Foveated sensing in KNN-convolutional neural networks based on isotropic cortical magnification

Nicholas M. Blauch (nblauch@fas.harvard.edu)
Harvard University

Talia Konkle (talia_konkle@harvard.edu
Harvard University

Abstract

Human vision prioritizes the center of gaze through spatially-variant retinal sampling, leading to magnification of the fovea in cortical visual maps. In contrast, deep neural network models (DNNs) typically operate on spatially uniform inputs, limiting their use in understanding the active and foveated nature of human vision. Prior works exploring foveated sampling in DNNs have introduced anisotropy in their attempts to wrangle retinal samples into a grid-like representation, sacrificing faithful cortical retinotopy and creating undesirable warped receptive field shapes that depend on eccentricity. Here, we offer an alternative approach by adapting the model architecture to enable realistic foveated sensing. First, we develop a spatially-variant input sensor derived from the assumption of isotropic cortical magnification. Second, as this produces a curved sensor manifold, we devise a novel method for hierarchical convolutional processing that defines receptive fields as k-nearest-neighborhoods on the sensor manifold. This approach allows us to build hierarchical KNN convolutional neural networks (KNN-CNNs) closely matched to their CNN counterparts. Architecturally, these models have more realistic cortical retinotopy and desirable receptive field properties, such as increasing size and approximately constant shape as a function of eccentricity. Training foveated KNN-CNNs end-to-end over natural images on a categorization task, we find that they provide improved performance relative to non-foveated CNNs when retinal resources are constrained relative to the native image resolution. Moreover, they exhibit increasing performance with multiple fixations that encode different parts of the image in high-resolution. Broadly, this model class offers a more biologically-aligned sampling of the visual world, enabling future computational work to model the active and spatial nature of human vision, and to build more neurally mappable models.

Introduction

The primate retina samples visual information most densely in the fovea —- corresponding to the center of gaze – where color-sensitive cones and ganglion cells are heavily concentrated. Sampling drops off quickly toward the periphery. This space-variant retinal sampling is thought to be closely related to the cortical magnification factor (CMF), which gives the extent of visual cortex spanned by a constant extent of the visual field, at each point in the visual field. In their seminal work, Daniel and Whitteridge (1961) found that the CMF decreases sharply with eccentricity, but is roughly constant at all points in the visual field of constant eccentricity. This is known as *isotropic* cortical magnification (see also Schwartz (1980); Rovamo and Virsu (1984)), as the sampling rate is the same at a given point regardless of which direction it is measured.

Many attempts have been made to model foveation in computer vision (Wang, Mayo, Deza, Barbu, & Conwell, 2021; Da Costa, Kornemann, Goebel, & Senden, 2024; Jérémie, Daucé, & Perrinet, 2024), demonstrating some intriguing benefits. However, these approaches – whether using a warped Cartesian space (Wang et al., 2021; Da Costa et al., 2024), or a log-polar image model (Jérémie et al., 2024) – introduce anisotropic sampling across space in their attempts to produce a rectangular, grid-like output image suitable for standard computer vision. This anisotropy is dependent on eccentricity, and the shape of unit receptive fields computed in these spaces thus varies with eccentricity, which is particularly troubling for convolutional architectures. Moreover, sampling rate has generally not been well tuned to the native image resolution, leading to oversampling or blurring of the fovea.

Method

Foveation with isotropic cortical magnification

Here, we address these challenges by sampling visual space with isotropic cortical magnification, drawing on the mathematical models of Schwartz (1980) and Rovamo and Virsu (1984). First, given the cortical magnification function $M(r) = 1/(r+a)$, we sample the image from the fovea to the periphery (radially) equally in the logarithmic dimension given by the CMF integral ($w = \log(r+a)$). Second, we determine the number of equally spaced samples to draw in a circle at each radius in order to preserve local isotropy; that is, ensuring that the distance between neighboring angles is equal to the distance between neighboring radii at any given point. This ensures locally consistent spacing (isotropy) throughout the visual field, while achieving magnification along the radial dimension. Together, this sampling strategy produces points that are approximately evenly distributed in the sensor manifolds ("cortical" space) described by the models of Schwartz (1980) and Rovamo and Virsu (1984) (Figure 1B-C). The result of this process is a sensor that maps a standard (rectangular) image into a curved manifold (Figure 1B), which can be flattened into two hemifield representations (Figure 1C).

KNN-convolutional networks

Standard convolutional neural networks expect a rectangular image grid (e.g. 224 x 224 pixels) and can thus not easily operate over the curved sensor manifold. To address this issue, we designed a modified convolutional architecture. Here, filters are not specified as n x n x c kernels, but instead as k x c kernels, where the k samples are drawn spatially as the k-nearest points in the sensor manifold of the previous layer (Figure 1D). A hierarchical network can be constructed by assigning each layer the same sampling function, with a progressively decreasing resolution. At each layer, each unit's receptive field is defined by finding the k-nearest-neighbors in the sensor manifold (Figure 1D) of the previous layer. This allows us to create KNN-convolutional neural networks (KNN-CNNs) that are resource matched with an arbitrary CNN, with the same number of layers, parameters, channels, and feature map locations per layers, along with matched pooling.

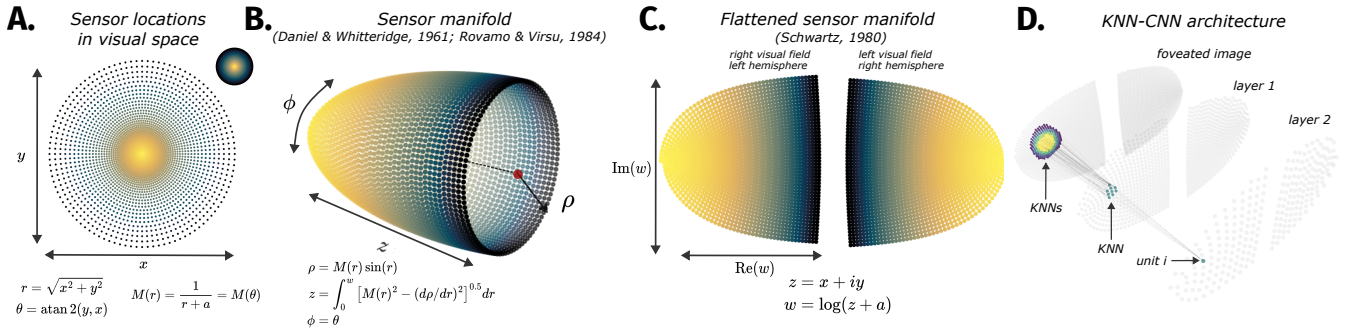

Figure 1: Foveated sensing via isotropic cortical magnification in KNN-convolutional neural networks. **A.** Sensor locations in visual space arising from isotropic sampling according to the cortical magnification function $M(r) = \frac{1}{r+a} = M(\theta)$ (Schwartz, 1980). We set $a = 1$ and a field-of-view of 16 degrees. **B.** The manifold of equally spaced sensor locations (Rovamo & Virsu, 1984). **C.** The complex log model serves as a flattened version of the sensor manifold (Schwartz, 1980). **D.** A visualization of the KNN-CNN architecture, showing the construction of a receptive field in the second layer. Note that KNNs are defined in the curved manifold space (**B**), but we plot the flat space to visualize the full manifold.

Results

We built a KNN-CNN matched to the classic AlexNet architecture (Krizhevsky, Sutskever, & Hinton, 2012), and trained both models on 1000-way ImageNet categorization (Deng et al., 2009) using the standard 224x224 sampling resolution on images. AlexNet achieved an accuracy of 57.7%, whereas our matched KNN-CNN reached an accuracy of 54.4%. Given that images were first resized to have their shortest side be 256 (the approximate native resolution), foveated sensing at a 224x224 resolution involves repeated sampling of the same pixels near the center of gaze. Thus, it is perhaps surprising that the KNN-CNN model still performs close to the non-foveated CNN, given that oversampling the fovea necessitates undersampling the periphery. We predicted foveation would become more useful when operating over a much higher resolution input than the sensor, as in the ambient light field in the real world. Thus, in the next experiment, we simulated this scenario by giving both a KNN-CNN and a matched-CNN the same constrained "pixel-budget" (64x64) to work with over a large field-of-view (90% of the image area). Here, we trained on a 100-category subset of ImageNet. During training, the model used 1 random fixation; at evaluation, we allowed each model to aggregate information by averaging category logits over 20 random fixations. We plot performance over the number of fixations (Figure 2A). This reveals a large advantage for our foveated KNN-CNN in this resource-constrained comparison.

After training, the first layer of the KNN-CNN yields orientation-tuned filters that, due to the architectual design, naturally increase in visual size with eccentricity, while being a constant size on the sensor manifold (Figure 2B-C), in line with empirical data (Dumoulin & Wandell, 2008; Motter, 2009).

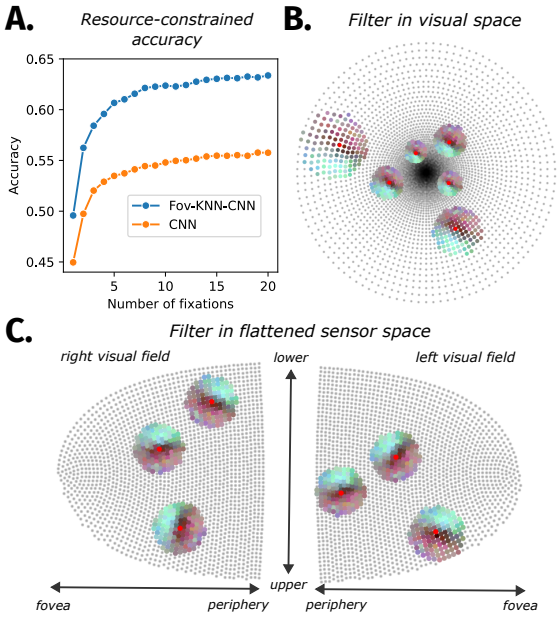

Figure 2: **A**. Fov-KNN-CNN outperforms standard CNN under strong resource constraints, and improves with more fixations. **B.** A convolutional kernel from a trained Fov-KNN-CNN plotted at various corresponding points in visual space. **C.** The same filter plotted in the corresponding flattened sensor space.

Discussion

This work presents a novel approach to foveated neural network processing based on cortical magnification. Our sensor ensures locally isotropic eccentricity-dependent sampling, while our KNN-CNN model allows for hierarchical convolutional processing, yielding biologically realistic receptive field properties. Our experiments reveal an advantage for foveation in our model relative to uniform sensing in a standard CNN when each is constrained to sense less pixels than are available in the environment. Our architecture provides a foundation for future computational work to better model the active and spatially-variant nature of human vision, with enhanced mappability to retinotopic brain areas.

Acknowledgments

We acknowledge support from NSF grant 2309041 to T.K, and gratitude to Brad Motter, Sarthak Chandra, George Alvarez, and the members of the Harvard Vision Lab for helpful conversations and feedback on this work.

References

Da Costa, D., Kornemann, L., Goebel, R., & Senden, M. (2024, April). Convolutional neural networks develop major organizational principles of early visual cortex when enhanced with retinal sampling. *Scientific Reports*, *14*(1), 8980. doi: 10.1038/s41598-024-59376-x

Daniel, P. M., & Whitteridge, D. (1961, December). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, *159*(2), 203–221. doi: 10.1113/jphysiol.1961.sp006803

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR* (pp. 248–255). IEEE. doi: 10.1109/CVPR.2009.5206848

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. doi: 10.1016/j.neuroimage.2007.09.034

Jérémie, J.-N., Daucé, E., & Perrinet, L. U. (2024). Retinotopic Mapping Enhances the Robustness of Convolutional Neural Networks. *arXiv*. doi: 10.48550/arxiv.2402.15480

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007

Motter, B. C. (2009, May). Central V4 Receptive Fields Are Scaled by the V1 Cortical Magnification and Correspond to a Constant-Sized Sampling of the V1 Surface. *The Journal of Neuroscience*, *29*(18), 5749–5757. doi: 10.1523/JNEUROSCI.4496-08.2009

Rovamo, J., & Virsu, V. (1984, January). Isotropy of cortical magnification and topography of striate cortex. *Vision Research*, *24*(3), 283–286. doi: 10.1016/0042-6989(84)90133-0

Schwartz, E. L. (1980, January). Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. *Vision Research*, *20*(8), 645–669. doi: 10.1016/0042-6989(80)90090-5

Wang, B., Mayo, D., Deza, A., Barbu, A., & Conwell, C. (2021, December). *On the use of Cortical Magnification and Saccades as Biological Proxies for Data Augmentation* (No. arXiv:2112.07173). arXiv.