Unsupervised Visual Domain Adaptation: A Deep Max-Margin Gaussian Process Approach

Minyoung Kim[1,2,3], Behnam Gholami[2,*], Pritish Sahu[2,*], Vladimir Pavlovic[2,3]

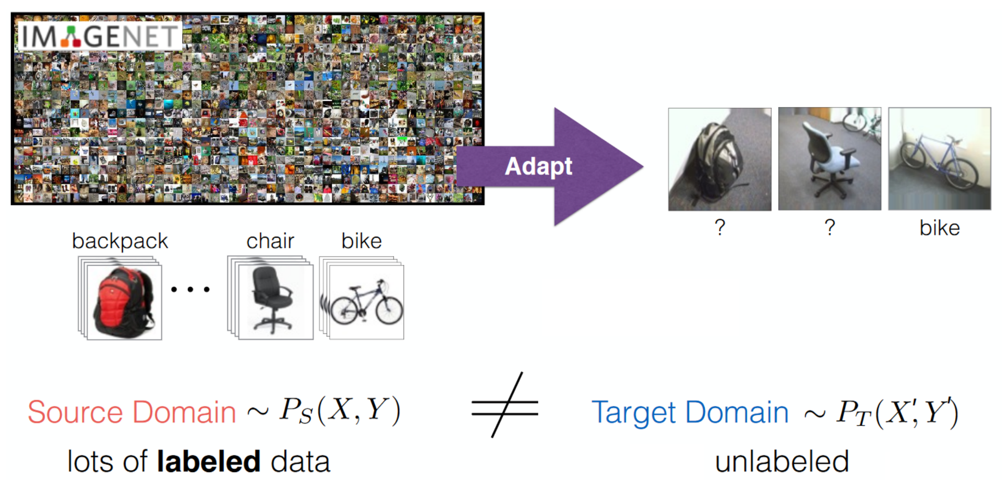
[1] SeoulTech, South Korea, [2] Rutgers University, NJ, USA, [3] Samsung AI Center, Cambridge, UK, (*Equal contribution)

RUTGERS SEOULTECH SAMSUNG AI Center - Cambridge

CVPR LONG BEACH, CA • June 16-20, 2019

Problem Definition and Contribution

Goal: Given a **Source** domain with labeled samples and a **Target** domain with unlabeled samples, improve a **Target** predictive function in the **Target** domain using knowledge from the **Source** domain.

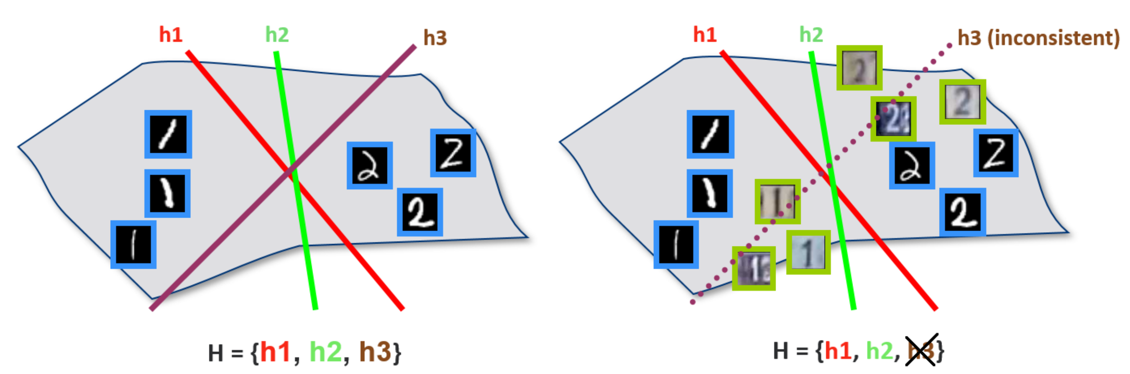


Key Contributions:

- Modeling the hypothesis space H of classifiers as Gaussian Process (GP) posterior of classifiers
- Imposing consensus in H by max-margin class separation principle
- Providing quantitative score of prediction uncertainty

Problem Formulation

Maximum Classifier Discrepancy (MCD)



$$e_T(h) \le e_S(h) + \sup_{h_1,h_2\in H} \mathbb{E}_{z\sim T}\big[I(h_1(z) \neq h_2(z))\big] + \text{const}$$

- $f_j(z)$:= model's confidence toward class j
- $f_j(z) \sim GP(0, K_j(.,.)),\ j = 1,\dots,K$
- Class Decision: class$(\mathbf{z}) = \arg\max_{1\le j\le K} f_j(\mathbf{z})$

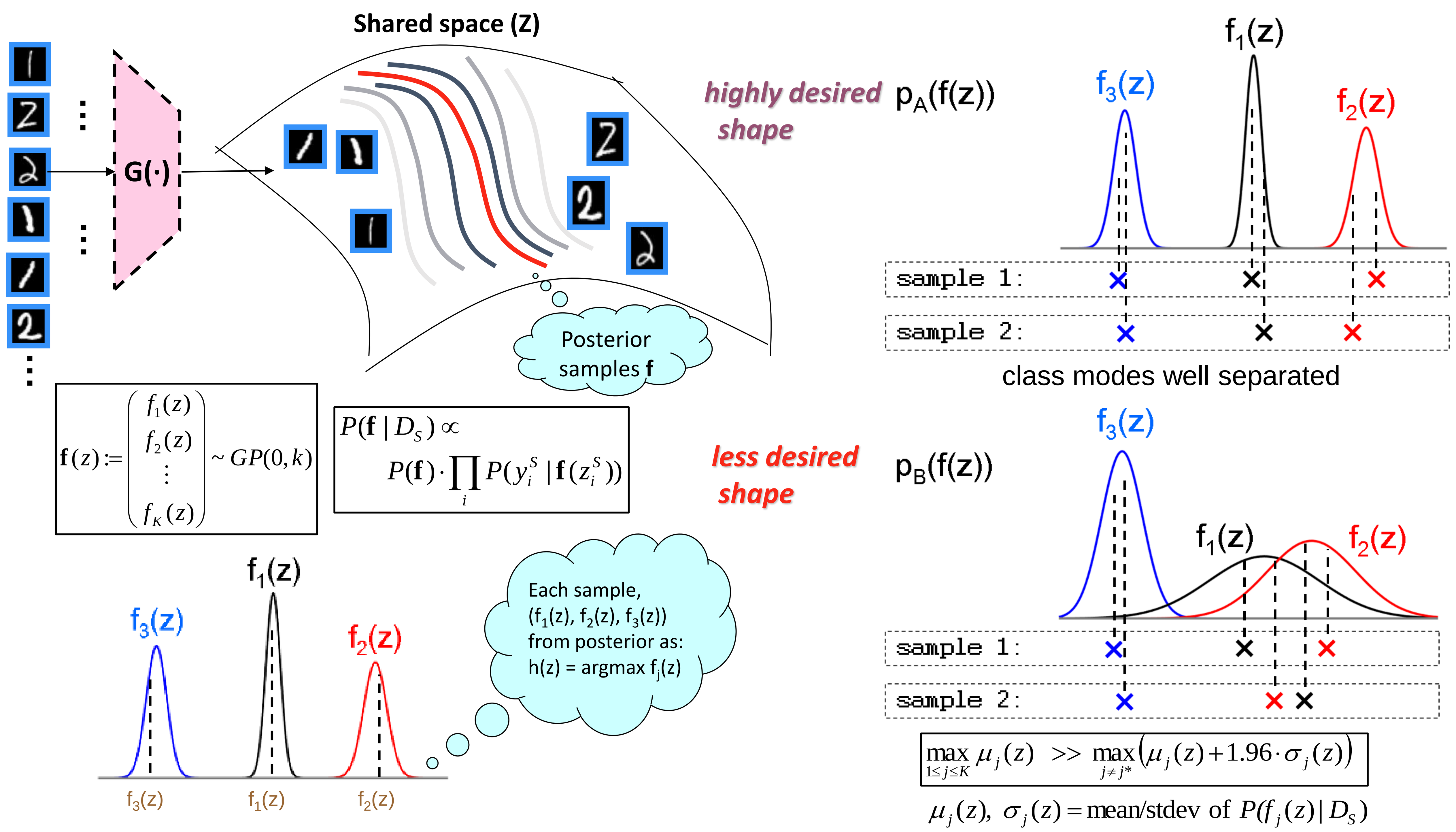
Optimization(alternate the two):

1. Fix $G(.)$ and compute GP posterior on **Source** samples to get μ, σ
2. Fix μ, σ and update G on **Target** as

$$\min_{G} \Big(\max_{j\neq j^*} \mu_j(\mathbf{z}) - \max_{1\le j\le K} \mu_j(\mathbf{z}) + 1 + \alpha \max_{1\le j\le K} \sigma_j(\mathbf{z}) \Big)_+$$

Method

- Defining the hypothesis space H of classifiers as the posterior distribution of Gaussian Process (GP) classification on **Source** samples
- Regularize the posterior to be maximally consistent over **Target** using the **Max-margin Principle**



Experiments & Results

Dataset

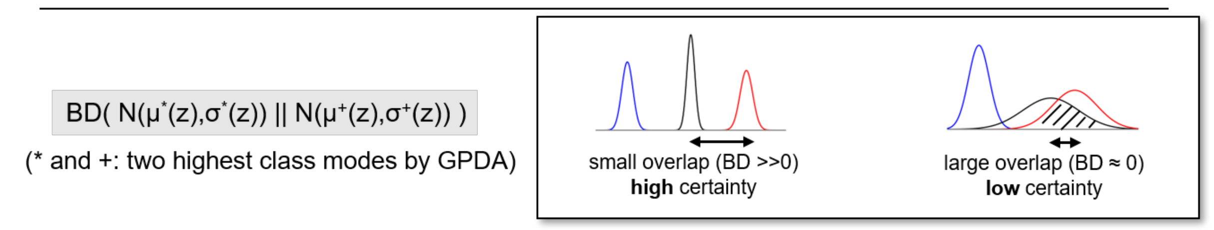
Statistics of the datasets

Dataset	Images	classes	Domains	Description
Digit	140,000	10	4	mnist, svhn, usps
Traffic Signs	150,000	43	2	Synthetic and German Traffic Signs
Visda	280,000	12	2	Synthetic and Real Images

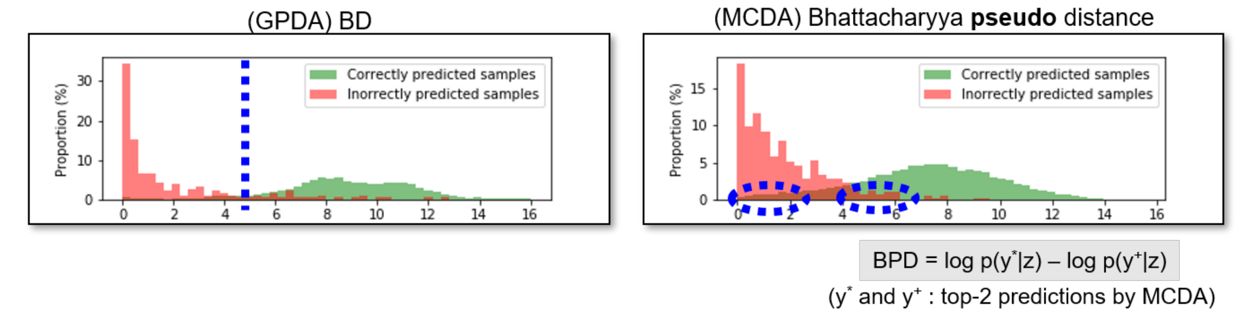
METHOD	SVHN to MNIST	SYNSIG to GTSRB	MNIST to USPS	MNIST* to USPS*	USPS to MNIST
Source Only	67.1	85.1	76.7	79.4	63.4
MMD † [34]	71.1	91.1	-	81.1	-
DANN † [16]	71.1	88.7	$77.1^{1.8}$	85.1	$73.0^{0.2}$
DSN † [9]	82.7	93.1	91.3	-	-
ADDA [60]	$76.0^{1.8}$	-	$89.4^{0.2}$	-	$90.1^{0.8}$
MCDA ($n=2$)	$94.2^{2.6}$	$93.5^{0.4}$	$92.1^{0.8}$	$93.1^{1.9}$	$90.0^{1.4}$
MCDA ($n=3$)	$95.9^{0.5}$	$94.0^{0.4}$	$93.8^{0.8}$	$95.6^{0.9}$	$91.8^{0.9}$
MCDA ($n=4$)	$96.2^{0.4}$	$94.4^{0.3}$	$94.2^{0.7}$	$96.5^{0.3}$	$94.1^{0.3}$
GPDA	$\mathbf{98.2}^{0.1}$	$\mathbf{96.19}^{0.2}$	$\mathbf{96.45}^{0.15}$	$\mathbf{98.11}^{0.1}$	$\mathbf{96.37}^{0.1}$

BD Distance

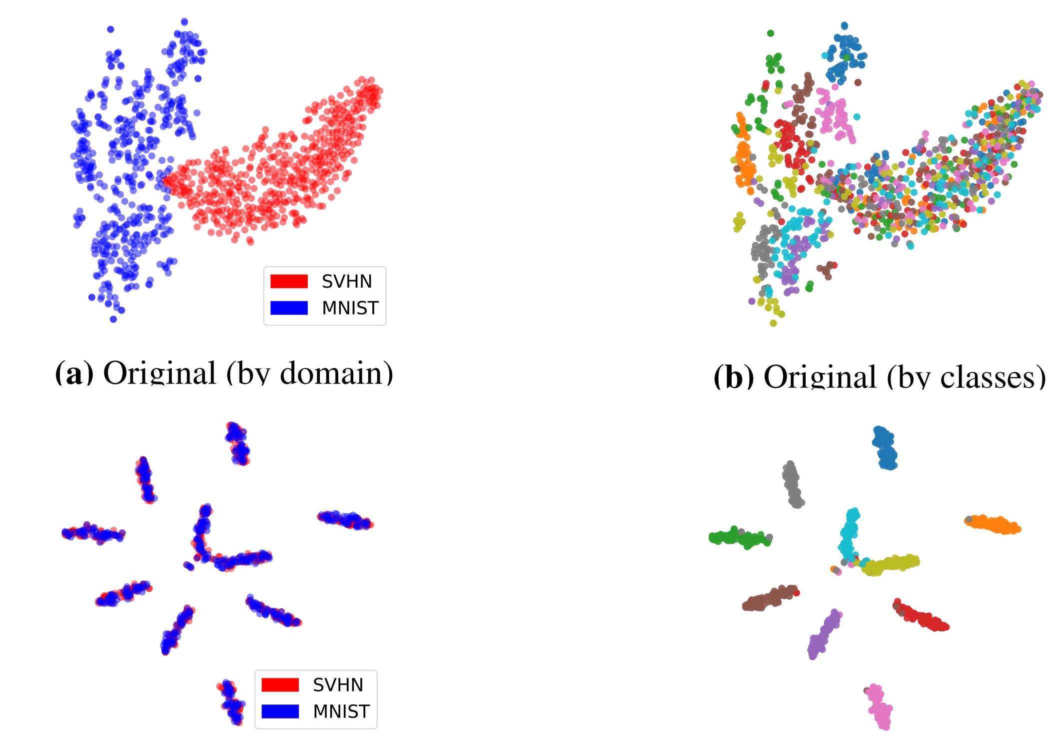
Prediction certainty = Bhattacharyya Dist. (eg, larger BD = less overlap = more certain prediction)



Histograms of BDs for (correctly predicted points vs. incorrect ones)



t-SNE Visualization



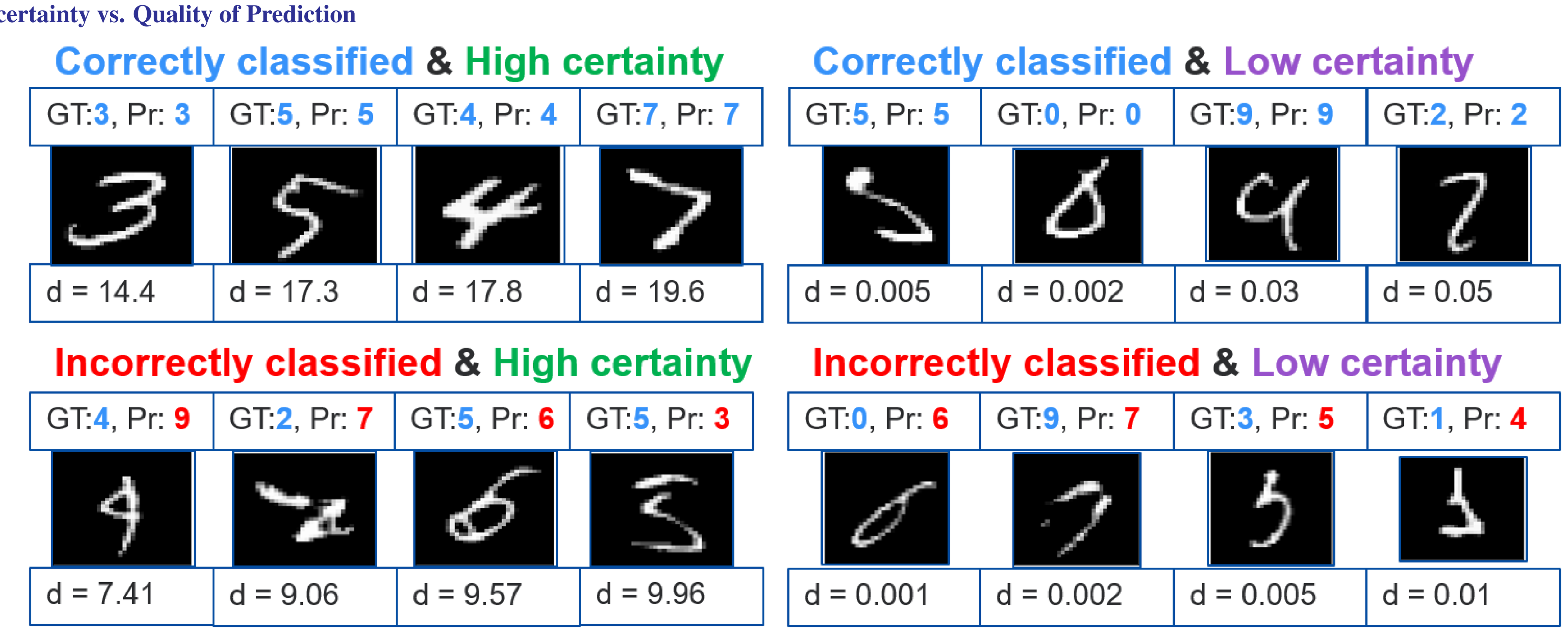
(a) Original (by domain) (b) Original (by classes)

(c) GPDA (by domain) (d) GPDA (by classes)

VisDA

Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	mean	Ave. ranking
Source Only	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4	7.25
MMD [34]	87.1	63.0	76.5	42.0	90.3	42.9	**85.9**	53.1	49.7	36.3	**85.8**	20.7	61.1	4.41
DANN [16]	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4	5.58
SimNet [44]	**94.3**	**82.3**	73.5	47.2	87.9	49.2	75.1	**79.7**	85.3	**68.5**	81.1	**50.3**	72.9	3.83
MCDA ($n=2$)	81.1	55.3	83.6	65.7	87.6	72.7	83.1	73.9	85.3	47.7	73.2	27.1	69.7	4.25
MCDA ($n=3$)	90.3	49.3	82.1	62.9	**91.8**	69.4	83.8	72.8	79.8	53.3	81.5	29.7	70.6	4.08
MCDA ($n=4$)	87.0	60.9	**83.7**	64.0	88.9	**79.6**	84.7	76.9	88.6	40.3	83.0	25.8	71.9	3.00
GPDA (ours)	83.0	74.3	80.4	**66.0**	87.6	75.3	83.8	73.1	**90.1**	57.3	80.2	37.9	**73.31**	**2.75**

Uncertainty vs. Quality of Prediction



GT = Ground-truth class label, Pr = Predicted class label, d = BD