A CNN-Based Computer Vision Interface for Prosthetics' Control

Emanuele Lindo Secco[1(✉)], Daniel David McHugh[1], and Neil Buckley[1,2]

[1] Robotics Laboratory, School of Mathematics, Computer Science and Engineering, Liverpool Hope University, L16 9JD Liverpool, UK
{seccoe,13005235,bucklen}@hope.ac.uk

[2] AI Laboratory, School of Mathematics, Computer Science and Engineering, Liverpool Hope University, L16 9JD Liverpool, UK

Abstract. In this paper we present a CNN-based Interface for the control of prosthetic and robotic hand: a CNN visual system is trained with a set of images of daily life object in order to classify and recognize them. Such a classification provides useful information for the configuration of prosthetic and robotic hand: following the training, in fact, a low cost embedded computer combined with a low cost camera on the device (i.e. a prosthetic or robotic hand) can drive the device in order to approach and grasp whatever object belong to the training set.

Keywords: Prosthetics · AI · CNN · Auto-grasping

1 Introduction

In today's world, there are more than 2.1 million people in the US alone that live with a prosthetic limb this number is expected to double by 2050 in recent studies the number of people each year that become an amputee in the US is 185,000. This results in around 300 to 500 amputations performed every day, with this many people becoming amputees the need for smarter and better prosthetics is a must as more people become amputees (Access Prosthetics, 2020).

One of the possible ways that prosthetics can become smarter is with the use of *Artificial Intelligence* (AI) this area in computer science is growing massively as the number of useful applications that stem of AI are endless. For example, most car industries now use robotics and AI to build cars as they are far quicker than humans and are more accurate this results in better produces for the business this is one of many uses that AI can offer. Another very important part of AI is the creation of intelligent machines that react and learn as humans do so when combined with prosthetics the idea of smart prosthetic that can move and think without the user having to interact with the device becomes a very real possibility.

This paper will highlight how AI and prosthetics could be used together to create smarter prosthetics that will help users interact with the real world better by improving these devices. Amputees could be given a better quality of life with the help of AI not only that, but with the help of AI the field of prosthetics could advance must quicker leading to better prosthetics devices for amputees.

X. Gao et al. (Eds.): MobiHealth 2021, LNICST 440, pp. 41–59, 2022.
https://doi.org/10.1007/978-3-031-06368-8_3

2 Materials and Methods

In this section of the paper, the techniques of controlling a prosthetic will be listed this will give information about what they are and how they are used in prosthetics then the two projects that the researcher has done will be explained in detail explaining what the project is and how it works.

2.1 Current Techniques for Controlling a Prosthetic

Prosthetics have been around for many years now and have changed many different times as our technology advances to become more precise and general smarter devices overall, the main aim for prosthetics is to mimic a real limb both in function and appearance so the user has the same freedom as a real limb would. One of the main advancements over the years within prosthetics is how the user can control a prosthetics. In this part of the paper, threepossible methods are discussed on how a prosthetics can be controlled.

1 – EMG- The first method that will be discussed is myoelectric prosthetics this has been used in prosthetics for many years now as it is one of the easiest and best ways to identify movement within the arm and translate it to the prosthetics limb. Myoelectric prosthetics are unique in the way they allow the user to control their limb as once the prosthetic is attached to the user it starts to detect and collect muscle and nerve activity from the body which then is translated to the limbs motors to perform the action that the user requires in a natural way that looks realistic and performs well. The way the prosthetic detects muscle and nerve activity from the body is a method know as *Electromyography* (EMG) this is when one or more small needles which are called electrodes are inserted into the arm and attached to the muscle.

When these electrodes detect electrical activity within the muscle, a wave is created on an oscilloscope (which is a monitor that displays electrical activity) from these waves the person can see if the user is moving their whole hand or just one finger by the number of waves on the screen. Also, these electrodes allow us to detect how intense this electrical activity is which tell us the amount of muscle contraction, which is happening.

One device that has been used for prosthetics that use's EMG is called MyoBand it was created by ThalmicLabs which is a band that is made from eight EMG sensors that when worn on the arm will start to read electrical activity this data then will be sent over Bluetooth to a dongle that will be inserted into another device like a computer that will read the transmitted data and perform the action that relates to what the user is doing with their arm. This device works just the same as one that would be attached to a prosthetic limb as proven by (McHugh, D. 2019). In which the MyoBand was connected to a prosthetics hand, and by detecting certain gestures using EMG, the prosthetic hand would move to different grasp types (Fig. 1).

2 – Computer Vision - The second method we want to focus on, is *Computer Vision* (CV): this is the method that most researchers are using to create smarter prosthetics this method has been used for many years now in different fields and given great results. An example would be most robotic arms in the manufacturing industry have computer vision as this allows them to see with the help from a camera and allows them to detect an object in their vision also they will be able to recognize different objects.

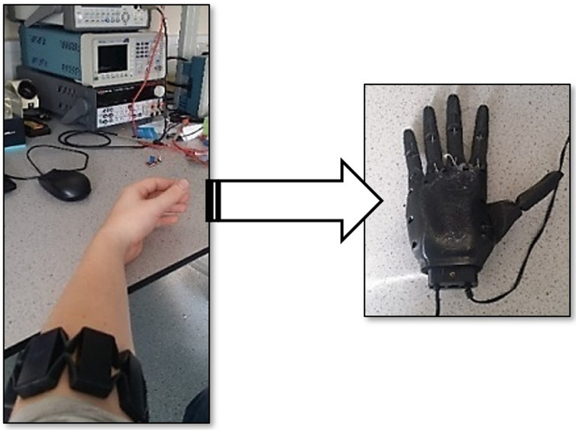

Fig. 1. The MyoBand connected to the open bionicsrobotic hand

As Computer Vision has been used for many years within robotics it's only natural that researchers would try to use it on prosthetics given the results when applied to robots, two main areas within computer vision that are being explored for prosthetics hands are object recognition/detection and gesture recognition. These are very important areas within prosthetics as by mastering these fields the ideal that smarter prosthetics could become more mainstream looks more realistic. Object recognition/detection is one of the main areas within prosthetics with the focus of giving prosthetic hands/arms the ability to see an object this should allow the hand/arm to be able to change into one of the set grasp types in order for the user to interact with the object more naturally and should result in faster response's times. When compared to a human this is the same with gesture recognition as well, but instead of using an object to get the hand/arm to perform the action the user will make a certain gesture that will tell the hand/arm to move in a certain way.

How this is done is using Artificial intelligence or AI to tell the hand that when certain object is in view it should perform this grasp type in order to interact with this object but a simple AI wouldn't be able to perform to the degree that is required for the hand to grasp an object properly, so certain methods within AI are used to make the AI perform better and allow it to handle complex tasks which increases its speed when compared to other methods. This method is called deep learning which use's artificial neural networks to be able to determent what the object is based on certain criteria for example; one network that could be used is a *Convolutional Neural Network* or CNN. This is one model that is used to automatically learn an object's features in order to identify that object this is done by feeding the model thousands of training images of different objects and getting the model to learn the different feature that makes certain objects different this is called feature extraction and is used in many different models.

For gesture recognition, the same model and the same methods can be used this is because feature extraction can work with any object like hands so it can be used for gestures as well. Some example work that other researchers have done can be seen here (Hu et al., 2018) and (Ghazaei et al., 2017). Another method that is used alongside feature extraction is edge detection this is an image processing technique for finding boundaries of objects within images this helps the feature extraction as it highlights that an object is in a certain area by the boundaries. Also, the model will be able to classify the object at the same time this is another important part of a deep learning model as it speeds up the process of correctly identifying objects.

3 – Brain Computer Interface (BCI or BMI) - The third method is one of the newest out of all three; it has been used mostly in the medical field to help with neuronal rehabilitation among other subjects in the medical field. But due to the potential of this topic, many researchers have tried to use it in their areas to solve problems that current technology can't or to find a better way to perform a solved problem, and this method is called Brain Computer Interface (BCI).

One of these fields is prosthetics as researchers believe that with BCI technology users will have better control of their prosthetic limb that is just as responsive as a human limb would be not only that but if this technology can be mastered then many people that have lost their function in one limb could be restored using a prosthetic. BCI works by acquiring electrical activities from the brain and nerves this step is very complex as the volume of signals that the brain fires off in a single second is unimaginable, so the task to read and understand what signals control muscle movement in the hand is very difficult. But with the help from AI, this task is possible to some degree this involves using deep learning with big data to gather vast amounts of data then trying to understand and classify what each signal means.

In one study the use of BCI and Myoelectric has been used together to create a system where the user could control the prosthetic by thought with the help of a MyoBand to collect the electrical signals (Hotson et al., 2016) not only this. But other researchers have tried to bring the sense of touch into prosthetics using BCI, which is on a different level compared to the other two methods (Kwok, 2013).

2.2 Convolutional Neural Networks (CNN)

It is important to understand how the CNN works to understand how the AI can detect and recognize each gesture as the CNN is the brain of the AI, to do these certain layers are used to extract the feature needed in order for the AI to make the right choice. Something to note with CNNs are they aren't all the same, some will have different steps, and some will have fewer steps it depends on how complex the CNN is, the database of images that were collected will be inputted into the CNN for training so the model can make a prediction when it sees the same gesture being performed by the user but that doesn't mean all images will be used for training some are used to test if the AI can correctly identify the gesture.

A simple overview of a CNN for classification could have this kind of architecture:

- *Input layer* - This will hold the pixel values of the images.
- *Convolution layer* - This will compute the output of neurons that are connected to the inputs, each computing the dot product of the inputs.
- *Pooling layer* - This will perform a down-sampling operation along the spatial dimensions.
- *Fully-connected layer* - This will compute the class scores, which provides the AI will the answer.

In this project each image has been pre-processed meaning the images have been turned into greyscale images so the pixel values have a range of 0 to 255 this means if you placed a pixel matrix over one of the images used you would be able to see its pixel values which help the AI understand what that pixel is. Another smaller matrix of numbers is created to perform convolution on the image, how convolution is done is by overlaying the smaller matrix over the image matrix and multiplying the numbers to create the dot of that value this will be saved into a new table called the convolved feature or feature map which is the most important information from that image that the AI will use later on.

Unlike with a normal CNN were each neuron is connected to all neurons, it would be better to use a method called Local Connectivity which connects each neuron to only a local region of input volume. The spatial extent of this connectivity is a hyper-parameter called the Receptive Field of neuron which is the size of the filter which is used by the CNN this is important to understand as the connectivity along the depth axis is always equal to the depth of the input volume it is also important to remember the asymmetry in spatial dimensions, e.g. (height and width) and the depth dimension.

An example of this could be that, suppose that the input has the sizes of [32 $\times$ 32 $\times$ 3], and the receptive field or filter size has a size of [5 $\times$ 5] then each neuron in the Conv layer should have a weight of [5 $\times$ 5 $\times$ 3] this totals into 75 weights and + 1 bias parameter also it's important the notice that the extent of the connectivity along the depth axis must be 3 since this is the depth of the input in this case.

To compute the spatial size of the output, a function can be performed which is the input size (W), the receptive field size/filter size (F), the stride which has been applied (S) and the amount of zero-padding that has been used (P). The formula below calculates how many neurons can fit

$$(W - F + 2P)/S + 1 \tag{1}$$

Parameter Sharing is a scheme to control the number of parameters in the neural network. By using this scheme the number of parameters used can be strongly reduced

$$W_2 = (W_1 - F + 2P)/S + 1 \tag{2}$$

where $\mathrm{H_2 = (H_1 - F + 2P)/S + 1}$ and $\mathrm{D_2 = K}$. Pooling layer this step is between the convolution layer and is used to reduce the dimensionality of the feature map as using higher dimensions can cause some issues as it confuses the CNN in later steps with the amount of noise, so to help the AI it's better to get rid of the dimensions without losing any important information in the process.

Max pooling is one of the methods used which uses a filter matrix of any size and a stride to down-sample every depth of the input through this method around 75% of the input will be down-sampled. The formula's for pooling layer holds:

$$W_2 = (W_1 - F)/S + 1 \tag{3}$$

where $H_2 = (H_1 - F)/S + 1$ and $D_2 = D_1$. The next step is the fully-connected layer this is when the neurons have full connections to all activations in the previous layer. These activations can be computed with a matrix multiplication which is followed by a bias offset.

2.3 Gesture Recognition

Gesture recognition is the ability for computers to capture and understand human gestures as commands and perform certain actions depending on the gestures an example could be a wave of the hand to start up the system, or it could be a peace sign to put the system to sleep. The amount of gestures a human can do is unlimited as a gesture is defined as any physical movement, which is non-verbal; gesture recognition has been around. For many years now and has become more popular as the potential use becomes more evident in today's world some of the most popular examples of gesture recognition would be the Wii, X-box Kinect and PlayStation Move.

In our first project gesture recognition was used in order to detect five different hand gestures from the user using a live camera feed and a keyboard input from the user, when the program starts it will only display the first screen which will be used to capture the gestures from the user, but in total three screens can be displayed at once. The first screen will be a live-feed of what the camera sees with the Region of Interest (ROI) being a blue square box this is important as the user must perform the gesture within that region in order for the AI to detect what gesture the user is performing. Otherwise, the AI won't work as it will only detect within that area this is because of the amount of resource it takes for the AI to work so the ROI will be half of the main screen size, but this can be changed to fit the whole screen if the user requires it to be.

The second screen is a real-time grey-scaled camera feed this is called the Binary thresholding screen this will activate when a certain key is pressed and can be used to save new gestures that the user can use to train the AI on. This gives the user an easy way to add new gestures as some of the pre-processing is done, but the main use of this screen is when the space bar is pressed it will give the user the prediction score and the predicted gesture the AI thinks the user has performed based on what gesture was seen on the first screen in the ROI. Finally, the third screen shows the contour matrix in real-time this shows the users how the AI sees the hand and how it can detect the different gestures (Fig. 2).

Fig. 2. Screenshots of the three stages of fist gesture recognition

Once the AI was able to predict the gesture, it would then send a character to the prosthetic hand which would then move the actuators into the correct position so that the prosthetic hand would be mimicking the user's gesture.

For this project to work three main technologies have been used these are OpenCV, Keras and TensorFlow each of these have been used to handle certain parts of the project and played an important role in the final version not only that but the model that was used is very important as it is the brain of the AI which allows it to predict the gesture the user has done.

The VGG-16 model works by sending the image through a stack of convolution layers then it uses a filter with a small receptive field normally the size is 3×3 which is the smallest size to capture all directions of the image. The convolution stride is fixed to 1 pixel the spatial padding of convolution layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 convolution layer, spatial pooling is carried out by five max-pooling layers, which follow some of the convolution layers not all the convolution layers are followed by max-pooling. Max-pooling is performed over a 2×2 pixel window with stride 2.

2.4 Object Detection and Recognition

Object detection and recognition is one of the main topics when computer vision is brought up these two subjects have been used in many different applications from face recognition to live video detection normally these are used together as when used together an AI can see that an object is in view and it can understand what the object is. But they both perform very different job in object detection the AI will only be able to see if any object is present so it will highlight that an object is in view while object recognition will understand what that object is so it will tell you what the object is and normally give you a percentage score which will inform the user how accurate the AI thinks it is.

In the second project an AI was used to detect and recognise multiple objects this ranged from humans to household objects it uses a live-feed camera to detect and recognise any object in real-time that comes within view of the camera this was built using python and runs on *Jupyter notebook* which is an IDE that can be downloaded from Anaconda. When the program runs, it will start by loading up a screen, and this is what the camera can see which can be resized depending on what the user wants. In this project the screen was around half of the monitor, this was due to the amount of resources that were available on the system as the larger the screen, the more resources it would take to run the AI. When the screen is loaded, the user can then start to place objects within the view of the camera which will allow the AI to see if it can first detect that an object is present then see if the AI can recognize what the object is.

In this project, many different libraries have been used some of these are for support to help with images like Pillow or to handle the camera like OpenCV while other libraries are the main backbone of the project like TensorFlow and Protobuf. One of the main libraries that this project uses is called Protobuf this is developed by Google and used to configure the model and training parameter.

This project uses a special kind of model which is called You Only Look Once (YOLO) this is one of the newer models that has been used for object detection and has started to grow quite rapidly within the machine learning family. YOLO models are known for their speed as when compared to other models like R-CNN they often perform much faster which is one of the reasons why it was used in this project as speed is very important when you are using a real-time detection and recognition AI.

YOLO first works on splitting the input image into a grid of cells these cells are responsible for predicting the bounding box if the center of a bounding box falls within it, each cell will predict a bounding box and share its X, Y coordinate, height, width and confidence score also class prediction is based on each cell's information. While the image is being split into different bounding boxes multiply convolution layers, and max-pooling layers are processing the image to decide what the probability is that an object is inside each of the cells.

However not all bounding boxes would have an object within them if this is the case one of the jobs is to remove these bounding boxes as they can lead to producing bad results this is done based on the predicted confidence score of each cell that if the score is less than a certain threshold which is set to 0.24 then remove this cell as it should be redundant as it doesn't allow the AI to detect the object.

The first important loss function in the YOLO model is the confidence loss function this allows the model to measure the 'objectness' of each cell which will produce a value which is the confidence value for the whole boundary box. If an object is detected in the boundary box then this function will be used (Eq. 4), if no object is detected in the boundary box then this function will be used (Eq. 5).

$$\sum_{i=0}^{S2} \sum_{j=0}^{B} 1obj\ ij\left(Ci - \acute{\mathrm{C}}\mathrm{i}\right)^2 \tag{4}$$

where $\acute{C}\,i$ is the box confidence score of box j in cell i. When $obj\ ij = 1$ then the box j in cell i is responsible for detecting the object.

$$\lambda\ noobj \sum_{i=0}^{S2} \sum_{j=0}^{B} 1noobj\ ij\left(Ci - \acute{\mathrm{C}}\mathrm{i}\right)^2 \tag{5}$$

where *1noobj ij* is the complement of *1obj ij,* $\acute{C}\,i$ is the box confidence score of box j in cell I and $\lambda\,noobj$ weights down the loss when detecting background. Another loss function in the YOLO model is the classification loss function this is used if an object has been detected; the classification loss at each cell is the squared error of the class conditional probabilities for each class (Eq. 6).

$$\sum_{i=0}^{S2} 1obj\ i \sum_{C_\varepsilon\ classes} \left(\mathrm{pi(C)} - \tilde{\mathrm{P}}\mathrm{i(C)}\right)^2 \tag{6}$$

where *1 obj i* = 1 if an object appears in cell *i*, otherwise 0 and $\widetilde{P}\,i(C)$ *denotes the conditional class probability for class C on cell i.* Finally, another loss function in the YOLO model is the localization loss which measures the errors in the boundary boxes location and sizes this function is only used when the boxes have objects within them (Eq. 7)

$$\begin{aligned} &\lambda\ coord \sum_{i=0}^{S2} \sum_{j=0}^{B} 1obj\ ij\left[(xi - Xi)^2 + (yi - Yi)^2\right] \\ &+ \lambda\ coord \sum_{i=0}^{S2} \sum_{j=0}^{B} 1obj\ ij\left[\left(\sqrt{wi} - \sqrt{Wi}\right)^2 + \left(\sqrt{hi} - \sqrt{Hi}\right)^2\right] \end{aligned} \tag{7}$$

where *1obj ij* = 1 then the box *j* in cell *i* is responsible for detecting the object and the λ *coord* increase the weight for the loss in the boundary box coordinates.

The last important function in the YOLO model is the loss function for an iteration of t which is an objective function that is a multi-part function that tells the model what to do if a bounding box doesn't have any objects within it. Its confidence of objectness needs to be reduced and shown as a first loss term this will tell the model that no object is present within this box as bounding boxes coordinate prediction need to align with prior information a loss term reducing the difference between prior and predicted is added for a few iterations. If a bounding box k is responsible for a truth box, the predictions need to be aligned with the truth values which are represented as the third loss term the **λ** values are the pre-defined weightages for each of the loss terms (Eq. 8).

$$\begin{aligned} losst = &\sum_{i=0}^{W} \sum_{j=0}^{H} \sum_{k=0}^{A} 1MaxIOU < Thresh\ \lambda noobj * (-boijk)^2 \\ &+ 1t < 12800\ \lambda prior * \sum r\varepsilon(x, y, w, h)(priorrk - brijk)^2 \\ &+ 1truthk\ (\lambda coord * \sum r\varepsilon(x, y, w, h)(truthr - brijk)^2 \\ &+ \lambda obj * (IOUktruth - boijk)^2 \\ &+ \lambda class * \left(\sum_{C=1}^{C}(truthc = bcijk)^2\right) \end{aligned} \tag{8}$$

To finalize the methodology section, both projects that have been successfully created for this paper and have been explained in great detail this includes a piece of brief information about the subject, details on how the project was created, some information about the issues faced while working on the project and finally technology used to create the project. In the next section of the paper, the results of the project will be displayed. This will include images of the projects working and a detailed explanation of each result.

3 Results

In this section of the paper, the results that have been collected from both projects will be displayed. This will demonstrate how projects have done and will show that both projects were able to work accurately.

3.1 Gesture Recognition

All of the five gestures used in this project will be displayed below this is to show how well each of the gestures is recognized by the AI after that a table will display the overall accuracy of the project for each of the gestures used in the project.

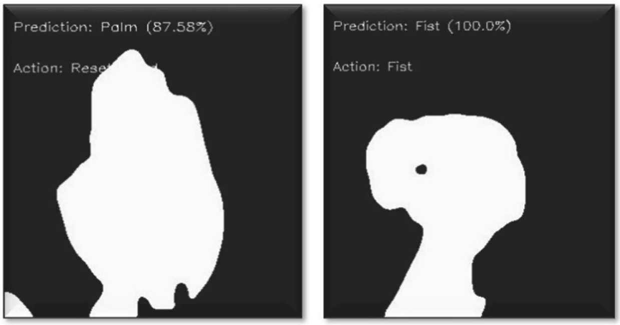

Fig. 3. On the left and right panels, the recognition of the palm and fist, respectively

As can be seen in Fig. 3 (left panel) this is the *palm gesture* that is included in the five gestures that were used in the project, as you can see when the test was performed the AI was able to correctly predict what gesture the used performed and give the researcher a very good prediction score. In this case the score was 87% which is really good this means that the AI is 87% sure that the gesture performed is the Palm gesture not only that, but the action that follows the gesture is working too as each gesture results in a different action taking place, in this case, the Palm gesture should result in the hand being reset. What this result has shown is the AI can understand what gesture the user is performing, which allows the correct action to be translated to the hand.

As can be observed in Fig. 3 (right panel) this is the *fist gesture* that is included in the five gestures that were used in the project, as you can see when the test was performed the AI was able to correctly predict what gesture the used performed and give the researcher a very good prediction score in this case the score was 100% which is great this means that the AI is 100% sure that the gesture performed is the Fist gesture. This result is unusual as getting 100% in prediction AI is not really common, if the AI is showing too many 100% it could indicate that the AI has some issues, or it could just be this one test that went really well, not only that but the action that follows the gesture is working too as each gesture results in a different action taking place, in this case, the Fist gesture should result in the hand being reset. What this result has shown is the AI can understand what gesture the user is performing, which allows the correct action to be translated to the hand.

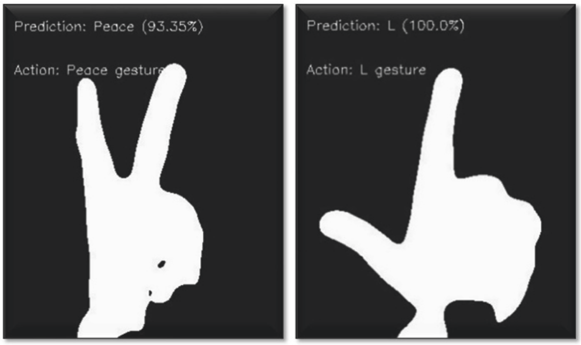

Fig. 4. On the left and right panels, the recognition of the peace gesture and L-shape gesture, respectively

In Fig. 4 (left panel) the *peace gesture* that is included in the five gestures that were used in the project, as you can see when the test was performed the AI was able to correctly predict what gesture the used performed and give the researcher a very good prediction score. In this case the score was 93% which is really good this means that the AI is 93% sure that the gesture performed is the Peace gesture not only that, but the action that follows the gesture is working too as each gesture results in a different action taking place, in this case, the Peace gesture should result in the hand being reset. What this result has shown is the AI can understand what gesture the user is performing, which allows the correct action to be translated to the hand.

Fig. 5. Recognition of the okay gesture

Figure 4 (right panel) shows the L gesture that is included in the five gestures that were used in the project, as you can see when the test was performed the AI was able to correctly predict what gesture the used performed and give the researcher a very good prediction score. In this case, the score was 100% which is good this means that the AI is 100% sure that the gesture performed is the L gesture like with the fist gesture this could be a very good test result or something wrong with the AI more testing will confirm this to the researcher. Not only that but the action that follows the gesture is working too as each gesture results in a different action taking place in this case, the L gesture should result in the hand being reset. What this result has shown is the AI can understand what gesture the user is performing, which allows the correct action to be translated to the hand.

Figure 5 displays the *Okay gesture* that is included in the five gestures that were used in the project, as you can see when the test was performed the AI was able to correctly predict what gesture the used performed and give the researcher a very good prediction score. In this case, the score was 100% which is good this means that the AI is 100% sure that the gesture performed is the Okay gesture like with the fist gesture this could be a very good test result or something wrong with the AI more testing will confirm this to the researcher. Not only that but the action that follows the gesture is working too as each gesture results in a different action taking place in this case the Okay gesture should result in the hand being reset. What this result has shown is the AI can understand what gesture the user is performing, which allows the correct action to be translated to the hand.

Table 1. Accuracy of the five gestures' recognition [%]

Trial	Gesture 1 - Palm	Gesture 2 - Fist	Gesture 3 - Peace	Gesture 4 - L	Gesture 5-Okay
1	87	100	93	100	100
2	75	90	80	69	80
3	88	93	77	73	74
4	98	81	68	71	84
5	74	89	82	87	90
Average	84	92	80	80	86

Table 1 is a table that holds the results of each gesture that has been tested five times using our AI not only that, but at the bottom of the table, each gesture has been given an average score based on the results from the tests. All the results are quite good none of the prediction scores are less than 50% which is really good as a score which is 50% or lower tells us that the AI isn't really sure amount the prediction even if the prediction is correct, but our AI is often confident about the prediction which is why these scores are higher as the AI is confident about the result. Not only are the individually scores well the Average scores are good as well especially the Fist gesture which performed the best overall this could be because of how simple the gesture is compared to some of the others like the *Okay* and *L gesture*, or it could be that the training data was better.

Overall the gesture results have proven that the gesture recognition project was successful as the AI was able to understand what gesture the user had performed and was able to provide a good prediction score with the correct action for each of the gestures. The only part of the project that couldn't be tested was how well it would communicate with the robot hand, at the early stages of this project a robotic hand was used to see how communication could be done and the results were good the AI was able to send a letter when a certain gesture was performed and the hand would respond but this tested was done very early, and no evidence was recorded.

3.2 Objects Detection

As the number of objects that can be detected and recognized by this projects AI is many, it makes sense only to show a few objects to prove that the AI works for different objects to show that the AI works only five single objects will be shown and one multiple objects. Six results will be displayed each with an explanation about the result after that a table will show the accuracy of these five objects when tested multiple times.

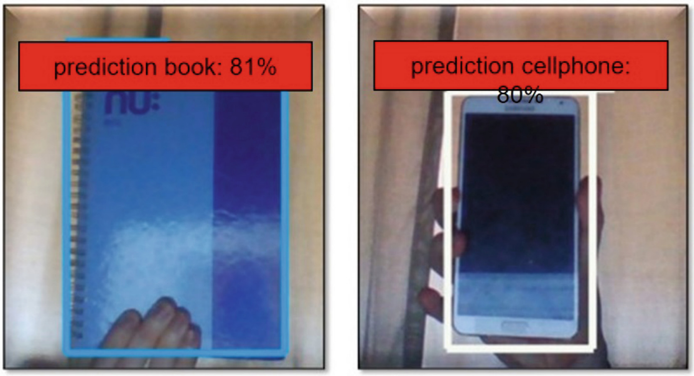

Fig. 6. Object recognition performance while detecting a book (left panel) and a mobile phone (right panel)

In Fig. 6 (left panel) which was the first object that was tested on the AI out of the five single objects as you can see from the image above the AI has been able to detect that an object is present and the AI has also been able to recognize what the object is correctly. This can be seen by the border that the AI creates around the object, in this result, the AI believes this object is a book with a confident score of 81% which is a very high score this tells the researcher that the AI is certain that this object is a book.

Figure 6 (right panel) is the second object that was tested on the AI out of the five single objects as you can see from the image above the AI has been able to detect that an object is present and the AI has also been able to recognise what the object is correctly. This can be seen by the border that the AI creates around the object, in this result, the AI believes this object is a phone with a confident score of 80% which is a very high score this result tells the researcher that the AI is certain that this object is a phone.

Figure 7 (top left panel) is the third object that was tested on the AI out of the five single objects as you can see from the image above the AI has been able to detect that an object is present and the AI has also been able to recognize what the object is correctly .

This can be seen by the border that the AI creates around the object, in this result the AI believes this object is scissors with a confident score of 69% which is still a good score not as good as the other two but still good enough for this object project this result tells the researcher that the AI is certain that this object is a pair of scissors.

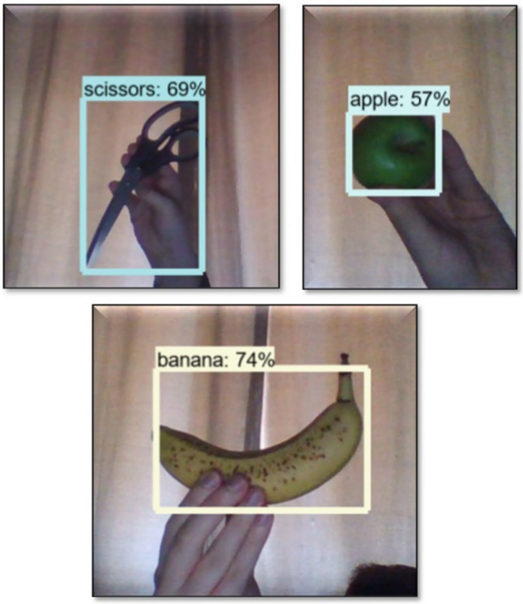

Fig. 7. Object recognition performance while detecting a scissor (top left panel), an apple (top right panel) and a banana (bottom panel)

As it can be noticed in Fig. 7 (top right panel), this was the fourth object that was tested on the AI out of the five single objects as you can see from the image above the AI has been able to detect that an object is present and the AI has also been able to recognize what the object is correctly. This can be seen by the border that the AI creates around the object, in this result the AI believes this object is an apple with a confident score of 57% which is still a good score not as good as the other two but still good enough for this object project as any score over 50% is useable this result tells the researcher that the AI is certain that this object is an apple.

As it can be observed in Fig. 7 (bottom panel), this was the 5th object that was tested on the AI out of the five single objects as you can see from the image above the AI has been able to detect that an object is present and the AI has also been able to recognise what the object is correctly. This can be seen by the border that the AI creates around the object, in this result the AI believes this object is a banana with a confident score of 74% which is a very good score not as good as the first and second objects but still good for this object project this result tells the researcher that the AI is certain that this object is a banana.

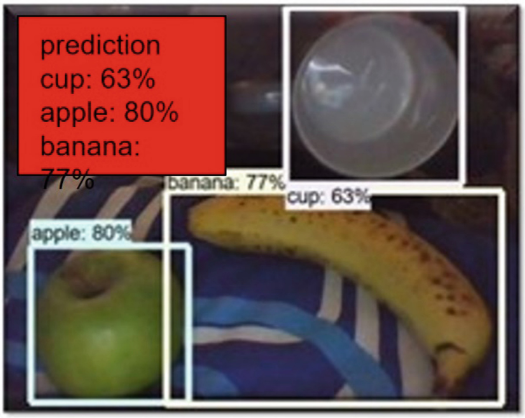

Fig. 8. Object recognition performance while detecting multiple objects

Figure 8 is the result from the multiple objects results that were taken from the AI project as you can see the AI has been able to detect that more than one object is present and has also been able to correctly recognize each of the objects that the AI can see. This can be seen by the borders that the AI creates around the objects, in this result, the AI believes that three objects are present which is an apple with an 80% confident score, banana with a 77% confident score and a cup with a 63% confident score. All these results are good scores as the AI has to work harder to understand the borders of each object and what each object is, this result tells the researcher that the AI can detect and recognize more than one object at one time.

Table 2. Accuracy of the object recognition, single object configuration [%]

Trial	Object 1 book	Object 2 phone	Object 3 scissors	Object 4 apple	Object 5 banana
1	81	80	69	57	74
2	85	95	79	66	69
3	78	82	61	73	59
4	95	72	81	86	86
5	73	77	83	69	74
Average	82	81	75	70	72

Table 2 holds the results of some of the single object which have been tested five times using our AI not only that but at the bottom of the table, each object has been given an average score based on the results from the tests. The individual scores of the objects are very good as these scores prove that the AI can accurately recognise these objects with the highest being 95% which is the phone and book object and the lowest being 57% which is the apple object. Not only are the individual scores good the average scores are very good as well this proves that the AI can understand what these objects are most of the time which is useful as it allows the researcher to know which object is easier for object recognition and which are harder.

Table 3. Accuracy of the object recognition, multiple objects configuration [%]

Test number	Book, phone & cup	Phone, apple & cup	Cup, banana & book	Apple, banana & cup	Banana, book & phone
1	72	80	84	67	70
2	68	78	76	70	88
3	69	72	60	86	81
4	91	58	79	88	63
5	70	85	68	73	87
Average	74	75	73	77	78

Table 3 holds the results of all of the multiply object that have been tested these objects have been tested five times using our AI not only that but at the bottom of the table, each object has been given an average score based on the results from the tests unlike the single objects test the individual scores here are an average of all three objects that performed in that group.

The individual scores of the multiply objects are very good as these scores prove that the AI can accurately recognize these objects with the highest being 91% which is the Book, Phone and Cup group and the lowest being 58% which is the Phone, Apple and Cup group. Not only are the individual scores good the average scores are very good as well this proves that the AI will understand what these objects are most of the time which is useful as it allows the researcher to know which object is easier for object recognition and which are harder.

By comparing Tables 2 and 3 together certain statements can be made about the AI, one of the statements that can be said is that this AI performs better at single object detection and recognition compared to multiply object detection and recognition this is shown by the average scores in both tables as in the single objects table the overall scores are higher. Which tells the researcher that single object has a higher chance of being detected and recognised by the AI not only that, but the individual scores are higher as well, which further supports this statement.

Overall the object detection and recognition project was a success it did everything that the researcher wanted it was able to detect most objects that were placed in front of the camera, and when it did detect an object, it was able to tell us what that object was with a good prediction score. The only part of the project that wasn't done was linking the AI to a robotic hand, but given that most of the work has already been done it wouldn't have been hard to write a program that could talk with the hand.

To finalize the result section, both projects have provided very good results that show how well each of the projects performed when tested this can be seen in the figures and the tables that are shown above also each of the results have been talked about in detail to explain what they mean and finally a brief overview of how the project went in terms of the results has been detailed. In the next section of the paper, the conclusion will be discussed this will summaries the projects and include the researcher's thoughts about the projects and discuss the future works of this project and the field itself.

4 Conclusion

In this section of the paper, the paper will be summarized this will include what the researcher's thoughts about the projects are also the future works of this project will be talked about plus the future works of this field will be discussed as well.

4.1 Summary

This paper set out to prove that with the help of AI, prosthetic devices could become smarter, leading to better ways that prosthetic limbs could interact with the real world. This was first shown with the gesture project by using an AI to detect and understand what each gesture did it allowed the researcher to control a robotic hand with ease this resulted in the robotic hand being able to switch to different grasp position which allowed the robotic hand to interact with different objects in the real world. The object project followed the same idea as the gesture project but used objects instead of hand gestures, using an AI to detect and recognize different objects which would then move a robotic hand so it could interact with that object in the real world.

Both projects have shown that the use of AI and prosthetics together are key in order to achieve smarter prosthetic as they far surpass any prosthetic device without AI assistants, that being said one of the main issues with AI-assisted devices at this time is how to implement the AI. Therefore it can talk with the device or how to implement the AI into the device itself as having a camera attached to the device isn't really suitable as the idea of a prosthetic is to mimic a real limb so having a visible camera closely would backfire of this idea, but other ways methods could be used in order to allow the AI to see and talk with the prosthetic limb.

The final thoughts about how the projects went now that the paper is finished are both projects did what the researcher wanted and fully prove what the paper set out to do while both projects presented certain issues while in development which is to be expected the researcher was able to overcome them and create two AI's that provide great results. Unfortunately, certain parts of the project couldn't be done due to the lack of equipment which the researcher required nevertheless both projects were still able to provide very good results this ended with the researcher feeling very satisfied with both projects.

4.2 Future Research

In terms of both projects, the next step would be to use the AI with a robotic hand to see how well both can talk to the robotic hand and to see what results can be collected from these tests this would include creating programs. So that the hand could understand the signals from the AI also another idea that could be implemented for the gesture project is having a multiple gesture recognition system like with the object project this way the user could perform two gestures at one time which could lead to more gestures being programmed into the hand. But other research that could be explored is the use of BCI with prosthetics this field is still quite new compared to the other fields as the use of thoughts to control anything is still out of our hands as researchers don't fully understand how to use thoughts to move complex objects like hands or arms, so this field is one of the most exciting for prosthetics as it opens many different doors which the other methods could not.

The possible applications from these projects are quite different as both are very different projects, from the gesture project if some UI could be created then one of the applications could be a smart home controller where one gesture could turn on the device while the other gestures control certain parameters of the device like volume, brightness, lock and unlock. Another application could be for medical use with a focus on rehabilitation as it could be used to collect data on how well the user can performing each hand gesture this would tell you if the user can perform the gesture and how well they can perform the gesture (see also Maereg et al., 2020; McHugh et al., 2020; Secco et al., 2020; Myers et al., 2021).

The performance of these systems clearly depends on the type of technology that is used vs the inter-subjects and intra-subject variability of the data: for example, in (Maereg et al., 2020) we showed that a hand gesture recognition system based on near-infrared sensing wristband allows an accuracy - over all the subjects involved in the study – of 98% with a std of 1.8%. However, such a system requires a proper training of the recognition system and a proper set-up of the sensors whose architecture involves a higher redundancy vs the proposed system in this paper.

Focusing on this CNN CNN-based Computer Vision Interface, it should be also mentioned the importance of considering the introduction of a set of features in the pre-processing of the data in order to improve the data organization and the structure of the information which are initially provided into the CNN (Buckley et al., 2021). Moreover, a proper further analysis of the Yolo drawbacks and of the advantages and disadvantages vs other techniques should be performed (Howard et al., 2022).

For the object project, many different applications could be used the first being robotics as many robots already use the same technology it could be easily integrated to work with a pick and place robot as the AI can already detect objects the only difference is some of the code would need to be changed. So, the AI knows what to do when it detects the objects. Another application could be tracking objects which could be used in surveillance so when an object comes within view of a security camera it could be able to detect what object is in front of the camera and be able to track where that object has moved this could be used on multiple cameras. So, the AI could track a person that is moving from one camera's view to another's camera view this could help identify where a person is going as the AI could highlight them making it easier to see where they are heading which could be used in police work.

Acknowledgements. This work was presented in thesis form in fulfilment of the requirements for the MSC in Robotic Engineering for the student Daniel David McHugh under the supervision of E.L. Secco from the Robotics Laboratory, School of Mathematics, Computer Science and Engineering, Liverpool Hope University.

References

McHugh, D.: Development of an interface between openbionics robotic hand and wearable Myo band. Liverpool, 1–21 (2019a)

McHugh, D.: Shows MyoBand connected to prosthetic hand (2019b). Accessed 1 July 2019

Ghazaei, G., Alameer, A., Degenaar, P., Morgan, G., Nazarpour, K.: Deep learning-based artificial vision for grasp classification in myoelectric hands. J. Neural Eng., **14**(3), 036025 (2017). Accessed 5 July 2019

Hotson, G., et al.: Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. J. Neural Eng. **13**(2), 026017 (2016)

Hu, Z., Hu, Y., Liu, J., Wu, B., Han, D., Kurfess, T.: 3D separable convolutional neural network for dynamic hand gesture recognition. Neurocomputing, **318**, 151–161 (2018). https://www.sciencedirect.com/science/article/pii/S0925231218309925. Accessed 5 July 2019

Maereg, A.T., Lou, Y., Secco, E.L., King, R.: Hand gesture recognition based on near-infrared sensing wristband. In: Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020), pp. 110–117 (2020). ISBN: 978–989–758–402–2. https://doi.org/10.5220/0008909401100117

McHugh, D., Buckley, N., Secco, E.L.: A low cost visual sensor for gesture recognition via AI CNNS. In: Intelligent Systems Conference (IntelliSys), Amsterdam, The Netherlands (2020)

Secco, E.L., Scilio, J.: Development of a symbiotic GUI for robotic and prosthetic hand. In: Intelligent Systems Conference (IntelliSys), Amsterdam, The Netherlands (2020)

Myers, K., Secco, E.L.: A low-cost embedded computer vision system for the classification of recyclable objects. In: Sharma, H., Saraswat, M., Kumar, S., Bansal, J.C. (eds.) CIS 2020. LNDECT, vol. 61, pp. 11–30. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4582-9_2

Access Prosthetics. 15 limb loss statistics that may surprise you - access prosthetics (2020). https://accessprosthetics.com/15-limb-loss-statistics-may-surprise/. Accessed 5 June 2019

Kwok, R.: Neuroprosthetics: once more, with feeling. Nature **497**(7448), 176–178 (2013)

Buckley, N., Sherrett, L., Secco, E.L.: A CNN sign language recognition system with single & double-handed gestures. In: IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1250–1253 (2021)

Howard, A.M., Secco, E.L.: A low-cost human-robot interface for the motion planning of robotic hands. In: Arai, K. (ed.) IntelliSys 2021. LNNS, vol. 296, pp. 450–464. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-82199-9_30