A Zero-Shot Framework for Sketch Based Image Retrieval [ECCV `18]

CS688 Paper Presentation 2

Doheon Lee

20183398

2018. 11. 26

KAIST

Review : Adversarial Metric Learning

- **Metric**
  - **Measure similarity between two images**
  - **Mathematical measurements are not intuitive.**

- **Generating hard negative using GAN.**
  - **Better than using existing data for metric learning**

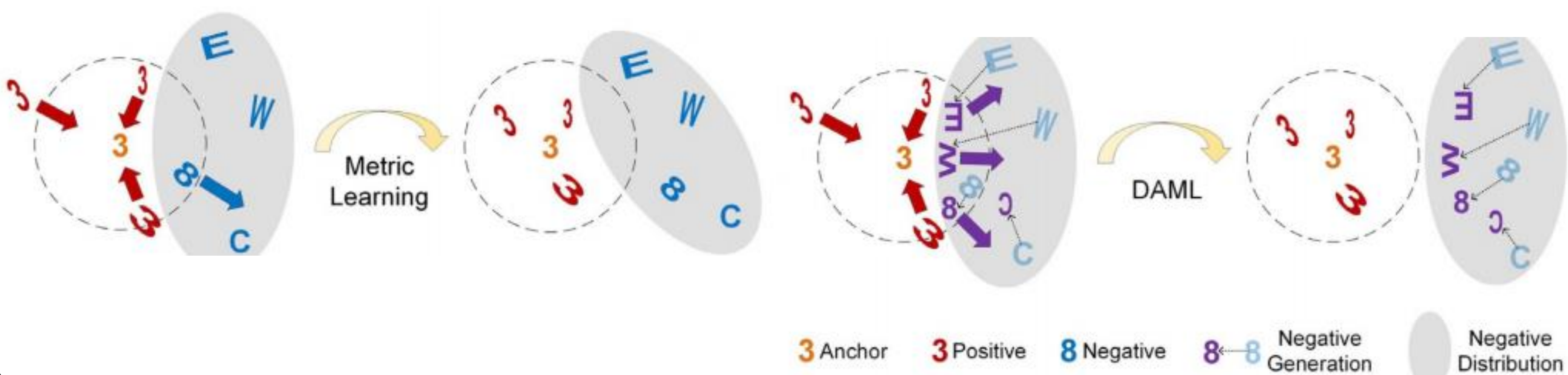


Table of Contents

- Introduction
- Background
- Main Contribution
- Experiment & Result

Introduction

Image Retrieval

- Text based image retrieval
  - Search image by textual description

- Content based image retrieval
  - Search image similar to query image
  - Sketch-based Image Retrieval (SBIR)

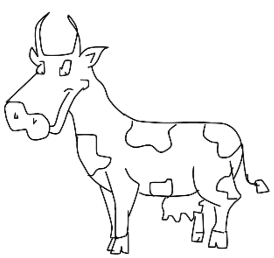
KAIST

Problems in Coarse Evaluation

- SBIR is usually used for fine-grained IR.
  - Current methods are focused on class –based retrieval.
  - Shape or attributed-based retrieval are important.

Problems in Coarse Evaluation

- **Get credit when fetches an image in same class.**
  - **No need to match outlines and shape**
  - **Simply learning a class specific mapping**



Query

Images

Fine-grained Evaluation

- **Evaluate by comparing the estimated rank.**
  - **Annotating rank list by human.**

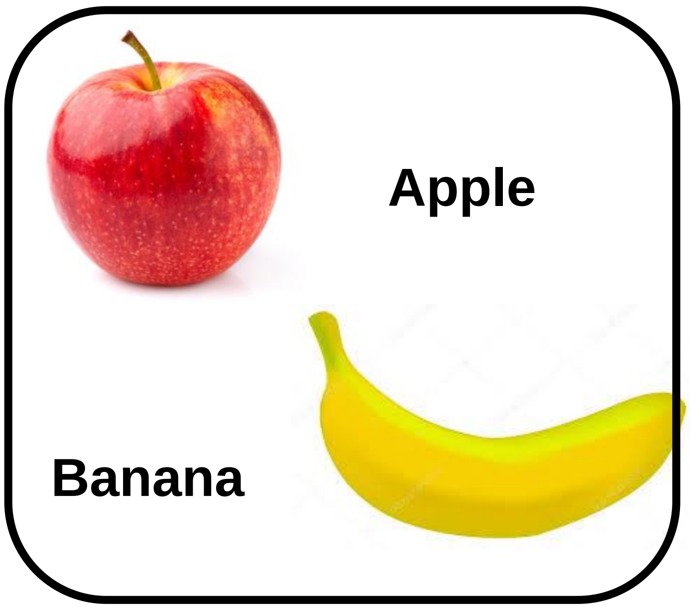
  → Human biased and requires human labor

Coarse-grained evaluation in the zero-shot setting.

Related Work

Zero-shot Learning

- **Learning to recognize images of novel classes**

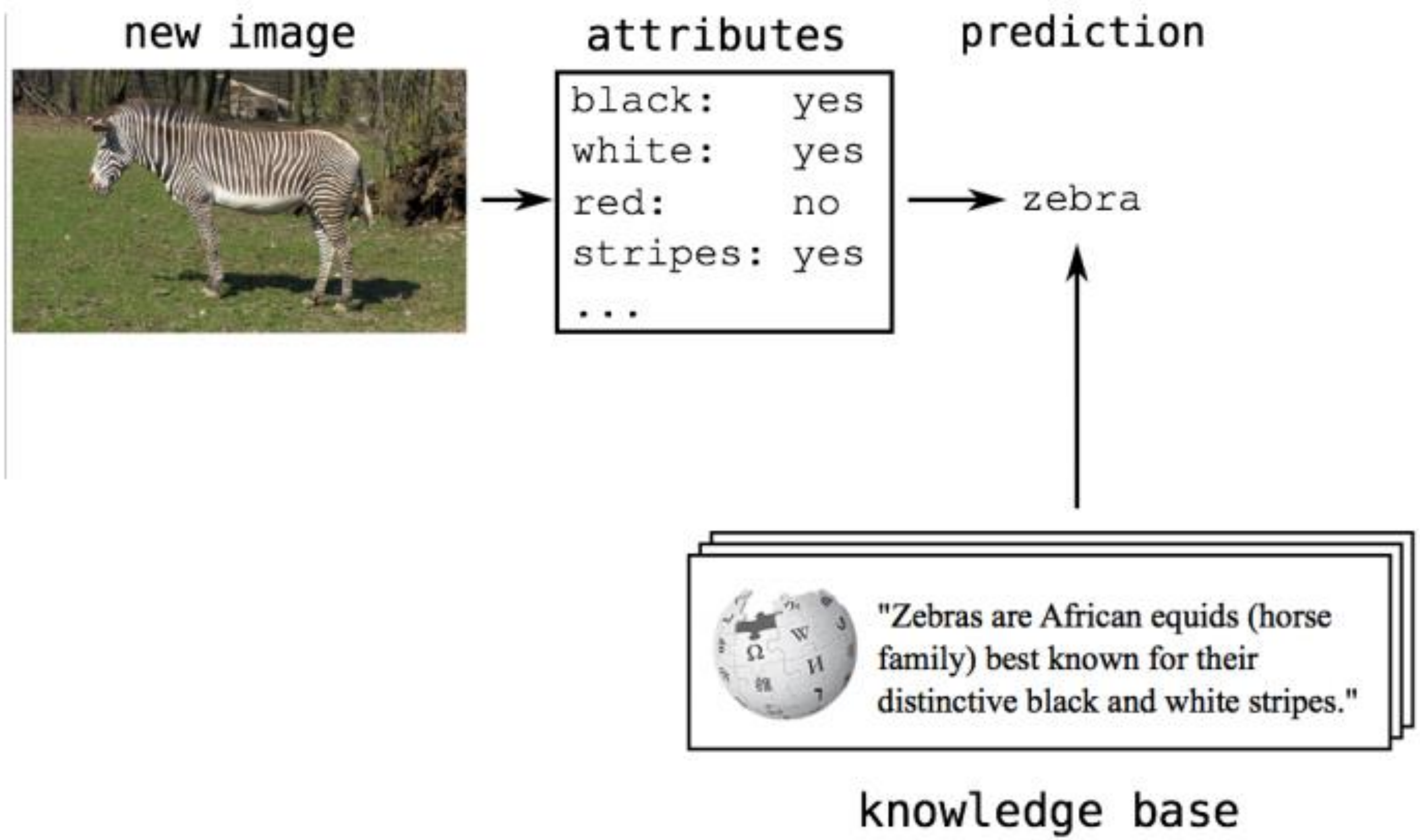


Training Set

Test Set

Zero-shot Learning

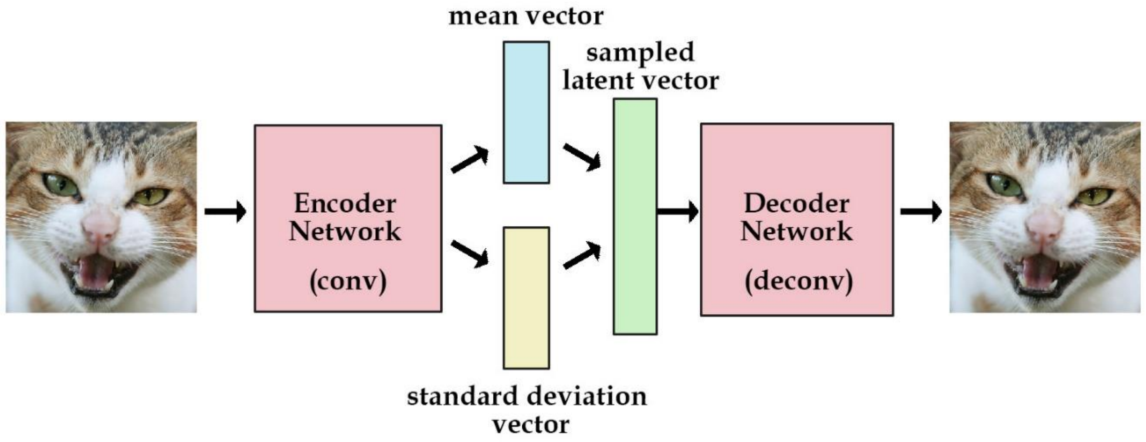
Attribute Based Classification: Example



Slide from ECCV 16 ZSL Tutorial

Variational Autoencoder

- **Find latent features from data**
- **Encoder**
  - **Encodes data (x) to latent variable (z)**
- **Decoder**
  - **decodes latent variable (z) to data(x)**



Main Contribution

Main Contribution

- **Proposed a new benchmark for zero-shot SBIR**
- **Proposed a generative approach for the SBIR task**

New Benchmark

- **Modified "Sketchy" dataset**
  - **Dataset contains images with 6 sketch each**
  - **125 classes : 104 train, 21 test**

Table 1. Statistics of the proposed dataset split of Sketchy database for ZS-SBIR task

Dataset Statistics	#
Train classes	104
Test classes	21
Train Images	10400
Train Sketches	62787
Avg. sketches per image	6.03848
Test Sketches	12694
DB images for training	62549
DB images for testing	10453

New Benchmark

- **Current SBIR works are class-based.**

Table 2. Precision and mAP are estimated by retrieving 200 images. - indicates that the authors do not present results on that metric. 1:Using 128 bit hash codes

Method	Precision@200		mAP@200	
	Traditional	Zero-Shot	Traditional	Zero-Shot
Baseline	-	0.106	-	0.054
Siamese-1	-	0.243	-	0.134
Siamese-2	0.690	0.251	0.518	0.149
Coarse-grained triplet	0.761	0.169	0.573	0.083
Fine-grained triplet	-	0.155	-	0.081
DSH[1]	0.866	0.153	0.783	0.059

Generative Model for ZS-SBIR

- **Sketch gives a basic outline of the image.**
  - **Additional details are generated from the latent prior vector**
  - **Training by sketch-image pairs to model probability density function:** $p(x_{img}|x_{sketch};\theta)$

x: features

- **The trained result can generate image features.**

Conditional VAE

- **Variational lower bound for p(x)**

$$p(x) \geq \mathcal{L}(\phi, \theta; x)$$

q: variational distribution (Gaussian)

$$= -D_{KL}\left(q_\phi(z|x) || p_\theta(z)\right) + \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]$$

- **Conditional probability** $p(x_{img}|x_{sketch})$

$$\mathcal{L}(\phi, \theta; x_{img}, x_{sketch}) =$$

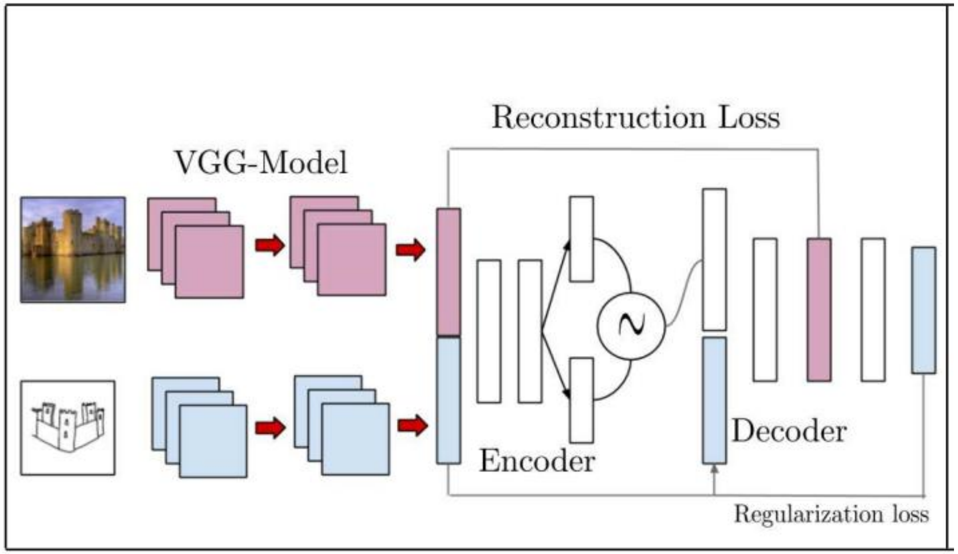
$$-D_{KL}\left(q_\phi\left(z|x_{img}, x_{sketch}\right) || p_\theta\left(z|x_{sketch}\right)\right) + \mathbb{E}\left[\log p_\theta\left(x_{img}|z, x_{sketch}\right)\right]$$

Conditional VAE

- **Regularization loss for preserving latent alignments of the sketch**

$$\mathcal{L}_{recons} = \lambda . \|f_{NN}(\widehat{x}_{img}) - x_{sketch}\|_2^2$$

Generated feature



Conditional Adversarial AE

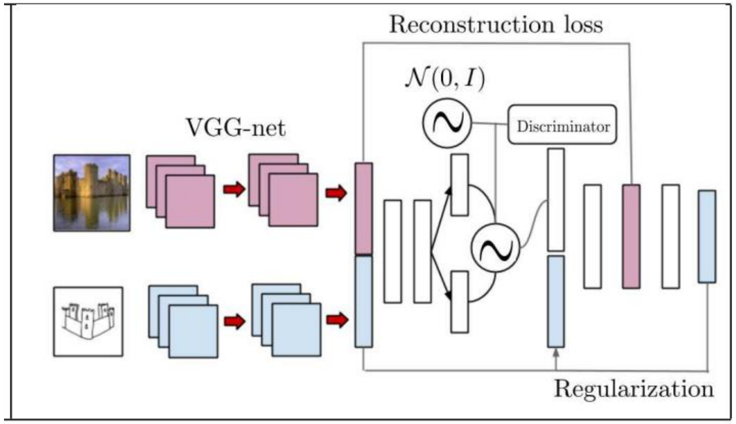
- **Using GAN model replaced KL-Divergence term.**
  - **Network Minimize loss**

E: encoder

$$\mathbb{E}_z\left[\log p_\theta\left(x_{img} \middle| z, x_{sketch}\right)\right] + \mathbb{E}_{x_{img}}\left[\log\left(1 - \mathcal{D}(E(x_{img}))\right)\right]$$

  - **Discriminator $\mathcal{D}$ maximize following terms**

$$\mathbb{E}_z\left[\log\left[\mathcal{D}(z)\right]\right] + \mathbb{E}_{x_{img}}\left[\log\left[1 - \mathcal{D}\left(E(x_{img})\right)\right]\right]$$



Experiment & Result

Experiment benchmark

- **The experiments are done in proposed zero-shot benchmark**

- **Features are generated from decoder part.**
  - **Sampled features are clustered using K-means.**

VGG-16 features

$$\mathcal{D}(x_I^{db}, \mathcal{I}_{x_S}) = min_{k=1}^{K} cosine\left(\theta(x_I^{db}), C_k\right)$$

Cluster Center

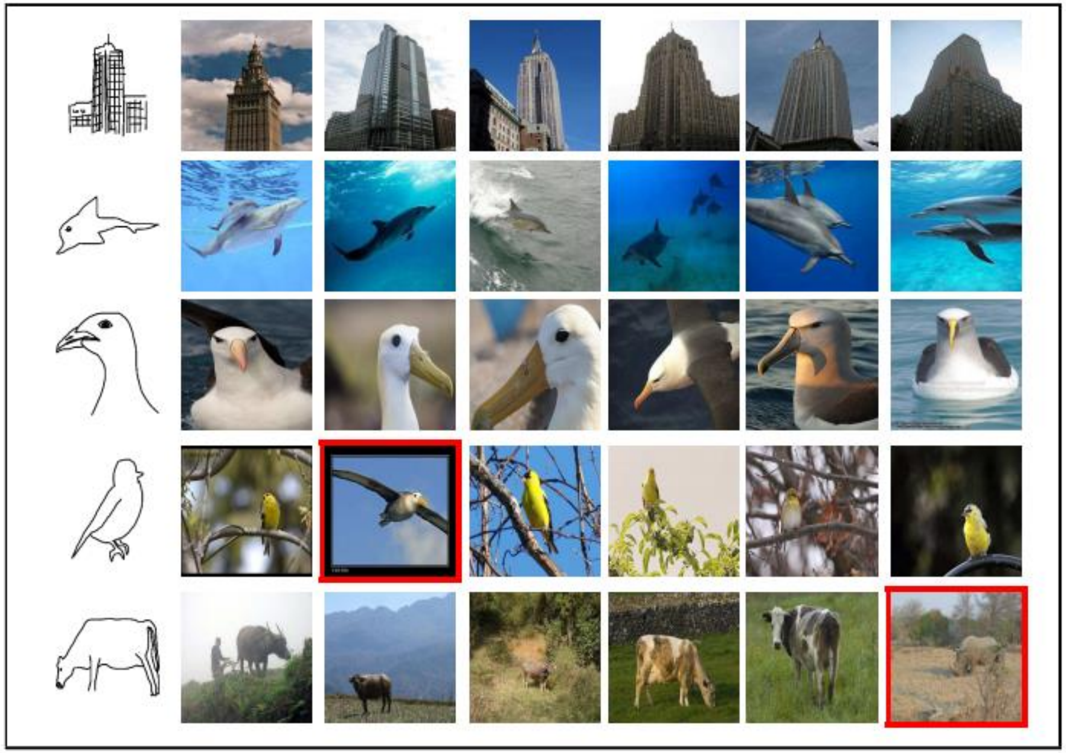
Result

Table 3. The Precision and MAP evaluated on the retrieved 200 images in ZS-SBIR on the proposed split

Type	Evaluation Methods	Precision@200	mAP@200
SBIR methods	Baseline (VGG-16)	0.106	0.054
	Siamese-1	0.243	0.134
	Siamese-2	0.251	0.149
	Coarse-grained triplet	0.169	0.083
	Fine-grained triplet	0.155	0.081
	DSH	0.153	0.059
ZSL methods	Direct Regression	0.066	0.022
	ESZSL	0.187	0.117
	SAE	0.238	0.136
Ours	CAAE	**0.260**	**0.156**
	CVAE	**0.333**	**0.225**

Deep Sketch Hashing

Result



Preserved Attribute

Fig. 3. Top 6 images retrieved for some input sketches using CVAE in the proposed zero-shot setting. Note that these sketch classes have never been encountered by the model during training. The red border indicates that the retrieved image does not belong to sketch's class. However, we would like to emphasize that the retrieved false positives do match the outline of the sketch