2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Assessing Vision Quality in Retinal Prosthesis Implantees through Deep Learning: Current Progress and Improvements by Optimizing Hardware Design Parameters and Rehabilitation

Alexandros Benetatos*, Nikos Melanitis† and Konstantina S. Nikita†

***Abstract*— Retinal prosthesis (RP) is used to partially restore vision in patients with degenerative retinal diseases. Assessing the quality of RP-acquired (i.e., prosthetic) vision is needed to evaluate RP impact and prospects. Spatial distortions caused by electrical stimulation of the retina in RP, and the low number of electrodes, have limited the prosthetic vision: patients mostly localize shapes and shadows rather than recognizing objects. We simulate prosthetic vision and evaluate vision on image classification tasks, varying critical hardware parameters: total number and size of electrodes. We also simulate rehabilitation by re-training our models on prosthetic vision images. We find that electrode size has little impact on vision while at least 400 electrodes are needed to sufficiently restore vision (more than 65% classification accuracy on a complex visual task after rehabilitation). Argus II, a currently available implant, produces a low-resolution vision leading to low accuracy $(21.3\%$ score after rehabilitation) in complex vision tasks. Rehabilitation produces significant improvements (accuracy improvement of up to 30% on complex tasks, depending on the number of electrodes) in the attained vision, boosting our expectations for RP interventions and motivating the establishment of rehabilitation procedures for RP implantees.**

***Index Terms*— retinal prosthesis, Argus II, visual rehabilitation, prosthetic vision, visual recognition tasks**

I. Introduction

Retinitis pigmentosa, macular degeneration and other degenerative retinal diseases are causing irreversible vision loss to more than 100 million people worldwide [1]. Through retinal prosthesis (RP) we are able to restore vision by electrically stimulating the retina to evoke neuronal responses that are interpreted by the brain as visual perceptions. However, current retinal implants still provide a limited vision: implantees can mostly localize shapes and shadows rather than recognize actual objects [2]. Still, integrating models that predict retinal response to RP interventions will lead to improved vision [3].

Spatial distortions, caused from the stimulation of ganglion cells' axons in the region of the activated electrodes (axonal stimulation), impede the generation of precise and localized visual perceptions. The low number of electrodes in current retinal implants leads to poor and low resolution vision.

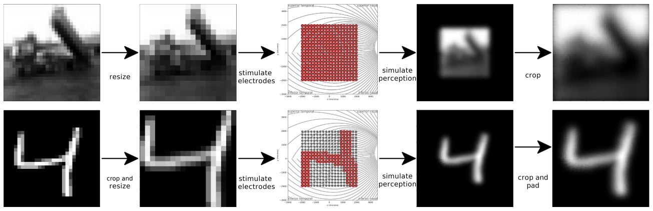

Fig. 1. Data pre-processing in CIFAR (top-row) and MNIST (bottom-row): To obtain the samples of the patients' prosthetic vision, we crop (MNIST only) and resize the dataset sample to the implant's grid size (2nd column). We stimulate the electrodes (3rd column) and get the simulated visual perception (4th column). We crop the black area around the implant and (MNIST only) we add padding to match the style of the original images (5th column). Then, we feed these images to the HCA model. In both rows, the implants have a 19×19 grid of electrodes of $50\mu m$ radius.

In this paper, we identify critical implant design parameters that influence prosthetic vision (i.e., vision attained by RP). We evaluate the impact of the number and size of electrodes. We introduce a novel approach to evaluate the quality of prosthetic vision using image understanding tasks, as digit and object recognition. We find that current RP technology (Argus II [4]) provides low resolution vision and does not enable implantees to handle such tasks. To show the improvements in RP-attained vision and in implantees autonomy through a rehabilitation process, we simulate rehabilitation by retraining our models on distorted prosthetic vision images.

II. Methods

In order to evaluate how the implants' number and size of electrodes affect prosthetic vision, we follow a three step process: (i) we set up a model to simulate object or digit recognition, (ii) we simulate prosthetic vision in square, fixed area, retinal implants of various parameters, in a subset (10%) of the MNIST [5] (digit recognition) and the CIFAR 10 [6] (object recognition) datasets, (iii) we evaluate the effect of the spatial distortions in prosthetic vision and retrain our models (step (i)) on the distorted images to simulate implantees rehabilitation to the prosthetic visual percepts.

A. Simulating implantees ability to classify images

Over the last decade, since DanNet's breakthrough [7], Deep Convolutional Neural Networks (DCNNs) have been the state of the art for image classification. DCNNs can be usually divided into the backbone and the classification head. The backbone primarily consists of convolution and pooling layers in a hierarchical layout where lower layer outputs are fed to upper layers inputs. The final output of the backbone is

*Alexandros Benetatos is with the School of Electrical and Computer Engineering, National and Technical University of Athens, Greece alexandrosbene@gmail.com

†N. Melanitis and K. S. Nikita are with the School of Electrical and Computer Engineering, National Technical University of Athens, Greece

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

converted into a fixed size vector using max pooling and fed to the classification head, which is a multi-layer perceptron.

We model Human Classification Ability (HCA) with a state of the art DCNN. More specifically, we use a ResNet-34 model [8], with 21.5 million parameters, pre-trained on ImageNet [9]. Using a pre-trained and deep convolutional network (ResNet) we get out-of-the-box rich visual representations which can be used in different visual tasks [10]. The intuition behind this is that, using large and diverse image datasets like ImageNet for training, the convolutional layers learn to emphasize on more important visual features. This results in a richer and denser feature vector being generated by the backbone, that are used for the classification task [10]. We apply the transfer learning paradigm: we use this feature vector with a new output layer for the classification head which we train on our datasets (MNIST, CIFAR 10). Then, we also train the full model (backbone and classification head) for some epochs (fine-tuning).

B. Simulating the visual perception induced by custom retinal implants

We use pulse2percept library for python to simulate prosthetic vision attained by retinal implants [11]. Pulse2percept allows us to simulate the visual perception induced by electrode stimulations in a custom designed implant. We apply the ScoreBoard and AxonMap models to simulate the resulting visual perception [12]. The simple ScoreBoard model assumes that the stimulation of a grid of electrodes in the retina results in a grid of independent focal spots of light (Gaussian blobs). The more complex and realistic AxonMap model takes into consideration the spatial distortions caused by the excitation of ganglion axon pathways around the stimulated electrodes [12].

1) Custom square retinal implants of $4000 \times 4000\ \mu m^2$ *area:* We simulate 182 implant configurations of a constant area ($4000 \times 4000\ \mu m^2$) but with varying electrodes number and size. We simulated implants with square electrode grids sizing from 5×5 to 25×25 electrodes with a step of 2 electrodes per side and electrode sizes varying from 10 μm to 400 μm with a step of 10 μm and no overlap between neighboring electrodes. We simulated both ScoreBoard and AxonMap models.

2) Case Study (Argus II): We simulate Argus II, a commercially available retinal implant, in both ScoreBoard and AxonMap models. Argus II consists of a 6×10 grid of electrodes of size 200 μm with a distance of 575 μm between the centers of each electrode resulting in a $3075 \times 5375\ \mu m^2$ area implant [4].

3) Images Pre-Processing: In order to improve the visual acuity in prosthetic vision, we process and enlarge the input images to crop any background and stimulate the maximum number of electrodes with useful stimuli (Fig. 1). In detail, we cropped the padded images from the MNIST dataset in a square shape tightly around the numeric characters and then sub- or up-sampled them to the size of the implant (Fig. 1 - top row). For the CIFAR 10 dataset, we only sub- or up-sampled the images since there is no padding around them to crop (Fig. 1 - bottom row). Then, we further processed the simulated prosthetic vision images: we cropped again tightly around the implant's area, added padding only for the MNIST dataset results to match the style of the original padded images, and reshaped all the images to a fixed 224×224 pixels size to use them as inputs in our HCA model.

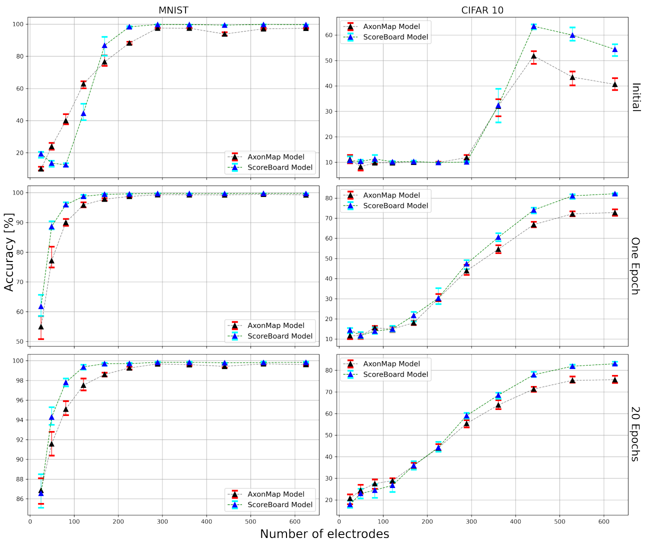

Fig. 2. To evaluate the impact of the number and size of electrodes, the rehabilitation process and the spatial distortions (AxonMap model), we plot HCA model testing accuracy. In MNIST we reach a saturation point at around 150 electrodes. In the more complex CIFAR 10 we need at least 350 electrodes. Also, the small deviation between max and min classification accuracy achieved by varying the electrodes size, shown by the top and bottom bars surrounding every triangle in the plots, suggests that electrode size has a minimal impact (at most 8% after the rehabilitation period). For very easy visual tasks (MNIST) the ScoreBoard model approaches the AxonMap model relatively well, while for more complex tasks (CIFAR 10) the two models diverge as the number of electrodes increase.

C. Evaluating the simulated visual perceptions before and after a simulated rehabilitation period

We evaluate prosthetic vision by measuring the classification accuracy of the HCA model on the prosthetic vision images. We simulate a rehabilitation period by retraining the HCA model on the prosthetic vision images.

In detail, we retrain the top layer of our HCA model, independently for each implant configuration and retina model (i.e., ScoreBoard and AxonMap models). We evaluate classification accuracy after each training epoch. We stop training after accuracy improvement has saturated, at 20 epochs. We treat the one-epoch model as an approximation of the patient's visual classification ability after just a short period of time (e.g. hours - days) and the best-out-of-20-epochs model as an approximation of the visual classification accuracy after a longer rehabilitation period (e.g. months).

III. Results and Discussion

A. Expected recognition rates decrease as the number of electrodes decrease and level of distortion increases

Applying a level of distortion to an image causes the recognition rates in classification tasks for both humans and

DCNNs to smoothly decrease as the level of distortions increases [13]. This smooth transition from the less distorted prosthetic vision images (higher number of electrodes) to the more distorted images (lower number of electrodes) is observed in our results (Fig. 2) where, excluding the untrained HCA model in CIFAR 10 dataset, we observe a smooth transition from lower to higher number of electrodes.

B. The HCA model

The HCA model (i.e., the ResNet-34 DCNN) achieved a high classification accuracy after training on MNIST (99.60%) and CIFAR10 (96.48%) sets. In MNIST dataset we first trained the top layer of the model for 15 epochs, we chose the model with the best evaluation accuracy and continued training the whole model for 10 epochs, for fine-tuning, choosing the best evaluation model as the HCA model. In CIFAR10, we converted the images from rgb to grayscale and trained our model as in MNIST case, but instead trained the top layer for 30 epochs and fine-tuned for 20: CIFAR 10 represents a more difficult task than MNIST, thus our model needs more training to reach similar levels of performance.

The correlation between human's visual perception system and DCNNs architecture has been known for years. The first biomimetic visual computational models, resulting in the current DCNNs, originate from the study of Hubel and Wiesel [14] on the architecture and hierarchy of organic cells in the visual cortex [15]. They found that simple cells respond to lines of particular orientation in certain locations of the image, while complex cells receive outputs from multiple simple cells leading to spatially invariant responses. The computational analogs are the convolution and pooling layers of DCNNs. Every upper layer builds on previous layers, responding in a combination of previous activations, creating a hierarchical computation model of visual information.

However, this is the first time, to our knowledge, DCNNs are used to simulate humans' ability to recognize and classify images. Another novelty lies in retraining DCNNs to simulate a rehabilitation process and study the possible impact of spatial visual distortions occurring from retinal implants, in implantees' ability to recognize and classify objects in prosthetic vision images.

C. Custom square retinal implants of 4000×4000 μm^2 *area*

1) RP-attained visual acuity is nearly insensitive to electrode size: We observe that the size of an implant's electrodes has no substantial impact on the quality of prosthetic vision, leading to only a minimal difference in the classification accuracy. More specifically, we observe a maximum of 8% difference between the best- and worst-performing implant configuration, assuming a constant number of electrodes (Fig. 2). However, the relation between electrodes' size and the evaluation results is not clear. Between 80 and 400 electrodes there is, in some cases at least, a slight increase in accuracy as the size of the electrodes increase, however, in other cases there is a decrease in accuracy, and mostly there is very low accuracy variation with varying electrode size.

We conclude that the accuracy variation with electrode size can be mostly attributed to the stochastic nature of our rehabilitation simulations.

Consequently, in retinal implant development, we can specify electrode size to satisfy other design parameters like manufacturing cost and size of implant.

2) More than 400 *electrodes are needed for complex visual tasks:* We observe (Fig. 2) that increasing electrodes' number beyond a point (around 150 in MNIST and 500 in CIFAR 10) offers minimal improvement in the classification task (saturation point). The different saturation points can be explained by the complexity of each task: MNIST is one of the easiest datasets in computer vision while CIFAR 10 is a much more difficult one - recognizing digits (MNIST) is less resolution-sensitive than object classification (e.g. cars, birds frogs) in CIFAR 10.

In easy visual tasks under ideal conditions (MNIST) we achieve good performance with 80 electrodes. However, for more complex classification tasks (CIFAR 10) we need more than 350 electrodes (Fig. 2).

Saturation in performance occurs since in prosthetic vision, for any number of electrodes, the images are distorted and have low sharpness (Fig. 1) and so the HCA model (and thus the implantees) cannot reach a higher score for the visual tasks: further improvements in electrodes' number and consequently prosthetic vision resolution offer negligible performance improvements.

3) ScoreBoard model is a good approximation of AxonMap model in some cases: In easy visual tasks (MNIST) the ScoreBoard model is a good approximation of the AxonMap model, especially when the number of electrodes is large (more than 150) (Fig. 2 - top row). In more complex tasks (CIFAR 10) we see that the two models diverge as the number of electrodes increases (Fig. 2 - bottom row).

4) Informal validation: We did an informal validation (sanity-check) of our hypothesis on rehabilitation: that the one-epoch model represents short term and the best-out-of-20-epochs model long-term rehabilitation. We provided ten volunteers with twenty prosthetic-vision simulated CIFAR 10 samples for a $25 \times 25 - 30$ μm implant (i.e., implant with electrode grid size is 25×25 electrodes of 30 μm radius) and a $15 \times 15 - 30$ μm implant for five days. Each participant labeled the images. Each day (i) we logged the classification accuracy results and (ii) "trained" the participants by revealing the true image labels for each sample they examined. By the fifth day the participants had reached a saturation point in the classification task and the relative classification accuracy improvement, for both implants we examined, between day 1 and day 7, was around 30% percent. These results correlate with our simulated evaluation results (Fig. 1 - bottom row) showing a 30% improvement over the rehabilitation period.

In any case, the most important aspect of our results is not the actual absolute classification accuracy for the visual task, but rather the relative results about the insensitivity to the electrode's size, the saturation point for the number

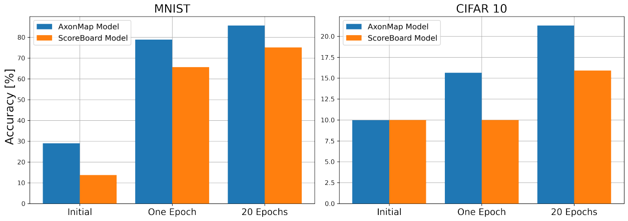

Fig. 3. Evaluation accuracy for Argus II attained vision: We observe that results for the more complex CIFAR 10 are quite close to random guessing. Still, rehabilitation noticeably improves the RP-attained vision

of electrodes and the improvement from the rehabilitation process.

D. Case Study (Argus II)

We repeat our simulations for the Argus II implant. We see that the classification task accuracy lies between the 7×7 and 9×9 custom square implant accuracy (as expected from Argus II size). In addition, we observe a correlation between the classification accuracy on MNIST and the more difficult classification of letters by actual Argus II-implantees [16]. More specifically, reports show that in the more difficult task of letter identification, separated into three classification subtasks with about 10 letters each, patients with Argus II devices implanted for 8 to 35 months achieved a mean accuracy of about 60% [16], close (considering the difficulty difference) to the accuracy in Fig. 3 after some rehabilitation period (78% to 85% accuracy).

However, in more complex classification tasks, such as the CIFAR 10, we see that the low resolution of Argus II is not sufficient to capture the needed visual information.

We conclude that current prosthetic vision has low acuity for complex tasks such as CIFAR 10 or even MNIST, as evidenced by the low accuracy score in our simulations. Still, rehabilitation can noticeably improve the RP-attained vision (i.e., 55% improvement in MNIST and 11% in CIFAR 10) (Fig. 3).

IV. Conclusion and Future Work

We examined the effect of retinal implants' design parameters on prosthetic vision. We found that the size of the electrodes has effectively no impact on performance in visual tasks and, consequently, on the quality of prosthetic vision. Furthermore, increasing the number of electrodes leads to improved vision. We observe that at least 350-400 electrodes are required to handle (i.e., around 60% classification score) a complex visual task (e.g., object recognition). While the classification score on the visual tasks increases as the number of electrodes increases, the score improvement diminishes in higher electrodes' numbers. We observed that more than 500 electrodes offer minimal improvements in the prosthetic vision acuity. We show that current RP interventions (Argus II) provide a low-quality vision that is adequate for easy visual tasks (MNIST) but not for more complex tasks (CIFAR 10). Still, performance can be substantially improved through rehabilitation.

In future work, larger datasets with more object classes (e.g. CIFAR 100 [6], ImageNet [9]) should be examined. Our evaluation was based on relatively easy visual tasks, compared to real-world situations; in our classification tasks, we know that the object always belongs to one of the classes, whose number is limited. Moreover, we process MNIST digits so that each digit covers the entire implant area - digit recognition 'in-the-wild' is harder, unless special functionalities are implemented in the RP intervention.

Moreover, investigating the impact of the shape and area covered by retinal implants is of great importance to attain a comprehensive overview of all design parameters that may influence prosthetic vision. Lastly, concerning the enormous computational resources needed to simulate even 6000 images for 182 implants using AxonMap model, GPU acceleration would enable the easy investigation of distortions, caused by ganglion axonal excitation, in prosthetic vision.

References

[1] K. Pennington and M. DeAngelis, "Epidemiology of age-related macular degeneration (amd): associations with cardiovascular disease phenotypes and lipid factors," *Eye and Vision*, vol. 3, 12 2016.

[2] L. da Cruz *et al.*, "Five-year safety and performance results from the argus ii retinal prosthesis system clinical trial," *Ophthalmology*, vol. 123, no. 10, pp. 2248–2254, 2016.

[3] N. Melanitis and K. S. Nikita, "Biologically-inspired image processing in computational retina models," *Computers in Biology and Medicine*, vol. 113, p. 103399, 2019.

[4] Y. H.-L. Luo and L. da Cruz, "The argus® ii retinal prosthesis system," *Progress in Retinal and Eye Research*, vol. 50, pp. 89–107, 2016.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278 – 2324, 12 1998.

[6] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.

[7] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "High-performance neural networks for visual object classification," *CoRR*, vol. abs/1102.0183, 2011.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, 06 2016.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[10] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, (Cambridge, MA, USA), p. 3320–3328, MIT Press, 2014.

[11] M. Beyeler, G. Boynton, I. Fine, and A. Rokem, "pulse2percept: A python-based simulation framework for bionic vision," *pulse2percept: A Python-based simulation framework for bionic vision*, p. 148015, 7 2017.

[12] M. Beyeler, D. Nanduri, J. D. Weiland, A. Rokem, G. M. Boynton, and I. Fine, "A model of ganglion axon pathways accounts for percepts elicited by retinal implants," *Scientific Reports*, vol. 9, pp. 1–16, 12 2019.

[13] G. Chen, Y. Liu, and Q. Zhao, "Perceptual decision making of humans and deep learning machines: a behavioral study," 02 2016.

[14] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, pp. 106–54, 1 1962.

[15] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, pp. 193–202, 2004.

[16] L. D. Cruz *et al.*, "The argus ii epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss," *British Journal of Ophthalmology*, vol. 97, pp. 632–636, 5 2013.