frontiers | Frontiers in Virtual Reality

TYPE Original Research
PUBLISHED 27 January 2025
DOI 10.3389/frvir.2025.1499830

OPEN ACCESS

EDITED BY
Andrea Sanna,
Polytechnic University of Turin, Italy

REVIEWED BY
Aikaterini Bourazeri,
University of Essex, United Kingdom
Lia Morra,
Polytechnic University of Turin, Italy

*CORRESPONDENCE
Mikołaj Łysakowski,
mikolaj.lysakowski@put.poznan.pl

RECEIVED 21 September 2024
ACCEPTED 07 January 2025
PUBLISHED 27 January 2025

CITATION
Łysakowski M, Gapsa J, Lyu C, Bohné T, Tadeja SK and Skrzypczyński P (2025) Enhancing augmented reality with machine learning for hands-on origami training. *Front. Virtual Real.* 6:1499830. doi: 10.3389/frvir.2025.1499830

Enhancing augmented reality with machine learning for hands-on origami training

Mikołaj Łysakowski[1]*, Jakub Gapsa[2], Chenxu Lyu[3], Thomas Bohné[3], Sławomir Konrad Tadeja[3] and Piotr Skrzypczyński[1,4]

[1]Center for Artificial Intelligence and Cybersecurity, Poznań University of Technology, Poznań, Poland, [2]Faculty of Mechanical Engineering, Poznań University of Technology, Poznań, Poland, [3]Department of Engineering, University of Cambridge, Cambridge, United Kingdom, [4]Institute of Robotics and Machine Intelligence, Poznań University of Technology, Poznań, Poland

This research explores integrating augmented reality (AR) with machine learning (ML) to enhance hands-on skill acquisition through origami folding. We developed an AR system using the YOLOv8 model to provide real-time feedback and automatic validation of each folding step, offering step-by-step guidance to users. A novel approach to training dataset preparation was introduced, which improves the accuracy of detecting and assessing origami folding stages. In a formative user study involving 16 participants tasked with folding multiple origami models, the results revealed that while the ML-driven feedback increased task completion times, it also made participants feel more confident throughout the folding process. However, they also reported that the feedback system added cognitive load, slowing their progress, though it provided valuable guidance. These findings suggest that while ML-supported AR systems can enhance the user experience, further optimization is required to streamline the feedback process and improve efficiency in complex manual tasks.

KEYWORDS

augmented reality, machine learning, edge computing, assembly task, education

1 Introduction

Augmented reality (AR) is increasingly transforming education by enabling immersive, hands-on learning experiences, particularly in scenarios where human instructors are unavailable or traditional learning environments are inadequate (Zonaphan et al., 2022). By overlaying digital content onto the physical world, AR facilitates interactive and engaging training, making it a powerful tool for skill development in both education and practical applications (Zambri and Kamaruzaman, 2020).

The integration of AR with machine learning (ML) introduces new possibilities for automated feedback in manual skill acquisition. Building on our previous work (Łysakowski et al., 2024), we explore how an AR system powered by a YOLOv8 model can provide real-time detection and evaluation of user actions, specifically within the context of origami folding. The novelty of this research lies in the deployment of a state-of-the-art object detection algorithm on a resource-constrained AR device (HoloLens 2) to enable on-device step validation without requiring external computation. Unlike previous systems, which focus on predefined animations or step sequences, our approach evaluates both user actions and outcomes in real time, directly influencing the learning process.

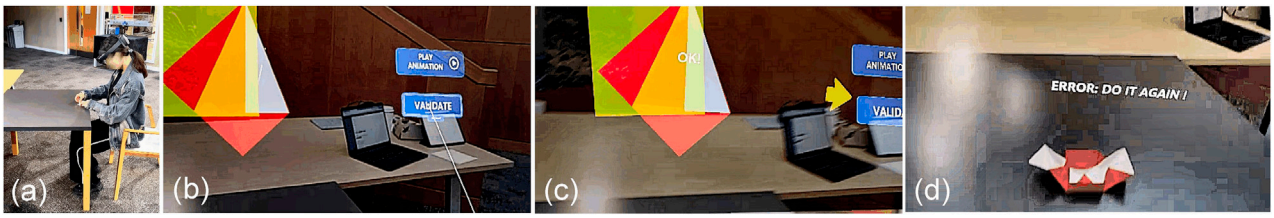

FIGURE 1
A user wearing a HoloLens 2 head-mounted display while folding an origami model **(A)**, and screenshots from the AR application showing: **(B)** an animated model guiding the folding process, **(C)** a correctly completed folding step, and **(D)** an incorrectly completed folding step.

Origami, the Japanese art of paper folding, is known to enhance manual and cognitive skills by improving fine motor abilities, hand-eye coordination, and spatial awareness (Supple et al., 2021). These benefits make origami a valuable tool in education, therapy, and personal development, with research supporting its effectiveness across various fields (Zhao et al., 2020). For instance, Herbas Torrico, 2021 highlighted how origami was used during COVID-19 lockdowns as an educational tool, emphasizing its role in cognitive development. The pandemic also revealed origami's versatility, with Monnier and Winters (2022) demonstrating its calming and creative benefits during the lockdown, though also noting that printed instructions were challenging to follow unless participants had prior knowledge of basic folds. Tutors play an essential role in guiding learners through complex folding techniques, ensuring proper understanding of key concepts (Andreass, 2011).

Existing AR-based origami tutorials, such as Wiwatwattana et al. (2016), lack the ability to recognize user actions or provide automated feedback, limiting their effectiveness. To address this gap, our system integrates state detection and validation using YOLOv8, an advanced computer vision model designed for real-time edge applications. This ensures that origami folding steps are evaluated dynamically, enabling interactive, step-by-step guidance (Figure 1B). Correct folds are validated through the neural network, allowing progression to subsequent steps (Figure 1C), while incorrect folds trigger corrective feedback (Figure 1D).

This work builds upon our previous conference paper (Łysakowski et al., 2024), which primarily demonstrated the technical feasibility of using YOLOv8 for step validation in AR-based origami folding on a limited number of models. In contrast, this study significantly expands the scope by incorporating a broader range of origami designs and conducting a mixed-methods experiment comparing user performance and perceptions with and without ML-driven feedback. By focusing on the effectiveness of real-time ML-driven feedback in a practical setting, this work transitions from a proof-of-concept to a comprehensive evaluation of AR-ML integration for skill acquisition.

2 Related work

Immersive technologies enable learners to interact with virtual simulations that replicate real-world tasks, offering opportunities to develop practical skills in a controlled, repeatable environment (Zonaphan et al., 2022). This is especially valuable for tasks that demand manual dexterity and precision, allowing learners to practice and refine their abilities beyond the constraints of traditional methods.

Integrating neural networks with AR significantly enhances these technologies by offering real-time feedback and adaptive learning paths. This integration is crucial for mastering complex manual tasks where precision and accuracy are vital. In the automotive industry, for example, AR systems enhanced with ML models guide workers through the assembly of intricate engine components. As demonstrated by Zogopoulos et al. (2021), these systems use image-based state tracking to identify parts and tools, overlaying step-by-step instructions directly onto physical components.

Similarly, in industrial maintenance, AR systems combined with predictive ML models offer technicians a virtual training environment to anticipate and address potential faults before they occur (Palmarini et al., 2018). These systems leverage historical data to predict common issues, guiding users through diagnostic and repair procedures with real-time feedback, thus enhancing both learning outcomes and operational efficiency (Danielsson et al., 2020).

In the medical field, AR and ML are revolutionizing surgical training. Systems like those developed by Pauly et al. (2015) allow surgical trainees to practice in a simulated, risk-free environment. ML algorithms analyze surgical movements, providing immediate feedback on technique, which is crucial for developing the precision required in real-life surgeries (Khandelwal et al., 2019).

Traditional crafts, such as carpentry, are also finding new possibilities through the integration of AR and ML. Palmarini et al. (2018) explored AR systems to guide users through processes like cutting, assembling, and finishing woodwork, with ML used for object recognition and tracking to ensure accuracy at each stage. With the use of immersive technologies, such training can be extended, offering learners a virtual workshop where they can experiment and learn without the risk of wasting materials.

Origami, which requires intricate manual skills, serves as an exemplary case for AR-enhanced learning. Wiwatwattana et al. (2016) introduced Origami Guru, an AR application designed to assist users in paper folding. However, limitations such as the lack of real-time feedback and action recognition reduced its effectiveness. Our work addresses these limitations by incorporating ML, particularly object detection frameworks like YOLOv8, to provide real-time feedback and assessment. This approach allows users to practice origami in a virtual environment that replicates the physical

world, offering instant corrections and ensuring precise execution of each folding step.

Research such as PlayGAMI (Grandhi and Chang, 2019) and the work of Watanabe and Kinoshita (2012) have explored the potential of real-time tracking in origami using visual markers on paper or comparing silhouettes of the folded paper for each step, though these methods are less adaptable to standard origami materials. Similarly, Shimanuki et al. (2020) employed single-camera setups to analyze origami operations, but this pipeline requires rigid constraints, such as uniform backgrounds and simplistic assumptions about paper geometry. By contrast, our system applies YOLOv8, leveraging its ability to generalize across diverse backgrounds and lighting conditions without relying on strict environmental controls.

Recently, the mixed-reality system Origami Sensei (Chen et al., 2023) achieved automatic origami step recognition based solely on the paper's appearance, using computer vision to identify folding steps and provide real-time projections and verbal instructions. Unlike our approach, which leverages YOLOv8 directly on a HoloLens 2 for real-time, on-device feedback and step validation, Origami Sensei relies on external devices like a tablet and projector, introducing latency and hardware dependencies. By implementing our system entirely within the AR glasses, we eliminate external components, achieving a more seamless and portable user experience. The application of object detection on AR headsets has been a significant area of research (Farasin et al., 2020; Goka et al., 2022; Kim et al., 2023; Łysakowski et al., 2023a). For instance, Farasin et al. (2020) investigated a two-stage network strategy for object detection and tracking that relies on offloading computation to a high-performance server. While this method benefits from server-side processing power, it restricts the headset to merely capturing video frames and displaying the processed results, necessitating reliable and fast network connections like Wi-Fi or LTE. In contrast, our approach focuses on achieving real-time object detection directly on the headset using the advanced YOLOv8s network, eliminating the need for server-side processing. This software design reduces latency, which is essential for maintaining user immersion (Chen et al., 2018).

Furthermore, simpler algorithms that are designed to execute in real-time on AR headsets, as demonstrated by Malek et al. (2022), highlight the practicality of performing computations directly on the device for certain tasks. Additionally, frameworks such as Vuforia (PTC, 2023) and EasyAR (EasyAR, 2024) can also be utilized. Our approach strikes a balance by employing a robust, cutting-edge model on the AR headset, combining the benefits of real-time, on-device processing with the adaptability and broad applicability of sophisticated object detection frameworks.

While YOLOv8 provides critical real-time detection capabilities for a hands-on AR application, it faces challenges such as varying environmental conditions and the computational constraints of AR devices. Łysakowski et al. (2023b) highlight these issues, noting that factors like lighting and view angles can impact detection accuracy. To mitigate these challenges, alternative object detection algorithms can be considered. For instance, the Single Shot MultiBox Detector (SSD) discussed by Liu et al. (2015) offers a balance between speed and accuracy, making it suitable for scenarios where slightly higher processing latency is acceptable. This approach gives users more time to process visual and contextual information, reducing cognitive overload and allowing for better decision-making (Johri et al., 2024). Alternatively, Faster R-CNN (Ren et al., 2015) provides higher detection accuracy at the cost of speed, making it a viable option for tasks requiring meticulous object recognition and classification, such as detailed medical simulations or complex industrial processes.

The integration of ML with AR for hands-on training, as explored in this work, marks a significant advancement by enabling autonomous, real-time evaluation of each step's correctness directly on AR headgear. This approach surpasses traditional AR tutorials by offering intelligent, interactive feedback, enhancing user independence.

3 System architecture and implementation

Our system is implemented within the Unity game engine, leveraging its robust capabilities to develop an interactive AR application aimed at assisting users with origami folding lessons. At the start, users can select the primary color of the paper sheet, assuming the opposite side is white, and then choose one of the available origami models to fold. Initially, our application presents an animation demonstrating the complete folding process (see Figure 1B), followed by interactive, step-by-step instructions. Next, the users physically fold the paper according to these instructions and, after completing each step, press a virtual button to confirm their progress (see Figure 1C). The completion of this process results in a folded origami model.

The system architecture (Figure 2) integrates a neural network-based algorithm that predicts the state of the folded paper after each step. Specifically, a YOLOv8s (Yaseen, 2024) neural network, a state-of-the-art object detection framework known for its speed and accuracy in identifying objects within images, is employed to detect and classify the various stages of the paper folding process. The YOLO ("You Only Look Once") model operates by dividing an input image into a grid and predicting bounding boxes and class probabilities for objects within the image in a single pass, making them well-suited for real-time applications (Redmon et al., 2016). Since predictions are made on demand, the system uses a compact neural network model with input images resized to 320 × 320 pixels, achieving an inference time of approximately 500 m, which strikes a balance between accuracy and performance.

The diagram (Figure 3) outlines the interactive origami folding process supported by animation and real-time validation. It features a sequence starting with an animation button to demonstrate each folding step, followed by a validation process that uses camera-based feedback to assess fold result, displaying success or error messages accordingly.

We chose the YOLOv8s model for its efficiency in balancing speed and accuracy, making it well-suited for real-time or near-real-time applications, such as detecting the stages of paper folding in our AR origami system. YOLOv8s offers a compact architecture that delivers high detection accuracy while maintaining fast inference times (Jocher et al., 2023), which is crucial for processing on resource-constrained devices like the Microsoft HoloLens 2 (HL2) mixed reality head-mounted display (HMD). Additionally, YOLOv8s' ability to perform well with relatively small input image sizes (in our case 320 × 320 pixels) ensures that our system can

10.3389/frvir.2025.1499830

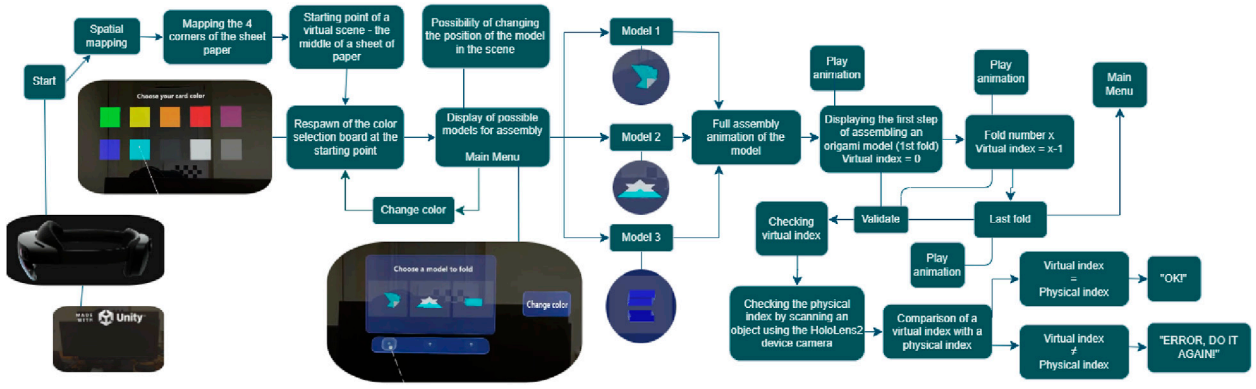

FIGURE 2
Data and control flow in the origami folding application with ML-based outcome assessment.

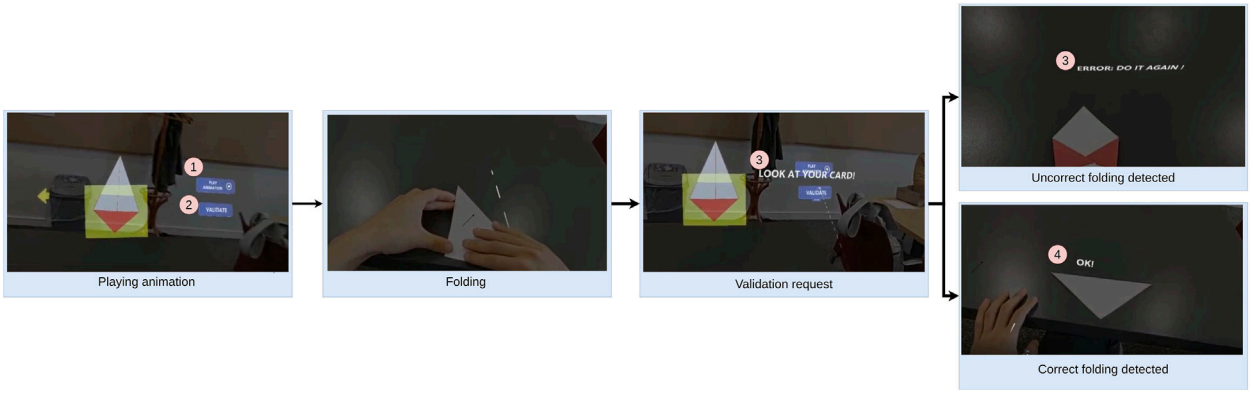

FIGURE 3
The diagram illustrates the step-by-step origami folding process with feedback. The initial step illustrates the (1) "Play animation" button, which initiates the animation demonstrating the correct method for executing the current step of the folding process. Once the folding process is complete, the user is required to press the button labeled (2), which initiates the validation process. Subsequently, the user is prompted by the text (3) "Look at your card" to move their head to make the paper sheet visible in the camera. In the event of a successful fold detection, the text element (4), "OK!", is displayed. Conversely, in the event of an unsuccessful fold, the text (3), "Error! Do it again!", is displayed.

quickly and accurately classify folding stages without compromising the user experience. The model's robust performance across various datasets also ensures that it can generalize well to unseen paper colors and different lighting conditions, further justifying its use in this application. We further enhanced system performance by employing the Unity Sentis library, a successor to the Barracuda neural network inference library, for on-device inference on the HL2 HMD.

The origami models were designed using Blender software, following a detailed analysis and preparation of physical models. The folding mechanics were studied, and a virtual grid was created to accurately represent the fold lines (Figure 4A). Each folding step was modeled as a distinct entity, with transitions between steps smoothly animated within Unity. These animations were carefully synchronized so that the final state of one step seamlessly aligns with the initial state of the next, as illustrated in Figure 4B. Although an animation method similar to the one proposed by Agui et al. (1983), which utilizes folding rules, could have been applied, the keyframe-based method supported by Blender offers several advantages. Blender allows for precise control over fold geometry, integrates seamlessly with modern 3D rendering tools, including Unity (Suhail et al., 2024), and supports intuitive adjustments of fold animations through a graphical interface, reducing the complexity of manually assigning keyframe coordinates.

The YOLOv8s model was trained on a custom dataset comprising images of the folded models captured from different angles and distances to closely resemble real-world conditions (see Figure 5). To capture these images the frontal camera of the HL2 HMD was used, which is the only RGB camera available on this device.

To streamline the dataset preparation process, we utilized the Segment Anything Model (SAM) by Kirillov et al. (2023), an advanced image segmentation tool developed by Meta AI, and the Track Anything algorithm (Yang et al., 2023), which is based on the Segment Anything Model. The ability of SAM to perform zero-shot learning, which allows it to identify objects without needing specific training examples, makes it an ideal fit for our application. We used prompts based on points on the folded objects

10.3389/frvir.2025.1499830

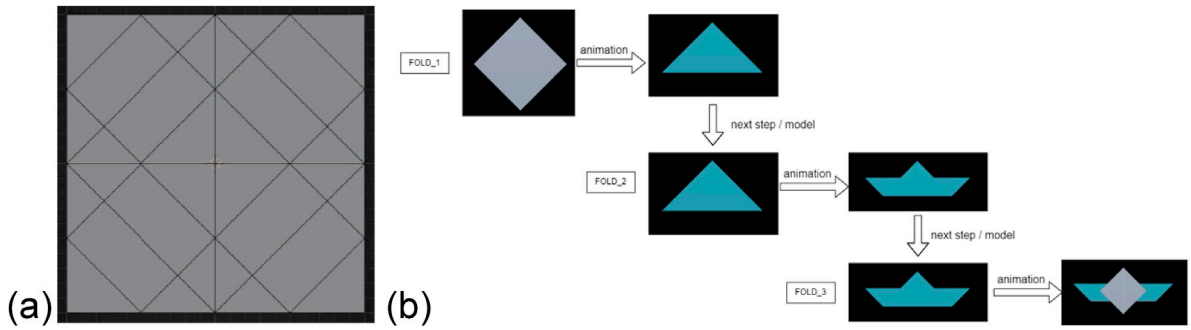

FIGURE 4
Representation of an example models' grid in Blender **(A)** and a diagram depicting the step-by-step process of folding the model **(B)**.

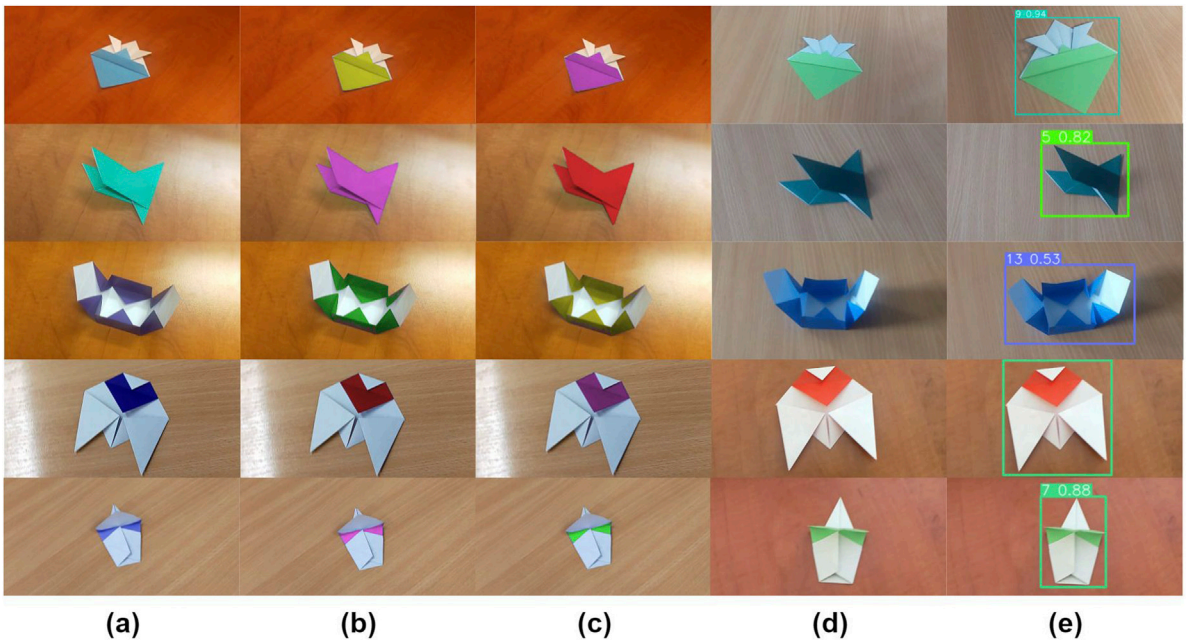

FIGURE 5
Example images for the samurai hat, bird, box, fly and yacht origami models. Images from the train/val set with color augmentation **(A–C)**. Examples from the test set **(D)** and final detection **(E)**.

to guide SAM in achieving precise segmentation. Then, Track Anything facilitated the generation of the ground truth data. After manually labeling the first frame of each stage, we used Track Anything to segment the folded models from the remaining images in the sequence, considering different color on both sides of the folded paper sheet. The obtained segmentation masks were used to create bounding boxes, which were then saved in the appropriate format. This automated pipeline significantly expedited the process of data acquisition, labeling, and model training, enabling the efficient training of the YOLOv8 neural network.

Additionally, color augmentation was applied in the HSV (Hue, Saturation, Value) space to increase the dataset's variability by changing the color of the other (non-white) side of the paper. This process resulted in approximately 150 training images and 70 validation images per stage for each origami model. Additionally, 30 images per stage were collected for the test set, featuring new paper colors and various lighting conditions to ensure robust generalization. The dataset covers all folding stages, and standard augmentation techniques—such as image rotation, translation, scaling, and mosaics—were applied during training. The test set did not include color replacement augmentation.

Figure 5 displays the outcomes of the color augmentation process for the origami models, followed by example images from the test set and successful detections made by the model. Finally, we tested the detection procedure using the YOLOv8s network on a separate test set of images, and the resulting Average Precision (*AP*) values are shown in Table 1. Average Precision indicates how well the model predicts the correct folding stage across various threshold values, while AP_{50}

10.3389/frvir.2025.1499830

TABLE 1 Numerical results of YOLOv8s on the five origami models averaged over all stages of folding.

Model/Dataset	AP_{50}^{val}	AP^{val}	AP_{50}^{test}	AP^{test}
Samurai Hat	0.995	0.966	0.973	0.934
Bird	0.988	0.968	0.969	0.896
Box	0.995	0.982	0.927	0.86
Fly	0.995	0.99	0.97	0.785
Yacht	0.994	0.945	0.985	0.925

represents the precision at an Intersection over Union (IoU) threshold of 50%, measuring the overlap between the predicted and actual paper folds.

For the experiments described, a dataset was compiled featuring the "Yacht", "Samurai Hat", "Bird", "Fly", and "Box" origami models (Figure 7). These models consist of 9, 11, 12, 13, and 19 folding stages, respectively, with each stage corresponding to a specific, predefined paper shape. All these models were prepared according to instructions from the Origami Guide website[1]. They were selected from the extensive number of designs available on the website to represent varying levels of complexity, both in terms of the number of stages and visual appearance, which posed additional challenges for the vision-based automatic validation procedure. All physical models were folded from 15 × 15 cm sheets of standard Toyo Origami Paper. We intentionally used paper sheets that are white on one side and colored (e.g., red) on the other side to enhance the visibility of the folds to the HL2 camera and facilitate the automatic validation process.

The numerical results demonstrate good performance of the YOLOv8s model across all five origami models. The AP_{50}^{val} values on the validation set are consistently high, exceeding 0.98 for all models, indicating that the model can accurately detect and classify the folding stages. Similarly, the overall AP^{val} values for the validation set remain above 0.94, reflecting robust performance across varying IoU thresholds.

In the test set, the AP_{50}^{test} values slightly decrease, particularly for the *Box* model (0.927), which suggests some challenges in detecting this model's folding stages under different conditions, such as varying paper colors and lighting. The AP^{test} for the *Fly* model drops significantly, indicating difficulties with detection at higher IoU levels. Nevertheless, the AP^{test} values for all other models, though slightly lower than the validation set, remain strong, showing that the model generalizes well to unseen data.

4 Research design and methodology

The objective of this research was to explore the impact of deep learning-based feedback within an AR application on the accuracy, efficiency, and user satisfaction of origami folding. In order to do so, we carried out a user study with 16 participants during which they all experienced both conditions in a randomized, balanced order. The results of this mixed-method study allowed us to better understand the influence and limitations of the ML-based feedback) on origami folding tasks previously used to study intricacies of manual skill development (Chen et al., 2023).

4.1 Participants

The study involved 16 participants, hereinafter referred to as P1-P16, in total recruited through the opportunity sampling method from university students and staff. The youngest participant was 20, and the oldest reported being 29 years old (M = 24.19, SD = 2.81). Each participant has engaged in a series of origami folding tasks under two conditions: with and without the assistance of deep learning-based feedback.

To better understand the participants' baseline familiarity with origami, their prior experiences were grouped into four clusters, as summarized in Table 2. The clustering process revealed that most participants had some childhood exposure to origami (Figure 6), though the depth of experience varied. Nine participants revealed a general experience, aligning with the activity's common introduction during the early years. However, some recalled specific motivations, such as therapeutic uses or entrepreneurial activities, while others had minimal or no practical engagement. These varying experiences may have influenced participants' engagement with the AR origami system.

4.2 Experimental design and setup

A within-subject design was used to ensure that each participant experienced both conditions (i.e., feedback and no feedback). Participants folded four origami models with varying complexity and with each model being folded in a balanced order utilizing Latin square method. Furthermore, the balanced Latin square method was also employed to determine the order of conditions, i.e., with and without automated feedback, minimizing learning effects across the trials.

A configuration file with the designed experimental setup was prepared and uploaded to the AR system for each participant. The software then accessed this information, demonstrating the desired order of origami models for each user. A black desk placed nearby and facing a uniformly colored wall was used for the experiment to minimize the source of distraction in the background and strengthen the contrast between the folding paper and the desk surface (see Figure 1A). For coherency, the same paper and lighting conditions are maintained throughout. The camera stream was recorded during the experiment as a reference to assess folding quality.

4.3 Origami models

To evaluate our AR system's effectiveness across various task complexities, we selected five origami models, as illustrated in Figure 7. One of the models, namely, (T) *Bird*, was used during the training session with each participant as it provided a relatively moderate challenge in terms of folding difficulty. The remaining

1 https://origami.guide/

TABLE 2 Participants' familiarity with origami, based on prior experience revealed during interviews.

Type of experience	Participants	Description
General childhood experience	P2, P4, P6, P7, P8, P9, P12, P14, P15	Engaged in origami during childhood without recalling specific details or models
No or minimal experience	P5, P10, P13	Little to no exposure to origami, limited to basic activities like folding paper planes
Specific childhood experience	P3, P11, P16	Recalled detailed childhood engagement, including folding specific models (e.g., boxes in school, therapeutic activities, or entrepreneurial endeavors)
Awareness without experience	P1	Awareness of origami concepts but lacked practical engagement

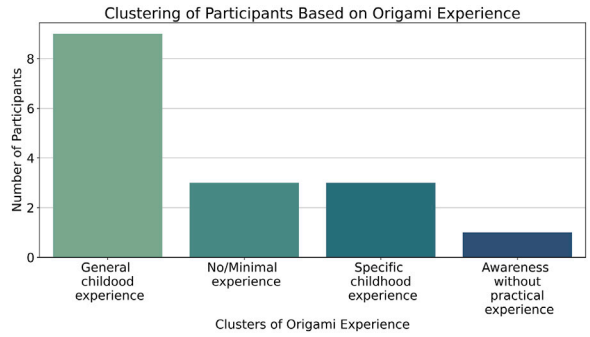

FIGURE 6
Clustering of participants based on origami experience.

models (A) (B) (C) (D) were folded by the participants in randomized, balanced order.

These models were chosen to represent a range of folding complexities, understood as the number of steps required for their completion and folding difficulty. This approach aligns with the existing definitions of task complexity in the context of AR assistance systems, where the number of steps and their difficulty can be considered key complexity indicators (Bock et al., 2024). By varying the complexity of the origami models, we were able to assess our system's ability to support hands-on training across a broader spectrum of origami folding tasks.

4.4 Task and procedure

The main task of the participants was to fold four origami models of varying complexities. Each model was folded twice: once with the machine-learning-based feedback and once without it, resulting in the folding of eight origami models. The whole experimental procedure consisted of three subsequent phases:

Training session: Participants folded a simple model ((T) *Bird*, see Figure 7) using the AR system with and without deep learning-based feedback to familiarize themselves with its features. Experimental task: Each participant folded all four models (see Figure 7) under both conditions (with and without feedback), resulting in eight origami models (i.e., each model was folded twice with and without feedback). After each condition, participants completed two questionnaires to assess their cognitive load (Hart and Staveland, 1988) and flow (Engeser and Rheinberg, 2008) with and without machine-learning-based feedback. The order of experiencing the conditions by the participants (P1-P8) was balanced utilizing the Latin square design. Post-assessment: After the task, participants took part in a semi-structured discussion to assess their confidence, understanding, and overall experience with the folding process.

4.5 Data collection and analysis protocol

During the user study, we collected quantitative data from the HL2 headset that allowed us to obtain performance measurements, further enhanced by qualitative information obtained through questionnaires, participants' feedback and our own observations. In quantitative data included the *task completion time*, i.e., the time taken to fold each model completely is recorded automatically from within the AR system.

The *accuracy and aesthetics* of each fold were assessed by three annotators who had not participated in the experiment and independently assigned scores to the completed origami models in a range of $\{0, 1, 2\}$ based on video recordings of the final models.

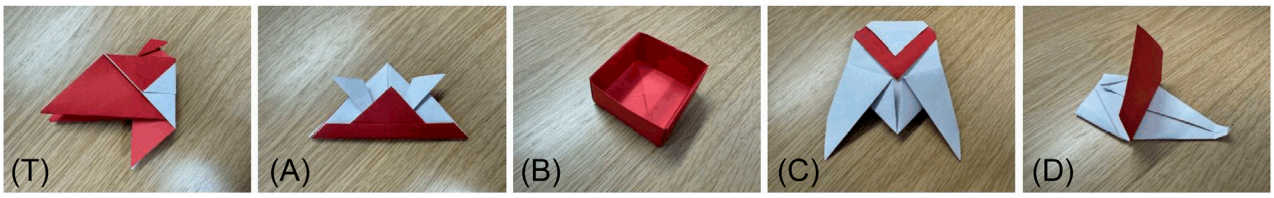

FIGURE 7
Five origami models: (T) *Bird* requiring 12 folds, **(A)** *Samurai Hat* requiring 11 folds, **(B)** *Box* requiring 19 folds, **(C)** *Fly* requiring 13 folds, **(D)** *Yacht* requiring nine folds.

10.3389/frvir.2025.1499830

TABLE 3 The task completion times (i.e., origami folding) for each participant (P1-P16) measured in [s].

Participant	No feedback				With feedback			
P1	228.83	315.66	273.75	214.94	396.83	578.66	221.94	245.77
P2	380.64	299.14	222.57	192.61	507.93	292.16	246.91	191.04
P3	237.82	144.50	183.17	123.62	409.04	340.55	573.68	286.56
P4	221.11	246.20	324.25	304.18	432.70	256.91	499.20	540.71
Participant	with feedback				no feedback			
P5	337.60	759.19	315.83	411.78	344.78	246.54	196.43	174.77
P6	1229.36	719.57	622.24	536.24	391.96	287.56	243.40	270.76
P7	712.59	791.96	668.12	323.87	315.39	218.11	402.70	255.34
P8	485.95	391.57	533.06	707.22	169.57	192.59	210.07	204.38
Participant	no feedback				with feedback			
P9	226.33	396.42	297.17	340.79	338.92	521.05	316.96	621.21
P10	212.25	182.06	107.21	127.95	461.57	233.74	214.84	211.33
P11	295.52	214.96	172.38	166.49	397.95	266.79	597.78	325.77
P12	279.80	232.07	371.78	367.00	614.70	348.39	510.36	518.33
Participant	with feedback				no feedback			
P13	468.71	715.34	281.14	380.57	232.84	237.24	147.67	218.61
P14	1111.19	378.62	347.21	368.94	572.09	201.18	134.93	141.70
P15	646.87	374.47	728.62	472.52	303.94	183.31	214.24	194.66
P16	661.53	543.63	377.67	686.01	178.09	137.66	165.36	262.15

Here, {0} denoted non-complete or wrongly folded model, {1} denoted semi-correctly folded model, and {2} was given to satisfactorily folded model. The final mark was taken as an average of the three annotators' scores.

To gather insights into the *user experience:*, we collected qualitative data through a set of questionnaires. First, we utilized the *NASA Task Cognitive Load* (NASA TLX) (Hart and Staveland, 1988) survey to gauge subjective cognitive workload during a given task execution, which is an often used approach in AR research (Dudley et al., 2018; Bozzi et al., 2023). While NASA TLX was criticized for posing certain limitations (McKendrick and Cherry, 2018), we decided to use it as it provides a structured way to capture crucial subjective qualitative insights concerning AR interface. Second, we measured perceived flow levels, a metric indicative of engagement and skillfulness in immersive interfaces (Engeser and Rheinberg, 2008; Laakasuo et al., 2022; Tadeja et al., 2021b; Bozzi et al., 2023). By combining these metrics, we aimed to ascertain how the AR interface coupled with machine learning-based feedback influences the overall user experience.

5 Results

Here, we present the results of the statistical analysis of the gathered quantitative (i.e., task completion times and quantified questionnaire responses) and qualitative (i.e., participants' feedback and comments as well as our own observations) data.

5.1 Task completion times

We analyzed the log-transformed task completion times using the *analysis of variance* (ANOVA) test. The results showed statistically significant differences ($(F(1.0, 126.0) = 96.702, \eta_p^2 = 0.434, p < .001)$) between trials with and without feedback. The results shown in Table 3 reveal that task completion took longer when participants folded origami with automated feedback than when folding without it. Specifically, the average time for task completion with feedback was 478.27 [s] (SD = 202.25 [s]), compared to 243.02 [s] (SD = 83.78 [s]) without feedback. This difference highlights that integrating the feedback mechanism led to longer task durations.

The range of completion times provides additional insights into this disparity. The shortest observed time with feedback (191.04 [s]) was notably higher than the shortest time without feedback (107.21 [s]). Similarly, the most extended task duration with feedback (1229.36 [s]) was more than double the longest time without feedback (572.70 [s]). This indicates that while the automatic verification functionality offers detailed guidance, it slows users down as they pause to process and respond to the

10.3389/frvir.2025.1499830

TABLE 4 Questionnaire results of NASA TLX and SFS indicating relatively high levels of cognitive load (Prabaswari et al., 2019) and flow (Tadeja et al., 2021a) experienced by the participants.

Participant	TLX		SFS flow		SFS anxiety	
#	[0–100]		[1–7]		[1–7]	
Automatic Feedback	*no*	*yes*	*no*	*yes*	*no*	*yes*
P1	55.33	67.00	3.90	2.80	2.50	3.50
P2	75.00	74.00	3.80	3.90	3.00	4.00
P3	51.00	51.00	7.00	6.80	6.00	4.50
P4	48.33	70.00	3.80	4.40	3.50	3.00
P5	49.00	63.00	4.70	4.20	5.00	3.50
P6	54.67	48.00	6.30	5.40	3.00	2.50
P7	62.00	56.33	5.90	6.70	6.50	7.00
P8	36.67	39.33	6.10	5.60	4.50	2.50
P9	50.00	62.33	3.40	3.50	2.00	2.50
P10	59.67	58.00	4.80	3.60	3.50	3.50
P11	36.00	50.00	4.10	4.60	4.50	5.00
P12	63.33	66.67	5.20	4.60	3.00	3.00
P13	44.00	62.67	5.90	4.50	4.50	4.50
P14	72.33	35.67	5.10	5.40	3.50	2.50
P15	48.00	47.67	4.40	4.70	3.50	3.50
P16	46.00	76.33	5.70	4.80	3.50	2.50
M	53.21	58.00	5.00	4.72	3.88	3.6
SD	10.78	11.53	1.03	1.05	1.18	1.18
p-value	0.249		0.454		0.518	

feedback. This may particularly affect less experienced participants, as evidenced by the larger spread of task times when feedback was active.

In general, the analysis of task completion times suggests that while potentially helpful in improving accuracy and fold quality, the feedback system introduces additional overhead that extends task duration in both experimental setups. Thus, repeating the origami folding task under the automated feedback was still slower than when ML-based verification was not used. Participants may have needed to adjust their folding strategy or revise their actions based on the system's guidance, leading to longer times. The variance in task duration (as indicated by the standard deviations) also points to differing levels of dependency on the feedback, with some participants taking longer to fold when the system was in place.

Overall, these findings indicate that while feedback is useful, its integration should aim to balance guidance with efficiency, perhaps by allowing more experienced users to skip certain steps or streamline the validation process, especially when they previously folded the same or similar models. It also underlines that to be a robust alternative to a no-feedback system, further optimization in both hardware and software apparatus is required.

5.2 Questionnaires results

The *SFS Flow* scores reflect the degree to which participants felt immersed and engaged in the task (Table 4). For trials with and without a feedback, the Shapiro-Wilk tests showed no significant departure from normality, hence ANOVA was performed. Without feedback, participants reported, on average, slightly higher flow ($M = 5.00$) than with feedback ($M = 4.72$), although this difference was also not significant ($F(1.0, 30.0) = 0.575, \eta_p^2 = .019, \quad p = 0.454$). The slight drop in flow when using feedback may indicate that the system's interventions occasionally interrupted the natural flow of the task, especially for more experienced users who might prefer a smoother, uninterrupted folding process. For less experienced participants, the feedback may have introduced helpful pauses that did not significantly detract from their engagement. At the same time, in both cases, the participants reported relatively high flow levels, with P1 and P3 reporting the lowest and highest scores of 40% and 97% under the feedback condition, respectively. This demonstrates the participants' engagement while using our AR system (Tadeja et al., 2021a).

The *SFS Anxiety* scores reveal how anxious participants felt while completing the task. Interestingly, participants reported lower anxiety when feedback was active ($M = 3.6$) compared to no feedback ($M = 3.88$). The Shapiro-Wilk test indicated a departure from normality for the anxiety score in trials with feedback. Consequently, an ANOVA could not be performed. Instead, a Mann-Whitney U test was conducted, which showed no statistically significant difference in the anxiety level ($U = 107.0, \quad p = 0.431$). The slightly lower anxiety under feedback condition, with seven participants reporting a drop in scores in comparison with no feedback (Table 4) suggests that participants may have felt more confident in their folding accuracy due to the system's validation, even if the validation process increased cognitive workload. This aligns with anecdotal feedback, where some participants mentioned feeling more comfortable and confident using the validation system.

The NASA TLX (Hart and Staveland, 1988) shown in Table 4 are inconsistent with data drawn from the time analysis as the cognitive load experienced during folding origami with and without machine learning-based feedback was roughly similar. Overall, *NASA TLX* scores indicate the participants' perceived mental workload as relatively high in all cases and conditions, ranging from "somewhat high" [30 – 49] to "high" [50 – 79] (Prabaswari et al., 2019) with both the lowest of 35.67/100 (P14) and the highest score of 76.33/100 (P16) experienced with active feedback. The Shapiro-Wilk tests indicated that there was no statistically significant departure from normality. On average, the workload was slightly higher when using feedback ($M = 58.00$) compared to no feedback ($M = 53.21$). Analysis of variance (ANOVA) revealed that this difference was not statistically significant ($F(1.0, 30.0) = 1.382, \eta_p^2 = .044, \quad p = 0.249$). The higher workload when feedback was present suggests that participants found it cognitively more demanding to integrate and respond to the system's instructions. Moreover, the participants rated their "performance" contributing more towards the overall cogitative load in the no feedback condition (Figure 8), which suggests that feedback gave them more comfort in assessing their own work.

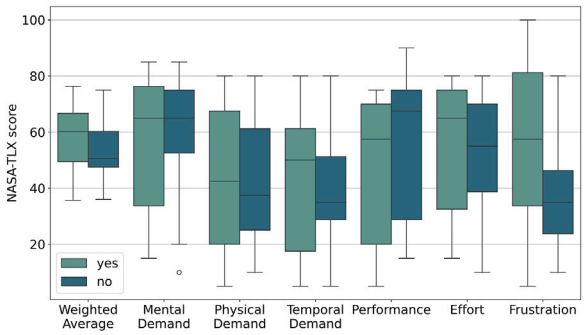

FIGURE 8
The weighted (leftmost) average and "raw" NASA TLX scores using box plots with lower scores indicating better outcomes. The most visible difference is the higher perceived temporal demand and frustration when folding the origami models with automated feedback.

However, the fact that this increase was not significant suggests that the feedback system added some complexity but not enough to overwhelm most of the users.

From the analysis of the questionnaires, we draw conclusions that are threefold. First, the feedback system added a limited amount of cognitive load, as indicated by the higher NASA TLX scores when the system was used with feedback. While the increase was not significant, it shows that users needed to devote more mental effort to follow the system's guidance, possibly due to interruptions in their natural workflow. At the same time, the users may feel more frustrated when using the feedback during task repetition. Second, the slight reduction in flow with feedback may indicate that the validation process disrupted the seamless progression of the task. More experienced users might have been affected by the system's frequent checks, which interrupted their folding rhythm. Third, the system seemed to reduce anxiety among a considerable number of participants, as they potentially felt reassured by the feedback. This finding suggests that while the feedback may slow down the task, it provides valuable reassurance, which can improve the user experience.

Overall, the questionnaire results suggest that while the feedback system does introduce some cognitive overhead and can interrupt flow, it may also positively contribute to users' confidence by reducing anxiety. To further improve the system, adjustments could be made to streamline the feedback process, reducing its impact on cognitive load and flow while maintaining its beneficial effects on user confidence and overall experience.

5.3 Machine learning feedback accuracy

We present in Table 5 average folding accuracy scores across four different origami models: (A) *Samurai Hat*, (B) *Box*, (C) *Fly*, and (D) *Yacht*. We populated Table 5 based on the experimental outcomes under the condition in which the participants were assisted with a machine-learning-based validation mechanism (feedback) integrated with the AR interface. Three evaluators non-directly involved in the capturing of experimental data conducted the *accuracy and aesthetics* assessment, assigning a score from the set of $\{0, 1, 2\}$, with $\{2\}$ representing the satisfactorily folded model, $\{1\}$ denoted semi-correctly folded model and $\{0\}$ denoted non-complete or wrongly folded model. These results illustrate how effective the automatic validation process was in ensuring the accuracy of the folded models.

We can observe a general trend where more origami models requiring more folds tend to lead to lower accuracy. However, the latter is not strictly dependent on the number of folds, as models with fewer stages and more intricate folds like the (A) *Samurai Hat* can still result in lower accuracy. This finding is consistent with prior work indicating that both the number of steps (e.g., folds) and their difficulties can contribute to the overall complexity of AR-based guidance (Bock et al., 2024). In terms of the origami model, we can observe the following.

- (A) *Samurai Hat*: the average scored accuracy was 1.33. Despite having a moderate number of folding stages, participants show the lowest level of performance.
- (B) *Box*: the average accuracy for this origami model was 1.40. This model has more stages, which increases the likelihood of errors, but the accuracy is comparable to simpler models like the (C) *Fly* or (D) *Yacht*. This suggests that participants may struggle similarly with models of different complexity levels due to specific challenges rather than just the number of stages.
- (C) *Fly*: the accuracy for this origami model was 1.44, which is higher than that of both (A) *Samurai Hat* and (B) *Box*. While this model has more folding stages than *Samurai Hat*, its design does not have folds that are complicated and hard to demonstrate on the animations (like tucking the paper inside), which improves the overall accuracy.
- (D) *Yacht*: with the fewest number of stages, this origami model obtained high average accuracy of 1.54. This suggests that simpler models are easier to execute with higher precision, as fewer steps result in fewer opportunities for errors.

What is worth noticing is that longer task completion times, as seen in Table 3, may lead to slightly better accuracy scores. For instance, participants who scored 2.00 on the (A) *Samurai Hat* (P6) or (D) *Yacht* (P1) took more time to fold the models, indicating that taking more time to complete the task could result in fewer errors. Furthermore, the NASA TLX scores shown in Table 4 suggest low to no correlation with accuracy scoring as indicated by Pearson correlation coefficient of $r = 0.26$. This suggests that the cognitive load experienced by the participants did not influence the accuracy of their origami folds.

5.4 Users' feedback, comments and observations

The feedback from 16 participants who used our AR-based system for origami folding provides insights into the strengths and areas for improvement in the user experience. The comments have been organized by theme, highlighting various aspects of the system's performance. We also discuss the observations of participants' respective behaviors made during the experiments.

Feedback on the validation process. Several users raised concerns regarding the validation process. P1 expressed a

TABLE 5 Accuracy assessment by three evaluators for each participant. Rows represent origami models and their corresponding averages. Columns represent participants (P1-P16) with their average scores.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	Avg
(A)	1.00	1.00	1.33	2.00	1.33	2.00	0.00	0.00	2.00	1.67	0.00	1.33	2.00	1.67	2.00	2.00	1.33
(B)	1.00	1.67	1.33	1.33	1.00	2.00	1.00	1.00	1.33	1.67	2.00	1.00	2.00	1.00	2.00	1.00	1.40
(C)	1.33	2.00	2.00	0.67	2.00	1.00	1.00	1.00	2.00	2.00	0.67	0.67	1.67	1.00	2.00	2.00	1.44
(D)	2.00	1.67	2.00	1.00	1.00	1.67	0.67	1.00	2.00	2.00	1.67	1.67	1.33	1.00	2.00	2.00	1.54
Avg	1.33	1.58	1.67	1.25	1.33	1.67	0.67	0.75	1.83	1.83	1.08	1.17	1.75	1.17	2.00	1.75	

desire to see both the initial and final stages of a fold, preferring to play the animation only when needed. He also felt that the validation process should focus on key milestones rather than every minor step. P1 also noted a lack of trust in the system, stating that when the machine flagged an error, he did not question his own actions but simply retried the step repeatedly. P9, P13 and P15 echoed this sentiment, finding that the validation would be helpful but did not meet expectations in its current form. On the other hand, P14 felt more comfortable and confident using the validation system, suggesting that it provided a positive sense of guidance and facilitated a period of contemplation during the process.

Some participants suggested that the validation process should be optional or more flexible. P3 and P11, both with experience in origami, suggested that users should be able to skip certain steps, particularly smaller, repetitive actions. P11 also found the validation frustrating, as he felt it slowed down experienced users who already knew how to fold the model.

Visual and interaction challenges Several users highlighted visual issues with the system. P4 noted that it was difficult to distinguish between one part of a folded paper and overlapping sections. P11, while experienced in origami, found it challenging to see fold lines clearly, which made it hard to assess whether a fold was done correctly.

Participants also noted interaction challenges, particularly with the buttons used for validation. P6 described the button as "not responsive," while P7 suggested that a built-in validation system would be more intuitive, allowing users to check their work without needing to move their hands or head excessively. Participant P13 indicated that the verification step was less physically comfortable and proposed that it be executed automatically, obviating the need to use a button.

Animation and user preferences Several users provided suggestions regarding the animation and user interface. P2 and P15 requested a clearer indication of where folds should occur, suggesting the inclusion of dotted lines or visual guides on the paper itself. P1 that the system should display both the initial and final stages of a fold, with P1 emphasizing that animations should only play when necessary. P15 also recommended unifying symmetric folds to improve clarity.

In terms of animation control, P5 and P14 preferred automatic animation, while P3 and P11 expressed a desire to skip steps if needed. Moreover, P2 hoped for user interface elements that more clearly indicated where each fold should be made, further enhancing the system's guidance.

5.4.1 Observations of participants' behaviors

Nearly all participants made errors in the final step of the model (A) *Samurai Hat*, which involves tucking paper inside the model. This may be attributed to the animation, which does not clearly demonstrate this action. Furthermore, when users fail the validation too many times, they tend to lose trust in the system, believing it is malfunctioning rather than recognizing their own mistakes. Also, some participants attempted multiple consecutive steps at once, which frequently led to validation errors. In addition, participants varied in their folding techniques. Some folded the paper while holding it in the air, while others validated the steps by holding the paper in their hands. Occasionally, validation was performed with the paper pressed flat on the desk. Moreover, in non-validation (no feedback) mode, some participants watched several steps ahead before beginning to fold and then attempted to complete these steps all at once.

6 Discussion and further work

The findings from this study reveal key insights into the performance and usability of our AR-based origami folding system, highlighting several areas for improvement. Central to these observations is the interplay between system reliability, user trust, and task complexity, which significantly influenced user engagement and accuracy.

A recurring issue was the loss of trust in the validation system due to repeated failures. Participants who experienced multiple validation errors attributed these issues to the system, leading to frustration and disengagement. This highlights a critical relationship between feedback reliability and cognitive load. As shown by the increased perceived cognitive load (Figure 8), unreliable feedback not only undermines user confidence but also diminishes the system's intended utility. To address this, future iterations must prioritize consistent feedback mechanisms that enhance trust and foster sustained engagement.

The observed challenges in the interaction between users and the validation system, such as improper handling of paper during validation and suboptimal camera positioning, underscore the importance of user-centered design in AR applications. The HL2 camera's positioning above the user's eyes often resulted in misaligned or incomplete views, particularly when users looked down. This observation emphasizes the need for dynamic camera calibration or augmented detection algorithms capable of compensating for varied perspectives, ensuring robust performance regardless of user behavior.

The complexity of the models, measured by the number of folding stages, had an impact on accuracy. Simpler models, such as the (D) *Yacht*, which had fewer stages, tended to result in higher accuracy scores. In contrast, more complex models like the (B) *Box* and (C) *Fly*, with a greater number of stages and more intricate folds, led to lower accuracy. This suggests that model complexity should be assessed not just by the number of steps but by the intricacy of individual folds, as errors often occur during symmetric or multi-fold steps. Moreover, the tendency of users to skip ahead during folding underscores the need for adaptive guidance that aligns with individual user preferences and expertise levels.

Scalability also emerged as a consideration for the system, given the need to acquire real images for each origami model and step. To address this, we streamlined the annotation process using the Segment Anything model, significantly reducing manual effort. While this approach improved efficiency, synthetic datasets generated from the animation system could further reduce data requirements. However, synthetic images may not fully capture the complexities of real-world data, potentially reducing model performance in real-world scenarios. Combining them with real-world data and applying techniques like domain adaptation is often necessary to ensure practical effectiveness (Man and Chahl, 2022).

This study acknowledges several limitations that could influence the effectiveness and utility of the proposed AR-based origami folding system.

First, while the NASA TLX is widely used, it has limitations in capturing cognitive load nuances in AR contexts (McKendrick and Cherry, 2018). Future work will incorporate physiological measures, such as pupillary data, to better quantify cognitive demands (Chen et al., 2011).

Second, the feedback system provided limited information about fold correctness, potentially reducing its utility. Adding detailed error feedback (e.g., mistake types and corrective steps) could enhance user experience but requires research into optimal delivery formats (e.g., audio, imagery, text, animations). Additionally, the interplay between feedback mechanisms, the AR interface, and device design may impact usability. Future studies will explore these factors and alternative design approaches.

Moreover, discrepancies between training data and real-world conditions could affect model performance. While data augmentation was used, the training set lacked variability in hand visibility, background clutter, and object positioning, with all origami samples presented flat on a table. This domain shift might reduce feedback accuracy and user trust. Collecting data during real folding sessions with diverse users could address this but requires significant resources. Future efforts should focus on gathering realistic datasets and evaluating model robustness under varied conditions.

A limitation is also the system's sensitivity to domain shifts, as the training data primarily comprised controlled conditions with uniform backgrounds and consistent lighting. Real-world environments with varied lighting, cluttered backgrounds, and diverse user behaviors (e.g., folding styles and hand positioning) could degrade feedback accuracy. Future work should prioritize domain adaptation techniques and incorporate more diverse, realistic datasets to enhance model robustness and maintain performance across different contexts.

Our findings suggest that while the validation system shows promise, several improvements are needed. These include better visual clarity (e.g., clearer fold lines and paper layer differentiation during complex folds), more flexible validation (e.g., allowing step skipping or milestone focus for advanced users), and improved interaction design, including optimized camera positioning and button layouts to minimize errors.

Finally, the small sample size limits generalizability but offers valuable preliminary insights. Larger, more diverse participant groups in future studies will help validate and refine the system.

Addressing these issues will increase user trust and usability, balancing task complexity, time, and cognitive load to improve performance in AR-supported origami folding.

7 Conclusion

This article aims to validate the ability of an ML-supported AR application to enhance hands-on learning of a manual task. We investigated this through a rigorous evaluation of its usability and impact on learning outcomes. The findings contribute to the broader understanding of how AR and ML can be leveraged to support complex manual tasks in educational and professional settings.

Our work illustrates that AR support, which provides immersive and interactive experiences for training in various domains, can be augmented by a machine learning model to automate task evaluation upon completion. Similar to the recent work of Chen et al. (2023), where origami serves as a proxy task for studying the potential of teaching intricate manual skills, our approach highlights the broader applicability of AR-enhanced systems to hands-on tasks such as manual assembly or maintenance processes. While promising, the model may still lack accuracy and robustness, particularly when dealing with variations in incorrectly folded patterns. Future efforts will focus on improving the model's resilience to these variations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The University Research Ethics Committee (UREC), University of Cambridge, England. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

ML: Data curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing–original draft, Writing–review

and editing. JG: Data curation, Software, Visualization, Writing–original draft, Writing–review and editing. CL: Data curation, Investigation, Software, Visualization, Writing–original draft, Writing–review and editing. TB: Funding acquisition, Project administration, Resources, Validation, Writing–review and editing. ST: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Validation, Writing–original draft, Writing–review and editing. PS: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Poznan University of Technology grant 0214/SBAD/0248. The study was supported by funding provided through an unrestricted gift by Meta. Work of Piotr Skrzypczyński and the publication costs were funded from PUT internal grant 0214/SBAD/0248.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agui, T., Takeda, M., and Nakajima, M. (1983). Animating planar folds by computer. *Comput. Vis. Graph. Image Process.* 24, 244–254. doi:10.1016/0734-189X(83)90046-4

Andreass, B. (2011). Origami art as a means of facilitating learning. *Procedia - Soc. Behav. Sci.* 11, 32–36. doi:10.1016/j.sbspro.2011.01.028

Bock, L., Bohné, T., and Tadeja, S. K. (2024). Decision support for augmented reality-based assistance systems deployment in industrial settings. *Multimedia Tools Appl.* doi:10.1007/s11042-024-19861-x

Bozzi, L. O. S., Samson, K. D. G., Tadeja, S., Pattinson, S., and Bohné, T. (2023). "Towards augmented reality guiding systems: an engineering design of an immersive system for complex 3D printing repair process," in *2023 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, 384–389. doi:10.1109/VRW58643.2023.00084

Chen, K., Li, T., Kim, H.-S., Culler, D. E., and Katz, R. H. (2018). "Marvel: enabling mobile augmented reality with low energy and low latency," in *Proceedings of the ACM conference on embedded networked sensor systems (SenSys)*.

Chen, Q., Mishra, R., El-Zanfaly, D., and Kitani, K. (2023). "Origami sensei: mixed reality ai-assistant for creative tasks using hands," in Companion Publication of the 2023 ACM designing interactive systems conference. *DIS '23 companion*, 147—–151. doi:10.1145/3563703.3596625

Chen, S., Epps, J., and Chen, F. (2011). "A comparison of four methods for cognitive load measurement," in *Proceedings of the 23rd Australian computer-human interaction conference* (Canberra, Australia: OzCHI '11), 76—–79. doi:10.1145/2071536.2071547

Danielsson, O., Holm, M., and Syberfeldt, A. (2020). Augmented reality smart glasses in industrial assembly: current status and future challenges. *J. Industrial Inf. Integration* 20, 100175. doi:10.1016/j.jii.2020.100175

Dudley, J. J., Schuff, H., and Kristensson, P. O. (2018). "Bare-handed 3d drawing in augmented reality," in *Proceedings of the 2018 designing interactive systems conference* (New York, NY, USA: Association for Computing Machinery), DIS '), 18, 241–252. doi:10.1145/3196709.3196737

EasyAR (2024). EasyAR MEGA library. Available at: https://www.easyar.com/.

Engeser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation Emot.* 32, 158–172. doi:10.1007/s11031-008-9102-4

Farasin, A., Peciarolo, F., Grangetto, M., Gianaria, E., and Garza, P. (2020). Real-time object detection and tracking in mixed reality using microsoft hololens. *15th Int. Jt. Conf. Comput. Vis. Imaging Comput. Graph. Theory Appl.* 4, 165–172. doi:10.5220/0008877901650172

Goka, R., Ueda, K., Yamaguchi, S., Kimura, N., Iseya, K., Kobayashi, K., et al. (2022). "Development of tomato harvest support system using mixed reality head mounted display," in *IEEE 4th global conference on life sciences and technologies*, 167–169. doi:10.1109/LifeTech53646.2022.9754831

Grandhi, U., and Chang, I. Y. (2019). "PlayGAMI: augmented reality origami creativity platform," in *ACM SIGGRAPH 2019 appy hour* (Los Angeles, CA: SIGGRAPH '19). doi:10.1145/3305365.3329729

Hart, S., and Staveland, L. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.*, 139–183. doi:10.1016/s0166-4115(08)62386-9

Herbas Torrico, B. C. (2021). "Teaching work measurement through origami in developing countries," in *2nd south American international conference on industrial engineering and operations management*.

Jocher, G., Qiu, J., and Chaurasia, A. (2023). Ultralytics YOLO.

Johri, A., Sayal, A., N, C., Jha, J., Aggarwal, N., Pawar, D., et al. (2024). Crafting the techno-functional blocks for metaverse - a review and research agenda. *Int. J. Inf. Manag. Data Insights* 4, 100213. doi:10.1016/j.jjimei.2024.100213

Khandelwal, P., Srinivasan, K., and Roy, S. S. (2019). "Surgical education using artificial intelligence, augmented reality and machine learning: a review," in *IEEE international conference on consumer electronics - taiwan (ICCE-TW)*, 1–2.

Kim, M., Lee, K., Balan, R., and Lee, Y. (2023). "Bubbleu: exploring augmented reality game design with uncertain AI-based interaction," in *Proceedings of the 2023 CHI conference on human factors in computing systems*.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). Segment anything. *arXiv* 2304.02643

Laakasuo, M., Palomäki, J., Abuhamdeh, S., Lappi, O., and Cowley, B. U. (2022). Psychometric analysis of the flow short scale translated to Finnish. *Sci. Rep.* 12, 20067. doi:10.1038/s41598-022-24715-3

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C., et al. (2015). SSD: single shot multibox detector. *Corr. abs/1512*, 02325. doi:10.48550/arXiv.1512.02325

Łysakowski, M., Gapsa, J., and Skrzypczyński, P. (2024). "Origami unfolded: augmented reality and machine learning integration for manipulative training scenarios," in A Metaverse for the good: the European metaverse research network conference *(barcelona)*.

Łysakowski, M., Żywanowski, K., Banaszczyk, A., Nowicki, M. R., Skrzypczyński, P., and Tadeja, S. (2023a). "Using AR and YOLOv8-based object detection to support real-world visual search in industrial workshop: lessons learned from a pilot study," in *Ieee int. Symp. On mixed and augmented reality adjunct (ISMAR-Adjunct)*, 154–158.

Łysakowski, M., Żywanowski, K., Banaszczyk, A., Nowicki, M. R., Skrzypczyński, P., and Tadeja, S. K. (2023b). Real-time onboard object detection for augmented reality: enhancing head-mounted display with YOLOv8. *IEEE Int. Conf. Edge Comput. Commun.*, 364–371. doi:10.1109/edge60047.2023.00059

Malek, K., Mohammadkhorasani, A., and Moreu, F. (2022). Methodology to integrate augmented reality and pattern recognition for crack detection. *Computer-Aided Civ. Infrastructure Eng.* 37, 43–56. doi:10.1111/mice.12932

Man, K., and Chahl, J. (2022). A review of synthetic image data and its use in computer vision. *J. Imaging* 8, 310. doi:10.3390/jimaging8110310

McKendrick, R. D., and Cherry, E. (2018). A deeper look at the nasa tlx and where it falls short. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 62, 44–48. doi:10.1177/1541931218621010

Monnier, R., and Winters, L. (2022). Playing in limbo. *J. Play Adulthood* 4, 32–49. doi:10.5920/jpa.1022

Palmarini, R., Erkoyuncu, J. A., Roy, R., and Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. *Robotics Computer-Integrated Manuf.* 49, 215–228. doi:10.1016/j.rcim.2017.06.002

Pauly, O., Diotte, B., Fallavollita, P., Weidert, S., Euler, E., and Navab, N. (2015). Machine learning-based augmented reality for improved surgical scene understanding. *Comput. Med. Imaging Graph.* 41, 55–60. doi:10.1016/j.compmedimag.2014.06.007

Prabaswari, A. D., Basumerda, C., and Utomo, B. W. (2019). The mental workload analysis of staff in study program of private educational organization. *IOP Conf. Ser. Mater. Sci. Eng.* 528, 012018–018. doi:10.1088/1757-899x/528/1/012018

PTC (2023). Vuforia engine library. Available at: https://library.vuforia.com/.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: unified, real-time object detection. *CVPR*, 779–788. doi:10.1109/cvpr.2016.91

Ren, S., He, K., Girshick, R. B., and Sun, J. (2015). "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in neural information processing systems 28: annual conference on neural information processing systems 2015, december 7-12, 2015, Montreal, quebec, Canada*. Editors C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 91–99.

Shimanuki, H., Watanabe, T., Asakura, K., Sato, H., and Ushiama, T. (2020). Anomaly detection of folding operations for origami instruction with single camera. *IEICE Trans. Inf. Syst. 103-D* E103.D, 1088–1098. doi:10.1587/transinf.2019edp7242

Suhail, N., Bahroun, Z., and Ahmed, V. (2024). Augmented reality in engineering education: enhancing learning and application. *Front. Virtual Real.* 5. doi:10.3389/frvir.2024.1461145

Supple, B., O'Neill, S., Pentek, A., and Hao, G. (2021). Beyond paper folding: origami and focused play to enhance interdisciplinary learning and teaching in universities. *All Irel. J. High. Educ.* 13. doi:10.62707/aishej.v13i3.591

Tadeja, S. K., Lu, Y., Rydlewicz, M., Rydlewicz, W., Bubas, T., and Kristensson, P. O. (2021a). Exploring gestural input for engineering surveys of real-life structures in virtual reality using photogrammetric 3D models. *Multimedia Tools Appl.* 80, 31039–31058. doi:10.1007/s11042-021-10520-z

Tadeja, S. K., Rydlewicz, W., Lu, Y., Bubas, T., Rydlewicz, M., and Kristensson, P. O. (2021b). Measurement and inspection of photo-realistic 3-D VR models. *IEEE Comput. Graph. Appl.* 41, 143–151. doi:10.1109/MCG.2021.3114955

Watanabe, T., and Kinoshita, Y. (2012). "Folding support for beginners based on state estimation of origami," in *TENCON 2012 IEEE region 10 conference*, 1–6. doi:10.1109/TENCON.2012.6412167

Wiwatwattana, N., Laphom, C., Aggaitchaya, S., and Chattanon, S. (2016). "Origami guru: an augmented reality application to assist paper folding," in *Information Technology: new generations* (Springer), 1101–1111.

Yaseen, M. (2024). What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector

Yang, J., Gao, M., Li, Z., Gao, S., Wang, F., and Zheng, F. (2023). Track anything: segment anything meets videos. *arXiv, cs.cv*. doi:10.48550/arXiv.2304.11968

Zambri, A. A., and Kamaruzaman, M. F. (2020). "The integration of augmented reality (ar) in learning environment," in *2020 sixth international conference on e-learning (econf)*, 194–198. doi:10.1109/econf51404.2020.9385487

Zhao, F., Gaschler, R., Kneschke, A., Radler, S., Gausmann, M., Duttine, C., et al. (2020). Origami folding: taxing resources necessary for the acquisition of sequential skills. *PLoS One* 15, e0240226. doi:10.1371/journal.pone.0240226

Zogopoulos, V., Birem, M., De Geest, R., Hofman, R., Jorissen, L., Vanherle, B., et al. (2021). Image-based state tracking in augmented reality supported assembly operations. *Procedia CIRP* 104, 1113–1118. doi:10.1016/j.procir.2021.11.187

Zonaphan, L., Northus, K., Wijaya, J., Achmad, S., and Sutoyo, R. (2022). Metaverse as A Future of education: a systematic review. *2022 8th Int. HCI UX Conf. Indonesia (CHIuXiD)* 1, 77–81. doi:10.1109/CHIuXiD57244.2022.10009854