Numerosity representation in a deep convolutional neural network

Cihua Zhou\textsuperscript{1,2}, Wei Xu\textsuperscript{1}, Yujie Liu\textsuperscript{1}, Zhichao Xue\textsuperscript{1}, Rui Chen\textsuperscript{1}, Ke Zhou\textsuperscript{1} \& and Jia Liu\textsuperscript{1}

Abstract

Enumerating objects in the environment (i.e., “number sense”) is crucial for survival in many animal species, and foundational for the construction of more abstract and complex mathematical knowledge in humans. Perhaps surprisingly, deep convolutional neural networks (DCNNs) spontaneously emerge a similar number sense even without any explicit training for numerosity estimation. However, little is known about how the number sense emerges, and the extent to which it is comparable with human number sense. Here, we examined whether the numerosity underestimation effect, a phenomenon indicating that numerosity perception acts upon the perceptual number rather than the physical number, can be observed in DCNNs. In a typical DCNN, AlexNet, we found that number-selective units at late layers operated on the perceptual number, like humans do. More importantly, this perceptual number sense did not emerge abruptly, rather developed progressively along the hierarchy in the DCNN, shifting from the physical number sense at early layers to perceptual number sense at late layers. Our finding hence provides important implications for the neural implementation of number sense in the human brain and advocates future research to determine whether the representation of numerosity also develops gradually along the human visual stream from physical number to perceptual number.

Keywords

deep convolutional neural network, numerosity perception, numerosity underestimation, connectedness, perceptual organization

Received 23 December 2020; accepted 23 March 2021

Introduction

Humans and many animal species are equipped with an innate ability to approximately estimate the number of objects (i.e., numerosity) in their environment, namely the number sense (Burr & Ross, 2008; Cantlon et al., 2009; Dehaene, 1997, 2011; Feigenson, Dehaene, & Spelke, 2004). For animal species, the sense of number is critical for survival and reproduction in choosing zones with more food or joining a larger group of companions to avoid being eaten (e.g., Agrillo, Parrish, & Beran, 2016; Gross et al., 2009; Piffer, Agrillo, & Hyde, 2012). In humans, it is widely believed that the construction of more abstract and complex mathematical skills rests upon the basic number sense (e.g., Piazza, 2010). Atypical development of the number sense in the early stage of development may result in developmental dyscalculia, a neurodevelopmental disorder that causes specific mathematical learning difficulties (Bugden & Ansari, 2016; Butterworth, 2018; Price, Holloway, Rasen, Vesterinen, & Ansari, 2007). This shared ability, therefore and obviously, has evolutionary and developmental advantages for both animal species and human
beings. Perhaps surprisingly, recent evidence shows that brain-inspired deep convolutional neural networks (DCNNs) also spontaneously emerge an ability to sense the number of objects in a set, even when it was originally trained to classify objects in natural images, indicating that this biological instinct is no longer limited to living organisms (Nasr, Viswanathan, & Nieder, 2019). However, little is known about to what extent it is comparable with human’s number sense.

The distinction between biological and artificial systems is the origin of number sense. In biological systems, the number sense is thought from genetic predisposition, as the newborn babies, as well as day-old chicks and fishes, can discriminate between different numbers of objects (Agrillo & Bisazza, 2018; Butterworth, 2018; Rugani, 2018). In contrast, there are no predisposed abilities in DCNNs, and the number sense must derive from natural images used to train the networks (but see Kim, Jang, Back, Song, & Paik, 2021). Such distinction may lead to different mechanisms underlying the number sense between biological and artificial systems. To test this possibility, previous studies have found that the number sense in DCNNs also follows the Weber–Fechner law as observed in the biological systems, that the discrimination threshold of numerosity tends to increase linearly with the number of items in a set, or remains constant across a large range of numerosity on a logarithmic scale (e.g., Anobile, Cicchini, & Burr, 2016; Castelli, Glaser, & Butterworth, 2006; Cicchini, Anobile, & Burr, 2016; Dehaene & Changeux, 1993; Franconeri, Bemis, & Alvarez, 2009; He, Zhang, Zhou, & Chen, 2009; He, Zhou, Zhou, He, & Chen, 2015; Ross & Burr, 2010).

However, this similarity could be explained by another alternative account to numerosity perception, which claims that the numerosity is represented indirectly from judgements of the non-numerical magnitude properties (i.e., density, size, or cumulative surface area), rather than directly from the sense of number (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Durgin, 1995, 2008; Morgan, Raphael, Tibber, & Dakin, 2014; Tibber, Greenwood, & Dakin, 2012). Because these non-numerical magnitude properties also obey the Weber–Fechner law, it remains unclear whether the numerosity representation in the DCNN is based on the direct sense of number, or alternatively based on non-numerical magnitude properties. In most circumstances, the numerosity and continuous magnitude properties confound with each other. When dot arrays differ in numerosity, the continuous magnitude properties of these stimuli, such as the total area and density, differ as well. Therefore, to answer this question, a critical test is to examine whether DCNNs show numerosity underestimation effects.

The numerosity underestimation effects speak to a phenomenon observed in humans, that connecting pairs of dots with lines in a dot-display reduces the estimate of numerosity (Anobile et al., 2016; Fornaciai, Cicchini, & Burr, 2016; Franconeri et al., 2009; He et al., 2009; He et al., 2015). From the perspective of the “number sense” theory, numerosity perception should be based on the perceptual objects segmented by perceptual organization processes (He et al., 2015). Consequently, two items connected with a line tend to be perceived as one single object in a perceptual sense, rather than pairs of two objects in a physical sense, which then reduce the estimate of numerosity. On the contrary, if numerosity is determined by the continuous magnitude property such as the density, adding lines to the dot-display should increase the density and thereby increase the numerosity.

Here we tested whether the self-emerged number sense in the DCNN operated on the perceptual number (e.g., number of perceptual units or regions segmented by perceptual organization processes), or physical number (e.g., number of physical dots irrespective of whether they are connected or not). To do so, like previous studies (Kim et al., 2021; Nasr et al., 2019), we first identified the potential number-selective units in the AlexNet, a typical DCNN pre-trained for object classification, by using dot-displays without any connections. Then, we used the new dot-displays with pairs of connected dots as the input and tested whether these number-selective units could exhibit the numerosity underestimation effects, like humans do. Specifically, we designed different sets of dot-displays with either two pairs of dots (two-connected condition) connected, or no dots connected. Then we submitted these dot-displays to the AlexNet and obtained the response profile of each number unit to each numerosity in different stimulus sets. We hypothesized that, if the AlexNet shares the same perceptual number representation as that of humans, the number units should also “perceive” the dot-connected displays as being less numerous, compared with dot-displays with no dots connected.

Materials and methods

Neural network model

The DCNN model used in the present study is the AlexNet, which consists of five convolutional layers (Conv1 to Conv5) and three fully connected layers (FC1 to FC3). The ReLU (Rectified Linear Units) is applied to the output of each convolutional and fully connected layer. A max-pooling layer is attached after the Conv1, Conv2, and Conv5 layers. The layers of Conv1 to Conv5 have 193600, 139968, 64896, 43264,
43264, 4096, 4096, and 1000 units, respectively. Other detailed descriptions for the architecture and hyperparameters of this model can be found in Krizhevsky, Sutskever, and Hinton (2017). The AlexNet was pretrained to successfully perform object classification on the ILSVRC2012 ImageNet dataset, so that the model parameters were refined to extract hierarchical visual features of natural images (Krizhevsky et al., 2017).

**Stimulus sets**

Four sets of dot-displays were used in the present study. The discovery set contained zero-connected dot-displays with randomly distributed dots, which were used to identify the potential numerosity-selective units in the DCNN (Figure 1A). The second, third, and fourth sets were designed for the main analysis. The zero-connected set was designed in a similar way to the discovery set, with each dot-display containing only discrete dots (Figure 1B). This zero-connected set was used to validate the numerosity-selective units after identifying them using the discovery set, and also served as a baseline condition for estimating the numerosity underestimation effects. The two-connected set contained dot-displays in which two pairs of dots, each connected with a line-segment, formed a dumbbell-like configuration (Figure 1C). The zero-connected-control set contained

![Stimuli examples](image-url)

**Figure 1.** Stimuli used in the present study, and response profile of each numerosity-selective PN-set. (A–D) Illustrations of the dot-displays (with 5 dots as exemplars) from the discovery (A), zero-connected (B), two-connected (C), and zero-connected-control (D) sets, respectively. (E) Average tuning response of each PN-set to every numerosity of the zero-connected set. Only the results of the FC2 layer were shown here for demonstration purposes. (F) Distribution of preferred numerosities in the FC2 layer. (G) Standard deviation (sigma) of best-fitting Gaussian function for tuning curves of each PN-set in the FC2 layer, for both linear and logarithmic scales. Ideal line refers to the PNs for each PN-set defined by the discovery set. (H) Estimated PN (green line) of each PN-set in the FC2 layer for the zero-connect set. Ideal line refers to the hypothesized PN. (I) Estimated PN of each PN-set at each layer for the zero-connect set (Conv1 to FC2).
dot-displays similar to those in the third set, with the exception that each of the two line-segments had one end attached to a dot with the other end hanging freely, forming a pin-like pattern (Figure 1D). Critically, to keep the low-level statistical information of the stimuli similar, the dot distributions of the corresponding dot-displays from the second, third, and fourth sets were exactly the same, as illustrated in Figure 1A.

Each set contained 14,000 dot-displays, 1000 for each of 14 numerosities (5 to 18). The dots and line-segments were black drawn on a white background (224 × 224 pixels). The diameter of the dot was 12 pixels. The dots (and line-segments) were randomly distributed in the display following the constraints that no two dots were closer than 18 pixels and the minimum distance from the dot to the border of the image was 8 pixels. In the two-connected and zero-connected-control sets, each of the two line-segments had a width of 2 pixels and a length ranging from 18 to 86 pixels. The connections between pairs of dots were designed to avoid line-segments crossing each other or other dots.

**Overall analytic strategy**

In the present study, we adopted a Region of interest (ROI) analysis, a prevalent approach in fMRI studies of visual processing (Poldrack, 2007). Specifically, we first used dot-display without any connections to identify the number-selective units in the DCNN. Then, on these number-selective units, we continued to explore how they behaved when presented with the new dot-displays (zero-connected, two-connected, and zero-connected-control sets).

**Defining the numerosity-selective units**

In each layer of the pre-trained AlexNet, using the DNNBrain toolbox (Chen et al., 2020) (https://github.com/BNUCNL/dnnbrain/), we first obtained the activation values after the ReLU operation of every unit in response to every dot-display of the discovery set. Because the FC3 has no function of ReLU and is a layer for classification output, our analysis was then limited to the layers of Conv1 to FC2. Next, activation values of all units for each numerosity were submitted to a one-way ANOVA with numerosity (14 levels: 5–18) as the factor, in a similar manner to that used in Nasr et al. (2019). A unit that exhibited a significant main effect of numerosity (p < 0.01, i.e., a significant change of the response across numerosities) was a potential candidate unit that might be numerosity-selective. The preferred numerosity (PN) of this candidate unit was defined as the numerosity that induced the largest average response among the responses for all stimuli. All candidate units that had the same PN formed a PN-set. A total of 14 PN-sets were then obtained for each layer.

Subsequently, we performed a validation analysis to confirm whether these candidate units were truly numerosity-selective. If the units identified by one stimulus-set were truly numerosity-selective, they should exhibit the similar tuning patterns of numerosity when testing with another independent stimulus-set. For each PN-set in each layer, we used the dot-displays from the zero-connect set as the new input and obtained a mean activation value of this PN-set for each numerosity of the zero-connected set. Two criteria were then used to define a reliable numerosity-selective PN-set. First, for each PN-set, the new PN calculated based on the zero-connected set should be consistent with the PN defined by using the discovery set. Second, because the number sense follows the Weber–Fechner law, the width of the tuning curve of each PN-set should increase linearly with the number of items in a set, or remain constant across numerosities on a logarithmic scale. The Gaussian functions were then fitted to the mean activation values of each PN-set on two different scales, a linear scale (f(x) = x) and a logarithmic scale (f(x) = log2(x)), to obtain the estimated PN and width of the tuning curve. For each PN-set, the estimated PN and tuning widths were defined as the mean and standard deviation (sigma) of the Gaussian fit, respectively. A regression analysis was used to test whether the slope (i.e., increase of tuning width as a factor of numerosity) was significantly different than zero.

Only the candidate PN-set that had an estimated PN consistent with its originally defined PN and followed the Weber–Fechner law was finally defined as the numerosity-selective PN-set. All the following analysis was therefore carried out on these PN-sets, unless otherwise specified.

**Analysing the response profiles of the PN-sets for three stimulus-sets**

Using the same method described above, in each layer, we first obtained the tuning curves of each PN-set to the numerosities from the zero-connected, two-connected, and zero-connected-control sets. The tuning curve of each PN-set was normalized to the 0 to 1 range and fitted by the Gaussian function. For each of the zero-connected, two-connected, and zero-connected-control sets, we obtained the estimated PN of each PN-set at each layer. It should be noted that the estimated PN of a PN-set for one stimulus-set referred to the numerosity that induced the largest average response among the responses for all stimuli within this stimulus-set. Importantly, for a PN-set having a
pre-defined PN of N, the dot-displays of the two-connected set that can induce the maximal response of this PN-set, should contain a greater number of physical dots than N, if there was a numerosity underestimation effect.

Note that the PN-sets preferring smaller or larger numerosities, compared with other numerosities, exhibited a response pattern of monotonic increase or decrease as the stimulus numerosity increased, respectively. Thus, only part of the entire tuning curve of these PN-sets was submitted to a Gaussian fit. To avoid the poor fitting at the boundary of the test range, we excluded the two numerosities (5 and 18) from further analysis and limited our fitting procedure only on the numerosities of 6 to 17.

**Estimating numerosity underestimation effects**

Based on the estimated PNs obtained in the above step, we submitted the estimated PNs for the two-connected and zero-connected sets to a Wilcoxon signed-ranks test to evaluate the numerosity underestimation effect in each layer. For each PN-set, the underestimation effect was calculated as the difference in the estimated PN between the two-connected and zero-connected sets. Then at each layer, the underestimation effects for all PN-sets were averaged to obtain a mean underestimation effect for this layer.

In addition, to rule out an alternative explanation that the line-segments themselves may contribute to the numerosity underestimation effects, we also carried out a control analysis by comparing the estimated PNs from the two-connected and zero-connected-control sets using a Wilcoxon signed-ranks test.

**Results**

First, we tested at which layer the number sense can self-emerge in the pre-trained AlexNet. We totally identified 51.6%, 71.3%, 60.6%, 52.4%, and 32.3%, 66.4%, 20.6% of all units as the potential numerosity-selective units at the layers from Conv1 to FC2, respectively. The proportion of number-selective units seemed to be greater than that reported in previous research (Nasr et al., 2019), which might be due to only one stimulus set or the different parameters of the AlexNet being used in our study. However, similar to those found in previous studies (Kim et al., 2021; Nasr et al., 2019), the distribution of preferred numerosities covered the entire range (5 to 18), with relatively more units preferring the numerosity of 5 or 18 (Figure 1F, results of FC2 layer as an example).

Next, in each layer, all candidate units that had the same PN were pooled to form a PN-set for each PN. If these potential PN-sets were truly numerosity selective, their response profiles should be invariant to the visual appearance of the dot-displays. Therefore, we carried out a validation analysis using a completely new zero-connected stimulus set (Figure 1B), to confirm whether they showed reliable and invariant selectivity for numerosity. As evidenced in Fig 1I, the validation analysis revealed a trend that the numerosity representation became progressively very reliable as the layer increased. First, the PN-sets at the Conv1 and Conv2 layers defined by using the discovery set didn’t show consistent numerosity selective responses across stimulus-sets, as their estimated PNs for the new zero-connected set were dramatically different from those for the discovery set. And their tuning responses to each PN were also different between both stimulus sets. For example, as illustrated in S2 Figure, though PN-sets at the Conv 1 layer showed some tuning responses to numerosities for the discovery set, they didn’t exhibit clear tuning patterns to numerosities for the zero-connected set. On the contrary, the PN-sets at the layers of Conv3 to FC2 showed the similar estimated PNs for the zero-connected set as those for the discovery set, and they also exhibited similar tuning patterns to numerosities for both the discovery and zero-connected sets. Figure 1E and S3 Figure showed the typical response profiles of PN-sets at the FC2 layer as an example (also see S2 Figure for results of the Conv5 layer). The tuning responses of the PN-sets at the FC2 layer followed the Weber–Fechner law (Figure 1E), as the width of the tuning curve of each PN-set increased linearly with numerosity on a linear scale (regression analysis: slope = 0.11, p < 0.05, partial \( \eta^2 = 0.24 \)) and remained constant across numerosities on a logarithmic scale (regression analysis: slope = 0.005, p > 0.05) (Figure 1G). We thus defined the PN-sets at the layers of Conv3 to FC2 as the truly numerosity selective PN-sets.

Then, we explored whether these numerosity selective PN-sets represented the number based on physical or perceptual sense, and how the number sense evolved through the hierarchy of the DCNN. In the layers from Conv3 to FC2, we compared the responses of these numerosity selective PN-sets for the zero-connected and two-connected sets to test the numerosity underestimation effects. We found that, at the layers of Conv3 and Conv4, there were no significant differences in the estimated PNs of the PN-sets between the zero-connected and two-connected sets, reflecting a representation based on the number of physical dots in a dot array (all \( ps > 0.05 \)) (Figure 2B, S4A and S5A Figures). However, and most importantly, at the other three layers from Conv5 to FC2, the estimated PNs for the two-connected set were significantly higher than those for the zero-connected set, representing a number representation based on the perceptual numbers (all \( ps < 0.01 \), all Cohen’s ds > 1.3) (Figure 2B, S6A and S7A Figures). Thus, significant numerosity
underestimation effects were only found at the layers of Conv5, FC1, and FC2, and there was a trend that the numerosity underestimation effects became progressively much greater as the layer increased (mean underestimation effect: 0.69, 1.40, and 2.10 for Conv 5, FC1, and FC2 layers, respectively). Figure 2A showed the results of each PN-set at the FC2 layer as an example. The 'perceived' numerosity of the dot-displays in the two-connected set was systematically underestimated, as reflected as a rightward shift of the red curve in Figure 2A and upward shift of the red line in Figure 2B. That is, for the PN-set with a pre-defined preferred numerosity, the dot-displays in the two-connected set that induced the maximal response of this PN-set contained more physical dots than those in the zero-connected set.

Finally, since the two-connected and zero-connected set differed not only in the number of perceptual units, but also in the presence of line-segments, one may argue that the line-segments themselves would increase the effective number of items to be counted and thus result in the numerosity underestimation effects. Therefore, at the layers of Conv5, FC1, and FC2, we compared the responses of these numerosity selective PN-sets for the zero-connected-control and two-connected sets to address this issue. We found that the two-connected condition still had significantly greater numerosity underestimation effects than the zero-connected-control condition (all ps < 0.01, all Cohen's ds > 1.69). It thus ruled out this alternative explanation (Figure 3A, results of the FC2 layer as an example; Figure 3B, S4B, S5B, S6B, and S7B Figures), suggesting that the line-segments themselves cannot fully account for the underestimation of perceived numerosity in our study.

Discussion

In the literature, of particular relevance to the present study, several computational models regarding the
nature of the number sense have been put forward (Dakin et al., 2011; Dehaene & Changeux, 1993; Hannagan, Nieder, Viswanathan, & Dehaene, 2018; Stoianov & Zorzi, 2012; Testolin, Dolfi, Rochus, & Zorzi, 2020; Testolin, Zou, & McClelland, 2020; Verguts & Fias, 2004; Zorzi & Testolin, 2018). Most of them were of high cognitive fidelity and relied either on implementing the predetermined number detectors in the architecture, or on learning through extensive exposure to numerical stimuli. However, their biological implementation in the brain remains challenging. An alternative to these models is the biologically inspired DCNN, designed to mimic the retinotopic and hierarchical nature of the primate visual system, which has a relatively high biological fidelity to date. Several prominent studies did find the number detectors emerged in the DCNN, even when the neural network was trained only for object recognition (DeWind, unpublished results; Kim et al., 2021; Nasr et al., 2019). However, these studies have mostly focused on demonstrating that the number detectors exhibited the tuning characteristics following the Weber–Fechner law, similar to those observed in the primate studies. There is still a lack of understanding of how the number is represented and developed in the DCNN.

In the present study, we first replicated the previous findings that numerosity-selective units spontaneously emerged in the AlexNet pre-trained for object classification (Kim et al., 2021; Nasr et al., 2019). Then, via a systematic manipulation of the connection between dots, our study went beyond these studies, showing further that the number-selective units at the later layers of the AlexNet operated on the perceptual number, rather than the physical number. The existence of numerosity underestimation effect in the DCNN provided a new support for the similarity of visual numerosity processing in the DCNN and that in humans. Note that future studies, in which the number-selective units are defined by using the 2-connected set and their responses for the 0-connected set are tested to look for an overestimation effect, could be carried out to further validate our hypothesis.

However, this conclusion seemed to be counterintuitive. On the one hand, the DCNN used in the present study was trained only for object classification, without any explicit requirements of numerosity judgement. One may question why the DCNN has to develop a human-like number sense when learning to recognize natural objects. However, from the perspective of cognitive development, object recognition relies on the object individuation process, also known as the figure-ground segmentation. The figure-ground segmentation process segregates the to-be-recognized object from others, and the background (Johnson, 2001; Ostrovsky, Meyers, Ganesh, Mathur, & Sinha, 2009; Xu, Carey, & Welch, 1999). And importantly, its output also serves as the basis for numerosity perception. In this sense, numerosity perception may share with object recognition the same intermediate processing stage, i.e., the object individuation process or figure-ground segmentation. In fact, we have proposed that numerosity perception requires the prior
segmentation of the visual stimuli into perceptual objects (He et al., 2015). Moreover, numerous human studies have established that perceptual organization plays a crucial role in the object individuation process, by segmenting a visual image into a potential collection of perceptual units according to the Gestalt principles (for a review, see Peterson, 2015). Therefore, if the DCNN can learn numerosity representations that are more consistent with those used by humans, we would predict that grouping pairs of dots into single units should also lead to lower estimates of the number of dots for the DCNN. This is indeed what we have observed in the DCNN, which closely follows human numerosity perception. In short, our results provided new computational evidence for the notion in cognitive science that perceptual organization may precede number estimation (Anobile et al., 2016; Fornaciai et al., 2016; He et al., 2015). To date, there are various CNN models available for object classification tasks, for instance, AlexNet, VGGNet (Simonyan & Zisserman, 2015), GoogLeNet (Szegedy et al., 2015), ResNet (He, Zhang, Ren, & Sun, 2016), and InceptionResNet-v2 (Szegedy, Ioffe, Vanhoucke, Alemi, & Aaai, 2017). All of them have achieved high accuracies in classifying natural images. It will be interesting for future studies to establish whether the numerosity underestimation effect found in the AlexNet could be generalized to other kinds of CNNs.

On the other hand, in humans and non-human primates, it is generally accepted that both feedforward and feedback processes contribute to perceptual organization (for a review, see Self & Roelfsema, 2015). However, a fundamental limitation of the DCNN is its feedforward architecture, which only mimics the bottom-up function of the primate brain. For this reason, there is an ongoing debate on whether the DCNN possesses the ability of perceptual organization (Serre, 2019). Previous neurophysiology study on figure-ground segmentation has suggested that different perceptual organization processes may be mediated by distinct neural circuits. For instance, the boundary detection relied largely on feedforward connections through the visual hierarchy, whereas the region-filling process depended more on the feedback from higher visual areas (Poort et al., 2012). Since the uniform connectedness plays a vital role in boundary detection, we then argued that the perception of uniform connectedness might be achieved via a feedforward manner in the DCNN. In fact, the DCNN contains a multilayer hierarchical structure mimicking the primate brain, and can better represent increasingly more complex features across layers. In humans, other principles of the perceptual organization, such as subjective contour and closure, have also been found to lead to the numerosity underestimation effect (He et al., 2015; Kirjakovski & Matsumoto, 2016). For the DCNN, future investigations are still needed to address whether similar numerosity underestimation could occur when dots are grouped by these organization principles, and if not, whether adding the recurrent connections to the current feedforward architecture would provide a better fit to human numerosity perception.

In addition, we found that the perceptual number sense did not emerge abruptly in the DCNNs. Instead, the numerosity perception in the DCNN evolved a physical to perceptual sense of number through the hierarchy. First, the numerosity perception emerged early at the layer of Conv3, but its internal representation was determined only by the number of physical dots in a display, irrespectively of the connections between them. Then, the number sense developed progressively along the hierarchy, evolving a gradual transformation from the sense of physical number to the sense of perceptual number. Finally, by the end of the DCNN, numerosity representation operated on the perceptual number. This finding provided an important implication for the neural implementation of the number sense in humans, a fundamental question not well answered in the cognitive neuroscience of numerosity perception. If the DCNN shares the similar internal representation of number with humans, we might hypothesize that the processing of numerosity in the human brain might also evolve a shift from a physical number sense to perceptual number sense along the visual stream. While most of the recent fMRI and ERP studies focused on whether numerosity representation is based on the direct sense of number, or on non-numerical continuous magnitude properties (DeWind, Park, Woldorff, & Brannon, 2019; Fornaciai, Brannon, Woldorff, & Park, 2017; Fornaciai & Park, 2017; Lyons, Ansari, & Beilock, 2015; Park, DeWind, Woldorff, & Brannon, 2015; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Sokolowski, Fias, Mousa, & Ansari, 2017), there was one study that began to address the difference in the activations of several brain regions between the sensory and perceptual representation of numerosity (Fornaciai & Park, 2018). Their initial findings did support our hypothesis by showing that activations in V2 were modulated by a physical sense of number while activations in V3 represented the perceptual encoding of numerosity. However, a systematic investigation based on the entire cortical hierarchy, with the application of more sensitive measure of neural representation, is still needed. Our finding thus advocates future research to determine whether the representation of numerosity in the human brain also develops gradually along the visual stream, from physical sense to perceptual sense.
Taken together, our results provide compelling evidence suggesting that the numerosity perception of the DCNN acts upon the perceptual number, like humans.

Declaration of conflicting interest
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and publication of this article: this work was supported by the National Key Research and Development Program of China (2019YFA0709503); National Nature Science Foundation of China (31671133); National Basic Research Program of China (2018YFC0810602); and Fundamental Research Funds for the Central Universities.

ORCID iD
Ke Zhou https://orcid.org/0000-0003-4773-758X

Supplemental material
Supplementary material for this article is available online.

References
Agrillo, C., & Bisazza, A. (2018). Understanding the origin of number sense: A review of fish studies. Philosophical Transactions of the Royal Society B-Biological Sciences, 373(1740). https://doi.org/10.1098/rstb.2016.0511
Agrillo, C., Parrish, A. E., & Beran, M. J. (2016). How illusory is the solitaire illusion? Assessing the degree of misperception of numerosity in adult humans. Frontiers in Psychology, 7. https://doi.org/10.3389/Fpsyg.2016.01663
Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. Perception, 45(1–2), 5–31. https://doi.org/10.1177/0301006615602599
Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the ‘Approximate Number System’ in children with persistent Developmental Dyscalculia. Developmental Science, 19(5), 817–833. https://doi.org/10.1111/desc.12324
Burr, D., & Ross, J. (2008). A visual sense of number. Current Biology, 18(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052
Butterworth, B. (2018). The implications for education of an innate numerosity-processing mechanism. Philosophical Transactions of the Royal Society B-Biological Sciences, 373(1740), 16. https://doi.org/10.1098/rstb.2017.0118
Cantlon, J. F., Libertus, M. E., Pinel, P., Dehaene, S., Brannon, E. M., & Pelphrey, K. A. (2009). The neural development of an abstract concept of number. Journal of Cognitive Neuroscience, 21(11), 2217–2229. https://doi.org/10.1162/jocn.2008.21159
Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. Proceedings of the National Academy of Sciences of the United States of America, 103(12), 4693–4698. https://doi.org/10.1073/pnas.060044103
Chen, X., Zhou, M., Gong, Z., Xu, W., Liu, X., Huang, T., … Liu, J. (2020). DNNBrain: A unifying toolbox for mapping deep neural networks and brains. Frontiers in Computational Neuroscience, 14(105). https://doi.org/10.3389/fncom.2020.580632
Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. Nature Communications, 7. https://doi.org/10.1038/Ncomms12536
Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A universal common metric for approximate number and density. Proceedings of the National Academy of Sciences of the United States of America, 108(49), 19552–19557. https://doi.org/10.1073/pnas.1113195108
Dehaene, S. (1997). Conscious and unconscious processing of numerical information. Journal of Psychophysiology, 11(4), 366–366.
Dehaene, S. (2011). Origins of proto-mathematical intuitions. in Dehaene, S. & Brannon, E. M. (Eds.), Space, time and number in the brain: searching for the foundations of mathematical thought, pp.187–189. Elsevier.
Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities – A neuronal model. Journal of Cognitive Neuroscience, 5(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390
DeWind, N. K. (unpublished results). The number sense is an emergent property of a deep convolutional neural network trained for object recognition. bioRxiv. https://doi.org/10.1101/609347
DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. Cortex, 114, 76–89. https://doi.org/10.1016/j.cortex.2018.03.027
Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. Journal of Experimental Psychology-Human Perception and Performance, 21(1), 149–169. https://doi.org/10.1037/0096-1523.21.1.149
Durgin, F. H. (2008). Texture density adaptation and visual number revisited. Current Biology, 18(18), R855–R856. https://doi.org/10.1016/j.cub.2008.07.053
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. Trends in Cognitive Sciences, 8(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002
Fornaciai, M., Brannon, E. M., Woldorff, M. G., & Park, J. (2017). Numerosity processing in early visual cortex. Neuroimage, 157, 429–438. https://doi.org/10.1016/j.neuroimage.2017.05.069
Fornaciai, M., Cicchini, G. M., & Burr, D. C. (2016). Adaptation to number operates on perceived rather than physical numerosity. Cognition, 151, 63–67. https://doi.org/10.1016/j.cognition.2016.03.006
Fornaciai, M., & Park, J. (2017). Distinct neural signatures for very small and very large numerosities. Frontiers in Human Neuroscience, 11(21). https://doi.org/10.3389/fnhum.2017.00021
Fornaciai, M., & Park, J. (2018). Neural sensitivity to numerosity in early visual cortex is not sufficient for the representation of numerical magnitude. *Journal of Cognitive Neuroscience, 30*, 1–15. https://doi.org/10.1162/jocn_a_01320

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113*(1), 1–13. https://doi.org/10.1016/j.cognition.2009.07.002

Gross, H. J., Pahl, M., Si, A., Zhu, H., Tautz, J., & Zhang, S. W. (2009). Number-based generalisation in the honeybee. *Plos One*, 4(1). https://doi.org/ARTN e4263

Hannagan, T., Nieder, A., Viswanathan, P., & Dehaene, S. (2018). A random-matrix theory of the number sense. *Philosophical Transactions of the Royal Society B-Biological Sciences, 373*(1740). https://doi.org/10.1098/rstb.2017.0253

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review, 16*(3), 509–517. https://doi.org/10.3758/pbr.16.3.509

He, L. X., Zhou, K., Zhou, T. G., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), E5647–E5655. https://doi.org/10.1073/pnas.1512408112

Johnson, S. P. (2001). Visual development in human infants: Binding features, surfaces, and objects. *Visual Cognition, 8*(3–5), 565–578. https://doi.org/10.1080/13506280143000124

Kim, G., Jang, J., Baek, S., Song, M., & Paik, S.-B. (2021). Visual number sense in untrained deep neural networks. *bioRxiv Science Advances, 7*, eabdi6127. https://doi.org/10.1126/sciadv.abdi6127

Kirjakovski, A., & Matsumoto, E. (2016). Numerosity underestimation in sets with illusory contours. *Vision Research, 122*, 34–42. https://doi.org/10.1016/j.visres.2016.03.005

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90. https://doi.org/10.1145/3065386

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping, 36*(2), 475–488. https://doi.org/10.1002/hbm.22641

Morgan, M. J., Raphael, S., Tibber, M. S., & Dakin, S. C. (2014). A texture-processing model of the ‘visual sense of number’. *Proceedings of the Royal Society B-Biological Sciences, 281*(1790). https://doi.org/10.1098/rspb.2014.1137

Nasr, K., Viswanathan, P., & Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Science Advances, 5*(5), eaav7903. https://doi.org/10.1126/sciadv.aav7903

Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., & Sinha, P. (2009). Visual parsing after recovery from blindness. *Psychological Science, 20*(12), 1484–1491. https://doi.org/10.1111/j.1467-9280.2009.02471.x

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2015). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex, 26*(2), 748–763. https://doi.org/10.1093/cercor/bhv017

Peterson, M. (2015). Low-level and high-level contributions to figure-ground organization. The Oxford Handbook of Perceptual Organization. Oxford: Oxford University Press (pp. 259–280). https://doi.org/10.1093/oxfordhb/9780199686858.013.059

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14*(12), 542–551. https://doi.org/10.1016/j.tics.2010.09.008

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piffer, L., Agrillo, C., & Hyde, D. C. (2012). Small and large number discrimination in guppies. *Animal Cognition, 15*(2), 215–221. https://doi.org/10.1007/s10071-011-0447-9

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 1*(1), 67–70. https://doi.org/10.1093/scan/nsm006

Poort, J., Raudies, F., Wannig, A., Lamme, V. A. F., Neumann, H., & Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron, 75*(1), 143–156. https://doi.org/10.1016/j.neuron.2012.04.032

Price, G. R., Holloway, I., Rasanen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology, 17*(24), R1042–R1043. https://doi.org/10.1016/j.cub.2007.10.013

Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision, 10*(2), 10. https://doi.org/10.1167/10.2.10

Rugani, R. (2018). Towards numerical cognition’s origin: Insights from day-old domestic chicks. *Philos Trans R Soc Lond B Biol Sci, 373*(1740), 20160509. https://doi.org/10.1098/rstb.2016.0509

Self, M. W., & Roelfsema, P. R. (2015). Scene perception in early vision: Figure-ground organization in the lateral geniculate nucleus. *Proceedings of the National Academy of Sciences, 112*(22), 6784–6785. https://doi.org/10.1073/pnas.1507097112

Serre, T. (2019). Deep learning: The good, the bad, and the ugly. *Annual Review of Vision Science, Vol 5*, 5, 399–426. https://doi.org/10.1146/annurev-vision-091718-014951

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*. https://doi.org/abs/1409.1556v5

Sokolowski, H. M., Fias, W., Mousa, A., & Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: A functional neuroimaging
meta-analysis. *Neuroimage, 146*, 376–394. https://doi.org/10.1016/j.neuroimage.2016.10.028

Stoianov, I., & Zorzi, M. (2012). Emergence of a ‘visual number sense’ in hierarchical generative models. *Nature Neuroscience, 15*(2), 194–196. https://doi.org/10.1038/nn.2996

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A., & Aaai. (2017). *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*. Palo Alto: Assoc Advancement Artificial Intelligence.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., . . . Rabinovich, A. (2015). *Going Deeper with Convolutions*. Paper presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Testolin, A., Dolfi, S., Rochus, M., & Zorzi, M. (2020). Visual sense of number vs. sense of magnitude in humans and machines. *Scientific Reports, 10*(1). https://doi.org/10.1038/S41598-020-66838-5

Testolin, A., Zou, W. Y., & McClelland, J. L. (2020). Numerosity discrimination in deep neural networks: Initial competence, developmental refinement and experience statistics. *Developmental Science, 23*(5), e12940. https://doi.org/10.1111/desc.12940

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision, 12*(6). https://doi.org/10.1167/12.6.8

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience, 16*(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Xu, F., Carey, S., & Welch, J. (1999). Infants’ ability to use object kind information for object individuation. *Cognition, 70*(2), 137–166. https://doi.org/10.1016/S0010-0277(99)0007-4

Zorzi, M., & Testolin, A. (2018). An emergentist perspective on the origin of number sense. *Philosophical Transactions of the Royal Society B-Biological Sciences, 373*(1740), 20170043. https://doi.org/10.1098/rstb.2017.0043