Supplementary Material

Contents

1 Bayesian Posterior Update Formula

As discussed in Section IV-B of the paper, we parametrize the object-level state probability distribution as the product of a Gaussian distribution for the geometric change, l, and a Beta distribution for the stationarity score, v:

$$\begin{aligned} p(l,v \mid \mathbf{z}_1 \dots \mathbf{z}_T) &:= q(l,v \mid \mu_T, \sigma_T, \alpha_T, \beta_T) \\ &:= \mathcal{N}(l \mid \mu_T, \sigma_T^2)\text{Beta}(v \mid \alpha_T, \beta_T) \end{aligned} \tag{1}$$

where μ_T and σ_T^2 represent the mean and variance of l respectively, and α_T and β_T are the number of observed inlier and outlier measurements with respect to the model. Moreover, $\{\mathbf{z}_t\}_{t=1\dots T}$ are the measurement features extracted from the object and its associated observation for timestamp t. Each measurement, $\mathbf{z}_t$, contains the magnitude of geometric change, Δ_t, and the stationarity class, s_t, for the object:

$$\mathbf{z}_t = \{\Delta_t \in \mathbb{R}, s_t \in \{0,1\}\}. \tag{2}$$

Following Section IV-C of the paper, the full measurement likelihood distribution is the product of a Gaussian-Uniform mixture and k Bernoulli distributions:

$$\begin{aligned} & p(\mathbf{z}_T \mid l, v, \mathbf{z}_1 \dots \mathbf{z}_{T-1}) \propto p(\Delta_T \mid l, v)\, p(s_T \mid v) \\ &:= (v\mathcal{N}(\Delta_T \mid l, \tau^2) + (1-v)\mathcal{U}(\Delta_T \mid -\Delta_{\max}, \Delta_{\max}))\text{Bernoulli}(s_T \mid v)^k \\ &= (v\mathcal{N}(\Delta_T \mid l, \tau^2) + (1-v)\mathcal{U}(\Delta_T \mid -\Delta_{\max}, \Delta_{\max}))\, v^{ks_T}(1-v)^{k(1-s_T)} \end{aligned} \tag{3}$$

Where τ^2 is the geometric measurement variance and $\Delta_{\max}$ is the maximum change valid. The factor k is used to balance the relative importance between the geometric consistency likelihood $p(\Delta_T \mid l, v)$ and the stationarity likelihood $p(s_T \mid v)$ to adjust the model's behaviour. The value of k can be adaptively selected based on the semantic class of the object and whether an inlier or and outlier measurement has been received. As we will show below, k acts as a weight in the Beta stationarity update rule for the posterior.

Assuming the prior has the parametrization $q(l, v \mid \mu, \sigma, \alpha, \beta)$, we would like to match the first and second moments in l and v of the true posterior

$$p(l,v \mid \Delta_T, s_T, \mu, \sigma, \alpha, \beta) \propto p(\Delta_T \mid l, v)\, p(s_T \mid v)\, q(l,v \mid \mu, \sigma, \alpha, \beta) \tag{4}$$

to those of our approximated posterior:

$$q(l,v \mid \mu', \sigma', \alpha', \beta'). \tag{5}$$

Substituting the parametrized prior (1) and the measurement likelihood (3), the true posterior (4) becomes

$$\begin{aligned} & p(l,v \mid \Delta_T, s_T, \mu, \sigma, \alpha, \beta) \propto \\ & \left(v\tfrac{1}{\sqrt{2\pi}\tau}e^{-\frac{(\Delta_T - l)^2}{2\tau^2}} + (1-v)\mathcal{U}(\Delta_T)\right)\left(\tfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(l-\mu)^2}{2\sigma^2}}\tfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}v^{a-1}(1-v)^{b-1}v^{ks_T}(1-v)^{k(1-s_T)}\right), \end{aligned} \tag{6}$$

which can be written as the sum of an inlier term and an outlier term :

$$p(l,v \mid \Delta_T, s_T, \mu, \sigma, \alpha, \beta) \propto p_{\text{in}} + p_{\text{out}}$$

where

$$\begin{aligned} p_{\text{in}} &= v\frac{1}{\sqrt{2\pi}\tau}e^{-\frac{(\Delta_T - l)^2}{2\tau^2}}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(l-\mu)^2}{2\sigma^2}}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}v^{\alpha-1+ks_T}(1-v)^{\beta-1+k(1-s_T)} \\ p_{\text{out}} &= (1-v)\mathcal{U}(\Delta_T)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(l-\mu)^2}{2\sigma^2}}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}v^{\alpha-1+ks_T}(1-v)^{\beta-1+k(1-s_T)} \end{aligned} \tag{7}$$

Collecting like terms,

$$\begin{aligned} p_{\text{in}} &= \eta_1 \mathcal{N}\left(l \mid m, \gamma^2\right) \mathcal{N}\left(\Delta_T \mid \mu, \tau^2 + \sigma^2\right) \text{Beta}(\alpha + k s_T + 1, \beta + k(1 - s_T)) \\ p_{\text{out}} &= \eta_2 \mathcal{U}(\Delta_T) \mathcal{N}\left(l \mid \mu, \sigma^2\right) \text{Beta}(\alpha + k s_T, \beta + k(1 - s_T) + 1) \end{aligned} \tag{8}$$

where we define

$$\begin{aligned} \gamma^2 &= \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \\ m &= \gamma^2 \left(\frac{\mu}{\sigma^2} + \frac{\Delta_T}{\tau^2}\right) \\ \eta_1 &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha + k s_T + 1)\Gamma(\beta + k(1 - s_T))}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + k + 1)} \\ \eta_2 &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha + k s_T)\Gamma(\beta + k(1 - s_T) + 1)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + k + 1)} \end{aligned} \tag{9}$$

Here, m is the updated mean of l after fusing in Δ_T, and γ^2 is the updated variance of l after fusing in Δ_T. We can further write the true posterior (6) as a weighted mixture of two Gaussian-Beta distributions

$$\begin{aligned} &p(l, v \mid \Delta_T, s_T, \mu, \sigma, \alpha, \beta) \\ &\quad \propto C_1 \mathcal{N}\left(l \mid m, \gamma^2\right) \text{Beta}(v \mid \alpha + k s_T + 1, \beta + k(1 - s_T)) \\ &\qquad + C_2 \mathcal{N}\left(l \mid \mu, \sigma^2\right) \text{Beta}(v \mid \alpha + k s_T, \beta + k(1 - s_T) + 1) \end{aligned} \tag{10}$$

by defining

$$\begin{aligned} C_1 &= \eta_1 \mathcal{N}\left(\Delta_T \mid \mu, \tau^2 + \sigma^2\right) \\ C_2 &= \eta_2 \mathcal{U}(\Delta_T). \end{aligned} \tag{11}$$

Intuitively, the weights, C_1 and C_2, are the probability of the measurement, $\mathbf{z}_T$, being an inlier and outlier, respectively. The first Gaussian-Beta distribution models the effect of fusing an inlier measurement, where the geometric change, l, is updated with a new mean, m, and variance, γ^2. The positive probabilistic counter, α, in the Beta stationarity distribution first gains 1 for the consistent geometric measurement. Then, both counters, α and β, are adjusted by the k-weighted stationarity measurement. The second Gaussian-Beta distribution models the effect of fusing an outlier measurement, where we discard the outlying geometric change measurement, Δ_T, and only update the Beta stationarity distribution. In practice, we limit the Beta probabilistic counters, α and β, to a maximum threshold to ensure fast response when changes occur.

Finally, we match the first and second moments with respect to l of the true posterior (10) to the approximated posterior (5):

$$\mu' = C_1 m + C_2 \mu, \tag{12}$$

and

$$\sigma'^2 + \mu'^2 = C_1 \left(m^2 + \gamma^2\right) + C_2 \left(\mu^2 + \sigma^2\right). \tag{13}$$

Similarly, we match the first and second moments with respect to v of the true posterior (10) to the approximated posterior (5):

$$\frac{\alpha'}{\alpha' + \beta'} = C_1 \frac{\alpha + k s_T + 1}{\alpha + \beta + k + 1} + C_2 \frac{\alpha + k s_T}{\alpha + \beta + k + 1}, \tag{14}$$

and

$$\frac{(\alpha'+1)\,\alpha'}{(\alpha'+\beta'+1)\,(\alpha'+\beta')} = C_1\left(\frac{(\alpha+ks_T+2)(\alpha+ks_T+1)}{(\alpha+\beta+k+2)(\alpha+\beta+k+1)}\right) + C_2\left(\frac{(\alpha+ks_T+1)(\alpha+ks_T)}{(\alpha+\beta+k+2)(\alpha+\beta+k+1)}\right) \quad (15)$$

Solving the system of equations (12)-(15), we obtain the parameters for the approximated posterior $q(l,v \mid \mu',\sigma',a',b')$:

$$\begin{aligned}
\mu' &= C_1 m + C_2 \mu \\
\sigma' &= \left[C_1\left(m^2+\gamma^2\right) + C_2\left(\mu^2+\sigma^2\right) - \mu'^2\right]^{\frac{1}{2}} \\
\alpha' &= \frac{(C_1\theta\phi + C_2\theta\psi - \theta^2)}{(\theta^2 - C_1\phi - C_2\psi)} \\
\beta' &= \frac{(C_1\theta\phi + C_2\theta\psi - \theta^2)(C_1\kappa + C_2\zeta - 1)}{(C_1\phi + C_2\psi - \theta^2)(C_1\kappa + C_2\zeta)}
\end{aligned} \quad (16)$$

where

$$\begin{aligned}
\kappa &= \frac{\alpha+ks_T+1}{\alpha+\beta+k+1} \\
\zeta &= \frac{(\alpha+ks_T)}{(\alpha+\beta+k+1)} \\
\theta &= C_1\kappa + C_2\zeta \\
\phi &= \frac{(\alpha+ks_T+2)(\alpha+ks_T+1)}{(\alpha+\beta+k+1)(\alpha+\beta+k+2)} \\
\psi &= \frac{(\alpha+ks_T+1)(\alpha+ks_T)}{(\alpha+\beta+k+1)(\alpha+\beta+k+2)}
\end{aligned} \quad (17)$$

2 TorWIC_9-2 Results

We further evaluate our framework, POCD, on another route from the TorWIC dataset. The groundtruth schematics of the route is shown in Figure 1. Qualitative map reconstruction results are shown in Figure 2 and quantitative results are listed in Table 1. Again, POCD produces the most accurate map comparing to the baseline methods.

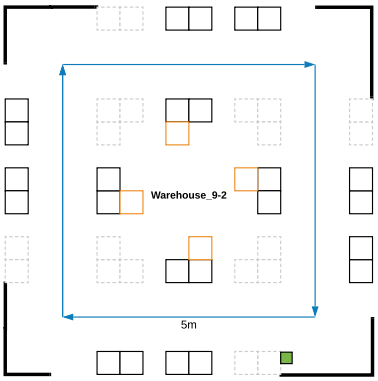

Figure 1: The TorWIC_9-2 ground truth schematic. Green box: fixed AprilTag for drift reference; black: stationary box or fence; orange: additional boxes; dotted grey: original position of shifted boxes.

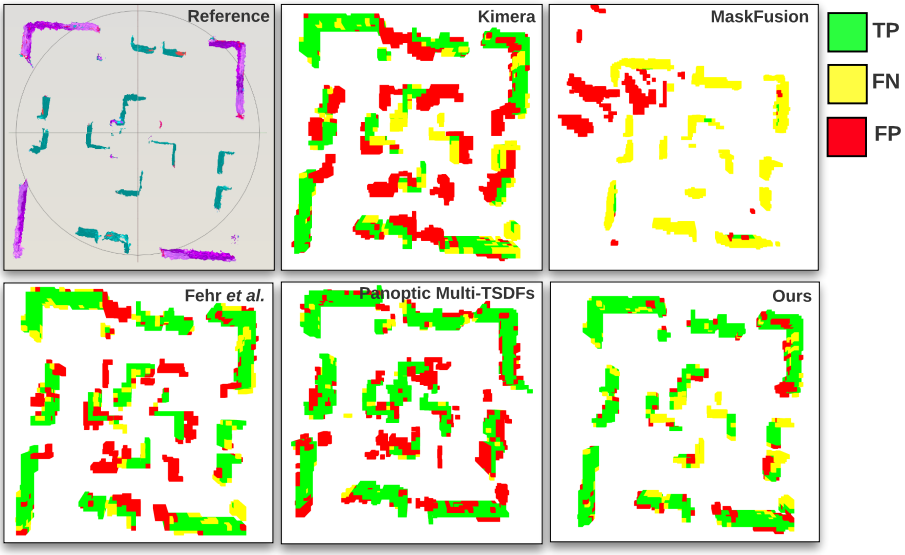

Figure 2: Bird's-eye-view qualitative 3D reconstruction results of the TorWIC_9-2 route. The reconstruction produced by POCD is compared against that of Kimera, MaskFusion, Fehr *et al.*, and Panoptic Multi-TSDFs. The green sections represent true positives, the yellow sections represent false negatives, and the red sections represent false positives. The first image is the reference map of the routes' final configurations. We use a voxel size of 20 cm for the TorWIC routes.

Table 1: Quantitative map reconstruction results on the TorWIC_9-2 dataset.

TorWIC_9-2	**Precision ↑**	**Recall (TPR) ↑**	**FPR ↓**
Kimera	53.6	76.3	7.9
MaskFusion	24.8	6.9	**1.1**
Fehr *et al.*	72.5	80.3	3.8
Panoptic Multi-TSDFs	72.2	**83.7**	4.3
POCD(ours)	**83.0**	78.8	1.9
Improvement	10.5	-4.9	-0.8

3 List of Parameters in POCD

We list the parameters used in POCD for the TorWIC and ToyCar dataset in Table 3. Symbols, descriptions and values are provided.

Table 2: Parameters used in the POCD framework for the TorWIC and ToyCar datasets.

Parameter	**Description**	**TorWIC Dataset**	**ToyCar Dataset**
v	voxel size	5 cm	2 cm
s	stationarity class	0: box, 1: fence	0: toy cars, 1: cubes/airplane
λ_1	association weight for position difference	1	1
λ_2	association weight for orientation difference	1	1
λ_3	association s for semantic consistency	1000	1000
λ_{diff}	scaling factor for geometric change estimation	1.6	1.8
θ_{dist}	maximum centroid-distance for association	0.9 m	0.9 m
θ_{sim}	percent of outliers from ICP during association	10%	10%
θ_{cutoff}	cutoff cost for feasible association	3.6	3.6
θ_{vis}	threshold of points within camera FOV to label an unassociated object as *unobserved*	20%	20%
θ_{stat}	stationarity threshold for object pruning	0.4	0.45
θ_{depth}	cutoff threshold for depth information	3 m	4 m
ν	initial stationarity expectation	semi-static: 0.67, dynamic: 0.67	semi-static: 0.67, dynamic: 0.67
ν_{max}	max stationarity expectation	0.9999	0.9999
μ	initial geometric change expectation	0 cm	0 cm
σ	initial geometric change expectation	0.5 m	0.5 m
Δ_{max}	maximum geometric change cutoff	4 m	0.6 m
τ	measurement standard deviation	20 cm	3 cm
k	weights for stationarity class measurements	outlier measurements, dynamic-class: 3 inlier measurements, dynamic-class: 0 outlier measurements, static-class: 0 inlier measurements, static-class: 3	outlier measurements, dynamic-class: 3 inlier measurements, dynamic-class: 0 outlier measurements, static-class: 0 inlier measurements, static-class: 3

4 Dataset Information

In this section, we provide additional information on the released TorWIC dataset[1].

4.1 Robot and Sensors

The dataset was collected on the OTTO 100 Autonomous Mobile Robot[2], remote controlled by a human operator at walking speed. We record sensor measurements from an Intel RealSense D435i RGB-D camera, a wheel encoder, an IMU unit, and a Hokuyo UAM501 2D laser scanner, all rigidly mounted on the platform. Figure 3 shows the robot platform and the sensor frames, and Table 3 lists the specifications and formats of the sensor measurements.

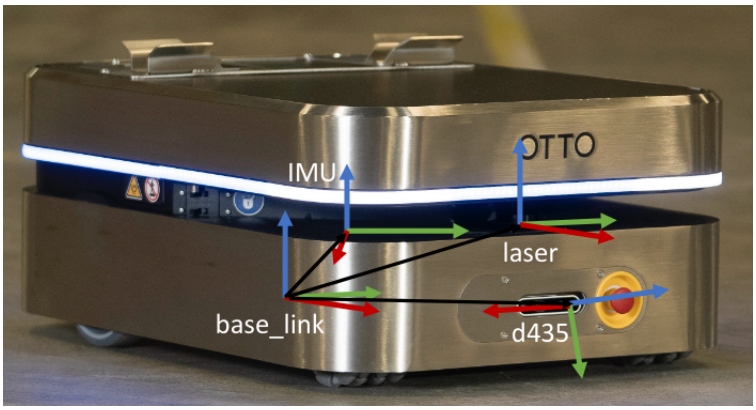

Figure 3: An image of the Otto 100 platform with the sensors used to collect the dataset.

4.1.1 List of Available Data

Table 3: Available Sensor Data and Specifications

Sensor	Data	FPS	Resolution	FOV	Format
Intel RealSense D435i	colour	15	640x360	H:87 V:58 D:95	RGB image
Intel RealSense D435i	depth	15	640x360	H:87 V:58 D:95	16-bit image
2D Hokuyo LiDAR	laser scan	15	1080	270°	PCD point cloud file
Base	pose	15	-	-	text file id sec nsec p_x p_y p_z q_x q_y q_z q_w
Base	odometry	24	-	-	text file id sec nsec p_x p_y p_z q_x q_y q_z q_w v_x v_y v_z ω_x ω_y ω_z
IMU	accel & gyro	50	-	-	text file id sec nsec a_x a_y a_z ω_x ω_y ω_z

4.1.2 Sensor Calibration

The RealSense camera was calibrated with the Intel's OEM calibration tool. The extrinsics for the sensors are factory calibrated. It is assumed that the calibrations remain intact for all trajectories. The sensor calibrations are provided in the data Google Drive link[3] in the text file. Sensor extrinsics are also provided in the ROS bags under the `tf_static` topic.

[1] https://github.com/Viky397/TorWICDataset
[2] https://ottomotors.com/100#stats
[3] https://drive.google.com/drive/folders/12-h2OPmlmxLk0Y9C3Hr5glkalUp66oEJ?usp=sharing

4.1.3 Sensor Synchronization

The sensors on the OTTO 100 platform are not synchronized with each other. For our dataset, we use the RealSense image timestamps as the reference, and take the measurements with the closest timestamp from the LiDAR and the poses. The provided odometry and IMU data is not sub-sampled. Please contact us if you need the unprocessed, raw data (as ROS bags).

4.2 Dataset Scenarios

The dataset provides 18 trajectories in 18 scenarios, including the baseline setup. Each trajectory contains the robot traversing through a static configuration of the environment, starting and finishing at the fixed April-Tag. Users can stitch the trajectory together with the provided script[1] to create routes with structural changes in the scene. A high level overview of the scenarios and trajectories is listed in Table 4, and more details can be found in the scenario setup document [4]. Two sample frames are shown in Figure 4.

Table 4: Dataset Trajectory Breakdown

Changes	Number of trajectories	Total Number of Frames	Description
Baseline	1	2377	Default configuration with all box walls and fences forming a square.
Box shifts and rotations	9	37746	Various box walls are rotated and shifted.
Removing boxes	4	14457	Sections of the box walls are removed.
Moving fences	3	12148	Fences are shifted outwards and inwards, along with the box walls.
Adding new boxes	1	3979	The four fences are covered with stacks of boxes.

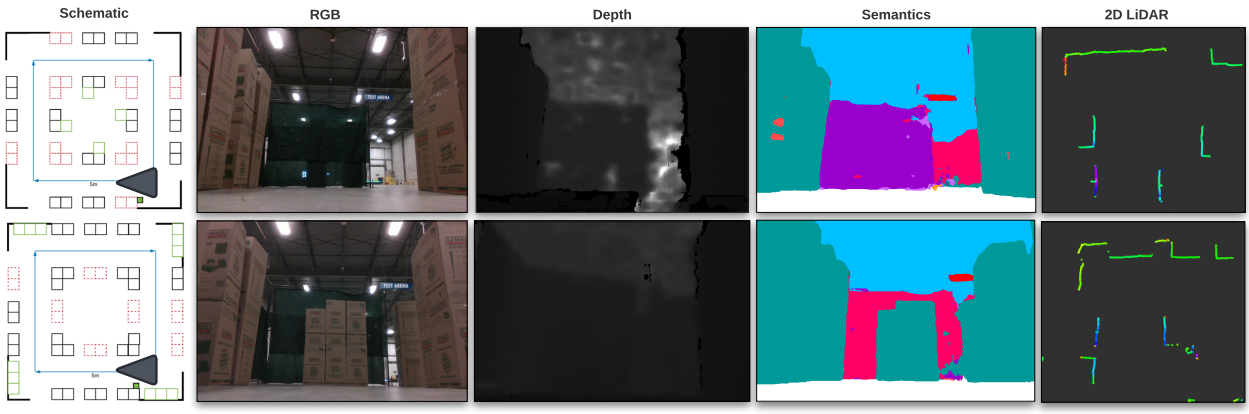

Figure 4: An example of two frames captures by the robot at the AprilTag in two scenarios (Scenario_2-2 and Scenario_4-1). Changes include 3 stacks of boxes added in front of the fence, and an additional box wall to the right of the fence. Note that the fence in the distance is labelled as a Miscellaneous Static Feature: purple in the top row, and Wall/Fence/Pillar: hot pink in the bottom row due to semantic aliasing in our trained segmentation network.

4.3 Robot Global Poses

The poses of the robot were obtained offline using a proprietary LiDAR-based SLAM solution based on the Hokuyo laser scans.

[4] https://github.com/Viky397/TorWICDataset/blob/main/TorWIC_Dataset.pdf

4.4 Depth Images

The depth images from RealSense D435i RGB-D camera have been aligned to the colour images for per-pixel correspondence. The measurements are in millimeters. The `uint16` values can be converted into float values and multiplied by 0.001 to get depth in meters.

4.5 Ground-truth Segmentation for Fine-tuning

The provided semantic segmentation masks are not perfect. A semantic segmentation model was trained to produce the semantic masks for the dataset. Unfortunately, the full training data is proprietary and cannot be released. However, we release a subset of this data[5] such that users can fine-tune their models, if needed. Within the training set, there are 79 folders with unique ID's. Within each folder, there are 3 sets of images, each within its own sub-folder. Each image folder contains an image of the individual semantic mask, the source RGB image, the combined semantic mask image, the combined semantic indexed image, and an annotation `.json` file. For training purposes, the combined indexed image should be used (named `combined_indexedImage.png`). Each pixel holds the class ID of the semantic class corresponding to Table 5. The provided ROS bags of the dataset contain colourized masks that correspond to the Class ID column in Table 5.

Table 5: Semantic Class Lookup Table

Semantic Class	**uint16 Class ID**	**Colour**	**RGB**
Background	0	black	[0,0,0]
Driveable Ground	1	white	[255,255,255]
Ceiling	2	baby blue	[0,191,255]
Ego Vehicle	3	bright green	[0,255,0]
Wall/Fence/Pillar	4	hot pink	[255,0,102]
Miscellaneous Static Feature	5	purple	[153,0,204]
Shelf/Rack	6	dark blue	[51, 51, 204]
Goods Materials	7	teal	[0, 153, 153]
Fixed Machinery	8	baby pink	[255, 204, 255]
Cart/Pallet Jack	9	orange	[255,153,0]
Pylons	10	yellow	[255,255,0]
Text Region	11	bright red	[255,0,0]
Miscellaneous Non-Static Feature	12	baby purple	[204, 102, 255]
Person	13	watermelon	[255, 77, 77]
Forklift/Truck	14	dark green	[0, 153, 51]
Miscellaneous Dynamic Feature	15	grey	[191, 191, 191]

[5] https://drive.google.com/file/d/1ovm4ycVrQfpuseI2Kc8TofS-LI0Nly_I/view?usp=sharing