A. Full Experimental Results on MVTec, Digits Dataset and Office-Home Dataset.

In this section, we present the full experiment results on MVTec AD, digits and Office-Home dataset in the form of tables and bar plots.

A.1. MVTec Dataset

The full results on MVTec dataset are presented in Table 1. The first row denotes the domain adaptation setting. For example, L→C denotes that the case that the source domain is "Leather" and the target domain is "Carpet". The full results are also plotted in Fig. 2.

A.2. Digits Anomaly Detection

The full experimental results of MNIST→USPS are shown in Table 2. The digit in the first column in each row is regarded as the normal class. Fig. 3 is the corresponding bar plot.

The full experimental results of MNIST→SVHN are shown in Table 3. The digit in the first column in each row is regarded as the normal class. The full results are also plotted in Fig. 4.

A.3. Objects Recognition Anomaly Detection

Full results of Product→Clipart are presented in Table 4. The object category in the first column in each row is regarded as the normal class. Recall that the evaluation metric is AUROC. These numbers are also plotted in Fig. 6.

Similarly, full results of Clipart→Product are presented in Table 5. The results are also plotted in Fig. 5.

B. Proof of Information-Theoretic Lower Bound:

The proof for Thm. 1 is as follows:

Proof. Y_S are defined as the labeling function from X to Y for the source domain and Y_T as the map for the target domain. We assume that Y is 1 when the data is anomalous and 0 otherwise. Since the JS distance is a metric, we have the following inequality:

$$d_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{Y_T}) \leq d_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{\hat{Y}}) + d_{\mathrm{JS}}(\mathcal{D}^{\hat{Y}}, \mathcal{D}^{Y_T}) \quad (7)$$

We define $\varepsilon_S(h \circ g) = \varepsilon_S(\hat{Y})$ as

$$\varepsilon_S(h \circ g) = \mathbb{E}_X(|Y_S(X) - h \circ g(X)|)$$

and similarly for $\varepsilon_T(h \circ g)$. We can bound $d_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{\hat{Y}})$ by $\sqrt{\varepsilon_S(h \circ g)}$ ([19, 34])

$$\begin{aligned}
&d_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{\hat{Y}}) = \sqrt{D_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{\hat{Y}})} \leq \sqrt{\frac{1}{2}\|\mathcal{D}^{Y_S} - \mathcal{D}^{\hat{Y}}\|_1} \\
&= \Big(\frac{1}{2}\big(\big|\Pr(Y_S = 0) - \Pr(\hat{Y} = 0)\big| \\
&+ \big|\Pr(Y_S = 1) - \Pr(\hat{Y} = 1)\big|\big)\Big)^{\frac{1}{2}} \\
&= \big|\Pr(Y_S = 1) - \Pr(\hat{Y} = 1)\big|^{\frac{1}{2}} = \big|\mathbb{E}_X(Y_S) - \mathbb{E}_X(\hat{Y})\big|^{\frac{1}{2}} \\
&\leq \varepsilon_S(h \circ g)^{\frac{1}{2}}
\end{aligned}$$

With Eq. (7), we get:

$$d_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{Y_T}) \leq \sqrt{\varepsilon_S} + \sqrt{\varepsilon_T} \quad (8)$$

This can be rewritten as:

$$\varepsilon_S\,(h \circ g) + \varepsilon_T\,(h \circ g) \geq \frac{1}{2}\, d_{\mathrm{JS}}(\mathcal{D}^{Y_S}, \mathcal{D}^{Y_T})^2$$

□

C. Proof of Generalization Upper Bound:

We start with introductions of notations and definitions. Recall

$$\tilde{\mathcal{H}} := \{\mathrm{sgn}\,(|h(\mathbf{x}) - h'(\mathbf{x})| - t) \,|\, h, h' \in \mathcal{H}, t \in [0, 1]\}$$

Let $\widehat{D}$ denote the empirical distribution from samples $x \sim \mathcal{D}$ of size n. The empirical Rademacher complexity is defined as follows [34]:

Definition 1 (Empirical Rademacher Complexity)**.** . *Let $\mathcal{H}$ be a family of functions mapping from X to $[a, b]$. Let $\mathbf{S} = \{\boldsymbol{x}_i\}_{i=1}^n$ denote a fixed sample of size n with elements in X. Then, the empirical Rademacher complexity of $\mathcal{H}$ with respect to the sample X is defined as:*

$$\mathrm{Rad}_{\mathbf{S}}(\mathcal{H}) := \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i h\left(\mathbf{x}_i\right)\right]$$

where $\sigma = \{\sigma_i\}_{i=1}^n$ and σ_i are i.i.d. uniform random variables taking values in $\{+1, -1\}$.

We then have the following lemmas [34]:

Lemma 1. *Let $\mathcal{H} \subseteq [0, 1]^X$, then for all $\delta > 0$, w.p. at least $1 - \delta$, the following inequality holds for all $h \in \mathcal{H}$: $\varepsilon_S(h) \leq \widehat{\varepsilon}_S(h) + 2\,\mathrm{Rad}_{\mathbf{S}}(\mathcal{H}) + 3\sqrt{\log(2/\delta)/2n}$, where n is the number of samples in* $\mathbf{S}$.

Lemma 2. *$\forall\delta > 0$, w.p. at least $1 - \delta$, the following inequality holds:*

$$d_{\tilde{\mathcal{H}}}(\mathcal{D}, \widehat{\mathcal{D}}) \leq 2\,\mathrm{Rad}_{\mathbf{S}}(\tilde{\mathcal{H}}) + 3\sqrt{\log(2/\delta)/2n}.$$

Table 1. Experimental results on MVTec dataset. The metric is AUROC(%). The first column denotes the domain adaptation setting, e.g. L→C denotes that the source domain is "Leather" and the target domain is "Carpet". Results are also plotted in Fig. 2.

Model	DSVDD	AGT	BiOST	OCSVM (S+T)	OCSVM (T)	**IRAD (OC)**	IF (S+T)	IF (T)	**IRAD (I)**
L→C	47.5±2.9	48.1±1.2	47.2±1.9	16.8±2.2	17.0±3.2	44.0±2.1	47.1±1.8	47.4±2.5	**51.9±3.1**
C→L	45.6±4.1	41.3±2.9	46.5±1.4	32.2±2.3	33.9±2.8	37.6±2.8	35.9±1.4	36.3±3.4	**48.0±2.4**
L→W	58.1±2.4	59.1±2.8	60.0±3.7	52.5±1.3	56.5±4.5	58.1±3.2	54.9±2.7	58.6±3.1	**71.2±2.5**
W→L	42.3±2.5	41.3±2.9	42.6±2.4	34.9±2.7	33.9±2.8	38.3±2.4	35.9±2.4	36.3±3.4	**44.8±2.5**
W→C	46.3±1.9	48.1±1.2	48.9±0.8	19.4±2.8	17.0±3.2	48.2±2.6	43.8±1.9	47.4±2.5	**49.5±1.2**
C→W	56.9±3.0	59.1±2.8	55.7±1.6	55.5±2.9	56.5±4.5	57.7±1.9	56.1±2.2	58.6±3.1	**60.4±2.7**
Average	49.4	49.5	50.15	35.2	35.8	47.3	45.6	47.4	**54.3**

Table 2. Anomaly detection with domain adaptation on MNIST (source domain) and USPS (target domain). The evaluation metric is AUROC in percent. The highest numbers are in bold. Results are also plotted in Fig. 3.

Model	DSVDD	AGT	BiOST	OCSVM (S+T)	OCSVM (T)	**IRAD (OC)**	IF (S+T)	IF (T)	**IRAD (IF)**
0	36.5±10.1	94.3±1.7	75.9±4.2	74.5±0.6	13.5±0.8	94.8±2.6	22.9±1.4	95.1±0.4	**96.0±0.6**
1	74.3±7.1	**99.0±0.1**	97.6±1.8	4.0±0.2	13.6±0.5	97.5±1.5	97.3±0.6	98.7±0.1	**99.0±0.2**
2	57.2±4.7	45.9±3.3	72.8±9.6	45.1±0.8	33.1±1.1	63.1±1.2	48.6±2.8	74.0±1.8	**82.2±0.7**
3	59.5±9.0	56.1±3.2	69.7±11	61.2±1.1	37.5±1.4	85.3±2.5	39.3±1.8	84.5±2.2	**88.7±2.6**
4	68.3±9.0	73.2±1.7	79.1±10	20.6±0.6	30.5±0.9	88.0±3.6	71.5±0.4	81.3±1.4	**88.3±1.4**
5	48.7±3.4	41.2±2.0	79.5±3.6	66.0±0.4	41.8±1.1	66.4±3.8	32.3±0.9	70.0±1.6	**81.1±2.7**
6	65.1±6.1	65.3±3.8	90.0±4.0	42.3±0.7	36.5±0.6	94.0±1.9	59.4±1.4	95.7±0.8	**96.3±1.0**
7	62.7±5.0	69.4±2.3	66.8±6.6	37.8±1.6	48.3±2.0	90.8±1.8	61.5±1.6	91.8±1.3	**95.6±1.6**
8	53.1±10.6	68.7±3.8	78.3±9.5	46.4±0.5	28.7±1.2	83.7±3.4	51.0±1.4	79.1±1.3	**83.7±2.3**
9	62.7±4.4	76.4±2.3	84.7±11	28.1±0.7	36.6±0.7	87.8±1.4	69.6±1.6	93.1±0.8	**94.9±0.5**
Average	58.8	68.9	79.4	42.6	32.0	85.1	55.3	86.3	**90.6**

With Lemma 1 and Lemma 2, we can derive:

Lemma 3. *For* $\forall\delta > 0$*, w.p. at least* $1-\delta$*, for* $\forall h \in \tilde{\mathcal{H}}$*:*

$$d_{\tilde{\mathcal{H}}}(\mathcal{D}_S, \mathcal{D}_T) \le d_{\tilde{\mathcal{H}}}(\widehat{\mathcal{D}}_S, \widehat{\mathcal{D}}_T) + 2\,\mathrm{Rad}_{\mathbf{S}}(\tilde{\mathcal{H}}) + 2\,\mathrm{Rad}_{\mathbf{T}}(\tilde{\mathcal{H}}) + 3\sqrt{\log(4/\delta)/2n} + 3\sqrt{\log(4/\delta)/2n_t}$$

Proof. The triangular inequality of $d_{\tilde{\mathcal{H}}}(\cdot,\cdot)$ is written as:

$$d_{\tilde{\mathcal{H}}}(\mathcal{D}, \mathcal{D}') \le d_{\tilde{\mathcal{H}}}(\mathcal{D}, \widehat{\mathcal{D}}) + d_{\tilde{\mathcal{H}}}(\widehat{\mathcal{D}}, \widehat{\mathcal{D}}') + d_{\tilde{\mathcal{H}}}(\widehat{\mathcal{D}}', \mathcal{D}').$$

By Lemma 2, it follows that with probability $\ge 1-\delta/2$, the following two inequalities hold:

$$d_{\tilde{\mathcal{H}}}(\mathcal{D}_S, \widehat{\mathcal{D}}_S) \le 2\,\mathrm{Rad}_{\mathbf{S}}(\tilde{\mathcal{H}}) + 3\sqrt{\log(4/\delta)/2n}$$
$$d_{\tilde{\mathcal{H}}}(\mathcal{D}_T, \widehat{\mathcal{D}}_T) \le 2\,\mathrm{Rad}_{\mathbf{T}}(\tilde{\mathcal{H}}) + 3\sqrt{\log(4/\delta)/2n_t}$$

These two inequalities can be combined as with a union bound to obtain the inequality in the lemma. □

Lemma 4. *Let* $\langle\mathcal{D}_S, f_S\rangle$ *and* $\langle\mathcal{D}_T, f_T\rangle$ *be the source and target domains respectively. For any function class* $\mathcal{H} \subseteq [0,1]^{\mathcal{X}}$*, and* $\forall h \in \mathcal{H}$*, the following inequality holds:*

$$\varepsilon_T(h) \le \varepsilon_S(h) + d_{\tilde{\mathcal{H}}}(\mathcal{D}_S, \mathcal{D}_T) + \min\left\{\mathbb{E}_{\mathcal{D}_S}\left[|f_S - f_T|\right], \mathbb{E}_{\mathcal{D}_T}\left[|f_S - f_T|\right]\right\}$$

Finally, the proof of generalization upper bound Thm. 2 is given as:

Proof. Following Lemma 4, we have:

$$\varepsilon_T(h) \le \varepsilon_S(h) + d_{\tilde{\mathcal{H}}}(\mathcal{D}_S, \mathcal{D}_T) + \min\left\{\mathbb{E}_{\mathcal{D}_S}\left[|f_S - f_T|\right], \mathbb{E}_{\mathcal{D}_T}\left[|f_S - f_T|\right]\right\}$$

The probabilistic bounds for $\varepsilon_S(h)$ are given in Lemma 1 and Lemma 3. Applying them to the inequality above finishes the proof. The term $O(\sqrt{\log(4/\delta)/2n})$ goes away since $n_T \ll n$. □

D. Training and Implementation Details

The network configurations for MVTec dataset and Office-Home dataset is stated in section 4.1 and section 4.3. For experiments on digits dataset, the generator network has 5 transpose convolution layers; the discriminator is a 4-layer CNN followed by an FC layer; the encoder has 5 CNN layers. The training time is ∼10 minutes for one digit experiment and 1 hour for one experiment on Office-Home/MvTec dataset on an RTX 2080. The optimizer is Adam with lr = 10^{-4} and no weight decay. About training/validation sets: the training set contains *normal data*

Table 3. Results on MNIST→SVHN with highest numbers in bold. The metric is AUROC in percent. Results are also plotted in Fig. 4.

Model	DSVDD	AGT	BiOST	OCSVM (S+T)	OCSVM (T)	**IRAD (OC)**	IF (S+T)	IF (T)	**IRAD (IF)**
0	51.3±1.3	53.2±0.5	56.1±1.7	50.4±0.1	47.1±0.2	56.5±3.2	49.0±0.5	54.4±0.4	**56.7±2.0**
1	51.7±1.2	50.7±0.5	56.7±1.8	49.5±0.1	51.0±0.2	52.5±4.2	50.4±0.4	51.8±0.6	**61.0±2.0**
2	51.0±0.8	50.6±0.3	53.8±1.0	49.0±0.1	49.6±0.1	52.7±1.4	50.9±0.4	51.2±0.2	**56.0±0.2**
3	51.7±0.3	50.0±0.4	53.7±1.1	48.8±0.1	49.7±0.3	54.1±0.6	51.0±0.1	50.6±0.3	**55.8±0.9**
4	50.4±0.7	50.9±0.5	54.7±1.5	49.7±0.1	51.3±0.3	51.7±2.8	49.7±0.5	51.7±0.5	**55.9±1.1**
5	50.5±0.4	51.0±0.2	53.9±1.1	49.8±0.1	49.5±0.2	51.1±0.7	42.9±0.2	51.3±0.4	**54.1±0.8**
6	49.2±0.5	51.3±0.4	55.9±1.4	50.8±0.1	49.8±0.2	53.3±1.0	48.8±0.3	52.4±0.5	**56.6±1.4**
7	50.4±1.1	49.5±0.6	56.3±2.0	50.1±0.2	51.3±0.2	49.4±2.1	50.1±0.5	51.9±0.4	**57.0±1.3**
8	50.5±2.3	50.7±0.3	53.0±0.8	50.5±0.2	50.0±0.3	52.7±2.1	49.1±0.3	51.3±0.2	**54.2±0.9**
9	49.8±0.4	51.7±0.5	54.4±1.1	50.7±0.2	48.8±0.3	52.6±0.9	49.2±0.3	52.3±0.4	**55.9±1.4**
Average	50.6	50.9	54.8	49.9	49.8	52.6	49.1	51.9	**56.3**

Table 4. Experimental results of Product→Clip Art. The evaluation metric is AUROC in percent. Results are also plotted in Fig. 6.

Model	DSVDD	AGT	BiOST	OCSVM (S+T)	OCSVM (T)	**IRAD (OC)**	IF (S+T)	IF (T)	**IRAD (IF)**
Bike	51.1±2.7	55.7±2.6	52.7±0.8	50.0±0.1	50.0±0.1	54.1±5.1	45.5±1.8	57.7±3.9	**85.7±2.8**
Calculator	53.4±8.5	79.7±4.2	65.2±1.0	49.4±0.7	50.0±0.0	52.1±2.3	46.4±1.5	**81.5±3.9**	79.2±1.8
Drill	53.5±4.8	54.5±3.3	47.0±0.5	47.1±1.5	48.2±1.2	51.8±3.5	58.8±2.8	63.6±6.0	**71.2±5.5**
Hammer	50.3±1.7	64.4±2.3	43.7±0.9	49.3±0.7	49.2±0.5	60.3±1.6	56.8±1.1	61.9±3.4	**77.0±6.0**
Kettle	44.3±6.5	56.3±3.2	47.7±1.5	48.7±0.7	47.7±1.7	66.7±0.9	57.0±2.1	57.7±3.3	**70.0±4.9**
Knives	64.3±4.3	68.9±3.9	63.1±1.5	48.7±0.9	49.5±0.6	50.7±2.3	36.1±2.7	67.8±5.0	**70.3±3.5**
Pan	49.2±5.8	56.4±3.9	49.3±1.5	49.9±0.5	50.0±0.0	57.3±3.5	59.8±1.3	60.0±5.3	**72.8±3.7**
Paper	51.4±1.9	60.8±3.8	45.1±2.6	49.0±0.8	48.7±0.7	54.7±1.9	58.4±3.1	61.1±5.6	**61.8±0.8**
Scissors	49.0±8.7	66.5±3.7	38.6±0.8	48.5±0.6	48.5±1.3	58.2±0.2	59.0±1.1	62.9±3.0	**70.0±3.3**
Soda	48.8±5.8	57.3±8.7	56.9±0.8	49.9±0.4	50.0±0.1	54.2±1.5	50.9±1.8	56.4±7.8	**63.2±4.9**
Average	51.5	62.0	50.9	49.0	49.1	56.0	52.8	63.0	**72.1**

Table 5. Results of Clip Art→Product with the best numbers in bold. The metric is AUROC in percent. Results are also plotted in Fig. 5.

Model	DSVDD	AGT	BiOST	OCSVM (S+T)	OCSVM (T)	**IRAD (OC)**	IF (S+T)	IF (T)	**IRAD (I)**
Bike	49.4±11.6	54.0±2.5	43.0±0.6	46.2±1.2	46.5±2.2	78.8±5.9	51.4±2.0	65.5±3.69	**90.3±2.6**
Calculator	48.6±6.7	56.5±5.2	69.0±0.6	50.0±0.1	50.0±0.1	58.9±4.8	46.3±3.0	57.6±6.3	**82.2±1.8**
Drill	52.8±9.5	33.9±2.1	66.4±0.7	50.0±0.1	50.0±0.1	54.1±3.0	34.4±1.3	64.4±5.1	**73.0±5.4**
Hammer	44.7±9.0	79.4±1.2	50.1±0.7	47.8±0.6	48.7±0.5	53.5±4.1	81.9±1.5	80.0±1.1	**84.5±2.8**
Kettle	49.1±11.1	52.0±3.1	63.0±1.0	50.0±0.1	50.0±0.1	63.4±2.7	45.4±1.5	55.6±5.0	**75.8±8.5**
Knives	57.2±1.8	47.3±3.3	48.8±2.2	48.1±1.2	49.4±0.8	54.5±1.8	48.7±1.8	36.0±1.5	**63.9±2.4**
Pan	50.2±7.6	48.4±3.6	57.7±1.4	50.0±0.1	50.0±0.0	54.6±1.5	45.0±2.0	60.9±2.4	**76.0±4.5**
Paper	48.0±9.3	**74.9± 3.2**	27.4±4.0	50.0±0.1	50.0±0.1	63.7±3.2	68.0±2.0	70.6±4.0	67.4±3.4
Scissors	51.3±10.1	65.0±1.2	56.4±0.6	49.5±0.4	49.5±0.7	65.4±1.9	63.0±0.9	59.0±1.5	**68.9±4.0**
Soda	52.9±12.0	48.0±9.0	50.2±1.2	48.5±1.0	47.9±1.2	51.2±2.1	34.1±2.5	51.0±13	**53.3±1.8**
Average	50.4	55.9	53.2	49.0	49.2	59.8	51.8	60.0	**73.5**

only; the validation set contains normal and anomalous data from the *source domain only.*

E. Experiments with different numbers of target-domain training data

The full results of experiments with different numbers of target-domain training data are presented in Fig. 10.

F. Examples of images in Office-Home dataset for evaluation

Fig. 11 displays some examples of images in the Clip Art and Product domain from Office-Home dataset.

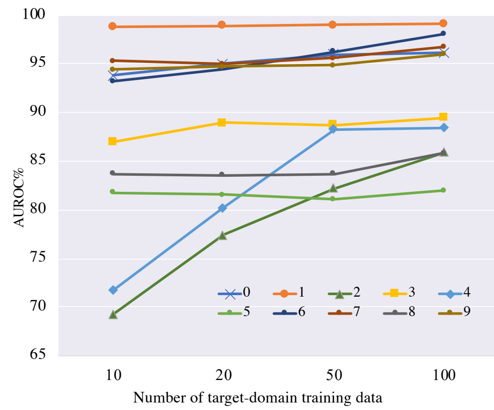

Figure 10. AUROC on MNIST-USPS experiments of all ten categories with the number of target-domain training data $n_t = 10, 20, 50, 100$. The model performance increases as more target-domain data are available for training.

Figure 11. Examples of ten categories in Clip Art and Product domain from Office-Home dataset.