A Appendix

A.1 Dataset Details.

CIFAR-10 [6], CIFAR-100 [6], and LSUN (FIX) [7, 10] are used for the one-class classification task. CIFAR-10 and CIFAR-100 consist of 50,000 training and 10,000 test images with 10 and 20 (superclass) image classes, respectively. LSUN (FIX) is the fixed version [10] of the testing set of LSUN [7], consisting of 10,000 images of 10 different scenes, in which 8,000 and 2,000 images are used for training and testing, respectively.

SVHN [8] and ImageNet (FIX) [3, 5, 10] are used for cross-dataset anomaly detection with an auxiliary anomaly set. Specifically, SVHN consists of 26,032 test images with 10 digits, and ImageNet (FIX) [10] consists of 10,000 test images with 200 classes from a subset of the full ImageNet dataset [3].

A.2 Data Augmentation Details

We follow SimCLR [2] augmentations, including Inception crop [9], horizontal flip, color jitter, and grayscale for random augmentations, as well as the rotation as shifting transformation used in CSI [10]. The details of each type of transformation are demonstrated as follows.

Inception Crop. We randomly crop the area of each original training image with the uniform distribution from 0.08 to 1.0 and make a random aspect ratio with 3/4 to 4/3 of the original aspect ratio. After the crop, cropped images are resized to the original image size.

Horizontal Flip. We flip each image horizontally with 50% of probability.

Color Jitter. We make a distortion of the hue, brightness, and saturation of each image. Specifically, we transform the RGB (red, green, blue) color space into the HSV (hue, saturation, value) color space and add noise to the HSV channels. Then, we apply color jitter with 80% of probability.

Grayscale. We randomly convert the image into a grayscale image with 20% of probability.

Rotation. Like CSI, we use a random rotation from {0°, 90°, 180°, 270°} as the shifting transformation.

A.3 Evaluation Metrics.

For evaluation, we measure the effectiveness of the proposed normality score in distinguishing in- and out-of-distribution images with **Area Under the Receiver Operating Characteristic curve (AUROC)**. Let TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. The ROC curve is a graph plotting the true positive rate = TP / (TP+FN) against the false positive rate = FP / (FP+TN) by varying a threshold.

Table 1. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-10 with $\gamma_l = 0.01$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 90%, which implies the hard pair.

Class	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
Plane	-	**81.7**	97.0	99.2	97.7	99.1	98.9	95.0	**78.7**	**87.9**	92.8
Car	99.4	-	100.0	100.0	100.0	100.0	100.0	99.9	99.0	94.6	99.2
Bird	93.1	99.1	-	97.3	90.4	92.7	95.7	92.0	98.5	99.7	95.4
Cat	97.4	98.0	93.4	-	90.8	**67.2**	91.9	**87.8**	98.5	98.4	91.5
Deer	98.6	99.9	95.4	97.6	-	96.2	99.1	**74.7**	99.4	99.8	95.6
Dog	99.5	99.7	97.0	91.2	92.8	-	97.9	**85.3**	99.7	99.6	95.8
Frog	99.1	98.8	97.3	97.4	98.8	97.8	-	99.4	99.0	99.8	98.6
Horse	99.5	99.7	99.1	99.4	94.2	97.3	99.8	-	99.8	99.6	98.7
Ship	96.1	94.3	99.7	99.8	99.7	99.8	99.8	99.7	-	97.3	98.5
Truck	97.6	**84.3**	99.9	99.9	99.9	99.9	99.9	99.5	97.4	-	97.6

Table 2. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-10 with $\gamma_l = 0.05$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 90%, which implies the hard pair.

Class	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
Plane	-	95.3	97.9	99.6	99.0	99.7	99.6	98.3	**87.4**	96.8	97.1
Car	99.8	-	100.0	100.0	100.0	100.0	100.0	100.0	99.5	96.4	99.5
Bird	95.5	99.7	-	97.8	93.1	96.2	97.0	97.0	99.4	99.9	97.3
Cat	98.8	99.4	95.0	-	94.9	**70.2**	94.0	93.7	99.4	99.5	93.9
Deer	99.6	99.9	96.6	97.9	-	97.1	98.9	**84.1**	99.9	99.9	97.1
Dog	99.7	99.8	98.1	91.7	95.8	-	98.9	92.1	99.8	99.8	99.2
Frog	99.7	99.7	98.1	98.1	99.1	98.9	-	99.6	99.7	99.9	99.2
Horse	99.8	100.0	99.5	99.5	96.1	98.2	99.9	-	100.0	99.9	99.2
Ship	97.7	98.4	99.9	100.0	99.9	100.0	100.0	99.9	-	99.2	99.4
Truck	98.9	92.9	100.0	100.0	100.0	100.0	100.0	99.9	99.0	-	99.0

A.4 Detailed OOD Detection Results

For the one-class classification task under scenario-1 and scenario-2, We report the results of each individual normal class on CIFAR-10, CIFAR-100, and LSUN (FIX) from Table 1 to Table 9. The detailed analysis is demonstrated as follows.

Table 1 presents the confusion matrix of AUROC values of HSCL with a labeled ratio $\gamma_l = 0.01$ on CIFAR-10, where bold denotes the hard pairs with AUROC score less than 90%. The results align with the human intuition that classes from the same superclass are likely to be confused with each other. For example, "Car", "Ship", "Plane", and "Truck" classes of "Vehicle" superclass, as well as "Cat", "Dog", "Horse", and "Deer" classes of "Animal" superclass are easy to be confused to each other.

Table 2 and Table 3 present the confusion matrix of AUROC values of HSCL with labeled ratios $\gamma_l = 0.05$ and $\gamma_l = 0.10$ on CIFAR-10, respectively. With the

Table 3. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-10 with $\gamma_l = 0.10$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 90%, which implies the hard pair.

Class	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
Plane	-	98.1	98.5	99.8	99.5	99.9	99.8	99.2	91.9	98.5	98.4
Car	99.9	-	100.0	100.0	100.0	100.0	100.0	100.0	99.7	97.4	99.7
Bird	96.1	99.9	-	98.6	95.7	98.1	97.9	98.5	99.6	100.0	98.3
Cat	99.2	99.7	96.1	-	96.3	**74.0**	95.6	95.7	99.6	99.7	95.1
Deer	99.8	100.0	97.1	98.2	-	98.0	99.2	**88.5**	99.9	100.0	97.8
Dog	99.7	99.9	98.2	91.8	96.8	-	98.9	94.2	99.9	99.9	97.7
Frog	99.8	99.9	98.5	98.8	99.2	99.4	-	99.8	99.9	99.9	99.5
Horse	99.9	100.0	99.5	99.6	97.0	98.6	99.9	-	100.0	100.0	99.4
Ship	98.3	99.3	99.9	99.9	99.9	100.0	100.0	99.9	-	99.4	99.6
Truck	99.3	95.5	100.0	100.0	100.0	100.0	100.0	99.9	99.4	-	99.3

Table 4. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-10 with contamination ratio $\gamma_p = 0.05$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 90%, which implies the hard pair.

Class	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
Plane	-	95.2	98.0	99.8	99.2	99.8	99.7	98.9	**85.6**	97.5	97.1
Car	99.8	-	100.0	100.0	100.0	100.0	100.0	100.0	99.3	96.1	99.5
Bird	94.2	99.5	-	97.2	92.2	95.5	96.4	96.8	98.7	99.9	96.7
Cat	98.5	99.2	93.0	-	92.8	**74.0**	**85.6**	93.9	99.0	99.4	92.8
Deer	99.6	99.8	95.6	97.7	-	97.2	96.4	**86.0**	99.7	99.8	96.9
Dog	99.2	99.7	97.0	89.3	97.2	-	97.9	94.1	99.6	99.6	97.1
Frog	99.6	99.6	97.6	97.4	97.4	98.5	-	99.3	99.5	99.8	98.8
Horse	99.8	99.9	99.3	99.4	95.4	97.9	99.9	-	100.0	99.9	99.1
Ship	96.5	98.4	99.8	99.9	99.8	100.0	100.0	99.9	-	98.9	99.2
Truck	98.9	92.3	100.0	100.0	100.0	100.0	100.0	99.9	98.9	-	98.9

increase of the labeled ratio, only some extreme hard pairs, *i.e.*, "Plane-Ship", "Cat-Dog" and "Deer-Horse" have AUROC scores less than 90%.

Table 4 and Table 5 present the confusion matrix of AUROC values of HSCL with contamination ratios $\gamma_p = 0.05$ and $\gamma_p = 0.10$ on CIFAR-10, respectively. With the increase of the contamination ratio, some extreme hard pairs, *i.e.*, "Plane-Ship", "Cat-Dog", "Cat-Frog", "Deer-Horse", and "Truck-Car", have degraded performance.

Table 6 and Table 7 present the confusion matrix of AUROC values of HSCL with contamination ratios $\gamma_p = 0.05$ and $\gamma_p = 0.10$ on LSUN (FIX), respectively. The class index from 1 to 10 represents "Bedroom", "Kitchen", "Living room", "Dining room", "Bridge", "Tower", "Restaurant", "Conference room", "Classroom", and "Church outdoor", respectively. Unlike CIFAR-10, each scene

Table 5. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-10 with contamination ratio $\gamma_p = 0.10$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 90%, which implies the hard pair.

Class	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
Plane	-	89.7	97.5	99.7	99.3	99.5	99.7	98.9	**83.5**	94.1	95.8
Car	99.7	-	100.0	100.0	100.0	100.0	100.0	100.0	99.2	95.6	99.4
Bird	95.0	99.7	-	97.0	90.6	93.7	94.6	93.9	99.5	100.0	96.0
Cat	99.1	99.5	94.0	-	93.6	**65.7**	90.9	93.5	99.5	99.8	92.8
Deer	99.8	100.0	96.4	98.1	-	97.1	98.5	**80.6**	99.9	100.0	96.7
Dog	99.7	99.9	97.4	90.7	95.4	-	98.1	91.3	99.9	99.9	96.9
Frog	99.6	99.8	97.9	97.9	98.8	98.4	-	99.5	99.8	99.9	99.1
Horse	99.8	100.0	99.1	99.4	94.9	97.1	99.8	-	100.0	100.0	98.9
Ship	96.7	96.5	99.8	100.0	99.9	100.0	100.0	99.9	-	97.8	98.9
Truck	98.7	**88.4**	100.0	100.0	100.0	100.0	100.0	100.0	98.4	-	98.4

Table 6. Confusion matrix of AUROC (%) of our HSCL for one-class classification on LSUN (FIX) with contamination ratio $\gamma_p = 0.05$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 80%, which implies the hard pair.

Class	1	2	3	4	5	6	7	8	9	10	Mean
1	-	95.6	93.4	94.7	81.2	87.5	81.1	**72.7**	95.6	95.9	88.6
2	99.3	-	91.3	98.7	97.7	99.1	99.3	99.2	99.0	**76.0**	95.5
3	98.1	**77.3**	-	96.2	90.1	98.4	99.3	98.1	95.3	**74.3**	91.9
4	92.5	92.0	92.2	-	**79.4**	89.2	92.5	88.5	80.8	94.8	89.1
5	84.1	94.5	90.9	82.0	-	87.6	88.0	84.8	84.4	92.6	87.7
6	82.9	95.9	95.9	85.5	**76.7**	-	**74.1**	**68.8**	**74.0**	97.3	83.5
7	80.2	96.8	96.8	93.8	**75.4**	**76.7**	-	**75.6**	88.2	98.0	86.8
8	**78.8**	95.1	95.3	87.2	**76.3**	**74.5**	**74.7**	-	88.3	96.7	84.1
9	92.5	94.0	90.2	**72.7**	**70.9**	**76.8**	88.1	80.0	-	95.0	84.5
10	99.4	**63.1**	88.7	99.1	98.1	99.5	99.6	99.4	99.1	-	94.0

category of LSUN (FIX) has much diversity, which makes different categories easy to get confused with each other.

Table 8 and Table 9 present the confusion matrix of AUROC values of HSCL with contamination ratios $\gamma_p = 0.05$ and $\gamma_p = 0.10$ on CIFAR-100, respectively. The superclass index from 1 to 20 represents "Aquatic mammals", "Fish", "Flowers", "Food containers", "Fruit and vegetables", "Household electrical devices", "Household furniture", "Insects", "Large carnivores", "Large man-made outdoor things", "Large man-made outdoor things", "Large man-made outdoor things", "Medium-sized mammals", "Non-insect invertebrates", "People", "Reptiles", "Small mammals", "Trees", "Vehicles 1", and "Vehicles 2", respectively. Since each superclass contains 5 categories, it enlarges the intra-class diversity, which also makes different superclasses easy to get confused with each other.

Table 7. Confusion matrix of AUROC (%) of our HSCL for one-class classification on LSUN (FIX) with contamination ratio $\gamma_p = 0.10$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 80%, which implies the hard pair.

Class	1	2	3	4	5	6	7	8	9	10	Mean
1	-	96.1	92.5	96.3	84.8	89.0	83.7	**76.1**	96.0	95.5	90.0
2	99.3	-	90.7	98.0	96.9	98.9	99.3	99.0	98.1	**76.3**	95.2
3	98.9	**77.5**	-	97.0	93.2	98.8	99.5	98.8	95.7	**70.3**	92.2
4	94.9	92.9	93.0	-	**78.5**	90.7	93.8	89.6	82.0	95.0	90.1
5	84.7	94.6	90.5	81.1	-	88.1	88.5	85.7	83.6	91.5	87.6
6	85.2	95.7	96.0	87.8	**79.3**	-	**75.1**	**68.9**	**75.0**	97.2	84.5
7	81.7	96.9	97.2	92.3	**75.5**	**75.5**	-	**71.6**	86.0	98.5	86.1
8	**72.1**	96.2	96.5	87.4	**78.3**	**73.9**	**76.1**	-	88.4	97.4	85.1
9	93.4	94.3	90.6	**72.9**	**71.1**	**77.0**	88.0	**79.6**	-	94.4	84.6
10	99.7	**61.1**	88.5	99.5	98.8	99.7	99.9	99.8	99.3	-	94.0

A.5 Compare with PU-Learning

Positive unlabeled (PU) learning is a special case of semi-supervised learning, where the training dataset contains a few positive labeled samples and a large number of unlabeled samples. Unlike PU-learning, our setting assumes both a few positive (normal) and negative (abnormal) samples are available. We compared with one PU-learning method, VPU [1], on CIFAR-10 with the same setting as our method using ResNet-18. The results are shown in Table 10. Since the proposed method can take the advantage of abnormal samples, it achieves better performance.

A.6 Ablation Study of Different Data Combinations

We conduct experiments on the "Plane" Class of CIFAR-10 to see the influence of different data combinations. Since we focus on the semi-supervised setting with contaminated samples, we vary the percentage of the labeled normal (LN), labeled abnormal (LA), and contaminated unlabeled abnormal (CUA) samples from 5% to 10%. The results are shown in Table 11. We can see that labeled abnormal samples have more influence than labeled normal samples. Since abnormal samples are rare in training, increasing the percentage of LA can provide more discrimination information.

A.7 Efficiency Analysis

In the training stage, the proposed HSCL uses 250 epochs, while Elsa [4] requires 500 epochs and CSI requires 1,000 epochs. In addition, Elsa needs an extra fine-tuning stage with 50 epochs and a prototype selection stage that employs spherical k-means clustering with additional complexity $O(T \cdot K \cdot N \cdot D)$, where T is the number of iteration, K is the number of clusters, N is the number of

Table 8. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-100 with contamination ratio $\gamma_p = 0.05$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 80%, which implies the hard pair.

Supclass	1 11	2 12	3 13	4 14	5 15	6 16	7 17	8 18	9 19	10 20	Mean
	-	85.7	97.4	94.1	96.2	97.6	96.9	93.8	85.1	88.3	
1	81.3	82.1	87.1	89.4	95.8	83.0	87.5	92.8	93.3	92.9	90.5
	85.1	-	92.7	94.9	94.3	95.6	95.2	92.5	93.7	85.8	
2	**76.6**	91.9	92.5	88.9	95.9	83.9	91.4	82.3	92.6	93.3	90.5
	98.9	97.3	-	98.2	91.8	99.1	98.7	95.9	97.6	98.6	
3	95.1	97.4	98.0	97.4	97.4	98.7	98.0	94.6	98.5	98.8	97.4
	91.4	94.0	94.0	-	87.6	81.4	86.8	95.4	95.5	89.2	
4	93.9	89.8	96.0	92.4	91.8	89.9	94.4	95.4	89.9	88.6	91.4
	97.9	96.6	91.5	95.2	-	97.1	97.9	98.0	98.1	98.4	
5	97.8	94.9	98.0	97.0	96.7	97.7	97.5	98.2	98.9	98.7	97.2
	96.2	96.1	97.8	83.9	94.0	-	**75.5**	97.6	98.9	84.7	
6	93.3	93.1	98.0	95.8	94.6	92.6	94.0	97.2	83.3	85.1	92.2
	96.3	94.5	96.1	90.7	95.5	81.6	-	95.5	97.4	88.6	
7	94.0	93.0	96.8	93.6	95.4	94.7	94.0	97.4	87.2	88.6	93.2
	87.4	92.2	86.7	91.1	90.9	95.7	94.5	-	88.5	92.3	
8	91.0	89.2	90.0	83.3	95.4	89.1	89.6	90.6	90.0	91.1	90.5
	88.0	97.3	97.0	96.9	97.2	99.3	98.2	95.4	-	95.6	
9	93.1	89.5	89.9	95.8	96.5	96.7	92.0	95.7	97.9	97.7	95.2
	92.4	95.3	99.3	96.9	99.0	95.9	96.2	98.2	96.3	-	
10	84.0	93.3	97.0	97.5	98.4	97.0	91.6	91.5	88.5	85.8	94.4
	96.5	96.7	99.1	99.3	99.7	99.3	99.4	99.3	97.0	90.3	
11	-	97.5	98.2	98.6	99.1	98.3	98.1	95.8	98.5	98.0	97.8
	80.6	91.5	94.6	92.6	93.6	94.1	95.1	94.2	86.6	87.2	
12	86.5	-	89.8	92.7	90.8	92.6	91.3	92.9	90.7	91.2	91
	87.8	94.4	95.8	95.5	96.7	96.9	97.4	93.9	86.2	96.3	
13	96.0	88.9	-	93.8	**73.8**	94.4	88.5	94.7	97.1	97.3	92.9
	79.8	82.4	88.8	85.0	88.5	92.8	89.0	**74.8**	85.0	88.0	
14	81.3	87.0	86.2	-	93.4	**77.5**	84.7	83.8	85.9	88.7	85.4
	95.9	96.3	93.2	95.8	94.7	95.3	97.3	97.4	95.2	97.8	
15	97.5	91.2	86.1	96.9	-	96.7	94.3	99.7	98.1	98.2	95.7
	74.5	**72.6**	93.2	84.6	91.9	90.7	91.5	86.4	90.0	88.7	
16	80.9	88.6	88.1	**77.0**	93.1	-	88.5	88.9	90.0	90.0	86.8
	76.5	89.3	90.5	93.6	92.5	97.0	94.1	88.7	**74.1**	**75.4**	
17	84.6	85.2	**78.7**	86.1	85.8	89.3	-	82.4	83.8	83.3	85.8
	99.2	98.4	99.2	99.7	99.5	99.7	99.8	99.7	99.4	98.4	
18	97.6	99.5	99.6	99.5	99.9	99.6	99.4	-	99.4	98.9	99.3
	99.1	99.3	99.8	99.1	99.9	97.4	97.6	99.7	99.7	91.0	
19	97.5	98.5	99.6	99.6	99.4	99.7	96.0	98.4	-	82.6	97.6
	97.1	97.8	99.3	97.1	99.3	96.2	96.3	98.6	99.2	86.7	
20	94.8	96.3	98.8	98.0	98.9	98.1	95.8	95.4	**75.2**	-	95.7

Table 9. Confusion matrix of AUROC (%) of our HSCL for one-class classification on CIFAR-100 with contamination ratio $\gamma_p = 0.10$. The last column shows the mean result over all abnormal classes. Bold denotes the values under 80%, which implies the hard pair.

	1	2	3	4	5	6	7	8	9	10	
Supclass	11	12	13	14	15	16	17	18	19	20	Mean
	-	86.9	97.0	93.5	96.0	97.4	96.8	94.1	83.8	86.9	
1	**79.6**	81.2	86.7	90.0	95.1	84.0	86.6	90.8	92.7	91.9	90.0
	83.7	-	92.5	94.3	93.6	95.5	95.5	92.8	93.3	84.6	
2	**73.6**	91.5	92.3	88.0	96.3	83.2	91.1	**76.6**	92.4	92.7	89.6
	98.5	96.8	-	97.5	88.3	98.8	98.6	96.0	97.4	98.3	
3	95.6	96.8	97.6	96.9	96.4	98.0	97.4	96.0	98.4	98.8	97.0
	89.5	93.3	93.5	-	85.5	82.2	86.4	95.0	94.9	86.0	
4	91.8	87.4	95.7	91.5	91.2	90.0	93.5	94.0	86.7	85.4	90.2
	97.8	96.8	91.2	95.4	-	97.1	98.0	98.0	98.1	98.3	
5	97.9	95.1	97.9	97.2	96.5	97.6	97.6	98.5	98.9	98.8	97.2
	94.9	95.0	97.7	83.9	93.6	-	**73.1**	97.5	98.7	80.7	
6	91.3	92.1	98.0	95.5	94.1	92.4	92.9	96.5	**79.8**	80.6	91.0
	95.8	94.6	95.8	91.8	95.6	83.2	-	95.2	97.3	88.2	
7	93.2	93.1	96.6	92.9	95.9	94.4	93.8	97.2	86.0	87.5	93.0
	83.1	89.9	85.1	90.3	90.6	95.1	93.1	-	85.0	89.2	
8	88.3	85.1	87.4	80.4	94.3	86.8	87.3	87.7	87.7	88.7	88.2
	86.7	97.0	97.4	96.9	97.4	99.3	98.3	96.2	-	95.5	
9	92.3	87.8	90.3	95.9	96.8	96.6	92.3	92.3	97.5	97.5	95.0
	91.6	95.8	99.5	97.2	99.1	96.8	96.7	98.5	95.6	-	
10	83.6	92.4	96.6	97.5	98.9	97.2	91.6	90.8	88.4	85.5	94.4
	94.3	96.3	99.0	98.8	99.5	98.7	98.8	99.0	95.2	87.0	
11	-	95.7	97.6	98.0	99.0	97.6	97.2	94.5	97.4	96.9	96.9
	79.1	91.6	93.9	92.9	94.4	94.7	95.1	94.3	84.5	87.2	
12	86.6	-	87.8	92.7	89.7	92.4	89.8	88.7	88.4	89.7	90.2
	86.9	95.1	95.0	96.2	96.4	97.8	98.0	94.8	86.1	96.5	
13	95.6	88.9	-	94.1	**72.5**	94.8	88.7	93.0	97.2	97.8	92.9
	72.5	**78.4**	82.4	87.8	84.8	95.5	90.4	**72.9**	**72.2**	88.1	
14	80.0	82.1	**79.0**	-	91.7	**73.9**	**78.6**	**64.8**	82.6	86.8	81.3
	94.6	96.0	92.3	95.0	93.0	94.9	96.9	97.3	95.2	97.5	
15	97.5	90.3	85.8	96.6	-	96.2	94.5	99.6	97.7	98.0	95.2
	73.6	**74.0**	91.1	83.8	90.2	90.8	91.5	86.6	89.5	86.7	
16	**77.3**	87.2	87.1	**77.1**	92.1	-	88.1	88.7	88.7	88.5	85.9
	76.0	87.7	92.5	90.7	93.1	90.0	88.4	90.5	80.3	**63.8**	
17	80.6	82.9	81.9	89.1	86.1	90.3	-	**76.3**	**71.8**	**71.7**	83.4
	99.0	98.4	99.1	99.6	99.5	99.7	99.8	99.7	99.3	97.8	
18	97.1	99.3	99.5	99.5	99.8	99.6	99.3	-	99.0	98.6	99.1
	98.7	99.2	99.5	99.2	99.8	98.7	98.6	99.4	99.4	92.9	
19	97.1	98.3	99.4	99.1	99.4	99.5	97.4	97.5	-	85.5	97.8
	97.0	97.3	99.3	97.3	99.5	96.0	95.7	99.1	99.0	84.5	
20	92.8	95.9	98.8	98.2	99.0	98.5	94.7	94.0	**72.6**	-	95.2

Table 10. Comparison with PU-learning on the "Plane" Class of CIFAR-10.

Method	Labeled (%)	AUROC
VPU	5%	74.0%
Ours	5% (with abnormal)	97.1%

Table 11. Comparison with different data combinations on Class 0 of CIFAR-10.

LN (%)	LA (%)	CUA (%)	AUROC
5%	5%	5%	97.1
10%	5%	5%	97.5
5%	10%	5%	98.4
5%	5%	10%	95.8

training samples, and D is the sample dimension. Moreover, in the inference stage, CSI also needs to use 1-nearest neighbor to calculate the normality score that requires $O(N \cdot D)$ complexity, while HSCL just needs $O(D)$ with the learned prototypes. Compared with CSI and Elsa, it shows the high efficiency of our HSCL framework.

A.8 Limitations

As shown in the results, our HSCL also has difficulty in distinguishing the hard pairs, where normal and abnormal classes are quite similar. In future work, we hope to overcome such limitations and learn discriminative representations to detect hard examples of anomalies.

A.9 Broader Impact

Our research models the complementary contrastive relations with semi-supervised learning for anomaly detection and achieves promising performance even without clean training data. The positive impact is obvious. The framework can be easily applied in many real-world tasks, such as detecting financial fraud, manufacturing inspection, autonomous driving, and medical diagnosis. Negative impacts of our research are difficult to predict, however, the method may cause overconfidence in extreme hard examples. Once an abnormal sample is misclassified as a normal sample, this may bring much trouble to the real situation. Such a phenomenon should be analyzed for most existing anomaly detection approaches.

References

1. Chen, H., Liu, F., Wang, Y., Zhao, L., Wu, H.: A variational approach for learning from positive and unlabeled data. Advances in Neural Information Processing Systems **33**, 14844–14854 (2020)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
4. Han, S., Song, H., Lee, S., Park, S., Cha, M.: Elsa: Energy-based learning for semi-supervised anomaly detection. arXiv preprint arXiv:2103.15296 (2021)
5. Hendrycks, D., Mazeika, M., Kadavath, S., Song, D.: Using self-supervised learning can improve model robustness and uncertainty. arXiv preprint arXiv:1906.12340 (2019)
6. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
7. Liang, S., Li, Y., Srikant, R.: Enhancing the reliability of out-of-distribution image detection in neural networks. arXiv preprint arXiv:1706.02690 (2017)
8. Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning (2011)
9. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
10. Tack, J., Mo, S., Jeong, J., Shin, J.: Csi: Novelty detection via contrastive learning on distributionally shifted instances. Advances in neural information processing systems **33**, 11839–11852 (2020)