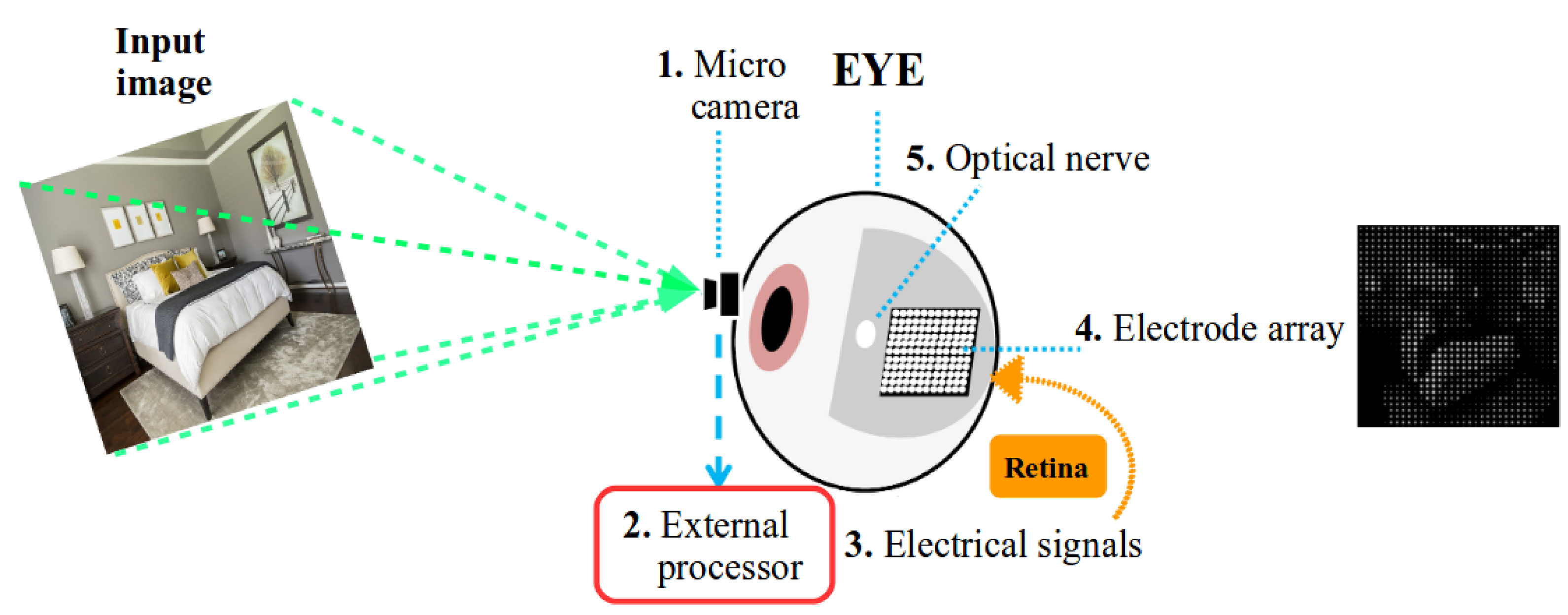
# Indoor Scenes Understanding for Visual Prosthesis with Fully Convolutional Networks

Instituto Universitario de Investigación en Ingeniería de Aragón Universidad Zaragoza [1]

Centro Universitario de la Defensa Zaragoza [2]

**Melani Sanchez-Garcia** [1], Ruben Martinez-Cantin [1,2] and Jose J. Guerrero [1]

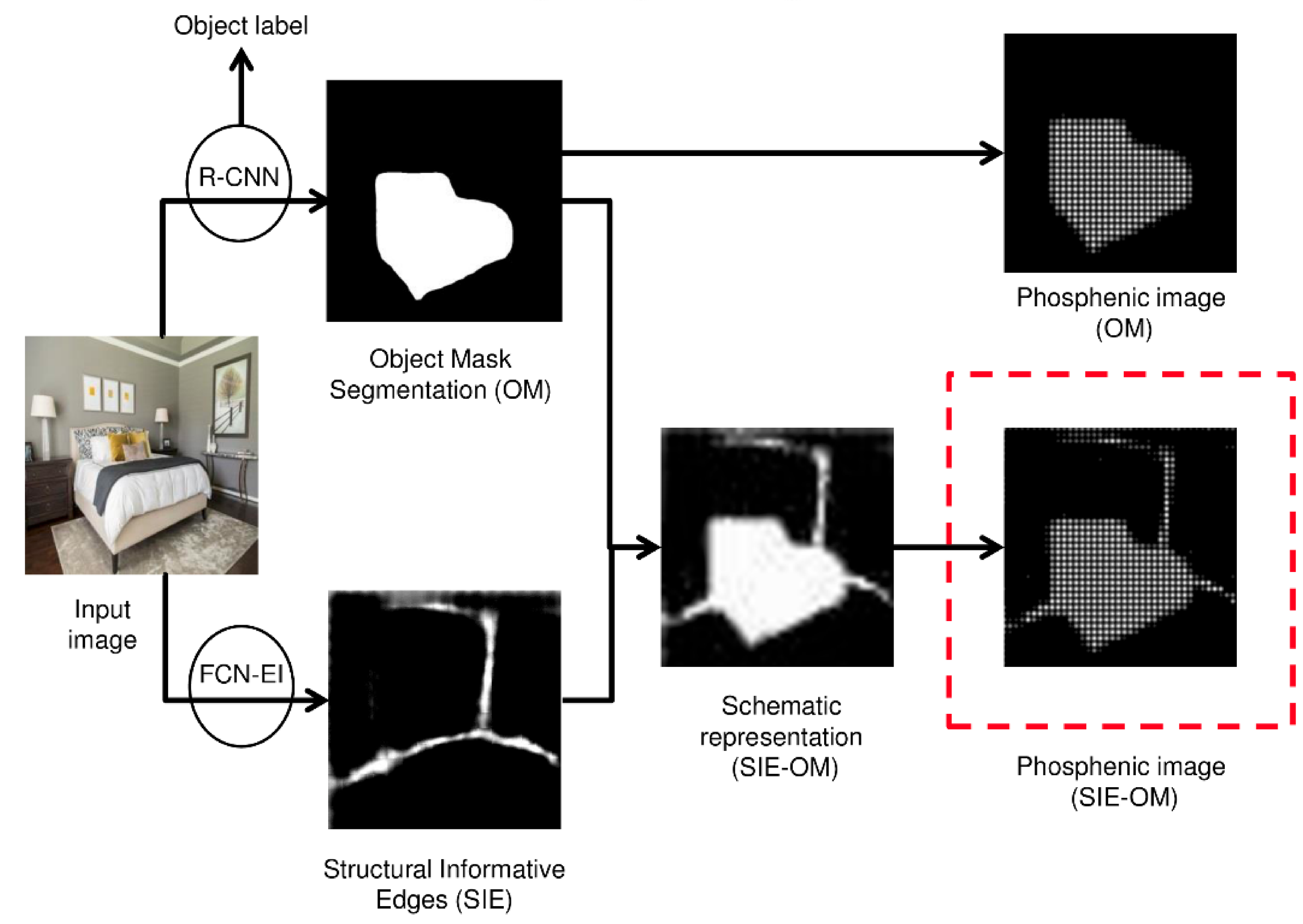
## Motivation




- Low resolution → ~ 100 pixels
- Restricted dynamic range → ~ 8 gray levels
- Limited visual field → 20º

Loss of visual information

## Overview




## Phosphene image generation using Deep Learning

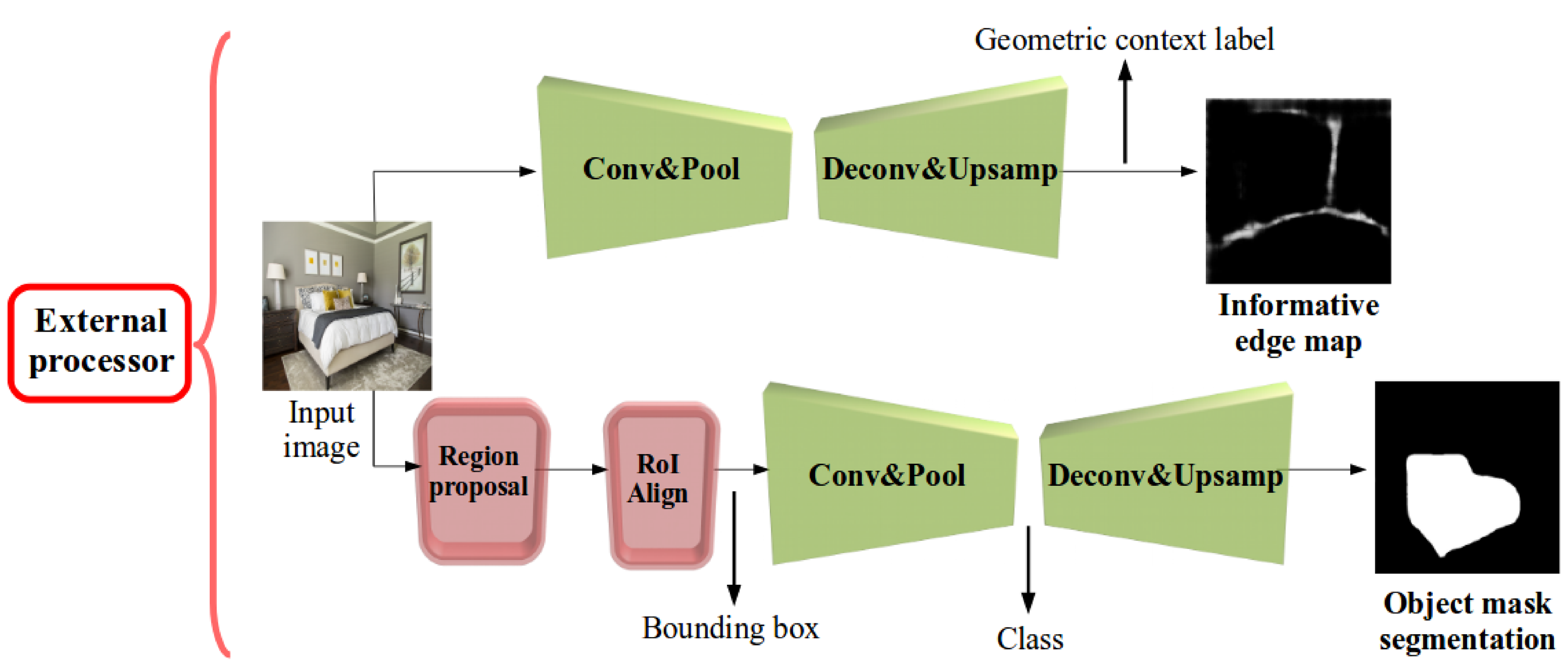
We propose a new phosphene image generation **"SIE-OM"** based on useful information of indoor environments ( structural informative edges (SIE) and relevant object masks (OM)) using **Convolutional Neural Networks (CNNs)**.

### SIE

For **Structural Informative Edges**, we use the framework of [1] which uses a Fully Convolutional Network (FCN) for pixel classification to estimate probability maps representing the room structural edges.

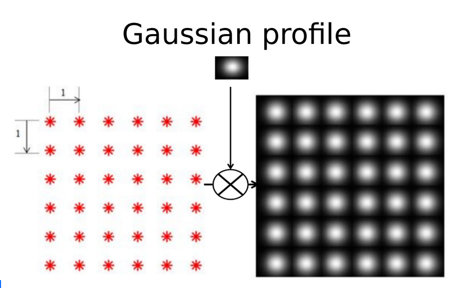
### OM

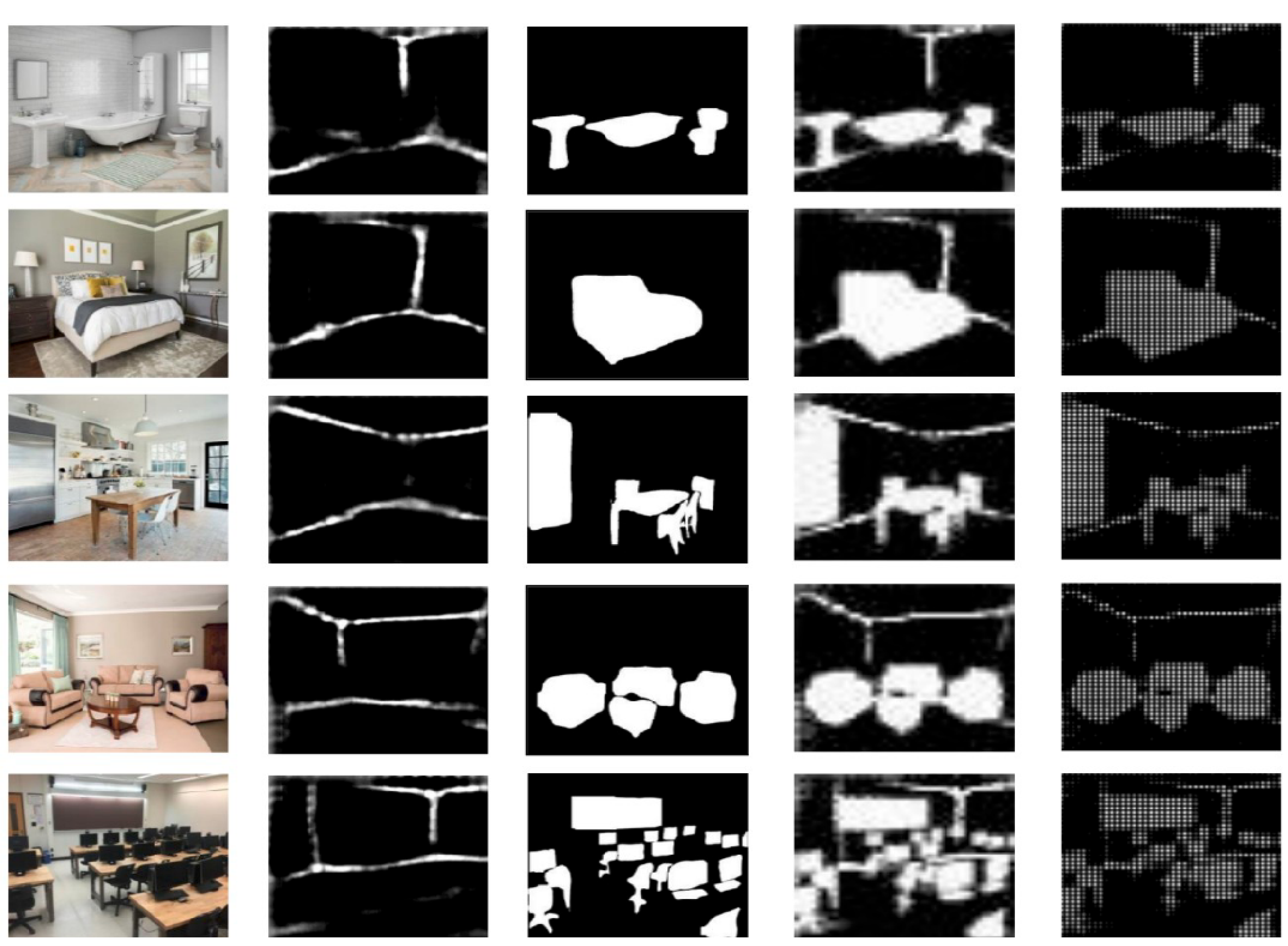
For **Object Masks**, we use the new approach of [2] which using selective search, it identifies a manageable number of bounding-box object region candidates (RoI). Then it extracts CNN features from each region independently for classification and extract a binary mask in pixel level of the object.




### Visual perception stimulation

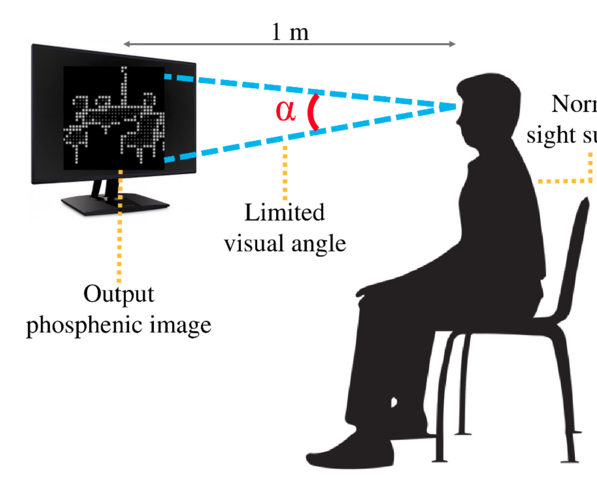
**Phosphenes** are idealized representations of the percepts feasible in the current implants. They are approximated as **grayscale circular dots** with a **Gaussian luminance profile**.



## Evaluation and results

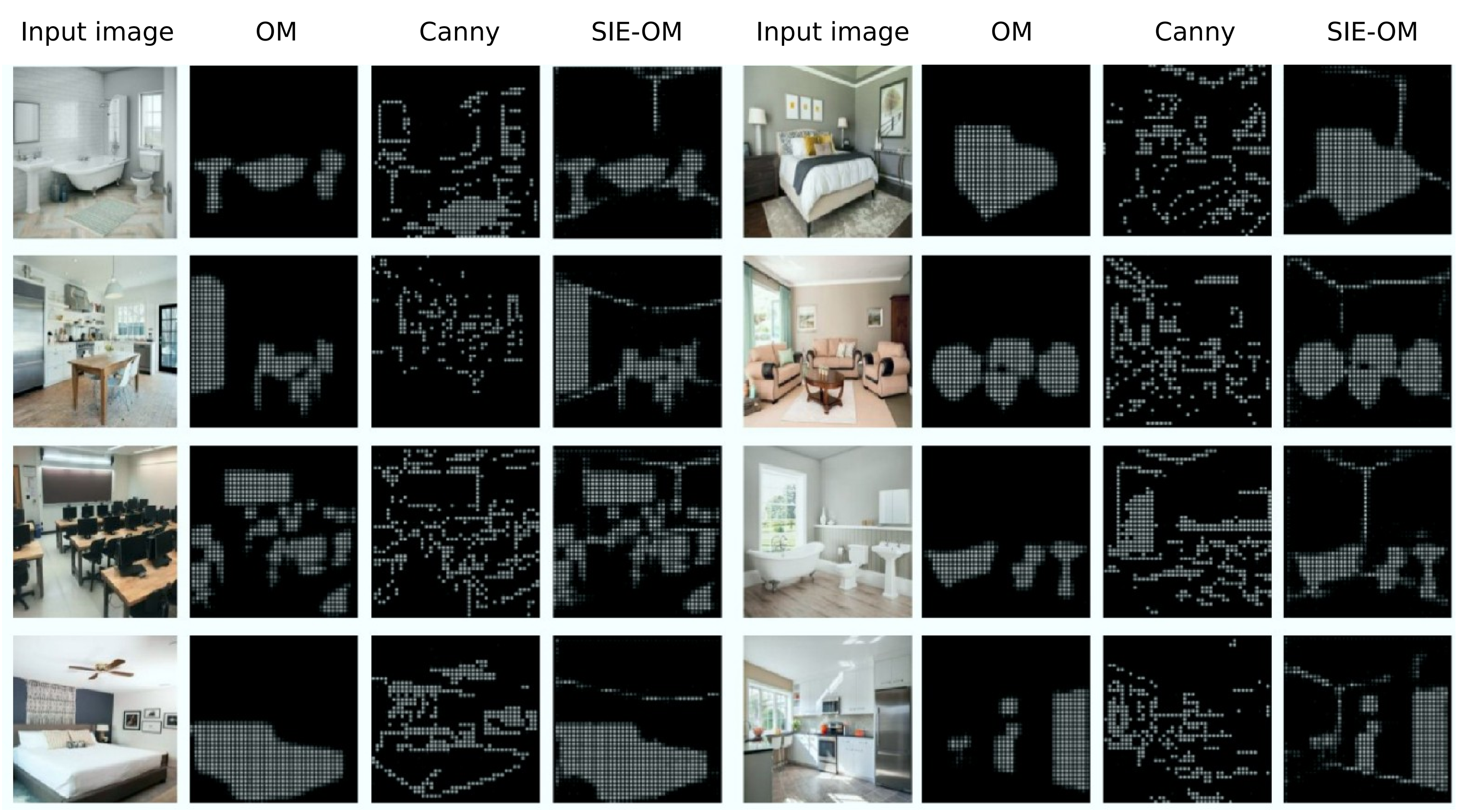
The experiment was carried out with twelve people with normal sight in **Simulated Prosthetic Vision (SPV)**.




We evaluated two tasks:

- **Object identification**
- **Indoor room recognition**

We compared our approach with OM and a baseline method used in SPV:




| Method | Object Identified | | Object Not Identified | | % Correct Object Identification | % Room Recognized | % Level of Confidence | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | %C | %I | %C | %I | | | DY | PY | M | PN | DN |
| Canny | 7 | 2 | 72 | 17 | 79 | 32 | 0 | 14 | 9 | 23 | 55 |
| OM | 13 | 5 | 74 | 10 | 87 | 41 | 9 | 32 | 23 | 23 | 14 |
| SIE-OM | 13 | 1 | 75 | 10 | 88 | 55 | 27 | 18 | 27 | 5 | 23 |

## Conclusions

- Deep learning algorithms can make better use of the limited resolution by highlighting salient features for SPV.

- **Best results** obtained with **SIE-OM** in the comprehension of the environment compared to OM and the baseline method.

- The structural informative edges become an interesting source of information of the scene providing sense of scale or perspective of the objects and depth.

## References

[1] Mallya, A. and Lazebnik, S. (2015). Learning informative edge maps for indoor scene layout prediction. In Proceedings of the IEEE International Conference on Computer Vision, pages 936–944.

[2] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In Computer Vision (ICCV), 2017 IEEE International Conference on, pages 2980–2988. IEEE.

### Acknowledgements

This work was supported by project DPI2015-65962-R (MINECO/FEDER, UE) and BES-2016-078426 (MINECO)