



ENHANCING THE PERFORMANCE OF MULTI-OBJECT TRACKING IN TRAFFIC STREAM VIDEOS THROUGH INITIAL VELOCITY AND FRAME-SKIPPING STRATEGIES

Chia-Chun Lin¹, Albert Y. Chen²

1) Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, Taiwan. Email: r12521522@ntu.edu.tw

2) Professor, Department of Civil Engineering, National Taiwan University, Taiwan. Email: albertchen@ntu.edu.tw

Abstract: Multi-object tracking in videos is an important task in various domains, such as traffic engineering and construction management. This paper proposes two methods, Grid Mean State and InCo-Skip, to improve multi-object tracking performance, particularly under frame-skipping scenarios. The study focuses on traffic flow counting, using YOLOv8 for vehicle tracking. Initial tests show that while car tracking remains accurate, motorcycles suffer a significant accuracy degradation when homogeneous frame skipping is applied. Grid Mean State addresses the issue by utilizing velocity vectors from earlier frames, and InCo-Skip provides an alternative skipping strategy to balance computational efficiency and accuracy. The combined methods show a substantial enhancement in counting accuracy, achieving up to 28.2% improvement for motorcycles under challenging conditions.

Keywords: Object tracking, Kalman filter, Initial velocity, Frame skip

1. INTRODUCTION

Object tracking is a critical task in various fields, including traffic analysis (Chen et al., 2020), construction safety management (Lin et al., 2021), infrastructure inspection (Wang et al., 2021), and mixed reality (Kinoshita et al., 2022). Accurately tracking objects, especially in real-time, presents challenges in achieving a balance between computational resources and accuracy.

In traditional Multi-Object Tracking (MOT) algorithms, such as ByteTrack (Zhang et al., 2022), the Kalman filter is widely used for motion prediction, where the state vector includes parameters such as position, aspect ratio, and velocity. However, the assumption of zero initial velocity during the first few frames can lead to unsuccessful tracking, particularly under Homogenous frame-skipping scenarios. This issue is amplified when tracking smaller objects like motorcycles, where motion dynamics change rapidly.

As a result, this study aims to address the limitations of the zero initial velocity assumption by proposing a method that improves the Kalman filter's initial velocity estimation. At the same time, this study hopes to introduce better frame-skipping strategies to mitigate the amplification of initial velocity inaccuracies during frame skipping.

2. LITERATURE REVIEW

2.1 Introduction to Kalman Filter

The Kalman Filter (Kalman, 1960), first introduced in the 1960s, was originally applied in spacecraft navigation and signal processing to estimate system states with noise and uncertainty. Its principle lies in modeling the motion of a target through a state-space model, which helps to predict future states while reducing noise interference by continuously updating the Kalman Filter through measurements. This method is most effective under the assumption that both the system model and the noise follow Gaussian distributions.

The Kalman filter is based on two primary equations: the prediction and the update.

(1) Prediction

$$\hat{x}_k = F_k x_{k-1} + B_k u_k \quad (1)$$

$$P_k = F_k P_{k-1} F_k^T + Q_k \quad (2)$$

Equation (1) predict the state \hat{x}_k at the next time step based on the previous state \hat{x}_{k-1} and control inputs u_k . In equation (2), P_k is the covariance matrix that reflects the uncertainty in the state prediction, with F representing the state transition matrix and Q_k representing the process noise.

(2) Update

$$K_k = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \quad (3)$$

$$\hat{x}_k = \hat{x}_k + K_k (z_k - H_k \hat{x}_k) \quad (4)$$

$$P_k = (I - K_k H_k) P_k \quad (5)$$

From Equation (3) and (4), K_k is the Kalman gain, which determines how much the predictions should be adjusted based on the new measurements z_k . H represents the measurement matrix, and R accounts for measurement noise. At last, Equation (5) updates the covariance matrix P_k by applying the K_k .

2.2 Kalman Filter in Multi-Object Tracking (MOT) and ByteTrack

The Kalman Filter, as introduced earlier, has evolved significantly since its initial application in aerospace and signal processing. By the 2010s, it became a cornerstone in MOT algorithms due to its ability to estimate and predict the states of objects moving through space (Li et al., 2010; Pathan et al., 2009; Weng et al., 2006). In MOT, the Kalman Filter models each object's motion using a state vector, which typically represents the object's position, velocity, and sometimes shape attributes. A common form of the state vector in MOT is:

$$\text{state vector} = [x, y, a, h, \dot{x}, \dot{y}, \dot{a}, \dot{h}]$$

Where:

x and y represent the center coordinates of the bounding box.

a denotes the aspect ratio of the bounding box.

h is the height of the bounding box.

\dot{x} and \dot{y} are the velocities in the x and y directions, respectively.

\dot{a} is the rate of change of the aspect ratio.

\dot{h} is the rate of change of the height.

This allows the Kalman Filter to predict both the object's future position and its motion behavior. However, a key challenge in these algorithms is the estimation of initial velocity values, which significantly affect tracking performance. Many algorithms often default to a zero initial velocity, which may cause either failure of tracking, or tracking delays and inaccuracies until the velocity is correctly estimated after several frames.

Recent advances in MOT, such as ByteTrack, combine the Kalman Filter with IoU (Intersection over Union) (Rezatofighi et al., 2019) metrics for object association. As shown in Figure 1, IoU is used to compare bounding boxes across frames to link detections, while the Kalman Filter predicts motion between frames. Although this combination enhances tracking performance, the issue of poor initial velocity estimation remains, especially under conditions where frame skipping is required to reduce computational load. Frame skipping reduces the number of frames available for updates, making accurate initial velocity estimation even more critical for maintaining tracking accuracy.

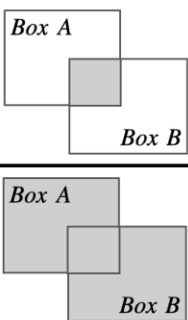
$$IoU (Box A, Box B) = \frac{\text{Intersection}}{\text{Union}}$$


Figure 1. IoU Calculation Principle

Numerous studies (Linderoth et al., 2011; Weiner, 1981; Zhao & Huang, 2017) have highlighted the importance of accurate initialization of the Kalman Filter. While these studies focus on general initialization strategies for the Kalman filter, they do not specifically address initial velocity assumptions. Additionally, research on frame-skipping strategies (Park et al., 2020), introduces algorithms to determine when to skip frames, which adds computational complexity rather than simplifying it.

In response to these issues, our study focuses on developing methods to improve initial velocity assumption and testing new frame-skipping strategies to maintain accuracy while minimizing additional computational overhead across various tracking scenarios.

3. METHODOLOGY

In this section, we describe two key methods employed in our tracking system: Grid Mean State and InCo-Skip Method. These methods serve to enhance accuracy of initial trajectory estimation and efficiency in multi-object tracking systems, ensuring that velocity estimations are initialized correctly and frame-skipping strategies are optimized.

3.1 Grid Mean State

The Grid Mean State method is used to initialize the velocity vectors in our tracking system. This method is divided into two primary phases, namely the Initial Tracking Phase and Full Video Tracking Phase. The process is illustrated in Figure 2. In Initial Tracking Phase, we calculate Grid Mean State by leveraging the state vectors of objects being successfully tracked. Then we use Grid Mean State to replace the initial zero velocity assumption in subsequent Full Video Tracking Phase.

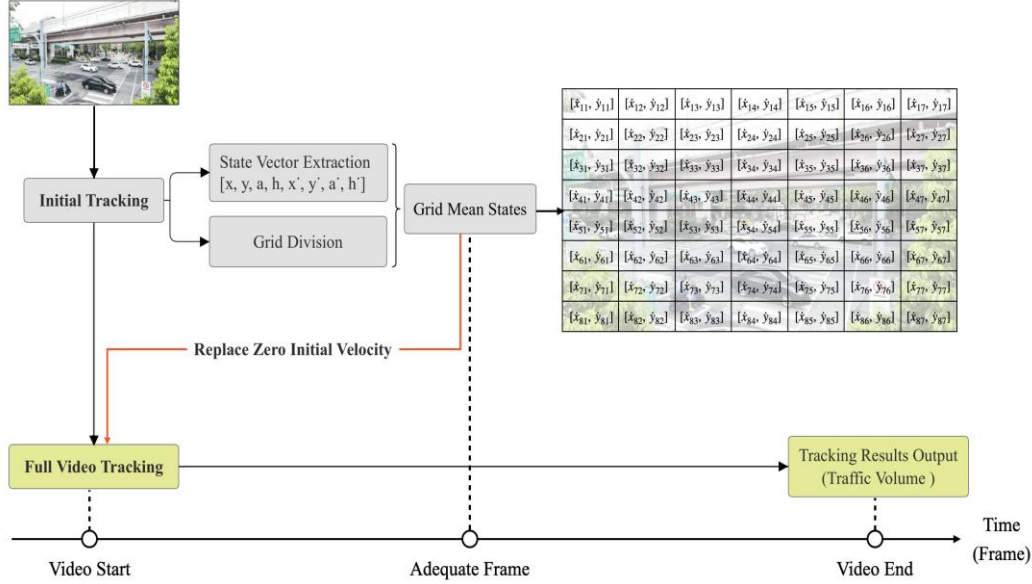


Figure 2. Grid Mean State Calculation Process

The procedure can be broken down into the following steps:

(1) State vector extraction

In the Initial Tracking Phase, YOLOv8 is used to detect and track objects in the video frame. For each tracked object, the state vector at a given frame t is defined as follows:

$$\text{state vector } (t) = [x, y, a, h, \dot{x}, \dot{y}, \dot{a}, \dot{h}]$$

(2) Grid division and object assignment

The entire frame is divided into a grid of predefined size, such as 16x9 (for a 16:9 aspect ratio video), where each grid cell corresponds to a specific section of the frame. Each grid cell is represented by its coordinates (i, j) , where i is the horizontal index, and j is the vertical index.

Once the frame is divided into grids, the position (x, y) of each successfully tracked object is used to determine which grid cell the object belongs to at each time step. Specifically, for an object located at coordinates (x, y) , we can determine the grid it belongs to by Equation (6).

$$i = \left\lfloor \frac{x}{\text{grid width}} \right\rfloor, \quad j = \left\lfloor \frac{y}{\text{grid height}} \right\rfloor \quad (6)$$

(3) Grid Mean State computation

For each grid cell (i, j) , we compute the Grid Mean State based on the state vectors of all objects that appear in that grid during the Initial Tracking Phase. The mean state is calculated for the velocity components $[\dot{x}, \dot{y}, \dot{a}, \dot{h}]$. If no objects are tracked in a particular grid, the mean state for that grid is set to zero.



Figure 3. Visualization of Grid Mean State with Velocity Arrows

(4) Usage of Grid Mean States for full video tracking

The calculated Grid Mean State is then used to replace the initial zero velocity assumption in subsequent Full Video Tracking Phase. This ensures that the initial trajectory estimation is more accurate, particularly for small or fast-moving objects. By improving the accuracy of the first few frames of tracking, we allow the Kalman filter's updates to take over and maintain accurate object trajectories in later frames.

3.2 InCo-Skip (Inhomogeneous and Counterintuitive Frame Skip)

In video tracking, reducing computational load is crucial, especially when dealing with real-time processing or large datasets. A common approach is the Homogeneous Skip method, where alternate frames (1, 3, 5, etc.) are processed, skipping every other frame. While this method effectively cuts down computation by 50%, it introduces a major drawback related to initial tracking accuracy. Specifically, with homogeneous frame skipping, the first frame's position must be used to predict the object's position in the third frame. Due to the zero-initial velocity assumption in the Kalman filter, the predicted position in the third frame aligns with the position in the first frame, and the comparison for object matching is made between the first and third frames rather than the first and second frames. This mismatch significantly reduces the success rate of initial tracking, particularly for small or fast-moving objects, which might not be tracked at all.

In response to this issue, InCo-Skip is introduced. Instead of uniformly skipping every other frame, this method processes consecutive pairs of frames (e.g., frames 1 and 2), then skips the next two frames (3 and 4), and continues by processing frames 5 and 6, skipping frames 7 and 8, and so on. This ensures that the tracking system has at least two consecutive frames for comparison which is similar to not skipping any frames at all during initial tracking. Once initial tracking is successful, skipping frame pairs is less detrimental to accuracy because the Kalman filter updates the object state using its learned velocity and direction from the previous frames. Even after skipping frame pairs, the filter can make stable predictions and maintain tracking continuity.

We named this method InCo-Skip, which stands for Inhomogeneous and Counterintuitive Frame Skip. The name reflects both the non-uniform skipping pattern and the method's unconventional

approach of skipping consecutive frames while maintaining tracking stability. The Figure 4 shows the comparison between Homogeneous Skip and InCo-Skip.

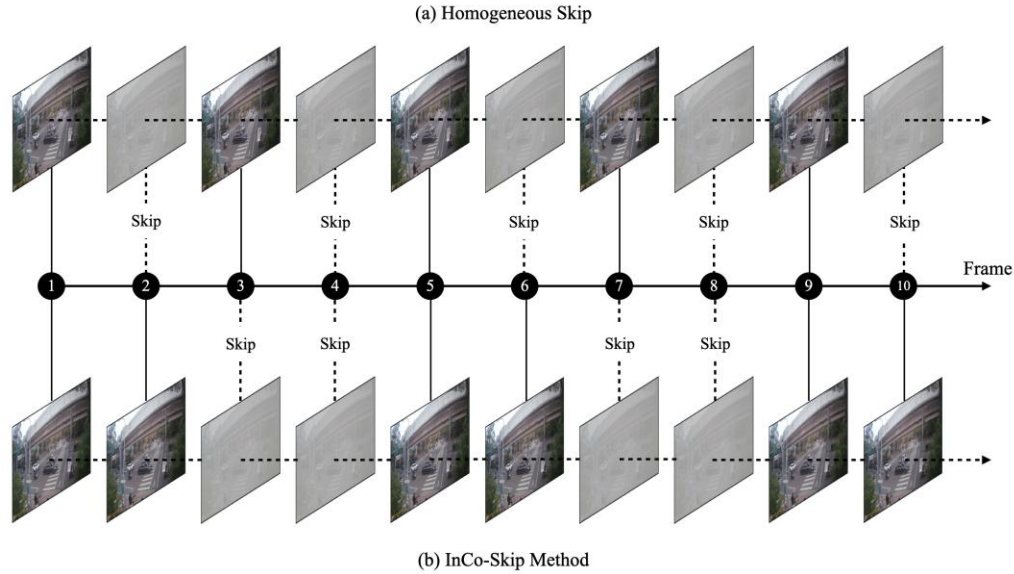


Figure 4. (a) Homogeneous Skip and (b) InCo-Skip Method

4. RESULTS AND DISCUSSIONS

We evaluate the effectiveness of the two methods introduced in Section 3, Grid Mean State and InCo-Skip, in improving tracking accuracy, particularly for motorcycles under Homogenous frame-skipping scenarios. The primary metric we are concerned with is Counting Accuracy, which refers to how accurately the tracking system, based on YOLOv8, can count vehicles compared to manual counts.

Tracking accuracy is also influenced by the IoU threshold, which is a crucial parameter in the tracking process. A well-chosen IoU threshold can enhance tracking performance and improve the overall accuracy of vehicle counts. Therefore, in the upcoming tests, we will compare the effects of different IoU values on tracking accuracy to understand their influence on our evaluation of traffic flow.

This study conducts initial testing on a five-minute video at 30 frames-per-second (fps), capturing traffic flow as shown in Table 1. The accuracy for cars remains consistent even under Homogeneous Skip. However, accuracy of motorcycles drops significantly, likely due to smaller object sizes and faster movement speeds. This observation motivates the need for more robust methods to ensure that frame skipping, which is essential for real-time processing or large datasets, does not lead to substantial accuracy loss. Thus, we test and evaluate the Grid Mean State and InCo-Skip methods to see how they mitigate this issue.

Table 1. Accuracy comparison of car and motorcycle in No Skip and Homogeneous Skip

Vehicle Type	*IoU threshold	Accuracy (%)		Difference
		No Skip	Homogeneous Skip	
Car	0.20 ~ 0.01	98.6	98.6	0
	0.20	98.5	54.9	↓ 43.6
Motorcycle	0.10	98.5	63.4	↓ 35.1
	0.05	98.5	66.2	↓ 32.3
	0.01	98.5	77.5	↓ 21.0

* In YOLOv8, the value of $1 - \text{IoU}$ is used as the match threshold.

4.1 Effect of Grid Mean State on motorcycle accuracy under frame-skipping scenarios

This section compares motorcycle counting accuracy using a zero-initial velocity state versus the Grid Mean State method under frame-skipping scenarios. As shown in Table 4.2, applying the Grid Mean State improves motorcycle accuracy under Homogenous frame-skipping conditions, accuracy increases from 54.9% to 60.6% at $\text{IoU} = 0.20$. However, the effect of the Grid Mean State diminishes at lower IoU thresholds.

Conversely, in the InCo-Skip scenario, the Grid Mean State method shows clear improvements across all IoU thresholds, with an increase in accuracy of up to 18.3% at $\text{IoU} = 0.20$. However, as the IoU value decreases, the increase in accuracy becomes less significant. In summary, we observe that regardless of whether in the Homogeneous Skip or InCo-Skip scenario, the Grid Mean State has a more significant impact at higher IoU values.

Table 2. Motorcycle accuracy using Grid Mean State vs. Zero Initial Velocity

Skip Method	IoU threshold	Accuracy (%)		Difference
		Zero Initial Velocity	Grid Mean State	
Homogenous	0.20	54.9	60.6	↑ 5.7
	0.10	63.4	63.4	0
	0.05	66.2	64.8	↓ 1.4
	0.01	77.5	77.5	0
InCo-Skip	0.20	39.4	57.7	↑ 18.3
	0.10	78.9	87.3	↑ 8.4
	0.05	88.7	94.4	↑ 5.7
	0.01	95.8	95.8	0

4.2 Effect of InCo-Skip vs. Homogeneous Skip on motorcycle accuracy

In this section, this study compares the performance of the InCo-Skip method against Homogeneous Skip in terms of motorcycle counting accuracy, focusing on the impact of different IoU thresholds and the initial velocity assumption (zero velocity or Grid Mean State). As shown in Table 3, the results highlight some key patterns.

The InCo-Skip method underperforms compared to Homogeneous Skip at a higher IoU threshold (0.20), with a 15.5% and 2.9% decrease in accuracy. However, as the IoU threshold decreases, InCo-Skip shows a remarkable improvement, outperforming Homogeneous Skip by up to 29.6% at IoU=0.05. This suggests that InCo-Skip is particularly effective in lower IoU value.

Table 3. Motorcycle accuracy using Homogeneous Skip vs. InCo-Skip

Initial Velocity	IoU threshold	Accuracy (%)		Difference
		Homogeneous Skip	InCo-Skip	
Zero	0.20	54.9	39.4	↓ 15.5
	0.10	63.4	78.9	↑ 15.5
	0.05	66.2	88.7	↑ 22.5
	0.01	77.5	95.8	↑ 18.3
Grid Mean State	0.20	60.6	57.7	↓ 2.9
	0.10	63.4	87.3	↑ 23.9
	0.05	64.8	94.4	↑ 29.6
	0.01	77.5	95.8	↑ 18.3

4.3 Combined performance of Grid Mean State and InCo-Skip methods

In this section, we evaluate the combined effect of using both the Grid Mean State and InCo-Skip methods on motorcycle counting accuracy. As observed in previous sections, the Grid Mean State method shows better accuracy improvements at higher IoU thresholds, while the InCo-Skip method performs better at lower IoU thresholds. Therefore, the combination of these two methods aims to leverage the strengths of both, mitigating each other's limitations.

Table 4 demonstrates that when using both methods, the motorcycle counting accuracy improves across all IoU thresholds compared to using only the Zero Initial Velocity with Homogeneous Skip. At higher IoU thresholds, the improvement is modest, but at lower IoU thresholds, the accuracy gains are more pronounced, showing notable improvements in counting accuracy of up to 28.2%.

Table 4. Motorcycle accuracy with both methods vs. Zero Initial Velocity + Homogeneous Skip

IoU threshold	Accuracy (%)		Difference
	Baseline	Both Methods	
0.20	54.9	57.7	↑ 2.8
0.10	63.4	87.3	↑ 23.9
0.05	66.2	94.4	↑ 28.2
0.01	77.5	95.8	↑ 18.3

To provide a clear recommendation for practical use, our results indicate that setting the IoU threshold between 0.05 and 0.10 yields the most consistent improvement in tracking accuracy when both the Grid Mean State and InCo-Skip methods are applied. By combining these two methods, it effectively balances the strengths of each, reducing the shortcomings observed when using either method in isolation.

5. CONCLUSIONS

In this paper, this study proposed two methods, Grid Mean State and InCo-Skip, to address the challenge of maintaining tracking accuracy for motorcycles in frame-skipping scenarios. Our results showed that while car accuracy remained high, motorcycle accuracy significantly dropped during homogeneous skipping. The Grid Mean State method was more effective at higher IoU thresholds, while InCo-Skip performed better at lower IoU thresholds.

By combining both methods, we achieved improved accuracy across all conditions, highlighting the complementary strengths of each approach. These findings suggest that our methods provide a practical solution for maintaining high tracking accuracy while accommodating frame skipping demands, which is particularly beneficial in real-time applications and when handling large datasets.

6. FUTURE WORK

In this study, the primary focus was on evaluating the effectiveness of the proposed methods based on counting accuracy, using a single test video for performance assessment. While this approach provided initial insights into the methods' potential, there are several areas for further exploration and improvement.

Future work should involve testing on a wider range of publicly available datasets like MOT17 and BDD100K to validate the methods' robustness across diverse scenarios. Additionally, while counting accuracy was the main metric, future research should incorporate other multi-object tracking (MOT) performance metrics, such as ID switch, MOTA (Multiple Object Tracking Accuracy), MOTP (Multiple Object Tracking Precision), and Track Fragmentation. These metrics will provide a deeper understanding of the proposed methods' strengths and limitations, assessing not only accuracy but also stability in more challenging environments.

ACKNOWLEDGMENTS

The authors thank the support from the National Science and Technology Council of Taiwan for grant NSTC 113-2628-E-002-026-MY3.

REFERENCES

- Chen, A. Y., Chiu, Y.-L., Hsieh, M.-H., Lin, P.-W., & Angah, O. (2020). Conflict analytics through the vehicle safety space in mixed traffic flows using UAV image sequences. *Transportation Research Part C: Emerging Technologies*, 119, 102744.
- Kalman, R. E. (1960). *A new approach to linear filtering and prediction problems*.
- Kinoshita, A., Fukuda, T., & Yabuki, N. (2022). Enhanced Tracking Method with Object Detection for Mixed Reality in Outdoor Large Space. *Legal Depot D/2022/14982/02*, 457.
- Li, X., Wang, K., Wang, W., & Li, Y. (2010). A multiple object tracking method using Kalman filter. *The 2010 IEEE International Conference on Information and Automation*, 1862–1866.
- Lin, Z.-H., Chen, A. Y., & Hsieh, S.-H. (2021). Temporal image analytics for abnormal construction activity identification. *Automation in Construction*, 124, 103572.
- Linderoth, M., Soltesz, K., Robertsson, A., & Johansson, R. (2011). Initialization of the Kalman filter without assumptions on the initial state. *2011 IEEE International Conference on Robotics and Automation*, 4992–4997.
- Park, J. W., Kim, J., & Lee, H.-J. (2020). Fast Object Detection Using a Frame Skip Method. *2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 1–2.
- Pathan, S. S., Al-Hamadi, A., & Michaelis, B. (2009). Intelligent feature-guided multi-object tracking using Kalman filter. *2009 2nd International Conference on Computer, Control and Communication*, 1–6.
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 658–666.
- Wang, M., Kumar, S. S., & Cheng, J. C. P. (2021). Automated sewer pipe defect tracking in CCTV videos based on defect detection and metric learning. *Automation in Construction*, 121, 103438.
- Weiner, L. B. (1981). Kalman Filter Initialization With Large Initial Uncertainty And Strong Measurement Nonlinearity. *Conference Proceedings Southeastcon '81.*, 150–151. <https://doi.org/10.1109/SECON.1981.673417>
- Weng, S.-K., Kuo, C.-M., & Tu, S.-K. (2006). Video object tracking using adaptive Kalman filter. *Journal of Visual Communication and Image Representation*, 17(6), 1190–1208.
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). Bytetrack: Multi-object tracking by associating every detection box. *European Conference on Computer Vision*, 1–21.
- Zhao, S., & Huang, B. (2017). On initialization of the Kalman filter. *2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP)*, 565–570.