ISSN (p) : 2407-1323
ISSN (e) : 2442-4404
DOI : https://doi.org/10.25124/jett.v9i1.4414
| Vol. 9 | No. 1 | Halaman : 1138 - 1144
JETT
Jurnal Elektro Telekomunikasi Terapan Juli 2022

A METHOD FOR PROBLEM SOLVING FOGGY CITYSCAPES IMBALANCED DATASET

Ahmad Alfan[1], Suryo Adhi Wibowo[2]

[1,2]Program Studi S1 Teknik Telekomunikasi, Fakultas Teknik Elekro, Telkom University
[1]ahmdalfan@student.telkomuniversity.ac.id, [2]suryoadhiwibowo@telkomuniverisity.ac.id

Abstract

Imbalanced dataset is the major problem we all will face in the process of developing a deep learning model. There were many approaches to solve this very problem such as heuristic data sampling and modifying loss function for model training. To find the solution, we chose the Foggy Cityscapes dataset for the experiment since this dataset has an imbalanced object class distribution. We proposed a method to solve the imbalanced dataset namely instance-level downsampling as an extension of the traditional downsampling method. The algorithm of this method will selectively keep and drop the certain image in the dataset by evaluating the majority and minority object class proportion inside a single image. After comparing the model evaluation using the Mean Average Precision (mAP) metric, the model which was trained with a balanced dataset has more balanced knowledge or is less biased across the object classes of interest.

Keywords: imbalanced dataset, data balancing, object detection, foggy cityscapes, deep learning

1. INTRODUCTION

One of the most crucial steps in Deep Learning model development is data preprocessing. Model is the compact version of our dataset. Hence, a dataset is the main source of knowledge for the model whether it leads to good or bad learning. Limited resource of data collection would lead to the lack of data in certain classes which is then known as imbalanced data. Imbalanced data has no good effect on the model itself. The majority class will dominate the model knowledge and make the model biased. Once the model is biased, for example in the object detection model, minority object class will be barely detected. In the case of Autonomous Vehicle (AV), undetected objects potentially break the action system decision and finally could harm the passenger.

This very problem has gotten big attention from researchers, and they had proposed different approaches. There are 2 main approaches which are traditional data balancing methods and weighted loss function methods such as Cross-Entropy Loss and Focal Loss. Traditional data balancing methods focus on refining the dataset by adding more data (upsampling) or removing certain data (downsampling). On the other hand, weighted loss function methods focus on the model learning process by giving the majority class less priority and the minority class more priority. Weighted loss function such as Focal Loss has been applied to many if not all object detection models right now. But, as our observations, if the data has a big difference in terms of amount, the weighted loss function would still not be able to balance the model performance between majority and minority class [1-6]. So, traditional data balancing methods would still come in handy.

Based on the explained problem, in this research, we are proposing a method which is the extension of traditional data balancing methods to increase the model generalization across all object classes of interest in the Foggy Cityscapes dataset [7], [8]. Instead of doing traditional sampling by randomly picking samples from the dataset as Utomo et. al. did in their research [2], we prefer to have more control by applying instance-level analysis against each image inside the dataset. This dataset holds thousands of synthetic foggy images with imbalanced object classes distribution, and we will only focus on car and person class. In the evaluation step, we will compare the performance

ISSN (p) : 2407-1323
ISSN (e) : 2442-4404
DOI : https://doi.org/10.25124/jett.v9i1.4414
| Vol. 9 | No. 1 | Halaman : 1138 - 1144

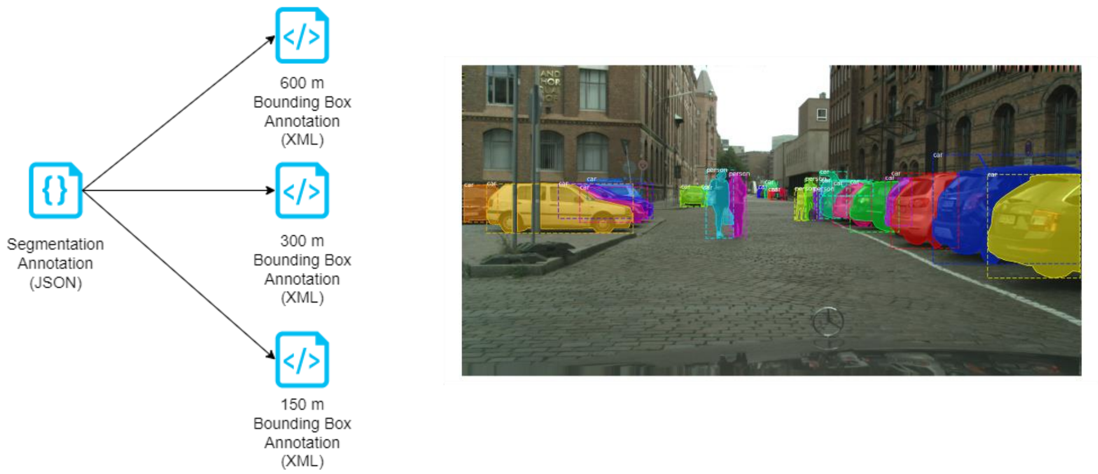
between the 2 EfficientDet D0 models. EfficientDet D0 is the smallest version of an efficient neural network architecture that has higher accuracy yet less computational complexity (low latency). Low latency makes it possible to do real-time inference which is an important requirement of video analysis for AV. Bidirectional Feature Pyramid Network (BiFPN) in EfficientDet also enables the model to detect different objects scales which can be difficult under hazy scenes. This research has an important contribution to the development of smart surveillance system and object detection performance in AV. The outcome of this research is an understanding of how our method can solve the imbalanced dataset problem by referring to Mean Average Precision (mAP) as the performance metric [9], [10].

2. SYSTEM DESIGN

2.1 Foggy Cityscapes Dataset

Foggy Cityscapes consists of 15,000 synthetic foggy images for train, validation, and test sets. The original clear images dataset has 5,000 images, each image synthesized to have 3 different levels of maximum visibility distance which are 600 m, 300 m, and 150 m as shown in Figure 1 [7], [8]. To train the object detection model using this dataset, we need to convert the segmentation annotation into bounding box annotation as shown in Figure 2(b). Since the annotation was created for the Cityscapes dataset (clear weather), we need to copy one annotation into 3 annotation files for 3 different visibility distance levels as shown in Figure 2(a).

Figure 1. Foggy Cityscapes dataset samples



a) Convert 1 JSON annotation into 3 XML

b) Segmentation to bounding box annotation

Figure 2. Annotation conversion

ISSN (p) : 2407-1323
ISSN (e) : 2442-4404
DOI : https://doi.org/10.25124/jett.v9i1.4414
| Vol. 9 | No. 1 | Halaman : 1138 - 1144
JETT
Jurnal Elektro Telekomunikasi Terapan Juli 2022

2.2 Similar Classes Merging

Foggy Cityscapes has many object classes that can be grouped as car or person as we can see in Table 1 [8]. Since we have fewer data of person class, merging rider class to person will make the dataset more balanced. Something to consider is to make sure any class that we will merge has a similar appearance [4].

Table 1. Foggy Cityscapes dataset class definitions

Group	Classes
flat	road, sidewalk, parking, rail track
human	person, rider
vehicle	car, truck, bus, on rails, motorcycle, bicycle, caravan, trailer
construction	building, wall, fence, guard rail, bridge, tunnel
object	pole, pole group, traffic sign, traffic light
nature	vegetation, terrain
sky	sky
void	ground, dynamic, static

2.3 Instance-Level Downsampling

Previous merging results in the 2.2 section can further be polished with instance-level downsampling. This is the method that we propose to make the dataset balanced. To understand the algorithm, Figure 3 shows the flowchart of this method. Figure 4 also shows an example of which image will be kept and dropped.

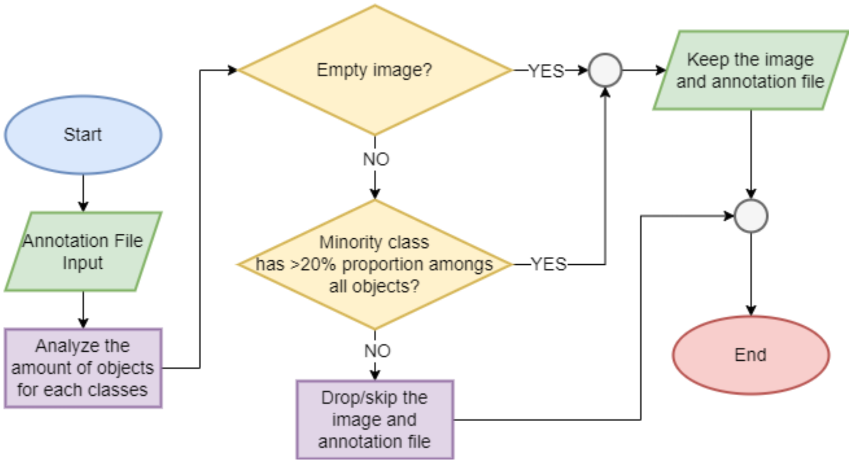


Figure 3. Instance-level downsampling flowchart

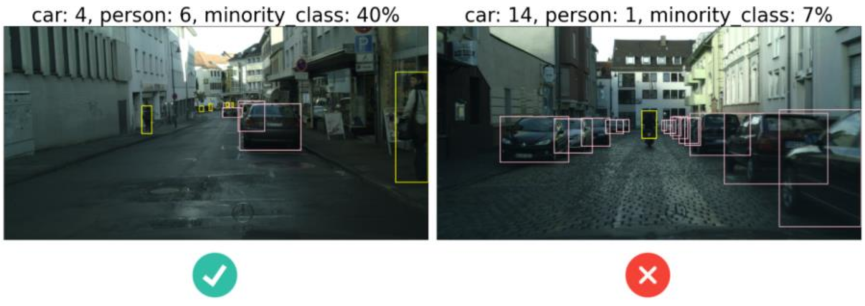


Figure 4. Example of kept and dropped image

ISSN (p) : 2407-1323
ISSN (e) : 2442-4404
DOI : https://doi.org/10.25124/jett.v9i1.4414
| Vol. 9 | No. 1 | Halaman : 1138 - 1144
JETT
Jurnal Elektro Telekomunikasi Terapan Juli 2022

From the flowchart, we can see that only empty images (without objects) and images with >20% minority class proportion will be kept. These constraints will selectively choose every image that has mild imbalanced data. Table 2 shows us the degrees of the imbalanced dataset based on its minority class proportion [5].

Table 2. Degree of the imbalanced dataset

Degree of imbalanced	The proportion of Minority Class
Mild	20-40% of the dataset
Moderate	1-20% of the dataset
Extreme	<1% of the dataset

We chose 20% as the limit because it will produce a balanced final dataset without eliminating too many object samples. Since downsampling means removing some of the data, we certainly don't want to have a too-small dataset, this would lead to overfitting or underfitting [3].

2.4 Mean Average Precision (mAP)

Mean Average Precision (mAP) is one of the most popular performance metrics for evaluating object detection models [9]. The mAP itself is the mean of Average Precision (AP) from all images. AP is the average of Precision (P) from each object inside a single image. In this research, we use the PASCAL VOC mAP standard which is mAP@0.5 or mAP with Intersection over Union (IoU) threshold of 50% [9], [10]. PASCAL VOC mAP can be calculated with this formula

$$mAP@0.5 = \frac{1}{N}\sum_{i=1}^{N} AP_i \,;\; IoU \geq 0.5 \tag{1}$$

where,

$$AP = \frac{1}{M}\sum_{i=1}^{M} P_i \tag{2}$$

$$P = \frac{TP}{TP + FP} \tag{3}$$

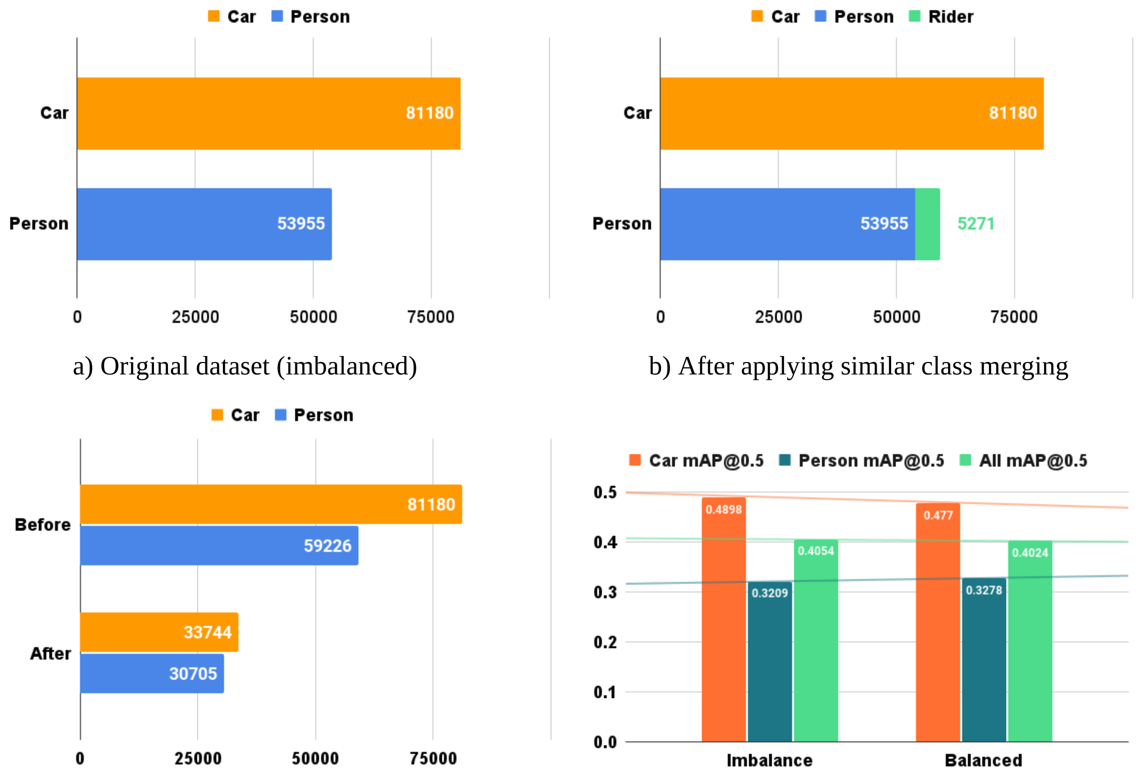
$$IoU = \frac{Intersection\ between\ predicted\ and\ groud\ truth\ box}{Union\ between\ predicted\ and\ groud\ truth\ box} \tag{4}$$

In Equation 1, N is the number of images, and in Equation 2, M is the number of objects in a single image. Variable i is the iteration value for each element of AP or P. In Equation 3, precision can be calculated once we know the number of True Positive (TP) and False Positive (FP) predictions. Finally, comparing the balance between car and person mAP will be the objective of this research because it represents whether to model is biased to a certain class or not.

3. RESULT ANALYSIS

In this section, we will analyze the model improvement using the proposed method. But first, we need to understand how the dataset looks like before we apply any approaches. As shown in Figure 5(a), car and person classes have a big gap which is very clear that the dataset is imbalanced.

ISSN (p) : 2407-1323
ISSN (e) : 2442-4404
DOI : https://doi.org/10.25124/jett.v9i1.4414
| Vol. 9 | No. 1 | Halaman : 1138 - 1144



a) Original dataset (imbalanced)

b) After applying similar class merging

c) Instance-level downsampling comparison

d) mAP comparison between trained models

Figure 5. Analysis visualization

3.1 Similar Classes Merging

In Figure 5(b), we can see the similar class merging result, the person class now have more objects and decreased the imbalanced gap with the car class. Although, because the rider only has 5,271 objects in the dataset, the result was still imbalanced if we compare it to car objects.

3.2 Instance-Level Downsampling

To balance the dataset even further, we applied instance-level downsampling. Figure 5(c) shows the comparison between before and after instance-level downsampling. After applying instance-level downsampling, we can see that the dataset is more balanced. But, because this method is downsampling, now we have a smaller dataset after removing certain images. A smaller and balanced dataset is preferred rather than having a large dataset with an extreme imbalance which can produce a biased model.

3.3 Performance Comparison Using EfficientDet D0

Evaluating the trained model will prove this method. So, we employ EfficientDet D0 as an object detection model and use Tensorflow Object Detection API as the framework [11], [12]. We trained the model on an imbalanced dan balanced dataset. Both model trainings used Focal Loss to prove that the weighted loss function is not enough to get a well-generalized model. After training with 15,000 steps, batch size of 8, learning rate 1e-5 until 1e-4, and input size 640×640, we can see in Figure 5(d), the comparison of PASCAL VOC mAP between the 2 datasets. The overall mAP for the balanced dataset is lower, but if we see the gap difference between car and person mAP, it's clear that the balanced dataset win. The balanced dataset becomes less biased to car class. We can see car mAP decreased from 0.4898 to 0.477 and person mAP increased from 0.3209 to 0.3278. This result indicates that the balanced dataset pushed the model understanding to become more balanced across 2 classes. The three trend lines moving closer to each other also prove that the model tries to balance the mAP.

Another way to analyze the improvement is by looking at the inference result. As we can see in Figure 6, the model with an imbalanced and balanced dataset detected the same number of objects inside an image. But car and person objects in the balanced dataset model have a small difference of confidence score percentage compared to imbalanced one. Depending on the threshold, an imbalanced confidence score could lead to a false decision if the system doesn't treat car and person equally.

a) Trained with imbalanced dataset   b) Trained with balanced dataset

Figure 6. Model inference results.

4. CONCLUSION

Based on the result analysis, it's proven that our method was able to improve the model performance. As measured by the mAP metric, the 1st model which was trained with an imbalanced dataset showed a biased knowledge with a higher tendency to detect cars confidently, but less sure about detecting a person. On the other hand, the 2nd model which was trained with a balanced dataset showed a more balance model knowledge across the 2 classes. So, it's clear that the weighted loss function itself is sometimes not enough if the imbalanced case is severe. For future research, based on the underlying concept of downsampling, we must state that our method works best with a big amount of original dataset (thousands). A small amount of imbalanced dataset could produce an even smaller balanced dataset which will not be good for the model.

ACKNOWLEDGMENT

This research was supported by Ministry of Education, Culture, Research and Technology under grant of Matching Fund Kedaireka No. 3541/E3/SK.09/KL/2021

DAFTAR PUSTAKA

[1] T., Lin, P., Goyal, R., Girshick, K., He and P. Dollar. 2017. Focal Loss for Dense Object Detection. Venice, Italy. IEEE International Conference on Computer Vision (ICCV).

[2] Utomo, Syam Suryo, Cahyanto, Triawan Adi, and Prakoso, Bakhtiar Hadi. 2020. Penggunaan Algoritma Random Over Sampling Untuk Mengatasi Masalah Imbalance Data Pada Klasifikasi Gizi Balita. Jember, Indonesia. Universitas Muhammadiyah Jember.

[3] Kumar, Benai. 2020. 10 Techniques to deal with Imbalanced Classes in Machine Learning. [Online] Available at https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/ [accessed on November 28th, 2021].

[4] Potyraj, Emily. 2021. 4 Ways to Improve Class Imbalance for Image Data. [Online] Available at https://towardsdatascience.com/4-ways-to-improve-class-imbalance-for-image-data-9adec8f390f1 [accessed on November 29th, 2021].

[5] Google Developers. 2017. Imbalanced Data. [Online] Available at https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data [accessed on December 5th, 2021].

[6] Brems, Matt. 2020. 5 Strategies for Handling Unbalanced Classes. [Online] Available at https://blog.roboflow.com/handling-unbalanced-classes/ [accessed on December 5th, 2021].

[7] Sakaridis, Christos, Dai, Dengxin, and Van Gool, Luc. 2018. Semantic Foggy Scene Understanding with Synthetic Data. International Journal of Computer Vision. Springer Science and Business Media LLC.

[8] Cordts, M., Omran, M., Ramos, S., Rehfeld, et al. (2016). The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[9] Hui, Jonathan. 2018. mAP (mean Average Precision) for Object Detection. [Online] Available at https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173 [accessed on December 22nd, 2021].

[10] Everingham, M., Van Gool, L., Williams, C., Winn, J., and Zisserman, A. 2012. The PASCAL Visual Object Classes Challenge 2012 (VOC2012). http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html

[11] Tan, M., Pang, R., and Le, Q.V. 2020. EfficientDet: Scalable and Efficient Object Detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[12] Yu, Hongkun, Chen, Chen, Du, Xianzhi, Li, Yeqing, et al. (2020). TensorFlow Model Garden. https://github.com/tensorflow/models.