Bachelor Thesis

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

Weakly Supervised Data Augmentation for LiDAR Based 3D Object Detection

Petr Šebek

Supervisor: prof. Ing. Tomáš Svoboda, Ph.D.

May 2021

BACHELOR‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Šebek Petr** Personal ID number: **483508**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Cybernetics and Robotics**

II. Bachelor’s thesis details

Bachelor’s thesis title in English:

Weakly Supervised Data Augmentation for LiDAR Based 3D Object Detection

Bachelor’s thesis title in Czech:

Slabě supervisovaná příprava LiDARových dat pro detekci 3D objektů

Guidelines:

Three-dimensional Light Detection and Ranging (3D LiDAR) sensors has recently got attraction in automotive [3] or in terrain robotics [4]. The sensor produce unorganized cloud of 3D points less regular than images which complicates using of (convolutional) deep networks. Object detection from 3D LiDAR data is less mature than from RGB data [3]. On the other hand full 3D information offers interesting possibilities in data augmentation which proved to improve recognition performance [1, 2]. Augmentation methods [1, 2] are uniformed, unsupervised. They statistically simulate small rotation and translation disturbances and noisify LiDAR reflections.
Within a related project we are developing a 3D point cloud tracking method that keeps object detected even when heavily occluded by other traffic actors. This thesis shall investigate possibilities of incorporating the results of 3D tracking as kind of weak supervision for data augmentation. We hope that modeling occlusions increases detection accuracy.
The thesis starts with evaluating the uniformed methods [1, 2] a baseline for comparison. Weakly supervised 3D LiDAR data augmentation may also have some potential in improving our LiDAR simulator [5].

Bibliography / sources:

[1] Martin Hahner, Dengxin Dai, Alexander Liniger, Luc Van Gool. Quantifying Data Augmentation for LiDAR based 3D Object Detection. arXiv:2004.01643 [cs.CV] (2020).
[2] Jaeseok Choi, Yeji Song, Nojun Kwak. Part-Aware Data Augmentation for 3D Object Detection in Point Cloud arXiv:2007.13373 [cs.CV] (2020).
[3] Selected chapters from: Joel Janai, Fatma Güney, Aseem Behl and Andreas Geiger (2020), Computer Vision for Autonomous Vehicles, Foundations and Trends in Computer Graphics and Vision: Vol. 12, No. 1–3, pp 1–308. DOI: 10.1561/0600000079.
[4] T. Rouček et al. Darpa subterranean challenge: Multi-robotic exploration of underground environments. In International Conference on Modelling and Simulation for Autonomous Systems, Springer 2019.
[5] P. Vacek, O. Jašek, K. Zimmermann, T. Svoboda. Learning to Predict Lidar Intensities. IEEE Transactions on Intelligent Transportation Systems, accepted, to appear in 2021.

Name and workplace of bachelor’s thesis supervisor:

prof. Ing. Tomáš Svoboda, Ph.D., Vision for Robotics and Autonomous Systems, FEE

Name and workplace of second bachelor’s thesis supervisor or consultant:

Date of bachelor’s thesis assignment: **07.01.2021** Deadline for bachelor thesis submission: **21.05.2021**

Assignment valid until: **30.09.2022**

prof. Ing. Tomáš Svoboda, Ph.D. Supervisor’s signature	prof. Ing. Tomáš Svoboda, Ph.D. Head of department’s signature	prof. Mgr. Petr Páta, Ph.D. Dean’s signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt

Student's signature

Acknowledgements

I would like to give thanks to my supervisor, Tomáš Svoboda, for all his advice and for guidance of this thesis. I would also like to acknowledge Partik Vacek for dataset preparation and his advice about neural networks training.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, May 21, 2021

Abstract

We propose two local 3D point cloud augmentations, Insertion and Movement simulation. Insertion method inserts objects bounding box to point cloud. This augmentation helps especially in case of an unbalanced dataset, as it increases number of exemplars for weakly represented classes. Movement simulation simulates positions of all moving objects in the future, based on their speed and direction from 3D tracking. For both of these augmentations, we design an algorithm that simulates realistic occlusion. For Movement simulation, we design an additional algorithm, which can fill parts of the scene that are uncovered by an objects movement.

Keywords: data augmentation, LiDAR data, deep learning

Supervisor: prof. Ing. Tomáš Svoboda, Ph.D.
ČVUT,
Resslova 307/9,
Praha

Abstrakt

Navrhli jsme dvě lokální metody pro rozšíření datasetu: Vložení a Simulace pohybu. Vložení přidává do mračna bodů nové objekty. Tato metoda pomáhá nejvíce na nevyvážených datasetech, kde dokáže zvýšit zastoupení malých tříd. Simulace pohybu simuluje pozice všech pohybujících se objektů na základě jejich rychlosti a směru z 3D sledování. Pro obě tyto metody jsme navrhli algoritmus, který simuluje realistickou viditelnost. Pro Simulaci pohybu jsme navíc navrhli algoritmus, který zaplní tu část mračna bodů, která se odkryje posunutím objektů.

Klíčová slova: rozšíření datasetu, LiDARová data, hluboké učení

Překlad názvu: Slabě supervisovaná příprava LiDARových dat pro detekci 3D objektů

Contents

Figures

Tables

Chapter 1

Introduction

Supervised learning needs a large number of training samples. Training samples need to be manually annotated, which is extremely time-consuming and therefore expensive. This usually results in an insufficient number of training samples. Data augmentations have the ability to increase the number of training samples with much less expense than capturing and annotating new samples. Augmentations create new samples, which could be expected in the real world, but are not represented in the original dataset by slightly changing the original ones. These new samples do not need to be manually annotated, because the annotations are taken from the original samples.

Augmentations for 3D point clouds are different than augmentations which can be applied to RGB data, which is well studied as mentioned in article [4]. Cartesian coordinates of each point are known which gives us the ability to make augmentations that would not be possible with RGB data. We can thus apply augmentations based on several geometrical transformations, locally or globally. Local augmentation is applied just to a certain part of a scene, global one is applied to the whole scene.

Based on articles [4, 2] we experimented with 5 global uninformed augmentations: Translation, Rotation, Mirror reflection, Random noise, and Random point removal. We propose two new local informed augmentations: *Insertion* and *Movement* simulation.

Training a neural network on 3D point clouds brings problem with usage of usual convolutions. Point clouds are unorganized without any a structure, which makes a direct application of usual convolutions impossible, because they are designed for extracting features from relations between neighboring pixels or voxels. Another problem is that the information stored in 3D point clouds is much sparser then in an RGB images. This is not preferable for the training of the neural network, because most of the space taken by point clouds are empty. We train models for semantic segmentation, which means that the deep neural network is trained to assign a class label to each part of the point cloud. We opted for the semantic segmentation task over the detection task. Training models for the detection task takes longer than for the semantic segmentation task. If augmenting data improves segmentation, there is a high probability that it will improve the detection results as well.

Chapter 2

Augmentation methods

2.1 Data representation

LiDAR sensors measure distance and intensity of reflection from surfaces in their surroundings. Measured data can be seen as a list of 3D coordinates with associated intensities. Whole point cloud we call scene, exemplar of scene can be seen in figure 2.1. Colors of points correspond with points class in table 2.1.

Class	Color
Background	Red
Vehicle	Yellow
Pedestrian	Purple
Bikes	Blue
Road blocks	Gray
Road	Green
Not arrived	Black

Table 2.1: Colors of each class

Point clouds are unorganized and for that type of data we cannot use usual convolutions, due to lack of structures in point cloud. There are more solutions for this problem, e.g. Kernel Point Convolution (KPconv) [7], which can adjust kernel shape during training, or voxelization, which makes 3D convolution usage possible. However, training 3D convolutional neural network on point cloud is very time-consuming. In our case, if we want the voxel size to be 10 cm with point cloud length, width, and height (200 m, 100 m, 10 m), then one point cloud contains 200 millions voxels, but on average it can be filled with points maximally 107,000 voxels, due to the average number of points in point cloud. That means the majority of the point cloud is free space, where the 3D convolution does not have any response. Therefore, we take another approach and make a 2D picture from the point clouds. Information in the 2D picture is much denser, which eliminates spaces where convolutions have no response and training is much faster.

We transform point clouds into a 360° horizontal field of view (FoV) inspired

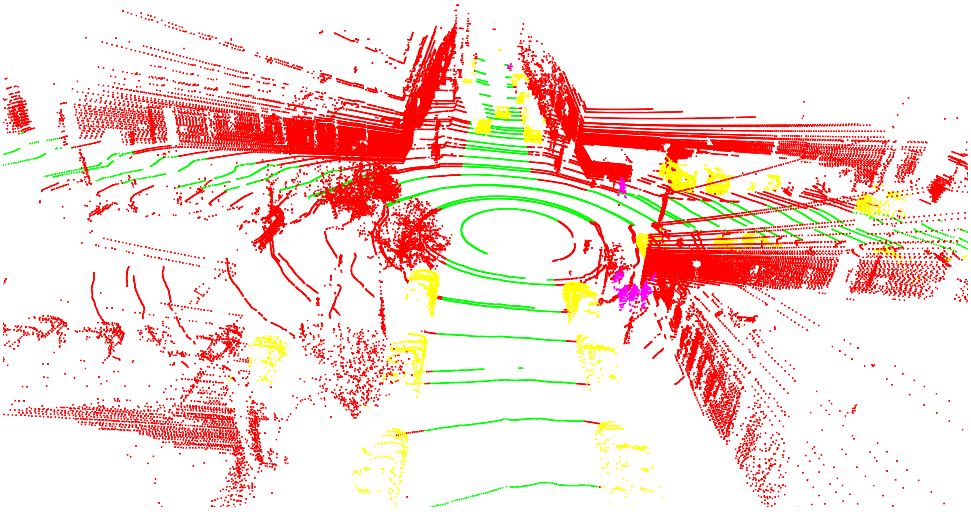


Figure 2.1: Visualization of one scene made from 3D point cloud. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks

by article [8], which is a 2D picture of the point cloud from the place where the LiDAR is. Therefore, we spherically project the 3D point clouds to the field of view. Our FoV representation has 3 channels. The first channel contains the distance between LiDAR and the measured point. The second channel contains the reflections intensity and in the third is a mask containing 0, if no point is projected to corresponding pixel, or 1 if the pixel represents some point. If more points are projected to the same pixel, the pixel stores information about the closest point. If the pixel does not represent any point the value in the first and second channel is set to 0. These pixels belong to a new class called Not arrived. Figure 2.2 is field of view made from point cloud in figure 2.1.

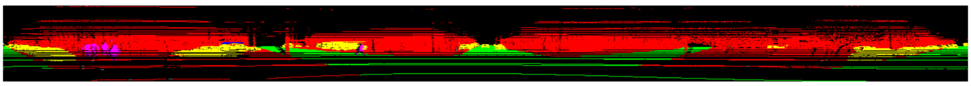


Figure 2.2: 360° horizontal field of view. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks

In order to make field of view we transform Cartesian coordinates to Spherical coordinates (equation 2.1), which we then project to the field of view (equation 2.2). In our case, the field of view has 112 rows and 1,440 columns, which means that it has a 0.25° resolution in azimuth.

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \varphi &= \arctan 2(y, x) + \pi \\ \theta &= \arccos\left(\frac{z}{r}\right) \end{aligned} \tag{2.1}$$

$$\begin{aligned}\text{row}_{\text{point}} &= \text{floor}\left(\frac{112 \cdot (\theta_{\text{point}} - \theta_{\text{min}})}{\theta_{\text{max}} - \theta_{\text{min}}}\right)\\ \text{column}_{\text{point}} &= \text{floor}\left(\frac{1,440 \cdot \varphi_{\text{point}}}{2 \cdot \pi}\right)\end{aligned} \tag{2.2}$$

where θ_{min} and θ_{max} is minimal and maximal elevation angle in scene. We use this approach because Argoverse dataset is measured in cities. In the city there are many high objects near by (buildings on sides of streets or trees on sidewalks), which means that the probability that the highest LiDAR beam has no reflection is very low.

For comparison we trained models on baseline dataset, which is generated with same pipeline as is described above, while only difference is that θ_{min} and θ_{max} is minimal and maximal elevation angle in the whole dataset, so the range of elevation angle is for sure in all FoVs the same.

2.2 Uninformed augmentation

Uninformed augmentations do not consider context of the scene [4, 2]. They are used, in our case, globally on the whole scene. From article [4] we try Translation, Rotation, and Mirror reflection. From article [2] we try Add Noise to Partition and Sparsify Partition, which is used locally in the article. However, we use it globally, which means that we call them Random noise and Random points removal.

2.2.1 Translation

Translation is applied on Cartesian coordinates, i.e. directly on the point cloud. This augmentation makes the field of view which shows surroundings from different points of view. With rising translation, the field of view deformation increases, on the other hand the difference between the new and original field of view increases. We use translation with a mean value of 0.2 m, uniformly from interval $(-0.2414; -0.0414) \bigcup (0.0414; 0.2414)$ m in coordinates X and Y. 0.2 meters we set based on article [4], where this value of translation gives the best results.

2.2.2 Rotation

Rotation is applied to Spherical coordinates. This augmentation adds context to edges from the original field of view so that the neural network can learn from these parts fully. Furthermore, the neural network may no longer expect that the road is in the middle and on the edges of field of view. Due to this augmentation, the road may be anywhere and the neural network cannot presume road location and it is forced to adapt. We add a random number from interval $(0; 2\pi)$ to the azimuth angle of all points in a scene. A random number is generated for each scene separately.

2.2.3 Mirror reflection

This augmentation is applied on field of view (FoV) representation. It mirrors the picture around a random column. This augmentation does not add any deformation, therefore it provides a plausible field of view, which is different from the original one. Due to this the neural network is trained with more silhouettes of different objects. That should help especially for detection of pedestrians. Pedestrians have many silhouettes due to different positions of hands and legs.

2.2.4 Random noise

This augmentation is applied on field of view. It adds noise to the 1. and 2. channel. Noise in distance should simulate objects with different details than objects which are in the original dataset. Noise in intensity should simulate different materials and colors. We use uniform noise. The values of noise in distance are from interval (-3;3) cm and in intensity (-3;3)% from maximum.

2.2.5 Random points removal

This augmentation is applied on field of view. Due to this augmentation the neural network should learn more general information rather than details. Furthermore it should simulate weather conditions such as rain or snow. We remove 5% of points from original field of view.

2.3 Informed augmentation

Informed augmentations consider context of the scene. In our case Insertion add objects on places, where objects are mostly located and Movement simulation using information about speed and direction of moving objects in the scene.

2.3.1 Insertion

Insertion is an augmentation, which adds bounding boxes of certain class to point clouds, so the neural network can learn from more objects. Consequently, Insertion can increase percentage of a certain class, which is also beneficial for training. We conduct experiments which insert vehicles, pedestrians, and bikes.

For realistic Insertion, we need to address the following problems:

- **Insertion object:** We need to have point clouds, which belong only to one object. For that we exploit bounding box annotations in datasets. Vehicle point clouds are cut out from Argoverse, pedestrian and bikes point clouds are used from KITTI. Details will be discussed in subsubsection 2.3.1.1.

- **Insertion place:** Objects should be located appropriately, for example vehicles and bikes should be on the road and pedestrians on the sidewalk. Therefore we have to get full information on where these areas are in the scenes. For that we make a Road map, which we will further discuss in subsubsection 2.3.1.2.

- **Collision prevention:** Objects should be inserted without any collision with objects which are already in scene. Our method is described in subsubsection 2.3.1.3.

- **Realistic visibility:** After insertion we need to simulate occlusions - which part of the inserted object will be occluded and which part of the scene will the object occlude. Subsubsection 2.3.1.4 is focused on solving this problem.

2.3.1.1 Cutting the bounding box out

From a randomly chosen scene, we cut out a objects bounding box based on its center (in Cartesian coordinates), orientation (in quaternions), and its length, width, and height. This means that we must make a decision for, whether the points are in 3 specific intervals, meaning that it is the intersection of 6 half-spaces. Then we remove all points which are in the bounding box, but are not annotated as the same class as the object has, mostly it is points annotated as the Road. We use bounding boxes which consist of at least 51 points. In figure 2.3 can be seen relation between number of reflections from vehicle and their distance.

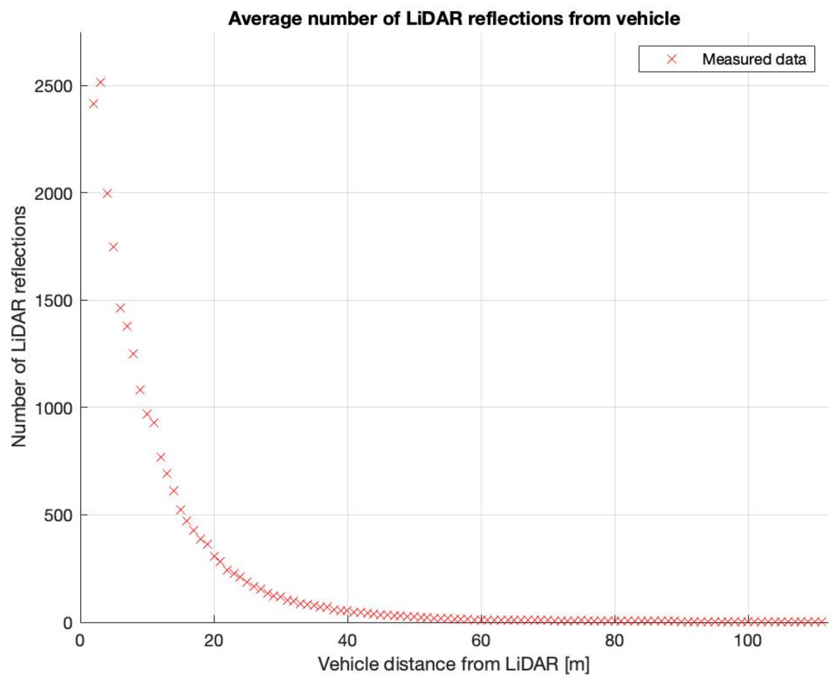


Figure 2.3: Average number of LiDAR reflection from vehicle based on their distance from LiDAR

2.3.1.2 Road map

Scenes are recorded in sequences (around 156 scenes in a sequence) and each scene has its own transformation matrix into the global coordinate system. Therefore, we can combine Road annotations across scenes and make a dense Road map. We know where the roads are even when it is not captured in the scene point cloud due to this. In maps, each pixel represents a square 1 m by 1 m in real world units. On average, Road maps made from one sequence are of size 448×426 m. Example of Road map made from one sequence can be seen in figure 2.4. Then we combine the Road maps across sequences to reduce the number of places, where the Road annotations are sparser (usually beginnings and ends of the sequences). The large Road map made by merging sequence Road maps is 2,856 m by 2,317 m large, visualization of this Road map can be seen in figure 2.5.

Pedestrians are usually located near the road (on sidewalks), therefore we suggest to make a “Pedestrian area”, which is around the road (maximally 2 m far from the road). With this method the “Pedestrian area” is even in places were the road should continue but it is not captured in dataset, however it will simulate crosswalk. Visualization of the large Road map can be seen in figure 2.6.

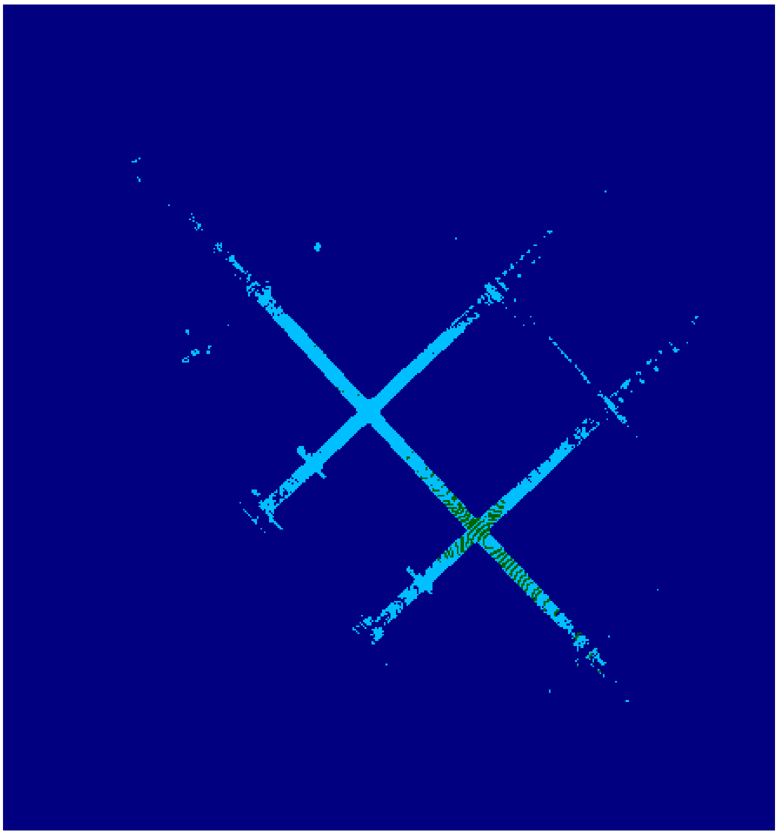


Figure 2.4: Visualization of Road map made from one sequence. light blue - Road annotation from sequence, green - Road annotation from one scene, dark blue - not drivable area

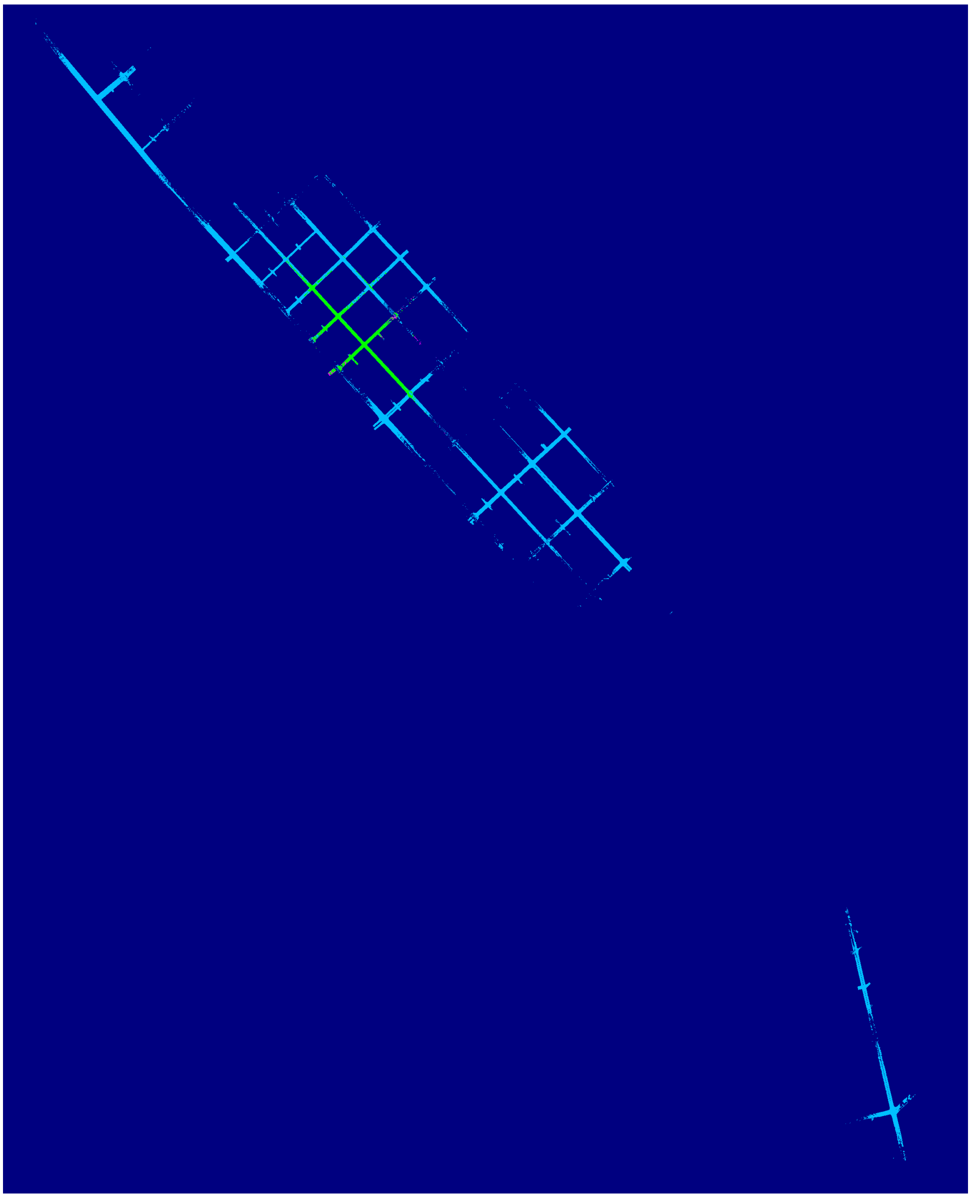


Figure 2.5: Visualization of Road map made from all sequences. light blue - Road annotation from all sequences, green - Road annotation from one sequence, dark blue - not drivable area

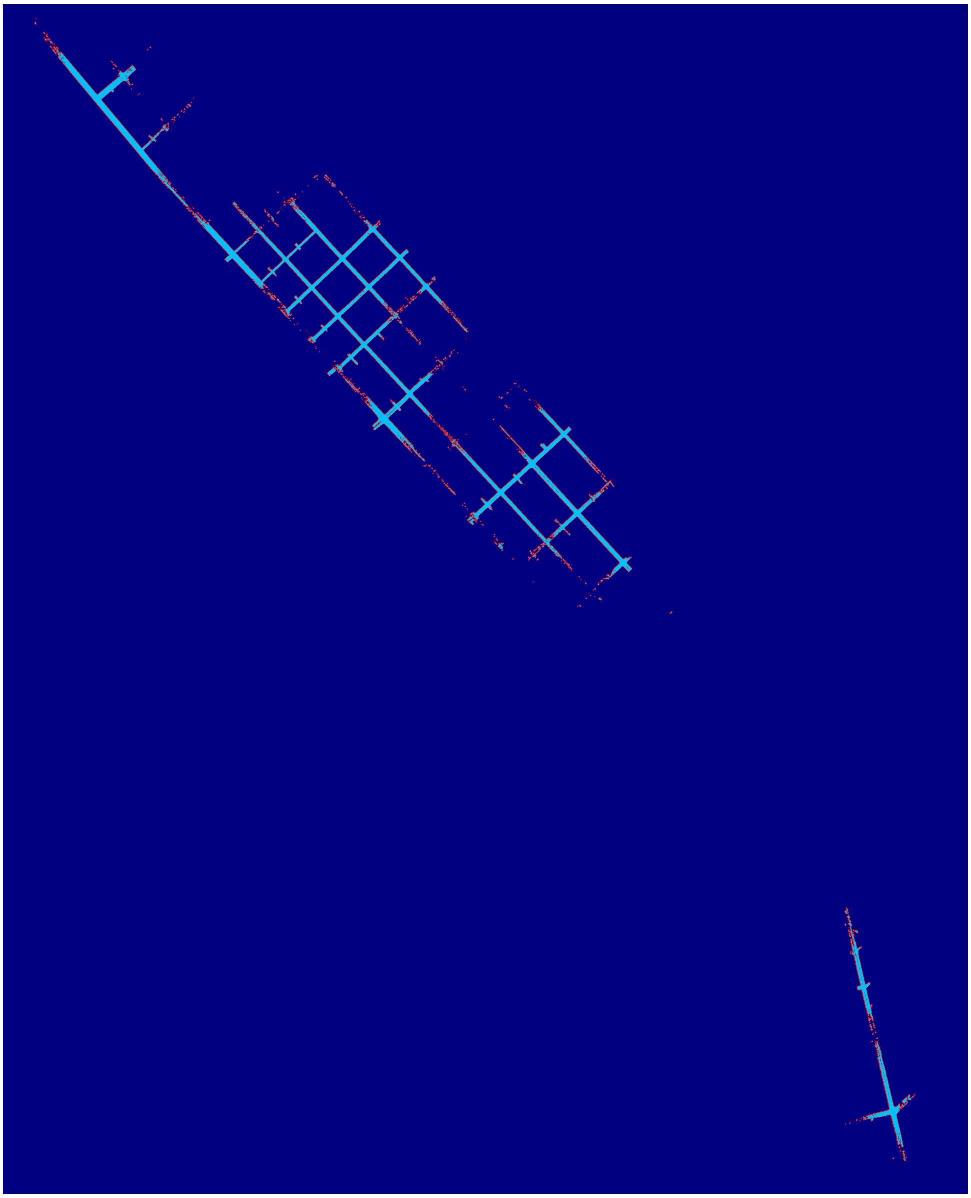


Figure 2.6: Visualization of Road map with "Pedestrian area". light blue - Road annotation from all sequence, red - "Pedestrian area", dark blue - not drivable area

2.3.1.3 Possible placement for bounding box insertion

For this part, we need the scene point cloud, Road map (from 2.3.1.2) and bounding box, which we want to add (from 2.3.1.1).

1. We project all points from the point cloud to the plane X/Y. If some point, which is not annotated as the Road, is projected to a place where the road should be in the Road map, we remove this part of the road from the Road map, because it is not available.

2. We rotate the bounding box and all its points by 1° around the scene Z coordinate axis. We rotate around the scene Z coordinate axis because we want to maintain the distance between the bounding box and the center of the scene and perspective of the bounding box. We want to sustain these two requirements in order to make the outcome realistic. If we repeat this step 360×, we will move to another cut bounding box.

3. We project all points from the bounding box to plane X/Y. If all points are projected to the available road in the Road map, we continue to step 4, otherwise we go back to step 2.

4. We are searching for points that are annotated as the Road and are closest to the center of the bounding box. We are doing this in projection to plane X/Y. We gradually increase the size of the circle by 0.1 meters until at least one point annotated as the Road is inside this circle. From these points we compute the average in Z coordinates, it is the level of the road. We adjust Z coordination of the bounding box according to the road level.

$$\text{bounding box center}_z = \text{road level} + \frac{\text{bounding box height}}{2} \tag{2.3}$$

   Due to this correction, the vehicle will be touching the road surface as it would be in the real world.

5. We cut a cuboid with the same dimensions and rotation as the bounding box (from step 2), from scene point. If there are only points that are annotated as Road in the cuboid, we continue to step 6, otherwise we go back to step 2.

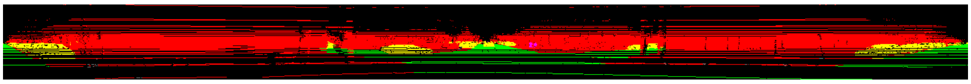
6. We compute if there are any points in the intersection between the bounding box from step 2 and every single bounding box annotated in the scene. If all intersections are empty, placement is suitable for insertion, otherwise we go back to step 2.

From our point of view if steps 5 and 6 are fulfilled it provides us enough insurance that we can place the bounding box in the scene. Even when we only check if there are any points in the intersection and we do not check if the intersections have any dimensions. Bounding box is a rough approximation of an object, so it can have a slight intersection with another bounding box, but the object would still not interfere with objects in scene.

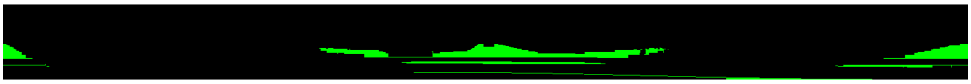
2.3.1.4 Visibility

For solving visibility we are using FoV representations. However there is one problem, that must be solved. LiDAR data is sparse, which results in black spots in, for example, vehicles and background in FoV representation, as can be seen in the original field of view in the figure 2.7. We need to reduce the black spots in order to secure proper results of our method. To solve this problem, we applied a closing [1] to the field of view. Closing is a combination of dilatation and erosion [5], while dilatation is applied first. For dilatation and erosion we use the same core. We use a rectangular core with the size of 5 rows and 3 columns. We chose this vertical shape because most of the obstacles on the road are vertical. Dilatation places core centre to each pixel and sets its new value to maximum value at the core. Erosion makes the opposite operation - it sets new values of pixels as the minimum value at the core. Due to this effect we need to use closing in masks of each class, because indexes of classes do not correspond to their relations. When we merge closing results we set classes priority decreasingly: Pedestrian, Road blocks, Bikes, Vehicle, Background, Road. Classes which have a higher probability that their closed mask can cover other objects have lower priority. For example in the figure 2.7 Pedestrian will be covered by Background.

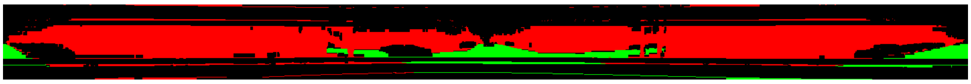
[1] `https://scikit-image.org/docs/dev/api/skimage.morphology.html#skimage.morphology.closing`



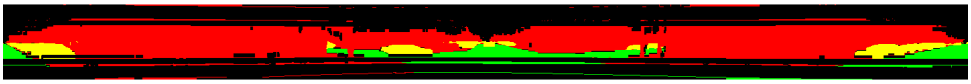
(a) : Original 360° field of view



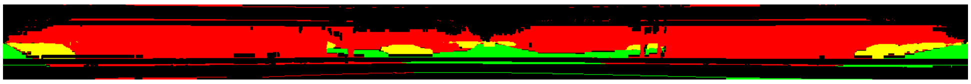
(b) : Road mask



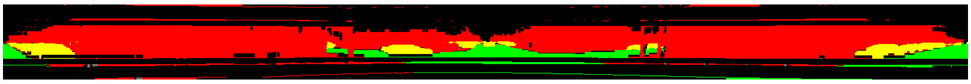
(c) : Background mask combined with previous mask (b)



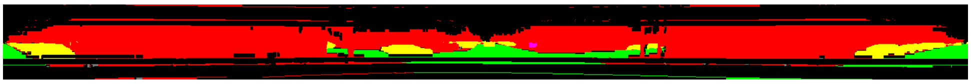
(d) : Vehicle mask combined with previous mask (c)



(e) : Bikes mask combined with previous mask (d)



(f) : Road blocks mask combined with previous mask (e)



(g) : Pedestrian mask combined with previous mask (f) - Final closed 360° field of view

Figure 2.7: Visualization of making closed 360° field of view. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks

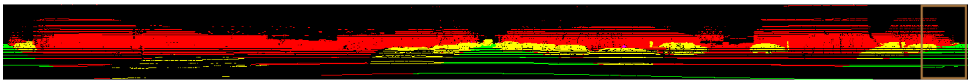
The distance in pixels, which are added by closing are computed as average distance of neighbour pixels with corresponding class, which are not further then 1 column and 2 rows from added pixel (shape of dilatation and erosion core). We use the same process on the bounding box point cloud. Both these closed field of views are used for simulating realistic occlusions. The process goes as follows:

1. If a certain pixel in bounding box closed 360° field of view (bbox-FoV) represents at least one point, we continue to step 2, otherwise we move to a other pixel and stay on step 1.

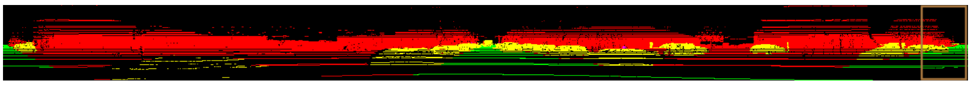
2. If in scene closed 360° field of view (scene-FoV) there is in the corresponding pixel a higher value in distance than in bbox-FoV or this pixel does not represent any point in scene, we remove all points in the scene that are projected to the corresponding pixel, as they would be covered by the added object. Then we add points from the bounding box which are projected to the corresponding pixel to scene point cloud. If in scene-FoV the distance is smaller than in bbox-FoV, we do not remove or add any point, because some object in the scene covers this part of the object in the bounding box.

3. If more than 20 points are visible from the whole object we use this augmented scene, otherwise we try a different location of the bounding box (2.3.1.3).

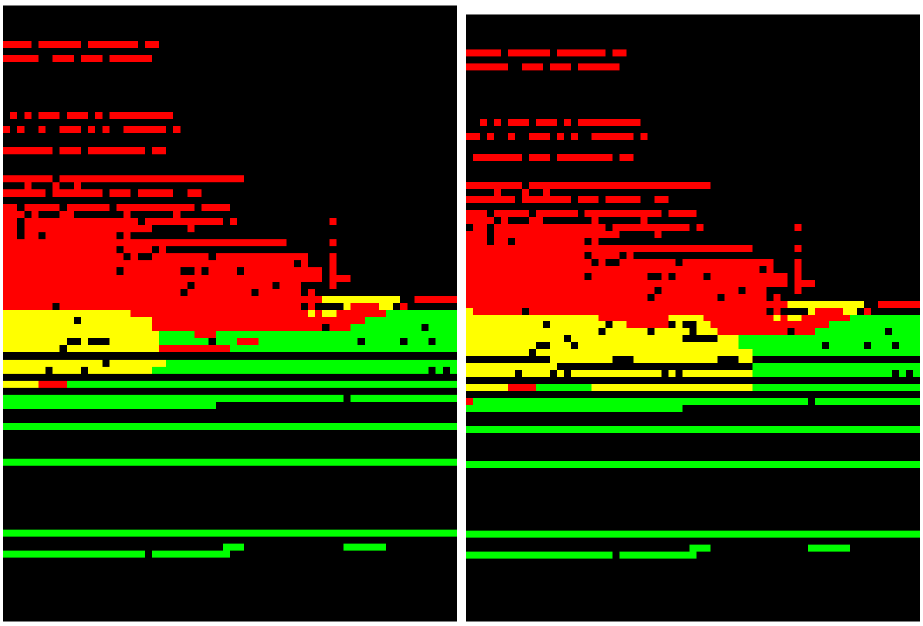
In figures 2.8, 2.9, and 2.10 a comparison between original and augmented field of view is illustrated.



(a) : Original 360° field of view



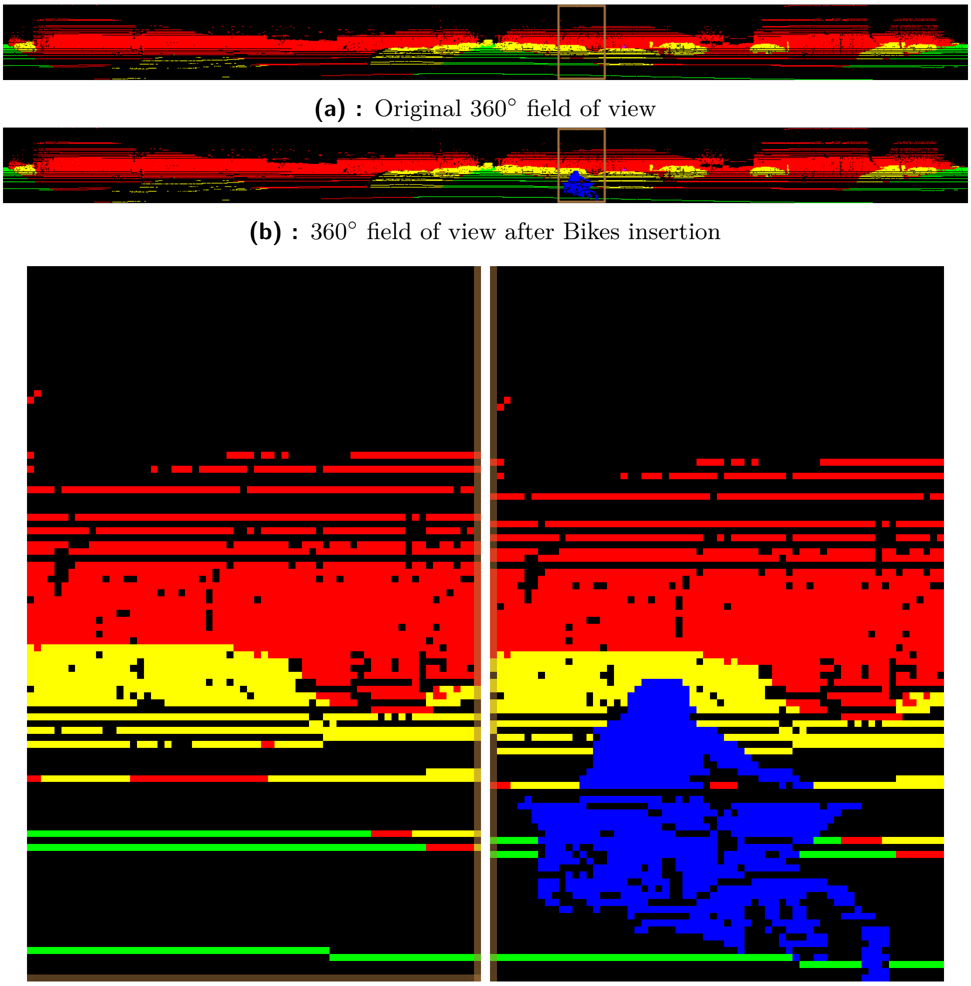
(b) : 360° field of view after Vehicle insertion.



(c) : Detail of original 360° field of view (a)

(d) : Detail of 360° field of view after Vehicle (yellow) insertion (b)

Figure 2.8: Visualization of Vehicle insertion. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks



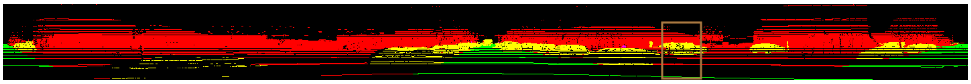
(a) : Original 360° field of view

(b) : 360° field of view after Bikes insertion

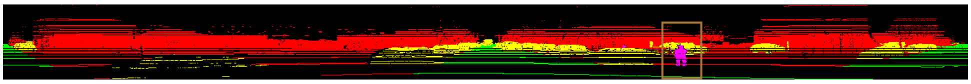
(c) : Detail of original 360° field of view (a)

(d) : Detail of 360° field of view after Bikes (blue) insertion (b)

Figure 2.9: Visualization of Bikes insertion. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks



(a) : Original 360° field of view



(b) : 360° field of view after Pedestrian insertion

(c) : Detail of original 360° field of view (a)

(d) : Detail of 360° field of view after Pedestrian (purple) insertion (b)

Figure 2.10: Visualization of Pedestrian insertion. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks

2.3.2 Movement simulation

This augmentation is based on 3D tracking. It moves all moving objects in the direction of their movement, except for the measuring vehicle (ego). In this augmentation, we assume that every object can continue moving at its current velocity for at least 0.3 seconds, without colliding with other objects.

For time optimization we are using the method below just on moving objects. Therefore, we need to know which objects are stationary and which are moving. This decision is based on the object position in the global coordination system. If an object moves more then 0.5 m in X or Y coordinates in the whole sequence, we assume the object as moving. Our method consists of the following steps:

1. Remove all moving objects from the scene.

2. Fill parts of scene, which is uncovered by the removal of moving objects.

3. Add moving objects moved by their velocity vector back to point cloud with realistic visibility.

Therefore, we need to address the following problems:

- **Object velocity vector:** We need to know how fast objects are moving and in which direction. For that, we use Argoverse tracking data, which is included in the dataset. We will discuss more in subsubsection 2.3.2.1.
- **Filling uncovered part of scene:** Remove objects uncover a part of the scene, which needs to be filled by some scene background. Therefore, we combine the point clouds of all static parts of scenes in sequence (stationary scene). Subsubsection 2.3.2.3 is focused on solving this problem.
- **Realistic visibility:** After we move objects we need to add them, with realistic occlusions, so we use the same method as in subsubsection 2.3.1.4.

2.3.2.1 Object velocity vector

In every scene, we compute the velocity vector for each moving object. Velocity vector is calculated as the difference in position between the previous and present scene (preferred) or the present and following scene, if the object is not annotated in the previous scene. This vector represents the object position change per 0.1 seconds. By this vector, we move all object points in the case of simulation by 0.1 s or we multiply the vector by 3 if we simulate by 0.3 s.

2.3.2.2 Composed stationary scene

For making composed stationary scene we remove points, which belong to moving objects, from every individual scene in sequence. These individual scenes are glued together by transforming them to the global coordinate system. An example of composed stationary scene can be seen in figure 2.11.

Figure 2.11: Visualization of composed stationary scene. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks

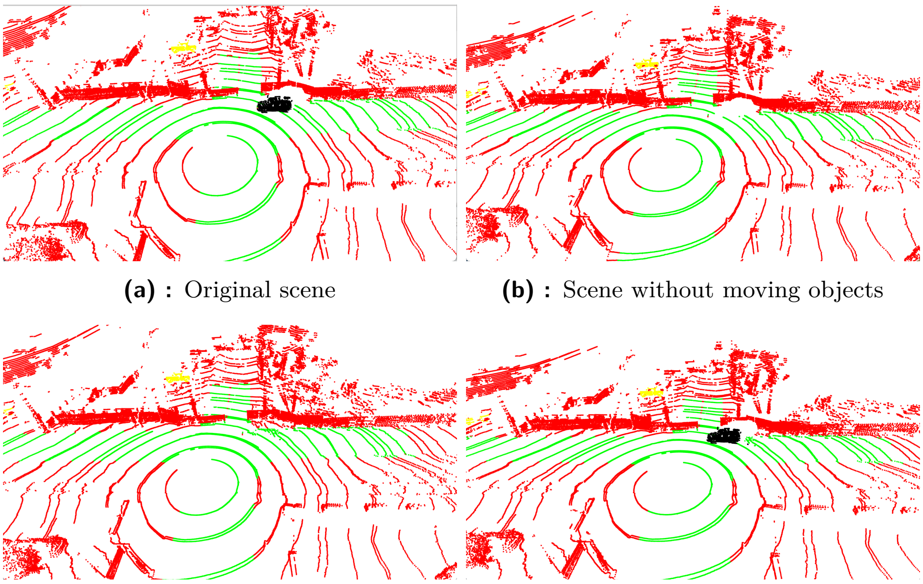
2.3.2.3 Filling uncovered part of scene

1. We transform stationary scene (created in 2.3.2.2) to scene coordination system and crop them to the same dimension as the scene and we make field of view from this point cloud.

2. From points, which belong to moving objects in the original scene, we make a 360° field of view.

3. If pixel in the field of view (from step 2) represent points we try to add the same amount of points, which are projected to a corresponding pixel from stationary scene.

In figure 2.12 Movement simulation pipeline is to be seen. Moving object (in this case vehicle) is colored black. In figure 2.12a illustrated original scene. Figure 2.12b is original scene from, which we removed moving objects. In figure 2.12c can be seen a scene, which is filled from the stationary scene, and in last figure 2.12d is the final scene, where the moving object is added with realistic visibility.



(a) : Original scene **(b) :** Scene without moving objects

(c) : Scene with filled uncovered scene **(d) :** Scene with moved object

Figure 2.12: Movement simulation pipeline. red - Background, yellow - stationary Vehicle, purple - stationary Pedestrian, blue - stationary Bikes, gray - Road blocks, black - moving object

Chapter 3

Implementation details

3.1 Datasets

As the main dataset we use Argoverse [1], from which we use whole point clouds (scenes). However, for some of our augmentation (Insertion), we want to have some additional object point clouds. For this reason, we also use KITTI [3] dataset. From this we use just the parts, which belong just to specific objects (pedestrian and cyclist).

3.1.1 Argoverse dataset

Argoverse [1] dataset was made by Argo AI[1]. The dataset was captured in two cities (Miami and Pittsburgh, USA) by two 32 beam LiDARs with a maximal range up to 200 m. These two LiDARs rotate at 10 Hz. The point cloud has a 50° range in vertical angle, because each LiDAR has 40° vertical field of view, however, they have a 30° overlap between each other.

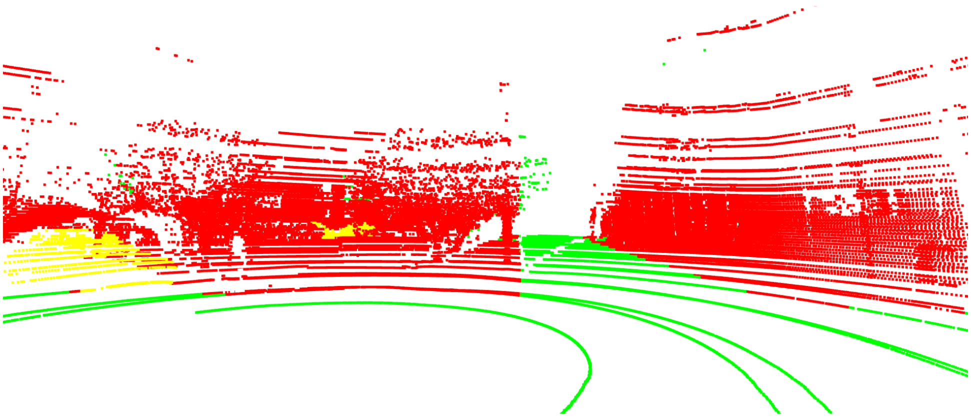
From the Argoverse dataset we took data to make two datasets of our own (large and small version). The small version contains 2,142 training scenes (12 sequences from Argoverse dataset), 506 validation scenes, and 506 test scenes, which we use for evaluating uninformed methods from articles [4, 2]. The large version contains 5,701 training scenes (33 sequences from Argoverse dataset), 1,306 validation scenes, and 1,238 test scenes. Each scene has about 107,000 points and all scenes, which we use include the Road annotations.

We changed the Argoverse dataset for our experiments, because of the following reasons:

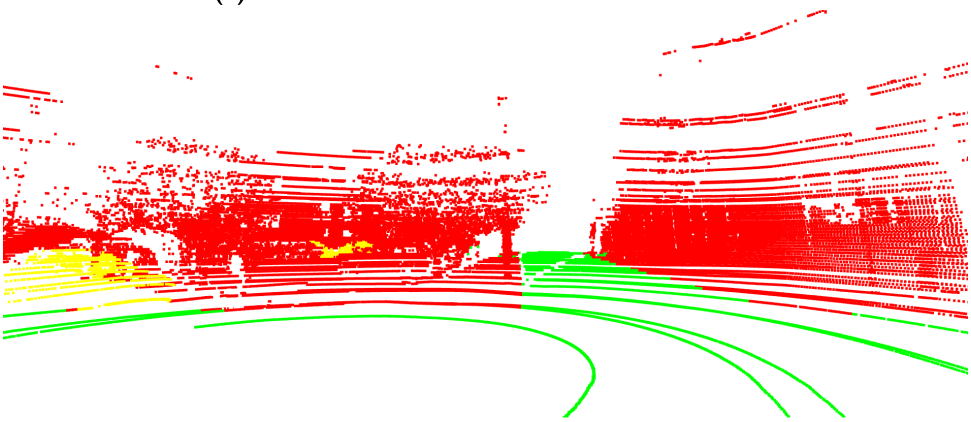
- **Dataset task:** We want to train our models for semantic segmentation, however, originally the dataset was designed for object detection. Therefore, we assign a label to each point in bounding box based on the bounding box annotations contained in the dataset.
- **Number of classes:** The original dataset contains 15 classes. We reduce them to 6 (Background, Vehicle, Pedestrian, Bikes, Road blocks, and Road) by merging the original classes.

[1] www.argoverse.org

- **Road annotation:** In original dataset was as Road annotated even points, which is above road surface for example, parts of tree crowns. Therefore we correct the Road annotations by re-annotating all points that are more than 30 cm above the X/Y plane from the Road to the Background. In figure 3.1 can be seen comparison between scene before and after correction of Road annotations.



(a) : Scene before correction of Road annotations.



(b) : Scene after correction of Road annotations.

Figure 3.1: Visualization of correction of Road annotations. red - Background, yellow - Vehicle, purple - Pedestrian, blue - Bikes, gray - Road blocks

- **Point cloud measurements:** We remove all points, which are further from LiDAR than 100 meters in the X coordinate, 50 meters in the Y coordinate and more than 10 meters above LiDAR from point clouds. We remove these points because objects which are more than 100 m from LiDAR reflect a small amount of LiDAR beams, therefore these objects could make a hard case for the neural network, which could even worsen training results.

The percentage of each class in the dataset is in table 3.1.

Class	Large dataset	Small dataset
Background	82.22%	83.62%
Vehicle	10.92%	9.53%
Pedestrian	0.31%	0.3%
Bikes	0.01%	0.42%
Road blocks	0.06%	0.05%
Road	6.48%	6.08%

Table 3.1: Percentages of each class in point clouds

In the small version of the dataset we find odd Bikes annotations, as can be seen in figure 3.2. We think that these odd annotations are wrong and we fix them in the large version of the dataset. This is the reason why the percentage of Bikes in point cloud (table 3.1) and in field of view (table 3.2) is much higher in the small dataset then in the large dataset.

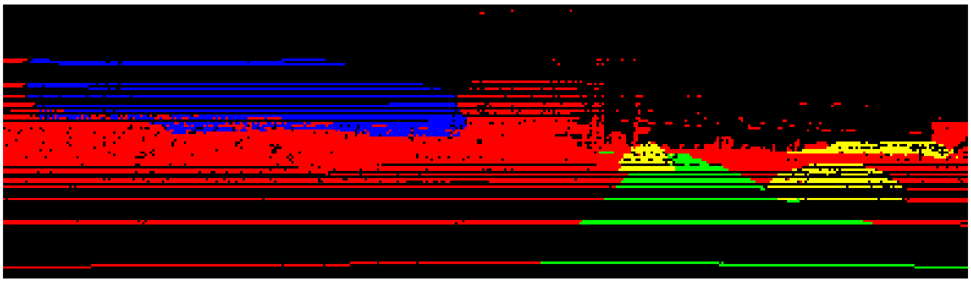


Figure 3.2: Odd Bikes annotation. Background - red, Vehicle - yellow, Pedestrian - purple, Bikes - blue, Road blocks - gray

These odd Bikes annotation makes major part of Bikes annotation in the small dataset as can be seen in table 3.1 percentage after fixing this annotation drop from 0.42% to 0.01%, that means 97.6% of Bikes annotation was corrupted.

3.1.2 KITTI

KITTI [3] dataset was captured in a city (Karlsruhe, Germany). One 64 beam LiDAR with maximal range up to 120 m and rotating at 10 Hz was used for scanning surrounding objects. An average scene contains 100,000 points. LiDAR has 26.8° vertical field of view.

KITTI has a 2× smaller vertical angle in the field of view than Argoverse has with the same amount of LiDAR beams, that means LiDAR rows are closer to each other and therefore an object in KITTI dataset is usually measured by more LiDAR beams. From KITTI dataset we use 2,667 pedestrians and 1,010 bikes.

3.2 Neural network

To implement our neural network, we use Python3[2] especially PyTorch[3]. We use a convolution neural network (CNN) based on U-Net architecture [6], used for semantic segmentation, for training. In our architecture we use convolution with padding in order to sustain picture size. Also, we add one 2D convolution, with Batchnorm and Relu right in the middle of the network. For all training, we maintain the same neural network to compare results. The scheme of our neural network is on figure 3.3.

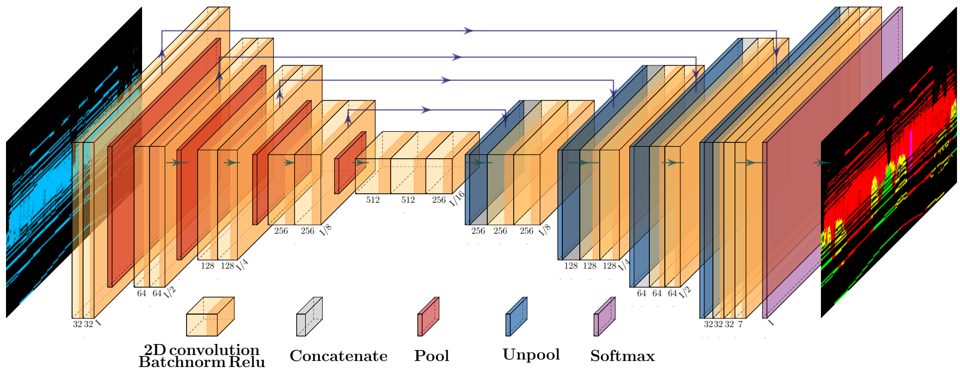


Figure 3.3: Neural network scheme

At the beginning of each training, we randomly initialize parameters of the neural network by Xavier uniform distribution[4], which compute uniform bounds as $[-a, a]$ according to 3.1

$$a = \sqrt{3} \cdot \sqrt{\frac{2}{\mathrm{n_{in}} + \mathrm{n_{out}}}}, \tag{3.1}$$

where $\mathrm{n_{in}}$ is number of input values and $\mathrm{n_{out}}$ is number of output value. This initialization should eliminate the vanishing or exploding gradient problem.

Due to random initialization, each training of the neural network can have a different result. We want to know what is the average result on the original dataset and on the augmented dataset, so we train the neural network five times (with the same dataset) and average their results. The learning rate is set to 0.001, the batch size is 16, Adam optimizer was used[5] and as a criterial function, we use weighted cross entropy loss [6].

Weights in the cross entropy depend on the class percentage in the training dataset, which can be seen in table 3.2. Two of the informed augmentations

[2] `https://www.python.org/download/releases/3.0/`

[3] `https://pytorch.org`

[4] `https://pytorch.org/docs/stable/_modules/torch/nn/init.html#xavier_uniform_`

[5] `https://pytorch.org/docs/stable/_modules/torch/optim/adam.html#Adam`

[6] `https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html#crossentropyloss`

change percentage of the same classes in the field of view significantly (Pedestrian and Bikes insertion) and consequently the weights in cross entropy loss. The weights which we use in cross entropy loss can be found in the table 3.3.

Class	Large dataset	Small dataset	Pedestrian insertion	Bikes insertion
Background	21.694%	22.472%	21.676%	21.682%
Vehicle	3.436%	3.003%	3.434%	3.431%
Pedestrian	0.093%	0.086%	0.134%	0.092%
Bikes	0.005%	0.133%	0.005%	0.048%
Road blocks	0.020%	0.021%	0.020%	0.020%
Road	2.318%	2.225%	2.317%	2.309%
Not arrived	72.434%	72.060%	72.413%	72.417%

Table 3.2: Percentages of each class in field of view

Class	Weight
Background	$1.8 \cdot 10^{-4}$
Vehicle	$1.2 \cdot 10^{-3}$
Pedestrian	$4.3 \cdot 10^{-2}$
Bikes	$7.6 \cdot 10^{-1}$
Road blocks	$1.9 \cdot 10^{-1}$
Road	$1.7 \cdot 10^{-3}$
Not arrived	$5.5 \cdot 10^{-5}$

(a) : Weights for the large dataset

Class	Weight
Background	$6.6 \cdot 10^{-4}$
Vehicle	$4.9 \cdot 10^{-3}$
Pedestrian	$1.7 \cdot 10^{-1}$
Bikes	$1.1 \cdot 10^{-1}$
Road blocks	$7.0 \cdot 10^{-1}$
Road	$6.6 \cdot 10^{-3}$
Not arrived	$2.0 \cdot 10^{-4}$

(b) : Weights for the small dataset

Class	Weight
Background	$1.8 \cdot 10^{-4}$
Vehicle	$1.2 \cdot 10^{-3}$
Pedestrian	$3.0 \cdot 10^{-2}$
Bikes	$7.7 \cdot 10^{-1}$
Road blocks	$2.0 \cdot 10^{-1}$
Road	$1.7 \cdot 10^{-3}$
Not arrived	$5.5 \cdot 10^{-5}$

(c) : Weights for Pedestrian insertion

Class	Weight
Background	$5.6 \cdot 10^{-4}$
Vehicle	$3.6 \cdot 10^{-3}$
Pedestrian	$1.3 \cdot 10^{-1}$
Bikes	$2.6 \cdot 10^{-1}$
Road blocks	$6.0 \cdot 10^{-1}$
Road	$5.3 \cdot 10^{-3}$
Not arrived	$1.7 \cdot 10^{-4}$

(d) : Weights for Bikes insertion with 1:1 ratio

Table 3.3: Weights for the cross entropy loss.

Weights of each class are computed as the inverted value of class percentage in FoV representation. Divided by the sum of the inverted value of percentage in FoV representation over all classes (equation 3.2).

$$w_{\text{class}} = \frac{1}{P_{\text{class}} \sum\limits_{\text{all classes}} \frac{1}{P_i}}, \tag{3.2}$$

where w_{class} is weight of certain class, P_i is percentage of i^{th} class in FoV representation.

Our codes are available at `https://gitlab.fel.cvut.cz/students/sebek-petr`.

Chapter 4

Experiments

As mentioned in Chapter 3, we train the neural network five times. Then we average these five runs and we get the result, which is less dependent on the first initialization. Results of every run can be analyzed in Appendix B. Values in the confusion matrices are the number of pixels, which are annotated as class in the corresponding row and are predicted as class in the corresponding column, in these tables GT stands for Ground Truth and PL stands for Predicted Label. For evaluating the results, we compute Recall, Precision, and Intersection Over Union (IOU) using equations 4.1.

$$\begin{aligned} \text{Recall} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \\ \text{Precision} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \\ \text{IOU} &= \frac{\text{True Positives}}{\text{False Positives} + \text{True Positives} + \text{False Negatives}} \end{aligned} \tag{4.1}$$

All augmentations are applied on every scene in the small or the large dataset. That means that every augmentation adds 2,142 new scenes in the case of the small dataset and 5,701 new scenes in the case of a the large dataset. Our goal is to improve IOU as much as possible, however, improving a higher value of IOU is much harder then a lower value of IOU, therefore the improvement from 0.8 to 0.85 is more valuable than from 0.2 to 0.25, even when the value difference is the same i.e. 0.05.

4.1 Uninformed augmentation

For evaluating uninformed methods we train the models on the small dataset (2,142 training scenes). Each training consists of 150 epochs.

4.1.1 Baseline

Baseline is trained just on the small dataset. In table 4.1 is the average training result and in appendix B.1.1 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15259667	413486	60890	18276	57744	343188	2143
Vehicle	723943	3505808	10465	669	6433	14141	1219
Pedestrian	61232	21132	34992	65	894	682	33
Bikes	11041	4687	4733	42665	178	244	58
Road blocks	30020	6240	1141	8	8446	536	19
Road	352922	46152	2781	67	8461	1703598	120
Not arrived	4395	8654	886	57	231	1728	58830488

Class	Recall	Precision	IOU
Background	0.945	0.928	0.880
Vehicle	0.822	0.875	0.736
Pedestrian	0.294	0.321	0.179
Bikes	0.671	0.693	0.516
Road blocks	0.182	0.119	0.075
Road	0.806	0.826	0.688
Not arrived	1.000	1.000	1.000

Mean IOU: 0.582 with Not arrived
Mean IOU: 0.512 without Not arrived

Table 4.1: Average baseline result on the small dataset

From the results, we can see that the predictions on Road blocks and Pedestrians are weak. Road blocks are usually small objects, which reflects just a small amount of LiDAR beams, and pedestrians are usually heavily occluded. This confirms the fact that Road blocks with Pedestrian has the smallest two percentage in FoV in the small dataset 3.2 (0.021% for Road blocks and 0.086% for Pedestrians). These factors most probably cause the not ideal detection results.

4.1.2 Translation

Translation adds another 2,142 scenes to the dataset, so the models are trained on a dataset that contains 4,282 scenes, the ratio between original and new scenes is 1:1. In table 4.2 is the average training result and in appendix B.1.2 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15318871	402968	67790	14283	35933	314705	844
Vehicle	510261	3718562	14756	308	3078	15304	407
Pedestrian	57444	20995	39368	11	560	644	10
Bikes	12757	4814	4563	41168	160	130	16
Road blocks	28550	6981	1769	3	8624	480	3
Road	327499	41931	749	15	3468	1740299	140
Not arrived	1092	8362	1654	25	74	942	58834289

Class	Recall	Precision	IOU	IOU improvement
Background	0.948	0.942	0.896	0.016
Vehicle	0.872	0.885	0.783	0.047
Pedestrian	0.331	0.316	0.190	0.011
Bikes	0.647	0.745	0.526	0.010
Road blocks	0.186	0.171	0.096	0.021
Road	0.823	0.840	0.711	0.023
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.600 with Not arrived

Mean IOU: 0.534 without Not arrived

Table 4.2: Average result with Translation

This augmentation improves the detection of all classes. It most improves the detection of Vehicles (0.047).

4.1.3 Rotation

Rotation adds another 2,142 scenes to the dataset, so the models are trained on a dataset that contains 4,282 scenes, the ratio between original and new scenes is 1:1. In table 4.3 is the average training result and in appendix B.1.3 can be found result of each training.

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15348421	401276	40603	12532	31852	319874	835
Vehicle	660786	3577863	6174	887	2563	13846	559
Pedestrian	61497	21103	35318	32	416	650	15
Bikes	10477	4891	4262	43567	157	232	22
Road blocks	28943	7933	1137	22	7987	381	7
Road	350639	41609	740	66	2849	1718032	168
Not arrived	1730	4965	277	75	122	438	58838832

Class	Recall	Precision	IOU	IOU improvement
Background	0.950	0.932	0.889	0.009
Vehicle	0.839	0.881	0.754	0.018
Pedestrian	0.297	0.407	0.205	0.026
Bikes	0.685	0.765	0.565	0.049
Road blocks	0.172	0.176	0.095	0.020
Road	0.813	0.837	0.701	0.013
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.601 with Not arrived

Mean IOU: 0.535 without Not arrived

Table 4.3: Average result with Rotation

Rotation improves the detection of all classes mostly on Bikes (0.049) and on Pedestrians (0.026).

4.1.4 Mirror reflection

Mirror reflection adds another 2,142 scenes to the dataset, so the models are trained on a dataset that contains 4,282 scenes, the ratio between original and new scenes is 1:1. In table 4.4 is the average training result and in appendix B.1.4 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15194277	502423	54376	13787	47274	342109	1148
Vehicle	569364	3668617	7384	541	2513	13771	487
Pedestrian	57564	22129	37765	81	742	731	19
Bikes	15752	5795	3743	37929	121	236	31
Road blocks	26091	10386	1126	24	8336	433	15
Road	310455	46012	1065	121	5952	1750326	171
Not arrived	1807	4332	269	34	167	533	58839297

Class	Recall	Precision	IOU	IOU improvement
Background	0.941	0.939	0.887	0.007
Vehicle	0.861	0.861	0.756	0.020
Pedestrian	0.317	0.359	0.202	0.023
Bikes	0.596	0.726	0.483	-0.033
Road blocks	0.180	0.134	0.082	0.007
Road	0.828	0.830	0.708	0.020
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.588 with Not arrived

Mean IOU: 0.520 without Not arrived

Table 4.4: Average result with Mirror reflection

This augmentation, as we expected, improves the detection of Pedestrians (0.023), but it worsens the detection of Bikes (-0.033).

4.1.5 Random noise

Random noise adds another 2,142 scenes to the dataset, so models are trained on dataset that contains 4,282 scenes, ratio between original and new scenes is 1:1. In table 4.5 is the average training result and in appendix B.1.5 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15319155	434675	42868	13245	30354	313797	1299
Vehicle	480991	3747806	12591	1953	2342	16228	767
Pedestrian	64242	21943	31822	34	439	533	18
Bikes	9011	6024	3237	44887	148	265	34
Road blocks	29304	8481	859	3	7272	487	3
Road	369162	39626	963	30	2366	1701834	121
Not arrived	2076	5147	273	60	191	1069	58837624

Class	Recall	Precision	IOU	IOU improvement
Background	0.948	0.941	0.895	0.015
Vehicle	0.879	0.879	0.784	0.048
Pedestrian	0.267	0.351	0.177	-0.002
Bikes	0.706	0.746	0.569	0.053
Road blocks	0.157	0.171	0.089	0.014
Road	0.805	0.837	0.696	0.008
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.601 with Not arrived

Mean IOU: 0.535 without Not arrived

Table 4.5: Average result with Random noise

This augmentation worsens detection of Pedestrians (-0.002), however improves detection for rest of the classes, especially Vehicles (0.048), Bikes (0.053) and Road blocks (0.014).

4.1.6 Random points removal

Random points removal adds another 2,142 scenes to the dataset, so models are trained on dataset that contains 4,282 scenes, the ratio between original and new scenes is 1:1. In table 4.6 is the average training result and in appendix B.1.6 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15511984	310068	35271	11798	23222	262194	856
Vehicle	785438	3455726	4100	1066	1520	14443	384
Pedestrian	67346	19704	31034	37	290	607	12
Bikes	9860	5258	4021	44128	132	172	36
Road blocks	31869	5898	1045	4	7154	435	5
Road	388506	37642	486	23	1512	1685802	130
Not arrived	1185	3246	80	9	58	1065	58840796

Class	Recall	Precision	IOU	IOU improvement
Background	0.960	0.924	0.890	0.010
Vehicle	0.811	0.901	0.744	0.008
Pedestrian	0.261	0.419	0.189	0.010
Bikes	0.694	0.776	0.577	0.061
Road blocks	0.154	0.212	0.098	0.023
Road	0.797	0.858	0.704	0.016
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.600 with Not arrived
Mean IOU: 0.534 without Not arrived

Table 4.6: Average result with Random points removal

Random points removal improves the detection of all classes. It helps with Bikes (0.061) and to Road blocks (0.023) the most.

4.1.7 Translation and Mirror reflection

Usage of Translation and Mirror reflection adds another 4,284 scenes to the dataset (2,142 translated and 2,142 mirror reflected), so the models are trained on dataset that contains 6,426 scenes, the ratio between original and new scenes is 1:2. In table 4.7 is the average training result and in appendix B.1.7 can be found result of each training.

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15448947	336149	37418	8494	27086	296424	875
Vehicle	598567	3638046	7154	389	1992	16040	490
Pedestrian	62987	19789	35165	29	383	668	11
Bikes	13990	5115	4353	39751	140	235	22
Road blocks	29437	7767	1067	14	7752	369	4
Road	353535	34889	678	22	2415	1722428	135
Not arrived	990	2852	121	10	99	955	58841413

Class	Recall	Precision	IOU	IOU improvement
Background	0.956	0.936	0.897	0.017
Vehicle	0.853	0.900	0.779	0.043
Pedestrian	0.295	0.413	0.207	0.028
Bikes	0.625	0.816	0.547	0.031
Road blocks	0.167	0.197	0.099	0.024
Road	0.815	0.846	0.709	0.021
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.606 with Not arrived
Mean IOU: 0.540 without Not arrived

Table 4.7: Average result with Translation and Mirror reflection

This combination of uninformed augmentations has the best results (0.028 improvement in mean IOU without Not arrived) from all combinations of uninformed augmentations that we try. It improves detection of all classes, mostly Vehicles (0.043) and Bikes (0.031).

4.1.8 Translation and Mirror reflection used simultaneously

Usage of Translation, Mirror reflection, and their simultaneous usage adds another 6,426 scenes to the dataset (2,142 translated, 2,142 mirror reflected, and 2,142 simultaneous usage), so the models are trained on dataset that contains 8,568 scenes, the ratio between original and new scenes is 1:3. In table 4.8 is the average training result and in appendix B.1.8 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15402750	354449	26399	4166	9103	358050	475
Vehicle	645547	3595343	6788	218	1215	13223	343
Pedestrian	64874	19980	33226	43	193	706	9
Bikes	34132	4084	4413	20691	79	195	14
Road blocks	30381	11114	656	21	3722	515	1
Road	421634	40036	558	21	1322	1650341	189
Not arrived	1562	2837	47	14	13	311	58841654

Class	Recall	Precision	IOU	IOU improvement
Background	0.953	0.928	0.888	0.008
Vehicle	0.843	0.893	0.766	0.030
Pedestrian	0.279	0.460	0.210	0.031
Bikes	0.325	0.822	0.304	-0.212
Road blocks	0.080	0.243	0.064	-0.011
Road	0.781	0.816	0.663	-0.025
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.556 with Not arrived

Mean IOU: 0.482 without Not arrived

Table 4.8: Average result with Translation and Mirror reflection used simultaneously

This combination of uninformed augmentations has a worse mean IOU than referential learning. This is mostly due to the fact that it has a worse detection of Bikes by 0.212. The neural network starts to label the the pixels in Bikes as Background in more than half of the cases, because of this, the recall on Bikes gets worse, which we will disuse in section 4.4. On the other hand, this is the best augmentation for the detection of Pedestrians.

4.1.9 All geometrical augmentations used simultaneously

Usage of Translation, Rotation, Mirror reflection, and their simultaneous usage adds another 8,568 scenes to the dataset (2,142 translated, 2,142 rotated, 2,142 mirror reflected, and 2,142 simultaneous usage), so models are trained on dataset that contains 10,710 scenes, ratio between the original and new scenes is 1:4. In table 4.9 is the average training result and in appendix B.1.9 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15337962	386578	34740	4306	11164	380201	442
Vehicle	598086	3638046	9481	1112	1928	13779	245
Pedestrian	62575	20046	35516	30	238	620	7
Bikes	32264	4566	4628	21889	72	177	11
Road blocks	30691	10418	578	2	4160	560	1
Road	398116	34403	652	30	1517	1679207	177
Not arrived	8485	16775	1089	1386	2000	77179	58739527

Class	Recall	Precision	IOU	IOU improvement
Background	0.949	0.931	0.887	0.007
Vehicle	0.853	0.885	0.768	0.032
Pedestrian	0.298	0.411	0.209	0.030
Bikes	0.344	0.759	0.311	-0.205
Road blocks	0.090	0.211	0.066	-0.009
Road	0.794	0.784	0.651	-0.037
Not arrived	0.998	1.000	0.998	-0.002

Mean IOU: 0.556 with Not arrived
Mean IOU: 0.482 without Not arrived

Table 4.9: Average result with all geometrical augmentations used simultaneously

This combination of augmentations has a similar outcome as the previous combination of Translation and Mirror reflection. The mean IOU is worse and the detection of Bikes is worse by 0.205 caused by the decrease of recall value, however the precision gets better in this class.

4.1.10 All uninformed augmentations used simultaneously

Usage of all uninformed augmentations and their simultaneous usage adds another 12,852 scenes to the dataset (2,142 translated, 2,142 rotated, 2,142 mirror reflected, 2,142 noised, 2,142 with points removed and 2,142 simultaneous usage), so models are trained on dataset that contains 14,994 scenes, ratio between original and new scenes is 1:6. In table 4.10 is the average training result and in appendix B.1.10 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15300969	396619	30855	6189	12090	408271	400
Vehicle	512081	3720660	12377	541	1832	14985	201
Pedestrian	65415	20083	32469	47	251	752	14
Bikes	31409	4261	2752	24908	99	169	8
Road blocks	28944	12455	662	21	3637	690	1
Road	414533	33770	628	20	1821	1663157	174
Not arrived	1458	1310	84	7	16	397	58843168

Class	Recall	Precision	IOU	IOU improvement
Background	0.947	0.936	0.889	0.009
Vehicle	0.873	0.888	0.786	0.050
Pedestrian	0.273	0.413	0.195	0.016
Bikes	0.392	0.794	0.354	-0.162
Road blocks	0.078	0.200	0.058	-0.017
Road	0.787	0.797	0.655	-0.033
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.563 with Not arrived
Mean IOU: 0.490 without Not arrived

Table 4.10: Average result with all augmentations used simultaneously

Using all augmentations simultaneously also results in a worse mean IOU. The detection of Bikes gets worse by 0.162 and Road blocks (by 0.017). On the other hand, it has the biggest increase in the detection of Vehicles (0.05) from all tested augmentations.

4.2 Informed augmentation

In this section we focus on two local augmentations the Insertion and the Movement simulation. Both augmentations are trained on the large dataset (5,701 training scene), augmentations are used on all scenes in dataset (same as in section 4.1), so single usage of one augmentation adds 5,701 new scenes to dataset. This dataset is more than 2.5× larger, therefore we decrease the number of training epochs from 150, which we use for evaluation uninformed methods (section 4.1), to 100.

4.2.1 Baseline

Baseline is trained just on the large dataset. In table 4.11 is the average training result and in appendix B.2.1 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37865942	810978	202281	30694	136641	934471	2398
Vehicle	1050827	9606977	15553	2269	9279	52832	604
Pedestrian	88444	92075	105833	1609	521	1735	17
Bikes	2593	10787	28150	675	323	642	22
Road blocks	10779	4479	2014	6	19749	954	4
Road	612180	100023	9036	1836	14660	4610097	5
Not arrived	4946	9066	1564	60	550	120	143207321

Class	Recall	Precision	IOU
Background	0.947	0.955	0.907
Vehicle	0.895	0.904	0.817
Pedestrian	0.365	0.295	0.194
Bikes	0.016	0.025	0.010
Road blocks	0.520	0.109	0.099
Road	0.862	0.824	0.727
Not arrived	1.000	1.000	1.000

Mean IOU: 0.536 with Not arrived

Mean IOU: 0.459 without Not arrived

Table 4.11: Average baseline result on the large dataset

The baseline result on the large dataset has a similar pattern as the baseline result on the small dataset 4.1.1. Classes, which have small percentage in the field of view, have not ideal prediction, i.e. Bikes, Road blocks, and Pedestrian.

4.2.2 Vehicle insertion

Vehicle insertion adds another 5,701 scenes to the dataset, so the models are trained on dataset that contains 11,402 scenes, the ratio between original and new scenes is 1:1. In table 4.12 is the average training result and in appendix B.2.2 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38166487	795331	105813	17212	91778	806315	470
Vehicle	799076	9866654	10923	925	4430	56117	217
Pedestrian	88421	89153	109812	905	319	1617	8
Bikes	1848	12045	28048	233	318	687	12
Road blocks	11555	2950	1463	3	21044	969	1
Road	591424	101941	5296	285	6722	4642164	5
Not arrived	1435	2430	412	22	315	92	143218921

Class	Recall	Precision	IOU	IOU improvement
Background	0.955	0.962	0.920	0.013
Vehicle	0.919	0.908	0.840	0.023
Pedestrian	0.378	0.424	0.249	0.055
Bikes	0.005	0.013	0.004	-0.006
Road blocks	0.554	0.171	0.150	0.051
Road	0.868	0.843	0.747	0.020
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.559 with Not arrived
Mean IOU: 0.485 without Not arrived

Table 4.12: Average result with Vehicle insertion

This augmentation improves detection of all classes, except Bikes.

4.2.3 Pedestrian insertion

Pedestrian insertion adds another 5,701 scenes to the dataset, so models are trained on dataset that contains 11,402 scenes, the ratio between original and new scenes is 1:1. In table 4.13 is the average training result and in appendix B.2.3 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38229226	725903	123318	28341	106160	769910	548
Vehicle	927864	9748365	12015	1000	4605	44341	152
Pedestrian	90714	80949	113990	2956	356	1259	9
Bikes	1261	10030	31039	68	316	470	8
Road blocks	10763	3359	1473	17	21400	972	1
Road	669309	109204	4980	1386	8049	4554901	8
Not arrived	1446	7446	645	57	163	79	143213790

Class	Recall	Precision	IOU	IOU improvement
Background	0.956	0.958	0.917	0.010
Vehicle	0.908	0.913	0.835	0.018
Pedestrian	0.393	0.405	0.248	0.054
Bikes	0.002	0.002	0.001	-0.009
Road blocks	0.563	0.153	0.137	0.038
Road	0.852	0.848	0.739	0.012
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.554 with Not arrived
Mean IOU: 0.480 without Not arrived

Table 4.13: Average result with Pedestrian insertion

This augmentation improves detection of all classes, except Bikes.

4.2.4 Bikes insertion

Bikes insertion adds another 5,701 scenes to the dataset, so models are trained on dataset that contains 11,402 scenes, the ratio between original and new scenes is 1:1. In table 4.14 is the average training result and in appendix B.2.4.1 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38226218	793826	105144	6446	83005	768428	339
Vehicle	805811	9869590	9724	558	3789	48759	109
Pedestrian	93858	81810	110223	2503	231	1599	10
Bikes	1393	9613	25575	5586	314	704	7
Road blocks	11292	3171	1400	114	21173	835	0
Road	605956	96039	3214	581	5695	4636347	5
Not arrived	1653	2088	377	34	105	72	143219297

Class	Recall	Precision	IOU	IOU improvement
Background	0.956	0.962	0.921	0.014
Vehicle	0.919	0.910	0.842	0.025
Pedestrian	0.380	0.435	0.253	0.059
Bikes	0.129	0.333	0.101	0.091
Road blocks	0.557	0.188	0.163	0.064
Road	0.867	0.850	0.752	0.025
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.576 with Not arrived
Mean IOU: 0.505 without Not arrived

Table 4.14: Average result with Bikes insertion with ratio 1:1

This augmentation has the highest mean IOU of all types of informed augmentation. It improves the detection of all classes and it is the only type of augmentation, which helps Bikes detection. Improvement of IOU on one class has a positive effect on the rest of the classes. We think that increasing percentage of this class in FoV helps with better separation between Pedestrian and Bikes, which improves results of both classes.

Then we try an experiment with a higher ratio between new and original scenes. From the original scene, we make 10 new scenes. It means we add 57,010 new scenes, so the models are trained on a dataset with 62,711 scenes and the ratio between original and new scenes is 1:10. Because we increase dataset nearly 6× models converge faster, therefore each training consists of 25 epochs, with 6 validations in each epoch. In table 4.15 is the average training result and in appendix B.2.4.2 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38533226	593205	72669	8274	75044	700746	241
Vehicle	864219	9814205	7284	3365	3433	45718	117
Pedestrian	97532	82690	105522	2595	211	1675	8
Bikes	1325	9428	24750	6656	305	724	4
Road blocks	12425	2788	1403	106	20472	791	0
Road	648092	96129	2435	444	5161	4595571	3
Not arrived	874	1065	91	12	64	53	143221468

Class	Recall	Precision	IOU
Background	0.964	0.960	0.926
Vehicle	0.914	0.926	0.852
Pedestrian	0.364	0.493	0.265
Bikes	0.154	0.321	0.115
Road blocks	0.539	0.200	0.170
Road	0.859	0.860	0.754
Not arrived	1.000	1.000	1.000

Mean IOU: 0.583 with Not arrived

Mean IOU: 0.513 without Not arrived

Table 4.15: Average result with Bikes ratio with ratio 1:10

From results, we can see that additional data helps to improve results even more on all classes.

4.2.5 Movement simulation

Movement simulation adds another 5,701 scenes to the dataset, so models are trained on dataset that contains 11,402 scenes, the ratio between original and new scenes is 1:1. We perform two experiments with this augmentation. First, we simulate position of objects in 0.1 s and the other position in 0.3 s. In table 4.16 is the average training result of Movement simulation 0.1 s and in appendix B.2.5.1 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38038694	807056	171288	16436	92519	856581	832
Vehicle	774222	9887086	13744	1197	6194	55637	261
Pedestrian	90294	87383	109249	720	427	2147	12
Bikes	2822	10021	29587	29	258	466	8
Road blocks	10934	3711	2186	4	20127	1019	3
Road	581428	96791	4543	354	8633	4656083	4
Not arrived	2372	4350	1138	60	543	109	143215054

Class	Recall	Precision	IOU	IOU improvement
Background	0.951	0.963	0.918	0.011
Vehicle	0.921	0.908	0.842	0.025
Pedestrian	0.376	0.348	0.216	0.022
Bikes	0.001	0.002	0.000	-0.010
Road blocks	0.530	0.162	0.141	0.042
Road	0.871	0.836	0.743	0.016
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.551 with Not arrived

Mean IOU: 0.477 without Not arrived

Table 4.16: Average result with Movement simulation 0.1s

In table 4.17 is the average training result of Movement simulation 0.3 s and in appendix B.2.5.2 can be found result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38222964	651937	114972	18875	121874	852117	667
Vehicle	1139930	9477914	13885	2602	11460	92287	263
Pedestrian	90933	87383	104993	1569	1664	3689	3
Bikes	2443	10982	28091	364	405	905	3
Road blocks	10766	4156	1333	22	20354	1354	0
Road	645415	69924	2980	386	9430	4619699	3
Not arrived	5414	5097	890	61	576	178	143211410

Class	Recall	Precision	IOU	IOU improvement
Background	0.956	0.953	0.913	0.006
Vehicle	0.883	0.920	0.819	0.002
Pedestrian	0.362	0.396	0.233	0.039
Bikes	0.008	0.013	0.005	-0.005
Road blocks	0.536	0.133	0.118	0.019
Road	0.864	0.830	0.734	0.007
Not arrived	1.000	1.000	1.000	0.000

Mean IOU: 0.546 with Not arrived

Mean IOU: 0.470 without Not arrived

Table 4.17: Average result with Movement simulation 0.3s

From the results we can see that simulation of 0.1 s is better overall. Interestingly, simulation of 0.1 s improves Vehicles much more than simulation of 0.3 s. However, simulation of 0.3 s improves pedestrian detection more then simulation of 0.1 s. This augmentation deforms objects. This deformation increases with the amount of the shift. Vehicles are faster than pedestrians, therefore the simulation of 0.3 s can be too deforming. On the other hand, pedestrians do not move much in 0.1 s and their new location can be similar to the original position.

4.3 Range in elevation angle

In this section we compare different approaches for choosing minimal and maximal elevation angle in FoV representation (mentioned in section 2.1). Experiments have the same parameters as in section 4.2 e.i., models are trained on the large dataset (5,701 training scenes) five times, where each training consists of 100 epochs.

4.3.1 Dynamic range

This approach sets the range of elevation angle separately for each scene, based on points maximal and minimal elevation angle in scene. We use this approach in all our experiments except the experiment in subsection 4.3.2.

In table 4.18 is the average training result with dynamic range of elevation angle and in appendix B.2.1 can be found the result of each training.

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37865942	810978	202281	30694	136641	934471	2398
Vehicle	1050827	9606977	15553	2269	9279	52832	604
Pedestrian	88444	92075	105833	1609	521	1735	17
Bikes	2593	10787	28150	675	323	642	22
Road blocks	10779	4479	2014	6	19749	954	4
Road	612180	100023	9036	1836	14660	4610097	5
Not arrived	4946	9066	1564	60	550	120	143207321

Class	Recall	Precision	IOU
Background	0.947	0.955	0.907
Vehicle	0.895	0.904	0.817
Pedestrian	0.365	0.295	0.194
Bikes	0.016	0.025	0.010
Road blocks	0.520	0.109	0.099
Road	0.862	0.824	0.727
Not arrived	1.000	1.000	1.000

Mean IOU: 0.536 with Not arrived
Mean IOU: 0.459 without Not arrived

Table 4.18: Average result with dynamic range of elevation angle

4.3.2 Static range

This approach sets the range of elevation angle globally for all scenes, based on points maximal and minimal elevation angle in the whole dataset. In the table 4.19 is the average training result with static range of elevation angle and in appendix B.3 can be found result of each training.

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	31424999	780588	164172	30497	158589	938942	1749
Vehicle	1034765	8103456	16192	2786	11125	48196	446
Pedestrian	64895	86828	83504	2465	846	1590	13
Bikes	1555	12315	23812	61	443	372	14
Road blocks	6443	3885	1518	5	19887	1185	1
Road	429108	90014	7798	3171	16910	4338071	9
Not arrived	5899	8075	1835	174	722	118	151734578

Class	Recall	Precision	IOU
Background	0.938	0.953	0.897
Vehicle	0.879	0.892	0.794
Pedestrian	0.348	0.281	0.183
Bikes	0.002	0.002	0.001
Road blocks	0.604	0.099	0.093
Road	0.888	0.814	0.738
Not arrived	1.000	1.000	1.000

Mean IOU: 0.529 with Not arrived
Mean IOU: 0.451 without Not arrived

Table 4.19: Average result with static range of elevation angle

4.4 Results discussion

4.4.1 Uninformed augmentation

In table 4.20 is comparison between single usage of uninformed augmentations.

Augmentation / Class	Baseline IOU	Translation IOU	Rotation IOU	Mirror Inversion IOU	Random Noise IOU	Random Points Removal IOU
Background	0.880	**0.896**	0.889	0.887	0.895	0.890
Vehicle	0.736	0.783	0.754	0.756	**0.784**	0.744
Pedestrian	0.179	0.190	**0.205**	0.202	0.177	0.189
Bikes	0.516	0.526	0.565	0.483	**0.569**	0.577
Road blocks	0.075	0.096	0.095	0.082	0.089	**0.098**
Road	0.688	**0.711**	0.701	0.708	0.696	0.704
Not arrived	1.000	1.000	1.000	1.000	1.000	1.000
Mean	0.582	0.600	**0.601**	0.588	**0.601**	0.600
Mean without Not arrived	0.512	0.534	**0.535**	0.520	**0.535**	0.534

Table 4.20: Comparison of single usage of uninformed methods

In our experiments Rotation and Random points removal improve the mean IOU (from 0.512 to 0.535 without Not arrived) the most out of all single used uninformed augmentations.

These augmentations are based on articles [4, 2] and therefore we compare our results with the results in articles. Articles [4] focus just on Vehicle class. Article [2] represents results on Vehicle, Bikes, and, Pedestrian class. Both articles address the problem as detection (per bounding box), however we address the problem as semantic segmentation (per pixel), therefore in comparison we are focusing on the position of each augmentation in relative improvement. In table 4.21, the result comparison between our results and results from articles [4, 2] on Vehicle class is shown. Values in the table represent relative improvement of IOU on Vehicle class.

Augmentation	Our	Works [4, 2]
Translation	6.4%	$\sim$ 11%
Rotation	2.4%	$\sim$ 14%
Mirror reflection	2.7%	$\sim$ 10%
Random noise	6.5%	$\sim$ 2.5%
Random points removal	1.1%	$\sim$ 2.5%

Table 4.21: Uninformed augmentation results comparison on Vehicle class

All these augmentations improve the detection of Vehicle class in articles and in our experiments. In the articles Rotation helps most on Vehicle detection, which in our case is in the 4^{th} place. On the other hand, Random noise in our experiments helps most and in the articles it is in the 4^{th} place. Translation has a similar impact in articles and in our experiments (in both at 2^{nd} place). Position of Mirror reflection and Random points removal are also similar in our experiments and in the articles (Mirror reflection 3^{rd} place and Random points removal on 5^{th} place).

In table 4.22 is shown result comparison on Pedestrian class between our experiments and experiments in article [2]. Values in the table represent percentage improvement of IOU on pedestrian class.

Augmentation	Our	Works [2]
Random noise	-1.1%	$\sim$ 6%
Random points removal	5.6%	$\sim$ 7.5%

Table 4.22: Uninformed augmentation results comparison on Pedestrian class

From the comparison, we can see that Random points removal is very helpful in both approaches. On the other hand, Random noise in our experiments worsens the detection of pedestrians.

From the result in tables 4.21 and 4.22 we can see that all augmentations, without one exception (Random noise on Pedestrian class), which improve the results on detection task also improve the results on segmentation task. Therefore, we assume that it should work vice versa.

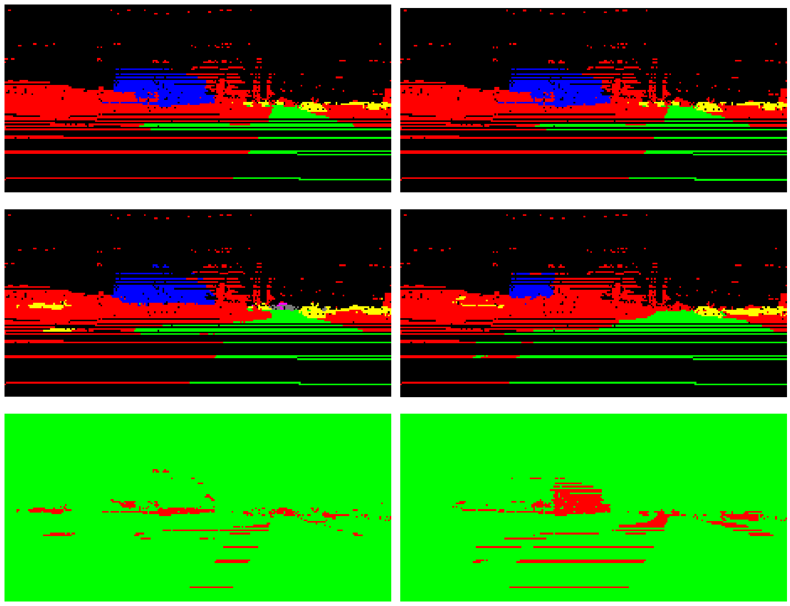
In table 4.23 is comparison between multiple uninformed augmentations usage.

Augmentation / Class	Baseline IOU	Translation and Mirror Reflection IOU	Translation and Mirror Inversion simultaneously IOU	All Geometrical Augmentations simultaneously IOU	All Augmentations simultaneously IOU
Background	0.880	**0.897**	0.888	0.887	0.889
Vehicle	0.736	0.779	0.766	0.768	**0.786**
Pedestrian	0.179	0.207	**0.210**	0.209	0.195
Bikes	0.516	**0.547**	0.304	0.311	0.354
Road blocks	0.075	**0.099**	0.064	0.066	0.058
Road	0.688	**0.709**	0.663	0.651	0.655
Not arrived	**1.000**	**1.000**	**1.000**	0.988	**1.000**
Mean	0.582	**0.606**	0.556	0.556	0.563
Mean without Not arrived	0.512	**0.540**	0.482	0.482	0.490

Table 4.23: Comparison of multiple uninformed augmentations usage

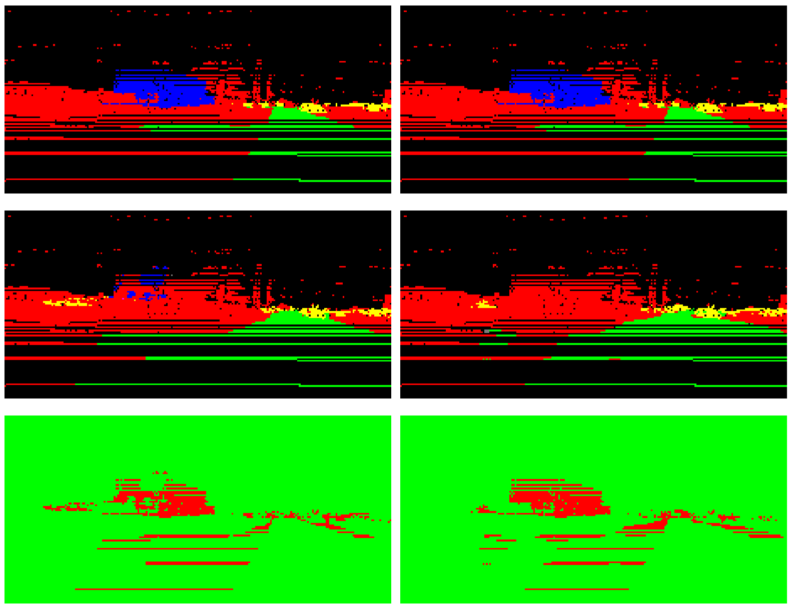
Combination of Translation and Mirror reflection helps most of single or combined usage of uninformed augmentations. It improves the mean IOU (without Not arrived) from 0.512 to 0.540.

As we mentioned in chapter 3 our Bikes class is in the small dataset corrupted by a mistake in dataset, which is illustrated in figure 3.2. Bikes class detection is the main reason why our experiments with simultaneous usage of augmentations have worse IOU than baseline. When we take a closer look at prediction of models, which is trained with the usage of simultaneous augmentations. We find out that these models predict these odd annotations as Background (as can be seen in figure 4.1), which is from our point of view the correct label for these annotations and it explains changes in confusion matrices compared to reference.



(a) : Baseline model

(b) : Translation and Mirror reflection used simultaneously



(c) : All geometrical augmentations used simultaneously

(d) : All uninformed augmentations used simultaneously

Figure 4.1: Comparison of predictions for odd Bikes annotation. Annotation (upper), prediction (middle), match (lower). Annotation and prediction: Background - red, Vehicle - yellow, Pedestrian - purple, Bikes - blue, Road blocks - gray. Match: correct prediction - green, wrong prediction - red

4.4.2 Informed augmentation

Insertion augmentations are overall more helpful then Movement simulation. This is probably cause because Insertion can increase percentage of certain class in dataset and neural network can learn more from this class. In figure 4.24 is comparison of Insertion augmentations.

Class \ Augmentation	Baseline IOU	Vehicle insertion IOU	Pedestrian insertion IOU	Bikes insertion 1:1 IOU	Bikes insertion 1:10 IOU
Background	0.907	0.920	0.917	0.921	**0.926**
Vehicle	0.817	0.840	0.835	0.842	**0.852**
Pedestrian	0.194	0.249	0.248	0.253	**0.265**
Bikes	0.010	0.004	0.001	0.101	**0.115**
Road blocks	0.099	0.150	0.137	0.163	**0.170**
Road	0.727	0.747	0.739	0.752	**0.754**
Not arrived	1.000	1.000	1.000	1.000	1.000
Mean	0.536	0.559	0.554	0.576	**0.583**
Mean without Not arrived	0.459	0.485	0.480	0.505	**0.513**

Table 4.24: Comparison of Insertion augmentations

Bikes insertion is the most helpful augmentation from all informed augmentations. Usually, classes with small percentage in the dataset tend to have not ideal detection results and Bikes has the lowest percentage in the original large dataset. Bikes insertion increases Bikes percentage in the dataset and we think this is the reason it is the only augmentation, which helps Bikes detection. Better Bikes detection consequently helps the detection of other classes. This augmentation has even better results if the ratio between the augmented and original data is higher.

Class \ Augmentation	Baseline IOU	Movement simulation 0.1s IOU	Movement simulation 0.3s IOU
Background	0.907	**0.918**	0.913
Vehicle	0.817	**0.842**	0.819
Pedestrian	0.194	0.216	**0.233**
Bikes	**0.010**	0.000	0.005
Road blocks	0.099	**0.141**	0.118
Road	0.727	**0.743**	0.734
Not arrived	1.000	1.000	1.000
Mean	0.536	**0.551**	0.546
Mean without Not arrived	0.459	**0.477**	0.470

Table 4.25: Comparison of Movement simulation augmentations

Movement simulation has a lower impact on training than Insertion. From the results we can see that simulation of 0.1 s is more beneficial for faster moving objects, e.g. vehicles. On the other hand simulation of 0.3 s is more beneficial for slower moving objects, e.g. pedestrian. We think that simulation of 0.1 s does not move pedestrian enough to make a real difference in scene respectively in FoV. Simulation of 0.3 s it is probably to deforming for faster moving vehicles.

4.4.3 Range in elevation angle

In table 4.26 can be seen comparison between dynamic and static approach.

Augmentation / Class	Dynamic range IOU	Static range IOU
Background	**0.907**	0.897
Vehicle	**0.817**	0.794
Pedestrian	**0.194**	0.183
Bikes	**0.010**	0.001
Road blocks	**0.099**	0.093
Road	0.727	**0.738**
Not arrived	1.000	1.000
Mean	**0.536**	0.529
Mean without Not arrived	**0.459**	0.451

Table 4.26: Comparison of ranges of elevation angle in FoV representation

From comparison, we can see that results of both approaches are very similar. We think that the decrease of IOU with the static approach is caused by less fortunate training. From that we determine that our premise that highest LiDAR beam in cities has at least one reflection is right.

Chapter 5

Conclusion

In this thesis, we conducted augmentations, which were used on 3D point clouds. Five of them were uninformed augmentations (section 2.2) based on articles [4, 2]. These augmentations do not take into account the context of the scene and they are applied to the whole scene. We proposed two new informed augmentations, which use the context of the scene and they are applied locally. Insertion augmentation inserts additional objects to an appropriate place in scene. Movement augmentation method exploits 3D tracking results and considers the objects speed and direction. Therefore, it can simulate the positions of objects in the future. For both methods we proposed an algorithm for resolving occlusions and visibility.

From the results of uninformed augmentations, we find out that the augmentations which improve the results of the detection task also improve the results of the semantic segmentation task. From that, we predict that it should work vice versa, i.e. if an augmentation improves result of the semantic segmentation task it should improve the result of detection task.

Insertion augmentation proved to be the best among all augmentations, which we tried. This method improves the mean IOU by 0.054. Best uninformed augmentation is the combination of Translation and Mirror reflection, which improves the mean IOU just by 0.028. From our experiments, we determine that Insertion helps most if applied to the least represented class i.e., with the lowest percentage in the dataset.

Movement simulation improves the mean IOU by 0.018. This augmentation moves moving objects in the positions, where they are expected in the future, based on their speed and direction. Our experiments show that the time which this augmentation simulates should be different for each class. For a faster moving class, e.g., Vehicles, the time should be lower (around 0.1 s) and for a slower moving class, e.g., Pedestrian, time should be higher (around 0.3 s).

Appendix A

Bibliography

[1] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps, 2019.

[2] Jaeseok Choi, Yeji Song, and Nojun Kwak. Part-aware data augmentation for 3d object detection in point cloud. *arXiv:2007.13373*, 1:1–6, 2020.

[3] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[4] Martin Hahner, Dengxin Dai, Alexander Liniger, and Luc Van Gool. Quantifying data augmentation for lidar based 3d object detection. *arXiv:2004.01643*, 1:1–7, 2020.

[5] Nickson Joram. Morphological operations in image processing [online]. Available: `https://himnickson.medium.com/morphological-operations-in-image-processing-cb8045b98fcc`. Accessed May 20, 2021.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[7] Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J. Guibas. Kpconv: Flexible and deformable convolution for point clouds, 2019.

[8] Wei Wang, Muhamad Risqi U. Saputra, Peijun Zhao, Pedro Gusmao, Bo Yang, Changhao Chen, Andrew Markham, and Niki Trigoni. Deeppco: End-to-end point cloud odometry through deep parallel neural network, 2020.

Appendix B

Results of each training

B.1 Uninformed augmentations

B.1.1 Baseline

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15285126	390942	41885	22889	34820	377722	2012
Vehicle	596561	3641738	5323	1161	4276	12728	893
Pedestrian	62912	21334	33320	175	515	736	42
Bikes	13305	4567	4613	40638	136	291	60
Road blocks	32288	4873	940	6	7883	406	17
Road	305424	49916	737	77	3412	1754413	126
Not arrived	2594	13295	403	56	83	1496	58828515

Class	Recall	Precision	IOU
Background	0.946	0.938	0.890
Vehicle	0.854	0.882	0.767
Pedestrian	0.280	0.382	0.193
Bikes	0.639	0.625	0.462
Road blocks	0.170	0.154	0.088
Road	0.830	0.817	0.700
Not arrived	1.000	1.000	1.000

Mean IOU: 0.586 with Not arrived

Mean IOU: 0.517 without Not arrived

Table B.1: Result of 1. baseline training on the small dataset

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15213975	418329	45265	17776	51391	404545	4115
Vehicle	712153	3520286	5804	262	4623	16901	2651
Pedestrian	63165	22693	31508	41	813	761	53
Bikes	9860	6735	3882	42524	156	341	112
Road blocks	29813	6161	1287	3	8133	994	22
Road	295761	52916	702	77	4172	1760328	149
Not arrived	817	2908	161	5	96	2140	58840315

Class	Recall	Precision	IOU
Background	0.942	0.932	0.881
Vehicle	0.826	0.874	0.738
Pedestrian	0.265	0.356	0.179
Bikes	0.669	0.701	0.520
Road blocks	0.175	0.117	0.076
Road	0.833	0.805	0.693
Not arrived	1.000	1.000	1.000

Mean IOU: 0.584 with Not arrived
Mean IOU: 0.514 without Not arrived

Table B.2: Result of 2. baseline training on the small dataset

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15327369	399779	66388	16890	49830	293362	1778
Vehicle	744567	3486454	7794	326	6926	15763	850
Pedestrian	58978	18534	40077	67	818	523	37
Bikes	12319	2934	5327	42532	208	220	70
Road blocks	31978	4857	997	15	8071	463	32
Road	389081	42941	1415	27	5166	1675319	156
Not arrived	4712	7772	755	17	106	1517	58831563

Class	Recall	Precision	IOU
Background	0.949	0.925	0.881
Vehicle	0.818	0.880	0.736
Pedestrian	0.337	0.326	0.199
Bikes	0.669	0.710	0.525
Road blocks	0.174	0.113	0.074
Road	0.792	0.843	0.691
Not arrived	1.000	1.000	1.000

Mean IOU: 0.586 with Not arrived
Mean IOU: 0.518 without Not arrived

Table B.3: Result of 3. baseline training on the small dataset

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15290720	406584	50371	20732	37014	348947	1028
Vehicle	850185	3383701	9115	1239	4324	13456	660
Pedestrian	60780	20481	36869	18	266	605	15
Bikes	7112	5725	4490	45895	58	307	23
Road blocks	29693	7565	791	9	7851	500	4
Road	342293	40803	1971	120	3962	1724863	93
Not arrived	9455	10665	970	91	90	2400	58822771

Class	Recall	Precision	IOU
Background	0.946	0.922	0.876
Vehicle	0.794	0.873	0.712
Pedestrian	0.310	0.353	0.197
Bikes	0.722	0.674	0.535
Road blocks	0.169	0.147	0.085
Road	0.816	0.825	0.695
Not arrived	1.000	1.000	1.000

Mean IOU: 0.586 with Not arrived
Mean IOU: 0.517 without Not arrived

Table B.4: Result of 4. baseline training on the small dataset

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15181147	451797	100544	13094	115667	291364	1783
Vehicle	716250	3496864	24289	358	12018	11858	1043
Pedestrian	60327	22622	33188	28	2062	785	22
Bikes	12610	3476	5357	41740	335	63	29
Road blocks	26330	7747	1692	9	10295	318	22
Road	432053	44188	9081	38	25597	1603069	79
Not arrived	4401	8630	2143	119	784	1087	58829278

Class	Recall	Precision	IOU
Background	0.940	0.924	0.872
Vehicle	0.820	0.867	0.728
Pedestrian	0.279	0.188	0.127
Bikes	0.656	0.754	0.540
Road blocks	0.222	0.062	0.051
Road	0.758	0.840	0.663
Not arrived	1.000	1.000	1.000

Mean IOU: 0.569 with Not arrived
Mean IOU: 0.497 without Not arrived

Table B.5: Result of 5. baseline training on the small dataset

B.1.2 Translation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15299743	424289	57794	14871	41494	316468	737
Vehicle	415004	3816599	12270	143	3370	15011	283
Pedestrian	55017	21086	41896	24	444	554	13
Bikes	8922	4941	4563	44855	197	123	9
Road blocks	28860	6368	1282	4	9420	477	2
Road	307615	34553	776	4	3836	1767165	156
Not arrived	662	3666	329	5	91	768	58840921

Class	Recall	Precision	IOU
Background	0.947	0.949	0.901
Vehicle	0.895	0.885	0.802
Pedestrian	0.352	0.352	0.214
Bikes	0.705	0.749	0.570
Road blocks	0.203	0.160	0.098
Road	0.836	0.841	0.722
Not arrived	1.000	1.000	1.000

Mean IOU: 0.615 with Not arrived

Mean IOU: 0.551 without Not arrived

Table B.6: Result of 1. training with Translation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15177219	429595	116526	15294	47409	368768	585
Vehicle	487522	3723651	30801	174	2375	17972	185
Pedestrian	51075	23890	42368	10	927	762	2
Bikes	12994	4538	4506	41259	166	138	9
Road blocks	26046	7199	2607	4	9899	658	0
Road	280931	47915	742	2	3226	1781206	83
Not arrived	1643	28264	5966	52	73	2072	58808372

Class	Recall	Precision	IOU
Background	0.939	0.946	0.892
Vehicle	0.874	0.873	0.775
Pedestrian	0.356	0.208	0.151
Bikes	0.649	0.726	0.521
Road blocks	0.213	0.154	0.098
Road	0.843	0.820	0.711
Not arrived	0.999	1.000	0.999

Mean IOU: 0.593 with Not arrived
Mean IOU: 0.525 without Not arrived

Table B.7: Result of 2. training with Translation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15380719	352799	58741	22666	30543	308717	1211
Vehicle	538849	3694373	8252	833	3960	15652	761
Pedestrian	62311	19319	36204	5	526	648	21
Bikes	9173	5038	4801	44303	118	148	29
Road blocks	30584	5718	1325	3	8180	596	7
Road	351781	34623	669	40	4923	1721899	170
Not arrived	951	3309	620	26	64	693	58840779

Class	Recall	Precision	IOU
Background	0.952	0.939	0.897
Vehicle	0.867	0.898	0.789
Pedestrian	0.304	0.327	0.187
Bikes	0.696	0.653	0.508
Road blocks	0.176	0.169	0.095
Road	0.814	0.841	0.706
Not arrived	1.000	1.000	1.000

Mean IOU: 0.597 with Not arrived
Mean IOU: 0.530 without Not arrived

Table B.8: Result of 3. training with Translation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15404326	334845	55135	9851	37722	312801	716
Vehicle	546737	3684576	10477	243	4464	15817	366
Pedestrian	62645	19559	35508	5	579	733	5
Bikes	12892	4256	4234	41854	225	133	16
Road blocks	26666	9015	2125	2	8205	399	1
Road	325564	42477	781	14	3702	1741398	169
Not arrived	1475	3313	737	42	126	576	58840173

Class	Recall	Precision	IOU
Background	0.954	0.940	0.899
Vehicle	0.864	0.899	0.788
Pedestrian	0.298	0.326	0.184
Bikes	0.658	0.805	0.567
Road blocks	0.177	0.149	0.088
Road	0.824	0.841	0.712
Not arrived	1.000	1.000	1.000

Mean IOU: 0.606 with Not arrived
Mean IOU: 0.540 without Not arrived

Table B.9: Result of 4. training with Translation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15332351	473312	50755	8737	22498	266772	971
Vehicle	563196	3673613	11982	151	1223	12072	443
Pedestrian	56172	21121	40864	13	324	527	13
Bikes	19804	5300	4711	33570	95	110	20
Road blocks	30598	6608	1507	3	7417	273	7
Road	371608	50091	780	16	1656	1689828	126
Not arrived	732	3262	621	1	19	605	58841202

Class	Recall	Precision	IOU
Background	0.949	0.936	0.892
Vehicle	0.862	0.868	0.762
Pedestrian	0.343	0.367	0.216
Bikes	0.528	0.790	0.463
Road blocks	0.160	0.223	0.103
Road	0.799	0.858	0.706
Not arrived	1.000	1.000	1.000

Mean IOU: 0.591 with Not arrived
Mean IOU: 0.523 without Not arrived

Table B.10: Result of 5. training with Translation

B.1.3 Rotation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15350055	426434	33504	10122	25258	309274	749
Vehicle	728549	3517567	3054	363	1376	11425	346
Pedestrian	64304	21939	31914	17	275	566	19
Bikes	12067	4814	3657	42701	100	244	27
Road blocks	31086	6370	838	8	7693	408	10
Road	379892	41718	321	36	1641	1690326	171
Not arrived	1606	6891	133	97	128	399	58837188

Class	Recall	Precision	IOU
Background	0.950	0.927	0.884
Vehicle	0.825	0.874	0.737
Pedestrian	0.268	0.435	0.199
Bikes	0.671	0.800	0.575
Road blocks	0.166	0.211	0.102
Road	0.800	0.840	0.694
Not arrived	1.000	1.000	1.000

Mean IOU: 0.599 with Not arrived

Mean IOU: 0.532 without Not arrived

Table B.11: Result of 1. training with Rotation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15413204	400604	32793	11788	30161	266374	472
Vehicle	537104	3702986	7992	216	3106	11024	252
Pedestrian	65617	19565	32822	48	407	571	4
Bikes	10685	4619	3897	44089	149	154	17
Road blocks	30109	7055	878	77	8032	262	0
Road	400936	47832	772	91	2785	1661495	194
Not arrived	2102	6023	190	73	80	485	58837489

Class	Recall	Precision	IOU
Background	0.954	0.936	0.896
Vehicle	0.869	0.884	0.780
Pedestrian	0.276	0.414	0.198
Bikes	0.693	0.782	0.581
Road blocks	0.173	0.180	0.097
Road	0.786	0.856	0.694
Not arrived	1.000	1.000	1.000

Mean IOU: 0.607 with Not arrived

Mean IOU: 0.541 without Not arrived

Table B.12: Result of 2. training with Rotation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15248400	447493	67021	11759	37739	341725	1259
Vehicle	580918	3653153	7919	206	2881	16686	917
Pedestrian	53779	22500	41598	3	488	638	28
Bikes	10033	5263	4752	42925	257	358	22
Road blocks	27232	8901	2040	2	7880	352	6
Road	315058	42089	1124	50	3123	1752467	194
Not arrived	1798	4318	496	45	171	288	58839326

Class	Recall	Precision	IOU
Background	0.944	0.939	0.889
Vehicle	0.857	0.873	0.762
Pedestrian	0.349	0.333	0.206
Bikes	0.675	0.781	0.567
Road blocks	0.170	0.150	0.087
Road	0.829	0.830	0.708
Not arrived	1.000	1.000	1.000

Mean IOU: 0.603 with Not arrived

Mean IOU: 0.537 without Not arrived

Table B.13: Result of 3. training with Rotation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15384614	377024	37684	18961	33493	302546	1074
Vehicle	676164	3560609	8485	1253	1616	13857	696
Pedestrian	60040	20445	37401	64	421	646	17
Bikes	8172	4689	4716	45850	40	109	34
Road blocks	28472	8620	907	18	7983	404	9
Road	340370	36902	750	83	2384	1733471	145
Not arrived	1259	4055	252	54	157	571	58840094

Class	Recall	Precision	IOU
Background	0.952	0.932	0.891
Vehicle	0.835	0.887	0.755
Pedestrian	0.314	0.415	0.218
Bikes	0.721	0.692	0.546
Road blocks	0.172	0.173	0.094
Road	0.820	0.845	0.713
Not arrived	1.000	1.000	1.000

Mean IOU: 0.602 with Not arrived
Mean IOU: 0.536 without Not arrived

Table B.14: Result of 4. training with Rotation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15345836	354829	32013	10034	32609	379454	621
Vehicle	781199	3455000	3423	2397	3837	16239	585
Pedestrian	63746	21070	32857	30	490	830	11
Bikes	11431	5071	4288	42271	241	297	11
Road blocks	27817	8723	1026	5	8349	481	12
Road	316940	39507	734	71	4316	1752401	136
Not arrived	1889	3539	317	107	77	447	58840066

Class	Recall	Precision	IOU
Background	0.950	0.927	0.884
Vehicle	0.811	0.889	0.736
Pedestrian	0.276	0.440	0.204
Bikes	0.665	0.770	0.554
Road blocks	0.180	0.167	0.095
Road	0.829	0.815	0.698
Not arrived	1.000	1.000	1.000

Mean IOU: 0.596 with Not arrived
Mean IOU: 0.529 without Not arrived

Table B.15: Result of 5. training with Rotation

B.1.4 Mirror reflection

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15346354	391827	52705	8321	31882	323405	902
Vehicle	490740	3746579	9347	261	1682	13620	451
Pedestrian	60650	20052	36976	159	566	613	18
Bikes	26866	4621	4640	27076	66	310	31
Road blocks	24912	11642	1444	24	7826	555	10
Road	322945	41575	1399	108	2065	1745844	169
Not arrived	1903	5977	235	20	100	553	58837654

Class	Recall	Precision	IOU
Background	0.950	0.943	0.898
Vehicle	0.879	0.887	0.791
Pedestrian	0.311	0.346	0.196
Bikes	0.426	0.753	0.373
Road blocks	0.169	0.177	0.095
Road	0.826	0.837	0.712
Not arrived	1.000	1.000	1.000

Mean IOU: 0.581 with Not arrived

Mean IOU: 0.511 without Not arrived

Table B.16: Result of 1. training with Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15183081	494095	62045	12793	42851	359109	1422
Vehicle	398173	3842465	6096	521	2067	12750	608
Pedestrian	52171	24413	40790	57	897	690	16
Bikes	11527	6464	3189	41933	150	319	28
Road blocks	24739	11507	1335	46	8350	405	31
Road	281803	48786	941	162	5696	1776547	170
Not arrived	983	3046	173	24	55	416	58841745

Class	Recall	Precision	IOU
Background	0.940	0.952	0.897
Vehicle	0.901	0.867	0.792
Pedestrian	0.343	0.356	0.212
Bikes	0.659	0.755	0.543
Road blocks	0.180	0.139	0.085
Road	0.840	0.826	0.714
Not arrived	1.000	1.000	1.000

Mean IOU: 0.606 with Not arrived

Mean IOU: 0.541 without Not arrived

Table B.17: Result of 2. training with Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15075818	615139	46109	20452	49158	347554	1166
Vehicle	649648	3588912	3790	1229	2317	16410	374
Pedestrian	57764	22497	37410	34	732	577	20
Bikes	9697	6288	2878	44500	132	93	22
Road blocks	27143	9629	807	11	8468	346	9
Road	315120	39220	793	100	4778	1753900	194
Not arrived	897	4007	97	41	26	308	58841066

Class	Recall	Precision	IOU
Background	0.933	0.934	0.876
Vehicle	0.842	0.837	0.724
Pedestrian	0.314	0.407	0.216
Bikes	0.700	0.671	0.521
Road blocks	0.182	0.129	0.082
Road	0.830	0.828	0.707
Not arrived	1.000	1.000	1.000

Mean IOU: 0.589 with Not arrived

Mean IOU: 0.521 without Not arrived

Table B.18: Result of 3. training with Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15249784	431633	59841	13924	66937	331976	1301
Vehicle	883874	3358323	5893	467	2969	10624	530
Pedestrian	59452	21499	36674	124	402	858	25
Bikes	17460	5065	3817	37064	79	104	21
Road blocks	26513	10329	913	32	8287	318	21
Road	333637	53288	1084	162	11723	1714035	176
Not arrived	1594	3605	215	35	137	379	58840477

Class	Recall	Precision	IOU
Background	0.944	0.920	0.873
Vehicle	0.788	0.865	0.701
Pedestrian	0.308	0.338	0.192
Bikes	0.583	0.715	0.473
Road blocks	0.179	0.092	0.064
Road	0.811	0.833	0.697
Not arrived	1.000	1.000	1.000

Mean IOU: 0.572 with Not arrived

Mean IOU: 0.500 without Not arrived

Table B.19: Result of 4. training with Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15116350	579424	51181	13445	45542	348503	951
Vehicle	424388	3806806	11797	227	3533	15454	475
Pedestrian	57786	22185	36976	33	1117	921	16
Bikes	13210	6540	4193	39073	182	356	56
Road blocks	27151	8827	1131	8	8750	541	5
Road	298774	47192	1112	75	5500	1761306	146
Not arrived	3659	5029	628	50	517	1013	58835546

Class	Recall	Precision	IOU
Background	0.936	0.948	0.890
Vehicle	0.893	0.850	0.772
Pedestrian	0.311	0.346	0.196
Bikes	0.614	0.738	0.505
Road blocks	0.189	0.134	0.085
Road	0.833	0.828	0.710
Not arrived	1.000	1.000	1.000

Mean IOU: 0.594 with Not arrived

Mean IOU: 0.526 without Not arrived

Table B.20: Result of 5. training with Mirror reflection

B.1.5 Random noise

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15389432	392887	54052	15025	30252	272238	1510
Vehicle	594103	3631286	15579	3677	2831	14095	1109
Pedestrian	61769	22021	34409	24	450	326	35
Bikes	8184	7275	2682	45070	195	178	26
Road blocks	29648	8248	944	1	7202	363	7
Road	402321	37735	1486	84	3748	1668592	139
Not arrived	1504	5816	484	224	126	699	58837589

Class	Recall	Precision	IOU
Background	0.953	0.933	0.892
Vehicle	0.852	0.885	0.767
Pedestrian	0.289	0.314	0.177
Bikes	0.709	0.703	0.545
Road blocks	0.155	0.161	0.086
Road	0.789	0.853	0.695
Not arrived	1.000	1.000	1.000

Mean IOU: 0.594 with Not arrived

Mean IOU: 0.527 without Not arrived

Table B.21: Result of 1. training with Random noise

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15304487	447191	39258	14197	37216	312117	930
Vehicle	334764	3895955	10837	174	1979	18358	613
Pedestrian	67078	22313	28216	107	540	774	6
Bikes	8843	6128	2979	45101	149	389	21
Road blocks	29311	8797	435	8	7462	399	1
Road	378372	41065	1007	33	1570	1691912	146
Not arrived	5651	7957	406	58	456	894	58831020

Class	Recall	Precision	IOU
Background	0.947	0.949	0.901
Vehicle	0.914	0.880	0.812
Pedestrian	0.237	0.339	0.162
Bikes	0.709	0.756	0.577
Road blocks	0.161	0.151	0.084
Road	0.800	0.836	0.691
Not arrived	1.000	1.000	1.000

Mean IOU: 0.604 with Not arrived
Mean IOU: 0.538 without Not arrived

Table B.22: Result of 2. training with Random noise

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15358294	472862	27259	10407	26356	259114	1104
Vehicle	464899	3771187	4628	4991	2055	14474	446
Pedestrian	67884	21199	29255	12	329	339	16
Bikes	11222	4937	3600	43606	127	83	35
Road blocks	31127	7192	532	3	7040	515	4
Road	376936	42603	631	13	1797	1692048	77
Not arrived	1520	5972	228	10	157	1609	58836946

Class	Recall	Precision	IOU
Background	0.951	0.942	0.898
Vehicle	0.885	0.872	0.783
Pedestrian	0.246	0.442	0.188
Bikes	0.686	0.739	0.552
Road blocks	0.152	0.186	0.091
Road	0.800	0.860	0.708
Not arrived	1.000	1.000	1.000

Mean IOU: 0.603 with Not arrived
Mean IOU: 0.536 without Not arrived

Table B.23: Result of 3. training with Random noise

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15375334	359999	33777	12519	33274	339192	1301
Vehicle	657671	3562297	21197	664	2764	17352	735
Pedestrian	64227	22583	30928	19	541	716	20
Bikes	8478	6156	3091	45362	166	310	47
Road blocks	28026	10000	708	1	7246	432	0
Road	380336	32115	425	8	2607	1698509	105
Not arrived	857	2878	105	1	137	1095	58841369

Class	Recall	Precision	IOU
Background	0.952	0.931	0.889
Vehicle	0.836	0.891	0.759
Pedestrian	0.260	0.343	0.173
Bikes	0.713	0.774	0.590
Road blocks	0.156	0.155	0.084
Road	0.803	0.825	0.687
Not arrived	1.000	1.000	1.000

Mean IOU: 0.597 with Not arrived

Mean IOU: 0.530 without Not arrived

Table B.24: Result of 4. training with Random noise

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15168230	500437	59995	14081	24674	386328	1651
Vehicle	353522	3878305	10715	259	2082	16864	933
Pedestrian	60255	21603	36303	12	338	510	13
Bikes	8329	5627	3837	45300	106	366	45
Road blocks	28408	8171	1680	5	7414	728	7
Road	307846	44613	1269	16	2108	1758112	141
Not arrived	848	3116	142	9	79	1048	58841200

Class	Recall	Precision	IOU
Background	0.939	0.952	0.897
Vehicle	0.910	0.869	0.800
Pedestrian	0.305	0.319	0.185
Bikes	0.712	0.759	0.581
Road blocks	0.160	0.201	0.098
Road	0.832	0.812	0.698
Not arrived	1.000	1.000	1.000

Mean IOU: 0.608 with Not arrived

Mean IOU: 0.543 without Not arrived

Table B.25: Result of 5. training with Random noise

B.1.6 Random point removal

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15506524	343980	27319	10974	20366	245650	583
Vehicle	632428	3610708	2680	894	1584	14181	205
Pedestrian	68870	20729	28498	87	224	619	7
Bikes	10302	5095	4131	43868	151	44	19
Road blocks	29718	7807	1000	8	7413	466	1
Road	396793	36099	301	25	1003	1679734	150
Not arrived	1168	3192	59	0	98	811	58841114

Class	Recall	Precision	IOU
Background	0.960	0.932	0.897
Vehicle	0.847	0.896	0.772
Pedestrian	0.239	0.445	0.184
Bikes	0.690	0.785	0.580
Road blocks	0.160	0.240	0.106
Road	0.795	0.865	0.707
Not arrived	1.000	1.000	1.000

Mean IOU: 0.607 with Not arrived

Mean IOU: 0.541 without Not arrived

Table B.26: Result of 1. training with Random point removal

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15567151	255544	26456	9636	23567	271991	1051
Vehicle	1028120	3211466	3673	174	1164	17523	560
Pedestrian	72564	18862	26550	35	363	637	23
Bikes	11186	5618	4007	42408	58	262	71
Road blocks	33198	4836	742	0	7168	459	10
Road	390897	27952	366	17	1551	1693234	88
Not arrived	1082	2997	38	8	29	1237	58841051

Class	Recall	Precision	IOU
Background	0.964	0.910	0.880
Vehicle	0.753	0.910	0.701
Pedestrian	0.223	0.429	0.172
Bikes	0.667	0.811	0.577
Road blocks	0.154	0.211	0.098
Road	0.801	0.853	0.704
Not arrived	1.000	1.000	1.000

Mean IOU: 0.590 with Not arrived
Mean IOU: 0.522 without Not arrived

Table B.27: Result of 2. training with Random point removal

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15493323	302857	60118	14393	22289	261320	1096
Vehicle	727629	3506593	9054	3907	1293	13548	656
Pedestrian	64666	18305	35304	13	228	509	9
Bikes	8464	4825	3996	46109	72	89	55
Road blocks	33806	4015	1569	13	6536	463	11
Road	423751	40481	638	18	1171	1647927	119
Not arrived	1908	3603	234	36	90	1148	58839423

Class	Recall	Precision	IOU
Background	0.959	0.925	0.890
Vehicle	0.823	0.904	0.756
Pedestrian	0.297	0.318	0.181
Bikes	0.725	0.715	0.562
Road blocks	0.141	0.206	0.091
Road	0.779	0.856	0.689
Not arrived	1.000	1.000	1.000

Mean IOU: 0.596 with Not arrived
Mean IOU: 0.528 without Not arrived

Table B.28: Result of 3. training with Random point removal

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15461173	352305	30608	12084	24153	274112	961
Vehicle	873697	3369438	3135	177	1368	14569	296
Pedestrian	68299	19951	29911	47	190	621	15
Bikes	8641	6017	3912	44566	162	294	18
Road blocks	31697	6269	905	2	7126	410	4
Road	347337	46002	565	25	1514	1718526	136
Not arrived	881	3513	33	0	24	1081	58840910

Class	Recall	Precision	IOU
Background	0.957	0.921	0.884
Vehicle	0.790	0.886	0.717
Pedestrian	0.251	0.433	0.189
Bikes	0.701	0.783	0.587
Road blocks	0.154	0.206	0.097
Road	0.813	0.855	0.715
Not arrived	1.000	1.000	1.000

Mean IOU: 0.598 with Not arrived

Mean IOU: 0.531 without Not arrived

Table B.29: Result of 4. training with Random point removal

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15531749	295655	31857	11907	25736	257900	592
Vehicle	665317	3580427	1962	180	2192	12398	204
Pedestrian	62334	20675	34911	3	448	653	10
Bikes	10710	4736	4062	43692	219	172	19
Road blocks	30926	6567	1009	0	7531	377	3
Road	383755	37678	564	34	2325	1689590	159
Not arrived	887	2928	37	3	52	1049	58841486

Class	Recall	Precision	IOU
Background	0.961	0.931	0.897
Vehicle	0.840	0.907	0.773
Pedestrian	0.293	0.469	0.220
Bikes	0.687	0.783	0.577
Road blocks	0.162	0.196	0.097
Road	0.799	0.861	0.708
Not arrived	1.000	1.000	1.000

Mean IOU: 0.610 with Not arrived

Mean IOU: 0.545 without Not arrived

Table B.30: Result of 5. training with Random point removal

B.1.7 Translation and Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15389416	374739	31336	7447	21659	329983	816
Vehicle	506157	3734698	3603	451	1528	15908	335
Pedestrian	62740	20799	34407	92	258	725	13
Bikes	20322	6424	3128	33254	147	313	22
Road blocks	27500	9846	985	1	7732	348	1
Road	344353	36483	252	36	1309	1731513	159
Not arrived	1030	2458	39	10	129	749	58842027

Class	Recall	Precision	IOU
Background	0.953	0.941	0.899
Vehicle	0.876	0.892	0.792
Pedestrian	0.289	0.467	0.217
Bikes	0.523	0.805	0.464
Road blocks	0.167	0.236	0.108
Road	0.819	0.833	0.703
Not arrived	1.000	1.000	1.000

Mean IOU: 0.598 with Not arrived

Mean IOU: 0.531 without Not arrived

Table B.31: Result of 1. training with Translation and Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15424975	320502	40727	10875	27445	330019	853
Vehicle	713546	3522275	6967	263	1919	17159	551
Pedestrian	62235	18849	36927	2	353	659	9
Bikes	9405	5032	4163	44585	134	248	43
Road blocks	30950	5731	1458	0	7915	349	10
Road	317818	28453	632	7	2155	1764938	102
Not arrived	1283	2869	106	29	31	1396	58840728

Class	Recall	Precision	IOU
Background	0.955	0.931	0.892
Vehicle	0.826	0.902	0.758
Pedestrian	0.310	0.406	0.213
Bikes	0.701	0.800	0.596
Road blocks	0.171	0.198	0.101
Road	0.835	0.835	0.716
Not arrived	1.000	1.000	1.000

Mean IOU: 0.611 with Not arrived
Mean IOU: 0.546 without Not arrived

Table B.32: Result of 2. training with Translation and Mirror reflection

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15446740	362630	32282	8087	24540	279766	1351
Vehicle	561535	3677642	4925	398	2692	14708	780
Pedestrian	65192	19321	33562	36	355	551	17
Bikes	15230	4901	4714	38364	141	247	13
Road blocks	29872	7598	1087	30	7484	337	5
Road	363991	40981	670	42	2696	1705571	154
Not arrived	770	2217	120	5	142	802	58842386

Class	Recall	Precision	IOU
Background	0.956	0.937	0.898
Vehicle	0.863	0.894	0.782
Pedestrian	0.282	0.434	0.206
Bikes	0.603	0.817	0.531
Road blocks	0.161	0.197	0.097
Road	0.807	0.852	0.708
Not arrived	1.000	1.000	1.000

Mean IOU: 0.603 with Not arrived
Mean IOU: 0.537 without Not arrived

Table B.33: Result of 3. training with Translation and Mirror reflection

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15513721	305844	38336	6907	34110	255920	558
Vehicle	734477	3502931	8821	265	1868	14150	168
Pedestrian	63957	18731	35227	4	447	661	7
Bikes	12517	4110	4997	41772	160	48	6
Road blocks	29295	7998	859	6	7979	276	0
Road	396587	37363	1131	16	4016	1674854	138
Not arrived	1160	3653	151	2	121	1064	58840291

Class	Recall	Precision	IOU
Background	0.960	0.926	0.892
Vehicle	0.822	0.903	0.755
Pedestrian	0.296	0.394	0.203
Bikes	0.657	0.853	0.590
Road blocks	0.172	0.164	0.092
Road	0.792	0.860	0.702
Not arrived	1.000	1.000	1.000

Mean IOU: 0.605 with Not arrived
Mean IOU: 0.539 without Not arrived

Table B.34: Result of 4. training with Translation and Mirror reflection

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15469886	317030	44413	9158	27676	286436	797
Vehicle	477121	3752687	11458	569	1953	18276	616
Pedestrian	60813	21246	35705	11	506	744	9
Bikes	12480	5111	4766	40784	121	322	26
Road blocks	29570	7663	948	33	7652	539	8
Road	344926	31167	707	12	1903	1735267	123
Not arrived	708	3064	190	6	73	768	58841633

Class	Recall	Precision	IOU
Background	0.958	0.944	0.906
Vehicle	0.880	0.907	0.807
Pedestrian	0.300	0.364	0.197
Bikes	0.641	0.806	0.556
Road blocks	0.165	0.192	0.097
Road	0.821	0.850	0.717
Not arrived	1.000	1.000	1.000

Mean IOU: 0.611 with Not arrived
Mean IOU: 0.547 without Not arrived

Table B.35: Result of 5. training with Translation and Mirror reflection

B.1.8 Translation and Mirror reflection used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15375018	391541	28486	5369	12585	342087	310
Vehicle	531570	3710781	5415	63	1130	13621	100
Pedestrian	60512	20058	37586	2	256	615	5
Bikes	33195	4642	3476	22080	103	114	0
Road blocks	29266	11943	700	17	3994	490	3
Road	452646	41602	683	4	1557	1617419	194
Not arrived	1658	3351	51	4	27	264	58841087

Class	Recall	Precision	IOU
Background	0.952	0.933	0.891
Vehicle	0.871	0.887	0.784
Pedestrian	0.316	0.492	0.238
Bikes	0.347	0.802	0.320
Road blocks	0.086	0.203	0.064
Road	0.765	0.819	0.654
Not arrived	1.000	1.000	1.000

Mean IOU: 0.564 with Not arrived
Mean IOU: 0.492 without Not arrived

Table B.36: Result of 1. training with Translation and Mirror reflection used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15411355	321310	24231	4014	9534	384691	261
Vehicle	795624	3439809	10187	219	2418	14295	128
Pedestrian	64909	19743	33021	103	334	917	7
Bikes	34782	4060	4217	20192	96	262	1
Road blocks	33054	8427	448	11	3885	588	0
Road	386264	33000	591	10	1408	1692661	171
Not arrived	2056	4993	105	6	18	397	58838867

Class	Recall	Precision	IOU
Background	0.954	0.921	0.882
Vehicle	0.807	0.898	0.739
Pedestrian	0.277	0.454	0.208
Bikes	0.317	0.822	0.297
Road blocks	0.084	0.220	0.065
Road	0.801	0.808	0.673
Not arrived	1.000	1.000	1.000

Mean IOU: 0.552 with Not arrived

Mean IOU: 0.477 without Not arrived

Table B.37: Result of 2. training with Translation and Mirror reflection used simultaneously

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15488996	292487	23480	3288	8960	337446	739
Vehicle	661428	3580671	6220	422	743	12376	820
Pedestrian	67134	19783	31258	12	187	644	16
Bikes	34302	3704	4825	20612	21	106	40
Road blocks	29488	12025	614	31	3616	636	3
Road	466701	44813	766	36	1490	1600099	200
Not arrived	1634	1875	20	28	0	193	58842692

Class	Recall	Precision	IOU
Background	0.959	0.925	0.889
Vehicle	0.840	0.905	0.772
Pedestrian	0.263	0.465	0.202
Bikes	0.324	0.844	0.306
Road blocks	0.078	0.241	0.063
Road	0.757	0.820	0.649
Not arrived	1.000	1.000	1.000

Mean IOU: 0.554 with Not arrived

Mean IOU: 0.480 without Not arrived

Table B.38: Result of 3. training with Translation and Mirror reflection used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15357914	394137	27698	4169	7056	363843	579
Vehicle	772405	3471057	6141	231	1060	11391	395
Pedestrian	66027	18442	33865	51	60	579	10
Bikes	35871	4067	4814	18588	94	151	25
Road blocks	29644	12160	835	36	3308	428	2
Road	397170	40590	497	27	1062	1674575	184
Not arrived	1476	2056	46	30	14	339	58842481

Class	Recall	Precision	IOU
Background	0.951	0.922	0.880
Vehicle	0.814	0.880	0.733
Pedestrian	0.284	0.458	0.213
Bikes	0.292	0.804	0.273
Road blocks	0.071	0.261	0.059
Road	0.792	0.816	0.672
Not arrived	1.000	1.000	1.000

Mean IOU: 0.547 with Not arrived

Mean IOU: 0.472 without Not arrived

Table B.39: Result of 4. training with Translation and Mirror reflection used simultaneously

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15380471	372772	28104	3994	7381	362185	489
Vehicle	466709	3774400	5978	157	725	14436	275
Pedestrian	65791	21875	30403	50	131	776	8
Bikes	32513	3948	4734	21985	81	345	4
Road blocks	30453	11019	684	11	3811	435	0
Road	405393	40176	256	31	1097	1666952	200
Not arrived	990	1912	17	2	10	364	58843147

Class	Recall	Precision	IOU
Background	0.952	0.939	0.896
Vehicle	0.885	0.893	0.801
Pedestrian	0.255	0.433	0.191
Bikes	0.346	0.838	0.324
Road blocks	0.082	0.288	0.068
Road	0.788	0.815	0.669
Not arrived	1.000	1.000	1.000

Mean IOU: 0.564 with Not arrived
Mean IOU: 0.492 without Not arrived

Table B.40: Result of 5. training with Translation and Mirror reflection used simultaneously

B.1.9 All geometrical augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15431759	316204	30671	4811	7401	364078	472
Vehicle	664856	3575016	8605	515	851	12570	267
Pedestrian	65688	19200	33488	14	172	460	12
Bikes	32390	4452	4060	22343	70	271	24
Road blocks	32466	9379	440	0	3647	481	0
Road	455197	31311	565	17	879	1625940	196
Not arrived	1394	1788	33	6	22	282	58842917

Class	Recall	Precision	IOU
Background	0.955	0.925	0.887
Vehicle	0.839	0.903	0.770
Pedestrian	0.281	0.430	0.205
Bikes	0.351	0.806	0.324
Road blocks	0.079	0.280	0.065
Road	0.769	0.811	0.652
Not arrived	1.000	1.000	1.000

Mean IOU: 0.558 with Not arrived

Mean IOU: 0.484 without Not arrived

Table B.41: Result of 1. training with all geometrical augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15370342	344693	36380	1476	11799	390327	379
Vehicle	687450	3548029	8924	448	2151	15395	283
Pedestrian	64795	20080	33158	3	94	894	10
Bikes	37329	4917	4234	16901	78	150	1
Road blocks	28891	12035	669	0	4134	684	0
Road	390952	28749	502	97	2212	1691392	201
Not arrived	36747	75867	5012	6821	9857	384328	58327810

Class	Recall	Precision	IOU
Background	0.951	0.925	0.883
Vehicle	0.832	0.879	0.747
Pedestrian	0.279	0.373	0.190
Bikes	0.266	0.656	0.233
Road blocks	0.089	0.136	0.057
Road	0.800	0.681	0.582
Not arrived	0.991	1.000	0.991

Mean IOU: 0.526 with Not arrived

Mean IOU: 0.449 without Not arrived

Table B.42: Result of 2. training with all geometrical augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15312970	407169	32295	4343	9200	389005	414
Vehicle	695934	3534620	12970	1436	2414	15097	209
Pedestrian	62548	20029	35503	108	99	741	6
Bikes	33311	3992	5273	20893	61	72	8
Road blocks	30878	10225	489	7	4157	655	2
Road	367816	34043	508	14	1234	1710334	156
Not arrived	1025	1690	122	95	10	422	58843078

Class	Recall	Precision	IOU
Background	0.948	0.928	0.883
Vehicle	0.829	0.881	0.746
Pedestrian	0.298	0.407	0.208
Bikes	0.328	0.777	0.300
Road blocks	0.090	0.242	0.070
Road	0.809	0.808	0.679
Not arrived	1.000	1.000	1.000

Mean IOU: 0.555 with Not arrived

Mean IOU: 0.481 without Not arrived

Table B.43: Result of 3. training with all geometrical augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15215460	474724	40607	5772	14051	404371	411
Vehicle	445990	3790995	10431	286	2138	12618	222
Pedestrian	59434	20858	37755	1	547	436	3
Bikes	30422	4515	4767	23734	92	69	11
Road blocks	29373	11090	764	1	4596	588	1
Road	370620	38710	1028	16	1512	1702039	180
Not arrived	1595	2651	117	9	33	298	58841739

Class	Recall	Precision	IOU
Background	0.942	0.942	0.890
Vehicle	0.889	0.873	0.787
Pedestrian	0.317	0.395	0.214
Bikes	0.373	0.796	0.341
Road blocks	0.099	0.200	0.071
Road	0.805	0.803	0.672
Not arrived	1.000	1.000	1.000

Mean IOU: 0.568 with Not arrived

Mean IOU: 0.496 without Not arrived

Table B.44: Result of 4. training with all geometrical augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15359281	390103	33749	5130	13371	353227	535
Vehicle	496200	3741571	6479	2878	2088	13219	245
Pedestrian	60412	20064	37676	25	279	573	5
Bikes	27871	4957	4808	25577	60	325	12
Road blocks	31848	9363	529	6	4267	395	5
Road	405996	39205	657	9	1752	1666332	154
Not arrived	1664	1879	162	2	78	566	58842091

Class	Recall	Precision	IOU
Background	0.951	0.937	0.894
Vehicle	0.878	0.889	0.791
Pedestrian	0.317	0.448	0.228
Bikes	0.402	0.761	0.357
Road blocks	0.092	0.195	0.067
Road	0.788	0.819	0.671
Not arrived	1.000	1.000	1.000

Mean IOU: 0.573 with Not arrived
Mean IOU: 0.501 without Not arrived

Table B.45: Result of 5. training with all geometrical augmentations used simultaneously

B.1.10 All uninformed augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15259353	385196	38449	4016	13544	454423	415
Vehicle	543652	3684218	18721	885	1352	13665	187
Pedestrian	60011	20764	36878	83	307	977	14
Bikes	29513	5200	3666	24853	19	355	4
Road blocks	26469	14233	628	20	3988	1075	0
Road	388789	36350	631	6	1133	1687023	173
Not arrived	1117	1416	53	1	40	576	58843239

Class	Recall	Precision	IOU
Background	0.945	0.936	0.887
Vehicle	0.864	0.888	0.780
Pedestrian	0.310	0.372	0.204
Bikes	0.391	0.832	0.362
Road blocks	0.086	0.196	0.063
Road	0.798	0.782	0.653
Not arrived	1.000	1.000	1.000

Mean IOU: 0.564 with Not arrived

Mean IOU: 0.491 without Not arrived

Table B.46: Result of 1. training with all uninformed augmentations used simultaneously

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15413426	348672	21858	13447	10427	347215	351
Vehicle	604963	3633190	8126	396	1848	13970	187
Pedestrian	70824	19239	27888	5	367	681	30
Bikes	30574	3398	3155	26214	133	123	13
Road blocks	30589	11272	725	4	3353	470	0
Road	488089	28075	585	33	1294	1595828	201
Not arrived	1664	881	63	9	5	290	58843530

Class	Recall	Precision	IOU
Background	0.954	0.926	0.887
Vehicle	0.852	0.898	0.777
Pedestrian	0.234	0.447	0.182
Bikes	0.412	0.654	0.338
Road blocks	0.072	0.192	0.055
Road	0.755	0.815	0.644
Not arrived	1.000	1.000	1.000

Mean IOU: 0.555 with Not arrived

Mean IOU: 0.481 without Not arrived

Table B.47: Result of 2. training with all uninformed augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15348616	387805	29442	4503	5553	378885	592
Vehicle	402546	3838112	7293	136	1013	13315	265
Pedestrian	63521	19747	35050	25	82	604	5
Bikes	33141	3993	2180	24154	70	60	12
Road blocks	29100	12726	653	1	3302	628	3
Road	392948	39199	449	13	792	1680523	181
Not arrived	1466	1524	178	14	5	426	58842829

Class	Recall	Precision	IOU
Background	0.950	0.943	0.899
Vehicle	0.900	0.892	0.812
Pedestrian	0.294	0.466	0.220
Bikes	0.380	0.837	0.354
Road blocks	0.071	0.305	0.061
Road	0.795	0.810	0.670
Not arrived	1.000	1.000	1.000

Mean IOU: 0.574 with Not arrived
Mean IOU: 0.503 without Not arrived

Table B.48: Result of 3. training with all uninformed augmentations used simultaneously

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15237393	407102	35100	4646	11040	459776	339
Vehicle	589945	3631325	19017	1139	2792	18302	160
Pedestrian	68891	19777	29323	106	89	838	10
Bikes	31828	4844	2625	24048	99	159	7
Road blocks	31630	9794	839	73	3346	728	3
Road	372056	28795	577	40	2796	1709657	184
Not arrived	1500	1558	66	3	7	263	58843045

Class	Recall	Precision	IOU
Background	0.943	0.933	0.883
Vehicle	0.852	0.885	0.767
Pedestrian	0.246	0.335	0.165
Bikes	0.378	0.800	0.345
Road blocks	0.072	0.166	0.053
Road	0.809	0.781	0.659
Not arrived	1.000	1.000	1.000

Mean IOU: 0.553 with Not arrived

Mean IOU: 0.479 without Not arrived

Table B.49: Result of 4. training with all uninformed augmentations used simultaneously

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	15246061	454323	29427	4335	19888	401057	305
Vehicle	419300	3816459	8729	151	2158	15675	208
Pedestrian	63829	20890	33209	17	412	664	13
Bikes	31993	3872	2138	25274	176	151	6
Road blocks	26934	14250	469	7	4200	552	1
Road	430783	36433	898	10	3092	1642758	131
Not arrived	1545	1171	63	8	23	434	58843198

Class	Recall	Precision	IOU
Background	0.944	0.940	0.890
Vehicle	0.895	0.878	0.796
Pedestrian	0.279	0.443	0.207
Bikes	0.397	0.848	0.371
Road blocks	0.090	0.140	0.058
Road	0.777	0.797	0.649
Not arrived	1.000	1.000	1.000

Mean IOU: 0.567 with Not arrived

Mean IOU: 0.495 without Not arrived

Table B.50: Result of 5. training with all uninformed augmentations used simultaneously

B.2 Informed augmentations

B.2.1 Baseline

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37711679	797046	285198	46219	134981	1005512	2774
Vehicle	878943	9772193	17648	1838	9834	57075	813
Pedestrian	81789	91968	112392	1806	317	1947	17
Bikes	1981	10878	29764	53	208	294	17
Road blocks	8960	4667	2946	22	20038	1354	1
Road	527750	99183	7582	3324	12031	4697964	5
Not arrived	6773	11161	1820	99	752	358	143202666

Class	Precision	Recall	IOU
Background	0.943	0.962	0.909
Vehicle	0.910	0.906	0.831
Pedestrian	0.387	0.246	0.177
Bikes	0.001	0.001	0.001
Road blocks	0.527	0.112	0.102
Road	0.878	0.815	0.732
Not arrived	1.000	1.000	1.000

Mean IOU: 0.536 with Not arrived

Mean IOU: 0.459 without Not arrived

Table B.51: Result of 1. baseline training on the large dataset

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37590981	930043	198199	35952	143274	1082493	2467
Vehicle	1059540	9596252	21202	7847	8308	44360	835
Pedestrian	74643	91728	118817	3003	914	1121	10
Bikes	3815	10076	27758	370	560	586	30
Road blocks	10772	4837	1558	0	20130	685	6
Road	438368	113519	11544	4109	21575	4758721	3
Not arrived	5180	5629	1102	63	1273	39	143210343

Class	Precision	Recall	IOU
Background	0.940	0.959	0.904
Vehicle	0.894	0.893	0.807
Pedestrian	0.409	0.313	0.215
Bikes	0.009	0.007	0.004
Road blocks	0.530	0.103	0.094
Road	0.890	0.808	0.735
Not arrived	1.000	1.000	1.000

Mean IOU: 0.537 with Not arrived
Mean IOU: 0.460 without Not arrived

Table B.52: Result of 2. baseline training on the large dataset

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37551637	1008361	212876	33650	148729	1026259	1897
Vehicle	941783	9710649	12571	847	9368	62602	524
Pedestrian	85277	99917	101224	946	487	2377	8
Bikes	3549	9930	28617	405	251	421	22
Road blocks	11107	4414	2094	5	19295	1062	11
Road	605890	95000	10494	699	15432	4620306	18
Not arrived	4541	8150	1768	8	340	91	143208731

Class	Precision	Recall	IOU
Background	0.939	0.958	0.902
Vehicle	0.904	0.888	0.812
Pedestrian	0.349	0.274	0.181
Bikes	0.009	0.011	0.005
Road blocks	0.508	0.100	0.091
Road	0.864	0.809	0.717
Not arrived	1.000	1.000	1.000

Mean IOU: 0.530 with Not arrived
Mean IOU: 0.451 without Not arrived

Table B.53: Result of 3. baseline training on the large dataset

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38161049	663163	179954	19453	117287	840846	1657
Vehicle	1063674	9594785	18095	508	10906	50156	220
Pedestrian	95883	89739	102049	220	387	1950	8
Bikes	877	12072	29057	4	282	892	11
Road blocks	10231	4441	1982	2	20320	1012	0
Road	716104	93757	9111	116	12077	4516674	0
Not arrived	5951	9130	1773	12	351	88	143206324

Class	Precision	Recall	IOU
Background	0.954	0.953	0.911
Vehicle	0.894	0.917	0.826
Pedestrian	0.352	0.298	0.192
Bikes	0.000	0.000	0.000
Road blocks	0.535	0.126	0.113
Road	0.845	0.835	0.724
Not arrived	1.000	1.000	1.000

Mean IOU: 0.538 with Not arrived
Mean IOU: 0.461 without Not arrived

Table B.54: Result of 4. baseline training on the large dataset

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38314368	656279	135181	18200	138938	717246	3197
Vehicle	1310197	9361010	8253	305	7979	49971	629
Pedestrian	104628	87024	94687	2070	504	1280	43
Bikes	2744	10983	25555	2544	317	1020	32
Road blocks	12826	4039	1492	1	18966	660	4
Road	772788	98657	6452	932	12188	4456821	1
Not arrived	2285	11261	1357	118	38	25	143208545

Class	Precision	Recall	IOU
Background	0.958	0.946	0.908
Vehicle	0.872	0.915	0.807
Pedestrian	0.326	0.347	0.202
Bikes	0.059	0.105	0.039
Road blocks	0.499	0.106	0.096
Road	0.833	0.853	0.728
Not arrived	1.000	1.000	1.000

Mean IOU: 0.540 with Not arrived
Mean IOU: 0.463 without Not arrived

Table B.55: Result of 5. baseline training on the large dataset

B.2.2 Vehicle insertion

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37929405	1026720	148101	12551	99045	767020	567
Vehicle	494285	10170345	15404	1055	4658	52341	256
Pedestrian	87825	86997	113020	369	557	1464	4
Bikes	628	11551	29487	16	459	1046	8
Road blocks	10850	2100	1531	0	22322	1184	1
Road	604661	117664	4242	73	6960	4614231	8
Not arrived	1079	1490	756	31	262	43	143219968

Class	Precision	Recall	IOU
Background	0.949	0.969	0.921
Vehicle	0.947	0.891	0.849
Pedestrian	0.389	0.362	0.231
Bikes	0.000	0.001	0.000
Road blocks	0.588	0.166	0.149
Road	0.863	0.849	0.748
Not arrived	1.000	1.000	1.000

Mean IOU: 0.557 with Not arrived

Mean IOU: 0.483 without Not arrived

Table B.56: Result of 1. training with Vehicle insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37867038	1003642	99266	11959	115258	885583	663
Vehicle	865030	9802759	4482	618	4311	60728	416
Pedestrian	88329	96635	102381	501	442	1932	16
Bikes	996	13415	27425	298	256	783	22
Road blocks	10949	2652	1333	3	22046	1003	2
Road	584584	94316	3798	597	8922	4655617	5
Not arrived	1500	1967	214	43	1091	215	143218599

Class	Precision	Recall	IOU
Background	0.947	0.961	0.912
Vehicle	0.913	0.890	0.820
Pedestrian	0.353	0.429	0.240
Bikes	0.007	0.021	0.005
Road blocks	0.580	0.145	0.131
Road	0.871	0.830	0.739
Not arrived	1.000	1.000	1.000

Mean IOU: 0.550 with Not arrived
Mean IOU: 0.475 without Not arrived

Table B.57: Result of 2. training with Vehicle insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38480450	572314	83922	20258	70143	755904	418
Vehicle	917508	9743883	13030	617	3526	59571	209
Pedestrian	93899	81369	112511	731	143	1577	6
Bikes	3922	10536	28345	34	229	127	2
Road blocks	12749	2949	1867	3	19584	836	0
Road	632042	96976	8520	219	4995	4605083	4
Not arrived	1793	2714	383	4	50	100	143218585

Class	Precision	Recall	IOU
Background	0.962	0.959	0.924
Vehicle	0.907	0.927	0.847
Pedestrian	0.388	0.453	0.264
Bikes	0.001	0.002	0.001
Road blocks	0.516	0.198	0.167
Road	0.861	0.849	0.747
Not arrived	1.000	1.000	1.000

Mean IOU: 0.564 with Not arrived
Mean IOU: 0.492 without Not arrived

Table B.58: Result of 3. training with Vehicle insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38399526	593912	77779	15949	87890	808074	279
Vehicle	1002956	9660092	8048	488	4497	62165	98
Pedestrian	90329	87502	107933	1959	337	2171	5
Bikes	2258	10226	28607	637	411	1045	11
Road blocks	13156	2706	1188	6	19896	1034	2
Road	586885	89954	3581	270	5737	4661411	1
Not arrived	1161	4192	165	21	88	90	143217912

Class	Precision	Recall	IOU
Background	0.960	0.958	0.921
Vehicle	0.900	0.925	0.838
Pedestrian	0.372	0.475	0.264
Bikes	0.015	0.033	0.010
Road blocks	0.524	0.167	0.145
Road	0.872	0.842	0.749
Not arrived	1.000	1.000	1.000

Mean IOU: 0.561 with Not arrived
Mean IOU: 0.488 without Not arrived

Table B.59: Result of 4. training with Vehicle insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38156020	780071	119999	25347	86554	814994	424
Vehicle	715601	9956193	13651	1848	5159	45782	110
Pedestrian	81724	93264	113215	966	116	942	9
Bikes	1440	14500	26377	182	238	438	20
Road blocks	10072	4344	1400	5	21375	791	1
Road	548950	110799	6339	268	6996	4674480	7
Not arrived	1644	1789	544	12	85	14	143219541

Class	Precision	Recall	IOU
Background	0.954	0.966	0.923
Vehicle	0.927	0.908	0.848
Pedestrian	0.390	0.402	0.247
Bikes	0.004	0.006	0.003
Road blocks	0.563	0.177	0.156
Road	0.874	0.844	0.753
Not arrived	1.000	1.000	1.000

Mean IOU: 0.561 with Not arrived
Mean IOU: 0.488 without Not arrived

Table B.60: Result of 5. training with Vehicle insertion

B.2.3 Pedestrian insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38369643	663018	110976	18481	103699	716704	888
Vehicle	736366	9950898	7609	373	4935	37821	342
Pedestrian	99502	76579	111032	1923	296	893	11
Bikes	1308	10240	30916	135	372	220	4
Road blocks	12188	3771	1165	8	19858	996	2
Road	687043	127254	5585	1766	7370	4518819	2
Not arrived	2512	2104	1181	156	309	6	143217361

Class	Recall	Precision	IOU
Background	0.960	0.961	0.924
Vehicle	0.927	0.918	0.856
Pedestrian	0.383	0.414	0.248
Bikes	0.003	0.006	0.002
Road blocks	0.523	0.145	0.128
Road	0.845	0.857	0.740
Not arrived	1.000	1.000	1.000

Mean IOU: 0.557 with Not arrived

Mean IOU: 0.483 without Not arrived

Table B.61: Result of 1. training with Pedestrian insertion

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37648696	1008444	204795	65064	126881	929029	500
Vehicle	521946	10140725	18501	1319	6618	49183	52
Pedestrian	84674	88640	108864	6102	392	1561	3
Bikes	1266	9186	31898	51	532	260	2
Road blocks	9708	2311	2411	15	22758	785	0
Road	540748	120555	4048	1491	13684	4667309	4
Not arrived	1287	2840	862	18	235	8	143218379

Class	Recall	Precision	IOU
Background	0.942	0.970	0.915
Vehicle	0.944	0.892	0.847
Pedestrian	0.375	0.293	0.197
Bikes	0.001	0.001	0.000
Road blocks	0.599	0.133	0.122
Road	0.873	0.826	0.737
Not arrived	1.000	1.000	1.000

Mean IOU: 0.546 with Not arrived

Mean IOU: 0.470 without Not arrived

Table B.62: Result of 2. training with Pedestrian insertion

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38257340	692144	107915	34002	101714	789644	650
Vehicle	1101641	9579681	12272	2716	2348	39500	186
Pedestrian	84521	85347	115920	2439	353	1635	21
Bikes	980	10609	30138	119	380	932	37
Road blocks	9444	5511	1190	49	20710	1077	7
Road	695821	106801	3455	2405	5318	4534024	15
Not arrived	1942	28631	274	81	96	164	143192441

Class	Recall	Precision	IOU
Background	0.957	0.953	0.914
Vehicle	0.892	0.912	0.821
Pedestrian	0.399	0.427	0.260
Bikes	0.003	0.003	0.001
Road blocks	0.545	0.158	0.140
Road	0.848	0.845	0.734
Not arrived	1.000	1.000	1.000

Mean IOU: 0.553 with Not arrived

Mean IOU: 0.478 without Not arrived

Table B.63: Result of 3. training with Pedestrian insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38405193	622907	98220	12816	106861	737027	385
Vehicle	846136	9826782	7873	117	4765	52543	128
Pedestrian	86827	81680	118750	1191	445	1338	5
Bikes	1806	8808	32304	22	184	71	0
Road blocks	12812	2960	1053	9	20017	1137	0
Road	675279	100214	4903	381	7143	4559908	11
Not arrived	805	2383	204	27	59	162	143219989

Class	Recall	Precision	IOU
Background	0.961	0.959	0.923
Vehicle	0.915	0.923	0.850
Pedestrian	0.409	0.451	0.273
Bikes	0.001	0.002	0.000
Road blocks	0.527	0.144	0.127
Road	0.853	0.852	0.743
Not arrived	1.000	1.000	1.000

Mean IOU: 0.560 with Not arrived

Mean IOU: 0.486 without Not arrived

Table B.64: Result of 4. training with Pedestrian insertion

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38465262	643003	94686	11342	91647	677149	320
Vehicle	1433235	9243739	13822	475	4360	42661	52
Pedestrian	98048	72501	115385	3127	296	872	7
Bikes	946	11308	29940	14	115	871	1
Road blocks	9664	2242	1550	4	23660	868	0
Road	747655	91196	6913	888	6730	4494446	11
Not arrived	685	1272	706	7	118	59	143220782

Class	Recall	Precision	IOU
Background	0.962	0.944	0.910
Vehicle	0.861	0.918	0.800
Pedestrian	0.398	0.439	0.264
Bikes	0.000	0.001	0.000
Road blocks	0.623	0.186	0.167
Road	0.840	0.862	0.740
Not arrived	1.000	1.000	1.000

Mean IOU: 0.554 with Not arrived

Mean IOU: 0.480 without Not arrived

Table B.65: Result of 5. training with Pedestrian insertion

B.2.4 Bikes insertion

B.2.4.1 Ratio 1:1

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38320883	600324	99722	5931	108296	848028	225
Vehicle	946408	9726118	8163	820	6450	50305	80
Pedestrian	90397	83259	112769	1936	390	1484	1
Bikes	1265	10712	26341	3632	600	645	0
Road blocks	11062	2726	803	277	22275	845	0
Road	563502	83279	3254	452	8093	4689253	6
Not arrived	1562	1477	52	14	65	95	143220364

Class	Recall	Precision	IOU
Background	0.958	0.960	0.921
Vehicle	0.906	0.926	0.844
Pedestrian	0.389	0.449	0.263
Bikes	0.084	0.278	0.069
Road blocks	0.586	0.152	0.138
Road	0.877	0.839	0.750
Not arrived	1.000	1.000	1.000

Mean IOU: 0.569 with Not arrived
Mean IOU: 0.498 without Not arrived

Table B.66: Result of 1. training with Bikes insertion with ratio 1:1

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38306400	717880	92041	4946	83270	778266	606
Vehicle	810627	9856530	13569	475	4315	52667	161
Pedestrian	98903	82807	103947	2935	202	1424	18
Bikes	2202	8526	25260	6217	477	508	5
Road blocks	12213	2549	1144	62	21455	565	0
Road	616942	93388	2724	500	7661	4626617	7
Not arrived	895	716	221	41	61	13	143221682

Class	Recall	Precision	IOU
Background	0.958	0.961	0.922
Vehicle	0.918	0.916	0.846
Pedestrian	0.358	0.435	0.244
Bikes	0.144	0.410	0.119
Road blocks	0.565	0.183	0.160
Road	0.865	0.847	0.748
Not arrived	1.000	1.000	1.000

Mean IOU: 0.577 with Not arrived

Mean IOU: 0.507 without Not arrived

Table B.67: Result of 2. training with Bikes insertion with ratio 1:1

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37746534	1405502	95013	2813	75072	658249	226
Vehicle	765099	9919083	5804	199	2092	45991	76
Pedestrian	96051	82708	108641	961	99	1769	7
Bikes	1064	11461	28094	1733	146	687	10
Road blocks	10301	2618	1610	26	22587	846	0
Road	682399	86568	3031	86	4115	4571632	8
Not arrived	1068	907	164	8	82	50	143221350

Class	Recall	Precision	IOU
Background	0.944	0.960	0.909
Vehicle	0.924	0.862	0.805
Pedestrian	0.374	0.448	0.256
Bikes	0.040	0.297	0.037
Road blocks	0.595	0.217	0.189
Road	0.855	0.866	0.755
Not arrived	1.000	1.000	1.000

Mean IOU: 0.564 with Not arrived

Mean IOU: 0.492 without Not arrived

Table B.68: Result of 3. training with Bikes insertion with ratio 1:1

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38309520	620715	159320	11771	80071	801638	374
Vehicle	737844	9940154	9777	561	3284	46662	62
Pedestrian	84853	85620	116173	2185	310	1081	14
Bikes	858	9487	19833	12075	211	713	18
Road blocks	11346	4077	1690	106	19935	830	4
Road	596829	102552	4202	1202	4743	4638308	3
Not arrived	4085	6517	844	104	176	72	143211831

Class	Recall	Precision	IOU
Background	0.958	0.964	0.925
Vehicle	0.926	0.923	0.859
Pedestrian	0.400	0.373	0.239
Bikes	0.280	0.431	0.204
Road blocks	0.525	0.183	0.157
Road	0.867	0.845	0.748
Not arrived	1.000	1.000	1.000

Mean IOU: 0.590 with Not arrived

Mean IOU: 0.522 without Not arrived

Table B.69: Result of 4. training with Bikes insertion with ratio 1:1

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38447755	624711	79626	6770	68320	755959	268
Vehicle	769080	9906067	11307	739	2807	48174	170
Pedestrian	99086	74657	109586	4502	154	2238	13
Bikes	1578	7879	28347	4277	140	971	3
Road blocks	11538	3886	1757	100	19617	1090	0
Road	570110	114409	2862	665	3867	4655925	1
Not arrived	659	824	608	5	142	133	143221258

Class	Recall	Precision	IOU
Background	0.962	0.964	0.928
Vehicle	0.922	0.923	0.857
Pedestrian	0.378	0.468	0.264
Bikes	0.099	0.251	0.076
Road blocks	0.516	0.206	0.173
Road	0.871	0.852	0.756
Not arrived	1.000	1.000	1.000

Mean IOU: 0.579 with Not arrived

Mean IOU: 0.509 without Not arrived

Table B.70: Result of 5. training with Bikes insertion with ratio 1:1

B.2.4.2 Ratio 1:10

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38600527	563253	81516	10329	65220	662435	129
Vehicle	877622	9799800	7062	3373	3412	46971	104
Pedestrian	96073	79830	111032	1674	58	1559	10
Bikes	1278	8886	25793	6254	114	869	1
Road blocks	13726	1997	1153	59	20296	753	4
Road	701297	89097	3566	537	4343	4548995	4
Not arrived	129	366	93	8	20	85	143222928

Class	Recall	Precision	IOU
Background	0.965	0.958	0.926
Vehicle	0.913	0.929	0.854
Pedestrian	0.383	0.482	0.271
Bikes	0.145	0.281	0.106
Road blocks	0.534	0.217	0.183
Road	0.851	0.865	0.751
Not arrived	1.000	1.000	1.000

Mean IOU: 0.584 with Not arrived

Mean IOU: 0.515 without Not arrived

Table B.71: Result of 1. training with Bikes insertion with ratio 1:10

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38670664	459730	61289	7764	97636	686199	127
Vehicle	1008247	9665270	10107	891	3577	50224	28
Pedestrian	105123	78268	102555	1661	407	2218	4
Bikes	1498	8492	25587	6361	655	601	1
Road blocks	10524	2573	1179	58	23090	564	0
Road	732086	74319	1913	512	6410	4532599	0
Not arrived	2615	1156	35	3	185	43	143219592

Class	Recall	Precision	IOU
Background	0.967	0.954	0.924
Vehicle	0.900	0.939	0.851
Pedestrian	0.353	0.506	0.263
Bikes	0.147	0.369	0.118
Road blocks	0.608	0.175	0.157
Road	0.848	0.860	0.745
Not arrived	1.000	1.000	1.000

Mean IOU: 0.580 with Not arrived
Mean IOU: 0.509 without Not arrived

Table B.72: Result of 2. training with Bikes insertion with ratio 1:10

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38628555	548576	73458	6435	52650	673141	594
Vehicle	945150	9735536	8001	4204	2825	42269	359
Pedestrian	103486	83175	100133	1916	58	1453	15
Bikes	1243	8842	27682	4571	173	682	2
Road blocks	13429	3322	1776	15	18186	1260	0
Road	677191	88681	2498	210	2208	4577040	11
Not arrived	718	1350	100	3	45	11	143221402

Class	Recall	Precision	IOU
Background	0.966	0.957	0.926
Vehicle	0.907	0.930	0.849
Pedestrian	0.345	0.469	0.248
Bikes	0.106	0.263	0.082
Road blocks	0.479	0.239	0.190
Road	0.856	0.864	0.754
Not arrived	1.000	1.000	1.000

Mean IOU: 0.578 with Not arrived
Mean IOU: 0.508 without Not arrived

Table B.73: Result of 3. training with Bikes insertion with ratio 1:10

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38361441	790937	65049	3445	72527	689821	189
Vehicle	645482	10038861	4732	4566	3091	41570	42
Pedestrian	85972	84863	116143	1756	244	1255	3
Bikes	1077	9884	23311	7735	269	919	0
Road blocks	12302	3187	1516	11	20100	872	0
Road	599580	119248	2209	244	4461	4622093	4
Not arrived	347	300	49	14	41	15	143222863

Class	Recall	Precision	IOU
Background	0.959	0.966	0.928
Vehicle	0.935	0.909	0.855
Pedestrian	0.400	0.545	0.300
Bikes	0.179	0.435	0.145
Road blocks	0.529	0.200	0.169
Road	0.864	0.863	0.760
Not arrived	1.000	1.000	1.000

Mean IOU: 0.594 with Not arrived
Mean IOU: 0.526 without Not arrived

Table B.74: Result of 4. training with Bikes insertion with ratio 1:10

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38404947	603531	82037	13401	87191	792135	167
Vehicle	844598	9831559	6520	3791	4263	47560	53
Pedestrian	97008	87316	97749	5970	289	1892	12
Bikes	1533	11038	21380	8362	316	549	17
Road blocks	12147	2861	1395	387	20688	510	0
Road	530308	109301	1992	721	8387	4697130	0
Not arrived	561	2156	179	34	31	113	143220555

Class	Recall	Precision	IOU
Background	0.961	0.963	0.926
Vehicle	0.916	0.923	0.851
Pedestrian	0.337	0.463	0.242
Bikes	0.194	0.256	0.124
Road blocks	0.545	0.171	0.149
Road	0.878	0.848	0.759
Not arrived	1.000	1.000	1.000

Mean IOU: 0.579 with Not arrived
Mean IOU: 0.509 without Not arrived

Table B.75: Result of 5. training with Bikes insertion with ratio 1:10

B.2.5 Movement simulation

B.2.5.1 0.1 seconds simulation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37851557	983705	142592	20304	105832	878844	575
Vehicle	721460	9945728	9785	318	7497	53326	230
Pedestrian	96862	90527	100123	381	456	1868	19
Bikes	2910	11810	28034	6	296	126	13
Road blocks	11125	3739	1792	7	20338	984	3
Road	592718	104726	6144	170	13496	4630578	7
Not arrived	2246	4602	1100	9	824	137	143214711

Class	Recall	Precision	IOU
Background	0.947	0.964	0.914
Vehicle	0.926	0.892	0.833
Pedestrian	0.345	0.346	0.209
Bikes	0.000	0.000	0.000
Road blocks	0.535	0.137	0.122
Road	0.866	0.832	0.737
Not arrived	1.000	1.000	1.000

Mean IOU: 0.545 with Not arrived

Mean IOU: 0.469 without Not arrived

Table B.76: Result of 1. training with Movement simulation 0.1 seconds

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38186626	616639	286730	8269	86271	797952	922
Vehicle	1081289	9572210	22719	1295	3645	56799	387
Pedestrian	80732	78994	126816	1080	171	2438	5
Bikes	2816	11510	28113	49	409	286	12
Road blocks	10038	4108	3197	4	19639	1002	0
Road	617761	74975	3644	118	5880	4645459	2
Not arrived	1854	3888	1568	42	995	219	143215063

Class	Recall	Precision	IOU
Background	0.955	0.955	0.914
Vehicle	0.891	0.924	0.830
Pedestrian	0.437	0.268	0.199
Bikes	0.001	0.005	0.001
Road blocks	0.517	0.168	0.145
Road	0.869	0.844	0.748
Not arrived	1.000	1.000	1.000

Mean IOU: 0.548 with Not arrived
Mean IOU: 0.473 without Not arrived

Table B.77: Result of 2. training with Movement simulation 0.1 seconds

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38291264	696750	82805	7405	119353	784797	1035
Vehicle	979385	9689277	8748	523	10250	49796	365
Pedestrian	100856	89205	97562	603	633	1365	12
Bikes	1881	11403	28588	24	329	962	8
Road blocks	10956	3863	1557	2	20727	878	5
Road	633934	110881	3354	64	13103	4586502	1
Not arrived	2494	7017	927	15	281	3	143212892

Class	Recall	Precision	IOU
Background	0.958	0.957	0.918
Vehicle	0.902	0.913	0.831
Pedestrian	0.336	0.436	0.234
Bikes	0.001	0.003	0.000
Road blocks	0.546	0.126	0.114
Road	0.858	0.846	0.741
Not arrived	1.000	1.000	1.000

Mean IOU: 0.548 with Not arrived
Mean IOU: 0.473 without Not arrived

Table B.78: Result of 3. training with Movement simulation 0.1 seconds

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38258247	694091	105802	22915	63737	838165	452
Vehicle	604913	10062361	7865	2035	3392	57655	123
Pedestrian	92370	89604	105174	323	201	2549	15
Bikes	4503	8119	29618	20	114	816	5
Road blocks	11993	3812	1386	5	19751	1040	1
Road	586412	91978	4039	736	4655	4660014	5
Not arrived	2266	3405	508	117	309	115	143216909

Class	Recall	Precision	IOU
Background	0.957	0.967	0.927
Vehicle	0.937	0.919	0.865
Pedestrian	0.362	0.413	0.239
Bikes	0.000	0.001	0.000
Road blocks	0.520	0.214	0.179
Road	0.871	0.838	0.746
Not arrived	1.000	1.000	1.000

Mean IOU: 0.565 with Not arrived
Mean IOU: 0.493 without Not arrived

Table B.79: Result of 4. training with Movement simulation 0.1 seconds

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37605777	1044098	238514	23287	87405	983151	1177
Vehicle	484063	10165857	19606	1815	6188	60611	204
Pedestrian	80653	88588	116572	1216	675	2519	13
Bikes	2003	7266	33583	49	146	144	4
Road blocks	10562	3037	3000	3	20184	1195	7
Road	476317	101398	5537	683	6034	4757865	5
Not arrived	3004	2839	1589	121	308	71	143215697

Class	Recall	Precision	IOU
Background	0.941	0.973	0.916
Vehicle	0.947	0.891	0.848
Pedestrian	0.402	0.279	0.197
Bikes	0.001	0.002	0.001
Road blocks	0.531	0.167	0.145
Road	0.890	0.820	0.744
Not arrived	1.000	1.000	1.000

Mean IOU: 0.550 with Not arrived
Mean IOU: 0.475 without Not arrived

Table B.80: Result of 5. training with Movement simulation 0.1 seconds

B.2.5.2 0.3 seconds simulation

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38288708	500611	135157	39615	134029	883793	1496
Vehicle	1070926	9520914	17926	3545	12278	112465	290
Pedestrian	100022	82145	103658	1003	388	3017	3
Bikes	3024	10961	28236	67	296	607	4
Road blocks	11210	3887	1583	22	19795	1490	1
Road	603633	41667	4398	853	9650	4687632	6
Not arrived	10533	8120	1610	117	919	191	143202139

Class	Recall	Precision	IOU
Background	0.958	0.955	0.916
Vehicle	0.887	0.936	0.836
Pedestrian	0.357	0.354	0.216
Bikes	0.002	0.001	0.001
Road blocks	0.521	0.112	0.101
Road	0.877	0.824	0.738
Not arrived	1.000	1.000	1.000

Mean IOU: 0.544 with Not arrived

Mean IOU: 0.468 without Not arrived

Table B.81: Result of 1. training with Movement simulation 0.3 seconds

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	37691547	978102	137654	16360	197468	961993	285
Vehicle	640966	9987832	15835	5400	14294	73782	235
Pedestrian	69834	94134	110181	5358	6850	3877	2
Bikes	1524	10728	27236	1560	846	1300	1
Road blocks	9091	5000	1351	13	20910	1623	0
Road	608368	119058	2247	275	9192	4608699	0
Not arrived	871	11530	342	111	755	389	143209631

Class	Recall	Precision	IOU
Background	0.943	0.966	0.912
Vehicle	0.930	0.891	0.835
Pedestrian	0.380	0.374	0.232
Bikes	0.036	0.054	0.022
Road blocks	0.550	0.084	0.078
Road	0.862	0.815	0.721
Not arrived	1.000	1.000	1.000

Mean IOU: 0.543 with Not arrived

Mean IOU: 0.467 without Not arrived

Table B.82: Result of 2. training with Movement simulation 0.3 seconds

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38367479	569830	86307	14880	71723	872964	226
Vehicle	1683672	8931503	8716	1664	5042	107401	346
Pedestrian	97286	80368	107462	638	178	4303	1
Bikes	2967	10184	29024	148	275	597	0
Road blocks	13639	3071	1000	62	18491	1725	0
Road	620744	45793	2603	598	4099	4674001	1
Not arrived	1137	1527	582	40	68	133	143220142

Class	Recall	Precision	IOU
Background	0.960	0.941	0.905
Vehicle	0.832	0.926	0.780
Pedestrian	0.370	0.456	0.257
Bikes	0.003	0.008	0.002
Road blocks	0.487	0.185	0.155
Road	0.874	0.826	0.738
Not arrived	1.000	1.000	1.000

Mean IOU: 0.548 with Not arrived

Mean IOU: 0.473 without Not arrived

Table B.83: Result of 3. training with Movement simulation 0.3 seconds

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38437614	498793	107347	13134	116516	808968	1037
Vehicle	1192829	9419930	16212	792	10434	97800	347
Pedestrian	95814	87415	101849	316	546	4295	1
Bikes	806	11688	29206	15	312	1161	7
Road blocks	11487	3309	1336	0	20968	888	0
Road	684795	52027	3449	47	13597	4593924	0
Not arrived	1837	2933	1285	27	1037	77	143216433

Class	Recall	Precision	IOU
Background	0.961	0.951	0.916
Vehicle	0.877	0.935	0.827
Pedestrian	0.351	0.391	0.227
Bikes	0.000	0.001	0.000
Road blocks	0.552	0.128	0.116
Road	0.859	0.834	0.734
Not arrived	1.000	1.000	1.000

Mean IOU: 0.546 with Not arrived
Mean IOU: 0.470 without Not arrived

Table B.84: Result of 4. training with Movement simulation 0.3 seconds

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	38329473	712352	108398	10386	89638	732867	295
Vehicle	1111260	9529393	10739	1609	15255	69988	100
Pedestrian	91710	92856	101816	532	358	2956	8
Bikes	3894	11350	26756	32	297	863	3
Road blocks	8406	5513	1398	17	21608	1045	1
Road	709536	91076	2203	160	10615	4534239	10
Not arrived	12695	1378	631	12	101	104	143208708

Class	Recall	Precision	IOU
Background	0.959	0.952	0.914
Vehicle	0.887	0.912	0.818
Pedestrian	0.351	0.404	0.231
Bikes	0.001	0.003	0.001
Road blocks	0.569	0.157	0.140
Road	0.848	0.849	0.737
Not arrived	1.000	1.000	1.000

Mean IOU: 0.549 with Not arrived
Mean IOU: 0.473 without Not arrived

Table B.85: Result of 5. training with Movement simulation 0.3 seconds

B.3 Static range of elevation angle

PL / GT	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	31618496	716267	122699	15935	115416	909318	1409
Vehicle	654656	8496659	9288	3546	9271	43061	488
Pedestrian	76581	85506	74684	1244	503	1610	15
Bikes	1830	11649	24118	12	521	409	35
Road blocks	9071	3005	1002	8	18500	1340	0
Road	393660	109399	7431	4625	9281	4360688	1
Not arrived	9384	15615	1840	478	650	124	151723312

Class	Recall	Precision	IOU
Background	0.944	0.965	0.913
Vehicle	0.922	0.900	0.836
Pedestrian	0.311	0.310	0.184
Bikes	0.000	0.000	0.000
Road blocks	0.562	0.120	0.110
Road	0.893	0.820	0.747
Not arrived	1.000	1.000	1.000

Mean IOU: 0.541 with Not arrived

Mean IOU: 0.465 without Not arrived

Table B.86: Result of 1. training with static range of elevation angle

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	31295967	804111	166119	28730	166691	1036656	1266
Vehicle	1371970	7764204	15435	1448	16474	47059	379
Pedestrian	55717	87889	89633	5179	723	988	14
Bikes	790	11387	25467	99	371	453	7
Road blocks	5137	5348	1412	4	19971	1053	1
Road	398625	81904	7549	4093	17748	4375157	9
Not arrived	6378	8884	2052	116	727	6	151733240

Class	Recall	Precision	IOU
Background	0.934	0.945	0.886
Vehicle	0.842	0.886	0.760
Pedestrian	0.373	0.291	0.196
Bikes	0.003	0.002	0.001
Road blocks	0.607	0.090	0.085
Road	0.896	0.801	0.733
Not arrived	1.000	1.000	1.000

Mean IOU: 0.523 with Not arrived
Mean IOU: 0.443 without Not arrived

Table B.87: Result of 2. training with static range of elevation angle

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	31594720	739458	209865	35880	115723	802353	1541
Vehicle	1488152	7640206	28703	3454	5677	50474	303
Pedestrian	62259	76861	98873	739	307	1087	17
Bikes	1476	9608	26949	8	252	266	15
Road blocks	7120	4295	2581	0	18022	908	0
Road	541480	72325	10403	880	10679	4249318	0
Not arrived	6267	9040	2885	62	448	177	151732524

Class	Recall	Precision	IOU
Background	0.943	0.937	0.887
Vehicle	0.829	0.893	0.754
Pedestrian	0.412	0.260	0.190
Bikes	0.000	0.000	0.000
Road blocks	0.547	0.119	0.109
Road	0.870	0.832	0.740
Not arrived	1.000	1.000	1.000

Mean IOU: 0.526 with Not arrived
Mean IOU: 0.447 without Not arrived

Table B.88: Result of 3. training with static range of elevation angle

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	31204174	843072	165437	31580	213332	1039132	2813
Vehicle	580004	8554964	14089	3259	9647	54268	738
Pedestrian	65480	92687	76843	2504	852	1772	5
Bikes	1716	14236	21167	180	768	503	4
Road blocks	4693	3718	621	10	22476	1406	2
Road	333090	93854	6097	4160	21778	4426096	10
Not arrived	4041	4257	1464	216	1287	199	151739939

Class	Recall	Precision	IOU
Background	0.931	0.969	0.905
Vehicle	0.928	0.891	0.833
Pedestrian	0.320	0.269	0.171
Bikes	0.005	0.004	0.002
Road blocks	0.683	0.083	0.080
Road	0.906	0.801	0.740
Not arrived	1.000	1.000	1.000

Mean IOU: 0.533 with Not arrived

Mean IOU: 0.455 without Not arrived

Table B.89: Result of 4. training with static range of elevation angle

GT \ PL	Background	Vehicle	Pedestrian	Bikes	Road blocks	Road	Not arrived
Background	31411642	800035	156744	40363	181783	907253	1720
Vehicle	1079047	8061249	13447	2227	14556	46121	322
Pedestrian	64438	91200	77490	2659	1845	2497	14
Bikes	1966	14697	21362	6	303	229	11
Road blocks	6196	3061	1975	4	20466	1221	3
Road	478689	92588	7513	2100	25068	4279098	29
Not arrived	3425	2580	936	2	498	85	151743877

Class	Recall	Precision	IOU
Background	0.938	0.951	0.894
Vehicle	0.875	0.889	0.789
Pedestrian	0.323	0.277	0.175
Bikes	0.000	0.000	0.000
Road blocks	0.622	0.084	0.080
Road	0.876	0.817	0.732
Not arrived	1.000	1.000	1.000

Mean IOU: 0.524 with Not arrived

Mean IOU: 0.445 without Not arrived

Table B.90: Result of 5. training with static range of elevation angle