Trung M. Bui, PhD

Computer Vision & Machine Learning Engineer — Production Systems

Morgan Hill, CA • +1 669-326-2460 • bmtrungvp@gmail.com
linkedin.com/in/trung-m-bui • github.com/mtbui2010 • **U.S. Green Card**

Summary

Computer Vision and Machine Learning Engineer with 7+ years shipping production CV systems and 11+ years total CV experience including PhD research. Specializes in **deep learning models for object detection, segmentation, and 6D pose estimation**, with hands-on experience training, fine-tuning, and deploying transformer-based architectures at real-time speeds. Published in *IEEE TIP* (Q1) and *IEICE Transactions*. Hands-on experience with VLMs (CLIP, BLIP, LLaVA, GroundingDINO, Gemini Vision), multi-object tracking (SORT / DeepSORT / ByteTrack), vision-language-action policies, RAG systems, and prompt engineering. Strong production track record: models running 3+ years at 92% accuracy, 30+ FPS on edge hardware.

Experience

Senior Computer Vision & ML Engineer *March 2019 – Present*

Korea Electronics Technology Institute (KETI), Seongnam, South Korea

- Led perception ML for a **production CV system** that improved success from 70% to **92% in cluttered real-world scenes**, processing hundreds of items per hour. System in production 3+ years across multiple deployment sites.
- Designed a **vision transformer architecture** with fine-grained attention for precision detection and pose estimation, achieving **20% precision improvement over CNN baselines**. Published in *IEICE Transactions*, Nov 2025.
- Built end-to-end **training pipelines**: dataset curation (synthetic + real + production failures), augmentation strategies, hyperparameter tuning, evaluation harnesses, continuous improvement loops.
- Trained and shipped deep learning models for **object detection, segmentation, and 6D pose estimation** on RGB-D data; handled long-tail challenges via active learning and open-vocabulary detection.
- Optimized inference for edge deployment with **TensorRT, ONNX, INT8 quantization** on NVIDIA Jetson, sustaining **30+ FPS** with **50% latency reduction** versus FP32 baseline.
- Integrated **VLMs (CLIP, BLIP, LLaVA, GroundingDINO, Gemini Vision)** for zero-shot recognition, open-vocabulary detection, and semantic grounding; implemented **RAG pipelines** combining foundation models with internal knowledge bases.
- Developed **prompt engineering strategies** for LLMs and VLMs to elicit structured outputs and complex reasoning; built multi-agent framework with LangGraph for production-grade orchestration.
- Led cross-functional collaboration with engineering, hardware, and customer deployment teams; mentored 3 junior engineers; conducted code and design reviews.
- Contributed core CV components to national R&D projects with multi-million USD funding.

Selected Projects

Fine-Aware Vision Transformer for Precision Detection *Published, IEICE 2025*

- Designed a novel transformer architecture with fine-grained attention module for cluttered RGB-D scenes; standard ViT struggled with global pooling treating background as equally important as foreground.
- Achieved **20% precision improvement** over CNN baselines; trained on 200K+ annotations with custom augmentation; L1 loss on pose regression + BCE on quality scoring.
- Deployed on NVIDIA Jetson with TensorRT, sustained 30+ FPS at production accuracy.

Tracker Lab — Multi-Object Tracking from Scratch *2025*

github.com/mtbui2010/vision_tracking

- Implemented **SORT, DeepSORT, ByteTrack, and a custom tracker from scratch** — Kalman filter, Hungarian assignment, and motion/appearance/IoU cost functions derived in notebooks, no third-party tracking libraries.
- Detection with fine-tuned **YOLOv11** (MOT17/DanceTrack); ReID embeddings via **OSNet**; custom **MOTA / IDF1 / HOTA** metrics benchmarked against MOT17-val.
- Full-stack research platform: FastAPI orchestration + Next.js 14 / Canvas UI, side-by-side tracker comparison, frame-by-frame algorithm stepping; deployed via Docker, Vercel, and RunPod serverless GPU workers.

VLA Manipulation Policies (Isaac Lab + MuJoCo) *2025*

github.com/mtbui2010/robot_sim_vla

- Evaluated **OpenVLA-OFT** vision-language-action policies for a Franka Panda arm across Isaac Lab and MuJoCo simulators.
- Reached **98.75% success on the LIBERO benchmark** (80 episodes), edging the 97.1% paper baseline with only 2 trials/task; built fine-tuning infra and a LeRobot diffusion-policy baseline.

GroundingDINO Annotation & Training Tool *2024*

github.com/mtbui2010/groundingdino_tool

- Streamlit-based web tool for semi-automatic image labeling and open-vocabulary detector fine-tuning.
- Pipeline: YOLOv8n initial detection GroundingDINO open-vocabulary refinement FastSAM segmentation human verification fine-tuning loop.
- Designed for the **long-tail problem**: closes label gaps for novel objects without requiring large pre-labeled datasets.

Multimodal AI System with VLM-Grounded Reasoning *2024 – Present*

github.com/mtbui2010/carerobotagent

- Multi-agent framework integrating **VLMs with perception pipelines** for scene understanding and structured decision-making.
- Implemented **RAG** over internal knowledge base; multi-backend LLM support (OpenAI, Anthropic, Ollama); episodic + semantic memory (SQLite + ChromaDB).
- Tested in AI2-THOR simulator with switchable planning strategies (ReAct, Chain-of-Thought, Hierarchical, Self-Refine).

Single Image Dehazing Using Color Ellipsoid Prior *IEEE TIP 2018*

github.com/mtbui2010/Single-Image-Dehazing

- Novel algorithm for single-image dehazing; PhD research basis.
- Published in *IEEE Transactions on Image Processing* (Q1 journal). Outperformed existing methods on standard benchmarks.

Selected Publications

- **T. M. Bui**, J. Hwang, S. Jun, W. Kim, D. Shin. "A Fine-Aware Vision Transformer for Precision Grasp Pose Detection." *IEICE Transactions on Information and Systems*, Nov. 2025.
- **T. M. Bui**, W. Kim. "Single Image Dehazing Using Color Ellipsoid Prior." *IEEE Transactions on Image Processing*, Feb. 2018. *(Q1 journal)*
- **T. M. Bui**, Y. Kim, S. J. Moon, M. Cho, M. Seo, D. Shin. "Development of a Mobile Assistive Robot for Daily Living Support." *Ubiquitous Robots*, 2025.
- **T. M. Bui**, H. N. Tran, W. Kim, S. Kim. "Segmenting Dark Channel Prior in Single Image Dehazing." *IET Electronics Letters*, March 2014.

Education

Ph.D. in Computer Vision *2014 – 2019*

Kyung Hee University, South Korea

Thesis: *Single Image Dehazing Using Color Ellipsoid Prior* (basis of IEEE TIP 2018 publication)

M.Eng. in Computer Vision *2011 – 2014*

Kyung Hee University, South Korea

B.Eng. in Electrical and Electronics Engineering *2005 – 2010*

Ho Chi Minh City University of Technology, Vietnam — Excellent Engineer Training Program

Technical Skills

Computer Vision: object detection, semantic / instance segmentation, 6D pose estimation, multi-object tracking (SORT / DeepSORT / ByteTrack), person re-identification, RGB-D perception, point cloud processing, video understanding
Deep Learning: vision transformers, CNN architectures, attention mechanisms, transfer learning, self-supervised learning, hyperparameter tuning, model adaptation
Multimodal AI: VLMs (CLIP, BLIP, LLaVA, GroundingDINO, Gemini Vision), LLM integration, prompt engineering, RAG, agent frameworks (LangGraph)
Production ML: large-scale training pipelines, dataset curation, augmentation, evaluation harnesses, shadow deployment, canary rollout, model versioning, drift detection
Frameworks: PyTorch, TensorFlow, Hugging Face, ONNX, TensorRT, CUDA
Languages: Python (expert), C++, C
Edge Deployment: TensorRT optimization, INT8 quantization, NVIDIA Jetson, real-time inference (30+ FPS), sub-200ms latency, model quantization & pruning
Classical CV: OpenCV, calibration, RANSAC, ICP, feature matching, geometric vision
Infrastructure: Docker, Linux, Git, distributed training, Streamlit, FastAPI
Certifications: Visual Perception for Self-Driving Cars (Coursera, 2023); Improving Deep Neural Networks (DeepLearning.AI, 2021)