Disjuctive Normal Shape Boltzmann Machine

Ertunc Erdil[1], Fitsum Mesadi[2], Tolga Tasdizen[2], Mujdat Cetin[1]

[1] Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey
[2] Department of Electrical and Computer Engineering University of Utah, Salt Lake City, UT, USA

March 9, 2017

Outline

Motivation

- Shape modeling has a variety of applications in computer vision and image processing including object detection, image segmentation, shape matching, inpainting and graphics.
- A strong shape model should contain two important properties:
  - Realism: The model should capture the correct shape distribution.
  - Generalization: The samples generated from the learned distribution should cover unseen shapes.

Motivation

- Shape Boltzmann machine (a type of Deep Boltzmann machine) is a powerful tool for shape modelling; however, has some drawbacks in representation of local shape parts.
- Disjunctive Normal Shape Model (DNSM) is a strong shape representation that can effectively represent local parts of objects.

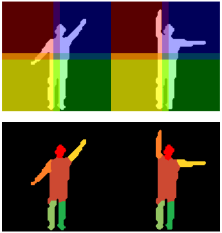
Shape Boltzmann Machine

DNSM

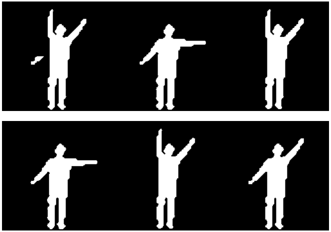
Contributions

- We propose a new shape model called Disjunctive Normal Shape Boltzmann Machine (DNSBM) which exploits
  - the property of shape Boltzman machine (SBM) for learning complex binary shape distributions
  - the property of DNSM for representing local parts of shapes.

Training Images

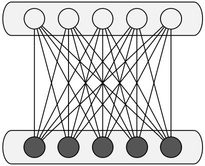


Samples



Restricted Boltzmann Machine (RBM)

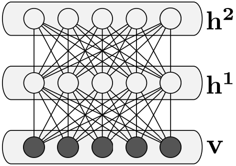
RBM is a model that includes a number of hidden variables $\mathbf{h}$ each connected to all image pixels (units in the visible layer $\mathbf{v}$)



RBMs can approximate any binary distribution if an exponential number of hidden units and a large amount of training data are available.

Deep Boltzmann Machine (DBM)

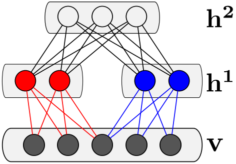
DBM is capable of learning more complex structures in the data using additional hidden units



DBMs require a large number of binary images to learn the shape distributions.

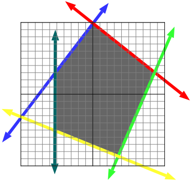
Shape Boltzmann Machine (SBM)

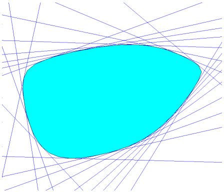
- SBM is a shape model based on RBM and DBM that accurately captures the properties of binary shapes.
- SBM is capable of learning shape distributions even when the size of the training set is limited, by exploiting information from local shape representations.



DNSM

- DNSM [1] approximates the characteristic function of a shape as a union of convex polytopes which themselves are represented as intersections of half-spaces.
- A polytope can be represented by intersection of half-spaces.





[1] Mesadi et al., "Disjunctive normal shape and appearance priors with applications to image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015

DNSM

- Consider the characteristic function of a D-dimensional shape $f : \mathbf{R}^D \rightarrow B$ where $B = \{0, 1\}$.
- Let $\Omega^+ = \{\mathbf{x} \in \mathbf{R}^D : f(\mathbf{x}) = 1\}$ represent the foreground region.
- $\Omega^+ \approx \bigcup_{i=1}^{N} P_i$ where $P_i = \bigcap_{j=1}^{M_i} H_{ij}$ and

$$h_{ij}(\mathbf{x}) = \begin{cases} 1, & \text{if} \sum_{k=1}^{D} \delta_{ijk} x_k + c_{ij} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

DNSM

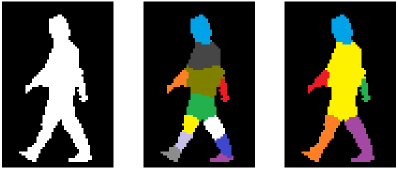
Converting the disjunctive normal form to a differentiable shape representation requires the following steps:

- Replace the disjunction with negations and conjunctions $f(\mathbf{x}) \approx \bigvee_{i=1}^{N} \bigwedge_{j=1}^{M_i} h_{ij}(\mathbf{x}) = \neg \bigwedge_{i=1}^{N} \neg \bigwedge_{j=1}^{M_i} h_{ij}(\mathbf{x})$.
- Conjunctions of binary functions are equivalent to their product and negation is equivalent to subtraction from 1, $f(\mathbf{x})$ can also be approximated as $1 - \prod_{i=1}^{N}(1 - \prod_{j=1}^{M_i} h_{ij}(\mathbf{x}))$.
- Relax the discriminants h_{ij} to sigmoid functions $\sigma_{ij} = 1/(1 + e^{-(\sum_{k=1}^{D} \delta_{ijk} x_k + c_{ij})})$.
- $f(\mathbf{x}) = 1 - \prod_{i=1}^{N} \left(1 - \prod_{j=1}^{M_i} \sigma_{ij}\right)$, where $\mathbf{x} = \{x, y\}$ for $(2D)$ shapes and $\mathbf{x} = \{x, y, z\}$ for $(3D)$ shapes.

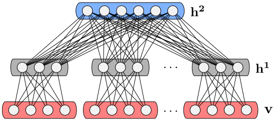
DNSM

- The only free parameters of the model are δ_{ijk} and c_{ij}.

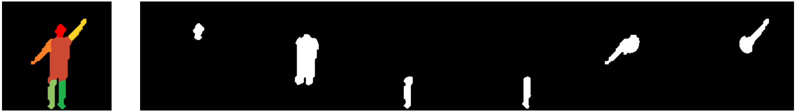
$$E(\mathbf{\Delta}^t) = \int_{\mathbf{x}\in\Omega} (f(\mathbf{x}) - q_t(\mathbf{x}))^2 d\mathbf{x} + \eta \sum_{i}^{N} \sum_{r\neq i}^{N} \int_{\mathbf{x}\in\Omega} g_i(\mathbf{x}) g_r(\mathbf{x}) d\mathbf{x}$$



DNSBM



- The units in each block of $\mathbf{v}$ are fully connected with the units in the corresponding block of $\mathbf{h^1}$.
- Each unit of $\mathbf{h^1}$ is also connected to all units of $\mathbf{h^2}$.
- While the connections between $\mathbf{v}$ and $\mathbf{h^1}$ capture the dependencies between pixels, the connections between $\mathbf{h^1}$ and $\mathbf{h^2}$ capture the inter-relations of local shape parts.

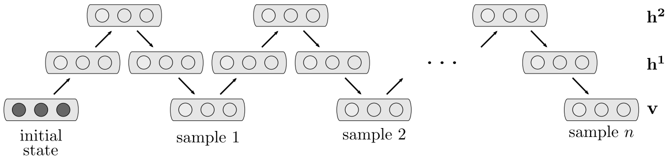


DNSBM

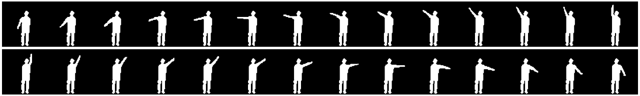
- Learning of the model involves maximizing $\log p(\mathbf{v};\theta)$ of the observed data $\mathbf{v}$ with respect to its parameters $\theta = \{b, W_1, W_2, c_1, c_2\}$.
- In DNSBM, each connected red-gray block pair between $\mathbf{v}$ - $\mathbf{h^1}$ and each connected gray-blue pair between $\mathbf{h^1}$ - $\mathbf{h^2}$ forms an RBM.
- We train DNSBM by training each RBM in DNSBM from bottom-up in a greedy manner using approximate gradient descent.

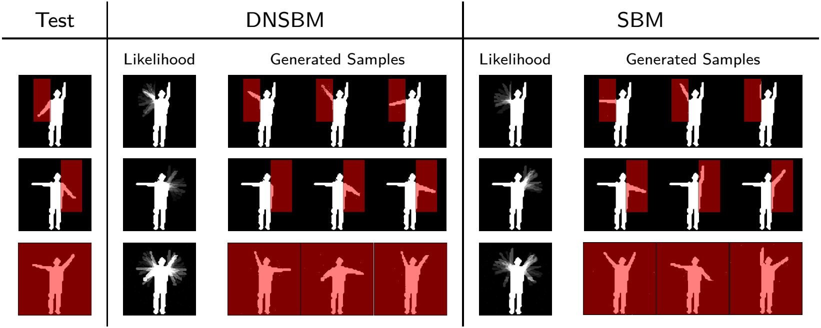
DNSBM

- In DNSBM, all hidden units are conditionally independent given the visible units.
- Similarly, all visible units are conditionally independent given the hidden units.
- This allows us to perform Gibbs sampling: $\mathbf{h}^1 \sim p(\mathbf{h}^1|\mathbf{h}^2, \mathbf{v})$, $\mathbf{h}^2 \sim p(\mathbf{h}^2|\mathbf{h}^1)$, $\mathbf{v} \sim p(\mathbf{v}|\mathbf{h}^1)$.

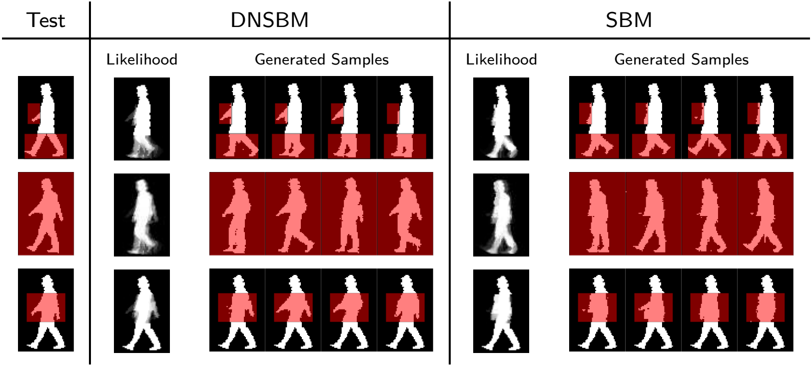


Standing Person data set





Walking Silhouette data set



Some unrealistic samples

- DNSBM

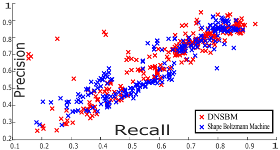
- SBM

Quantitative Results

- Dice score results

	DNSBM	SBM
Walking silhouette	**0.6526**	0.6112
Standing Person	**0.5935**	0.5825

- Precision-recall curve



- Imputation Score: $p(\mathbf{v}_u|\mathbf{v}_o) = \sum_s p(\mathbf{v}_u|\mathbf{h}^s)$ where $\mathbf{h}^s \sim p(\mathbf{h}|\mathbf{v}_o)$

	DNSBM	SBM
Imputation score	**0.085**	0.014

Conclusion

- We have presented a shape model, DNSBM, that is based on the SBM and the DNSM.
- DNSBM is able to represent physically meaningful local shape parts individually and exploits this representation when the training set size is limited.
- We have shown the performance of DNSBM on two data sets for shape completion.
- The proposed method exhibits better performance than SBM.
- The source code is available at `github.com/eerdil/dnsbm`

THANK YOU!