ReGlove: A Soft Pneumatic Glove for Activities of Daily Living Assistance via Wrist-Mounted Vision

Rosh Ho[1] and Jian Zhang[1]

Abstract— This paper presents ReGlove, a system that converts low-cost commercial pneumatic rehabilitation gloves into vision-guided assistive orthoses. Chronic upper-limb impairment affects millions worldwide, yet existing assistive technologies remain prohibitively expensive or rely on unreliable biological signals. Our platform integrates a wrist-mounted camera with an edge-computing inference engine (Raspberry Pi 5) to enable context-aware grasping without requiring reliable muscle signals. By adapting real-time YOLO-based computer vision models, the system achieves 96.73 % grasp classification accuracy with sub-40.00 ms end-to-end latency. Physical validation using standardized benchmarks shows 82.71 % success on YCB (YCB Object and Model Set) object manipulation and reliable performance across 27.00 Activities of Daily Living (ADL) tasks. With a total cost under $250.00 and exclusively commercial components, ReGlove provides a technical foundation for accessible, vision-based upper-limb assistance that could benefit populations excluded from traditional Electromyography (EMG)-controlled devices.

I. INTRODUCTION

Upper-limb impairment resulting from stroke, spinal cord injury, or neuromuscular disorders affects over 5.00 million Americans, significantly impacting independence and quality of life. While sophisticated robotic orthoses exist commercially, their high cost (often exceeding $10 000.00) and complexity limit widespread adoption, particularly for chronic conditions requiring long-term use.

This work explores an alternative paradigm: functionally enhancing mass-produced, low-cost pneumatic rehabilitation gloves with vision-based control to create accessible assistive devices. Commercial pneumatic gloves present an attractive starting point, costing under $50.00 while offering inherent compliance and safety through soft actuation. However, they typically operate through simple manual controls or require reliable surface electromyography (sEMG) signals—a significant limitation for patients with weak or noisy muscle activation due to neurological damage.

Recent advances in computer vision for prosthetic control demonstrate that visual context can robustly inform grasp selection [1], [2]. However, these approaches have not been systematically applied to orthotic applications using commercial components. The ReGlove system bridges this gap by integrating established computer vision techniques with affordable, commercially available hardware.

This paper presents three key contributions: (1) An integrated hardware-software architecture that transforms commercial pneumatic gloves into vision-guided orthoses using readily available components; (2) A lightweight perception pipeline based on YOLO architectures that achieves real-time grasp classification on edge computing hardware; and (3) A comprehensive performance evaluation establishing baseline functionality across standardized benchmarks including YCB object manipulation and Activities of Daily Living (ADL) tasks. Through this proof-of-concept, we demonstrate a viable pathway toward assistive devices that balance capability with accessibility.

II. RELATED WORK

A. Actuation for Hand Assistance

Hand assistive devices primarily employ cable-driven or pneumatic actuation. Cable-driven systems [3], [4] transmit force from proximal motors through tendon-like mechanisms, offering precise control but suffering from mechanical complexity, cable management issues, and limited compliance.

Pneumatic actuators, used in commercial rehabilitation gloves, provide inherent compliance and safety through soft, inflatable chambers [5]. Clinical evidence supports their efficacy in improving hand function, with randomized trials showing significant improvements in active range of motion and grip strength for chronic stroke patients [6], [7]. Their commercial availability and low cost (<$50.00) make them a practical foundation for accessible assistive technology.

Alternative approaches include shape-memory alloys [8] and motorized exoskeletons, but these face challenges in reliability, weight, and cost that limit practical deployment.

B. Control Modalities

Traditional control methods include manual triggers and sEMG. Manual control requires use of the contralateral limb, making it impractical for independent use. sEMG-based control can enable more natural actuation but often fails for patients with weak or noisy signals due to neuromuscular degeneration [9].

Vision-based control, successfully demonstrated in prosthetic systems [1], [10], offers a promising alternative

by relying on object context rather than biological signals. Prior work primarily used computationally intensive architectures like VGG-16, limiting real-time performance on low-power hardware. We adapt this approach using modern YOLO architectures optimized for edge deployment, making vision-based control practical for orthotic applications where EMG may be unreliable.

III. SYSTEM DESIGN

The ReGlove system integrates a pneumatic glove with a vision-based control pipeline (Fig. 2). A wrist-mounted camera captures the visual scene, a Raspberry Pi 5 runs the grasp classifier, and an ESP32 microcontroller operates the pneumatic components. A binary intent signal (tactile switch or sEMG) initiates the control loop.
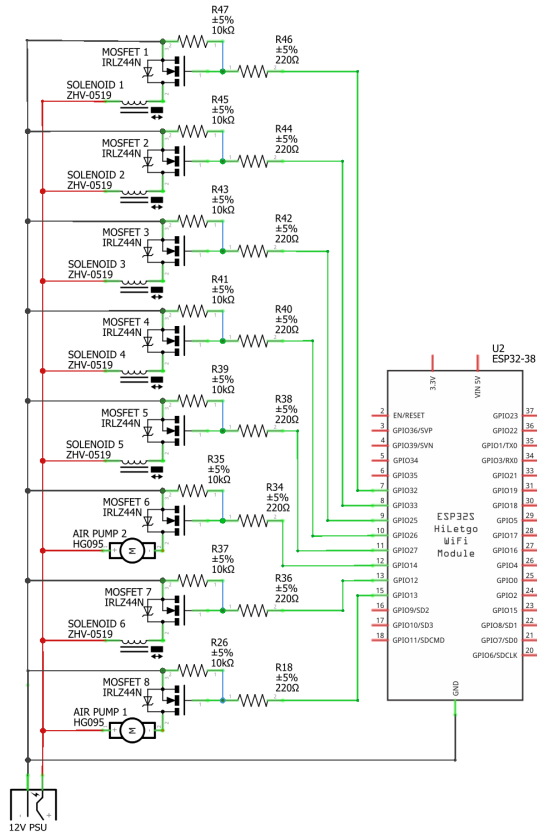
A. Hardware Implementation

Fig. 1. Complete wiring schematic for the pneumatic control system, illustrating connections between the Raspberry Pi 5, ESP32 microcontroller, solenoid valves, air pumps, and power supply components. The diagram shows both digital control signals and pneumatic pathways.

The pneumatic subsystem uses a commercial rehabilitation glove with ethylene-vinyl acetate bellows actuators, providing one degree of freedom per finger for bidirectional flexion and extension. We employ two HG095 mini air pumps ($6.00\,1\text{min}^{-1}$ flow rate) for inflation and vacuum generation, and six ZHV-0519 three-way solenoid valves for individual finger control. The complete wiring schematic is shown in Fig. 1.

Safety Considerations: The system incorporates multiple safety features including an exhaust solenoid that actively regulates pressure during flexion cycles, preventing over-pressurization and ensuring fail-safe operation. This design eliminates risk of actuator failure or user injury from excessive pressure buildup, maintaining compliance with soft robotic safety standards for human-worn devices.

Thumb Adaptation: The commercial glove's single-DOF design limits thumb opposition. We address this with a custom 3D-printed thermoplastic polyurethane (TPU) brace that maintains partial abduction while allowing pneumatic flexion, preserving capability for most functional grasp types [11]. The hand configuration with and without the brace and glove is shown in Fig. 4.

Pneumatic Circuit: The system employs a semi-closed loop design with separate inflation and deflation subloops. During extension, the inflation pump activates while selected finger solenoids open; during flexion, the vacuum pump activates with reversed valve states. An exhaust solenoid regulates pressure between cycles.

Control Inputs: While the system architecture supports multiple input modalities (sEMG, EEG, EOG), we use a simple tactile switch for benchtop validation to isolate vision system performance. This allows future drop-in replacement with sEMG once IRB approval is secured for clinical studies.

The total hardware cost is approximately $235.00 (Table I), with detailed specifications in supplementary materials.

TABLE I
Hardware Cost Breakdown (as of October 2025)

Component	Cost (USD)
Pneumatic glove with finger control	$17.00
ZHV-0519 three-way solenoid valves (×6)	$19.50
Vinyl tubing (4 × 5 mm)	$7.50
HG095 12 V DC, 6 L min⁻¹ air pumps (×2)	$3.46
ESP32-WROOM-32D Microcontroller	$4.29
Raspberry Pi 5 (8 GB)	$81.19
Logitech c270 (wrist-mounted camera)	$24.00
MyoWare sEMG sensors	$39.90
IRLZ44N MOSFET (×8)	$8.96
12 V rechargeable battery	$28.99
Total	$234.79

Note: Costs are approximate and vary based on supplier.

B. Vision Pipeline & Model Development

We used a grasp classification system using three publicly available datasets: DeepGrasping (885.00 images) [1], ImageNet subset (5180 images), and Hand-Cam (250 images) [2]. To address class imbalance, we applied extensive data augmentation including geometric transformations, photometric adjustments, and occlusion modeling, yielding approximately 2000 images per grasp type (pinch, power, three-jaw chuck, tool, key).
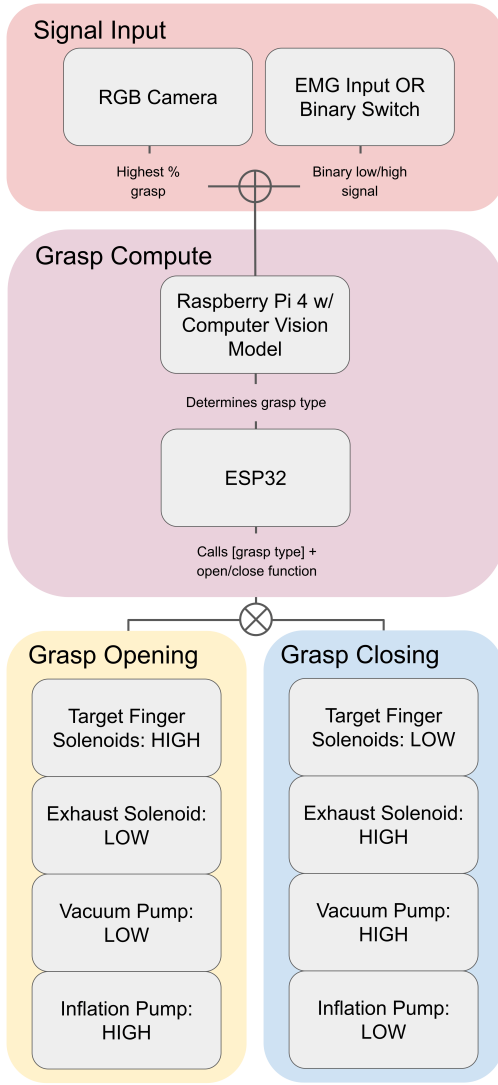
Fig. 2. End-to-end system workflow. The wrist-mounted camera captures the visual scene and streams RGB frames to the Raspberry Pi 5 for inference using the lightweight YOLO-based grasp classifier. The predicted grasp type is forwarded to the ESP32 microcontroller, which manages valve-switching logic for the pneumatic circuit and actuates the glove accordingly. A binary intent signal (tactile switch or sEMG) initiates the control loop, while the pumps and solenoid manifold generate positive or negative pressure to drive finger extension or flexion. This diagram summarizes the integration of sensing, inference, pneumatic routing, and actuation within the complete assistive architecture.

We evaluated multiple architectures under identical training conditions:

- VGG-16 & VGG-16 + Depth: Baseline models replicating prior work [1]
- YOLO v11 & v12: Modern lightweight object detectors optimized for edge deployment

Depth augmentation using synthetic depth maps from DepthAnything [12] did not improve performance, likely due to inconsistency in synthetic depth quality. Both YOLO variants significantly outperformed VGG-based approaches (Table II), with YOLO v11 achieving 96.67 %

accuracy versus 82.59 % for VGG-16. YOLO's superior performance stems from architectural features that preserve spatial structure (Spatial Pyramid Pooling - Fast, Feature Pyramid Network/Path Aggregation Network layers) and integrated augmentation mechanisms that improve robustness to lighting and background variation.

Given its optimal accuracy-latency tradeoff, we selected YOLO v11 for system integration, achieving 0.90 ms inference latency on Raspberry Pi 5—well below the 10.00 ms to 20.00 ms threshold for human-perceptible feedback [13].

TABLE II
Grasp Classification Model Performance Comparison

Model	Accuracy (%)	Inference Time (ms)
VGG-16	82.59	7.24 ± 0.45
VGG-16 + Depth	79.91	7.32 ± 0.52
YOLO v11	96.67	0.90 ± 0.15
YOLO v12	96.45	0.50 ± 0.08

IV. EXPERIMENTAL RESULTS

A. Grasp Classification Performance

The YOLO v11 model achieved a mean grasp classification accuracy of 96.67 % (95.00 % CI: 95.20 % to 97.80 %) on the test set. Analysis of the confusion matrix (supplementary Fig. S1) revealed that most misclassifications occurred between geometrically similar pinch and three-jaw chuck grasps. Performance degradation was primarily observed for scale-ambiguous objects where visual cues alone were insufficient to infer absolute size.

The model's inference latency of 0.90±0.15 ms enables real-time operation, with total image preprocessing and classification completing in under 2.00 ms. This represents a 8.00 × speedup compared to VGG-16 while maintaining superior accuracy.

B. Physical Grasping Performance

We evaluated physical grasping capability using standardized benchmarks to assess functional utility.

1) YCB Object Set: Using the YCB Gripper Assessment Protocol [14], ReGlove achieved an overall success rate of 82.71 % (215.50/260.50 points). Performance was robust for objects with defined edges and surfaces (cups, blocks, utensils) but lower for small, smooth, or low-friction items (marbles, coins, washers). This performance gap primarily reflects mechanical limitations of the compliant ethylene-vinyl acetate actuators rather than perception errors. Full results are available in supplementary materials (Table S-III).

2) Activities of Daily Living (ADL): On a subset of 27.00 ADL tasks based on Matheus & Dollar [15], the system achieved a mean performance score of 2.65±0.28 out of 3.00 (0.00=failed, 3.00=excellent). The system excelled at tasks involving power or tripod grasps (pouring liquids, manipulating utensils) but struggled with

fine manipulation requiring precise fingertip control (unwrapping tablets, rotating small bolts). A comparative analysis of execution times for these tasks is shown in Fig. 3.

Multi-phase operations revealed limitations in sequential grasp switching, highlighting the need for more sophisticated control hierarchies. Complete task-by-task results are provided in supplementary materials (Table S-IV, Figure S2).

C. Integrated System Performance

The complete assistive system achieved end-to-end latency of 38.00 ± 6.40 ms from image capture to glove actuation, confirming real-time responsiveness for interactive use. The system reliably executed all five grasp types under live inference conditions without performance degradation during extended operation.

During 90.00-minute continuous testing sessions, the waist-mounted pneumatic unit maintained stable operation without overheating or pressure drift. Average power consumption was 10.30 ± 1.20 W, compatible with commercially available 12.00 V portable battery packs for untethered operation.

TABLE III

Summary of System Performance Evaluation

Metric	Performance
Software Performance	
Grasp Classification Accuracy	96.67 %
Inference Latency	0.90 ± 0.15 ms
Hardware Performance	
YCB Object Success Rate	82.71 %
ADL Task Score (0.00 to 3.00)	2.65 ± 0.28
Integrated System	
End-to-End Latency	38.00 ± 6.40 ms
Average Power Draw	10.30 ± 1.20 W
Continuous Operation Duration	90.00 minutes

V. DISCUSSION

The ReGlove system demonstrates that commercial pneumatic rehabilitation gloves can be effectively converted into vision-guided assistive orthoses through integration with modern computer vision and low-cost computing hardware. This approach offers a affordable (under $250.00), non-invasive pathway toward functional hand assistance that circumvents the limitations of EMG-based control.

A. Technical Performance and Significance

The system's 96.67 % grasp classification accuracy and 38.00 ms end-to-end latency compare favorably with prior vision-based prosthetic systems requiring more complex hardware [1], [2]. More significantly, by relying exclusively on visual context rather than biological signals, the approach extends accessibility to patient populations with unreliable EMG due to neuromuscular degeneration [9].

The performance gap between software perception (96.67 % accuracy) and physical execution (82.71 % YCB success) highlights the mechanical limitations of commercial pneumatic gloves rather than perception shortcomings. This suggests that relatively simple hardware improvements—such as high-friction fingertip coatings or reinforced actuator segments—could significantly enhance functional performance without increasing system complexity or cost.

B. Limitations and Design Considerations

Several important limitations warrant discussion. The current "pause-and-select" control paradigm requires users to position their hand and trigger a single, static grasp. This does not support dynamic tasks requiring mid-manipulation grasp adjustments or provide mechanisms for user correction of mispredicted grasps.

The system's performance with small, smooth objects remains limited by the compliant nature of pneumatic actuation. While this compliance enhances safety, it reduces precision for fine manipulation tasks. Future iterations could incorporate variable-stiffness mechanisms or hybrid actuation approaches to balance safety and dexterity.

Our benchtop validation used a healthy operator, which allowed controlled testing of core functionality but leaves open questions about real-world performance with impaired users. The simplified binary intent detection (tactile switch) served as a reliable trigger for technical validation but may not reflect the control challenges faced by target users.

C. Future Directions

Building on this proof-of-concept, several research directions appear promising:

- Multi-modal control integration: Subsequent iterations will incorporate sEMG as the primary intent detection modality, operating in concert with the existing vision-based grasp classification. This hybrid approach will enable more natural actuation paradigms while maintaining the robustness of visual context awareness. Additionally, implementation of closed-loop force control will enhance manipulation precision and user experience.
- Hardware refinement: Improved actuator geometry, high-friction surfaces, and variable-stiffness mechanisms to enhance grip stability and fine manipulation capability.
- Control hierarchy expansion: Temporal grasp sequencing and gesture prediction to enable complex, multi-phase tasks like opening containers or using tools.
- Clinical translation: Formal studies with stroke and SCI patients to quantify ADL improvement, user acceptance, and long-term usability.
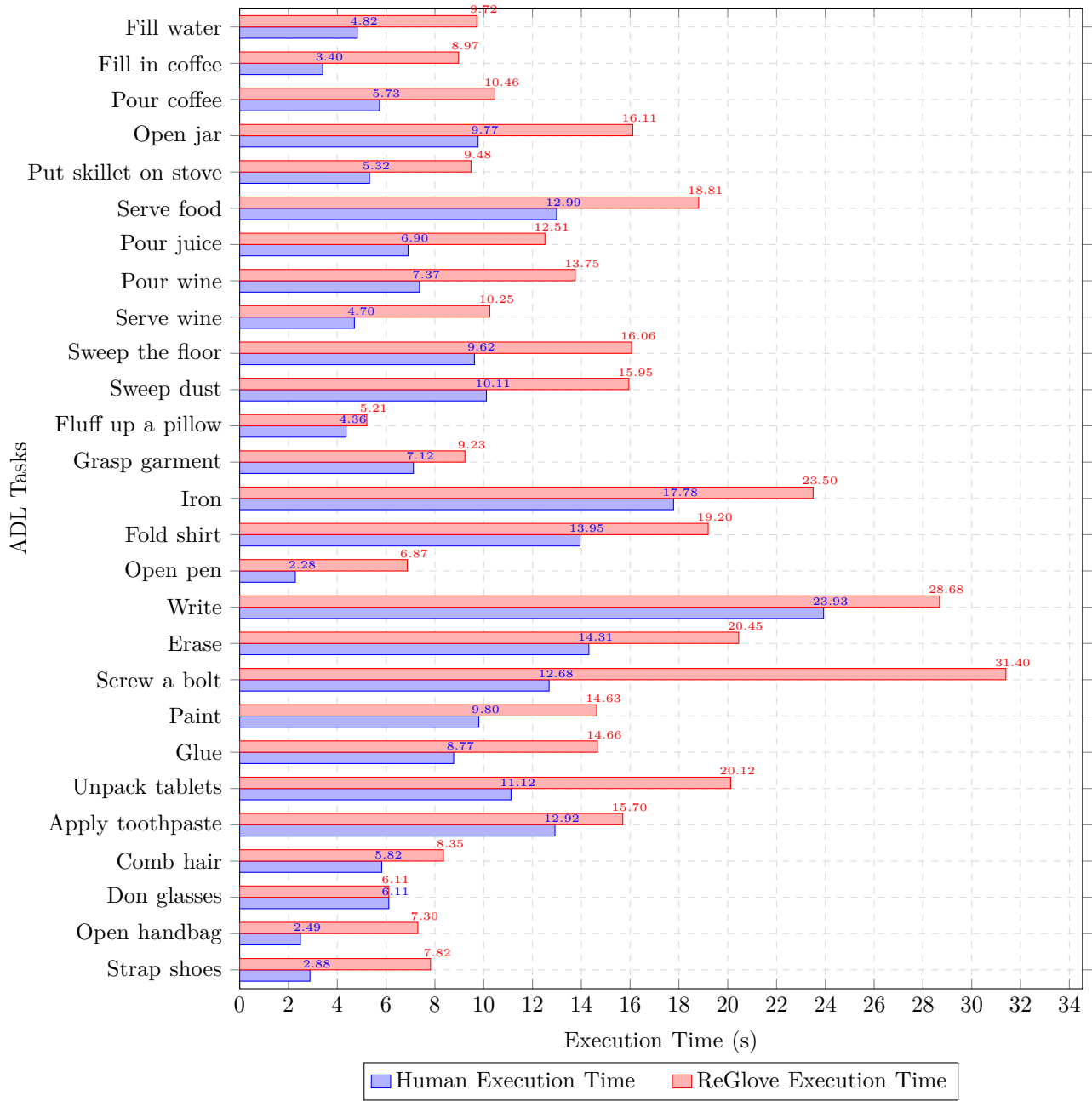
Fig. 3. Comparative analysis of human versus ReGlove execution times across 27.00 ADL tasks. Blue bars represent average human performance, while red bars show ReGlove-assisted performance.

- System integration: Miniaturization of pneumatic components and development of fully self-contained wearable form factors.

The modular architecture supports incremental improvement in each of these areas while maintaining the core benefits of affordability and accessibility.

VI. CONCLUSION

This work presents ReGlove, an end-to-end demonstration of vision-guided pneumatic hand assistance using exclusively commercial components and open-source software. The system achieves real-time dexterous grasping with 96.67 % classification accuracy and 82.71 % physical success on standardized benchmarks, while maintaining a total cost under $250.00.

By bridging affordable rehabilitation hardware with modern computer vision, ReGlove offers a practical pathway toward restoring functional hand capability for individuals with chronic upper-limb impairment. The approach demonstrates that intelligent assistive technology need not be complex or expensive to be effective, providing a foundation for future development of accessible devices that can significantly impact quality of life for underserved populations.

Fig. 4. Hand configuration comparisons: (a) bare hand, (b) hand with 3D printed thumb brace, (c) complete orthosis glove worn over thumb brace. The brace maintains functional thumb positioning while allowing pneumatic flexion.

FUNDING

SUPPLEMENTARY MATERIALS

Additional materials are available as ancillary files with this arXiv submission, including:

- Confusion matrix analysis (Fig. S1)
- Complete YCB benchmark results (Table S-III)
- Detailed ADL task performance (Table S-IV, Figure S2)
- Hardware specifications and wiring diagrams

References

[1] J. DeGol, A. Akhtar, B. Manja, and T. Bretl, "Automatic grasp selection using a camera in a hand prosthesis," in Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 431–434.

[2] L. Taverne, M. Cognolato, T. Bützer, R. Gassert, and O. Hilliges, "Video-based prediction of hand-grasp preshaping with application to prosthesis control," pp. 4975–4982, 2019.

[3] K. B. Kim, H. Choi, B. Kim, B. B. Kang, S. Cheon, and K.-J. Cho, "Exo-glove poly iii: Grasp assistance by modulating thumb and finger motion sequence with a single actuator," Soft Robotics, vol. 12, 2025.

[4] C. G. Rose and M. K. O'Malley, "Hybrid rigid-soft hand exoskeleton to assist functional dexterity," IEEE Robotics and Automation Letters, vol. 4, pp. 73–80, 2019.

[5] D. Y.-L. Lim, H.-S. Lai, and R. C.-H. Yeow, "A bidirectional fabric-based soft robotic glove for hand function assistance in patients with chronic stroke," Journal of Neuroengineering and Rehabilitation, vol. 20, p. 120, 2023.

[6] S. Fardipour and M. Hadadi, "Investigation of therapeutic effects of wearable robotic gloves on improving hand function in stroke patients: A systematic review," Current Journal of Neurology, 2022.

[7] M.-J. Ko, Y.-C. Chuang, L.-J. Ou-Yang, Y.-Y. Cheng, Y.-L. Tsai, and Y.-C. Lee, "The application of soft robotic gloves in stroke patients: A systematic review and meta-analysis of randomized controlled trials," Brain Sciences, vol. 13, p. 900, 2023.

[8] T. Bützer, O. Lambercy, J. Arata, and R. Gassert, "Fully wearable actuated soft exoskeleton for grasping assistance in everyday activities," Soft Robotics, vol. 8, pp. 128–143, 2021.

[9] J. Xu, R. Wang, S. Shang, A. Chen, L. Winterbottom, T.-L. Hsu, W. Chen, K. Ahmed, P. L. L. Rotta, X. Zhu, D. M. Nilsen, J. Stein, and M. Ciocarlie, "Chatemg: Synthetic data generation to control a robotic hand orthosis for stroke," IEEE Robotics and Automation Letters, vol. 10, pp. 907–914, 2025.

[10] Y. Zhang, Y. Xie, Q. Zhao, X. Xu, H. Deng, and N. Yi, "Vision-based grasping method for prosthetic hands via geometry and symmetry axis recognition," Biomimetics, vol. 10, p. 242, 2025.

[11] J. Z. Zheng, S. De La Rosa, and A. M. Dollar, "An investigation of grasp type and frequency in daily household and machine shop tasks," in 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 4169–4175.

[12] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth anything: Unleashing the power of large-scale unlabeled data," arXiv preprint, 2024.

[13] J. Jerald and M. C. Whitton, "Relating scene-motion thresholds to latency thresholds for head-mounted displays," Proceedings of the IEEE Virtual Reality Conference, pp. 211–218, 2009.

[14] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the yale-cmu-berkeley object and model set," IEEE Robotics & Automation Magazine, vol. 22, pp. 36–52, 2015.

[15] K. Matheus and A. M. Dollar, "Benchmarking grasping and manipulation: Properties of the objects of daily living," in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010, pp. 5020–5027.