Confidence Trigger Detection: Accelerating Real-time Tracking-by-detection Systems

Zhicheng Ding*
Fu Foundation School of Engineering and Applied Science,
Columbia University, New York, USA
*Corresponding author: zhicheng.ding@columbia.edu

Siyang Li
Lubin School of Business, Pace University, New York, USA
lisiyang98@hotmail.com

Qikai Yang
Department of Computer Science, University of Illinois Urbana-Champaign, Urbana, USA
qikaiy2@illinois.edu

Zhixin Lai
Department of Computer Science, Cornell University, Ithaca, USA
zl768@cornell.edu

Panfeng Li
Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, USA
pfli@umich.edu

Edward Wong
Tandon School of Engineering, New York University, Brooklyn, USA
ewong@nyu.edu

Abstract—Real-time object tracking necessitates a delicate balance between speed and accuracy, a challenge exacerbated by the computational demands of deep learning methods. In this paper, we introduce Confidence-Triggered Detection (CTD), a novel approach that strategically skips object detection for frames exhibiting high similarity, leveraging tracker confidence scores. CTD not only enhances tracking speed but also preserves accuracy, surpassing existing tracking algorithms. Through extensive evaluation across various tracker confidence thresholds, we identify an optimal trade-off between tracking speed and accuracy, providing crucial insights for parameter fine-tuning and enhancing CTD’s practicality in real-world scenarios. Furthermore, our experiments across diverse detection models underscore the robustness and versatility of the CTD framework, demonstrating its potential to enable real-time tracking in resource-constrained environments.

Keywords—Multiple objects tracking, Object detection, Real-time tracking

I. INTRODUCTION

The field of object detection has witnessed a dramatic shift in recent years, with deep learning architectures achieving unprecedented performance. However, video analysis presents a unique challenge that goes beyond the analysis of isolated frames. Unlike static images, videos possess an inherent temporal dimension, where the relationships between frames hold crucial information. Capturing these temporal dynamics is essential for a comprehensive understanding of the visual content in videos [1-4]. This is particularly important in the context of multiple object tracking (MOT), which has numerous applications in surveillance systems, intelligent robotics, and autonomous vehicles [5,15-16].

Most recent MOT approaches leverage a tracking-by-detection paradigm [6]. This involves three key steps: (1) object detection in each frame, (2) object tracking based on information from previous frames, and (3) data association using location and feature data from both detector and tracker [7]. While algorithms like SORT perform well in terms of precision and accuracy, they often suffer from ID switches and struggle with occlusions. DeepSORT addresses this with a better association metric, but real-world deployment necessitates both high accuracy and real-time performance.

Despite the surge in proposed MOT algorithms, achieving real-time performance remains a significant challenge. The MOTChallenge leaderboard exemplifies this inherent trade-off: faster (higher Hz) algorithms typically exhibit lower Multiple Object Tracking Accuracy (MOTA), while higher MOTA results tend to be associated with slower processing speeds.

Frame skipping is a commonly used technique to improve speed. It involves processing only a subset of video frames, either by skipping a fixed number or waiting for the tracker to be ready for the next assignment. While this reduces computational load and improves speed, it comes at the cost of reduced tracking accuracy. Critical information might be missed, leading to tracking errors such as objects shifting significantly or being entirely lost.

Inspired by frame skipping and other confidence-based work [20-24], this paper presents a novel Confidence-Triggered Detection (CTD) approach. CTD leverages the confidence score associated with object association to determine when to trigger object detection. This allows the system to strategically skip frames where objects exhibit minimal movement, minimizing unnecessary computations. Conversely, a significant discrepancy between the tracker’s predicted location and the detector’s output triggers a new detection. This approach aims to achieve a balance between processing speed and tracking accuracy compared to traditional fixed-interval frame skipping.

In summary, our paper makes the following primary contributions:

- We present CTD, a novel general real-time tracking framework designed to enhance tracking speed while maintaining high accuracy.
Our evaluation of the CTD framework under varying confidence thresholds reveals an optimal tradeoff between tracking speed and accuracy. This assessment provides critical insights for tuning CTD parameters to suit specific tracking contexts, thereby bolstering its practicality and performance in real-world deployments.

We perform extensive experiments to assess the performance of CTD across a range of detection models which demonstrate the robustness and versatility of the CTD framework.

II. RELATED WORK

A. Tracking-by-detection

The dominant paradigm in MOT is tracking-by-detection. This approach leverages two key components: an object detector and a tracker. In each frame, the detector identifies and localizes objects [17-19]. The tracker then predicts the current location of previously detected objects based on their past states. Finally, an association step links detected objects from the current frame with corresponding tracks using information from both the detector and tracker. This approach offers improved accuracy compared to alternatives, as the tracker's predictions can be continuously refined by up-to-date detection.

B. Real-time Object Tracking

Defining a universal speed threshold for real-time performance in MOT is challenging due to variations in video frame rates captured by different cameras [25-28]. However, a system is generally considered real-time if its processing speed surpasses the input video frame rate. If the processing speed falls behind, a delay accumulates, causing the system output to become out of sync with the actual video and events.

Frame skipping is a common way for real-time object tracking systems where the system processes only a subset of video frames [29]. One approach involves skipping a fixed number of frames after each tracking assignment. However, this method can lead to missing critical frames, especially for fast-moving objects, resulting in significant accuracy drops [30-31].

Our proposed CTD approach also incorporates frame skipping. However, unlike fixed-interval skipping, CTD leverages a confidence score to selectively trigger new detection. This allows the system to strategically skip frames with minimal object movement while maintaining tracking accuracy. We continuously run the tracker in all frames to monitor the discrepancy between the tracker's predictions and the latest detection. This difference serves as a measure of confidence, and a significant discrepancy triggers a new detection to correct potential biases before they accumulate significantly. This approach aims to achieve a balance between processing speed and tracking accuracy compared to traditional fixed-interval frame skipping.

C. DeepSORT

DeepSORT [3] is an extension of the SORT [1] algorithm that incorporates appearance information for object matching. This enables Deep SORT to track objects even during extended occlusions. Additionally, it utilizes the Mahalanobis distance metric to incorporate motion information into the association process.

We leverage DeepSORT for two main reasons in our approach. Firstly, DeepSORT can track objects during occlusions ensuring that objects can still be matched after skipping a certain number of frames. Secondly, we utilize the Mahalanobis distance calculated by Deep SORT to infer low confidence scores and trigger new detections when necessary.

III. METHODS

This section details the methodology employed in our object-tracking system. We first provide a high-level overview of the tracking-by-detection framework utilized. Then, we explain how the Confidence-Triggered Detection (CTD) approach, a core component responsible for strategically skipping object detection for enhanced processing speed in greater detail. Finally, we explain how the system infers the confidence score using Mahalanobis distance.

A. Tracking-by-detection Framework

Our system leverages a tracking-by-detection architecture, as illustrated in Figure 1. This architecture shares similarities with DeepSORT and integrates three key components: Kalman Filter, Hungarian Assignment, and the CTD module.

![Fig. 1: illustrates the Confidence-Triggered Detection (CTD) framework.](image)

The confidence score is derived by combining the detection results from the previous frame with the current frame's object location, as predicted by the Kalman Filter. When the confidence score falls below a predefined threshold (Confidence Score Threshold) or the number of consecutively skipped frames exceeds a predetermined threshold (Maximum Skip Frame Threshold), a new object detection will be triggered and the previous detection result will be replaced. Otherwise, object detection is skipped for this frame, and thus improves processing speed.
Kalman Filter: When a new frame arrives, the Kalman Filter predicts the object's location in the current frame based on its detection information from the previous frame. This prediction helps maintain object tracking during periods when detection is skipped.

Hungarian Assignment: This algorithm plays a crucial role in object association and ID attribution. It essentially determines whether an object detected in the current frame corresponds to the same object tracked in the previous frame.

Confidence-Triggered Detection (CTD): This innovative method integrates the Mahalanobis Distance and a derived Confidence Score with Detector Trigger Modules for object detection. The Confidence Score, deduced from the Mahalanobis Distance using a chi-squared distribution, dictates whether to maintain tracking or invoke the detector for new object detection, as elaborated in Section III-C. Specifically, when the confidence score falls below a predefined threshold or the number of consecutively skipped frames exceeds a predetermined threshold, new object detection will be triggered and the previous detection result will be replaced. Otherwise, object detection is skipped for this frame. The intricacies of the CTD approach, including detector operations and decision mechanisms, are further expounded in Section III-B.

B. Confidence-Triggered Detection (CTD) Approach

The CTD approach aims to achieve a balance between processing speed and tracking accuracy by strategically skipping object detection in certain frames. Here's how it operates:

Initialization: The system starts with a counter set to a value exceeding the maximum frame-skipping threshold. Additionally, the confidence score is initialized to 0%. The initial settings ensure object detection occurs in the first frame and establish initial object locations.

Confidence Score Evaluation: The Mahalanobis Distance is calculated between two bounding boxes. The predicted bounding of the current frame by the Kalman filter and bounding boxes of the last frame are detected by the detector. By applying a chi-squared distribution to the Mahalanobis Distance, as outlined in Section III-C, we obtain a confidence score, which will be used to determine if triggering a new round of detection within the Detection Trigger Module.

Detection Trigger Module: If the confidence score surpasses a predefined threshold and the frame-skipping counter hasn't reached its limit, the system skips detection and relies solely on the Kalman filter's prediction for the current frame (detection is skipped). This approach improves processing speed. Conversely, if the confidence score falls below the threshold or the frame-skipping counter reaches its limit, a new detection is triggered in the current frame to potentially correct the discrepancies between the predicted and actual object location.

Data Association: The Hungarian algorithm is employed for the data association task, effectively matching the detection results from the current frame with the predicted object locations provided by the Kalman filter. This process ascertains the continuity of object identity between consecutive frames, ensuring that each detected object is accurately aligned with its corresponding track.

By strategically skipping detection based on confidence scores, the CTD approach achieves a desirable balance between processing speed and tracking accuracy.

C. Confidence Score Inference

The Mahalanobis Distance is a metric used to assess the similarity between a data point and a distribution defined by its mean and covariance. In the context of multi-object tracking, it can be employed to measure the discrepancy between a predicted bounding box and the detected objects in a video frame, defined as:

\[
M(i, j)^2 = (x_j - \bar{y}_j)(S_i)^{-1}(x_j - \bar{y}_j)
\]

where \(i\) denotes the information at the \(i\)-th frame and \(j\) denotes the information at \(j\)-th frame. The \(x_j\) and \(S_i\) represent the mean and covariance of detected bounding boxes. The Mahalanobis distance follows a chi-square distribution [8]. Given probability level \(p\) and degrees of freedom \(v\), we can calculate the distance threshold corresponding to that probability using the inverse cumulative distribution function (CDF) [9]:

\[
d = F^{-1}(p|v)
\]

where \(p\) can be calculated using CDF, shown as equation (3).

\[
d = F(x|v) = \int_0^x \frac{1}{2\sqrt{\pi}v/2} \exp\left(-\frac{x^2}{2v^2}\right) \, dx
\]

where \(\Gamma(\cdot)\) represents the Gamma function. In our paper, we set the degrees of freedom \(v\) to 4. We evaluate the model's performance using different probability values \(q\).

IV. EXPERIMENT AND RESULTS

This section assesses the efficacy of the CTD approach in improving real-time object-tracking performance. We detail our experimental setup, encompassing datasets, models, and frame-by-frame analysis. Then we conduct comparative assessments against existing techniques and demonstrate the superior performance of the CTD. Furthermore, we analyze the accuracy-speed trade-off across different confidence score thresholds, which yields valuable insights into optimizing CTD's parameters for different tracking scenarios. Ultimately, we perform extensive experiments to assess the performance of CTD across diverse detection models which demonstrates the robustness and versatility of the CTD framework.

A. System Setup

We developed a tracking-by-detection system following the framework outlined in Section III-A. DeepSORT was chosen as the tracker due to its real-time capabilities, occlusion handling, and competitive MOTA scores. YOLOv3 Tiny [10] served as the object detector due to its balance between speed and accuracy. DeepSORT integrated with YOLOv3 Tiny achieved the best processing speeds compared to the existing algorithms.

We employed 2D video clips sourced from the MOTChallenge dataset [11,32-36]. This dataset offers a comprehensive range of challenges, featuring diverse target
motions, camera perspectives, and pedestrian densities, rendering it ideal for rigorous evaluation.

C. Comparison with Existing Methods

Following an exhaustive exploration of multiple confidence score thresholds, we determined that establishing the threshold at 30% achieves a commendable equilibrium between precision and computational efficiency. Subsequently, we conducted comparative analyses between our proposed method CTD and existing non-skippable and skippable tracking approaches. The results are summarized in Table 1. As a non-skippable approach, TraByDetNs operates object detection for each frame, contrasting with skippable methods such as TSDA_OAL [12], TC_SIAMESE [13], and GMPHD [14], which perform detection intermittently across frames. Our proposed method, CTD, outperforms all evaluated methods in terms of processing speed and achieves superior MOTA (Multi-Object Tracking Accuracy), and FP (False Positive) metrics compared to other skippable methods.

D. Trade-Off Between Accuracy and Speed

Although the CTD method enhances overall processing speed by judiciously skipping object detection. However, this optimization strategy may potentially compromise tracking accuracy (MOTA). To thoroughly assess the performance of CTD, we conducted evaluations using various confidence thresholds. This allowed us to delve into the nuanced trade-off between speed (FPS) and accuracy (MOTA) across diverse scenarios. Specifically, we varied the maximum skip frame threshold from 1 to 10 and explored confidence thresholds ranging from 100% down to 0%.

Notably, we established two extreme cases for threshold settings to provide additional context for our evaluation. A 100% threshold mandates detection in each frame, representing the baseline condition. On the other hand, the 0% threshold mirrors fixed frame skipping, whereby detection is solely triggered based on reaching the maximum skip frame threshold, disregarding confidence scores.

In Figures 3(a) and 3(b), we present the variations in processing speed (FPS) and accuracy (MOTA) under varying...
confidence thresholds. As the confidence threshold is elevated, the incremental speed advantage is reduced, and correspondingly, the detriment to accuracy becomes less pronounced. This observation suggests a trade-off between speed and accuracy that is modulated by the confidence threshold, with higher thresholds diminishing returns in speed but offering a buffering effect against accuracy degradation.

Figure 3(c) illustrates the direct relationship between speed (FPS) and accuracy (MOTA). It highlights that CTD consistently achieves superior accuracy levels at the same speed compared to the fixed frame skipping strategy (0% confidence threshold), underscoring the efficacy of CTD. Furthermore, it offers valuable guidance for selecting an appropriate confidence threshold. By referring to Figure 3(c), one can determine the threshold associated with the highest speed while maintaining the desired level of accuracy.

The analysis demonstrates superior accuracy levels compared to the fixed frame skipping strategy. Furthermore, it yields valuable insights into optimizing CTD's parameters for different tracking scenarios, enhancing its adaptability and effectiveness in real-world applications.

E. A Comparison between CTD with Different Trackers

The CTD framework's adaptability was assessed across various detection models, such as YOLOv3, MobileNet, and SqueezeNet. According to the data in Table 2, a consistent speed-accuracy balance is maintained across all models when incorporated with CTD. Notably, a mere 10% reduction in the confidence score threshold to 90% achieves a significant speed enhancement of about 30%, with only a negligible 3% compromise in accuracy. When the confidence score threshold is further relaxed to 10%, there is an impressive increase in speed by approximately 145%, with an acceptable decrease in accuracy of 27%. These outcomes highlight two principal advantages of the CTD framework: its plug-and-play compatibility with different platforms and models, and its ability to deliver substantial improvements in processing speed while sustaining satisfactory accuracy levels.

Table 2: illustrates the tracking outcomes for frames 55 to 61 utilizing the CTD approach. The analysis is illustrated with cropped image frames, where white bounding boxes denote predictions made by the Kalman filter, and blue bounding boxes indicate detection

| Confidence Threshold | YOLOv3 Tiny | MobileNet | SqueezeNet |
|----------------------|------------|-----------|------------|
| Speed Gain | Accua | Rcy Lose | Speed Gain | Accua | Rcy Lose | Speed Gain | Accua | Rcy Lose |
| 100% (baseline) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 90% | 29.2% | -2.1% | 29.4% | -3.0% | 26.6% | -2.8% |
| 10% | 148.1% | -26.2% | 145.5% | -28.1% | 141.3% | -27.8% |

V. CONCLUSIONS

This work presents the CTD approach, a novel technique that significantly improves tracking speed while maintaining accuracy in tracking-by-detection systems. The CTD leverages the tracker's confidence score to strategically activate object detection. This paper offers three key contributions:

- Enhanced Speed: CTD demonstrably increases tracking speed compared to conventional methods by selectively skipping detection in frames with low confidence scores.
- Optimal Accuracy and Speed Trade-Off: CTD achieves an optimal tradeoff between accuracy and speed through comprehensive experimentation. The experiments also provide valuable guidance for optimizing CTD in real-world deployments.
- Adaptable Framework: CTD has demonstrated great robustness and versatility through extensive experiments across a range of detection models.

By achieving a significant speedup with minimal accuracy loss, the adaptable framework CTD paves the way for real-time object tracking in resource-constrained environments.

REFERENCES

[1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” CoRR, vol. abs/1602.00763, 2016.

[2] X. Chen, A. T. Z. Kasgari and W. Saad, “Deep Learning for Content-Based Personalized Viewport Prediction of 360-Degree VR Videos,” in IEEE Networking Letters, vol. 2, no. 2, pp. 81-84, June 2020
[1] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” Proceedings - International Conference on Image Processing, ICIP, vol. 2017-Septe, pp. 3645–3649, 2018.

[2] X. Chen, K. Li, T. Song, and J. Guo “Few-Shot Name Entity Recognition on Stackoverflow.” arXiv.org, 15 Apr. 2024, arxiv.org/abs/2404.09405.

[3] C. H. Huang, E. Boyer, B. D. C. Angonese, N. Navab, and S. Ilic, “Toward user-specific tracking by detection of human shapes in multi-camera systems,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015.

[4] I. Bouchrika, J. N. Carter, and M. S. Nixon, “Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras,” Multimedia Tools and Applications, 2016.

[5] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 3645–3649.

[6] G. Chang, “Robust Kalman filtering based on Mahalanobis distance as outlier discriminator,” Journal of Geodesy, 2014.

[7] K. Pearson, On the Criterion that a Given System of Deviations from the Previsited规的, Case Table of Correlated System of Variates is Such That Can Be Reasonably Supposed to have Arisen from Random Sampling. New York, NY: Springer New York, 1992, pp. 11–28.

[8] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” New York, NY: Springer New York, 1992, pp. 11–28.

[9] D. Han, J. Ju, H. ko, D. Kim, and B. Ku, “Online multi-person tracking with two-stage data association and online appearance model learning,” IET Computer Vision, vol. 11, 07 2016.

[10] Y. Young Chul, Y.-M. Song, K. Yoon, and M. Jeon, “Online multi-object tracking using selective deep appearance matching,” 06 2018, pp 212.

[11] Y.-M. Song, K. Yoon, Y.-C. Yoon, K. C. Yow, and M. Jeon, “Online multi-object tracking with gmphd filter and occlusion group management,” IEEE Access, vol. 7, pp. 165 103–165 121, 2019.

[12] P. Li, Y. Lin, and E. Schultz-Fellenz, “Contextual hourglass network for semantic segmentation of high resolution aerial imagery,” CoRR, 2019.

[13] P. Li, M. Abouelenien, and R. Mihalea, “Deception detection from linguistic and physiological data streams using bimodal convolutional neural networks,” CoRR, 2023.

[14] A. Shree, K. Jia, Z. Xiong, S. F. Chow, R. Phan, P. Li, and D. Curro, “Image analysis,” US Patent App. 17/638,773, 2022.

[15] W. Castillo, B. Scott, A. Firl, D. R. Cutts, J. M. Igner, D. Rethage, D. Curro, and P. Li, “Techniques for enhanced image capture using a computer-visual network,” US Patent App. 17/163,043, 2021.

[16] Z. Li, P. Zhou, Y. Zhang, and L. Gao, “Joint coverage and power control in highly dynamic and massive uav networks: An aggregative game-theoretic learning approach,” ArXiv, vol. abs/1907.08363, 2019.

[17] X. Liu, R. Wang, D. Sun, J. Li, C. Youn, Y. Lyu, J. Zhan, D. Wu, X. Xu, M. Liu, X. Lei, Z. Xu, Y. Zhang, Z. Li, Q. Yang, and T. Abdelzaher, “Influence Pathway Discovery on Social Media” IEEE 9th International Conference on Collaboration and Internet Computing (CIC), pp. 105-109, 2023.

[18] W. He, A. M. Sainju, Z. Jiang, D. Yan, and Y. Zhou, “Earth imagery segmentation on terrain surface with limited training labels: A semi-supervised approach based on physics-guided graph co-training,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 13, no. 2, pp. 1–22, 2022.

[19] Y. Li and L. Liu, “Enhancing diffusion-based point cloud generation with smoothness constraint,” arXiv preprint arXiv:2404.02596, 2024.

[20] W. He, Z. Jiang, M. Kirby, Y. Xie, X. Jia, D. Yan, and Y. Zhou, “Quantifying and reducing registration uncertainty of spatial vector labels on earth imagery,” in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 554–564.

[21] W. He, Z. Jiang, T. Xiao, Z. Xu, S. Chen, R. Fick, M. Medina, and C. Andreatini, “A hierarchical spatial transformer for massive point samplings in continuous space,” Advances in Neural Information Processing Systems, vol. 36, 2024.

[22] Y. Ji et al., “Prediction of COVID-19 Patients’ Emergency Room Revisit using Multi-Source Transfer Learning,” 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), Houston, TX, USA, 2023, pp. 138-144, doi: 10.1109/ICHI57859.2023.00028.

[23] Z. Luo, Y. Ji, A. Gupta, Z. Li, A. Frisch and D. He, “Towards Accurate and Clinically Meaningful Summarization of Electronic Health Record Notes: A Guided Approach,” 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Pittsburgh, PA, USA, 2023, pp. 1-5, doi: 10.1109/BHI58755.2023.1031411.

[24] A. Zaman, Z. Huang, W. Li, H. Qin, D. Kang, and X. Liu, “Artificial intelligence-aided grade crossing safety violation detection methodology and a case study in new jersey,” Transportation Research Record: Journal of the Transportation Research Board, vol. 2677, no. 10, p. 688–706, May 2023.

[25] W. Li, J. Wang, X. Qin, G. Jing, and X. Liu, “Artificial intelligence-aided detection of rail defects based on ultrasonic imaging data,” Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 238, no. 1, p. 118–127, Nov 2023.

[26] D. Yang, Y. Bai, X. Jia, Y. Liu, X. Cao, and W. Yu, “Cheating suﬃx: Targeted attack to text-to-image diﬀusion models with multi-modal priors,” arXiv preprint arXiv:2402.01369, 2024.

[27] J. Li, K. Gao, Y. Bai, J. Zhang, S-t. Xia, and Y. Wang, “Fmm-attack: A flow-based multi-modal adversarial attack on video-based Ims,” arXiv preprint arXiv:2403.13507, 2024.

[28] Y. Wang, Z. Wu, C. Li, and A. Wu, “Extraction and recovery of spatio-temporal structure in latent dynamics alignment with diﬀusion model,” Advances in Neural Information Processing Systems, vol. 36, 2024.

[29] Y. Wang, Q. Luo, Y. Ding, D. Wang, and H. Deng, “Dem-net: dependency-aware multi-interest network with self-supervised graph learning for click-through rate prediction,” arXiv preprint arXiv:2109.12512, 2021.

[30] H. Wang, Z. Tang, S. Zhang, C. Shen, and T.-H. Chang, “Embracing uncertainty: A diﬀusion generative model of spectrum eﬃciency in 5g networks,” in 2023 International Conference on Wireless Communications and Signal Processing (WCSP), IEEE, 2023, pp. 880–885.

[31] Z. Tang, D. Rybin, and T.-H. Chang, “Zeroth-order optimization meets human feedback: Provably learning via ranking oracles,” in The Twelfth International Conference on Learning Representations, 2024.

[32] Z. Chen, L. Jing, L. Yang, Y. Li, and B. Li, “Class-level confidence based 3d semi-supervised learning,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2025, pp. 633–642.

[33] Z. Chen, L. Jing, Y. Liang, Y. Tian, and B. Li, “Multimodal semi-supervised learning for 3d objects,” arXiv preprint arXiv:2110.11601, 2021.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.