ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Vision-Guided Robotic Grinding with Deep Learning-Based Bead Segmentation and Digital Twin Verification*

Seong Hyeon Kim and Hyo-Young Kim

Abstract— **Weld bead grinding is a critical post-processing step in metal fabrication, yet conventional robotic grinding based on teach-pendant programming lacks adaptability to variations in bead geometry and position. This paper presents a vision-guided robotic grinding system that combines deep learning-based weld bead segmentation, automated grinding path generation, and digital twin-based pre-verification. A U-Net model with a ResNet34 encoder and ImageNet pre-training segments weld bead regions from RGB images captured by an Intel RealSense D415 camera mounted on a Staubli RX160 manipulator, achieving a mean Intersection over Union (IoU) of 0.9311 and a Dice coefficient of 0.9641. The segmented bead contours are transformed into the robot coordinate frame through hand-eye calibration and forward kinematics, enabling automated generation of grinding waypoints along the bead centerline. The CHOMP algorithm plans collision-free trajectories within MoveIt, and all planned motions are validated in a digital twin environment built on NVIDIA Isaac Sim 5.0, integrated with ROS through a distributed multi-container architecture. Experimental results demonstrate that the proposed system effectively generates adaptive grinding paths for varying weld bead geometries and verifies them in simulation before physical deployment.**

I. Introduction

Welding is among the most prevalent joining methods in metal fabrication across shipbuilding, automotive, and structural steel industries. The weld beads remaining on workpiece surfaces after welding must be removed through grinding to ensure dimensional accuracy and surface quality, as residual beads can act as stress concentrators that compromise fatigue life [1]. Conventionally, weld bead grinding has been performed manually using handheld angle grinders, which is labor-intensive, produces inconsistent results, and exposes workers to hazardous conditions including noise, vibration, and metal dust [2].

Robotic grinding systems have been increasingly adopted to address these issues [3]. However, conventional robotic grinding typically relies on teach-pendant programming, which fundamentally lacks the flexibility to accommodate variations in weld bead geometry, position, and height that naturally occur in production environments [4]. Each new workpiece configuration requires re-teaching, making the approach impractical for high-mix, low-volume manufacturing scenarios.

Recent advances in computer vision and deep learning have enabled vision-guided robotic systems that can autonomously adapt to workpiece variations [5]. The U-Net architecture [6], with its encoder-decoder structure and skip connections, is well suited for pixel-level weld bead segmentation. Combined with hand-eye calibration [7], such systems can automatically generate grinding paths tailored to each workpiece. Furthermore, digital twin technology [8] allows virtual commissioning of planned trajectories before physical execution, while the CHOMP algorithm [9] generates smooth, collision-free trajectories through continuous optimization.

This paper presents an integrated vision-guided robotic grinding system that combines: (1) a ResNet34-backbone U-Net with ImageNet transfer learning for weld bead segmentation, (2) coordinate transformation and grinding path generation through hand-eye calibration and forward kinematics, (3) CHOMP-based trajectory planning within MoveIt, and (4) digital twin verification using NVIDIA Isaac Sim 5.0 with ROS integration for pre-deployment validation.

II. System Architecture

The hardware platform consists of a Staubli RX160 6-DOF industrial manipulator and an Intel RealSense D415 RGB-D camera mounted on the end-effector in an eye-in-hand configuration. An NVIDIA RTX 5070 GPU is used for deep learning inference and simulation rendering.

The software architecture is based on a distributed ROS topology. Robot control and trajectory execution operate under ROS 1 Noetic using MoveIt and the staubli_val3_driver. Vision processing runs inside a Docker container (Ubuntu 22.04, ROS 2 Humble, Python 3.10) to satisfy PyTorch Nightly CUDA 12.8 wheel requirements for the RTX 5070 GPU. Simulation is handled through ROS 2 Jazzy on the host system (Ubuntu 24.04). A bidirectional ROS 1–ROS 2 bridge enables communication between the heterogeneous middleware layers.

The processing pipeline operates as follows: the RGB-D camera captures workpiece surface images; the vision module segments weld bead regions using the trained U-Net model; the segmented bead coordinates undergo a three-stage coordinate transformation— camera intrinsic-based deprojection, hand-eye calibration, and forward kinematics—to produce 3D positions in the robot base frame. Approximately 30 grinding waypoints are generated along the bead centerline with appropriate tool orientations. These waypoints are processed by the CHOMP planner within MoveIt to produce smooth, collision-free trajectories. Before

* This study was supported by the Korea Institute of Industrial Technology under the project (KITECH EH-25-0004 and KITECH JB-26-0002) and by the Ministry of Trade, Industry and Energy (MOTIE) and KIAT (RS-2025-24536353) and by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (RS-2026-25493046).

S. H. Kim is with the Korea Institute of Industrial Technology, Cheonan, Republic of Korea (corresponding author to provide phone: 82-41-589-8562; e-mail: shkim@kitech.re.kr).

H.-Y. Kim is with Tech University of Korea, Siheung, Republic of Korea (e-mail: kimhy@tukorea.ac.kr).

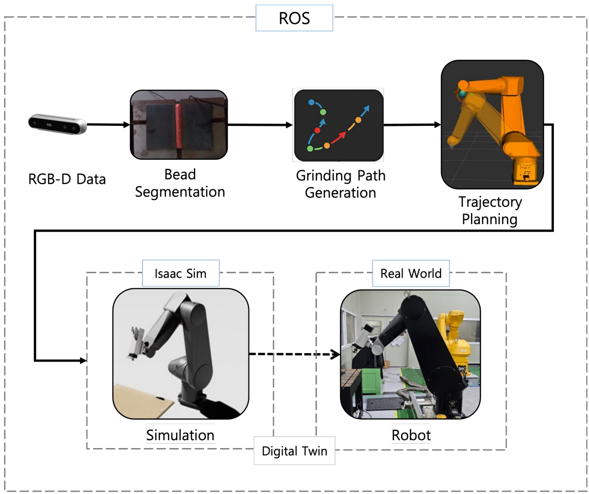

Figure 1. Schematic Diagram for Vision-Based Robotic Grinding Path Generation and Verification System

physical execution, all planned motions are validated in the NVIDIA Isaac Sim 5.0 digital twin environment.

III. Weld Bead Segmentation

A U-Net segmentation model [6] is employed for pixel-wise classification of weld bead regions from RGB images. The encoder uses a ResNet34 [10] backbone with four residual block groups (64, 128, 256, 512 channels), initialized with ImageNet pre-trained weights for effective transfer learning. The decoder follows the standard U-Net design with four upsampling stages, skip connections, and a final 1×1 convolution with sigmoid activation to produce a binary segmentation mask.

The training dataset comprises 197 RGB images of welded metal plates captured at distances of 100–300 mm, split into 70% training, 15% validation, and 15% testing at 1024×1024 resolution. Data augmentation (horizontal flipping, shift-scale-rotate, brightness-contrast adjustment, Gaussian noise, motion blur) and a composite loss (BCEWithLogitsLoss + Dice loss) were applied. The model was trained for 50 epochs with batch size 6 using Adam (lr = 1e−4) with a plateau-based scheduler. On the test set, the model achieved a mean IoU of 0.9311 and Dice coefficient of 0.9641, with 29 ms inference time on the RTX 5070 GPU.

IV. Path Generation and Digital Twin Verification

To convert the segmented bead coordinates from the 2D image plane to the 3D robot workspace, a multi-stage coordinate transformation pipeline is employed. This involves three sequential steps: (1) image-to-camera frame transformation using the camera intrinsic matrix, (2) camera-to-end-effector transformation through hand-eye calibration, and (3) end-effector-to-base frame transformation via forward kinematics of the Staubli RX160.

Once the bead contour points are expressed in the robot base frame, the bead centerline is extracted from the binary mask using morphological skeletonization. Approximately 30 waypoints are sampled at uniform intervals along the skeleton, each encoding position (x, y, z) and tool orientation aligned with the surface normal. Inverse kinematics is computed for each waypoint, and the CHOMP algorithm [9] generates smooth, collision-free trajectories by minimizing a cost functional that balances trajectory smoothness with obstacle avoidance through functional gradient descent.

Before physical execution, the planned grinding trajectories are verified in a digital twin environment built on NVIDIA Isaac Sim 5.0, which replicates the physical robot cell with accurate geometric and kinematic models imported via URDF. The trajectory commands from the CHOMP planner in MoveIt are bridged from ROS 1 to ROS 2 and executed on the virtual robot, enabling collision detection, joint limit verification, and path coverage inspection. The simulation confirmed that the planned trajectories could be executed without collisions or joint limit violations, reducing commissioning time and risk before physical deployment.

V. Conclusion

This paper presented a vision-guided robotic grinding system that integrates deep learning-based weld bead segmentation, automated grinding path generation, CHOMP-based trajectory planning, and digital twin-based verification. The ResNet34-backbone U-Net segmentation model achieved a mean IoU of 0.9311 and Dice coefficient of 0.9641, demonstrating reliable weld bead detection under varying conditions. The coordinate transformation pipeline through hand-eye calibration and forward kinematics enables automatic generation of grinding waypoints along the bead centerline without manual intervention. The digital twin environment in Isaac Sim provides safe trajectory verification before physical deployment, confirming collision-free execution and joint limit compliance. Future work will focus on closed-loop force control for adaptive grinding and extension to complex multi-pass weld bead geometries.

References

[1] Z. Zhu, X. Tang, C. Chen, F. He, and M. Li, “Review on robot-assisted polishing: Status and future trends,” Robot. Comput.-Integr. Manuf., vol. 80, 2023.

[2] H. Zhou, S. Ma, G. Wang, Y. Deng, and Z. Liu, “A hybrid control strategy for grinding and polishing robot based on adaptive impedance control,” Proc. Inst. Mech. Eng. C, 2021.

[3] M. Gonzalez, J. Rodriguez, and P. Fernandez, “A robotic grinding station based on an industrial manipulator and vision system,” PLOS ONE, 2021.

[4] Y. Wang, Z. Zhang, and L. Chen, “Trajectory planning method with grinding compensation strategy for robotic propeller blade sharpening application,” Robot. Comput.-Integr. Manuf., 2023.

[5] J. Chen, H. Liu, and W. Zhang, “Vision sensing-based online correction system for robotic weld grinding,” Chin. J. Mech. Eng., vol. 36, 2023.

[6] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, Springer, 2015, pp. 234–241.

[7] R. Y. Tsai and R. K. Lenz, “A new technique for fully autonomous and efficient 3D robotics hand/eye calibration,” IEEE Trans. Robot. Autom., vol. 5, no. 3, pp. 345–358, 1989.

[8] M. Grieves and J. Vickers, “Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems,” in Transdisciplinary Perspectives on Complex Systems, Springer, 2017, pp. 85–113.

[9] M. Zucker et al., “CHOMP: Covariant Hamiltonian optimization for motion planning,” Int. J. Robot. Res., vol. 32, no. 9–10, pp. 1164–1193, 2013.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, 2016, pp. 770–778.