IMT SCHOOL FOR ADVANCED STUDIES LUCCA

ROTATION-INVARIANT RESTRICTED BOLTZMANN MACHINE USING SHARED GRADIENT FILTERS

THE UNIVERSITY OF EDINBURGH

Mario Valerio Giuffrida, Sotirios A. Tsaftaris

valerio.giuffrida@imtlucca.it, s.tsaftaris@ed.ac.uk

25th INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS (ICANN) 2016

Motivations

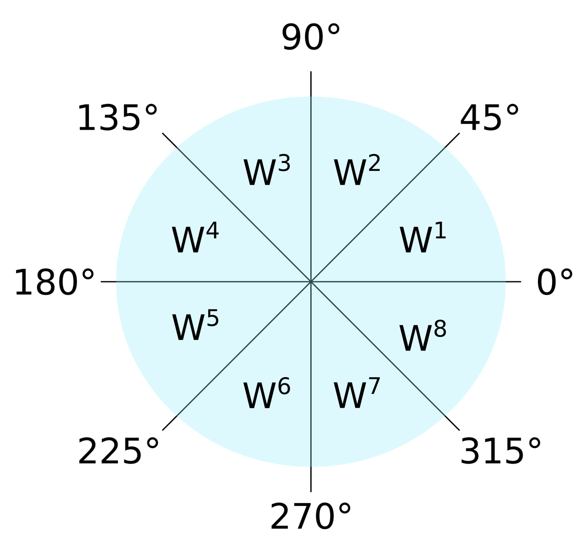
- Invariance allows describing transformed versions of an image with the same representation
- RBM cannot accommodate such variability in the dataset
- Current approaches deal with invariance using data augmentation
- Main problems are: propagation of nuisance due to pixel interpolation, unnecessary time and/or memory consumptions
- We present ERI-RBM, a model that can learn rotation invariance directly from data, without transforming the input image

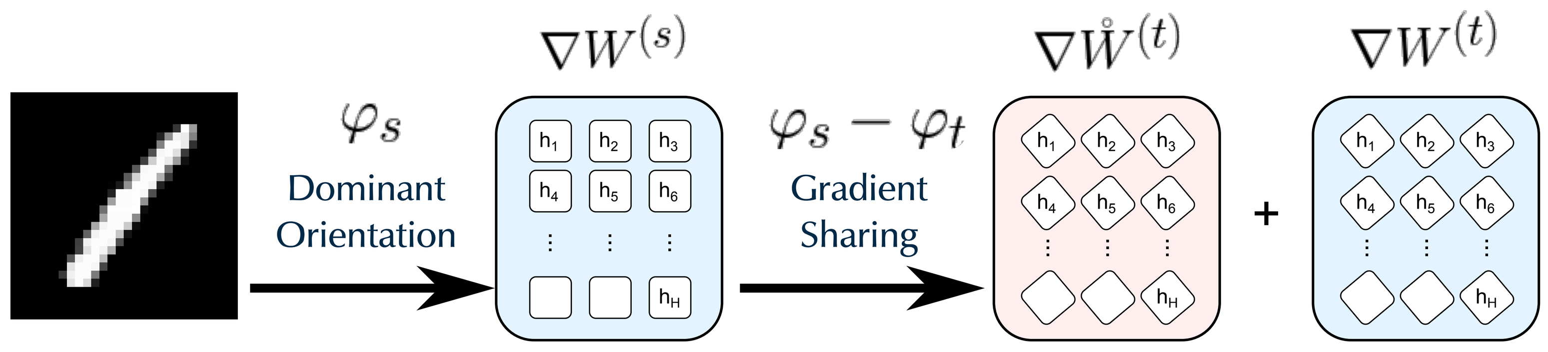
Contributions

- Learned representation with ERI-RBM is compact
- Invariance is achieved using a single-layer network
- We introduce a set of Weight Matrices that are associated with a dominant orientation of the images
- The contribution of each Weight Matrix is shared during the Contrastive Divergence, by rotating the learned filters by a suitable angle

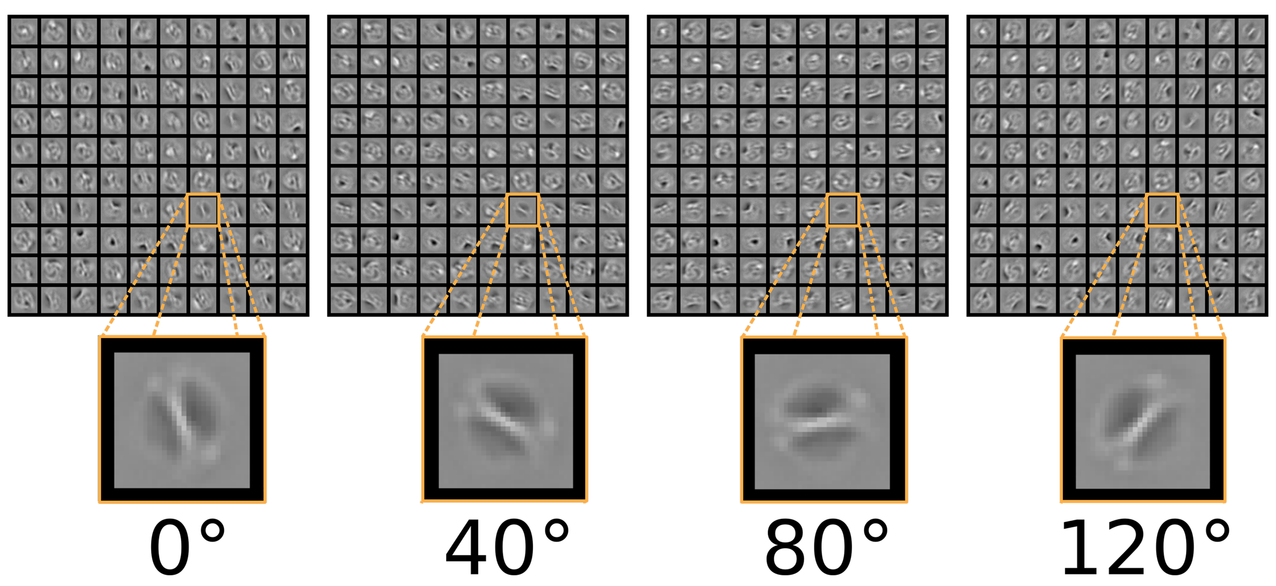
Explicit Rotation-Invariant Restricted Boltzmann Machine

- The interval [0,360[is split in S angles
- Dominant orientation of images is computed via histograms of gradients
- Each angle in S is associated with a different Weight Matrix
- Gradient computed for each image uses a specific Weight Matrix, determined by the dominant orientation
- Contribution provided by the gradient is shared amongst the other matrices by rotating the learned filters
- Rotated filters are added up to the corresponding Weight Matrix





Learned Filters (S=9)



The Revised Energy Function

$$E(v,h;s) = -h^T W^{(s)} v - c^T v - \left[b^{(s)}\right]^T h$$

Conditional Probabilities

$$p(v_j = 1|h;s) = \sigma\left(c_j + h^T W^{(s)}_{\bullet,j}\right) \quad p(h_k = 1|v;s) = \sigma\left(b^{(s)}_k + W^{(s)}_{k,\bullet} v\right)$$

The Shared Gradient Term

$$\theta = \varphi_s - \varphi_t$$

$$\nabla \mathring{W}^{(t)} = R_\theta(\nabla W^{(s)}) \equiv \begin{pmatrix} R_\theta\left(\nabla W^{(s)}_{1,\bullet}\right) \\ R_\theta\left(\nabla W^{(s)}_{2,\bullet}\right) \\ \vdots \\ R_\theta\left(\nabla W^{(s)}_{H,\bullet}\right) \end{pmatrix}$$

Rotation Matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Final Update Rule

$$\nabla W^{(s)} := \nabla W^{(s)} + \nabla \mathring{W}^{(t)}$$

Experimental Results

	RBF SVM $C = 10, \gamma = 0.1$	Linear SVM $C = 0.1$	Softmax	K-NN K=3
RBM (H=100)	87.37%	59.27%	57.80%	82.69%
D-RBM (H=100, S=4)	83.44%	58.95%	56.80%	78.84%
D-RBM (H=100, S=9)	79.18%	53.62%	50.76%	73.56%
D-RBM (H=100, S=18)	69.84%	49.20%	46.58%	63.61%
O-RBM (H=100 S=18)	87.37%	58.99%	57.80%	82.69%
ERI-RBM (H=100, S=4)	78.49%	60.27%	58.31%	74.97%
ERI-RBM (H=100, S=9)	91.27%	74.87%	73.02%	88.48%
ERI-RBM (H=100, S=18)	**92.08%**	**77.69%**	**75.84%**	**89.34%**
TI-RBM (H=100, S=18)	80.63%	69.10%	68.20%	73.60%

- *Dominant-RBM (D-RBM)*: many RBMs are learned separately and the dataset is split w.r.t. the dominant orientation of images
- *Oriented-RBM (O-RBM):* one RBM is learned with input images that are pre-aligned with respect to their dominant orientation
- *Transformation-Invariant RBM (TI-RBM)*: Sohn et. al 2012

Conclusions

- We compared our proposed ERI-RBM with baseline (RBM, D-RBM, and O-RBM) and state of the art (TI-RBM) methods
- Using four different classifiers, we showed our method outperformed all the others (almost +5% better than O-RBM)
- Performance improvement from S=9 to S=18 is little (SVM), which indicates that finer discretisation is not needed
- Accuracy of 92% with 100 hidden units indicates ERI-RBM reaches the highest performance with compact representation
- Future work proved that ERI-RBM learns rotation invariant features with a score of 0.9, the highest amongst the compared methods (arXiv pre-print available)

Acknowledgments

- We thank NVIDIA Corporation for providing us a Titan X GPU