Everything Counts: A Taxonomy of Deep Learning Approaches for Object Counting

Kai Heinrich
Otto-von-Guericke-Universität Magdeburg

Andreas Roth
Technische Universität Dresden

Patrick Zschech
Friedrich-Alexander-University of Erlangen-Nürnberg

Everything Counts: A Taxonomy of Deep Learning Approaches for Object Counting

Research paper

Heinrich, Kai, TU-Dresden, Dresden, kai.heinrich@tu-dresden.de

Roth, Andreas, TU-Dresden, Dresden, andreas.roth@mailbox.tu-dresden.de

Zschech, Patrick, TU-Dresden, Dresden, patrick.zschech@tu-dresden.de

Abstract

Many tasks, like process fault detection, disease diagnostic or maintaining security in public places are connected to the process of counting objects from observing image data like camera footage or microscopic images. However, the pivotal process of counting objects based on image data is subject to a lot of error sources when done manually. Therefore, a lot of approaches exist that aim to aid the automation of the counting process. Since most of those methods are of black box nature and have to be applied to real world scenarios, we provide a taxonomy that reflects both: the method characteristic as well as the counting problem characteristics in order to derive a systematization of the scientific field and to provide a tool that acts as a set of guidelines for choosing appropriate counting methods for different situations. We first conduct a literature review, where counting problem characteristics and solutions are extracted and later transformed into a taxonomy. We finally showcase the taxonomy using four case studies that are based on publicly available datasets for the sake of scientific comprehension.

Keywords: Counting, Object Detection, Deep Learning, Survey, Taxonomy.

1 Introduction

The forecasting of features that influence economical outcomes like profit by utilizing available internal and external data plays an important role in the age of digitization. The service and manufacturing industry can benefit from disruptive and innovative processes and technologies that follow up the digital transformation process (Shen, Chao and Zhao, 2015). Especially the technology provided within the scientific field of Artificial Intelligence (AI) with its subsidiary field of Deep Learning Systems (DLS) aims to provide solutions to problems that are either very cumbersome or simply impossible for humans to tackle (Goodfellow, Bengio and Courville, 2016). One of these problems is the counting of objects based on image data. This basic task provides a foundation for many prediction problems including yield prognosis (e.g., counting crops in agrarian management), failure analysis and maintenance (e.g., counting small tears in a machine as a sign of wear and tear), security surveillance (e.g., surveillance of large crowds of people) or disease diagnostics (e.g., counting cells based on microscopic images). These tasks are often very labor intensive and subject to errors when executed manually by humans, since they are monotonic and repetitive which leads to tiredness and lack of concentration (Cohen et al., 2017). A suitable automation of the counting process with a low error rate significantly improves problem solutions that are based on counting. However, there are several difficulties connected to the task of counting automation, such as other objects that cover the objects of interest or a skewed perspective (C. Zhang, 2015). While most counting systems promise improved results in terms of prediction performance, they come with high barriers in terms of application interpretability and required computational power (Larochelle et al., 2007; Arel, Rose and Karnowski, 2010). While

DLS can provide promising solutions to the counting problem, they remain a black box and therefore one cannot fully understand their behavior given different environments and data scenarios (Heinrich and Fleissner, 2018). Different environmental factors that characterize a specific counting problem, like perspective or object density, lead to different requirements towards the algorithm. In addition to the plain number of objects, an algorithm can, for example, be required to output numbers for only interesting parts of the image and therefore needing localization information. The construction, evaluation and systematization of AI technologies like DLS in the context of applications, tasks and domains should be considered a vital research interest in IS research (Agarwal and Dhar, 2014). While the problem of constructing decision support systems that can be applied to the task of counting is somewhat established for specific problems, the efforts of evaluation and systematization of new analytic tools to harness business value from growing amounts of data in order to achieve full or semi-automated processes are somewhat sparse (H. Chen, Chiang and Storey, 2012; Sharma, Mithas and Kankanhalli, 2014). This is partially due to the reason that there is a lack of a comprehensive descriptive system that organizes knowledge in a structured manner to aid the construction of descriptive theories (Gregor, 2006). Therefore, our research goal is to provide a taxonomy for object counting with DLS based on image data. In order to achieve that goal, we employ a two-step-approach: We first review related work on DLS for the task of object counting and secondly, based on this systematic review, we conduct a content analysis of the knowledge base and propose a systematization using taxonomy development methods. The paper is structured as follows: In the next section we provide a description of criteria for our literature review and the subsequent conceptualization of approaches for object counting. We present a taxonomy as a form of conceptualization as well as details about the development process and the different dimensions and characteristics of the taxonomy in Section 3. We then proceed to apply the taxonomy to different cases in Section 4. The last section provides a summary, research limitations and an outlook.

2 Research Methodology

The research goals (RG) of this paper are: (1) to provide a classification scheme for object counting with DLS and (2) to apply this taxonomy to a spectrum of different scenarios to showcase the different method requirements that depend on the nature of the counting problem. In order to provide a taxonomy as a solution to RG (1) we conduct a literature review with content analysis and taxonomy development (Webster and Watson, 2002; Nickerson, Varshney and Muntermann, 2013). The development of a taxonomy is considered a valid research method in IS research and has been applied to several domains (Jöhnk, Röglinger, Thimmel and al, 2017; Zschech, 2018). We then proceed to project four real world cases, counting humans, cars, cells and plants, onto our taxonomy space to address RG (2). We chose the cases because they (a) span a wide spectrum of counting problem environments and are therefore suitable to showcase the developed taxonomy and (b) we can project the cases using well-known image datasets from scientific literature that exhibit certain characteristics which influence the choice of DLS. In addition to that, the datasets are all publicly available and well-described which enables us to conduct a transparent and comprehensible case projection. The research process is depicted in Figure 1.

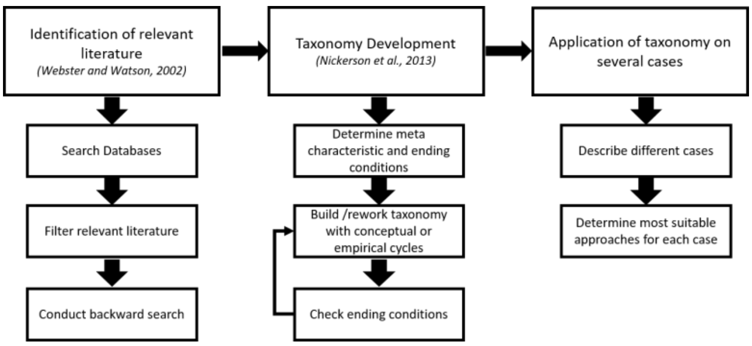

Figure 1. Research Process for RG (1) and RG (2)

For the identification of relevant literature, we conducted a search with the terms “count” and either of the keywords “deep learning” or “neural network” in the databases Science Direct, EBSCOhost, IEEE Xplore and arXiv. Did however not consider SpringerLink, Google Scholar or ACM for search limitations regarding our inclusion criteria. Also an additional search approach revealed duplicate counts in the 90^{th} percentile. Both, the terms and the database choices cover the area of interest as suggested by RG (1). The search was conducted in a way, that the terms had to be present in at least one of the following meta data of the paper: abstract, title or keywords, where we explicitly defined the search string in a way, that the term “count” had to be present in the title as a necessary condition. Initially 321 relevant papers were identified. We then proceeded to apply subjective filtering by screening the title (131 papers remain) and then the abstract (87 papers remain) for relevance. After removing duplicates (83 papers remain) and conducting an additional backward search (16 papers added), we yield a final set of 99 relevant papers. An additional forward search based on the 99 papers revealed no additional results. The subjective filter was implemented based on some relevancy checks: (a) counting objects had to be a core topic of the paper, which prevented the inclusion object tracking or object recognition heavy papers; (b) the paper needed to be an original work rather than a survey, so that only papers that describe a method in-depth will be included in the taxonomy; (c) the method was based on image rather than video data; (d) the article had to include artificial neural networks (ANN), both, standard and DLS, since that was the focus of our research goal; (e) the focus was set on single image counting rather than on sequence counting (e.g., in videos) and (f) we excluded literature that only used artificial neural networks to pre-process the data.

3 Taxonomy for Object Counting

3.1 Taxonomy Development

A taxonomy consists of a set of dimensions that in turn consist of mutually exclusive characteristics (Nickerson et al., 2013). The characteristics need to cover the topic of the respective dimension completely. That requires our objects of interest to be assigned to exactly one characteristic in each dimension. The objects of interest in our case are scientific contributions that describe methods for counting objects based on image data using neural network models. Taxonomies are developed by determining dimensions and characteristics as a consequence of an initial meta characteristic, which is derived from the target audience or the research goal, underlying the purpose of creating the taxonomy. Since the target audience are researchers that want to solve a counting problem, the meta characteristic focuses on the application of the method. Therefore, the taxonomy should reflect on the method in terms of application environment rather than detailed differences in the algorithm themselves. The subjective and objective ending conditions of the development process are checked with every iteration. The minimal subjective ending conditions are a concise, robust, comprehensive, extendible, and explanatory taxonomy and are directly applied from Nickerson et al., 2013. The ending condition is fulfilled when (a) the taxonomy consists of mutually exclusive and complete characteristics, (b) all objects of interest are assigned and (c) there is no additional change in the last iteration of the development process. We started the iteration process with an inductive approach, since we had a large amount of objects (99). During the first two iterations, we identified common characteristics in a subgroup of ten and five objects respectively to create five dimensions: *Architecture*, *Number of Objects*, *Density of Objects*, *Output* and type of *Training Data*. After these first two empirical-to-conceptual iterations, we conducted a conceptual-to-empirical iteration. It was found that the counting problem itself exerts a substantial influence on the choice of method (e.g., the choice of method depends heavily on density and number of objects). In addition to that, it was found that researcher were able to further optimize results based on a static background. We therefore added the *Background Dynamic* as an additional dimension.

Dimension	Characteristics				
Architecture	One-Step (22)		Multiple Steps (ANN) (16)		Multiple Steps (ANN+other) (61)
Number of Objects	0-15 (32)	15-50 (23)	50-200 (21)	200-1000 (14)	1000< (4)
Density of Objects	No Overlapping (31)			Overlapping (68)	
Background Dynamic	Static (43)			Dynamic (56)	
Output	Estimation Value (10)	Absolute Count (59)	Absolute Count by Region (9)	Bounding Box (6)	Centroid Coordinate (15)
Training Data	Absolute Count (38)		Centroid Coordinate (35)		No label for counting (19)
Type of Counting	Detection based (15)		Feature based (69)		Cluster based (15)

Figure 2. Taxonomy of Deep Learning counting methods with literature count in parenthesis

We continued by conducting empirical-to-conceptual cycles based on subgroups of ten objects in iterations 4-6 and groups of 15 objects in iterations 7-9. In the last iteration the remaining nine objects were classified and besides the identification of further characteristics, the dimension *Type of Counting* was added. Since no more objects could be assigned and no more revisions occurred during the last iteration, the subjective and objective sets of ending conditions could be checked.

Using the recommended setting for the concise criterion is to have between five and nine dimensions, which is fulfilled by our seven dimension design. The criterion for robustness is also fulfilled, since the taxonomy enables researchers to distinguish well between different methods for object counting. Especially the dimensions *Architecture* and *Type of Counting* help to identify a suitable method. After the last two iterations did not lead to considerable change in the taxonomy, the criterion of a comprehensive taxonomy can also be verified. Since we can add additional dimensions to the taxonomy without violating the criterion of a concise taxonomy, the criterion of extensibility can also be seen as fulfilled. Above that the taxonomy can be deemed explanatory since the inclusion of the dimensions *Architecture, Type of Counting* and *Output* help to describe information processing towards output generation. The objective criteria are also fulfilled, since the dimensions are mutually exclusive, no more changes were needed during the last iteration and we classified the entire set of objects. Therefore all ending conditions can be seen as fulfilled which terminates the taxonomy development process at our last iteration. The final taxonomy is depicted in Figure 2 and the dimensions and characteristics are explained in the next subsection. We indicate the evidence strength of a characteristic by giving the number of relevant papers *x* out of the 99 papers that constitute our knowledge base in parenthesis (e.g. *x/99*) in the description as well as in the summary in Figure 2.

3.2 Taxonomy Dimensions and Characteristics

3.2.1 Architecture

This concept distinguishes the counting methods based on extent and type of systems used in the data analysis process that is responsible for counting. First, we distinguish systems that only use one step processes (22/99), resulting in a *One-Step* ANN that can be trained as a whole. Typically those networks were based on CNN architectures (Capobianco and Marinai, 2017; Hao Lu et al., 2017; Kang, Ma and Chan, 2017). Those methods typically use local and/or global features to determine the object density per pixel or partial and whole images. Another one-step ANN architecture can be used to solve the classification task of whether an object is present in a given part of the image or not (Achler, Vural and Amir, 2009; Song, Qiao and Corbetta, 2017).

Second, there are counting systems that use *multiple steps that only involve ANN algorithms* (16/99). Those architectures usually consist of the extraction of partial image features and then using them within a regression model (Aich and Stavness, 2017; Marsden et al., 2017; Tang, Pan and Zhou, 2017). This machine based feature extraction provides advantages in scenarios with overlapping and strong variations in object size (Boominathan, Kruthiventi and Babu, 2016). Other ANN-based multi-step approaches involve sequential models like Long-Term-Short-Term networks (Yao, Han, Wan and Hou, 2017) or parallel architectures that use multiple networks as base learners while having a final “deciding network” in the second step (Boominathan et al., 2016; Onoro-Rubio and López-Sastre, 2016; Y. Zhang et al., 2016).

The last characteristic involves *multiple ANNs as well as other methods* (61/99). This category acts as a collection for the versatile preprocessing possibilities of ANNs and other image processing methods. ANNs are used to either manipulate images to achieve better results (e.g., scaling or color related features), extract features (e.g., abstract or low level features like edges or texture) and for regression and combination of previous results (Cho, Chow and Leung, 1999; Ikeda et al., 2008; Ferrari, Lombardi and Signoroni, 2017).

3.2.2 Number of Objects

This concept, while it should not be considered isolated, gives us a description of the counting problem at hand that we try to solve using ANN algorithms. The success of various counting methods depends largely on that concept. For counting crowds of people, detection-based methods can be used for a small number of objects but not if large crowds are present in images, where you would use feature-based methods (Sindagi and Patel, 2017). We divided the found literature up into small number of objects ranging from *0-15* (32/99) and *15-50* (23/99 papers), medium sized *50-200* (21/99), large *200-1000* (14/99) and “outlier” or “extra-large” with the number of objects being larger than *1000* (4/99). The numbers are in regard to the approximate object count as identified by the authors when not given directly in the publication. Some methods were built to be more flexible and can be applied to various object numbers. The classification of an object into a class was conducted based on the average numbers of objects in the used image datasets. We could, however, not determine such numbers for five publications, since there was no indication (neither sufficient image data nor numbers) with regard to the number of objects.

3.2.3 Density of Objects

The concept of density is another characteristic attribute of the counting problem and we can distinguish between *no overlapping* objects (31/99) and *overlapping* objects (68/99). While further breaking down this dimension by analyzing the degree of overlapping would be an interesting feature, we cannot objectively determine that from the given objects and therefore omit this feature from further inclusion. It was found that for counting problems with considerable overlapping, feature-based methods are preferred and more successful in predicting object numbers (Lempitsky and Zisserman, 2010). Only five publications could be found that applied detection- or segmentation-based methods to images with considerable amounts of overlapping. The core idea of those methods is the concept of distinction of objects which is a task that grows harder when the amount of overlapping increases. The inference of the object count from local and/or global features applied by feature-based methods has a clear advantage in the case of high density. For the case of small amounts or absence of overlapping typical examples are counting cells on microscopic images (Bertucco et al., 1998; Embleton, Gibson and Heaney, 2003) or counting cars/humans from an aerial perspective (Pavlidis, Morellas and Papanikolopoulos, 2000; Yifei Xue, Tiejun Wang and Skidmore, 2017).

3.2.4 Background Dynamic

Another important influence factor found in the literature is the background dynamic, which can be distinguished into *static* (43/99) and *dynamic backgrounds* (56/99). In this context, static is expressed as a class of methods that only work with a background that has certain static features. “Static” can have multiple meanings in this context: We can assume that over a sequence of images only the foreground changes (e.g., fixed surveillance cameras) where we can ignore the background and remove it via background subtraction methods (Amin et al., 2008; Yoshinaga, Shimada and Taniguchi, 2010; French, Fisher, Mackiewicz and Needle, 2015). As another option for static backgrounds we can assume a region of interest as part of an image that is henceforth declared as foreground (Laki et al., 2011; Hu, Zheng, Chen and Chen, 2015). This enables some methods to ignore small changes in the background like moving cloud shadows or constant non-moving foreground objects (Hou and Pang, 2008; Garcia-Bunster, Torres-Torriti and Oberli, 2012). Another variation of static backgrounds are mono-colored backgrounds, where only noise removal is necessary (e.g., when observing cells under a microscope) (Usaj, Torkar, Kanduser and Miklavcic, 2011; Adagale and Pawar, 2013; Ferrari et al., 2017). The assumption of static background is often times required to distinguish between foreground and background objects, where only the number or density of the foreground objects is of interest. Methods that do not require static background because of the foreground/background distinction are classified into the characteristic “dynamic” (Pavlidis et al., 2000; Mundhenk, Konjevod, Sakla and Boakye, 2016; Nazeer, Wong and Nichol, 2017).

3.2.5 Output

The counting methods found in the literature offer a range of different outputs, depending on the specifics of the problem at hand. The characteristics are given and described in ascending order of degree of detail of the output information. In some cases the methods only supply a rough *estimation value* (10/99) without positioning or a total count. This is the case for public transportation surveillance where only a rough percentage estimate is needed, e.g., of how crowded a departure platform is (Cho et al., 1999; Chow and Cho, 2002), reaching from 0-100%.

Some methods output an *absolute count* of objects in an image (59/99), again without positioning information. To circumvent this problem, some approaches divide the image and provide the *absolute count by region* (8/99). With using overlapping regions, more robust models are generated (Cohen et al., 2017; Han, Wan, Yao and Hou, 2017). When regions are overlapped on purpose, certain parts of an image are considered multiple times by an algorithm which makes the results more stable and robust towards output error (Cohen et al., 2017).

Other output methods that take positioning into account, involve determining a count per pixel by either returning a *bounding box* that forms a rectangular shape around the object (6/99) or using centroid coordinate by pixel density (15/99). When using a bounding box the object is framed by a rectangle and then the rectangle coordinates are returned. Another approach to return a count per pixel via centroid coordinate is using a density function as depicted by a heat map in Figure 3, where a higher pixel density is expressed by red color whereas low densities are expressed by blue color (Lempitsky and Zisserman, 2010; Kang et al., 2017; Zeng et al., 2017). This avoids the cumbersome task of detection and localization of objects, because the focus lies on each pixel instead of image regions. The advantage of this method is that, in addition to calculating a total object number, the algorithm can output numbers for arbitrary regions of interest. In addition to that, the density map can reveal high object densities within the different regions of an image (Shao, Wang, Xue and Zhang, 2015).

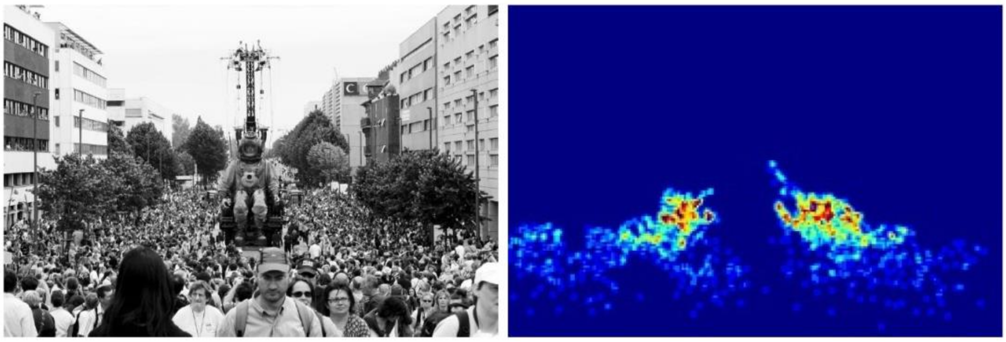

Figure 3. Example of density based count per pixel method (Marsden et al., 2017)

3.2.6 Training Data

The different options of available training data can be characterized by the effort it takes to annotate the unlabeled image data. We present the different characteristics of this dimension from lowest to highest effort. The first annotation tier with the lowest effort is given by annotating the *total count* (38/99). This is particularly useful when we only need an estimate or the total number as output without positioning information.

For the purpose of annotating (labeling) with positioning information, some kind of *centroid coordinate* (35/99) is needed. This involves manually "pointing out" the objects by humans, which is the natural way for a person to count a large number objects. The absolute count of objects in an image can usually only be annotated instantly by humans for a small number of objects, where the total number can be "seen" instantly. Therefore, central coordinate labeling results in an object instance that is marked with a dot (usually in the middle) which is preferred to giving the total number when we have a large amount of objects due to the reasons described above. This type of training data can be used for all output types since some outputs require the image to be divided up into regions, which in turn requires positioning information such as centroid coordinates (Marsden et al., 2017). It was also found that this kind of data is needed in order to create density maps using Gaussian filter techniques (Boominathan et al., 2016).

The last tier of approaches do *not require the annotation* of training data at all (19/99). Those methods are divided up into cluster-based counting (Bertucco et al., 1998; Biswas et al., 2017), where the image is split into segments and the number of segments is given as an output and detection-based counting which instead of annotations needs example images of the objects that are to be counted (Achler et al., 2009; Biswas et al., 2017). Seven objects could not be characterized unanimously in this category.

3.2.7 Type of counting

The last dimension gives an insight on how the different approaches work. The first option is to use segmentation or *cluster-based* methods (15/99). After background and noise removal, which are essential pre-processing steps for this kind of approach, every foreground segment is declared as one object instance (Bertucco et al., 1998; Manik, Saini and Vadera, 2016). In addition to de-noising, often times smoothing and image sharpening algorithms are applied to avoid erroneous outputs and improve prediction quality (Zhou, Xiao and Ma, 2010). The foreground/background segmentation is either conducted by a classification algorithm or by a threshold transforming images into binary black and white images with white pixels representing foreground objects (Schofield, Mehta and Stonham, 1996). Those methods can only be applied when we experience low to non-overlapping, since a combined foreground segment would be limited to one object, which is not the case when overlapping is present.

The *detection-based methods* (15/99) use a multi-step approach where objects are detected first and then localized and counted afterwards. Depending on the task at hand, the detector is either trained to recognize the whole object (Biswas et al., 2017) or only parts of it (e.g., only the head of a human body) (Song et al., 2017). The detectors are trained with single image training data to determine whether the object of interest is present or not. However a shortcoming of those approaches is that they cannot successfully be used with strong overlapping and large number of objects (Loy, Chen, Gong and Xiang, 2013; Sindagi and Patel, 2017).

Feature-based methods (69/99) can circumvent that downside by extracting local and/or global features of the image. This can either be done by manually crafting features (Embleton et al., 2003) or to use annotated data to train feature extractors (Aich and Stavness, 2017; Capobianco and Marinai, 2017). Manually extracted features include shape, color, area, diameter, roughness of texture (Lal, Behera, Sethy and Rath, 2017) or even additional meta annotations like perspective or weather data (Nazeer et al., 2017). After feature extraction, common machine learning algorithms like SVM, linear regression or ANNs are used to predict the desired output (Aik and Zainuddin, 2009; Shao et al., 2015; Aich and Stavness, 2017).

4 Application of the counting taxonomy on different cases

From the systematic literature review we could identify several application domains of DLS counting methods. In order to classify cases based on our taxonomy, we choose case collections that span a wide area of application domains, have some evidence in literature and provide publicly available datasets for comprehensibility: **humans** (50/99), **cars** (4/99), **cells** (22/99) and **plants** (8/99).

The taxonomy is divided into dimensions ***CP*** that describe the counting problem (*Number of Objects*, *Density of Objects*, *Background Dynamic* and *Training Data*) and dimensions ***SOL*** that describe the solution method for the counting problem (*Architecture*, *Output*, *Type of Counting*). We determine the number of objects by using the average over all images in the dataset. The density is classified as “overlapping” whenever there is considerable presence of overlapping in a majority of the dataset images and as “no overlapping” otherwise. Whenever we have changing perspectives and scenes, we declare the background as “dynamic” and “static” otherwise. The training data is determined by the given format of the dataset. While the characteristics of the counting problem are determined by the dataset and therefore can be projected without ambiguity, the characteristics of the solution dimensions are projected for each case based on the most suitable option which is extracted as a consensus majority vote from literature. When we observe the *Type of Counting* for example, it is technically possible to use detection-based methods for dense crowd counting. However, literature suggests that feature-based methods yield better results in a majority of cases. The dimension *Output* depends on the business goal definition (e.g., “What do we need?”) rather than the counting problem and the characteristics were projected based on the problems connected to the datasets in literature. In addition, while the projection onto the dimension *Architecture* was conducted by a majority vote, the application of more complex architectures are subject to economic parameters like cost of computational power and data science capabilities, rather than being subject to a specific counting problem. Therefore, we use a different shading pattern and bold font for the solution dimensions in Figures 4 and 6 to indicate that other options are possible but the marked characteristic reflect the majority of decisions as found in related literature.

4.1 Case I: Humans

The majority (50/99) of cases examined in the literature review were based on image data that involved counting humans in various situations. Results from human image analysis are utilized in the fields of sociology, psychology or public safety (Loy et al., 2013). Especially in situations where there is a high density of humans in small spaces, an automated algorithmic counting procedure is needed since it cannot be done in a manual fashion, which can lead to critical situations in the case of public safety. Taking into account image parameters like perspective or overlapping, standard procedures fail to comply with requirements that are set to guarantee safety for large crowds (Sindagi and Patel, 2017). Observing small and controllable scenarios like bus stops or elevators, we can use segmentation or detection-based models (Schofield et al., 1996; Garcia-Bunster et al., 2012). However, when we encounter a scenario with large human crowds in large open spaces feature-based methods or related methods that involve the perspective are applied (Kang, Dhar and Chan, 2016). The datasets representing the human case are chosen in a way that they increase in “difficulty”, which is reflected by increasing degrees of overlapping, background changes and an increasing number of objects, as can be seen in Figure 4.

The **UCSD**[1] dataset contains 2000 surveillance camera images (238x158) in a pedestrian area. The people are marked by centroid coordinates in 20% of the images, while the rest of labelling was done by linear interpolation (Chan, Liang and Vasconcelos, 2008). The images contain 49.885 persons (= 15 persons on average per image) from one perspective with a marked region of interest. The static perspective enables the application of static background methods as well as detection-based methods due to the rather small number of people per image, although there is some minor overlapping in some images.

The **MALL**[2] dataset was made publicly available due to the missing variations in the USCD dataset. The dataset contains images from surveillance cameras within a shopping mall that are characterized by difficult lighting and hue, reflecting surfaces and strong variation in number of persons and movement patterns. In addition to that, the images contain skewed perspectives as well as several objects like plants that can cover up humans on the images. The datasets consists of 2000 images (320x240) with 62.325 persons (=31 persons on average per image) marked by centroid coordinates. The large amount of overlapping due to the high density of objects suggests using feature-based methods. Although the perspective does not change in the case of a security camera, the background extraction could be subject to fail calculations when large groups of people decide to not move for a couple of minutes and are therefore classified as background.

The **WorldExpo 10**[3] dataset focuses on cross-scene counting, where 3980 images were taken from 108 security cameras resulting in 199.923 persons being marked by centroid coordinate across multiple scenes (= 50 persons on average per image). The dataset is characterised by skewed perspectives and strong variation in movement patterns, caused by the nature of an exposition. In comparison to the MALL dataset, we have multiple perspectives so that a background extraction cannot be achieved. Methods working with a dynamic background are therefore suggested. In addition to that, large crowds in exposition queues suggest that overlapping and large numbers of people occur in at least some of the images so that feature-based methods are applicable.

Another example is given by the **Shanghaitech[4]** dataset. This dataset consists of 1198 coordinate annotated images with 330.164 person (=276 persons on average per image). The images are divided into two subset Part_A contains 482 publicly available internet images and Part B contains 716 images from the center of Shanghai. The challenges of this dataset are a very high variance of objects per image and a non-uniform distribution of object density levels. Given those challenges and the scenario, the images cannot be divided into foreground and background, which together with the high level of overlapping and high densities implies the use of feature-based models.

In summary, we can state that for this task in most of the cases feature-based methods are used, resulting in the application of multi-step neuronal networks that help to auto-craft the features rather than manually engineer them. It is found that in general DLS algorithms can help with image feature engineering, substantially reducing the effort for manual feature annotation or generation and therefore are suitable for those kind of cases despite the harsh requirements of the counting scenarios that need to be met. Figures 4 (a)-(d) show the projection of those four datasets onto the dimensions of the developed taxonomy.

---

[1] http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm

[2] http://personal.ie.cuhk.edu.hk/~ccloy/downloads_mall_dataset.html

[3] http://www.ee.cuhk.edu.hk/~xgwang/expo.html

[4] http://cs-chan.com/downloads_crowd_dataset.html

Dimension	Characteristics				
Architecture	One-Step	**Multiple Steps (ANN)**	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	**Centroid Coordinate**
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	**Detection-based**	**Feature-based**	Cluster-based		

(a) **Case I (Humans):** UCSD

Dimension	Characteristics				
Architecture	One-Step	**Multiple Steps (ANN)**	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	**Centroid Coordinate**
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	**Feature-based**	Cluster-based		

(b) **Case I (Humans):** MALL

Dimension	Characteristics				
Architecture	One-Step	**Multiple Steps (ANN)**	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	**Centroid Coordinate**
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	**Feature-based**	Cluster-based		

(c) **Case I (Humans):** WorldExpo 10

Dimension	Characteristics				
Architecture	One-Step	**Multiple Steps (ANN)**	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	**Centroid Coordinate**
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	**Feature-based**	Cluster-based		

(d) **Case I (Humans):** Shanghaitech

Figure 4. Projected characteristics for counting humans (CP = solid shade, SOL = striped shade)

4.2 Case II: Cells

Counting cells from microscopic images is an important indicator for disease diagnostics (Nazlibilek et al., 2014). The process of manually counting the cells is very cumbersome and depends largely on the expertise of the medical associate who is conducting the blood test (Adagale and Pawar, 2013). In the context of cells, we are often times interested in a specific type of cell so that cell type classification often times precedes the counting process (Swolin et al., 2003). We can assume that the majority of cell images have a static background and small to none overlapping so that all types of counting can be applied. Figure 5 depicts some examples of cell images.

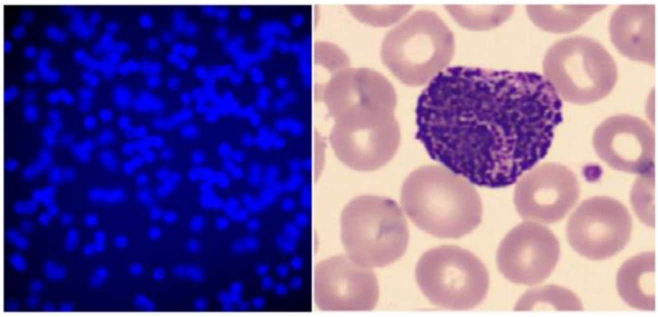

Figure 5. Artificial VGG Cells image (left) and microscopic cell image (right)

The **VGG Cells**[5] dataset consists of 200 computer generated images (256x256) with 174 annotated cells per image on average (Cohen et al., 2017). The images are generated to be able to simulate blur, pixel errors and light falloff and are annotated with centroid coordinate and for the purpose of cell margin detection with exact fitting bounding boxes. The advantage of using computer generated images is that the true cell count is present in the database and the dataset can therefore be used as a stand-

[5] http://www.robots.ox.ac.uk/~vgg/research/counting/index_org.html

ard benchmark for cell counting methods. Figure 6 (a) shows the taxonomy projection for the VGG cells dataset.

Dimension	Characteristics				
Architecture	One-Step	Multiple Steps (ANN)	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	Centroid Coordinate
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	Feature-based	Cluster-based		

(a) **Case II (Cells):** VGG Cells

Dimension	Characteristics				
Architecture	One-Step	Multiple Steps (ANN)	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	Centroid Coordinate
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	Feature-based	Cluster-based		

(b) **Case III (Cars):** TRANCOS

Dimension	Characteristics				
Architecture	One-Step	Multiple Steps (ANN)	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	Centroid Coordinate
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	Feature-based	Cluster-based		

(c) **Case III (Cars):** COWC

Dimension	Characteristics				
Architecture	One-Step	Multiple Steps (ANN)	Multiple Steps (ANN+other)		
Number of Objects	0-15	15-50	50-200	200-1000	1000<
Density of Objects	No Overlapping	Overlapping			
Background Dynamic	Static	Dynamic			
Output	Estimation Value	Absolute Count	Absolute Count by Region	Bounding Box	Centroid Coordinate
Training Data	Absolute Count	Centroid Coordinate	No label for counting		
Type of Counting	Detection-based	Feature-based	Cluster-based		

(d) **Case IV (Plants):** LCC 2017

Figure 6. Projected characteristics for cells, cars and plants (CP = solid shade, SOL = striped shade)

4.3 Case III: Cars

The next case to be described is the counting of cars. Traffic surveillance systems that are used to prevent accidents, guarantee traffic safety, maintain traffic flow and prevent parking problems have to cope with an increasing number of cars, which leads to the demand for intelligent traffic surveillance system (Biswas et al., 2017). As seen in the previous cases, we also have varying difficulty levels in traffic image data with regard to counting. Figure 7 (left) shows an image from a traffic live cam and Figure 7 (right) shows image data recorded from a satellite.

The **TRANCOS**[6] dataset is based on 1244 images of different streets in the metropolitan area of Madrid with 46.796 centroid coordinate annotations of cars (= 38 cars on average per image). In addition to that, every image has a marked region of interest to indicate the actual street (Guerrero-Gómez-Olmedo et al., 2015). There is a strong degree of overlapping in traffic jam situations that occur mostly in rush hour traffic. In comparison to that, only a few cars are to be seen in the late evening or morning hours. This variation and the possible degree of overlapping leads to feature-based methods once again. In comparison to other scenarios, however, foreground extraction is possible since the perspective of static traffic cams does not change in this case. The classifiers to distinguish fore- and background need to be able to deal with the different hours of the day and cloud situations. The **COWC**[7] satellite image dataset contains 32.176 annotated cars on 32 HD images (= 1000 cars on average per image) from different parts of the world like Canada, New Zealand or Germany (Mundhenk et al., 2016).

[6] http://agamenon.tsc.uah.es/Personales/rlopez/data/trancos/

[7] https://gdo152.llnl.gov/cowc/

Figure 7. Traffic cam image (left) and satellite image of parking spaces (right)

Other objects that are similar to cars (e.g., boats, or trailers) can also be identified in the images. The size of a car in an image is between 24 and 48 pixels. By having an aerial perspective, overlapping is not present in the images. That enables the use of detection-based methods. Those methods can be trained further by providing positive and negative examples of the entity "car". Since every perspective is unique, there is no possibility of using static background methods. In general depending on the perspective, both feature and detection-based methods can be used in this case. Figures 6 (b)-(c) show the taxonomy projection for the two car datasets.

4.4 Case IV: Plants

The counting of plants can be divided into different categories. The first category is counting the number of fruit for yield prognosis and optimizing the harvesting process (S. W. Chen et al., 2017). The second category involves counting plants or trees from an aerial perspective for the estimation of water and fertilizer supply (Cheang, Cheang and Tay, 2017). Traditionally, drones are used to record high resolution images of farm areas. The downside to this image recoding process is the limited altitude capability of the drone in comparison to satellite data, which can cover a larger area but provides a lower resolution and is hard to update. The last category involves counting leaves for determination of the plant's phenotype which helps with predicting growth intervals and crop yield (Dobrescu, Giuffrida and Tsaftaris, 2017). This process is usually carried out by manually using image manipulation software (Aich and Stavness, 2017). Some example images of those categories are shown in Figure 8.

The **LCC 2017**[8] dataset acts as a standard benchmark for the classification of plants with the subtask of leave counting being part of the CVPPP challenge. The dataset contains 27 images of Tabaco plants and 783 images of Arabidopsis plants. Every image contains a plant in the centre and the number of leaves is given for a training dataset. In addition to that, each pixel is annotated with a binary classification of whether it belongs to the plant or not. Those inputs can be used to train algorithms with different output structures (see Section 3.2.5). The binary pixel annotation can be used to determine the foreground part of the image using semantic segmentation (Aich and Stavness, 2017). In general, the images of plants show a high degree of overlapping of leaf objects so that feature-based methods using foreground/background separation are the preferred method. Figure 6 (d) shows the taxonomy projection for the LCC 2017 dataset. Summarizing the case study results, we can state that the feature-based models are widely used for counting objects, especially in situations with a high degree of overlapping and high object density. This lies within the nature of using global and/or local features to provide pixel counts rather than identifying each single object. However, feature-based methods require an extensive feature engineering process to be successful. This process can be shortened substantially by using DLS architectures like combined or single CNN and LSTM networks that already generate higher order features in the subsequent layers they provide within their architecture.

[8] https://www.plant-phenotyping.org/CVPPP2017-challenge

Figure 8. Image of oranges hanging on a tree (left), satellite image of a palm plantation (center), image of leaves from a plant (right)

5 Conclusion & Outlook

The work presented in this paper aimed to contribute to the knowledge base of both machine learning and information systems by developing a taxonomy for the general problem of counting objects from image data. The presented taxonomy poses a contribution to descriptive theory, as information systems that (partially) deal with counting problems can be constructed based on the concise overview and the suggested connections between the solution and the problem space of counting objects. We developed the taxonomy by first conducting a literature review and extracting information on both the used method as well as the nature of the counting problem. We then transformed those concepts into dimensions and characteristics of a taxonomy for counting objects with DLS. The taxonomy was then showcased by projecting four cases containing multiple publicly available datasets onto the taxonomy space. Limitations of this work include the narrow focus on deep learning models and the focus on single image counting. In addition to that, the taxonomy based on its meta characteristic is focused on the application of counting methods to put it into a business context, as proposed by IS research. Therefore, the dimensions lack technical details with regard to the algorithms and could be further improved by making more detailed distinctions. Future research should address these gaps and especially provide an extension of the taxonomy for image sequences.

Given the digitization of manufacturing and service processes, where large amounts of unstructured image data can be produced, the topic of counting objects will gain further importance for diagnostics, error prevention and process improvement. Besides the development of new algorithms, the feasibility of inference is of great importance for further research projects, especially when employing methods that require a lot of computational power like Deep Learning on the one hand but on the other hand are required to support critical decisions within a short period of time.

References

Achler, T., D. C. Vural and E. Amir. (2009). “Counting objects with biologically inspired regulatory-feedback networks.” In: 2009 International Joint Conference on Neural Networks (pp. 36–40).

Adagale, S. S. and S. S. Pawar. (2013). “Image segmentation using PCNN and template matching for blood cell counting.” In: 2013 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1–5).

Agarwal, R. and V. Dhar. (2014). “Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research.” Information Systems Research, 25(3), 443–448.

Aich, S. and I. Stavness. (2017). “Leaf Counting with Deep Convolutional and Deconvolutional Networks.” ArXiv:1708.07570 [Cs].

Aik, L. E. and Z. Zainuddin. (2009). “A kernelized Probabilistic Neural Network approach for counting pedestrians.” In: 2009 IEEE International Symposium on Industrial Electronics (pp. 2065–2068).

Amin, I. J., A. J. Taylor, F. Junejo, A. Al-Habaibeh and R. M. Parkin. (2008). "Automated people-counting by using low-resolution infrared and visual cameras." Measurement, 41(6), 589–599.

Arel, I., D. C. Rose and T. P. Karnowski. (2010). "Deep machine learning-a new frontier in artificial intelligence research [research frontier]." IEEE Computational Intelligence Magazine, 5(4), 13–18.

Bertucco, L., G. Nunnari, C. Randieri, V. Rizza and A. Sacco. (1998). "A cellular neural network based system for cell counting in culture of biological cells." In: Proceedings of the 1998 IEEE International Conference on Control Applications (Cat. No.98CH36104) (Vol. 1, pp. 341–345 vol.1).

Biswas, D., Hongbo Su, C. Wang, J. Blankenship and A. Stevanovic. (2017). "An Automatic Car Counting System Using OverFeat Framework." Sensors (14248220), 17(7), 1–13.

Boominathan, L., S. S. S. Kruthiventi and R. V. Babu. (2016). "CrowdNet: A Deep Convolutional Network for Dense Crowd Counting." ArXiv:1608.06197 [Cs].

Capobianco, S. and S. Marinai. (2017). "Deep neural networks for record counting in historical handwritten documents." Pattern Recognition Letters.

Chan, A. B., Z.-S. J. Liang and N. Vasconcelos. (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).

Cheang, E. K., T. K. Cheang and Y. H. Tay. (2017). "Using Convolutional Neural Networks to Count Palm Trees in Satellite Images." ArXiv:1701.06462 [Cs].

Chen, H., R. H. L. Chiang and V. C. Storey. (2012). "Business Intelligence and Analytics: From Big Data to Big Impact." MIS Q., 36(4), 1165–1188.

Chen, S. W., S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, … V. Kumar. (2017). "Counting Apples and Oranges With Deep Learning: A Data-Driven Approach." IEEE Robotics and Automation Letters, 2(2), 781–788.

Cho, S.-Y., T. W. S. Chow and C.-T. Leung. (1999). "A neural-based crowd estimation by hybrid global learning algorithm." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(4), 535–541.

Chow, T. W. S. and S.-Y. Cho. (2002). "Industrial neural vision system for underground railway station platform surveillance." Advanced Engineering Informatics, 16(1), 73–83.

Cohen, J. P., G. Boucher, C. A. Glastonbury, H. Z. Lo and Y. Bengio. (2017). "Count-ception: Counting by Fully Convolutional Redundant Counting." ArXiv:1703.08710 [Cs, Stat].

Dobrescu, A., M. V. Giuffrida and S. A. Tsaftaris. (2017). "Leveraging multiple datasets for deep leaf counting." ArXiv:1709.01472 [Cs].

Embleton, K. V., C. E. Gibson and S. I. Heaney. (2003). "Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method." Journal of Plankton Research, 25(6), 669–681.

Ferrari, A., S. Lombardi and A. Signoroni. (2017). "Bacterial colony counting with Convolutional Neural Networks in Digital Microbiology Imaging." Pattern Recognition, 61(Supplement C), 629–640.

French, G., M. Fisher, M. Mackiewicz and C. Needle. (2015). "Convolutional Neural Networks for Counting Fish in Fisheries Surveillance Video" (pp. 7.1-7.10). British Machine Vision Association.

Garcia-Bunster, G., M. Torres-Torriti and C. Oberli. (2012). "Crowded pedestrian counting at bus stops from perspective transformations of foreground areas." IET Computer Vision, 6(4), 296–305.

Goodfellow, I., Y. Bengio and A. Courville. (2016). Deep learning. Cambridge, Massachusetts: MIT Press.

Gregor, S. (2006). "The nature of theory in information systems." MIS Quarterly, 611–642.

Guerrero-Gómez-Olmedo, R., B. Torre-Jiménez, R. López-Sastre, S. Maldonado-Bascón and D. Oñoro-Rubio. (2015). "Extremely overlapping vehicle counting." In: Iberian Conference on Pattern Recognition and Image Analysis (pp. 423–431). Springer.

Han, K., W. Wan, H. Yao and L. Hou. (2017). "Image Crowd Counting Using Convolutional Neural Network and Markov Random Field." ArXiv:1706.03686 [Cs].

Hao Lu, Zhiguo Cao, Yang Xiao, Bohan Zhuang and Chunhua Shen. (2017). "TasselNet: counting maize tassels in the wild via local counts regression network." Plant Methods, 13, 1–17.

Heinrich, K. and V. Fleissner. (2018). “Deep Intelligent Systems for Time Series Prediction: Champion or Lame Duck? - Evidence from Crude Oil Price Prediction.” AMCIS 2018 Proceedings.

Hou, Y. L. and G. K. H. Pang. (2008). “Automated people counting at a mass site.” In: 2008 IEEE International Conference on Automation and Logistics (pp. 464–469).

Hu, X., H. Zheng, Y. Chen and L. Chen. (2015). “Dense crowd counting based on perspective weight model using a fisheye camera.” Optik - International Journal for Light and Electron Optics, 126(1), 123–130.

Ikeda, N., A. Saitoh, T. Isokawa, N. Kamiura and N. Metsui. (2008). “A neural network approach for counting pedestrians from video sequence images.” In: 2008 SICE Annual Conference (pp. 2485–2488).

Jöhnk, J., M. Röglinger, M. Thimmel and et al. (2017). “How to Implement Agile IT Setups: A Taxonomy of Design Options.” In: 25th European Conference on Information Systems 2017. Proceedings (p. 15).

Kang, D., D. Dhar and A. B. Chan. (2016). “Crowd Counting by Adapting Convolutional Neural Networks with Side Information.” ArXiv:1611.06748 [Cs].

Kang, D., Z. Ma and A. B. Chan. (2017). “Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks - Counting, Detection, and Tracking.” ArXiv:1705.10118 [Cs].

Laki, A., I. Rattalino, F. Corinto, K. Iván, D. Demarchi and P. Civera. (2011). “An integrated LOC hydrodynamic focuser with a CNN-based camera system for cell counting application.” In: 2011 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 301–304).

Lal, S., S. K. Behera, P. K. Sethy and A. K. Rath. (2017). “Identification and counting of mature apple fruit based on BP feed forward neural network.” In: 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS) (pp. 361–368).

Larochelle, H., D. Erhan, A. Courville, J. Bergstra and Y. Bengio. (2007). “An empirical evaluation of deep architectures on problems with many factors of variation.” In: Proceedings of the 24th international conference on Machine learning (pp. 473–480). ACM.

Lempitsky, V. and A. Zisserman. (2010). “Learning to count objects in images.” In: Advances in Neural Information Processing Systems (pp. 1324–1332).

Loy, C. C., K. Chen, S. Gong and T. Xiang. (2013). “Crowd Counting and Profiling: Methodology and Evaluation.” In: S. Ali, K. Nishino, D. Manocha, & M. Shah (Eds.), Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective (pp. 347–382). New York, NY: Springer New York.

Manik, S., L. M. Saini and N. Vadera. (2016). “Counting and classification of white blood cell using Artificial Neural Network (ANN).” In: 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES) (pp. 1–5).

Marsden, M., K. McGuinness, S. Little, C. E. Keogh and N. E. O’Connor. (2017). “People, Penguins and Petri Dishes: Adapting Object Counting Models To New Visual Domains And Object Types Without Forgetting.” ArXiv:1711.05586 [Cs].

Mundhenk, T. N., G. Konjevod, W. A. Sakla and K. Boakye. (2016). “A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning.” ArXiv:1609.04453 [Cs].

Nazeer, M., M. S. Wong and J. E. Nichol. (2017). “A new approach for the estimation of phytoplankton cell counts associated with algal blooms.” Science of The Total Environment, 590–591(Supplement C), 125–138.

Nazlibilek, S., D. Karacor, T. Ercan, M. H. Sazli, O. Kalender and Y. Ege. (2014). “Automatic segmentation, counting, size determination and classification of white blood cells.” Measurement, 55(Supplement C), 58–65.

Nickerson, R. C., U. Varshney and J. Muntermann. (2013). “A method for taxonomy development and its application in information systems.” European Journal of Information Systems, 22(3), 336–359.

Onoro-Rubio, D. and R. J. López-Sastre. (2016). “Towards perspective-free object counting with deep learning.” In: European Conference on Computer Vision (pp. 615–629). Springer.

Pavlidis, I., V. Morellas and N. Papanikolopoulos. (2000). “A vehicle occupant counting system based on near-infrared phenomenology and fuzzy neural classification.” IEEE Transactions on Intelligent Transportation Systems, 1(2), 72–85.

Schofield, A. J., P. A. Mehta and T. J. Stonham. (1996). "A system for counting people in video images using neural networks to identify the background scene." Pattern Recognition, 29(8), 1421–1428.

Shao, J., D. Wang, X. Xue and Z. Zhang. (2015). "Learning to Point and Count." ArXiv:1512.02326 [Cs].

Sharma, R., S. Mithas and A. Kankanhalli. (2014). "Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations." European Journal of Information Systems, 23(4), 433–441.

Shen, F., J. Chao and J. Zhao. (2015). "Forecasting Exchange Rate Using Deep Belief Networks and Conjugate Gradient Method." Neurocomput., 167(C), 243–253.

Sindagi, V. A. and V. M. Patel. (2017). "A survey of recent advances in CNN-based single image crowd counting and density estimation." Pattern Recognition Letters.

Song, D., Y. Qiao and A. Corbetta. (2017). "Depth driven people counting using deep region proposal network." In: 2017 IEEE International Conference on Information and Automation (ICIA) (pp. 416–421).

Swolin, B., P. Simonsson, S. Backman, I. Löfqvist, I. Bredin and M. Johnsson. (2003). "Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks – evaluation of DiffMasterTM Octavia." Clinical & Laboratory Haematology, 25(3), 139–147.

Tang, S., Z. Pan and X. Zhou. (2017). "Low-Rank and Sparse Based Deep-Fusion Convolutional Neural Network for Crowd Counting." Mathematical Problems in Engineering, 1–11.

Usaj, M., D. Torkar, M. Kanduser and D. Miklavcic. (2011). "Cell counting tool parameters optimization approach for electroporation efficiency determination of attached cells in phase contrast images." Journal of Microscopy, 241(3), 303–314.

Webster, J. and R. T. Watson. (2002). "Analyzing the past to prepare for the future: Writing a." MIS Quarterly, 26(2), 13–23.

Yao, H., K. Han, W. Wan and L. Hou. (2017). "Deep Spatial Regression Model for Image Crowd Counting." ArXiv:1710.09757 [Cs].

Yifei Xue, Tiejun Wang and A. K. Skidmore. (2017). "Automatic Counting of Large Mammals from Very High Resolution Panchromatic Satellite Imagery." Remote Sensing, 9(9), 1–16.

Yoshinaga, S., A. Shimada and R. Taniguchi. (2010). "Real-time people counting using blob descriptor." Procedia - Social and Behavioral Sciences, 2(1), 143–152.

Zeng, L., X. Xu, B. Cai, S. Qiu and T. Zhang. (2017). "Multi-scale Convolutional Neural Networks for Crowd Counting." ArXiv:1702.02359 [Cs].

Zhang, C. (2015). "Design of Human Capability Maturity Analysis System Online P-CMM Model" (pp. 302–305). IEEE.

Zhang, Y., D. Zhou, S. Chen, S. Gao and Y. Ma. (2016). "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network." In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 589–597).

Zhou, X., X. Xiao and C. Ma. (2010). "A study of automatic recognition and counting system of urine-sediment visual components." In: 2010 3rd International Conference on Biomedical Engineering and Informatics (Vol. 1, pp. 78–81).

Zschech, P. (2018). "A TAXONOMY OF RECURRING DATA ANALYSIS PROBLEMS IN MAINTENANCE ANALYTICS." In: 26th European Conference on Information Systems 2018. Proceedings. Portsmouth.