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

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

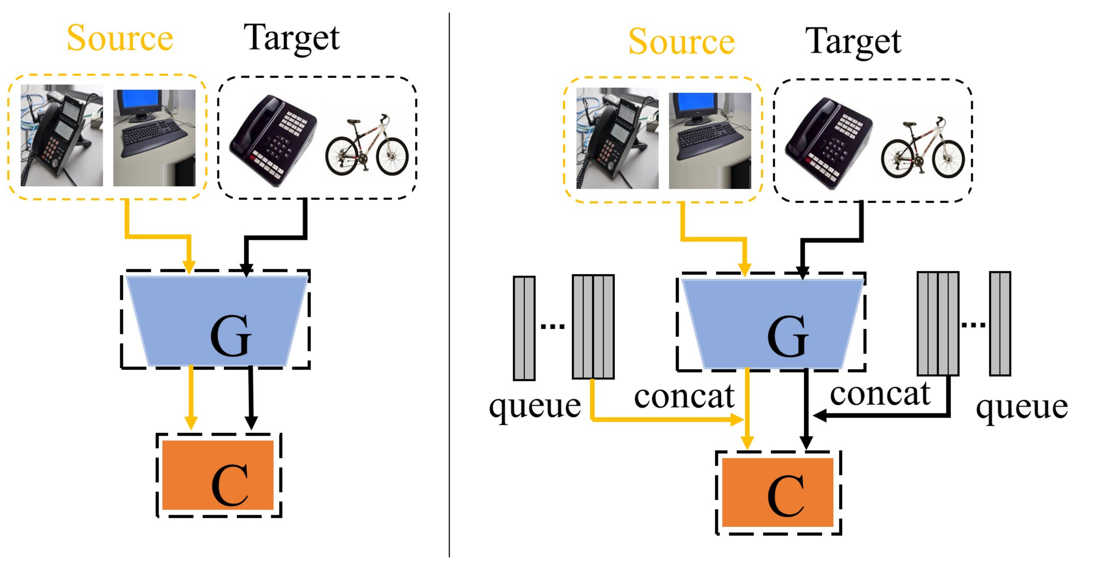
[1]School of Information Science and Engineering, Yunnan University, China
[2]National Pilot School of Software, Yunnan University, China
[3]Yunnan Key Laboratory of Software Engineering, Yunnan University, China
[4]Guangzhou University, China [5]CISiP, Universiti Malaya, Malaysia

Summary

- A discriminator-free adaptation network: Multi-batch Nuclear-norm Adversarial Network (MBAN)
- Three technical 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

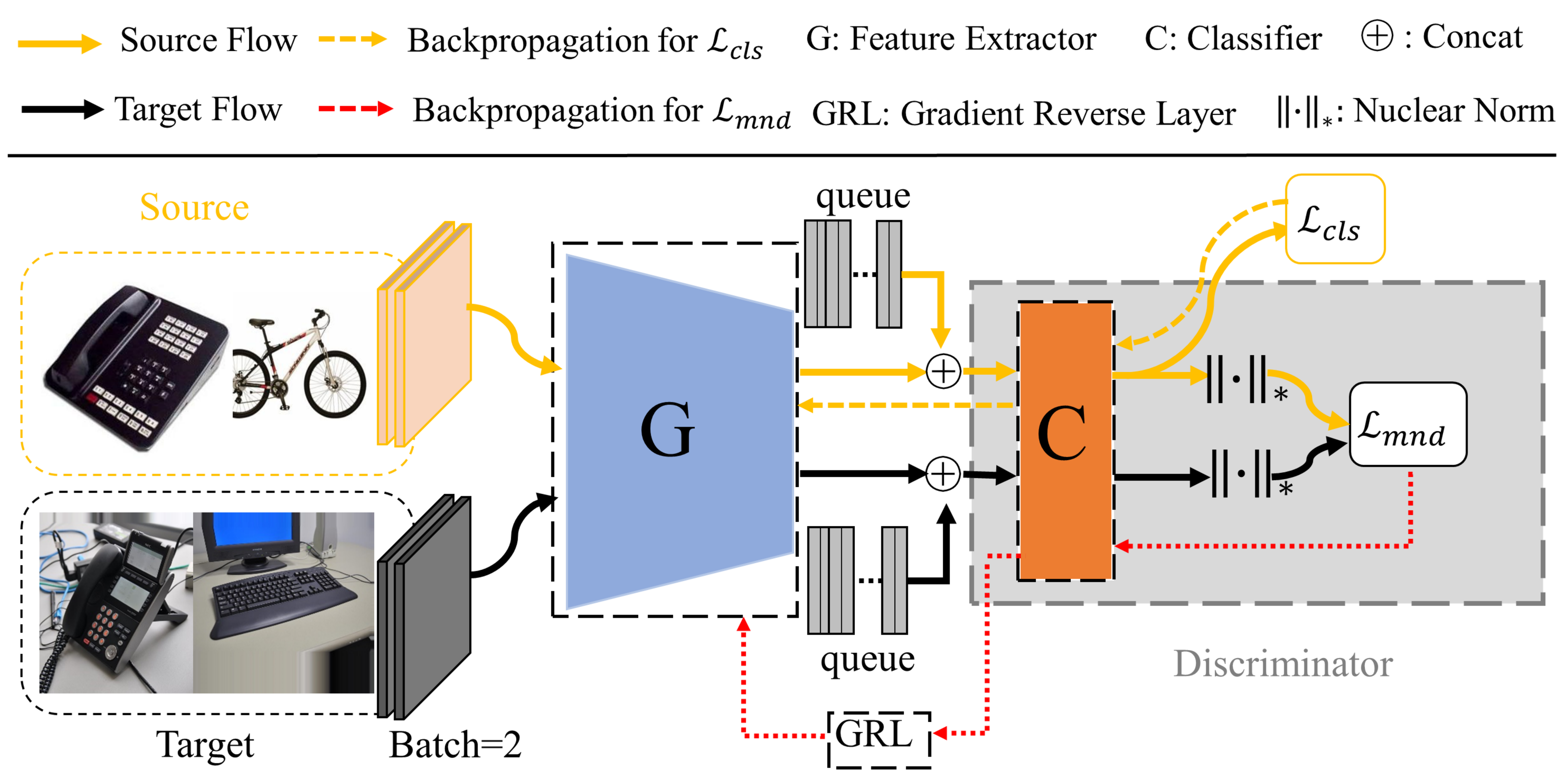
- DALN reuses a classifier as the discriminator, which achieves feature alignment via min-max the NWD of the output matrix
- MBAN builds a dynamic feature queue, which generates a large and consistent output matrix



(a) DALN (b) MBAN(Ours)

The main idea of MBAN

- The classifier is coupled with a new discrepancy (MND) as a domain discriminator, enabling it leverages the predicted discriminative information to align class-level features.



Feature queue

- Build a dynamic feature queue. This allows us to reuse previously mini-batch to generate a large and consistent outputs
- The features in the queue are progressively replaced. The current mini-batch is enqueued and the oldest mini-batch is removed from the queue
- The queue size is an independent hyperparameter

Probability rescaling on target domain

Temperature scaling to calibrate the predicted outputs on target domain, which can control the smoothness of the predicted outputs. The probability that the *i*-th sample belongs to *j*-class can be calibrated as:

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

where Z_{ij} denotes the logit output of the fully connected layer in the classifier, T denotes the temperature parameter for probability rescaling.

Multi-batch nuclear-norm discrepancy

The Multi-batch Nuclear-norm discrepancy (MND), which gives higher scores to source domain and lower scores to target domain.

$$\mathrm{L}_{mnd} = \frac{1}{n_s^b}\left\|\mathbf{Y}_s\right\|_* - \frac{1}{n_t^b}\left\|\mathbf{Y}_t\right\|_*, \ \mathbf{Y}_s = C\left(\left[G(\mathbf{X}_s)\oplus\mathbf{A}_s\right]\right) \quad \mathbf{Y}_t = C\left(\left[G(\mathbf{X}_t)\oplus\mathbf{A}_t\right]\right) \quad (2)$$

where the matrix $\mathbf{A}$ represents the cached feature queue, and $\hat{Y}$ denotes the predicted outputs of the concatenated feature matrix.

Overall function

$$\mathrm{L}_{mban} = \mathrm{L}_{cls} - \beta\mathrm{L}_{mnd}, \quad \mathrm{L}_{cls} = \frac{1}{n_s}\sum_{i=1}^{n_s}\mathrm{L}_{ce}\left(x_i^s, y_i^s\right) \quad (3)$$

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

Results

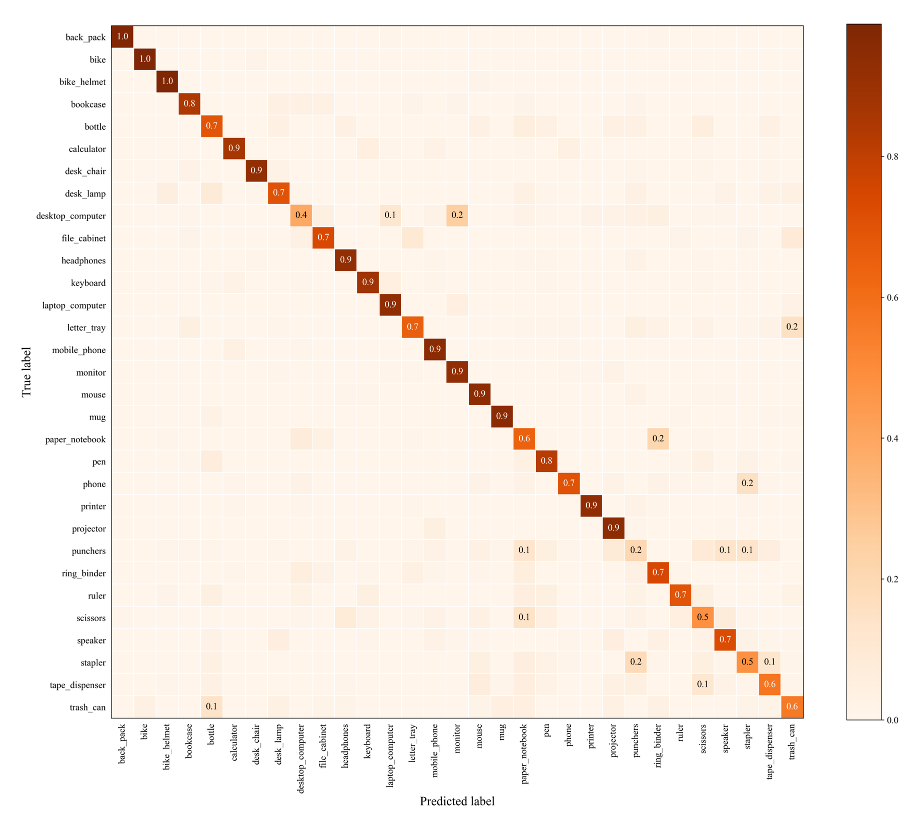
Table : Accuracy (%) on VisDA-2017 (ResNet-101) for UDA

Method	plane	bcybl	bus	car	horse	knife	mcyle	persn	plant	sktb	train	truck	Avg
ResNet-101	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
DANN	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
MCD	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
CDAN	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38.0	73.9
BNM	89.6	61.5	76.9	55.0	89.3	69.1	81.3	65.5	90.0	47.3	**89.1**	30.1	70.4
DSAN	90.9	66.9	75.7	62.4	88.9	77.0	**93.7**	75.1	92.8	67.6	**89.1**	39.4	75.1
MDD	94.7	79.7	75.0	55.1	92.2	67.1	83.1	77.2	84.9	84.6	82.6	51.7	77.3
MCC	88.1	80.3	80.5	**71.5**	90.1	93.2	85.0	71.6	89.4	73.8	85.0	36.9	78.8
DADA	92.9	74.2	82.5	65.0	90.9	93.8	87.2	74.2	89.9	71.5	86.5	48.7	79.8
GATE	-	-	-	-	-	-	-	-	-	-	-	-	74.8
SCDA†	93.1	84.6	78.2	52.2	90.8	95.2	81.0	77.2	91.1	80.5	**89.1**	43.5	79.7
DALN	**96.0**	**86.3**	74.3	50.0	92.4	94.7	83.5	76.4	91.0	87.2	88.4	47.4	80.6
InfoMLP	94.5	85.4	77.2	65.2	**94.8**	82.3	86.1	**81.4**	**93.0**	77.4	88.6	50.5	81.4
MBAN	95.4	84.4	80.5	54.7	92.0	**96.4**	87.8	78.7	88.8	**89.7**	86.8	**57.3**	**82.7**

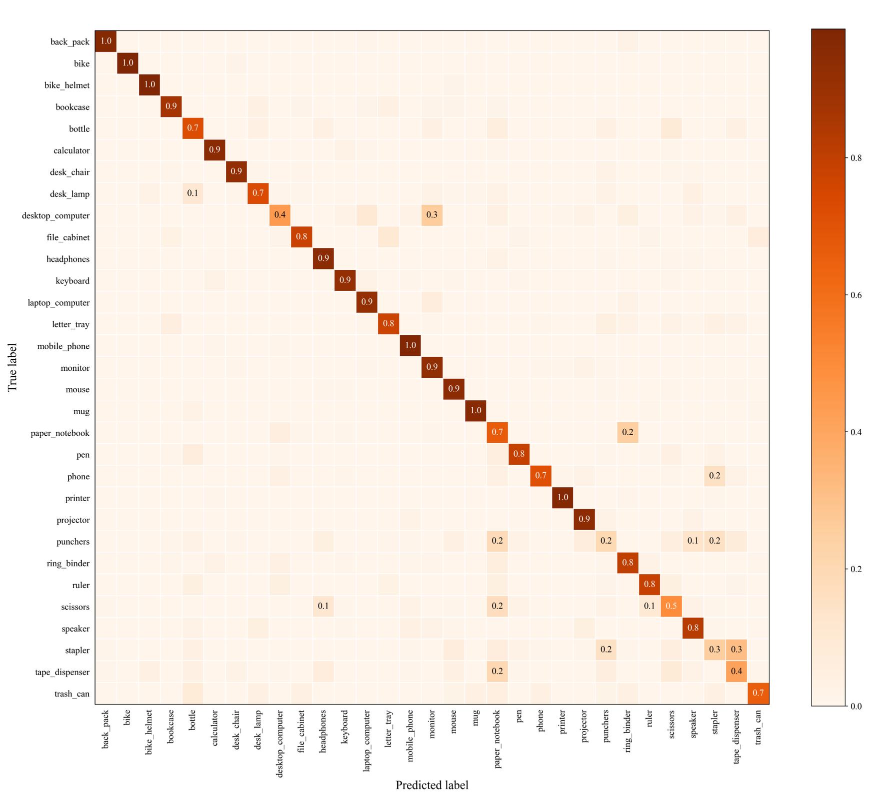
Table : Accuracy (%) on Office-31 (ResNet-50) for UDA

Method	A→W	D→W	W→D	A→D	D→A	W→A	AVG
ResNet-50	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	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	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	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	91.5	98.5	**100.0**	90.3	70.9	71.6	87.1
MDD	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	**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	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	90.5	98.7	**100.0**	91.3	73.4	75.9	88.3
MetaAlign	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	94.2	98.7	99.8	95.2	75.7	76.2	90.0
DALN	95.2	99.1	**100.0**	**95.4**	76.4	76.5	90.4
InfoMLP	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	**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**

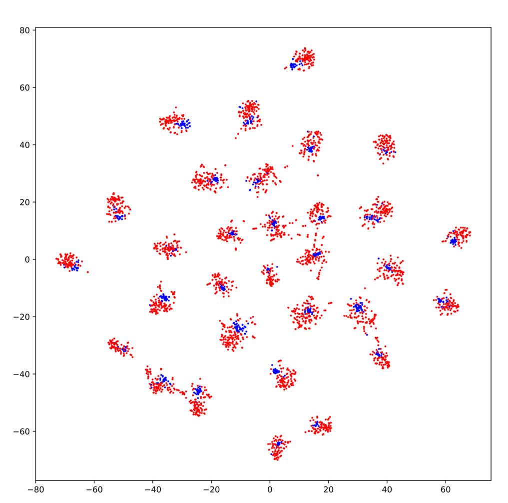
Analysis



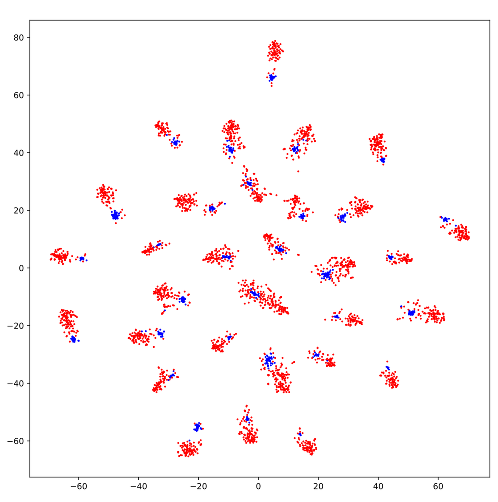
(a) DALN



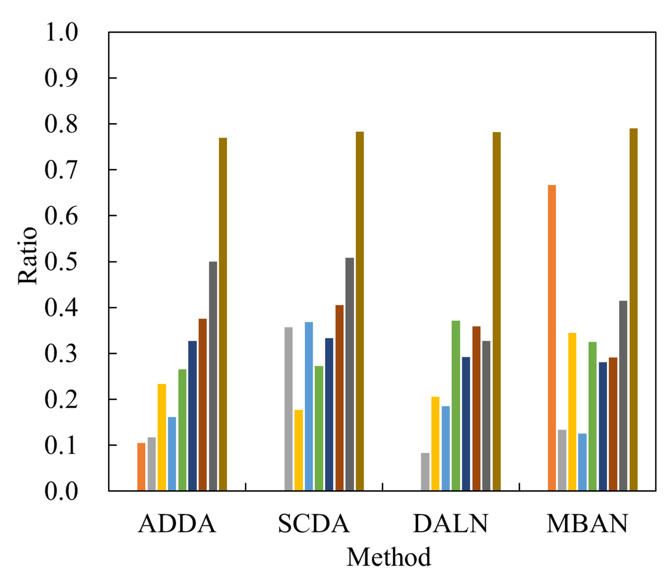
(b) MBAN(Ours)



(c) DALN



(d) MBAN(Ours)



(e) Determinacy

Mail: peiwang@mail.ynu.edu.cn