scientific reports

Check for updates

OPEN

Acquisition of similar properties by filters in the same stream of a multistream convolutional neural network

Hiroshi Tamura[1,2,3]✉

Functional modular organization is observed in a variety of cortical areas in the brain. In the visual cortex of primates, adjacent neurons often respond to the same visual submodality, such as color or orientation, and have a similar preferred orientation or preferred color. However, it remains unclear why functional modular organization emerges in the cerebral cortex. In the present study, I constructed and trained a multistream convolutional neural network to examine whether filters in the same stream acquire similar properties. Although filters in the same stream were able to develop any structures, they acquired similar degrees of orientation and color selectivity and preferred similar orientations and colors. The deletion of filters in a single stream that had similar degrees of stimulus selectivity resulted in larger decreases in classification accuracy than the deletion of those that did not. By contrast, the deletion of filters in a single stream that shared a preferred stimulus resulted in similar decreases in classification accuracy to the deletion of those that did not. Together, these findings suggest that filters with similar degrees of stimulus selectivity in the same stream are required for optimal task performance of the multistream convolutional neural network, and probably of the brain.

In the cerebral cortex, adjacent neurons often share response properties, and clusters of neurons that share response properties are elongated vertically and form functional columns[1]. In the primary visual cortex of primates, neighboring neurons often encode similar types of visual submodality and form orientation or color modules[2–4]. In addition, neighboring neurons often share stimulus preferences, such as preferred orientation and preferred color[4,5].

One possible reason for the appearance of functional modular organization involves an anatomical constraint. Because adjacent neurons are likely to receive inputs from the same axons[6], they are likely to share response properties and form functional modular organizations. However, in the rodent visual cortex, adjacent neurons are tuned to a variety of stimulus parameters[7], suggesting that such an anatomical constraint may not explain the appearance of functional modular organization.

Another possible reason for the appearance of functional modular organization involves computational requirements, because modular design has several computational advantages[8]. However, the importance of functional modular organization in the brain remains unclear[9–11]. For example, the finding from a comparison of animals with and without ocular dominance columns questioned the importance of functional modular organization based on stimulus preference[10]. Furthermore, the importance of modular organization based on visual submodality, such as orientation or color modules, also remains unclear.

In the present study, I constructed a multistream convolutional neural network (CNN) to explore whether filters in the same stream of the first convolutional layer (c1) acquire similar properties. The underlying rationale was as follows: if c1 filters in the same stream acquire similar properties in a multistream CNN, then the presence of similar types of filters in the same stream of c1 is required for information processing. CNNs are hierarchically organized feed-forward networks that consist of multiple sets of layers. Each set of layers undergoes convolution, thresholding, and pooling[12]. The filter weights for convolution were initially set to random values and modified during training. In the present study, because no constraints were placed on inputs to c1 filters in the same stream of multistream CNN, these filters were—in principle—able to develop any structures. Nonetheless, c1 filters in the same stream tended to encode similar types of visual submodality and preferred similar stimulus

[1]Graduate School of Frontier Biosciences, The University of Osaka, Suita, Osaka 565-0871, Japan. [2]Center for Information and Neural Networks, Suita, Osaka 565-0871, Japan. [3]Cognitive Neuroscience Group, Graduate School of Frontier Biosciences, The University of Osaka, 1-4 Yamadaoka, Suita, Osaka 565-0871, Japan. ✉email: tamura.hiroshi.fbs@osaka-u.ac.jp

parameters. Moreover, the deletion of c1 filters in the same stream that shared visual submodality types resulted in a larger decrease in classification accuracy than the deletion of those that did not. By contrast, the deletion of c1 filters in the same stream that preferred similar stimulus parameters resulted in a similar degree of decrease in classification accuracy to the deletion of those that did not. These results suggest that c1 filters with a similar degree of stimulus selectivity in the same stream are required for the optimal task performance of multistream CNN.

Results

A multistream CNN with convergence (mcAlexNet; Fig. 1) was constructed based on AlexNet[12]. mcAlexNet consisted of five hierarchically organized convolutional layers (c1–c5) and three pooling layers (Max-pool). c1 contained 32 parallel streams (s0–s31), each equipped with two filters. c2 contained 16 parallel streams (s0–s15), each equipped with 12 filters. Each c2 filter received converging inputs that came exclusively from four filters in two streams of c1. c3 contained eight parallel streams (s0–s7), each equipped with 48 filters. Each c3 filter received converging inputs that came exclusively from 24 filters in two streams of c2. c4 contained four parallel streams (s0–s3), each equipped with 64 filters. Each c4 filter received converging inputs that came exclusively from 96 filters in two streams of c3. Finally, c5 contained two parallel streams (s0 and s1), each equipped with 128 filters. Each c5 filter received converging inputs that came exclusively from 128 filters in two streams of c4.

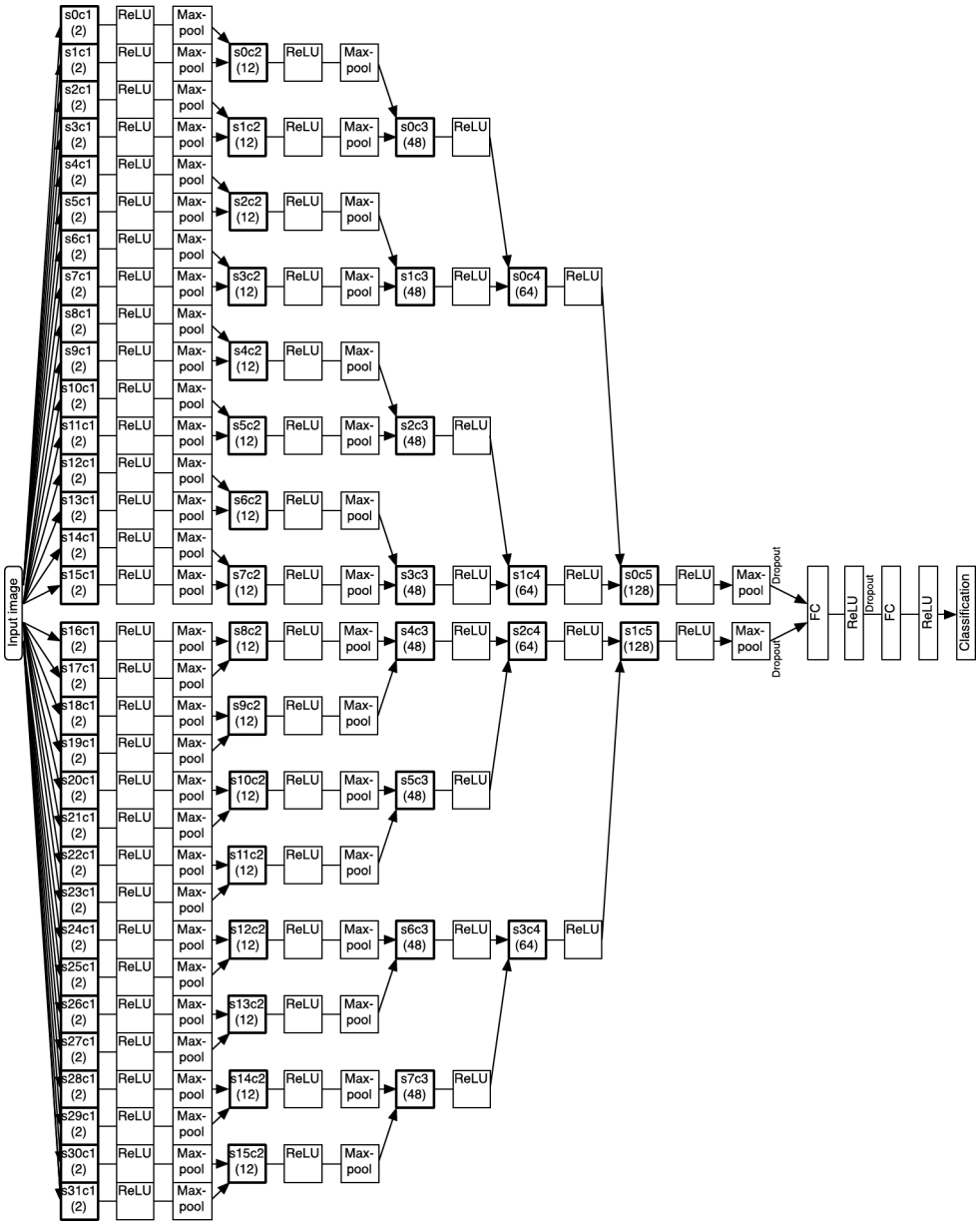


Fig. 1. Architecture of the multistream convolutional neural network with convergence (mcAlexNet). The numbers of filters in the convolutional layers are indicated in parentheses.

Outputs from the two streams of c5 were concatenated and fed into fully connected layers (FCs), and then fed to the output layer for classification. Converging architecture was introduced to define distances between the filters without constraining inputs, and was based on the anatomical organization of the cerebral cortex in which outputs from adjacent neurons tend to converge onto common targets[6,13–15].

mcAlexNet was trained for the classification of 1000 object-image categories using the ImageNet database[16]. The present study was based on 16 model instances of mcAlexNet. Each model instance was trained using randomly initialized parameters. After training, the top-5 accuracy of mcAlexNet was 0.442 to 0.455 using the validation set. This performance was lower than that of both the original AlexNet[12] and my previous study using two-stream fully parallel AlexNet[17]; however, the filters were considered well trained and matured for the present purpose.

Similarities in the properties of filters in the same stream of c1 in mcAlexNet

After training, the c1 filters in mcAlexNet acquired a variety of kernels. Figure 2A shows that many filters in s0–s15 of c1 were orientation selective, whereas those in s16–s31 were color selective, suggesting the segregation of information based on the visual submodality in c1 of mcAlexNet (i.e., orientation information was processed by filters in one group of 16 streams, whereas color information was processed by filters in the other group of 16 streams). This result is consistent with that of my previous study[17].

Another salient feature of Fig. 2A involves the similarities in kernel structures of filters in the same stream of c1 in mcAlexNet. For example, both filters in stream 2 (s2-0 and s2-1; Fig. 2A) were orientation selective. Orientation selectivity was quantified using the orientation index (OI), which was calculated using circular variance[18]. The OI had a value between zero and one, and a larger OI reflected stronger orientation selectivity. The OI of the two filters in s2 was high (s2-0, 0.842; s2-1, 0.853; Fig. 2B, top left) and the absolute difference in OI (ΔOI) between them was small (0.011). Both filters preferred a vertical orientation (preferred orientation (pO): s2-0, 1.24°; s2-1, 1.69°; Fig. 2B, top center) and the absolute difference in pO (ΔpO) between them was small (0.45°). Color selectivity of the two filters was low. Color selectivity was evaluated using the color index (CI), which was calculated using the correlation coefficient (r) of filter weights among the red (R), green (G), and blue (B) channels[17]. The CI had a value between zero and one, and a larger value reflected greater color selectivity. The two filters had low CI (s2-0, 0.0030; s2-1, 0.0028; Fig. 2B, top right) and the absolute difference in CI (ΔCI) between them was small (0.0002). Because the two filters were not color selective, preferred color was not examined.

To further quantify similarities between two filters of c1 in mcAlexNet, a set of responses of filters to 1,000 images was calculated, and a distance measure (dR) between filters in the set of responses was obtained. The dR had a value between zero and one. If the responses to the 1,000 images were similar to each other, dR was zero. The dR was examined using filters that had receptive field at the center. For the two filters (s2-0 and S2-1), Pearson's r was −0.621 and dR was 0.811 (Fig. 2B, bottom left). Thus, their responses to stimulus sets differed significantly despite the similarities in OI, CI and pO of the two filters.

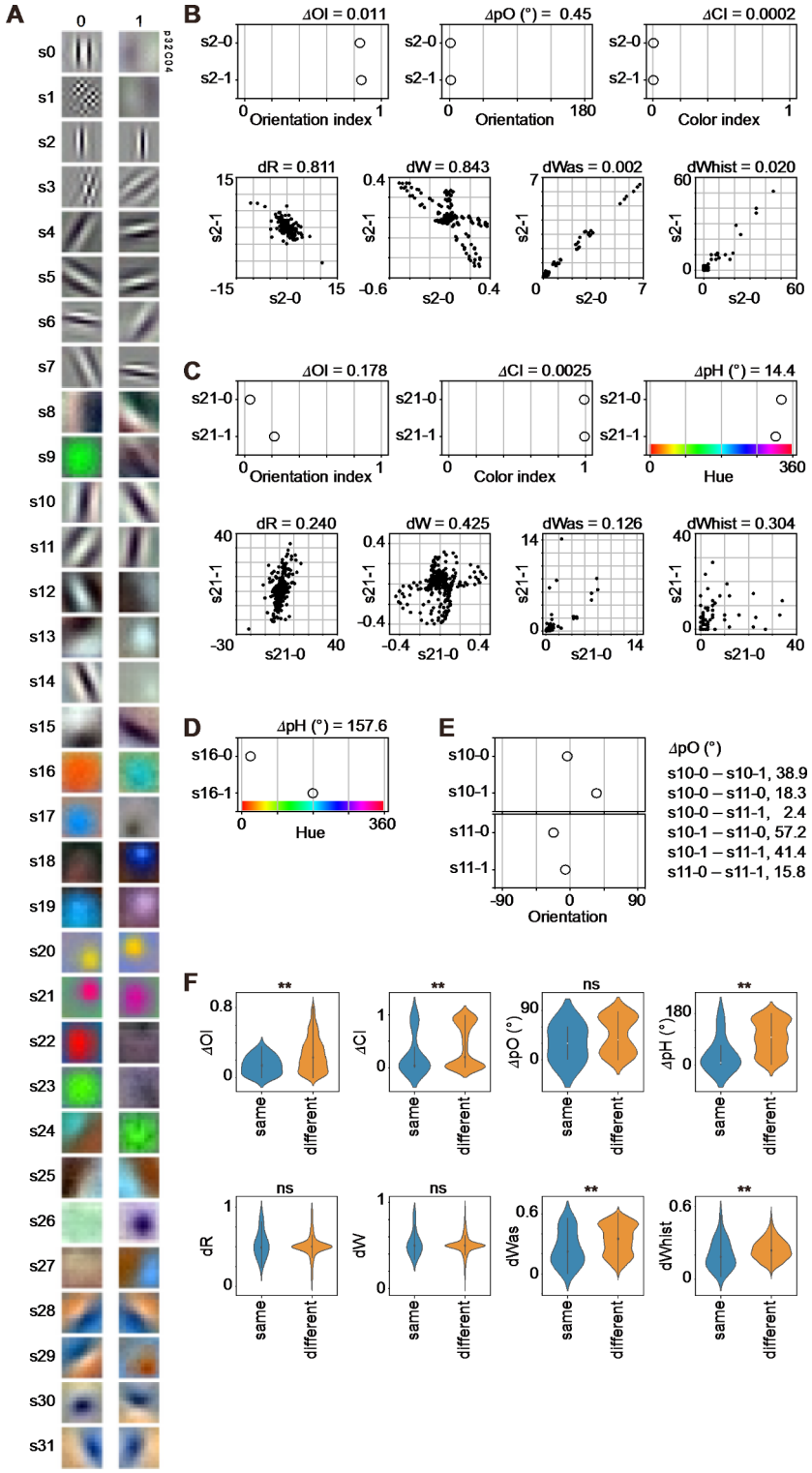
Filter weight structure was quantified using three measures, and the distance was calculated for each measure to quantify similarities between two filters of c1 in mcAlexNet. One measure was based on the pixel-by-pixel filter weights (W). Another measure was based on the amplitude spectrum of a filter weight after two-dimensional discrete Fourier transform (Was). A third measure was based on the pixel histogram of filter weight (Whist). From these measures, the distances (dW, dWas, and dWhist) between filters were calculated. dW quantified the overall similarity in filter structure, whereas dWas quantified the similarity in filter shape, ignoring the spatial phase. Furthermore, dWhist quantified the similarity in the frequency distribution of pixel values, ignoring the filter shape. Distances had a value between zero and one. If two filters were similar to each other, distance was zero. The dW between the two filters (s2-0 and S2-1) was 0.843 (Fig. 2B, bottom center-left), suggesting a low similarity in overall filter structure. By contrast, both dWas and dWhist between the two filters were small (dWas, 0.002, Fig. 2B, bottom center-right; dWhist, 0.020, Fig. 2B, bottom right), suggesting a high similarity in filter shape between the two filters, ignoring the spatial phase. Indeed, the absolute difference in preferred spatial phase (ΔpP) between them—a value between 0° (in-phase) and 180° (anti-phase)—was 153.8° (see Supplementary Table 1 for details). Together, these analyses suggest that the two filters (s2-0 and s2-1) have similar degrees of orientation and color indices and prefer similar orientations but encode different spatial phases.

Note that some of the measures used to characterize filter properties were correlated with each other (Supplementary Fig. 1). However, the degree of correlation differed among model instances. It is therefore important to characterize filters using a variety of measures.

Two filters in another stream (s21-0 and s21-1) of Fig. 2A were not orientation selective (OI: s21-0, 0.038; s21-1, 0.216; ΔOI, 0.178; Fig. 2C, top, left), but were color selective (CI: s21-0, 0.990; s21-1, 0.992; ΔCI, 0.0025; Fig. 2C, top center). Because neither filter was orientation selective, pO was not examined. Color preference was evaluated using the direction in a color wheel (preferred hue [pH]). Both filters preferred magenta (pH: s21-0, 331.7°; s21-1, 317.3°). The absolute difference in pH (ΔpH) between the two filters was 14.4° (Fig. 2C, top right). The filter-response and filter-weight distances (dR, dW, dWas, and dWhist) were relatively small (0.126–0.425; Fig. 2C, bottom), suggesting modest similarities in responses and filter weight structures between the two filters.

In some cases, however, two filters in the same stream of c1 in mcAlexNet showed different structures. For example, one filter (s16-0) of Fig. 2A preferred orange (pH, 21.8°), whereas the other filter (s16-1) of the same stream preferred cyan (pH, 179.4°; Fig. 2D). Another interesting feature in Fig. 2A is that filters in some of the adjacent streams of c1 in mcAlexNet displayed similar structures. For example, all four filters from streams 10 (s10-0 and s10-1) and 11 (s11-0 and s11-1) preferred near-vertical (0°) orientations (pO, −22.0–35.2°; Fig. 2E).

Similarities in filter properties were compared between two filters in the same stream ($n=32$) and two filters from different streams ($n=1{,}984$). The ΔOI of two filters in the same stream was smaller than that from different



streams ($p = 2.83 \times 10^{-4}$; Mann–Whitney U test; Fig. 2F, top left), and the ΔCI of two filters in the same stream was also smaller than that from different streams ($p = 0.0047$; Fig. 2F, top center-left). The ΔpO did not differ between two filters in the same stream and those from different streams ($p = 0.192$; Fig. 2F, top center-right), whereas the ΔpH of two filters in the same stream was smaller than that from different streams ($p = 0.0041$; Fig. 2F, top right). The dR ($p = 0.747$; Fig. 2F, bottom left) and dW ($p = 0.972$; Fig. 2F, bottom center-left) did not differ between two filters in the same stream and those from different streams. However, the dWas ($p = 0.0054$; Fig. 2F, bottom center-right) and dWhist ($p = 0.0088$; Fig. 2F, bottom right) of two filters in the same stream were smaller than those from different streams. These results indicate that two filters in the same stream tend to have more similar properties than two filters from different streams in this model instance.

To further quantify the similarities in filter properties of the same stream, the relationship between the similarities in filter properties and filter-distance of c1 in mcAlexNet was examined. The filter-distance between

◂**Fig. 2**. Comparisons of properties between two filters in the same stream in the first convolutional layer (c1) of a multistream convolutional neural network with convergence (mcAlexNet) of a representative model instance. (**A**) Visualization of weight for c1 filters. The minimum and maximum weight values are scaled for visualization. (**B**, **C**) Comparisons of properties between two filters in s2 (B) and in s21 (**C**). In B, the orientation index (OI, top left), preferred orientation (pO, top center), and color index (CI, top right) are plotted for the two filters (s2-0 and s2-1), and the absolute differences in OI (ΔOI), pO (ΔpO), and CI (ΔCI) between the two filters are provided for each panel. In C, the OI (top left), CI, (top center), and preferred hue (pH, top right) are plotted for the two filters (s21-0 and s21-1), and ΔOI, ΔCI, and the absolute difference in pH (ΔpH) between the two filters are provided for each panel. Both B and C show comparisons between the two filters of responses to a set of 1,000 images (R, bottom left), pixel-by-pixel filter weights (W, bottom center-left), amplitude spectrum of filter weights (Was, bottom center-right), and pixel weight histogram (Whist, bottom right). Distances using the R (dR), W (dW), Was (dWas), and Whist (dWhist) are provided for each panel. (**D**) A comparison of pH between two filters in s16. ΔpH is provided. (**E**) A comparison of pO between four filters in s10 and s11. ΔpOs are provided. (**F**) Comparisons of ΔOI (top left), ΔCI (top center-left), ΔpO (top center-right), and ΔpH (top right) between filters in the same stream and those each from different streams, and comparisons of dR (bottom left), dW (bottom center-left), dWas (bottom center-left), and dWhist (bottom right) between filters in the same stream and those from different streams. Double asterisks indicates significant differences between filters in the same stream and those from different streams with $p < 0.01$. "ns" indicates non-significant differences between filters in the same stream and those from different streams.

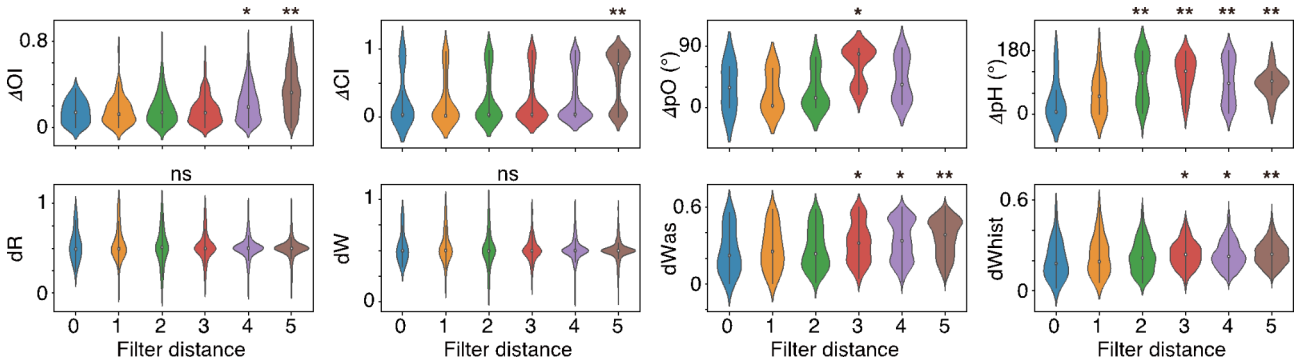


Fig. 3. Comparisons of similarities in filter properties in the first convolutional layer (c1) of a multistream convolutional neural network with convergence (mcAlexNet) across filter-distance groups for the representative model instance in Fig. 2. Comparisons of absolute differences in the orientation index (ΔOI, top left), color index (ΔCI, top center-left), preferred orientation (ΔpO, top center-right), and preferred hue (ΔpH, top right) across filter-distance groups are shown. In addition, comparisons of the distances of set of responses (dR, bottom left), pixel-by-pixel filter weights (dW, bottom center-left), amplitude spectrum of filter weights (dWas, bottom center-right), and pixel weight histograms (dWhist, bottom right) across the filter-distance groups are shown. Double and single asterisks indicate significant differences between the 0th and corresponding filter-distance group with $p < 0.01$ and $p < 0.05$, respectively. "ns" indicates non-significant differences across filter-distance groups.

filters in the same stream was defined as zero. The filter-distance between filters from different streams was defined using the following rule, which was based on the anatomical organization of the cerebral cortex[6,13–15]. The filter-distance between two filters from different streams was 1 if their outputs were integrated for the first time by filters in c2. For example, the filter-distance between a filter from s0 and a filter from s1 was 1. The filter distances between two filters from different streams were 2, 3, 4, or 5 if their outputs were integrated for the first time by filters in c3, c4, c5, or FC, respectively. Consequently, the filter-distance between a filter from s0–s1 and a filter from s2–s3 was 2, that between a filter from s0–s3 and a filter from s4–s7 was 3, that between a filter from s0–s7 and a filter from s8–s15 was 4, and that between a filter from s0–s15 and a filter from s16–s31 was 5 (Supplementary Fig. 2).

Similarities in OI and CI differed among filter-distance groups in the model instance shown in Fig. 2. The ΔOI differed among filter-distance groups ($p = 1.84 \times 10^{-63}$; Kruskal–Wallis test; Fig. 3, top left), and that of the 0th filter-distance group was smaller than those of the 4th ($p = 0.0157$; Mann–Whitney U test) and 5th ($p = 3.05 \times 10^{-8}$) filter-distance groups. The maximum range of filter-distance in which ΔOI did not differ from that of the 0th filter-distance group was defined as the size of the orientation-index module, which was 3 for this model instance. The ΔCI also differed among filter-distance groups ($p = 5.40 \times 10^{-73}$; Fig. 3, top center-left), and the ΔCI of the 0th filter-distance group was smaller than that of the 5th filter-distance group ($p = 1.23 \times 10^{-6}$). The size of the color-index module, which was defined similarly to that of the orientation-index module, was 4.

Similarities in pO and pH also differed among filter-distance groups. The ΔpO differed among filter-distance groups ($p = 8.13 \times 10^{-6}$; Kruskal–Wallis test; Fig. 3, top center-right), and that of the 0th filter-distance group was smaller than that of the 3rd filter-distance group ($p = 0.0330$; Mann–Whitney U test). The number of filter pairs

in the 5th filter-distance group of ΔpO was zero because neither filter in pairs of the 5th filter-distance group showed orientation selectivity. Note that the comparison of ΔpO between two filters in the same stream and those from the different streams did not reach the significant difference (see Fig. 2F top center-right), suggesting that "different" group included heterogenous filter pairs and analysis using filter-distance group revealed the difference in ΔpO. The ΔpH also differed among filter-distance groups ($p = 0.00196$; Fig. 3, top right), and that of the 0th filter-distance group was smaller than those of the 2nd–5th filter-distance groups ($p = 0.00211$–0.00932). The size of the orientation- and color-tuning module were 2 and 1, respectively.

In contrast to the aforementioned results, the dR ($p = 0.896$; Kruskal–Wallis test; Fig. 3, bottom left) and dW ($p = 0.933$; Fig. 3, bottom center-left) did not differ among filter-distance groups. However, the dWas ($p = 1.28 \times 10^{-11}$; Fig. 3, bottom center-right) and dWhist ($p = 0.0010$; Fig. 3, bottom right) differed among filter-distance groups. The dWas and dWhist of the 0th filter-distance group were smaller than those of the 3rd–5th filter-distance groups ($p = 0.0276$–8.10×10^{-4}). These results suggest that filters with smaller filter distances tend to have more similar shapes in this model instance.

Because analysis of relationship between the similarities in filter properties and filter-distance in the network captured similarity in properties of two filters in the same stream well, population analysis was performed using this measure. Sixteen model instances, each trained with random initial values, were examined. Significant difference in ΔOI among filter-distance groups ($p < 0.05$, Kruskal–Wallis test) was consistently observed in all 16 instances. The median ΔOI across filter pairs was calculated for each filter-distance group of an individual model instance. The median ΔOI differed among filter-distance groups ($p = 2.91 \times 10^{-12}$, $\chi^2 = 63$, $n = 16$, Friedman test, Fig. 4, top left). The median ΔOI of the 0th filter-distance group was smaller than that of the 2nd–5th filter-distance groups ($p = 3.05 \times 10^{-5}$–6.10×10^{-5}, $n = 16$, Wilcoxon signed-rank test) and increased gradually with filter-distance. Significant difference in ΔCI among filter-distance groups was also consistently observed in all 16 instances. The median ΔCI differed among filter-distance groups ($p = 4.97 \times 10^{-11}$, $\chi^2 = 57.0$, Fig. 4, top center-left). The median ΔCI of the 0th filter-distance group was smaller than those of the 2nd–5th filter-distance groups ($p = 3.05 \times 10^{-5}$–4.27×10^{-4}), with a rather abrupt increase in the 5th filter-distance group. These results indicate that c1 filters in the same stream of mcAlexNet tend to have more similar degrees of selectivity than those from different streams with larger filter distances.

Significant differences in ΔpO and ΔpH among filter-distance groups were observed in 11 and eight instances, respectively. Both the median ΔpO ($p = 3.88 \times 10^{-8}$; $\chi^2 = 42.9$, Fig. 4, top center-right) and ΔpH ($p = 0.0027$; Fig. 4, top right) differed among filter-distance groups. The median ΔpO of the 0th filter-distance group was smaller than those of the 3rd–5th filter-distance groups ($p = 3.05 \times 10^{-5}$–2.44×10^{-4}), and the median ΔpH of the 0th filter-distance group was smaller than those of the 2nd–5th filter-distance groups ($p = 0.0017$–0.034). These results indicate that c1 filters in the same stream of mcAlexNet tend to have more similar preferences to orientation and color than those from different streams with larger filter distances.

Significant differences in both dR and dW among filter-distance groups were observed in only one instance. The median dR did not differ among filter-distance groups ($p = 0.113$, $\chi^2 = 8.89$; Fig. 4, bottom left); however, the median dW differed among filter-distance groups ($p = 2.21 \times 10^{-4}$, $\chi^2 = 24.0$; Fig. 4, bottom center-left). The dW of the 0th filter-distance group was smaller than those of the 1st–5th filter-distance groups ($p = 0.0052$–0.018). By contrast, significant differences in dWas and dWhist among filter-distance groups were consistently

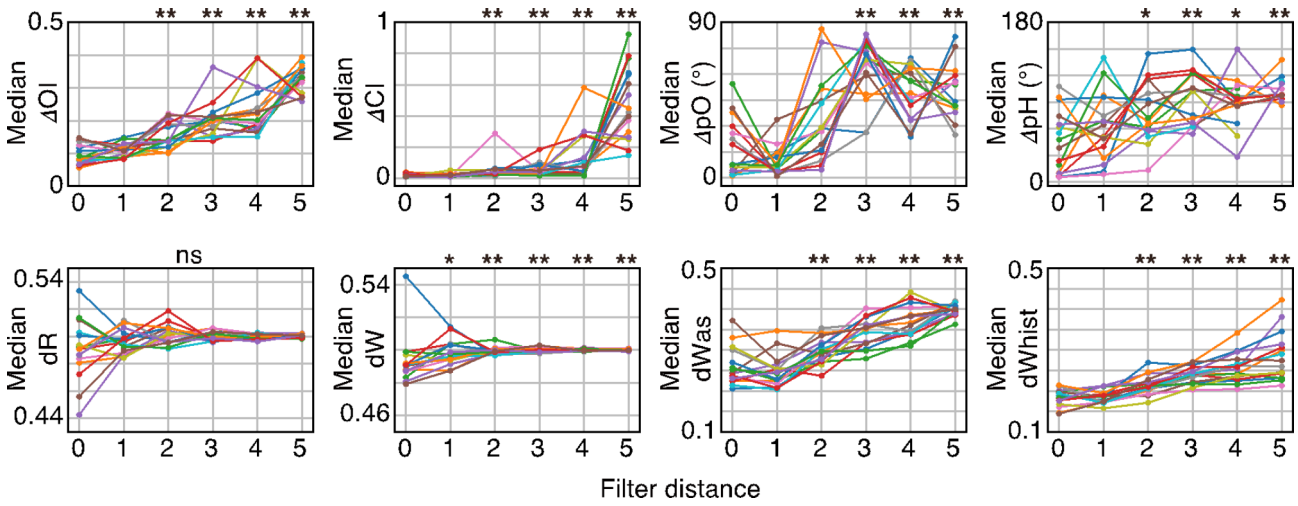


Fig. 4. Comparisons of similarities in filter properties in the first convolutional layer (c1) of a multistream convolutional neural network with convergence (mcAlexNet) across filter-distance groups. Comparisons of median absolute differences in the orientation index (ΔOI, top left), color index (ΔCI, top center-left), preferred orientation (ΔpO, top center-right), and preferred hue (ΔpH, top right) across filter-distance groups are shown. In addition, comparisons of the median distance of set of responses (dR, bottom left), pixel-by-pixel filter weights (dW, bottom center-left), amplitude spectrum of filter weights (dWas, bottom center-right), and pixel weight histograms (dWhist, bottom right) across the filter-distance groups are shown. Each line represents a model instance; the results from 16 model instances are plotted. Other conventions are as in Fig. 3.

observed in all 16 instances. The median dWas ($p = 4.61 \times 10^{-13}$, $\chi^2 = 66.9$; Fig. 4, bottom center-right) and dWhist ($p = 9.09 \times 10^{-14}$, $\chi^2 = 70.3$; Fig. 4, bottom right) differed among filter-distance groups, and those of the 0th filter-distance group were smaller than those of the 2nd–5th filter-distance groups ($p = 3.05 \times 10^{-5}$–0.0021). Both median dWas and dWhist increased gradually with filter-distance. These results indicate that c1 filters in the same stream of mcAlexNet tend to have more similar shapes than those from different streams with larger filter distances.

The size of the index module, which was the combination of the orientation- and color-index modules, was compared with that of the tuning module, which was the combination of the orientation- and color-tuning modules. Because the size of the orientation-index module (2.31 ± 1.01, mean ± standard deviation [SD], $n = 16$) did not differ from that of the color-index module (2.81 ± 1.17, $n = 16$; $p = 0.210$, Mann–Whitney U test), the two were combined as the index module. Similarly, because the size of the orientation-tuning module (1.89 ± 0.93, $n = 9$) did not differ from that of the color-tuning module (1.6 ± 0.89, $n = 5$; $p = 0.480$), the two were combined as the tuning module. The size of the index module (2.56 ± 1.09) was larger than that of the tuning module (1.79 ± 0.89, $p = 0.041$; Fig. 5).

Similarities in properties of filters in the same stream of c1 in a fully parallel multistream AlexNet

In the first analysis, I examined mcAlexNet, which has multistream architecture with convergence. Converging connectivity was introduced to define the distances between filters in the network, to analyze relationships between similarities in filter properties and filter-distance. Next, to clarify whether converging connectivity contributed to the appearance of similar properties in two filters of the same stream, I constructed another multistream CNN, this time with multistream architecture without convergence (mAlexNet; see Supplementary Fig. 3). In mAlexNet, 32 streams were organized in a fully parallel manner from the first convolutional layer (c1) to the last convolutional layer (c5) without convergence. Outputs from the 32 streams of c5 were concatenated

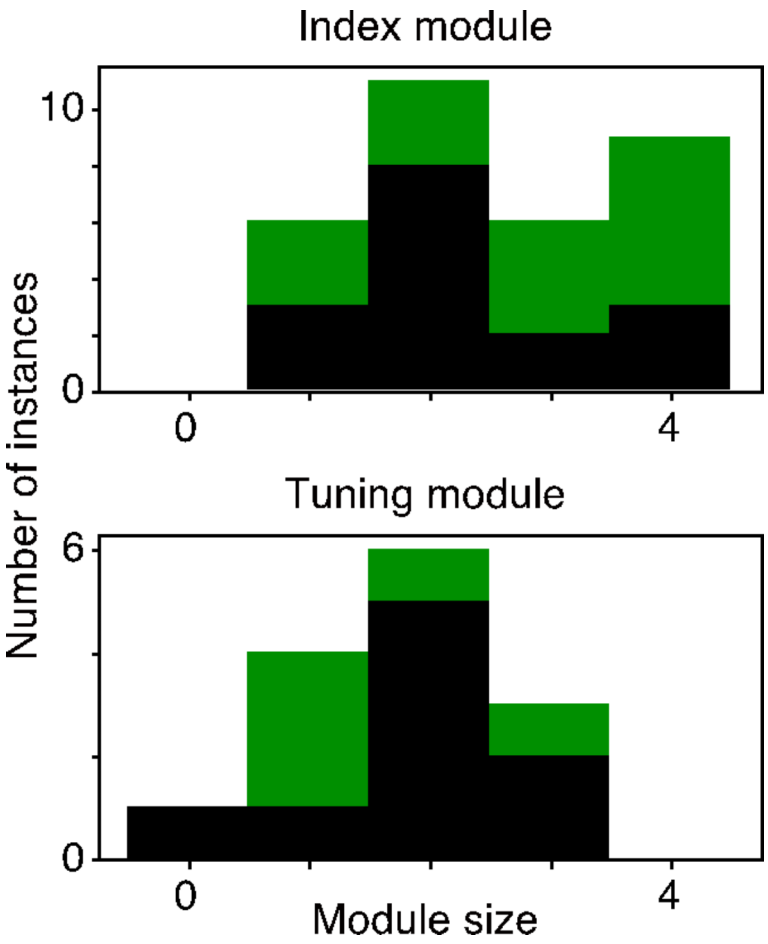


Fig. 5. Comparisons of module sizes between the index and tuning modules in the first convolutional layer (c1) of a multistream convolutional neural network with convergence (mcAlexNet). Top, Frequency distribution of the size of the index module (black columns, orientation-index module; green columns, color-index module). Bottom, Frequency distribution of the size of the tuning module (black columns, orientation-tuning module; green columns, color-tuning module). The analysis was limited to model instances with significant differences in similarity measures among filter-distance groups ($p < 0.05$, Kruskal–Wallis test).

and fed into FCs, and then fed to the output layer for classification. In the present study, eight model instances of mAlexNet, each trained with random initial values, were examined. After training, the top-5 accuracy of mAlexNet was 0.348 to 0.362 using the validation set.

Similar to the results obtained with mcAlexNet, c1 filters in the same stream of mAlexNet displayed kernel structures that were similar to each other (Supplementary Fig. 3). Significant differences between two filters in the same stream and two filters from different streams ($p < 0.05$, Kruskal–Wallis test) in ΔOI and ΔCI were observed in four and eight instances, respectively. Moreover, significant differences in ΔpO and ΔpH were observed in four and three instances, respectively. The median ΔOI, ΔCI, ΔpO, and ΔpH of two filters in the same stream were smaller than those from different streams ($p = 0.0078$ for the four measures, Wilcoxon signed-rank test; Fig. 6, top). In mAlexNet, significant differences in dR and dW were observed in eight and six instances, respectively, and significant differences in dWas and dWhist were observed in three and eight instances, respectively. The median dR, dW, dWas, and dWhist of two filters in the same stream were smaller than those from different streams ($p = 0.0078$–0.015; Fig. 6, bottom). These results suggest that converging connectivity is not important for the appearance of similar properties in filters in the same stream.

In mcAlexNet, significant differences in dR and dW among the filter-distance groups were observed in only one instance, suggesting that the presence/absence of convergence in the networks results in the differences in similarities of dR and dW among filter-distance groups. Another difference between mcAlexNet and mAlexNet was that many c1 filters of mAlexNet failed to develop filter structure and were flat (Supplementary Fig. 3), indicating that convergence is important for the development of appropriate filter structures in multistream CNNs. Interestingly, if a filter in a stream was flat, the other filter in the same stream was also flat.

Relationships between classification accuracies and similarities in filter properties

Because similar properties of filters in the same stream emerged spontaneously in c1 of mcAlexNet, such organization likely plays an important role in the classification of input images. If this is the case, the degree of similarity in properties of filters in the same stream is likely to be related to classification accuracy in mcAlexNet. To examine this possibility, correlations between top-5 accuracy and median ΔOI, ΔCI, ΔpO, or ΔpH of filter pairs in the same stream of c1 in mcAlexNet were calculated using all 16 model instances. The median ΔOI was negatively related to top-5 accuracy ($r = -0.506$, Fig. 7, top left), indicating that if filter pairs in the same stream in c1 of a model instance displayed higher similarity in OI (i.e., a smaller median ΔOI) than those of other model instances, then the classification accuracy of the model instance was higher than that of the other model instances.

There was a negative correlation between the median ΔpO and top-5 accuracy ($r = -0.418$, Fig. 7, top center-right) but this correlation was not significant. The median ΔCI and ΔpH were not related to top-5 accuracy

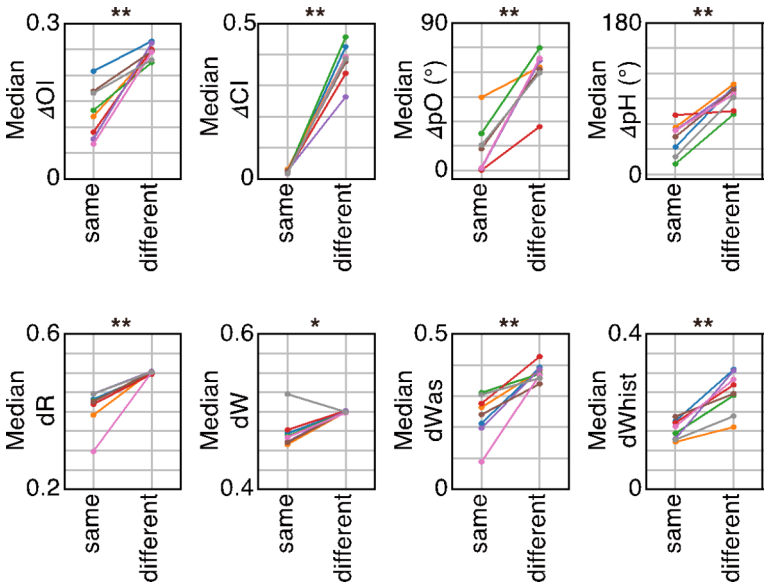


Fig. 6. Comparisons of similarities in filter properties in the first convolutional layer (c1) of another type of multistream convolutional neural network (mAlexNet) between two filters in the same stream and two filters from different streams. Comparisons of the median absolute differences in orientation index (ΔOI, top left), color index (ΔCI, top center-left), preferred orientation (ΔpO, top center-right), and preferred hue (ΔpH, top right) between two filters in the same stream and those from different streams are shown. Moreover, comparisons of the median distances of set of responses (dR, bottom left), pixel-by-pixel filter weights (dW, bottom center-left), amplitude spectrum of filter weights (dWas, bottom center-right), and pixel weight histograms (dWhist, bottom right) between two filters in the same stream and those from different streams are shown. Each line represents a model instance; the results from eight model instances were plotted. Other conventions are as in Fig. 3.

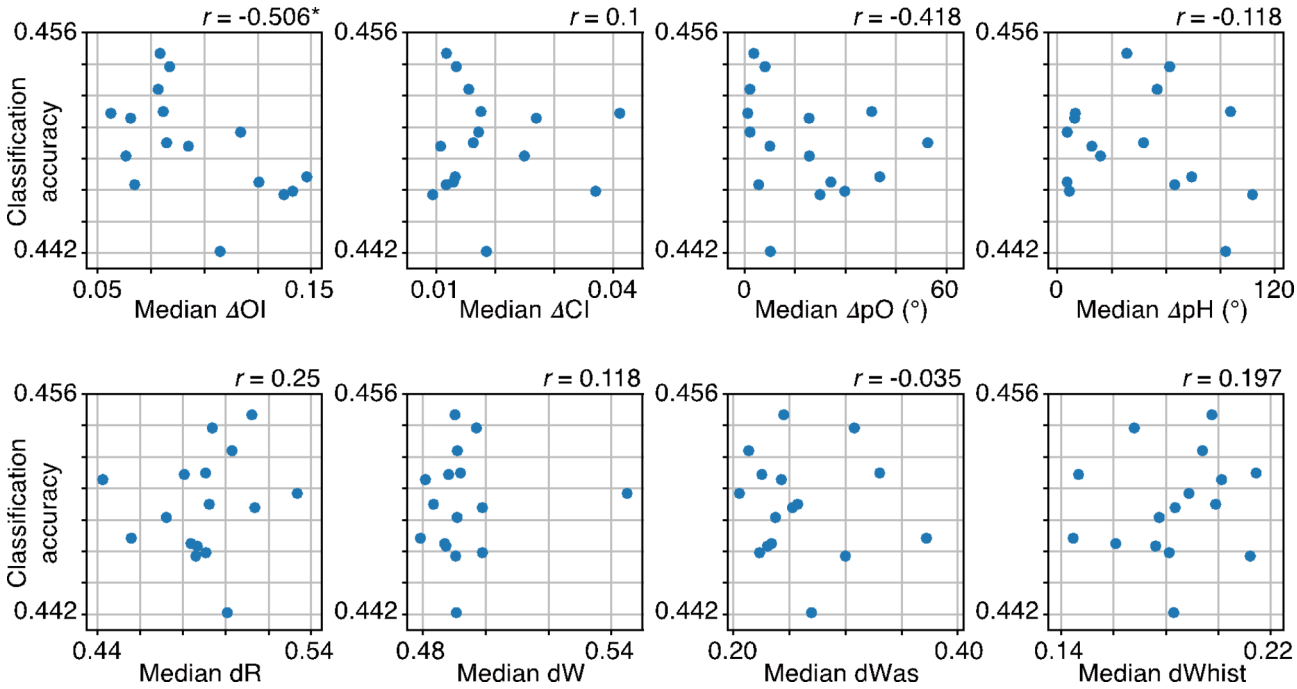


Fig. 7. Relationships between classification accuracy (top-5 accuracy) and the median absolute differences or median distances in properties between two filters in the same stream of the first convolutional layer (c1) of a multistream convolutional neural network with convergence (mcAlexNet). Relationships between top-5 accuracy and the medians of absolute differences in orientation index (ΔOI, top left), color index (ΔCI, top center-left), preferred orientation (ΔpO, top center-right), and preferred hue (ΔpH, top right) are shown. Moreover, relationships between top-5 accuracy and the median distances of set of responses (dR, bottom left), pixel-by-pixel filter weights (dW, bottom center-left), amplitude spectrum of filter weights (dWas, bottom center-right), and pixel weight histograms (dWhist, bottom right) are shown. Each point represents a model instance; the results from 16 model instances are plotted. A correlation coefficient (r) is provided for each panel. A single asterisk indicates significant correlations between two measures with $p < 0.05$.

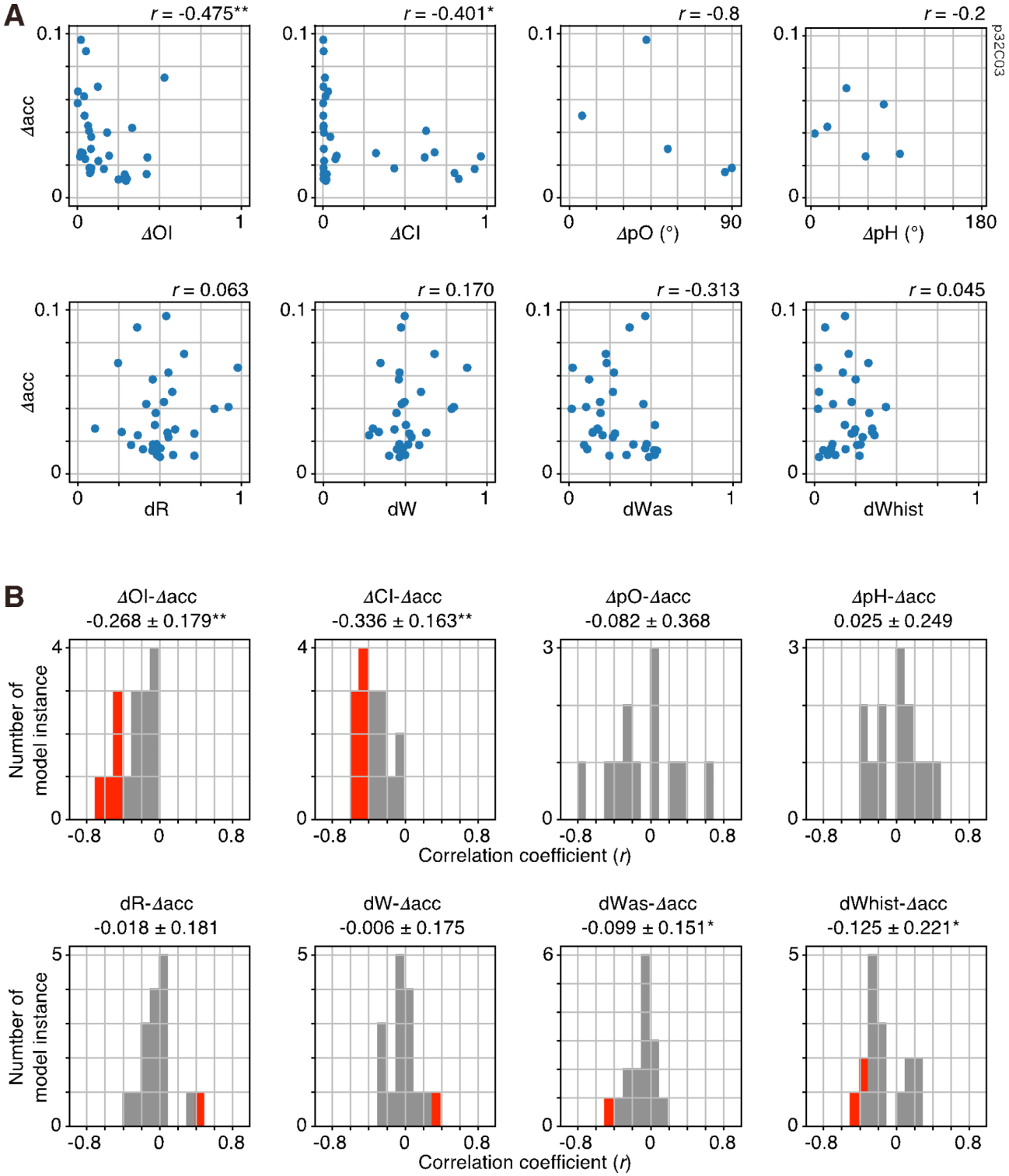
($r = 0.1$ and -0.118, respectively, Fig. 7, top center-left and top right). There were also no correlations between top-5 accuracy and median dR, dW, dWas, or dWhist ($r = -0.035$–0.25, Fig. 7, bottom). These results suggest that two filters in the same stream having similar orientation selectivity may play an important role in the classification of input images.

Because the median ΔOI was negatively related to classification accuracy in the aforementioned analysis, I examined the relationship between similarities in filter properties and classification accuracies by deleting a single stream of c1 in mcAlexNet. If similarities in properties of two filters in the same stream play an important role in the classification of input images, the effect on classification accuracy of deleting a stream with two filters with similar properties is likely to be more marked than that of deleting a stream with filters with dissimilar properties. To delete a stream of c1, output values of the max-pool layer of a stream after c1 were set to zero. The classification accuracy (top-5 accuracy) was calculated for the original mcAlexNet and the single-stream-deleted mcAlexNet. Changes in top-5 accuracy (Δacc) were evaluated by calculating the difference between the top-5 accuracy calculated using original mcAlexNet and that calculated using single-stream-deleted mcAlexNet.

In a model instance, by deleting a single stream of c1 in mcAlexNet, top-5 accuracy decreased from 0.449 to 0.353–0.439, and Δacc was 0.010–0.096 ($n = 32$; Fig. 8A, vertical axis), indicating that the deletion of a single stream decreases classification performance, and that the degree of deletion differs among deleted streams.

A relatively large decrease in classification accuracy was observed if a stream with two filters with small ΔOI was deleted (Fig. 8A, top left). The ΔOI correlated negatively with Δacc in this model instance ($r = -0.475$, Spearman's correlation, $n = 32$; Fig. 8A, top left). The ΔCI also correlated negatively with Δacc ($r = -0.401$, $n = 32$; Fig. 8A, top center-left). There was a strong negative correlation between ΔpO and Δacc ($r = -0.8$, $n = 5$; Fig. 8A, top center-right), but this correlation was not significant, likely because of the small sample size. There was a weak relationship between ΔpH and Δacc ($r = -0.2$, $n = 6$), which was also not significant (Fig. 8A, top right). Note that analyses were limited to streams if both filters were selective to orientation (OI ≥ 0.5; for analyses using ΔpO) or color (CI ≥ 0.5; for analyses using ΔpH). There were no or weak relationships between dR and Δacc as well as between similarities in filter weights (dW, dWas, or dWhist) and Δacc; these relationships were not significant (Fig. 8A, bottom).

A similar tendency was observed in the 16 model instances; that is, similarities in selectivity indexes (OI and CI) were negatively correlated with Δacc. The correlation between ΔOI and Δacc was significant ($p < 0.05$) in five instances. The average r was -0.268 ± 0.179 (mean ± SD, $n = 16$; Fig. 8B, top left), and the frequency distribution



of r was shifted from zero to negative ($p = 3.05 \times 10^{-5}$; Wilcoxon signed-rank test). Correlation between ΔCI and Δacc was significant in seven instances. Average r was -0.336 ± 0.163 ($n = 16$; Fig. 8B, top center-left), and the frequency distribution was shifted from zero to negative ($p = 3.05 \times 10^{-5}$). These results suggest that a stream having two filters with similar degrees of orientation selectivity or color selectivity plays an important role in the classification of input images.

Although there was a strong negative correlation between ΔpO and Δacc in one model instance, the correlation coefficient was widely distributed between −0.8 and 0.6, and the frequency distribution of r between ΔpO and Δacc ($r = 0.082 \pm 0.368$, $n = 12$) was not shifted from zero ($p = 0.508$; Fig. 8B, top center-right). The frequency distribution of r between ΔpH and Δacc ($r = 0.025 \pm 0.249$, $n = 13$) was also not shifted from zero ($p = 0.755$ Fig. 8B, top right). No significant correlations were observed between ΔpO and Δacc or ΔpH and Δacc. This analysis was limited to model instances in which five or more streams had two filters with orientation

◀**Fig. 8**. Effects of deleting a stream of the first convolutional layer (c1) in a multistream convolutional neural network with convergence (mcAlexNet) on classification accuracy (top-5 accuracy). (**A**) Relationships between changes in classification accuracy (Δacc) and absolute differences in orientation index (ΔOI, top left), color index (ΔCI, top center-left), preferred orientation (ΔpO, top center-right), or preferred hue (ΔpH, top right) between two filters in a deleted stream are shown. Relationships between Δacc and the distances of set of responses (dR, bottom left), pixel-by-pixel filter weights (dW, bottom center-left), amplitude spectrum of filter weights (dWas, bottom center-right), or pixel weight histograms (dWhist, bottom right) between two filters in a deleted stream are shown. Each point represents a deleted stream; the results from 32 streams are plotted. For ΔpO and Δacc, the results from five streams in which both filters were orientation selective (orientation index ≥ 0.5) are plotted. For ΔpH and Δacc, the results from six streams in which both filters were color selective (color index ≥ 0.5) are plotted. The correlation coefficient (r) is provided for each panel. Double and single asterisks indicate significant correlations between two measures, with $p < 0.01$ and $p < 0.05$, respectively. (**B**) Frequency distributions of r between Δacc and filter similarity measures of deleted streams. Frequency distributions of r between Δacc and ΔOI (top left,), ΔCI (top center-left), ΔpO (top center-right), ΔpH (top right), dR (bottom left), dW (bottom center-left), dWas (bottom center-right), or dWhist (bottom right) of filters in deleted streams are shown. Red columns indicate significant correlations ($p < 0.05$). The mean ± standard deviation of r across model instances is provided for each panel. The results from 16 model instances are plotted. For ΔpO and Δacc, results from 12 instances in which five or more streams had two filters with orientation selectivity (orientation index ≥ 0.5) are plotted. For ΔpH and Δacc, results from 13 instances in which five or more streams had two filters with color selectivity (color index ≥ 0.5) are plotted. Double and single asterisks indicate significant shifts from 0 with $p < 0.01$ and $p < 0.05$, respectively.

or color selectivity. The results suggest that the contributions of two filters in the same stream that prefer similar orientations or colors in the classification of input images did not differ from those that prefer different orientations or colors.

A significant but positive correlation was observed between dR and Δacc in one instance, and the frequency distribution of r was not shifted from zero ($r = -0.018 \pm 0.181$, $n = 16$; $p = 0.404$; Fig. 8B, bottom left). Similarly, a significant but positive correlation was observed between dW and Δacc in one instance, and the frequency distribution of r was not shifted from zero ($r = -0.006 \pm 0.175$, $n = 16$; $p = 0.669$; Fig. 8B, bottom center-left). For dWas and dWhist, a weak tendency toward negative correlation was observed. A significant negative correlation between dWas and Δacc was observed in one instance, and the frequency distribution of r was shifted from zero ($r = -0.099 \pm 0.151$, $n = 16$; $p = 0.029$; Fig. 8B, bottom center-right). Similarly, a significant negative correlation between dWhist and Δacc was observed in two instances, and the distribution of r was also shifted from zero ($r = -0.125 \pm 0.221$, $n = 16$; $p = 0.044$; Fig. 8B, bottom right). These results suggest that two filters in the same stream with similar filter shapes but without close similarities in spatial phase may contribute weakly to the classification of input images.

The number of filters and the acquisition of similar filter properties in the same stream of c1 in mcAlexNet

To examine whether the acquisition of similar filter properties in the same stream of c1 in mcAlexNet is dependent on the number of filters, a network with a smaller number of filters than mcAlexNet was constructed (smcAlexNet). smcAlexNet consisted of five hierarchically organized convolutional layers (c1–c5) and three pooling layers (Max-pool). c1 contained eight parallel streams (s0–s7), each equipped with two filters. c2 contained four parallel streams (s0–s3), each equipped with 12 filters. c3 contained two parallel streams (s0 and s1), each equipped with 48 filters. c4 contained two parallel streams (s0 and s1), each equipped with 32 filters. Finally, c5 contained two parallel streams (s0 and s1), each equipped with 32 filters. The other architectures of smcAlexNet were similar to those of mcAlexNet. Thus, the number of filters in convolutional layers of smcAlexNet was reduced to 25% of the number of filters of mcAlexNet. In the present study, eight model instances of smcAlexNet, each trained with random initial values, were examined. After training, the top-5 accuracy of smcAlexNet was 0.353 to 0.366 using the validation set. The analysis was limited to the median ΔOI, ΔCI, ΔpO, and ΔpH, which captured similarities in filter properties in the same stream of c1 of mcAlexNet.

The properties of two filters in the same stream of c1 in smcAlexNet were less similar to each other than those of mcAlexNet. The median ΔCI ($p = 3.70 \times 10^{-4}$, Mann–Whitney U test; Fig. 9), ΔpO ($p = 0.022$), and ΔpH ($p = 0.0071$) of smcAlexNet were larger than those of mcAlexNet, whereas the median ΔOI ($p = 0.697$) did not differ between mcAlexNet and smcAlexNet. These results suggest that the acquisition of similar filter properties in the same stream of c1 in mcAlexNet is dependent on the number of filters, and indicate that properties of filters in the same stream of c1 become less similar to each other if the network contains a smaller number of filters.

Discussion

An important finding of the present study was that c1 filters in the same stream of mcAlexNet, which had 32 streams in c1, acquired preferences for the same visual submodality after training. This result is consistent with my previous study using two-stream fully parallel AlexNet[17], in which I revealed that color information and shape information are segregated into two different streams.

Another important and new finding of the present study was that two C1 filters in the same stream preferred similar orientations and/or colors after training. For example, if one filter preferred a vertical orientation, the other filter in the same stream also preferred a vertical orientation.

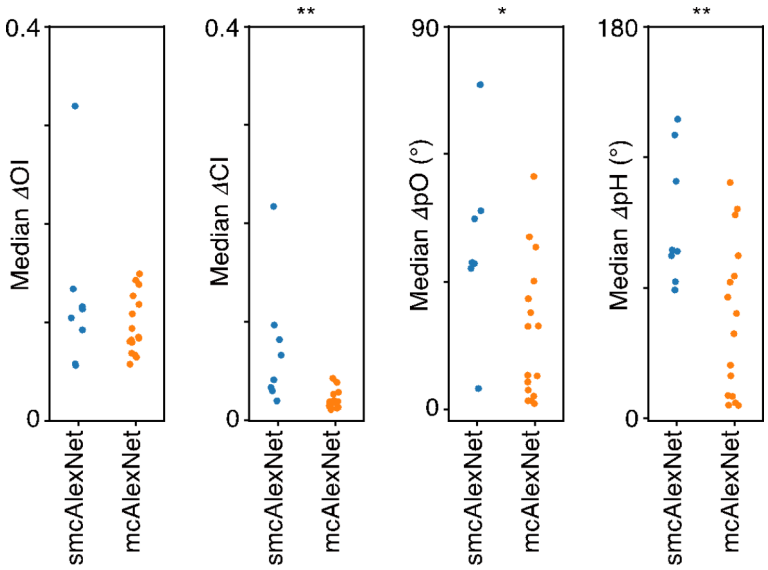


Fig. 9. Comparisons of similarities in filter properties in the same stream of the first convolutional layer (c1) between a multistream convolutional neural network with convergence (mcAlexNet, orange) and a small mcAlexNet (smcAlexNet, blue). Comparisons of the median absolute differences in orientation index (ΔOI, $n = 16$ for mcAlexNet, $n = 8$ for smcAlexNet; left), color index (ΔCI, $n = 16$ for mcAlexNet, $n = 8$ for smcAlexNet; center-left), preferred orientation (ΔpO, $n = 16$ for mcAlexNet, $n = 7$ for smcAlexNet; center-right), and preferred hue (ΔpH, $n = 16$ for mcAlexNet, $n = 8$ for smcAlexNet; right). Note that analyses were limited to streams if both filters were selective to orientation (orientation index ≥ 0.5) for analysis using ΔpO, or to color (color index ≥ 0.5) for analysis using ΔpH. As a result, in one model instance of smcAlexNet, no filter pairs were selective to orientation; this instance was excluded from the analysis of ΔpO. Other conventions are as in Fig. 3.

In the present study, by performing the experimental deletion of two filters in the same stream, the functional importance of the clustering of filters dealing with the same visual submodality was clarified. By contrast, in my previous study[17], the functional importance of the clustering of filters dealing with the same visual submodality was not investigated. Furthermore, relationships between performance and similarity in stimulus preference (pO, or pH) of two filters in the same stream were also investigated in the present study.

In the present study, two filters in the same stream of c1 of mcAlexNet acquired similar properties. In mcAlexNet, filter weights were initially set to random values and there were no constraints on inputs to two filters in the same stream of c1. In principle, two filters in the same stream were therefore able to have any filter structures. However, two filters in the same stream acquired similar properties, such as OI, CI, pO, pH. The spontaneous appearance of filters with similar properties in mcAlexNet is consistent with a self-organization model of functional modular organization in the brain[19].

The acquisition of similar properties in c1 filters of the same stream in mcAlexNet is reasonable because modular design has several computational advantages[8]. The presence of filters with similar OI and CI in the same stream allows for the analysis of input images in a certain visual submodality, and improves resolution in the submodality by circumventing combinatorial explosion. The single-stream-deletion experiment further confirmed that filters with similar OI and CI in the same stream play an important role in the classification of input images.

I must note that filters with similar degrees of stimulus selectivity in the same stream are required for the optimal task performance of mcAlexNet. The result cannot be generalized to other types of neural networks because the performance of the multi-stream architecture was inferior to that of conventional single-stream architecture.

Filters in the same stream of c1 also acquired similar pO or pH, which suggests that the presence of these filters in the same stream is beneficial for information processing in mcAlexNet. However, the single-stream-deletion experiment failed to reveal the importance of filters with similar pO and pH in the same stream for classification of input images, suggesting that filters with similar stimulus preferences in the same stream are as important as those with different preferences. In future studies using mcAlexNet with more parallel streams, which may increase the sample size, it may be possible to reveal the relationship between the similarity in stimulus preference and Δacc.

Although the present study revealed the acquisition of similar properties by filters in the same stream of c1 in mcAlexNet after training, the mechanisms of acquisition were unclear. The outputs of c1 filters from the 0th and 1st filter-distance groups were integrated for the first time at c2 filters, and the filter properties of the 0th and 1st filter-distance group were similar to each other. These findings suggest the importance of sharing output targets in the acquisition of similar filter properties in mcAlexNet.

Similar properties in c1 filters that emerged after training were not caused by the sharing of inputs. An analysis of the similarities in filter properties using R, W, Was and Whist revealed that filters in the same stream of c1 displayed similar, but not completely identical, properties; this suggests that two filters in the same stream of c1 may receive independent or unique inputs from each other, and create similar filter properties *de novo*.

The c1 filters in the same stream of mcAlexNet acquired preferences to the same visual submodality, and the deletion experiment revealed the importance of this organization for image classification. The results indicate the functional importance of orientation module and color module in the primary visual cortex. By contrast, although c1 filters in the same stream of mcAlexNet acquired preferences to similar orientations and colors, the deletion experiment failed to reveal the functional importance of this clustering. Together, these results suggest that the clustering of neurons with similar stimulus preferences may not play an important role in the primary visual cortex. Nonetheless, it has been suggested that the stochasticity inherent in neural activity can be reduced by integrating outputs from neurons with similar stimulus preferences[20,21]. The brain may thus make effective use of such functional organization.

Some properties observed in c1 filters of mcAlexNet have been reported in the brain. Comparisons of dW and dWas between two filters in the same stream and two filters from different streams suggest that two filters in the same stream of c1 display similarities in shape but differ in spatial phase. Indeed, the ΔpP of two filters in the same stream was not different from that of two filters from different streams in 15 of the 16 model instances of mcAlexNet. Even in the one model instance with a significant difference, the ΔpP of two filters in the same stream was larger than that of two filters from different streams (Supplementary Table 1). The analysis suggested that two filters in the same stream preferred a variety of spatial phases. Consistent with these findings, adjacent neurons in the cat visual cortex prefer similar orientation, but prefer a variety of phases[22]. Furthermore, the results from the present study revealed that dR of two filters in the same stream did not differ from that of two filters from different streams. A low similarity in responses to a set of stimulus images between adjacent neurons has been observed in primate visual cortical areas[23].

smcAlexNet, which had a smaller number of filters than mcAlexNet, displayed less similar properties of filters in c1 compared with mcAlexNet. This result might be relevant to the differences in visual cortex functional organization between rodents and primates. Adjacent neurons in the rodent visual cortex are tuned to a variety of stimulus orientations[7], whereas those in the primate visual cortex are tuned to a specific stimulus orientation[5]. The present results suggest that, compared with the primate cortex, the relatively small number of neurons in the rodent cortex may cause the dissimilar orientation preferences of adjacent neurons in the rodent visual cortex.

Methods

mcAlexNet, mAlexNet, and smcAlexNet were constructed and trained as described in a previous report[17]. The networks consisted of five hierarchically organized convolutional layers (c1–c5) and three pooling layers. c1 of mcAlexNet contained 32 parallel streams (s0–s31), each equipped with two filters. c2 contained 16 parallel streams (s0–s15), each equipped with 12 filters. Each c2 filter only received converging inputs from four filters in two streams of c1. c3 contained eight parallel streams (s0–s7), each equipped with 48 filters. Each c3 filter only received converging inputs from 24 filters in two streams of c2. c4 contained four parallel streams (s0–s3), each equipped with 64 filters. Each c4 filter only received converging inputs from 96 filters in two streams of c3. Finally, c5 contained two parallel streams (s0 and s1), each equipped with 128 filters. Each c5 filter only received converging inputs from 128 filters in two streams of c4. Outputs from the two streams of c5 were concatenated and fed into FCs, and then fed to the output layer for classification (Fig. 1).

mAlexNet was organized in a fully parallel manner from the first convolutional layer (c1) to the last convolutional layer (c5) without convergence. Outputs from the 32 streams of c5 were concatenated and fed into FCs, and then fed to the output layer for classification. Each stream of c1, c2, c3, c4, and c5 of mAlexNet contained two, six, 12, eight, and eight filters, respectively (Supplementary Fig. 3).

smcAlexNet contained smaller number of filters than mcAlexNet, but its architecture was similar to that of mcAlexNet. c1 of smcAlexNet contained eight parallel streams, each equipped with two filters. c2 contained four parallel streams (s0–s3), each equipped with 12 filters. Each c2 filter only received converging inputs from four filters in two streams of c1. c3 contained two parallel streams (s0 and s1), each equipped with 48 filters. Each c3 filter only received converging inputs from 24 filters in two streams of c2. c4 contained two parallel streams (s0 and s1), each equipped with 32 filters. Each c4 filter only received inputs from 48 filters in one stream of c3. Finally, c5 contained two parallel streams (s0 and s1), each equipped with 32 filters. Each c5 filter only received inputs from 32 filters in one stream of c4. Outputs from the two streams of c5 were concatenated and fed into FCs, and then fed to the output layer for classification.

The orientation and color selectivity of each filter of c1 were quantified using selectivity indices. If a filter did not develop any structure (i.e., a flat kernel; for an example, see Supplementary Fig. 3), the filter was excluded from the analyses. Orientation selectivity was evaluated using circular variance[18], as follows:

$$OI = 1 - \frac{\left| \sum {}_i R(\theta_i)\, e^{\frac{2\pi\, i\theta_i}{180^\circ}} \right|}{\sum {}_i R(\theta_i)},$$

Here, OI is the orientation index and $R(\theta_i)$ is the filter weight amplitude at the i-th orientation. Filter weight amplitude was calculated by summating the amplitude within ±15°, and was examined with an interval of 30°. The OI had a value between zero and one, and a larger OI reflected stronger orientation selectivity. The ΔOI was the absolute difference in OI between two filters and represented the similarity in OI. The pO was obtained by calculating the average vector in the complex plain. The ΔpO was the absolute difference in pO between two filters and represented the similarity in pO.

Color selectivity was evaluated by calculating the *r* of the filter weight among R, G, and B channels[17]. If a filter was not color selective, weight values were correlated among channels. The smallest *r* among the three (r_{min}) was selected, and the CI was obtained using the following formula:

$$CI = r_{\min} \times (-0.5) + 0.5$$

The CI had a value between zero and one, and a larger CI reflected greater color selectivity. The ΔCI was the absolute difference in CI between two filters and represented the similarity in CI. Color preference was quantified using RGB values of a single pixel that was selected in the following way. The minimum and maximum weight values across all 64 c1 filters were transformed to RGB values by scaling them to 0 and 255, respectively. Next, the RGB values of a filter were summated for each R, G, and B channel. Among the summated RGB values, the largest color channel was selected. The pixels of the color channel with the largest value were then selected. From the RGB values of the pixels, the pH of the filter was calculated by converting the RGB color space to HSV color space. The ΔpH was the absolute difference in pH between two filters and represented the similarity in pH.

The responses of filters were examined using 1,000 images. From each of 1,000 categories of the validation set of ImageNet database[16], one stimulus image was randomly selected to create a set of 1,000 stimulus images. The set was consistently used in the present study. Filter outputs (responses) were calculated for each stimulus image, and a set of responses of filters to 1,000 images was obtained. The similarity in the set of responses between two filters was evaluated by calculating the distance measure (dR), which was obtained using the following formula:

$$dR = (-0.5) \times r_R + 0.5$$

Here, r_R is the Pearson's *r* value of the set of responses between filters. The dR had a value between zero and one. If the responses to the stimulus set were similar to each other, dR was zero. The dR was examined using filters that had receptive field at the center.

Filter weight structure was evaluated using three measures. One measure was W, which was the pixel-by-pixel filter weight, and was sensitive to exact *x*–*y* positions. Another measure was Was, which was the amplitude spectrum calculated after two-dimensional discrete Fourier transform of filter weights, and was not sensitive to shifts in *x*–*y* positions of filter weights. For the third measure, Whist, filter weights were discretized and a histogram was obtained; this measure was not sensitive to filter structure. From these measures, distances (dW, dWas, and dWhist) were calculated using the following formulae:

$$dW = (-0.5) \times r_W + 0.5$$

Here, r_W is the Pearson's *r* of W between two filters.

$$dWas = (-0.5) \times r_{Was} + 0.5$$

Here, r_{Was} is the Pearson's *r* of Was between two filters.

$$dWhist = (-0.5) \times r_{Whist} + 0.5$$

Here, r_{Whist} is the Pearson's *r* of Whist between two filters.

To examine the contribution of each stream of c1 to image classification, a single-stream-deletion experiment was performed. To delete a stream of c1, all output values of the max-pool layer of the stream after c1 were set to zero. The classification accuracy (top-5 accuracy) was calculated using the validation set for the original mcAlexNet and single-stream-deleted mcAlexNet. The Δacc was the difference between the top-5 accuracy calculated using the original mcAlexNet and that calculated using single-stream-deleted mcAlexNet (i.e., Δacc = top-5 accuracy calculated using original mcAlexNet − top-5 accuracy calculated using single-stream-deleted mcAlexNet).

Statistical analysis

All data were pooled for the statistical analyses. Analyses were performed using Python libraries (pandas, numpy, scipy, scikit-learn, matplotlib, and seaborn). The statistical tests used in the present study were the Mann–Whitney U test (two-tailed), Wilcoxon signed-rank test (two-tailed), Friedman test for repeated samples, and Kruskal–Wallis H-test. All *r* values are Spearman's rank correlation unless otherwise stated. The statistical threshold for *p*-values was set at 0.05. Median values were calculated to represent populations, whereas the mean ± SD was calculated to represent population measures of module size and *r* values.

Data availability

Parts of the datasets generated and/or analyzed during the current study are available at the Osaka University Knowledge Archive (https://hdl.handle.net/11094/96433; https://doi.org/10.60574/96433). The remaining data are available from the corresponding author on reasonable request.

Received: 16 July 2024; Accepted: 30 December 2024
Published online: 02 January 2025

References

1. Mountcastle, V. B. The columnar organization of the neocortex. *Brain* **120**, 701–722 (1997).

2. Livingstone, M. S. & Hubel, D. H. Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* **4**, 309–356 (1984).
3. Garg, A. K., Li, P., Rashid, M. S. & Callaway, E. M. Color and orientation are jointly coded and spatially organized in primate primary visual cortex. *Science* **364**, 1275–1279 (2019).
4. Chatterjee, S., Ohki, K. & Reid, R. C. Chromatic micromaps in primary visual cortex. *Nat. Commun.* **12**, 2315 (2021).
5. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215–243 (1968).
6. Rockland, K. S. & Ichinohe, N. Some thoughts on cortical minicolumns. *Exp. Brain Res.* **158**, 265–277 (2004).
7. Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
8. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (W.H. Freeman, 1982).
9. Purves, D., Riddle, D. R. & LaMantia, A. S. Iterated patterns of brain circuitry (or how the cortex gets its spots). *Trends Neurosci.* **15**, 362–368 (1992).
10. Horton, J. C. & Adams, D. L. The cortical column: a structure without a function. *Philos. Trans. R Soc. Lond. B* **360**, 837–862 (2005).
11. Kaas, J. H. Evolution of columns, modules, and domains in the neocortex of primates. *Proc. Natl. Acad. Sci. U S A.* **109** (Suppl 1), 10655–10660 (2012).
12. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **27**, 1097–1105 (2012).
13. Gilbert, C. D. & Wiesel, T. N. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* **9**, 2432–2442 (1989).
14. Man, K., Kaplan, J., Damasio, H. & Damasio, A. Neural convergence and divergence in the mammalian cerebral cortex: from experimental neuroanatomy to functional neuroimaging. *J. Comp. Neurol.* **521**, 4097–4111 (2013).
15. Wertz, A. et al. PRESYNAPTIC NETWORKS. Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science* **349**, 70–74 (2015).
16. Deng, J. et al. ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (2009).
17. Tamura, H. An analysis of information segregation in parallel streams of a multi-stream convolutional neural network. *Sci. Rep.* **14**, 9097 (2024).
18. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation selectivity in macaque V1: diversity and laminar dependence. *J. Neurosci.* **22**, 5639–5651 (2002).
19. von der Malsburg, C. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* **14**, 85–100 (1973).
20. Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
21. Bush, P. C. & Mainen, Z. F. Columnar architecture improves noise robustness in a model cortical network. *PLoS ONE* **10**, e0119072 (2015).
22. DeAngelis, G. C., Ghose, G. M., Ohzawa, I. & Freeman, R. D. Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J. Neurosci.* **19**, 4046–4064 (1999).
23. Tamura, H. Pairwise correlations of spiking activity changes along the ventral visual cortical pathway of macaque monkeys. *bioRxiv* 220301 (2017).

Acknowledgements
I thank Bronwen Gardner, PhD, from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

Author contributions
HT designed the research, conducted the experiments, analyzed the data, and wrote the paper.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-84981-1.

Correspondence and requests for materials should be addressed to H.T.