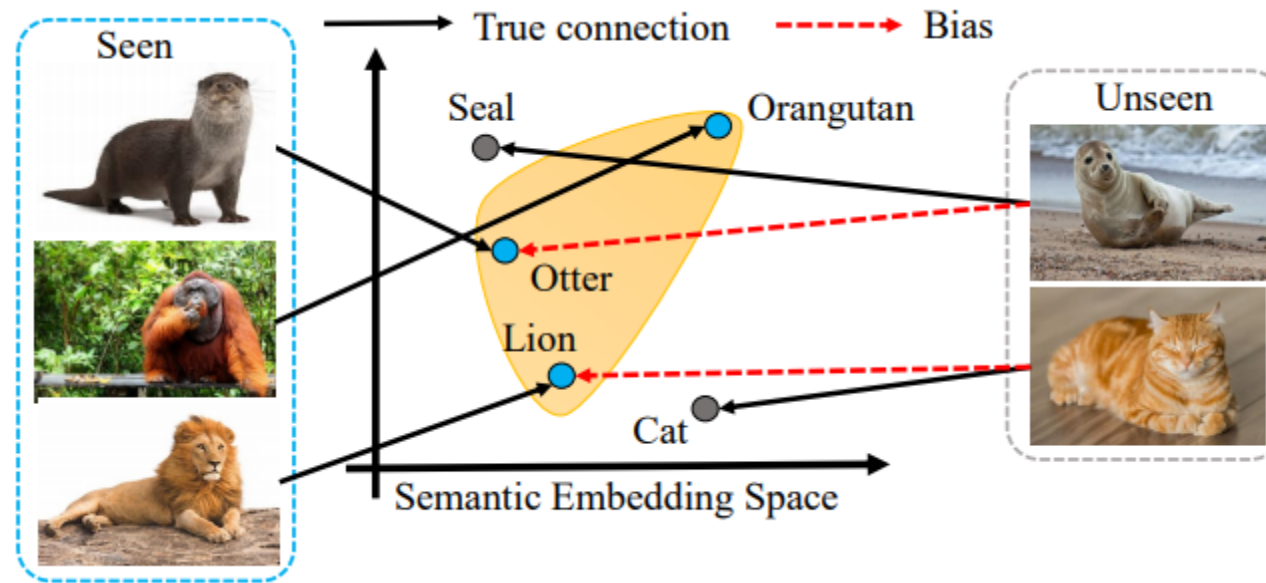


Towards Recognizing Unseen Categories in Unseen Domains

Mancini, M., Akata, Z., Ricci, E., & Caputo, B. (2020). Towards recognizing unseen categories in unseen domains. In Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16 (pp. 466-483). Springer International Publishing.

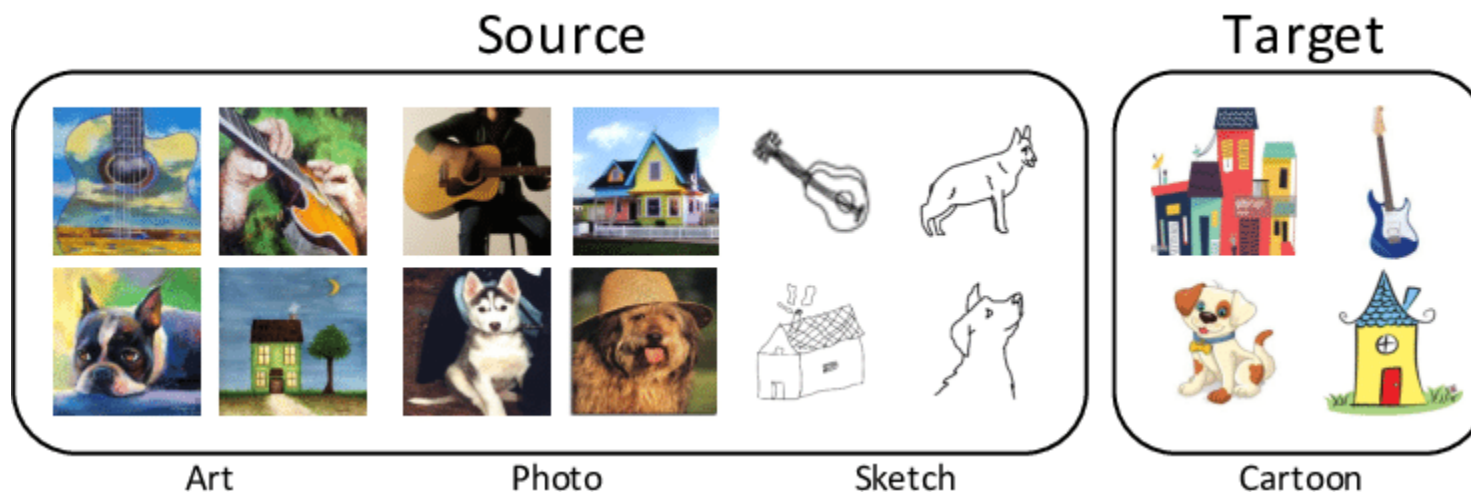
Introduction

- Zero-shot learning : unseen data를 입력 받아도, seen data로 학습된 지식을 전이하여 unseen data를 unseen class로 예측



Introduction

- Zero-shot learning : unseen data를 입력 받아도, seen data로 학습된 지식을 전이하여 unseen data를 unseen class로 예측
- Domain Generalization : train과 test의 distribution shift (domain shift)를 완화할 수 있는 generalize된 기법을 고안하는 분야



Introduction

- Zero-shot learning : unseen data를 입력 받아도, seen data로 학습된 지식을 전이하여 unseen data를 unseen class로 예측
- Domain Generalization : train과 test의 distribution shift (domain shift)를 완화할 수 있는 generalize된 기법을 고안하는 분야

- ZSL + DG가 문제를 **동시에** 해결할 순 없을까!?
=> Mixup을 사용해 semantic & domain shift 이 동시에 되며
구분이 되도록 학습을 진행하자!

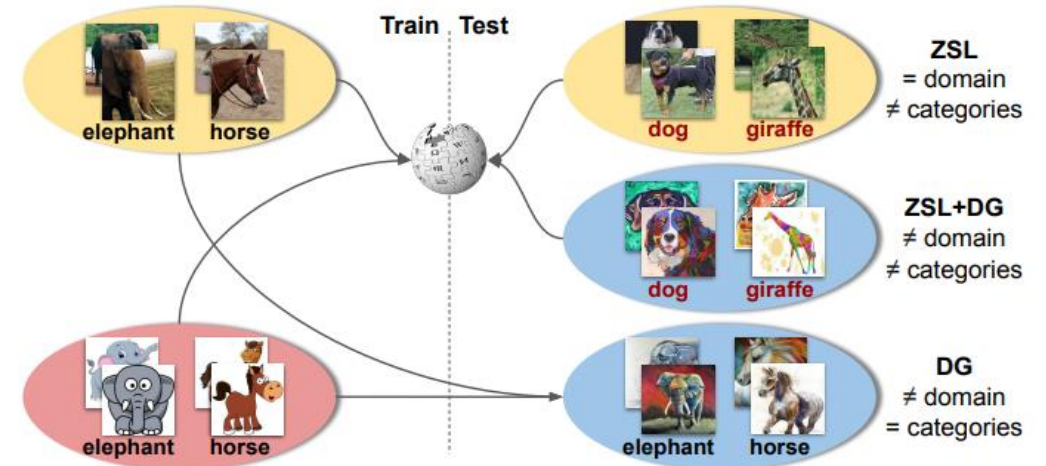


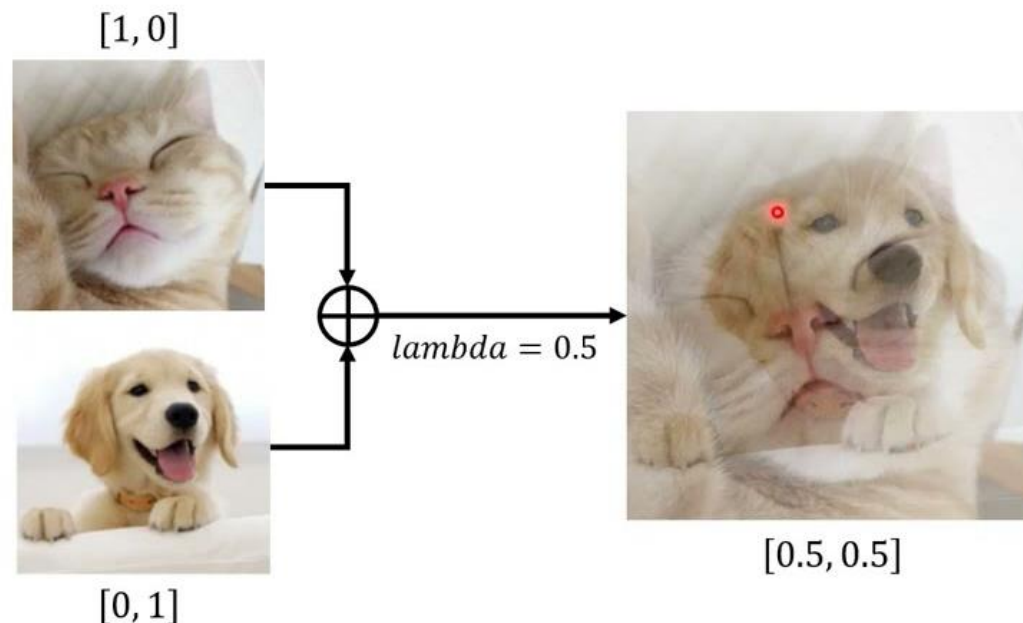
Fig. 1. Our ZSL+DG problem. During training we have images of multiple categories (e.g. elephant, horse) and domains (e.g. photo, cartoon). At test time, we want to recognize unseen categories (e.g. dog, giraffe), as in ZSL, in unseen domains (e.g. paintings), as in DG, exploiting side information describing seen and unseen categories.

Mixup^[1]

- Mixup은 17년 제안된 Vicinal Risk Minimization(VRM) 기반의 학습(훈련 데이터셋의 근방(vicinal) 분포도 함께 활용)
- 근접한 정도에 대한 분포를 beta distribution(확률에 대한 확률)으로 하며, 가지고 있는 training dataset에 대한 분포에서 어디쯤에 위치할지를 lambda 값으로 조정하면서 결정하는 형태
- Mixup을 통해 두 클래스 간의 decision boundary가 유연해짐 => 과적합이 덜 발생

$$\begin{aligned}\hat{x} &= \lambda x_i + (1 - \lambda)x_j \\ \hat{y} &= \lambda y_i + (1 - \lambda)y_j\end{aligned}$$

$\lambda \in [0, 1]$ 는 $Beta(\alpha, \alpha)$ 에서 추출합니다.



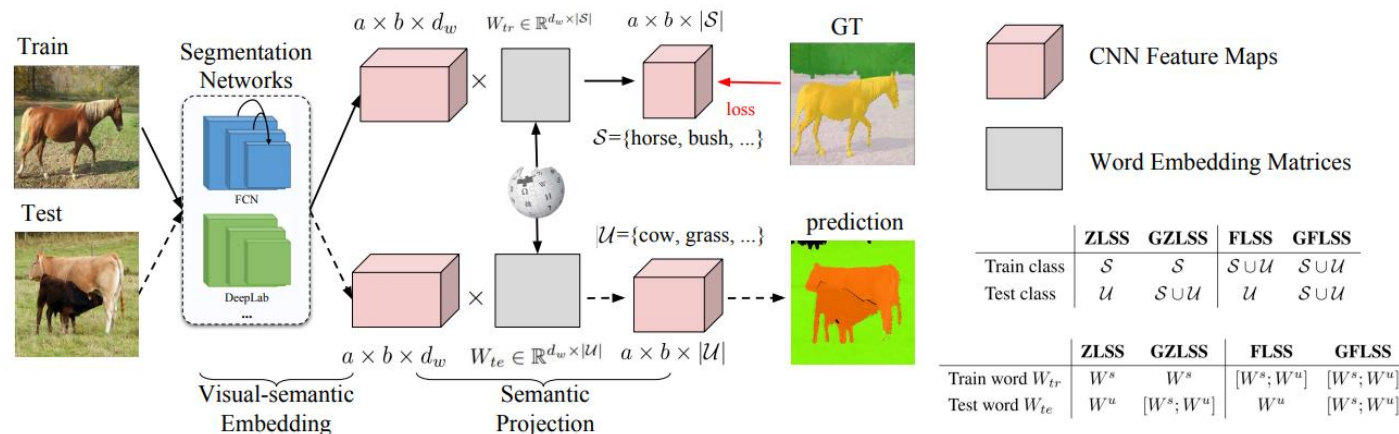
Semantic projection layer^[2]

- SPNet에서 처음 도입되었으며 feature vector와 word embedding 사이의 similarity를 계산하여 class를 예측

$$p(\hat{y}_{ij} = s | x; W^s) = \frac{\exp(w_s^\top \phi(x)_{ij})}{\sum_{c \in \mathcal{S}} \exp(w_c^\top \phi(x)_{ij})} \quad (1) \quad (\Phi : \text{feature embedding} / i, j : \text{픽셀} / W_c : W_s \text{ normalized to have unit length})$$

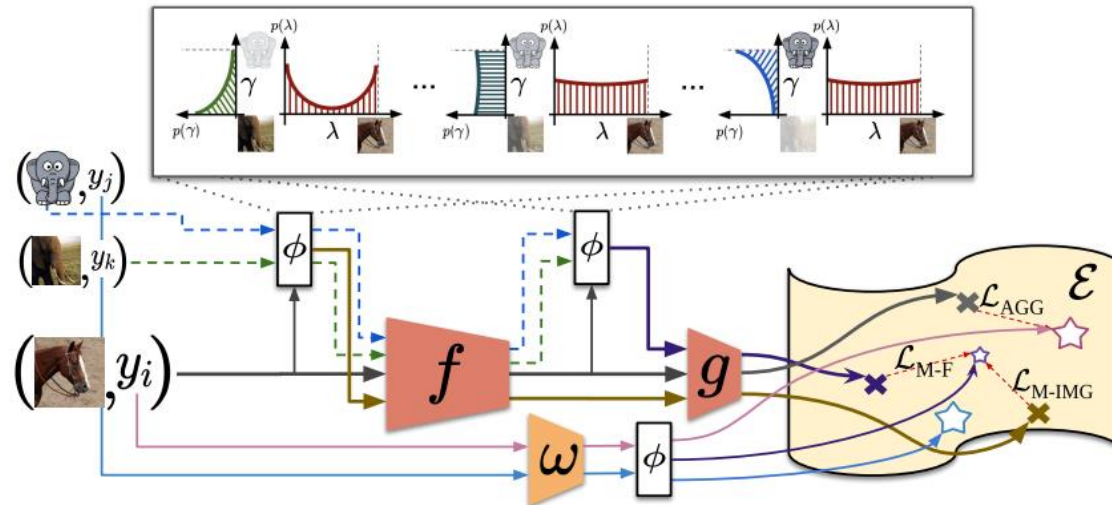
- seen class에 대한 bias가 생기기에 inference시에 calibration factor를 추가

$$\operatorname{argmax}_{u \in \mathcal{S} \cup \mathcal{U}} p(\hat{y}_{ij} = u | x; [W^s; W^u]) - \gamma \mathbb{I}[u \in \mathcal{S}] \quad (\gamma : \text{calibration factor})$$



CuMix - Method

- 본 논문은 Unseen Domain에서 Unseen Data를 구별하는 방법을 최초로 제안
- Mixup 기법을 통해 feature & image level에서 다양한 domain과 class들을 혼합하여 학습
- 이때, domain 영역과 class에 대한 영역을 구분하고 나누기 위해 학습 과정에서 다양한 loss 함수를 적용



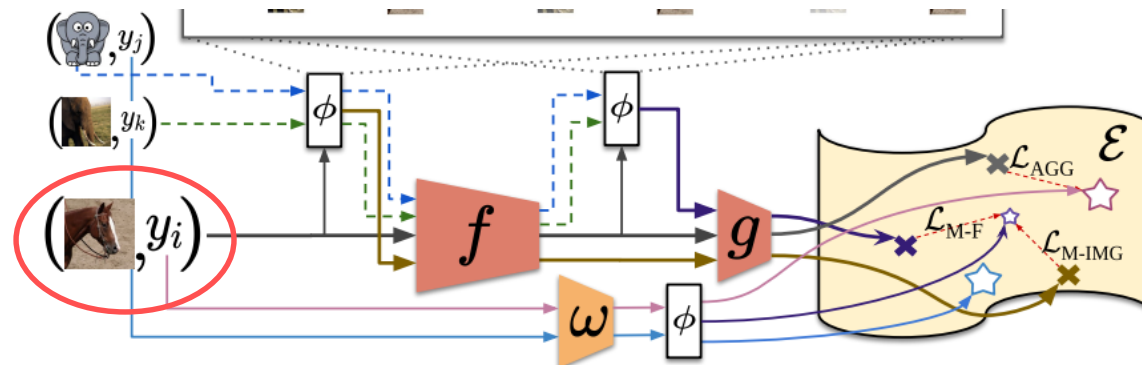
CuMix - Predict

- semantic projection layer를 통해 이미지 x 의 label y^* 예측

$$y^* = \operatorname{argmax}_y \omega(y)^\top g(f(x)). \quad f: \text{img mapping 함수} / g: \text{feature semantic projection embedding} / \omega: \text{DG 처리 임베딩}$$

- Loss 함수는 cross Entropy 사용

$$\mathcal{L}_{\text{AGG}}(x_i, y_i) = \sum_{y \in \mathcal{Y}^s} \ell(\omega(y)^\top g(f(x_i)), y_i)$$



CuMix – Train Dataset Argument

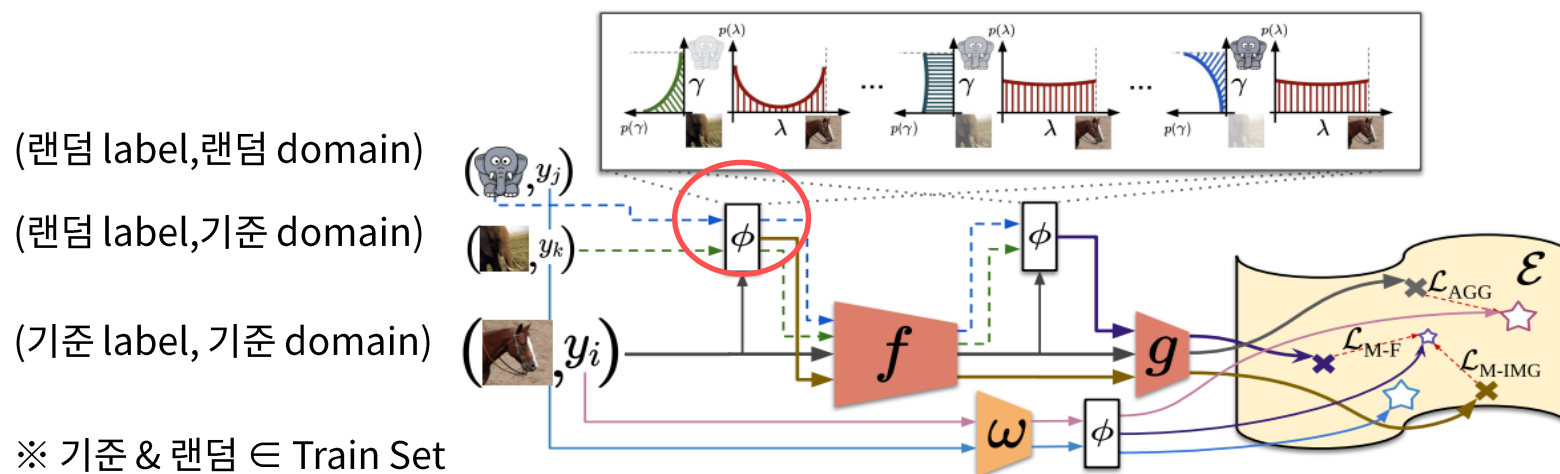
- 본 논문에서는 triplet Mixup 방식을 통해 다양한 형태의 domain & class를 미리 시뮬레이션

$$\phi(a_i, a_j, a_k) = \lambda a_i + (1 - \lambda)(\gamma a_j + (1 - \gamma)a_k) \quad \text{r: 베르누이 분포 (기존 도메인 or 랜덤 도메인)}$$

- 처음부터 막 섞으면 학습이 제대로 진행되지 않기에 다음과 같은 조건을 통해 람다와 감마 분포 결정

- $s < N$: $\alpha = 0, \beta = s / N * \beta_{\max}$ (싱글 도메인에 대해서 분류 학습 진행)
- $N \leq s < 2N$: $\alpha = s - N / N, \beta = \beta_{\max}$ (동일한 도메인끼리 무작위로 mixup해 unseen class를 위한 class 분포 학습)
- $s \geq 2N$: $\alpha = 1$ (서로 다른 도메인의 샘플을 mixup해 unseen domain에 대한 domain shift 학습)

※ N : warm-up step epoch / s : 현재 epoch / α (감마 조절 함수) / β (람다 조절 함수, 베타 분포(하이퍼파라미터))



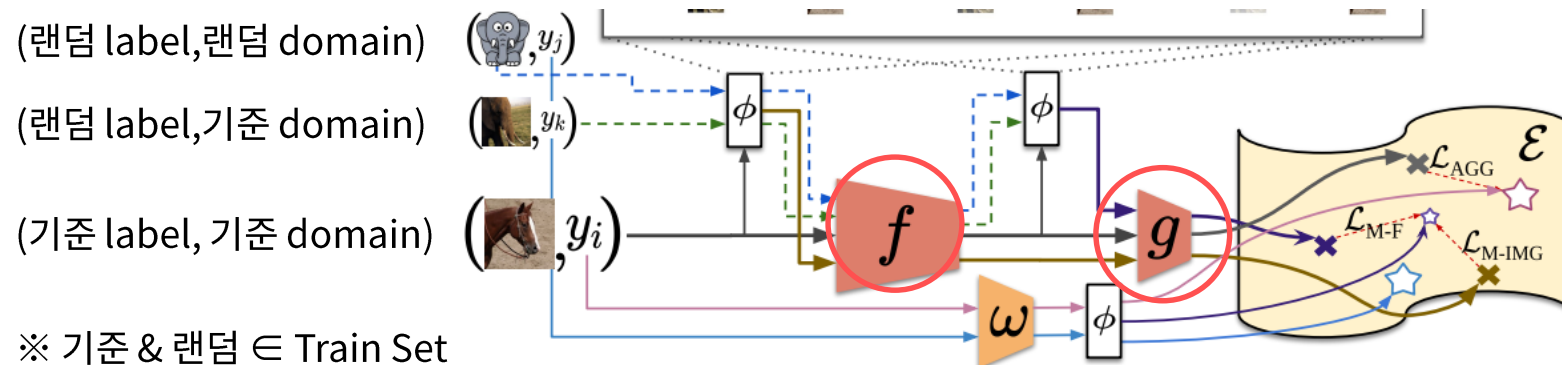
CuMix – Train Semantic & Domain-shift

- 본 논문에서는 triplet Mixup 방식을 통해 다양한 형태의 domain & class를 미리 시뮬레이션

$$\phi(a_i, a_j, a_k) = \lambda a_i + (1 - \lambda)(\gamma a_j + (1 - \gamma)a_k) \quad \text{r: 베르누이 분포 (기존 도메인 or 랜덤 도메인)}$$

- Random Class 혹은 Domain을 구현한 data와 기존 data를 Mixup한 이미지의 feature를 embedding

$$\mathcal{L}_{\text{M-IMG}}(x_i, y_i, d_i) = \mathcal{L}_{\text{AGG}}(\phi(x_i, x_j, x_k), \phi(\bar{y}_i, \bar{y}_j, \bar{y}_k)).$$



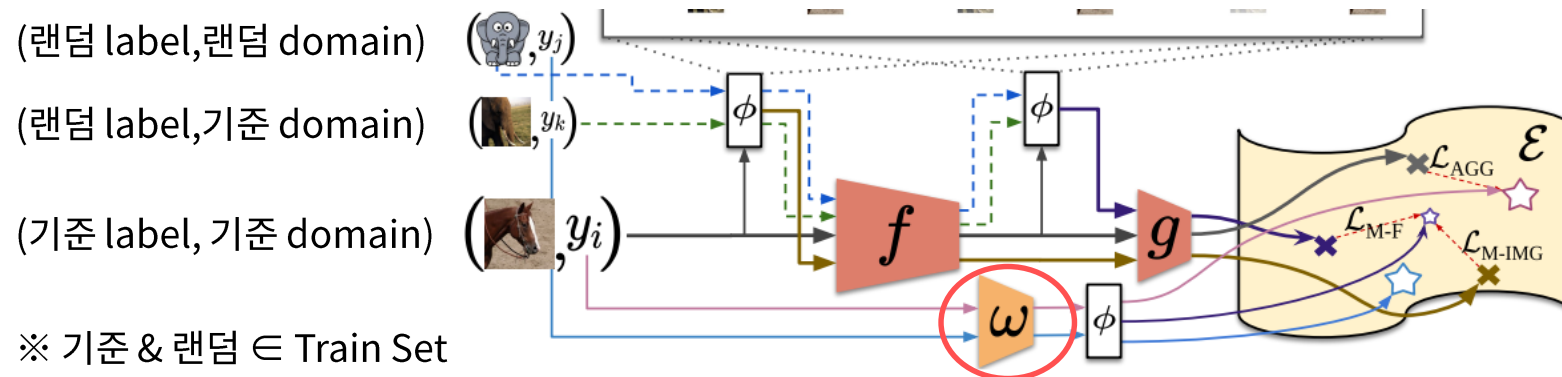
CuMix – Train Unseen Data Classification

- 본 논문에서는 triplet Mixup 방식을 통해 다양한 형태의 domain & class를 미리 시뮬레이션

$$\phi(a_i, a_j, a_k) = \lambda a_i + (1 - \lambda)(\gamma a_j + (1 - \gamma)a_k) \quad \text{r: 베르누이 분포 (기존 도메인 or 랜덤 도메인)}$$

- 기준과 랜덤 각각에 대한 feature를 뽑고 이를 Mixup 함으로써 각각에 대한 분포를 학습

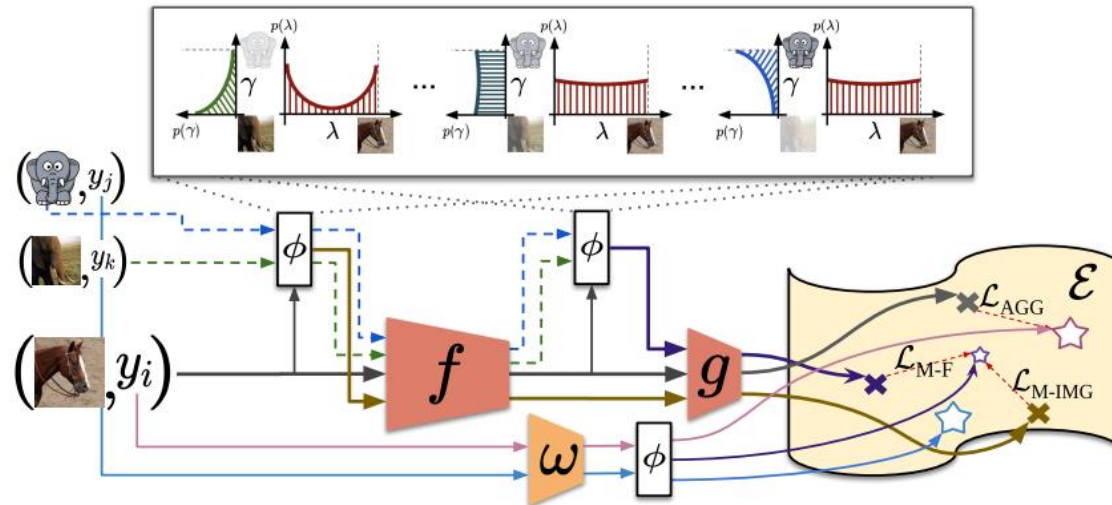
$$\mathcal{L}_{M-F}(x_i, y_i, d_i) = \sum_{y \in \mathcal{Y}^s} \ell \left(\omega(y)^\top g(\phi(f(x_i), f(x_j), f(x_k))), \phi(\bar{y}_i, \bar{y}_j, \bar{y}_k) \right)$$



CuMix – Train Model

- 최종적으로 CuMix 방식은 다음과 같은 loss 식을 이용해 학습 수행

$$\mathcal{L}_{\text{CuMIX}}(\mathcal{S}) = |\mathcal{S}|^{-1} \sum_{(x_i, y_i, d_i) \in \mathcal{S}} \mathcal{L}_{\text{AGG}}(x_i, y_i) + \eta_{\text{I}} \mathcal{L}_{\text{M-IMG}}(x_i, y_i, d_i) + \eta_{\text{F}} \mathcal{L}_{\text{M-F}}(x_i, y_i, d_i) \quad \eta : \text{loss에 대한 weight}$$



CuMix – Experiment

- f : Resnet-18 / ω : fully-connected classifier / setting $\beta_{\max} = 0.6$ / $\eta_I = 0.1$ / $\eta_F = 3$ / $N = 10$
- 4개 중 3개 source domain, 1개 target domain
- DG를 했을 경우 평균적으로 CuMix가 가장 좋았다

Table 1. Domain Generalization accuracies on PACS with ResNet-18.

Target	AGG	DANN [11]	MLDG [20]	CrossGrad [39]	MetaReg [3]	JiGen [5]	Epi-FCR [21]	CuMix
Photo	94.9	94.0	94.3	94.0	94.3	96.0	93.9	95.1
Art	76.1	81.3	79.5	78.7	79.5	79.4	82.1	82.3
Cartoon	73.8	73.8	77.3	73.3	75.4	75.3	77.0	76.5
Sketch	69.4	74.3	71.5	65.1	72.2	71.4	73.0	72.6
Average	78.5	80.8	80.7	80.7	77.8	80.4	81.5	81.6

CuMix – Experiment

- f : Resnet-18 / ω : fully-connected classifier / setting $\beta_{\max} = 0.6$ / $\eta_I = 0.1$ / $\eta_F = 3$ / $N = 10$
- 600'000 images / 345 categories / 6 domains
- Train : 245 category & 5 domain / Val : 55 category & 5 domain / Test : 45 category & 1 target domain
- Zeroshot & DG를 했을 경우 기존의 방법을 동시에 적용하는 것보다 제안하는 방식이 더 좋은 성능을 보인다.

Table 3. ZSL+DG scenario on the DomainNet dataset with ResNet-50 as backbone.

Method	Clipart	Infograph	Painting	Quickdraw	Sketch	Avg.
SPNet	26.0	16.9	23.8	8.2	21.8	19.4
<i>mixup</i> +SPNet	27.2	16.9	24.7	8.5	21.3	19.7
Epi-FCR+SPNet	26.4	16.7	24.6	9.2	23.2	20.0
CuMix	27.6	17.8	25.5	9.9	22.6	20.7