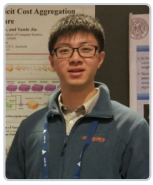
Lidong Yu

(+86) 151-6523-1608 | yvlidong@gmail.com | Beijing, China
Google Scholar | lidongyv.github.io



EDUCATION

Beijing Institute of Technology (985) Aug 2016 – Jun 2019
M.S. in Computer Science & Technology, School of Computer Science Beijing

Ocean University of China (985) Aug 2012 – Jun 2016
B.S. in Computer Science & Technology, School of Computer Science Qingdao

WORK EXPERIENCE

Neolix Technology (Beijing) Co., Ltd. Aug 2025 – Present
Head of AD Algorithms, AD Division II Beijing
Algorithm org of 38 people, ~8,700 GPUs

- **Perception**: group query-based multi-task model for object detection, lane detection, and occupancy network; E2E-based traffic light detection and navigation model
- **PNC & Two-stage Planner**: static trajectory prediction + downstream dynamic trajectory optimization
- **E2E & World Model**: two-stage pretrain with LLM distillation and perception model inception; AR-based trajectory prediction with multi-task FT and CoT debugger; post-training with RL
- **Infra**: data closed loop; offline large model for auto-labeling; auto-balance for multi-task training; multi-platform storage and training; closed-loop evaluation and simulator

NVIDIA Jul 2024 – Aug 2025
Principal Engineer, Free IC of the Autonomous Vehicle Department Beijing

- **Modeling**: Roadnet (16 people, ~1,200 GPUs, 3 mo) HD→Mapless, segment→query→token-based; Planner (13 people, ~1,500 GPUs, 3 mo) classical PNC→model-based, two-stage E2E with AR; E2E (13 people, ~3,500 GPUs, 6 mo) Alpamayo mass-production ORIN delivery for Mercedes-Benz SOP, one-stage E2E with Cosmos CoT and Alpamayo WM-RL
- **Infra**: perception continuous training with data auto-balance & mining framework, multi-task training optimization with distillation; E2E-Alpamayo backbone pretrain / FT / post-training framework

Horizon Robotics Jun 2023 – Jul 2024
Head of Static Perception Algorithms, "Saturn V" Project Beijing
Led a 15-person modeling + 5-person post-processing team, ~1,200 GPUs

- Owned static-perception model + post-processing + data infra; segment-based → query-based lane detection → group-wise multi-task; static E2E + navigation model
- **Infra** (7 devs + 5 contractors + 20 downstream): built data closed loop 0→1; re-architected the loop (issue turnaround D+10→D+3); automated labeling tool (4D GT), 1k→30k intersections/month; auto data mining, redundancy 80%→5%

DeepRoute.ai (Shenzhen) Apr 2022 – Jun 2023
Head of Static Perception Models Beijing
Core team of 4, 600 GPUs

- Delivered China's first Mapless solution (2023) and the AITO (Wenjie) POC; lane-segment solution for urban point-to-point and highway auto-cruise, >85% pass rate at ramps/intersections; lane segment → query-based → AR-based lane detection
- **Infra**: >200M-image data mining platform, ~12M-frame auto mining; auto-balance framework for multi-task training

NIO (Shanghai) Mar 2021 – Apr 2022
Principal Engineer, Autonomous Driving Algorithms Beijing

- China's first-generation in-house AD algorithms for NIO ET7 and ES8; data-driven Mono3D detection-tracking model; data closed loop and model evaluation framework for object detection, lane detection, and post-processing

UISEE Technology (Beijing) May 2019 – Apr 2020
Senior Researcher, Nanjing AI Research Institute Nanjing

- In 2019, first in China to deploy a Shadow-Mode data closed-loop system + IL-based E2E on-vehicle; imitation learning, RL, data generation, and simulation for AD; joint 3D object detection and tracking

PUBLICATIONS

1. 3D Visual Illusion Depth Estimation. **NeurIPS 2025**
2. Diving into the Fusion of Monocular Priors for Generalized Stereo Matching. **ICCV 2025**
3. Sparse Point Guided 3D Lane Detection. **ICCV 2023**
4. Parameterized Cost Volume for Stereo Matching. **ICCV 2023**
5. FoggyStereo: Stereo Matching with Fog Volume Representation. **CVPR 2022**
6. Deep Stereo Matching with Explicit Cost Aggregation Sub-Architecture. **AAAI 2018**
7. Content-Aware Inter-Scale Cost Aggregation for Stereo Matching. **arXiv 2020**
8. DRL-based Trajectory Tracking for Motion-related Modules in Autonomous Driving. **arXiv 2023**

PATENTS

1. 3D reconstruction method, apparatus, system, and storage medium. (CN201980002779.1)

2. Image prediction and vehicle behavior planning method, apparatus, system, and storage medium. (CN202010196263.X)
3. Multi-object tracking method, apparatus, system, and storage medium. (CN201911334843.4)
4. Single-object tracking method, apparatus, system, and storage medium. (CN201911194289.4)
5. Training method and system for generative adversarial networks, electronic device, and storage medium. (CN201980002600.2)

PROJECT DETAILS

Neolix — AD Algorithms

Aug 2025 – Present

- **Perception** — refactored single-task / multi-task formulation
  - Used a group query-based model to unify single-task and multi-task training, moving from a multi-head design to a data-friendly single-head multi-task design
  - One query-based model for object detection, lane detection, occupancy network, E2E traffic light detection, and navigation model
  - Single head allows self-configured data balance without affecting other tasks; rebalances task weight by BEV-token / image-token at token level rather than task-loss level
  - Refactored perception tasks to fit E2E: LiDAR early fusion, object detection, MOT upgraded to object prediction, and a future tokenizer supporting pre/post-training
  - Upgraded lane and traffic-light detection to a navi-based static E2E model, reducing learning difficulty and providing a strong prior for upstream decisions
- **E2E & World Model**
  - Two-stage pretrain — LLM distillation from Cosmos / VGGT (CoT, 3D-geometry estimation, 3D OD, lane tasks, failsafe, scene recognition); perception model inception (AD task-force knowledge for RL, plus a visual debugger)
  - Controllable iterative training & release — pretrain backbone with frozen head; fine-tune perception head with frozen backbone; add data at any time
  - Auto-regression-based trajectory prediction — multi-task FT with pre/post perception / perceiver; CoT debugger; multi-modal diffusion; post-training with RL
  - Generalized world model — pretrain (LLM distillation for backbone; Cosmos→CoT language inception; D4RT unified parallel AR + 3D insight); render (generalized generation rather than single-scene reconstruction; DiT & 3D-GS only for render); conditional inception (language / layout / trajectory)
- **Infra** — AI-native & RaaS transformation
  - Auto triage with VLM; evaluation and simulation via agent skills; efficient meetings with personal AI agents; vibe-coding-based auto research
  - Data closed loop for E2E and WM: auto-labeling with offline perception models; auto mining with VLM
  - Agent-based front-end / back-end development and customer service; auto checker for the simulator

NVIDIA — AD / E2E / World Model

Jun 2024 – Jul 2025

- **Perception Model → Two-stage E2E**
  - Segment-based lane detection (what-you-see-is-what-you-get): optimized super-resolution, improved geometric precision, BEV multi-scale fusion, dynamic multi-resolution DCN
  - Query-based lane detection (static E2E): moved beyond WYSIWYG, enhanced reasoning, navi-info input, added intersection crossing, split/merge, lane change, and other upstream navi decisions
  - Token-based lane detection: enabled the two-stage→one-stage Alpamayo upgrade, tokenizing all perception elements via AR, supporting upstream conditional SA and downstream debugger visualization
- **PNC → One-stage Alpamayo**
  - Model-ized the planning module; single-mode → multi-mode regression; regression → AR → diffusion formulation
  - Perception moved downstream for the one-stage upgrade: Cosmos / Qwen / Llama backbone distillation; CoT, 3D-geometry, and AD multi-task inception training; removed the perception tokenizer (re-purposed as a debugger task); changed planning from perception-based to image-stream-based
  - Planning FT and post-training: AR, diffusion, and multi-modal trajectory prediction; RL post-training; closed-loop training with Cosmos-WM
- **World Model** — Cosmos pretrain → render research; distillation from VGGT and Qwen

Horizon "Saturn V" — Road-Structure Understanding

Aug 2023 – Aug 2024

Light-map PNC solution and a road-understanding module under crowdsourced mapping.

- **Algorithm R&D — continuous elements (lane lines)**
  - WYSIWYG strategy for road edges: the model outputs only lane lines and road edges (no centerlines); designed data-cleaning and labeling rules for occluded-intersection scenarios
  - Replaced segment-based with query-based, removing post-processing and shifting effort upstream at comparable compute cost; F1 improved by 15%
  - Large recall gains on lateral intersection sides and distant weakly-visible regions; unlike the entry-point strategy, supports building lanes directly
  - Query-based temporal strategy moved tracking into the network; PV refinement reduced strong dependence on BEV features and introduced a new multi-modal fusion approach; query grouping enabled case-level fine-tuning
- **Discrete elements (arrow, crosswalk, junction)**
  - Replaced the two-stage detector with a query-based solution at comparable score and >30% higher compute efficiency
  - Query grouping closed the single-task / multi-task metric gap and enabled one-model multi-task training with same-source data
- **Data closed loop**
  - Built full-element, multi-task, same-source temporal datasets for road structure; optimized data collection, labeling rules, and balancing for the query-based approach
  - Co-developed the automated labeling tool (4D GT); helped architect and implement, from 0 to 1, the large closed-loop systems (automated collection, road-test link, vehicle-cloud collaboration, and road-coverage systems)

DeepRoute — Lane Detection under MapFree

May 2022 – Jun 2023

Data-driven system for lane detection.

- **GT generation**: built an HD-map→local-BEV GT system (2M images, 12M lane instances)
- **Projection-GT-friendly model**: optimized segment-based lane detection (AP 13%→60%, EPE 1.5m→0.18m); labeling-rule definition and data cleaning; direction-head linking with SR for higher precision
- **Query-based lane detection**: multi-level regression over lane–laneline–point chains; full-stack development of data-topology definition, algorithm, evaluation, and visualization tools
- **Language-model-based topology learning**: auto-regressive keypoint generation; data-topology definition and algorithm development
- **Segment-based topology learning**: MaskFormer-based topology-keypoint segmentation; "reverse transformer" converting query-based point-chain generation into a segment algorithm
- Data-driven automatic model iteration and evaluation

NIO — 3D OD & Lane (Mass Production)

Mar 2021 – Apr 2022

- **Responsibilities**
  - Developed rule-based association post-processing for 7-camera + 1-LiDAR tracking association
  - Learning-based association: designed a learnable cost metric and matching scheme to replace rules and bipartite-graph matching
  - ReID data labeling and pseudo-label generation
  - On-vehicle deployment and testing of post-processing; ET7 NAD & SUD feature delivery
  - Monocular depth estimation: data labeling, algorithm development, quantized deployment and evaluation
  - Ground model: data labeling, algorithm development, quantized deployment, road-test regression
  - 3D lane-line solution support and post-processing R&D
- **Results**
  - Association at 20% 3D error rate: 0.03 ID-switch, 86% recall — overall performance exceeding EQ4
  - 0.1% split rate for distant small objects beyond 150 m; robust association in congestion and parking scenarios
  - Full-image dense depth at 6.8% rel. error; ground at 1.3% rel. error
  - Passed production HOMO testing; LKA and LKW feature tests passed
  - Supported ET7 SUD & NAD road testing and mass-production deployment

Research Projects

Mar 2021 – Aug 2025

- Deep stereo matching with explicit cost aggregation ↗
- FoggyStereo: stereo matching with fog volume representation ↗
- Diving into the fusion of monocular priors for generalized stereo matching ↗
- 3D visual illusion depth estimation ↗
- Sparse point guided 3D lane detection ↗

UISEE — AD Research

Jul 2019 – Dec 2020

- Occluded 3D object detection and tracking with RepPoint formulation ↗
- Clustering-based feature disentanglement for ReID, SOT, and MOT ↗
- Relative navigation and shadow-mode data collection with trajectory imitation ↗