nature neuroscience

Emergence of a 'visual number sense' in hierarchical generative models

Ivilin Stoianov[1] & Marco Zorzi[1,2]

Numerosity estimation is phylogenetically ancient and foundational to human mathematical learning, but its computational bases remain controversial. Here we show that visual numerosity emerges as a statistical property of images in 'deep networks' that learn a hierarchical generative model of the sensory input. Emergent numerosity detectors had response profiles resembling those of monkey parietal neurons and supported numerosity estimation with the same behavioral signature shown by humans and animals.

Many animal species have evolved a capacity to estimate the number of objects seen[1]. Numerosity estimation is foundational to mathematical learning in humans[2,3], and susceptibility to adaptation suggests that numerosity is a primary visual property[4]. Nonetheless, the nature of the computations underlying this "visual sense of number"[4] remains controversial[5]. Variability in object size prevents a simple solution based on the summation of their surface area (cumulative surface area), which is a main perceptual correlate of numerosity. A prominent theory[6] requires object size normalization as key preprocessing stage for numerosity estimation. Others circumvent the problem, assuming the use of "occupied area" independent of object size[7].

Here we show that visual numerosity emerges as a statistical property of images through unsupervised learning. We used deep networks, multilayer neural networks that contain top-down connections and learn to generate sensory data rather than to classify it[8,9]. Stochastic hierarchical generative models are appealing because they develop increasingly more complex distributed nonlinear representations of the sensory input across layers[9]. These features make deep networks particularly attractive for the purpose of neuro-cognitive modeling.

The deep network had one 'visible' layer encoding the sensory data and two hierarchically organized 'hidden' layers (**Fig. 1**). The training database consisted of 51,200 unlabeled binary images containing up to 32 randomly placed objects with variable surface area, such as those in **Supplementary Figure 1a**. Crucially, learning concerned only efficient coding of the sensory data (that is, maximizing the likelihood of reconstructing the input) and not number discrimination, as information about object numerosity was not provided (**Supplementary Methods** and **Supplementary Fig. 1**).

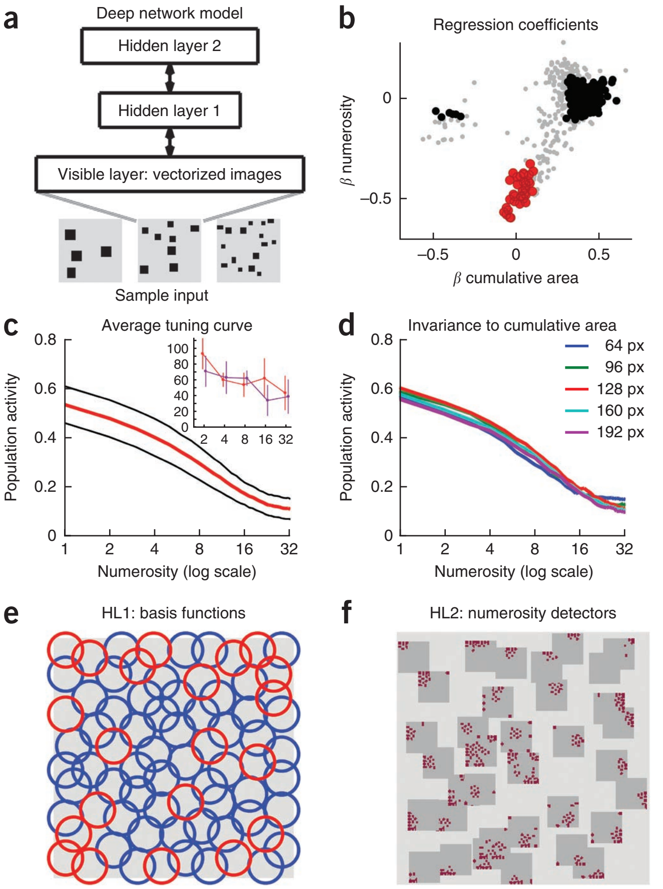

Figure 1 Deep network model and number-sensitive neurons. (**a**) Architecture of the deep network model and sample input images (samples with 4, 8 and 16 objects and equal cumulative area). (**b**) Regression coefficients for log(numerosity) and log(cumulative area) of neurons in the second hidden layer. Selectivity is indexed by large absolute value of one coefficient combined with near-zero value of the other. Red, numerosity detectors; black, cumulative-area detectors; gray, non-selective neurons. (**c**) Population activity of numerosity detectors (mean activation value) as a function of number of objects (±1 s.d. bands represent variability across images). Inset (adapted from ref. 10): corresponding response (mean firing rate ± s.e.m.) of a number-sensitive neuron in the monkey LIP area (red and purple represent different experimental blocks). (**d**) Population activity of numerosity detectors (mean activation value), showing invariance to cumulative area in pixels (px). (**e**) Spatial properties of off-center (blue) and on-center (red) basis functions in hidden layer 1 (HL1) (samples in **Supplementary Fig. 2**) superimposed on the image space (gray area). (**f**) Spatial selectivity of numerosity detectors in hidden layer 2 (HL2), represented as 30 × 30 pixel plots superimposed on the image space (light gray area). Each colored point in a neuron's receptive field (dark gray squares) represents an HL1 center-surround neuron.

[1]Dipartimento di Psicologia Generale, Università di Padova, Italy. [2]Center for Cognitive Science, Università di Padova, Italy. Correspondence should be addressed to M.Z. (marco.zorzi@unipd.it).

Received 26 August 2011; accepted 26 October 2011; published online 8 January 2012; doi:10.1038/nn.2996

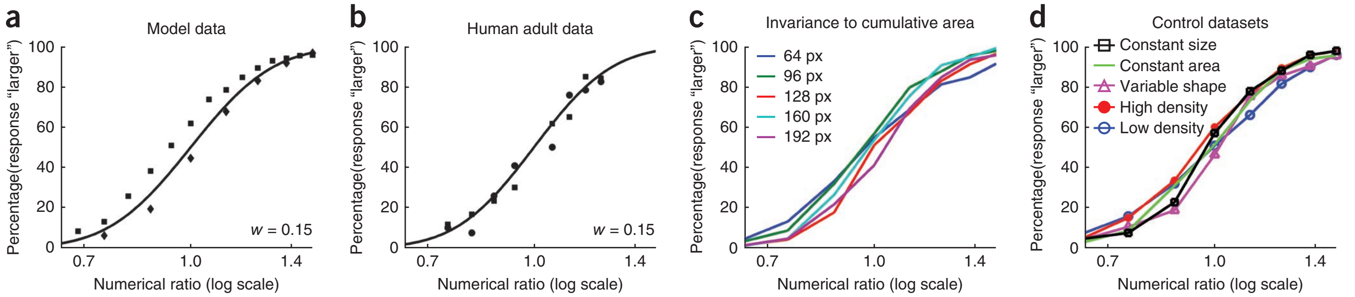

Figure 2 Numerosity comparison task. Probability of the response "larger" as a function of the log-ratio of input numerosity and reference. (**a**) Numerosity discrimination on the test dataset with 8 (diamonds) or 16 (squares) as reference, indexed by a Weber fraction of $w = 0.15$ (sigmoid fit). (**b**) Human adult data (replotted from ref. 3) in numerosity comparison (squares, 16 as reference; circles, 32). (**c**) Invariance to cumulative area in pixels (px). (**d**) Performance on control data sets A–D: constant object area (black), constant cumulative area (green), variable object features (purple) and variable density (red and blue).

We first sought sensitivity to numerosity information after learning in terms of internal coding by hidden neurons, controlling for the confounding cumulative surface area. Compressed monotonic coding that resembles a scalar variable is the simplest number code found in the lateral intraparietal (LIP) area of the monkey brain[10]. We found distinct populations of neurons in the second hidden layer (HL2) that noisily estimated numerosity and cumulative area, respectively (**Fig. 1b** and **Supplementary Methods**). The numerosity detectors in particular showed response profiles consistent with the neurophysiological data (**Fig. 1c**). Average activity across numerosity detectors was well explained by log(numerosity) of the stimulus (regression $R^2 = 0.82$) and was invariant to cumulative area (**Fig. 1d**), suggesting that population coding can support numerosity estimation.

We then assessed whether HL2 neurons could support numerosity comparison[2,3,11]. A linear classifier, fed with HL2 activity, was trained on the image dataset to decide whether a visual numerosity was larger than a reference number (either 8 or 16) (**Supplementary Methods**). The classifier scored 93% on a novel test set of 51,200 images. This set was also used to thoroughly assess numerosity discrimination, which is modulated by numerical ratio in humans and animals[1–3,11]. Probability of the response "larger", plotted as a function of the log ratio of the two numbers (test numerosity/reference), followed a classic sigmoid curve (**Fig. 2a**). Notably, the curves for the two reference numbers were identical, in accordance with Weber's law for numbers[1] and in excellent agreement with human behavioral studies[2,3,11] (**Fig. 2b**). The response distributions were used to compute an index of number discriminability (also known as number acuity[2,3]), the internal Weber fraction[11] w (**Supplementary Methods**). More intuitively, $2w$ represents the proportion by which a numerosity must differ from the reference to be discriminable with about 95% confidence[11]. The model's w was 0.15, which is in line with the mean values observed in human adults[3,11]. Crucially, numerosity estimation was invariant to cumulative area (**Fig. 2c**).

We also generated four more test sets to assess the model's numerosity estimation ability under specific conditions, as in animal studies[12,13] (**Fig. 2d** and **Supplementary Methods**). Set A contained objects with fixed size and shape (squares of 3 × 3 pixels) for all numerosities, set B had equal cumulative surface area (100 pixels) for all numerosities (object size therefore decreased with increasing numerosity), set C had objects with variable features (shape, size and orientation) in each image and set D had two density levels for each numerosity. The w values for these sets were 0.13, 0.14, 0.14 and 0.17, respectively. These results show that numerosity estimation in the model, like that in animals and humans[1,13], is invariant to cumulative area, density and object features (**Fig. 2c,d**).

Analyses of the network computations revealed that most of the first hidden layer (HL1) neurons were center-surround detectors that uniformly covered the image space (**Fig. 1e**; see examples in **Supplementary Fig. 2**). Also, the numerosity detectors in HL2 were spatially selective (**Fig. 1f**). They received strong input from HL1 neurons with spatially aligned receptive fields. They also received inhibition from a few HL1 neurons that encoded cumulative area, thereby providing a normalization signal. Thus, the numerosity detectors encoded local, size-invariant numerosity. The population activity of HL2 numerosity detectors was well predicted by a linear combination of the population activity of the two types of HL1 neuron (**Supplementary Fig. 3**), and it adequately supported numerosity comparison when used as the sole input to a classifier (**Supplementary Methods**). Simulations with a simplified mathematical model confirmed these analyses (**Supplementary Methods** and **Supplementary Fig. 4**). We emphasize that the response properties of the hidden neurons were not stipulated in any way but represent an emergent property of the image data obtained without supervision.

Unsupervised 'deep learning' discovered statistical features that efficiently coded a large set of images[8]. Visual numerosity, a high-order feature, was progressively extracted across hidden layers, and it was coded invariantly from other visual properties only in the deepest layer of a hierarchical generative model[9]. The emergent monotonic encoding is consistent with single-cell recordings in monkey LIP[10] and functional magnetic resonance imaging blood oxygen level–dependent (BOLD) modulation in the human homolog of LIP[14]. The model computed numerosity through the combination of local computations and a simple global image statistic (cumulative area), without explicit individuation and size normalization of visual objects (compare refs. 6,15). The numerosity detectors were spatially selective, which is consistent with the properties of LIP neurons[10] and with numerosity adaptation[4]. Thus, local visual numerosities are invariants that can support various numerosity-related estimates, and they form the basis of a "visual sense of number". Though the adequacy of the proposed neural mechanism should be further tested in new behavioral and neurophysiological studies, its relative simplicity fits well with the long phylogenetic history of numerosity estimation[1]. Future studies should also assess whether sensitivity to numerosity can emerge when this dimension is a less salient stimulus feature in the training data, such as in natural images. One overarching implication of our findings is that learning a hierarchical generative model was the key to understanding the neural mechanism underlying numerosity perception and thus to bridging the gap between neurons and behavior.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

This study was supported by the European Research Council (grant no. 210922 to M.Z.)

AUTHOR CONTRIBUTIONS

M.Z. and I.S. conceived the experiments, discussed the results and wrote the paper. I.S. wrote the code, ran the model and analyzed data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Nieder, A. *Nat. Rev. Neurosci.* **6**, 177–190 (2005).
2. Halberda, J., Mazzocco, M.M.M. & Feigenson, L. *Nature* **455**, 665–668 (2008).
3. Piazza, M. *et al. Cognition* **116**, 33–41 (2010).
4. Burr, D. & Ross, J. *Curr. Biol.* **18**, 425–428 (2008).
5. Durgin, F.H. *Curr. Biol.* **18**, R855–R856 (2008).
6. Dehaene, S. & Changeux, J. *J. Cogn. Neurosci.* **5**, 390–407 (1993).
7. Allik, J. & Tuulmets, T. *Percept. Psychophys.* **49**, 303–314 (1991).
8. Hinton, G.E. & Salakhutdinov, R.R. *Science* **313**, 504–507 (2006).
9. Hinton, G.E. *Trends Cogn. Sci.* **11**, 428–434 (2007).
10. Roitman, J.D., Brannon, E. & Platt, M. *PLoS Biol.* **5**, e208 (2007).
11. Piazza, M. *et al. Neuron* **44**, 547–555 (2004).
12. Brannon, E.M. & Terrace, H.S. *Science* **282**, 746–749 (1998).
13. Nieder, A., Freedman, D. & Miller, E.K. *Science* **297**, 1708–1711 (2002).
14. Santens, S. *et al. Cereb. Cortex* **20**, 77–88 (2010).
15. Verguts, T. & Fias, W. *J. Cogn. Neurosci.* **16**, 1493–1504 (2004).