Logical Vision: Meta-Interpretive Learning for Human-Like Vision

Wang-Zhou Dai[1], Stephen Muggleton[2],
Alireza Tamaddoni-Nezhad[2] and Zhi-Hua Zhou[1]

[1] National Key Laboratory for Novel Software Technology, Nanjing University
[2] Department of Computing, Imperial College London

Abstract. Progress in statistical learning in recent years has enabled computers to recognize objects with near-human ability. However, recent studies have revealed particular drawbacks in current computer vision systems which suggest there exist considerable differences between the way these systems function compared with human visual cognition. We are investigating a framework referred to as Logical Vision which is demonstrated on learning visual concepts constructively and symbolically. It first constructively extracts logical facts of mid-level features, then generative Meta-Interpretive Learning technique is applied to learn high-level notions. Experiments conducted on learning simple shapes (e.g. polygons) demonstrated that this technique outperforms some of existing object recognition methods based on statistical machine learning. We are now investigating methods for extending these initial experiments to higher-level inference from real-world images and videos (e.g. microscopic videos of bacteria).

1 Introduction

Present Computer Vision approaches are mainly based on statistical analysis of digital images [7]. For example, state-of-the-art methods allow identification of surface description and depth and also simple object recognition (e.g. using Neural Networks). However, these techniques fail to cope with high-level visual analysis and are unable to account for human-like vision (e.g. partial occlusion, light source identification and shadow prediction) and higher-level inference (e.g. intention of agents, properties of objects not directly observed within image). Deep Neural Networks (DNNs) [6, 2] have demonstrated impressive and state-of-the-art results on many pattern recognition tasks, especially image classification problems [8]. However, recent studies revealed some major differences between statistics-based computer vision systems and human visual cognition [1, 14]. For example, it is easy to produce images that are completely unrecognizable to humans, though state-of-the-art visual learning algorithms believe them to be recognizable objects with over 99% confidence [1]. Moreover, humans can typically learn from a single visual example [9], unlike statistical learning which depends on hundreds or thousands of images. Humans achieve this ability using background knowledge, which plays a critical role. By contrast, statistics-based computer vision algorithms have no general mechanisms for incorporating background knowledge. In this paper we consider a novel visual concept learning framework, called *Logical Vision* [4] which uses background knowledge on mid-level symbols to guide the sampling of low-level features. A generalized Meta-Interpretive Learning (MIL) [11] is used then

Table 1. Predictive accuracy of learning simple geometrical shapes on single object datasets.

ACC	tri	quad	pen	hex	reg	r_tri
HOG	0.83 ± 0.04	0.76 ± 0.01	0.73 ± 0.03	0.75 ± 0.07	0.63 ± 0.08	0.74 ± 0.04
dense-SIFT	0.82 ± 0.05	0.66 ± 0.06	0.64 ± 0.04	0.71 ± 0.03	0.71 ± 0.05	0.77 ± 0.07
LBP	0.87 ± 0.05	0.69 ± 0.04	0.67 ± 0.03	0.73 ± 0.03	0.65 ± 0.05	0.75 ± 0.05
CNN	0.91 ± 0.01	0.75 ± 0.00	0.75 ± 0.00	0.84 ± 0.02	0.59 ± 0.06	0.85 ± 0.04
C+d+L	0.82 ± 0.01	0.75 ± 0.00	0.76 ± 0.01	0.76 ± 0.01	0.64 ± 0.05	0.80 ± 0.04
LV$_{Poly}$	$\mathbf{1.00 \pm 0.00}$	$\mathbf{0.99 \pm 0.01}$	$\mathbf{1.00 \pm 0.00}$	$\mathbf{0.99 \pm 0.01}$	$\mathbf{1.00 \pm 0.00}$	$\mathbf{1.00 \pm 0.00}$

to learn high-level visual concepts. MIL also enhances the constructive paradigm of Logical Vision through its ability to learn recursive theories, inventing predicates and learning from a single example.

2 The proposed framework

The input for Logical Vision consists of a set of geometrical primitives B_P, one or a set of images $\mathcal{I}$ as background knowledge, and a set of logic facts E representing the examples as the target visual concepts. The task is to learn a hypothesis H that defines the target visual concept where $B_P, \mathcal{I}, H \models E$. The purpose of mid-level features extraction is to obtain necessary logical facts B_A representing mid-level features of $I \in \mathcal{I}$. This procedure is realized by repeatedly executing a "conjecturing and sampling" procedure which uses the mid-level feature conjectures to guide the sampling of low-level features. The resulting features are then used to revise previously constructed conjectures. After obtaining mid-level features B_A, Logical Vision uses a generalized Meta-Interpretive Learner to learn target visual concepts. The input of generalized Meta-Interpretive Learning (MIL) [11] consists of a generalized Meta-Interpreter B_M and domain specific primitives B_P together with two sets of ground atoms as background knowledge B_A and examples E respectively. The output of MIL is a revised form of the background knowledge containing the original background knowledge B_A, domain specific primitives B_P augmented with additional ground atoms representing a hypothesis H.

3 Experiments

Table 1 compares the predictive accuracies of an implementation of Logical Vision (LV_{Poly}) versus several statistics-based computer vision algorithms on the task of learning simple geometrical concepts. We used a popular statistics-based computer vision toolbox VLFeat [15] to implement the statistical learning algorithms. The experiments are carried with different kinds of features. Because the sizes of datasets are small, we used support vector machine (libSVM [3]) as classifier. The parameters are selected by 5-fold cross-validation. The features we have used in the experiments are as follows: **HOG**, Histogram of Oriented Gradients [5], **Dense-SIFT**, Scale Invariant Feature Transform [10], **LBP**, Local Binary Pattern [12], **CNN**, Convolutional Neural Network (CNN) [13]. We also compare with a combinations of above feature sets (i.e. **C+d+L**).

4 Conclusion and further works

By using the proposed Logical Vision approach, we were able to extract logical facts of mid-level features and learn high-level visual concepts from images constructively and symbolically. The experimental results showed the advantage of the proposed framework compared to traditional computer vision learning methods. We are currently applying Logical Vision for the task of microscopic video/image analysis. The goal of this project is to learn high-level descriptions from microscopic videos, e.g. hypotheses about the movment and interactions of bacteria.

References

1. Ahn, N., Yosinski, J., Clune, J.: Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA (2015)
2. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning 2(1), 1–127 (2009)
3. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
4. Dai, W.Z., Muggleton, S., Zhou, Z.H.: Logical Vision: Meta-interpretive learning for simple geometrical concepts. In: Short Paper Proceedings of the 25th International Conference on Inductive Logic Programming. National Institute of Informatics, Tokyo (2015)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the 13rd IEEE Conference on Computer Vision and Pattern Recognition. pp. 886–893. San Diego, CA (2005)
6. Hinton, G.E.: Learning multiple layers of representation. Trends in Cognitive Sciences 11(10), 428–434 (2007)
7. Krig, S.: Computer Vision Metrics: Survey, Taxonomy, and Analysis. Apress, Berkely, CA (2014)
8. Krizhevsky, A., Sutskever, I., E. Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc. (2012)
9. Lake, B.M., Salakhutdinov, R., Gross, J., Tenenbaum, J.B.: One shot learning of simple visual concepts. In: Proceedings of the 33rd Annual Conference of the Cognitive Science Society. pp. 2568–2573 (2011)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
11. Muggleton, S., Lin, D., Tamaddoni-Nezhad, A.: Meta-interpretive learning of higher-order dyadic datalog: Predicate invention revisited. Machine Learning 100(1), 49–73 (2015)
12. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA (2015)
14. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I.J., Fergus, R.: Intriguing properties of neural networks. CoRR abs/1312.6199 (2013)
15. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2008)

.