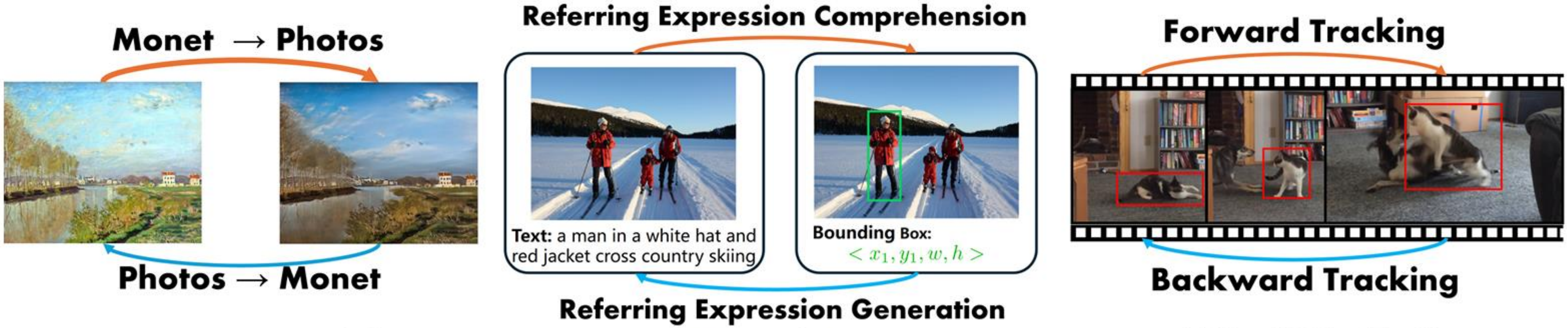
Variational Inference for Cyclic Learning

Zhuojun Zou and Jie Hao
Institute of Automation, Chinese Academy of Sciences,
Beijing, China
✉zouzhuojun2018@ia.ac.cn

Motivation

Despite its power to eliminate paired annotations via cycle-consistency, cyclic learning remains fragmented across domains with task-specific heuristics and no unified theoretical foundation.



(a) Image-to-Image Translation (b) REC and REG (c) Visual Object Tracking

Variational Inference

Goal: Learn bidirectional mappings $\mathbf{y} = f_\phi(\mathbf{x}), \quad \mathbf{x} = g_\theta(\mathbf{y})$

Key Insight: Treat intermediate points as latent variables with Dirac conditional distributions: $p_\theta(\mathbf{x}|\mathbf{y}) = \delta(\mathbf{x} - g_\theta(\mathbf{y})); \quad p_\phi(\mathbf{y}|\mathbf{x}) = \delta(\mathbf{y} - f_\phi(\mathbf{x}))$

Maximize the Log-likelihood:

$$\log p_\theta(\mathbf{x}) = \int q_\phi(\mathbf{y}|\mathbf{x}) \log \frac{p_\theta(\mathbf{x},\mathbf{y})}{p_\theta(\mathbf{y}|\mathbf{x})} d\mathbf{y} = \ell_\theta(\mathbf{x}) + D_{KL}(q_\phi(\mathbf{y}|\mathbf{x}) || p_\theta(\mathbf{y}|\mathbf{x})),$$

ELBO for cyclic learning:

$$\ell_{\theta,\phi}(\mathbf{x},\mathbf{y}) = \underbrace{\int q_\phi(\mathbf{y}|\mathbf{x}) \log p_\theta(\mathbf{x}|\mathbf{y}) d\mathbf{y}}_{\text{Reconstruction}} - \underbrace{D_{KL}(q_\phi(\mathbf{y}|\mathbf{x}) \,||\, p(\mathbf{y}))}_{\text{Prior matching}} + \underbrace{\int q_\theta(\mathbf{x}|\mathbf{y}) \log p_\phi(\mathbf{y}|\mathbf{x}) d\mathbf{x}}_{\text{Reconstruction}} - \underbrace{D_{KL}(q_\theta(\mathbf{x}|\mathbf{y}) \,||\, p(\mathbf{x}))}_{\text{Prior matching}}$$

Approximated Loss:

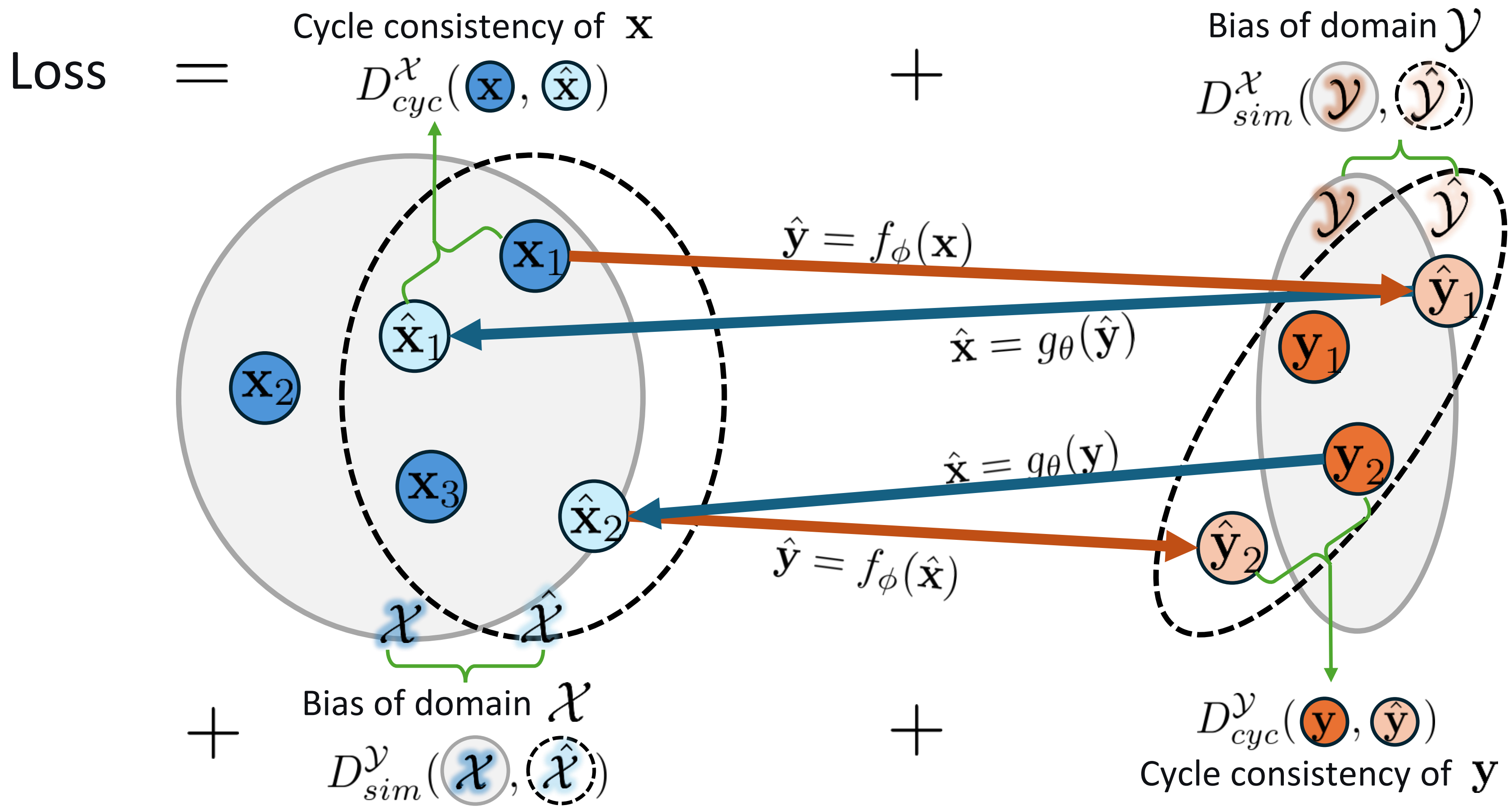
$$\arg\max_{\theta,\phi} \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x})} \left[\log p_\theta(\mathbf{x}|\mathbf{y})\right] = \arg\min_{\theta,\phi} D_{cyc}\left(\mathbf{x}, g_\theta\left(f_\phi(\mathbf{x})\right)\right).$$

$$\arg\max_{\phi} \; -D_{KL}(q_\phi(\mathbf{y}|\mathbf{x}) || p(\mathbf{y})) = \arg\min_{\phi} D_{sim}(f_\phi(\mathbf{x}), \mathcal{Y}).$$

Single-step Joint Optimization

End-to-end optimization of both mappings simultaneously using a unified loss:

$$\mathcal{L}(\mathbf{x},\mathbf{y}) = D^{\mathcal{X}}_{cyc}(\mathbf{x}, g_\theta(f_\phi(\mathbf{x}))) + D^{\mathcal{X}}_{sim}(f_\phi(\mathbf{x}), \mathcal{Y}) + D^{\mathcal{Y}}_{cyc}(\mathbf{y}, f_\phi(g_\theta(\mathbf{y}))) + D^{\mathcal{Y}}_{sim}(g_\theta(\mathbf{y}), \mathcal{X}),$$



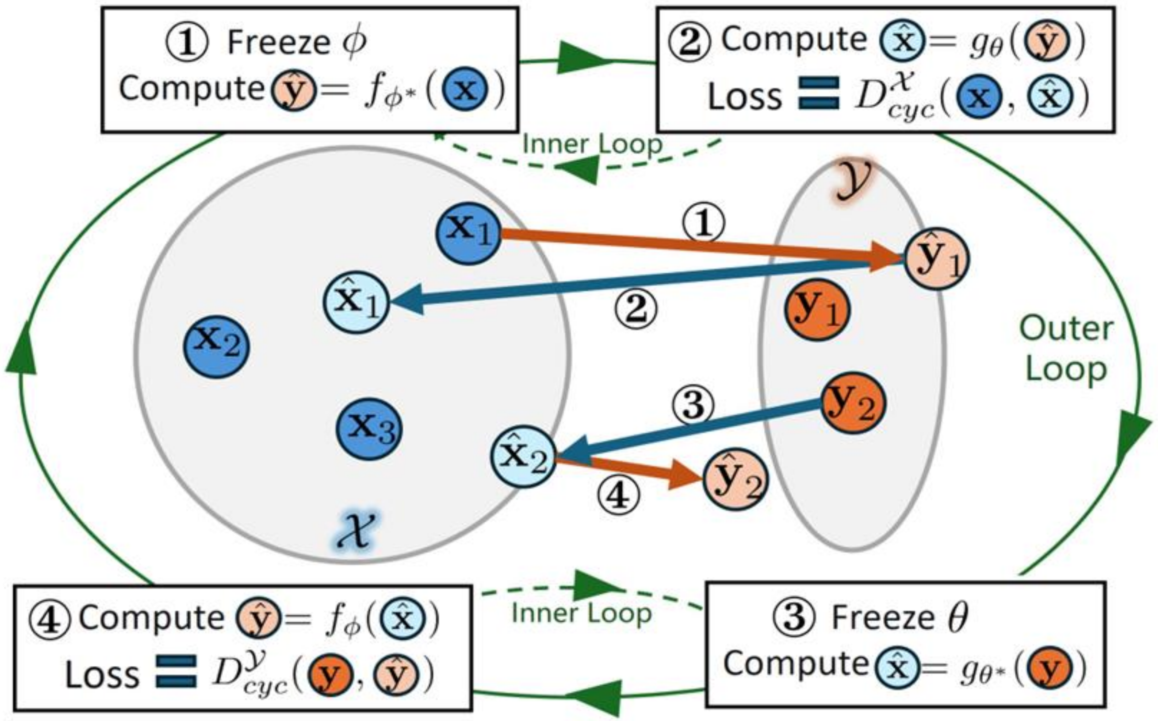
Dual EM-iteration Optimization

We propose an EM algorithm that **alternately infers pseudo-labels from one mapping and updates the other via cycle loss**, ensuring progressive cycle-consistency alignment.

Algorithm 1: An EM approach for cyclic tasks

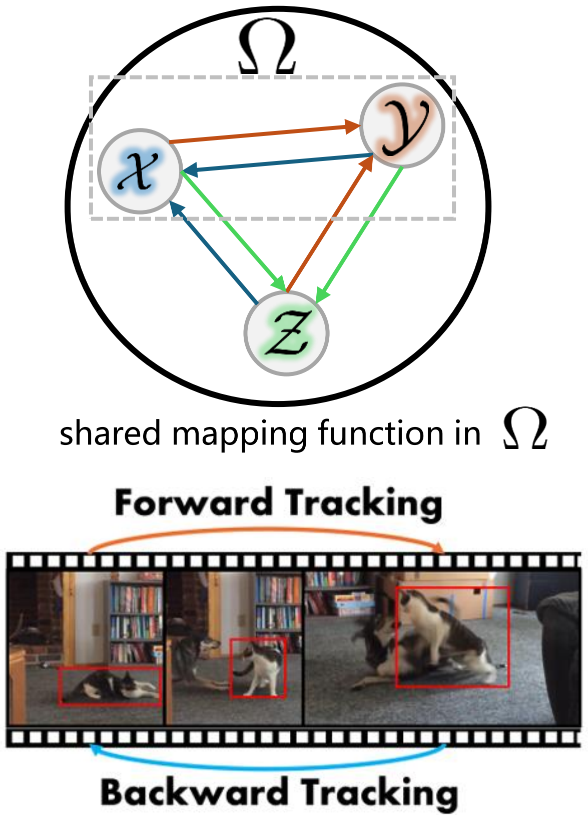
---

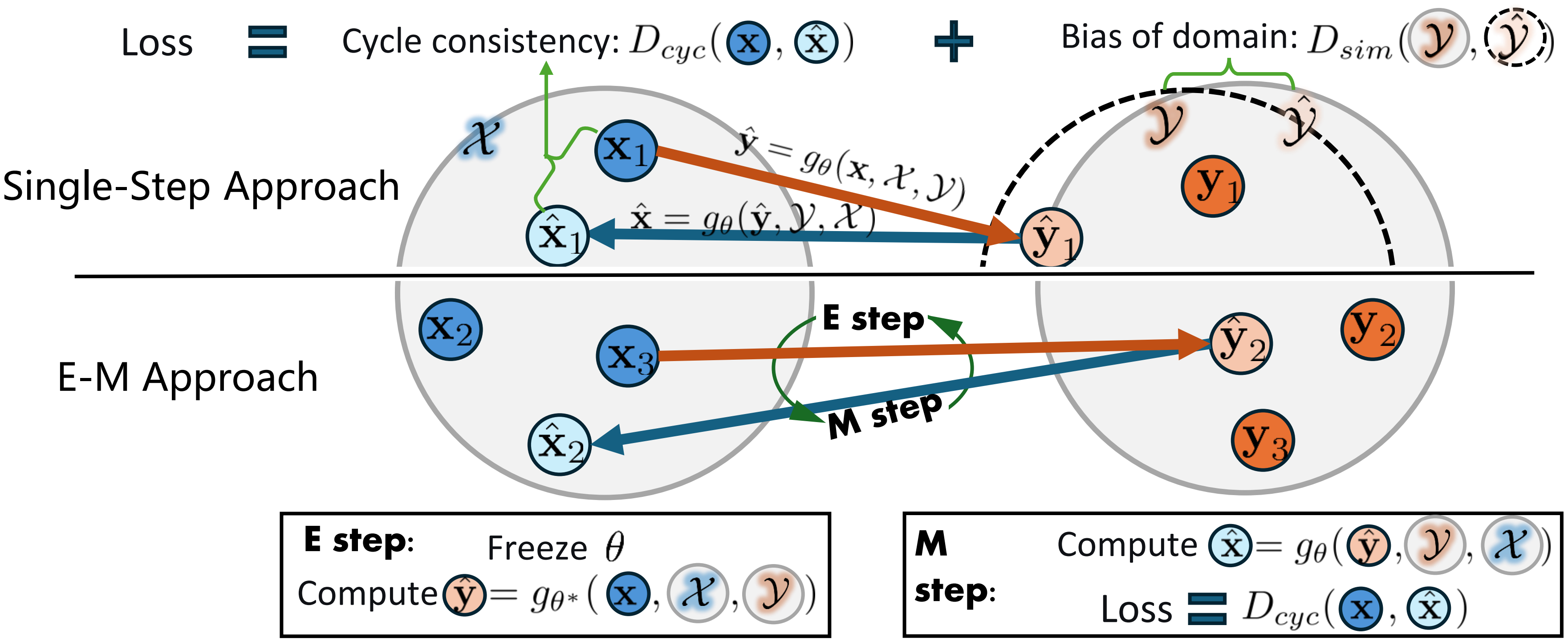
while *not converge* **do**
    while *insufficient loss decrease* **do**
        Get pseudo-labels $\hat{\mathbf{y}} = f_\phi(\mathbf{x})$ for $\mathbf{x}$;
        Update θ via $\mathcal{L}(\theta) = D^{\mathcal{X}}_{cyc}(\mathbf{x}, g_\theta(\hat{\mathbf{y}}))$;
    while *insufficient loss decrease* **do**
        Get pseudo-labels $\hat{\mathbf{x}} = g_\theta(\mathbf{y})$ for $\mathbf{y}$;
        Update ϕ via $\mathcal{L}(\phi) = D^{\mathcal{Y}}_{cyc}(\mathbf{y}, f_\phi(\hat{\mathbf{x}}))$;



Extending to Self-Cyclic Learning

For self-cyclic tasks where a single model handles both directions (e.g., forward-backward tracking), our framework offers two simplified optimizations: using only **half the loss terms** (single-step) or **a single EM iteration**.

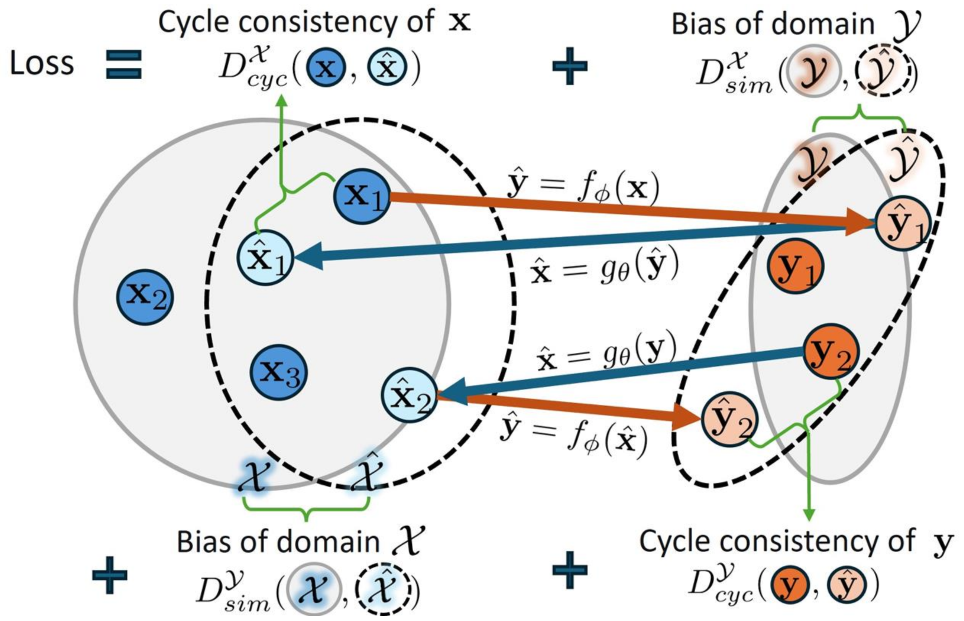




Application: Unpaired Image Translation

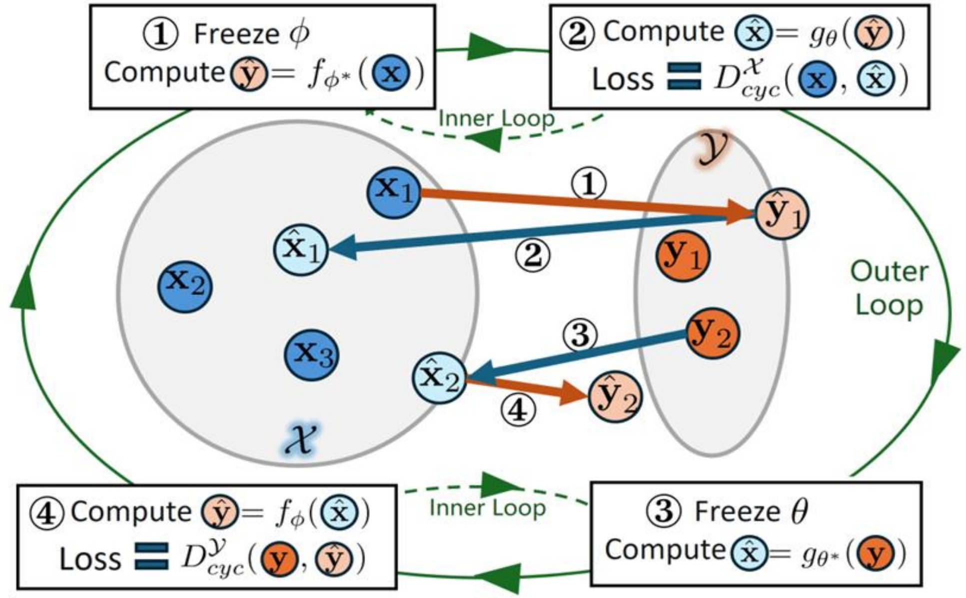
CycleGAN: The corresponding components in CycleGAN can be mapped onto those in our framework as:

$$D^{\mathcal{X}}_{cyc}(\mathbf{x}, g_\theta(f_\phi(\mathbf{x}))) \Rightarrow \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\|g_\theta(f_\phi(\mathbf{x})) - \mathbf{x}\|_1]$$
$$D^{\mathcal{X}}_{sim}(f_\phi(\mathbf{x}), \mathcal{Y}) \Rightarrow \mathcal{L}_{\text{GAN}}(f_\phi, D_{\mathcal{Y}}, \mathcal{X}, \mathcal{Y})$$
$$D^{\mathcal{Y}}_{cyc}(\mathbf{y}, f_\phi(g_\theta(\mathbf{y}))) \Rightarrow \mathbb{E}_{\mathbf{y}\sim p_{data}(\mathbf{y})}[\|f_\phi(g_\theta(\mathbf{y})) - \mathbf{y}\|_1]$$
$$D^{\mathcal{Y}}_{sim}(g_\theta(\mathbf{y}), \mathcal{X}) \Rightarrow \mathcal{L}_{\text{GAN}}(g_\theta, D_{\mathcal{X}}, \mathcal{Y}, \mathcal{X})$$



CycleGN: The proposed CycleGN operates without GANs and adopts EM iteration. Its correspondence to our algorithm is:

$$① \Rightarrow \hat{\mathbf{y}} = f_\phi(\mathbf{x});$$
$$② \Rightarrow \text{Update } \theta \text{ via } \mathcal{L}_{\text{cyc}}(g_\theta) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\|g_\theta(\hat{\mathbf{y}}) - \mathbf{x}\|_1]$$
$$③ \Rightarrow \hat{\mathbf{x}} = g_\theta(\mathbf{y});$$
$$④ \Rightarrow \text{Update } \phi \text{ via } \mathcal{L}_{\text{cyc}}(f_\phi) = \mathbb{E}_{\mathbf{y}\sim p_{data}(\mathbf{y})}[\|f_\phi(\hat{\mathbf{x}}) - \mathbf{y}\|_1]$$



Application: Unpaired Image Translation

CycleGAN: The corresponding components in CycleGAN can be mapped onto those in our framework as:

$$D^{\mathcal{X}}_{cyc}(\mathbf{x}, g_\theta(f_\phi(\mathbf{x}))) \Rightarrow \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\|g_\theta(f_\phi(\mathbf{x})) - \mathbf{x}\|_1]$$
$$D^{\mathcal{X}}_{sim}(f_\phi(\mathbf{x}), \mathcal{Y}) \Rightarrow \mathcal{L}_{\text{GAN}}(f_\phi, D_\mathcal{Y}, \mathcal{X}, \mathcal{Y})$$
$$D^{\mathcal{Y}}_{cyc}(\mathbf{y}, f_\phi(g_\theta(\mathbf{y}))) \Rightarrow \mathbb{E}_{\mathbf{y}\sim p_{data}(\mathbf{y})}[\|f_\phi(g_\theta(\mathbf{y})) - \mathbf{y}\|_1]$$
$$D^{\mathcal{Y}}_{sim}(g_\theta(\mathbf{y}), \mathcal{X}) \Rightarrow \mathcal{L}_{\text{GAN}}(g_\theta, D_\mathcal{X}, \mathcal{Y}, \mathcal{X})$$

CycleGN: The proposed CycleGN operates without GANs and adopts EM iteration. Its correspondence to our algorithm is:

$$① \Rightarrow \hat{\mathbf{y}} = f_\phi(\mathbf{x});$$
$$② \Rightarrow \text{Update } \theta \text{ via } \mathcal{L}_{\text{cyc}}(g_\theta) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\|g_\theta(\hat{\mathbf{y}}) - \mathbf{x}\|_1]$$
$$③ \Rightarrow \hat{\mathbf{x}} = g_\theta(\mathbf{y});$$
$$④ \Rightarrow \text{Update } \phi \text{ via } \mathcal{L}_{\text{cyc}}(f_\phi) = \mathbb{E}_{\mathbf{y}\sim p_{data}(\mathbf{y})}[\|f_\phi(\hat{\mathbf{x}}) - \mathbf{y}\|_1]$$

Loss	GAN	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN	✓	0.40	0.10	0.06
BiGAN/ALI	✓	0.19	0.06	0.02
SimGAN	✓	0.20	0.10	0.04
Feat. loss + GAN	✓	0.06	0.04	0.01
CycleGAN	✓	**0.52**	**0.17**	**0.11**
CycleGN (ours)	×	**0.52**	0.14	0.10

Labels to Photo Results

Loss	GAN	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN	✓	0.45	0.11	0.08
BiGAN/ALI	✓	0.41	0.13	0.07
SimGAN	✓	0.47	0.11	0.07
Feat. loss + GAN	✓	0.50	0.10	0.06
CycleGAN	✓	**0.58**	**0.22**	**0.16**
CycleGN (ours)	×	0.51	0.16	0.10

Photo to Labels Results

Application: Unsupervised Object Tracking

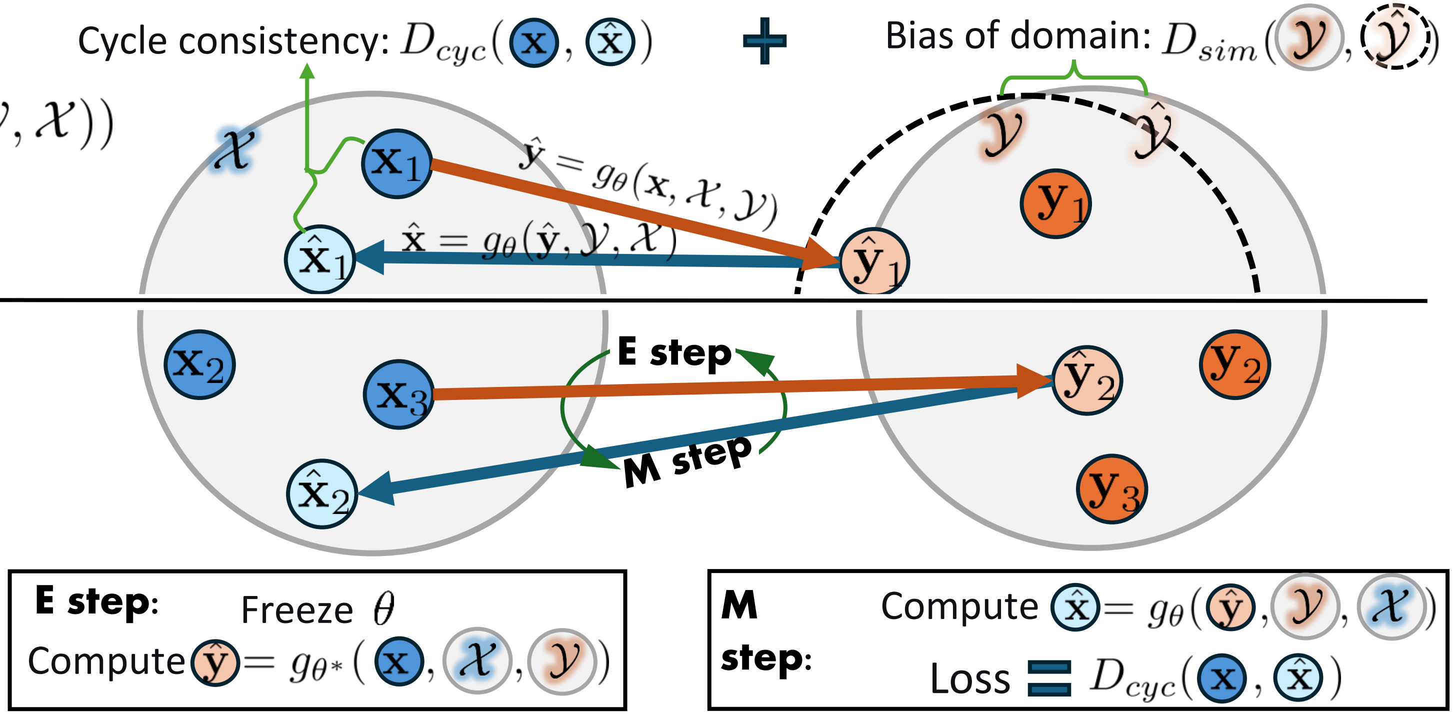
CycleTrack: The proposed Loss function:

$$\mathcal{L}(T,\mathbf{x}) = \mathcal{L}_b(\mathbf{x}, T(T(\mathbf{x},\mathcal{X},\mathcal{Y}),\mathcal{Y},\mathcal{X})) + \mathcal{L}_b(T(\mathbf{x},\mathcal{X},\mathcal{Y}),\tilde{\mathbf{y}}),$$
$$s.t. \quad \tilde{\mathbf{y}} = \arg\max_{\mathbf{y}\in \mathrm{BOX}_{\mathcal{Y}}} \mathrm{IoU}(\mathbf{y}, T(\mathbf{x},\mathcal{X},\mathcal{Y})),$$

Single-Step Approach

The corresponding components:

$$D_{cyc}(\mathbf{x}, g_\theta(g_\theta(\mathbf{x}))) \Rightarrow \mathcal{L}_b(\mathbf{x}, T(T(\mathbf{x},\mathcal{X},\mathcal{Y}),\mathcal{Y},\mathcal{X}))$$
$$D_{sim}(g_\theta(\mathbf{x}),\mathcal{Y}) \Rightarrow \mathcal{L}_b(T(\mathbf{x},\mathcal{X},\mathcal{Y}),\tilde{\mathbf{y}})$$



CycleTrack-EM: The EM counterpart of CycleTrack is designed as:

$$\mathbf{E\ step} \Rightarrow \hat{\mathbf{y}} = g_\theta(\mathbf{x});$$
$$\mathbf{M\text{-}step} \Rightarrow \text{Update } \theta \text{ via } \mathcal{L}_b(\mathbf{x}, T(\hat{\mathbf{y}},\mathcal{Y},\mathcal{X}))$$

E-M Approach

Application: Unsupervised Object Tracking

CycleTrack: The proposed Loss function:

$$\mathcal{L}(T, \mathbf{x}) = \mathcal{L}_b(\mathbf{x}, T(T(\mathbf{x}, \mathcal{X}, \mathcal{Y}), \mathcal{Y}, \mathcal{X})) + \mathcal{L}_b(T(\mathbf{x}, \mathcal{X}, \mathcal{Y}), \tilde{\mathbf{y}}),$$
$$s.t. \quad \tilde{\mathbf{y}} = \arg\max_{\mathbf{y} \in \text{BOX}_{\mathcal{Y}}} \text{IoU}(\mathbf{y}, T(\mathbf{x}, \mathcal{X}, \mathcal{Y})),$$

The corresponding components:

$$D_{cyc}(\mathbf{x}, g_\theta(g_\theta(\mathbf{x}))) \Rightarrow \mathcal{L}_b(\mathbf{x}, T(T(\mathbf{x}, \mathcal{X}, \mathcal{Y}), \mathcal{Y}, \mathcal{X}))$$
$$D_{sim}(g_\theta(\mathbf{x}), \mathcal{Y}) \Rightarrow \mathcal{L}_b(T(\mathbf{x}, \mathcal{X}, \mathcal{Y}), \tilde{\mathbf{y}})$$

CycleTrack-EM: The EM counterpart of CycleTrack is designed as:

$$\textbf{E step} \Rightarrow \hat{\mathbf{y}} = g_\theta(\mathbf{x});$$
$$\textbf{M-step} \Rightarrow \text{Update } \theta \text{ via } \mathcal{L}_b(\mathbf{x}, T(\hat{\mathbf{y}}, \mathcal{Y}, \mathcal{X}))$$

Method	LaSOT		TrackingNet	
	AUC	Precision	AUC	Precision
ResPUL	-	-	54.6	48.5
LUDT+	30.5	28.8	56.3	49.5
USOT*	35.8	34.0	61.5	56.6
ULAST*-off	46.8	44.8	64.9	58.5
ULAST*-on	47.1	45.1	65.4	59.2
CycleTrack	51.0	49.7	75.9	71.5
CycleTrack-EM	**56.5**	**57.9**	**77.3**	**74.4**

Unsupervised Tracking Results

Method	LaSOT		TrackingNet	
	AUC	Precision	AUC	Precision
ResPUL	-	-	54.6	48.5
LUDT	26.2	23.4	54.3	46.9
USOT	33.7	32.3	59.9	55.1
ULAST-off	42.9	40.5	-	-
ULAST-on	43.3	40.7	-	-
CycleTrack	45.0	42.2	65.6	59.0
CycleTrack-EM	**51.2**	**49.9**	**69.1**	**64.7**

Strictly-Unsupervised Tracking Results

Thank you.

Speaker: Zhuojun Zou
Institute of Automation, Chinese Academy of Sciences,
Beijing, China
zouzhuojun2018@ia.ac.cn