

Crowdsourcing annotation system of object counting dataset for deep learning algorithm

Tjeng Wawan Cenggoro^{1,4}, Fidelson Tanzil¹, Ayu Hidayah Aslamiah¹, Ettikan Kandasamy Karuppiah³, Bens Pardamean^{2,4}

¹ Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

² Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia 11480

³ NVIDIA Corporation

⁴ Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia 11480

Email: wcenggoro@binus.edu

Abstract. Deep Learning is currently the state-of-the-art technique for various Computer Vision tasks, including object counting. Despite of its high performance, Deep Learning requires a gigantic amount of training data to show its best result. Getting this massive data in reasonable time requires a proper strategy such as crowdsourcing. However, in case of object counting, we found no crowdsourcing system able to effectively collect necessary data. To tackle this problem, we develop a crowdsourcing system to annotate image for object counting dataset. This system is also equipped with validation system to ensure the quality of collected dataset.

1. Introduction

In the last decade, the research of deep learning [1] for computer vision is significantly gain popularity. Almost all of computer vision researches used deep learning for algorithm analysis, especially in problems such as image classification [2-6], object detection [7-9], and object counting [10, 11]. The reason behind this popularity is the fact that Deep Learning algorithm is able to outperform other type machine learning algorithm in most computer vision tasks, given a massive training data. Due to the deep learning requirement for massive training data, an effective data collection strategy is necessary to acquire large data in a reasonable time. The most effective strategy that can be used to create this massive data is via crowdsourcing strategy, which utilizes large number of humans to annotate dataset, then validate the collected data. Especially in computer vision, the frequently used datasets, ImageNet for image classification [12] and Microsoft COCO [13] for object detection, are developed by using crowdsourcing strategy. To the extent of our knowledge, there is no effective crowdsourcing strategy for massive dataset in object counting problem. Even though there are dataset that can be used for object counting problem [14-17], the number of available data is still not as large as ImageNet or Microsoft COCO. Whilst, the object counting dataset that has been discussed above was not developed by crowdsourcing strategy, thus it is difficult to scale the data size and ensure its quality. In this research, we propose creation of information system that can be used to



collect dataset for object counting using crowdsourcing paradigm. The collected dataset is annotated via system using dots. This strategy is more feasible to be done for object counting case [10]. This system is implemented with validation strategy that is designed for object counting so that the quality of produced dataset can be guaranteed.

2. Related Works

The Crowdsourcing has become an important process in the research workflow of deep learning. The two most influencing dataset in deep learning now, ImageNet [12] and Microsoft COCO [13], were developed by crowdsourcing paradigm. ImageNet can even be considered as essential dataset for deep learning in computer vision. There are many researches in computer vision using Convolutional Neural Network [18] which trained by ImageNet as the base for developing new model [19-22]. The popular usage of crowdsourcing paradigm is due to deep learning algorithm requirements for massive data training. This requirement is essential for deep learning to achieve optimal performance that is able to outperform other algorithms, which has been shown by Alex Krizhevsky in 2012 [2] when he won ImageNet, a large-scale images classification competition. Crowdsourcing paradigm enable researchers to collect data quicker, cheaper and easier [23], which is required by deep learning. While crowdsourcing has an ability to provide tremendous benefits for massive data, specific strategies are needed to validate the quality of collected data. In this case, manual data validation is not an option as it requires a lengthy process and laborious work for a massive data. To date, most validation strategy for crowdsourcing is used for categorical data [24-28]. Particularly in computer vision, validation strategy for crowdsourcing can only be found in image classification cases [12] and object detection [13, 29]. In contrast to object counting cases, currently available datasets are not constructed using crowdsourcing paradigm. For example, in dataset of UCF CC 50 [14] and TRANCOS [15], the dataset are produced by manual labor. Those process caused the collected data for both datasets are not large enough for training deep learning algorithm (50 for UCF CC 50 and 1244 images for TRANCOS). Other examples are UCSD Pedestrian dataset [17] and bacterial cell [16] which are created by using specific algorithm. The usage of algorithm enables massive dataset collection in a short time. However, the result will be biased to the underlying assumption in the employed algorithm. In regard to the information system built in this study, the most similar research is a study from the WalMartLabs team that built Chimera [30]. Chimera is an information system that combines machine learning, rule-based classification, and crowdsourcing for product categorization. Meanwhile, if we also consider non-academically-published annotation systems, Amazon Mechanical Turk (MTurk) is currently the most commonly used annotation system. MTurk can be used to annotate various dataset types, such as survey, image classification, and object detection. However, MTurk is not explicitly accommodating dataset annotation specifically for object counting.

3. System Design

The proposed system in this paper is designed with the goal to collect a massive object counting dataset in a reasonable time. Consequently, the system design heavily utilizes crowdsourcing paradigm. The annotation task in the system is also crafted specifically for creating object counting dataset. Therefore, this system is suitable for any organization that in a condition of requiring big data to train deep learning model for object counting.

Fig. 1 shows the design of the developed annotation system using use case and activity diagram. As depicted in fig. 1a, the system built in this research consists of four main modules: upload images, view statistics, discard/approve annotation, and mark object location on images. The workflow among these modules can be seen in fig. 1b. As depicted in fig. 1a, the actors in this system are divided into two types, administrator actor and worker actor. Workers are responsible to annotate images by marking it with dots, where each of these dots pointed to the center of each object in the image. This annotation task design requires only a little computer literacy, thus almost anyone who can operate computer is qualified to be worker. This task design is intended to recruit workers as much as possible, which is essential to accelerate crowdsourcing task.

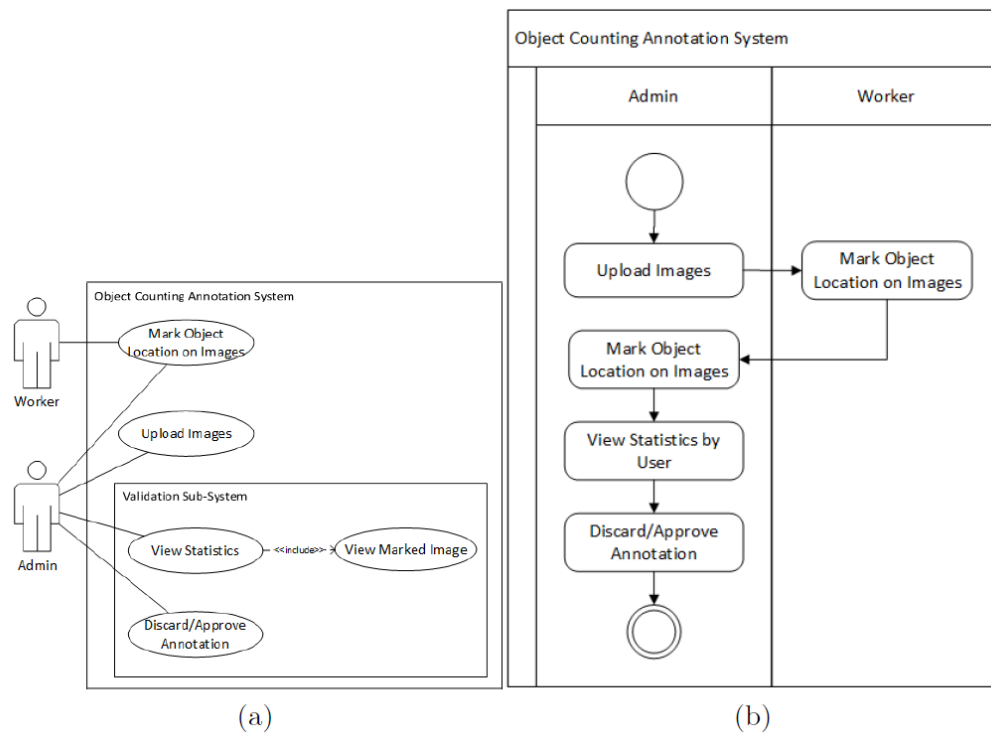


Figure 1: (a) Use Case Diagram of the Annotation System; (b) Activity Diagram of the Annotation System.

On the other hand, administrator is responsible to supply the system with images to be annotated and assess the quality of worker actors' annotations by using the validation sub-system. Administrator then have a right to whether keep the annotation submitted by worker or discard it based on the information given in the validation sub-system. Administrator can also annotate images as ground truths for validation.

The sub-system for validation in this annotation system is formed by two main modules: view statistics and discard/approve annotation. The view statistics module is used by administrator to judge the quality of each worker annotation. This module displays four charts: distance summary, dots count summary, distance to ground truth, and dots count differences to ground truth, as shown in fig. 2a to 2d. Based on these charts, administrator can assess five worst annotations submitted by a worker. The administrator can also view the marked images using view marked images module. If the administrator actor thinks that an annotation should be discarded, they can use discard/approve annotation module.

The distance summary chart in fig. 2a displays the average distance summary of all dots marked by a worker actor to the dots submitted by other worker in a same image. The blue and green bar indicates the average distance between an observed worker with two other workers. For the effectiveness of data collection, the annotation system is designed to allow an image to be marked by at most three workers. A single value in the line chart of distance summary indicates the average distance to all other workers marking the same image. For simplicity, the chart only shows top five submitted annotations sorted by the line chart value descending. The other submitted annotations can only be displayed after the top five annotations is discarded or approved by administrator. In a similar way to distance summary chart, the dots count summary chart displays the information of differences in total number of dots marked in an image by a worker compared to other workers, as shown in fig. 2b. For a more detailed information, the administrator can also display one of the marked images with markers from two other workers as in fig. 2e or with ground truth markers as in fig. 2f. For simplicity,

the we use coins as for the objects to be marked in the images, as shown in fig. 2f and 2e. Although the sample object we choose has a uniform shape, the proposed system can be generalized to any kind of object with arbitrary shape. The worker can mark the arbitrary-shaped object at the center of the object.

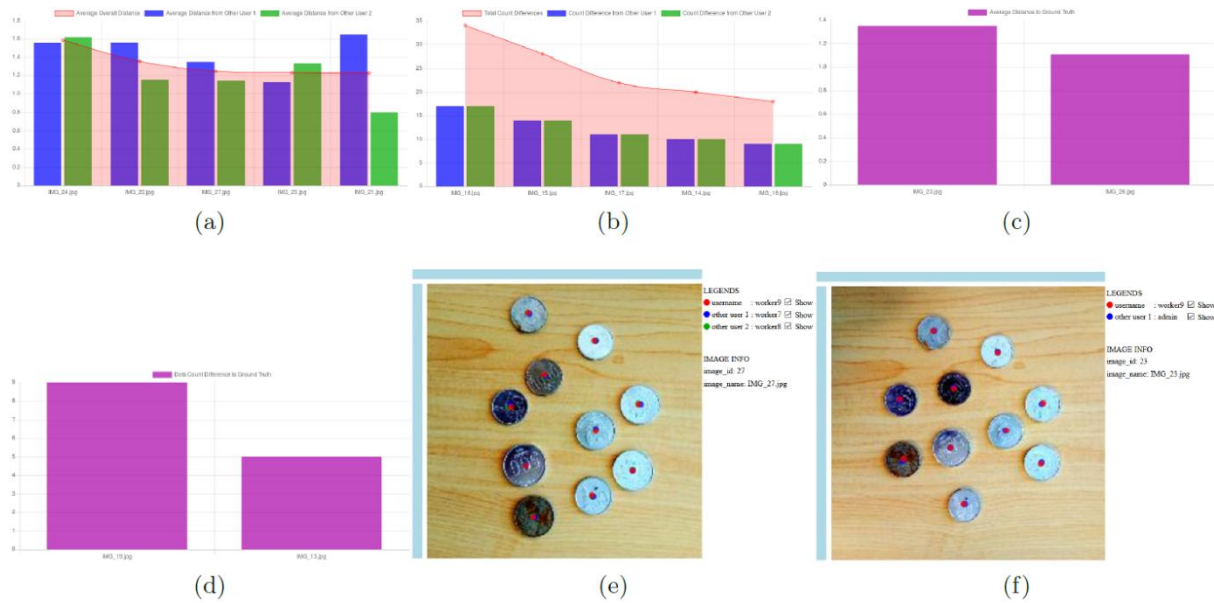


Figure 2: (a) Distance Summary Chart; (b) Dots Count Summary Chart; (c) Distance to Ground Truth Chart; (d) Dots Count Differences to Ground Truth Chart; (e) Marked Image View with Two Other Workers Markers; (f) Marked Image View with Ground Truth Markers.

The quality of a worker annotations can also be assessed by inferring the distance to ground truth and dots count differences to ground truth chart, as displayed respectively in fig. 2c and 2d. Each of these two charts displays same type of information to distance summary and dots count summary chart respectively. However, these two charts compare the observed worker annotations to the ground truths submitted for validation by administrator instead.

In the view statistics module, administrator can see the worst five annotations. By seeing only these five annotations, administrator can quickly discard any invalid annotation without the needs to check all data one by one. In addition, administrator can also discard or approve all annotations submitted by a worker. The choice for discarding/approving data one by one or all annotations can be decided by detecting these three cases:

- (i) worker only makes mistakes in a few data,
- (ii) worker abuses the system (annotate all data wrongly),
- (iii) worker correctly annotates all data

In case (i), administrator can discard data one by one so that only wrong annotations that are discarded. In case (ii), administrator needs to discard all data. In case (iii), administrator can accept all annotations. Given the vital role of these three cases, we can conclude that the ability to detect these cases can be used as a success/failure indicator of our validation system.

4. Results and Discussion

Fig. 3 shows the statistics of a worker with only one wrong annotation, which relates to case (i) in previous section. We can see that from fig. 3a and fig. 3a, only the first image shows significant

differences to the other two workers that annotate the same image. On the other hand, the statistics view in fig. 3c shows negligible errors in dots distance to ground truths (only about 1 pixel). The statistics view in fig. 3d also supports the dots distance to ground truths chart by showing no difference of dots count to ground truth images. We can see from the marked image shown in fig. 3e that the observed worker fails to mark four most bottom-right objects. From these statistics view, we can easily recognize that this is case (i), thus administrator can discard only wrong annotations.

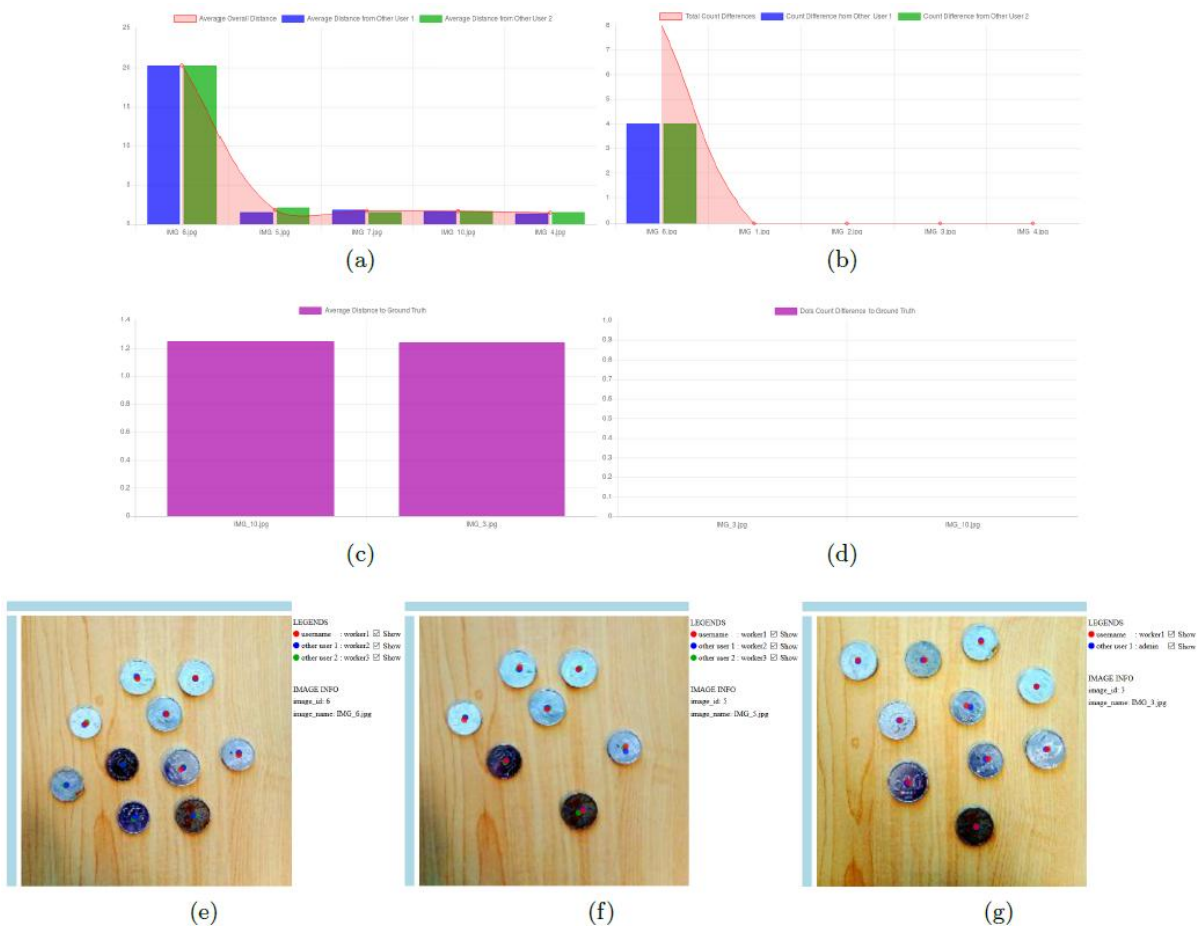


Figure 3: Statistics view for worker with one invalid annotation: (a) Distance Summary Chart; (b) Dots Count Summary Chart; (c) Distance to Ground Truth Chart; (d) Dots Count Differences to Ground Truth Chart; (e) Markers Comparison to Other Worker Actors for the First Image; (f) Markers Comparison to Other Workers for the Second Image; (g) Markers Comparison to One of the Ground Truths.

From the statistics view displayed in fig. 4, we can see clearly that the case belongs to case (ii). Fig. 4a and 4b shows that all five worst annotations have significant differences compared to other workers. Not to mention that there are also clear differences to ground truths as displayed in fig. 4c and 4d. The view of marked image in fig. 4e and 4f shows that the observed worker is spamming the annotation system by just randomly marking the images.

By looking at the statistics view in fig. 5, we can detect that the displayed case is case (iii). It is clear that all the worker annotations displayed in fig. 5 consistently show negligible differences to both other workers annotations and ground truths. Therefore, administrator can safely approve all the annotations.

In fig.6a and 6b, we can see that there are significant differences of the observed worker annotation to only the first other worker annotation for the first image. However, the statistics in other images in fig. 6a and 6b shows negligible differences. We can also see that the annotations of this observed worker actor are consistent with the provided ground truths as displayed in fig. 6c and 6d. Therefore, we can imply that the first other worker annotation for the first image is probably incorrect. Administrator can subsequently check the statistics of this worker to see if the pattern in fig. 3 or fig. 4 is emerging.

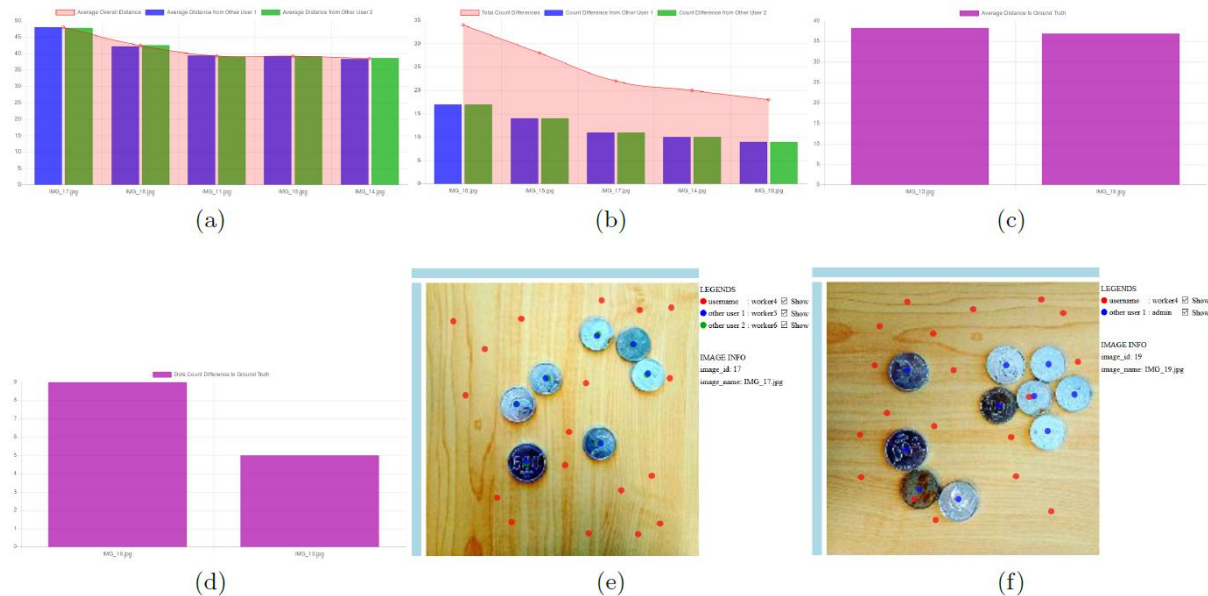


Figure 4: Statistics view for worker with all invalid annotations (a) Distance Summary Chart; (b) Dots Count Summary Chart; (c) Distance to Ground Truth Chart; (d) Dots Count Differences to Ground Truth Chart; (e) Markers Comparison to Other Workers; (f) Markers Comparison to Ground Truths.

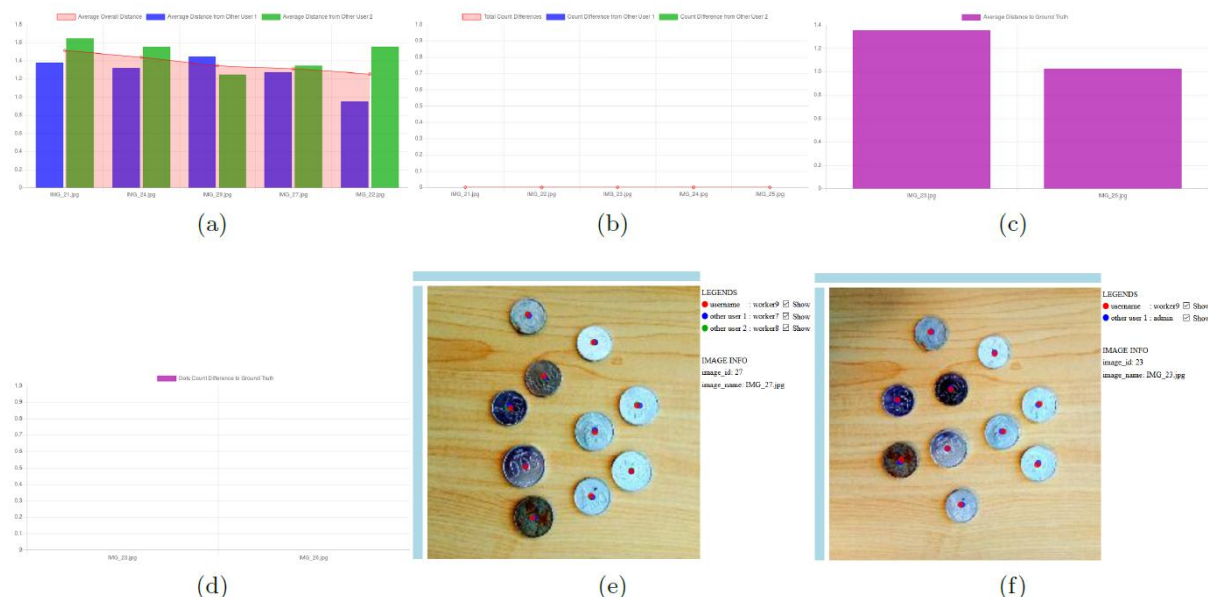


Figure 5: (a) Distance Summary Chart; (b) Dots Count Summary Chart; (c) Distance to Ground Truth Chart; (d) Dots Count Differences to Ground Truth Chart; (e) Marked Image View with Two Other Workers Markers; (f) Marked Image View with Ground Truth Markers.

We can see that by using the provided statistics, administrator can easily and quickly determine the quality of annotations, thus enable administrator to decide whether to accept the annotations, discard all annotations of a worker, or discard annotations one by one starting from the worst annotation. This enables administrator to effectively validate the collected dataset without the needs to check all of annotations one by one, which is not practical for a massive dataset. The quality of validated dataset can also be guaranteed because administrator can easily discard annotations starting from the worst quality.

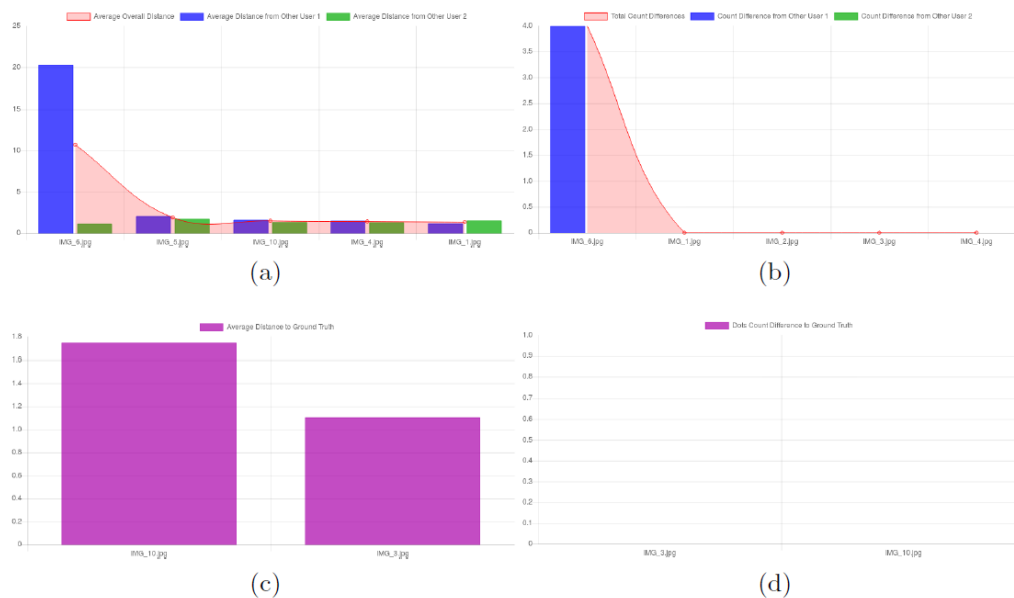


Figure 6: Statistics view for invalid annotation from other worker (a) Distance Summary Chart; (b) Dots Count Summary Chart; (c) Distance to Ground Truth Chart; (d) Dots Count Differences to Ground Truth Chart.

5. Conclusion

In this paper, we show a design of effective annotation system for collecting large object counting dataset. The effectiveness of the system is attained by combining crowdsourcing strategy with a specially-designed system for object counting dataset validation. We show that the developed system enables administrator to validate a large-scale dataset without requiring a laborious work. This success is indicated by the ability of validation system to show cases that guides administrator to effectively discard/approve annotations. For the future, it would be interesting to compare the validation strategy in this paper with other related validation strategies.

Acknowledgement

This research is funded by Bina Nusantara University Applied Technology Grant 2018. We would like to give our gratitude to NVIDIA Bina Nusantara University AI R&D Center for a useful discussion on annotation system management. We also thank to Lecturer Specialists Team, School of Computer Science, Bina Nusantara University, for their input to this paper.

References

- [1] LeCun Y, Bengio Y and Hinton G 2015 Nature 521 436-444 ISSN 0028-0836 (Preprint arXiv:1312.6184v5) URL <http://dx.doi.org/10.1038/nature14539>
- [2] Krizhevsky A and Hinton G E 2012 Advances in Neural Information Processing Systems

- [3] Simonyan K and Zisserman A 2015 The International Conference on Learning Representations 2015 1-14 (Preprint 1409.1556) URL <http://arxiv.org/abs/1409.1556>
- [4] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol 07-12-June pp 1-9 ISBN 9781467369640 ISSN 10636919 (Preprint 1409.4842)
- [5] He K, Zhang X, Ren S and Sun J 2016 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) vol 7 (IEEE) pp 770-778 ISBN 978-1-4673-8851-1 ISSN 1664-1078 (Preprint 1512.03385) URL <http://ieeexplore.ieee.org/document/7780459/>
- [6] Szegedy C, Iosifidou S, Vanhoucke V and Alemi A 2016 arXiv preprint 42 11-24 (Preprint 1602.07261) URL <http://arxiv.org/abs/1602.07261>
- [7] Ren S, He K, Girshick R and Sun J 2017 IEEE Transactions on Pattern Analysis and Machine Intelligence 39 1137-1149 ISSN 01628828 (Preprint 1506.01497)
- [8] Redmon J, Divvala S, Girshick R and Farhadi A 2016 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE) pp 779-788 ISBN 978-1-4673-8851-1 ISSN 01689002 (Preprint 1506.02640) URL <http://arxiv.org/abs/1506.02640> <http://ieeexplore.ieee.org/document/7780460/>
- [9] He K, Gkioxari G, Dollar P and Girshick R 2017 2017 IEEE International Conference on Computer Vision (ICCV) (IEEE) pp 2980-2988 ISBN 978-1-5386-1032-9 ISSN 0006-291X (Preprint 1703.06870) URL <http://arxiv.org/abs/1703.06870> <http://ieeexplore.ieee.org/document/8237584/>
- [10] Lempitsky V and Zisserman A 2010 Advances in Neural Information Processing Systems 1324-1332 ISSN 0956-7976 URL <http://papers.nips.cc/paper/4043-learning-to-count-objects-in-images.pdf>
- [11] Oñoro-Rubio D and Lopez-Sastre R J 2016 ECCV 2016: 14th European Conference, Amsterdam, The Netherlands (Lecture Notes in Computer Science vol 9911) ed Leibe B, Matas J, Sebe N and Welling M (Cham: Springer International Publishing) pp 615-629 ISBN 978-3-319-46477-0 URL http://link.springer.com/10.1007/978-3-319-46478-7_38
- [12] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg A C and Fei-Fei L 2015 International Journal of Computer Vision 115 211-252 ISSN 15731405 (Preprint 1409.0575) URL <http://dx.doi.org/10.1007/s11263-015-0816-y>
- [13] Lin T y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P and Zitnick C L 2014 European Conference on Computer Vision 2014 pp 740-755 (Preprint 1405.0312) URL <http://arxiv.org/abs/1405.0312> http://link.springer.com/10.1007/978-3-319-10602-1_48
- [14] Idrees H, Saleemi I, Seibert C and Shah M 2013 IEEE Conference on Computer Vision and Pattern Recognition (IEEE) pp 2547-2554 ISBN 978-0-7695-4989-7 URL <http://ieeexplore.ieee.org/document/6619173/>
- [15] Guerrero-Gomez-Olmedo R, Torre-Jimenez B, Lopez-Sastre R, Maldonado-Bascon S and Onoro-Rubio D 2015 Iberian Conference on Pattern Recognition and Image Analysis vol 2 pp 423-431 URL http://link.springer.com/10.1007/978-3-319-19390-8_48
- [16] Lehmussola A, Ruusuvuori P, Selinummi J, Huttunen H and Yli-Harja O 2007 IEEE Transactions on Medical Imaging 26 1010-1016 ISSN 0278-0062 URL <http://ieeexplore.ieee.org/document/4265752/>
- [17] Chan A B, Zhang-Sheng John Liang and Vasconcelos N 2008 2008 IEEE Conference on Computer Vision and Pattern Recognition (IEEE) pp 1-7 ISBN 978-1-4244-2242-5 URL <http://ieeexplore.ieee.org/document/4587569/>
- [18] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W and Jackel L D 1989 Backpropagation Applied to Handwritten Zip Code Recognition
- [19] Girshick R, Donahue J, Darrell T and Malik J 2014 2014 IEEE Conference on Computer Vision

- and Pattern Recognition (IEEE) pp 580-587 ISBN 978-1-4799-5118-5 (Preprint arXiv:1311.2524v5) URL <http://ieeexplore.ieee.org/document/6909475/>
- [20] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren M P and Ng A Y 2017 arXiv preprint (Preprint 1711.05225) URL <http://arxiv.org/abs/1711.05225>
- [21] Carreira J and Zisserman A 2017 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE) pp 4724-4733 ISBN 978-1-5386-0457-1 (Preprint 1705.07750) URL <http://arxiv.org/abs/1705.07750> <http://ieeexplore.ieee.org/document/8099985/>
- [22] Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhutdinov R, Zemel R and Bengio Y 2015 IEEE Transactions on Neural Networks 5 157-166 ISSN 10459227 (Preprint 1502.03044) URL <http://ieeexplore.ieee.org/document/279181/> <http://arxiv.org/abs/1211.5063> <http://arxiv.org/abs/1502.03044>
- [23] Snow R, O'Connor B, Jurafsky D and Ng A Y 2008 Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP '08 October (Morristown, NJ, USA: Association for Computational Linguistics) p 254 URL <http://portal.acm.org/citation.cfm?doid=1613715.1613751>
- [24] Whitehill J, Ruvolo P, Wu T, Bergsma J and Movellan J 2009 Advances in Neural Information Processing Systems 22 1-9
- [25] Liu Q, Peng J and Ihler A 2012 Advances in neural information processing systems 1-9
- [26] Hong C 2017 arXiv preprint (Preprint 1706.03930) URL <http://arxiv.org/abs/1706.03930>
- [27] Raykar V C and Yu S 2012 Journal of Machine Learning Research 13 491-518
- [28] Khan Khattak F and Salleb-Aouissi A 2011 Workshop Comput. Social Sci. Wisdom Crowds 1-5
- [29] Su H, Deng J and Fei-fei L 2012 Twenty-Sixth AAAI Conference on Artificial Intelligence pp 40-46
- [30] Sun C, Rampalli N, Yang F and Doan A 2014 Proceedings of the VLDB Endowment 7 1529-1540 ISSN 21508097 URL <http://pages.cs.wisc.edu/~anhai/papers/chimera-vldb14.pdf> <http://dl.acm.org/citation.cfm?doid=2733004.2733024>