Image Segmentation by Deep Learning of Disjunctive Normal Shape Model Shape Representation

Mehran Javanmardi
University of Utah
Salt Lake City, UT, USA
mehran@sci.utah.edu

Ricardo Bigolin Lanfredi
University of Utah
Salt Lake City, UT, USA
ricbl@sci.utah.edu

Mujdat Cetin
Sabanci University
Istanbul, Turkey
mcetin@sabanciuniv.edu

Tolga Tasdizen
University of Utah
Salt Lake City, UT, USA
tolga@sci.utah.edu

Abstract

Segmenting images with low-quality, low signal to noise ratio has been a challenging task in computer vision. It has been shown that statistical prior information about the shape of the object to be segmented can be used to significantly mitigate this problem. However estimating the probability densities of the object shapes in the space of shapes can be difficult. This problem becomes more difficult when there is limited amount of training data or the testing images contain missing data. Most shape model based segmentation approaches tend to minimize an energy functional to segment the object. In this paper we propose a shape-based segmentation algorithm that utilizes convolutional neural networks to learn a posterior distribution of disjunction of conjunctions of half spaces to segment the object. This approach shows promising results on noisy and occluded data where it is able to accurately segment the objects. We show visual and quantitative results on datasets from several applications, demonstrating the effectiveness of the proposed approach. We should also note that inference with a CNN is computationally more efficient than density estimation and sampling approaches.

1. Introduction

Prior statistical knowledge of the shapes of objects to be segmented has shown to be an important part of the many segmentation algorithms that deal with low quality data. Shape priors also are essential for segmenting objects with occlusions. For instance, consider the scenario of segmenting pedestrians that are occluded by fences of a sidewalk. Most semantic segmentation algorithms [2, 4, 22, 31] can do a very efficient segmentation of the part of the pedestrians that are not occluded, however they are incapable of segmenting the whole shape. This can be achieved by incorporating prior information about the shape and appearance of the human body into the segmentation algorithm. Other scenarios in which prior shape and appearance information could be critical include cases where the surrounding objects are hardly differentiated from the object of interest due to low image contrast and low signal to noise ratio. This issue is of importance especially in medical image segmentation where the tissue boundaries are not clearly visible [27]. These limitations promote segmentation approaches that efficiently utilize the appearance and shape priors. In this work, we present a novel deep learning framework that allows fitting parametric implicit shape models which is further learned for image segmentation.

2. Related Work

One of the pioneers to incorporate shape priors in segmentation of images is the active contour or snakes model by Kass et al. [18]. In the snakes model, a contour is a spline which is constrained by regularizing terms such as its length and curvature. Another popular approach to segment object specific shapes is the Active Shape Model (ASM) proposed by Cootes et al. [8]. Active shape models are mostly used in applications where a more informative object specific shape prior model needs to be learned. Active shape models are parametric deformable models which statistically capture the variations of shapes in the provided training data. These models are also called the Point Distribution Models (PDM), which iteratively deform to fit an unseen object in a new image. The PDM is constructed using principal component analysis which restricts their ap-

plication to problems with unimodal shape densities.

There have been multiple approaches that extend ASMs to handle multimodal distributions by learning manifolds [14] or using a mixture of Gaussians to model the shape distribution [7]. The drawbacks of these models could be seen in the use of parametric probability distributions which could impede analysis of complex shape variations. Another disadvantage of ASMs could be considered as their need for annotating landmark points across all training examples which could be a very time consuming task. Active Appearance Models (AAM) [6] are another extension of ASMs which capture appearance variations as well as shape variations, however like ASMs they consider a Gaussian-like distribution to model appearances as well as linear analysis tools like PCA which hinders multimodal analysis.

Shape prior based level sets were introduced to overcome the shortcomings of ASMs. One of the advantages of level set based approaches over active shape models is their implicit nature which enables them to handle topological changes in addition to eliminating the need for landmarks in training. Principal component analysis is used on the signed distance functions of level sets in [27, 21] to capture shape variability. However this approach has the disadvantage of linear combinations of shapes not corresponding to valid shapes, because the space of signed distance functions is a nonlinear manifold. Other works [19, 9], propose a level set paradigm where nonparametric density estimation based shape priors are used. These methods extend the use of shape priors to multi-modal shape densities.

In [15], Foulonneau et al. use Legendre moments from silhouettes as shape descriptors and the distance between shape descriptors to estimate the prior shape density. There are different methods [5, 30, 26] that include learning based data terms with nonparametric shape priors. Utilizing more elaborate data terms can result in significant improvement of segmentation accuracy especially in scenarios that foreground and background have complex densities. Erdil et al. [10] propose an algorithm which exploits nonparametric shape and features priors. Learned discriminative class-dependent geometric and appearance based features are extracted and are incorporated with shape priors into the segmentation process. Erdil et al. [12] propose a Markov Chain Monte Carlo (MCMC) sampling approach which is capable of segmenting objects with occlusion in a low SNR image using nonparametric shape priors.

There are other works in the literature that exploit prior-based information for image segmentation. Zhang et al. [32, 33] propose sparse shape composition which has multiple medical image segmentation utility. Patenaude et al. [24] perform subcortical brain segmentation using a Bayesian model for shape and appearance priors. Nguyen et al. [28] use support vector machines to propose a classifier-based implicit shape representation. Hamarneh et al. [16] propose an algorithm to perform watershed segmentation using prior shape and appearance information. Wang et al. [29] utilize adaptive shape priors in a graph cut image segmentation paradigm. Deep learning methods such as Boltzmann machines have been also used to apply shape priors [13, 3, 11].

Ramesh et al. [25] propose the Disjunctive Normal Shape Model (DNSM) for image segmentation which is formed by disjunction of a number of polytopes where each polytope is formed by conjunction of half spaces. Mesadi et al. [23] reformulate DNSM in a Bayesian framework for image segmentation. We reformulate the DNSM such that it could used on top of a Convolutional Neural Network (CNN) where the whole network is jointly trained. The joint training of DNSM and a CNN delivers some benefits to the whole algorithm, (i) There will be no need for an elaborate initialization of DNSM parameters, (ii) It makes the algorithm fully automatic by getting rid of the need for seed points in the original DNSM, (iii) The algorithm becomes more robust to noise and occlusions since the original DNSM uses a simple Chan-Vese energy for segmentation. More details on DNSM are given in the next section.

3. Approach

In this section we discuss the general formulation of DNSMs and then we show how it can be integrated into a convolutional neural network to be jointly trained for accurate image segmentation.

3.1. DNSM Shape representation

Disjunctive Normal Shape Model is an implicit parametric shape model by Ramesh et al. [25]. DNSM is disjunction of **N** convex polytopes where each polytope itself is the conjunction of **M** half spaces in the D-Dimensional space. This disjunction is able to form arbitrary shapes in arbitrary dimensional space.

Consider a characteristic function $f : \mathbb{R}^D \rightarrow B$ where $B = \{0, 1\}$. Let $\Omega_f = \{\mathbf{x} \in \mathbb{R}^D : f(\mathbf{x}) = 1\}$. If we approximate Ω_f as union of **N** polytopes we will have $\tilde{\Omega}_f = \bigcup_{i=1}^{\mathbf{N}} P_i$, where the i'th polytope, P_i, is defined as intersection of **M** half spaces. We can write $P_i = \bigcap_{j=1}^{\mathbf{M}} H_{ij}$. We can define arbitrary half space H_{ij} with the perceptron equation as follows:

$$H_{ij}(\mathbf{x}) = \begin{cases} 1, & \sum_{k=0}^{D} w_{ijk} x_k + b_{ij} \geq 0 \\ 0, & otherwise \end{cases} \quad (1)$$

where w_{ijk} and b_{ij} are the weights and bias terms and D is the input dimension. Note that we can include the bias term into w_{ijk}s which will make the input $D+1$ dimensional. By Boolean logic any function can be written as disjunction of conjunction which is known as the disjunctive normal form

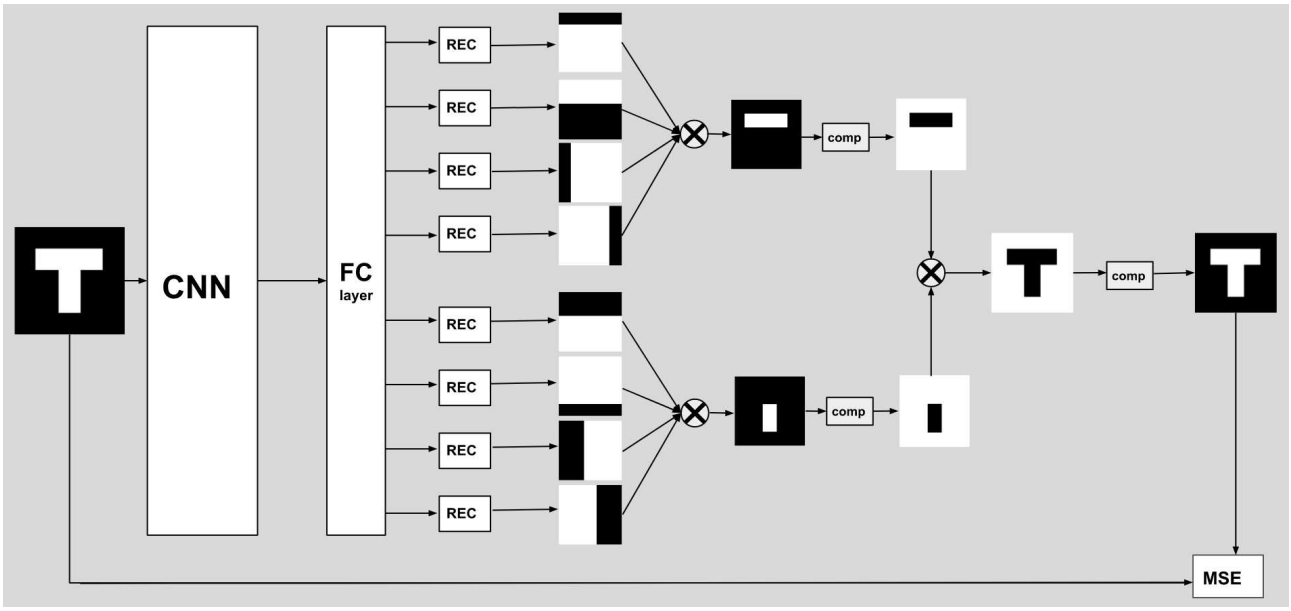

Figure 1. The proposed architecture for joint training. The CNN learns to produce the coefficients for the half spaces, which are further used by the shape reconstruction block, REC, to generate the half spaces in 2D. Further these half spaces are multiplied together to create polytopes. The disjunction of these polytopes finally represents the whole shape which is enabled by image complement and multiplication. The network can be jointly trained using a simple Mean Squared Error on the input shape and the output reconstructed shape. The final fully connected layer of the CNN is shown separately to emphasize that it corresponds to the parameter vector of the DNSM.

[17]. Therefore the characteristic function for $\tilde{\Omega}_f$ can be written as:

$$\tilde{f}(\mathbf{x}) = \bigvee_{i=1}^{\mathbf{N}} \underbrace{\left(\bigwedge_{j=1}^{\mathbf{M}} H_{ij}(\mathbf{x}) \right)}_{d_i(\mathbf{x})} \tag{2}$$

To convert the disjunctive normal form of the characteristic function to a differentiable form we approximate the conjunction of the binary variables with the product of their approximations. We can write $d_i(\mathbf{x}) = \bigwedge_{j=1}^{\mathbf{M}} H_{ij}(\mathbf{x})$ is equivalent to $d_i(\mathbf{x}) = \prod_{j=1}^{\mathbf{M}} H_{ij}(\mathbf{x})$. As for the disjunction we can use De Morgan's law to transform disjunction into a conjunction and use the product approximation. We can write $\bigvee_{i=1}^{\mathbf{N}} d_i(\mathbf{x})$ as the negation of conjunctions $\neg \bigwedge_{i=1}^{\mathbf{N}} \neg d_i(\mathbf{x})$. Since we are dealing with binary entities we can write negations as $\neg\Psi = 1 - \Psi$ which results in $1 - \prod_{i=1}^{\mathbf{N}}(1 - d_i(\mathbf{x}))$. We relax the binary perceptron H_{ij} with logistic sigmoid function:

$$\sigma_{ij}(\mathbf{x}) = \frac{1}{1 + e^{\sum_{k=1}^{\mathbf{D+1}} w_{ijk} x_k}} \tag{3}$$

Now we can write the differentiable characteristic function as :

$$\tilde{f}(\mathbf{x}) = 1 - \prod_{i=1}^{\mathbf{N}} (1 - \prod_{j=1}^{\mathbf{M}} \sigma_{ij}(\mathbf{x})) \tag{4}$$

where $\mathbf{x}$ is $(x, y, 1)$ for the 2D space and $(x, y, z, 1)$ for the 3D space.

3.2. Network Representation

In the previous section we investigated how DNSM represents shape and studied how it can be reformulated as a differentiable function. The original work on DNSM [25] requires an elaborate initialization of the model weights, w_{ijk} and accurate seed points for each polytope of the DNSM. This requires user interference, which makes the method semi-automatic. In this section we use the differentiable shape representation of DNSM, Equation 4, incorporated on top of a convolutional neural network which enables fully automatic learning of model weights through annotated training data regardless of how they are initialized.

3.2.1 Learning Model Weights by a CNN

As depicted in Figure 1 we use a convolutional neural network to learn the model weights for our DNSM shape representation. The CNN take as input an image and outputs $\mathbf{N} \times \mathbf{M} \times \mathbf{F}$ coefficients which correspond to the w_{ijk} and b_{ij} in Equation 3. $\mathbf{N}$ and $\mathbf{M}$ are hyperparameters of the DNSM model which accordingly correspond to the number of groups (polytopes) and the number of discriminants (half spaces) per group. $\mathbf{F}$ is the number of variables needed to represent a half space in arbitrary dimension. For instance in 2D the variables needed are $(x, y, 1)$ which results in $\mathbf{F} = 3$.

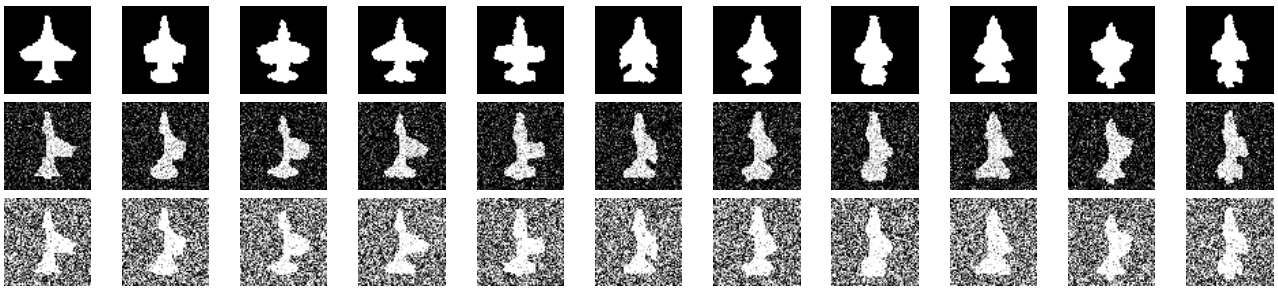

Figure 2. The aircraft dataset [19]. The first row is the training images used in a leave-one-out fashion. The second row corresponds to the testing images with low amount of noise and the third row corresponds to the testing images with high amount of noise.

Figure 1 depicts the proposed architecture to represent the shape of a toy example letter "T". The Model has 2 groups ($\mathbf{N} = 2$) and each group has 4 discriminants ($\mathbf{M} = 4$). Since the input image is in 2D, we will need 3 variables to represent the half space, therefore $\mathbf{F} = 3$. As can be seen in the Figure 1, there are 8 arrows exiting the CNN, each one containing 3 coefficients to represent a half space in 2D. The coefficients for each half space are fed into a 'REC' block which reconstructs the 2D half spaces from the input coefficients.

3.2.2 DNSM as a Neural Network

The half spaces generated by the convolutional neural network and the shape reconstruction blocks from last section correspond to σ_{ij} in Equation 3. Since Equation 3 is a sigmoid function the reconstructed half spaces have intensities between zero and one, therefore we can treat them as relaxed binary images. We can construct the characteristic function $\tilde{f}(\mathbf{x})$ in Equation 4 by simple multiplication of images followed by 'COMP' blocks which perform image complementation which corresponds to subtraction of the input image from an all ones image. This network is differentiable and is used in an end-to-end training with the CNN to learn shapes from a training set. We use a Mean Squared Error (MSE) of the output of the network and the input image to train the network.

3.3. Interpretation

The joint network can be interpreted as a maximum a posteriori (MAP) estimator. Let us assume that we want to estimate the shape y given an input image x. We can write:

$$p(y|x) \propto p(x|y) \times p(y) \tag{5}$$

where $p(y)$ corresponds to the shape prior which is enabled by learning of DNSM weights through a CNN for a specific dataset. During training, the CNN learns how the polytopes are placed to generate the whole shape and even in more details how each specific polytope is generated through its discriminants. During testing, the network applies these priors through CNN by generating the coefficients for the DNSM such that it meets constraints learned during training. For instance during training it learns that airplanes are symmetric, therefore in testing, if model generates a polytope on one side of the image, there must be corresponding weights from CNN that enforce polytopes on the other side.

3.4. Training

In the last section and Figure 1 we observed how the model generates a shape, given an input image. This shape will have a continuous value between zero and one. We can refer to the generated shape as the output probability map of the network. This probability map can be used to train the network. For instance we can train the network for image segmentation where we use a MSE of the output probability map and the ground truth binary mask of the training data. We could also use an image reconstruction loss to train the network if the target shapes are not binary. In this case the loss would be the MSE of the output shape reconstruction and the input image. WE should note that in this case the input images should be scaled such that they have values between zero and one. We use an image segmentation task for our experiments on airplane dataset, walking silhouettes dataset and dendritic spine dataset and use an image reconstruction loss to train the network for our experiments on MNIST digit dataset.

4. Experiments

In this section we present experimental results of the proposed approach on a variety of datasets. We perform experiments on test images that contain a high amount of additive white noise. We also validate the capability of the proposed algorithm in segmenting objects with occlusions and missing parts. We perform multiple experiments on aircraft dataset [19], MNIST handwritten digits dataset [20], the walking silhouettes dataset [9] and the dendritic spine dataset. All experiments are performed in Tensorflow [1] and on an NVIDIA Titan X GPU.

For all the experiments performed in this paper we choose 24 groups (polytopes) and for each group we choose 8 discriminants (half spaces). We choose these values by performing cross validation on the walking silhouettes

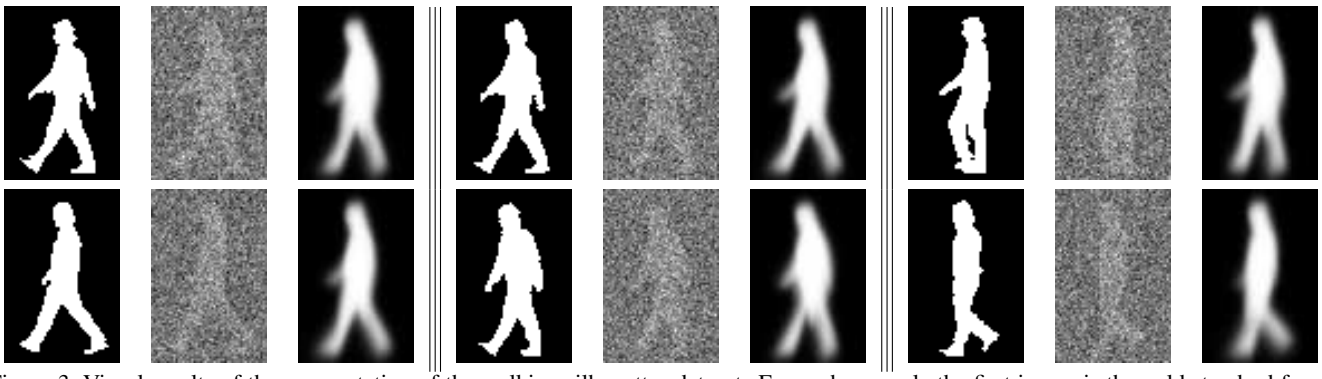

Figure 3. Visual results of the segmentation of the walking silhouettes dataset. For each example the first image is the gold standard for the segmentation, the second image is the corrupted image with high level of noise and the third image is the output probability map of the proposed approach. Quantitative results for the first three images of the first row are given in Table 3.

Table 1. Quantitative results on the first testing images of the dataset with different amount of noise. The numbers reported are the Dice coefficient.

	Kim et al. [19]	Erdil et al. [12]	Proposed Approach
image 1, low noise	91.12	**94.49**	94.23
image 2, low noise	92.69	94.75	**98.51**
image 3, low noise	89.97	94.66	**96.96**
image 1, high noise	87.15	89.86	**94.80**
image 2, high noise	90.31	93.25	**98.14**
image 3, high noise	88.03	90.05	**97.83**

dataset and keep them fixed for all the other datasets. The convolutional neural network used is a network of two layers of convolutional layers followed by ReLU non-linearity and max pooling layers. The output of the second max pooling layer is then fed into two consecutive fully connected layers where the corresponding coefficients for the DNSM model are generated.

4.1. Aircraft Dataset

The aircraft dataset [19] contains 11 binary images of aircrafts which are synthetically generated. We follow Erdil et al. [12] to evaluate our proposed architecture on this dataset. The training set and two different testing sets are shown in Figure 2. To generate the testing sets all the images in the training set have been cropped such that the left wing for aircrafts are missing. Two different testing sets are generated by adding different noise levels to the occluded images. The middle row in Figure 2 corresponds to the testing set with low level noise and the bottom row corresponds to the testing set with high level of noise. We perform our experiments in a leave-one-out setting. We train 11 different models where for each model 10 images are chosen for training and one is left for testing.

Visual results of the proposed approach are shown in Figures 4 and 5 for two different noise levels. Figure 4 depicts the results on the first three images of the testing dataset

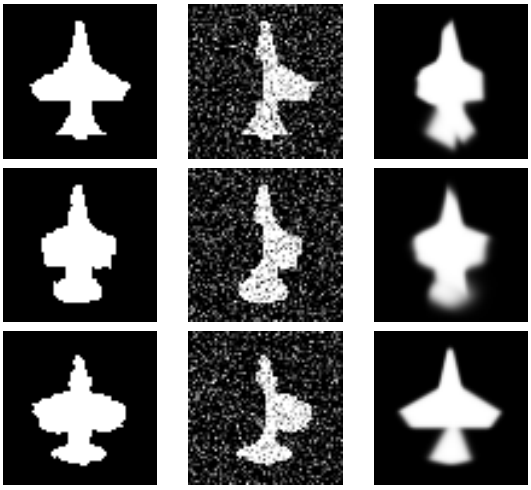

Figure 4. Visual segmentation results of the proposed approach on the high SNR testing dataset. The first column depicts the ground truth for segmentation, the second column shows the testing images with low amount of noise and cropped left wing, the third column depicts the final segmentation of our approach on the noisy images

with high SNR. Figure 5 contains the results from the low SNR testing dataset. Quantitative results for the first three testing images are given in Table 1. We report the Dice coefficient as the quantitative measure for our segmentation results. We compare our results quantitatively with Kim et al. [19] and Erdil et al. [12] and observe better performance in most cases.

4.2. Walking Silhouettes Dataset

The Walking Silhouettes dataset [9] contains 151 binary images of a person while walking. For this experiment we follow Erdil et al. [10] and choose a subset of 30 images out of 151 for our experiments. Out of the 30 image we use the same 16 training images as [12] and test on the remaining 14 images. The test images are corrupted with a high amount of additive white noise to show the robustness of the proposed method in dealing with noisy images. We

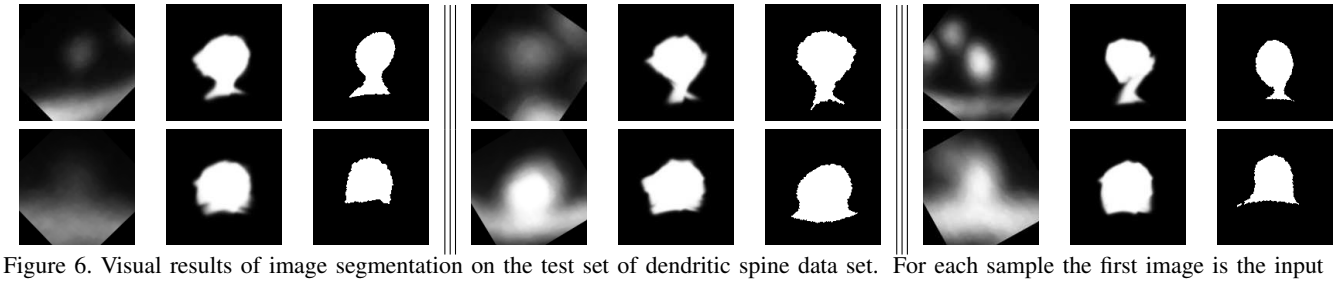

Figure 6. Visual results of image segmentation on the test set of dendritic spine data set. For each sample the first image is the input image to be segmented, the second image is the proposed approach's output probability map and the third image is the gold standard for segmentation. The three samples in the first row correspond to mushroom spines and the three samples on the second row correspond to stubby spines.

Table 2. Quantitative results on the dendritic spine dataset. The numbers reported are the Dice coefficient.

	Foulonneau et al. [15]	Kim et al. [19]	Chen et al. [5]	Erdil et al. [10] with geometric priors	Erdil et al. [10] with appearance priors	Proposed Approach
Dice Score	73.48	64.24	72.38	74.74	74.92	**75.25**

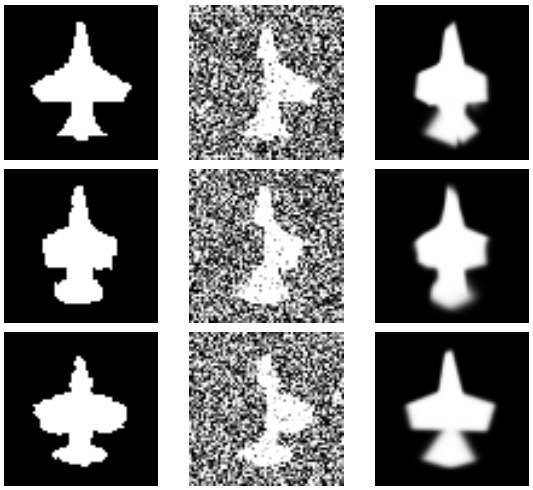

Figure 5. Visual segmentation results of the proposed approach on the low SNR testing dataset. The first column depicts the ground truth for segmentation, the second column shows the testing images with high amount of noise and cropped left wing, the third column depicts the final segmentation of our approach on the noisy images

present quantitative and visual results for three of the testing images in Figure 3 and Table 3.

Table 3. Quantitative results on three test images of the Walking Silhouettes dataset shown in Figure 3 with high amount of additive noise. The numbers reported are the Dice coefficient.

	Kim et al. [19]	Erdil et al. [12]	Proposed Approach
image 1	87.13	88.10	**91.38**
image 2	89.87	89.87	**93.30**
image 3	85.43	86.58	**92.06**

4.3. Dendritic Spine Data Set

In this section we evaluate our approach on the dendritic spine data set. The data set is obtained from Neuronal Structure and Foundation Laboratory of Champalimaud Neuroscience Foundation in Lisbon.

In general dendritic spines are categorized into 4 groups, mushroom spines, thin spines, stubby spines and filopodia spines. We follow [10] to perform our experiments and use samples only from the mushroom and stubby classes. The data set contains 88 mushroom and 27 stubby spines. We use a subset of 8 mushroom spines and 8 stubby spines as our training set and evaluate our method on the remaining 80 mushroom and 19 stubby spines. Visual results of the segmentation are given in Figure 6 for both mushroom and stubby spines where the first row correspond to mushroom samples and the second row depicts stubby spine examples. We also compare our results quantitatively in Table 2 with different methods. As can be observed we achieve the highest Dice coefficient for segmenting the spines on this data set.

4.4. MNIST Digits Dataset

The MNIST [20] is a large dataset of handwritten digits containing 60000 images for training and 10000 images for testing. It contains 10 different classes corresponding to different modes of shape density. In our experiments for this dataset we predict 10 different reconstructed images for a single input which corresponds to all ten different modes of shape density. We also predict probabilities with every reconstructed image which represents the probability that the reconstructed image belongs to that specific class of shape density.

Visual results of our approach are given in Figure 7. We train our model using all the 60000 images provided in the training set and test on the unseen 10000 testing images. We add additive Gaussian noise to the images and randomly remove some part of the digits in the testing image to show our approach's resilience to noise and missing data. We also occlude some parts of the digits with a random block to evaluate our approach's capacity to deal with noise and

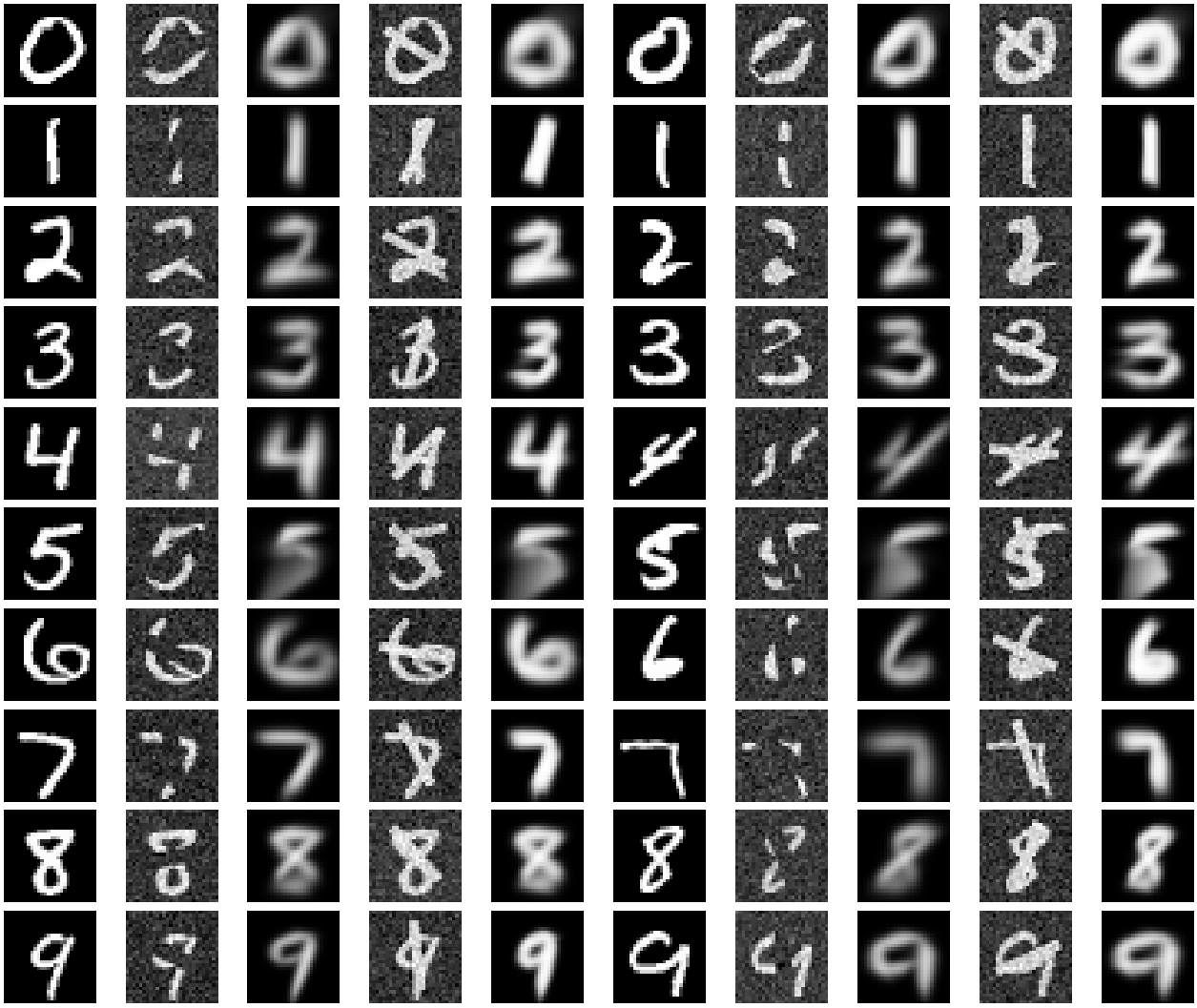

Figure 7. Visual results of the image reconstruction of noisy occluded images. The first column for each sample corresponds to the gold standard for the image reconstruction, the second column corresponds to the noisy missing data image, the third column is the output of our proposed approach for the input of second column, the fourth column depicts the noisy occluded image and the last column corresponds to the proposed approach's output when the input is the fourth column. Each row contains 2 examples from the same class.

occlusion. The reconstructed images from noisy occluded images are given in Figure 7. The reconstructed image in the class with the highest probability is shown.

5. Conclusion

We have presented a framework for image segmentation with shape models. We adapt the Disjunctive Normal Shape Models for shape representation which is enabled by a convolutional neural network to perform image segmentation. We have shown superior performance of our proposed approach on four different datasets with quantitative and qualitative results for segmentation and reconstruction. The experimental results demonstrate the proposed approach's efficiency in dealing with high amount of noise in the data as well as occlusions and missing information. The proposed approach also benefits from fast inference via CNNs which is computationally more efficient that other methods such as density estimation and sampling.

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016. 4

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 1

[3] F. Chen, H. Yu, R. Hu, and X. Zeng. Deep learning shape priors for object segmentation. In *Computer Vision and Pat-*

tern Recognition (CVPR), 2013 IEEE Conference on, pages 1870–1877. IEEE, 2013. 2

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv preprint arXiv:1606.00915*, 2016. 1

[5] S. Chen and R. J. Radke. Level set segmentation with both shape and intensity priors. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 763–770. IEEE, 2009. 2, 6

[6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001. 2

[7] T. F. Cootes and C. J. Taylor. A mixture model for representing shape variation. *Image and Vision Computing*, 17(8):567–573, 1999. 2

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 1

[9] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International journal of computer vision*, 69(3):335–351, 2006. 2, 4, 5

[10] E. Erdil, M. U. Ghani, L. Rada, A. O. Argunsah, D. Unay, T. Tasdizen, and M. Cetin. Nonparametric joint shape and feature priors for image segmentation. *IEEE Transactions on Image Processing*, 26(11):5312–5323, 2017. 2, 5, 6

[11] E. Erdil, F. Mesadi, T. Tasdizen, and M. Cetin. Disjunctive normal shape boltzmann machine. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 2357–2361. IEEE, 2017. 2

[12] E. Erdil, S. Yildirim, M. Cetin, and T. Tasdizen. Mcmc shape sampling for image segmentation with nonparametric shape priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 411–419, 2016. 2, 5, 6

[13] S. A. Eslami, N. Heess, C. K. Williams, and J. Winn. The shape boltzmann machine: a strong model of object shape. *International Journal of Computer Vision*, 107(2):155–176, 2014. 2

[14] P. Etyngier, F. Segonne, and R. Keriven. Shape priors using manifold learning techniques. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 2

[15] A. Foulonneau, P. Charbonnier, and F. Heitz. Multi-reference shape priors for active contours. *International journal of computer vision*, 81(1):68, 2009. 2, 6

[16] G. Hamarneh and X. Li. Watershed segmentation using prior shape and appearance knowledge. *Image and Vision Computing*, 27(1-2):59–68, 2009. 2

[17] M. Hazewinkel. *Encyclopaedia of mathematics: C an updated and annotated translation of the soviet mathematical encyclopaedia*, volume 2. Springer Science & Business Media, 2013. 3

[18] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988. 1

[19] J. Kim, M. Çetin, and A. S. Willsky. Nonparametric shape priors for active contour-based image segmentation. *Signal Processing*, 87(12):3021–3044, 2007. 2, 4, 5, 6

[20] 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. 4, 6

[21] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Computer vision and pattern recognition, 2000. Proceedings. IEEE conference on*, volume 1, pages 316–323. IEEE, 2000. 2

[22] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1

[23] F. Mesadi, E. Erdil, M. Cetin, and T. Tasdizen. Image segmentation using disjunctive normal bayesian shape and appearance models. *IEEE transactions on medical imaging*, 2018. 2

[24] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922, 2011. 2

[25] N. Ramesh, F. Mesadi, M. Cetin, and T. Tasdizen. Disjunctive normal shape models. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 1535–1539. IEEE, 2015. 2, 3

[26] A. Soğanlı, M. G. Uzunbaş, and M. Çetin. Combining learning-based intensity distributions with nonparametric shape priors for image segmentation. *Signal, Image and Video Processing*, 8(4):789–798, 2014. 2

[27] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE transactions on medical imaging*, 22(2):137–154, 2003. 1, 2

[28] H. Van Nguyen and F. Porikli. Support vector shape: A classifier-based shape representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(4):970–982, 2013. 2

[29] H. Wang, H. Zhang, and N. Ray. Adaptive shape prior in graph cut image segmentation. *Pattern Recognition*, 46(5):1409–1414, 2013. 2

[30] R. Yang, M. Mirmehdi, X. Xie, and D. Hall. Shape and appearance priors for level set-based left ventricle segmentation. *IET Computer Vision*, 7(3):170–183, 2013. 2

[31] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015. 1

[32] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Sparse shape composition: A new framework for shape prior modeling. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1025–1032. IEEE, 2011. 2

[33] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical image analysis*, 16(1):265–277, 2012. 2