Supplementary Material for "Generative-Discriminative Complementary Learning"

September 6, 2019

S1. Proof of Theorem 1

According to the triangle inequality of total variation (TV) distance, we have

$$d_{TV}(P_{XY}, Q_{XY}) \leq d_{TV}(P_{XY}, P_{Y|X}Q_X) + d_{TV}(P_{Y|X}Q_X, Q_{XY}). \tag{1}$$

Using the definition of TV distance, we have

$$\begin{aligned} d_{TV}(P_{Y|X}P_X, P_{Y|X}Q_X) &= \frac{1}{2}\int |p_{Y|X}(y|x)p_X(x) - p_{Y|X}(y|x)q_X(x)|\mu(x,y) \\ &\overset{(a)}{\leq} \frac{1}{2}\int |p_{Y|X}(y|x)|\mu(x,y)\int |p_X(x) - q_X(x)|\mu(x) \\ &\leq c_1 d_{TV}(P_X, Q_X), \end{aligned} \tag{2}$$

where p and q are densities, μ is a (σ-finite) measure, c_1 is a constant, and (a) follows from the Hölder inequality.

Similarly, we have

$$d_{TV}(P_{Y|X}Q_X, Q_{Y|X}Q_X) \leq c_2 d_{TV}(P_{Y|X}, Q_{Y|X}), \tag{3}$$

where c_2 is a constant. Combining (1), (2), and (3), we have

$$\begin{aligned} d_{TV}(P_{XY}, Q_{XY}) &\leq c_1 d_{TV}(P_X, Q_X) + c_2 d_{TV}(P_{Y|X}, Q_{Y|X}) \\ &\leq c_1 d_{TV}(P_X, Q_X) + c_2 d_{TV}(P_{Y|X}, Q'_{Y|X}) + c_2 d_{TV}(Q'_{Y|X}, Q_{Y|X}). \end{aligned} \tag{4}$$

Since we have no access to $P_{Y|X}$, by simply adapting the proof of Theorem 1 in [Thekumparampil et al.2018], we bound $d_{TV}(P_{Y|X}, Q'_{Y|X})$ using complementary conditional probabilities as

$$d_{TV}(P_{Y|X}, Q'_{Y|X}) = \max_{S_1,\ldots,S_K \subseteq \mathcal{X}} \sum_{y\in\mathcal{Y}} \{P(y|S_y) - Q'(y|S_y)\}$$

$$
\begin{aligned}
&= \max_{S_1,\ldots,S_K \subseteq \mathcal{X}} \langle \mathbf{1}, P(\cdot|\{S_y\}_{y\in\mathcal{Y}}) - Q'(\cdot|\{S_y\}_{y\in\mathcal{Y}})\rangle \\
&\overset{(a)}{=} \max_{S_1,\ldots,S_K \subseteq \mathcal{X}} \langle \mathbf{1}, \boldsymbol{M}^{-1}(P(\cdot|\{S_{\bar{y}}\}_{\bar{y}\in\mathcal{Y}}) - Q'(\cdot|\{S_{\bar{y}}\}_{\bar{y}\in\mathcal{Y}}))\rangle \\
&\overset{(b)}{\leq} \|\boldsymbol{M}^{-\intercal}\|_1 \max_{S_1,\ldots,S_K \subseteq \mathcal{X}} \|P(\cdot|\{S_{\bar{y}}\}_{\bar{y}\in\mathcal{Y}}) - Q'(\cdot|\{S_{\bar{y}}\}_{\bar{y}\in\mathcal{Y}}))\|_1 \\
&= \|\boldsymbol{M}^{-1}\|_\infty d_{TV}(P_{\bar{Y}|X}, Q'_{\bar{Y}|X}), \qquad (5)
\end{aligned}
$$

where $P(\cdot|\{S_y\}) = [P(Y = 1|S_1), \cdots, P(Y = K|S_K)]^\intercal$, $P(\cdot|\{S_{\bar{y}}\}) = [P(\bar{Y} = 1|S_1), \cdots, P(\bar{Y} = K|S_K)]^\intercal$, (a) follows from $P(\cdot|\{S_{\bar{y}}\}) = \boldsymbol{M}P(\cdot|\{S_y\})$, and (b) follows from the fact that $\mathbf{1}^\intercal Ax \leq \|Ax\|_1 \leq \|A\|_1\|x\|_1$. By combining (4) and (5), we have

$$
\begin{aligned}
d_{TV}(P_{XY}, Q_{XY}) \leq\, & c_1 d_{TV}(P_X, Q_X) + c_2\|\boldsymbol{M}^{-1}\|_\infty d_{TV}(P_{\bar{Y}|X}, Q'_{\bar{Y}|X}) \\
& + c_2 d_{TV}(Q_{Y|X}, Q'_{Y|X}) \qquad (6)
\end{aligned}
$$

According to the relations between total variation (TV), KL divergence (d_{KL}), and Jensen-Shannon divergence (d_{JS}), we can rewrite (6) as

$$
\begin{aligned}
d_{TV}(P_{XY}, Q_{XY}) \leq\, & 2c_1\sqrt{d_{JS}(P_X, Q_X)} + c_2\|\boldsymbol{M}^{-1}\|_\infty \sqrt{d_{KL}(P_{\bar{Y}|X}, Q'_{\bar{Y}|X})} \\
& + \ c_2\sqrt{d_{KL}(Q_{Y|X}, Q'_{Y|X})}, \qquad (7)
\end{aligned}
$$

which follows from the Pinsker's inequality. By replacing $2c_1$ in (7) with a new constant c_1 (using the same notation for simplicity), we can obtain the inequality in Theorem 1. From the theorem, we can see that if the complementary labels are highly-biased, it may cause $\boldsymbol{M}$ to be rank-deficient. In this case, our algorithm may not minimize the distance between P_{XY} and Q_{XY} efficiently.

S2. Illustration of Our Objective Function (Eq. (5))

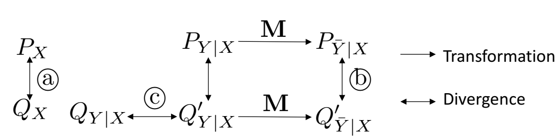

Figure 1: Illustration of the divergence terms that are minimized in Eq. (5). $P_{Y|X}$ ($P_{\bar{Y}|X}$) is the conditional distribution of ordinal (complementary) label given features on the real data. $Q'_{Y|X}$ ($Q'_{\bar{Y}|X}$) is the conditional distribution of ordinal (complementary) label produced by the classification network C in Eq. (5). $Q_{Y|X}$ is the conditional distribution of ordinal label given features induced by our generator G. From the figure, we can see that minimizing ⓑ leads to reduced divergence between $P_{Y|X}$ and $Q'_{Y|X}$. Therefore, the objective function minimizes the divergence between $P_{Y|X}$ and $Q_{Y|X}$ further because of ⓒ. Combined with ⓐ, our objective minimizes divergence between P_{XY} and Q_{XY}.

S3. Quality of synthetic data

Method \ r_l	0.2	0.4	0.6	0.8	1.0
CIFAR10					
Ordinary label, IS	5.16 ± 0.066	5.99 ± 0.058	6.19 ± 0.070	6.27 ± 0.070	6.53 ± 0.082
$CCGAN$, IS	5.28 ± 0.048	5.90 ± 0.065	6.27 ± 0.094	6.27 ± 0.067	6.48 ± 0.052
Ordinary label, FID	54.33	39.18	35.18	32.91	28.40
$CCGAN$, FID	50.75	37.47	33.86	34.55	31.63
CIFAR100					
Ordinary label, IS	5.11 ± 0.038	6.80 ± 0.084	7.59 ± 0.154	7.94 ± 0.133	7.82 ± 0.09
$CCGAN$, IS	4.80 ± 0.042	6.36 ± 0.059	6.73 ± 0.095	7.17 ± 0.085	7.22 ± 0.115
Ordinary label, FID	65.00	44.14	41.49	36.25	34.34
$CCGAN$, FID	79.13	44.01	43.63	36.21	34.63
VGGFACE100					
Ordinary label, IS	19.18 ± 0.254	29.19 ± 0.235	48.99 ± 0.533	54.59 ± 0.390	67.77 ± 0.568
$CCGAN$, IS	16.49 ± 0.243	28.10 ± 0.368	45.82 ± 0.746	52.97 ± 0.470	62.30 ± 0.409
Ordinary label, FID	100.48	66.00	42.98	38.07	26.26
$CCGAN$, FID	113.78	59.98	36.45	31.661	27.79

Table 1: This table shows the Inception Score and FID socore on CIFAR10, CIFAR100 and VGGFACE100 dataset. r_l denotes the proportion of sampled labeled data for training from the training set S. All these scores are under the uniformed M setting.

S4. More Generated Images

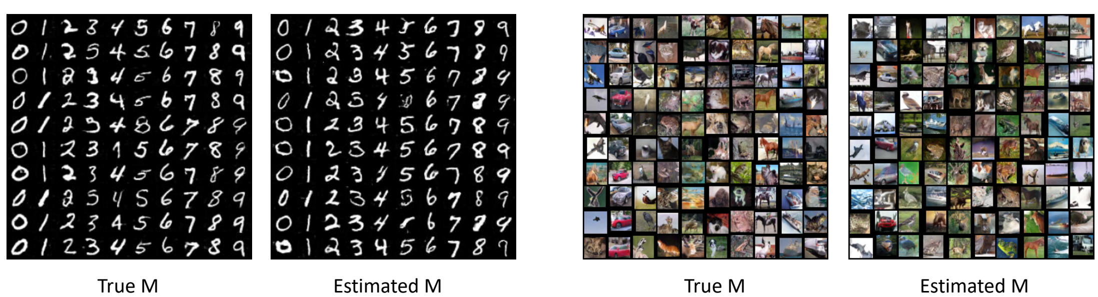

Figure 2: Synthetic results for MNIST and CIFAR10. We set $r_l = 1$ here. It shows the generated data with true M and esitimated M

References

[Thekumparampil et al.2018] Thekumparampil, K. K.; Khetan, A.; Lin, Z.; and Oh, S. 2018. Robustness of conditional gans to noisy labels. In *Advances in Neural Information Processing Systems*, 10271–10282. 1

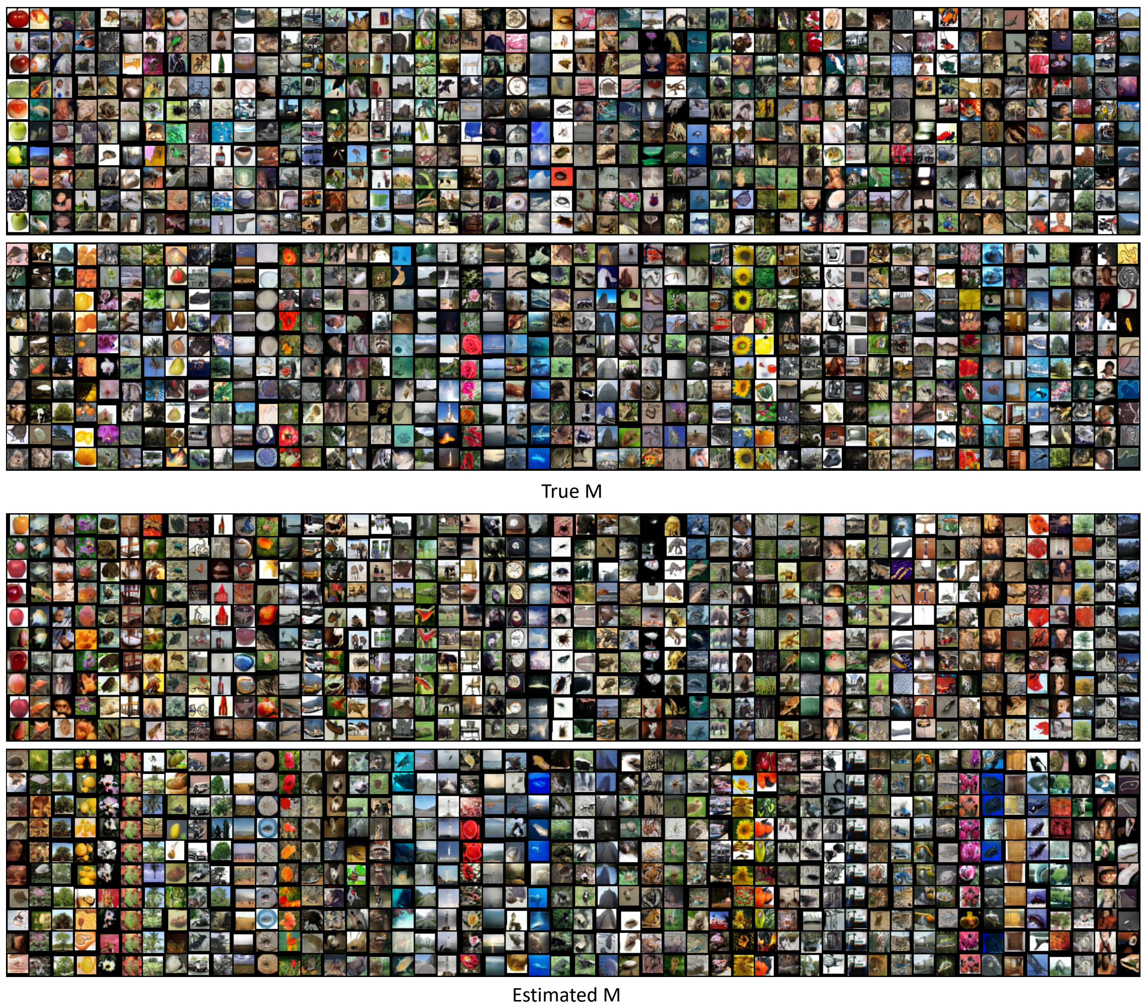

Figure 3: Synthetic results for CIFAR100. We set $r_l = 1$ here. It shows the generated data with true $\boldsymbol{M}$ and esitimated $\boldsymbol{M}$

Figure 4: Synthetic results for MNIST and VGGFACE100. We set $r_l = 1$ here. It shows the generated data with true $\boldsymbol{M}$ and esitimated $\boldsymbol{M}$