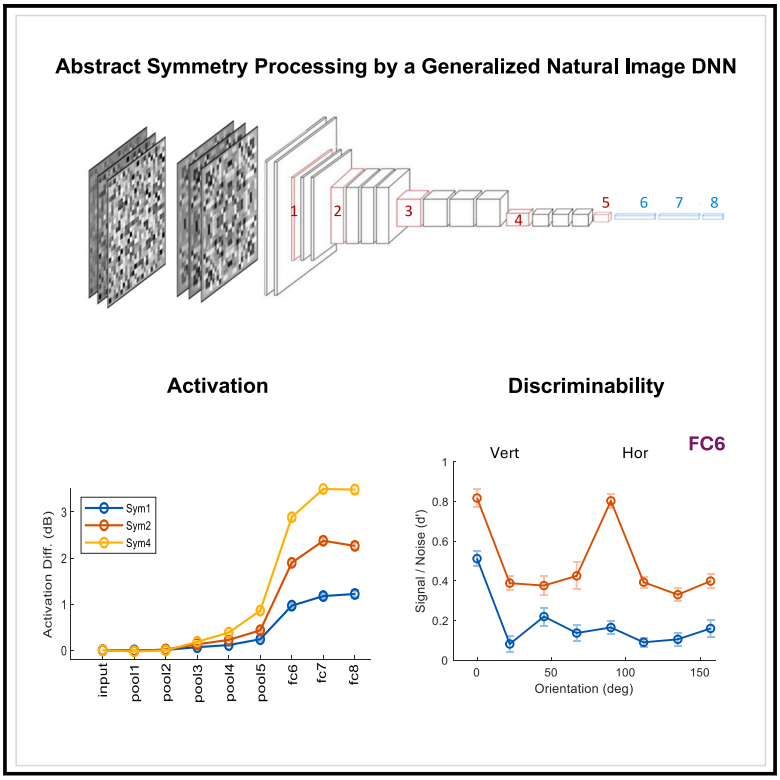
iScience

Article

The canonical deep neural network as a model for human symmetry processing

Graphical abstract



Highlights

- The abstract property of symmetry transcends recognition by definable shape templates
- We tested a DNN trained on natural images with object-free random-dot symmetry images
- Like in the human brain, symmetry coding only emerged in later layers of the DNN
- Feedforward DNN symmetry responses were homologous with human global processing properties

Authors

Yoram S. Bonneh, Christopher W. Tyler

Correspondence

yoram.bonneh@gmail.com

In brief

Applied sciences; Computer science; Signal processing

Bonneh & Tyler, 2025, iScience *28*, 111540
January 17, 2025
https://doi.org/10.1016/j.isci.2024.111540

CellPress

iScience

CellPress
OPEN ACCESS

Article

The canonical deep neural network as a model for human symmetry processing

Yoram S. Bonneh[1,2,5,*] and Christopher W. Tyler[3,4]
[1]School of Optometry and Vision Science, Faculty of Life Science, Bar-Ilan University, Ramat-Gan 5290002, Israel
[2]The Leslie and Susan Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat Gan, Israel
[3]Division of Optometry, Health Sciences, City University of London, London EC1V 0HB, UK
[4]Smith-Kettlewell Eye Research Institute, San Francisco, CA 94115, USA
[5]Lead contact
*Correspondence: yoram.bonneh@gmail.com
https://doi.org/10.1016/j.isci.2024.111540

SUMMARY

A key property of our environment is the mirror symmetry of many objects, although symmetry is an abstract global property with no definable shape template, making symmetry identification a challenge for standard template-matching algorithms. We therefore ask whether Deep Neural Networks (DNNs) trained on typical natural environmental images develop a selectivity for symmetry similar to that of the human brain. We tested a DNN trained on such typical natural images with object-free random-dot images of 1, 2, and 4 symmetry axes. Symmetry coding was negligible in the earliest DNN layers. The strongest discriminability occurred in the first fully connected layer, FC6, plausibly analogous to the human lateral occipital complex (LOC), matching many structural properties of human symmetry processing. These results support the homology between the feedforward DNN trained on natural images and the global processing of the extended visual hierarchy as it has evolved in the human brain.

INTRODUCTION

The concept of reflection symmetry

Reflection symmetry is an abstract property of images that is expressed by the presence of similar elements related to mirror-image inversion. The symmetry is defined as global when the midpoints of the separations between multiple such elements are aligned to form a symmetry axis (particularly when there is no local symmetry information along this axis). The present study addresses the question of whether computational neural networks trained to analyze structure in natural images have the capability to extract pure symmetry information independently of the presence of objects in the image. To assess this capability, we rely on two levels of operational definition. The concept of pure symmetry is operationally defined as the mirror-symmetry relations occurring in a random pattern that has been reflected in one or more axes. Such patterns contain no recognizable objects, and the goal of the system is to be able to distinguish between such a symmetric image and one that does not contain such mirror-symmetry relationships (beyond any that would occur in a random image by chance alone). Thus, the second level of operational definition is that pure symmetry processing is defined as the ability to discriminate between symmetric images and their random counterparts. In the present design, the control random images are obtained by randomly scrambling the symmetry images in order to equate the probabilities of repeated elements that are inherent in the symmetric images (see STAR Methods for details).

The deep neural network model of hierarchical perceptual organization in the human brain

Deep neural network (DNN) models developed for image classification have been suggested as biologically inspired models for visual processing.[1] Such DNN models have provided important insights into the constraints on image processing, and shown interesting relationships to the hierarchical neural organization of image processing in the human brain as well as to behavioral data on human perception. Examples of such relationships are: general object recognition at the neural level, similarity perception,[2,3] scene recognition,[4] crowding,[5] face processing,[6,7] visual search asymmetry,[8] and other aspects of visual perception.[9–14] The validity of the correspondence between DNN performance and human brain organization has also been questioned in some cases.[15,16]

For the Gestalt perceptual property of reflection symmetry, one DNN study found that, unlike in human perception, symmetry did not affect object discrimination by a DNN, suggesting that it developed object coding without reference to symmetry relations.[17] Another study[18] decoded symmetry of natural images from a five-layer DNN specifically trained for symmetry detection with high correlation to the human rating of symmetry of natural images. It is as yet unclear whether a DNN trained on natural images, in general, could have a selectivity for symmetry in unnatural images such as random dot patterns, and to what degree its performance would resemble human observers. We are interested in this capability because symmetry is a mid-level object property that is largely independent of the edge processing

and regional segmentation processes that are typically considered as the ubiquitous basis for early visual processing.

Consequently, we took a deeper look into the function of DNN symmetry processing by probing a pre-trained ImageNet visual DNN to discriminate abstract symmetry images, which excluded other object information because they were based on random texture seeds. In this way, we could test whether a DNN trained on the typical range of image structures in the natural environment would be sensitive to the presence of various forms of the Gestalt property of long-range reflection symmetry uncontaminated with other image features. The results can assess whether the pre-trained DNN is an effective model of the mechanisms of symmetry perception in the human brain, where symmetry signals are found only in upper levels of the visual hierarchy.[19] The key question for the present study was whether the DNN exhibited similar properties, thus shedding light on the processes underlying the human symmetry-processing characteristics. For this purpose, we first review the relevant properties of human symmetry perception, particularly those that distinguish it from local form and contour perception processes that are well-known to be the key features employed by DNNs trained on natural images (e.g., Olshausen & Field[20]).

Symmetry as a canonical paradigm of mid-level vision

Mid-level visual processing is conceived as the encoding of featural properties of the world at a level of complexity above the visual processing at the local, first-order, and second-order processes that have been known for half a century or more, but prior to the high-level processes of full object perception and categorization. Such visual processing divisions are somewhat arbitrary from a purely theoretical perspective based on the information available in the optic array (the plenoptic function,[21]; but are neurophysiologically inspired by the processes that have evolved to take place in the human brain. Although these processes have local symmetries,[22] they are not well suited to extracting the long-range symmetry structure of objects in the world such as are found in many living organisms and much of the human constructed environment of buildings, vehicles, pottery, and so on. These long-range pattern symmetries are formally defined in the Appendix and require specialized long-range neural mechanisms for their encoding.

More specifically, local processing is conceptualized as the kind that takes place in the retina, including specialized spectral encoding, areal summation, luminance gain control, local lateral inhibition, local contrast and color processing, and sustained/transient specialization. First-order spatiotemporal processing can be conceptualized as the kind that takes place in the first levels of the primary visual cortex,[23] elaborating the retinal signals to coding for orientation, scale (spatial frequency), phase (edge encoding), motion speed and direction, binocular disparity and the corresponding three-dimensional properties of stereo-motion direction and orientation.[24] Second-order processing can be conceptualized as the kinds of spatial pattern interactions that take place in higher retinotopic cortical representations, although due to recurrent feedback interactions, their signature can also be found in the primary visual cortex. These second-order processes are spatial contextual interactions of contrast, motion, and color as revealed by many visual illusions, Gestalt grouping and figure/ground organizations,[25,26] and long-range spatial disparity structure encoding.[27,28] One particular study has identified neural responses specific to global symmetry relations in images,[29] but otherwise the neural basis for symmetry processing is unknown. How this progression from local to global processing is structured in the neuroanatomy of the visual processing hierarchy in the context of the DNN model is addressed in the Materials and Methods section.

Human symmetry perception

The detection of mirror symmetry is an important probe into the global properties of form perception. Imposing symmetry on unpatterned (random-element) stimulus fields ensures that it is processed as the abstract transformation rather than via the shapes of local pattern elements (except close to the axis of symmetry). The ability to match an unanticipated pattern within a presentation time of less than 100 msec (and therefore less time than required for an eye movement), reveals important parallel-processing properties of human pattern processing. The match between pairs of symmetric pattern elements is invisible in the Fourier amplitude spectrum, being an obscure property of the phase spectrum, that depends on the location of the symmetry axis. Symmetry cannot, therefore, be extracted from differential activation strength in a particular spatial frequency channel, but requires specialized global matching properties.

It is often said that symmetry is extracted by selective processing of signal along the axis of reflection. However, Tyler et al.[30] showed that, with dense random arrays of dots, the symmetry could be detected even with the randomization of an axial strip up to 5 deg wide, requiring pattern matching far beyond any plausible local axis information. Moreover,[31] found that the sensitivity to symmetry as a function of eccentricity (in uniform random-dot noise) showed a reverse eccentricity scaling, with the axial randomization width becoming narrower with eccentricity, rather than wider as would be predicted by a cortically scaled receptive-field model of symmetry processing. These results again support a global pattern-matching model rather than a local receptive-field model of symmetry processing. Tyler & Hardage[32] then extended this result using eccentricity-scaled stimuli to show that patches of noise scaled out to the far periphery, from 32° to 64° eccentricity, could support symmetry detection (with dynamic noise masks) for presentation durations as short as ~50 msec, equally short at any eccentricity (Figure 1A). These conditions involve pattern symmetry at separations of 64° on the retina, implying a truly long-range pattern-matching capability for identifying patches of symmetry pattern.

Other nonlinear and global properties of the noise-matching process revealed by symmetry detection tasks were also reported.[32] Patterns of one-tenth density could be detected more readily than those of full density, a property of second- order (rectifying) rather than first-order (linear filter) systems. This interpretation was tested by using patterns of ***opposite polarity*** across the symmetry axis (anti-symmetry), a manipulation to which a second-order mechanism should be impervious. Sensitivity was slightly reduced at high density, though still uniform across eccentricity. At low density, there was no significant reduction in sensitivity for opposite-polarity relative to same-polarity stimuli, supporting a dominant role in the second-order

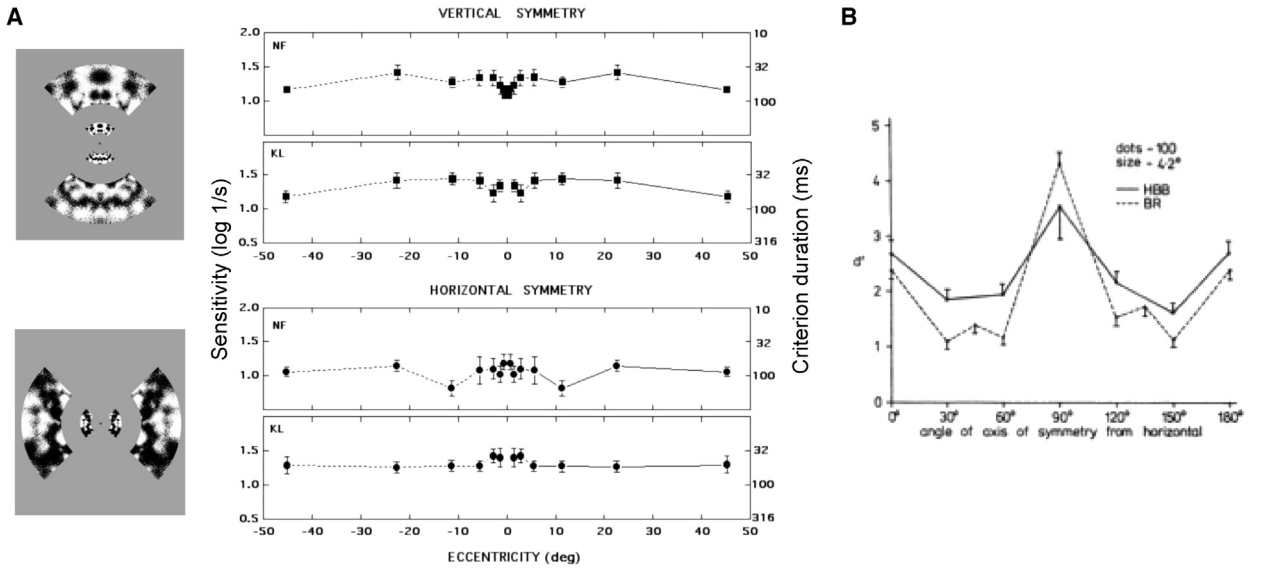


Figure 1. Human symmetry detection data

(A) Human data for the detectability of symmetry scaled in proportion to eccentricity across the visual field to on the horizontal and vertical meridians (from Tyler & Hardage, 1996). Error bars denote ±1 SEM across trials.

(B) Orientation selectivity in the human performance of random-dot symmetry discrimination as a function of the orientation of the symmetry axis (from Barlow & Reeves, 1979).

processing of symmetry. In summary, the human symmetry detection results combine to reveal nonlinear, long-range, global mechanisms of pattern matching in noise stimuli, well outside those predicted from standard simple filter models. To extend this comparison to another relevant dimension, we reproduce a figure from Barlow & Reeves[33] of the d' for random-dot symmetry discrimination as a function of the orientation of the symmetry axis (Figure 1B) showing that there is enhanced discernibility for the vertical and horizontal axis orientations.

Here, we ask whether the canonical form deep neural network(DNN), as trained on the standard vocabulary of natural images of environmental objects[34] would exhibit discrimination results comparable to the human response to image symmetry, not only with respect to the standard forms of abstract (random-dot) symmetry but also for the more extreme forms of human long-range symmetry processing, with the axis occluded (as in Figure 1A, lower left panel) and with opposite polarity pattern on either side of the symmetry axis.

RESULTS

Effect of symmetry across the deep neural network layers

The primary measure of the activation difference between symmetry and fully random images (see Figures 2A and 2C) across the probed gni-DNN layers (pooling and FC layers, see Figure 3B) is plotted in Figure 4A. It is computed as the L2 difference in dB between the response maps in a layer of pairs of images, symmetric and fully random, averaged across N = 1000 image pairs and plotted relative to the 1-axis symmetry results at the input layer (see STAR Methods). As shown, the differential symmetry activation is very low for the first two layers, then progressively increases up to later layers for the 1-, 2- and 4-axis symmetries, with overall strength increasing with a number of axes. The pattern is similar for the three levels of symmetry, showing a strong increase between layers Pool5 and FC6. The same activation difference analysis for the block-randomized control images used later as the symmetry discrimination baseline is shown in Figure 4B. (All SEM error bars in Figures 4A and 4B are smaller than the symbols due to the large N, and the mean values reflect significant differences between the DNN responses for the three symmetry types across layers.) These images have residual local regularities that could make them distinguishable from fully random images, as is shown by the increasing activation differences from layer Pool4 and higher, for the 2- and 4-axis block-randomized control images (Sym2-br, Sym4-br). Note, however, that the 1-axis symmetry has no significant signal and no discriminability from random in any layer, indicating that the symmetry provides no local cue that the gni-DNN can use to discriminate it from random.

However, to assess the symmetry-specific discriminability of the model, it is helpful to compute d-prime measures between the activation differences for symmetry (Figure 4A) and the appropriate baseline of block-randomized controls (Figure 4B, see Methods/Stimuli). The d-prime measures can show significant differences even in conditions with very low activation differences when their variability is low. The d-prime discriminability measures between the data shown in Figure 4A with the corresponding baseline data in Figure 4B are shown in Figure 4C. Positive values reflect a larger activation difference for the symmetry images compared to the baseline of the corresponding block-randomized control images. As shown, a large and significant

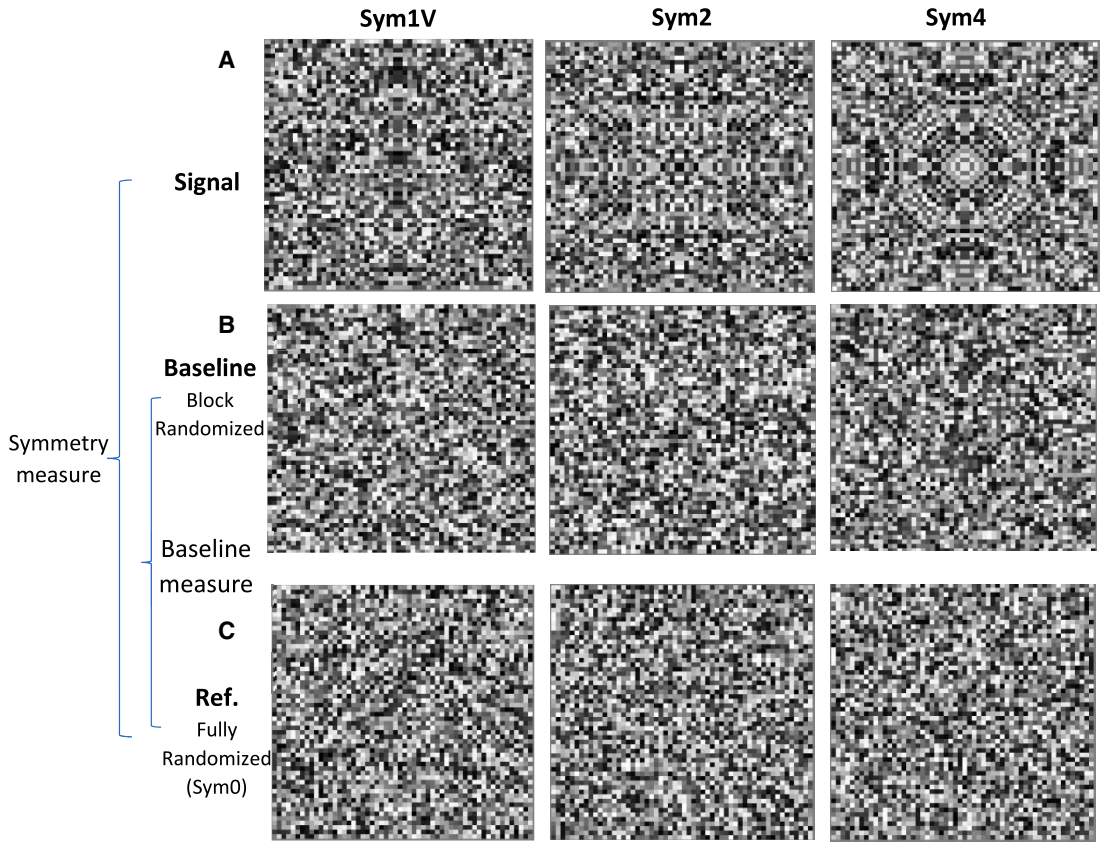


Figure 2. Examples of stimuli used

The stimuli consisted of 1-, 2-, and 4-axis random-dot symmetry images (Sym1V, Sym2, and Sym4, respectively, top row, (A). The reference stimuli (bottom row, C) were fully random images (3 examples shown). In addition to the symmetry images used as "Signal," block-randomized control versions of the symmetry images (middle row, (B) were paired with fully randomized images as a reference and were used as a baseline for comparison. Texel size was 4x4 pixels as shown (except for the axis orientation study), with the blocks for order randomization being 4x4 texels in size resulting in 14x14 blocks, and the images were 224 x 224 pixels in size. Note that the differences between the fully random and the baseline (block-randomized) control images are subtle and difficult to detect, but were sufficient to generate differences in the model's response relative to fully random images.

symmetry effect started from layer Pool3, slightly reducing in layers Pool4 and Pool5, then increasing again to peak at layer FC6. In all layers from layer Pool3 and up, the d-primes increased between 1-, 2-, and 4-axis symmetry with a maximum close to d' = 1 (Figure 4C, blue, orange, and yellow bars). However, it is important to note that this level of discriminability implies that the DNN would perform at only about 75% correct on the basis of the output from one of these layers for the patterns of perfect symmetry, which the human visual system would discriminate essentially perfectly in exposures of the order of 100 ms.[35]

To complete the picture, we also computed the discriminability between the block-randomized control measures (Figure 2B) and a baseline of fully random images (Figure 2C), to produce the d-prime measures shown in Figure 4D. These measures show some similarity to the d-primes for the symmetry images of Figure 4C, but the values are much lower and the advantage of the 4-axis case is relatively stronger here.

Effect of symmetry type and axis

For a closer look at the effect of symmetry type, as well as preferred axis, we focused on two layers that showed a strong effect in Figure 4C: a "low" layer, Pool3, and a "high" layer, FC6. The results are shown in Figure 5 in the same format as in Figure 4, with the activation results on the left and discriminability on the right. The L2 distances increased from a low level to about 3 dB for increasing numbers of symmetry axes in FC6 but only about 1/10th of this response for Pool3 (Figures 5A–5C, full curves). The block-randomized control patterns show a much weaker response with a similar increase up to about 1 dB in FC6 and an essentially flat function in Pool3 (dashed curves). The d' plots (Figures 5B–5D) show that the discriminability does indeed increase as more axes of symmetry are introduced. However, although there is a big increase in the symmetry signal from 2-axis to 4-axis symmetry, there is also a corresponding increase in the response to the block-randomized control patterns, such that the net discriminability is not significantly increased by doubling the number of axes of symmetry from 2 to 4. The significant discriminability of the symmetric from the randomized images implies that, somehow, the gni-DNN training on natural images gives it the surprising capability of extracting symmetry information in random images that do not contain any of the objects on which it was trained.

A

B

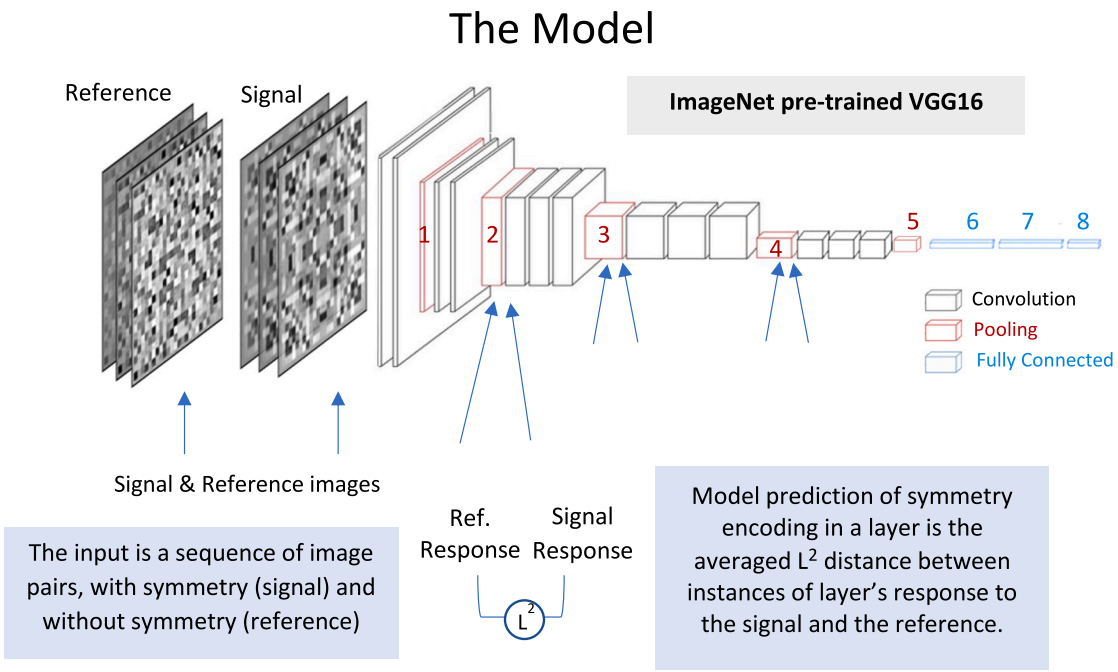


Figure 3. The model and pre-training ImageNet samples

(A) Examples of the ImageNet images used to train the DNN. Circles indicate those with a significant degree of image symmetry of predominantly 1-axis (white), 2-axis (yellow), 4-axis (pink), or 8-axis (red) symmetry.

(B) The generalized natural image DNN model of symmetry perception. The model uses an ImageNet pre-trained VGG16 network assumed to have learned the regularities of natural images, including any naturally occurring symmetries. To test this, pairs of random-dot images with and without symmetry were used to probe the model and compute the L2 metric between the response maps of different model layers as a measure of symmetry.

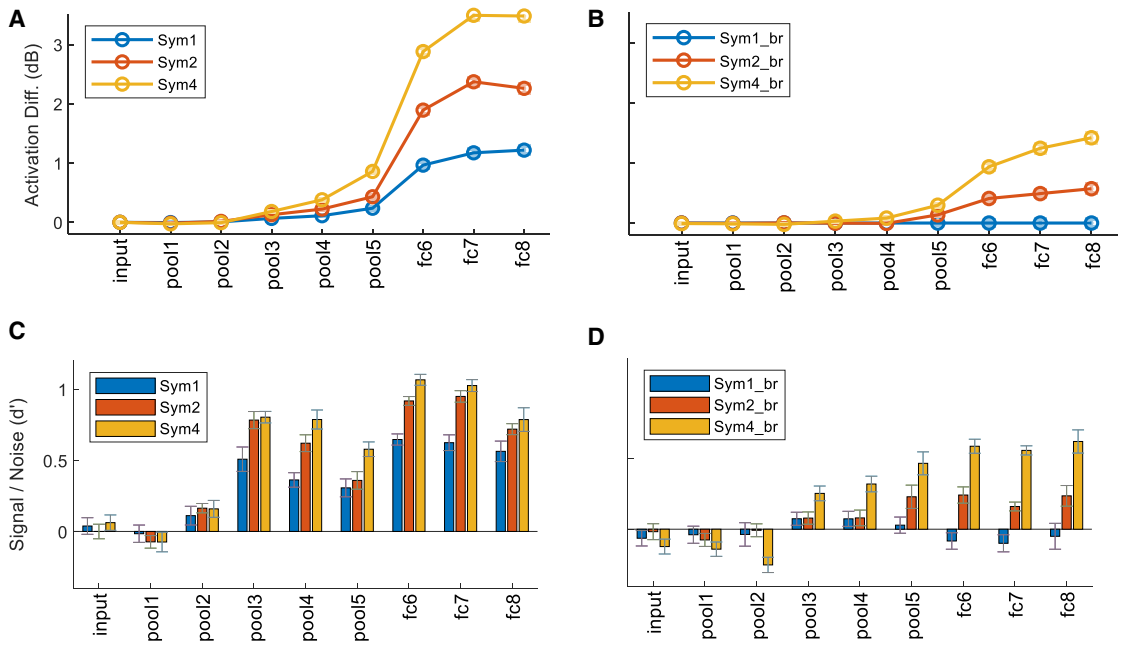


Figure 4. The effect of symmetry across the gni-DNN layers, for the different symmetry types

The results were obtained with a texel size of 4x4 pixels, for vertical 1-axis, 2- axis, and 4-axis symmetry, referenced to fully random images (left column, A and C) and for an additional baseline computed from block-randomized control images for the three types (right column, B and D). The activation difference in dB for the different layers relative to the 1-axis symmetry results at the input layer are plotted in the upper row (A and B). The discriminability measure (SNR as d') is plotted on the lower row (C and D), with the SNR between the symmetry measures and the block-randomized baseline in (c), and the SNR between the block-randomized control images and the fully random images (i.e., fully random vs. fully random activation diff) in (D). Error bars denote 1 SEM across $N = 1000$ images for the activation values in (A and B), and across 10 bins of 100 images in (C and D).

It is interesting that layer FC6 of the gni-DNN shows significantly greater discriminability in Figure 5D for vertical than horizontal symmetry (while, as expected because they have no orientation bias, the block-randomized control versions did not show a difference). Thus, training the DNN on real-world images gives it a preferential capability for recognizing vertical symmetries.

Effect of symmetry density

The effect of the "symmetry density" of the images is assessed for layer FC6 in Figure 6, where a randomly selected proportion of the dots in the image was replaced with random dots containing no symmetry information. The three curves for the different numbers of symmetry axes separate as expected toward the 100% symmetry level, but it is interesting that they converge to about the same low, but significant, symmetry response level by the 40% proportion of symmetry. Beyond this point, the curves increase roughly with the square of the symmetry percentage, as expected for some form of correlation process.

Orientation specificity

The orientation specificity was explored further by reducing the texel size to 1 pixel and windowing the targets in a circular aperture (see Figures 7A and 7B). This allowed the targets to be rotated through 8 orientations to test the orientation specificity of the gni-DNN, with the results shown in Figures 7C–7F. All curves are plotted relative to the random condition for the corresponding dot size. As shown, the strongest symmetry effect was found in the cardinal orientations (0° and 90°), necessarily similar for the 2-axis symmetry (solid red lines) but with differential results for 1-axis symmetry (solid blue lines) depending on the gni-DNN level. Performance was similar in both axis orientations in Pool3, but also for the block-randomized control patterns (dashed lines in Figures 7C and 7E). By layer FC6, however, the response was much stronger for the vertical (0°) orientation for the 1-axis symmetry but showed little variation for its block-randomized versions. Thus, the discriminability at FC6 retained this strong vertical axis preference, whereas it was a weaker effect in Pool3.

Two-axis symmetry showed a strong cardinal axis preference at both processing levels. Note, however, that the combined results in layer FC6 are compatible only with a single (vertical) axis preference since the 2-axis symmetry is ambiguous in that it has an axis on the vertical for both cardinal orientations, whereas the 1-axis symmetry is discriminative in that only the vertical axis shows the enhancement.

Effect of image size

One way to assess the performance of early versus later layers of the gni-DNN is to study the effect of image size, as shown in Figure 8 for the discrimination performance with the original image format of Figure 2. The pattern of results is similar for all three levels of symmetry. In the early layer (Pool3), discrimination performance improved dramatically up to an image width of about

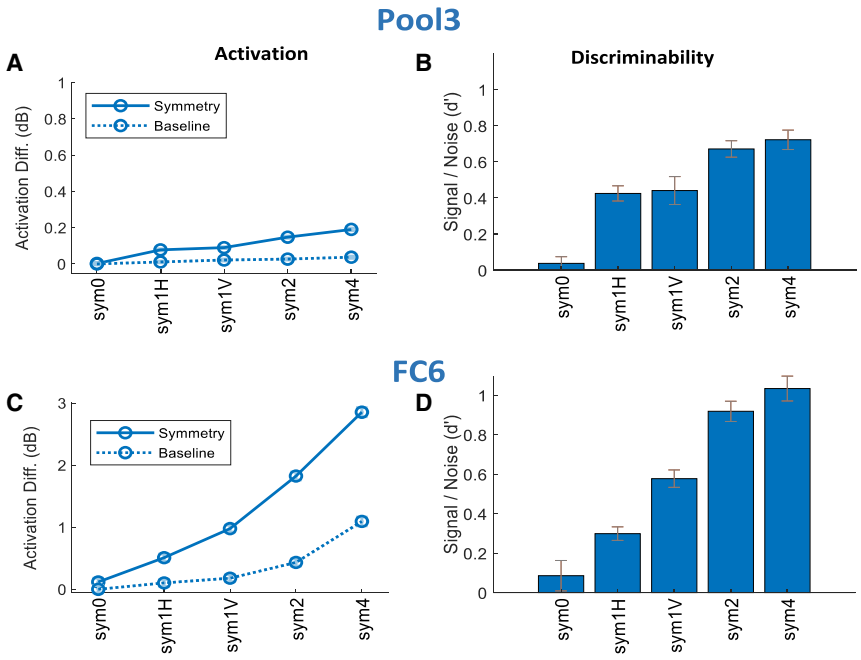


Figure 5. The basic symmetry effect

The L2 difference between the activation maps of a layer was computed for each pairing of random-dot symmetry images with a fully randomized reference image (N = 500 images). To control for local n-gram structure, the same comparison was run for block-randomized (4x4 texel blocks) control images. The results are shown for layers Pool3 (A and B) and FC6 (C and D). The left plots (A and C) show the activation difference in dB relative to a baseline of Sym0 (fully random) images, for 1-axis H,V, 2-, and 4-axis symmetry images (solid lines), and the corresponding block-randomized versions referenced to sym0 images in dotted lines. The right column (B and D) plots the discriminability measures (signal-to-noise as d-prime) for the data in (A and C). As shown, the patterns of results for the early (Pool3) and later (FC6) channels were similar, though with a much smaller difference in dB for the Pool3 results, and a slightly worse discriminability power (d-prime). Error bars denote ±1 SEM across images.

14 texels but leveled off beyond that, implying that increasing the image area (or total number of texels) by a further factor of 16 has little advantage. Thus, even though the total length of the symmetry axes is quadrupling over this range, the performance only improved by 30% or less beyond 24 texels, implying that the gni-DNN is only performing local processing up to layer Pool3. The FC6 performance, on the other hand, shows weaker discrimination for the smaller sizes but continues to improve strongly over the range beyond 14 texels, with improvements of 230–300%. Thus, the later fully connected layer draws from a much larger area of the pattern and ultimately wins out in terms of discrimination performance.

Effect of a gap around the symmetry axis

The final issue was to probe the role of paraxial symmetry information with an analysis of the effect of gap size on the symmetry response. In order to eliminate the possibility of local processing of symmetry, by the relative activation in classical symmetric vs. asymmetric receptive fields (corresponding to the enhanced activation along the symmetry for cosine wavelet analysis of symmetry patterns), the introduction of a gap between the mirror-symmetric fields showed that symmetry was still perceivable In human studies, though at substantially reduced levels.[32,33] The latter study also included a condition in which the dot contrast polarity was inverted across the symmetry axis. Nevertheless, performance was quite similar to the same-polarity condition for all gap sizes, indicating that human mirror symmetry processing has a large second-order (polarity-insensitive) component.

For comparison with this result, the gni-DNN performance was tested with stimuli composed of strips of 1-axis symmetric random pattern separated by various distances (Figure 9A). Gaps from 0 to 32 texels were tested on vertical 1-axis symmetry. Note that the Cartesian coordinates of these stimuli correspond to the cortical projection of the radial stimuli of the stimuli of Tyler & Hardage,[32] which were designed to take into account the approximately polar transform of information from the retina to the cortex. As for the previous gni-DNN analysis, the gap-stimulus analysis included discriminability from a baseline of block-randomized control images, as well as from a baseline of fully random images.

The result is that the d-prime for 1-axis symmetry discrimination (Sym1 relative to Sym1-br) for Pool3 (Figure 9B) began at a high level and was rapidly reduced as gap size was increased, falling to zero at a gap size of 4 texels and remaining close zero for all larger gap sizes. For FC6, a similar pattern was observed for small gap sizes, but the symmetry discrimination stayed almost constant for the 4-texel gap size and larger, at around a d-prime of 0.6 (Figure 9C). We made two consistency checks, showing a very similar d-prime effect for fully random baseline images (Sym1/Sym0) as for the block-randomized control (Sym1/Sym1-br), as shown by the blue vs. red plots in Figure 9, as well as close to zero effect for all gap sizes for block-randomized images relative to fully random (Sym1-br/Sym0, in yellow). Moreover, the results for the inverse-polarity symmetry images (Figures 9D and 9E) were very similar to the results for the same-polarity ones (compare Figures 9C and 9E). In sum, the strong discriminability for 1-axis symmetry with large gaps in FC6 (blue curve in Figure 9C) is clear evidence that the gni-DNN can process symmetry by mechanisms beyond the differential activation of local cosine receptive fields, such as long-range pattern matching, for example.

The manipulation of inverting the contrast polarity of one of the pair of symmetric fields in the symmetry stimulus is interesting because it probes whether the symmetry processing is first-order (polarity-specific) or second-order (polarity-blind). If the symmetry were processed in the brain through the symmetry properties of even-symmetric simple-cell receptive fields, it

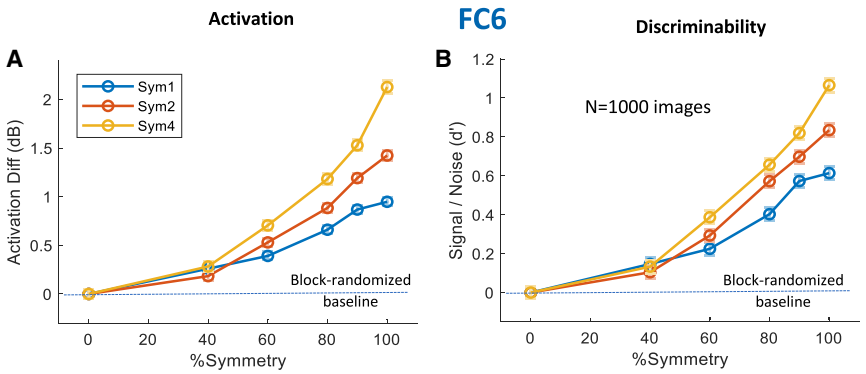


Figure 6. The effect of the symmetry type and density

The symmetry density (percentage of symmetric texels) was manipulated by mixing symmetric images and their block randomized counterparts with some proportion of fully randomized images, referenced to fully randomized images. The plots compare 1-axis vertical, 2-, and 4-axis symmetry. (A) Activation differences as a function of percent-symmetry referenced to fully randomized images and presented in dB relative to the zero-symmetry condition for each type.

(B) The same data presented in d-prime values, as the difference between the symmetry measures and the corresponding block-randomized control images expressed in SD units. Note the steep increase of the symmetry effect with the percentage of symmetry points from 60%. Data were averaged across *N* = 1000 images of 56x56 texels. Error bars denote ±1 SEM.

would be polarity-specific. (Note that, if the local processing was based on a combination of even- and odd-symmetric receptive fields, or Gabor filters of all possible carrier phases, the net activation would not be symmetry-specific. It has to encode a matching reflected *pattern* of activations across receptive field positions and sizes in order to discriminate symmetry per se.). If, on the other hand, it were processed by more specialized pattern-processing mechanisms later in the visual processing hierarchy,[19] it might be expected to be polarity blind, as was largely found by Tyler & Hardage.[32] And if the gni-DNN is a valid reflection of human visual processing, it should be expected to exhibit a similar degree of polarity blindness as human visual processing. It is interesting, therefore, to find that the gni-DNN discrimination performance is virtually identical as a function of gap size for opposite-as same-polarity symmetry patterns at both low and high levels in the DNN hierarchy (Figure 9).

DISCUSSION

The goal of this study was to assess the performance of a canonical 8-level deep neural network to process the formal property of reflection symmetry in object-neutral symmetry images in the absence of any object-specific information in the set of test images.

Processing levels

The performance of the generalized natural image Deep Neural Network (gni-DNN) with object-neutral, pure symmetry images showed that it was able to pick up information about the symmetry property *per se*, in the absence of any object-specific information in the set of test images. In one sense this is surprising since few random images of the natural environment would have image symmetry about the central axis. On the other hand, it turned out that many of the objects in the training set had inherent symmetries, such as the bilateral symmetry of most organisms, or the artificial symmetry of constructed objects and the symmetric viewpoints chosen to depict them (see Figure 3A).

In the 8-level DNN employed in this study, the first two main layers (Pool1, Pool2) showed no significant symmetry response (Figure 3). This is an interesting result that matches the symmetry response of the human brain determined by fMRI studies, which shows no differential symmetry/random response in the first three layers of the human visual hierarchy (V1-3,[19,36,37] even though they give the strong visual impression that there is a special appearance along the axis of symmetry, which evokes comparison with a totem pole of stacked symmetric faces and creatures. In fact, the main human symmetry response in those studies occurs beyond the primary visual hierarchy in the lateral occipital complex, a mid-level visual integration area that encodes many elaborated object properties. There is very little symmetry response in the rest of the brain if no explicit task is assigned to the participants.

Beyond the first two layers, the gni-DNN response builds up from layer Pool3 to a sudden increase at layer FC6, then levels off for layers beyond that (Figure 4). The build-up is reminiscent of the corresponding level in the human cortex where symmetry processing becomes manifest through increased fMRI activation. However, the gni-DNN performance differs in that the activation continues unabated to subsequent levels beyond FC6, but symmetry activation is not evident at higher levels in the human brain. The inference is that the gni-DNN continues to operate in mid-level vision mode as far as this version is implemented, while the human brain continues the abstraction process to higher-level object recognition and scene processing, for which the symmetry is no longer relevant once the objects have been recognized.

N-gram structure

A unique aspect of this study is that it identifies the n-gram structure implicit in the symmetry operation. Although the only operation used to generate a symmetric image from the purely random base arrays was to repeat the half-image after a symmetry-reflection flip, the result includes local structure at the n-gram level as explained in the introduction, which allows the symmetry to be picked up from an increased response variance from the local repeated elements in the image as a whole.

A significant aspect of the block-randomization discrimination procedure is that it also removes the consistent symmetry information extending along the horizontal and vertical symmetry axes at the scale of the blocks in the full-size images. Thus, to the extent that symmetry detection is mediated by such local

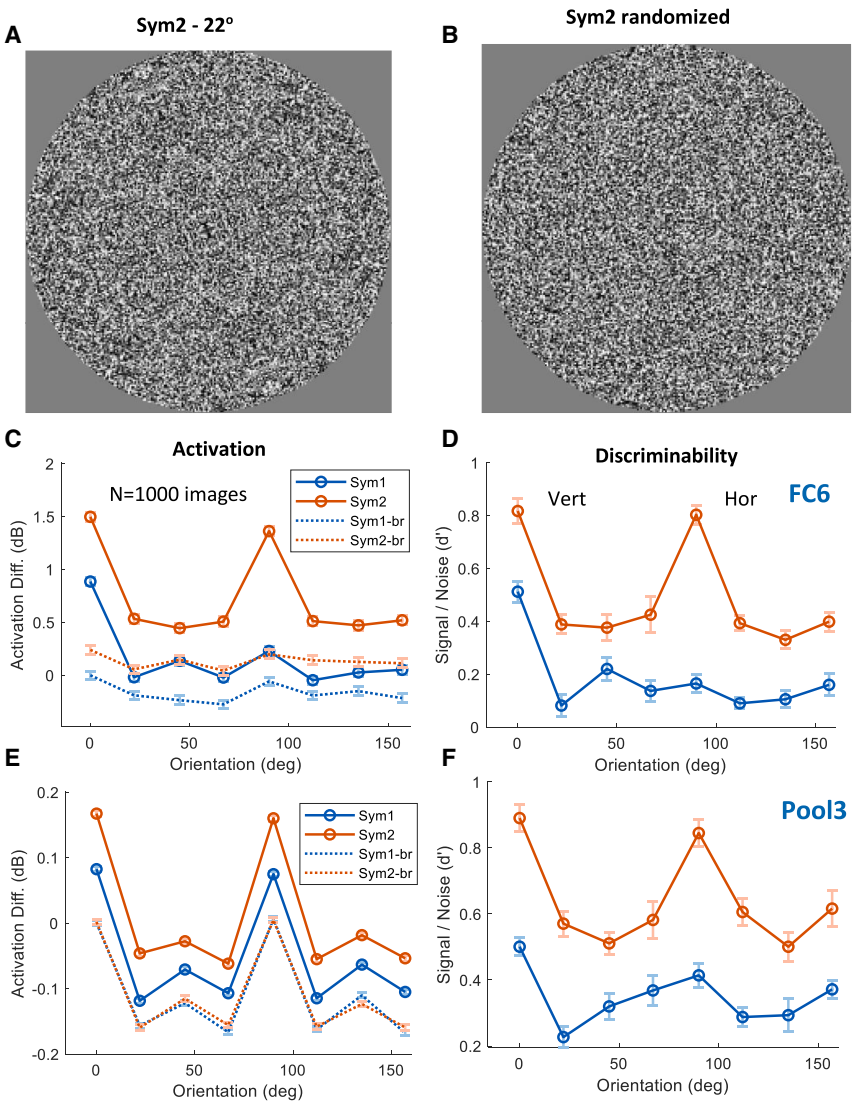


Figure 7. The effect of symmetry axis orientation on the symmetry effect

The random dot stimuli were rotated in 8 orientations in a circular window, with 1 pixel per texel to avoid rotation artifacts. Error bars denote ±1 SEM.

(A) Example of stimulus image of 2-axis symmetry (Sym2) rotated by 22.5°; (B) a similar Sym2 image, block-randomized (16x16 pixel blocks) before rotation, used for computing a baseline.

(C and E) Activation differences between symmetry and fully randomized reference images for 1- and 2-axis symmetry as a function of symmetry axis orientation, expressed in dB relative to the block-randomized condition of Sym1, for layers FC6 in (C) and Pool3 in (E), plotted in solid lines, with the corresponding baseline measures (block-randomized control stimuli) plotted in dotted lines.

(D and F) The corresponding discriminability measures are computed as the difference between the symmetry effect and the control effect in SD units (signal to noise ratio as d-prime), for layer FC6 in (D) and Pool3 in (F). As shown, the strongest symmetry effect was found in both cardinal orientations of 2-axis symmetry, while the 1-axis symmetry conditions showed that it was much stronger for the vertical for in layer FC6, but less so in layer Pool3.

paraxial content, it should be eliminated under this manipulation, as is the case for block-randomized control discrimination for single-axis symmetry. However, some local symmetry information of this type will remain from the oblique symmetry axes in the 4-axis symmetry case. The data of Figure 4 show that a low level of symmetry detection performance remains for 2- and 4-axis detection for Pool5 and beyond. Since the symmetry information is eliminated by block-randomization in the 1-axis case, this multi-axis performance is attributable to the residual variance cue of higher overall variance in the symmetry than the fully random patterns.

Mid-vs. high-level processing

The remainder of the study focuses on differences in symmetry processing between the mid-level layers (represented by Pool3) and the high-level layers (represented by the FC6 layer) of the gni-DNN. The response to the symmetry vs. random patterns is roughly proportional to the number of symmetry axes in the image (1, 2, or 4). In terms of discriminability (Figures 4B and 4D), consider first the 1-axis case, which was tested in two orientations. Although the mid-level response (Pool3) was the same for the two orientations, the high-level response was significantly stronger for the vertical than the horizontal orientation, implying that the vertical axis preference is a function of the longer-range processing available at FC6, whereas the processing is predominantly local at Pool3.

The increased number of axes provides enhanced discriminability in both Pool3 and FC6, but it is notable that there is no significant improvement for the 4-axis over the 2-axis discriminability. This saturation effect is attributable to the significant increase in block-randomized discriminability for the 4-axis over the 2-axis patterns, implying that the 4-axis advantage in basic symmetry response is attributable to the local symmetry information that is not canceled by the block-randomization control procedure.

However, it should be recognized that these d-prime levels are not sufficient for the gni-DNN to provide an accurate symmetry/random decision for each individual image, as could the human visual system. This level of performance is more similar to the human performance found for short (e.g., 50 msec) exposures by[33] and,[31] while these targets would be perfectly discriminable for the entire stimulus set with longer (e.g., >100 ms) exposures. In this sense, the gni-DNN is operating equivalently to the primary response of the human visual hierarchy, not to the full processing of image structure the human brain.

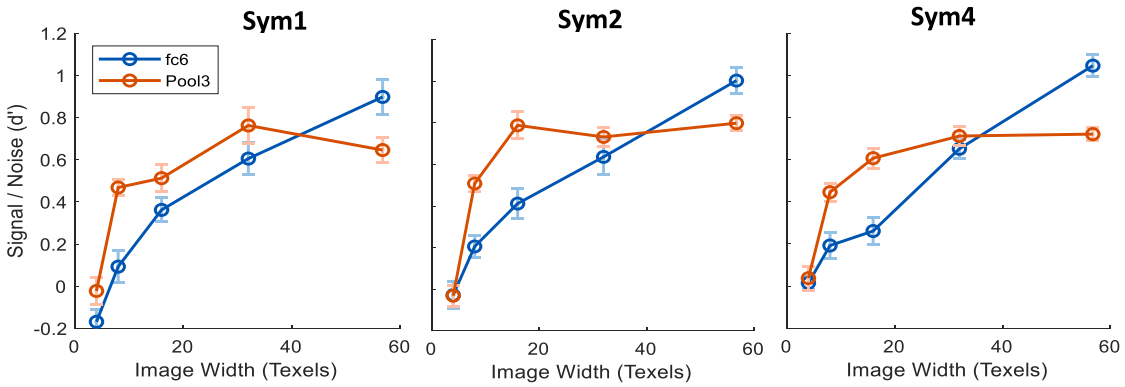


Figure 8. The effect of image size on the symmetry effect

Image size was varied from 4 to 56 texels of image width. The discriminability (signal-to-noise, as d-prime) is plotted for 1-, 2-, and 4-axis symmetry relative to corresponding block-randomized control images for each symmetry type for two layers: Pool3 (red symbols) and FC6 (blue symbols). Error bars denote ±1 SEM across 10 bins of 150 images (N = 1500 images per condition). As shown, the symmetry effect increased with size, but for the early layer (Pool3) it saturated early, while for the higher layer (FC6) the integration spanned the whole range, starting slowly but ultimately reaching a higher level of discriminability.

Orientation specificity

Like humans, the gni-DNN shows a substantial performance boost for symmetry on the cardinal meridians (horizontal and vertical axes; Figures 4 and 5). This anisotropy could only derive from the natural-image training of the network since the symmetry patterns were simply rotated versions of each other. The weaker performance in the oblique orientations is consistent with the oblique effect in human vision, as has been established both for contrast sensitivity[38,39] and for symmetry detection.[35,40] The latter authors found that both cardinal orientations were more detectable than other orientations in long (2 s) presentations, but that the vertical axis symmetry predominated for brief (150 ms) presentations.

Range of symmetry discriminability

The final issue to be examined was the degree to which the symmetry response was derived from paraxial vs. long-range information, as developed in the introduction. This question was addressed by restricting the symmetry and control patterns to narrow strips of 1-axis and 2-axis texels and moving them apart so that they were separated by gaps of various sizes, without changing the amount of pattern information in the image as a whole. Under this manipulation, the symmetry responses dropped precipitously as soon as a small gap was introduced between the strips, implying that the gni-DNN was drawing the majority of the symmetry information locally from around the axis (see Figure 6). This suggests the interpretation that the gni-DNN was deriving the symmetry estimate from something such as even-vs. odd-symmetric receptive fields lying along the axis, with the predicted proportion of activation response rising from 0.5 even-symmetric activation for random patterns to 1.0 for fully symmetric patterns. The fact that the behavior was similar for inverse-polarity symmetry could be attributed to the even/odd ratio deviating in the other direction, from 0.5 for random patterns to a 0.0 even-symmetric proportion on the axis for fully inverse-symmetric patterns.

Beyond the 4-texel range, a residual symmetry-specific response remained in the FC6 layer at about the same level all the way out to gaps of 32 texels. This behavior implies the operation of a long-range process similar to that identified by Barlow & Reeves[32] for humans, which not only extended across the entire visual field as far as could be tested but was only marginally degraded for the long-range detection of inverse-polarity symmetry. The latter property of undiminished long-range discriminability, for both the human and the gni-DNN performance, implies that the image symmetry is processed by a second-order, contrast-polarity-insensitive processing mechanism. Thus, although humans are perceptually aware of whether the dots and the local patterns they form are black or white, the symmetry structure (and presumably other form descriptors) is processed as a sign-insensitive property to both humans and machines.

Conclusion

One of the key properties of the world around us is the mirror symmetry of many kinds of objects found in it. Notably, symmetry is an abstract visual property with no definable template matching involved. To assess whether DNNs trained on a general natural-image database (gni) typical of the cultural environment experienced by human brains would encode the symmetry of image structure, we probed the symmetry-specific performance of standard feedforward gni-DNN with pure (random-dot) symmetry images that would not typically be encountered in that environment. The results reveal that the gni-DNN discrimination performance was similar to that of humans in many respects, including the effects of symmetry axis orientation, number of symmetry axes, and texture separation around the symmetry axis, even in extreme cases of widely separated strips with symmetry present only in matching or polarity-inverted form. In particular, the results showed that there was no significant coding for abstract symmetry in the earliest layers of the gni-DNN, and its strongest discriminability was found in the first fully connected layer, FC6, plausibly analogous to the human Lateral Occipital Complex (LOC), where symmetry coding is first identified in human fMRI studies.[19] In more detail, this area can be distinguished from more elaborated processing of the

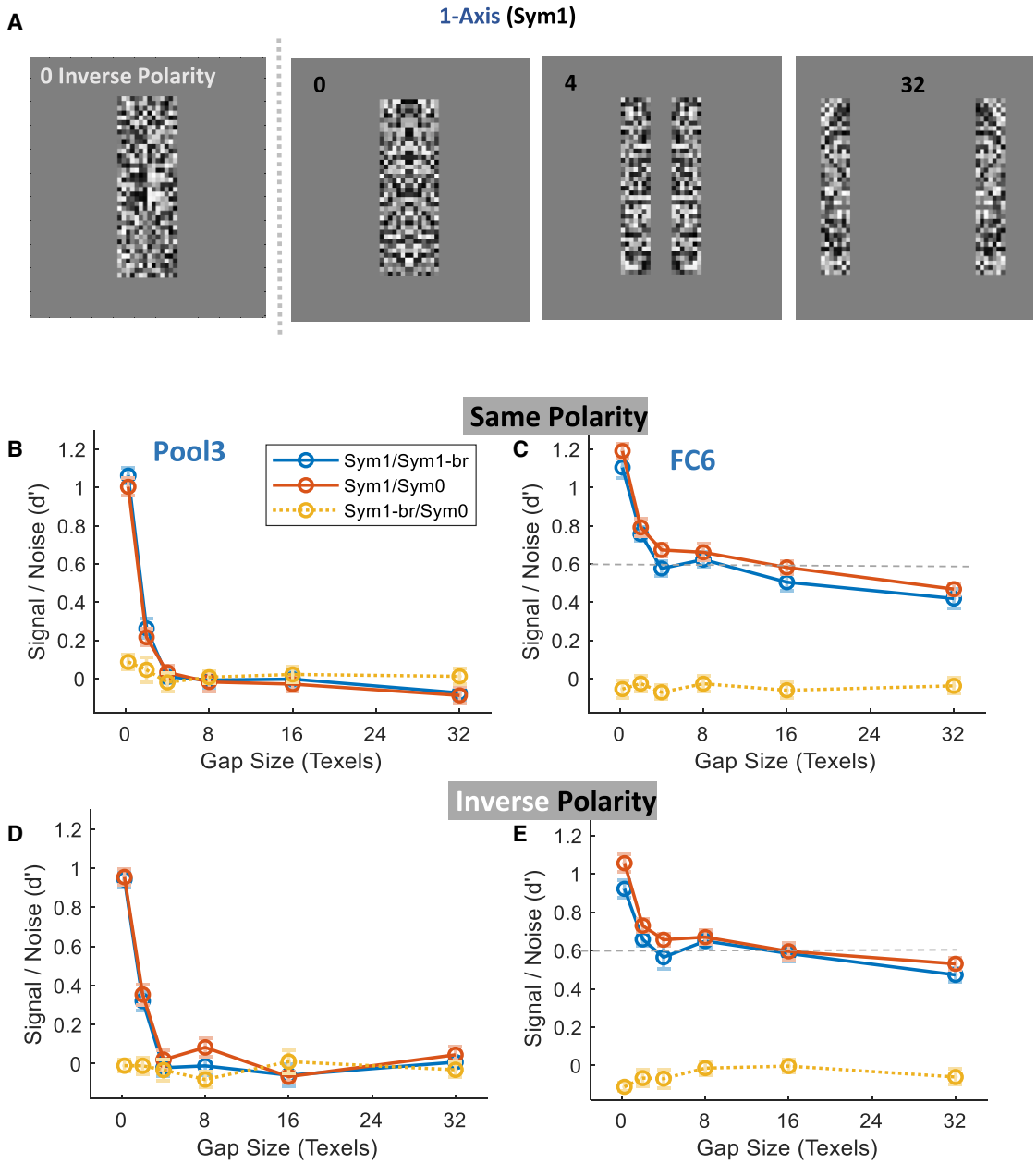


Figure 9. Effect of vertical gaps on the symmetry effect

Gaps of 0, 2, 4, 8, 16, and 32 texels (of 4x4-pixels) were tested in vertical 1-axis symmetry (Sym1), both same polarity and inverse polarity, to explore non-local effects.

(A) Examples of stimuli used, with 0-, 4-, and 32-texel gaps of the same polarity (3 right panels) and no gap inverse polarity (left panel).

(B–D) Discriminability (signal-to-noise, as d-prime) is plotted for Sym1 relative to the corresponding block-randomized control images (Sym1-br) as well as relative to fully random images (Sym0), with Sym1-br relative to Sym0 for comparison of the local vs. long-range selectivity; (A and B) same-polarity and (D and E) inverse-polarity symmetry; (B and D) for Pool3 and (C and E) for the FC6 layers. Error bars denote ±1 SEM across 10 bins of 200 images (N = 2000 images per condition). As shown, the symmetry effect decreased rapidly with the gap size, reached zero for the 4-texel gap in Pool3, but maintained an almost constant level of discriminability around d' = 0.6 in FC6 (dashed lines in c). The discriminability of Sym1-br control stimuli relative to Sym0 was close to zero for all gap sizes.

posterior fusiform gyrus, which is specialized for face processing and other complex object stimuli, but is insensitive to image symmetry per se.[41] The same study identifies a cortical specialization for image symmetry as being co-located with the occipital face area (OFA), a subregion of the LOC, which would therefore be the portion of LOC homologous with layer FC6 of the gni-DNN. The posterior fusiform gyrus (PFG), on the other hand, would be excluded by virtue of its insensitivity to image

symmetry per se. (In,[41] the PFG area was sensitive to object inversion with the same symmetry but insensitive to changes from symmetric to asymmetric 3D poses of the same object, while the OFA responded in the opposite manner.) In this way, the present study expands the range of homology between the feedforward gni-DNN trained on natural images and the early visual hierarchy of the human brain with respect to an important aspect of visual object processing, the abstract processing of object-independent symmetry structure.

Limitations of the study

This study shares the general limitation that there is no absolute proof of the underlying mechanism for any neuroscientific performance result, be it neurophysiological, biochemical, psychophysical, or through mathematical modeling. The inferred mechanism is always based on inferences from empirical evidence relative to control conditions that are designed to eliminate all factors other than the one that is the focus of a particular study. However, it is difficult to define the perfect control that truly eliminates all possible factors for any particular stimulus, as has become evident in functional imaging studies, in particular. In the present work, we have come closer than any other study of symmetry processing with the block-randomized controls. However, just as the brain is a black box, the gni-DNN is a black box whose processing has to be inferred from the distinction between the test and control stimuli, which is designed to equate the responses of all known types of local cortical receptive fields (other than the explicitly second-order symmetry neuronal response coding reported[29]). The discriminability for long-range symmetry structure relative to these block-randomized controls in Figure 9 is therefore the strongest evidence that we can envisage for the operation of long-range symmetry processing by the gni-DNN.

Given the relative and implicit measure of symmetry we obtain, it should be noted that we do not claim that the gni-DNN can "detect" symmetry but that it "has a selectivity for symmetry similar to that of the human brain," and that it "encode(s) the symmetry of image structure." These claims are based on differences in the main activation patterns in the fully connected layers, which are not seen in the earlier pooling layers. Nevertheless, the model's selectivity for random-dot symmetry could also be obtained from a single image by computing the L2 difference between the activation patterns for that image and a block-randomized version of that image (here "activation diff"). If statistically significant, this "activation diff" would then provide a measure of the selectivity of the gni-DNN for the symmetry of this single image.

Another limitation of the study is that we did not explore measures other than the L2 for the differences in activation patterns evoked by symmetric and random images, such as simple activity distribution properties or the existence of "symmetry sensitive units." For example, one of our reviewers suggested that in the upper layer (FC8), units associated with the image *category* of symmetric objects (rather than the structural symmetry per se) might be more activated in response to random-dot symmetry compared to fully random. However, we consider this an overinterpretive projection of the human designation of the category membership onto the inner workings of the DNN, which would have to develop its category discrimination based on local bottom-up information. This evaluation is consistent with the current results because the present measures do not just separate random from non-random but reveal similarities to various structural characteristics of human symmetry perception and how they build up across the DNN layers.

As with most DNN studies, the architecture of the gni-DNN assessed in the present work had homogeneous spatial sampling across the image space, whereas a more realistic model of human visual processing would incorporate the cortical scaling factor of a radial gradient of progressively coarser spatial resolution in peripheral vision (as reflected in the stimuli employed for Figure 1A). On the other hand, this polar transformation that human visual processing incorporates from retina to cortex implies that the human cortical processing is effectively Cartesian thereafter, making it analogous to the DNN with Cartesian input to its initial layer. Thus, in specific reference to Figure 1A, it is the case that the psychophysical performance is uniform with eccentricity when the stimuli are scaled in proportion to the distance from fixation, whereas performance falls with eccentricity when the stimuli are uniform in a manner similar to the initial parts of the discrimination functions of Figure 9.

However, in contrast to the retino-cortical scaling, human perception is strongly Cartesian across the primary perceptual field of conscious awareness (such that a physically regular grid appears to have regular spacing, not the profound distortion of its neural projection onto the primary visual cortex). This visual appearance implies that the accessible perceptual representation retains the homogeneous spatial relationships despite the polar sampling of the retinal input.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yoram Bonneh (yoram.bonneh@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The code to reproduce the results in this study as well as pre-computed data are available at OSF (https://doi.org/10.17605/OSF.IO/8F7YT).
- Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

The authors would like to thank Ron Dekel for his initial inspiration and for suggesting the research approach.

AUTHOR CONTRIBUTIONS

Conceptualization, YSB and CWT; analysis and code: YSB; writing: CWT; review and editing: CWT and YSB.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Deep neural network model
- METHOD DETAILS
  - Stimuli
- QUANTIFICATION AND STATISTICAL ANALYSIS

Appendix

Although the above contextual interactions are sometimes classified as mid-level vision, here we adopt a more discriminative definition of mid-level vision as referring to properties that are represented by long-range abstract relationships rather than being progressive concatenations of low-level features such as the ubiquitous Gabor filters. These long-range relationships are the Gestalt principles, which have never been mathematically formalized to our knowledge, although initial attempts have been made by[42] and.[43] Thus, we treat mid-level (spatial) vision as consisting of relationships such as

$$f(r') = T[f(r)]$$

where r is a region of space, r' is a displaced replication of that spatial region, and T[ ] is some transformation operator.

Thus, for example, this relationship would form a plane-perspective operation if T[ ] was a scale change, or could represent a closure principle if r' defined line extension for a curved line element and was a complete set. In the present study, the transformation to be considered is the mirror symmetry of a 2D pattern at location r' relative to that at location r, such that $r(\phi, \theta) = r(\phi, -\theta)$, where ϕ, θ are frontoparallel Cartesian axes. The same principle can be generalized to the specification of relative proximity (when applied to the spacing between elements) and common motion (when the long-range transformation is on a spatiotemporal variable). Further long-range Gestalt relationships that are less amenable to such formalization are those of 3D structural inference, amodal completion, and figure/ground relationships.

Received: June 21, 2022
Revised: December 13, 2023
Accepted: December 3, 2024
Published: December 5, 2024

REFERENCES

1. Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365. https://doi.org/10.1038/nn.4244.
2. Kubilius, J., Bracci, S., and Op de Beeck, H.P. (2016). Deep Neural Networks as a Computational Model for Human Shape Sensitivity. PLoS Comput. Biol. *12*, e1004896. https://doi.org/10.1371/journal.pcbi.1004896.
3. Jozwik, K.M., Kriegeskorte, N., Storrs, K.R., and Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. Front. Psychol. *8*, 1726. https://doi.org/10.3389/fpsyg.2017.01726.
4. Groen, I.I., Greene, M.R., Baldassano, C., Fei-Fei, L., Beck, D.M., and Baker, C.I. (2018). Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. Elife *7*, e32962. https://doi.org/10.7554/eLife.32962.
5. Volokitin, A., Roig, G., and Poggio, T. (2017). Do deep neural networks suffer from crowding? In Advances in Neural Information Processing Systems, pp. 5629–5639. https://doi.org/10.1167/18.10.902.
6. Grill-Spector, K., Weiner, K.S., Gomez, J., Stigliani, A., and Natu, V.S. (2018). The functional neuroanatomy of face perception: From brain measurements to deep neural networks. Interface Focus *8*, 20180013. https://doi.org/10.1098/rsfs.2018.0013.
7. Grossman, S., Gaziv, G., Yeagle, E.M., Harel, M., Mégevand, P., Groppe, D.M., Khuvis, S., Herrero, J.L., Irani, M., Mehta, A.D., and Malach, R. (2019). Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. Nat. Commun. *10*, 4934. https://doi.org/10.1038/s41467-019-12623-6.
8. Gupta, S.K., Zhang, M., Wu, C.-C., Wolfe, J.M., and Kreiman, G. (2021). Visual Search Asymmetry: Deep Nets and Humans Share Similar Inherent Biases. In 35th Conference on Neural Information Processing Systems (NeurIPS 2021), pp. 1–14.
9. Sun, E.D., and Dekel, R. (2021). ImageNet-trained deep neural networks exhibit illusion-like response to the Scintillating grid. J. Vis. *21*, 15. https://doi.org/10.1167/jov.21.11.15.
10. Gruber, L.Z., Haruvi, A., Basri, R., and Irani, M. (2018). Perceptual dominance in brief presentations of mixed images: Human perception vs. deep neural networks. Front. Comput. Neurosci. *12*, 57. https://doi.org/10.3389/fncom.2018.00057.
11. Geirhos, R., Schütt, H.H., Medina Temme, C.R., Bethge, M., Rauber, J., and Wichmann, F.A. (2018). Generalisation in humans and deep neural networks. In Advances in Neural Information Processing Systems, pp. 7538–7550.
12. Dekel, R. (2017). Human perception in computer vision. Preprint at arXiv. https://doi.org/10.48550/arXiv.1701.04674.
13. Bonneh, Y., and Dekel, R. (2020). Deep Neural Networks as a computational model for early vision: Lateral masking and contour integration. J. Vis. *20*, 1356. https://doi.org/10.1167/jov.20.11.1356.
14. Golan, T., Raju, P.C., and Kriegeskorte, N. (2020). Controversial stimuli: Pitting neural networks against each other as models of human cognition. Proc. Natl. Acad. Sci. USA *117*, 29330–29337. https://doi.org/10.1073/PNAS.1912334117.
15. Jiahui, G., Feilong, M., di Oleggio Castello, M.V., Nastase, S.A., Haxby, J.V., and Gobbini, M.I. (2023). Modeling naturalistic face processing in humans with deep convolutional neural networks. Proc. Natl. Acad. Sci. USA *120*, e2304085120. https://doi.org/10.1073/pnas.2304085120.
16. Nonaka, S., Majima, K., Aoki, S.C., and Kamitani, Y. (2021). Brain hierarchy score: Which deep neural networks are hierarchically brain-like? iScience *24*, 103013. https://doi.org/10.1016/j.isci.2021.103013.
17. Pramod, R.T., and Arun, S.P. (2016). Do computational models differ systematically from human object perception? In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-Decem, pp. 1601–1609. https://doi.org/10.1109/CVPR.2016.177.
18. Brachmann, A., and Redies, C. (2016). Using convolutional neural network filters to measure left-right mirror symmetry in images. Symmetry *8*, 144. https://doi.org/10.3390/sym8120144.
19. Tyler, C.W., Baseler, H.A., Kontsevich, L.L., Likova, L.T., Wade, A.R., and Wandell, B.A. (2005). Predominantly extra-retinotopic cortical response to pattern symmetry. Neuroimage *24*, 306–314. https://doi.org/10.1016/j.neuroimage.2004.09.018.
20. Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607–609. https://doi.org/10.1038/381607a0.
21. Adelson, E.H., and Bergen, J.R. (1991). The Plenoptic Function and the Elements of Early Vision. In Computational Models of Visual Processing, M. Landy and J.A. Movshon, eds., pp. 3–20. https://doi.org/10.7551/mitpress/2002.003.0004.
22. Tyler, C.W. (1997). Symmetries, structure and schemata in perceptual coding. In Computational and Biological Mechanisms of Visual Coding, M. Jenkin and L. Harris, eds. (Cambridge University Press).
23. Hubel, D.H., and Livingstone, M.S. (1985). Complex-unoriented cells in a subregion of primate area 18. Nature *315*, 325–327. https://doi.org/10.1038/315325a0.

24. Tyler, C.W. (1983). Sensory processing of binocular disparity. In Basic and Clinical Aspects of Binocular Vergence Eye Movements, C. Schorm and K.J. Ciuffreda, eds. (Butterworths), pp. 199–295.
25. Craft, E., Schütze, H., Niebur, E., and Von Der Heydt, R. (2007). A neural model of figure-ground organization. J. Neurophysiol. *97*, 4310–4326. https://doi.org/10.1152/jn.00203.2007.
26. Likova, L.T., and Tyler, C.W. (2008). Occipital network for figure/ground organization. Exp. Brain Res. *189*, 257–267. https://doi.org/10.1007/s00221-008-1417-6.
27. Tyler, C.W. (1973). Stereoscopic vision: Cortical limitations and a disparity scaling effect. Science *181*, 276–278. https://doi.org/10.1126/science.181.4096.276.
28. Tyler, C.W. (1975). Spatial of Binocular Disparity Sensitivity. Vis. Res. *15*, 583–590.
29. Lee, T.S., Mumford, D., Romero, R., and Lamme, V.A. (1998). The role of the primary visual cortex in higher level vision. Vis. Res. *38*, 2429–2454. https://doi.org/10.1016/S0042-6989(97)00464-1.
30. Tyler, C.W., Hardage, L., and Miller, R.T. (1995). Multiple mechanisms for the detection of mirror symmetry. Spatial Vis. *9*, 79–100.
31. Tyler, C.W. (1999). Human symmetry detection exhibits reverse eccentricity scaling. Vis. Neurosci. *16*, 919–922. https://doi.org/10.1017/S0952523899165118.
32. Tyler, C.W., and Hardage, L. (1996). Mirror symmetry detection: Predominance of second-order pattern processing throughout the visual field. In Human Symmetry Perception and Its Computational Analysis, pp. 157–171.
33. Barlow, H.B., and Reeves, B.C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. Vis. Res. *19*, 783–793. https://doi.org/10.1016/0042-6989(79)90154-8.
34. Fei-Fei, L., Deng, J., and Li, K. (2010). ImageNet: Constructing a large-scale image database. J. Vis. *9*, 1037. https://doi.org/10.1167/9.8.1037.
35. Corballis, M.C., and Roldan, C.E. (1975). Detection of symmetry as a function of angular orientation. J. Exp. Psychol. Hum. Percept. Perform. *1*, 221–230. https://doi.org/10.1037//0096-1523.1.3.221.
36. Sasaki, Y., Vanduffel, W., Knutsen, T., Tyler, C., and Tootell, R. (2005). Symmetry activates extrastriate visual cortex in human and nonhuman primates. Proc. Natl. Acad. Sci. USA *102*, 3159–3163. https://doi.org/10.1073/pnas.0500319102.
37. Van Meel, C., Baeck, A., Gillebert, C.R., Wagemans, J., and Op de Beeck, H.P. (2019). The representation of symmetry in multi-voxel response patterns and functional connectivity throughout the ventral visual stream. Neuroimage *191*, 216–224. https://doi.org/10.1016/j.neuroimage.2019.02.030.
38. Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. Psychol. Bull. *78*, 266–278. https://doi.org/10.1037/h0033117.
39. Mitchell, D.E., Freeman, R.D., and Westheimer, G. (1967). Effect of orientation on the modulation sensitivity for interference fringes on the retina. J. Opt. Soc. Am. *57*, 246–249. https://doi.org/10.1364/JOSA.57.000246.
40. Wenderoth, P., and Welsh, S. (1998). Effects of pattern orientation and number of symmetry axes on the detection of mirror symmetry in dot and solid patterns. Perception *27*, 965–976. https://doi.org/10.1068/p270965.
41. Chen, C.C., Kao, K.L.C., and Tyler, C.W. (2007). Face configuration processing in the human brain: The role of symmetry. Cereb. Cortex *17*, 1423–1432. https://doi.org/10.1093/cercor/bhl054.
42. Tyler, C.W. (2004). Beyond fourth-order texture discrimination: Generation of extreme-order and statistically-balanced textures. Vis. Res. *44*, 2187–2199. https://doi.org/10.1016/j.visres.2004.03.032.
43. Michaelsen, E. (2014). Gestalt algebra-a proposal for the formalization of gestalt perception and rendering. Symmetry *6*, 566–577. https://doi.org/10.3390/sym6030566.
44. Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–14.
45. Maunsell, J.H., and Van Essen, D.C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. J. Neurophysiol. *49*, 1148–1167. https://doi.org/10.1152/jn.1983.49.5.1148.
46. Gross, C.G. (2008). Single neuron studies of inferior temporal cortex. Neuropsychologia *46*, 841–852. https://doi.org/10.1016/j.neuropsychologia.2007.11.009.
47. Julesz, B., and Bergen, J.R. (1983). Human Factors and Behavioral Science: Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures. Bell Syst. Tech. J. *62*, 1619–1645. https://doi.org/10.1002/j.1538-7305.1983.tb03502.x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Code (MATLAB) and pre-computed data	This paper	https://doi.org/10.17605/OSF.IO/8F7YT
Software and algorithms		
MATLAB R2022a	Mathworks	www.mathworks.com

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Deep neural network model

As a DNN model for pure symmetry, we used a standard VGG-16 Convolutional Neural Network architecture[44] that has been pre-trained on approximately 1.2 million images from the ImageNet Dataset[34] to classify images into 1000 object categories. A sampling of these images is depicted in Figure 3A, illustrating that they are not only images from nature, but predominantly images from the human cultural sphere, in which a substantial proportion have strong reflection symmetries on multiple axes. It is also noteworthy that these images were often photographed from a symmetric viewpoint, so that they contain strong image symmetry (which would seldom be the case if photographed from a nonselective viewpoint). There is thus a substantial degree of image symmetry in the image training set.

The model is depicted in Figure 3B, where the zero-symmetry reference and signal symmetry images are shown at left, and the various processing layers of the hierarchical DNN are depicted by the stacks of square plates at right. A stack of plates of the same size represents the convolution layers of one level in the hierarchy with different parameters for the input processing at that level, such as different orientations, sizes, and so forth of the local processing units at that level. Each level represents a smaller array of processing units encoding a more abstracted representation than the previous level. There are 8 levels pooling from each layer to the next, with 5 layers of max-pooling (in red, labeled as Pooling 1–5) and 3 levels of fully-connected layers (labeled as Fully Connected 6–8). In the analyses we probe the 5 pooling layers Pool1 to Pool5 and the last 3 fully connected layers FC6 to FC8. In addition, we also probe the input layer for validation of the analyses.

In the context of this DNN architecture, we can make the comparison with the neuroanatomy of the biological image processing in the human brain. We note that this organization consists of a hierarchy of processing layers involving progressive convergence and specialization of the information processing from one layer to the next, in the form of the receptive field organization of each layer. In overview, this convergence and specialization operates for 5 parallel processing layers (retina, lateral geniculate nucleus, V1, V2, V3) before breaking into an array of more specialized aggregation centers for color, motion, depth structure, and so on (e.g., Maunsell & van Essen[45]), many of which are located in the association cortex of Brodmann area 19, now often known as the lateral occipital cortex (LOC), which has much in common with the region designated as V4 in the macaque monkey, although the older literature[46] would suggest that this full convergence is not achieved until the next neuroanatomical layer, which is the inferotemporal cortex (IT) in macaque. Thus, it is roughly at the level of the 5^{th} or 6^{th} processing layer that the primate visual cortex can be regarded as fully connected, making LOC or IT the plausible neuroanatomical analog of the first fully-connected layer of the visual processing hierarchy of Figure 3.

METHOD DETAILS

To assess pure symmetry independent of recognizable objects, we used the standard ImageNet-trained VGG16 network model[44] to compute the average L2 distance from zero symmetry, for sets of random-dot symmetry images with one, two, and four axes of symmetry. To identify the fact that this DNN was trained on a general database of natural images, we term it a gni-DNN. The various stimulus conditions are described in Figure 2 and below, while the model used for the analyses is depicted in Figure 3 and described in detail below (see Model).

Stimuli

Example stimuli are shown in Figure 2, with detailed information on stimulus parameters appearing in the figure caption. The stimuli consist of purely random spatial noise with a uniform intensity distribution (Sym0, Figure 2C), vertical reflection symmetry around a central axis for the same kind of noise (Sym1 Vertical), and two and four axis reflections (Sym2 and Sym4), all shown in Figure 2A.

Thus, in MATLAB terminology, these symmetries are given by the following, where R is a random $n * n$ matrix of decimal numbers in [0 1], and k = texel size (and where $k = 4$ in all experiments except one):

$$Sym1V = [R, R(2 * k * n : -1 : 1, :)];$$

$$Sym1H = [R, R(2 * k * n : -1 : 1, :)]';$$

$$Sym2 = [R, R(:, k * n : -1 : 1); R(k * n : -1 : 1, :), R(k * n : -1 : 1, k * n : -1 : 1)];$$

$$Sym4 = [R1, R1(:, k * n : -1 : 1); R1(k * n : -1 : 1, :), R1(k * n : -1 : 1, k * n : -1 : 1)];$$

where $R1 = \mathrm{kron}(R, \mathrm{ones}(k))$; and R is symmetric around the diagonal, $R(i,j) = R(j,i)$ for all i,j.

It is noticeable in Figure 2A that the symmetry structure becomes progressively more salient as the number of axes increases. The gni-DNN processing for discriminating the symmetric stimuli was assessed with respect to that for the null stimulus of Sym0.

As described in the introduction, two kinds of symmetry information can be distinguished in these stimuli. One is the local structure around the axis of reflection, which is constrained to have even symmetry (cosine phases of its Fourier component representation) at the scale of the local cortical receptive fields. The other is the long-range symmetry relations, the fact that the patterns in local patches away from the symmetry axis should match after a reflection operation. Although this long-range information is formally equivalent to the local structure constraint in terms of a complete Fourier description, the long-range symmetry relations are inaccessible to the kinds of local (complex Gabor) processing units known to exist as receptive fields in the human cortex, and thus presumably also in a DNN trained on the natural image array (as the brain has been). This distinction is an empirical question of the kind that the present study is designed to address. In order to differentiate between the discriminability of the local and long-range aspects of the symmetry information by the gni-DNN, we introduced strips of random dots extending various distances around the symmetry axes of the symmetry stimuli in order to eliminate the local axis information and restrict the processing to the long-range information alone.

Control stimuli

One issue with such stimuli is that, in addition to the symmetry structure of matching local elements around and equidistant from the symmetry axes, the symmetry stimuli have pattern redundancies not present in the random pattern (Sym0). For example, the Sym4 pattern consists of 8 identical triangular patches in an alternating symmetrical arrangement. It would thus be possible to discriminate the symmetrical patterns from the random one (and from each other) purely on the basis of the elemental histogram statistics, without recourse to any higher-order (n-gram) spatial relationships (see Julesz & Bergen[47]).

To avoid this potential confound, the basic comparisons were rerun with patterns of the form shown in Figure 2B, where block-randomized control patterns (BR) are made by breaking the symmetry image into texel blocks of 4x4 pixels and re-arranging them in a random permutation order, which gives similar perceptual impressions to the purely random ones, but eliminates the differences in histogram between the symmetric and null patterns for each symmetry type. The random permutation of image blocks was made by the function randblock() (Jos van der Geest, https://www.mathworks.com/matlabcentral/fileexchange/17981-randblock, MATLAB Central File Exchange).

This novel manipulation, combined with the randomization along the symmetry axes, is a stringent test of the ability of the gni-DNN to extract pure long-range symmetry information, such as (a) its presence along extended axes of symmetry, (b) the mirror reflection of salient patterns at long distances from the axes, and (c) the tendency to form recognizable shapes with multiple numbers of symmetry axes. All these types of extended symmetry structure are eliminated by the block-randomization procedure, which confines the residual information to local self-similar blocks in random locations. Although strongly degraded, this local n-gram self-similarity is thus still present in the block-randomized images (though completely absent in the fully random images). The block-randomized controls were then used as a rigorous base for the discriminability of the long-range information in the full symmetry relative to the scrambled symmetry images.

It is therefore an empirical question whether this kind of information is accessible to the mid-level processing at the intermediate stages of the gni-DNN, which in turn are likely to mimic aspects of the mid-level processing of human vision, or of biological vision in general. This is an interesting issue because mid-level biological vision is not necessarily accessible to conscious perception and will not necessarily be detectable in psychophysical studies. Such kinds of higher-order regularities have, however, been shown by Tyler[42] to be perceptually accessible, though not specifically in this novel block-randomized symmetry paradigm. These mid-level vision attributes, if detectable in the gni-DNN output, could then be a target for direct neural response investigations. (Note that the randomization of the blocks will preserve the local mirror-reflection relationships within each block, though strongly disguised by the random global reordering, whereas different properties were preserved to higher order in the previous papers. Control studies were therefore run to determine the gni-DNN discriminability of the block-randomized stimuli relative to the full random ones.).

QUANTIFICATION AND STATISTICAL ANALYSIS

The basic computation for the DNN symmetry effect for a given layer (see Figure 3B) was as follows: (1) Compute the activation difference for pairs of images, a ''signal'' of one symmetry image and a ''reference'' of a fully randomized image, by (a) computing the layer activation map for each image used as the DNN input; and (b) computing the L2 (or Euclidean) distance between the response maps of the two images in a pair, resulting in one positive scalar value per image pair for that layer; (2) Compute a scrambled baseline control activity level per condition by using ''block-randomized'' symmetry images as ''signal'' in the procedure of (1) above. This was done to assess local regularities associated with the symmetry images as explained in Stimuli (above), rather than the full effect of mirror symmetry; (3) Compute the discriminability between the signal activation difference (S) described in (1) and the baseline used as ''noise'' (N) described in (2) for multiple image pairs (500–2000 depending on the experiment). This was implemented by first converting both vectors to log ratios (dB), and then computing a d-prime (d') or signal to noise ratio (SNR) values in this metric (log distance between the means expressed in standard deviation (SD) units as $\log(S_{AV}/N_{AV})/SD$, where SD (scalar) is, as in Cohen's D, the Euclidean norm of S_{SD} and N_{SD} divided by $\sqrt{2}$. For computing error bars for the d-prime measures, we divided the data into 10 bins of image pairs and computed the average and standard error of the means (SEM) across the d-prime values for these bins. This computational scheme was used for the various discriminative conditions analyzed in the study.

To summarize the terms, we use: (a) ''Activation Diff'' – activation L2 differences, in arbitrary unis, plotted in dB ''relative to'' some appropriate value; (b) ''Relative to'' - used for plotting Activation Diffs (see (a)), determines the zero in the plot (c) ''Referenced to'' for computing ''activation diffs'', e.g., from fully random images; (d) ''Discriminability''/SNR/d-prime - between Activation Diffs of *Symx* and *Symx_br*; (e) ''Baseline'' – for discrimination, i.e., the Activation Diff measure from which to compute discriminability.

The analyses were run on a GTX 2080Ti GPU, in MATLAB (Natick, Massachusetts) using its Deep Learning Toolbox, with the simulations broken into up to 20 min segments and repeated as many times as needed. Results are discussed where they are statistically significant by the criterion that they exceed differences of 3 SEMs, for the larger of the two SEMs of the means being compared.