Computer Vision-Based Smart Artificial Hand for Upper Limb Amputee

Anish Sambhare, Ajay Lohar, Rushank Suryawanshi and Narendra Bhagat
Department of Electronics and Telecommunication Engineering,
Bharatiya Vidya Bhavan's Sardar Patel Institute of Technology, India

Keywords: Convolutional Nueral Network(CNN), Prosthetic Hand Control, Machine Learning Algorithms, Feature Extraction, Real Time Applications, Cost-Effective Prosthetics.

Abstract: This study introduces a novel approach for classifying grips in prosthetic hands to facilitate object manipulation using Convolutional Neural Networks (CNN). The experimental findings indicate that this method outperforms traditional learning models. Initially, three machine learning algorithms were assessed: Decision Tree, SVM, and Random Forest, to classify various objects such as bottles, cell phones, and cups. The classification accuracies achieved were 76%, 90%, and 95%, respectively. Traditional machine learning preprocessing techniques proved to be quite complex, making CNNs a more attractive option due to their ability to perform feature extraction and classification without the need for extensive preprocessing. The CNN developed in this study achieved a training accuracy of 99% and a testing accuracy of 97.5%, surpassing contemporary models like YOLO v3 and Faster R-CNN. The integration of data-augmented training and dropout regularization enhances the model's robustness and generalizability. This allows the prosthetic hand to achieve precise grip control in a cost-effective and sensor-free manner, making the system a dependable choice for real-time applications, thereby improving accessibility and functionality in prosthetic design.

1 INTRODUCTION

A prosthetic hand serves as an important device for the upper limb amputation as it allows a person to perform activities of daily living and regain some motor functions, such as the ability to grasp an object. Enhanced quality of life and independence, improved psychological well-being, promotion of self-confidence and social assimilation, and the ability to use and move in everyday settings are their benefits.

AI technology makes prosthetics more human-like through sensory feedback, better prediction, and intuitive motion, thus, amputees can do complicated tasks and their lives are improved (Chopra and Emran, 2024).Computer vision, one of the AI wings, is the key to making machines able to visually analyze the data thus, the robots can do functions such as image processing, object recognition, and tracking very quickly and precisely (Kutlugun and Eyüpoğlu, 2020).This paper deals with the development of CNN, architectures, applications in different dimensions, problems like generalization and security, as well as ways forward and hardware implementation (Ștefan–Adrian Ionescu and Poboroniuc, 2023).

This work presents "Action Image" for robotic grasping, accomplishing 84% best-case success in the real world through a CNN employing simulations with RGB, Depth, and RGB-D inputs (Somer M. Nacy and Baqer, 2017).The proposed study presents a two-stage CNN-based strategy for object detection and appearance synthesis, which in turn, enhances accuracy, explainability, and real-time capabilities of robotics in comparison to other techniques like PoseCNN or DOPE, such as pose estimation (Xiaotong Chen and Jenkins, 2019).

This project introduces an affordable 3D-printed upper limb prosthetic from recycled materials that is able to perform essential movements like gripping and rotating (Zewen Li and Zhou, 2021).Catalyzed by the discovery of the high expenditure required for myoelectric prosthetics, this lightweight, low-power artificial limb applies electromyography (EMG) signal processing and 3D printing. Meanwhile, continued development of the EMG scheme, biocompatible materials, compact motors, and stable power will bring it towards commercialization (Divya Pradip Roy and Hoque, 2021).

The main contribution of this research are as follows.

1. A cost-efficient computer vision-based artificial

Sambhare, A., Lohar, A., Suryawanshi, R. and Bhagat, N.
Computer Vision-Based Smart Artificial Hand for Upper Limb Amputee.
DOI: 10.5220/0013619200004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 413-419
ISBN: 978-989-758-763-4

hand, leveraging advancements in deep learning, embedded systems, and convolutional neural networks (CNNs).

2. The proposed system integrates a camera module for real-time object detection, classification, and grip formation, ensuring affordability and reduced maintenance.
3. By enabling precise and responsive control through a computer vision-based approach, this solution enhances user interaction, functionality, and mobility, addressing key limitations of traditional prosthetics.

2 RELATED WORK

Many Studies have been carried out to investigate the development of prosthetic hands with artificial intelligence, using different technologies and approaches to extend their functionality. These research works explore various ways to merge the newest AI methods to make the hands more controllable, adaptable, and overall functional, contributing significantly to the industry's progress.

In the research by (Ujjwal Sharma and Singh, 2023), object detection developments are explored, with YOLO being a main focus. Faster R-CNN ResNet was able to reach 77.4% mAP at 6 FPS, while YOLO v2 paired with a (416x416) resolution had an equal value of accuracy (77.2%) and speed (68 FPS). Later, YOLO v2 was upgraded to 78.4% mAP at 60 FPS (480x480). YOLO v3, which also used DarkNet-53 like ResNet-50, showed equivalent accuracy but was faster. Dataset specifics are not provided.

(Xia Zhao and Parmar, 2024) discussed the enormous contributions of Convolutional Neural Networks (CNNs) in improving computer vision tasks such as image classification, object detection, and video prediction. CNNs surpass traditional methods by delivering accurate results. The challenges in this field involve training with large datasets, model complexity, and high computational cost. Future research will focus on optimizing architectures and reducing the dependency on labeled data to improve performance.

(Ross Girshick and Malik, 2014) introduced a straightforward and reliable object detection method designed to work with CNNs, evaluated on the PASCAL VOC 2011 and 2012 datasets. The model improved mAP by more than 30% and achieved 47.9% accuracy in the segmentation task. This approach, integrating CNNs and region proposals, offers a fast alternative to more complicated ensemble systems.

(Chunyuan Shi and Liu, 2020) explored CNNs for recognizing grasp patterns in prosthetic hands, reporting mono-modal accuracies of 80% for RGB, 85.4% for grayscale, and 89.8% for depth images. The fusion of grayscale and depth data increased the recognition rate to 94.6%. Additionally, Vision-EMG achieved a 50% reduction in grasp-and-pick-up time compared to Coded-EMG, highlighting superior performance.

(Meena Laad and Saiyed, 2024) compared two object detection CNNs: SSD with MobileNetV1 and Faster-RCNN with InceptionV2, using a custom dataset of 444 images (355 for training, 89 for testing). While SSD was faster, it showed inferior performance compared to Faster-RCNN, which, though slower, was more accurate.

(Shripad Bhatlawande and Gadgil, 2023) conducted research into robotic grasping using RGB-D data. The study utilized the Cornell Grasp Dataset and applied graph segmentation and morphological image processing (MIP) with a Random Forest (RF) classifier. The method achieved 94.26% accuracy in grasping detection, outperforming other algorithms in both speed and accuracy.

(Douglas Morrison and Leitner, 2020) introduce the EGAD dataset, a more diverse tool for assessing robot arm interactions with objects, particularly for grasp-centric tasks. The GG-CNN algorithm associated with EGAD succeeds 58% of the time, indicating the challenges of natural grasp depth and orientation. More complex datasets like EGAD show greater limitations, offering opportunities for algorithm improvement.

(Cloutier and Yang, 2013) revisit various prosthetic hand control techniques, focusing on anticipatory pattern recognition, fuzzy clustering, neural networks, and ENG control. The study uses EMG signals for motion classification, achieving accuracy rates between 86% and 98%. ENG interfaces provide a more natural control method through the Peripheral Nervous System.

3 METHODOLOGY

3.1 Data Collection

The step of collecting data is video shooting of objects in the real world which will be interpreted by a computer vision model as grip recognition through, they are further processed into 300 frames of 15-second length, to make sure that the data is of good quality. These frames are classified among power grip, precision, grip or pinch grip.
Class A: Power Grip – Utilized mostly by objects

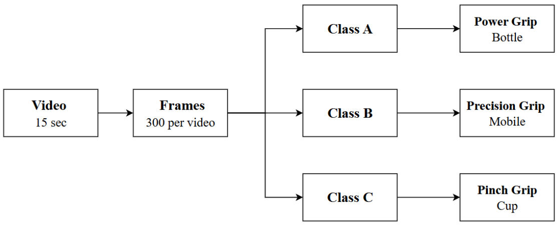

Figure 1: Frame Extraction and Classification Flowchart.

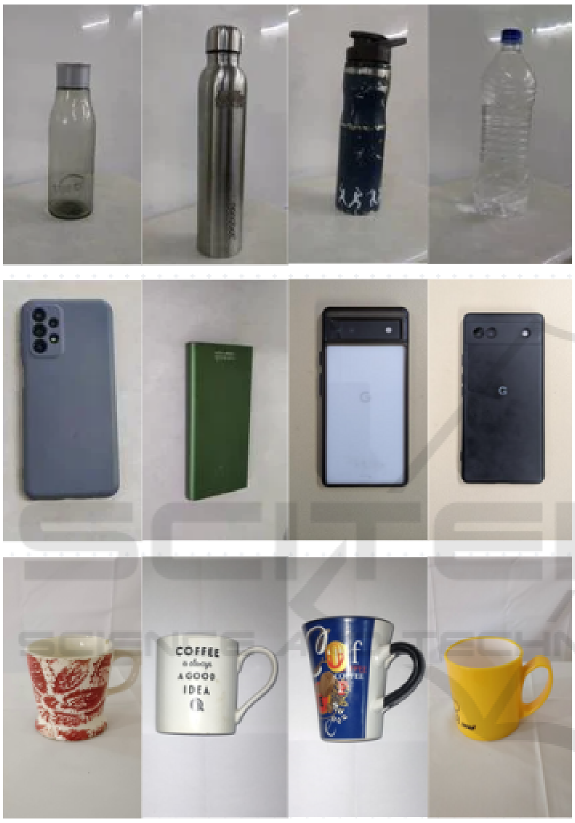

Figure 2: Dataset

such as bottles, which require a strong, whole-hand grasp to be effectively handled.
Class B: Precision Grip – The hugs used for phones which require a delicate firm grip by means of only the thumb and fingers.
Class C: Pinch Grip – Small objects, such as cups, have to follow a more gentle fingertip grip.
Fig. 1 shows the frame extraction and classification flowchart.

The categorization data is absolutely necessary for teaching a robot vision model to 1) detect objects, 2) pick out suitable gripping methods, and 3) alter hand configurations so that physical interaction with the real world is smooth.
Hence Fig. 2 shows the dataset collected.

3.2 Model Training

1. Data Processing : There are three groups of images of bottles, cups, and mobiles that are both used during training (70%), validation (15%), and testing (15%) to ensure the model is trained, validated, and performs well while avoiding overfitting. Data augmentation is used, such as rotations, movements, magnifications, and flips, and suddenly the model learns how to be able to deal with different data, and normalization is also applied during preprocessing. Pixel intensity values initially populated the range of 0 to 255 are brought to the interval 0–1 using Equation:

$$I_{\text{norm}} = \frac{I}{255} \tag{1}$$

where (I = pixel intesity) Normalization guarantees that the input scaling is consistent and in this way, the speed of convergence is increased and the numerical instability is reduced during training.The model receives the pre-processed data in mini-batches, where the images are resized to 224x224 pixels for memory efficiency and consistency.

In CNN, Convolutional layers are responsible for feature extraction from input images. The convolution operation is mathematically represented as:

$$Y[i,j] = \sum_{m}\sum_{n} X[i+m, j+n] \cdot K[m,n] \tag{2}$$

where X is the input, K is the kernel, and Y is the output.Deeper CNN layers are capable of learning high-level features, such as object parts, through complex convolutions for recognition.

Max pooling is a layer of deep learning networks that cut down the image dimensions while yet being able to maintain dimensionality. This can be seen as:

$$Y[i,j] = \max(X[m,n]) \tag{3}$$

Max pooling brings about the reduction of the computational cost, the prevention of overfitting, and the provision of spatial invariance, reference is made to categorical cross-entropy which is the loss function used during training is given by:

$$L = -\sum_{c=1}^{C} y_c \cdot \log(p_c) \tag{4}$$

where C is the number of classes, y_c is the actual label, and p_c is the probability of the predicted class being correct. This particular loss is very well-suited for multi-class classification, as it punishes the incorrect predictions more hard, thus leading to a better accuracy rate.The last layer in the CNN has the softmax activation function applied and subsequently, it converts the logits into probabilities of respective classes:

$$p_c = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}} \quad (5)$$

So, z_c stands for the class score of the observation, c. Softmax is probabilistic and hence, fits the process of multi-class classification.

A specific CNN pipeline guarantees the model is trained on both seen and unseen data, thus it leads to accurate and flexible grasp detection.

‘

2. Training Algorithm : The algorithm is trained through supervised learning for image classification and it consists of three phases: data preparation, model training, and evaluation with CNN that is used for feature extraction and spatial hierarchy capture. The testing of the model and its reliability for object detection is done after the training. Thus, the following algorithm ensure that the model also performs well on unseen data.

Algorithm 1 CNN Framework

```
Input : Dataset containing images of bottles, cups, and mobiles
Output : GripType → A (Power), B (Precision), C (Pinch)
Initialize IMAGE_SIZE = (224, 224), BATCH_SIZE = 32
Split DT into $DT_{train}$, $DT_{valid}$, and $DT_{test}$ with
    Train:Val:Test ratio = 35:35:30
for each image in DT do
    $DT_{preprocessed}$ = ImagePreprocessing(DT)
        Rescale pixel values to [0,1]
        Apply data augmentation (rotation, shift, shear, zoom, flip) for training set
        Resize images to IMAGE_SIZE
end for
Initialize CNN model with Conv2D layers (32, 64, 128 filters), MaxPooling2D layers, Flatten layer, Dense layers (128 neurons, 3 output classes), and Dropout (0.5)
Train $CNN_{model}$ using $DT_{train}$
    Optimize using Adam optimizer
    Use categorical_crossentropy loss
    Train for specified epochs
while True do
    PredictedGripType = $CNN_{model}(DT_{test})$
    if prediction confidence > threshold then
        Return grip type (power/precision/pinch)
    else
        Return error
    end if
end while
Function: ImagePreprocessing(DT)
for each image in DT do
    Normalize pixel values
    Resize to IMAGE_SIZE
    if training set then
        Apply augmentation
    end if
    Return preprocessed image
end for
```

3.3 Hardware Integration

The hardware integration for the artificial hand system facilitated smooth interaction between the control system and the servo motors that manage hand movements. Initially, specific controllers were selected for both internal and external operations, but issues with memory limitations and library compatibility affected performance. These obstacles led the design team to explore alternative solutions, which improved the system's hardware and software capabilities.

The answer was a new driver-based approach that allowed seamless communication between the control system and the servo motors. This advancement enabled precise and autonomous movements of the fingers and thumb. The driver served as a calibrated control solution, addressing the shortcomings of the original controller and greatly enhancing reliability and efficiency. This transformation, depicted in Fig. 3, illustrates the overall project concept and emphasizes the driver's essential role in achieving high-level performance.

The Python program analyzed classification outputs to manage hand movements, allowing for various grips such as the power grip, precision grip, and pinch grip. The system dynamically modified grip shapes and sizes based on the objects detected, ensuring natural and accurate manipulation. This feature enabled the artificial hand to replicate human hand motions, improving usability for upper-limb amputees.

The system setup, detailed in Fig. 4, consisted of a desktop computer running the software, testing hardware, a camera module, and a power supply unit. The prosthetic hand included servo motors and internal mechanisms, showcasing the collaboration between mechanics and electronics.

This innovative integration resulted in a responsive and adaptable artificial hand, capable of executing a range of tasks with precision. It enhanced the user experience by supporting natural hand movements, making it highly functional for amputees.

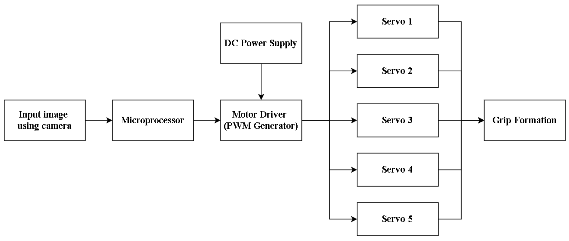

Figure 3: System Flow Diagram

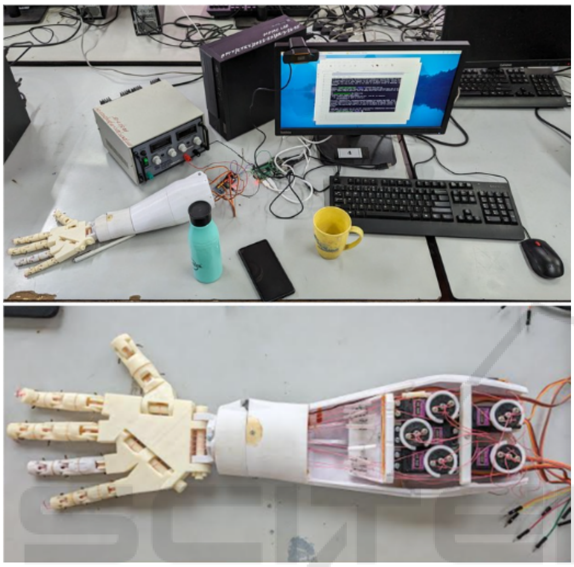

Figure 4: System Setup

4 RESULTS AND DISCUSSIONS

Originally, the model was initialized with three image categories including those of bottles, cell phones, and cups acquired from a database via the use of machine learning algorithms namely Decision Tree, Support Vector Machine (SVM), and Random Forest. Respective accuracies of 76%, 90%, and 95% were recorded, with Random Forest being the best among the three in the classification results. Nonetheless, the image extraction and pre-processing steps were best with obstacles, Proper pre-processing —using suitable parameters—was the main criterion for motivating the disambiguation of the images into the fewer, most peculiar types. Although some modifications have been made, a few images were incorrectly identified because of insufficient preprocessing. This erratic behavior shows a view of the flaws of the typical machine learning methods while handling the complex image datasets in cases where the dynamic preprocessing pipelines are required. This turned the attention to deep learning methods, which by definition include feature extraction and classification, thus making preprocessing the least necessary.

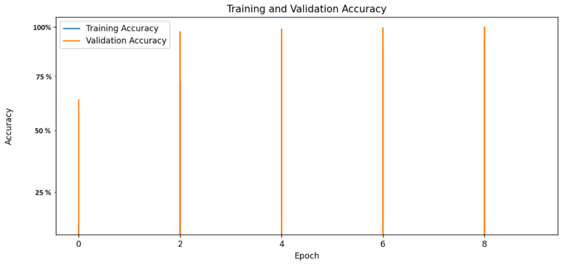

Figure 5: Model Accuracy

The Convolutional Neural Networks (CNNs) architecture was utilized to minimize the drawbacks of conventional machine learning techniques. CNNs are employed due to their ability of high-accuracy image classification as well as their usability in real-time applications for CNNs. There were many convolutions over the initial layers with ReLU as the activation function, max-pooling layers, and fully connected layers optimized for feature extraction and classification part. Using this approach significantly reduced preprocessing while simultaneously enhancing categorization accuracy. CNNs are also “plug-n-play” systems whereby they can be embedded in other systems thus efficiently working dependably in real-time situations, as confirmed by the studies already done.

Online learning algorithms such as the Decision Tree, SVM, and the Random Forest have glorious performance over structured datasets but they require plenty of preprocessing along the way and have serious problems when it comes to image data. Deep learning, particularly CNNs, not only feature extraction and classification together but also almost less preprocessing while still giving better results and being more reliable than other methods. CNNs are the best alternative for systems of information flow charts and AI supervised applications, which are ensured by their adaptability and dependability. Thus,CNN is deployed as an architecture because its superiority and flexibility are the two main reasons.

The proposed Convolutional Neural Network (CNN) model exhibited best-in-class performance in object grip classification, obtaining a training accuracy of 99% after only 10 epochs. This high accuracy points to the power of the selected architecture to extract the subtle visual features necessary for perfect object classification. The model employs the technique of multi-layer feature extraction in which layers become smaller, first of all, it detects low-level aspects and then proceeds to higher contextual signals. Hence, dropout regularization, in addition to improving the model’s generalizability, mitigated the overfitting, which was shown by the consistently smooth validation accuracy values in the training process.The

Table 1: Comparison of Proposed CNN Model with Reference Literature

Model	Accuracy
CNN Model (This Study)	97.5% (Test)
YOLO v2[1]	77.2%
YOLO v3[1]	77.2%
Faster R-CNN[1]	93%
ResNet	93.75[1]%
CNNs (General)[2]	93%
CNNs Region Proposals[3]	94.6%
CNNs for Grasp Recognition[4]	94.6%
Faster R-CNN with MobileNetV1[5]	94.26%

[1](Ujjwal Sharma and Singh, 2023)[2](Xia Zhao and Parmar, 2024)[3](Ross Girshick and Malik, 2014)[4](Chunyuan Shi and Liu, 2020)[5](Meena Laad and Saiyed, 2024)

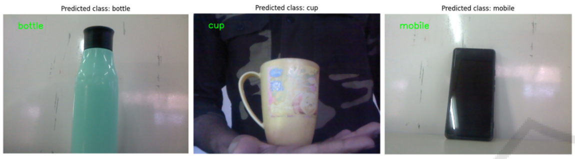

Figure 6: Output Prediction

testing phase confirmed the model's strength in real situations, as it reached out to an accuracy of 97.5% with the symmetrical loss of 0.05. These results reflect the algorithm's capability to maintain a balance between precision and efficiency, hence, it is a go-to for real-time prosthetic grip detection systems. The utilization of data augmentation methods, among the most reliable techniques, has substantially sustained the model against changes such as shifting, resizing, and illumination. Fig. 5 shows models training and validation accuracy.

With a CNN model proposed, the results are far better than other state-of-the-art models located in the literature. Showing a training accuracy of 99% and a testing accuracy of 97.5%, the CNN model competes by offering better performances than YOLO, Faster R-CNN, and other deep learning structures. The capability of better accuracy with less preprocessing is a very effective option in real-time applications for example in the case of prosthetics hand control systems which need to be controlled by the user. Fig. 6 confirms how the model does deal with real-world objects' classification and, thus, perfectly gives the proof of the proposed method's distinguishing different object types.

The most noteworthy aspect of the proposed system is its smart computer vision-based prosthetic hand that accurately replicates many different hand movements including power grip, precision grip, and pinch grip. The CNN model's high accuracy allows not only the work in the software domain but also the control of the prosthetic hand. Hence, embedded systems as well as software domains have their own standout. The embedded system acts as a major player in achieving short and hence fast processing time and high functionality of the actuator system, making the control of the prosthetic hand completely transparent. If compared to the real sensor-based systems that employ EMG sensors, which can be quite costly and susceptible to failures, the addition of a camera that will detect grips gives the main advantage. Among them, the camera-based solution gives a cost-effective option maintaining high precision without resorting to bulky sensors, and this improves both robustness and reliability. This unique way of doing visual and embedded systems engineering work is a real option for the smart prosthetic hands development that becomes usual in these cases.

5 CONCLUSIONS

The intended addition to this study is a robust methodology, which leads to grip classification for prosthetic hands using the celebrated power of Convolutional Neural Networks (CNNs). During the early phases of the study, conventional classification techniques like Decision Tree, Support Vector Machine (SVM), or Random Forest were studied, producing a 76%, 90%, and 95% accuracy, respectively. However, they required too much preprocessing and failed to manage image datasets.

By using CNNs, we can overcome the preprocessing burden: there is no need for an extra feature extractor because of the inherent feature extraction and classification capabilities of CNNs, thus increasing efficiency and accuracy. This architecture stands with a surreal 99% training accuracy and 97.5% testing accuracy on a proposed model superior to state-of-the-art models such as YOLO v3 and Faster R-CNN. It is dropout regularization and data augmentation that guarantees that the model is robustly generalized to handle different conditions in the environment and dynamics of objects.

The application area of this CNN-based model goes beyond controlling a prosthetic hand and appears instead to be real-time applications, such as in robots, assistive devices, and automation systems, where accuracy and flexibility are essential. Cost-effectiveness would increase access and utility, as expensive sensors would be replaced by affordable vision-based methods.

As such, this study suggests the promise of CNNs over conventional methods for grip classification in quality with very reliable, real-time, low-cost solutions towards prosthetics and beyond. This aspect will make for future work towards optimizing this system in conjunction with advanced embedded systems to boost performance and real-world application enablement.

REFERENCES

Chopra, S. and Emran, T. B. (2024). Advances in ai-based prosthetics development: Editorial. *Journal of Biomedical Research*.

Chunyuan Shi, Dapeng Yang, J. Z. and Liu, H. (2020). Computer vision-based grasp pattern recognition with application to myoelectric control of dexterous hand prosthesis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(9):2090–2099.

Cloutier, A. and Yang, J. J. (2013). Control of hand prostheses: A literature review. In *Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. ASME.

Divya Pradip Roy, Md Zahirul Alam Chowdhury, F. A. and Hoque, M. E. (2021). Design and development of a cost-effective prosthetic hand for upper limb amputees. In *2021 Biomedical Engineering International Conference (BMEiCON-2021)*. IEEE.

Douglas Morrison, P. C. and Leitner, J. (2020). Egad! an evolved grasping analysis dataset for diversity and reproducibility in robotic manipulation. *IEEE Robotics and Automation Letters*, 5(3):4368–4375.

Kutlugun, E. and Eyüpoğlu, C. (2020). Artificial intelligence methods used in computer vision. In *2020 5th International Conference on Computer Science and Engineering (UBMK)*, pages 214–218. IEEE.

Meena Laad, R. M. and Saiyed, N. (2024). Unveiling the vision: A comprehensive review of computer vision in ai and ml. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. IEEE.

Ross Girshick, Jeff Donahue, T. D. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 580–587. IEEE.

Shripad Bhatlawande, Swati Shilaskar, U. D. and Gadgil, V. (2023). Grip-vision: Enhancing object handling with computer vision-based grip classification. In *2023 2nd International Conference on Futuristic Technologies (INCOFT)*. IEEE.

Somer M. Nacy, M. A. T. and Baqer, I. A. (2017). A novel approach to control the robotic hand grasping process by using an artificial neural network algorithm. In *Journal of Intelligent Systems*, volume 26, pages 215–231. De Gruyter.

Ujjwal Sharma, T. G. and Singh, J. (2023). Real-time image processing using deep learning with opencv and python. *Journal of Pharmaceutical Negative Results*, 14:1905–1908.

Xia Zhao, Limin Wang, Y. Z. X. H. M. D. and Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57:99.

Xiaotong Chen, Rui Chen, Z. S. Z. Y. Y. L. R. I. B. and Jenkins, O. C. (2019). Grip: Generative robust inference and perception for semantic robot manipulation in adversarial environments. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3988–3995. IEEE.

Zewen Li, Fan Liu, W. Y. S. P. and Zhou, J. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019.

Ștefan–Adrian Ionescu and Poboroniuc, M. (2023). 3d printed upper limb prosthesis controlled by the healthy hand. In *2023 IEEE Conference on Robotics and Automation*. IEEE.