

Characterization of Visual Object Representations in Rat Primary Visual Cortex

Sebastiano Vascon^{1,2,*}[0000–0002–7855–1641], Ylenia Parin^{1,*},
Eis Annavini^{3,*}[0000–0002–3507–4873], Mattia D’Andola³[0000–0003–4554–8822],
Davide Zoccolan³[0000–0001–7221–4188], and
Marcello Pelillo^{1,2}[0000–0001–8992–9243]

¹ DAIS, Ca’ Foscari University of Venice, Via Torino 155, 30170, Mestre (VE)

² ECLT, Ca’ Foscari University of Venice, San Marco 2940, 30124, Venice (VE)

³ Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Via Bonomea 265, Trieste (TS)

Abstract. For most animal species, quick and reliable identification of visual objects is critical for survival. This applies also to rodents, which, in recent years, have become increasingly popular models of visual functions. For this reason in this work we analyzed how various properties of visual objects are represented in rat primary visual cortex (V1). The analysis has been carried out through supervised (classification) and unsupervised (clustering) learning methods. We assessed quantitatively the discrimination capabilities of V1 neurons by demonstrating how photometric properties (luminosity and object position in the scene) can be derived directly from the neuronal responses.

Keywords: rat’s visual system, core object recognition, objects classification

1 Introduction

For most animals, recognition of visual objects is of paramount importance. The visual system of many species has adapted to quickly and effortlessly detect and classify objects in spite of major variation (or transformation) in their appearance. This set of abilities is called Core Object Recognition [2] and is typical of primate species, where a hierarchy of visual cortical areas, known as the ventral stream, supports shape processing and image understanding [3]. In recent years, some authors [21,22] have investigated whether such core ability also exists in rats, by exploiting machine-learning tools, such as information theory and pattern classifiers, which have proved to be invaluable tools to understand how object vision works in primates [2]. Indeed, rodents have become increasingly interesting model organisms to study the mammalian visual system [23,8,9,10]. In particular, rodent object-processing abilities are supposed to be located along a progression of cortical areas, starting in primary visual cortex (V1), and extending to lateral extrastriate areas LM, LI, and LL [8,20], which are thought to

* equal contribution

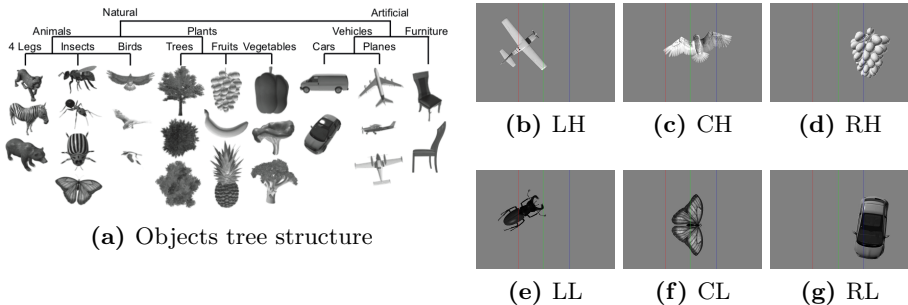


Fig. 1: Stimulus set. The image on the left shows the semantic hierarchical structure of the objects belonging to the stimulus set. The images on the right show an example stimulus for each combination of position (first letter of the label) and luminosity (the second letter), e.g., “LH” stands for left and high luminosity, “RL” stands for right and low luminosity, “CL” stands for center and low luminosity, etc.

be an homologous of the primate ventral stream. Recent work by [21] has shown that, indeed, visual object representations along this progression become more explicit, i.e.: 1) information about low-level visual properties, such as luminance, is gradually lost; and 2) object identity becomes more easily readable through linear classifiers, even in the presence of changes in object appearance.

In this work, we tried to understand at a deeper level what visual properties are encoded in the activity of a population of rat V1 neurons, using both unsupervised and supervised machine learning algorithms. The focus on V1 was motivated by the fact that this cortical area is the entry stage of visual information in cortex (future work will aim at providing a similar characterization in higher-order visual cortical areas).

Specifically, in this study, we applied for the first time the Dominant Set[16] clustering algorithm (DS) to understand the structure of visual object representations in a visual cortical area. The choice of the DS has been driven also by its recent success in related fields, like in brain connectomic [5,6] or neuroscience [17], making it a good candidate for the task at hand. Furthermore, we applied an array of supervised algorithms to show that V1 neuronal responses can be used to predict with great accuracy the photometric information on the scene presented to the rat.

The article is organized as follows: in section 2 we provide a description of the experimental methods that were used to produce the stimulus set and to record the responses of V1 neurons; in section 3 we describe the analysis we carried out to understand the organization of visual stimuli in terms of V1 neuron responses; in section 4 we show how V1 neuronal responses can be used to classify some key visual properties of the stimuli, i.e., their location within the visual field and their luminosity; the section 5 concludes the paper with some future perspectives.

2 Materials & Methods

In this section, we describe the steps that were performed to build the dataset and, whether non-conventional, the methodologies used to analyze the data.

2.1 Stimulus set and data acquisition

For our experiments, we built a rich and ecological stimulus set using a large number of objects, organized in a semantic hierarchy (Figure 1). To build the stimulus set we used 40 3D models of real world objects⁴, both natural and artificial, each rendered in 36 different poses, randomly chosen around four main views (frontal, lateral, top, and 45° in azimuth and elevation), at one of three possible sizes (30-35-40°) chosen at random, in one of three possible positions (0°, $\pm 15^\circ$), also chosen at random, and rotated in plane of either 0, 90 or $\pm 45^\circ$ for a total of 1440 stimuli. To further characterize the stimulus set we extracted a set of low and mid level features (such as position, contrast, and orientation) of the stimuli as they were presented on-screen to the rat: for the scope of the current work we will focus on the *position* of the center of mass, and on the *luminosity*. Stimuli were presented on a gray background to anesthetized naïve Long-Evans rats for 150 ms while collecting extracellular neuronal activity from all the layers of primary visual cortex (V1) using multi-shank, 64-channel silicon electrode arrays⁵. We recorded extracellular potentials using an RZ2 BioAmp signal processor⁶ at a sampling frequency of 24.4141 kHz. We characterized the neurons by carefully mapping the positions of each unit’s receptive field (RF), rotating the rat afterwards in order to center the RFs on the screen and thus achieve maximal response to the stimuli.

2.2 Data preprocessing

We filtered the raw extracellular potentials with a band-pass filter (0.5–11 kHz) to extract neurons’ spiking activity, and the resulting action potentials (spikes) were extracted using an Expectation-Maximization clustering algorithm [18] that separates the spikes produced by different neurons according to their shape. Then, we estimated the optimal spike count window for each neuron using its firing rate averaged over the 10 best stimuli [1,21], and we used it to compute the average number of spikes produced by a neuron in response to each stimulus, across its repeated presentations. Finally, we scaled the spike counts of each neuron to zero mean and unitary variance to obtain the population vectors for the stimulus set [12]. This led to a vector of size 177 for each visual stimulus that was used in all unsupervised and supervised analysis, where 177 is the total number of units (single- and multi-unit) obtained through spike sorting.

⁴ TurboSquid <https://www.turbosquid.com/>

⁵ NeuroNexus Technologies, Ann Arbor, MI, USA

⁶ Tucker-Davis Technologies, Alachua, FL, USA

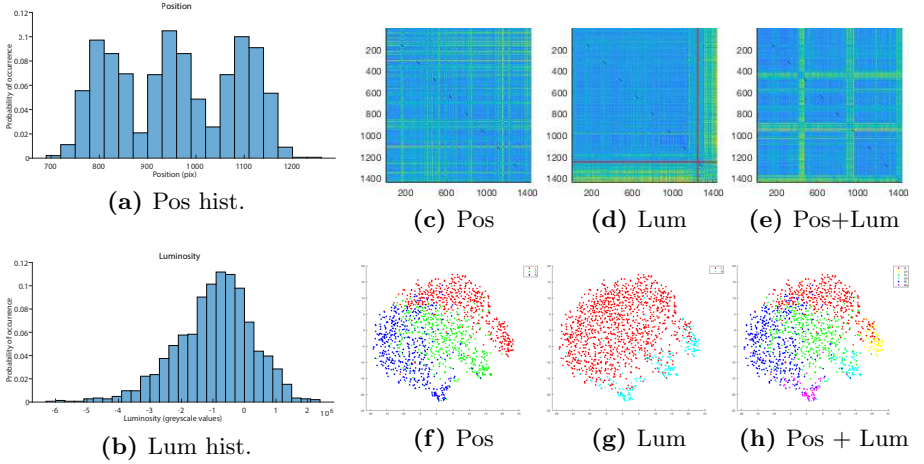


Fig. 2: Feature distribution and binning visualization. Figures 2a- 2b report position and luminosity histograms. Figures 2c- 2d show distance matrices of neuronal population vectors ordered by position and luminosity. Figure 2e shows the same distance matrix in which rows and columns are ordered according to the binning of the combined feature. Figures 2f- 2h report tSNE [13] 2-dimensional maps where point color refers to the feature binning.

2.3 Characterization of visual features

To quantify the low- and mid-level features of the stimuli, we saved them as they were presented during the experiments and we extracted the position of the center of mass of each stimulus and the total luminosity w.r.t. the background, defined as $L_{tot} = \sum(I(i, j) - 128)$, where $I(i, j)$ is the matrix of pixel intensities in greyscale values. As shown in Figure 2a, the distribution of the position of the stimuli along the x axis was, as expected from the presentation protocol, trimodal, with the three peaks corresponding to the three main visual field positions used to show the stimuli during the experiment. This naturally leads to partition the set of stimuli in three classes, according to their position: left, right, or center. The luminosity instead shows a unimodal distribution that does not suggest a clear categorization; for this reason we visually inspected the distance matrix of the neuronal population vectors corresponding to each object, ordered according to the luminosity of the objects (see figure 2d). We then set a threshold (red lines in figure 2d) by hand at the point where the stimuli clearly separate. The final distribution of samples (objects) per class is reported in Table 1.

3 Characterize object representations through clustering

First, we analyzed the space of neuronal responses obtained from V1 through unsupervised (clustering) methods. The aim was to assess how performing is the

	Luminosity		Tot. Pos.
	Low lum.	High lum.	
Right	411	69	480
Pos. Center	406	74	480
Left	407	73	480
Tot. Lum.	1224	216	1440

Table 1: Classes distributions.

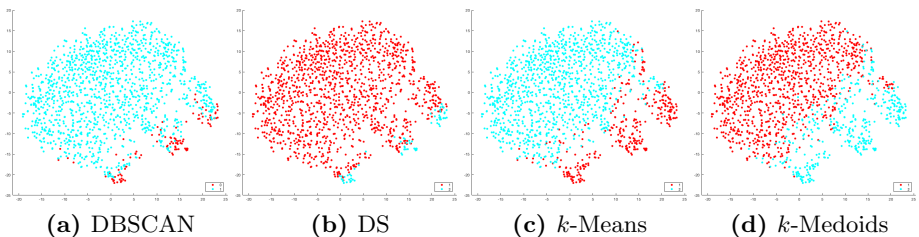
Algorithm	K	SIL
DBSCAN eps=14.9 minPts=4	2	0.525
DS $\sigma = 504$	2	0.565
K-Means k=2	2	0.448
K-Medoids k=2	2	0.394

Table 2: Best Average Silhouette value for each compared clustering method and number of clusters extracted.

rat's neuronal embedding of the visual stimuli in terms of automatic grouping. The rationale is that stimuli having similar photometric characteristics (position and/or luminosity) should lie close to each other in the embedding, while being well separated from those having different properties. To check whether the neuronal mapping was meaningful in this regard, we tested different clustering algorithms, considering their best parameter setting under both internal and external indexes. The performances of each method have been stressed to their limit, so as to provide a guideline for future studies relying on similar methods. To have an intuition of the complexity of this task we first visually inspected the distance matrices of all the stimuli per classes (see figure 2c-2e). The distance matrix is a symmetric matrix of size $n \times n$ (where $n = 1440$) containing the Euclidean distances of all the neuronal population vectors for each pair of stimuli. Then the matrix is sorted accordingly to the visual feature (position, luminosity and the binned position+luminosity) under exam. As one can note, the matrix relative to the class position resembles a random matrix, and the three classes (left, center, right) are not clearly identifiable. Instead, the other two matrices, related to luminosity and the combination between position and luminosity, clearly report the two and six classes in which features are binned.

3.1 Experiments

The experiments that have been carried out compared the performances of supervised clustering algorithms like k -means [14] and k -medoids[11] (where the number of cluster is known in advance) and unsupervised techniques like *DBSCAN*[7] and *Dominant Set*[16], where no a-priori information on the underlying structures is available. We performed two different experiments considering on a

**Fig. 3:** Best partition according to SIL. Each point of the map is colored according to the cluster it belongs. Clusters are obtained with different methods through Silhouette maximization.

Alg.	Position				Luminosity				Position & Luminosity			
	K	AMI	P	ARI	K	AMI	P	ARI	K	AMI	P	ARI
DBSCAN	2	0.008	0.388	0.012	2	0.43	0.924	0.633	2	0.128	0.31	0.136
DS	17	0.179	0.746	0.121	2	0.475	0.933	0.675	17	0.268	0.672	0.171
Kmeans	6	0.255	0.692	0.233	2	0.336	0.89	0.534	6	0.378	0.638	0.299
Kmedoids	3	0.189	0.557	0.179	2	0.257	0.85	0.415	6	0.313	0.543	0.21

Table 3: Best ARI value for each compared clustering algorithm and stimulus classes. The best clustering algorithm for each feature and external index is highlighted in bold.

first instance an internal criterion, the *Silhouette* [19], and later an external one, the *Adjusted Rand Index* (ARI). The silhouette is a specific measure for each object and accounts for how well each object lies within its cluster and how well is separated from the others. In order to provide a global measure, the *overall average silhouette width* (SIL) is taken [19]. The ARI is a measure that accounts for the agreement of two partitions, the predicted from a clustering method and the annotation: the higher its value, the better the algorithm has separated data. For all the clustering methods, the Euclidean metric has been used to compute distances/similarities.

Internal criterion For each method we searched in its parameters space the setting that maximizes the SIL. Maximizing the average silhouette means finding the parameters of a partitioning algorithm that separate and merge points in the best way possible provided their similarities or dissimilarities. In case a clustering method collapsed to an unwanted solution (one single cluster), we looked for the highest SIL value which separates the objects into at least two clusters with minimum density equal to the size of the less represented class of objects (in our case 69, see Table 1 for the classes distributions). We amend that this particular selection criterion is not completely fair, because it mixes some prior-information on the structure of the data with an internal index, but has the positive effect to find a reasonable solution for all the clustering algorithms at hand. The quantitative and qualitative results are reported in Table 2 and in Figure 3 respectively.

External criterion The test on the external criteria has been performed following the same schema as in the experiment on the internal measures, but instead of maximizing the SIL we looked for the parameters that maximize the ARI for each class of stimuli (see Table 1). Other external indexes that are computed are the Adjusted Mutual Information (AMI) and the Purity (P). The AMI is similar to the ARI but quantifies the commonalities between two partitioning from an information-theoretic perspective. The purity index takes into account how the labels are organized inside of each cluster. We performed these experiments to understand which method has the potential of grouping as expected the different neuronal mappings with respect to the single classes. The quantitative and qualitative results are reported in Table 3 and in Figure 4 respectively.

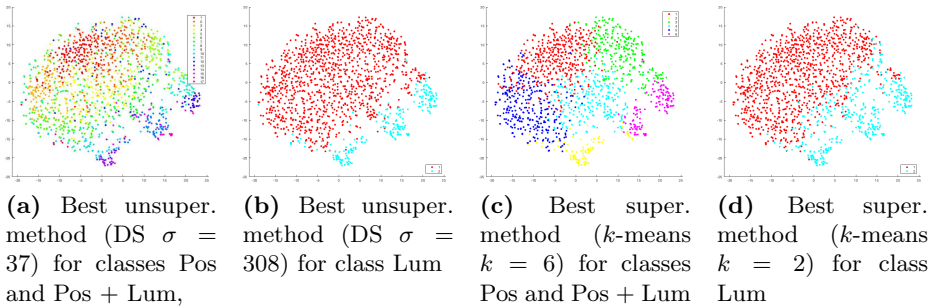


Fig. 4: Best partition according to ARI. For each stimulus feature we selected the best partition produced by supervised (k -means and k -medoids) and unsupervised method (DS and DBSCAN) by maximizing ARI metric. Each point of the map is colored according to the cluster it belongs to.

3.2 Results

In general the values of SIL, ARI and AMI are quite low indicating that the task is very complex, data is very noisy and visual stimuli are not perfectly mapped to the neuronal response. This can be easily seen in Figures 2f-2h, where the tSNE [13] projection shows a mixing of classes, in particular on the upper part of the projection for the position (Figure 2f) and position+luminosity (see Figure 2h). The luminosity classes, instead, are well separated from each other. From both perspectives (internal and external indexes) the embedding found using the responses from V1 was sufficiently able to group the different classes.

Considering the internal index (see Table 2), the method that best performs was the DS, outperforming also the supervised methods like k -means and k -medoids. The motivations are due to the fact that DS only depends on the similarity matrix, which is not the case for the other techniques that also rely on other assumptions (like the number of clusters or global densities). Regarding the results for the external criterion (see Table 3) the best unsupervised clustering method is the DS among all features. The top purity is reached by the DS and this can be explained by the higher number of clusters that are generated. In terms of supervised clustering, the best performing method is k -means in all the considered metrics. These results suggest us that, in case of absent a-priori information on the number of clusters, the DS method can be considered as a more-than-valid alternative to standard approaches (like DBSCAN). Furthermore, knowledge on the number of clusters can be fruitfully used by supervised clustering algorithms (like k -means).

4 Inferring object properties with supervised learning

As seen in section 3, the analysis of the neuronal embedding was meaningful under different criteria to analyze how the space is partitioned. This lead to

Alg	Position				Luminosity				Position + Luminosity			
	ACC	AUC	mF1	MF1	ACC	AUC	mF1	MF1	ACC	AUC	mF1	MF1
Lin SVM	0.906	0.961	0.859	0.859	0.913	0.922	0.913	0.818	0.756	0.500	0.268	0.181
Ker SVM	0.942	0.979	0.912	0.912	0.939	0.970	0.939	0.877	0.758	0.536	0.275	0.181
ECOC LSVM	0.908	0.963	0.862	0.862	0.912	0.922	0.912	0.813	0.937	0.966	0.810	0.778
k-NN ($k = 9$)	0.913	0.960	0.870	0.870	0.931	0.937	0.931	0.855	0.935	0.944	0.805	0.763

Table 4: Classifier performances.

a second set of experiments in terms of discrimination power of the features extracted from the V1 area. We considered separately the three classes of photometric characteristics *position*, *luminosity* and *position+luminosity* and carried out several tests by training and testing standard classifiers (Linear/Kernel SVM⁷, Error Correcting Output Code Linear SVM⁸ [4] and k-NN) on the V1 embedding to confirm its discrimination capability. The rationale is that similar visual stimuli will lie in close proximity and vice-versa different ones will be located far away. With this assumption, a classifier should be able to find a boundary to discriminate between the classes.

4.1 Experiments & Results

Considering a class of visual stimuli (see Table 1 for a details on classes) we performed a 5-fold cross validation to find the best parameter setting of each classifier. The folds have been created in a stratified way ensuring that each class is represented with the same proportion of the dataset. The training and testing have been performed randomly generating 10 different splits of the data and consequently averaging the performances. To evaluate the performances we used four indexes that are common in classification tasks: *accuracy* (ACC), *average area under the ROC curve* (AUC), the *micro F-measure* (mF1) and the *macro F-measure* (MF1) [15]. The results are reported in Table 4. It is evident how the neuronal responses can be used successfully to classify visual stimuli; in fact, we achieved a very high ACC, AUC and (in general) F-score. As expected, due to the class imbalance (see Table 1), the MF1 is a bit lower than the mF1. This is particularly evident for the luminosity and position+luminosity classes, in which a strong imbalance (the larger class is $\simeq 6$ times bigger w.r.t. the smaller one) is reported. Regarding the class position and the class luminosity, the best performing method was the Kernel SVM followed by the k -NN. It is worth to note that the simple k -NN is the second best choice for all the three classes; this gives us an indication of the difficulties in finding a linear separator, hence on the non linear separability of the space. This motivates also the fact that the Linear SVM performs poorly w.r.t. the Kernel SVM. Furthermore, in the case of position+luminosity the best performing method was the ECOC L-SVM followed by the k -NN. This is explained by the fact that, in that particular case, we increased the number of classes from 2-3 to 6, needing more hyperplanes to separate them. The ECOC is based on an ensemble of Linear SVMs trained in

⁷ Software at <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

⁸ Software at <https://www.mathworks.com/help/stats/fitcecoc.html>

one-vs-one mode which creates all the possible intersecting hyperplanes w.r.t. the classes. For this reason it outperforms the other classifiers, while being not so far from the performances of the Linear SVM for the Position and for the Luminosity, both cases having fewer classes. Concerning the stability of the results we reported a maximum mean standard deviation of $\simeq 1\%$ considering all the 10 runs.

5 Conclusions

In this paper, we investigated how visual stimuli are mapped into the representational space of V1 neurons focusing on two low-level properties (luminosity and position within the visual field). We thus quantified the extent to which these properties were accurately represented in the V1 population space, using supervised and unsupervised learning methods. We found that, indeed, both luminosity and position and their combination are naturally mapped in the V1 representation, and that these features can be accurately extracted using pattern classifiers. Among the clustering methods, DS showed the greatest accuracy at inferring the structure of the representation. Among the classifiers, the SVM with nonlinear kernel achieved the highest accuracy. In both cases, this testifies of the complexity of the representation and of the not complete linear discriminability of the data.

As future work, we will try different distance functions and will test whether other higher-level visual features, e.g. orientation, are encoded. Moreover, the same data processing pipeline will be applied to higher-order visual areas, e.g. LM-LI-LL, to understand the differences with V1.

Acknowledge

This work was supported by a European Research Council Consolidator Grant (DZ, project n. 616803-LEARN2SEE).

References

1. Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J.J., Zecchina, R., Zoccolan, D.: Shape Similarity, Better than Semantic Membership, Accounts for the Structure of Visual Object Representations in a Population of Monkey Inferotemporal Neurons. *PLOS Computational Biology* **9**(8), 1–21 (2013)
2. DiCarlo, J.J., Cox, D.D.: Untangling invariant object recognition. *Trends in Cognitive Sciences* **11**(8), 333–341 (2007)
3. DiCarlo, J.J., Zoccolan, D., Rust, N.C.: How does the brain solve visual object recognition? *Neuron* **73**(3), 415–434 (2012)
4. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research* **2**, 263–286 (1994)
5. Dodero, L., Vascon, S., Giancardo, L., Gozzi, A., Sona, D., Murino, V.: Automatic white matter fiber clustering using dominant sets. In: 2013 International Workshop on Pattern Recognition in Neuroimaging. pp. 216–219 (June 2013)

6. Doderio, L., Vascon, S., Murino, V., Bifone, A., Gozzi, A., Sona, D.: Automated multi-subject fiber clustering of mouse brain using dominant sets. *Frontiers in Neuroinformatics* **8**, 87 (2015)
7. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. pp. 226–231. KDD'96, AAAI Press (1996)
8. Glickfeld, L.L., Olsen, S.R.: Higher-Order Areas of the Mouse Visual Cortex. *Annual Review of Vision Science* **3**(1), 251–273 (2017)
9. Glickfeld, L.L., Reid, R.C., Andermann, M.L.: A mouse model of higher visual cortical function. *Current opinion in neurobiology* **24**, 28–33 (2014)
10. Huberman, A.D., Niell, C.M.: What can mice tell us about how vision works? *Trends in neurosciences* **34**(9), 464–473 (2011)
11. Kaufman, L., Rousseeuw, P.: Clustering by means of medoids. In: *Statistical Data Analysis Based on the L1 Norm and Related Methods*, pp. 405–416. North-Holland; Amsterdam (1987)
12. Kiani, R., Esteky, H., Mirpour, K., Tanaka, K.: Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology* **97**(6), 4296–4309 (2007)
13. van der Maaten, L., Hinton, G.: Visualizing Data using t-SNE . *Journal of Machine Learning Research* **9**, 2579–2605 (2008)
14. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. pp. 281–297. University of California Press, Berkeley, Calif. (1967)
15. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA (2008)
16. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. *IEEE transactions on pattern analysis and machine intelligence* **29**(1), 167–172 (2007)
17. Pennacchietti, F., Vascon, S., Nieuws, T., Rosillo, C., Das, S., Tyagarajan, S., Diaspro, A., del Bue, A., Maria Petrini, E., Barberis, A., Cella Zanacchi, F.: Nanoscale molecular reorganization of the inhibitory postsynaptic density is a determinant of gabaergic synaptic potentiation. *Journal of Neuroscience* (2017)
18. Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., Tolias, A.S., Solomon, S., Buzsáki, G., Carandini, M., Harris, K.D.: Spike sorting for large, dense electrode arrays. *Nature neuroscience* **19**(4), 634–641 (2016)
19. Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53 – 65 (1987)
20. Sereno, M.I., Allman, J.: Cortical visual areas in mammals. The neural basis of visual function **4**, 160–172 (1991)
21. Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F.B., Vanzella, W., Riggi, M., Buffolo, F., Panzeri, S., Zoccolan, D.: Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. *eLife* **6**, 1–39 (2017)
22. Vermaercke, B., Gerich, F.J., Ytebrouck, E., Arckens, L., Op de Beeck, H.P., Van den Bergh, G.: Functional specialization in rat occipital and temporal visual cortex. *Journal of neurophysiology* **112**(8), 1963–1983 (2014)
23. Zoccolan, D.: Invariant visual object recognition and shape processing in rats. *Behavioural brain research* **285**, 10–33 (2015)