A Network Model for Multiple Object Scales in Cluttered Backgrounds: Multiple Patches per Object

Xiaoying Song*, Wenqiang Zhang*, Juyang Weng†*
*School of Computer Science, Fudan University, Shanghai, 201203 China
†Department of Computer Science and Engineering, Cognitive Science Program, Neuroscience Program,
Michigan State University, East Lansing, MI, 48824 USA
Email: 09210240073@fudan.edu.cn, wqzhang@fudan.edu.cn, weng@cse.msu.edu

***Abstract*—WWN-5 was designed for objects of different scales. In this work, we study how different features of different scales at different locations can be shared by different objects of different scales in WWN-5— allowing each object to use multiple patches.**

Index Terms - object recognition, scale invariance, cluttered background, neural networks.

How does the brain deal with objects of different scales present in cluttered natural backgrounds? One approach is to search for a given object appearance pattern from the input image using different scales. Cresceptron 1997 [4] and many other methods later used large-to-small scale-based search over the entire image for each object learned. This requires that the model of an object to be detected and recognized is pre-given. When the number of objects is large, this approach becomes intractably expensive.

Inspired by the brain-anatomy, we propose a "brain-like" network model that develops distributed emergent representations inside a network where different concepts (e.g., types, scales, and locations) can mutually collaborate and compete in parallel. For example, at nearby locations it is impossible to detect objects of similar scales even of different types since they are not consistent with experience.

Where-What Network (WWN) [1], [3], [2] is a class of networks to model not only the brain's dorsal pathway (where or how information) and ventral pathway (what information) for vision, but also their development through experiences. WWN-1 and WWN-4 were meant for object of a single scale. WWN-5 was designed for objects of different scales. In this work, we study how different features of different scales at different locations can be shared by different objects of different scales in WWN-5— allowing each object to use multiple patches.

Suppose that a sufficient number of neurons are available in multiple internal areas, so that object appearances of different scales at different image location are sufficiently sampled by different neuronal areas. During training, a hypothetical brain area called pulvinar supplies the rough location and scale, but not the type, so that neurons whose receptive fields are away from the object of interest are suppressed, being prevented from firing and updating. Near the pulvinar supervised location and scale, neurons are allowed to fire and compete.

To enable the system to learn potentially a wide variety of objects, we do not hand craft any fixed feature types (e.g. edge or Gabor features). The features in the WWN-5 are all automatically developed (emerge from experience) based on the theory of Lobe Component Analysis (LCA) which aims to develop dually optimal representations for the high-dimensioned neuronal inputs, in the sense of space (least-mean error in representation) and time (least-distance toward the best representation at any time after "birth" from to a limited number of training samples). Inside WWN-5, each internal area Y_i has a different range of default receptive field. The more sensory areas the WWN-5 has, potentially better the scales are sampled. In this work, we study how local features in early areas can help the later areas so that object recognition is not based on a rigid template, but rather, an hierarchy of responses with increasing tolerance to distortions. We report experimental results to show how WWN-5 deals with a variety of learned concepts: location, type, and scale in cluttered backgrounds.

Although the type concept here in the experiments is about the object category, in principle the type can represent other concepts, such as living things vs. nonliving things, viewing angle, lighting direction, usage, price, etc. However, such wider capabilities have yet to be experimentally demonstrated.

References

[1] Z. Ji, J. Weng, and D. Prokhorov. Where-what network 1: "Where" and "What" assist each other through top-down connections. In *Proc. IEEE International Conference on Development and Learning*, pages 61–66, Monterey, CA, Aug. 9-12 2008.

[2] M. Luciw and J. Weng. Where What Network 3: Developmental top-down attention with multiple meaningful foregrounds. In *Proc. IEEE International Joint Conference on Neural Networks*, pages 4233–4240, Barcelona, Spain, July 18-23 2010.

[3] M. Luciw and J. Weng. Where What Network 4: The effect of multiple internal areas. In *2010 IEEE 9th international conference on development and learning*, pages 311–316, 2010.

[4] J. Weng, N. Ahuja, and T. S. Huang. Learning recognition and segmentation using the Cresceptron. *International Journal of Computer Vision*, 25(2):109–143, Nov. 1997.