

# A Mixed Reality System for Robust Manikin Localization in Childbirth Training

Haojie Cheng , Chang Liu , Abhiram Kanneganti , Mahesh Arjandas Choolani ,  
Arundhati Tushar Gosavi , and Eng Tat Khoo\* 

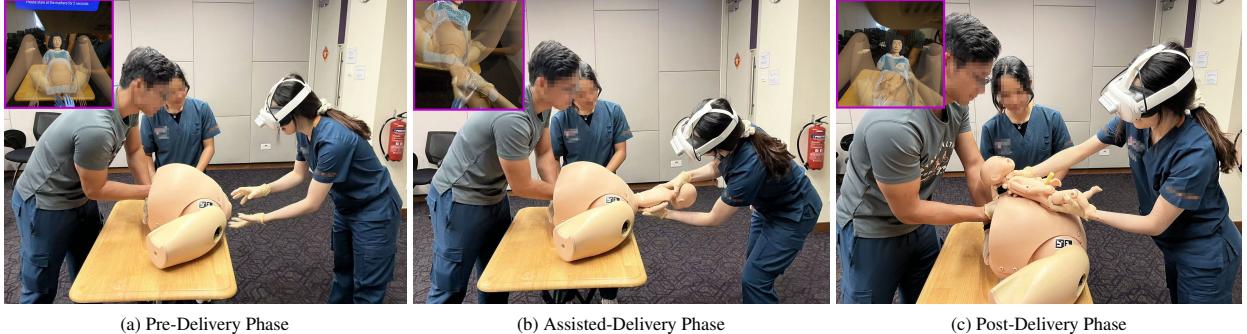


Fig. 1: Mixed Reality-assisted childbirth training using a manikin setup. The primary trainee wears a Meta Quest headset with a front-facing RGB-D camera for real-time manikin localization. Two supporting trainees assist by pushing the neonatal manikin and stabilizing the manikin. The top-left image shows the trainee's first-person view with virtual guidance.

**Abstract**— Opportunities for medical students to gain practical experience in vaginal births are increasingly constrained by shortened clinical rotations, patient reluctance, and the unpredictable nature of labour. To alleviate clinicians' instructional burden and enhance trainees' learning efficiency, we introduce a mixed reality (MR) system for childbirth training that combines virtual guidance with tactile manikin interaction, thereby preserving authentic haptic feedback while enabling independent practice without continuous on-site expert supervision. The system extends the passthrough capability of commercial head-mounted displays (HMDs) by spatially calibrating an external RGB-D camera, allowing real-time visual integration of physical training objects. Building on this capability, we implement a coarse-to-fine localization pipeline that first aligns the maternal manikin with fiducial markers to define a delivery region and then registers the pre-scanned neonatal head within this area. This process enables spatially accurate overlay of virtual guiding hands near the manikin, allowing trainees to follow expert trajectories reinforced by haptic interaction. Experimental evaluations demonstrate that the system achieves accurate and stable manikin localization on a standalone headset, ensuring practical deployment without external computing resources. A large-scale user study involving 83 fourth-year medical students was subsequently conducted to compare MR-based and virtual reality (VR)-based childbirth training. Four senior obstetricians independently assessed performance using standardized criteria. Results showed that MR training achieved significantly higher scores in delivery, post-delivery, and overall task performance, and was consistently preferred by trainees over VR training. Although validated only in the context of obstetric delivery, the system demonstrates strong potential for broader manikin-based procedural training and other healthcare education scenarios.

**Index Terms**—Mixed reality, Head-mounted displays, Manikin localization, Real-time tracking, Childbirth training.

## 1 INTRODUCTION

Witnessing the start of life in the labour ward is the hallmark of every medical student's Obstetrics & Gynaecology (O&G) rotation [1]. In recent years, however, there has been a trend away from demonstrating competence in independently conducting routine vaginal births (VBs) towards one of participation or observation due to competing factors such as limited posting durations, patient aversion to male students [2]. This reduction in opportunities for medical students to participate in

VBs is worrying considering that non-obstetricians and -midwives are frequently first responders to unexpectedly precipitous VBs occurring out-of-hospital or in acute care settings [3,4] and that obstetrics comprises a large and continually rising proportion of medicolegal liabilities in developed countries [5,6]. Traditional group-based teaching methods include a combination of didactic lectures, manikin simulation sessions, and delivery suite tours [7–9]. While lectures and materials can be standardized, practical exposure to childbirth remains limited and unpredictable. These constraints underscore the need for more efficient training methods that ensure students gain both theoretical knowledge and hands-on experience in childbirth procedures.

The growing adoption of immersive technologies, particularly virtual reality (VR) and mixed reality (MR), is reshaping medical education by offering scalable, interactive training solutions [10–12]. In childbirth training, where clinical exposure is often limited, these technologies present valuable alternatives. VR enables the simulation of fully virtual delivery suites with interactive guidance for childbirth procedures [13]. Previous research [14] demonstrated that students who trained with a VR-based childbirth simulator achieved significantly greater knowledge gains compared to those who underwent traditional manikin-based simulation. However, a key limitation of VR training lies in its fully virtual environment, which precludes interaction with the physical

• Haojie Cheng, Chang Liu, Mahesh Arjandas Choolani, Arundhati Tushar Gosavi, and Eng Tat Khoo are with National University of Singapore.

E-mail: [hjcheng@nus.edu.sg](mailto:hjcheng@nus.edu.sg), [chang\\_lc@nus.edu.sg](mailto:chang_lc@nus.edu.sg), [obghead@nus.edu.sg](mailto:obghead@nus.edu.sg), [obggat@nus.edu.sg](mailto:obggat@nus.edu.sg), [etkhoo@nus.edu.sg](mailto:etkhoo@nus.edu.sg).

• Abhiram Kanneganti is with National University Health System, Singapore.  
E-mail: [abhiram\\_kanneganti@nuhs.edu.sg](mailto:abhiram_kanneganti@nuhs.edu.sg).

• \* Corresponding author.

surroundings and lacks tactile feedback. Other studies have also shown that tactile feedback is crucial in learning procedural skills [15–19]. To bridge this gap, MR offers a hybrid approach that combines virtual guidance with physical interaction [20–23]. In childbirth training, MR allows students to perform procedures on physical manikins while receiving real-time visual cues, such as hand manoeuvre guidance. This integration retains the tactile realism of manikin-based training while enhancing it with interactive and adaptive instruction.

However, integrating MR into childbirth training for large-scale medical education presents several technical and practical challenges. A key difficulty lies in accurately localizing physical manikins within the MR environment. During delivery simulation, expulsion forces and abundant lubrication, coupled with the non-rigid structure and silicone material of the manikin, which makes robust localization particularly difficult. To be suitable for widespread use in medical education settings, the MR system must be simple to set up and consistently deployable, enabling large cohorts of students to engage with the system within their limited rotation times. It must also be autonomous and easy to operate, allowing medical students with no technical background to use the system independently without the assistance of specialized personnel or the need for complex calibration procedures.

To address these challenges, we present the first MR system for childbirth training, targeting normal delivery as the most common and fundamental clinical scenario. The system emphasizes essential competencies in the correct execution of delivery maneuvers, confident post-delivery handling, and accurate neonatal placement on the maternal abdomen. To enable spatial alignment between virtual and physical models, an external RGB-D camera is integrated with the headset and calibrated through an improved eye-to-hand method. Accurate localization is then achieved through a coarse-to-fine registration pipeline: the virtual maternal model is first aligned to the manikin using fiducial markers to define a structurally correct delivery region of interest (ROI), and the pre-scanned neonatal head is subsequently registered within this ROI. This process allows virtual guiding hands to be overlaid on the manikin with anatomical accuracy, enabling trainees to follow expert trajectories reinforced by haptic feedback. A large-scale user study was conducted with 83 fourth-year medical students, whose performance was evaluated by four senior obstetricians. The results demonstrated that MR training led to greater engagement and significantly higher scores in delivery, post-delivery, and overall handling tasks compared to VR training.

The main contributions of our paper are outlined as follows:

- We propose the first MR-based childbirth training system that integrates virtual guidance with tactile manikin interaction, allowing trainees to practice independently without on-site expert supervision.
- We present the first integration of an external RGB-D camera with a Quest headset in passthrough mode, enabling real-time communication and stable MR interaction.
- We develop a coarse-to-fine localization strategy combining multi-marker alignment of the maternal manikin with point cloud registration of neonatal head to compute accurate virtual guiding hand positions.
- We conducted a large-scale MR user study for childbirth training, which to our knowledge provides initial empirical insights into how MR-based training differs from VR-based approaches in this domain.

## 2 RELATED RESEARCH

### 2.1 Limitations of Conventional Childbirth Training

Childbirth training for medical and midwifery students relies heavily on labour ward placements, but opportunities to observe or assist in deliveries are often limited by unpredictable clinical schedules and the availability of consenting patients [24, 25]. The spontaneous and time-sensitive nature of labour further restricts access to meaningful learning experiences [26, 27].

To compensate for limited clinical exposure, conventional childbirth education often relies on slideshows, instructional videos, and manikin-based simulations. Slideshow-based teaching remains common in preclinical settings due to its simplicity, but it does not adequately prepare students for the dynamic and tactile nature of childbirth [3]. Instructional videos help familiarize students with the delivery environment and improve engagement [28], yet they are passive and do not support skill development through active practice [8]. Physical manikins are widely used to simulate childbirth and obstetric emergencies [7], offering valuable procedural practice. However, they depend on the presence of experienced instructors to guide learners and provide timely feedback, which limits scalability and consistency across large student groups [9]. These limitations highlight the need for more interactive and scalable training systems that provide realistic procedural experience alongside individualized support.

### 2.2 VR and MR Applications in Childbirth Training

To address the limitations of conventional methods, immersive technologies such as VR and MR have been increasingly adopted to enhance procedural training and learner engagement in childbirth training. Chang et al. [29] developed the IFOREAL simulator with 3D anatomical animations and delivery tools, improving perceived usefulness but hindered by unfamiliar controls. To further enhance intuitiveness, Nugraha et al. integrated Leap Motion-based hand tracking for natural hand interactions [30]. More recently, Liu et al. developed a comprehensive VR childbirth simulator with guided hand manoeuvres and conducted a large-scale empirical study with medical students [14]. While the system significantly improved learning outcomes compared to manikin training, students reported lower ratings in usability and feedback, reflecting VR's ongoing limitations in tactile realism and instructor presence.

To bridge the gap between virtual guidance and physical interaction, several MR systems have been explored for childbirth training. Mazalan et al. [31] proposed a marker-based MR application using Vuforia to display 3D childbirth animations on smartphones with interactive quizzes, but its limited immersion and lack of tactile realism positioned it more as a reference tool than a simulator. Building on interactivity, Ballit et al. [32] developed an MR simulation on HoloLens 2 that applied a hyperelastic mass-spring model to reproduce fetal and pelvic tissue deformation, enabling pseudo-haptic practice of forceps delivery. Yet, in the absence of user studies or clinical context, its educational efficacy remains uncertain. Complementarily, Ljungblad et al. [20] employed MagicLeap 2 for MR anatomical visualization, allowing learners to inspect a virtual fetus within the maternal pelvis. While well received for conceptual understanding, it provided little support for procedural training or psychomotor skill acquisition. Collectively, these efforts underscore that existing MR systems tend to emphasize either accessibility or visualization, but fall short of delivering immersive, validated, hands-on childbirth training.

### 2.3 Vision-based Manikin Tracking Techniques in MR

Accurate tracking of physical manikins is essential for immersive interaction in MR-based childbirth simulations. Marker-based tracking is a mature and widely adopted technique in medical MR systems due to its high spatial accuracy, low computational overhead, and robustness on mobile HMDs. One common approach employs planar fiducial markers to register patient anatomy or instruments with virtual overlays to assist with navigation tasks such as ventriculostomy and neuronavigation [33, 34]. To enhance robustness and minimize drift, several studies have introduced multi-marker fusion or fixed marker rigs, which improve tracking consistency under head or patient movement [35, 36]. Some systems further integrate marker detection with the optical pipeline of HoloLens or other wearable devices, enabling direct marker recognition without external cameras [37–39]. Building on these insights, effective deployment requires careful planning of marker placement to ensure visibility and stability during interaction, especially in constrained clinical or training environments.

To reduce reliance on physical markers, deep learning-based 6-DoF object pose estimation has gained traction in recent years. Methods

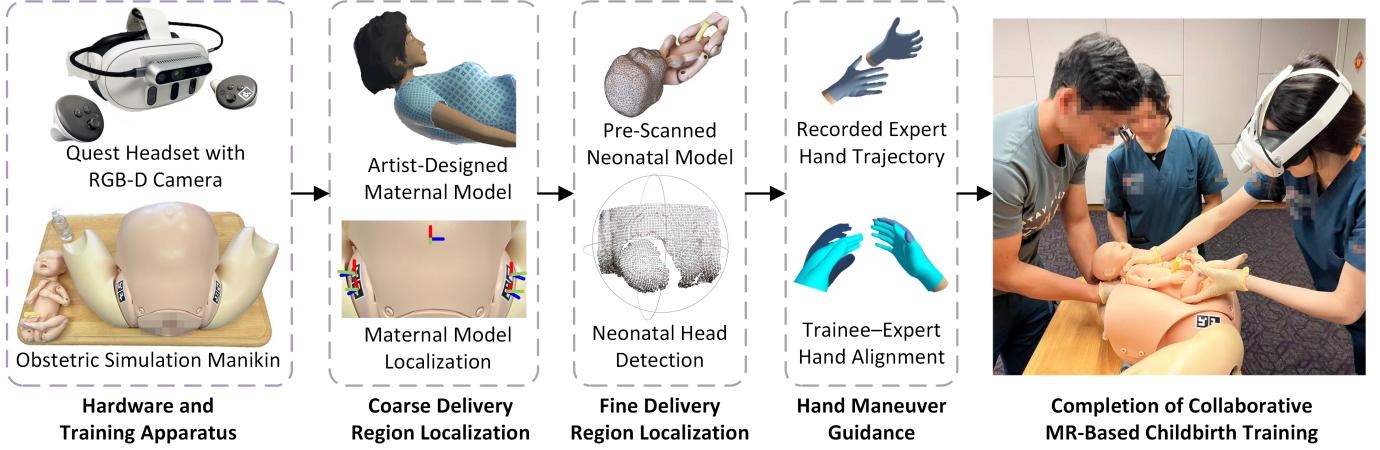


Fig. 2: Workflow of our MR visualization framework for immersive childbirth training, employing a coarse-to-fine delivery region localization strategy that sequentially estimates maternal and neonatal poses to ensure accurate hand maneuver guidance for collaborative training.

such as FoundationPose [40], MegaPose [41], NeRF-Pose [42], Any-6D [43] and OSOP [44] infer object poses directly from RGB or RGB-D inputs. These models are primarily designed for rigid objects, which makes them difficult to adapt to deformable training manikins used in childbirth education. To address this limitation, human pose estimation techniques have been explored for tracking flexible manikins [45, 46]. Some studies further extend these capabilities to specialized scenarios, such as infant pose estimation [47–49]. However, they remain highly sensitive to occlusions, particularly when the trainee’s hands or instruments cover parts of the manikin. In addition, their dependence on high-performance GPU inference makes real-time deployment on mobile HMDs infeasible without incurring significant latency or requiring offloading to external servers.

### 3 METHODOLOGY

#### 3.1 Design of Our MR Childbirth Training System

Our work aims to integrate the benefits of virtual feedback and physical manikin training to develop an interactive MR childbirth training system. Even in the absence of professional obstetricians, our MR system can assist medical students and junior doctors in familiarizing themselves with and performing the correct procedures during childbirth, as shown in Fig. 2.

Compared to the HoloLens headset commonly used in MR tasks [22, 37, 50], the Quest headset demonstrates superior performance in key aspects such as screen field of view, resolution, refresh rate and processor performance [51]. Furthermore, the Quest headset supports seamless switching between VR and MR modes. This flexibility allows direct comparisons between VR and MR experiences in our user study under identical hardware conditions, which is not feasible with the HoloLens headset due to its lack of VR support. Although recent firmware updates provide limited access to the Quest’s camera, the interface suffers from low transmission speed and the absence of spatially aligned RGB-D streams, making it unsuitable for real-time MR interaction. To address this limitation, our system first establishes a customized data access pipeline and integrates an external RGB-D camera onto the Quest headset, enabling reliable, spatially aligned real-world scene capture in real time to track the 3D pose of the maternal and neonatal manikins.

A central requirement for effective normal delivery training is that trainees learn the correct sequence of hand maneuvers. To support this goal, our system must first localize the delivery region of the physical obstetric manikin with sufficient accuracy to guide procedural interaction. We adopt a coarse-to-fine localization strategy: the maternal manikin is first used to define a coarse delivery region, and the neonatal manikin’s head position is then employed to refine this region to a clinically meaningful scale. To meet physicians’ requirements for large-scale, user-friendly deployment, the entire childbirth training system

operates on a standalone headset without the need for external servers. For maternal manikin tracking, we employ a fiducial marker-based approach, which provides efficient and reliable registration even when no exact virtual replica of the maternal model is available. In contrast, the neonatal manikin is composed entirely of silicone and undergoes significant deformation during simulated delivery, making marker attachment infeasible. To address this challenge, we implement a point cloud registration method on the neonatal head, which refines the delivery region dynamically. Once the delivery region is established, the system overlays virtual guiding hands in the corresponding spatial location. Trainees are instructed to follow the guiding trajectory and practise the correct hand maneuvers on the physical manikin, thereby combining tactile realism with real-time procedural guidance in MR.

#### 3.2 Multi-Coordinate System Integration

Accurate spatial registration is fundamental to enabling realistic interaction in MR-based childbirth training. Since the Quest headset, the mounted RGB-D camera, and the obstetric manikin each operate in separate local coordinate systems, a consistent unification is required to ensure that virtual overlays align precisely with the physical manikin. All components are integrated into the headset coordinate system, which serves as the global reference for both virtual rendering and physical interaction.

To localize the maternal manikin in the headset frame, we first adopt the headset coordinate frame  $H$  as the global reference, since rendering and user input are all anchored to the HMD, expressing all poses in  $H$  avoids downstream frame conversions. We then introduce two time-invariant links to reduce the estimation degrees of freedom: the RGB-D camera  $C$  is rigidly mounted on the headset, yielding a constant transform  $\mathbf{T}_C^H$ , and fiducial markers  $F$  are rigidly attached to the maternal manikin  $M$  at a known location, yielding a constant  $\mathbf{T}_M^F$ . At runtime, the reliably observable quantity is the marker pose in the camera frame,  $\mathbf{T}_F^C(t)$ , obtained via fiducial detection. By composing transforms along the kinematic chain from  $M$  to  $H$ , the manikin pose in the headset frame follows as:

$$\mathbf{T}_M^H(t) = \mathbf{T}_C^H \cdot \mathbf{T}_F^C(t) \cdot \mathbf{T}_M^F. \quad (1)$$

Consequently, once  $\mathbf{T}_F^C(t)$  is estimated,  $\mathbf{T}_M^H(t)$  is uniquely determined by Eq. 1. Fusion process across multiple visible markers and the calibration of  $\mathbf{T}_C^H$  are detailed in Secs. 3.4.1 and 3.3, respectively.

#### 3.3 Spatial Registration of Headset and RGB-D Camera

Since the Quest headset and the RGB-D camera occupy different spatial positions and operate in distinct coordinate frames, camera calibration is required to ensure precise virtual-to-physical alignment in MR display. Conventional calibration methods typically rely on capturing

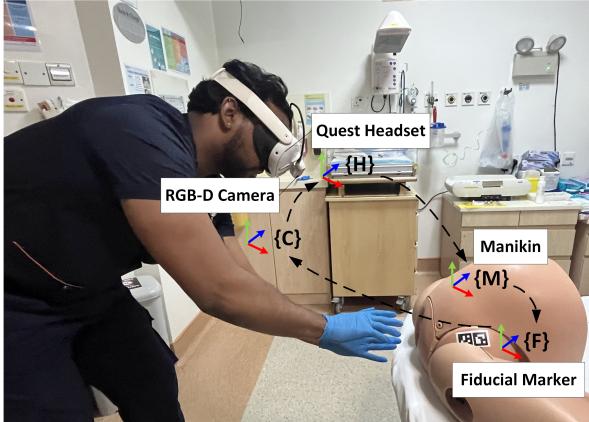


Fig. 3: System setup and transformation of world coordinate systems between Quest headset and manikin.

fiducial markers simultaneously across cameras and solving for their relative transformation [52]. However, the Quest interface does not provide access to reliable RGB-D streams or intrinsic parameters of its onboard cameras, which makes direct cross-camera calibration with the external RGB-D device infeasible. Consequently, we estimate the relative transformation between the headset and the RGB-D camera through a custom calibration procedure described below.

Since the Quest headset provides access to the controller's pose through its built-in development modules. Meanwhile, the RGB-D camera can capture images containing fiducial markers. Building on this setup, we pioneer the application of dynamic pose tracking using the Quest controller and fiducial markers to extend the eye-to-hand calibration algorithm [53]. The RGB-D camera and fiducial marker should be rigidly attached to the Quest headset and Quest controller respectively, to maintain the fixed relative pose. According to the rigid body kinematics theory [53], the coordinate system transformation within the setup between two adjacent frames  $t_i$  and  $t_j$  can be described as:

$$\mathbf{T}_H^U(t_i, t_j) \cdot \mathbf{T}_C^H = \mathbf{T}_C^H \cdot \mathbf{T}_F^C(t_i, t_j), \quad (2)$$

where  $\mathbf{T}_H^U(t_i, t_j) = \mathbf{T}_H^U(t_j) \cdot (\mathbf{T}_H^U(t_i))^{-1}$  and  $\mathbf{T}_F^C(t_i, t_j) = \mathbf{T}_F^C(t_j) \cdot (\mathbf{T}_F^C(t_i))^{-1}$ . Specifically,  $\mathbf{T}_H^U(t_i, t_j) \in \mathbb{R}^{4 \times 4}$  is the relative transformation of the user controller  $U$  with respect to headset  $H$ , and  $\mathbf{T}_M^R(t_i, t_j) \in \mathbb{R}^{4 \times 4}$  is the relative transformation of RGB-D camera  $C$  with respect to the fiducial marker  $F$ . The relative pose between  $H$  and  $C$  can be subsequently calculated as:

$$\mathbf{T}_C^H = \underset{\mathbf{T}_C^H}{\operatorname{argmin}} \sum_{i,j} \|\mathbf{T}_H^U(t_i, t_j) \cdot \mathbf{T}_C^H - \mathbf{T}_C^H \cdot \mathbf{T}_F^C(t_i, t_j)\|. \quad (3)$$

This constrained nonlinear homogeneous equation can be solved by Tsai-Lenz algorithm [54]. In practical applications, the fiducial marker can be affixed to the Quest controller at any position to ensure a rigid connection. After completing the precomputed camera calibration, the real-world scene information captured by the RGB-D camera can be accurately aligned and rendered in the passthrough mode of Quest headset, ensuring correct spatial positioning and seamless integration.

### 3.4 Design of Obstetric Manikin Tracking

To enable procedural guidance and assessment on a standalone headset, our tracker delivers stable, low-latency manikin poses within a coarse-to-fine pipeline that anchors the delivery region from the maternal reference and refines it using neonatal head pose for precise virtual guidance.

#### 3.4.1 Fiducial Marker-based Maternal Manikin Tracking

The fiducial marker offers an optimal trade-off between tool cost, implementation feasibility and tracking stability. However, fiducial markers

on the curved manikin surface often produce temporally unstable 3D pose estimations with RGB-only detection methods [55]. To address this, we optimized the detection process by adding depth information. Specifically, the RGB image is used to detect the 2D positions of the fiducial marker corners, while the camera's intrinsic parameters for both RGB and depth sensors are utilized to calculate the 3D positions of the marker corners. These 3D positions are then integrated to derive more robust and accurate parameters of the fiducial marker.

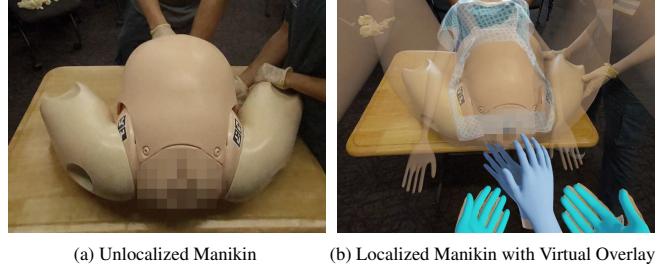


Fig. 4: First-person headset view of manikin localization. The light-blue hands indicate the tracked hands of the trainee, while the dark-blue hands represent the expert's virtual guiding hands.

To robustly achieve accurate maternal model alignment in MR, multiple fiducial markers are strategically placed on the maternal manikin's surface (Fig. 4) to expand viewpoint coverage and reduce the impact of single-marker occlusion. After defining the marker coordinate system, the relative poses of all markers with respect to the virtual maternal model's center are precomputed during initialization, thereby establishing a fixed geometric prior and avoiding per-session recalibration. The fixed spatial relationship between the  $i$ -th marker  $F_i$  and the model is denoted by  $\mathbf{T}_{F_i}^M$  so that each detected marker yields a consistent candidate estimate of the model pose.

During runtime, the headset-mounted RGB-D camera continuously detects the live poses of visible markers. Given each marker pose  $\mathbf{T}_{F_i}$  in the RGB-D camera coordinate system, the associated local model pose  $\mathbf{T}_M^{F_i}$  is obtained by chaining the measurement with the precomputed marker-to-model transform. When multiple markers are visible during training, their per-marker estimates are fused to determine the optimal model pose  $\mathbf{T}_M$ , leveraging redundancy to mitigate measurement noise and transient occlusions. A distance-weighted scheme is adopted, with  $w^i$  inversely proportional to the squared camera–marker distance, so that closer observations contribute more to the fused estimate:

$$\mathbf{T}_M^F = \mathcal{F} \left( \sum_{i=1}^N \tilde{w}^i \cdot \mathbf{T}_M^{F_i} \right), \quad \tilde{w}^i = \frac{w^i}{\sum_{j=1}^N w^j}, \quad (4)$$

where  $\mathcal{F}(\cdot)$  denotes the one Euro filter [56], which suppresses high-frequency jitter while preserving responsiveness required for interactive guidance. After localizing the physical maternal manikin, the preset local pose of the virtual guiding hands  $\mathbf{G}$  relative to the virtual maternal model and the calibrated transform  $\mathbf{T}_C^H$  in Eq. (3) are incorporated to express guidance in the headset frame:

$$\mathbf{T}_{\text{hand}}^H = \mathbf{G} \cdot \mathbf{T}_C^H \cdot \mathbf{T}_F^C \cdot \mathbf{T}_M^F. \quad (5)$$

In our MR system, the 3D model of the maternal manikin is designed by digital artists and does not perfectly replicate the shape or dimensions of the physical manikin. As a result, the estimated hand positions derived from Eq. (5) can only serve as coarse approximations. Therefore, upon determining the initial coarse hand pose  $\mathbf{T}_{\text{hand}}^H \in \mathbb{R}^{4 \times 4}$ , a delivery ROI is instantiated based on the physical maternal manikin's dimensions, thereby constraining subsequent neonatal head localization and reducing computational complexity.

#### 3.4.2 Markerless Neonatal Manikin Detection

To refine and update the virtual guiding hand placement based on the actual position of the neonatal manikin, it is necessary to perform a

secondary localization step once the neonate begins to emerge from the vaginal opening. Since normal delivery tasks do not require precise or continuous tracking of the neonatal pose, this localization only needs to be performed once, and can be accomplished without relying on computationally intensive data-driven methods or additional external computing devices.

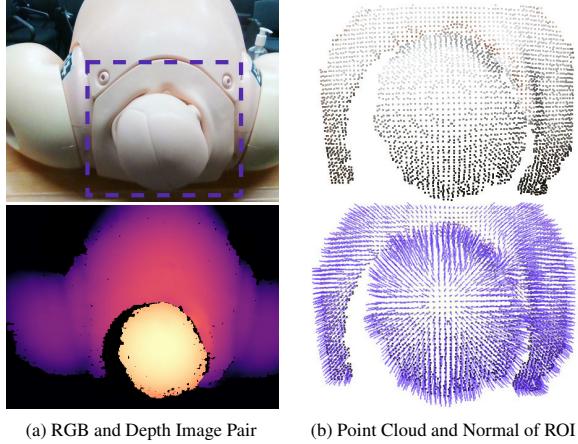


Fig. 5: Point cloud generation of the neonatal head ROI. The purple dashed box in the RGB image indicates a coarse search region near the delivery area for localizing the neonatal head.

Although fiducial markers are widely used for tracking physical objects because of their simplicity and reliability, they are not well suited to a neonatal manikin in childbirth training. The manikin’s silicone surface lacks sufficiently planar regions, which prevents secure adhesion of markers. During simulated delivery, the head and body also experience substantial compression and shear that further destabilize any attachment. In addition, the lubricant applied to facilitate delivery produces specular reflections that interfere with detection and gradually degrades the printed markers. Taken together, these factors make marker-based tracking unreliable for the neonatal manikin.

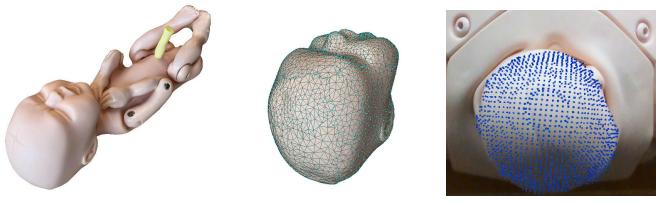


Fig. 6: Virtual neonatal head model acquisition and point cloud registration for localization in an MR environment.

To robustly detect the initial emergence of the neonatal manikin, we estimate the head pose using a fast two-stage point-cloud registration pipeline. A high-resolution template point cloud of the neonatal head  $\mathbb{P}$  is preacquired with a structured-light scanner and registered offline to the MR system’s global frame. At runtime, the current coarse hand pose  $\mathbf{T}_{\text{hand}}^H$  (Eq. 5) defines a delivery ROI near the outlet (Fig. 5), which bounds the search to the clinically relevant area, reduces computation, and suppresses spurious correspondences from background geometry. Within this ROI, we adopt Fast Global Registration (FGR) [57] for coarse initialization because the neonatal head frequently enters with unknown orientation and partial visibility. FGR estimates a global rigid transform directly from feature correspondences without an initial guess, satisfying our time-constrained, no-prior requirement and furnishing a suitable initializer for local refinement. The coarse alignment is then refined with point-to-plane ICP [58], which is well matched to the relatively rigid head geometry in the localized ROI, offering fast convergence and improved metric accuracy under small residual misalignments, thereby meeting the latency budget on a standalone

headset. The resulting neonatal head pose  $\mathbf{T}_b$  (Fig. 6) is used to update the relative trainee-hand transform  $\hat{\mathbf{G}}$  in conjunction with the current maternal model pose  $\mathbf{T}_M^F$  (Eq. 4), re-anchoring virtual guidance to the observed contact geometry. The updated hand pose in the headset coordinate system is then computed by the same transform chaining as in Sec. 3.4:

$$\hat{\mathbf{T}}_{\text{hand}}^H = \hat{\mathbf{G}} \cdot \mathbf{T}_C^H \cdot \mathbf{T}_F^C \cdot \mathbf{T}_M^F. \quad (6)$$

### 3.5 MR-Guided Hand Manoeuvres for Normal Delivery

To enable effective training of normal delivery manoeuvres, we first recorded the hand trajectories of an experienced obstetrician interacting with a physical neonatal manikin using high-precision motion sensors. These trajectories were treated as expert demonstrations and subsequently used to generate animated virtual hand manoeuvres. During the MR simulation, the virtual expert hand model is rendered near the head of the physical neonatal manikin, whose spatial location is continuously tracked. A user interface component provides real-time visual prompts to guide the trainee’s hands to spatially align with the expert hand posture, thereby facilitating skill acquisition through direct motion imitation.

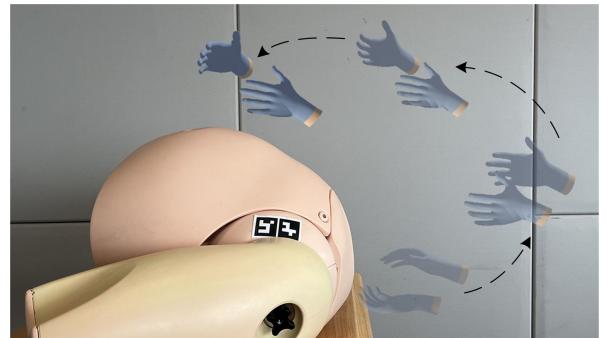


Fig. 7: Illustration of expert hand trajectory across the sequential manoeuvres of normal delivery under MR guidance.

To account for practical variability in initial neonatal manikin poses, differing spatial arrangements between the trainee and the manikin, and other unpredictable training conditions, we designed a robust interaction mechanism that balances hand tracking precision and temporal smoothness. This design ensures consistent guidance quality and trainee comfort across diverse scenarios. A proximity-based trigger mechanism is implemented by attaching collision detectors to both the trainee’s wrist and the virtual expert hand. When the Euclidean distance between the two falls below 5 cm, the expert hand animation is activated. Multiple checkpoints are embedded along the animated trajectory to continuously monitor alignment accuracy. If the deviation between the trainee’s hand and the virtual trajectory exceeds a predefined threshold at any checkpoint, the animation is paused and a corrective prompt is issued, instructing the trainee to re-align with the expert hand before resuming.

This real-time hand guidance framework enables context-aware correction and promotes fine-grained motor skill acquisition, even under uncertain training conditions. By integrating adaptive animation control with spatial proximity sensing, the system provides an intelligent feedback loop that enhances both the fidelity and usability of MR-based procedural training.

## 4 EVALUATIONS

We evaluate our MR framework through both system-level registration quality analysis and trainee-level training assessment. The former examines calibration accuracy and model alignment, while the latter investigates usability and training outcomes.

## 4.1 MR Registration Quality Assessment

### 4.1.1 Assessment of Camera Calibration Accuracy

To evaluate the alignment accuracy between the RGB-D camera and Quest coordinate systems, a fiducial marker was rigidly attached to the center of the Quest controller's local coordinate system. The controller was then moved along arbitrary trajectories within the region visible to both the RGB-D camera and Quest headset. During this process, the raw movement trajectory of the fiducial marker is recorded by the RGB-D camera (green points in Fig. 8) and the movement trajectory of the same point is captured by the Quest headset (blue points in Fig. 8).

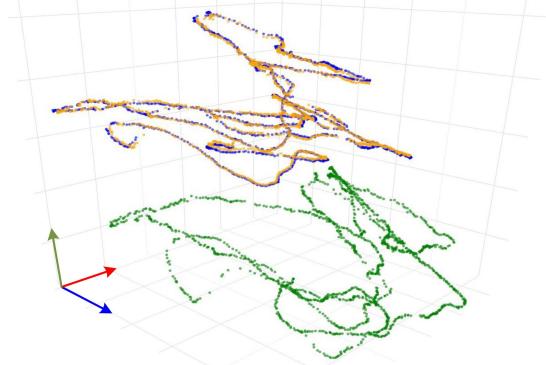


Fig. 8: Example assessment of camera calibration accuracy. **Green points** denote the original RGB-D camera trajectory, **yellow points** denote the calibrated RGB-D camera trajectory and **blue points** denote the Quest headset trajectory.

Using the relative pose transformation  $\mathbf{T}_C^H$  in Eq. (3), the raw RGB-D trajectory was transformed into the corrected movement trajectory (yellow points in Fig. 8). After 10 independent trials, the RMSE between the corrected RGB-D trajectory and the Quest trajectory was calculated to be  $3.6 \pm 0.2\text{mm}$ . This calibration accuracy satisfies our requirements of childbirth training tasks and verifies the feasibility of hardwired connection between the Quest headset and RGB-D camera for MR task development.

### 4.1.2 Assessment of Manikin Alignment

To evaluate manikin tracking performance, we conducted a controlled experiment with the setups summarized in Tab. 1. During childbirth training, trainees primarily focus on the birthing region, and natural head movements can intermittently move fiducial markers out of the camera's view. To address this and the limited surface area for marker placement, the manikin was instrumented with either two markers (one per side) or four markers (two per side). Given the manikin's curved surface, depth data was also incorporated to examine whether it improves detection accuracy and stability. In practice, markers only need to be placed approximately symmetrically on the manikin surface to ensure visibility, without strict constraints on exact positions. The system does not assume a specific maternal manikin model either. When switching to a new manikin, the relative pose between the maternal manikin and the attached markers is computed once following the procedure described in Sec. 3.4.1. This calibration step is performed offline prior to training and does not involve trainee interaction.

In this experiment, only the maternal manikin alignment was assessed, as the neonatal head remains spatially constrained to the maternal structure. The delivery region estimated from the maternal alignment directly defines the subsequent neonatal head registration, which is performed only once per trial. Although the virtual expert hands are initially adjusted based on the neonatal head pose, their relative transformation to the maternal manikin remains fixed thereafter. Thus, accurate maternal alignment is the principal determinant of both the spatial fidelity of the delivery region and the reliability of neonatal head registration.

This experiment involves three key metrics, including Absolute Pose Error (APE), Jitter Frames Percentage (JFP) and Frames Per Second

(FPS). APE evaluates global tracking accuracy by measuring errors in manikin translation and rotation. To obtain reliable ground truth for the APE metric, we rigidly attached a Quest Controller to the maternal manikin and used its recorded location and rotation as the reference pose. JFP assesses tracking stability through two sub-metrics. The first sub-metric, i.e., marker loss rate, accounts for the percentage of frames with undetected fiducial markers caused by occlusion or markers moving out of the camera's field of view. The second sub-metric, i.e., pose jittering rate, captures temporal inconsistencies in consecutive pose estimates caused by detection instability or noise. The pose jittering in our experiment is defined as a deviation exceeding  $5\text{mm}$  or  $5^\circ$  between consecutive frames. FPS measures the tracking efficiency of maternal manikin on the Quest 3 headset's built-in processor.

The experimental results in Tab. 1 show that increasing the number of markers leads to a reduction in APE, but only marginally in RGB-only setups. For instance, the shift from 2 to 4 markers in the RGB configuration results in a 4.4% reduction in positional error and a 7.3% reduction in rotational error. However, the inclusion of depth data substantially improves the tracking performance. Specifically, the 4-marker RGB-D configuration results in an 86% reduction in positional error and an 80% reduction in rotational error compared to the 4-marker RGB configuration, with a noticeable reduction in the standard deviation. This demonstrates that depth data is crucial in improving tracking accuracy, especially under real-world conditions where the manikin's curved surface could otherwise hinder marker detection.

Table 1: Controlled experiment for tracking performance evaluation of the maternal manikin. ( $\uparrow$  indicates higher values are better and  $\downarrow$  indicates lower values are better.)

Setup	APE (mm/ $^\circ$ ) $\downarrow$	JFP (%/%) $\downarrow$	FPS $\uparrow$
2 Markers + RGB	$25.34 \pm 5.05$	$14.27 \pm 2.82$	<b><math>115.68 \pm 5.19</math></b>
	$19.25 \pm 0.85$	$22.31 \pm 3.08$	
2 Markers + RGB-D	$3.89 \pm 0.36$	$14.13 \pm 2.90$	$108.76 \pm 4.22$
	$3.98 \pm 0.32$	$6.98 \pm 1.17$	
4 Markers + RGB	$24.22 \pm 4.70$	$4.75 \pm 1.58$	$110.29 \pm 2.75$
	$17.84 \pm 0.78$	$8.37 \pm 2.67$	
4 Markers + RGB-D	<b><math>3.45 \pm 0.26</math></b>	<b><math>4.71 \pm 0.64</math></b>	$106.21 \pm 2.23$
	$3.65 \pm 0.28$	$2.96 \pm 0.53$	

For JFP, both depth data and marker count help improve tracking stability, but their effects differ between the two sub-metrics. The marker loss rate is primarily influenced by the number of markers, with 4-marker setups consistently reducing the loss rate by approximately 67% compared to 2-marker setups. In contrast, the pose jittering rate benefits more significantly from the inclusion of depth data, with reductions of nearly 69% when depth information is added to a 2-marker setup, and over 65% in a 4-marker setup. These results highlight the substantial improvement in detection stability achieved through depth data, particularly on curved surfaces where RGB-only setups struggle to maintain consistent pose estimates. Finally, while the inclusion of depth data and additional markers slightly reduces FPS, the frame rate of 106 FPS in the 4-marker RGB-D setup remains comfortably within the real-time performance threshold, ensuring smooth and uninterrupted tracking during childbirth training scenarios.

## 4.2 User Study Design

A large-scale user study was conducted in collaboration with the School of Medicine to evaluate immersive technologies for childbirth training by comparing MR and VR. The study protocol was reviewed and approved by the Institutional Review Board (IRB) of our university. The study focused on the normal delivery task and examined how physical interaction in MR may influence training outcomes relative to a fully virtual environment.

### 4.2.1 Pipeline of User Study

The study was conducted during the first week of the Year 4 undergraduate medical students' six-week O&G clerkship and comprises two sessions: a *learning session* on Monday and an *assessment session*



Fig. 9: Experimental setup of the user study. (a) Learning stage with VR-based childbirth training. (b) Learning stage with MR-based childbirth training integrating headset guidance with manikin interaction. (c) Assessment stage where both VR and MR groups performed the delivery task on a manikin without headset assistance, and were recorded by the video camera for the DOPS assessment.

on Friday. 14 clinical groups, each consisting of six students, were recruited and cluster-randomised prior to the study.

**Learning session.** The learning session started off with participants watching a pre-recorded briefing by a medical expert in a meeting room. They were then distributed with a sticker of their assigned unique study record number (SRN) on, scanned the QR code and filled up their demographic information on their phone. Upon completion, they proceeded to a separate room for VR or MR simulations with on-site student helpers facilitating the session. The VR simulation (Fig. 9(a)) took on average 11 minutes, whereas the MR simulation (Fig. 9(b)) took slightly longer, averaging around 15 minutes, due to additional physical interaction and coordination with supporting trainees. Participants belonging to the MR group were additionally required to complete a preference questionnaire after finishing the MR simulation.

**Assessment Session.** On the day of the assessment session, participants gathered in a seminar room with 4 sets of Laerdal Prompt Flex manikins. All participants received surgical masks for anonymity and wrist tags indicating the SRN assigned to them. Following that, a medical expert demonstrated to them how to manipulate the manikin, focusing on inserting the neonatal manikin in and pushing it through the maternal manikin without hinting on the hand manoeuvre. After the demonstration, participants were grouped into teams of three, rotating through roles: expulsion of the neonate, stabilising the maternal manikin and neonate reception and handling. Participants then had five minutes to get familiar using the manikin. Four video cameras operated by study facilitators were set up to record the performance from fixed tripod positions focused on the manikin (Fig. 9(c)). Each participant was allowed two attempts to perform the hand manoeuvre assisting neonatal delivery. Before starting, they displayed the SRN on the wrist tag to the camera and then informed the facilitator by saying “I am ready”. Upon completing the procedure, participants indicated their completion by saying “I have completed”. After all recordings were captured, the session concluded, marking the end of the study.

#### 4.2.2 Assessment Tool

Two complementary tools were used: a *Preference Questionnaire* for subjective perceptions of the MR system, and the *Direct Observation of Procedural Skills* (DOPS) for expert evaluation of training performance in the assessment session.

**Preference Questionnaire.** The Preference Questionnaire was crafted based on Witmer and Singer’s Presence Questionnaire and the System Usability Scale. This concise questionnaire, designed to assess participants’ immediate perceptions of the MR system, contains five 5-point Likert-scale questions, with 1 indicating ‘Not At All’ and 5 indicating ‘Completely’. The five rating questions were thoughtfully chosen to obtain comprehensive aspects of the MR simulation experience, and their key ideas can be extracted as listed in Table 2. As the questionnaire focuses on MR-specific interaction characteristics, it was

administered only in the MR condition and not intended for a direct VR–MR comparison.

**Direct Observation of Procedural Skills.** DOPS was adapted as the primary assessment tool for the user study. It is a validated method in medical education used for evaluating medical students’ readiness, comprehension of procedures, and accuracy in technical skills. The technique grading rubric designed by the medical expert in the study team comprises four major categories, including hand placement prior to delivery, hand manoeuvre during delivery, handling of neonatal post-delivery, and overall carefulness during the procedure. (The detailed rubric is shown in the supplementary material.) Each sub-category is scored on a 3-point scale, with 2 points for correct independent execution, 1 point for inadequate performance, and 0 points for omission.

Table 2: Preference questionnaire used for MR evaluation. Each item was rated on a 5-point Likert scale.

Dimension	Description
Engagement	How involved were you in the MR (passthrough) experience?
Presence	How well could you concentrate on the assigned tasks rather than the mechanisms used to perform them?
Ease of Use	How well could you move or manipulate objects in the MR environment?
Preference	I prefer the MR (passthrough) environment compared to the VR environment.
Interest in Reuse	I think I would like to practise using this simulation frequently.

Table 3: Summary of assessment categories and sub-categories adapted from DOPS.

Category	Sub-category (3-point Scale)
Hand Placement	Q1: Hand placement readiness Q2: Guarding of the perineum
Delivery	Q3: Identification of extension Q4: Restitution with delivery of the anterior shoulder
Post-Delivery	Q5: Confident handling of the neonate Q6: Placing the neonate on the abdomen of the maternal model
Overall	Q7: Overall handling and respect for tissue (any damage caused)

Four professional assessors from the O&G Department, not affiliated with the study team, individually graded the video recordings

of the performances. Prior to distribution for grading, videos were screened by the study team to exclude cases where the onsite instructor unintended intervention occurred or where camera obstruction blocked the visibility of key actions. To minimise grading bias potentially introduced by assessor fatigue or reduced attention span over time, all video recordings were randomly sequenced, ensuring a balanced mix of participants from different groups and assessment dates. Final scores were computed as the average among the ratings from four assessors.

#### 4.2.3 Demographics

83 Year 4 undergraduate medical students from the Yong Loo Lin School of Medicine, National University of Singapore were recruited for this study. After accounting for absences, incomplete simulations and disqualified video recordings, 72 validated video recordings were included for grading, comprising 37 in the VR group and 35 in the MR group. The final cohort had an average age of 22.43 (SD = 0.90), with 52.78% female participants and 47.22% male participants. 98.61% of the participants reported no prior experience with normal childbirth, with only one in the cohort having witnessed it once. Regarding VR usage, 72.22% of the participants had used VR once a year, 26.39% had never used VR, and only one reported relatively frequent usage of once a month.

#### 4.2.4 Results

**Preference Questionnaire Statistics.** 35 participants from the MR intervention group completed the Preference Questionnaire immediately after their MR simulation. The results of all five questions were not normally distributed based on the Shapiro-Wilk test. The *Engagement* item received the highest mean score among the five items (Mean = 4.23, SD = 0.91). More than 80% of the participants felt they were highly involved during the MR experience, highlighting MR's capability in capturing learners' attention and sustaining active participation. Moderately correlated with the *Engagement* item (Spearman's  $\rho = 0.579$ ), the *Presence* item received the second highest rating (Mean = 3.74, SD = 1.01). Despite potential distractions from their peers and the surroundings in the passthrough mixed reality mode, the participants still managed to concentrate on the given tasks and followed the hand manoeuvre throughout. The *Ease of Use* item received a slightly lower score (Mean = 3.66, SD = 0.80) than *Engagement* and *Presence*. However, there was no major usability issue as none of the participants rated 1 and only two rated 2. 94.3% of the participants did not encounter difficulty during their MR experience.

Moreover, as the participants experienced both the fully virtual environment and the passthrough MR environment, the *Preference* item was designed to determine whether they prefer training in a more isolated setting or a more interactive setting. The results showed that 5 participants preferred the former and 22 participants enjoyed the latter, while 8 remained neutral (Mean = 3.74, SD = 1.17). More vigorous discussions were observed in the MR group, driven by the inherently collaborative effort of manipulating the manikin. Lastly, the *Interest in Reuse* item indicated that 20 participants (57.1%) reported a strong willingness to use the designed MR system for future training (Mean = 3.51, SD = 1.17), reflecting its potential as a viable tool for skill acquisition and pointing to the need for further refinements to better align with medical students' learning habits and practice workflows.

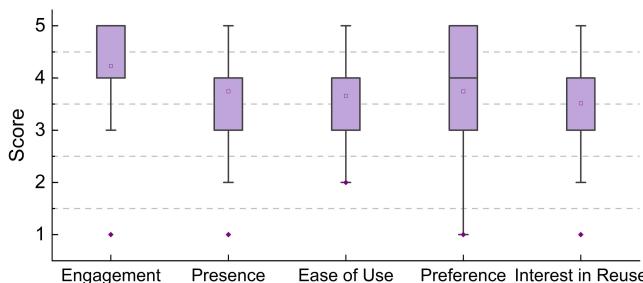


Fig. 10: Statistical analysis of the preference questionnaire for MR evaluation.

**DOPS Statistics.** Assumption of the same baseline performance was made since all participants were in their first week of their Year 4 O&G clerkship and none had experience assisting an actual childbirth. The statistical distributions of the sub-category scores in Table 3 are visualized in Fig. 11. To enable group-level comparison, the scores of corresponding sub-categories were aggregated into composite scores. Since the Shapiro-Wilk test indicated that some of the aggregated scores were not normally distributed, a non-parametric Mann-Whitney test was used to evaluate pairwise differences between the VR and MR groups during the practical assessment. As shown in Fig. 11 and Table 4, *Hand Placement* yielded comparable score distributions between VR and MR. This finding is consistent with expert interpretations that pre-delivery guidance did not differ substantially, as the physical manikin in MR offered minimal additional tactile input. In contrast, significant group differences were identified in *Delivery*, *Post-Delivery* and *Overall Handling*. The consistently higher medians observed in MR, particularly in *Post-Delivery*, suggest enhanced competence in critical tasks such as holding and positioning the newborn. Moreover, the narrower IQR in *Overall Handling* reflects stronger inter-rater consensus, underscoring the reliability of expert evaluations and the greater proficiency conferred by MR training.

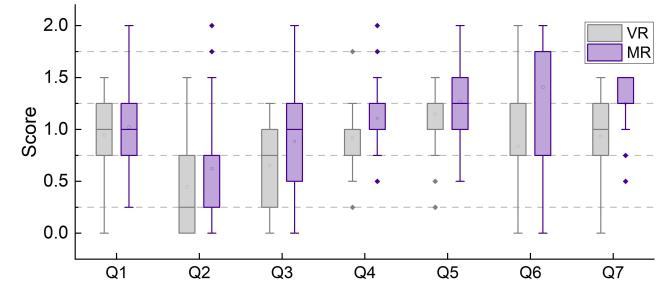


Fig. 11: Statistical analysis of the DOPS across the categories in Table 3 comparing VR and MR conditions.

Furthermore, the total scores across all categories in Table 4 were calculated to determine the overall extent to which the MR group outperformed the VR group. While the median of the total score of the VR group was 5.5 (IQR = 3), the MR group achieved a median of 7.5 (IQR = 2). The result indicated a statistically significant difference ( $p = 0.003$ ,  $U = 387$ ), demonstrating the MR group's clear advantage over the VR group. Although the animation of the guiding hands is identical in both VR and MR simulations, instead of holding nothing in the air and simply following playback in VR, the compulsory action in MR, including feeling the tension of the neonate being delivered, sensing its weight in their hands, and eventually placing it back on the maternal manikin to complete the procedure. These requirements encourage sustained attentional engagement, as insufficient or imprecise manipulation makes it difficult for the neonate to be delivered. Consequently, the MR condition reinforces trainees' focus on the hand manoeuvres and the procedural sequence, strengthening their memory of both the individual actions and the overall delivery process.

Table 4: Statistical comparison of DOPS assessment scores using the Mann-Whitney test.

Category	Median (VR / MR)	IQR (VR / MR)	U	p-value
Hand Placement	1.5 / 1.5	1.00 / 1.00	552.5	0.284
Delivery	1.8 / 2.0	1.00 / 0.88	443.0	$p < 0.05$
Post-Delivery	2.0 / 2.8	1.25 / 1.25	327.0	$p < 0.001$
Overall Handling	1.0 / 1.3	0.50 / 0.25	365.5	$p = 0.001$

To investigate whether participants' subjective preference for MR training influenced their actual skill acquisition, we computed Spearman's rank correlation between the rated preference total (Fig. 10) and the graded DOPS total (Fig. 11). While the result suggested there was very weak correlation between the preference ratings and DOPS scores

(Spearman's  $\rho = 0.051$ ), on the other hand, it suggested that participants' personal preference towards the MR system did not influence their skill acquisition. It further implies that the design of this MR simulation is capable of effectively training the students, provided they follow the guided procedure, and has potential as an efficient tool for self-learning complicated hand manoeuvres.

## 5 DISCUSSION AND FUTURE WORK

The effectiveness of the proposed MR system relies on the accuracy and temporal stability of the underlying spatial registration pipeline. In close-range manipulation scenarios such as childbirth delivery, even minor deviations can undermine the perceived alignment. Our findings from the user study indicate that the immersive interaction is driven by the fidelity of spatial coupling between visual guidance and physical interaction. By enabling trainees to directly map visual cues to physical contact locations without additional mental transformation, accurate spatial coupling constitutes a key system-level mechanism through which immersive interaction translates into effective skill acquisition. This perspective also helps clarify the limitations of prior works [14, 20] that rely on visually correct but spatially decoupled overlays, in which visual guidance is not spatially aligned with the physical contact space. Such systems implicitly assume that visual guidance alone is sufficient for skill acquisition, even when visual instructions and haptic interaction occur in misaligned reference frames. Our results suggest that this decoupling introduces spatial ambiguity during close-range manipulation, helping explain the limited transfer reported in VR-based training for tasks dominated by fine-grained hand-object interaction. Importantly, the proposed system does not require room-scale tracking infrastructure, which simplifies deployment and makes it more suitable for large-scale training scenarios.

Based on discussions with obstetric specialists, we identify several clinically motivated directions that are critical for bringing MR-based training closer to real delivery experience. First, the effectiveness of the proposed system is bounded by the assumption of rigid spatial registration. In normal childbirth scenarios, interaction is dominated by the global pose of the neonatal head, enabling rigid alignment to provide stable guidance. In contrast, complex cases such as breech presentation or shoulder dystocia violate this assumption due to non-rigid deformation and articulated motion, indicating the need for non-rigid neonatal pose estimation. Another implication concerns the use of pre-recorded expert hand trajectories. While this design ensures reproducibility and simplifies deployment, it assumes a fixed correspondence between expert demonstrations and the user's physical interaction space. This assumption, though sufficient for the current study, may contribute to performance variability across users with different anthropometric characteristics, motivating anthropometry-aware guidance. Finally, MR-based training sessions reveal that accurate spatial alignment alone is insufficient to constrain safe manipulation. Some trainees applied excessive pulling forces, exposing a sim-to-real gap caused by the absence of force-aware feedback. This boundary of visual guidance suggests that spatial fidelity should be complemented by mechanisms that convey physical constraints, such as force feedback or stress visualization.

## 6 CONCLUSION

This research demonstrates the feasibility and pedagogical value of integrating MR into obstetric training. By aligning virtual models with physical manikins through a coarse-to-fine strategy, the system enables structurally faithful guidance that reinforces correct hand manoeuvres during delivery practice. Technical evaluations confirmed the accuracy, stability and real-time performance of the manikin localization pipeline on a standalone headset, validating the practicality of the approach. Importantly, a large-scale user study with medical students revealed that MR training not only enhanced task performance but was also consistently preferred to VR-based training. Taken together, these results highlight MR as a promising paradigm for simulation-based medical education. The proposed system reduces reliance on continuous expert supervision, offers scalable access to high-fidelity training, and provides a generalizable framework that could be extended to other manikin-based healthcare training scenarios.

## REFERENCES

- [1] K. Singh, E. Yong, and P. Wong, "The teaching of obstetrics and gynaecology in singapore from 1905 to the present," *Ann Acad Med Singapore*, vol. 34, no. 6, pp. 121C–125C, 2005. [1](#)
- [2] A. Akkad, S. Bonas, and P. Stark, "Gender differences in final year medical students' experience of teaching of intimate examinations: a questionnaire study," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 115, no. 5, pp. 625–632, 2008. [1](#)
- [3] S. Dotters-Katz, A. Panzer, M. Givens, M. Smid, and A. Chuang, "Preclinical labor-and-delivery shadowing: The impact on medical students' perceptions of obstetrics and gynecology," *American Journal of Perinatology Reports*, vol. 8, no. 2, pp. e64–e67, 4 2018. [Online]. Available: <http://dx.doi.org/10.1055/s-0038-1641723> [1, 2](#)
- [4] H. Morgan, D. Marzano, M. Lanham, T. Stein, D. Curran, and M. Hammoud, "Preparing medical students for obstetrics and gynecology milestone level one: a description of a pilot curriculum," *Medical Education Online*, vol. 19, no. 1, p. 25746, 1 2014. [Online]. Available: <http://dx.doi.org/10.3402/meo.v19.25746> [1](#)
- [5] E. Wilson, S. Janssens, D. G. Hewett, B. Jolly, and M. Beckmann, "Simulation training in obstetrics and gynaecology: What's happening on the frontline?" *Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol. 56, no. 5, pp. 496–502, 2016. [1](#)
- [6] G. Jones, C. Mitchell, J. Hirst, and D. Anumba, "The royal college of obstetricians and gynaecologists," *Understanding the relationship between social*, vol. 23, p. 96, 2024. [1](#)
- [7] K. Daniels, J. Arafah, A. Clark, S. Waller, M. Druzin, and J. Chueh, "Prospective randomized trial of simulation versus didactic teaching for obstetrical emergencies," *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, vol. 5, no. 1, pp. 40–45, 2 2010. [Online]. Available: <http://dx.doi.org/10.1097/SIH.0b013e3181b65f22> [1, 2](#)
- [8] L. Birch, N. Jones, P. Doyle, P. Green, A. McLaughlin, C. Champney, D. Williams, K. Gibbon, and K. Taylor, "Obstetric skills drills: Evaluation of teaching methods," *Nurse Education Today*, vol. 27, no. 8, pp. 915–922, 11 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.nedt.2007.01.006> [1, 2](#)
- [9] R. Sutherland, K. Reid, D. Kok, and M. Collins, "Teaching a fishbowl tutorial: sink or swim?" *The Clinical Teacher*, vol. 9, no. 2, pp. 80–84, 3 2012. [Online]. Available: <http://dx.doi.org/10.1111/j.1743-498X.2011.00519.x> [1, 2](#)
- [10] M. Barrie, J. J. Socha, L. Mansour, and E. S. Patterson, "Mixed reality in medical education: narrative literature review," in *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, vol. 8, no. 1. Sage Publications Sage CA: Los Angeles, CA, 2019, pp. 28–32. [1](#)
- [11] T. Turso-Finnich, R. O. Jensen, L. X. Jensen, L. Konge, and E. Thinggaard, "Virtual reality head-mounted displays in medical education: a systematic review," *Simulation in Healthcare*, vol. 18, no. 1, pp. 42–50, 2023. [1](#)
- [12] H. W. Ng, A. Koh, A. Foong, J. Ong, J. H. Tan, E. T. Khoo, and G. Liu, "Real-time spoken language understanding for orthopedic clinical training in virtual reality," in *International Conference on Artificial Intelligence in Education*. Springer, 2022, pp. 640–646. [1](#)
- [13] C.-Y. Chang, P. Panjaburee, and S.-C. Chang, "Effects of integrating maternity vr-based situated learning into professional training on students' learning performances," *Interactive Learning Environments*, vol. 32, no. 5, pp. 2121–2135, 2024. [1](#)
- [14] C. Liu, F. F.-Y. Tan, S. Zhao, A. Kanneganti, G. A. Tushar, and E. T. Khoo, "Facilitating virtual reality integration in medical education: A case study of acceptability and learning impact in childbirth delivery training," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–14. [1, 2, 9](#)
- [15] M. W. Benjamin and O. Sabri, "Using haptic feedback in a virtual reality bone drilling simulation to reduce plunge distance," *Cureus*, vol. 13, no. 9, 2021. [2](#)
- [16] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, pp. 21–53, 2013. [2](#)
- [17] V. Y. Cheung, Y. M. Tang, and K. K. Chan, "Medical students' perception of the application of a virtual reality training model to acquire vaginal examination skills," *International Journal of Gynecology & Obstetrics*, vol. 161, no. 3, pp. 827–832, 2023. [2](#)
- [18] C.-H. Yang, S.-F. Liu, C.-Y. Lin, and C.-F. Liu, "Immersive virtual reality-

- based cardiopulmonary resuscitation interactive learning support system,” *IEEE Access*, vol. 8, pp. 120 870–120 880, 2020. 2
- [19] K. Rangarajan, H. Davis, and P. H. Pucher, “Systematic review of virtual haptics in surgical simulation: a valid educational tool?” *Journal of surgical education*, vol. 77, no. 2, pp. 337–347, 2020. 2
- [20] L. W. Ljungblad, D. Murphy, and H. E. Fonkalsrud, “A mixed reality for midwifery students: a qualitative study of the technology’s perceived appropriateness in the classroom,” *BMC medical education*, vol. 25, no. 1, p. 337, 2025. 2, 9
- [21] T. Sielhorst, T. Obst, R. Burgkart, R. Riener, and N. Navab, “An augmented reality delivery simulator for medical training,” in *International workshop on augmented environments for medical imaging-MICCAI Satellite Workshop*, vol. 141, 2004, pp. 11–20. 2
- [22] H. Cheng, C. Xu, X. Chen, Z. Chen, J. Wang, and L. Zhao, “Realistic volume rendering with environment-synced illumination in mixed reality,” in *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2023, pp. 423–428. 2, 3
- [23] H. Cheng, C. Xu, Z. Chen, J. Wang, Y. Chen, and L. Zhao, “A mixed reality framework for interactive realistic volume rendering with dynamic environment illumination,” in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2023, pp. 741–742. 2
- [24] D. Cotter, M. J. Turner, F. M. McAuliffe, and M. F. Higgins, “Medical students learning experiences of the labour ward: a qualitative research study,” *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 206, pp. 204–207, 2016. 2
- [25] D. Grasby and J. A. Quinlivan, “Attitudes of patients towards the involvement of medical students in their intrapartum obstetric care,” *Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol. 41, no. 1, pp. 91–96, 2001. 2
- [26] J. C. Chang, M. R. Odrobina, and K. McIntyre-Seltman, “The effect of student gender on the obstetrics and gynecology clerkship experience,” *Journal of women’s health*, vol. 19, no. 1, pp. 87–92, 2010. 2
- [27] B. Mavis, P. Vasilenko, R. Schnuth, J. Marshall, and M. C. Jeffs, “Medical students’ involvement in outpatient clinical encounters: A survey of patients and their obstetricians–gynecologists,” *Academic Medicine*, vol. 81, no. 3, pp. 290–296, 2006. 2
- [28] C.-Y. Chang, H.-Y. Sung, J.-L. Guo, B.-Y. Chang, and F.-R. Kuo, “Effects of spherical video-based virtual reality on nursing students’ learning performance in childbirth education training,” *Interactive Learning Environments*, vol. 30, no. 3, pp. 400–416, 2019. 2
- [29] C.-W. Chang, S.-C. Yeh, M. Li, and E. Yao, “The introduction of a novel virtual reality training system for gynecology learning and its user experience research,” *IEEE Access*, vol. 7, pp. 43 637–43 653, 2019. 2
- [30] A. S. Nugraha, A. N. Faza, W. Indayani, H. Harryanto, and A. Setiawan, “A natural childbirth training simulation in virtual environment for prospective midwife,” in *2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE, 2018, pp. 1–5. 2
- [31] N. F. A. B. Mazalan and S. Y. Tan, “An interactive mobile augmented reality for obstetrics learning,” in *Proceedings of the 2023 7th International Conference on Virtual and Augmented Reality Simulations*, ser. ICVARS ’23. New York, NY, USA: Association for Computing Machinery, 2023, p. 84–92. [Online]. Available: <https://doi.org/10.1145/3603421.3603433> 2
- [32] A. Ballit, M. Hirvert, C. Rubod, and T.-T. Dao, “Fast soft-tissue deformations coupled with mixed reality toward the next-generation childbirth training simulator,” *Medical Biological Engineering Computing*, vol. 61, no. 8, pp. 2207–2226, 6 2023. [Online]. Available: <http://dx.doi.org/10.1007/s11517-023-02864-5> 2
- [33] M. Schneider, C. Kunz, A. Pal’A, C. R. Wirtz, F. Mathis-Ullrich, and M. Hlaváč, “Augmented-reality-assisted ventriculostomy,” *Neurosurgical focus*, vol. 50, no. 1, p. E16, 2021. 2
- [34] T. Frantz, B. Jansen, J. Duerinck, and J. Vandemeulebroucke, “Augmenting microsoft’s hololens with vuforia tracking for neuronavigation,” *Healthcare technology letters*, vol. 5, no. 5, pp. 221–225, 2018. 2
- [35] T. Fick, J. van Doormaal, E. Hoving, L. Regli, and T. van Doormaal, “Holographic patient tracking after bed movement for augmented reality neuronavigation using a head-mounted display,” *Acta Neurochirurgica*, vol. 163, pp. 879–884, 2021. 2
- [36] M. Lai, S. Skryman, C. Shan, D. Babic, R. Homan, E. Edström, O. Persson, G. Burström, A. Elmi-Terander, B. H. Hendriks *et al.*, “Fusion of augmented reality imaging with the endoscopic view for endonasal skull base surgery; a novel application for surgical navigation based on intraoperative cone beam computed tomography and optical tracking,” *Plos one*, vol. 15, no. 1, p. e0227312, 2020. 2
- [37] S. Eom, D. Sykes, S. Rahimpour, and M. Gorlatova, “Neurolens: Augmented reality-based contextual guidance through surgical tool tracking in neurosurgery,” in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 355–364. 2, 3
- [38] F. Zhang, L. Chen, W. Miao, and L. Sun, “Research on accuracy of augmented reality surgical navigation system based on multi-view virtual and real registration technology,” *IEEE Access*, vol. 8, pp. 122 511–122 528, 2020. 2
- [39] K. W. Ng, Y. Gao, M. S. Furqan, Z. Yeo, J. Lau, K. Y. Ngiam, and E. T. Khoo, “Holopocus: Portable mixed-reality 3d ultrasound tracking, reconstruction and overlay,” in *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer, 2023, pp. 111–120. 2
- [40] B. Wen, W. Yang, J. Kautz, and S. Birchfield, “Foundationpose: Unified 6d pose estimation and tracking of novel objects,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 868–17 879. 3
- [41] Y. Labb  , L. Manuelli, A. Mousavian, S. Tyree, S. Birchfield, J. Tremblay, J. Carpentier, M. Aubry, D. Fox, and J. Sivic, “Megapose: 6d pose estimation of novel objects via render & compare,” *arXiv preprint arXiv:2212.06870*, 2022. 3
- [42] F. Li, S. R. Vutukur, H. Yu, I. Shugurov, B. Busam, S. Yang, and S. Ilic, “Nerf-pose: A first-reconstruct-then-regress approach for weakly-supervised 6d object pose estimation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2123–2133. 3
- [43] T. Lee, B. Wen, M. Kang, G. Kang, I. S. Kweon, and K.-J. Yoon, “Any6d: Model-free 6d pose estimation of novel objects,” *arXiv preprint arXiv:2503.18673*, 2025. 3
- [44] I. Shugurov, F. Li, B. Busam, and S. Ilic, “Osop: A multi-stage one shot object pose estimation framework,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6835–6844. 3
- [45] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299. 3
- [46] J. Jiang, P. Streli, X. Luo, C. Gebhardt, and C. Holz, “Manikin: biomechanically accurate neural inverse kinematics for human motion estimation,” in *European Conference on Computer Vision*. Springer, 2024, pp. 128–146. 3
- [47] D. Groos, L. Adde, R. St  en, H. Ramampiaro, and E. A. Ihlen, “Towards human-level performance on automatic pose estimation of infant spontaneous movements,” *Computerized Medical Imaging and Graphics*, vol. 95, p. 102012, 2022. 3
- [48] D. G. Kyrollos, A. Fuller, K. Greenwood, J. Harrold, and J. R. Green, “Under the cover infant pose estimation using multimodal data,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023. 3
- [49] D. Scherfgren and J. Schild, “Estimating the pose of a medical manikin for haptic augmentation of a virtual patient in mixed reality training,” in *Proceedings of the 23rd symposium on virtual and augmented reality*, 2021, pp. 33–41. 3
- [50] K. W. Ng, M. S. Furqan, Y. Gao, K. Y. Ngiam, and E. T. Khoo, “Holovein—mixed-reality venipuncture aid via convolutional neural networks and semi-supervised learning,” *Electronics*, vol. 12, no. 2, p. 292, 2023. 3
- [51] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. St  hmer, T. J. Cashman, B. Tekin, J. L. Sch  nberger *et al.*, “Hololens 2 research mode as a tool for computer vision research,” *arXiv preprint arXiv:2008.11239*, 2020. 3
- [52] S. Garrido-Jurado, R. Mu  oz-Salinas, F. J. Madrid-Cuevas, and M. J. Mar  n-Jim  nez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014. 4
- [53] Y. C. Shiu and S. Ahmad, “Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form  $ax = xb$ ,” 1987. 4
- [54] R. Y. Tsai, R. K. Lenz *et al.*, “A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration,” *IEEE Transactions on robotics and automation*, vol. 5, no. 3, pp. 345–358, 1989. 4
- [55] B. Wieczorek,   . Wargula, M. Kukla, A. Kubacki, and J. G  recki, “The

- effects of aruco marker velocity and size on motion capture detection and accuracy in the context of human body kinematics analysis,” *Technical Transactions*, vol. 117, no. 1, 2020. [4](#)
- [56] G. Casiez, N. Roussel, and D. Vogel, “1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 2527–2530. [4](#)
- [57] Q.-Y. Zhou, J. Park, and V. Koltun, “Fast global registration,” in *European conference on computer vision*. Springer, 2016, pp. 766–782. [5](#)
- [58] Y. Chen and G. Medioni, “Object modelling by registration of multiple range images,” *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992. [5](#)