A Deep Learning–Driven Virtual Patient Framework for Predicting Functional Visual Prosthesis Performance

Jonathan Skaza
skaza@ucsb.edu

2025-06-05

The Team

Jonathan Skaza

Shravan Murlidaran

Apurv Varshney

Ziqi Wen

William Yang Wang

Miguel P Eckstein

Michael Beyeler

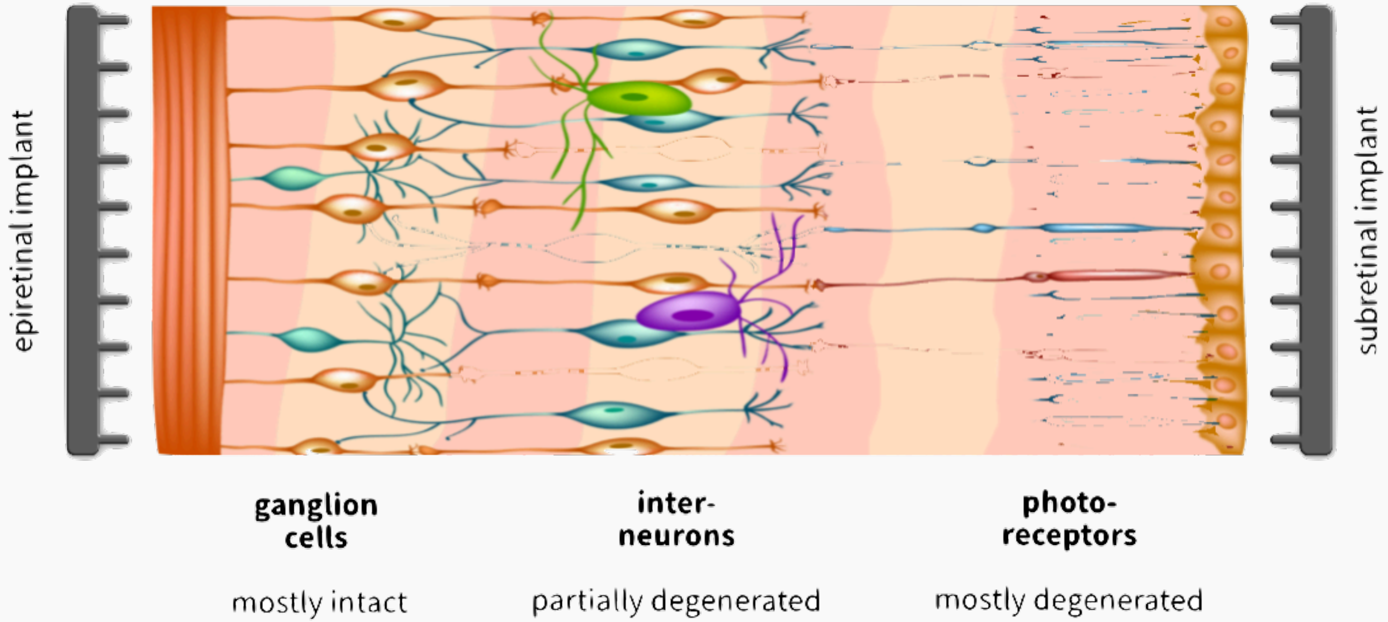
VIU
VISION AND IMAGE UNDERSTANDING LAB

UC Noyce Initiative

Motivation

- Many live with uncorrectable vision loss (e.g., retinitis pigmentosa, macular degeneration)
- Advanced therapies help early-stage disease but not late-stage
- Visual prostheses (“bionic eyes”) provide an alternative
  - epiretinal, subretinal, suprachoroidal, cortical
- Typically involve a camera that captures images, a processing unit (VPU), and an implant containing micro-electrodes that stimulate the remaining visual pathways in the eye or brain
- Evoke visual percepts (“phosphenes”)
- Outcomes often fall short (crude percepts) $\longrightarrow$ need for better prediction

Implant Sites

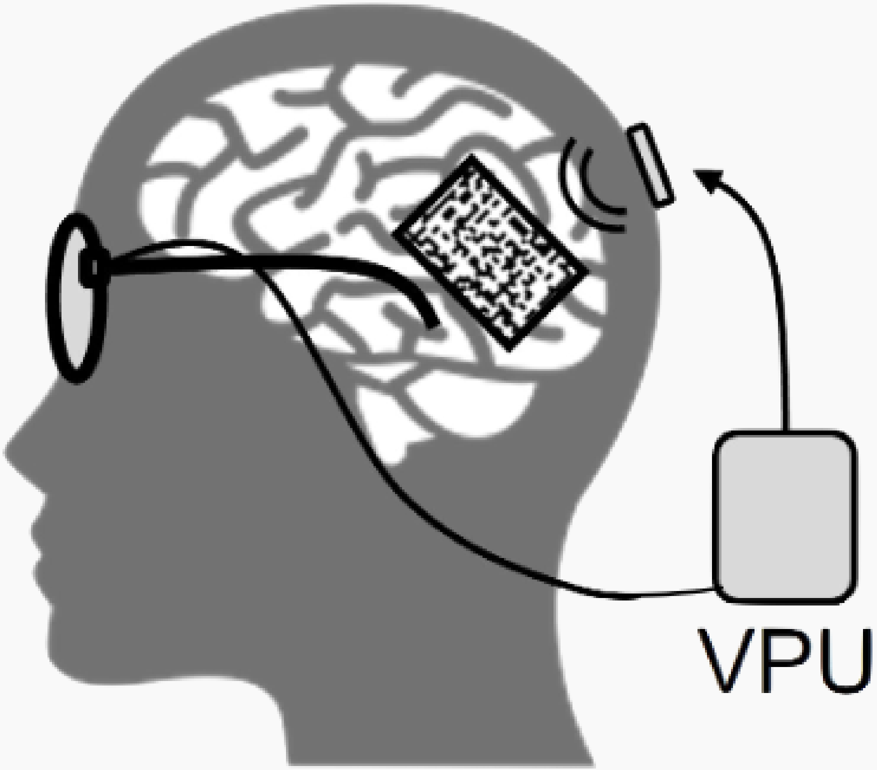


Epiretinal vs. Subretinal (Stanford Artificial Retina Project, 2025)

Visual Prosthesis User

Argus II user (Luo & Da Cruz, 2016)

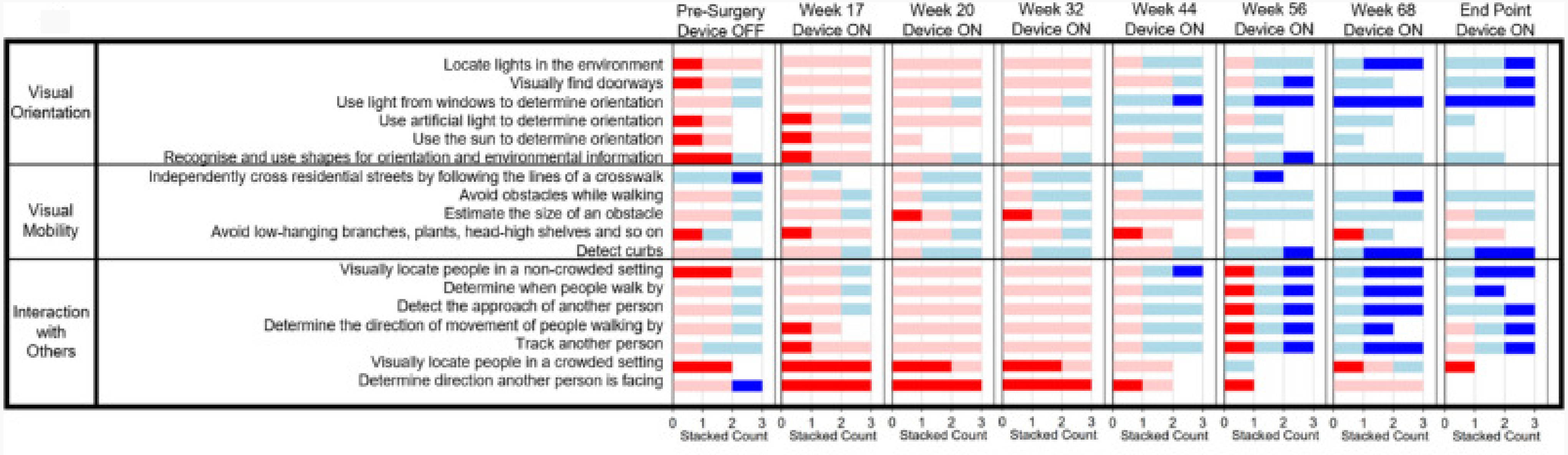
Visual Prosthesis



Functional Low-Vision Observer Rated Assessment (FLORA)

- FLORA evaluates real-world functional vision in ultra-low vision users of retinal prostheses (Geruschat et al., 2015)
- Components
  - Self-report interview
  - Observer-rated tasks
  - Narrative case summary
- Task difficulty (Impossible → Easy) and degree of vision used (None → Vision Only)

Functional Low-Vision Observer Rated Assessment (FLORA)



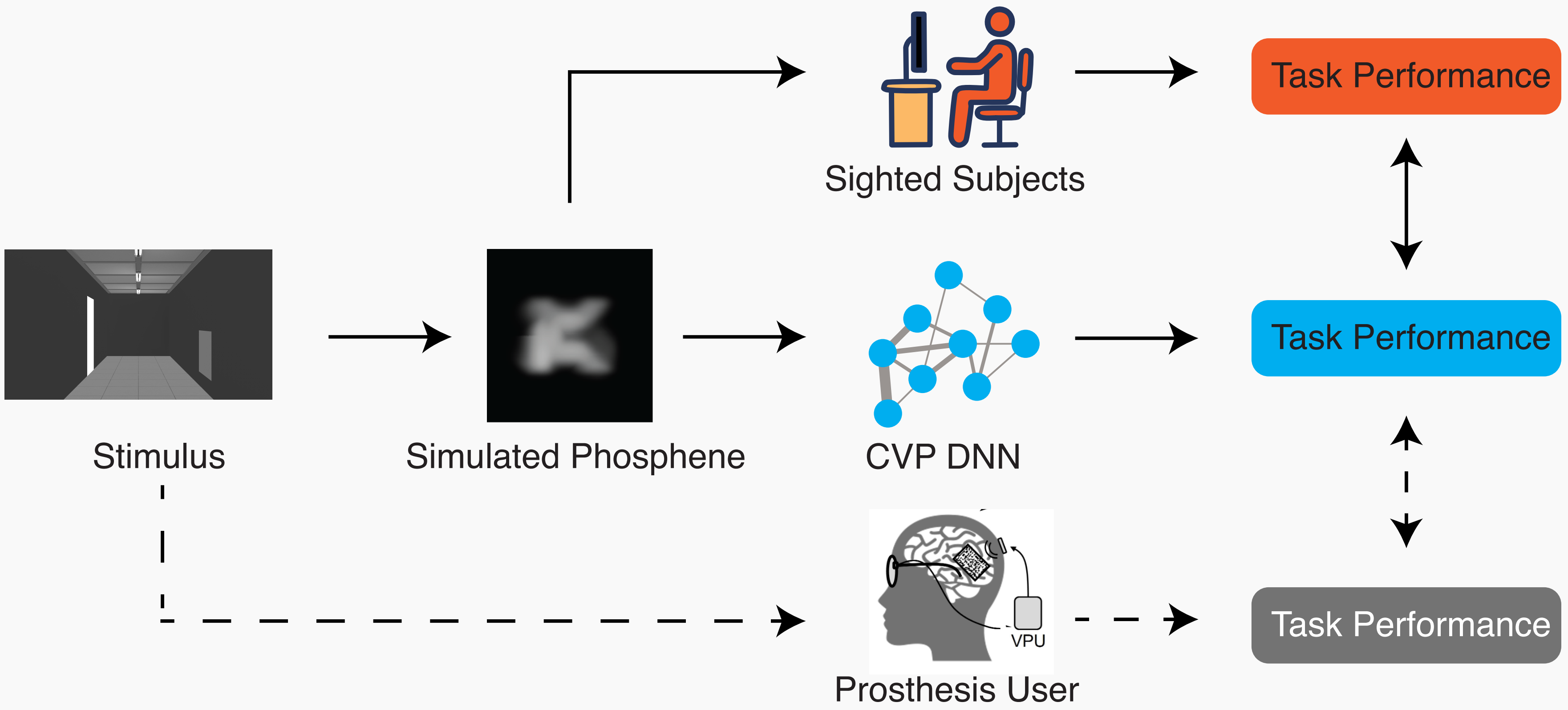
FLORA task difficulty progression (Karapanos et al., 2021)

- No scalable predictive tool for pre-implantation evaluation
- Currently, the most accurate information about a person's capabilities with prosthetic vision comes from experimental testing post-implantation

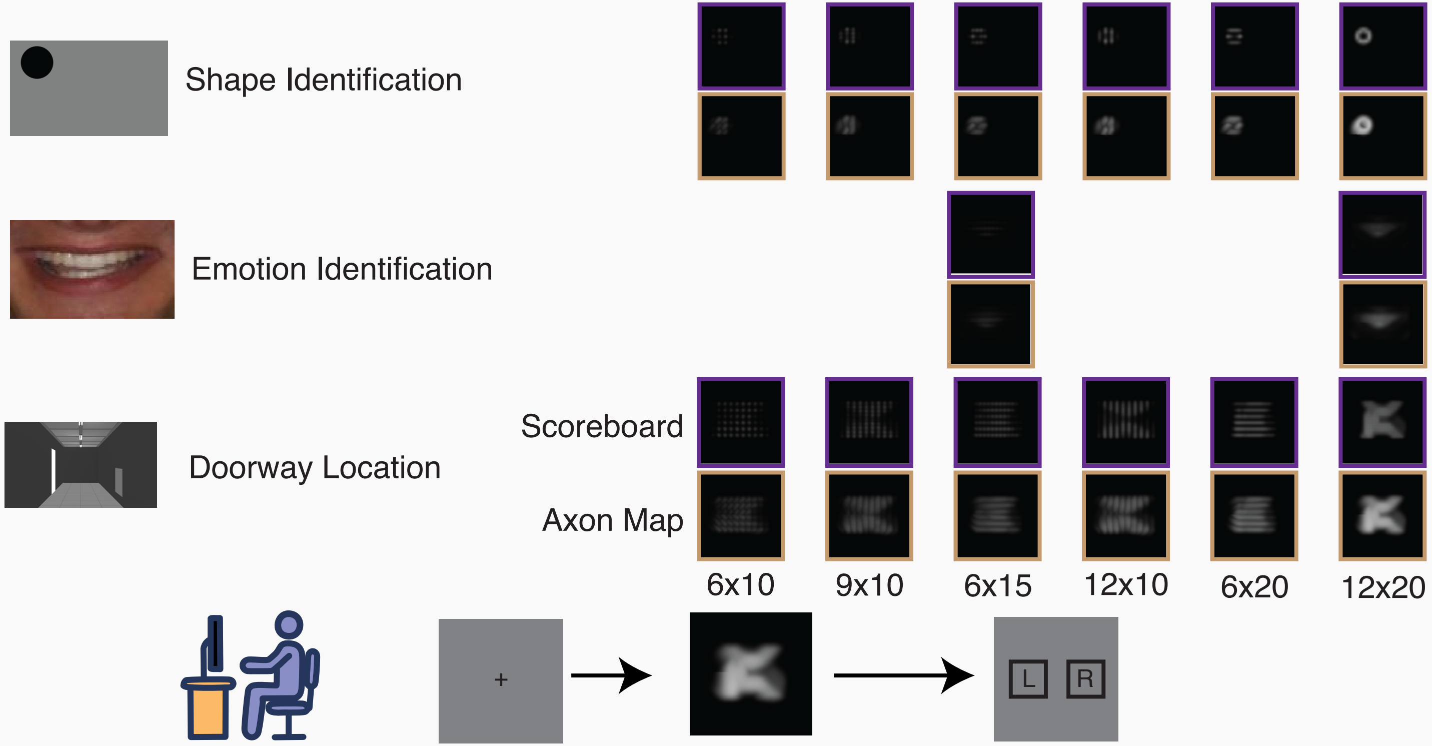
Current Methods to Simulate Phosphenes

- Scoreboard: each electrode produces a localized point of light (a “dot”) (Hayes et al., 2003; Thompson et al., 2003)
- Axon-map: incorporates the anatomical layout of retinal ganglion cell axon pathways, simulating perceptual distortions caused by current spread along axonal trajectories (Beyeler et al., 2019)
- See **pulse2percept** Python package (Beyeler et al., 2017)

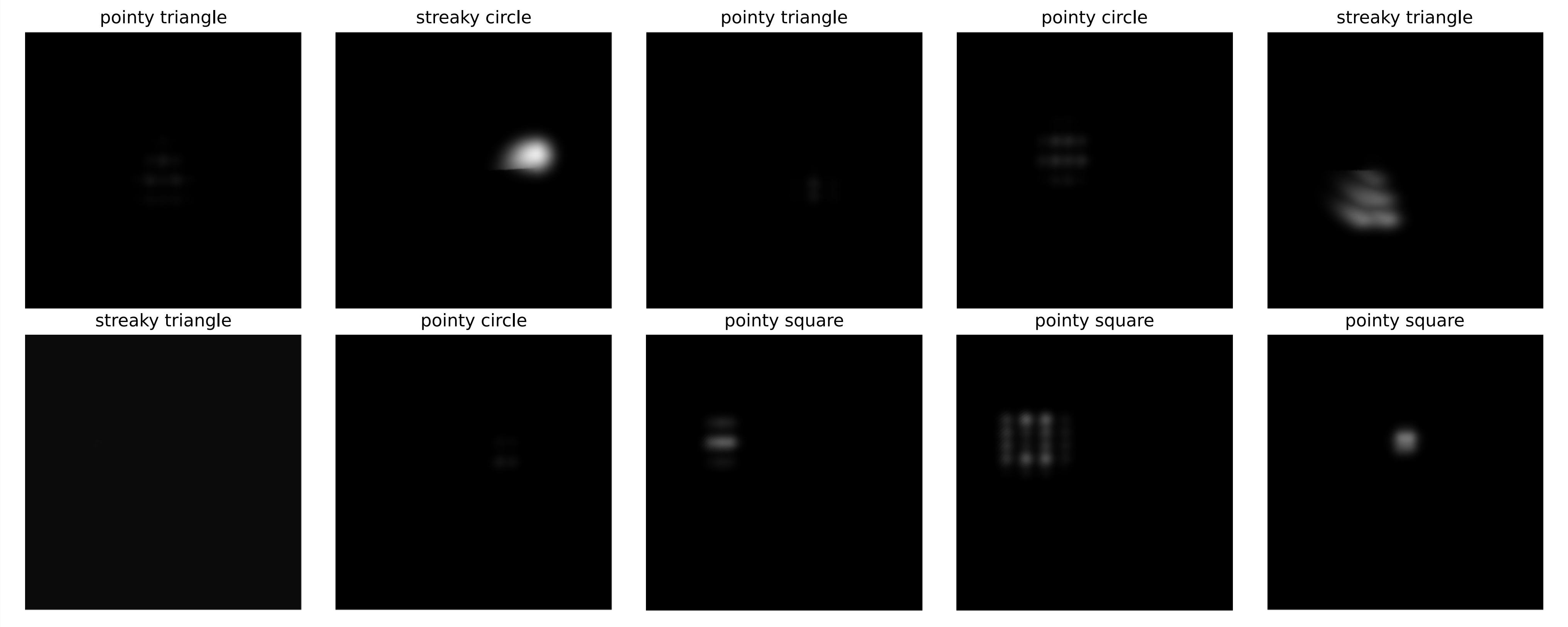
Computational Virtual Patient (CVP) Pipeline



- 12 normally-sighted subjects per condition, forced fixation, forced choice

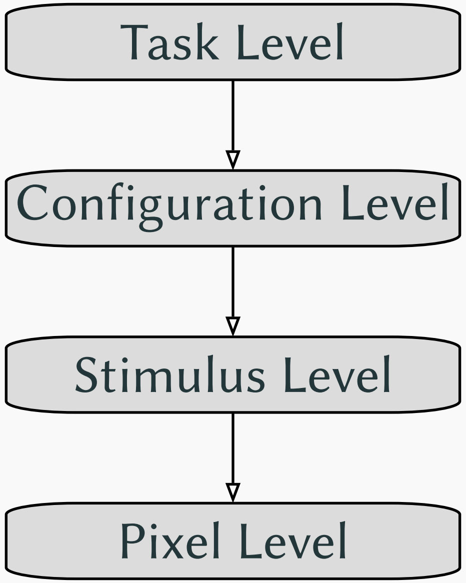


Example Shape Phosphenes

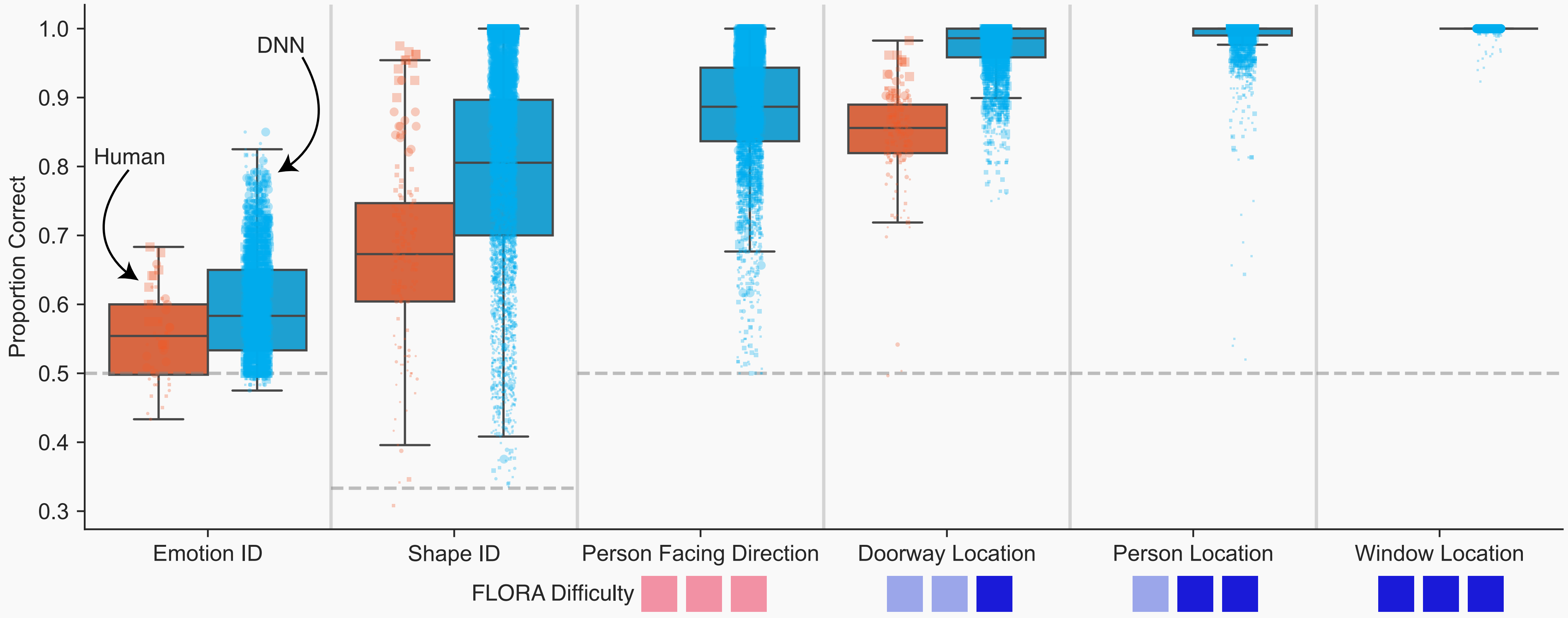


Human-DNN Alignment Evaluation

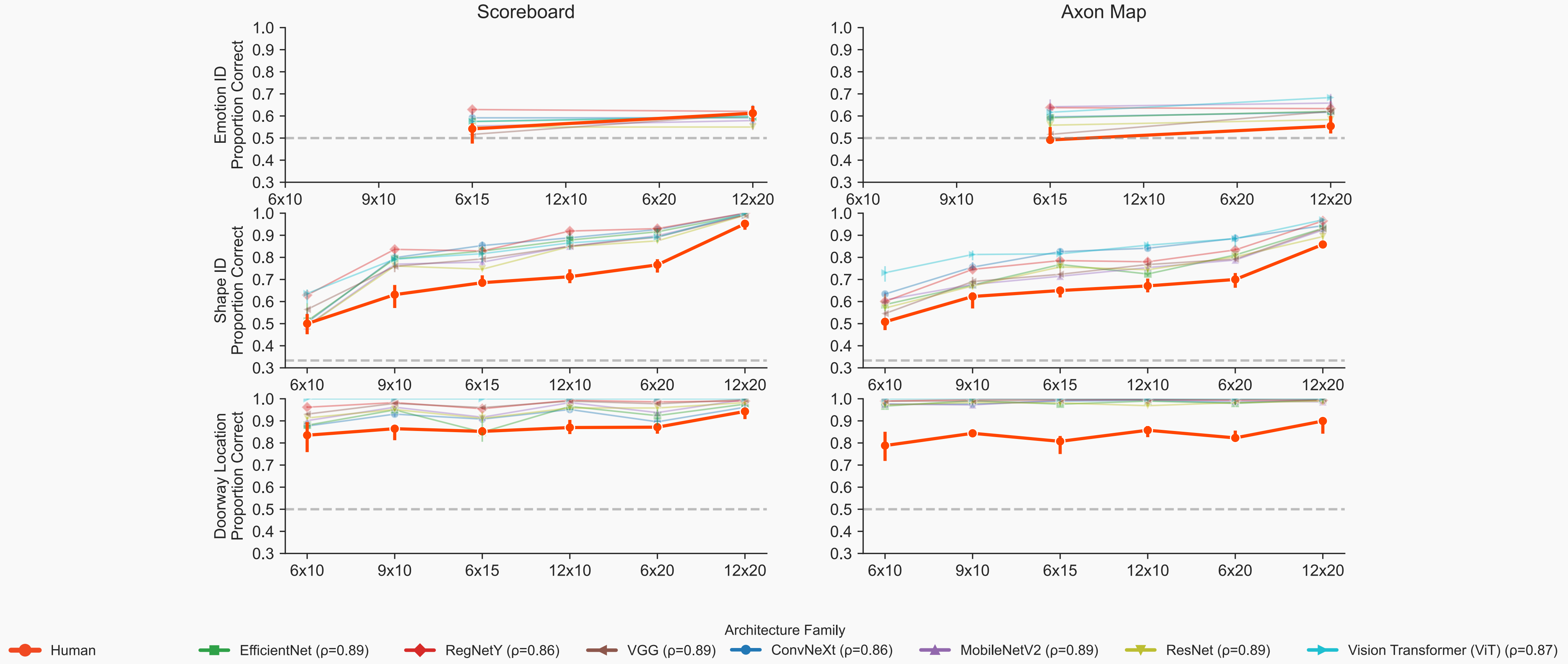
Comparison Focus	Key Question	Evaluation Metric
Task Difficulty	Do DNNs and humans find the same tasks difficult?	Proportion Correct
Configuration Sensitivity	Do DNNs and humans show similar responses to changes in implant configuration?	Spearman Correlation
Stimulus-Level Agreement	Do DNNs and humans make similar decisions on individual stimuli?	F1 Score, Jaccard Index
Attribution	Do DNNs and humans rely on similar features when making decisions?	Saliency Map Similarity



Task Difficulty ✓



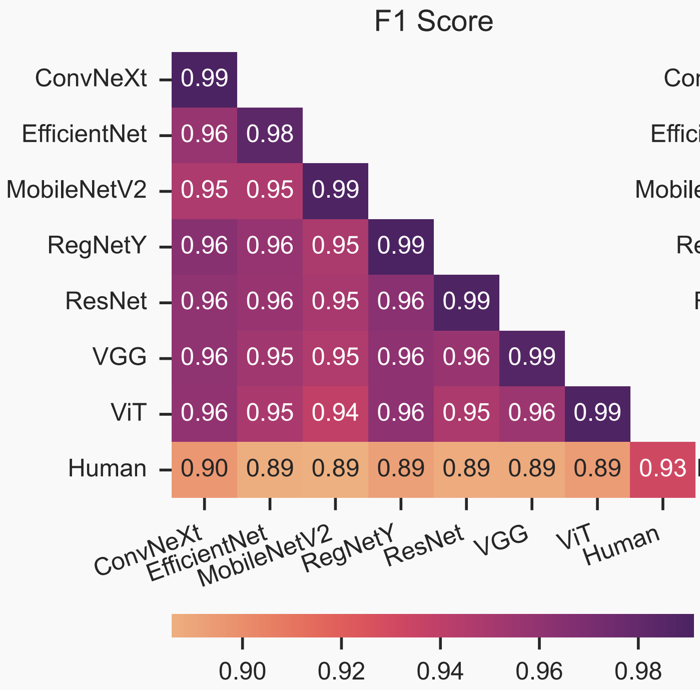
Configuration Sensitivity ✓

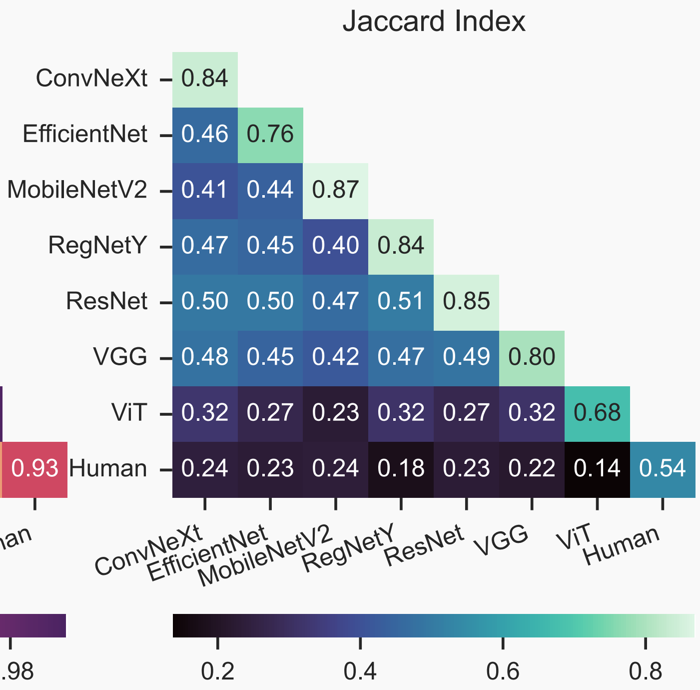


Stimulus-Level Agreement ❌

$$\text{F1} = \frac{2\,|\,\text{human}_{\text{correct}} \cap \text{model}_{\text{correct}}\,|}{|\,\text{human}_{\text{correct}}\,| + |\,\text{model}_{\text{correct}}\,|}$$

$$\text{Jaccard} = \frac{|\,\text{human}_{\text{incorrect}} \cap \text{model}_{\text{incorrect}}\,|}{|\,\text{human}_{\text{incorrect}} \cup \text{model}_{\text{incorrect}}\,|}$$



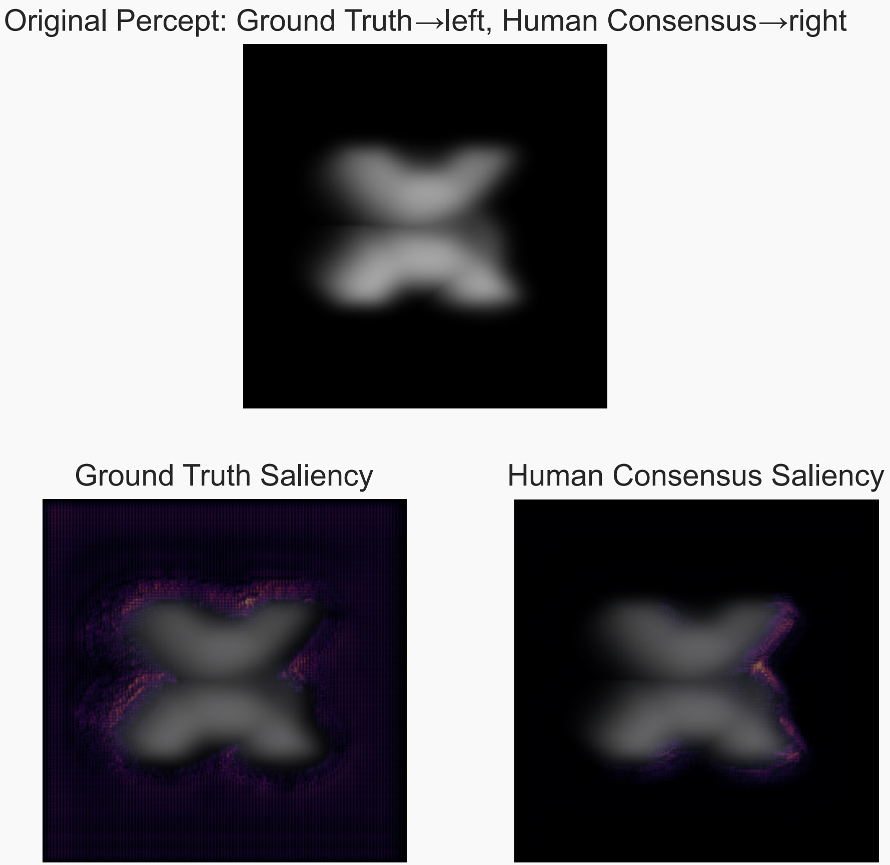


A Note on DNN Architectures

- ViT (Transformer-based models):
  - Weakest alignment with humans (Jaccard index = 0.14)
  - Rely more on global context and exhibit less sensitivity to spatially local cues due to weak locality and translation-invariance inductive biases (Naseer et al., 2021)
  - Global processing benefits natural images but is suboptimal for coarse, spatially structured phosphene patterns
- ConvNeXt (Convolutional architecture):
  - Strongest human alignment
  - Convolutional inductive biases (local spatial filtering, shared weights, translation invariance) mirror early human visual processing principles (Kubilius et al., 2021)
  - Robust feature hierarchies support coarse-to-fine representations similar to primate visual system (Yamins & DiCarlo, 2016)

Attribution Example

- Fit a CNN to predict human consensus, another to predict ground truth
- Compute saliency on held-out example (Simonyan et al., 2013)



Conclusions & Future Work

- Computational Virtual Patients (CVPs) can predict prosthetic-vision capabilities across multiple tasks and devices, offering a novel pre-implantation evaluation tool
- This framework could accelerate visual-prosthesis development and set more accurate expectations for implant recipients
- Attribution methods enable insight into decision strategies and error patterns under prosthetic simulation
- Fully trained DNNs did not align with human performance, suggesting that users rely on pre-existing visual processing rather than forming new perceptual mappings
- Future work: more tasks, integrate video-based DNN models and VR simulations for dynamic task assessment, different types of implants

Send me an email if interested — preprint available soon!

References

Beyeler, M., Boynton, G. M., Fine, I., & Rokem, A. (2017). pulse2percept: A Python-based simulation framework for bionic vision. *Proceedings of the 16th Python in Science Conference (Scipy)*, 81–88.

Beyeler, M., Nanduri, D., Weiland, J. D., Rokem, A., Boynton, G. M., & Fine, I. (2019). A model of ganglion axon pathways accounts for percepts elicited by retinal implants. *Scientific Reports*, *9*(1), 9199.

Geruschat, D. R., Flax, M., Tanna, N., Bianchi, M., Fisher, A., Goldschmidt, M., Fisher, L., Dagnelie, G., Deremeik, J., Smith, A., & others. (2015). FLORA™: Phase I development of a functional vision assessment for prosthetic vision users. *Clinical and Experimental Optometry*, *98*(4), 342–347.

Hayes, J. S., Yin, V. T., Piyathaisere, D., Weiland, J. D., Humayun, M. S., & Dagnelie, G. (2003). Visually guided performance of simple tasks using simulated prosthetic vision. *Artificial Organs*, *27*(11), 1016–1028.

References

Karapanos, L., Abbott, C. J., Ayton, L. N., Kolic, M., McGuinness, M. B., Baglin, E. K., Titchener, S. A., Kvansakul, J., Johnson, D., Kentler, W. G., & others. (2021). Functional vision in the real-world environment with a second-generation (44-channel) suprachoroidal retinal prosthesis. *Translational Vision Science & Technology*, *10*(10), 7.

Kubilius, J., Schrimpf, M., Nayebi, A., & al. (2021). Brain-like object recognition with high-performing shallow recurrent ANNs. *Advances in Neural Information Processing Systems*, *34*, 12885–12900.

Luo, Y. H.-L., & Da Cruz, L. (2016). The Argus® II retinal prosthesis system. *Progress in Retinal and Eye Research*, *50*, 89–107.

Naseer, M., Ranasinghe, K., Khan, S., & al. (2021,). Intriguing Properties of Vision Transformers. *Neurips*.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *Arxiv Preprint Arxiv:1312.6034*.

References

Stanford Artificial Retina Project. (2025,). *The Stanford Artificial Retina Project*. `https://med.stanford.edu/artificial-retina.html`

Thompson, R. W., Barnett, G. D., Humayun, M. S., & Dagnelie, G. (2003). Facial recognition using simulated prosthetic pixelized vision. *Investigative Ophthalmology & Visual Science*, *44*(11), 5035–5042.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365.