# Ultra-fast categorization of images containing animals *in vivo* and *in computo*

**Jean-Nicolas JEREMIE, Laurent PERRINET**

Institut de Neurosciences de la Timone

## Ultra-fast image categorization

A well-known task in the in the study of vision is the categorization of an animal in a scene. Applied to arbitrary natural scenes, it constitutes a challenging problem due to the large variations in shape, pose, size, texture, and position of animals that could be present in the scene. Yet, biological visual systems are able to perform such detection [1] in a very short amount of time as the scene is flashed to the subject [2] and robustly to geometrical tranforms [3]. It was shown that a feedforward architecture may be enough to perform such task Serre *et al.* [4].

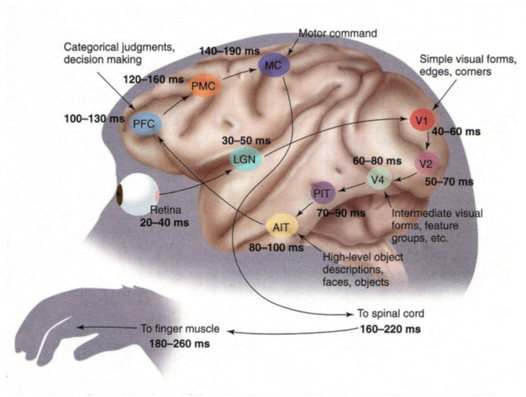


Figure 1: Hierarchy and latency of visual areas involved in ultra-rapid image categorization (from [5]).

## The dataset maker

The label's set of the ImageNet dataset is based on a large semantical database of English words: WordNet. The nouns, verbs, adjectives, and adverbs in this database are grouped into a graphical set of cognitive synonyms (synsets), each expressing a distinct concept.

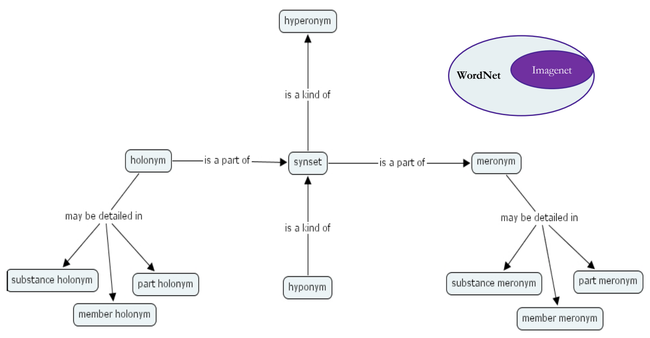


Figure 2: Definitions of the connection between synsets in Wordnet [6]

I used the hyperonym link (a synset is a kind of another synset) to select a specific subset of labels in the ImageNet dataset. I coded a Dataset Maker library based on this relationship with the "animal" synset. As a control we also defined independently an "artifact" synset.

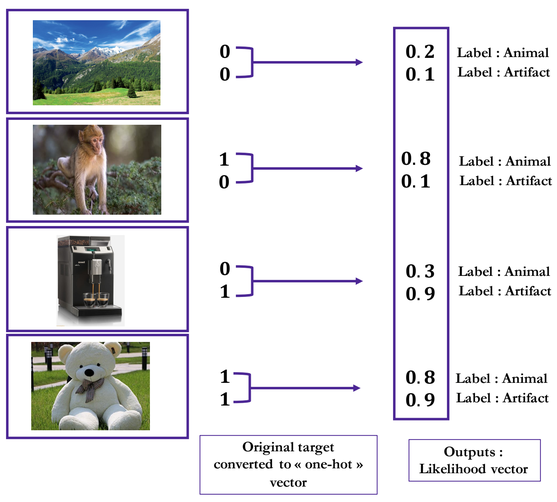


Figure 3: Outputs of the network with two independent labels. The network is trained for each supervision pair using the binary cross entropy loss.

## Transfer learning

Transfer Learning is a method which uses an existing DCNN network pre-trained on a specific task (such as VGG16 trained on ImageNet) and that modifies this network by re-training a subset of its weights on a different task. Here, I retrained the "head" of the network for the two independent categorization tasks defined in the dataset maker ("animal?" and "artifact?").

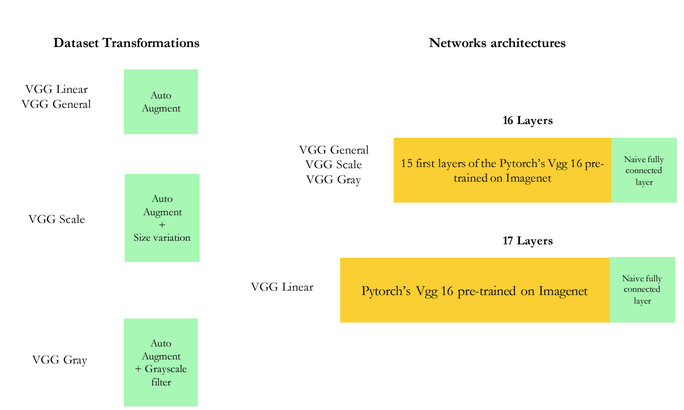


Figure 4: Transfer Learning strategies

As a control, I tested the networks on the dataset of Serre *et al.* [4]. This contains a total of 600 targets (images containing an animal) and 600 distractors (images not containing an animal). The networks obtain an accuracy on the "animal" synset similar to those found in the model and neurophysiological data of Serre *et al.* [4] (about 83%).

| | VGG General | VGG Linear | VGG Scale | VGG gray |
|---|---|---|---|---|
| **Imagenet** | **0.974** | 0.969 | 0.946 | 0.968 |
| **Serre** | 0.842 | 0.836 | **0.852** | 0.849 |


Figure 5: Accuracy of Transfer Learning accuracy for two datasets : Imagenet Datasetmaker and that of Serre *et al.* [4].

## "If it is animal, it is not artifact"

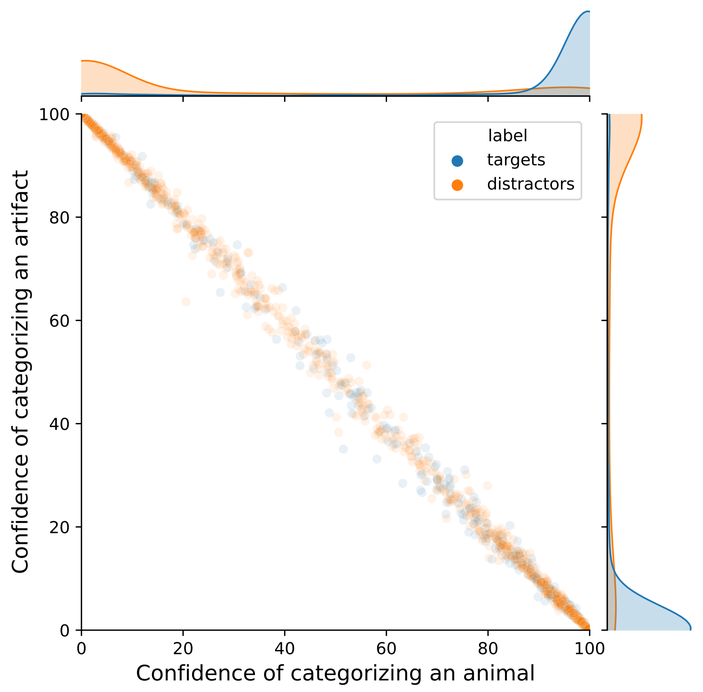


Figure 6: Application of the re-trained networks to the dataset of Serre *et al.* [4]. Here, by exposing the predictions for the "animal" and "artifact" labels, we highlight a bias in the composition of the dataset. Although the outputs are independent, "animal" images confidently correspond to "non-artifact" images (and *vice versa*), thus facilitating the overall detection.

## Geometrical transforms

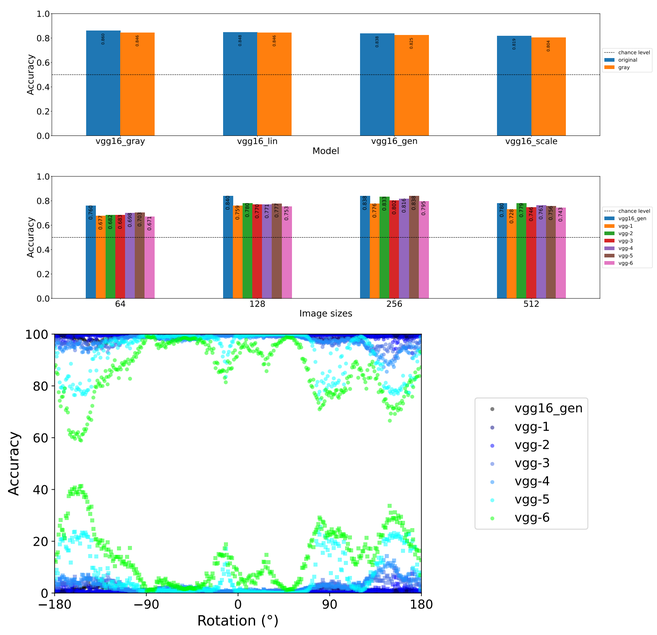


Figure 7: Robustness of the categorization to different image transformations in the models are similar to that reported in psychophysics [3]. We modeled different levels of complexity by pruning the network ('vgg16_lin') in the feature layer, from one block of layers ('vgg16_1') to 6 ('vgg16_6'). (Top) Use of grayscale images; (Middle) Changing the resolution; (Bottom) Rotating the images. Such experiment demonstrate that low-level features may be sufficient to categorize animals [7].

## Perspectives

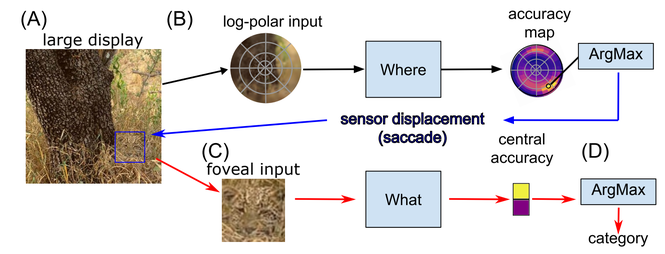


Figure 8: A dual-pathway model implementing saccades which could be extended from digits images (MNIST) [8] to natural images.

In an endeavour to push the limits of biologically inspired computer vision algorithms [9], one of my next challenges will be to infer target location simultaneously with target detection. To solve this problem, the mammalian brain seems to exploit two different visual pathways. Based on this principle, a dual-pathway artificial neuron architecture has recently been proposed and applied to simple digit images. Building on this work and using the deep learning tools developed here, we aim to extend this application to generic cognitive tasks with arbitrary natural scenes.

## References

1. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *nature* **381,** 520–522 (1996).
2. Kirchner, H. & Thorpe, S. J. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research* **46,** 1762–1776. doi:10.1016/j.visres.2005.10.002 (May 2006).
3. Rousselet, G. A., Macé, M. J.-M. & Fabre-Thorpe, M. Is It an Animal? Is It a Human Face? Fast Processing in Upright and Inverted Natural Scenes. *Journal of Vision* **3,** 440–455 (2003).
4. Serre, T., Oliva, A. & Poggio, T. A Feedforward Architecture Accounts for Rapid Categorization. *Proceedings of the National Academy of Sciences* **104,** 6424–6429. doi:10.1073/pnas.0700622104 (Apr. 2007).
5. Thorpe, S. J. & Fabre-Thorpe, M. Neuroscience. Seeking Categories in the Brain. *Science (New York, N.Y.)* **291,** 260–263. doi:10.1126/science.1058249 (Jan. 2001).
6. Fellbaum, C. *WordNet: An Electronic Lexical Database.* 449 pp. (A Bradford Book, Cambridge, MA, USA, May 22, 1998).
7. Perrinet, L. U. & Bednar, J. A. Edge Co-Occurrences Can Account for Rapid Categorization of Natural versus Animal Images. *Scientific Reports* **5,** 11400. doi:10.1038/srep11400 (Sept. 2015).
8. Daucé, E., Albigès, P. & Perrinet, L. U. A dual foveal-peripheral visual processing model implements efficient saccade selection. *Journal of Vision* **20,** 22–22. doi:10.1167/jov.20.8.22 (June 5, 2020).
9. .*Biologically Inspired Computer Vision.* (eds Cristóbal, G., Perrinet, L. U. & Keil, M. S.) doi:10.1002/9783527680863 (Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim, Germany, Oct. 7, 2015).