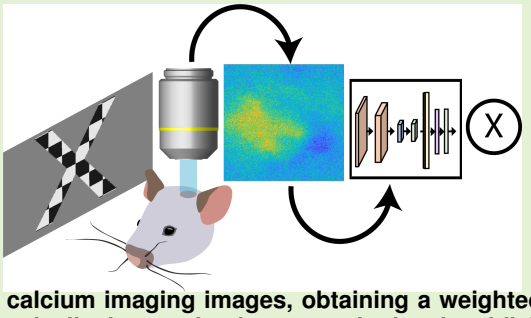
Predicting visual stimuli from cortical response recorded with wide-field imaging in a mouse

D. De Luca, S. Moccia, L. Lupori, R. Mazziotti, T. Pizzorusso, and S. Micera *Fellow, IEEE*

***Abstract*— Neural decoding of the visual system is a subject of research interest, both to understand how the visual system works and to be able to use this knowledge in areas such as computer vision or brain-computer interfaces. Spike-based decoding is often used, but it is difficult to record data from the whole visual cortex and it requires proper pre-processing. We here propose a decoding method that combines wide-field calcium brain imaging, which allows us to obtain large-scale visualization of cortical activity with a high signal-to-noise ratio, and convolutional neural networks (CNNs). A mouse was presented with 10 different visual stimuli, and the activity from its primary visual cortex (V1) was recorded. A CNN we designed was then compared to other existing commonly used CNNS, that were trained to classify the visual stimuli from wide-field calcium imaging images, obtaining a weighted F1 score of more than 0.70 on the test set, showing it is possible to automatically detect what is present in the visual field of the animal.**



***Index Terms*— Deep learning, transfer learning, visual cortex, visual prostheses, wide-field imaging**

I. Introduction

THE study of visual system has always been of great interest among researchers, as vision constitutes about 70% of the perceptual information acquired by humans, and understanding neural coding is crucial to comprehend how the brain interprets environmental stimuli [1]. Decoding of visual information aims to derive a visual stimulus from the response generated by the stimulus itself and is of critical importance not only to study how the brain processes information, but also to advance artificial vision technology used in virtual reality and brain-computer interfaces [2] [3].

Mice are frequently employed in visual cortex decoding for a variety of reasons, despite their tiny size and low-resolution vision. The mouse visual cortex shares several characteristics with the human visual cortex, including retinotopy, different types of receptive fields, orientation tuning, and plasticity in ocular dominance [4] [5]. Additional factors include the availability of transgenic animals, the ability to record vast amounts of data, and their relatively simple maintenance [6].

Spike signals are the most common type of neural data used in animal visual decoding studies [3]. It is possible to obtain spike data using intracortical probes, or through two-photon calcium imaging paired with proper algorithms [7]. As a consequence, plenty of state-of-the-art algorithms exist to analyze and decode spike data, and most of the visual neural decoding studies are spike-based [8] [9]. Concurrently, a variety of algorithms, from linear to Bayesian-based to deep neural network techniques, can produce excellent decoding performance [3]. Intracortical microelectrodes are the most practical when accessing deeper levels of the cortex, and recordings frequently yield good decoding performance. However, they do not enable us to record the activity throughout the entire visual cortex surface, so that only a fraction of the visual field's cortical activity is recorded because of the visual cortex's retinotopic architecture [4] [10] [11].

Calcium imaging is a recent and still less explored technique, which is typically used to record from large portions of the brain [3]. A commonly applied decoding strategy is to gather data using two-photon Ca^{2+} imaging and to convert the signal into spikes to exploit the many available algorithms [12] [13] [14]. Wide-field calcium imaging, to our knowledge, has never been used in the literature to perform visual cortex decoding. This technique allows us to record from the whole primary visual cortex (V1) of the animal with high signal-to-noise-ratio (SNR) [15], enough to eliminate the need for averaging the signals over multiple trials and enabling us to gather data on which to use machine learning

Submitted on February 1^{st}, 2023. This work was partially supported by the association "Sulle Ali Di Un Sogno ONLUS".

D.D.L., S. Moccia, and S. Micera are with the BioRobotics Institute and Department of Excellence in Robotics AI, Pisa, Italy (e-mails: daniela.deluca@santannapisa.it, sara.moccia@santannapisa.it, silvestro.micera@santannapisa.it).

S. Micera further inherits the position of Bertarelli Foundation Chair in Translational NeuroEngineering, Centre for Neuroprosthetics and Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

R.M., L.L., and T.P. are with the Institute of Neuroscience (IN-CNR), Scuola Normale Superiore, Pisa, Italy (e-mails: leonardo.lupori@sns.it, raffaelemario.mazziotti@unifi.it, tommaso.pizzorusso@in.cnr.it)

An earlier version of this paper was presented at 2020 IEEE SENSORS Conference and was published in its Proceedings (doi: 10.1109/SENSORS52175.2022.9967250)

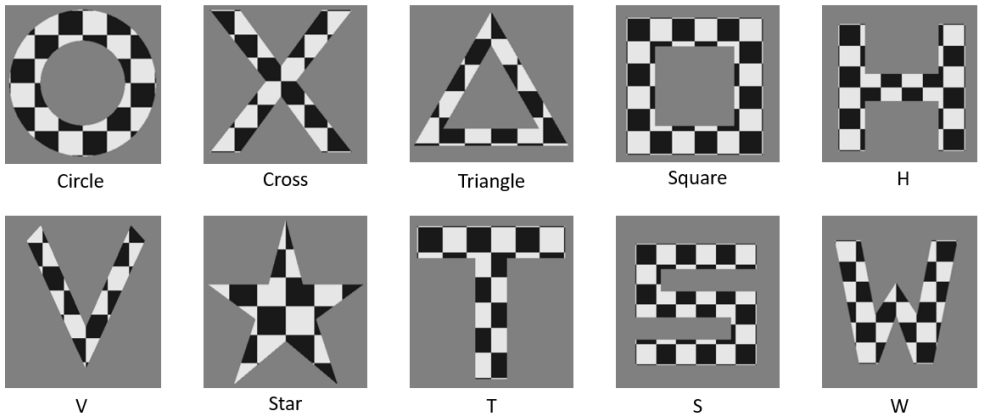


Fig. 1: Visual stimuli delivered to the mouse.

methods. It exploits engineered proteins that have the property of expressing fluorescence in response to certain neuronal events [15] [16]. Genetically encoded calcium indicators (GECIs) are a class of these proteins and are specific for detecting alterations in intracellular calcium, which indirectly indicates neural activity [17]. Among these, the GCamP family is the most common in neuroscience [18]. *In vivo* experiments usually involve the head fixation of the animal, whose skull can be thinned to achieve partial transparency or removed and substituted with transparent materials [18]. A light source, e.g., a LED, is then used along with a light-sensing device connected to a microscope to capture fluorescence [17] [18].

Here, we propose to exploit wide-field calcium imaging to record the V1 response of a mouse and deep learning architectures to perform V1 response decoding. We study convolutional neural networks (CNNs) to classify V1 responses evoked by 10 classes of visual stimuli passively presented to a transgenic mouse. We rely on CNNs as they are the standard not only for medical imaging sensors analysis [19] [20] and for brain signal classification [21] [22], but are also extensively used for image classification across several sensor modalities [23] [24] [25], as their versatility allows us to optimize the usage of sensor data. In the literature, CNN-based approaches have been adopted to analyze wide-field calcium imaging recordings in the mouse cortex to classify sleep stages [26] and to detect mild traumatic brain injury [27].

In this work, we designed a CNN to perform visual cortex responses classification and compared its performance against that of CNNs commonly used in medical image analysis: Inception V3 [28], VGG 16 [29], and Inception-ResNet V2 [30]. The contributions of this study can be summarized as follows:

- Collection of a dataset of V1 responses to standardized visual stimuli, acquired via wide-field calcium imaging *in vivo*.
- Design of a CNN to perform V1 responses decoding from wide-field calcium images recorded *in vivo*.
- Comparison of the performance of state-of-the-art CNNs used for classification tasks in medical images.

II. Materials and Methods

The visual stimuli shown to the mouse and the procedure of acquisition of the cortical response of the mouse are described in Sec. II-A. Section II-B presents an overview of the dataset analyzed in this work and data augmentation methodologies, Sec. II-C describes the CNN architectures and Sec. II-D our experimental protocol to compare the CNNs performance.

A. Data acquisition

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). In this work, we performed experiments on a GCaMP6f transgenic mouse, which was deeply anesthetized with Isoflurane 3% and kept with the head fixed, while a screen was positioned 20 cm from its right eye. Anesthesia only affects the temporal dynamics of the cortical response of the animals and does not alter the topography of the activation pattern in V1, which is taken into account by our classification algorithms [31].

The visual stimuli consisted of 60 repetitions of 10 different shapes, shown in Fig. 1, which were delivered at the center of the screen. Their outlines were filled with a checkerboard that flickered at a temporal frequency of 5 Hz and a spatial frequency of 0.08 cpd, on a gray background. This choice was made to enhance the response amplitude of V1 neurons, sensitive to contrast and orientation [32] [33]. Each trial consisted of the presentation of a stimulus with a frequency of 0.5 Hz. The sequence was: 500 ms of a pre-stimulus gray background, 500 ms of stimulus duration, and 1 s of post-stimulus gray background to let the cortical response return to baseline.

A custom Leica fluorescence microscope (Leica Microsystems) was used to visualize V1 activity, and was equipped with a Leica Z6 APO coupled with a Leica PlanApo 2.0x (10447178) objective. An I3 cube (excitation BP 450-490 nm dichroic 510 nm emission LP 515 nm) was used to detect fluorescence. Images were acquired with a 12-bit depth acquisition camera (PCO edge 5.5) at 10 fps with a resolution of 270x320 pixels, for a total of 20 frames per trial.

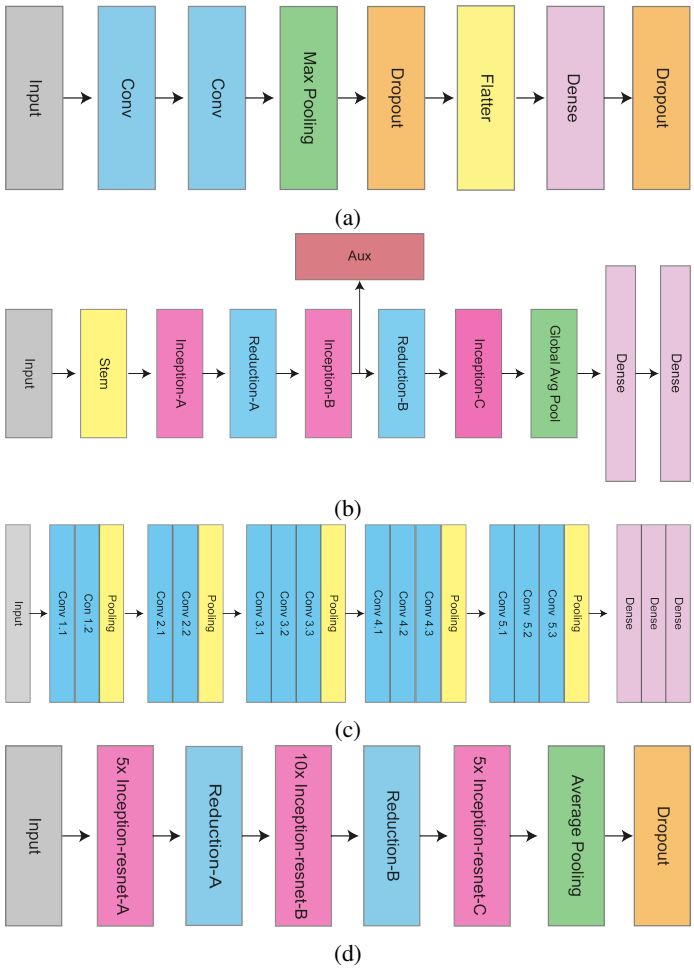


Fig. 2: Architectures of the CNNs trained and tested in this work. (a) The CNN we designed, (b) Inception V3, (c) VGG 16, (d) Inception-ResNet V2

B. Data preprocessing

For every trial, we visualized $\Delta F/F_0$, where ΔF and F_0 are the variation of fluorescence with respect to the baseline and the baseline fluorescence, respectively. F_0 has been calculated as the mean fluorescence intensity across the frames corresponding to the response to the pre-stimulus gray background. The mean across the three central frames corresponding to the presentation of the stimulus (frames 7, 8, and 9) was computed to obtain a final, smoothed image to be included in the dataset.

The images were visually inspected to discard trials where cortical activity was not present or only partially visible. A sample of the final dataset is shown in Fig. 3. After this data curation process, the selected samples were 260, divided as follows: 25 for Circle, 29 for Cross, 24 for H, 22 for S, 26 for T, 28 for V, 28 for W, 28 for Square, 19 for Star and 31 for Triangle.

The images were prepared to train the CNNs with a factor of 2 downsampling and mean subtraction.

C. Convolutional neural networks

This Section describes the proposed CNN and the CNNs models we considered as competitors.

1) Our CNN: The CNN we developed has 9 layers, as summarized in Table I and Fig. 2a. Of these, 2 are convolutional with 32 and 64 channels, respectively. Their kernels have a size of 3x3 and they both have a hyperbolic tangent function as activation function. Dropout layers are included, to attenuate the chance of overfitting. This architecture proved to be effective for this classification task, while being a lightweight model with low computational and memory requirement [34] [35]. The hyperparameters (learning rate, optimizer, number of epochs, batch size) were adjusted through an empirical, iterative process rather than a systematic tuning procedure.

2) Inception V3: The first CNN we compared against our model is Inception V3 [28]. This CNN (Fig. 2b) is the third evolution of the GoogLeNet CNN, winner of the ILSVRC 2014 competition. The innovation proposed by Google was the Inception module, which allowed to reduce the number of network parameters while increasing the network depth. Multi-scale feature extraction is allowed by parallel filters of different sizes [36]. Inception V3 achieved a top-1 accuracy of 77.9% and a top-5 accuracy of 93.7% on the ImageNet [37] dataset. Those are the fraction of test images for which the classification was correct or among the highest probable labels provided by the CNN [36], respectively.

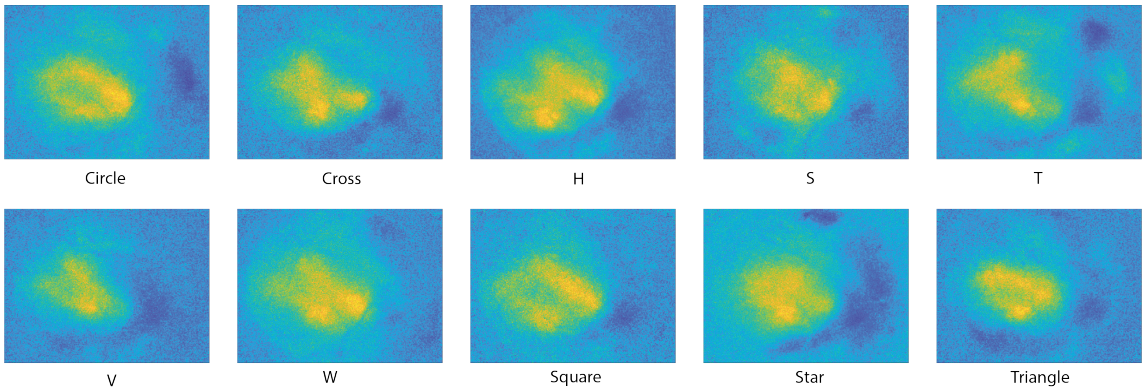


Fig. 3: Example of cortical responses for each of the visual stimuli, recorded via wide-field calcium imaging.

3) VGG 16: VGG 16 (Fig. 2c) was introduced by the Oxford's Visual Geometry Group (VGG) for ILSVRC in 2014 [29]. This CNN is composed of 16 layers and is an improvement of the AlexNet CNN [38], where the large-sized kernel filters are replaced with sequences of 3x3 filters [36]. VGG 16 classified the ImageNet dataset with a top-1 accuracy of 71.5% and a top-5 accuracy of 89.8%.

4) Inception-ResNet V2: Inception-ResNet V2 (Fig. 2d) is a hybrid architecture that combines inception module and residual connections, linking the input and output of the Inception V4 blocks [30]. Inception-ResNet V2 achieved a top-1 accuracy of 80.4% and a top-5 accuracy of 95.3% on ImageNet.

D. Training strategy and performance assessment

The dataset was divided into 5 stratified folds. During training, 25% of the training set (52 samples) was used as the validation set. Offline and on-the-fly data augmentation were performed by applying geometrical transformations to the datasets [39]. Offline data augmentation was performed by applying 4 different geometrical transformations to the images, which were a random rotation in a range of ± 20 degrees, a random spatial shear of a factor between 0 and 0.2, random corrections of brightness with a maximum variation of 40%, random corrections of contrast with a lower and upper bound of variation of respectively 20% and 50%. The augmentation factor was set to 8 so the training set went from 166 to 5312 images. Of these, 20% (1664 images) were used as validation set during training.

TABLE I: Architecture of the convolutional neural network we designed to perform V1 responses classification.

Name	Kernel size	Output shape
Input	-	135x160x1
Convolutional layer	3x3	133x158x32
Convolutional layer	3x3	131x156x64
Max pooling layer	2x2	65x78x64
Dropout	-	65x78x64
Flattening layer	-	324480
Dense layer	-	128
Dropout	-	128
Output	-	10

On-the-fly data augmentation was also performed during training, which may contribute to reduce the chance of overfitting [40]. The geometrical transformations were the following: random rotation in the range of ± 5 degrees, random vertical and horizontal shift of a factor of ± 0.02, random zoom of a factor of ± 0.3, random shear with intensity in a range of ± 2.

To reduce the impact of the small size of the dataset, we also exploited transfer learning [41] [42]. This method involves training a neural network on a larger dataset and using the resulting weights as the starting weights for a new neural network. Big data enables CNNs to extract low-level spatial properties that are shared by multiple image datasets, which accounts for this method's efficacy [40].

All the experiments were performed on a Dell XPS 8940, Intel Core i7-10700 processor, 2.90GHz CPU, NVIDIA GeForce GTX 1660 Ti GPU, 16 GB RAM.

1) Ablation study: We performed 3 experiments to test different training strategies, with the aim of individuating the most suitable for the CNN we designed. **A**: The CNN was trained on the original dataset (i.e., without data augmentation) using Adadelta as optimizer for 100 epochs and batch size equal to 16, maximizing the accuracy on the validation set. We used the categorical cross-entropy as loss function.

B: The same training procedure was repeated with the dataset augmented both offline and on-the-fly.

C: To compensate for the limited size of the dataset, we pre-trained the CNN to classify the MNIST handwritten digits dataset [43], and fine-tuned it with our images. The MNIST dataset is composed of 60000 training samples and 10000 test samples, which are 28x28 pixels greyscale images. We chose this dataset because it shares similar characteristics in terms of semantic content with our calcium imaging dataset. Each MNIST digit is indeed a single, bright object displayed on a dark background. The MNIST digits were resized to 135x160 pixels to match the wide-field frames, normalized between 0 and 1, and fed into the CNN for training. The CNN was trained for 25 epochs on batches of 32 samples, enough to reach an accuracy of 95% on the test set. Categorical cross-entropy was used as loss function, and Adadelta as optimizer to maximize the MNIST classification accuracy. The weights of the model trained on MNIST digits were loaded in the CNN and fine-tuning was done in two steps. The first step was to freeze the

weights of the first convolutional layer of the network and to train the rest of the model (Adadelta optimizer, accuracy as metrics, 100 epochs and size 16 batches). The second step consisted of unfreezing the weights of the first layer and fine-tuning the model again, for 20 epochs, using Adam optimizer with a learning rate of 10^{-4}. A repeated stratified 5 fold cross-validation with 5 repetitions was performed.

2) Existing CNNs: Transfer learning was also exploited to train the CNNs we used to compare our model. In all 3 cases we used pre-trained CNNs on the ImageNet dataset, removed their last fully connected layers, and fine-tuned the CNNs on our wide-field calcium imaging dataset augmented both offline and on-the-fly. Inception V3 was frozen until its 288^{th} layer, VGG 16 until its 3^{rd} convolutional block (3 layers) and Inception-ResNet V2 until its 288^{th} layer. The CNNs were fine-tuned for 25 epochs using Adam as optimizer, with a learning rate of $5 * 10^{-5}$. The loss function to be minimized was the categorical cross-entropy, and the metrics to be maximized was the accuracy on the validation set. To compare the classification performance achieved by the state-of-the-art CNNs and the one we designed, we implemented a repeated stratified 5 fold cross-validation with 5 repetitions.

3) Performance assessment: The classification performance of our CNN trained according to the 3 different strategies previously described was evaluated using the accuracy (Eq. 1) achieved on the test set across the stratified 5 folds. The performance of the CNN we designed was compared against that of the state-of-the-art CNNs by computing the weighted F1 score (Eq. 2).

$$\textbf{accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of images}} \tag{1}$$

$$\textbf{F1 score} = 2\frac{\text{Precision * Recall}}{\text{Precision + Recall}} \tag{2}$$

The receiver operating characteristic (ROC) curves were computed for each stratified fold, and the area (AUC) under the macro-averaging ROC curve was calculated.

Statistical significance between results was evaluated using the one-way ANOVA test after performing the Shapiro-Wilk normality test (significance level α = 0.05 in both tests), to assess the distribution of the data.

III. Results

A. Training strategy selection for our CNN

As shown in Fig. 4, the mean classification accuracy across folds resulting from experiment **A** was 51.54%±6.48%. This result indicates that our CNN can perform classification above the chance level, but the accuracy is relatively low when no data augmentation is applied.

The classification performance improved significantly ($p < 0.05$) by training the model according to experiment **B**, achieving an accuracy across folds of 71.93%±6.73%, with a median of 73.08%.

Experiment **C**, that consisted in applying transfer learning to the model pre-trained on the MNIST dataset, classified the visual stimuli with a performance of 78.46%±3.31% accuracy on the test set. The ROC curves relative to experiment **C**, in Fig. 5, show an area under the curve (AUC) of 0.97.

B. Comparison with existing CNNs

The F1 scores achieved by the tested classifiers across the 5 repetitions of the repeated stratified 5 folding are displayed in the boxplots in Fig. 6, while Fig.7 shows the confusion matrices and ROC curves resulting from the first repetition of the repeated cross-validation. Inception V3 achieved an F1 score of 0.15±0.03 (mean ± std), with an AUC of 0.81±0.02; VGG16 achieved an F1 score of 0.72±0.09, with an AUC of 0.94±0.04; Inception-ResNet v2 achieved an F1 score of 0.75±0.02, with an AUC of 0.95±0.02; our CNN achieved an F1 score of 0.77±0.02, with an AUC of 0.97±0.00.

Inception V3 performed significantly worse than the other networks, including our CNN ($p < 0.05$). The CNN we

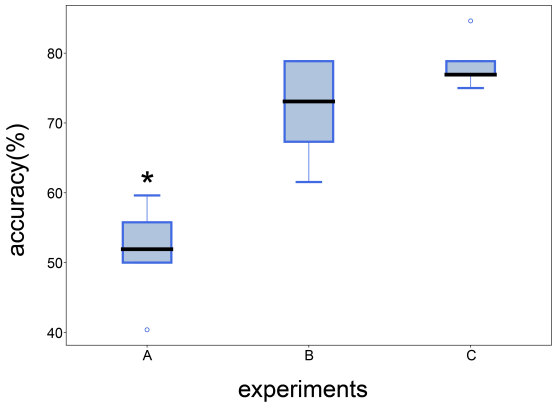


Fig. 4: Performance of our CNN in terms of accuracy on the test set. **A**: CNN trained on the original dataset. **B**: CNN trained on the dataset augmented offline and online. **C**: CNN trained on the MNIST dataset and fine-tuned on the wide-field dataset augmented offline and online.

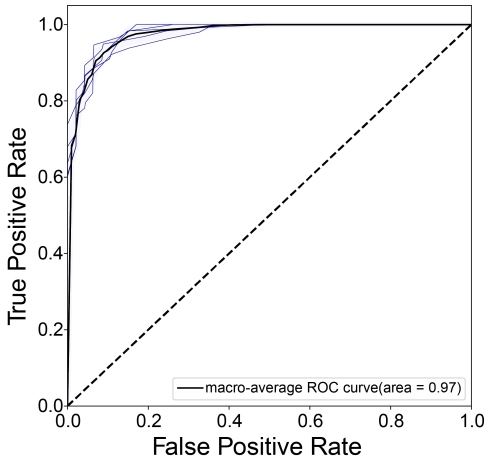


Fig. 5: ROC curves for our CNN trained on the MNIST dataset and fine-tuned on the wide-field dataset augmented offline and online (experiment **C**). The thin lines refer to the ROC curves of the 5 stratified folds, while the thick line to the macro average ROC curve.

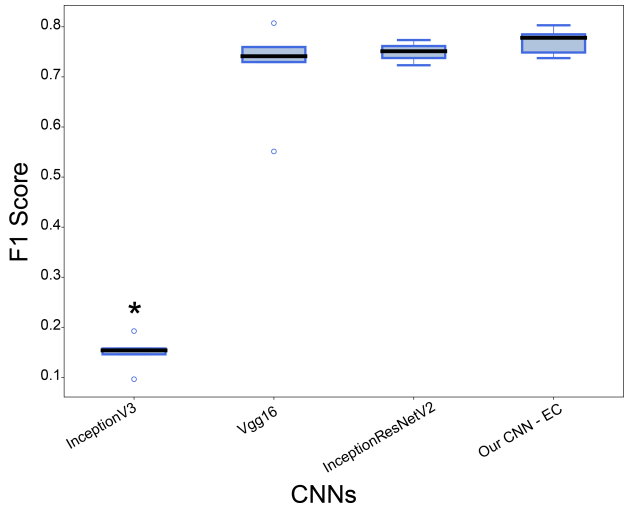


Fig. 6: F1 score of the trained and tested existing CNNs and the CNN we proposed, trained according to experiment C.

developed did not differ with statistical significance from VGG 16 and Inception-ResNet V2 ($p > 0.05$).

IV. Discussion

We designed a light, 9-layered CNN and performed experiments to individuate the most suitable training strategy for this task. The first experiment (**A**) consisted of training and testing the model on the base (i.e., not augmented) version of the dataset. The mean accuracy across folds ($51.54\% \pm 6.48\%$) was above the chance level, and we hypothesized it could increase along with the size of the dataset. Data augmentation is a commonly used approach when datasets are not large enough for CNN to extract generalizable features [40]. Experiment **B** was performed to investigate the effects of offline and on-the-fly data augmentation. The mean classification accuracy across folds increased significantly ($p < 0.05$), showing that the size of the dataset has a prominent role when training this model. Experiment **C** was performed by applying transfer learning and pre-training the CNN on the MNIST dataset before fine-tuning it on our wide-field calcium images. It resulted in a slight improvement with respect to experiment **B**. The difference between **B** and **C** is not statistically significant but, as shown in Fig.4, the standard deviation is much lower in **C**, making it the most robust model among the ones we designed. These results show the importance of the size of the dataset when training such a CNN.

We compared the CNN we designed against 3 commonly used algorithms: Inception V3, VGG 16, and Inception-ResNet V2. The goal was to investigate the performance of models that have become, in recent years, widely used in the biomedical field [19]. On the one hand, these models represent a good benchmark for evaluating the performance of a newly designed CNN. On the other hand, if an existing model performed better, it would be reasonable to deploy it for future experiments. These CNNs were not trained directly on our dataset, but a pre-trained version of them was used. this choice was made to shorten training time and to avoid overfitting problems due to an extremely small dataset and very complex models.

Inception V3 resulted in a weighted F1 score significantly lower than the other CNNs. Its AUC was above 0.80, which is considered satisfactory [44], but the confusion matrix shows a tendency of the CNN to label images as “Cross” and a complete misclassification of “S” pictures. In general, the matrix results in poor diagonality. VGG 16, Inception-ResNet, and our CNN yielded consistent results with each other. All of these achieved a weighted F1 score above 0.70 and very good AUCs above 0.90. In particular, our CNN resulted in an AUC of 0.97. Their confusion matrices are more diagonal and do not show preference toward any class. However, they share a tendency of misclassification between “Circle” and “Square” images, which is reasonable given the very similar cortical responses elicited in the mouse by these two shapes (shown in Fig.3).

These results suggest that, from the classification performance point of view, there is no significant difference between the lightweight model we developed and other computationally more complex models that are commonly used in medical image analysis. While for more complex tasks they can be the gold standard, their use might not be recommended for this application, since they are inherently complicated and require expensive processing. We could dramatically cut on training time, memory requirements, and computing resources by using a lighter design [35]. A lightweight model is the best option when the aim is to promote computational speed and artificial intelligence sustainability [45]. Similar choices have also been made in the literature [26] [27].

Limitations of this method include the difficulties in the skull preparation of the animals, which has to be removed or made transparent. In either case, the operation requires a great deal of care because poor results lead to the opacity of the recorded images [15] [18] [16]. Another challenge is the need for large datasets to properly train the CNNs, and it is strictly related to the availability of the animals and the quality of the obtained images. As previously discussed, the latter can be mitigated by data augmentation and transfer learning, but the original dataset size remains the primary contributor for successful classification of a CNN [40].

V. Conclusion

The goal of this study was to assess the possibility of decoding the V1 response of a transgenic, anesthetized mouse passively looking at simple visual stimuli, by combining wide-field calcium imaging to record V1 activity and deep-learning architectures for classification. The visual stimuli were divided into 10 different shapes and displayed on the screen in random order. For each of them, the V1 evoked activation was recorded via wide-field calcium imaging. We obtained a set of 10 classes of images, that were used to train and test the CNNs. Our findings suggest that, with these techniques, it is possible to univocally associate V1 responses to the visual stimulus that evoked them, and that lightweight CNNs are suitable to perform this classification task. Moreover, the application of data augmentation and transfer learning highly improves the performance of the models, suggesting a high dependence on the dataset size.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2023.3335613

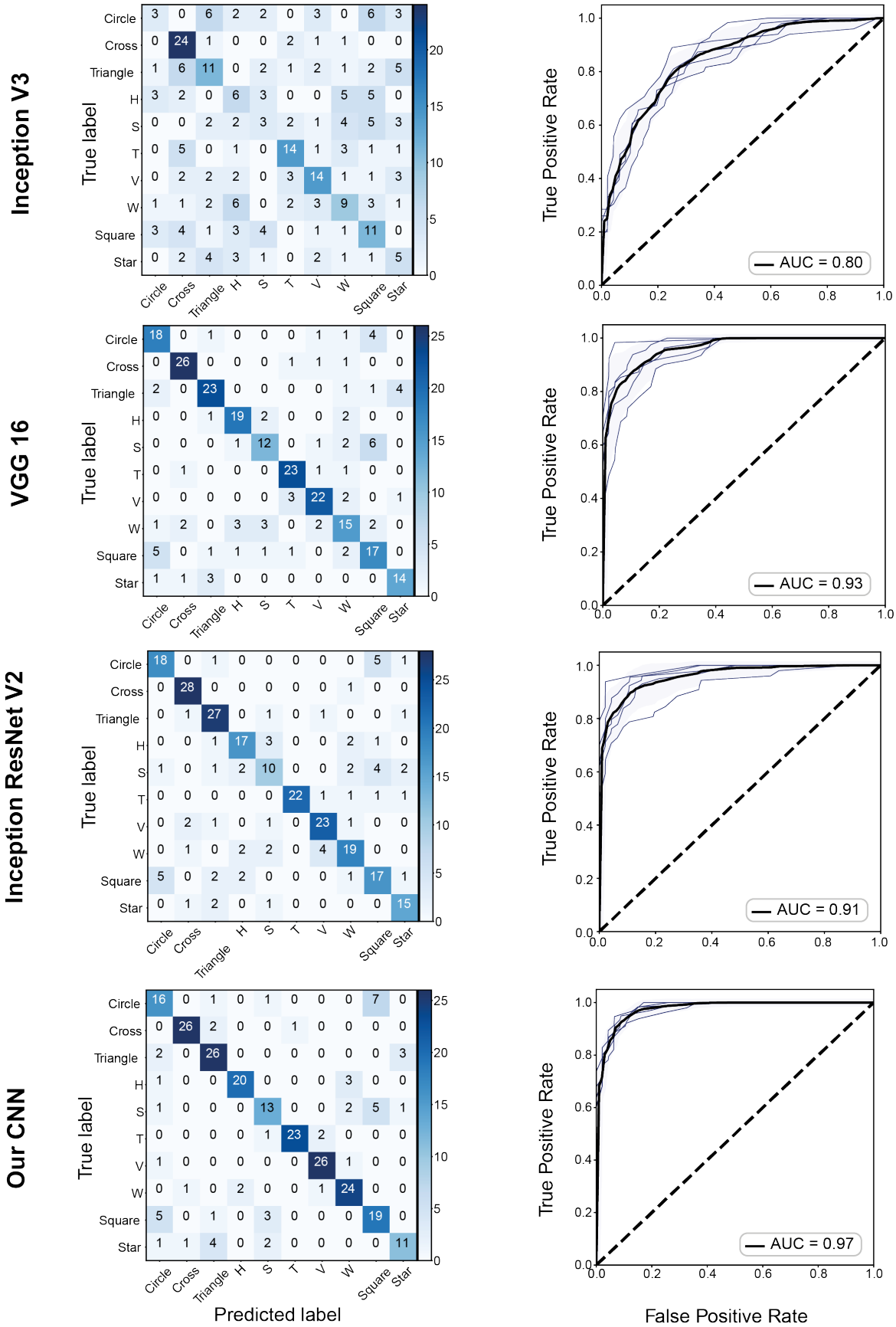


Fig. 7: Confusion matrix and receiver operating characteristic curves of the trained and tested CNNs resulting from the first repetition of the repeated cross-validation.

Future work will be partially devoted to more extensive data collection, increasing both the number of animals involved in the recordings and the variety of visual stimuli.

REFERENCES

[1] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How does the brain solve visual object recognition?" *Neuron*, vol. 73, no. 3, pp. 415–434, 2012.

[2] L. Xu, M. Xu, T.-P. Jung, and D. Ming, "Review of brain encoding and decoding mechanisms for eeg-based brain–computer interface," *Cognitive Neurodynamics*, vol. 15, pp. 569–584, 2021.

[3] Y.-J. Zhang, Z.-F. Yu, J. Liu, T.-J. Huang *et al.*, "Neural decoding of visual information across different neural recording modalities and approaches," *Machine Intelligence Research*, pp. 1–16, 2022.

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

[5] C. M. Niell and M. P. Stryker, "Highly selective receptive fields in mouse visual cortex," *Journal of Neuroscience*, vol. 28, no. 30, pp. 7520–7536, 2008.

[6] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of physiology*, vol. 148, no. 3, p. 574, 1959.

[7] C. Stringer and M. Pachitariu, "Computational processing of neural recordings from calcium imaging data," *Current Opinion in Neurobiology*, vol. 55, pp. 22–31, 2019.

[8] C. Stringer, M. Michaelos, D. Tsyboulski, S. E. Lindo, and M. Pachitariu, "High-precision coding in visual cortex," *Cell*, vol. 184, no. 10, pp. 2767–2778, 2021.

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

[10] M. E. Garrett, I. Nauhaus, J. H. Marshel, and E. M. Callaway, "Topography and areal organization of mouse visual cortex," *Journal of Neuroscience*, vol. 34, no. 37, pp. 12 587–12 600, 2014.

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

[12] B. F. Grewe, D. Langer, H. Kasper, B. M. Kampa, and F. Helmchen, "High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision," *Nature Methods*, vol. 7, no. 5, pp. 399–405, 2010.

[13] M. Pachitariu, C. Stringer, and K. D. Harris, "Robustness of spike deconvolution for neuronal calcium imaging," *Journal of Neuroscience*, vol. 38, no. 37, pp. 7976–7985, 2018.

[14] S. Garasto, A. A. Bharath, and S. R. Schultz, "Visual reconstruction from 2-photon calcium imaging suggests linear readout properties of neurons in mouse primary visual cortex," *bioRxiv*, p. 300392, 2018.

[15] C. Ren and T. Komiyama, "Characterizing cortex-wide dynamics with wide-field calcium imaging," *Journal of Neuroscience*, vol. 41, no. 19, pp. 4160–4168, 2021.

[16] B. L. Sabatini and L. Tian, "Imaging neurotransmitter and neuromodulator dynamics in vivo with genetically encoded indicators," *Neuron*, vol. 108, no. 1, pp. 17–32, 2020.

[17] G. Christine and K. Arthur, "Imaging calcium in neurons," *Neuron*, vol. 73, no. 5, pp. 862–885, 2012.

[18] A. K. Nietz, L. S. Popa, M. L. Streng, R. E. Carter, S. B. Kodandaramaiah, and T. J. Ebner, "Wide-field calcium imaging of neuronal network dynamics in vivo," *Biology*, vol. 11, no. 11, p. 1601, 2022.

[19] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[20] S. P. Singh, L. Wang, S. Gupta, B. Gulyas, and P. Padmanabhan, "Shallow 3d cnn for detecting acute brain hemorrhage from medical imaging sensors," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14 290–14 299, 2020.

[21] M. Li, C. Ma, W. Dang, R. Wang, Y. Liu, and Z. Gao, "Dscnn: Dilated shuffle cnn model for ssvep signal classification," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 12 036–12 043, 2022.

[22] S. Madhavan, R. K. Tripathy, and R. B. Pachori, "Time-frequency domain deep convolutional neural network for the classification of focal and non-focal eeg signals," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3078–3086, 2019.

[23] J. M. Gandarias, A. J. Garcia-Cerezo, and J. M. Gómez-de Gabriel, "Cnn-based methods for object recognition with high-resolution tactile sensors," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6872–6882, 2019.

[24] D. G. León, J. Gröli, S. R. Yeduri, D. Rossier, R. Mosqueron, O. J. Pandey, and L. R. Cenkeramaddi, "Video hand gestures recognition using depth camera and lightweight cnn," *IEEE Sensors Journal*, vol. 22, no. 14, pp. 14 610–14 619, 2022.

[25] M. F. Qureshi, Z. Mushtaq, M. Z. U. Rehman, and E. N. Kamavuako, "E2cnn: An efficient concatenated cnn for classification of surface emg extracted from upper limb," *IEEE Sensors Journal*, vol. 23, no. 8, pp. 8989–8996, 2023.

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

[27] F. Koochaki, F. Shamsi, and L. Najafizadeh, "Detecting mtbi by learning spatio-temporal characteristics of widefield calcium imaging data using deep learning," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 2917–2920.

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] C. Peng, Y. Liu, X. Yuan, and Q. Chen, "Research of image recognition method based on enhanced inception-resnet-v2," *Multimedia Tools and Applications*, vol. 81, no. 24, pp. 34 345–34 365, 2022.

[31] I. E. Aasebø, M. E. Lepperød, M. Stavrinou, S. Nøkkevangen, G. Einevoll, T. Hafting, and M. Fyhn, "Temporal processing in the visual cortex of the awake and anesthetized rat," *Eneuro*, vol. 4, no. 4, 2017.

[32] P. G. Fahey, T. Muhammad, C. Smith, E. Froudarakis, E. Cobos, J. Fu, E. Y. Walker, D. Yatsenko, F. H. Sinz, J. Reimer *et al.*, "A global map of orientation tuning in mouse visual cortex," *BioRXiv*, p. 745323, 2019.

[33] G. Van den Bergh, B. Zhang, L. Arckens, and Y. M. Chino, "Receptive-field properties of v1 and v2 neurons in mice and macaque monkeys," *Journal of Comparative Neurology*, vol. 518, no. 11, pp. 2051–2070, 2010.

[34] D. De Luca, S. Moccia, L. Lupori, R. Mazziotti, T. Pizzorusso, and S. Micera, "Convolutional neural network classifies visual stimuli from cortical response recorded with wide-field imaging in mice," *Journal of Neural Engineering*, vol. 20, no. 2, p. 026031, 2023.

[35] F. Lei, X. Liu, Q. Dai, and B. W.-K. Ling, "Shallow convolutional neural network for image classification," *SN Applied Sciences*, vol. 2, no. 1, pp. 1–8, 2020.

[36] I. Patrini, M. Ruperti, S. Moccia, L. S. Mattos, E. Frontoni, and E. De Momi, "Transfer learning for informative-frame selection in laryngoscopic videos through learned features," *Medical & Biological Engineering & Computing*, vol. 58, pp. 1225–1238, 2020.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[39] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[40] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[41] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G.-R. Xue, Y. Yu, and Q. Yang, "Heterogeneous transfer learning for image classification," in *Twenty-fifth AAAI Conference on Artificial Intelligence*, 2011.

[42] M. Shaha and M. Pawar, "Transfer learning for image classification," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 2018, pp. 656–660.

[43] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[44] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, 2013, vol. 398.

[45] A. van Wynsberghe, "Sustainable ai: Ai for sustainability and the sustainability of ai," *AI and Ethics*, vol. 1, no. 3, pp. 213–218, 2021.