gipsa-lab

Grenoble | images | parole | signal | automatique | laboratoire

# Zero Shot Learning

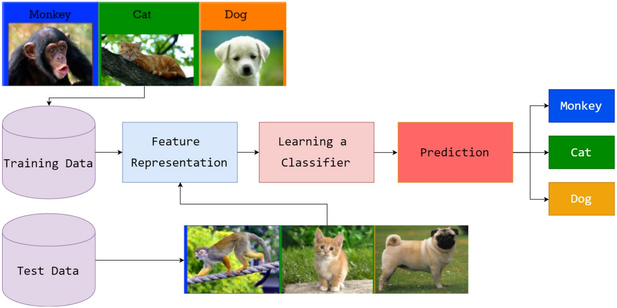
## Zero Shot Learning


Dawood ALCHANTI

Laboratoire Gipsa-lab, UMR 5216, CNRS, UGA
11 rue des mathématiques 38420 Grenoble, France
*dawood.alchanti@gipsa-lab.fr*


2019/05/13

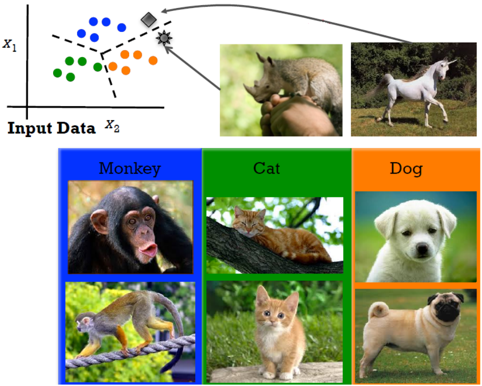
# Conventional Supervised Learning




## Supervised Learning Setting

1. Sufficient labeled training instances are needed for each class.
2. Can only classify the instances belonging to classes covered by the training data.

# Generalization of Conventional Supervised Setting




## Main Drawback

1. Lack the ability to deal with new unseen classes!.




## Your turn:

- Can you recognize the **Wampimuk**?

# How it work?

## Your turn:

- Can you recognize the **Wampimuk**?

## Solution: via Semantic Transfer

- *Extra information is needed*:
  - **Wampimuk** := “small, horns, furry, cute”.
- Textual description, visual information, semantic knowledge, ..

# Why do we care about ZSL?

## Advantages

Reduce the time and the effort for annotating new emerging categories, especially fine-grained ones.

## Auxiliary Information or Semantic Information

- Contains information about Seen $\mathcal{S}$ and Unseen Class Set $\mathcal{U}$.
- Should be related to the instances in the feature space $\mathcal{X}$..
- Form a bridge or a link between $\mathcal{S}$ and $\mathcal{U}$.

## Definition

Zero-Shot Learning (ZSL): Given **labeled** training **instances** $D^{tr}$ belonging to the **seen** classes $\mathcal{S}$. ZSL aims to learn a classifier $f(.)^u$: $\chi \mapsto \mathcal{U}$ that can classify testing instances $X^{te}$ belonging to **unseen** classes $\mathcal{U}$, such that the **label space**, $\mathcal{L} = \mathcal{S} \cap \mathcal{U} = \emptyset$, is **disjoint**.

## Main Idea

**Transfer knowledge** contained in the **source domain/task** to the target domain for learning model in the **target task**.

## Recognition Setting

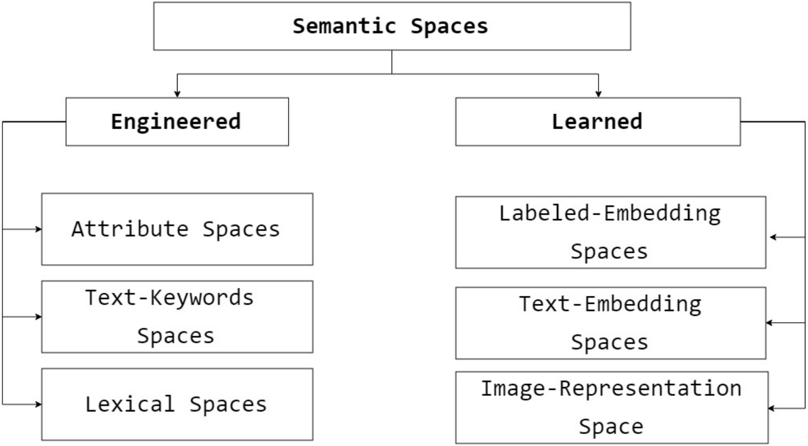
Via metadata transfer relating unseen classes to some known categories.

# ZSL Formalization

## Some Terminology

- **Feature space** $\mathcal{X} \in \mathbb{R}$, **Semantic Space** $\mathcal{T} \in \mathbb{R}$.
- **Each class** $c_i$ **has** a corresponding vector representation which referred to as: ***Class Prototype*** $t_i$.
- $(x_i^{tr}, y_i^{tr})$ is the ith labeled training instance, where $x_i^{tr} \in \mathcal{X} \in \mathbb{R}^d$ and $y_i^{tr} \in \mathcal{S}$.
- $x_i^{te}$ is the ith unlabeled test instance, where $x_i^{te} \in \mathcal{X} \in \mathbb{R}^d$.
- $T^s = \{t_i^s\}_{i=1}^{N_s}$: Set of seen prototypes, $t_i^s$: Class Prototypes for Seen Class $c_i^s$.
- $T^u = \{t_i^u\}_{i=1}^{N_u}$: Set of unseen prototypes, $t_i^u$: Class Prototype for Unseen Class $c_i^u$.

# Semantic Space



# **Engineered**: Attribute Space as Semantic Space

otter

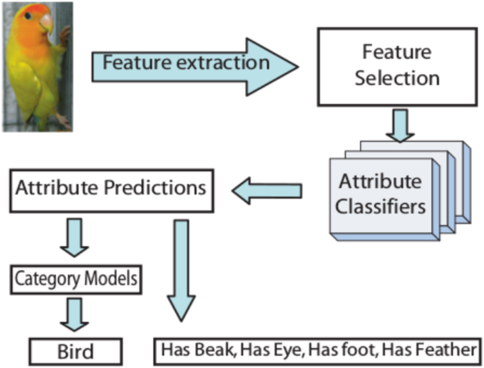
black: yes
white: no
brown: yes
stripes: no
water: yes
eats fish: yes

polar bear

black: no
white: yes
brown: no
stripes: no
water: yes
eats fish: yes

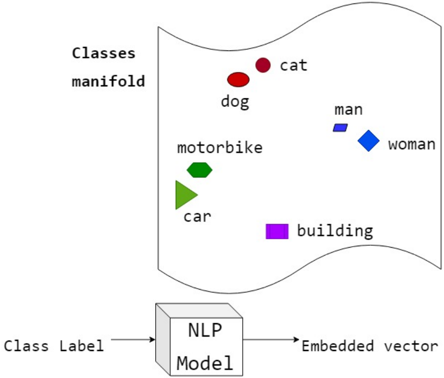
zebra

black: yes
white: yes
brown: no
stripes: yes
water: no
eats fish: no




### Attribute and Category Classifiers

Learn **attribute classifiers** for each attribute. To **predict** object **category**: use the predicted **attributes as features**.

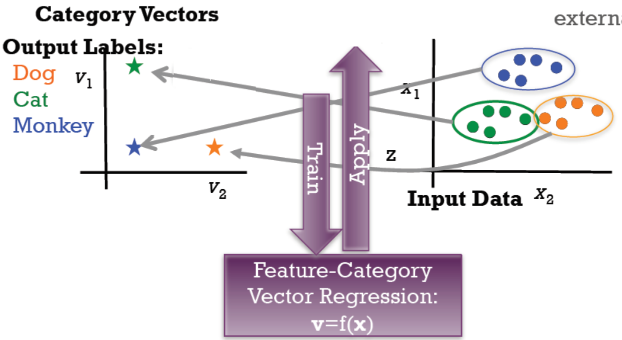



## Various methods to obtain embedding vector

- Word2Vector, Glove, RNN, ...
- Skip-gram: Co-occurrence over Wikipedia.

Key is to embed categories as vectors

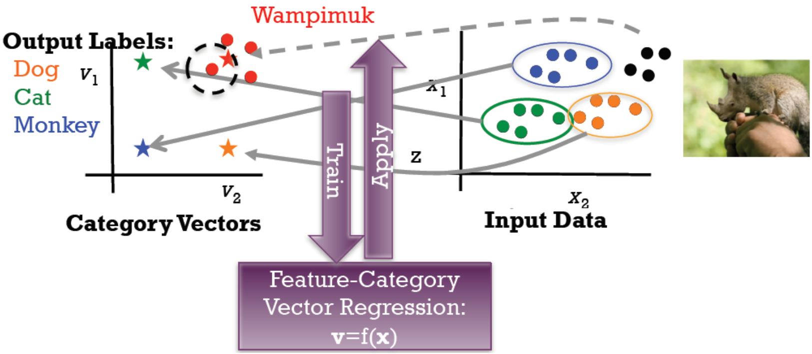
Task: Transferred from external metadata source




# Learn Mapping or Compatibilty Function

Task: Transferred from external metadata source

Map: Transferred from past experience

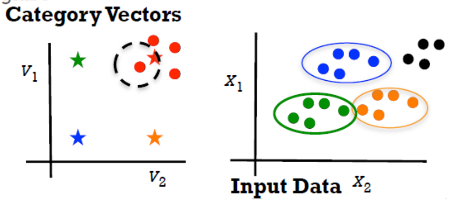



Feature-category vector map can generalize to new categories.

# ZSL Algorithm # 1: Regression/Classification

- 1. Obtain category vectors, v.
  - Attribute description (Wampimuk := small, cute, furry, horns )
  - Word-vector (Digested wikipedia co-occurrence count)
- 2. Train:
  - Given some known class-category vectors v and images x:
  - Learn image -> category vector classifiers/regressors v=f(x).
    - E.g., SVM / OLS,SVR. Deep Neural Net.
- 3. Test:
  - Specify vec v* for new class to recognise
  - Map test data f(x*) to cat vec space
  - NN matching of v* vs f(x*)
- Pros/Cons:
  - + Easy and fast!
  - - Category separability

Lampert, CVPR'09
Socher, NIPS'13
Xu, ECCV'16

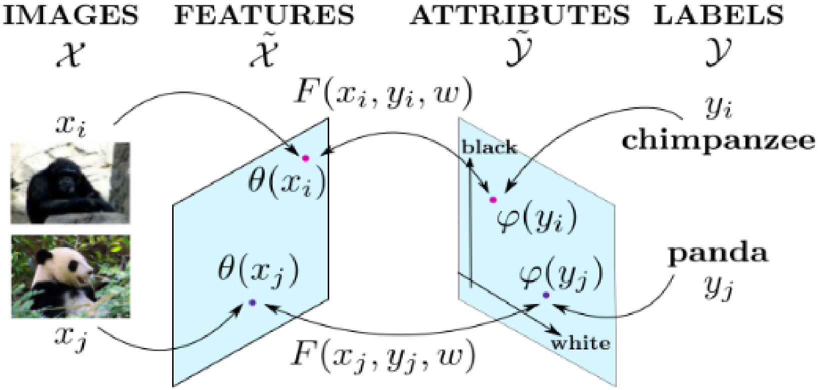


# ZSL Algorithm # 2: Energy Function Ranking

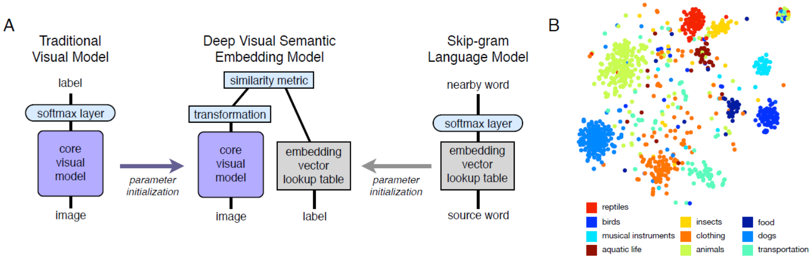
- Given: Training image data **x**, category vectors **v**.
- 1. Train an energy function $E_W(\mathbf{x},\mathbf{v})$.
  - E.g. Bilinear: $E_W(\mathbf{x},\mathbf{v})=\mathbf{x}^T W \mathbf{v}$.
  - W such that $E_W(\mathbf{x},\mathbf{v})$ is large when data and category match. x=v.
  - W such that $E_W(\mathbf{x},\mathbf{v})$ is small when data and category mismatch. x!=v.
  - E.g., Max margin ranking objective.
- 2. Testing: Classify example x* that may be a novel class.
  - Consider vectors **v*** for classes to recognize.
  - Evaluate $E(\mathbf{x}^*,\mathbf{v}^*)$ for each v*.
  - Max response gives classification. $\mathrm{argmax}_{v^*} E(x^*,v^*)$

Frome, NIPS'13
Akata, PAMI'16
Yang, ICLR'15
Romera-Paredes, ICML'15

- Pros/Cons
  - + Train for separability => higher accuracy.
  - - More complex and slower optimisation (Except [ Romera-Paredes, ICML'15 ])


IMAGES
$\mathcal{X}$
FEATURES
$\tilde{\mathcal{X}}$
ATTRIBUTES
$\tilde{\mathcal{Y}}$
LABELS
$\mathcal{Y}$
$F(x_i, y_i, w)$
$x_i$
$\theta(x_i)$
black
$y_i$
chimpanzee
$\varphi(y_i)$
$\theta(x_j)$
$\varphi(y_j)$
panda
$y_j$
$x_j$
white
$F(x_j, y_j, w)$

# ZSL Algorithm # 3: Deep Visual-Semantic Embedding




## Joint learning

A **non-linear model** that jointly optimize the mapping function where visual feature embedding and vector class embedding are learned in an end-to-end setting.

# References


- Akata et al, PAMI, Label-Embedding for Image Classification, 2015
- Ba et al, ICCV, Predicting Deep Zero-Shot Convolutional Neural Networks using Textual Descriptions, 2015
- Belkin et al, JMLR, Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples, 2006
- Frome et al, NIPS, DeViSE: A Deep Visual-Semantic Embedding Model, 2013
- Fu et al, IEEE T PAMI, Transductive Multi-View Zero-Shot Learning, 2015.
- Fu & Sigal, CVPR, Semi-supervised Vocabulary-informed Learning, 2016.
- Kordumova et al, ICMR, Pooling Objects for Recognizing Scenes Without Examples, 2016
- Gan et al, AAAI, Exploring Semantic Inter-Class Relationships (SIR) for Zero-Shot Action Recognition, 2015
- Lazardiou et al, ACL, Hubness and Pollution: Delving into Cross-Space Mapping for Zero-Shot Learning, 2016
- Mensink et al, CVPR, COSTA: Co-Occurrence Statistics for Zero-Shot Classification, 2014
- Mukherjee & Hospedales, EMNLP, Gaussian Visual-Linguistic Embedding for Zero-Shot Recognition, 2016
- Norouzi et al, ICLR, Zero-Shot Learning by Convex Combination of Semantic Embeddings, 2014

# References


- Ren et al, ACM MM, Joint image-text representation by gaussian visual semantic embedding, 2016
- Rohrbach et al, NIPS, Transfer Learning in a Transductive Setting, 2013
- Rohrbach et al, CVPR, Evaluating Knowledge Transfer and Zero-Shot Learning in a Large-Scale Setting, 2011
- Romera-Paredes & Torr, ICML, An embarrassingly simple approach to zero-shot learning, 2015
- Socher et al, NIPS, Zero-Shot Learning Through Cross-Modal Transfer, 2013
- Xu et al, ECCV, Multi-task zero-shot action recognition with prioritized data augmentation, 2016
- Xu et al, arXiv:1511.04458, Transductive Zero-Shot Action Recognition by Word-Vector Embedding
- Yang & Hospedales, ICLR, A Unified Perspective on Multi-Domain and Multi-Task Learning, 2015
- Yang & Hospedales, CVPR, Multivariate Regression on the Grassmannian for Predicting Novel Domains, 2016
- Yang & Hospedales, arXiv:1605.06391, Deep Multi-task Representation Learning: A Tensor Factorisation Approach, 2016
- Yang & Hospedales, arXiv, Unifying Multi-Domain Multi-Task Learning: Tensor and Neural Network Perspectives, 2016


Some of the presentation slides are based on Timothy Hospedales talks at `https://yandexdataschool.com/conference`

## Thank you for your attention!

## Questions & Discussion