Supplemental Material for Active Decision Boundary Annotation with Deep Generative Models

Miriam Huijser
Aiir Innovations
Amsterdam, The Netherlands
https://aiir.nl/

Jan C. van Gemert
Delft University of Technology
Delft, The Netherlands
http://jvgemert.github.io/

1. Projecting query point on decision boundary

Projecting query point $\mathbf{z}^*$ on the decision boundary parameterized by $\mathbf{w}$ and b yields $\mathbf{z}^p$, see figure 1.

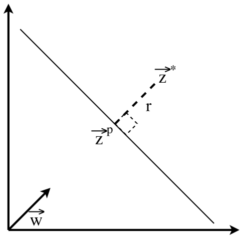

Figure 1: $\mathbf{z}^p$ is the projection of query point $\mathbf{z}^*$ on the decision boundary.

The projected point is defined as

$$\mathbf{z}^p = \mathbf{z}^* - yr\frac{\mathbf{w}}{|\mathbf{w}|}, \tag{1}$$

where y is the class label of $\mathbf{z}^*$. Furthermore, $\mathbf{z}^p$ lies on the decision boundary and thus satisfies

$$\mathbf{w}^\intercal \mathbf{z}^p + b = 0. \tag{2}$$

Substituting Eq. (1) in Eq. (2) yields:

$$\mathbf{w}^\intercal \left(\mathbf{z}^* - yr\frac{\mathbf{w}}{|\mathbf{w}|} \right) + b = 0. \tag{3}$$

Now, we solve for r:

$$r = \frac{\mathbf{w}^\intercal \mathbf{z}^* + b}{y|\mathbf{w}|}. \tag{4}$$

Substituting r from Eq. (4) in Eq. (1) gives:

$$\begin{aligned} \mathbf{z}^p &= \mathbf{z}^* - y\frac{\mathbf{w}^\intercal \mathbf{z}^* + b}{y|\mathbf{w}|} \cdot \frac{\mathbf{w}}{|\mathbf{w}|} \\ &= \mathbf{z}^* - \frac{(\mathbf{w}^\intercal \mathbf{z}^* + b)\mathbf{w}}{|\mathbf{w}||\mathbf{w}|} \\ &= \mathbf{z}^* - \frac{(\mathbf{w}^\intercal \mathbf{z}^* + b)\mathbf{w}}{\mathbf{w}^\intercal \mathbf{w}}. \end{aligned} \tag{5}$$

2. Results measured in Average Precision

As requested by one of our anonymous reviewer we also evaluate our results measured in Average Precision, see Tables 1, 2, 3, 4.

Experiment 5: Full dataset evaluation

	Sample	Boundary (ours)
MNIST	99.78 ± 0.02	**99.84 ± 0.03**
SVHN	90.65 ± 0.19	91.22 ± 1.95
Shoe-Bag	98.89 ± 0.18	**99.34 ± 0.13**

Table 4: Average Precision results for sample-based active learning and boundary active learning for all datasets averaged over 150 queries (maximum possible score is 150), averaged over all class pairs. The experiments are repeated 5 times and significant results are shown in bold. Significance is measured with a paired t-test with p $<$ 0.05.

Experiment 1: Evaluating various query strategies (average precision)

	MNIST 0 vs. 8		SVHN 0 vs. 8		Shoe-Bag	
Strategy	Sample	Boundary (ours)	Sample	Boundary (ours)	Sample	Boundary (ours)
Uncertainty	98.9 ± 0.2	**99.4 ± 0.2**	91.1 ± 0.8	**93.8 ± 0.7**	98.9 ± 0.2	**99.3 ± 0.1**
Uncertainty-dense	94.5 ± 11.0	**96.5 ± 10.7**	77.1 ± 5.7	**89.4 ± 2.1**	84.6 ± 4.0	**96.9 ± 1.1**
5 Cluster centroid	98.6 ± 0.1	**99.6 ± 0.03**	75.5 ± 5.4	**83.1 ± 1.2**	95.1 ± 0.8	**99.3 ± 0.1**
Random	98.4 ± 0.4	**99.2 ± 0.3**	89.3 ± 1.4	**93.7 ± 0.8**	98.1 ± 0.4	**99.3 ± 0.1**

Table 1: Average precision averaged over 150 iterations.

Experiment 3: Evaluating annotation noise

Sampling noise	MNIST 0 vs. 8		SVHN 0 vs. 8		Shoe-Bag	
(# images)	Sample	Boundary (ours)	Sample	Boundary (ours)	Sample	Boundary (ours)
0	99.0 ± 0.2	**99.5 ± 0.1**	90.9 ± 1.4	**93.7 ± 0.8**	98.9 ± 0.2	**99.3 ± 0.2**
1	99.0 ± 0.2	**99.5 ± 0.1**	90.9 ± 1.4	**93.4 ± 0.7**	98.9 ± 0.2	**99.3 ± 0.2**
2	99.0 ± 0.2	**99.4 ± 0.2**	90.9 ± 1.4	**92.3 ± 1.6**	98.9 ± 0.2	99.1 ± 0.6
3	99.0 ± 0.2	**99.3 ± 0.1**	90.9 ± 1.4	**92.5 ± 0.8**	98.9 ± 0.2	99.1 ± 0.6
4	99.0 ± 0.2	**99.1 ± 0.1**	90.9 ± 1.4	91.3 ± 0.9	98.9 ± 0.2	**99.1 ± 0.3**
5	99.0 ± 0.2	99.0 ± 0.1	90.9 ± 1.4	86.7 ± 9.7	98.9 ± 0.2	98.8 ± 0.4

Table 2: Average Precision results for noisy boundary active learning with uncertainty sampling for MNIST (classifying 0 and 8), SVHN (classifying 0 and 8) and Handbags vs. Shoes averaged over 150 queries (maximum possible score is 150). Each experiment is repeated 15 times. For each row, the significantly best result is shown in bold, where significance is measured with a paired t-test with $p < 0.05$. Noise has been added to the boundary annotation points; not to the image labels.

Experiment 4: Evaluating a human oracle

	MNIST 0 vs. 8		SVHN 0 vs. 8		Shoe-Bag	
Annotation	Sample	Boundary (ours)	Sample	Boundary (ours)	Sample	Boundary (ours)
Human oracle	-	-	63.0 ± 7.2	64.6 ± 7.6	-	-
SVM oracle	94.7 ± 2.5	**96.1 ± 2.7**	74.0 ± 5.8	74.9 ± 5.4	94.9 ± 1.5	**95.8 ± 1.6**

Table 3: Average Precision results for a human and a SVM oracle for sample-based active learning and our boundary active learning for MNIST (classifying 0 and 8), SVHN (classifying 0 and 8) and Shoe-Bag averaged over 10 queries (maximum possible score is 10). The experiments are repeated 15 times and significant results per row are shown in bold for $p < 0.05$.