On Fast Deep Nets for AGI Vision

Jürgen Schmidhuber, Dan Cireşan, Ueli Meier, Jonathan Masci, and Alex Graves

The Swiss AI Lab IDSIA
University of Lugano & SUPSI, Switzerland

Abstract. Artificial General Intelligence will not be general without computer vision. Biologically inspired adaptive vision models have started to outperform traditional pre-programmed methods: our fast deep / recurrent neural networks recently collected a string of 1st ranks in many important visual pattern recognition benchmarks: IJCNN traffic sign competition, NORB, CIFAR10, MNIST, three ICDAR handwriting competitions. We greatly profit from recent advances in computing hardware, complementing recent progress in the AGI theory of mathematically optimal universal problem solvers.

Keywords: AGI, Fast Deep Neural Nets, Computer Vision, Hardware Advances vs Theoretical Progress.

1 Introduction

Computer vision is becoming essential for thousands of practical applications. For example, the future of search engines lies in image and video recognition as opposed to traditional text search. Autonomous robots such as driverless cars depend on vision, too. Generally speaking, the "G" in "AGI" will be undeserved without excellent computer vision.

AGI research is currently driven by two types of progress. On the one hand, the new millennium brought the first universal problem solvers [10, 21] that are theoretically optimal in asymptotic and other senses, putting AGI research on a sound mathematical footing for the first time, although such approaches are currently not yet practically feasible. On the other hand, due to ongoing hardware advances, the computing power per Swiss Franc is still growing by a factor of 100-1000 per decade, greatly increasing the practical feasibility of less general methods invented in the previous millennium. This paper reflects the second type of progress, exploiting graphics cards or GPUs (mini-supercomputers normally used for video games) which are 100 times faster than today's CPU cores, and a million times faster than PCs of 20 years ago, to train biologically plausible deep neural nets on vision tasks.

Excellent object recognition results illustrate the benefits of this pragmatic approach. As of January 2011, our neural computer vision team has collected a string of 1st ranks in many important and highly competitive international visual pattern recognition benchmarks.

1. IJCNN's online Traffic Sign Recognition Benchmark (1st & 2nd rank; 1.02% error rate), January 2011 [4].

J. Schmidhuber, K.R. Thórisson, and M. Looks (Eds.): AGI 2011, LNAI 6830, pp. 243–246, 2011.

2. NORB data set, NY University, 2004 [13]. Our team set the new record (2.53% error rate) in February 2011 [3].
3. CIFAR-10 data set of Univ. Toronto, 2009 [11]. Our team set the new record (19.51% error rate) in 2011 [3].
4. MNIST data set of NY University, 1998 [12]. Our team set the new record (0.35% error rate) in 2010 [2], and tied it again in January 2011 [3].
5. Three Handwriting Recognition Competitions at ICDAR 2009, all won by our multi-dimensional LSTM recurrent neural networks trained by *Connectionist Temporal Classification* (CTC) [7, 8]: Arabic Handwriting Competition of Univ. Braunschweig, Handwritten Farsi/Arabic Character Recognition Competition, French Handwriting Competition based on data from the RIMES campaign.

Remarkably, none of the above requires the traditional sophisticated computer vision techniques developed over the past six decades or so. Instead, our biologically rather plausible systems are inspired by human brains, and learn to recognize objects from numerous training examples. We use supervised, artificial, feedforward or recurrent [9, 7, 8] (deep by nature) neural networks with many non-linear processing stages, partially inspired by early hierarchical neural systems such as Fukushima's Neocognitron [5]. We sometimes (but not always) profit from sparse network connectivity and techniques such as weight sharing & convolution [12, 1, 25], max-pooling [17], and contrast enhancement [6] like the one automatically generated by unsupervised *Predictability Minimization* [18, 22, 24].

2 Neural Network ReNNaissance

Our NNs are now outperforming all other methods including the theoretically less general and less powerful support vector machines based on statistical learning theory [27] (which for a long time had the upper hand, at least in practice). Such results are currently contributing to a second *Neural Network ReNNaissance* (the first one happened in the 1980s and early 90s).

3 Outlook

The methods discussed above are passive learners - they do not learn to actively search for the most informative image parts. Humans, however, use sequential gaze shifts for pattern recognition. This can be more efficient than the fully parallel one-shot approach. That's why we intend to combine the fast deep / recurrent nets above with variants of what to our knowledge was the first artificial fovea sequentially steered by a learning neural controller [23], using a variant of reinforcement learning to create saccades and find targets in a visual scene.

4 Conclusion

The first decades of attempts at AGI have been dominated by heuristic approaches, e.g., [15, 16, 26, 14]. In recent years things have changed, however. The new millennium brought the first mathematically sound, asymptotically optimal, universal problem solvers, providing a new, rigorous foundation for the previously largely heuristic field of General AI and embedded cognitive agents, identifying the limits of both human and artificial intelligence, and providing a yardstick for any future approach to general cognitive systems [19, 10, 20]. The field is indeed becoming a real formal science.

On the other hand, however, one cannot dispute the significance of hardware progress on the road to practical AGI, as illustrated by our recent practical successes mentioned in this paper, achieved by methods which are combinations of algorithms mostly developed in the previous millennium, but greatly profiting from dramatic advances in computational power per Swiss Franc, obtained in the new millennium.

We are confident that theory and practice will converge in the not-so-distant future.

Acknowledgements. This work was funded in part by EU project 270247, "A Neurodynamic Framework for Cognitive Robotics: Scene Representations, Behavioural Sequences, and Learning," and by SNF grant 200021-111968.

References

[1] Behnke, S.: Hierarchical Neural Networks for Image Interpretation. LNCS, vol. 2766. Springer, Heidelberg (2003)
[2] Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Deep big simple neural nets for handwritten digit recogntion. Neural Computation 22(12), 3207–3220 (2010)
[3] Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J.: High-performance neural networks for visual object classification. arxiv 1102.0183 (2011)
[4] Ciresan, D.C., Meier, U., Masci, J., Schmidhuber, J.: A committee of neural networks for traffic sign classification. In: International Joint Conference on Neural Networks to appear (2011)
[5] Fukushima, K.: Neocognitron: A self-organizing neural network for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36(4), 193–202 (1980)
[6] Fukushima, K.: Neocognitron for handwritten digit recognition. Neurocomputing 51, 161–180 (2003)
[7] Graves, A., Fernández, S., Schmidhuber, J.: Multi-dimensional recurrent neural networks. In: Proceedings of the 17th International Conference on Artificial Neural Networks (September 2007)
[8] Graves, A., Schmidhuber, J.: Offline handwriting recognition with multidimensional recurrent neural networks. In: Advances in Neural Information Processing Systems, vol. 21, MIT Press, Cambridge (2009)
[9] Hochreiter, S., Schmidhuber, J.: Flat minima. Neural Computation 9(1), 1–42 (1997)
[10] Hutter, M.: Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability. Springer, Berlin (2004); (On J. Schmidhuber's SNF grant 20-61847)
[11] Krizhevsky, A.: Learning multiple layers of features from tiny images. Master's thesis, Computer Science Department, University of Toronto (2009)
[12] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)
[13] LeCun, Y., Huang, F.J., Bottou, L.: Learning methods for generic object recognition with invariance to pose and lighting. In: Proc. of Computer Vision and Pattern Recognition Conference (2004)
[14] Mitchell, T.: Machine Learning. McGraw Hill, New York (1997)
[15] Newell, A., Simon, H.: GPS, a program that simulates human thought. In: Feigenbaum, E., Feldman, J. (eds.) Computers and Thought, pp. 279–293. McGraw-Hill, New York (1963)

[16] Rosenbloom, P.S., Laird, J.E., Newell, A.: The SOAR Papers. MIT Press, Cambridge (1993)
[17] Scherer, D., Müller, A., Behnke, S.: Evaluation of pooling operations in convolutional architectures for object recognition. In: International Conference on Artificial Neural Networks (2010)
[18] Schmidhuber, J.: Learning factorial codes by predictability minimization. Neural Computation 4(6), 863–879 (1992)
[19] Schmidhuber, J.: The new AI: General & sound & relevant for physics. In: Goertzel, B., Pennachin, C. (eds.) Artificial General Intelligence, pp. 175–198. Springer, Heidelberg (2006); also available as TR IDSIA-04-03, arXiv:cs.AI/0302012
[20] Schmidhuber, J.: New millennium AI and the convergence of history. In: Duch, W., Mandziuk, J. (eds.) Challenges to Computational Intelligence. Studies in Computational Intelligence, vol. 63, pp. 15–36. Springer, Heidelberg (2007); arXiv:cs.AI/0606081
[21] Schmidhuber, J.: Ultimate cognition *à la* Gödel. Cognitive Computation 1(2), 177–193 (2009)
[22] Schmidhuber, J., Eldracher, M., Foltin, B.: Semilinear predictability minimization produces well-known feature detectors. Neural Computation 8(4), 773–786 (1996)
[23] Schmidhuber, J., Huber, R.: Learning to generate artificial fovea trajectories for target detection. International Journal of Neural Systems 2(1 & 2), 135–141 (1991)
[24] Schraudolph, N.N., Eldracher, M., Schmidhuber, J.: Processing images by semi-linear predictability minimization. Network: Computation in Neural Systems 10(2), 133–169 (1999)
[25] Simard, P., Steinkraus, D., Platt, J.: Best practices for convolutional neural networks applied to visual document analysis. In: Seventh International Conference on Document Analysis and Recognition, pp. 958–963 (2003)
[26] Utgoff, P.: Shift of bias for inductive concept learning. In: Michalski, R., Carbonell, J., Mitchell, T. (eds.) Machine Learning, vol. 2, pp. 163–190. Morgan Kaufmann, Los Altos (1986)
[27] Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)