Generative ScatterNet Hybrid Deep Learning (G-SHDL) Network

Amarjot Singh and Nick Kingsbury

Signal Processing Group, University of Cambridge

Contributions

We propose a G-SHDL network inspired by the circuitry of the visual cortex. The advantages of the proposed network are:

- Computationally efficient hybrid architecture.
- Rapid learning of useful features with RBMs using structural priors.
- Advantages over unsupervised and supervised learning (small datasets).

G-SHDL Network

The G-SHDL network is composed of a handcrafted front-end (ScatterNet) which extracts invariant features from the input image that are then used by the Convolutional RBM (CRBM) (unsupervised) to **rapidly** learn hierarchical features. The features from the last layer of the RBM are used by the CRF (supervised) to achieve segmentation.

Rapid learning of CRBM

Convolutional RBM (CRBM) learns M filters of size $N_H \times N_H$ using Markov Chain Monte Carlo (MCMC) sampling which is computationally intensive. The learning is accelerated by initializing the RBM filters with PCA structural priors obtained as:

$$\min_{W} \|X - WW^T X\|_F^2, \; s.t. \; WW^T = I_M \quad (1)$$

where X are patches sampled from N the hand-crafted features, I_M is an identity matrix of size $N_H \times N_H$. The solution represents M leading principal eigenvectors of XX^T obtained using Eigen decomposition.

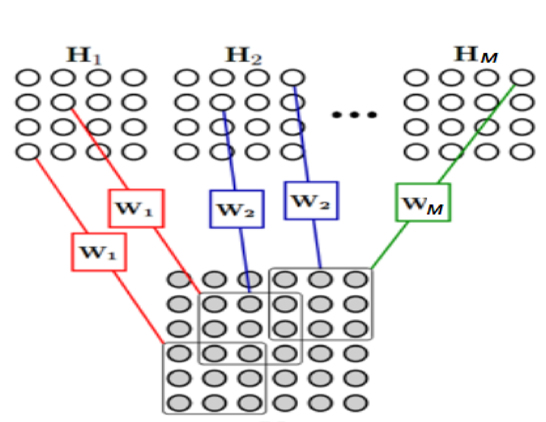

Figure 1: A CRBM which learns M filters: $W_1, ..., W_M$ from the visible units V and hidden units H. 4 stacked CRBM layers learn 200, 150, 100 and 50 filters with PCA structural priors.

G-SHDL Architecture, Optimization, and Segmentation

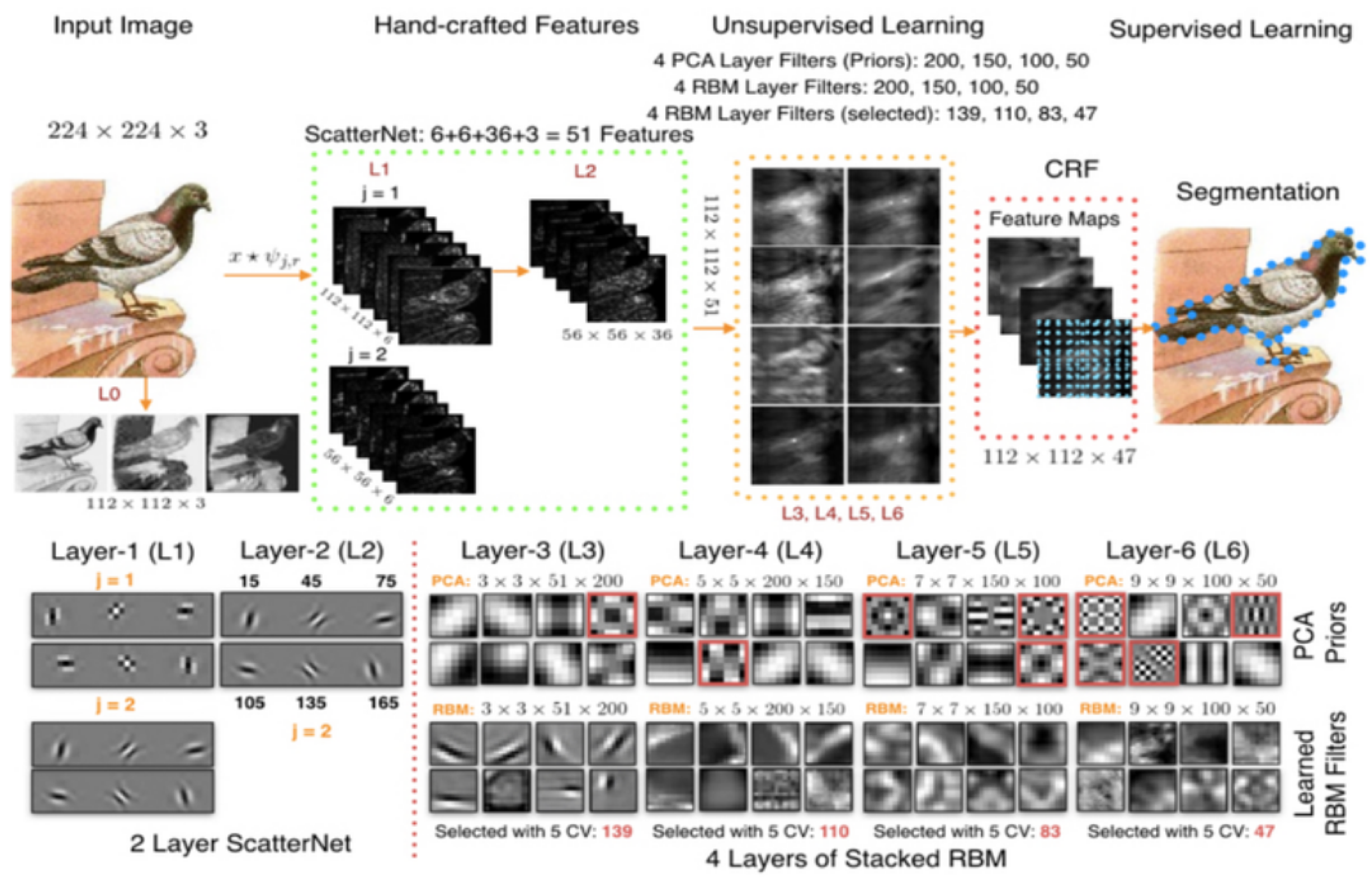

Figure 2: The figure presents the architecture for the G-SHDL network.

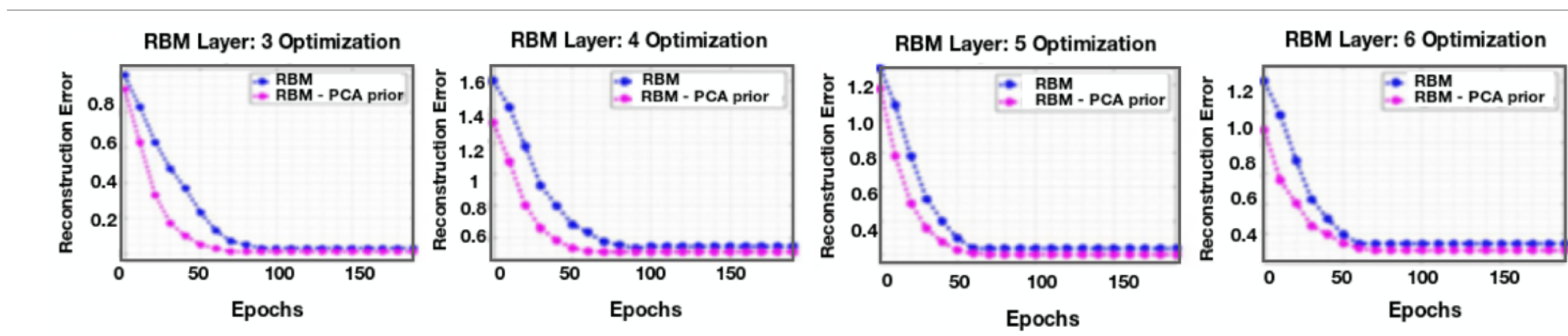

Figure 3: The figure shows the improvement in learning for RBMs at each layer with the use of PCA structural priors.

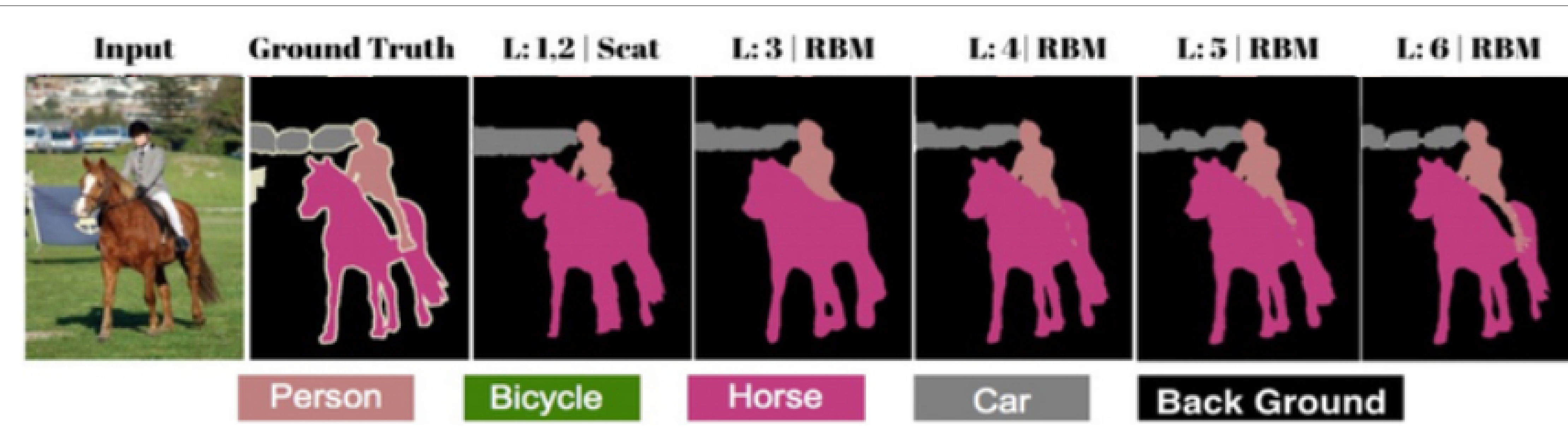

Figure 4: Figure shows an image (MSRC) with their ground truth and segmentation obtained at L2 to L6 of the G-SHDL network.

Results

The G-SHDL is applied on the Stanford Background (SB) and MSRC datasets for semantic segmentation. The optimal filters for the four stacked CRBM layers are selected using 5-CV as shown in Table 1.

Table 1. 5 fold cross validation performed on the training dataset of Stanford background (SB) dataset to select optimal filters for L3 to L6 RBM layers. L(size) = No. of Filters (a, a is equivalent to $a \times a$)

Filters	L3 (size)	43 (size)	L5 (size)	L6 (size)
PCA	200 (3,3)	150 (5,5)	100 (7,7)	50 (9,9)
RBM	200 (3,3)	150 (5,5)	100 (7,7)	50 (9,9)
Selected	139	110	83	47

The segmentation pixel accuracy (PA) for each layer of the G-SHDL network is shown in Table 2.

Table 2. PA (%) on SB dataset for each module computed with CRF. The increase in accuracy with the addition of each layer is also shown. HC: Hand-crafted. RBM Layers: L3, L4, L5 and L6.

Dataset	HC	L3	L4	L5	L6
Accuracy	68.4	72.3	74.8	76.7	**78.21**

The segmentation pixel accuracy (PA) is presented for both the dataset, SB and MSRC, in Table 3, with advantages over supervised learning in Table 4.

Table 3. PA (%) and comparison on both datasets. Unsup: Unsupervised, Semi: Semi-supervised and Sup: Supervised.

Dataset	G-SHDL	Semi	Unsup	Sup
SB	78.21	77.76	68.1	**80.2**
MSRC	83.90	83.6	74.7	**89.0**

Table 4. Comparison of G-SHDL on PA (%) with Recurrent CNN (rCNN) against different training dataset sizes on SB dataset.

Arch.	50	100	200	300	400	500	572
G-SHDL	**40.3**	**59.9**	**66.4**	**72.6**	75.7	78.20	78.21
rCNN	15.6	34.5	41.1	66.9	**76.2**	**79.87**	**80.2**

Conclusions

G-SHDL is a computationally efficient network architecture that learns useful features for semantic segmentation, rapidly (PCA priors) using unlabeled data. This network can be useful for applications with limited labeled datasets and computational resources.

References

1. A Singh and N G Kingsbury (2017), Scatternet Hybrid Deep learning (SHDL) Network For Object Classification, IEEE MLSP workshop 2017.