Joint Embeddings of Shapes and Images via CNN Image Purification

Yangyan Li[*]   Hao Su[*]   Charles R. Qi   Noa Fish
Daniel Cohen-Or   Leonidas J. Guibas
([*]Joint First Authors)

Joint Embeddings of Shapes and Images via CNN Image Purification

Deep learning is so cool for so many problems…

Deep learning, yay or nay?

A piece of cake, elementary math…

$$Y = f(X)$$

What the hell is the f?

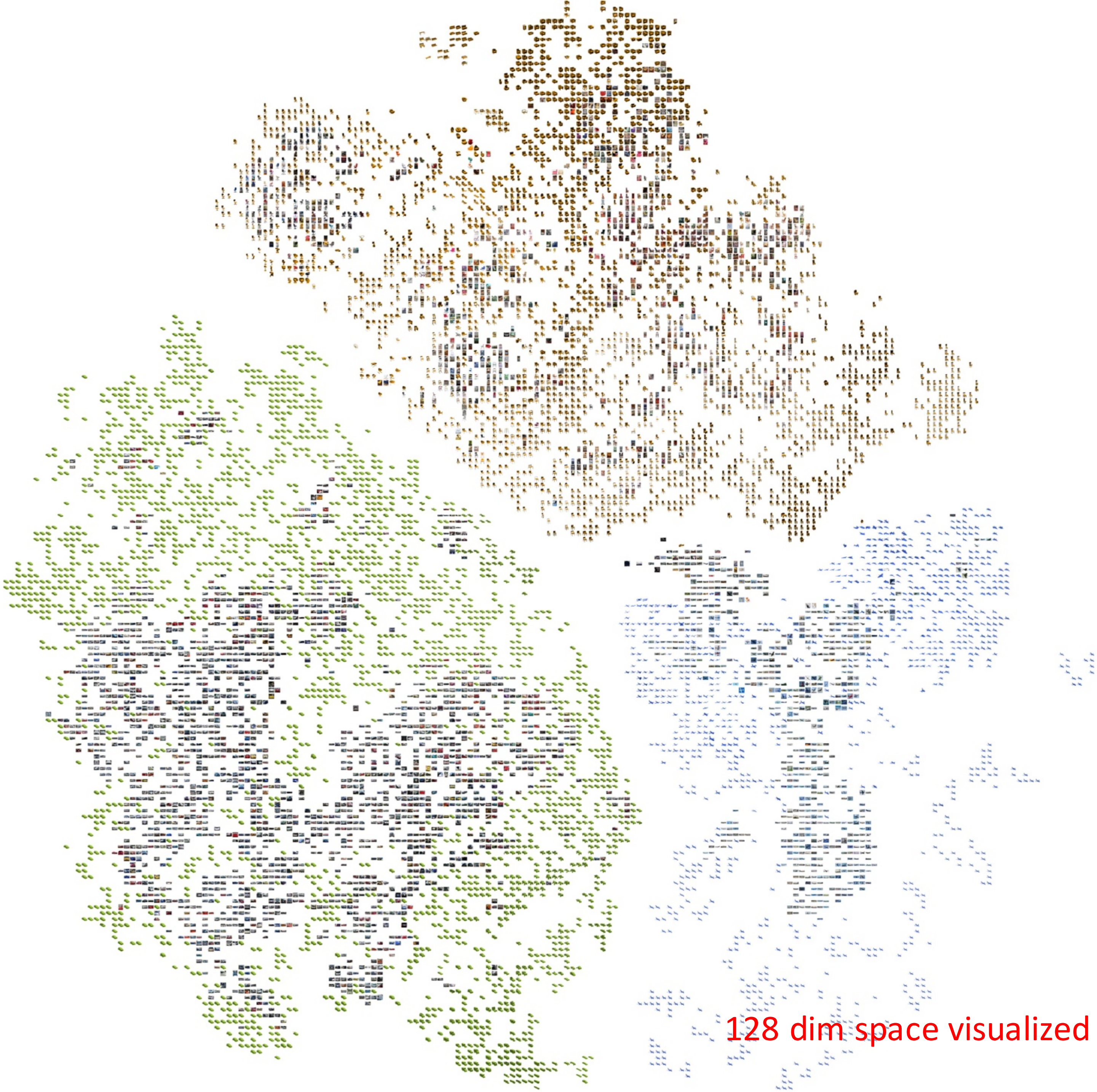
✗ ✓

It eats, a lot!

✗ ✓

A “FoodTech” for Deep Learning

Joint Embeddings of Shapes and Images via CNN Image Purification



128 dim space visualized by t-SNE

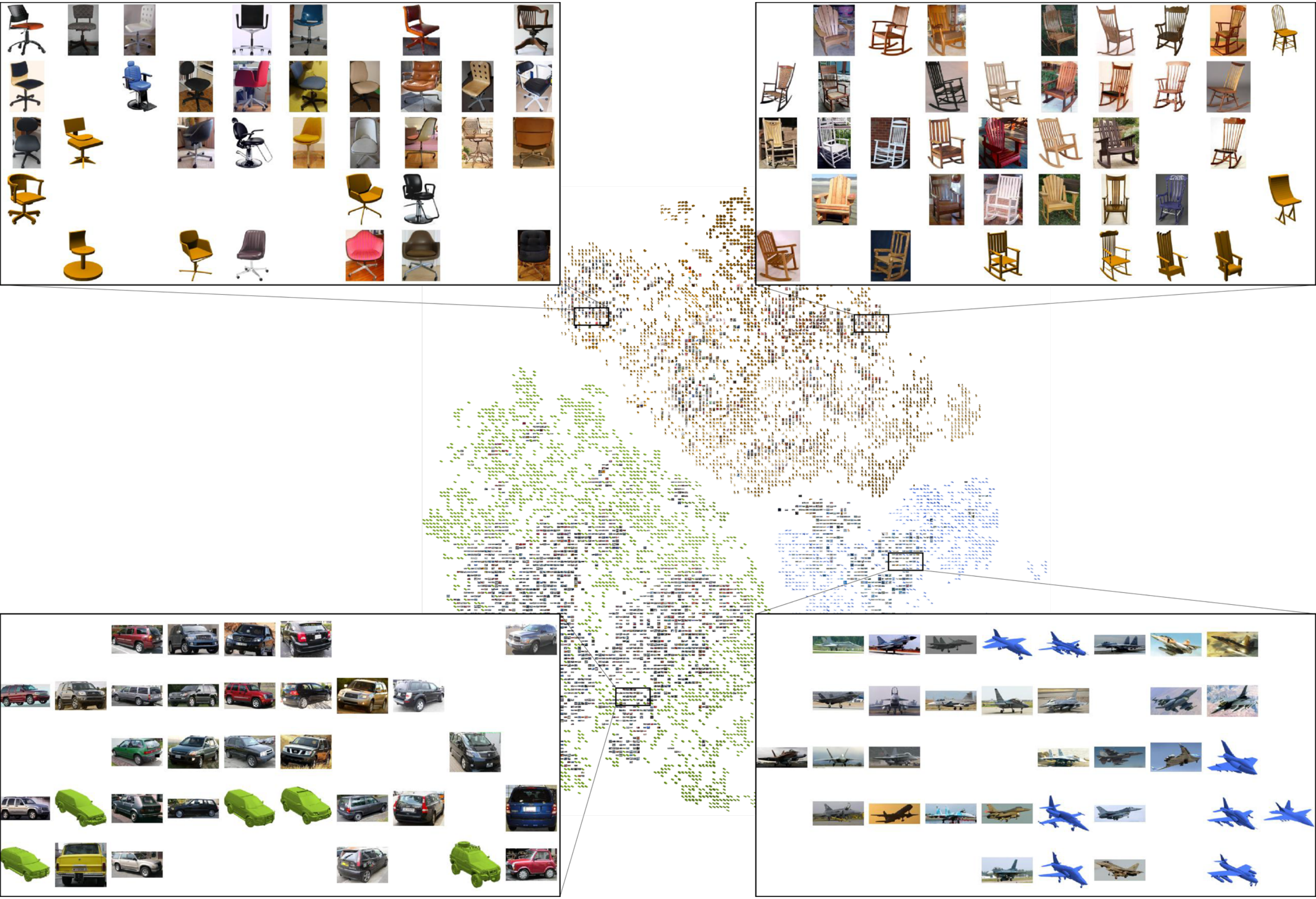
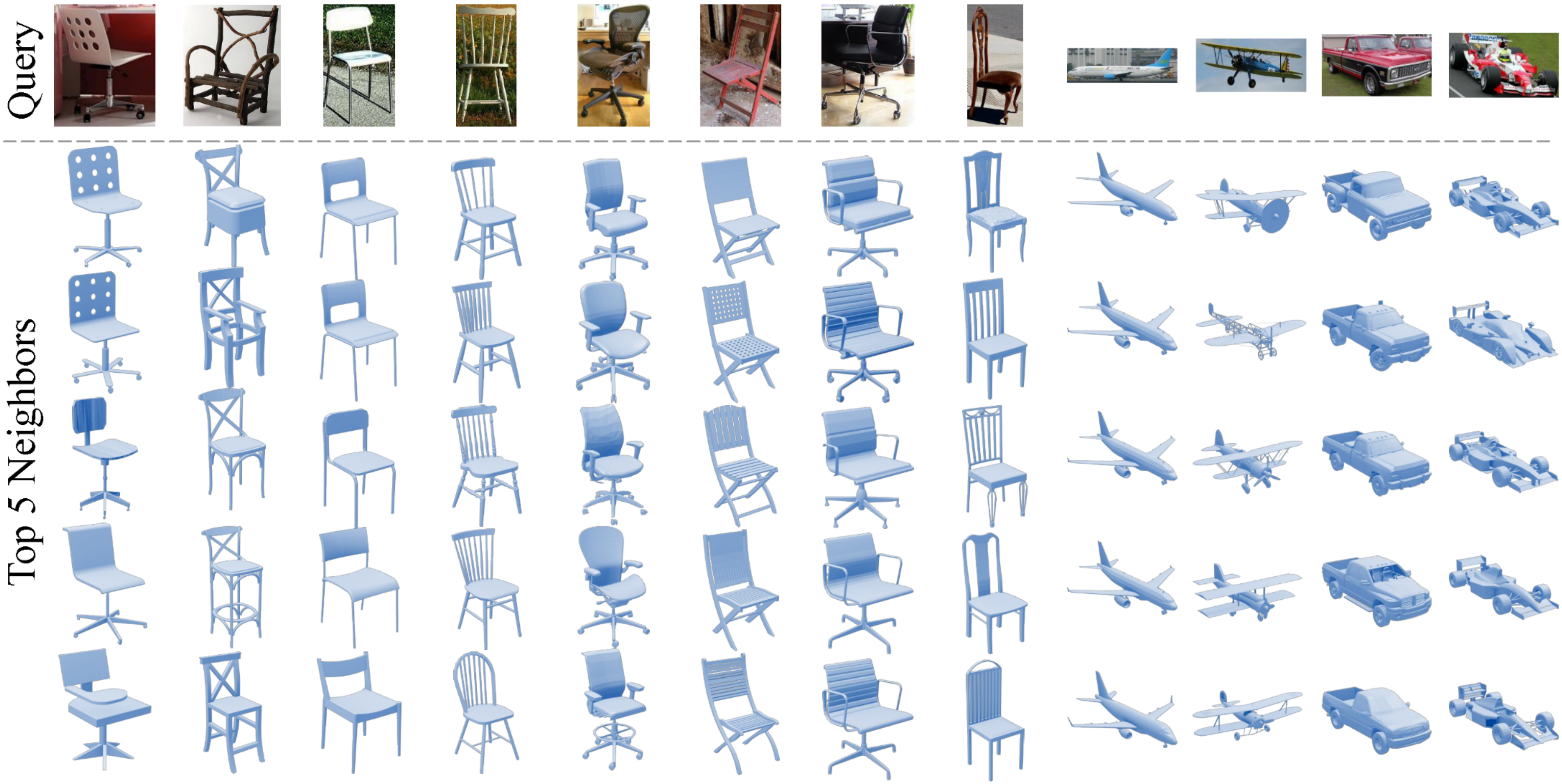
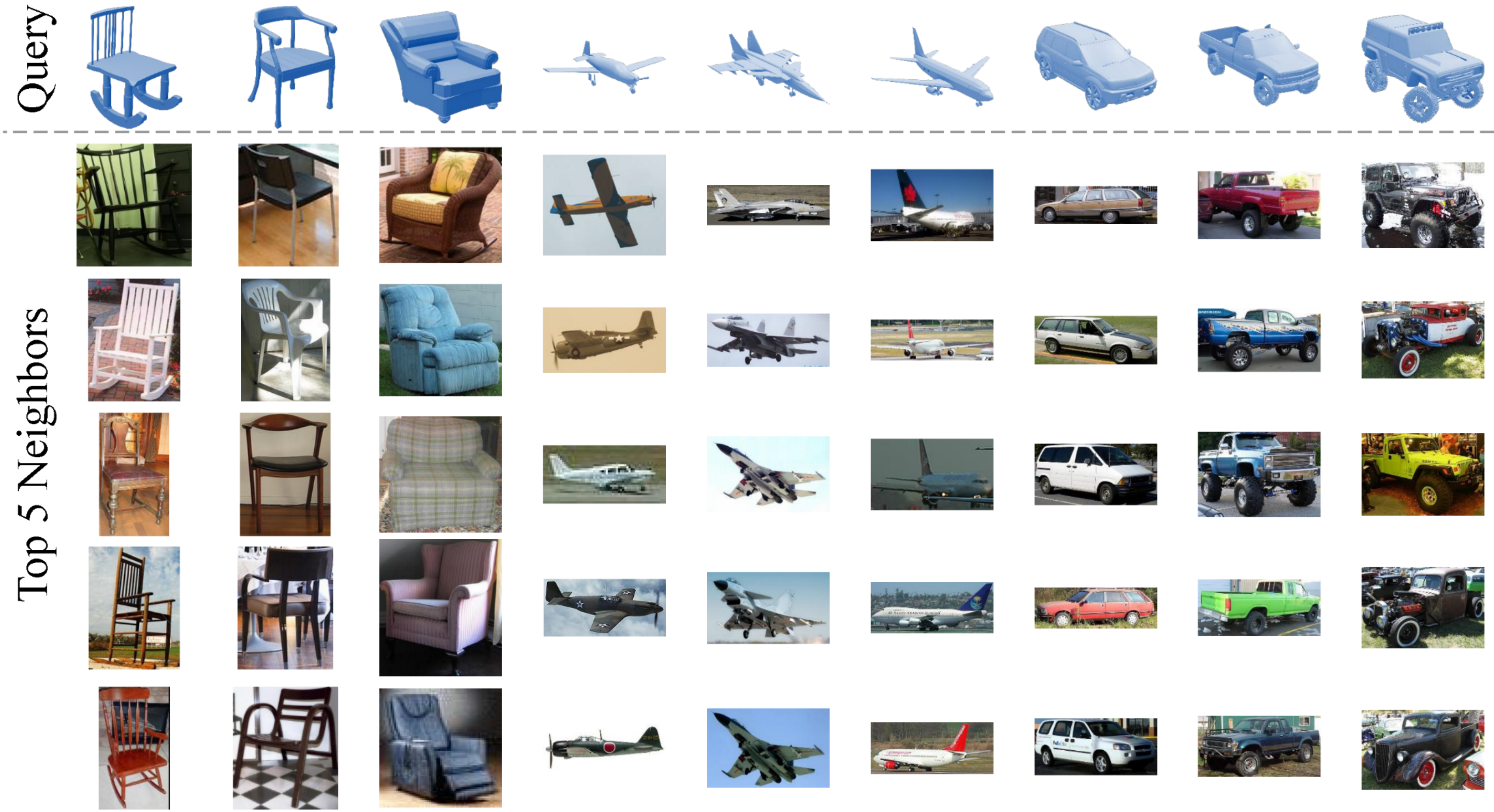


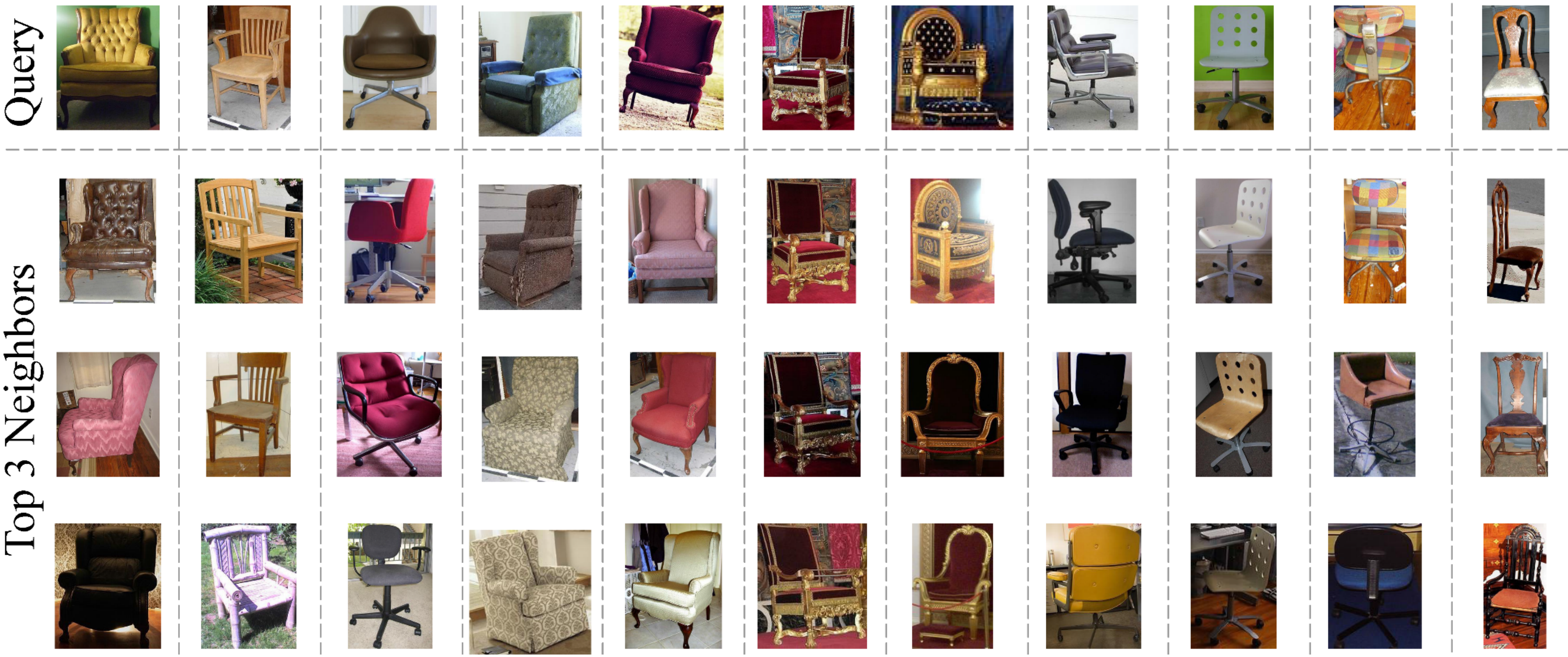
Image based Shape Retrieval

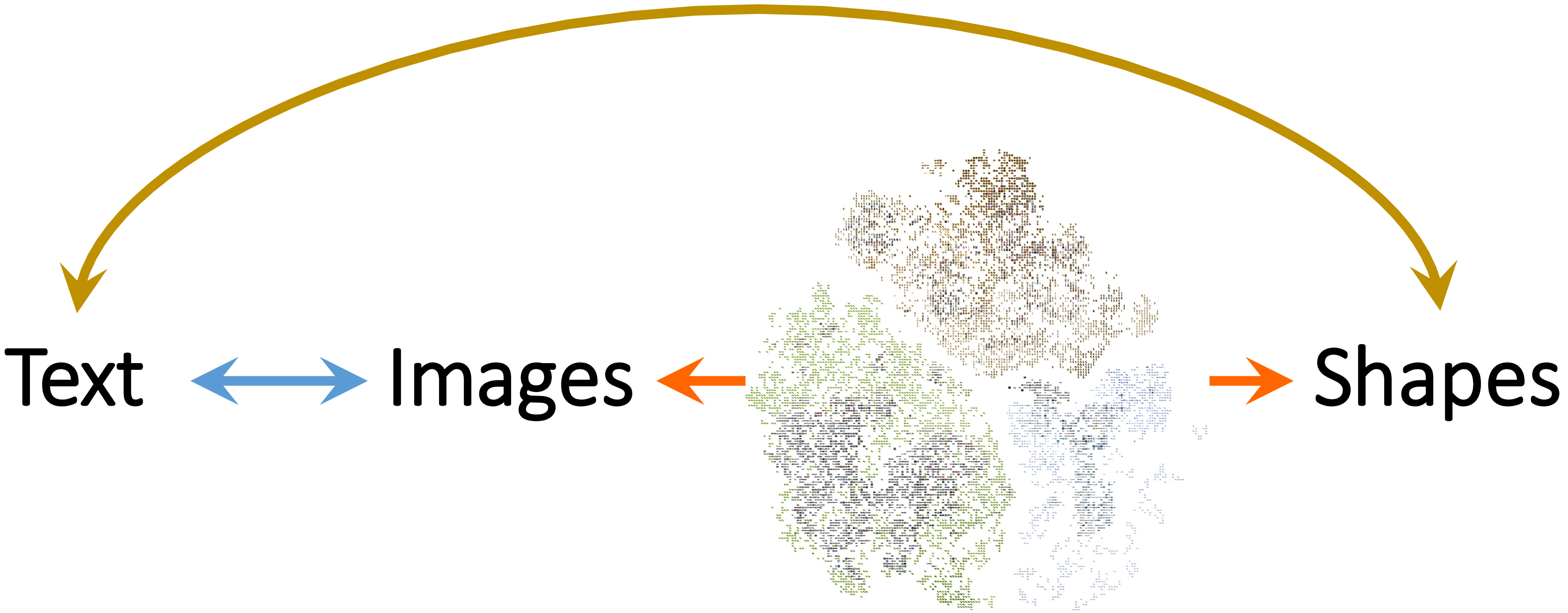


Shape based Image Retrieval



Cross-View Image Retrieval



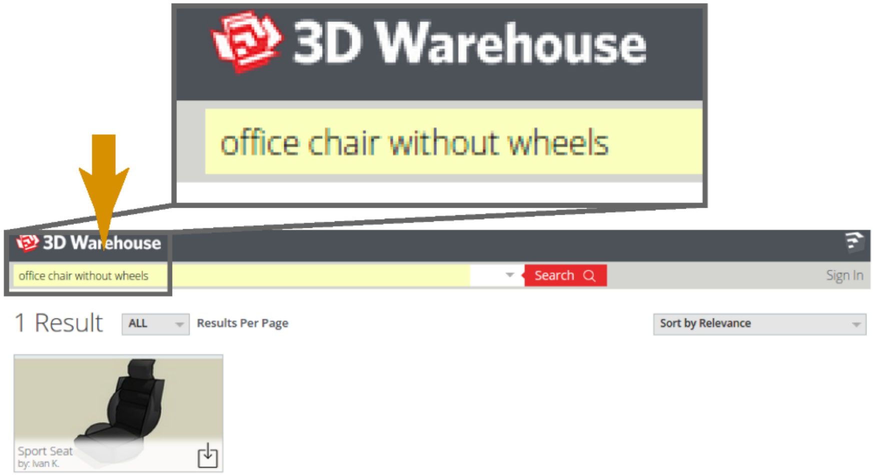


Text

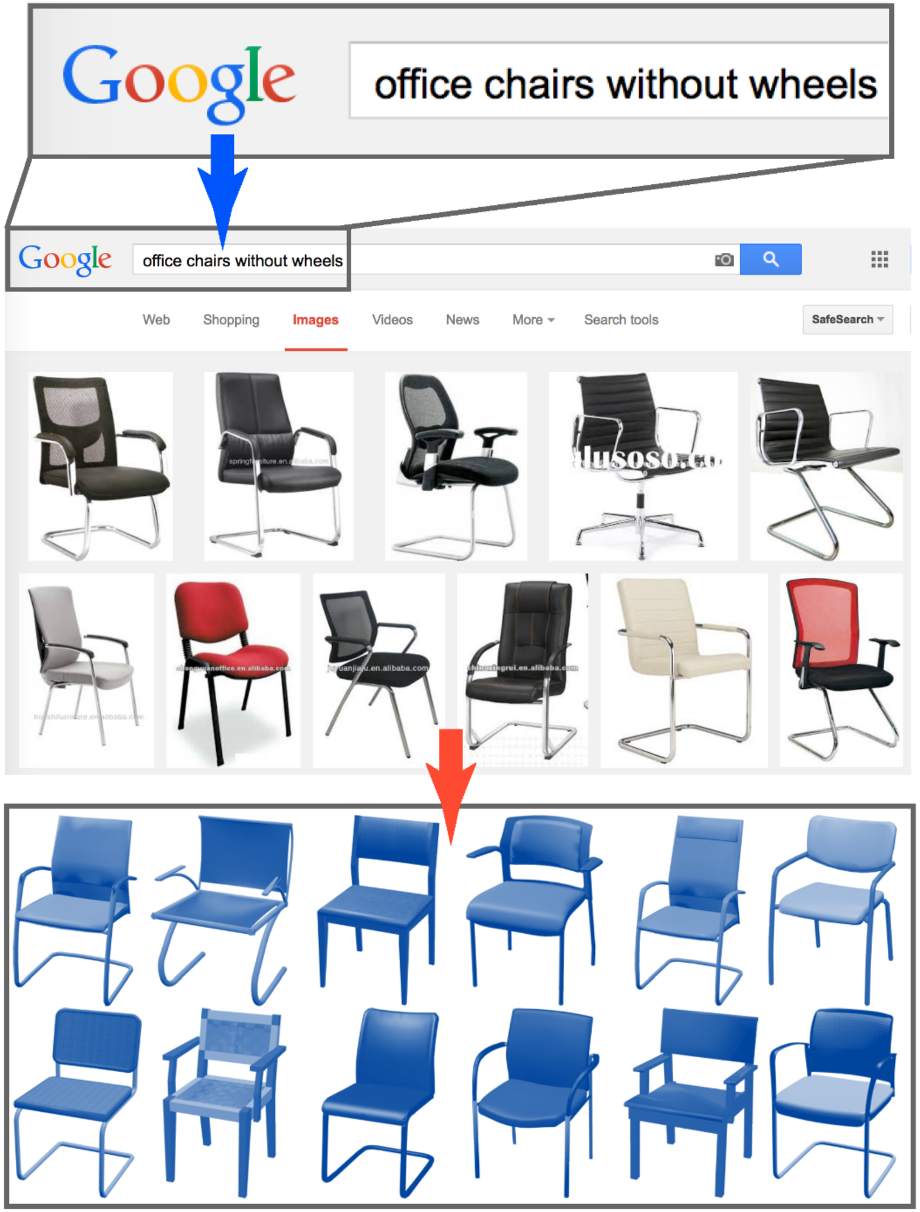
Images

Shapes

Text based Shape Retrieval



office chair without wheels



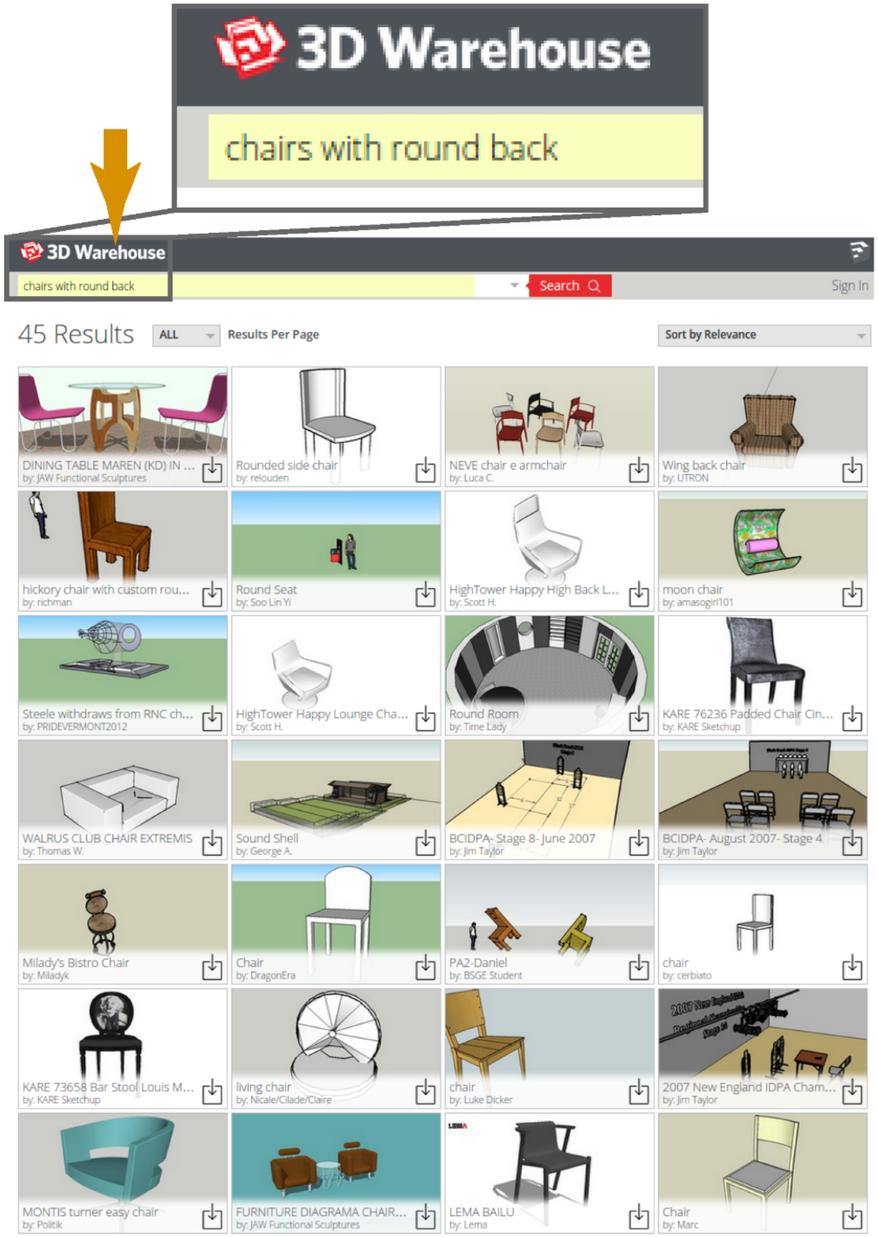
office chairs without wheels

Text to Shape Query (TS)

Text to Image Query (TI)

Image to Shape Query (IS)

Text based Shape Retrieval



3D Warehouse

chairs with round back

Search

Sign In

45 Results ALL Results Per Page Sort by Relevance

DINING TABLE MAREN (KD) IN ...
by: JAW Functional Sculptures

Rounded side chair
by: relouden

NEVE chair e armchair
by: Luca C.

Wing back chair
by: UTRON

hickory chair with custom rou...
by: richman

Round Seat
by: Soo Lin Yi

HighTower Happy High Back L...
by: Scott H.

moon chair
by: amasogirl101

Steele withdraws from RNC ch...
by: PRIDEVERMONT2012

HighTower Happy Lounge Cha...
by: Scott H.

Round Room
by: Time Lady

KARE 76236 Padded Chair Cin...
by: KARE Sketchup

WALRUS CLUB CHAIR EXTREMIS
by: Thomas W.

Sound Shell
by: George A.

BCIDPA- Stage 8- June 2007
by: Jim Taylor

BCIDPA- August 2007- Stage 4
by: Jim Taylor

Milady's Bistro Chair
by: Miladyk

Chair
by: DragonEra

PA2-Daniel
by: BSGE Student

chair
by: cerbiato

KARE 73658 Bar Stool Louis M...
by: KARE Sketchup

living chair
by: Nicale/Cilade/Claire

chair
by: Luke Dicker

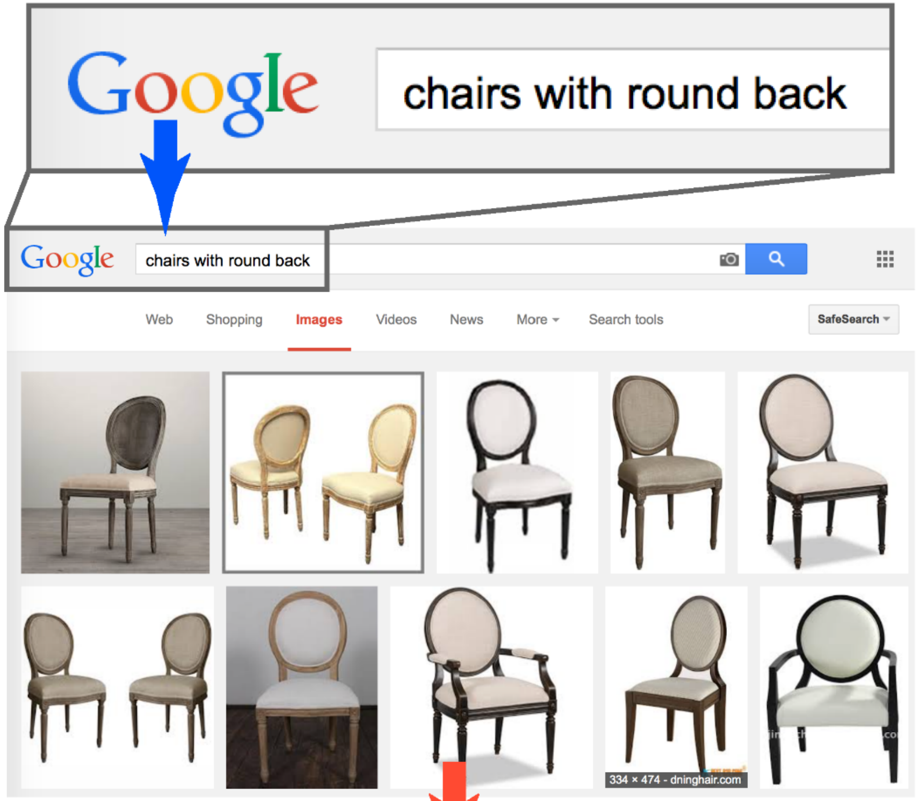
2007 New England IDPA Cham...
by: Jim Taylor

MONTIS turner easy chair
by: Politik

FURNITURE DIAGRAMA CHAIR...
by: JAW Functional Sculptures

LEMA BAILU
by: Lema

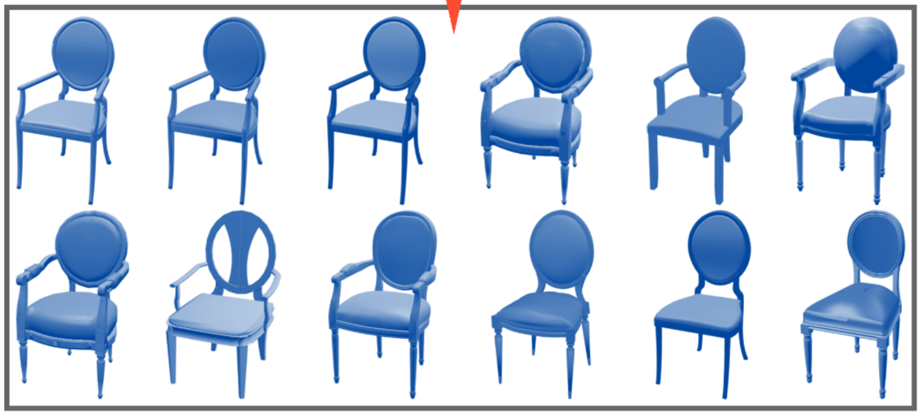
Chair
by: Marc

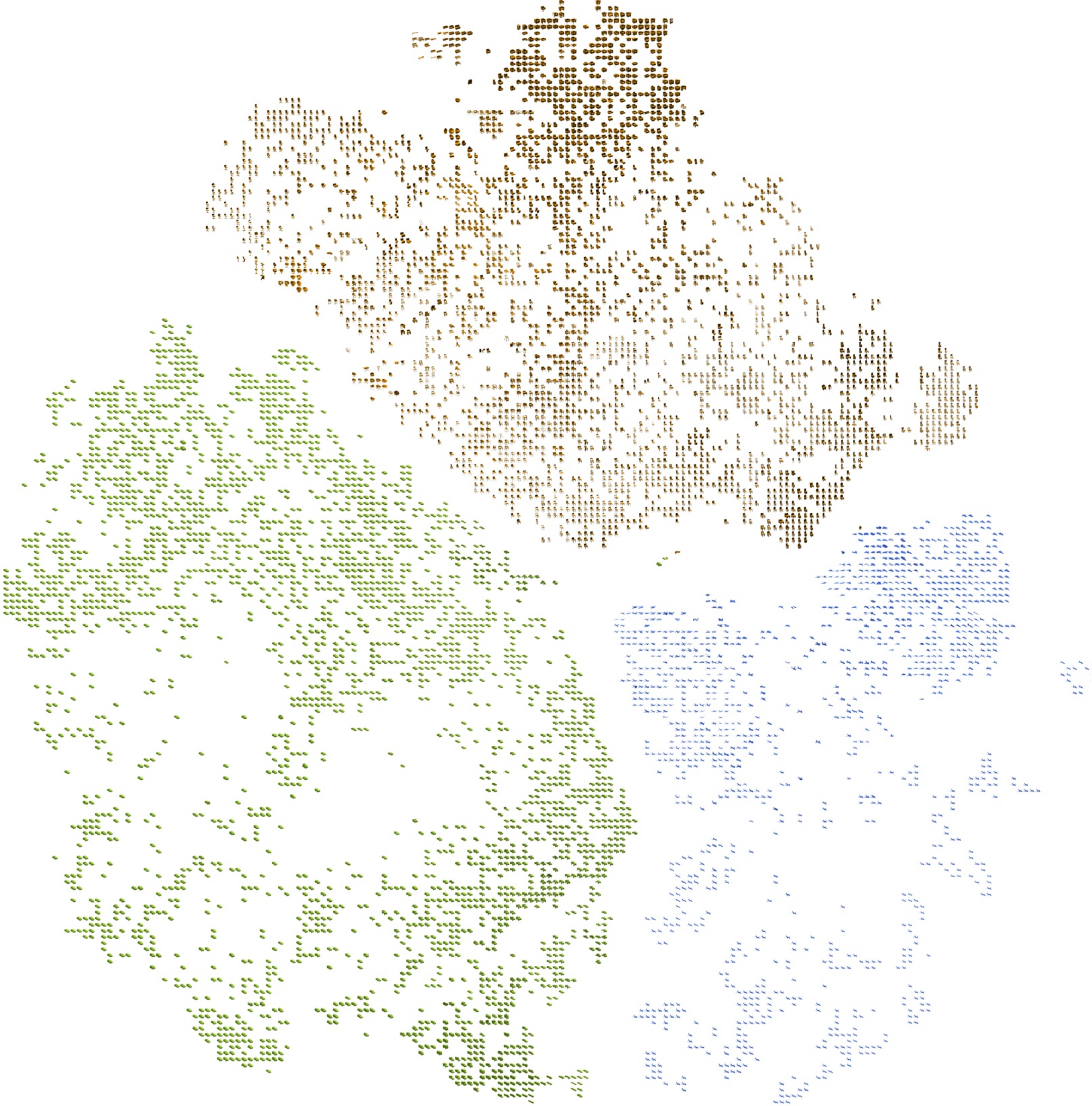


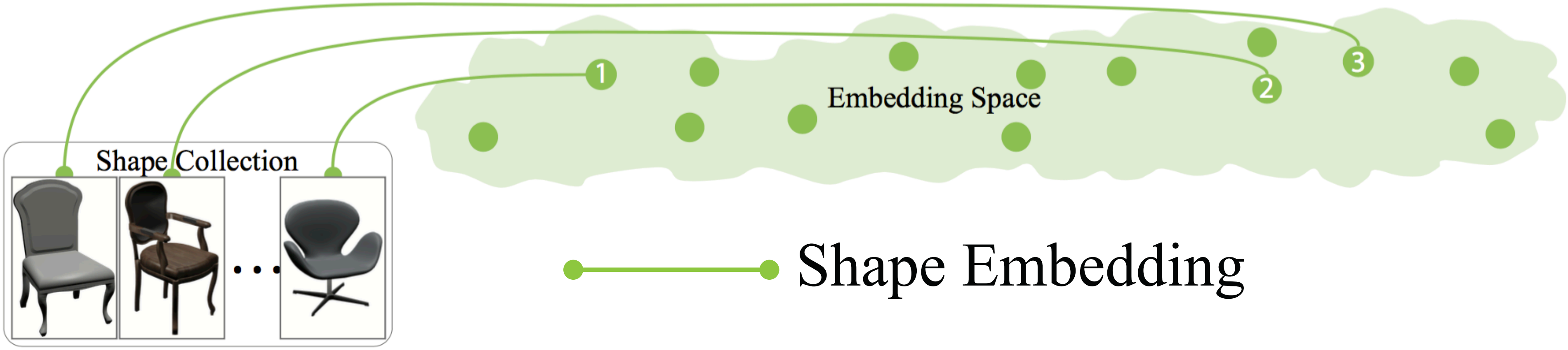
Google chairs with round back

Web Shopping Images Videos News More Search tools SafeSearch

334 × 474 - dninghair.com



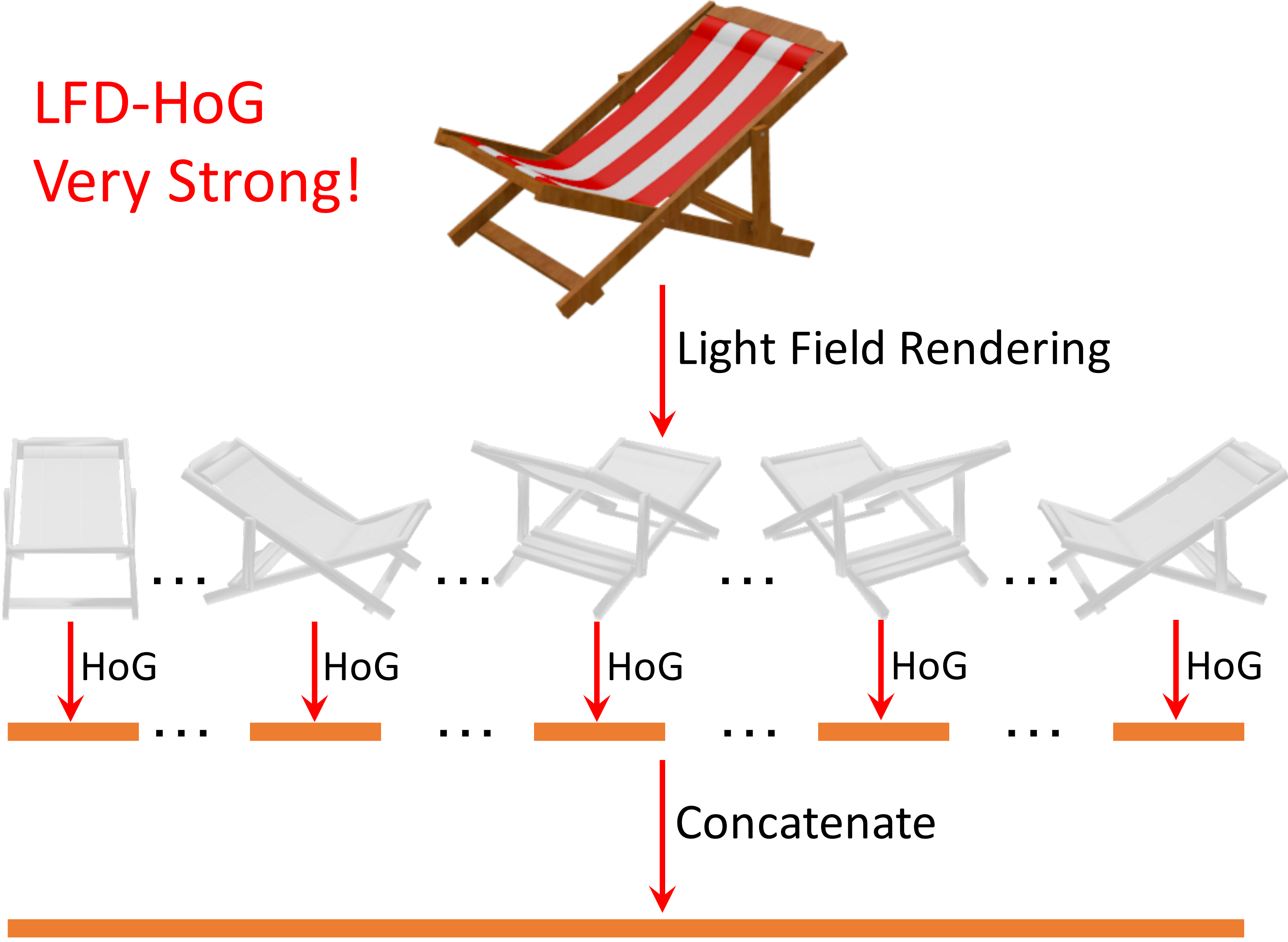




$$Similarity_{(s_i,s_j)} = \|\mathcal{P}_i - \mathcal{P}_j\|$$

Many choices for $\mathcal{P}_i$:
Shape Histograms, Spin Images, Spherical Harmonics, Shape Distributions, etc.

LFD-HoG
Very Strong!



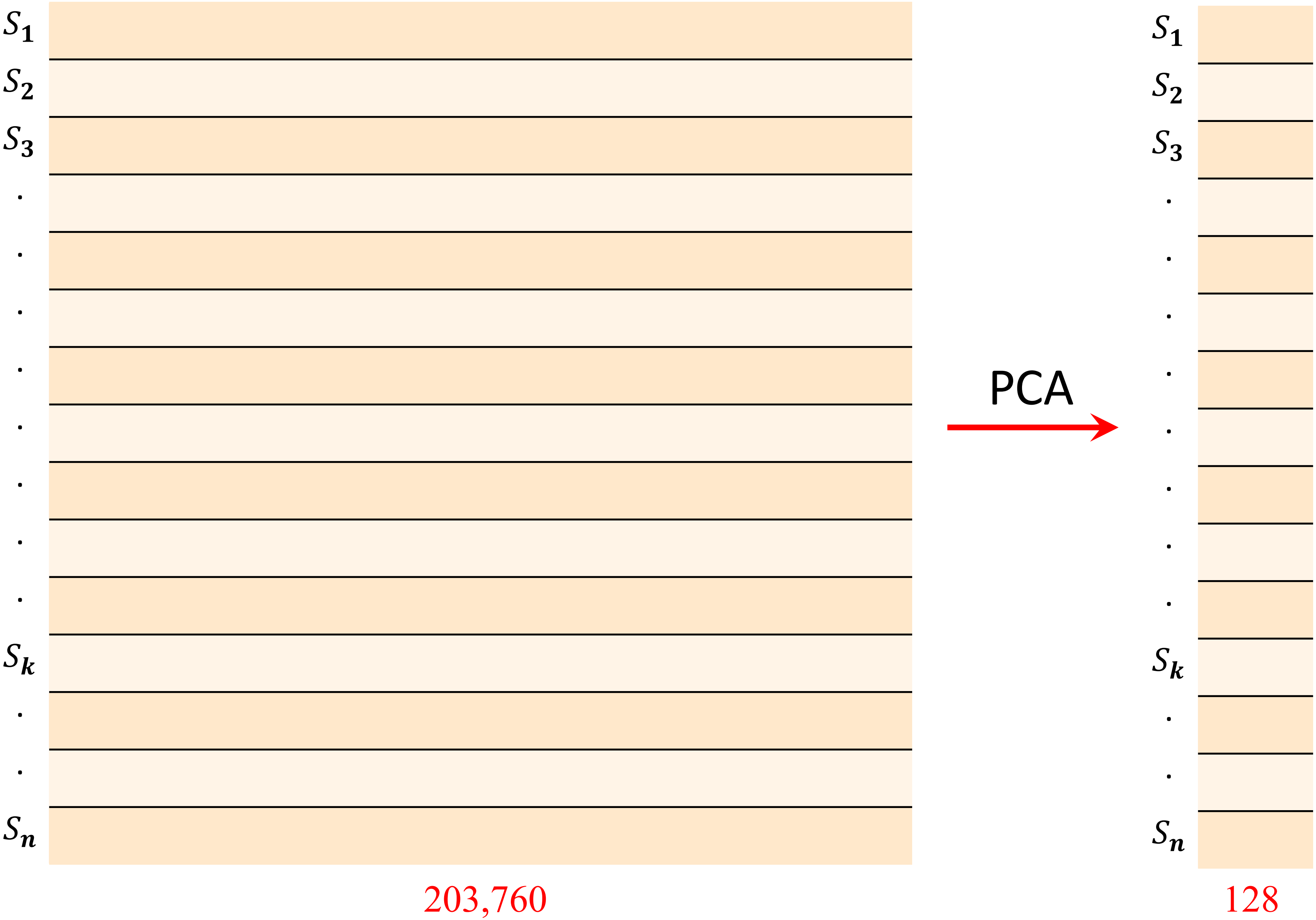
Light Field Rendering

... ...

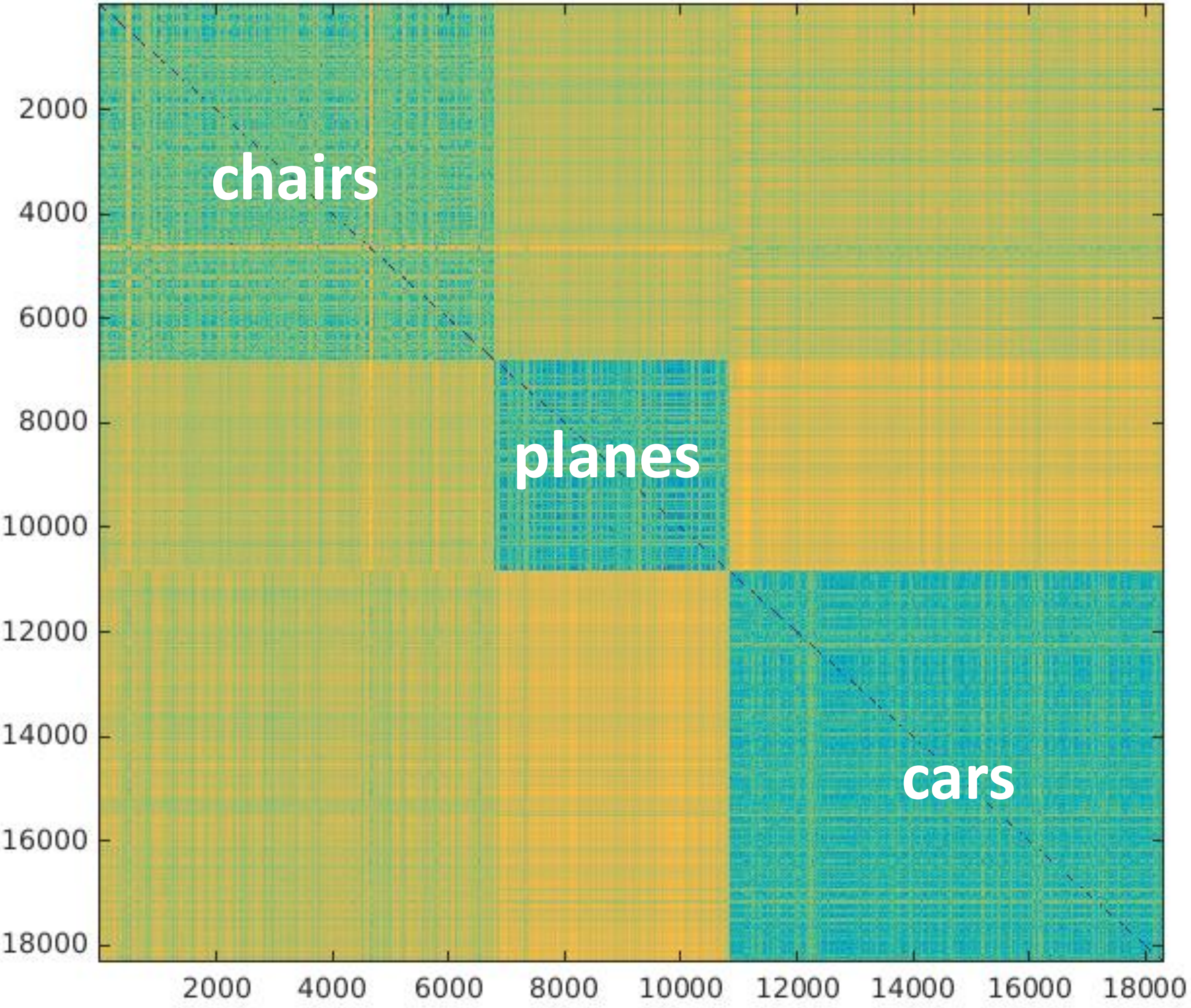
HoG HoG HoG HoG HoG

... ...

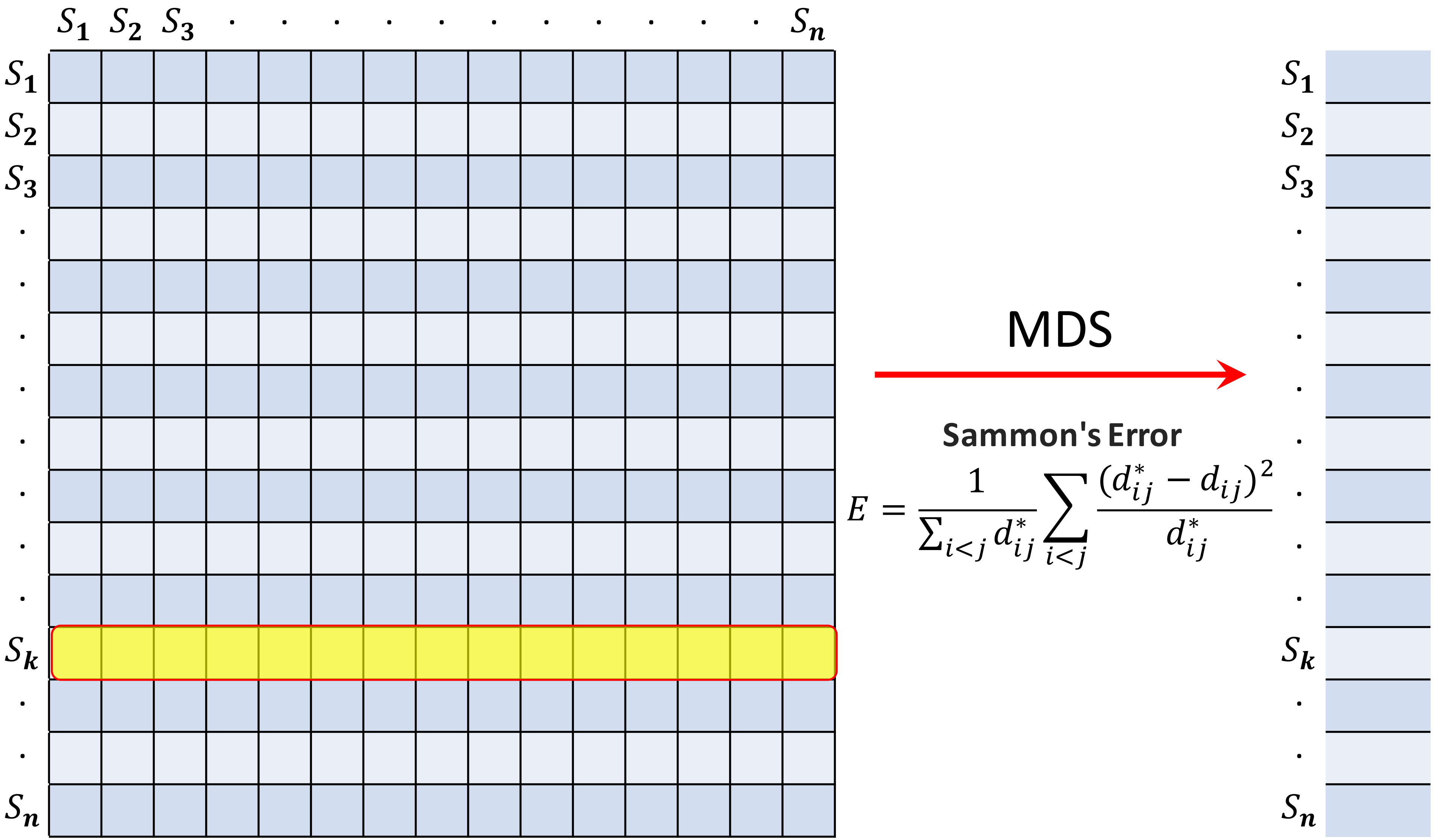
Concatenate



S_1
S_2
S_3
S_k
S_n
203,760
PCA
S_1
S_2
S_3
S_k
S_n
128



Distance Matrix: $d(S_i, S_j)$ in the $(i, j) - th$ element



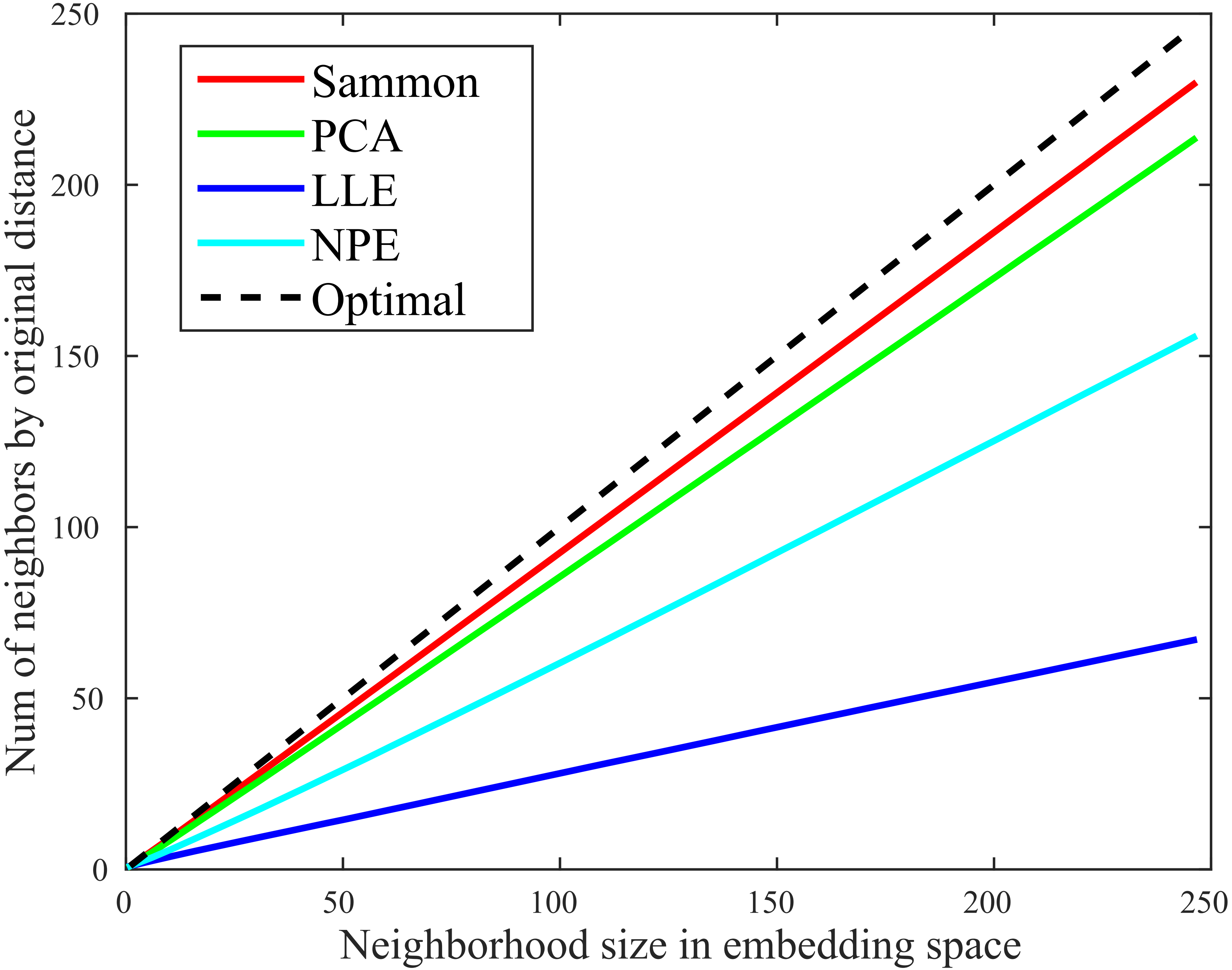
MDS

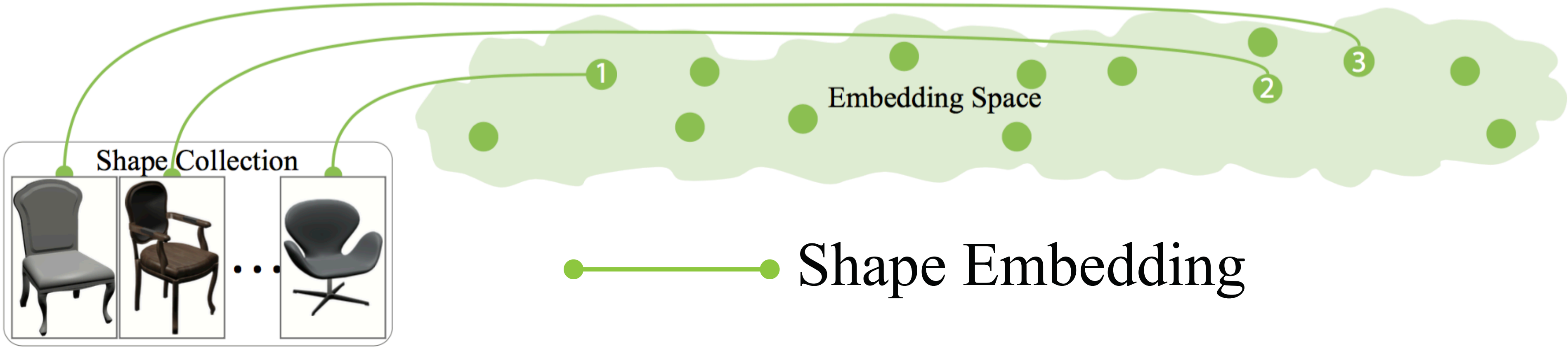
Sammon's Error

$$E = \frac{1}{\sum_{i<j} d_{ij}^*} \sum_{i<j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*}$$

Distance Matrix: $d(S_i, S_j)$ in the $(i, j) - th$ element

Each row can serve as the embedding point


Sammon
PCA
LLE
NPE
Optimal
Num of neighbors by original distance
0
50
100
150
200
250
Neighborhood size in embedding space
0
50
100
150
200
250

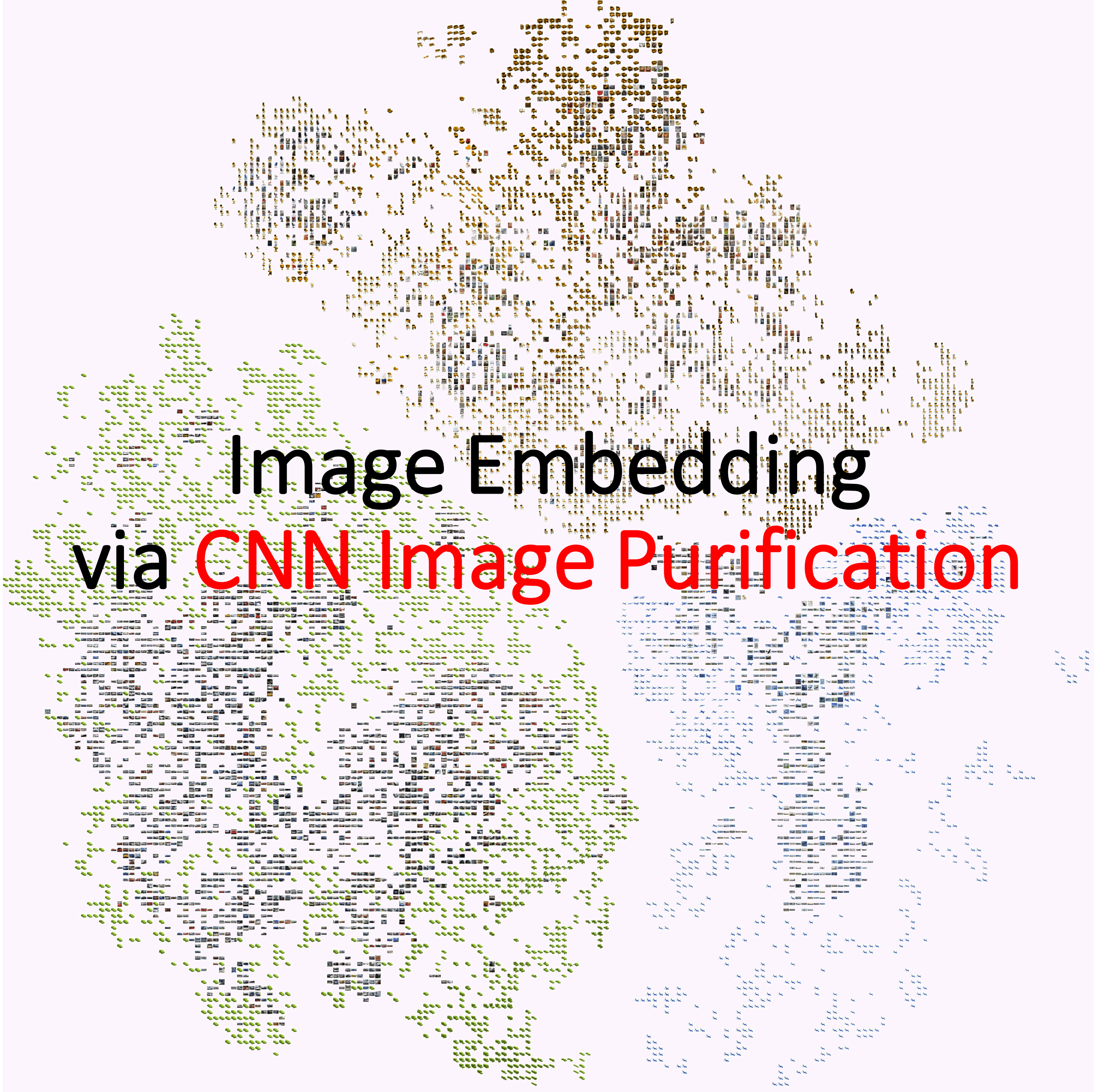


$$Similarity_{(s_i,s_j)} = \|\mathcal{P}_i - \mathcal{P}_j\|$$

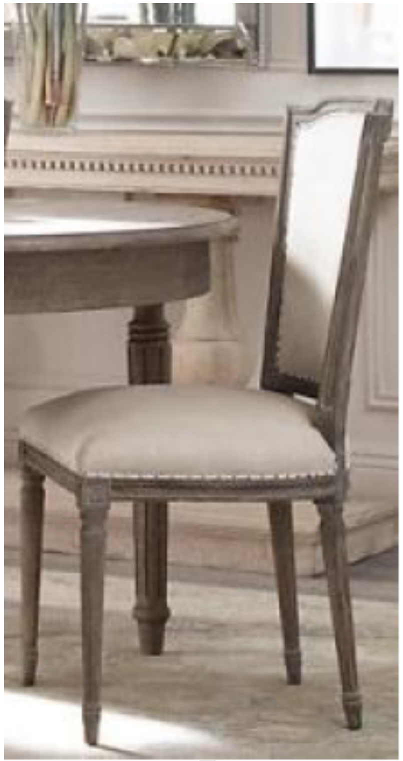
Our choice of embedding point $\mathcal{P}_i$:

1. Extract Light Field HoG Descriptors
2. Compute Distance Matrix
3. MDS with Sammon's Error

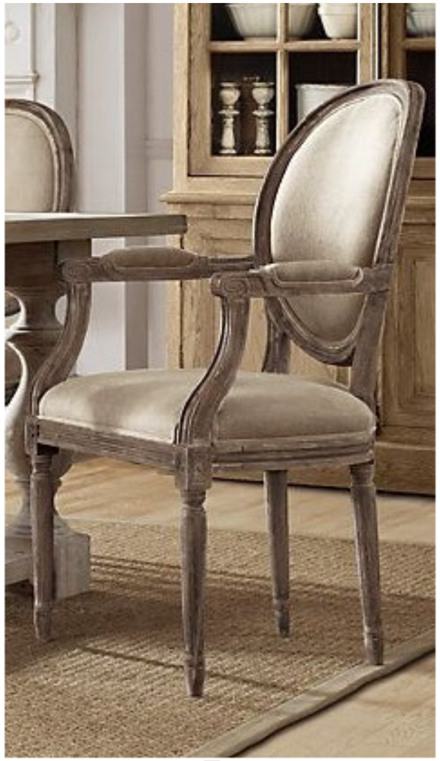
Image Embedding via CNN Image Purification



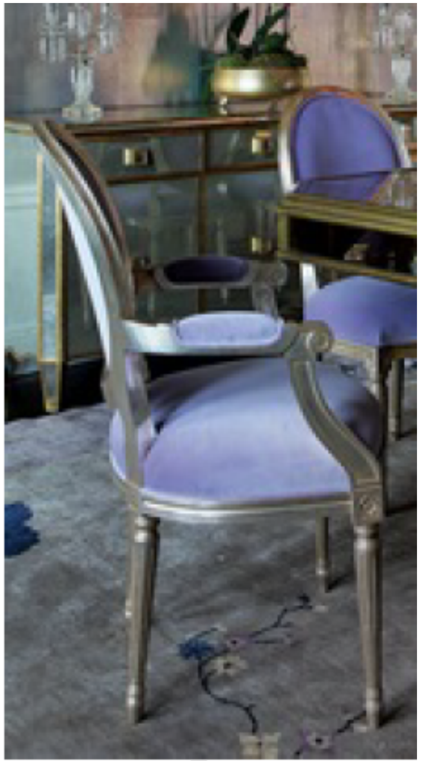
Deep learning, yay or nay?



I_1



I_2



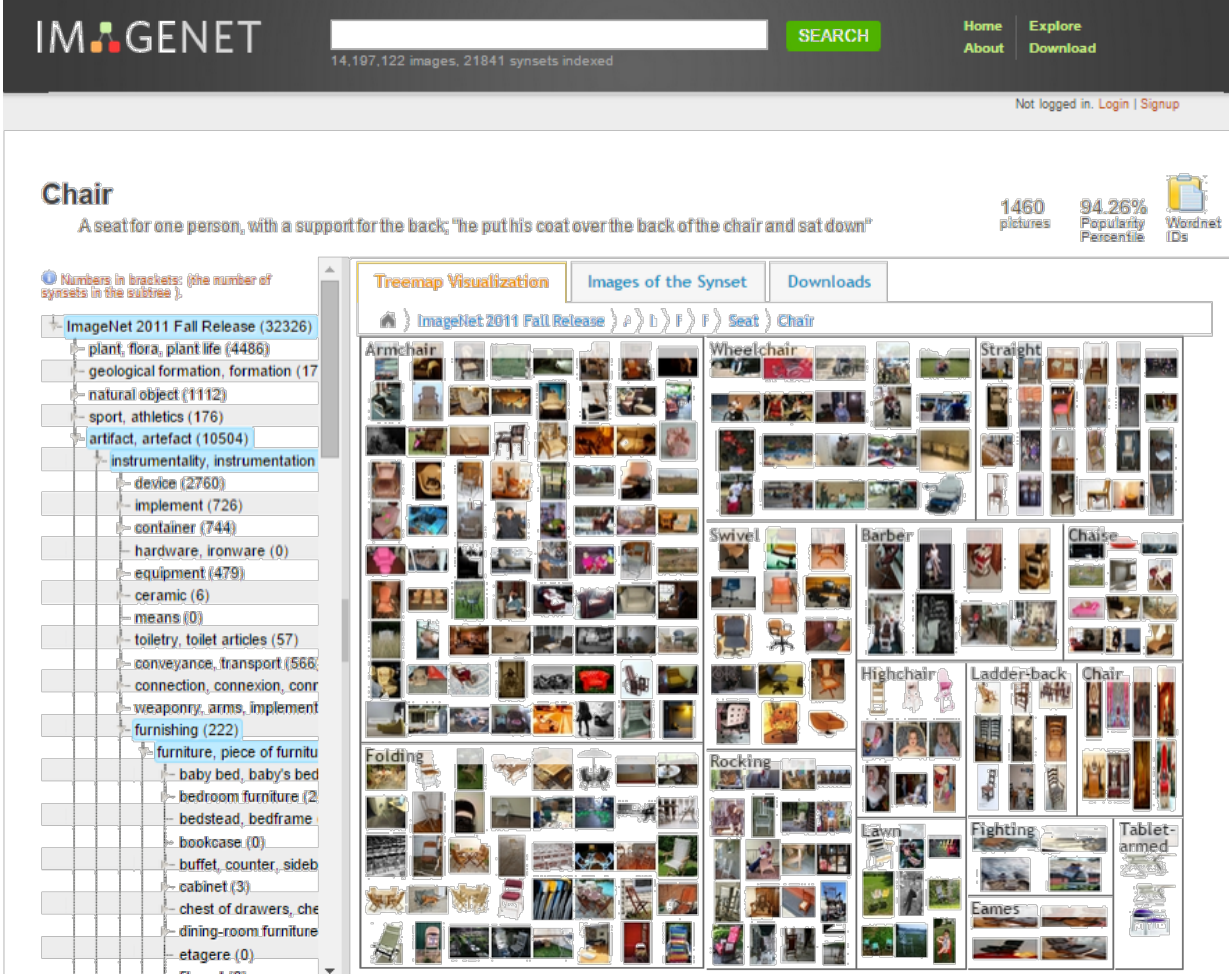
I_3

$$\mathcal{P}_i = f(I_i)$$

$$\|\mathcal{P}_2 - \mathcal{P}_3\| < \|\mathcal{P}_1 - \mathcal{P}_2\|$$

A piece of cake,
elementary math…

What the hell is
the f?



14,197,122 images, 21841 synsets indexed

Not logged in. Login | Signup

Chair

A seat for one person, with a support for the back; "he put his coat over the back of the chair and sat down"

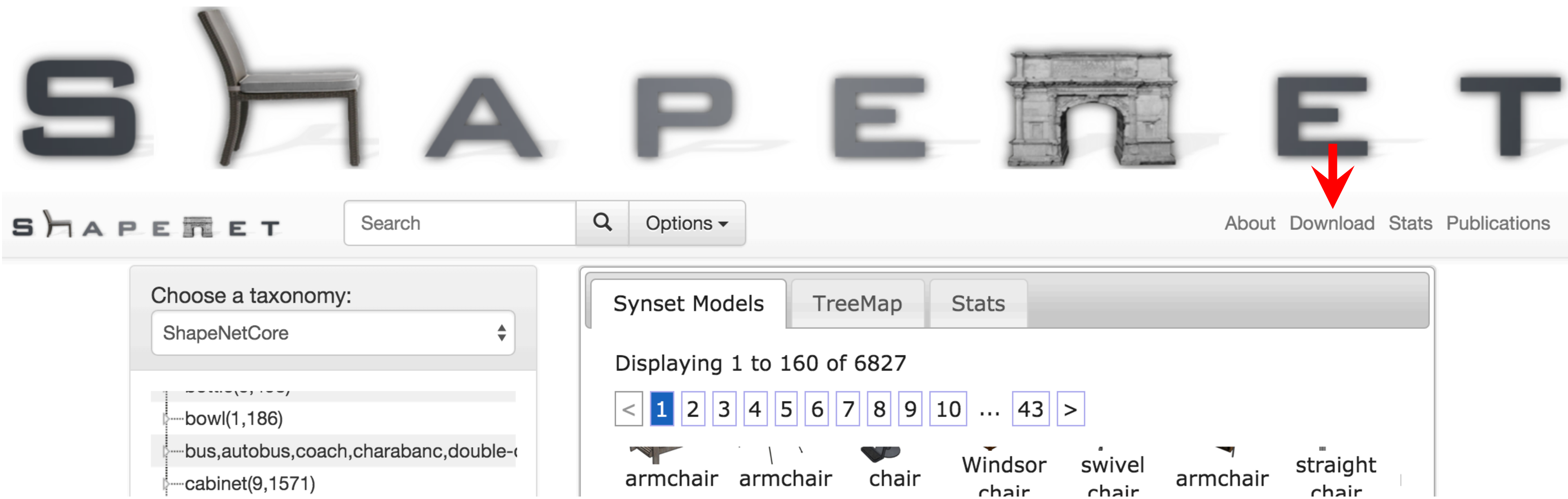
1460 pictures

94.26% Popularity Percentile

Wordnet IDs

Numbers in brackets: (the number of synsets in the subtree).

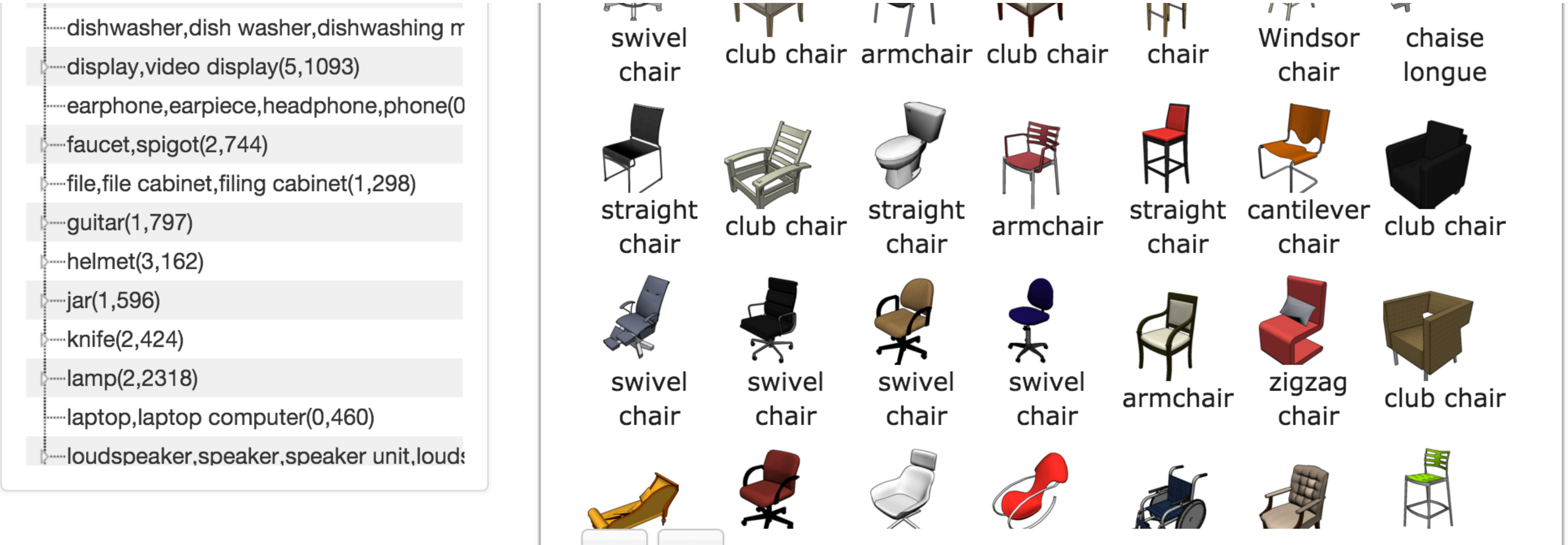
- ImageNet 2011 Fall Release (32326)
  - plant, flora, plant life (4486)
  - geological formation, formation (17
  - natural object (1112)
  - sport, athletics (176)
  - artifact, artefact (10504)
    - instrumentality, instrumentation
      - device (2760)
      - implement (726)
      - container (744)
      - hardware, ironware (0)
      - equipment (479)
      - ceramic (6)
      - means (0)
      - toiletry, toilet articles (57)
      - conveyance, transport (566
      - connection, connexion, conn
      - weaponry, arms, implement
      - furnishing (222)
        - furniture, piece of furnitu
          - baby bed, baby's bed
          - bedroom furniture (2
          - bedstead, bedframe
          - bookcase (0)
          - buffet, counter, sideb
          - cabinet (3)
          - chest of drawers, che
          - dining-room furniture
          - etagere (0)



Choose a taxonomy: ShapeNetCore

Displaying 1 to 160 of 6827

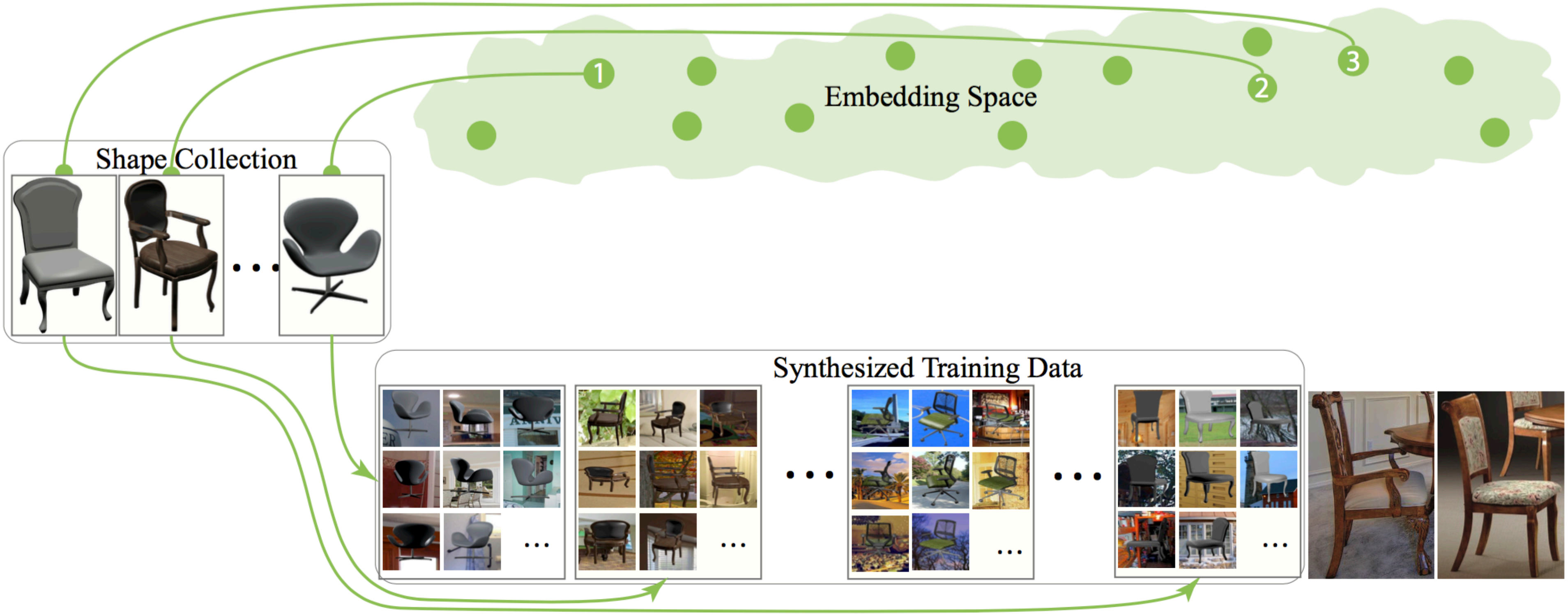
http://shapenet.org



- ✓ A model is worth a thousand images!
- ✓ Rendering: $\mathrm{Image} = f(Properties)$
- ✓ Computer Vision: $\mathrm{Properties} = f^{-1}(Image)$

It eats, a lot!

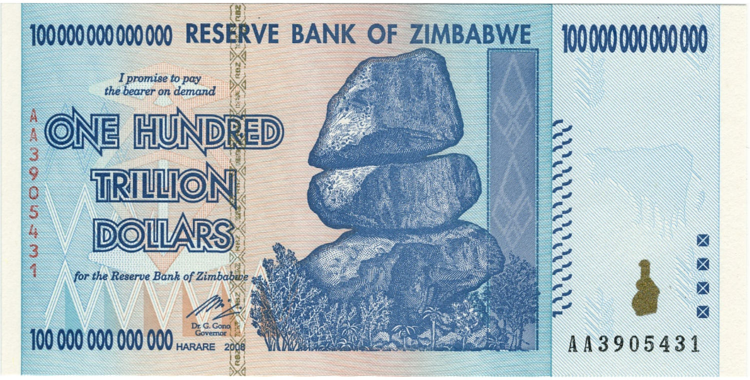
✓



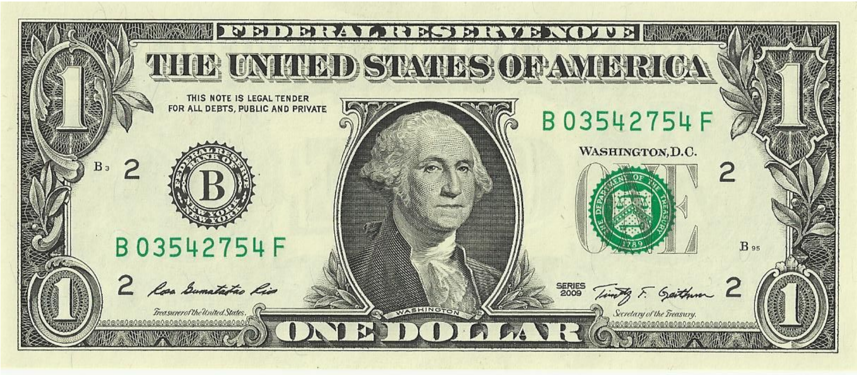
Shape Embedding

Image Synthesis

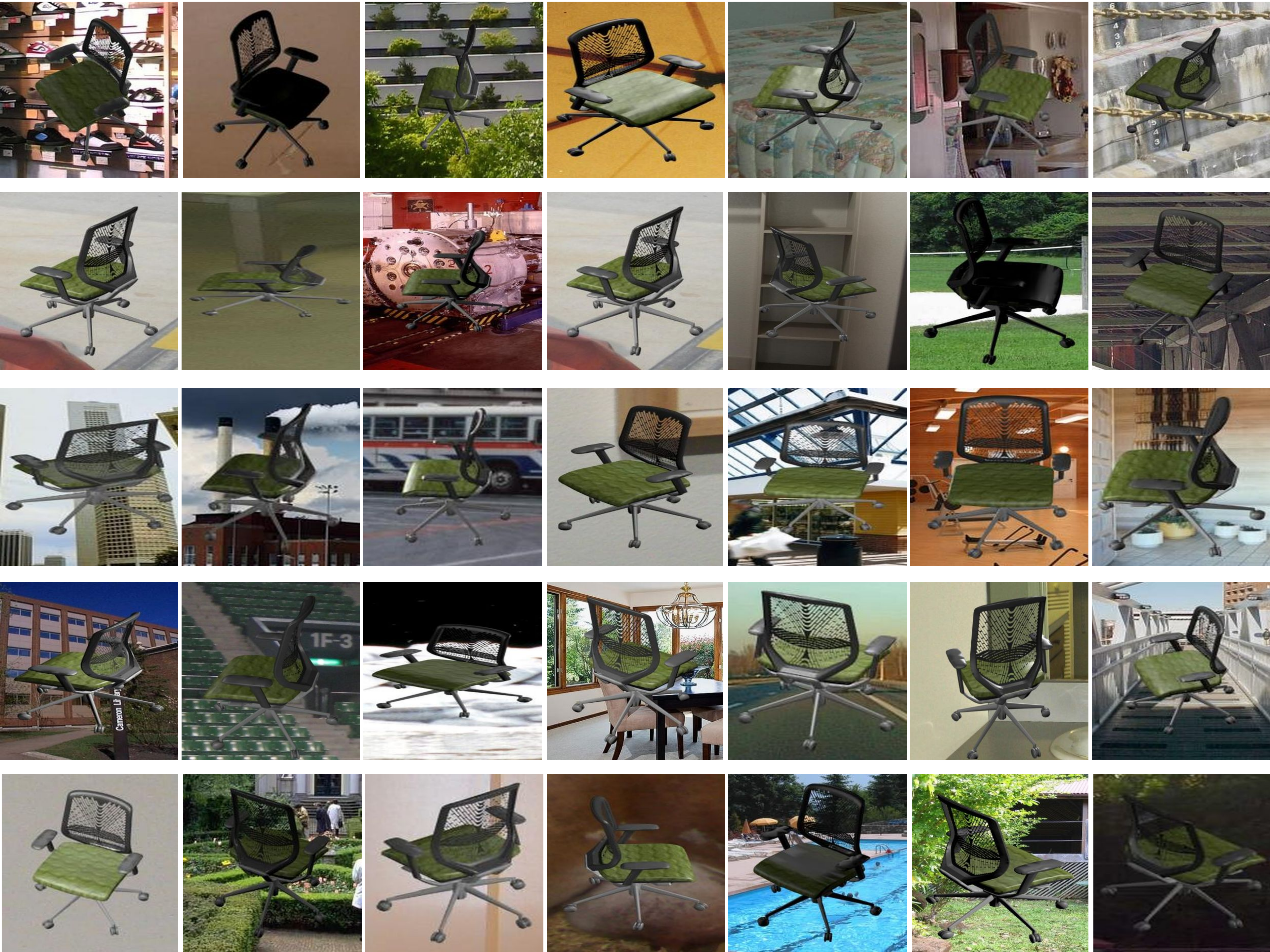
Many image-point pairs $(I_{S_i}, \mathcal{P}_i)$

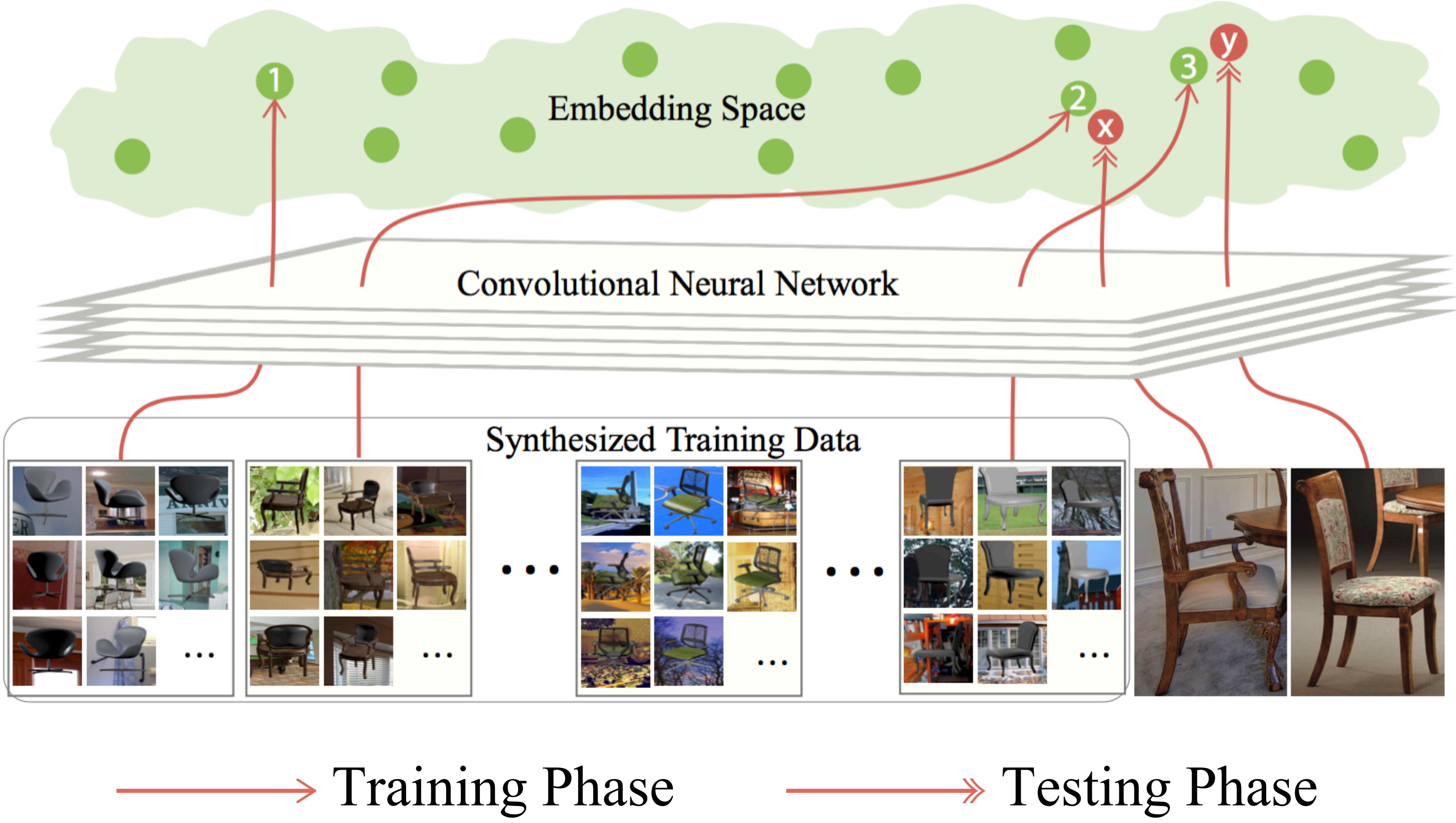

100 000 000 000 000
RESERVE BANK OF ZIMBABWE
I promise to pay the bearer on demand
ONE HUNDRED TRILLION DOLLARS
for the Reserve Bank of Zimbabwe
Dr G. Gono Governor
HARARE 2008
AA3905431

$\neq 10^{14} *$


FEDERAL RESERVE NOTE
THE UNITED STATES OF AMERICA
THIS NOTE IS LEGAL TENDER FOR ALL DEBTS, PUBLIC AND PRIVATE
B 03542754 F
WASHINGTON, D.C.
SERIES 2009
Treasurer of the United States.
Secretary of the Treasury.
ONE DOLLAR

It’s not only the number…


1F-3
Cameron Li ary



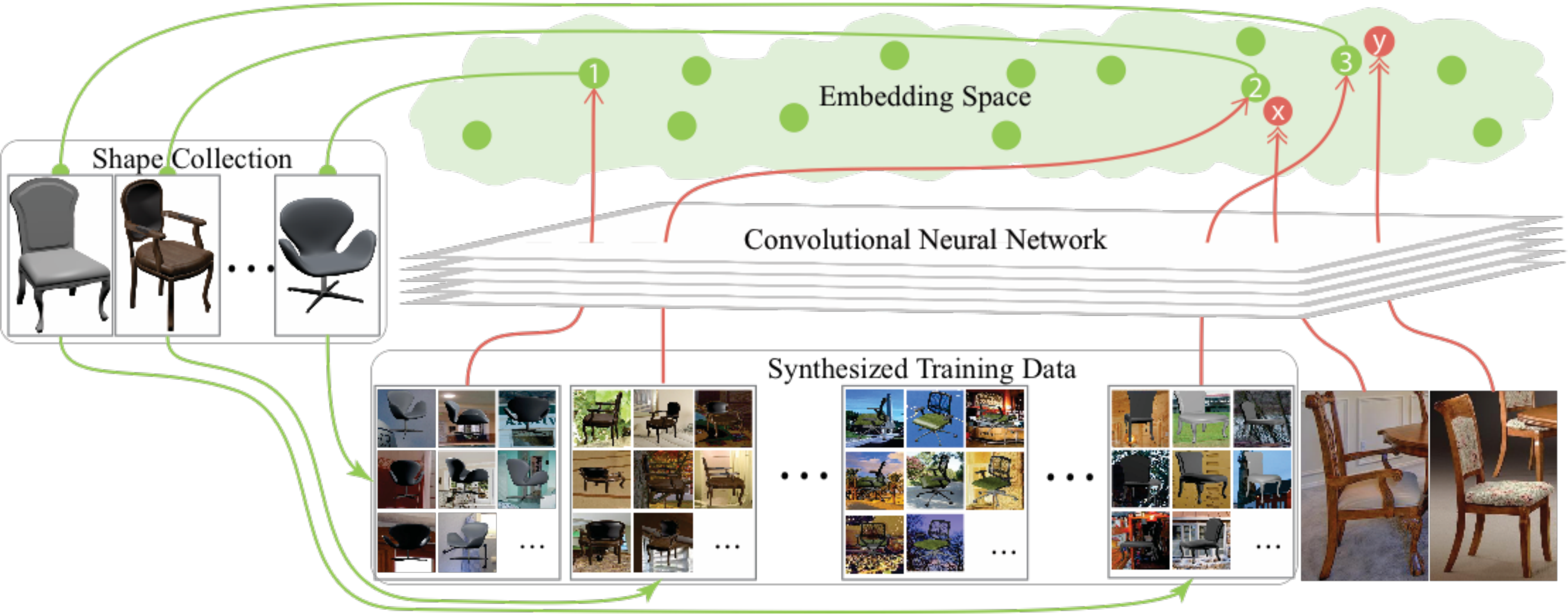
Input: many image-point pairs $(I_{S_i}, \mathcal{P}_i)$

Task: learn the function $\mathcal{P}_i = f(I_{S_i})$

Hey, wake up!

Here comes the most important slide!

●—● Shape Embedding →Precious High Quality Supervision



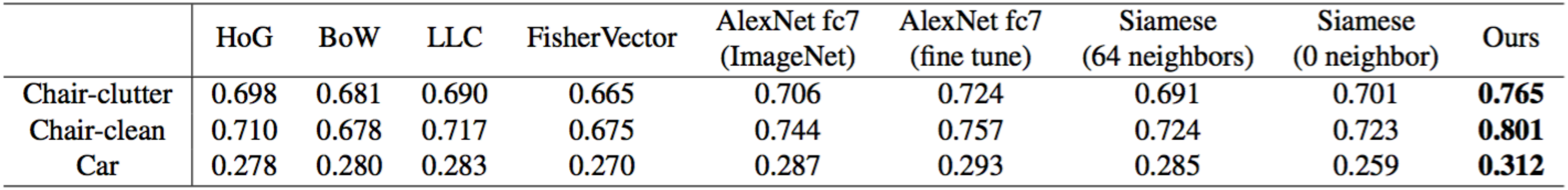
→ Image Synthesis →Messy but Nutritional Training Data

→ Training Phase

↠ Testing Phase

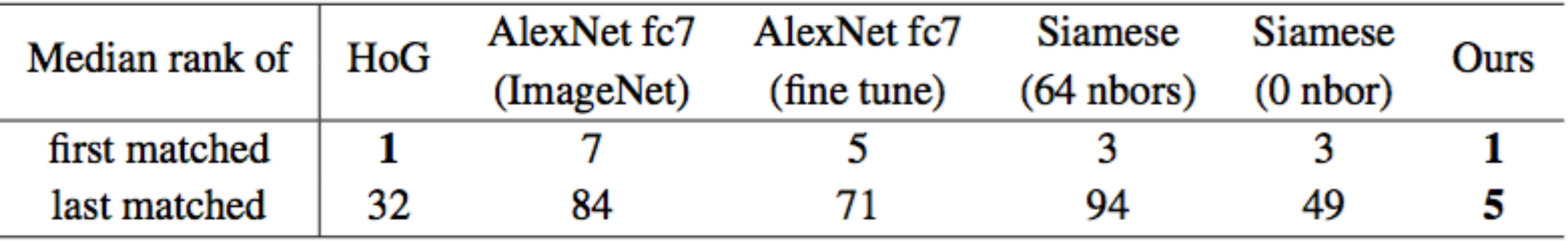
→$\mathcal{P}_i = f(I_{s_i})$, the hell function

Quantitative Evaluation



	HoG	BoW	LLC	FisherVector	AlexNet fc7 (ImageNet)	AlexNet fc7 (fine tune)	Siamese (64 neighbors)	Siamese (0 neighbor)	Ours
Chair-clutter	0.698	0.681	0.690	0.665	0.706	0.724	0.691	0.701	**0.765**
Chair-clean	0.710	0.678	0.717	0.675	0.744	0.757	0.724	0.723	**0.801**
Car	0.278	0.280	0.283	0.270	0.287	0.293	0.285	0.259	**0.312**

AUC of image to image retrieval precision-recall curve



Median rank of	HoG	AlexNet fc7 (ImageNet)	AlexNet fc7 (fine tune)	Siamese (64 nbors)	Siamese (0 nbor)	Ours
first matched	**1**	7	5	3	3	**1**
last matched	32	84	71	94	49	**5**

First and last image match rankings in shape to image retrieval

Quantitative Evaluation

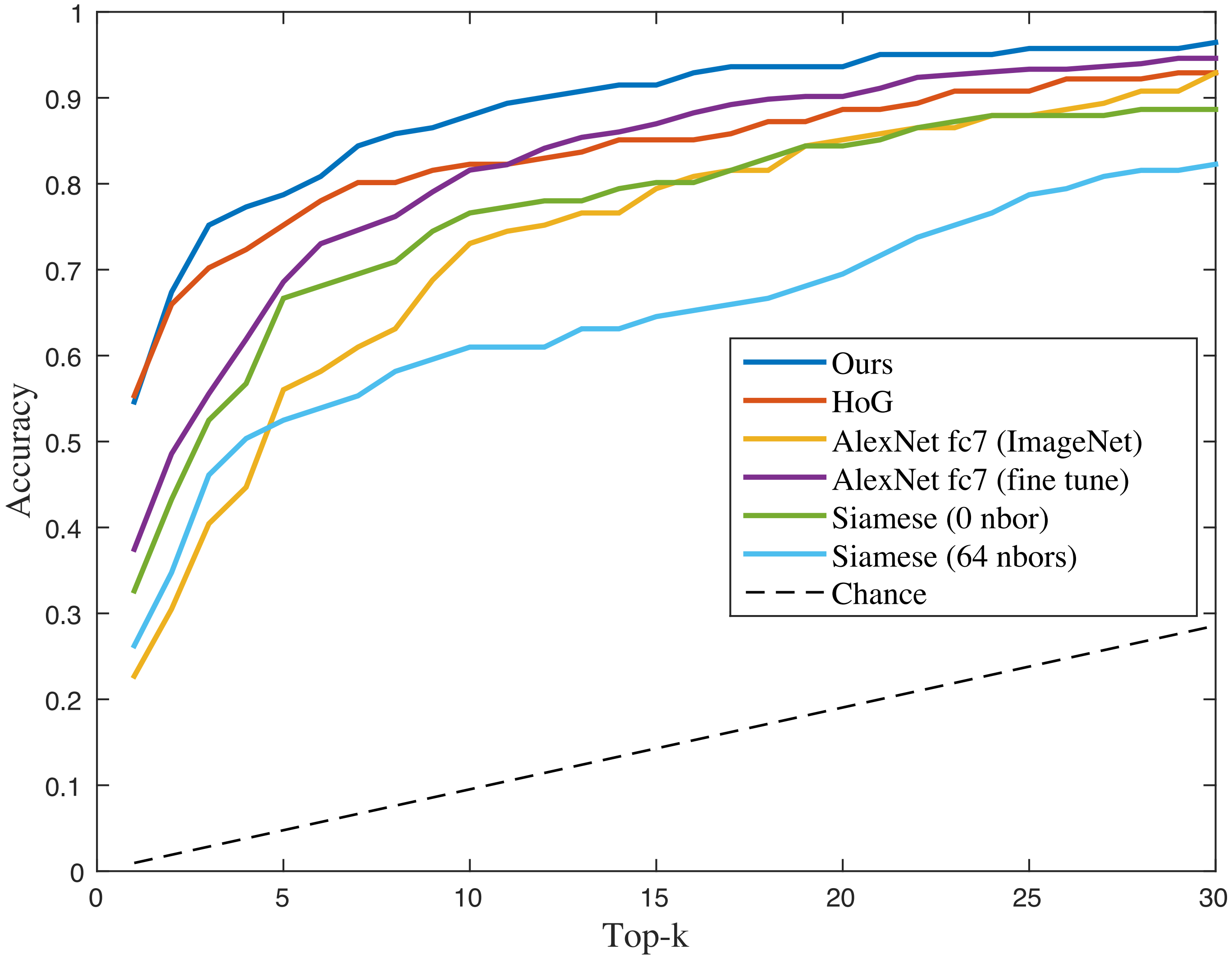
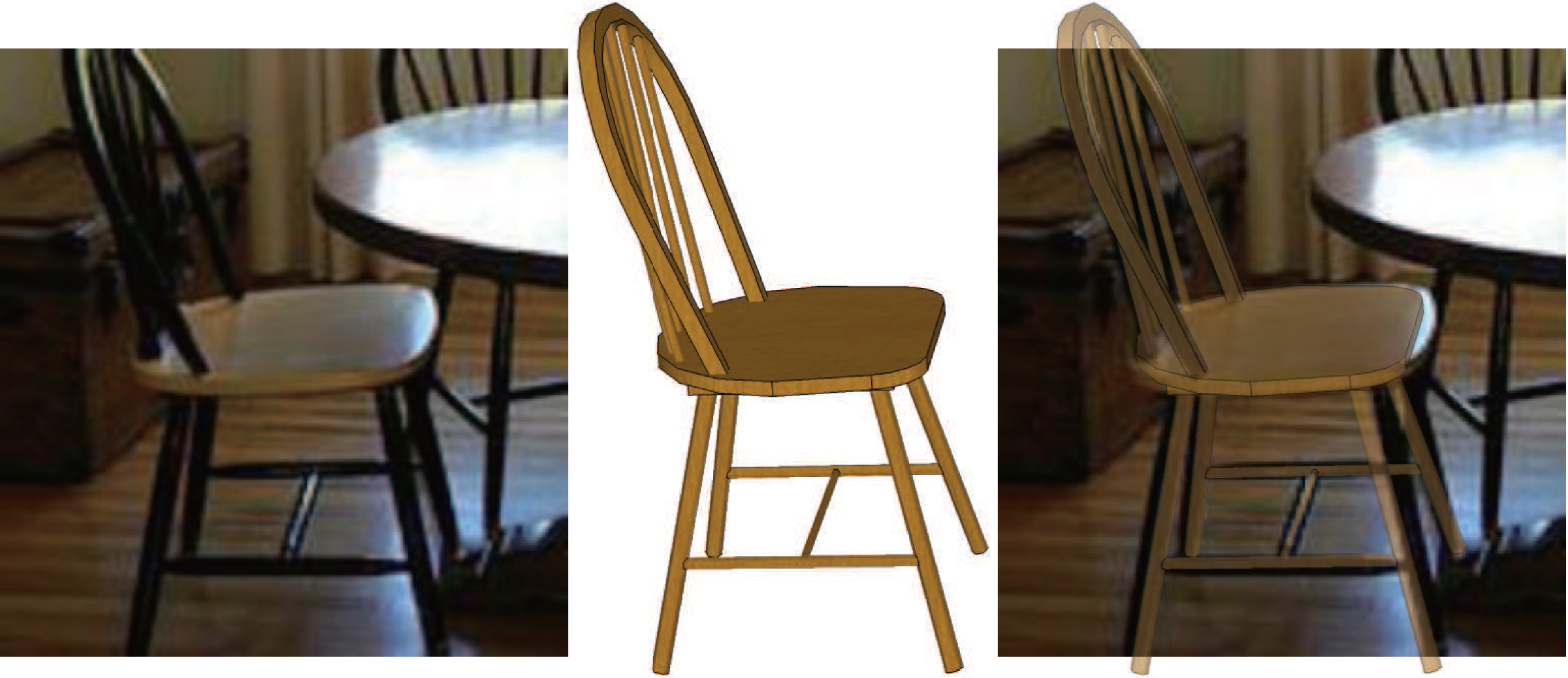


Image to shape retrieval

Key Steps towards 3D Reconstruction

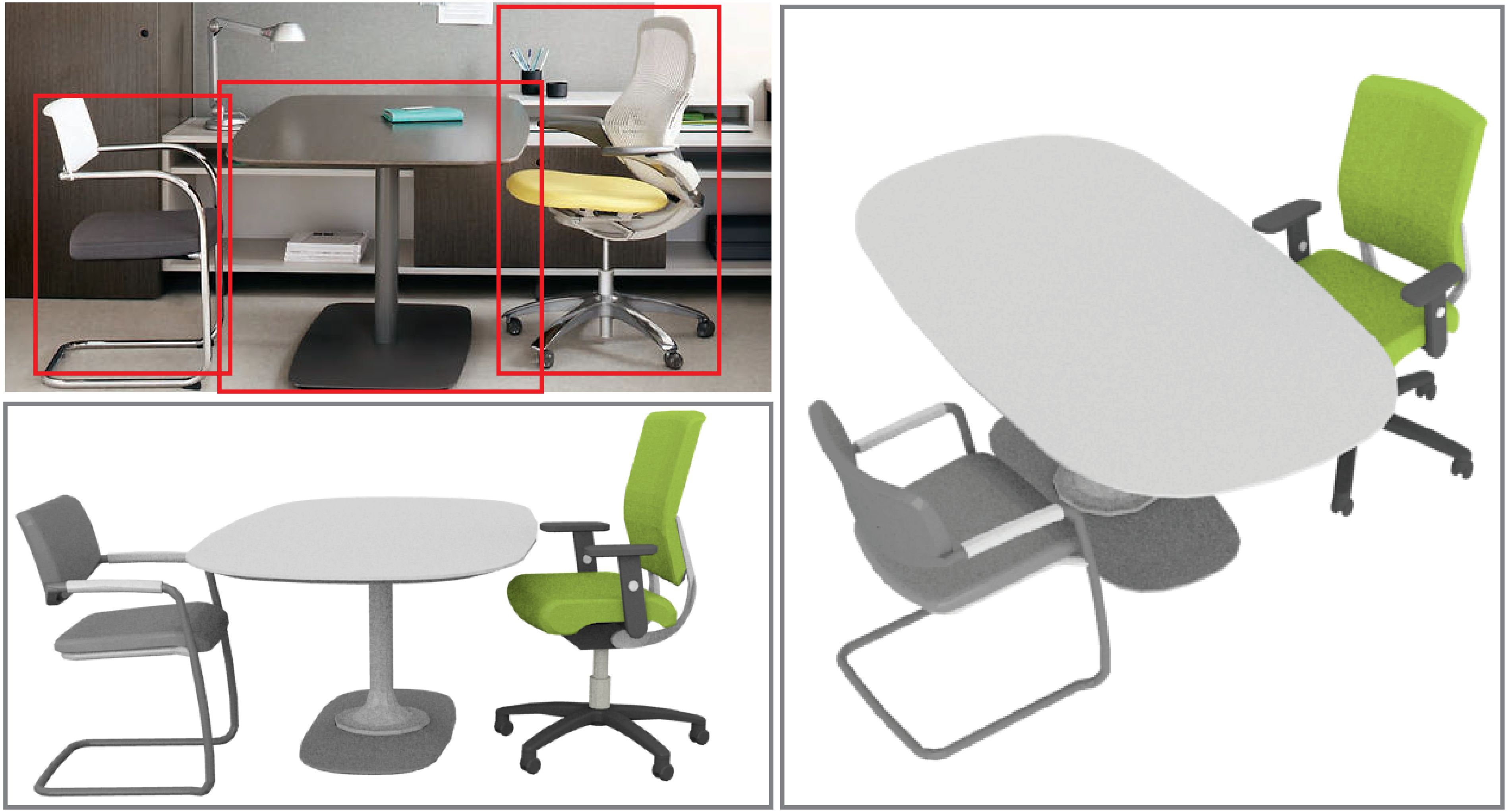


Similar Shape Retrieval

+

Viewpoint estimation

Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views, ICCV 2015 Oral

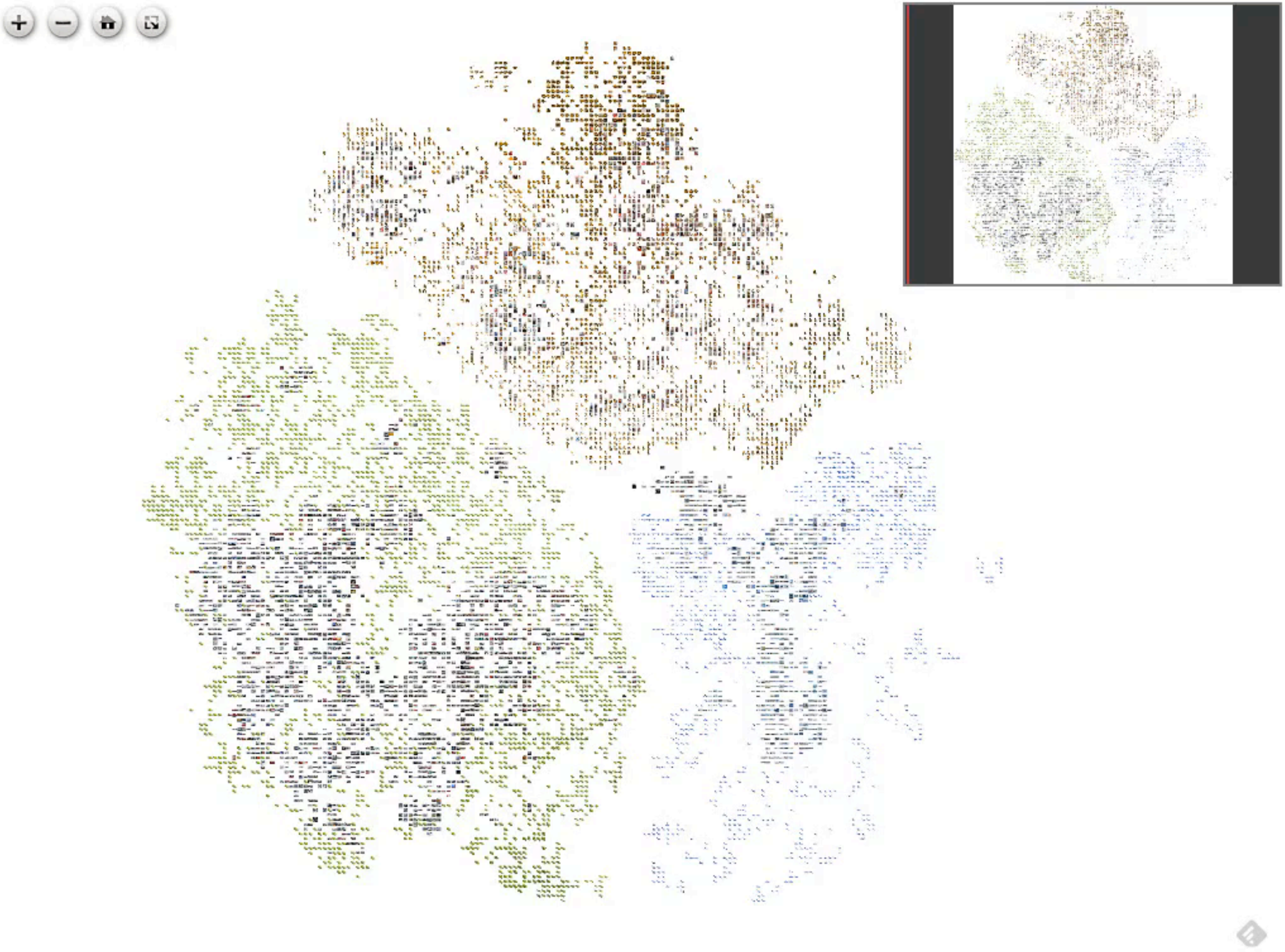


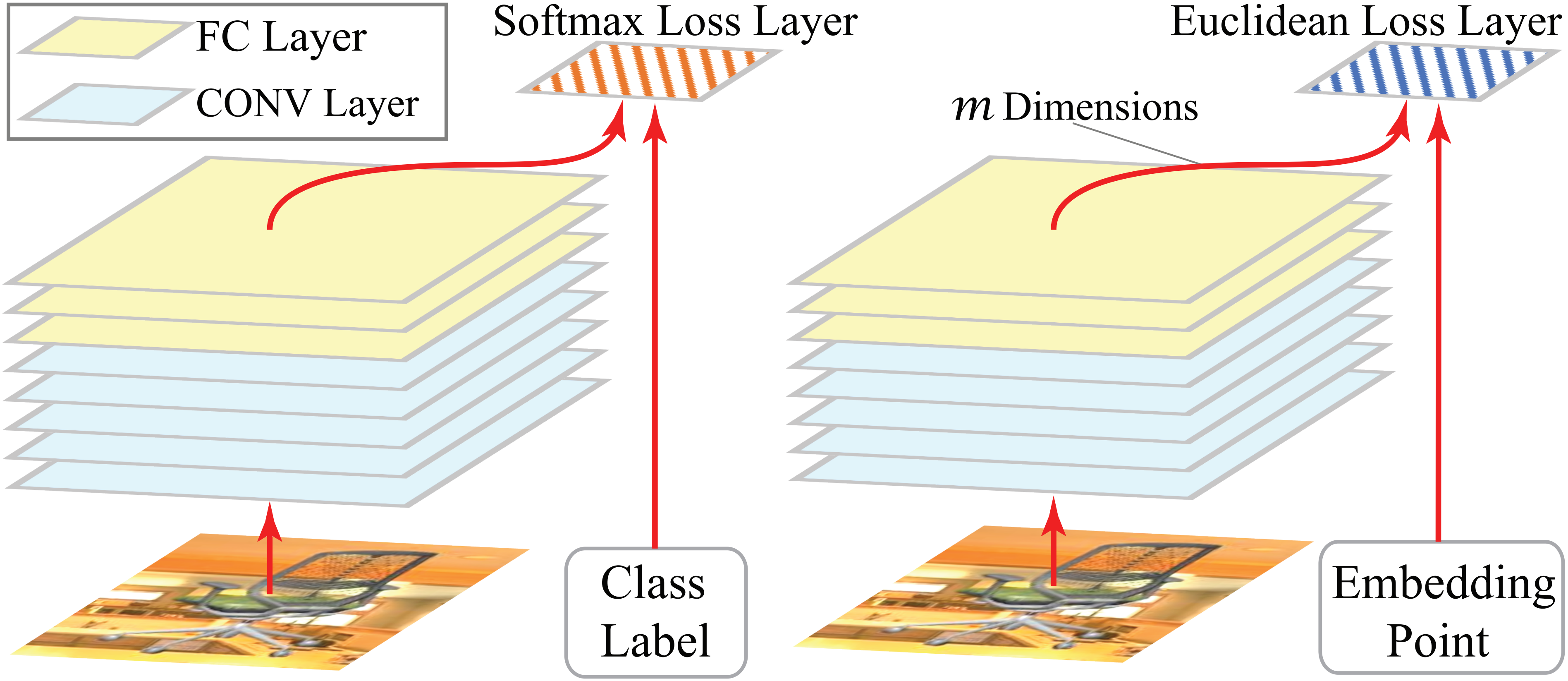
Stay Cool with http://shapenet.github.io/JointEmbedding/
http://shapenet.github.io/RenderForCNN/

Take Home Messages

- Train with synthetic, and act on real
- Asymmetry between synthesis and learning
    - Analogy to encoding/decoding in cryptology
- Promising directions
    - Occlusion patterns
    - Contextual information (depth images)

Thank you!




FC Layer
CONV Layer
Softmax Loss Layer
Class Label
Euclidean Loss Layer
m Dimensions
Embedding Point