A Review of Generalized Zero-Shot Learning Methods

Farhad Pourpanah, *Member, IEEE,* Moloud Abdar, Yuxuan Luo, Xinlei Zhou, Ran Wang, *Member, IEEE,* Chee Peng Lim, and Xi-Zhao Wang, *Fellow, IEEE*

Abstract—Generalized zero-shot learning (GZSL) aims to train a model for classifying data samples under the condition that some output classes are unknown during supervised learning. To address this challenging task, GZSL leverages semantic information of the seen (source) and unseen (target) classes to bridge the gap between both seen and unseen classes. Since its introduction, many GZSL models have been formulated. In this review paper, we present a comprehensive review on GZSL. Firstly, we provide an overview of GZSL including the problems and challenges. Then, we introduce a hierarchical categorization for the GZSL methods and discuss the representative methods in each category. In addition, we discuss the available benchmark data sets and applications of GZSL, along with a discussion on the research gaps and directions for future investigations.

Index Terms—Generalized zero shot learning, deep learning, semantic embedding, generative adversarial networks, variational auto-encoders

A Presentation by Yara mohammadi

Advanced Deep Learning course
University of Tehran - AI Department
Dr. Amin Sadeghi

Zero-shot Learning

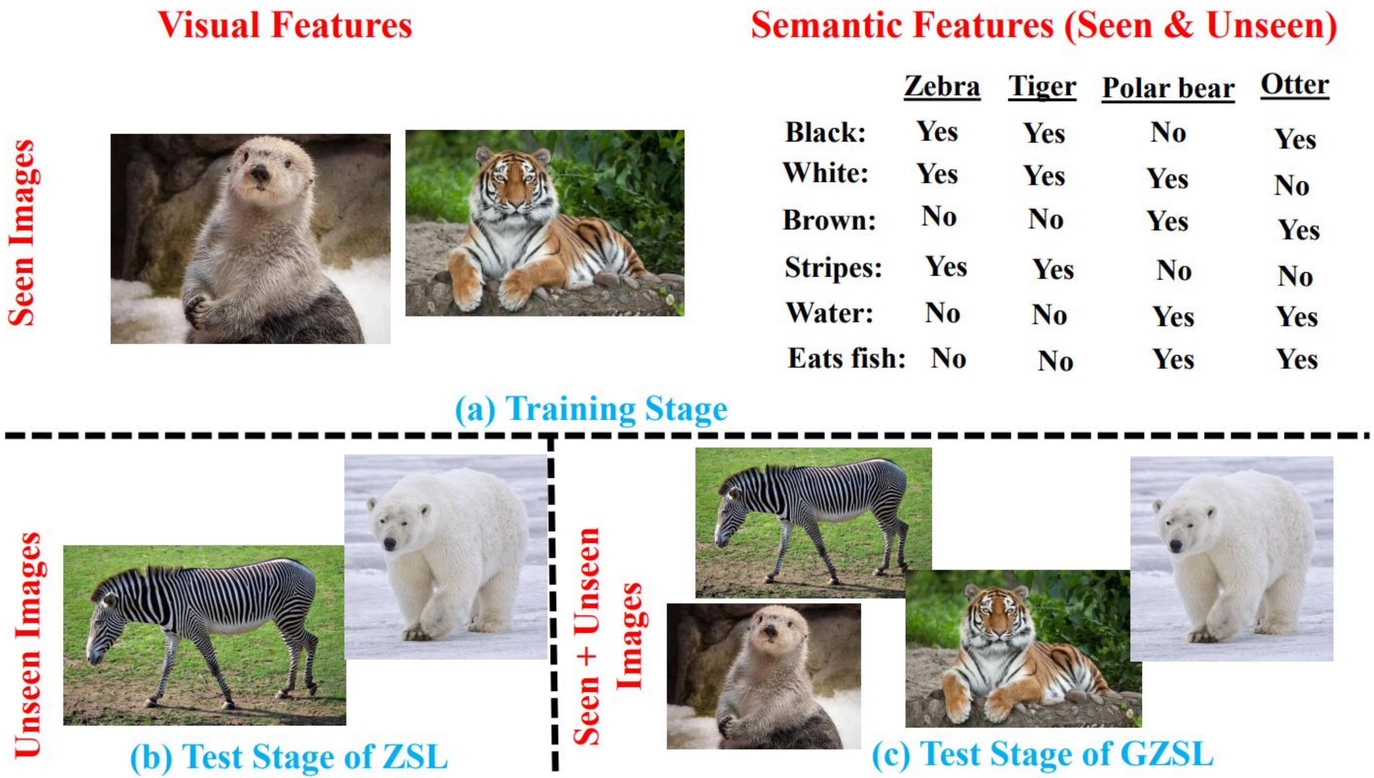
- Audience Task: Recognize the Wampimuk
- Solution: Semantic Transfer
  - Domain Ontology
    - Wampimuk: small, horns, furry, cute
  - Wikipedia Page

Pattern recognition with no training examples

Why care?

- Hard to annotate large-scale labels
- Lack of sufficient labeled samples (endangered birds)
- New samples observed in progress (COVID-19)
- ZSL is more similar to human behaviour

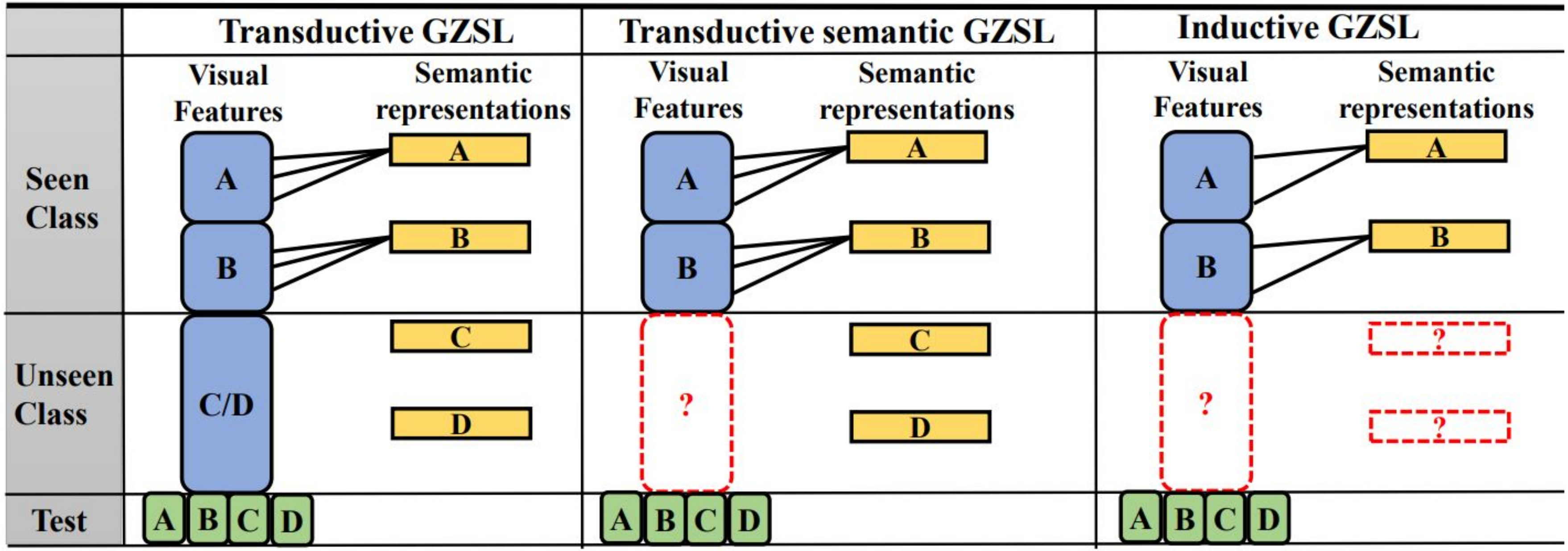
ZSL vs. GZSL



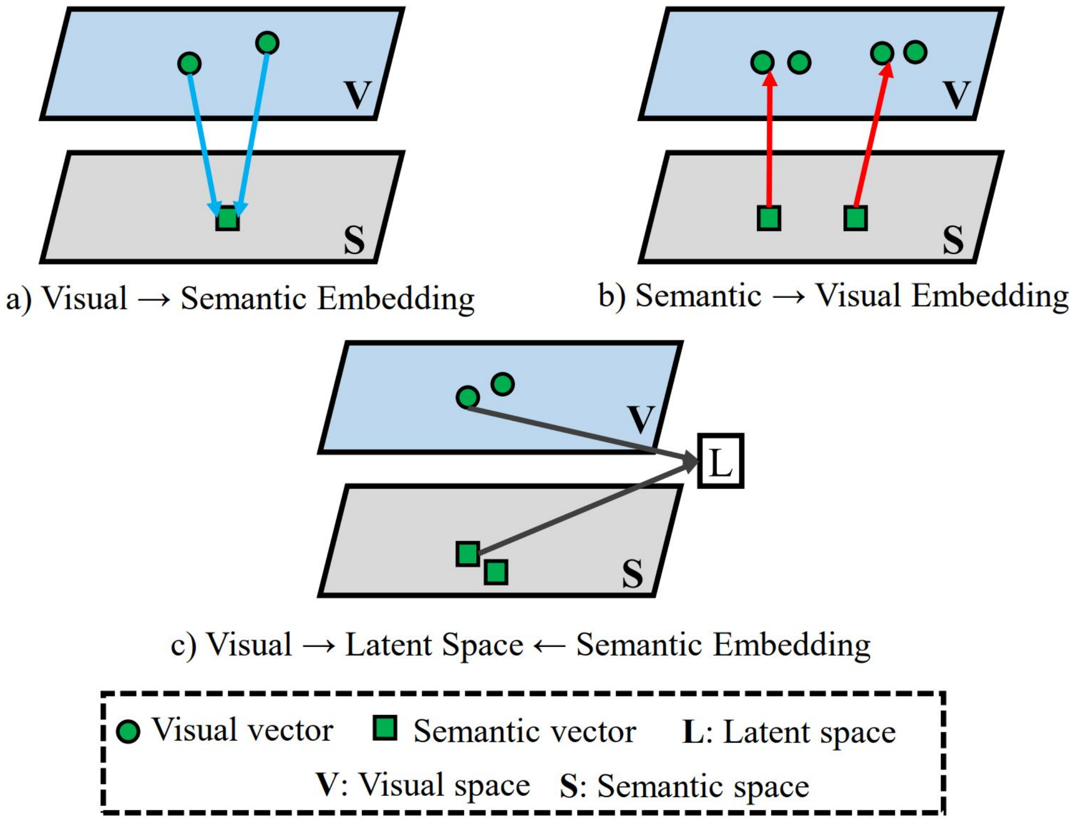
Visual Features
Semantic Features (Seen & Unseen)
Seen Images
Zebra Tiger Polar bear Otter
Black: Yes Yes No Yes
White: Yes Yes Yes No
Brown: No No Yes Yes
Stripes: Yes Yes No No
Water: No No Yes Yes
Eats fish: No No Yes Yes
(a) Training Stage
Unseen Images
(b) Test Stage of ZSL
Seen + Unseen Images
(c) Test Stage of GZSL

Problem Formulation

How to train model?



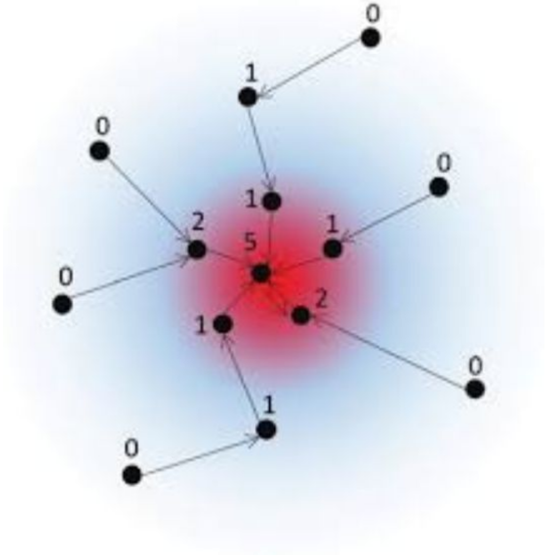
Embedding spaces



Challenges

Hubness

- Aspect of curse of dimensionality
- Visual to Semantic space & Nearest Neighbor

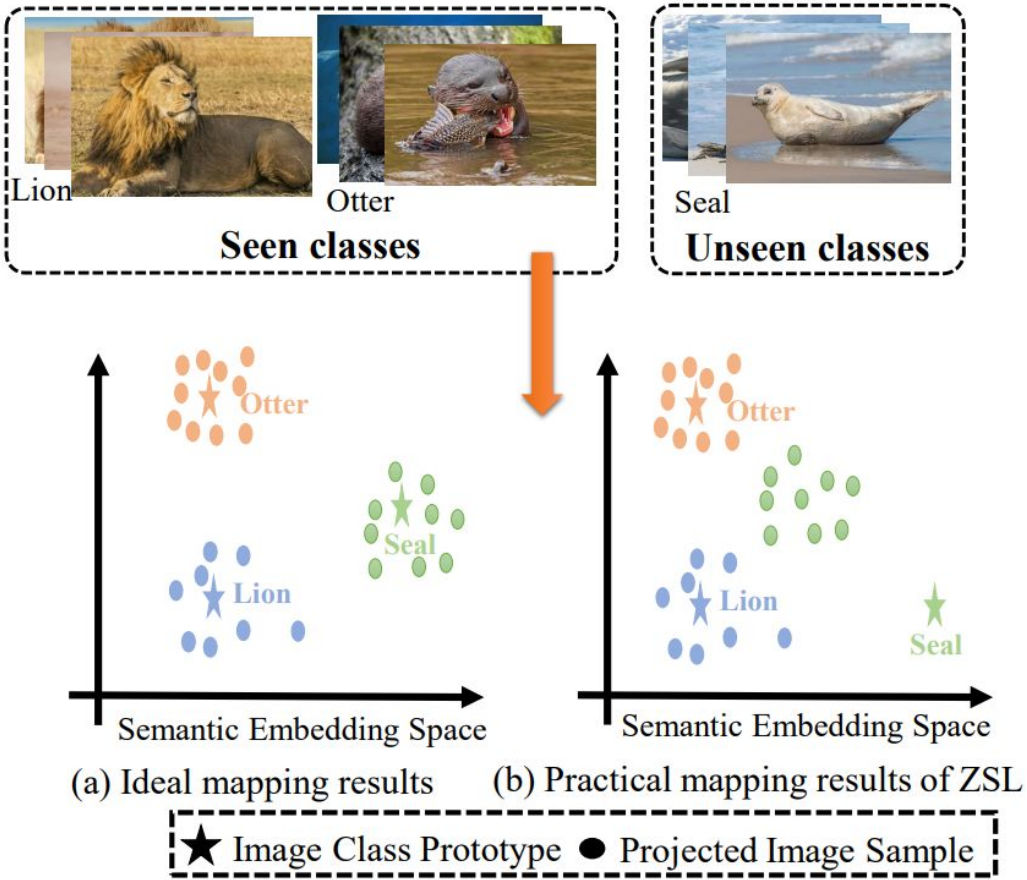


Challenges

Domain Shift

- Visual & Semantic have different spaces

- Seen & unseen classes have different distribution



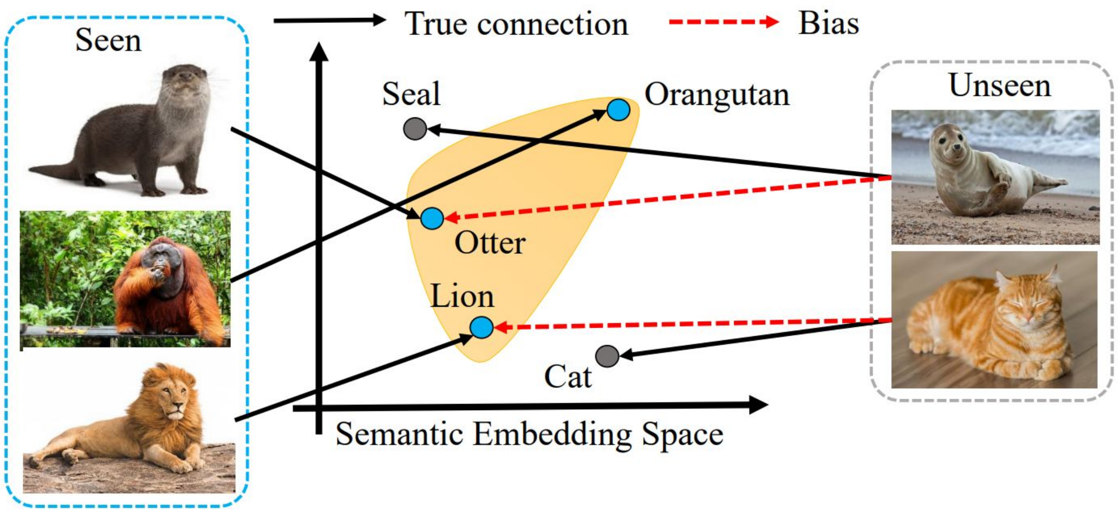
Challenges

Biased towards seen classes

- ZSL methods

Solution:

- Calibrated Stacking
- Novelty Detector



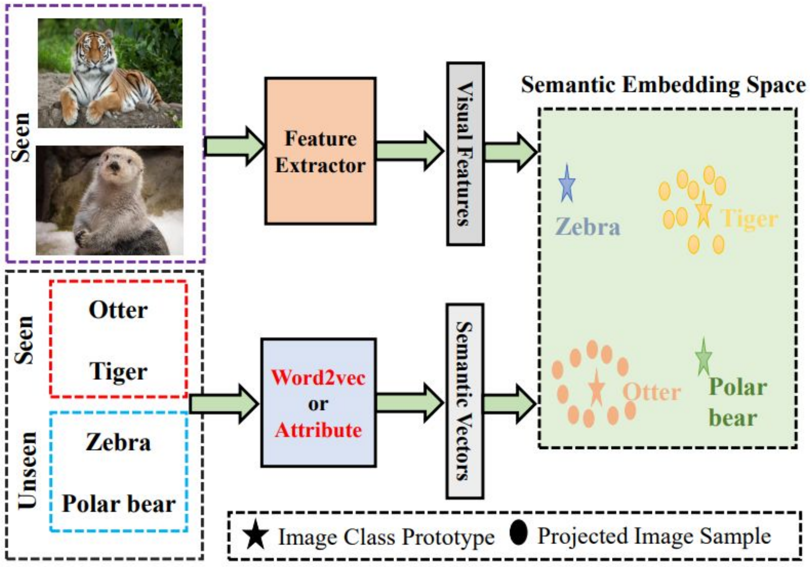
Performance Indicator

$$HM = 2 \times \frac{Acc_s \times Acc_u}{Acc_s + Acc_u}$$

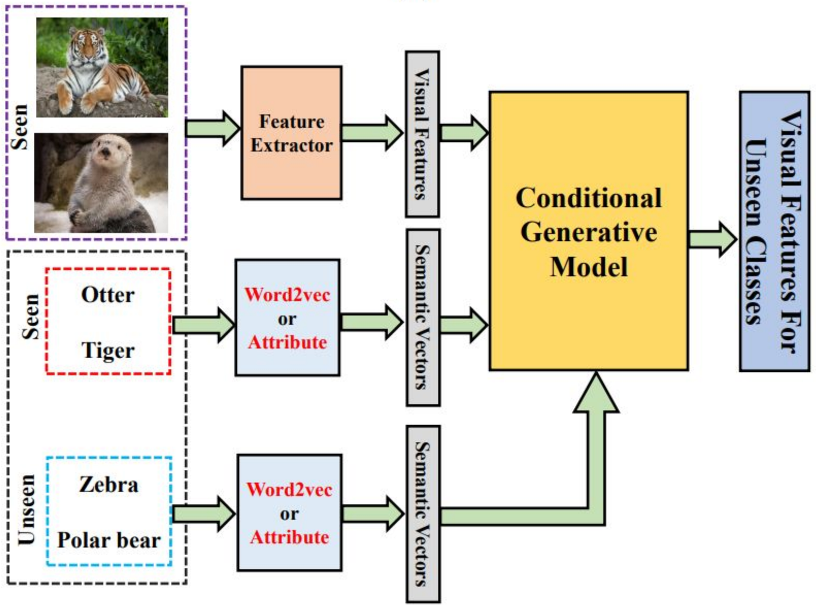
$$HM = HarmonicMean$$
$$Acc_s = SeenClassesAccuracy$$
$$Acc_u = UnseenClassesAccuracy$$

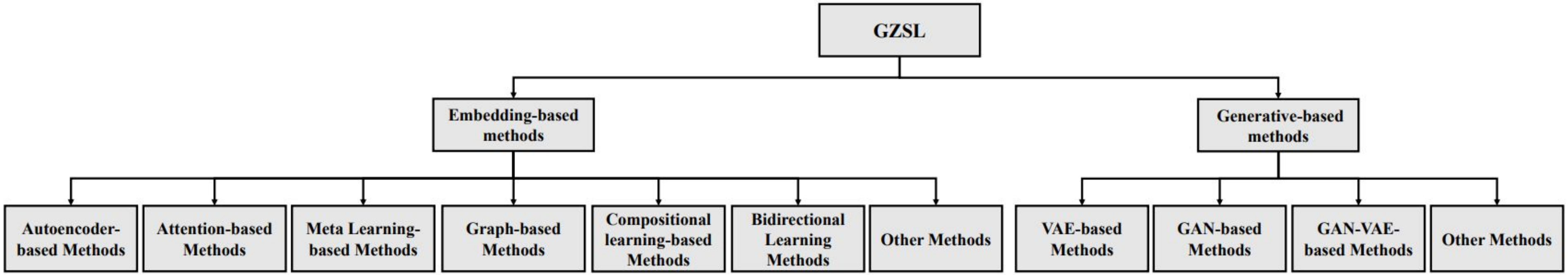
Inductive & Semantic Transductive GZSL Methods

Embedding Based Methods



Generative Based Methods

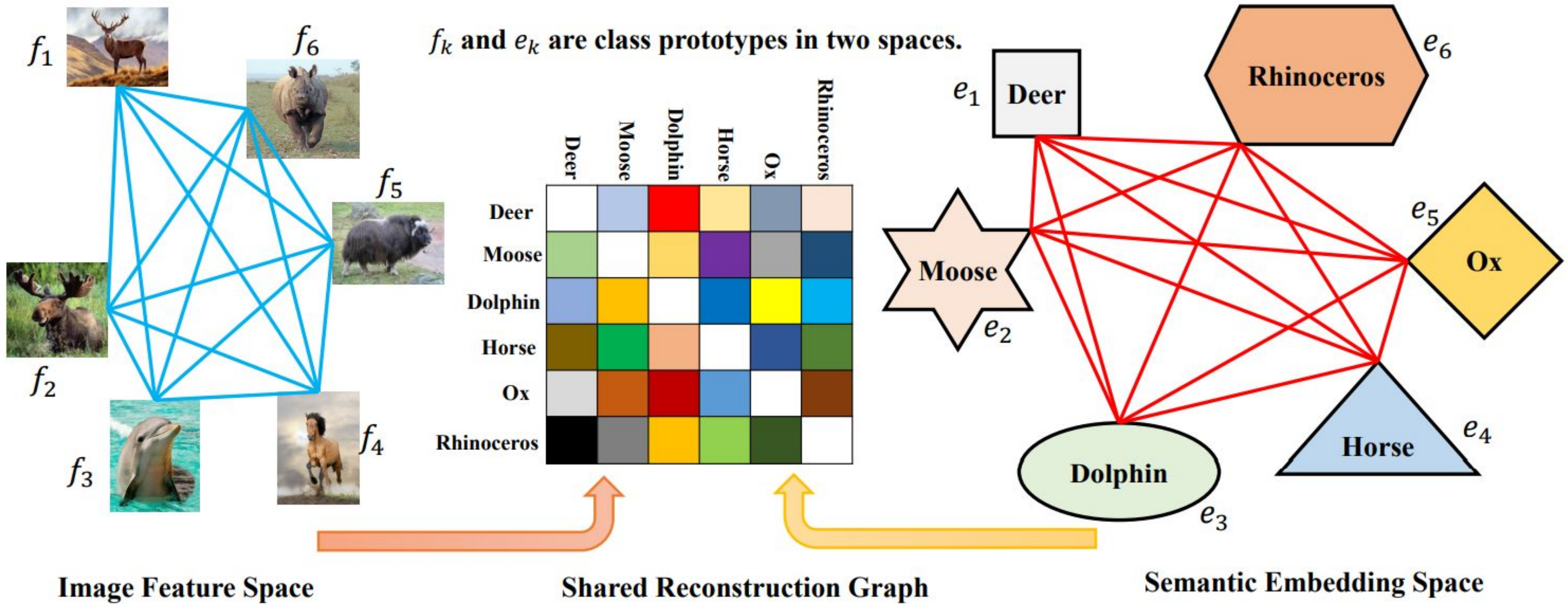



GZSL
Embedding-based methods
Generative-based methods
Autoencoder-based Methods
Attention-based Methods
Meta Learning-based Methods
Graph-based Methods
Compositional learning-based Methods
Bidirectional Learning Methods
Other Methods
VAE-based Methods
GAN-based Methods
GAN-VAE-based Methods
Other Methods

Embedding-based Methods

Graph-based Methods

SRG (2017)



Meta Learning-based Methods

Learning to learn strategy!

Divide training classes into 2 sets:

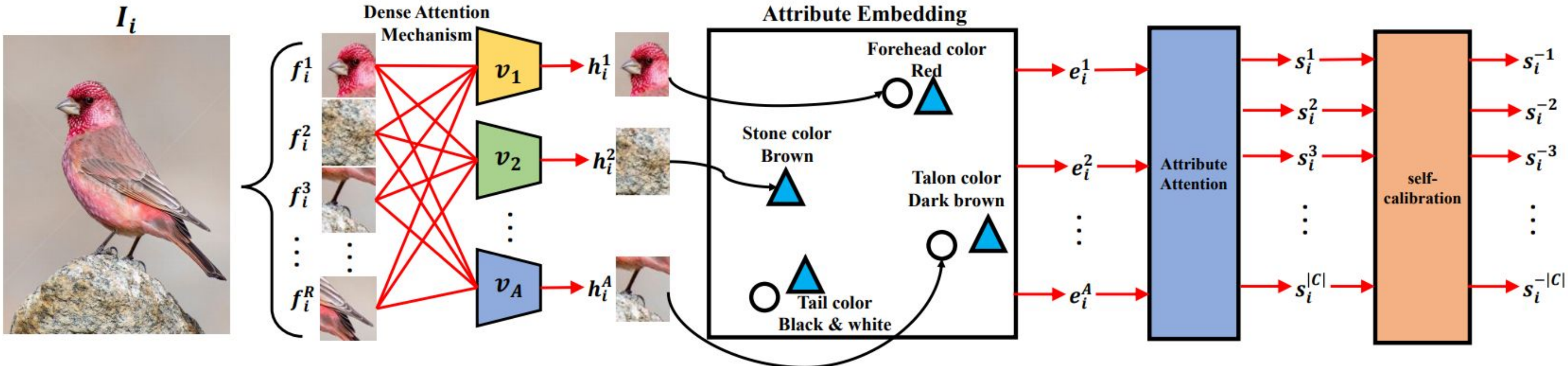
Support (seen) & Query (unseen)

Learn different tasks by randomly selecting a set

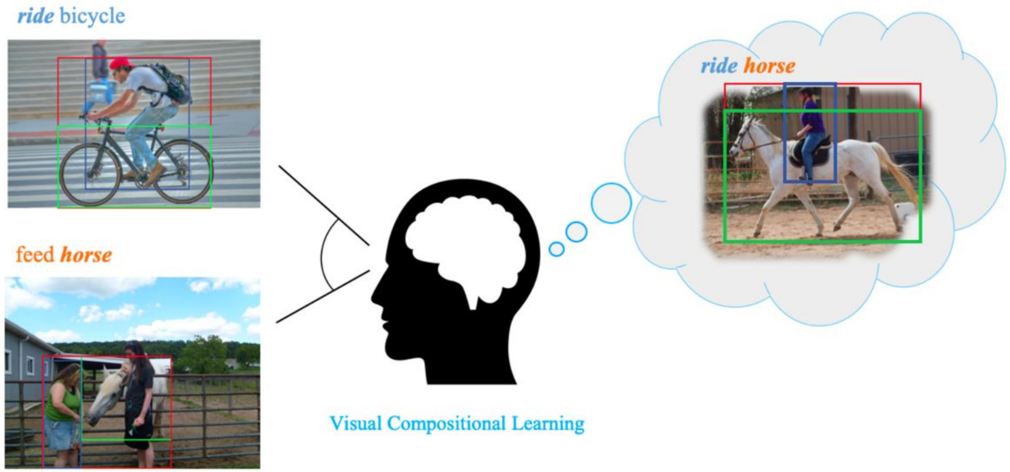
Helps transfer knowledge from unseen to seen (alleviate Bias problem)

Attention-based Methods

DAZLE (2020)

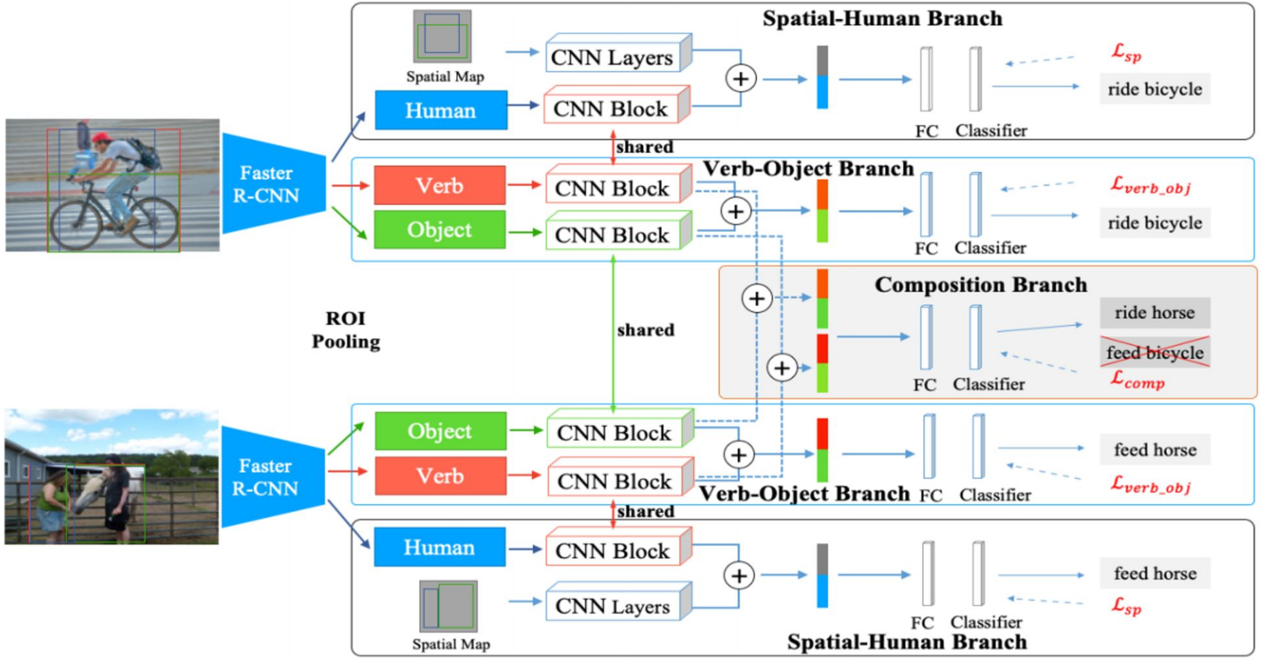


Compositional learning-based Methods



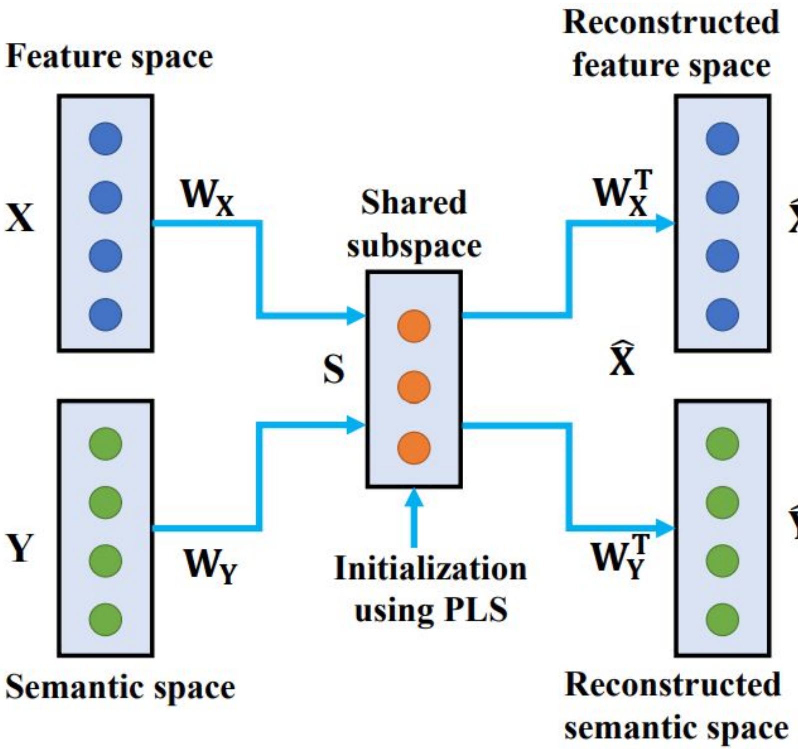
Compositional learning-based Methods

Visual Compositional Learning for HOI Detection (2020)



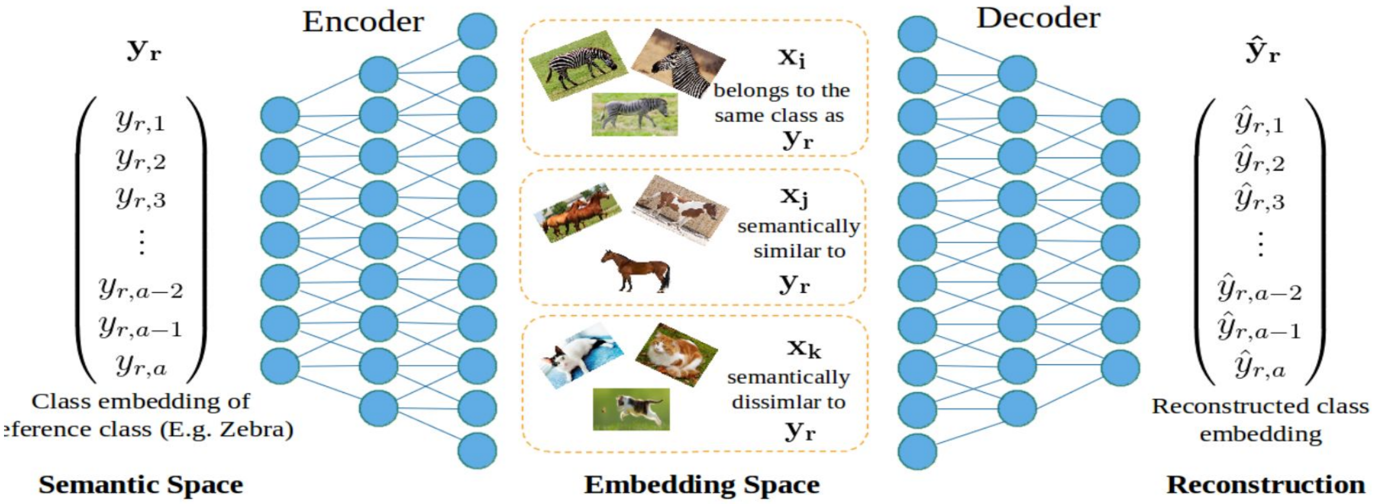
Bidirectional Learning Methods

JIL (2019)



Autoencoder-based Methods

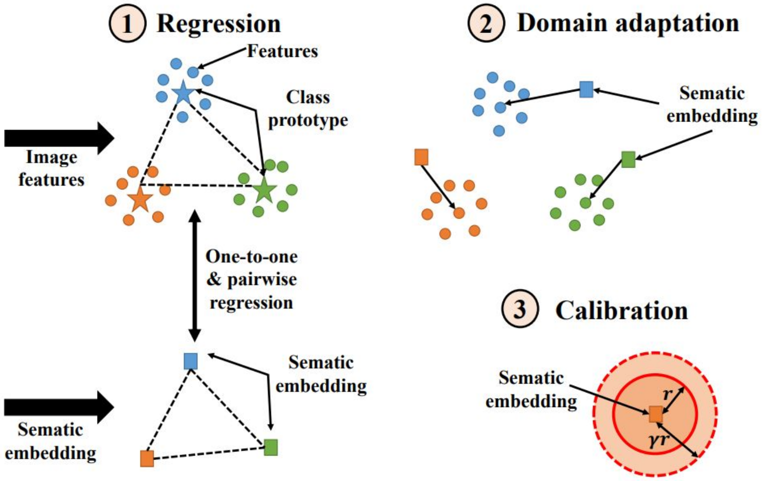
Preserving semantic relations for ZSL(2018)



Other methods

Consider inter-class and intra-class relations

ZSL using relational matching, adaptation & calibration (2019)



Generative-Based Methods

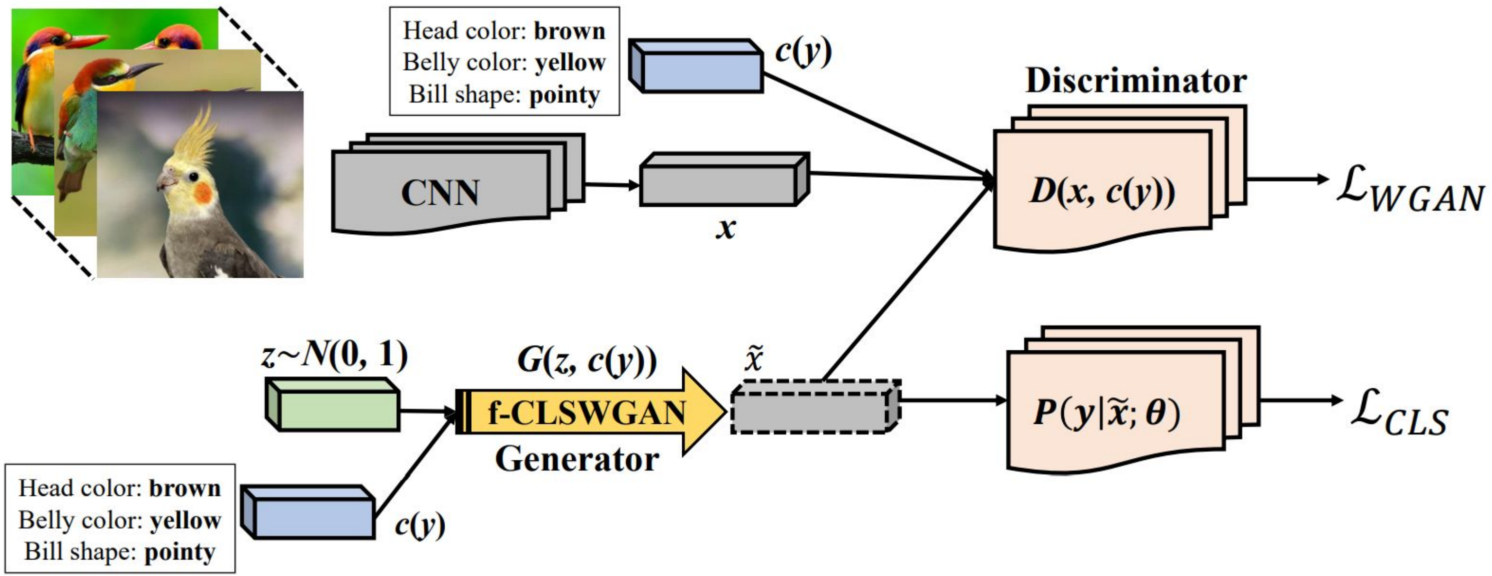
Generative Based Methods

Generated samples should be:

- Semantically related to real samples
- Discriminative so classification is easy

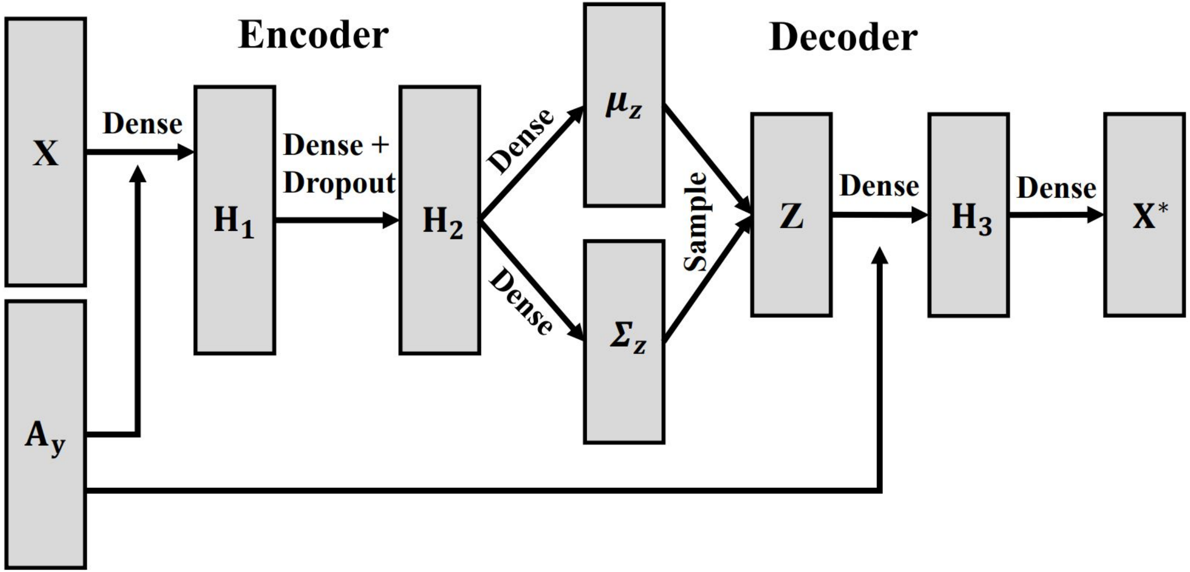
Generative Adversarial Networks

f-CLSWGAN (2018)



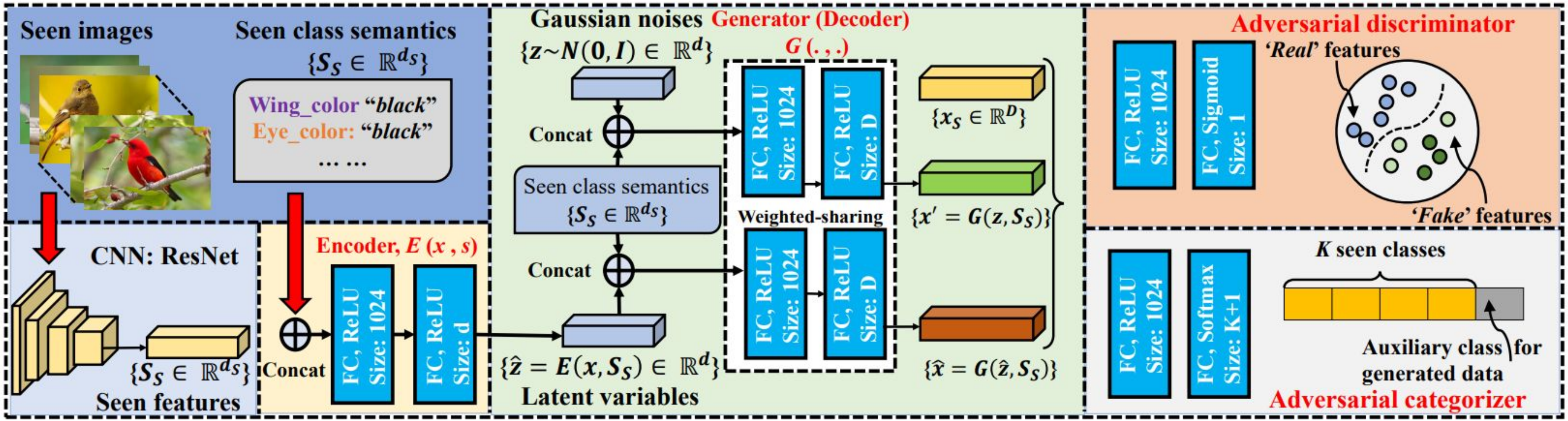
Variational Autoencoders

CVAE-ZSL (2018)



Combined GANs & VAEs

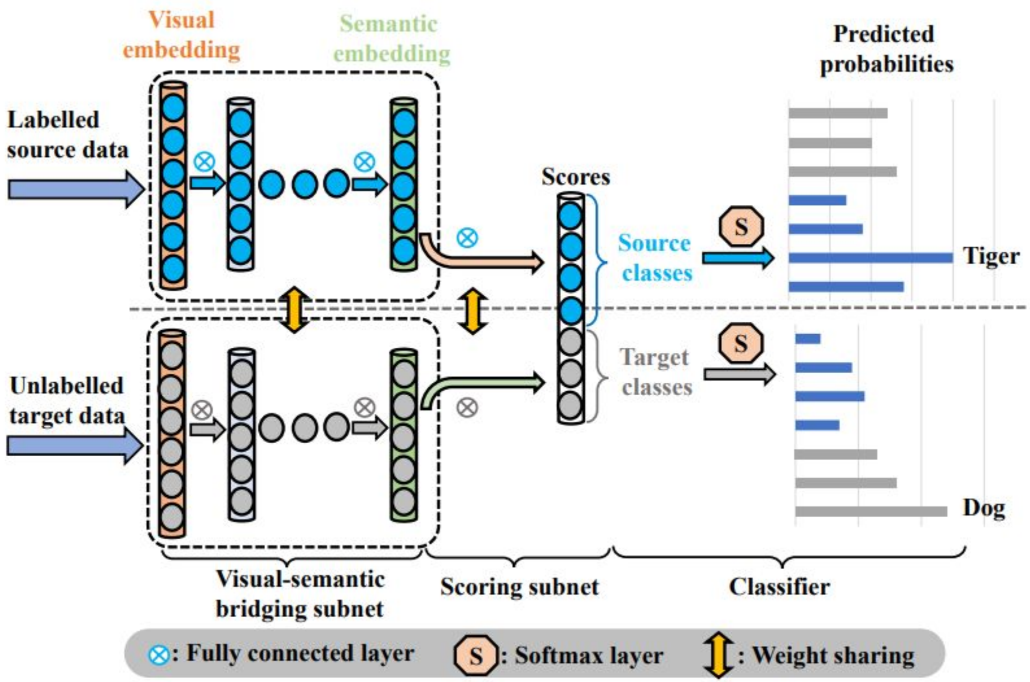
VAEGAN (2020)



Transductive GZSL Methods

Embedding-based

QFSL (2018)



Generative-based

Comparisons

- Embedding-based
    - Easy to implement
    - Poor performance under GZSL problem (low HM)
        - Knowledge transfer limited to semantic loss
        - Lack of visual samples -> bias problem
        - Search in HD space -> Hubness problem
    - Semantic embedding models:
        - shrink variance & reduce discriminability
    - Visual embedding models:
        - Direct projection causes info loss, & overfitting on seen classes
    - Latent embedding models:
        - Better in learning structural differences of 2 spaces & ajdusting to unseen classes
    - Still has domain shift & bias problem

Comparisons

- Generative-based
  - Complex & difficult to train
  - More balanced performance (high HM)
    - Knowledge transfer limited to :
      - learning distribution of visual features
      - Constraining to retain info between generated features & real semantic features
    - Synthesizing unseen visual features -> lower bias
  - Have access to unseen data under transductive settings -> Violation of GZSL setting

Comparisons

- Transductive
  - Access to distribution of unseen classes
  - Solve bias & shift problems
  - Higher HM score

- Inductive & Semantic transductive
  - Lower HM score

Applications

- Image:
  - Classification
    - ImageNet,, CUB-200-2011, aPascal-aYahoo, AnimalWithAttribute, SUN attribute dataset, North America Birds, DeepFashion …
  - Object detection, Segmentation, Image annotation
- Video:
  - Action & Gesture recognition
    - Multi-modal (audio + video + text)!
- NLP:
  - Single/Multi-label Text Classification, noisy text description, ...

Thank you