



A Systematic Evaluation of Interneuron Morphology Representations for Cell Type Discrimination

Sophie Laturnus^{1,4} · Dmitry Kobak¹ · Philipp Berens^{1,2,3,4}

Published online: 4 May 2020
© The Author(s) 2020

Abstract

Quantitative analysis of neuronal morphologies usually begins with choosing a particular feature representation in order to make individual morphologies amenable to standard statistics tools and machine learning algorithms. 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 effort, 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, two-dimensional persistence images and morphometric statistics, which continued to perform well even when neurons were only partially traced. Combining these feature representations together led to further performance increases suggesting that they captured non-redundant information. The same representations performed well in an unsupervised setting, implying that they can be suitable for dimensionality reduction or clustering.

Keywords Neuroanatomy · Benchmarking · Cell types · Mouse · Visual cortex

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)

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12021-020-09461-z>) contains supplementary material, which is available to authorized users.

✉ Philipp Berens
philipp.berens@uni-tuebingen.de

Sophie Laturnus
sophie.laturnus@uni-tuebingen.de

Dmitry Kobak
dmitry.kobak@uni-tuebingen.de

¹ Institute for Ophthalmic Research, University of Tübingen, Tübingen, Germany

² Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany

³ Bernstein Center for Computational Neuroscience, University of Tübingen, Tübingen, Germany

⁴ Center for Integrative Neuroscience, University of Tübingen, Tübingen, Germany

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; Xi 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 (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; Peng et al. 2014; Bria et al. 2016). As a consequence, more and more full reconstructions of neurons are becoming available in public databases, 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; Yanbin et al. 2013; Yanbin et al. 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 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 types of interneurons. We used carefully curated anatomical data from three studies, encompassing over 500 retinal and cortical interneurons with complete axonal and dendritic reconstructions and expert annotated cell type labels (Helmstaedter et al. 2013; Jiang et al. 2015; Scala et al. 2019). 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. By combining different representations together, we also studied to what extent they captured complementary information about cell morphologies. In addition, we investigated how robust these representations are if only parts of a neuron are reconstructed and how useful they remain in an unsupervised setting.

Results

Morphological Feature Representations

We analyzed the discriminability between different morphological cell types in adult mouse retina and adult mouse cortex (Fig. 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) (Fig. 1a). 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 based on

biocytin stainings (Jiang et al. 2015; Scala et al. 2019). We analyzed the neurons separated by layer (V1 L2/3: $n = 108$ neurons in 7 classes, Fig. 1b; V1 L4: $n = 92$ neurons in 7 classes, Fig. 1c; V1 L5: $n = 93$ neurons in 6 classes, Fig. 1d). All four 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 investigated 62 feature representations that we grouped into four different categories: density maps, morphometric statistics, morphometric distributions, and persistence images (Fig. 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 24 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 (Fig. 2c). We also considered a feature representation that joins all of them into a 24-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 (Fig. 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 (Fig. 2e). After binning, this becomes a 400-dimensional feature vector; Fig. 2f shows two principal components (PCs) across all bipolar cells of type 1 and 5O, indicating that PC1 discriminates the types very well.

Finally, we used persistence images, a recently introduced quantification of neural morphology based on topological ideas (Adams et al. 2017; Li et al. 2017; Kanari et al. 2018; Kanari et al. 2019). We used four different distance functions (also called filter functions) to construct one- and two-dimensional persistence images, resulting in eight different persistence representations. The same two bipolar cell types can be well discriminated based on PC1 of the two-dimensional radial-distance-based persistence images (Fig. 2g, h).

See “Methods” for a complete list and detailed definitions of the investigated feature representations.

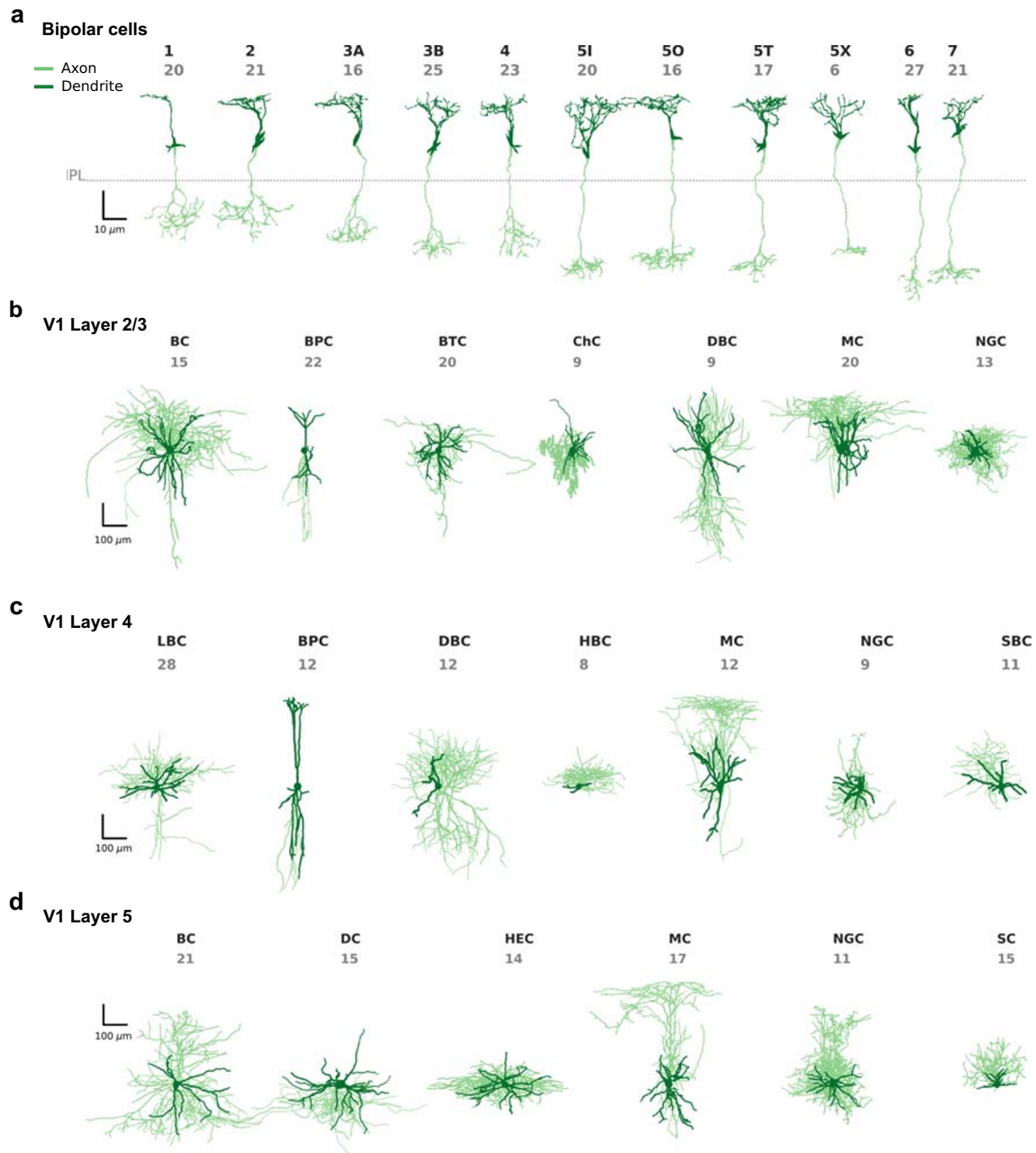


Fig. 1 Exemplary cells of each cell type for all four data sets. Axons are shown in light green, dendrites in dark green. **a** Mouse retinal bipolar cells (cone-connecting) from Helmstaedter et al. (2013). The dashed line shows the onset of the inner plexiform layer (IPL). The cell types used for analysis are types 1, 2, 3A, 3B, 4, 5I, 5O, 5T, 5X, 6, and 7. Cell types 8 and 9 were excluded from further analysis due to insufficient sample sizes. **b** Layer 2/3 inhibitory interneurons in primary visual cortex of adult mice (Jiang et al. 2015). BC: basket cells, BPC: bipolar cells, BTC: bitufted cells, ChC: chandelier

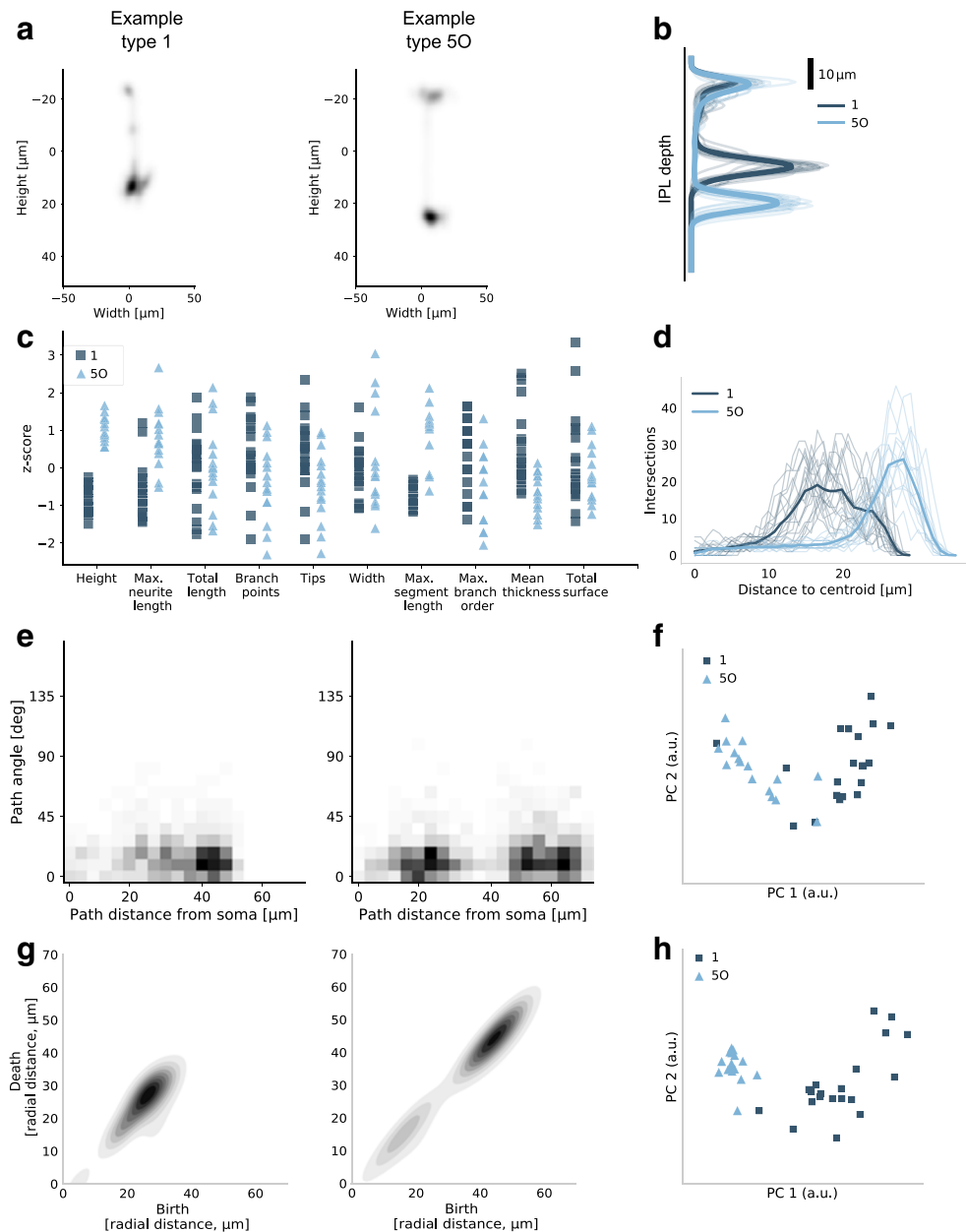
cells, DBC: double bouquet cells, MC: Martinotti cells, NGC: neurogliaform cells. **c** Layer 4 inhibitory interneurons in primary visual cortex of adult mice (Scala et al. 2019). LBC: large basket cells, BPC: bipolar cells, DBC: double bouquet cells, HBC: horizontal basket cells, MC: Martinotti cells, NGC: neurogliaform cells, SBC: small basket cells. **d** Layer 5 inhibitory interneurons in primary visual cortex of adult mice (Jiang et al. 2015). BC: basket cells, DC: deep-projecting cells, HEC: horizontally elongated cells, MC: Martinotti cells, NGC: neurogliaform cells, SC: shrub cells

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 Fig. 3).

Fig. 2 Selected feature representations for retinal bipolar cells of type 1 and type 50. **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 50 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



As an example, Fig. 4 shows the performance of one particular feature representation (XZ density map 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, 21 pairs in the V1 L4 data set, and 15 pairs in the V1 L5 data set (for morphometric statistics and 2D persistence, see Fig. S1). We used the cross-validated log-loss as the main measure of performance, because it is a proper scoring rule used by logistic regression, it is unaffected by class imbalance and it penalizes confident but wrong decisions. Zero loss means perfect classification, while chance-level performance (for balanced classes) corresponds to the loss of $\ln(2) \approx 0.69$. For each pair of types, we also computed cross-validated classification accuracy, F1 score, and Matthews correlation

coefficient. In our data, the relationships between log-loss and these other performance measures were monotonic and approximately quadratic (Fig. S2). As a rule of thumb, a log-loss of 0.2 roughly corresponded to 94% accuracy, a log-loss of 0.4 corresponded to 82% accuracy, and a log-loss of 0.6 corresponded to 67% accuracy.

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. types 1 and 2, or types 3A and 3B) are hard to distinguish (Fig. 1), and indeed the log-loss values were generally higher close to the diagonal than far away from it (Fig. 4). In fact, to distinguish between bipolar types 1/2, 3A/3B/4, and 5I/5O/5T, the original studies used

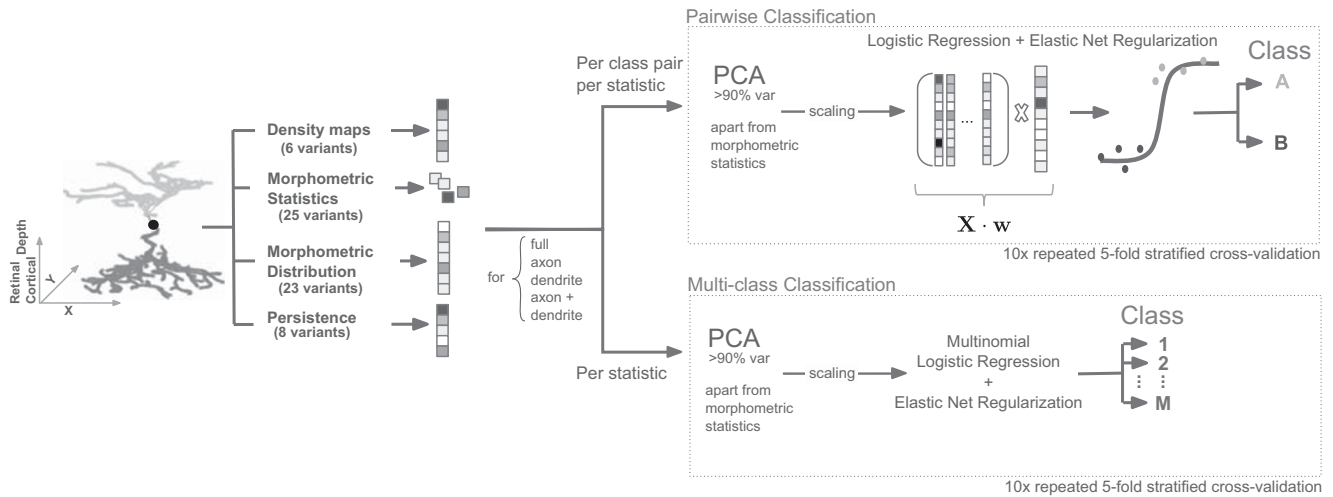


Fig. 3 Processing pipeline. Inhibitory interneurons were soma-centered. Retinal bipolar cells were soma-centered in x and y while $z = 0$ was chosen to correspond to the inner plexiform layer (IPL) onset. The z direction of each cell was aligned with cortical/retinal depth, whereas the x and y direction were left

tiling of the retina and synaptic input patterns in addition to the morphological information (Helmstaedter et al. 2013; Behrens et al. 2016).

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 $4 \times 3 = 12$ average log-losses for each of the 62 feature representations. Figure 5 shows a summary for the seven top performing features (see “Methods” for how they were selected). The performance using the dendritic features was consistently poor for the bipolar cells and the V1 L5 interneurons (close to chance level) and generally much lower than using the axonal features (see also Fig. S3). 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 (Markram et al. 2004; Ascoli et al. 2008; Helmstaedter et al. 2013; DeFelipe et al. 2013; Sümbül et al. 2014), as can be seen in

unchanged. Several different feature representations were extracted automatically and used for pairwise and multi-class classifications using logistic regression regularized with elastic net. The performance was assessed using 10 times repeated 5-fold stratified cross-validation

Fig. 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 about 86% of the total traced neuritic length (4.55 m out of 5.27 m). 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 (Fig. S3a).

Using the performance of the full-neuron features (Fig. 5a), we found that the top performing feature representation were XZ density maps, with a mean log-loss of 0.18 ± 0.08 (mean±SD across $n = 4$ datasets), followed by Z density maps (0.19 ± 0.05), 2D persistence images constructed using z -projection as a distance function (0.22 ± 0.05), morphometric statistics (0.23 ± 0.09), and YZ density maps (0.25 ± 0.07). The 2D persistence images that were based on other distance functions performed

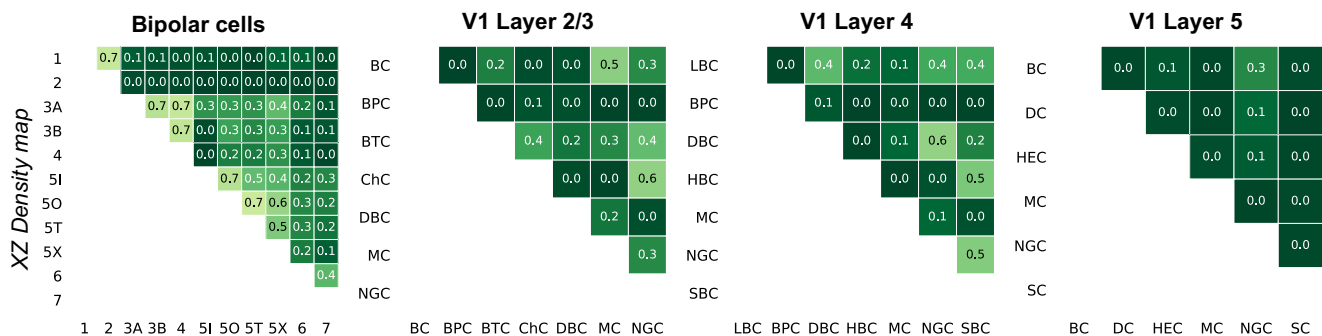


Fig. 4 Cross-validated log-loss for each pair of morphological types in each data set using XZ density maps on full neurons as predictors in logistic regression. Zero log-loss corresponds to perfect prediction,

$\ln(2) \approx 0.69$ corresponds to random guessing. For the classification results of other feature representations see Fig. S1 and <https://doi.org/10.5281/zenodo.3716519>. For abbreviations see Fig. 1

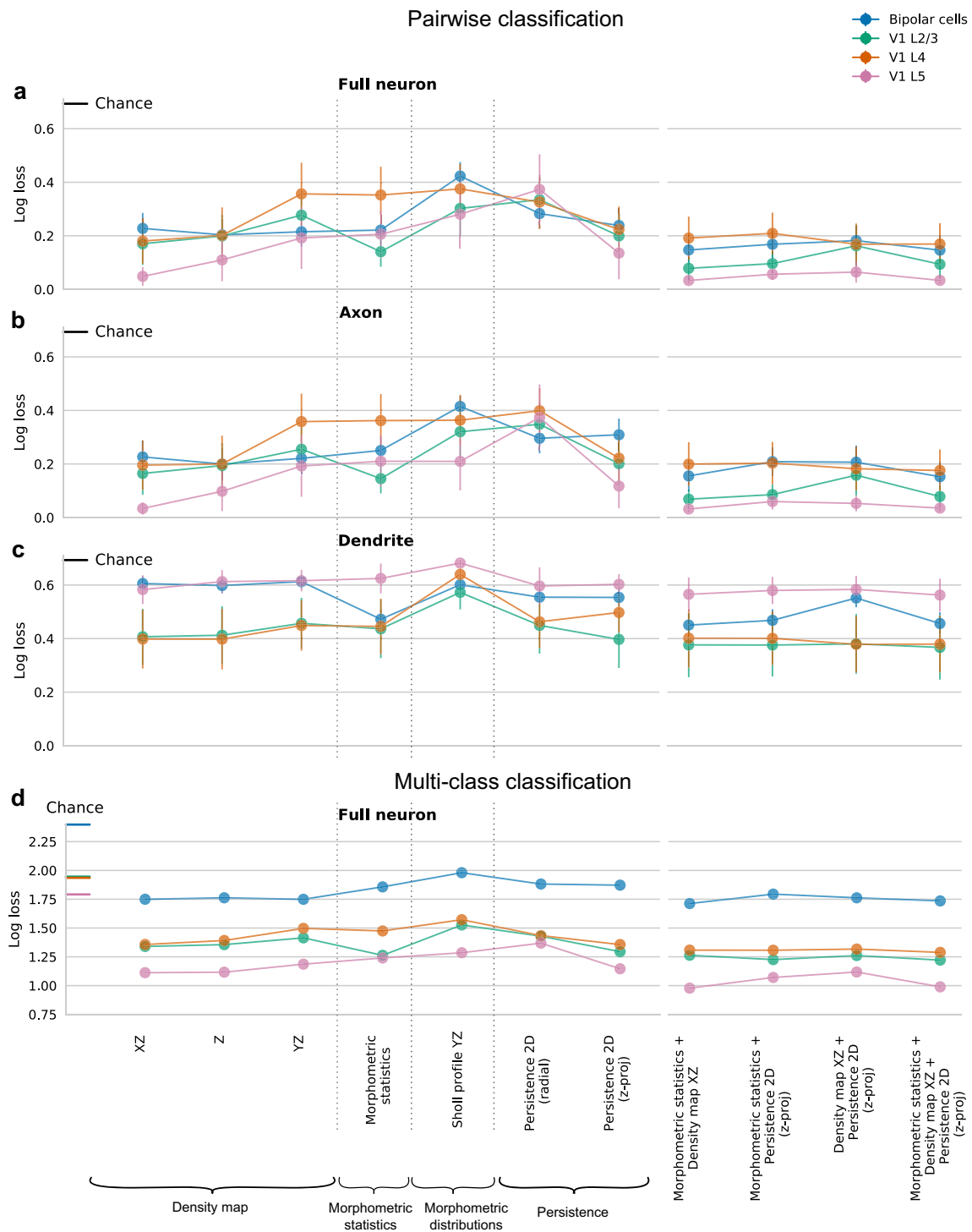


Fig. 5 **a–c** Pairwise classification performance of the top performing feature representations based on the full-neuron (a), axonal (b), and dendritic (c) features for each data set. Feature representations are grouped into density maps, morphometric statistics, morphometric distributions, persistence 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 Fig. S3a 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 Fig. S3b for the results using axonal, dendritic, and combined axonal+dendritic feature representations

considerably worse (radial distance: 0.31 ± 0.04 , path length: 0.46 ± 0.09 , branch order: 0.55 ± 0.03). The best performing morphometric distribution was the YZ Sholl intersection profile (0.37 ± 0.07).

To make a statistical comparison of the performance between two different feature representations A and B , we computed the mean difference $\delta(A, B)$ in log-loss across all 112 pairs of neural types (pooling pairs across the four 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 difference in the performance between XZ density maps and morphometric statistics ($\delta = 0.05 \pm 0.05$, $z = 0.92$, $p = 0.36$, z -test), between morphometric statistics and 2D z -projection-based persistence ($\delta = 0.08 \pm 0.17$, $z = 0.5$, $p = 0.62$), or between 2D z -projection-based persistence and XZ density maps ($\delta = 0.03 \pm 0.03$, $z = 1.16$, $p = 0.24$).

Among the density maps, the Z density maps did not perform differently from XZ density maps ($\delta = 0.006 \pm 0.02$, $z = 0.37$, $p = 0.71$), while YZ density maps performed much worse in the three V1 data sets ($\delta = 0.14 \pm 0.05$, $z = 2.84$, $p = 0.005$, average across V1 pairs only) but very similar in the bipolar data set ($\delta = 0.01 \pm 0.02$, $z = 0.77$, $p = 0.44$). Indeed, the y direction is mostly meaningless in the V1 data as the slices are flattened during the biocytin staining process (Farhoodi et al. 2019).

Next, we asked if combining feature representations can improve the performance. We pooled morphometric statistics and XZ density maps, morphometric statistics and z -projection-based 2D persistence, XZ density maps and z -projection-based 2D persistence, and all three of these feature sets, yielding four additional combined feature sets (Fig. 5, right). We found that combinations of features tended to outperform individual feature sets. For example, XZ density maps combined with morphometric statistics outperformed both XZ density maps and morphometric statistics on their own ($\delta = 0.06 \pm 0.02$, $z = 2.35$, $p = 0.02$; $\delta = 0.1 \pm 0.03$, $z = 3.57$, $p = 0.0004$). This suggests that these feature representations, despite their similar individual performance, contained some non-redundant information. Combining all three feature representations together yielded further improvements in some cases (compared to pairwise combinations: $p = 0.01$ when adding morphometric statistics to XZ density maps and persistence, $p = 0.004$ for XZ density maps, $p = 0.9$ for persistence), but the differences were small ($\delta < 0.04$ for the three comparisons).

Top Performing Features are Consistent Across Classification Schemes

As an alternative to the pairwise classification approach, we also used multi-class classification. We used multinomial logistic regression with exactly the same pipeline of regularization and cross-validation as above. For each of the full-neuron feature representations and each of the data sets, we obtained the cross-validated multi-class log-loss (Fig. 5d; see Fig. S3b for the axonal, dendritic, and combined axonal+dendritic feature representations and Fig. S4 for conversion to accuracy). 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 Fig. 5d. Note also that 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 and L4 data sets, 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 (Fig. 5d). For three out of four data sets (bipolar, V1 L4 and V1 L5), XZ density maps performed the best (bipolar data set log-loss: 1.75, V1 L4: 1.36, V1 L5: 1.11), followed by z -projection-based 2D persistence images (1.87/1.36/1.15) and morphometric statistics (1.86/1.47/1.24). For the V1 L2/3 data set, morphometric statistics showed the smallest loss (1.26), very closely followed by the persistence images (1.3) and the density maps (XZ: 1.34/ Z: 1.36). *Height* was one of the most relevant morphometric statistics across all data sets (Fig. S6). Combining morphometric statistics with XZ density maps led to a clear improvement in all cortical data sets and was on par with further adding 2D persistence. For the bipolar data, combining features did not improve performance compared to XZ density maps alone.

To make sure that our conclusions were not dependent on the choice of the classification approach or the performance metric, we repeated the experiments using two other pairwise classifiers: k -nearest neighbour (with $k = 3$) and decision trees. In each case, we used the classification accuracy, F1 score and Matthews correlation coefficient (MCC) as performance metrics (note that log-loss is not meaningful for these classifiers). The selection of top performing feature representations was very consistent, with XZ and Z density maps always ranked the first (Fig. 6).

Fig. 6 Ranked top five feature representations for each classification scheme using different performance measures on full-neuron data. All measures and all classification schemes selected the same top-5 features

| | LogitNet | | | | 3-NN | | | Decision Tree | | |
|-----------------------|----------|----------|----|-----|----------|----|-----|---------------|----|-----|
| DM xz | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| DM z | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| DM yz | | | | | | | | | | |
| Persistence 2D radial | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 |
| Persistence 2D path | | | | | | | | | | |
| Persistence 2D z-proj | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| Morphometrics | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 |
| | Log loss | Accuracy | F1 | MCC | Accuracy | F1 | MCC | Accuracy | F1 | MCC |

The Best Feature Representations are Robust Against Partial Tracings

Accurate morphological reconstructions often become more and more difficult to achieve 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 (z-projection) 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, and 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 (Fig. 7). In addition, we shuffled the labels of each pairwise comparison to estimate the chance-level distribution of log-losses (Fig. 7a, 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 0.14 ± 0.06 to 0.17 ± 0.07 , mean \pm 95CI across all 21 pairs, for morphometric statistics,

and very similarly for the density map and the 2D persistence). After that, all representations were noticeably losing in performance.

Using Morphological Features for Unsupervised Learning

So far we used supervised learning and assumed cell type labels 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. 2019). 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 the best performing feature representations identified above 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 first used XZ density maps (reduced to a set of 6–18 PCs capturing 90% of the variance) as an input to t-SNE with perplexity 50 (Fig. 8a). The resulting embeddings corresponded well to the pairwise classification performance for XZ density maps that we presented earlier (Fig. 4). For example, horizontally elongated cells (HECs) and shrub cells (SCs) in the V1 L5 data set that were both easily distinguishable from other types in the classification task, formed clear clusters, away from other cell types. In contrast, the embedding for basket cells (BCs) and neurogliaform cells (NGCs), the only cell pair with a high log-loss for density maps, showed some overlap. Similarly,

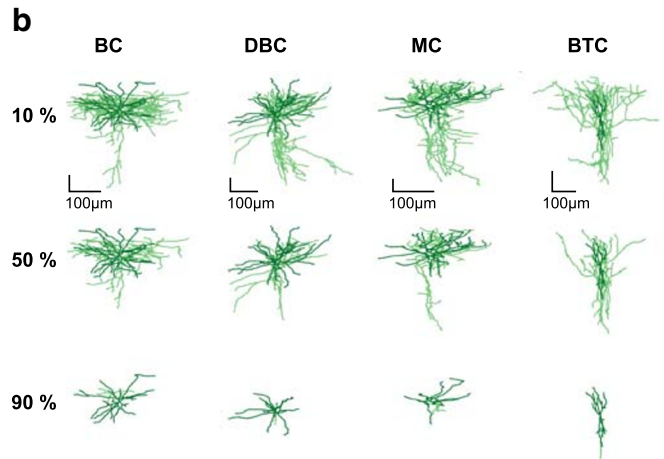
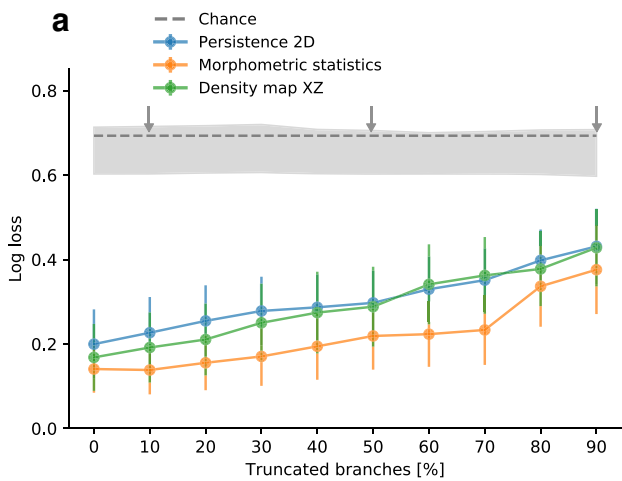


Fig. 7 a Cross-validated log-loss of XZ density maps, morphometric statistics and z -projection-based 2D persistence as a function of truncation level. Branches were truncated to mimic what happens when neurons are only partially traced. The classification was performed on all pairs of types in V1 L2/3 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 during the cross-validation (shading intervals go from the minimum to the maximum obtained chance-level values). 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 Fig. 1 for abbreviations

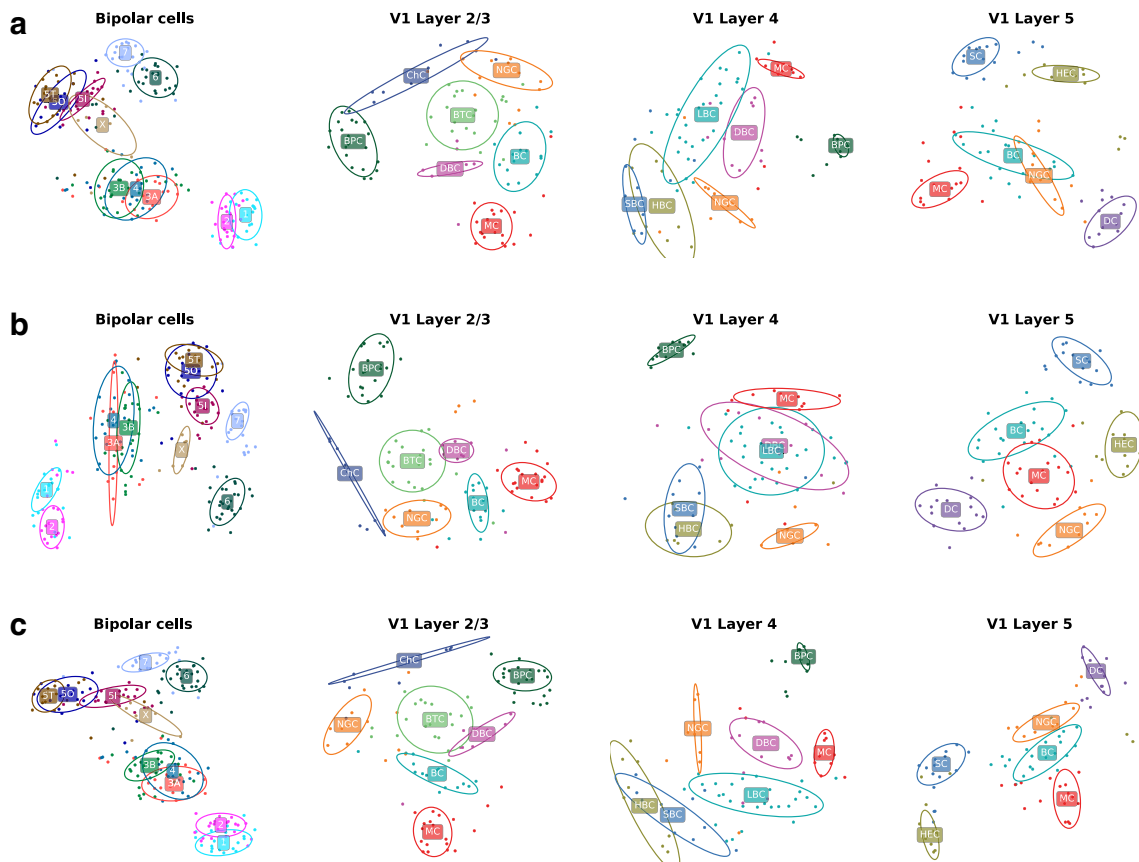


Fig. 8 a T-SNE embeddings of all four data sets using embeddings using the XZ density maps combined with morphometric statistics. **b** T-SNE embeddings using the XZ density maps combined with morphometric statistics. **c** T-SNE embeddings using XZ density maps

combined with z -projection-based 2D persistence images. 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 Fig. 1

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 (Fig. 8a).

We then combined XZ density maps with either morphometric statistics (11–12 PCs) or with the best performing 2D persistence image (z -projection; 4–6 PCs). To do so, we reduced each feature representation to a set of PCs capturing 90% of the variance and normalized each set of PCs by the standard deviation of the respective PC1, to put both sets roughly on the same scale. We pooled the scaled PCs together and used this feature representation as input for t-SNE (Fig. 8b and c). For some of the data sets (e.g. V1 L2/3 and V1 L5), the combination of XZ density maps with one of the two other feature representations yielded an arguably superior t-SNE embeddings with less overlap between types.

The embeddings shown in Fig. 8 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 (Fig. S5). Moreover, dendritic features resulted in t-SNE embeddings with far less structure than the full-neuron features, suggesting that there is less of an interesting variability in the dendritic morphologies compared to the axonal ones.

Discussion

Here we benchmarked existing morphological representations in the context of supervised cell type classification on well-curated data sets encompassing over 500 full reconstructions of interneurons in the mouse visual system. We found that density maps, z -projection-based 2D persistence images 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 these predictors work well independent of the used classification scheme or the performance metric suggesting that they are a good starting point for morphological analysis.

Previous literature

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 (Yanbin et al. 2013;

Polavaram et al. 2014; Yanbin et al. 2015; Gouwens et al. 2019) and have been shown to perform well in a one-vs-rest classification of cortical neurons (Mihaljević et al. 2015; Mihaljević et al. 2018). Persistence, in turn, has lately been shown to distinguish pyramidal neuron types in juvenile rat somatosensory cortex (Kanari et al. 2019).

These studies, however, did not directly compare 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 density maps contain enough information to accurately discriminate most inhibitory cell types. This implies that the spatial extent and overall shape of the axonal arbour, as a consequence of a neuron's connectivity, are more relevant than precise branching characteristics; a finding that has already been proposed for dendrites (Cuntz et al. 2007; Cuntz et al. 2008; Cuntz et al. 2010; Cuntz 2012; van Pelt and van Ooyen 2013).

Various forms of persistence-based measures performed consistently worse than density maps, except when the z -projection was used as a distance measure. This distance measure captures a neuron's orientation towards the pia and its relative spatial extent across layers (Li et al. 2017; Kanari et al. 2018), again supporting the point that spatial extent is most meaningful for inhibitory cell type classification. Morphometric statistics performed similarly well on some of the data sets but showed a somewhat lower performance than density maps. For example, they failed to distinguish double-bouquet cells and Martinotti cells in layer 4 (see Fig. S1). We did not extensively evaluate combinations of feature representations, but combining the best representations often lead to small but significant improvements in performance. However, strong regularization has to be implemented to avoid overfitting to a data set of limited size (see Fig. 5).

The axonal morphology of neurons in our study contained more information about the cell type than the dendritic morphology, in agreement with the existing literature (Mihaljević et al. 2015; Jiang et al. 2015; Ofer et al. 2018). Our unsupervised analysis also demonstrated far more variability within axonal features compared to the dendritic features (Fig. S5), which is in line with classical expert-based cell type naming conventions (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% (630/8129) 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 robust classification performance of truncated morphologies and the good performance of density maps suggest that full reconstructions might not be necessary for morphological cell type identification.

Limitations

Our study has several limitations. First, we always used cell type labels provided in the original publications, treating them as ground truth. For cortical interneurons it has been shown that there is a considerable inter-expert variability between assigned cell type labels (DeFelipe et al. 2013), affecting the outcome of any supervised learning task. Ideally, one would use consensus labels between multiple experts or different data modalities for a benchmark evaluation, but such data sets are currently even harder to obtain than the data sets used in this study.

Next, tissue shrinkage and staining method can affect the measured morphology (Farhoodi et al. 2019). In our study all three data sets obtained through biocytin staining (V1 L2/3, V1 L4, V1 L5) showed flattening of the cortical slice (*y*-direction) which made XY density maps perform worse in comparison to other projections. We did not observe this effect for the bipolar cells that have been obtained through EM imaging. Obviously, a feature representation can only be good for classification, if the data contains the relevant information in the first place. Therefore, it is important to be aware of biases or distortions in the experimental protocol before deciding on which feature representation to use.

Further, we believe that a meaningful comparison between different numerical descriptions of morphology is only possible through 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. The resulting data set of 505 interneurons is comparable to the sample sizes used in related studies (Mihaljević et al. 2015; Mihaljević et al. 2018). As these interneurons are only locally projecting, our study does not provide guidance as to

which features are useful for discriminating neurons based on their long-range projection patterns (Costa et al. 2016; Econom et al. 2018; Gerfen et al. 2018).

As the 505 neurons were split into 31 types, the median sample size per class was only 16 cells. It is difficult to fit machine learning algorithms 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 well as nonlinear non parametric models since fitting more complicated models can be challenging with these 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.

Finally, we restricted our benchmarking effort to the most prominent and 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 (Panico and Sterling 1995) and did not explicitly quantify the amount of layer-specific arborization (DeFelipe et al. 2013; Gouwens et al. 2019) because this concept only applies to cortical neurons and layer boundaries were not available for our data. However, given the good performance of density maps and *z*-projection-based 2D persistence images, it is possible that including layer-specific information could improve the performance of morphometric statistics.

Outlook

Our study serves to provide a starting point for future work on algorithmic cell type discrimination based on anatomical data, for example in the context of large-scale efforts to map every cell type in the brain as pursued by the NIH BRAIN initiative. It allows experimenters to make an informed choice which cell type representations are useful to automatically distinguish interneurons based on their morphology. Of course, how far our results generalize to other species and brain regions remains to be seen. The resulting representations also make it possible to relate

anatomical descriptions of neurons to data from other modalities such as e.g. gene expression patterns (Cadwell et al. 2017).

The representations investigated here are purely descriptive and do not provide deeper mechanistic insight, compared e.g. to generative models of the growth process of neurons during development (van Pelt and Schierwagen 2004; Cuntz et al. 2010; Memelli et al. 2013; Wolf et al. 2013; Fard et al. 2018; Farhoodi et al. 2019). Ideally, a mechanistically grounded feature representation would perform at least on par with the representations identified here for cell type discrimination while yielding parameters that are more easily interpretable. Potential starting points for such a representations are growth models proposed by van Pelt and Schierwagen (2004) and Cuntz et al. (2010), which have a manageable amount of parameters and show systematic parameter differences for dendrites of different cell types. This make them promising candidates for further research.

Methods

Data

We used data from Helmstaedter et al. (2013), Scala et al. (2019), and Jiang et al. (2015), splitting the latter data set into two parts by cortical layer. All neurons were labelled by human experts in the original studies. We confirmed the quality of all reconstructions through inspection. Our study investigated a total of 5.27 meters of traced neurites from $n = 505$ 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) of adult mouse primary visual cortex (Jiang et al. 2015). We obtained the reconstructions (as .ASC files) and their cell type labels from the authors.
3. **V1 Layer 4.** Manually traced biocytin stainings of $n = 92$ inhibitory interneurons of 7 types in layer 4 (L4) of adult mouse primary visual cortex Scala et al. (2019).

We obtained the reconstructions (as .ASC files) and their cell type labels from the authors.

4. **V1 Layer 5.** Manually traced biocytin stainings of $n = 94$ inhibitory interneurons of 6 types in layer 5 (L5) of 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 files) and their cell type labels from the authors.

For data availability see [Information Sharing Statement](#).

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 node; 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 centroid of the convex hull of 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 were hence assigned according to the majority vote over all sub-segment types within this branch.

All cortical interneurons were soma-centered and their z -coordinate (height) was oriented along the cortical depth. We re-sampled all neurons to consist of equidistant points with 1 micron spacing along all neurites through linear interpolation. As the y -coordinate (oriented in the direction along the thickness of the cortical slice, i.e. the viewing direction of the microscope) sometimes contained sudden “jumps”, we used a 3rd order Savitzky-Golay filter with window size of 21 microns to smooth these (as implemented in `scipy.signal`). We did not account for the shrinkage of slice thickness (depth) that happens during the staining. 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). Smoothing and re-sampling was not done for this data set.

Feature Representations

We calculated 62 different feature representations for each cell. These representations can be grouped into four categories: *density maps*, *morphometric statistics*, *morphometric distributions*, and *persistence*. All feature representations were separately computed for axons, for dendrites, and for the whole neuron (without distinguishing axons from dendrites) yielding $62 \cdot 3 = 186$ representations per neuron (see Fig. 2).

Density Maps

We sampled equidistant points with 25 nm spacing along each neurite of the traced skeletons and normalized the resulting point clouds within each data set for each modality to lie between 0 and 1 (for the range values used for this normalisation see the linked Github repository). 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.

Morphometric Statistics

For each cell we computed a set of 24 single-valued summary statistics:

- number of branch points** Count of points at which the neurites branch.
- number of tips** Count of end-points.
- cell height** Extent (max–min) of the cell in the z direction in microns. This direction corresponds to the cortical/retinal depth.
- cell width** Extent (max–min) of the cell in the x direction in microns. In biocytin data this direction corresponds to the width of the slice.
- cell depth** Extent (max–min) of the cell in the y direction in microns. In biocytin data this direction corresponds to the depth of the slice and is flattened due to the staining process.

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

average thickness The average radius across all neurites in microns (soma is excluded).

total length Total path length of all neurites in microns.

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.

maximal neurite length Path length of the longest neurite from tip to soma in microns.

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. This ordering scheme is also called *centrifugal* order.

maximum segment Euclidean length of the longest segment in microns.

median intermediate segment The median path length of intermediate segments in microns.

median terminal segment The median path length of terminal segments in microns.

median path angle “Path angle” is the angle between two consecutive sub-segments (not including the sub-segments that meet at a branching point). Median path angle refers to the median across all such consecutive sub-segments. It is bounded between $[0, 180]$ degree.

maximal path angle Maximal path angle represents the 99.5 percentile across all such consecutive sub-segments. We used the percentile to ameliorate the influence of reconstruction quality on this measure. It is bounded between $[0, 180]$ degree.

median tortuosity The median $\log(\text{tortuosity})$ 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. Tortuosity follows a skewed distribution, we therefore used the log-transformed value.

maximal tortuosity The 99.5 percentile of $\log(\text{tortuosity})$ across all neurite segments. We used the percentile to ameliorate the influence of reconstruction quality on this measure.

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.

average branch angle Average branch angle, see above.
maximal branch angle Maximal branch angle, see above.
maximal degree Maximal number of neurites meeting at a single branch point.
tree asymmetry The tree asymmetry measures how far a tree-graph is away from a perfectly balanced tree-graph. As a measure we use the weighted sum over the proportional sums of absolute deviations (PSADs) of each branch point:

$$\sum w_p \cdot \text{PSAD}(p)$$

where the sum is over all branch points p , $w_p \in \{0, 1\}$, and $w_p = 1$ iff the sub-tree emerging from branch point p has more than 3 leaves. The PSAD is a measure of topological tree asymmetry and is defined for one branching node p as

$$\text{PSAD}(p) = \frac{m}{2(m-1)(n-m)} \cdot \sum_m |r_i - \frac{n}{m}|$$

where m is the out-degree of node p , n is the number of leaves of the sub-tree starting at p , and r_i is the number of leaves of the i -th sub-tree of p . For a more detailed definition see Verwer and van Pelt (1986).

In addition to the 24 features listed above, we grouped all of them together into one vector which we named *morphometric statistics*.

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. For more details see Bird and Cuntz (2019).
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 Euclidean segment lengths (20 bins from 0 to the maximal segment length within each data set).

thickness Histogram of nodes' radii, the soma is excluded (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 each data set)

Euclidean distances to soma Histogram of the Euclidean distance of each branch point and tip to the soma (20 bins from 0 to maximal path length within each data set).

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 radii around the soma (Sholl 1953). We used 36 steps from soma to maximal radial distance from soma.

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 Euclidean distance 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 \times branch orders 2D histogram of branch angles as a function of branch orders (across all branch points).

branch angles \times path distances 2D histogram of branch angles as a function of path distances to the soma in microns (across all branch points).

path angle \times branch order 2D histogram of path angles as a function of branch orders (across all nodes).

path angle \times path distance 2D histogram of path angles as a function of path distances to the soma in microns (across all nodes).

thickness \times branch order 2D histogram of neurite radii as a function of branch orders (across all nodes).

thickness \times path distance 2D histogram of neurite radii as a function of path distances to the soma in microns (across all nodes).

Persistence Diagrams

Persistence diagrams originated from algebraic topology but recently have been proposed as a representation for neural morphologies (Kanari et al. 2018). We briefly outline the underlying algorithm here. Starting from each tip, one records the “birth time” of each branch as the distance of the tip from the soma. Hereby, the distance is measured according to some filter function f . 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 as the distance of the branch point from the soma. This results in a 2D point cloud of (*birth time*, *death time*) for each branch, the so called *persistence diagram*, in which only the longest branch survived until the soma and has a death time of 0. Depending on the filter function f , different aspects of the neuron’s topology can be captured. Here we employed four different filter functions, following (Kanari et al. 2018):

radial distance returns the Euclidean distance of node x to the soma s ;

path length returns the sum over the lengths of all segments along the path connecting node x with the soma s ;

branch order returns the branch order of node x ;

z -projection returns the difference of z -coordinates between node x and the soma s projected onto the z -axis.

Euclidean distance is undefined for persistence diagrams themselves. To circumvent this problem we converted each persistence diagram into a 1D or 2D image and used Euclidean distance on the results as this procedure has been shown to work well in the neural domain (Kanari et al. 2019; Li et al. 2017). To obtain a 2D Gaussian persistence image we performed kernel density estimation of the 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_{birth} and \max_{death} refer to the maxima across all cells within each data set (for actual values see the linked Github repository). For the purposes of downstream analysis, we treated this as a set of 10,000 features.

The 1D Gaussian persistence vector we 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 $[0, \max_{\text{birth}}]$.

We compared our implementations of morphometric statistics and persistence diagrams to the existing implementations (NeuronM: <https://github.com/BlueBrain/NeuroM>, TMD: <https://github.com/BlueBrain/TMD>) and made sure that they are in full agreement. Note that our implementation defines the soma at its centre (not its surface) which can slightly affect some of the features.

Classification

Each feature representation was used as a predictor for pairwise and multinomial classification. Morphometric statistics were z -scored. Except for morphometric statistics, we reduced all representations using principal component analysis (PCA) and kept as many principal components as needed to capture at least 90% of the variance on the training set (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). We divided all PCA components by the standard deviation of the respective first principle component to put all features roughly on the same scale and to allow for combination of features. To avoid overfitting when using combinations of feature representations, we performed the same PCA reduction on morphometric statistics as well.

For binary classification, we used logistic regression with an elastic net regularization. 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, standardize=False)
m.scoring = make_scorer(log_loss,
                       greater_is_better=False, needs_proba=True)
```

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 comparison between different classification schemes we also computed the mean test-set macro F1 scores, which is the unweighted mean of the F1 scores for each class, and the mean test-set Matthews correlation coefficient (Brian 1975).

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.

The processing pipeline including preprocessing, feature extraction and classification was automated using `DataJoint` (Yatsenko et al. 2015).

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 identified their superset (six features). We also included the best performing morphometric distribution to have at least one feature representation of each category investigated. This led to the seven features shown in Fig. 5.

Statistical Analysis of Differences

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

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

where \mathcal{P} denotes the set of pairs of types across the four data sets, $|\mathcal{P}| = 112$ 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:

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

where $\mathcal{P}_{-\tau}$ is the set of pairs without type τ . This yields $n = 31$ estimates of $\delta_{-\tau}$, with the jackknife estimate of the standard error given by

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

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

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

Other Classification Schemes

For comparison with different classification schemes we used the 3-nearest neighbour classifier and the decision tree classifier of the Python `scikit-learn` implementation:

```
from sklearn.neighbors import
    KNeighborsClassifier
from sklearn.tree import
    DecisionTreeClassifier

m_neighbour = KNeighborsClassifier
    (n_neighbours=3)
m_tree = DecisionTreeClassifier
    (random_state=17)
```

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 as many principal components needed to keep 90% of the variance each. As done during classification, we scaled each set of PCs by the standard deviation of the respective PC1, to put both sets roughly on the same scale. Then we stacked them together to obtain a combined representation of each cell. Exact (non-approximate) t-SNE was run with perplexity 50 and random initialization using the `scikit-learn` implementation:

```
TSNE(perplexity=50, method='exact',
    random_state=42)
```

The same procedure was applied for embedding the combination of density maps and z -projection-based 2D persistence images.

To plot the coverage ellipses for each cell type (Fig. 8), 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`.

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. 7). 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 (z -projection) for each degree 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 designed the study. SL performed the analysis with supervision from DK and PB. All authors wrote the manuscript.

Information Sharing Statement

Our analysis code and notebooks reproducing our figures are available at: <http://github.com/berenslab/morphology-benchmark>. All preprocessed morphological reconstructions used in this study are available at Zenodo at <https://doi.org/10.5281/zenodo.3716519>. Most of the original reconstructions of the V1 interneurons of layer 2/3 and layer 5 are available at <http://neuromorpho.org> (archive TOLIAS), but the dataset used in this paper differs in several cells. The original reconstructions of V1 layer 4 interneurons are also part of the data deposited to Zenodo at <https://doi.org/10.5281/zenodo.3336165>.

The pairwise and multi-class classification results of all feature representations, modalities, performance measures and all classifiers as well as all necessary data to reproduce our figures are deposited as csv files to Zenodo at <https://doi.org/10.5281/zenodo.3716519>.

Acknowledgments Open Access funding provided by Projekt DEAL. We thank Ulrike von Luxburg, Debarghya Ghoshdastidar, Xiaolong Jiang, Federico Scala and Andreas Tolias for discussions. This work was funded by the German Ministry of Education and Research (FKZ 01GQ1601), the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2064/1 – 390727645; BE5601/4-1, SFB 1233 “Robust Vision”, Project number 276693517), and the National Institute of Mental Health (U19MH114830). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L. (2017). Persistence images: a stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1), 218–252.
- Ascoli, G.A., Alonso-Nanclares, L., Anderson, S.A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsáki, G., Cauli, B., DeFelipe, J., Fairén, A., et al. (2008). Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, 9(7), 557.
- Ascoli, G.A., Donohue, D.E., Halavi, M. (2007). Neuromorpho.org: a central resource for neuronal morphologies. *Journal of Neuroscience*, 27(35), 9247–9251.
- Baden, T., Berens, P., Franke, K., Rosón, M., Iroslav, R., Omán, B., Bethge, M., Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature*, 529(7586), 345.
- Behrens, C., Timm, S., Haverkamp, S., Thomas, E., Philipp, B. (2016). Connectivity map of bipolar cells and photoreceptors in the mouse retina. *eLife*, 5, e20041.
- Bird, A.D., & Cuntz, H. (2019). Dissecting sholl analysis into its functional components. *Cell Reports*, 27(10), 3081–3096.
- Bria, A., Iannello, G., Onofri, L., Peng, H. (2016). Terafly: real-time three-dimensional visualization and annotation of terabytes of multidimensional volumetric images. *Nature Methods*, 13(3), 192.
- Brian, W. (1975). Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2), 442–451.
- Briggman, K.L., Helmstaedter, M., Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337), 183.
- Cadwell, C.R., Scala, F., Li, S., Livrizzi, G., Shen, S., Sandberg, R., Jiang, X., Tolias, A.S. (2017). Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using patch-seq. *Nature Protocols*, 12(12), 2531.
- Cajal, S.R. (1899). *Textura del sistema nervioso del hombre y de los vertebrado: estudios sobre el plan estructural y composición histológica de los centros nerviosos adicionados de consideraciones fisiológicas fundadas en los nuevos descubrimientos*. Moya.
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 1.
- Costa, M., Manton, J.D., Ostrovsky, A.D., Prohaska, S., Jefferis, G.S.X.E. (2016). NBLAST: Rapid, Sensitive comparison of neuronal structure and construction of neuron family databases. *Neuron*, 91(2), 293–311.
- Cuntz, H. (2012). The dendritic density field of a cortical pyramidal cell. *Frontiers in Neuroanatomy*, 6, 2.
- Cuntz, H., Borst, A., Segev, I. (2007). Optimization principles of dendritic structure. *Theoretical Biology and Medical Modelling*, 4(1), 21.

- Cuntz, H., Forstner, F., Borst, A., Häusser, M. (2010). One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Computational Biology*, 6(8), e1000877.
- Cuntz, H., Forstner, F., Haag, J., Borst, A. (2008). The morphological identity of insect dendrites. *PLoS Computational Biology*, 4(12), e1000251.
- DeFelipe, J., López-Cruz, P.L., Benavides-Piccione, R., Bielza, C., Larrañaga, P., Anderson, S., Burkhalter, A., Cauli, B., Fairén, A., Feldmeyer, D., et al. (2013). New insights into the classification and nomenclature of cortical gabaergic interneurons. *Nature Reviews Neuroscience*, 14(3), 202.
- Econom, M.N., Viswanathan, S., Tasic, B., Siljka Yao, Bas, E., Winnubst, J., Menon, V., Graybiel, L.T., Nguyen, T.N., Smith, K.A., Yao, Z., et al. (2018). Distinct descending motor cortex pathways and their roles in movement. *Nature*, 563(7729), 79.
- Econom, M.N., Clack, N.G., Lavis, L.D., Gerfen, C.R., Svoboda, K., Myers, E.W., Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual neurons. *eLife*, e10566, 5.
- Efron, B., & Hastie, T. (2016). *Computer age statistical inference* Vol. 5. Cambridge: Cambridge University Press.
- Fard, P.K., Pfeiffer, M., Bauer, R. (2018). A generative growth model for thalamocortical axonal branching in primary visual cortex. bioRxiv.
- Farhoodi, R., Lansdell, B.J., Kording, K.P. (2019). Quantifying how staining methods bias measurements of neuron morphologies. *Frontiers in Neuroinformatics*, 13, 36.
- Friedman, J., Hastie, T., Tibshirani, R. (2001). The elements of statistical learning, volume 1 Springer series in statistics New York.
- Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1.
- Gerfen, C.R., Econom, M.N., Chandrashekar, J. (2018). Long distance projections of cortical pyramidal neurons. *Journal of Neuroscience Research*, 96(9), 1467–1475.
- Gillette, T.A., & Ascoli, G.A. (2015). Topological characterization of neuronal arbor morphology via sequence representation: I-motif analysis. *BMC Bioinformatics*, 16(1), 216.
- Gillette, T.A., & Grefenstette, J.J. (2009). On comparing neuronal morphologies with the constrained tree-edit-distance. *Neuroinformatics*, 7(3), 191–194.
- Gillette, T.A., Hosseini, P., Ascoli, G.A. (2015). Topological characterization of neuronal arbor morphology via sequence representation: II-global alignment. *BMC Bioinformatics*, 16(1), 209.
- Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., Li, Y., Schwarz, L.A., Li, A., Hu, B., et al. (2016). High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nature Communications*, 7, 12142.
- Gouwens, N.W., Sorensen, S.A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S.M., Feng, D., Anastassiou, C., Barkan, E., et al. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature Neuroscience*.
- Helmstaedter, M., Briggman, K.L., Turaga, S.C., Jain, V., Seung, H.S., Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461), 168.
- Heumann, H., & Wittum, G. (2009). The tree-edit-distance, a measure for quantifying neuronal morphology. *Neuroinformatics*, 7(3), 179–190.
- Jefferis, G.S.X.E., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfing, T., Maurer, C.R. Jr., Luo, L. (2007). Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell*, 128(6), 1187–1203.
- Jiang, X., Shen, S., Cadwell, C.R., Berens, P., Sinz, F., Ecker, A.S., Patel, S., Tolias, A.S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, 350(6264), aac9462.
- Kanari, L., Dłotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., Markram, H. (2018). A topological representation of branching neuronal morphologies. *Neuroinformatics*, 16(1), 3–13.
- Kanari, L., Ramaswamy, S., Shi, Y., Morand, S., Meystre, J., Perin, R., Abdellah, M., Wang, Y., Hess, K., Markram, H. (2019). Objective morphological classification of neocortical pyramidal cells. *Cerebral Cortex*.
- Li, Y., Wang, D., Ascoli, G.A., Mitra, P., Wang, Y. (2017). Metrics for comparing neuronal tree shapes based on persistent homology. *PLoS One*, 12(8), e0182184.
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., Caizhi, W. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10), 793.
- Memelli, H., Torben-Nielsen, B., Kozłowski, J. (2013). Self-referential forces are sufficient to explain different dendritic morphologies. *Frontiers in Neuroinformatics*, 7, 1.
- Mihaljević, B., Benavides-Piccione, R., Bielza, C., DeFelipe, J., Larrañaga, P. (2015). Bayesian network classifiers for categorizing cortical gabaergic interneurons. *Neuroinformatics*, 13(2), 193–208.
- Mihaljević, B., Larrañaga, P., Benavides-Piccione, R., Hill, S., DeFelipe, J., Bielza, C. (2018). Towards a supervised classification of neocortical interneuron morphologies. *BMC Bioinformatics*, 19(1), 511.
- Ofer, N., Shefi, O., Yaari, G. (2018). Axonal tree morphology and signal propagation dynamics improve neuronal classification. bioRxiv.
- Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L.F., Dalgleish, H., Carandini, M., Harris, K.D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. bioRxiv.
- Panico, J., & Sterling, P. (1995). Retinal neurons and vessels are not fractal but space-filling. *Journal of Comparative Neurology*, 361(3), 479–490.
- Peng, H., Ruan, Z., Long, F., Simpson, J.H., Myers, E.W. (2010). V3D Enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nature Biotechnology*, 28(4), 348.
- Peng, H., Tang, J., Xiao, H., Bria, A., Zhou, J., Butler, V., Zhou, Z., Gonzalez-Bellido, P.T., Oh, S.W., Chen, J., et al. (2014). Virtual finger boosts three-dimensional imaging and microsurgery as well as terabyte volume image visualization and analysis. *Nature Communications*, 5, 4342.
- Polavaram, S., Gillette, T.A., Parekh, R., Ascoli, G.A. (2014). Statistical analysis and data mining of digital reconstructions of dendritic morphologies. *Frontiers in Neuroanatomy*, 8, 138.
- Ragan, T., Kadiri, L.R., Venkataraju, K.U., Bahlmann, K., Sutun, J., Taranda, J., Arganda-Carreras, I., Kim, Y., Sebastian Seung, H., Osten, P. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature Methods*, 9(3), 255.
- Rousseeuw, P.J., & Driessen, K.V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212–223.
- Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular diversity and specializations among the cells of the adult mouse brain. *Cell*, 174(4), 1015–1030.

- Scala, F., Kobak, D., Shan, S., Bernaerts, Y., Laternus, S., Cadwell, C.R., Hartmanis, L., Froudarakis, E., Castro, J., Tan, Z.H., et al. (2019). Neocortical layer 4 in adult mouse differs in major cell types and circuit organization between primary sensory areas. *bioRxiv* page 507293.
- Schultz, S.R., Copeland, C.S., Foust, A.J., Quicke, P., Schuck, R. (2017). Advances in two-photon scanning and scanless microscopy technologies for functional neural circuit imaging. *Proceedings of the IEEE*, 105(1), 139–157.
- Scorcioni, R., Polavaram, S., Ascoli, G.A. (2008). L-measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, 3(5), 866.
- Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., et al. (2016). Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*, 166(5), 1308–1323.
- Sholl, D.A. (1953). Dendritic organization in the neurons of the visual and motor cortices of the cat. *Journal of Anatomy*, 87(4), 387.
- Stegle, O., Teichmann, S.A., Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, 16(3), 133.
- Stringer, C., & Pachitariu, M. (2019). Computational processing of neural recordings from calcium imaging data. *Current Opinion in Neurobiology*, 55, 22–31.
- Sümbül, U., Song, S., McCulloch, K., Becker, M., Lin, B., Sanes, J.R., Masland, R.H., Seung, H.S. (2014). A genetic and computational approach to structurally classify neuronal types. *Nature Communications*, 5, 3512.
- Tasic, B., Yao, Z., Graybiel, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729), 72.
- Uylings, H.B.M., & van Pelt, J. (2002). Measures for quantifying dendritic arborizations. *Network: Computation in Neural Systems*, 13(3), 397–414.
- van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605.
- van Pelt, J., & Schierwagen, A. (2004). Morphological analysis and modeling of neuronal dendrites. *Mathematical Biosciences*, 188(1-2), 147–155.
- van Pelt, J., & van Ooyen, A. (2013). Estimating neuronal connectivity from axonal and dendritic density fields. *Frontiers in Computational Neuroscience*, 7, 160.
- Verwer, R.W.H., & van Pelt, J. (1986). Descriptive and comparative analysis of geometrical properties of neuronal tree structures. *Journal of Neuroscience Methods*, 18(1-2), 179–206.
- Wolf, S., Grein, S., Queisser, G. (2013). Employing NeuGen 2.0 to automatically generate realistic morphologies of hippocampal neurons and neural networks in 3D. *Neuroinformatics*, 11(2), 137–148.
- Xi, C., Teichmann, S.A., Meyer, K.B. (2018). From tissues to cell types and back: single-cell gene expression analysis of tissue architecture. *Annual Review of Biomedical Data Science*, 1, 29–51.
- Yanbin, L., Carin, L., Coifman, R., Shain, W., Roysam, B. (2015). Quantitative arbor analytics: unsupervised harmonic co-clustering of populations of brain cell arbors based on L-measure. *Neuroinformatics*, 13(1), 47–63.
- Yanbin, L., Trett, K., Shain, W., Carin, L., Coifman, R., Roysam, B. (2013). Quantitative profiling of microglia populations using harmonic co-clustering of arbor morphology measurements. In *2013 IEEE 10th international symposium on biomedical imaging* (pp. 1360–1363): IEEE.
- Yatsenko, D., Reimer, J., Ecker, A.S., Walker, E.Y., Sinz, F., Berens, P., Hoenselaar, A., Cotton, R.J., Siapas, A.S., Tolias, A.S. (2015). Datajoint: managing big scientific data using MATLAB or Python. *bioRxiv*.
- Yuan, J., Gong, H., Li, A., Li, X., Chen, S., Zeng, S., Luo, Q. (2015). Visible rodent brain-wide networks at single-neuron resolution. *Frontiers in Neuroanatomy*, 9, 70.
- Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., Van Der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular architecture of the mouse nervous system. *Cell*, 174(4), 999–1014.
- Zeng, H., & Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9), 530.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.