Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] see Section 6
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] see our Supplementary File
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] see Section 5
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 5
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We include our source code in the Supplementary File
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Proofs

For the following proofs, we treat the variables as continuous variables and always use the integral. If one or some of the variables are discrete, it is straight-forward to replace the corresponding integral(s) with summation sign(s) and the proofs still hold.

A.1 Remark 1

Proof.

i) If there exists a representation z defined by the mapping $p(z|x)$ that aligns both the marginal and conditional distribution, then $\forall d, d', y$ we have:

$$\begin{aligned} p(y, z|d) &= p(z|d)p(y|z, d) \\ &= p(z|d')p(y|z, d') = p(y, z|d'). \end{aligned} \tag{8}$$

By marginalizing both sides of Eq 8 over z, we get $p(y|d) = p(y|d')$.

ii) If $p(y|d)$ is unchanged w.r.t. the domain d, then we can always find a domain invariant representation, for example, $p(z|x) = \delta_0(z)$ for the deterministic case (that maps all x to 0), or $p(z|x) = \mathcal{N}(z; 0, 1)$ for the probabilistic case.

These representations are trivial and not of our interest since they are uninformative of the input x. However, the readers can verify that they do align both the marginal and conditional distribution of data.

□

A.2 Remark 2

Proof.

- If $I(z, d) = 0$, then $p(z|d) = p(z)$, which means $p(z|d)$ is invariant w.r.t. d.
- If $p(z|d)$ is invariant w.r.t. d, then $\forall z, d$:

$$\begin{aligned} p(z) &= \int p(z|d')p(d')\mathrm{d}d' = \int p(z|d)p(d')\mathrm{d}d' \\ &\quad (\text{since } p(z|d') = p(z|d)\forall d') \\ &= p(z|d)\int p(d')\mathrm{d}d' = p(z|d) \\ &\implies I(z, d) = 0 \end{aligned} \tag{9}$$

□

A.3 Theorem 1

Proof.

i) Marginal alignment: $\forall z$ we have:

$$\begin{aligned} p(z|d) &= \int p(x|d)p(z|x)\mathrm{d}x \\ &= \int p(f_{d',d}(x')|d)p(z|f_{d',d}(x'))\left|\det J_{f_{d',d}}(x')\right|\mathrm{d}x' \end{aligned}$$

(by applying variable substitution in multiple integral: $x' = f_{d,d'}(x)$)

$$\begin{aligned} &= \int p(x'|d')\left|\det J_{f_{d',d}}(x')\right|^{-1} p(z|x') \\ &\qquad\qquad \left|\det J_{f_{d',d}}(x')\right|\mathrm{d}x' \end{aligned}$$

(since $p(f_{d',d}(x')|d) = p(x'|d')\left|\det J_{f_{d',d}}(x')\right|^{-1}$ and $p(z|f_{d',d}(x')) = p(z|x')$)

$$\begin{aligned} &= \int p(x'|d')p(z|x')\mathrm{d}x' \\ &= p(z|d') \end{aligned} \tag{10}$$

ii) Conditional alignment: $\forall z, y$ we have:

$$\begin{aligned} p(z|y, d) &= \int p(x|y, d)p(z|x)\mathrm{d}x \\ &= \int p(f_{d',d}(x')|y, d)p(z|f_{d',d}(x'))\left|\det J_{f_{d',d}}(x')\right|\mathrm{d}x' \end{aligned}$$

(by applying variable substitution in multiple integral: $x' = f_{d,d'}(x)$)

$$= \int p(x'|y,d') \left|\det J_{f_{d',d}}(x')\right|^{-1} p(z|x') \left|\det J_{f_{d',d}}(x')\right| \mathrm{d}x'$$

(since $p(f_{d',d}(x')|y,d) = p(x'|y,d') \left|\det J_{f_{d',d}}(x')\right|^{-1}$ and $p(z|f_{d',d}(x')) = p(z|x')$)

$$= \int p(x'|y,d')p(z|x')\mathrm{d}x'$$
$$= p(z|y,d') \tag{11}$$

Note that

$$p(y|z,d) = \frac{p(y,z|d)}{p(z|d)} = \frac{p(y|d)p(z|y,d)}{p(z|d)} \tag{12}$$

Since $p(y|d) = p(y) = p(y|d'), p(z|y,d) = p(z|y,d')$ and $p(z|d) = p(z|d')$, we have:

$$p(y|z,d) = \frac{p(y|d')p(z|y,d')}{p(z|d')} = p(y|z,d') \tag{13}$$

□

B Discussion on the choice of the distance metric between representations

As discussed in Section 3.2, we enforce the representation network g_θ to be invariant under the domain transformation $f_{d,d'}$ (with any two domains d, d'), meaning that $g_\theta(x) = g_\theta(f_{d,d'}(x))$.

In our implementation, we use the squared error distance as the distance between $g_\theta(x)$ and $g_\theta(f_{d,d'}(x))$. Admittedly, this distance tends to have the side effect of making the norm of the representation smaller. However, as visualized in Section 5.3, it does successfully align the distributions of the representation.

We also considered other distances such as contrastive distance and the cosine distance. We present below in Table 4 an ablation study with difference choices of the distance metrics, for the Rotated Mnist experiment with the target domain $\mathcal{M}_{75}$. Note that in this Rotated Mnist dataset, domains $\mathcal{M}_{75}$ and $\mathcal{M}_0$ are (equally) the hardest target domains. Therefore, we choose $\mathcal{M}_{75}$ for this ablation study to compare the performance of the variants.

Table 4: **Ablation study**: Rotated MNIST experiments with $\mathcal{M}_{75}$ as the target domain.

Distance Metric	Accuracy
Squared Error Distance	97.1±0.3
Contrastive Distance	95.8±0.9
Cosine Distance	90.1±0.3

As the Squared Error Distance gives the best performance and also is the most stable one in practice, we decide to use it for our implementation.