This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

SurgVLA-Struct: Structured Surgical Reward Modeling for Across-Task Vision-Language Robotic Manipulation

Erli Zhang, Shuojue Yang, Daiyun Shen, Zhengyang Li, and Yueming Jin*

Department of Biomedical Engineering, National University of Singapore, Singapore
ymjin@nus.edu.sg

Abstract. Vision-language-action (VLA) models offer a promising paradigm for surgical robot autonomy by mapping visual observations and language instructions to continuous control. However, existing surgical VLA systems remain limited in manipulation complexity and often suffer from instability and compounding errors in long-horizon settings, particularly when trained solely on demonstrations. Most are optimized for single procedural objectives rather than unified control across diverse procedures. To address this gap, we present **SurgVLA-Struct**, a reinforcement-refined framework for surgical robotic manipulation across tasks in the SurRoL simulator. Our approach comprises three key components: (1) a structured 10-task benchmark with 7 newly introduced language-conditioned variants incorporating obstacle constraints, extended horizons, and attribute-driven instructions; (2) a unified continuous-control backbone with chunked 7-DoF action prediction under a shared action space, enabling stable and consistent language-conditioned manipulation without task-specific decoders; and (3) structured surgical reward modeling within a GRPO framework using task-invariant, simulator-verifiable objectives encoding precision, efficiency, trajectory stability, and contact safety. Extensive experiments show that SurgVLA-Struct consistently outperforms classical reinforcement learning baselines and state-of-the-art VLA models, validating stable language-conditioned manipulation across diverse tasks.

Keywords: Surgical Data Science · Vision-Language-Action · Reinforcement Learning · Surgical Robot Learning · Across-Task Manipulation

1 Introduction

Robot-assisted minimally invasive surgery has enhanced surgical precision and dexterity in complex procedures through improved motion control and tremor reduction [14,17,7,23]. However, most platforms remain teleoperated, and scalable autonomy for multi-step surgical manipulation remains unresolved [10,18]. A key step toward surgical autonomy is developing end-to-end systems that translate high-level surgeon instructions into executable robot actions. Such systems must

* Corresponding author.

generalize across diverse manipulation tasks, reason over long horizons, and enforce safety-aware interaction under precision-critical and contact-sensitive constraints [15,16,22]. Enabling instruction-driven, multi-task surgical manipulation therefore requires tightly integrated vision, language, and articulated control.

Traditional surgical robot learning has largely relied on imitation learning (IL) and reinforcement learning (RL). RL-based approaches explore multi-task manipulation under predefined task interfaces [20,12,4], improving policy generalization across simulated procedures, but typically depend on manually specified goals and engineered rewards, and lack native language conditioning. Procedure-specific systems such as SutureBot [5] achieve strong autonomous performance within defined tasks but do not generalize across diverse manipulation settings. In parallel, recent vision-language-action (VLA) models [24,9,3,2] enable end-to-end instruction-conditioned control across diverse robotic tasks, yet surgical vision-language systems [11,13] primarily focus on navigation, workflow assistance, or scene understanding rather than fine-grained articulated manipulation. Together, these limitations motivate a unified framework integrating language-conditioned reasoning with robust, multi-task surgical manipulation under long-horizon and precision-critical constraints.

To address this gap, we construct a structured benchmark of 10 manipulation tasks based on SurRoL [21], with 7 expanded scenarios introducing obstacle constraints, extended horizons, and attribute-driven instructions to better approximate constraint-sensitive surgical manipulation. Building on this setting, we propose **SurgVLA-Struct**, a unified VLA framework that adopts a shared continuous action space with chunked action prediction, enabling stable language-conditioned control across tasks without task-specific decoders. However, imitation learning alone aligns actions to demonstrations without explicitly optimizing execution efficiency or collision-aware behavior, particularly under extended interaction horizons. We therefore introduce structured reward modeling within a GRPO-based reinforcement framework [19], incorporating simulator-computable objectives that penalize redundant motion and unsafe contact while rewarding precise task completion and trajectory stability. Extensive experiments demonstrate that SurgVLA-Struct consistently outperforms classical RL baselines and state-of-the-art VLA models in across-task surgical manipulation.

2 Method

2.1 Problem Formulation and Multi-Task Data Curation

We study instruction-driven surgical robotic manipulation in a unified multi-task setting, where a single shared policy is jointly trained across all tasks under a common observation and action interface. At time step t, the robot observes an RGB image o_t and receives a natural-language instruction l. The policy π_θ predicts articulated control commands conditioned on both inputs. Instead of a single action, it outputs a fixed-length action chunk of size $K = 8$:

$$a_{t:t+K-1} = \pi_\theta(o_t, l), \tag{1}$$

where each $a_t \in \mathbb{R}^7$ denotes a 7-degree-of-freedom (7-DoF) end-effector command. During execution, only the first predicted action is applied before replanning at the next timestep, forming a receding-horizon control loop that improves stability and responsiveness, reducing error accumulation in long-horizon and contact-sensitive manipulation.

We curate a structured benchmark on SurRoL [21] comprising 10 language-conditioned manipulation tasks: 3 original tasks and 7 new variants increasing complexity. Expand variants introduce obstacle-aware spatial constraints; Place tasks impose precision-critical long-horizon control (e.g., gauze-based hemostatic stabilization); PegTransferColor and Location variants require grounding object attributes in natural language. Together, these extensions enforce long-horizon dependency, collision sensitivity, and instruction-specific action selection.

For each task, 500 oracle demonstrations are generated under randomized object placement, obstacle configurations, and target variations (5,000 total). We use 400 per task for training and 100 for testing, with non-overlapping object configurations and initializations between the two splits. Rollouts capture RGB observations, 7-DoF actions, and proprioceptive states, stored in RLDS format for reproducible VLA training.

2.2 Pipeline Overview

As illustrated in Fig. 1, SurgVLA-Struct adopts a two-phase training paradigm combining supervised alignment and reinforcement refinement. RGB observations, language instructions, and proprioceptive states are fused by a vision-language backbone, while an action head predicts $K = 8$ future actions optimized with an L_1 loss against oracle demonstrations.

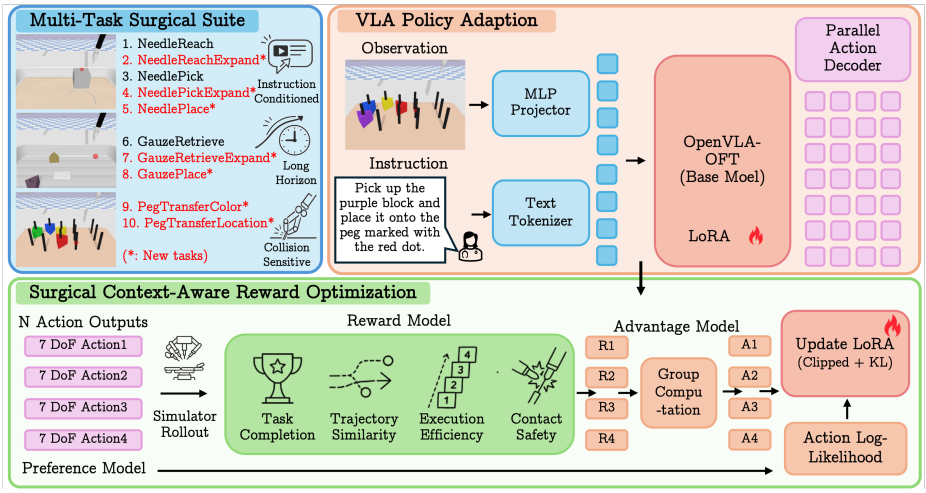


Fig. 1. Overview of the proposed SurgVLA-Struct Framework.

We then refine the policy using Group Relative Policy Optimization (GRPO), initialized from the supervised checkpoint. Only LoRA parameters are updated

while the pretrained backbone remains frozen, preserving multimodal alignment and stabilizing optimization. Multiple rollouts from identical initial states form trajectory groups for relative advantage estimation.

2.3 Vision-Language-Action Learning and Supervised Alignment

We adopt OpenVLA-OFT [8] as the base architecture, which replaces tokenized action prediction with parallel continuous regression to directly output real-valued commands, avoiding quantization artifacts and preserving the fine-grained control resolution needed for precision-sensitive, contact-aware surgical manipulation. The shared continuous regression head further enables consistent control across heterogeneous task dynamics without task-specific output branches, supporting stable multi-step prediction under long-horizon and constraint-sensitive settings.

Given fused multimodal features $\mathbf{z}_t = f_{\text{fusion}}(o_t, l)$, an action head predicts the mean action $\boldsymbol{\mu}_t = f_\phi(\mathbf{z}_t)$. We optimize the policy using supervised imitation:

$$\mathcal{L}_{\text{sup}}(\phi) = \mathbb{E}\big[\|\mathbf{a}_t - \mathbf{a}_t^{\text{gt}}\|_1\big], \tag{2}$$

where $\mathbf{a}_t^{\text{gt}}$ denotes oracle actions. This stage establishes stable alignment between perception, language understanding, and continuous articulated control, providing a strong initialization for subsequent reinforcement refinement.

2.4 Surgically Structured Reward Design for GRPO Refinement

While supervised imitation aligns trajectories, it does not optimize execution quality under contact-sensitive and long-horizon constraints. We therefore refine the policy using Group Relative Policy Optimization (GRPO), which leverages group-relative advantages from identical initial states to promote outcome-oriented improvement and stabilize updates in sparse-reward settings.

Let $\mathcal{G}$ denote rollout groups, where each group $g \in \mathcal{G}$ contains 8 trajectories generated from the same initial state under the previous policy $\pi_{\phi_{\text{old}}}$, each executed for up to 100 environment steps. For each trajectory τ^g, we compute its cumulative reward and define the group-relative advantage $A^g = R^g - \bar{R}^g$, where $\bar{R}^g$ is the mean reward within group g. The resulting scalar advantage is broadcast uniformly to every timestep, and to every action within the corresponding 8-step chunk, of that trajectory. During reinforcement refinement, we parameterize the stochastic policy as a diagonal Gaussian distribution to enable likelihood-based updates. The importance ratio is then computed using Gaussian action likelihoods:

$$r_t^g(\phi) = \frac{\pi_\phi(\mathbf{a}_t \mid o_t, l)}{\pi_{\phi_{\text{old}}}(\mathbf{a}_t \mid o_t, l)} \tag{3}$$

Policy variance, clipping, and KL regularization follow the default GRPO configuration of our OpenVLA-OFT-based implementation, unmodified in this work; at test time the policy executes the deterministic mean action $\boldsymbol{\mu}_t$, consistent with the supervised action head.

Training-Time Surgical Rewards. Standard binary success signals are insufficient for surgical manipulation, where precision, efficiency, and contact regulation are critical. We therefore design four complementary simulator-verifiable reward components used exclusively during training within the simulator. These rewards rely on simulator state access and are not required at inference time.

(1) Task Completion. We use a binary success reward to ensure outcome-level supervision:

$$R_{\text{completion}} = \mathbf{1}[\text{task success}].$$

(2) Trajectory Similarity. To encourage stable motion patterns without enforcing strict imitation, we measure deviation from oracle actions:

$$R_{\text{similarity}} = \exp\left(-\frac{1}{m}\sum_{t=1}^{m}\|\mathbf{a}_t - \mathbf{a}_t^{\text{gt}}\|_2^2\right),$$

where $m = \min(T, T_{\text{gt}})$, with T the rollout length and T_{gt} the oracle length. This term promotes stable motion without penalizing shorter yet successful rollouts.

(3) Execution Efficiency. In surgical procedures, unnecessary instrument motion and prolonged manipulation increase operative time, cognitive load, and risk exposure. We therefore penalize prolonged execution beyond oracle duration:

$$R_{\text{efficiency}} = \exp\left(-\frac{\max(0,\, T - T_{\text{gt}})}{T_{\text{gt}}}\right).$$

Beyond merely matching oracle timing, this reward encourages temporally efficient and purposeful motion, discouraging dithering, hesitation, or redundant corrective actions. By explicitly modeling execution compactness, the policy is guided toward stable long-horizon behaviors that achieve task objectives with controlled and deliberate motion, aligning with real surgical workflow constraints.

(4) Contact Safety. Delicate surgical environments require strict regulation of interaction forces to prevent unintended tissue trauma or excessive pressure on surrounding structures. We penalize accumulated contact force during execution:

$$R_{\text{contact}} = \exp\left(-\frac{F_{\text{total}}}{F_0}\right),$$

where F_{total} aggregates contact forces and F_0 is the mean oracle contact force. This reward introduces force-awareness into policy learning, encouraging controlled interaction and collision-sensitive planning rather than purely outcome-driven behavior. By guiding contact levels toward expert demonstrations, this term promotes safety-conscious manipulation aligned with surgical practice.

All reward components are task-invariant and defined under a shared simulator interface, ensuring consistent optimization signals across tasks without task-specific reward engineering. The total training reward is an unweighted sum, $R = R_{\text{completion}} + R_{\text{similarity}} + R_{\text{efficiency}} + R_{\text{contact}}$, treating all four components equally rather than tuning task-specific coefficients. Together, these objectives provide physically grounded feedback modeling task success, motion stability, efficiency, and contact regulation, enabling transferable policy learning across diverse surgical tasks.

3 Experiments and Results

3.1 Experiment Setting

Task Construction We evaluate on 10 language-conditioned single-PSM tasks in SurRoL [21], including 7 new obstacle-aware, long-horizon, and attribute-driven variants. Each method is evaluated over 100 randomized episodes per task averaged across 5 seeds; results are reported as mean±std success rate (%), for which the corresponding 95% confidence interval is approximately $\pm 1.24\times$ the reported standard deviation ($t_{0.025,4}/\sqrt{5}$).

Implementation Details Stage I initializes from OpenVLA-7B and applies LoRA fine-tuning (rank 32) using AdamW with learning rate 5×10^{-4}, batch size 4, for 50,000 steps. Stage II performs GRPO refinement for another 50,000 steps with the same learning rate, updating LoRA parameters only. All experiments are implemented and trained in PyTorch on a single NVIDIA A6000 GPU. Code and task configurations will be made publicly available to support reproducibility.

3.2 Results

We compare SurgVLA-Struct with classical RL baselines (BC [1], DEX [6], and GC-GPT-DT [4]), multiple OpenVLA [9] training variants (full fine-tuning, LoRA-based supervised fine-tuning, and GRPO-only optimization), and recent VLA foundation models, including $\pi_{0.5}$ [3] and GR00T [2]. $\pi_{0.5}$ and GR00T are initialized from their official publicly released pretrained checkpoints, using each model's default image and proprioceptive encoders without architectural modification. All VLA-based methods, including $\pi_{0.5}$ and GR00T, are fine-tuned on the same 400-trajectory training set via LoRA (rank 32, matching Stage I) applied to the default target modules of each official implementation, under matched training budgets; we note that applying identical LoRA rank across architecturally distinct backbones is a simplification adopted for controlled comparison and may not equalize each baseline's achievable performance ceiling. RL baselines use visual and state inputs only, whereas VLA-based models additionally condition on language instructions. The PegTransferColor and Location tasks require attribute-specific language grounding and cannot be evaluated for RL baselines; GC-GPT-DT additionally relies on structured goal conditioning that is incompatible with free-form attribute grounding in these tasks.

As shown in Table 1 and Table 2, SurgVLA-Struct achieves consistent performance across all tasks under a single shared policy, maintaining higher success rates than RL baselines under obstacle constraints, extended horizons, and attribute-driven instructions. Compared to $\pi_{0.5}$ and GR00T, SurgVLA-Struct shows qualitatively more conservative, contact-aware behavior as task complexity increases, particularly on obstacle-present Expand variants, consistent with the contact-safety reward term. Gains are consistent across five seeds (std $\leq$2%), and the shared policy improves without task-specific decoders or degradation on

Table 1. Mean±Std success rate (%) on Needle manipulation tasks (5 seeds). *denotes newly introduced tasks.

Model	Needle Reach	Needle ReachExpand*	Needle Pick	Needle PickExpand*	Needle Place*
BC [1]	100.0 ± 0.0	97.2 ± 0.8	22.4 ± 1.6	20.2 ± 1.8	15.4 ± 1.6
DEX [6]	100.0 ± 0.0	100.0 ± 0.0	92.4 ± 0.9	85.2 ± 1.6	76.4 ± 1.8
GC-GPT-DT [4]	100.0 ± 0.0	100.0 ± 0.0	97.4 ± 0.9	91.2 ± 0.8	85.4 ± 1.7
OpenVLA [9]	100.0 ± 0.0	100.0 ± 0.0	90.2 ± 1.7	87.4 ± 1.8	77.2 ± 1.6
OpenVLA-OFT only	100.0 ± 0.0	100.0 ± 0.0	92.2 ± 0.8	90.4 ± 0.9	78.2 ± 1.8
OpenVLA-GRPO only	96.4 ± 0.9	95.2 ± 1.6	87.4 ± 1.8	89.2 ± 1.7	83.4 ± 1.6
$\pi_{0.5}$ [3]	100.0 ± 0.0	100.0 ± 0.0	93.2 ± 0.8	90.4 ± 1.6	81.2 ± 1.8
GR00T [2]	100.0 ± 0.0	98.4 ± 0.9	89.2 ± 1.6	87.4 ± 1.8	76.2 ± 1.7
Ours	$\mathbf{100.0 \pm 0.0}$	$\mathbf{100.0 \pm 0.0}$	$\mathbf{99.2 \pm 0.8}$	$\mathbf{93.4 \pm 0.9}$	$\mathbf{89.4 \pm 0.8}$

Table 2. Mean±Std success rate (%) on Gauze and language-conditioned Peg manipulation tasks (5 seeds). *denotes newly introduced tasks; – indicates inapplicable baselines.

Model	Gauze Retrieve	Gauze RetrieveExpand*	Gauze Place*	PegTransfer Color*	PegTransfer Location*
BC [1]	7.2 ± 1.6	4.4 ± 0.9	1.2 ± 0.8	–	–
DEX [6]	73.4 ± 1.8	70.2 ± 1.6	63.4 ± 1.8	–	–
GC-GPT-DT [4]	95.2 ± 0.8	91.4 ± 1.6	80.2 ± 1.8	–	–
OpenVLA [9]	82.4 ± 1.8	80.2 ± 1.7	78.4 ± 1.6	10.2 ± 1.8	7.4 ± 1.6
OpenVLA-OFT only	82.2 ± 1.7	81.4 ± 0.9	80.4 ± 1.6	11.2 ± 1.8	8.4 ± 1.6
OpenVLA-GRPO only	78.4 ± 1.8	83.2 ± 1.6	81.4 ± 1.7	4.2 ± 0.8	2.4 ± 0.9
$\pi_{0.5}$ [3]	85.2 ± 1.8	86.4 ± 1.6	80.2 ± 1.7	18.4 ± 1.8	16.2 ± 1.6
GR00T [2]	81.4 ± 1.6	80.2 ± 1.8	75.4 ± 1.7	8.2 ± 1.8	6.4 ± 1.6
Ours	$\mathbf{95.4 \pm 0.9}$	$\mathbf{93.2 \pm 0.8}$	$\mathbf{91.4 \pm 0.9}$	$\mathbf{26.2 \pm 1.6}$	$\mathbf{25.4 \pm 1.8}$

simpler tasks when jointly trained with harder ones, suggesting limited negative transfer.

On simpler tasks (e.g., NeedleReach and GauzeRetrieve), most methods approach saturated performance, but pure imitation grows increasingly prone to failure as difficulty rises: Expand variants require collision-sensitive navigation where baselines frequently make unintended contact; Place tasks demand temporally extended precision where error accumulation causes late-stage failures; and PegTransfer variants require precise attribute grounding and strict instruction-following, where most baselines fail. Structured surgical reward modeling refines the aligned policy by penalizing redundant motion and unsafe contact while rewarding precise, stable completion, enabling robust manipulation across diverse tasks (Fig. 2).

3.3 Ablation Studies

Table 3 evaluates each reward component's contribution on four challenging tasks. Binary success improves over SFT-only training, confirming the value of explicit outcome supervision; trajectory similarity yields modest further gains by stabilizing motion; execution-length regularization substantially improves complex multi-stage tasks, particularly PegTransfer variants, where inefficient motion otherwise accumulates and degrades stability; and collision-aware penalties

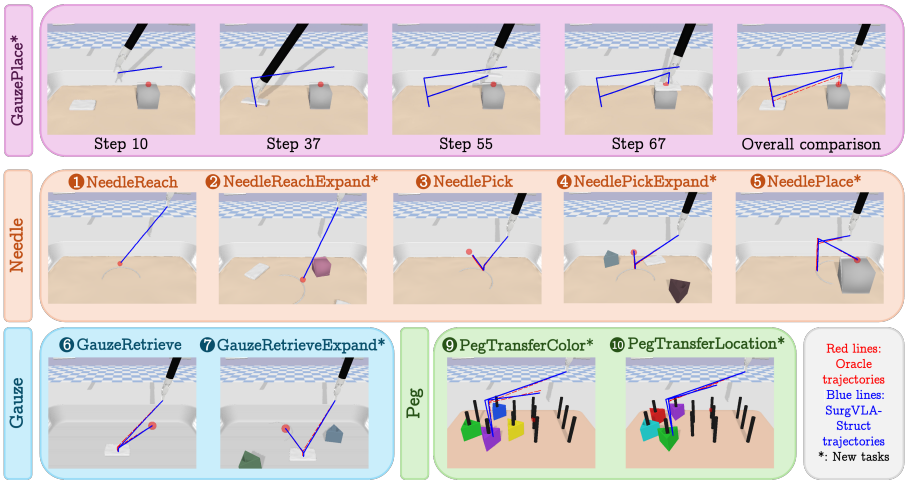


Fig. 2. Qualitative trajectory comparison across 10 SurRoL tasks. We illustrate intermediate steps for GauzePlace as a representative long-horizon example. Red lines denote oracle trajectories and blue lines denote SurgVLA-Struct generated trajectories.

give a further boost, underscoring the importance of explicit contact regulation. Together, these results show that while SFT provides cross-modal grounding, task-invariant structured rewards are essential for stable and physically consistent behavior across tasks.

Table 3. Ablation on reward design. Each row cumulatively adds the next reward term; incremental gain (%) over the preceding row shown in parentheses.

Method	Needle Place	Gauze Place	PegTransfer Color	PegTransfer Location
SFT only (OpenVLA-OFT)	77	78	10	7
SFT + GRPO (+ Success)	80 (+3↑)	80 (+2↑)	13 (+3↑)	10 (+3↑)
SFT + GRPO (+ Similarity)	82 (+2↑)	81 (+1↑)	14 (+1↑)	11 (+1↑)
SFT + GRPO (+ Length)	85 (+3↑)	85 (+4↑)	20 (+6↑)	18 (+7↑)
SFT + GRPO (+ Collision, Full)	89 (+4↑)	91 (+6↑)	26 (+6↑)	25 (+7↑)

4 Conclusion

We present **SurgVLA-Struct**, a unified VLA framework for surgical robotic manipulation across tasks, together with a structured SurRoL benchmark featuring expanded language-conditioned variants modeling extended temporal structure, obstacle-aware interaction, and constraint-sensitive control. SurgVLA-Struct employs a shared continuous-action backbone trained under a single policy and refined through task-invariant structured reward modeling within a GRPO framework, explicitly encoding execution efficiency and contact-aware regulation beyond imitation. Across 10 tasks under matched budgets, SurgVLA-Struct consistently outperforms classical RL baselines and strong VLA models, demonstrating that unified continuous control with structured surgical reward refinement enables stable and scalable instruction-driven surgical manipulation in simulation.

Limitations and Future Work. The current evaluation is conducted entirely within simulation, on a single-PSM setup; reported success rates, and the interaction and efficiency behaviors they reflect, should be read as simulator-scoped rather than as evidence of clinical safety or utility. Two of the four reward terms, trajectory similarity and execution efficiency, are computed relative to oracle demonstrations, so reward computation is not fully decoupled from the imitation data; exploring oracle-free formulations is left to future work. Our framework builds on SurRoL, a dVRK-compatible surgical simulation platform previously used in sim-to-real surgical robot learning studies [6], providing a foundation for future physical transfer. Excluded SurRoL tasks involve bimanual settings beyond the unified single-arm action space or lack oracle policies required for Stage-I SFT initialization. Future work will investigate domain randomization, camera variation adaptation, and validation on physical dVRK platforms to bridge the sim-to-real gap and assess real-world deployment feasibility.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bain, M., Sammut, C.: A framework for behavioural cloning. Machine Intelligence **15** (1995)
2. Bjorck, J., Castañeda, F., Cherniadev, N., Da, X., Ding, R., Fan, L., Fang, Y., Fox, D., Hu, F., Huang, S., Jang, J., Jiang, Z., Kautz, J., Kundalia, K., Lao, L., Lin, Z., Lin, Z., Liu, G., Llontop, E., Magne, L., Mandlekar, A., Narayan, A., Nasiriany, S., Reed, S., Tan, Y.L., Wang, G., Wang, Z., Wang, J., Wang, Q., Xiang, J., Xie, Y., Xu, Y., Xu, Z., Ye, S., Yu, Z., Zhang, A., Zhang, H., Zhao, Y., Zheng, R., Zhu, Y.: Gr00t n1: An open foundation model for generalist humanoid robots (2025)
3. Black, K., Brown, N., Darpinian, J., Dhabalia, K., Driess, D., Esmail, A., Equi, M.R., Finn, C., Fusai, N., Galliker, M.Y., Shi, L.X., Smith, L., Springenberg, J.T., Stachowicz, K., Tanner, J., Vuong, Q., Walke, H., Walling, A., Wang, H., Yu, L., Zhilinsky, U.: $\pi_{0.5}$: a vision-language-action model with open-world generalization. In: Proceedings of The 9th Conference on Robot Learning. Proceedings of Machine Learning Research, vol. 305, pp. 17–40. PMLR (2025)
4. Fu, J., Long, Y., Chen, K., Wei, W., Dou, Q.: Multi-objective cross-task learning via goal-conditioned gpt-based decision transformers for surgical robot task automation. In: IEEE International Conference on Robotics and Automation (ICRA) (2024)
5. Haworth, J., Chen, J.T., Nelson, N., Kim, J.W., Moghani, M., Finn, C., Krieger, A.: Suturebot: A precision framework and benchmark for autonomous end-to-end suturing. In: Advances in Neural Information Processing Systems (NeurIPS) (2025)
6. Huang, T., Chen, K., Li, B., Liu, Y., Dou, Q.: Demonstration-guided reinforcement learning with efficient exploration for task automation of surgical robot. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 4640–4647 (2023)
7. Iftikhar, M.: Artificial intelligence: Revolutionizing robotic surgery. Annals of Medicine and Surgery **74**, 101941 (2024)

8. Kim, M.J., Finn, C., Liang, P.: Fine-tuning vision-language-action models: Optimizing speed and success. arXiv preprint arXiv:2502.19645 (2025)
9. Kim, M.J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E.P., Sanketi, P.R., Vuong, Q., Kollar, T., Burchfiel, B., Tedrake, R., Sadigh, D., Levine, S., Liang, P., Finn, C.: Openvla: An open-source vision-language-action model. In: Proceedings of The 8th Conference on Robot Learning. Proceedings of Machine Learning Research, vol. 270, pp. 2679–2713. PMLR (2025)
10. Lee, A., Baker, T., Bederson, J., Rapoport, B.: Levels of autonomy in fda-cleared surgical robots: A systematic review. npj Digital Medicine **7**, 60 (2024)
11. Li, S., Wang, J., Dai, R., Ma, W., Ng, W.Y., Hu, Y., Li, Z.: Robonurse-vla: Robotic scrub nurse system based on vision-language-action model. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3986–3993 (2025)
12. Long, Y., Wei, W., Huang, T., Wang, Y., Dou, Q.: Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning. IEEE Robotics and Automation Letters (2023)
13. Ng, C.K., Bai, L., Wang, G., Wang, Y., Gao, H., Yuan, K., Jin, C., Zeng, T., Ren, H.: Endovla: Dual-phase vision-language-action model for autonomous tracking in endoscopy. In: Conference on Robot Learning (CoRL) (2025)
14. Picozzi, P., et al.: Advances in robotic surgery: A review of new surgical robotics platforms. Electronics **13**(23), 4675 (2024)
15. Reiley, C.E., Lin, H.C., Yuh, D.D., Hager, G.D.: A review of methods for objective surgical skill evaluation. Surgical Endoscopy **25**(2), 356–366 (2011)
16. Richter, F., Orosco, R.K., Yip, M.C.: Surgical robot autonomy: Task-constrained motion planning and control. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 12345–12351. IEEE (2021)
17. Sadagopan, N.: Beyond ai and robotics: The dawn of surgical automation. AI Surgery **1**, 34 (2024)
18. Schmidgall, S., et al.: Will your next surgeon be a robot? autonomy and ai in robotic surgery. Science Robotics (2025)
19. Shao, Z., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
20. Thananjeyan, B., Balakrishna, A., Seita, D., Hsu, J., Ferroni, G., Okamura, A.M., Goldberg, K.: Multilateral surgical pattern cutting in 2d orthotropic gauze with deep reinforcement learning. In: Robotics: Science and Systems (RSS) (2020)
21. Xu, Z., Li, X., Lu, Z., Liu, D., Dou, Q., Heng, P.A.: Surrol: An open-source reinforcement learning centered and dvrk compatible platform for surgical robot learning. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2021)
22. Yang, S., Li, Q., Shen, D., Gong, B., Dou, Q., Jin, Y.: Deform3DGS: Flexible Deformation for Fast Surgical Scene Reconstruction with Gaussian Splatting . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15006. Springer Nature Switzerland (October 2024)
23. Zhao, Z., Fang, F., Yang, X., Xu, Q., Guan, C., Zhou, S.K.: See, predict, plan: Diffusion for procedure planning in robotic surgical videos. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer (2024)
24. Zitkovich, B., Yu, T., Xu, S., Xu, P., Xiao, T., Xia, F., Wu, J., Wohlhart, P., Welker, S., Wahid, A., Vuong, Q., Vanhoucke, V., Tran, H., Soricut, R., Singh, A., Singh, J., Sermanet, P., Sanketi, P.R., Salazar, G., Ryoo, M.S., Reymann, K.,

Rao, K., Pertsch, K., Mordatch, I., Michalewski, H., Lu, Y., Levine, S., Lee, L., Lee, T.W.E., Leal, I., Kuang, Y., Kalashnikov, D., Julian, R., Joshi, N.J., Irpan, A., Ichter, B., Hsu, J., Herzog, A., Hausman, K., Gopalakrishnan, K., Fu, C., Florence, P., Finn, C., Dubey, K.A., Driess, D., Ding, T., Choromanski, K.M., Chen, X., Chebotar, Y., Carbajal, J., Brown, N., Brohan, A., Arenas, M.G., Han, K.: Rt-2: Vision-language-action models transfer web knowledge to robotic control. In: Proceedings of The 7th Conference on Robot Learning. Proceedings of Machine Learning Research, vol. 229, pp. 2165–2183. PMLR (2023)