Sample, Don't Assume: Unconstrained Receptive Field Estimation

Niklas Mueller (n.muller@uva.nl)
Psychology Research Institute, University of Amsterdam

Paolo Papale (p.papale@nin.knaw.nl)
Department of Vision & Cognition, Netherlands Institute for Neuroscience (KNAW)

Iris I. A. Groen (i.i.a.groen@uva.nl)
Institute for Informatics, University of Amsterdam

Abstract

The spatial tuning of (populations of) neurons, as reflected in their (population) receptive field (pRF), is one of the most fundamental properties determining neural responses in visual cortex. pRF geometry is typically modeled as a 2D isotropic Gaussian, effectively assuming the pRF samples a circular 'aperture' in the visual field. However, it has been found that using a more complex geometry can improve neural predictions. Thus, it remains unclear what assumptions to make about the geometry of pRFs. Here, we show that removing any geometrical assumptions, and instead estimating pRFs in a fully data-driven way, leads to significant improvements in neural predictions. We combine linear encoding models with random sampling of pixels from feature maps of convolution deep neural networks to estimate unconstrained pRFs from monkey multi-unit electrophysiology recordings. Our new method not only improves neural predictions but also allows for both quantitative pRF mapping (parameter estimation) and qualitative inspection of the pRF geometry from the obtained pixel importance maps.

Keywords: Population receptive field (pRF) mapping; deep neural networks; monkey electrophysiology; encoding models

Introduction

Neurons in visual cortex respond preferably to a specific location in the visual field: their receptive field (RF). When sampling the activity of multiple neurons at once (e.g. multi-unit activity or BOLD), this is referred to as a population RF (pRF). Simple mathematical models are capable of predicting neural data and estimating the parameters describing the pRFs by assuming an isotropic Gaussian geometry (Dumoulin & Wandell, 2008). However, the true geometrical nature of pRFs in the primate visual cortex is debated. Several models have been proposed, each introducing a mechanism that aims to explain a specific observation in human fMRI (e.g., Aqil et al., 2021; Kay et al., 2013; Lerma-Usabiaga et al., 2020; Silson et al., 2018; Zuiderbaan et al., 2012) or monkey electrophysiology (Klink et al., 2021), by increasing the complexity of the pRF model. It therefore remains unclear what the optimal complexity of the pRF geometry is and whether a Gaussian is the best geometrical estimate.

Here, we introduce a fully data-driven, fast, and model-free approach for estimating the spatial structure of pRFs. We build encoding models using feature maps from convolutional deep neural networks (DNNs) (Bashivan et al., 2019; St-Yves & Naselaris, 2018) to predict electrophysiology data from the macaque ventral stream. Leveraging the 2D structure of convolutional feature maps, we iteratively obtain encoding scores of randomly sampled sets of pixels to estimate the importance of each pixel to encoding performance, yielding a 2D importance map per recording site. This allows us to estimate the pRF with high spatial detail without making assumptions about its underlying geometry. Crucially, we can still estimate conventional pRF parameters from the importance maps.

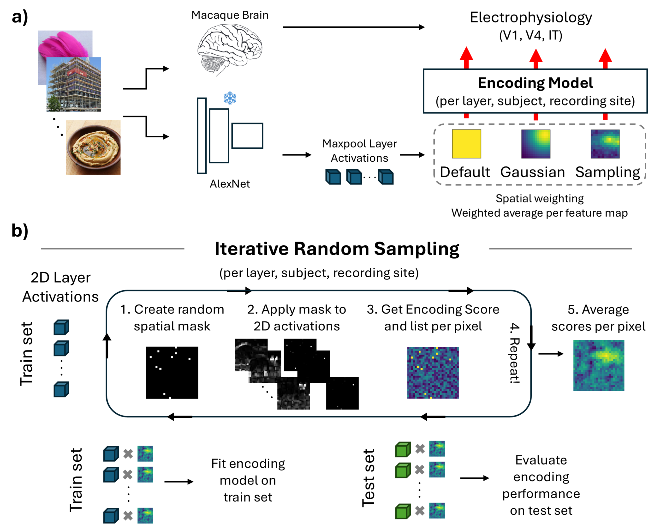

Figure 1: Schematic overview of **a)** linear encoding models using DNN feature maps (here, AlexNet); and **b)** the iterative random sampling method: after fitting the default encoding model, encoding performance is evaluated anew using information from only a random selection of pixels for many iterations. The average encoding score per pixel across iterations yields the sampled pRF.

Methods

Electrophysiology data We used the THINGS ventral stream spiking dataset (TVSD) (Papale et al., 2025) which contains preprocessed, averaged (within a time window per region of interest (ROI) to capture only feedforward processing), and normalized electrophysiology data of areas V1, V4, and IT from two macaques viewing $\sim 22k$ natural images of common objects (Hebart et al., 2019).

DNN feature extraction For each image, resized to 224 x 224 pixels, we extracted the activations from an ImageNet-trained (Russakovsky et al., 2015) AlexNet (Krizhevsky et al., 2012) from the three maxpool layers (features.[2, 5, 12]).

Neural encoding models We built cross-validated linear encoding models for each layer, subject, and recording site separately (**Fig. 1a**). We compare three models, each applying a different spatial weighting (implemented as a weighted average per pixel within each feature map) to the DNN feature maps: (1) *Default*: equal weights for all pixels; (2) *Gaussian*: weights follow a 2D Gaussian distribution; (3) *Sampling*: weights are estimated using random spatial sampling (see below). For each model, we fit a cross-validated linear regression for each recording site, regressing the neural data across training images onto the spatially weighted DNN activations and subsequently calculating Pearson's correlation between the predicted and the measured neural data on test images.

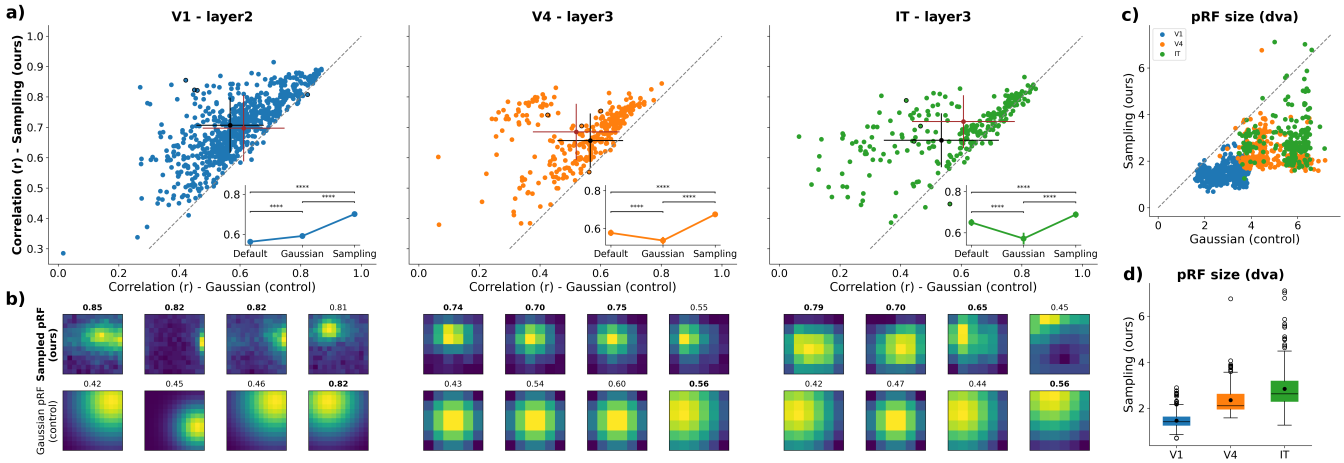

Figure 2: **Randomly sampled receptive fields improve encoding performance a)** Encoding performance per site (for two subjects) when using a 2D Gaussian (x-axis) or the sampled pRFs (y-axis) as spatial weights on the test set, for the best layer per ROI. **b)** Sampled pRFs and best Gaussian pRFs and their encoding scores on the test set (best model in bold) for selected sites. **c)** Estimated pRF sizes per ROI for the Gaussian and Sampling method and **d)** across ROIs for the sampling method.

Random spatial sampling To estimate the pixel-wise importance of the DNN feature maps for the encoding performance per site, we iteratively sampled random subsets of pixels (**Fig. 1b**) where each sampled pixel is weighted by 1 while the other pixels are weighted by 0. For each random sample, we obtain an encoding score (using the fitted regression of the default model) on the training set which we assign back to each pixel included in the sample. After many iterations, we average the assigned scores per pixel, yielding an importance map for each site, which we use as spatial weights, and fit a final regression using these spatial weights.

Gaussian model As a control model, we used isotropic 2D Gaussians as spatial weights. The best fitting (yielding the highest encoding score on training images) parameters (x, y, σ) per site are identified via a grid search on the training set.

Receptive field mapping To estimate the pRF parameters for our sampling method, we performed a grid search over the x, y, and σ parameters of an isotropic 2D Gaussian for each site, such that the resulting Gaussian had the maximum correlation with the importance map of that site.

Results

Unconstrained pRFs improve encoding performance We find that our random sampling method is able to identify locations in the feature maps (as a proxy for the visual field) that contribute more strongly to the encoding model performance compared to other locations. When using the resulting pRF estimates as spatial weights, our method outperforms the default model (using no spatial sampling) and the Gaussian model (using two-dimensional Gaussians as spatial weights) (see **Fig. 2a**). Both the Gaussian and the sampling models improve over the default model for sites in V1. However, for V4 and IT, the Gaussian model performs worse than the default, and the sampling model improves over both.

Gaussian models overestimate pRF sizes The pRF parameter estimation by the Gaussian control model systematically overestimates the size of the pRFs compared to using our sampling method (see **Fig. 2c**). The resulting sampled pRF sizes increase along the ventral hierarchy (see **Fig. 2d**).

Discussion

We show that dropping the assumption that pRFs have an isotropic Gaussian shape and instead sampling the pRF improves encoding performance drastically when predicting neural responses to natural images in macaque visual cortex. Despite not assuming a Gaussian geometry, our method still allows to estimate the parameters of conventional pRF models. Crucially, without changes to the sampling procedure, the parameters of any given pRF model (e.g. CSS, Kay et al., 2013) can be estimated on the sampled pRFs.

The spatial resolution of the pRF mapping for models using DNN features (St-Yves & Naselaris, 2018; Bashivan et al., 2019, e.g.) is limited by the DNN input size and the convolutions applied in the DNN. However, we note that our method can be adapted to sample pixels directly in the input image, which removes the resolution constraints of the convolutions.

Overall, unconstrained pRF estimation using DNN feature maps improves neural prediction performance in macaque visual cortex. The resulting pRF estimates further allow for fast and flexible parameter estimation of any given pRF model.

Acknowledgments

This work was supported by the Interdisciplinary PhD Program by the Data Science Center of the University of Amsterdam.

References

Aqil, M., Knapen, T., & Dumoulin, S. O. (2021). Divisive normalization unifies disparate response signatures throughout the human visual hierarchy. *Proceedings of the National Academy of Sciences*, *118*(46), e2108713118.

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science*, *364*(6439), eaav9436.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660.

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS one*, *14*(10), e0223792.

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of neurophysiology*, *110*(2), 481–494.

Klink, P. C., Chen, X., Vanduffel, W., & Roelfsema, P. R. (2021). Population receptive fields in nonhuman primates from whole-brain fmri and large-scale neurophysiology in visual cortex. *Elife*, *10*, e67304.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Lerma-Usabiaga, G., Benson, N., Winawer, J., & Wandell, B. A. (2020). A validation framework for neuroimaging software: The case of population receptive fields. *PLoS computational biology*, *16*(6), e1007924.

Papale, P., Wang, F., Self, M. W., & Roelfsema, P. R. (2025). An extensive dataset of spiking activity to reveal the syntax of the ventral stream. *Neuron*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, *115*(3), 211-252. doi: 10.1007/s11263-015-0816-y

Silson, E. H., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2018). Differential sampling of visual space in ventral and dorsal early visual cortex. *Journal of Neuroscience*, *38*(9), 2294–2303.

St-Yves, G., & Naselaris, T. (2018). The feature-weighted receptive field: an interpretable encoding model for complex feature spaces. *NeuroImage*, *180*, 188–202.

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center–surround configurations in population receptive fields using fmri. *Journal of vision*, *12*(3), 10–10.