Improving Vision in Retinal Prostheses with Artificial Intelligence

Jacob Perala
Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA 56267
pera0064@morris.umn.edu

ABSTRACT

Retinal prosthetics hope to restore functional vision to the millions of people around the world experiencing diseases causing degradation of vision. Approaches in artificial intelligence, namely computer vision and deep learning models, have aided in the improvement of the vision restored in retinal prosthetic devices and of specific challenges including indoor and outdoor images, facial recognition, and collision avoidance. This paper will provide a survey on how these models have improved the quality of vision in retinal prosthetics and aided those with specific retinal degenerative diseases. Techniques using computer vision are largely responsible for processing regions of interest in a scene, and have proven effective at providing a better understanding of environments in prosthetic vision. The needs of those undergoing therapy vary and points towards medical challenges involved that need to be kept in mind when simulating prosthetic vision for testing. The application of artificial intelligence in these devices has proven to be effective when compared to previous methods of processing and have the potential to provide functional sight for patients.

Keywords

Neural Network, CNN, Computer Vision, Retinal Prosthetic

1. INTRODUCTION

Hoping to aid those with retinal degenerative diseases causing vision loss, a retinal prosthetic (RP) is designed to restore functional vision. In the last decade, advancements in processing power and artificial intelligence have improved vision in RP systems. This paper presents a survey of how artificial intelligence can be used in retinal prostheses to, at least partially, restore vision to the visually impaired. To achieve this, researchers have applied computer vision, neural networks, and other deep learning methodologies to improve upon previous models.

Retinal degenerative diseases affect millions and creates an economic burden costing hundreds of billions of dollars in the United States alone [12]. Both the economic burden and the cost of comfort have driven researchers towards a viable RP, resulting in over five hundred implanted prosthetics over the last fifteen years with three regulation approved devices discussed by Ayton et al [4]. These devices are cleared for treatment of two diseases, retinitis pigmentosa and age-related macular degeneration, which cause the degradation of light sensitive cells in the retina.

This paper first introduces background for the topic including the pipeline to create an image for a retinal prosthetic, a few of the biological requirements of the eye, and an overview of some artificial intelligence approaches. In section 3, a description of the relevant subsets of AI (neural networks and computer vision) will be introduced in addition to their role in retinal prosthetics. Section 4 will delve further into the methodologies and testing used, which can differ in machine learning approaches and if the testing was performed with an implanted patient or by simulating prosthetic vision. Concerns that arise when researching the use of a RP in the real world, in comparison to controlled environments or simulations, will be discussed in section 5 followed by future work and closing thoughts.

2. BACKGROUND

The earliest attempts to restore vision date back to the 1700's when Charles Le Roy attempted to cure a patient's blindness with electrical stimulation [4]. Since then a myriad of improvements have been made in both the ability to process and to deliver information for patients with a retinal prosthetic. This section will cover these improvements and the role they play in the context of a retinal prosthetic (RP), in addition to necessary biological components of the eye and system architecture of current models.

2.1 Basic Functions of the Retina

Responsible for receiving and processing light, the retina is a thin layer of tissue lining the interior of the back of the eye. One of many layers in the retina, the photoreceptive layer is composed of rods and cones that are responsible for sensing light, and producing black-and-white or color vision. After the light has reached the retina, the light is registered by the photoreceptive cells and triggers a series of chemical reactions. These reactions are passed on to the ganglion cells, commonly referred to as retinal ganglion cells (RGCs), which is the final output of neurons in the retina. It is the responsibility of the RGCs to convert these chemical signals into neural impulses which are sent to the brain. The brain then registers and decodes these impulses to determine the image and its features.

2.2 System Architecture of Current Models

Current RP models typically consist of four system com-

UMM CSci Senior Seminar Conference, December 2021 Morris, MN.
.

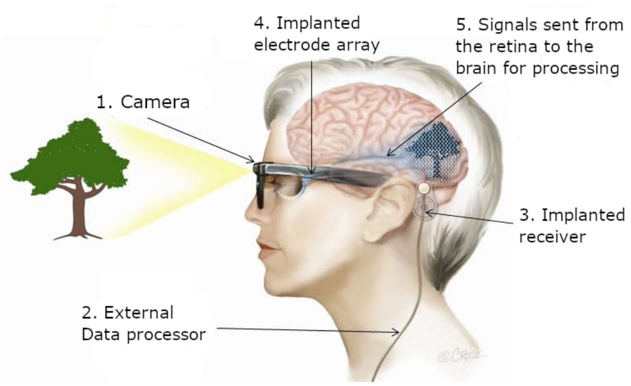
4. Implanted electrode array
5. Signals sent from the retina to the brain for processing
1. Camera
3. Implanted receiver
2. External Data processor

Figure 1: Retinal prosthetic system structure [1]

ponents, shown in labels 1–4 in Figure 1. First, a camera is used to capture images in the field of view. Once these images are gathered, the data is passed to an external unit for processing. The processing device is responsible for two tasks. The first is deciphering the images received by the camera system. Once processed, the device computes the patterns needed to stimulate the implanted electrode array and cells in the retina. Upon completion, the processing unit will send the pattern through a wire to a receiver used to activate the implant. Inside of the eye is an implant, attached to the retina, which serves as the mechanism for stimulation of the cells. These signals for the implant could be sent through a wire passing through the skin behind the ear. Some are run externally to a pair of glasses used to stimulate the implant; others are run subdermally (beneath the skin) to the implant itself. The stimulated RGCs create neural impulses and follow the biological processes outlined in 2.1 sending the signals to the brain for image processing - see label 5 in Figure 1. When the user would like to deactivate the prosthetic, a switch can be toggled on the external device.

2.3 Types of Retinal Prostheses

There are a number of shared approaches when considering the development of a RP; the placement of the implant can lead to differences in the vision provided, including distortions in the image [4, 7]. This needs to be taken into account when simulating prosthetic vision for testing, which is discussed further in section 4.1.

Epiretinal prostheses place an electrode array implant on the inner retinal nerve fibers. Although most commonly used in the real world, this method does have its downsides, namely the distortions that may occur when targeting the cells for stimulation [12, 4]. This is due to the direct stimulation of RGCs which can result in the accidental stimulation of unwanted cells in the region. Despite potential distortions this method continues to see use in the world, most likely due to its ease of access for surgeons and reduction in risk of damaging the retina. Positioned further inside the eye, and closest to the damaged cells, a subretinal prosthetic places the implant behind the damaged retina [4]. This approach sends signals to the middle and outer parts of the retina, taking advantage of the network in the retina and potentially reducing distortions. Considered a novel approach by [4], suprachoroidal models place the implant further inside the eye than subretinal models. These approaches can simplify the surgical complexity reducing risk. The distance for signals to travel and evoke the RGCs is greater in this method, however, which may result in more distortions and require stronger electrical signals. Regardless, clinical trials regarding this method have been promising [4].

2.4 Producing Phosphenes

The replication of highly detailed vision is not yet achievable, instead what researchers hope to produce when providing therapy to the visually impaired are known as phosphenes. A phosphene is the presence of light in vision, despite no light entering the eye. Phosphenes are not uncommon in individuals and can be invoked in a number of ways. To experience this most individuals may use a mechanical method by exerting pressure on, or rubbing, ones closed eyes producing patterns of light. The activation of RGCs allows the appearance of these phosphenes to appear in patients with a RP; the presence of which allows these colorless shapes of light to be ”built” into the shape of an object (see Figure 4 for a simulated version). In the case of a RP, phosphenes are likely produced via stimulating the electrodes of the implant near the retina. However, non–invasive techniques like magnetic or chemical stimulations have begun to be explored, albeit less intensively than more common implanted methods.

2.5 Interpreting Regions of Interest

When examining an image there are often regions of interest (ROI) that warrant examination such as facial features, objects in or out of motion, and even subtle features such as depth. When processing an image, time becomes a dominant constraint when considering the capability of retinal prosthetics. At present it is difficult to recreate a captured scene with precision, much less in real time. As such, quickly identifying and relaying these points of interest serves as a prime focus for artificial intelligence models when determining what information to send to the implant. The rest of this subsection describes processes and themes used to quickly and effectively identify ROIs in an image.

2.5.1 Image Segmentation

Also known as pixel-level classification, image segmentation divides an image into groupings of pixels. A collection of pixels, more formally known as an image object, are grouped based on a similar property such as color, depth, or intensity. This process results in a series of image objects, producing grouped attributes of interest composing the image. One attribute provided in these segments is the outline of the image object gathered via edge detection, a process tasked with identifying the contours of objects in images based on differences from other pixels in the region.

When considering pursuits in artificial vision, edge detection and image segmentation can play an effective role in providing understanding of a patient's environment, including recognition of objects such as furniture or hazards like potential collisions. Utilized in the design of automated robots and self–driving cars, the data provided from processed image objects helps determine what information in a scene is necessary when considering the time constraints of retinal prosthetics.

2.5.2 Object & Facial Recognition

The presence of objects in an image itself warrants attention, but when considering the applications of computer vision systems the nature of the object can also be crucial. Some objects may provide relevant details about the scenery such as stationary objects like benches and buildings; others may be dynamic, people or vehicles for example which may be in motion. These details illuminate potential hazards as well as areas of interest in the image. Regardless of whether the result of this information is utilized by a machine or person, this ability to discern dangers or interests grants greater situational awareness. This increased situational awareness allows for the completion of tasks such as identifying an object before grabbing it or making note of a potential collision, two abilities that were previously absent or impossible for the visually impaired.

Facial structures can be an object composed of numerous points of interest capable of fine tuned movement. This level of detail can be a challenge to recreate, but there are promising developments such as methodology proposed in [13] for low vision systems like RPs.

2.5.3 *Scene Reconstruction*

A component of computer vision systems, scene reconstruction aims to recreate a scene including details like depth. This recovery of depth can serve a myriad of purposes such as aiding in object identification and detailed recreation of environments [6]; however, the recreation of an environment does not always need to be highly detailed. When considering applications limited on time a scene may be reconstructed in such a way that the edges of objects, produced via image segmentation, reproduce a passable representation of the scene[7, 11]. A continuing challenge that appears in artificial vision is the ability to determine which objects in a scene are most significant.

We can see how a scene can be composed in Figure 2, based on methodologies proposed in [11]. A camera is used to capture an image, in this case an indoor scene. The inputted image is then analyzed in two parts: the objects in the image and the layout. To define the layout of the image previous work suggests detecting what are known as structurally informative edges (SIE). These are the edges and intersections that are formed by walls, floor, and other structures that give a room orientation. In the Figure 2 we can see an SIE extraction of a floor and two walls meeting. The objects for the scene are segmented using a convolutional neural network, a type of artificial intelligence described in 3.2. The model classifies images using probability scores to refine the dimensions of the object and apply masks, or shapes of the object for highlighting, for each object. Once the masks have been applied the contours of the masked objects are highlighted and then replaced or stacked to create an image containing the objects. The result, known as object mask segmentation (OMS), is combined with the layout segmentation, SIE, to create a full scene. This result can then be converted into a format simulating prosthetic vision.

2.5.4 *Optical Flow Estimation*

Already constrained by time however, dynamic objects in an environment pose numerous challenges for artificial vision systems. Also known as scene flow estimation, optical flow estimation is used to approach the task of tracking objects in a scene. Image segmentation can help in this process, tracking the silhouette of the object, like the approach taken in [6].

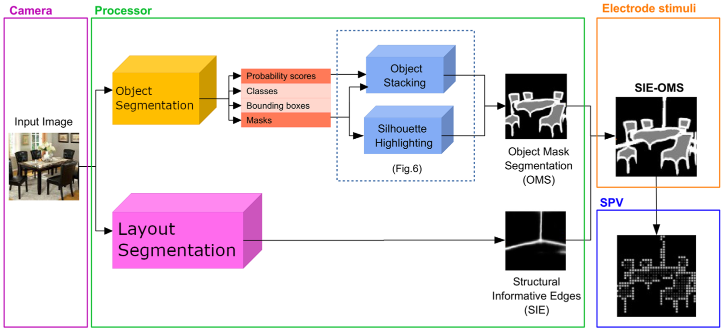

Figure 2: Scene reconstruction flow. Based on [11]

It is important to note that optical flow captures the appearance of motion and not the actual motion of an object. This difference can be explained with the barber pole illusion portrayed in Figure **??**. A barber pole is a cylindrical sign with helix shaped stripes angling towards the top of the sign. As the sign spins it appears the stripes are moving up the sign, this is the optical flow of the sign whereas the rotational movement of the sign (similar to a top) is the motion field. Although actual motion of illusions such as the barber pole will not be accurately represented by optical flow, the appearance of motion is often a fairly accurate portrayal and is useful for RPs.

One of the challenges associated with dynamic objects is the occurrence of occlusions, where one object may be obscured by another. The presence of such occlusions can make it difficult to separate objects in the scene resulting in poor image segmentation. Occluding one object with another can be expected when dealing with dynamic objects, especially in uncontrollable environments like outdoor scenes. To work with this challenge optical flow algorithms may be used, possibly to understand the depth of the scene and layering of objects. Understanding the depth of the environment allows the possibility to prioritize recognition of an object based on its layer in comparison to other objects; the less occluded the layer the greater chance there is for recognition [11].

3. NEURAL NETWORKS IN RETINAL PROSTHESES

Computer vision (CV) techniques are the focus of recent approaches in RP systems, utilizing various neural networks, usually convolutional neural networks (CNNs). Both technologies have been studied for decades with the first NN developed in 1950 known as the SNARC [9]. Later, in the sixties, the strive for computer vision started, a system meant to mimic the visual system of humans including three-dimensional structures. What has largely held back the advancement of these technologies since they were pioneered is the lack of computing power. However, continuing advancements have allowed the progression and application of these technologies.

3.1 Computer Vision

Utilized in applications such as self-driving cars, medical diagnoses, and manufacturing, computer vision processes information from media based data including images and video. Using this data, the computer is able to garner a better understanding of the objects or scene it is presented with as input. With a deeper understanding the computer is more capable of completing automated tasks or presenting findings to an operator. This process can be broken down into three steps 1) input of image based media, often gathered via a camera 2) processing of image data, utilizing a NN for pattern finding 3) the processor returns the findings to the requester for further actions. Often using NNs as the basis for the model, CV requires large amounts of data to extract features or patterns unlike the human visual system which passively collects information. This means applications of computer vision may excel at tasks like detecting defects in parts on an assembly line but are less effective at generalized tasks such as object identification in varying settings which is necessary for providing understanding to patients undergoing retinal prosthesis.

3.2 Neural Networks

A subset of machine learning, neural networks (NNs) consist of a series of nodes, each used to perform a calculation. These calculations are used to identify patterns and classify data, inspired by the structure of biological neural systems. A typical neural net is composed of an input layer, an output layer, and may contain zero or more hidden layers. A layer which gets its name due to the implication that the nodes and computation are private to the NN. The input layer is responsible for the initial intake of data for processing, which is then passed to the following layers via the edges or synapses that connect the nodes or neurons. The information is typically passed to the hidden layers which apply computational transformations. Each node in this case will contain the input, a weight, and bias. Weight being a parameter of NNs that reflects the importance of a feature when computing an output and bias being a constant representing the difference between the function's output and expected output. The computed value of the transformations is then passed to an activation function which the output and if the information should be passed to the next node. If the value is satisfactory the node activates, passing the information to the next neuron in the series. The inclusion of an activation function adds non-linearity to the model allowing the capability to solve non-trivial problems.

Artificial intelligence models like neural networks require large data sets and training to learn what features or patterns to analyze. To begin training the model the weights mentioned earlier are randomized and initially the output will likely not be very accurate. At this point we can begin feeding training data into the model and calculate the outputs in a process named forward propagation. The goal at this stage is to determine how well the model is able to predict the output. This is done with a loss function, also known as a cost function and is used to compute the error of the output versus the expected answer. At this point the goal shifts to minimizing the loss, which means better predictions from the model. One way to improve the model is to alter the earlier mentioned weights via a process named backpropagation. Abstracting away some of the mathematics, backpropagation sends the error information backwards updating the weights in the model in hopes of reducing loss. The optimization of these weights is what is referred to as training, or teaching the model. This is done using optimization techniques like gradient descent which are outside the scope of this paper. The result of these practices is the model learning to be more accurate. One vulnerability in the training process however is what is known as overfitting, or the model being too well fitted to the training data. Because the model is so well prepared due to its training it is inaccurate or unable to process new incoming data. Consider the scenario where a model is meant to recognize dogs and is provided with a series of pictures of dogs, all with the characteristic of spots and drooping ears. This could result in the model assuming *all* dogs need to have spots and droopy ears, limiting the flexibility of the model.

The subset of neural networks we will be concerned with are known as convolutional neural networks (CNN). These fall under the umbrella of deep learning, which are generally described as a neural network with a large amount of hidden layers, IBM says at least three [2]. In addition to more hidden layers convolutional neural networks also have different structures. These neural networks are first composed of a convolutional layer, although there can be more than one of this type of layer. This is also where the greater part of computation is completed and serves as the building blocks for the CNN. There are two parts needed here, the data, and a filter. In the case of an image input we could consider data such as height, width, color, and depth in a color image. With this data we can apply a feature detector sometimes called a kernel or filter, a small matrix of weights applied to the entire image. It will move across the image to checking if basic pieces of the feature are present, perhaps a spot of light or a particular . This is known as a convolution. The feature detector looks at a patch of the image at a sliding its way through the image until having examined the totality. For each segment what is known as a dot product is produced and is a calculation of the input pixels and the filter. Once the feature detector has made its way through the image what is outputted is a collection of dot products, known as a feature map. In the end the convolutional layer will convert the image into numerical values so the model can extract patterns.

Following the convolutional layer is the pooling layer which is also referred to as downsampling. Here, the model reduces parameters in the input. In some ways this layer is similar to the convolutional layer with the largest difference being that when the pooling process sweeps over the image the filter does not have any weights. Instead, the filter applies an aggregate function, or calculation, to the values which then are then used to populate the output array. There are two commonly used types of pooling, the more popular max pooling and then average pooling. Max pooling will select the most prominent feature from the feature map, whereas average pooling calculates an average of the present features. Unlike in the convolutional layer which looks at the entire image, the pooling layer removes unnecessary noise from the image. Although this action is responsible for the majority of lost data it does benefit a CNN by improving efficiency, reducing complexity, and reducing risk of overfitting the model.

The final layer of these neural networks are the fully-connected layer. Here the classification of the extracted features happens based on the what was gathered from previous layers. Typically the fully-connected layer uses an activation function such as softmax, a function that normalizes the outputs by assigning probabilities to the sum of the weighted values to classify inputs. A generic pipeline for a CNN can be seen in Figure 3 using a stuffed animal as an example. In the case of classification models, like those used in image recognition, the outputs may be a list of probabilities displaying the models best estimations.

4. METHODS

In this section different options for testing prosthetic vision is discussed and how they can be used to give researchers a platform for experimenting with new approaches. This is followed by applications of artificial intelligence and how they have been used to improve artificial vision.

4.1 Testing

When it comes to testing the effectiveness of generating visuals for the visually impaired there are a few methods. The first approach revolves around working with patients who have been implanted with a RP. Individuals implanted with a RP are rare however, making testing those who are currently undergoing treatment difficult for any groups not currently developing a RP.

For interested research parties who are unable to find implanted participants, a simulation tool, such as pulse–2–percept [5], can be used to represent the retina. Using such a simulation platform allows researchers to view the output of the shape of the phosphenes, in the case of pulse–2–percept, rather than a pixelated image like some simulated prosthetic vision (SPV).

An example of SPV can be seen in [11] which uses the SIE-OMS processing discussed in 2.5.3 depicting a series of indoor scenes, a portion of which are displayed in Figure 4. Examining the simulated outcomes of the images, we see the outlines of the objects are not smooth and are missing pieces of light or phosphenes. This is to mimic the dropout or degeneration of light sensitive cells patients undergoing prosthesis may experience depending on the stage of the disease. Dropout rates are at the discretion of the researching party, although dropout rates tend to range between 10% and 30% [8, 10].

As an intermediary measure between simulations and implants, some researchers opt to test retinal prosthetic devices on participants who still possess their vision such as work in [7] from the University of California. Although this approach expands the potential pool of participants for study, it also requires the need to set parameters. Factors such as the distance of RGC bodies and distance from the stimulation site can differ between patients [7]. As a result, researchers are required to operate in broad spectrums to more accurately account for varying conditions.

4.2 Improving Artificial Vision With AI

Using NNs with CV techniques researchers have improved RP's by refining facial features, improving environment representation, and reducing collisions with hazard detection. Here a discussion of recent approaches and results will be explored.

Facial Recognition

Faces are a frequent and detailed part of everyday life and it can be useful to recognize those who are a common presence in our lives. Work done in [13] shows how the magnification of facial features like the nose and inclusion of external features like hair can help with facial recognition. To accomplish this, Jing Wang et al. presented SPV images of celebrities and public officials in China to twelve patients who retained their sight. Patients would then be required to identify the subject as hastily as possible or respond in the negative. The images were processed using four different techniques to magnify regions of interest, all utilizing CV feature extraction to identify the face in the image and magnify its features. Three of the techniques proposed use magnification of features to increase recognizability in comparison to the previously used style of directly lowering the resolution. This was shown to improve recognition accuracy in all three magnification techniques. An increase in computational power has allowed the use of these techniques, however some are still too computationally expensive for use in a RP system at present. Research in [13] demonstrates a variation of the Viola Jones Facial Recognition technique using statistical face recognition proved to increase identification accuracy in participants while usable in a RP system. Both the Viola Jones approach and face matting recognition (the third magnification technique explored) offered greater detail, but are too computationally expensive. Despite requiring stronger computational power in the external processing unit, their work has displayed how CV can be used to improve magnification of features and increase the ability to distinguish between similar faces for those with a RP.

Reconstructing Scenes

Introduced in section 2.5.3, advancements have been made in how scenes can be reconstructed using structurally informative edges as layout features. Using a CNN to classify objects and highlight their contours, methodologies used in [11] demonstrate how segmentation can be used to effectively reconstruct an indoor scene. A pool of eighteen subjects between the ages of 20 and 57 with normal vision were shown indoor scenes portrayed as SPV images on a computer screen for 10 seconds. The participant would then verbally communicate their identification of the room, resulting in a not answered or NA response if the 10 second limit was surpassed. The images were processed using direct and edge methods, two techniques previously used in retinal prostheses, as well as the SIE-OMS method proposed using structurally informative edges and object masking. These trials showed that the proposed method significantly outperforms earlier approaches with participants correctly identifying objects 62.78% of the time compared to direct or edge methods which measured 36.83% and 19.17% respectively. Additionally, a noticeable improvement in the ability to recognize the room presented was shown when using SIE-OMS and resulted in fewer NA responses from participants.

Natural scenes taking place outdoors pose their own challenges and the greater goal of providing understanding of a scene or environment remains an issue. Research proposed by Han et al. explores how utilizing deep learning to simplify scenes using segmentation can provide greater understanding of a scene when compared to previous methods relying on saliency or depth. A pool of 45 students with normal sight acted as participants in the study, which was performed remotely. The students were asked to watch videos which had been converted to an SPV format and determine whether cars and people were present in the scene as well as provide a level of confidence in their answer. The images were processed using four strategies including saliency, depth, object segmentation, and a combination of the three using CV algorithms. A comparison of accuracy and precision between the techniques is displayed in Table 1. In this study accuracy measures the number of correct predictions and precision measures the number of correct predictions divided

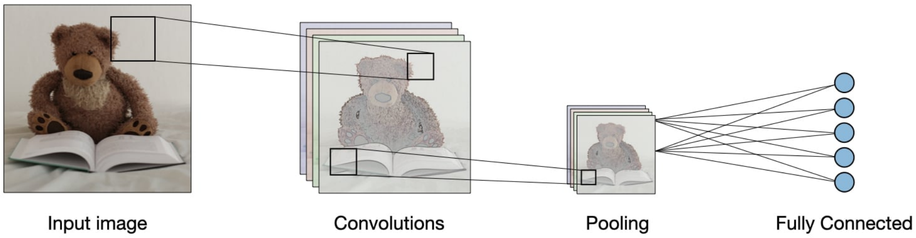

Figure 3: Convolutional neural network pipeline [3]

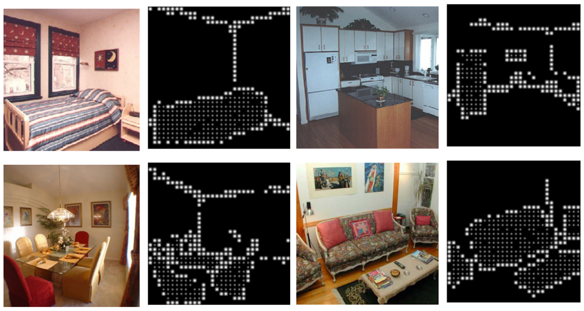

Figure 4: Simulated prosthetic vision of indoor scenes [11]

Condition	Accuracy	Precision
Saliency	0.51	0.53
Depth	0.54	0.56
Segmentation	0.68	0.73
Combination	0.66	0.72

Table 1: Accuracy and precision of vehicle and person detection by participants [7]

by the number of trials with a person or vehicle present. Results showed segmentation techniques outperformed previous methodologies when it came to identifying people and vehicles in outdoor settings. The combination strategy was used to determine the effectiveness of combining the information from the saliency, depth, and segmentation models; this proved not to have any effect however, and was only slightly worse than the segmentation strategy.

5. PRACTICAL CONSIDERATIONS

Developing a RP comes with the inherent issue of meeting individual needs in a changing world. Experiences from one visually impaired person to the next can differ by medical status, environments, and tasks to be performed. This section delves into a few of these diversities and how to address them.

5.1 Variations in Environment

Environments needing to be reproduced for the visually impaired can vary in numerous ways, the separation between indoor and outdoor scenes perhaps the most distinguishable. Such environments may differ by number of dynamic objects (if there are objects capable of movement at all), contrast created by light sources and shadows, or their layout [7]. An outdoor scene may include pedestrians and vehicles in motion, whereas one's office may be composed of static objects such as a desk and office supplies. Certain workplaces may also be predictable spaces with repeating tasks, further simplifying vision tasks.

5.2 Patient Differences

When considering the use of a retinal prosthetic, biological differences in the patient can lead to different results and effectiveness. Although there are varying diseases and parts of the eye which can lead to loss of vision, only two diseases are eligible for RP treatment. The diseases currently eligible for retinal prostheses are retinitis pigmentosa and age related macular degeneration, retinal degenerative diseases which cause the breakdown of the photoreceptive layer [4, 11, 7].

The nature of these diseases means that patients undergoing treatment are in different stages of degeneration and there is no universal solution when considering therapy. Differences in placement of the implant, amount of RGCs remaining, and stimulation method may result in differences in the shape as well as presence of phosphenes.

6. CONCLUSIONS

A survey of processes used to improve the vision of retinal prosthetics using AI has been presented, displaying the growing capability of a modern prosthetic for therapeutic treatment. Advancements in artificial intelligence and hardware have driven modern techniques which allow the application of neural networks and computer vision to operate within the time constraints required of a RP. Additionally, the differences in patients undergoing treatment is explored, outlining challenges when developing a device for implantation. Despite advancements in segments of technology previously inhibiting the growth of RP's, many of the challenges still remain such as segmentation in moving images. Some techniques which could prove successful are still not fast enough to be practical (such as the facial recognition study described in 4.2). Progress in the fields of processors and artificial intelligence have provided great strides towards a functional prosthetic to produce vision in the last decade.

Acknowledgments

I would like to thank Professor Elena Machkasova and Professor Kristin Lamberty for their guidance and insights in this work.

7. REFERENCES

[1] Bionic vision australia | prototype bionic eye. https://tinyurl.com/42n65fau. Accessed: 10-23-2021.

[2] What is deep learning. https://tinyurl.com/yckurtrz. Accessed: 11-17-2021.

[3] Amidi, A., and Amidi, S. Convolutional neural networks cheatsheet. https://stanford.edu/ shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks.

[4] Ayton, L. N., Barnes, N., Dagnelie, G., Fujikado, T., Goetz, G., Hornig, R., Jones, B. W., Muqit, M., Rathbun, D. L., Stingl, K., Weiland, J. D., and Petoe, M. A. An update on retinal prostheses. *International Federation of Clinical NeurophysiologY 131*, 6 (2020), 1383–1398.

[5] Beyeler, M., Boynton, G., Fine, I., and Rokem, A. pulse2percept: A python-based simulation framework for bionic vision.

[6] Guo, F., Yang, Y., Xiao, Y., Gao, Y., and Yu, N. Recognition of moving object in high dynamic scene for visual prosthesis. *IEICE TRANSACTIONS on Information and Systems E102-D* (2019), 1321–1331.

[7] Han, N., Srivastava, S., Xu, A., Klein, D., and Beyeler, M. Deep learning–based scene simplification for bionic vision. In *Augmented Humans Conference 2021* (New York, NY, USA, 2021), AHs'21, Association for Computing Machinery, p. 45–54.

[8] McKone, E., Robbins, R. A., He, X., and Barnes, N. Caricaturing faces to improve identity recognition in low vision simulations: How effective is current-generation automatic assignment of landmark points? *PloS one 13*, 10 (2018), e0204361.

[9] Russell, S., and Norvig, P. *Artificial Intelligence A Modern Approach*, fourth ed.

[10] Sanchez-Garcia, M., Martinez-Cantin, R., and Guerrero, J. Structural and object detection for phosphene images, 09 2018.

[11] Sanchez-Garcia, M., Martinez-Cantin, R., and Guerrero, J. J. Semantic and structural image segmentation for prosthetic vision. *PLoS ONE 15* (2020).

[12] Shim, S., Eom, K., Jeong, J., and Kim, S. J. Retinal prosthetic approaches to enhance visual perception for blind patients. *Micromachines 11*, 5 (2020).

[13] Wang, J., Wu, X., Lu, Y., Wu, H., Kan, H., and Xinyu, C. Face recognition in simulated prosthetic vision: Face detection-based image processing strategies. *Journal of neural engineering 11* (06 2014), 046009.