Gradient Matching Generative Networks for Zero-Shot Learning

Mert Bulent Sariyildiz
Bilkent University
Department of Computer Engineering
mert.sariyildiz@bilkent.edu.tr

Ramazan Gokberk Cinbis
Middle East Technical University (METU)
Department of Computer Engineering
gcinbis@metu.edu.tr

CVPR 2019 (Oral)

2019.10.15

발표자 : 김용규

Introduction

Zero-shot learning (ZSL)

- Seen classes 로부터 학습한 classification 모델을 가지고 unseen classes에 대해서 추측

Introduction

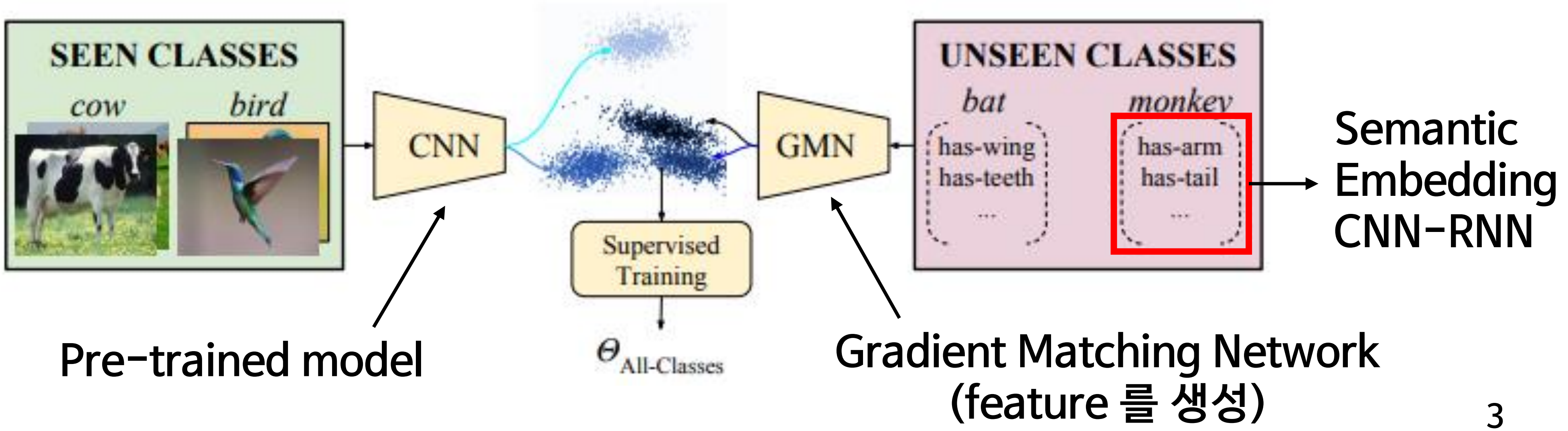
Discriminative model 문제점

- 분포 차이로 인해 domain shift 문제가 발생 -> 정확도가 떨어짐

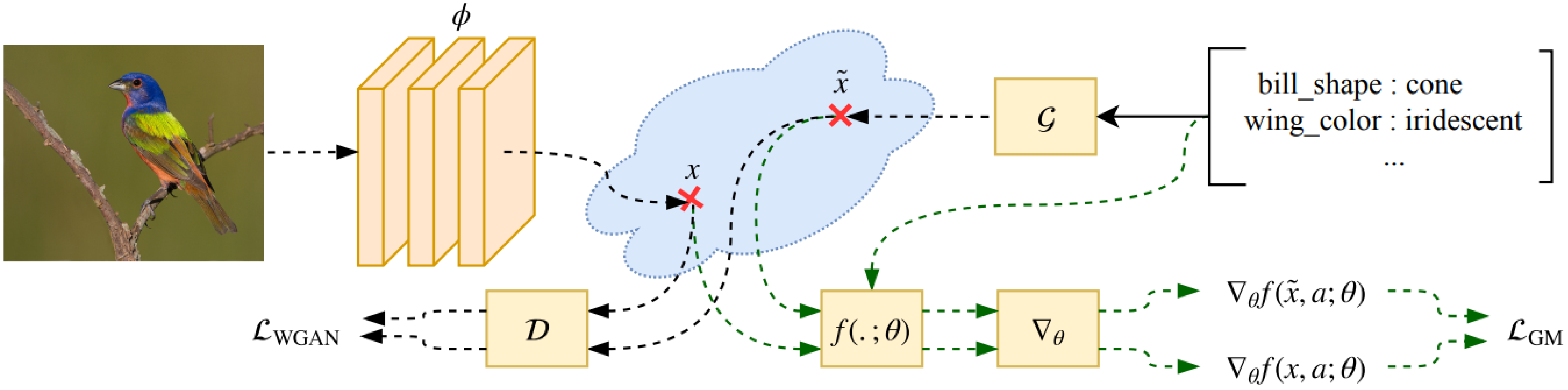
해결책

- Unseen lass label 정보를 통해 feature를 생성 : Unsupervised learning -> Supervised learning

목적 : label data (semantic embedding)를 통해 유의미한 feature 생성 -> good classification accuracy

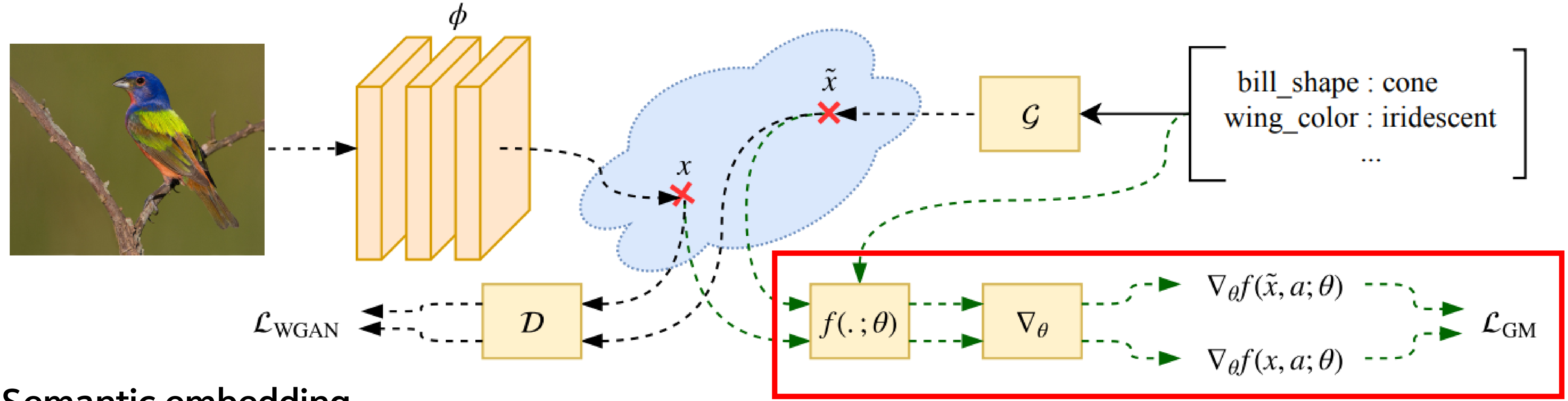


Method

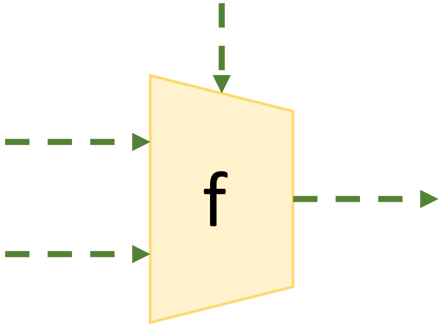


- Seen data & Unseen data // Train data & Test data
- WGAN Loss
- Gradient Matching Loss

Method



Semantic embedding



The degree of confidence

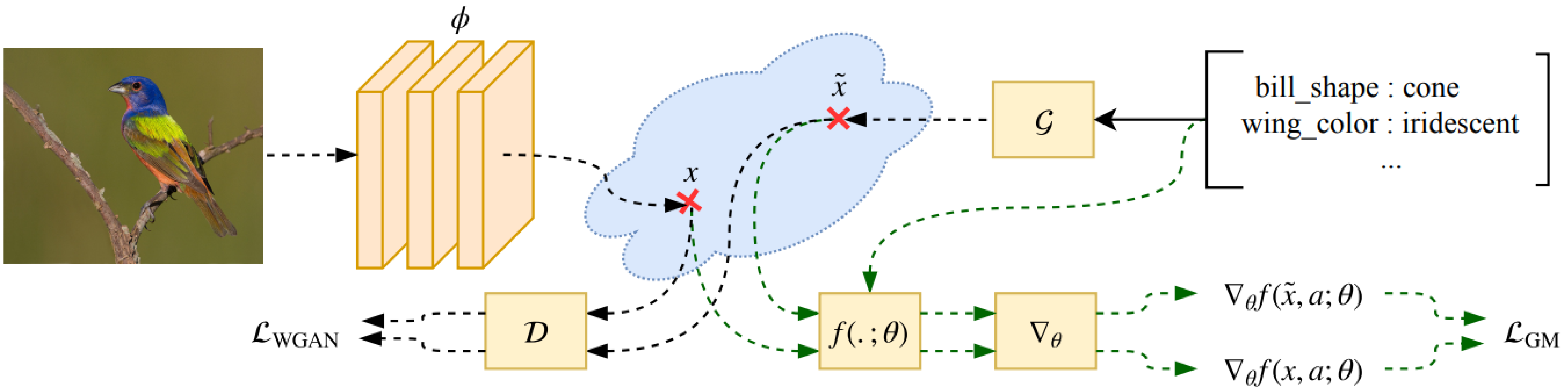
$$f(x, a; W, b) = x^{\mathrm{T}} W a + b$$

GT : Feature x 의 class에 해당하는 confidence score

Cross Entropy loss..?

Gradient Matching Loss

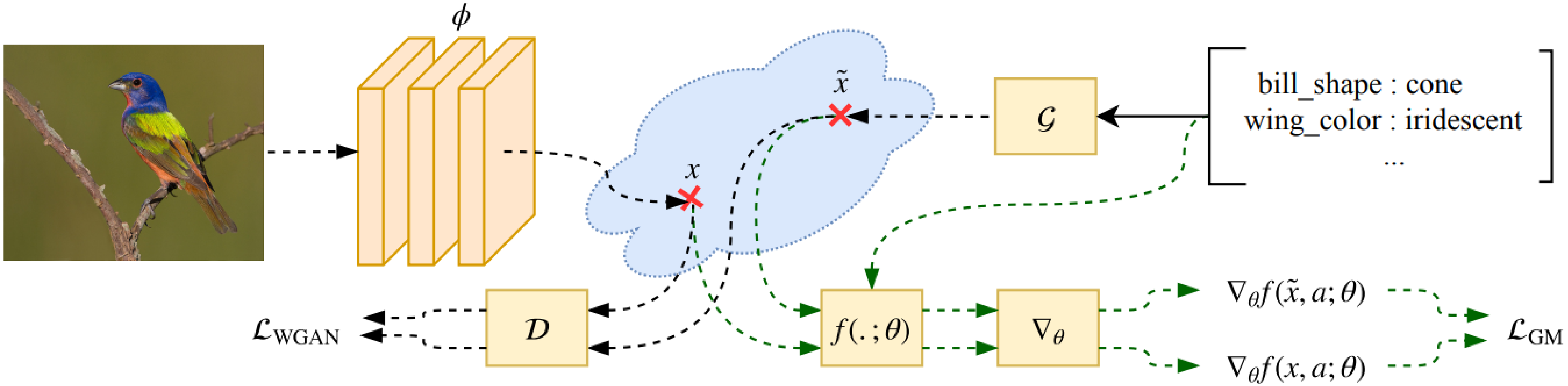
Method



Gradient Matching Loss

- G가 true class manifold를 잘 학습한 경우, classification model parameter에 대한 loss function의 partial derivative 사이에 correlation이 높다.
- G가 true class manifold를 잘 학습한 것과 두 partial derivative 사이에 연관이 있다.
- 따라서, x 의 partial derivative를 GT로 사용해서 같아지도록 학습하자.

Method



Gradient Matching Loss

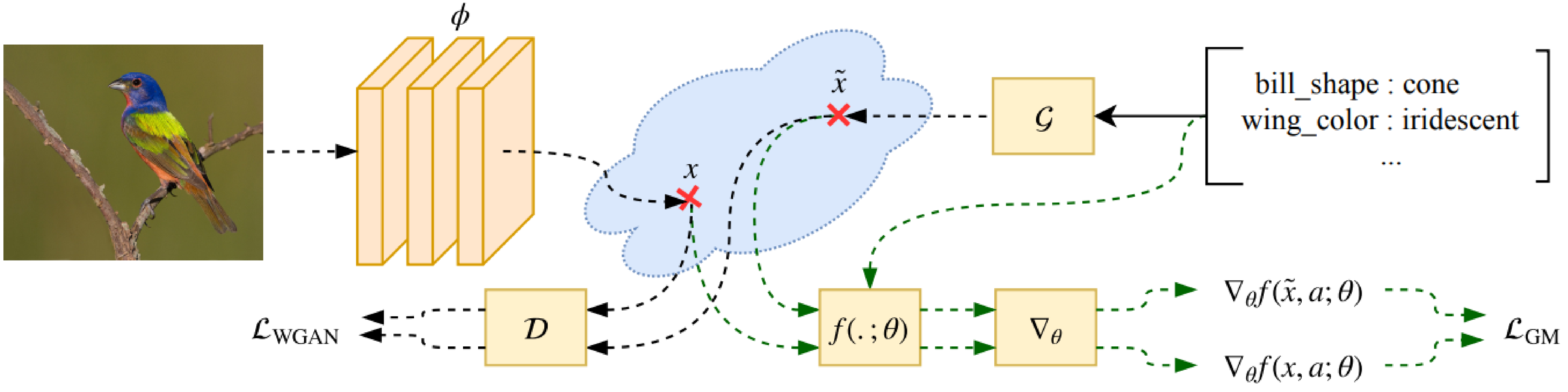
$$g_{\mathrm{r}}(\theta) = \mathop{\mathbb{E}}_{(x,a)\sim\mathcal{D}_{\mathrm{s}}} \left[\nabla_{\theta_f} \mathcal{L}_{\mathrm{CLS}}(f, x, a; \theta_f = \theta)\right],$$

$$g_{\mathrm{s}}(\theta) = \mathop{\mathbb{E}}_{\tilde{x}\sim\mathcal{G}(a\sim\mathcal{A}_{\mathrm{s}})} \left[\nabla_{\theta_f} \mathcal{L}_{\mathrm{CLS}}(f, \tilde{x}, a; \theta_f = \theta)\right]$$

$$\mathcal{L}_{\mathrm{GM}} = \mathop{\mathbb{E}}_{\theta} \left[1 - \frac{g_{\mathrm{r}}(\theta)^{\mathrm{T}} g_{\mathrm{s}}(\theta)}{\|g_{\mathrm{r}}(\theta)\|_2 \|g_{\mathrm{s}}(\theta)\|_2}\right]$$

Local minimum으로 향하지 않도록 하기 위해서
absolute scale 보다 direction이 더 중요

Method



$$\theta_{\mathcal{G}}^*, \theta_{\mathcal{D}}^* = \arg \min_{\theta_{\mathcal{G}}, \theta_{\mathcal{D}}} \{\mathcal{L}_{\text{WGAN}} + \beta \mathcal{L}_{\text{GM}}\} + \boldsymbol{L_{CLS}}$$

$$\mathcal{L}_{\text{WGAN}}^{\text{S}} = \mathop{\mathbb{E}}_{\tilde{x} \sim \mathcal{G}(a \sim \mathcal{A}_{\text{s}})} [\mathcal{D}(\tilde{x})] - \mathop{\mathbb{E}}_{x \sim \mathcal{X}_{\text{s}}} [\mathcal{D}(x)] + \lambda \mathcal{L}_{\text{GP}}$$

$$\mathcal{L}_{\text{WGAN}}^{\text{S+U}} = \mathop{\mathbb{E}}_{\tilde{x} \sim \mathcal{G}(a \sim \mathcal{A}_{\text{all}})} [\mathcal{D}(\tilde{x})] - \mathop{\mathbb{E}}_{x \sim \mathcal{X}_{\text{all}}} [\mathcal{D}(x)] + \lambda \mathcal{L}_{\text{GP}}$$

$$\mathcal{L}_{\text{cWGAN}}^{\text{S}} = \mathbb{E}\left[\mathcal{D}(x, a)\right] - \mathbb{E}\left[\mathcal{D}(\tilde{x}, a)\right] + \lambda \, \mathbb{E}\left[(\|\nabla_{\hat{x}} \mathcal{D}(\hat{x}, a)\|_2 - 1)^2\right].$$

$$(\mathcal{L}_{\text{WGAN}}^{\text{S}}, \mathcal{L}_{\text{WGAN}}^{\text{S+U}}, \mathcal{L}_{\text{cWGAN}}^{\text{S}})$$

Experiments

ZSL : Unseen class 비교
GZSL : Unseen class + Seen class 비교

u : unseen class data 비교
s : seen class data 비교

	Zero-Shot Learning			Generalized Zero-Shot Learning								
	CUB	SUN	AWA	CUB			SUN			AWA		
Method	T-1	T-1	T-1	u	s	h	u	s	h	u	s	h
Train only with real samples ($\mathcal{D}_s$)	56.8	60.7	62.3	26.9	67.6	38.4	23.4	36.3	28.4	13.4	78.1	22.9
$\mathcal{L}^{S}_{WGAN} + \mathcal{L}_{CLS}$	58.3	61.4	70.0	47.0	71.0	56.5	47.7	**41.2**	44.2	47.8	78.7	59.5
$\mathcal{L}^{S}_{cWGAN}$	60.6	62.6	72.0	55.9	71.1	62.6	53.6	41.1	46.5	55.2	79.1	65.0
$\mathcal{L}^{S}_{WGAN} + \mathcal{L}_{GM}$	61.9	63.8	70.4	55.8	70.7	62.4	53.8	40.9	46.5	52.1	78.8	62.7
$\mathcal{L}^{S}_{cWGAN} + \mathcal{L}_{GM}$	**64.6**	64.1	73.9	57.9	**71.2**	63.9	55.2	40.8	46.9	63.2	78.8	70.1
$\mathcal{L}^{S+U}_{WGAN} + \mathcal{L}_{GM}$ (transductive)	**64.6**	**64.3**	**82.5**	**60.2**	70.6	**65.0**	**57.1**	40.7	**47.5**	**70.8**	**79.2**	**74.8**

Experiments

	Zero-Shot Learning			Generalized Zero-Shot Learning								
	CUB	**SUN**	**AWA**	**CUB**			**SUN**			**AWA**		
Method	**T-1**	**T-1**	**T-1**	**u**	**s**	**h**	**u**	**s**	**h**	**u**	**s**	**h**
Zhang et al. [46] '18	52.6	61.7	67.4	31.5	40.2	35.3	41.2	26.7	32.4	38.7	74.6	51.0
Bucher et al. [25] '17	57.8	60.4	66.3	28.8	55.7	38.0	40.5	37.2	38.8	2.3	**90.2**	4.5
Xian et al. [26] - `DEVISE` '18	60.3	60.9	66.9	52.2	42.4	46.7	38.4	25.4	30.6	35.0	62.8	45.0
Xian et al. [26] - `ALE` '18	61.5	62.1	68.2	40.2	59.3	47.9	41.3	31.1	35.5	47.6	57.2	52.0
Verma et al. [28] '18	59.6	63.4	69.5	41.5	53.3	46.7	40.9	30.5	34.9	56.3	67.8	61.5
Felix et al. [27] - `cycle-WGAN` '18	57.8	59.7	65.6	46.0	60.3	52.2	48.3	33.1	39.2	56.4	63.5	59.7
Felix et al. [27] - `cycle-CLSWGAN` '18	58.4	60.0	66.3	45.7	**61.0**	52.3	49.4	**33.6**	40.0	56.9	64.0	60.2
$\mathcal{L}^{S}_{\text{cWGAN}} + \mathcal{L}_{\text{CLS}}$ [26]	62.2	62.7	69.4	51.1	54.9	52.9	50.6	30.3	37.3	57.5	66.8	61.8
$\mathcal{L}^{S}_{\text{cWGAN}} + \mathcal{L}_{\text{GM}}$ *(Ours)*	**64.3**	**63.6**	**71.9**	**56.1**	54.3	**55.2**	**53.2**	33.0	**40.7**	**61.1**	71.3	**65.8**
$\mathcal{L}^{S}_{\text{cWGAN}} + \mathcal{L}_{\text{GM}}$ ‡ *(Ours)*	64.6	64.1	73.9	57.9	71.2	63.9	55.2	40.8	46.9	63.2	78.8	70.1