Novelty Assessment Report

Paper: PointRePar : SpatioTemporal Point Relation Parsing for Robust Category-Unified 3D Tracking
PDF URL: https://openreview.net/pdf?id=DLcnrY5Uqo
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

3D single object tracking (SOT) remains a highly challenging task due to the inherent crux in learning representations from point clouds to effectively capture both spatial shape features and temporal motion features. Most existing methods employ a category-specific optimization paradigm, training the tracking model individually for each object category to enhance tracking performance, albeit at the expense of generalizability across different categories. In this work, we propose a robust category-unified 3D SOT model, referred to as SpatioTemporal Point Relation Parsing model (PointRePar), which is capable of joint training across multiple categories while excelling in unified feature learning for both spatial shapes and temporal motions. Specifically, the proposed PointRePar captures and parses the latent point relations across both spatial and temporal domains to learn superior shape and motion characteristics for robust tracking. On the one hand, it models the multi-scale spatial point relations using a Mamba-based U-Net architecture with adaptive point-wise feature refinement. On the other hand, it captures both the point-level and box-level temporal relations to exploit the latent motion features. Extensive experiments across three benchmarks demonstrate that our PointRePar not only outperforms the existing category-unified 3D SOT methods significantly, but also compares favorably against the state-of-the-art category-specific methods. Codes will be released.

Core Task Landscape

This paper addresses: **Category-Unified 3D Single Object Tracking from Point Clouds**

A total of **50 papers** were analyzed and organized into a taxonomy with **20 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Siamese Network-Based Tracking Paradigms**
- **Motion-Centric Tracking Paradigms**
- **Category-Unified Tracking**
- **Temporal Context and Memory Mechanisms**
- **Feature Representation and Enhancement**
- **Class-Agnostic and Open-Vocabulary Tracking**
- **Annotation-Efficient and Weakly-Supervised Tracking**
- **Multi-Object Tracking in Point Clouds**
- **Motion Detection and Prediction in Dynamic Scenes**
- **Instance Segmentation and Scene Understanding**
- ... and 2 more categories

Complete Taxonomy Tree

- Category-Unified 3D Single Object Tracking from Point Clouds Survey Taxonomy
- Siamese Network-Based Tracking Paradigms
  - Transformer-Enhanced Siamese Trackers (6 papers)
  - [1] 3d siamese transformer network for single object tracking on point clouds (Hui, 2022) View paper
  - [2] PTT: Point-track-transformer module for 3D single object tracking in point clouds (Jiayao Shan, 2021) View paper
  - [4] Real-time 3D single object tracking with transformer (Jiayao Shan, 2022) View paper
  - [8] Revisiting Siamese-Based 3D Single Object Tracking With a Versatile Transformer (Jiaming Liu, 2025) View paper
  - [31] Feature-concatenated transformer for 3D object tracking in point clouds (Qi Shen, 2023) View paper
  - [44] GLT-T: Global-Local Transformer Voting for 3D Single Object Tracking in Point Clouds (Nie, 2023) View paper
  - Alternative Siamese Architectures (5 papers)
  - [5] OST: Efficient one-stream network for 3D single object tracking in point clouds (Xiantong Zhao, 2024) View paper
  - [12] 3d-siamrpn: An end-to-end learning method for real-time 3d single object tracking using raw point cloud (Z. Fang, 2020) View paper
  - [15] PillarTrack: Redesigning Pillar-based Transformer Network for Single Object Tracking on Point Clouds (Xu Weisheng, 2024) View paper
  - [50] SPAN: siampillars attention network for 3D object tracking in point clouds (Yi Zhuang, 2022) View paper
- Motion-Centric Tracking Paradigms
  - Matching-Free Motion Prediction (3 papers)
  - [3] A lightweight and detector-free 3d single object tracker on point clouds (Yan Xia, 2023) View paper
  - [6] Beyond 3d siamese tracking: A motion-centric paradigm for 3d single object tracking in point clouds (Chaoda Zheng, 2022) View paper
  - [47] An Effective Motion-Centric Paradigm for 3D Single Object Tracking in Point Clouds (Chaoda Zheng, 2023) View paper
  - Hybrid Motion-Appearance Approaches (3 papers)

- [9] TM2B: Transformer-Based Motion-to-Box Network for 3D Single Object Tracking on Point Clouds (Anqi Xu, 2024) View paper
- [32] Mbptrack: Improving 3d point cloud tracking with memory networks and box priors (Tian-Xing Xu, 2023) View paper
- [45] Enhancing 3D Single Object Tracking with Efficient Point Cloud Segmentation (Yushi Yang, 2024) View paper

- Category-Unified Tracking ★ (3 papers)
  - [0] PointRePar : SpatioTemporal Point Relation Parsing for Robust Category-Unified 3D Tracking (Anon et al., 2026) View paper
  - [16] Towards category unification of 3D single object tracking on point clouds (Nie, 2024) View paper
  - [23] TrackAny3D: Transferring Pretrained 3D Models for Category-unified 3D Point Cloud Tracking (Wang Mengmeng, 2025) View paper
- Temporal Context and Memory Mechanisms (4 papers)
  - [18] Exploit Spatiotemporal Contextual Information for 3D Single Object Tracking via Memory Networks (Mengmeng Wang, 2024) View paper
  - [19] 3d single-object tracking in point clouds with high temporal variation (Wu Qiao, 2024) View paper
  - [48] Modeling continuous motion for 3d point cloud object tracking (Luo, 2024) View paper
  - [49] M3SOT: Multi-frame, Multi-field, Multi-space 3D Single Object Tracking (Liu, 2023) View paper
- Feature Representation and Enhancement
  - Point Cloud Completion and Densification (2 papers)
  - [36] Implicit and efficient point cloud completion for 3D single object tracking (Pan Wang, 2023) View paper
  - [41] Exploiting more information in sparse point cloud for 3D single object tracking (Yu-Bo Cui, 2022) View paper
  - Positional and Structural Feature Learning (2 papers)
  - [20] Structure aware 3D single object tracking of point cloud (Xiaoyu Zhou, 2021) View paper
  - [27] PPE: Point position embedding for single object tracking in point clouds (Yuanzhi Su, 2023) View paper
  - Multi-Scale and Hierarchical Approaches (3 papers)
  - [35] Learning Adaptive Conceptual Prototypes for 3D Single Object Tracking (Jie Xiao, 2025) View paper
  - [40] CDTracker: Coarse-to-Fine Feature Matching and Point Densification for 3D Single-Object Tracking (Yu-An Zhang, 2024) View paper
  - [42] Correlation Pyramid Network for 3D Single Object Tracking (Mengmeng Wang, 2023) View paper
- Class-Agnostic and Open-Vocabulary Tracking (2 papers)
  - [7] Toward class-agnostic tracking using feature decorrelation in point clouds (Shengjing Tian, 2024) View paper
  - [37] Open3dtrack: Towards open-vocabulary 3d multi-object tracking (Ayesha Ishaq, 2025) View paper
- Annotation-Efficient and Weakly-Supervised Tracking (3 papers)
  - [21] MixCycle: Mixup Assisted Semi-Supervised 3D Single Object Tracking with Cycle Consistency (Qiao Wu, 2023) View paper
  - [24] Self-supervised class-agnostic motion prediction with spatial and temporal consistency regularizations (KeWei Wang, 2024) View paper
  - [28] Weakly supervised class-agnostic motion prediction for autonomous driving (Ruibo Li, 2023) View paper
- Multi-Object Tracking in Point Clouds
  - Detection-Based Multi-Object Tracking (2 papers)
  - [14] Center-based 3d object detection and tracking (Tianwei Yin, 2021) View paper
  - [34] 3D multi-object tracking in point clouds based on prediction confidence-guided data association (Hai Wu, 2021) View paper
  - Joint Detection and Tracking (2 papers)
  - [13] Pointtracknet: An end-to-end network for 3-d object detection and tracking from point clouds (Sukai Wang, 2020) View paper
  - [33] Exploring simple 3d multi-object tracking for autonomous driving (Chenxu Luo, 2021) View paper
  - Category-Level Multi-Object State Tracking (1 papers)
  - [26] Category-Level Multi-Object 9D State Tracking Using Object-Centric Multi-Scale Transformer in Point Cloud Stream (Jingtao Sun, 2024) View paper
- Motion Detection and Prediction in Dynamic Scenes (3 papers)
  - [10] Ratrack: moving object detection and tracking with 4d radar point cloud (Zhijun Pan, 2024) View paper
  - [22] Moving object detection and tracking with 4d radar point cloud (Z Pan, 2023) View paper
  - [39] Any motion detector: Learning class-agnostic scene dynamics from a sequence of lidar point clouds (Filatov, 2020) View paper
- Instance Segmentation and Scene Understanding (2 papers)
  - [11] Any3DIS: Class-Agnostic 3D Instance Segmentation by 2D Mask Tracking (Phuc Nguyen, 2025) View paper
  - [30] Unsupervised class-agnostic instance segmentation of 3d lidar data for autonomous vehicles (Lucas Nunes, 2022) View paper
- Specialized Detection and Tracking Applications
  - Waterborne and Specialized Environment Tracking (1 papers)
  - [38] Obstacle tracking for unmanned surface vessels using 3-D point cloud (Jon MuhoviÄ, 2019) View paper
  - Multi-Modal Fusion Tracking (2 papers)
  - [29] Complexer-yolo: Real-time 3d object detection and tracking on semantic point clouds (MartÃn SimÃ³n, 2019) View paper
  - [46] 3D object tracking using RGB and LIDAR data (Alireza Asvadi, 2016) View paper
- Foundational Detection and Localization Methods
  - Density-Based and Clustering Detection (1 papers)
  - [25] Density-based clustering for 3d object detection in point clouds (Syeda Mariam Ahmed, 2020) View paper
  - IoU-Based Localization (1 papers)
  - [17] DeepPCT: Single object tracking in dynamic point cloud sequences (Hao Liu, 2022) View paper

Narrative

Core task: category-unified 3D single object tracking from point clouds. The field has evolved from early category-specific methods toward more general frameworks that can handle diverse object types within a unified architecture. The taxonomy reveals several major branches: Siamese Network-Based Tracking Paradigms, which leverage template-matching strategies inherited from 2D tracking (e.g., 3D SiamRPN[12], Siamese Transformer Tracking[1]); Motion-Centric Tracking Paradigms, which emphasize motion patterns and temporal dynamics (Motion Centric Paradigm[6], Motion to Box[9]); and Category-Unified Tracking, which aims to eliminate per-class specialization. Additional branches address temporal context and memory mechanisms for long-term consistency, feature representation enhancements to handle sparse and irregular point clouds, class-agnostic and open-vocabulary approaches that generalize beyond training categories, and annotation-efficient methods that reduce supervision requirements. Parallel branches cover multi-object tracking, motion detection in dynamic scenes, instance segmentation, and foundational detection methods that underpin many tracking systems.

Recent work has increasingly focused on unifying tracking across object categories and reducing reliance on category-specific tuning. A handful of studies explore how to build trackers that generalize to arbitrary object types, balancing the need for discriminative features with computational efficiency. PointRePar[0] sits squarely within the Category-Unified Tracking branch, alongside works like Category Unification[16] and TrackAny3D[23], which similarly pursue category-agnostic designs. Compared to these neighbors, PointRePar[0] emphasizes reparameterization techniques to achieve unified representations without sacrificing per-category performance, contrasting with approaches that rely heavily on large-scale pretraining or open-vocabulary embeddings. This line of work addresses a key trade-off: how to maintain strong discriminative power across diverse object shapes and sizes while avoiding the overhead of category-specific modules, a challenge that remains central to advancing practical 3D tracking systems.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Towards category unification of 3D single object tracking on point clouds

Authors: Nie, Jiahao, He Zhiwei, Jiahao Nie, Lv, et al. (15 authors total) | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Category-specific models are provenly valuable methods in 3D single object tracking (SOT) regardless of Siamese or motion-centric paradigms. However, such over-specialized model designs incur redundant parameters, thus limiting the broader applicability of 3D SOT task. This paper first introduces unified models that can simultaneously track objects across all categories using a single network with shared model parameters. Specifically, we propose to explicitly encode distinct attributes associat...

Relationship Analysis

Both papers belong to the Category-Unified Tracking category, focusing on training single models to track objects across multiple categories using shared parameters. They overlap in addressing the core challenge of unified feature learning across diverse object categories in 3D point cloud tracking. The key difference is that the original paper (PointRePar) emphasizes spatiotemporal point relation parsing through Mamba-based U-Net architecture with dynamic feature aggregation and long-term temporal modeling, while the candidate paper (CUTrack) focuses on deformable group vector-attention mechanisms (AdaFormer) to adaptively encode varying size and shape information, along with unified model inputs and learning objectives.

---

2. TrackAny3D: Transferring Pretrained 3D Models for Category-unified 3D Point Cloud Tracking

Authors: Wang Mengmeng, Wang Haonan, Mengmeng Wang, Li Yulong, Haonan Wang, et al. (14 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

3D LiDAR-based single object tracking (SOT) relies on sparse and irregular point clouds, posing challenges from geometric variations in scale, motion patterns, and structural complexity across object categories. Current category-specific approaches achieve good accuracy but are impractical for real-world use, requiring separate models for each category and showing limited generalization. To tackle these issues, we propose TrackAny3D, the first framework to transfer large-scale pretrained 3D mode...

Relationship Analysis

Both papers belong to the Category-Unified Tracking category, training single models across multiple object categories for 3D point cloud tracking. They share the goal of addressing category-unified tracking through shared parameters, with PointRePar focusing on spatiotemporal point relation parsing using Mamba-based architectures and dynamic feature aggregation, while TrackAny3D takes a transfer learning approach by adapting large-scale pretrained 3D models (RECON) with parameter-efficient fine-tuning techniques and mixture-of-geometry-experts. The key distinction lies in their architectural foundations: PointRePar builds a custom Mamba-based U-Net for multi-scale spatial relation parsing, whereas TrackAny3D leverages frozen pretrained Transformers with lightweight adapters for knowledge transfer.

Contributions Analysis

Overall novelty summary. The paper proposes PointRePar, a category-unified 3D single object tracking framework that jointly trains across multiple object categories while learning unified spatial and temporal features. Within the taxonomy, it resides in the Category-Unified Tracking leaf, which contains only three papers total. This is a relatively sparse research direction compared to the more crowded Siamese Network-Based and Motion-Centric branches, suggesting that category-unified approaches remain an emerging area. The sibling papers in this leaf (Category Unification and TrackAny3D) similarly pursue unified tracking but differ in their technical strategies for achieving generalization.

The taxonomy reveals that PointRePar sits at the intersection of multiple research threads. Its spatial relation parsing connects to Feature Representation and Enhancement branches (particularly Multi-Scale and Hierarchical Approaches), while its temporal modeling relates to Temporal Context and Memory Mechanisms. The Siamese and Motion-Centric paradigms, which dominate the field with over 15 papers combined, represent alternative tracking philosophies that PointRePar aims to unify. The taxonomy's scope notes clarify that category-unified methods explicitly encode category information, distinguishing them from class-agnostic approaches that track arbitrary objects without category-specific parameters.

Among 30 candidates examined, the first contribution (category-unified framework with spatiotemporal parsing) shows one refutable candidate among 10 examined, indicating some prior work overlap in the unified tracking space. The second contribution (U-shaped Mamba architecture with dynamic aggregation) examined 10 candidates with none clearly refuting it, suggesting relative novelty in the specific architectural design. The third contribution (long-term temporal parsing with Gaussian perturbation) similarly found no clear refutations among 10 candidates. The limited search scope means these statistics reflect top-30 semantic matches rather than exhaustive coverage of the field.

Based on the limited literature search, the architectural components appear more novel than the high-level category-unified framing, which has established precedents in the sparse Category-Unified Tracking leaf. The analysis covers top-30 semantic matches and does not extend to broader architectural surveys or domain-specific tracking literature outside point cloud methods. The taxonomy structure suggests PointRePar addresses an active but underpopulated research direction where unified approaches remain less explored than category-specific paradigms.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: PointRePar: Category-Unified 3D SOT Framework with Spatiotemporal Point Relation Parsing

Description: The authors introduce PointRePar, a category-unified 3D single object tracking model that enables joint training across multiple object categories while achieving robust performance through spatiotemporal point relation parsing. Unlike category-specified methods that train separately per category, this framework learns generalizable patterns across categories.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. TrackAny3D: Transferring Pretrained 3D Models for Category-unified 3D Point Cloud Tracking

URL: View paper

Prior Art Analysis

TrackAny3D[23] demonstrates that prior work exists in category-unified 3D single object tracking. Both papers address the same fundamental problem: enabling a single model to track objects across multiple categories without category-specific training. TrackAny3D[23] explicitly states it is 'the first framework to transfer large-scale pretrained 3D models for category-agnostic 3D sot' and achieves 'state-of-the-art performance on category-agnostic 3D sot.' The candidate paper was published at ICCV (a major venue), while the original paper is under review at ICLR 2026, suggesting TrackAny3D[23] predates the original work. Both papers compare against the same baseline methods (mocut, siamcut) and address identical limitations of category-specific approaches.

Evidence

Evidence 1 - **Rationale**: Both papers claim to propose novel category-unified 3D SOT frameworks. TrackAny3D[23] explicitly claims to be 'the first framework' for this task, which directly challenges the novelty of PointRePar's category-unified approach. - **Original**: we propose a robust category-unified 3d sot model, referred to as spatiotemporal point relation parsing model (pointrepar), which is capable of joint training across multiple categories while excelling in unified feature learning for both spatial shapes and temporal motions. - **Candidate**: we propose trackany3d, the first framework to transfer large-scale pretrained 3d models for category-agnostic 3d sot. we first integrate parameterefficient adapters to bridge the gap between pretraining and tracking tasks while preserving geometric priors.

Evidence 2 - **Rationale**: Both papers identify the same fundamental limitation of existing methods: category-specific training paradigms that lack generalization. This shows they address the identical research gap. - **Original**: most existing methods employ a category-specified optimization paradigm, training the tracking model individually for each object category to enhance tracking performance, albeit at the expense of generalizability across different categories. - **Candidate**: current category-specific approaches achieve good accuracy but are impractical for real-world use, requiring separate models for each category and showing limited generalization.

Evidence 3 - **Rationale**: Both papers acknowledge the same prior work (cutrack/mocut by nie et al., 2024) as the existing category-unified approach, demonstrating they are working on the same established problem space. - **Original**: recent studies have initiated the exploration of the category-unified training paradigm for the 3d sot task. a prominent example is cutrack(nie et al., 2024), which introduces deformable grouping vector attention mechanism to dynamically adapt to objects of diverse sizes and shapes across categories... - **Candidate**: only recently has mocut [30] focused on this issue. however, mocut's solution often relies on manually designed rules and hyperparameter tuning.

Evidence 4 - **Rationale**: Both papers evaluate on three benchmarks and claim state-of-the-art performance on category-unified/category-agnostic 3D SOT, showing they target the same evaluation criteria and problem definition. - **Original**: extensive experiments across three benchmarks demonstrate that ourpointreparnot only outperforms the existing category-unified 3d sot method cutrack significantly, but also compares favorably against the state-ofthe-art category-specified methods. - **Candidate**: experiments on three commonlyused benchmarks show that trackany3d establishes new state-of-the-art performance on category-agnostic 3d sot, demonstrating strong generalization and competitiveness.

Evidence 5 - **Rationale**: Both papers report identical numerical results for the same baseline methods (siamcut, mocut) on the same dataset (NuScenes), demonstrating they are evaluating against the exact same prior work and using the same experimental setup. - **Original**: siamcut (nie et al., 2024)40.96/44.9131.42/53.8053.91/52.6563.29/58.2141.03/38.0140.41/48.54 mocut (nie et al., 2024)57.32/66.0133.47/63.1261.75/64.3860.90/61.8457.39/56.0751.19/64.63 - **Candidate**: siamcut [30] 40.96 / 44.91 31.42 / 53.80 53.91 / 52.65 63.29 / 58.21 41.03 / 38.01 40.41 / 48.54 mocut [30] 57.32 / 66.01 33.47 / 63.12 61.75 / 64.38 60.90 / 61.84 57.39 / 56.07 51.19 / 64.63

---

2. Spatial-temporal relation networks for multi-object tracking

URL: View paper

Brief Assessment

Spatial Temporal Relations[76] focuses on multi-object tracking (MOT) in 2D video with spatial-temporal relation networks for object association across frames, not category-unified 3D single object tracking with point cloud relation parsing.

---

3. BEVFormer: Learning Birdâ□s-Eye-View Representation From LiDAR-Camera via Spatiotemporal Transformers

URL: View paper

Brief Assessment

BEVFormer[71] focuses on multi-modality fusion (LiDAR-camera) for BEV representation learning across multiple perception tasks, not on category-unified 3D single object tracking with spatiotemporal point relation parsing.

---

4. Multi-person articulated tracking with spatial and temporal embeddings

URL: View paper

Brief Assessment

Multi Person Embeddings[74] focuses on multi-person pose estimation and tracking in 2D images/videos using spatial and temporal embeddings for human body parts. The original paper addresses 3D single object tracking in point clouds with category-unified training across object categories. These are fundamentally different tasks in different domains (2D multi-person tracking vs. 3D single object tracking).

---

5. Shasta: Modeling shape and spatio-temporal affinities for 3d multi-object tracking

URL: View paper

Brief Assessment

Shasta[73] addresses multi-object tracking (MOT) with multiple objects per scene, while PointRePar focuses on single object tracking (SOT) with one target per scene. The tasks, problem formulations, and technical approaches differ fundamentally.

---

6. Unified Multi-Modal Object Tracking Through Spatialâ□□Temporal Propagation and Modality Synergy

URL: View paper

Brief Assessment

Unified Multi Modal[78] focuses on multi-modal object tracking across RGB-D, RGB-T, and RGB-E tasks with cross-modal fusion, not category-unified 3D point cloud tracking with spatiotemporal relation parsing.

---

7. SCGTracker: Spatio-temporal correlation and graph neural networks for multiple object tracking

URL: View paper

Brief Assessment

SCGTracker[75] focuses on multiple object tracking (MOT) with spatio-temporal correlations between multiple objects, while PointRePar addresses single object tracking (SOT) with category-unified training across object categories. These are fundamentally different tracking paradigms with distinct technical challenges.

---

8. Delving into Dynamic Scene Cue-Consistency for Robust 3D Multi-Object Tracking

URL: View paper

Brief Assessment

Dynamic Scene Consistency[79] focuses on 3D multi-object tracking (MOT) using spatial cue-consistency between multiple objects in crowded scenes, while PointRePar addresses single object tracking (SOT) with category-unified training across object categories. These are fundamentally different tasks with distinct technical approaches.

---

9. Standing between past and future: Spatio-temporal modeling for multi-camera 3d multi-object tracking

URL: View paper

Brief Assessment

Past Future Modeling[77] addresses multi-camera 3D multi-object tracking (MOT) with spatio-temporal modeling for trajectory prediction and occlusion handling, while the original paper focuses on category-unified 3D single object tracking (SOT) with spatiotemporal point relation parsing using Mamba-based architectures for point cloud feature learning.

---

10. L4P: Towards Unified Low-Level 4D Vision Perception

URL: View paper

Brief Assessment

Low Level 4D[72] focuses on unified low-level 4D vision perception tasks (depth estimation, optical flow, 2D/3D tracking) using video data, while the original paper addresses category-unified 3D single object tracking on point clouds with spatiotemporal relation parsing. These are fundamentally different problem domains and data modalities.

Contribution 2: U-shaped Spatial Relation Parsing Mamba with Dynamic Feature Aggregation

Description: The authors propose a Dynamic Feature Aggregation (DFA) mechanism that adaptively refines point features and a U-shaped Spatial Relation Parsing Mamba (USRPM) architecture that captures multi-scale spatial dependencies through hierarchical Mamba-based encoding with bidirectional scanning.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. MCNet: A multi-level consistency network for 3D point cloud self-supervised learning

URL: View paper

Brief Assessment

Multi Level Consistency[63] focuses on multi-level consistency learning for self-supervised point cloud representation, not on dynamic feature aggregation mechanisms or U-shaped Mamba architectures for supervised tracking tasks.

---

2. Two-stream multi-level dynamic point transformer for two-person interaction recognition

URL: View paper

Brief Assessment

Two Stream Interaction[70] focuses on two-person interaction recognition using point clouds from RGB+D videos, not on 3D object tracking or general point cloud spatial relation parsing with Mamba architectures.

---

3. Multi-Level Cross-Attention Point Cloud Completion Network

URL: View paper

Brief Assessment

Multi Level Completion[65] focuses on point cloud completion using multi-level cross-attention, not on spatial relation parsing with Mamba-based architectures or dynamic feature aggregation for 3D tracking tasks.

---

4. Mdcsnet: multi-scale dynamic spatial information fusion with criticality sampling for point cloud classification

URL: View paper

Brief Assessment

Multi Scale Dynamic[61] focuses on multi-scale dynamic spatial information fusion with criticality sampling for point cloud classification tasks, not 3D object tracking with temporal modeling and U-shaped Mamba architectures.

---

5. PointMM: point cloud semantic segmentation CNN under multi-spatial feature encoding and multi-head attention pooling

URL: View paper

Brief Assessment

Multi Spatial Encoding[68] focuses on semantic segmentation using multi-spatial feature encoding and multi-head attention pooling, not on 3D object tracking with Mamba-based architectures or dynamic feature aggregation for temporal motion modeling.

---

6. GAF-Net: geometric contextual feature aggregation and adaptive fusion for large-scale point cloud semantic segmentation

URL: View paper

Brief Assessment

Geometric Feature Aggregation[64] focuses on semantic segmentation of large-scale point clouds using geometric contextual features and adaptive fusion between encoder-decoder, not on spatiotemporal tracking with Mamba-based architectures for dynamic feature refinement.

---

7. 2-s3net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network

URL: View paper

Brief Assessment

Attentive Feature Fusion[62] focuses on sparse semantic segmentation for LiDAR point clouds using multi-branch feature fusion and adaptive feature selection, not on 3D object tracking with Mamba-based architectures for spatiotemporal relation parsing.

---

8. Infrastructure-side point cloud object detection via multi-frame aggregation and multi-scale fusion

URL: View paper

Brief Assessment

Multi Frame Aggregation[69] focuses on infrastructure-side multi-frame point cloud detection for autonomous driving, not single object tracking with U-shaped Mamba architectures and dynamic feature aggregation for spatial relation parsing.

---

9. A hierarchical framework for threeâ□□dimensional pavement crack detection on point clouds with multiâ□□scale abnormal region filtering and multimodal interaction fusion

URL: View paper

Brief Assessment

Hierarchical Pavement Crack[67] focuses on pavement crack detection from depth and normal images with multi-level feature extraction, not on point cloud tracking with Mamba-based architectures or dynamic feature aggregation mechanisms.

---

10. Point Cloud Semantic Segmentation with Transformer and Multi-Scale Feature Extraction

URL: View paper

Brief Assessment

Multi Scale Feature[66] focuses on semantic segmentation using transformer and dilated convolutions for multi-scale feature extraction, not on 3D object tracking with Mamba-based U-Net architectures and dynamic feature aggregation for spatial relation parsing.

Contribution 3: Long-term Temporal Relation Parsing with Conditional Gaussian Perturbation

Description: The authors develop a temporal modeling approach that captures both point-level motion through Temporal Scan Mamba and box-level trajectory patterns through Long-term Motion Trajectory Rectification. They also introduce Conditional Gaussian Perturbation (CGP), a density-aware noise injection method that simulates prediction errors conditioned on scene sparsity to improve robustness.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Extraction and temporal segmentation of multiple motion trajectories in human motion

URL: View paper

Brief Assessment

Multiple Motion Trajectories[56] focuses on extracting motion trajectories from human body parts (hands/feet) in ballet videos for activity recognition, not on 3D point cloud tracking with temporal modeling architectures like Mamba or density-aware noise injection methods.

---

2. A framework for food traceability information extraction based on a video surveillance system

URL: View paper

Brief Assessment

Food Traceability Extraction[59] focuses on extracting traceability information from video surveillance for food tracking using photogrammetry methods and temporal trajectories, which is fundamentally different from the original paper's 3D point cloud tracking with Temporal Scan Mamba and Conditional Gaussian Perturbation for robustness enhancement.

---

3. The Temporal Game: A New Perspective on Temporal Relation Extraction

URL: View paper

Brief Assessment

Temporal Game[52] focuses on temporal relation extraction in natural language processing through point-wise comparisons of temporal entities, not on 3D object tracking with point clouds, motion trajectories, or Conditional Gaussian Perturbation for handling scene sparsity.

---

4. Representing pairwise spatial and temporal relations for action recognition

URL: View paper

Brief Assessment

Pairwise Spatial Temporal[53] focuses on pairwise spatial-temporal relationships between quantized features for action recognition in videos, not on 3D point cloud tracking with point-level motion (Temporal Scan Mamba) and box-level trajectory features (Long-term Motion Trajectory Rectification) or density-aware noise injection (Conditional Gaussian Perturbation).

---

5. Global Tracking based Multi-Object Tracking in Complex Environments

URL: View paper

Brief Assessment

Global Tracking Complex[58] focuses on multi-object tracking in video sequences using Kalman filters and transformer architectures for global association across frames. It does not address point cloud tracking, point-level motion modeling via Temporal Scan Mamba, box-level trajectory rectification, or density-aware noise injection methods like Conditional Gaussian Perturbation for 3D single object tracking.

---

6. Motion Prompting: Controlling Video Generation with Motion Trajectories

URL: View paper

Brief Assessment

Motion Prompting[55] focuses on motion trajectory conditioning for video generation using point tracks, not on 3D object tracking with temporal relation parsing and Conditional Gaussian Perturbation for robustness in sparse scenes.

---

7. Selfâ□□Supervised Learning of Part Mobility from Point Cloud Sequence

URL: View paper

Brief Assessment

Part Mobility Learning[51] focuses on self-supervised learning of part mobility from point cloud sequences using trajectory-based analysis and PointRNN architecture. The original paper addresses 3D object tracking with temporal modeling for motion prediction, which is a fundamentally different task and application domain.

---

8. Exploiting spatial-temporal context for trajectory based action video retrieval

URL: View paper

Brief Assessment

Trajectory Action Retrieval[57] focuses on action video retrieval using trajectory representations with spatial-temporal context, not on 3D object tracking with point-level motion modeling, box-level trajectory rectification, or density-aware noise injection methods like CGP.

---

9. Using LSTM with Trajectory Point Correlation and Temporal Pattern Attention for Ship Trajectory Prediction

URL: View paper

Brief Assessment

Trajectory Point Correlation[54] focuses on ship trajectory prediction using attention mechanisms for spatial-temporal correlations in maritime navigation data (lon, lat, sog, cog). The original paper addresses 3D object tracking with point-level motion via Temporal Scan Mamba and box-level trajectory features through Long-term Motion Trajectory Rectification, plus density-aware noise injection (CGP). These are fundamentally different application domains and technical approaches.

---

10. PTT: Point-Trajectory Transformer for Efficient Temporal 3D Object Detection

URL: View paper

Brief Assessment

Point Trajectory Transformer[60] focuses on trajectory encoding for temporal 3D object detection in autonomous driving, not on category-unified 3D tracking with density-aware noise injection for robustness enhancement.

Appendix: Text Similarity Detection

Textual similarity detection checked 31 papers and found 2 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Towards category unification of 3D single object tracking on point clouds

Detected in: Core Task (sibling)

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] PointRePar : SpatioTemporal Point Relation Parsing for Robust Category-Unified 3D Tracking View paper
- [1] 3d siamese transformer network for single object tracking on point clouds View paper
- [2] PTT: Point-track-transformer module for 3D single object tracking in point clouds View paper
- [3] A lightweight and detector-free 3d single object tracker on point clouds View paper
- [4] Real-time 3D single object tracking with transformer View paper
- [5] OST: Efficient one-stream network for 3D single object tracking in point clouds View paper
- [6] Beyond 3d siamese tracking: A motion-centric paradigm for 3d single object tracking in point clouds View paper
- [7] Toward class-agnostic tracking using feature decorrelation in point clouds View paper
- [8] Revisiting Siamese-Based 3D Single Object Tracking With a Versatile Transformer View paper
- [9] TM2B: Transformer-Based Motion-to-Box Network for 3D Single Object Tracking on Point Clouds View paper
- [10] Ratrack: moving object detection and tracking with 4d radar point cloud View paper
- [11] Any3DIS: Class-Agnostic 3D Instance Segmentation by 2D Mask Tracking View paper
- [12] 3d-siamrpn: An end-to-end learning method for real-time 3d single object tracking using raw point cloud View paper
- [13] Pointtracknet: An end-to-end network for 3-d object detection and tracking from point clouds View paper
- [14] Center-based 3d object detection and tracking View paper
- [15] PillarTrack: Redesigning Pillar-based Transformer Network for Single Object Tracking on Point Clouds View paper
- [16] Towards category unification of 3D single object tracking on point clouds View paper
- [17] DeepPCT: Single object tracking in dynamic point cloud sequences View paper
- [18] Exploit Spatiotemporal Contextual Information for 3D Single Object Tracking via Memory Networks View paper
- [19] 3d single-object tracking in point clouds with high temporal variation View paper
- [20] Structure aware 3D single object tracking of point cloud View paper
- [21] MixCycle: Mixup Assisted Semi-Supervised 3D Single Object Tracking with Cycle Consistency View paper
- [22] Moving object detection and tracking with 4d radar point cloud View paper
- [23] TrackAny3D: Transferring Pretrained 3D Models for Category-unified 3D Point Cloud Tracking View paper
- [24] Self-supervised class-agnostic motion prediction with spatial and temporal consistency regularizations View paper
- [25] Density-based clustering for 3d object detection in point clouds View paper
- [26] Category-Level Multi-Object 9D State Tracking Using Object-Centric Multi-Scale Transformer in Point Cloud Stream View paper
- [27] PPE: Point position embedding for single object tracking in point clouds View paper
- [28] Weakly supervised class-agnostic motion prediction for autonomous driving View paper
- [29] Complexer-yolo: Real-time 3d object detection and tracking on semantic point clouds View paper
- [30] Unsupervised class-agnostic instance segmentation of 3d lidar data for autonomous vehicles View paper
- [31] Feature-concatenated transformer for 3D object tracking in point clouds View paper
- [32] Mbptrack: Improving 3d point cloud tracking with memory networks and box priors View paper
- [33] Exploring simple 3d multi-object tracking for autonomous driving View paper
- [34] 3D multi-object tracking in point clouds based on prediction confidence-guided data association View paper
- [35] Learning Adaptive Conceptual Prototypes for 3D Single Object Tracking View paper
- [36] Implicit and efficient point cloud completion for 3D single object tracking View paper
- [37] Open3dtrack: Towards open-vocabulary 3d multi-object tracking View paper
- [38] Obstacle tracking for unmanned surface vessels using 3-D point cloud View paper
- [39] Any motion detector: Learning class-agnostic scene dynamics from a sequence of lidar point clouds View paper
- [40] CDTracker: Coarse-to-Fine Feature Matching and Point Densification for 3D Single-Object Tracking View paper

- [41] Exploiting more information in sparse point cloud for 3D single object tracking View paper
- [42] Correlation Pyramid Network for 3D Single Object Tracking View paper
- [43] PillarTrack:Boosting Pillar Representation for Transformer-based 3D Single Object Tracking on Point Clouds View paper
- [44] GLT-T: Global-Local Transformer Voting for 3D Single Object Tracking in Point Clouds View paper
- [45] Enhancing 3D Single Object Tracking with Efficient Point Cloud Segmentation View paper
- [46] 3D object tracking using RGB and LIDAR data View paper
- [47] An Effective Motion-Centric Paradigm for 3D Single Object Tracking in Point Clouds View paper
- [48] Modeling continuous motion for 3d point cloud object tracking View paper
- [49] M3SOT: Multi-frame, Multi-field, Multi-space 3D Single Object Tracking View paper
- [50] SPAN: siampillars attention network for 3D object tracking in point clouds View paper
- [51] Selfâ□Supervised Learning of Part Mobility from Point Cloud Sequence View paper
- [52] The Temporal Game: A New Perspective on Temporal Relation Extraction View paper
- [53] Representing pairwise spatial and temporal relations for action recognition View paper
- [54] Using LSTM with Trajectory Point Correlation and Temporal Pattern Attention for Ship Trajectory Prediction View paper
- [55] Motion Prompting: Controlling Video Generation with Motion Trajectories View paper
- [56] Extraction and temporal segmentation of multiple motion trajectories in human motion View paper
- [57] Exploiting spatial-temporal context for trajectory based action video retrieval View paper
- [58] Global Tracking based Multi-Object Tracking in Complex Environments View paper
- [59] A framework for food traceability information extraction based on a video surveillance system View paper
- [60] PTT: Point-Trajectory Transformer for Efficient Temporal 3D Object Detection View paper
- [61] Mdcsnet: multi-scale dynamic spatial information fusion with criticality sampling for point cloud classification View paper
- [62] 2-s3net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network View paper
- [63] MCNet: A multi-level consistency network for 3D point cloud self-supervised learning View paper
- [64] GAF-Net: geometric contextual feature aggregation and adaptive fusion for large-scale point cloud semantic segmentation View paper
- [65] Multi-Level Cross-Attention Point Cloud Completion Network View paper
- [66] Point Cloud Semantic Segmentation with Transformer and Multi-Scale Feature Extraction View paper
- [67] A hierarchical framework for threeâ□dimensional pavement crack detection on point clouds with multiâ□scale abnormal region filtering and multimodal interaction fusion View paper
- [68] PointMM: point cloud semantic segmentation CNN under multi-spatial feature encoding and multi-head attention pooling View paper
- [69] Infrastructure-side point cloud object detection via multi-frame aggregation and multi-scale fusion View paper
- [70] Two-stream multi-level dynamic point transformer for two-person interaction recognition View paper
- [71] BEVFormer: Learning Birdâ□s-Eye-View Representation From LiDAR-Camera via Spatiotemporal Transformers View paper
- [72] L4P: Towards Unified Low-Level 4D Vision Perception View paper
- [73] Shasta: Modeling shape and spatio-temporal affinities for 3d multi-object tracking View paper
- [74] Multi-person articulated tracking with spatial and temporal embeddings View paper
- [75] SCGTracker: Spatio-temporal correlation and graph neural networks for multiple object tracking View paper
- [76] Spatial-temporal relation networks for multi-object tracking View paper
- [77] Standing between past and future: Spatio-temporal modeling for multi-camera 3d multi-object tracking View paper
- [78] Unified Multi-Modal Object Tracking Through Spatialâ□Temporal Propagation and Modality Synergy View paper
- [79] Delving into Dynamic Scene Cue-Consistency for Robust 3D Multi-Object Tracking View paper