Personal Statement

Kyuhwan Shim *Seoul National University*

1 Introduction

My academic and professional journey has been shaped by a deep interest in deep learning, robotics, and multimodal AI. I have been particularly intrigued by how AI can perceive, reason, and act as an autonomous agent in the real world, integrating vision, language, and action to make informed decisions. This led me to explore Vision-Language-Action models, where I investigated how robots can process human instructions and dynamically execute actions in complex environments.

Through my studies and research experiences at Seoul National University, Universität Bonn, and the AI industry, I have developed expertise in robotic with a background of data science and deep learning. My work has focused on enabling AI systems to generalize across diverse environments by leveraging multimodal representations and adaptive learning techniques. Moving forward, I am eager to contribute to the advancement of physical AI and intelligent robotic system.

2 Academic and Research Experience

I have pursued research at the intersection of computer vision, natural language processing (NLP), and robotics, leading to multiple publications and real-world AI applications.

2.1 Vision-Language-Action Modeling.

At the Graduate School, I worked on **CLIP-RT[Kang et al., 2024]**, a VLA model which extends the idea of CLIP to robot learning to learn language-conditioned policies from natural language. Different from other VLA models, CLIP-RT learns to predict robotic actions represented in natural language (e.g., "Move arm left") based on contrastive imitation learning. Furthermore, CLIP-RT is a discriminative VLA model that predicts the language-based motion primitive in the predefined list of motion primitives.

My primary contributions to CLIP-RT included the development of **Stochastic Trajectory Diversification (STD)[Kim et al., 2024]**, a novel data augmentation strategy designed to enhance the diversity and robustness of robotic demonstrations. This technique significantly improved the generalization ability of CLIP-RT and was recognized with the Excellence Paper Award at Korea Software Congress 2024. Additionally, I was responsible for **experiment setup, data preprocessing, and post-processing**, ensuring high-quality training data for policy learning.

Additionally, during my undergraduate studies, I contributed to **KIRINO [Choi et al., 2024]**, a chatbot system utilizing retrieval-augmented generation (RAG) to enhance user interactions. While this work primarily focused on NLP, it provided me with valuable insights into retrieval-augmented reasoning, which later influenced my research in robot learning.

2.2 International Research and Robotics Experience

My research experience extended internationally through the Robocup2024@Home DSPL League as a core member of Team TIDYBOY, where I developed an automated labeling tool and managed the dataset pipeline to streamline data processing for real-time perception. Additionally, I played a key role in optimizing the team's workflow to ensure effective collaboration. As part of our efforts, we improved object detection and scene understanding using YOLOv10, enhancing both latency and accuracy in complex environments through multimodal integration. Our team's work led to a third-place finish in the competition.

Furthermore, I participated in a research project as a visiting student researcher at the Autonomous Intelligent Systems Group of Universität Bonn, where I worked on multimodal robotic perception and diffusion-based action generation as part of a broader study on Physical AI. This experience reinforced my interest in integrating AI-driven reasoning with real-world robot navigation and manipulation.

3 Industry and Practical AI Experience

Beyond my academic research, I have applied AI to real-world systems, working at the intersection of machine learning, robotics, and embedded AI.

Machine Learning Engineer at Nota Inc. At Nota Inc., I developed vision-based AI models for intelligent traffic systems, optimizing inference pipelines for real-time deployment on edge devices using AWS and GCP. My work focused on designing object detection models to improve robustness and accuracy across diverse environments, reinforcing my expertise in scalable deep learning architectures.

Co-Founder of OCONNECT As a co-founder of OCONNECT, I led the development of ONESTEP, an easy-to-insert rotary socket designed for accessibility in low-visibility conditions. This innovation was recognized with the James Dyson Award and received funding from the Didimdol R&D Startup Growth Project, Youth Entrepreneurship Academy, and Pre-Startup Package. Through this experience, I gained valuable insights into translating technology into human-centric solutions.

Data Science Competitions I have actively participated in AI and data science competitions, applying machine learning, computer vision, and structured data analytics to real-world problems. I developed a domain-adaptive segmentation algorithm for the Samsung AI Challenge, assist to designd a credit risk rating model for the Gangseo-gu Data Challenge (2023), and created an eco-friendly travel route recommender system integrating geospatial data and NLP, which won the NIA Minister Award at BIGCONTEST(2022).

These experiences honed my ability to extract meaningful insights from data, optimize AI models for deployment, and tackle challenges from a multidisciplinary perspective. Competing in diverse challenges has shaped me into a well-rounded researcher, bridging theoretical AI with practical applications.

4 Conclusion

My long-term research goal is to develop **adaptive AI systems that can reason, learn, and interact in the physical world**. In pursuit of this, I have explored vision-language-action (VLA) models, including leading the VLAVLA project, which extends conventional VLA approaches by integrating auxiliary learning tasks. This framework strengthens the relationship between visual perception, actions, and language, enhancing contextual understanding and generalization in robotic decision-making.

With a strong foundation in deep learning and multimodal AI, I am committed to advancing AI-driven robotic intelligence. My experience in representation learning, affordance-based reasoning, and robot learning equips me to contribute meaningfully to this field. I have pursued a **-shaped generalist** approach, combining expertise in machine learning with interdisciplinary collaboration to address real-world AI challenges.

As robotics continues to evolve into tightly integrated hardware-software systems, optimizing AI models for real-world performance remains a critical challenge. I aim to develop **efficient and scalable AI techniques** that enable robots to operate reliably in complex environments with limited computational resources. Through this program, I look forward to collaborating with researchers to drive innovations in physical intelligence and AI-powered robotic decision-making.

References

Geon Choi, Ganghun Kim, Hyunjae Kim, Kyuhwan Shim, and Myoungwan Koo. Kirino: An interactive chatbot system for user persona. *Korea Computer Congress*, 2024. URL `https://underthelights.github.io/kirino_page/`.

Gi-Cheon Kang, Junghyun Kim, Kyuhwan Shim, Jun Ki Lee, and Byoung-Tak Zhang. Clip-rt: Learning language-conditioned robotic policies from natural language supervision. *Conference on Robot Learning (CoRL) 2024, LangRob Workshop*, 2024. URL `https://clip-rt.github.io/`.

Junghyun Kim, Gi-Cheon Kang, Kyuhwan Shim, Jun Ki Lee, and Byoung-Tak Zhang. Stochastic trajectory diversification for language-based robot behavior learning. *Korea Software Congress*, 2024. URL `https://clip-rt.github.io/`.