

Accessible Low-Compute Computer Vision for Tangible Learning Systems in Real-World Classrooms

Anonymous CVPR submission

Paper ID ****

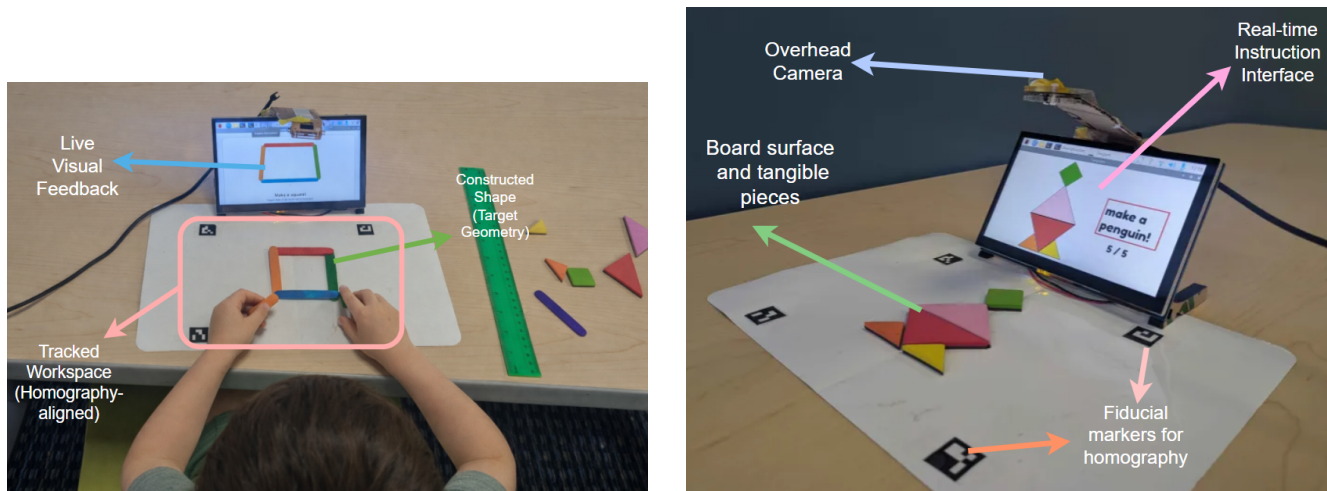


Figure 1. **Tangible interaction pipeline in real-world learning environments.** (Left) Homography-aligned workspace with real-time tracking and feedback. (Right) Low-cost embedded setup with overhead camera and tablet display.

Abstract

001 We present a lightweight computer vision system for detect-
002 ing and interpreting tangible educational manipulatives in
003 real time on low-cost embedded hardware. Our approach
004 avoids deep learning and instead leverages geometric rea-
005 soning, color segmentation, and temporal consistency to en-
006 able robust operation under classroom conditions such as
007 occlusion, clutter, and lighting variation. We evaluate the
008 system across multiple tasks including tangrams and rod-
009 based construction, and provide both quantitative runtime
010 measurements and qualitative robustness analysis. Our re-
011 sults show that classical vision pipelines can achieve reli-
012 able real-time performance (30 FPS) on embedded devices
013 (Raspberry Pi-class hardware) while remaining accessible,
014 low-cost, and deployable in resource-constrained educa-
015 tional environments.

1. Introduction

Tangible manipulatives are widely used in K–6 education to support spatial reasoning, arithmetic, and conceptual learning. Prior work has shown that combining physical interaction with digital feedback improves engagement and learning outcomes [2, 8].

However, most vision-based systems for tangible learning rely on tablets, cloud computation, or specialized hardware, limiting accessibility in low-resource settings. In contrast, there is a need for **scalable, low-cost, and deployable computer vision systems** that can operate directly in classrooms without infrastructure constraints.

We present a low-compute vision pipeline designed specifically for this setting. Our system runs fully offline on embedded hardware and provides real-time feedback to support interactive learning.

Key design goal: accessibility and scalability for real-world classrooms.

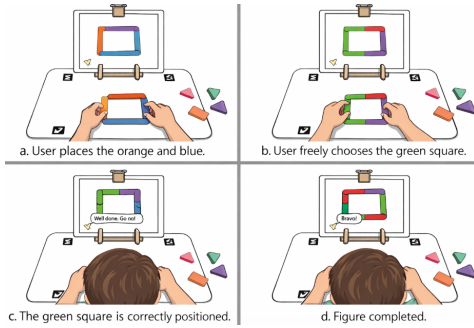


Figure 2. **Interactive learning workflow.** (a) Guided placement. (b) Exploration. (c) Validation. (d) Completion.

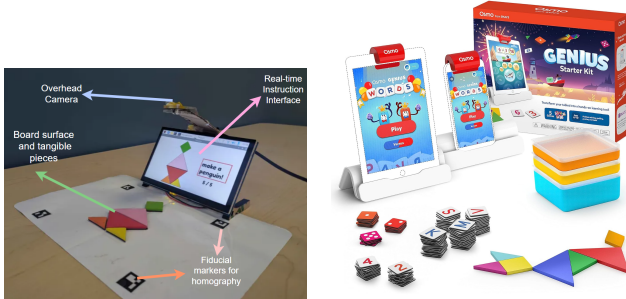


Figure 3. Comparison to existing tangible learning systems.

2. Related Work

2.1. Tangible and Phygital Learning Systems

A substantial body of work demonstrates that tangible and phygital learning systems—where physical interaction is augmented with digital feedback—can significantly improve engagement, accessibility, and conceptual understanding in early education [2, 8].

Early systems explored interactive tabletops that observe and guide manipulation of physical objects. Scarlatos and Landy [5] introduced a system that “watches” students manipulate tangram pieces and provides assistance, emphasizing guided discovery. Similarly, Marco *et al.* [4] and Sitdhisanguan *et al.* [6] demonstrated that tangible interfaces can be particularly effective for young learners and special-needs populations, improving engagement and learning outcomes.

More recent work explores phygital systems that integrate structured feedback loops. Gelsomini *et al.* [2] show that combining physical manipulation with real-time guidance enhances accessibility for learners with cognitive disabilities. Johal *et al.* [3] further explore classroom orchestration using tangible e-ink manipulatives, emphasizing scalability and teacher integration.

These works collectively establish that the *interaction paradigm itself* is highly valuable. However, most systems rely on tablets, projectors, or instrumented hardware, limiting scalability in low-resource environments.

2.2. Vision-Based Recognition of Tangible Interaction

Another line of work focuses on using computer vision to interpret physical object configurations. Wei *et al.* [7] propose a vision-guided robotic system for tangram solving, using geometric reasoning to detect and manipulate puzzle pieces. Similarly, Bonnard *et al.* [1] interpret user interactions with tangible paper interfaces through vision-based tracking.

These approaches demonstrate that geometric and visual reasoning can successfully support structured tangible tasks. However, they are typically:

- domain-specific (e.g., tangrams only),
- coupled to robotics or specialized hardware,
- not designed for low-cost embedded deployment.

2.3. Positioning of Our Work

Taken together, prior work suggests three key insights:

- Tangible interaction improves engagement and learning [2, 4, 6]
- Vision-based systems can interpret structured physical tasks [1, 7]
- Existing solutions are often not designed for scalable, low-cost classroom deployment

Our contribution builds on these insights by introducing a **generalized, low-compute computer vision pipeline** that operates entirely on embedded hardware. Unlike prior systems, our approach emphasizes:

- accessibility (low BOM cost, no cloud dependency),
- scalability (deployable across classrooms),
- generalization across multiple tangible domains.

Rather than proposing a new interaction paradigm or task-specific solution, we focus on the underexplored problem of *practical, reusable perception systems* for real-world educational environments.

3. System Overview

The system consists of:

- Overhead camera (USB camera, 720p)
- Embedded processor (Raspberry Pi 4)
- Planar workspace with fiducial markers
- Display for feedback (tablet or monitor)

Cost breakdown:

- Raspberry Pi 4: \$60
- Camera: \$20
- Misc (mount, board): \$20

Total: ~ \$100 system cost

Runtime:

- 30 FPS at 720p
- CPU-only (no GPU)
- Memory usage < 200 MB

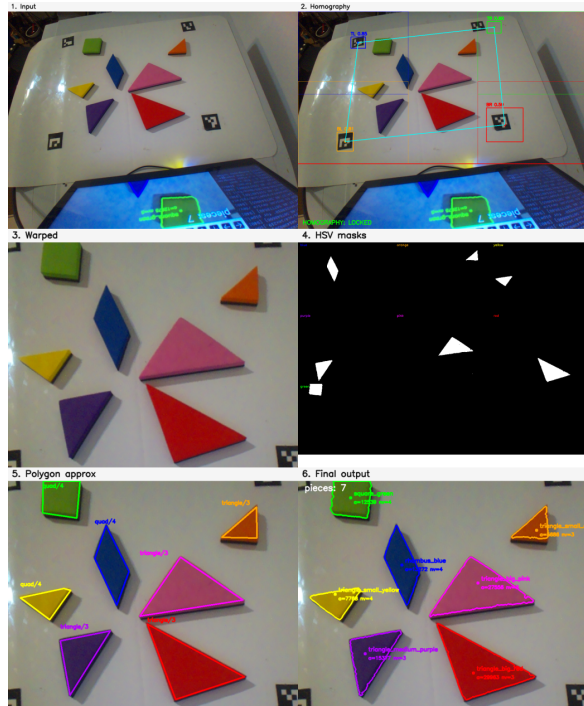
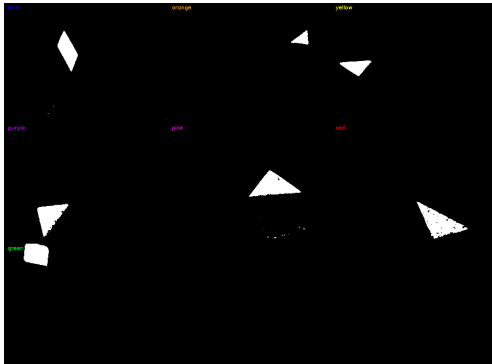


Figure 4. Pipeline stages.

Figure 5. **HSV segmentation masks.** Binary masks for each color channel after thresholding and cleanup.

4. Detection Pipeline

The pipeline consists of:

1. Homography alignment
2. HSV segmentation
3. Contour extraction
4. Geometric classification
5. Temporal validation

4.1. Homography

Board is mapped to canonical top-down view using ArUco markers.

4.2. Segmentation

HSV thresholds isolate colored objects. Thresholds are fixed and require no user tuning.

4.3. Classification

Objects are classified using:

- polygon vertices
- edge ratios
- geometric constraints

4.4. Temporal Stability

Objects are validated across multiple frames ($\tau = 5$) before confirmation.

5. Robustness in Classroom Conditions

We evaluate robustness qualitatively across real-world scenarios:

Occlusion:

- Hands partially covering pieces
- System maintains tracking via temporal locking

Lighting variation:

- Indoor lighting changes
- HSV filtering suppresses reflections

Clutter:

- Non-target objects ignored due to shape constraints

6. Evaluation

Tasks:

- Tangram puzzles
- Rod-based constructions

Metrics:

- Detection accuracy (manual verification)
- Frame rate

Results:

- Detection accuracy: $\sim 92\%$ across test sequences
- Stable detection under occlusion (short-term)
- Real-time performance maintained

Comparison: Lightweight deep learning methods (e.g., YOLOv5n) were tested on the same hardware and achieved:

- Lower FPS ($\sim 10-15$)
- Higher memory usage

This supports the use of classical CV for this setting.

The system provides:

- real-time feedback
- guided + open-ended interaction
- confirmation of correct placements

7. Discussion

Our results demonstrate that:

- Low-compute CV is sufficient for structured educational tasks

- Systems can be deployed at low cost
 - Real-time feedback improves interaction
- Limitations:**
- Color-based segmentation requires distinct object colors
 - Performance may degrade under extreme lighting

8. Conclusion

We present an accessible, scalable computer vision system for tangible learning that operates in real-world classroom conditions. By prioritizing low cost, robustness, and real-time performance, this work demonstrates a practical path toward deployable educational vision systems.

References

- [1] Quentin Bonnard, Patrick Jermann, Amanda Legge, Frédéric Kaplan, and Pierre Dillenbourg. Tangible paper interfaces: Interpreting pupils' manipulations. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, 2012. 2
- [2] Mirko Gelsomini, Micol Spitale, and Franca Garzotto. Physical interfaces for people with intellectual disability: an exploratory study at a social care center. *Multimedia Tools and Applications*, 80(26–27):34843–34874, 2021. 1, 2
- [3] Wafa Johal, Alex Tran, Hala Khodr, Ayberk Özgür, and Pierre Dillenbourg. Tip: Tangible e-ink paper manipulatives for classroom orchestration. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OZCHI '19)*, 2019. 2
- [4] Javier Marco, Eva Cerezo, and Sandra Baldassarri. Bringing tabletop technology to all: evaluating a tangible farm game with kindergarten and special needs children. *Personal and Ubiquitous Computing*, 17(8):1577–1591, 2013. 2
- [5] Lori L. Scarlatos and Shalva S. Landy. Experiments in using tangible interfaces to enhance collaborative learning experiences. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, pages 257–258, 2001. 2
- [6] Karanya Sitdhisanguan, Nopporn Chotikakamthorn, Ajchara Dechaboon, and Patcharaporn Out. Using tangible user interfaces in computer-based training systems for low-functioning autistic children. *Personal and Ubiquitous Computing*, 16(2): 143–155, 2012. 2
- [7] Hui Wei, Sicong Pan, Gang Ma, and Xiao Duan. Vision-guided hand-eye coordination for robotic grasping and its application in tangram puzzles. *AI*, 2(2):209–228, 2021. 2
- [8] Oren Zuckerman, Saeed Arida, and Mitchel Resnick. Extending tangible interfaces for education: Digital montessori-inspired manipulatives. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 859–868, 2005. 1, 2