Learning Compositional Visual Concepts with Mutual Consistency

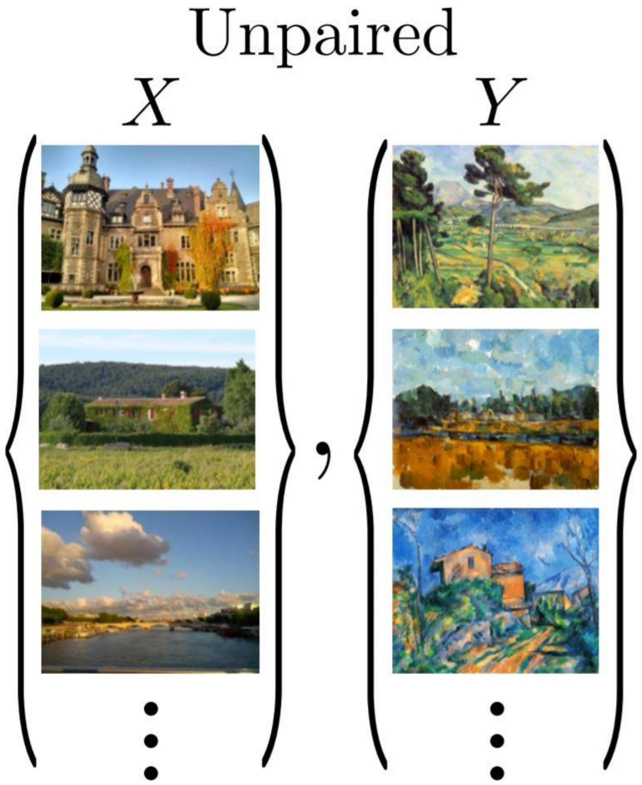
Yunye Gong, Srikrishna Karanam, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, and Peter C. Doerschuk
CVPR 2018 (spotlight)

Concept GAN

Presented by: Anand Bhattad (Vision Lunch, April 24 2018)

Image Translation (Pix to Pix and Cycle GAN)

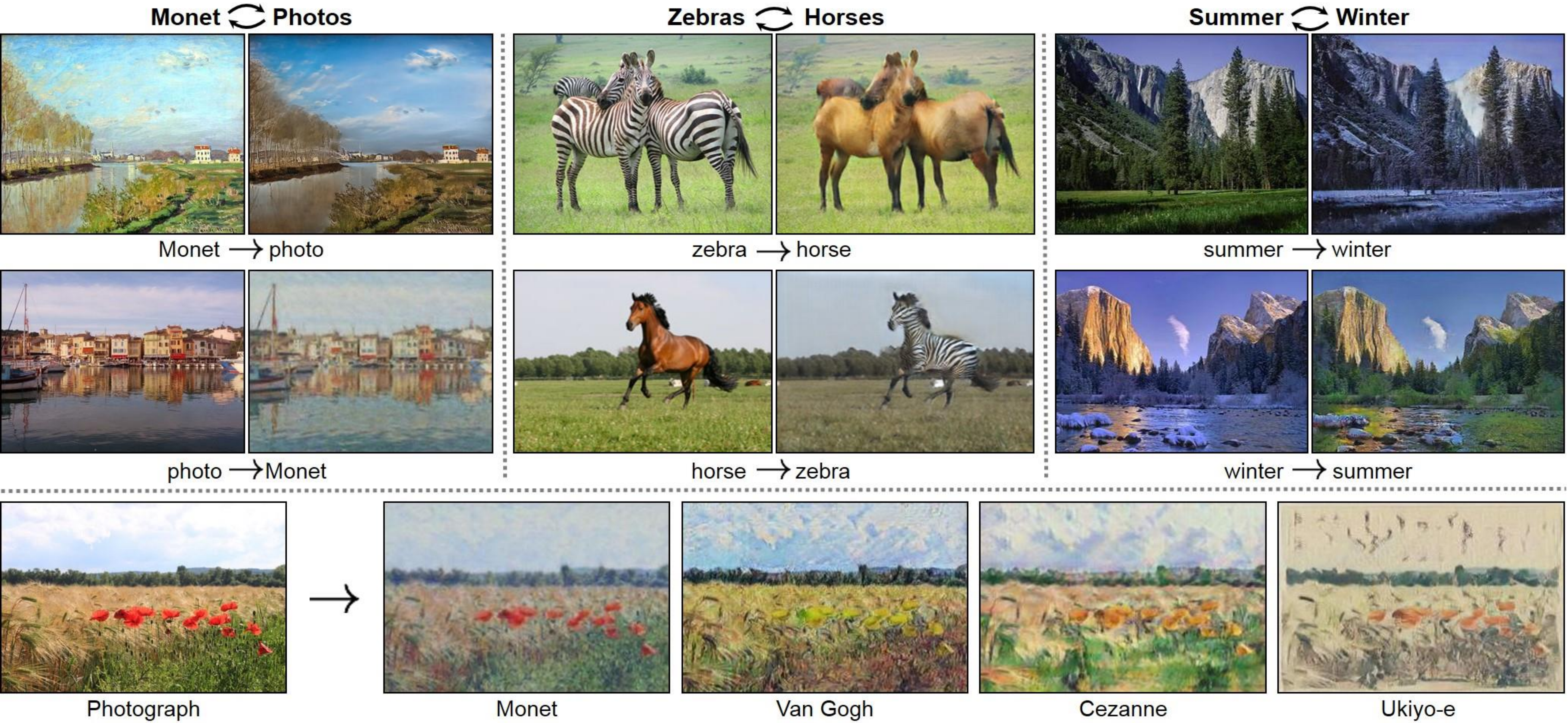
Pix2Pix



Cycle GAN

[Source: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial]()

Cycle GAN



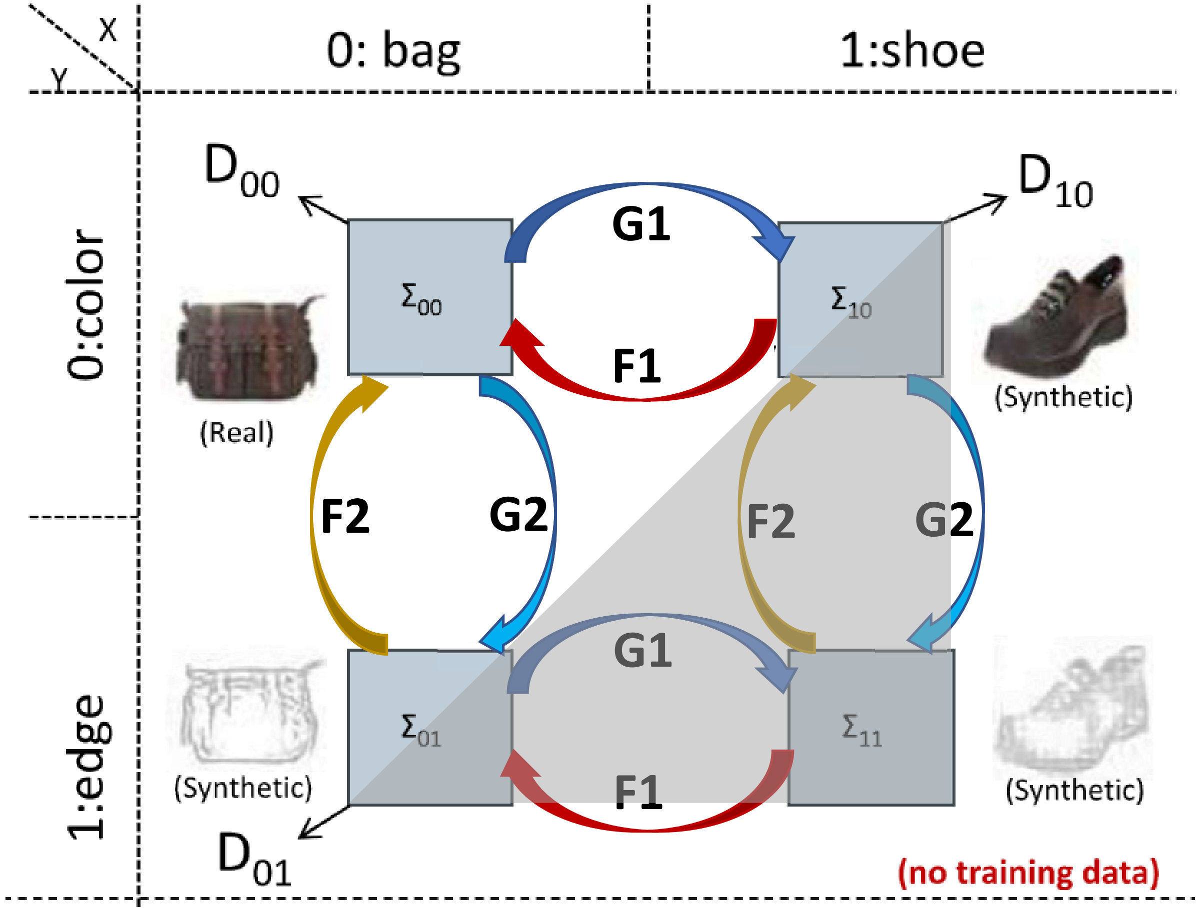
[Source: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial]

New Unseen Samples from Translations?

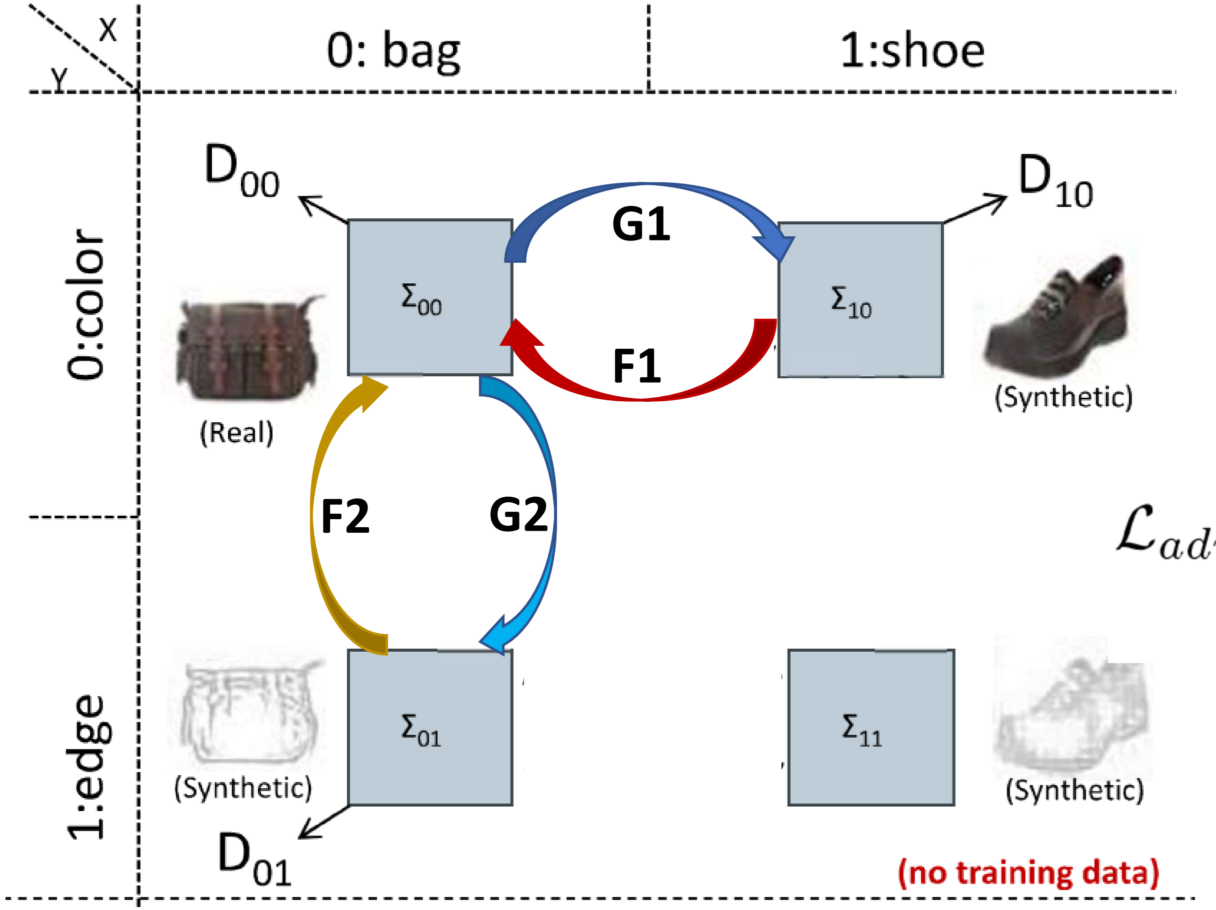
- Colored Bags
- Edge Bags
- Colored Shoes
- Edge Shoes

Overview: Cycle GAN

Cycle GAN: Learn and Compose concepts **separately**



Losses: Adversarial



$$\boxed{\mathcal{L}_{adv}(G_1, D_{10}, \Sigma_{00}, \Sigma_{10})}$$

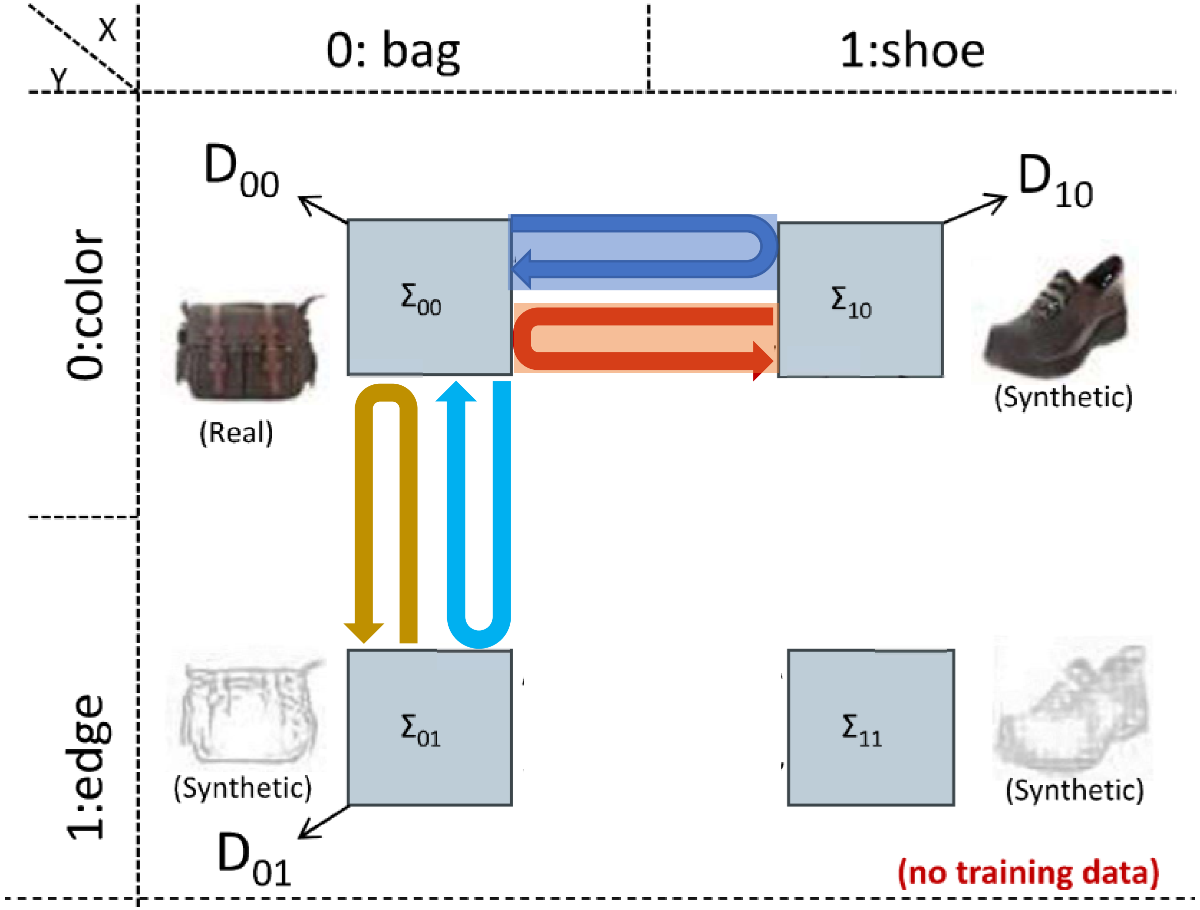
$$\boxed{\mathcal{L}_{adv}(F_1, D_{00}, \Sigma_{10}, \Sigma_{00})}$$

$$\boxed{\mathcal{L}_{adv}(G_2, D_{01}, \Sigma_{00}, \Sigma_{01})}$$

$$\boxed{\mathcal{L}_{adv}(F_2, D_{00}, \Sigma_{01}, \Sigma_{00})}$$

$$\mathcal{L}_{adv}(G_1, D_{10}, \Sigma_{00}, \Sigma_{10}) = \mathbb{E}_{\sigma_{10} \sim P_{10}}[\log D_{10}(\sigma_{10})] + \mathbb{E}_{\sigma_{00} \sim P_{00}}[\log(1 - D_{10}(G_1(\sigma_{00})))]$$

Losses: Cycle Consistency

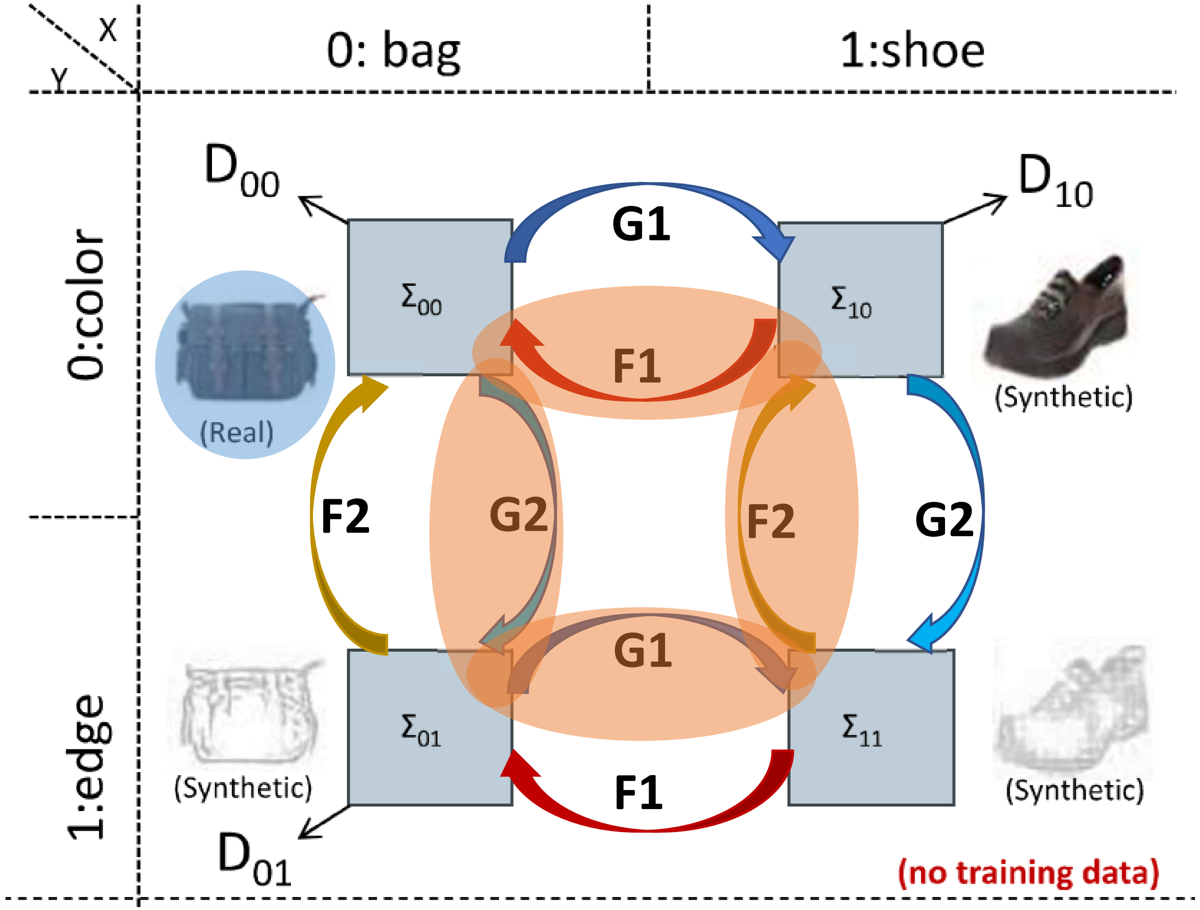


$$\mathcal{L}_{\text{cyc2}}(G1, F1)$$

$$\mathcal{L}_{\text{cyc2}}(G2, F2)$$

$$\mathcal{L}_{\text{cyc2}}(G1, F1) = \mathbb{E}_{\sigma_{00} \sim P_{00}}[||(F_1 \circ G_1)\sigma_{00} - \sigma_{00}||_1] + \mathbb{E}_{\sigma_{10} \sim P_{10}}[||(G_1 \circ F_1)\sigma_{10} - \sigma_{10}||_1]$$

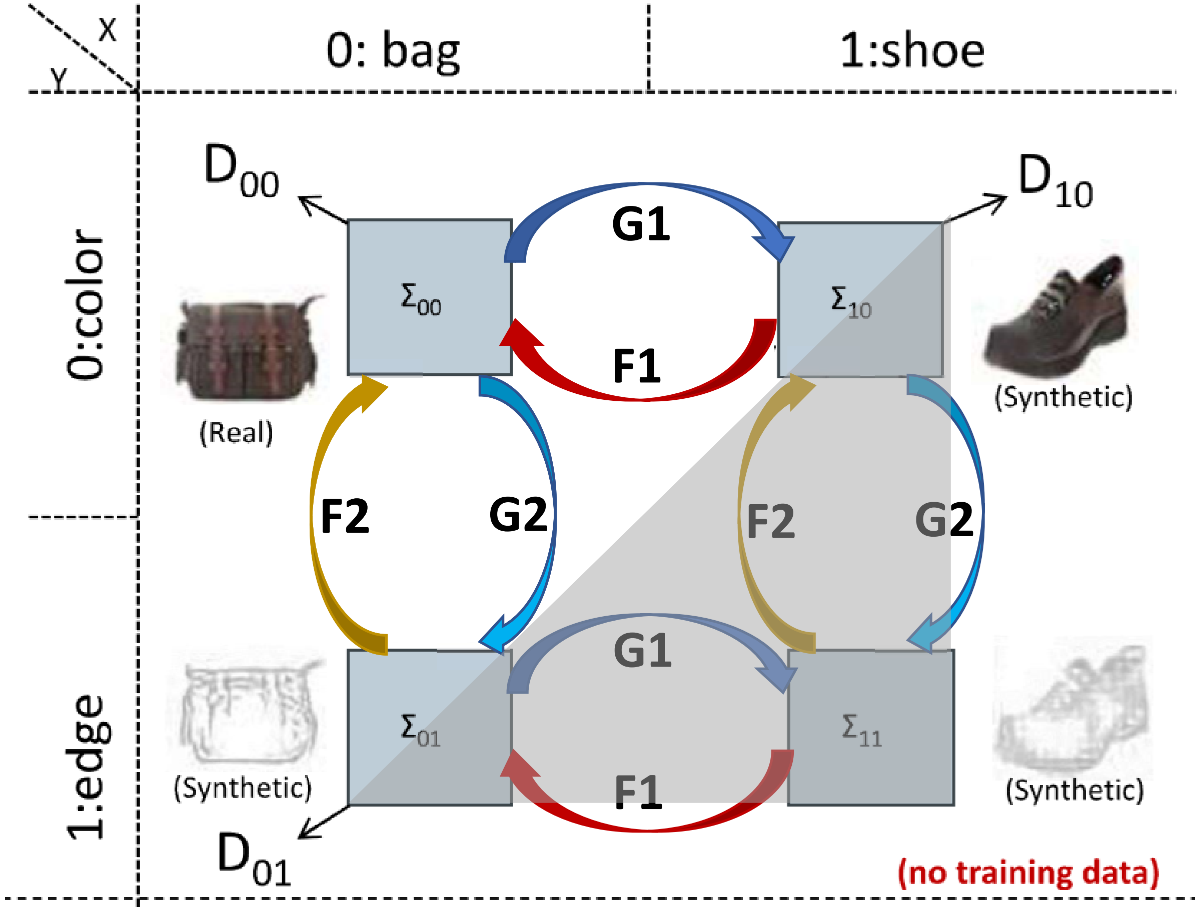
Can we generate sketch shoes by combining them?



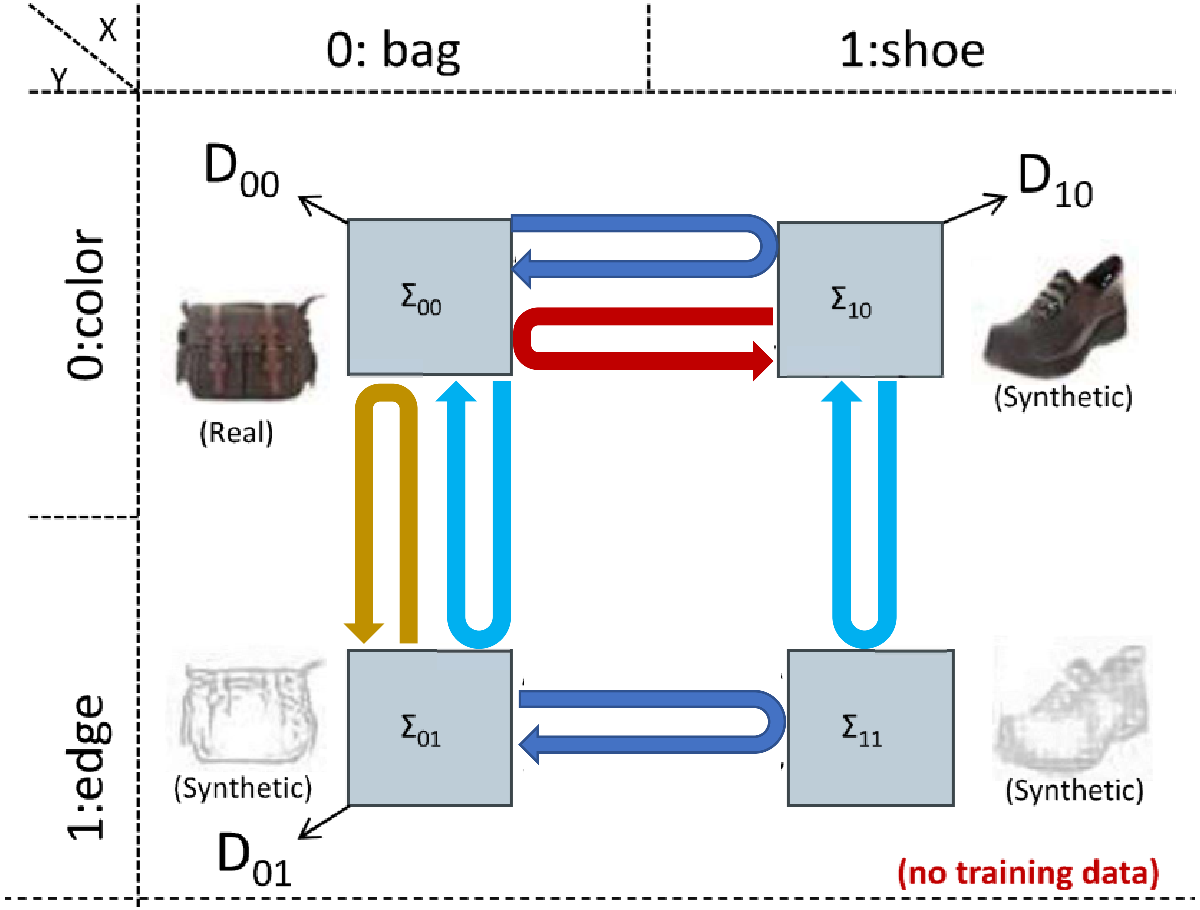
Proposed Model: Concept GAN

Cycle GAN: Learn and Compose concepts **separately**

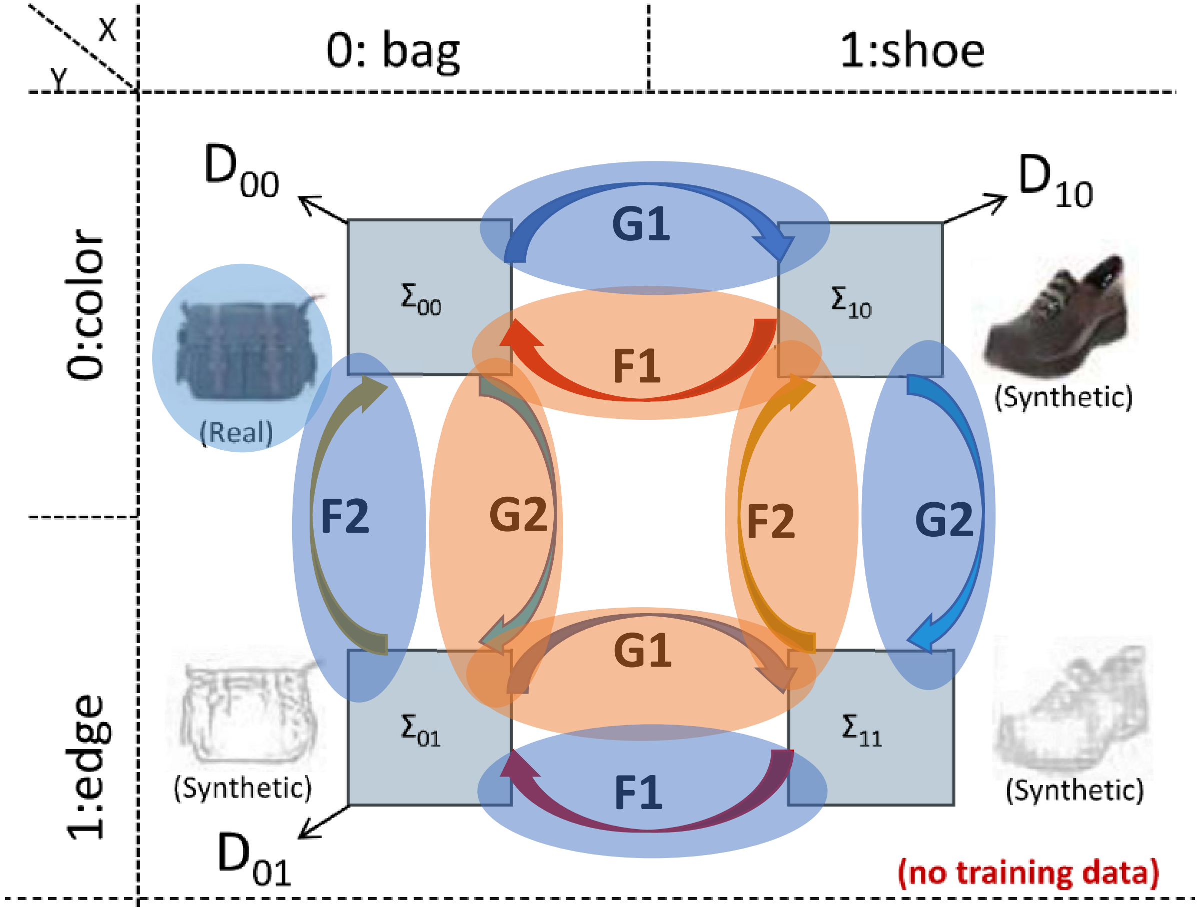
Concept GAN: Learn and Compose concepts **simultaneously**



Losses: Cycle Consistency



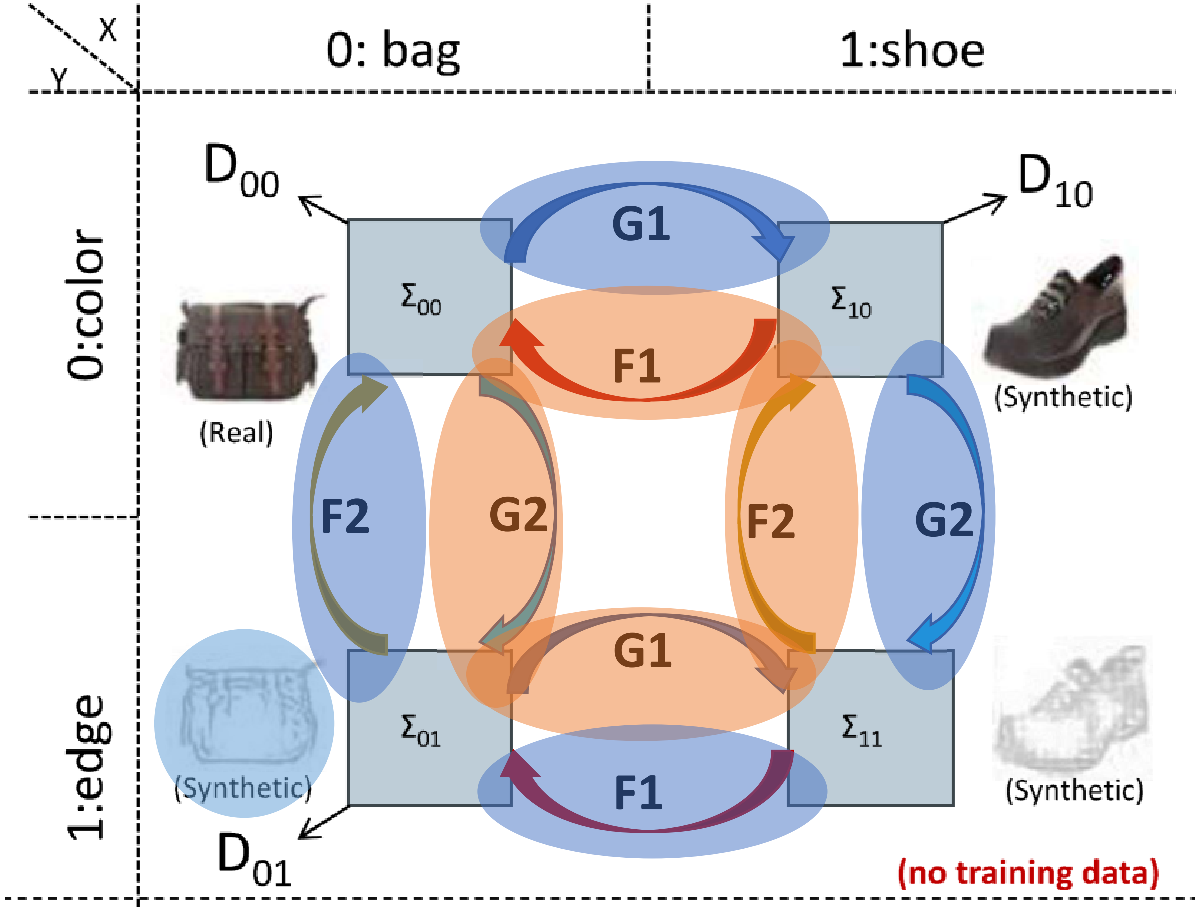
Losses: Cycle Consistency (Dist=4)



$$\boxed{\mathcal{L}_{cyc4}(G, F, \Sigma_{00})}$$

$$\begin{aligned}\mathcal{L}_{cyc4}(G, F, \Sigma_{00}) &= \mathbb{E}_{\sigma_{00}\sim P_{00}}[\|(F_2 \circ F_1 \circ G_2 \circ G_1)(\sigma_{00}) - \sigma_{00}\|_1] \\ &+ \mathbb{E}_{\sigma_{00}\sim P_{00}}[\|(F_1 \circ F_2 \circ G_1 \circ G_2)(\sigma_{00}) - \sigma_{00}\|_1].\end{aligned}$$

Losses: Cycle Consistency (Dist=4)

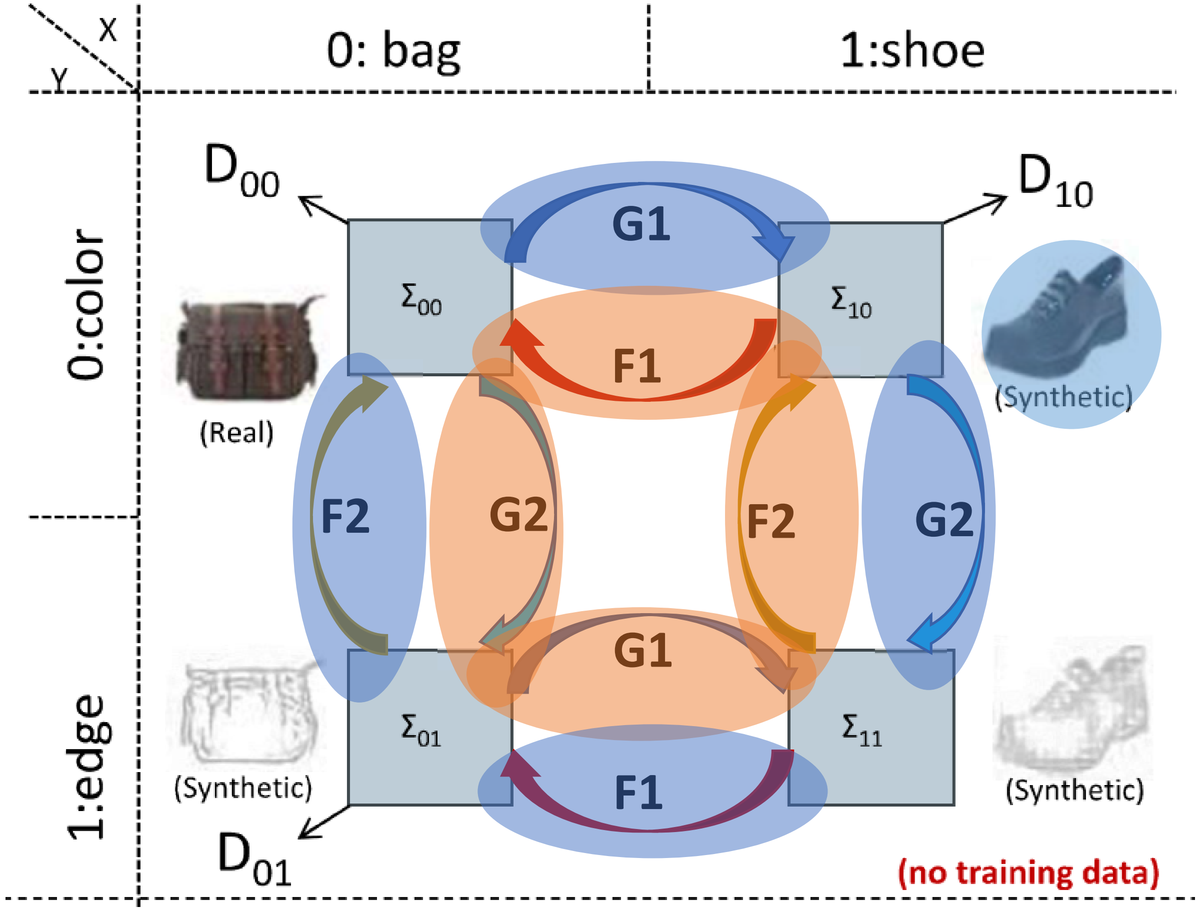


$$\mathcal{L}_{cyc4}(G, F, \Sigma_{00})$$

$$\boxed{\mathcal{L}_{cyc4}(G, F, \Sigma_{01})}$$

$$\begin{aligned}\mathcal{L}_{cyc4}(G, F, \Sigma_{00})& \\ &= \mathbb{E}_{\sigma_{00}\sim P_{00}}[\|(F_2 \circ F_1 \circ G_2 \circ G_1)(\sigma_{00}) - \sigma_{00}\|_1] \\ &+ \mathbb{E}_{\sigma_{00}\sim P_{00}}[\|(F_1 \circ F_2 \circ G_1 \circ G_2)(\sigma_{00}) - \sigma_{00}\|_1].\end{aligned}$$

Losses: Cycle Consistency (Dist=4)



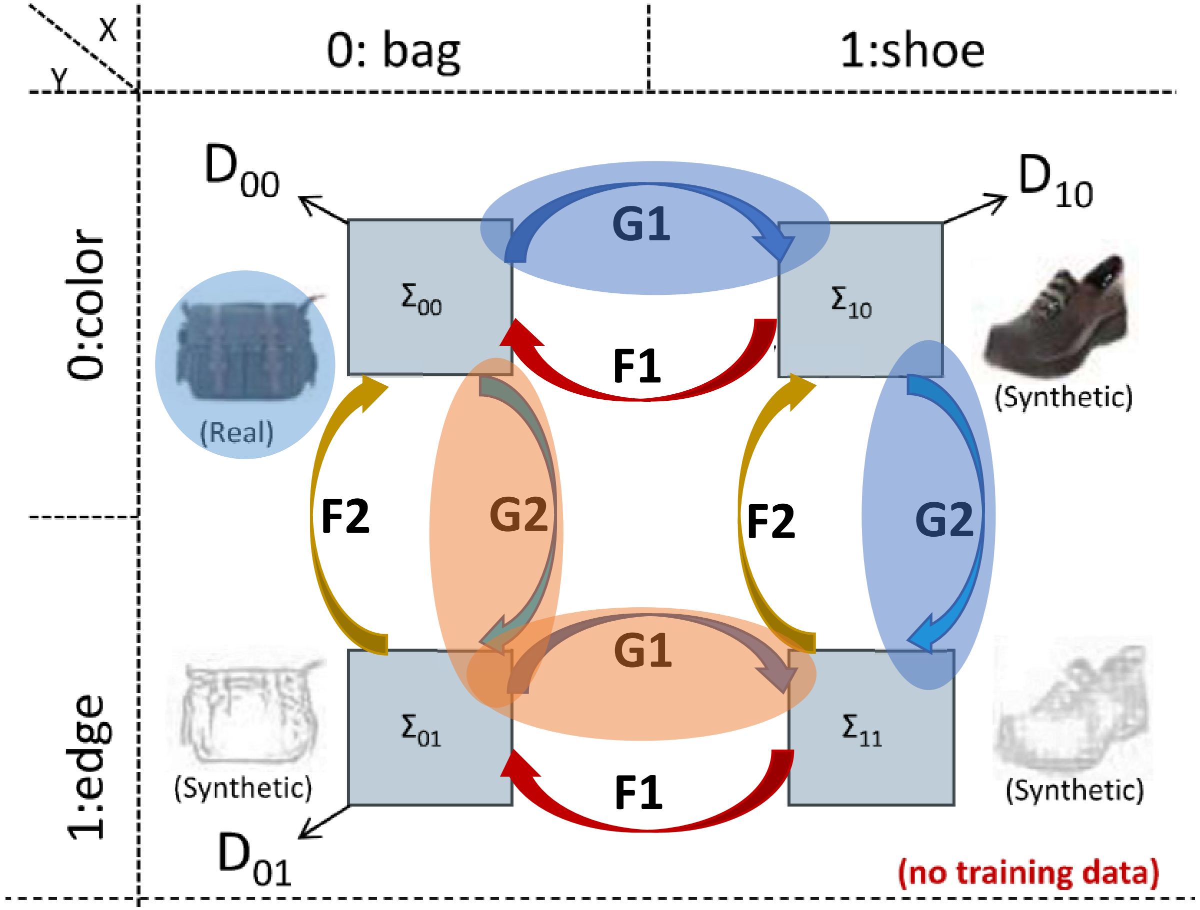
$\mathcal{L}_{cyc4}(G, F, \Sigma_{00})$

$\mathcal{L}_{cyc4}(G, F, \Sigma_{01})$

$\boxed{\mathcal{L}_{cyc4}(G, F, \Sigma_{10})}$

$$\begin{aligned}\mathcal{L}_{cyc4}&(G, F, \Sigma_{00})\\ &= \mathbb{E}_{\sigma_{00} \sim P_{00}}[\|(F_2 \circ F_1 \circ G_2 \circ G_1)(\sigma_{00}) - \sigma_{00}\|_1]\\ &+ \mathbb{E}_{\sigma_{00} \sim P_{00}}[\|(F_1 \circ F_2 \circ G_1 \circ G_2)(\sigma_{00}) - \sigma_{00}\|_1].\end{aligned}$$

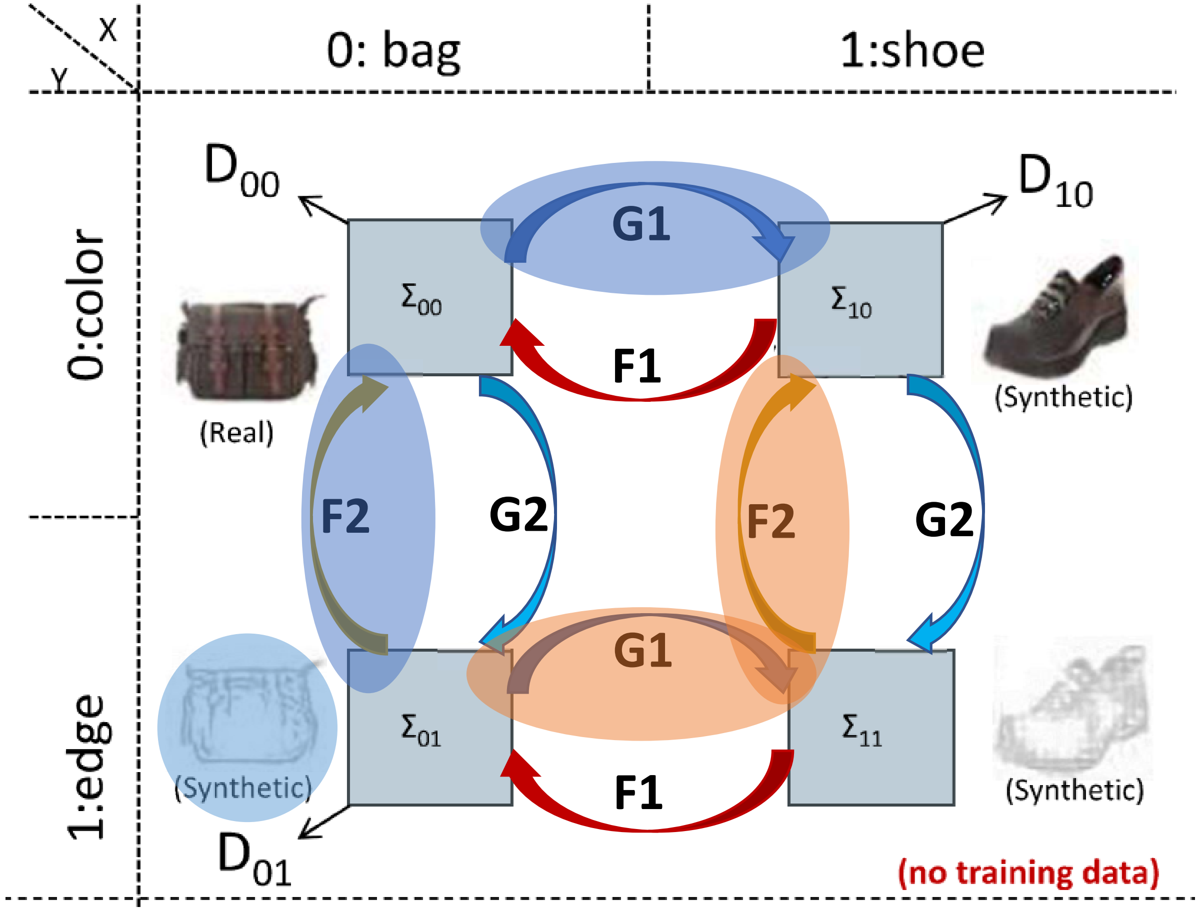
Loss: Commutative



$$\boxed{\mathcal{L}_{comm}(G_1, G_2, \Sigma_{00})}$$

$$\mathcal{L}_{comm}(G_1, G_2, \Sigma_{00}) = \mathbb{E}_{\sigma_{00} \sim P_{00}} [\| (G_2 \circ G_1)(\sigma_{00}) - (G_1 \circ G_2)(\sigma_{00}) \|_1]$$

Loss: Commutative



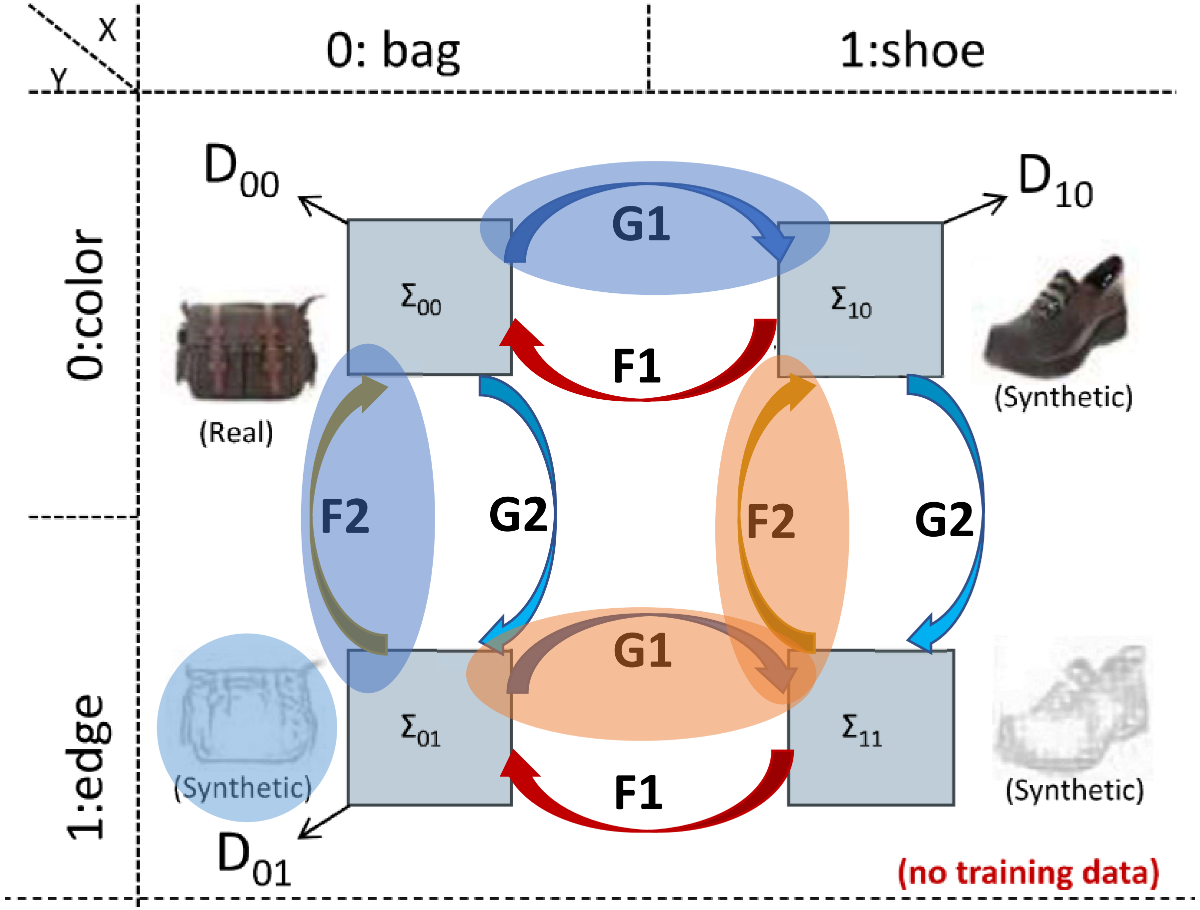
$$\mathcal{L}_{comm}(G_1, G_2, \Sigma_{00})$$

$$\boxed{\mathcal{L}_{comm}(G_1, F_2, \Sigma_{01})}$$

$$\mathcal{L}_{comm}(F_1, G_2, \Sigma_{10})$$

$$\mathcal{L}_{comm}(G_1, G_2, \Sigma_{00}) \\ = \mathbb{E}_{\sigma_{00} \sim P_{00}} [\|(G_2 \circ G_1)(\sigma_{00}) - (G_1 \circ G_2)(\sigma_{00})\|_1]$$

Loss: Commutative



$$\mathcal{L}_{comm}(G_1, G_2, \Sigma_{00})$$

$$\boxed{\mathcal{L}_{comm}(G_1, F_2, \Sigma_{01})}$$

$$\mathcal{L}_{comm}(F_1, G_2, \Sigma_{10})$$

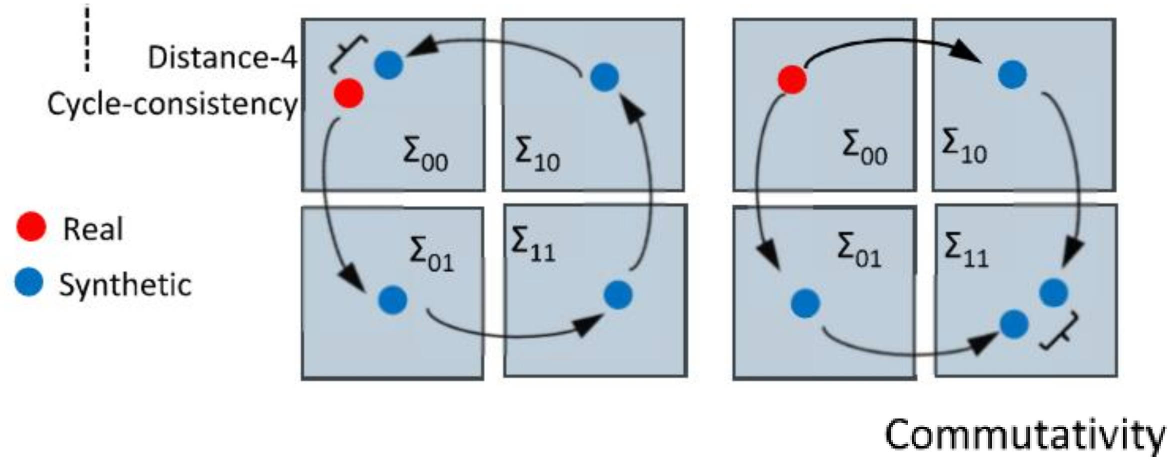
$$\mathcal{L}_{comm}(G_1, G_2, \Sigma_{00}) \\ = \mathbb{E}_{\sigma_{00} \sim P_{00}}[\|(G_2 \circ G_1)(\sigma_{00}) - (G_1 \circ G_2)(\sigma_{00})\|_1]$$

Concept GAN

$$\mathcal{L}(G, F, D, \Sigma) = \mathcal{L}_{ADV} + \lambda\mathcal{L}_{CYC} + \mu\mathcal{L}_{COMM}$$

$$\mathcal{L}_{CYC} = \mathcal{L}_{CYC2} + \mathcal{L}_{CYC4}$$

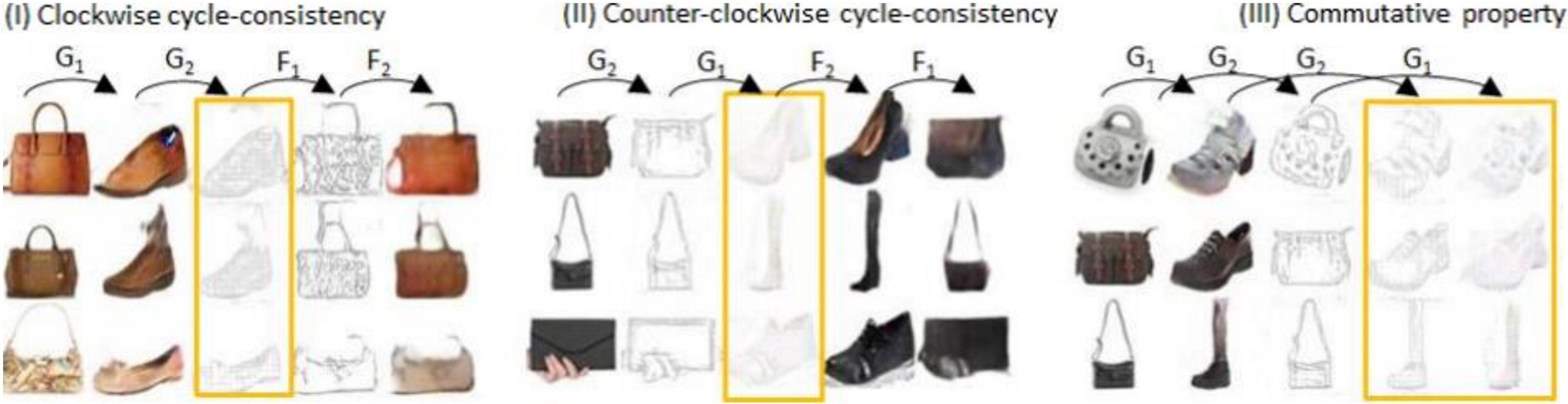
$$G^*, F^* = \arg\min_{G,F}\max_{D}\mathcal{L}(G, F, D, \Sigma)$$



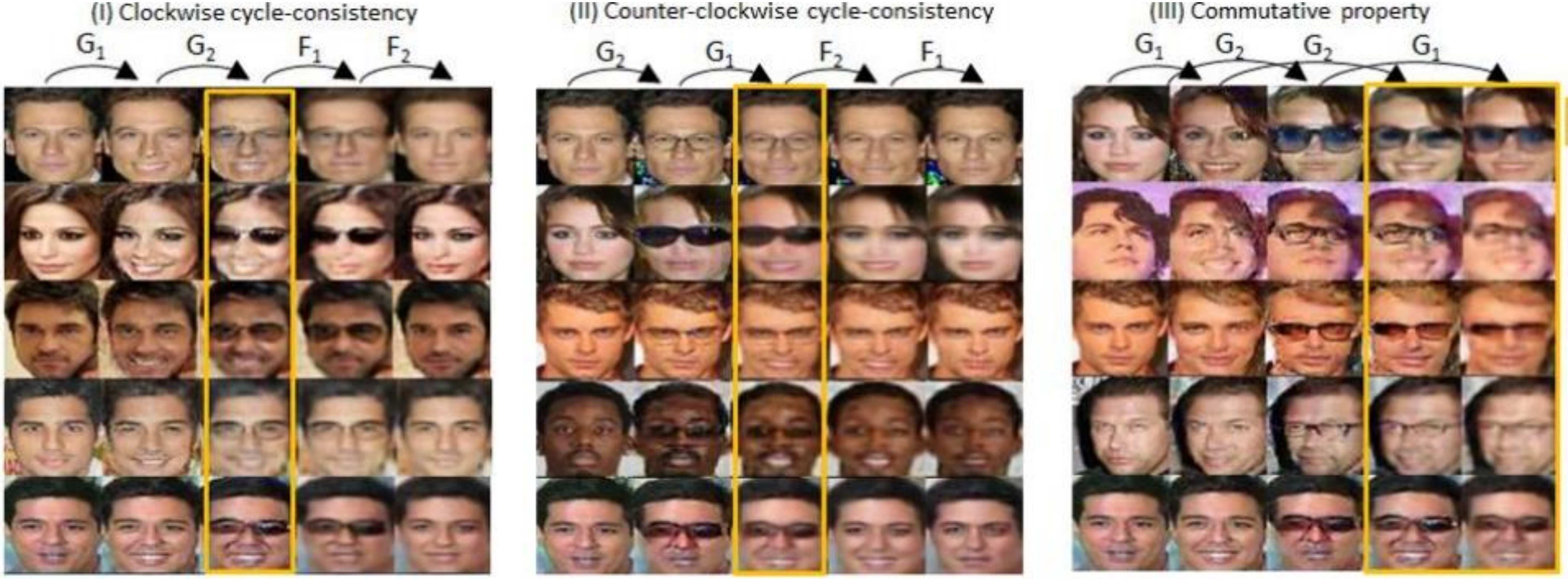
Proposed: ConceptGAN

Comparing Results

Qualitative Results



Qualitative Results



Quantitative Results (Classification)

Test Set: edge Shoes only

Classifier	Val	CycleGAN	Ours
C1: “color/shoe” vs. “edge/shoe”	99	0	**99**
C2: “edge/handbag” vs. “edge/shoe”	99	**99**	98
Both C1 and C2	N/A	0	**98**

Test Set: with eyeglasses, with bangs

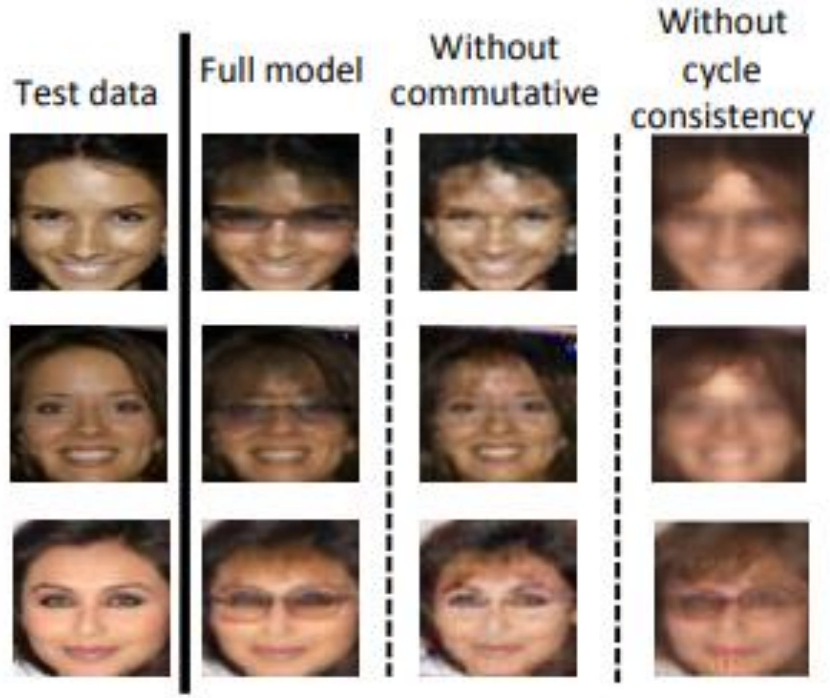
Classifier	Val	CycleGAN	Ours
C1: “with” vs. “no” eyeglasses	98	93	**98**
C2: “with” vs. “no” bangs	93	61	**67**
Both C1 and C2	N/A	56	**66**

18 60

$\mathcal{L}_{cyc4}$ $\mathcal{L}_{comm}$

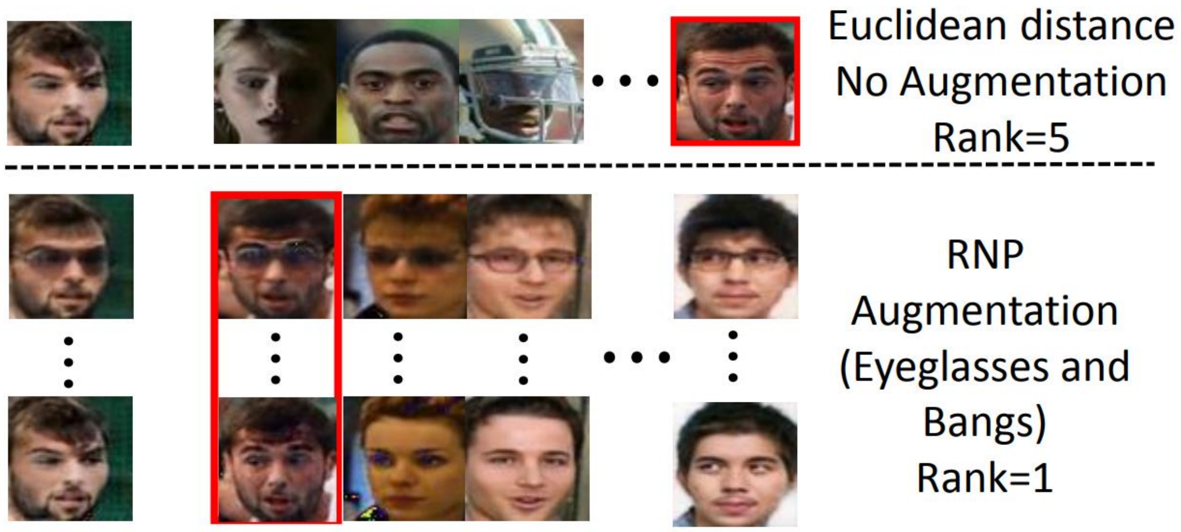
(Accuracy if L_x removed)

Ablation Study



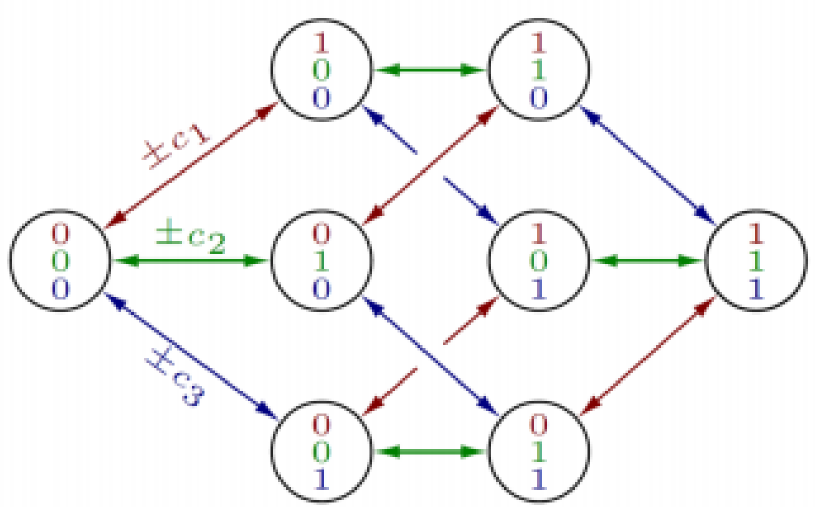
Re-ID

Attributes	Smiling & Eyeglasses			Bangs & Eyeglasses			Smiling, Bangs, & Eyeglasses		
Ranking Method	l_2	RNP	SRID	l_2	RNP	SRID	l_2	RNP	SRID
Augmentation	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
CaffeFace	8.3	10.7	12.8	7.9	12.3	16.9	11.5	13.3	16.6
VGGFace	38.6	43.9	49.4	49.8	59.4	61.5	44.4	54.8	58.6

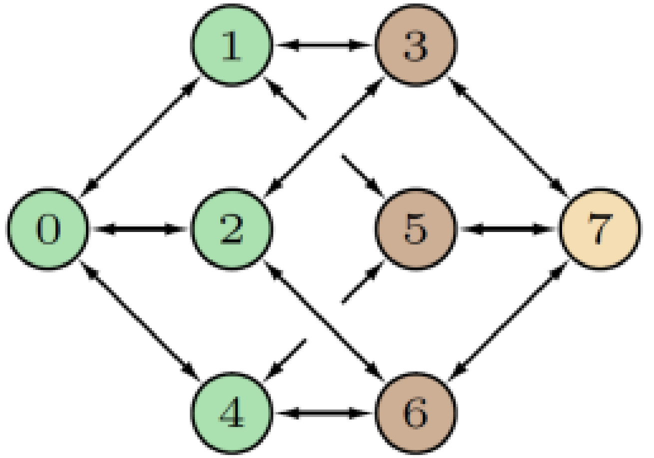


Ranking method	l_2	SRID
Augmentation	No	Yes
LFW	9.5	13.1
MS-Celeb1M	11.7	14.8

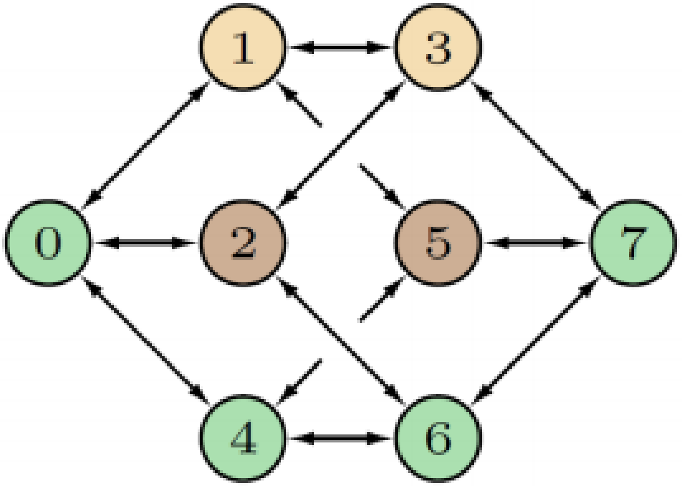
Generalizing to n(=3) Concepts



(a)



(b)



(c)

Learning from 2 attributes in 2 experiments

Experiment1: (smile, eyeglasses)

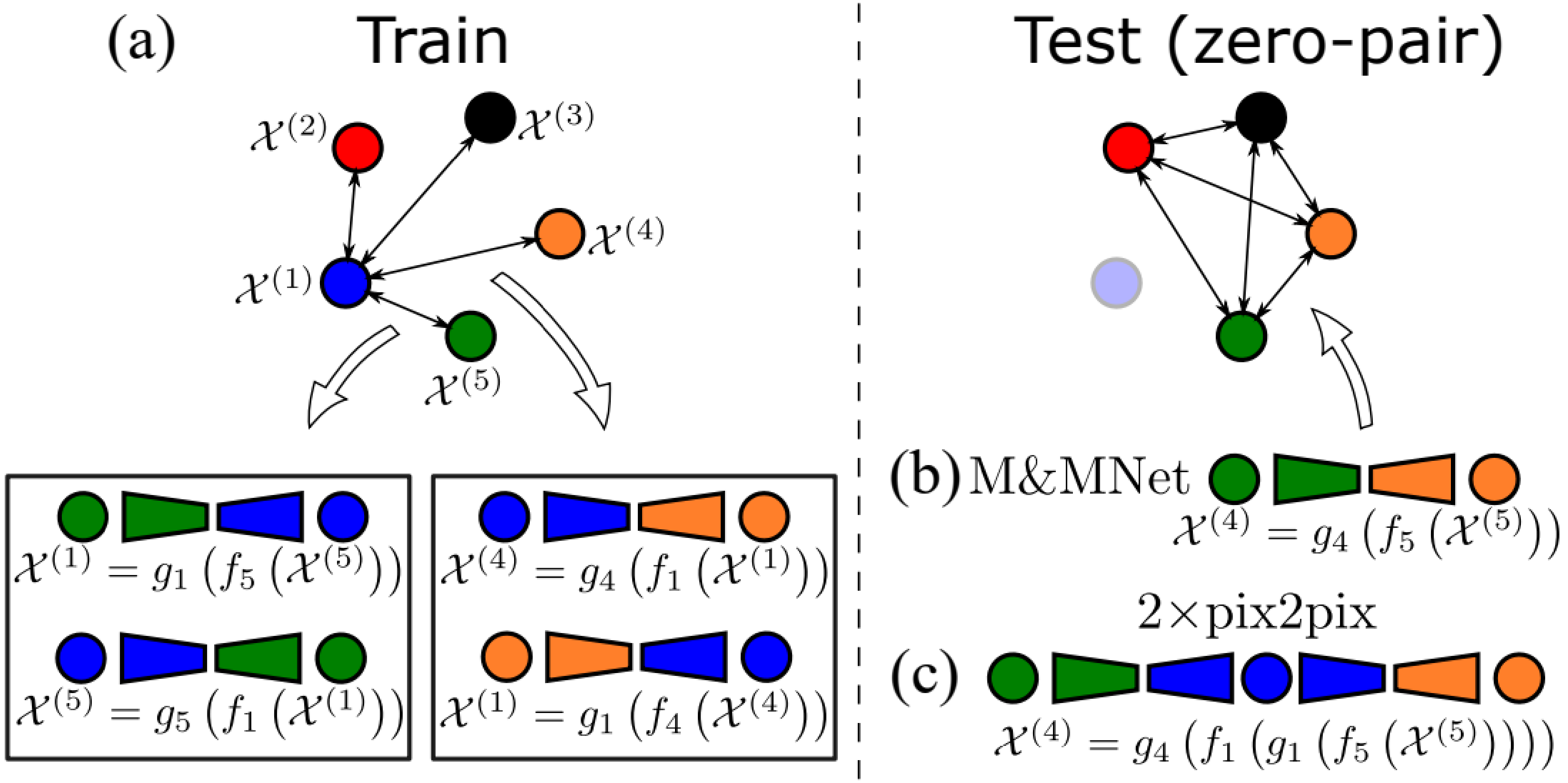
Experiment2: (bangs, eyeglasses)

(smile, eyeglasses, bangs)

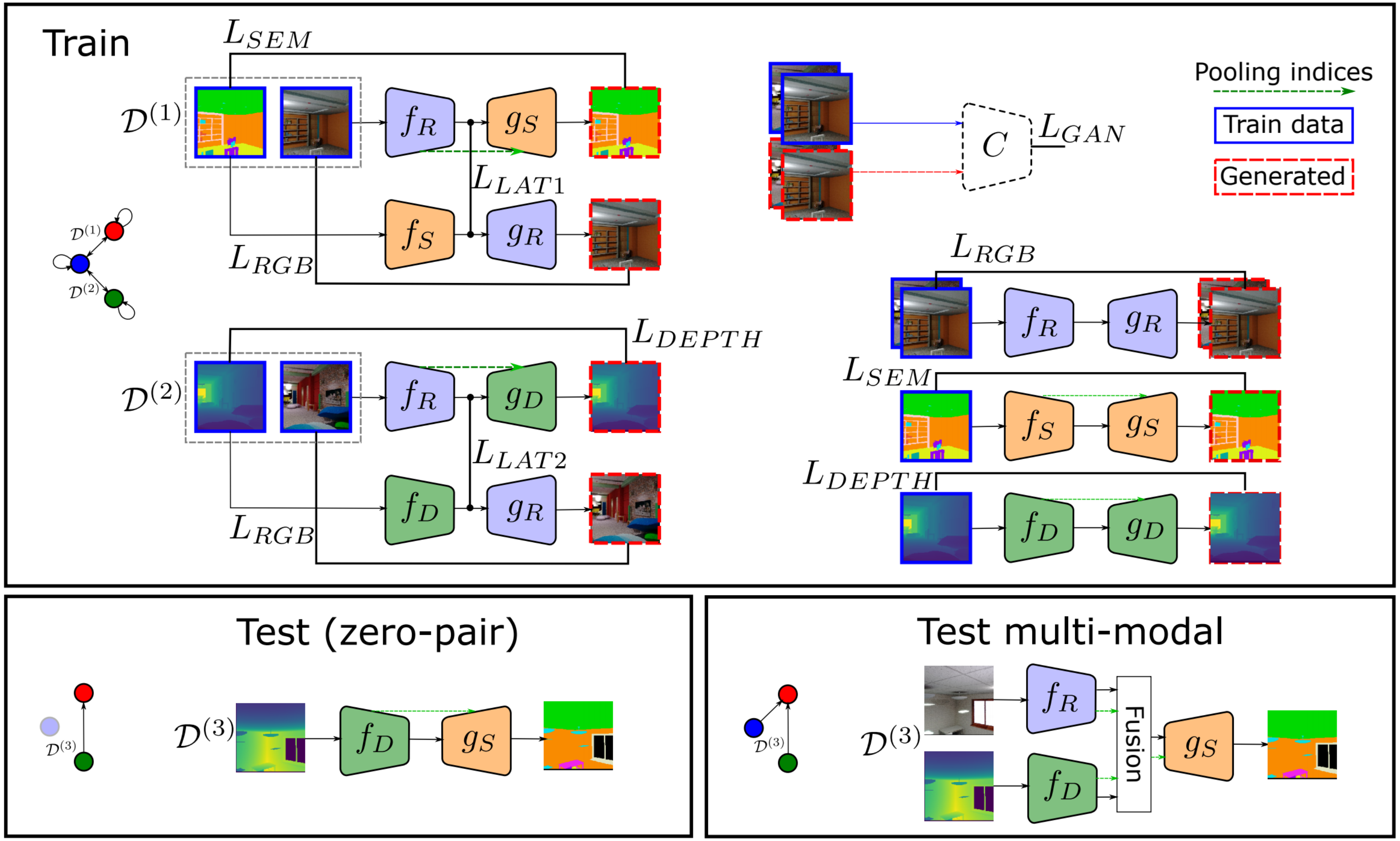
Mix and match networks: encoder-decoder alignment for zero-pair image translation

Yaxing Wang, Joost van de Weijer, Luis Herranz
CVPR 2018

Task


(a)
Train
$\mathcal{X}^{(1)}$
$\mathcal{X}^{(2)}$
$\mathcal{X}^{(3)}$
$\mathcal{X}^{(4)}$
$\mathcal{X}^{(5)}$
$\mathcal{X}^{(1)} = g_1\left(f_5\left(\mathcal{X}^{(5)}\right)\right)$
$\mathcal{X}^{(5)} = g_5\left(f_1\left(\mathcal{X}^{(1)}\right)\right)$
$\mathcal{X}^{(4)} = g_4\left(f_1\left(\mathcal{X}^{(1)}\right)\right)$
$\mathcal{X}^{(1)} = g_1\left(f_4\left(\mathcal{X}^{(4)}\right)\right)$
Test (zero-pair)
(b) M&MNet
$\mathcal{X}^{(4)} = g_4\left(f_5\left(\mathcal{X}^{(5)}\right)\right)$
2×pix2pix
(c)
$\mathcal{X}^{(4)} = g_4\left(f_1\left(g_1\left(f_5\left(\mathcal{X}^{(5)}\right)\right)\right)\right)$

Overview



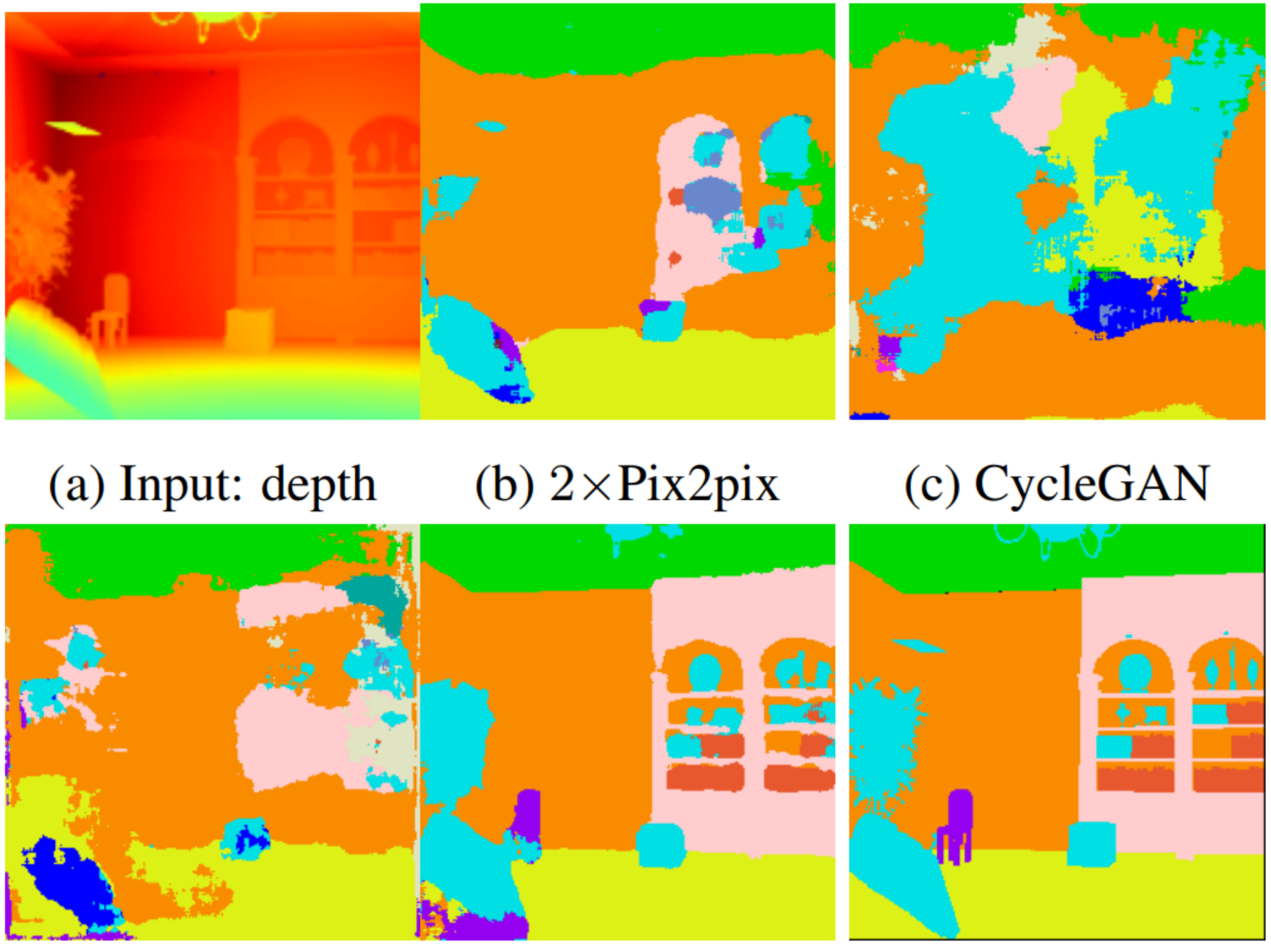
Train
L_{SEM}
$\mathcal{D}^{(1)}$
f_R
g_S
L_{LAT1}
f_S
g_R
L_{RGB}
$\mathcal{D}^{(1)}$
$\mathcal{D}^{(2)}$
L_{DEPTH}
$\mathcal{D}^{(2)}$
f_R
g_D
L_{LAT2}
f_D
g_R
L_{RGB}
C
L_{GAN}
Pooling indices
Train data
Generated
L_{RGB}
f_R
g_R
L_{SEM}
f_S
g_S
L_{DEPTH}
f_D
g_D
Test (zero-pair)
$\mathcal{D}^{(3)}$
$\mathcal{D}^{(3)}$
f_D
g_S
Test multi-modal
$\mathcal{D}^{(3)}$
$\mathcal{D}^{(3)}$
f_R
f_D
Fusion
g_S

Losses

$$L = \lambda_R L_{RGB} + \lambda_S L_{SEG} + \lambda_D L_{DEPTH} + \lambda_A L_{LAT}$$

$$\begin{aligned}
L_{LAT} &= L_{LAT1} + L_{LAT2} \\
L_{LAT1} &= \mathbb{E}_{(x,y)\sim p_{\mathcal{D}(1)}(x,y)}\left[\left\| f_R(x) - f_S(y) \right\|_2\right] \\
L_{LAT2} &= \mathbb{E}_{(x,z)\sim p_{\mathcal{D}(2)}(x,z)}\left[\left\| f_R(x) - f_D(z) \right\|_2\right]
\end{aligned}$$

Depth –to- Semantic segmentation



(a) Input: depth (b) 2×Pix2pix (c) CycleGAN

(d) D→R→S (e) Proposed (f) Ground truth

Depth-to-Semantic segmentation

Method	Conn.	L_{SEM}	Bed	Book	Ceiling	Chair	Floor	Furniture	Object	Picture	Sofa	Table	TV	Wall	Window	mIoU	Global
Baselines																	
CycleGAN [34]	SC	CE	2.79	0.00	16.9	6.81	4.48	0.92	7.43	0.57	9.48	0.92	0.31	17.4	15.1	6.34	14.2
2×pix2pix [10]	SC	CE	34.6	1.88	70.9	20.9	63.6	17.6	14.1	0.03	38.4	10.0	4.33	67.7	20.5	25.4	57.6
M&MNet $D \rightarrow R \rightarrow S$	PI	CE	0.02	0.00	8.76	0.10	2.91	2.06	1.65	0.19	0.02	0.28	0.02	58.2	3.3	5.96	32.3
M&MNet $D \rightarrow R \rightarrow S$	SC	CE	25.4	0.26	82.7	0.44	56.6	6.30	23.6	5.42	0.54	21.9	10.0	68.6	19.6	24.7	59.7
Zero-pair																	
M&MNet $D \rightarrow S$	PI	CE	50.8	18.9	89.8	31.6	88.7	48.3	44.9	62.1	17.8	49.9	51.9	86.2	79.2	55.4	80.4
Multi-modal																	
M&MNet $(R, D) \rightarrow S$	PI	CE	49.9	25.5	88.2	31.8	86.8	56.0	45.4	70.5	17.4	46.2	57.3	87.9	79.8	57.1	81.2