FineGAN: Unsupervised Hierarchical Disentanglement for Fine-Grained Object Generation and Discovery

Krishna Kumar Singh* Utkarsh Ojha* Yong Jae Lee
University of California, Davis

CVPR 2019 (Oral)

2019.09.03

Yonggyu Kim

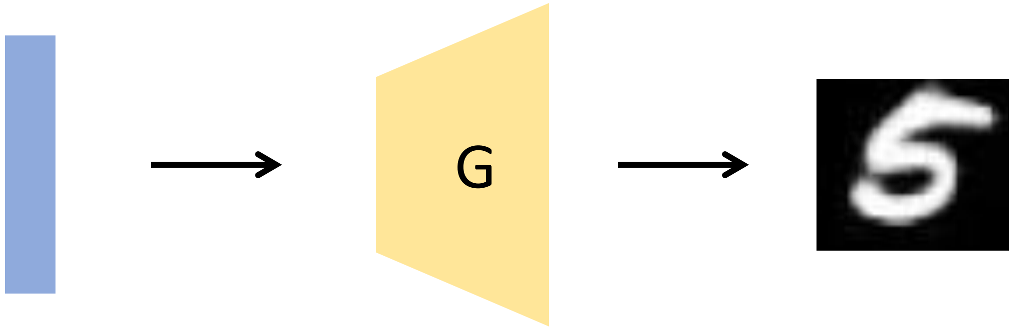
DAVIAN
Data and Visual Analytics Lab

Previous (InfoGAN)

▪ Motivation

Unsupervised learning으로 Representation이 disentangle 할 수 있도록 학습하여, 원하는 image를 생성하길 원함.

<u>즉, vector representation(distribution)이 유의미하도록 만들자.</u>



Previous (InfoGAN)

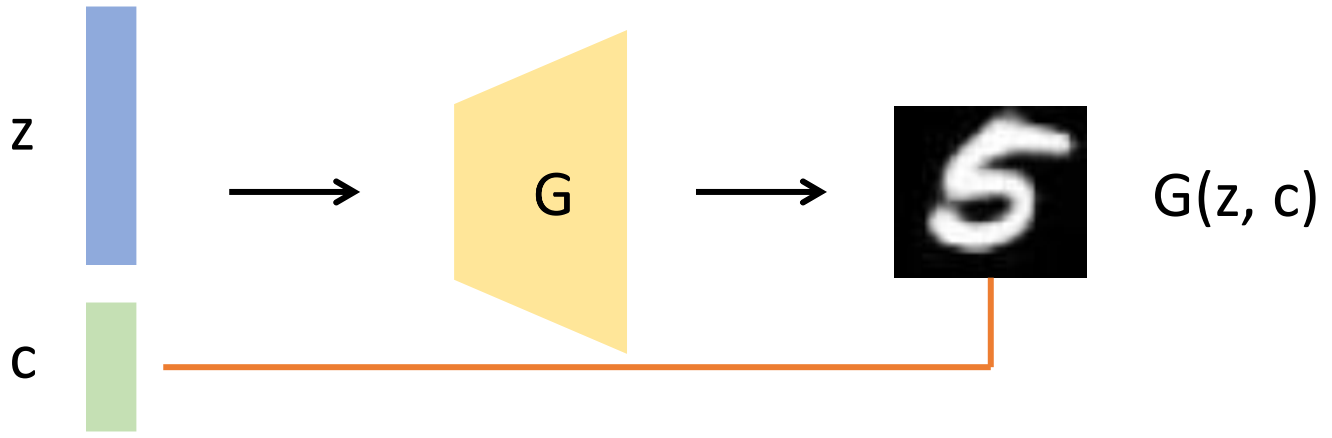
- Main idea

$$\min_{G}\max_{D} V_I(D,G) = V(D,G) - \lambda I(c; G(z,c))$$

$$I(X;Y) = H(X) - H(X|Y)$$

- 의존 관계 크다.
- G가 바뀌면 c도 바뀐다.
- 둘 간의 공유하는 정보량이 높아야 한다.

- 기존 GAN처럼 G, D를 학습하되, latent code c ~ P(c)에 대해서 c 와 G(z, c)가 의존관계에 놓이게 되도록 학습



Previous (InfoGAN)

- Main idea

$$\min_G \max_D V_I(D,G) = V(D,G) - \lambda I(c; G(z,c))$$

$$I(c; G(z,c)) = H(c) - H(c|G(z,c))$$

$$H(X|Y) = -\int_x \int_y p(x,y) \log \frac{p(x,y)}{p(y)} dydx = \int_x \int_y p(x,y) \log p(y|x) dydx$$

$$= \mathbb{E}_{x \sim P_X}[\mathbb{E}_{y \sim P_Y}[\log P(Y|X)]]$$

$$= \mathbb{E}_{x \sim G(z,c)} [\mathbb{E}_{c' \sim P(c|x)}[\log P(c'|x)]] + H(c)$$

$$E_{x \sim G(z,c)} \left[\int \log p(c'|x) P(c') dc' \right]$$

$$= E_{x \sim G(z,c)} \left[\int \log \left(p(c'|x) \cdot \frac{P(c')}{Q(c'|x)} \cdot Q(c'|x) \, dc' \right) \right]$$

$$= E_{x \sim G(z,c)} \left[\int \log p(c'|x) \cdot Q(c'|x) dc' - \int \log \frac{Q(c'|x)}{P(c')} \cdot Q(c'|x) dc' \right]$$

$$= \mathbb{E}_{x \sim G(z,c)} [D_{KL}(P(\cdot|x) || Q(\cdot||x)) + \mathbb{E}_{c' \sim P(c|x)}[\log Q(c'|x)]] + H(c)$$

$$\geq \mathbb{E}_{x \sim G(z,c)} [\mathbb{E}_{c' \sim P(c|x)}[\log Q(c'|x)]] + H(c)$$

DAVIAN
Data and Visual Analytics Lab

Previous (InfoGAN)

- Lemma 5.1 & 증명

- For random variables X, Y and function f(x,y) under suitable regularity conditions:

$$\mathbb{E}_{x\sim X, y\sim Y|x}[f(x,y)] = \mathbb{E}_{x\sim X, y\sim Y|x, x'\sim X|y}[f(x',y)]$$

$$\begin{aligned}
L_I(G,Q) &= E_{c\sim P(c), x\sim G(z,c)}[\log Q(c|x)] + H(c) && \longleftarrow \text{Goal eq} \\
&= E_{c\sim P(c), x\sim P_G(x|z,c)}[\log Q(c|x)] + H(c) \\
&= E_{c\sim P(c), x\sim P_G(x|z,c), c'\sim P(c|x)}[\log Q(c'|x)] + H(c) && \longleftarrow \text{Lemma 5.1} \\
&= E_{x\sim P_G(x|z,c), c'\sim P(c|x)}[\log Q(c'|x)] + H(c) \\
&= E_{x\sim P_G(x|z,c)}[E_{c'\sim P(c|x)}[\log Q(c'|x)]] + H(c) \\
&\leq I(c; G(z,c))
\end{aligned}$$

DAVIAN
Data and Visual Analytics Lab

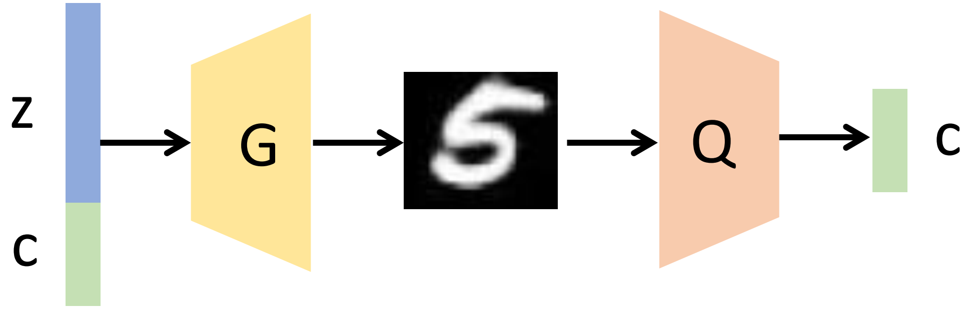
Previous (InfoGAN)

- Result

$$\min_{G,Q} \max_{D} V_{\text{InfoGAN}}(D,G,Q) = V(D,G) - \lambda L_I(G,Q)$$

$$L_I(G,Q) = \underbrace{E_{c\sim P(c),x\sim G(z,c)}[\log Q(c|x)]}_{\text{Reconstruction loss!}} + H(c)$$

- Intuitions
  - V(D,G) is objective function of GAN
  - 이에 더불어 G, Q 는 L_I도 최대화해야 함
  - 즉, Q 는 G(z,c) 를 다시 c로 잘 바꿔야 하고,
  - G 는 Q 가 잘 바꿀 수 있도록 x=G(z,c) 를 생성해야 한다.



Previous (InfoGAN)

- Result

(a) Varying c_1 on InfoGAN (Digit type)

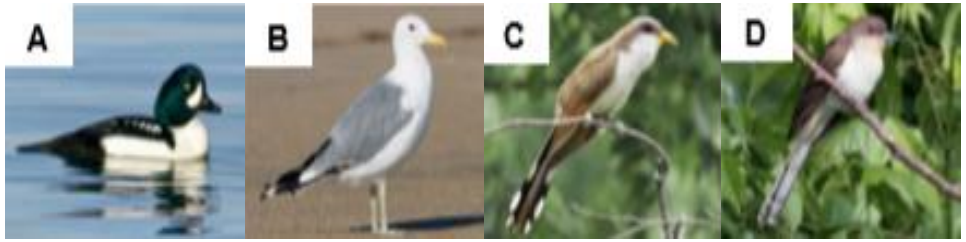
(b) Varying c_1 on regular GAN (No clear meaning)

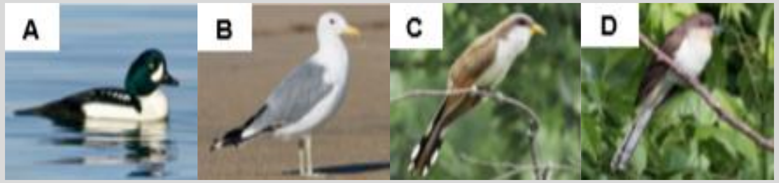
(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

(d) Varying c_3 from -2 to 2 on InfoGAN (Width)

Motivation

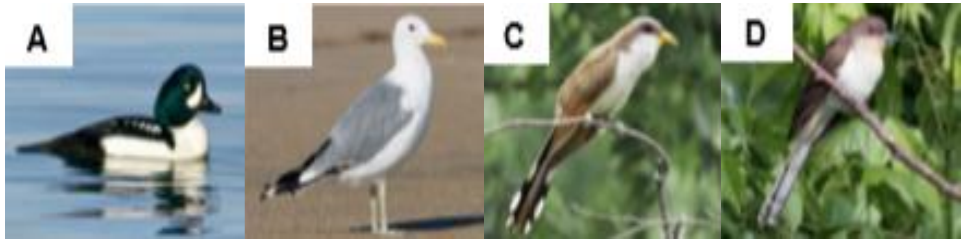
- Image grouping task

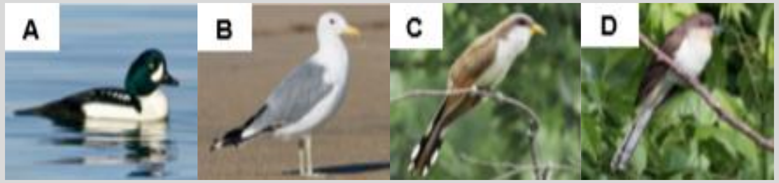




Motivation

- Image grouping task

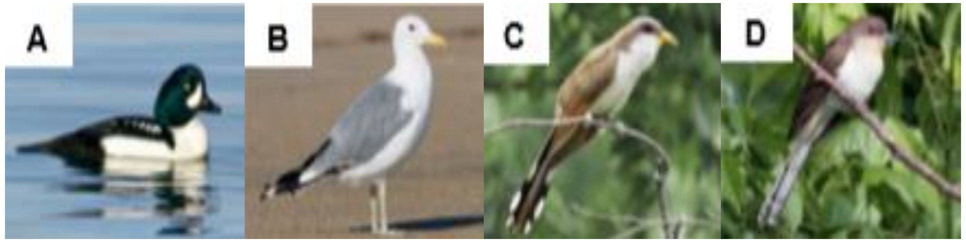


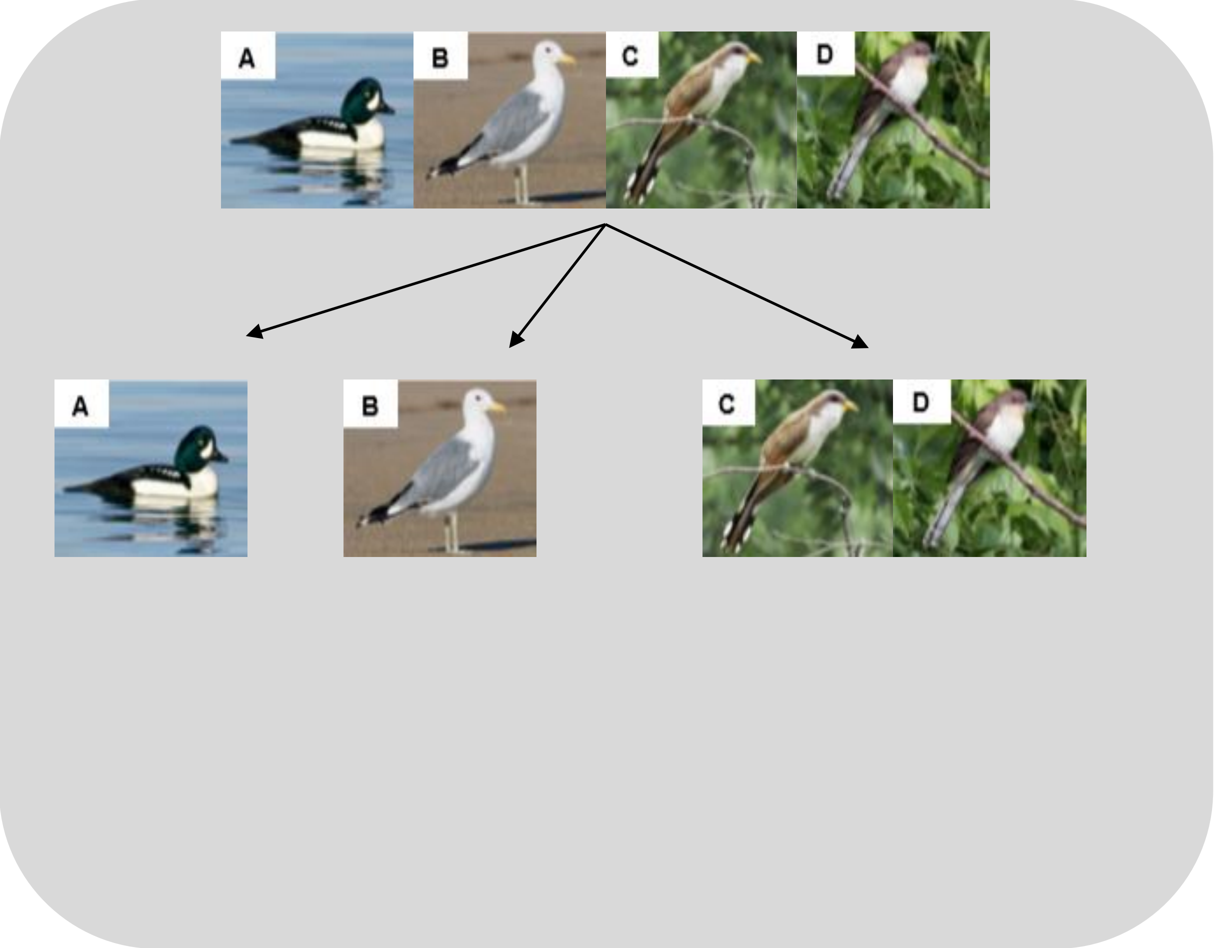


배경(Background)

Motivation

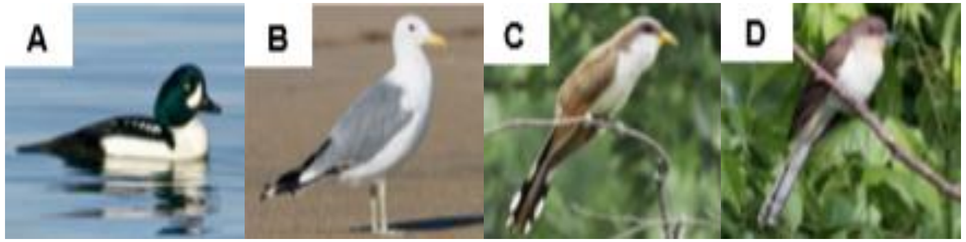
- Image grouping task

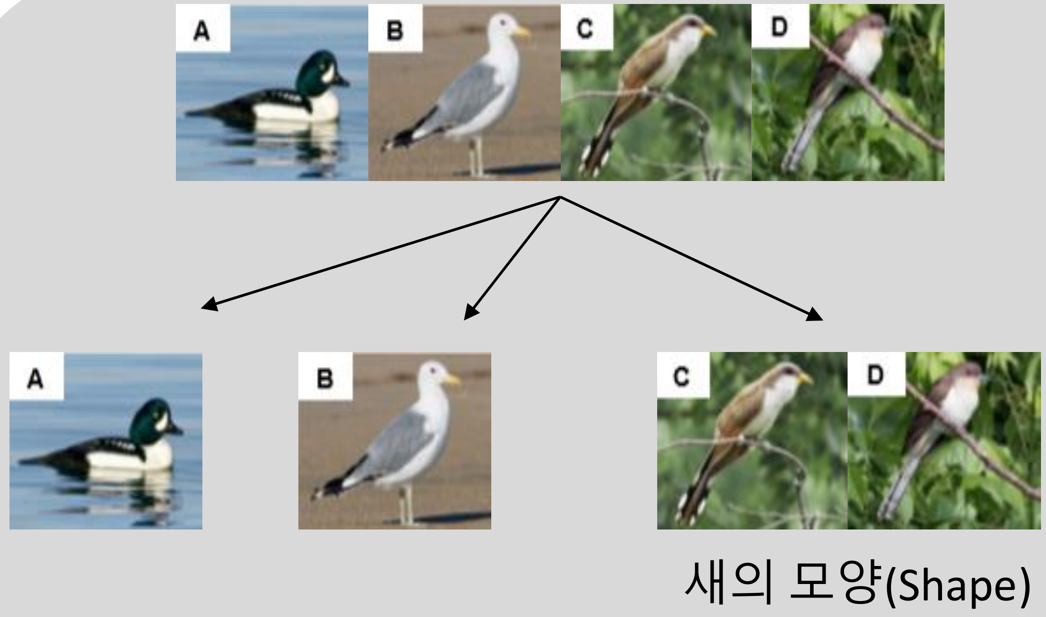




Motivation

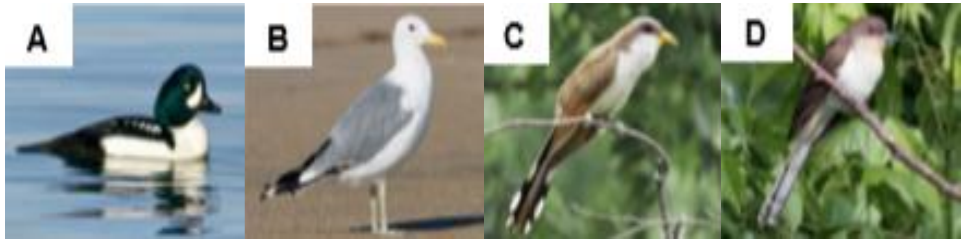
- Image grouping task

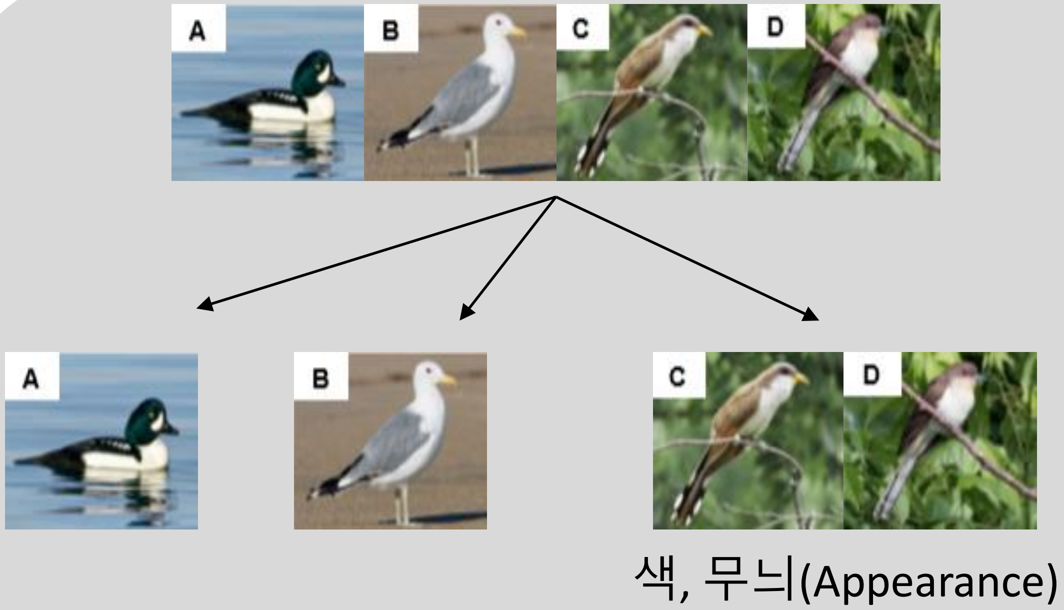




Motivation

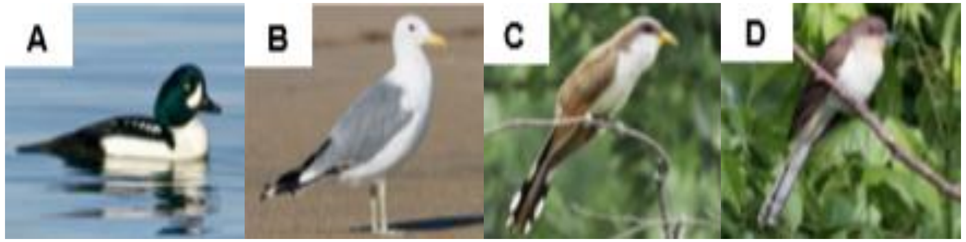
- Image grouping task

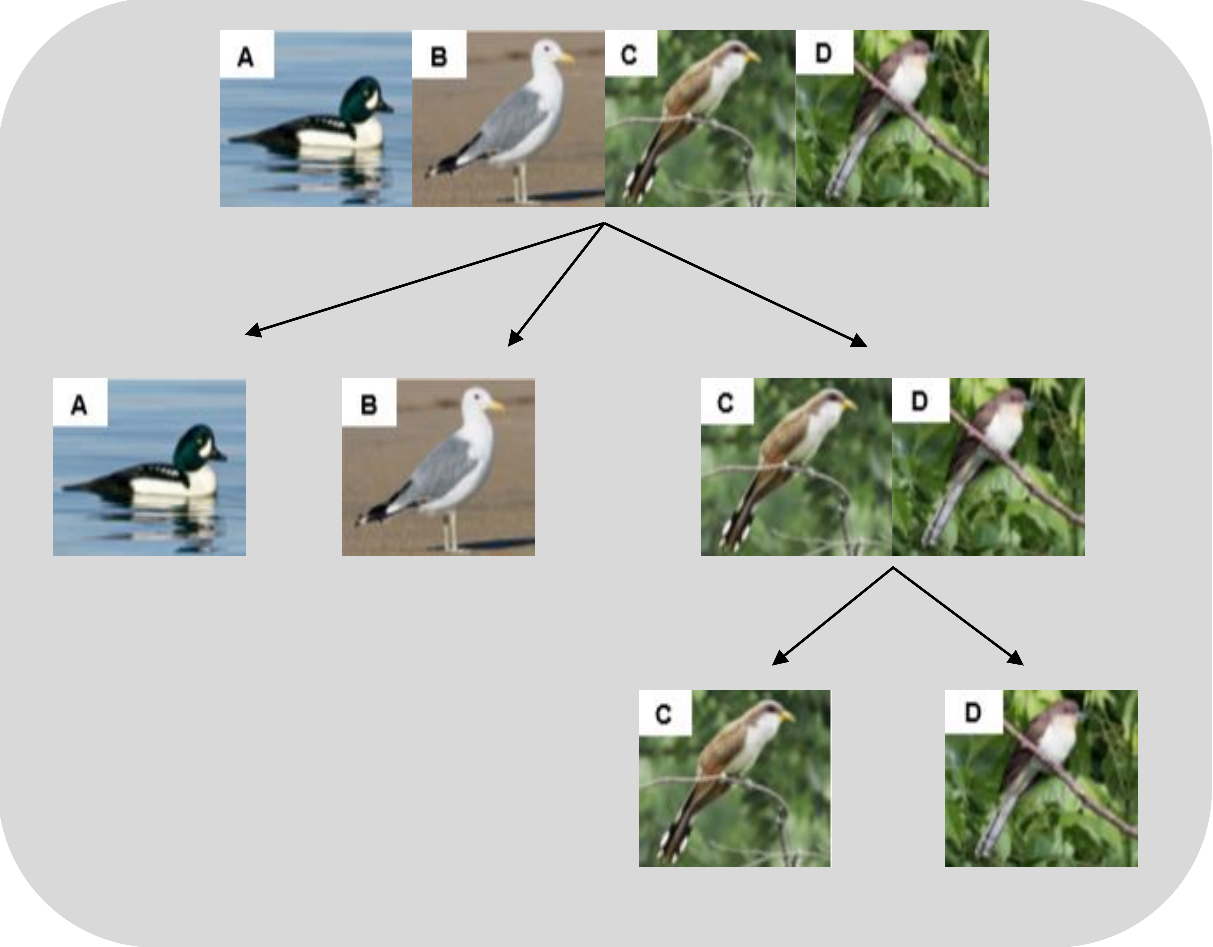




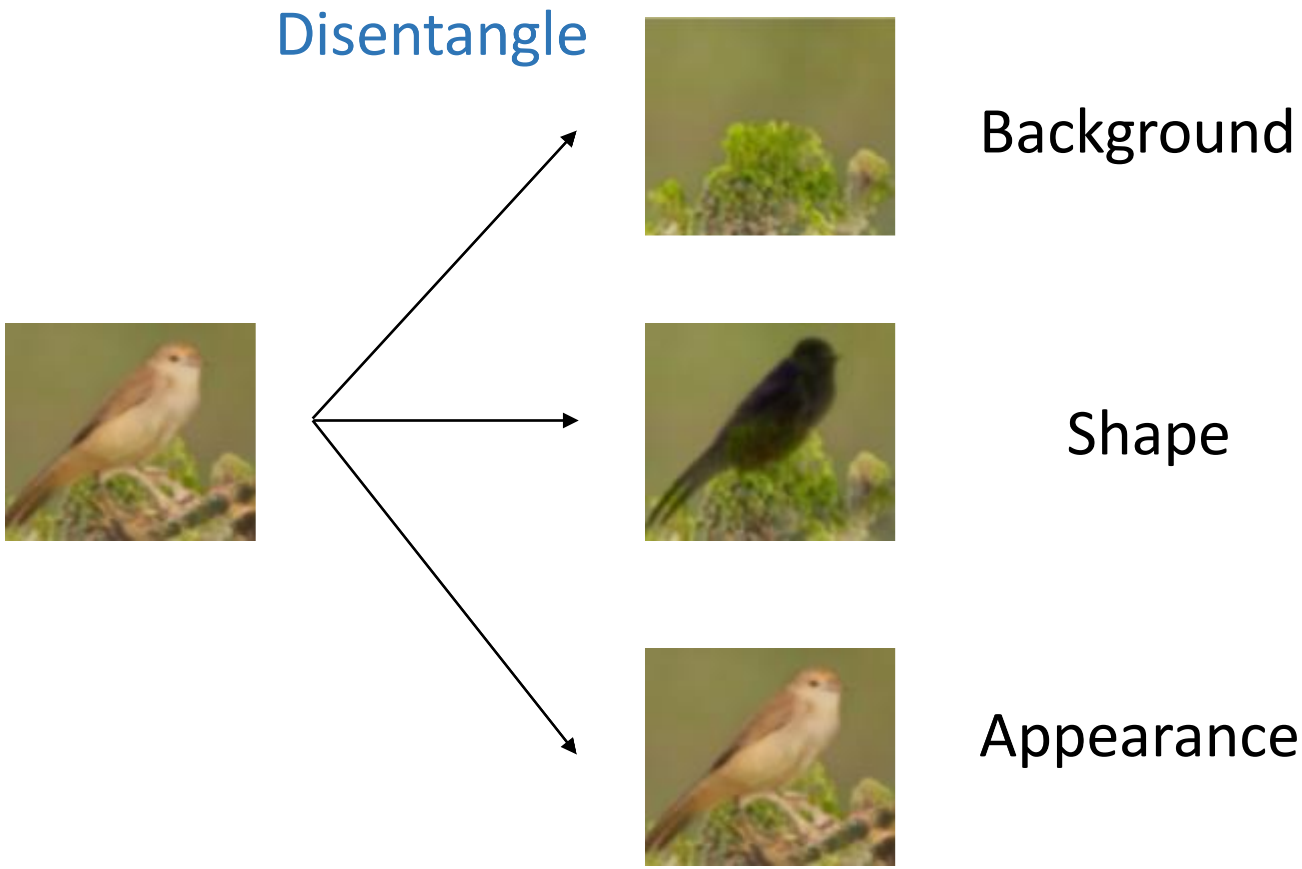
Motivation

- Image grouping task

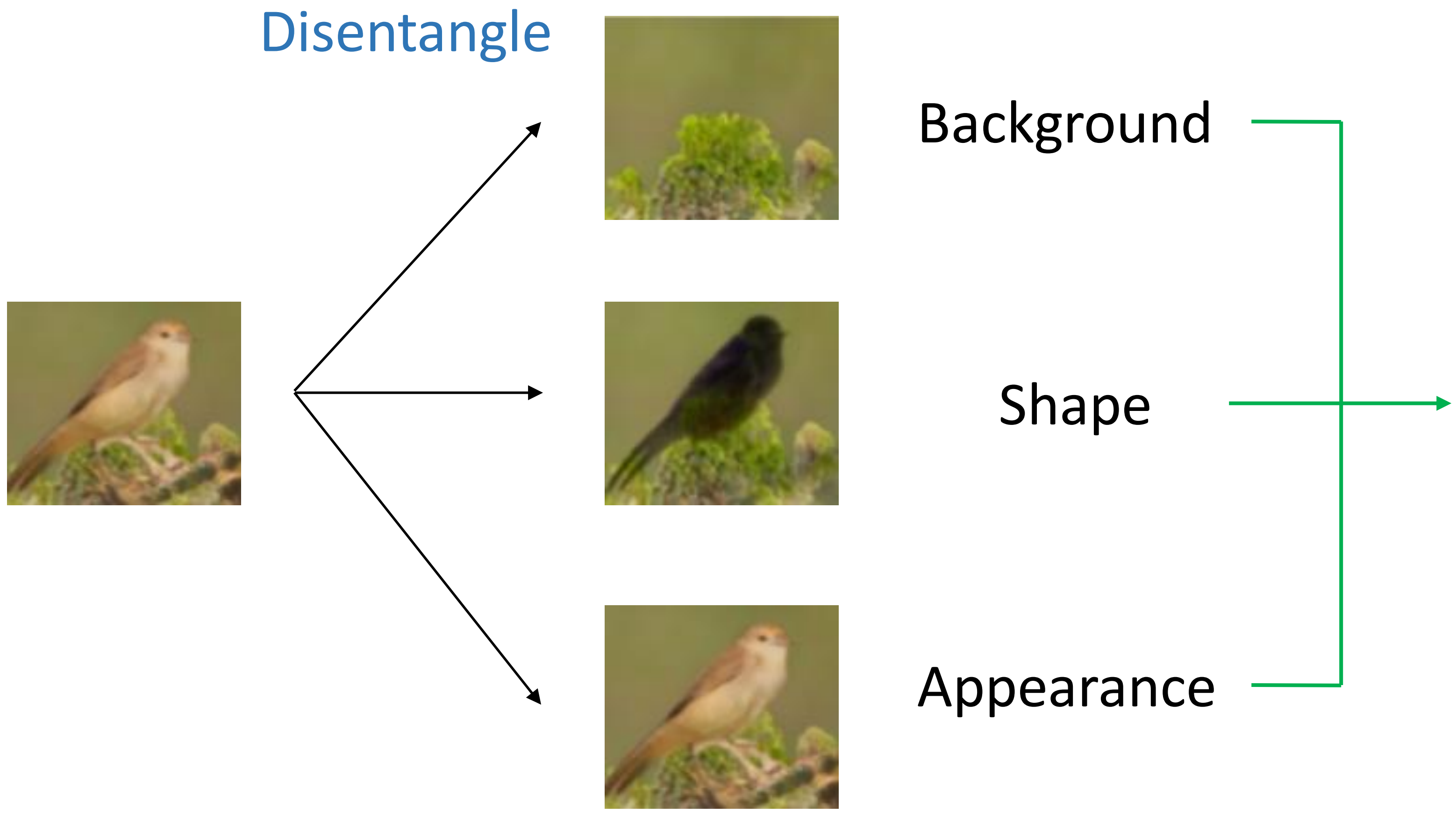




Motivation



Motivation



1. Image Generation
   <u>(without any manual fine-grained annotations)</u>

2. <u>Unsupervised</u> fine-grained object clustering (or “discovery”)

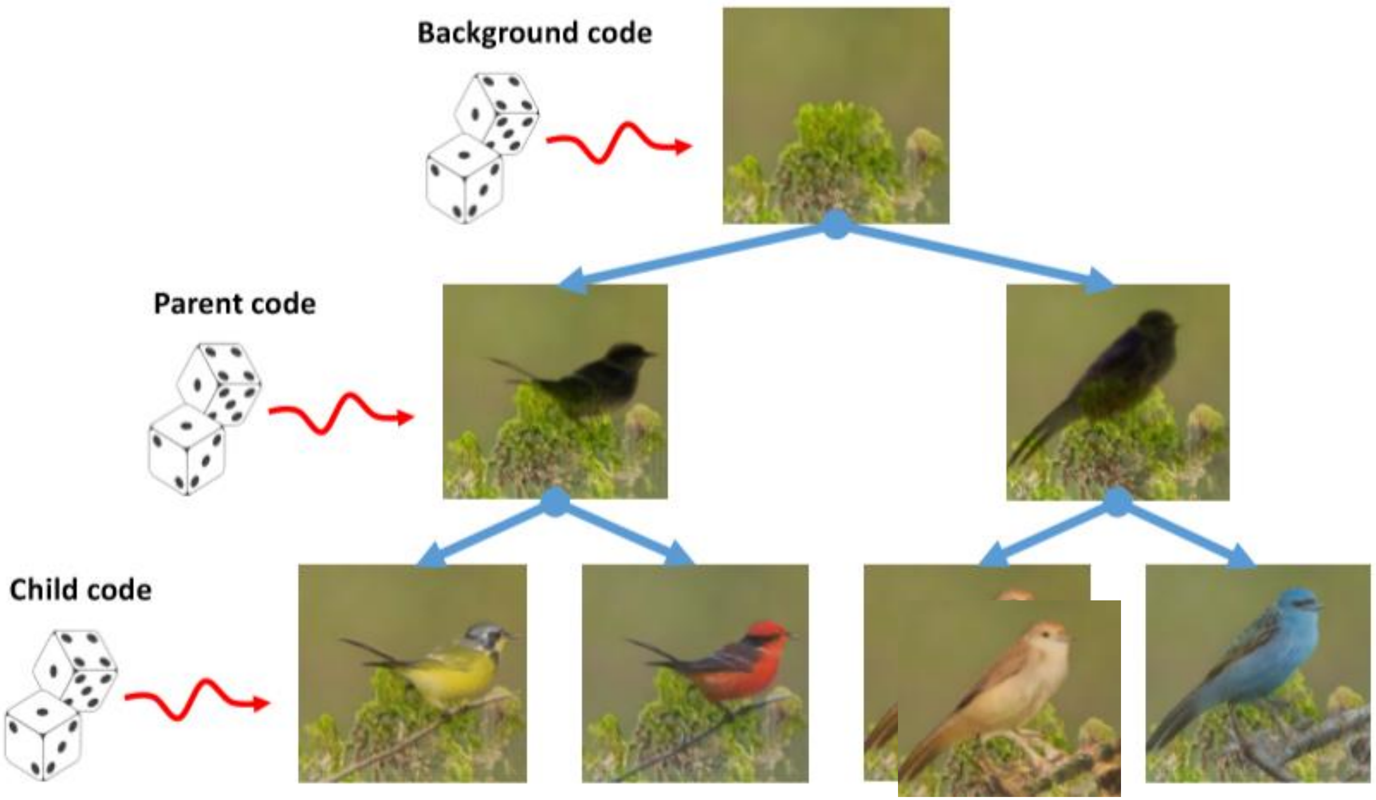
DAVIAN
Data and Visual Analytics Lab

Main Idea & Contribution

Authors hypothesize that a generative model with the capability of hierarchically generating images with fine-grained details can also be useful for fine-grained grouping of real images.

Main Idea & Contribution

1. Image Generation (without any manual fine-grained annotations)



FineGAN은 unsupervised 방식으로 fine-grained object의 background, shape, appearance를 계층적으로 잘 생성하도록 학습

Main Idea & Contribution

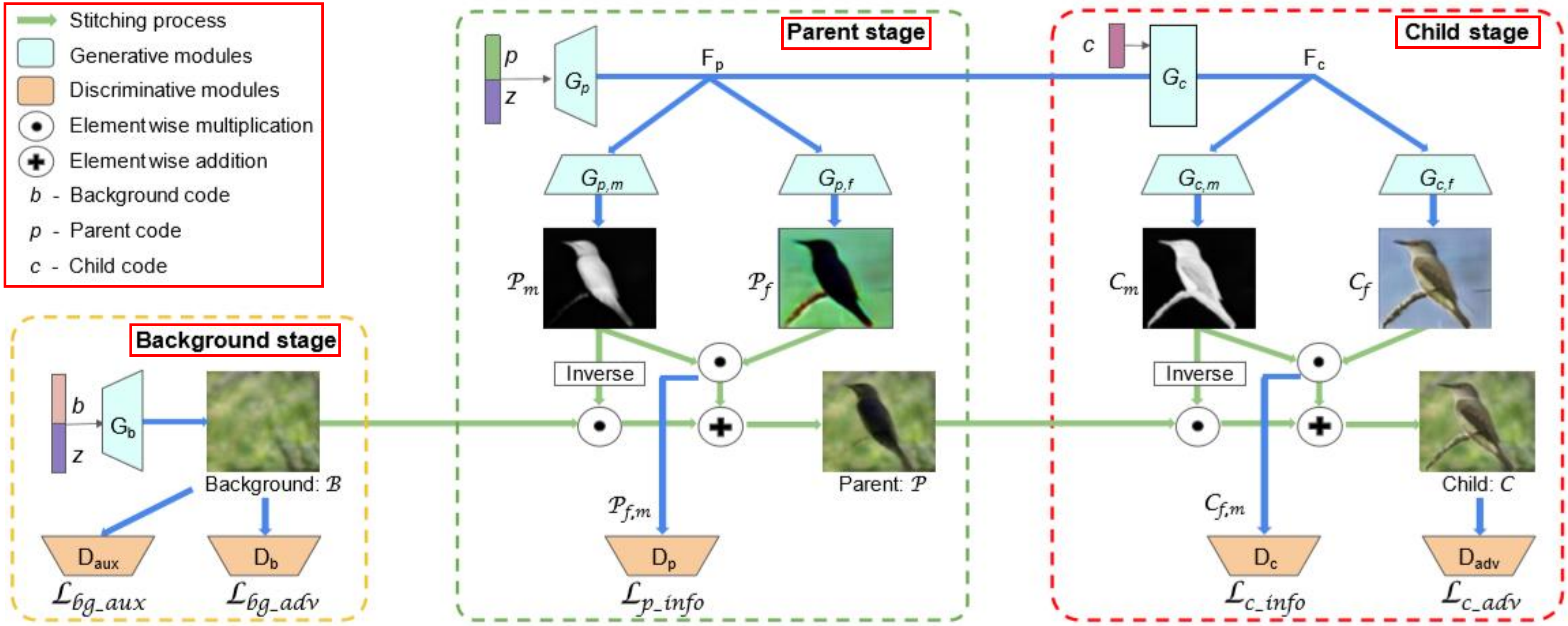
2. Unsupervised fine-grained object clustering (or “discovery”)

- This is the first attempt to cluster fine-grained categories in the unsupervised setting (Because, unsupervised object category discovery focuses only on clustering entry-level categories. (e.g. birds vs cars vs dogs))

- FineGAN learns disentangled representation to cluster real images for unsupervised fine-grained object category discovery.

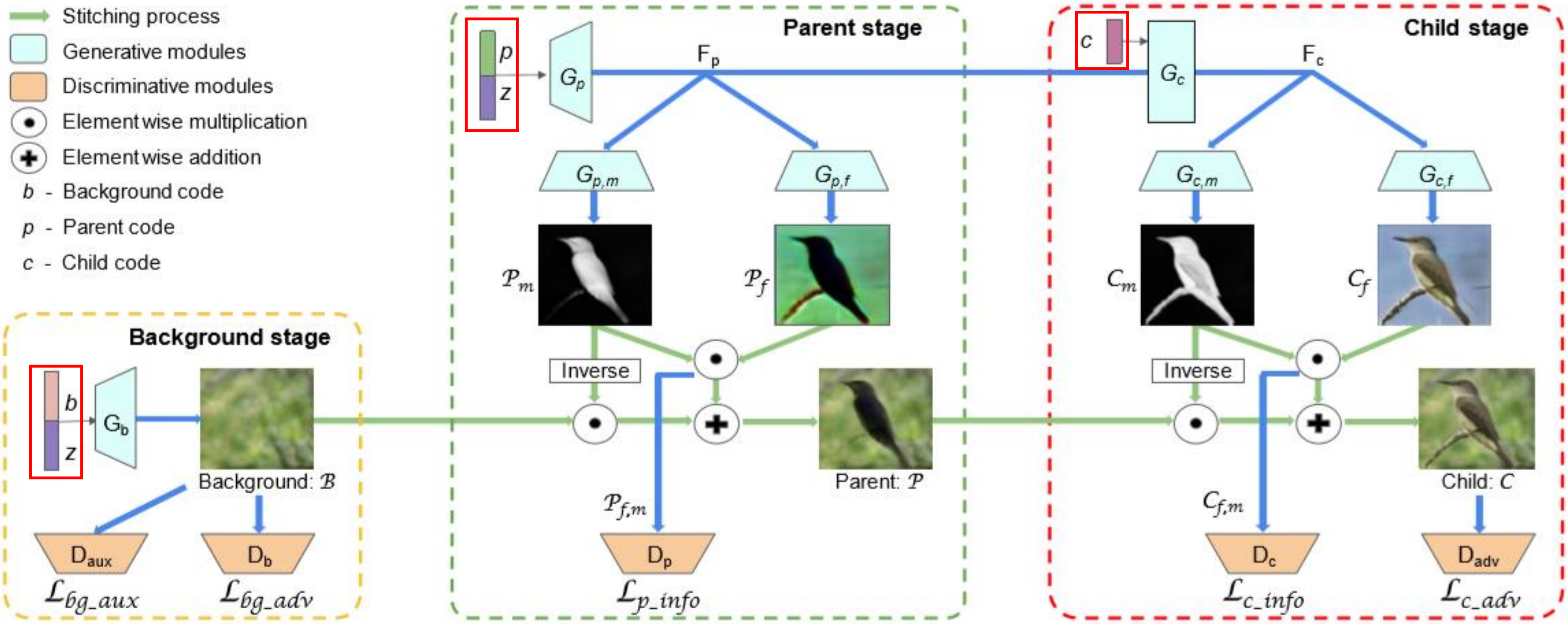
DAVIAN
Data and Visual Analytics Lab

Method



Stitching process
Generative modules
Discriminative modules
Element wise multiplication
Element wise addition
b - Background code
p - Parent code
c - Child code
Background stage
Parent stage
Child stage
p
z
G_p
F_p
$G_{p,m}$
$G_{p,f}$
$\mathcal{P}_m$
$\mathcal{P}_f$
Inverse
Parent: $\mathcal{P}$
$\mathcal{P}_{f,m}$
D_p
$\mathcal{L}_{p\_info}$
c
G_c
F_c
$G_{c,m}$
$G_{c,f}$
C_m
C_f
Inverse
Child: C
$C_{f,m}$
D_c
D_{adv}
$\mathcal{L}_{c\_info}$
$\mathcal{L}_{c\_adv}$
b
z
G_b
Background: $\mathcal{B}$
D_{aux}
D_b
$\mathcal{L}_{bg\_aux}$
$\mathcal{L}_{bg\_adv}$

Method



continuous noise vector $z \sim \mathcal{N}(0, 1)$

background code $b \sim \mathrm{Cat}(K = N_b, p = 1/N_b)$

parent code $p \sim \mathrm{Cat}(K = N_p, p = 1/N_p)$

child code $c \sim \mathrm{Cat}(K = N_c, p = 1/N_c)$

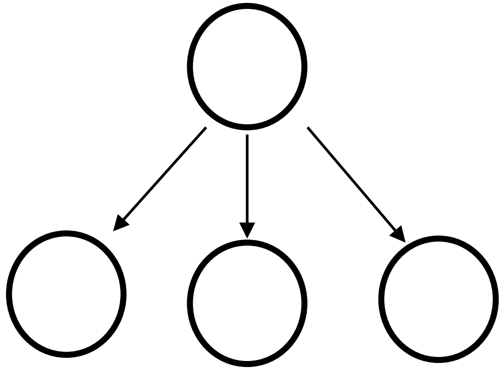
Method

- Relationship between latent code

Data에 implicit hierarchy가 존재한다고 가정
Hierarchy를 발견하기 위해 2가지 constraint를 걸어줌.

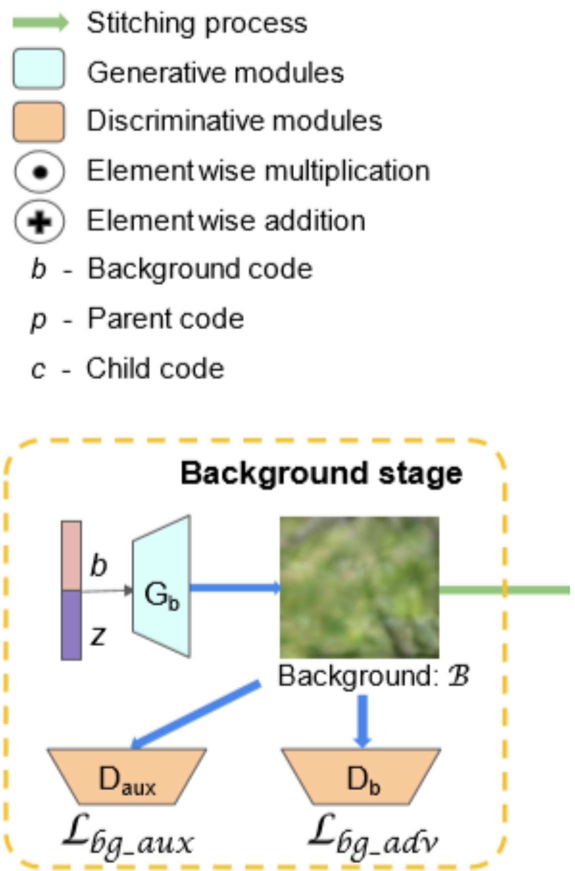
1) $(N_p < N_c)$

2) For each parent code, authors tie a fixed number of child codes (multiple child codes share the same parent code.)



Object 와 background 사이에 correlation(ducks in water)을 없애주기 위해
학습할 때, Background code 수 = child code 수

Method



Different(unknown) background classes

Adversarial loss(patch level) :

$$\mathcal{L}_{bg\_adv} = \min_{G_b} \max_{D_b} \mathbb{E}_x[\log(D_b(x))] + \mathbb{E}_{z,b}[\log(1 - D_b(G_b(\boxed{z}, \boxed{b})))]$$

Intra-class background details

Auxiliary background classification loss :

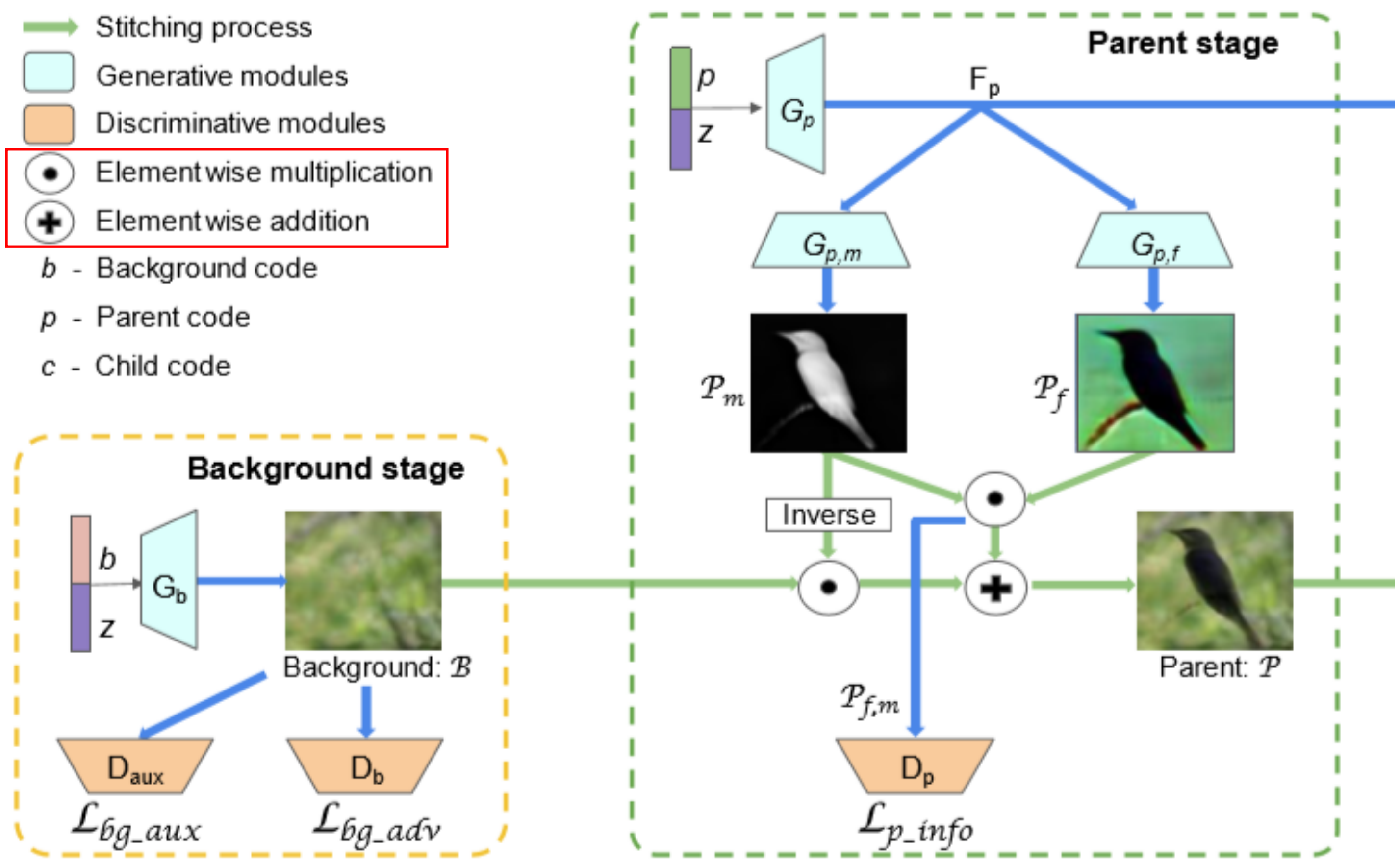
$$\mathcal{L}_{bg\_aux} = \min_{G_b} \mathbb{E}_{z,b}[\log(1 - D_{aux}(G_b(z, b)))]$$

Foreground : 1
Background : 0

Background loss :

$$\mathcal{L}_b = \mathcal{L}_{bg\_adv} + \mathcal{L}_{bg\_aux}$$

Method



$$\mathcal{P}_{f,m} = \mathcal{P}_m \odot \mathcal{P}_f \text{ and } \mathcal{B}_m = (1 - \mathcal{P}_m) \odot \mathcal{B}$$

$$\mathcal{P} = \mathcal{P}_{f,m} + \mathcal{B}_m$$

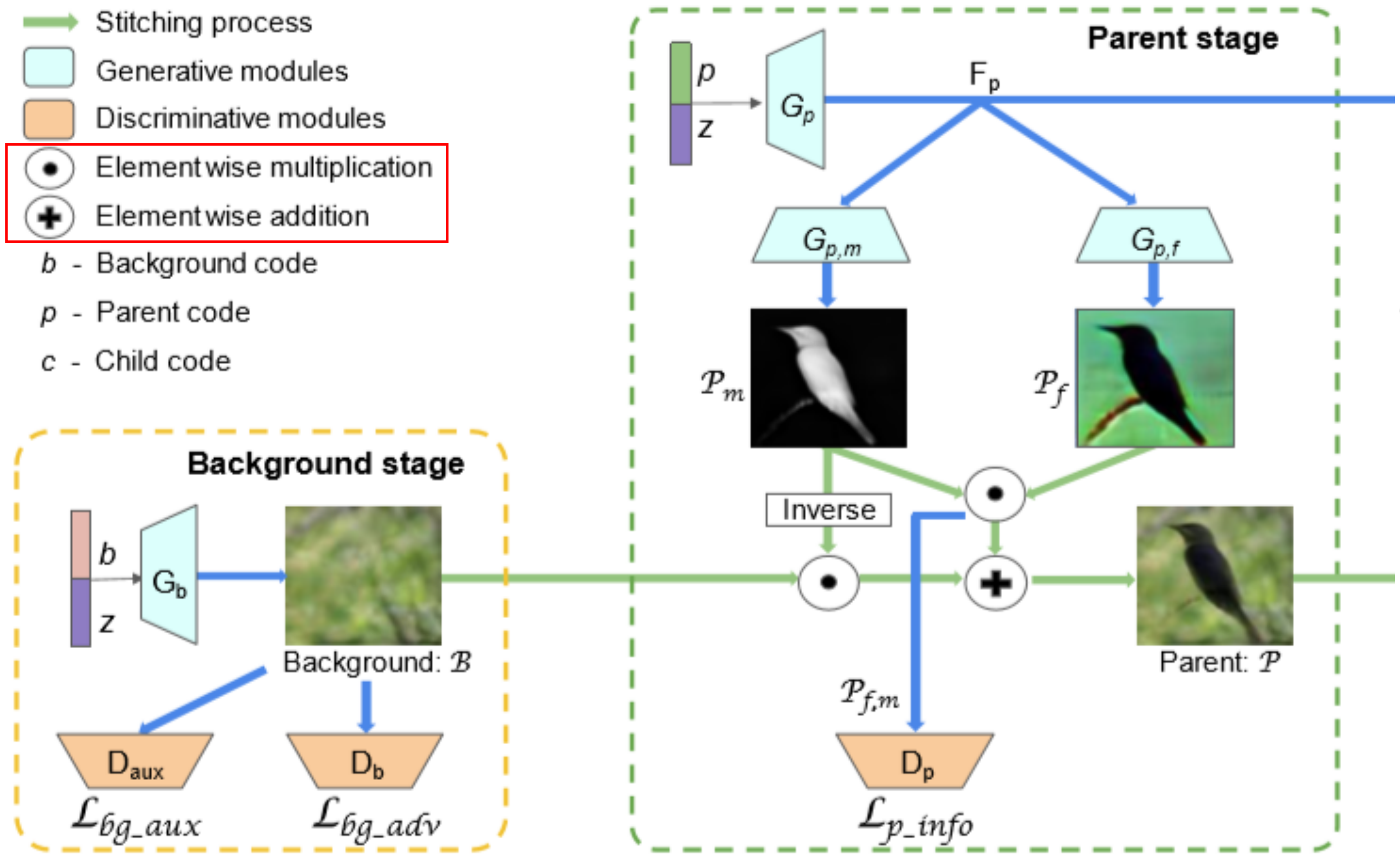
The true distribution for $\mathcal{P}_{f,m}$ or $\mathcal{P}$ is unknown

→ We cannot use the standard GAN objective to train D_p.

we maximize the mutual information $I(p, \mathcal{P}_{f,m})$, with D_p approximating the posterior $P(p|\mathcal{P}_{f,m})$:

$$\mathcal{L}_p = \mathcal{L}_{p\_info} = \max_{D_p, G_{p,f}, G_{p,m}} \mathbb{E}_{z,p}[\log D_p(p|\mathcal{P}_{f,m})]$$

Method



$\mathcal{P}_{f,m} = \mathcal{P}_m \odot \mathcal{P}_f$ and $\mathcal{B}_m = (1 - \mathcal{P}_m) \odot \mathcal{B}$

$\mathcal{P} = \mathcal{P}_{f,m} + \mathcal{B}_m$

The true distribution for $\mathcal{P}_{f,m}$ or $\mathcal{P}$ is unknown

→ We cannot use the standard GAN objective to train D_p.

we maximize the mutual information $I(p, \mathcal{P}_{f,m})$, with D_p approximating the posterior $P(p|\mathcal{P}_{f,m})$:

$$\mathcal{L}_p = \mathcal{L}_{p\_info} = \max_{D_p, G_{p,f}, G_{p,m}} \mathbb{E}_{z,p}[\log D_p(p|\mathcal{P}_{f,m})]$$

DAVIAN
Data and Visual Analytics Lab

Method

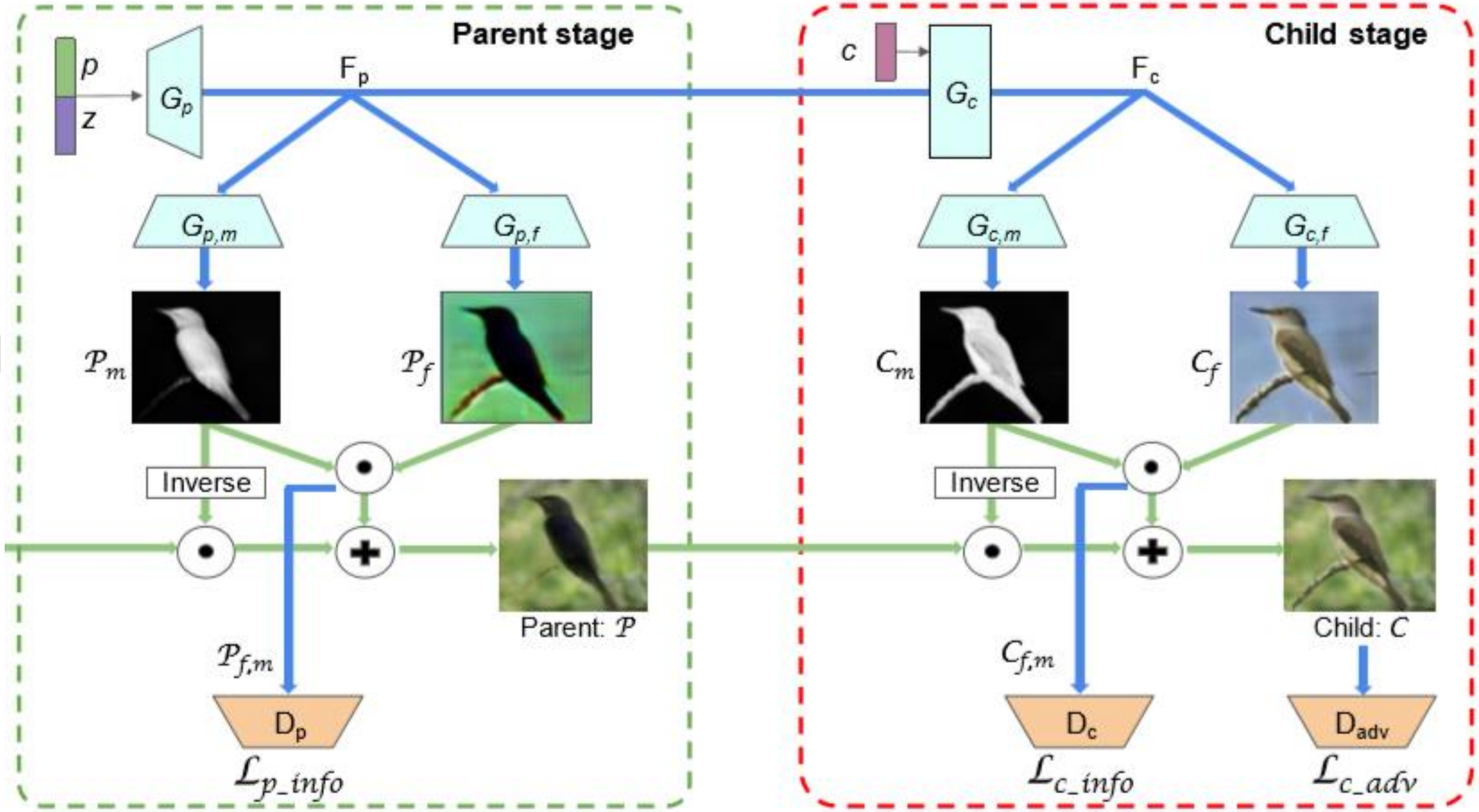
$$\mathcal{P}_{c,m} = (1 - \mathcal{C}_m) \odot \mathcal{P} \quad \mathcal{C}_{f,m} = \mathcal{C}_m \odot \mathcal{C}_f$$

$$\mathcal{C} = \mathcal{C}_{f,m} + \mathcal{P}_{c,m}$$

$$\mathcal{L}_{c\_info} = \max_{D_c, G_{c,f}, G_{c,m}} \mathbb{E}_{z,p,c}[\log D_c(c|\mathcal{C}_{f,m})]$$

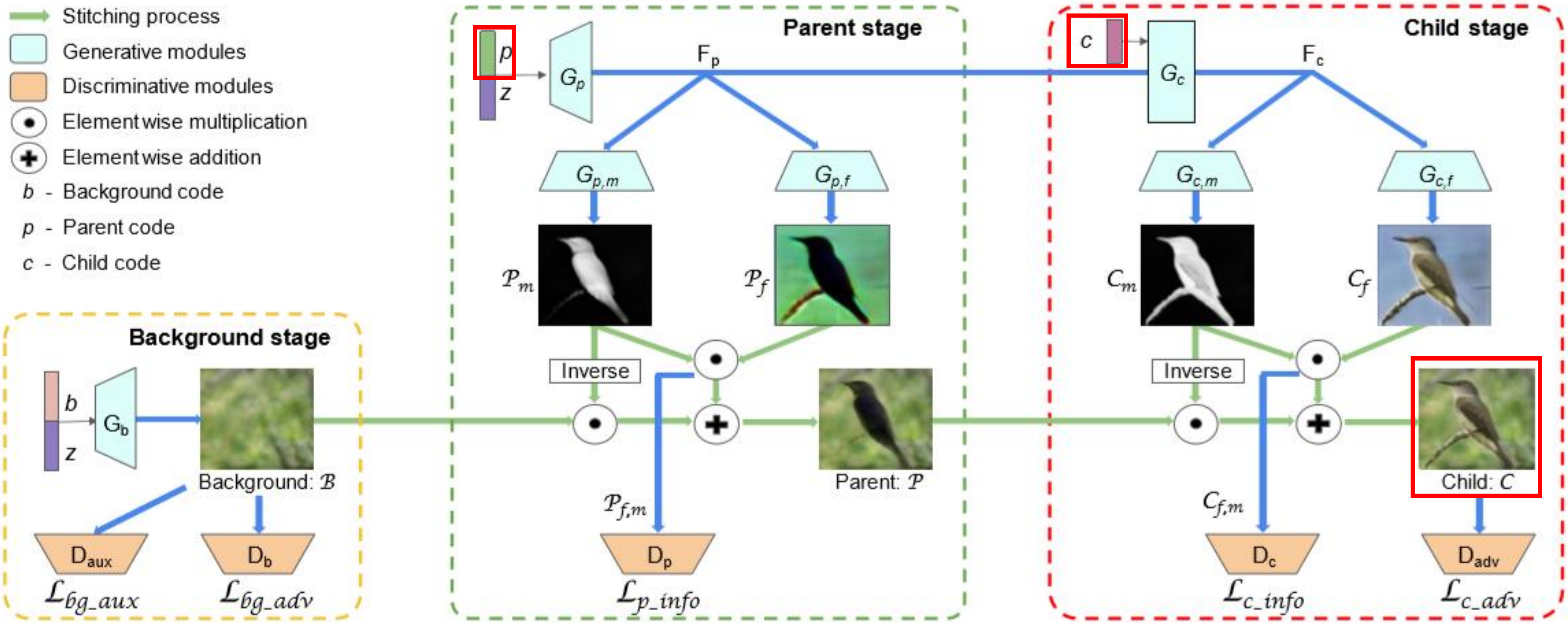
$$\mathcal{L}_{c\_adv} = \min_{G_c} \max_{D_{adv}} \mathbb{E}_x[\log(D_{adv}(x))] + \mathbb{E}_{z,b,p,c}[\log(1 - D_{adv}(\mathcal{C}))]$$

$$\mathcal{L}_c = \mathcal{L}_{c\_adv} + \mathcal{L}_{c\_info}$$



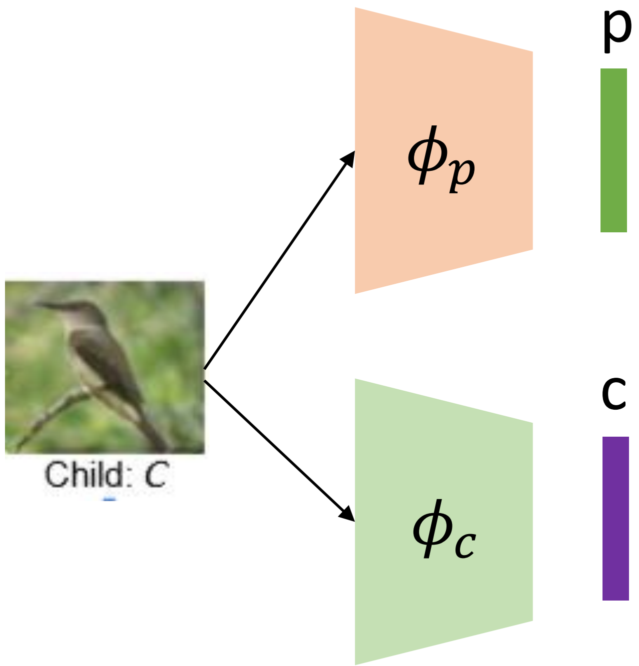
DAVIAN
Data and Visual Analytics Lab

Method

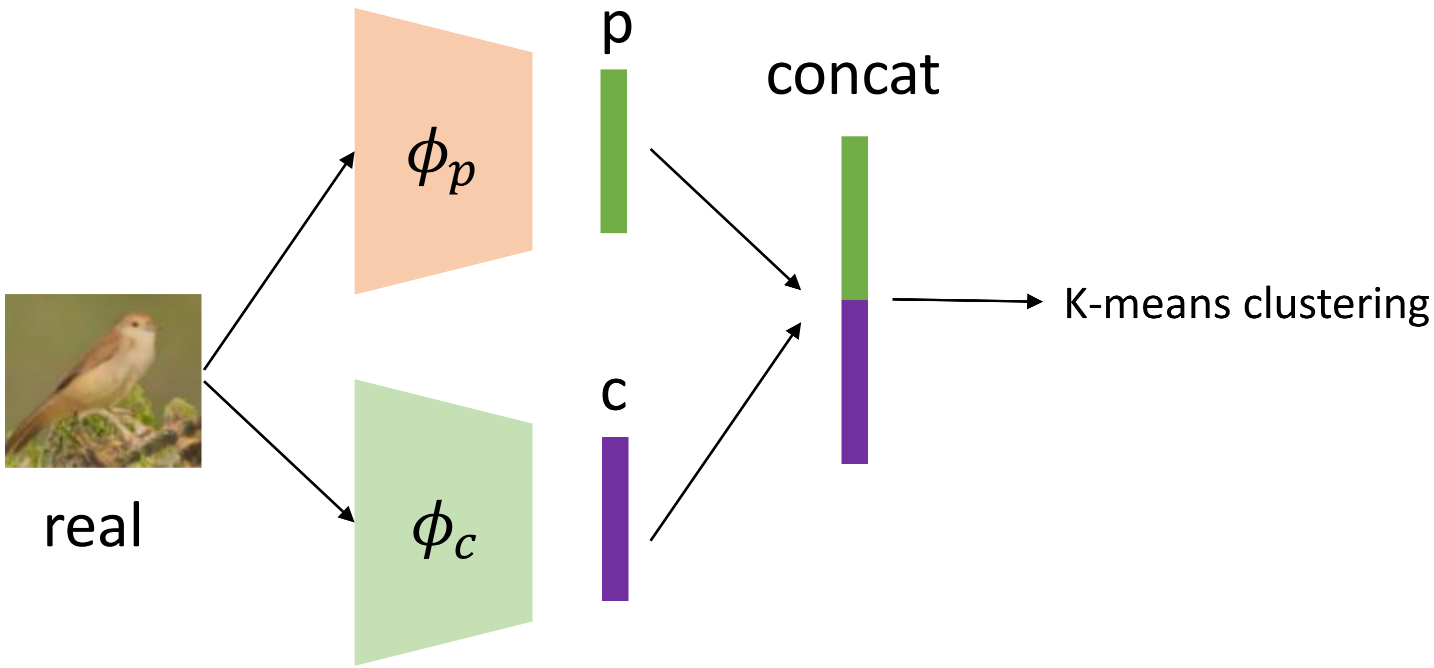

Stitching process
Generative modules
Discriminative modules
Elementwise multiplication
Elementwise addition
b - Background code
p - Parent code
c - Child code
Background stage
b
z
G_b
Background: $\mathcal{B}$
D_{aux}
D_b
$\mathcal{L}_{bg\_aux}$
$\mathcal{L}_{bg\_adv}$
Parent stage
p
z
G_p
F_p
$G_{p,m}$
$G_{p,f}$
$\mathcal{P}_m$
$\mathcal{P}_f$
Inverse
Parent: $\mathcal{P}$
$\mathcal{P}_{f,m}$
D_p
$\mathcal{L}_{p\_info}$
Child stage
c
G_c
F_c
$G_{c,m}$
$G_{c,f}$
$\mathcal{C}_m$
$\mathcal{C}_f$
Inverse
Child: $\mathcal{C}$
$\mathcal{C}_{f,m}$
D_c
D_{adv}
$\mathcal{L}_{c\_info}$
$\mathcal{L}_{c\_adv}$

Method

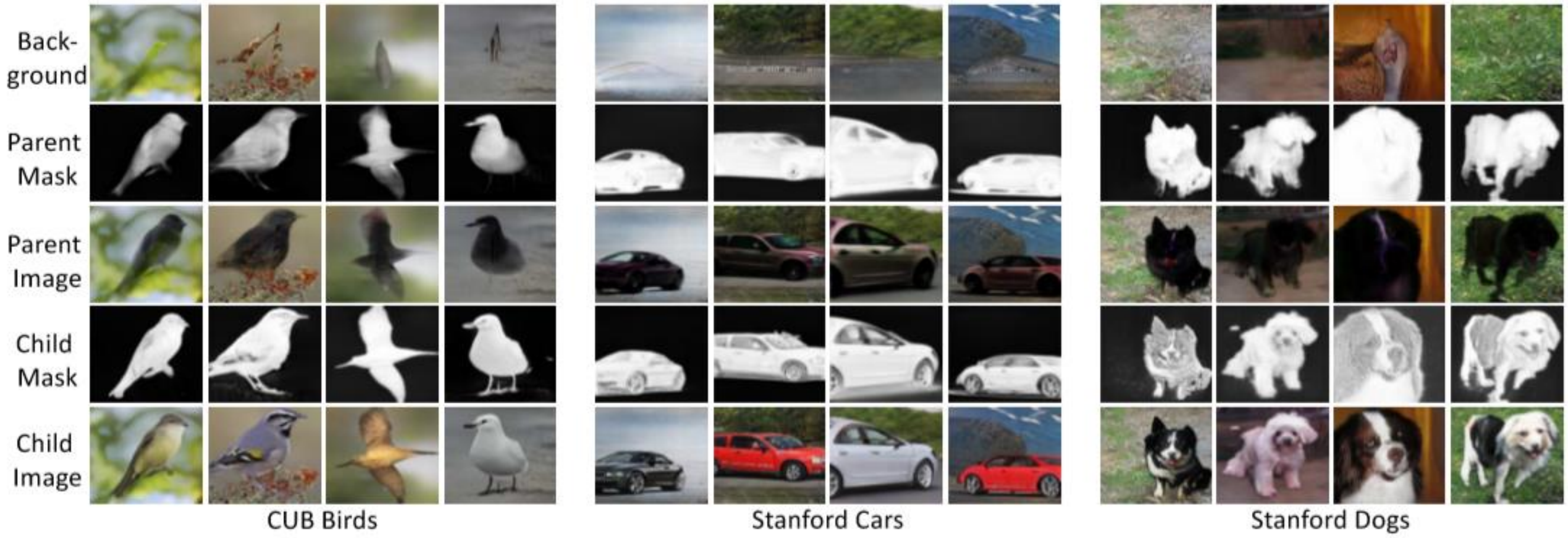
Train



Test



Experiments



Experiments

DAVIAN
Data and Visual Analytics Lab

Experiments

	IS			FID		
	Birds	Dogs	Cars	Birds	Dogs	Cars
Simple-GAN	31.85 ± 0.17	6.75 ± 0.07	20.92 ± 0.14	16.69	261.85	33.35
InfoGAN [9]	47.32 ± 0.77	43.16 ± 0.42	28.62 ± 0.44	13.20	29.34	17.63
LR-GAN [52]	13.50 ± 0.20	10.22 ± 0.21	5.25 ± 0.05	34.91	54.91	88.80
StackGANv2 [57]	43.47 ± 0.74	37.29 ± 0.56	**33.69 ± 0.44**	13.60	31.39	16.28
FineGAN (ours)	**52.53 ± 0.45**	**46.92 ± 0.61**	32.62 ± 0.37	**11.25**	**25.66**	**16.03**

	N_p=20	N_p=10	N_p=40	N_p=5	N_p=mixed
Inception Score (CUB)	**52.53**	52.11	49.62	46.68	51.83

N_c = 200 / (6, 5), (3, 20), (11, 10)

	NMI			Accuracy		
	Birds	Dogs	Cars	Birds	Dogs	Cars
JULE [53]	0.204	0.142	0.232	0.045	0.043	0.046
JULE-ResNet-50 [53]	0.203	0.148	0.237	0.044	0.044	0.050
DEPICT [15]	0.290	0.182	0.329	0.061	0.052	0.063
DEPICT-Large [15]	0.297	0.183	0.330	0.061	0.054	0.062
Ours	**0.403**	**0.233**	**0.354**	**0.126**	**0.079**	**0.078**

DAVIAN
Data and Visual Analytics Lab

감사합니다.

참고 : https://www.youtube.com/watch?v=_4jbgniqt_Q&t=948s (PR-22 InfoGAN, 차준범님 강의)