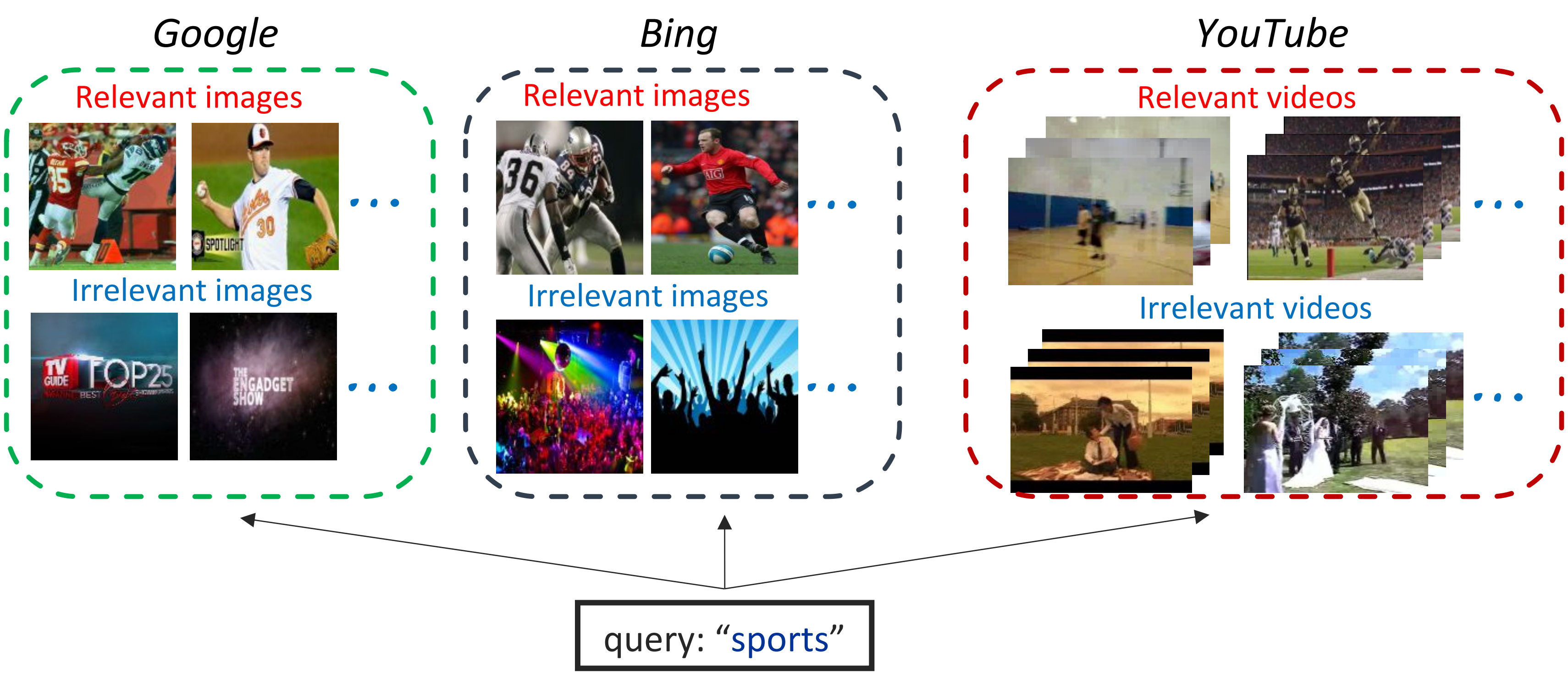
NANYANG TECHNOLOGICAL UNIVERSITY

Weakly Supervised Domain Generalization

presented by **Niu Li**
IGS ROSE
Nanyang Technological University (NTU), Singapore

Learning from Web data is increasingly popular but remains challenging.

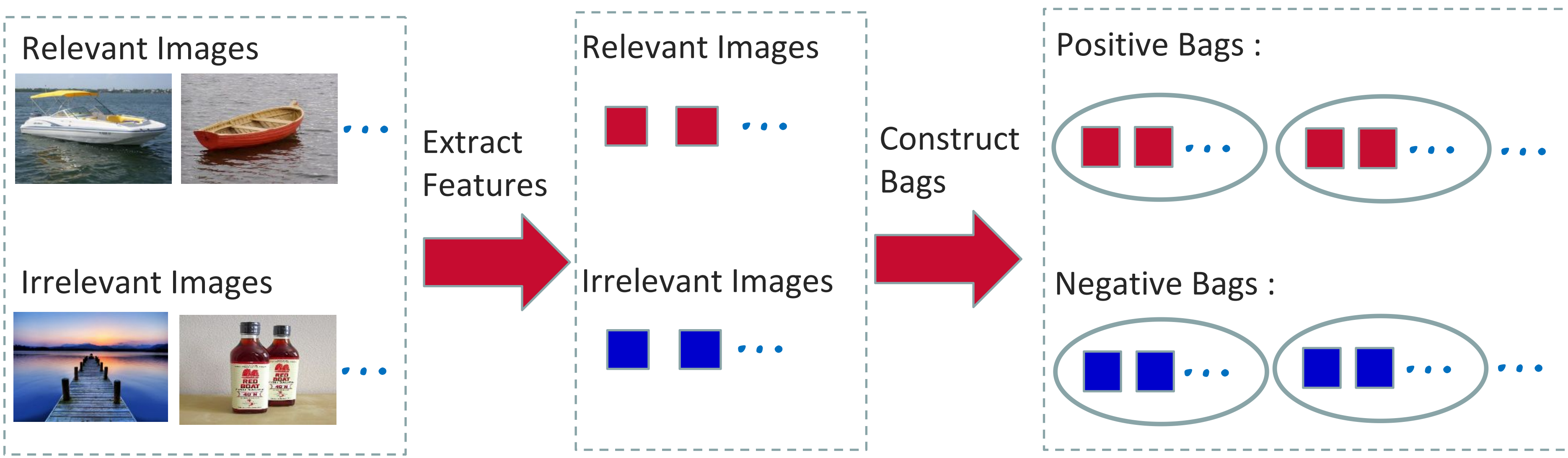
- Label Noise → Multi-instance Learning
- Unseen Target Domain → Domain Generalization



NANYANG TECHNOLOGICAL UNIVERSITY

Background: Multi-instance Learning

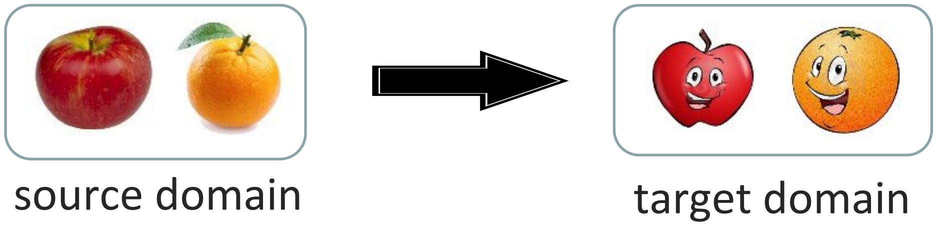
By treating each cluster as a "bag" and the images in each bag as "instances", multi-instance learning (MIL) methods are used for visual recognition.



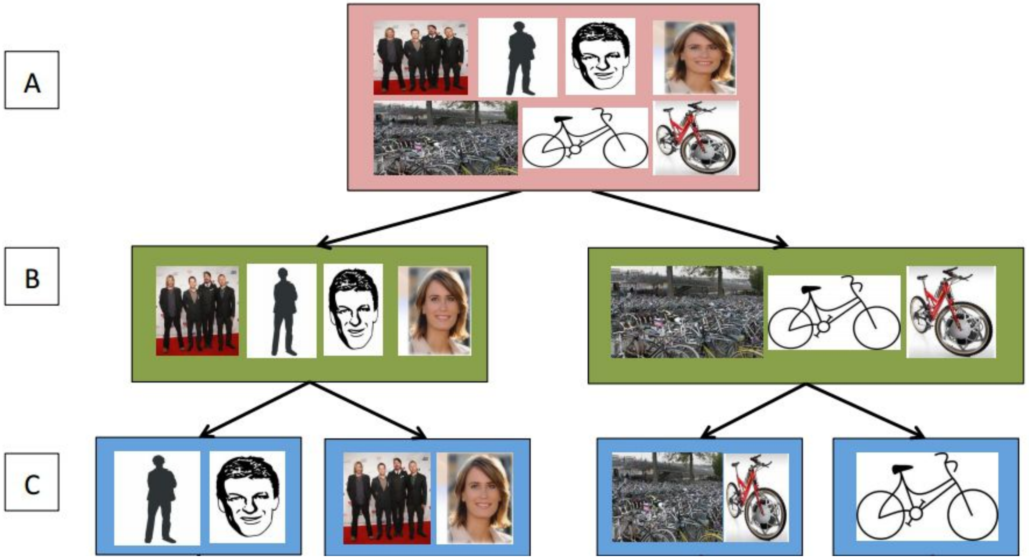
NANYANG TECHNOLOGICAL UNIVERSITY

Background: Domain Generalization

- Domain Adaptation



- Domain generalization is to generalize source domain to **unknown** target domain. Source domain may contain multiple latent domains characterized by different hidden factors (e.g., pose, illumination).



Formulation

Preliminary: multi-class multi-instance formulation

$$\min_{\substack{\mathbf{h}\in\mathcal{H}\\ \mathbf{w}_c,\xi_l}} \quad \frac{1}{2}\sum_{c=1}^{C}\|\mathbf{w}_c\|^2 + C_1\sum_{l=1}^{L}\xi_l$$

bag-level loss

$$\text{s.t.} \quad \frac{1}{|\mathcal{B}_l|}\sum_{i\in I_l} h_i((\mathbf{w}_{Y_l})'\phi(\mathbf{x}_i) - (\mathbf{w}_{\tilde{c}})'\phi(\mathbf{x}_i))$$

select indicator

$$\geq \eta - \xi_l, \quad \forall l, \tilde{c} \neq Y_l,$$

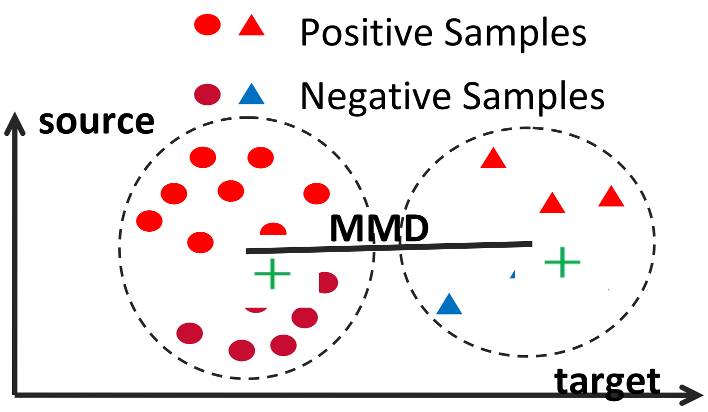
$$\xi_l \geq 0, \qquad \forall l,$$

NANYANG TECHNOLOGICAL UNIVERSITY

Formulation

Precompute probabilities: disover latent domains

$$\max_{\pi_{i,m}} \sum_{m \neq \tilde{m}} \| \frac{1}{N_m} \sum_{i=1}^{N} \pi_{i,m} \phi(\mathbf{x}_i) - \frac{1}{N_{\tilde{m}}} \sum_{i=1}^{N} \pi_{i,\tilde{m}} \phi(\mathbf{x}_i) \|^2$$



NANYANG TECHNOLOGICAL UNIVERSITY

Formulation

Final formulation: weakly supervised domain generalization

$$\min_{\substack{\mathbf{h}\in\mathcal{H}\\ \mathbf{w}_{c,m},\xi_l}} \quad \frac{1}{2}\sum_{c=1}^{C}\sum_{m=1}^{M}\|\mathbf{w}_{c,m}\|^2 + C_1\sum_{l=1}^{L}\xi_l$$

$$-C_2\,\rho(\mathbf{B}, \mathbf{K}\circ(\mathbf{h}\mathbf{h}'))$$

MMD-based regularizer

$$\text{s.t.} \quad \frac{1}{|\mathcal{B}_l|}\sum_{i\in I_l} h_i \left(\sum_{m=1}^{M}\hat{\beta}_{i,m}(\mathbf{w}_{Y_l,m})'\phi(\mathbf{x}_i) - (\mathbf{w}_{\tilde{c},\tilde{m}})'\phi(\mathbf{x}_i)\right)$$

$$\geq \eta - \xi_l, \qquad \forall l, \tilde{m}, \tilde{c} \neq Y_l,$$

$$\xi_l \geq 0, \qquad \forall l.$$

calculated from $\pi_{i,m}$

Experimental Results

1. Video Event Recognition training dataset: Flickr test dataset: Kodak, CCV
2. Image Classification training dataset: Bing test dataset: Caltech-256

	Method	Testing Dataset		
		Kodak	CCV	Caltech-256
basic baseline	SVM [9]	34.36	40.84	70.93
multi-instance baselines	sMIL [6]	38.46	41.34	71.33
	mi-SVM [2]	37.95	46.38	71.47
	MIL-CPB [30]	38.97	46.29	71.6
	KI-SVM [32]	40.00	42.85	71.20
domain generalization baselines	DICA [34]	42.05	44.10	70.80
	LRESVM [42]	41.94	48.12	72.93
	[23] (Match)	37.13	41.37	71.07
	[23] (Ensemble)	37.42	41.40	70.08
	[19] (Match)	40.93	44.44	71.47
	[19](Ensemble)	42.39	47.51	72.40
subcategory baselines	Sub-Cate [22]	38.59	47.93	72.27
	MMDL [41]	40.51	48.87	72.80
our special cases	WSDG_sim1	42.56	47.47	71.87
	WSDG_sim2	43.59	49.93	74.00
Ours	WSDG	**45.64**	**51.18**	**75.20**

Thanks for your attention!

NANYANG
TECHNOLOGICAL
UNIVERSITY