# A. Proofs

*Derivation of* (1).

$$\begin{aligned}
\mathcal{E}_U(\phi, h) &= \int \mathrm{d}x p_U(x) \left|h(\phi(x)) - f(x)\right| \\
&= \int \mathrm{d}x \int \mathrm{d}z p_U^\phi(z) \phi_U(x|z) \left|h(\phi(x)) - f(x)\right| \\
&= \int \mathrm{d}z p_U^\phi(z) \int \mathrm{d}x \phi_U(x|z) \left|h(z) - f(x)\right| \\
&= \int \mathrm{d}z p_U^\phi(z) \left|h(z) - \int \mathrm{d}x \phi_U(x|z) f(x)\right| \\
&\doteq \int \mathrm{d}z p_U^\phi(z) \left|h(z) - f_U^\phi(z)\right| \\
&\doteq \int \mathrm{d}z p_U^\phi(z) r_U(z; \phi, h)
\end{aligned}$$

where we use the following fact: For any fixed $z$, $h(z) \in \{0, 1\}$, if $h(z) = 0$ then $|h(z) - f(x)| = f(x) - h(z)$ for all $x$. Similarly, when $h(z) = 1$, we have $|h(z) - f(x)| = h(z) - f(x)$ for all $x$. Thus we can move the integral over $x$ inside the absolute operation. □

*Proof of Proposition 3.1.* First we have

$$\rho_U = \int \mathrm{d}x p_U(x) f(x) = \int \mathrm{d}x \int \mathrm{d}z p_U^\phi(z) \phi_U(x|z) f(x) = \int \mathrm{d}z p_U^\phi(z) f_U^\phi(z) \,.$$

When $\mathcal{E}_S(\phi, h) = 0$ we have

$$\left|\int \mathrm{d}z p_S^\phi(z) h(z) - \rho_S\right| = \left|\int \mathrm{d}z p_S^\phi(z) h(z) - \int \mathrm{d}z p_S^\phi(z) f_S^\phi(z)\right| \leq \int \mathrm{d}z p_S^\phi(z) \left|h(z) - f_S^\phi(z)\right| = \mathcal{E}_S(\phi, h) = 0$$

thus $\int \mathrm{d}z p_S^\phi(z) h(z) = \rho_S$.

Applying the fact that $p_S^\phi(z) = p_T^\phi(z)$ for all $z \in \mathcal{Z}$,

$$\begin{aligned}
\mathcal{E}_T(\phi, h) &= \int \mathrm{d}z p_T^\phi(z) \left|h(z) - f_T^\phi(z)\right| \geq \left|\int \mathrm{d}z p_T^\phi(z) h(z) - \int \mathrm{d}z p_T^\phi(z) f_T^\phi(z)\right| \\
&= \left|\int \mathrm{d}z p_S^\phi(z) h(z) - \int \mathrm{d}z p_T^\phi(z) f_T^\phi(z)\right| = \left|\rho_S - \rho_T\right| \,,
\end{aligned}$$

which concludes the proof. □

*Proof of Proposition 3.2.* Let $p_S$ be the uniform distribution over $[0, 1]$ and $p_T$ be the uniform distribution over $[2, 3]$. The labeling function $f$ is set as $f(x) = 1$ iff $x \in [0, \rho_S] \cup [2, 2 + \rho_T]$ such that the definition of $\rho_S$ and $\rho_T$ is preserved. We construct the following mapping $\phi$: For $x \in [0, 1]$ $\phi(x) = x$. For $x \in [2, 2+\rho_T]$ $\phi(x) = (x-2)\rho_S/\rho_T$. For $x \in [2+\rho_T, 3]$ $\phi(x) = 1 - (3 - x)(1 - \rho_S)/(1 - \rho_T)$. $\phi$ maps both source and target data into $[0, 1]$ with $p_S^\phi$ to be uniform over $[0, 1]$ and $p_T^\phi(z) = \rho_T/\rho_S$ when $z \in [0, \rho_S]$ and $p_T^\phi(z) = (1 - \rho_T)/(1 - \rho_S)$ when $z \in [\rho_S, 1]$. Since $p_S^\phi(z) = 1$ for all $z \in [0, 1]$ we can conclude that $\sup_{z\in\mathcal{Z}} p_T^\phi(z)/p_S^\phi(z) \leq \max\left\{\frac{\rho_T}{\rho_S}, \frac{1-\rho_T}{1-\rho_S}\right\}$. □

*Proof of Theorem 4.3.* Instead of working with Assumption 4.2 we first extend Construction 4.1 with the following addition

**Construction A.1.** (*Connectedness from target domain to source domain.*) Let $C_T \subset \mathcal{X}$ be a set of points in the raw data space that satisfy the following conditions:

1. $\phi(C_T) \subset \phi(C_0 \cup C_1)$.

2. For any $x \in C_T$, there exists $x' \in C_T \cap (C_0 \cup C_1)$ such that one can find a sequence of points $x_0, x_1, ..., x_m \in C_T$ with $x_0 = x$, $x_m = x'$, $f(x) = f(x')$ and $d_{\mathcal{X}}(x_{i-1}, x_i) < \frac{\Delta}{L}$ for all $i = 1, ..., m$.
3. $p_T(C_T) \geq 1 - \delta_3$.

We now proceed to prove bound based on Constructions 4.1 and A.1. Later on we will show that Assumption 4.2 indicates the existence of Construction A.1 so that the bound holds with a combination of Constructions 4.1 and Assumption 4.2.

The third term of (2) can be written as

$$\begin{aligned}
&\int \mathrm{d}z p_S^\phi(z) \left(\frac{p_T^\phi(z)}{p_S^\phi(z)} - 1\right) r_S(z;\phi,h) \\
&\leq \inf_{B\subseteq\mathcal{Z}} \int_B \mathrm{d}z p_S^\phi(z) \left(\frac{p_T^\phi(z)}{p_S^\phi(z)} - 1\right) r_S(z;\phi,h) + \int_{B^c} \mathrm{d}z p_S^\phi(z) \left(\frac{p_T^\phi(z)}{p_S^\phi(z)} - 1\right) r_S(z;\phi,h) \\
&\leq \inf_{B\subseteq\mathcal{Z}} \left(\sup_{z\in B} \frac{p_T^\phi(z)}{p_S^\phi(z)} - 1\right) \int_B \mathrm{d}z p_S^\phi(z) r_S(z;\phi,h) + \int_{B^c} \mathrm{d}z p_T^\phi(z) r_S(z;\phi,h) \\
&\leq \inf_{B\subseteq\mathcal{Z}} \left(\sup_{z\in B} \frac{p_T^\phi(z)}{p_S^\phi(z)} - 1\right) \mathcal{E}_S(\phi,h) + p_T^\phi(B^c) \\
&\leq \beta\mathcal{E}_S(\phi,h) + \delta_1 \,.
\end{aligned} \tag{12}$$

For the second term of (2), plugging in $r_U(z;\phi,h) = \left|h(z) - f_U^\phi(z)\right|$ gives

$$\begin{aligned}
&\int \mathrm{d}z p_T^\phi(z) \left(r_T(z;\phi,h) - r_S(z;\phi,h)\right) \\
&= \int \mathrm{d}z p_T^\phi(z) \left(\left|h(z) - f_T^\phi(z)\right| - \left|h(z) - f_S^\phi(z)\right|\right) \\
&= \int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \\
&= \int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \left(\mathbb{1}\left\{z \in \phi(C_0)\right\} + \mathbb{1}\left\{z \in \phi(C_1)\right\} + \mathbb{1}\left\{z \in \left(\phi(C_0) \cup \phi(C_1)\right)^c\right\}\right) \\
&= \int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \mathbb{1}\left\{z \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&+ \int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \mathbb{1}\left\{z \in \left(\phi(C_0) \cup \phi(C_1)\right)^c\right\}
\end{aligned} \tag{13}$$

Applying $\left|f_T^\phi(z) - f_S^\phi(z)\right| \leq f_T^\phi(z) + f_S^\phi(z)$ to the first part of (13) gives

$$\begin{aligned}
&\int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \mathbb{1}\left\{z \in \phi(C_0)\right\} \\
&\leq \int \mathrm{d}z p_T^\phi(z) f_T^\phi(z) \mathbb{1}\left\{z \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z) f_S^\phi(z) \mathbb{1}\left\{z \in \phi(C_0)\right\} \\
&= \int \mathrm{d}z p_T^\phi(z) \int \mathrm{d}x \phi_T(x|z) f(x) \mathbb{1}\left\{z \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z) f_S^\phi(z) \mathbb{1}\left\{z \in \phi(C_0)\right\} \\
&= \int \mathrm{d}x f(x) \int \mathrm{d}z p_T^\phi(z) \phi_T(x|z) \mathbb{1}\left\{z \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z) f_S^\phi(z) \mathbb{1}\left\{z \in \phi(C_0)\right\} \\
&= \int \mathrm{d}x f(x) p_T(x) \mathbb{1}\left\{\phi(x) \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z) f_S^\phi(z) \mathbb{1}\left\{z \in \phi(C_0)\right\} \\
&= \int \mathrm{d}x p_T(x) \mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z) f_S^\phi(z) \mathbb{1}\left\{z \in \phi(C_0)\right\}
\end{aligned} \tag{14}$$

Similarly, applying $\left|f_T^\phi(z) - f_S^\phi(z)\right| = \left|(1 - f_T^\phi(z)) - (1 - f_S^\phi(z))\right| \le (1 - f_T^\phi(z)) + (1 - f_S^\phi(z))$ to the second part of (13) gives

$$\begin{aligned}
&\int \mathrm{d}z p_T^\phi(z) \left|f_T^\phi(z) - f_S^\phi(z)\right| \mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&\le \int \mathrm{d}z p_T^\phi(z)(1 - f_T^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&= \int \mathrm{d}z p_T^\phi(z)\left(1 - \int \mathrm{d}x \phi_T(x|z) f(x)\right)\mathbb{1}\left\{z \in \phi(C_1)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&= \int \mathrm{d}x(1 - f(x)) \int \mathrm{d}z p_T^\phi(z)\phi_T(x|z)\mathbb{1}\left\{z \in \phi(C_1)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&= \int \mathrm{d}x(1 - f(x))p_T(x)\mathbb{1}\left\{\phi(x) \in \phi(C_1)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&= \int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 0, \phi(x) \in \phi(C_1)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \qquad (15)
\end{aligned}$$

Combining the second part of (14) and the second part of (15)

$$\begin{aligned}
&\int \mathrm{d}z p_T^\phi(z) f_S^\phi(z)\mathbb{1}\left\{z \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&= \int \mathrm{d}z \frac{p_T^\phi(z)}{p_S^\phi(z)} p_S^\phi(z) f_S^\phi(z)\mathbb{1}\left\{z \in \phi(C_0)\right\}\left(\mathbb{1}\left\{z \in B\right\} + \mathbb{1}\left\{z \in B^c\right\}\right) \\
&+ \int \mathrm{d}z \frac{p_T^\phi(z)}{p_S^\phi(z)} p_S^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\}\left(\mathbb{1}\left\{z \in B\right\} + \mathbb{1}\left\{z \in B^c\right\}\right) \\
&\le (1 + \beta)\int \mathrm{d}z p_S^\phi(z) f_S^\phi(z)\mathbb{1}\left\{z \in \phi(C_0)\right\} + (1 + \beta)\int \mathrm{d}z p_S^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&+ \int \mathrm{d}z p_T^\phi(z)\mathbb{1}\left\{z \in B^c\right\}\left(\mathbb{1}\left\{z \in \phi(C_0)\right\} + \mathbb{1}\left\{z \in \phi(C_1)\right\}\right) \\
&\le (1 + \beta)\int \mathrm{d}x p_S(x)\mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0)\right\} + (1 + \beta)\int \mathrm{d}x p_S(x)\mathbb{1}\left\{f(x) = 0, \phi(x) \in \phi(C_1)\right\} + p_T(B^c) \\
&\le (1 + \beta)\int \mathrm{d}x p_S(x)\left(\mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0) \vee f(x) = 0, \phi(x) \in \phi(C_1)\right\}\right) + \delta_1 \qquad (16)
\end{aligned}$$

For $i \in \{0, 1\}$ if $x \in C_i$ then $f(x) = i$ and $\phi(x) \in C_i$. So if $f(x) = 1, \phi(x) \in \phi(C_0)$ or $f(x) = 0, \phi(x) \in \phi(C_1)$ holds we must have $x \notin C_0 \cup C_1$. Therefore, following (16) gives

$$\begin{aligned}
&\int \mathrm{d}z p_T^\phi(z) f_S^\phi(z)\mathbb{1}\left\{z \in \phi(C_0)\right\} + \int \mathrm{d}z p_T^\phi(z)(1 - f_S^\phi(z))\mathbb{1}\left\{z \in \phi(C_1)\right\} \\
&\le (1 + \beta)\int \mathrm{d}x p_S(x)\mathbb{1}\left\{x \notin C_0 \cup C_1\right\} + \delta_1 \\
&= (1 + \beta)(1 - p_S(C_0 \cup C_1)) + \delta_1 \\
&\le (1 + \beta)\delta_2 + \delta_1 \qquad (17)
\end{aligned}$$

Now looking at the first part of (14) and the first part of (15)

$$\begin{aligned}
&\int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0)\right\} + \int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 0, \phi(x) \in \phi(C_1)\right\} \\
&= \int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0), x \in C_T\right\} + \int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0), x \notin C_T\right\} \\
&+ \int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 0, \phi(x) \in \phi(C_1), x \in C_T\right\} + \int \mathrm{d}x p_T(x)\mathbb{1}\left\{f(x) = 0, \phi(x) \in \phi(C_1), x \notin C_T\right\} \\
&\le \int \mathrm{d}x p_T(x)\left(\mathbb{1}\left\{f(x) = 1, \phi(x) \in \phi(C_0), x \in C_T\right\} + \mathbb{1}\left\{f(x) = 0, \phi(x) \in \phi(C_1), x \in C_T\right\}\right) + p_T(C_T^c)
\end{aligned}$$

$$\leq \int \mathrm{d}x p_T(x) \mathbb{1}\left\{x \in C_T\right\} \mathbb{1}\left\{f(x)=1, \phi(x) \in \phi(C_0) \vee f(x)=0, \phi(x) \in \phi(C_1)\right\} + \delta_3 \,. \tag{18}$$

Next we show that the first part of (18) is 0. Recall that $\phi(C_T) \subset \phi(C_0 \cup C_1)$ and if $x \in C_T$ there exists $x' \in C_T \cap (C_0 \cup C_1)$ with a sequence of points $x_0, x_1, ..., x_m \in C_T$ such that $x_0 = x$, $x_m = x'$, $f(x) = f(x')$ and $d_{\mathcal{X}}(x_{i-1}, x_i) < \frac{\Delta}{L}$ for all $i = 1, ..., m$. So for $x \in C_T$ and $f(x) = i$, we pick such $x'$. Since $\phi$ is $L$-Lipschitz and $\phi(C_T) \subset \phi(C_0 \cup C_1)$ we have $\phi(x_0), \phi(x_1), ..., \phi(x_m) \in \phi(C_0 \cup C_1)$ and $d_{\mathcal{Z}}(\phi(x_{i-1}), \phi(x_i)) < \Delta$ for all $i = 1, ..., m$. Applying the fact that $\inf_{z_0 \in \phi(C_0), z_1 \in \phi(C_1)} d_{\mathcal{Z}}(z_0, z_1) \geq \Delta > 0$ we know that if $\phi(x) = \phi(x_0) \in \phi(C_j)$ for some $j \in \{0, 1\}$ then $\phi(x') = \phi(x_m) \in \phi(C_j)$. From $x' \in C_0 \cup C_1$ and $f(x') = f(x) = i$ we have $\phi(x') \in \phi(C_i)$. Since $C_0 \cap C_1 = \emptyset$ we can conclude $i = j$ and thus $\phi(x) \in \phi(C_i)$ if $f(x) = i$ for any $x \in C_T$. Therefore, if $x \in C_T$, neither $f(x) = 1, \phi(x) \in \phi(C_0)$ nor $f(x) = 0, \phi(x) \in \phi(C_1)$ can hold. Hence the first part of (18) is 0.

So far by combining (17) and (18) we have shown that the sum of (14) and (15) (which are the first two parts of (13)) can be upper bounded by $\delta_1 + (1+\beta)\delta_2 + \delta_3$. For the third part of (13) we have

$$\begin{aligned}
&\int \mathrm{d}z p_T^{\phi}(z) \left| f_T^{\phi}(z) - f_S^{\phi}(z) \right| \mathbb{1}\left\{z \in (\phi(C_0) \cup \phi(C_1))^c\right\} \\
&\leq \int \mathrm{d}z p_T^{\phi}(z) \mathbb{1}\left\{z \in (\phi(C_0) \cup \phi(C_1))^c\right\} \\
&= \int \mathrm{d}z \frac{p_T^{\phi}(z)}{p_S^{\phi}(z)} p_S^{\phi}(z) \mathbb{1}\left\{z \in (\phi(C_0) \cup \phi(C_1))^c\right\} \left(\mathbb{1}\left\{z \in B\right\} + \mathbb{1}\left\{z \in B^c\right\}\right) \\
&\leq \int \mathrm{d}z \frac{p_T^{\phi}(z)}{p_S^{\phi}(z)} p_S^{\phi}(z) \mathbb{1}\left\{z \in (\phi(C_0) \cup \phi(C_1))^c\right\} \mathbb{1}\left\{z \in B\right\} + \int \mathrm{d}z p_T^{\phi}(z) \mathbb{1}\left\{z \in B^c\right\} \\
&\leq (1+\beta) \int \mathrm{d}z p_S^{\phi}(z) \mathbb{1}\left\{z \in (\phi(C_0) \cup \phi(C_1))^c\right\} + \delta_1 \\
&= (1+\beta) \left(1 - \int \mathrm{d}z p_S^{\phi}(z) \mathbb{1}\left\{z \in \phi(C_0) \cup \phi(C_1)\right\}\right) + \delta_1 \\
&= (1+\beta) \left(1 - \int \mathrm{d}x p_S(x) \mathbb{1}\left\{x \in \phi^{-1}\left(\phi(C_0) \cup \phi(C_1)\right)\right\}\right) + \delta_1 \\
&= (1+\beta) \left(1 - p_S\left(\phi^{-1}\left(\phi(C_0) \cup \phi(C_1)\right)\right)\right) + \delta_1 \\
&\leq (1+\beta) \left(1 - p_S\left(C_0 \cup C_1\right)\right) + \delta_1 \\
&\leq (1+\beta)\delta_2 + \delta_1 \,.
\end{aligned} \tag{19}$$

Putting (19) into (13) gives

$$\int \mathrm{d}z p_T^{\phi}(z) \left(r_T(z; \phi, h) - r_S(z; \phi, h)\right) \leq 2\delta_1 + 2(1+\beta)\delta_2 + \delta_3 \,. \tag{20}$$

Plugging (12) and (20) into (2) gives the result of Theorem 4.3 under Constructions 4.1 and A.1.

It remains to show that Assumption 4.2 implies the existence of a Construction A.1. To prove this, we first write $\phi(C_T) \subset \phi(C_0 \cup C_1)$ as $C_T \subset \phi^{-1}(\phi(C_0 \cup C_1))$. By Construction 4.1 we have $p_S(C_0 \cup C_1) \geq 1 - \delta_2$. From (19) we have

$$\begin{aligned}
p_T\left(\phi^{-1}(\phi(C_0 \cup C_1))\right) &= \int \mathrm{d}x p_T(x) \mathbb{1}\left\{x \in \phi^{-1}(\phi(C_0 \cup C_1))\right\} \\
&= \int \mathrm{d}z p_T^{\phi}(z) \mathbb{1}\left\{z \in \phi(C_0 \cup C_1)\right\} \geq (1+\beta)\delta_2 + \delta_1 \,.
\end{aligned}$$

Setting $B_S = C_0 \cup C_1$ and $B_T = \phi^{-1}(\phi(C_0 \cup C_1)$ in Assumption 4.2 gives a construction of Construction A.1, thus concluding the proof.

□

*Proof of Corollary 4.5.* Based on the statement of Corollary 4.5 it is obvious that Construction 4.1 can be made with $\delta_1 = 0$, $\delta_2 = 0$ and a finitely large $L$. (Here we implicitly assume that $\phi$ is bounded on $\mathcal{X}$). It remains to show that Assumption 4.2

holds with $\delta_3 = 0$. As $\delta_1 = \delta_2 = 0$, any $B_S$ and $B_T$ will be supersets of $\mathrm{Supp}(p_S)$ and $\mathrm{Supp}(p_T)$ respectively. So it sufficies to consider $B_S = \mathrm{Supp}(p_S)$ and $B_T = \mathrm{Supp}(p_T)$.

Now we verify that $C_T = \mathrm{Supp}(p_T)$ satisfies the requirements in Assumption 4.2. According to Assumption 4.4, for any $x \in \mathrm{Supp}(p_T)$, there must exist $S_{T,i,j}$ such that $x \in S_{T,i,j}$, $S_{T,i,j}$ is connected, $f(x') = i$ for all $x' \in S_{T,i,j}$ and $S_{T,i,j} \cap \mathrm{Supp}(p_S) \neq \emptyset$. Pick $x' \in S_{T,i,j} \cap \mathrm{Supp}(p_S)$. Such $x'$ satisfies $x' \in C_T \cap B_S$ with our choice of $C_T$ and $B_S$. Since $S_{T,i,j}$ is connected we can find a sequence of points $x_0, ..., x_m \in S_{T,i,j}$ with $x_0 = 0$, $x_m = x'$ and $d_{\mathcal{X}}(x_{i-1}, x_i) < \epsilon$ for any $\epsilon > 0$. As $S_{T,i,j}$ is label consistent we have $f(x) = f(x')$. Picking $\epsilon = \frac{\Delta}{L}$ concludes the fact that $C_T = \mathrm{Supp}(p_T)$ satisfies the requirements in Assumption 4.2.

Since $p_T(\mathrm{Supp}(p_T)) = 1$ we have $\delta_3 = 0$. As a result, $\mathcal{E}_T(\phi, h) \le (1+\beta)\mathcal{E}_S(\phi, h)$ holds according to Theorem 4.3, which concludes the proof of Corollary 4.5.

□

*Derivation of* (6). The Fenchel Dual of $\bar{f}_\beta(u)$ can be written as

$$\begin{aligned} \bar{f}^*_\beta(t) &= \begin{cases} t f'^{-1}(t) - \bar{f}_\beta(f'^{-1}(t)) & \text{if } t \le f'(\frac{1}{1+\beta})\,, \\ +\infty & \text{if } t > f'(\frac{1}{1+\beta})\,. \end{cases} \\ &= \begin{cases} t f'^{-1}(t) - f(f'^{-1}(t)) + C & \text{if } t \le f'(\frac{1}{1+\beta})\,, \\ +\infty & \text{if } t > f'(\frac{1}{1+\beta})\,. \end{cases} \\ &= \begin{cases} f^*(t) + C_{f,\beta} & \text{if } t \le f'(\frac{1}{1+\beta})\,, \\ +\infty & \text{if } t > f'(\frac{1}{1+\beta})\,. \end{cases}, \end{aligned}$$

where $C_{f,\beta} = f(\frac{1}{1+\beta}) - f'(\frac{1}{1+\beta})\frac{1}{1+\beta} + f'(\frac{1}{1+\beta})$.

Therefore, the modified $\bar{f}_\beta$-divergence can be written as

$$D_{f,\beta}(p,q) = \sup_{T:\mathcal{Z}\mapsto \mathrm{dom}(f^*)\cap(-\infty, f'(\frac{1}{1+\beta})]} \mathbb{E}_{z\sim q}\left[T(z)\right] - \mathbb{E}_{z\sim p}\left[f^*(T(z))\right] - C_{f,\beta}\,.$$

□

*Derivation of* (7). According to Nowozin et al. (2016), the GAN objecitve uses $f(u) = u\log u - (1+u)\log(1+u)$. Hence $f^*(t) = -\log(1-e^t)$, $f'(u) = \log\frac{u}{u+1}$ and $f'(\frac{1}{1+\beta}) = \log\frac{1}{2+\beta}$. So we need to parameterize $T : \mathcal{Z} \mapsto \big(-\infty, \log\frac{1}{2+\beta}\big]$. $T(z) = \log\frac{g(z)}{2+\beta}$ with $g(z) \in (0,1]$ satisfies the range constraint for $T$. Plugging $T(z) = \log\frac{g(z)}{2+\beta}$ into (6) gives the result of (7).

□

## B. Experiment Details

**Synthetic datasets** For source distribution, we sample class 0 from $\mathcal{N}([-1,-0.3], diag(0.1, 0.4))$ and class 1 from $\mathcal{N}([1, 0.3], diag(0.1, 0.4))$. For target distribution, we sample class 0 from $\mathcal{N}([-0.3,-1], diag(0.4, 0.1))$ and class 1 from $\mathcal{N}([0.3, 1], diag(0.4, 0.1))$. For label classifier, we use a fully-connect neural net with 3 hidden layers $(50, 50, 2)$ and the latent space is set as the last hidden layer. For domain classifier (critic) we use a fully-connect neural net with 2 hidden layers $(50, 50)$.

**Image datasets** For MNIST we subsample 2000 data points and for USPS we subsample 1800 data points. The subsampling process depends on the given label distribution (e.g. shift or no-shift). For label classifier, we use LeNet and the latent space is set as the last hidden layer. For domain classifier (critic) we use a fully-connect neural net with 2 hidden layers $(500, 500)$.

In all experiments, we use $\lambda = 1$ in the objective (4) and ADAM with learning rate 0.0001 and $\beta_1 = 0.5$ as the optimizer. We also apply a l2-regularization on the weights of $\phi$ and $h$ with coefficient $0.001$.

**More discussion on synthetic experiments.** The only unexcepted failure is WDANN1-2, which achieves only 20% accuracy in 2-out-of-5 runs. Looking in to the low accuracy runs we found that the l2-norm of the encoder weights is clearly higher than the successful runs. Large l2-norm of weights in $\phi$ likely results in a high Lipschitz constant $L$, which is undesirable according to our theory. We only implemented l2-regularization to encourage Lipschitz continuity of the encoder $\phi$, which might be insufficient. How to enforce Lipschitz continuity of a neural network is still an open question. Trying more sophisticated approaches for Lipschitz continuity can a future direction.

**Choice of $\beta$.** Since a good value of $\beta$ may depend on the knowledge of target label distribution which is unknown, we experiment with different values of $\beta$. Empirically we did not find any clear pattern of correlation between value of $\beta$ and performance as long as it is big enough to accommodate label distribution shift so we would leave it as an open question. In practice we suggest to use a moderate value such as $2$ or $4$, or estimate based on prior knowledge of target label distribution.