Argus III: A Novel Image Optimization and Augmentation Framework to Enable an Improved Patient Experience for the Next Generation Epiretinal Prosthesis

William Huang

Abstract

Bionic eyes, commonly referred to as retinal prostheses, provide a promising solution to restore vision to the 200 million people who suffer from retinal degenerative diseases worldwide. However, current FDA-approved implants are limited to low-resolution grayscale images, making object identification and localization difficult for patients. Furthermore, patients will face a steep learning curve to adapt to the prostheses.

To address these challenges, this study develops a novel methodology framework that consists of three integral components: 1) an optimal transportation theory (OT)-based virtual magnifier to localize and enlarge regions of interest (ROIs) while preserving important features and curvatures; 2) a real-time image optimization framework to encode the maximum amount of spatial and color information to patients through attention mechanisms as well as color scheme comparisons; and 3) an autoencoder-OT model to augment the optimized images.

Computational experiments through distortion maps showed that the magnifier enlarged the ROIs with minimal area and angle distortion. Further, users were able to select important features and optimize ROI densities according to their preference through a "digital knob" user interface. In contrast to current schemes, the image optimization framework demonstrated better visual quality, was computationally efficient (less than 380 ms on tested cases), and allowed for optimal color mapping through comparison studies. A prototype processing system confirmed the effectiveness of the proposed optimization framework over current prostheses. Finally, the AE-OT model augmented images from 6 datasets to generate an image library for patient training.

This research offers an accurate, scalable, and optimized architecture that will enable the next generation epiretinal prosthesis.

Keywords: Retinal prostheses, virtual processing unit, virtual magnifier, image optimization, image augmentation, manifold learning, conformal geometry, optimal transportation theory, region-contrast saliency maps, color quantization

1 Introduction

Millions of people worldwide suffer from photoreceptor loss due to retinal degenerative diseases such as retinitis pigmentosa (RP) and age-related macular degeneration (AMD). Together, these diseases account for the leading cause of vision loss. Wet AMD causes leakages in the macula of the retina, severely impairing a patient's ability to see with sharp vision (Kozarsky, Kozarsky). Further, RP destroys the rod cells in the photoreceptors of the eye, limiting the ability to see in dimly lit environments (Mills, Mills). Moreover, the brain initially compensates for the dark spots in the patient's vision, hindering early detection of the diseases until it has reached its late stages. Currently, there are no effective treatments and cures to reverse the effects of RP or AMD. Therefore, doctors can only prescribe medication to reduce the swelling or recommend the use of sunglasses (Kozarsky, Kozarsky).

Creating a bionic eye, or retinal prosthesis in scientific terms has been proven to be one of the most promising solutions to restore vision to patients suffering from AMD or RP. People with AMD and RP retain 90% of their inner retinal neurons, allowing for the possibility to elicit visual percepts via electrical stimulation (Horsager et al., 2009). Further, recent studies have shown positive results in potentially restoring vision through electrical stimulation. During the electrical stimulation of tissue via external electrodes, the resulting impulse induces depolarization, causing the firing of action potentials in the neurons. These signals are then carried through the optic nerve to the visual cortex of the brain, which converts the signals into images for the blind to see.

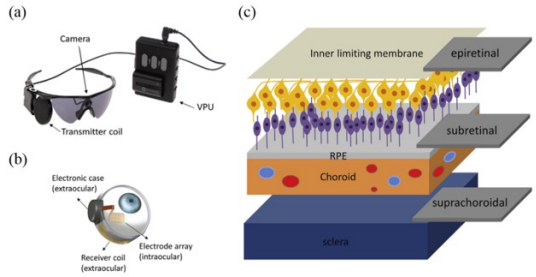

Figure 1: *(a) External components (b) internal components of the Argus II system; (c) illustration of the implantation sites of the visual cortex, epiretinal, subretinal, and supra-choroidal prostheses. (Yue et al., 2016).*

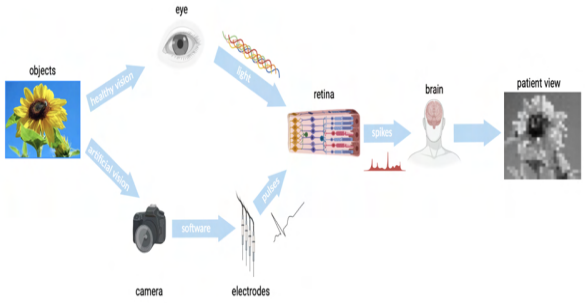

Figure 2: *Current bionic eye strategy to restore vision.*

The stimulation of the retina has been proven most successful out of all regions of the visual pathway (Brindley and Lewin, 1968; Normann et al., 2009; Sakaguchi et al., 2009). There are currently two main methods to electrically stimulate the retina: the epiretinal approach and the subretinal approach. The epiretinal approach is a microelectrode array

placed directly on the nerve fiber layer, while the subretinal approach is in direct contact with the bipolar cells. Both methods have shown positive phosphene activation (Yue et al., 2016). Concretely, several groups have created retinal prostheses capable of generating electrical stimuli to the retina (eg. Argus II and alpha-IMS) (Figure 1). A typical epiretinal prosthesis contains three main external components: a video-camera, virtual processing unit (VPU), and an external coil for transmission (Luo and da Cruz, 2016). These components work together to produce fast, real-time imaging of a patient's surroundings. An image is first captured through a camera, which is then converted into electrical parameters conveying spatial-temporal information. This data is then sent to the chip through RF telemetry and then finally stimulates the retinal neurons through a microelectrode array. Figure 2 [1] shows the existing strategy for vision restoration.

Although these retinal prostheses can successfully stimulate a patient's retina, they have a major limiting factor. Because of the compact environment in which these prostheses need to be surgically placed in, the size constraint is severely limited. In addition, the heat released from the retinal prosthesis can damage the retina. Other factors include power consumption, limitations on micro fabrications, and incision size on the eyeball. The current image sizes of retinal prostheses range from 60 to over 1500 electrode arrays. Each electrode ranges from 260 um to 520 um in diameter (Horsager et al., 2009). This is significantly smaller than the size of pictures people see, which reach up to around 65,000 pixels. Because of the limited resolution, patients will only be able to see a blurry, pixelated image of their surroundings. Further, recent clinical trials have shown that patients with the Argus II implant had a visual acuity of 20/1262, which is far below the limit for legal blindness. (20/200). Although recipients with such visual acuity had a significant improvement of activities of daily living and mobility by assessing a variety of daily visual tasks, such as letter and large geometric shapes recognition, word reading, object localization, and outdoor movement detection, it's extremely difficult for them to perform more sophisticated visual tasks such as object and face recognition. Therefore, any image taken by the camera will have to be processed and downsampled by the VPU of the retinal prosthetic before being converted into electrical currents to ensure that the maximum spatial-temporal information is viewable by the patient.

There are three ways that images can be downsampled. They are to scale down the whole image, zoom in on a region of interest (ROI), or have high resolution in the ROI with a blurry peripheral vision (full field, ROI, fisheye). van Rheede et al. (2010) compared these three paradigms when coupled with visual acuity, block, and wayfinding tasks and monitored subjects' responses to each. A visual acuity task focuses on analyzing a patient's

[1] Unless otherwise noted, all images, data tables, and graphs were created by the author.

ability to resolve visual details using the Snellen chart, which is the chart used in many ophthalmologist clinics. The block task measures the patient's ability to match blocks of different shapes and colors to their respective outlines. Finally, the wayfinding task analyzes the ability to perform common real-life tasks, such as walking under bridges or moving through a tunnel. Overall, the full field and ROI models performed significantly better than the fisheye one (van Rheede et al., 2010). Another primary image processing component being applied to retinal prosthetics is edge detection. Edge detection algorithms focus on detecting areas of an image in which the image brightness drastically changes. In other words, the algorithm will define boundaries around objects in images, which is applicable to helping blind people see. Sharmili et al. (2011) discovered a simple and sensitive edge detection algorithm capable of producing superior results to Sobel and Canny's detection algorithms.

A more recent research direction tries to take advantage of visual attention mechanisms by investigating various attention models. Examples include pixelization models (Li et al., 2005), importance maps and ROI processing (Boyle, 2008), and salient segmentation methods based on the Itti saliency model (Wang et al., 2016). Li et al. (2005) proposed a pixelization model that assigned higher resolution where prominent features including contrast, edge, orientation, and symmetry occured. Boyle (2008) applied ROI processing to binary images, and their results indicated that the method used in zoom processing improved scene recognition performance. Recently, Wang et al. (2016) proposed image processing strategies using a saliency segmentation method based on the Itti saliency model. Their results indicated that using the saliency model proved advantageous to object detection and that the proposed strategies improved recognition accuracy of objects. The above studies have suggested that saliency models are a promising direction to assist prosthesis recipients in locating objects and improving their visual perception.

However, little research has been conducted to improve the resolution of salient objects in images while retaining their hue and critical details. Besides the limited spatial information, current retinal prosthetic devices only encode illuminance, meaning the images lack key color information. Medical research recently found that changes in stimulation parameters could alter the color perception in patients, leading to the possibility of encoding color information for retinal prostheses (Yue et al., 2016).

Due to the aforementioned physical and computational limitations, patients will continue to be limited to low-resolution images and therefore face challenges of recognizing objects in images, particularly at the early stages of prosthesis implementation. Therefore, it is necessary to develop a training program to help patients adapt to these scenes. A critical component of this program would need an image library generated by VPU systems. Such

image library is currently not available for prosthesis recipients.

Generative adversarial networks (GANs) and variational autoencoders (VAEs) are one of the most effective approaches for unconditional image augmentation (Lei et al., 2020). GANs train an unconditional generator that regresses real images from random noises and a discriminator that measures the difference between produced samples and input images. However, GANs have two major drawbacks. 1) GANs are sensitive to hyperparameters. 2) GANs suffer from mode collapse, a phenomenon where the generator can only produce one type of a small set of outputs. More generally, the fundamental principles of deep learning remains primitive. A recent study of nine different GAN models and variants determined that gradient-descent-based GAN optimization is not always locally convergent (Mescheder et al., 2018). Variational autoencoders use an encoder to map data from a data distribution to a Gaussian latent distribution, which is then mapped back to the data distribution via a decoder. However, VAEs have a drawback as well, as they generate ambiguous images on multimodal data distributions.

The goal of this study is to develop a solution to encode the maximum amount of spatial and color information to patients suffering from retinal degenerative diseases, through attention mechanisms as well as color scheme comparisons. If successful, patients will be able to see higher quality images even with a limited number of electrodes. In addition, this paper proposes developing a simple, general, and scalable augmentation model capable of creating an image library contributing to a critical training program for prosthesis users. Together, these will significantly improve the quality of life of patients suffering from RP or AMD.

Following the introduction, Section 2 of this paper will present the methodology. The methodology consists of: 1) a novel virtual magnifier capable of preserving important features and details; 2) an improved framework for image content optimization; and 3) an autoencoder-optimal transportation model for image augmentation. Section 3 will present the results and findings. Section 4 presents the discussion and conclusion.

2 Materials and Methods

2.1 Proposed Image Optimization and Augmentation Framework

The goal of the proposed image optimization and augmentation framework aims to accomplish three tasks: 1) localize and "magnify" key segment(s) in an image frame while preserving important features and curvatures of objects in the segment(s); 2) optimize the magnified segments to encode the maximum amount of spatial and color information to the patients through attention mechanisms as well as color scheme comparisons; and 3) augment the optimized images to enable new patients to quickly adapt to the prosthesis.

To achieve these objectives, there are several challenges. First, to highlight a specific

segment in an image frame, it requires altering the feature density of the original image, involving a complex mapping between two density distributions. Second, displaying the key segment that meets the resolution requirement of prostheses requires a critical trade-off of spatial and color information. Third, augmenting images commonly faces an issue of mode collapse where the image is not from the same distribution as the original image. To address

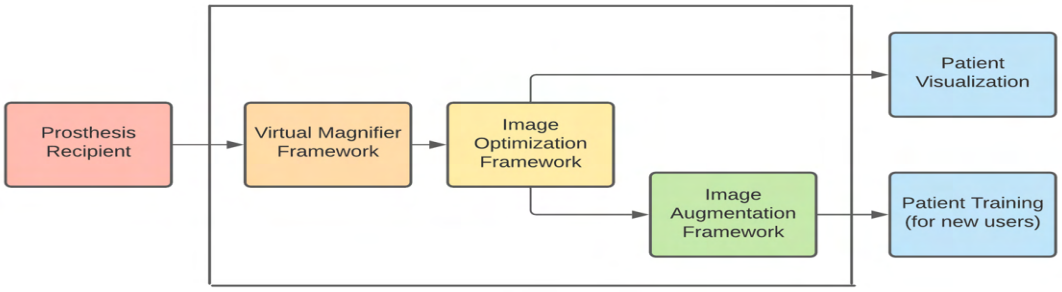

Figure 3: *Overall methodological framework for image optimization and augmentation through a) Virtual Magnifier. b) Optimization Framework. c) Augmentation Framework.*

these challenges, this paper proposes an image optimization and augmentation framework (Figure 3). Specifically, the methodology framework consists of a (1) virtual magnifier, (2) image optimization model, and (3) image augmentation model. The virtual magnifier increases the image density of key segments and maps the original image to a new one through the optimal transportation theory (OT). To increase computational efficiency, a Fast Fourier Transform (FFT) based OT algorithm was applied. The image optimization model localizes main objects in specified segments while reducing unnecessary background noise through region-contrast based saliency maps. The environment of salient objects was framed by borders identified by an edge detection operator. A novel color depth mapping technique was developed through MiniBatchKmeans clustering and color space selection. The resulting image was down-sampled using bicubic interpolation to reduce image size while preserving color quality. Finally, to enable unconditional image generation, an autoencoder-OT model was adopted to create an image library for patients.

2.2 Smart Image Segmentation through Virtual Magnifier to Preserve Critical Features Before Resolution Reduction

Due to its limited resolution, the VPU of a retinal prosthesis has to choose a ROI from a captured image by a camera. A common strategy is to maintain the image density while carving out ROIs or segments. Zhao et al. (2012) developed a conformal magnifier to magnify a ROI using conformal mapping. It allows enlarging features of a ROI while deforming remaining areas without any cropping. By using conformal mapping, the ROI is magnified with minimal distortion, while the transition region is a smooth and continuous deformation

between the focus and context regions. A recent improvement of the conformal magnifier is to quantitatively specify the density of ROI and map the image with an initial density to the desired density. With this development, it is possible to develop a smart image segmentation approach to optimize the density of magnification and best preserve the critical features of objects in the segments or ROIs before resolution reduction.

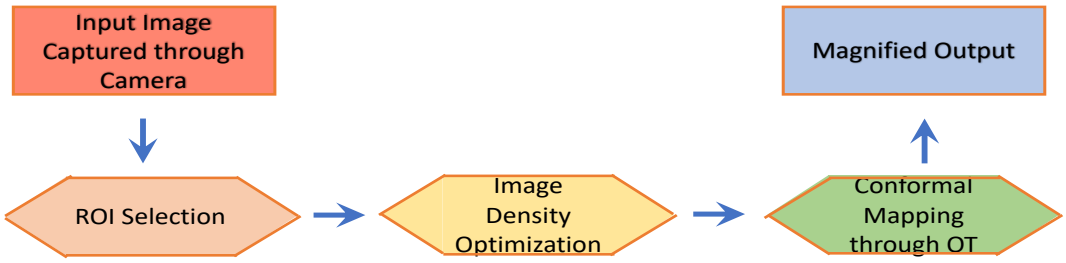

Figure 4: *The pipeline of the smart image segmentation through the virtual magnifier.*

The proposed framework of smart image segmentation through a virtual magnifier is given in Figure 4. Once the ROI is localized or focused by the camera, the density of the ROI will be determined and optimized before the ROI is to be magnified. Then, the conformal mapping will be obtained through OT theory to map the original image density to the optimized density to enlarge the ROI. I will briefly introduce conformal mapping and the OT theory.

2.2.1 Theoretical background of conformal mapping

Conformal geometry is the intersection of complex analysis, algebraic topology, Riemann surface theory, algebraic curves, differential geometry, and partial differential equation. Recently, computational conformal geometry has been developed to obtain conformal mapping between surfaces. A conformal map is a function that locally preserves both angles and the shapes of infinitesimally small figures, but not necessarily their size or curvature. I first need to introduce the definition of manifold.

Definition 1 (Manifold)**.** *A manifold is a topological space M covered by a set of open sets $\{U_\alpha\}$. A homeomorphism $\phi_\alpha : U_\alpha \to \mathbb{R}^n$ maps U_α to the Euclidean space $\mathbb{R}^n$. (U_α, ϕ_α) is called a coordinate chart of M. The set of all charts $\{(U_\alpha, \phi_\alpha)\}$ form the atlas of M. Suppose $U_\alpha \cap U_\beta \neq \emptyset$, then*

$$\phi_{\alpha\beta} = \phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha (U_\alpha \cap U_\beta) \to \phi_\beta (U_\alpha \cap U_\beta)$$

is a transition map.

If all transition maps $\phi_{\alpha\beta} \in C^\infty (\mathbb{R}^n)$ are smooth, then the manifold is a differential manifold or a smooth manifold.

In general, manifold is a topological space that locally resembles Euclidean space near each point, meaning that every point has a neighbourhood homeomorphic to an open subset

of Euclidean space. Globally it may be not homeomorphic to a Euclidean space. With the definition of manifold, the canonical representation of conformal map can be introduced through the Poincare uniformization theorem:

Theorem 1 (Poincare Uniformization Theorem)**.** *Let* $(\Sigma, \mathbf{g})$ *be a compact 2 -dimensional Riemannian manifold. Then there is a metric* $\tilde{\mathbf{g}} = e^{2\lambda}\mathbf{g}$ *conformal to* $\mathbf{g}$ *which has constant Gauss curvature.*

Mathematically, suppose $f(z) : \mathbb{C} \to \mathbb{C}$ is a complex valued function on the complex plane, its real representation is $f(x + iy) = u(x, y) + iv(x, y), z = x + iy$. The complex differential operators are defined as

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) \tag{1}$$

$f(z)$ is conformal if and only if $\frac{\partial f}{\partial \bar{z}} = 0$. Different methods have been developed to compute this conformal maps. For example, harmonic map minimizes the elastic deformation energy. But it only applies to genus zero surfaces. The holomorphic differential map finds harmonic vector fields on surfaces and it can be applied to genus one surfaces. A more recent and advanced method is to use Ricci flow, which was invented by Hamilton and used by Perelman to prove Poincare conjecture. It can find conformal maps for high genus surfaces.

Another advantage of applying conformal mapping is that it enables global parameterization of general surfaces without partitioning to minimize distortions. Note that conventional methods use local parameterization and one often has to decompose the surface into patches.

2.2.2 Theoretical background of optimal mass transportation

Optimal mass transportation theory (OT) has become increasingly popular in deep learning. I briefly introduce the theoretical background of OT (Gu et al., 2016).

Given two distributions with measure (Ω, μ) and (Ω^*, ν) with densities $d\mu(x) = f(x)dx$ and $d\nu(y) = g(y)dy$, such that $\mu(\Omega) = \mu(\Omega^*)$. A mapping $T : \Omega \to \Omega^*$ is measure-preserving, if for any Borel set $B \subset \Omega^*$, it satisfies

$$\int_{T^{-1}(B)} d\mu = \int_B d\nu \tag{2}$$

A measure-preserving map is denoted as $T_{\#}\mu = \nu$. The cost function $c : \Omega \times \Omega^* \to \mathbb{R}$ defines the cost for transporting a unit mass from distribution $f(x)$ to distribution $g(y)$, the total transportation cost for a transportation map T is given by

$$\int_\Omega c(x, T(x))d\mu(x) \tag{3}$$

The *optimal transport map* is the measure-preserving map T that minimizes the total transport cost,

$$\min_{T_{\#}\mu=\nu} \int_{\Omega} c(x, T(x)) d\mu(x) \tag{4}$$

Brenier's theorem shows that for a quadratic cost function $c(x, y) = 1/2|x - y|^2$, the optimal transportation map is the gradient map of a convex function $u : \Omega \to \mathbb{R}$, that is, the Brenier potential. The Brenier potential satisfies the Monge-Ampere equation:

$$det(D^2)(x) = \frac{f(x)}{g \circ \bigtriangledown u(x)} \tag{5}$$

In 2D cases (e.g., images), the Monge-Ampere equation can be written in term of Poisson equation:

$$u_{xx}u_{yy} - u_{xy}^2 = \frac{f}{g \circ Du}, \tag{6}$$

or equivalently,

$$\triangle u = \sqrt{u_{xx}^2 + u_{yy}^2 + 2u_{xx}^2 + 2\frac{f}{g \circ Du}} \tag{7}$$

By sampling Ω by regular grid points with horizontal and vertical steps, define finite difference operators can be defined to solve discrete Poisson equation and find the optimal transportation map in Eq. 5.

2.2.3 Smart image segmentation through virtual magnifier

The Poisson equation 7 provides a fast solution to find the optimal mass transportation from one distribution $f(x)$ to another $g(y)$. This lead us to develop a smart image segmentation based on optimal mass transportation mapping. Let $f(x)$ denote the original image distribution and $g(y)$ denote the distribution of the ROI with enlarged density for critical features or objects. The density $g(y)$ can be optimized so as to facilitate the patient to recognize objects in the low-resolution images from the VPU of prosthesis. The density parameter can be viewed as the times of magnification for glasses. The proposed pipeline in Fig. 4 will optimize the density parameter for the VPU.

2.3 Image Optimization Framework to Encode Color and Spatial Information into Retinal Prostheses

This section develops a solution to encode the maximum amount of spatial and color information to prosthesis recipients.

To overcome the challenge of the limited information that prostheses can pass to patients, this study proposes a real-time image processing pipeline that can be run on any image and

Algorithm 1 Image segmentation through optimal mass transport map.

Require: ROI $\mathcal{A}$, source density $f(i,j)$ and target density $g(i,j)$ of $\mathcal{A}$, feature threshold ϵ
Ensure: The optimal mass transport map $T : (\Omega, f) \to (\Omega^*, g)$
  $g(i,j) \leftarrow$ initial density
  $\epsilon \leftarrow \epsilon*$
  while $\epsilon < \epsilon*$ **do**
    Find T by solving Eq.7: $\triangle u = \sqrt{u_{xx}^2 + u_{yy}^2 + 2u_{xx}^2 + 2\frac{f}{g \circ Du}}$
    Magnify $\mathcal{A}$ through $T : (\Omega, f) \to (\Omega^*, g)$
    Update ϵ
    Increase density $g(i,j)$
  end while

is capable of identifying the key color and spatial information in the image. Further, the model will be downsampled, divided, and segmented to resemble the constraints of the retinal prostheses. The input image of the algorithm was first passed through a salient object detection to assist patients in identifying the main objects (Figure 5). Subsequently, the input image was also converted to a grayscale image to generate an edge map for contrast. The resulting images were then weighted and overlaid to form the combined image (E. in the figure). Finally, the combined image was quantized with its color features and then downsampled to fit the size of the images passed through retinal prostheses.

2.3.1 Generating Grayscale map and Edge Extraction

Since the resolution of the images generated by the algorithm must be significantly lower than the input stream, identifying the edges of objects in a pixelated image will greatly help patients see the boundaries of objects in reduced images. To accomplish this, I first generate grayscale maps of the color images. Let R,G, and B represent red, green, and blue respectively. the intensity I of any pixel is replaced as I = 0:299R + 0:578G + 0:144B. Then, I apply edge detection to the grayscale images. There are 4 main operators for edge detection: Sobel, Canny, Laplacian, and Prewitt. The Sobel operator has the ability to smooth and provide an edge response simultaneously by convolving with a small, integer filter. The Canny method is the most widely used, and most complicated. Finally, the Laplacian only uses one kernel, meaning it calculates the derivatives in one pass, and the Prewitt model is used for detecting vertical and horizontal edges (Tsankashvili, Tsankashvili). The edge detection method adopted in this study is the Canny operator due to its low error rate, ability to localize edge points, and lack of false edges. To reduce noise in the image, a Gaussian filter is adopted. A Gaussian kernel filter with size (2k+1) x (2k+1) is defined as

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp -\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma}; \tag{8}$$

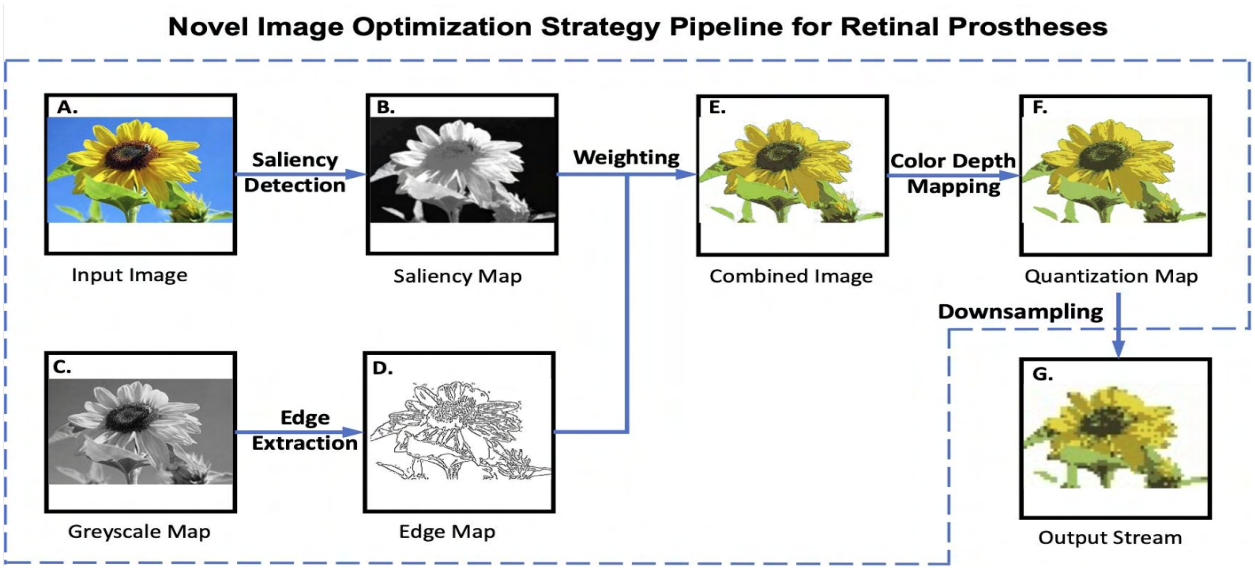

Figure 5: *Proposed image processing strategy based on saliency detection, edge detection, color depth mapping, and downsampling. A-B For each image, a saliency map (B) is built by my global saliency detection algorithm); Edge extraction: C-D where C is a gray map, which is used to extract the edge map (D) by Canny operator); Weighting: A, B and D - E (the output stream (E) is obtained by the original image (A) saliency maps (B) and edge maps (D) of the input stream). E-F (the image (F) is generated through color depth mapping and K-means clustering).*

where $1 \leq i, j \leq (2k+1)$.

Because the edges in an image may be directed in any direction, additional filters are needed to detect the horizontal, vertical, and diagonal edges. If θ denotes the angle of the edge, it is equal to $\tan^{-1} \frac{G_y}{G_x}$

Subsequently, it is now necessary to thin the predicted edges by using a non-maximum suppression technique. This is accomplished by passing a 3x3 filter over the output of the previous two steps with a center value of 0. Finally, to account for remaining edges caused by the variations in noise and color, threshold values are needed to remove "non-edges." For example, any pixels with an intensity gradient above the maximum value is kept, while any pixel below the minimum value is removed.

2.3.2 Global Region Contrast Salient Object Detection

Because humans are more likely to pay attention to regions with higher contrast to their surroundings Eihhauser and Konig (2003), this study proposes removing unnecessary background information and helping patients focus on the main objects in the images through saliency maps. In contrast to the widely known histogram contrast saliency maps, this study adopts a region contrast (RC) saliency map because it is computationally cost-efficient (Cheng et al., 2015). The detailed procedure is as follows.

Determining region contrast: The first step in determining the contrast is to use a graph-based image segmentation method (Felzenszwalb and Huttenlocher, 2004). Given a region r_k the saliency value can be measured by finding its color contrast to other regions in the image through

$$S(r_k) = \sum_{r_k \neq r_i} w(r_i) D_r(r_k, r_i); \tag{9}$$

where $w(r_i)$ is the weight of region r_i and $Dr(,)$ is the color distance between the two regions. And the number of pixels is weighted to add contrast to larger regions. Given two regions, r_1 and r_2, the color distance can be defined as

$$D(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D_r(c_{1,i}; c_{2,j}); \tag{10}$$

with $f(c_{k,i})$ is the probability of the $i-th$ color $c_{k,i}$ among all n_k colors in the $k-th$ region where $k = \{1, 2\}$. Using the histogram contrast approach (HC) for each region is inefficient since each region typically contains a small number of colors in the color histogram of the whole image. Therefore this algorithm uses a sparse histogram representation, as it poses as a computationally efficient method. To maximize the spatial information, a spatial weighting term is introduced to (4) to prioritize the closer regions, and decrease importance of farther regions. For a region r_k the spatially weighted region can be defined as (Cheng et al., 2015; Felzenszwalb and Huttenlocher, 2004):

$$S(r_k) = w_s(r_k) \sum_{r_k \neq r_i} \exp -\frac{D_s(r_k, r_i)}{\sigma_S^2} w(r_i) D_r(r_k, r_i); \tag{11}$$

where the spatial distance D_r is considered between regions r_k and r_i. The spatial distance between the regions is defined as the Euclidean distance between their centroid values.

2.3.3 Color Quantization Algorithm and Color Space Comparison

This study proposes a color depth mapping technique using various color spaces to cluster the colors in the image. The color depth of each pixel in the downsampled image will be reduced to 3 bits (8 colors), while the original colors will be projected to the closest color among the 8. For a given image with M number of colors, the image is partitioned into n distinct clusters with each cluster representing a unique color using the MiniBatchKmeans clustering algorithm. Moreover, not only is the distance between the individual pixel values and centroid decreased, but also the time it takes for each centroid calculation is lowered in comparison to the traditional K-means clustering algorithm. Processing time reduction is critical to improving a patient's experience.

Because the commonly used RGB color space combines the intensity with its color information, it can be insufficient to perform color clustering. The CIELAB or L*a*b color space is the most widely used model to calculate the distance from the centroids, as it has high perceptual accuracy and matches human perception of lightness. However, the YCbCr color space, commonly used in video and photography systems, has not been explored for color transformation. Because the Y or luma component is non-linearly encoded with RGB values, it is possible to separate the CbCr (color) component to cluster it individually and add the Y (intensity) component after. To find each color space's performance, this study calculates the mean-squared error (MSE), as well as the structural similarity (SSIM) of each when paired to images of varying color and number of objects.

2.3.4 Optimal Image Downsampling for Color Preservation

Due to the size restriction in the microelectrode array, visual perception is limited to low-resolution images. The bicubic-interpolation and the nearest-neighbor interpolation are commonly used to downsample images. The former presents a blurring effect that better preserves the color features of an image, at the expense of spatial details. In this study, I adopt the bicubic interpolation to increase the quality of the color features.

2.4 An Autoencoder-Optimal Transportation (AE-OT) Model for Image Augmentation

Due to the physical and computational limitations of retinal prosthesis, new patients will have steep learning curves to adapt to the devices and recognize objects in images from the VPU. Therefore, it is necessary to develop a training program to help patients adapt to these scenes. A critical component of this program would need an image library generated by VPU systems. Such image library is currently not available for prosthesis recipients. The existing libraries for computer vision research such as ImageNet only provide images with resolutions much higher than those produced by retinal prosthesis.

Converting the existing image libraries suitable for prosthesis recipients is extremely time-consuming and expensive. A fast and economical strategy is to learn models from a small set of images and generate new ones for patient training. Generative adversarial networks (GANs) and variational autoencoders (VAEs) are one of the most effective approaches for unconditional image augmentation (Lei et al., 2020). GANs train an unconditional generator that regresses real images from random noises and a discriminator that measures the difference between produced samples and input images. However, GANs are sensitive to hyperparameters and often suffer from mode collapse, a phenomenon where the generator can only produce one type of a small set of outputs.

More generally, the fundamental principles of deep learning remains primitive. A recent study of nine different GAN models and variants determined that gradient-descent-based GAN optimization is not always locally convergent (Mescheder et al., 2018). Variational autoencoders use an encoder to map data from a data distribution to a Gaussian latent distribution, which is then mapped back to the data distribution via a decoder. However, VAEs can generate ambiguous images on multimodal data distributions.

To overcome these issues, this study will apply an autoencoder-optimal transportation (AE-OT) model, which is the state-of-the-art image augmentation method. But AE-OT models have stringent requirements about input image quality. Deep learning is very sensitive to image noise and this issue is particularly severe for low-resolution images as in the case of my study. The common strategy is to pre-process the input images, which is tedious and time-consuming. My proposed methodological framework nicely circumvents this issue through the virtual magnifier. It pre-processes the input images by increasing the density of critical objects. It essentially automates the process of image pre-processing. Figure 6 illustrates the proposed new framework to augment the images for training.

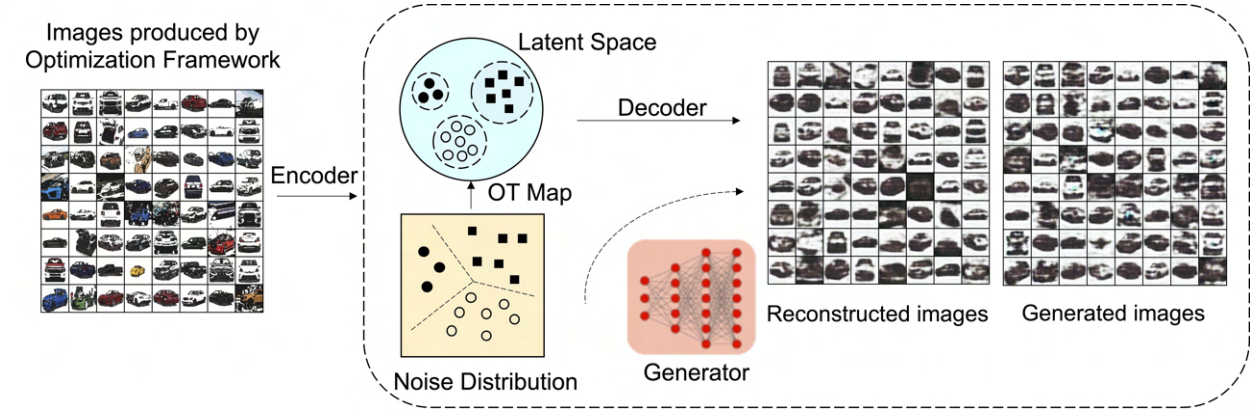

Figure 6: *The AE-OT framework to augment images produced from the Optimization Framework.*

The fundamental hypothesis of GANs can be explained by the manifold distribution hypothesis, which states that a specific class of natural data is concentrated on a low-dimensional manifold embedded in a high-dimensional background space (Tenenbaum et al., 2000). The real data distribution ν is concentrated on a manifold Σ embedded in the ambient space χ. (Σ, ν) together show the intrinsic structure of the real data sets. GAN models compute a generator map g_θ from the latent space $\mathcal{Z}$ to the manifold Σ, where θ represents the parameter of a deep neural network.

The latent space $\mathcal{Z}$ consists of Gaussian distribution ζ. The generator map g_θ pushes forward or map ζ to the generated distribution $\mu(\theta)$ in the manifold Σ. The discriminator calculates a distance between the real data distribution ν and the generated distribution $\mu(\theta)$, such as the Wasserstein distance $Wc(\mu(\theta), \nu)$.

I first briefly introduce the theoretical basis of AE-OT model. GANs mainly accomplish (1) manifold learning—namely, computing the decoding/encoding maps between the latent space $\mathcal{Z}$ and the ambient space χ; and (2) probability distribution transformation, which involves transformation between the given white noise and the data distribution. The generator map is decomposed as

$$g_\theta = h \circ T \tag{12}$$

where $h : \mathcal{Z} \to \Sigma$ is the decoding map from the latent space to the data manifold Σ in the ambient space, and the probability distribution transformation map $T : Z \to Z$.

The decoding map h is for manifold learning, and the map T is for measure transportation. OT provides rigorous and powerful ways to compute the optimal mapping to transform one probability distribution into another distribution, and to determine the distance between them. OT theory can be utilized to compute the probability distribution transformation map T, the Wasserstein distance $Wc(\mu(\theta), \nu)$ between the generated data distribution $\nu(y)$ and the real data distribution $\mu(x)$, and the minimal total transportation cost. This cost of an OT map is called the Wasserstein distance:

$$W_c(\mu, \nu) = \underset{T_{\#}\mu=\nu}{min} \int_\Omega c(x, T(x)) d\mu(x) \tag{13}$$

OT interpretation of GANs makes part of the black box transparent and OT-baed GANs reduce the training difficulty and avoid mode collapses. Therefore, this study adopts the AE-OT model. More importantly, I propose to improve the performance of the AE-OT model by pre-processesing the input images by increasing the density of critical objects through the virtual magnifier.

3 Computational Experiments and Results

This section presents the numerical and visual results to demonstrate the effectiveness of the proposed methodology framework. The following results are obtained for three major modules in the framework: the innovative virtual magnifier, novel image optimization framework, and the modified AE-OT model.

Experimental Setup: The research requires a laptop, PyCharm Community Edition 2019.3.1, Visual Studios, and OpenGL and OpenCV for visualization and user interface. All the algorithms were developed using C++ and Python on Windows and Mac platforms. All the experiments were conducted on a Windows SurfacePro70380 laptop with Intel Core i7-1065G7 1.50 GHz CPU and 16 GB memory and a MacBook Pro with a 2.9 GHz Dual-Core Intel Core i5 and 8 GB memory. To train the autoencoder model, a Dell Precision T-5820

workstation with an Intel Xeon W-2125 Processor (4 Core, 4.0GHz, 8.25MB Cache, 120W), 32GB (2x16GB) 2666MHz DDR4 RDIMM ECC was used along with the NVIDIA Quadro p6000 GPU. The paper was written in LaTeX through Overleaf.

3.1 Virtual Magnifier

Evaluation through Visual Test: As the proposed virtual magnifier is developed to help patients focus in on ROI's, this study first evaluates the pipeline visually. The algorithm successfully warped the images to enlarge specified objects while maintaining local details and angles (Figure 7). Further, it was able to magnify objects in crowded images, as seen in the car example where the group of people in the back was enlarged. The densities shown in the gaussian mixture align with the densities selected in the second column, therefore confirming the effectiveness of the magnifier.

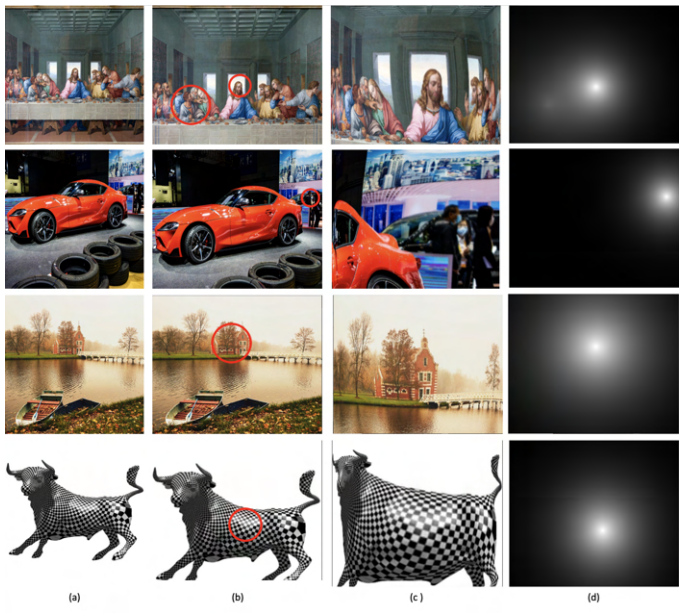

Figure 7: *Visual demonstration of virtual magnifier. (a) Input Image, (b) means and standard deviations of Gaussian densities, (c) Optimal transport (OT) map, and (d) Gaussian mixture.*

Evaluation through Accuracy Test: In order to measure the accuracy of my proposed algorithm, I implemented area-preserving histograms as a metric. For each ROI on an image, I computed its original area and the final planar area, then plotted the histogram of the logarithms of the area distortion factors. The histograms highly concentrated on the origin, showing that the virtual magnifier preserves the ROIs with high accuracy.

Evaluation through Scalability Test: The proposed algorithm was tested with different resolutions, from 256 x 256 up to 1024 x 1024. The running times are reported in Table 1. This experiment proves the magnifier's scalability and stability. In other mapping methods like Sinkhorn and the convex geometric algorithm, the computational efficiency is slow, as

they cannot handle OT problems above moderate complexity.

Table 1: *Scalability test on 5 different images at 256 x 256, 512 x 512, and 1024 x 1024.*

Image	256	512	1024
Car	5.702	6.886	8.772
The Last Supper	24.121	26.894	34.796
Female	3.95	4.008	4.487
Male	3.628	4.036	5.502
Sunflower	9.21	9.617	10.228

3.2 Image Optimization Framework to Encode Color and Spatial Information into Retinal Prostheses

3.2.1 Salient Object Detection Experimental Comparisons

This study compares the region-contrast saliency maps (adopted in this study) to four other saliency maps reported in (Cheng et al., 2015; Ma and Zhang, 2009; Harel et al., 2007; Achanta et al., 2009) respectively. The saliency maps generated by the region contrast algorithm are of much higher quality visually, because of its ability to distinguish and portray the objects in contrast to its background (Figure 8).

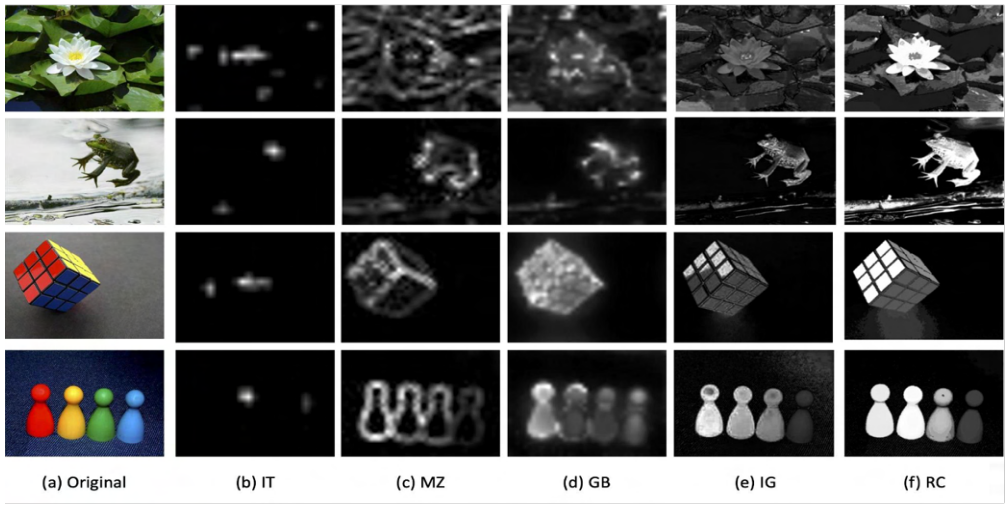

Figure 8: *Visual comparison of saliency maps: (a) input image, (b) saliency maps from (Cheng et al., 2015), (c) saliency maps from (Ma and Zhang, 2009), (d) saliency maps from (Harel et al., 2007), (e) saliency maps from (Achanta et al., 2009), (f) region-contrast method in this study.*

Table 2 compares the region-contrast saliency maps to maps from (Cheng et al., 2015; Ma and Zhang, 2009; Harel et al., 2007; Achanta et al., 2009) based on the MSE and SSIM values. I also find that SSIM is a better metric. For example, MSE incorrectly concludes that maps such as the frog generated with the IT approach is of greater accuracy. Especially

Table 2: *MSE and SSIM of the maps from the method proposed, IT (Cheng et al., 2015), MZ (Ma and Zhang, 2009), GB (Harel et al., 2007), and IG (Achanta et al., 2009)*

	IT		MZ		GB		IG		RC	
image	MSE	SSIM	MSE	SSIM	MSE	SSIM	MSE	SSIM	MSE	SSIM
lotus	7779.45	0.05	5906.95	0.2	3394.31	0.36	4426.91	0.25	728.21	0.58
frog	378707.37	0	39097.61	0.02	35239.6	0.02	36923.46	0.05	34303.75	0.16
rubik's cube	9927.65	0.01	8434.6	0.06	9575.38	0.15	8572.88	0.09	7849.32	0.16
pins	4285.96	0.04	3966.07	0.06	3184.76	0.25	2584.39	0.12	2242.94	0.25

in this study, where the saliency maps are used for object display to patients, comparing the images visually is a crucial aspect for comparison.

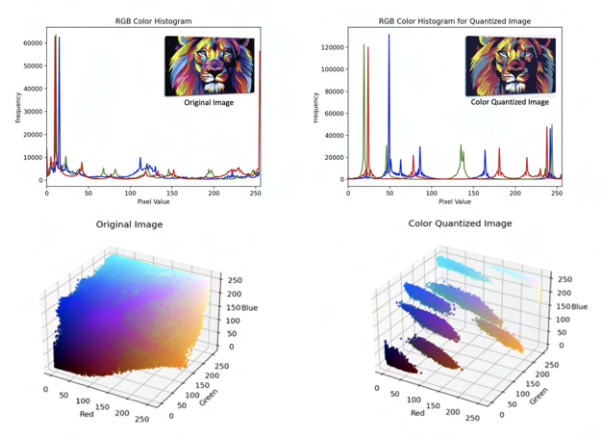

Figure 9: *3D Maps and histograms for the lion original image compared to the color quantized image for the lion at 8 colors.*

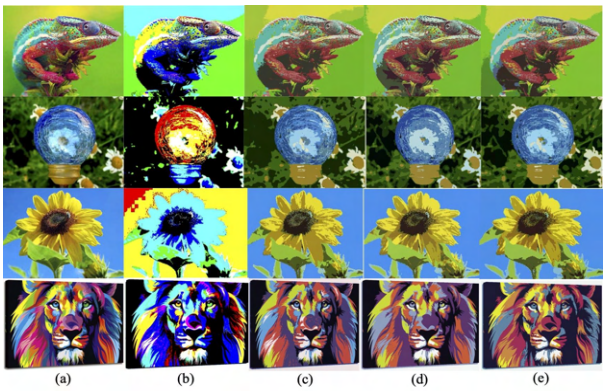

Figure 10: *Representation of (a) original image, (b) color depth mapping through manually defining colors, (c) color clustering through RGB color space, (d) color clustering through CIELAB color space, (e) color clustering through YCbCr color space.*

3.2.2 Color Quantization Analysis

Figure 9 demonstrates the effectiveness of MiniBatchKmeans, as the pixels are clustered to 8 batches when downsampled to 3 bits (8 colors). To study the effectiveness of the individual color spaces, I first need to compare them visually. As a reference, I also generated color images clustered to self-defined colors by computing the distance of each pixel RGB value to those colors (eg: red (255, 0, 0), blue (0, 0, 255)) (Fig.3b). The clustering results showed little difference among the three color spaces based on the image quality (Figure 10).

This is also supported by MSE and SSIM metrics shown in Table 3, where the output performance varies with input image complexity, color contrast, and pixel intensity. However, clustering through machine learning (Fig.10c) is still much more effective than clustering through manually defined colors (Fig.10b).

Finally, to show the effectiveness of the proposed framework, another validation study

was conducted. Figure 11 shows its advantage over the current prosthesis output, which fails to add color and identify the salient objects.

Table 3: *MSE and SSIM when tested on clustering in RGB, L*a*b, and YCbCr color spaces Color space clustering comparison for 8 colors*

	RGB		LAB		YCbCr	
image	MSE	SSIM	MSE	SSIM	MSE	SSIM
chameleon	576.6020863	0.687294967	769.2845623	0.678478287	597.5151649	0.681415638
light	425.7828083	0.566199211	468.1894139	0.576582033	385.7437528	0.605459881
sunflower	335.0843716	0.736839507	369.1024077	0.725594999	352.0570889	0.73564814
lion	1017.107602	0.605669868	1271.245059	0.609444552	1259.462023	0.565489579

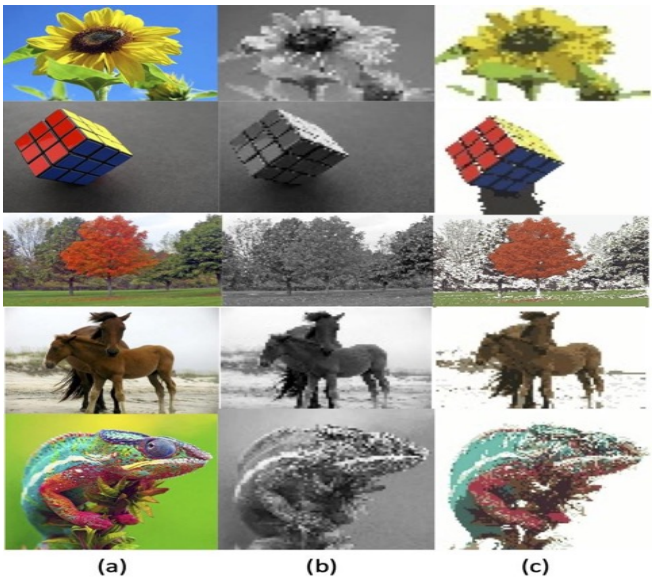

Figure 11: *(a) Original Image, (b) Current Prosthetic Output, My Algorithm Output.*

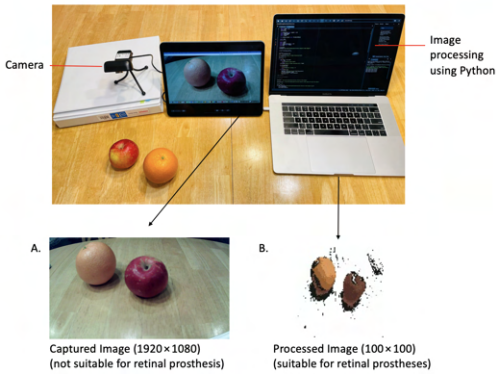

Figure 12: *Illustration of the simulated virtual processing unit (VPU) platform.*

3.3 Prototype System Experimental Setup

I set up a prosthetic vision simulation platform which consists of a webcam (HD 1080p, 30 fps, Dericam) and a MacBook Pro (Dual-Core Intel Core i5 @2.9 GHz, 8 GB). The images are captured by the webcam mounted on a tripod and transmitted to the computer. The software written in Python has been developed to receive and

Table 4: *Speed comparison when downsampled to 55 x 37*

Functions	Algorithm Test
Edge detection with Canny operator	8.1 ms
Region contrast saliency object detection	0.0043 ms
Image merging	36ms
Color uniform quantization	340ms
RGB $\longrightarrow$ Defined color space	3.6ms
Bicubic downsampling interpolation	0.34ms
Total Time	388.04 ms

process the images according to algorithms developed from the aforementioned optimization framework. A prototype frame for prostheses was also 3D-printed to store the camera (Figure 14). The simulated prosthesis vision (SPV) experiment shown in Figure 12 is tested on an apple and an orange.

Overall optimization framework implementation speed: The key algorithm processes of my proposed framework and the time taken to process each are stated in Table 4. Execution time results show that the region-based saliency object detection algorithm can realize running time of approximately .0043 ms on the cropped images with 275 $\times$ 183 resolution, and the color depth mapping realizes a time of approximately 3.6 ms, when converted to the YCbCr space for color clustering. With a performance speed of around 390 ms, the proposed prototype system proves that the framework not allows for real-time video transfer to patients, but also identifies the salient object while reducing the number of colors.

3.4 Autoencoder-Optimal Transport (AE-OT) Model Performance

This section visualizes the computational output of the auto-encoder model for the purpose of generating a library of low-resolution images for patient training. This paper shows the results generated using datasets of my family, cars, and pears (Figure 13). However, I also augmented images from pet, flower, and other publicly available datasets. The training time of the models was extensive, ranging from 5-10 hours. The results show the generality of the AE-OT model and its efficacy in image augmentation, confirming its potential to be applied to retinal prostheses users.

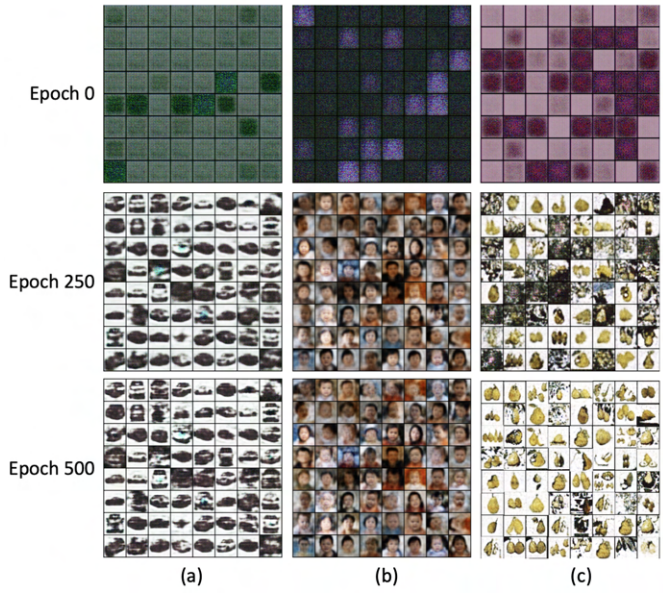

Figure 13: *AE-OT training on (a) Car dataset, (b) My family dataset, (c) Pear dataset.*

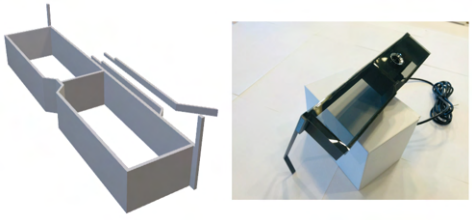

Figure 14: *CAD prototype model for prosthesis frame.*

4 Discussion and Conclusion

This study develops a novel image optimization and augmentation framework to improve the user experience for users of retinal prostheses. A prototype system is also developed to demonstrate the proposed methodology framework. The computational experiments and prototype system shows that my developed method are able to 1) localize and "magnify" key segment(s) in an image frame while preserving important features and curvatures of objects in the segment(s); 2) optimize the magnified segments to encode the maximum amount of spatial and color information to the patients through attention mechanisms as well as color scheme comparisons; and 3) augment the optimized images to enable new patients to quickly adapt to the prosthesis.

The virtual magnifier utilizes image density as a "digital knob" to adjust the size of region of interests. This enable a hardware knob for patients to tune the virtual magnifier conforming to patient preference. The results show that magnifier is accurate, efficient, and scalable. The image optimization framework was developed through edge detection, salient object detection, color quantization, and bicubic interpolation. A region-contrast based computation was adopted by combining color and intensity difference features, aiming to quickly and effectively detect foreground objects in images. When compared to other saliency map computation techniques, the region contrast maps were more effective in salient object identification. This study also proposed a color quantization algorithm that performs pixel color depth mapping through a MiniBatchKmeans clustering algorithm in 3 color spaces. Finally, the AE-OT model successfully augments images to potentially generate an image library for patient training. This suggests that the real-time image optimization and augmentation strategy based on the proposed framework may provide a promising method for the development of the image processing module in future retinal prostheses. Further research can be devoted to reducing the training time of image augmentation, enhancing the performance of the AE-OT model by increasing the dimension of the latent space, and providing patients with thermal imaging as an additional cue for salient object detection.

Acknowledgments

I would like to thank the great mentoring from Dr. Lan Yue at the U.S. Food and Drug Administration (FDA) for introducing me to this research and for providing insightful feedback. Finally, I am beyond grateful for my science research teacher, Ms. Melissa Grace Klose, for her encouragement and guidance.

References

Achanta, R., S. Hemami, F. Estrada, and S. Susstrunk (2009). Frequency-tuned salient region detection. In *2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL*, pp. 1597–1604.

Boyle, J. (2008). Region-of-interest processing for electronic visual prostheses. *Journal of Electronic Imaging 17*, 013002.

Brindley, G. S. and W. S. Lewin (1968). The sensations produced by electrical stimulation of the visual cortex. *The Journal of physiology 196*(2), 479–493.

Cheng, M., N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu (2015). Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*(3), 569–582.

Eihhauser, W. and P. Konig (2003). Does luminance-constrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience 17*, 1089–1097.

Felzenszwalb, P. and D. Huttenlocher (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision 59*, 167–181.

Gu, X., F. Luo, J. Sun, and S. Yau (2016). Variational principles for minkowski type problems, discrete optimal transport, and discrete monge–ampère equations. *Asian Journal of Mathematics 20*(2), 383–398.

Harel, J., C. Koch, and P. Perona (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems 19*, 545–552.

Horsager, A., S. Greenwald, J. Weiland, M. Humayun, R. Greenberg, M. McMahon, G. Boynton, and I. Fine (2009). Predicting visual sensitivity in retinal prosthesis patients. *Investigative Ophthalmology and Visual Science 50*(4), 1483–1491.

Kozarsky, A. Macular degeneration (amd): Symptoms, causes, treatment, prevention.

Lei, N., D. An, Y. Guo, K. Su, S. Liu, Z. Luo, S.-T. Yau, and X. Gu (2020). A geometric understanding of deep learning. *Engineering 6*(3), 361–374.

Li, R., X. Zhang, and G. Hu (2005). A computational pixelization model based on selective attention for artificial visual prosthesis. pp. 654–662.

Luo, Y. H. and L. da Cruz (2016). The argus ii retinal prosthesis system. *Progress in Retinal and Eye Research 50*(89-107).

Ma, Y.-F. and H.-J. Zhang (2009). Contrast-based image attention analysis by using fuzzy growing. In *ACM International Conference on Multimedia*.

Mescheder, L. M., A. Geiger, and S. Nowozin (2018). Which training methods for gans do actually converge? In *ICML*.

Mills, M. Retinal diseases: Age-related macular degeneration and retinitis pigmentosa.

Normann, R., B. Greger, P. House, S. Romero, F. Pelayo, and E. Fernandez (2009). Toward the development of a cortically based visual neuroprosthesis. *Journal of Neural Engineering 6*, 035001.

Sakaguchi, H., M. Kamei, T. Fujikado, and et. al. (2009). Artificial vision by direct optic nerve electrode (av-done) implantation in a blind patient with retinitis pigmentosa. *Journal of Artificial Organs 12*, 206–209.

Sharmili, N., P. S. Ramaiah, and G. Swamynadhan (2011). Image compression and resizing for retinal implant in bionic eye. *International Journal of Computer Science & Engineering Survey 2*, 30–37.

Tenenbaum, J. B., V. de Silva, and J. C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science 290*(5500), 2319–2323.

Tsankashvili, N. Comparing edge detection methods.

van Rheede, J. J., C. Kennard, and S. L. Hicks (2010). Simulating prosthetic vision: Optimizing the information content of a limited visual display. *Journal of vision 10*(14).

Wang, J., H. Li, W. Fu, Y. Chen, L. Li, Q. Lyu, T. Han, and X. Chai (2016). Image processing strategies based on a visual saliency model for object recognition under simulated prosthetic vision. *Artificial Organs 40*, 94–100.

Yue, L., J. Weiland, B. Roska, and M. Humayun (2016). Retinal stimulation strategies to restore vision: Fundamentals and systems. *Progress in Retinal and Eye Research 53*, 21–47.

Zhao, X., W. Zeng, X. Gu, A. Kaufman, W. Xu, and K. Mueller (2012, 11). Conformal magnifier: A focus+context technique with minimal distortion. *Visualization : proceedings of the ... IEEE Conference on Visualization. IEEE Conference on Visualization 18*(11), 1928–1941.