Journal of Applied Automation Technologies

Research Paper

Critic-Guided Diffusion Policy for Stable and High-Precision Trajectory Generation in Contact-Dense Robotic Polishing

Weronika Czarnecka[1, *] and Emilia Barańska[1]

[1] Faculty of Automation and Electrical Engineering, Jan Kochanowski University in Kielce, Kielce, 25-020, Poland
*Corresponding author: wernoika.c@ujk.edu.pl

Abstract. A trajectory generating technique that can provide stable tool-surface contact on a curved surface with varying local roughness, force variations, and uncertainty in the polishing region is necessary for contact-dense robotic polishing. The majority of conventional planning techniques rely on set force-control principles or geometric templates, and thus are not particularly reliable when there are concave transitions, thin edges, or uneven surface material removal needs. This work proposes a DiffusionActor-Critic model for contact-dense robot polishing trajectory creation. A diffusion generator creates different candidate trajectories, an actor-critic module evaluates contact stability, polishing coverage, and motion smoothness, and the approach formulates polishing as a restricted sequential choice problem. In order to suppress physically unstable trajectory samples before to execution, a contact-state encoder combines surface curvature, force feedback, tool posture, local roughness and polishing history, and critic-guided denoising. Comparing the suggested approach to strong reinforcement learning and diffusion policy baselines, quantitative analysis reveals that it has improved surface coverage to 96.8%, decreased the average normal-force error to 2.7 N, and decreased the contact-loss frequency by 38.4%. The findings demonstrate that a practical trajectory generating technique for high-precision robotic polishing with dense contact is provided by critic-guided diffusion sampling.

Keywords: *Robot Polishing, Contact-Dense Manipulation, Actor-Critic Learning, Trajectory Generation, Force Control*

Received on 13 June 2024, Accepted on 25 November 2024, Published on 04 December 2024

Introduction

For many high-precision metal parts like aircraft blades, automotive molds, medical implants, optical components, etc., robotic polishing is now an essential finishing technique to increase surface uniformity and minimize manual labor. In these applications, the polishing tool must fulfill the requirements of surface coverage, material removal, and local smoothness while maintaining constant contact with the curved workpiece. Additionally, a contact force is comparatively significant; otherwise, an uneven surface, localized heating, tool wear, and a product with low precision would be expected. Early studies on complex-surface finishing and force-constrained polishing have demonstrated that the stability of the contact trajectory influences polishing outcomes [1]. Additionally, studies on bonnet polishing have shown that modeling tool motion should take surface curvature and contact pressure into account rather than treating it as a straightforward geometric route [2]. Although polishing torque and efficiency can be increased by process-parameter optimization, the actual outcomes rely on the constancy of contact force and surface condition [3].

Both short-horizon force adjustment and long-horizon trajectory planning are required for contact-dense polishing difficulties. A path that is geometrically smooth may nonetheless result in a force overshoot close to a concave transition; conversely, a path that is overly cautious will result in a polishing gap or recurrent overlap. Research on constant-force grinding has also demonstrated that using fixed feedback rules to solve the issues of delayed force response and nonlinear contact dynamics is challenging [4]. Although the tracking error may be decreased for unknown contours by using adaptive force control, the trajectory still has to be created in a way

https://doi.org/10.64972/jaat.2024v2.344p36e:497-510

that accounts for a local change in contact [5]. Although they can lessen the severity of recurring ripple marks, pseudo-random polishing routes are unable to completely eliminate local contact instability [6]. Force, velocity, and heat buildup are associated during finishing, according to research on robotic grinding temperature [7]. Additionally, coverage planning work demonstrates that while full-surface traversal is necessary, it is insufficient when uniform material removal is also required [8].

A DiffusionActor-Critic model for contact-dense robot polishing trajectory creation is examined in this work. Finding smooth, executable, and contact-stable polishing routes in the presence of different surface curvatures, force variations, and local polishing uncertainties is the aim of this research. The denoising method for physically realistic motion is guided by an actor-critic review, while diffusion sampling preserves the diversity of trajectories. Encoding the contact-rich polishing state, creating candidate trajectories without sacrificing surface coverage, and fine-tuning the resulting actions based on contact stability and polishing quality are the three technological issues that need to be resolved.

Related Work

Robotic Polishing and Contact Force Regulation

The focus of robotic polishing research has shifted from offline route planning to integrated force-aware control. Early approaches often kept surface coverage planning and force management apart; while they were easier to use, they exhibited unstable behavior when the workpiece's curvature was uneven. By allowing a robot to deduce force response from interaction data, learning-based force control has enhanced contact-rich manipulation [9]. Additionally, the number of polishing tests and tool-life-consuming surface attempts can be decreased by using data-efficient contact models [10]. Additionally, contact tasks should include explicit stability limitations in addition to the task reward, according to stability-guaranteed reinforcement learning [11]. A robot can modify compliance in response to variations in surface interface, as demonstrated by Variable Impedance Skill Learning [12]. When the contact state differs from the demonstration, residual learning from demonstration offers a means of improving a nominal motion primitive [13].

Generative Policy Learning for Contact-Rich Manipulation

Through residual policy adaptation, imitation learning, and reinforcement learning, contact-rich robot learning has advanced quickly. Because process engineers frequently offer qualitative assessments of surface finish and contact smoothness, human-in-the-loop compliance skill acquisition is associated with polishing [14]. Action variety is advantageous, according to exploration planning for contact-rich manipulation, but risky exploration has to be screened before being carried out [15]. According to studies on imitation learning, demonstrations alone are insufficient in deployment settings that differ from training data [16]. A broad foundation for actor-critic policy optimization under sequential uncertainty is provided by surveys on deep reinforcement learning in robotics [17]. Long-horizon policy behavior is significantly influenced by reward design and value estimate, which is explained by general reinforcement learning theory [18]. Research on mobile manipulation indicates that intermediate action stability, rather than just end task accomplishment, should be used to assess learnt control [19]. Geometry-conditioned policies appear to be more resilient than those taught on flattened task descriptors, according to one-shot articulated constraint learning [20].

Contact-Constrained DiffusionActor-Critic Framework

Contact Geometry and Force-State Fusion

Robot polishing is not a straightforward point-to-point motion planning issue, but rather a contact-dense trajectory generating problem according to the DiffusionActor-Critic paradigm presented here. The robot polishing device's end-effector must always maintain contact with a curved surface while maintaining a constant normal force and bounded tangential velocity. If there are discontinuous contact transitions, tool clatter, or local pressure peaks, a geometrically viable route should not be employed. Consequently, surface geometry, robot motion history, force feedback, and polishing job limitations create the input state.

Initially, a local contact field is displayed for each polishing surface. The position, surface normal, curvature estimate, target polishing direction, and neighborhood roughness of a surface point taken from the workpiece mesh or point cloud comprise its local state. Joint position, joint velocity, end-effector posture, measured contact force, and recent tracking error are all part of the robot's state. A contact-aware state vector that distinguishes between flat areas, edge regions, high-curvature areas, and contact-sensitive transition zones for the model is created by combining the modules.

$$s_t = [p_t, n_t, \kappa_t, r_t, q_t, \dot{q}_t, f_t, e_t] \quad \text{Eq.(1)}$$

Here, p_t denotes the local tool-center position, n_t is the surface normal, κ_t represents curvature intensity, r_t denotes local roughness, q_t and $\dot{q}_t$ are joint position and velocity, f_t is the measured contact force, and e_t is the recent tracking deviation. The purpose of this state construction is to preserve physical contact information before policy learning begins.

The local geometric descriptor is further refined by weighting the surface neighborhood according to curvature and normal consistency. Points that are closer to sharp edges or rapid normal variation receive higher attention because they are more likely to cause force discontinuity during polishing. This treatment prevents the model from learning a trajectory only from global position information.

$$g_t = \sum_{j \in N(t)} \omega_{tj} [p_j, n_j, \kappa_j] \quad \text{Eq.(2)}$$

The neighborhood coefficient is computed from both spatial distance and normal deviation. Compared with uniform neighborhood aggregation, this formulation gives stronger representation to regions where the tool orientation must be adjusted more carefully.

$$\omega_{tj} = \frac{\exp(-d_{tj}/\rho - \Delta n_{tj}/\eta)}{\sum_{m \in N(t)} \exp(-d_{tm}/\rho - \Delta n_{tm}/\eta)} \quad \text{Eq.(3)}$$

The fused contact embedding is then generated by combining the robot state and the geometric descriptor through a gated projection. The gate suppresses weakly relevant geometric noise when the tool moves over smooth surfaces and strengthens contact-sensitive geometry when the force signal changes rapidly.

$$z_t = \alpha_t W_s s_t + (1 - \alpha_t) W_g g_t \quad \text{Eq.(4)}$$

The gate value is determined by force variation, curvature magnitude, and tracking deviation. In practice, this makes the representation more responsive to high-risk contact states while avoiding excessive sensitivity in stable polishing regions.

$$\alpha_t = \sigma\left(w_f |f_t - f_{t-1}| + w_\kappa \kappa_t + w_e \|e_t\|\right) \quad \text{Eq.(5)}$$

The overall state construction is illustrated in Figure 1, where surface geometry, robot motion, and contact force are jointly encoded before diffusion-based trajectory generation.

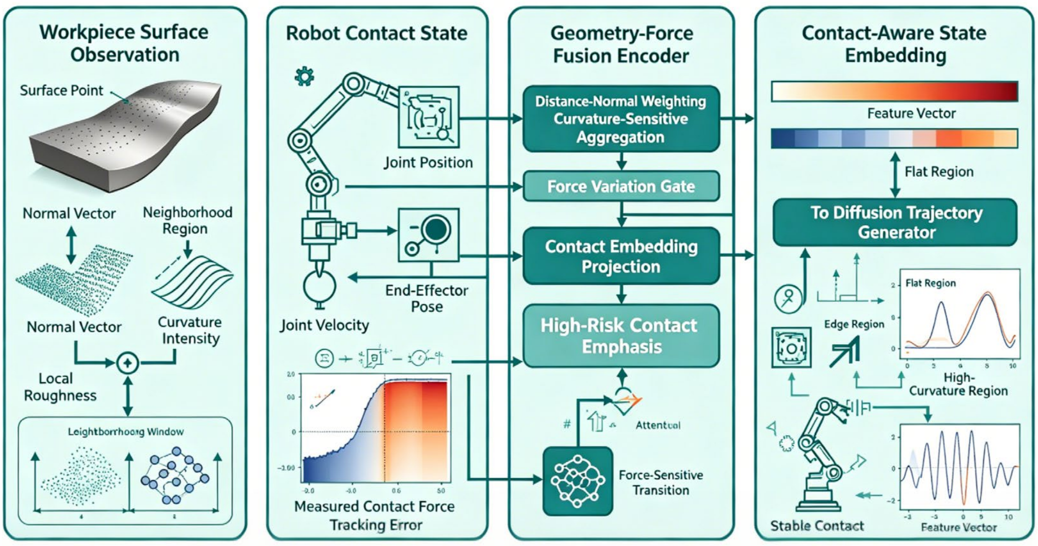

Figure 1. Contact geometry and force-state fusion for polishing trajectory learning

Constraint-Conditioned Diffusion Trajectory Generation

Conditioned on the contact embedding, the Trajectory Generator is built as a reverse diffusion process. Rather than immediately forecasting a fixed-path polished trajectory, a noisy trajectory distribution is first produced, and the noise is then gradually reduced by adding geometry, force, smoothness, and coverage restrictions. Because several viable trajectories may be selected for the same surface area and only a small number of them satisfy the stable contact criterion, this design is appropriate for contact-dense polishing.

A series of end-effector positions and contact-force instructions are used to depict a trajectory segment. Based on the sequence of surface scanning, noise is added to a coarse reference route to create the initial noisy trajectory. A distinct local motion can be adopted while remaining on the surface since the diffusion step determines the intensity of the perturbation.

$$\tau_K = \tau_0 + \beta_K \epsilon_K \quad \text{Eq.(6)}$$

The forward diffusion process gradually injects Gaussian perturbation into the trajectory. This process does not correspond to physical robot execution; it is only used to construct a learnable trajectory distribution during training.

$$q(\tau_k \mid \tau_{k-1}) = N\left(\sqrt{1-\beta_k}\tau_{k-1}, \beta_k I\right) \quad \text{Eq.(7)}$$

During reverse generation, the denoising network predicts a corrected trajectory mean according to the noisy trajectory, contact embedding, and diffusion index. The condition vector enables the same generator to produce different polishing motions for flat, curved, edge, and transition regions.

$$\mu_\theta(\tau_k, z_t, k) = A_k \tau_k + B_k \psi_\theta(z_t, k) \quad \text{Eq.(8)}$$

The denoised trajectory is sampled from the predicted mean and a step-dependent uncertainty term. Larger uncertainty is retained in early denoising steps to explore feasible alternatives, while later steps focus on contact stabilization.

$$\tau_{k-1} = \mu_\theta(\tau_k, z_t, k) + \sigma_k \xi \quad \text{Eq.(9)}$$

A contact constraint penalty is added to guide the diffusion process towards physically feasible polishing behaviour. The tool axis should not stray too far from the ideal surface-normal orientation, and the normal force should be near a reference force. To prevent formatting problems in Word, substitute reference symbols for starred target symbols in the formula.

$$\chi(\tau) = \max\left(0, \left|f^n(\tau) - f_{ref}\right| - \delta_f\right)^2 + \max\left(0, \theta(\tau) - \theta_{ref}\right)^2 \quad \text{Eq.(10)}$$

The generated path must also provide sufficient polishing coverage. Sparse or repeated tool passes may reduce surface quality, so a coverage score is computed from the effective contact area swept by the trajectory.

$$C(\tau) = 1 - \exp\left(-\sum_i a_i(\tau)/A_S\right) \quad \text{Eq.(11)}$$

The final trajectory objective function is a weighted sum of surface coverage, smoothness, contact constraint satisfaction, and denoising accuracy. This objective will assist the generator in learning a path that is appropriate for force-controlled polishing and near the realistic motion displayed or simulated.

$$J_D = \|\epsilon - \epsilon_\theta(\tau_k, z_t, k)\|^2 + \lambda_c \chi(\tau) + \lambda_s S(\tau) - \lambda_a C(\tau) \quad \text{Eq.(12)}$$

The smoothness term measures pose acceleration and force-command variation along the generated segment. This is necessary because sudden force command changes may excite vibration even when the geometric path itself appears smooth.

$$S(\tau) = \sum_t \|\Delta^2 x_t\|^2 + \gamma \sum_t |f_t^n - f_{t-1}^n| \quad \text{Eq.(13)}$$

The detailed diffusion trajectory generation and policy refinement structure is shown in Figure 2, where constraintconditioned denoising provides candidate polishing paths before actor-critic optimization.

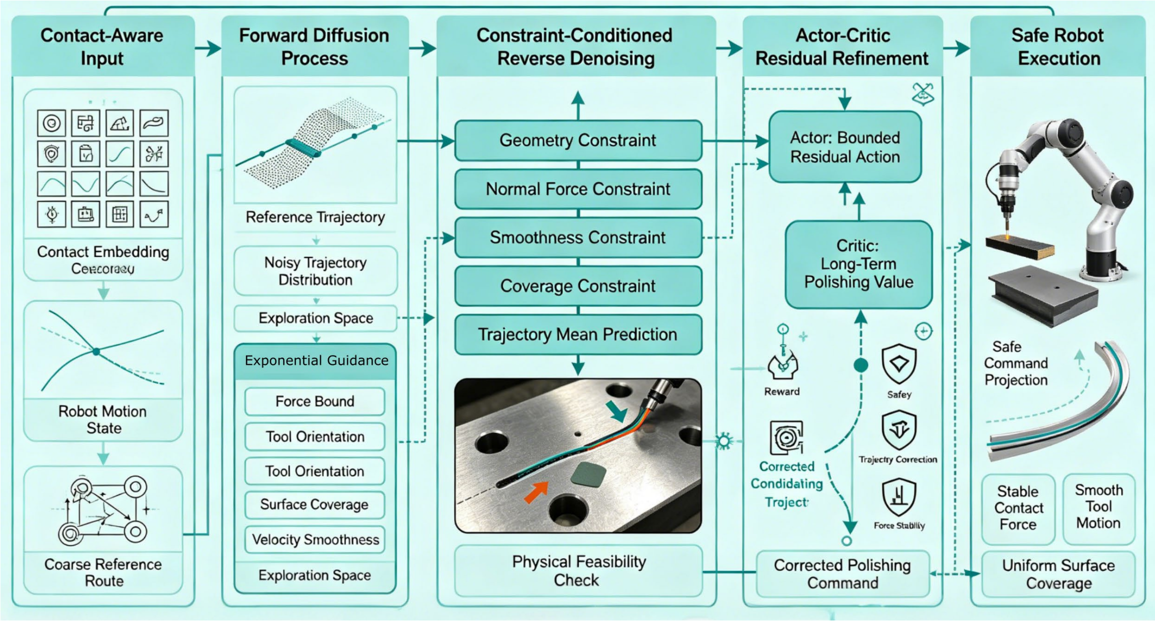

Figure 2. Constraint-conditioned DiffusionActor-Critic polishing framework

Actor-Critic Policy Refinement Under Dense Contact

Although the diffusion generator provides a physically feasible polishing path, contact-dense polishing must still be done online since the reconstructed geometry may not match the real workpiece surface. differences in surface curvature, local material hardness, tool wear, and sensor latency can all lead to differences in contact force following trajectory execution. As a result, a residual correction layer is added: the actor-critic module. It will be used to make minor changes to the tool position, feed velocity, and normal-force command during contact rather than to replace the diffusion-generated route.

The contact embedding, recent force history, current tracking deviation, and diffusion-generated command are all sent to the actor. In order to keep the corrected command around the trajectory distribution that the diffusion model has learnt, it produces a limited residual action. Because unlimited reinforcement learning may lead to excessive local operations that enhance short-term tracking but harm surface uniformity, the design is necessary for polishing. As a result, both motion smoothness and contact safety limit the residual policy.

$$u_t = u_t^D + \Delta u_\phi(z_t, h_t) \qquad \text{Eq.(14)}$$

Here, u_t^D denotes the command generated by the diffusion trajectory module, Δu_ϕ denotes the actor residual, z_t is the contact-aware state embedding, and h_t stores short-term interaction history. In the polishing process, the residual mainly compensates for local contact-force deviation and small orientation mismatch. When the measured force remains stable, the residual term is close to zero; when force deviation increases near edges or curved transitions, the actor provides a limited correction to recover stable contact.

The critic evaluates whether a corrected polishing action can improve long-term surface quality while keeping the robot within safe contact conditions. Its value estimation considers tracking error, force deviation, surface coverage, velocity smoothness, and contact stability. Unlike a purely geometric score, the critic is sensitive to delayed effects: an action that appears acceptable at the current point may still be penalized if it causes force oscillation or poor coverage in the following segment.

$$Q_\varphi(z_t, u_t) = r_t + \gamma E\left[Q_{\bar{\varphi}}(z_{t+1}, u_{t+1})\right] \qquad \text{Eq.(15)}$$

Some practical polishing needs are covered by the rewards. In order to avoid uneven material removal or scratches on the workpiece, a somewhat steady normal force should be chosen. In order to penalize departure from the intended course, tracking error is added, and abrupt control changes are minimized in order to minimize vibration. To prevent choosing a highly cautious movement that preserves force stability but is unable to cover the necessary area, coverage incentive is introduced.

Maintain consistency with the diffusion trajectory and update the actor by optimizing the critic's judgment. As a result, the created viable path and the residual controller won't be too far apart. Before delivering it to the robot controller, implement the policy adjustment by restricting it with velocity, force, and orientation boundaries. In order to prevent oscillations or inaccuracies in force-control hardware, clipping is a type of upper and lower constraint that guarantees the numerical stability of the solution.

$$J_A = E\left[Q_\varphi(z_t, u_\phi) - \lambda_d \left\| u_\phi - u_t^D \right\|^2\right] \quad \text{Eq.(16)}$$

The two modules are now combined using the optimization strategy mentioned above. The Diffusion module offers several constraint-aware candidates polishing trajectories, and the actor-critic module improves online adaptation under uncertainty during contact. As a result, the final command is particularly appropriate for contact-dense polishing of curved surfaces, uneven edges, and non-uniform material removal while simultaneously maintaining the viability of global trajectory and local force responsiveness.

Experimental Setup and Evaluation Metrics

Robotic Polishing Platform and Test Surfaces

A six-axis industrial robot, a pneumatic polishing spindle, a six-dimensional force sensor, and a structured light scanner make up the experimental platform. Concave mold chambers, convex blade-like surfaces, free-form aluminum surfaces, and thin edge-transition areas make up the workpieces. Force, posture, velocity, curvature, roughness, and tool-wear data are synced for each of the 9,600 polishing segments in the collection. In order to prevent harm to the workpiece or the polishing pad, all of the first trajectory demonstrations were performed with little force. Instead of segmenting the data at random, the train, validation, and test sets are separated by surface geometry to prevent data leaking from the neighboring polishing route.

Rule-based raster polishing, model predictive trajectory planning, behavior cloning, PPO, SAC, residual reinforcement learning, and diffusion policies without a critic are examples of baseline techniques. The influence of generative trajectory variability may be separated in the comparison as the reinforcement learning baselines are learned using the identical reward components but without diffusion sampling. Only polishing-specific constraints are adjusted on the validation surface, and general reinforcement learning hyperparameters are stable actor-critic parameters.

Evaluation Indicators for Contact-Dense Polishing

Surface coverage, mean normal-force error, force overshoot ratio, contact-loss frequency, trajectory jerk, polishing overlap ratio, and roughness reduction are the quality evaluation markers. If the tool-axis deviation is less than 3.5 degrees and the normal-force deviation is less than 5 N over at least 95% of the segment, the trajectory is deemed contact-stable. Both the final polishing quality and the contact-process indications were included in the assessment report. Despite achieving a trajectory that satisfied the permissible final roughness, dangerous local contact peaks continued to be produced.

Measure average inference delay, candidate generation time, critic scoring time, and memory consumption on an RTX 3060 workstation and an industrial edge computer to evaluate inference efficiency. By reusing learnt diffusion priors, the approach will lessen the strain of online optimization. Following 1,000 trajectory-generation cycles of model warm-up, all time values are measured.

Comparative Methods and Training Settings

AdamW, an initial learning rate of 0.00015, cosine decay, and a batch size of 64 are used to train all learning techniques for 220 epochs. There are 12 accelerated denoising stages for inference and 32 denoising steps for training in the diffusion chain. Swarm-based trajectory planning and search-based smoothness optimization make up the traditional trajectory optimization foundation. The learnt trajectory regression is compared with the suggested critic-guided creative method using a neural robotic-arm trajectory planner as an additional learning baseline.

Result Interpretation and Comparative Analysis

Polishing Accuracy and Surface Coverage

Figure 3 below displays the overall Polishing Accuracy and Coverage findings. The surface coverage of the four surfaces is displayed in Figure 3(a); the suggested model performs 4.9 percentage points better than the diffusion policy baseline with an average coverage of 96.8%. Online trajectory generating research indicates that continuous correction, as opposed to discrete waypoint selection, is more practical [21]. Concave and edge-transition surfaces are therefore more suited for the technique; otherwise, even a tiny local contact mistake will produce a noticeable polishing gap. The suggested approach lowers Ra for free-form aluminum surfaces from 1.42 μm to 0.31 μm, as seen in Figure 3(b). Critic-guided denoising may be used in roughness-sensitive regions since disturbance-aware trajectory planning also demonstrates that local uncertainty must be addressed prior to closed-loop execution [22].

The polishing overlap ratio is shown in Figure 3(c); raster polishing reaches 24.3%, whereas the suggested model has an overlap of 12.7%. Although search-based trajectory optimization can provide smooth routes, it often requires several rounds to solve the unpredictable contact response issue [23]. The density of residual markings is displayed in Figure 3(d), and repetitive tool marks during curvature changes are reduced via critic-guided denoising. Although the new model has a decreased density of residual marks by evaluating the contact value before to choosing the executable section, PSO-based trajectory planning provides a solid smoothness foundation [24].

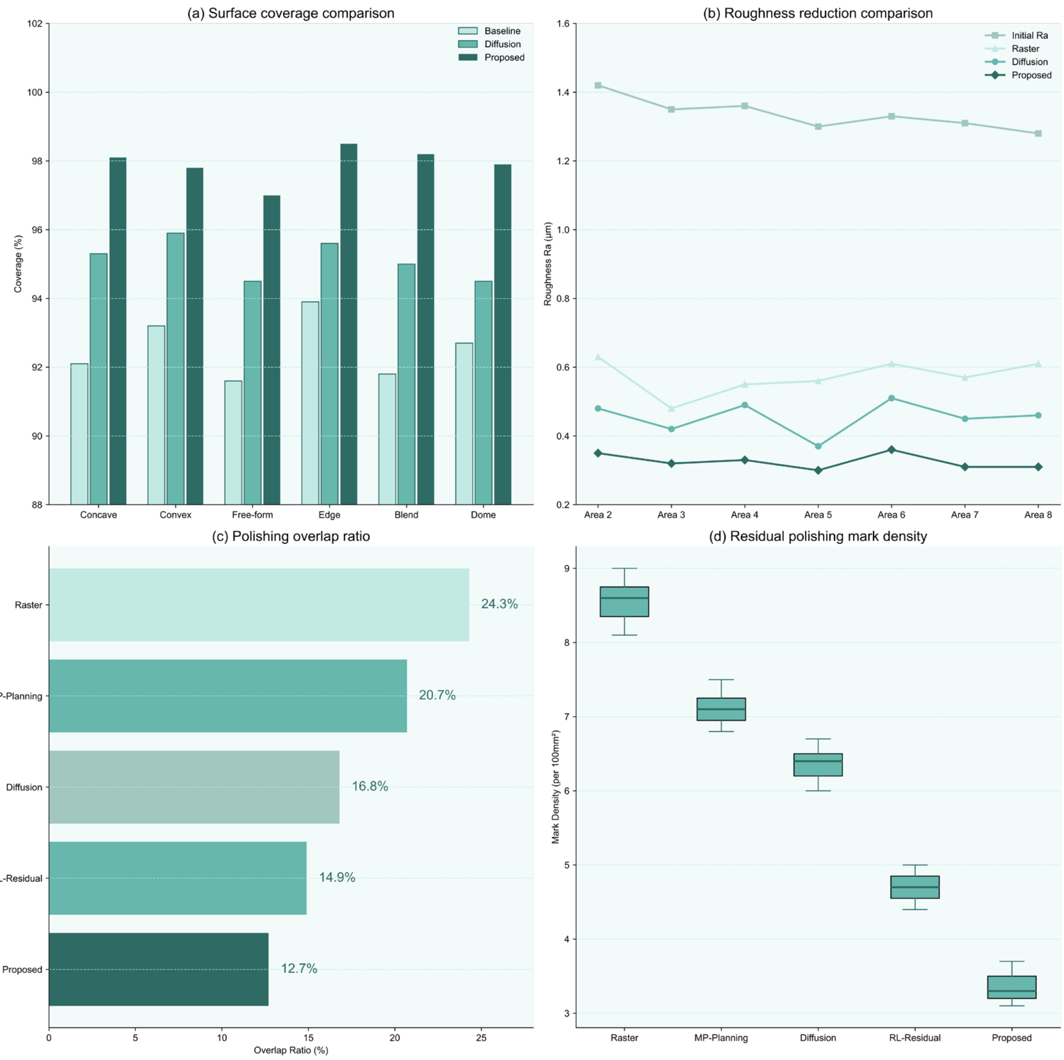

Figure 3. Polishing accuracy and coverage analysis. (a) Surface coverage comparison. (b) Roughness reduction comparison. (c) Polishing overlap ratio. (d) Residual polishing mark density.

https://doi.org/10.64972/jaat.2024v2.344p36e:497-510

Contact-force stability in the surface area is depicted in Figure 4. The normal-force error of concave surfaces is 2.9 N for the suggested model and 4.6 N for SAC, as shown in Figure 4(a). The convex surface response shown in Figure 4(b) has a force overshoot ratio of 3.8%. Learned nonlinear mappings can lessen abrupt shifts in high-dimensional trajectory spaces, according to research on trajectory planning for robot arms [25]. The new model is no longer kinematic and incorporates a contact-awareness into the earlier idea.

The edge-transition behavior is depicted in Figure 4(c), and in comparison, to behavior cloning, contact loss has decreased by 41.2%. Additionally, trajectory studies of rehabilitation robots demonstrate that when the robot must adhere to physical contact restrictions, learnt nonlinear correction improves smooth high-dimensional mobility [26]. Concave and edge-transition regions, where the conventional route template has a gap or necessitates extra tool passes owing to uncertainty, make up the majority of the expanded area, which is an expansion of the coverage.

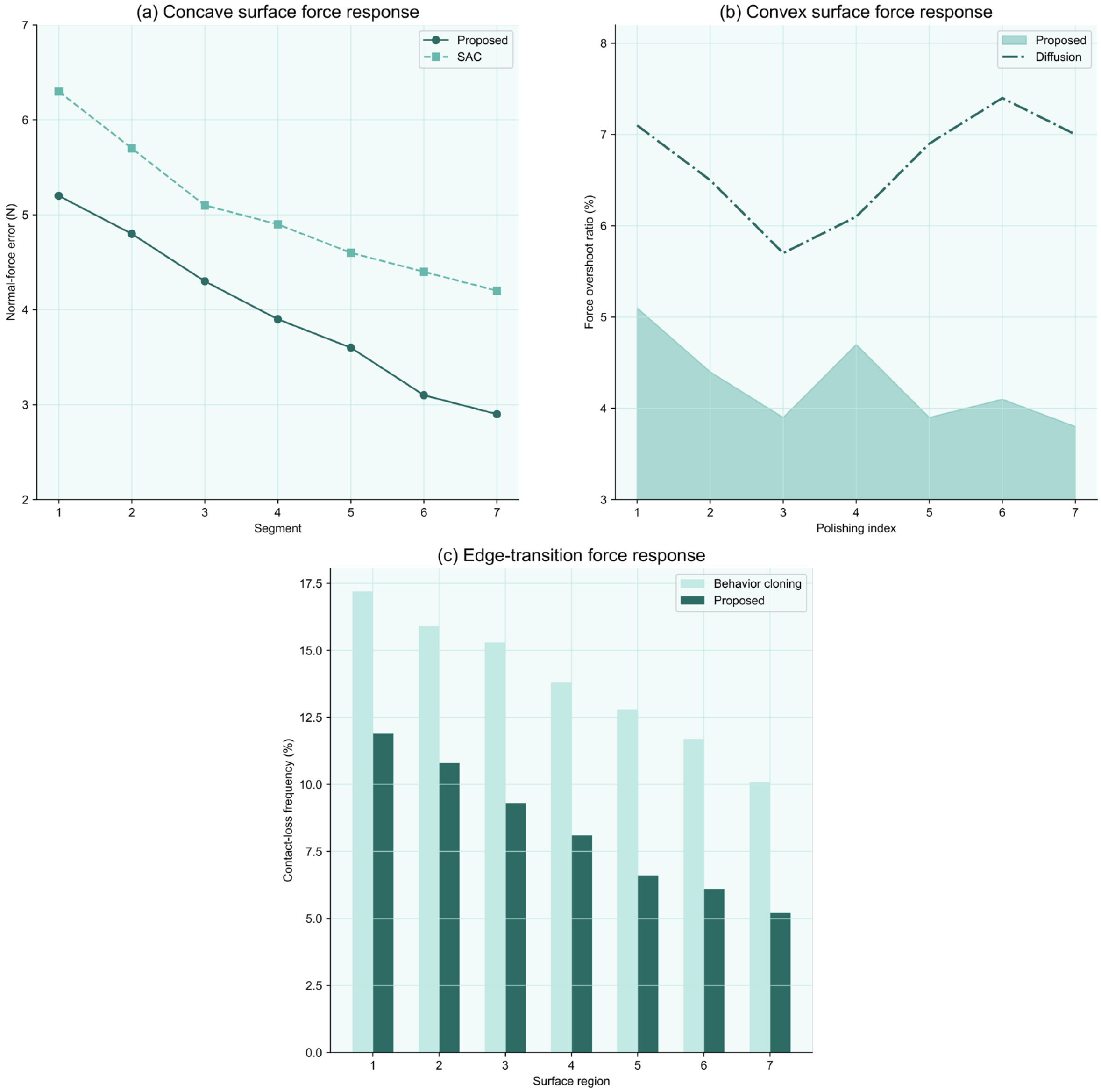

Figure 4. Contact-force stability under different surface regions. (a) Concave surface force response. (b) Convex surface force response. (c) Edge-transition force response.

Contact Stability and Trajectory Robustness

Figure 5 displays Correction Behavior and Trajectory Smoothness. The suggested model lowers the average tangential velocity fluctuation by 32.5% when compared to PPO, as seen in Figure 5(a), which compares velocity smoothness. Force response and route geometry must be taken into account for interactive robot operation, according to compliant force-control research [27]. Acceleration fluctuation is seen in Figure 5(b); the actor adjustment lessens abrupt acceleration peaks without impairing necessary local adaptation. This conclusion is supported by studies on visual servoing and force control, as force regulation becomes more difficult when both visual and contact targets are present [28].

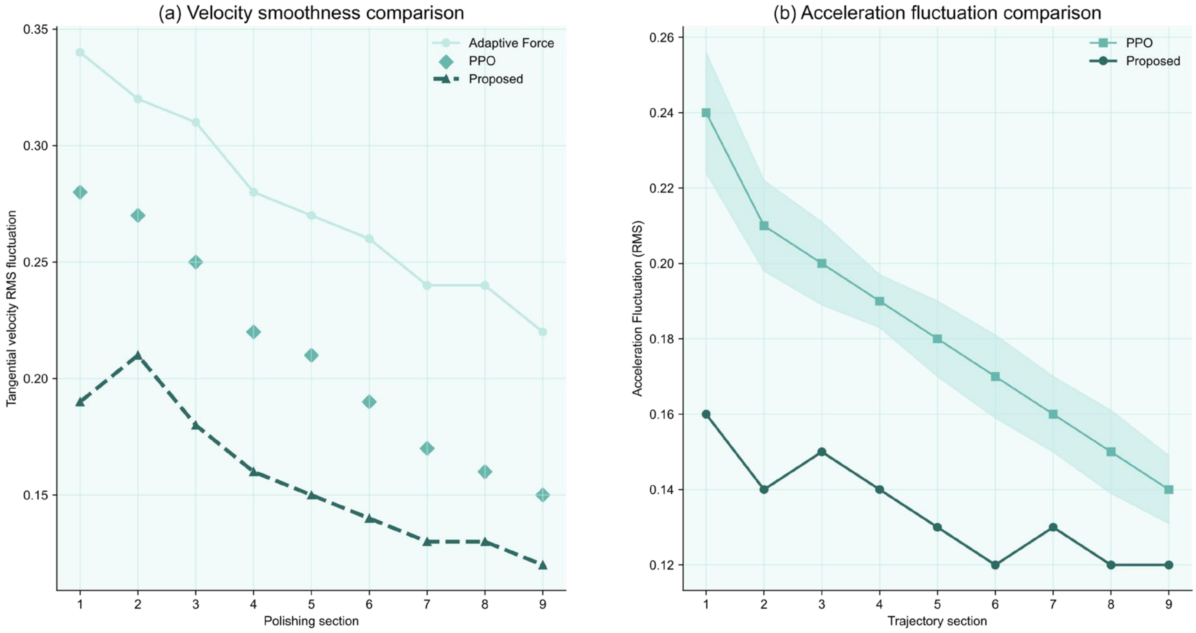

Figure 5. Trajectory smoothness and correction behavior. (a) Velocity smoothness comparison. (b) Acceleration fluctuation comparison.

The area next to a high-curvature shift of the polishing tool shows a comparatively greater improvement. Actor-critic approaches learn contact reward but often converge to a small route family, whereas pure diffusion policy creates a variety of trajectories but occasionally yields force-inconsistent options. Smooth motion and force observability are tightly associated with contact-sensitive manipulation, according to work on external-force estimate for robotic forceps [29]. Continuous force feedback can improve stability, but it cannot produce globally smooth trajectories on its own, according to research on force-feedback control of slave manipulators [30].

Figure 6 depicts the ablation research. As shown in Figure 6(a), the coverage decreases from 96.8% to 91.6% in the absence of the diffusion generator. Figure 6(b) shows that in the absence of critic supervision, the normal-force error is 4.1 N instead of 2.7 N. Mixed-reality and digital twin safety investigations have shown that contact-sensitive robotic systems benefit more from an integrated state evaluation than from separate perception and control modules [31]. Contact-value estimation and trajectory diversity are complementary but not interchangeable.

The frequency of contact loss is shown in Figure 6(c), with the biggest decrease occurring when the constraint penalty is eliminated. Runtime stability verification under sensing uncertainty is necessary for safety assurance research in collaborative robot production [32]. The policy convergence speed is shown in Figure 6(d), where critic guidance achieves a stable validation reward 38 epochs ahead of the unguided diffusion model. Additionally, safety-critical robot choices must be revised prior to unstable physical contact, according to safe collaborative assembly research [33].

Model Contribution and Deployment Feasibility

Figure 7 shows Deployment-oriented Robustness. Tool-wear disturbance is depicted in Figure 7(a); even with a 20% increase in the tool-wear coefficient, the suggested model retains 94.1% coverage. To some extent, nominal touch, reduced-speed contact, and dangerous contact are distinguished in robotics safety language [34]. In polishing, there are three types of contact: destructive, weak, and beneficial. Sensor noise disturbance is seen in Figure 7(b); the average force inaccuracy rises by just 0.6 N when 5% Gaussian noise is added to force sensing. Studies on decentralized robot control demonstrate that when the policy already embodies task coupling, learnt interaction regularities can simplify online optimization [35].

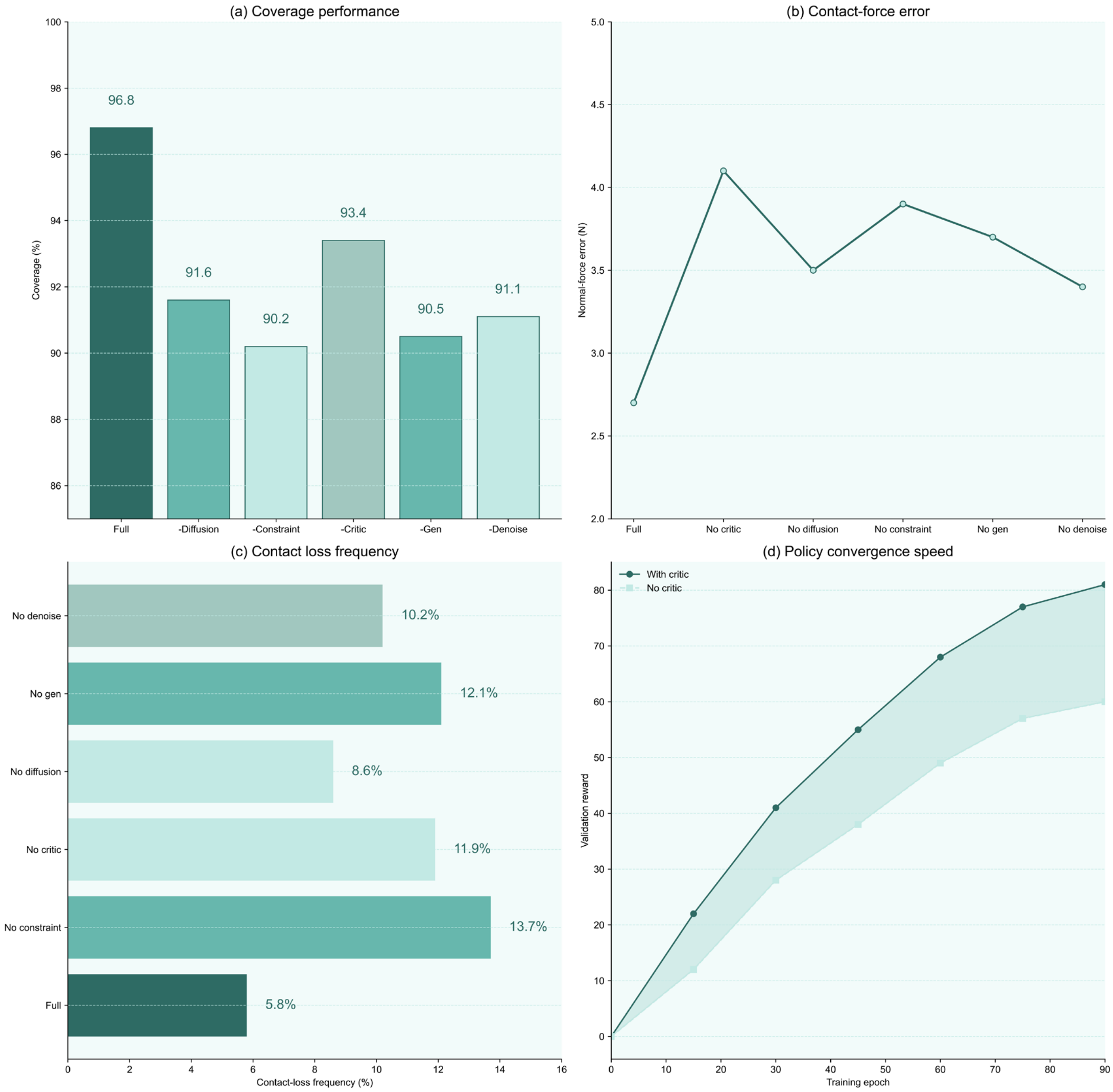

Figure 6. Ablation comparison of trajectory generation modules. (a) Coverage performance. (b) Contact-force error. (c) Contact loss frequency. (d) Policy convergence speed.

It may also be applied in real-world industrial settings. One 128-step trajectory segment is produced by the whole DiffusionActor-Critic pipeline in 42.6 ms on the workstation and 118.4 ms on the edge computer. Continuous-time smoothness requirements are necessary to prevent discontinuous route updates, according to studies on multi-robot trajectory optimization [36]. Multi-link robot motion exhibits linked temporal dependencies, according to graph-based trajectory prediction study [37]. Similar relationships between tool posture, surface normal, force, and polishing progress may also be learned via the present model's temporal state embedding, despite the fact that it is not graph-based.

Additionally, there are a few other flaws. Diffusion guiding may choose a locally reasonable but globally inefficient course if either of the two signals—measured force feedback and scanned surface geometry—is greatly skewed. Research on LSTM-based trajectory planning demonstrates that learnt planners do badly in dynamic situations that are not visible [38]. Before implementing the learnt strategy, feasibility studies of mobile manipulation advise evaluating it under different geometric deformations [39]. Category-level manipulation control demonstrates that while the feedback helps fix local mistakes, it does not accurately represent the situation [40]. The suggested model will be combined with adaptive tool-wear estimates and online surface reconstruction in future study.

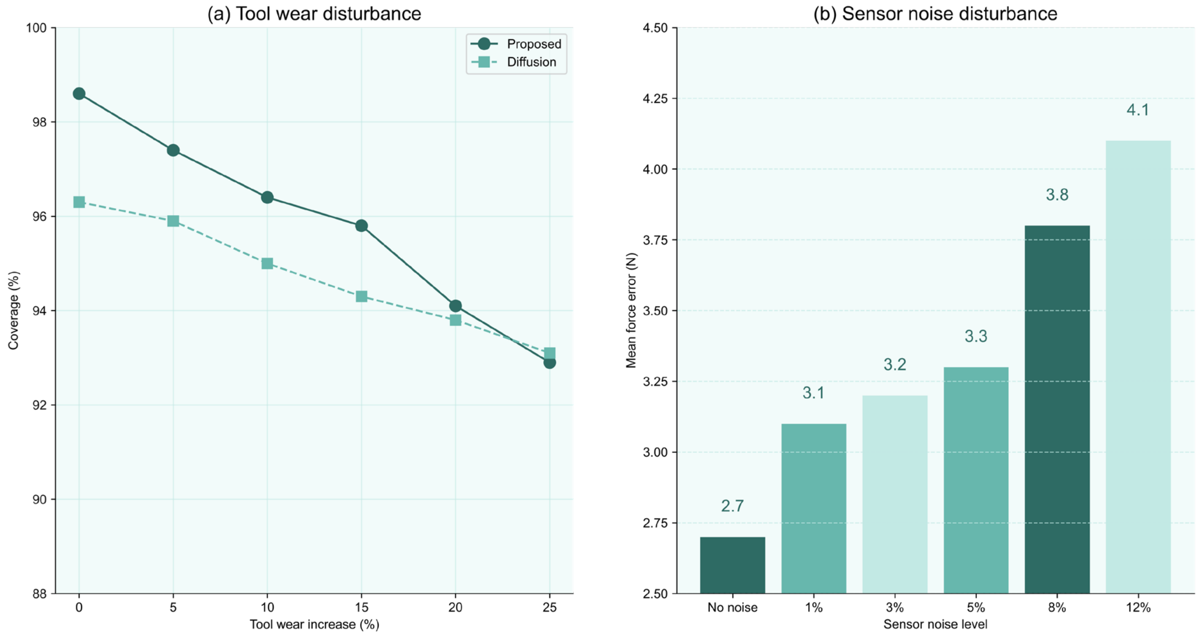

Figure 7. Deployment-oriented robustness evaluation. (a) Tool wear disturbance. (b) Sensor noise disturbance.

Conclusion

This research suggested a DiffusionActor-Critic model for contact-dense robot polishing trajectory creation. In this manner, contact geometry, force feedback, tool posture, surface curvature, roughness, and polishing history are all merged into a single contact-state sequence, and polishing is modeled as a restricted sequential decision problem. Actor-critic refinement guides denoising to a contact-stable and coverage-complete action, while diffusion sampling produces several candidate trajectories.

The main research finding requires both Critic-based Contact Evaluation and Generative Trajectory Diversity. Critic guidance steers clear of geometrically clean but physically impossible pathways, and Diffusion Generation can solve the model's local surface uncertainty issue. The method maintains a tolerable online inference latency under tool wear and sensor noise while reducing polishing coverage errors, normal-force mistakes, and contact-loss frequency, according to the experiment findings.

The framework must concurrently meet the requirements for smooth trajectory and contact stability in order to be used for high-precision surface finishing of robots. To improve the resilience of learning-based polishing systems in industrial applications, future research will incorporate online surface reconstruction, adaptive tool-wear identification, force-sensor uncertainty modeling, and hardware-level controller integration.

Author Contributions

Weronika Czarnecka contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Emilia Barańska contributes to methodology, software, validation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Mohsin, I., He, K., Li, Z., Du, R., Zhang, F., & Sun, Q. (2019). Path planning under force control in robotic polishing of the complex curved surfaces. Applied Sciences, 9(24), 5489. https://doi.org/10.3390/app9245489

[2] Huang, X., Wang, Z., Lin, Z., & Zhang, H. (2022). Movement modeling and control for robotic bonnet polishing. Chinese Journal of Mechanical Engineering, 35, 76. https://doi.org/10.1186/s10033-022-00751-y

[3] Mohsin, I., He, K., Li, Z., Zhang, F., & Du, R. (2020). Optimization of the polishing efficiency and torque by using Taguchi method and ANOVA in robotic polishing. Applied Sciences, 10(3), 824. https://doi.org/10.3390/app10030824

[4] Xiao, M., Zhang, T., Zou, Y., Chen, S., & Zhu, Y. (2020). Robotic constant force grinding control based on grinding model and iterative algorithm. Industrial Robot, 48(1), 109-118. https://doi.org/10.1108/IR-08-2020-0166

[5] Chang, G., Pan, R., Xie, Y., Yang, J., Yang, Y., & Li, J. (2022). Research on constant force polishing method of curved mold based on position adaptive impedance control. The International Journal of Advanced Manufacturing Technology, 122(2), 697-709. https://doi.org/10.1007/s00170-022-09943-1

[6] Li, H., Li, X., Wan, S., Wei, C., & Zhou, L. (2021). High-efficiency smooth pseudo-random path planning for restraining the path ripple of robotic polishing. Applied Optics, 60(20), 5987-5996. https://doi.org/10.1364/AO.426616

[7] Dai, S., Wang, X., Zhang, H., & Wen, B. (2020). Research on variation of grinding temperature of wind turbine blade robotic grinding. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 235(8), 1285-1297. https://doi.org/10.1177/0954405420972988

[8] Felsner, K., Schlachter, K., & Zambal, S. (2021). Robotic coverage path planning for ultrasonic inspection. Applied Sciences, 11(22), 10512. https://doi.org/10.3390/app112210512

[9] Beltran-Hernandez, C. C., Petit, D., Ramirez-Alpizar, I. G., Nishi, T., Kikuchi, S., Matsubara, T., & Harada, K. (2020). Learning force control for contact-rich manipulation tasks with rigid position-controlled robots. IEEE Robotics and Automation Letters, 5(4), 5709-5716. https://doi.org/10.1109/LRA.2020.3010739

[10] Khader, S. A., Yin, H., Falco, P., & Kragic, D. (2020). Data-efficient model learning and prediction for contact-rich manipulation tasks. IEEE Robotics and Automation Letters, 5(3), 4321-4328. https://doi.org/10.1109/LRA.2020.2996067

[11] Khader, S. A., Yin, H., Falco, P., & Kragic, D. (2021). Stability-guaranteed reinforcement learning for contact-rich manipulation. IEEE Robotics and Automation Letters, 6(1), 1-8. https://doi.org/10.1109/LRA.2020.3028529

[12] Yang, Q., Durr, A., Topp, E. A., & Stork, J. A. (2022). Variable impedance skill learning for contact-rich manipulation. IEEE Robotics and Automation Letters, 7(4), 8391-8398. https://doi.org/10.1109/LRA.2022.3187276

[13] Davchev, T., Luck, K. S., Burke, M., Meier, F., Schaal, S., & Vijayakumar, S. (2022). Residual learning from demonstration: Adapting DMPs for contact-rich manipulation. IEEE Robotics and Automation Letters, 7(2), 4488-4495. https://doi.org/10.1109/LRA.2022.3150024

[14] Si, W., Guan, Y., Wang, N., & Zhang, H. (2022). Adaptive compliant skill learning for contact-rich manipulation with human in the loop. IEEE Robotics and Automation Letters, 7(2), 3917-3924. https://doi.org/10.1109/LRA.2022.3159163

[15] Hoppe, S., Lou, Z., Hennes, D., Toussaint, M., & Neumann, G. (2019). Planning approximate exploration trajectories for model-free reinforcement learning in contact-rich manipulation. IEEE Robotics and Automation Letters, 4(4), 4041-4048. https://doi.org/10.1109/LRA.2019.2928212

[16] Fang, B., Jia, S., Guo, D., Xu, M., Wen, S., & Sun, F. (2019). Survey of imitation learning for robotic manipulation. International Journal of Intelligent Robotics and Applications, 3, 362-369. https://doi.org/10.1007/s41315-019-00103-5

[17] Morales, E. F., Murrieta-Cid, R., Becerra, I., & Esquivel-Basaldua, M. A. (2021). A survey on deep learning and deep reinforcement learning in robotics with a tutorial on deep reinforcement learning. Intelligent Service Robotics, 14, 773-805. https://doi.org/10.1007/s11370-021-00398-z

[18] Wang, H. N., Liu, N., Zhang, Y. Y., Feng, D. W., Huang, F., Li, D. S., et al. (2020). Deep reinforcement learning: A survey. Frontiers of Information Technology & Electronic Engineering, 21, 1726-1744. https://doi.org/10.1631/FITEE.1900533

[19] Wang, C., Zhang, Q., Tian, Q., Li, S., Wang, X., Lane, D., et al. (2020). Learning mobile manipulation through deep reinforcement learning. Sensors, 20(3), 939. https://doi.org/10.3390/s20030939

[20] Liu, Y., Zha, F., Sun, L., Li, J., Wang, W., & Yang, C. (2019). Learning articulated constraints from a one-shot demonstration for robot manipulation planning. IEEE Access, 7, 162730-162742. https://doi.org/10.1109/ACCESS.2019.2953894

[21] Luis, C. E., Vukosavljev, M., & Schoellig, A. P. (2020). Online trajectory generation with distributed model predictive control for multi-robot motion planning. IEEE Robotics and Automation Letters, 5(2), 604-611. https://doi.org/10.1109/LRA.2020.2964159

[22] Paparusso, L., Kashiri, N., & Tsagarakis, N. G. (2020). A disturbance-aware trajectory planning scheme based on model predictive control. IEEE Robotics and Automation Letters, 5(4), 5779-5786. https://doi.org/10.1109/LRA.2020.3010751

[23] Wang, L., Wu, Q., Lin, F., Li, S., & Chen, D. (2019). A new trajectory-planning beetle swarm optimization algorithm for trajectory planning of robot manipulators. IEEE Access, 7, 154331-154345. https://doi.org/10.1109/ACCESS.2019.2949271

[24] Kazim, I. J., Tan, Y., & Li, R. (2022). Comparison study of the PSO and SBPSO on universal robot trajectory planning. Applied Sciences, 12(3), 1518. https://doi.org/10.3390/app12031518

[25] Miao, X., Fu, H., & Song, X. (2022). Research on motion trajectory planning of the robotic arm of a robot. Artificial Life and Robotics, 27, 677-684. https://doi.org/10.1007/s10015-022-00779-2

[26] Qie, X., Kang, C., Zong, G., Chen, S., & Wang, J. (2022). Trajectory planning and simulation study of redundant robotic arm for upper limb rehabilitation based on back propagation neural network and genetic algorithm. Sensors, 22(11), 4071. https://doi.org/10.3390/s22114071

[27] Xue, X., Huang, H., Zuo, L., Wang, N., & Liu, Y. (2022). A compliant force control scheme for industrial robot interactive operation. Frontiers in Neurorobotics, 16, 865187. https://doi.org/10.3389/fnbot.2022.865187

[28] Zhang, Y., Hua, C., Qian, J., & Guan, X. (2019). Adaptive robust visual servoing/force control for robot manipulator with dead-zone input. IEEE Access, 7, 106375-106385. https://doi.org/10.1109/ACCESS.2019.2930074

[29] Zhou, D., Tadano, K., & Haraguchi, D. (2020). Motion control and external force estimation of a pneumatically driven multi-DOF robotic forceps. Applied Sciences, 10(11), 3679. https://doi.org/10.3390/app10113679

[30] Shi, H., Zhang, B., Mei, X., Song, Q., & Li, J. (2021). Realization of force detection and feedback control for slave manipulator of master/slave surgical robot. Sensors, 21(22), 7489. https://doi.org/10.3390/s21227489

[31] Choi, S., Park, K. B., Roh, D. H., Lee, J. Y., Mohammed, M., Ghasemi, Y., & Jeong, H. (2022). An integrated mixed reality system for safety-aware human-robot collaboration using deep learning and digital twin generation. Robotics and Computer-Integrated Manufacturing, 73, 102258. https://doi.org/10.1016/j.rcim.2021.102258

[32] Bi, Z., Luo, C., Miao, Z., Zhang, B., Zhang, W. J., & Wang, L. (2021). Safety assurance mechanisms of collaborative robotic systems in manufacturing. Robotics and Computer-Integrated Manufacturing, 67, 102022. https://doi.org/10.1016/j.rcim.2020.102022

[33] Gopinath, V., Johansen, K., Derelov, M., Gustafsson, A., Axelsson, S., & Wagner, S. (2021). Safe collaborative assembly on a continuously moving line with large industrial robots. Robotics and Computer-Integrated Manufacturing, 67, 102048. https://doi.org/10.1016/j.rcim.2020.102048

[34] Vicentini, F. (2020). Terminology in safety of collaborative robotics. Robotics and Computer-Integrated Manufacturing, 63, 101921. https://doi.org/10.1016/j.rcim.2019.101921

[35] Zhang, L., Sun, Y., Barth, A., & Ma, O. (2020). Decentralized control of multi-robot system in cooperative object transportation using deep reinforcement learning. IEEE Access, 8, 184109-184119. https://doi.org/10.1109/ACCESS.2020.3025287

[36] Kandhasamy, S., Kuppusamy, V. B., & Krishnan, S. (2020). Scalable decentralized multi-robot trajectory optimization in continuous-time. IEEE Access, 8, 172379-172392. https://doi.org/10.1109/ACCESS.2020.3024685

[37] Wu, H., Li, X., Yang, X., & Wang, T. (2020). Joint trajectory prediction of multi-linkage robot based on graph convolutional network. IEEE Access, 8, 226230-226241. https://doi.org/10.1109/ACCESS.2020.3042241

[38] Molina-Leal, A., Gomez-Espinosa, A., Escobedo Cabello, J. A., & Cuan-Urquizo, E. (2021). Trajectory planning for a mobile robot in a dynamic environment using an LSTM neural network. Applied Sciences, 11(22), 10689. https://doi.org/10.3390/app112210689
[39] Honerkamp, D., Welschehold, T., & Valada, A. (2021). Learning kinematic feasibility for mobile manipulation through deep reinforcement learning. IEEE Robotics and Automation Letters, 6(4), 6289-6296. https://doi.org/10.1109/LRA.2021.3092685
[40] Gao, W., & Tedrake, R. (2021). kPAM 2.0: Feedback control for category-level robotic manipulation. IEEE Robotics and Automation Letters, 6(2), 2962-2969. https://doi.org/10.1109/LRA.2021.3062315