

# CS233, CME251: Geometric and Topological Data Analysis

Leonidas Guibas  
Computer Science Department  
Stanford University



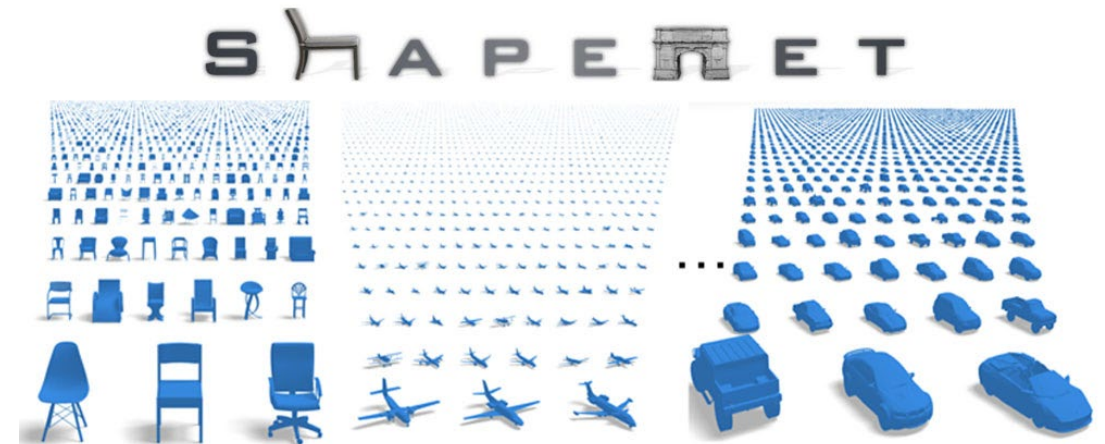
Lecture 4  
6 April 2022



# Last Time: Visual Datasets

# DataSets

- ◆ Relate geometry and topology to data semantics
- ◆ Explain how big visual datasets including ImageNet and ShapeNet are organized



## What is WordNet?

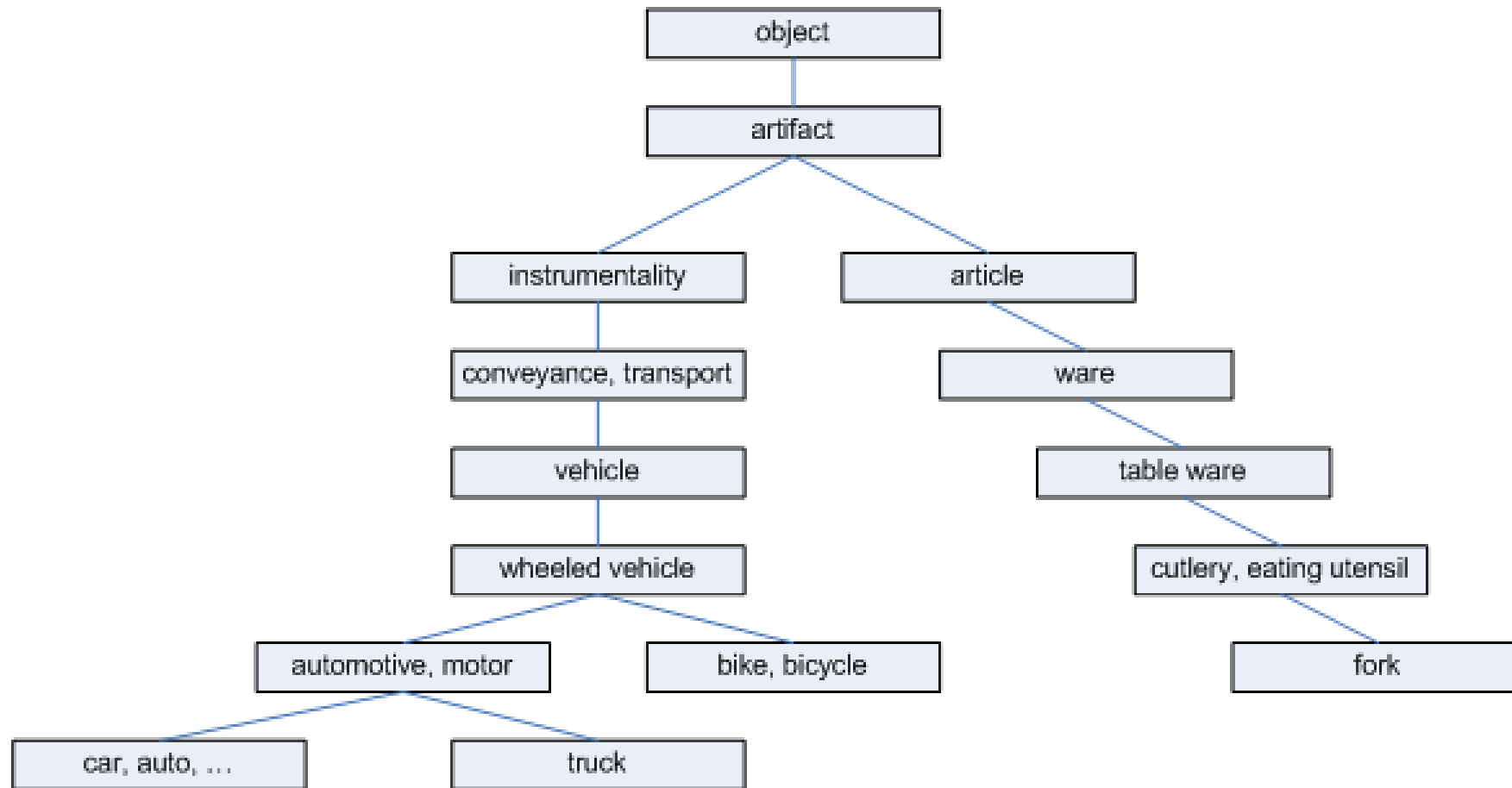


Original paper  
by  
**[George  
Miller, et al  
1990]** cited  
over 5,000  
times

Organizes over  
150,000 words  
into 117,000  
categories  
called *synsets*.

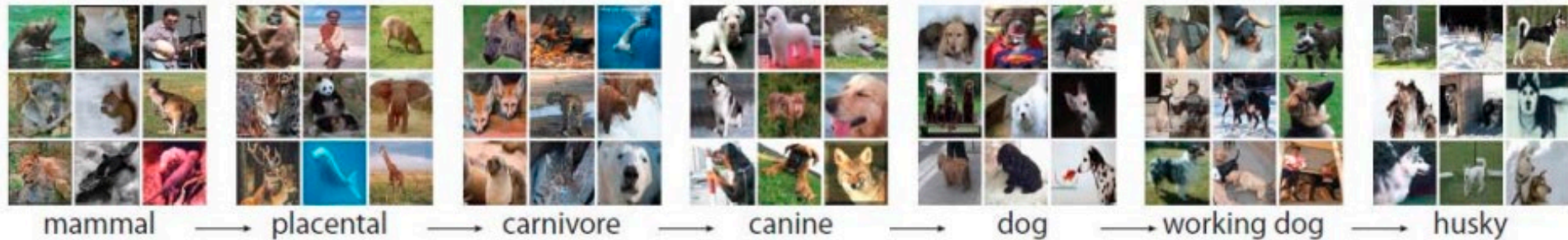
Establishes  
ontological and  
lexical  
relationships in  
NLP and related  
tasks.

# Taxonomy: is-a Relationship



# IMAGENET is a knowledge ontology

- Taxonomy (with WordNet backbone)



- [S: \(n\) Eskimo dog, husky](#) (breed of heavy-coated Arctic sled dog)
    - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
    - [S: \(n\) working dog](#) (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
    - [S: \(n\) dog, domestic dog, Canis familiaris](#) (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) "*the dog barked all night*"
      - [S: \(n\) canine, canid](#) (any of various fissiped mammals with nonretractile claws and typically long muzzles)
        - [S: \(n\) carnivore](#) (a terrestrial or aquatic flesh-eating mammal) "*terrestrial carnivores have four or five clawed digits on each limb*"
- 
- [S: \(n\) placental, placental mammal, eutherian, eutherian mammal](#) (mammals having a placenta; all mammals except monotremes and marsupials)
    - [S: \(n\) mammal, mammalian](#) (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
      - [S: \(n\) vertebrate, craniate](#) (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
        - [S: \(n\) chordate](#) (any animal of the phylum Chordata having a notochord or spinal column)
          - [S: \(n\) animal, animate being, beast, brute, creature, fauna](#) (a living organism characterized by voluntary movement)
            - [S: \(n\) organism, being](#) (a living thing that has (or can develop) the ability to act or function independently)
              - [S: \(n\) living thing, animate thing](#) (a living (or once living) entity)
                - [S: \(n\) whole, unit](#) (an assemblage of parts that is regarded as a single entity) "*how big is that part compared to the whole?*"; "*the team is a unit*"
                - [S: \(n\) object, physical object](#) (a tangible and visible entity; an entity that can cast a shadow) "*it was full of rackets, balls and other objects*"
                  - [S: \(n\) physical entity](#) (an entity that has physical existence)
                    - [S: \(n\) entity](#) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Slide Credit: Fei-Fei Li, Jia Deng

# ImageNet Challenge Tasks

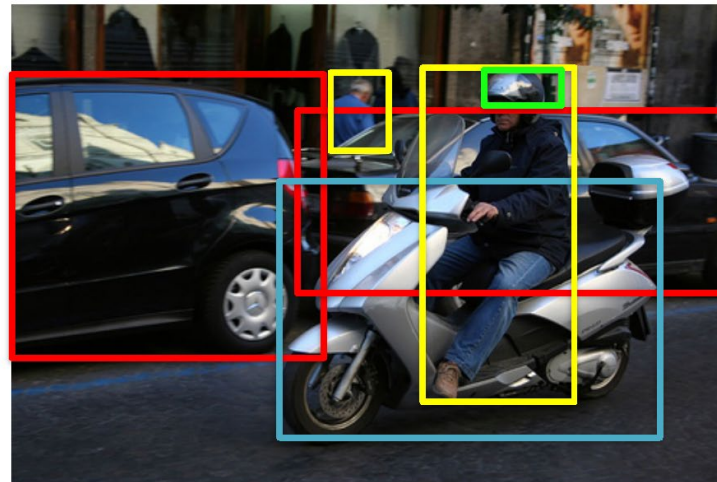
Steel drum



Objects: 1000 classes  
Training: 1.2M images  
Validation: 50K images  
Test: 100K images

Classification

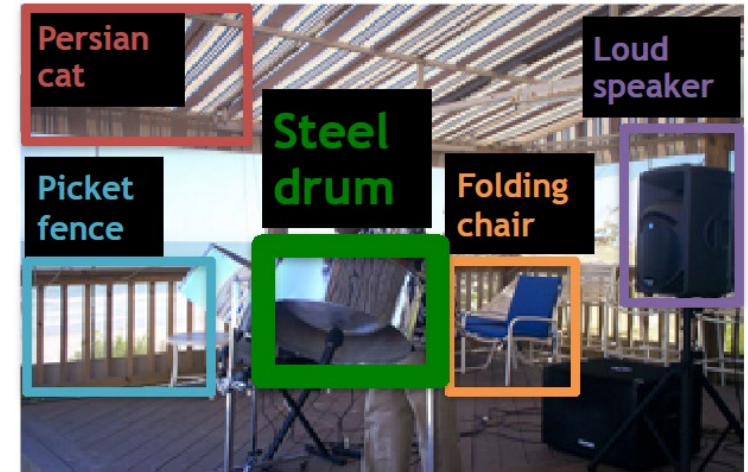
Classification + Localization



Person  
Car  
Motorcycle  
Helmet

Objects: 200 classes  
Training: 450K images, 470K bounding boxes  
Validation: 20K images, all bounding boxes  
Test: 40K images, all bounding boxes

Output



Object Detection

# 3D Opportunities: Encoding Knowledge



Among all digital representations we have of a real artifact, 3D is the most faithful to the actual physical object

# ShapeNet (>3M Models) <https://www.shapenet.org/>

The screenshot shows the ShapeNet website interface. At the top, there is a search bar with the text "Search" and a magnifying glass icon, followed by "Options" with a dropdown arrow. To the right are navigation links: "Home", "About", "Download", and "Statistics". Below the search bar, the word "chair" is displayed in bold, followed by a definition: "a seat for one person, with a support for the back; 'he put his coat over the back of the chair and sat down'". Below the definition are links for "ImageNet" and "MetaData".

On the left side, there is a "Choose a taxonomy:" section with a dropdown menu set to "ShapeNetCore". Below this is a scrollable list of categories with their respective counts, such as "airplane,aeroplane,plane(12,4501)", "aquarium,fish tank,marine museum(0,4)", and "chair(23,7083)".

On the right side, there is a "Synset models" section. It displays "Displaying 1 to 40 of 7080" models. Below this is a pagination bar with numbers 1 through 177. The main area shows a grid of 40 3D chair models, each with a label underneath. The labels include: club chair, cantilever chair, armchair, straight chair, straight chair, club chair, deck chair, rex chair, straight chair, club chair, club chair, swivel chair, butterfly chair, armchair, armchair, club chair, recliner, cantilever chair, swivel chair, swivel chair, armchair, folding chair, rocking chair, club chair, and several more.



Stanford:  
Leonidas Guibas  
Pat Hanrahan  
Silvio Savarese



Princeton:  
Tom Funkhouser  
Jianxiong Xiao



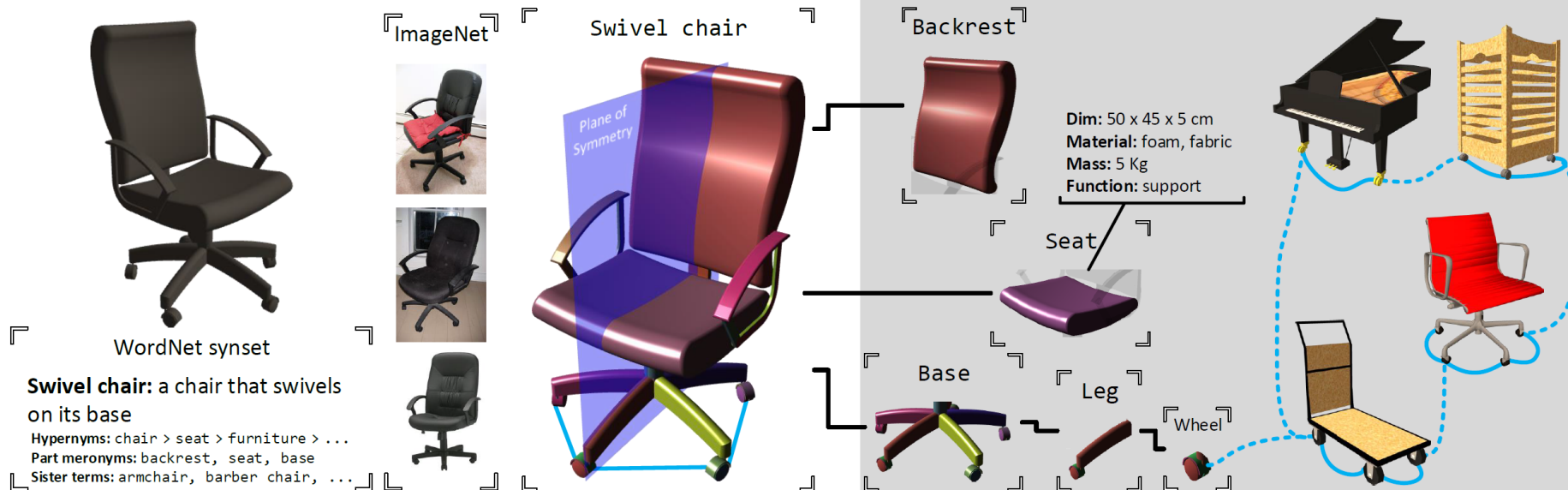
UT Austin:  
Qixing Huang

# Object Knowledge: ShapeNet

Parts, symmetries, keywords, physical properties, materials, affordances, ...

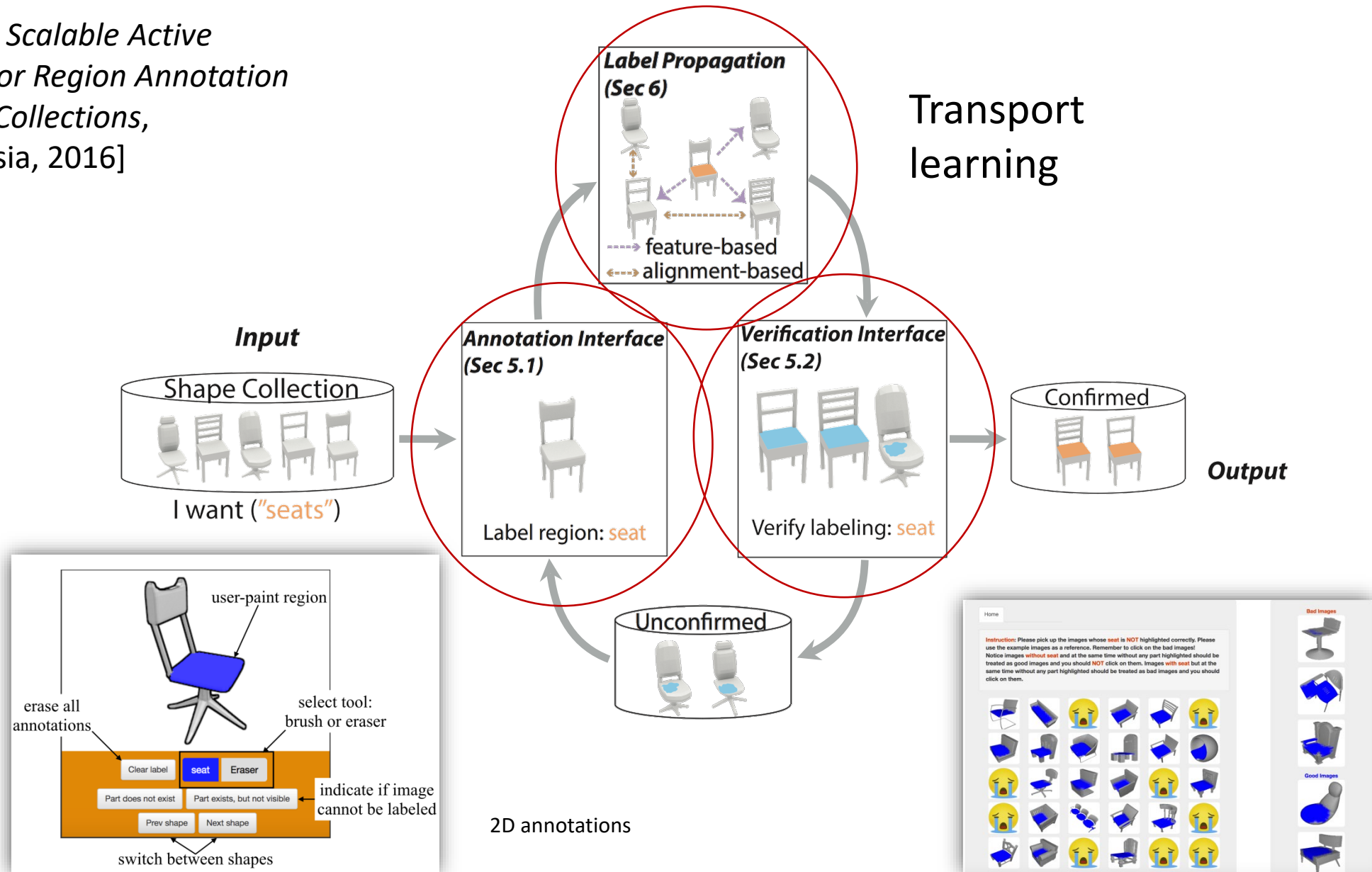


Link to WordNet Taxonomy   Alignment+Symmetry   Part Hierarchy   Part Correspondences

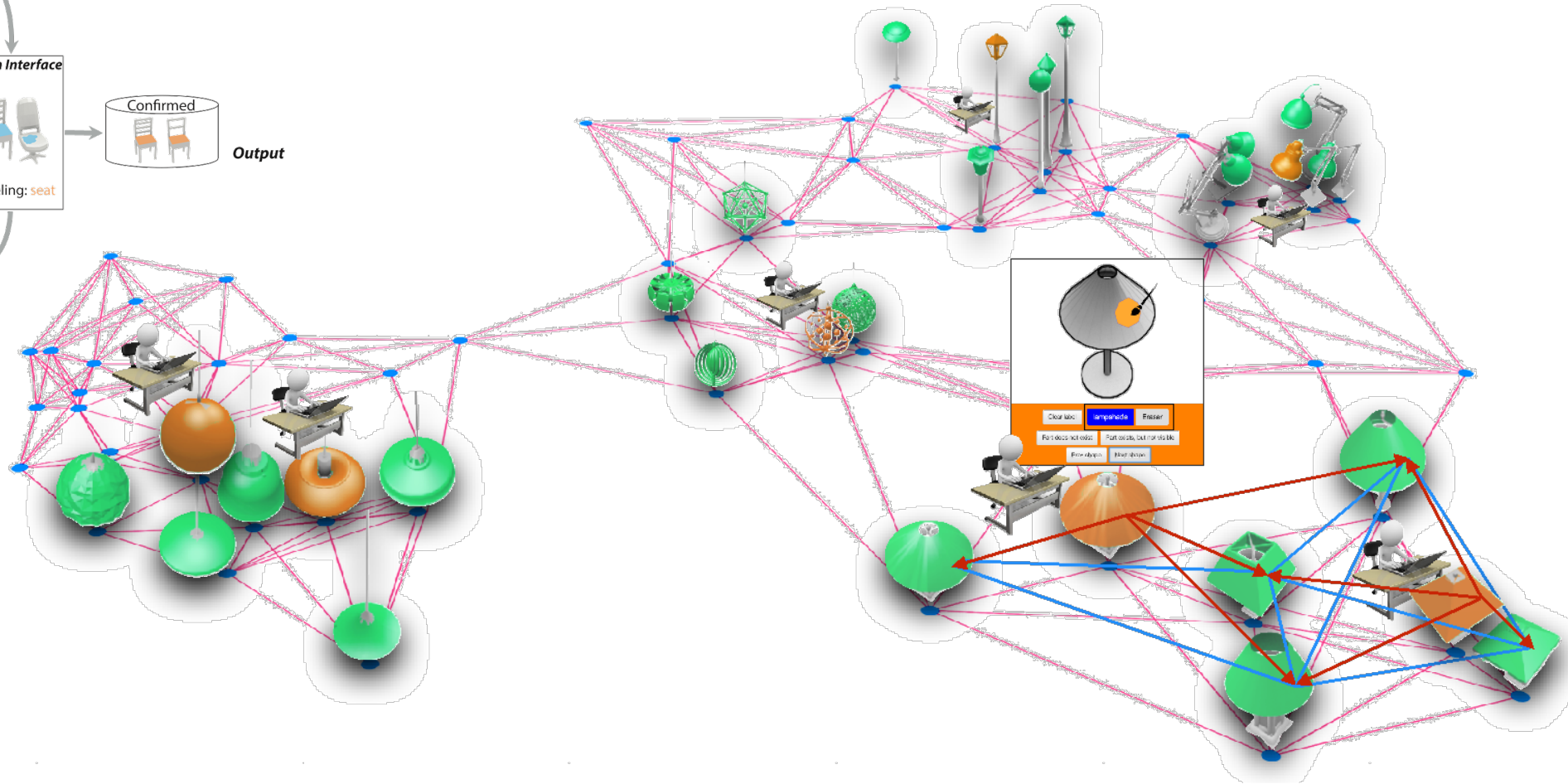
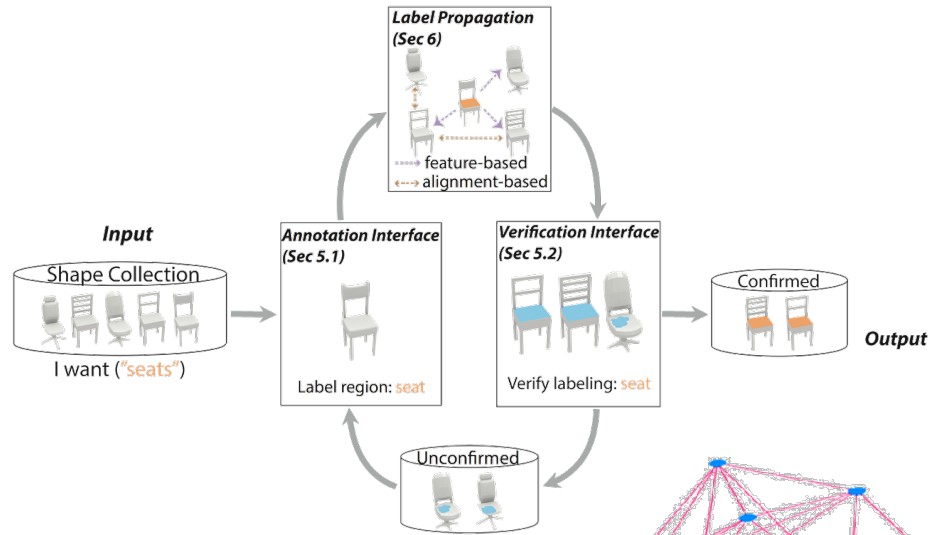


# Model Part Annotation

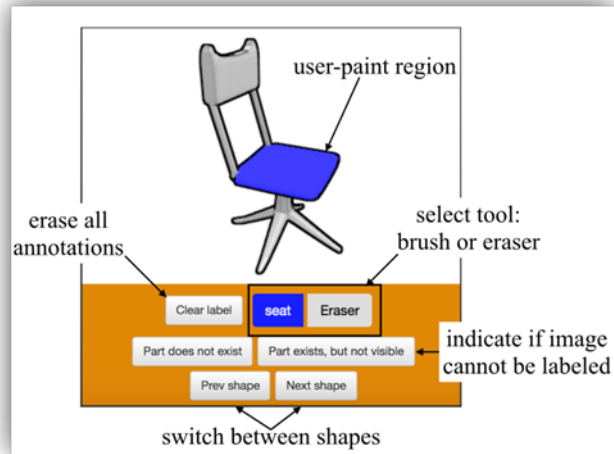
[Li Yi et al., *A Scalable Active Framework for Region Annotation in 3D Shape Collections*, SIGGRAPH Asia, 2016]



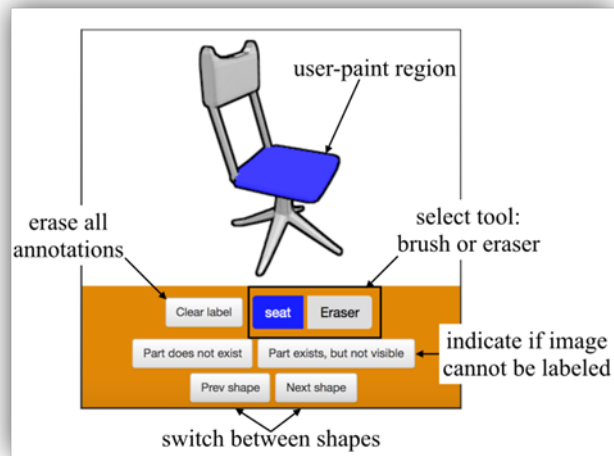
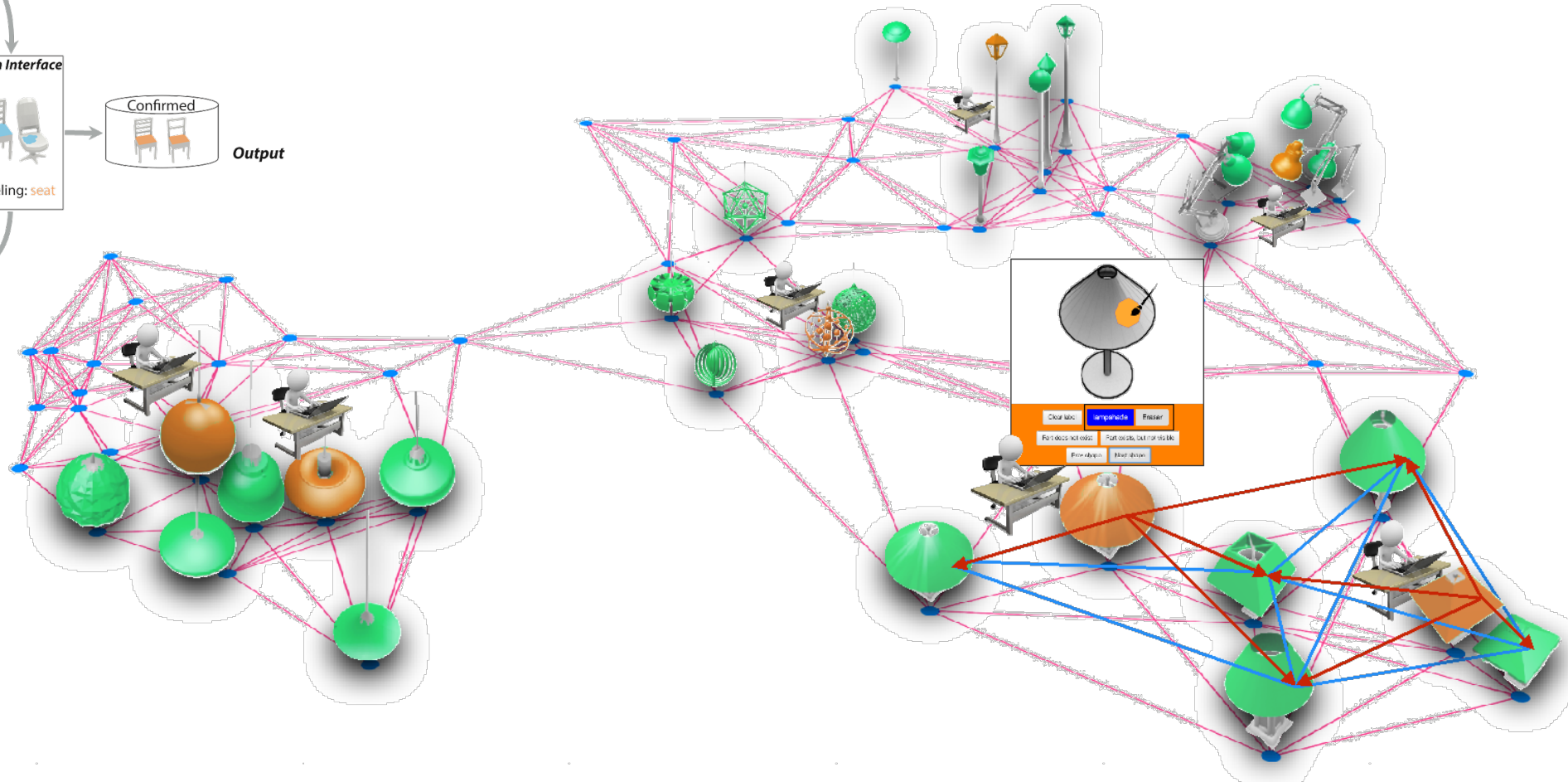
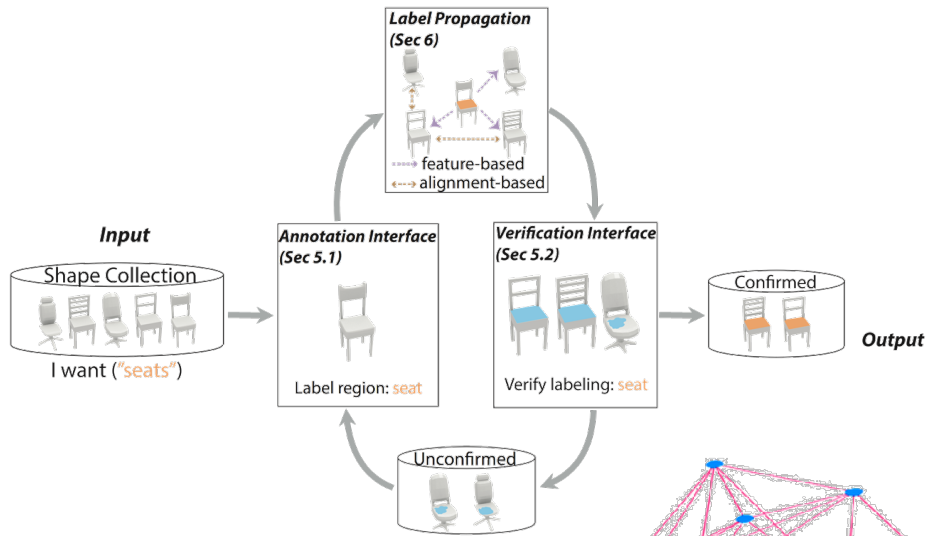
# Crowdsourcing and Algorithmic Propagation



Joint learning between humans and algorithms

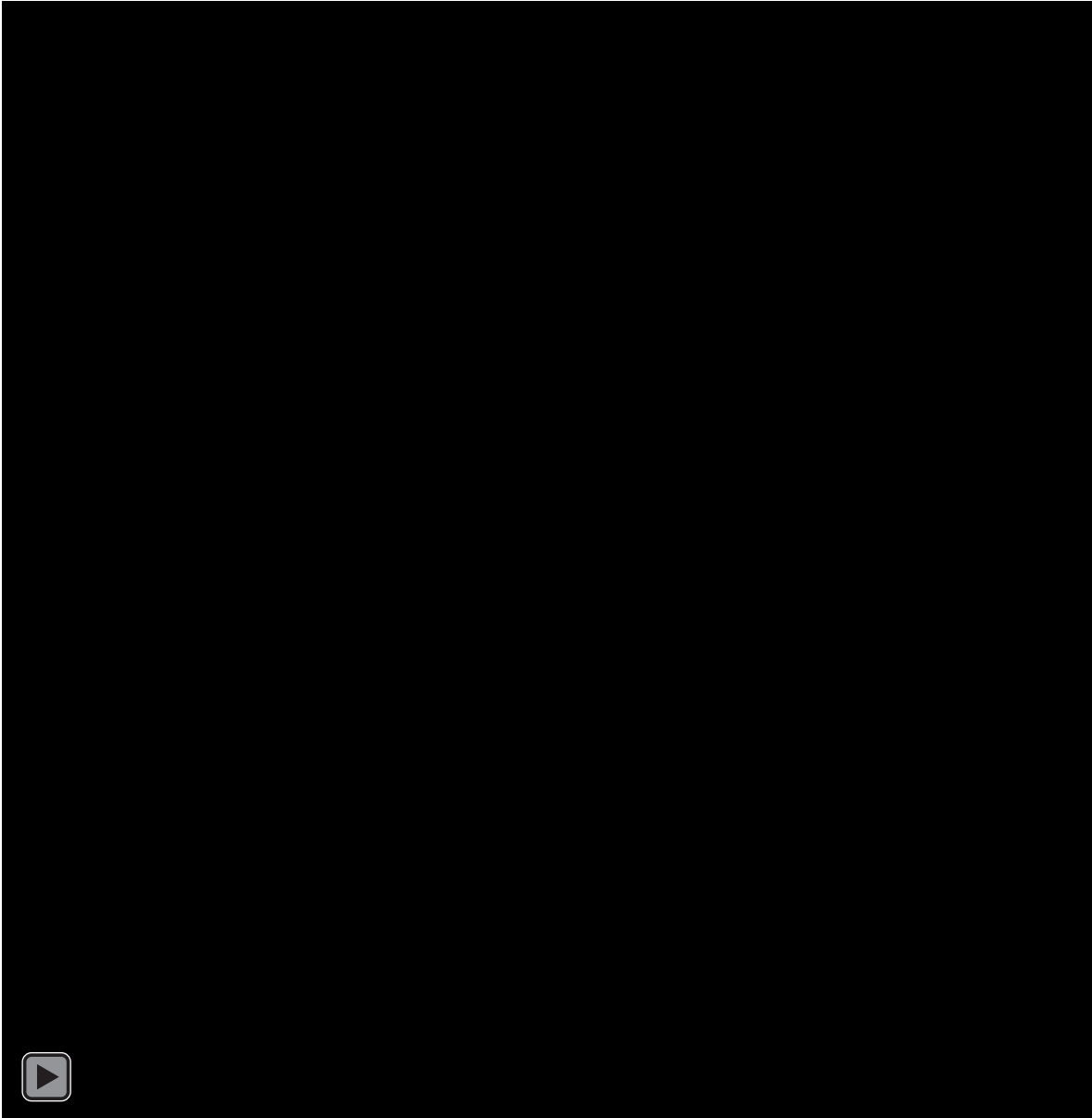


# Crowdsourcing and Algorithmic Propagation

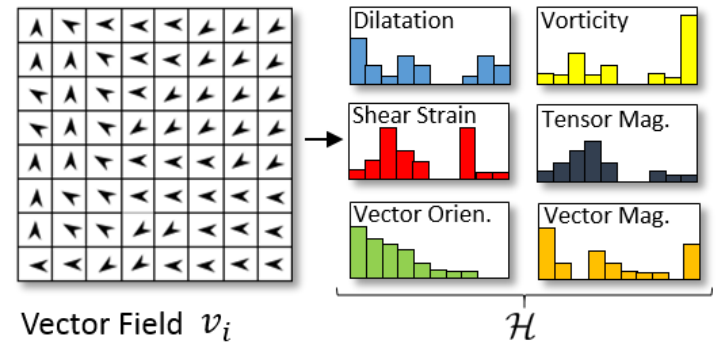


Joint learning between humans and algorithms

# Object Function / Interaction Knowledge



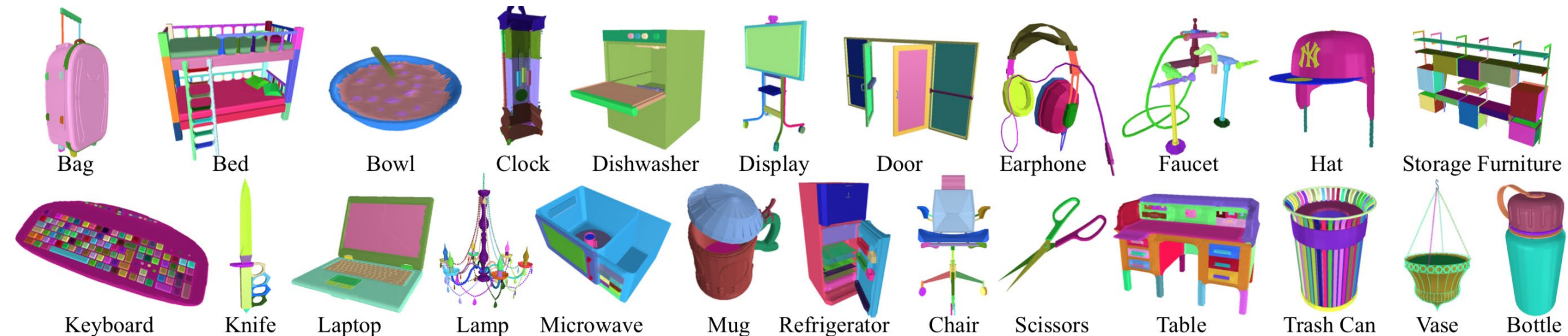
Vector Field to Histograms



# Digitizing Human Physical Interactions and Manual Skills



# PartNet: Fine-Grained Parts



- ◆ Subset of ShapeNetCore

- ◆ 24 common indoor categories, 26,671 shapes, 573,585 parts
- ◆ Avg 18 Part/shape, Max 230

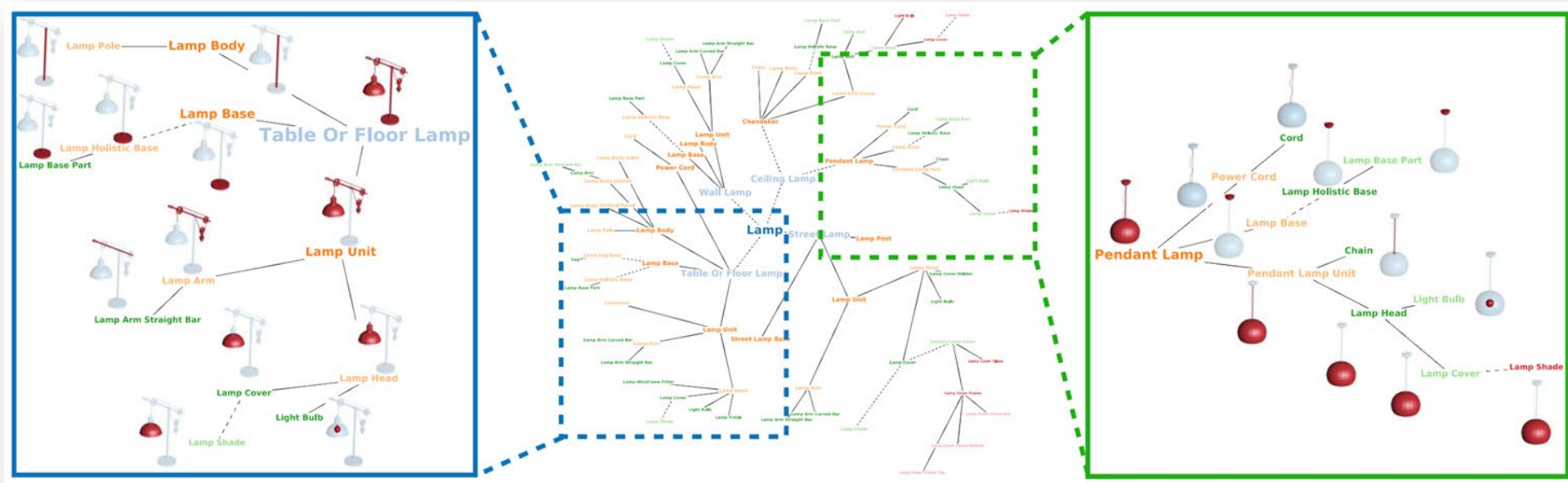
- ◆ Human-annotated:

- ◆ more fine-grained parts + instance-level parts

K. Mo, S. Zhu, A. Chang, L. Yi, S. Tripathi, L. Guibas and H. Su. *PartNet: A Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding*. CVPR 2019.

# PartNet Data: Hierarchical Segmentation

- ◆ Provides hierarchical segmentation of shapes: parts at multiple scales
- ◆ All shapes from the same category conform to a consistent part template



“and-or” graphs



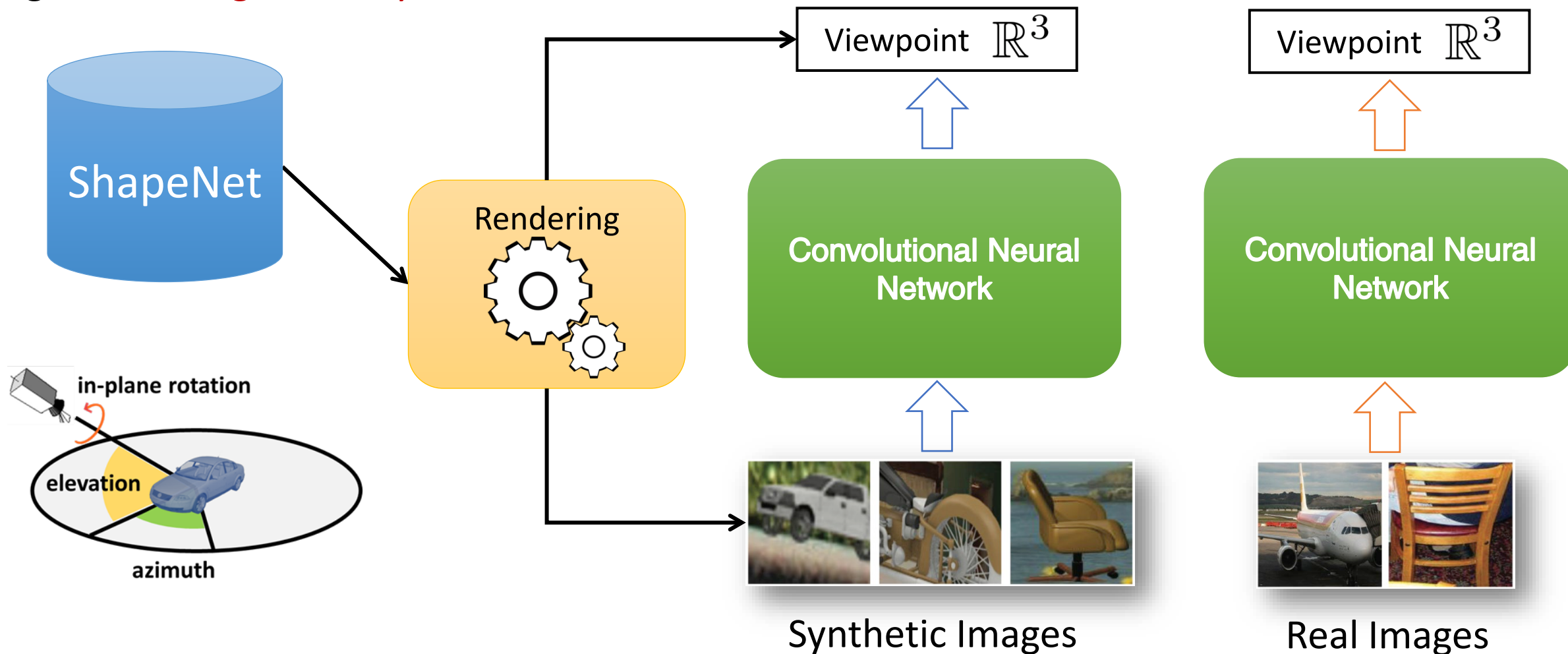
# PartNet-Mobility: Annotate Part Motions (SAPIEN)



F. Xiang, Y. Qin, K. Mo, Y. Xia, H. Zhu, F. Liu, M. Liu, H. Jiang, Y. Yuan, H. Wang, L. Yi, L. Guibas and H. Su,  
*SAPIEN: A SimulATED Part-based Interactive Environment*. CVPR 2020.

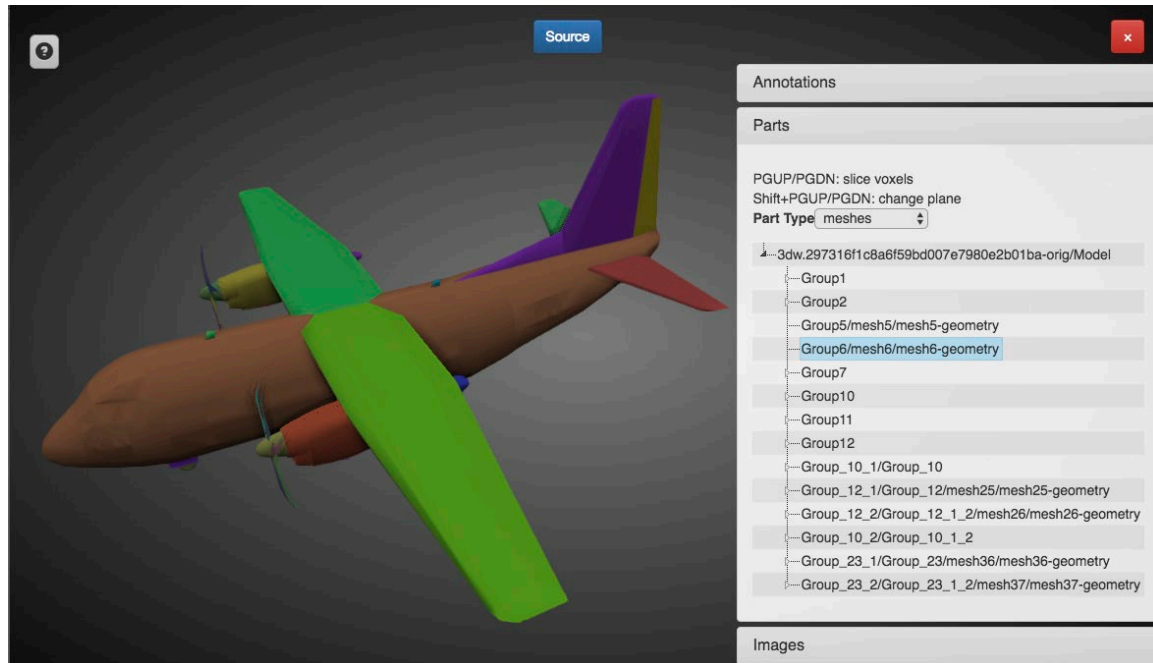
# Synthetic Data for ML

Leverage ShapeNet annotations to generate **large-scale synthetic** dataset



# Exploit Scene Graphs of Models in the Wild

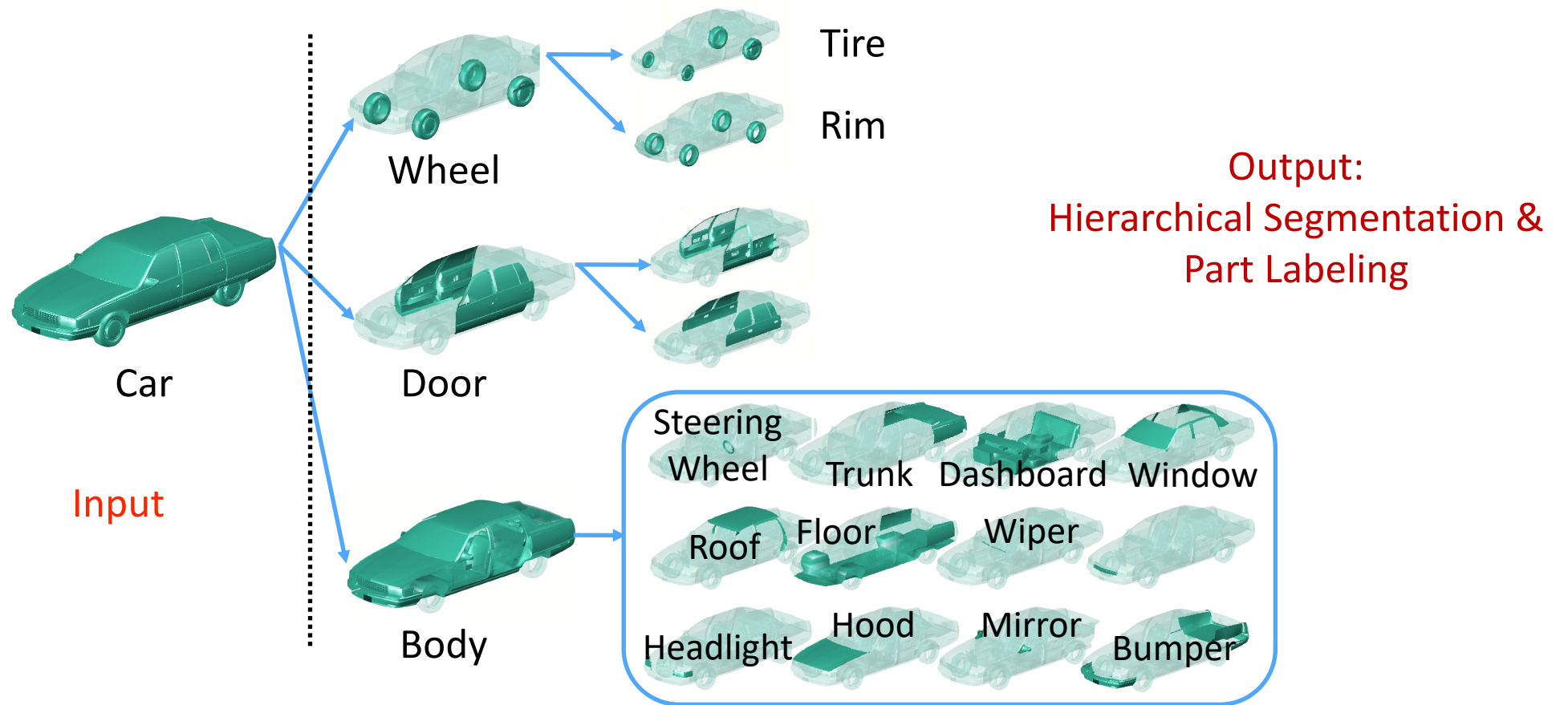
CAD data include *scene graphs*:  
Part geometry + Hierarchical structure



Inconsistent and unlabeled

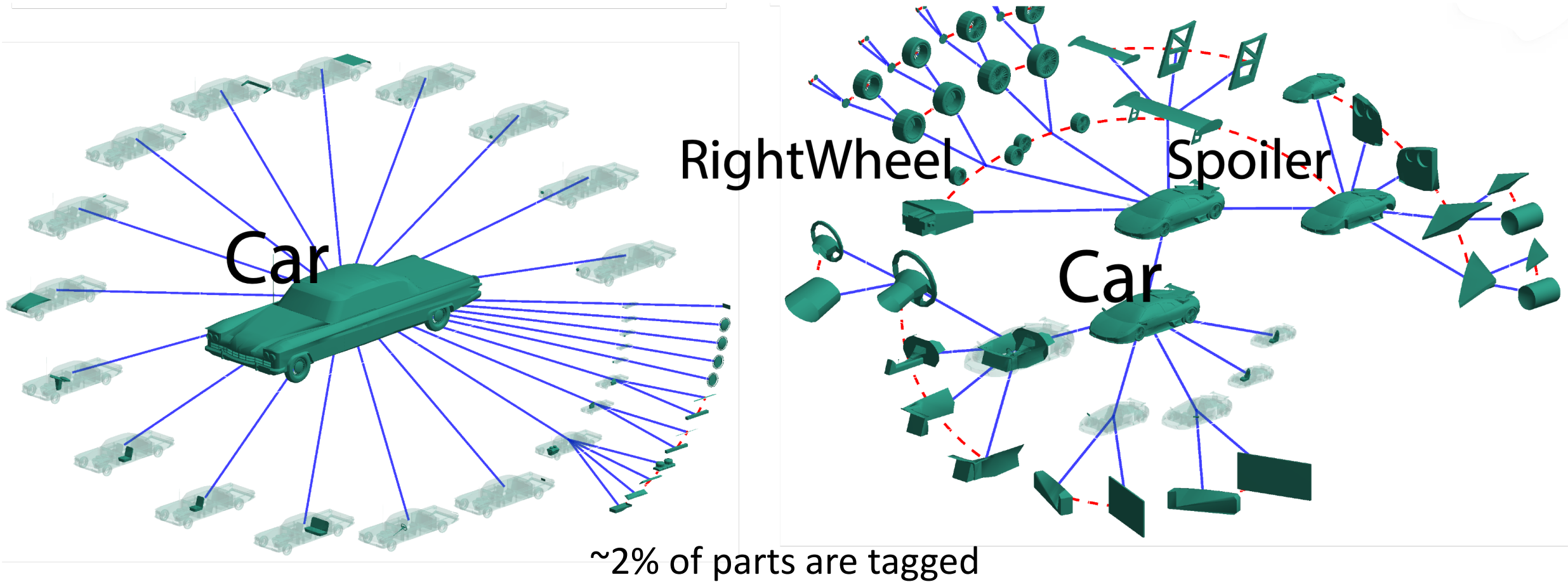
# An Application of Horizontal Networks

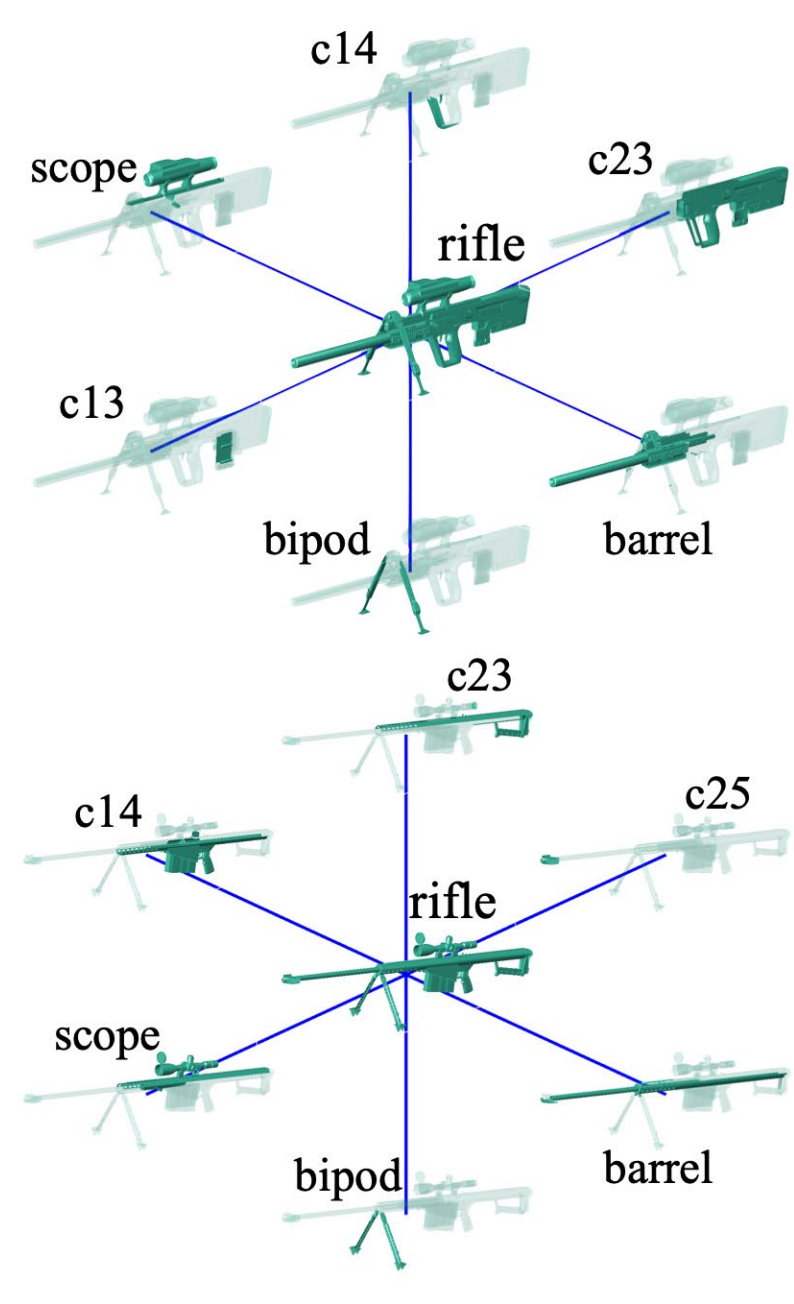
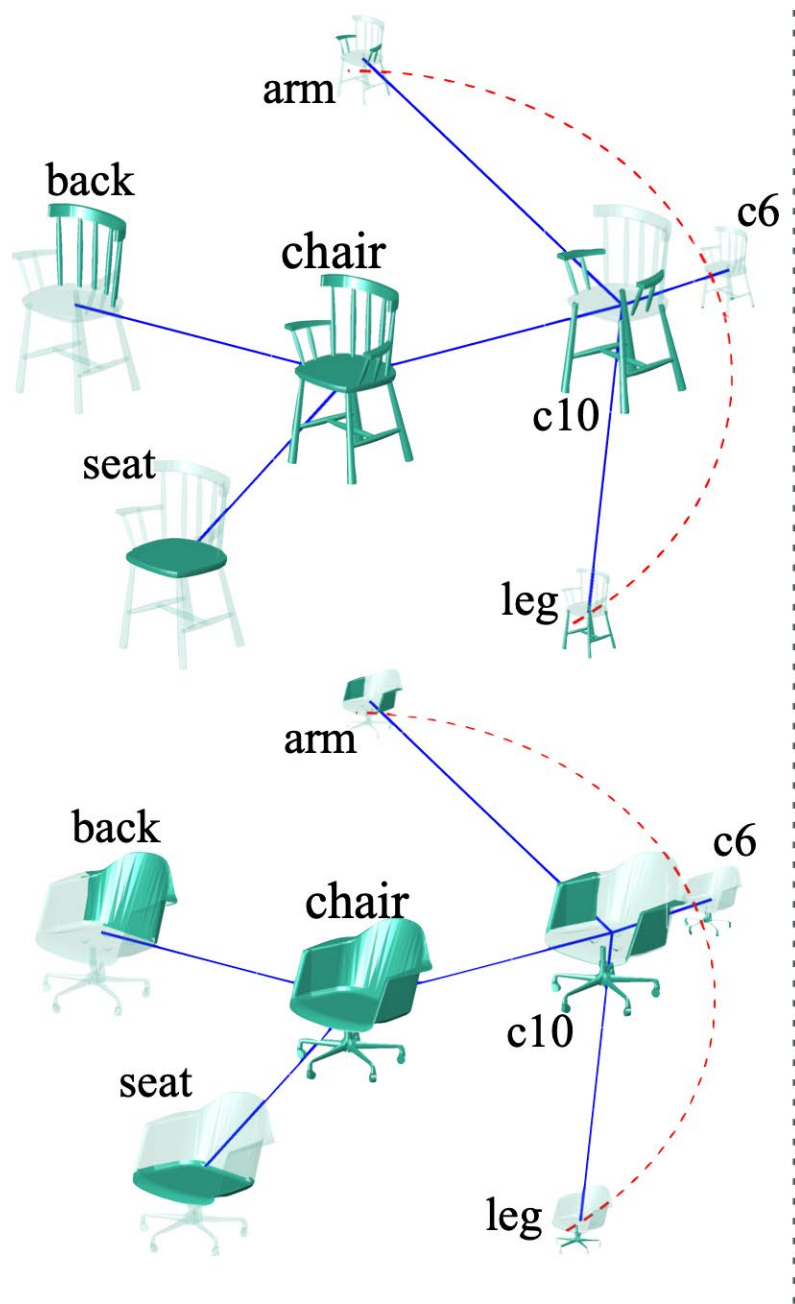
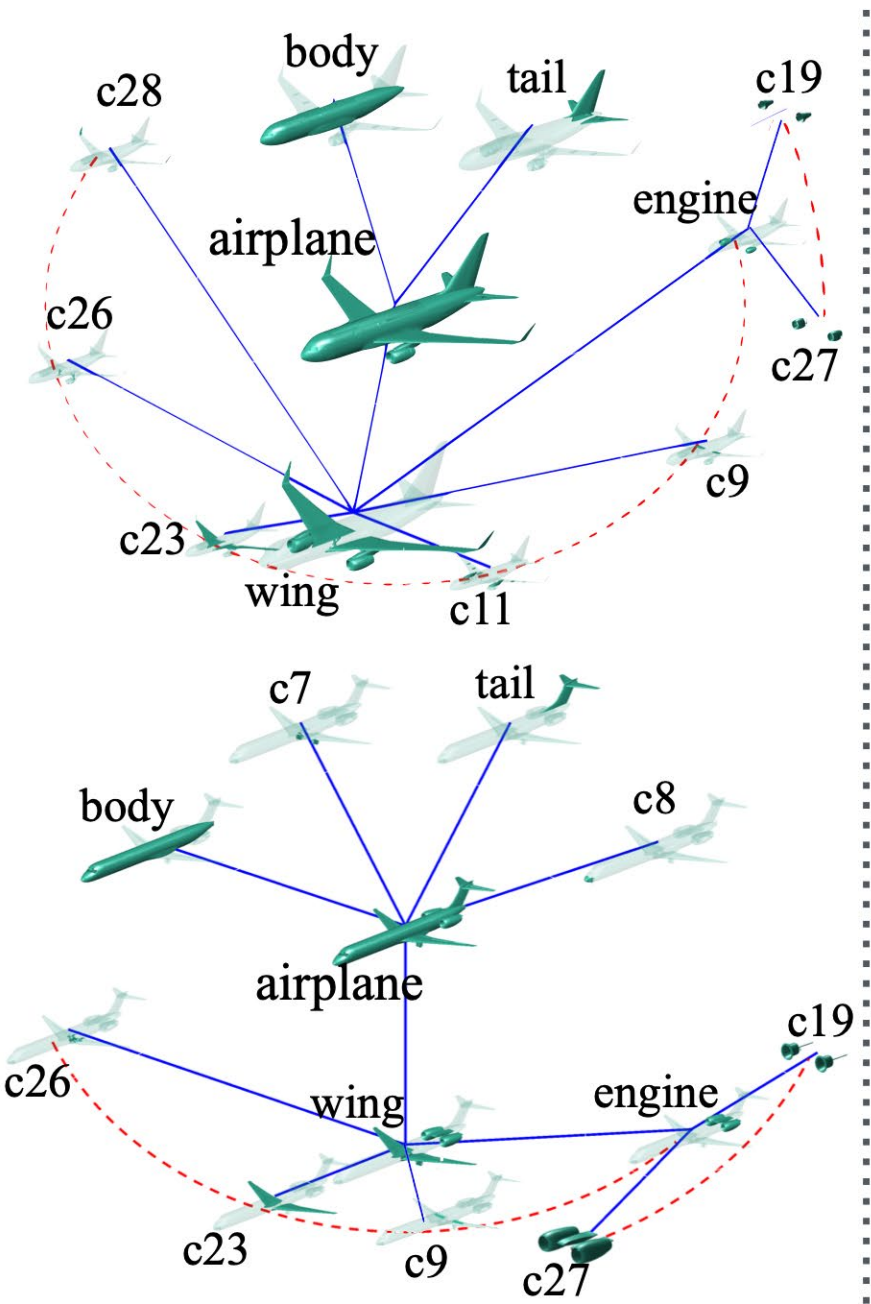
- Learning hierarchical shape segmentation and labeling in a weakly supervised manner from online repositories
- Based on “horizontal” information diffusion



# Challenges - Heterogeneous Data

*Different Parts, Different Hierarchy, Sparse Tagging*



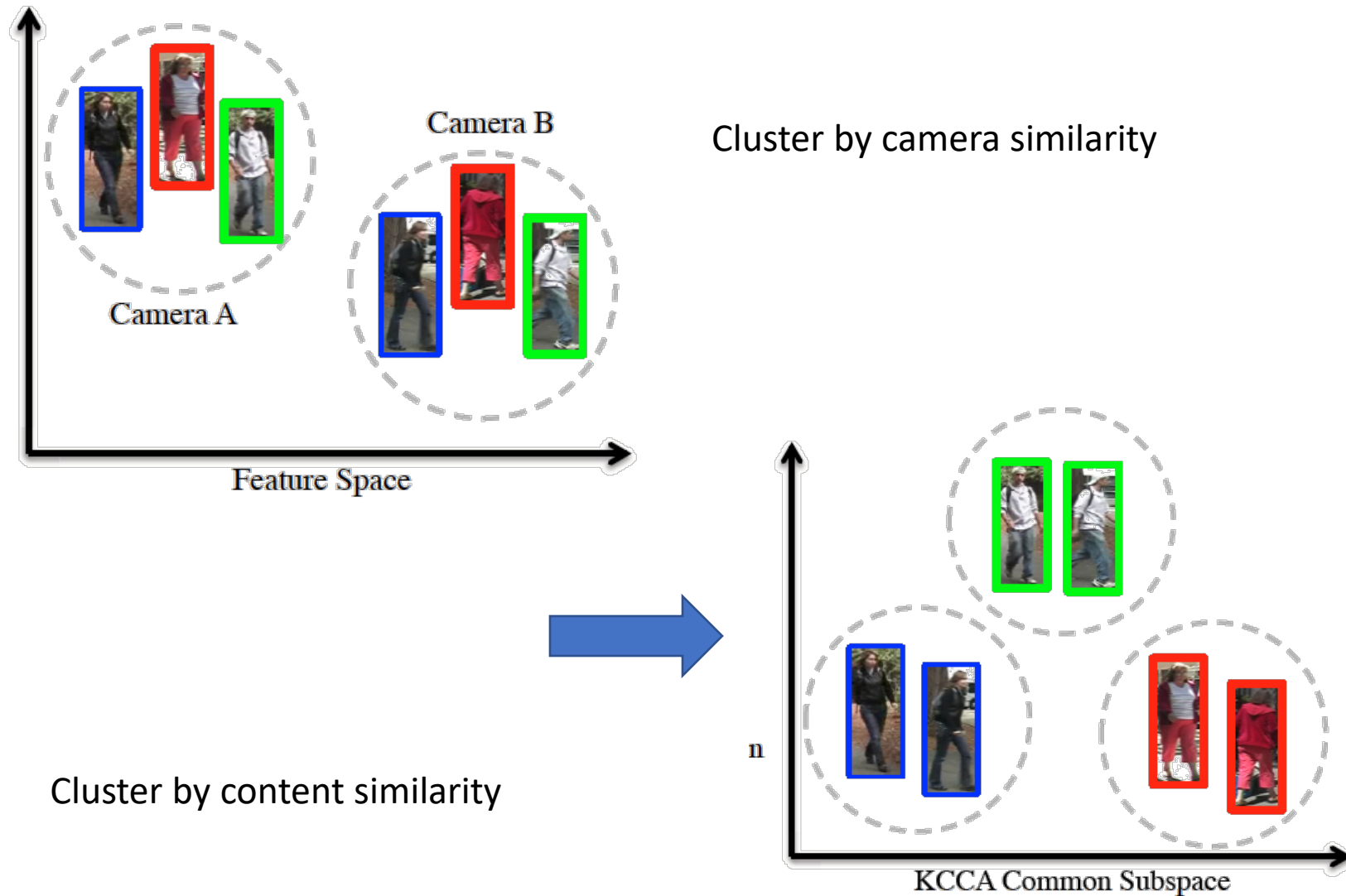


# Today: CCA & MDS

Canonical Correlation Analysis  
Multi-dimensional Scaling

# CCA: Shared Structure Across Different Data Sets

# Different Views of the Same Data



# Covariance and Correlation

# Covariance and Correlation

- **Pearson correlation** coefficient between two random variables  $X, Y$

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_x)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

$$\mu_X = \bar{x}$$

mean

$$\sigma_X$$

standard deviation

- Note that, by the Cauchy-Schwarz inequality

$$E[(X - \mu_x)(Y - \mu_Y)]^2 \leq E[(X - \mu_x)^2]E[(Y - \mu_Y)^2]$$

so

$$-1 \leq \rho_{X,Y} \leq 1$$

# Covariance and Correlation

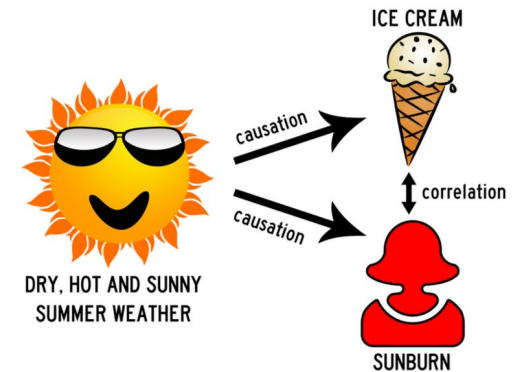
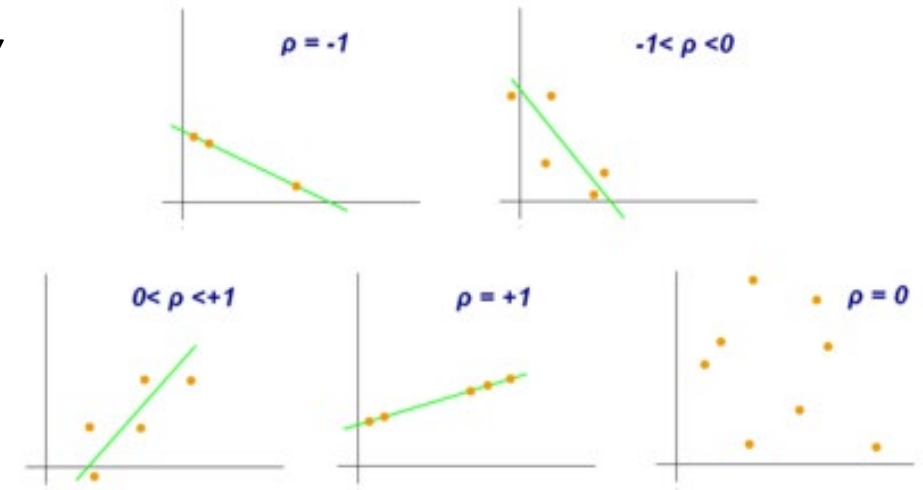
- Correlation measures a **linear association** between  $X, Y$

$$\rho = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$$

- Note that,  $X, Y$  independent, then  $\text{corr}(X, Y) = 0$ 
  - but the opposite is not true

[ $U$  uniform in  $[0, 2\pi]$ ,  $X = \sin U$ ,  $Y = \cos U$ ]

- High correlation not the same as causality



# Empirical Correlation

- Empirical version, for  $n$  measurements  $x_i, y_i$  of  $X$  and  $Y$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n s_x s_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

- We'll use centered versions,

$$\bar{x} = 0, \bar{y} = 0 \quad \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}}$$

$$= \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

# Canonical Correlations for Two Sets of Variates

# Canonical Correlations

- Canonical correlation analysis seeks a pair of linear transformations, one for each of the sets of variables  $X, Y$ , such that when the set of variables is transformed, the corresponding coordinates are maximally correlated.

$$S = ((\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)) \quad \begin{array}{l} S_x = (\mathbf{x}_1, \dots, \mathbf{x}_n) \\ S_y = (\mathbf{y}_1, \dots, \mathbf{y}_n) \end{array}$$

- Consider projections of  $X$  and  $Y$

$$\begin{array}{l} \mathbf{x} \rightarrow \langle \mathbf{w}_x, \mathbf{x} \rangle \\ \mathbf{y} \rightarrow \langle \mathbf{w}_y, \mathbf{y} \rangle \end{array}$$

- So we get

$$\begin{array}{l} S_{x, \mathbf{w}_x} = (\langle \mathbf{w}_x, \mathbf{x}_1 \rangle, \langle \mathbf{w}_x, \mathbf{x}_2 \rangle, \dots, \langle \mathbf{w}_x, \mathbf{x}_n \rangle) \\ S_{y, \mathbf{w}_y} = (\langle \mathbf{w}_y, \mathbf{y}_1 \rangle, \langle \mathbf{w}_y, \mathbf{y}_2 \rangle, \dots, \langle \mathbf{w}_y, \mathbf{y}_n \rangle) \end{array}$$

# 1<sup>st</sup> Canonical Correlation

- We choose  $w_x, w_y$  to maximize the correlation between these two vectors

$$\rho = (\rho_1 =) \max_{w_x, w_y} \text{corr}(S_x w_x, S_y w_y) = \max_{w_x, w_y} \frac{\langle S_x w_x, S_y w_y \rangle}{\|S_x w_x\| \|S_y w_y\|}$$

- Or, in empirical form, if we write  $E[f(\mathbf{x}, \mathbf{y})] = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i, \mathbf{y}_i)$

We can re-write the correlation expression as

$$\rho = \max_{w_x, w_y} \frac{E[\langle w_x, \mathbf{x} \rangle \langle w_y, \mathbf{y} \rangle]}{\sqrt{E[\langle w_x, \mathbf{x} \rangle^2] E[\langle w_y, \mathbf{y} \rangle^2]}}$$

# Covariance Formulation

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{E[\langle \mathbf{w}_x, \mathbf{x} \rangle \langle \mathbf{w}_y, \mathbf{y} \rangle]}{\sqrt{E[\langle \mathbf{w}_x, \mathbf{x} \rangle^2] E[\langle \mathbf{w}_y, \mathbf{y} \rangle^2]}}$$

- Can re-write as

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{E[\mathbf{w}_x^T \mathbf{x} \mathbf{y}^T \mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x^T \mathbf{x} \mathbf{x}^T \mathbf{w}_x] E[\mathbf{w}_y^T \mathbf{y} \mathbf{y}^T \mathbf{w}_y]}}$$

- so that

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x^T E[\mathbf{x} \mathbf{y}^T] \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T E[\mathbf{x} \mathbf{x}^T] \mathbf{w}_x \mathbf{w}_y^T E[\mathbf{y} \mathbf{y}^T] \mathbf{w}_y}}$$

- If we now write

$$C(\mathbf{x}, \mathbf{y}) = E \left[ \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^T \right] = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = C$$

- we get

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x^T C_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T C_{xx} \mathbf{w}_x \mathbf{w}_y^T C_{yy} \mathbf{w}_y}}$$

# Solution by Eigenanalysis

- Note that the expression  $\frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}}$

is invariant to re-scalings of  $\mathbf{w}_x$  or  $\mathbf{w}_y$

- We can therefore solve the optimization problem

$$\max_{\mathbf{w}_x, \mathbf{w}_y} \mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y$$

subject to the constraints

$$\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x = 1 \text{ and } \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y = 1.$$

# Lagrange Multipliers

$$L(\lambda, \mathbf{w}_x, \mathbf{w}_y) = \mathbf{w}_x^T C_{xy} \mathbf{w}_y - \frac{\lambda_x}{2} (\mathbf{w}_x^T C_{xx} \mathbf{w}_x - 1) - \frac{\lambda_y}{2} (\mathbf{w}_y^T C_{yy} \mathbf{w}_y - 1)$$

- Setting derivatives of  $L$  to 0 w.r.t.  $\mathbf{w}_x, \mathbf{w}_y$  we get (after some manipulation)

$$\frac{\partial L}{\partial \mathbf{w}_x} = C_{xy} \mathbf{w}_y - \lambda_x C_{xx} \mathbf{w}_x = 0 \quad \frac{\partial L}{\partial \mathbf{w}_y} = C_{yx} \mathbf{w}_x - \lambda_y C_{yy} \mathbf{w}_y = 0$$

multiply by  $\mathbf{w}_x^T$ , multiply by  $\mathbf{w}_y^T$   
subtract

$$\begin{aligned} 0 &= \mathbf{w}_x^T C_{xy} \mathbf{w}_y - \mathbf{w}_x^T \lambda_x C_{xx} \mathbf{w}_y - \mathbf{w}_y^T C_{yx} \mathbf{w}_x + \mathbf{w}_y^T \lambda_y C_{yy} \mathbf{w}_y \\ &= \mathbf{w}_y^T \lambda_y C_{yy} \mathbf{w}_y - \mathbf{w}_x^T \lambda_x C_{xx} \mathbf{w}_y \end{aligned}$$

# Lagrange Multipliers

- Leading to  $\lambda_x \mathbf{w}_x^T C_{xx} \mathbf{w}_x = \lambda_y \mathbf{w}_y^T C_{yy} \mathbf{w}_y$  and  $\lambda_x = \lambda_y (= \lambda)$

- Assuming  $C_{yy}$  is invertible, we can use the  $w_y$  derivative constraint above to get

$$\mathbf{w}_y = \frac{C_{yy}^{-1} C_{yx} \mathbf{w}_x}{\lambda}$$

$$\frac{\partial L}{\partial \mathbf{w}_x} = C_{xy} \mathbf{w}_y - \lambda_x C_{xx} \mathbf{w}_x = 0$$

$$\frac{\partial L}{\partial \mathbf{w}_y} = C_{yx} \mathbf{w}_x - \lambda_y C_{yy} \mathbf{w}_y = 0$$

- so now from the  $w_x$  derivative constraint we get

$$\frac{\partial L}{\partial \mathbf{w}_x} = C_{xy} \mathbf{w}_y - \lambda_x C_{xx} \mathbf{w}_x = 0$$

- or

# Eigenvalue Problem

- A generalized eigenvalue problem

$$C_{xy}C_{yy}^{-1}C_{yx}w_x = \lambda^2 C_{xx}w_x$$
$$Ax = \lambda Bx$$

- Can symmetrically also get

$$C_{yx}C_{xx}^{-1}C_{xy}w_y = \lambda^2 C_{yy}w_y$$

- One can go back and forth between  $w_x$  and  $w_y$   $w_y = \frac{C_{yy}^{-1}C_{yx}w_x}{\lambda}$

# Solving the Eigenvalue Problem

- If  $C_{xx}$  is invertible, then can reduce to a standard symmetric eigenvalue problem

$$C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \mathbf{w}_x = \lambda^2 \mathbf{w}_x$$

- Numerically, all these inversions and multiplications lose precision
- Alternate approach:
  - $C_{xx}$  and  $C_{yy}$  are symmetric positive definite; use complete Cholesky decomposition –  $R_{xx}$  lower triangular so that

$$C_{xx} = R_{xx} R_{xx}^T$$

- Now let  $\mathbf{u}_x = R_{xx}^T \mathbf{w}_x$

- Can re-write

$$C_{xy} C_{yy}^{-1} C_{yx} R_{xx}^{-T} \mathbf{u}_x = \lambda^2 R_{xx} \mathbf{u}_x$$

$$R_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} R_{xx}^{-T} \mathbf{u}_x = \lambda^2 \mathbf{u}_x$$

# Centered Variables, CCA vs PCA

- For centered variates, the covariance  $C_{xy} = x^T y$
- Canonical correlation analysis attempts to answer the question “which directions accounts for most of the covariance between the two data sets?” The goal is to find directions  $w_x, w_y$  so as to maximize

$$w_x^T C_{xy} w_y = (xw_x)^T (yw_y)$$

- subject to

$$\|xw_x\| = 1, \|yw_y\| = 1$$

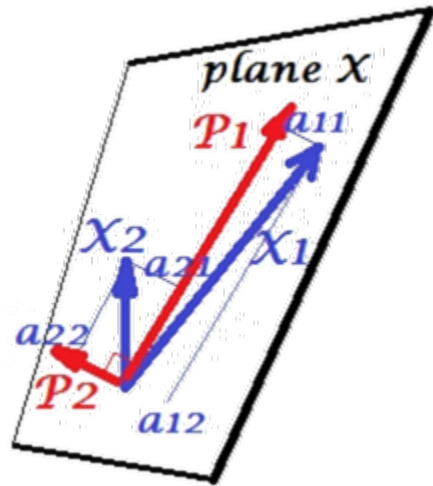
- In PCA we have a single variate and seek the direction that “maximizes the variance in the data”

$$w_x^T C_{xx} w_x = (xw_x)^T (xw_x)$$

- subject to

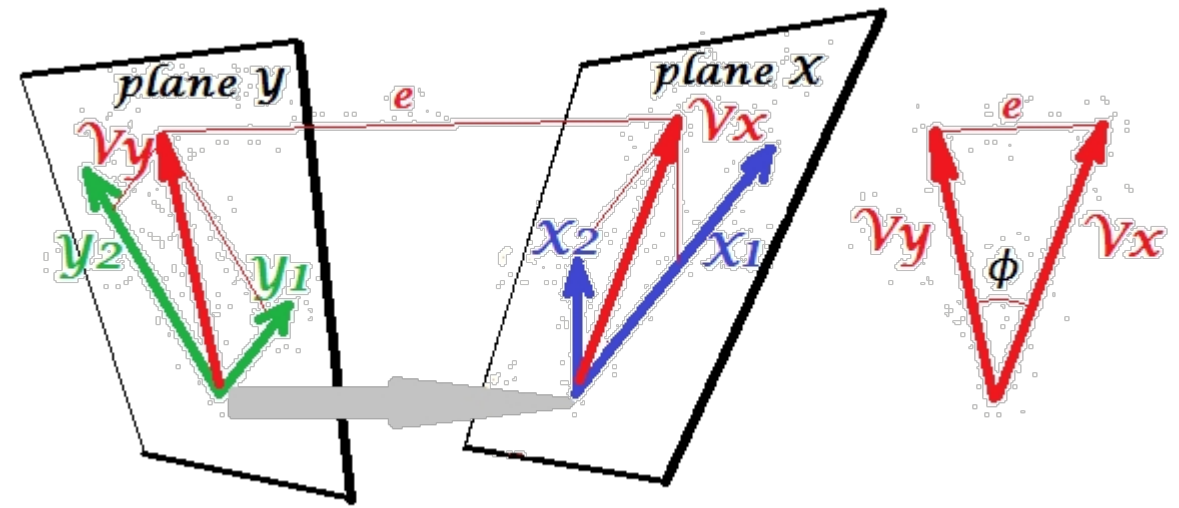
$$\|w_x\| = 1$$

# CCA vs PCA

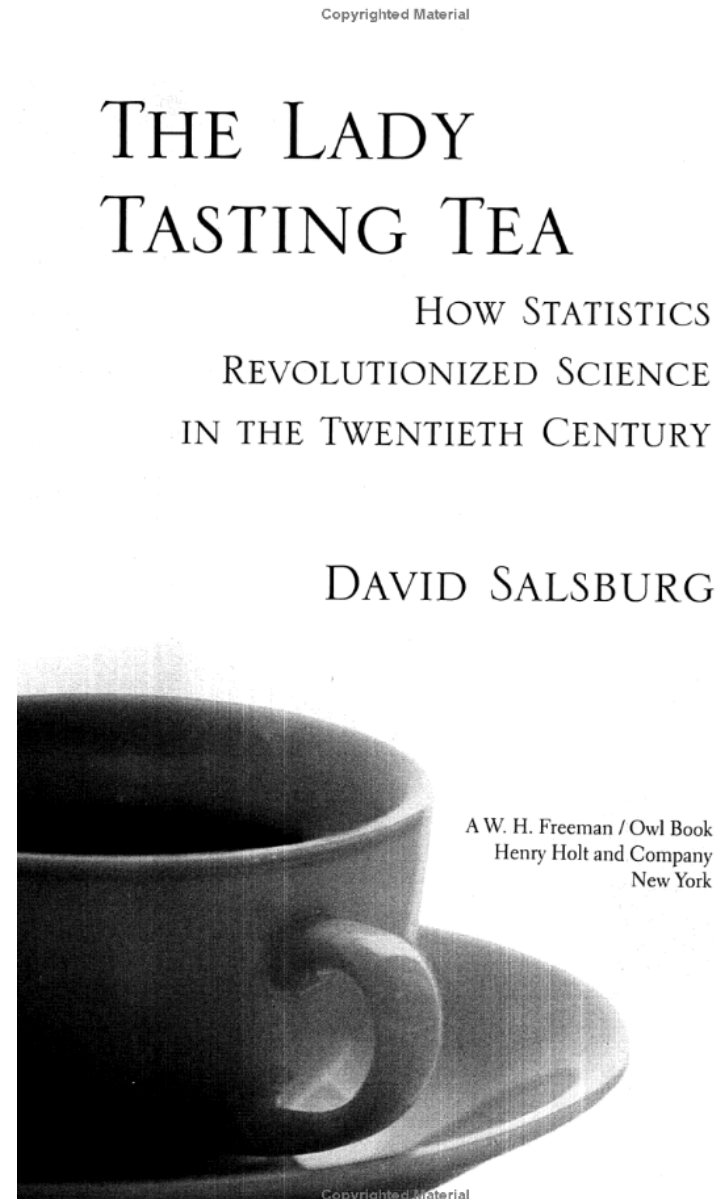
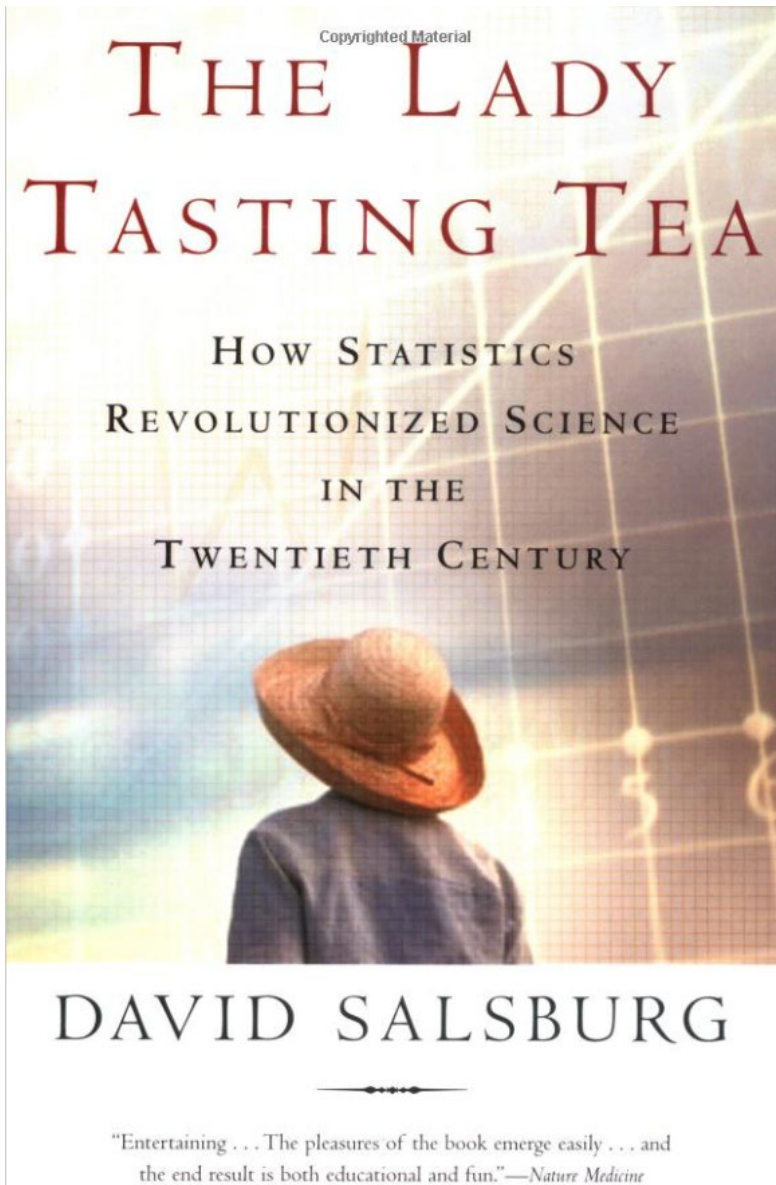


PCA

CCA



# History of Statistics



# Example 1

# CCA Example 1: Scores Data

Example:  $n = 88$  students took tests in each of 5 subjects: mechanics, vectors, algebra, analysis, statistics. (From Mardia et al. (1979) “Multivariate analysis” .) Each test is out of 100 points

The tests on mechanics, vectors were closed book and those on algebra, analysis, statistics were open book. There’s clearly some correlation between these two sets of scores:

|     | alg   | ana   | sta   |
|-----|-------|-------|-------|
| mec | 0.547 | 0.409 | 0.389 |
| vec | 0.610 | 0.485 | 0.436 |

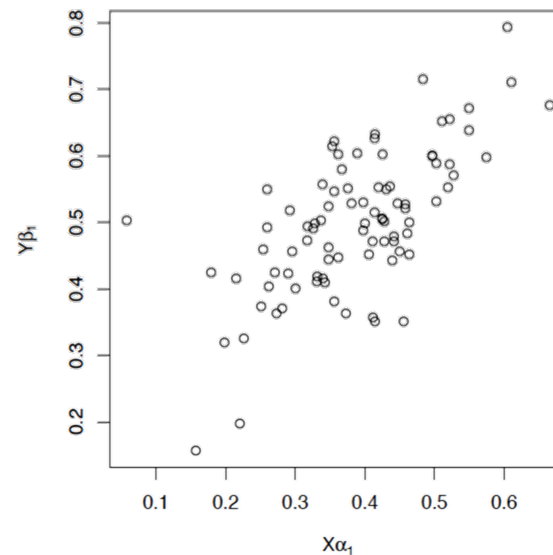
Canonical correlation analysis attempts to explain this phenomenon using the variables in each set **jointly**. Here  $X$  contains the closed book test scores and  $Y$  contains the open book test scores, so  $X \in \mathbb{R}^{88 \times 2}$  and  $Y \in \mathbb{R}^{88 \times 3}$

# CCA Example 1: Scores Data

The first canonical directions (multiplied by  $10^3$ ):

$$\alpha_1 = \begin{pmatrix} 2.770 \\ 5.517 \end{pmatrix} \begin{matrix} \text{mec} \\ \text{vec} \end{matrix}, \quad \beta_1 = \begin{pmatrix} 8.782 \\ 0.860 \\ 0.370 \end{pmatrix} \begin{matrix} \text{alg} \\ \text{ana} \\ \text{sta} \end{matrix}$$

The first canonical correlation is  $\rho_1 = 0.663$ , and the variates:



# Higher Order CCA

# Higher-Order Canonical Correlates

- We defined the 1<sup>st</sup> canonical correlation  $\rho_1$  through the projection vectors / directions

$$\mathbf{w}_x = \mathbf{w}_x^{(1)} \text{ and } \mathbf{w}_y = \mathbf{w}_y^{(1)}$$

- Given the first  $k-1$  directions, the  $k$ -th canonical correlation is defined via vectors  $\mathbf{w}_x^{(k)}$  and  $\mathbf{w}_y^{(k)}$ , so that we maximize

$$\max (\mathbf{x}\mathbf{w}_x^{(k)})^T (\mathbf{y}\mathbf{w}_y^{(k)})$$

- but under orthogonality constraints to the previous directions

$$\|\mathbf{x}\mathbf{w}_x^{(k)}\| = 1, \|\mathbf{y}\mathbf{w}_y^{(k)}\| = 1$$

$$(\mathbf{x}\mathbf{w}_x^{(k)})^T (\mathbf{x}\mathbf{w}_x^{(j)}) = 0, j = 1, 2, \dots, k-1$$

$$(\mathbf{y}\mathbf{w}_y^{(k)})^T (\mathbf{y}\mathbf{w}_y^{(j)}) = 0, j = 1, 2, \dots, k-1$$

# Equivalent Higher Order CCA Formulations

$$\begin{array}{ccc}
 \max_{\mathbf{a}_i, \mathbf{b}_i} \sum_i \mathbf{a}_i^T \mathbf{X} \mathbf{Y}^T \mathbf{b}_i & \longleftrightarrow & \max_{\mathbf{A}, \mathbf{B}} \text{trace}(\mathbf{A}^T \mathbf{X} \mathbf{Y}^T \mathbf{B}) \\
 \text{s.t. } \mathbf{a}_i^T \mathbf{X} \mathbf{X}^T \mathbf{a}_j = \delta_{i=j} \forall j \leq i & & \text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{X}^T \mathbf{A} = \mathbf{I}_d \\
 \mathbf{b}_i^T \mathbf{Y} \mathbf{Y}^T \mathbf{b}_j = \delta_{i=j} \forall j \leq i & & \mathbf{B}^T \mathbf{Y} \mathbf{Y}^T \mathbf{B} = \mathbf{I}_d \\
 & & \updownarrow \\
 \begin{bmatrix} \mathbf{0} & \mathbf{X} \mathbf{Y}^T \\ \mathbf{Y} \mathbf{X}^T & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{a}_i \\ \mathbf{b}_i \end{pmatrix} = \rho_i \begin{bmatrix} \mathbf{X} \mathbf{X}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Y} \mathbf{Y}^T \end{bmatrix} \begin{pmatrix} \mathbf{a}_i \\ \mathbf{b}_i \end{pmatrix} & \longleftrightarrow & \min_{\mathbf{A}, \mathbf{B}} \frac{1}{2} \|\mathbf{A}^T \mathbf{X} - \mathbf{B}^T \mathbf{Y}\|_F^2 \\
 & & \text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{X}^T \mathbf{A} = \mathbf{I}_d \\
 & & \mathbf{B}^T \mathbf{Y} \mathbf{Y}^T \mathbf{B} = \mathbf{I}_d
 \end{array}$$

# Principal Angles Between Subspaces

- Goal: capture geometric configuration of two subspaces with few and intuitive numbers
- Principal angles intuition
  - Measure angles between ‘most similar’ directions within subspaces
  - Capture relative ‘orientation’ of two subspaces
  - Recursive definitions with decreasing similarity
- Consider subspaces spanned by the cols of  $X$  and  $Y$  respectively

■ Definition

$$\cos \theta_i = \max_{\mathbf{u}_i \in \mathcal{Y}_1} \max_{\mathbf{v}_i \in \mathcal{Y}_2} \mathbf{u}_i^T \mathbf{v}_i$$

$$\text{s.t. } \|\mathbf{u}_i\|_2 = \|\mathbf{v}_i\|_2 = 1$$

$$\forall j < i : \mathbf{u}_i^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{v}_j = 0$$



$$\max \text{trace}_{\mathbf{A}, \mathbf{B}} (\mathbf{A}^T \mathbf{X}^T \mathbf{Y} \mathbf{B})$$

such that

$$\mathbf{A}, \mathbf{B} \in \mathcal{O}_p$$

‘align’ them

# CCA and Principal Angles

- CCA between RVs in column form  $\rightarrow$  principal angles of row spaces spanned by data matrices

$$\rho_i = \cos \theta_i$$

- For  $i = 1, 2, \dots$

- Note 1: The number of canonical directions/variates is

$$r = \min \{ \text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y}) \}$$

- Note 2:

- If  $\mathbf{X}$  and  $\mathbf{Y}$  are orthogonal ( $\mathbf{X}^T \mathbf{Y} = 0$ ) then all principal angles are  $90^\circ$  and the corresponding canonical correlations are 0
- If  $\mathbf{X}$  and  $\mathbf{Y}$  intersect in a  $d$ -dimensional subspace, then the first  $d$  principal angles are 0

# CCA Computation via Sphering/Whitening

For any symmetric invertible matrix  $A \in \mathbb{R}^{n \times n}$ , there is a matrix  $A^{1/2} \in \mathbb{R}^{n \times n}$ , called the (symmetric) **square root** of  $A$ , such that  $A^{1/2}A^{1/2} = A$

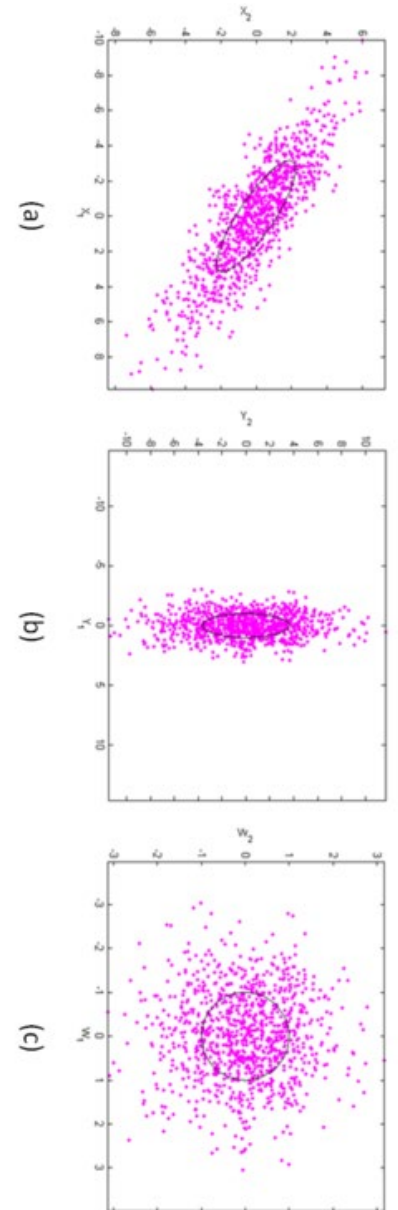
We write the inverse of  $A^{1/2}$  as  $A^{-1/2}$ . Note  $A^{-1/2}AA^{-1/2} = I$ . (Why?)

Given centered matrices  $X \in \mathbb{R}^{n \times p}$  and  $Y \in \mathbb{R}^{n \times q}$ ,<sup>2</sup> we define  $V_X = X^T X \in \mathbb{R}^{p \times p}$  and  $V_Y = Y^T Y \in \mathbb{R}^{q \times q}$ . Then

$$\tilde{X} = XV_X^{-1/2} \in \mathbb{R}^{n \times p} \quad \text{and} \quad \tilde{Y} = YV_Y^{-1/2} \in \mathbb{R}^{n \times q}$$

are called the **sphered** versions of  $X$  and  $Y$ .<sup>3</sup> Note that the sample covariance of  $\tilde{X}$  and  $\tilde{Y}$  is

$$\text{cov}(\tilde{X}) = I/n \quad \text{and} \quad \text{cov}(\tilde{Y}) = I/n$$



# Transformed Problem

As suggested by the previous slide, we will take  $\tilde{X} = XV_X^{-1/2}$  and  $\tilde{Y} = YV_Y^{-1/2}$ , and we'll solve the problem

$$\tilde{\alpha}_1, \tilde{\beta}_1 = \underset{\|\tilde{X}\tilde{\alpha}\|_2=1, \|\tilde{Y}\tilde{\beta}\|_2=1}{\operatorname{argmax}} (\tilde{X}\tilde{\alpha})^T (\tilde{Y}\tilde{\beta})$$

Recall that then  $\alpha_1 = V_X^{-1/2}\tilde{\alpha}_1$  and  $\beta_1 = V_Y^{-1/2}\tilde{\beta}_1$ .

So why is this **simpler**? Note that the constraint says

$$1 = (\tilde{X}\tilde{\alpha})^T (\tilde{X}\tilde{\alpha}) = \tilde{\alpha}^T V_X^{-1/2} X^T X V_X^{-1/2} \tilde{\alpha} = \tilde{\alpha}^T \tilde{\alpha}$$

i.e.,  $\|\tilde{\alpha}\|_2 = 1$ . Similarly,  $\|\tilde{\beta}\|_2 = 1$ . Hence our problem can be **rewritten** as:

$$\tilde{\alpha}_1, \tilde{\beta}_1 = \underset{\|\tilde{\alpha}\|_2=1, \|\tilde{\beta}\|_2=1}{\operatorname{argmax}} \tilde{\alpha}^T M \tilde{\beta}$$

where  $M = \tilde{X}^T \tilde{Y} = V_X^{-1/2} X^T Y V_Y^{-1/2} \in \mathbb{R}^{p \times q}$ . The same is true for further directions

# SVD to the Rescue

Now comes the **singular value decomposition** to the rescue (again!). Let  $r = \min\{p, q\}$ . Then we can decompose

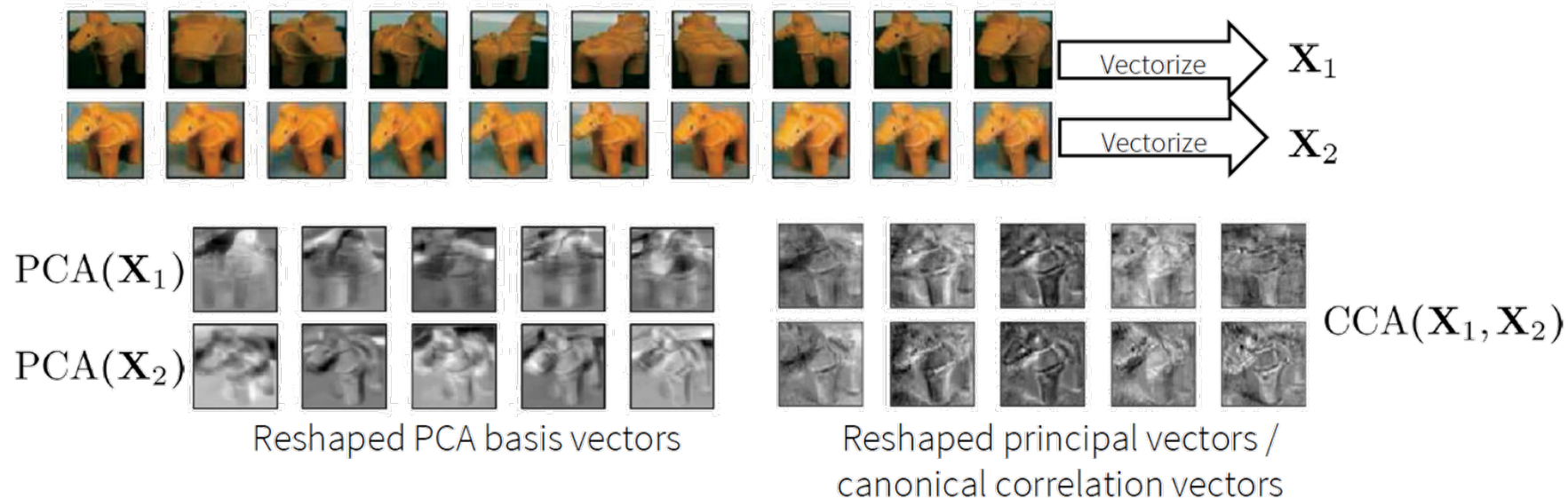
$$M = UDV^T$$

where  $U \in \mathbb{R}^{p \times r}$ ,  $V \in \mathbb{R}^{q \times r}$  have orthonormal columns, and  $D = \text{diag}(d_1, \dots, d_r) \in \mathbb{R}^{r \times r}$  with  $d_1 \geq \dots \geq d_r \geq 0$ . Further:

- ▶ The transformed canonical directions  $\tilde{\alpha}_1, \dots, \tilde{\alpha}_r \in \mathbb{R}^p$  and  $\tilde{\beta}_1, \dots, \tilde{\beta}_r \in \mathbb{R}^q$  are the columns of  $U$  and  $V$ , respectively
- ▶ The **canonical directions**  $\alpha_1, \dots, \alpha_r \in \mathbb{R}^p$  and  $\beta_1, \dots, \beta_r \in \mathbb{R}^q$  are the columns of  $V_X^{-1/2}U$  and  $V_Y^{-1/2}V$ , respectively;
- ▶ the **canonical variates**  $X\alpha_1, \dots, X\alpha_r \in \mathbb{R}^n$  and  $Y\beta_1, \dots, Y\beta_r \in \mathbb{R}^n$  are the columns of  $XV_X^{-1/2}U \in \mathbb{R}^{n \times r}$  and  $YV_Y^{-1/2}V \in \mathbb{R}^{n \times r}$ , respectively
- ▶ The **canonical correlations**  $\rho_1 \geq \dots \geq \rho_r$  are equal to  $d_1 \geq \dots \geq d_r$ , the diagonal entries of  $D$

# More Examples

# CCA Example 2: Object Recognition



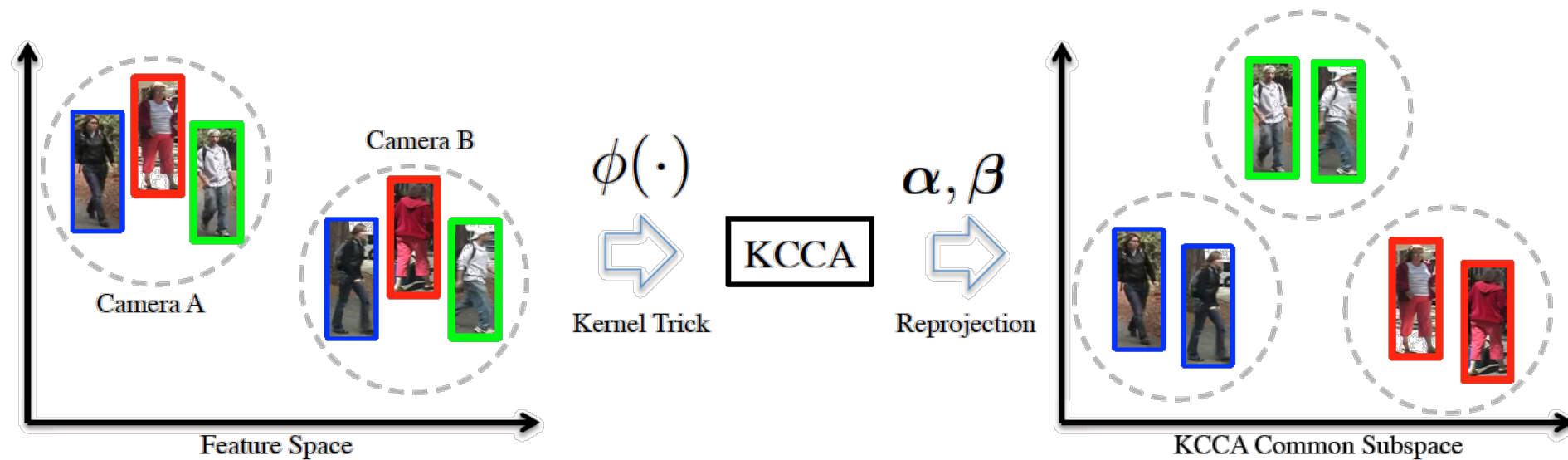
- “Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlation” Kim, Kittler, Cipolla [PAMI 07]
- Idea: Find low-dimensional subspace embedding s.t.
  - within-class CCA is maximized
  - Between-class CCA is minimized

# CCA Example 3: KCCA for Matching People

- Lisanti, Giuseppe, Iacopo Masi, and Alberto Del Bimbo. "Matching people across camera views using kernel canonical correlation analysis." Proceedings of the International Conference on Distributed Smart Cameras. ACM, 2014.



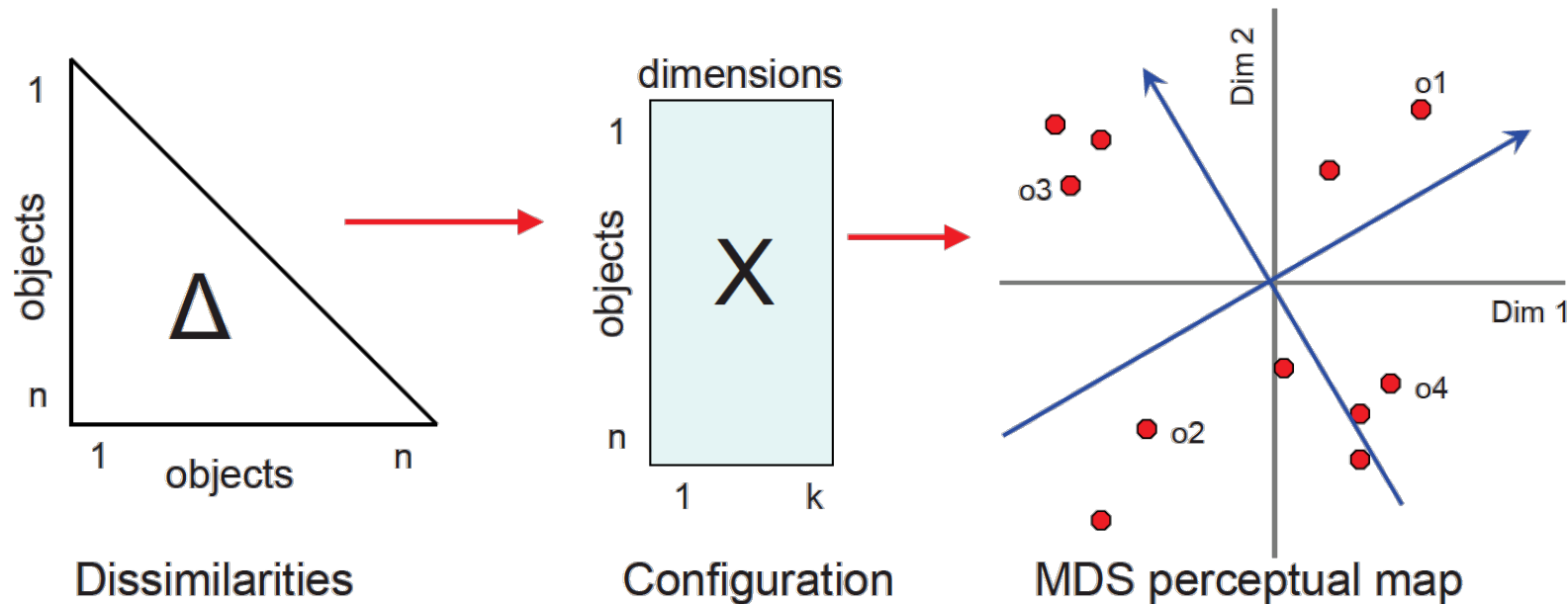
# Common Subspace Through KCCA



# Multidimensional Scaling

# Multidimensional Scaling (MDS)

- A “distance preserving” embedding of the data into a Euclidean space
  - Sometimes distances are observed directly (e.g., similarity ratings)
  - Sometimes they can be calculated from a data table (e.g., Euclidean distances, correlations)



# Formally (Metric MDS) ...

- Given a (symmetric) matrix of pairwise “dis-similarities” between  $n$  objects / data sets

$$M = \left( \delta_{ij} \right)_{n \times n}$$

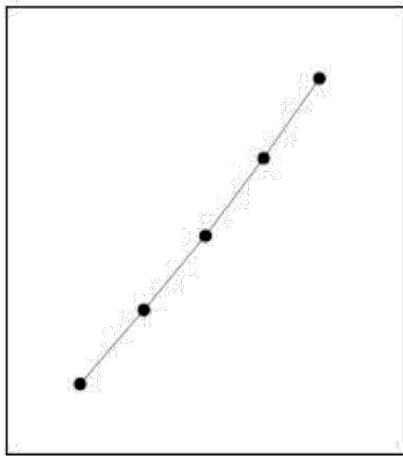
No need to satisfy the triangle inequality

- Find  $n$  points in low-dimensional space  $R^d$ , so that their distance matrix is as close as possible to  $M$
- Low  $d$  (=2,3) allows us to visualize the data directly

# Distances and Dimensionality

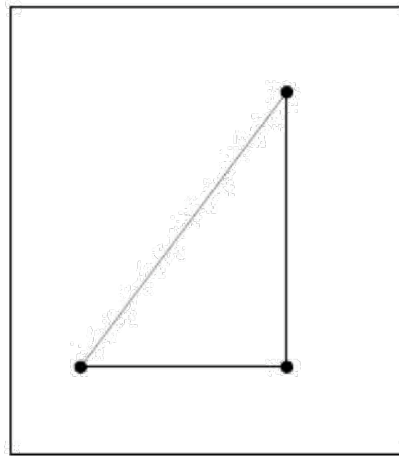
- How do distances/dissimilarities determine dimensionality?

$$D = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}$$



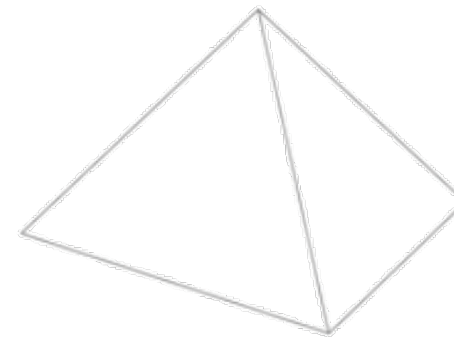
k=1

$$D = \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 5 \\ 4 & 5 & 0 \end{bmatrix}$$



k=2

$$D = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$



k=3

# MDS: Motivating Example

- Travel times by train between French cities

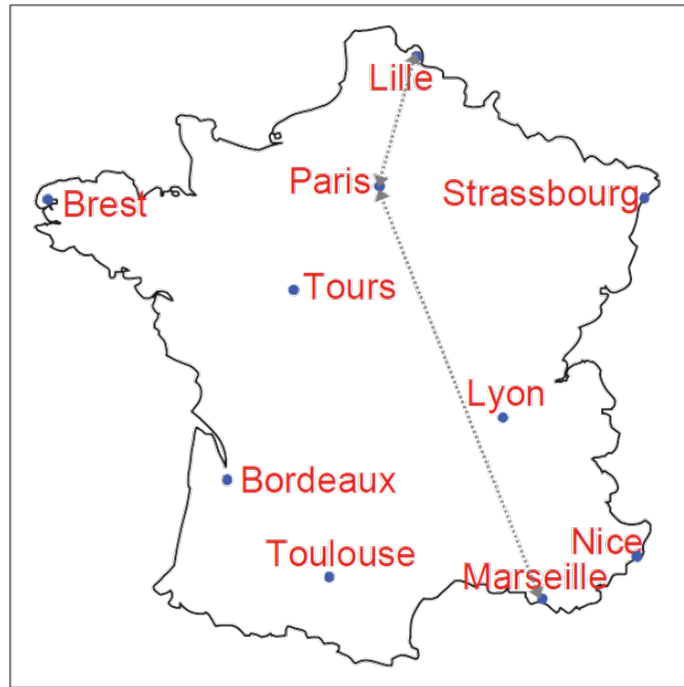
|            | Bor-<br>deaux | Brest | Lille | Lyon | Mar-<br>seille | Nice  | Paris | Strassb<br>ourg | Tou-<br>louse | Tours |
|------------|---------------|-------|-------|------|----------------|-------|-------|-----------------|---------------|-------|
| Bordeaux   | 0             |       |       |      |                |       |       |                 |               |       |
| Brest      | 9:58          | 0     |       |      |                |       |       |                 |               |       |
| Lille      | 6:39          | 7:11  | 0     |      |                |       |       |                 |               |       |
| Lyon       | 8:05          | 7:11  | 4:52  | 0    |                |       |       |                 |               |       |
| Marseille  | 5:47          | 8:49  | 6:12  | 1:35 | 0              |       |       |                 |               |       |
| Nice       | 8:30          | 13:36 | 8:20  | 4:33 | 2:26           | 0     |       |                 |               |       |
| Paris      | 2:59          | 4:17  | 1:04  | 2:01 | 3:00           | 5:52  | 0     |                 |               |       |
| Strasbourg | 8:08          | 10:16 | 6:54  | 4:36 | 7:04           | 11:15 | 4:01  | 0               |               |       |
| Toulouse   | 2:02          | 13:52 | 9:42  | 4:25 | 3:26           | 6:29  | 5:14  | 10:56           | 0             |       |
| Tours      | 2:36          | 5:38  | 4:17  | 4:21 | 5:13           | 9:04  | 1:13  | 6:03            | 6:06          | 0     |

- Considerations:
  - The recovered configuration ( $\mathbf{X}$  = map) should be 2D
  - NSEW not relevant to distances— may have to rotate
  - Travel time not necessarily  $\sim$  map distance (TGV)
  - May need to consider other relations between dissimilarity and distance in MD space

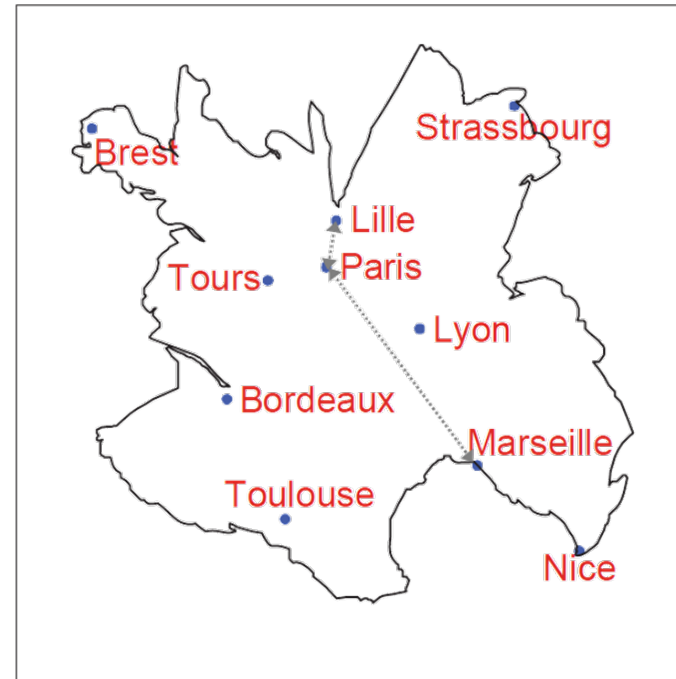
Source: Patrick Groenen, "Past, Present, and Future of Multidimensional Scaling", CARME, 2011

# MDS: Motivating Example

- Travel times by train between French cities



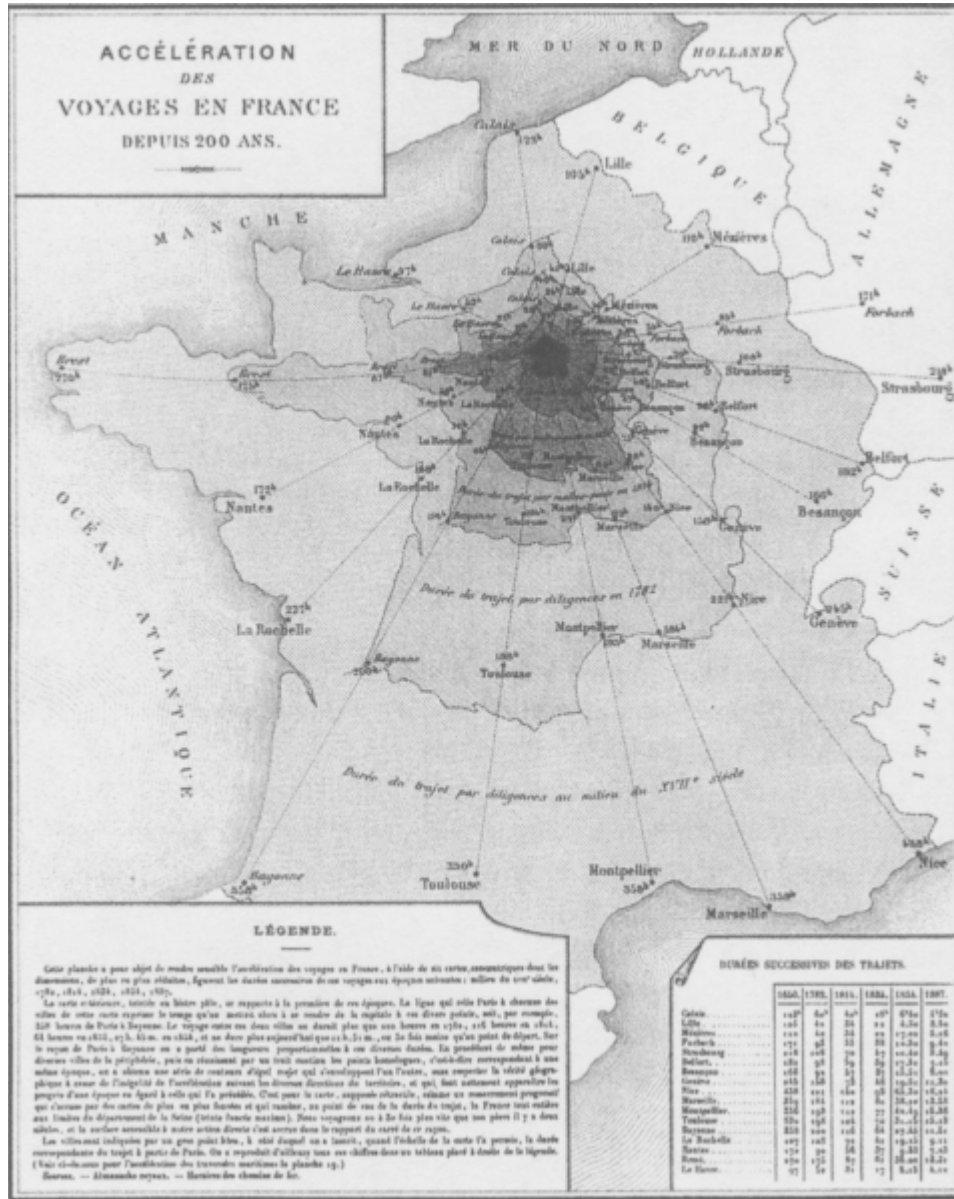
Actual map of France



MDS recovered map

Marseille, Lille, Lyon, ... closer to Paris in travel time (TGV)

# MDS: Motivating Example



- Travel time only partially related to map distance

Emile Cheysson (1888) – anaphoric map, showing decrease in time to travel from Paris over 200 years

- Other considerations relating distance and dissimilarity

# MDS Has Many Uses

- Psychology (perception, cognition)
- Political science (voting behavior, court decisions)
- Sociology (social network analysis)
- Archeology (artifact similarity)
- Biology/Chemistry (molecular structure, species analysis)
- Document retrieval & classification
- Graph layout
- Pattern recognition
- Dimension reduction
- ...

# Obtaining Dissimilarity Data

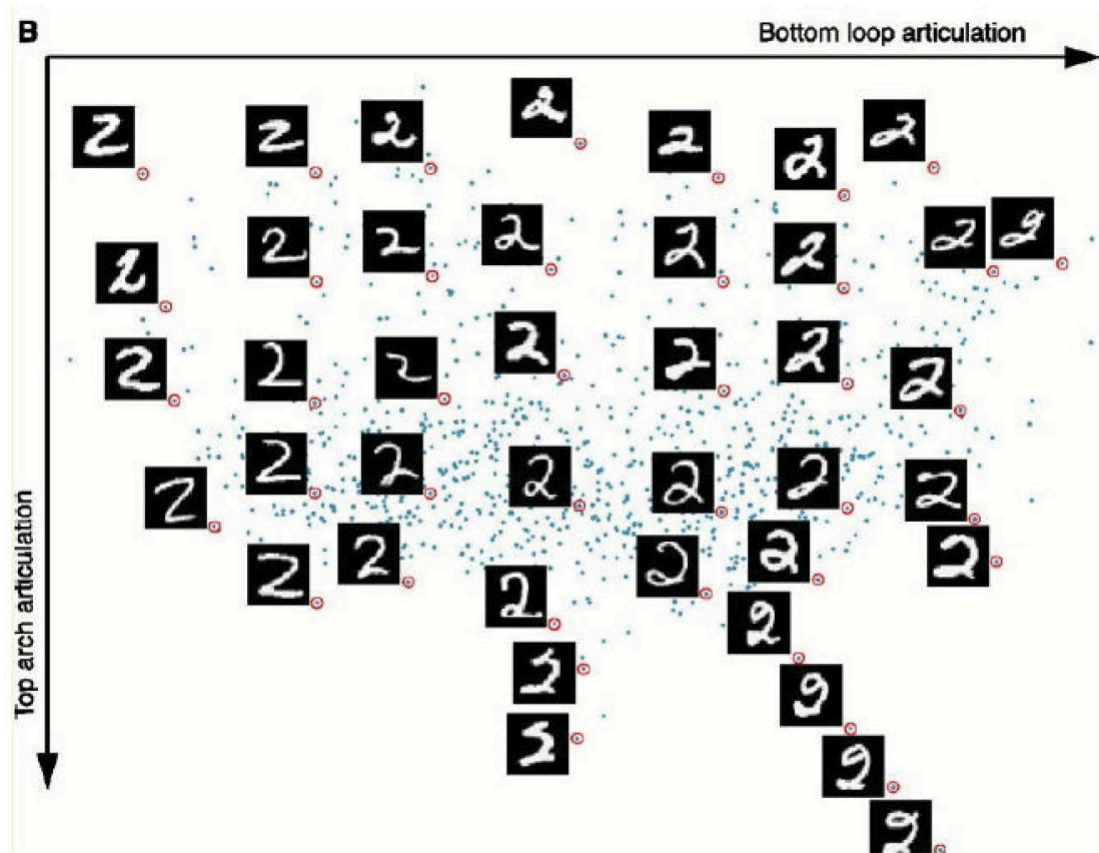
- Direct ratings (flavor comparisons)
- Confusion data (Morse dots and dashes)
- Co-occurrence data (Amazon recommendations)
- Sorting into groups

Sometimes distances are naturally defined – but sometimes we seek subjective dissimilarities

The screenshot shows a software interface for defining subjective dissimilarities between flavors. On the left, there is a list of flavors: Butter Pecan, Chocolate Chip, Raspberry Ripple, Coffee, Orange Sherbert, and Pistachio. The 'Orange Sherbert' flavor is currently selected. The interface is divided into two main sections, each with three comparison containers. The top section is titled 'Favorites' and contains three containers: 'Name: Favorites' (with 'Raspberry Ripple' and 'Orange Sherbert' listed), 'Name: Neutral' (with 'Pistachio' listed), and 'Name: Occasional' (with 'Butter Pecan' listed). The bottom section is titled 'Never Eat' and contains three containers: 'Name: Never Eat' (with 'Chocolate Chip' and 'Raspberry Ripple' listed), 'Name: Wife Likes' (with 'Pistachio' listed), and 'Name: Grandparents Like' (with 'Orange Sherbert' listed). At the bottom left, there is a button labeled 'Add Another Container', and at the bottom right, there is a button labeled 'Finish'.

Can every distance matrix be realized in a Euclidean space?

# Example: Pattern Recognition

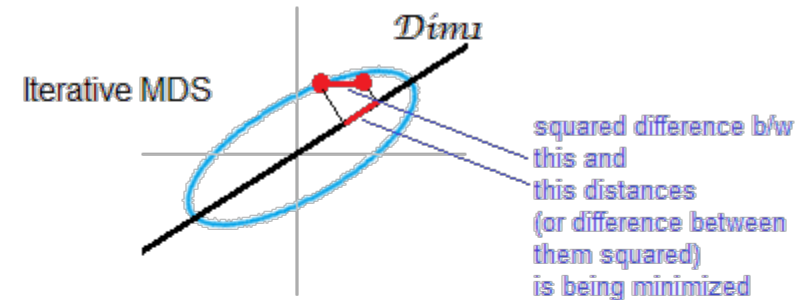


MDS of judged similarity of  
handwritten "2"s

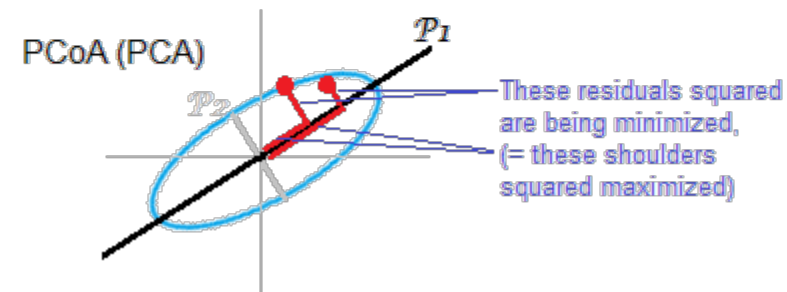
Goal: determine features  
important in pattern  
recognition

# Classic Metric MDS

- ◆ Sometimes we can model our data as points in a high-dimensional Euclidean space – and we are looking for an embedding to a lower-dimensional space that preserves (absolute or relative) distances (in the high-d space) as much as possible.
- ◆ In this case the problem has a clean geometric solution.



Is this the same as PCA?



# Classic Metric MDS

- ◆ To go from dimension  $D$  down to dimension  $d$
- ◆ Given data  $X \in R^{D \times n}$

$$X = \begin{pmatrix} | & | & \dots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \\ | & | & \dots & | \end{pmatrix} \quad \text{and} \quad M = \left( \text{dist}^2(\mathbf{x}_i, \mathbf{x}_j) \right)_{n \times n}$$

- ◆ We look for  $X'$ ,

$$X' = \begin{pmatrix} | & & | \\ \mathbf{x}'_1 & \dots & \mathbf{x}'_n \\ | & & | \end{pmatrix} \in R^{d \times n}$$

- ◆ We can assume the  $\mathbf{x}_i'$  are centered

# Classic Metric MDS

- ◆ So that we minimize  $\|M' - M\|$  (related to the *stress* of the system)
- ◆ where  $M' = \left( \text{dist}^2(\mathbf{x}_i', \mathbf{x}_j') \right) = \left( \left\| \mathbf{x}_i' - \mathbf{x}_j' \right\|^2 \right) \in R^{n \times n}$
- ◆  $M'$  is the Euclidean distances matrix for points  $\mathbf{x}_i'$ .

$$\min \|M' - M\|$$

# The Math Details

♦ Ideally we want  $M' = \left( \left\| \mathbf{x}'_i - \mathbf{x}'_j \right\|^2 \right) = M$

$$\left( \langle \mathbf{x}'_i - \mathbf{x}'_j, \mathbf{x}'_i - \mathbf{x}'_j \rangle \right) = M$$

$$\left( \left\| \mathbf{x}'_i \right\|^2 + \left\| \mathbf{x}'_j \right\|^2 - 2 \langle \mathbf{x}'_i, \mathbf{x}'_j \rangle \right) = M$$

$$\begin{pmatrix} \left\| \mathbf{x}'_1 \right\|^2 & \left\| \mathbf{x}'_1 \right\|^2 & \dots & \left\| \mathbf{x}'_1 \right\|^2 \\ \left\| \mathbf{x}'_2 \right\|^2 & \left\| \mathbf{x}'_2 \right\|^2 & \dots & \left\| \mathbf{x}'_2 \right\|^2 \\ \vdots & \vdots & \dots & \vdots \\ \left\| \mathbf{x}'_n \right\|^2 & \left\| \mathbf{x}'_n \right\|^2 & \dots & \left\| \mathbf{x}'_n \right\|^2 \end{pmatrix} + \begin{pmatrix} \left\| \mathbf{x}'_1 \right\|^2 & \left\| \mathbf{x}'_2 \right\|^2 & \dots & \left\| \mathbf{x}'_n \right\|^2 \\ \left\| \mathbf{x}'_1 \right\|^2 & \left\| \mathbf{x}'_2 \right\|^2 & \dots & \left\| \mathbf{x}'_n \right\|^2 \\ \vdots & \vdots & \dots & \vdots \\ \left\| \mathbf{x}'_1 \right\|^2 & \left\| \mathbf{x}'_2 \right\|^2 & \dots & \left\| \mathbf{x}'_n \right\|^2 \end{pmatrix} - 2 \begin{pmatrix} - & \mathbf{x}'_1 & - \\ \vdots & \vdots & \vdots \\ - & \mathbf{x}'_n & - \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{x}'_1 & \dots & \mathbf{x}'_n \\ | & & | \end{pmatrix}$$

want to get rid of these

$$X'^T \quad X'$$

# The Magic Matrix $J$

$$J = \begin{pmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ \vdots & & \ddots & \vdots \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}_{n \times n} = I - \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & 1 \end{pmatrix} = I - \frac{1}{n} K$$

$$(a \ a \ \cdots \ a) \cdot J = 0$$

$$J \cdot \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = 0$$

# So We Get to The Gram Matrix

Cleaning the system:

$$\times J / \left( \begin{array}{cccc} \|\mathbf{x}'_1\| & \|\mathbf{x}'_1\| & \dots & \|\mathbf{x}'_1\| \\ \|\mathbf{x}'_2\| & \|\mathbf{x}'_2\| & \dots & \|\mathbf{x}'_2\| \\ & & \vdots & \\ \|\mathbf{x}'_n\| & \|\mathbf{x}'_n\| & \dots & \|\mathbf{x}'_n\| \end{array} \right) + \left( \begin{array}{ccc} \|\mathbf{x}'_1\| & \|\mathbf{x}'_2\| & \dots & \|\mathbf{x}'_n\| \\ \|\mathbf{x}'_1\| & \|\mathbf{x}'_2\| & & \|\mathbf{x}'_n\| \\ \vdots & \vdots & \dots & \vdots \\ \|\mathbf{x}'_1\| & \|\mathbf{x}'_2\| & & \|\mathbf{x}'_n\| \end{array} \right) - 2X'^T X' = M / \times J$$

Note that  $X'K = KX'^T = 0$ ,  
as  $X'$  is centered

$$J = I - \frac{1}{n}K \quad -2JX'^T X'J = -2X'^T X' = JMJ$$

$$X'^T X' = -\frac{1}{2}JMJ =: B$$

$$X'^T X' = B$$

So from the distance matrix we can get the Gram (inner product) matrix.

# And Finally Use the Spectral Hammer

We will use the spectral decomposition of  $B$ :

$$X'^T X' = B = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}^T$$
  

$$X'^T X' = \underbrace{\begin{pmatrix} | & & | & \vdots & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_d & \vdots & \mathbf{v}_n \\ | & & | & \vdots & | \end{pmatrix}}_{n \times d} \begin{pmatrix} \sqrt{\lambda_1} & & & & \\ & \ddots & & & \\ & & \sqrt{\lambda_d} & & \\ & & & \ddots & \\ & & & & \sqrt{\lambda_n} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & & & & \\ & \ddots & & & \\ & & \sqrt{\lambda_d} & & \\ & & & \ddots & \\ & & & & \sqrt{\lambda_n} \end{pmatrix}^T \underbrace{\begin{pmatrix} | & & | & \vdots & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_d & \vdots & \mathbf{v}_n \\ | & & | & \vdots & | \end{pmatrix}^T}_{X'}^T$$

$d \times d$

# So We Get the $X'$

So we find  $X'$  by throwing away the last  $n-d$  eigenvalues

$$X' = \begin{pmatrix} \sqrt{\lambda_1} \mathbf{v}_1 \\ \dots & \dots & \dots \\ \sqrt{\lambda_d} \mathbf{v}_d \end{pmatrix}_{d \times n}$$

For this  $X'$  : 
$$X' = \arg \min_{X'} \|X'^T X' - B\|_{L^2}$$

This choice minimizes the inner product (and distance) loss

$$\|A\|_{L^2} = \sqrt{\sum_{i,j} A_{ij}^2}$$

# More General Metric MDS

- In general, we minimize directly the square loss on distances

$$\text{stress} = \sqrt{\frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j (d_{ij}^2)}} \quad \text{stress} = \mathcal{L}(\hat{d}_{ij}) = \left( \frac{\sum_{i < j} (\hat{d}_{ij} - f(d_{ij}))^2}{\sum_{i < j} d_{ij}^2} \right)^{\frac{1}{2}}$$

$f$  monotonic

- Sammon mapping

$$\text{Sammon's stress}(\hat{d}_{ij}) = \frac{1}{\sum_{l < k} d_{lk}} \sum_{i < j} \frac{(\hat{d}_{ij} - d_{ij})^2}{d_{ij}}$$

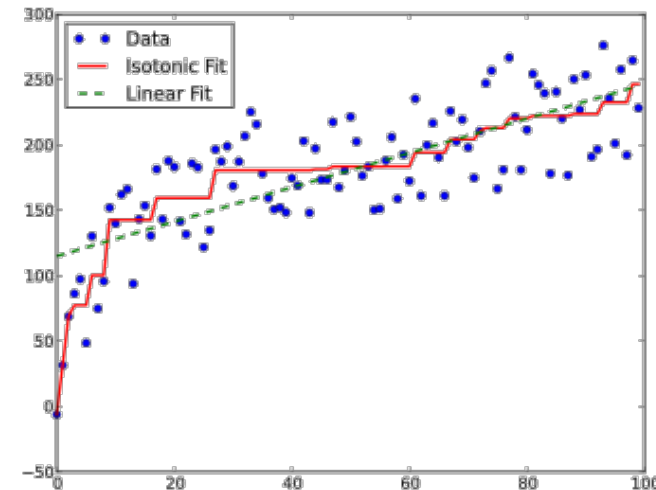
- This weighting system normalizes the squared-errors in pairwise distances by using the distance in the original space. As a result, Sammon mapping preserves the small  $d_{ij}$ , giving them a greater degree of importance in the fitting procedure than for larger values of  $d_{ij}$

Generally solved by gradient descent

# Non-Metric MDS

- Sometimes all we can say is that (dis)similarity is *ordinally* related to distance in MD space (only ordering of distances matters, not the actual values)
- If we only have ordering information, we can use *monotone (isotonic) regression* to find “disparities” that are compatible with the ordering constraints of the dissimilarities (not unique)
- We can then use alternating least squares (ALS) by repeating
  - a monotone regression step, followed by
  - a metric stress reduction step

$$\text{Stress}(\hat{\mathbf{D}}, \mathbf{X}) = \left[ \frac{\sum_{i < j} (\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2}{\sum_{i < j} d_{ij}^2(\mathbf{X})} \right]^{1/2}$$



# The 1st Assignment

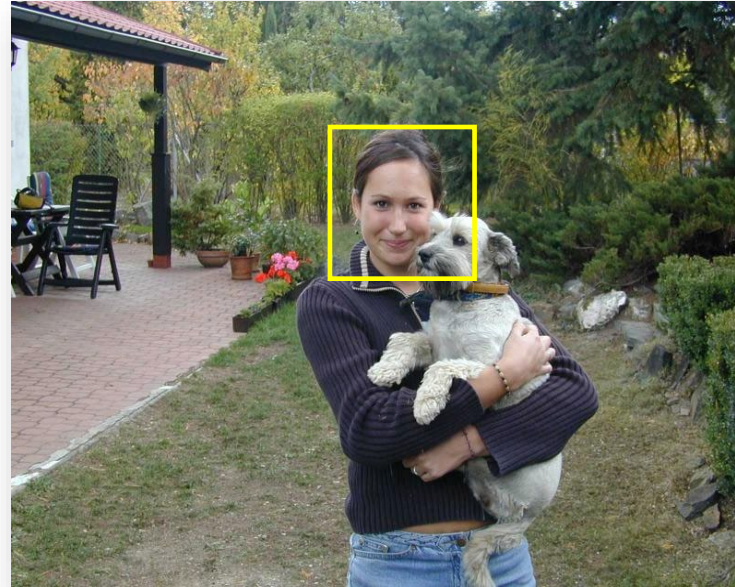
# Assignment 1: PCA + CCA

- Problem 1 (PCA): Face Reconstruction/Recognition: Eigenfaces
- Problem 2 (CCA): Word Recognition in Camera Images

PCA

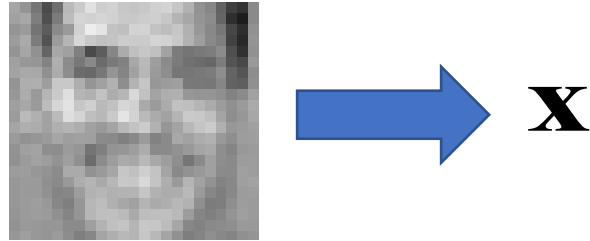
# Face Recognition

- Photo organization
- Surveillance
- ...
- Has to be coupled with face detection

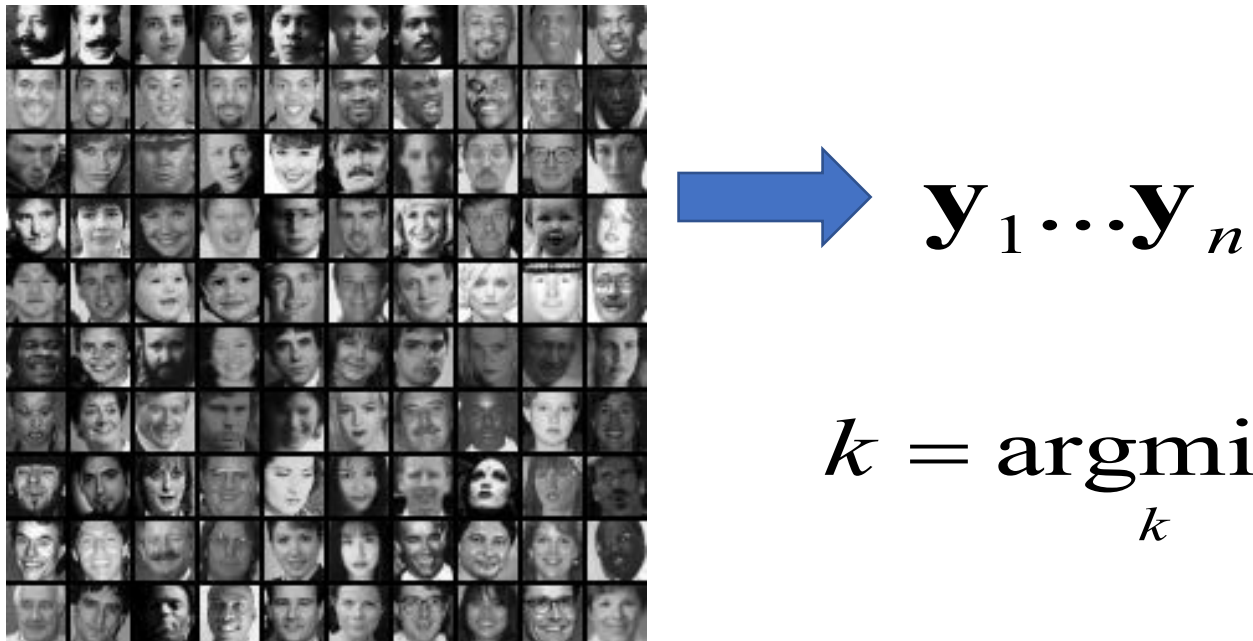


# Recognition: Embed into $\mathcal{R}^n$ !

- Treat image pixels as a long vector ...



- Face recognition by nearest neighbor



$$k = \operatorname{argmin}_k \left\| \mathbf{y}_k^T - \mathbf{x} \right\|$$

# Eigenfaces (PCA on Face Images)

- Dimension reduction — compute covariance matrix of face images
- Derive the principal components (“eigenfaces”)
  - $K$  eigenvectors with largest eigenvalues
- Represent all face images in the dataset as linear combinations of eigenfaces
  - Perform nearest neighbor on these coefficients

M. Turk and A. Pentland, [Face Recognition using Eigenfaces](#), CVPR 1991

# Training Images

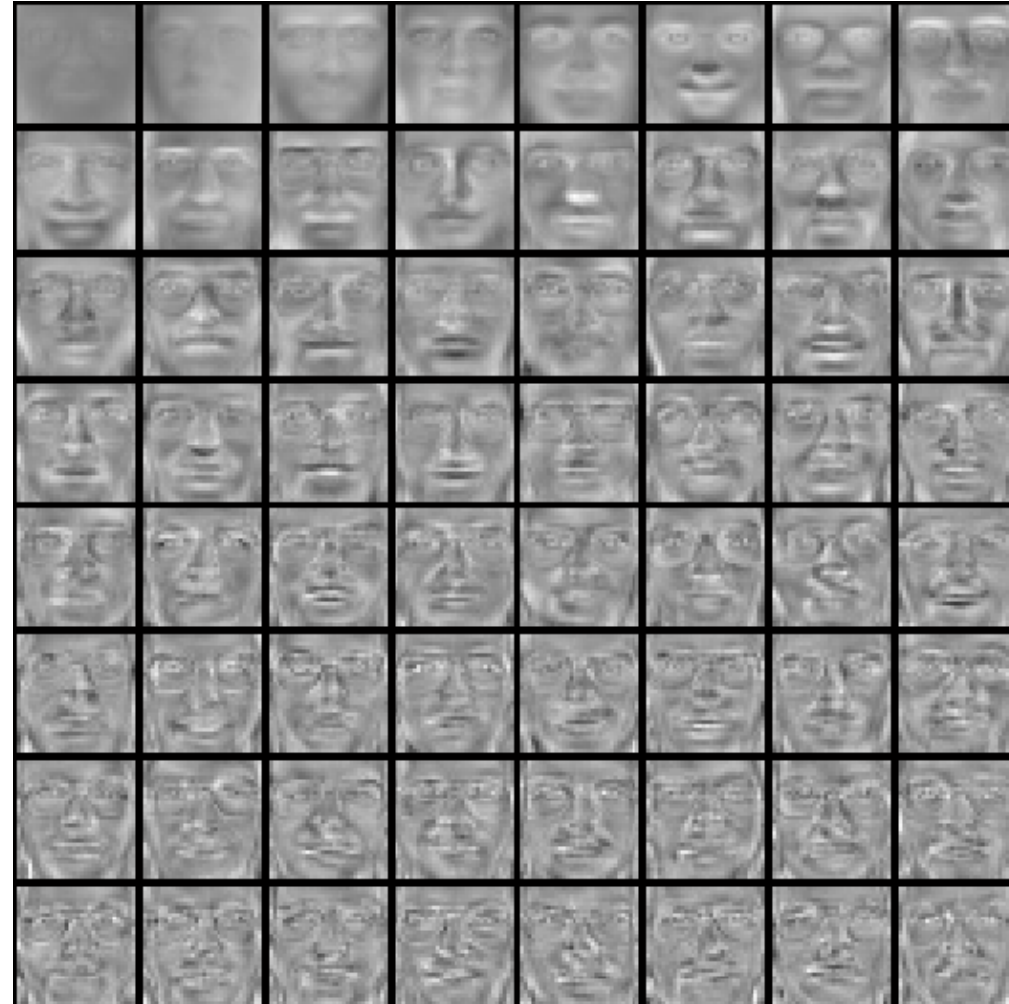


# Eigenfaces

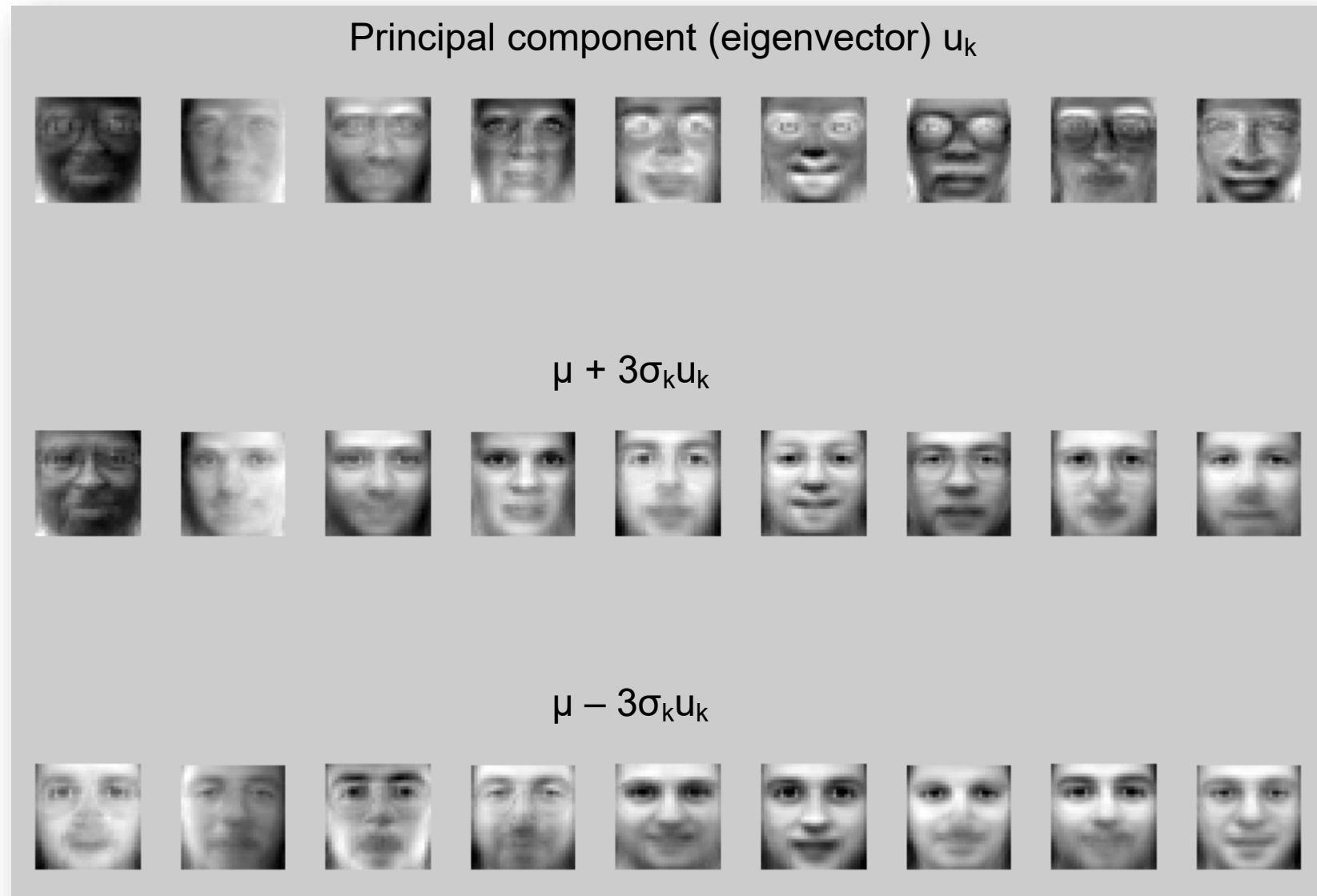
Mean:  $\mu$



Top eigenvectors:  $u_1, \dots, u_k$



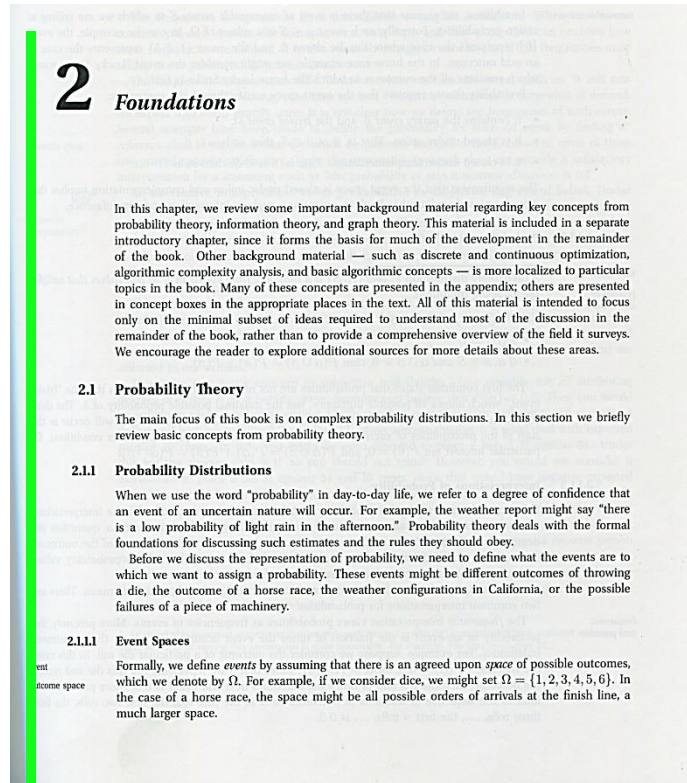
# Eigenface Visualization



CCA

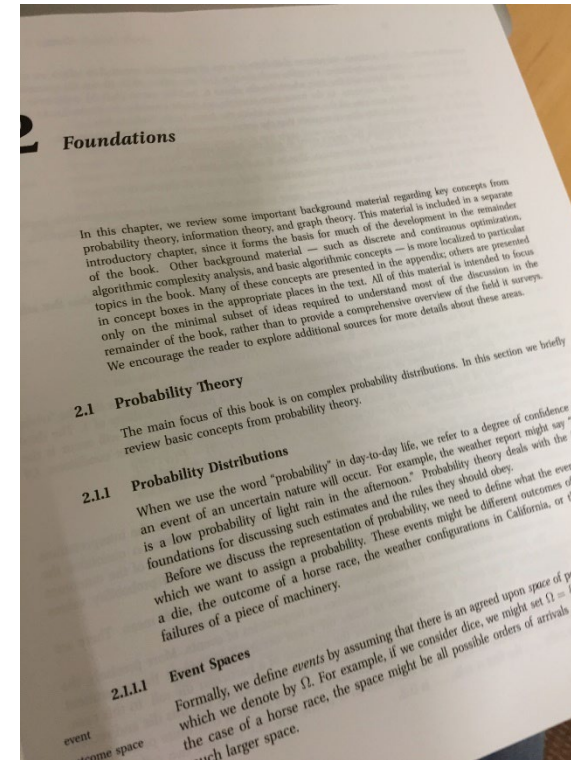
# Camera Text Recognition

Scanned



In this chapter, we review some important background material regarding key concepts from probability theory, information theory, and graph theory...

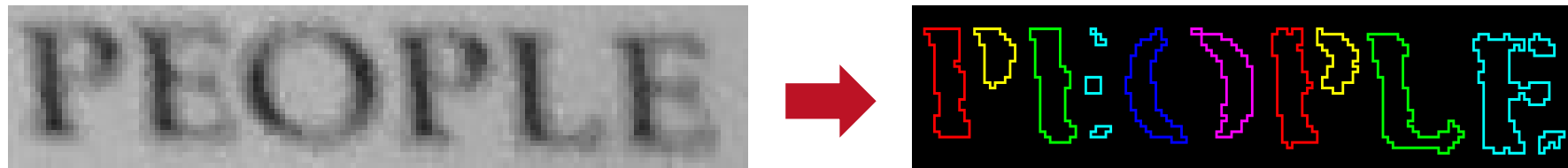
Camera



In this chapter we review some important background material regarding key concepts from probability theory, information theory, and graph theory...

# Why OCR Fails

- OCR reads text character by character, but character segmentation can be challenging.
  - View perspective distortions
  - Uneven lighting
  - Presence of noise and blur

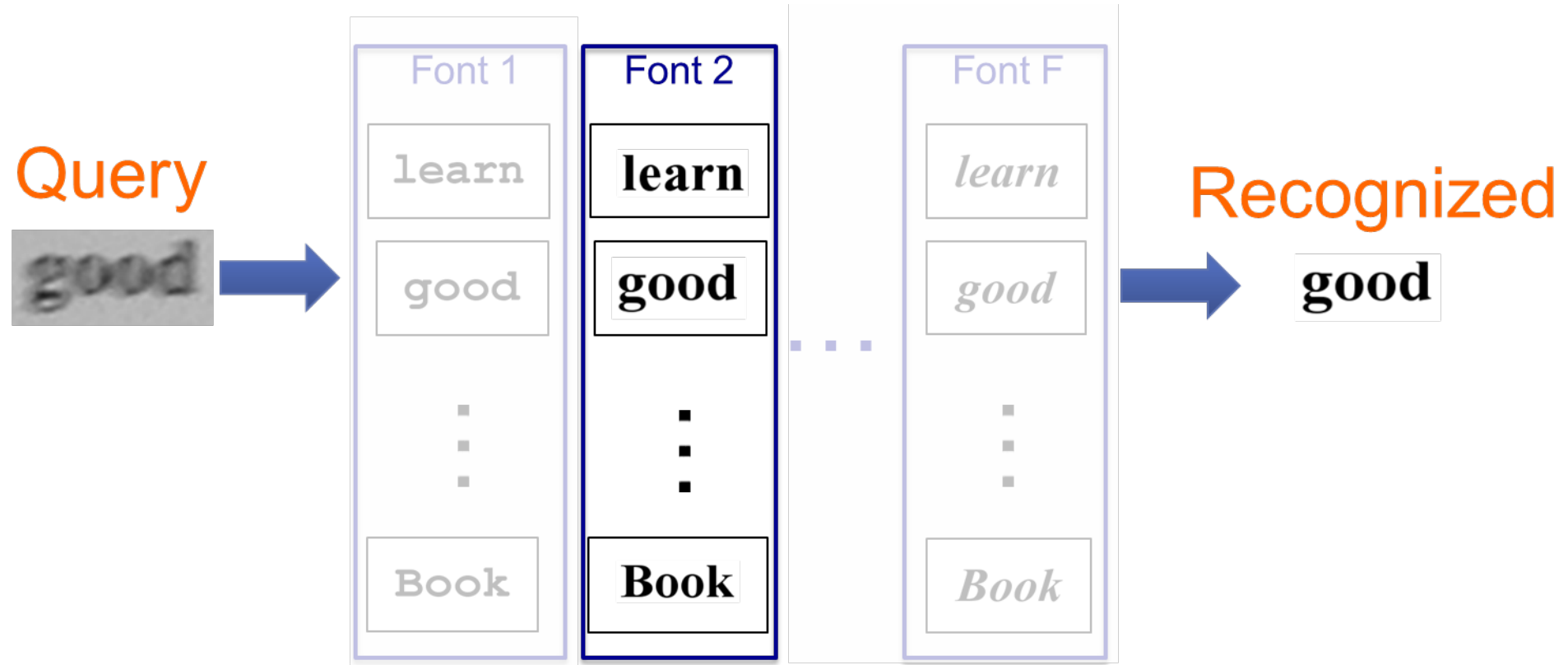


# Word Recognition via Image Retrieval



Huizhong Chen, Visual Word Recognition with Large-Scale Image Retrieval  
Stanford EE Ph.D. Thesis, 2015

# If Font Is Known, Task is Easier



# Inter-font Similarities

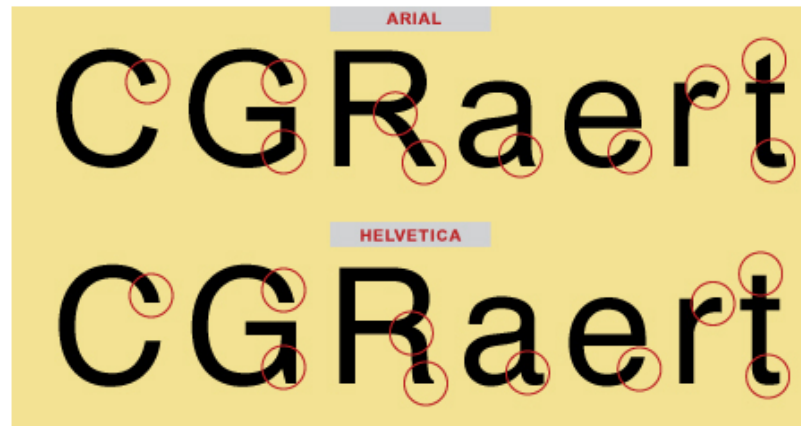
- Some fonts have visual similarities (e.g., **Arial**, **Helvetica**).

THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG

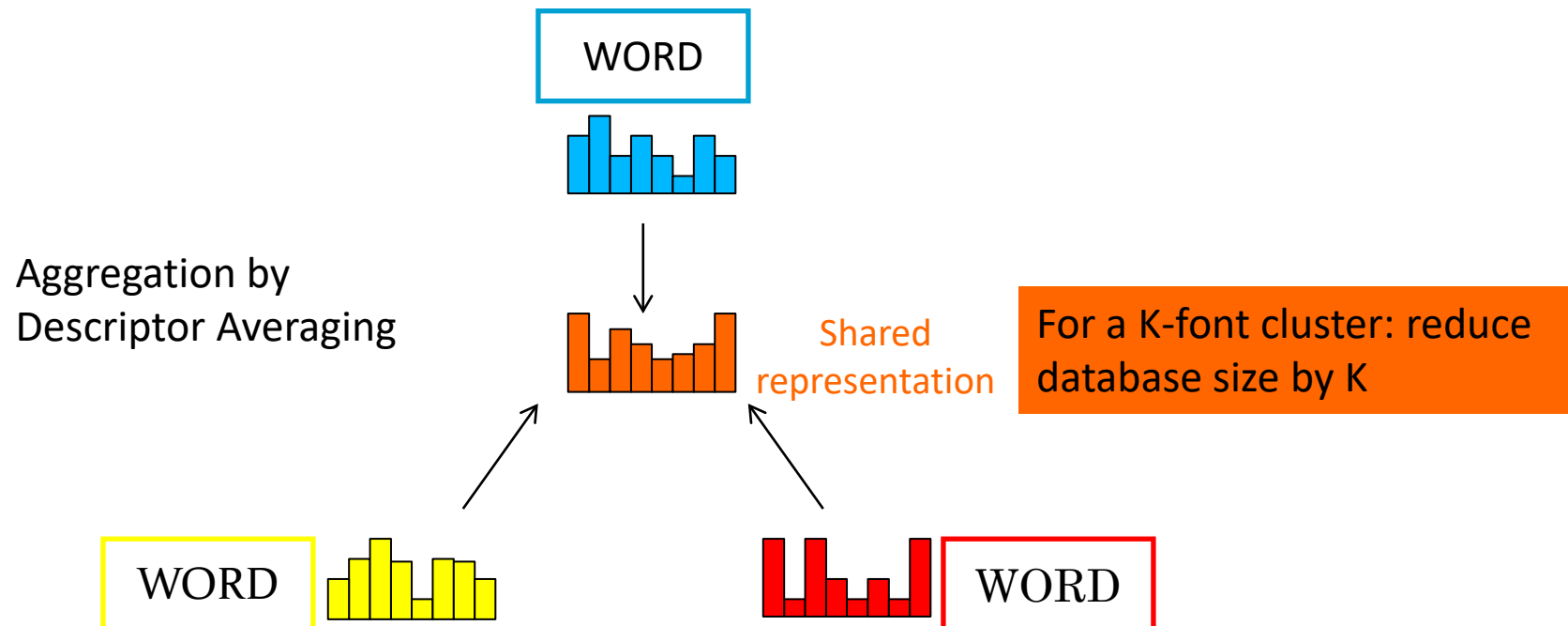
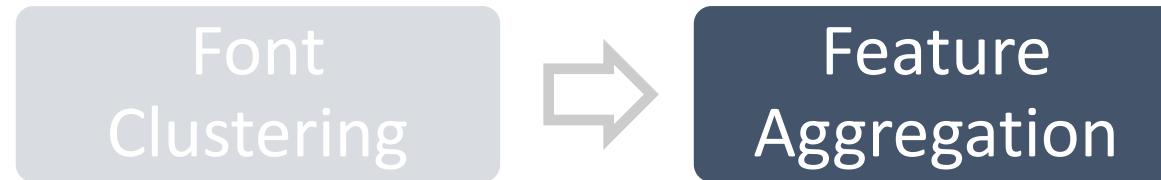
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG

the quick brown fox jumps over the lazy dog

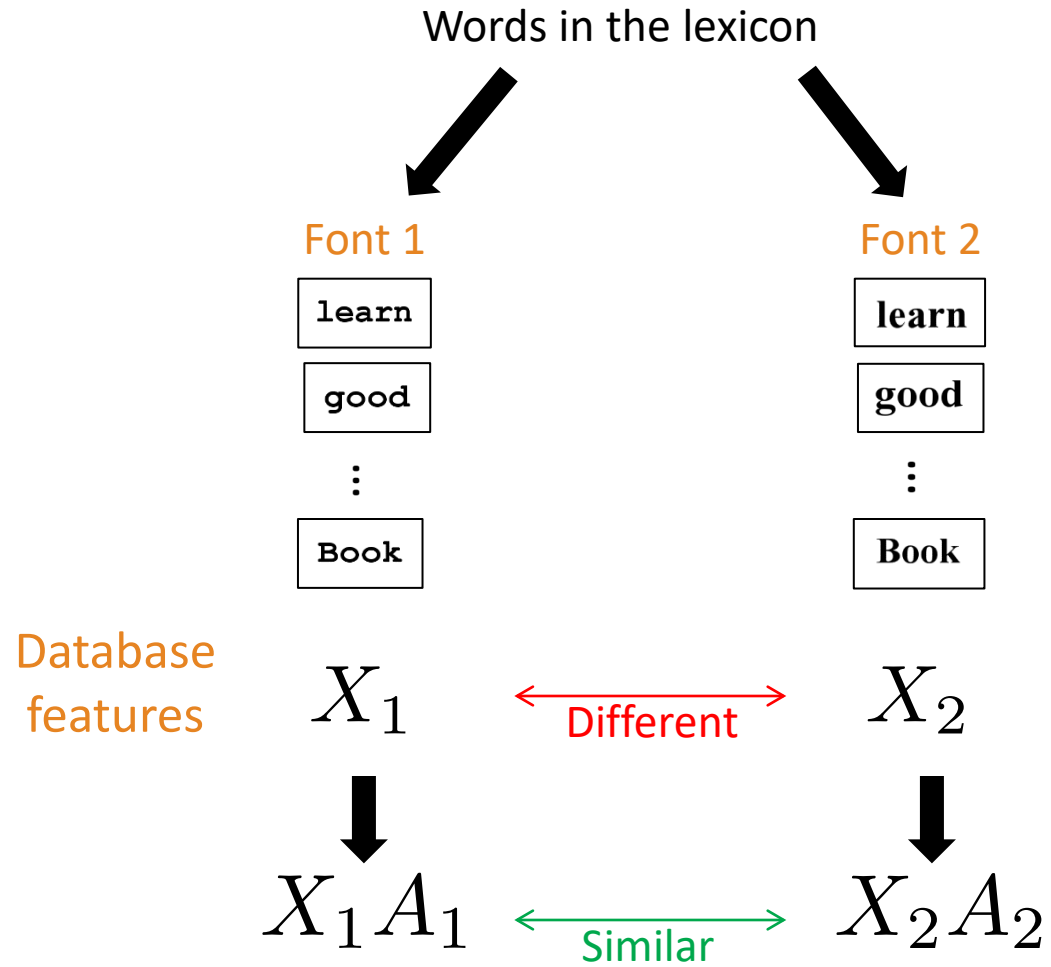
the quick brown fox jumps over the lazy dog



# Compact Data Base Representation



# Motivation for CCA



# Two View vs. Multi-View CCA

# That's All

