SCAPE: Shift-variant Cortical-implant Adaptive Phosphene Encoding

Michelle Appel*, Antonio Lozano†, Eleftherios Papadopoulos*, Umut Güçlü*, Yağmur Güçlütürk*,
*Donders Institute for Brain, Cognition and Behaviour,
Radboud University, Nijmegen, The Netherlands
†Institute of Bioengineering, Universidad Miguel Hernández de Elche, Spain
Email: michelle.appel@ru.nl

***Abstract*—Visual cortical implants aim to restore a form of sight by electrically stimulating neurons with electrode arrays. Each of the implant's contact point potentially generates a visual percept or "phosphene", in a specific location of the visual field, according to the visual area's retinotopy.**

To convey useful visual information to the implant's users, conventional computer vision image encoding methods typically apply uniform filters across the entire visual field, ignoring the uneven phosphene map distribution imposed by implant layouts. This mismatch can oversmooth detail in dense regions or introduce clutter where coverage is sparse.

To overcome this limitation, we present SCAPE (Shift-variant Cortical-implant Adaptive Phosphene Encoding), a framework that adapts image processing to local electrode density. Electrode cortical locations are projected into visual-field space to estimate sampling density via cortical magnification models or kernel density estimation. Nyquist principles then convert phosphene density into a spatial scale map, guiding shift-variant filtering whose kernel width matches local resolution limits.

We demonstrate an efficient separable Difference-of-Gaussians implementation, though SCAPE generalizes to other kernels. Integrated with a reconstruction decoder, SCAPE consistently preserves structural detail of the generated phosphene images and improves reconstruction quality across diverse datasets and implant schemes. By explicitly linking electrode layout to adaptive spatial filtering, SCAPE provides a computationally lightweight and principled foundation for enhancing prosthetic vision and accelerating translation toward clinical use.

I. Introduction

Visual cortical prostheses offer a promising path to restore vision in individuals with severe visual impairment by directly stimulating populations of neurons in visual cortex. These devices, such as the Utah array [1]–[3] and polyimide-based, ultraflexible cortical implants such as those developed by Neuralink and others [4]–[6], aim to bypass damaged retinal pathways and directly encode visual information into the brain through early cortical sites [7]. By stimulating cortical neurons in a spatially organized manner, these implants can evoke perceptual phosphenes that correspond to visual stimuli according to their retinotopy [8].

However, a central challenge in designing effective cortical prostheses is the limited number of stimulating electrodes, which fundamentally constrains the spatial resolution of visual information. This bottleneck necessitates careful consideration of how visual stimuli are processed and represented before delivery to the implant [9]–[12].

Current approaches to visual encoding for cortical implants often rely on uniform spatial filtering techniques, such as Sobel or Canny edge detection, to reduce the dimensionality of visual input. While these methods can help manage the high spatial resolution of natural images, they neglect implant-specific sampling density. This mismatch can produce oversmoothing in regions where electrode coverage is high, reducing available detail, or introduce spurious structure in sparse regions that cannot be faithfully represented by the implant [13]–[15].

In this work we introduce SCAPE (Shift variant Cortical prosthesis Adaptive Phosphene Encoding), an adaptive encoding framework that tailors spatial filtering to the local resolvability of each implant configuration. SCAPE first estimates the sampling density from electrode or phosphene positions using analytic magnification models or kernel density estimation. It then maps density to a local spatial scale via Nyquist principles and applies shift variant filtering, implemented here as a difference of Gaussians, to the input image before phosphene rendering.

The main contributions of this paper are:

- A principled method for local sampling density estimation and shift variant spatial filtering tailored to cortical implant layouts.
- Comprehensive evaluation of SCAPE in simulation across multiple electrode configurations, including high density, Utah array, Neuralink-type shanks, and receptive field-based schemes.
- Benchmarking performance with representational similarity analysis, and reconstruction accuracy of an Attention UNet decoder.

II. Related Work

Encoding strategies for prosthetic vision have evolved from simple heuristic preprocessing to learned encoders and patient-specific pipelines. However, most prior methods treat the visual field uniformly, overlooking the spatial inhomogeneity imposed by cortical magnification and implant geometry. Below we review these approaches and highlight the need for adaptive methods such as SCAPE.

A. Image Processing for Visual Prostheses

Early and recent encoders have focused on extracting salient features or learning mappings to electrode activations, but

they rarely incorporate the sampling constraints of individual implants. Ignoring these constraints risks oversmoothing detail in dense regions or overwhelming sparse regions with clutter. SCAPE addresses this gap by explicitly linking electrode density to filter scale.

1) Early Heuristic Methods: Under severe electrode count and bandwidth limits, early pipelines enhanced contrast, equalized histograms, or applied gradient-based edge detectors such as Sobel and Canny to emphasize contours in low-resolution phosphene maps [16], [17]. Outputs were often binarized or morphologically filtered to reduce noise and highlight obstacles or object boundaries. These methods were computationally efficient but processed the field uniformly, with no adjustment to electrode layout or cortical magnification. This uniformity limited their ability to balance detail and clutter.

2) More Recent Learned Encodings: Later approaches framed encoding as an optimization problem. Relic et al. trained convolutional networks end-to-end with a differentiable phosphene simulator on MNIST [15]. Granley et al. proposed hybrid neural autoencoders that invert neuroscientific forward models to produce patient-specific stimulation patterns [18]. De Ruyter van Steveninck et al. extended this to jointly optimize spatial filtering and electrode selection in an end-to-end learned encoder [19]. These methods improved fidelity over heuristics but still did not explicitly modulate spatial scale based on local sampling density or magnification, leaving a gap that SCAPE fills.

B. Simulated Prosthetic Vision Pipelines

Simulation frameworks have been crucial for both developing and benchmarking encoders. Early work rendered static Gaussian phosphenes without retinotopy or magnification due to limited tools. Later immersive VR-SPV systems allowed user studies but still relied on non-differentiable heuristics and uniform phosphene shapes [13]. Recently, van der Grinten et al. released a differentiable simulator in PyTorch that integrates retinotopic projection, cortical magnification, current spread, and temporal dynamics [10]. This enables gradient-based optimization of encoders in realistic implant layouts. However, even with such simulators, most encoding methods still apply uniform processing, rather than adapting filter scale to local density.

Simulation frameworks have been crucial for both developing and benchmarking encoders. Early work rendered static Gaussian phosphenes without retinotopy or magnification due to limited tools. Later immersive VR-SPV systems allowed user studies but still relied on non-differentiable heuristics and uniform phosphene shapes [13]. Recently, van der Grinten et al. released a differentiable simulator in PyTorch that integrates retinotopic projection, cortical magnification, current spread, and temporal dynamics [10]. This enables gradient-based optimization of encoders in realistic implant layouts. In a similar state-of-the-art effort, Fine and Boynton introduced a virtual patient model [20] which allows to explore the physiological constraints of targeting V1 to generate phosphene percepts, and offers insights into both the limitations and opportunities of encoding visual information through phosphene vision. However, even with such simulators, most encoding methods still apply uniform processing, rather than adapting filter scale to the phosphene map's local density.

C. Adaptive and Spatially-Varying Encoding Approaches

Recent methods aim for patient-specific adaptation. Granley et al. introduced a human-in-the-loop Bayesian optimization framework to tune deep encoders to subjective feedback [21]. Hybrid autoencoders have inverted biophysical forward models to generate customized stimulation patterns [18], [19]. These approaches personalize encoders but do so globally, without explicitly varying processing scale across the visual field. To our knowledge, no prior method adapts filtering continuously to local electrode density or Nyquist limits. SCAPE directly addresses this unmet need by coupling density estimation with shift-variant filtering.

III. Methods

A. Phosphene Simulation Framework

To assess SCAPE under conditions that approximate clinical prosthetic vision we employ the simulator of van der Grinten et al. [10]. This pipeline models key aspects of cortical stimulation, including the retinotopic projection of electrodes into visual-field coordinates and the rendering of each electrode's percept as a Gaussian phosphene. The resulting phosphene image captures the spatial layout of perceptual activations and serves as the input for SCAPE's adaptive encoding stages.

1) Electrode Placement: We begin with E electrodes implanted in cortical tissue, each with a two-dimensional coordinate on the cortical surface:

$$\{(x_i, y_i)\}_{i=1}^{E}.$$

These positions are obtained from clinical implant schematics or patient-specific models.

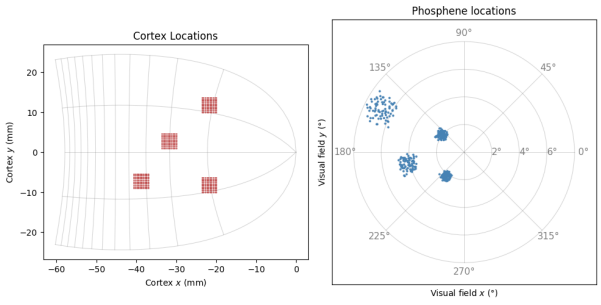

Fig. 1: Cortical electrode locations of 4 Utah Arrays (left) plotted in cortical x, y coordinates with overlaid retinotopic grid, and corresponding phosphene centers (right) in visual-field polar coordinates.

2) Retinotopic Projection and Phosphene Centers: Each cortical electrode (x_i, y_i) is mapped into visual-field coordinates $(\mu_{x,i}, \mu_{y,i})$ using the inverse of the Wedge–Dipole transform introduced by Polimeni *et al.* [22]. This model captures cortical magnification, in which the foveal region is allocated a larger cortical representation than the periphery. In

complex notation the forward mapping from visual-field polar coordinates (r, θ) to cortical coordinate w is

$$w = k\big[\ln\big(r\,e^{i\alpha\theta} + a\big) - \ln\big(r\,e^{i\alpha\theta} + b\big)\big],$$

where r is eccentricity in degrees, θ is polar angle, k scales degrees to cortical millimeters, and a, b, α are fitted parameters. Analytically inverting this relation yields

$$r\,e^{i\alpha\theta} = \frac{b\,e^{\phi} - a}{1 - e^{\phi}}, \quad \phi = \frac{w}{k},$$

from which

$$r = \left|\tfrac{b\,e^{\phi}-a}{1-e^{\phi}}\right|, \quad \theta = \frac{1}{\alpha}\arg\!\big(\tfrac{b\,e^{\phi}-a}{1-e^{\phi}}\big).$$

Applying this inverse transform to each (x_i, y_i) and adding optional Gaussian angular noise and dropout produces $N \le E$ phosphene centers

$$\{(\mu_{x,i}, \mu_{y,i})\}_{i=1}^{N},$$

expressed in degrees of visual angle. These centers form the basis for SCAPE's density estimation and adaptive filtering [10]. An example implant scheme is shown in Figure 1, where the left panel plots the cortical electrode locations of 4 Utah Arrays in x, y coordinates and the right panel shows the corresponding visual-field polar coordinates of the phosphenes.

3) Gaussian Blob Rendering: Empirical reports indicate that electrically evoked phosphenes are most often perceived as localized flashes of light with an approximately circular appearance [3], [7], [10], [23]. Although some studies describe elongated or irregular forms, for simplicity we model each phosphene as an isotropic Gaussian. Note that SCAPE's core computations (density estimation and adaptive filtering) rely only on phosphene centers; the Gaussian shape is used downstream for visualization, evaluation, and amplitude normalization.

Formally, each phosphene center $(\mu_{x,i}, \mu_{y,i})$ in visual-field coordinates generates a two-dimensional Gaussian blob

$$G_i(x, y) = \exp\!\Big(-\frac{(x - \mu_{x,i})^2 + (y - \mu_{y,i})^2}{2\,\sigma^2}\Big),$$

where (x, y) are degrees of visual angle and σ is the nominal phosphene radius. The simulator rasterizes these Gaussians onto a regular image grid by converting $(\mu_{x,i}, \mu_{y,i})$ into pixel positions according to the chosen field-of-view and resolution, evaluating G_i at every grid point, and summing across all N phosphenes:

$$I_{\text{raw}}(x, y) = \sum_{i=1}^{N} G_i(x, y).$$

This raw phosphene map is then used for visualization, metric evaluation, and amplitude equalization but does not influence SCAPE's density or filter-scale computations. An example of this rendering is shown in Figure 2, where the left panel plots the centers of the phosphenes in visual-field coordinates and the right panel shows the corresponding Gaussian blobs.

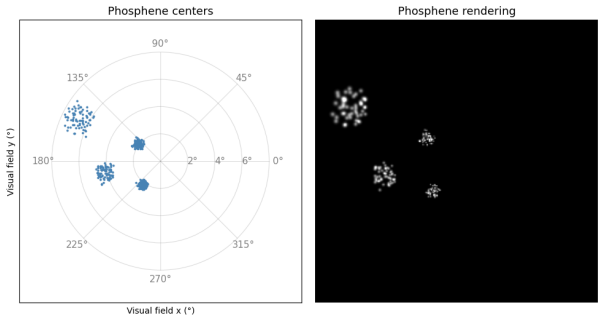

Fig. 2: Simulation of phosphene responses for a 4-Utah-array cortical implant. **Left:** Phosphene center locations in visual-field coordinates, shown on a polar grid spanning ±8° of eccentricity. **Right:** Corresponding Gaussian-blob rendering of phosphenes for a nominal stimulus amplitude (80 µA). This rendering illustrates how spatially discrete electrode activations translate into blurred perceptual spots that vary with cortical sampling density.

4) Activation-to-Electrode Mapping: To evoke a specific percept with a cortical implant, we ultimately need to decide which electrodes to turn on and how much current to send through each. A convenient intermediate representation is a continuous *activation map*

$$A(x, y) \in [0, 1]$$

defined over the visual field. Intuitively, brighter regions of this map correspond to stronger intended stimulation, and darker regions to weaker or no stimulation.

Given the N electrode centers $\{(\mu_i^x, \mu_i^y)\}_{i=1}^{N}$, we sample this map to produce a raw activation vector $\mathbf{a} = (a_1, \ldots, a_N)$. Two common sampling strategies are:

- **Point sampling:** $a_i = A(\mu_i^x, \mu_i^y)$.
- **Local pooling:** $a_i = \max_{(x,y)\in R_i} A(x, y)$, where R_i is a small neighborhood around (μ_i^x, μ_i^y) (e.g. a disk whose radius reflects local magnification).

In our simulator implementation, we build a direct correspondence between each electrode and pixel(s) of the activation map by precomputing, for every electrode i, a "distance map" on the visual-field grid. The simulator then uses these distance maps to (a) find the nearest pixel index for point sampling, or (b) define a binary mask for region pooling, entirely in pixel-space.

In a real implant system, the same correspondence can instead be obtained analytically: one applies the inverse retinotopic mapping (e.g. the wedge-dipole transform) to compute exactly which image coordinates fall under each electrode's receptive field, without requiring any precomputed pixel lookup.

Because $A(x, y)$ was normalized into $[0, 1]$, we then apply a global stimulus scale S (in Amperes) to obtain the actual per–electrode currents

$$I_i = S\,a_i.$$

Choosing S appropriately ensures that we stay within safe charge-delivery limits while preserving the relative pattern

of activation. This current vector $\{I_i\}$ is then handed to the phosphene simulator, which applies temporal filtering, thresholding, and Gaussian-blob rendering to generate the final perceptual image.

In the end, our goal is to design or optimize the activation map $A(x, y)$ (and thus the stimulus vector $\{I_i\}$) so that the resulting phosphenes are as clear and interpretable as possible.

5) Amplitude Equalization: Phosphene brightness does not scale linearly with electrode current. Local electrode density and current spread interact with multiple nonlinear stages, such as overlapping Gaussian-blob summation, saturation in sigmoid transforms, thresholding and temporal filtering, which cause some regions of the percept to appear disproportionately bright or dim. Because these nonlinear effects accumulate, there is no practical closed-form inverse that maps a target brightness profile back to electrode currents.

To achieve a uniform perceptual dynamic range, we apply a simple gain-learning step after simulation. Each phosphene i has an initial amplitude $A_i = 1$. We first drive all electrodes at the same nominal current S, render the raw percept

$$I_{\text{raw}}(x, y) = \sum_{i=1}^{N} S\, G_i(x, y),$$

and measure each blob's peak intensity,

$$m_i = \max_{x,y}\ G_i(x, y).$$

We then optimize the gains $\{A_i\}$ by minimizing

$$\mathcal{L}(A) = \frac{1}{N} \sum_{i=1}^{N} \big(A_i\, m_i - m^*\big)^2,$$

updating each A_i by gradient descent with learning rate η and clamping to $[A_{\min}, A_{\max}]$. After convergence the normalized percept

$$I_{\text{norm}}(x, y) = \sum_{i=1}^{N} A_i\, S\, G_i(x, y)$$

has phosphenes with comparable peak brightness, improving visual consistency for evaluation and downstream decoding.

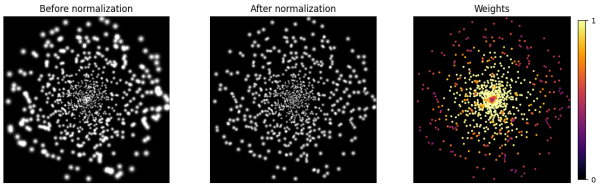

Fig. 3: Dynamic amplitude normalization example. *Left:* raw percept obtained by driving every electrode at the same current S. *Center:* normalized percept after learning per-electrode gains A_i. *Right:* learned gains $A_i \in [A_{\min}, A_{\max}]$. Normalization yields a more even brightness profile and reduces total current draw, conserving implant power and minimizing excessive charge delivery.

This calibration step yields a more uniform perceptual phosphene map, improving clarity and interpretability of encoded images. Compressing the amplitude range also lowers the total current required, which reduces neural fatigue and extends implant battery life. Although amplitude equalization is not required by SCAPE's core processing, we apply it here to standardize perceptual output for evaluation and comparison.

B. SCAPE Adaptive Encoding

The fundamental challenge in cortical prosthetic vision is that electrode arrays sample the visual field nonuniformly. Regions with dense electrode coverage can resolve fine detail while sparse regions cannot. SCAPE addresses this by adapting image filtering to the local sampling density. First, we estimate a continuous density map from the simulator-derived phosphene centers. Next, we convert density into a spatial scale map via sampling-theorem principles. Finally, we apply a shift-variant filter whose parameters vary continuously across the image. This adaptive encoding preserves maximal detail where it is supported and reduces visual clutter where it is not.

Figure 4 illustrates the full SCAPE pipeline at a glance: from electrode layout to phosphene centers (Panel 1), to density and σ-mapping (Panels 2–3), to the final shift-variant filter output (Panel 4).

1) Density Estimation: To guide adaptive filtering, SCAPE first computes a smooth sampling density $d(x, y)$ over the visual field. This density reflects how densely the implant probes each region and sets the local spatial resolution limit. Two practical estimators are used.

a) Analytic Cortical Magnification: In an idealized implant that uniformly samples cortex around the fovea, one can predict density purely from known retinotopy. Writing eccentricity $r = \sqrt{x^2 + y^2}$ in degrees, the dipole-model magnification

$$M(r) = \frac{k}{2\pi}\Big(\frac{1}{r + a} - \frac{1}{r + b}\Big)$$

(with parameters k, a, b fit to human data [10]) gives a nominal density

$$d_{\text{analytic}}(x, y) = \frac{M(r)}{r}.$$

We then scale d_{analytic} so that its integral equals the total number of phosphenes N:

$$\iint d_{\text{analytic}}(x, y)\, \mathrm{d}x\, \mathrm{d}y = N.$$

b) Adaptive Kernel Density Estimation: For real implants with nonuniform phosphene layouts, we derive density directly from the simulator-produced phosphene centers $\{(\mu_{x,i}, \mu_{y,i})\}$. We place a two-dimensional Gaussian kernel at each center, choosing its bandwidth h_i based on local spacing. Specifically, let $d_{(i,k)}$ be the distance from point i to its kth nearest neighbor; then

$$h_i = \alpha\, d_{(i,k)},$$

where α (typically 1.0) controls smoothing. The density is

$$d_{\text{KDE}}(x, y) = \sum_{i=1}^{N} \frac{1}{2\pi\, h_i^2} \exp\Big(-\frac{(x - \mu_{x,i})^2 + (y - \mu_{y,i})^2}{2\, h_i^2}\Big).$$

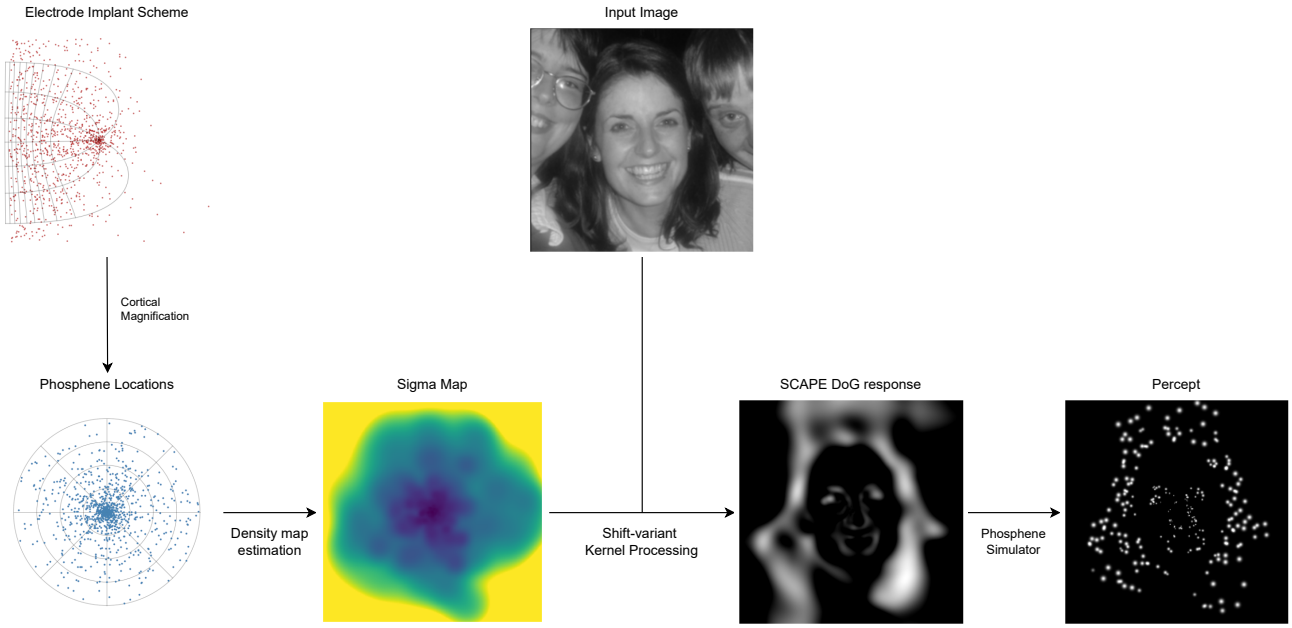

Fig. 4: Overview of the SCAPE pipeline. Starting from a cortical implant layout (top left) we obtain phosphene centers via the simulator, estimate a local density map, convert it into a spatial scale (σ) map, and then apply shift-variant filtering to produce an activation map for phosphene rendering.

Finally we normalize so that

$$\iint d_{\mathrm{KDE}}(x,y)\,\mathrm{d}x\,\mathrm{d}y = N.$$

This adaptive KDE provides a flexible density estimate that captures local variations in electrode coverage.

2) σ-Mapping via Nyquist Principles: Once a smooth density map $d(x,y)$ is available, SCAPE computes a spatial-scale map $\sigma(x,y)$ that indicates the precise local sampling limit. By the Nyquist sampling theorem [24], the highest spatial frequency resolvable at (x,y) is

$$f_{\mathrm{Nyq}}(x,y) = \tfrac{1}{2}\sqrt{\frac{d(x,y)}{\pi}},$$

since a local density of d phosphenes per unit area implies an average spacing of $\Delta \approx \sqrt{1/d}$. To match this limit exactly, we set

$$\sigma(x,y) = \frac{\kappa}{f_{\mathrm{Nyq}}(x,y)} = 2\,\kappa\sqrt{\frac{\pi}{d(x,y)}},$$

where $\kappa > 0$ (typically 1) adjusts the transition sharpness. The resulting σ-map directly specifies the standard deviation of the local processing kernel at each location, thereby guiding where and at what scale spatial frequencies should be preserved or suppressed.

3) Shift-variant Filtering: Having obtained a continuous filter-scale map $\sigma(x,y)$, SCAPE applies spatially adaptive filtering to the input image. At each location (x,y), a local kernel $K(\cdot;\sigma(x,y))$ is convolved with the image, yielding an output

$$I_{\mathrm{filt}}(x,y) = \iint K\big(u,v;\sigma(x,y)\big)\, I_{\mathrm{in}}(x-u,y-v)\,\mathrm{d}u\,\mathrm{d}v.$$

By tying the kernel's parameters to the local sampling density, shift-variant filtering preserves fine detail where electrodes are dense and reduces clutter where they are sparse. In the following sections we describe a concrete implementation using a difference of Gaussians and discuss how other kernel families can be incorporated within the same framework.

a) Difference-of-Gaussians Example: As a concrete illustration of SCAPE's adaptive filtering, we approximate the Laplacian-of-Gaussian (LoG) operator, widely used to model center–surround receptive fields in early vision [25], with a pair of Gaussian kernels. At each location (x,y), the LoG of the input image I would be

$$\mathrm{LoG}_\sigma(x,y) = \nabla^2\big[G_\sigma * I\big](x,y),$$

where G_σ is a Gaussian of standard deviation $\sigma(x,y)$ and $*$ denotes convolution. Computing a full LoG at every pixel with its own σ has complexity $\mathcal{O}(n^2)$ per pixel, where n is the kernel size, making it impractical for real-time use. Instead, we approximate it by a Difference-of-Gaussians (DoG):

$$\mathrm{DoG}_\sigma(x,y) = \big[G_{\sigma_1} - G_{\sigma_2}\big] * I\,(x,y), \quad \sigma_2 = \lambda\,\sigma_1,$$

with $\lambda > 1$ (typically $\lambda = 1.6$) chosen so that $\mathrm{DoG} \approx \mathrm{LoG}$.

To implement this shift-variant DoG efficiently, we factor each Gaussian G_σ into separable one-dimensional kernels. This reduces the complexity to $\mathcal{O}(n)$ per pixel and allows us to modulate kernel width dynamically according to the local $\sigma(x,y)$ map. In practice this involves two passes of row- and column-wise convolutions with varying standard deviations, yielding real-time performance even on mobile hardware. This computational tractability is important for prosthetic vision pipelines, where efficiency and low latency are critical. This

separable DoG serves as a simple yet powerful example of SCAPE's core idea: by varying filter scale across the image, we capture fine edges where phosphene density is high and suppress noise where it is low.

b) Extending to Other Kernels: While the DoG serves as a straightforward example, SCAPE's shift-variant framework accommodates any spatial filter family. For instance, one can replace the Gaussian kernels with orientation-tuned Gabor filters to emphasize contours aligned with cortical receptive-field preferences. More generally, the kernel $K(\cdot; \sigma(x, y))$ may be parameterized by a small set of basis functions (wavelets, steerable filters) whose shape adapts with σ. In future work one can even learn these kernels end-to-end: by embedding SCAPE into a differentiable reconstruction pipeline, the per-location filter weights can be optimized jointly with a decoder network. This flexibility allows SCAPE to capture both classical center–surround processing and more complex feature tuning while retaining its core principle of local, density-driven adaptation.

C. Reconstruction Decoder Integration

Human perceptual evaluation of phosphene encodings requires extensive behavioral studies and cannot be conducted at scale. Phosphene images are also fundamentally different from natural scenes, being sparse assemblies of localized blobs rather than continuous luminance patterns. To approximate how much visual information survives encoding, we train a convolutional decoder to reconstruct the original scene from the phosphene map.

This decoder plays a role loosely analogous to downstream visual cortex: it must infer edges, textures, and object layouts from an abstract representation. Unlike the brain, however, the decoder can adjust all its parameters through gradient-based learning and is not constrained by biological priors. Despite these fundamental differences, a consistent improvement in reconstruction accuracy suggests that the encoding has preserved more of the scene's essential structure.

Our protocol is based on the end-to-end autoencoder framework of de Ruyter van Steveninck et al. [19], but here we fix the encoder to SCAPE and train only the decoder. By comparing reconstruction error and perceptual feature losses under identical training conditions, we derive a quantitative measure of how readily SCAPE's output can be interpreted, guiding future behavioral and clinical validation.

1) Attention-UNet Architecture: The reconstruction decoder employs an Attention-UNet, an extension of the original U-Net [26] with integrated attention and channel-recalibration mechanisms. Key components include:

- **Squeeze-and-Excitation (SE) blocks** to adaptively weight feature channels [27].
- **Dilated residual blocks** in the bottleneck for multi-scale context aggregation [28].
- **Spatial attention gates** on skip connections to emphasize salient regions [29].

Down-sampling is achieved via max-pooling and up-sampling via transposed convolutions. A final 1×1 convolution with sigmoid activation produces a reconstructed grayscale image. This configuration balances context integration and selective feature focus, making it effective for recovering scenes from sparse phosphene representations.

D. Evaluation Metrics

1) Representational Similarity Analysis: Representational Similarity Analysis (RSA) is a widely used framework in neuroscience for comparing the structure of internal representations across different systems [30]. Instead of requiring spatial alignment or direct correspondence between individual activation patterns, RSA abstracts each representation into a pairwise dissimilarity matrix. This approach enables comparisons across modalities with very different spatial resolutions and scales.

In classic applications, each stimulus is represented as a vector r_i of activations in a brain region (for example, responses in V1). The representational dissimilarity matrix (RDM) is then defined by

$$D_{ij} = \delta(r_i, r_j),$$

where δ is a dissimilarity metric such as correlation distance

$$\delta_{\mathrm{corr}}(r_i, r_j) = \frac{1 - \mathrm{corr}(r_i, r_j)}{2}.$$

This RDM characterizes how different the representations of all pairs of stimuli are relative to each other.

This principle motivates our use of RSA for phosphene encoding. Phosphenes are, in a sense, direct artificial activations of V1. Although our encoding is constrained by electrode sampling, it still forms a structured response space. By building RDMs of phosphene renderings, we can ask whether the relational geometry of the stimulus space is retained. If it is, then different stimuli remain discriminable in the same way, even if their absolute fidelity is degraded.

Practically, for a set of N images, we flatten each phosphene rendering into a vector

$$r_i = \mathrm{vec}(I_p(i)),$$

and compute the correlation distance matrix

$$D_p(i, j) = \frac{1 - \mathrm{corr}(r_i, r_j)}{2}.$$

Similarly, we compute an RDM for the original images I_o. To quantify second-order similarity, we correlate the upper triangles of these matrices:

$$\rho = \mathrm{corr}_{\mathrm{Spearman}}\Big(\mathrm{vec}(D_p^{\mathrm{upper}}),\ \mathrm{vec}(D_o^{\mathrm{upper}})\Big).$$

High ρ indicates that the pairwise relationships between stimuli are well preserved by the encoding, even if pixelwise error is high. Low ρ suggests that the encoding distorts the relational structure, potentially impairing discriminability.

In this work we restrict RSA to **pixel-space representations** of images, which provides a direct low-level reference for evaluating structural fidelity of phosphene encodings. Although RSA can also be applied in higher-level feature spaces (e.g. pretrained neural network embeddings or behavioral similarity judgments), these were not used in the present analysis.

2) Reconstruction Performance: As a final evaluation axis, we assess how effectively a decoder network can reconstruct the original input image from the phosphene representation. Unlike the low-level fidelity metrics applied directly to the phosphene maps, this approach yields reconstructed natural images, making it more appropriate to apply conventional perceptual and pixel-based quality measures.

Specifically, for each phosphene-encoded input, the trained decoder produces a reconstructed image $\hat{I}_o$. We then compare $\hat{I}_o$ to the original reference image I_o using a suite of complementary metrics. These include:

- **Mean Squared Error (MSE)** – measures average pixel-wise error. Lower is better.
- **Structural Similarity Index (SSIM)** – assesses structural fidelity of luminance patterns. Lower values indicate smaller loss and thus better performance in our implementation.
- **Peak Signal-to-Noise Ratio (PSNR)** – quantifies the ratio between signal power and reconstruction error. Higher is better.
- **LPIPS** – learned perceptual similarity metric based on deep feature distances. Lower is better.
- **DISTS** – deep image structure and texture similarity, combining structural and perceptual cues. Lower is better.
- **MDSI** – mean deviation similarity index, designed for full-reference image quality assessment. Lower is better.
- **VSI** – visual saliency-induced index, emphasizing perceptual saliency in quality assessment. Lower is better.

This selection covers both intensity-based reconstruction metrics (MSE, SSIM, PSNR) and perceptual similarity metrics (LPIPS, DISTS, MDSI, VSI). Together they provide a multifaceted view of how much visual detail each encoding preserves in a form usable for end-to-end reconstruction.

Because the reconstructions are natural images, the metrics are applied without further adaptation. All metrics except PSNR are treated as loss functions, where lower values indicate better reconstruction quality.

IV. Experiments

The experiments are designed to evaluate how well SCAPE adapts to different visual scenes, implant layouts, and spatial sampling densities compared to non-adaptive baselines. We assess performance across multiple datasets with diverse image statistics and test a range of implant schemes that vary in electrode count and distribution. All methods are processed through the same prosthetic vision simulation framework to ensure a fair comparison. Performance is quantified using complementary approaches: low-level fidelity metrics applied to phosphene renderings, representational similarity analysis to assess preservation of stimulus relationships, and reconstruction-based evaluation to gauge how well a learned decoder can recover the original scene from each encoding. This multi-faceted evaluation allows us to capture both the structural accuracy of the encoded images and their potential usability for downstream perception.

A. Datasets

We evaluate SCAPE using three publicly available datasets chosen to span complementary domains of visual content and statistics that are relevant for prosthetic vision research:

- **MS COCO**: diverse natural scenes and everyday objects, providing varied spatial frequencies, textures, and complex layouts. This dataset serves as a benchmark for general-purpose encoding performance in uncontrolled environments.
- **LaPa**: human face images, representing structured, high-contrast features and smooth shading. Faces are particularly important for functional prosthetic vision, since face recognition and interpretation are critical for social interaction.
- **SUN**: indoor and outdoor scene photographs, ranging from cluttered to open environments. This dataset tests how well encoders generalize across large-scale scene structures and semantic diversity, which reflects the types of navigation and recognition tasks prosthesis users encounter in daily life.

All images are first converted to grayscale to reflect the fact that brightness is the dominant cue available in prosthetic vision, where phosphenes primarily convey intensity rather than color [31]. In early visual cortex, luminance signals are more strongly represented than chromatic signals, making brightness the most relevant dimension for encoding. Preliminary experiments with color confirmed that intensity alone preserved the majority of task-relevant structure.

Grayscale conversion is performed using **Rec. 709 perceptual luminance** (luma) weighting, which aligns with human brightness perception:

$$Y = 0.2126\,R + 0.7152\,G + 0.0722\,B, \tag{1}$$

where R, G, and B are normalized to $[0, 1]$. After conversion, perceptual normalization is applied to ensure a consistent dynamic range across images before encoding.

Scheme	View angle (deg)	Electrodes (N)	Eccentricity range (deg)
Uniform 1024	16.0	1024	0.001–7.984
Neuralink	25.0	4224	0.000–12.095
1 Utah array	0.4	94	0.009–0.203
4 Utah arrays	16.0	320	1.632–7.931

TABLE I: Implant layouts and basic properties. Counts reflect the effective number of electrodes inside the simulator field of view. Eccentricity is reported in degrees of visual angle.

B. Implant Schemes

We evaluate SCAPE across four implant layouts that span large differences in sampling density, spatial extent, and geometric regularity. This set stresses density adaptation both in the foveal region and in the periphery, and it includes layouts used in recent prosthetic vision studies.

a) Uniform 1024: This layout provides a dense, near-uniform sampling of the visual field and serves as a standard reference in the literature on differentiable prosthetic vision. It allows direct comparison to prior work that uses a similar order of magnitude in channel count [19]. It also exposes whether SCAPE preserves fine structure when the density is sufficient across most of the field of view.

b) Neuralink shank: This layout represents a high-site-count configuration that covers a wide field of view with thousands of closely spaced sites. Such arrays have been demonstrated in the Neuralink brain–machine interface platform, which integrates flexible polymer shanks with thousands of recording and stimulation channels [4]. Related developments in ultraflexible electrode arrays further highlight the feasibility of maintaining long-term, high-density interfaces in cortical tissue [5]. Within our framework, this scheme stresses computational efficiency and scale selection, since local density varies strongly with eccentricity and along the shank geometry. It therefore provides a test case for whether SCAPE scales robustly to thousands of sites while maintaining stability and efficiency.

c) One Utah array: This layout matches a single-array setup and concentrates sampling near fixation with a very small field of view. It reflects ongoing experimental configurations used in current research programs at the Institute of Bioengineering of the Miguel Hernández University. It tests SCAPE under severe channel limits and minimal coverage, which is useful for understanding performance in constrained early clinical or preclinical settings.

d) Four Utah arrays: This layout models a realistic multi-array configuration with gaps between arrays and moderate channel count. It introduces strong spatial inhomogeneity due to array borders and inter-array spacing, which is a common feature of practical cortical implants. It therefore tests whether SCAPE can suppress clutter in sparse regions while preserving detail within array footprints.

Together these schemes span two orders of magnitude in channel count, from 94 to 4224. They also span field of view from a fraction of a degree to more than twelve degrees of eccentricity. This diversity is important because SCAPE is designed to set filter scale from local sampling density. The chosen set probes that mechanism in both uniform and highly inhomogeneous layouts, which is essential for a fair test of adaptive encoding.

C. Baselines

We compare SCAPE against simple, widely used encoding strategies that do not adapt their spatial scale to local sampling density. Each baseline produces an activation map that is passed through the same simulator and amplitude normalization as SCAPE, ensuring that differences in outcome are attributable to the encoding strategy itself. Together, these baselines span a spectrum of encoding strategies: a contour-focused method (Canny), a bandpass but nonadaptive method (fixed DoG), and a structure-free random control.

a) Canny edge detection: Canny edge maps are a standard choice in prosthetic vision pipelines because they preserve object boundaries under severe spatial resolution limits and often improve recognition in simulated conditions. We include a Canny baseline with standard smoothing and hysteresis thresholding, applied uniformly across the field of view. To enhance the visibility of extracted contours in the simulated percepts, edges are dilated with a 3×3 kernel [19]. This baseline serves as a strong nonadaptive feature extractor centered on high-contrast contours.

b) Fixed-scale Difference of Gaussians: We implement a fixed-scale Difference of Gaussians (DoG) filter as a second baseline. This method applies Gaussian smoothing at a fixed scale and subtracts it from the original image, producing a bandpass representation that emphasizes edges and local contrast. In contrast to SCAPE, the filter scale is uniform across the image and cannot adjust to local variations in sampling density or spatial frequency content. We use a fixed filter size of $\sigma = 3$, chosen to produce feature sizes comparable to the dilated Canny edges.

c) Random control: As a structure-free control, we generate a random activation pattern that matches the total activation energy of the method under test. This ensures that any observed improvements are due to meaningful spatial structure in the encoding rather than trivial differences in overall current or luminance.

Representative examples of these baselines are shown in Figure 6, together with SCAPE for comparison. All methods use identical preprocessing, simulator settings, and amplitude normalization, ensuring that differences in performance reflect only the encoding strategy rather than downstream rendering effects.

D. SCAPE Configuration

SCAPE builds a local density map from phosphene coordinates using adaptive kernel density estimation and converts that density into a spatial scale map $\sigma(x, y)$. The scale map controls a shift variant Difference of Gaussians that runs with separable one dimensional passes.

a) Density estimation: All experiments use adaptive KDE over phosphene centers in visual field coordinates. Bandwidth h_i for point i equals α times the distance to its k-th nearest neighbor. We set $k = 16$ and $\alpha = 1.0$. The resulting density map is normalized so that its surface integral equals the total number of phosphenes of the scheme under test.

b) Mapping density to scale: We convert density $d(x, y)$ into a local scale through a Nyquist motivated rule. We use

$$\sigma_{\text{fov}}(x, y) \;=\; \frac{2}{\pi\sqrt{2}\,\beta}\,\frac{1}{\sqrt{d(x, y)}} \quad \text{with} \quad \beta = 0.55,$$

which yields a scale that decreases in regions with higher sampling density and increases in sparse regions. For filtering on an image grid we convert σ_{fov} from degrees to pixels using the simulator field of view and resolution.

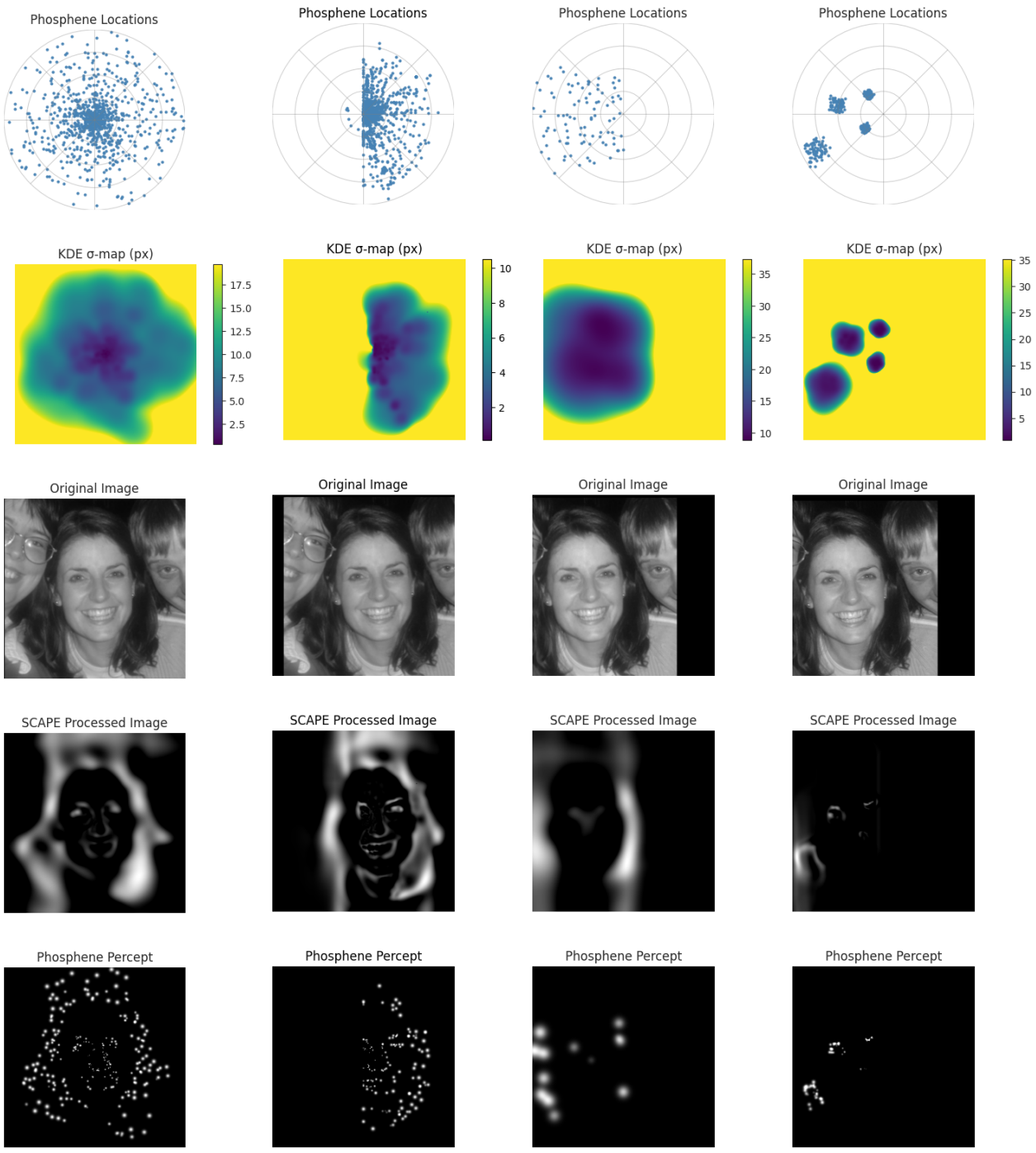

Fig. 5: **Implant schemes used for evaluation.** Each column illustrates one of the four electrode layouts tested in this study: Uniform 1024, Neuralink-type shank [4], [5], single Utah array, and four Utah arrays. For each scheme we show (top to bottom): electrode locations in visual field coordinates, the corresponding kernel density estimate (KDE) map of local sampling density, an example input image, the SCAPE-processed representation, and the resulting phosphene percept. The field of view differs between layouts and is scaled here for visualization. Together these schemes span large variations in electrode count, spatial distribution, and coverage, providing a diverse testbed for evaluating adaptive encoding.

c) Stimulus alignment: To ensure consistency across implant schemes, we center each input stimulus on the centroid of the $\sigma(x, y)$ map. This alignment guarantees that the most informative regions of the stimulus overlap with the highest density regions of the implant layout, and it standardizes comparisons across schemes.

d) Filter family and execution: We use a Difference of Gaussians with ratio $\lambda = 1.6$ to approximate a Laplacian of Gaussian while remaining efficient. The filter runs as two separable Gaussian passes per branch, first horizontal then

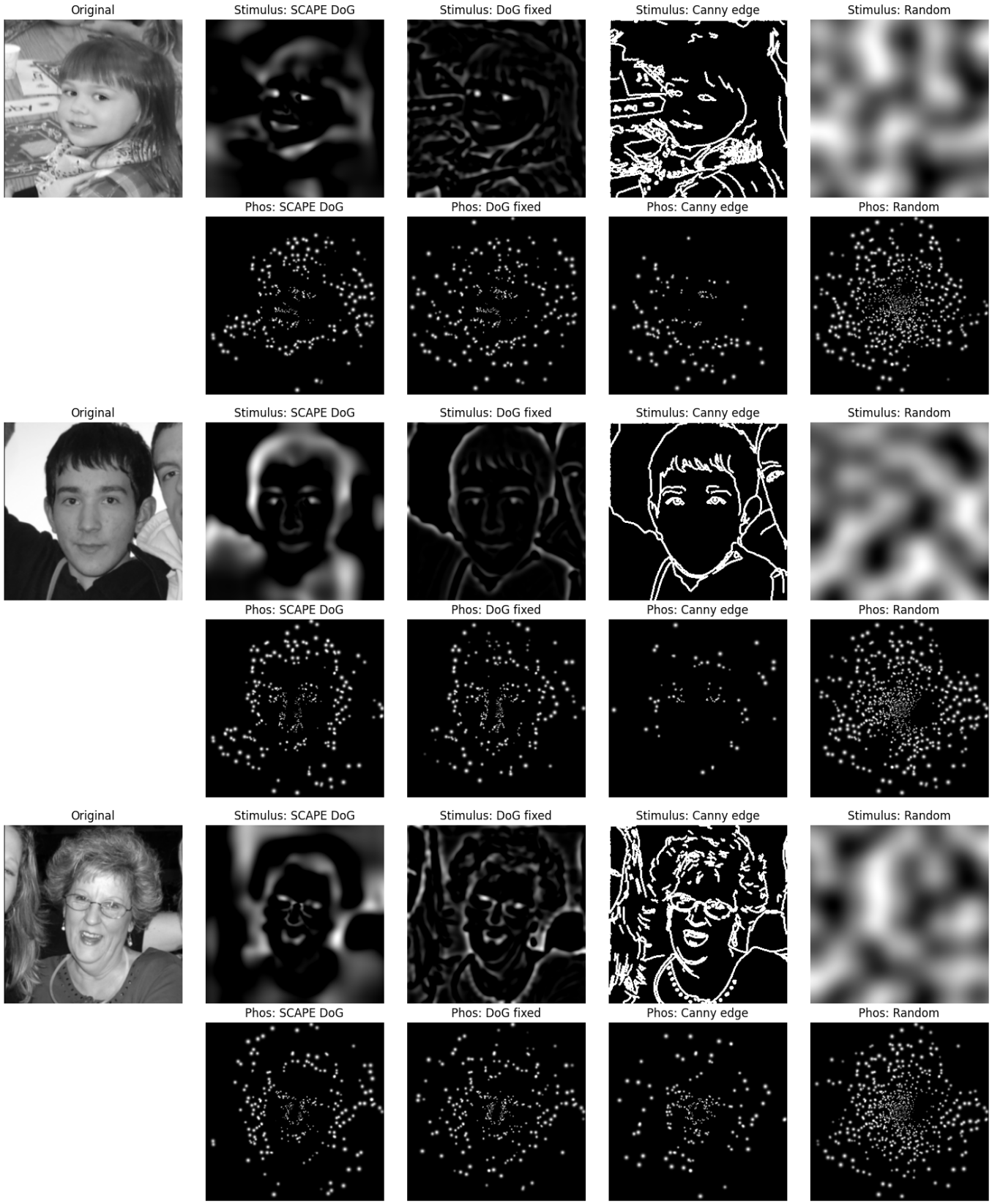

Fig. 6: **Qualitative comparison of baseline and SCAPE encoders.** Representative examples from the LaPa dataset. Top: stimulus representations after encoding. Bottom: corresponding phosphene renderings for the default implant scheme with 1024 electrodes. SCAPE balances edge emphasis with adaptive clutter suppression, while fixed DoG and Canny either oversmooth or fragment the structure. Random activation patterns are added as a control.

vertical, at $\sigma_1(x, y)$ and $\sigma_2(x, y) = \lambda\,\sigma_1(x, y)$. Gaussian weights are truncated where their value falls below a small threshold and each one dimensional kernel is normalized to sum to one. Padding uses reflection at image borders. The half kernel radius is capped at a fixed maximum to bound runtime and memory.

e) Retinotopic model: Retinotopic projection follows the dipole form of the Polimeni wedge–dipole family with standard parameters. This model sets the visual field coordinate system used by the KDE and by the degree to pixel conversion.

f) Sampling to electrodes: The shift variant DoG produces an activation map on the simulator grid. Electrode currents are obtained by sampling the activation map at phosphene centers. The same sampling rule is used for every implant scheme.

E. Amplitude Normalization

To keep comparisons fair across encoders, we apply the same dynamic amplitude equalization to every method. The calibration runs once per implant scheme after mapping activations to electrode currents and before simulation. The resulting per electrode gains are reused for all images and all encoders within that scheme and across datasets. The procedure follows the definition in the Methods section and is tuned for stability.

We optimize the gains with a learning rate of 0.002 for 2000 steps, constrain each gain to $[0, \texttt{amplitude}]$, and use a small scale factor of 1×10^{-4} to set the update magnitude. In practice this removes simulator induced brightness bias, reduces extreme responses, and keeps total delivered current comparable, so the results reflect differences in encoding rather than artifacts of the simulation.

a) Representational similarity analysis: For each dataset and implant scheme we construct representational dissimilarity matrices (RDMs) for the original images and for each encoded set. Each image is converted to a grayscale vector, and pairwise dissimilarity is computed as correlation distance

$$D_{ij} = 1 - \mathrm{corr}(r_i, r_j).$$

This produces an RDM that summarizes the relational structure of the stimulus set.

To compare encoders, we correlate the upper triangular entries of the reference RDM with those of the encoded RDM using Spearman correlation ρ. Higher ρ indicates closer alignment of relational geometry and thus better preservation of stimulus similarity structure. For visualization, we also report dissimilarity as $1 - \rho$, so that lower values correspond to better preservation.

To obtain stable estimates, each method is evaluated on 500 randomly sampled images per dataset. Standard errors are estimated by 5000 bootstrap resamples, and statistical significance is assessed with 10,000 permutations. We report mean ρ (and $1 - \rho$ when plotting) per dataset and implant scheme for SCAPE and all baselines.

b) Reconstruction performance: We measure how much scene information survives each encoding by training a fixed decoder to reconstruct the original image from phosphene renderings. For every encoder, dataset, and implant scheme we train an Attention U-Net on the corresponding phosphene maps using the same preprocessing, the same simulator settings, and the same training schedule. Inputs and targets are single-channel perceptual luminance images with the normalization described above. Evaluation uses a held-out split and reports MSE, SSIM, PSNR, LPIPS, DISTS, and VSI. Scores are computed per image and summarized as the mean with standard error. We also include representative reconstructions to illustrate qualitative differences that are not fully captured by the metrics.

For practical reasons we restrict reconstruction experiments to the Uniform 1024 scheme. Decoder training is specific to each implant layout, and running it for all schemes would multiply training cost substantially. The 1024 scheme provides a stable, widely used reference case in the literature and offers the fairest setting to compare encoders while keeping evaluation tractable.

V. Results

A. Qualitative Comparison of Encoders

We begin by comparing encoders qualitatively to build intuition for their behavior before turning to quantitative analyses. Representative examples of SCAPE, fixed-scale DoG, Canny, and random control are shown in Figure 6 (see Section IV). Each method produces distinct patterns in both the encoded stimulus and the resulting phosphene rendering under the default 1024-electrode implant scheme.

Several systematic differences emerge. **SCAPE** emphasizes local contrast while suppressing spurious responses in low-density regions, yielding more interpretable percepts. **Fixed-scale DoG** captures edges but fails to adapt to varying density, leading to oversmoothing in dense regions and excessive detail in sparse ones. **Canny** extracts sharp contours but discards shading, resulting in fragmented and brittle percepts. Finally, the **random control** produces unstructured patterns with no relation to the input, serving as a sanity check.

These qualitative examples illustrate why adaptive scale selection is important: SCAPE strikes a balance between edge emphasis and clutter suppression that is not achieved by fixed or contour-based baselines. The following sections quantify these observations using representational dissimilarity analysis and reconstruction performance.

B. Representational Dissimilarity Analysis

We next assessed the representational fidelity of different phosphene encoding strategies using RSA as defined in Section IV. Figure 7 shows the dissimilarity of phosphene encodings relative to RDMs across three datasets (COCO, LaPa, SUN) and four implant schemes. Lower values correspond to better preservation of the relational structure of the original images.

Across all conditions, **SCAPE (adaptive DoG)** achieved the lowest dissimilarity, indicating that its density-adaptive filtering preserves similarity structure most effectively. The **fixed-scale DoG** baseline performed moderately well but failed to adapt across sampling densities, leading to higher dissimilarity in sparse layouts. **Canny edges** and the **random control**

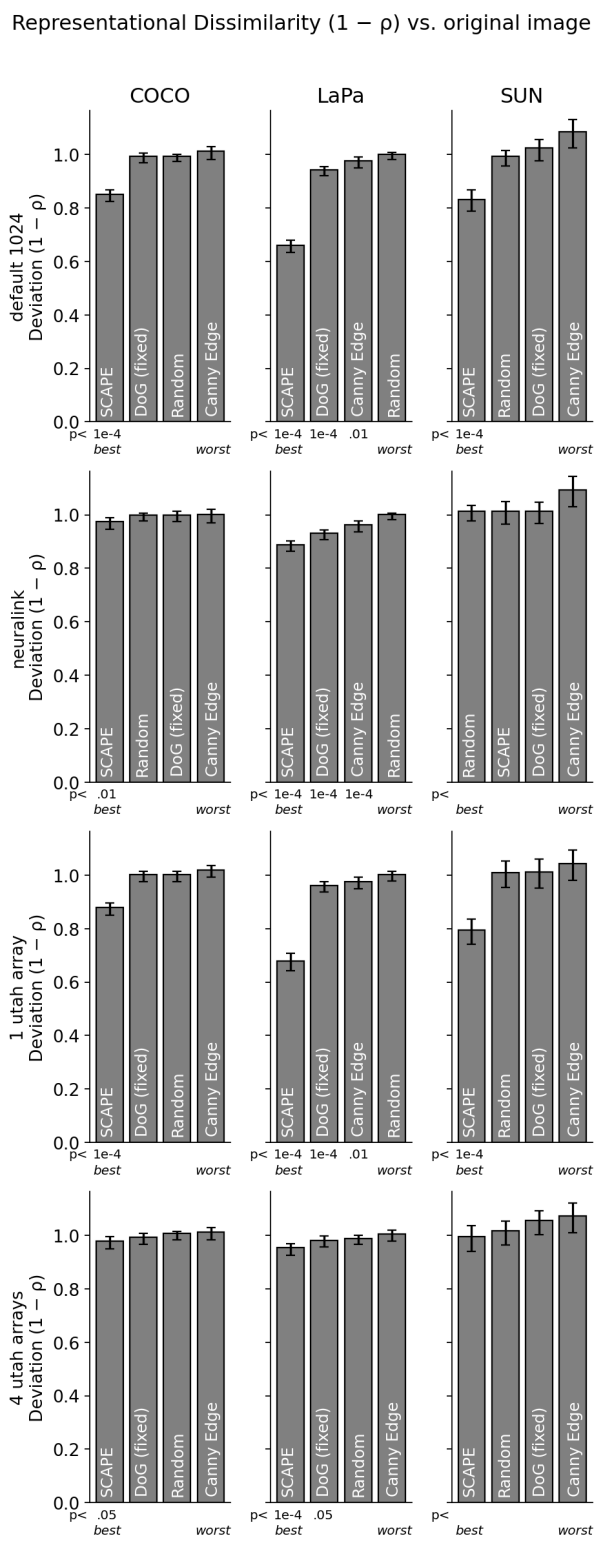

Fig. 7: **Representational similarity analysis.** Dissimilarity $(1 - \rho)$ between phosphene encodings and original image RDMs across COCO, LaPa, and SUN datasets. Bars show mean dissimilarity, error bars denote bootstrap confidence intervals, and p-values (permutation test) are reported below each bar. Lower values correspond to greater preservation of structure.

consistently produced substantially higher dissimilarities, reflecting poor alignment with stimulus similarity. Permutation-based p-values (shown below each bar) confirm that SCAPE significantly outperforms the nonadaptive baselines across datasets and implant schemes.

C. Reconstruction Performance

To complement the representational analysis, we evaluated how effectively a learned decoder could reconstruct natural images from phosphene representations. As noted in Section IV, reconstruction experiments were restricted to the Uniform 1024 scheme. Decoder training is specific to each implant layout, and running separate decoders for all schemes would multiply training cost substantially. The 1024 layout provides a stable and widely used reference case, making it the most appropriate benchmark for encoder comparison.

Quantitative results are summarized in Tables IIa–IIc, grouped into intensity-based metrics (MSE, SSIM, PSNR) and perceptual similarity metrics (LPIPS, DISTS, MDSI, VSI). Across all datasets, **SCAPE** consistently achieved the lowest MSE and the highest PSNR values, indicating superior fidelity in reproducing low-level intensities. SSIM scores showed a similar pattern on COCO and SUN, with more modest but consistent gains on LaPa. Perceptual metrics confirmed these findings: SCAPE generally outperformed the nonadaptive baselines across datasets, although individual metrics occasionally favored Canny (e.g., MDSI on COCO and SUN). Despite such cases, SCAPE provided the best overall balance across metrics, highlighting its robustness to diverse visual statistics.

Qualitative reconstructions further illustrate these trends. As shown in Figures 8–10, SCAPE reconstructions preserve continuous shading and object structure, enabling more interpretable outputs. By contrast, Canny emphasizes sparse contours at the expense of smooth surfaces, which often fragments object boundaries, while the random control produces severely degraded reconstructions lacking usable structure. Taken together, these results demonstrate that adaptive scale selection enables SCAPE to retain substantially more perceptually relevant information than nonadaptive encoding strategies. While decoder reconstructions are not direct behavioral measures, the ability to recover more naturalistic scenes suggests that human users may also benefit in perceptual and functional tasks, motivating future behavioral and clinical validation.

VI. Discussion and Conclusion

A. Summary of Contributions

This work introduced SCAPE, a principled framework for encoding visual information in cortical prostheses that adapts spatial filtering to the local phosphene map sampling density imposed by electrode layouts. Unlike conventional pipelines that apply uniform filters across the visual field, SCAPE derives a continuous density map from electrode or phosphene locations, converts this into a spatial scale map using Nyquist principles, and applies shift-variant filtering whose kernel width matches local resolution limits. In an efficient

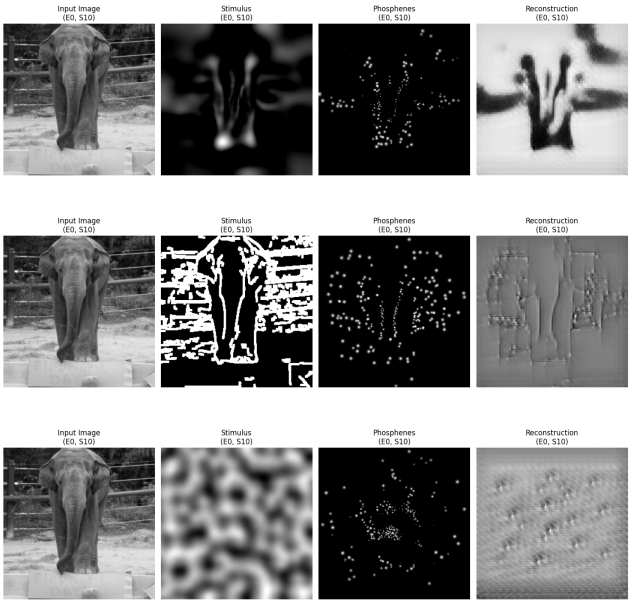

Fig. 8: Reconstruction examples on COCO under the Uniform 1024 scheme. Each row shows one encoding method: SCAPE (top), Canny edges (middle), and random control (bottom). Columns depict the pipeline from input image to activation map, phosphene rendering, and final reconstruction.

Fig. 9: Reconstruction examples on LaPa faces under the Uniform 1024 scheme. Each row shows one encoding method: SCAPE (top), Canny edges (middle), and random control (bottom). Columns depict the pipeline from input image to activation map, phosphene rendering, and final reconstruction.

TABLE II: Reconstruction metrics on the Uniform 1024 scheme across datasets (results restricted to Uniform 1024; see Section IV for rationale).

(a) COCO

Processing model	Intensity reconstruction			Perceptual reconstruction			
	MSE	SSIM	PSNR	LPIPS	DISTS	MDSI	VSI
SCAPE	**0.062**	**0.592**	**12.718**	**0.620**	**0.424**	0.559	**0.151**
Canny	0.070	0.635	12.012	0.623	0.489	**0.545**	0.156
Random	0.071	0.635	11.999	0.695	0.636	0.600	0.179

(b) LaPa

Processing model	Intensity reconstruction			Perceptual reconstruction			
	MSE	SSIM	PSNR	LPIPS	DISTS	MDSI	VSI
SCAPE	**0.057**	**0.500**	**13.180**	**0.564**	**0.382**	**0.515**	**0.101**
Canny	0.072	0.554	12.285	0.579	0.413	0.518	0.129
Random	0.076	0.566	12.003	0.625	0.516	0.560	0.147

(c) SUN

Processing model	Intensity reconstruction			Perceptual reconstruction			
	MSE	SSIM	PSNR	LPIPS	DISTS	MDSI	VSI
SCAPE	**0.054**	**0.599**	**13.177**	0.616	**0.433**	0.563	**0.154**
Canny	0.064	0.636	12.359	**0.607**	0.486	**0.548**	0.156
Random	0.067	0.638	12.171	0.676	0.582	0.602	0.189

implementation, we demonstrated a separable Difference-of-Gaussians operator that achieves real-time performance while preserving the encoded images' detail in dense regions and suppressing clutter in sparse ones.

Across multiple cortical implant schemes and datasets, SCAPE consistently produced more interpretable percepts and preserved relational structure more effectively than non-adaptive baselines. Both representational similarity analysis and reconstruction experiments confirmed these gains across natural scenes, faces, and indoor–outdoor environments. These findings establish SCAPE as a general and efficient strategy for phosphene vision encoding that can be integrated with existing neural interface pipelines [9], phosphene simulators [10], [20], [32], extended with alternative kernels, and adapted to patient-specific layouts.

B. Relation to Prior Work

Research on prosthetic vision encoding spans from early heuristics to recent learned approaches. Early pipelines emphasized contrast enhancement and edge extraction, which offered computational efficiency but treated the visual field uniformly, limiting their ability to balance detail in dense regions and clutter in sparse ones. Learned encoders have demonstrated improved perceptual quality, particularly when combined with differentiable simulators, but typically apply spatial filtering globally rather than adapting to local sampling constraints. Simulation studies further highlight that excessive detail can overwhelm sparse layouts, while overly uniform filtering can erase useful structure [13], [14].

SCAPE complements these directions by offering a middle ground: lightweight and more interpretable heuristics, but grounded in electrode density and Nyquist theory rather than hand-tuned parameters. It also integrates naturally with modern differentiable simulators [10], enabling hybrid pipelines where phosphene density-aware preprocessing forms the first stage and downstream networks refine the encoding.

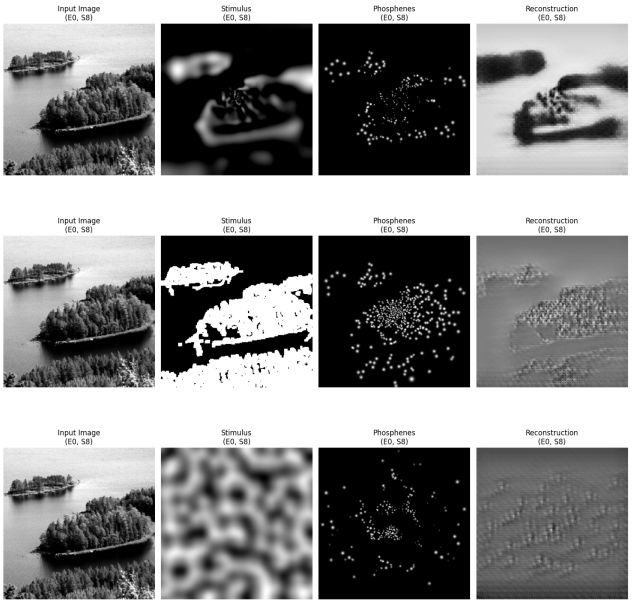

Fig. 10: Reconstruction examples on SUN scenes under the Uniform 1024 scheme. Each row shows one encoding method: SCAPE (top), Canny edges (middle), and random control (bottom). Columns depict the pipeline from input image to activation map, phosphene rendering, and final reconstruction.

C. Clinical and Practical Implications

Electrode arrays might sample the visual field inhomogeneously, yet most encoding pipelines ignore this constraint. By linking electrode density to filter scale, SCAPE ensures that each region is represented at an appropriate level of detail. This adaptivity improves interpretability of phosphene patterns, reduces spurious activations, and preserves fine structure where the implant supports it. Such improvements are directly relevant for functional tasks such as object recognition, face perception, and scene navigation—benchmarks in clinical evaluation of prosthetic vision [7], [33]. Moreover, by providing cleaner and more interpretable inputs from the outset, SCAPE may accelerate training outcomes and give patients a head-start in the process of perceptual learning and adaptation with a new implant [34], [35].

In order to exploit SCAPE's advantages, accurate phosphene maps will need to be obtained. While high-channel count implants have been demonstrated to generate shape perception [8], they come with challenges. Practically, the acquisition of the phosphene maps in human participants relies mostly on manually mapping the generated phosphenes using the user´s report after single or multiple electrode stimulation. This procedure will become greatly time consuming and tedious for the implant user when channel numbers go up to the hundreths of thousands. However, new methodologies based in dimensionality reducion of resting-state neural data promise a greatly accelerated and semi-automated procedure even for high electrode counts [36]. These methods, combined with receptive field map priors obtained from anatomical scans [37], [38] or from aligning functional atlas [39], together with pre-surgical optimal planning [40], will help obtaining high-quality phosphene maps estimations that can be directly used (or fine-tuned with new behavioral reports) to directly inform our SCAPE framework.

SCAPE's separable DoG implementation also achieves real-time performance on modest hardware, making it suitable for the stringent power and latency constraints of implantable systems. These properties position SCAPE as a practical foundation for next-generation cortical implant pipelines, bridging the gap between simulation and clinical translation.

D. Limitations

Several limitations must be noted. First, all results were obtained in simulation, which cannot capture the full variability of human percepts. Clinical evaluation will be essential to assess whether the observed improvements translate into functional gains. Second, usability was evaluated indirectly through decoder reconstructions, which provide a convenient proxy but not behavioral outcomes such as recognition or navigation. Behavioral studies in VR will be useful to test the perceptual outcomes and to establish whether the improvements observed in simulation may translate into perceptual and functional benefits, and ultimately to assess these effects in real clinical investigation settings.

Third, our study focused solely on static images and spatial encoding. Temporal dynamics such as persistence, adaptation, and fading of phosphenes [3], [7], [23], [41]–[43] were not used, though recent simulators now incorporate these mechanisms [10], [20]. Finally, SCAPE is presented as a principled but non-learned approach, leaving open how it compares to task-optimized or human-in-the-loop encoders [21] under behavioral evaluation.

E. Future Directions

Future work should validate SCAPE in behavioral and clinical settings, ideally through immersive VR-based prosthetic vision platforms. Incorporating temporal dynamics of phosphene perception is another priority, extending SCAPE's adaptivity from space into time. Another direction is to integrate depth information into the encoding pipeline, allowing scale to be modulated not only by electrode density but also by viewing distance and scene geometry. Hybrid approaches also hold promise, where density-driven filtering is refined through task-driven learning or user feedback, and more sophisticated kernels capture orientation or feature selectivity. Finally, SCAPE's efficiency makes it a strong candidate for embedded deployment, warranting evaluation on processors and FPGAs under realistic power and latency constraints.

F. Conclusion

SCAPE introduces a density-adaptive encoding framework for cortical prosthetic vision that links electrode layout to spatial filtering through Nyquist-based scale mapping. Efficiently implemented with a separable Difference-of-Gaussians, it improves representational fidelity and reconstruction quality while remaining lightweight and compatible with clinical hardware. By grounding adaptivity in electrode density, SCAPE

provides a principled alternative to uniform filtering and a flexible foundation for integration with learned or patient-specific encoders, laying the groundwork for next-generation cortical implants that adapt across space today and across time, tasks, and users in the future.

REFERENCES

[1] E. M. Maynard, C. T. Nordhausen, and R. A. Normann, "The utah intracortical electrode array: A recording structure for potential brain-computer interfaces," *Electroencephalography and Clinical Neurophysiology*, vol. 102, no. 3, pp. 228–239, 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013469496951760

[2] R. A. Normann, E. M. Maynard, P. J. Rousche, and D. J. Warren, "A neural interface for a cortical vision prosthesis," *Vision Research*, vol. 39, no. 15, pp. 2577–2587, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0042698999000401

[3] E. M. Schmidt, M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. O'Rourke, and P. Vallabhanath, "Feasibility of a visual prosthesis for the blind based on intracortical micro stimulation of the visual cortex," *Brain*, vol. 119, no. 2, pp. 507–522, 04 1996. [Online]. Available: https://doi.org/10.1093/brain/119.2.507

[4] E. Musk and Neuralink, "An integrated brain-machine interface platform with thousands of channels," *Journal of Medical Internet Research*, vol. 21, no. 10, p. e16194, 2019.

[5] Z. Zhao, H. Zhu, X. Li, L. Sun, F. He, J. E. Chung, D. F. Liu, L. Frank, L. Luan, and C. Xie, "Ultraflexible electrode arrays for months-long high-density electrophysiological mapping of thousands of neurons in rodents," *Nature Biomedical Engineering*, vol. 7, no. 4, pp. 520–532, 2023.

[6] C. Orlemann, C. Boehler, R. N. Kooijmans, B. Li, M. Asplund, and P. R. Roelfsema, "Flexible polymer electrodes for stable prosthetic visual perception in mice," *Advanced Healthcare Materials*, vol. 13, p. 2304169, 2024. [Online]. Available: https://doi.org/10.1002/adhm.202304169

[7] E. Fernández, A. Alfaro, C. Soto-Sánchez, P. Gonzalez-Lopez, A. M. Lozano, S. Peña, M. D. Grima, A. Rodil, B. Gómez, X. Chen, P. R. Roelfsema, J. D. Rolston, T. S. Davis, and R. A. Normann, "Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex," *The Journal of Clinical Investigation*, vol. 131, no. 23, 12 2021. [Online]. Available: https://www.jci.org/articles/view/151331

[8] X. Chen, F. Wang, E. Fernández, and P. R. Roelfsema, "Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex," *Science*, vol. 370, no. 6521, pp. 1191–1196, 2020. [Online]. Available: https://doi.org/10.1126/science.abd7435

[9] A. Lozano, J. S. Suárez, C. Soto-Sánchez, J. Garrigós, J. J. Martínez-Alvarez, J. M. Ferrández, and E. Fernández, "Neurolight: A deep learning neural interface for cortical visual prostheses," *International Journal of Neural Systems*, vol. 30, no. 09, p. 2050034, 2020. [Online]. Available: https://doi.org/10.1142/S0129065720500347

[10] M. van der Grinten, J. de Ruyter van Steveninck, A. Lozano, L. Pijnacker, B. Rueckauer, P. Roelfsema, M. van Gerven, R. van Wezel, U. Güçlü, and Y. Güçlütürk, "Towards biologically plausible phosphene simulation for the differentiable optimization of visual cortical prostheses," *eLife*, vol. 13, p. e85812, feb 2024. [Online]. Available: https://doi.org/10.7554/eLife.85812

[11] M. Beyeler and M. Sanchez-Garcia, "Towards a smart bionic eye: Ai-powered artificial vision for the treatment of incurable blindness," *Journal of Neural Engineering*, vol. 19, no. 6, p. 061001, 2022. [Online]. Available: https://doi.org/10.1088/1741-2552/aca69d

[12] J. de Ruyter van Steveninck, T. van Gestel, P. Koenders, G. van der Ham, F. Vereecken, U. Güçlü, M. van Gerven, Y. Güçlütürk, and R. van Wezel, "Real-world indoor mobility with simulated prosthetic vision: The benefits and feasibility of contour-based scene simplification at different phosphene resolutions," *Journal of Vision*, vol. 22, no. 2, p. 1, 2022. [Online]. Available: https://doi.org/10.1167/jov.22.2.1

[13] J. Kasowski and M. Beyeler, "Immersive virtual reality simulations of bionic vision," in *Augmented Humans 2022*, ser. AHs 2022. ACM, Mar. 2022, p. 82–93. [Online]. Available: http://dx.doi.org/10.1145/3519391.3522752

[14] N. Han, S. Srivastava, A. Xu, D. Klein, and M. Beyeler, "Deep learning–based scene simplification for bionic vision," 2021. [Online]. Available: https://arxiv.org/abs/2102.00297

[15] L. Relic, B. Zhang, Y.-L. Tuan, and M. Beyeler, "Deep learning-based perceptual stimulus encoder for bionic vision," 2022. [Online]. Available: https://arxiv.org/abs/2203.05604

[16] W. Liu, W. Fink, M. Tarbell, and M. Sivaprakasam, "Image processing and interface for retinal visual prostheses," in *2005 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2005, pp. 2927–2930 Vol. 3.

[17] J. K. Yeonji Oh, Jonggi Hong, "Retinal prosthesis edge detection (rped) algorithm: Low-power and improved visual acuity strategy for artificial retinal implants," *PLoS ONE*, vol. 19, no. 6, p. e0305132, 2024.

[18] J. Granley, L. Relic, and M. Beyeler, "Hybrid neural autoencoders for stimulus encoding in visual and other sensory neuroprostheses," 2022. [Online]. Available: https://arxiv.org/abs/2205.13623

[19] J. de Ruyter van Steveninck, U. Güçlü, R. van Wezel, and M. van Gerven, "End-to-end optimization of prosthetic vision," *bioRxiv*, 2020. [Online]. Available: https://www.biorxiv.org/content/early/2020/12/21/2020.12.19.423601

[20] I. Fine and G. M. Boynton, "A virtual patient simulation modeling the neural and perceptual effects of human visual cortical stimulation, from pulse trains to percepts," *Scientific Reports*, vol. 14, no. 17400, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-65337-1

[21] J. Granley, T. Fauvel, M. Chalk, and M. Beyeler, "Human-in-the-loop optimization for deep stimulus encoding in visual prostheses," 2023. [Online]. Available: https://arxiv.org/abs/2306.13104

[22] J. Polimeni, M. Balasubramanian, and E. Schwartz, "Multi-area visuotopic map complexes in macaque striate and extra-striate cortex," *Vision Research*, vol. 46, no. 20, pp. 3336–3359, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0042698906001428

[23] W. H. Dobelle and M. G. Mladejovsky, "Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind," *J. Physiol.*, vol. 243, no. 2, pp. 553–576, Dec. 1974.

[24] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[25] R. A. Young, "The gaussian derivative model for spatial vision: I. retinal mechanisms," *Spatial Vision*, vol. 2, no. 4, pp. 273 – 293, 1987. [Online]. Available: https://brill.com/view/journals/sv/2/4/article-p273_3.xml

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[27] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2019. [Online]. Available: https://arxiv.org/abs/1709.01507

[28] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2016. [Online]. Available: https://arxiv.org/abs/1511.07122

[29] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 2018. [Online]. Available: https://arxiv.org/abs/1804.03999

[30] N. Kriegeskorte, M. Mur, and P. Bandettini, "Representational similarity analysis – connecting the branches of systems neuroscience," *Frontiers in systems neuroscience*, vol. 2, p. 4, 02 2008.

[31] W. Wang, R. Li, J. Ding, L. Tao, D.-P. Li, and Y. Wang, "V1 neurons respond to luminance changes faster than contrast changes," *Scientific Reports*, vol. 5, p. 17173, 12 2015.

[32] M. Beyeler, G. M. Boynton, I. Fine, and A. Rokem, "pulse2percept: A python-based simulation framework for bionic vision," in *Proceedings of the 16th Python in Science Conference (SciPy)*, 2017, pp. 81–88. [Online]. Available: https://doi.org/10.25080/shinma-7f4c6e7-00c

[33] K. Stingl, K. U. Bartz-Schmidt, D. Besch, C. K. Chee, C. L. Cottriall, F. Gekeler, M. Groppe, T. L. Jackson, R. E. MacLaren, A. Koitschev, A. Kusnyerik, J. Neffendorf, J. Nemeth, M. A. N. Naeem, T. Peters, J. D. Ramsden, H. Sachs, A. Simpson, M. S. Singh, B. Wilhelm, D. Wong, and E. Zrenner, "Subretinal visual implant alpha IMS–Clinical trial interim report," *Vision Res.*, vol. 111, no. Pt B, pp. 149–160, Jun. 2015.

[34] R. A. Normann, B. Greger, P. House, S. F. Romero, F. Pelayo, and E. Fernandez, "Toward the development of a cortically based visual neuroprosthesis," *Journal of Neural Engineering*, vol. 6, no. 3, p. 035001, 2009, erratum in: J Neural Eng. 2009 Aug;6(4):049802. [Online]. Available: https://doi.org/10.1088/1741-2560/6/3/035001

[35] M. Beyeler, "Learning to see again: The role of perceptual learning and user engagement in sight restoration," *Journal of Vision*, vol. 24, no. 10, p. 200, 2024. [Online]. Available: https://doi.org/10.1167/jov.24.10.200

[36] A. Lozano, X. Chen, M. La Grouw, B. Li, F. Wang, M. van der Grinten, C. Soto-Sánchez, A. Morales-Gregorio, E. Fernández, and

P. R. Roelfsema, "Large-scale rf mapping without visual input for neuroprostheses in macaque and human visual cortex," *medRxiv*, 2024. [Online]. Available: https://doi.org/10.1101/2024.12.22.24319047

[37] N. C. Benson, O. H. Butt, R. Datta, P. D. Radoeva, D. H. Brainard, and G. K. Aguirre, "The retinotopic organization of striate cortex is well predicted by surface topology," *Current Biology*, vol. 22, no. 21, pp. 2081–2085, 2012, erratum in: Curr Biol. 2012 Dec 4;22(23):2284. [Online]. Available: https://doi.org/10.1016/j.cub.2012.09.014

[38] F. L. Ribeiro, S. Bollmann, and A. M. Puckett, "Predicting the retinotopic organization of human visual cortex from anatomy using geometric deep learning," *NeuroImage*, vol. 244, p. 118624, 2021, epub 2021 Oct 1. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2021.118624

[39] M. Rosenke, R. van Hoof, J. van den Hurk, K. Grill-Spector, and R. Goebel, "A probabilistic functional atlas of human occipito-temporal visual cortex," *Cerebral Cortex*, vol. 31, no. 1, pp. 603–619, 2021. [Online]. Available: https://doi.org/10.1093/cercor/bhaa246

[40] R. van Hoof, A. Lozano, F. Wang, P. C. Klink, P. R. Roelfsema, and R. Goebel, "Optimal placement of high-channel visual prostheses in human retinotopic visual cortex," *Journal of Neural Engineering*, vol. 22, no. 2, 2025, *Equal contribution. [Online]. Available: https://doi.org/10.1088/1741-2552/adaeef

[41] M. Bak, J. P. Girvin, F. T. Hambrecht, C. V. Kufta, G. E. Loeb, and E. M. Schmidt, "Visual sensations produced by intracortical microstimulation of the human occipital cortex," *Med. Biol. Eng. Comput.*, vol. 28, no. 3, pp. 257–259, May 1990.

[42] J. R. Bartlett, R. W. Doty, sr., B. B. Lee, N. Negrão, and W. H. Overman, jr., "Deleterious effects of prolonged electrical excitation of striate cortex in macaques," *Brain Behavior and Evolution*, vol. 14, no. 1-2, pp. 46–66, 04 2008. [Online]. Available: https://doi.org/10.1159/000125575

[43] M. S. Beauchamp, D. Oswalt, P. Sun, B. L. Foster, J. F. Magnotti, S. Niketeghad, N. Pouratian, W. H. Bosking, and D. Yoshor, "Dynamic stimulation of visual cortex produces form vision in sighted and blind humans," *Cell*, vol. 181, no. 4, pp. 774–783.e5, May 2020.