PAPER • OPEN ACCESS

Convolutional neural network classifies visual stimuli from cortical response recorded with wide-field imaging in mice

To cite this article: Daniela De Luca *et al* 2023 *J. Neural Eng.* **20** 026031

View the article online for updates and enhancements.

Journal of Neural Engineering

Convolutional neural network classifies visual stimuli from cortical response recorded with wide-field imaging in mice

Daniela De Luca[1] , Sara Moccia[1,*] , Leonardo Lupori[3] , Raffaele Mazziotti[2,4] , Tommaso Pizzorusso[2,3] and Silvestro Micera[1,5]

[1] The BioRobotics Insistute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pisa, Italy
[2] Institute of Neuroscience (IN-CNR), Pisa, Italy
[3] BIO@SNS Lab, Scuola Normale Superiore, Pisa, Italy
[4] Department of Neuroscience, Psychology, Drug Research and Child Health, University of Florence, Florence, Italy
[5] Bertarelli Foundation Chair in Translational Neuroengineering, EPFL, Geneva, Switzerland
* Author to whom any correspondence should be addressed.

E-mail: sara.moccia@santannapisa.it

Keywords: visual stimuli decoding, optic nerve, visual prostheses, deep learning, wide-field imaging

Abstract

Objective. The optic nerve is a good location for a visual neuroprosthesis. It can be targeted when a subject cannot receive a retinal prosthesis and it is less invasive than a cortical implant. The effectiveness of an electrical neuroprosthesis depends on the combination of the stimulation parameters which must be optimized, and an optimization strategy might be performing closed-loop stimulation using the evoked cortical response as feedback. However, it is necessary to identify target cortical activation patterns and to associate the cortical activity with the visual stimuli present in the visual field of the subjects. Visual stimuli decoding should be performed on large areas of the visual cortex, and with a method as translational as possible to shift the study to human subjects in the future. The aim of this work is to develop an algorithm that meets these requirements and can be leveraged to automatically associate a cortical activation pattern with the visual stimulus that generated it. *Approach.* Three mice were presented with ten different visual stimuli, and their primary visual cortex response was recorded using wide-field calcium imaging. Our decoding algorithm relies on a convolutional neural network (CNN), trained to classify the visual stimuli from the correspondent wide-field images. Several experiments were performed to identify the best training strategy and investigate the possibility of generalization. *Main results.* The best classification accuracy was 75.38% ± 4.77%, obtained pre-training the CNN on the MNIST digits dataset and fine-tuning it on our dataset. Generalization was possible pre-training the CNN to classify Mouse 1 dataset and fine-tuning it on Mouse 2 and Mouse 3, with accuracies of 64.14% ± 10.81% and 51.53% ± 6.48% respectively. *Significance.* The combination of wide-field calcium imaging and CNNs can be used to classify the cortical responses to simple visual stimuli and might be a viable alternative to existing decoding methodologies. It also allows us to consider the cortical activation as reliable feedback in future optic nerve stimulation experiments.

1. Introduction

In the years, many neural prostheses aimed at vision restoration have been investigated [1]. Visual prostheses may target different levels of the visual pathways depending on the cause of vision loss and residual functions of the visual system. Recently, researchers are investigating electrical stimulation of the optic nerve to restore vision in blind subjects [2–4]. This nerve holds great potential specifically as a visual prosthesis target for subjects affected by outer retinal lesions or degeneration diseases, and for all those that cannot benefit from a retinal device [2–4], since optic nerve implants bypass the retina and are less invasive than cortical implants [4]. The optic nerve is a compact conduit made up of the myelinated axons of the retinal ganglion cells, so it processes information relative to the whole visual field [5].

Its electrical stimulation may thus allow subjects to experience visual perception throughout their visual field using a limited number of electrodes [6].

The stimulation protocol of neural prostheses needs to be optimized to be effective, and current research is focused on finding the best strategy to do it. The stimulation protocol is the combination of all the parameters involved in the delivery of the electric stimuli, namely the active sites of the electrode, current intensity, frequency, and pulse width [6]. Optimizing such a stimulation protocol means choosing the optimal combination of current parameters and active electrode sites that elicits the desired effect. The effect of the electrical stimulation may not be known *a priori* and searching through a huge stimulation-parameter space is often required [7]. Optimization strategies are necessary to reduce the number of animal experiments, which would be large by performing manual tuning or grid searches throughout all the possible combinations of stimulation parameters. In the case of an optic nerve visual prosthesis, an optimization strategy could be to perform closed-loop stimulation relying on the feedback provided by the visual cortex. This means that the stimulation parameters could be continually adapted, for example, by exploiting genetic algorithms that maximize the similarity between the response elicited in the visual cortex by the electrical stimulation of the optic nerve and the target cortical activation [6]. Recent studies performed on rabbits proved that intraneural optic nerve stimulation can selectively activate the visual cortex [4, 8], but further animal experiments are now necessary to 1) assess the possibility of univocally associating visual inputs and cortical activation patterns and 2) understand which are the activation patterns that have to be resembled to make sure the animal is experiencing a certain visual perception.
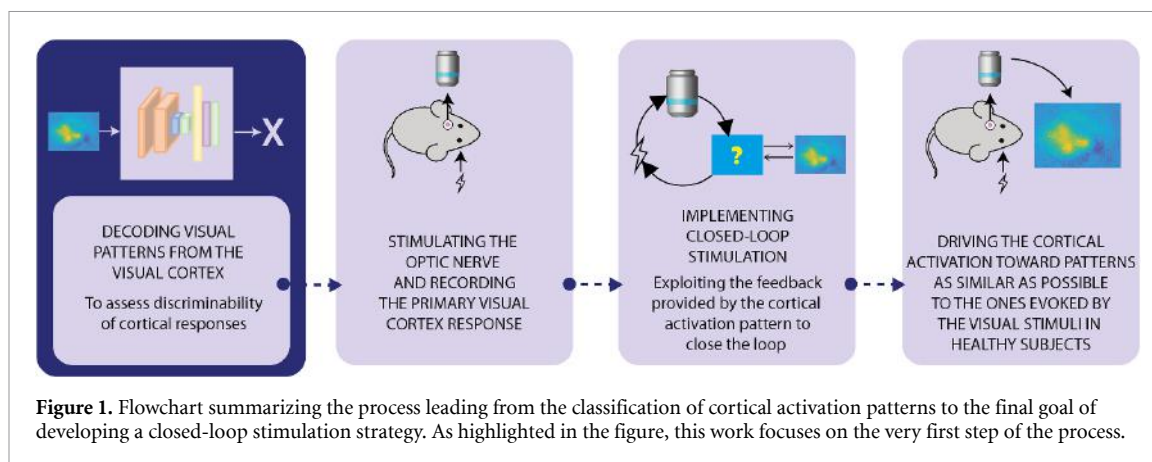
In this work, we focus on the first point and investigate whether the cortical response can be regarded as an indirect measure of visual percepts, developing an algorithm that automatically associates cortical activation with the class of stimuli that generated it. Successful classification of evoked cortical responses might pave the way for the long-term goal of the current study. This aims, in the future, to drive the cortical activation towards patterns that are as similar as possible to the ones evoked by the visual stimulus that must be replicated [6], as schematized in figure 1.

When choosing a technique to record cortical activation, it is useful to consider that the primary visual cortex has a retinotopic organization [9–11], for which each point of the visual field is mapped to a specific point of the visual cortex, and close points in the visual field are close in the visual cortex [10]. It is desirable to record from its whole surface, and wide-field calcium imaging is suitable for this purpose [12], as opposed to other recording techniques that focus on limited areas of the visual cortex as discussed in section 2. This technique takes advantage of genetically encoded fluorescent indicators, which are engineered proteins that alter their fluorescence intensity when several neuronal events occur [12, 13]. Some of these indicators, called genetically encoded calcium indicators (GECIs), are sensitive to fluctuations in intracellular calcium, an indirect indicator of neural activity [14]. Improvements in technology allowed *in vivo* imaging of neural activity and, among GECIs, the GCamP family is the most used in neuroscience applications [15]. Imaging is performed after the head fixation of the animal. The images can be captured with a closed skull by thinning the bone or using glass coverslips achieving partial transparency, or an open skull by replacing the bone flap with transparent materials [15]. The images are obtained by exploiting a light source, usually provided by LEDs, combined with a light-sensing device, usually attached to a microscope, to capture emission fluorescence [14, 15]. Often performed in mice, wide-field calcium imaging has been widely used in the literature to evaluate neural dynamics across broad brain areas, especially to explore the interaction between them, thanks to its high signal-to-noise ratio (SNR) and spatiotemporal resolution. The main application of this technique has been the understanding of large-scale cortex dynamics in behavioral and cognitive processes, from sensorimotor integration to decision-making tasks [12]. The variety of applications for which wide-field calcium imaging is useful proves its great versatility. In [16–18] it was used to investigate motor-evoked cortical activity during motor training for after-stroke functional recovery. In [19] it is used to explore the interaction between all the cortical regions engaged in locomotion. In [10, 11] wide-field calcium imaging is used to explore the organization of the mouse visual cortex, hence the idea of considering this technique to capture the activity of almost all this area. Furthermore, the SNR of wide-field calcium imaging is high enough to allow trial-by-trial analyses without the need of averaging across several trials, allowing us to collect a dataset on which to apply machine learning techniques.

Using wide-field calcium imaging allows us to exploit data analysis and processing pipelines similar to those needed to analyze and process functional magnetic resonance imaging (fMRI) data [20], facilitating the comparability of the studies performed on animals and on humans, and could allow us to smoothly move from animal models to human subjects in the future. To our knowledge, wide-field calcium imaging has never been used to exploit retinotopy and perform visual stimuli decoding.

We here investigate whether patterns of cortical activation recorded with this technique can be automatically associated with the visual stimuli that generated them. The first goal of this study is to assess the possibility to discriminate the visual stimuli delivered

Figure 1. Flowchart summarizing the process leading from the classification of cortical activation patterns to the final goal of developing a closed-loop stimulation strategy. As highlighted in the figure, this work focuses on the very first step of the process.

to the animal, based on the cortical recordings. This is important, since the impossibility to do that would imply the impossibility to rely on the cortical activation as feedback. The second goal is the automatic association of the cortical activation pattern to the stimulus that generated it. We propose a learning-based decoding algorithm applied to wide-field calcium images to classify cortical responses to visual stimuli delivered to 3 anesthetized GCaMP6f transgenic mice. Our algorithm relies on deep learning, and in particular on convolutional neural networks (CNNs), which have become one of the leading choices for medical image analysis [21], providing us with a reliable tool to perform image classification. Based on this premise, the contributions of this work can be summarized as follows:

- Collecting the first dataset of wide-field calcium images showing the cortical activity from the primary visual cortex (V1) of 3 GCaMP6f transgenic mice in response to standardized visual stimuli.
- Designing a CNN able to automatically classify wide-field images, associating a visual stimulus to its evoked cortical activity encoded in the image.
- Exploiting transfer learning to generalize the classification performance from one mouse to the others.

To our knowledge, such an approach has never been proposed in the literature to perform decoding of the visual cortex.

2. Related work on visual cortex decoding

Neural decoding of visual information is a topic of great interest for researchers, who make a great effort to understand how neurons represent external stimuli [22]. Despite their small dimensions and low-resolution vision, it is common to perform visual cortex decoding in mice for several reasons. The mouse visual cortex is similar to the human in several aspects such as retinotopy, receptive field types, orientation

tuning and ocular dominance plasticity [9, 23]. Other reasons include the possibility of recording large-scale data, the availability of transgenic animals and their relative ease of maintenance [24]. In visual decoding research in animals, spike signals are the most used neural data [25]. They can be directly obtained by implanting intracortical microelectrodes, or performing two-photon calcium imaging and applying proper transcoding algorithms [26]. Being widely used, spike recording benefits from the existence of cutting-edge technologies [25]. Consequently, spike-based visual neural decoding is predominant and many studies exist that take advantage of these recording techniques [27, 28]. For the same reason, plenty of algorithms exist that can achieve great decoding performance, from linear to bayesian-based and to deep neural network methods [25]. Intracortical microelectrodes recordings achieve typically very high decoding performance, but they do not allow to capture the activity of the whole surface of the visual cortex. These electrodes are most convenient when the main goal is to reach deeper layers of the cortex, but they only provide recordings from limited portions of the surface. Due to the retinotopic organization of the visual cortex, the cortical activity relative to only a portion of the visual field is captured [9–11].

Calcium imaging is a relatively recent and not yet much explored technique, employed to record neuronal activity from large brain areas [25]. The fluorescence signal is often recorded via two-photon Ca^{2+} imaging and converted into spiking signal to apply the many techniques available to perform spike-based decoding [29–31]. Using two-photon Ca^{2+} imaging it has been demonstrated that natural images evoke responses of only small populations of neurons in V1. These neurons, however, are sparsely active in the visual cortex and not clustered in a small area.

In [9], for example, sparsely distributed high-responding neurons are selected and used to decode natural and artificial movie scenes. Another study aimed at reconstructing visual scenes from two-photon recordings of a neural population in the mouse primary visual cortex by exploiting the
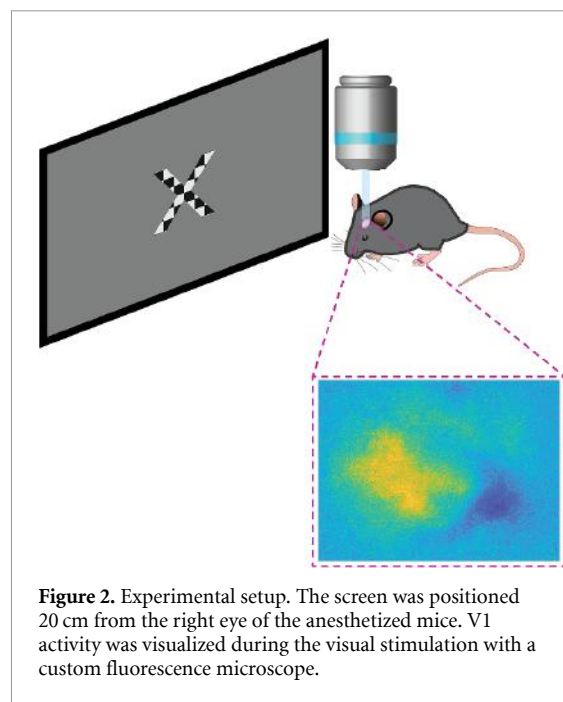
total spike count and a linear decoder [31]. Supervised machine learning was applied to perform visual decoding of neuronal calcium responses to 118 naturalistic scenes in [32]. They tested several machine learning architectures and brain areas exploiting a dataset from the Allen Brain Institute consisting of 25 000 neurons. The highest decoding accuracy, independently from neuron type and cortical depth, was achieved in the primary visual cortex using a 1D CNN. The Allen Brain Observatory dataset [33] from the Allen Brain Institute, consisting of neuronal calcium images of mouse V1 responses to artificial visual stimuli, was used to decode orientations, spatial frequencies, moving direction and speed using a linear support vector machine (SVM) [34]. The Neuropixel dataset from the Allen Brain Institute [35] contains instead spike responses of hundreds of neurons from the mouse visual cortex to natural and artificial images, that were classified using a deep neural network in [28]. In [36], it is demonstrated that natural images can be reliably reconstructed from a small number of sparse high-responding neurons. High-precision decoding was performed in [27], where spiking activity from up to 50 000 neurons of mice V1 and higher visual areas was used to decode stimulus orientation with thresholds of 0.35 degrees and 0.37 degrees respectively.

Visual stimuli decoding is also of great interest in humans, for whom techniques such as fMRI, functional near-infrared spectroscopy (fNIRS), or electroencephalogram (EEG) are used. In fact, despite the high performance of spike activity classification, this method lacks translationality [20]. Analyses of fMRI data have been performed to classify object categories [37], hand gestures [38], visual features [39], images [40] and colors, and to reconstruct seen and imagined objects [41]. EEG was also exploited for visual stimuli classification, using CNNs [42–44].

A long-term goal of the present study is to design experiments to perform visual stimuli decoding, moving from animal experiments to human subjects. This transition is not straightforward. On the one hand, recording single-neuron activity is highly impractical in humans, due to the difficulty of the surgeries and ethical reasons [20]. On the other hand, it is difficult to implement fMRI mice experiments [20]. Wide-field calcium imaging is a translational method that allows this transition to be made more linearly. It is possible to perform analyses on wide-field calcium imaging data that are comparable with fMRI data, so that it is theoretically possible to design similar experiments when moving the study to human subjects [20].

3. Materials and methods

In section 2, details about the animal experiments, the presentation of visual stimuli and the collection of the

Figure 2. Experimental setup. The screen was positioned 20 cm from the right eye of the anesthetized mice. V1 activity was visualized during the visual stimulation with a custom fluorescence microscope.

datasets are provided. This is followed by the description of the CNN and the experiments to identify the best training strategy, in section 3.2. Finally, the metrics used to assess the performance of the CNN are described in section 3.2.5

3.1. Image acquisition and preprocessing

Animal experiments were performed in accordance with the European Directives (2010/63/EU) and were approved by the Italian Ministry of Health (authorization number 621/2020-PR). Three different experimental sessions were carried out on three mice, which were anesthetized (Isoflurane 3%) and positioned, with their head fixed, in front of a screen placed 20 cm from their right eye, as schematized in figure 2. Being anesthetized, the mice could not move their eyes. To avoid dryness, a proper gel was applied. The visual stimuli were then delivered passively to the animals. The stimuli consisted of ten different geometrical shapes, which are shown in figure 3. These were displayed in the center of the screen, and their outlines were filled with a checkerboard that flickered with a temporal frequency of 5 Hz and a spatial frequency of 0.08 cycles per degree, on a gray background. This configuration was chosen to enhance the response of V1 neurons [45–47]. During the first experimental session, one stimulus every 2 s was delivered. After 500 ms of a pre-stimulus gray background, the stimulus remained visible for 500 ms, enough to elicit a V1 response detectable with wide-field calcium imaging [12]. Each stimulus was followed by 1 s of a post-stimulus gray background, to let the fluorescence signal return to baseline. During the two remaining experimental sessions, the sequence consisted of 1 s of a pre-stimulus gray background,
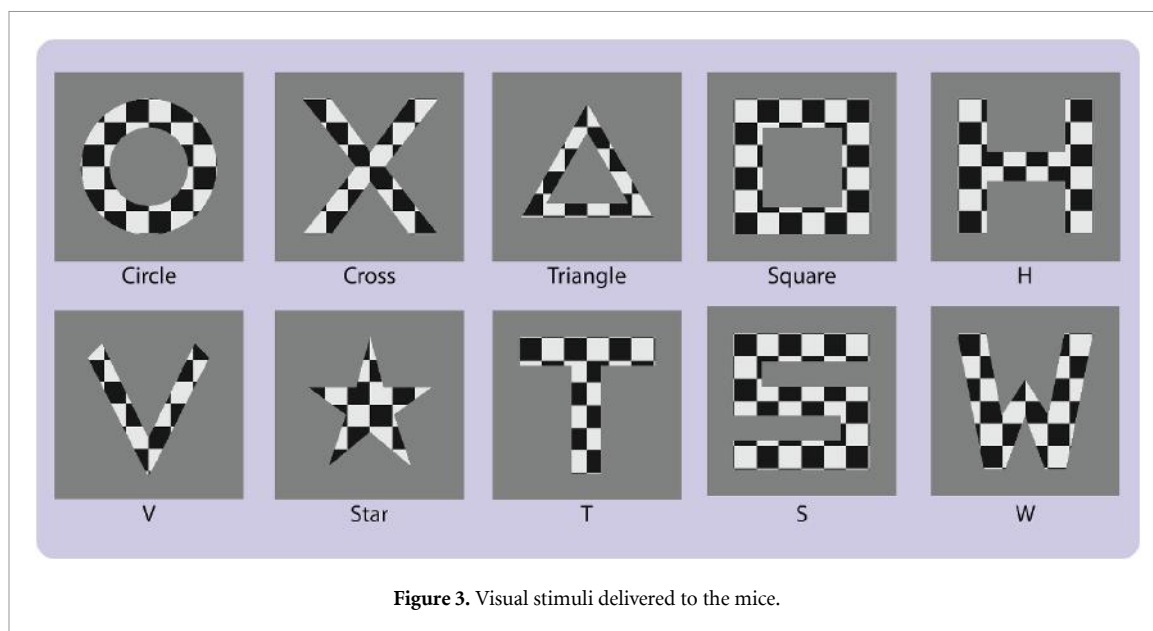
Figure 3. Visual stimuli delivered to the mice.

1 s of visual stimulus, and 5 s of a post-stimulus gray background.

V1 activity was visualized using a custom Leica fluorescence microscope (Leica Microsystems), equipped with a Leica Z6 APO coupled with a Leica PlanApo 2.0 × (10 447 178) objective. Fluorescence was detected through an I3 cube (excitation BP 450–490 nm dichroic 510 nm emission LP 515 nm). Images were captured with a 12-bit depth acquisition camera (PCO edge 5.5) at 10 frames per second with a resolution of 270 × 320 pixels.

Every trial resulted in a sequence of frames showing the activity of V1 during the presentation of the stimulus. For each frame, a no-stimulus response was calculated as the mean response across the frames corresponding to the pre-stimulus gray background presentation. This was regarded as the baseline and subtracted from each frame. The variation of fluorescence was then normalized with respect to the baseline. The visual cortex response was averaged across the three central frames, and the obtained images made up the dataset used to train the CNN. Examples for each mouse are displayed in figures 4(a)–(c), respectively, where a colormap is used for visual purposes. Mouse 2 and Mouse 3 datasets were of lower quality compared to Mouse 1 dataset. Some of the images displayed the evident presence of blood vessels and opaque areas caused by bone regrowth [48]. Data curation was also performed on the datasets, which were visually inspected to discard images displaying zero or only partially visible cortical activity. Their final compositions are summarized in table 1.

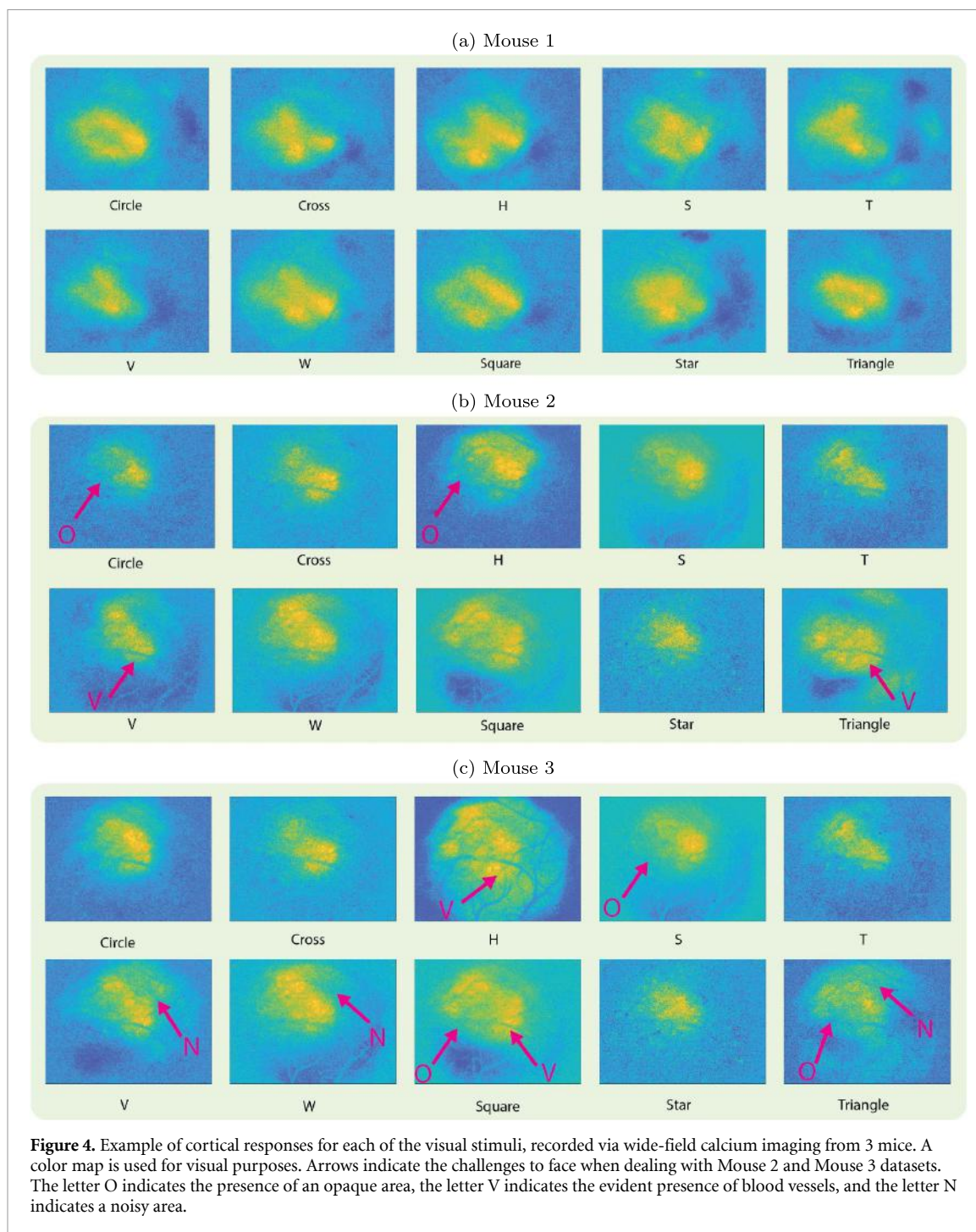### 3.2. Visual stimuli classification

3.2.1. CNN
Our CNN was composed of nine layers, as summarized in table 2. The first two layers were convolutional,

with 32 and 64 filters, respectively, with kernel size 3x3, followed by a max pooling layer. A dropout layer randomly omitting hidden units from the network with 0.25 probability was then included to prevent overfitting [49]. It was followed by a flattening layer, a fully-connected layer with 128 neurons, and another dropout layer, with a 0.5 probability of omitting units. Finally, the output layer had 10 units as the number of classes and was activated by the softmax activation function. All the other layers were activated with the hyperbolic tangent function.

3.2.2. Training CNN
The CNN was trained maximizing the accuracy of the validation set and minimizing the categorical cross-entropy. For each of the 3 datasets, 80% was selected through random stratified sampling to be used as training set, and the remaining 20% as test set. Then, the training sets were further split into training and validation sets with an 80:20 ratio. The datasets underwent data augmentation on the training samples, both offline and on-the-fly, using common transformations [50]. Data augmentation is a technique used to improve the generalization power of machine learning algorithms. The underlying assumption is that it is possible to extract more generalizable features from a larger dataset [51]. Offline data augmentation was performed by applying 4 different geometric transformations: random rotation in a range of ±20 degrees, random spatial shear in the range 0%–20%, random corrections of brightness in the range 0%–40%, random corrections of contrast in the range 20%–50%. The augmentation factor was set to 8 so that the Mouse 1 training set went from 166 to 5312 images, Mouse 2 training set from 93 to 2976 images, and Mouse 3 training set from 189 to 6048 images. Data augmentation was also performed on-the-fly during training, to attenuate

Figure 4. Example of cortical responses for each of the visual stimuli, recorded via wide-field calcium imaging from 3 mice. A color map is used for visual purposes. Arrows indicate the challenges to face when dealing with Mouse 2 and Mouse 3 datasets. The letter O indicates the presence of an opaque area, the letter V indicates the evident presence of blood vessels, and the letter N indicates a noisy area.

the chance of overfitting. The transformations were the following: random rotation in the range of ± 5 degrees; random vertical and horizontal shift in the range of $\pm 2\%$, random zoom of a factor in the range of $\pm 30\%$, random spatial shear in the range of $\pm 2\%$. Figure 5 shows examples of the performed geometric transformations on a picture extracted from the dataset. Stratified 5-fold cross-validation was used for classification performance assessment. All the training experiments were performed on a Dell XPS 8940, Intel Core i7-10 700 processor, 2.90 GHz CPU, NVIDIA GeForce GTX 1660 Ti GPU, 16 GB RAM.

Several experiments were performed to assess the best training strategy to classify the cortical responses and to investigate the possibility of generalizing the result to different animals. The first issue was addressed using only the dataset from Mouse 1 which, as shown in figure 4, was the one with the highest quality images in terms of SNR, contrast and absence of opaque areas and visible blood vessels. This choice was made to avoid attributing classification errors to poor model training and architecture, rather than to a noisy dataset.

Mouse 1

The experiments performed on the first mouse were aimed at identifying the best training strategy. The CNN was trained testing the impact on the

Table 1. Size and composition of the datasets from the 3 mice after data curation.

(a) Mouse 1

Shape	Samples
Circle	25
Cross	29
'H'	24
'S'	22
'T'	26
'V'	28
'W'	28
Square	28
Star	19
Triangle	31
Total	260

(b) Mouse 2

Shape	Samples
Circle	17
Cross	11
'H'	15
'S'	14
'T'	18
'V'	14
'W'	15
Square	16
Star	13
Triangle	12
Total	145

(c) Mouse 3

Shape	Samples
Circle	26
Cross	29
'H'	34
'S'	35
'T'	29
'V'	37
'W'	31
Square	25
Star	24
Triangle	25
Total	295

Table 2. Convolutional neural network.

Name	Kernel size	Output shape [pixels]
Input	—	$135 \times 160 \times 1$
Convolutional layer	3×3	$133 \times 158 \times 32$
Convolutional layer	3×3	$131 \times 156 \times 64$
Max pooling layer	2×2	$65 \times 78 \times 64$
Dropout	—	$65 \times 78 \times 64$
Flattening layer	—	324480
Dense layer	—	128
Dropout	—	128
Output	—	10

classification accuracy of different combinations of offline and on-the-fly data augmentation and transfer learning.

- **Experiment 1 (BASE):** We trained the CNN using the original dataset recorded from Mouse 1, without data augmentation. The optimizer was Adadelta, and the model was trained for 100 epochs with a batch size equal to 16. The accuracy on the validation set was maximized using the categorical cross-entropy as loss function.
- **Experiment 2 (OFF-AUG):** The same training strategy was adopted, but with the offline augmented version of the dataset.
- **Experiment 3 (OOTF-AUG):** The same training strategy was adopted, but the dataset was augmented both offline and on-the-fly. The latter is a technique that allows the enrichment of the variety of the dataset without any storage use [51].
- **Experiment 4 (PT-BASE):** Transfer learning allows us to improve learning in one domain, exploiting information from a related domain [52], and is used to address limited availability of training data. The effects of this technique were explored by pre-training the CNN to classify the MNIST handwritten digits dataset [53] and fine-tuned with our wide-field calcium images. The MNIST dataset consists of 60 000 training samples and 10 000 test samples, both of which are grayscale images with a size of 28×28 pixels. This dataset was used for two main reasons: it comprises ten classes as our dataset, and the images are compatible with ours in terms of size, semantic information and in being grayscale. These were resized to 135×160 pixels to match the size of our images, normalized between 0 and 1, and provided to the CNN. The network was pre-trained for 25 epochs on batches of 32 samples, with categorical cross-entropy as loss function and Adadelta as optimizer. MNIST classification accuracy was optimized and reached more than 95% on the test set. The weights of the pre-trained CNN were loaded in the model. Initially, the weights of one out of two convolutional layers were frozen and the rest of the model was re-trained with the original Mouse 1 dataset, i.e. without data augmentation. During this first step Adadelta was used as optimizer, the classification accuracy as metrics, and it was trained on 100 epochs using mini-batches of 16 samples. The second step consisted of unfreezing weights of the first layer and re-training the whole CNN, using Adam optimizer with a lower learning rate of 10^{-4} for fine-tuning.
- **Experiment 5 (PT-OFF-AUG):** The same training strategy was adopted, but with the offline augmented version of the dataset.
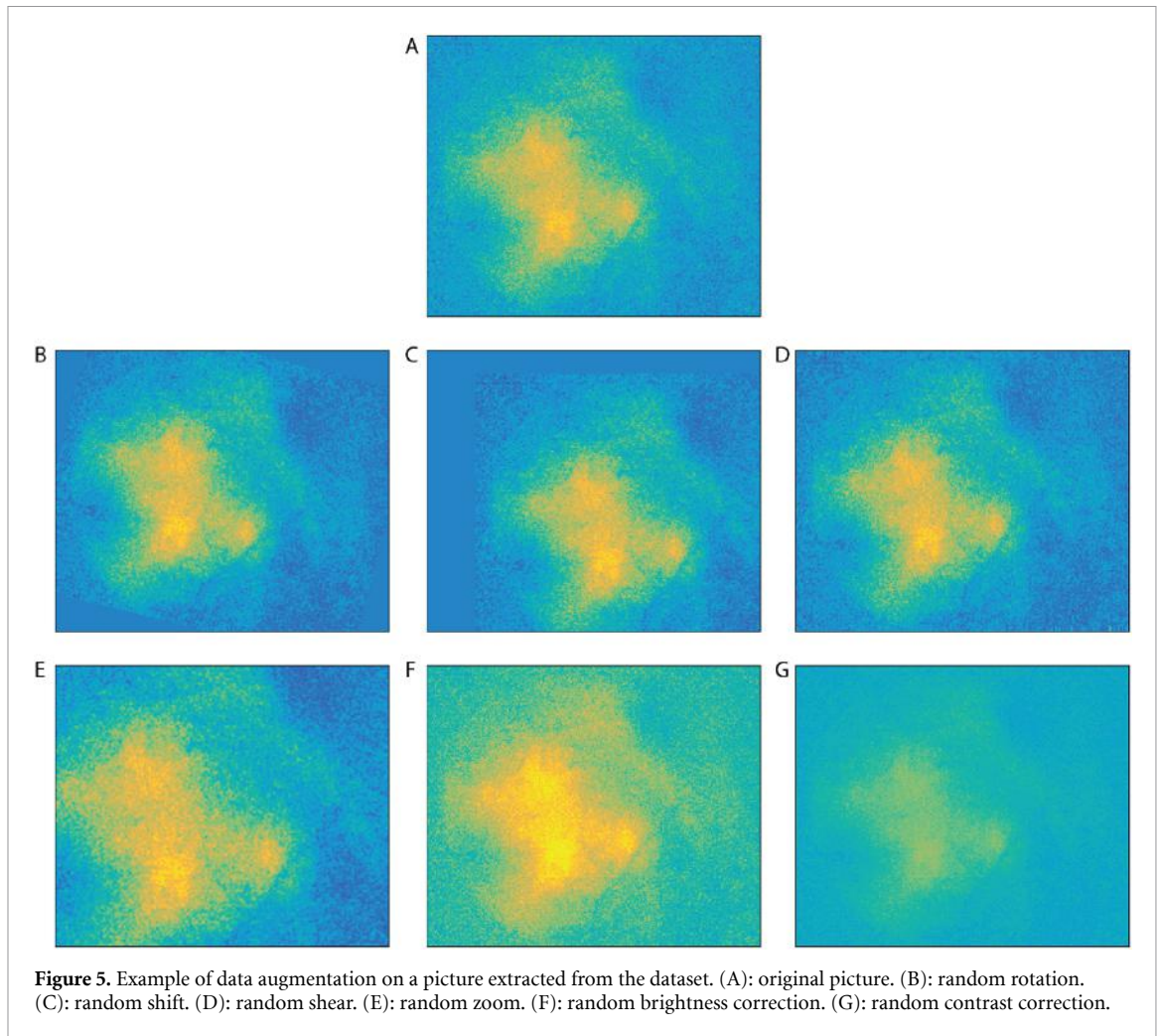
Figure 5. Example of data augmentation on a picture extracted from the dataset. (A): original picture. (B): random rotation. (C): random shift. (D): random shear. (E): random zoom. (F): random brightness correction. (G): random contrast correction.

- **Experiment 6 (PT-OOTF-AUG):** The same training strategy was adopted, but the dataset was augmented both offline and on-the-fly.
- **Experiment 7 (PT-OOTF-AUG +1c):** This experiment was performed to determine whether the network complexity was the best suited for this task. A convolutional layer was added to the CNN to increase its complexity, and the CNN was pre-trained with the MNIST dataset and fine-tuned on Mouse 1 dataset, augmented both offline and on-the-fly.
- **Experiment 8 (PT-OOTF-AUG -1c):** This experiment was performed to explore a lower network complexity. A convolutional layer was removed from the CNN and the same training strategy of the previous experiment was performed.

Mouse 2 and Mouse 3
We performed experiments to investigate the ability of the CNN to generalize and classify the cortical activity of the other animals.

- **Experiments 9–10 (M2-TEST and M3-TEST):** The CNN trained on Mouse 1 dataset was tested on Mouse 2 and Mouse 3 datasets.

- **Experiments 11–12 (M2-PT-OOTF-AUG and M3-PT-OOTF-AUG):** Two CNNs were trained and tested separately on Mouse 2 and Mouse 3 datasets, following the procedure of previously described Experiment 8.
- **Experiments 13–14 (M2-TL and M3-TL):** Transfer learning was applied starting from the CNN previously trained to classify the Mouse 1 dataset. The weights of this network were loaded as the initial weights, and fine-tuning was performed in the same way as in experiment 8, separately for Mouse 2 and Mouse 3.

3.2.5. Performance assessment
The performance of the model after the different training strategies was evaluated using classification accuracy as metrics:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of images}}.$$

The receiver operating characteristic (ROC) curves were computed for each stratified fold and macro-averaged. Additionally, t-distributed stochastic neighbor embedding (t-SNE) plots were generated. The activation of the last hidden layer of

the CNN was visualized in two dimensions to examine the internal features learned by the model [54]. Statistical significance between the performance resulting from the different training strategies was assessed using two statistical tests, both with significance level (α) set to 0.05. One-way ANOVA test was used when the data met the assumption of a normal distribution, verified performing the Shapiro-Wilk normality test (α set to 0.05). The Kruskal–Wallis test was used otherwise. Human-based classification was also performed. A total of 15 subjects (age 27 ± 2 years) were asked to classify a subset of 20 images from the Mouse 1 dataset. Of these, 15 were chosen that were misclassified by the network, and 5 that were correctly classified. The choice of the single pictures was semi-random, i.e. making sure that at least one image per class was present and prioritizing classes of images that were misclassified more often. This classification was performed by proposing a questionnaire to the subjects. A brief explanation of the task was provided, as well as the original visual stimuli delivered to the mouse (figure 3). Before answering the questions, they could also observe 3 examples of visual stimuli with the correspondent cortical pattern recorded with wide-field imaging, to learn how to associate one with the other.

4. Results

4.1. Classification of Mouse 1 images
Figure 6 displays the performance in terms of accuracy achieved from experiments 1 (**BASE**) to 8 (**PT-OOTF-AUG -1c**). Experiment **BASE** resulted in a mean classification accuracy of 51.54% ± 6.48% across the 5 stratified folds, above chance level. Offline data augmentation (**OFF-AUG**) led to an increase in the accuracy to 73.08% ± 5.84% and a median accuracy of 71.15%, while the integration of on-the-fly data augmentation (**OOTF-AUG**) led to a performance of 71.93% ± 6.73% accuracy, but with a median of 73.08%. Experiment **PT-BASE** consisted of pre-training the CNN to perform MNIST dataset classification and fine-tuning it with Mouse 1 dataset. The accuracy dropped to 17.30% ± 4.39%. Offline data augmentation (performed in **PT-OFF-AUG**) increased the accuracy to 71.54% ± 3.73%. On-the-fly data augmentation was added in experiment **PT-OOTF-AUG**, and the CNN classified the visual stimuli with a performance of 78.46% ± 3.31% accuracy on the test set. ROC curves relative to experiment **PT-OOTF-AUG** are displayed in figure 7, with an area under the curve of 0.97. The activation of the last hidden layer of the CNN trained with **PT-OOTF-AUG** is visualized for each stratified fold in the t-SNE plots displayed in figure 8, where colored point clouds represent the different classes. Experiments **PT-OOTF-AUG +1c** and **PT-OOTF-AUG -1c**, where a convolutional layer was added and removed respectively, resulted in accuracies of 75.38% ± 4.77% and

69.23% ± 5.71%. The difference between this result and experiment **PT-OOTF-AUG** accuracy is not statistically significant, showing that such changes in complexity are not crucial in affecting the performance of the CNN.

4.2. Classification of Mouse 2 and Mouse 3 images
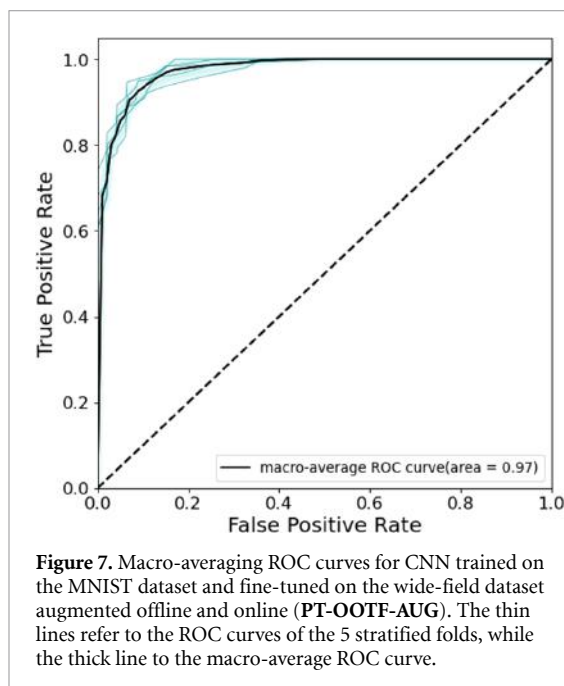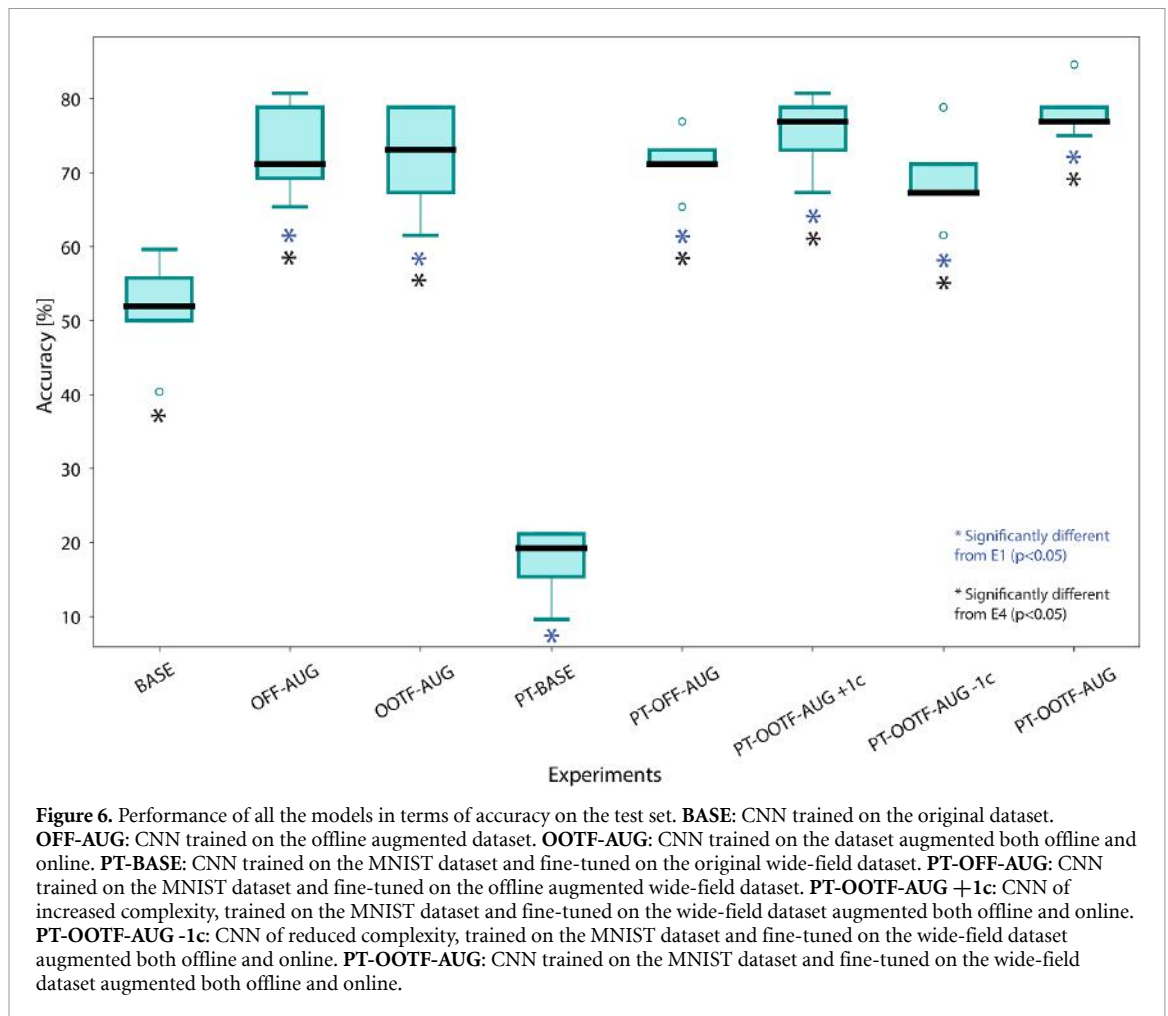The best-performing CNN on the Mouse 1 dataset, namely PT-OOTF-AUG, was used to investigate the possibility of generalization on Mouse 2 and Mouse 3 datasets. Its performance is shown in figure 9. Testing the model directly on Mouse 2 (**M2-TEST**) and Mouse 3 (**M3-TEST**) datasets resulted in accuracies that did not deviate from the chance level (6.90% ± 4.36% and 7.79% ± 1.73% respectively). When pre-training and fine-tuning the model according to the procedure of experiment **PT-OOTF-AUG** (MNIST pre-training), the accuracy increased to 47.58% ± 10.09% (median: 51.72%) for Mouse 2 (**M2-PT-OOTF-AUG**) and to 44.75% ± 3.80% (median: 45.76%) for Mouse 3 (**M3-PT-OOTF-AUG**). Transfer learning from a CNN pre-trained on Mouse 1 dataset resulted in an accuracy of 64.14% ± 10.81% when applied to Mouse 2 (**M2-TL**), and of 51.53% ± 6.48% for Mouse 3 (**M3-TL**).

4.3. Human-based classification
The outcome of human-based classification is summarized in table 3. Here, the classification results of both the CNN and the subjects are shown, along with the probability scores. In a CNN, this score is associated with the classification and expresses the probability of class membership for each of the visual stimuli. A high probability score associated with a correct classification and a low probability score associated with an incorrect classification are desirable [55, 56]. In the human-based classification case, this score has been expressed considering the majority of the subjects that gave the most voted answer. Among the 5 images correctly classified by the CNN, 4 were also correctly classified by humans. The exception is a 'Square' image, which was incorrectly classified by the subjects, but with a relatively low probability score (6/15), which generally indicates a higher likelihood of misclassification. The remaining 15 images were all misclassified by the CNN, while the subjects correctly classified them in 9 cases out of 15, but often with low probability scores (less than half of the subjects gave the right answer).

5. Discussion

Research on visual prosthetics has identified the optic nerve as a possible target for an implant [2–4]. In fact, despite the relatively low resolution achievable with respect to retinal and cortical prostheses, electrodes implanted in the optic nerve would allow the user to experience visual perception relative to the full field

Figure 6. Performance of all the models in terms of accuracy on the test set. **BASE**: CNN trained on the original dataset.
OFF-AUG: CNN trained on the offline augmented dataset. **OOTF-AUG**: CNN trained on the dataset augmented both offline and online. **PT-BASE**: CNN trained on the MNIST dataset and fine-tuned on the original wide-field dataset. **PT-OFF-AUG**: CNN trained on the MNIST dataset and fine-tuned on the offline augmented wide-field dataset. **PT-OOTF-AUG +1c**: CNN of increased complexity, trained on the MNIST dataset and fine-tuned on the wide-field dataset augmented both offline and online. **PT-OOTF-AUG -1c**: CNN of reduced complexity, trained on the MNIST dataset and fine-tuned on the wide-field dataset augmented both offline and online. **PT-OOTF-AUG**: CNN trained on the MNIST dataset and fine-tuned on the wide-field dataset augmented both offline and online.

Figure 7. Macro-averaging ROC curves for CNN trained on the MNIST dataset and fine-tuned on the wide-field dataset augmented offline and online (**PT-OOTF-AUG**). The thin lines refer to the ROC curves of the 5 stratified folds, while the thick line to the macro-average ROC curve.

of view [57]. However, the design of a neural prosthesis necessarily involves the choice of stimulation parameters that must be optimized. One possibility is to perform closed-loop stimulation using the cortical response evoked by the electrical stimulation as feedback, comparing it to the natural activation of the visual cortex [6] as schematized in figure 10. Hence, we need to verify the hypothesis that the cortical activation pattern can be used as an indicator of the visual perception of the animal, not to exclude *a priori* the possibility of using such activation as feedback. To do this, we propose recording and decoding large-scale cortical dynamics in response to visual stimuli, as successful decoding would indicate the possibility of discrimination between several visual stimuli based on the activity patterns recorded via wide-field calcium imaging. Such an insight would encourage us and the researchers in the field to further study the strengths of the evoked response in the visual cortex as a feedback signal.

We passively delivered simple visual stimuli to three anesthetized transgenic mice, recorded their V1 activity via wide-field calcium imaging, and classified the extracted frames, to investigate whether the information they convey could be exploited to automatically detect what was present in the visual field of the animal. We obtained above-chance classification accuracy for all the mice we analyzed, which suggests that the retinotopic organization of V1 can
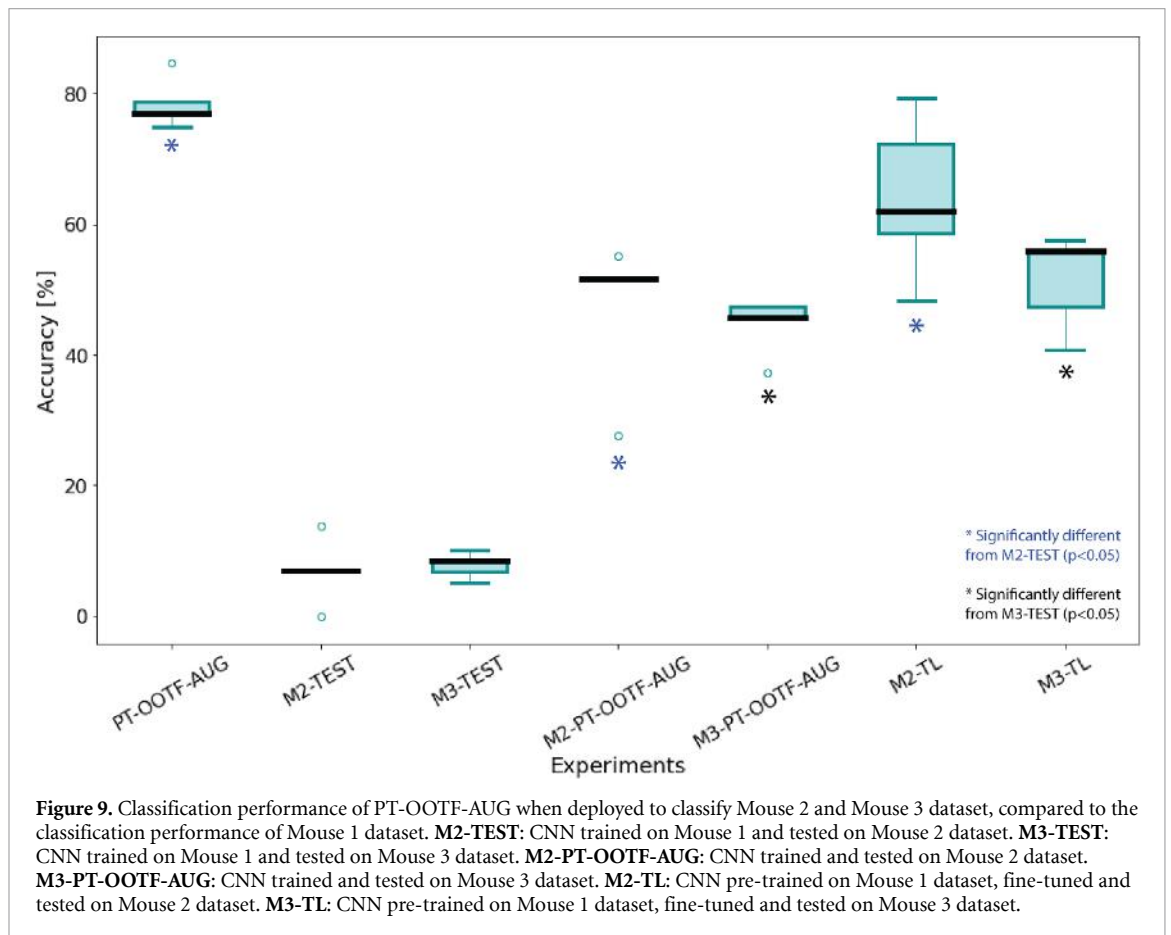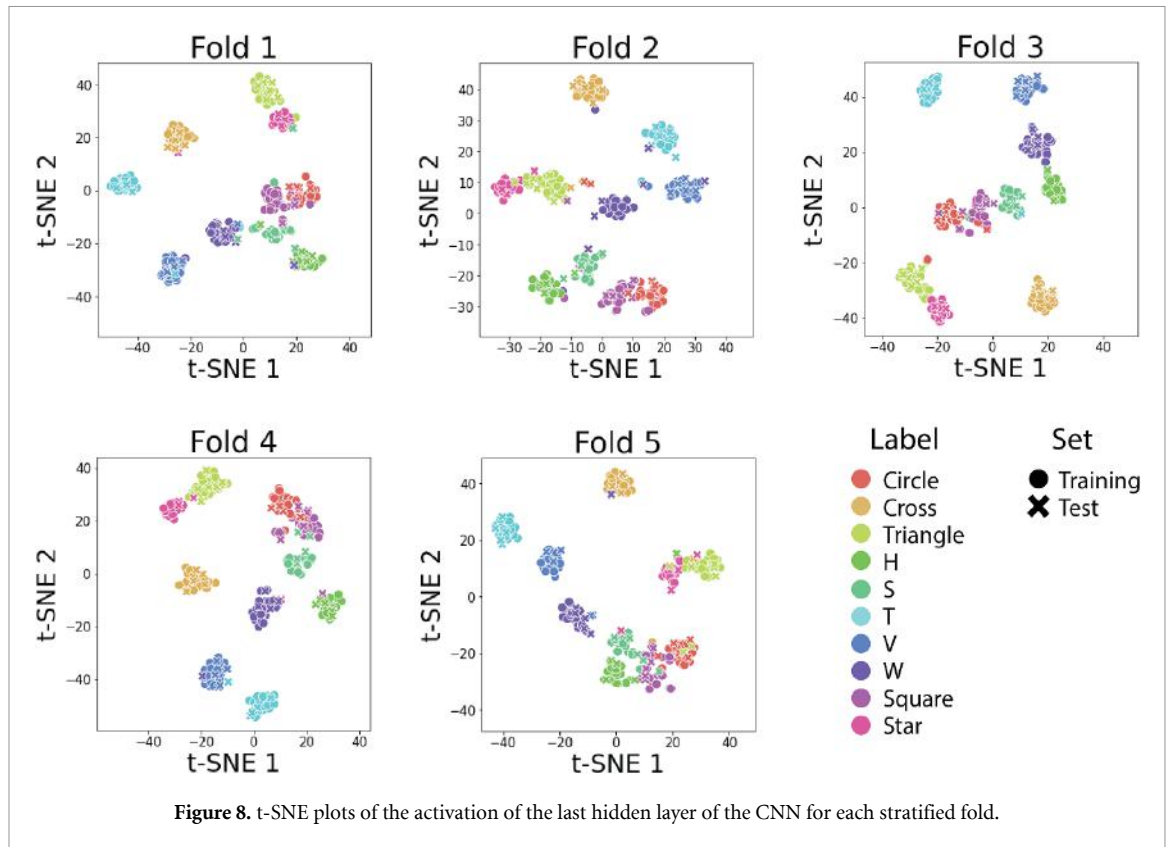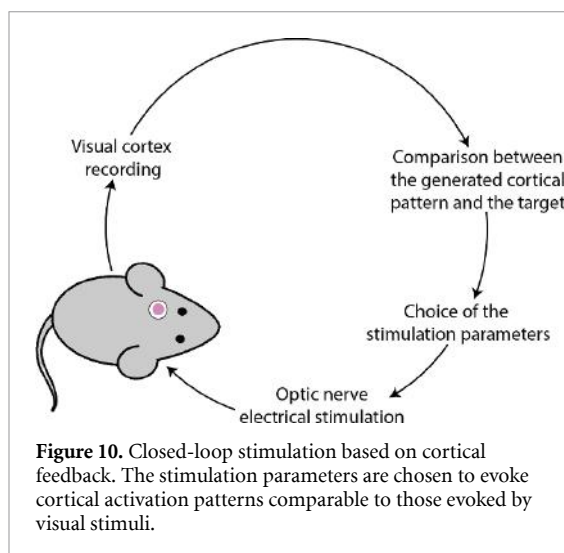
Figure 8. t-SNE plots of the activation of the last hidden layer of the CNN for each stratified fold.

Figure 9. Classification performance of PT-OOTF-AUG when deployed to classify Mouse 2 and Mouse 3 dataset, compared to the classification performance of Mouse 1 dataset. **M2-TEST**: CNN trained on Mouse 1 and tested on Mouse 2 dataset. **M3-TEST**: CNN trained on Mouse 1 and tested on Mouse 3 dataset. **M2-PT-OOTF-AUG**: CNN trained and tested on Mouse 2 dataset. **M3-PT-OOTF-AUG**: CNN trained and tested on Mouse 3 dataset. **M2-TL**: CNN pre-trained on Mouse 1 dataset, fine-tuned and tested on Mouse 2 dataset. **M3-TL**: CNN pre-trained on Mouse 1 dataset, fine-tuned and tested on Mouse 3 dataset.

Table 3. Human-based classification: results.

Class	CNN prediction	Probability	Subjects prediction	Probability
Circle	Circle	0,99	Circle	6/15
Circle	Square	0,90	Circle	13/15
Circle	Square	0,57	Circle	4/15
Circle	Square	0,57	Circle	5/15
Cross	T	0,54	Cross	14/15
H	H	1,00	H	7/15
H	W	0,80	W	7/15
S	Square	0,69	Square	8/15
S	W	0,42	Circle	6/15
T	S	0,47	T	13/15
V	V	1,00	V	14/15
V	W	0,32	Star	5/15
W	V	0,98	W	8/15
Square	Square	0,96	H	6/15
Square	Circle	0,92	Circle	5/15
Square	Circle	0,59	W	9/15
Star	Circle	0,55	Star	5/15
Star	Triangle	0,51	Star	5/15
Triangle	Triangle	1,00	Triangle	8/15
Triangle	Star	0,50	Triangle	6/15

Figure 10. Closed-loop stimulation based on cortical feedback. The stimulation parameters are chosen to evoke cortical activation patterns comparable to those evoked by visual stimuli.

be leveraged (1) to make inferences about what the animal is seeing, and (2) to rely on the feedback provided by V1 when designing future optic nerve stimulation experiments. In our hypothesis, the retinotopic organization is the reason for successful classification, since what is recorded in V1 through imaging and is classified by the CNN is a sort of projection of the visual field of the mouse. In the future, it will be reasonable to adjust the parameters of the electrical optic nerve stimulation to obtain a pattern shaped like the desired visual stimulus.

The first aspect to investigate was the feasibility of such a decoding method, and the first experiment(**BASE**) consisted of training and testing the CNN on the dataset divided into 5 stratified folds. The mean classification accuracy across the 5 stratified folds ($51.54\% \pm 6.48\%$) revealed the potential

of this choice, being largely above the chance level, although not fully satisfactory. The main hypothesis explaining this result was that the dataset was too limited to accurately perform the classification task.

Data augmentation is a commonly used strategy to tackle this issue. This technique reduces overfitting and improves the generalization capabilities of the network by acting at the root of the problem, under the assumption that from larger datasets it is possible to extract more information [51]. The dataset was augmented offline to perform experiment **OFF-AUG**, which resulted in a significant increase of the mean accuracy, with a difference of about 20%, showing the crucial role that the number of images plays in the training process of the CNN. Additionally, on-the-fly data augmentation was performed during training and without any data storage in experiment **OOTF-AUG**. The resulting accuracy was still significantly higher than **BASE**, but not significantly higher than **OFF-AUG**, despite a slight improvement. The difference between applying only offline and also on-the-fly data augmentation is not significant, but these results show how data availability highly affects the classification performance of this CNN.

Another paradigm we tested to limit the effect of the poor dataset size is transfer learning [58, 59]. This technique consists in training a neural network on a very large dataset and using the obtained weights as the initial weights of a new neural network instead of initializing them randomly, to train it in a new classification task. The effectiveness of this method is explained by the fact that big data allow the networks to learn low-level spatial features that are shared by many image datasets [51]. In our case, the number of classes of the MNIST dataset [53] and the semantic information of its images made it appropriate to

pre-train our CNN. First, fine-tuning was performed with the original, not augmented version of the dataset (**PT-BASE**). This was the experiment that resulted in the worst classification performance, with a mean accuracy lower than experiment **BASE**. The hypothesis that best explains this result is that the low number of image data was not enough to properly adjust the weights of the CNN, which probably did not deviate far enough from those fitted to the MNIST classification problem. Supporting this hypothesis, the mean accuracy drastically increased when applying offline data augmentation, confirming that an adequate number of data samples is crucial when training the CNN, especially in fine-tuning a pre-trained model. On-the-fly data augmentation was also embedded in the training procedure, leading to a further improvement in the mean classification accuracy that reached its maximum (78.46%), with minimal standard deviation (3.31%), proving to be the most robust model.

Once established that transfer learning was the best strategy to train the CNN, the focus of the experiments shifted to network complexity.

Existing models such as VGG16 were excluded due to their intrinsic complexity and high computational cost, in favor of a lighter architecture which allows us to reduce computing and memory resources and training time [60]. The ideal choice is a lightweight model, to prioritize both computation velocity and artificial intelligence sustainability [61]. Similar choices have been made in the literature, where lightweight CNNs with few convolutional layers have been used to perform wide-field calcium imaging data classification in mice, to classify sleep stages [62] and detect mild traumatic brain injury [63]. The effects of complexity variations were explored by changing the number of convolutional layers of our CNN, to determine whether the chosen complexity was the most suitable for this classification task. First, the CNN was pre-trained and fine-tuned with an additional convolutional layer (3 in total, **PT-OOTF-AUG +1c**), and at a later stage with only 1 convolutional layer (**PT-OOTF-AUG -1c**). The CNN performance resulting from these experiments was slightly lower, but not significantly, than the one resulting from **PT-OOTF-AUG**. This means that the CNN complexity was already appropriate for wide-field images classification but, concurrently, the number of convolutional layers is not a crucial factor when choosing the right architecture for this classification task.

Ideally, such an algorithm should perform well for every V1 wide-field calcium recording performed in the same conditions. The classification accuracy of Mouse 1 dataset was satisfactory, so the following step was to investigate the possibility of classifying wide-field calcium images recorded from other mice with the same CNN. The CNN was first tested on the image datasets recorded from Mouse 2 and Mouse 3 (**M2-TEST** and **M3-TEST**) to verify it. The classification

accuracy did not diverge from the chance level in both cases, showing that the network was not able to immediately generalize from one mouse to the others. A possible explanation can be found in the image datasets themselves. As seen in figure 4, the frames extracted from Mouse 2 and Mouse 3 recordings (figure 4(a)) are much noisier than those extracted from Mouse 1. In fact, in many Mouse 1 trials, it is possible to recognize the shapes with the naked eye. This is nearly impossible looking at the frames extracted from Mouse 2 and Mouse 3, characterized by the presence of opaque areas and clearly visible blood vessels. Therefore, this outcome was not unexpected. Conversely, above-chance classification accuracy was obtained by replicating experiment **PT-OOTF-AUG** on Mouse 2 (**M2-PT-OOTF-AUG**) and Mouse 3 (**M3-PT-OOTF-AUG**) datasets separately, relying on transfer learning and pre-training the CNN to classify the MNIST digits. The mean classification accuracy was still considerably lower than the one obtained with Mouse 1. This suggests that the quality of the data is extremely important for this classification task, much more than the size of the dataset. Indeed, Mouse 1 dataset is composed of 260 frames, Mouse has 145 and Mouse 3 has 295, and all of them underwent the same data augmentation process. Nevertheless, the resulting accuracy does not reflect the dataset size, but rather the clarity of the recorded images. The last experiment was performed exploiting transfer learning once again. The weights obtained training the CNN on Mouse 1 in Experiment **PT-OOTF-AUG** were used as initial weights to train the network to classify the images recorded from Mouse 2 (**M2-TL**) and Mouse 3(**M3-TL**). The mean classification accuracy increased in both cases showing that, although not able to automatically generalize, the weights of a CNN already trained to perform classification of wide-field recording from a mouse can be helpful when training the CNN to classify frames recorded from other animals. This means that generalization is achievable with transfer learning, which can be exploited to leverage features learned first from the MNIST dataset, and then those learned by training the network on an initial mouse. A similar procedure was adopted by Esteva *et al* in 2017 [64]. They pre-trained a CNN on the 2014 ImageNet Challenge and then fine-tuned it on medical images to classify skin cancer. This demonstrated the possibility of exploiting images even quite different from the ones to be classified, to improve the generalization capability of CNNs.

A total of 15 subjects were asked to classify a subset of 20 pictures extracted from Mouse 1 dataset, to compare their performance with the CNN. No frames were selected from Mouse 2 and Mouse 3 datasets. They were too distorted and the area where the shape should be most distinguishable was opaque so guessing the underlying visual stimulus was too complicated. What emerged from their answers is that humans misclassified fewer pictures with respect to

the CNN, although often with low probability scores. The fact that humans were able to guess what stimulus was being delivered to the mouse just by looking at its V1 recordings, leads to the conclusion that the CNN has potential for improvement. Further work will be dedicated to training the CNN on larger datasets. In fact, the limited data availability appears to be the major drawback of these experiments. We will collect new data from more animals including new classes, and investigate the use of learning-based data augmentation techniques based on deep learning, including autoencoders and generative models [51]. Out-of-distribution detection will also be included in this framework [65].

6. Conclusion

The goal of this study is to assess the possibility to use V1 activation as feedback to optimize the stimulation parameters for an optic nerve visual prosthesis. In this paper, we proposed an algorithm that exploits retinotopy to perform automatic decoding of visual stimuli from V1 in mice. These results suggest that a CNN can be exploited, which is sensitive not only to the size of the dataset used for training but mostly to its quality. In fact, applying the same training strategy, the best classification performance of 78.46% $\pm$ 3.31% classification accuracy was obtained when the least noisy dataset was used, i.e. Mouse 1 dataset. Transfer learning can be leveraged to overcome this issue and to allow the classification of lower-quality datasets. Pre-training the CNN on Mouse 1 dataset significantly improved the classification performance on Mouse 2 and Mouse 3 datasets after fine-tuning.

In conclusion, this strategy represents a potentially viable alternative to existing decoding methodologies, as well as a translational decoding technique that can allow us to shift these experiments to human subjects.

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iDs

Daniela De Luca ⓘ https://orcid.org/0000-0002-9623-0438
Sara Moccia ⓘ https://orcid.org/0000-0002-4494-8907
Leonardo Lupori ⓘ https://orcid.org/0000-0003-0458-5635
Raffaele Mazziotti ⓘ https://orcid.org/0000-0001-5344-5079
Tommaso Pizzorusso ⓘ https://orcid.org/0000-0001-5614-0668
Silvestro Micera ⓘ https://orcid.org/0000-0003-4396-8217

References

[1] Zrenner E 2013 Fighting blindness with microelectronics *Sci. Trans. Med.* **5** 210s16

[2] Veraart C, Raftopoulos C, Mortimer J T, Delbeke J, Pins D, Michaux G, Vanlierde A, Parrini S and Wanet-Defalque M C 1998 Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode *Brain Res.* **813** 181–6

[3] Brelén M E, Duret F, Gérard B, Delbeke J and Veraart C 2005 Creating a meaningful visual perception in blind volunteers by optic nerve stimulation *J. Neural Eng.* **2** S22

[4] Gaillet V, Cutrone A, Artoni F, Vagni P, Mega Pratiwi A, Romero S A, Lipucci Di Paola D, Micera S and Ghezzi D 2020 Spatially selective activation of the visual cortex via intraneural stimulation of the optic nerve *Nat. Biomed. Eng.* **4** 181–94

[5] Curcio C A and Allen K A 1990 Topography of ganglion cells in human retina *J. Comp. Neurol.* **300** 5–25

[6] Romeni S, Zoccolan D and Micera S 2021 A machine learning framework to optimize optic nerve electrical stimulation for vision restoration *Patterns* **2** 100286

[7] Losanno E, Badi M, Wurth S, Borgognon S, Courtine G, Capogrosso M, Rouiller E M and Micera S 2021 Bayesian optimization of peripheral intraneural stimulation protocols to evoke distal limb movements *J. Neural Eng.* **18** 066046

[8] Gaillet V, Borda E, Zollinger E G and Ghezzi D 2021 A machine-learning algorithm correctly classifies cortical evoked potentials from both visual stimulation and electrical stimulation of the optic nerve *J. Neural Eng.* **18** 046031

[9] Kampa B M, Roth M M, Göbel W and Helmchen F 2011 Representation of visual scenes by local neuronal populations in layer 2/3 of mouse visual cortex *Front. Neural Circuits* **5** 18

[10] Garrett M E, Nauhaus I, Marshel J H and Callaway E M 2014 Topography and areal organization of mouse visual cortex *J. Neurosci.* **34** 12587–600

[11] Zhuang J, Ng L, Williams D, Valley M, Li Y, Garrett M and Waters J 2017 An extended retinotopic map of mouse cortex *eLife* **6** e18372

[12] Ren C and Komiyama T 2021 Characterizing cortex-wide dynamics with wide-field calcium imaging *J. Neurosci.* **41** 4160–8

[13] Sabatini B L and Tian L 2020 Imaging neurotransmitter and neuromodulator dynamics *in vivo* with genetically encoded indicators *Neuron* **108** 17–32

[14] Christine G and Arthur K 2012 Imaging calcium in neurons *Neuron* **73** 862–85

[15] Nietz A K, Popa L S, Streng M L, Carter R E, Kodandaramaiah S B, Ebner T J 2022 Wide-field calcium

imaging of neuronal network dynamics *in vivo Biology* **11** 1601

[16] Conti E *et al* 2020 Synergic effect of optogenetic stimulation and motor training boosts recovery of motor functionality after stroke supported by segregation of motor representation *bioRxiv Preprint* https://doi.org/10.1101/2020.03.05.974972 (posted online 6 March 2020, accessed 28 July 2022)

[17] Conti E *et al* 2021 Restoration of motor-evoked cortical activity is a distinguishing feature of the most effective rehabilitation therapy after stroke *bioRxiv Preprint* https://doi.org/10.1101/2020.03.05.974972 (posted online 31 January 2021, accessed 28 July 2022) pp 2020–03

[18] Conti E *et al* 2022 Combining optogenetic stimulation and motor training improves functional recovery and perilesional cortical activity *Neurorehabil. Neural Repair.* **36** 107–18

[19] West S L *et al* 2021 Wide-field calcium imaging of dynamic cortical networks during locomotion *Cerebral Cortex* **32** 2668–87

[20] Cramer J V, Gesierich B, Roth S, Dichgans M, Düring M and Liesz A 2019 *In vivo* widefield calcium imaging of the mouse cortex for analysis of network connectivity in health and brain disease *NeuroImage* **199** 570–84

[21] Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, van der Laak J A W M, van Ginneken B and Sánchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88

[22] DiCarlo J J, Zoccolan D and Rust N C 2012 How does the brain solve visual object recognition? *Neuron* **73** 415–34

[23] Niell C M and Stryker M P 2008 Highly selective receptive fields in mouse visual cortex *J. Neurosci.* **28** 7520–36

[24] Hubel D H and Wiesel T N 1959 Receptive fields of single neurones in the cat's striate cortex *J. Physiol.* **148** 574

[25] Zhang Y J *et al* 2022 Neural decoding of visual information across different neural recording modalities and approaches *Mach. Intell. Res.* **19** 1–16

[26] Stringer C and Pachitariu M 2019 Computational processing of neural recordings from calcium imaging data *Curr. Opin. Neurobiol.* **55** 22–31

[27] Stringer C, Michaelos M, Tsyboulski D, Lindo S E and Pachitariu M 2021 High-precision coding in visual cortex *Cell* **184** 2767–78

[28] Iqbal A, Dong P, Kim C M and Jang H 2019 Decoding neural responses in mouse visual cortex through a deep neural network *2019 Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) pp 1–7

[29] Grewe B F, Langer D, Kasper H, Kampa B M and Helmchen F 2010 High-speed *in vivo* calcium imaging reveals neuronal network activity with near-millisecond precision *Nat. Methods* **7** 399–405

[30] Pachitariu M, Stringer C and Harris K D 2018 Robustness of spike deconvolution for neuronal calcium imaging *J. Neurosci.* **38** 7976–85

[31] Garasto S, Bharath A A and Schultz S R 2018 Visual reconstruction from 2-photon calcium imaging suggests linear readout properties of neurons in mouse primary visual cortex *bioRxiv Preprint* https://doi.org/10.1101/300392 (posted online 12 April 2018, accessed 31 August 2022) p 300392

[32] Ellis R J and Michaelides M 2018 High-accuracy decoding of complex visual scenes from neuronal calcium responses *BioRxiv Preprint* https://doi.org/10.1101/271296 (posted online 26 March 2018, accessed 26 August 2022) p 271296

[33] Shen H 2016 Brain-data gold mine released: massive survey of mouse visual-cortex activity aims to reveal brain's computational rules *Nature* **535** 209–11

[34] Cai L, Wu B and Ji S 2018 Neuronal activities in the mouse visual cortex predict patterns of sensory stimuli *Neuroinformatics* **16** 473–88

[35] de Vries S E J *et al* 2018 A large-scale, standardized physiological survey reveals higher order coding throughout the mouse visual cortex *bioRxiv Preprint* https://doi.org/10.1101/359513 (posted online 29 June 2018, accessed 2 June 2022) p 359513

[36] Yoshida T and Ohki K 2020 Natural images are reliably represented by sparse and variable populations of neurons in visual cortex *Nat. Commun.* **11** 1–19

[37] Haxby J V, Gobbini M I, Furey M L, Ishai A, Schouten J L and Pietrini P 2001 Distributed and overlapping representations of faces and objects in ventral temporal cortex *Science* **293** 2425–30

[38] Dinstein I, Gardner J L, Jazayeri M and Heeger D J 2008 Executed and observed movements have different distributed representations in human aIPS *J. Neurosci.* **28** 11231–9

[39] Kamitani Y and Tong F 2005 Decoding the visual and subjective contents of the human brain *Nat. Neurosci.* **8** 679–85

[40] Brouwer G J and Heeger D J 2009 Decoding and reconstructing color from responses in human visual cortex *J. Neurosci.* **29** 13992–4003

[41] Horikawa T and Kamitani Y 2017 Generic decoding of seen and imagined objects using hierarchical visual features *Nat. Commun.* **8** 1–15

[42] Jiao Z, You H, Yang F, Li X, Zhang H and Shen D 2019 *Decoding EEG by Visual-Guided Deep Neural Networks* (IJCAI) pp 1387–93

[43] Kalafatovich J, Lee M and Lee S W 2020 Decoding visual recognition of objects from eeg signals based on attention-driven convolutional neural network *2020 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC)* pp 2985–90

[44] Bagchi S and Bathula D R 2022 EEG-ConvTransformer for single-trial EEG-based visual stimulus classification *Pattern Recognit.* **129** 108757

[45] Jimenez L O, Tring E, Trachtenberg J T and Ringach D L 2018 Local tuning biases in mouse primary visual cortex *J. Neurophysiol.* **120** 274–80

[46] Fahey P G *et al* 2019 A global map of orientation tuning in mouse visual cortex *bioRxiv Preprint* https://doi.org/10.1101/745323 (posted online 23 August 2019, accessed 24 November 2022) p 745323

[47] Van den Bergh G, Zhang B, Arckens L and Chino Y M 2010 Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys *J. Comp. Neurol.* **518** 2051–70

[48] Goldey G J, Roumis D K, Glickfeld L L, Kerlin A M, Reid R C, Bonin V, Schafer D P and Andermann M L 2014 Removable cranial windows for long-term imaging in awake mice *Nat. Protocols* **9** 2515–38

[49] Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R 2012 Improving neural networks by preventing co-adaptation of feature detectors (arXiv:12070580)

[50] Chlap P, Min H, Vandenberg N, Dowling J, Holloway L and Haworth A 2021 A review of medical image data augmentation techniques for deep learning applications *J. Med. Imaging Radiat. Oncol.* **65** 545–63

[51] Shorten C and Khoshgoftaar T M 2019 A survey on image data augmentation for deep learning *J. Big Data* **6** 1–48

[52] Weiss K, Khoshgoftaar T M and Wang D 2016 A survey of transfer learning *J. Big Data* **3** 1–40

[53] Deng L 2012 The mnist database of handwritten digit images for machine learning research *IEEE Signal Process. Mag.* **29** 141–2

[54] Van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605

[55] Moccia S, Wirkert S J, Kenngott H, Vemuri A S, Apitz M, Mayer B, De Momi E, Mattos L S and Maier-Hein L 2018 Uncertainty-aware organ classification for surgical data science applications in laparoscopy *IEEE Trans. Biomed. Eng.* **65** 2649–59

[56] Jiang H, Kim B, Guan M and Gupta M 2018 To trust or not to trust a classifier *Advances in Neural Information Processing Systems* vol 31

[57] Mirochnik R M and Pezaris J S 2019 Contemporary approaches to visual prostheses *Mil. Med. Res.* **6** 1–9

[58] Zhu Y *et al* 2011 Heterogeneous transfer learning for image classification *25th AAAI Conf. on Artificial Intelligence*

[59] Shaha M and Pawar M 2018 Transfer learning for image classification *2018 2nd Int. Conf. on Electronics, Communication and Aerospace Technology (ICECA)* (IEEE) pp 656–60

[60] Lei F, Liu X, Dai Q and Ling B W K 2020 Shallow convolutional neural network for image classification *SN Appl. Sci.* **2** 1–8

[61] van Wynsberghe A 2021 Sustainable AI: AI for sustainability and the sustainability of AI *AI Ethics* **1** 213–8

[62] Zhang X, Landsness E C, Chen W, Miao H, Tang M, Brier L M, Culver J P, Lee J-M and Anastasio M A 2022 Automated sleep state classification of wide-field calcium imaging data via multiplex visibility graphs and deep learning *J. Neurosci. Methods* **366** 109421

[63] Koochaki F, Shamsi F and Najafizadeh L 2020 Detecting mtbi by learning spatio-temporal characteristics of widefield calcium imaging data using deep learning *2020 42nd Annual Int. Conf. of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 2917–20

[64] Esteva A, Kuprel B, Novoa R A, Ko J, Swetter S M, Blau H M and Thrun S 2017 Dermatologist-level classification of skin cancer with deep neural networks *Nature* **542** 115–8

[65] Hendrycks D and Gimpel K 2016 A baseline for detecting misclassified and out-of-distribution examples in neural networks (arXiv:161002136)