Closed loop vision

Zhibin Liao, Supervised by
Dr. Gustavo Carneiro and Dr. Tat-Jun Chin

Introduction

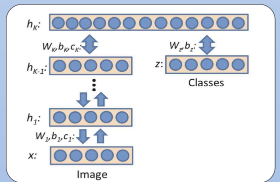
In machine learning, pure feed-forward visual classification approaches consisting of recognition tasks of increasing complexity. The main problem of the feed-forward approach is lack of feedback mechanism which makes it hard to recover from errors.

This project is aimed at study a new closed loop inference approach consisting of a feedback feed-forward mechanism using deep belief networks (DBN). Experiments performed indicate our closed loop inference show promising improvements compare to pure feed-forward DBN classifier.

Deep belief networks

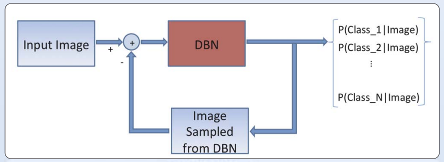
The main advantages of building closed loop inference on DBN :

1) DBN possesses the ability to process images using both feed-forward and feedback mechanisms.
2) The pair-wise layers are restricted boltzmann machines (RBM), which makes the number of layers flexible to change and training process easy to monitor.



Closed loop inference

Closed loop inference requires a trained DBN as foundation. Generally it estimates the posterior probability from inputs and then recover image from posterior probability . The difference between original image and recovered image are fed back to DBN in an iterative process that runs until convergence.



The major steps of closed loop inference are:

1) Building a **TRANSITION MATRIX** that summaries the posterior distributions that obtained from training set. It outlines the general likelihoods of labeled images to be classed as one other. Noise is added to prevent over-confidence.
2) The **PREDICTION** step computes the posterior probability with the noised transition matrix.

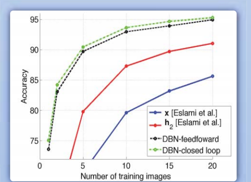
Algorithm Closed loop inference

1: Set $\mathbf{x}^{(0)} = \mathbf{x} \in [0,1]^{M \times M}$, representing the input image.
2: **Transition matrix:** compute transition matrix, representing $p(y = i | y = j)$ (for $i, j \in \{1, ..., Z\}$).
3: **for** t = 1:T **do**
4: **Prediction:** compute the prediction distribution $p(y^{(t)} | \{\mathbf{x}^{(l)}\}_{l=0}^{t-1})$.
5: **Data association:** search for the image that maximizes $p(\mathbf{x}^{(t)} | \{\mathbf{x}^{(l)}\}_{l=0}^{t-1})$ by useing $\mathbf{x}^{(t)}$.
6: **Correction:** correct the current prediction by estimating $p(y^{(t)} | \{\mathbf{x}^{(l)}\}_{l=0}^{t})$.
7: **end for**
8: The class for image $\mathbf{x}$ is the one that maximizes.

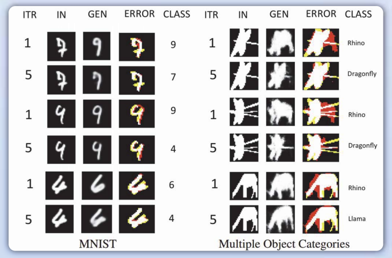
3) The **DATA ASSOCIATION** generates multiple samples of reconstruction images by performing Gibbs samplings. The difference between each sample image and original image is amplified in recovered image, and then feed-forward to compute posterior distributions. The reconstruction with the highest posterior probability gets selected.
4) The **CORRECTION** step takes the prediction in 2) and selected reconstruction in 3) to compute to fix the class prediction.

Experiments

On MNIST dataset [1], we find our closed loop inference system obtains recognition improvements in the order of 5% to 10% with some of the runs improved more than 20%. The best error rate we get is 2.30% while the underlying feed-forward DBN classifier does 2.51%. We then trained an asymmetric DBN [3] and get more competitive results: **1.69% error rate by asymmetric DBN and 1.59% by our method.**



On Multiple Object Categories dataset [2], our results are significantly outperform the ones reported by Eslami et al. [2]. Our closed loop inference **reduces error of feed-forward analysis by 5% on average.** With 20 training images per class, Eslami et al. reported 90.8% accuracy. In comparison, we have 94.9% and 95.2% accuracies for feed-forward inference and closed loop inference respectively.



Conclusion

The presented closed loop inference shows improvements comparing to pure feed-forward inference. However, there are still issues to be investigated. E.g. the introduced noise in transition matrix does not appear to be as useful as we expected. Further improvements include performing parallel computing on GPU to improve efficiency.

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[2] S. A. Eslami, N. Heess, and J. Winn. The shape boltzmann machine: a strong model of object shape. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEEConference on, pages 406–413. IEEE, 2012.

[3] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural computation, 18(7):1527–1554, 2006.

[4] G. Carneiro, T. Chin, and Z. Liao. Closed loop vision. Submitted to ICCV 2013.

School of Computer Science
University of Adelaide SA, Australia