Multi-batch Nuclear-norm Adversarial Network for Unsupervised Domain Adaptation

Pei Wang[1,2], Yun Yang[2,3]*, Zhenyu Yu[4,5]

[1]School of Information Science and Engineering, Yunnan University, China

2National Pilot School of Software, Yunnan University, China

3Yunnan Key Laboratory of Software Engineering, Yunnan University, China

4Guangzhou University, China 5Universiti Malaya, Malaysia

Contributions

- **Feature Queue**: cache features to generate a large and consistent outputs

- **Probability Rescaling**: avoid the negative effect of overconfident and noisy predictions

- **Multi-batch Nuclear-norm Discrepancy**: enhance the transferability and discriminability of the learned features

Code@: *https://github.com/peiwang0518/Multi-BAN*

Conceptual comparison

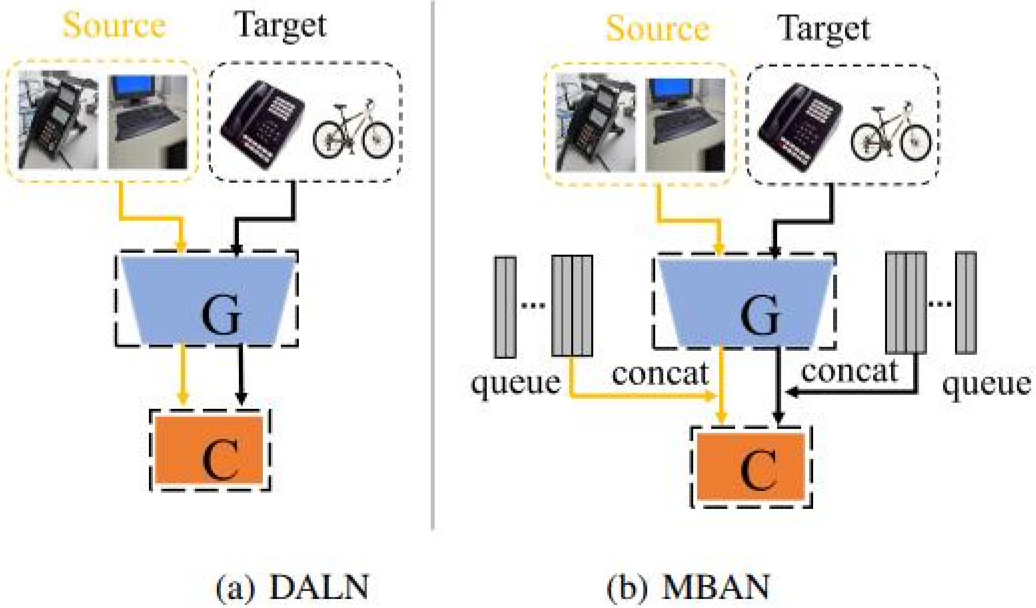


Fig. 1: Conceptual comparison of two adversarial learning paradigms.

Multi-batch Nuclear-norm Adversarial Network

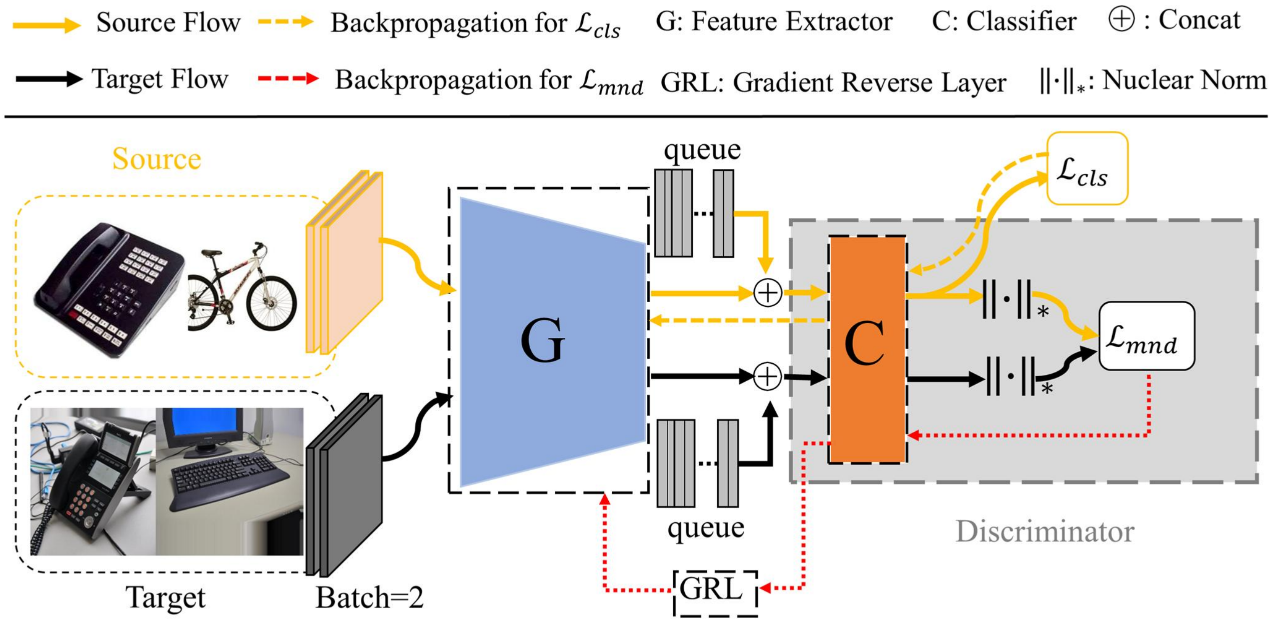


Fig. 2: The schematic of MBAN.

多批量核范数对抗领域自适应网络

（1） Feature queue

The current mini-batch is enqueued and the oldest mini-batch is removed from the queue

（2） Probability rescaling on target domain

$$\hat{Y}_{ij} = \frac{\exp\left(Z_{ij} / T\right)}{\sum_{j'=1}^{K} \exp\left(Z_{ij'} / T\right)},$$

T temperature parameter for probability rescaling

（3） Multi-batch nuclear-norm discrepancy (MND)

$$\mathcal{L}_{mnd} = \frac{1}{n_s^b}\left\|\widehat{\mathbf{Y}}_s\right\|_* - \frac{1}{n_t^b}\left\|\widehat{\mathbf{Y}}_t\right\|_*,$$

$$\widehat{\mathbf{Y}}_s = C\left(\left[G(\mathbf{X}_s) \oplus \mathbf{A}_s\right]\right)$$

$$\widehat{\mathbf{Y}}_t = C\left(\left[G(\mathbf{X}_t) \oplus \mathbf{A}_t\right]\right)$$

Overall function

$$\mathcal{L}_{mban} = \mathcal{L}_{cls} - \beta \mathcal{L}_{mnd},$$

Min supervised training loss on source domain

$$\mathcal{L}_{cls} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathcal{L}_{ce}\left(x_i^s, y_i^s\right)$$

min-max game

$$\min_G \max_C \mathcal{L}_{mnd}.$$

Theoretical insight

Expected error on target domain

$$蝌_t(h) \leq {}_s(h) + 2IW_1(P,Q) + {}_{ideal}(h^*),$$

where $\eta^* = \varepsilon_s(h^*) + \varepsilon_t(h^*)$ *is the error of ideal joint hypothesis* $h^* \in \mathcal{H}$ *on source and target domain, which is a sufficiently small constant.*

Results

TABLE I: Accuracy (%) on (a) Office-Home (ResNet-50) and (b) VisDA-2017 (ResNet-101) for UDA. † denotes that the results are cited from [12]. * denotes that the results are reproduced using the publicly released code.

(a) Office-Home.

Method	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	AVG
ResNet-50 [1]	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DANN [8]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
CDAN [3]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
DSAN [7]	54.4	70.8	75.4	60.4	67.8	68.0	62.6	55.9	78.5	73.8	60.6	83.1	67.6
BNM [18]	52.3	73.9	80.0	63.3	72.9	74.9	61.7	49.5	79.7	70.5	53.6	82.2	67.9
MDD [22]	54.9	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	60.2	82.3	68.1
MCC [19]	55.1	75.2	79.5	63.3	73.2	75.8	66.1	52.1	76.9	73.8	58.4	83.6	69.4
GATE [23]	54.6	76.9	79.8	66.1	73.5	74.2	65.3	54.8	80.6	73.9	59.5	83.7	70.2
MetaAlign [5]	59.3	76.0	80.2	65.7	74.7	75.1	65.7	56.5	81.6	74.1	61.1	85.2	71.3
SCDA [25]	57.5	76.9	80.3	65.7	74.9	74.5	65.5	53.6	79.8	**74.5**	59.6	83.7	70.5
DALN [12]	57.8	**79.9**	82.0	66.3	76.2	77.2	66.7	55.5	81.3	73.5	60.4	85.3	71.8
InfoMLP [21]	59.6	77.3	79.5	67.4	75.9	74.6	66.1	56.4	81.0	**74.5**	61.4	84.4	71.5
MBAN(Ours)	**60.0**	79.4	**82.9**	**67.9**	**77.8**	**79.0**	**66.8**	**57.7**	**82.8**	74.3	**61.8**	**85.6**	**73.0**

(b) VisDA-2017.

Method	AVG
ResNet-101 [1]	52.4
DANN [8]	57.4
MCD [9]	71.9
CDAN [3]	73.9
BNM [18]	70.4
DSAN [7]	75.1
MDD*	77.3
MCC [19]	78.8
DADA [11]	79.8
GATE [23]	74.8
SCDA† [12]	79.7
DALN [12]	80.6
InfoMLP [21]	81.4
MBAN(Ours)	**82.7**

TABLE II: Accuracy (%) on Office-31 for UDA (ResNet-50).

Method	A→W	D→W	W→D	A→D	D→A	W→A	AVG
ResNet-50 [1]	68.4±0.5	96.7±0.5	99.3±0.1	68.9±0.2	62.5±0.3	60.7±0.3	76.1
DANN [8]	82.0±0.4	96.9±0.2	99.1±0.1	79.7±0.4	68.2±0.4	67.4±0.5	82.2
CDAN [3]	94.1±0.1	98.6±0.1	**100.0±0.0**	92.9±0.2	71.0±0.3	69.3±0.3	87.7
DSAN [7]	93.6±0.2	98.3±0.1	**100.0±0.0**	90.2±0.7	73.5±0.5	74.8±0.4	88.4
BNM [18]	91.5	98.5	**100.0**	90.3	70.9	71.6	87.1
MDD [22]	94.5±0.3	98.4±0.1	**100.0±0.0**	93.5±0.2	74.6±0.3	72.2±0.1	88.9
MCC [19]	**95.5±0.2**	98.6±0.1	**100.0±0.0**	94.4±0.3	72.9±0.2	74.9±0.3	89.4
DADA [11]	92.3±0.1	**99.2±0.1**	**100.0±0.0**	93.9±0.2	74.4±0.1	74.2±0.1	89.0
GATE [23]	90.5	98.7	**100.0**	91.3	73.4	75.9	88.3
MetaAlign [5]	93.0±0.5	98.6±0.0	**100.0±0.0**	94.5±0.3	75.0±0.3	73.6±0.0	89.2
SCDA [25]	94.2	98.7	99.8	<u>95.2</u>	75.7	76.2	90.0
DALN [12]	<u>95.2</u>	<u>99.1</u>	**100.0**	**95.4**	76.4	76.5	90.4
InfoMLP [21]	93.3±0.5	99.0±0.1	**100.0±0.0**	93.2±0.3	76.7±0.2	76.2±0.3	89.7
MBAN(w/ output)	95.0±0.7	98.6±0.3	99.8±0.2	95.0±0.3	**77.5±0.5**	<u>77.9±0.3</u>	<u>90.6</u>
MBAN(Ours)	**95.5±0.4**	98.6±0.3	<u>99.9±0.2</u>	95.1±0.3	<u>77.4±0.7</u>	**78.2±0.2**	**90.8**

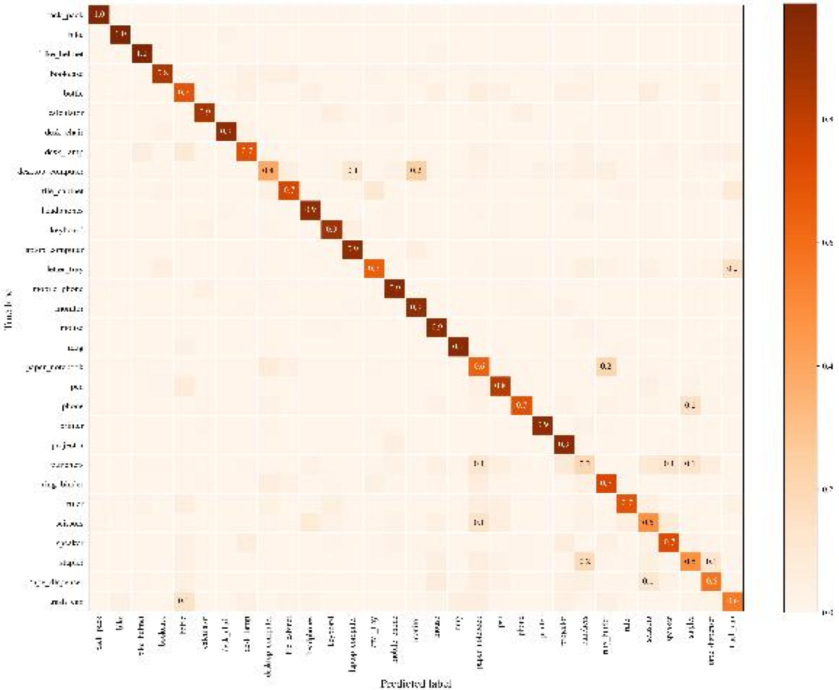
Analysis

Ablation study

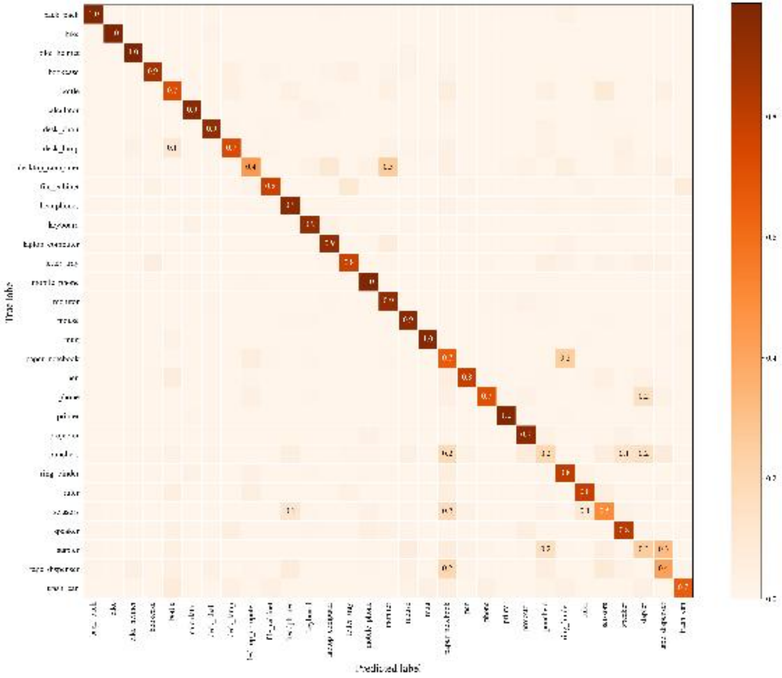
TABLE II: Accuracy (%) on Office-31 for UDA (ResNet-50).

Method	A→W	D→W	W→D	A→D	D→A	W→A	AVG
ResNet-50 [1]	68.4±0.5	96.7±0.5	99.3±0.1	68.9±0.2	62.5±0.3	60.7±0.3	76.1
DANN [8]	82.0±0.4	96.9±0.2	99.1±0.1	79.7±0.4	68.2±0.4	67.4±0.5	82.2
CDAN [3]	94.1±0.1	98.6±0.1	**100.0±0.0**	92.9±0.2	71.0±0.3	69.3±0.3	87.7
DSAN [7]	93.6±0.2	98.3±0.1	**100.0±0.0**	90.2±0.7	73.5±0.5	74.8±0.4	88.4
BNM [18]	91.5	98.5	**100.0**	90.3	70.9	71.6	87.1
MDD [22]	94.5±0.3	98.4±0.1	**100.0±0.0**	93.5±0.2	74.6±0.3	72.2±0.1	88.9
MCC [19]	**95.5±0.2**	98.6±0.1	**100.0±0.0**	94.4±0.3	72.9±0.2	74.9±0.3	89.4
DADA [11]	92.3±0.1	**99.2±0.1**	**100.0±0.0**	93.9±0.2	74.4±0.1	74.2±0.1	89.0
GATE [23]	90.5	98.7	**100.0**	91.3	73.4	75.9	88.3
MetaAlign [5]	93.0±0.5	98.6±0.0	**100.0±0.0**	94.5±0.3	75.0±0.3	73.6±0.0	89.2
SCDA [25]	94.2	98.7	99.8	95.2	75.7	76.2	90.0
DALN [12]	95.2	99.1	**100.0**	**95.4**	76.4	76.5	90.4
InfoMLP [21]	93.3±0.5	99.0±0.1	**100.0±0.0**	93.2±0.3	76.7±0.2	76.2±0.3	89.7
MBAN(w/ output)	95.0±0.7	98.6±0.3	99.8±0.2	95.0±0.3	**77.5±0.5**	77.9±0.3	90.6
MBAN(Ours)	**95.5±0.4**	98.6±0.3	99.9±0.2	95.1±0.3	77.4±0.7	**78.2±0.2**	**90.8**

Confusion Matrix

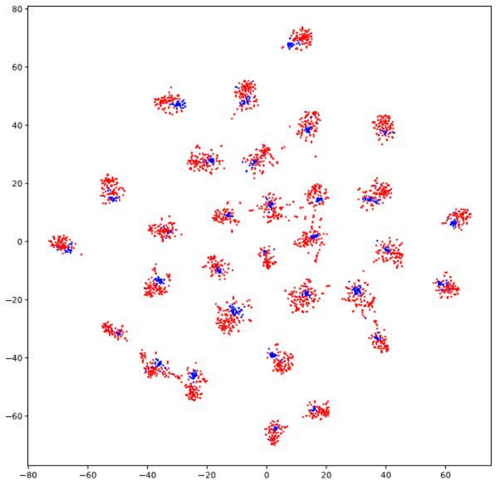


DALN

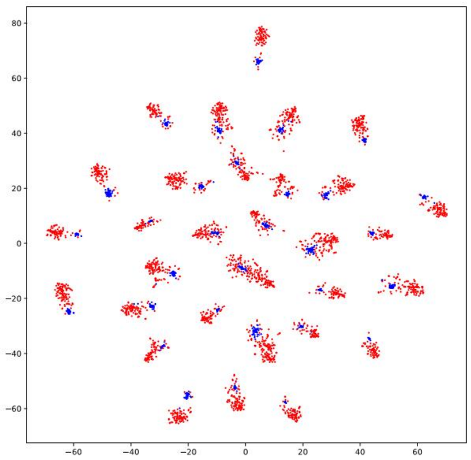


MBAN

Feature visualization

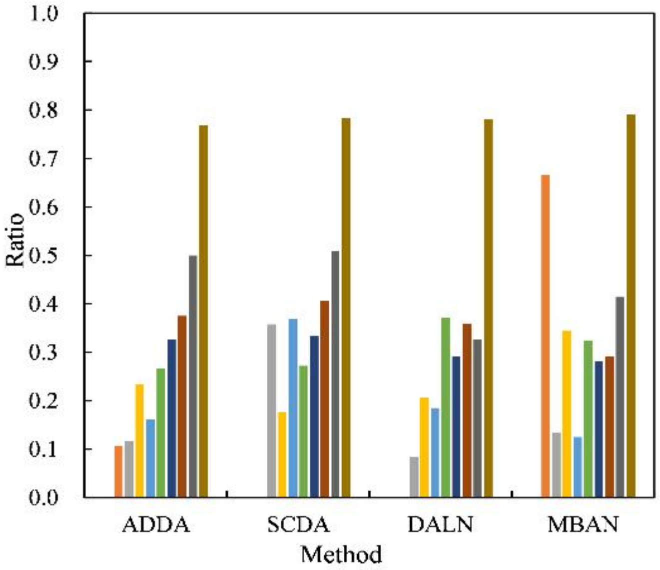


(b) DALN



(b) MBAN

Determinacy



Determinacy

Parameter Sensitivity

TABLE III: Accuracy (%) on Office-31 for sensitivity of queue length c. (ResNet-50).

queue length c	1	2	3	4
Avg	90.3	90.6	90.8	90.7

Thanks!

Multi-batch Nuclear-norm Adversarial Network for Unsupervised Domain Adaptation

Pei Wang, Yun Yang*, Zhenyu Yu

Code

Email