**Combining grayscale and recolored segmentation**

To provide background information for the segmented objects, we would like to add more information on the recolored segmentation maps. We tried several mixing methods and concluded that combining gray-scale 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 [1], as the saliency maps can include the region of interest in the background. We first thresholded the saliency map so that it only retains 30% of the most salient region and then add 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 gray-scale 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 suggest that 30 would be the best value to include information in the grayscale map that can be differentiated from the recolored segmentations.

To make the grayscale map much clearer in low percentage in the combined map, we tried increasing the contrast of the grayscale maps or thresholding the grayscale map into binary. However, they both 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.

**Simulated Prosthetic vision:**

The processed images were then transformed into simulated prosthetic images using open-source library pulse2percept. The images would be the input stimuli to the pulse2percept simulator. The simulator downscaled the processed images into the electrode array size and assign each pixel in the processed image to an electrode. The current on the electrodes is determined by the gray-scale value of the pixel. 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 [1].

We used the axon map model in 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 [2], we tried several possible pair of rho and axlambda, and chose the one (rho = 70 , axlambda = 30) that show the difference in coloring.