

Sparse long-range connections in visual cortex for cost-efficient small-world networks

Seungdae Baek^{1†}, Youngjin Park^{1†} and Se-Bum Paik^{1,2*}

¹Department of Bio and Brain Engineering, ²Program of Brain and Cognitive Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

[†]These authors contributed equally to this work

*Correspondence and requests for materials should be addressed to Se-Bum Paik (spaik@kaist.ac.kr).

Abstract

The brain performs visual object recognition using much shallower hierarchical stages than artificial deep neural networks employ. However, the mechanism underlying this cost-efficient function is elusive. Here, we show that cortical long-range connectivity(LRC) may enable this parsimonious organization of circuits for balancing cost and performance. Using model network simulations based on data in tree shrews, we found that sparse LRCs, when added to local connections, organize a small-world network that dramatically enhances object recognition of shallow feedforward networks. We found that optimization of the ratio between LRCs and local connections maximizes the small-worldness and task performance of the network, by minimizing the total length of wiring needed for integration of the global information. We also found that the effect of LRCs varies by network size, which explains the existence of species-specific LRCs in mammalian visual cortex of various sizes. Our results demonstrate a biological strategy to achieve cost-efficient brain circuits.

Highlights

- Long-range connections (LRCs) enhance the object recognition of shallow networks
- Sparse LRCs added to dense local connections organize a small-world type network
- Small-worldness of networks modulates the balance between performance and wiring cost
- Distinct LRCs in various species are due to the size-dependent effect of LRCs

27 **Significance statement**

28 The hierarchical depth of the visual pathway in the brain is constrained by biological factors, whereas
 29 artificial deep neural networks consist of super-deep structures (i.e., as deep as computational power
 30 allows). Here, we show that long-range horizontal connections (LRCs) observed in mammalian visual
 31 cortex may enable shallow biological networks to perform cognitive tasks that require deeper artificial
 32 structures, by implementing cost-efficient organization of circuitry. Using model simulations based on
 33 anatomical data, we found that sparse LRCs, when added to dense local circuits, organize “small-world”
 34 type networks and that this dramatically enhances image classification performance by integrating both
 35 local and global components of visual stimulus. Our findings show a biological strategy of brain circuitry
 36 to balance sensory performance and wiring cost in the networks.

37

38 **One sentence summary**

39 Cortical long-range connections organize a small-world type network to achieve cost-efficient functional
 40 circuits under biological constraints

41

42 Introduction

43 Recently, computer vision techniques for object recognition using deep neural networks (DNNs) have
 44 been remarkably developed. These now show performance comparable to that of humans¹⁻⁵, implying
 45 that deep artificial neural network models can provide collective knowledge of visual processing in the
 46 brain⁶⁻⁸. However, there exist fundamental differences between the two kinds of systems, particularly
 47 regarding the structure required for visual feature abstraction. The brain consists of much shallower
 48 hierarchical stages than found in artificial deep neural networks designed for recognizing images with
 49 complex feature components. For example, ResNet³, one of the state-of-the-art DNNs for object
 50 recognition, consists of more than 150 hierarchical layers, thus requires high-level computing power to
 51 manage its super-deep structure. However, the ventral visual pathway (from retina to IT) in the brain is
 52 composed of less than 10 hierarchical stages⁹ (Fig 1a). This is presumably due to reasons such as the
 53 restricted volume of the physical space^{10,11} and limited metabolic resources^{12,13}. This raises questions
 54 about the distinct strategy implemented in the brain (as compared to artificial networks) that enables cost-
 55 efficient visual processing under these constraints.

56 Long-range horizontal connection (LRC), a distinct anatomical structure observed in the primary
 57 visual cortex (V1) of various mammals (Fig 1b) may provide an important insight into this issue. LRCs
 58 are found in the visual cortex of monkeys¹⁴, cats^{15,16}, tree shrews^{17,18}, gray squirrels¹⁹, ferrets²⁰, and rats²¹.
 59 From their extraordinarily long wiring (up to 2-3 mm), LRCs are distinguished from local connections of a
 60 short lateral spread (up to 0.5 mm)²²⁻²⁶ (Fig 1c). Previous studies have suggested that LRCs may play a
 61 particular role in visual processing, compensating for their large wiring cost, and possibly contributing to
 62 layer-wide functions such as contextual modulation at early stages²⁷⁻²⁹. However, the exact functions of
 63 LRCs for visual processing are still unknown.

64 Here, we suggest that LRCs added to local cortical connections can organize a “small-world
 65 network”³⁰ in visual cortex, which enables cost-efficient object recognition under resource constraints (Fig
 66 1d). The small-world, a term first introduced by Watts and Strogatz³⁰, is a network of high local clustering
 67 and short average global path length. We hypothesized that this may describe a characteristic structure
 68 of neural circuits required to encode both local and global feature components of visual images.
 69 Particularly in circuits of large V1, we assumed that mere local connections would not be sufficient to
 70 organize a small-world network, but LRCs could act as shortcuts for cortical inter-neural communications,
 71 organizing the entire V1 circuitry as a “small-world” (Fig 1d). On the other hand, LRCs may not contribute
 72 significantly in a small V1 network where local connections suffice.

73 To validate this idea, we implemented a model neural network of V1 and performed simulations

for image classification capability of the network. We found that addition of the LRCs significantly enhances object recognition performance of a shallow network, and that the degree of performance enhancement is strongly correlated to the small-worldness of the circuit. Our key prediction that the effect of LRCs is dependent on the size of the network was validated by our analysis of animal data on the existence of LRCs in diverse mammal species. Taken together, our results provide a comprehensive understanding of the functional role of LRCs in V1.

Results

LRCs enhance object recognition performance of shallow networks

To test the contribution of LRCs to object recognition of shallow neural networks, we designed a three-layer convergent neural network as a simplified model of the ventral visual pathway (Fig 2a). The model network consisted of 784 input units, one hidden layer of 784 units, and 10 output units for classification. Neurons in the input and the hidden layers were linked by local feedforward connections, following observed convergent projections of the early visual pathway³¹. Model long-range connections were added in the lateral connections of the hidden layer.

We trained and tested the model network using CIFAR-10 datasets for natural image classification. Three types of network conditions with different connectivity were compared (Fig 2b): (1) feedforward only (FF), (2) feedforward with LRCs (FF+LRC), and (3) fully-connected (FC). As expected, we found that purely feedforward networks without LRCs show lower classification accuracy (23.5%) (Fig 2c) than that of a fully connected network (33.7%) (Fig 2c, black dashed line) which may indicate the maximum capability of this shallow network. Notably, we found that adding only a small number of LRCs to the FF network (Fig 2c) significantly enhances the network performance, comparable to that of the FC. When the performance of FF+LRC becomes equivalent to that of the FC (Fig 2c, left inset, Mann-Whitney U-test: $p = 0.15$), it uses only 5% of the number of total learnable connections compared to that of the FC network (Fig 2c, right inset). This indicates that LRCs can noticeably enhance the object recognition of shallow networks in a cost-efficient manner. We also confirmed that the performance improvement by LRCs was fairly consistent across different conditions of feedforward connectivity such as sampling ratio and convergence range (Supplementary Fig 1).

To investigate the cost-efficiency of LRCs more specifically, we varied the ratio of LRCs (the number of LRCs per total number of connections) in the network while the total number of connections were unchanged. We added LRCs while disconnecting the same number of local feedforward

connections, and measured changes in the classification performance of the network (Fig 2d). We found that the accuracy was dramatically improved by additional LRCs, and that the performance was maximized when sparse (20%) LRCs were combined with the dense local feedforward network (Fig 2e, brown asterisk) (Fig 2e inset, Mann-Whitney U-test: $*p < 10^{-14}$). This result indicates that a certain portion of LRCs inserted in a shallow FF network can optimize the cost-efficiency of the network.

Next, to investigate the contribution of the LRCs to images of a different spatial frequency spectrum, we generated new datasets of various frequency components (“shape”, “position”, and “shape-position”), by organizing different combinations of MNIST datasets (Fig. 2f, see Supplementary Fig 2 for details). This demands that the network encodes either local features (“shape”), global features (“position”), or both, to classify each class correctly. Using these datasets, we trained the network and examined the classification performance while varying the ratio of the LRCs again. We found that when the LRC ratio was 0%, the network showed higher performance for “shape” datasets than for the other two datasets (Fig. 2g and 2h, LRC ratio = 0%). On the other hand, the network with a large portion of LRCs showed better performance for “position” and “shape-position” datasets (Fig. 2g and 2h, LRC ratio = 97%). In particular, the network with a certain portion (20%) of LRCs showed the maximum average accuracy for the three datasets, implying a consistent classification capability for various types of images (Fig. 2g, brown asterisk). In addition, this result supports our hypothesis that LRCs enhance the network performance by integrating global features of images that cannot be captured by local connections only.

Combination of LRCs and local connections organizes a small-world type network

It is notable that networks with combined sparse LRCs and dense local connections are advantageous for encoding a wide range of the frequency spectrum of natural images. Interestingly, this connectivity architecture is comparable to a “small-world network^{30,32}”, in terms of both structure and purpose. A small-world type network is defined as a network of high local clustering and short average global path length³⁰. One is generally achieved using a number of local clusters with sparse long shortcuts (Fig 1d, bottom). Small-world networks have advantages of both high “local” interactions between adjacent nodes and short “global” distance between distant nodes. From this similarity, we hypothesized that a “small-worldness” index would efficiently describe the ability of our network to classify both “local” and “global” components in natural images (Fig 3a).

To test this hypothesis, we investigated correlations between the classification performance of the model network and a network structure described by the clustering coefficient (C), average path

length (L), and small-worldness (SW) index³². We found that performance enhancement was strongly correlated to the corresponding network parameters. Change in the clustering coefficient (C) is strongly correlated with performance for a “shape” dataset (Fig 3b), and a reciprocal of the shortest path length ($1/L$) is strongly correlated with performance for a “position” dataset (Fig 3c). As a result, the small-worldness, a value defined as C divided by L , shows a significant correlation with performance for a “shape-position” dataset (Fig 3d). We confirmed that the addition of LRCs to an FF network increased the image classification performance until the small-worldness reached the maximum. This suggests that the combination of sparse LRCs and dense local feedforward connections allows emergence of a new small-world network, and enables consistent object recognition of various types of images.

Size-dependency of small-worldness predicts the existence of species-specific LRCs

LRCs are observed in various mammalian species^{14–21}, but not in rodents smaller than rats (Fig 4a). Mice and rats are genetically close relatives with neuroanatomical and functional similarities, including organization of functional circuits in V1^{33–35}. Despite their genetic proximity, rats have LRCs clearly distinct from their local lateral spread, whereas mice do not (Supplementary Fig 3). Here we propose that the contribution of LRCs to small-worldness varies according to the size of the networks, and that this can explain the existence of species-specific LRCs.

Retino-cortical feedforward projections organize a precise retinotopy in most species^{34,36,37}, and visual information from a local retinal area is projected onto a local cortical space. If the size of the cortex is large, two distant cortical neurons that receive inputs from distinct retinal spaces cannot communicate because their local connections do not overlap. However, addition of LRCs in this large cortex enables communication of distant neurons by reducing the average path length in cortical space. In this condition, LRCs increase dramatically the small-worldness of a cortical network (as shown in Fig 3b-d). In contrast, if the size of the cortex is relatively small, as in mice, introducing LRCs would not further decrease the path length, because cortical neurons already share information with neighboring neurons via local connections. Thus small-worldness will remain unchanged even though a large number of LRCs are added. This scenario implies that LRCs may not affect the performance of networks of small size.

To validate this prediction, we performed connectivity analysis of networks of various sizes, simultaneously with the image-classification performance test: The CIFAR-10 classification performance was estimated while the small-worldness was measured in model cortical networks of various sizes (Fig 4b). When the model cortex size was large (gray squirrel^{35,38}, 82 mm²), introduction of LRCs increased

both classification accuracy and small-worldness (Fig 4c, rightmost). In contrast, when the cortex size was small (mouse^{35,39}, 3 mm²), the performance and small-worldness did not increase further regardless of how many LRCs were added (Fig 4c, leftmost). In general, we found a strong correlation between the increment of performance (ΔPerf) and small-worldness (ΔSW) in relation to the variation of V1 size (Fig 4d). Notably, ΔPerf and ΔSW were positive only when the cortex size exceeded a certain threshold (~ 5 mm², Fig 4e; Supplementary Fig 4). These results provide a simple explanation of the existence of species-specific LRCs.

Discussion

We showed that cortical long-range connections can play an important role in developing a cost-efficient circuit under structural constraints. In shallow networks with large processing area, LRCs act as shortcuts for the communication of distant neurons, which forms a small-world network. In networks of small area, however, LRCs could not contribute in such ways and this provides a possible explanation for the existence of species-specific LRCs in various mammals.

An ability to recognize natural images is a fundamental function in animals that is required for survival. Because natural images contain a significant portion of low-frequency component⁴⁰, a successful visual circuit should be able to encode this “global” information⁴¹. This requires sufficiently deep hierarchical structures and a large number of connections⁴², unless the convergent range of feedforward projections is large enough (i.e., comparable to fully connected networks). Any of the scenarios above require a significant amount of wiring cost. However, due to the restricted volume of the physical space and limited metabolic resources, the visual pathway in the brain does not include such a deep structure. This makes it difficult to encode the global features of visual input with localized feedforward projections. In this case, implementing a wide range of convergence between the layers also may not be an appropriate choice because it may lead to a significant loss of high-frequency “local” information. Here, according to our model, the combination of local convergent projections and sparse LRCs may enable the brain to address this issue by capturing both local and global information, using a minimum of wiring cost.

In the current study, for simplicity, we neglected detailed biological parameters except the size of V1 in the analysis of the small-worldness of the network (Fig 4). However, there are a number of other factors that affect the small-worldness condition of networks in the visual pathway: cell density, the convergent range of feedforward projections, the nonlinear magnification factor in the retinotopy, and the

recurrent and feedback circuits. These are all possible factors affecting the information processing. Detailed effects of these factors need to be considered in subsequent studies.

An additional question about the formation of LRCs during development remains. It has been suggested that LRCs develop without visual experience, because they are observed before eye-opening^{43,44}. This suggests that LRCs, a characteristic architecture that links cortical neurons of similar tuning, may emerge spontaneously without training. Previous studies have reported that feedforward afferents from the periphery may play important roles in early development of the cortex^{45,46}, and that LRCs in sensory cortices can be induced by feedforward projections from the retina⁴⁷. In a previous study⁴⁸, which is relevant to the current findings but contains independent findings, we showed that long-range horizontal connections in V1 may emerge from spatio-temporally structured retinal waves generated spontaneously, implying that cortical LRCs may originate from early peripheral activities before eye-opening. Further investigation of this scenario from the analysis of developing V1 in young animals may provide further support for our model.

Overall, our results suggest that cortical long-range connections enable organization of small-world type networks for cost-efficient visual recognition under biological constraints. This finding offers a simple but powerful model that explains the role of cortical long-range connections and an underlying mechanism for species-specific variants. This reveals a biological strategy of the brain to balance functional performance and resource cost.

Methods

Neural network model for image classification

All simulations were performed using the MATLAB deep learning toolbox. To implement a simplified model of the ventral visual pathway, we used the multilayer perceptron with three layers, consisting of an input layer of 784 units, one hidden layer of 784 units, and an output layer of 10 units for classification. The schematic architecture of the model is shown in Fig 2a.

The feedforward connectivity between the input and hidden layers were set to local convergent structure, following the study of Hubel and Wiesel^{31,49}. The i^{th} neuron in the hidden layer has a circular receptive field, that is

$$\text{if } d_{ij} < c, w_{ij}^b = 1$$

$$\text{otherwise, } w_{ij}^b = 0$$

where j is the index of the neuron in the input layer, and d_{ij} is the Euclidean distance between neurons i and j when the input layer is projected to the hidden layer. Here, c denotes the size of the receptive field, which is equivalent to maximum distance d_{ij} , which was initially set to 4. The term w_{ij}^b is a Boolean parameter that represents the presence of connection between neuron i and j . Connections between the hidden and output layer were randomly connected with probability 0.5.

Lateral connections could only be added in the hidden layer. The length distribution of lateral connections followed the V1 data¹⁷ observed in tree shrew V1 (Fig 1c). To match the statistics observed in the biological data, we assumed the condition that the 15° visual images were shown to a head-fixed tree shrew. Considering the magnification factor and V1 size in a tree shrew^{17,35,50}, the scaling factor was set to 0.1 mm/unit, so that one unit of the input and hidden layers matched 0.1 mm in the cortical space of the biological data. Among the lateral connections, long-range connections were defined as lateral connections longer than 1 mm⁵¹.

To perform image classification, our model was trained using stochastic gradient descent, even though it is not strictly a biologically plausible learning rule⁵². The batch size was set to 200 images and the weight decay factor was set to 0. All connection weights were initialized from a zero-mean Gaussian distribution with a standard deviation of 0.01. The learning rate was set to a constant value of 0.1. Other hyperparameters such as the number of epochs were chosen to provide reasonable accuracy in the image classification task.

245

246 **Image datasets**

247 The model was trained for performing the image classification task. Depending on the purpose, we used
 248 four kinds of image datasets. To test the ability of the model network for overall image recognition
 249 performance, the CIFAR-10 dataset⁵³ was used as an example of a natural image. Before training the
 250 network, the image was converted to greyscale using the MATLAB `rgb2gray` function.

251 To investigate the frequency specific role of LRCs for image classification, we generated three
 252 datasets by modifying the size and position of handwritten digits (MNIST). Details are as follows:

253 1) The “shape” dataset was designed by resizing a hand-written digital image to 8×8 pixels in the center
 254 of 28×28 pixels. The “local” dataset consists of eight categories depending on the number (selected
 255 from 1 to 8). For this dataset, only local information (shape) of the digits was required for classification.

256 2) The “position” dataset was made using the following procedure. First, the 28×28 -pixel area was
 257 divided into four 14×14 areas. Second, two out of the four areas were chosen. Third, a randomly selected
 258 8×8 -pixel digit was inserted into a random position in the chosen area. The “global” dataset also
 259 consisted of eight categories depending on the position where the digit was located. The type of digit is
 260 irrelevant for the classification.

261 3) The “shape-position” dataset was generated by the following procedure. First, the 28×28 -pixel area
 262 was divided into four 14×14 areas. Second, two diagonally aligned areas were chosen (either 45° or
 263 135°). Third, an 8×8 pixel digital image (either “7” or “9”) was located in a random position of the selected
 264 area. The reason we chose “7” and “9” among ten numbers was to make the difficulty of the task match
 265 that of the previous tasks. The “shape-position” dataset consisted of eight categories depending on the
 266 type and location of the digits.

267 For the “shape” dataset, the local shape of the digits was the only information related to the
 268 classification. For the “position” dataset, the shape of the numbers is irrelevant and only the position
 269 where two numbers are placed was important information for the classification. For the “shape-position”
 270 dataset, the network should perceive both the shape and position of the digits in order to classify the
 271 images correctly. For all three datasets, each category contained 1000 generated images. We properly
 272 tuned the difficulty of the three tasks to be similar by adjusting the background noise level and position
 273 of the numbers. This worked because they did not have to be exactly the same.

274

275 **Test of image recognition performance according to the connectivity**

276 To examine the exact role of a long-range lateral connection (Fig 2), we varied the ratio of the LRCs in
 277 the network and tested the image recognition performance after training of the modified MNIST dataset
 278 (shape, position, shape-position). Note that, in order to fairly compare the performance of networks that
 279 had different structures, the total number of connections had to be controlled because it is obvious that
 280 the network that has more learnable parameters should perform better. To keep the number of learnable
 281 parameters the same while varying the network connectivity, we ablated the feedforward connections
 282 between the input and hidden layer while adding the same number of LRCs (described in Fig 2b). The
 283 connections between the hidden and output classification layer were not ablated because those
 284 connections were thought to be required for the classification, but not for the image abstraction.

285

286 **Network connectivity indices**

287 Overall, the network parameters we used in the paper followed earlier definitions³² with small
 288 modifications. The clustering coefficient C is the average density of connections between the neighbors
 289 and is described as

$$290 \quad C = \log\left(\frac{1}{N_{hidden}} \sum_j c_j\right)$$

291 where N_{hidden} is the number of hidden neurons and c_j is the local coefficient of j hidden neurons. Here,
 292 c_j is determined by the following equation:

$$293 \quad c_j = \frac{1}{N_{neighbor,j}} \left(\sum_i \frac{1}{d_{ii'}} + \sum_{j'} \frac{1}{d_{jj'}} \right)$$

294 where $N_{neighbor,j}$ is the number of neurons connected with j hidden neurons, $d_{ii'}$ is the inter-neighbor
 295 distance between source i input neuron and destination i' input neuron, which are connected with j
 296 hidden neuron. Similarly, $d_{jj'}$ is the inter-neighbor distance between source j' hidden neuron and
 297 destination j' hidden neuron, which are connected with j hidden neuron.

298 The characteristic path length L is defined as:

$$L = \frac{1}{N_{input}(N_{input} - 1)} \sum_{i \neq i'} l_{ii'}$$

where N_{input} is the number of input neurons and $l_{ii'}$ is the shortest path between i and i' . The term $l_{ii'}$ is defined as the shortest path from a random input neuron i to another input neuron i' . Please note that the shortest path must include at least one hidden neuron because the network is a layered structure.

The small-worldness was defined as the ratio of the clustering coefficient C and path length L . However, this index is largely influenced by network parameters such as the number of units. To avoid this problem, the small-worldness SW was modified to the ratio of the normalized clustering coefficient C and normalized path length L as follows:

$$SW = \frac{C - C_{rand}}{C_{regular} - C_{rand}} \times \frac{L_{regular} - L_{rand}}{L - L_{rand}}$$

where C_{rand} and L_{rand} are the clustering coefficient and characteristic path length on a random network consisting of only randomly generated LRCs. The terms $C_{regular}$ and $L_{regular}$ are the clustering coefficient and characteristic path length on a random network consisting of only FFs.

Network model for various species

To model the visual pathways of four different species (mouse, rat, tree shrew, and gray squirrel), we varied the size of the processing layer (also the number of neurons) according to the relative V1 size of each species. All the other parameters (including input cell density, sparsity of feedforward connections, and V1 cell density) were assumed to be consistent across species. The anatomical data of averaged V1 size used in this study were from mouse^{35,39} (3 mm²), rat^{35,54} (7 mm²), tree shrew⁵⁰ (73 mm²), and gray squirrel^{35,38} (82 mm²). Based on the tree shrew model (784 hidden units) we defined earlier, we set 900 units for gray squirrel, 25 units for the mouse and 64 units for the rat models.

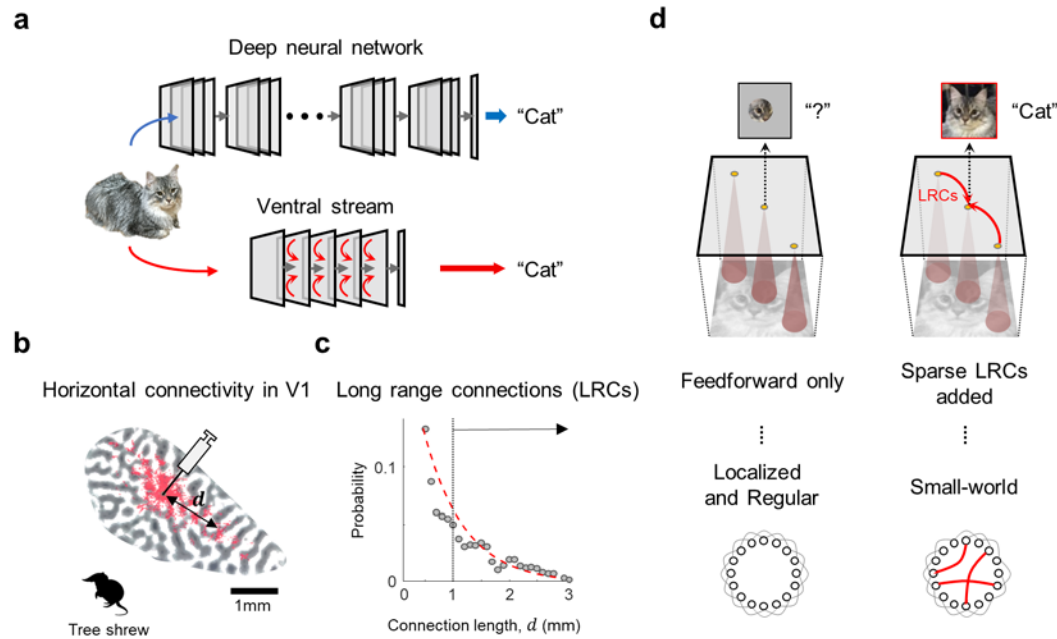
References

1. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (*NIPS 2012*) 1097–1105 (2012).
2. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv.1409.1556*. (2014).
3. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016). doi:10.1109/CVPR.2016.90
4. Zhang, Q. *et al.* Recent advances in convolutional neural network acceleration. *Neurocomputing* **323**, 37–51 (2019).
5. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
6. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365 (2016).
7. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci.* **111**, 8619–8624 (2014).
8. Bashivan, P., Kar, K. & DiCarlo, J. J. Neural population control via deep image synthesis. *Science* **364**, eaav9436 (2019).
9. DiCarlo, J. J. & Cox, D. D. Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007).
10. Horvát, S. *et al.* Spatial embedding and wiring cost constrain the functional layout of the cortical network of rodents and primates. *PLoS Biol.* **14**, 1–30 (2016).
11. Samu, D., Seth, A. K. & Nowotny, T. Influence of Wiring Cost on the Large-Scale Architecture of Human Cortical Connectivity. *PLoS Comput. Biol.* **10**, e1003557 (2014).
12. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).
13. Van den Heuvel, M. P., Bullmore, E. T. & Sporns, O. Comparative Connectomics. *Trends Cogn. Sci.* **20**, 345–361 (2016).
14. Stettler, D. D., Das, A., Bennett, J. & Gilbert, C. D. Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* **36**, 739–750 (2002).
15. Gibson, J. R. & Wiesel, T. N. Morphology and intracortical projections of functionally identified neurons in cat visual cortex. *Nature* **280**, 120–125 (1979).
16. Das, A., Gilbert, C. & D. Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. *Nature* **375**, 780–784 (1995).
17. Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17**, 2112–2127 (1997).
18. Chisum, H. J., Mooser, F. & Fitzpatrick, D. Emergent properties of layer 2/3 neurons reflect the collinear arrangement of horizontal connections in tree shrew visual cortex. *J. Neurosci.* **23**, 2947–2960 (2003).
19. Van Hooser, S. D., Heimel, J. A., Chung, S. & Nelson, S. B. Lack of patchy horizontal connectivity in primary visual cortex of a mammal without orientation maps. *J. Neurosci.* **26**, 7680–7692 (2006).
20. Durack, J. C. & Katz, L. C. Development of horizontal projections in layer 2 / 3 of ferret visual cortex.

- 359 *Cereb. Cortex* **6**, 178–183 (1996).
- 360 21. Rumberger, A., Tyler, C. J. & Lund, J. S. Intra- and inter-areal connections between the primary visual
361 cortex V1 and the area immediately surrounding V1 in the rat. *Neuroscience* **102**, 35–52 (2001).
- 362 22. Kisvárdy, Z. F. *et al.* Synaptic targets of HRP-filled layer III pyramidal cells in the cat striate cortex. *Exp.*
363 *Brain Res.* **64**, 541–552 (1986).
- 364 23. Kisvárdy, Z. F. & Eysel, U. T. Cellular organization of reciprocal patchy networks in layer III of cat visual
365 cortex (area 17). *Neuroscience* **46**, 275–286 (1992).
- 366 24. Malach, R., Amir, Y., Harel, M. & Grinvald, A. Relationship between intrinsic connections and functional
367 architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex.
368 *Proc. Natl. Acad. Sci.* **90**, 10469–10473 (1993).
- 369 25. Voges, N., Schüz, A., Aertsen, A. & Rotter, S. A modeler's view on the spatial structure of intrinsic
370 horizontal connectivity in the neocortex. *Prog. Neurobiol.* **92**, 277–292 (2010).
- 371 26. Stepanyants, A., Martinez, L. M. & Ferecsko, A. S. The fractions of short- and long-range connections.
372 *Proc. Natl. Acad. Sci.* **106**, 3555–3560 (2009).
- 373 27. Angelucci, A. *et al.* Circuits for local and global signal integration in primary visual cortex. *J. Neurosci.* **22**,
374 8633–8646 (2002).
- 375 28. Piech, V., Li, W., Reeke, G. N. & Gilbert, C. D. Network model of top-down influences on local gain and
376 contextual interactions in visual cortex. *Proc. Natl. Acad. Sci.* **110**, E4108–E4117 (2013).
- 377 29. Capparelli, F., Pawelzik, K. & Ernst, U. Constrained inference in sparse coding reproduces contextual
378 effects and predicts laminar neural dynamics. *PLoS Comput. Biol.* **15**, e1007370 (2019).
- 379 30. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- 380 31. Hubel, D. & Wiesel, T. Receptive fields, binocular interaction and functional architecture in the cat's visual
381 cortex. *J. Physiol.* **160**, 106–154 (1962).
- 382 32. Latapy, M., Magnien, C. & Vecchio, N. Del. Basic notions for the analysis of large two-mode networks.
383 *Soc. Networks* **30**, 31–48 (2008).
- 384 33. Kaschube, M. Neural maps versus salt-and-pepper organization in visual cortex. *Curr. Opin. Neurobiol.* **24**,
385 95–102 (2014).
- 386 34. Jang, J., Song, M. & Paik, S. B. Retino-cortical mapping ratio predicts columnar and salt-and-pepper
387 organization in mammalian visual cortex. *Cell Rep.* **30**, 3270–3279.e3 (2020).
- 388 35. Van Hooser, S. D. Orientation selectivity without orientation maps in visual cortex of a highly visual
389 mammal. *J. Neurosci.* **25**, 19–28 (2005).
- 390 36. Lee, K. S., Huang, X. & Fitzpatrick, D. Topology of on and off inputs in visual cortex enables an invariant
391 columnar architecture. *Nature* **533**, 90–94 (2016).
- 392 37. Ringach, D. L. Haphazard wiring of simple receptive fields and orientation columns in visual cortex. *J.*
393 *Neurophysiol.* **92**, 468–476 (2004).
- 394 38. Kay, R. F. & Ross, C. *Anthropoid origins: new visions*. (Kluwer Academic/Plenum Publishers, 2004).
- 395 39. Mazade, R. & Alonso, J. M. Thalamocortical processing in vision. *Vis. Neurosci.* **34**, E007 (2017).
- 396 40. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev.*

- 397 *Neurosci.* **24**, 1193–1216 (2001).
- 398 41. Turner, M. H., Sanchez Giraldo, L. G., Schwartz, O. & Rieke, F. Stimulus- and goal-oriented frameworks
399 for understanding natural vision. *Nat. Neurosci.* **22**, 15–24 (2019).
- 400 42. Luo, W., Li, Y., Urtasun, R. & Zemel, R. Understanding the effective receptive field in deep convolutional
401 neural networks. *Adv. Neural Inf. Process. Syst.* 4898–4906 (2016).
- 402 43. Ko, H., Mrsic-Flogel, T. D. & Hofer, S. B. Emergence of feature-specific connectivity in cortical
403 microcircuits in the absence of visual experience. *J. Neurosci.* **34**, 9812–9816 (2014).
- 404 44. Ruthazer, E. S. & Stryker, M. P. The role of activity in the development of long-range horizontal
405 connections in area 17 of the ferret. *J. Neurosci.* **16**, 7253–7269 (1996).
- 406 45. Jin, J., Wang, Y., Swadlow, H. A. & Alonso, J. M. Population receptive fields of on and off thalamic inputs
407 to an orientation column in visual cortex. *Nat. Neurosci.* **14**, 232–240 (2011).
- 408 46. Paik, S. & Ringach, D. L. Retinal origin of orientation maps in visual cortex. *Nat. Neurosci.* **14**, 919–925
409 (2011).
- 410 47. Sharma, J., Angelucci, A. & Sur, M. Induction of visual orientation modules in auditory cortex. *Nature* **404**,
411 841–847 (2000).
- 412 48. Kim, J., Song, M. & Paik, S. Spontaneous retinal waves generate long-range horizontal connectivity in
413 visual cortex. *bioRxiv.997015* (2020). doi: <https://doi.org/10.1101/2020.03.18.997015>
- 414 49. Sailamul, P., Jang, J. & Paik, S. B. Synaptic convergence regulates synchronization-dependent spike
415 transfer in feedforward neural networks. *J. Comput. Neurosci.* **43**, 189–202 (2017).
- 416 50. Keil, W. *et al.* Response to comment on ‘Universality in the evolution of orientation columns in the visual
417 cortex’. *Science* **336**, 413–413 (2012).
- 418 51. Trachtenberg, J. T. T. & Stryker, M. P. P. Rapid anatomical plasticity of horizontal connections in the
419 developing visual cortex. *J. Neurosci.* **21**, 3476 (2001).
- 420 52. Park, Y., Choi, W. & Paik, S. B. Symmetry of learning rate in synaptic plasticity modulates formation of
421 flexible and stable memories. *Sci. Rep.* **7**, 5671 (2017).
- 422 53. Krizhevsky, A. & Hinton, G. Learning multiple layers of features from tiny images. (2009).
- 423 54. Espinoza, S. G. & Thomas, H. C. Retinotopic organization of striate and extrastriate visual cortex in the
424 hooded rat. *Brain Res.* **272**, 137–144 (1983).
- 425

426 Figures



427

428 Figure.1. Cortical long-range horizontal connectivity for cost-efficient small-world networks

429 (a) Hierarchical structures of deep neural networks and the ventral visual pathway. (b) Long-range horizontal connections
 430 (LRCs) observed in the visual cortex of tree shrews (redrawn from Bosking 1997, Copyright 1997 Society for Neuroscience).
 431 Red dots represent synaptic boutons with retrograde labeling by injection. (c) Distribution of the length of the horizontal
 432 connections in tree shrew V1 (gray circle; Bosking 1997). The dashed curve represents an exponential function fitted to data.
 433 (d) Illustration of the circuits in shallow networks with and without LRCs. Sparse LRCs added to local feedforward projections
 434 organize a "small-world network".

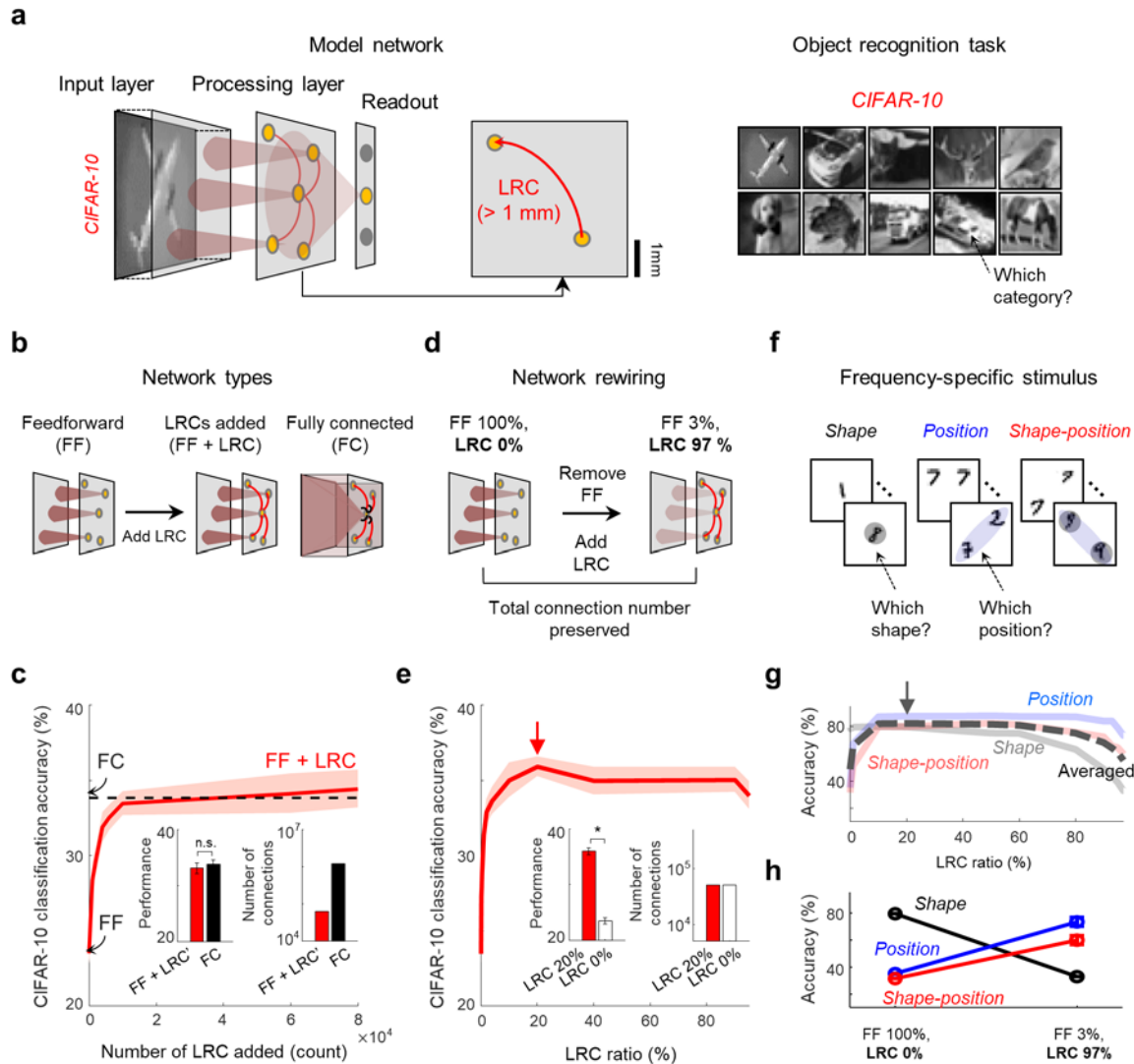


Figure 2. Long-range connections enhance image classification of shallow networks

(a) A simplified model of shallow networks in the visual pathway (left). Examples of the CIFAR-10 dataset used for image classification task (right). (b) Three types of shallow networks with different connectivity. (c) Image classification performance of the networks in (b). LRCs added to local feedforward networks drastically enhance classification accuracy (red solid line), comparable with that of FC (black dashed line). Performance of the FF + LRC (N = 10,000 LRCs) and the FC networks (left inset; Mann-Whitney U-test: $p = 0.15$). This FF + LRC network uses only 5% of the number of connections in the FC (right inset). (d) The network rewiring process. Feedforward connections (FF) are removed and the same number of LRCs are added to preserve the total number of connections. (e) Classification accuracy for CIFAR-10 datasets during the rewiring process. Red arrow indicates the maximum accuracy achieved. Classification performance increases until the ratio of LRCs reaches 20% of the total number of connections (left inset; Mann-Whitney U-test: $*p < 10^{-14}$), where the amount of total wiring in the network is unchanged (right inset). (f) Modified MNIST datasets for three types of classification test. The "shape" dataset requires identification of local shape of the digits, and the "position" dataset requires classification of the position of digits. The "shape-position" dataset requires both. (g) Classification accuracy for each dataset with variation of the LRC ratio. Average accuracy for all three tasks (black dashed line). Black arrow indicates the maximum accuracy. (h) Classification accuracy for each dataset of the network with 0% LRC and with 97% LRCs. Error bars and shaded area represent the standard deviation of the accuracy for 20 trials.

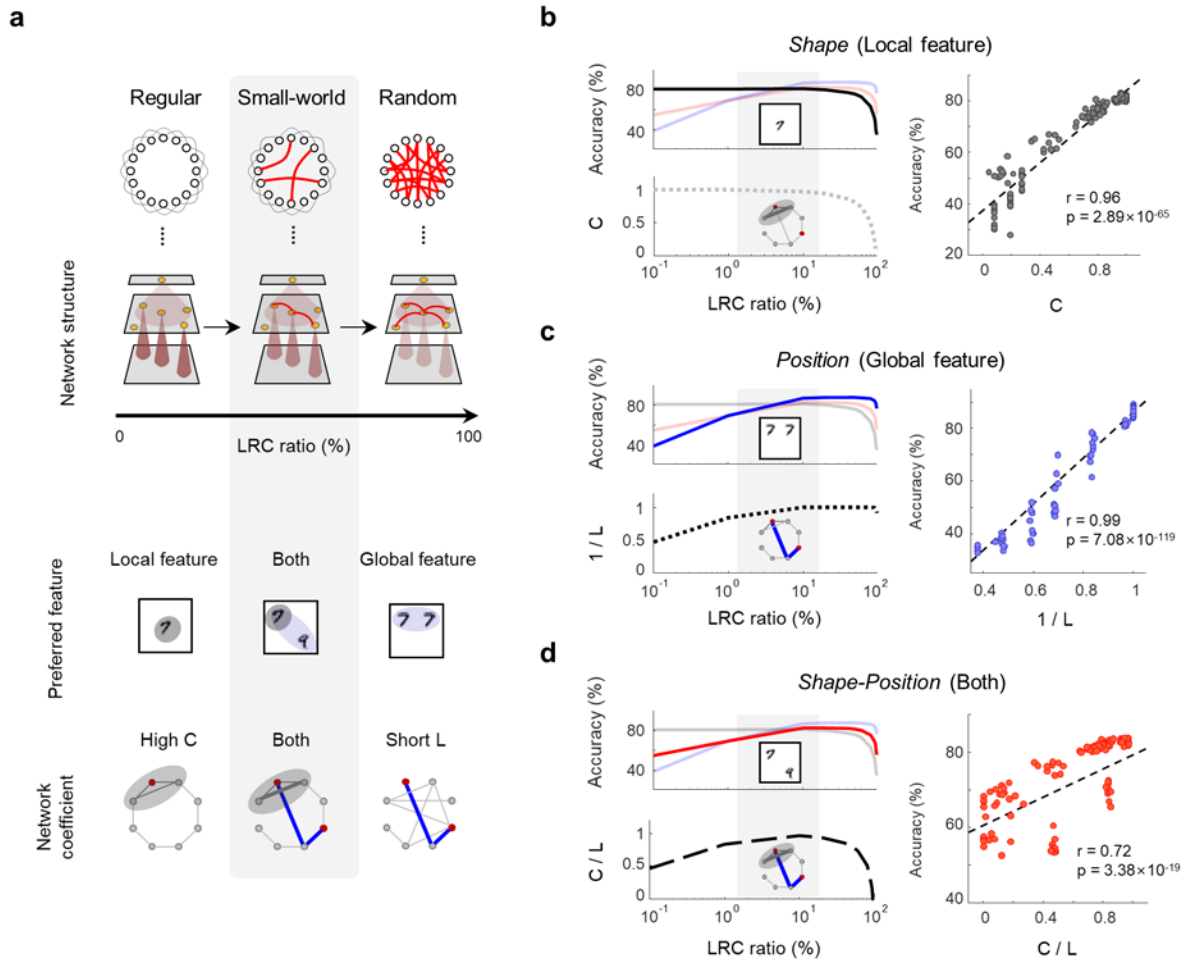


Figure 3. Network-theory approach to long-range connectivity

(a) Addition of LRCs to a local FF network modulates the network connection structure toward a small-world network, enabling correct classification of both local and global features in images. (b) Classification accuracy for “*Shape*” dataset (left top) and the clustering coefficient (C , left bottom) of the network during the rewiring process are strongly correlated (right). (c) Classification accuracy for “*Position*” dataset (left top) and reciprocal of average path length ($1/L$, left bottom) are also correlated (right). (d) Classification accuracy for “*Shape-Position*” dataset (left top) and the small-worldness index (C/L , left bottom) are correlated significantly (right). Gray shaded area represents “small-world” regime.

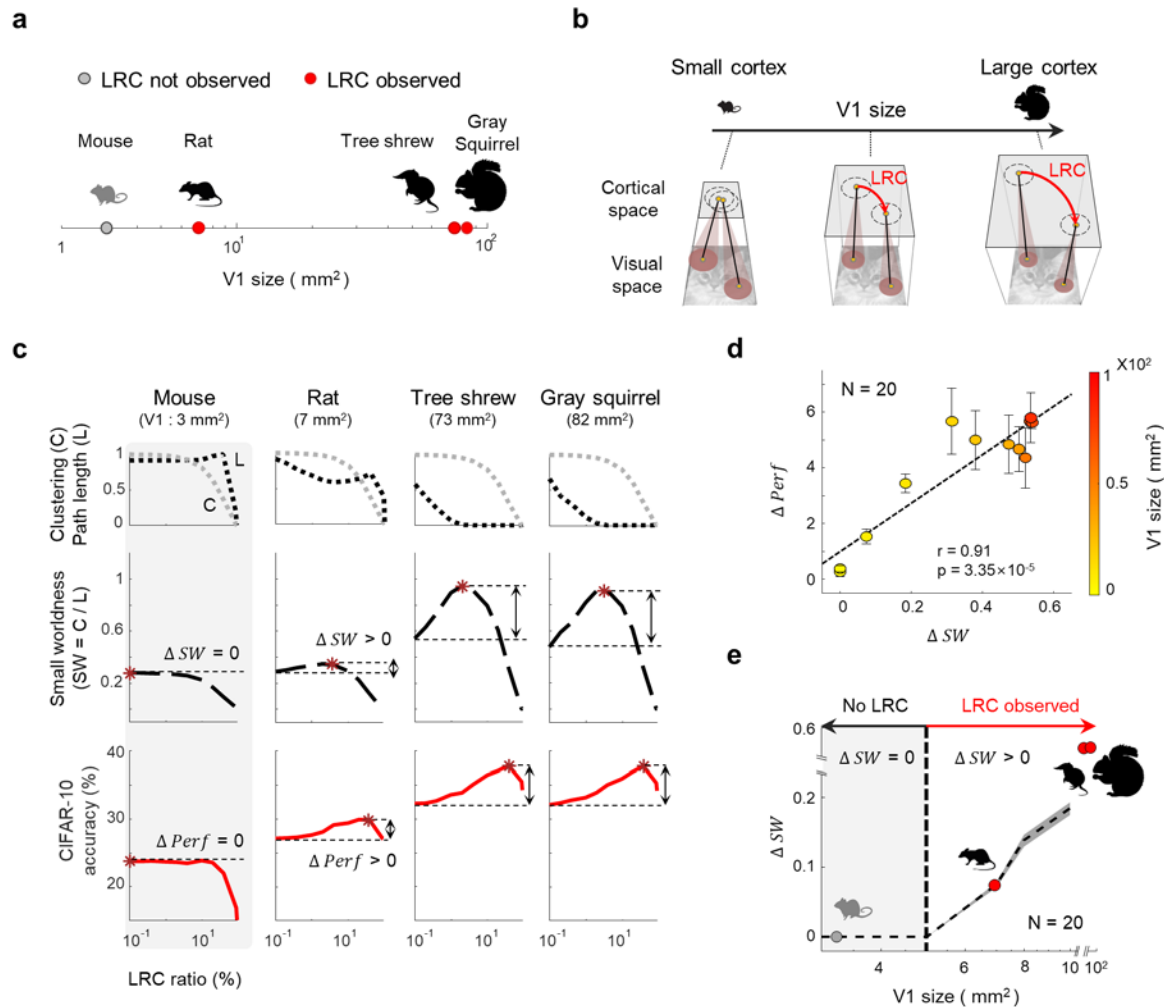
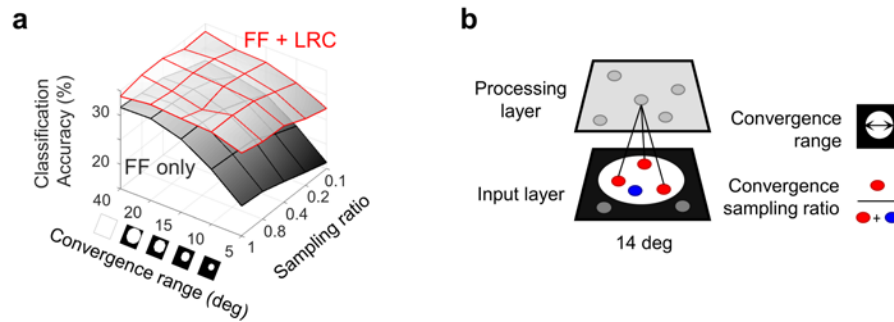


Figure 4. Long-range connections do not increase the small-worldness in small cortices

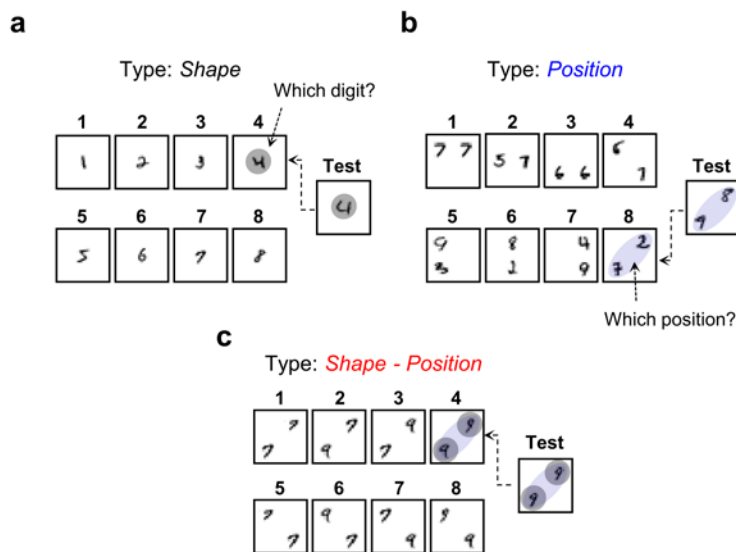
(a) Species with and without LRCs. (b) Illustration of the size-dependent effect of LRCs. Shaded circles represent local visual area (receptive field) projected onto target cortical neurons indicated as yellow dots. Note that LRCs in a large cortex can link two distant cortical neurons, the receptive fields of which are not overlapped. (c) The CIFAR-10 classification accuracy and the network parameters as a function of LRC ratio (mouse, rat, tree shrew, gray squirrel). Changes in performance accuracy and small-worldness, $\Delta Perf$ (accuracy) and ΔSW (small-worldness), were defined as the difference between the maximum and the initial value. Gray dashed lines represent the clustering coefficient (C) and black dashed line indicate the path length (L). Brown asterisks indicate the maximum points of accuracy and small-worldness curves. (d) Correlation between $\Delta Perf$ and ΔSW of networks for cortices of various sizes. (e) Changes of small-worldness, ΔSW , as a function of V1 size. Note that ΔSW becomes positive only when the cortex size exceeds 5 mm². Error bars represent the standard deviation for 20 trials in (d) and (e).

Supplementary Information



Supplementary Fig.1. Effect of the LRCs on various feedforward connectivity

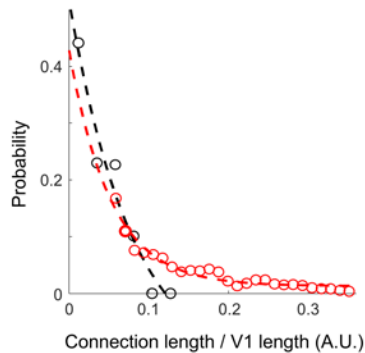
(a) Classification accuracy of the network for CIFAR-10 datasets in relation to the variation of structural parameters of the feedforward projections. The black surface represents performance of the feedforward network (FF only), and the red surface represents performance of the network with $n=10,000$ (1.2% of connections of fully connected networks in Fig 2) of LRCs added (FF + LRC). Performance is improved by LRCs, regardless of the condition of feedforward connectivity. (b) The convergence range of feedforward projections is defined as the diameter of sampling in the input layer. In the current model, 1 pixel of CIFAR-10 image was scaled to 0.5 degree in visual space. The sampling ratio is defined as the connection probability of neurons in the input layer within the convergence range.



Supplementary Fig.2. Modified MNIST datasets for the test of local and global feature classification

(a) The "shape" dataset was designed by arranging a hand-written digit of 8×8 pixels in the center of a 28×28 -pixel image. The dataset consists of eight categories by value of the number in the center, chosen randomly from 1 to 8. (b) The "position" dataset was made using the following procedure. First, two digits of 8×8 pixels were randomly chosen. Second, these two digits were allocated in a 28×28 image, with one of the following position alignments: horizontal (top, middle, bottom), vertical (left, middle, right), or diagonal (45° , 135°). This dataset also consists of eight categories only depending on the positions of digits. (c) The "shape-position" dataset was made by the following procedure. First, a 28×28 -pixel area was divided into four 14×14 areas. Second, two diagonally aligned areas were selected (either 45° or 135°). Third, one of two digits, either "7" or

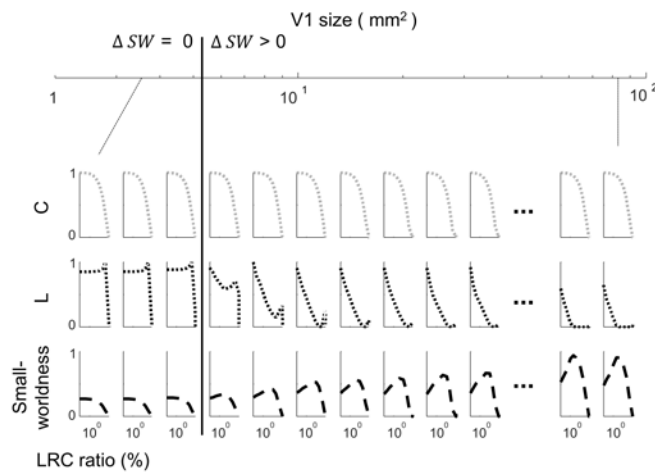
492 “9”, composed of 8×8 pixels was inserted into each selected area. This dataset consists of eight categories depending on
493 both the shape and position of the digits.



494
495 **Supplementary Fig.3. Length of lateral connections in tree shrews and mice**

496 Distribution of the length of lateral connections in tree shrews (red circles; Bosking 1997) and mice V1 (black circles; Seeman
497 2018), normalized by the size of each V1.

498



499
500 **Supplementary Fig.4. Changes in the network coefficients and small-worldness in relation to cortex size**

501 Change of the cluster coefficient (C), average shortest path length (L), and small-worldness during the rewiring process for V1
502 of various sizes.