Computer Vision and Image Understanding 265 (2026) 104669

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Computer vision for high-level control of prosthetic limbs: A literature review☆

Gianmarco Cirelli [a], Christian Tamantini [b,*], Loredana Zollo [a], Francesca Cordella [a]

[a] Research Unit of Advanced Robotics and Human-Centred Technologies, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo, 21, Rome, 00118, Italy
[b] Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Via Giandomenico Romagnosi, 18a, Rome, 00196, Italy

ARTICLE INFO

Keywords:
Upper limb prosthesis
Lower limb prosthesis
Computer vision
Assistive technologies
Wearable robots

ABSTRACT

Integrating computer vision into powered prosthetic devices is garnering attention as a potential avenue for enhancing functionality. These methodologies have been introduced to address challenges such as prosthetic abandonment, which can be attributed to the unreliability and lack of intuitiveness of myoelectric control. This paper provides insights into the current state, challenges, and future directions of integrating computer vision into prosthetic limb control systems. To this aim, a literature review was conducted, identifying 50 relevant studies (33 on upper-limb and 17 on lower-limb prostheses), sourced from Scopus, PubMed, and IEEE Xplore, with the search updated to May 2025. The hardware implementation aspects, including camera sensor positioning and computational units, were synthesized, as well as the computer vision approaches implemented with a specific emphasis on their implications for limb functional performance.

The review identified strengths and suggested areas for improvement, emphasizing the need for studies on optimal camera placement and the feasibility of embedded computational units for real-world testing.

Even if the reported results are promising, the performed analysis outlined the need to conduct usability studies and introduce vision-based approaches into lower-limb prosthesis control, which should be tested in real-world scenarios. Validation in real settings is crucial for these technologies to move beyond laboratory settings.

Contents

☆ This article is part of a Special issue entitled: 'ACVR 2024' published in Computer Vision and Image Understanding.

* Corresponding author at: Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Via Giandomenico Romagnosi, 18a, Rome, 00196, Italy.
E-mail address: christian.tamantini@cnr.it (C. Tamantini).

https://doi.org/10.1016/j.cviu.2026.104669
Received 16 May 2025; Received in revised form 2 October 2025; Accepted 22 January 2026
Available online 2 February 2026

1. Introduction

Wearable robots, including powered prostheses and exoskeletons, are designed to be integrated with the human body to replace lost limb functions or enhance physical capabilities. However, challenges such as limited intelligence, functionality, and reliance on basic sensors hinder effective human–robot interaction, preventing users from perceiving these devices as natural extensions of themselves (Xia et al., 2024). To overcome these limitations, additional sensory information have been integrated to enable more intuitive control, improve dexterity, expand task capabilities, and could reduce user cognitive effort (Gionfrida et al., 2024). In details, incorporating environmental awareness into wearable robots could enhance user safety by enabling threat detection and obstacle navigation (Medioni and Trivedi, 2017; Wang and Zhu, 2023). It allows these systems to transition from reactive responses to proactive planning, simulating responses to future stimuli over extended time frame (Nelson and MacIver, 2006).

It is particularly evident in the field of prosthetic control, whether for upper or lower limb devices. Its reliability remains in fact a significant challenge, as it directly impacts the user ability to achieve consistent, intuitive, and functional movements in daily life (Fleming et al., 2021; Ledoux and Goldfarb, 2017). This leads to the abandonment of prosthetic systems and highlights the need to find strategies to increase reliability (Cordella et al., 2016).

The limb loss is a devastating experience with functional and psychological repercussions (Şimsek et al., 2020; Yamamoto et al., 2019; Amtmann et al., 2015) severely impairing the amputees' capability to carry out Activities of Daily Living (ADLs) and work-related tasks (Sinha et al., 2011; Tamantini et al., 2021). It implies a restriction in the independence and quality of life (Manz et al., 2022). Powered limb prostheses aim to replace a missing body part to restore lost functions (Organization et al., 2011; Xiloyannis et al., 2021).

Despite advancements in prosthetic technology, the rejection rate remains high at 44% (Salminger et al., 2022), with 76.9% of upper-limb prosthetic users prioritizing cosmetic over functional use (Jang et al., 2011). Discomfort and functional limitations significantly impact prosthesis adoption (Smail et al., 2021; McDonald et al., 2021). Electromyographic (EMG) signals are commonly used to detect motion intent in powered upper-limb prostheses (Roche et al., 2014; Losey et al., 2018). However, the performance of EMG-driven control depends on balancing motion classes with system stability (Yadav and Veer, 2023) and is limited by algorithm constraints and the difficulty amputees face in generating consistent contractions (Farina et al., 2023; Schone et al., 2024). Performance further degrades over time due to factors like sweating, muscle fatigue, and prosthetic socket repositioning (Yeung et al., 2022; Gulati et al., 2021). To address the limitations of muscle-based control, computer vision has been introduced to improve the robustness of prosthetic control and object manipulation. By simulating human vision, it enhances environmental understanding, extracts object features, and refines the prosthetic hand preshape. This innovation marks a significant step toward integrating advanced sensory capabilities, improving functionality and usability for users (Yang and Liu, 2021).

Similarly, powered lower-limb prostheses offer several advantages over passive ones, including improved natural locomotion, reduced compensatory behaviors (Armannsdottir et al., 2018), and lower metabolic costs during ambulation (Sun et al., 2023). Trans-femoral and trans-tibial amputees with passive prostheses consume 60% and 30% more energy, respectively, and experience walking speeds 10% to 65% slower than non-amputees (Asif et al., 2021). However, drawbacks such as increased weight due to sensors, actuators, and batteries, along with the complexity of developing adaptable control systems, remain challenging (Huang et al., 2021; Yip et al., 2023). Effective control must respond to environmental changes (Young and Hargrove, 2015; Gehlhar et al., 2023), where computer vision enhances terrain prediction and contextual feature extraction, such as ramp angles and stair heights, to aid control (Liu et al., 2015; Massalin et al., 2017).

In prosthetics, the term control broadly refers to the mechanisms that enable a prosthesis to translate the user's intentions into functional actions. This control is typically described across different levels (Gentile et al., 2022). High-level control focuses on decoding the amputee's biological signals and converting them into meaningful commands, such as selecting a grasp type or planning the movement to be performed. Low-level control, on the other hand, is responsible for generating the actual control signals that drive the motors of the robotic prosthesis. In addition, by continuously monitoring the external environment with which the prosthesis interacts, low-level control laws can embed mechanisms that resemble human reflexes, enabling the system to react promptly to external events—for example, by adapting grip force upon contact or stabilizing the limb against unexpected perturbations (Stefanelli et al., 2023). In this context, visual information enriches high-level control methodologies by providing predictive and contextual information about the environment or the objects with which the user intends to interact. This allows the prosthesis to understand its surroundings, anticipate user goals, and select appropriate strategies for action, such as recognizing objects, inferring grasping strategies, detecting terrain characteristics, and planning movements proactively.

Despite significant advances in prosthetic control and signal decoding, a comprehensive understanding of how computer vision contributes to these systems is still lacking. While individual studies have explored vision-based strategies for object recognition, scene understanding, or terrain adaptation, the field remains fragmented, and a clear synthesis of approaches and trends is missing. This review addresses this gap by providing a structured literature analysis of the development of computer vision in prosthetic applications, clarifying how visual input has been integrated into control frameworks. In particular, it emphasizes the role of computer vision as a key enabler of high-level prosthetic control, based on the rationale that effective control requires not only decoding user intent but also anticipating external conditions.

The scientific literature is classified to trace emerging trends in hardware, such as the positioning of cameras and the deployment of systems on portable platforms compared to benchtop testing. Software contributions are examined in terms of the size of input data, computer vision algorithms, and implemented models, while validation strategies are compared with respect to subject enrollment, the distinction between healthy and amputee participants, evaluation metrics, and protocol activities. By mapping these dimensions, the review not only summarizes the current state of the art but also delineates the trajectory of research, highlighting the challenges that remain open and the directions toward which vision-based prosthetic control is moving.

The paper is structured as follows: Section 2 details the methodology used for literature reviewing, the results of which are shown in Section 3. Section 4 highlights in detail the characteristics of the selected scientific papers, dividing the approaches for lower limb prostheses from those for upper limb ones. Lastly, Section 5 concludes this review, outlining strengths and possible future directions in the field of prosthetic control based on computer vision systems.

2. Methodology

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards were followed in the execution of the literature review (Liberati et al., 2009). English-language full-text journal Scientific papers were chosen after a thorough search of the IEEE Xplore, PubMed, and Scopus databases. The following keywords and logic operators were combined: ("Computer vision" OR "Artificial vision" OR "Embodied intelligence") AND ("upper-limb" OR "Hand-wrist" OR "Hand" OR "lower-limb" OR "Knee") AND ("Prosthesis" OR "Prosthetic" OR "Prosthetic control"). No filter was applied on the publication date of the scientific papers, and all results of each database were included up to 15 May 2025.

The inclusion criteria are:

- The study should be published in an international journal using a peer-review procedure. The rigorous selection process guarantees the authenticity and dependability of the works that are taken into account for the study.
- The scientific paper discusses the application of computer vision in the control of limb prostheses (non-limb prostheses, e.g. eye prostheses, are not included).
- The vision sensor must be worn by the user.

After removing duplicates, two independent reviewers screened each article based on its title and abstract to confirm that the studies met the predefined inclusion criteria. The methodological quality of the selected studies was then evaluated using the AXIS (Appraisal Tool for Cross-Sectional Studies) methodology to assess the risk of bias (Downes et al., 2016). This technique evaluates several aspects of the included works such as the study design, the methodology, and the clarity and transparency of the presentation of the results.

Both the screening process of the resulting scientific works and the AXIS assessment were independently carried out by the first two authors (GC, CT). In case of a disagreement, the last author (FC) decided whether a paper should be included or not in this review and how their risk was assessed.

Data extraction was executed on the included scientific papers, based on the following checklist: (i) authors; (ii) objective of the study; (iii) type of prosthesis (when used); (iv) typology and placement of the adopted computer vision system; (v) exploited computer vision algorithm; (vi) dataset size and online availability at the time of the review (in case the dataset is available, the link is provided); *(vii)* characteristics of the participants involved in the study; *(viii)* tasks executed in the experimental protocol and performance metrics considered.

3. Results

The review process result is shown in Fig. 1 by the PRISMA flow diagram. Following database searching and removal of duplicates, a total of 343 papers were initially identified for the analysis.

Among these, 285 papers were excluded after reading the title and abstract, and the remaining papers underwent full review. Out of these, 8 papers were further excluded since they did not meet the inclusion criteria. For example, Shi et al. (2024, 2020) are not considered because the vision sensor is not worn by the user. Therefore, the screening process returned 50 papers that were fully studied and included in the analysis presented in this review.

According to the AXIS tool assessment, the included studies attained an average score of 85.1%. The majority of studies excelled in key areas (100% yes), such as clearly defining their objectives, providing sufficient methodological details for replication, and ensuring that their results aligned with their analyses. However, notable variability was observed in aspects like obtaining ethical approval or participant consent (73.2% yes), the presence of a sample representative of the target population(amputees, 44.2% yes), and the discussion of study limitations (58.3% yes). Despite these inconsistencies, the overall risk of bias remains low.

It is essential to highlight that this review primarily focuses on examining the technological and methodological approaches used for the validation of computer vision-based control systems in prosthetic devices. Given this emphasis, the methodological rigour observed in relevant areas suggests that the included studies offer a solid basis for understanding current advancements, identifying challenges, and guiding future research toward real-world implementation.

Fig. 2 shows the bibliometric network of the papers selected for this review that had at least 2 co-occurrences produced using VOSviewer software (Arruda et al., 2022). The bibliometric network provides an intuitive visual representation of trends and thematic connections among the selected papers, helping to represent the occurrences of the keywords. The items are represented by using their labels and a circle, where the size depends on the number of occurrences. Specifically, the higher the weight of an item, the larger the label and its circle. Therefore, it is evident that in the selected scientific papers, the items "*computer vision*", "*prosthetics*", and "*deep learning*" have the highest weight, i.e., the largest label and circle sizes, underlining the pertinence of the papers to the aim of this review.

The remainder of the Section presents the findings of the review, organized separately for upper- and lower-limb applications. This division reflects the fundamentally different role that computer vision assumes in each domain: for lower-limb prostheses, the focus is primarily on terrain perception and classification, whereas for upper-limb prostheses it centers on object detection and grasp classification. Moreover, the technical challenges associated with locomotion and manipulation differ considerably, making an independent treatment more appropriate. Cross-domain aspects are nevertheless addressed in Section 4, where joint considerations are drawn and comparative analyses are provided.

3.1. Upper limb prosthesis

The 33 selected papers on upper-limb prosthetics share the goal of developing multimodal control strategies that fuse exteroceptive and proprioceptive information to manage the robotic hand preshaping for object grasping. Specifically, Semiautonomous Control Strategies (SCS) were introduced to reduce user burden by allowing the prosthesis to autonomously handle tasks such as hand orientation or gesture selection. The amputee remains responsible for activating the control pipeline, modifying the control output, or restarting the process. Typically, user motion intention is decoded via surface EMG.

Despite variations among studies, a general system architecture can be outlined, as shown in Fig. 3. Visual information from the CVS is processed by the computer vision algorithm to extract object features. These features, combined with user intention, are input to the control system to manage robotic hand preshaping, such as grasp type and/or wrist orientation. The system then outputs control commands and feedback to the user.

Tables 1, 2 provide an overview of the selected papers, detailing the research aims, prostheses devices, CVS hardware and algorithmic components, sensor positioning, and subjects involved in experimental validation. Accuracy and computational load, defined as the time in milliseconds required to execute the algorithm, are reported where available.

3.1.1. Computer vision system

The Computer Vision System (CVS) consists of a camera sensor and a computational unit, with variations in technology (RGB or RGB-D), sensor positioning, and technical characteristics such as image resolution, acquisition frequency, and computational unit type.

RGB cameras operate in the visible light domain and typically use pixels consisting of charge-coupled semiconductor or complementary metal-oxide semiconductor devices. They are generally low-cost,

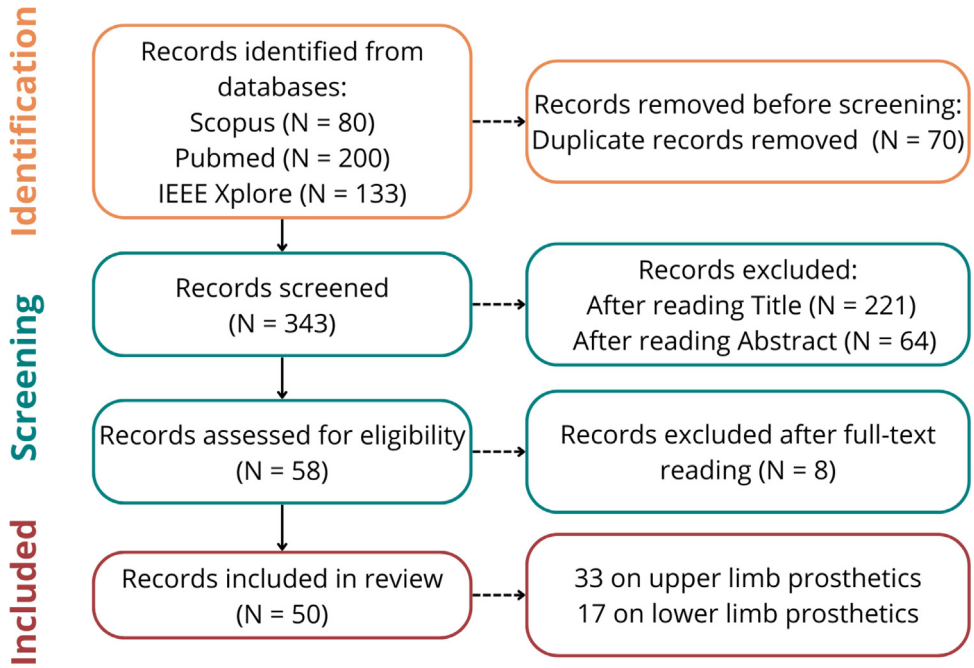

Fig. 1. PRISMA flow chart.

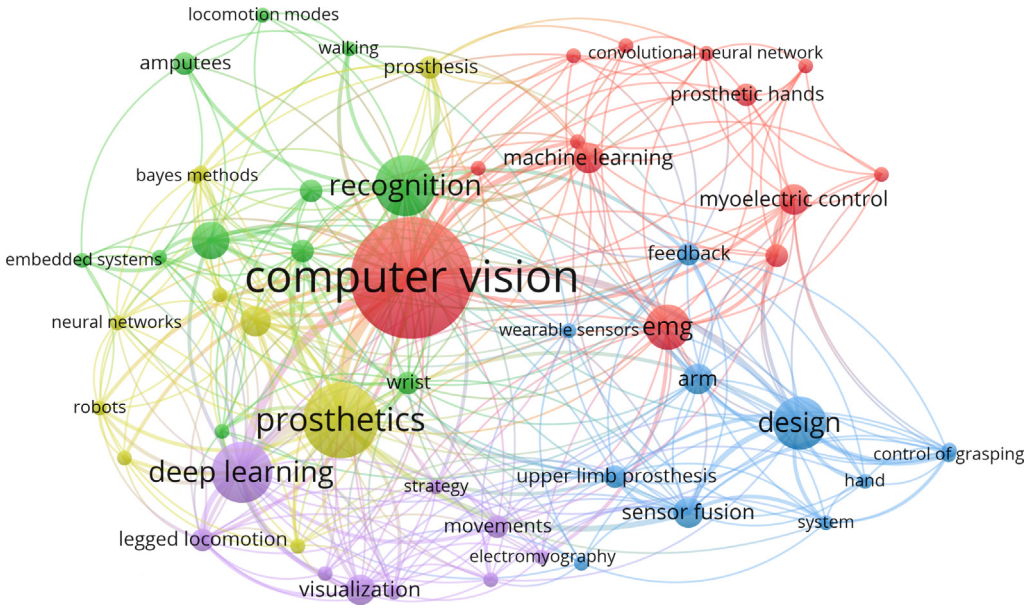

Fig. 2. Bibliometric network showing the map of co-occurring keywords with VOSviewer.

provide good resolution, and are compact enough for prosthetic integration. They are typically positioned at the palm (He et al., 2020; Castro et al., 2022; Weiner et al., 2022; Starke et al., 2022; Cirelli et al., 2023), back (Došen et al., 2010; Ghazaei et al., 2017; He et al., 2019; Cui et al., 2022), or ulnar border (Zhong et al., 2020b); in Cirelli et al. (2023) the positioning of the camera was justified considering the level of occlusion of the hand in the image. Some works use egocentric views via augmented reality glasses with stereo RGB cameras (Markovic et al., 2014; Roy et al., 2021; Fejér et al., 2021; Cognolato et al., 2022; Wang et al., 2022; Zandigohar et al., 2024; Zhang et al., 2024; Kyberd et al., 2023; Gardner et al., 2020).

RGB-D cameras combine RGB sensors with laser and LED-based depth sensors to estimate the distance between the sensor and surfaces. Unlike passive RGB sensors, these are active devices, emitting a signal that is reflected and detected (Tamantini et al., 2022). Due to their bulk and power demands, they are typically used in setups where the vision sensor is not integrated into the prosthesis, but placed at the head (McMullen et al., 2013; Markovic et al., 2015; Mouchoux et al., 2021; Park et al., 2022; Karrenbach et al., 2022; Peng et al., 2024; A Powered Prosthetic Hand with Vision System, 2025), back of the hand Castro and Dosen (2022), Peng et al. (2024), or elbow (Huang et al., 2022).

As mentioned earlier, the CVS includes a computational processing unit, which in most studies is a personal computer (Došen et al., 2010; Došen and Popović, 2011; McMullen et al., 2013; Markovic et al., 2014, 2015; Ghazaei et al., 2017; He et al., 2019; Zhong et al., 2020b; He et al., 2020; Roy et al., 2021; Mouchoux et al., 2021; Castro and Dosen, 2022; Huang et al., 2022; Park et al., 2022; Karrenbach et al., 2022;

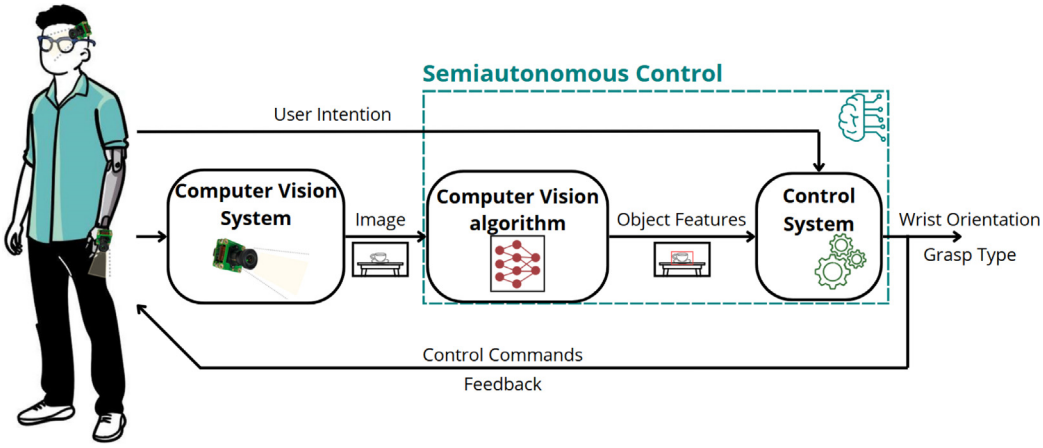

Fig. 3. Block diagram of the system with the upper-limb prosthesis.

Cognolato et al., 2022; Wang et al., 2022; Cui et al., 2022; Huang et al., 2023; Deshmukh et al., 2023; Zandigohar et al., 2024; Zhang et al., 2024; Kyberd et al., 2023; A Powered Prosthetic Hand with Vision System, 2025; Gardner et al., 2020), with only a few considering its portability (Ragusa et al., 2021; Fejér et al., 2021; Castro et al., 2022; Cirelli et al., 2023; Peng et al., 2024; Weiner et al., 2022; Starke et al., 2022). Portability is essential for integration into prostheses and use in real-life settings. Regarding image resolution, there is no standard to follow, as shown by the variety in Tables 1, 2. The resolution should instead strike an optimal balance between image quality and computational cost.

3.1.2. Computer vision algorithm

Software approaches in the literature vary according to the type of information extracted by the algorithm (e.g., image segmentation, geometrical feature extraction, or object detection through deep learning) and the number of hand gestures and objects considered. Deep learning methods can be further distinguished by the type of training used (custom or public datasets, transfer learning) and their reported performance (accuracy and frames per second).

The information acquired by the CVS is processed through computer vision algorithms to extract object features. Early strategies relied on classical segmentation to identify objects in the image, with distance estimation used to manage wrist rotation, grasp type, and hand opening (Došen and Popović, 2011; Došen et al., 2010; McMullen et al., 2013). An extension of this approach fits the segmented object with geometric primitives to refine the grasp (Markovic et al., 2014, 2015; Mouchoux et al., 2021; Castro and Dosen, 2022). These solutions often use RGB-D cameras to estimate the distance and real size of the object, with geometric fitting typically involving the RANSAC (Random Sample Consensus) algorithm (Fischler and Bolles, 1981). The number of grasp types corresponds to the selected primitives. However, these methods are not easily integrated for everyday use, as they require bulky RGB-D systems that cannot be easily embedded in prosthetic hands.

To develop embeddable solutions for prosthetic hands, the focus has shifted to approaches compatible with RGB cameras. Deep learning, is increasingly used in various fields (Shrestha and Mahmood, 2019), including computer vision (Voulodimos et al., 2018). Indeed, Convolutional Neural Networks (CNNs) are ideal for object detection and classification tasks, as they analyze structured data like images through convolutional operations for feature extraction and pattern recognition (Li et al., 2021). For this reason, CNNs have been widely used in the past decade to recognize objects in RGB images and select the appropriate grasping configurations (Ghazaei et al., 2017; He et al., 2019; Zhong et al., 2020b; He et al., 2020; Roy et al., 2021; Ragusa et al., 2021; Fejér et al., 2021; Castro et al., 2022; Huang et al., 2022; Weiner et al., 2022; Starke et al., 2022; Park et al., 2022; Karrenbach et al., 2022; Cognolato et al., 2022; Wang et al., 2022; Cui et al., 2022; Cirelli et al., 2023; Huang et al., 2023; Deshmukh et al., 2023; Zandigohar et al., 2024; Zhang et al., 2024; Peng et al., 2024; A Powered Prosthetic Hand with Vision System, 2025). Within this category, different families of algorithms have been exploited in this fields:

- *One-stage detectors*, such as YOLO (Redmon et al., 2016) or SSD (Liu et al., 2016), which directly predict object bounding boxes and classes in a single step. These approaches are fast and well suited to real-time applications, and have been exploited in prosthetic control to achieve high frame rates in grasp recognition tasks (Cirelli et al., 2023; Deshmukh et al., 2023).
- *Lightweight CNNs*, such as MobileNet (Sinha and El-Sharkawy, 2019), designed to balance accuracy and computational load, enabling deployment on embedded devices with limited resources (Zhong et al., 2020b; Ragusa et al., 2021).

Like any machine or deep learning algorithm, CNNs and detectors require datasets for training to discriminate between object classes. A major challenge in this domain is the availability of large and relevant datasets. Gold standard datasets for object detection and segmentation, such as ImageNet (Deng et al., 2009; Krizhevsky et al., 2012), ALOI (Amsterdam Library of Object Images) (Geusebroek et al., 2005), and Microsoft COCO (Common Objects in Context) (Lin et al., 2014), can be found in literature, but they are not specifically tailored to objects that can be manipulated by prosthetic hands. To cope with this limitation, three main strategies have been proposed:

- *output suppression*: since the largest datasets, such as the COCO dataset, contain images related to classes that are not of interest to prosthetic applications, these classes were suppressed in the model output (Park et al., 2022; Cirelli et al., 2023; Deshmukh et al., 2023; Peng et al., 2024).
- *custom dataset*: custom datasets, specific for the application, were built, considering only objects of interest (He et al., 2019; Zhong et al., 2020b; He et al., 2020; Huang et al., 2022; Weiner et al., 2022; Starke et al., 2022; Cui et al., 2022; Zhang et al., 2024; A Powered Prosthetic Hand with Vision System, 2025) or directly labeling the associated hand gesture (Roy et al., 2021; Ragusa et al., 2021; Castro et al., 2022; Karrenbach et al., 2022; Huang et al., 2023).
- *Transfer learning*: since neural networks in general, including convolutional ones, require a large amount of data to train, to avoid falling into overfitting problems, transfer learning could be a solution. The idea is to fine-tune a CNN already trained on another dataset, to "transfer" the previously acquired knowledge

Table 1
Main information from the first 18 papers included in this review related to upper limb prostheses. GS = Grasp Selection; WO = Wrist Orientation; IR = Intention Recognition; DS = Dataset Size (images); DA = Data Availability; NA = Not Available; AB = Able-Bodied; TRA = Trans-Radial Amputee; THA = Trans-Humeral Amputee; EP = Epileptic; NS = Non-Sighted.

REF	Aim	Prosthesis	Computer Vision				Subject
			Hardware	Position-ing	Algorithm & Performance	Dataset	
Došen and Popović (2011)	GS & WO	/	RGB, 320 × 240 px, benchtop test	/	Segmentation pipeline, 8 objects, 4 grasps & P/S	DS:NA, DA:NA	5 AB
Došen et al. (2010)	GS	CyberHand	RGB, 320 × 240 px, benchtop test	Hand (back)	Segmentation pipeline, 18 objects, 4 grasps	DS:NA, DA:NA	13 AB
McMullen et al. (2013)	GS	Custom (17 DoF)	RGB-D, benchtop test	Head	Segmentation pipeline, spherical objects	DS:NA, DA:NA	2 EP
Markovic et al. (2014)	GS & WO	IH2 Azzurra (Prensilia, 2025)	RGB, 640 × 480 px, 30 Hz, benchtop test	Head	Segmentation pipeline & fitting with geometric primitives, 20 objects, 4 grasps & P/S	DS:NA, DA:NA	13 AB
Markovic et al. (2015)	GS & WO	Michelangelo (Ottobock, 2025a)	RGB-D, 640 × 480 px RGB, 320 × 240 px depth images, 30 Hz, benchtop test	Head	Segmentation pipeline & fitting with geometric primitives, 10 objects, 2 grasps & P/S	DS:NA, DA:NA	10 AB, 1 TRA
Ghazaei et al. (2017)	GS	i-Limb Ultra (Össur, 2025)	RGB, 48 × 36 px, benchtop test	Hand (back)	CNN, 21 objects, 4 grasps, 84%, 150 ms	DS:5112, DA:🔗	2 TRA
He et al. (2019)	GS & WO	Gripper (3 DoF)	RGB, 416 × 416 px, benchtop test	Hand (back)	CNN, 5 objects, 6 gestures, 99%, 28.6 FPS(36 ms)	DS:600, DA:NA	1 AB
Zhong et al. (2020b)	WO	Gripper (2 DoF)	RGB, 224 × 224 px, benchtop test	Hand (external)	CNN (feature extraction) + BNN (uncertainties quantification), 3 objects, 82.42%, 17.6 ms	DS:NA, DA:NA	2 AB
He et al. (2020)	GS	Gripper (3 DoF)	RGB, benchtop test	Hand (palm)	CNN, 10 objects, 33 ms (30 FPS)	DS:600, DA:NA	1 AB
Roy et al. (2021)	WO	/	RGB, 572 × 218 px, benchtop test	Head	Segmentation pipeline, 6 objects, 2 grasps, 22 ms	DS:NA, DA:NA	15 AB
Mouchoux et al. (2021)	GS & WO	Michelangelo (Ottobock, 2025a)	RGB-D, 1920 × 1080 px RGB, 640 × 480 px depth images, 30 Hz, benchtop test	Head	Segmentation pipeline & fitting with geometric primitives, 6 objects, 2 grasps & P/S	DS: NA, DA:upon request	3 AB, 2 TRA
Ragusa et al. (2021)	GS	/	RGB, 5 MP, Jetson TX2, Jetson Nano, smartphones	/	CNNs, 17 objects, 3 affordance classes	DS:30000, DA:NA	12 AB
Fejér et al. (2021)	GS	/	RGB, 480 × 480 px, RGBD, Xilinx ZCU102 FPGA	Head + Hand (NA)	SIFT (recognition of keypoints) avg. precision = 0.84 , avg. recall = 0.94	DS: 3860, DA:🔗	/
Castro et al. (2022)	GS	Kwawu Arm 2.0	RGB, 96 × 96 px, Raspberry Pi 3	Hand (palm)	CNN, 25 objects, 5 grasps, 99%, 250 ms + 350 ms	DS: 25713, DA:🔗	/
Castro and Dosen (2022)	GS & WO	Michelangelo (Ottobock, 2025a)	RGB-D, 424 × 240 px depth images, 90 Hz, benchtop test	Hand (back)	Segmentation pipeline & fitting with geometric primitives, 16 objects, 2 grasps & P/S, 97%, 5–16 Hz	DS:NA, DA:NA	10 AB
Huang et al. (2022)	GS & WO	Custom (8 DoF)	RGB-D, benchtop test	Elbow	CNN, 7 objects, 4 grasps, 93.3%	DS:NA, DA:NA	10 AB, 3 THA
Weiner et al. (2022), Starke et al. (2022)	GS & WO	KIT prosthesis (Weiner et al., 2018)	RGB, 76 × 76 px, on-board system with a 400 MHz ARM Microcontroller	Hand (palm)	CNN, 13 objects, 2 grasps, 96.51%, 115 ms	DS:3900, DA:NA	1 AB

Table 2
Main information from the other 15 papers included in this review related to upper limb prostheses. GS = Grasp Selection; WO = Wrist Orientation; IR = Intention Recognition; DS = Dataset Size (images); DA = Data Availability; NA = Not Available; AB = Able-Bodied; TRA = Trans-Radial Amputee; THA = Trans-Humeral Amputee; EP = Epileptic; NS = Non-Sighted.

REF	Aim	Prosthesis	Computer Vision				Subject
			Hardware	Position-ing	Algorithm & Performance	Dataset	
Park et al. (2022)	GS	Custom	RGB-D, 640 × 480 px, benchtop test	Head	CNN, 4 objects, 3 grasps, 87%, 25 ms	DS:5200, DA:NA	10 AB
Karren-bach et al. (2022)	WO	/	HTC Vive Pro Eye(depth), Leap Motion Controller, benchtop test	Head	CNN, 30 objects, 4 wrist gestures, 75.2%	DS:NA, DA:NA	6 AB
Cognolato et al. (2022, 2020)	GS	/	MeganePro dataset, benchtop test	Head	CNN, 18 objects, 10 grasps, 85%	DS:NA, DA:🔗	30 AB, 15 TRA
Wang et al. (2022)	GS	/	MeganePro (Cognolato et al., 2020) dataset, benchtop test	Head	CNN, 18 objects, 10 grasps, 91.59%, 40 ms	DS:1186, DA:🔗	30 AB
Cui et al. (2022)	IR	Gripper (1 DoF)	RGB, 1920 × 1080 px, benchtop test	Hand (finger)	CNN, 6 objects, 2 classes, 92.4%	DS:3000, DA: upon request	12 AB
Cirelli et al. (2023)	GS & WO	/	RGB, 320 × 240 px, Raspberry Pi 4B	Hand (palm)	CNN, 16 objects, 10 grasps & P/S, 97.85%, 480 ms	DS:330000, DA:🔗	1 AB
Huang et al. (2023)	GS	/	RGB-D object dataset (Lai et al., 2011), Hit-GPRec dataset (Shi et al., 2020), benchtop test	/	DL-Net, 300 objects (RGB-D), 121 objects (Hit), 4 grasps	DS:207921(Lai et al., 2011), DA:🔗	/
Deshmukh et al. (2023)	GS	/	RGB 224 × 224 px, benchtop test	/	CNN, 17 objects, 20 grasps, 94.55%	DS:NA, DA:NA	20 AB, 2 TRA
Zandigo-har et al. (2024)	GS	Custom	RGB 1280 × 720 px, 60 Hz, benchtop test	Head	CNN, 54 objects (11 classes), 14 grasps, 81.46%	DS:89700, DA:NA	5 AB
Zhang et al. (2024)	GS	Custom	RGB 1920 × 1080 px, 30 Hz, benchtop test	Head	CNN, objects not specified, 8 grasps, 96.4%, 40 ms	DS: 6340, DA:upon request	12 AB
Peng et al. (2024)	GS & WO	Custom (Fan et al., 2025)	RGB-D, 1280 × 720 px for depth, 1920 × 1080 px RGB, 90 Hz, NVIDIA Jetson Orin NX 480 × 480 px	Head + Hand (palm)	CNN, 1 object, 1 grasp & PS, FE, 40 ms	DS:NA, DA:NA	8 AB, 4 NS
Kyberd et al. (2023)	GS & WO	not specified	RGB, 720 × 540 px. benchtop test	Head	Segmentation pipeline, 33 ms	DS:NA, DA:NA	14 AB, 4 TRA
A Powered Prosthetic Hand with Vision System (2025)	GS & WO	Ispire RH56BF3-2R	RGB-D, benchtop test	Head	CNN, and 3D reconstruction, 8 object, 8 grasp	DS:NA, DA:NA	7 AB, 3 TRA
Gardner et al. (2020)	GS	Bebionic V2 hand Ottobock (2025b)	RGB, 640 × 480 px, benchtop test	Head	Segmentation pipeline & KNN, 3 objects, 3 grasps	DS:NA, DA:NA	10 AB, 1 TRA

to other classes. In this way, in Ghazaei et al. (2017), Cognolato et al. (2022), Wang et al. (2022), Zandigohar et al. (2024) it was possible to use a small custom dataset to test the authors' approach.

Among the most widely used datasets in prosthetic applications, the Megane Pro dataset is one of the largest and freely available. It includes sEMG, eye tracking, and scene camera data from 30 able-bodied subjects and 15 trans-radial amputees (Cognolato et al., 2020), who grasped and manipulated 18 objects using 10 common grasps. This dataset was used in Cognolato et al. (2022), Wang et al. (2022) for fine-tuning pre-trained models. In contrast, the RGB-D Object (Lai et al., 2011) and Hit-GPRec (Shi et al., 2020) datasets consist of images of everyday objects, labeled with hand gestures. The former includes 300 objects labeled in 4 gestures (palmar wrist neutral, palmar wrist pronated, pinch, and tripod), while the latter contains 121 objects labeled in cylindrical, lateral, spherical, and tripod grasps. In Huang et al. (2023), the same model was trained and tested on these datasets separately.

Despite these resources, there is no standard regarding the number of objects or gestures, and a wide variability is observed across the reviewed works (see Tables 1, 2). Moreover, only a minority of the

studies clearly report the size of the datasets employed, and even fewer make them publicly available, which limits reproducibility and cross-study comparisons.

3.1.3. High-level prosthesis control

The differences in control strategies presented in the literature relate to the type of task (grasp selection, wrist orientation, or both) and the validation metrics, as detailed below.

In the selected papers, computer vision in prosthetic control is used for grasp selection (Došen et al., 2010; McMullen et al., 2013; Ghazaei et al., 2017; He et al., 2020; Ragusa et al., 2021; Castro et al., 2022; Park et al., 2022; Huang et al., 2023; Deshmukh et al., 2023; Zandigohar et al., 2024; Zhang et al., 2024; Gardner et al., 2020), wrist orientation (Zhong et al., 2020b; Roy et al., 2021; Karrenbach et al., 2022), or both (Došen and Popović, 2011; Markovic et al., 2014, 2015; He et al., 2019; Mouchoux et al., 2021; Castro and Dosen, 2022; Huang et al., 2022; Weiner et al., 2022; Starke et al., 2022; Cognolato et al., 2022; Wang et al., 2022; Cirelli et al., 2023; Peng et al., 2024; Kyberd et al., 2023; A Powered Prosthetic Hand with Vision System, 2025). Exceptions include two papers that focus on sub-tasks: (Fejér et al., 2021) proposed a Scale Invariant Feature Transform (SIFT) algorithm to map data from two cameras (one on the head, one on the prosthesis) to retrieve object coordinates, while Cui et al. (2022) aimed to determine when the prosthesis should perform the grasping task using posture and visual information. Regardless of the task, the control system always takes user motion intention and object features as input, providing control commands and sometimes visual feedback (e.g., LEDs or screens integrated into the prosthesis).

User motion intention is typically decoded using surface EMG, although in some cases it simply triggers the control pipeline (Došen and Popović, 2011; Došen et al., 2010; Markovic et al., 2014, 2015; Ghazaei et al., 2017; He et al., 2019; Zhong et al., 2020b; He et al., 2020; Castro et al., 2022; Castro and Dosen, 2022; Weiner et al., 2022; Starke et al., 2022; Kyberd et al., 2023). In McMullen et al. (2013), Zhang et al. (2024), intracranial electroencephalography (iEEG) was used to decode motion intention, with the approach tested in McMullen et al. (2013) on two human subjects undergoing intracranial recordings for epilepsy surgery. In Park et al. (2022), the distance between the robotic hand and the object triggers the grasp action, while in Cui et al. (2022), intention is recognized using inertial sensors (IMUs). Specifically, posture angle data from three IMUs placed on the forearm, arm, and hand trigger a miniature camera to capture the image of the target object. Recently, the combination of EMG and CVS for prosthetic control has been explored to improve accuracy compared to EMG classifiers alone (Mouchoux et al., 2021; Huang et al., 2022; Cognolato et al., 2022; Wang et al., 2022; Cirelli et al., 2023; Deshmukh et al., 2023; Zandigohar et al., 2024).

In Gardner et al. (2020), movement intention is decoded by combining mechanomyography (MMG), IMUs, and computer vision. MMG detects muscle activation, triggering the system when a threshold is exceeded and no major arm motion is detected. A head-mounted camera identifies the object, while inertial sensors track arm movement. These signals are fused and classified to predict the intended grasp, enabling intuitive control of a robotic prosthesis.

In Peng et al. (2024), the intention for movement in the developed system for visually impaired amputees is primarily encoded through voice interaction module. The user issues the voice instruction (VI) of "environment recognition", and the system is able to identify objects within the global field of vision through the environmental perception module, subsequently broadcasting this information to the user.

In A Powered Prosthetic Hand with Vision System (2025), a Motion Trajectory Regression-based Grasping Intent Estimation (MTR-GIE) was defined to predict user intent in multi-object environments by regressing wrist trajectories and spatially segmenting candidate objects, without the use of EMG sensors. Moreover, the Spatial Geometry-based Gesture Mapping (SG-GM) method enables the prosthetic hand to perform smooth, human-like finger movements based solely on visual input. It works by modeling the angles of each finger as a function of the distance between the hand and the object, using gesture data collected from real human grasps. These functions allow the prosthetic to dynamically adjust its posture as it approaches the object, achieving natural, continuous gesture transitions. In the remaining works, the intention recognition module is not specified or simply not yet considered (Ragusa et al., 2021; Fejér et al., 2021; Karrenbach et al., 2022; Huang et al., 2023).

Regarding the multimodal nature of the proposed control strategies, Weiner et al. (2022), Starke et al. (2022) combined vision, surface EMG with IMUs data embedded in the prosthetic hand to estimate its orientation and enhance wrist rotation management. In contrast, Mouchoux et al. (2021) integrated the IMU into the socket to measure the prosthetic hand's absolute orientation relative to the local coordinate system, enabling wrist adjustments based on the participant's forearm orientation.

Few approaches in the literature propose feedback systems for users, with most relying on visual information to convey insights into the control system output (Markovic et al., 2014; Mouchoux et al., 2021; Castro et al., 2022; Weiner et al., 2022; Starke et al., 2022; Cirelli et al., 2023; Zhang et al., 2024). For example, in Markovic et al. (2014), Zhang et al. (2024), Mouchoux et al. (2021) Augmented Reality (AR) glasses was used: in the first two works, these displayed information about the object to handle and the type of grasp to execute, while in the last one, the interface provided visual feedback during object selection and detection of myoelectric inputs. Detected objects appeared as yellow holograms with green faces indicating the predicted approach side, and turning all green when selected. In Castro et al. (2022), Cirelli et al. (2023), LEDs provided feedback: the former displayed the selected grasp, while the latter indicated object recognition and discrepancies between outputs from the EMG classifier and SCS. In Weiner et al. (2022), Starke et al. (2022), a monitor on the prosthetic hand displayed the camera image, showing bounding boxes around recognized objects and their associated grasp types. While AR offers advanced feedback, its application in daily life is challenging due to the high cost, technical requirements, and potential user discomfort (Zhan et al., 2020; Dargan et al., 2023). Similarly, integrating a display on the hand would require substantial modifications to commercial prosthetic designs and user training. Conversely, LEDs offer a simpler and more intuitive feedback mechanism for users.

In Peng et al. (2024), the system provides feedback for non-sighted amputees using auditory cues, vibrations, and spatial sound sources. A Voice Interaction module communicates environmental conditions and, upon receiving the target object information from the user, activates a sound source beneath the specified object. Additionally, vibration feedback assists during the approach phase along with audio prompts, and confirms a successful grasp by activating all vibration motors in the feedback bracelet.

To assess the effectiveness of the control strategies, specific tasks and metrics have been defined per each scientific work. Specifically, 13 performance indicators were computed in the validation of the proposed approaches (see Table 3 reporting the metrics along with their unit of measurements and description) when performing manipulation tasks.

Table 4 details the tasks and metrics chosen for the validation of proposed control strategies.

3.2. Lower-limb prosthesis

In the 17 selected studies, the common objective is to leverage visual information from a CVS to extract environmental features in front of the user, enhancing the control of active DoFs in a robotic leg. However, while feasibility studies exist, only Hong et al. (2023) and Zhang et al. (2020) presents the full integration of a CVS into active lower-limb prostheses for high-level control.

Table 3
Metrics chosen for the validation of the approaches proposed in the reviewed papers.

Metric	Ext. Name	Description
RSE [%]	Relative Size Error	The absolute value of the object size error, normalized by the correct object size.
ADE [cm]	Average Distance Error	Difference between the reference distance and the estimated distance of the hand from the object
AP [%]	Accuracy in Pre-shaping	Accuracy in the estimation of the grasp type and the size of the object
SR [%]	Success Rate	Percentage of success of the task
TAT [s]	Time to Accomplish Task	Total time spent to accomplish the task.
APAE [cm]	Average Prosthesis Aperture Error	Average size estimation error of the object, which is linked with the aperture of the hand
NPC	Number of Prediction Changes	Number of changes of the predicted object orientations along one grasping trajectory
A [%]	Accuracy	Accuracy in correctly recognizing grasp or object classes
IT [ms]	Inference Time	Time necessary for the model to process an image and make a prediction
GFS	Grasping Functionality Score	Output from the YCB Gripper Assessment Protocol (Calli et al., 2015)
AE [°]	Angular Error	Difference between the reference orientation angle and the one computed by the algorithm
AES [°]	Angular Estimation Stability	Standard deviation in orientation estimation of the object
PHAM	Prosthetic Hand Assessment Measure	Weighted sum of the 2D translational displacement of the hand, 3D deviation of the chest, and 3D deviation of the shoulder

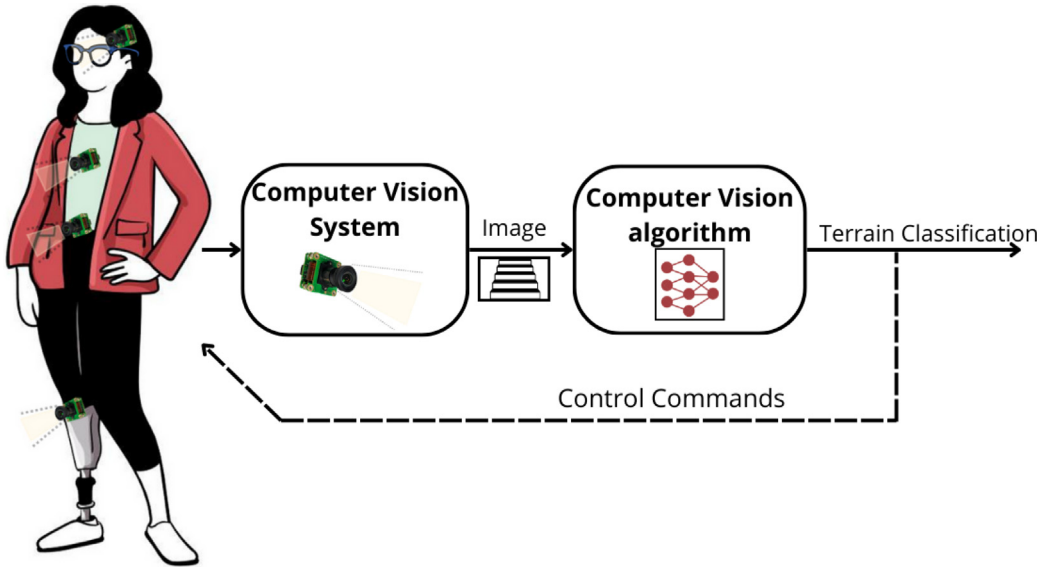

Fig. 4. Block diagram of the system with the lower-limb prosthesis.

It is feasible to establish a general system architecture, as shown in Fig. 4, despite differences among the proposed approaches. Table 5 summarizes the selected scientific papers on the application of CVS in lower-limb prosthetic control. Similar to the upper limb studies, the table details each work aim, prosthetic device, CVS hardware and algorithmic components, sensor positioning on the user, and the subjects involved in experimental validation. Accuracy and computational burden, defined as the time in milliseconds needed to execute the algorithm, are reported when available.

3.2.1. Computer vision system

As outlined for upper-limb prosthetics, approaches utilizing CVS for lower-limb prostheses also vary in terms of technology used (RGB or RGB-D), sensor positioning, and technical characteristics, such as image resolution, acquisition frequency, and computational unit. The primary visual sensing modalities explored in the literature are RGB (Zhong et al., 2020a; Laschowski et al., 2022; Li et al., 2022; Sharma and Rombokas, 2022; Li et al., 2023; Kurbis et al., 2024b,a) and RGB-D (Krausz et al., 2015; Zhang et al., 2019a,b; Krausz and Hargrove, 2021; Krausz et al., 2019; Al-Dabbagh and Ronsse, 2022; Chen et al., 2022; Hong et al., 2023; Zhang et al., 2020; Song et al., 2025). Regarding sensor placement, cameras have been positioned at the knee (Zhang et al., 2019a,b; Zhong et al., 2020a; Li et al., 2022; Chen et al., 2022; Zhang et al., 2020), chest (Krausz et al., 2015; Al-Dabbagh and Ronsse, 2022; Laschowski et al., 2022; Kurbis et al., 2024b), waist (Krausz and Hargrove, 2021; Krausz et al., 2019; Hong et al., 2023), and head (Zhong et al., 2020a; Li et al., 2022; Sharma and Rombokas, 2022; Song et al., 2025). In particular, Zhong et al. (2020a), Li et al. (2022)

Table 4
Tasks and metrics used to assess the performance of the control strategies and obtained results. CSI = Cluttered Scene Interaction.

REF	Task [Metrics]	Performance
Došen and Popović (2011)	Size test (static) [RSE]	RSE = 11%
	Size & Orientation test (dynamic) [RSE, ADE, AE]	RSE = 22%, ADE = 2.9 cm, AE = 1.52°
Došen et al. (2010)	Pick & Place [AP, SR, TAT]	ADE = 2.9 cm, AE = 1.52°
McMullen et al. (2013)	Reach & Grasp [SR]	SR = 71.4% (Subject 1) & 67.7% (Subject 2)
Markovic et al. (2014)	Pick & Place [AP, SR, TAT]	AP = 94%, SR = 81%, TAT = 3.47 s
Markovic et al. (2015)	Pick & Place [TAT, APAE, AE]	TAT = 5.9 s ± 1.9 , APAE = 0.75 ± 1.1 cm, AE = 9 ± 5°
Ghazaei et al. (2017)	Pick & Place [SR, TAT]	SR = 88%, TAT = 16.7 ± 9.3 s (Subject 1) & 19.3 ± 25.7 s (Subject 2)
He et al. (2019)	Pick & Place [/]	/
Zhong et al. (2020b)	Pick & Place [SR, NPC, TAT]	SR = 100%, NPC = 3, TAT = 3.5 s
He et al. (2020)	/	/
Roy et al. (2021)	Reach & Grasp [SR]	SR = 78.4%
Mouchoux et al. (2021)	CSI [TAT]	TAT (CSI) = 50.5 s (AB) & 57.0 s (A)
	SHAP [/]	/
Ragusa et al. (2021)	Performance on diff. Hw/Sw [A, IT]	A = 0.99 background, 0.85 grasp & 0.93 no-grasp, IT = 21 ms jetson TX2, 100 ms Jetson nano, 333 ms phone
Fejér et al. (2021)	/	/
Castro et al. (2022)	/	/
Castro and Dosen (2022)	Pick and Place [TAT]	TAT = 6.8 ± 1.8s
	CSI [TAT]	TAT = 7.8 ± 1.9s
Huang et al. (2022)	3C Assembly task [SR, TAT]	SR = 85% , TAT >1 min
Weiner et al. (2022), Starke et al. (2022)	YCB Gripper Assessment Protocol [GFS]	GFS = 91.8%
	ADLs [SR, TAT]	SR = 88.6%, TAT <8*TAT for AB
Park et al. (2022)	Reach & Grasp [SR]	SR = 89%
Karrenbach et al. (2022)	Pick & Place [TAT, PHAM]	TAT = 4.73s, PHAM <5
Cognolato et al. (2022, 2020)	Reach & Grasp [A]	A = 78.64 ± 6.13%
	Reach & Grasp [A]	A = 74.12 ± 8.87%
Wang et al. (2022)	Pick & Place [A]	A = 90.06%
Cui et al. (2022)	ADLs [SR]	SR = 92.4%
Cirelli et al. (2023)	Static condition test [A, AE, AES]	A = 99.80%, AE = 16.26 ± 8.62°, AES = 0.2°
	Approaching Condition test [SR]	SR = 100%
Huang et al. (2023)	Test with samples never seen (BOC) [A]	A-WWC = 99.20% (RGBD) & 83.53% (Hit)
	test with different angles (WWC) [A]	A-BOC 74.18%(RGBD) & 76.52%(Hit)
Deshmukh et al. (2023)	/	/
Zandigohar et al. (2024)	Pick & Place [A]	A = 95.3%
Zhang et al. (2024)	ADLs [TAT, SR]	TAT = 39.6 ± 6.0s, 24.5 ± 10.1s, 50.5 ± 6.0s, 26.2 ± 11.2s, SR = 100% except for one subject
Peng et al. (2024)	Reach & Grasp [SR, TAT]	SR = 90.6%, TAT = 32.47s
Kyberd et al. (2023)	SHAP [TAT]	TAT = 4.8 ± 1.5s (AB), TAT = 20 ± 12s (A)
A Powered Prosthetic Hand with Vision System (2025)	Reach & Grasp [SR, TAT, A]	SR = 95.43% (single-object), 88.75% (multi-object) , TAT = 1.47 ± 0.15s (Human), TAT = 3.07 ± 0.41s (Prosthesis), A = 94.35% (multi-object)
Gardner et al. (2020)	Pick & Place [A]	A = 90.46%

employed a vision fusion strategy, integrating data from augmented reality glasses with that from lower-limb cameras.

Since the limited number of works integrating computer vision into the prosthetic control loop, data processing is usually performed on personal computers. However, some studies have tested portable solutions: in Zhong et al. (2020a), Li et al. (2023) and Song et al. (2025), image analysis was validated online using a Single-Board Computer (SBC), and in Kurbis et al. (2024b,a), an iOS application on a smartphone was developed for real-time stair recognition.

3.2.2. Computer vision algorithm

As mentioned before, the information acquired by the camera sensor is processed with computer vision algorithms to extract features of the environment in front of the user. The primary objectives of these strategies include terrain classification (Zhang et al., 2019a,b; Zhong et al., 2020a; Krausz and Hargrove, 2021; Krausz et al., 2019; Al-Dabbagh and Ronsse, 2022; Laschowski et al., 2022; Li et al., 2022; Sharma and Rombokas, 2022; Chen et al., 2022; Li et al., 2023; Zhang et al., 2020), and more specialized tasks such as stair recognition (Kurbis et al., 2024b,a) and segmentation (Krausz et al., 2015). Exceptions are presented in Sharma and Rombokas (2022) and Hong et al. (2023), where environmental information is combined with data from IMUs to enhance the prediction and estimation of joint-space trajectories in unstructured environments. In the former, the joint kinematics, obtained with 17 IMUs placed on the body of the subject, is integrated with visual information coming from a head-mounted camera. These two information are fed into 2 different Long Short-Term Memory (LSTM) networks, and their outputs are combined using fusion layers to predict knee and ankle joint angles. In the latter, data from three IMUs (attached to the thigh, shank, and foot of the healthy limb) are

Table 5
Main information from the papers included in this review related to lower limb prostheses. SR = Stair Recognition; TC = Terrain Classification; JAE = Joint Angle Estimation; DS = Dataset Size (images); DA = Data Availability; NA = Not Available; AB = Able-Bodied; A = Amputee.

REF	Aim	Prosthesis	Computer Vision				Subject
			Hardware	Position-ing	Algorithm & Performance	Dataset	
Krausz et al. (2015)	SR	/	RGB-D, (320 × 240 px, 160 × 120 px, 120 × 90 px, 80 × 60 px), 30 Hz, benchtop test	1.5 m (chest)	Segmentation pipeline, 98.8%, 1 class, stair distance error 6 cm	DS:NA, DA:NA	1AB
Zhang et al. (2019a)	TC	/	RGB-D, 224 × 171 px, 15–25 Hz, benchtop test	Knee	CNN, 5 classes, simulation 100%, indoor 99.3%, outdoor 98.5%, 20ms	DS:7500, DA:NA	6AB, 3A
Zhang et al. (2019b)	TC	/	RGB-D, 100 × 100 px, 15–25 Hz, benchtop test	Knee	CNN+HMM, 5 classes, indoor 97.3% and 67 ms, outdoor 96.4% and 733 ms	DS:7500, DA:NA	5AB, 3A
Zhang et al. (2020)	TC & JAE	Custom (2 DoFs)	RGB-D, 100 × 100 px, 30 Hz, benchtop test	Knee	CNN+HMM, 6 classes, 94%, 23 ms	DS:7500, DA:NA	1AB, 4A
Zhong et al. (2020a)	TC	/	RGBs, 1240 × 1080 px, 25 Hz, Raspberry Pi 3B	Head and Knee	CNN+ Bayesian MLP, 6 classes, 95.38%, 80ms	DS:327000, DA:🔗	7AB, 1A
Krausz and Hargrove (2021), Krausz et al. (2019)	TC	RIC/ Michigan	RGB, 224 × 171 px, up to 45 Hz. benchtop test	Waist	LDA, 5 classes, 99%	DS:NA, DA:NA	12AB, 1A
Al-Dabbagh and Ronsse (2022)	TC	/	RGB-D, 1280 × 720 px, up to 90 Hz, benchtop test	Chest	SVM, 5 classes, 95.0%, >66 ms	DS:NA, DA:NA	8AB
Laschowski et al. (2022)	TC	/	Smartphone,1280 × 720 px, 30 Hz	Chest	CNNs, 12 classes, 72.9%, 2.2ms	DS:923000, DA:🔗	1AB
Li et al. (2022)	TC	/	RGB, 960 × 540 px, 50 Hz, RGB-D, 30 Hz, benchtop test	Head and Knee	CNN, 5 classes, 96.0%, 5.7ms	DS:27868, DA: 🔗	3AB, 2A
Sharma and Rombokas (2022)	JAE	/	RGB, 320 × 240 px, 30 Hz, benchtop test	Head	LSTM, <0.13 normalized RMSE	DS:NA, DA:NA	23AB
Chen et al. (2022)	TC	/	RGB-D, 224 × 171 px, 15–25 Hz, benchtop test	Knee	UDA, 5 classes, 98.1%, 26ms	DS:7500, DA:NA	6AB, 3A
Li et al. (2023)	TC	/	RGB, 224 × 224 px, Arduino UNO	Shank and Thigh	CNN, 3 classes, 97.4%, 67.0ms	DS:923000, DA:🔗	5AB, 2A
Kurbis et al. (2024b,a)	SR	/	Smartphone, 224 × 224 px, 30 Hz	Chest	CNN, 4 classes, 98.4%, 2.75ms	DS:51500, DA:🔗	/
Hong et al. (2023)	JAE	Custom (2 DoFs)	RGB-D, 1024 × 1024 px, 90 Hz, benchtop test	Waist	CNN, 5 obstacle classes	DS:1600, DA:NA	1AB, 2A
Song et al. (2025)	JAE	/	RGB-D, 1024 × 1024 px, 90 Hz, Nvidia Jetson AGX Orin	Head	CNN, 5 obstacle classes	DS:NA, DA:NA	AB

combined with vision-based information to generate obstacle avoidance trajectories. A Mask R-CNN is trained to identify and extract geometric features (e.g., height and distance of obstacles) for five obstacle types. This data is then input into improved dynamic movement primitives with type-2 fuzzy models (T2FDMPs) to generate joint-space trajectories, enabling a custom 2-DoF powered lower-limb prosthesis to perform obstacle avoidance.

Concerning the computer vision algorithms explored in the literature, in Krausz et al. (2015) a segmentation pipeline was used to segment stairs in the image. Nevertheless, for the terrain classification task, CNNs have an important role, as demonstrated by their use in several proposed approaches (Zhang et al., 2019a,b; Zhong et al., 2020a; Laschowski et al., 2022; Li et al., 2022, 2023; Kurbis et al., 2024b,a; Hong et al., 2023; Zhang et al., 2020; Song et al., 2025). Within this context, different families of CNN-based algorithms have been exploited: lightweight CNNs, such as MobileNet (Sinha and El-Sharkawy, 2019) or StairNet (Kurbis et al., 2024a), which provide

real-time performance suitable for navigation tasks,are the most exploited. Instead One-stage detector have been primarily exploited in obstacle avoidance tasks, where it is necessary to quickly detect obstacles to be avoided, and real-time performance is critical (Hong et al., 2023; Zhang et al., 2020).

Concerning the computer vision algorithms explored in the literature, in Krausz et al. (2015) a segmentation pipeline was used to segment stairs in the image. Nevertheless, for the terrain classification task, CNNs have an important role, as demonstrated by their use in several proposed approaches (Zhang et al., 2019a,b; Zhong et al., 2020a; Laschowski et al., 2022; Li et al., 2022, 2023; Kurbis et al., 2024b,a; Hong et al., 2023; Zhang et al., 2020; Song et al., 2025). As outlined in Table 5, more traditional machine learning approaches have also been explored in the literature for lower-limb prosthetic control. For example, in Krausz and Hargrove (2021), Krausz et al. (2019), the researchers applied Linear Discriminant Analysis (LDA) to classify five different types of terrain, using features extracted from various sensing modalities such as IMUs, an RGB-D camera mounted on the waist, EMG sensors, and goniometers. LDA helped to reduce the dimensionality of the feature set, achieving a high classification accuracy of 99% under offline conditions. On the other hand, in Al-Dabbagh and Ronsse (2022), a Support Vector Machine (SVM) was used to classify the same five terrain types. The researchers obtained an average accuracy of 95%, with the additional challenge of predicting the terrain class for the next three walking steps, under partial occlusion conditions, and with data collected in both indoor and outdoor environments.

As outlined in 3.1.2, machine and deep learning models require a large and well-labeled dataset to be trained, so researchers constructed custom datasets to train Zhang et al. (2019a), Krausz and Hargrove (2021), Krausz et al. (2019), Al-Dabbagh and Ronsse (2022), Laschowski et al. (2022), Li et al. (2022), Chen et al. (2022), Hong et al. (2023) or fine-tuned pre-trained models (Zhang et al., 2019b; Zhong et al., 2020a; Zhang et al., 2020). It is important to focus on (Laschowski et al., 2022), where the researchers created the ExoNet dataset, which consists of 923,000 RGB images labeled with 12 different terrain transition classes. The images were collected using a camera mounted on the chest of a subject walking and recording 52 h of video in both indoor and outdoor conditions. This dataset was used in Li et al. (2023) and Kurbis et al. (2024b,a), where the images were manually labeled with specific terrain classes to train CNNs.

Another possibility to fulfill the limits of supervised learning approaches is the generation of a labeled synthetic dataset by simulation, as proposed in Chen et al. (2022). Researchers mitigated the gap between real and simulated data using Unsupervised Domain Adaptation (UDA), which was used to better generalize models trained in a label domain, i.e. simulation, to a label-free target domain, i.e. real world.

3.2.3. High-level prosthesis control

As previously mentioned, a control strategy integrating Computer Vision Systems (CVS) for powered lower-limb prostheses was proposed and validated only in Hong et al. (2023) and Zhang et al. (2020). In the former, a custom device with two actuated DoFs (knee and ankle) was controlled to avoid obstacles recognized and segmented using a Mask R-CNN model. Three subjects, i.e. two amputees and an able-bodied one, were enrolled for the experimental validation, which consists of avoiding different obstacles during walking. The performance was evaluated using three metrics: average time for trajectory generation, distance between the obstacle avoidance and original trajectories (healthy limb), and norm of accelerations for different obstacle avoidance terms. Results showed good tracking of the desired trajectories, with knee and ankle position errors of less than 0.02 rad and 0.1 rad, respectively. A comparison with classical Dynamic Movement Primitives revealed that the time cost of T2FDMPs was longer, but they produced smoother trajectories with the smallest acceleration norm.

The latter study (Zhang et al., 2020) builds upon the environmental classification methods previously introduced in Zhang et al. (2019a) and Zhang et al. (2019b), extending them to real-time integration with the prosthesis control system. Specifically, an RGB-D and IMU were mounted on the prosthesis with a CNN–HMM pipeline for environmental classification and obstacle detection. A powered transfemoral prosthesis with active knee and ankle was controlled accordingly, enabling predictive adaptation to terrain changes and obstacle crossing. Experimental validation with four amputees and one able-bodied subject included walking on everyday terrain and overcoming static and dynamic obstacles. Results demonstrated classification accuracies above 94% for terrain recognition, correct and timely locomotion mode transitions (>98% accuracy), and safer, more natural obstacle negotiation compared to passive prostheses, with significantly reduced hip torques and lower tumbling risks.

4. Discussions

Based on the characteristics and results of the reviewed literature, it is evident that several challenges and pitfalls still need to be addressed. In particular, the discussion is organized into thematic subsections to present current strengths, recurring limitations, and open challenges for future research.

4.1. Optimizing camera choice and placement and designing embedded solutions for real-world testing

Concerning the computer vision system, the same considerations can be made for both upper and lower-limb prosthesis applications, starting from the choice of the sensor and its positioning. Fig. 5 highlights the heterogeneity in the choice of camera positions, and in the type of visual sensor exploited in literature. Although depth information would be very useful for improving the control of the prosthesis, this type of sensor is usually bulkier and more power-demanding compared to RGB cameras, limiting its use to approaches where the vision sensor is not integrated into the prosthetic device (Massalin et al., 2017). Even if the general trend regarding the development of RGB-D cameras is moving toward a reduction in size and weight, it must be considered that information coming from these sensors, i.e. point clouds of the framed scene, requires considerable computing power in embedded systems to be exploited in the real-time control of a prosthesis. This is the reason why the validation of all the approaches that considered RGB-D cameras (38%) was done in controlled benchtop conditions, which are far from real-world applications (McMullen et al., 2013; Markovic et al., 2015; Mouchoux et al., 2021; Fejér et al., 2021; Castro and Dosen, 2022; Huang et al., 2022; Park et al., 2022; Karrenbach et al., 2022; Huang et al., 2023; Peng et al., 2024; A Powered Prosthetic Hand with Vision System, 2025; Zhang et al., 2019a,b; Al-Dabbagh and Ronsse, 2022; Li et al., 2022; Chen et al., 2022; Hong et al., 2023; Zhang et al., 2020), except for (Song et al., 2025).

However, this choice is adopted also when RGB cameras were exploited: only the 24% of the works bring the implemented solution embedded into a compact system and in an operative scenario (Ragusa et al., 2021; Fejér et al., 2021; Castro et al., 2022; Cirelli et al., 2023; Peng et al., 2024; Zhong et al., 2020a; Li et al., 2023; Kurbis et al., 2024b,a; Weiner et al., 2022; Starke et al., 2022; Song et al., 2025). Most of the existing studies remain at a preliminary stage and rely on benchtop conditions, which provide greater computational resources. This limits the understanding of how systems would perform in real-time calculations and on devices with limited computing power, which are essential for seamless and practical use. Importantly, computational cost should not be overlooked regardless of the sensor type, with input resolution representing a key parameter that significantly affects it. Notably, only one study in the literature systematically compares the impact of two different image resolutions on both algorithm accuracy and execution time (Cirelli et al., 2023), aiming to reduce computational burden while preserving high performance. A more thorough

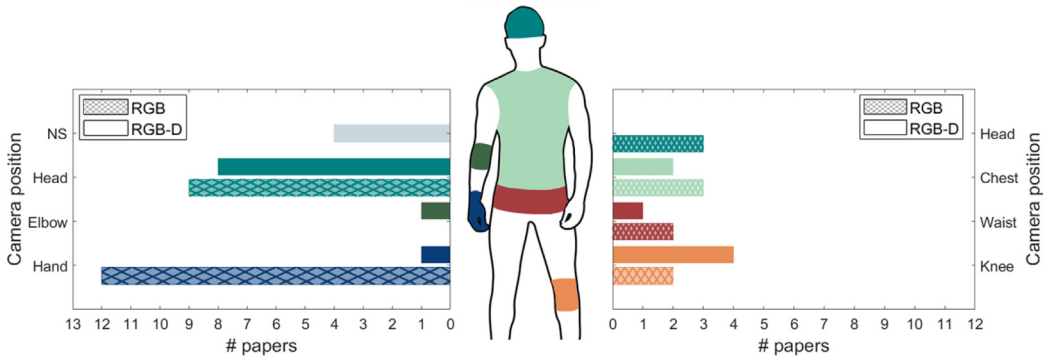

Fig. 5. Bar plots representing the different camera sensor positions explored in the literature for upper-limb (left) and lower-limb (right) prosthetic applications.

investigation on this front is clearly needed to guide future research in real-world implementation.

Regarding the positioning of the vision systems, the head seems to be the most frequent position for the computer vision system (see Fig. 5). Specifically, 21 works proposed methodologies starting from visual information coming from an egocentric view. Among the advantages of using such a configuration, there is the possibility of avoiding motion artifacts: whenever the camera is integrated inside the prosthetic device it moves together with the limb, creating distorted visual information that is not useful for algorithms processing input images (Zhang et al., 2019a). In addition, the integration of camera sensors at head level allows a much larger field of view to be captured (Roy et al., 2021; Sharma and Rombokas, 2022), which enables more usable information to be obtained to solve context prediction tasks (Li et al., 2022). However, the wide field of view may also introduce challenges in identifying the relevant area of interest. Finally, many commercial vision systems are compatible with the reviewed vision-based approaches, simplifying the design of the prosthetic limb that does not require ad-hoc modifications to accommodate the cameras. Nevertheless, integrating cameras at head level presents significant limitations due to several factors: (i) occlusion issues, as the prosthetic device itself may block the view by covering the objects the user intends to interact with; (ii) challenges in interfacing with the prosthetic device, given the physical separation between them; (iii) the need for accurate calibration to geometrically align the recorded visual data with the prosthesis; (iv) potential drawbacks in terms of overall size, wearability, and user comfort.

Integrating vision directly into limb prostheses mitigates these issues (Cirelli et al., 2023; Al-Dabbagh and Ronsse, 2022) and simplifies wiring for power and communication, reducing the need for complex setups (Weiner et al., 2022; Starke et al., 2022; Li et al., 2023). Integrating vision directly into the prosthetic limb facilitates calibration, namely the alignment of visual data with the prosthesis reference frame, thereby enabling accurate mapping of exteroceptive information into the workspace of the robotic system (Castro and Dosen, 2022; Zhang et al., 2019b). This configuration also provides a more compact and comfortable solution, as users only need to wear the prosthetic limb without requiring additional supports or devices. Nevertheless, on-prosthesis integration is not free from drawbacks. A major limitation is the reduced field of view compared to head-mounted systems, which may restrict the contextual information available for decision-making. Moreover, the prosthesis motion itself can introduce instability and motion artifacts in the sensor data, particularly during fast or abrupt movements (Zhong et al., 2020a). These aspects need to be carefully considered when designing vision-based control strategies, and may require advanced filtering, sensor fusion, or predictive approaches to ensure robust performance.

Only in Zhong et al. (2020a), a comparative analysis of camera positioning in terrain classification tasks was conducted, showing that lower limb cameras perform better for closer environments, while head-mounted cameras excel at predicting distant terrains. However, since amputees focus more on nearby terrains (Li et al., 2019), lower limb cameras are better suited to this population, as they are slightly influenced by the difference in gaits, and not by the height of the subject. Regarding upper limb prostheses, in Cirelli et al. (2023) the level of hand occlusion was analyzed at 4 different camera positions around the wrist, selecting the palm position as the optimal one. In this study, however, no reference is made to head-level positioning. Therefore, it is important to outline that a rigorous analysis of the impact of vision system positioning on the control strategies of prosthetic systems for both upper and lower limb applications is lacking in the literature.

4.2. Data generalization difficulty and the variability in input size

As regards the computer vision algorithms, CNNs are the most exploited in the literature, reaching 66%, followed by conventional segmentation pipelines (22%). From the reviewed works, it emerged that one-stage detectors and lightweight CNNs are the most widely adopted approaches, as they require relatively low computational power. Conversely, two-stage detectors (e.g., Faster R-CNN (Ren et al., 2016)) are generally disregarded, since in real-world and real-time applications, execution time is just as critical as algorithmic accuracy. Despite the chosen model, which can either be custom (Ghazaei et al., 2017; Kurbis et al., 2024b,a) or taken from the literature (Deshmukh et al., 2023; Zhang et al., 2019a), the significant variability in input size (i.e., image resolution) and training datasets makes it difficult to compare the results achieved by different approaches. Indeed, a major limitation lies in the challenge of data generalization due to the high dimensionality of vision input. While one of the key advantages of using computer vision in prosthetic control is its ability to capture not only the movements of the user but also rich contextual and environmental information, this benefit comes at the cost of increased data complexity. High-dimensional visual data demand significantly larger and more diverse datasets to ensure robust model generalization. In contrast to low-dimensional, body-centric signals such as EMG or IMU, vision-based systems must account for substantial variability in environmental context, object appearance, lighting conditions, and occlusions. As discussed in Section 3, current approaches often depend on samples from general-purpose datasets like ALOI and COCO, or on small-scale, custom datasets, highlighting the absence of standardized datasets specifically tailored for this application domain. The ExoNet dataset was built to mitigate the lack of a dataset for lower-limb prosthetic applications, and it was used to compare the performance of several common CNNs (Laschowski et al., 2022). On the other hand, the Megane Pro dataset could be a starting point for the training in the upper-limb prosthetic control field, as it collects at the same time sEMG, visual, and gaze tracking data from 15 amputees and 30 able-bodied subjects (Cognolato et al., 2020). Despite the emergence of these initial datasets, it is evident from Tables 1, 2, and 5 that only 12 out of 50 studies actually make their datasets publicly available, often omitting even basic details such as the size of the datasets used.

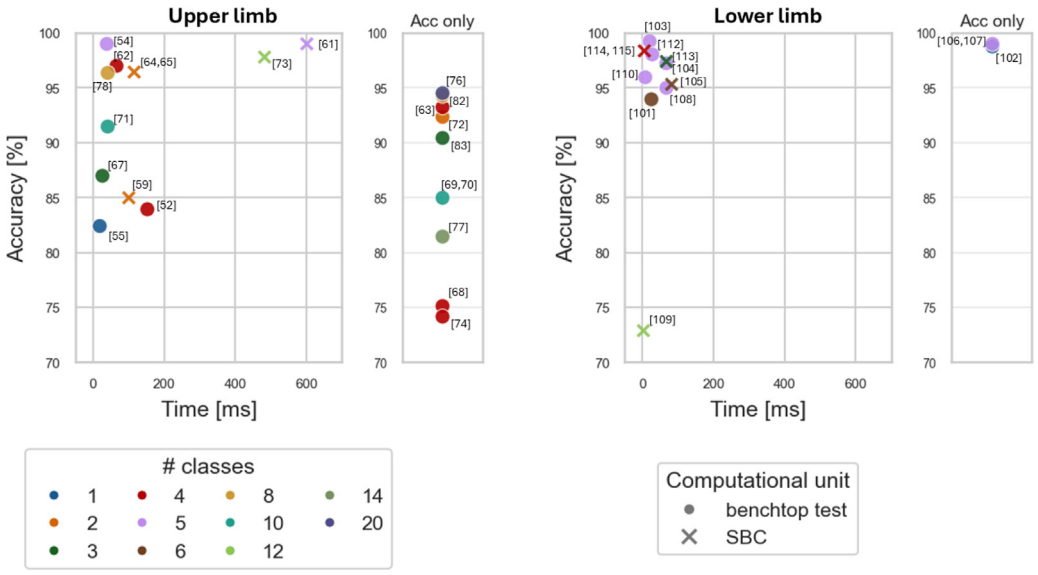

Fig. 6. Scatter plots of classification accuracy versus execution time for upper limb (left) and lower limb (right) applications. Studies reporting both accuracy and inference time metrics are shown in the principal panel (12 for the upper limb and 11 for the lower limb). Works that did not report execution time are represented in the "Acc only" panels, where only classification accuracy is indicated (10 for upper limb and 3 for lower limb). Studies that did not provide at least accuracy values were excluded from this plot (11 for the upper limb and 3 for the lower limb). Colors represent the number of classes considered, while marker type distinguishes between benchtop evaluations (circles) and implementations on embedded systems, i.e. on a Single-Board Computer (SBC, crosses). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Moreover, it is important to notice that the most representative classes considered in the terrain classification task for lower limb prosthetic application are level ground, upstairs, downstairs, up ramp, and down ramp (Zhang et al., 2019a,b; Al-Dabbagh and Ronsse, 2022; Li et al., 2022; Chen et al., 2022; Zhang et al., 2020). A subset of three classes is considered in Li et al. (2023), i.e. level ground, upstairs, and downstairs, while in Krausz and Hargrove (2021), Krausz et al. (2019), the sit-to-stand class is added to the 5 previously indicated. Instead in Zhong et al. (2020a), 6 classes were selected, considering also different types of terrain: up stairs, down stairs, brick, tile, grass, cement. Other exceptions are (Krausz et al., 2015) and Kurbis et al. (2024b,a), in which the objective of the work is the recognition of stairs and the extraction of their principal characteristics (Kurbis et al., 2024a). Lastly, in Laschowski et al. (2022) 12 classes were taken into account, considering a combination of three information: the different type of environment, i.e. upstairs, downstairs, level ground, door-wall; the state of locomotion (transition or steady); the condition after the transition state (level ground, door-wall, upstairs, downstairs, seat, other).

Although the first datasets usable in these contexts are starting to emerge, the significant variability in model input sizes (e.g., image resolution) still hinders direct comparisons across different approaches. Fig. 6 summarizes the performance of the selected works in terms of classification accuracy and inference time for upper-limb (left) and lower-limb (right) applications. The main panels show studies reporting both metrics (12 for upper limb, 11 for lower limb), while the 'Acc only' panels include works providing only the information about the classification accuracy (10 and 3, respectively). Colors are used to denote the number of classes, and marker styles are employed to distinguish between benchtop evaluations and implementations on SBC. For the upper limb, a large variability can be observed in the number of grasp classes considered, which makes direct comparison between works challenging. Nevertheless, some studies have demonstrated promising results, such as Cirelli et al. (2023), Deshmukh et al. (2023), achieving accuracies above 94%, even with a high number of hand gestures (>11) and objects (>16). In both cases, object recognition is performed using the YOLO (Redmon et al., 2016) object detector, trained either on the COCO dataset (Cirelli et al., 2023) or on a custom dataset (Deshmukh et al., 2023), and each recognized object is then associated with a predefined grasp type, enabling an effective grasp selection strategy. However, only in Cirelli et al. (2023) time inference is reported, and the solution was tested on a SBC, thus providing a more realistic indication of its feasibility for real-time deployment.

For the lower limb, the various reviewed studies exhibit comparable performance in terms of accuracy and processing time. As previously noted, the number of classes is usually fixed at five, which further increases the comparability of the reported results. Nevertheless, only a very limited number of works evaluate their algorithms under reduced computational power conditions (e.g., on SBC), as emerged from Fig. 6, leaving open questions on their actual suitability for embedded, real-time applications. However, while consistently maintaining high performance, the approach proposed in Chen et al. (2022), which incorporates an unsupervised domain adaptation with sim-to-real transfer, allows for better generalization of the trained models. This strategy achieves 98.06% ± 0.71% and 95.91% ± 1.09% in indoor and outdoor classification, respectively.

4.3. Study populations

A notable limitation in the current literature, which calls for more in-depth investigation in future studies, is the lack of evaluation of these approaches outside laboratory settings and on actual end users. Indeed, among the selected scientific papers, 43 involved able-bodied (AB) subjects, 28 in upper limb applications and 15 in lower limb ones, while only 20 studies enrolled Amputees (44.2% of yes in the risk of bias assessment AXIS tool). Specifically, 9 studies involved trans-radial amputees (TRA), and 1 involved a trans-humeral amputee (THA); for lower-limb applications, the amputation level was often unspecified. As mentioned in Section 3.1.3, only two studies deviated from this trend by testing their systems on epileptic (EP) (McMullen et al., 2013) and non-sighted (NS) individuals (Peng et al., 2024). It is therefore evident that there is a paucity of research examining the impact of vision-based prosthetic control on actual prosthesis users, with the majority of studies conducted in controlled laboratory settings. Moreover, as highlighted by the AXIS tool, none of the reviewed works justify the sample size (i.e. the number of subjects enrolled in the experiment)

from a statistical standpoint, and only the 73.2% of the studies involving human subjects have received approval from an ethical committee. Testing these systems in real-world environments could provide essential insights into usability, reliability, and user adaptation in dynamic and unpredictable contexts. Such evidence would be crucial to guide design improvements and ensure the practical applicability of these technologies. Moving forward, research in this field must prioritize enrolling end users, such as amputees, and statistically planning the expected outcomes, in order to enhance both the robustness and the clinical relevance of the findings.

4.4. Tasks, metrics and cognitive load reduction in upper limb prostheses

As Fig. 7 illustrates, while there is considerable variability in the choice of tasks and metrics across studies highlighting the current lack of standardization in the field, the most frequently adopted tasks are *Pick & Place* (33.3%) and its simplified variant *Reach & Grasp* (18.1%), and the most commonly reported performance metrics are the Success Rate (SR) and the Task Accomplishment Time (TAT). These tasks and metrics are widely employed not only due to their prevalence in prior studies, but also because they provide a structured and quantifiable means of evaluating key functional components of prosthetic use, such as movement accuracy, grasp stability, and execution time. Their standardized nature facilitates reproducibility and enables consistent benchmarking across studies. For these reasons, future research should continue to include these tasks and metrics to ensure comparability with existing literature and contribute to the establishment of consistent evaluation protocols. At the same time, extending the assessment framework to include more complex tasks, such as Activities of Daily Living (ADLs), would allow for a more comprehensive understanding of system performance in real-world scenarios (Cui et al., 2022; Zhang et al., 2024), including factors like user adaptability, cognitive load, and functional utility.

Concerning the control strategies, several approaches proposed a semiautonomous control strategy for upper-limb prostheses in which EMG signals are only used to trigger the control pipeline (Došen et al., 2010; Castro et al., 2022), while only recently the possibility of integrating information coming from both EMG and visual sensors was explored for improving prosthetic control (Cirelli et al., 2023; Deshmukh et al., 2023; Zandigohar et al., 2024). Moreover, although the literature demonstrated that multimodal control yields better results with respect to a pure EMG-based control (Mouchoux et al., 2021; Huang et al., 2022; Cognolato et al., 2022; Wang et al., 2022; Deshmukh et al., 2023; Gigli et al., 2018), there is a lack of an accurate and thorough study on the usability of the system by the user, which robustly quantifies the cognitive load required to the user (He et al., 2019; Huang et al., 2022; Cirelli et al., 2023). Given the promising performance gains demonstrated by vision-based multimodal strategies, further research in this area is well justified. To ensure their viability in real-world applications, it is crucial to move beyond purely technical validation and systematically evaluate their impact on user experience, including cognitive workload and long-term usability quantification.

To address this gap, future studies should include direct comparisons between traditional strategies and multimodal approaches integrating CV to reveal potential differences in user cognitive load. Assessments should combine subjective measures, such as the NASA-TLX or Borg Scale, with objective physiological measures, including relevant parameters that exhibit significant responses due to increased workload. As an example, heart rate variability, electroencephalographic activity, and galvanic skin response are well known to correlate with sustained attention and stress (Tamantini et al., 2025). Adopting these metrics would provide a more comprehensive evaluation of multimodal strategies, encompassing not only technical performance but also the overall user perception and required cognitive load.

4.5. Moving beyond feasibility studies in lower limb prostheses

Although all the works reviewed on lower-limb prosthesis control emphasize the importance of visual information for correctly classifying terrains and identifying the appropriate locomotion parameters, there ares currently only two studies in the literature that integrate CVS into an active lower-limb prosthesis (Hong et al., 2023; Zhang et al., 2020). Moreover, a multi-modal control strategy that exploits features extracted from different types of sensors, i.e. EMG, IMU, and vision, seems to be the best solution for improving the control of prosthetic limbs. For instance, it was demonstrated how a combination of vision features with EMG, IMU, and Goniometer feature sets is capable of minimizing inter- and intra-subject variability in terms of separability, repeatability, clustering, and overall performance, thus enabling better generalization, in a lower limb approach (Krausz and Hargrove, 2021; Krausz et al., 2019), achieving the best performance.

Based on these findings, future work on lower-limb prostheses should prioritize integrating vision with kinematic information, as this combination balances the richness of contextual data with real-time feasibility. In particular, integrating lightweight, low-power sensors, such as RGB cameras and IMUs, could offer a practical trade-off between environmental awareness and computational efficiency. In this context, simplifying the sensor setup by reducing the number of EMG electrodes, potentially limiting their use to detecting movement intention through a simple threshold-based trigger, may be advantageous. This is particularly relevant given the susceptibility of EMG signals to degradation caused by factors such as sweat, electrode displacement, and pressure variations within the socket, as well as safety concerns linked to decoding errors, which could compromise user stability and increase fall risk (Fleming et al., 2021).

To minimize system complexity, algorithms should focus on extracting high-level features (e.g., terrain class, slope direction, or obstacle presence) rather than processing raw video streams, thus reducing both data bandwidth and processing demands. Ultimately, selecting the right set of fused signals should aim to optimize task-relevant information while preserving wearability, power efficiency, and real-time responsiveness.

5. Conclusions and future perspectives

The integration of computer vision into the control pipeline of powered prosthetic limbs is receiving increasing attention due to its potential to enhance functionality and address issues like prosthetic abandonment, often attributed to the unreliability and non-intuitiveness of myoelectric control. This literature review was conducted using a search query in well-known electronic databases (Scopus, PubMed, and IEEE Xplore), focusing on scientific papers published in international journals up to May 2025. The search yielded 50 journal papers, with 33 and 17 dedicated to computer vision applications in upper- and lower-limb prosthetics, respectively, which are discussed in detail.

This review highlighted the main characteristics of computer vision systems and algorithms designed to support limb prosthesis control, outlining their potential and current limitations. Based on the literature analysis, a key unresolved issue remains the optimal placement of vision sensors. While both head-mounted and limb-mounted configurations have been proposed, comparative studies evaluating their impact on occlusions, field of view, and user comfort are lacking. Future work should focus on systematic benchmarking of sensor placements to identify configurations that best balance perception quality and wearability.

Another critical gap concerns real-world deployment. Although many studies demonstrate promising results in controlled settings, few leverage embedded computing platforms suitable for real-time execution in daily life scenarios. Future research should prioritize testing on resource-constrained hardware to assess the true feasibility and performance of proposed solutions in operational environments.

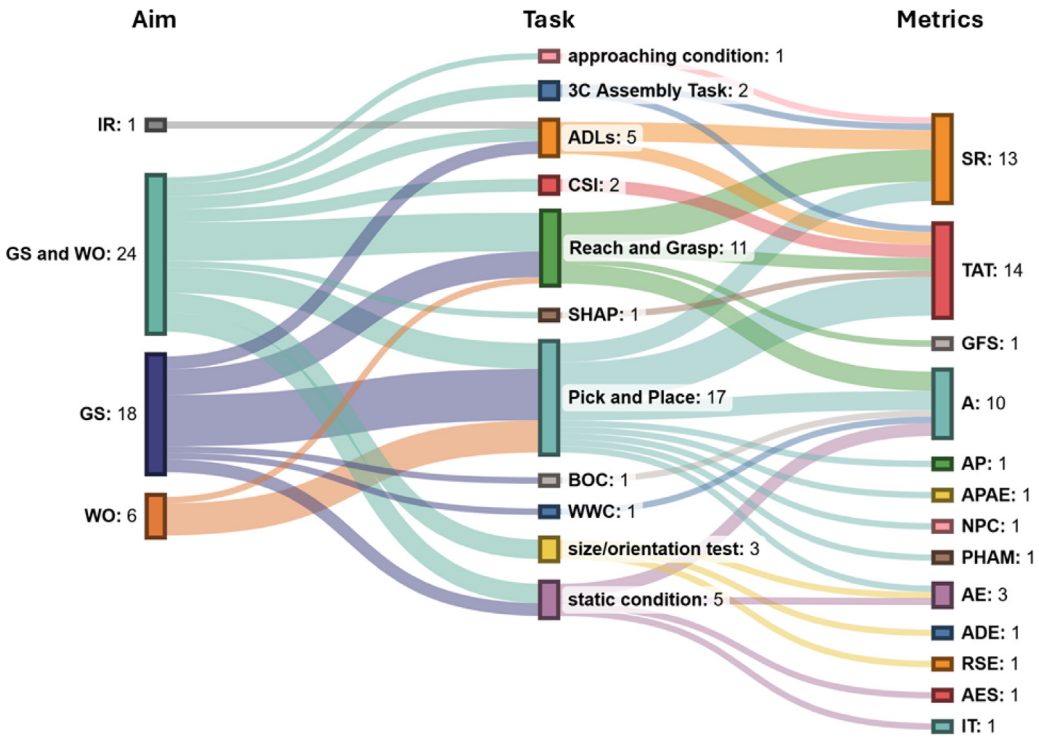

Fig. 7. Alluvial diagram illustrating the distribution of tasks and metrics across different study aims. The left column represents the different aim of studies (e.g. GS, WO), which are linked to the central column showing specific task types (e.g., ADLs, Reach and Grasp, Pick and Place). The right column depicts outcome measures (e.g., SR, TAT) and the definitions of these metrics are reported in Table 3. The width of each flow is proportional to the number of studies connecting the corresponding categories, highlighting the relationships between study design, task selection, and evaluation metrics. IR: Intention Recognition, GS: Grasp Selection, WO: Wrist Orientation, CSI: Cluttered Scene Interaction, BOC: Test with samples never seen, WWC: test with different angles; ADL: Activity of Daily Living; RSE: Relative Size Error; ADE: Average Distance Error; AP: Accuracy in Pre-shaping; SR: Success Rate; TAT: Time to Accomplish Task; APAE: Average Prosthesis Aperture Error; NPC: Number of Prediction Changes; A: Accuracy; IT: Inference Time; GFS: Grasping Functionality Score; AE: Angular Error; AES: Angular Estimation Stability; PHAM: Prosthetic Hand Assessment Measure.

The review also revealed a lack of publicly available, standardized datasets for key tasks such as object grasping and terrain classification. This hinders cross-study comparisons and slows algorithmic progress. To address this, the community should promote the creation and open sharing of annotated datasets, along with standardizing data formats and image resolutions to enable more robust benchmarking.

User-centered evaluation is another area requiring deeper investigation. Although some studies report technical performance metrics, usability and acceptability from the perspective of end users remain unexplored. Future studies should incorporate structured protocols to assess interface design, ease of integration into daily routines, and perceived cognitive workload.

Lastly, with regard to lower-limb prostheses, most current works focus on visual terrain recognition as a stand-alone task, without integrating this information into the control loop. There is a clear need for research efforts that go beyond feasibility and explore how visual input can directly inform control strategies, especially in dynamic, unstructured environments.

In summary, advancing vision-based prosthetic control will require emphasis on real-world usability, comparative evaluations, standardized datasets, and integration into functional prosthetic systems beyond simulations or benchtop testing, particularly for lower-limb systems, which remain under-investigated. Furthermore, promising directions for future work include the adoption of advanced vision methods capable of explicitly modeling hierarchical and part–object relationships, such as those recently explored in part–object relational visual saliency (Liu et al., 2021) and capsule networks with residual pose routing (Liu et al., 2024). By capturing compositional structures in complex visual scenes, these approaches may inspire new solutions for integrating richer visual understanding into prosthetic control systems.

CRediT authorship contribution statement

Gianmarco Cirelli: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Christian Tamantini:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Loredana Zollo:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Francesca Cordella:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the European Union - Next Generation EU - NRRP M4.C2 - Investment 1.5 Establishing and strengthening of Innovation Ecosystems for sustainability (Project n. ECS000 00024 Rome Technopole), in part by INAIL prosthetic center with BioInterNect (CUP: E57G23000280005), and in part by the Italian Ministry of Research, under the complementary actions to the NRRP “Fit4MedRob - Fit for Medical Robotics” Grant PNC0000007, (CUP: B53C22006990001).

Data availability

No data was used for the research described in the article.

References

A Powered Prosthetic Hand with Vision System, 2025. A powered prosthetic hand with vision system for enhancing the anthropopathic grasp. IEEE Trans. Neural Syst. Rehabil. Eng. : A Publ. IEEE Eng. Med. Biology Soc. PP, http://dx.doi.org/10.1109/TNSRE.2025.3567392.

Al-Dabbagh, A.H., Ronsse, R., 2022. Depth vision-based terrain detection algorithm during human locomotion. IEEE Trans. Med. Robot. Bionics 4 (4), 1010–1021.

Amtmann, D., Morgan, S.J., Kim, J., Hafner, B.J., 2015. Health-related profiles of people with lower limb loss. Arch. Phys. Med. Rehabil. 96 (8), 1474–1483.

Armannsdottir, A., Tranberg, R., Halldorsdottir, G., Briem, K., 2018. Frontal plane pelvis and hip kinematics of transfemoral amputee gait. effect of a prosthetic foot with active ankle dorsiflexion and individualized training–a case study. Disabil. Rehabil.: Assist. Technol. 13 (4), 388–393.

Arruda, H., Silva, E.R., Lessa, M., Proença Jr., D., Bartholo, R., 2022. Vosviewer and bibliometrix. J. Med. Libr. Assoc.: JMLA 110 (3), 392.

Asif, M., Tiwana, M.I., Khan, U.S., Qureshi, W.S., Iqbal, J., Rashid, N., Naseer, N., 2021. Advancements, trends and future prospects of lower limb prosthesis. IEEE Access 9, 85956–85977.

Calli, B., Walsman, A., Singh, A., Srinivasa, S., Abbeel, P., Dollar, A.M., 2015. Benchmarking in manipulation research: The ycb object and model set and benchmarking protocols. arXiv preprint arXiv:1502.03143.

Castro, M.N., Dosen, S., 2022. Continuous semi-autonomous prosthesis control using a depth sensor on the hand. Front. Neurorobotics 16, 814973.

Castro, M.C.F., Pinheiro, W.C., Rigolin, G., 2022. A hybrid 3D printed hand prosthesis prototype based on sEMG and a fully embedded computer vision system. Front. Neurorobotics 15, 751282.

Chen, C., Zhang, K., Leng, Y., Chen, X., Fu, C., 2022. Unsupervised sim-to-real adaptation for environmental recognition in assistive walking. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 1350–1360.

Cirelli, G., Tamantini, C., Cordella, L.P., Cordella, F., 2023. A semiautonomous control strategy based on computer vision for a hand–wrist prosthesis. Robotics 12 (6), 152.

Cognolato, M., Atzori, M., Gassert, R., Müller, H., 2022. Improving robotic hand prosthesis control with eye tracking and computer vision: A multimodal approach based on the visuomotor behavior of grasping. Front. Artif. Intell. 4, 744476.

Cognolato, M., Gijsberts, A., Gregori, V., Saetta, G., Giacomino, K., Hager, A.-G.M., Gigli, A., Faccio, D., Tiengo, C., Bassetto, F., et al., 2020. Gaze, visual, myoelectric, and inertial data of grasps for intelligent prosthetics. Sci. Data 7 (1), 43.

Cordella, F., Ciancio, A.L., Sacchetti, R., Davalli, A., Cutti, A.G., Guglielmelli, E., Zollo, L., 2016. Literature review on needs of upper limb prosthesis users. Front. Neurosci. 10, 209.

Cui, J.-W., Du, H., Yan, B.-Y., Wang, X.-J., 2022. Research on upper limb action intention recognition method based on fusion of posture information and visual information. Electronics 11 (19), 3078.

Dargan, S., Bansal, S., Kumar, M., Mittal, A., Kumar, K., 2023. Augmented reality: A comprehensive review. Arch. Comput. Methods Eng. 30 (2), 1057–1080.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.

Deshmukh, S., Khatik, V., Saxena, A., 2023. Robust fusion model for handling EMG and computer vision data in prosthetic hand control. IEEE Sensors Lett..

Došen, S., Cipriani, C., Kostić, M., Controzzi, M., Carrozza, M.C., Popović, D.B., 2010. Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation. J. Neuroeng. Rehabil. 7 (1), 1–14.

Došen, S., Popović, D.B., 2011. Transradial prosthesis: artificial vision for control of prehension. Artif. Organs. 35 (1), 37–48.

Downes, M.J., Brennan, M.L., Williams, H.C., Dean, R.S., 2016. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 6 (12), e011458.

Fan, S., Dai, J., Zhang, N., Zhang, T., Cheng, M., Liu, B., Jiang, L., 2025. Design, analysis, and experiment of a coupled-adaptive underactuated prosthetic handbased on linkage mechanisms. Anal. Exp. A Coupled-Adaptive Underactuated Prosthet. Handbased Link. Mech. (Accessed 16 June 2025).

Farina, D., Vujaklija, I., Brånemark, R., Bull, A.M., Dietl, H., Graimann, B., Hargrove, L.J., Hoffmann, K.-P., Huang, H., Ingvarsson, T., et al., 2023. Toward higher-performance bionic limbs for wider clinical use. Nat. Biomed. Eng. 7 (4), 473–485.

Fejér, A., Nagy, Z., Benois-Pineau, J., Szolgay, P., de Rugy, A., Domenger, J.-P., 2021. Implementation of scale invariant feature transform detector on FPGA for low-power wearable devices for prostheses control. Int. J. Circuit Theory Appl. 49 (7), 2255–2273.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Fleming, A., Stafford, N., Huang, S., Hu, X., Ferris, D.P., Huang, H.H., 2021. Myoelectric control of robotic lower limb prostheses: a review of electromyography interfaces, control paradigms, challenges and future directions. J. Neural Eng. 18 (4), 041004.

Gardner, M., Mancero Castillo, C.S., Wilson, S., Farina, D., Burdet, E., Khoo, B.C., Atashzar, S.F., Vaidyanathan, R., 2020. A multimodal intention detection sensor suite for shared autonomy of upper-limb robotic prostheses. Sensors 20 (21), 6097.

Gehlhar, R., Tucker, M., Young, A.J., Ames, A.D., 2023. A review of current state-of-the-art control methods for lower-limb powered prostheses. Annu. Rev. Control..

Gentile, C., Cordella, F., Zollo, L., 2022. Hierarchical human-inspired control strategies for prosthetic hands. Sensors 22 (7), 2521.

Geusebroek, J.-M., Burghouts, G.J., Smeulders, A.W., 2005. The amsterdam library of object images. Int. J. Comput. Vis. 61, 103–112.

Ghazaei, G., Alameer, A., Degenaar, P., Morgan, G., Nazarpour, K., 2017. Deep learning-based artificial vision for grasp classification in myoelectric hands. J. Neural Eng. 14 (3), 036025.

Gigli, A., Gregori, V., Cognolato, M., Atzori, M., Gijsberts, A., 2018. Visual cues to improve myoelectric control of upper limb prostheses. In: 2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob). IEEE, pp. 783–788.

Gionfrida, L., Kim, D., Scaramuzza, D., Farina, D., Howe, R.D., 2024. Wearable robots for the real world need vision. Sci. Robot. 9 (90), eadj8812.

Gulati, P., Hu, Q., Atashzar, S.F., 2021. Toward deep generalization of peripheral emg-based human-robot interfacing: A hybrid explainable solution for neurorobotic systems. IEEE Robot. Autom. Lett. 6 (2), 2650–2657.

He, Y., Kubozono, R., Fukuda, O., Yamaguchi, N., Okumura, H., 2020. Vision-based assistance for myoelectric hand control. IEEE Access 8, 201956–201965.

He, Y., Shima, R., Fukuda, O., Bu, N., Yamaguchi, N., Okumura, H., 2019. Development of distributed control system for vision-based myoelectric prosthetic hand. IEEE Access 7, 54542–54549.

Hong, Z., Bian, S., Xiong, P., Li, Z., 2023. Vision-locomotion coordination control for a powered lower-limb prosthesis using fuzzy-based dynamic movement primitives. IEEE Trans. Autom. Sci. Eng..

Huang, J., Li, Z., Xia, H., Chen, G., Meng, Q., 2022. Cross-modal integration and transfer learning using fuzzy logic techniques for intelligent upper limb prosthesis. IEEE Trans. Fuzzy Syst. 31 (4), 1267–1280.

Huang, H.H., Si, J., Brandt, A., Li, M., 2021. Taking both sides: seeking symbiosis between intelligent prostheses and human motor control during locomotion. Curr. Opin. Biomed. Eng. 20, 100314.

Huang, Z., Zheng, J., Zhao, L., Chen, H., Jiang, X., Zhang, X., 2023. DL-Net: Sparsity prior learning for grasp pattern recognition. IEEE Access 11, 6444–6451.

Jang, C.H., Yang, H.S., Yang, H.E., Lee, S.Y., Kwon, J.W., Yun, B.D., Choi, J.Y., Kim, S.N., Jeong, H.W., 2011. A survey on activities of daily living and occupations of upper extremity amputees. Ann. Rehabil. Med. 35 (6), 907.

Karrenbach, M., Boe, D., Sie, A., Bennett, R., Rombokas, E., 2022. Improving automatic control of upper-limb prosthesis wrists using gaze-centered eye tracking and deep learning. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 340–349.

Krausz, N.E., Hargrove, L.J., 2021. Sensor fusion of vision, kinetics, and kinematics for forward prediction during walking with a transfemoral prosthesis. IEEE Trans. Med. Robot. Bionics 3 (3), 813–824.

Krausz, N.E., Hu, B.H., Hargrove, L.J., 2019. Subject-and environment-based sensor variability for wearable lower-limb assistive devices. Sensors 19 (22), 4887.

Krausz, N.E., Lenzi, T., Hargrove, L.J., 2015. Depth sensing for improved control of lower limb prostheses. IEEE Trans. Biomed. Eng. 62 (11), 2576–2587.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25.

Kurbis, A.G., Kuzmenko, D., Ivanyuk-Skulskiy, B., Mihailidis, A., Laschowski, B., 2024a. StairNet: visual recognition of stairs for human–robot locomotion. BioMedical Eng. OnLine 23 (1), 20.

Kurbis, A.G., Mihailidis, A., Laschowski, B., 2024b. Development and mobile deployment of a stair recognition system for human-robot locomotion. IEEE Trans. Med. Robot. Bionics.

Kyberd, P., Popa, A.F., Cojean, T., 2023. A tool to assist in the analysis of gaze patterns in upper limb prosthetic use. Prosthesis 5 (3), 898–915.

Lai, K., Bo, L., Ren, X., Fox, D., 2011. A large-scale hierarchical multi-view rgb-d object dataset. In: 2011 IEEE International Conference on Robotics and Automation. IEEE, pp. 1817–1824.

Laschowski, B., McNally, W., Wong, A., McPhee, J., 2022. Environment classification for robotic leg prostheses and exoskeletons using deep convolutional neural networks. Front. Neurorobotics 15, 730965.

Ledoux, E.D., Goldfarb, M., 2017. Control and evaluation of a powered transfemoral prosthesis for stair ascent. IEEE Trans. Neural Syst. Rehabil. Eng. 25 (7), 917–924.

Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J., 2021. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans. Neural Netw. Learn. Syst..

Li, L., Wang, X., Meng, Q., Yu, H., 2023. A wearable computer vision system with gimbal enables position-, speed-, and phase-independent terrain classification for lower limb prostheses. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 4539–4548.

Li, M., Zhong, B., Liu, Z., Lee, I.-C., Fylstra, B.L., Lobaton, E., Huang, H.H., 2019. Gaze fixation comparisons between amputees and able-bodied individuals in approaching stairs and level-ground transitions: a pilot study. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC, IEEE, pp. 3163–3166.

Li, M., Zhong, B., Lobaton, E., Huang, H., 2022. Fusion of human gaze and machine vision for predicting intended locomotion mode. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 1103–1112.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann. Intern. Med. 151 (4), W–65.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer, pp. 740–755.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: European Conference on Computer Vision. Springer, pp. 21–37.

Liu, Y., Cheng, D., Zhang, D., Xu, S., Han, J., 2024. Capsule networks with residual pose routing. IEEE Trans. Neural Netw. Learn. Syst..

Liu, M., Wang, D., Huang, H., 2015. Development of an environment-aware locomotion mode recognition system for powered lower limb prostheses. IEEE Trans. Neural Syst. Rehabil. Eng. 24 (4), 434–443.

Liu, Y., Zhang, D., Zhang, Q., Han, J., 2021. Part-object relational visual saliency. IEEE Trans. Pattern Anal. Mach. Intell. 44 (7), 3688–3704.

Losey, D.P., McDonald, C.G., Battaglia, E., O'Malley, M.K., 2018. A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. Appl. Mech. Rev. 70 (1), 010804.

Manz, S., Valette, R., Damonte, F., Avanci Gaudio, L., Gonzalez-Vargas, J., Sartori, M., Dosen, S., Rietman, J., 2022. A review of user needs to drive the development of lower limb prostheses. J. Neuroeng. Rehabil. 19 (1), 119.

Markovic, M., Dosen, S., Cipriani, C., Popovic, D., Farina, D., 2014. Stereovision and augmented reality for closed-loop control of grasping in hand prostheses. J. Neural Eng. 11 (4), 046001.

Markovic, M., Dosen, S., Popovic, D., Graimann, B., Farina, D., 2015. Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis. J. Neural Eng. 12 (6), 066022.

Massalin, Y., Abdrakhmanova, M., Varol, H.A., 2017. User-independent intent recognition for lower limb prostheses using depth sensing. IEEE Trans. Biomed. Eng. 65 (8), 1759–1770.

McDonald, C.L., Westcott-McCoy, S., Weaver, M.R., Haagsma, J., Kartin, D., 2021. Global prevalence of traumatic non-fatal limb amputation. Prosthet. Orthot. Int. 45 (2), 105–114.

McMullen, D.P., Hotson, G., Katyal, K.D., Wester, B.A., Fifer, M.S., McGee, T.G., Harris, A., Johannes, M.S., Vogelstein, R.J., Ravitz, A.D., et al., 2013. Demonstration of a semi-autonomous hybrid brain–machine interface using human intracranial EEG, eye tracking, and computer vision to control a robotic upper limb prosthetic. IEEE Trans. Neural Syst. Rehabil. Eng. 22 (4), 784–796.

Medioni, G., Trivedi, M., 2017. Computer vision for assistive technologies. Comput. Vis. Image Underst. 154, 1–15.

Mouchoux, J., Carisi, S., Dosen, S., Farina, D., Schilling, A.F., Markovic, M., 2021. Artificial perception and semiautonomous control in myoelectric hand prostheses increases performance and decreases effort. IEEE Trans. Robot. 37 (4), 1298–1312.

Nelson, M.E., MacIver, M.A., 2006. Sensory acquisition in active sensing systems. J. Comp. Physiol. A 192, 573–586.

Organization, W.H., et al., 2011. World Report on Disability 2011. World Health Organization.

Össur, 2025. i-limb Ultra. URL https://www.ossur.com/en-us/prosthetics/arms/i-limb-ultra. (Accessed 16 June 2025).

Ottobock, 2025a. Michelangelo. URL https://www.ottobock.com/en-us/product/8E500. (Accessed 16 June 2025).

Ottobock, 2025b. Bebionic. URL https://www.ottobock.com/it-it/product/8E7*. (Accessed 16 June 2025).

Park, H.-J., An, B.-H., Joo, S.-B., Kwon, O.-W., Kim, M.Y., Seo, J., 2022. Grasping time and pose selection for robotic prosthetic hand control using deep learning based object detection. Int. J. Control. Autom. Syst. 20 (10), 3410–3417.

Peng, C., Yang, D., Zhao, D., Cheng, M., Dai, J., Jiang, L., 2024. Viiat-hand: a reach-and-grasp restoration system integrating voice interaction, computer vision, auditory and tactile feedback for blind amputees. IEEE Robot. Autom. Lett..

Prensilia, 2025. IH2 Azzurra. URL https://www.prensilia.com/it/ih2-azzurra/. (Accessed 16 June 2025).

Ragusa, E., Gianoglio, C., Dosen, S., Gastaldo, P., 2021. Hardware-aware affordance detection for application in portable embedded systems. IEEE Access 9, 123178–123193.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 779–788.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149.

Roche, A.D., Rehbaum, H., Farina, D., Aszmann, O.C., 2014. Prosthetic myoelectric control strategies: a clinical perspective. Curr. Surg. Rep. 2, 1–11.

Roy, R., Mahadevappa, M., Nazarpour, K., 2021. An electro-oculogram based vision system for grasp assistive devices—A proof of concept study. Sensors 21 (13), 4515.

Salminger, S., Stino, H., Pichler, L.H., Gstoettner, C., Sturma, A., Mayer, J.A., Szivak, M., Aszmann, O.C., 2022. Current rates of prosthetic usage in upper-limb amputees–have innovations had an impact on device acceptance? Disabil. Rehabil. 44 (14), 3708–3713.

Schone, H.R., Udeozor, M., Moninghoff, M., Rispoli, B., Vandersea, J., Lock, B., Hargrove, L., Makin, T.R., Baker, C.I., 2024. Biomimetic versus arbitrary motor control strategies for bionic hand skill learning. Nat. Hum. Behav. 1–16.

Sharma, A., Rombokas, E., 2022. Improving imu-based prediction of lower limb kinematics in natural environments using egocentric optical flow. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 699–708.

Shi, X., Guo, W., Xu, W., Yang, Z., Sheng, X., 2024. Semi-autonomous grasping control of prosthetic hand and wrist based on motion prior field. IEEE Robot. Autom. Lett..

Shi, C., Yang, D., Zhao, J., Liu, H., 2020. Computer vision-based grasp pattern recognition with application to myoelectric control of dexterous hand prosthesis. IEEE Trans. Neural Syst. Rehabil. Eng. 28 (9), 2090–2099.

Shrestha, A., Mahmood, A., 2019. Review of deep learning algorithms and architectures. IEEE Access 7, 53040–53065.

Şimsek, N., Öztürk, G.K., Nahya, Z.N., 2020. The mental health of individuals with post-traumatic lower limb amputation: a qualitative study. J. Patient Exp. 7 (6), 1665–1670.

Sinha, D., El-Sharkawy, M., 2019. Thin mobilenet: An enhanced mobilenet architecture. In: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference. UEMCON, IEEE, pp. 0280–0285.

Sinha, R., van den Heuvel, W.J., Arokiasamy, P., 2011. Factors affecting quality of life in lower limb amputees. Prosthet. Orthot. Int. 35 (1), 90–96.

Smail, L.C., Neal, C., Wilkins, C., Packham, T.L., 2021. Comfort and function remain key factors in upper limb prosthetic abandonment: findings of a scoping review. Disabil. Rehabil.: Assist. Technol. 16 (8), 821–830.

Song, C., Chen, C., Lv, Y., Zhang, W., Zhang, X., Xu, J., 2025. Attention-inspired path prediction and adaptive obstacle perception architecture for powered lower-limb prostheses. IEEE Sensors J..

Starke, J., Weiner, P., Crell, M., Asfour, T., 2022. Semi-autonomous control of prosthetic hands based on multimodal sensing, human grasp demonstration and user intention. Robot. Auton. Syst. 154, 104123.

Stefanelli, E., Cordella, F., Gentile, C., Zollo, L., 2023. Hand prosthesis sensorimotor control inspired by the human somatosensory system. Robotics 12 (5), 136.

Sun, H., He, C., Vujaklija, I., 2023. Design trends in actuated lower-limb prosthetic systems: a narrative review. Expert. Rev. Med. Devices 20 (12), 1157–1172.

Tamantini, C., Cordella, F., Lauretti, C., Zollo, L., 2021. The WGD—A dataset of assembly line working gestures for ergonomic analysis and work-related injuries prevention. Sensors 21 (22), 7600.

Tamantini, C., Cristofanelli, M.L., Fracasso, F., Umbrico, A., Cortellessa, G., Orlandini, A., Cordella, F., 2025. Physiological sensor technologies in workload estimation: A review. IEEE Sensors J..

Tamantini, C., Lapresa, M., Cordella, F., Scotto di Luzio, F., Lauretti, C., Zollo, L., 2022. A robot-aided rehabilitation platform for occupational therapy with real objects. In: Converging Clinical and Engineering Research on Neurorehabilitation IV: Proceedings of the 5th International Conference on Neurorehabilitation (ICNR2020), October 13–16, 2020. Springer, pp. 851–855.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., et al., 2018. Deep learning for computer vision: A brief review. Comput. Intell. Neurosci. 2018.

Wang, S., Zheng, J., Huang, Z., Zhang, X., Prado da Fonseca, V., Zheng, B., Jiang, X., 2022. Integrating computer vision to prosthetic hand control with sEMG: Preliminary results in grasp classification. Front. Robot. AI 9, 948238.

Wang, X., Zhu, Z., 2023. Context understanding in computer vision: A survey. Comput. Vis. Image Underst. 229, 103646.

Weiner, P., Starke, J., Hundhausen, F., Beil, J., Asfour, T., 2018. The kit prosthetic hand: design and control. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 3328–3334.

Weiner, P., Starke, J., Rader, S., Hundhausen, F., Asfour, T., 2022. Designing prosthetic hands with embodied intelligence: the kit prosthetic hands. Front. Neurorobotics 16, 815716.

Xia, H., Zhang, Y., Rajabi, N., Taleb, F., Yang, Q., Kragic, D., Li, Z., 2024. Shaping high-performance wearable robots for human motor and sensory reconstruction and enhancement. Nat. Commun. 15 (1), 1760.

Xiloyannis, M., Alicea, R., Georgarakis, A.-M., Haufe, F.L., Wolf, P., Masia, L., Riener, R., 2021. Soft robotic suits: State of the art, core technologies, and open challenges. IEEE Trans. Robot. 38 (3), 1343–1362.

Yadav, D., Veer, K., 2023. Recent trends and challenges of surface electromyography in prosthetic applications. Biomed. Eng. Lett. 13 (3), 353–373.

Yamamoto, M., Chung, K.C., Sterbenz, J., Shauver, M.J., Tanaka, H., Nakamura, T., Oba, J., Chin, T., Hirata, H., 2019. Cross-sectional international multicenter study on quality of life and reasons for abandonment of upper limb prostheses. Plast. Reconstr. Surgery–Global Open 7 (5), e2205.

Yang, D., Liu, H., 2021. Human-machine shared control: New avenue to dexterous prosthetic hand manipulation. Sci. China Technol. Sci. 64 (4), 767–773.

Yeung, D., Guerra, I.M., Barner-Rasmussen, I., Siponen, E., Farina, D., Vujaklija, I., 2022. Co-adaptive control of bionic limbs via unsupervised adaptation of muscle synergies. IEEE Trans. Biomed. Eng. 69 (8), 2581–2592.

Yip, M., Salcudean, S., Goldberg, K., Althoefer, K., Menciassi, A., Opfermann, J.D., Krieger, A., Swaminathan, K., Walsh, C.J., Huang, H., et al., 2023. Artificial intelligence meets medical robotics. Science 381 (6654), 141–146.

Young, A.J., Hargrove, L.J., 2015. A classification method for user-independent intent recognition for transfemoral amputees using powered lower limb prostheses. IEEE Trans. Neural Syst. Rehabil. Eng. 24 (2), 217–225.

Zandigohar, M., Han, M., Sharif, M., Günay, S.Y., Furmanek, M.P., Yarossi, M., Bonato, P., Onal, C., Padır, T., Erdoğmuş, D., et al., 2024. Multimodal fusion of emg and vision for human grasp intent inference in prosthetic hand control. Front. Robot. AI 11, 1312554.

Zhan, T., Yin, K., Xiong, J., He, Z., Wu, S.-T., 2020. Augmented reality and virtual reality displays: perspectives and challenges. Iscience 23 (8).

Zhang, K., Luo, J., Xiao, W., Zhang, W., Liu, H., Zhu, J., Lu, Z., Rong, Y., de Silva, C.W., Fu, C., 2020. A subvision system for enhancing the environmental adaptability of the powered transfemoral prosthesis. IEEE Trans. Cybern. 51 (6), 3285–3297.

Zhang, K., Xiong, C., Zhang, W., Liu, H., Lai, D., Rong, Y., Fu, C., 2019a. Environmental features recognition for lower limb prostheses toward predictive walking. IEEE Trans. Neural Syst. Rehabil. Eng. 27 (3), 465–476.

Zhang, X., Zhang, T., Jiang, Y., Zhang, W., Lu, Z., Wang, Y., Tao, Q., 2024. A novel brain-controlled prosthetic hand method integrating AR-SSVEP augmentation, asynchronous control, and machine vision assistance. Heliyon 10 (5).

Zhang, K., Zhang, W., Xiao, W., Liu, H., De Silva, C.W., Fu, C., 2019b. Sequential decision fusion for environmental classification in assistive walking. IEEE Trans. Neural Syst. Rehabil. Eng. 27 (9), 1780–1790.

Zhong, B., Da Silva, R.L., Li, M., Huang, H., Lobaton, E., 2020a. Environmental context prediction for lower limb prostheses with uncertainty quantification. IEEE Trans. Autom. Sci. Eng. 18 (2), 458–470.

Zhong, B., Huang, H., Lobaton, E., 2020b. Reliable vision-based grasping target recognition for upper limb prostheses. IEEE Trans. Cybern. 52 (3), 1750–1762.