Neural Networks 134 (2021) 76–85

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A brain-inspired network architecture for cost-efficient object recognition in shallow hierarchical neural networks

Youngjin Park [a,1], Seungdae Baek [a,1], Se-Bum Paik [a,b,*]

[a] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[b] Program of Brain and Cognitive Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

ARTICLE INFO

Article history:
Received 12 June 2020
Received in revised form 11 September 2020
Accepted 24 November 2020
Available online 28 November 2020

Keywords:
Visual cortex
Long-range horizontal connection
Object recognition
Shallow network
Artificial neural network
Cost-efficiency

ABSTRACT

The brain successfully performs visual object recognition with a limited number of hierarchical networks that are much shallower than artificial deep neural networks (DNNs) that perform similar tasks. Here, we show that long-range horizontal connections (LRCs), often observed in the visual cortex of mammalian species, enable such a cost-efficient visual object recognition in shallow neural networks. Using simulations of a model hierarchical network with convergent feedforward connections and LRCs, we found that the addition of LRCs to the shallow feedforward network significantly enhances the performance of networks for image classification, to a degree that is comparable to much deeper networks. We found that a combination of sparse LRCs and dense local connections dramatically increases performance per wiring cost. From network pruning with gradient-based optimization, we also confirmed that LRCs could emerge spontaneously by minimizing the total connection length while maintaining performance. Ablation of emerged LRCs led to a significant reduction of classification performance, which implies these LRCs are crucial for performing image classification. Taken together, our findings suggest a brain-inspired strategy for constructing a cost-efficient network architecture to implement parsimonious object recognition under physical constraints such as shallow hierarchical depth.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The current machine learning technique using the deep neural network (DNN) is able to implement high-accuracy visual object recognition comparable to that of humans (He et al., 2016; Krizhevsky, Sutskever, & Hinton, 2012; Lecun, Bengio, & Hinton, 2015; Simonyan & Zisserman, 2014; Zhang et al., 2019). However, even when the performance of a DNN appears to be similar to that of the brain in some visual tasks (Yamins & DiCarlo, 2016; Yamins et al., 2014), there is a critical difference between the two systems: the biological brain has physical constraints for circuit wiring. For example, an artificial neural network composed of model neurons (Rosenblatt, 1958) can create and annihilate inter-neural connections arbitrarily, whereas biological synapses cannot be controlled in such a way (Park, Choi, & Paik, 2017).

Above all, the brain is forced to consist of a limited number of layers and connections, due to physical constraints such as the spatial volume (Samu, Seth, & Nowotny, 2014), whereas artificial neural networks can implement an unlimited number of layers and connections as long as computational capacity allows. For example, ResNet (He et al., 2016), one of the state-of-the-art DNNs for object recognition, consists of more than 150 hierarchical layers, since the prodigious improvement of computing power allows the use of a super-deep structure. However, the ventral visual pathway (from the retina to IT) in the brain is composed of fewer than 10 hierarchical stages (Samu et al., 2014), presumably due to the restricted volume of the skull (Fig. 1(a)). What then is the distinct strategy of the brain as compared to the artificial network that enables cost-efficient visual processing under this physical constraint, and how can this be implemented in artificial neural networks?

We can find a clue from the anatomical structure observed in the visual cortex (Fig. 1(b) and (c)) of biological brains. Long-range horizontal connections (LRCs) are characteristic circuit structures observed in the visual cortex of various mammalian species including monkeys (Stettler, Das, Bennett, & Gilbert, 2002), cats (Das & Gilbert, 1995; Gibson & Wiesel, 1979), tree shrews (Bosking et al., 1997; Chisum, Mooser, & Fitzpatrick, 2003), ferrets (Durack & Katz, 1996; Van Hooser, Heimel, Chung, & Nelson, 2006), and rats (Rumberger, Tyler, & Lund, 2001). From their long range of connection (up to 2–3 mm), LRCs are distinguished from local connections of a short lateral spread (up to 0.5 mm) (Kisvárday & Eysel, 1992; Kisvárday et al., 1986; Malach, Amir, Harel, & Grinvald, 1993; Stepanyants, Martinez, & Ferecsko, 2009; Voges,

* Corresponding author.
E-mail address: sbpaik@kaist.ac.kr (S.-B. Paik).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.neunet.2020.11.013
0893-6080/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

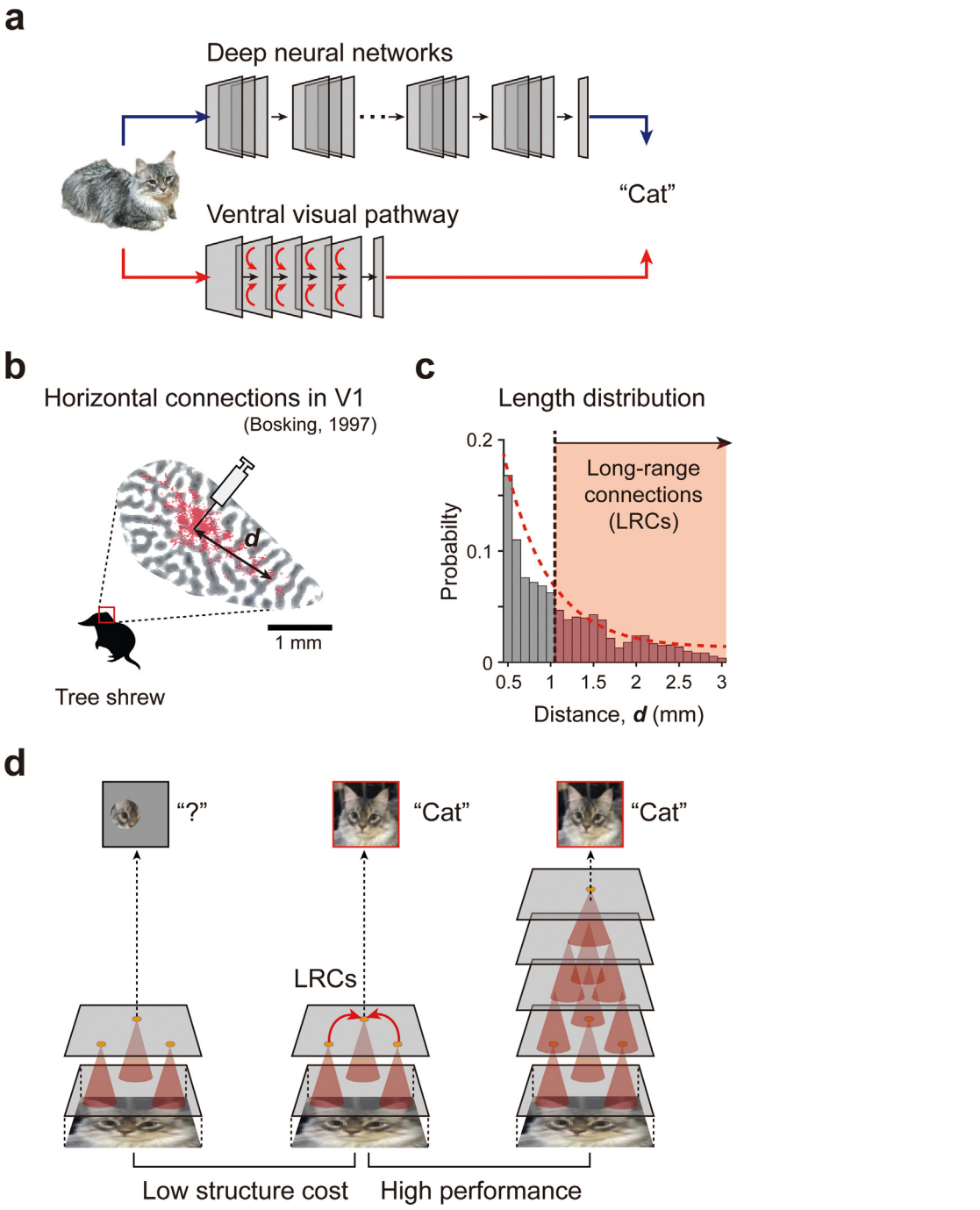

Fig. 1. Cortical long-range horizontal connectivity for cost-efficient circuit. (a) Artificial deep neural networks implement an excessive number of layers and connections as long as computational capacity allows, whereas the ventral visual pathway of the brain is composed of a much smaller number of hierarchies. (b) Long-range horizontal connections (LRCs) are observed in the visual cortex of various mammalian species. Red dots represent synaptic boutons labeled retrogradely. (c) Length of the lateral connections observed in tree shrews (Bosking et al., 1997). The red dashed line indicates an exponential fitting curve and the black solid line is the threshold for defining LRCs. (d) A scenario that the addition of LRCs to the shallow feedforward network may enhance the performance of networks for image classification, to a degree comparable to deeper networks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: (b) Redrawn from Bosking et al. (1997).
© 1997 Society for Neuroscience.

Schüz, Aertsen, & Rotter, 2010). Previous studies of the long-range connections have suggested their possible contribution to layer-wide functions, such as contextual modulation at an early stage (Angelucci et al., 2002; Capparelli, Pawelzik, & Ernst, 2019; Piech, Li, Reeke, & Gilbert, 2013). However, a comprehensive understanding of the exact role of LRCs for visual processing has been elusive.

Here, we suggest that the long-range connectivity may be crucial for cost-efficient visual processing under limited wiring resources (Fig. 1(d)). Using simulations of a model hierarchical network with convergent feedforward connections and LRC data from tree shrews, we investigated the contribution of the long-range connections to object recognition in shallow neural networks. First, we found that the addition of LRCs to the shallow feedforward network significantly enhances image classification

performance for MNIST and CIFAR-10 datasets to a degree that is comparable to much deeper networks. Second, we found that networks with sparse LRCs, similar to the networks in tree shrews, enable consistent recognition of various types of objects. Lastly, we found that LRCs can emerge spontaneously from network pruning for minimization of the total connection length while maintaining performance.

Taken together, the present results lead us to propose that long-range horizontal connectivity might be a key architecture to implement parsimonious object recognition under the physical constraints of the brain. Our results suggest that the brain develops its circuits with a biological strategy for balancing cost and performance, which may lead to the generation of a new cost-efficient artificial neural network architecture under structural constraints.

2. Materials and methods

2.1. Simplified three-layer network model

As a simplified model of the ventral visual pathway, we used a three-layer convergence network consisting of 784 input units, one hidden layer of 784 units, and ten output units for classification. Each layer is implemented as a 2D grid space, and the size of the input and hidden layer is set to 28 by 28 pixels, equivalent to the image size. For a biologically plausible model, in contrast to conventional CNN models, our model does not implement weight sharing. Instead, neurons in the hidden layer space have independent connection weights. The schematic architecture of the model is shown in Fig. 2(a), and a summary of the detailed parameters is provided in Table S1.

To match observed statistics in biological data, we assumed a condition that visual images were shown within 15-degrees of center vision in head-fixed tree shrews. Considering the magnification factor and the V1 size in tree shrews (Bosking et al., 1997; Van Hooser, 2005), we set the scaling factor as 0.125 mm/unit, so that one unit (pixel) of the input and hidden layers matches 0.125 mm in the cortical space of the biological data. For example, the definition of LRCs as lateral connections longer than 1 mm in the previous experiment (Trachtenberg & Stryker, 2001) was converted to that for model LRCs with connections longer than eight pixels. The definition for a local connection, which was defined as a connection shorter than 0.5 mm (Bosking et al., 1997), was converted to connections shorter than four pixels.

Feedforward connections between the input and hidden layers were set to a local convergence structure following previous studies (Hubel & Wiesel, 1962; Sailamul, Jang, & Paik, 2017). Each ith neuron in the hidden layer has a circular receptive field, that is,

$$\text{if } d_{ij} < \text{c}, \;\; w_{ij}^{b} = 1$$
$$otherwise, \;\; w_{ij}^{b} = 0$$

where j is the index of a neuron in the input layer and d_{ij} is the Euclidean distance between neurons i and j in the hidden layer and the input layer, respectively. c denotes the number of units in the input layer that is projected to the hidden layer by feedforward convergence, initially set to 4. w_{ij}^{b} is a boolean parameter that represents the presence of a connection between neurons i and j. Neurons in the hidden and output layers were randomly connected with a connection probability of 0.5. In this way, the total number of feedforward connections from the input to output layers was set to less than 3% of the fully-connected network, to simulate the condition that the wiring cost is tightly constrained.

The lateral connections were implemented only in the hidden layer as in previous model studies (Paik & Glaser, 2010; Ursino & La Cara, 2004; Yazdanbakhsh & Grossberg, 2004), and their length distribution was set to follow an exponential function fitted from V1 data observed in tree shrews (Bosking et al., 1997) (Fig. 1(b)),

$$Y = 0.4e^{-2.3d},$$

where d is the connection length (converted to unit) and Y is the connection probability. Once the total number of LRCs is determined, the weight matrix W_{lrc} is generated using the given length distribution. Here, all connection weights were initialized using a Gaussian distribution with a mean of zero and a standard deviation of 0.01. The overall network computation is as follows:

$$H = Max(0, W_{lrc}W_0X + W_0X + b_0)$$
$$O = softmax(W_1H + b_1)$$

In this equation, X is the input pixel value of the raw image, which is directly used for all computations without any preprocessing. H denotes the hidden layer activation, O is the output layer activation. The terms W_0 and W_1 are the feedforward (FF) connection matrix of each layer, b_0 and b_1 are the corresponding biases of the input and hidden layer, and W_{lrc} is the lateral connection matrix. Example connectivity matrices of model network, with and without LRCs, are shown in Fig. S1.

To perform image classification tasks, the model convergence network was trained using the adapted version of stochastic gradient descent in which the Boolean matrix w^b is multiplied to W_0, W_1 and W_{lrc} after every training epoch. Table 1 presents the hyper-parameters for network training. We set a batch size of 200 images and the weights of all learnable connections were initially selected from a Gaussian distribution of zero-mean with a standard deviation of 0.01. The learning rate was set to a constant of 0.1. The other hyper-parameters such as the number of epochs were chosen to provide an acceptable level of performance in the image classification task.

Table 1
Hyper-parameters of the network training.

Size of mini batch	200
Learning rate	0.1
STD of initial weight	0.01
Max. epoch in MNIST	100
Max. epoch in modified MNIST	300
Max. epoch in CIFAR-10	1000
Max. epoch in CIFAR-10 with length penalty	3000

2.2. Stimulus datasets

2.2.1. Modified MNIST

To separately examine the contribution of high and low-frequency information contained in sample images, we designed three types of modified MNIST datasets (Fig. S2), and 1000 test images were generated in each dataset. Details are as follows:

Type 1: shape. The "*shape*" dataset was designed by arranging a hand-written digit of 8 × 8 pixels in the center of a 28 × 28 pixels image. The dataset consists of eight categories depending on the number in the center (1 to 8). For this dataset, only local information (shape) of the digits is required for classification.

Type 2: position. The "*position*" dataset was made by 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 depending on the position only where the digits are located. Note that the shape of each number is irrelevant for classification.

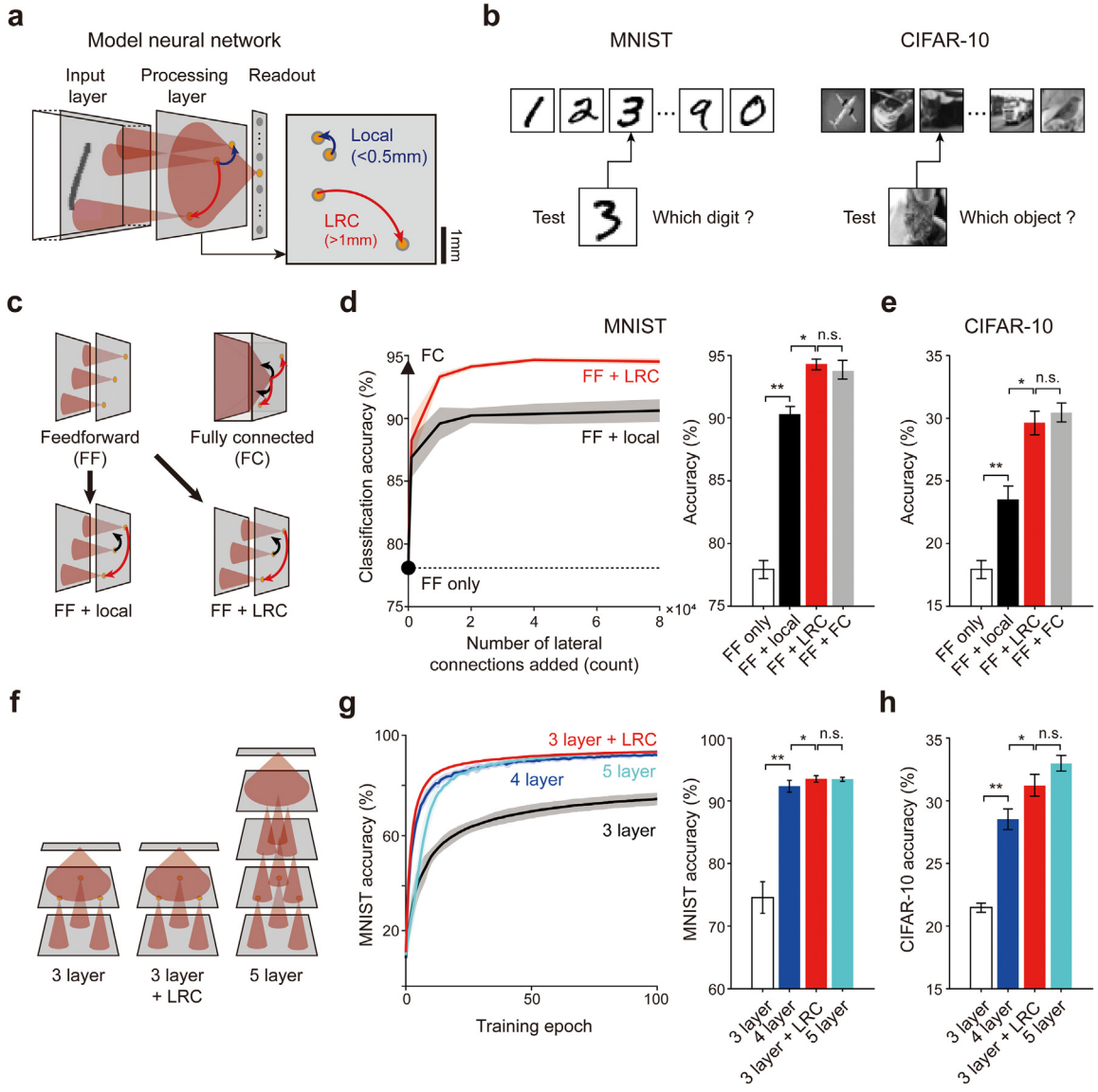

Fig. 2. Long-range connections enhance image classification of shallow networks. (a) Schematic structure of the simplified three-layer visual pathway model (*left*). The LRCs are defined as lateral connections longer than 1 mm and the local connections are set as connections shorter than 0.5 mm (*right*). (b) Illustration of the classification task using benchmark datasets (MNIST, CIFAR-10). Both the shape and position of the digits must be distinguished for successful classification. (c) Different conditions of the networks with various types of lateral connections. (d) The addition of LRCs to the initial feedforward network (red line) dramatically enhances the MNIST classification accuracy of the network. The performance was significantly higher when LRCs were added, compared to the case where the same number of local connections were added (black line). The black triangle represents the performance of the fully-connected network (FC) and the black circle indicates the initial feedforward network (FF) (*left*). When only a small portion (0.5% of FC) of connections were added as a form of LRCs to the initial feedforward network (red bar), the accuracy of the network became significantly higher than the FF only and FF+local networks (white and black bars, Mann–Whitney U-test: $^{*}p < 10^{-4}$, $^{**}p < 10^{-6}$), and comparable to the accuracy of the FC (gray bar) (*right*). (e) Performance for CIFAR-10 with the same network condition in (d) (white and black bars, Mann–Whitney U-test: $^{*}p < 10^{-4}$, $^{**}p < 10^{-6}$). (f) Networks of different depth (number of layers) and lateral connectivity compared in the test. (g) Classification accuracy of each type of the network (*left*). Performance of the three-layer network with LRCs (red curve) was comparable to, or even better than that of deeper networks (blue curve, four-layer; cyan curve, five-layer, Mann–Whitney U-test: $^{*}p < 10^{-3}$, $^{**}p < 10^{-6}$) (*right*). (h) Performance using CIFAR-10 with the same network condition in (g) (Mann–Whitney U-test: $^{*}p < 10^{-3}$, $^{**}p < 10^{-6}$). Shaded areas in (d) and (g), and error bars in (d–e) and (g–h) represent the standard deviation from 20 repeated trials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Type 3: shape-position. 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 "9", composed of 8 × 8 pixels was inserted into each selected area. This dataset consists of eight categories depending on both the shape and position of the digits. The reason we chose "7" and "9" among ten numbers is to adjust the difficulty of the task to be similar to that of the previous tasks. Note that this dataset requires both local information (shape) and global information (position) of the digits for classification.

2.2.2. Natural images

To further examine the role of the LRCs in image perception, natural images in the CIFAR-10 database (Sugawara & Nikaido, 2014) were used. The images were converted to grayscale using the MATLAB rgb2gray function.

2.3. Image classification performance with connectivity variation

2.3.1. The same number of connectivity

To examine changes in the image classification performance of the network according to the lateral connectivity structure,

we added the same number of either local or long-range connections to the initial feedforward network and compared the performance after training (Fig. S2(a)). The total number of learnable connections was counted by including both feedforward and lateral connections.

To examine the role of lateral connections (Fig. 3 and Fig. S5), we varied the ratio of lateral connections (either LRCs or local connections) in the network and tested the image classification performance after training of each dataset (*"shape"*, *"position"*, *"shape-position"*). To keep the number of learnable parameters the same while varying the network connectivity, we ablated feedforward connections between the input and hidden layer while adding the same number of lateral connections. The connections between the hidden and the output classification layer were not changed.

2.3.2. Same length of connectivity

The effects of local connections and LRCs with the same length were compared (Fig. S2(b)). The length of each connection was calculated as follows:

$$Length_n = \sum_{j,j'} d_{jj'}$$

where $d_{jj'}$ is the Euclidean distance of a lateral connection between source j and destination j' neurons.

2.4. Gradient-based connectivity optimization and pruning

We examined how random initial connectivity would change if the network was trained to minimize the total connection length while maximizing the image classification performance. The CIFAR-10 classification was used for the image recognition task. The objective function consists of a classification error minimization term and a connection length-penalty term as follows:

$$L = argmin_W \left((1-\lambda)\frac{1}{N}\sum_{i=1}^{N} L_i\left(f\left(x_i, W\right), y_i\right) + \lambda \sum_k \sum_l \frac{Len_{k,l} \otimes W_{k,l}^2}{2} \right)$$

where W is the weight matrix of the network, N is the number of connections, $Len_{k,l}$ represents the connection length matrix, k represents each layer, and λ represents the ratio between error minimization and length penalty. The value of λ was chosen so that small changes in both terms can modulate the loss function significantly. The stochastic-gradient descent method was applied for optimization. To simulate the spontaneous pruning of unnecessary connections during the training epochs, we disconnected connections with a weight value below 0.01 after the training.

3. Results

3.1. Long-range connections enhance object recognition in shallow networks

To test the object recognition ability of the shallow network with different types of lateral connections, a three-layer convergent feedforward network (Fig. 2(a)) was trained and tested with MNIST and CIFAR-10 image datasets for the classification task (Fig. 2(b)). Four different network conditions with various types of connectivity were modeled and tested (Fig. 2(c)) — feedforward only, feedforward with local connections, feedforward with LRCs, and fully connected.

First, we found that feedforward networks (FF) with no lateral connections showed much lower MNIST classification performance (Fig. 2(d), black circle) than that of the fully-connected (FC) network (Fig. 2(d) *left*, black triangle). However, adding only a small number of random LRCs to the feedforward network

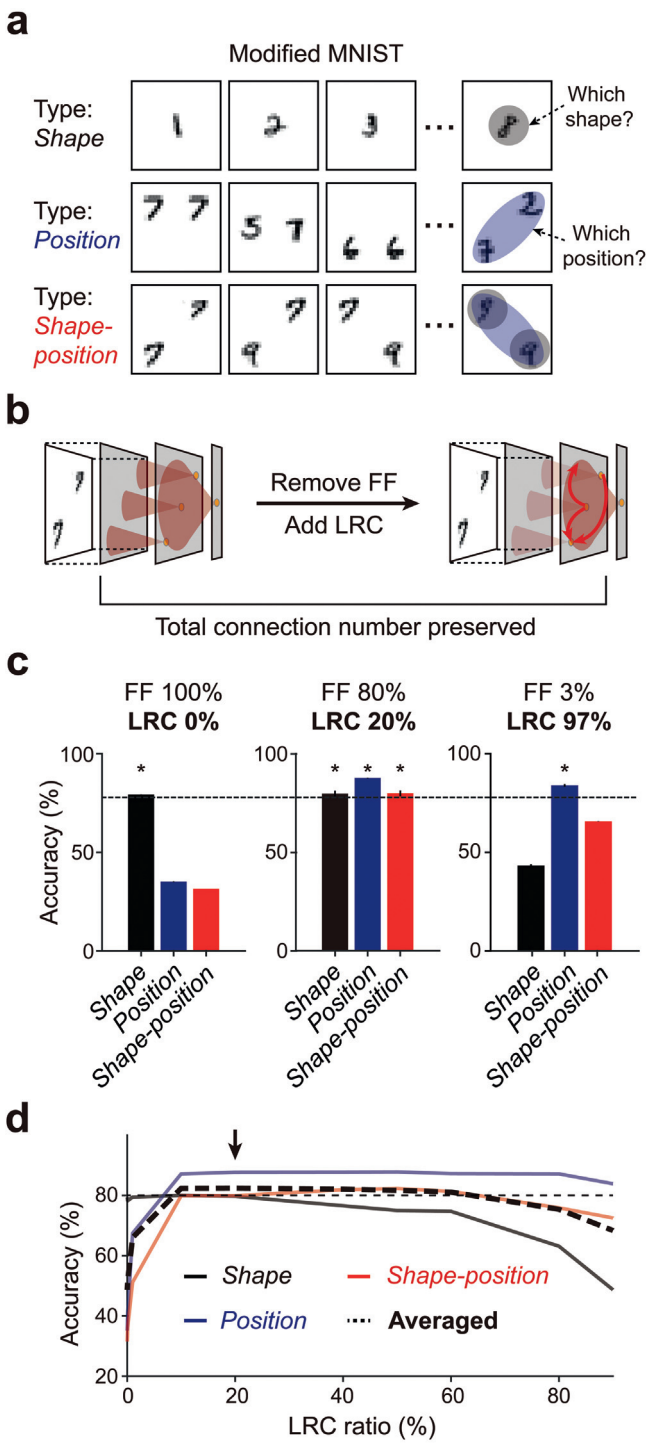

Fig. 3. Long-range connections integrate global features. (a) Three types of modified MNIST datasets. The *"shape"* dataset requires the local shape of digits to be distinguished, and the *"position"* dataset requires identification of the position of digits. The *"shape-position"* dataset requires both. (b) Network rewiring process. By deleting feedforward (FF) connections and adding the same number of LRCs, the total number of connections are preserved. (c) Classification performance for each dataset with a variation of the LRC ratio. The dotted line indicates 80% accuracy and asterisks on the bar graph indicate accuracy significantly higher than 80%. (d) The average performance of the network for all three tasks (black dotted line). Average performance is maximized when sparse LRCs (∼20%) are combined with dense local convergent connections. Error bars represent the standard deviation from 20 repeated trials.

Table 2
Classification accuracy of networks adding the same length of local and 682 long-range connections.

	LRC added	Local added
MNIST	93.32 ± 0.30%	90.62 ± 0.91%
CIFAR-10	27.78 ± 1.17%	24.12 ± 1.18%

Table 3
Classification accuracy of networks using subsets of CIFAR-10 with distinct 686 frequency spectra.

	High frequency dominant subset	Low frequency dominant subset
LRC 0%	28.97 ± 0.59%	28.59 ± 1.01%
LRC 97%	30.71 ± 1.15%	35.99 ± 0.94%

dramatically improved the classification performance up to that in the FC (Fig. 2(d) *left*, red solid line, and Fig. 2(d) *right*, red bars, Mann–Whitney U-test: $^{*}p < 10^{-4}$, $^{**}p < 10^{-6}$). In contrast, the addition of local connections only could not induce the same amount of change (Fig. 2(d) *left*, black solid line, and Fig. 2(d) *right*, black bars). The simulation using natural images (CIFAR-10) also show a similar tendency (Fig. 2(e)).

Next, we investigated whether this difference in performance is due to the longer total connections of LRCs when the same number of local and long-range connections were used (Fig. S3(a)). We examined the influence of lateral connections on the classification accuracy when local and long-range connections with the same total length were added to the feedforward network (Fig. S3(b)). Even in this condition, adding LRCs resulted in noticeably higher performance than adding the same length of local connections, despite that the number of added connections was sixteen times smaller when adding connections in the form of LRCs (Table 2; Fig. S3 (b) *left* (MNIST) and *right* (CIFAR-10), black and red triangles, Mann–Whitney U-test: $^{*}p < 10^{-4}$). Furthermore, from the variation of feedforward connectivity conditions, we confirmed that the performance improvement by LRCs appeared to be fairly consistent across different conditions of feedforward convergence range and sparseness of connections (Fig. S4).

Moreover, we found that the shallow three-layer network with LRCs shows comparable, or even better performance for the MNIST to that of deeper (four or five-layer) networks with much fewer learnable connections (Fig. 2(f) and (g)). The experiment using natural images (CIFAR-10) also showed similar results (Fig. 2(h)). These results indicate that long-range horizontal connectivity enables highly cost-effective object recognition in shallow neural networks.

3.2. Long-range connections integrate global features

We set another hypothesis that the LRCs enhance the network performance by integrating global features contained in the images that cannot be captured by local connections (Fig. 1(d)). To examine the spatial frequency-specific contribution of the LRCs, we designed three types of datasets by arranging various combinations of MNIST images (Fig. 3(a), *"shape", "position", and "shape-position"*), which requires the network to encode local features (high frequency), global features (low frequency), or both components for classification tasks. In addition, to investigate the contribution of the LRCs, we varied the ratio of the LRCs and local feedforward connections in model networks. We added the LRCs while disconnecting the same number of local feedforward connections, so that an equal number of learnable connections are kept (Fig. 3(b)). We then trained the network and measured the classification performance while varying the ratio of the LRCs in model networks.

As expected, a feedforward network without the LRCs showed the highest performance for *"shape"* datasets, but it could not classify the two other datasets that contained *"position"* information well (Fig. 3(c), *left*). A network with a large number of LRCs and sparse feedforward connections was superior to identify the *"position"* datasets, but was not effective at classifying the dataset with *"shape"* information (Fig. 3(c), *right*). Our additional tests show that local lateral connections added to feedforward networks can play a role similar to that of LRCs only when the given stimulus barely contains global features. In general, networks with local connections cannot achieve a level of performance accuracy similar to that of networks with LRCs (Fig. S5). These results indicate that local convergent connections contribute to encoding high-frequency (local) information such as the shape of the digits, and long-range connections are responsible for encoding low-frequency (global) information such as the position of objects. Simulations using natural images with different frequency spectra also showed similar results (Table 3 and Fig. S6).

Most importantly, the network with dense local feedforward connections and sparse LRCs showed good performance for all three datasets (Fig. 3(c), *middle*). The average performance for the three datasets was maximized around the point where the number of LRCs is ∼20% of dense local connections (Fig. 3(d), black arrow). Taken together, the combination of sparse LRCs and dense local convergent connections is highly effective for the classification of images with a wide-ranging frequency spectrum.

3.3. Long-range connections for balancing performance and wiring cost

Lastly, we tested whether LRCs can emerge spontaneously from random initialization, as a result of balancing performance and wiring cost. For this, we trained a model network with a new loss function, composed of the classification error and the connection length-penalty terms (Fig. 4(a)). Each term represents the object recognition performance and the total wiring cost of the network, respectively. The balancing ratio factor λ regulates the relative weight between the two terms. From an initial parameter search, we found the optimal range of λ for the current purpose, in which the performance-wiring cost ratio is maximized (Fig. S7). Under this condition, we observed that the classification performance started increasing in just several epochs of training, while the total connection length decreased as expected from the loss function. Note that there was no noticeable degradation of the final performance even though the length penalty term was added to the loss function (Fig. 4(b)). Interestingly, even when initial connections were mostly pruned to reduce wiring cost, a certain portion of very long connections survived until the end of the training, accepting a notable length penalty (Fig. 4(c) and (d)). We observed that the ratio of LRCs sharply dropped at the early stage of training, but eventually converged to a nonzero value in most conditions (Fig. 4(d)). We also confirmed that the emerged connectivity showed tuning-specific properties as observed in the visual cortex of various mammalian species (Fig. S8). This suggests that LRCs can emerge spontaneously from a randomly wired initial condition if both the performance and wiring cost are taken into account in the objective function.

We also investigated how the observed performance changes if connections in the trained network are removed by random ablation (Fig. 4(e)). A significant reduction of classification performance was observed when the developed LRCs were removed. In contrast, the performance degradation was noticeably smaller when only local connections were removed (Fig. 4(e), Mann–Whitney U-test: $^{*}p < 10^{-6}$), indicating that LRCs are crucial for performing image classification.

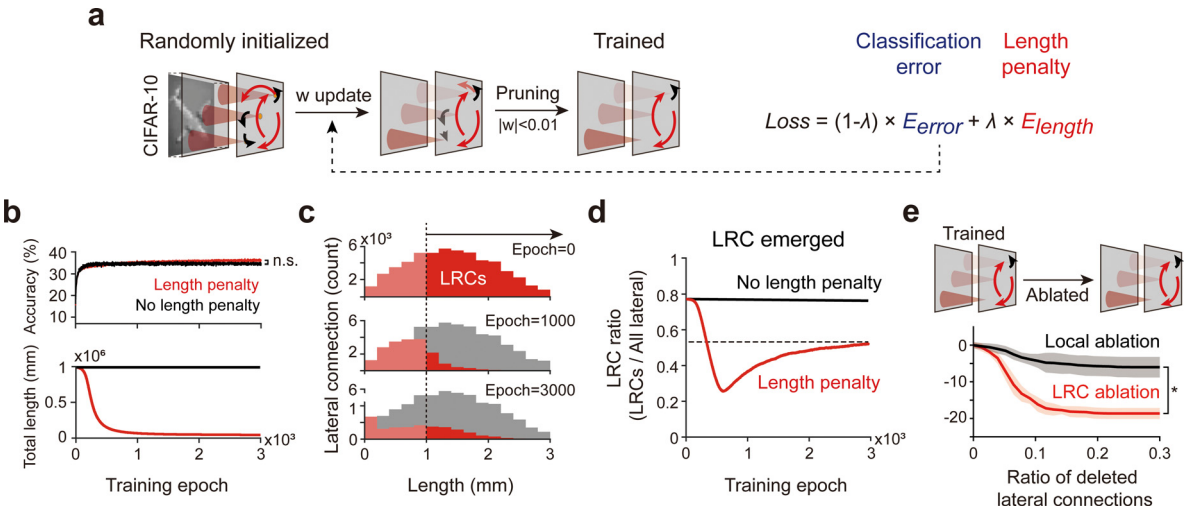

Fig. 4. LRCs emerge spontaneously while optimizing performance and wiring cost. (a) The network was trained with the objective function to minimize classification error and the total connection length. Connections weakened below the threshold ($|w| < 0.01$) were pruned during the training. (b) Changes of classification accuracy and the total connection length during the training (the balancing factor $\lambda = 10^{-3}$). (c) Changes in the length distribution of lateral connections during the training (red histogram, with length penalty; gray histogram, no length penalty). A certain portion of LRCs survives after training even with the length penalty (Epoch = 3000, red histogram). (d) The ratio of LRCs converged to the nonzero value (red curve) with the length penalty. (e) Ablation of the LRCs leads to a significantly larger reduction in classification performance, compared to the case where local connections are deleted (Mann–Whitney U-test: $^*p < 10^{-6}$). The shaded area and error bars represent the standard deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

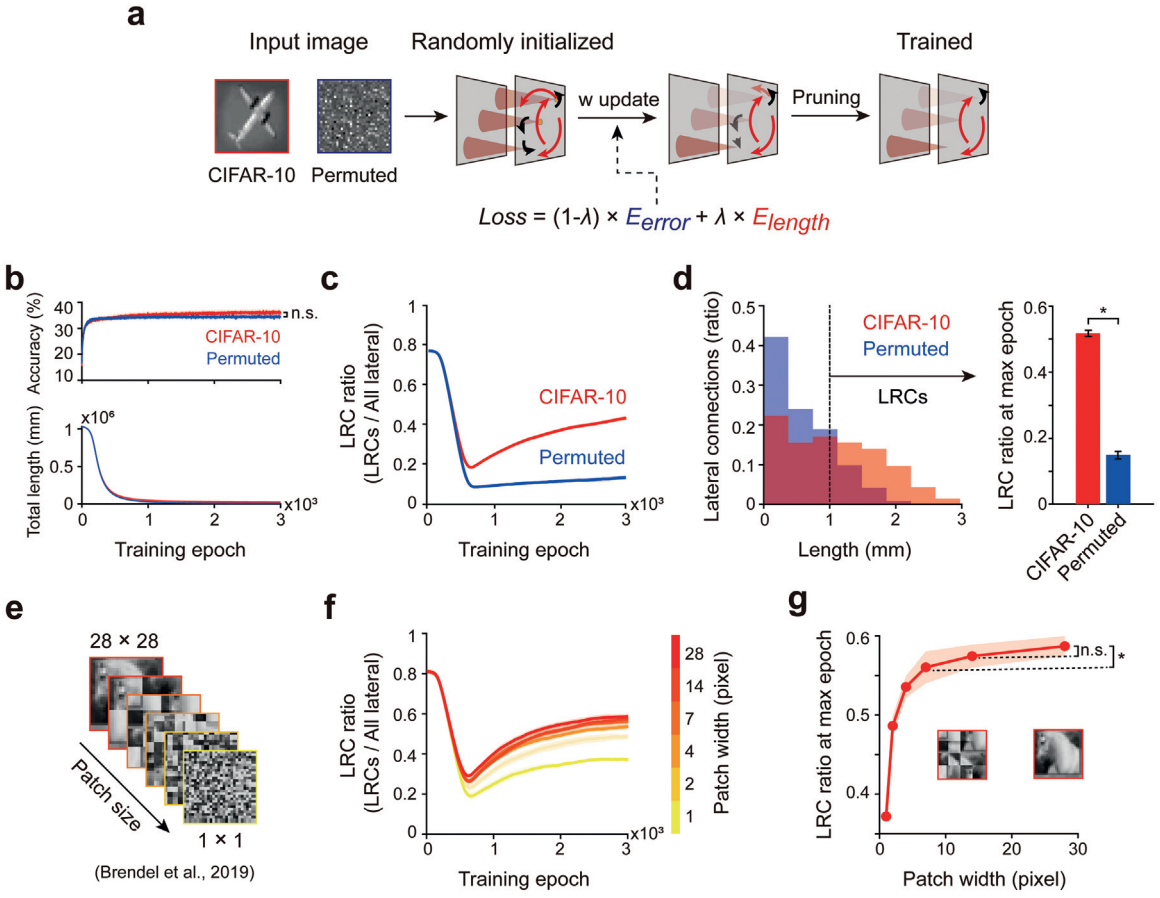

Fig. 5. The development of LRCs depends on the characteristics of visual stimuli. (a) Natural images from the CIFAR-10 dataset randomly permuted images of them were used for training. (b) The classification performance and the total wiring length appeared to be not meaningfully different between the two conditions. (c) Changes in the LRC ratio during the training (red line, CIFAR-10; blue line, permuted images). (d) Developed lateral distribution after the training (3000 epochs). Emerged LRCs for the permuted images were noticeably fewer than that for natural images (*right*, Mann–Whitney U-test: $^*p < 0.0005$). (e) Texturized images as inputs for training. Local patches (local information conserved) were maintained while their locations were randomized (global information scrambled) (Brendel & Bethge, 2019). (f) Changes in the LRC ratio during the training. Each color represents the patch width of the image that used for training. (g) The ratio of LRCs developed for the texturized images was significantly lower than that for the natural images (patch size = 7 × 7 and 28 × 28 pixels; Mann–Whitney U-test: $^*p < 0.05$). The shaded area and error bars represent the standard deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We then tested whether the development of LRCs is dependent on the characteristics of visual stimuli, by comparing results obtained when natural images from CIFAR-10 and pixel-by-pixel randomly permuted images of these images were used for training (Fig. 5(a)). We observed that the average performance of image classification was not different between these two conditions (Fig. 5(b)), but the developed LRCs for random images were noticeably fewer than those for the natural image training set (Fig. 5(c) and (d), *left*). The ratio of LRCs in the natural image training set was significantly higher than that in the permuted image set (Fig. 5(d), *right*, Mann–Whitney U-test: $^{*}p < 0.0005$). This suggests that the characteristics of visual stimuli can contribute critically to the development of lateral connectivity patterns. To test this scenario further, we used texturized images (Brendel & Bethge, 2019) as inputs for training (Fig. 5(e)). In this case, local patches (local information conserved) were maintained while their locations were randomized (global information scrambled). We found that the developed LRC ratio gradually decreased as the patch size decreased from 28×28 to 1×1 (Fig. 5(f)). For example, the ratio of LRCs developed for the texturized images (patch size $= 7 \times 7$ pixels) was significantly lower than that for the original images (Fig. 5(g), Mann–Whitney U-test: $^{*}p < 0.05$). These results further support the idea that the development of LRCs is particularly dependent on the global organization of the input image sets.

4. Discussion

In this study, we showed that a network with long-range connections can perform cost-efficient image classification. We first found that the addition of LRCs into the locally convergent feedforward network could dramatically enhance image perception performance. This dramatic increment was only observed when the training image included informative low-frequency components, which implies that LRCs contribute to improving object recognition by integrating global information contained in the input image. We further confirmed that LRCs could emerge spontaneously from randomly initialized networks through an optimization process with an objective function comprising both performance and wiring cost terms.

Considering that natural images contain a significant amount of low-frequency components (Simoncelli & Olshausen, 2001), a successful visual recognition system must capture these informative "global" correlations in visual images (Turner, Sanchez Giraldo, Schwartz, & Rieke, 2019). In a network that uses sufficiently deep layers and a large number of connections, the size of the receptive field gradually increases moving up in the hierarchy (Luo, Li, Urtasun, & Zemel, 2016). Therefore, even if receptive fields are localized in lower layers, the higher layer can sufficiently encode global information through deep structures of the network (Zeiler & Fergus, 2014). However, presumably due to the volume limitation, the brain's visual pathway is not deep enough, making it difficult to capture global information effectively with a local receptive field. Enlarging receptive fields in initial layers is not an appropriate solution because they cannot precisely capture high-frequency "local" information. Here, we showed that the combination of dense local convergent connections and a small number of long-range connections may address this issue.

Throughout this study, we performed simulations using convergent network models instead of convolutional neural networks (CNNs). Indeed, deep CNNs with recurrent connections can achieve much higher classification performance than the current model (Liang & Hu, 2015; Linsley, Kim, Veerabadran, Windolf, & Serre, 2018; Montobbio, Bonnasse-Gahot, Citti, & Sarti, 2019; Spoerer, McClure, & Kriegeskorte, 2017; Tang et al., 2018). However, a direct comparison of the current model with CNNs is inadequate due to the intrinsic differences in the network structures between the two. A convergence network model is implemented in the current study to provide a more biologically inspired architecture, including localized receptive fields of neurons as well as localized convergent feedforward projections with a weight distribution having considerable variance across the network space. In contrast, CNN models have artificial components designed for model simplicity and to reduce the computational cost, such as weight sharing, which are not directly comparable to biological equivalents.

Nevertheless, it would be interesting to suggest that LRCs enhance the cost-efficiency of the network by playing a role similar to that of weight sharing in a CNN. LRCs reduce the structural and computational cost of the network by removing redundant connections and adding shortcuts. Importantly, from "like-to-like" connections between similarly tuned units that exist periodically across the cortical surface (Bosking et al., 1997; Voges et al., 2010), LRCs generate periodic patches of functional modules that facilitate communication and the sharing of global information at different locations. This may have an effect similar to that of weight sharing by convolution in CNNs such that biological networks show enhanced performance even with a constrained circuit structure and lower energy consumption.

There have been similar approaches in model neural network studies (Liang & Hu, 2015; Linsley et al., 2018; Montobbio et al., 2019; Spoerer et al., 2017; Tang et al., 2018), in which lateral connections are added to model networks to improve the performance on visual tasks such as object recognition. However, while these models implemented lateral connections that were localized (<1 mm) and uniformly distributed in most cases, our proposed model pays attention to the distinct characteristics of LRCs in terms of the connectivity organization — extraordinarily long, sparse, and tuning-specific connectivity. It is commonly assumed that such long connections, which require a significant wiring cost, may develop due to their important functional role, which compensates for their high wiring cost. In the current study, we explore a scenario in which the network can take advantage of this distinct organization of LRCs. We show that "long and sparse" connections have higher cost-efficiency than "short and uniform" lateral connections and that this tuning-specific connectivity in the network can be reproduced spontaneously from circuit optimization for cost-efficiency. This distinguishes our current work from other studies of the roles of local lateral connections.

Lateral inhibition, often implemented by lateral connections, is an important factor, the function of which can be compared to that of the feedback function implemented by LRCs, considering that lateral inhibition can also improve object recognition by enhancing the contrast and sharpening visual signals (Roska, Nemeth, Orzo, & Werblin, 2000; Yantis & Abrams, 2014). An important difference is that lateral inhibition only works by suppressing the activities of neighboring neurons, while LRCs can contribute with both excitatory and inhibitory interactions. In addition, the effective range of lateral inhibition is relatively small, implying that it can control the gain function of a small, local patch in most cases. In contrast, LRCs can induce interactions between neurons that are far apart from each other. Lastly, lateral inhibition affects neighboring neurons uniformly or by simple rule, such as a decaying function of weight according to the distances between neurons. LRCs, on the other hand, make more specific pin-point connections between neurons depending on their tuning similarity. Overall, these distinct characteristics between lateral inhibition and LRC feedback suggest that they play different roles in information processing, which may together contribute to the consistent visual recognition of various types of stimulus information.

In the current study, we took a bi-directional approach to investigate the role of LRCs, that is, one that involved both goal-oriented and stimulus-oriented analysis. A goal-oriented framework is a top-down approach to describe the brain structure from the perspective of implementing higher-level functions. Examples of a goal-oriented approach can be found in the previous pioneering works comparing deep CNN models and the visual pathway (Bashivan, Kar, & DiCarlo, 2019; Yamins et al., 2014). This goal-oriented approach provides a plausible explanation of the functions of certain structures (i.e., receptive fields for invariant object recognition). However, it cannot explain why those particular structures (e.g. Gabor filter) are found after training of the network, and whether they are indeed the structures that most effectively process the given stimulus sets (e.g. a natural image). In contradistinction to the goal-oriented approach, the stimulus-oriented framework theoretically finds the brain structure that most effectively processes the given features of the stimulus. An example of a stimulus-oriented approach is the efficient tiling hypothesis (Nauhaus, Nielsen, & Callaway, 2016; Song, Jang, Kim, & Paik, 2018). Because stimulus-oriented frameworks also have limitations in explaining how high-level cognitive functions emerge, both stimulus-oriented and goal-oriented approaches should be used together to understand the system thoroughly. In our analyses using various types of controlled stimuli, we showed that high- and low-frequency components contained in the natural image could be processed by local convergence and long-range horizontal connections, respectively. Together with this bottom-up approach, we also tried a top-down approach by which we examined the optimized connectivity for efficient encoding of a natural image to confirm that the LRCs are developed as a consequence of optimization. Demonstrating either one of these results may not be sufficient to draw conclusions on the role of LRCs, but taken together, our results could provide evidence that the LRCs integrate global features to perceive natural images.

5. Conclusion

Overall, the present results indicate that long-range connectivity added to local feedforward projections might be a key architecture of shallow networks under physical constraints for parsimonious object recognition. Our results suggest a biologically-inspired strategy for balancing cost and performance, which may inspire a new cost-efficient artificial neural network architecture.

CRediT authorship contribution statement

Youngjin Park: Designed the model, Performed the simulations, Analyzed the data, Wrote the manuscript, Discussed and commented on the manuscript. **Seungdae Baek:** Designed the model, Performed the simulations, Analyzed the data, Wrote the manuscript, Discussed and commented on the manuscript. **Se-Bum Paik:** Conceived the project, Designed the model, Wrote the manuscript, Discussed and commented on the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and code that support the findings of this study are available from the authors upon request.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2019R1A2C4069863, NRF-2019M3E5D2A01058328) (to S.P.).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neunet.2020.11.013.Supplemental figures and legends are available in Supplemental Information.

References

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupe, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, *22*(19), 8633–8646.

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science*, *364*(6439), eaav9436.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, *17*(6), 2112–2127.

Brendel, W., & Bethge, M. (2019). Approximating CNNS with bag-of-local-features models works surprisingly well on imagenet. ArXiv Preprint ArXiv: 1904.00760.

Capparelli, F., Pawelzik, K., & Ernst, U. (2019). Constrained inference in sparse coding reproduces contextual effects and predicts laminar neural dynamics. *PLoS Computational Biology*, *15*(10), Article e1007370.

Chisum, H. J., Mooser, F., & Fitzpatrick, D. (2003). Emergent properties of layer 2/3 neurons reflect the collinear arrangement of horizontal connections in tree shrew visual cortex. *The Journal of Neuroscience*, *23*(7), 2947–2960.

Das, A., & Gilbert, C. D. (1995). Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. *Nature*, *375*(6534), 780–784.

Durack, J. C., & Katz, L. C. (1996). Development of horizontal projections in layer 2/3 of ferret visual cortex. *Cerebral Cortex*, *6*(2), 178–183.

Gibson, J. R., & Wiesel, T. N. (1979). Morphology and intracortical projections of functionally identified neurons in cat visual cortex. *Nature*, *280*(5718), 120–125.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* (pp. 770–778).

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal Physiology*, *160*(38), 106–154.

Kisvárday, Z. F., & Eysel, U. T. (1992). Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (area 17). *Neuroscience*, *46*(2), 275–286.

Kisvárday, Z. F., Martin, K. A. C., Freund, T. F., Maglóczky, Z., Whitteridge, D., & Somogyi, P. (1986). Synaptic targets of HRP-filled layer III pyramidal cells in the cat striate cortex. *Experimental Brain Research*, *64*(3), 541–552.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*: *Vol. 25*, (pp. 1097–1105).

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Liang, Ming, & Hu, Xiaolin (2015). Recurrent convolutional neural network for object recognition. In *2015 IEEE conference on computer vision and pattern recognition* (pp. 3367–3375). IEEE.

Linsley, D., Kim, J., Veerabadran, V., Windolf, C., & Serre, T. (2018). Learning long-range spatial dependencies with horizontal gated recurrent units. In *Advances in neural information processing systems* (pp. 152–164).

Luo, W., Li, Y., Urtasun, R., & Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 4898–4906).

Malach, R., Amir, Y., Harel, M., & Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proceedings of the National Academy of Sciences*, *90*(22), 10469–10473.

Montobbio, N., Bonnasse-Gahot, L., Citti, G., & Sarti, A. (2019). KerCNNs: Biologically inspired lateral connections for classification of corrupted images. ArXiv Preprint ArXiv:1910.08336.

Nauhaus, I., Nielsen, K. J., & Callaway, E. M. (2016). Efficient receptive field tiling in primate V1. *Neuron*, *91*(4), 893–904.

Paik, S.-B., & Glaser, D. A. (2010). Synaptic plasticity controls sensory responses through frequency-dependent gamma oscillation resonance. *PLoS Computational Biology*, *6*(9), Article e1000927.

Park, Y., Choi, W., & Paik, S. B. (2017). Symmetry of learning rate in synaptic plasticity modulates formation of flexible and stable memories. *Scientific Reports*, *7*(1), 5671.
Piech, V., Li, W., Reeke, G. N., & Gilbert, C. D. (2013). Network model of top-down influences on local gain and contextual interactions in visual cortex. *Proceedings of the National Academy of Sciences*, *110*(43), E4108–E4117.
Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.
Roska, B., Nemeth, E., Orzo, L., & Werblin, F. S. (2000). Three levels of lateral inhibition: A space–time study of the retina of the tiger salamander. *Journal of Neuroscience*, *20*(5), 1941–1951.
Rumberger, A., Tyler, C. J., & Lund, J. S. (2001). Intra- and inter-areal connections between the primary visual cortex V1 and the area immediately surrounding V1 in the rat. *Neuroscience*, *102*(1), 35–52.
Sailamul, P., Jang, J., & Paik, S. B. (2017). Synaptic convergence regulates synchronization-dependent spike transfer in feedforward neural networks. *Journal of Computational Neuroscience*, *43*(3), 189–202.
Samu, D., Seth, A. K., & Nowotny, T. (2014). Influence of wiring cost on the large-scale architecture of human cortical connectivity. *PLoS Computational Biology*, *10*(4), Article e1003557.
Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216.
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. ArXiv Preprint ArXiv:1409.1556.
Song, M., Jang, J., Kim, G., & Paik, S. B. (2018). Universality of the developmental origins of diverse functional maps in the visual cortex. BioRxiv, 354126.
Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent convolutional neural networks: A better model of biological object recognition. *Frontiers in Psychology*, *8*, 1551.
Stepanyants, A., Martinez, L. M., & Ferecsko, A. S. (2009). The fractions of short- and long-range connections. *Proceedings of the National Academy of Sciences*, *106*(9), 3555–3560.
Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*(4), 739–750.
Sugawara, E., & Nikaido, H. (2014). Properties of AdeABC and AdeIJK efflux systems of acinetobacter baumannii compared with those of the AcrAB-TolC system of *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, *58*(12), 7250–7257.
Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., ... Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(35), 8835–8840.
Trachtenberg, J. T. T., & Stryker, M. P. P. (2001). Rapid anatomical plasticity of horizontal connections in the developing visual cortex. *The Journal of Neuroscience*, *21*(10), 3476–3482.
Turner, M. H., Sanchez Giraldo, L. G., Schwartz, O., & Rieke, F. (2019). Stimulus- and goal-oriented frameworks for understanding natural vision. *Nature Neuroscience*, *22*(1), 15–24.
Ursino, M., & La Cara, G. E. (2004). A model of contextual interactions and contour detection in primary visual cortex. *Neural Networks*, *17*(5–6), 719–735.
Van Hooser, S. D. (2005). Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. *Journal of Neuroscience*, *25*(1), 19–28.
Van Hooser, S. D., Heimel, J. A., Chung, S., & Nelson, S. B. (2006). Lack of patchy horizontal connectivity in primary visual cortex of a mammal without orientation maps. *Journal of Neuroscience*, *26*(29), 7680–7692.
Voges, N., Schüz, A., Aertsen, A., & Rotter, S. (2010). A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Progress in Neurobiology*, *92*(3), 277–292.
Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365.
Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624.
Yantis, S., & Abrams, R. A. (2014). *Sensation and perception*. NY: Worth Publishers New York.
Yazdanbakhsh, A., & Grossberg, S. (2004). Fast synchronization of perceptual grouping in laminar visual cortical circuits. *Neural Networks*, *17*(5–6), 707–718.
Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818–833). Springer.
Zhang, Q., Zhang, M., Chen, T., Sun, Z., Ma, Y., & Yu, B. (2019). Recent advances in convolutional neural network acceleration. *Neurocomputing*, *323*, 37–51.