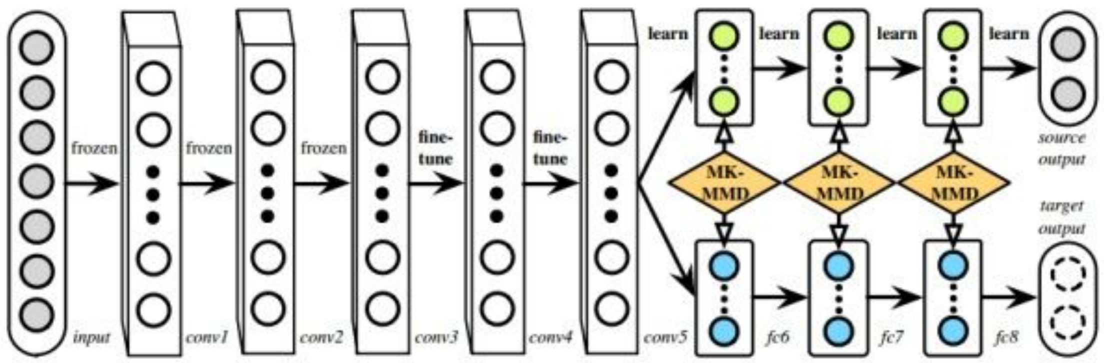
Minimum Class Confusion for Versatile Domain Adaptation

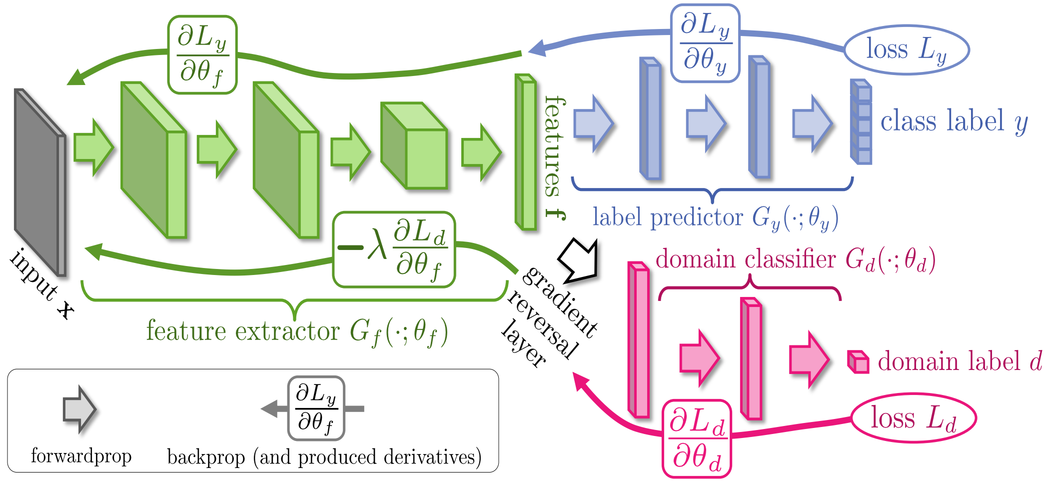
Ying Jin, Ximei Wang, Mingsheng Long, Jianmin Wang
Tsinghua University

Presented at ECCV2020

Domain Adaptation (DA)



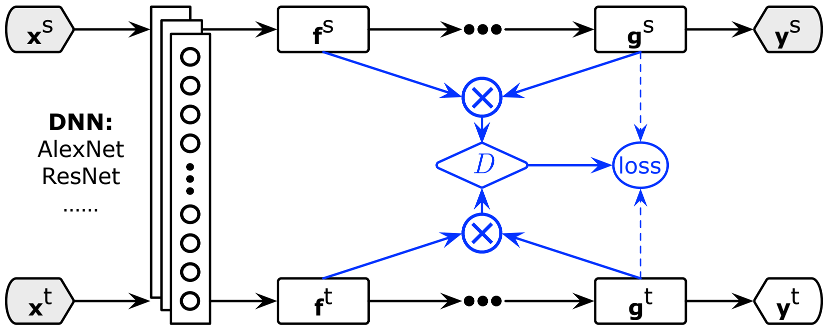
Moment Matching



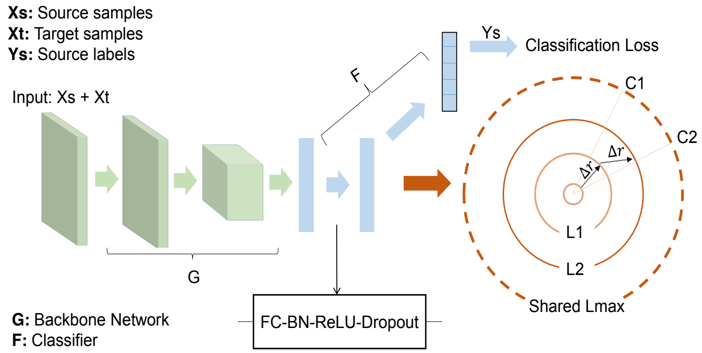
Adversarial Training

Long et al. ICML15
Ganin et al. JMLR16

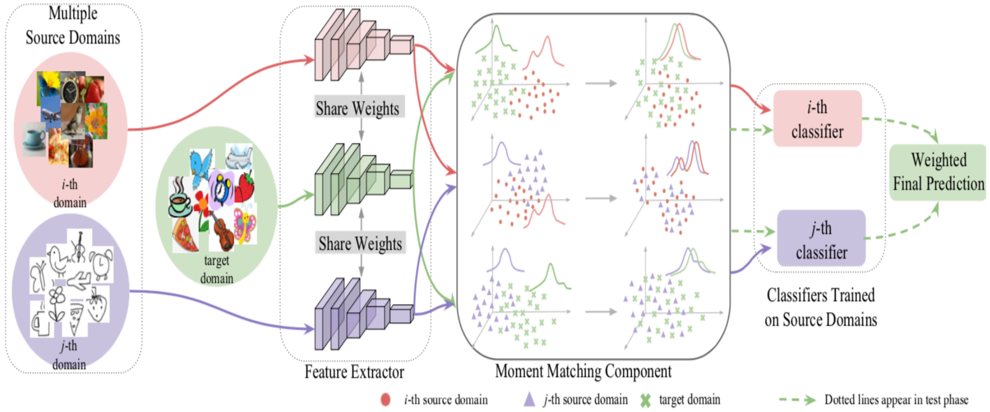
DA: Different Scenarios



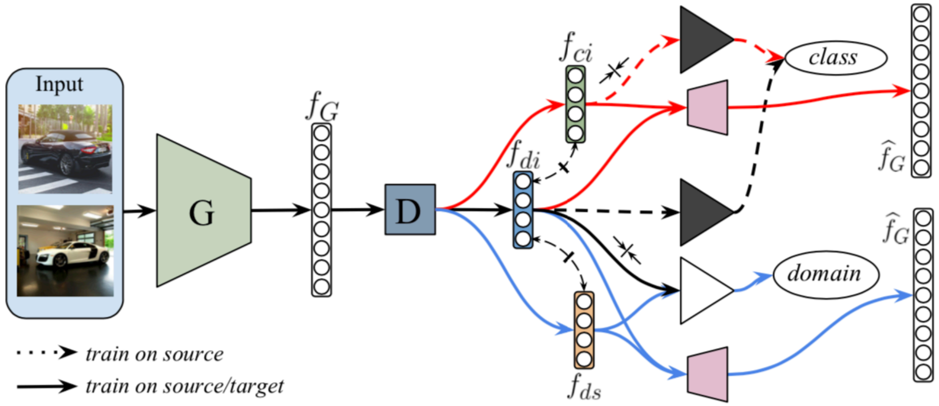
CDAN: Unsupervised DA (UDA)



AFN: UDA + Partial DA (PDA)



M^{3}SDA: Multi-Source DA (MSDA)

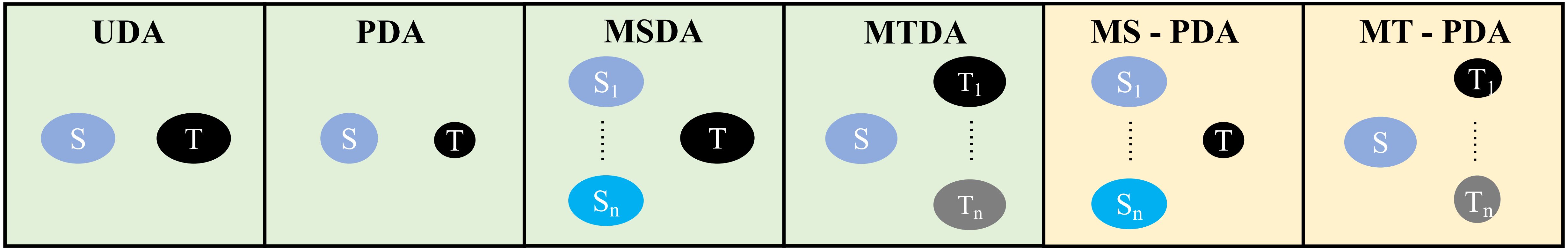


DADA: Multi-Target DA (MTDA)

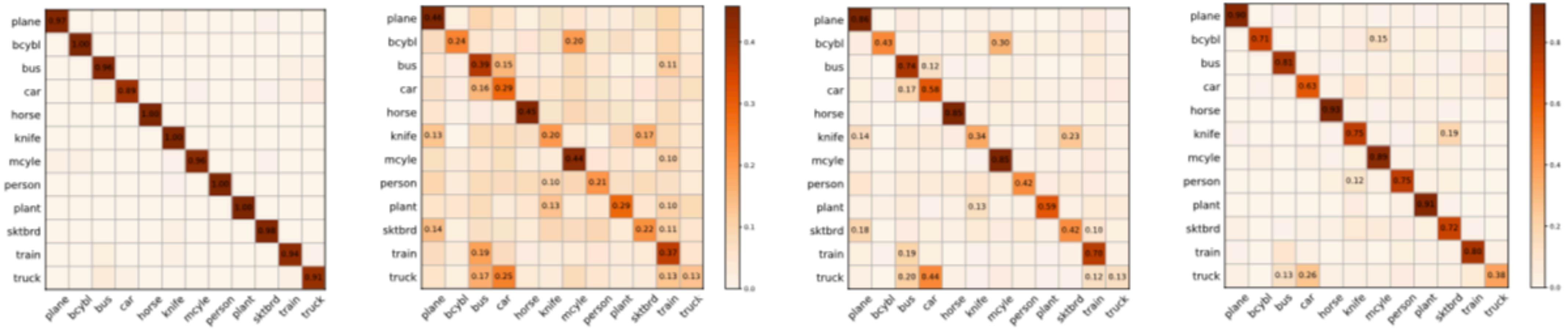
Long et al. NeurIPS18, Xu et al. ICCV19
Peng et al. ICCV19, Peng et al. ICML19

Versatile Domain Adaptation (VDA)

1. A variety of DA scenarios: closed-set, partial-set DA, multi-source and multi-target DA.
2. Existing DA methods: designed only for a specific scenario, and may underperform for scenarios they are not tailored to.
3. Practical applications, complicated data acquired in the real-world makes it difficult to confirm the label sets and domain configurations.
4. We need a versatile method which can tackle different scenarios at the same time.



Minimum Class Confusion (MCC)



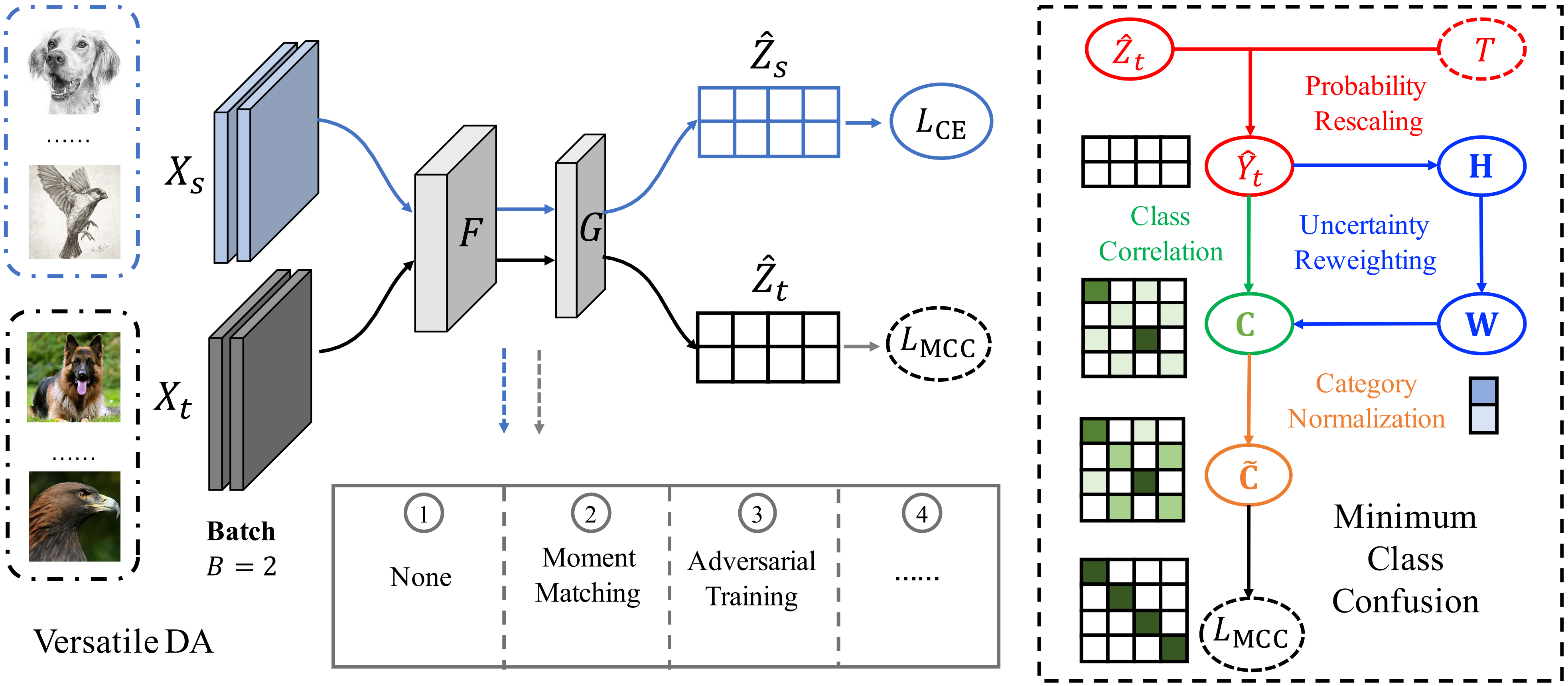
(a) Source (b) Target (c) MinEnt (d) MCC

The classifier trained on the source domain may confuse to distinguish the correct class from a similar class.

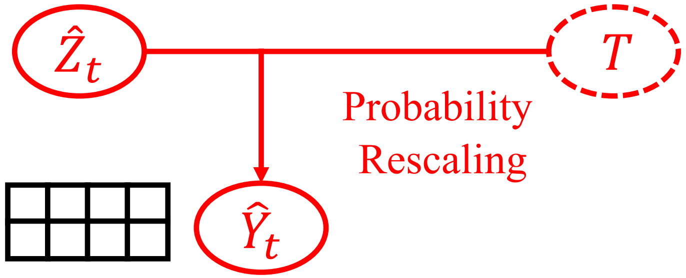
Class Confusion

Less Class Confusion, More Transfer Gains!

Minimum Class Confusion (MCC)



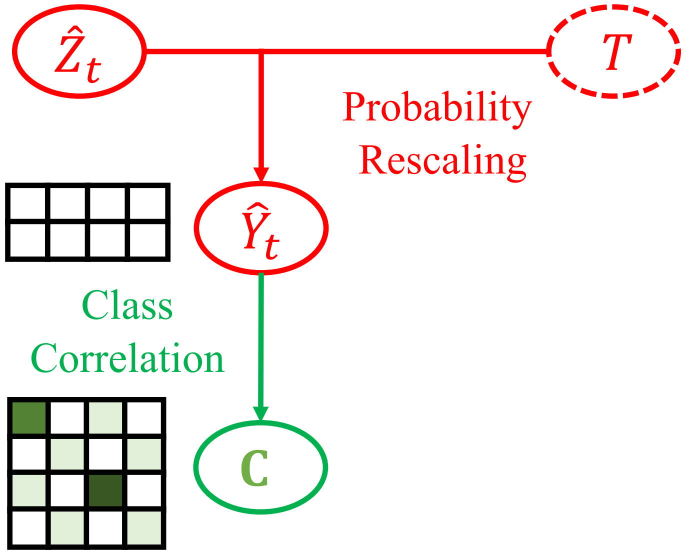
Minimum Class Confusion (MCC)



Temperature rescaling

$$\widehat{Y}_{ij} = \frac{\exp\left(Z_{ij}/T\right)}{\sum_{j'=1}^{|\mathcal{C}|} \exp\left(Z_{ij'}/T\right)},$$

Minimum Class Confusion (MCC)



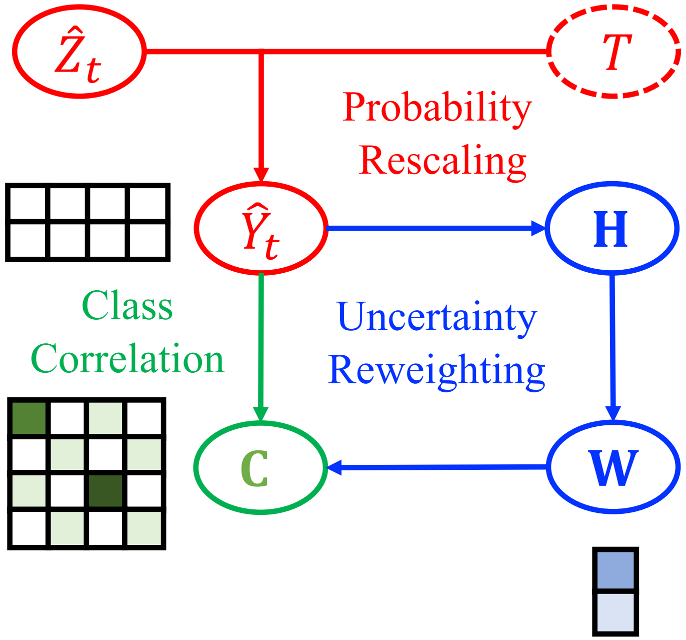
Temperature rescaling

$$\widehat{Y}_{ij} = \frac{\exp\left(Z_{ij}/T\right)}{\sum_{j'=1}^{|\mathcal{C}|} \exp\left(Z_{ij'}/T\right)}$$

Pair-wise class correlation

$$\mathbf{C}_{jj'} = \widehat{\mathbf{y}}_{\cdot j}^{\mathsf{T}} \widehat{\mathbf{y}}_{\cdot j'}$$

Minimum Class Confusion (MCC)



Temperature rescaling

$$\widehat{Y}_{ij} = \frac{\exp(Z_{ij}/T)}{\sum_{j'=1}^{|\mathcal{C}|} \exp(Z_{ij'}/T)}$$

Pair-wise class correlation

$$\mathbf{C}_{jj'} = \widehat{\mathbf{y}}_{\cdot j}^{\mathsf{T}} \widehat{\mathbf{y}}_{\cdot j'}$$

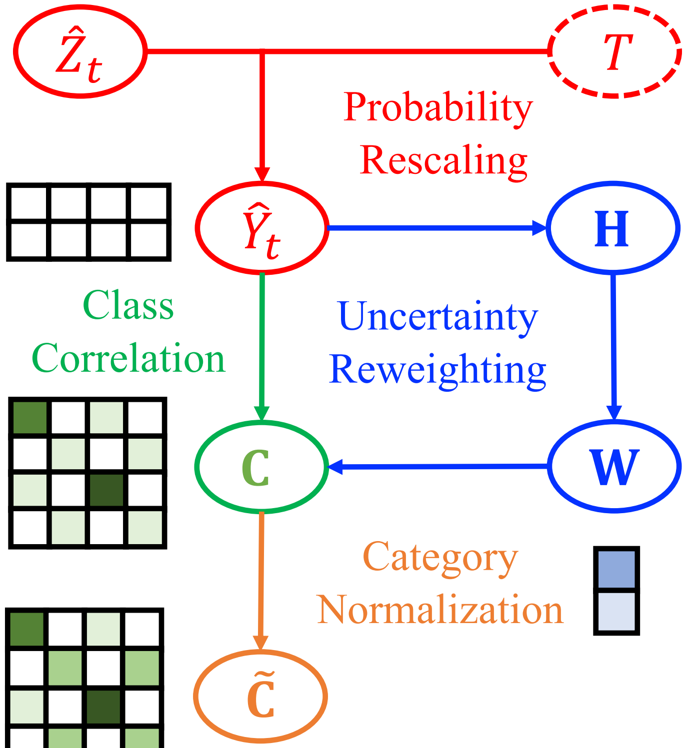
$$H(\widehat{\mathbf{y}}_{i\cdot}) = -\sum_{j=1}^{|\mathcal{C}|} \widehat{Y}_{ij} \log \widehat{Y}_{ij}.$$

Weighting Mechanism

$$W_{ii} = \frac{B\left(1+\exp(-H(\widehat{\mathbf{y}}_{i\cdot}))\right)}{\sum\limits_{i'=1}^{B}\left(1+\exp(-H(\widehat{\mathbf{y}}_{i'\cdot}))\right)}$$

$$\mathbf{C}_{jj'} = \widehat{\mathbf{y}}_{\cdot j}^{\mathsf{T}} \mathbf{W} \widehat{\mathbf{y}}_{\cdot j'}$$

Minimum Class Confusion (MCC)



Temperature rescaling

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

Pair-wise class correlation

$$\mathbf{C}_{jj'} = \hat{\mathbf{y}}_{\cdot j}^{\mathsf{T}}\hat{\mathbf{y}}_{\cdot j'}$$

$$H(\hat{\mathbf{y}}_{i\cdot}) = -\sum_{j=1}^{|\mathcal{C}|}\hat{Y}_{ij}\log\hat{Y}_{ij}$$

Weighting Mechanism

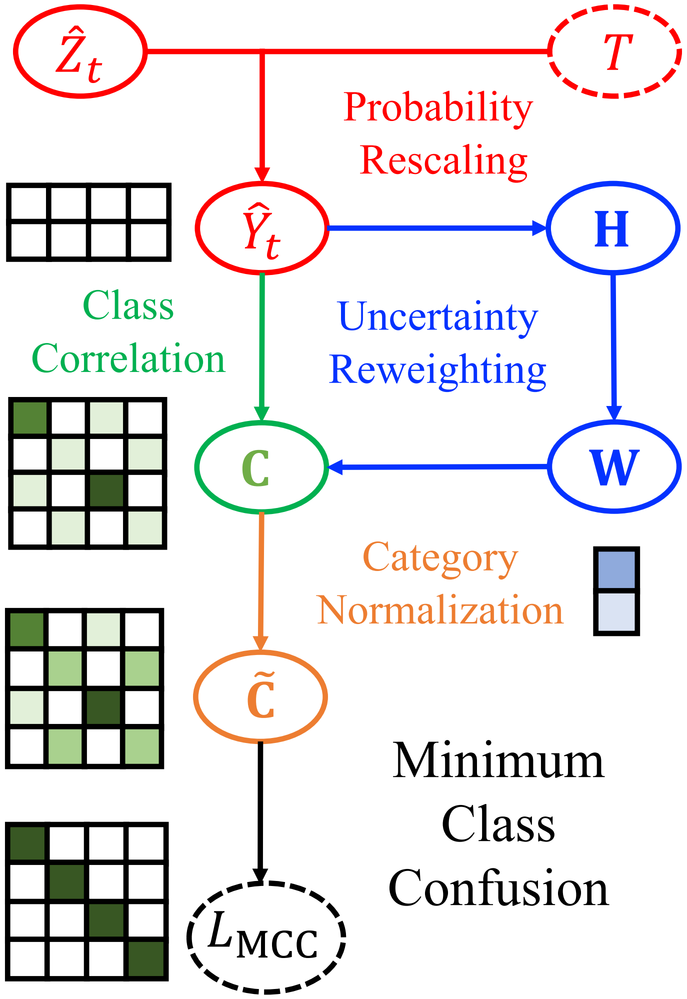
$$W_{ii} = \frac{B\left(1+\exp(-H(\hat{\mathbf{y}}_{i\cdot}))\right)}{\sum\limits_{i'=1}^{B}\left(1+\exp(-H(\hat{\mathbf{y}}_{i'\cdot}))\right)}$$

$$\mathbf{C}_{jj'} = \hat{\mathbf{y}}_{\cdot j}^{\mathsf{T}}\mathbf{W}\hat{\mathbf{y}}_{\cdot j'}$$

Category Normalization
in Random Walk

$$\tilde{\mathbf{C}}_{jj'} = \frac{\mathbf{C}_{jj'}}{\sum_{j''=1}^{|\mathcal{C}|}\mathbf{C}_{jj''}}$$

Minimum Class Confusion (MCC)



Temperature rescaling

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

Pair-wise class correlation

$$\mathbf{C}_{jj'} = \hat{\mathbf{y}}_{\cdot j}^{\mathsf{T}}\hat{\mathbf{y}}_{\cdot j'}$$

$$H(\hat{\mathbf{y}}_{i\cdot}) = -\sum_{j=1}^{|\mathcal{C}|}\hat{Y}_{ij}\log\hat{Y}_{ij}.$$

Weighting Mechanism

$$W_{ii} = \frac{B\left(1+\exp(-H(\hat{\mathbf{y}}_{i\cdot}))\right)}{\sum\limits_{i'=1}^{B}\left(1+\exp(-H(\hat{\mathbf{y}}_{i'\cdot}))\right)},$$

$$\mathbf{C}_{jj'} = \hat{\mathbf{y}}_{\cdot j}^{\mathsf{T}}\mathbf{W}\hat{\mathbf{y}}_{\cdot j'}$$

Category Normalization in Random Walk

$$\tilde{\mathbf{C}}_{jj'} = \frac{\mathbf{C}_{jj'}}{\sum_{j''=1}^{|\mathcal{C}|}\mathbf{C}_{jj''}}.$$

MCC Loss

$$L_{\text{MCC}}(\hat{\mathbf{Y}}_t) = \frac{1}{|\mathcal{C}|}\sum_{j=1}^{|\mathcal{C}|}\sum_{j'\neq j}^{|\mathcal{C}|}\left|\tilde{\mathbf{C}}_{jj'}\right|.$$

Minimum Class Confusion (MCC)

A Versatile Approach

$$\min_{F,G} \mathbb{E}_{(\mathbf{x}_s,\mathbf{y}_s)\in\mathcal{S}} L_{\mathrm{CE}}\left(\widehat{\mathbf{y}}_s, \mathbf{y}_s\right) + \mu \, \mathbb{E}_{\mathbf{X}_t\subset\mathcal{T}} L_{\mathrm{MCC}}(\widehat{\mathbf{Y}}_t).$$

A General Reguralizer

$$\min_{F,G} \max_{D} \mathbb{E}_{(\mathbf{x}_s,\mathbf{y}_s)\in\mathcal{S}} L_{\mathrm{CE}}\left(\widehat{\mathbf{y}}_s, \mathbf{y}_s\right) - \lambda \mathbb{E}_{\mathbf{x}\in\mathcal{S}\cup\mathcal{T}} L_{\mathrm{CE}}(D(\widehat{\mathbf{f}}), \mathbf{d}) + \mu \, \mathbb{E}_{\mathbf{X}_t\subset\mathcal{T}} L_{\mathrm{MCC}}(\widehat{\mathbf{Y}}_t).$$

Results – MTDA & MSDA

(a) MTDA

Method	c:	i:	p:	q:	r:	s:	Avg
ResNet	25.6	16.8	25.8	9.2	20.6	22.3	20.1
SE	21.3	8.5	14.5	13.8	16.0	19.7	15.6
MCD	25.1	19.1	27.0	10.4	20.2	22.5	20.7
DADA	26.1	20.0	26.5	12.9	20.7	22.8	21.5
MCC	**33.6**	**30.0**	**32.4**	**13.5**	**28.0**	**35.3**	**28.8**

(b) MSDA

Method	:c	:i	:p	:q	:r	:s	Avg
ResNet	47.6	13.0	38.1	13.3	51.9	33.7	32.9
MCD	54.3	22.1	45.7	7.6	58.4	43.5	38.5
DCTN	48.6	23.5	48.8	7.2	53.5	47.3	38.2
M^3SDA	58.6	26.0	52.3	6.3	62.7	49.5	42.6
MCC	**65.5**	**26.0**	**56.6**	**16.5**	**68.0**	**52.7**	**47.6**

A big margin on DomainNet, the largest and hardest dataset to date.

Results – PDA

Table 2: Accuracy (%) on Office-Home for PDA (ResNet-50).

Method (S:T)	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 [12]	38.6	60.8	75.2	39.9	48.1	52.9	49.7	30.9	70.8	65.4	41.8	70.4	53.7
DAN [20]	44.4	61.8	74.5	41.8	45.2	54.1	46.9	38.1	68.4	64.4	51.5	74.3	56.3
JAN [23]	45.9	61.2	68.9	50.4	59.7	61.0	45.8	43.4	70.3	63.9	52.4	76.8	58.3
PADA [3]	51.2	67.0	78.7	52.2	53.8	59.0	52.6	43.2	78.8	73.7	56.6	77.1	62.0
AFN [48]	58.9	76.3	81.4	70.4	**73.0**	77.8	72.4	55.3	80.4	75.8	60.4	79.9	71.8
MCC	**63.1**	**80.8**	**86.0**	**70.8**	72.1	**80.1**	**75.0**	**60.8**	**85.9**	**78.6**	**65.2**	**82.8**	**75.1**

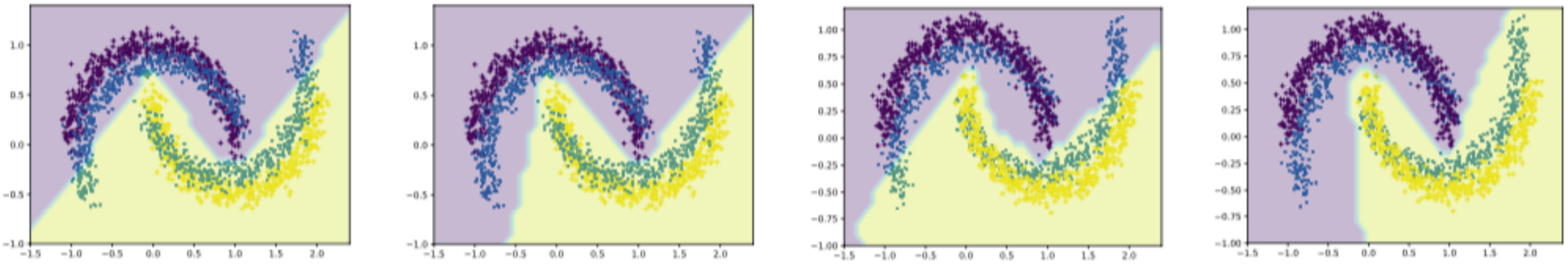
Results – UDA

Table 3: Accuracy (%) on VisDA-2017 for UDA (ResNet-101).

Method	plane	bcybl	bus	car	horse	knife	mcyle	person	plant	sktbrd	train	truck	mean
ResNet [12]	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
MinEnt [9]	80.3	75.5	75.8	48.3	77.9	27.3	69.7	40.2	46.5	46.6	79.3	16.0	57.0
DANN [7]	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
DAN [20]	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
MCD [36]	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 [21]	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
ADR [35]	87.8	79.5	83.7	65.3	92.3	61.8	88.9	73.2	87.8	60.0	85.5	32.3	74.8
AFN [48]	**93.6**	61.3	**84.1**	70.6	**94.1**	79.0	**91.8**	**79.6**	**89.9**	55.6	**89.0**	24.4	76.1
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**

Table 4: Accuracy (%) on Office-31 for UDA (ResNet-50).

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg
ResNet [12]	68.4±0.2	96.7±0.1	99.3±0.1	68.9±0.2	62.5±0.3	60.7±0.3	76.1
DAN [20]	80.5±0.4	97.1±0.2	99.6±0.1	78.6±0.2	63.6±0.3	62.8±0.2	80.4
RTN [22]	84.5±0.2	96.8±0.1	99.4±0.1	77.5±0.3	66.2±0.2	64.8±0.3	81.6
DANN [7]	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
JAN [23]	85.4±0.3	97.4±0.2	99.8±0.2	84.7±0.3	68.6±0.3	70.0±0.4	84.3
MADA [29]	90.0±0.1	97.4±0.1	99.6±0.1	87.8±0.2	70.3±0.3	66.4±0.3	85.2
MinEnt [9]	92.5±0.4	98.0±0.2	99.8±0.2	92.6±0.3	70.3±0.2	63.1±0.2	86.1
SimNet [33]	88.6±0.5	98.2±0.2	99.7±0.2	85.3±0.3	73.4±0.8	71.6±0.6	86.2
GTA [37]	89.5±0.5	97.9±0.3	99.8±0.4	87.7±0.5	72.8±0.3	71.4±0.4	86.5
CDAN [21]	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
AFN [48]	88.8±0.5	98.4±0.3	99.8±0.1	87.7±0.6	69.8±0.4	69.7±0.4	85.7
MDD [53]	94.5±0.3	98.4±0.3	**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**



(a) MinEnt [9] (b) MCC (c) DANN+MinEnt (d) DANN+MCC

Results – MSPDA & MTPDA

Table 5: Accuracy (%) for Multi-Source and Multi-target Partial DA.

(a) MSPDA

	:A	:C	:P	:R	Avg
DANN	58.3	43.6	60.7	71.2	58.5
PADA	62.8	51.8	71.7	79.2	66.4
M^3SDA	67.4	55.3	72.2	80.4	68.8
AFN	77.1	61.2	79.3	82.5	75.0
MCC	**79.6**	**67.5**	**80.6**	**85.1**	**78.2**

(b) MTPDA

	A:	C:	P:	R:	Avg
DANN	44.6	44.8	39.1	44.1	43.1
PADA	59.9	53.7	51.1	61.4	56.5
DADA	65.1	63.0	60.4	63.0	62.9
AFN	68.7	65.6	63.4	67.5	66.3
MCC	**73.1**	**72.1**	**69.4**	**68.3**	**70.7**

Results – Regularizer

Table 6: Accuracy (%) on VisDA-2017 as a *regularizer* for UDA (ResNet-101).

Method	plane	bcybl	bus	car	horse	knife	mcyle	persn	plant	sktb	train	truck	mean	
DANN [7]	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	
DANN + MinEnt [9]	87.4	55.0	75.3	**63.8**	87.4	43.6	**89.3**	72.5	82.9	**78.6**	85.6	27.4	70.7	22.0 %
DANN + BSP [5]	**92.2**	72.5	**83.8**	47.5	87.0	54.0	86.8	72.4	80.6	66.9	84.5	37.1	72.1	
DANN + MCC	90.4	**79.8**	72.3	55.1	**90.5**	**86.8**	86.6	**80.0**	**94.2**	76.9	**90.0**	**49.6**	**79.4**	
CDAN [21]	85.2	66.9	**83.0**	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38	73.9	
CDAN + MinEnt [9]	90.5	65.8	79.1	62.2	89.8	28.7	**92.8**	75.4	86.8	65.3	85.2	35.3	71.4	6.5 %
CDAN + BSP [5]	92.4	61.0	81.0	57.5	89.0	**80.6**	90.1	77.0	84.2	77.9	82.1	38.4	75.9	
CDAN + MCC	**94.5**	**80.8**	78.4	**65.3**	**90.6**	79.4	87.5	**82.2**	**94.7**	**81.0**	**86.0**	**44.6**	**80.4**	

Table 7: Accuracy (%) on Office-31 as a *regularizer* for UDA (ResNet-50).

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg	
DANN [7]	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	
DANN + MinEnt [9]	91.7±0.3	98.3±0.1	**100.0**±0.0	87.9±0.3	68.8±0.3	68.1±0.3	85.8	7.2 %
DANN + BSP [5]	93.0±0.2	98.0±0.2	**100.0**±0.0	90.0±0.4	71.9±0.3	73.0±0.3	87.7	
DANN + MCC	**95.6**±0.3	**98.6**±0.1	99.3±0.0	**93.8**±0.4	**74.0**±0.3	**75.0**±0.4	**89.4**	
CDAN [21]	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	
CDAN + MinEnt [9]	91.7±0.2	98.5±0.1	**100.0**±0.0	90.4±0.3	72.3±0.2	69.5±0.2	87.1	1.5 %
CDAN + BSP [5]	93.3±0.2	98.2±0.2	**100.0**±0.0	93.0±0.2	**73.6**±0.3	72.6±0.3	88.5	
CDAN + MCC	**94.7**±0.2	**98.6**±0.1	**100.0**±0.0	**95.0**±0.1	73.0±0.2	**73.6**±0.3	**89.2**	
AFN [48]	88.8±0.5	98.4±0.3	99.8±0.1	87.7±0.6	69.8±0.4	69.7±0.4	85.7	
AFN + MinEnt [9]	90.3±0.4	**98.7**±0.2	**100.0**±0.0	92.1±0.5	73.4±0.3	71.2±0.3	87.6	4.3 %
AFN + BSP [5]	89.7±0.4	98.0±0.2	99.8±0.1	91.0±0.4	71.4±0.3	71.4±0.2	86.9	
AFN + MCC	**95.4**±0.3	98.6±0.2	**100.0**±0.0	**96.0**±0.2	**74.6**±0.3	**75.2**±0.2	**90.0**	

Conclusion

- We propose Versatile Domain Adaptation (VDA) ;
- A novel loss function: Minimum Class Confusion (MCC)
    - A versatile domain adaptation method that can handle various DA scenarios
    - Strong performance in VDA
    - A general regularizer for existing DA methods

- We are looking forward to seeing
    - Effective methods for VDA
    - Other researchers combine our MCC with their methods to improve performance

Code is available at https://github.com/thuml/MCC

Questions ?

Thank you !