Retrospective on the SENSORIUM 2022 competition

Konstantin F Willeke, Paul G Fahey, Mohammad Bashiri, Laura Hansel, Christoph Blessing, Konstantin-Klemens Lurz, Max F Burg, Santiago A Cadena, Zhiwei Ding, Kayla Ponder, et al.

Retrospective on the SENSORIUM 2022 competition

Konstantin F. Willeke[*,1,3], Paul G. Fahey[*,2], Mohammad Bashiri[1,3], Laura Hansel[3], Christoph Blessing[3], Konstantin-Klemens Lurz[1,3], Max F. Burg[1,3], Santiago A. Cadena[1,3], Zhiwei Ding[2], Kayla Ponder[2], Taliah Muhammad[2], Saumil S. Patel[2], Kaiwen Deng[4], Yuanfang Guan[4], Yiqin Zhu[5], Kaiwen Xiao[5], Xiao Han[5], Simone Azeglio[6,7], Ulisse Ferrari[6], Peter Neri[7], Olivier Marre[6], Adrian Roggenbach[8], Kirill Fedyanin[9], Kirill Vishniakov[10], Maxim Panov[9], Subash Prakash[1], Kishan Naik[1], Kantharaju Narayanappa[1], Alexander S. Ecker[1,3], Andreas S. Tolias[2], Fabian H. Sinz[1,3]

[1] University of Tübingen; [2] Baylor College of Medicine, Houston [3] University of Göttingen; [4] University of Michigan; [5] Tencent AI Lab; [6] Sorbonne University; [7] Ecole Normale Superieure; [8] University of Zurich; [9] Technology Innovation Institute; [10] Mohamed bin Zayed University of Artificial Intelligence

[*] equal contribution

KONSTANTIN-FRIEDRICH.WILLEKE@UNI-TUEBINGEN.DE, PAUL.FAHEY@BCM.EDU, SINZ@CS.UNI-GOETTINGEN.DE

Editors: Marco Ciccone, Gustavo Stolovitzky, Jacob Albrecht

Abstract

The neural underpinning of the biological visual system is challenging to study experimentally, in particular as neuronal activity becomes increasingly nonlinear with respect to visual input. Artificial neural networks (ANNs) can serve a variety of goals for improving our understanding of this complex system, not only serving as predictive digital twins of sensory cortex for novel hypothesis generation *in silico*, but also incorporating bio-inspired architectural motifs to progressively bridge the gap between biological and machine vision. The mouse has recently emerged as a popular model system to study visual information processing, but no standardized large-scale benchmark to identify state-of-the-art models of the mouse visual system has been established. To fill this gap, we proposed the SENSORIUM benchmark competition. We collected a large-scale dataset from mouse primary visual cortex containing the responses of more than 28,000 neurons across seven mice stimulated with thousands of natural images, together with simultaneous behavioral measurements that include running speed, pupil dilation, and eye movements. The benchmark challenge ranked models based on predictive performance for neuronal responses on a held-out test set, and included two tracks for model input limited to either stimulus only (SENSORIUM) or stimulus plus behavior (SENSORIUM+). As a part of the NeurIPS 2022 competition track, we received 172 model submissions from 26 teams, with the winning teams improving our previous state-of-the-art model by more than 15%. Dataset access and infrastructure for evaluation of model predictions will remain online as an ongoing benchmark. We would like to see this as a starting point for regular challenges and data releases, and as a standard tool for measuring progress in large-scale neural system identification models of the mouse visual system and beyond.

Keywords

mouse visual cortex, system identification, neural prediction, natural images

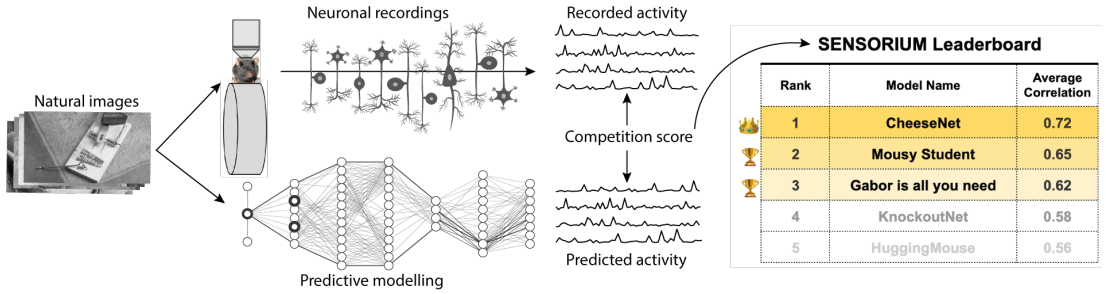

Figure 1: **A schematic illustration of the SENSORIUM competition.** We provide large-scale datasets of neuronal activity in the primary visual cortex of mice. Participants of the competition trained models on pairs of natural image stimuli and recorded neuronal activity.

Introduction

Understanding how the visual system processes visual information is a long standing goal in neuroscience. Neural system identification approaches this problem in a quantitative, testable, and reproducible way by building accurate predictive models of neural population activity in response to arbitrary input. If successful, these models can serve as functional digital twins for the visual cortex, allowing computational neuroscientists to derive new hypotheses about biological vision *in silico*, and enabling systems neuroscientists to test them *in vivo* (Walker et al., 2019; Ponce et al., 2019; Bashivan et al., 2019; Franke et al., 2022). In addition, highly predictive models are also relevant to machine learning researchers who use them to bridge the gap between biological and machine vision (Li et al., 2019; Safarani et al., 2021; Li et al., 2022; Sinz et al., 2019).

The work on predictive models of neural responses to visual inputs has a long history that includes simple linear-nonlinear (LN) models (Jones and Palmer, 1987; Heeger, 1992a,b), energy models (Adelson and Bergen, 1985), more general subunit/LN-LN models (Rust et al., 2005; Touryan et al., 2005; Schwartz et al., 2006; Vintch et al., 2015), and multi-layer neural network models (Zipser and Andersen, 1988; Lehky et al., 1992; Lau et al., 2002; Prenger et al., 2004). The deep learning revolution set new standards in prediction performance by leveraging task-optimized deep convolutional neural networks (CNNs) (Yamins et al., 2014; Cadieu et al., 2014; Cadena et al., 2019) and CNN-based architectures incorporating a shared encoding learned end-to-end for thousands of neurons (Antolík et al., 2016; Batty et al., 2017; McIntosh et al., 2016; Klindt et al., 2017; Kindel et al., 2019; Cadena et al., 2019; Burg et al., 2021; Lurz et al., 2021; Bashiri et al., 2021; Zhang et al., 2018; Cowley and Pillow, 2020; Ecker et al., 2018; Sinz et al., 2018; Walker et al., 2019; Franke et al., 2022).

The core idea of a neural system identification approach to improve our understanding of an underlying sensory area is that models that explain more of the stimulus-driven variability may capture nonlinearities that previous low-parametric models have missed (Carandini et al., 2005). Subsequent analysis of high performing models, paired with ongoing *in vivo* verification, can eventually yield more complete principles of brain computation. This motivates continually improving our models to explain as much as possible of the stimulus-driven variability and analyze these models to decipher principles of brain computations.

Standardized large-scale benchmarks are one approach to stimulate constructive competition between models compared on equal ground, leading to numerous incremental improvements that accumulate to substantial progress. In machine learning and computer vision, benchmarks have been an important driver of innovation in the last ten years. For instance, benchmarks such as the ImageNetChallenge (Russakovsky et al., 2015) helped jump start the revolution in artificial intelligence through deep learning. Similarly, neuroscience can benefit from more large-scale benchmarks to drive innovation and identify state-of-the-art models. This is especially true in the mouse visual cortex, which has recently emerged as a popular model system to study visual information processing, due to the wide range of available genetic and light imaging techniques for interrogating large-scale neural activity.

Existing neuroscience benchmarks vary substantially in the type of data, model organism, or goals of the contest (Schrimpf et al., 2018; Cichy et al., 2021; de Vries et al., 2019; Pei et al., 2021). For example, the **Brain-Score** benchmark (Schrimpf et al., 2018) ranks *task*-pretrained models that best match areas across primate visual ventral stream and other behavioral data, but do not provide neuronal training data. Instead, participants design objectives, learning procedures, network architectures, and input data that result in representations that are predictive of the withheld neural data. The **Algonauts** challenge (Cichy et al., 2021) competition ranks neural predictive models of human brain fMRI visual cortex activity in response to natural images and videos. Additionally, large data releases such as the mouse visual cortex dataset from **Allen Institute for Brain Science** (de Vries et al., 2019) are often not designed for a machine learning competition (consisting of only 118 natural images in addition to parametric stimuli and natural movies), and lack benchmark infrastructure for measuring predictive performance against a withheld test set. Lastly, the **Neural Latents** benchmark (Pei et al., 2021) also targets neuronal response prediction, but for cognitive, somatosensory, and motor areas with a focus on latent variable models.

To fill this gap, we created the `SENSORIUM` benchmark competition to facilitate the search for the best predictive model for mouse visual cortex. We collected a large-scale dataset from mouse primary visual cortex containing the responses of more than 28,000 neurons across seven mice stimulated with thousands of natural images, together with simultaneous behavioral measurements that include running speed, pupil dilation, and eye movements. Benchmark metrics will rank models based on predictive performance for neuronal responses on a held-out test set, and includes two tracks for model input limited to either stimulus only (`SENSORIUM`) or stimulus plus behavior (`SENSORIUM+`).

Our competition was part of the NeurIPS 2022 competition track, receiving 172 model submissions from 26 teams between May 20 and Oct 15, 2022. The winning teams substantially improved our previous state-of-the-art model in both competition tracks (`SENSORIUM`: +13.6%; `SENSORIUM+`: +18%). In this retrospective, we first describe the competition in detail, followed by the results of the competition, with descriptions from the winning teams outlining their approach. Finally, we reflect on the competition results as well as our lessons learned for future iterations.

The `SENSORIUM` Competition

The goal of the `SENSORIUM` 2022 competition and ongoing benchmark is to identify the best models for predicting sensory neural responses to arbitrary natural stimuli. At the start of

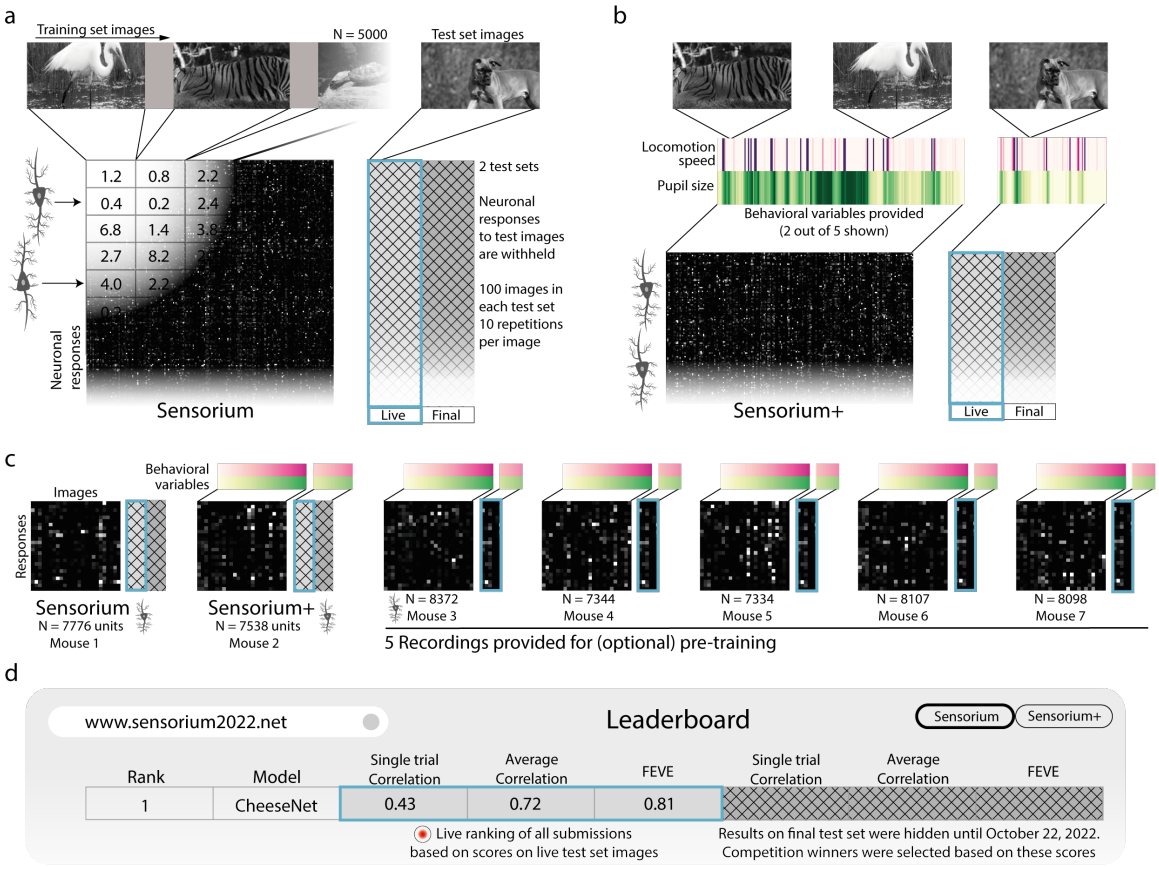

Figure 2: **Overview of the data and the competition structure.** **a**, Single recording from the SENSORIUM track. For ≈ 5000 training images, the neuronal activity of each neuron is provided. For 100 *live* and 100 *final* test set images shown 10 times each, neuronal responses are withheld. **b**, SENSORIUM+ track is the same as (**a**) but behavioral variables are available. **c**, Overview of seven dataset recordings. Five *pre-training* recordings are not part of the competition evaluation, but can be used to improve model performance. In the *public test* set, 100 images are shared with the *live test* set (blue frame), but neuronal responses are provided. **d**, *Live test* scores are displayed on the live leaderboard, while *final test* set scores were only revealed after the submissions closed.

the competition, a training dataset for refining model performance was publicly released (Fig. 2). For two animals, the neuronal responses to a set of competition test set images were permanently withheld. The competition test set images are divided into two exclusive groups: *live* and *final test*. Performance metrics computed on the *live test* images are used to maintain a public leaderboard on our website, while the performance metrics on the *final test* images were only used to score entries after the submission period has ended (Fig. 2d). By separating the *live test* and *final test* set performance metrics, we were able to provide feedback on *live test* set performance to participants wishing to submit updated predictions

(up to one submission per day), while protecting the *final test* set from overfitting over multiple submissions.

The competition has two tracks, `SENSORIUM` and `SENSORIUM+` , predicting two datasets with the same stimuli, but from two different animals and with differing model inputs.

`SENSORIUM`. In the first challenge, participants predict the average neuronal activity of 7,776 neurons in response to 10 repetitions of 200 unique natural images of our competition *live test* and *final test* image sets. The data provided for the test set includes the natural image stimuli but not the behavioral variables (Fig. 2a). Thus, this challenge focuses on stimulus-driven responses, treating other correlates of neural variability, such as behavioral state, as noise. This track resembles most of the current efforts in the community (Schrimpf et al., 2018) to identify stimulus-response functions without additional information about the brain and behavioral state.

`SENSORIUM+` In the second challenge, participants predict the single-trial neuronal activity of 7,538 neurons in response to 200 unique natural images of our competition *live test* and *final test* image sets. In this case, both the natural image stimuli and the accompanying behavioral variables are provided (Fig. 2b). As a significant part of response variability correlates with the animal's behavior and internal brain state (Niell and Stryker, 2010; Reimer et al., 2014; Stringer et al., 2019), their inclusion can result in models that capture single trial neural responses more accurately (Bashiri et al., 2021; Franke et al., 2022).

Data

The competition dataset was designed to compare neural predictive models that capture neuronal responses $\mathbf{r} \in \mathbb{R}^n$ of n neurons as a function $\mathbf{f}_\theta(\mathbf{x})$ of either only natural image stimuli $\mathbf{x} \in \mathbb{R}^{h \times w}$ (image **h**eight,**w**idth), or as a function $\mathbf{f}_\theta(\mathbf{x}, \mathbf{b})$ of both natural image stimuli and behavioral variables $\mathbf{b} \in \mathbb{R}^k$. We provide $k = 5$ variables: locomotion speed, pupil size, instantaneous change of pupil size (second order central difference), and horizontal and vertical eye position. See Fig. 2 and (Willeke et al., 2022) for an overview of the dataset.

Natural images. Natural images from ImageNet (Russakovsky et al., 2015) were cropped to fit a monitor with 16:9 aspect ratio, converted to gray scale, and presented to mice for 500 ms, preceded by a blank screen period between 300 and 500 ms (Fig. 2a,b)

Neuronal responses. We recorded the response of excitatory neurons in layer 2/3 of the right primary visual cortex in awake, head-fixed, behaving mice using calcium imaging. Activity was extracted and accumulated 50 - 550 ms after each stimulus onset.

Behavioral variables. During imaging, mice were head-mounted over a cylindrical treadmill, and an eye camera captured changes in the pupil position and dilation. Behavioral variables were similarly extracted and accumulated 50 - 550 ms after each stimulus onset.

Dataset. Our complete data corpus comprises seven recordings in seven animals (Fig. 2c), including the neuronal activity of over 28,000 neurons to 25,200 unique images, with 6,000–7,000 image presentations per recording (see (Willeke et al., 2022) for details). We report a conservative neuron count estimate, due to multiple segmented units from single neurons appearing in multiple, densely placed calcium imaging recording planes. Fig. 2c shows the uncorrected number of 54,569 units.

Five of the seven recordings, which we refer to as *pre-training recordings* (Fig. 2c, right), are provided solely for training and model generalization, and are not included in the

competition performance metrics. They contain 5,000 single presentations of natural images, randomly intermixed with 10 repetitions of 100 natural images, and all corresponding neuronal responses. The 100 repeated images and responses serve as a *public test* set.

In the two remaining *competition recordings* (Fig. 2c, left), the mice were also presented with 5,000 single presentations of training stimuli as well as the *public test* images. However, during the contest, we withheld the responses to the *public test* images and use them for the live leaderboard. We thus refer to these images as *live test* set. Furthermore, the competition recordings contain 10 repetitions of 100 additional natural *test* images that were randomly intermixed during the experiment. These *test* images are only present in the two competition recordings. The responses to these images were also withheld and used to determine the winner of the competition after submissions are closed (Fig. 2a,b). We refer to these images as our *final test* set. By providing both *live* and *final test* scoring, participants receive the benefit of iterative feedback while avoiding overfitting on the final scoring metrics.

In our first competition track (`SENSORIUM`, Fig. 2a), we withhold the behavioral variables, such that only the natural images can be used to predict the neuronal responses. For the other competition track (`SENSORIUM+`, Fig. 2b), as well as the five pre-training recordings, we are releasing all the behavioral variables. Lastly, we released the anatomical locations of the recorded neurons for all datasets. The complete corpus of data is available to download at https://sinzlab.org/sensorium2022.html.

Performance Metrics

Across the two benchmark tracks, three metrics of predictive accuracy are automatically and independently computed for the 100 *live test* set images and 100 *final test* set images, for which ground-truth neuronal responses are withheld (see (Willeke et al., 2022) for details)

Correlation to Average We calculate the *correlation to average* of 100 model predictions to the withheld, observed mean neural response across 10 repeated presentations of the same stimulus. This metric is computed for both the `SENSORIUM` and `SENSORIUM+` tracks to facilitate comparison. Correlation to average on the *final test* set served as the ultimate ranking score in the `SENSORIUM` track to determine competition winners.

Fraction of Explainable Variance Explained (FEVE) While correlation to average is a common metric, it is insensitive to affine transformations of either the neuronal response or predictions. This metric computes the ratio between the variance explained by the model and the explainable variance in the neural responses (Cadena et al., 2019). The explainable variance accounts for only the stimulus-driven variance and ignores the trial-to-trial variability in responses. This metric is computed for `SENSORIUM` but not `SENSORIUM+` , due to the lack of repeated trials with identical behavior fluctuations necessary to estimate explainable variance. For numerical stability, we compute the FEVE only for neurons with an explainable variance larger than 15% (N=4319 for `SENSORIUM` and N=4548 for `SENSORIUM+`).

Single Trial Correlation Lastly, to measure how well models account for trial-to-trial variations we compute the *single trial correlation* between predictions and single trial neuronal responses, without averaging across repeats. This metric is computed for both the `SENSORIUM` and `SENSORIUM+` tracks to facilitate comparison. Single trial correlation on the *final test* set served as the ultimate ranking score in the `SENSORIUM+` track to determine competition winners.

	Competition results					
	Live test set			Final test set		
	Single trial Correlation	Average Correlation	FEVE	Single trial Correlation	Average Correlation	FEVE
SENSORIUM						
LN Baseline	0.197	**0.363**	0.222	0.207	**0.377 (-28.6%)**	0.232
CNN Baseline	0.274	**0.513**	0.433	0.287	**0.528 (0%)**	0.439
#3: Azeglio et al.	0.307	**0.580**	0.549	0.319	**0.587 (+11.1%)**	0.516
#2: Zhu et al.	0.314	**0.589**	0.512	0.325	**0.598 (+13.2%)**	0.503
#1: Deng et al.	0.316	**0.594**	0.576	0.325	**0.600 (+13.6%)**	0.559
SENSORIUM+						
LN Baseline	**0.257**	0.373	-	**0.266 (-30.7%)**	0.385	-
CNN Baseline	**0.374**	0.571	-	**0.384 (0%)**	0.578	-
#3: Fedyanin et al.	**0.397**	0.605	-	**0.410 (+6.7%)**	0.618	-
#2: Deng et al.	**0.428**	0.643	-	**0.437 (+13.8%)**	0.650	-
#1: Roggenbach	**0.444**	0.625	-	**0.453 (+18.0%)**	0.632	-

Table 1: **Performance of the competition winners of both tracks.** For the final test set, the improvement over the CNN baseline model is shown in percentages.

Baseline

To establish baselines, we trained a simple linear-nonlinear model (LN Baseline) as well as a state-of-the-art convolutional neural network (CNN baseline, Lurz et al., 2021) model for both competitions. For each baseline, we trained a single model (based on one random seed) on only the training data from each competition track (for details, see Willeke et al. (2022)).

Results and Participation

During the four month submission period, 26 teams submitted a total of 172 models (SENSORIUM: 124, SENSORIUM+: 78). To our delight, our state-of-the-art baseline models were outperformed in both tracks by more than 15% (Table 1). We invited the winning teams of each track to describe both their successful and fruitless approaches.

SENSORIUM Rank 1: Deng & Guan

Our winning submission in the SENSORIUM and the SENSORIUM+ track only had minor changes to the SOTA model from the CNN Baseline (Lurz et al., 2021). In the core, we added a convolutional layer whose kernel size and strides were 4 before the first layer to replace the scaling operation in the original model. We set the channel number as 32 for the scaling

layer and increased the filtering numbers of the following 4 layers to 128, 256, 256, and 256. We did not change the readout and the other hyperparameters.

The key features of our winning solution were: for the SENSORIUM track, 1) we added the positions of the objects in the images as additional channels to the inputs; 2) we trained multiple models by using different train-validation splits and averaged the predictions of these models; 3) for the SENSORIUM+ track, we utilize the pretraining datasets in an ensemble way. We trained multiple models using different pretraining datasets and then averaged the predictions. The idea of adding object positions came from the contribution of "pupil behavior" in the SENSORIUM+ track. We hypothesized that the objects would catch the attention and their positions would reflect the pupil's behaviors. The object detection model was trained on the ILSVRC 2017 dataset. We sampled 250 from each category, converted them to gray-scale images, and fine-tuned the PyTorch YOLOv5 large model (Jocher et al., 2022). To label the objects in the competition data, we set the NMS confidence threshold as 0.05 and the IOU threshold as 0.5. The parameter image size was 256. The bounding boxes were merged into a larger one for each image, and the position was a vector (x, y, width, height) in the YOLO format. We gave the vector (0.5, 0.5, 1, 1) for the images without bounding boxes. Finally, there were 6 channels in our SENSORIUM model inputs: the image normalized by the provided mean and standard deviation, the centered first-channel image, and the object positions. For SENSORIUM+ , we removed the object positions. Our ablation studies as well as a description of unsuccessful attempts can be found in the appendix.

SENSORIUM Rank 2: Zhu, Xiao, & Han

Architecture. Our model was based on the CNN Baseline (Lurz et al., 2021) and is initialized in the same way. We used the shared core approach so that we can pre-train our model with data from all seven mice. The feature map in first layer looked like the gradient map of input gray image. Therefore, we use sobel operator to pre-extract the x, y-axis gradient map of the image, and input it into the model together with the image.

Training. We only used the data provided by the competition. To make full use of them, our model was trained in a two-stage manner. First, we update all parameters with all data until the validation score no longer improves. Second, we fine-tuned the core and readout networks of the target mice with smaller learning rate. We utilize self-distillation to generate more robust model. Specifically, we utilize the training set data to train a teacher model. Then, we use teacher model to predict neural responses for all data, and mix new image-response pairs data with real data. The student model with better performance can be obtained by being trained with mixed data. Our optimization object is to minimize the joint loss $Loss = L_{poisson} + \lambda L_{corr}$, where $L_{possion}$ donate Poisson loss, $L_{corr} = 1 - Corr(r_{av}, o_{av})$ is correlate loss, λ is a balance factor. We fix $\lambda = 1000$ in all experiments.

Inference. Model ensembling always works at inference time. However, directly ensembling multiple models will have large redundancy. Therefore, we designed a greedy ensemble strategy to achieve their best outcome. We trained more than 100 models with different seeds and number of core convolution layers. All of them were used to form our ensemble model. We then test each model one by one. If the validation score improves when we block any model, then this model will be removed. We repeat this process 3 times in total.

SENSORIUM Rank 3: Azeglio, Ferrari, Neri, & Marre

Our approach entailed building upon the baseline CNN model developed by the organizers (Lurz et al., 2021). This model comprises of a nonlinear core and a readout layer informed by biological retinotopy. Our focus was on the front-end modules situated between the input and the core, specifically two components. The first component was Scattering Networks, as introduced by Bruna and Mallat (2013), which enforce geometric constraints. These networks have two notable features: 1) their representations are both translation invariant and robust to minor deformations, and 2) deeper models can be achieved without the need for additional parameters as all parameters are fixed. The second component was VOneBlock, developed by Dapello and collaborators (Dapello et al., 2020), which implements biologically grounded constraints through a linear-nonlinear Poisson model composed of a Gabor filter bank with fixed weights, simple and complex cell nonlinearities, and neuronal stochasticity (independent Gaussian noise).

1, 2, 3... Ensembling! Based on the front-ends previously discussed, we decided to implement four distinct models: 1) Scattering front-end, baseline CNN core, and Gaussian readout; 2) VOneBlock front-end, baseline CNN core, and Gaussian readout; 3) Scattering front-end, Squeeze and Excitation CNN core (as described by Hu et al. (2018)) and Gaussian readout; and 4) VOneBlock front-end, Squeeze and Excitation CNN core, and Gaussian readout. Following training and evaluation of the various models, we combined them into an ensemble model by taking an average of their predictions. The performance of individual cores can be found in Table 4 in the appendix, along with a discussion of unsuccessful approaches. Code is available at https://github.com/sazio/sensorium.

SENSORIUM+ Rank 1: A. Roggenbach

In addition to the visual input, neural activity in the visual cortex also depends on the ongoing neural activity and the behavioral state (Stringer et al., 2019; Arieli et al., 1996; Syeda et al., 2022). The key addition of the model is to account for these non-sensory effects based on past neural activity. This is implemented by combining the output of the provided baseline model with a modulator network which consists of three parts.

First, a ten-dimensional network state of the activity of all neurons in the last known timestep is extracted. This low-dimensional projection is calculated by passing the neural activity through a reduced rank auto-regression network for the next time step (to remove the stimulus information which is not predictive for the next time step) and calculating a non-negative matrix factorization on this output. These features are linearly combined for each neuron in the modulator network. Second, the activity history of each neuron is passed through a filter bank with varying temporal kernels, resulting in five history features per neuron which are linearly combined. Third, the output of the provided core model is added to the previously described network state and history output, passed through a ReLU+1 non-linearity and multiplied with a scalar gain. This learned gain regressor is encouraged to be smooth by reading out the regressor with a half-normal distribution kernel and by applying a L2-penalty on the temporal difference.

Additionally, the pupil and running regressors for the core module are normalized between 0 and 1 and hyperparameters are slightly adjusted. Ensembling of five models with different

seeds and train/validation splits further increased the performance. Trained models and code are available at https://github.com/AdrianRoggenbach/adrian_sensorium.

SENSORIUM+ Rank 2: Deng & Guan

The model is identical to the rank 1 model of SENSORIUM.

SENSORIUM+ Rank 3: Fedyanin, Vishniakov, & Panov

We explored several directions to improve the CNN baseline model with varying success. Namely, we tried improving the feature extractor (core), simplifying a readout layer, incorporating geometric and color data augmentations, self-supervised pretraining, and model ensembling. In this section, we report validation set results for recording 27204-5-13.

Core Design Improvement. Initially, we discovered that full-resolution images gave worse results than images that were downscaled by the factor of 0.25. In our investigation, we found that size of the feature map produced by the core has much more pronounced effect on the final result than the input image size. We used this information to design a deeper encoder based on ResNet (He et al., 2016). We made the following changes to baseline ResNet-18 model: changing the number of layers from 18 to 9, changing stride of several layers from 2 to 1, replacing ReLU with ELU, and adding Dropout layers.

Table 2: Model performance with altered core module. Performance reported for a single recording.

	SENSORIUM	SENSORIUM+
Baseline (Lurz et al., 2021)	0.296	0.378
ResNet-18 (He et al., 2016)	0.236	0.310
ResNet-18, stride=1	0.288	0.284
ResNet-9	0.286	0.379
+ELU	0.300	0.400
+Dropout	0.311	0.409

Ensembling. We trained 20 basic models on both SENSORIUM and SENSORIUM+, using the different weight initializations, which gave improvement of **+2.4%** on SENSORIUM and **+3.2%** on SENSORIUM+ test sets.

Reflections

Competition results

It is noteworthy that the majority of the winning teams relied heavily on our CNN baseline model architecture, which remained mostly unchanged. Common successful strategies included:

- using additional transformations of the input data or temporal dependencies
- pre-training the core on the extra datasets, with fine-tuning on the respective competition dataset
- creating large model ensembles, together with improvements in model training
- adjustments of the core-architecture

These changes led to substantial gains in model performance, larger than 15% in both competition tracks. While we consider the improvements in model accuracy with these strategies as impressive (especially given the limited four month competition period), we look

forward to modeling approaches that differ more substantially from our previous state-of-the-art model. On that note, a recent publication by Li et al. (2023) utilized the benchmark data while describing an entirely novel modeling approach based on the Vision Transformer (Dosovitskiy et al., 2020).

Lessons learned for future iterations

We hope that this benchmark infrastructure serves as a catalyst for both computational neuroscientists and machine learning practitioners to advance the field of neuro-predictive modeling. Our broader goal is to continue to populate the underlying benchmark infrastructure with future iterations of dataset releases, new challenges, and additional metrics.

As is the case for benchmarks in general, by converging in this first iteration on a specific dataset, task, and evaluation metric in order to facilitate constructive comparison, `SENSORIUM` 2022 also becomes limited to the scope of those choices. In particular, we opted for simplicity for the first competition hosted on our platform in order to appeal to a broader audience across the computational neuroscience and machine learning communities. *A priori*, it is not clear how well the best performing models of this competition would transfer to a broader or more naturalistic setting where stimuli could be out of domain for the models. Having established our benchmarking framework, possible directions to extend in future challenges are:

- including cortical layers beyond L2/3 and areas in mouse visual cortex beyond V1
- replacing static image stimuli with dynamic movie stimuli in order to better capture the temporal evolution of representation and/or simulation
- replacing grayscale stimuli with coverage of UV- and green-sensitive cone photoreceptors
- increasing the number of animals and recordings in the test set beyond one per track to emphasize generalization across animals and brain states
- moving beyond passive stimulus viewing by incorporating a decision making paradigm
- including different or multiple sensory domains (e.g., auditory, olfactory, somatosensory, etc) and motor areas
- recording neural responses with different techniques (e.g., electrophysiology) that emphasize different population sizes and spatiotemporal resolution
- recording neural responses in different animal models, such as non-human primates.
- inverting model architecture to reconstruct visual input from neural responses.

We believe that predictive models have become an important tool for neuroscience research. In our view, systematically benchmarking and improving these models along with the development of accurate metrics will be of great benefit to neuroscience as a whole. We therefore invite the research community to join the benchmarking effort by continuing to participate in the benchmark, and by contributing new datasets and metrics to our benchmarking system. We would like to cast the challenge of understanding information processing in the brain as a joint endeavor in which we engage together as a whole community, iteratively re-defining what is the state-of-the-art in predicting neural activity and leveraging models to pursue the question of how the brain makes sense of the world.

Acknowledgments

KKL is funded by the German Federal Ministry of Education and Research through the Tübingen AI Center (FKZ: 01IS18039A). This work was supported by an AWS Machine Learning research award to FHS. MB and SAC were supported by the International Max Planck Research School for Intelligent Systems. This project has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (Grant agreement No. 101041669).

This research was supported by National Institutes of Health (NIH) via National Eye Institute (NEI) grant RO1-EY026927, NEI grant T32-EY002520, National Institute of Mental Health (NIMH) and National Institute of Neurological Disorders and Stroke (NINDS) grant U19-MH114830, and NINDS grant U01-NS113294. This research was also supported by National Science Foundation (NSF) NeuroNex grant 1707400. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NEI, NIMH, NINDS, or NSF.

This research was also supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC) contract no. D16PC00003, and with funding from the Defense Advanced Research Projects Agency (DARPA), Contract No. N66001-19-C-4020. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/IBC, DARPA, or the US Government.

Deng & Guan report that this study was also supported by NIH Grants: NIH/NIGMS R35GM133346.

AR acknowledges the use of Fenix Infrastructure resources, which are partially funded from the European Union's Horizon 2020 research and innovation programme through the ICEI project under the grant agreement No. 800858.

References

E H Adelson and J R Bergen. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am., 2(2):284–299, February 1985.

J Antolík, S B Hofer, J A Bednar, and T D Mrsic-flogel. Model constrained by visual hierarchy improves prediction of neural responses to natural scenes. PLoS Comput. Biol., pages 1–22, 2016.

Amos Arieli, Alexander Sterkin, Amiram Grinvald, and Ad Aertsen. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science, 273 (5283):1868–1871, sep 1996. doi: 10.1126/science.273.5283.1868. URL https://doi.org/10.1126%2Fscience.273.5283.1868.

Mohammad Bashiri, Edgar Walker, Konstantin-Klemens Lurz, Akshay Jagadish, Taliah Muhammad, Zhiwei Ding, Zhuokun Ding, Andreas Tolias, and Fabian Sinz. A flow-based latent state generative model of neural population responses to natural images. Adv. Neural Inf. Process. Syst., 34, December 2021.

Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image synthesis. Science (New York, N.Y.), 364(6439), 2019. ISSN 1095-9203. doi: 10.1126/science.aav9436.

Eleanor Batty, Josh Merel, Nora Brackbill, Alexander Heitman, Alexander Sher, Alan Litke, EJ Chichilnisky, and Liam Paninski. Multilayer recurrent network models of primate retinal ganglion cell responses. In International Conference on Learning Representations, 2017.

Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. IEEE transactions on pattern analysis and machine intelligence, 35(8):1872–1886, 2013.

Max F. Burg, Santiago A. Cadena, George H. Denfield, Edgar Y. Walker, Andreas S. Tolias, Matthias Bethge, and Alexander S. Ecker. Learning divisive normalization in primary visual cortex. PLOS Computational Biology, 17(6):e1009028, July 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009028.

Santiago A. Cadena, George H. Denfield, Edgar Y. Walker, Leon A. Gatys, Andreas S. Tolias, Matthias Bethge, and Alexander S. Ecker. Deep convolutional models improve predictions of macaque v1 responses to natural images. PLOS Computational Biology, 15(4):e1006897, April 2019. doi: 10.1371/journal.pcbi.1006897.

Charles F Cadieu, Ha Hong, Daniel L K Yamins, Nicolas Pinto, Diego Ardila, Ethan A Solomon, Najib J Majaj, and James J DiCarlo. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput. Biol., 10(12): e1003963, 2014.

Matteo Carandini, Jonathan B. Demb, Valerio Mante, David J. Tolhurst, Yang Dan, Bruno A. Olshausen, Jack L. Gallant, and Nicole C. Rust. Do we know what the early visual system does? J. Neurosci., 25(46):10577–10597, November 2005. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.3726-05.2005.

Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297, 2020.

R. M. Cichy, K. Dwivedi, B. Lahner, A. Lascelles, P. Iamshchinina, M. Graumann, A. Andonian, N. A. R. Murty, K. Kay, G. Roig, and A. Oliva. The algonauts project 2021 challenge: How the human brain makes sense of a world in motion, 2021. URL https://arxiv.org/abs/2104.13714.

BR Cowley and JW Pillow. High-contrast "gaudy" images improve the training of deep neural network models of visual cortex. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems 33, pages 21591–21603. Curran Associates, Inc., 2020.

Joel Dapello, Tiago Marques, Martin Schrimpf, Franziska Geiger, David Cox, and James J DiCarlo. Simulating a primary visual cortex at the front of cnns improves robustness to image perturbations. Advances in Neural Information Processing Systems, 33:13073–13087, 2020.

Saskia E. J. de Vries, Jerome A. Lecoq, Michael A. Buice, Peter A. Groblewski, Gabriel K. Ocker, Michael Oliver, David Feng, Nicholas Cain, Peter Ledochowitsch, Daniel Millman, Kate Roll, Marina Garrett, Tom Keenan, Leonard Kuan, Stefan Mihalas, Shawn Olsen, Carol Thompson, Wayne Wakeman, Jack Waters, Derric Williams, Chris Barber, Nathan Berbesque, Brandon Blanchard, Nicholas Bowles, Shiella D. Caldejon, Linzy Casal, Andrew Cho, Sissy Cross, Chinh Dang, Tim Dolbeare, Melise Edwards, John Galbraith, Nathalie Gaudreault, Terri L. Gilbert, Fiona Griffin, Perry Hargrave, Robert Howard, Lawrence Huang, Sean Jewell, Nika Keller, Ulf Knoblich, Josh D. Larkin, Rachael Larsen, Chris Lau, Eric Lee, Felix Lee, Arielle Leon, Lu Li, Fuhui Long, Jennifer Luviano, Kyla Mace, Thuyanh Nguyen, Jed Perkins, Miranda Robertson, Sam Seid, Eric Shea-Brown, Jianghong Shi, Nathan Sjoquist, Cliff Slaughterbeck, David Sullivan, Ryan Valenza, Casey White, Ali Williford, Daniela M. Witten, Jun Zhuang, Hongkui Zeng, Colin Farrell, Lydia Ng, Amy Bernard, John W. Phillips, R. Clay Reid, and Christof Koch. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nature Neuroscience, 23(1):138–151, December 2019. doi: 10.1038/s41593-019-0550-9. URL https://doi.org/10.1038/s41593-019-0550-9.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2020. URL https://arxiv.org/abs/2010.11929.

Alexander S. Ecker, Fabian H. Sinz, Emmanouil Froudarakis, Paul G. Fahey, Santiago A. Cadena, Edgar Y. Walker, Erick Cobos, Jacob Reimer, Andreas S. Tolias, and Matthias Bethge. A rotation-equivariant convolutional neural network model of primary visual cortex, 2018.

Katrin Franke, Konstantin F. Willeke, Kayla Ponder, Mario Galdamez, Na Zhou, Taliah Muhammad, Saumil Patel, Emmanouil Froudarakis, Jacob Reimer, Fabian H. Sinz, and Andreas S. Tolias. State-dependent pupil dilation rapidly shifts visual feature selectivity. Nature, 610(7930):128–134, September 2022. doi: 10.1038/s41586-022-05270-3. URL https://doi.org/10.1038/s41586-022-05270-3.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

D J Heeger. Half-squaring in responses of cat striate cells. Vis. Neurosci., 9(5):427–443, 1992a.

D J Heeger. Normalization of cell responses in cat striate cortex. Vis. Neurosci., 9(2):181–197, 1992b.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018.

Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, Kalen Michael, TaoXie, Jiacong Fang, Imyhxy, , Lorna, Zeng Yifu, Colin Wong, Abhiram

V, Diego Montes, Zhiqiang Wang, Cristi Fati, Jebastin Nadar, Laughing, UnglvKitDe, Victor Sonck, Tkianai, YxNONG, Piotr Skalski, Adam Hogan, Dhruv Nair, Max Strobel, and Mrinal Jain. ultralytics/yolov5: v7.0 - yolov5 sota realtime instance segmentation, 2022. URL https://zenodo.org/record/7347926.

J P Jones and L A Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. J. Neurophysiol., 58(6):1187–1211, December 1987.

William F Kindel, Elijah D Christensen, and Joel Zylberberg. Using deep learning to probe the neural code for images in primary visual cortex. Journal of vision, 19(4):29–29, 2019.

D A Klindt, A S Ecker, T Euler, and M Bethge. Neural system identification for large populations separating "what" and "where". In Advances in Neural Information Processing Systems, pages 4–6, 2017.

B Lau, G B Stanley, and Y Dan. Computational subunits of visual cortical neurons revealed by artificial neural networks. Proceedings of the National Academy of Sciences, 99(13): 8974–8979, 2002.

SR Lehky, TJ Sejnowski, and R Desimone. Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. The Journal of Neuroscience, 12(9):3568–3581, September 1992. doi: 10.1523/jneurosci.12-09-03568.1992. URL https://doi.org/10.1523/jneurosci.12-09-03568.1992.

Bryan M. Li, Isabel M. Cornacchia, Nathalie L. Rochefort, and Arno Onken. V1t: large-scale mouse v1 response prediction using a vision transformer, 2023. URL https://arxiv.org/abs/2302.03023.

Zhe Li, Wieland Brendel, Edgar Walker, Erick Cobos, Taliah Muhammad, Jacob Reimer, Matthias Bethge, Fabian Sinz, Zachary Pitkow, and Andreas Tolias. Learning from brains how to regularize machines. Advances in neural information processing systems, 32, 2019.

Zhe Li, Josue Ortega Caro, Evgenia Rusak, Wieland Brendel, Matthias Bethge, Fabio Anselmi, Ankit B Patel, Andreas S Tolias, and Xaq Pitkow. Robust deep learning object recognition models rely on low frequency information in natural images. bioRxiv, page 2022.01.31.478509, February 2022.

Konstantin-Klemens Lurz, Mohammad Bashiri, Konstantin Willeke, Akshay K Jagadish, Eric Wang, Edgar Y Walker, Santiago A Cadena, Taliah Muhammad, Erick Cobos, Andreas S Tolias, Alexander S Ecker, and Fabian H Sinz. Generalization in data-driven models of primary visual cortex. In Proceedings of the International Conference for Learning Representations (ICLR), page 2020.10.05.326256, October 2021.

Lane T McIntosh, Niru Maheswaranathan, Aran Nayebi, Surya Ganguli, and Stephen A Baccus. Deep learning models of the retinal response to natural scenes. Adv. Neural Inf. Process. Syst., 29(Nips):1369–1377, 2016.

Rafael Navarro, Pablo Artal, and David R Williams. Modulation transfer of the human eye as a function of retinal eccentricity. JOSA A, 10(2):201–212, 1993.

Cristopher M Niell and Michael P Stryker. Modulation of visual responses by behavioral state in mouse visual cortex. Neuron, 65(4):472–479, 2010.

D Pamplona, J Triesch, and CA Rothkopf. Power spectra of the natural input to the visual system. Vision research, 83:66–75, 2013.

Felix Pei, Joel Ye, David Zoltowski, Anqi Wu, Raeed H. Chowdhury, Hansem Sohn, Joseph E. O'Doherty, Krishna V. Shenoy, Matthew T. Kaufman, Mark Churchland, Mehrdad Jazayeri, Lee E. Miller, Jonathan Pillow, Il Memming Park, Eva L. Dyer, and Chethan Pandarinath. Neural latents benchmark '21: Evaluating latent variable models of neural population activity, 2021. URL https://arxiv.org/abs/2109.04463.

Carlos R Ponce, Will Xiao, Peter F Schade, Till S Hartmann, Gabriel Kreiman, and Margaret S Livingstone. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. Cell, 177(4):999–1009.e10, 2019.

Ryan Prenger, Michael C-K Wu, Stephen V David, and Jack L Gallant. Nonlinear V1 responses to natural scenes revealed by neural network analysis. Neural Netw., 17(5-6): 663–679, 2004.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In International conference on machine learning, pages 8748–8763. PMLR, 2021.

Jacob Reimer, Emmanouil Froudarakis, Cathryn R Cadwell, Dimitri Yatsenko, George H Denfield, and Andreas S Tolias. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. Neuron, 84(2):355–362, October 2014.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis., 115(3): 211–252, December 2015.

Nicole C Rust, Odelia Schwartz, J Anthony Movshon, and Eero P Simoncelli. Spatiotemporal elements of macaque v1 receptive fields. Neuron, 46(6):945–956, 2005.

Shahd Safarani, Arne Nix, Konstantin Willeke, Santiago Cadena, Kelli Restivo, George Denfield, Andreas Tolias, and Fabian Sinz. Towards robust vision by multi-task learning on monkey visual cortex. Adv. Neural Inf. Process. Syst., 34:739–751, December 2021.

Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J Majaj, Rishi Rajalingham, Elias B Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, et al. Brain-score: Which artificial neural network for object recognition is most brain-like? BioRxiv, page 407007, 2018.

Odelia Schwartz, Jonathan W Pillow, Nicole C Rust, and Eero P Simoncelli. Spike-triggered neural characterization. J. Vis., 6(4):484–507, July 2006.

F Sinz, A S Ecker, P Fahey, E Walker, E Cobos, E Froudarakis, D Yatsenko, X Pitkow, J Reimer, and A Tolias. Stimulus domain transfer in recurrent models for large scale cortical population prediction on video. In Advances in Neural Information Processing Systems 31, 2018.

Fabian H. Sinz, Xaq Pitkow, Jacob Reimer, Matthias Bethge, and Andreas S. Tolias. Engineering a less artificial intelligence. Neuron, 103(6):967–979, September 2019. doi: 10.1016/j.neuron.2019.08.034. URL https://doi.org/10.1016/j.neuron.2019.08.034.

Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brainwide activity. Science, 364(6437), 2019.

Atika Syeda, Lin Zhong, Renee Tung, Will Long, Marius Pachitariu, and Carsen Stringer. Facemap: a framework for modeling neural activity based on orofacial tracking. bioRxiv, pages 2022–11, 2022.

Jon Touryan, Gidon Felsen, and Yang Dan. Spatial structure of complex cell receptive fields measured with natural images. Neuron, 45(5):781–791, 2005.

B Vintch, J A Movshon, and E P Simoncelli. A convolutional subunit model for neuronal responses in macaque V1. J. Neurosci., 35(44):14829–14841, 2015.

Edgar Y Walker, Fabian H Sinz, Erick Cobos, Taliah Muhammad, Emmanouil Froudarakis, Paul G Fahey, Alexander S Ecker, Jacob Reimer, Xaq Pitkow, and Andreas S Tolias. Inception loops discover what excites neurons most using deep predictive models. Nat. Neurosci., 22(12):2060–2065, December 2019.

Konstantin F. Willeke, Paul G. Fahey, Mohammad Bashiri, Laura Pede, Max F. Burg, Christoph Blessing, Santiago A. Cadena, Zhiwei Ding, Konstantin-Klemens Lurz, Kayla Ponder, Taliah Muhammad, Saumil S. Patel, Alexander S. Ecker, Andreas S. Tolias, and Fabian H. Sinz. The sensorium competition on predicting large-scale mouse primary visual cortex activity, 2022. URL https://arxiv.org/abs/2206.08666.

D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences, 111(23):8619–8624, May 2014. doi: 10.1073/pnas.1403112111. URL https://doi.org/10.1073/pnas.1403112111.

Yimeng Zhang, T-S Tai Sing Lee, Ming Li, Fang Liu, Shiming Tang, Tai Sing, Lee Ming, Li Fang, Liu Shiming, T-S Tai Sing Lee, Ming Li, Fang Liu, and Shiming Tang. Convolutional neural network models of V1 responses to complex patterns. J. Comput. Neurosci., pages 1–22, 2018.

David Zipser and Richard A Andersen. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature, 331(6158): 679–684, 1988.

Appendix A. Additional material from the winning teams

SENSORIUM Rank 1: Deng & Guan

The ablation study results on the SENSORIUM model list in Table 3. They are the evaluations of the 5 datasets for pre-training. All of our modifications contributed to the performance improvement of the baseline model. We also found some unexpected contributions in these methods when comparing the ensemble model and the single model. The object positions significantly improved the performances in the single model, but only had minor effects in the ensemble model.

Model ablation study		
Method	Model type	Score
Best model	ensemble	0.6254±0.0235
Remove object bounding boxes	ensemble	0.6234±0.0239
Fewer filters	ensemble	0.6068±0.0236
Replace Conv-scale	ensemble	0.6007±0.0225
Remove centered image	ensemble	0.5964±0.0227
Best model	single	0.5895±0.0238
Remove object bounding boxes	single	0.5794±0.0241
Fewer filters	single	0.5745±0.0242
Replace Conv-scale	single	0.5706±0.0248
Remove centered image	single	0.5646±0.0225

Table 3: Ablation study for the core used in the SENSORIUM track.

We also tried several other methods to utilize the information of the objects but failed. These methods included constructing a matrix with the same shape as the image and assigning 0 or 1, or different weights between 0 and 1, for the elements outside and inside the bounding box. We tried adding this matrix as an additional channel and multiplying it on the original images to clip the image, but none of these experiments could outperform the baselines. One limitation of our current model is that it is not an end-to-end solution, but needs a separate model or extra manual effort to provide the object information. In future work, we may explore the model's ability to determine the regions of interest for Gaussian readout sampling by itself.

SENSORIUM Rank 2: Zhu, Xiao, & Han

Unsuccessful approaches We try to improve the performance via pre-training CNN core. The first idea come to our mind is training in CLIP (Radford et al., 2021) way. CLIP improves performance of backbone network by minimizing the similarity of two different modalities. We consider images and neural responses in mouse visual cortex as two intrinsically related modalities. Therefore, we feed the features extracted by the core into a fully connected

layer to match dimensions of neural responses, thereby minimizing their similarity. This pre-training method enables our model to converge faster, but unfortunately, it does not bring performance improvements.

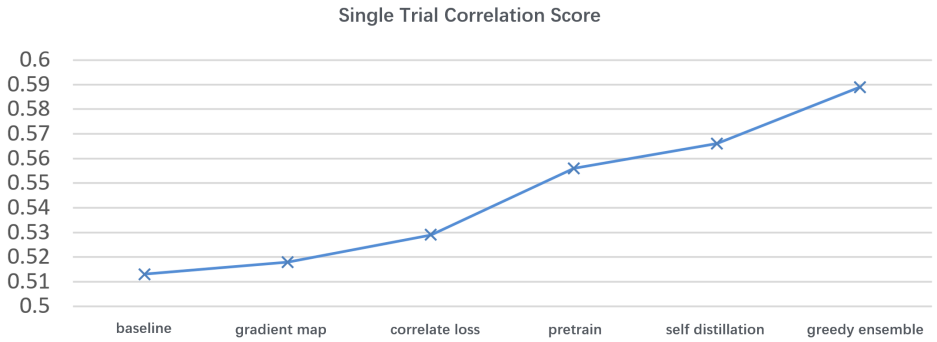

Figure 3: Ablation study for different training strategies in the SENSORIUM track.

SENSORIUM Rank 3: Azeglio, Ferrari, Neri, & Marre

Tried and Failed. Because the thin lens model does not take into account the spherical nature of the eye, it is not applicable to certain transformations of the input image. Thin lens approximations characterize the geometry of planar image projection, while spherical transformations map images onto a spherical surface. We incorporated spherical projection into our processing pipeline by sampling local power spectra across the visual field, and applying it as a preprocessing step to input images. We relied on the human Modulation Transfer Function (Navarro et al., 1993) estimated by Pamplona et al. (2013), which we appropriately scaled for application to mice. Compared with the baseline model, our results showed improvements in the third significant digit. Further attempts to extend our methods to SENSORIUM+ were unfruitful, possibly because this dataset lacks information about pupil position and dilation. Despite the above limitations, we remain interested in exploring the potential of this approach in future work.

Model	Single Trial Correlation	Correlation to Average	FEVE
Baseline CNN	0.29	0.543	0.482
Scattering CNN	0.31	0.56	0.495
Scattering SE[1]-CNN	0.31	0.559	0.492
VOneBlock CNN	0.30	0.557	0.487
VOneBlock SE-CNN	0.30	0.556	0.486
Ensemble	0.324	0.587	0.549

Table 4: Performance on the live test set of individual cores in the model ensemble.

SENSORIUM+ Rank 3: Fedyanin, Vishniakov, & Panov

Simplification of Readout Layer. We investigated how changing the shape of Σ in the Gaussian readout affects the final result. Our initial thought was that having a 2×2 covariance matrix for each of the neurons could be redundant. In our experiments in Table 5, we found that the shape of the Σ has almost no effect on the final result, and instead of having a 2×2 matrix for each neuron, fixing sigma to a single number seems to be sufficient.

Σ shape	$n \times 2 \times 2$	$n \times 2 \times 1$	$n \times 1 \times 1$	2×2	1×2	1×1
# params	31104	15552	7776	4	2	1
Acc	**0.311**	0.307	0.308	0.305	0.306	0.306

Table 5: Performance as a function of number of readout parameters. Different shapes of Σ with ResNet-9 encoder, where n represents the number of neurons.

Data augmentation. We tried augmenting the data, but with no further benefit (see Table 6). We guess one needs to match the readout shift for the geometrical augmentation precisely, but we didn't validate the hypothesis.

Aug	No Aug	Blur	ColorJitter	RRC2	RRC5	HFlip
Acc	**0.311**	0.305	0.296	0.103	0.167	0.258

Table 6: Model performance for various data augmentations. No Aug: ResNet-9 with no augmentations. RRC2: RandomResizedCrop with 20% crop lower bound. RRC5: RandomResizedCrop with 50% crop lower bound. Blur: Gaussian Blur. HFlip: Horizontal Flip. For each augmentation, the probability of application was set to 20%.

Self-supervised pre-training. For self-supervised pre-training we chose MoCo v2 (Chen et al., 2020) as our base framework. We perform pre-training only on the Core part of the network without Readout and Shifter. To do this, we add a pooling and linear layer, which produces the embedding of size 128. We tried different augmentations, but the pre-trained model didn't improve the final results.