applied sciences

MDPI

Article

Smoothness-Aware Hierarchical Planning and RCM-Constrained Execution for Robot-Assisted Neuroendoscopic Access: A Patient-Specific Computational Study †

Ya Li, Baoqing Pei * and Xueqing Wu

School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China
* Correspondence: pbq@buaa.edu.cn; Tel.: +86-136-9919-0667

† This article is an extended version of our paper entitled "RCM-Constrained Hierarchical Planning for Robot-Assisted Neuroendoscopic Access in Patient-Specific Multimodal Anatomical Scenes", accepted for publication in the proceedings of the 8th International Conference on Robotics and Computer Vision (ICRCV 2026), Jiangyin, China, 6–8 November 2026. The journal version adds smoothness-aware fine-tuning, held-out-patient validation, expanded methodological reporting, distributional and effect-size analyses, and additional RCM analyses.

Abstract

Robot-assisted neuroendoscopic access requires obstacle-responsive planning while maintaining a fixed cranial entry-point remote center of motion (RCM). We computationally evaluated a Hierarchical Planning-and-Execution Framework combining a proximal policy optimization (PPO) local planner generating Cartesian velocity commands with analytical RCM-constrained execution. Patient-specific scenes were reconstructed from multimodal imaging; seven patient anatomies supported development and within-cohort validation, and three were held out from training. Smoothness-Aware PPO+RCM was obtained by fine-tuning the original policy with an action-transition penalty. Under navigation-map perturbations, Smoothness-Aware PPO+RCM achieved a target error of 0.616 ± 0.236 mm, comparable to PPO+RCM ($p = 0.723$), while mean angular change decreased from $21.66 \pm 3.87°$ to $1.07 \pm 1.56°$ ($p < 0.001$). Full-trajectory time-averaged RCM error was 4.68 ± 4.70 mm; post-alignment RCM performance remained comparable to PPO+RCM ($p = 0.901$). Composite Constrained Task Success was 33/50 (66%). Across three held-out patients, mean target error ranged from 0.58 to 0.78 mm, post-alignment RCM error from 1.88 to 2.34 mm, and 17/30 scenes met the composite criterion, with no direct geometric intersections with simplified risk models. These findings support improved computational trajectory regularity with preserved targeting, geometric clearance, and RCM performance. Physical robot validation remains necessary.

Keywords: robot-assisted neuroendoscopy; remote center of motion; reinforcement learning; hierarchical motion planning; constrained execution; surgical robotics

Check for updates

Academic Editor: Alexander Zhbanov

Received: 10 September 2026
Revised: 20 October 2026
Accepted: 21 October 2026
Published: 29 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Robot-assisted neuroendoscopic access is a constrained motion-planning and control problem in which a rigid endoscope must be advanced from a cranial entry point to a deep intracranial target through a narrow patient-specific corridor. The motion must avoid vessels, deep nuclei, and other functionally important structures while maintaining the endoscope shaft about a fixed entry-point remote center of motion (RCM) at the skull opening [1–6]. Unlike unconstrained Cartesian point-to-point motion, tip translation, shaft orientation, and entry-point consistency are kinematically coupled; therefore, the planning

and execution modules must jointly support target convergence, anatomical clearance, and RCM-compatible robot motion [7–10].

Patient-specific image models provide the geometric and semantic information required for safe surgical approach planning [11,12]. However, these representations depend on segmentation and registration accuracy and may become spatially outdated because of intraoperative brain shift [13,14]. Consequently, a precomputed geometric path may be insufficient when the represented environment changes. Graph- and sampling-based planners remain effective for geometric path generation, but adaptation to updated local anatomy generally requires explicit replanning. In parallel, analytical RCM controllers can enforce the entry-point geometry but do not by themselves determine how the distal tip should move around patient-specific obstacles. This motivates an architecture in which adaptive local planning and deterministic RCM execution are treated as coupled but functionally distinct modules.

Beyond classical sampling-based planning, several complementary paradigms have been developed for constrained robotic motion. Within the rapidly exploring random tree (RRT) family, RRT* improves sampling-based solution quality through asymptotic optimality [15], whereas CHOMP, STOMP, and TrajOpt refine continuous trajectories through optimization of obstacle and smoothness costs [16–18]. Kinodynamic planning incorporates differential constraints and system dynamics [19], while model predictive control (MPC) can impose state/input constraints and has been applied to RCM-constrained surgical manipulation [20]. Safety-filter frameworks provide another modular strategy by separating nominal or learned decision making from explicit constraint-enforcing intervention [21]. Recent RCM surgical-robot studies have also emphasized prototype and experimental validation [22].

Deep reinforcement learning (DRL) provides a state-feedback alternative in which current observations are mapped directly to local actions, and has been investigated for robotic trajectory planning in unstructured environments and for surgical path-planning tasks involving steerable instruments [23,24]. Safety-aware DRL has also been explored for intraoperative robotic planning using updated anatomical observations [25]. Among policy-gradient methods, proximal policy optimization (PPO) offers a practical formulation for continuous control through a clipped surrogate objective [26]. PPO was selected here because it supports continuous actions and allows the local planner to respond to updated observations without solving a new optimization problem at every control step. The policy output is not responsible for enforcing the RCM constraint; that requirement is handled analytically by the lower-level execution layer.

A preliminary conference version of this work introduced the Hierarchical Planning-and-Execution Framework, coupling a PPO-based local planner with an RCM-constrained execution layer [27]. Validation in seven patient-derived computational environments showed accurate target reaching and low risk-zone exposure, but the learned local planner produced substantial step-to-step directional variation: PPO+RCM under navigation-map perturbations yielded a mean angular-change metric of $21.66 \pm 3.87°$, whereas the shortcut-pruned and B-spline-smoothed RRT+RCM baseline achieved $2.95 \pm 6.88°$. In addition, training and validation drew from the same seven-patient pool, so patient-wise transfer to anatomies excluded from policy training remained untested. To address these limitations, the present study fine-tunes the original PPO policy with action-transition regularization to reduce directional variation and extends the evaluation to three patient anatomies excluded from policy training. The action-transition regularization evaluated here is treated as an incremental planner-level extension rather than as a fundamentally new reinforcement-learning objective.

The present study makes three main contributions. First, it formalizes a hierarchical surgical-robot motion architecture in which PPO-based local planning, command mapping, and geometric RCM-constrained execution are separated at the velocity level. Second, Smoothness-Aware PPO+RCM adds action-transition regularization and fine-tunes the original policy to reduce directional variation without modifying the RCM execution layer. Third, the framework is evaluated using module controls, a smoothed RRT+RCM baseline, dynamic navigation-map perturbations, and three unseen-patient anatomies. The study evaluates computational target error, geometric clearance, trajectory regularity, and RCM consistency under the stated modeling assumptions.

2. Materials and Methods

2.1. Patient-Specific Scene Reconstruction

For each clinical case, multimodal imaging data were co-registered and fused to model the intracranial anatomy, hemorrhage target, and critical obstacles involved in approach planning. The reconstructed models served as the geometric and semantic basis for training and validating the Hierarchical Planning-and-Execution Framework.

MRI, CT, and MRA data were used to reconstruct the surgical environment. MRI provided anatomical information and was segmented into 95 anatomical regions using FastSurfer (version 2.1, Deep-MI Lab, German Center for Neurodegenerative Diseases, Bonn, Germany) [28], with the basal ganglia and adjacent nuclei treated as high-risk structures (Figure 1). CT was used to delineate the hemorrhage region using nnU-Net (version v2.0; German Cancer Research Center, Heidelberg, Germany), and the segmented lesion center was defined as the target. MRA was used to extract vascular structures by adaptive thresholding and morphological operations.

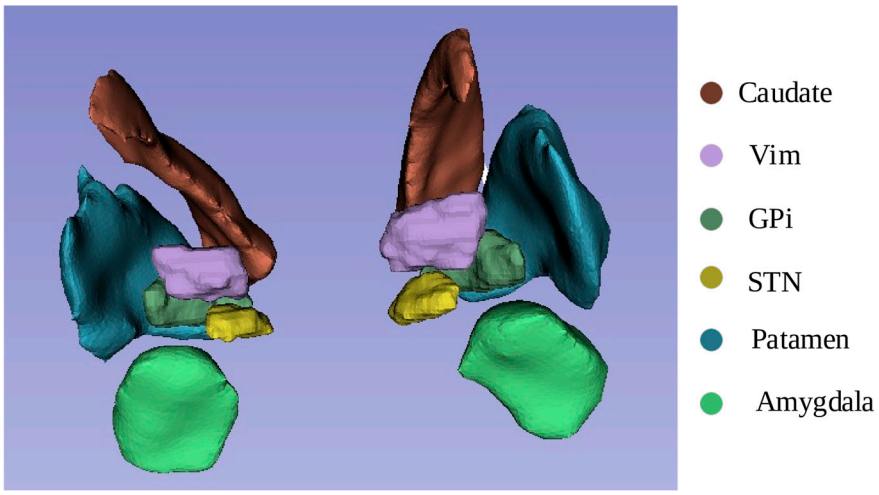


Figure 1. Critical nuclei surrounding the basal ganglia.

For efficient simulation and collision checking, vascular structures were reduced to centerline- and Voronoi-based geometric representations [29] (Figure 2), whereas the critical nuclei were approximated using oriented bounding boxes (OBBs) (Figure 3). This simplification reduced computational cost while retaining the spatial relationships needed for approach-planning validation. Accordingly, the collision and clearance results refer to these simplified geometric models rather than to exact tissue surfaces.

The reconstructed scene provides the geometric and semantic environment for learning-based local motion planning, collision-risk evaluation, and RCM-constrained robot execution.

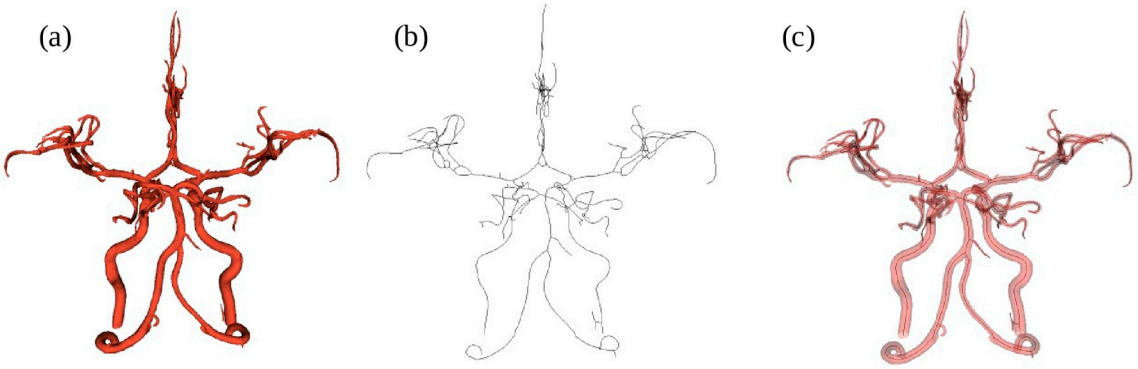


Figure 2. Vessel centerline extraction: (**a**) 3D vessel model; (**b**) extracted vessel centerline; (**c**) overlay of the centerline on the 3D vessel model.

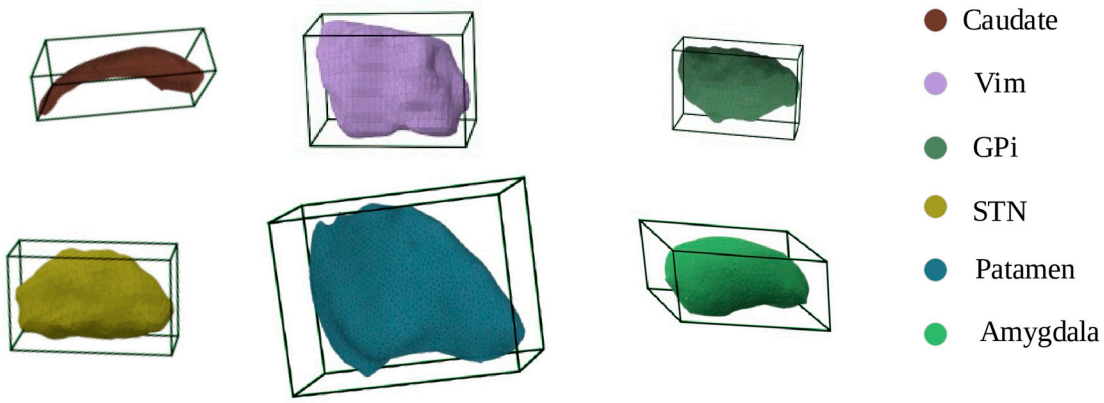


Figure 3. Schematic representation of oriented bounding boxes for the critical nuclei.

2.2. Hierarchical Planning-And-Execution Framework

The Hierarchical Planning-and-Execution Framework consists of three layers: a decision layer, a command-mapping layer, and a constraint-execution layer (Figure 4). The decision layer uses a PPO policy to generate a three-dimensional normalized action in the tool coordinate frame from the current task state. The command-mapping layer scales the normalized action into a desired Cartesian velocity and transforms it from the tool coordinate frame to the robot base frame. At the velocity level, the constraint-execution layer combines the end-effector motion task with RCM constraint feedback to generate the joint command for the next control cycle. This decomposition separates adaptive local decision-making from deterministic kinematic constraint handling, allowing the planning and execution modules to be evaluated and modified independently.

The computational environment used a fixed-base six-degree-of-freedom (6-DoF) UR5e manipulator(Universal Robots A/S, Odense, Denmark) equipped with a rigid neuroendoscope model. The PPO policy generated three-dimensional translational tool center point (TCP) commands at 30 Hz, with a maximum Cartesian command of 0.1 m/s. The endoscope control axis was defined by the tool0 and TCP reference points separated by 186 mm. The Unified Robot Description Format (URDF) joint-position limits were $\pm 2\pi$ rad for joints 1, 2, 4, 5, and 6 and $\pm\pi$ rad for joint 3, with nominal joint-velocity limits of π rad/s.

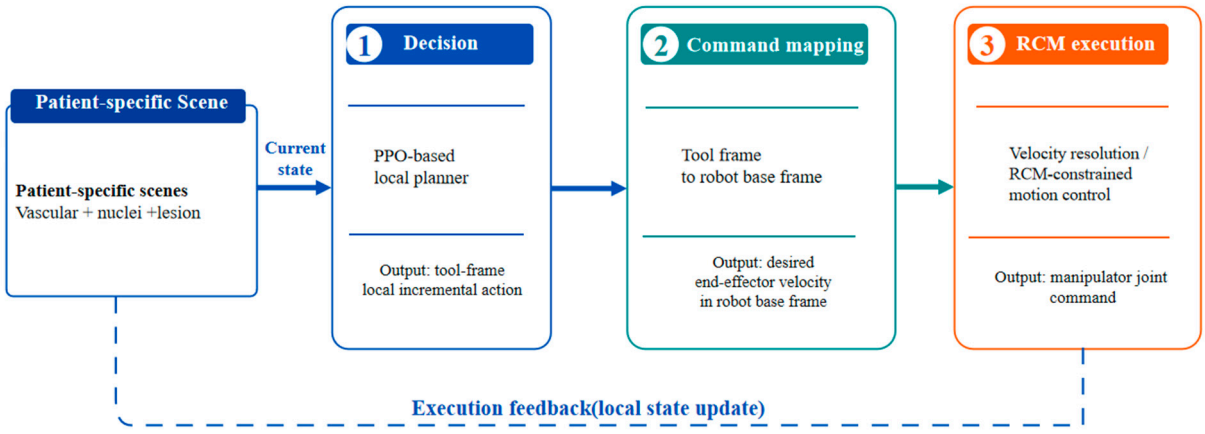


Figure 4. Three-layer structure of the Hierarchical Planning-and-Execution Framework.

2.2.1. State Representation

A 30-dimensional state vector characterizes the approach task by combining endoscope pose and motion, target-relative information, anatomical clearance, and RCM deviation (Table 1).

Table 1. State-space vector.

State Component	Dimension	Description
Tool-tip position	3	TCP position in the robot base frame
Tool orientation quaternion	4	Unit-quaternion representation of the TCP orientation
Tool linear and angular velocities	6	Current translational and rotational motion state of the TCP
Target position	3	Lesion target position in the robot base frame
Unit direction vector to target	3	Direction from the TCP to the target
Distance to target	1	Euclidean TCP-to-target distance
Distances to anatomical structures	7	Minimum distances from the TCP to the vascular structures and the surfaces of six critical nuclei
RCM error vector	3	Perpendicular deviation vector from the cranial entry point to the tool axis
Total	30	

Based on the current state, the PPO policy generates a local Cartesian action, which is mapped into the robot base frame and subsequently processed by the RCM-constrained execution module. The formulation of the constrained execution layer is described below.

2.2.2. RCM-Constrained Execution

The lower-level execution module converts the Cartesian velocity command generated by the upper-level planning policy into kinematically feasible robot motion while regulating deviation of the neuroendoscope shaft from the fixed cranial entry point. In the Hierarchical Planning-and-Execution Framework, the PPO-based decision layer determines the local target-directed motion, whereas the RCM-constrained execution layer explicitly enforces the entry-point constraint at the velocity level. The geometric formulation follows the RCM execution strategy used in the preliminary framework.

Let $q = [q_1, \ldots, q_n]^T$ denote the robot joint vector, p_e the fixed cranial entry point, and $p_i(q)$ and $p_{i+1}(q)$ two points defining the endoscope shaft segment. A virtual RCM point is parameterized along this segment by the scalar λ:

$$p_{RCM}(q, \lambda) = (1 - \lambda)p_i(q) + \lambda p_{i+1}(q), \qquad 0 \le \lambda \le 1 \tag{1}$$

Differentiating the RCM-point position gives

$$\dot{p}_{RCM} = \begin{bmatrix} (1 - \lambda)J_i(q) + \lambda J_{i+1}(q) & p_{i+1}(q) - p_i(q) \end{bmatrix} \begin{bmatrix} \dot{q} \\ \dot{\lambda} \end{bmatrix} = J_{RCM}(q, \lambda) \begin{bmatrix} \dot{q} \\ \dot{\lambda} \end{bmatrix} \tag{2}$$

Here, J_i and J_{i+1} are the translational Jacobians of the two shaft points, and $J_{RCM} \in \mathbb{R}^{3\times(n+1)}$ is the augmented Jacobian associated with the virtual RCM point. For an ideal fixed RCM, the virtual point coincides with the cranial entry point. The residual RCM position error is defined as

$$e_{RCM} = p_e - p_{RCM} \tag{3}$$

The feedback error e_{RCM} in Equation (3) is the vector from the virtual RCM point to the fixed cranial entry point. The reported RCM deviation is the shortest perpendicular distance from the cranial entry point to the current tool axis. The neuroendoscope tip position p_{tip} is determined by the robot configuration, and its translational velocity is expressed as

$$p_{tip} = f(q), \qquad \dot{p}_{tip} = J_{tip}(q)\dot{q}, \tag{4}$$

where J_{tip} is the translational Jacobian of the endoscope tip. The tip-motion task and the RCM task are combined into an augmented kinematic system:

$$\begin{bmatrix} \dot{p}_{tip} \\ \dot{p}_{RCM} \end{bmatrix} = J_{aug} \begin{bmatrix} \dot{q} \\ \dot{\lambda} \end{bmatrix} \tag{5}$$

The augmented Jacobian is

$$J_{aug} = \begin{bmatrix} J_{tip}(q) & 0_{3\times1} \\ (1 - \lambda)J_i(q) + \lambda J_{i+1}(q) & p_{i+1}(q) - p_i(q) \end{bmatrix} \tag{6}$$

The command-mapping layer provides the desired Cartesian tip velocity v^d_{tip} in the robot base frame. The planner-generated tip-motion command and the RCM-error correction are combined as

$$v^d_{aug} = K \begin{bmatrix} v^d_{tip} \\ e_{RCM} \end{bmatrix}, \qquad K = 20\, diag(0.01, 0.01, 0.01, 0.05, 0.05, 0.05) \tag{7}$$

The first three diagonal terms scale the planner-generated tip-motion command, whereas the last three terms weight the RCM-error correction. The augmented joint and interpolation velocities are obtained using the ordinary pseudoinverse (Moore–Penrose generalized inverse) of the augmented Jacobian:

$$\begin{bmatrix} \dot{q} \\ \dot{\lambda} \end{bmatrix} = J^{\dagger}_{aug} v^d_{aug}. \tag{8}$$

The augmented Jacobian was solved using the singular value decomposition (SVD)-based Moore–Penrose pseudoinverse; no explicit damped least-squares regularization or null-space joint-limit avoidance was implemented. The interpolation parameter was initialized at $\lambda_0 = 0.5$ and updated together with the joint motion; each update was accepted only when $0 < \lambda_{t+1} < 1$, otherwise the previous value was retained, thereby keeping the virtual RCM point within the 186-mm tool-axis segment.

2.2.3. Reward Design and Smoothness-Aware Extension

Let $d_{tar}(t)$ denote the current tip-to-target distance, $d_{tar}(0)$ the initial distance, and $d_{min}(t)$ the minimum tip-to-obstacle distance over the high-risk anatomical set. The original PPO reward combines target progress, anatomical-risk avoidance, action regularization, and terminal outcomes. In Equation (9), w_g, w_s, and w_a weight the goal-progress, safety-buffer, and action-magnitude terms, respectively, and R_{term} denotes the terminal reward assigned at target-reaching or collision termination.

$$R_t = w_g R_{goal}(t) + w_s R_{safe}(t) - w_a \parallel a_t \parallel_2^2 + R_{term}(t) \tag{9}$$

The normalized target-progress term is

$$R_{goal}(t) = \frac{d_{tar}(t-1) - d_{tar}(t)}{d_{tar}(0)} \tag{10}$$

which is positive when the tip moves toward the lesion center and negative when it moves away. Let $L_1 = 5$ mm be the critical-risk distance and $L_2 = 10$ mm the outer safety distance. The continuous buffer-zone penalty is

$$R_{safe}(t) = \begin{cases} -R_{col}, & d_{min}(t) \le L_1 \\ -R_{col}\frac{L_2 - d_{min}(t)}{L_2 - L_1 + \delta}, & L_1 < d_{min}(t) < L_2 \\ 0, & d_{min}(t) \ge L_2 \end{cases} \tag{11}$$

where R_{col} is the maximum proximity penalty, and δ prevents numerical division by zero. $L_1 = 5$ mm defines the inner modeled risk-zone distance and $L_2 = 10$ mm the outer buffer distance. Direct geometric collision terminates the episode during training. Reaching the training target threshold produces a positive terminal reward, whereas collision produces a negative terminal reward.

Smoothness-Aware PPO+RCM adds an action-transition regularization term:

$$R_{smooth}(t) = -w_{\Delta a} \parallel a_t - a_{t-1} \parallel_2^2 \tag{12}$$

The modified reward is therefore

$$R_t^{SA} = R_t - w_{\Delta a} \parallel a_t - a_{t-1} \parallel_2^2 \tag{13}$$

The additional term penalizes changes between successive normalized local commands rather than path length directly. In Equations (9), (12) and (13), $a_t \in [-1,1]^3$ denotes the normalized three-dimensional PPO action in the tool coordinate frame; the command-mapping layer subsequently scales and transforms this action into the desired Cartesian TCP velocity. Thus, the action-transition term penalizes $\parallel a_t - a_{t-1} \parallel_2^2$ in policy-action space rather than directly penalizing TCP velocity, acceleration, jerk, actuator effort, or tissue loading. Smoothness-Aware PPO+RCM was initialized from the trained PPO+RCM policy and fine-tuned while retaining the network architecture, scene randomization, action representation, and RCM execution layer.

2.3. Training Configuration and Validation Settings

2.3.1. Training Configuration

(1) Entry-point design. To construct representative neuroendoscopic approach scenarios, surgical approaches and cranial entry points were defined according to the 2020 Chinese expert consensus [30]. These configurations were used to support scene diversity during training and evaluation. The burr-hole entry positions are listed in Table 2.

Table 2. Burr-hole entry positions.

Region	No.	Burr-Hole Position
Left frontal-lobe approach	P01	Left superior frontal gyrus (SFG), anterior
Left frontal-lobe approach	P02	Left superior frontal gyrus (SFG), posterior
Left frontal-lobe approach	P03	Left middle frontal gyrus (MFG), anterior
Left frontal-lobe approach	P04	Left middle frontal gyrus (MFG), posterior
Left temporal-lobe approach	P05	Left superior temporal gyrus (STG), posterior
Left temporal-lobe approach	P06	Left middle temporal gyrus (MTG), posterior
Left insular-lobe approach	P07	Left insular lobe (IL)
Right frontal-lobe approach	P08	Right superior frontal gyrus (SFG), anterior
Right frontal-lobe approach	P09	Right superior frontal gyrus (SFG), posterior
Right frontal-lobe approach	P10	Right middle frontal gyrus (MFG), anterior
Right frontal-lobe approach	P11	Right middle frontal gyrus (MFG), posterior
Right temporal-lobe approach	P12	Right superior temporal gyrus (STG), posterior
Right temporal-lobe approach	P13	Right middle temporal gyrus (MTG), posterior
Right insular-lobe approach	P14	Right insular lobe (IL)

(2) Cranial scene initialization. For each patient dataset, the scene contained the lesion target, high-risk anatomical structures, cranial entry point, and robotic-arm model carrying the neuroendoscope. The robot base was fixed. To improve adaptability across patients and approaches, the cranial scene was randomized during training with respect to patient identity, global translation, global rotation, and cranial entry-point selection (Table 3). The translation and rotation ranges were used as engineering domain-randomization bounds rather than as clinically measured anatomical displacement ranges.

Table 3. Randomization conditions for cranial scene initialization.

Condition	Selection	Range
Patient environment	Random	ID 0–6
Translation	Random	≤100 mm along each of the X, Y, and Z directions
Rotation	Random	≤15° about each direction
Cranial entry point	Random	ID 0–13

(3) Reward-function configuration. The critical-risk distance L_1 was 5 mm, and the outer safety distance L_2 was 10 mm. Training used the reward formulation in Equations (9)–(13): normalized target progress, continuous anatomical-proximity penalty, action-magnitude regularization, terminal outcomes, and—only for Smoothness-Aware PPO+RCM—the action-transition penalty. Direct collision and collision-risk proportion were treated separately.

(4) PPO training parameters. The policy was implemented using Stable-Baselines3 (version 2.5.0; open-source software) with an MlpPolicy actor–critic architecture and a three-dimensional continuous action. The actor and critic used separate two-layer multilayer perceptrons with 64 hidden units per layer and Tanh activations. The actor parameterized a diagonal Gaussian action distribution, and the critic produced

a scalar state-value estimate. PPO used the Adam optimizer (learning rate 3×10^{-4}, $\beta_1/\beta_2 = 0.9/0.999$, $\varepsilon = 10^{-5}$), rollout length 1024, batch size 64, 10 epochs per update, discount factor 0.999, generalized advantage estimation (GAE) $\lambda = 0.95$, clipping range 0.05, entropy coefficient 0, value-loss coefficient 0.5, and gradient clipping at 0.5.

(5) Smoothness-Aware fine-tuning. Rather than training a new policy from random initialization, Smoothness-Aware PPO+RCM was initialized from the trained PPO+RCM model. The original reward was augmented with the action-transition term in Equation (12), using $w_{\Delta a} = 0.05$, $w_g = 2.5$, $w_s = 0.5$, $w_a = 0.05$. The policy was further optimized for 100,000 training steps under the same patient-scene randomization and RCM-constrained execution settings. Unless otherwise specified, all PPO parameters were identical to those of the original PPO+RCM model.

The original PPO policy was trained for 1,700,000 environment steps, completing 11,077 episodes. The training reward increased during the early and intermediate stages and reached a relatively stable plateau during the later stage (Figure S1). No fixed random seed was recorded, and independent repeated training runs were not performed.

(6) RCM-constrained execution-layer parameters. The constrained execution layer used the geometric-Jacobian formulation described in Section 2.2.2. The gain matrix was $K = 20 \cdot diag([0.01, 0.01, 0.01, 0.05, 0.05, 0.05])$, and the shaft interpolation parameter was initialized at $\lambda_0 = 0.5$. The ordinary Moore–Penrose pseudoinverse was used without an additional null-space objective.

(7) Training termination conditions. To improve training efficiency, three termination conditions were defined for each episode with a maximum length of 1024 steps: the robotic-arm control steps reached 1024; a direct collision with high-risk anatomical structures occurred during training; or the distance between the TCP and the target point was less than 4 mm and the approach was judged successful.

2.3.2. Validation Settings and Baselines

Five validation settings were used in the preliminary study. PPO+RCM in static scenes tested nominal feasibility. PPO+RCM in dynamically perturbed scenes tested response to navigation-map changes. RRT+RCM in dynamic scenes served as a sampling-based comparison using the same lower-level RCM controller. Straight-line+RCM retained constrained execution but removed adaptive local planning. PPO without RCM retained the learned planner but removed explicit entry-point-constrained execution. Each setting contained 50 independently initialized trials generated from the same seven patient-derived environments.

The dynamic perturbation model applied a random three-dimensional displacement with a maximum magnitude of 2 mm every 50 control steps at 30 Hz (approximately every 1.67 s) to the target and anatomical obstacle models. Cumulative displacement was bounded by 20 mm; when a proposed displacement would exceed this bound, a displacement in the opposite direction was applied. This model is interpreted as a navigation-map perturbation stress test rather than a biomechanical brain-shift model.

The Cartesian RRT baseline used a 3-mm step size, 25% goal-sampling probability, a 1000-iteration limit, a 4-mm goal threshold, and 1-mm edge-checking resolution. Sampling was restricted to an entry-to-target corridor with a 30-mm radius and 20-mm depth padding. Candidate paths were shortcut-pruned and smoothed using a clamped cubic B-spline; the smoothed path was retained only if it passed the same safety checks. In dynamic scenes, replanning was event-triggered rather than periodic. This baseline is used as a classical sampling-based reference rather than as the strongest available online planner.

2.3.3. Extended Validation Protocol

Two additional validation experiments were used to evaluate the extended framework. First, Smoothness-Aware PPO+RCM was tested under the same dynamic perturbation protocol used for the original PPO+RCM and RRT+RCM comparisons. The primary endpoint was trajectory smoothness, quantified by the mean angular change between consecutive trajectory segments. Secondary endpoints were target error, collision-risk proportion, minimum safety distance, path length, RCM consistency, and Composite Constrained Task Success. The lower-level RCM controller, environment configuration, and validation protocol were unchanged across the comparison. Because Smoothness-Aware PPO+RCM was obtained by fine-tuning the original PPO+RCM policy with only the additional action-transition penalty, the comparison isolates the effect of planner-level smoothness regularization.

Second, an unseen-patient validation was performed to assess patient-wise generalization. Three additional patient-specific anatomical datasets were completely excluded from policy training. Ten independently initialized test scenes were generated for each held-out patient, yielding 30 additional scenes. The trained Smoothness-Aware PPO+RCM policy was evaluated without patient-specific fine-tuning. Results were reported separately for each patient to avoid masking inter-patient anatomical variability through pooled scene-level averaging.

2.4. Evaluation Metrics and Statistical Analysis

The evaluation metrics were selected to characterize target error, collision-risk proportion, minimum clearance, trajectory length, mean angular change, RCM consistency, and simultaneous satisfaction of study-specific engineering criteria (Table 4). Direct collision denotes geometric intersection with the simplified risk models; collision-risk proportion denotes the fraction of trajectory samples within the 5-mm modeled risk zone. These measures quantify computational geometric behavior and should not be interpreted as clinical safety outcomes.

Because the static, dynamic, and method-comparison validation sets were independently initialized rather than scene-wise paired, between-group comparisons used two-sided Mann–Whitney U tests. Continuous outcomes are reported as mean ± standard deviation (SD) and, where distributional asymmetry is relevant, median [interquartile range, IQR]. Rank-biserial correlation (r_{rb}) is reported as a nonparametric effect size, with 95% confidence intervals estimated by bootstrap resampling. For each predefined comparison across the five primary geometric/trajectory metrics (target error, collision-risk proportion, minimum clearance, trajectory length, and mean angular change), Holm correction was applied to control the family-wise error rate. Scene-level statistical tests remain exploratory because multiple scenes were generated from a limited number of independent patient anatomies.

The Composite Constrained Task Success criterion requires final target error ≤ 0.75 mm, no direct geometric collision, minimum modeled clearance > 1 mm, and post-alignment mean RCM error ≤ 5 mm. These thresholds were defined as comparative engineering criteria rather than validated clinical tolerances. In particular, the 5-mm RCM threshold should not be interpreted as an acceptable cranial entry-point deviation for surgery.

To examine whether Composite Constrained Task Success depended strongly on the predefined engineering thresholds, a threshold-sensitivity analysis was performed for target-error thresholds of 0.50, 0.75, and 1.00 mm and post-alignment mean RCM-error thresholds of 2, 3, and 5 mm. The direct-collision criterion (no direct geometric collision) and minimum modeled clearance criterion (>1 mm) were kept unchanged. Endpoint-error classification

used 0.01-mm reporting precision throughout, and the 0.75-mm target-error/5-mm RCM combination was treated as the nominal criterion for the sensitivity analysis.

To complement the geometric mean-angular-change metric with time-dependent kinematic measures, TCP-level temporal motion metrics were calculated from the simulator-recorded translational TCP velocity at 30 Hz ($\Delta t = \frac{1}{30} s$) for PPO+RCM, RRT+RCM, and Smoothness-Aware PPO+RCM under the navigation-map perturbation protocol. For each trajectory, mean TCP speed was computed from the Euclidean norm of the three translational velocity components. TCP acceleration was estimated using the central finite difference $a_k = {}^{v_{k+1} - v_{k-1}}/_{2\Delta t}$ and jerk was obtained by applying the same central-difference operator to acceleration; derivative boundary samples were omitted. The root-mean-square (RMS) magnitudes of acceleration and jerk were used as trajectory-level temporal-variation measures. These metrics characterize simulated TCP kinematic regularity and do not represent physical actuator effort, hardware tracking error, or tissue interaction. The three temporal metrics were summarized as median [IQR] and compared using two-sided Mann–Whitney U tests, rank-biserial effect sizes with bootstrap 95% confidence intervals, and Holm correction across the three temporal metrics within each predefined method comparison.

Table 4. Evaluation metrics.

Category	Evaluation Metric	Detailed Description
Accuracy	Target error (mm)	Distance between the endoscope tip and the target center at termination.
Geometric safety	Direct collision	Geometric intersection between the endoscope model and a simplified high-risk anatomical structure.
	Collision-risk proportion (%)	Percentage of trajectory samples entering the 5-mm modeled risk zone around simplified critical structures.
	Minimum clearance (mm)	Minimum tip-to-obstacle clearance along the trajectory.
Efficiency	Trajectory length (mm)	Total length of the executed endoscope-tip trajectory.
Motion quality	Mean angular change (°)	Mean angular change between consecutive trajectory segments.
RCM consistency	Full-trajectory time-averaged RCM error (mm)	Mean perpendicular distance from the tool axis to the cranial entry point over the full trajectory.
	Post-alignment mean RCM error (mm)	Mean RCM error during the post-alignment phase, which begins after 10 consecutive samples with RCM error below 5 mm.
	Post-alignment P95 RCM error (mm)	The 95th percentile of RCM error after alignment.
	Post-alignment maximum RCM error (mm)	Maximum RCM error after alignment.
Composite	Composite Constrained Task Success	Final target error $\leq$ 0.75 mm, no direct collision, minimum clearance $>$ 1 mm, and post-alignment mean RCM error $\leq$ 5 mm. (These thresholds were defined as study-specific engineering criteria for comparative evaluation and should not be interpreted as validated clinical safety thresholds.)

3. Results

3.1. Five-Setting Validation

The five-setting validation results are summarized in Table 5 and Figure 5. Because several outcomes showed asymmetric distributions, both mean ± SD and median [IQR] are reported. PPO+RCM showed similar target error and geometric clearance in static and navigation-map-perturbation scenes. Relative to RRT+RCM under the perturbation protocol, PPO+RCM had lower target error but substantially larger mean angular change. The module-control experiments indicate complementary roles for local planning and RCM execution.

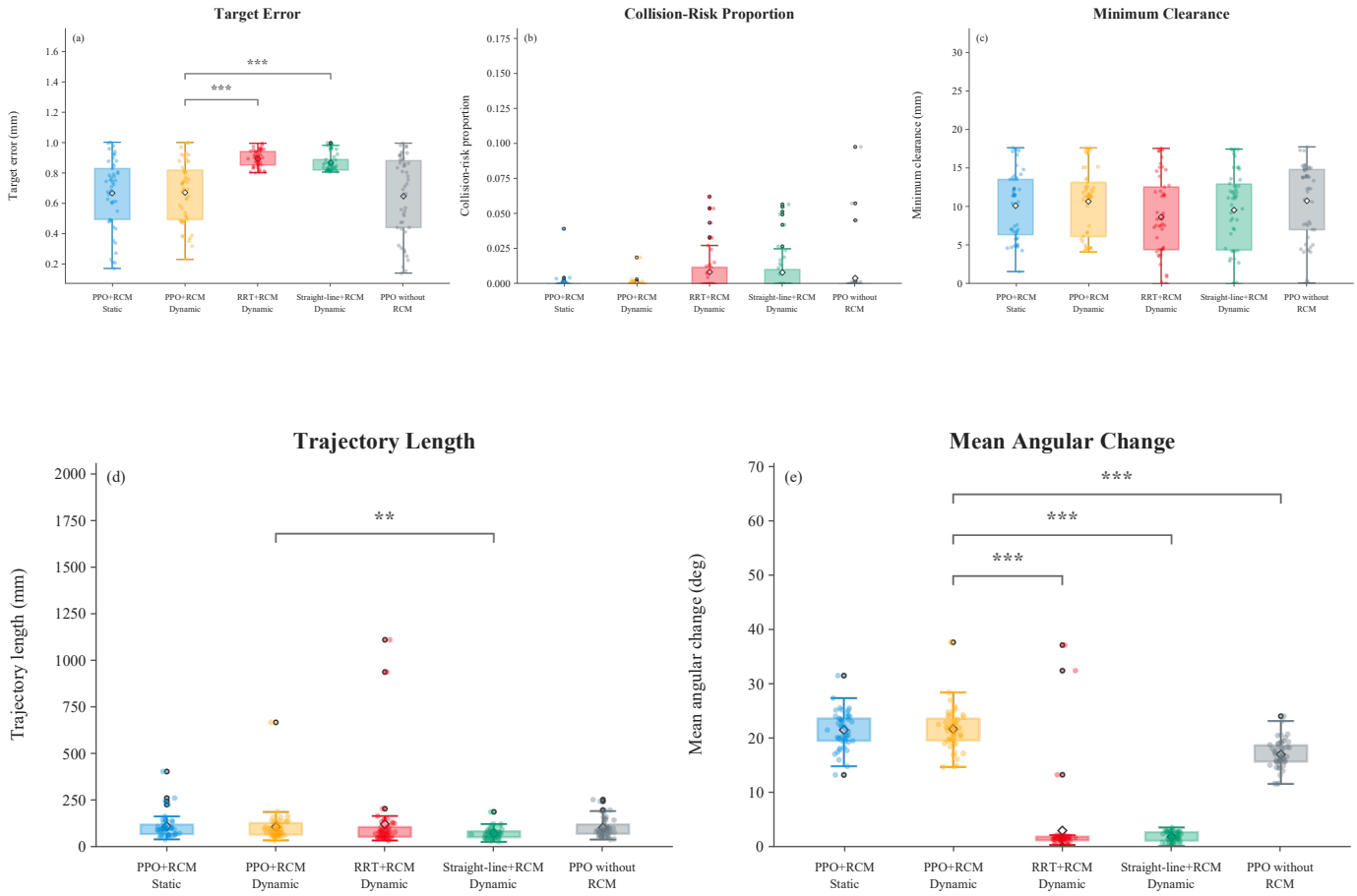


Figure 5. Five-setting comparison of (**a**) target error, (**b**) collision-risk proportion, (**c**) minimum clearance, (**d**) trajectory length, and (**e**) mean angular change. Colors distinguish the five validation settings labeled on the horizontal axes. Boxes span the interquartile range (25th–75th percentiles), whiskers indicate the range of non-outlying values, and diamonds indicate means. Colored dots represent individual trial results; black-outlined circles mark observations beyond the whiskers. Median markers are not shown. Solid brackets indicate pairwise comparisons assessed using two-sided Mann–Whitney U tests with Holm correction (** $p < 0.01$; *** $p < 0.001$). Non-significant comparisons are omitted.

Table 5. Five-setting validation results. For the five geometric/trajectory metrics, the first line reports mean ± SD and the second line median [IQR].

Validation Setting	Target Error (mm)	Collision-Risk Proportion (%)	Minimum Clearance (mm)	Trajectory Length (mm)	Mean Angular Change (°)	Post-Align. RCM (mm)	Constrained Success
PPO+RCM static	0.67 ± 0.23 0.72 [0.49, 0.83]	0.11 ± 0.56 0.00 [0.00, 0.00]	10.08 ± 4.24 11.34 [6.35, 13.45]	109.63 ± 62.98 97.69 [68.13, 119.09]	21.49 ± 3.33 21.48 [19.49, 23.58]	1.96 ± 1.82 1.96 [1.45, 2.42]	26/50
PPO+RCM dynamic	0.67 ± 0.20 0.70 [0.49, 0.82]	0.07 ± 0.27 0.00 [0.00, 0.00]	10.63 ± 4.31 11.29 [6.13, 13.06]	106.64 ± 88.92 94.39 [64.99, 126.81]	21.66 ± 3.87 21.83 [19.53, 23.57]	1.94 ± 1.66 1.93 [1.43, 2.72]	30/50
RRT+RCM dynamic	0.89 ± 0.05 0.89 [0.85, 0.94]	0.85 ± 1.61 0.00 [0.00, 1.19]	8.61 ± 5.22 8.35 [4.42, 12.48]	121.53 ± 192.10 78.49 [54.22, 105.37]	2.95 ± 6.88 1.45 [1.18, 1.83]	2.83 ± 3.29 2.12 [1.43, 2.81]	0/50
Straight-line+RCM dynamic	0.86 ± 0.06 0.84 [0.82, 0.89]	0.82 ± 1.61 0.00 [0.00, 1.03]	9.51 ± 5.26 11.05 [4.37, 12.86]	72.14 ± 28.93 69.57 [51.45, 82.57]	1.80 ± 0.91 1.82 [1.11, 2.65]	1.21 ± 1.38 1.63 [1.14, 2.27]	0/50
PPO without RCM dynamic	0.65 ± 0.26 0.70 [0.44, 0.88]	0.42 ± 1.68 0.00 [0.00, 0.00]	10.70 ± 4.90 12.80 [7.00, 14.72]	102.81 ± 50.88 89.86 [69.60, 120.05]	16.98 ± 2.61 16.70 [15.64, 18.64]	10.01 ± 9.17 7.12 [4.18, 9.22]	2/50

3.2. Target Error and Geometric Clearance

PPO+RCM maintained comparable target error between static and dynamic scenes (0.67 ± 0.23 vs. 0.67 ± 0.20 mm; medians 0.72 [0.49, 0.83] vs. 0.70 [0.49, 0.82] mm; U = 1213.0; $p = 1.000$). Under dynamic scenes, PPO+RCM had lower target error than RRT+RCM (0.67 ± 0.20 vs. 0.89 ± 0.05 mm; medians 0.70 [0.49, 0.82] vs. 0.89 [0.85, 0.94] mm; U = 382.0; $p < 0.001$; $r_{rb} = -0.682$, bootstrap 95% CI [−0.848, −0.504]).

Figure 6 illustrates the cumulative distribution of target error across the five validation settings and contextualizes the stricter 0.75-mm composite target-error threshold.

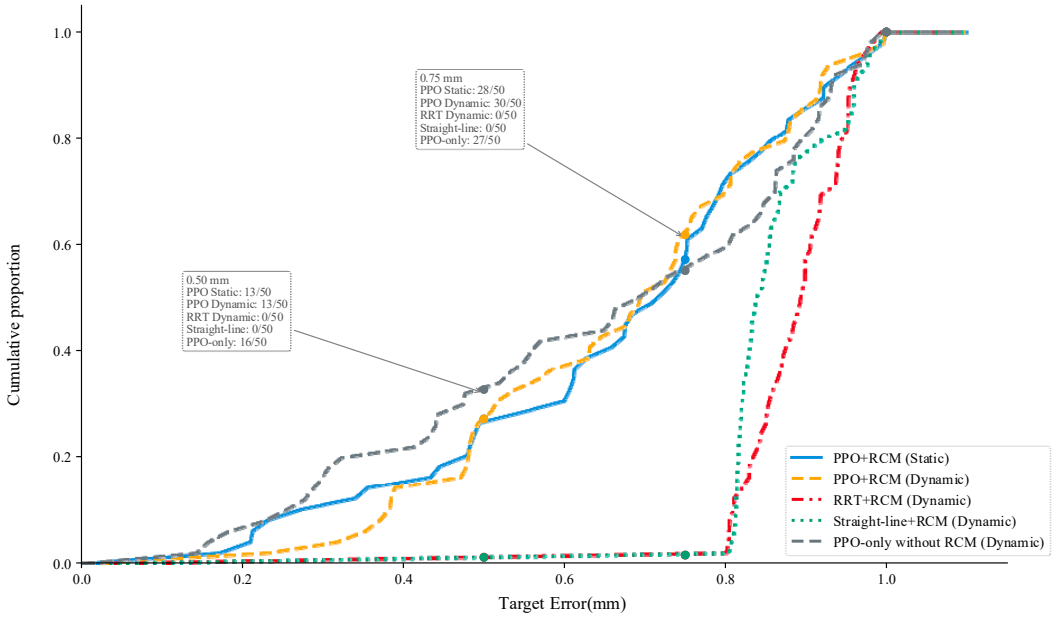


Figure 6. Target-error distributions for the five validation settings.

No direct geometric intersections with the simplified high-risk models were observed in the five validation settings. PPO+RCM had a low collision-risk proportion in both static and dynamic scenes (0.11 ± 0.56% vs. 0.07 ± 0.27%; both medians 0 [0, 0]%; $p = 1.000$), and minimum clearance was similar (10.08 ± 4.24 vs. 10.63 ± 4.31 mm; $p = 1.000$). Com-

pared with RRT+RCM, PPO+RCM showed a numerically lower collision-risk proportion ($0.07 \pm 0.27\%$ vs. $0.85 \pm 1.61\%$; $p = 0.137$), but the difference was not statistically significant. Minimum clearance was numerically larger with PPO+RCM (10.63 ± 4.31 vs. 8.61 ± 5.22 mm; $p = 0.137$).

3.3. RCM Consistency and Module Contribution

RCM deviation was evaluated as the perpendicular distance from the fixed cranial entry point to the current tool axis. Full-trajectory mean RCM errors were 5.73 ± 5.71 mm for PPO+RCM, 9.00 ± 7.43 mm for RRT+RCM, and 19.27 ± 16.25 mm for PPO without RCM. For all post-alignment samples, pooled mean ± SD values were 1.94 ± 1.66 mm for PPO+RCM and 2.83 ± 3.29 mm for RRT+RCM. At the trajectory level, the post-alignment mean RCM error was summarized as median [IQR]: 1.93 [1.43, 2.72] mm for PPO+RCM and 2.12 [1.43, 2.81] mm for RRT+RCM. PPO without RCM showed markedly larger RCM deviation (pooled post-alignment mean 10.01 ± 9.17 mm; trajectory-level median 7.12 [4.18, 9.22] mm), supporting the role of explicit RCM execution.

The explanatory controls clarified the role of each module. Straight-line+RCM preserved entry-point consistency and generated smooth trajectories, but it did not satisfy the strict constrained-success criterion because the target-error criterion was not satisfied. PPO without RCM preserved target convergence in many scenes, but the larger RCM deviation markedly reduced constrained task success. These findings support coupling learning-based local planning with explicit RCM-constrained execution rather than relying on either component alone.

Figure 7 shows the temporal convergence and stagewise distribution of RCM error, highlighting the difference between the initialization transient and the post-alignment constrained phase.

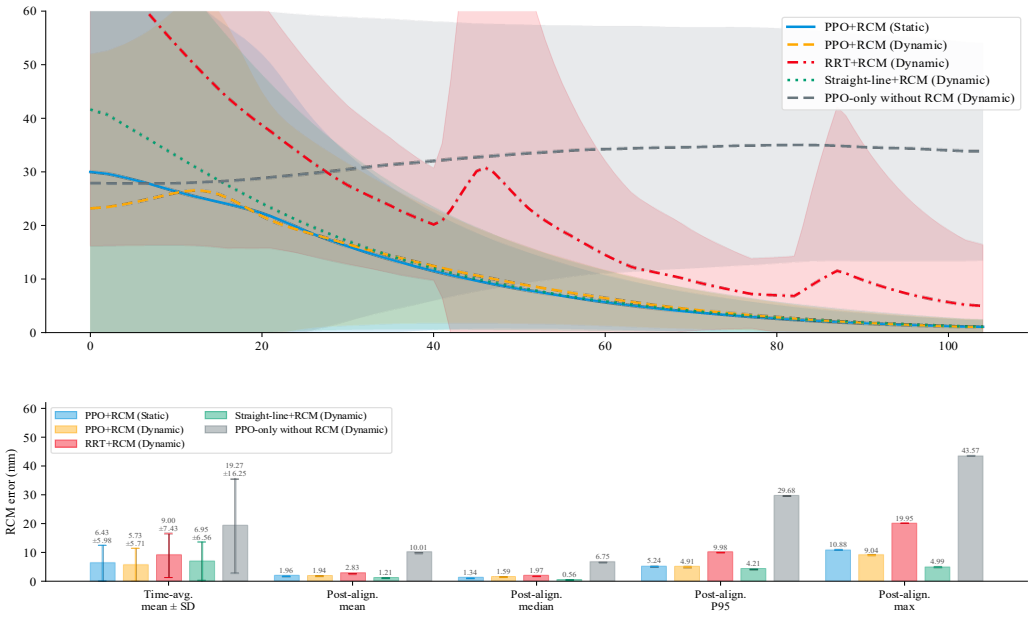


Figure 7. RCM-error convergence and stagewise statistics across the five validation settings. The upper panel shows the evolution of RCM error; colors and line styles identify the settings in the legend. Shaded bands indicate ±1 sample standard deviation around a centered five-step moving average. The lower panel presents the full-trajectory time-averaged mean ± SD and the post-alignment mean, median, 95th percentile (P95), and maximum RCM error. Colors correspond to the same validation settings.

3.4. Trajectory Length and Mean Angular Change

Trajectory-length distributions were right-skewed, particularly for RRT+RCM, so medians and IQRs were examined in addition to means. PPO+RCM static and dynamic scenes had similar trajectory lengths (109.63 ± 62.98 mm, median 97.69 [68.13, 119.09] mm vs. 106.64 ± 88.92 mm, median 94.39 [64.99, 126.81] mm; $p = 1.000$). Under dynamic conditions, RRT+RCM had a mean trajectory length of 121.53 ± 192.10 mm but a substantially lower median of 78.49 [54.22, 105.37] mm, reflecting a small number of very long trajectories (maximum 1108.21 mm). The PPO+RCM versus RRT+RCM trajectory-length difference was not statistically significant ($p = 0.137$).

Mean angular change was the clearest limitation of the original learned planner. PPO+RCM produced similar mean angular change in static and dynamic scenes (21.49 ± 3.33° vs. 21.66 ± 3.87°; medians 21.48 [19.49, 23.58]° vs. 21.83 [19.53, 23.57]°; $p = 1.000$). RRT+RCM showed substantially lower mean angular change under the dynamic protocol (2.95 ± 6.88°, median 1.45 [1.18, 1.83]°; $p < 0.001$ vs. PPO+RCM; $r_{rb} = 0.920$, 95% CI [0.800, 1.000]). This result identifies command/trajectory-direction irregularity as a specific computational weakness of the original PPO planner, rather than implying superiority of either method across all metrics.

3.5. Smoothness-Aware PPO+RCM Validation

Table 6 and Figure 8 compare PPO+RCM, RRT+RCM, and Smoothness-Aware PPO+RCM under the navigation-map perturbation protocol. Smoothness-Aware PPO+RCM achieved a target error of 0.616 ± 0.236 mm, median 0.644 [0.395, 0.826] mm, compared with 0.670 ± 0.201 mm, median 0.695 [0.492, 0.815] mm for PPO+RCM. The difference was small and not statistically significant ($p = 0.723$; $r_{rb} = -0.137$, 95% CI [−0.358, 0.085]). Against RRT+RCM, Smoothness-Aware PPO+RCM had lower target error ($p < 0.001$; $r_{rb} = -0.700$, 95% CI [−0.853, −0.528]). Differences in collision-risk proportion, minimum clearance, and trajectory length between Smoothness-Aware PPO+RCM and PPO+RCM were not significant ($p = 1.000$, 1.000, and 0.609, respectively). The collision-risk proportion was numerically lower than RRT+RCM but not statistically significant ($p = 0.117$).

The most pronounced change was mean angular change. Smoothness-Aware PPO+RCM decreased the mean angular change from 21.66 ± 3.87° to 1.07 ± 1.56°, with the median decreasing from 21.83 [19.53, 23.57]° to 0.53 [0.23, 1.08]°. The two distributions did not overlap in the validation data ($r_{rb} = -1.000$, bootstrap 95% CI [−1.000, −1.000]; $p < 0.001$). Relative to RRT+RCM, Smoothness-Aware PPO+RCM also had lower mean angular change (median 0.53° vs. 1.45°; $p < 0.001$; $r_{rb} = -0.485$, 95% CI [−0.696, −0.256]). These results support a great improvement in the geometric command/trajectory-direction metric.

RCM consistency was preserved after smoothness-aware fine-tuning (Figure 9). Full-trajectory time-averaged RCM error was 4.68 ± 4.70 mm for Smoothness-Aware PPO+RCM, compared with 5.73 ± 5.71 mm for PPO+RCM and 9.00 ± 7.43 mm for RRT+RCM. For all post-alignment samples, the corresponding pooled mean ± SD was 1.98 ± 1.74 mm for Smoothness-Aware PPO+RCM, 1.94 ± 1.66 mm for PPO+RCM, and 2.83 ± 3.29 mm for RRT+RCM. At the trajectory level, median [IQR] post-alignment mean RCM errors were 2.00 [1.42, 2.56], 1.93 [1.43, 2.72], and 2.12 [1.43, 2.81] mm, respectively. No difference was detected between Smoothness-Aware PPO+RCM and PPO+RCM ($p = 0.901$; $r_{rb} = -0.015$, 95% CI [−0.243, 0.205]) or RRT+RCM ($p = 0.581$; $r_{rb} = 0.065$, 95% CI [−0.303, 0.166]) at the trajectory-summary level. For all post-alignment samples, the P95 values were 5.46, 4.91, and 9.98 mm, and maxima were 8.06, 9.04, and 19.95 mm, respectively.

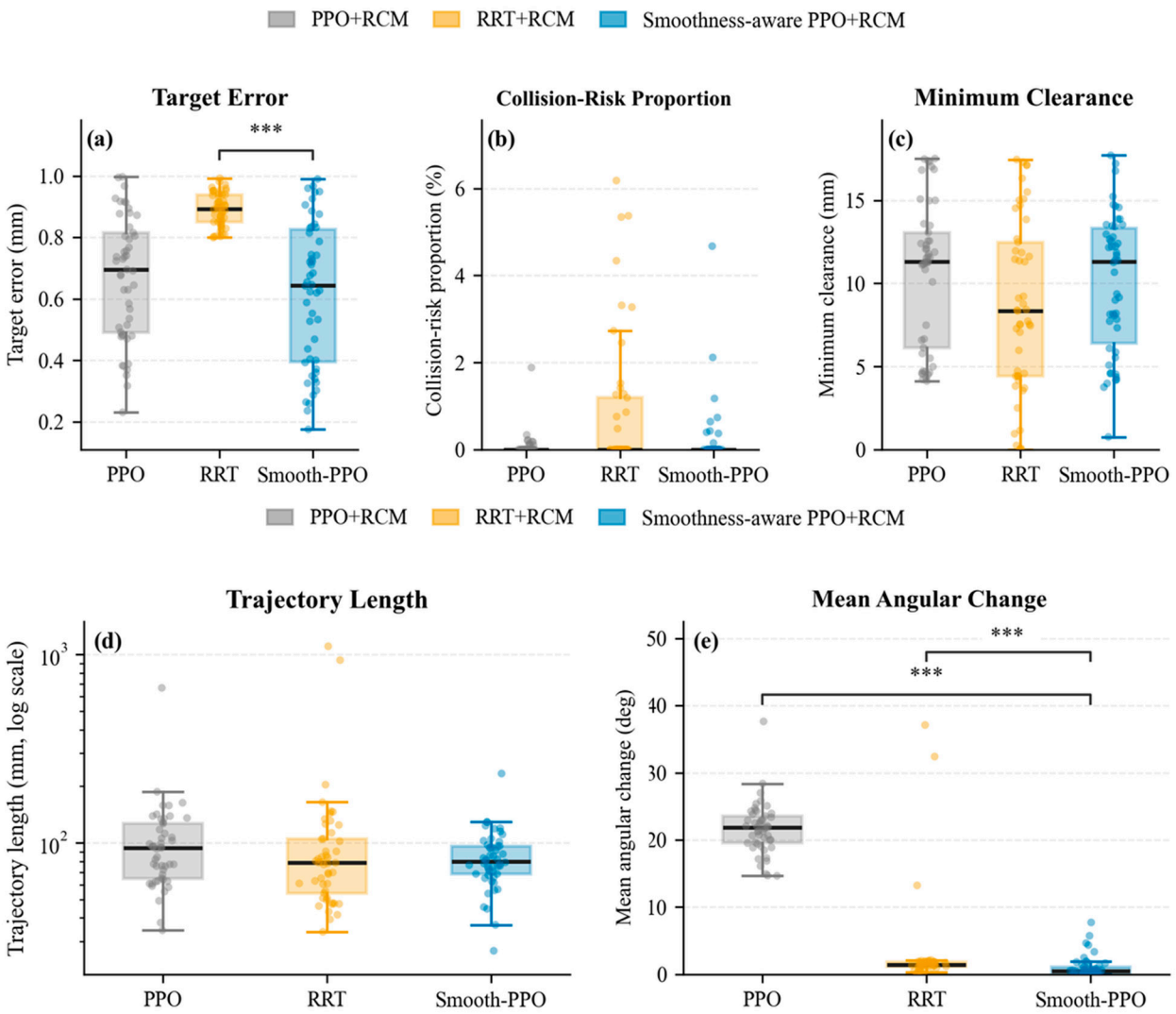


Figure 8. Comparison of (**a**) target error, (**b**) collision-risk proportion, (**c**) minimum clearance, (**d**) trajectory length, and (**e**) mean angular change under the navigation-map perturbation protocol. The trajectory-length axis in panel (**d**) uses a logarithmic scale. Brackets indicate pairwise comparisons; *** $p < 0.001$ after Holm correction.

Table 6. Smoothness-Aware PPO+RCM validation results. For the five geometric/trajectory metrics, the first line reports mean ± SD and the second line median [IQR].

Validation Setting	Target Error (mm)	Collision-Risk Proportion (%)	Minimum Clearance (mm)	Trajectory Length (mm)	Mean Angular Change (°)	Post-Align. RCM (mm)	Constrained Success
PPO+RCM dynamic	0.67 ± 0.20 0.70 [0.49, 0.82]	0.07 ± 0.27 0.00 [0.00, 0.00]	10.63 ± 4.31 11.29 [6.13, 13.06]	106.64 ± 88.92 94.39 [64.99, 126.81]	21.66 ± 3.87 21.83 [19.53, 23.57]	1.94 ± 1.66 1.93 [1.43, 2.72]	30/50
RRT+RCM dynamic	0.89 ± 0.05 0.89 [0.85, 0.94]	0.85 ± 1.61 0.00 [0.00, 1.19]	8.61 ± 5.22 8.35 [4.42, 12.48]	121.53 ± 192.10 78.49 [54.22, 105.37]	2.95 ± 6.88 1.45 [1.18, 1.83]	2.83 ± 3.29 2.12 [1.43, 2.81]	0/50
Smoothness-aware PPO+RCM dynamic	0.62 ± 0.24 0.64 [0.39, 0.83]	0.21 ± 0.74 0.00 [0.00, 0.00]	10.02 ± 4.16 11.31 [6.40, 13.31]	84.95 ± 31.52 79.62 [68.51, 96.33]	1.07 ± 1.56 0.53 [0.23, 1.08]	1.98 ± 1.74 2.00 [1.42, 2.56]	33/50

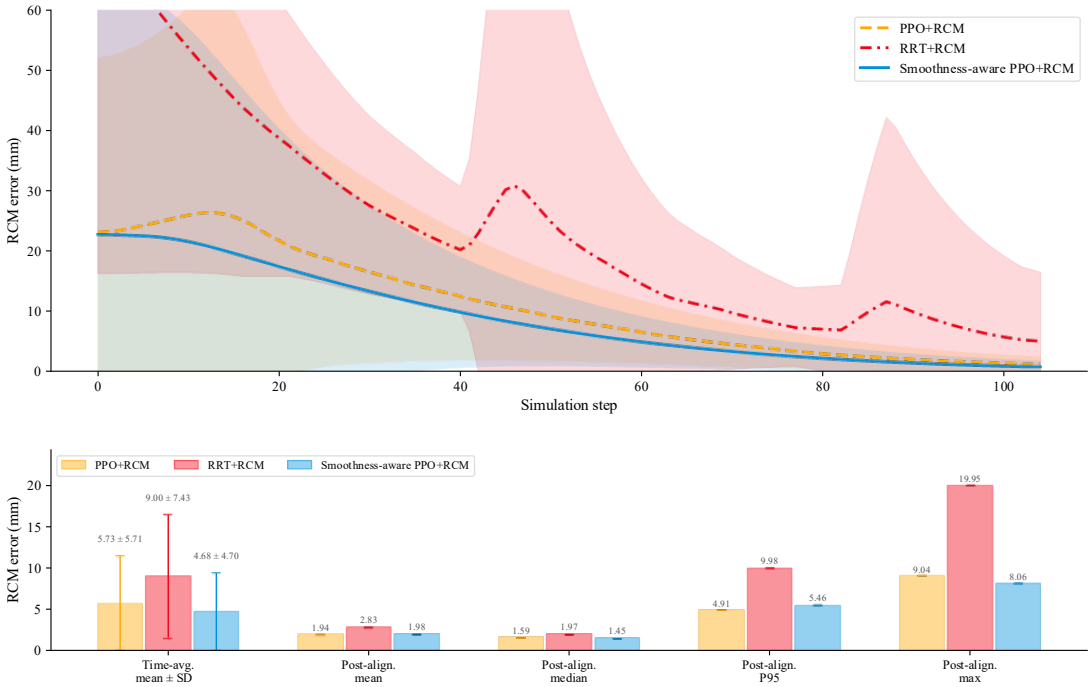


Figure 9. RCM-error evolution and stagewise statistics for PPO+RCM, RRT+RCM, and Smoothness-Aware PPO+RCM. The upper panel shows RCM-error evolution: orange dashed, red dash-dotted, and blue solid curves represent PPO+RCM, RRT+RCM, and Smoothness-Aware PPO+RCM, respectively. Shaded bands indicate ±1 sample standard deviation around a centered five-step moving average. The lower panel presents the full-trajectory time-averaged mean ± SD and the post-alignment mean, median, 95th percentile (P95), and maximum RCM error, using the same method colors.

3.5.1. Smoothness-Weight Sensitivity

To examine sensitivity to the selected action-transition weight, $w_{\Delta a}$ =0.025, 0.05, and 0.10 were evaluated under the same training and validation settings, with all other parameters unchanged. As shown in Table 7, $w_{\Delta a} = 0.05$ yielded the lowest mean angular change ($1.06 \pm 1.55°$), while target error, trajectory length, and post-alignment RCM performance remained comparable across the tested settings. Therefore, $w_{\Delta a} = 0.05$ was retained as the nominal setting for the main evaluation.

Table 7. Sensitivity of Smoothness-Aware PPO+RCM to the action-transition weight.

wΔa	Target Error (mm)	Trajectory Length (mm)	Mean Angular Change (°)	Post-Align. Mean RCM (mm)
0.025	0.598 ± 0.243	88.79 ± 36.72	1.84 ± 2.85	2.66
0.050	0.617 ± 0.237	84.96 ± 31.52	1.06 ± 1.55	1.98
0.100	0.633 ± 0.231	87.08 ± 34.44	2.07 ± 3.79	2.20

3.5.2. Temporal Motion Metrics

The TCP-level temporal metrics are summarized in Table 8. Smoothness-Aware PPO+RCM maintained a mean TCP speed comparable to PPO+RCM (median 27.10 vs. 27.75 mm/s; $p = 0.302$), while its RMS acceleration and RMS jerk were markedly lower (16.40 vs. 294.79 mm/s^2 and 82.60 vs. 7306.06 mm/s^3, respectively; both $p < 0.001$). The corresponding rank-biserial effect sizes were −0.949 (95% CI [−1.000, −0.856]) for RMS acceleration and −0.964 (95% CI [−1.000, −0.887]) for RMS jerk. Relative to RRT+RCM, Smoothness-Aware PPO+RCM also showed lower RMS acceleration (16.40 vs. 35.84 mm/s^2; $p < 0.001$) and RMS jerk (82.60 vs. 361.73 mm/s^3; $p < 0.001$), despite a slightly higher mean TCP speed (27.10 vs. 24.20 mm/s; $p = 0.016$). These results support the interpretation

that the reduction in mean angular change was accompanied by lower simulated TCP acceleration and jerk rather than being explained simply by slower motion.

Table 8. TCP-level temporal motion metrics under the navigation-map perturbation protocol. Values are median [IQR] across validation trajectories.

Validation Setting	Mean TCP Speed (mm/s)	RMS Acceleration (mm/s^2)	RMS Jerk (mm/s^3)
PPO+RCM	27.75 [26.06, 28.99]	294.79 [264.58, 309.71]	7306.06 [6592.27, 7814.62]
RRT+RCM	24.20 [22.84, 28.24]	35.84 [29.74, 40.62]	361.73 [317.13, 442.71]
Smoothness-Aware PPO+RCM	27.10 [24.43, 29.22]	16.40 [10.79, 27.95]	82.60 [42.39, 158.65]

3.6. Unseen-Patient Validation

The seven patient-specific anatomical environments used for model development and within-cohort validation do not constitute a strict test of transfer to unseen anatomy. Smoothness-Aware PPO+RCM was therefore evaluated on three patient cases that were completely excluded from training. Ten test scenes were generated for each held-out patient, resulting in 30 evaluation scenes without patient-specific fine-tuning.

Patient-wise results are summarized in Table 9. Across the three unseen patients, mean target error ranged from 0.58 to 0.78 mm and no direct geometric intersections with the simplified risk models were observed in the 30 scenes. Mean collision-risk proportion was 0.0067% for Patient 1 and 0% for Patients 2 and 3. Mean minimum clearance ranged from 4.36 to 11.86 mm, and mean angular change ranged from 0.37° to 1.56°. Whole-trajectory mean RCM errors were 4.41 ± 4.91, 2.71 ± 2.30, and 3.14 ± 2.02 mm for Patients 1–3, respectively; post-alignment mean RCM errors were 2.04 ± 0.94, 1.88 ± 0.79, and 2.34 ± 0.97 mm. Composite Constrained Task Success was 6/10, 6/10, and 5/10, yielding 17/30 successes overall.

The unseen-patient validation therefore supports retention of the measured computational targeting, geometric-clearance, command-regularity, and RCM characteristics across three anatomically distinct cases without patient-specific fine-tuning.

Among the 13 unsuccessful unseen-patient trials, 12 failed the target-error criterion, and one failed the post-alignment RCM criterion; the target-error criterion was therefore the predominant failure mode.

Table 9. Unseen-patient validation of Smoothness-Aware PPO+RCM across three held-out patient-specific anatomical environments. Each patient contributed 10 independently initialized scenes; no direct geometric intersections were observed in the 30 scenes. Values are mean ± SD unless otherwise indicated.

Patient	Target Error (mm)	Collision-Risk Proportion (%)	Minimum Clearance (mm)	Mean Angular Change (°)	Whole-Traj. RCM (mm)	Post-Align. RCM (mm)	Composite Success
Patient 1	0.75 ± 0.20	0.0067 ± 0.0064	4.36 ± 0.39	0.37 ± 0.22	4.41 ± 4.91	2.04 ± 0.94	6/10
Patient 2	0.58 ± 0.24	0	11.43 ± 0.18	0.90 ± 0.87	2.71 ± 2.30	1.88 ± 0.79	6/10
Patient 3	0.78 ± 0.18	0	11.86 ± 1.14	1.56 ± 2.42	3.14 ± 2.02	2.34 ± 0.97	5/10

Overall failure modes among the 13 unsuccessful trials: 12 target-error failures; 1 post-alignment RCM failure.

A representative Smoothness-Aware PPO+RCM trajectory is shown in Figure 10 together with the patient-specific anatomical risk structures. The global and enlarged local views illustrate the spatial relationship between the planned trajectory, target position, nearby vessels, and critical nuclei, including the right ventral intermediate nucleus of the thalamus (VIM-R) highlighted in the local view.

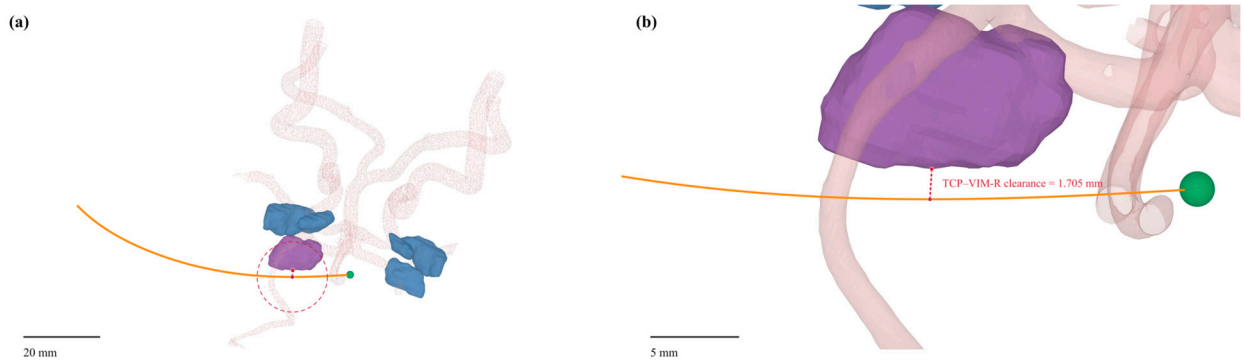


Figure 10. Representative Smoothness-Aware PPO+RCM trajectory in a patient-specific anatomical scene: (**a**) global view; (**b**) enlarged view of the region marked by the red dashed circle in (**a**). The orange curve represents the TCP trajectory, and the green marker denotes the target position. The purple structure represents the right ventral intermediate nucleus of the thalamus (VIM-R), blue structures indicate the other displayed critical nuclei, and translucent pink structures represent vessels. The red dashed segment in (**b**) indicates the illustrated TCP-to-VIM-R clearance of 1.705 mm. Scale bars: (**a**) 20 mm; (**b**) 5 mm.

3.7. Composite Constrained Task Success

Composite Constrained Task Success required target accuracy, direct-collision avoidance, minimum anatomical clearance, and post-alignment RCM consistency to be satisfied simultaneously. In the preliminary five-setting experiment, PPO+RCM achieved constrained success in 26/50 static trials and 30/50 dynamic trials; RRT+RCM and Straight-line+RCM achieved 0/50, and PPO without RCM achieved 2/50. With the smoothness-aware extension, Smoothness-Aware PPO+RCM achieved 33/50 constrained successes (66%), compared with 30/50 (60%) for dynamic PPO+RCM and 0/50 for RRT+RCM (Figure 11).

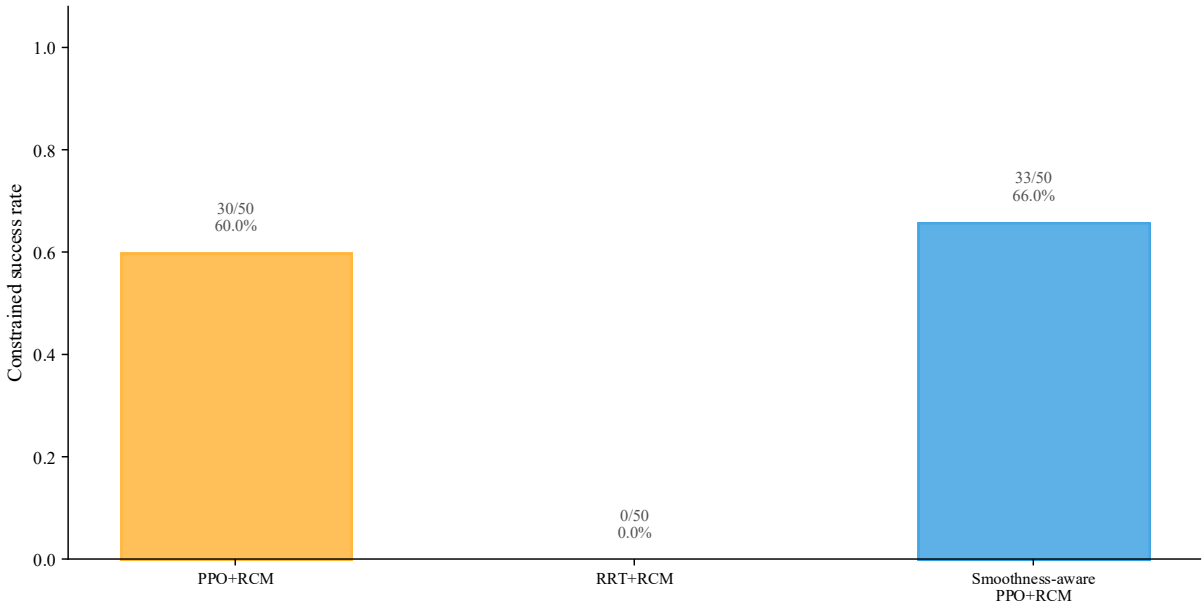


Figure 11. Composite Constrained Task Success for the three dynamic RCM-constrained planning methods. Success required final target error ≤ 0.75 mm, no direct collision, minimum clearance > 1 mm, and post-alignment mean RCM error ≤ 5 mm.

The threshold-sensitivity analysis is summarized in Table 10. At the nominal criterion (target error ≤ 0.75 mm and post-alignment mean RCM error ≤ 5 mm), success was 30/50

for PPO+RCM, 33/50 for Smoothness-Aware PPO+RCM, and 0/50 for RRT+RCM. With the target threshold held at 0.75 mm, tightening the RCM threshold from 5 to 3 and 2 mm reduced success to 26/50 and 16/50 for PPO+RCM and to 29/50 and 17/50 for Smoothness-Aware PPO+RCM. The smoothness-aware policy therefore retained slightly more composite successes than PPO+RCM across all three RCM thresholds at the 0.75-mm target threshold. However, the composite outcome was strongly threshold-dependent. RRT+RCM remained at 0/50 for target-error thresholds of 0.50 and 0.75 mm, but increased to 19/50, 36/50, and 42/50 when the target threshold was relaxed to 1.00 mm with RCM thresholds of 2, 3, and 5 mm, respectively. This pattern indicates that the 0/50 result primarily reflects failure to satisfy the strict composite target-error criterion rather than an inability of RRT+RCM to generate geometrically constrained trajectories.

Table 10. Sensitivity of Composite Constrained Task Success to target-error and post-alignment RCM-error thresholds.

Method	Target-Error Threshold (mm)	RCM ≤ 2 mm	RCM ≤ 3 mm	RCM ≤ 5 mm
PPO+RCM	0.50	7/50 (14%)	12/50 (24%)	13/50 (26%)
PPO+RCM	0.75	16/50 (32%)	26/50 (52%)	30/50 (60%)
PPO+RCM	1.00	25/50 (50%)	42/50 (84%)	47/50 (94%)
RRT+RCM	0.50	0/50 (0%)	0/50 (0%)	0/50 (0%)
RRT+RCM	0.75	0/50 (0%)	0/50 (0%)	0/50 (0%)
RRT+RCM	1.00	19/50 (38%)	36/50 (72%)	42/50 (84%)
Smoothness-Aware PPO+RCM	0.50	9/50 (18%)	15/50 (30%)	17/50 (34%)
Smoothness-Aware PPO+RCM	0.75	17/50 (34%)	29/50 (58%)	33/50 (66%)
Smoothness-Aware PPO+RCM	1.00	24/50 (48%)	42/50 (84%)	48/50 (96%)

4. Discussion

4.1. Performance of the Hierarchical Planning-And-Execution Framework

The results support the engineering rationale of separating adaptive local planning from explicit kinematic constraint execution. The original PPO+RCM configuration maintained essentially unchanged target error under the dynamic-map perturbation (0.67 ± 0.23 mm static vs. 0.67 ± 0.20 mm perturbation), while collision-risk proportion and minimum clearance were also stable. Compared with RRT+RCM, PPO+RCM had lower target error ($p < 0.001$), whereas the collision-risk-proportion difference was not statistically significant ($p = 0.137$). The module controls further separated the roles of the two layers: Straight-line+RCM preserved entry-point consistency but did not meet the strict target-error criterion, whereas PPO without RCM often reached the target but produced substantially larger RCM deviation. Thus, the planner and execution layer address complementary subproblems.

The comparison with RRT+RCM also clarifies a trade-off of the framework. PPO+RCM achieved lower target error, whereas the shortcut-pruned and B-spline-smoothed RRT trajectories exhibited substantially smaller mean angular change. Trajectory-length distributions contained strong right-tail outliers, especially for RRT+RCM, so medians and IQRs are reported alongside mean ± SD. RRT+RCM is retained as a classical sampling-based reference under the same RCM execution layer rather than as the strongest available online planner.

The composite-score sensitivity analysis further shows that this summary metric depends materially on the selected engineering thresholds. At the prespecified 0.75-mm target-error threshold, Smoothness-Aware PPO+RCM retained modestly more successes than PPO+RCM across RCM limits of 2–5 mm; however, when the target threshold was relaxed to 1.00 mm, the two PPO-based methods became similar and RRT+RCM achieved substantial nonzero success. Accordingly, the composite score is used here as a compact comparative engineering endpoint rather than evidence of clinical success or universal method superiority.

4.2. Effect of the Smoothness-Aware PPO+RCM Extension

The action-transition penalty directly addressed the command-regularity limitation of PPO+RCM. Smoothness-Aware PPO+RCM reduced mean angular change from $21.66 \pm 3.87°$ to $1.07 \pm 1.56°$, while target error, collision-risk proportion, minimum clearance, and trajectory length did not show detectable deterioration. The contribution is therefore interpreted as an incremental planner-level regularization of the hierarchical framework. Because mean angular change is a geometric proxy, temporal behavior was also evaluated using TCP speed, acceleration, and jerk.

The TCP-level temporal analysis is consistent with this interpretation: Smoothness-Aware PPO+RCM preserved a mean execution speed comparable to PPO+RCM while substantially reducing RMS acceleration and jerk, and it also showed lower RMS acceleration and jerk than RRT+RCM. These quantities remain simulator-level kinematic measures; they do not establish physical robot tracking smoothness, actuator loading, or tissue-interaction performance.

The smoothness-aware policy also preserved trajectory-level post-alignment RCM behavior. The per-trajectory median [IQR] post-alignment mean RCM error was 2.00 [1.42, 2.56] mm for Smoothness-Aware PPO+RCM and 1.93 [1.43, 2.72] mm for PPO+RCM, with no detected difference ($p = 0.901$). The full-trajectory time-averaged error for Smoothness-Aware PPO+RCM was 4.68 ± 4.70 mm, reflecting the initial alignment transient as well as subsequent constrained motion. Because the lower-level RCM controller was unchanged during fine-tuning, the reduced mean angular change can be attributed to planner-level regularization rather than altered kinematic constraint handling.

4.3. Generalization to Unseen Patient Anatomy

Randomization of scene pose and cranial entry point exposes the policy to heterogeneous tasks, but it does not by itself establish transfer to anatomy absent from training. In the three unseen patients, target error remained 0.58–0.78 mm, post-alignment RCM error 1.88–2.34 mm, and mean angular change 0.37–1.56°, with 17/30 scenes satisfying the composite engineering criterion. These results support limited patient-wise transfer evidence.

The generalization claim must nevertheless remain limited. Three patients are too few to represent the distribution of neuroanatomical variability, and ten randomized scenes from each patient are not equivalent to ten independent patients. The held-out results therefore support proof-of-concept patient-wise transfer but do not establish population-level robustness.

4.4. Interpretation of Full-Trajectory and Post-Alignment RCM Error

RCM error was recorded from the initial robot configuration, before the tool reached the cranial entry point. The alignment criterion was satisfied after the tool entered the cranial entry point. Therefore, the full-trajectory RCM error includes the initial approach and RCM alignment process, whereas the post-alignment RCM error characterizes the

stabilized constraint performance after alignment. RCM error is defined as the shortest perpendicular distance from the cranial entry point to the tool axis.

For dynamic PPO+RCM, the full-trajectory mean RCM error was 5.73 ± 5.71 mm, and the post-alignment pooled mean was 1.94 mm. Smoothness-Aware PPO+RCM showed the same pattern, with 4.68 ± 4.70 mm and 1.98 mm, respectively. The 5-mm post-alignment threshold used in the composite score is retained only as a study-specific engineering criterion and is not interpreted as a clinically validated tolerance.

4.5. Methodological Limitations and Future Robotic Validation

Several limitations remain. First, this is an exclusively computational study and does not model force interaction, deformable tissue mechanics, sensing noise, communication delay, registration error, or physical robot dynamics. The multimodal images were used to construct the computational scene rather than to evaluate image-acquisition or segmentation performance as study endpoints. Scanner-specific acquisition effects and uncertainty from co-registration and segmentation were not explicitly propagated through the planning and clearance analyses; therefore, the reported geometric metrics remain conditional on the reconstructed image models. Second, the repeated 2-mm perturbations are a navigation-map stress test rather than a biomechanical brain-shift model. Third, vessels and nuclei are simplified to centerline/Voronoi and OBB representations, so geometric collision and clearance metrics can underestimate or distort true surface contact risk. Fourth, the validation groups were independently randomized rather than evaluated on matched scene realizations, and scene-level statistical tests arise from a small number of independent patient anatomies. Fifth, the RRT comparator is not intended as the strongest modern online planner. Sixth, only three held-out patients were available. Seventh, the added TCP-level acceleration and jerk metrics characterize simulator-level kinematic regularity and do not establish physical tracking smoothness, actuator loading, or tissue-interaction performance. Finally, the original PPO training did not use a recorded fixed random seed and independent repeated training runs were not available; therefore, between-run variability attributable to stochastic initialization and optimization could not be quantified.

Future work should prioritize matched-scene computational comparisons with optimization/MPC or safety-filter baselines, uncertainty sensitivity analyses for registration and geometry, larger held-out patient cohorts, repeated training with controlled random seeds, and physical robot/phantom experiments using the same planning-and-execution interface. Experimental translation should quantify real-time tracking, command latency, joint limits, Jacobian conditioning, RCM entry-point motion, and contact/force effects before any claim of surgical readiness.

5. Conclusions

This study presents a computational evaluation of a Hierarchical Planning-and-Execution Framework that separates learning-based local motion planning from analytical RCM-constrained robot execution for neuroendoscopic access. The module-control experiments show that the two layers address complementary requirements: the planner supports target-directed, obstacle-responsive motion, while the execution layer maintains entry-point consistency. Smoothness-Aware PPO+RCM reduced mean angular change from 21.66 ± 3.87° to 1.07 ± 1.56° ($p < 0.001$) while maintaining comparable target error ($p = 0.723$). Full-trajectory time-averaged RCM error was 4.68 ± 4.70 mm, and the post-alignment RCM distribution remained comparable to PPO+RCM. In three held-out patients, mean target error remained 0.58–0.78 mm, post-alignment RCM error was 1.88–2.34 mm, and 17/30 scenes met the composite engineering criterion. These results support the computational feasibility of improving planner command regularity without degrading the

measured targeting, geometric-clearance, or RCM performance. Physical robot validation remains necessary before surgical translation.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app162110723/s1: Figure S1: Training reward curve of the original PPO policy over 1,700,000 environment steps and 11,077 completed episodes. A 113-episode moving average was used for visualization.

Author Contributions: Y.L.: conceptualization, methodology, software, validation, formal analysis, visualization, and writing—original draft. B.P.: supervision, project administration, methodology review, and writing—review and editing. X.W.: methodology review and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Beijing Natural Science Foundation (L252112).

Institutional Review Board Statement: The study was approved by the Biological and Medical Ethics Committee of Beihang University (protocol code BM20250192; approval date: 11 April 2025).

Informed Consent Statement: The requirement for informed consent was waived by the Biological and Medical Ethics Committee of Beihang University because this study involved retrospective analysis of existing de-identified imaging data and posed no additional risk to the participants.

Data Availability Statement: Model-training parameters, trained policy weights, non-patient-specific simulation configurations, analysis scripts, and analysis data supporting the reported numerical results are available from the corresponding author upon reasonable request for research purposes. Raw patient CT, MRI, and MRA data cannot be publicly released because of institutional, ethical, and patient-privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hallenberger, T.J.; Fischer, U.; Bonati, L.H.; Dutilh, G.; Mucklow, R.; Vogt, A.S.; Boeni-Eckstein, C.; Cardia, A.; Schubert, G.A.; Bijlenga, P.; et al. Early minimally invasive image-guided endoscopic evacuation of intracerebral hemorrhage (EMINENT-ICH): A randomized controlled trial. *Trials* **2024**, *25*, 692. [CrossRef] [PubMed]
2. Ma, H.; Peng, W.; Xu, S.; Liang, X.; Zhao, R.; Lv, M.; Guan, F.; Zhu, G.; Mao, B.; Hu, Z. Advancements of endoscopic surgery for spontaneous intracerebral hemorrhage. *World Neurosurg.* **2025**, *193*, 160–170. [CrossRef] [PubMed]
3. Yan, Z.; Jiang, L.; Li, G.; Xia, K.; Peng, L.; Hu, J.; Chen, S.; Zhang, J.; Huang, X. Efficacy study of neuronavigation-assisted stereotactic drilling of urokinase drainage versus craniotomy in the treatment of massive intracerebral haemorrhage in elderly patients. *Sci. Rep.* **2024**, *14*, 20439. [CrossRef] [PubMed]
4. Li, M.; Gao, Y.; Wang, C.; Li, A.; Liu, Z.; Zhang, Y.; Yang, L.; Yu, Y.; Zhen, X. A comparison between robot-assisted minimally invasive surgery and conventional craniotomy for intracerebral hemorrhage: Efficiency, complications and outcomes. *Front. Neurol.* **2026**, *16*, 1688309. [CrossRef] [PubMed]
5. Sadeghian, H.; Zokaei, F.; Hadian Jazi, S. Constrained kinematic control in minimally invasive robotic surgery subject to remote center of motion constraint. *J. Intell. Robot. Syst.* **2019**, *95*, 901–913. [CrossRef]
6. Bian, G.B.; Chen, Z.; Li, Z.; Wei, B.-T.; Liu, W.-P.; da Silva, D.S.; Wu, W.-Q.; de Albuquerque, V.H.C. Learning surgical skills under the RCM constraint from demonstrations in robot-assisted minimally invasive surgery. *Expert Syst. Appl.* **2023**, *225*, 120134. [CrossRef]
7. Marinho, M.M.; Bernardes, M.C.; Bó, A.P.L. Using general-purpose serial-link manipulators for laparoscopic surgery with moving remote center of motion. *J. Med. Robot. Res.* **2016**, *1*, 1650007. [CrossRef]
8. Zhang, W.; Wang, Z.; Ma, K.; Liu, F.; Cheng, P.; Ding, X. State of the art in movement around a remote point: A review of remote center of motion in robotics. *Front. Mech. Eng.* **2024**, *19*, 14. [CrossRef]
9. Colan, J.; Davila, A.; Fozilov, K.; Hasegawa, Y. A concurrent framework for constrained inverse kinematics of minimally invasive surgical robots. *Sensors* **2023**, *23*, 3328. [CrossRef] [PubMed]
10. Moustris, G.; Tzafestas, C. Modeling, Simulation and Control of the Double Delta Surgical Robot. *Machines* **2024**, *12*, 620. [CrossRef]
11. Koizumi, S.; Kin, T.; Shono, N.; Kiyofuji, S.; Umekawa, M.; Sato, K.; Saito, N. Patient-specific cerebral 3D vessel model reconstruction using deep learning. *Med. Biol. Eng. Comput.* **2024**, *62*, 3225–3232. [CrossRef] [PubMed]

12. Monfaredi, R.; Concepcion-Gonzalez, A.; Acosta Julbe, J.; Fischer, E.; Hernandez-Herrera, G.; Cleary, K.; Oluigbo, C. Automatic path-planning techniques for minimally invasive stereotactic neurosurgical procedures—A systematic review. *Sensors* **2024**, *24*, 5238. [CrossRef] [PubMed]
13. Li, C.; Evans, L.T.; Hong, J.; Davis, S.C.; Roberts, D.W.; Paulsen, K.D.; Fan, X. Fully automated image updating for brain shift compensation after dural opening. *J. Neurosurg.* **2026**, *144*, 206–216. [CrossRef] [PubMed]
14. Gong, W.; Luo, Y.; Yang, F.; Zhou, H.; Lin, Z.; Cai, C.; Lin, Y.; Chen, J. ETDformer: An effective transformer block for segmentation of intracranial hemorrhage. *Med. Biol. Eng. Comput.* **2025**, *63*, 2355–2372. [CrossRef] [PubMed]
15. Karaman, S.; Frazzoli, E. Sampling-based algorithms for optimal motion planning. *Int. J. Robot. Res.* **2011**, *30*, 846–894. [CrossRef]
16. Ratliff, N.; Zucker, M.; Bagnell, J.A.; Srinivasa, S.S. CHOMP: Gradient optimization techniques for efficient motion planning. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, 12–17 May 2009; pp. 489–494. [CrossRef]
17. Kalakrishnan, M.; Chitta, S.; Theodorou, E.; Pastor, P.; Schaal, S. STOMP: Stochastic trajectory optimization for motion planning. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 4569–4574. [CrossRef]
18. Schulman, J.; Duan, Y.; Ho, J.; Lee, A.; Awwal, I.; Bradlow, H.; Pan, J.; Patil, S.; Goldberg, K.; Abbeel, P. Motion planning with sequential convex optimization and convex collision checking. *Int. J. Robot. Res.* **2014**, *33*, 1251–1270. [CrossRef]
19. LaValle, S.M.; Kuffner, J.J. Randomized Kinodynamic Planning. *Int. J. Robot. Res.* **2001**, *20*, 378–400. [CrossRef]
20. Su, H.; Zhang, J.; She, Z.; Zhang, X.; Fan, K.; Zhang, X.; Liu, Q.; Ferrigno, G.; De Momi, E. Incorporating model predictive control with fuzzy approximation for robot manipulation under remote center of motion constraint. *Complex Intell. Syst.* **2022**, *8*, 2883–2895. [CrossRef]
21. Hsu, K.-C.; Hu, H.; Fisac, J.F. The Safety Filter: A Unified View of Safety-Critical Control in Autonomous Systems. *Annu. Rev. Control Robot. Auton. Syst.* **2024**, *7*, 47–72. [CrossRef]
22. Jian, Y.; Jin, Y.; Price, M.; Moore, J. A Parallel Robot with Remote Centre-of-Motion for Eye Surgery: Design, Kinematics, Prototype, and Experiments. *Int. J. Med. Robot. Comput. Assist. Surg.* **2024**, *20*, e2665. [CrossRef] [PubMed]
23. Xie, J.; Shao, Z.; Li, Y.; Guan, Y.; Tan, J. Deep Reinforcement Learning with Optimized Reward Functions for Robotic Trajectory Planning. *IEEE Access* **2019**, *7*, 105669–105679. [CrossRef]
24. Segato, A.; Sestini, L.; Castellano, A.; De Momi, E. GA3C Reinforcement Learning for Surgical Steerable Catheter Path Planning. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 2429–2435. [CrossRef]
25. Ao, Y.; Esfandiari, H.; Carrillo, F.; Laux, C.J.; As, Y.; Li, R.; Van Assche, K.; Davoodi, A.; Cavalcanti, N.A.; Farshad, M.; et al. SafeRPlan: Safe Deep Reinforcement Learning for Intraoperative Planning of Pedicle Screw Placement. *Med. Image Anal.* **2025**, *99*, 103345. [CrossRef] [PubMed]
26. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347.
27. Li, Y.; Pei, B.; Wu, X. RCM-Constrained Hierarchical Planning for Robot-Assisted Neuroendoscopic Access in Patient-Specific Multimodal Anatomical Scenes. In Proceedings of the 8th International Conference on Robotics and Computer Vision (ICRCV 2026), Jiangyin, China, 6–8 November 2026.
28. Henschel, L.; Conjeti, S.; Estrada, S.; Diers, K.; Fischl, B.; Reuter, M. FastSurfer—A fast and accurate deep learning based neuroimaging pipeline. *NeuroImage* **2020**, *219*, 117012. [CrossRef] [PubMed]
29. Antiga, L.; Piccinelli, M.; Botti, L.; Ene-Iordache, B.; Remuzzi, A.; Steinman, D.A. An image-based modeling framework for patient-specific computational hemodynamics. *Med. Biol. Eng. Comput.* **2008**, *46*, 1097–1112. [CrossRef] [PubMed]
30. China Emergency Neuroendoscopy Collaboration Group for Hypertensive Intracerebral Hemorrhage; Committee on Neuroendoscopy and Minimally Invasive Medicine, China Medicine Education Association; Chinese Society of Neurosurgery, Chinese Medical Association. 2020 Chinese Expert Consensus on Neuroendoscopic Surgical Treatment of Hypertensive Intracerebral Hemorrhage. *Natl. Med. J. China* **2020**, *100*, 2579–2585. (In Chinese) [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.