A systematic evaluation of neural morphology representations for cell type discrimination

Sophie Laturnus[1], Dmitry Kobak[1], and Philipp Berens[1,2,3]

[1] *Institute for Ophthalmic Research, University of Tübingen, Germany*
[2] *Institute for Biomedical Informatics, University of Tübingen, Germany*
[3] *Bernstein Center for Computational Neuroscience, University of Tübingen, Germany*
name.surname@uni-tuebingen.de

March 19, 2019

Abstract

The anatomy of a neuron has always been a defining feature of its cell type and quantitative analysis of neuronal morphologies usually begins with choosing a particular feature representation in order to make individual morphologies amenable to standard statistical tools. Many different feature representations have been suggested in the literature, ranging from density maps to intersection profiles, but they have never been compared side by side. Here we performed a systematic comparison of various representations, measuring how well they were able to capture the difference between known morphological cell types. For our benchmarking, we used several curated data sets consisting of mouse retinal bipolar cells and cortical inhibitory neurons. We found that the best performing feature representations were two-dimensional density maps and a small set of morphometric statistics which continued to perform well even when neurons were only partially traced. Combining both representations increased their performance further, suggesting that they captured some complementary information about the cell anatomy. The same combined representation performed well in an unsupervised setting, implying that it can be suitable for dimensionality reduction or clustering. Our study provides a thorough quantitative benchmark of morphology representations which forms the basis for the design and evaluation of morphological descriptors in future studies.

1 Introduction

The development of experimental methods for high-throughput single cell RNA sequencing (Zeisel et al., 2018; Saunders et al., 2018; Tasic et al., 2018; Cao et al., 2019) and large-scale functional imaging (Baden et al., 2016; Pachitariu et al., 2017; Schultz et al., 2017) has led to a surge of interest in identifying the building blocks of the brain – the neural cell types (Zeng and Sanes, 2017; Chen et al., 2018). Both data modalities are analyzed with specialized quantitative tools (Stegle et al., 2015; Stringer and Pachitariu, 2019) and produce data sets amenable to statistical analysis such as cell type identification by clustering.

At the same time, ever since the work of Santiago Ramón y Cajal (Ramón y Cajal, 1899), it was the anatomy of a neuron that has been considered the defining feature of a neural cell type. Like in genetics and physiology, recent years have seen a tremendous increase in the availability of anatomical data sets, due to advances in light and electron microscopy (Briggman et al., 2011; Helmstaedter et al., 2013; Economo et al., 2016) and associated tools for increasingly automated reconstruction (Peng et al., 2010, 2014; Bria et al., 2016). As a consequence, more and more full reconstructions of neurons are becoming available in public data bases, such as the Allen cell type atlas (http://celltypes.brain-map.org) or the NeuroMorpho database (http://neuromorpho.org). Anatomical analysis of neural cell types based on these reconstructions, however, requires accurate quantitative representations of the neuron morphologies. While many different representations have been developed in the literature, they have rarely been systematically compared with regard to their ability to discriminate different cell types. Two prominent examples of such representations are density maps (Jefferis et al., 2007) and morphometric

statistics (Uylings and van Pelt, 2002; Scorcioni et al., 2008; Polavaram et al., 2014; Lu et al., 2013, 2015), representing two ends of the spectrum: density maps ignore all fine details of a morphology, simply measuring the density of neurites; morphometric statistics, in turn, quantify the complex branching of axons and dendrites in a small set of single-valued summary statistics. Other spatial analyses such as Sholl intersection profiles (Sholl, 1953) can be seen as occupying an intermediate position on this spectrum. In addition, several novel feature representations based on graph theory and topology have been suggested in recent years (Heumann and Wittum, 2009; Gillette and Grefenstette, 2009; Gillette et al., 2015; Li et al., 2017; Kanari et al., 2018).

Here we benchmarked different representations of neural morphologies as to how well they were able to capture the difference between known morphological cell types. To this end, we used carefully curated anatomical data from two studies, encompassing over 400 retinal and cortical interneurons with complete axonal and dendritic reconstructions and expert annotated cell type labels (Helmstaedter et al., 2013; Jiang et al., 2015). In order to have a well-defined performance measure, we used a supervised learning framework: given the expert labels, we asked which morphological representations were most suitable for cell type discrimination using logistic regression. By combining different representations together, we also studied to what extent they capture complementary information about cell morphologies. In addition, we investigated how robust these representations are if only parts of a neuron are reconstructed.

2 Results

We analyzed the discriminability between different morphological cell types in adult mouse retina and adult mouse visual cortex (Figure 1). The retinal data set consisted of $n = 221$ retinal bipolar cells semi-automatically reconstructed from electron microscopy scans and sorted into 13 distinct cell types (Helmstaedter et al., 2013; Behrens et al., 2016) (see Figure 1A–B). In this study we only used the 11 cell types that included more than 5 neurons (remaining sample size $n = 212$). The cortical data consisted of inhibitory interneurons from primary visual cortex manually reconstructed from biocytin stainings (Jiang et al., 2015). We separately analyzed neurons from layer 2/3 (V1 L2/3 data set, $n = 108$ neurons in 7 classes, Figure 1C–D) and from layer 5 (V1 L5 data set, $n = 93$ neurons in 6 classes, Figure 1E–F). All three data sets comprised accurate and complete morphological reconstructions of dendrites and axons and included cell types that are morphologically close enough to pose a challenge for classification (see Discussion).

We used 54 feature representations that we grouped into four different categories: density maps, morphometric statistics, morphometric distributions, and persistence images (Figure 2). Each feature representation was computed using only axons, only dendrites, and using the full neuron (i.e. axons and dendrites together).

Density maps are one- or two-dimensional projections of the neural morphology. We used projections onto the x, y, and z axes as well as onto the xy, xz, and yz planes. Figure 2A shows the XZ density maps for two exemplary bipolar cells, one of type 1 and one of type 5O. Figure 2B shows Z density maps of all cells of these two types. This particular pair of cell types can be easily discriminated based on the Z projection alone.

We used 22 single-valued summary statistics of each neuron, such as width, height, total neurite length, number of tips, number of branch points, etc., many of which were different for the bipolar types 1 and 5O (Figure 2C). We also considered a feature representation that joins all of them into a 22-dimensional *morphometric statistics* vector.

We used 23 morphometric distributions of which 17 were one-dimensional and six were two-dimensional. As an example, the Sholl intersection profile (Sholl, 1953) describes the number of intersection of a 2D projection with concentric circles of different radius, and is very different for bipolar types 1 and 5O. (Figure 2D). An example of a two-dimensional distribution is the distribution of "path angle" (turning angle) vs. "path distance" (distance to soma along the neurite path) across all nodes in the traced morphology (Figure 2E). After binning, this becomes a 400-dimensional feature vector; Figure 2F shows two principal components (PCs) across all bipolar cells of type 1 and 5O, indicating that PC1 discriminates the types very well.

Lastly, we used one- and two-dimensional *persistence images*, recently introduced quantifications of neural morphology based on topological ideas (Li et al., 2017; Kanari et al., 2018, 2019). The same two

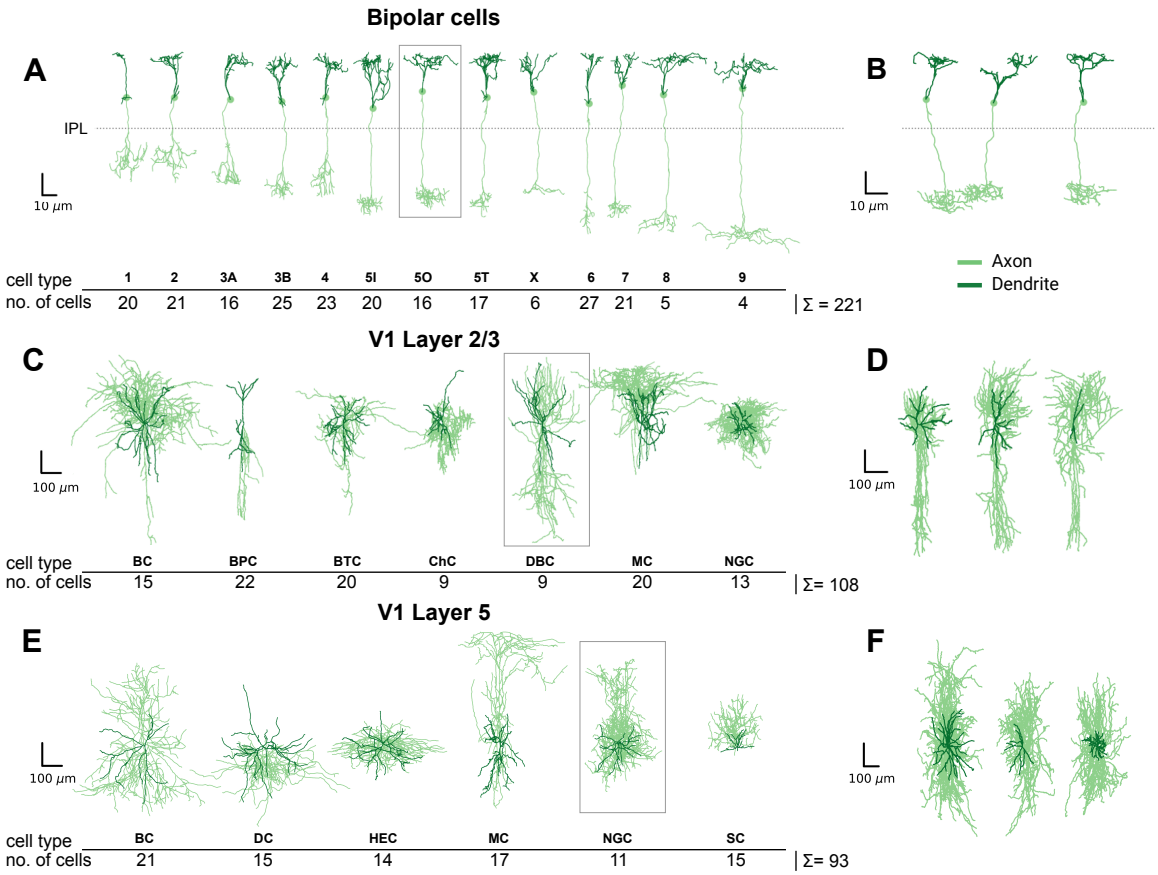

Figure 1: Exemplary cells of each cell type for all three data sets. Axons are shown in light green, dendrites in dark green. **(A)** Mouse retinal bipolar cells from Helmstaedter et al. (2013). The soma is indicated as a sphere. Dashed lines shows the onset of the inner plexiform layer (IPL). The cell types used for analysis are types 1, 2, 3A, 3B, 4, 5I, 5O, 5T, X, 6, and 7. Cell types 8 and 9 were excluded from further analysis due to insufficient sample sizes. **(B)** Three further examples of the type 5O. **(C)** Layer 2/3 inhibitory interneurons in primary visual cortex of adult mouse (Jiang et al., 2015). BC: basket cells, BPC: bipolar cells, BTC: bitufted cells, ChC: chandelier cells, DBC: double bouquet cells, MC: Martinotti cells, NGC: neurogliaform cells. **(D)** Three further examples of double bouquet cells. **(E)** Layer 5 inhibitory interneurons in primary visual cortex of adult mouse (Jiang et al., 2015). BC: basket cells, DC: deep-projecting cells, HEC: horizontally elongated cells, MC: Martinotti cells, NGC: neurogliaform cells, SC: shrub cells. **(F)** Three further examples of neurogliaform cells.

bipolar cell types can be well discriminated based on the first two PCs of the two-dimensional persistence images (Figure 2G–H).

We give a complete list of the investigated feature representations in the Methods section.

2.1 Predictive performance of feature representations

For each feature representation and for each pair of morphological types in a given data set, we built a binary classifier and assessed its performance using cross-validation. As a classifier, we used logistic regression regularized with elastic net penalty and PCA pre-processing. Nested cross-validation was used to tune the regularization strength and obtain an unbiased estimate of the performance (see Methods and Figure 3).

As an example, Figure 4 shows the performance of one particular feature representation (morphometric statistics of the full neuron) for all 55 pairs of neural types in the bipolar data set, 21 pairs in the V1 L2/3 data set, and 15 pairs in the V1 L5 data set. We used the cross-validated log-loss as the main measure of performance, because it is a proper scoring rule used by logistic regression and is unaffected

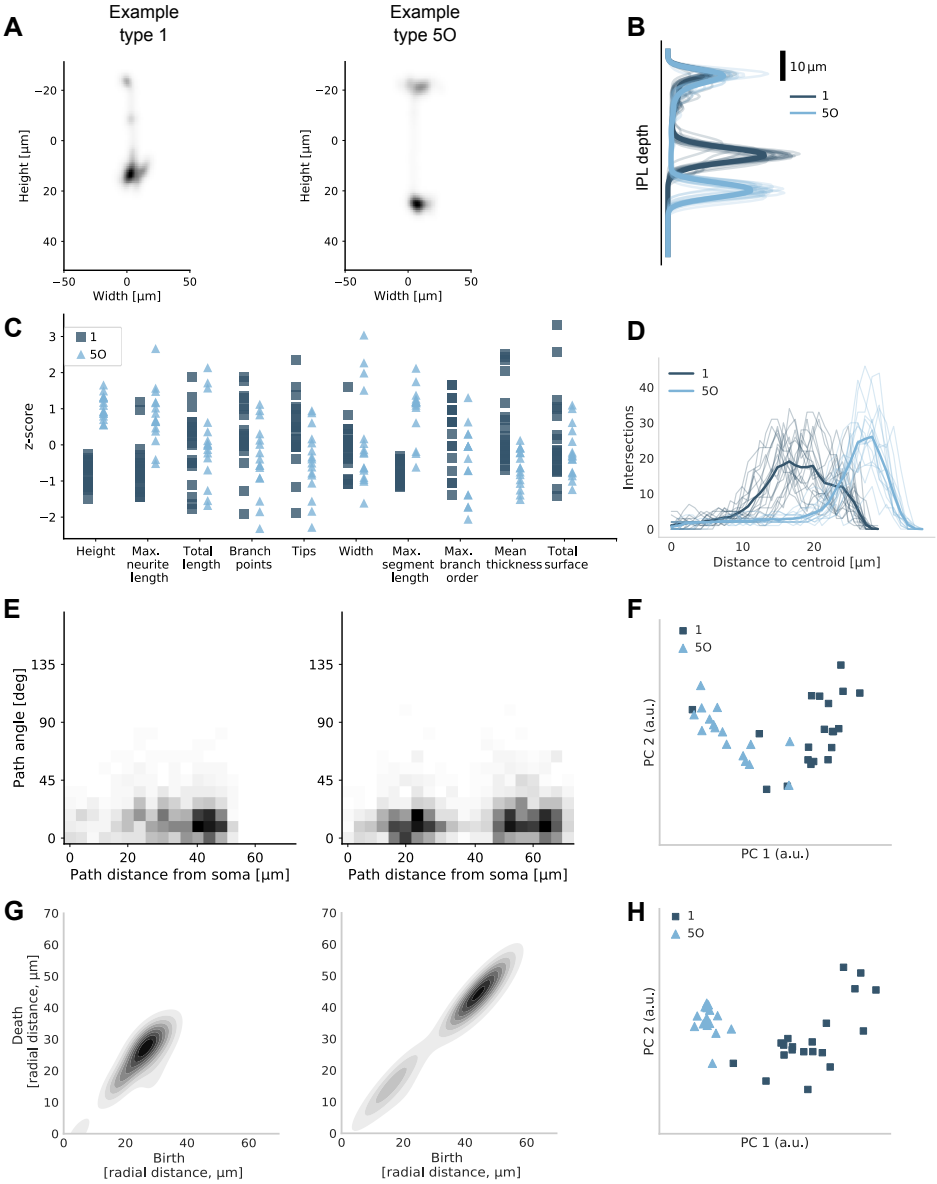

Figure 2: Selected feature representations for retinal bipolar cells of type 1 and type 5O. **(A)** Smoothed density map of XZ projection for two exemplary cells. **(B)** Smoothed density map of Z projection for all cells of these two types. The cells of type 5O stratify deeper in the inner plexiform layer (IPL) than cells of type 1. Bold lines show class means. **(C)** A selection of ten single-valued summary statistics that were included in the *morphometric statistics* vector. **(D)** Sholl intersection profile of the YZ projection for all cells of these two types. Bold lines show class means. **(E)** Two-dimensional distribution of path angles and path distances to the soma across all nodes for the same two exemplary cells shown in (A). **(F)** The first and the second principal components (PCs) of path-angle/path-distance histograms for all cells of these two types. **(G)** Two-dimensional persistence images for the same two exemplary cells shown in (A) and (E). **(H)** The first and the second PCs of 2D persistence images for all cells of these two types.

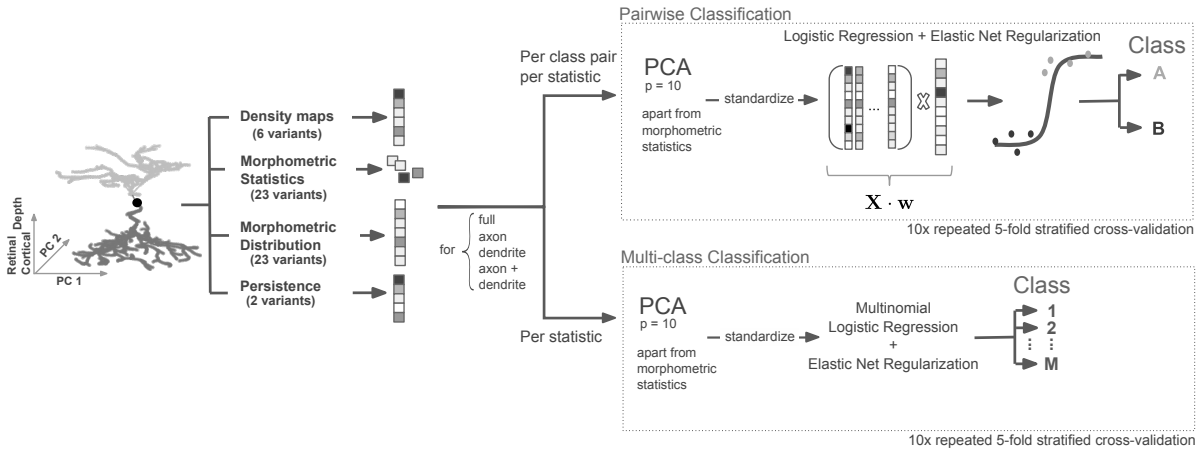

Figure 3: Processing pipeline. Inhibitory interneurons were soma-centered. Retinal bipolar cells were soma-centered in x and y while $z = 0$ was chosen to correspond to the the inner plexiform layer (IPL) onset. The z direction of each cell was aligned with cortical/retinal depth, and the x direction – with the widest spread orthogonal to z. Several different feature representations were extracted automatically and used for pairwise and multi-class classifications. For pairwise classification we used logistic regression regularized with elastic net, and for multi-class classification we used multinomial logistic regression. The performance was estimated using 10 times repeated 5-fold stratified cross-validation. The pipeline was implemented using `DataJoint` (Yatsenko et al., 2015).

by class imbalance. Zero loss means perfect classification, while chance-level performance (for balanced classes) corresponds to the loss of $\ln(2) \approx 0.69$. We also computed cross-validated accuracies, and found that for our data, a log-loss of 0.2 roughly corresponded to 90% accuracy, a log-loss of 0.4 corresponded to 80% accuracy, and a log-loss of 0.6 corresponded to 65% accuracy (Figure S1).

The matrix of pairwise classification performances for the bipolar data set can serve as a sanity check that our classification pipeline works as intended: bipolar types with close numbers (e.g. type 1 and 2, or type 3A and 3B) are hard to distinguish (Figure 1), and indeed the loss values are generally higher close to the diagonal than far away from it (Figure 4). In fact, to assign types to the bipolar cell types 1/2, 3A/3B/4, and 5I/5O/5T, the original studies used not only the morphology information, but also tiling of the retina and synaptic input patterns (Helmstaedter et al., 2013; Behrens et al., 2016).

2.2 Morphometric statistics and density maps perform best

For each feature representation, we averaged the log-losses across all pairs within each data set and within each "modality" (full-neuron/axon/dendrite), obtaining $3 \times 3 = 9$ average log-losses for each of the 54 feature representations. Figure 5 shows a summary for the nine top performing features (see Methods for how they were selected). The performance using the dendritic features was consistently poor and much lower than using the axonal features (see also Figure S1). Indeed, for cortical interneurons as well as for retinal bipolar cells, it is the axonal, and not the dendritic, geometry that primarily drives the definition of cell types (Helmstaedter et al., 2013; Sümbül et al., 2014; Ascoli et al., 2008; Markram et al., 2004; DeFelipe et al., 2013), as can be seen in Figure 1. In turn, the performance using the axonal features was practically indistinguishable from the performance of the full-neuron features, consistent with the statistics of our data, where axonal neurites make up 85% of the total traced neuritic length (3.8 m out of 4.3 m).

Using the performance of the full-neuron features (Figure 5A), we found that the top performing feature representation were XZ density maps, with the mean log-loss of 0.16 ± 0.06 (mean±SD across $n = 3$ datasets), followed by Z density maps (0.19 ± 0.03), morphometric statistics (0.19 ± 0.03), and YZ density maps (0.26 ± 0.04), and then by 2D persistence (0.35 ± 0.08). The best performing morphometric distribution was path angle vs. path distance, but it performed noticeably worse (0.48 ± 0.08). Among the density maps, XZ density map is arguably the most meaningful, as it captures neuronal morphology along the cortical/retinal depth (z) and in the most prominent laminar direction (x). It also showed the

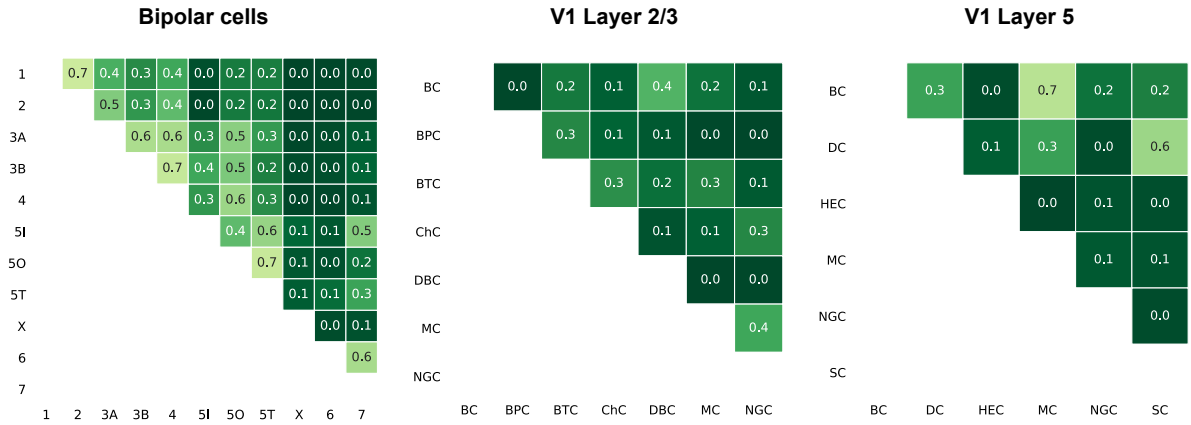

Figure 4: Cross-validated log-loss for each pair of morphological types in each data set using morphometric statistics as predictors in logistic regression. Zero log-loss corresponds to perfect prediction, $\ln(2) \approx 0.69$ corresponds to random guessing. For abbreviations see Figure 1.

best mean performance.

To make a statistical comparison between two different features A and B, we computed the difference $\delta(A, B)$ in log-loss across all 91 pairs of neural types (pooling pairs across the three data sets). The standard error of δ cannot be estimated directly because the pairs are not independent: e.g. the discriminative performances for bipolar types 1 and 2 and for bipolar types 1 and 3A include the same cells from type 1. We used a jackknife procedure *across types* (not across pairs) to estimate the standard error of each reported δ (see Methods). We found no evidence of significant difference in performance between morphometric statistics and XZ density maps ($\delta = 0.012 \pm 0.04$, mean across pairs with SD estimated via jackknife across types; $z = 0.29$, $p = 0.77$, z-test), but between morphometric statistics and 2D persistence ($\delta = 0.12 \pm 0.05$, $z = 2.47$, $p = 0.01$). Among the density maps, the Z density map did not show a significant difference from XZ ($\delta = 0.005 \pm 0.02$, $z = 0.26$, $p = 0.80$), while YZ was much worse than XZ in the two V1 data sets ($\delta = 0.13 \pm 0.03$, $z = 3.9$, $p = 0.0001$, average across V1 pairs only) but very similar in the bipolar data set ($\delta = 0.01 \pm 0.02$, $z = 0.59$, $p = 0.56$). Indeed, the y direction is mostly meaningless in the V1 data sets as biocytin slices are flattened during the staining process.

For completeness, we also performed all the classifications using axonal and dendritic features pooled together, but the resulting performance was very similar to the performance using axonal (or full-neuron) features alone (Figure S1).

2.3 Combining diverse feature representations can improve performance

Next, we asked if combining diverse feature representations can improve performance. We pooled morphometric statistics and XZ density maps, morphometric statistics and 2D persistence, XZ density maps and 2D persistence, and all three of these feature sets, yielding four additional combined feature sets (Figure 5, right). We found that any pairwise combination outperformed its constituent feature representation. In particular, morphometric statistics combined with XZ density maps (0.10 ± 0.04 for the full-neuron features, mean±SD across data sets) significantly outperformed both feature representations on their own ($\delta = 0.07 \pm 0.02$, $z = 2.88$, $p = 0.004$ for the combination against XZ density maps), suggesting that both capture some complementary information about neuron morphology. Combining XZ density maps with 2D persistence also led to a slight improvement compared to XZ density maps, but this improvement was not statistically significant ($\delta = 0.019 \pm 0.015$, $z = 1.23$, $p = 0.22$).

Finally, we investigated the performance when combining all three feature representations together. Compared to the XZ density maps combined with morphometric statistics, further adding 2D persistence led to no noticeable improvement ($\delta = 0.004 \pm 0.005$, $z = 0.87$, $p = 0.39$).

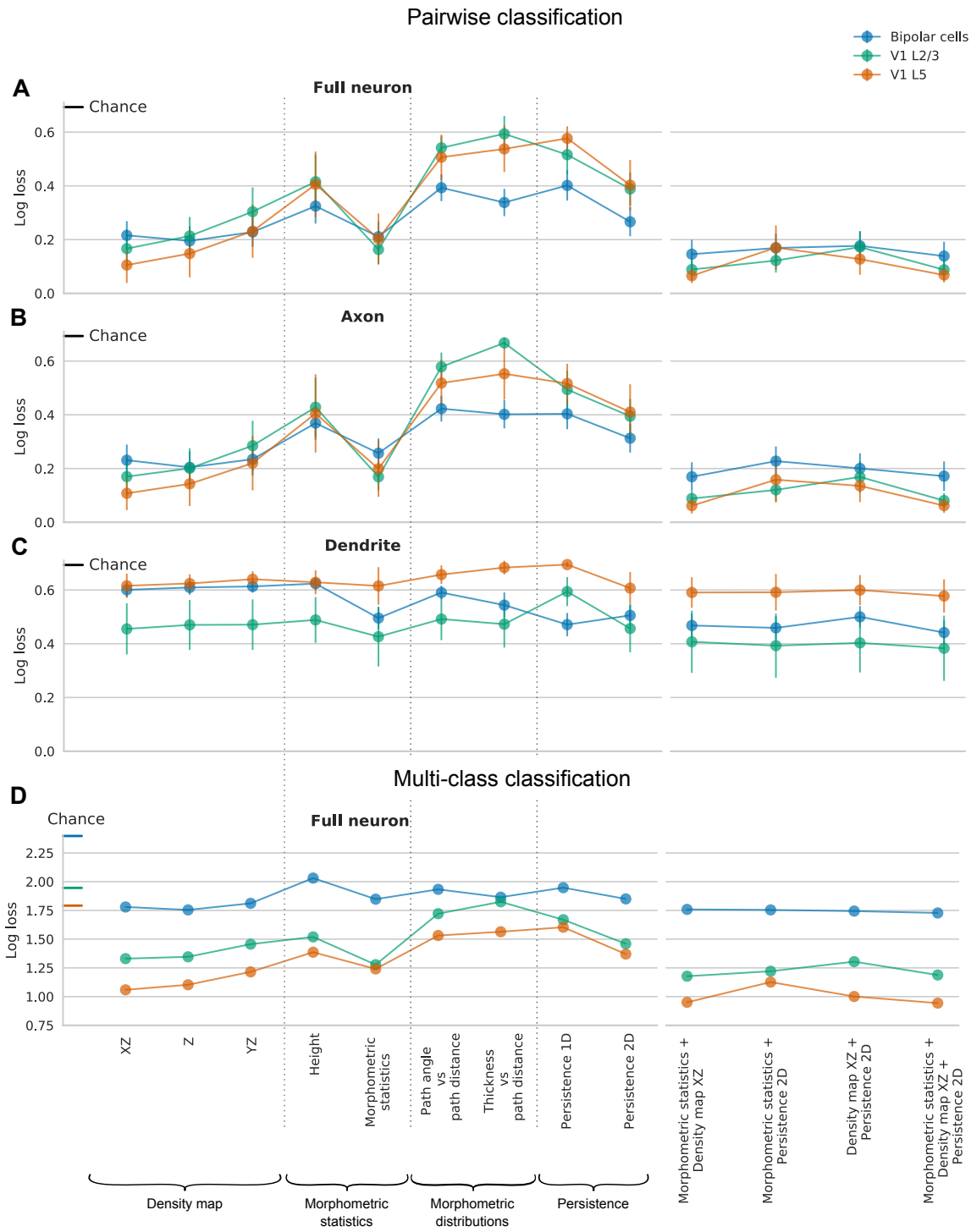

Figure 5: **(A–C)** Pairwise classification performance of the top performing feature representations based on the full-neuron (A), axonal (B), and dendritic (C) features for each data set. Feature representations are grouped into density maps, morphometric statistics, morphometric distributions, persistence images, and combinations of the top three feature representations. Each shown value is cross-validated log-loss, averaged across all pairs. Error bars correspond to 95% confidence intervals. Chance-level log-loss equals $\ln(2) \approx 0.69$ and is indicated in each panel. See Figure S1 for the results using combined axonal+dendritic feature representations. **(D)** Cross-validated log-loss of multinomial classification. Chance level for each data set is indicated on the y-axis. See Figure S2 for the results using combined axonal, dendritic, and axonal+dendritic feature representations.

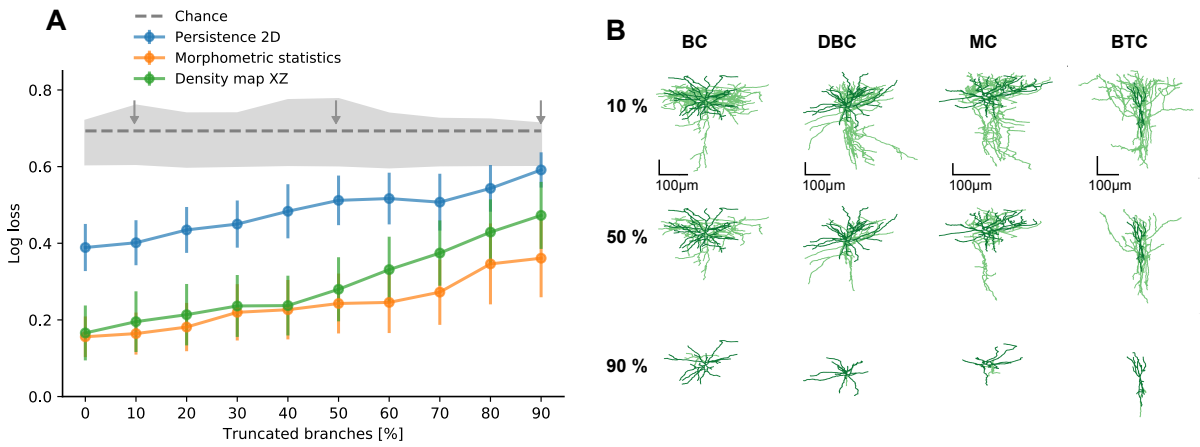

Figure 6: **(A)** Cross-validated log-loss of XZ density maps, morphometric statistics and 2D persistence as a function of truncation level. Branches were truncated to mimic what happens when neurons are only partially traced. The classification was performed on all pairs of types in V1 L2/3 data set. Dots and error bars show the means and 95% confidence intervals across all 21 pairs. Dashed grey line shows chance level at $\ln(2) \approx 0.69$. Grey shading shows the chance-level distribution of log-losses obtained by shuffling the labels. The arrows mark the levels of truncation shown in panel B. **(B)** XZ projections of four exemplary cells at three levels of truncation: 10%, 50%, and 90%. At 50% truncation the global structure of each cell is still preserved, whereas at 90% only the dendritic structures remain. See Figure 1 for abbreviations.

2.4 Multi-class classification

We investigated multi-class classification performance as an alternative to the pairwise classification presented above. We used multinomial logistic regression instead of the standard binomial one, and used exactly the same pipeline of regularization and cross-validation. For each of the 54 base full-neuron feature representations plus 4 combined ones, we obtained cross-validated multi-class loss in each data set (Figure 5D). Note that for each feature representation and data set, the performance is given by one single estimate, as opposed to the mean over all pairs that we reported above. Therefore only point estimates and no confidence intervals are shown in Figure 5D.

The values of multi-class loss are not directly comparable between data sets, because they are strongly influenced by the number of classes in a data set (K). The chance-level performance is given by $\ln(K)$ and is therefore different for each data set: $\ln(11) \approx 2.40$ for the bipolar data set, $\ln(7) \approx 1.95$ for the V1 L2/3 data set, and $\ln(6) \approx 1.79$ for the V1 L5 data set. For this reason, here we are not reporting averages across data sets.

The overall pattern was in good qualitative agreement with that obtained using pairwise classifications (Figure 5D). For two out of three data sets (bipolar and V1 L5), density maps performed the best (mean log loss between 1.75–1.81 and 1.06–1.22), followed by morphometric statistics (1.85/1.24), and for the third data set (V1 L2/3) morphometric statistics showed the smallest loss (1.28), very closely followed by the density maps (1.33–1.46). Combining morphometric statistics with XZ density maps and 2D persistence led to a clear improvement in all three data sets.

For the multi-class performances of the other "modalities" (axon, dendrite, and axon+dendrite) see Figure S2. They are also in good agreement with the pairwise performances: axonal and axonal+dendritic features perform very similar to the full-neuron features, while dendritic features perform much worse.

2.5 Morphometric statistics and density maps are robust against partial tracings

Accurate morphological reconstructions often become more and more difficult as one goes away from the cell soma, because the neurites become thinner and might have weaker staining which makes them easier to miss. We therefore assessed the robustness of using XZ density maps, morphometric statistics, and 2D persistence as predictors of cell type when neurons are only partially traced.

Partial tracings were simulated by subsequently removing 10 to 90% of the branches (in steps of 10%) of each reconstructed skeleton. On each truncation step, we removed the given fraction of branches with the highest branching order (see Methods). The branching order corresponds to the count of branch points that are passed when tracking the branch back to the soma, so the higher the branching order the more branching has occurred along this branch. This procedure cuts away most of the axonal neurites before reaching the dendrites that typically have branches of lower branch order, it therefore mimics what can happen in actual reconstructions. We used the V1 L2/3 data set for this analysis, performing all pairwise classifications between all pairs of cell types at each truncation step (Figure 6A). In addition, we shuffled the labels of each pairwise comparison to estimate the chance-level distribution of log-losses (Figure 6A, grey shading). Exactly the same cross-validation pipeline was run after shuffling the labels.

As expected, performance of each feature representation gradually decreased with increasing level of truncation. The decrease was rather moderate until around 30% truncation level (e.g. it grew from 0.16 ± 0.12 to 0.22 ± 0.17, mean$\pm$SD across all 21 pairs, for morphometric statistics, and very similarly for the density map). After that, all representations were noticeably losing in performance. At the highest levels of truncation, morphometric statistics outperformed other representations by a large margin. Even at 90% truncation, it had a log-loss of only 0.36 ± 0.24. This is in line with the good performance of dendritic morphometric statistics for the V1 L2/3 interneurons (Figure 5C) since it is mostly dendrites that remain at 90% truncation (Figure 6B).

2.6 Combining diverse features for unsupervised learning

So far we used supervised learning where cell type labels are assumed to be known. A more difficult and arguably more interesting task is to identify morphological cell types using unsupervised clustering (Sümbül et al., 2014; Gouwens et al., 2018). The small sample sizes of our data sets make it very challenging to obtain reliable clustering and to compare the clustering performance of various feature representations. Instead, we directly used one of the best performing feature representation identified in our classification study, a combination of morphometric statistics, XZ density map and 2D persistence, and performed unsupervised dimensionality reduction using t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten and Hinton, 2008). If the cell types are well-separated in the t-SNE embedding, then it is plausible that a clustering algorithm would identify them as separate types, given a large enough data set.

We reduced the XZ density maps, the 2D persistence image and morphometric statistics to 10 principal components each and normalized each set of 10 PCs by the standard deviation of the respective PC1, to put all three sets roughly on the same scale. We pooled the scaled PCs together and used this 30-dimensional feature representation as input for t-SNE (see Methods). Here we did not use cross-validation and embedded all cells from each data set at once (Figure 7).

The resulting embeddings correspond well to the pairwise classification performance that we presented earlier for morphometric statistics (Figure 4). For example, horizontally elongated cells (HECs) and neurogliaform cells (NGCs) in the V1 L5 data set that were both easily distinguishable from other types in the classification task, formed clear clusters, well-separated from other types. In contrast, retinal bipolar types that were hard to classify, such as types 1 and 2 or types 3A, 3B, and 4, formed joint clusters with a lot of overlap.

The combination of morphometric statistics, 2D persistence, and XZ density maps yielded t-SNE embeddings that had noticeably less overlap than any of these three feature representations on its own corresponding very well to their classification performance. The combination of morphometric statistics and XZ density maps (without 2D persistence) yielded a very similar embedding.

The embeddings shown in Figure 7 used full-neuron features that, as we saw above, are dominated by the axonal geometry. Applying the same procedure to the dendritic features yielded embeddings with far worse separation between cell types (Figure S3). Moreover, dendritic features resulted in t-SNE embeddings with far less structure than the full-neuron features (Figure S4), suggesting that there is less interesting variability in the dendritic morphologies compared to the axonal ones.

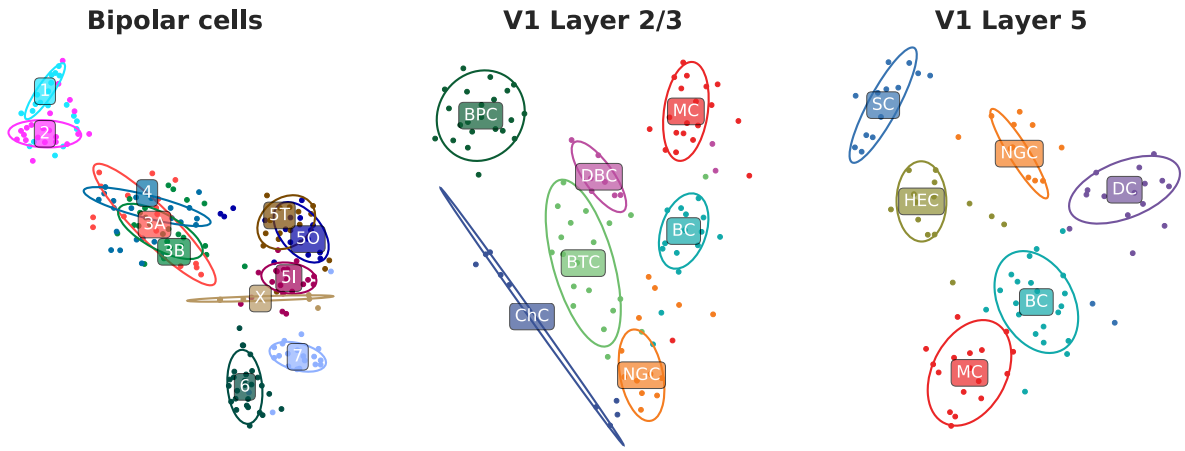

Figure 7: t-SNE embeddings of all three data sets (perplexity 30). We used a combined feature representation consisting of XZ density map, morphometric statistics and 2D persistence images. All three feature representations were reduced to 10 PCs and combined into a 30-dimensional feature set prior to t-SNE. The ellipses are 95% coverage ellipses for each type, assuming Gaussian distribution and using robust estimates of location and covariance. They are not influenced by single outliers. For abbreviations see Figure 1.

3 Discussion

Here we benchmarked existing morphological representations in the context of supervised cell type classification using well-curated data sets encompassing over 400 full reconstructions of neurons in the mouse visual system. We found that density maps and morphometric statistics yield the best predictions of cell type labels and showed that they do so even if substantial parts of the traced morphologies are removed. We demonstrated that the combination of density maps and morphometric statistics improved the prediction accuracy even further, suggesting that complementary information was captured by these two representations.

Previous literature has argued that on their own, density maps, morphometric statistics, and persistence work well for cell type identification and classification. Retinal cells, for example, can be successfully discriminated by their stratification depth within the inner plexiform layer (IPL) which can be seen as a z-projection of their neurite density (Helmstaedter et al., 2013; Sümbül et al., 2014). Morphometric statistics have been used in a wide variety of studies across species and brain areas (Lu et al., 2013, 2015; Gouwens et al., 2018) and have been shown to perform well in a one-vs-rest classification of cortical neurons (Mihaljević et al., 2015, 2018). Persistence, in turn, has lately been shown to distinguish 17 pyramidal neuron types in juvenile rat somatosensory cortex (Kanari et al., 2019).

These studies, however, have not directly compared different morphology representations, but rather focused on comparing classification schemes or establishing cell type related differences within their chosen morphological representation. Our study fills this gap by applying the same standardized classification procedure to each morphological representation, using well-curated data sets with well-defined cell types. This comparison revealed that the classical approaches of density maps and single-valued summary statistics yield the best performance and can be combined for further improved discriminatory power. On the other hand, persistence performed consistently worse. As a tool for analysis of neural morphologies, persistence was originally developed and validated using pyramidal cell dendrites and it possibly misses some information relevant to axons.

In our study, the axonal morphology of retinal bipolar cells and cortical interneurons contained more information about the cell type than its dendritic morphology, as described previously (Mihaljević et al., 2015; Jiang et al., 2015; Ofer et al., 2018). Indeed, also our unsupervised analysis demonstrated far more variability within axonal features compared to the dendritic features (Figure S4), in line with cell type naming conventions for these cell classes being based on axonal features (Markram et al., 2004; DeFelipe et al., 2013; Helmstaedter et al., 2013). Notwithstanding, dendritic reconstructions are more prevalent in the literature and in the available databases: at the time of writing only 55% (367/667) and

8% (630/8 129) of mouse cortical neurons in the Allen Cell Type atlas (http://celltypes.brain-map.org/data) and the NeuroMorpho library (Ascoli et al., 2007) are flagged as containing complete axonal reconstructions. This is because dendrites are usually thicker and more compact than axons and so are easier to stain and trace. This has obstructed acquisition of complete axonal reconstructions in mammals but might be remedied by recently developed whole brain imaging techniques (Ragan et al., 2012; Yuan et al., 2015; Economo et al., 2016; Gong et al., 2016). At the same time, the surprisingly robust classification performance of truncated morphologies (Figure 6) suggests that full reconstructions might not be necessary for cell type identification.

Limitations

The number of neurons (413) and neural types (24) within our study is relatively small. We believe that a meaningful comparison between different numerical descriptions of morphology is only possible when maintaining strict data consistence and quality criteria. This is why we restricted this work to data from only one species (mouse) of one developmental stage (adult) where morphological cell types are well established and supported by other studies based on electrophysiology (Jiang et al., 2015) or genetics (Shekhar et al., 2016). At the same time we wanted morphologies to be similar enough to pose a challenge and we required complete axonal and dendritic reconstructions. Despite these strict criteria, the resulting data set of 413 neurons is comparable to the sample sizes used in related studies (Mihaljević et al., 2015, 2018).

As the 413 neurons were split into 24 types, the median sample size per class was only 17 cells. It is difficult to fit statistical models in the $n \ll p$ regime where the number of dimensions p highly exceeds the number of samples n (Friedman et al., 2001). We used a simple linear model strongly regularized by PCA preprocessing and an elastic net penalty, as more complicated models would not be supported by our low sample sizes. This approach performs well when the leading principal components of the data have good discriminative power, but can also perform at chance level if the difference between types is restricted to low-variance directions. Thus, low classification performance for a given cell type pair does not necessarily imply that they could not be reliably separated with more available data.

We restricted our benchmarking effort to the most prominent and well-established morphological representations that have been independently employed by more than one research group. In particular, this excluded some methods based on graph theory (Heumann and Wittum, 2009; Gillette and Grefenstette, 2009) and sequence alignment (Gillette and Ascoli, 2015; Gillette et al., 2015; Costa et al., 2016) which can be promising candidates for further studies. The morphometric statistics that we used did not include everything that has been suggested in the literature either. For example, we did not use morphometric statistics such as fractal dimension because of their disputed relevance for our data (Panico and Sterling, 1995) and did not explicitly quantify the amount of layer-specific arborization (DeFelipe et al., 2013; Gouwens et al., 2018) because this concept only applies to cortical neurons and the same information is implicitly available through the Z density maps and other related spatial distributions.

Outlook

We showed that the combination of density maps and morphometric statistics yielded the best classification performance, suggesting that no feature representation captured all available information on its own. While immediately interpretable, both representations are purely descriptive and leave a deeper theoretical justification to be desired. Ideally, a theoretically grounded principled feature representation would perform at least as well for cell type discrimination as the feature combination that we considered here, and at the same time, it would consist of a much smaller set of interpretable parameters that could be used in a generative model of neural morphologies. A number of different approaches to generating neural morphologies have been suggested so far (van Pelt and Schierwagen, 2004; Cuntz et al., 2010; Memelli et al., 2013; Wolf et al., 2013; Farhoodi and Kording, 2018; Fard et al., 2018), but they mostly do not yield morphological representations that allow for straightforward cell type classification. We believe that research on interpretable generative models for neural morphologies will be an important direction for future research.

4 Methods

4.1 Data and code availability

We used data from Helmstaedter et al. (2013) and Jiang et al. (2015), splitting the latter data set in two parts by cortical layer. All neurons were labelled in the original studies by human experts. We confirmed the quality of all reconstructions through inspection. Our study investigated a total of 4.29 meters of traced neurites from $n = 413$ neurons.

1. **Bipolar cells.** This data set comprised $n = 221$ tracings of retinal bipolar cells in one mouse (p30) from electron-microscopy data (Helmstaedter et al., 2013). To allow for at least 5-fold cross-validation, we did not analyze cell types which had counts of 5 cells or fewer. This criterion excluded types 8 and 9 and resulted in $n = 212$ remaining morphologies in 11 types. The reconstructions (as .SWC files) as well as their cell type labels were obtained from the authors of Behrens et al. (2016) which explains the additional cell types 5O, 5I and 5T as compared to the original work.

2. **V1 Layer 2/3.** Manually traced biocytin stainings of $n = 108$ inhibitory interneurons of 7 types in layer 2/3 (L2/3) adult mouse primary visual cortex (Jiang et al., 2015). We obtained the reconstructions (as .ASC file) and their cell type labels from the authors.

3. **V1 Layer 5.** Manually traced biocytin stainings of $n = 94$ inhibitory interneurons of 6 types in layer 5 (L5) adult mouse primary visual cortex (Jiang et al., 2015). One deep-projecting cell lacked an axon so it was excluded from further analysis resulting in $n = 93$ remaining morphologies. We obtained the reconstructions (as .ASC file) and their cell type labels from the authors.

The reconstructions of the V1 interneurons are available at `http://neuromorpho.org` (archive `Tolias`). Check the "Include Auxiliary files" box to get the originally uploaded files with cortical depth alignment. The bipolar cell reconstructions will be made available upon publication of the manuscript.

Our analysis code is available at `http://github.com/berenslab/morphology-benchmark`. The processing pipeline including preprocessing, feature extraction and classification was automated using DataJoint (Yatsenko et al., 2015).

4.2 Preprocessing and nomenclature

Reconstructed morphologies were converted into SWC format using `NLMorphologyConverter` 0.9.0 (`http://neuronland.org`) where needed and further analysed in Python. The SWC format represents a morphology with a list of *nodes* (points) with each node described by its *id*, *3D position*, *radius*, *type* (1: soma, 2: axon, 3: dendrite), and *parent id*. Each node connects to its parent node with a straight line that we will call "sub-segment". Several nodes can connect to the same parent note; in this case this parent node is called a "branch point". A neurite path from one branch point to the next is called a "segment". The bipolar cells were missing explicit type labels for the soma, we therefore set every node of radius larger than 1 micron to be somatic. We generally allowed for only one somatic node so that within one reconstruction all somatic nodes were grouped and replaced by one node with position and radius being the mean across all original soma nodes. Especially in the initial branch segments it can occur that node type labels (1: soma, 2: axon, 3: dendrite) are not consistent between consecutive nodes. Node type labels within one branch we therefore assigned according to the majority vote over all sub-segment types within this branch. All interneurons were soma-centered and rotated such that the z coordinate (height) was oriented along the cortical depth. Bipolar cells were soma-centered in their x and y coordinates and IPL-centered in the z direction corresponding to the retinal depth (i.e. $z = 0$ corresponded to the outer border of the inner plexiform layer, IPL). All cells were rotated in the xy plane such that their x coordinate (width) was oriented along the first principal component of the xy point cloud, i.e. roughly corresponded to the cell's largest extent in the plane orthogonal to the cortical/retinal depth.

4.3 Feature representations

We calculated 54 different feature representations for each cell. These representations can be grouped into four categories: *density maps*, *morphometric statistics*, *morphometric distributions*, and *persistence*

images. All feature representations were separately computed for axons, for dendrites, and for the whole neuron (without distinguishing axons from dendrites) yielding $54 \cdot 3 = 162$ representations per neuron (see Figure 2).

4.3.1 Density maps

We sampled equidistant points with 25 nm spacing along each neurite of the traced skeletons and normalized the resulting point cloud by the maximal extent within each data set. For 2D density maps, the normalized point cloud was projected onto the xy, xz, and yz planes and binned into 100×100 bins spanning $[-0.1, 1.1]$. For 1D density maps the normalized point cloud was projected onto the x, y, and z axes and binned into 100 bins spanning $[-0.1, 1.1]$. We smoothed the resulting histograms by convolving them with a 11×11 (for 2D) or 11-bin Gaussian kernel with standard deviation $\sigma = 2$ bins. For the purposes of downstream analysis, we treated the density maps as vectors of 10 000 (for 2D) or 100 (for 1D) features. Overall we used 6 versions of density map representations.

4.3.2 Morphometric statistics

For each cell we computed a set of 22 single-valued summary statistics:

number of branch points Count of points at which the neurites branch.

cell width Extent of the cell in the x direction. This direction corresponds to the direction of the largest variance of the xy point cloud.

cell depth Extent of the cell in the y direction. This direction corresponds to the direction of the smallest variance of the xy point cloud.

cell height Extent of the cell in the z direction. This direction corresponds to the cortical/retinal depth.

number of tips Count of end-points.

number of stems Count of neurites extending directly from the soma.

maximal degree Maximal number of neurites meeting at a single branch point.

total length Total length of all neurites.

maximal neurite length Length of the longest neurite from tip to soma.

maximum branch order The maximum number of branch points passed when tracing a neurite from the tip back to the soma. Branch ordering starts with the soma having branch order 0 and each subsequent branching point increases the order by 1.

maximum segment length Length of the longest segment (w.r.t. Euclidean distance).

average thickness The average radius across all nodes.

surface Estimated total surface area of all neurites. Each neurite sub-segment is assumed to be a truncated cone, and its surface was computed as $\pi(r + R)\sqrt{(R - r)^2 + h^2}$ where h is the length of the sub-segment and r and R are the radii at both ends.

volume Estimated total volume of all neurites, computed as $\frac{1}{3}\pi h(r^2 + rR + R^2)$ for each sub-segment.

minimal tortuosity Minimal tortuosity across all neurite segments. Tortuosity describes the "bendiness" of a segment and is defined as the ratio of path length to the Euclidean distance between the end-points.

maximal tortuosity	Maximal tortuosity across all neurite segments.
average tortuosity	The average tortuosity across all neurite segments.
minimal branch angle	For each pair of branches meeting at a branching point, "branch angle" is the angle (in $[0, 180]$ degrees range) between the meeting sub-segments. Minimal branch angle refers to the minimal branch angle across all such pairs.
maximal branch angle	Maximal branch angle, see above.
average branch angle	Average branch angle, see above.
maximal path angle	"Path angle" is the angle between two consecutive sub-segments (not including the sub-segments that meet at a branching point). Maximal path angle is the maximum across all such consecutive sub-segments.
average path angle	Average path angle, see above.

In addition to the 22 features listed above, we grouped all of them together into one vector which we named *morphometric statistics*.

4.3.3 Morphometric distributions

For each cell we computed the following 17 one-dimensional morphometric distributions:

branch angles	Histogram of bifurcation angles between neurites (20 bins from 0 to 180).
branch orders	A vector of length K where K is the maximal branch order within each data set, with element i equal to the number of branch-points of order i.
path angles	Histogram of the path angles (20 bins from 0 to 180).
root angles	Histogram of the root angles of each segment. The root angle denotes the angle between the straight line connecting the segment's start and end nodes and the straight line that connects the segment's end point with the soma (20 bins from 0 to 180). It is indicative of the preferred growing direction of the neural arbor.
Euler root angles, α	Histogram of the Euler root angles (α). Here, the root angles are expressed as Euler angles around the x, y and z axis (20 bins from 0 to 180).
Euler root angles, β	See above.
Euler root angles, γ	See above.
segment lengths	Histogram of segment lengths in terms of Euclidean distance between the end-points of each segment (20 bins from 0 to the maximal segment length within the data set).
thickness	Histogram of nodes's radii (30 bins from 0 to the maximal radius within the data set).
path distance to soma	Histogram of the path length of each branch point and tip to the soma (20 bins from 0 to maximal neurite length within data set)
Euclidean distances to soma	Histogram of the Euclidean distance of each branch point and tip to the soma.

Sholl intersection xy Sholl intersection profile in the xy-plane. The Sholl intersection profile describes the number of intersection of a 2D projection with concentric circles of different radius. One difference to the classical procedure of Sholl (1953) is that we put the origin at the centroid of the convex hull not at the soma (36 steps from centroid to maximal radial distance from centroid).

Sholl intersection xz Sholl intersection profile in the xz-plane.

Sholl intersection yz Sholl intersection profile in the yz-plane.

3-star motif 10-dimensional vector with element i being the number of 3-star motifs in the sub-graph containing $i \cdot 10\%$ of all nodes closest to the soma. A 3-star motif represents a branch point where one branch forks into two branches. For each additional branch the 3-star motif is counted again (e.g. a neurite forking into three sub-branches is counted as having two 3-star motifs).

average 3-star motif The 3-star motif vector is calculated 100 times using different random nodes instead of the soma as the "center" of concentric spheres. The resulting vectors are then averaged.

average maximal distance Using the same sub-graphs, we compute the maximal neurite length in each sub-graph (10-dimensional vector). This procedure is repeated 100 times with random starting nodes and the resulting vectors are then averaged.

In addition, we used six two-dimensional morphometric distributions. The binning and normalization were the same as for the respective 1D distributions.

branch angles × branch orders 2D histogram of branch angles and branch orders (across all branch points).

branch angles × path distances 2D histogram of branch angles and path distances from soma (across all branch points).

path angle × branch order 2D histogram of path angles and branch orders (across all nodes, not only branch points).

path angle × path distance 2D histogram of path angles and path distances from soma (across all nodes).

thickness × branch order 2D histogram of neurite radii and branch orders (across all nodes).

thickness × path distance 2D histogram of neurite radii and path distances from soma (across all nodes).

4.3.4 Persistence image

The persistence image is a morphological representation introduced by Kanari et al. (2018). We briefly outline it here. Starting from each tip, one records the "birth time" of each branch as the radial distance of the tip from the soma. While moving away from the tips towards the soma, at each branch-point the "younger" branch, i.e. the one with a smaller birth time, is "killed" and its "death time" is recorded. This results in a 2D point cloud of (*birth time*, *death time*) for each branch with only the longest branch surviving until the soma having a death time of 0. To obtain the 2D persistence image we performed kernel density estimation of this point cloud using a 2D Gaussian kernel (`gaussian_kde` from the `scipy.stats` package with default settings). We evaluated the density estimate on a 100×100 equidistant grid spanning a $[0, \max_{\text{birth}}] \times [0, \max_{\text{death}}]$ rectangle. Here $\max_{\text{birth}}$ and $\max_{\text{death}}$ refer to the maxima across all cells within each data set (Bipolar cells: 89.8 μm and 74.0 μm, V1 L2/3: 915.7 μm

and 735.6 μm, V1 L5: 1615.7 μm and 647.9 μm). For the purposes of downstream analysis, we treated this as a set of $10,000$ features.

The 1D persistence image is obtained in a similar way. Namely, we performed a one-dimensional Gaussian kernel density estimation of the neurites' "living time" (birth$-$death) and sampled the resulting estimate at 100 equidistant points spanning $[0, \max_{\text{birth}}]$.

4.4 Classification

Each feature representation was used as a predictor for pairwise and multinomial classification. Except for morphometric statistics, we reduced all representations to 10 principal components to reduce overfitting (for cross-validation, PCA was computed on each outer-loop training set separately, and the same transformation was applied to the corresponding outer-loop test set).

For binary classification, we used logistic regression with an elastic net regularization. It minimizes the following loss function:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i(\mathbf{x}_i^\top\boldsymbol{\beta}) - \log(1+e^{-\mathbf{x}_i^\top\boldsymbol{\beta}})\Big] + \lambda\Big[\alpha\|\boldsymbol{\beta}\|_2^2/2 + (1-\alpha)\|\boldsymbol{\beta}\|_1\Big], \tag{1}$$

where $\mathbf{x}_i$ are predictors, y_i a binary response (that can be 0 or 1), and there are N training samples. Regularization parameter α was fixed to 0.5, which is giving equal weights to the lasso and ridge penalties. We used nested cross-validation to choose the optimal value of the regularization parameter λ and to obtain an unbiased estimate of the performance. The inner loop was performed using the `civisanalytics` Python wrapper around the `glmnet` library (Friedman et al., 2010) that does K-fold cross-validation internally (default: 3-fold). We kept the default setting which uses the maximal value of λ with cross-validated loss within one standard error of the lowest loss (`lambda_best`) to make the test-set predictions. We explicitly made the `civisanalytics` Python wrapper use the loss (and not accuracy) for λ selection:

```
from glmnet.scorer import make_scorer
from sklearn.metrics import log_loss

m = LogitNet(alpha=0.5, n_splits=3, random_state=17)
m.scoring = make_scorer(log_loss, greater_is_better=False, needs_proba=True)
```

Note that the default behavior of `glmnet` is to standardize all predictors. The outer loop was 10 times repeated stratified 5-fold cross-validation, as implemented in `scikit-learn` by

```
RepeatedStratifiedKFold(n_splits=5, n_repeats=10, random_state=17)
```

Model performance was assessed via mean test-set log-loss and test-set accuracy.

For multi-class classification we used multinomial logistic regression with an elastic net regularization. The parameters and the cross-validation procedure were the same as above.

4.4.1 Selection of top performing features

We identified the top five performing feature representations for each "modality" (full-neuron, axon, dendrite, as well as axon + dendrite) based on their mean binary classification performance across data sets and used their superset (9 features) for Figures 5, S1, and S2.

4.4.2 Statistical analysis of differences

We estimated the mean difference between two feature representations A and B as

$$\delta(A,B) = \frac{1}{|\mathcal{P}|}\sum_{p\in\mathcal{P}}\big(\ell(B,p) - \ell(A,p)\big),$$

where $\mathcal{P}$ denotes the set of pairs of types across the three data sets, $|\mathcal{P}| = 91$ is their total number, and $\ell(X, p)$ is cross-validated log loss of feature X for pair p. To estimate the standard error of $\delta(A, B)$ we use the jackknife procedure (Efron and Hastie, 2016). We repeat the procedure leaving out one type τ entirely, so that all pairs including that type are left out:

$$\delta_{-\tau} = \frac{1}{|\mathcal{P}_{-\tau}|} \sum_{p \in \mathcal{P}_{-\tau}} \big(\ell(B, p) - \ell(A, p)\big),$$

where $P_{-\tau}$ is the set of pairs without type τ. This yields $n = 21$ estimates of $\delta_{-\tau}$, with the jackknife estimate of the standard error given by

$$\widehat{\mathrm{SE}}(\delta(A, B)) = \Big[\frac{n-1}{n} \sum_{\tau} (\delta_{-\tau} - \bar{\delta})^2\Big]^{\frac{1}{2}},$$

where $\bar{\delta} = \frac{1}{n} \sum_{\tau} \delta_{-\tau}$. All reported p-values were obtained with a z-test using

$$z = \frac{\delta(A, B)}{\widehat{\mathrm{SE}}(\delta(A, B))}.$$

4.5 t-SNE visualization

For the t-SNE visualization (van der Maaten and Hinton, 2008) of each data set, we reduced density maps and morphometric statistics to 10 principal components each. We scaled each set of 10 PCs by the standard deviation of the respective PC1, to put the three sets roughly on the same scale. Then we stacked them together to obtain a 20-dimensional representation of each cell. Exact (non-approximate) t-SNE was run with perplexity 30 and random initialization using the `scikit-learn` implementation:

```
TSNE(perplexity=30, method='exact', random_state=42)
```

To plot the coverage ellipses for each cell type (Figure 7), we used robust estimates of location and covariance, so that the ellipses are not influenced by outliers. We used the minimum covariance determinant estimator (Rousseeuw and Driessen, 1999) as implemented in `MinCovDet()` in `scikit-learn`.

4.6 Robustness analysis

Morphological tracings are done manually and reconstruction quality can vary between protocols and experts. We assessed the robustness of the top performing feature representations by repeating the classification procedure on only partially traced neurons. We simulated incomplete tracings using the full reconstructions of the V1 L2/3 data set and assessed the performance of all pairwise classifications (Fig. 6). Incomplete tracings were simulated by successively removing 10–90% of all branches starting with the branches of the highest branch order. We then used the XZ density map, the morphometric statistics and the 2D persistence image for each grade of truncation as predictors.

To estimate the chance-level performance we shuffled the class labels for each pairwise classification at each truncation grade and repeated our classification pipeline with shuffled labels. The resulting distribution of chance-level test-set log loss values was in agreement with the theoretical value of $\ln(2)$.

Author contributions

PB and SL conceived the approach. SL performed the analysis with supervision from DK and PB. SL and DK wrote the manuscript with support from PB.

Acknowledgements

We thank Ulrike von Luxburg, Debarghya Ghoshdastidar, Xialong Jiang, Federico Scala and Andreas Tolias for discussions. This work was funded by the German Ministry of Education and Research

(FKZ 01GQ1601) and the German Research Foundation (DFG) under Germany´s Excellence Strategy (EXC 2064/1 – 390727645) and grants to PB (BE5601/4-1, SFB 1233 "Robust Vision" - Project number 276693517) and the National Institutes of Health BRAIN Initiative (1U19MH114830-01).

References

Giorgio A Ascoli, Duncan E Donohue, and Maryam Halavi. Neuromorpho.org: a central resource for neuronal morphologies. *Journal of Neuroscience*, 27(35):9247–9251, 2007.

Giorgio A Ascoli, Lidia Alonso-Nanclares, Stewart A Anderson, German Barrionuevo, Ruth Benavides-Piccione, Andreas Burkhalter, György Buzsáki, Bruno Cauli, Javier DeFelipe, Alfonso Fairén, et al. Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, 9(7):557, 2008.

Tom Baden, Philipp Berens, Katrin Franke, Miroslav Román Rosón, Matthias Bethge, and Thomas Euler. The functional diversity of retinal ganglion cells in the mouse. *Nature*, 529(7586):345, 2016.

Christian Behrens, Timm Schubert, Silke Haverkamp, Thomas Euler, and Philipp Berens. Connectivity map of bipolar cells and photoreceptors in the mouse retina. *eLife*, 5:e20041, 2016.

Alessandro Bria, Giulio Iannello, Leonardo Onofri, and Hanchuan Peng. Terafly: real-time three-dimensional visualization and annotation of terabytes of multidimensional volumetric images. *Nature Methods*, 13(3):192, 2016.

Kevin L Briggman, Moritz Helmstaedter, and Winfried Denk. Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337):183, 2011.

Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M Ibrahim, Andrew J Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J Steemers, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, page 1, 2019.

Xi Chen, Sarah A Teichmann, and Kerstin B Meyer. From tissues to cell types and back: single-cell gene expression analysis of tissue architecture. *Annual Review of Biomedical Data Science*, 1:29–51, 2018.

Marta Costa, James D Manton, Aaron D Ostrovsky, Steffen Prohaska, and Gregory SXE Jefferis. NBLAST: Rapid, sensitive comparison of neuronal structure and construction of neuron family databases. *Neuron*, 91(2):293–311, 2016.

Hermann Cuntz, Friedrich Forstner, Alexander Borst, and Michael Häusser. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Computational Biology*, 6 (8):e1000877, 2010.

Javier DeFelipe, Pedro L López-Cruz, Ruth Benavides-Piccione, Concha Bielza, Pedro Larrañaga, Stewart Anderson, Andreas Burkhalter, Bruno Cauli, Alfonso Fairén, Dirk Feldmeyer, et al. New insights into the classification and nomenclature of cortical gabaergic interneurons. *Nature Reviews Neuroscience*, 14(3):202, 2013.

Michael N Economo, Nathan G Clack, Luke D Lavis, Charles R Gerfen, Karel Svoboda, Eugene W Myers, and Jayaram Chandrashekar. A platform for brain-wide imaging and reconstruction of individual neurons. *eLife*, 5:e10566, 2016.

Bradley Efron and Trevor Hastie. *Computer age statistical inference*, volume 5. Cambridge University Press, 2016.

Pegah Kassraian Fard, Michael Pfeiffer, and Roman Bauer. A generative growth model for thalamocortical axonal branching in primary visual cortex. *bioRxiv*, 2018.

Roozbeh Farhoodi and Konrad P Kording. Sampling neuron morphologies. *bioRxiv*, 2018.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1, 2010.

Todd A Gillette and Giorgio A Ascoli. Topological characterization of neuronal arbor morphology via sequence representation: I-motif analysis. *BMC Bioinformatics*, 16(1):216, 2015.

Todd A Gillette and John J Grefenstette. On comparing neuronal morphologies with the constrained tree-edit-distance. *Neuroinformatics*, 7(3):191–194, 2009.

Todd A Gillette, Parsa Hosseini, and Giorgio A Ascoli. Topological characterization of neuronal arbor morphology via sequence representation: II-global alignment. *BMC Bioinformatics*, 16(1):209, 2015.

Hui Gong, Dongli Xu, Jing Yuan, Xiangning Li, Congdi Guo, Jie Peng, Yuxin Li, Lindsay A Schwarz, Anan Li, Bihe Hu, et al. High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nature Communications*, 7:12142, 2016.

Nathan W Gouwens, Staci A Sorensen, Jim Berg, Changkyu Lee, Tim Jarsky, Jonathan Ting, Susan M Sunkin, David Feng, Costas Anastassiou, Eliza Barkan, et al. Classification of electrophysiological and morphological types in mouse visual cortex. *bioRxiv*, 2018.

Moritz Helmstaedter, Kevin L Briggman, Srinivas C Turaga, Viren Jain, H Sebastian Seung, and Winfried Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500 (7461):168, 2013.

Holger Heumann and Gabriel Wittum. The tree-edit-distance, a measure for quantifying neuronal morphology. *Neuroinformatics*, 7(3):179–190, 2009.

Gregory SXE Jefferis, Christopher J Potter, Alexander M Chan, Elizabeth C Marin, Torsten Rohlfing, Calvin R Maurer Jr, and Liqun Luo. Comprehensive maps of drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell*, 128(6):1187–1203, 2007.

Xiaolong Jiang, Shan Shen, Cathryn R Cadwell, Philipp Berens, Fabian Sinz, Alexander S Ecker, Saumil Patel, and Andreas S Tolias. Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, 350(6264):aac9462, 2015.

Lida Kanari, Paweł Dłotko, Martina Scolamiero, Ran Levi, Julian Shillcock, Kathryn Hess, and Henry Markram. A topological representation of branching neuronal morphologies. *Neuroinformatics*, 16(1): 3–13, 2018.

Lida Kanari, Srikanth Ramaswamy, Ying Shi, Sebastien Morand, Julie Meystre, Rodrigo Perin, Marwan Abdellah, Yun Wang, Kathryn Hess, and Henry Markram. Objective morphological classification of neocortical pyramidal cells. *Cerebral Cortex*, 2019.

Yanjie Li, Dingkang Wang, Giorgio A Ascoli, Partha Mitra, and Yusu Wang. Metrics for comparing neuronal tree shapes based on persistent homology. *PloS One*, 12(8):e0182184, 2017.

Yanbin Lu, Kristen Trett, William Shain, Lawrence Carin, Ronald Coifman, and Badrinath Roysam. Quantitative profiling of microglia populations using harmonic co-clustering of arbor morphology measurements. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1360–1363. IEEE, 2013.

Yanbin Lu, Lawrence Carin, Ronald Coifman, William Shain, and Badrinath Roysam. Quantitative arbor analytics: unsupervised harmonic co-clustering of populations of brain cell arbors based on L-measure. *Neuroinformatics*, 13(1):47–63, 2015.

Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10):793, 2004.

Heraldo Memelli, Benjamin Torben-Nielsen, and James Kozloski. Self-referential forces are sufficient to explain different dendritic morphologies. *Frontiers in Neuroinformatics*, 7:1, 2013.

Bojan Mihaljević, Ruth Benavides-Piccione, Concha Bielza, Javier DeFelipe, and Pedro Larrañaga. Bayesian network classifiers for categorizing cortical gabaergic interneurons. *Neuroinformatics*, 13 (2):193–208, 2015.

Bojan Mihaljević, Pedro Larrañaga, Ruth Benavides-Piccione, Sean Hill, Javier DeFelipe, and Concha Bielza. Towards a supervised classification of neocortical interneuron morphologies. *BMC Bioinformatics*, 19(1):511, 2018.

Netanel Ofer, Orit Shefi, and Gur Yaari. Axonal tree morphology and signal propagation dynamics improve neuronal classification. *bioRxiv*, 2018.

Marius Pachitariu, Carsen Stringer, Mario Dipoppa, Sylvia Schröder, L Federico Rossi, Henry Dalgleish, Matteo Carandini, and Kenneth D Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, 2017.

Joseph Panico and Peter Sterling. Retinal neurons and vessels are not fractal but space-filling. *Journal of Comparative Neurology*, 361(3):479–490, 1995.

Hanchuan Peng, Zongcai Ruan, Fuhui Long, Julie H Simpson, and Eugene W Myers. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nature Biotechnology*, 28(4):348, 2010.

Hanchuan Peng, Jianyong Tang, Hang Xiao, Alessandro Bria, Jianlong Zhou, Victoria Butler, Zhi Zhou, Paloma T Gonzalez-Bellido, Seung W Oh, Jichao Chen, et al. Virtual finger boosts three-dimensional imaging and microsurgery as well as terabyte volume image visualization and analysis. *Nature Communications*, 5:4342, 2014.

Sridevi Polavaram, Todd A Gillette, Ruchi Parekh, and Giorgio A Ascoli. Statistical analysis and data mining of digital reconstructions of dendritic morphologies. *Frontiers in Neuroanatomy*, 8:138, 2014.

Timothy Ragan, Lolahon R Kadiri, Kannan Umadevi Venkataraju, Karsten Bahlmann, Jason Sutin, Julian Taranda, Ignacio Arganda-Carreras, Yongsoo Kim, H Sebastian Seung, and Pavel Osten. Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature Methods*, 9(3):255, 2012.

Santiago Ramón y Cajal. *Textura del sistema nervioso del hombre y de los vertebrados*. Moya, 1899.

Peter J Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.

Arpiar Saunders, Evan Z Macosko, Alec Wysoker, Melissa Goldman, Fenna M Krienen, Heather de Rivera, Elizabeth Bien, Matthew Baum, Laura Bortolin, Shuyu Wang, et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell*, 174(4):1015–1030, 2018.

Simon R Schultz, Caroline S Copeland, Amanda J Foust, Peter Quicke, and Renaud Schuck. Advances in two-photon scanning and scanless microscopy technologies for functional neural circuit imaging. *Proceedings of the IEEE*, 105(1):139–157, 2017.

Ruggero Scorcioni, Sridevi Polavaram, and Giorgio A Ascoli. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, 3(5):866, 2008.

Karthik Shekhar, Sylvain W Lapan, Irene E Whitney, Nicholas M Tran, Evan Z Macosko, Monika Kowalczyk, Xian Adiconis, Joshua Z Levin, James Nemesh, Melissa Goldman, et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*, 166(5):1308–1323, 2016.

D A Sholl. Dendritic organization in the neurons of the visual and motor cortices of the cat. *Journal of Anatomy*, 87(4):387, 1953.

Oliver Stegle, Sarah A Teichmann, and John C Marioni. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, 16(3):133, 2015.

Carsen Stringer and Marius Pachitariu. Computational processing of neural recordings from calcium imaging data. *Current Opinion in Neurobiology*, 55:22–31, 2019.

Uygar Sümbül, Sen Song, Kyle McCulloch, Michael Becker, Bin Lin, Joshua R Sanes, Richard H Masland, and H Sebastian Seung. A genetic and computational approach to structurally classify neuronal types. *Nature Communications*, 5:3512, 2014.

Bosiljka Tasic, Zizhen Yao, Lucas T Graybuck, Kimberly A Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N Economo, Sarada Viswanathan, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72, 2018.

Harry BM Uylings and Jaap van Pelt. Measures for quantifying dendritic arborizations. *Network: Computation in Neural Systems*, 13(3):397–414, 2002.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

Jaap van Pelt and Andreas Schierwagen. Morphological analysis and modeling of neuronal dendrites. *Mathematical Biosciences*, 188(1-2):147–155, 2004.

Sergei Wolf, Stephan Grein, and Gillian Queisser. Employing NeuGen 2.0 to automatically generate realistic morphologies of hippocampal neurons and neural networks in 3D. *Neuroinformatics*, 11(2): 137–148, 2013.

Dimitri Yatsenko, Jacob Reimer, Alexander S Ecker, Edgar Y Walker, Fabian Sinz, Philipp Berens, Andreas Hoenselaar, Ronald James Cotton, Athanassios S Siapas, and Andreas S Tolias. DataJoint: managing big scientific data using MATLAB or Python. *bioRxiv*, 2015.

Jing Yuan, Hui Gong, Anan Li, Xiangning Li, Shangbin Chen, Shaoqun Zeng, and Qingming Luo. Visible rodent brain-wide networks at single-neuron resolution. *Frontiers in Neuroanatomy*, 9:70, 2015.

Amit Zeisel, Hannah Hochgerner, Peter Lönnerberg, Anna Johnsson, Fatima Memic, Job Van Der Zwan, Martin Häring, Emelie Braun, Lars E Borm, Gioele La Manno, et al. Molecular architecture of the mouse nervous system. *Cell*, 174(4):999–1014, 2018.

Hongkui Zeng and Joshua R Sanes. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530, 2017.

Supplementary Figures

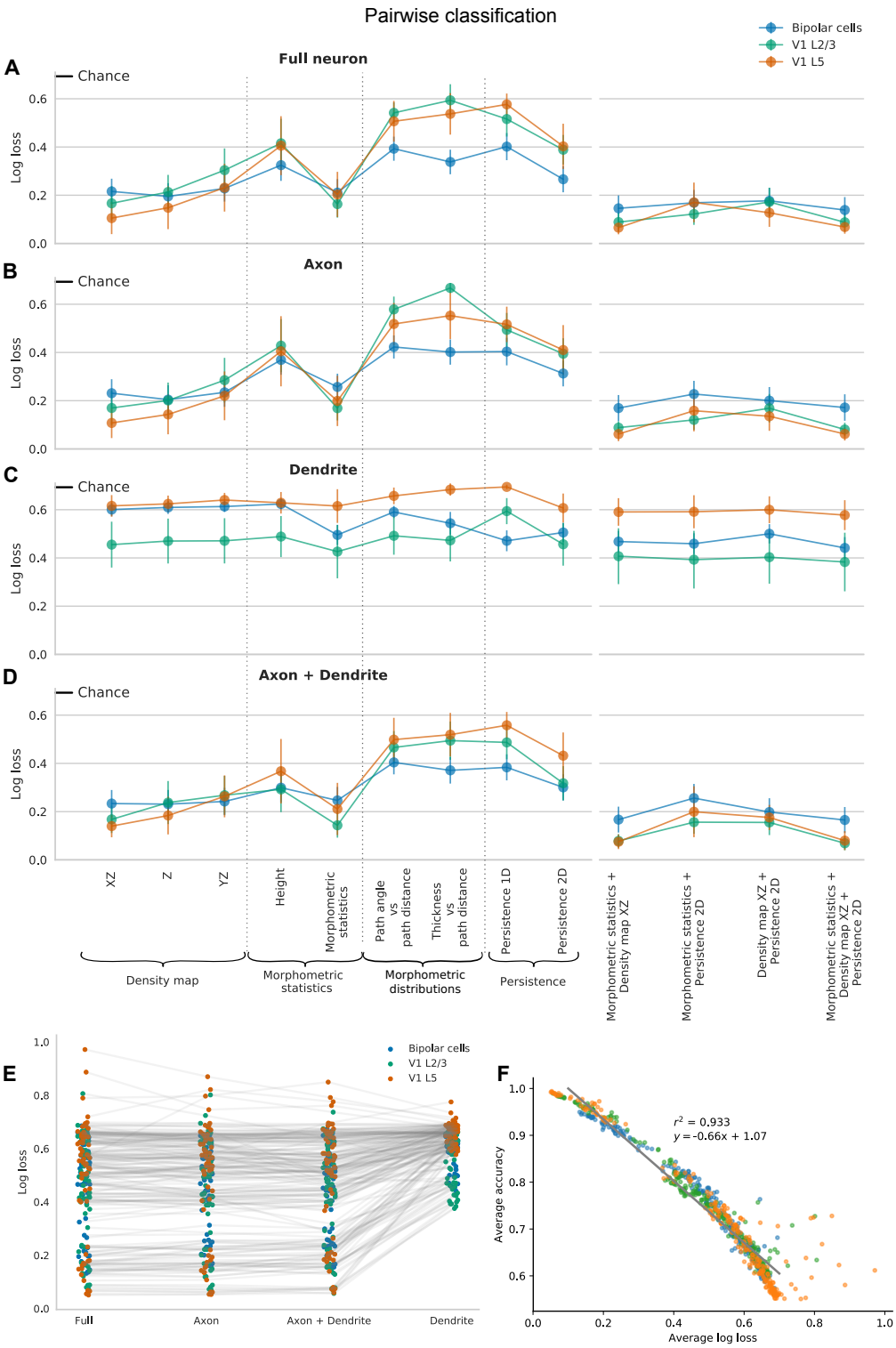

Figure S1: **(A–D)** Pairwise classification performance of the top performing feature representations based on features of different "modalities" (A–D) for each data set. Feature representations are grouped into density maps, morphometric statistics, morphometric distributions, persistence images, and combinations of the top three feature representations. Each shown value is cross-validated log-loss, averaged across all pairs. Error bars correspond to 95% confidence intervals. Chance-level log-loss equals $\ln(2) \approx 0.69$ and is indicated in each panel. **(E)** Comparison of average log-losses across pairs for each statistic in each modality and data set. Dendrites perform considerably worse than any of the other modalities. **(F)** Conversion from average log-loss to average accuracy. The grey line denotes a linear regression fit.

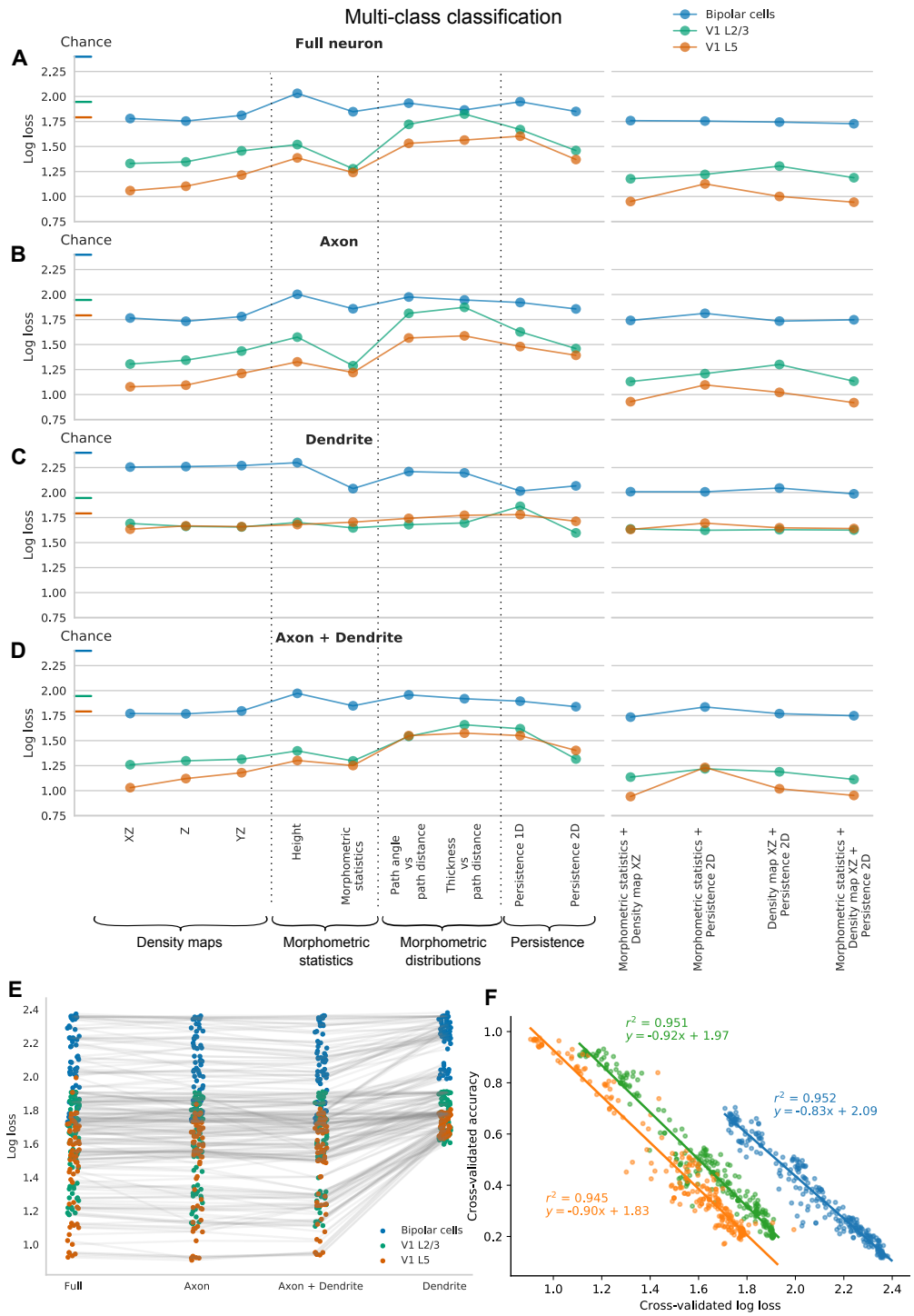

Figure S2: **(A–D)** Cross-validated log-loss of multinomial classification for different "modalities" (A–D) for each data set. Chance level for each data set is indicated on the y-axis. **(E)** Comparison of multi-class losses for each statistic in each modality and data set. Dendrites perform considerably worse than any of the other modalities. **(F)** Conversion from multi-class loss to accuracy for each data set. The colored lines denote a linear regression fit to the respective data set. V1 layer 5 (orange) is almost perfectly separable with accuracies close to 1 for some feature representations. This is also indicated by the emerging clusters in Figure 7.

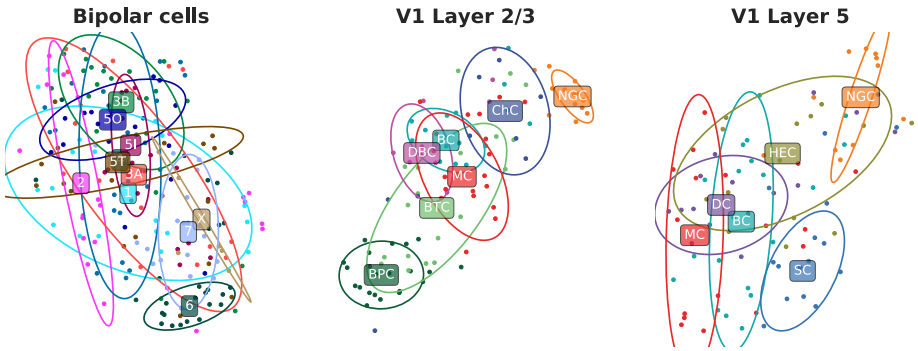

Figure S3: t-SNE embeddings of all three data sets (perplexity 30) for dendrites. We used a combined feature representation consisting of XZ density map, morphometric statistics and 2D persistence images. All three feature representations were reduced to 10 PCs and combined into a 30-dimensional feature set prior to t-SNE.

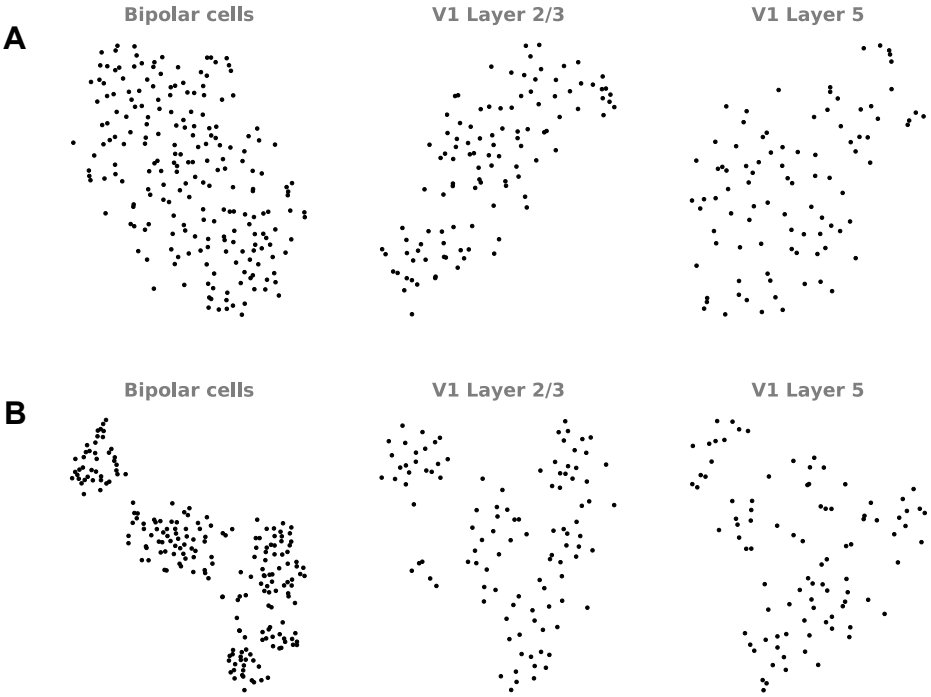

Figure S4: t-SNE embeddings of all three data sets (perplexity 30) for dendrites **(A)** and full neurons **(B)**. We used a combined feature representation consisting of XZ density map, morphometric statistics and 2D persistence images. All three feature representations were reduced to 10 PCs and combined into a 30-dimensional feature set prior to t-SNE. The embedding in (B) exhibits more structure than that obtained from purely dendritic features.