

## • 특집 • 자율제조(Autonomous Manufacturing)

# 6축 로봇 제어를 위한 개방형 시뮬레이션 기반 강화학습 경로 최적화 Path Optimization for 6-axis Robot Control Using Open Simulation-based Reinforcement Learning

김초아<sup>1</sup>, 백종우<sup>1</sup>, 이수한<sup>1</sup>, 이주연<sup>1,\*</sup>Cho A Kim<sup>1</sup>, Jong U Baek<sup>1</sup>, Su Han Lee<sup>1</sup>, and Ju Yeon Lee<sup>1,\*</sup><sup>1</sup> 서울과학기술대학교 기계공학부 (Department of Mechanical Engineering, Seoul National University of Science and Technology)# Corresponding Author / E-mail: [jylee@seoultech.ac.kr](mailto:jylee@seoultech.ac.kr), TEL: +82-2-970-6535

ORCID: 0000-0002-3311-4130

KEYWORDS: Robot path planning (로봇 경로 계획), Reinforcement learning (강화학습), Inverse kinematics (역기구학), Trajectory optimization (궤적 최적화), Industrial robot (산업용 로봇)

*The increasing adoption of industrial robot arms in advanced manufacturing has heightened the need for flexible trajectory planning methods that go beyond traditional offline programming (OLP) tools, which are often expensive, proprietary, and limiting. This study introduces an OLP-free pipeline designed to generate robot trajectory data and optimize paths for six-degree-of-freedom (6-DOF) robot arms using discrete reinforcement learning. Initially, five-axis NC code derived from CAD/CAM data is transformed into tool center point (TCP) trajectories through coordinate transformations. An analytical inverse kinematics solver then produces multiple joint solutions for each TCP pose, creating a discrete action space from which the learning agent can select feasible joint configurations along the trajectory. A reward function that considers variations in joint velocity and acceleration, as well as pose error, facilitates the simultaneous optimization of motion smoothness and tracking accuracy. The optimized trajectories are validated using an open-source physics simulator, showing enhanced motion stability, accuracy, and collision safety compared to conventional OLP-based paths. This proposed framework provides a flexible and cost-effective alternative to commercial OLP tools and lays a scalable foundation for future applications in automated and collaborative manufacturing systems.*

Manuscript received: February 5, 2026 / Revised: March 31, 2026 / Accepted: April 7, 2026

## 1. 서론

최근 산업용 로봇틱스 분야에서는 단순한 평면 작업을 넘어 복잡한 3차원 곡면 형상을 가공하거나 적층하는 기술이 핵심 과제로 부상하였다[1]. 특히 항공·우주와 같은 고부가가치 산업에서 블레이드나 일체형 디스크 등과 같은 복잡한 곡면 형상의 부품에서 정밀 가공의 필요성이 커지고 있다. 이러한 공정은 3, 4축 로봇의 자유도만으로는 곡면의 법선 벡터를 정밀하게 추종하거나 장애물을 회피하는 등의 복잡한 제어에 한계가 존재한다. 따라서 6축 이상의 다관절 로봇 도입이 필수적이다[2].

그러나 로봇의 관절 수가 증가하면서 하나의 직교 좌표계에서 TCP (Tool Center Point)에 대해 로봇이 가능한 자세가 무수히 많아지는 여유 자유도(Redundancy) 문제가 발생한다[3,4]. 이때, 무수히 많은 자세를 선택하여 경로를 생성하는 방식은 최종 제품의 품질을 좌우하는 결정적인 요인이 된다. 실제로 관절 공간에서의 속도 및 가속도 프로파일(Profile)을 연속적으로 보간할 때 로봇의 진동이 유의미하게 감소함이 보고되었다[5]. 또한, 금속 아크 적층 공정인 WAAM (Wire Arc Additive Manufacturing)에서 로봇의 TCP 속도 변동이 심한 구간의 비드(Bead) 높이와 폭이 불균일성을 초래한다는 연구 결과가 보고되었다[6,7]. 즉, 관절 경로의

불연속성은 제조 공정의 정밀도 저하, 장비 마모 가속화, 그리고 사이클 타임 증가를 초래하여 생산성을 저해하는 주요 원인이 된다.

이러한 문제를 해결하기 위해 부드러운 관절 경로를 생성하려는 연구가 지속되어 왔다. 기존 연구 중 일부는 자코비안(Jacobian) 기반 제어나 비선형 근 찾기(Nonlinear Root-finding) 등을 제안하며 수치 해석적 접근을 시도했다[8]. 하지만 이런 수식 기반 접근은 토크 최소화와 정밀한 경로 추종과 같이 상충하는 다중 목적을 동시에 달성하는 데에는 한계를 보였다.

이에 최근에는 복잡한 비선형 환경에서도 시행착오를 통해 공정에 최적인 정책을 스스로 학습하는 강화학습(Reinforcement Learning, RL) 기반의 접근이 주목받고 있다. 강화학습은 공정 목표에 맞춰 보상 함수를 유연하게 설계할 수 있고 용접이나 적층 같은 반복적이면서도 고난도의 최적화가 필요한 공정에 특히 효과적이다[9]. 그러나 강화학습의 특성상 필연적으로 수많은 시행착오를 전제로 하므로 실제 장비에서의 학습은 시간 소모와 파손 위험이 존재한다[10]. 또한, 가상 환경에서 정책을 학습하는 시뮬레이션 기반 기술(Sim-to-Real)의 활용 또한 필수적이다. 하지만 대부분의 상용 로봇팔 시뮬레이터들은 폐쇄형 아키텍처(Closed Architecture)와 독자적인 언어를 사용하고 있기 때문에 강화학습이 요구하는 대규모 반복 학습과 유연한 데이터 교환을 수행하는데 구조적인 제약이 따른다. 실제로 세계 시장 점유율 1위인 ABB나 FANUC와 같은 대표적인 제조사들은 RAPID나 KAREL 등 독자적인 언어와 제어 펄루프(Control Loop)를 사용하므로, 외부 Python 환경에서 저수준(Low-level) 제어 변수에 접근하거나 고속으로 데이터를 통신하는 데 한계가 있다[11,12].

따라서 강화학습을 통해 최종적으로 매끄러운 관절 경로를 산출하기 위해서는 우선 폐쇄적 아키텍처에 의존하지 않고 대규모 반복 학습을 수행할 수 있는 개방형 시뮬레이션 환경 구축이 선행되어야 한다.

본 연구는 이러한 한계를 해결하기 위해 오픈소스 기반 환경에서 경로 데이터 전처리부터 정책 학습에 이르는 전 과정을 유기적으로 통합한 시스템을 제안하고자 한다.

## 2. 관련 연구

### 2.1 OLP 기반 경로 생성의 한계

기존 상용 OLP (Offline Programming) 시스템은 폐쇄적인 구조로 인해 데이터 학습 기반 방법론의 직접적인 적용에 제한이 존재한다. 이러한 제약에도 불구하고 경로 최적화를 달성하기 위한 다양한 접근이 시도되어 왔다. 관련 연구들은 크게 1) 언어 장벽 완화를 목적으로 표준 경로를 제조사 고유 언어로 변환하는 방식과 2) 외부 PC 환경에서 계산된 경로를 미들웨어 또는 브리지 인터페이스를 통해 컨트롤러에 전달하는 방식의 두 범주로 구분된다.

먼저 변환/생성형 접근은 중립 경로 포맷(예: G-code/CL 등)을 기반으로 경로를 구성하고 이를 제조사 제어 언어(예: RAPID, KRL 등)로 자동 변환하는 것을 목표로 삼는다. 이 접근은 상용

OLP 내부 기능에 의존하지 않고 외부에서 생성한 오프라인 경로를 컨트롤러에 적용할 수 있다는 장점이 있다. 기존 연구에서는 RobMach를 통해 CAM 기반 G-code를 ABB RAPID로 변환하는 절차를 제시하여, 가공 궤적을 로봇 프로그램으로 생성·적용하는 과정이 구현되었다[13]. 또한 CAD/CAM 경로로부터 FANUC 로봇 프로그램을 자동 생성하는 Post-processor를 제안하여 실제 로봇 가공 적용 가능성이 보고된 바 있다[14]. 이와 같은 변환/생성형 접근은 경로를 제조사 코드로 변환하고 실행하는 문제에 초점이 놓여 있어 경로 자동화에 효과적이다. 하지만 대다수의 기존 연구는 공정 중 발생하는 환경 변화를 보정하는 데 한계가 있으며, 강화학습과 같은 데이터 기반 방법론을 통한 해결책은 아직 충분히 논의되고 있지 않다.

반면 미들웨어·브릿지형 접근은 외부 PC에서 경로 계획을 계산하고 이를 제조사의 통신 인터페이스로 전달하는 방식으로 구현된다. 기존 연구에서는 OPC UA 기반의 스킬(Semantic Skill) 모델을 제안하여, 상위 시스템이 로봇 동작을 호출하고 통합하도록 지원하는 프레임워크를 제시하였다[15]. 또한 KUKA 컨트롤러와 TCP/IP 기반 인터페이스(Eth.RSIXML)를 이용해 외부 PC에서 로봇의 모션을 구동하는 KCT (KUKA Control Toolbox) 흐름을 제안하고 이를 상용 컨트롤러 외부에서 경로 계획이 가능한 구조로 제시하였다[16]. 이러한 연구들은 외부 계산을 통해 호출한다는 점에서 확장성이 우수하나 통신 성능의 한계, 구현 복잡도, 시스템 통합 비용을 한계점으로 언급한다[15,16].

요약하면, 변환/생성형 접근과 브릿지형 접근은 모두 폐쇄형 생태계를 우회하는 방향으로 발전해 왔으나, 공통적으로 경로 생성 알고리즘의 투명성이 부족하고 강화학습 적용 시 시간 비용과 불안정성이 크다는 한계를 갖는다[17]. 특히 강화학습은 동일 환경에서의 대규모 반복과 데이터 수집, 보상 설계의 반복 수정이 필수적이다. 그러나 기존의 변환/생성형 접근은 학습 과정에 필요한 관측 가능성과 수정의 유연성을 충분히 제공하지 못한다[18]. 결과적으로 기존 연구들은 경로의 형식적인 변환에는 효과적이거나, 공정 환경을 반영하여 경로를 최적화하는 것에는 효과적이지 않다.

### 2.2 개방형 환경에서 학습 기반 경로 최적화 동향

최근에는 OLP에 의존하지 않고 오픈소스 시뮬레이터(예: Gazebo, PyBullet 등)를 기반으로 경로를 생성하고 반복적으로 검증하는 흐름이 강화되고 있다. 특히 PyBullet은 Python 환경에서 물리 시뮬레이션을 구성할 수 있어 강화학습 알고리즘과의 연동이 용이하다는 점에서 활용도가 높다[19].

다만 개방형 시뮬레이션 환경에서 학습된 정책이 곧바로 현실에서의 최적 정책으로 연결되는 것은 아니다. PyBullet 같은 개방형 환경의 대표적인 문제는 현실-시뮬레이션 격차(Sim-to-Real Gap)이다. 이 격차 때문에 시뮬레이션에서 학습된 정책이 현실에서 성능 저하를 보이는 문제는 반복적으로 보고되어 왔으며, 이를 완화하기 위해 Domain Randomization과 같은 방법이 제안되어 왔다[20]. 또한, 시뮬레이터는 충돌이나 속도 제한 같은 기계적 제약은 쉽게 설정할 수 있지만, 공정 품질과 직결되는

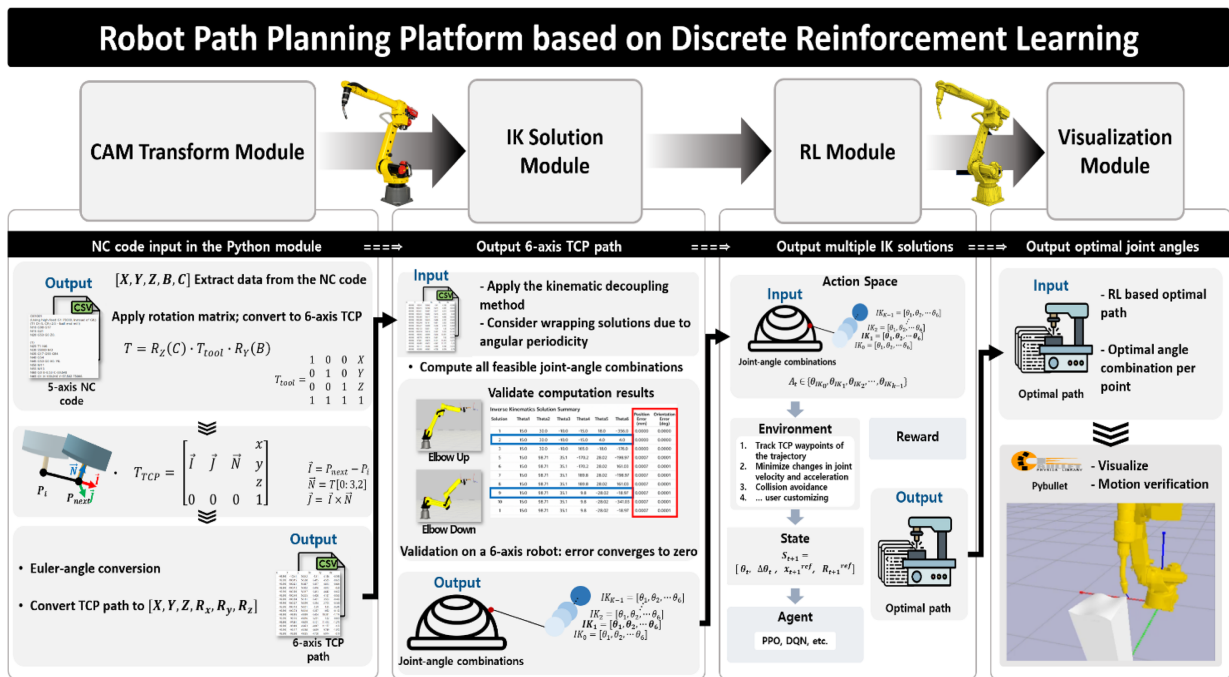


Fig. 1 System concept

복잡한 제약(예: 용접/적층 품질, 열 누적, 비드 형상)은 파라미터 튜닝 부담이 증가한다[21]. 따라서 학습 안정성과 실제 공정 적합성을 확보하기 위해서는 환경 구성과 학습 설계를 함께 고려해야 한다. 그 예시로, PyBullet 환경에서 DDPG (Deep Deterministic Policy Gradient) 심층강화학습 알고리즘으로 로봇 팔이 다양한 목표에 따라 움직일 수 있게 상태변수를 조정해 보상함수와 리플레이 버퍼(Replay Buffer)를 통합한 연구가 있다 [22]. 해당 연구에서는 복잡한 수동 파라미터 튜닝과 모델에 과도하게 의존하지 않는 적응성 있는 제어를 검증했다. 마찬가지로 로봇팔이 최소 시간, 최단 경로, 가속도 같은 다중 조건을 만족할 수 있도록 최적화 기반 방법인 PSO-DE를 PyBullet 환경에서 적용해 검증한 연구가 수행되었다[23].

이처럼 강화학습을 활용한 로봇 경로 최적화 연구가 개방형 환경에서도 활발히 수행되고 있지만 이러한 학습 기반 연구에서도 역기구학(Inverse Kinematics, IK) 다중해를 경로 전체 시퀀스 관점에서 다루는 사례는 상대적으로 제한적이다. 예를 들어 DQN을 이용해 고자유도 조작기의 IK 문제를 강화학습으로 푸는 접근을 제시한 연구는 있지만, 이는 주로 단일 자세를 목표로 하며 경로 전체에서의 관절 시퀀스 최적화와는 다르다[8]. 따라서 다중 IK 해를 경로 전체 시퀀스 관점에서 고려한 경로 최적화는 개방형 환경에서 여전히 체계적인 방법론이 요구된다.

결과적으로, 기존의 개방형 시뮬레이션 기반 연구들은 주로 연속 제어 입력의 학습, 단일 목표 자세 도달, 또는 일반적인 경로 추종 문제에 초점을 두고 있다. 그러나 실제 6축 산업용 로봇의 공정 경로에서는 각 웨이포인트마다 복수의 역기구학 해가 존재하며, 경로 품질은 개별 시점의 해 선택보다 전 구간을 걸친 관절 시퀀스의 연속성에 의해 크게 좌우된다. 이러한 점에서 본

연구는 개방형 환경을 사용하는 데 그치지 않고, 상용 OLP에 의존하지 않는 개방형 시뮬레이션 환경에서 이산 행동공간 강화 학습으로 경로 전체에서 속도 및 가속도 변화가 완만한 관절 시퀀스를 안정적으로 도출하는 것을 목표로 한다. 이를 위해 CAM 데이터로부터 로봇 TCP 경로를 구성하고, 웨이포인트 (WayPoint)별로 IK 다중해 후보 집합을 생성한 뒤, 이러한 후보의 선택과정을 강화학습 행동공간으로 지정한다. 이러한 접근은 3장에서 제시하는 통합 파이프라인 설계로 구체화된다.

### 3. 이산 공간 강화학습 기반 경로 설계 시스템

Fig. 1에 제시된 제안 시스템은 총 4단계의 파이프라인으로 구성된다. 첫째, CAM 경로 변환 모듈에서는 상용 소프트웨어에서 추출된 5축 NC 데이터를 6축 로봇의 제어에 적합한 TCP 위치 및 자세 좌표계로 변환한다. 둘째, IK 다중해 생성 모듈은 변환된 각 웨이포인트에 대해 해석적 기법으로 로봇팔의 다중해 집합을 산출한다. 셋째, 이산 공간 강화학습 최적화 모듈은 생성된 다중해 집합을 행동 공간으로 삼아 관절의 속도, 가속도와 궤적 오차를 최소화하는 경로를 산출한다. 마지막으로, 시뮬레이션 기반 검증 모듈에서는 학습된 정책을 가상 환경에 입력해 충돌 여부 및 동역학적 궤적 추종성을 평가한다. 각 모듈은 기능 확장 및 실험 재현이 용이하도록 입·출력 형식을 통일하고 모듈 간의 의존성이 최소화되는 구조로 설계하였다. 이러한 모듈들은 유기적으로 연동되어, 원시 CAM 데이터 입력부터 최종 검증까지 연속적인 구조를 구성한다. 본 절에서는 각 모듈의 역할에 대한 설명과 모듈에서 수행되는 구체적인 실행 절차를 기술한다.

Table 1 Input values for the CAM transformation module

|     |                        |
|-----|------------------------|
| $X$ | X-axis linear position |
| $Y$ | Y-axis linear position |
| $Z$ | Z-axis linear position |
| $B$ | Rotation around Y-axis |
| $C$ | Rotation around Z-axis |

Table 2 Output values for the CAM transformation module

|       |                         |
|-------|-------------------------|
| $X$   | X-axis displacement     |
| $Y$   | Y-axis displacement     |
| $Z$   | Z-axis displacement     |
| $R_x$ | Roll (X-axis rotation)  |
| $R_y$ | Pitch (Y-axis rotation) |
| $R_z$ | Yaw (Z-axis rotation)   |

### 3.1 CAM 경로 변환 모듈(CAM Transform Module)

본 모듈은 상용 CAD/CAM 소프트웨어와 로봇 제어 환경 간 데이터 비정합성(Non-compatibility)을 해소하는 전처리를 수행한다. 일반적으로 CAM 시스템에서 생성된 5축 NC (Numerical Control) 데이터는 공작 기계 기준의 좌표계를 따르므로, 이를 6축 산업용 로봇이 수행 가능한 TCP 위치 및 자세 정보로 변환하는 과정이 필수적이다[24]. 이를 위해 Liu 외(2025)가 제안한 기하학적 변환 알고리즘을 기반으로 Python 기반의 변환 모듈을 구현하였다[25].

본 모듈의 핵심 원리는 Table 1에서 제시된 끝점 위치와 회전축의 방향 벡터를 추출하여, 로봇의 베이스 좌표계에 맞는  $4 \times 4$  동차 변환 행렬(Homogeneous Transformation Matrix)로 재구성하는 것이다. 이 과정에서 NC 좌표계에서 정의된 선형 이동 및 회전 정보는 기하학적 변환을 통해 로봇 제어에 적합한 자세 표현으로 변환되며, 그 결과는 Table 2에 제시된 위치( $X, Y, Z$ ) 및 자세( $R_x, R_y, R_z$ ) 출력 변수로 표현된다. 이를 통해 본 모듈은 상용 OLP 없이도 복잡한 3차원 곡면 경로를 6축 로봇 제어에 적합한 경로 데이터로 자동 변환할 수 있다.

$$T_{tool,i} = \begin{bmatrix} 1 & 0 & 0 & x_i \\ 0 & 1 & 0 & y_i \\ 0 & 0 & 1 & z_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$T_{BC,i} = R_z(C_i) \cdot T_{tool,i} \cdot R_y(B_i) \quad (2)$$

$$T_{C,i} = T_{BC,i} \cdot R_y(\pi) \quad (3)$$

실행 절차는 다음과 같다. 식(1)-식(3)에 따라, 회전축 정보( $B, C$ )를 각각 Y축과 Z축을 기준으로 한 회전행렬  $R_y(B)$ ,  $R_z(C)$ 로 변환한 후, 강체 회전 이론을 따라 순차적으로 합성하여 초기

Table 3 CAM transformation &amp; geometric definitions (Eqs. 1-4)

|                 |                                                                |
|-----------------|----------------------------------------------------------------|
| $T_{tool,i}$    | Initial tool pose matrix of the $i$ -th waypoint               |
| $x_i, y_i, z_i$ | Tool tip coordinates extracted from NC code                    |
| $R_z(C_i)$      | Rotation matrix around the Z-axis (reflecting C-axis rotation) |
| $R_y(B_i)$      | Rotation matrix around the Y-axis (reflecting B-axis rotation) |
| $T_{BC,i}$      | Combined tool pose matrix with B and C rotations               |
| $T_{TCP,i}$     | Final calculated TCP transformation matrix                     |

공구 자세 행렬을 도출한다.

$$T_{TCP,i} = \begin{bmatrix} x_i \\ \vec{I}_i \vec{J}_i \vec{N}_i & y_i \\ z_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

이후 식(4)를 이용해 현재 웨이포인트에 대한 정보를 얻은 후, 다음 웨이포인트와의 위치 및 자세 변화를 이용해 경로의 진행 방향 벡터  $\vec{I}$ 를 계산하고, 공구축을 기준으로 법선 벡터  $\vec{N}$ 를 정의한다. 진행 방향 벡터와 법선 벡터의 외적으로 중법선 벡터  $\vec{J}$ 를 계산하고 정규화하여 웨이포인트마다 동일한 TCP 좌표계를 구성한다. 최종적으로 공구가 곡면 형상을 따라 수직 방향을 유지하면서도 연속적인 움직임을 갖는 6축 로봇 경로를 생성한다.

### 3.2 IK 다중해 생성 모듈(IK Solution Module)

본 모듈은 강화학습 에이전트가 탐색할 수 있는 최적의 행동 공간을 확보하기 위해, 목표 TCP에 대한 가능한 모든 IK 다중해를 산출하는 역할을 수행한다. 일반적인 6축 로봇은 기구학적 특성상 같은 TCP 위치와 자세에 대해 최대 8가지의 서로 다른 관절 조합이 존재한다. 그러나 본 연구에서는 해석적 IK로 얻은 기본 해에 대해 회전 관절의 주기성에 따른 등가 표현(예:  $\theta$ 와  $\theta \pm 360^\circ$ )이 추가로 허용될 수 있다. 그 결과, 기구학적 유효성과 관절 한계(Joint Limits)를 통과한 후보 해의 개수는 평균 14개로 확장되었다. 반면, 대부분의 기존 상용 플랫폼은 단일해만 제공하여 경로 최적화의 유연성을 제한하는 한계가 있다. 따라서 본 모듈은 강화학습 에이전트가 탐색할 수 있는 충분한 행동 공간(Action Space)을 확보하기 위해, 가능한 모든 IK 해를 산출하는 다중해 생성 역할을 담당한다. 이때 정확도에서 한계를 보이는 수치 해석적 근사법 대신, 로봇의 기하학적 구조를 활용한 해석적 기법을 적용하였다. 특히 손목 중심점(Wrist Center)과 말단 자세(Orientation)를 분리해 계산하는 기구학적 디커플링(Kinematic Decoupling) 원리를 통해 연산 속도와 해의 정확성을 동시에 확보한다[26]. 이를 통해 관절 공간 전체에서 최적의 거동을 선택할 수 있는 이산적인 후보 집합을 제공해 후속 강화학습 모듈의 핵심 입력 데이터로 활용된다. 해당 모듈에서 다중해를 산출하는 구체적인 절차는 다음과 같다.

### 3.2.1 손목 중심점(Wrist Center) 계산

TCP 위치  $x_p, y_p, z_p$ 와 말단 링크 길이  $d_6$ 를 이용하여 손목 관절의 중심 좌표인  $P_{wc}$ 를 역산한다. 이 과정은 6축 로봇의 말단 자세와 위치를 분리하여, 팔(Arm)에 해당하는 1-3축의 위치를 우선 결정하기 위한 단계이다.

$$P_{wc} = T_{06} \cdot \begin{bmatrix} 0 \\ 0 \\ -d_6 \\ 1 \end{bmatrix} = \begin{bmatrix} x_p - d_6 \cdot r_{13} \\ y_p - d_6 \cdot r_{23} \\ z_p - d_6 \cdot r_{33} \\ 1 \end{bmatrix} \quad (5)$$

### 3.2.2 관절 다중해 조합

기하학적 위상에 따라 발생하는 Shoulder Left/Right, Elbow Up/Down, Wrist Flip/Non-flip의 8가지 경우의 수를 조합하여 이론적으로 가능한 모든 해를 도출한다. 이는 동일한 TCP 위치에 도달할 수 있는 다양한 로봇의 자세군을 생성한다.

### 3.2.3 유효성 필터링

도출된 후보 해 중 로봇의 물리적 관절 한계를 넘거나 특이점(Singularity)에 근접한 해를 배제하여, 후속 강화학습 단계의 탐색 효율을 극대화한다.

### 3.3 이산 공간 강화학습 기반 최적 경로 선택 모듈(RL Module)

본 모듈은 웨이포인트마다 IK 다중해 후보 집합을 생성하고 관절 각도의 연속성을 극대화해 최종적으로는 최적 경로를 결정하는 역할을 수행한다. 일반적인 로봇 제어에서는 연속 공간(Continuous Space) 기반 강화학습이 사용되고 있으나, 탐색 범위가 방대하여 학습 수렴 속도가 느리고 하이퍼 파라미터(Hyperparameter) 변화에 민감해 불안정성 문제가 빈번히 발생한다는 한계가 존재한다. 본 연구에서는 이러한 한계를 극복하기 위해 문제를 다중해 집합 내에서 최적 해 조합 선택 문제(Optimal Solution Selection Problem)로 재정의하고, 이산 공간 강화학습(Discrete Reinforcement Learning) 기반의 알고리즘을 적용하였다[27]. 학습 메커니즘을 다음과 같은 마르코프 결정 과정(Markov Decision Process, MDP)를 기반으로 서술할 수 있다.

#### 3.3.1 행동 공간(Action Space, $\alpha$ )

행동 공간은 현재 타임스텝  $t$ 에서 선택 가능한  $k$ 개의 관절 각도 조합으로 정의된다. 에이전트는 후보 집합 내에서 하나의 해를 선택함으로써 로봇의 구동 형태(Configuration)를 결정한다.

$$A_t \in \{\theta_{ik0}, \theta_{ik1}, \theta_{ik2}, \dots, \theta_{ik_{k-1}}\} \quad (6)$$

#### 3.3.2 상태 공간(State Space, $s$ )

선택된 IK 해를 바탕으로 환경은 다음 상태  $S_{t+1}$ 로 전이된다. 상태 벡터에는 현재의 관절 각도  $\theta$ 와 각속도  $\Delta\theta$ 가 포함되어 경로의 연속성을 반영하며, 에이전트가 목표 경로를 추종할 수

#### Algorithm 1 Masked PPO for Optimal IK-Solution Sequencing

**Input:** Path CSV, DH Parameters,  $\Pi_{PPO}, V_\psi, N_{updates}, H$   
**Output:** Optimized joint trajectory  $\Theta^* = (\theta_1, \theta_2, \dots, \theta_T)$

```

1: // Phase 0: Initialization
2: Initialize  $\Pi_{PPO}$  and  $V_\psi$  with random weights
3: Set initial joint history  $\theta_{t-1} \leftarrow 0, \theta_{t-2} \leftarrow 0$ 
4: for update = 1 to  $N_{updates}$  do
5:   // Phase 1: Experience Collection with IK-Masking
6:    $D \leftarrow \emptyset, (s, info) \leftarrow Env.reset()$ 
7:   for  $t = 1$  to  $H$  do
8:      $m \leftarrow info.action\_mask$  ▷ Get valid IK indices
9:      $logits \leftarrow \Pi_{PPO}(s)$ 
10:     $logits[m == False] \leftarrow -\infty$  ▷ Mask invalid IK solutions
11:     $a_t \sim \text{Categorical}(\text{softmax}(logits))$ 
12:     $(s', r_t, done, trunc, info') \leftarrow Env.step(a_t)$ 
13:    Store  $(s, a_t, r_t, V_\psi(s), \log \Pi_{PPO}(a_t|s))$  in  $D$ 
14:    if  $done$  or  $trunc$  then
15:      break
16:    end if
17:     $s \leftarrow s'$ , Update joint history
18:  end for
19: // Phase 2: Network Update
20: Compute advantages  $\hat{A}_t$  and returns  $R_t$ 
21: Update  $\Pi_{PPO}$  by maximizing clipped surrogate objective
22: Update  $V_\psi$  by minimizing value loss
23: end for
24: // Phase 3: Final Trajectory Generation (Inference)
25:  $s \leftarrow Env.reset(), \Theta^* \leftarrow \emptyset$ 
26: while not done do
27:    $m \leftarrow Env.get\_action\_mask()$ 
28:    $a^* = \text{argmax}_{a:m_a=True} \Pi_{PPO}(a|s)$ 
29:    $(s, \theta_t, \dots) \leftarrow Env.step(a^*)$ 
30:   Append current joints  $\theta_t$  to  $\Theta^*$ 
31: end while
32: return  $\Theta^*$ 

```

Fig. 2 Discrete reinforcement learning algorithm

있도록 다음 포인트의 위치 정보  $x_{t+1}^{ref}$  및 툴 자세  $R_{t+1}^{ref}$ 를 함께 입력 받아 구성된다.

$$\theta_{IK} \rightarrow S_{t+1} = [\theta_t, \Delta\theta_t, x_{t+1}^{ref}, R_{t+1}^{ref}] \quad (7)$$

#### 3.3.3 보상 함수(Reward Function, $r$ )

보상 함수  $r_t$ 는 관절의 동적 특성과 기하학적 정밀도를 모두 고려하여 설계되었다. 각 항은 음의 보상을 부여함으로써 에이전트가 페널티를 최소화하도록 학습을 유도한다.

$$r_t = - \left( \omega_1 \cdot \frac{\|\theta_t - \theta_{t-1}\|^2}{v_{max}^2} + \omega_2 \cdot \frac{\|\Delta\theta_{t-1} - \Delta\theta_{t-2}\|^2}{a_{max}^2} + \omega_3 \cdot \frac{\|x_t^{reals} - x_t^{ref}\|^2}{\delta_{pos}} + \omega_4 \cdot \frac{\|log(R_t^{real} R_t^{ref})\|}{\delta_{ori}} \right) \quad (8)$$

강화학습의 전체 절차는 Fig. 2를 통해 제시하였으며, 학습에 사용된 주요 파라미터는 Table 4에 정리하였다. 본 연구에서는 PPO (Proximal Policy Optimization) 모델을 사용하여 학습을 수행했으며, 이를 통해 제한된 행동 공간 내에서 안정적인 정책 학습이 가능하도록 하였다. 최종적으로는 식(9)와 같이 감쇠율을 적용하며 기대 누적 보상을 최대화한 정책을 학습하도록 하였다.

$$\pi^*(s) = \text{argmax} E[\sum_{t=0}^T \gamma^t \cdot r_t] \quad (9)$$

Table 4 Reinforcement learning parameters

|                |                                             |                     |
|----------------|---------------------------------------------|---------------------|
| $\omega_1$     | Joint velocity smoothness weight            | 5                   |
| $\omega_2$     | Joint acceleration (Jerk) smoothness weight | 1                   |
| $\omega_3$     | TCP position tracking error weight          | 100                 |
| $\omega_4$     | TCP orientation tracking error weight       | 50                  |
| $v_{max}^2$    | Squared mean joint velocity                 | 342.5 <sup>2</sup>  |
| $a_{max}^2$    | Squared mean joint acceleration             | 2110.0 <sup>2</sup> |
| $\delta_{pos}$ | Mean position difference                    | 5.0                 |
| $\delta_{ori}$ | Mean orientation difference                 | 2.0                 |
| $x_t^{real}$   | Actual TCP position                         | Variable            |
| $x_t^{ref}$    | Reference TCP position                      | Variable            |
| $R_t^{real}$   | Actual TCP orientation                      | Variable            |
| $R_t^{ref}$    | Reference TCP orientation                   | Variable            |

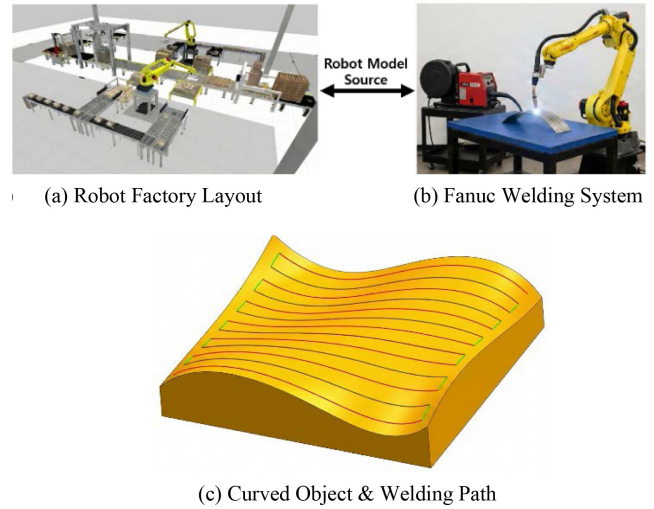


Fig. 3 Simulation environment for the robotic system

### 3.4 시뮬레이션 기반 검증 모듈(Visualization Module)

본 모듈은 강화학습을 통해 도출된 최적 경로 정책의 안전성과 실제 구동 가능성을 검증하기 위해, 물리 엔진 기반의 시뮬레이션을 구축하고 정량적 평가를 수행한다. 오픈 소스 물리 시뮬레이터인 PyBullet을 활용하여 실제 로봇의 동역학적 특성과 작업 환경을 시뮬레이션에 구현하였으며, 학습된 정책이 출력한 관절 시퀀스를 입력하여 검증을 수행한다. 시뮬레이터는 실제 동역학 파라미터를 기반으로 물리 연산을 수행함으로써 충돌 여부, 사이클 타임, 최대 관절 속도 등의 성능 지표를 정량적으로 평가한다. 이러한 평가 결과는 학습 과정에서의 피드백으로 활용되거나, 최종 공정 투입 가능성을 판단하는 근거로 사용된다. 이를 통해 관절의 거동의 연속성이나 장애물 회피 등 다양한 공정 조건에서의 제어 성능을 시각적으로 확인할 수 있으며, 실제 로봇 시스템에서 발생할 수 있는 시행착오 비용을 절감할 수 있다. 아울러 학습 및 평가 과정의 주요 지표는 로깅(Logging) 및 모니터링 시스템을 통해 저장되며, 스텝 단위 성능 지표는 시뮬레이터에서 계산되어 후처리 가능하도록 구성하였다.

## 4. 실험 설계 및 적용

본 장에서는 제안한 통합 프레임워크를 실제 공정에 적용하기 위한 실험 환경을 정의하고, 초기 CAM 데이터가 강화학습에 이진트의 입력으로 변환되어 시뮬레이션에 전달되기까지의 구체적인 데이터 흐름을 기술한다.

### 4.1 대상 공정 및 실험 환경

Fig. 3에 제시된 바와 같이, 본 연구의 대상 공정은 금속 와이어를 아크 용접으로 적층하는 WAAM 공정이다. 이 공정은 용접 토치가 적층면에 대해 지속적으로 수직(Normal Vector)을 유지해야 균일한 비드 품질을 확보할 수 있으므로, 단순한 위치 추종을 넘어서 고도화된 자세 제어가 필수적이다.

따라서 실험 대상은 가로 200 mm, 세로 200 mm 크기의 물결(Wave) 형태 곡면으로 설정하였다. 이 형상은 단순 평면 대비 기하학적 복잡성이 높다. 따라서 TCP가 곡면의 법선 벡터를 정확히 추종하는지, 급격한 곡률 변화 구간에서 관절 움직임이 부드러운지를 평가하기에 용이하다.

실험에 사용된 로봇 모델은 소형 아크 용접 및 적층 공정에 널리 사용되는 FANUC M-20iA (6축 산업용 로봇)이다. 본 실험에서는 PyBullet 기반의 물리 엔진 상에 해당 로봇의 URDF (Unified Robot Description Format)와 기구학적 파라미터, 관절 운동 범위를 정밀하게 묘사한 PyBullet 시뮬레이션 환경을 구축하여 검증을 수행하였다.

### 4.2 이산 공간 강화학습 기반 경로 설계 시스템 적용 사례

제안된 시스템은 각 모듈 단계별로 데이터를 순차적으로 가공하여 최적의 제어 입력을 생성한다. 각 단계의 상세 적용 내역은 다음과 같다.

#### 4.2.1 경로 변환 모듈 적용

본 절에서는 3.1절에서 설명한 경로 변환 모듈을 6축 로봇을 사용하는 WAAM 공정 형상에 적용하는 과정을 설명한다. 물결 곡면을 가지는 3D 형상에 대해 Fusion 360 CAM 기능을 사용하여 5축 NC 코드 형태의 경로를 생성하였다. 생성된 G-code에는 공구 위치 X, Y, Z, 회전축 B, C 및 이송 속도 등 다양한 정보가 포함되며, 본 연구에서는 경로 변환에 필요한 X, Y, Z, B, C 데이터만을 추출하여 CSV 형식으로 정리하였다. 위에서 제시한 변환 모듈을 통해 X, Y, Z, Rx, Ry, Rz 데이터에 대한 경로 CSV 파일을 생성 및 확보하였다. 경로 변환으로 생성된 TCP 경로는 총 4,320개의 웨이포인트를 가지며, 로봇 베이스를 기준으로 생성되었다.

#### 4.2.2 IK 다중해 모듈 적용

생성된 TCP 경로는 IK 다중해 모듈의 입력으로 사용된다.

Table 5 Application results of the inverse kinematics module

| [°] | 01    | 02    | 03    | 04     | 05     | 06     |
|-----|-------|-------|-------|--------|--------|--------|
| 1   | 95.08 | 9.70  | 0.94  | -179.0 | 77.5   | -184.5 |
| 2   | 95.08 | 9.70  | 0.94  | -179.0 | 77.5   | 175.5  |
| 3   | 95.08 | 9.70  | 0.94  | 181.0  | 77.5   | -184.5 |
| 4   | 95.08 | 9.70  | 0.94  | 181.0  | 77.5   | 175.5  |
| 5   | 95.08 | 9.70  | 0.94  | 1.01   | -77.5  | -364.5 |
| 6   | 95.08 | 9.70  | 0.94  | 1.01   | -77.5  | -4.54  |
| 7   | 95.08 | 9.70  | 0.94  | -178.6 | -77.5  | 355.5  |
| 8   | 95.08 | 75.82 | 60.21 | -178.6 | 136.8  | -183.3 |
| 9   | 95.08 | 75.82 | 60.21 | 181.4  | 136.8  | 176.7  |
| 10  | 95.08 | 75.82 | 60.21 | 181.4  | 136.8  | -183.3 |
| 11  | 95.08 | 75.82 | 60.21 | 1.01   | 136.8  | 176.7  |
| 12  | 95.08 | 75.82 | 60.21 | 1.01   | -136.8 | -363.3 |
| 13  | 95.08 | 75.82 | 60.21 | 1.5    | -136.8 | -3.27  |
| 14  | 95.08 | 75.82 | 60.21 | 1.5    | -136.8 | 356.7  |

각 웨이포인트에 대해 해석적 IK를 적용하여 로봇이 취할 수 있는 다중 관절각 해를 모두 산출하였고, 웨이포인트별 후보군을 CSV로 저장하였다. 본 실험 조건에서 웨이포인트당 후보 해의 개수는 평균 14개 수준으로 생성되었다. Table 5는 첫 번째 웨이포인트에 대해 생성된 다중 해의 예시를 보여준다.

#### 4.2.3 이산 공간 강화학습 모듈 적용

3.3절에서 제안한 이산 공간 강화학습 기반의 경로 최적화 기능을 적용하였다. 강화학습은 구축된 데이터 파이프라인 위에서 PPO 알고리즘을 사용하여 최적 경로 선택 정책을 학습하도록 설계하였다. 에이전트는 매 스텝마다 CSV 파일에 저장된 약 14개의 IK 후보군 중 하나를 행동으로 선택한다. 이때 웨이포인트별로 실제 유효 후보의 수가 서로 다를 수 있으므로, 본 연구에서는 행동공간의 크기를 고정된 뒤 유효하지 않은 후보가 선택되지 않도록 마스킹하는 방식으로 처리하였다. 따라서 에이전트는 각 스텝에서 실제로 존재하는 IK 후보들에 대해서만 선택을 수행하도록 구성되었다. 선택된 관절 각도는 PyBullet 시뮬레이터에 전달되며, 시뮬레이터는 즉시 충돌 여부와 경로 이탈 여부를 확인한다. 이 과정은 에피소드 당 4,320 스텝 동안 반복되며, 에이전트는 시행착오를 통해 누적 보상을 최대화하는 방향으로 정책을 업데이트한다. 학습이 완료되면 최종적으로 선택된 최종 관절 시퀀스를 CSV 형태의 데이터셋으로 저장한다. 제안 방식은 후보 해 집합만 교체하면 다른 CAD 형상 및 공정 경로에도 동일한 학습 파이프라인을 적용할 수 있어 확장성이 높다.

#### 4.2.4 시뮬레이션 모듈 적용

3.4절에서 도출된 최적 정책의 실제 구동 가능성을 검증하기 위해 PyBullet 기반 디지털 트윈에서 WAAM 공정용 용접 토치를 장착한 FANUC M-20iA 모델로 시뮬레이션 기반의 가시화 및

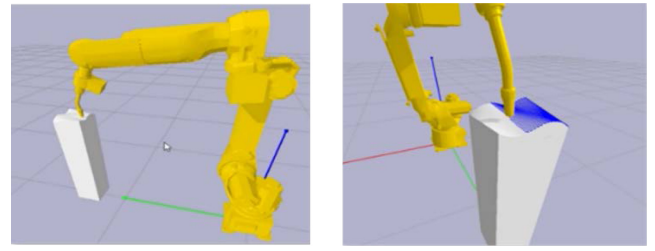


Fig. 4 Path tracking in PyBullet simulation

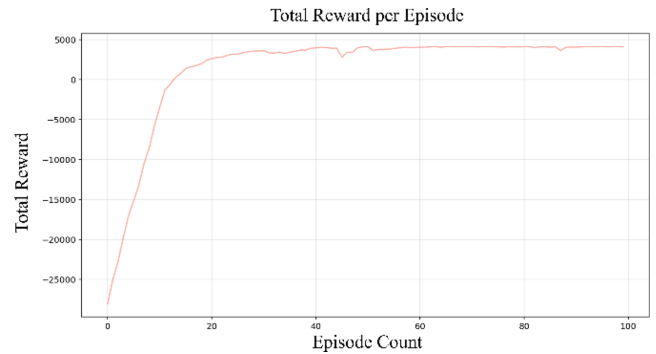


Fig. 5 Learning curve: Average cumulative reward per episode

유효성 검증을 수행하였다. 강화학습 모듈이 출력한 최종 결과물인 관절 시퀀스 CSV를 불러온 뒤, Fig. 4와 같이 곡면형상 CAD가 포함된 가상 셀에서 TCP가 공정 기준 경로를 추종하도록 구동하였다. 이때 공정 품질에 직접 영향을 주는 지표로서 TCP 위치 오차 및 자세 오차 기준을 적용하여 구간별 추종 성공여부를 판정하였다. 또한 속도/가속도 패널티 기반 정책을 통해 용접 중 공구 자세 및 동작의 연속성이 유지되는지 확인하였다. 이러한 시뮬레이션 연동이 실제 셀 적용 시 간섭 위험이 없는 경로인지도 검증하기 위해 시뮬레이터 내부의 충돌 감지 엔진(Collision Detection Engine)을 활성화해 주행 중 로봇 링크 간 자가 간섭과 토치와 비드 간 충돌을 실시간으로 모니터링하여 경로의 안정성을 확보하였다.

## 5 적용 및 검증 결과

### 5.1 강화학습 알고리즘 학습 안정성

Fig. 5의 가로축은 학습의 반복 횟수인 에피소드 수(Episode Count)를 나타내며 세로축은 각 에피소드에서 획득한 누적 보상(Total Reward)를 의미한다. 학습 초기 이산 공간 RL 모델은 학습 초기(0-20에피소드) 약  $-2.5 \times 10^4$ 의 낮은 보상 값에서 시작하여 가파르게 상승하는 양상을 보인다. 이후 약 40 에피소드 부근에서 약  $4.0 \times 10^3$  수준에 도달한 후 큰 변동 없이 수렴하는 등 높은 학습 안정성을 확인할 수 있다.

이는 수백만 번의 반복 학습이 필요한 연속 공간 탐색 방식과 비교했을 때, 기구학적으로 유효한 해 집합 내에서만 최적의

Table 6 Comparison of discrete RL and VC baseline

| Evaluation metric                                  | Discrete RL | Visual component |
|----------------------------------------------------|-------------|------------------|
| Position error (MAE) [mm]                          | 0.00135     | 0.00135          |
| Orientation error (MAE) [rad]                      | 0.128       | 0.128            |
| Angular velocity RMSE [rad/step]                   | 1.5147      | 1.5171           |
| Angular acceleration RMSE [rad/step <sup>2</sup> ] | 1.0723      | 1.0785           |

조합을 선택하도록 문제를 단순화했으며 강화학습의 탐색 효율성과 수렴 속도를 획기적으로 개선했음을 보여준다[28]. 결과적으로 학습된 정책은 전체 4,320개의 웨이포인트에 대해 기구학적 불연속성 없이 매끄러운 경로를 생성하는 데 성공하였다.

## 5.2 경로 추종 정밀도 평가

학습이 완료된 에이전트가 생성한 최종 경로의 품질을 위치·자세 오차와 관절 움직임의 부드러움 측면에서 평가하였다. 또한, 본 연구에서는 제안 방법의 결과를 비교하기 위하여 Visual Components (VC) 환경에서 동일 경로에 대해 포인트별 역기구학 해를 산출하여 생성한 기준 경로를 Baseline으로 사용하였다 (Table 6 참조).

정량적 평가 결과, 제안하는 시스템은 0.00135 mm의 위치 오차와 0.128 rad의 자세 오차를 기록하여 두 방법에서 동일하게 나타났으며, 두 경로가 동일한 기준에서 유사한 TCP 재현성을 갖는다는 점을 보여준다. 반면 경로의 부드러움을 나타내는 각 속도 및 각가속도의 RMSE (Root Mean Square Error)가 제안 플랫폼에서 각각 1.5147 rad/step, 1.0723 rad/step<sup>2</sup>로 VC 기반 Baseline보다 소폭 낮게 측정되었다. 이는 로봇이 급격한 감속이나 떨림 없이 매끄럽게 거동함을 의미하며, 결과적으로 실제 공정 투입 시 장비의 기계적 부하를 줄이고 적층 품질을 균일하게 유지할 수 있음을 시사한다. 다만 이러한 비교는 시뮬레이션 기반 Baseline에 대한 결과이므로, 실제 산업용 OLP와의 직접 비교로 일반화하기에는 한계가 있다.

## 5.3 시뮬레이션 기반 공정 가능성 검증

PyBullet 기반의 가상 공정 시뮬레이션을 통해 다음의 세 가지 요소를 최종 검증하였다.

### 5.3.1 경로 재현 및 추종성(Path Tracking)

학습된 모델이 생성한 경로는 시뮬레이터 상에서 끊임 없이 정상적으로 재현되었다. 특히 WAAM 공정의 핵심 요구사항인 TCP의 자세 제어 측면에서, 로봇의 툴 끝단이 물결 모양의 복잡한 곡면 형상에 대해 지속적으로 수직을 유지하며 안정적으로 경로를 추종함을 확인하였다. 이는 앞선 3.1절의 경로 변환 정밀도가 실제 구동 단계까지 유효하게 유지됨을 시각적으로 입증한다.

### 5.3.2 동작의 부드러움(Visual Smoothness)

시각적 모니터링 결과, 전체 경로 주행 과정에서 로봇 관절의 급격한 떨림이나 불연속적인 자세 변경은 관찰되지 않았다. 이는 4절에서 도출된 낮은 각가속도 RMSE 수치가 실제 동작에서도 부드러운 거동으로 발현됨을 의미하며, 이산 공간 기반의 해 선택 방식이 기구학적으로 안정된 경로를 생성함을 보여준다.

### 5.3.3 기구학적 안전성 및 충돌 여부(Kinematic Safety and Collision Avoidance)

시뮬레이션 전 구간에서 로봇 링크 간의 자가 간섭(Self-collision)이나 작업물 및 주변 환경과의 충돌은 전혀 발생하지 않았다. 이는 다중 IK 생성 단계에서의 유효성 필터링과 강화학습 에이전트의 제약 조건 학습이 성공적으로 수행되었음을 보여준다.

결론적으로 PyBullet 시뮬레이션 결과, 제안하는 시스템은 경로 재현성, 자세 유지, 그리고 기구학적 충돌 회피 측면에서 유의미한 결과를 보였다.

## 6. 결론

본 연구에서는 기존 상용 OLP 소프트웨어의 폐쇄적인 환경을 극복하고 Python 기반의 로봇 경로 생성 및 강화학습 프레임워크를 제안하였다. 이 시스템은 공정 요구사항에 유연하게 대응할 수 있는 경로 계획 생성을 목표로 5축 NC 데이터를 6축 로봇 경로로 변환하고, 다중 IK 해를 산출하여 이산 공간 강화학습을 통해 경로를 최적화하는 통합 파이프라인을 구축하였다. 또한 곡면을 포함하는 WAAM 공정 시뮬레이션을 통해 제안 시스템의 효용성을 입증하였다. 실험 결과, 이산 공간 강화학습 모델이 기구학적으로 유효한 해 내에서 탐색을 수행함으로써 학습의 수렴 안정성을 확보하였으며, 위치·자세 정밀도 유지와 동시에 동적 안정성을 확보함을 확인하였다.

본 연구의 핵심 의의는 로봇의 IK 다중해를 강화학습의 행동 공간으로 정의함으로써, 경로 추종과 구동의 부드러움이라는 복합적인 조건을 동시에 만족시켰다는 점이다. 또한, 전체 파이프라인을 개방형 코드로 구현하여, 연구자가 목적에 맞춰 보상 함수나 최적화 로직을 자유롭게 재설계할 수 있는 범용성 및 확장성을 확보하였다.

다만 본 연구는 다음과 같은 한계를 가진다. 첫째, 제안한 프레임워크의 검증은 PyBullet 기반 시뮬레이션 환경에서 수행되었으며, 실제 로봇 제어기, 통신 지연, 센서 오차, 기구 오차 및 공정 중 외란을 포함한 하드웨어 조건은 반영하지 못하였다. 둘째, 실험은 WAAM 공정의 단일 곡면 형상과 특정 로봇 모델을 대상으로 수행되어, 다른 공정 조건이나 로봇 구조에 대한 일반화 가능성은 추가 검증이 필요하다. 셋째, 본 연구는 경로 생성 및 기구학적 추종 가능성에 중점을 두었으며, 실제 적층 품질, 열 누적, 비드 형상과 같은 공정 품질 지표와의 직접적인 상관관계는 검증하지 못하였다. 따라서 향후 연구에서는 프레임워크를

통해 도출된 최적 경로를 실제 로봇 하드웨어에 적용하여 시뮬레이션과의 간극을 검증하고 보정하는 데 주력할 예정이다. 아울러 작업 공간 내 장애물 회피나 다중 로봇 협업과 같은 복합적인 제약 조건을 포함하는 다목적 최적화로 시스템을 고도화하고, 적용 가능한 공정의 범위를 확대해 나갈 계획이다.

결론적으로 본 연구는 고비용의 상용 도구 없이도 로봇 경로 계획을 수행할 수 있는 기반을 마련하였다. 이는 산업 및 교육 현장의 기술 접근성을 높여 데이터 기반의 지능형 자율 제조 환경을 구현하는 데 핵심적인 기여를 할 것으로 기대된다.

## ACKNOWLEDGEMENT

이 논문은 과학기술정보통신부 및 한국연구재단의 생애 첫 연구 연구결과로 수행되었음(No. RS-2023-00278548).

이 논문은 산업통상자원부 및 한국산업기술진흥원의 제조 데이터 표준 인공지능 활용 제품 전주기 탄소중립 지원 기술개발 연구결과로 수행되었음(No. RS-2025-10562969).

## REFERENCES

- Li, M., Wang, W., Zhang, J., Wang, C., Zou, L., Huang, Y., (2025), Robotic accurate grinding of complex surfaces with 3D-printed compliant tool featuring internal-blade structure, *Journal of Manufacturing Processes*, 148, 375-385.
- Gibson, B., Mhatre, P., Borish, M., Atkins, C., Potter, J., Vaughan, J., Love, L., (2022), Controls and process planning strategies for 5-axis laser directed energy deposition of Ti-6Al-4V using an 8-axis industrial robot and rotary motion, *Additive Manufacturing*, 58, 103048.
- Kim, J. S., Jung, H. I., Bae, J. H., Jang, Y. S., Cho, M. S., Park, J. H., (2012), Forward kinematics analysis of redundant robot arm using conformal geometric algebra, *Proceedings of the KSPE Spring Conference*, 279-280.
- Ahn, K., Jee, S., (2020), Five-axis tool path smoothing based on cubic B-spline curves, *Journal of the Korean Society for Precision Engineering*, 37(3), 217-224.
- Li, S., Zhang, X., (2022), Research on planning and optimization of trajectory for underwater vision welding robot, *Array*, 16, 100253.
- Bacomo, A., Chanal, H., Duc, E., Limousin, M., (2025), A method for analyzing the kinematic influence of a robot on the WAAM process, *International Journal of Advanced Manufacturing Technology*, 136(10), 4239-4255.
- Ahn, C. W., Kim, D. H., Kim, C. K., Cho, Y. T., (2020), Development of a software for path planning in wire arc additive manufacturing, *Journal of the Korean Society for Precision Engineering*, 37(2), 149-155.
- Malik, A., Lischuk, Y., Henderson, T., Prazenica, R., (2022), A deep reinforcement-learning approach for inverse kinematics solution of a high degree of freedom robotic manipulator, *Robotics*, 11(2), 44.
- Elguea-Aguinaco, I., Inziarte-Hidalgo, I., Bøgh, S., Arana-Arexolaleiba, N., (2024), A review on reinforcement learning for motion planning of robotic manipulators, *International Journal of Intelligent Systems*, 2024(1), 1636497.
- Salvato, E., Fenu, G., Medvet, E., Pellegrino, F. A., (2021), Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning, *IEEE Access*, 9, 153171-153187.
- Garcia del Castillo Lopez, J., (2019), Machina.NET: a library for programming and real-time control of industrial robots, *Journal of Open Research Software*, 7(1), 18.
- Raffaeli, R., Bilancia, P., Neri, F., Peruzzini, M., Pellicciari, M., (2022), Engineering method and tool for the complete virtual commissioning of robotic cells, *Applied Sciences*, 12(6), 3164.
- Pan, J., Fu, Z., Xiong, J., Lei, X., Zhang, K., Chen, X., (2022), Robmach: G-code-based off-line programming for robotic machining trajectory generation, *The International Journal of Advanced Manufacturing Technology*, 118(7), 2497-2511.
- Nagata, F., Yamane, Y., Okada, Y., Kusano, T., Watanabe, K., Habib, M. K., (2018), Development of post-processor approach for an industrial robot FANUC R2000iC, *Artificial Life and Robotics*, 23(2), 186-191.
- Profanter, S., Perzylo, A., Rickert, M., Knoll, A., (2021), A generic plug & produce system composed of semantic OPC UA skills, *IEEE Open Journal of the Industrial Electronics Society*, 2, 128-141.
- Chinello, F., Scheggi, S., Morbidi, F., Prattichizzo, D., (2010), KCT: a MATLAB toolbox for motion control of KUKA robot manipulators, *Proceedings of the IEEE International Conference on Robotics and Automation*, 4603-4608.
- Pan, Z., Polden, J., Larkin, N., Van Duin, S., Norrish, J., (2012), Recent progress on programming methods for industrial robots, *Robotics and Computer-Integrated Manufacturing*, 28(2), 87-94.
- Bilancia, P., Schmidt, J., Raffaeli, R., Peruzzini, M., Pellicciari, M., (2023), An overview of industrial robots control and programming approaches, *Applied Sciences*, 13(4), 2582.
- Yang, X., Ji, Z., Wu, J., Lai, Y.-K., (2021), An open-source multi-goal reinforcement learning environment for robotic manipulation with Pybullet, *Proceedings of the Annual Conference Towards Autonomous Robotic Systems*, 14-24.
- Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P., (2017), Domain randomization for transferring deep neural networks from simulation to the real world, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 23-30.

21. Peng, X. B., Andrychowicz, M., Zaremba, W., Abbeel, P., (2018), Sim-to-real transfer of robotic control with dynamics randomization, Proceedings of the IEEE International Conference on Robotics and Automation, 3803-3810.
22. Chen, X., (2025), Robotic arm control method based on reinforcement learning optimization, Proceedings of the IEEE 5th International Conference on Electronic Technology, Communication and Information, 221-225.
23. Luo, Y., Xue, Z., Song, X., Miao, Z., Hu, Y., (2025), Dynamic cooperative optimisation with differential evolution for trajectory planning of robotic arms, IET Cyber-Systems and Robotics, 7(1), e70016.
24. Suh, U. H., Choi, S. S., Park, H. M., Kim, T. S., Oh, K. Y., Yi, H., (2024), Passive mode control of 2 DOF wearable upper-limb rehabilitation robot, Journal of the Korean Society for Precision Engineering, 41(8), 591-596.
25. Liu, J., Lin, X., Huang, C., Cai, Z., Liu, Z., Chen, M., Li, Z., (2025), A study on path planning for curved surface UV printing robots based on reinforcement learning, Mathematics, 13(4), 648.
26. Zheng, Q., Jia, J., Zhu, P., Xiao, Y., Zhang, F., Ma, W., Zhao, S., (2022), Kinematics analysis and trajectory planning of ABB-IRB2400 robot, UPB Scientific Bulletin, Series D, 84(2).
27. Marchesini, E., Farinelli, A., (2020), Discrete deep reinforcement learning for mapless navigation, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 10688-10694.
28. Kober, J., Bagnell, J. A., Peters, J., (2013), Reinforcement learning in robotics: a survey, International Journal of Robotics Research, 32(11), 1238-1274.

**Cho A Kim**

B.Sc. candidate in the Department of Mechanical Engineering, Seoul National University of Science and Technology. Her research interest is Production/Technology.

E-mail: chocochoa@seoultech.ac.kr

**Jong U Baek**

B.Sc. candidate in the Department of Mechanical Engineering, Seoul National University of Science and Technology. His research interest is Smart Manufacturing.

E-mail: 20100119@seoultech.ac.kr

**Su Han Lee**

B.Sc. candidate in the Department of Mechanical Engineering, Seoul National University of Science and Technology. His research interest is Production/Technology.

E-mail: tngks70@seoultech.ac.kr

**Ju Yeon Lee**

Professor in the Department of Mechanical Engineering, Seoul National University of Science and Technology. Her research interest is DigitalTwin/CPS.

E-mail: jylee@seoultech.ac.kr