Object Instance Segmentation for Better Distinguishability in In-Door Scenes

Roman Beltiukov*
rbeltiukov@ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

Liubov Kurafeeva*
liubov_kurafeeva@ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

Francie Wei*
hwei@ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

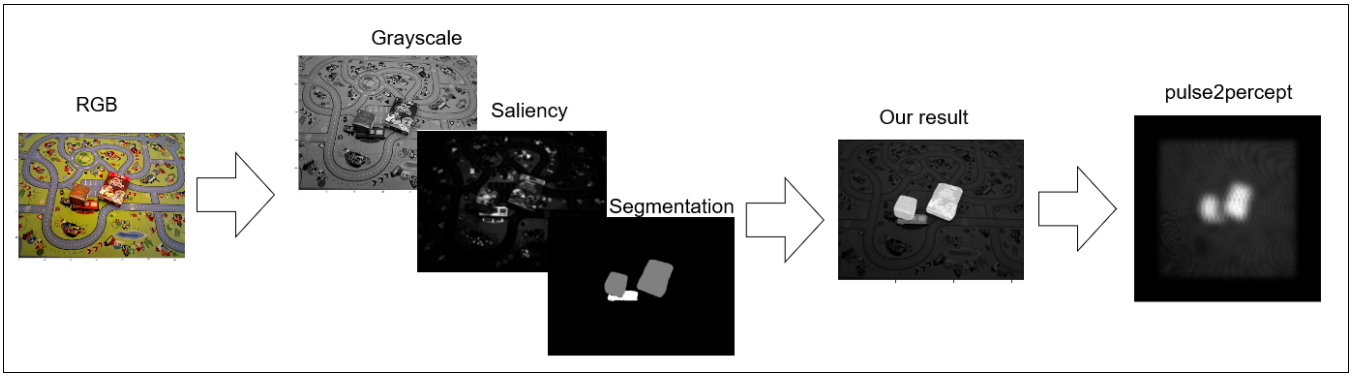


Figure 1: Proposed method of combining recolored segmentation results, saliency map and background to highlight objects and allow to better distinguish them.

ABSTRACT

Retinal degenerative diseases lead many people to lose their vision which significantly reduces their abilities to successfully participate in everyday life and the number of tasks they are able to accomplish. Furthermore, due to the low resolution of currently available certified bionic implants, several tasks (like objects differentiation and separation) are very hard or impossible. In this paper, we provide a method of combining the modern deep-learning models with graph theory to facilitate objects' location, identification, and visual separation from each other. We show that our approach steadily facilitates the mentioned task when objects are stacked or located near each other. We also provide ideas of possible improvements and extensions to this work to provide bionic implant users a better experience and facilitate their everyday lives.

KEYWORDS

bionic vision, objects identification, indoor scenes

1 INTRODUCTION

Due to different retinal degenerative diseases, many people lose their vision. Such loss significantly reduces their abilities to successfully participate in everyday life and the number of tasks they are able to accomplish. To facilitate their life and provide a way to handle new problems, implants (also called *bionic vision implants*) could be installed providing an ersatz-sight that is far from reality but still uses the same way to provide information to the human brain.

Despite continuous research and constant improvements in this area, modern approved bionic implants still provide very low resolution. This problem heavily restrains implant owners and reduces the number of tasks that could be reliably solved relying on visual information. One of these tasks is objects identification and visual differentiation. With such low resolution (up to 40 pixels for each axis), it is hard to visually disjunct near-located objects and to locate needed objects in indoor scenes.

In this work, we aim to provide a method to improve visual information available to low-resolution implant owners to allow them more easily understand the number of objects on the scene and visually separate them. We make the next contributions:

- We adjust and modify state-of-the-art computer vision algorithms based on neural networks and graph theory to modify visual observations and introduce more information to implant owners
- We evaluate our solution using open-source implementations of computational models of bionic vision and hold an unbiased comparison of our solution with plain grayscale image data.

2 RELATED WORK

Most retinal bionic vision implants utilize a stream from an external video camera and provide preprocessed images to the user. Being physically rather close to natural human sight, we can apply standard techniques of image preprocessing like rescaling, edge detection, Gaussian blur, and others. Several works [2, 7] that study bionic implant users states that increasing the contrast of objects and providing them simplified scenes increases their orientation skills and allows them to complete tasks more efficiently.

*All authors contributed equally to this research.

With new implants being developed, scientists suppose increasing of the computational power of Video Processing Units so it possibly can hold additional tasks connected with more complex image preprocessing methods, starting with seam carving [1] and up to introducing semantic segmentation of the whole image [6].

Most of the recent papers which tried to enhance bionic vision tried to apply different saliency detection algorithms and add this information to the image [10, 11, 13, 16, 18]. Saliency information is defined as information about image pixels that differ from surroundings, so the human eye would typically focus on them. To obtain a saliency map, there exist deep learning models [8] that provide qualitative predictions. However, despite saliency maps being useful, they are *purpose-agnostic* so they highlight the information no matter what the task is and therefore wouldn't help much to search certain objects.

Several papers also introduced using depth information [12, 14] to enhance the image and provide the user some additional context. Unfortunately, these papers focused on an outdoor environment where depth sensors indisputably would help locate dangerous objects. Still, very little information about indoor objects search and any tasks except the basic location are presented.

The idea of applying segmentation or object detection models to the video stream isn't new in this area, as several papers have tried to do this. The paper by Weiland et al. [19] tried to adjust the object detection algorithm to facilitate objects' location on the empty white table. However, objects were pretty separated and located far from each other. The paper by Horne et al. [6] also introduced semantic segmentation of the image, but they tried to segment all the images and highlight those areas which are important in terms of outdoor navigation. Besides that, in the recent work of Han et al. [5] a rather close approach was presented but with a focus on the outdoor environment.

Finally, the closest to our ideas Sanchez-Garcia et al. [15] work was published two years ago. This paper introduces fully convolutional networks to segment the objects in indoor scenes. Despite being very close to our work, the paper focuses on directly highlighting the objects considering they are already far from each other and do not involve the separation task of adjacent objects. The authors directly assign a constant brightness level to every object and do not scrutinize situations when objects are in front of each other or located very close. To further separate our work from this paper, we focus on small objects and everyday items that bionic implant users possibly want to find and locate inside the room or other indoor location.

3 METHODS

To provide better visual objects separation, we focus on adding visual highlighting information to the original picture changing original brightness of the objects. Initially (see subsection 3.2), we convert the original image to grayscale, then we use the saliency model to obtain saliency map information. After that, we run an instance segmentation model on the same original image to get instance segmentation masks of the predefined allowed list of classes). Given the masks, we construct a graph on a base of these masks where objects are vertices, and each edge corresponds to two objects being adjacent to each other. After creating the graph, we apply the graph coloring strategy to assign brightness to these objects and recolor them to separate the objects better (in the subsection 3.3). Given all these layers (saliency map, grayscale image, and colored segmentation results) in subsection 3.4 we combine them together to produce the final result to be provided to the user. Finally, in subsection 3.5 to evaluate the results, we processed our images with the pulse2percept (p2p) model and conducted a user study in order to measure the improvement.

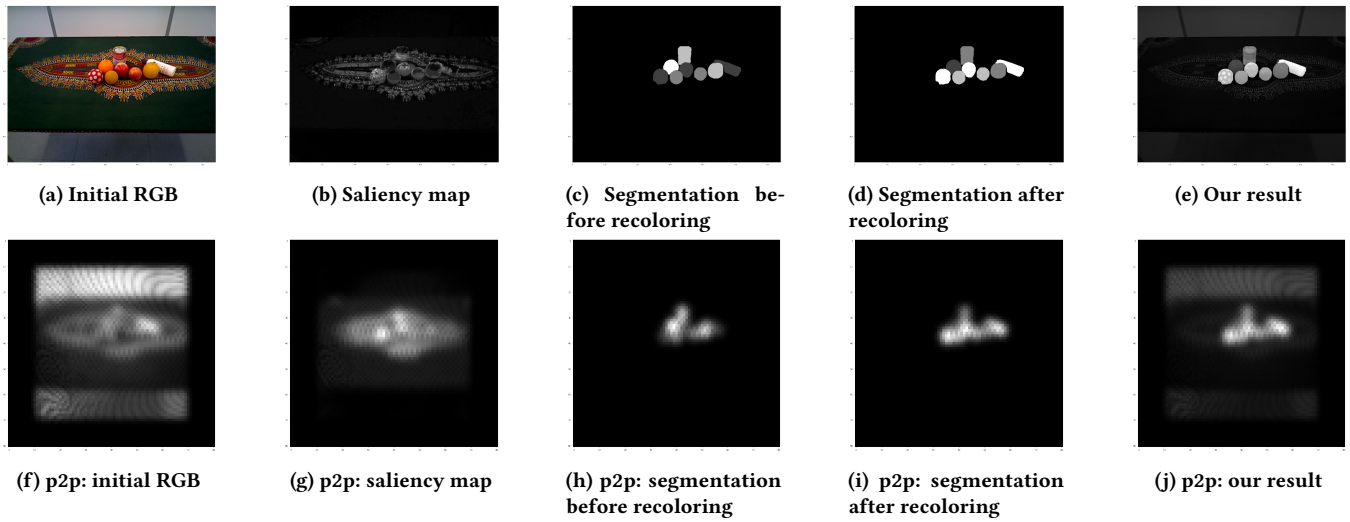
3.1 Dataset

We chose Object Clutter Indoor Dataset [17] to use as visual stimuli for our research. This dataset provides 96 different cluttered indoor scenes with different objects that you can find in the room. Each scene is located on either floor or the table and represents a typical scenario where the bionic implant user needs to find an object in a pile of similar objects or near some background. We chose a subset of pictures from the dataset to reduce calculations but tried to keep different backgrounds to make the research more representative. For each picture from the dataset, there exist RGB picture, point cloud, depth map, and label information, but we kept only RGB to represent the most popular RGB video cameras suited for implants. Initial RGB image and corresponding p2p visualization are provided on Figure 2a and Figure 2f.

3.2 Processing the image

As we design our solution for daily indoor scenarios, we need to keep original picture information and provide it to the user. Considering the current restrictions of bionic vision implants, we convert the image to grayscale using standard conversion weights. In addition to it, we extract saliency map information (see Figure 2b and Figure 2g for saliency map and p2p perception of saliency map correspondingly) from the original color image using the DeepGaze II model [8]. Saliency models predict how people look in images and provide valuable feedback on what places to highlight, as this would be interesting to pay attention to. Given that the original background would be dimmer than before, we believe it is important to highlight important places and structures besides the objects as this information can be critical.

In parallel to saliency reconstruction, we invoke a segmentation algorithm on the original color image. We use detectron2 model [20] with pretrained weights on LVIS dataset [4] named 'X101-FPN' in the detectron2 Model Zoo. The LVIS dataset is a dataset for large vocabulary instance segmentation that incorporates more than 1200 categories of objects, including typical indoor objects like fruits, pens, and others. This dataset is, to our knowledge, the best-suited dataset for indoor instance object segmentation with pre-trained models available in detectron2 Model Zoo. After segmentation, we filter the results and keep only masks of classes from the predefined allowed list of classes where we enlisted the most popular indoor objects categories available in the LVIS dataset. This list intentionally does not contain several classes (like table, carpet, and others) as these objects are usually not the target to be found by bionic vision users. In addition to filtering the classes, we also remove objects whose size is less than certain hyperparameters. Besides that, we also look for pair of objects which intersect heavily



(a) Initial RGB (b) Saliency map (c) Segmentation before recoloring (d) Segmentation after recoloring (e) Our result

(f) p2p: initial RGB (g) p2p: saliency map (h) p2p: segmentation before recoloring (i) p2p: segmentation after recoloring (j) p2p: our result

Figure 2: Stages of processing the picture and corresponding visualizations via pulse2percept.

and remove one of them. We calculate Intersection-Over-Union (IoU) metrics for that.

After segmentation masks are obtained, we construct a graph from these masks. As segmentation only provides us a list of objects to highlight, we need to properly assign them different brightness to make them easily visually separable from each other. To achieve this, we create an adjacency graph of these objects where each node corresponds to an object, and if two objects are close to each other on the image (so the distance between their closest points is less than the corresponding hyperparameter), then the edge between these objects is created. This graph later is subject to the graph coloring algorithm. This algorithm tries to assign different colors to adjacent nodes, which greatly fits our strategy of highlighting the adjacent objects differently to make them different. We use standard greedy algorithm [9] with largest_first strategy. After coloring is finished, the adjacent objects are assigned different color classes, and the whole graph (and therefore the whole image) uses the least possible amount of colors. As for the grayscale image only changes in brightness are available, reducing the overall amount of classes increases the brightness difference between the classes and makes objects of different color classes more distinguishable. Results of coloring and corresponding p2p perception are provided on Figure 2c and Figure 2h.

3.3 Recoloring

The main disadvantage of the graph coloring algorithm is a lack of ordering in coloring. For example, given object A with color class 1, for the coloring algorithm there is no difference in assigning classes 2 or 3 to the adjacent objects. However, as we want adjacent objects to be not only different but most possibly different, we want to assign the most different brightness values to adjacent objects. To achieve this, we use recoloring strategy on top of the resulting colored graph.

The primary purpose of recoloring strategy is to make static mapping between current color classes assigned to each node to new color classes where the difference between adjacent objects increases. This mapping is global for the whole graph, so if the strategy decides to change color class 1 to color class 5, this change would be applied to all color class 1 objects (so they become of color class 5), and no more color classes would be mapped to color class 5. As we plan to directly transform color classes to corresponding brightness levels, we value the situation where objects of color classes 1 and 5 are adjacent greater than where the same objects are of color classes 1 and 2. To reassign the labels, we created several different recoloring strategies.

The currently used recoloring strategy is based on top of the breadth-first search algorithm. It starts with a random node, assigns the biggest class to this node, and takes all neighbors whose classes are not reassigned yet. Next, all these neighbors classes are reassigned, so they differ the most from the current node and each other, and then the algorithm deletes assigned color classes from the list of available color classes and recursively starts from these nodes, and works until all classes are remapped.

We noticed that this algorithm, on average, produces better coloring results than the default coloring greedy strategy and other recoloring algorithms, but this statement is yet to be proved via user survey in future work.

After recoloring the graph, we provide these labels to the objects and transfer them to the single two-dimensional layer to be used later in combination with other layers. Results of recoloring and corresponding p2p perception are provided on Figure 2d and Figure 2i.

3.4 Combining the results

The resulting segmentation map is being rescaled to the [0.255] range to be treated as a valid greyscale image. This map alone theoretically could be passed to bionic implant users, but it would pose a risk of hiding the objects not being recognized but still crucial for the user.

To provide background information for the segmented objects, we would like to add more information on the recolored segmentation maps. Therefore, we tried several mixing methods and concluded that combining grayscale maps and recolored segmentation maps would bring the best result.

We first tried to combine the saliency map and the recolored segmentation map by modifying the method given in the previous paper [5], as the saliency maps can include the region of interest in the background. We first thresholded the saliency map to only retain 30% of the most salient region and then added them to the place where nothing is segmented in the recolored segmentation map. However, the combined map only adds small undistinguishable artifacts on the recolored segmentation map. When we transform it into simulated prosthetic images, it does not introduce significant changes to the simulated prosthetic images transformed from the recolored segmentation map.

We then tried to combine the grayscale map and the recolored segmentation map. The grayscale map would be x% percent of the combined image, and the recolored segmentation map would be (100-x%) percent of the image. We tried many x values and suggested that 30 would be the best value to include information in the grayscale map differentiated from the recolored segmentations. To make the grayscale map much clearer in low percentage in the combined map, we tried thresholding the grayscale map into binary. However, thresholding introduce significant changes in the shape of the background objects (like tables) that make the background hard to understand both the combined map and in simulated prosthetic vision images. We also tried adding contrast to the grayscale map, but the resulting prosthetic images are similar to the ones without adding contrast.

The final combination and corresponding p2p perception are provided on Figure 2e and Figure 2j. Additional examples of full processing pipeline could be found on Figure 5.

3.5 User study

To estimate the changes, we considered surveying objects' distinguishability on images as bionic implant users see them. The processed images from previous steps were then transformed into simulated prosthetic images using open-source library pulse2percept [3]. The images would be the input stimuli to the pulse2percept simulator. The simulator downscaled the processed images into the electrode array size and assigned each pixel in the processed image to an electrode. The grayscale value of the pixel determines the current on the electrodes. The size of the electrode array in our simulation is 32x32, which is a possible size of an electrode array in current or developing retinal implants, and shows good performance on identifying people and cars in the outdoors scene in the previous paper [5].

We used the axon map model in the pulse2percept library. The shape of the phosphene in the model is determined by parameters rho and axlambda, where rho is the exponential decay constant away from the axon and axlambda is the exponential decay constant along the axon. Since the actual phosphene shape varies among patients, we tried several possible pair of rho and axlambda and chose the one (rho = 70, axlambda = 30) that show the difference in coloring.

Table 1: Comparison of MAE and STDAE for original and modified pictures.

Mean Average Error (MAE) and standard deviation of average error (STDAE) for each picture.				
Image name	Original picture		Our picture	
	MAE	STDAE	MAE	STDAE
img0	6.685	1.039	5.218	1.166
img1	1.939	0.345	0.929	0.759
img2	5.491	0.663	5.397	0.493
img3	4.723	1.218	4.737	1.717
img4	5.115	0.858	3.628	0.723
img5	10.407	0.922	10.174	0.985
img6	10.647	1.219	8.937	1.693
img7	6.163	0.943	5.457	0.973
img8	2.746	0.659	1.906	0.904
img9	3.712	0.651	3.383	0.555
img10	7.029	0.170	5.676	0.692
img11	10.983	0.296	9.500	0.985
img12	4.966	0.417	3.972	0.985
img13	7.726	0.813	6.325	1.586
img14	3.738	0.603	3.000	1.008

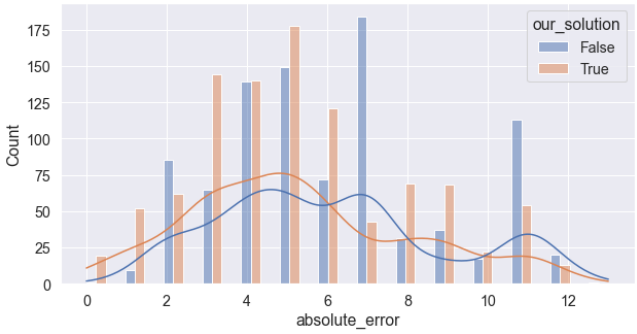


Figure 3: Combined histogram of mean absolute errors of objects estimation.

We created a subset of 15 randomly chosen images and applied our algorithm to produce versions of them with highlighted objects. Then we processed each image with the chosen pulse2percept model to represent a typical perception of these images by a bionic implant user. We then created a publicly available website where we asked people to estimate the number of objects on the given picture. We neither stated whether this picture was original or preprocessed nor described the research idea to the subjects.

4 RESULTS

We asked people to estimate the number of objects on randomly chosen 15 pictures. We didn't record the exact number of users, but observations over IP addresses of review submissions show that about 30 different persons participated in our review producing 1930 estimations over 30 different pictures. In Table 1 we collected mean

average error and standard deviation of error of users' estimations. For all pictures except img3, we noticed that error of predictions lowered, and users started to notice more objects on the picture. However, despite the stable decline, results don't change much in absolute values. We connect this with the dataset being rather clean in terms of additional objects and different backgrounds. Since the used dataset provides a simple background (as table, floor, or carpet), reviewers are not distracted by additional obstacles that would be less noticeable on processed images. Besides that, we think that due to surveyors being unaware of key changes in the picture and not being trained, collected results could be worse than for those who understand how highlighting would emphasize the objects.

On Figure 3 you can see the histogram of absolute errors combined for all pictures. This histogram also shows a slight improvement in recognizing the objects on the pictures, with a mean absolute error being shifted to zero compared with original images. The more detailed histograms of each picture are available in Figure 4 in papers appendices. We noticed that the histograms support our conclusions about the changes and show stable improvement for every image except img3.

The worse performance on img3 is closely connected with the used segmentation model. After investigating the reasons, we noticed that the segmentation model constantly highlights straight parallel lines as a possible book that leads to invalid estimations. We suggest that adjusting the model and its settings would provide better results due to the reduced number of false-positive segmented objects. In addition, we intentionally didn't finetune the model to the dataset in order to measure the performance in the wild, but in real life, such finetuning could be introduced for certain backgrounds, classes, or depending on person perception.

5 DISCUSSION

In this work, we combined different deep-learning models and graph theory to process pictures to facilitate objects finding, separation, and estimation for bionic implant users. We simulated bionic implant vision via open-source implementation of computational models of bionic vision, compared original pictures with their modified versions, and proved that our findings at worst do not complicate solving the tasks stated above but often facilitate it and improve human efficiency. We also made all the code and results available on https://github.com/maybe-hello-world/team_blue_291A.

5.1 Possible improvements

Despite being visually attractive, the results do not always introduce a significant change to the user's ability to solve the task. We want to mention two different possible problems to overcome to make the difference more noticeable:

(1) Segmentation model threshold attenuation. Several hyperparameters (like threshold and non-max suppression) need to be modified during the inference of the network. We found this step especially complicated during the research due to the lack of computing resources for that. Still, given the needed amount of time, the segmentation network could be adjusted to reduce the number of false-positive objects. For example, we had to exclude some classes from the segmentation process (like the 'toy' class) because they provided too many false positive masks. We believe that reducing the number of the model's errors would increase the solution's overall performance.
(2) Development of new recoloring strategies. For the paper, we implemented a relatively simple BFS-related recoloring strategy which does not always produce positive results according to our observations. We suggest providing a better recoloring approach would help to highlight the objects more qualitatively and facilitate their visual separation further.

5.2 Dataset for estimation

For the initial iteration of the research, we took the dataset with different objects stacked together to emphasize the possibility to visually differ them and estimate their total amount. To better demonstrate the method, we suggest testing this solution in the clumsy environment with different obstacles provided and objects not in the center of the frame. Initial LVIS dataset could be a good starting point for future work, though trying to spread the research on video stream estimation could require custom dataset collection.

5.3 Survey

Due to limited time and people to survey, we couldn't score all coloring and mixing strategies and collect feedback on a broader set of pictures. In the future, we suggest spending more time and resources on it to find the best possible strategy, better measure performance, and find situations where our solution works better.

5.4 Future work

For future work based on this research, we suggest overcoming the mentioned problems of our work and enhancing it to the degree of providing helpful assistance to bionic vision implant users. Stabilizing of segmentation and recoloring process together with network distillation to lower resource consumption would provide a better experience and can advance this research to the point of applicability for real devices. In addition to that, we see a potentially exciting improvement in implementing the audio interface for the segmentation network that would allow changing allowed classes for segmentation on the fly. That would provide an interesting experience of a user being able to highlight only needed objects and provide them theoretically better-than-human visual perception in certain situations.

REFERENCES

[1] Walid I. Al-Atabany, Tzyy Tong, and Patrick A. Degenaar. 2010. Improved content aware scene retargeting for retinitis pigmentosa patients. *BioMedical Engineering OnLine* 9, 1 (16 Sep 2010), 52. https://doi.org/10.1186/1475-925X-9-52
[2] Lauren N. Ayton, Nick Barnes, Gislin Dagnelie, Takashi Fujikado, Georges Goetz, Ralf Hornig, Bryan W. Jones, Mahiul M.K. Muqit, Daniel L. Rathbun, Katarina Stingl, James D. Weiland, and Matthew A. Petoe. 2020. An update on retinal prostheses. *Clinical Neurophysiology* 131, 6 (2020), 1383–1398. https://doi.org/10.1016/j.clinph.2019.11.029
[3] Michael Beyeler, Geoffrey M. Boynton, Ione Fine, and Ariel Rokem. 2017. pulse2percept: A Python-based simulation framework for bionic vision. *bioRxiv* (2017). https://doi.org/10.1101/148015 arXiv:https://www.biorxiv.org/content/early/2017/07/10/148015.full.pdf
[4] Agrim Gupta, Piotr Dollár, and Ross Girshick. 2019. LVIS: A Dataset for Large Vocabulary Instance Segmentation. arXiv:1908.03195 [cs.CV]

[5] Nicole Han, Sudhanshu Srivastava, Aiwen Xu, Devi Klein, and Michael Beyeler. 2021. Deep Learning–Based Scene Simplification for Bionic Vision. arXiv:2102.00297 [cs.CV]

[6] Lachlan Horne, Jose Alvarez, Chris McCarthy, Mathieu Salzmann, and Nick Barnes. 2016. Semantic labeling for prosthetic vision. *Computer Vision and Image Understanding* 149 (2016), 113–125. https://doi.org/10.1016/j.cviu.2016.02.015 Special issue on Assistive Computer Vision and Robotics - "Assistive Solutions for Mobility, Communication and HMI".

[7] Mark S Humayun, Jessy D Dorn, Ashish K Ahuja, Avi Caspi, Eugene Filley, Gislin Dagnelie, Joël Salzmann, Arturo Santos, Jacque Duncan, Saddek Mohand-Said, et al. 2009. Preliminary 6 month results from the argus tm ii epiretinal prosthesis feasibility study. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4566–4568.

[8] Matthias Kümmerer, Thomas S. A. Wallis, and Matthias Bethge. 2016. DeepGaze II: Reading fixations from deep features trained on object recognition. arXiv:1610.01563 [cs.CV]

[9] M. Kubale, Optymalizacja Dyskretna English, and American Mathematical Society. 2004. *Graph Colorings*. American Mathematical Society. https://books.google.ru/books?id=fokbCAAAQBAJ

[10] Heng Li, Tingting Han, Jing Wang, Zhuofan Lu, Xiaofei Cao, Yao Chen, Liming Li, Chuanqing Zhou, and Xinyu Chai. 2017. A real-time image optimization strategy based on global saliency detection for artificial retinal prostheses. *Information Sciences* 415-416 (2017), 1–18. https://doi.org/10.1016/j.ins.2017.06.014

[11] Heng Li, Xiaofan Su, Jing Wang, Han Kan, Tingting Han, Yajie Zeng, and Xinyu Chai. 2018. Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision. *Artificial Intelligence in Medicine* 84 (2018), 64–78. https://doi.org/10.1016/j.artmed.2017.11.001

[12] Chris McCarthy, Janine G Walker, Paulette Lieby, Adele Scott, and Nick Barnes. 2014. Mobility and low contrast trip hazard avoidance using augmented depth. *Journal of Neural Engineering* 12, 1 (nov 2014), 016003. https://doi.org/10.1088/1741-2560/12/1/016003

[13] N Parikh, L Itti, and J Weiland. 2010. Saliency-based image processing for retinal prostheses. *Journal of Neural Engineering* 7, 1 (jan 2010), 016006. https://doi.org/10.1088/1741-2560/7/1/016006

[14] Alejandro Perez-Yus, Jesus Bermudez-Cameo, Gonzalo Lopez-Nicolas, and Jose J. Guerrero. 2017. Depth and Motion Cues With Phosphene Patterns for Prosthetic Vision. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*.

[15] Melani Sanchez-Garcia., Ruben Martinez-Cantin., and Jose Guerrero. 2019. Indoor Scenes Understanding for Visual Prosthesis with Fully Convolutional Networks. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*. INSTICC, SciTePress, 218–225. https://doi.org/10.5220/0007257602180225

[16] Ashley Stacey, Yi Li, and Nick Barnes. 2011. A salient information processing system for bionic eye with application to obstacle avoidance. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 5116–5119. https://doi.org/10.1109/IEMBS.2011.6091267

[17] Markus Suchi, Timothy Patten, David Fischinger, and Markus Vincze. 2019. EasyLabel: A Semi-Automatic Pixel-wise Object Annotation Tool for Creating Robotic RGB-D Datasets. In *International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019*. IEEE, 6678–6684. https://doi.org/10.1109/ICRA.2019.8793917

[18] Jing Wang, Heng Li, Weizhen Fu, Yao Chen, Liming Li, Qing Lyu, Tingting Han, and Xinyu Chai. 2016. Image Processing Strategies Based on a Visual Saliency Model for Object Recognition Under Simulated Prosthetic Vision. *Artificial Organs* 40, 1 (2016), 94–100. https://doi.org/10.1111/aor.12498 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/aor.12498

[19] James D. Weiland, Neha Parikh, Vivek Pradeep, and Gerard Medioni. 2012. Smart image processing system for retinal prosthesis. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 300–303. https://doi.org/10.1109/EMBC.2012.6345928

[20] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

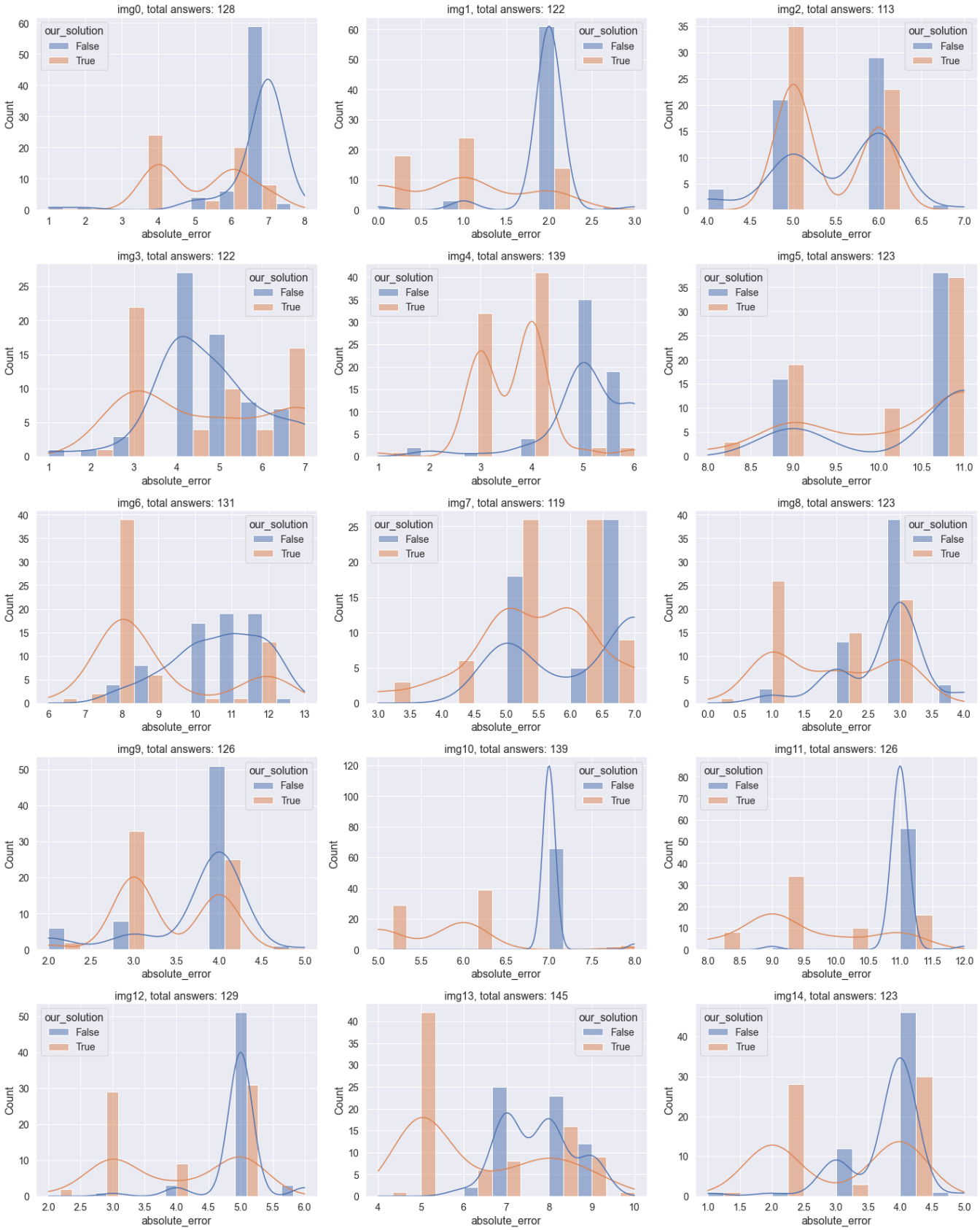


Figure 4: Histograms of errors for each picture from the test dataset

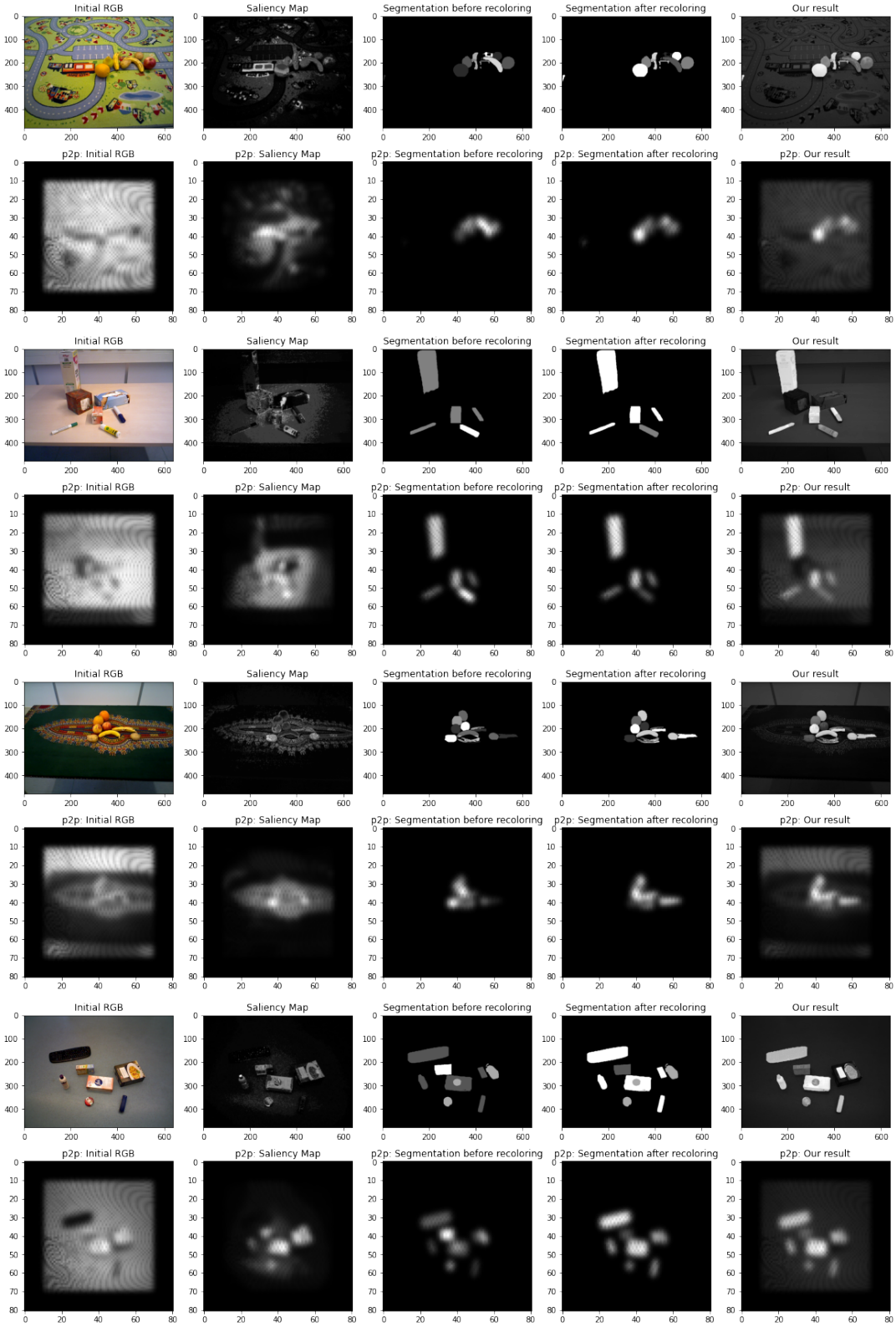


Figure 5: More examples of picture processing showing an improvement in objects detection.