer to another literature's reference, and the other literature's reference will refer to another literature's reference and so on. This is called 'branching method' that cause a domino effect will help us to gather more information because each literature is linked to one or more reference and each references linked have more or less the same idea as the root reference. By using the branching method, the information gathering process will be even less time-consuming and the process of obtaining the data will be much more effective and efficient.

b. Focus Group Discussions

Focus group discussions can be used as the data collection method for this research, regarding the fact that this thesis consists of 3 people working on this thesis. Through focus group discussions, researchers can obtain the reasons, motivation, argumentation or the basis of an opinion that came from an individual or a group. Some issues might be easier to be discussed in a group within a particular topic. From the discussed issue, the information can be gathered easier compared to finding facts alone compared using focus group discussions. Therefore, this data collection method is used to gather data effectively and efficiently.

3.1.3 Development Methodology

This crowd counting system project will be developed using Waterfall methodology. This methodology is chosen because Waterfall methodology flexibility and effectiveness (Bassil, 2011). Here are the steps in Waterfall methodology such as:

1. **Analysis:** Gathering the requirements thoroughly, then define and analyze the needs that have to be fulfilled by the program that were about to be built. This phase has to be done completely in order to produce a comprehensive design.
2. **Design:** In this phase, the developer will produce an intact system and decide the flow of software development until detailed algorithm.
3. **Implementation:** A phase where all of the design were converted into line of codes. These line of codes still in forms of modules that about to be integrated into a complete system.
4. **Testing:** In this phase, the modules made were being merged and tested. The testing were executed in purpose to know if the software in accordance with the functionality and the design.
5. **Deployment:** The client or user execute a test to know whether the system has been functioning properly or not. After the system functioning properly, then it is ready to be deployed.
6. **Maintenance:** The process of improvement the current state of the deployed system and troubleshooting problems that appeared after the deployment were executed.

3.2 System Design

3.2.1 Software Design Document

a. Software Description

This software will be used to observe how much crowd in a specific area inside a surveillance video footage using crowd counting system based on Convolutional Neural Network. By implementing the system in the web application, the data that will be retrieved from the video surveillance or the uploaded picture of where the crowd counting will be implemented will have more of an accurate result. Furthermore, by having the web application it will help the user where the application will be more user-friendly and accessible to all users.

References:

Ahn, S. (2016). Deep Learning Architectures and Applications. *Journal Of Intelligence And Information Systems*, *22*(2), 127-142. doi: 10.13088/jiis.2016.22.2.127

Algorithm Spotlight: Crowd Counter. (2018). Retrieved 27 December 2019, from https://algorithmia.com/blog/algorithm-spotlight-crowd-counter

Arteta, C., Lempitsky, V., Noble, J. A., & Zisserman, A. (2014, September). Interactive object counting. In *European conference on computer vision* (pp. 504-518). Springer, Cham.

Bassil, Y. (2011). A Simulation Model for the Waterfall Software Development Life Cycle. *International Journal Of Engineering & Technology (Ijet)*, *2*(5), 2. Retrieved from https://arxiv.org/pdf/1205.6904.pdf

Brostow, G., & Cipolla, R. (n.d.). Unsupervised Bayesian Detection of Independent Motion in Crowds. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR06). doi: 10.1109/cvpr.2006.320

Capka, D. (2019). Lesson 1 - Introduction to PHP and web applications. Retrieved 26 December 2019, from https://www.ict.social/php/basics/introduction-to-php-and-web-applications

Chan, A. B., Liang, Z. S. J., & Vasconcelos, N. (2008). Privacy preserving crowd monitoring: Counting people without people models or tracking. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-7). IEEE.

Change Loy, C., Gong, S., & Xiang, T. (2013). From semi-supervised to transfer counting of crowds. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2256-2263).

Dailey, K. (2013, April 29). The rise of CCTV surveillance in the US. Retrieved December 20, 2019, from https://www.bbc.com/news/magazine-22274770

Faisal, E., Sleit, A., & Alsayyed, R. (2018). Crowd Counting Mapping to make a Decision. International Journal of Advanced Computer Science and Applications, 9(2), 282–286. doi: 10.14569/ijacsa.2018.090239

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Idrees, H., Saleemi, I., Seibert, C., & Shah, M. (2013). Multi-source multi-scale counting in extremely dense crowd images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2547-2554).

Kong, D., Gray, D., & Tao, H. (2006). A viewpoint invariant approach for crowd counting. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 3, pp. 1187-1190). IEEE.

Krausz, B & Bauckhage, C. (2011). Automatic detection of dangerous mo- tion behavior in human crowds, in Advanced Video and Signal-Based Surveillance (AVSS), 8th IEEE International Conference on, pp. 224229,2011.

Kurama, V. (2019). Dense and Sparse Crowd Counting Methods and Techniques: A Review. Retrieved 27 December 2019, from https://nanonets.com/blog/crowd-counting-review/#what-is-crowd-counting

Lempitsky, V., & Zisserman, A. (2010). Learning to count objects in images. In *Advances in neural information processing systems* (pp. 1324-1332).

Lin, S.-F., Lin, C.-D., 2006. Estimation of the pedestrians on a crosswalk. In: International Joint Conference SICE-ICASE, 2006. IEEE, pp. 4931–4936, http:// dx.doi.org/10.1109/SICE.2006.314851

Loy, C., Chen, K., Gong, S., & Xiang, T. (2013). Crowd Counting and Profiling: Methodology and Evaluation. *Modeling, Simulation And Visual Analysis Of Crowds*, 347-382. doi: 10.1007/978-1-4614-8483-7_14

Luo, H., Sang, J., Wu, W., Xiang, H., Xiang, Z., Zhang, Q., & Wu, Z. (2018). Applied Sciences. A High-Density Crowd Counting Method Based on Convolutional Feature Fusion, 8(12), 1–12. doi: 10.3390/app8122367

Nakatsuka, M., Iwatani, H., & Katto, J. (2008). International Conference on Mobile Computing and Ubiquitous Networking. A Study on Passive Crowd Density Estimation Using Wireless Sensors, 1–6.

Pathan, S., Al-Hamadi, A., & Michaelis, B. (2010). Crowd behavior detection by statistical modeling of motion patterns, in Soft Computing and Pattern Recognition (SoCPaR), International Conference on, pp. 8186, 2010.

Rabaud, V., & Belongie, S. (2006). Counting crowded moving objects. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 1, pp. 705-711). IEEE.

Regazzoni, C. S., & Tesei, A. (1995). Signal Processing. Distributed Data Fusion for Real-Time Crowding Estimation, 47–63.

Roqueiro, D., & Petrushin, V. A. (2007). Counting people using video cameras. International Journal of Parallel, Emergent and Distributed Systems, 22(3), 193–209. doi: 10.1080/17445760601139096

Rukmanda, T., Sugeng, K., & Murfi, H. (2018). Modification of architecture learning convolutional neural network for graph. *AIP Conference Proceedings 2023*. doi: 10.1063/1.5064197

Saha, S. (2018). A Comprehensive Guide to Convolutional Neural Networks—the ELI5 way. From https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Topkaya, I., Erdogan, H., & Porikli, F. (2014). Counting people by clustering person detector outputs. *2014 11Th IEEE International Conference On Advanced Video And Signal Based Surveillance (AVSS)*. doi: 10.1109/avss.2014.6918687

Types of Web Applications: From a Static Web Page to a Progressive Web App - DZone Web Dev. (2019). Retrieved 27 December 2019, from https://dzone.com/articles/types-of-web-applications-from-a-static-web-page-t

Wang, M., & Wang, X. (2011). Automatic adaptation of a generic pedestrian detector to a specific traffic scene. In *CVPR 2011* (pp. 3401-3408). IEEE.

What is Web Application (Web Apps) and its Benefits. (2019). Retrieved 27 December 2019, from https://searchsoftwarequality.techtarget.com/definition/Web-application-Web-app

Wu, J., Li, Z., Qu, W., & Zhou, Y. (2019). One Shot Crowd Counting with Deep Scale Adaptive Neural Network. Electronics, 8(6), 701. doi: 10.3390/electronics8060701

Xu, M., Papageorgiou, D., Abidi, S., Dao, M., Zhao, H., & Karniadakis, G. (2017). A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLOS Computational Biology*, *13*(10), e1005746. doi: 10.1371/journal.pcbi.1005746

Yamashita, R., Nishio, M., Do, R., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Into Imaging*, *9*(4), 611-629. doi: 10.1007/s13244-018-0639-9

Zhang, C., Li, H., Wang, X., & Yang, X. (2015). Cross-scene crowd counting via deep convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 833–841. doi: 10.1109/cvpr.2015.7298684

Zhang, D., Peng, H., Haibin, Y., & Lu, Y. (2013). Crowd Abnormal Behavior Detection Based on Machine Learning. Information Technology Journal, 12(6), 1199–1205. doi: 10.3923/itj.2013.1199.1205

Zhao, T & Nevatia, R. (2004). Tracking multiple humans in complex situations. IEEE Trans. Pattern Anal. Mach. Intell., Volume 26(9):1208–1221.

Zhao, T & Nevatia, R. (2004). Tracking multiple humans in crowded environment. In *CVPR (2)*, 406–413.

Zou, Z., Cheng, Y., Qu, X., Ji, S., Guo, X., & Zhuo, P. (2019). Journal of LATEX Templates. Attend To Count: Crowd Counting with Adaptive Capacity Multi-Scale CNNs, 1–10. Retrieved from https://arxiv.org/pdf/1908.02797.pdf