Out-of-Distribution Generalization under Augmented Stimuli Reveals the Inductive Bias of Visual Cortex Digital Twins

Ayumi Kasagi[1*†], Takemi Hieda[1*], Yuki Hosaka[1], Ruixiang Li[1], Masato Taki[2, 3], Teppei Matsui[1†]

[1]Graduate School of Brain Science, Doshisha University, Kyoto, Japan
[2]Graduate School of Artificial Intelligence and Science, Rikkyo University, Tokyo, Japan
[3]AI as Science Team, RIKEN iTHEMS, Wako, Saitama, Japan
tematsui@mail.doshisha.ac.jp, akasagi@mail.doshisha.ac.jp

Abstract

An important goal in Neuro-AI is to develop a digital twin of the visual cortex. Currently, state-of-the-art models of the visual cortex require large amounts of training data, which are difficult to obtain for most neuroscience laboratories. Here, we propose an approach to alleviate this limitation by enhancing neuronal data quality through optimization of visual stimuli. We first evaluated various image-transformation methods *in silico* using CNN-based models of the mouse visual cortex. We then validated the selected methods *in vivo* using real mouse brain recordings. The *in vivo* experiments identified two methods that enhanced neuronal responses and accelerated the training of digital twins. Unexpectedly, one method (Sharpening) consistently failed to match the *in silico* predictions. This discrepancy was likely due to CNN's inductive bias toward high spatial frequencies. Consistently, models of the visual cortex with more favorable spectral sensitivity successfully predicted *in vivo* neuronal responses to Sharpening-transformed images. Taken together, our work makes the following contributions toward the development of a digital twin of the visual cortex: 1) Two *in vivo*-validated image-transformation methods that enhance data quality and accelerate model training. 2) Evidence that the RNN-based model is more aligned with the real visual cortex than CNN- or ViT-based models.

Introduction

Development of an accurate functional model ("digital twin") of the brain is a powerful approach in Neuro-AI toward achieving a novel mechanistic understanding of biological information processing (Doerig et al. 2023). Deep neural networks (DNNs) are the key technology for building state-of-the-art (SOTA) models of the brain based on real neuronal data (Lindsay 2021). However, due to their DNN backbones, SOTA models of the brain require large amounts of training data. In the case of visual cortex models, this requirement is further exacerbated by the weak and sparse neuronal activity elicited by natural images (Vinje and Gallant 2000; Yoshida and Ohki 2020).

To compensate for the low reliability of neuronal activity, current SOTA models of the visual cortex rely on massive experimental datasets for training (Yamins et al. 2014; Walker et al. 2019; Allen et al. 2022; Tong et al. 2023; Du et al. 2025; Papale et al. 2025; Wang et al. 2025). For example, DNN-based models of the mouse visual cortex are typically trained on the activity of tens of thousands of neurons in response to several thousand visual stimuli collected across multiple animals (Walker et al. 2019; Willeke et al. 2022; Tong et al. 2023; Du et al. 2025). For most neuroscience laboratories, acquiring such large-scale datasets is often impractical.

A promising alternative approach is to reduce the data requirement by improving data quality. Cowley and Pillow demonstrated such an approach with a novel image-transformation method ("Gaudy") that enhances the reliability of neuronal responses relative to the original, non-transformed natural scenes (Cowley and Pillow 2020). In the present study, we extend this approach by generalizing the method to grayscale natural scenes and adding *in vivo* validation. Extending the method to grayscale images is important because grayscale stimuli are widely used in most neuroscience studies of the mouse visual cortex (Walker et al. 2019; Yoshida and Ohki 2020; Du et al. 2025) and in many studies of the monkey visual cortex (Vinje and Gallant 2000; Yamins et al. 2014; Chang and Tsao 2017).

Figure 1 describes the overview of this work. We first screened a set of image-transformation methods using *in silico* models of the mouse primary visual cortex (V1)(*In Silico* Evaluation). Based on *In Silico* Evaluation, we identified candidates of image-transformation methods that can enhance the quality of neuronal data (Candidate Methods). We then validated the candidate methods *in vivo* using two-photon calcium imaging in the real mouse V1 (*In Vivo* Validation). While many *in silico* predictions were validated *in vivo*, we also found unexpected discrepancies between *in silico* and *in vivo* results, which led us to identify a previously unrecognized bias in commonly used models of the mouse V1.

Our main contributions are as follows:

- *In silico* experiments using convolutional neural network (CNN)-based models of the mouse V1 identified four image-transformation methods that can enhance neuronal responses to grayscale natural scenes.
- *In vivo* experiments in the real mouse V1 validated two of

[*]co-1st authors
[†]co-corresponding authors

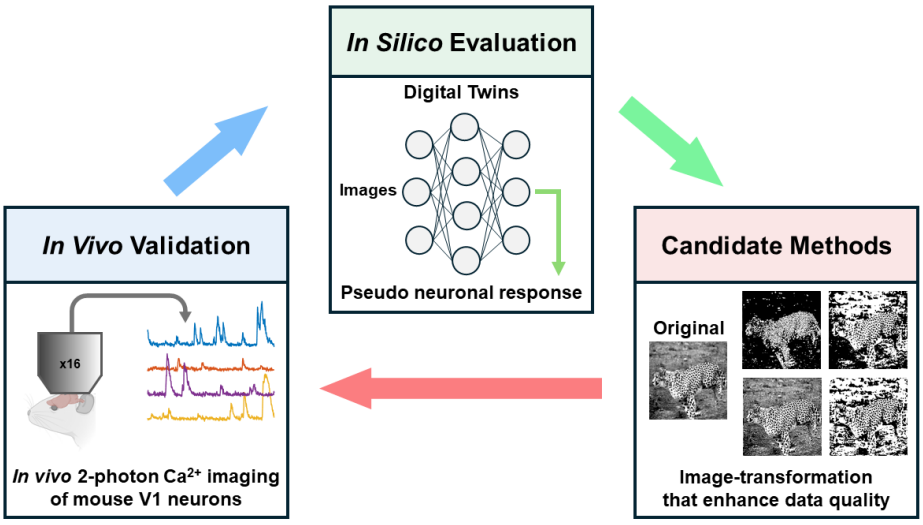

Figure 1: **Overview**

these methods (Diff Gauss and Gaudy) that can enhance neuronal data quality and accelerate the training of digital twins.

- *In vivo* results for one method (Sharpening) consistently contradicted the *in silico* predictions made by CNN-based models.
- The contradictory results for Sharpening were likely due to CNN's inductive bias toward high spatial frequencies, which was mitigated in ViT- and RNN-based models.
- The RNN-based model best predicted the *in vivo* results and is recommended as the base architecture for developing a digital twin of the visual cortex.

Methods

Image-transformation methods for grayscale natural images

A previous study demonstrated that increasing the contrast saturation of colored images can enhance neural responses and reduce the amount of data required to train a DNN-based model of the visual cortex (Cowley and Pillow 2020). Building upon this work, we prepared a set of 12 image-transformation methods applicable to grayscale images (Fig. 2). Eleven of these methods were adopted from standard image-processing libraries in MATLAB and OpenCV: CLAHE, Diff Gauss, Gamma Correction, Highpass Filter, Histeq, Imadjust, Linear Transformation, Local Threshold, Log Transformation, Retinex, and Sharpening. In addition, we included the Gaudy transformation introduced in the previous study, applied here to a single channel to perform contrast saturation (Cowley and Pillow 2020).

Mouse V1 models (digital twins)

We considered four models of the mouse primary visual cortex (V1): an in-house CNN-based model ("in-house CNN"),

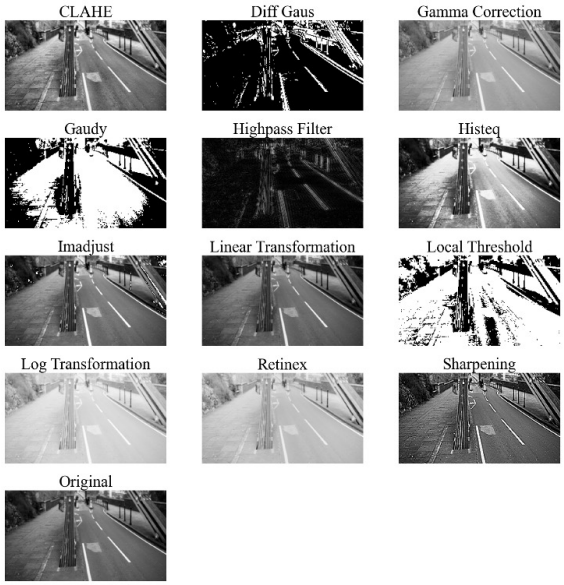

Figure 2: **Image-Transformation Methods**

a publicly available CNN-based model (Sensorium CNN), a Vision Transformer (ViT)-based model, and a recurrent neural network (RNN)-based model. Because these models were trained to reproduce actual mouse V1 responses, they can generate pseudo-neuronal activity patterns in response to previously unseen images, effectively serving as simulators of cortical responses.

- **In-house CNN** consists of two main components (Lurz et al. 2021; Tong et al. 2023): a core CNN block—shared across all neurons—primarily based on convolutional operations, and a readout layer that computed neuron-specific weighted sums of receptive fields from the ex-

tracted feature maps to output scalar neuronal activity values. The model was trained on a large-scale neuronal activity dataset from the Sensorium2022 competition (Willeke et al. 2022), which consists of V1 neuronal activities evoked by thousands of natural images recorded from head-fixed mice.

- **Sensorium CNN** has a CNN-based architecture described in (Willeke et al. 2022) and serves as the baseline model for the Sensorium2022 competition. We obtained the model code and pretrained weights from the official competition website (https://github.com/vlgiitr/Sensorium-2022).
- **ViT-based model** is a model of the mouse V1 based on a Vision Transformer (ViT) architecture (V1T; (Li et al. 2023)). It was trained on the same dataset provided by the Sensorium2022 competition. The model code and pretrained weights were obtained from the authors' website (https://huggingface.co/bryanlimy/V1T).
- **RNN-based model** is a recurrent CNN developed as a foundational neural network for modeling the mouse visual cortex (Wang et al. 2025). It was trained to predict mouse V1 responses to short segments of natural movies. The model code and pretrained weights were obtained from the authors' website (https://github.com/cajal/fnn).

In silico experiments for screening of image-transformation methods

We conducted *in silico* experiments to identify image-transformation methods that can enhance neuronal responses in the model mouse V1. In the initial screening of the image-transformation methods (Fig. 4), we used in-house CNN and Sensorium CNN, because CNNs are the most widely used architecture for modeling the visual response to static images in the mouse V1 (Walker et al. 2019; Willeke et al. 2022; Tong et al. 2023; Du et al. 2025). We also tested ViT- and RNN-based models subsequently (Fig. 8). For the visual stimulus, we used grayscale natural images provided in the Sensorium2022 dataset. Simulated neuronal responses of the CNN- and ViT-based models were obtained by feeding the natural image or its transformed-counterpart to the models. For the RNN-based model, which requires a movie as input, natural images were presented for 1 sec per image separated by 2 sec gray screens. Simulated neuronal response to each image was calculated by taking the mean of the activity during the presentation of the image.

Two metrics were defined to quantify the effects of image-transformations on the simulated responses. Δresponse measures the change in the overall strength of the simulated responses between original and transformed images, and calculated using the following equation:

$$\Delta\text{response}_u = \frac{r_{c,u} - r_{n,u}}{r_{n,u}}$$

where u denotes the neuronal unit, r_c and r_n are the model's mean outputs across transformed and original images, respectively.

Δvariance quantifies across-image variance of the simulated responses, indicating the dynamic range of the simulated responses. Δvariance was calculated using the following equation:

$$\Delta\text{variance}(u) = \frac{\text{Var}(r_c(u))}{\text{Var}(r_n(u))} - 1$$

In vivo experiments in mouse V1

Animals. All animal experiments were approved by the Institutional Animal Care and Use Committee of Doshisha University. Eleven C57BL/6J mice (P33-38, both sexes) were used. Mice were group-housed under a 12-hour light/-dark cycle with ad libitum access to food and water.

Virus injection and cranial window surgery. Mice were anesthetized throughout all surgical procedures by intraperitoneal injection of a mixture of midazolam, medetomidine, and butorphanol tartrate. Body temperature was maintained using a heating pad. Adeno-associated virus (AAV) carrying the genetically encoded calcium indicator GCaMP8m (Zhang et al. 2023) under the synapsin promoter was injected into the binocular region of the left V1. Two weeks after injection, and 1–3 days prior to imaging, a metal headplate was implanted and a craniotomy (4-mm diameter) was made over the virus-injected area.

***In vivo* two-photon calcium imaging and preprocessings.** Mice were head-fixed on a stereotaxic stage and placed under a two-photon microscope equipped with a 16× objective (Nikon, Japan). Time-series images were acquired at 15 Hz from cortical depths of 150–200 μm below the pial surface (layer 2/3). The field of view (FOV) measured 800 × 800 μm. An infrared camera recorded facial and body movements during imaging at a frame rate of 15 Hz.

Visual stimuli were displayed on a gamma-corrected LCD monitor positioned 20 cm in front of the mouse, covering the binocular visual field. The binocular region was targeted for its high spatial resolution (van Beest et al. 2021). Visual stimulus presentation was controlled using PsychoPy (Peirce et al. 2019). The original natural images were obtained from the Allen Brain Observatory dataset (de Vries et al. 2020) and transformed using one of the selected image-transformation methods.

At the beginning of each imaging session, we mapped orientation selectivity (square-wave drifting gratings, 12 directions) to assess the quality of visual responses within the FOV (Matsui and Ohki 2013). When robust orientation tuning was confirmed, experiments using natural images were performed. Each original or transformed image was presented for 1 sec, followed by a 2 sec gray-screen inter-stimulus-interval. All images were presented 10 times in pseudo-random order. In each FOV, responses were recorded for the set of original images and one or two types of transformed images (Fig. 3A).

Acquired time-series images were preprocessed using Suite2P (Pachitariu et al. 2016). Preprocessing included motion correction, extraction of neuronal time series, neuropil subtraction, and deconvolution to estimate spike rates (Fig. 3B-C).

***Analysis of in vivo visual responses*.** For each trial, the visual response was defined as the mean spike rate during the image-presentation period. Unlike the *in silico* neuronal

responses, *in vivo* neuronal responses have trial-to-trial variability. We therefore quantified the reliability of the visual response for each neuron (Tong et al. 2023). Specifically, we calculated across-image correlations of mean visual responses to even and odd presentations and then converted the correlations to the Spearman-Brown reliability. Reliability scores were computed separately for original and transformed images. A neuron was classified as reliably responsive if its reliability score exceeded 0.5 for either the original or transformed images. Only reliably responsive neurons were included in comparisons between responses to original and transformed images.

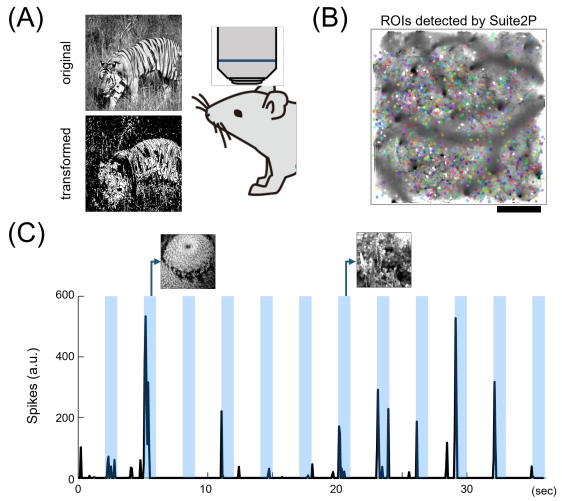

Figure 3: ***In vivo* two-photon calcium imaging in the mouse V1** (A) Neuronal activities from mouse V1 were recorded while the mouse viewed either original or transformed images. (B) Example FOV showing neuronal ROIs detected by Suite2P. Scale bar, 200 µm. (C) Example activity timecourse of a neuron. Shaded regions indicate stimulus-ON periods (1 sec) separated by stimulus-OFF periods (2 sec) showing the gray screen. Two example stimuli are shown.

Analysis of spatial frequency sensitivity

Profiles of sensitivity to spatial frequency were calculated for all models (in-house CNN, Sensorium CNN, ViT, and RNN) as well as for the real mouse V1. For all models, we followed the procedure of (Subramanian et al. 2023) and generated sinusoidal gratings with logarithmically spaced spatial frequencies ranging from 1 cycle per image-width (cpi) up to Nyquist limit ($\approx \min(H, W)/2$). Each grating was assigned a random orientation (0–π) and random phase (0–2π). For each frequency-bin, N stimuli (typically $N \in [200, 1000]$) were synthesized, using input sizes appropriate for each model. For the RNN-based model, the per-trial response was defined as the mean activity during the stimulus epoch minus the mean activity during the preceding baseline period. For each stimulus, the population response was quantified as the root-mean-square (RMS) of responses across neurons.

$$R_{\mathrm{rms}} = \sqrt{\frac{1}{U}\sum_{u=1}^{U} r_u^2}.$$

Within each frequency bin, we computed the mean population response and its standard error of the mean (SEM) across stimuli, yielding frequency–response curves. For the real mouse V1, we used neuronal data provided by (Han, Vermaercke, and Bonin 2022). From this dataset, we curated 5,612 reliably responding V1 neurons with well-estimated preferred spatial frequencies. The histogram of preferred spatial frequencies was then used as the profile of spatial frequency sensitivity for the mouse V1.

Results

In silico experiments identified four image-transformation methods with enhanced neuronal responses

Figure 4 summarizes the results of the *in silico* experiments comparing the neuronal responses of CNN-based models to original and transformed natural images. Most image-transformation methods increased the mean neuronal responses to transformed images (11 out of 12 methods for in-house CNN and 8 out of 12 for Sensorium CNN; Fig. 4A). Among these, four methods—Diff Gauss, Gaudy, Local Threshold, and Sharpening—produced particularly large increases compared with the other methods.

We also examined the dynamic range of the *in silico* visual responses by quantifying the variance of the responses across images (Fig. 4B). Notably, the same four methods that produced the largest increases in mean response also yielded substantial gains in response variance.

Next, following the procedure of a previous study(Cowley and Pillow 2020), we examined whether image transformation facilitates model training for the visual cortex. Five out of the twelve methods improved both the final prediction accuracy and the speed of convergence relative to those obtained with the original images (Fig. 5). Notably, four of these five methods corresponded to the same four methods that produced the largest improvements in *in silico* neuronal responses. This finding is consistent with previous reports showing that increases in neuronal response magnitude enhance the efficiency of model training (Cowley and Pillow 2020).

Taken together, the *in silico* experiments identified four image-transformation methods—Diff Gauss, Gaudy, Local Threshold, and Sharpening—as promising candidates for accelerating the construction of a digital twin of the mouse V1. The next section describes *in vivo* validation of these four methods in the real mouse brain.

In vivo validation of selected image-transformation methods

The *in silico* experiments identified four image-transformation methods that were predicted to elicit stronger and more reliable visual responses than non-transformed (original) images. To test these predictions

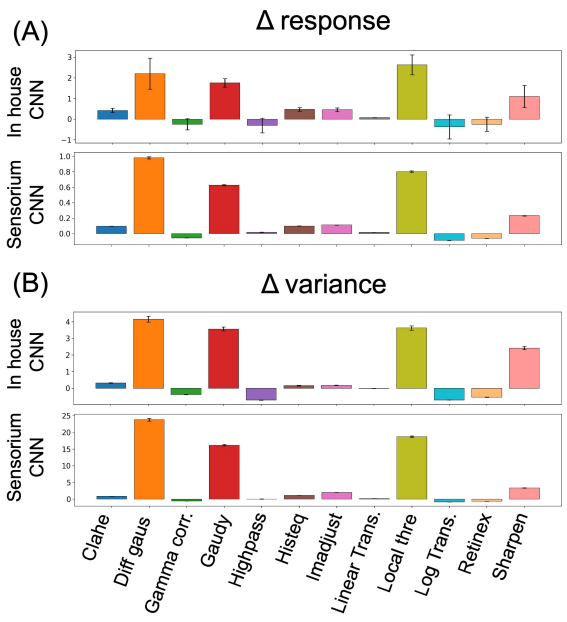

Figure 4: ***In silico* responses to transformed images relative to those of the original images** Δ response and Δ variance for two CNN models across 12 transformed image datasets. Error bars, standard error of the mean (SEM) across model neurons. In both models, the "Difference of Gaussian," "Gaudy," "Local Thresholding," and "Sharpen" datasets exhibited relatively high values, indicating that these transformations enhanced model responses in the *in silico* CNN experiments.

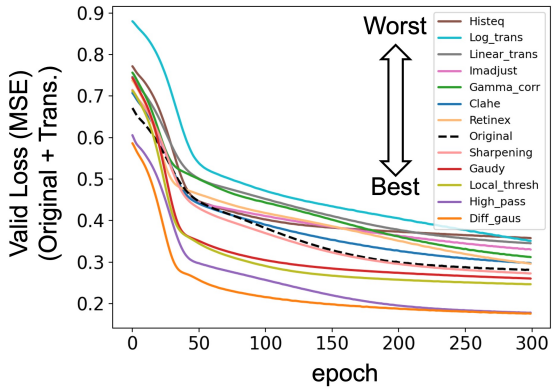

Figure 5: **Candidate methods promoted faster model learning** A ResNet-based model for neuronal activity prediction (Cowley and Pillow 2020) was trained to predict pseudo–V1 neuronal responses generated by the in-house CNN model. The same number of original and transformed images were sampled for training as in the previous study (Cowley and Pillow 2020)

in vivo, we recorded the activity of mouse V1 neurons using two-photon calcium imaging (Fig. 3). Representative examples of *in vivo* neuronal responses to transformed and original images are shown in Figure 6A. Diff Gauss and Gaudy produced higher mean responses and greater across-image variances for transformed than for original images, consistent with the *in silico* predictions (Fig. 6B). Local Threshold showed no significant increase in mean response and a slight decrease in variance. Sharpening, unexpectedly, resulted in significant decreases in both mean response and variance, contradicting the *in silico* predictions.

To further assess the impact of image-transformation on neuronal data quality, we quantified the fraction of reliably responsive neurons per FOV (Fig. 6C). Diff Gauss and Gaudy showed robust increases in the fraction of reliable neurons (increased in 5/5 and 4/5 FOVs, respectively). Local Threshold yielded mixed results (increased in 3/5 FOVs). In contrast, Sharpening consistently decreased the fraction of reliable neurons across all FOVs (5/5), again contradicting the *in silico* results. Taken together, the *in vivo* validation suggests that Diff Gauss and Gaudy are promising candidates for improving neuronal data quality.

Notably, the *in vivo* results for Sharpening consistently diverged from the *in silico* predictions, indicating that the CNN-based models used in the *in silico* experiments failed to generalize to out-of-distribution (OOD) images generated by this transformation. In the following section, we investigate the cause of this discrepancy, which led us to identify a possible effect of inductive bias in CNNs that affects models of the visual cortex.

Inductive bias of CNN caused the gap between the real mouse V1 and the CNN-based models

Why did the *in vivo* results for Sharpening contradict the *in silico* predictions? Comparisons between transformed and original images revealed that Sharpening specifically increased high spatial frequency information, unlike the other three methods, which affected a broader range of spatial frequencies (Fig. 7A). Changes in relative power spectra further showed that Sharpening selectively enhanced the power of high-frequency components, whereas the other three methods increased power across all frequencies (Fig. 7B). Based on these observations, we hypothesized that the contradiction between *in silico* and *in vivo* results arises from differences in spectral sensitivity between CNN-based models and the real mouse V1.

Consistent with this hypothesis, both the in-house CNN and Sensorium CNN exhibited spectral sensitivity profiles peaking at very high spatial frequencies (Fig. 7C), whereas the real mouse V1 was largely insensitive to such high-frequency components (Fig. 7D). Together, these findings support the hypothesis that differences in spectral sensitivity profiles between the real mouse V1 and CNN-based models underlie the contradictory results for Sharpening. Moreover, models with more mouse-like spectral sensitivity profiles are expected to better predict the *in vivo* results for Sharpening. We address this prediction in the next section.

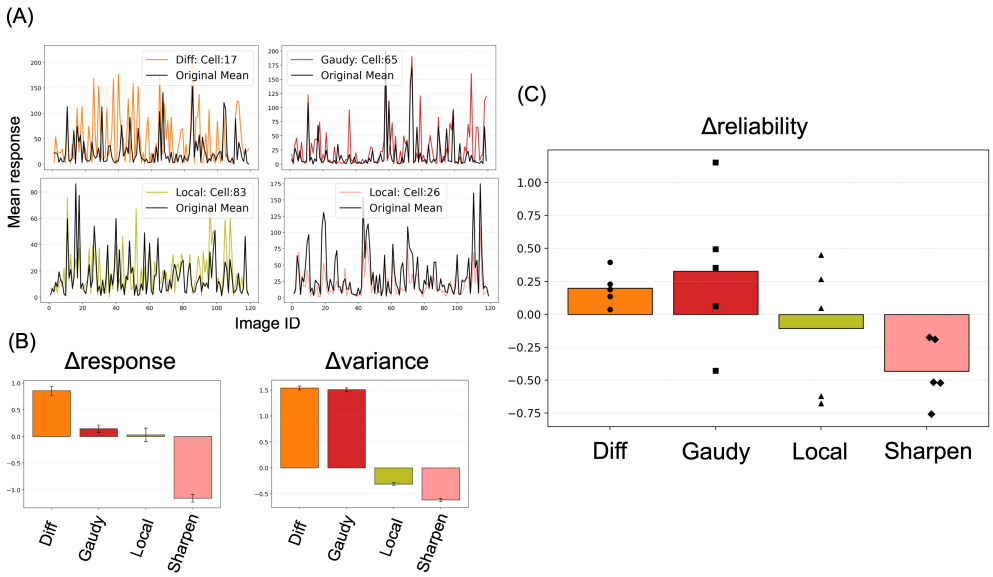

Figure 6: ***In vivo* results for four candidate methods** (A) Example neuronal responses to original and transformed images. (B) Changes of mean response and response variance relative to transformed vs. original images. Errobars, SEM across neurons. (C) Relative change of the fractions of reliable neurons per FOV. Each dot indicates FOV.

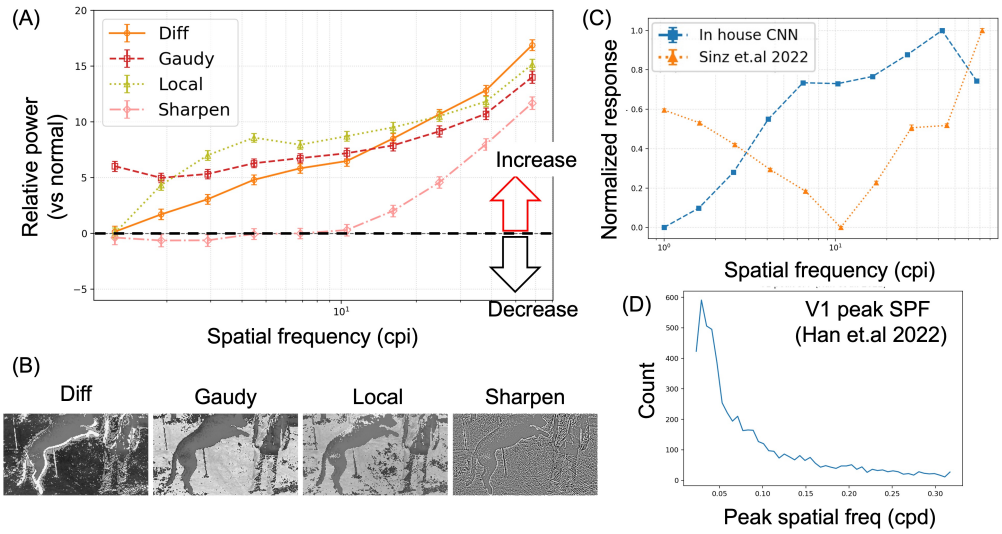

Figure 7: **CNN-based models are biased toward high frequencies** (A) Spatial frequency characteristics from four transformed image datasets that enhanced CNN responses in *in silico* experiments. The vertical axis shows relative power to original images; values above zero indicate enhancement. The sharpen method selectively boosts high-frequency components. Errorbars, SEM across images. (B) Difference images obtained by subtracting originals, showing that only edge features remain for sharpen. (C) Frequency responses of the two CNN models, both showing stronger activity at higher spatial frequencies. (D) Spectral sensitivity profile of the mouse V1 calculated using data from (Han, Vermaercke, and Bonin 2022).

RNN-based model was most aligned with the real mouse V1

If the inductive bias of CNNs caused the discrepancy between the real and modeled mouse V1, a natural solution would be to use architectures that do not strongly rely on convolutional operations. To test this, we examined a ViT-based model (Li et al. 2023) and an RNN-based model (Wang et al. 2025) of the mouse V1. Both models exhibited spectral sensitivity profiles with roughly unimodal shapes centered at low spatial frequencies and reduced sensitivity

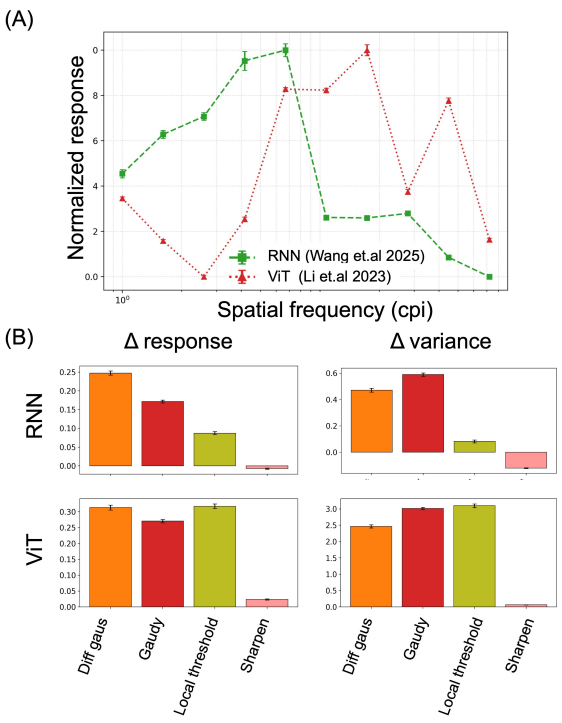

Figure 8: **RNN- & ViT-based models do not suffer from high frequency bias and are more aligned with the real mouse V1** (A) Frequency response profiles of RNN- and ViT-based models trained to predict mouse V1 activity. (B) Model responses to four transformed image datasets used in the *in silico* CNN experiments. Errorbars, SEM across model neurons.

at high frequencies (Fig. 8A). Thus, their spectral sensitivity profiles were more similar to that of the real mouse V1 than those of the CNN-based models. Consistent with this, both ViT- and RNN-based models predicted low neuronal responses for Sharpening (Fig. 8B), in agreement with the *in vivo* results. These findings support the notion that the contradiction between *in silico* and *in vivo* results for Sharpening arose from the inductive bias inherent to CNNs.

Finally, we compared the ViT- and RNN-based models directly. The RNN-based model exhibited a spectral sensitivity profile shifted toward lower spatial frequencies compared with that of the ViT-based model (Fig. 8A), showing a closer match to the real mouse V1. Notably, the ViT-based model predicted a large increase in response variance for Local Threshold, contradicting the *in vivo* results, which showed only minor changes (Fig. 8B). In contrast, the RNN-based model accurately predicted the *in vivo* results across all four image-transformation methods. Taken together, these results indicate that the RNN-based model is preferred over ViT- and CNN-based models as the base architecture for developing a digital twin of the mouse visual cortex.

Discussions

By combining *in silico* simulations and *in vivo* validation, we identified two image-transformation methods—Diff Gauss and Gaudy—that consistently enhanced the quality of neuronal responses to grayscale natural images. These methods may help reduce the amount of data required to train DNN-based models of the visual cortex. Our *in vivo* experiments also served as a test of OOD generalization for visual cortex models. Notably, CNN-based models consistently failed to reproduce the *in vivo* results for Sharpening-transformed images. This finding is striking, as CNNs have become the *de facto* standard for modeling the visual cortex (Lindsay 2021). The failed predictions are likely due to differences in the spectral sensitivity profiles between CNN-based models and the real mouse V1. Importantly, a previous study reported that CNNs are more sensitive to high spatial frequencies than humans (Subramanian et al. 2023), suggesting that our observation is not limited to the mouse V1.

In contrast to CNNs, the RNN-based model exhibited spectral sensitivity most similar to that of the real mouse V1 and best predicted the *in vivo* results. The RNN is also known to be important for modeling the primate ventral visual stream (Kar et al. 2019). Our findings corroborate and extend this observation to the mouse visual cortex, suggesting that even for modeling responses to static images, RNNs are a preferred architecture for constructing models of the visual cortex.

It was somewhat unexpected that Gaudy (Cowley and Pillow 2020) proved effective, given the difficulty of perceiving contrast-saturated grayscale images (Mooney 1957). One possible explanation is that our *in vivo* validation was conducted in the mouse V1, which primarily represents low-level image statistics rather than higher-level structures. Future *in vivo* experiments in higher visual areas (e.g., the primate inferior temporal cortex) may reveal reduced neuronal responses to Gaudy-transformed grayscale images.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Numbers JP25H01550 and JP25K17415 to AK, JP 25K18567 to RL, JP22H05116 to MT, JP24H02331 to TM and JST-PRESTO Grant Number JPMJPR25J3 to AK, JST-CREST Grant Number JPMJCR22N4 to TM and MT, and AMED Grant Number JP25m0625422 to TM and MT.

References

Allen, E. J.; St-Yves, G.; Wu, Y.; Breedlove, J. L.; Prince, J. S.; Dowdle, L. T.; Nau, M.; Caron, B.; Pestilli, F.; Charest, I.; et al. 2022. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1): 116–126.

Chang, L.; and Tsao, D. Y. 2017. The code for facial identity in the primate brain. *Cell*, 169(6): 1013–1028.

Cowley, B.; and Pillow, J. 2020. High-contrast "gaudy" images improve the training of deep neural network models of visual cortex. In Larochelle, H.; Ranzato, M.; Hadsell,

R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems 33*, 21591–21603. Curran Associates, Inc.

de Vries, S. E.; Lecoq, J. A.; Buice, M. A.; Groblewski, P. A.; Ocker, G. K.; Oliver, M.; Feng, D.; Cain, N.; Ledochowitsch, P.; Millman, D.; et al. 2020. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature neuroscience*, 23(1): 138–151.

Doerig, A.; Sommers, R. P.; Seeliger, K.; Richards, B.; Ismael, J.; Lindsay, G. W.; Kording, K. P.; Konkle, T.; Van Gerven, M. A.; Kriegeskorte, N.; et al. 2023. The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 24(7): 431–450.

Du, F.; Angel Núñez-Ochoa, M.; Pachitariu, M.; and Stringer, C. 2025. A simplified minimodel of visual cortical neurons. *Nature Communications*, 16(1): 5724.

Han, X.; Vermaercke, B.; and Bonin, V. 2022. Diversity of spatiotemporal coding reveals specialized visual processing streams in the mouse cortex. *Nature communications*, 13(1): 3249.

Kar, K.; Kubilius, J.; Schmidt, K.; Issa, E. B.; and DiCarlo, J. J. 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience*, 22(6): 974–983.

Li, B. M.; Cornacchia, I. M.; Rochefort, N. L.; and Onken, A. 2023. V1T: large-scale mouse V1 response prediction using a Vision Transformer. arXiv:2302.03023.

Lindsay, G. W. 2021. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, 33(10): 2017–2031.

Lurz, K.-K.; Bashiri, M.; Willeke, K.; Jagadish, A.; Wang, E.; Walker, E. Y.; Cadena, S. A.; Muhammad, T.; Cobos, E.; Tolias, A. S.; Ecker, A. S.; and Sinz, F. H. 2021. Generalization in data-driven models of primary visual cortex. In *International Conference on Learning Representations*.

Matsui, T.; and Ohki, K. 2013. Target dependence of orientation and direction selectivity of corticocortical projection neurons in the mouse V1. *Frontiers in Neural Circuits*, 7: 143.

Mooney, C. M. 1957. Age in the development of closure ability in children. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 11(4): 219.

Pachitariu, M.; Stringer, C.; Schröder, S.; Dipoppa, M.; Rossi, L. F.; Carandini, M.; and Harris, K. D. 2016. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*.

Papale, P.; Wang, F.; Self, M. W.; and Roelfsema, P. R. 2025. An extensive dataset of spiking activity to reveal the syntax of the ventral stream. *Neuron*, 113(4): 539–553.

Peirce, J.; Gray, J. R.; Simpson, S.; MacAskill, M.; Höchenberger, R.; Sogo, H.; Kastman, E.; and Lindeløv, J. K. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1): 195–203.

Subramanian, A.; Sizikova, E.; Majaj, N.; and Pelli, D. 2023. Spatial-frequency channels, shape bias, and adversarial robustness. *Advances in neural information processing systems*, 36: 4137–4149.

Tong, R.; da Silva, R.; Lin, D.; Ghosh, A.; Wilsenach, J.; Cianfarano, E.; Bashivan, P.; Richards, B.; and Trenholm, S. 2023. The feature landscape of visual cortex. *bioRxiv*.

van Beest, E. H.; Mukherjee, S.; Kirchberger, L.; Schnabel, U. H.; van der Togt, C.; Teeuwen, R. R. M.; Barsegyan, A.; Meyer, A. F.; Poort, J.; Roelfsema, P. R.; and Self, M. W. 2021. Mouse visual cortex contains a region of enhanced spatial resolution. *Nature Communications*, 12(1): 4029.

Vinje, W. E.; and Gallant, J. L. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456): 1273–1276.

Walker, E. Y.; Sinz, F. H.; Cobos, E.; Muhammad, T.; Froudarakis, E.; Fahey, P. G.; Ecker, A. S.; Reimer, J.; Pitkow, X.; and Tolias, A. S. 2019. Inception loops discover what excites neurons most using deep predictive models. *Nature neuroscience*, 22(12): 2060–2065.

Wang, E. Y.; Fahey, P. G.; Ding, Z.; Papadopoulos, S.; Ponder, K.; Weis, M. A.; Chang, A.; Muhammad, T.; Patel, S.; Ding, Z.; Tran, D.; Fu, J.; Schneider-Mizell, C. M.; da Costa, N. M.; Reid, R. C.; Collman, F.; Franke, K.; Ecker, A. S.; Reimer, J.; Pitkow, X.; Sinz, F. H.; Tolias, A. S.; and Consortium, M. 2025. Foundation model of neural activity predicts response to new stimulus types. *Nature*, 640(8058): 470–477.

Willeke, K. F.; Fahey, P. G.; Bashiri, M.; Pede, L.; Burg, M. F.; Blessing, C.; Cadena, S. A.; Ding, Z.; Lurz, K.-K.; Ponder, K.; Muhammad, T.; Patel, S. S.; Ecker, A. S.; Tolias, A. S.; and Sinz, F. H. 2022. The Sensorium competition on predicting large-scale mouse primary visual cortex activity. arXiv:2206.08666.

Yamins, D. L.; Hong, H.; Cadieu, C. F.; Solomon, E. A.; Seibert, D.; and DiCarlo, J. J. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23): 8619–8624.

Yoshida, T.; and Ohki, K. 2020. Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. *Nature communications*, 11(1): 872.

Zhang, Y.; Rózsa, M.; Liang, Y.; Bushey, D.; Wei, Z.; Zheng, J.; Reep, D.; Broussard, G. J.; Tsang, A.; Tsegaye, G.; et al. 2023. Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature*, 615(7954): 884–891.