A Study of Shape Modeling Against Noise

Cheng Long, Adrian Barbu

Department of Statistics

Florida State University

October 16, 2022

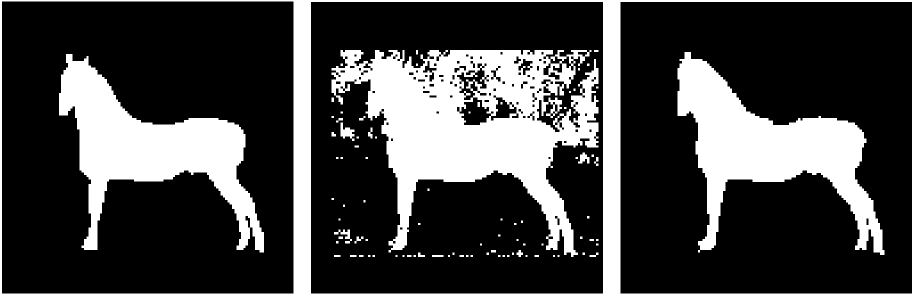
Outline

Shape Denoising

- We are interested in methods for modeling the shape (contour) of a specific class of objects, e.g. horses.
- In this study the shapes are represented as binary images of a certain size, e.g. 128×128
- The focus of our study is to study shape denoising - the task of recovering a shape corrupted by noise
- We will introduce six types of shape noise
- We will evaluate and compare the performance of seven shape modeling methods for shape denoising.

The Shape Denoising Problem

Shape denoising is the process of removing the noise from a shape, with the goal of obtaining a shape as close to the original shape as possible.

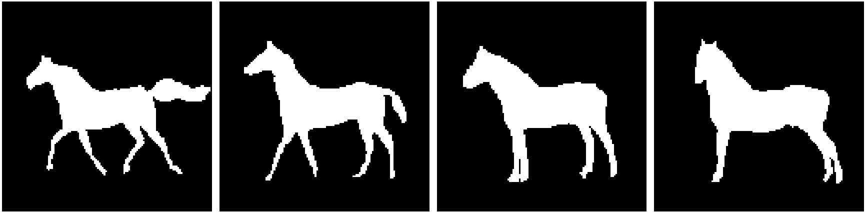


(a) Original shape (b) Noisy shape (c) Denoised shape

Figure: Shape denoising example. The noisy shape (b) has been obtained from the original shape (a) by a noise inducing process. A shape denoising method is used to obtain the denoised shape (c).

Shape Representation

Shapes are represented as binary images.

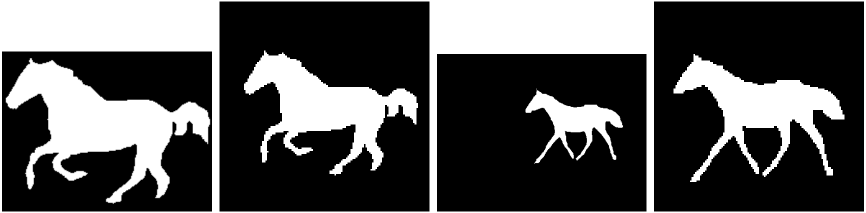


(a) sample 1 (b) sample 2 (c) sample 3 (d) sample 4

Figure: Examples of shapes used in this work.

Shape Alignment

- We used binary images of size 128×128.
- All binary images were aligned to have the objects centered and of approximately the same size.

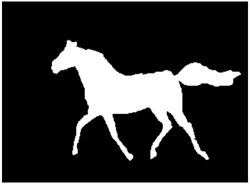


(a) mask image1 (b) aligned image1 (c) mask image2 (d) aligned image2

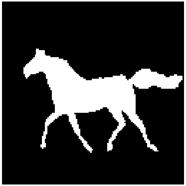
Figure: Two alignment examples from the Weizmann Horse Dataset.

Six types of noise

We study six different types of noise.

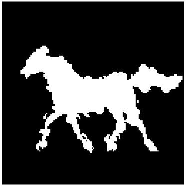


(a) original



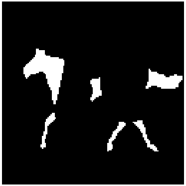
(b) aligned

(c) salt and pepper

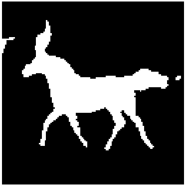


(d) circle

(e) real image



(f) occlusion

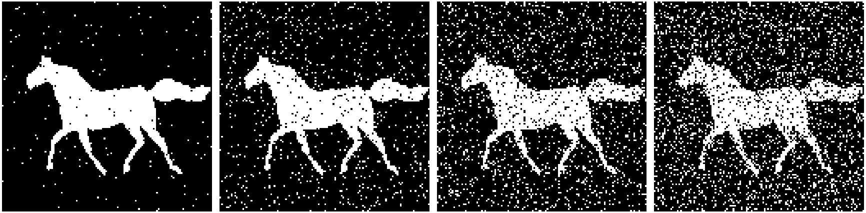


(g) detection image

(h) thresholded probability

Salt and Pepper Noise

The salt and pepper noise is obtained by flipping each pixel to its opposite value with a probability p.

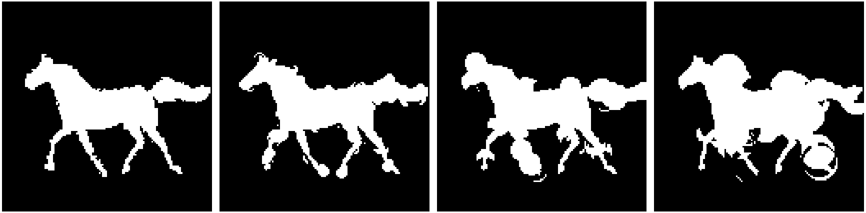


(a) $p = 0.01$ (b) $p = 0.05$ (c) $p = 0.10$ (d) $p = 0.15$

Figure: Salt and pepper noise with different levels p.

Circle Noise

The circle noise is obtained by adding semicircles or punching holes at random locations on the boundary between the foreground and the background.

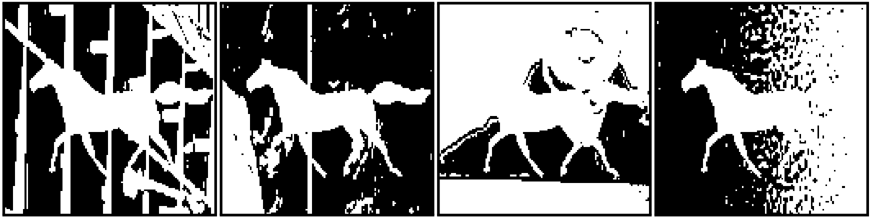


(a) $r = 1$ (b) $r = 3$ (c) $r = 6$ (d) $r = 10$

Figure: Circle noise with different levels.

Real Image Noise

- Noisy backgrounds are obtained by thresholding real images using various thresholds
- The shape background pixels are replaced with the noisy background

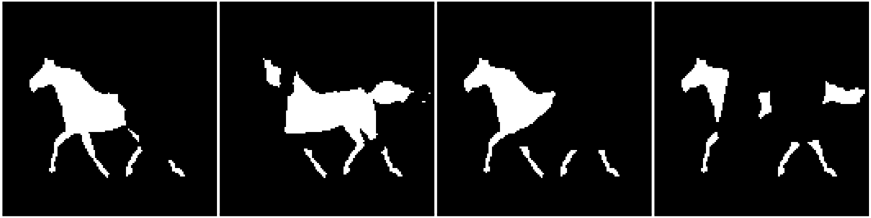


(a) sample 1 (b) sample 2 (c) sample 3 (d) sample 4

Figure: Examples of real image noise.

Occlusion Noise

The shape foreground pixels are occluded using thresholded real images.



(a) sample 1 (b) sample 2 (c) sample 3 (d) sample 4

Figure: Examples of occlusion noise.

Detection Image Noise

Detection image noise is obtained as the segmentation of an object from a color or grayscale image using a trained CNN (convolutional neural network).

(a) object shape (b) color image (c) detection image noise

Figure: Example of detection image noise in the Weizmann Horse Dataset.

Thresholded Probability Noise

Uses a color image I and a binary image M representing the object in the image.

- Denote $C_1 = \{(x, y), M(x, y) = 1\}$ as the foreground region and $C_0 = \{(x, y), M(x, y) = 0\}$ as the background region.
- k-means clustering is used to partition the image into k clusters, obtaining cluster indices for all pixels $L \in \{1, 2, ..., k\}^N$.
- For cluster $i \in \{1, 2, ..., k\}$, obtain the number of pixels that belong to foreground or background:

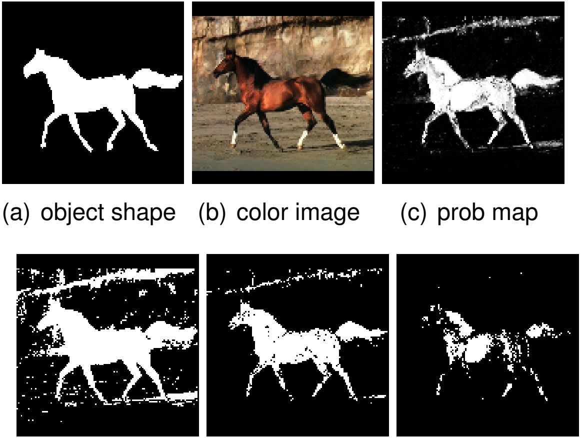
$$\begin{aligned} N_{i1} &= |\{(x, y)|L(x, y) = i \wedge (x, y) \in C_1\}|, \\ N_{i0} &= |\{(x, y)|L(x, y) = i \wedge (x, y) \in C_0\}| \end{aligned} \tag{1}$$

- Then the probability map of color image I can be computed as follows:

$$P(x, y) = \frac{N_{j1}}{N_{j1} + N_{j0}}, \text{ where } j = L(x, y). \tag{2}$$

Thresholded Probability Noise

Then binary noisy shapes are obtained by applying different thresholds to the probability map P.



(a) object shape (b) color image (c) prob map

(d) threshold 0.04 (e) threshold 0.5 (f) threshold 0.98

Clean Image Datasets

The **Weizmann Horse dataset** [Borenstein et al., 2004]:

- Contains 327 horse images and their corresponding mask images.
- 159 images were randomly selected as the training set S_{train}^{clean} and the other 168 images as the test set S_{test}^{clean}.

The Caltech-UCSD Birds 200 dataset [Welinder et al., 2010] contains photos of 200 bird species.

- We use 417 images of seven Flycatcher species in our experiments.
- 207 images were randomly selected as the training set S_{train}^{clean} and the other 210 images as the test set S_{test}^{clean}.

IoU

In our experiments, the criterion we use to estimate the performance of modeling against noises is Intersection over Union (IoU), also known as the Jaccard Index.

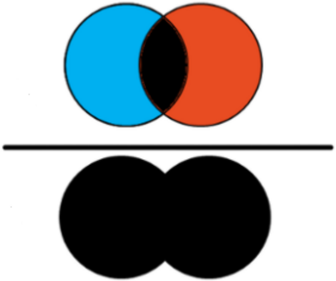


Figure: Computation of $IOU(A, B) = \frac{|A \cap B|}{|A \cup B|}$.

Noisy Image Datasets

We denote the set $\{S_{test}^{salt}, S_{test}^{circle}, S_{test}^{real}, S_{test}^{occlusion}, S_{test}^{detection}, S_{test}^{probability}\}$ as S_{test}^{all}.

Dataset	IoU(begin)	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Horse	S_{test}^{salt}	1000	1000	1000	1000	1000
	S_{test}^{circle}	77	1000	1000	1000	1000
	S_{test}^{real}	1000	1000	1000	1000	1000
	$S_{test}^{occlusion}$	479	622	690	938	1779
	$S_{test}^{detection}$	3	7	14	77	60
	$S_{test}^{probability}$	298	673	1095	1345	525
Bird	S_{test}^{salt}	1000	1000	1000	1000	1000
	S_{test}^{circle}	416	1000	1000	1000	1000
	S_{test}^{real}	1000	1000	1000	1000	1000
	$S_{test}^{occlusion}$	435	556	809	972	1000
	$S_{test}^{detection}$	22	30	44	56	15
	$S_{test}^{probability}$	843	962	934	519	70

Table: Number of noisy test images in each noise category.

Shape Denoising Methods Evaluated

The following methods have been evaluated for shape modeling and denoising:

- Active Shape Model (ASM) [Cootes et al., 1995]
- Deep Boltzman Machine (DBM) [Salakhutdinov and Hinton, 2009]
- Centered Convolutional DBM (CDBM) [Yang et al., 2021]
- Energy Based Model (EBM) [Pang et al., 2020]
- U-Net [Ronneberger et al., 2015]
- DeepLab V3+ [Chen et al., 2018]
- Masked Autoencoder (MAE) [He et al., 2021]

U-Net [Ronneberger et al., 2015]

- A Fully Convolutional Network (FCN) with a symmetric architecture.
- Encoder uses max pooling to shrink channel size
- Decoding part uses skip connections from the encoder

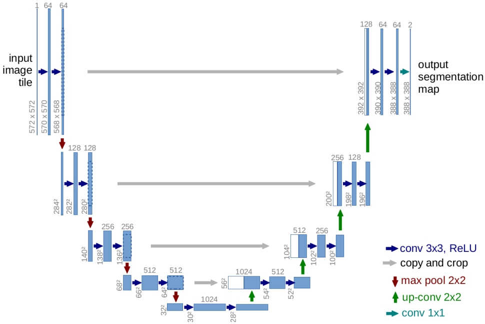


Figure: Original U-Net architecture

Masked Autoencoder (MAE)

The masked autoencoder (MAE) [He et al., 2021]

- Scalable self-supervised learner
- Trained to fill-in missing parts of an image

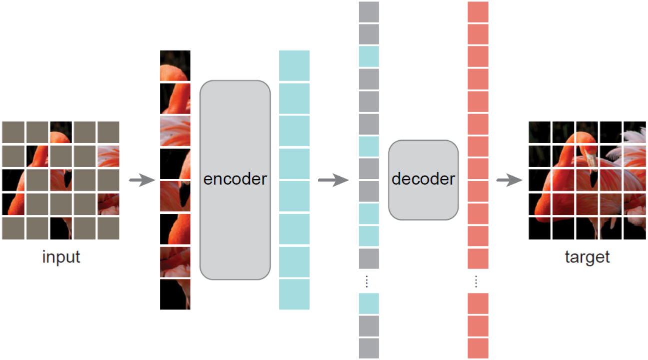


Figure: MAE architecture [He et al., 2021]

Salt and Pepper Noise Example

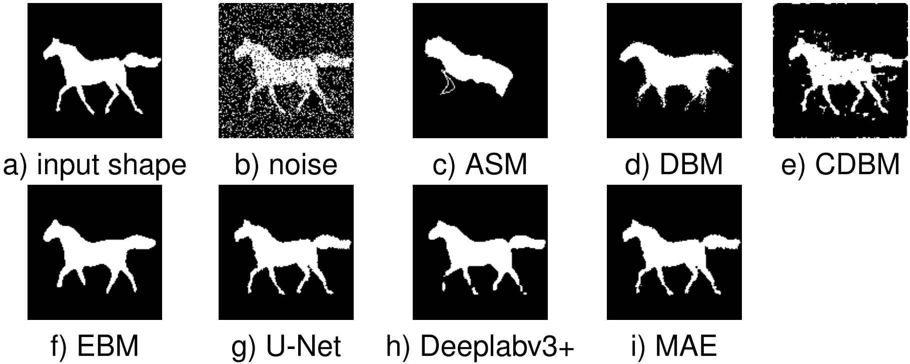


Figure: Example of results of different methods on a shape perturbed by salt and pepper noise.

Salt and Pepper Noise Results

- Performance comparison of all methods against salt and pepper noise

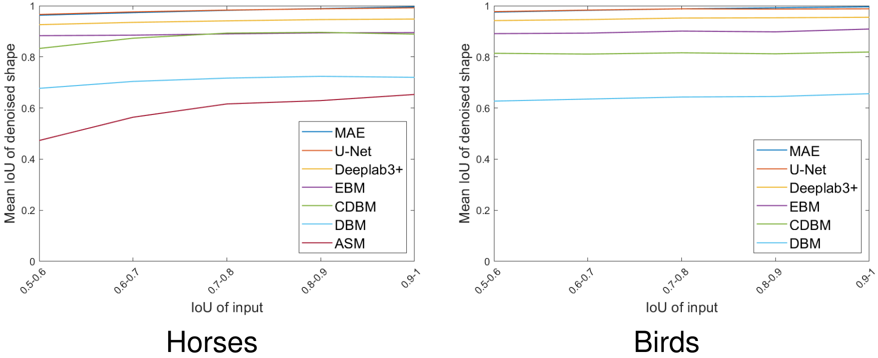


Figure: Performance on salt and pepper noise data S_{test}^{salt}.

Circle Noise Example

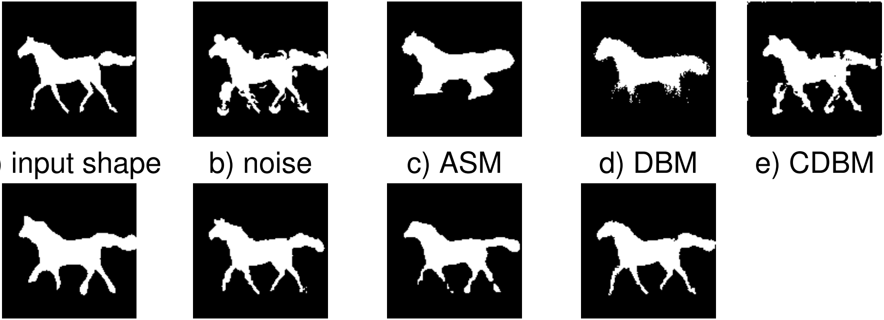
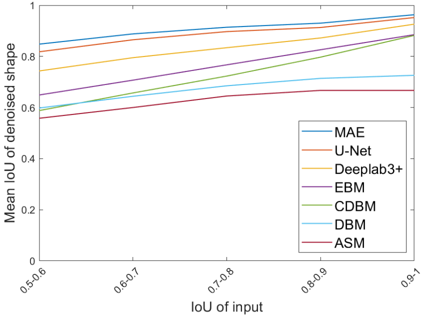


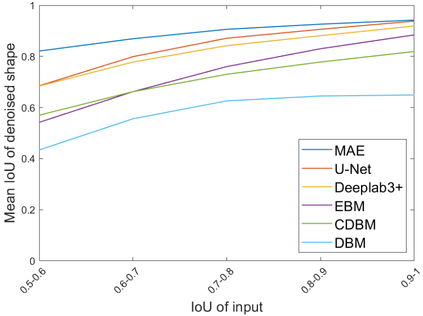
Figure: Example of results of different methods on a shape perturbed by circle noise.

Circle Noise Results

- Performance comparison of all methods against circle noise



Horses



Birds

Figure: Performance on circle noise data S_{test}^{circle}.

Real Image Noise Example

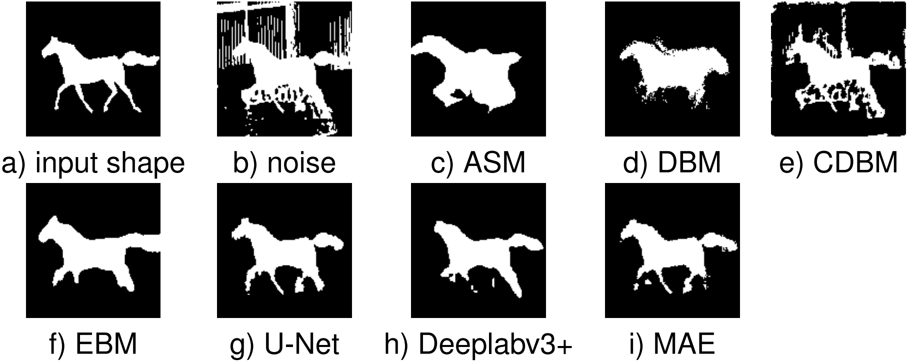
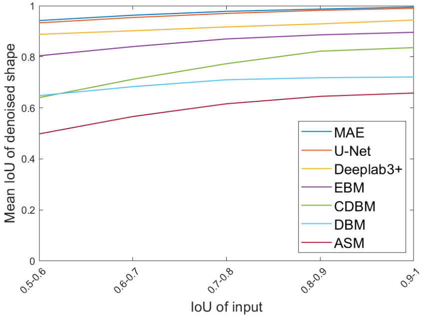
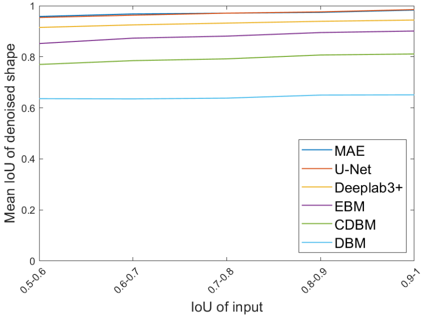


Figure: Example of results of different methods on a shape perturbed by real image noise.

- Performance comparison of all methods against real image noise



Horses



Birds

Figure: Performance on real image noise data S_{test}^{real}.

Occlusion Noise Example

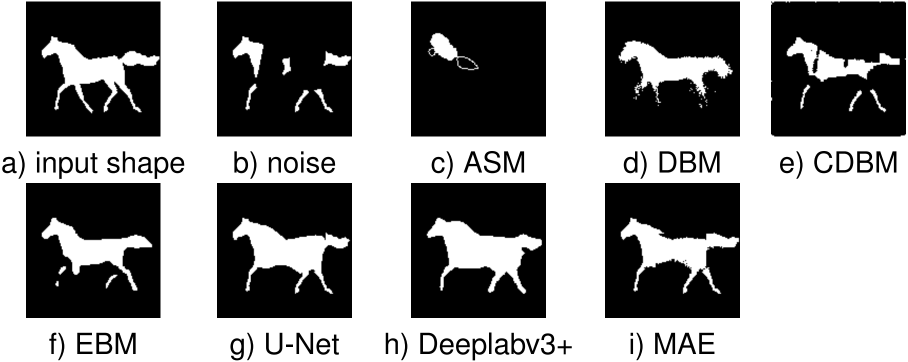
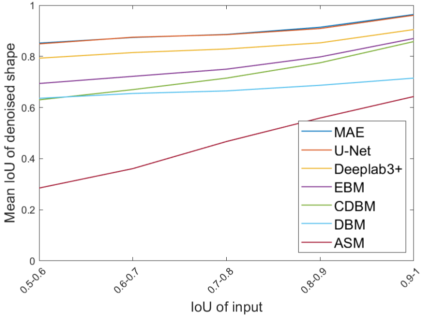


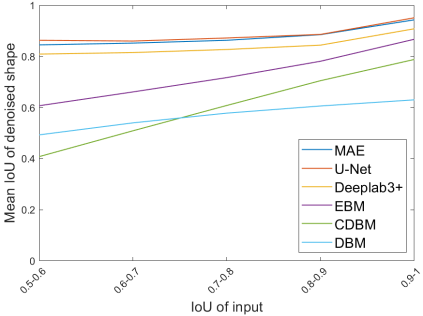
Figure: Example of results of different methods on a shape perturbed by occlusion noise.

Occlusion Noise Results

- Performance comparison of all methods against occlusion noise



Horses



Birds

Figure: Performance on real image noise data $S_{test}^{occlusion}$.

Detection Image Noise Example

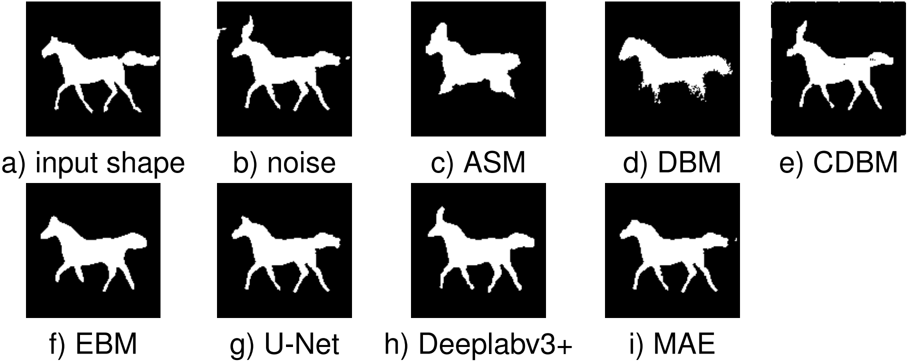
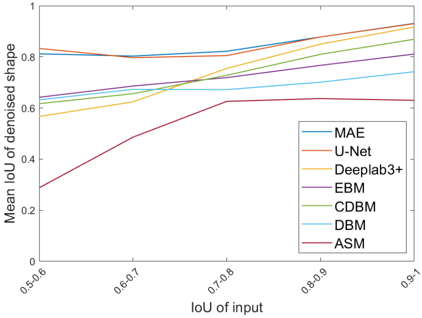


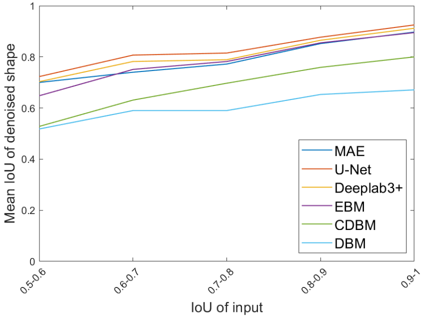
Figure: Example of results of different methods on a shape perturbed by detection image noise.

Detection Image Noise Results

- Performance comparison of all methods against detection image noise



Horses



Birds

Figure: Performance on real image noise data $S_{test}^{detection}$.

Thresholded Probability Noise Example

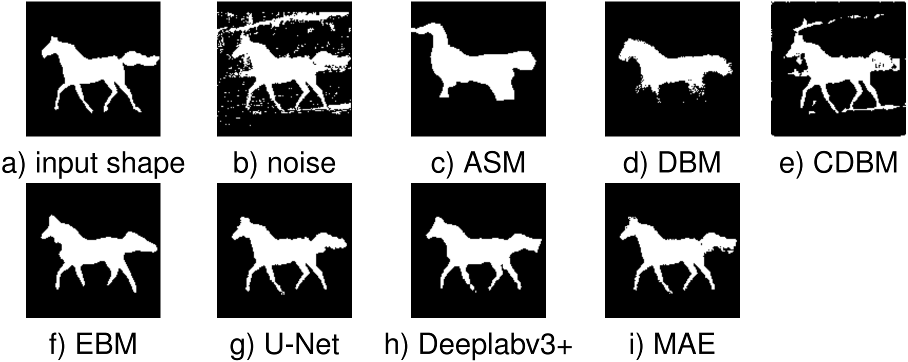


Figure: Example of results of different methods on a shape perturbed by thresholded probability noise.

Thresholded Probability Noise Results

- Performance comparison of each methods against thresholded probability noise

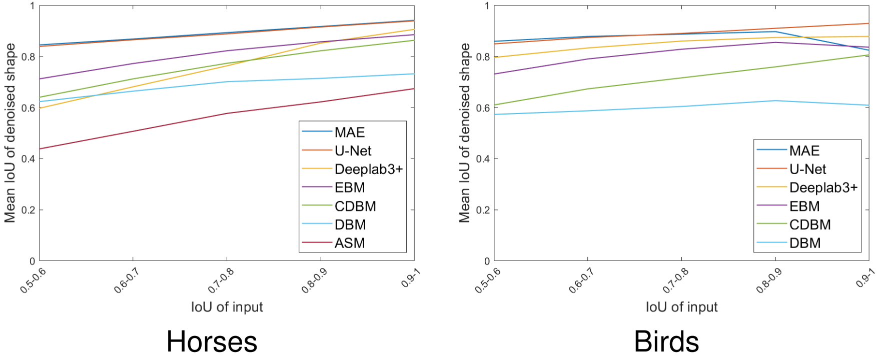


Figure: Performance on thresholded probability noise data $S_{test}^{probability}$.

Summary and Discussion

Comparing the methods:

- Experiments reveal that MAE and U-Net are the best shape denoising methods we evaluated for all six types of noise.
- DeepLabv3+ is the third best shape denoising method for the six noise types in most situations.
- EBM outperforms CDBM on all six noise types, especially when dealing with real image noise.

Comparing the noise types:

- The salt and pepper noise is the easiest to deal with, followed by real image noise.
- Circle noise and occlusion noise are more challenging than the above two, especially when the noise level is high.
- The most challenging noises among these six are the thresholded probability noise and detection image noise.

Conclusion and Future Work

We presented a study of methods for shape denoising

- Shapes are represented as binary images
- Shapes were aligned by translation and scaling
- We introduced six types of shape noise
- We evaluated seven shape modeling/denoising methods on these types of noise

Future work:

- Study shape denoising in the wild, where the shapes are not aligned
- Study methods trained on aligned shapes vs. methods trained on unaligned shapes

Borenstein, E., Sharon, E., and Ullman, S. (2004).
Combining top-down and bottom-up segmentation.
In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 46–46. IEEE.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018).
Encoder-decoder with atrous separable convolution for semantic image segmentation.
In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.

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

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. (2021).
Masked autoencoders are scalable vision learners.
arXiv preprint arXiv:2111.06377.

Pang, B., Han, T., Nijkamp, E., Zhu, S.-C., and Wu, Y. N. (2020).
Learning latent space energy-based prior model.
arXiv preprint arXiv:2006.08205.

Ronneberger, O., Fischer, P., and Brox, T. (2015).
U-net: Convolutional networks for biomedical image segmentation.

In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Salakhutdinov, R. and Hinton, G. (2009).
Deep boltzmann machines.
In *Artificial intelligence and statistics*, pages 448–455. PMLR.

Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., and Perona, P. (2010).
Caltech-UCSD Birds 200.
Technical Report CNS-TR-2010-001, California Institute of Technology.

Yang, J., Liu, S., and Wang, X. (2021).
Centered convolutional deep boltzmann machine for 2d shape modeling.
Personal and Ubiquitous Computing, pages 1–11.