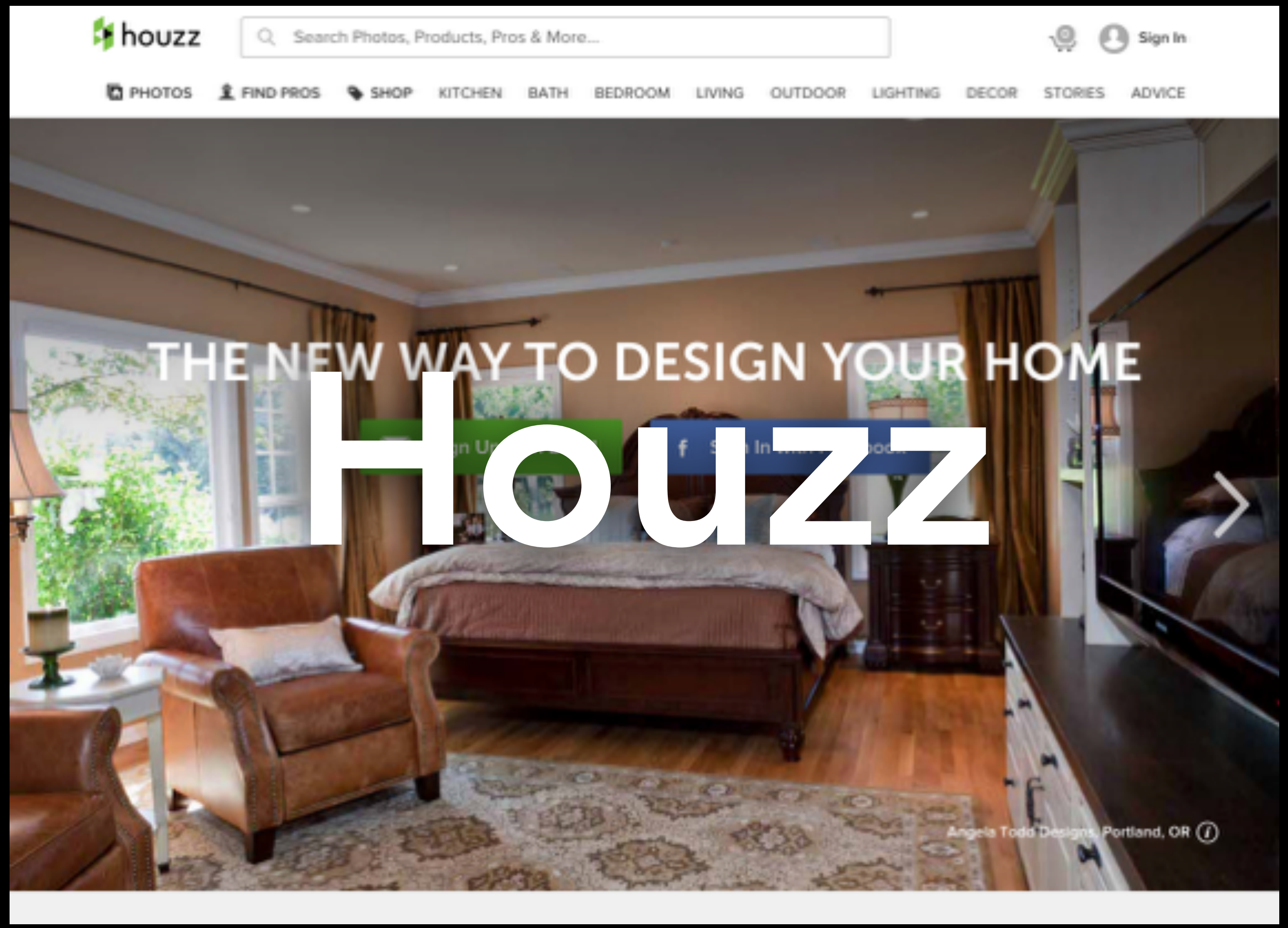
GROKSTYLE

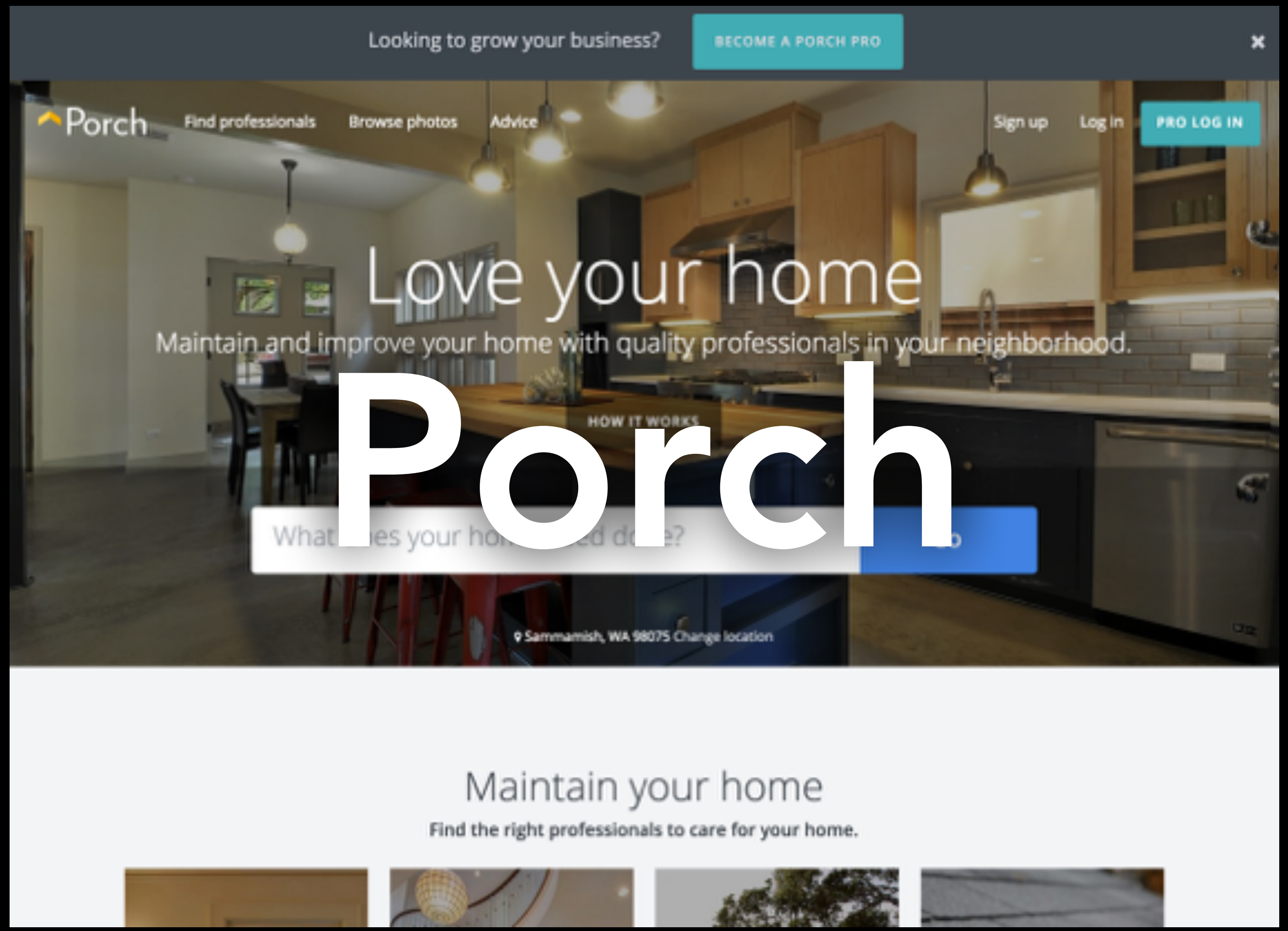
We're looking for interns!

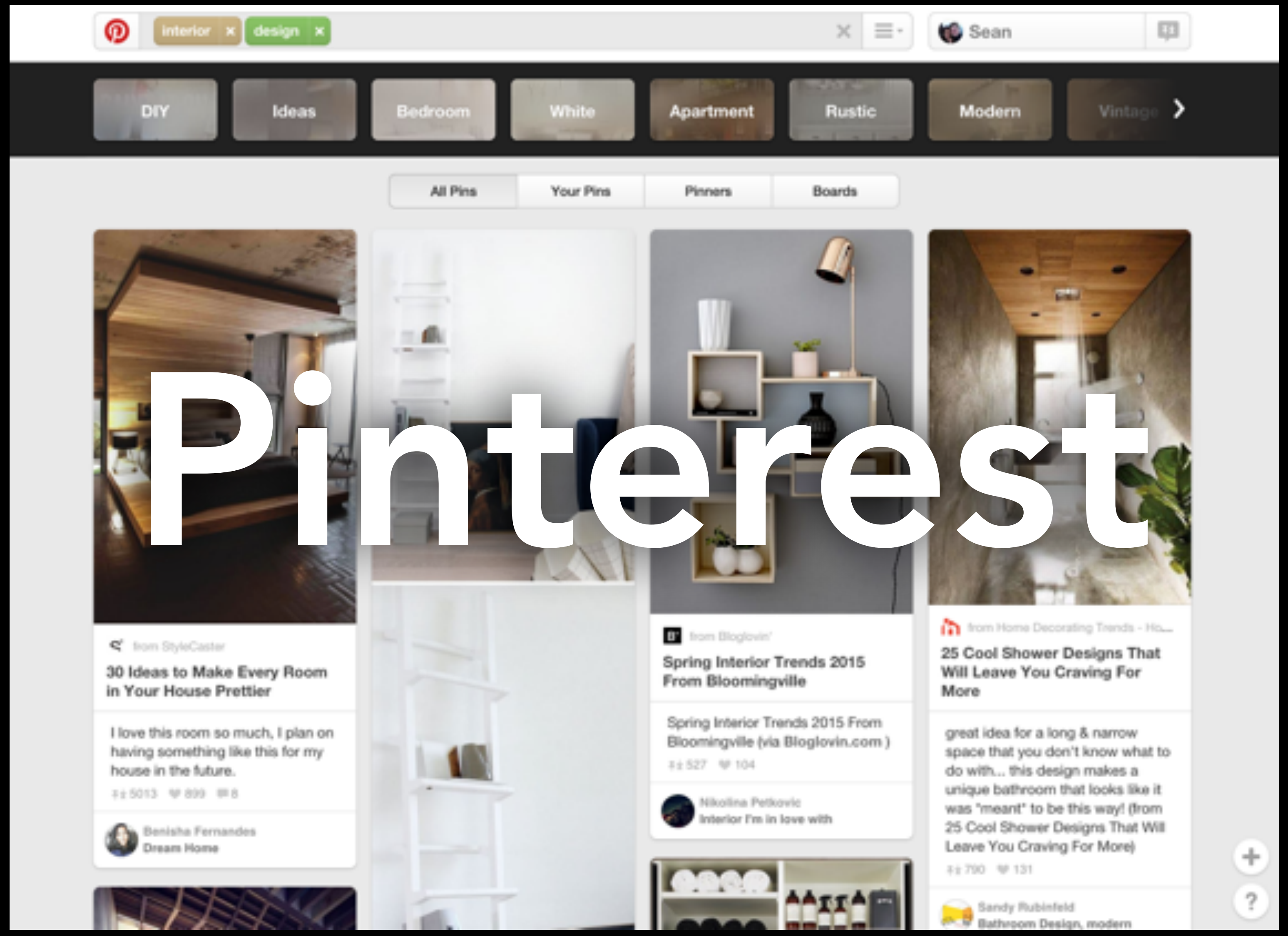
sbell@grokstyle.com

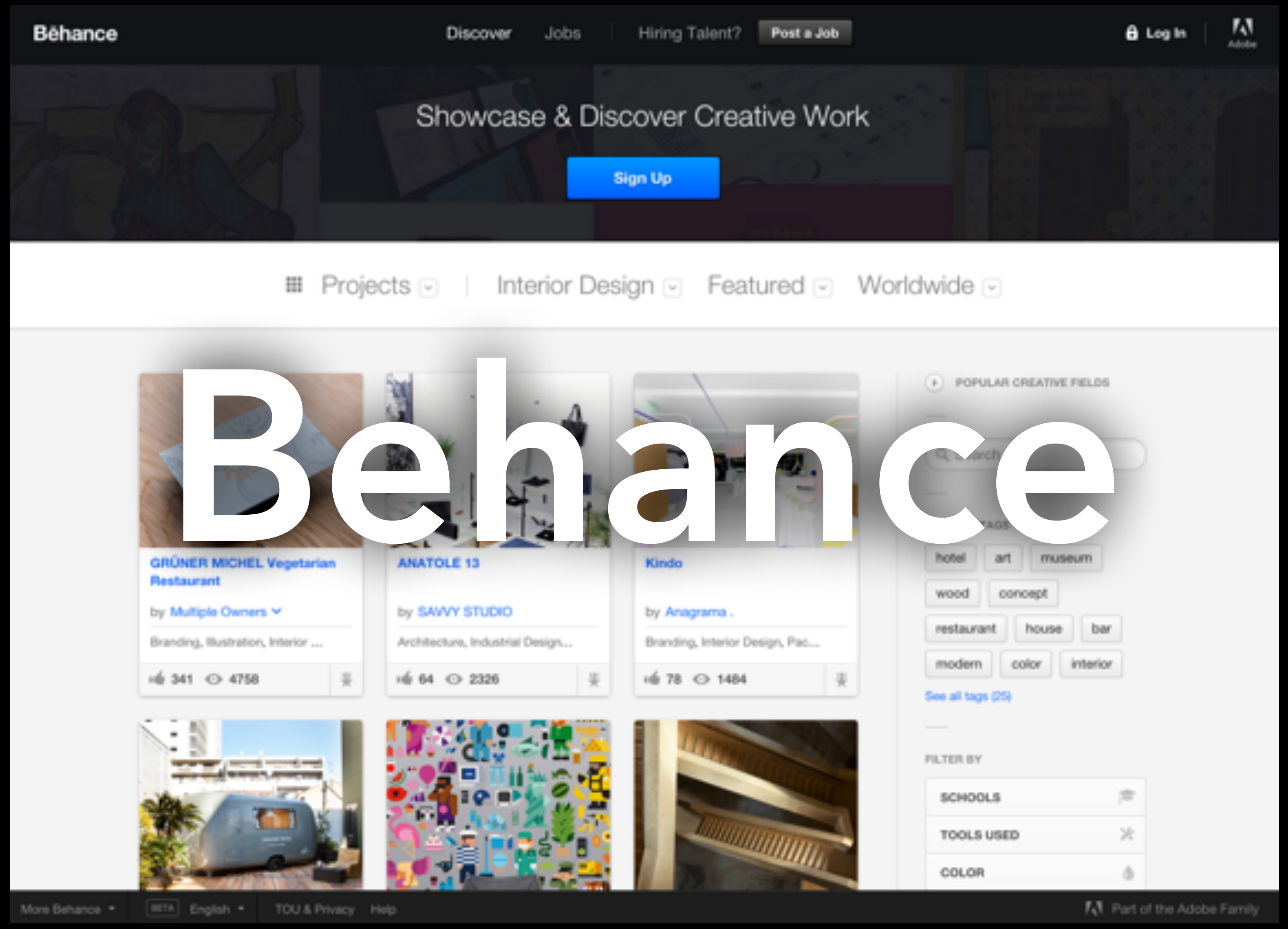
LEARNING VISUAL SIMILARITY FOR PRODUCT DESIGN WITH CONVOLUTIONAL NEURAL NETWORKS

SEAN BELL AND **KAVITA BALA**
CORNELL UNIVERSITY





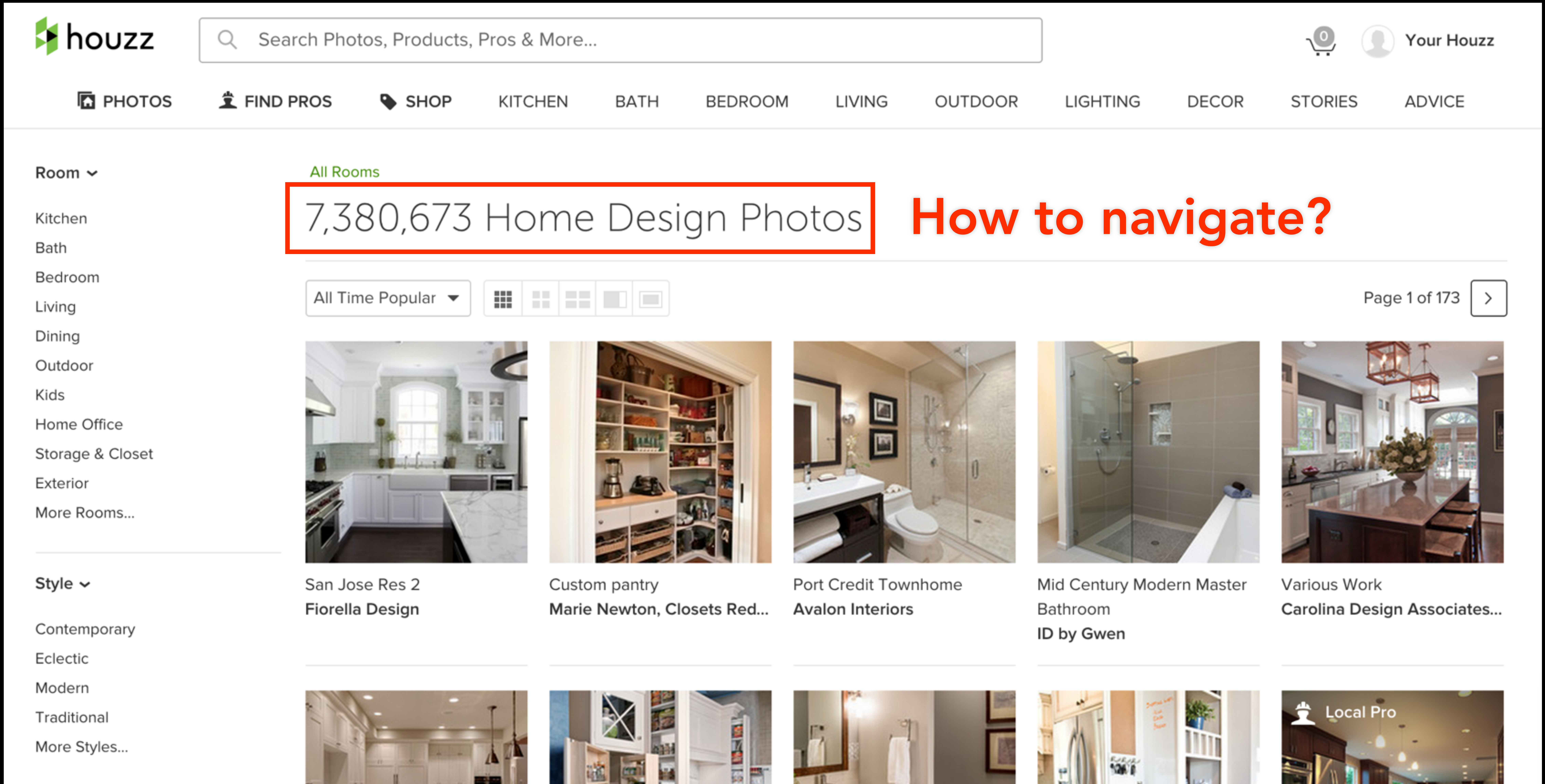




• • •

Millions of users

THE PROBLEM


houzz
Search Photos, Products, Pros & More...
0
Your Houzz
PHOTOS
FIND PROS
SHOP
KITCHEN
BATH
BEDROOM
LIVING
OUTDOOR
LIGHTING
DECOR
STORIES
ADVICE
Room
Kitchen
Bath
Bedroom
Living
Dining
Outdoor
Kids
Home Office
Storage & Closet
Exterior
More Rooms...
Style
Contemporary
Eclectic
Modern
Traditional
More Styles...
All Rooms
7,380,673 Home Design Photos
How to navigate?
All Time Popular
Page 1 of 173
San Jose Res 2
Fiorella Design
Custom pantry
Marie Newton, Closets Red...
Port Credit Townhome
Avalon Interiors
Mid Century Modern Master Bathroom
ID by Gwen
Various Work
Carolina Design Associates...
Local Pro

THE PROBLEM: "WHAT IS IT?"

Michelle Melwani_Mehra wrote:

What is the countertop? Thank you! ›

kammererk wrote:

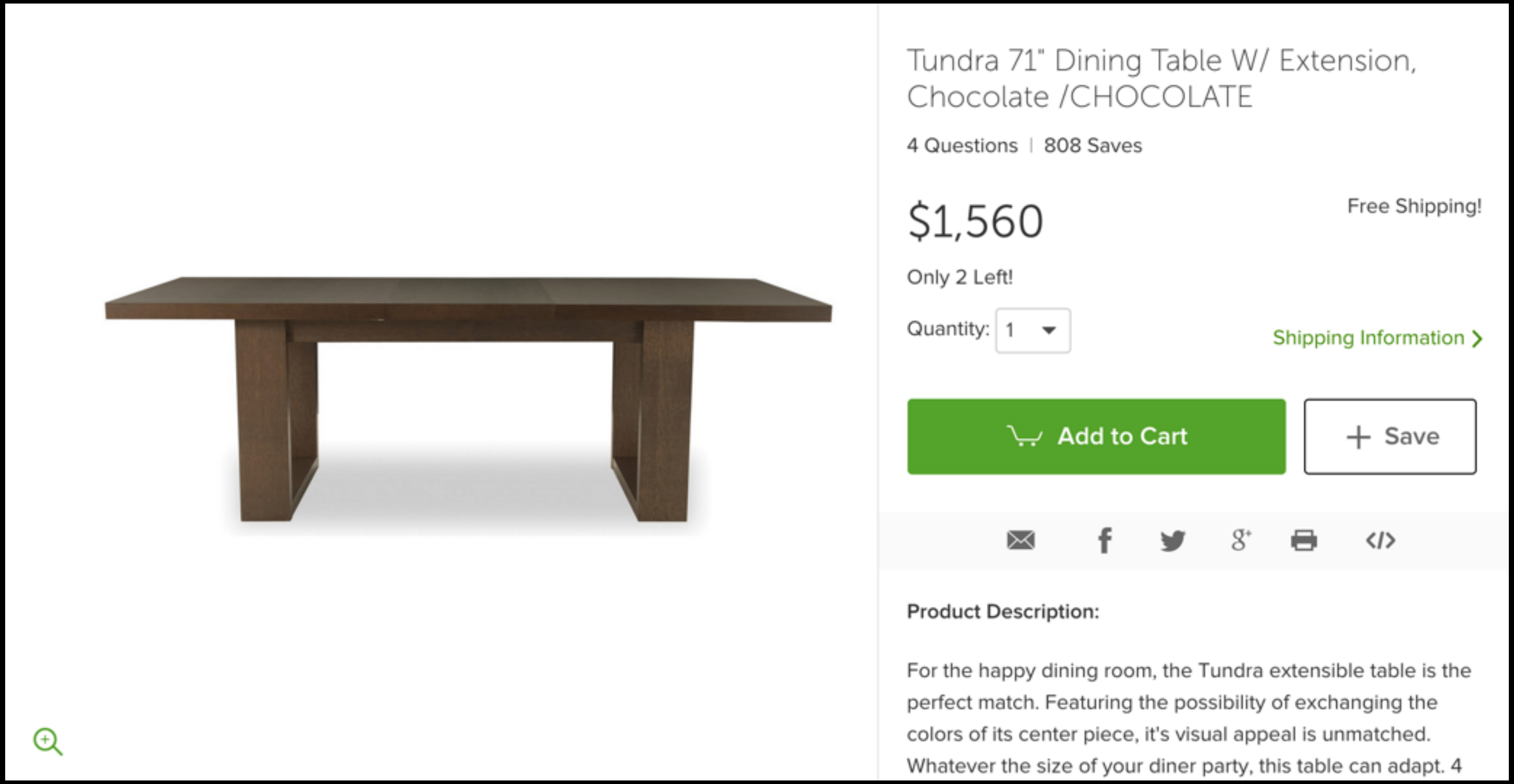
Can someone tell me the manufacturer/style of the range hood. Thanks. ›

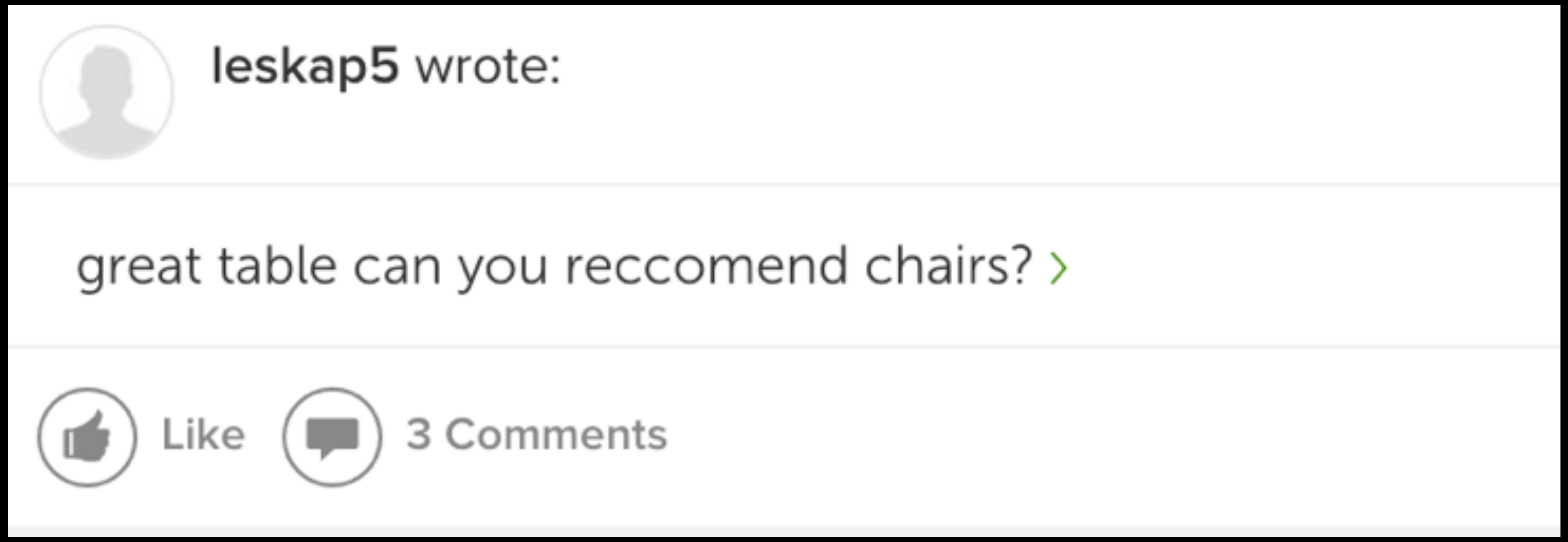
sjheff1 wrote:

What is the color of the subway tile you used in this kitchen? Thank You. ›

Can we answer this automatically?

THE PROBLEM: "WHERE IS IT USED?"





THE PROBLEM

(1) "What is this?"

(2) "Where is it used?"

Name: "Great Bowl O'Fire Sculptural Fire Bowl"

Category: Fire pit

Sold by: John T. Unger, LLC

THE PROBLEM

(1) "What is this?"

(2) "Where is it used?"

Name: "Great Bowl O'Fire Sculptural Fire Bowl"

Category: Fire pit

Sold by: John T. Unger, LLC

Challenge: determine whether these are the same product

(different resolution, viewpoint, color, lighting, occlusions)

TWO KINDS OF IMAGES

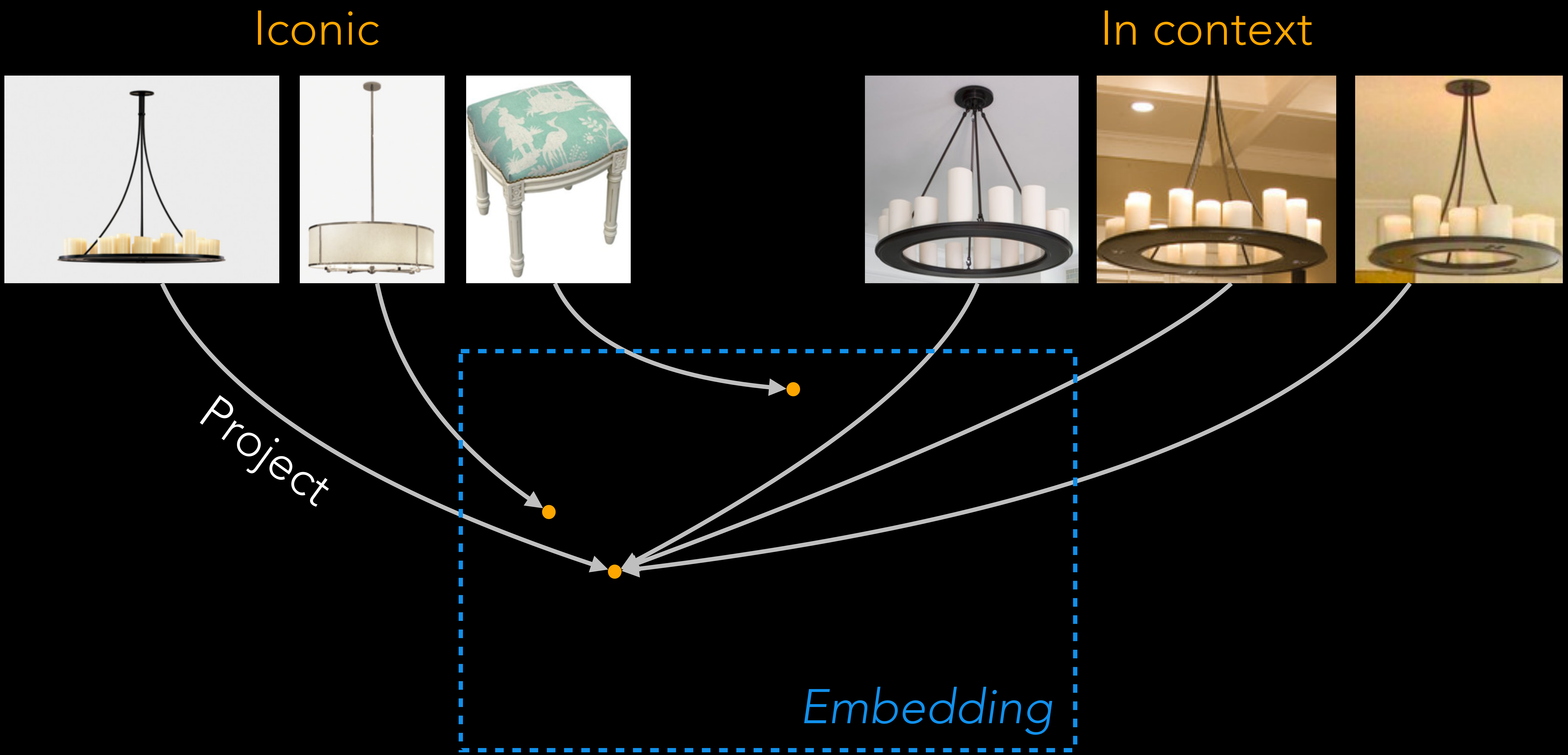
Iconic

(From a product website)

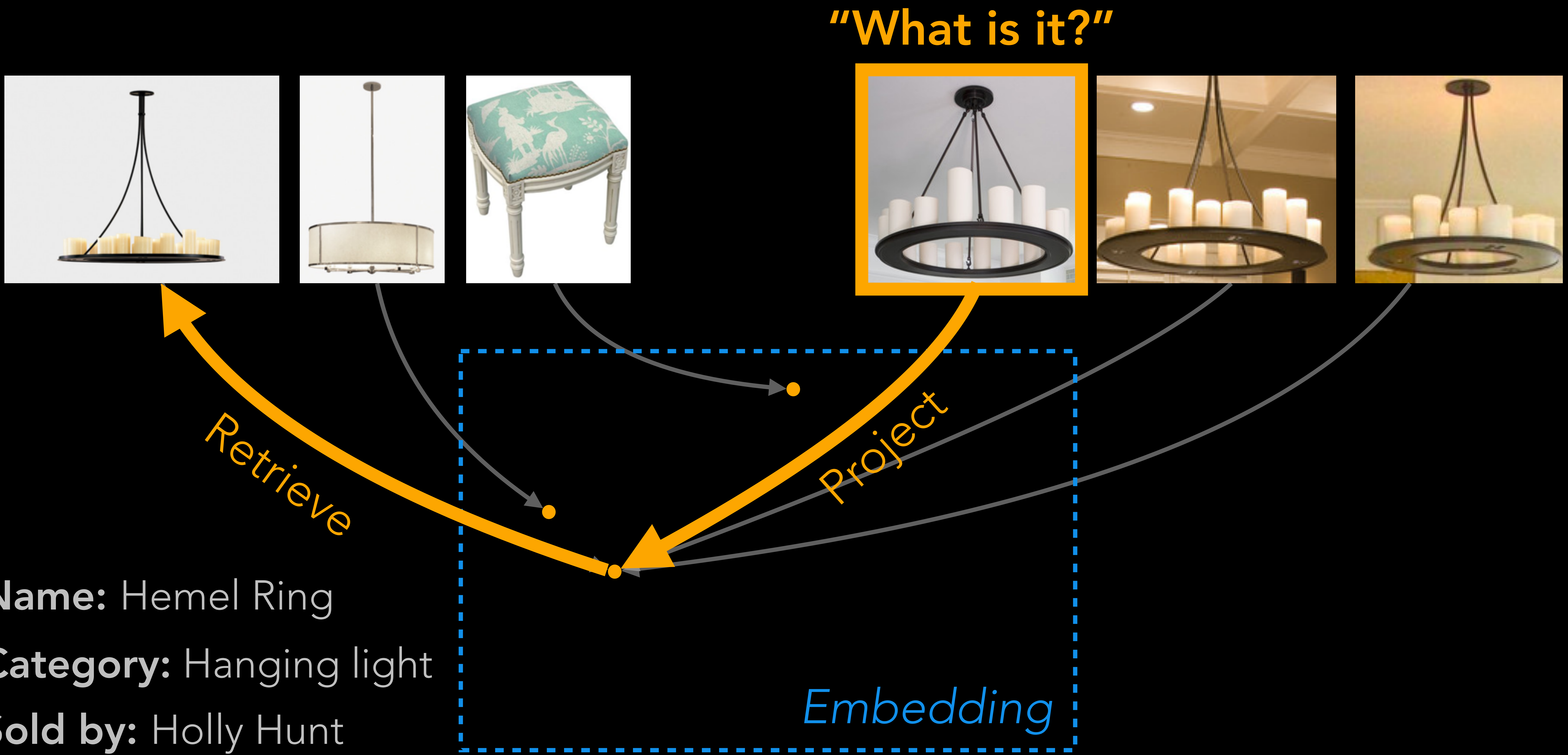
In context

(Cropped from a scene photo)

PROJECTING INTO A JOINT EMBEDDING



SEARCH USING THE EMBEDDING



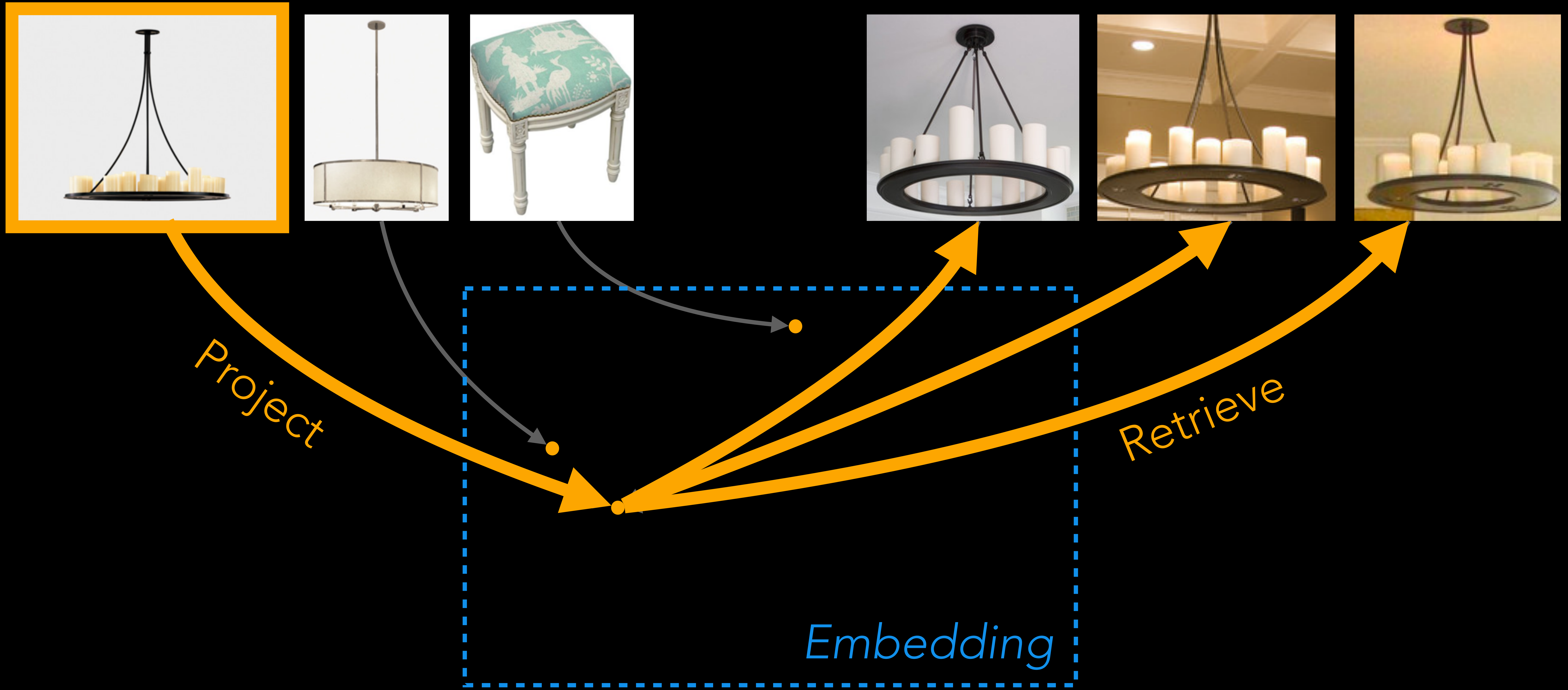
Name: Hemel Ring

Category: Hanging light

Sold by: Holly Hunt

SEARCH USING THE EMBEDDING

"Where is it used?"



APPROACH 1: DIRECTLY SOLVING FOR EMBEDDING

Learn the positions of all images directly from triplet judgements
[Agarwal 2007, Tamuz 2011, Kulis 2012, van der Maaten 2012, …]

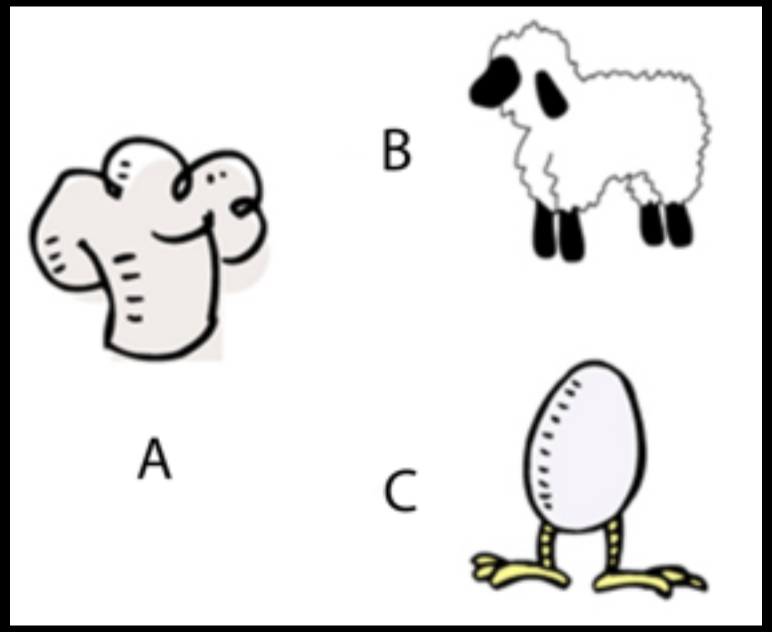
Need triplet judgements to project new items

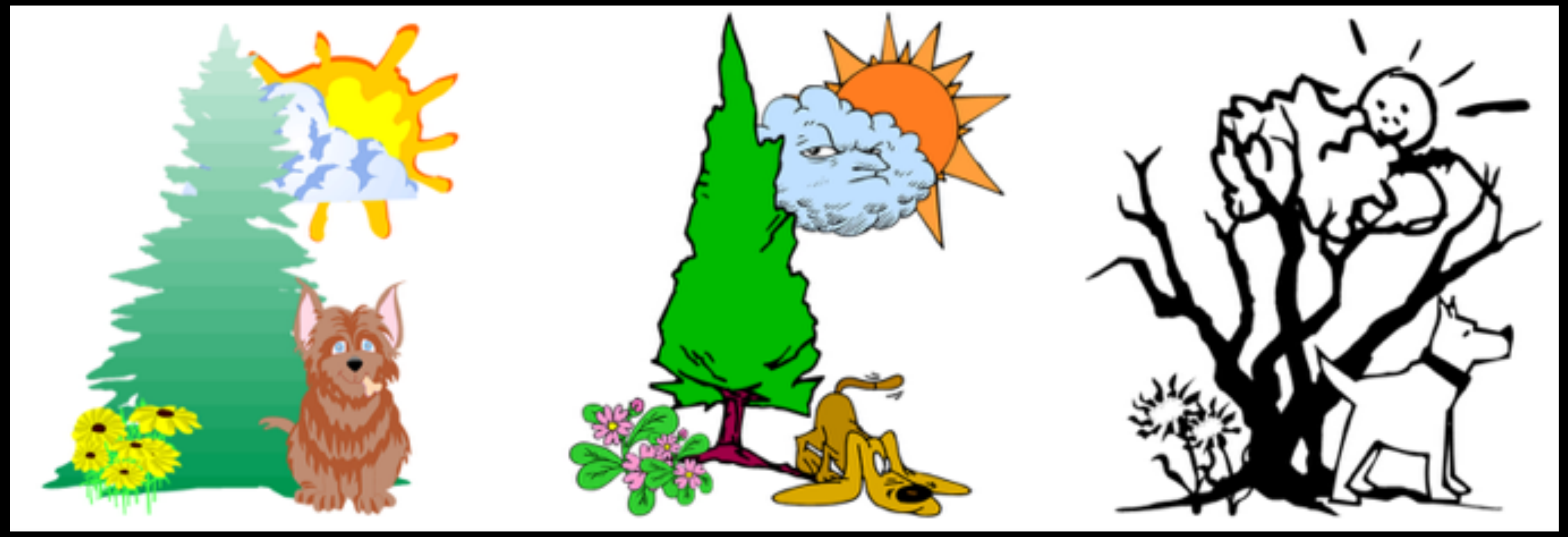
Generalized Nonmetric
Multidimensional Scaling
[Agarwal et al 2007]

APPROACH 2: METRIC LEARNING

Learn a distance function over features
[Chechik 2010, Garces 2014, O'Donovan 2014, ...]

Image → Extract Features (Designed, Fixed) → Project to Embedding (Learned) → Embedding position x

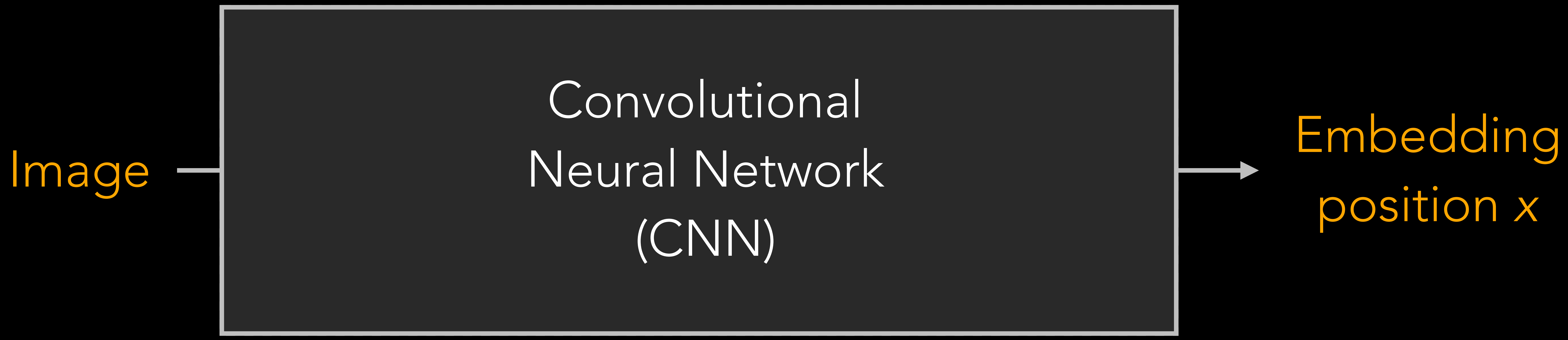




A Similarity Measure for Illustration Style [Garces 2014]

APPROACH 3: DEEP LEARNING (END-TO-END)

Learn end-to-end: a function from image to embedding
[Chopra 2005, Hadsell 2006, Wang 2014, …]



DEEP LEARNING HAS TAKEN OVER

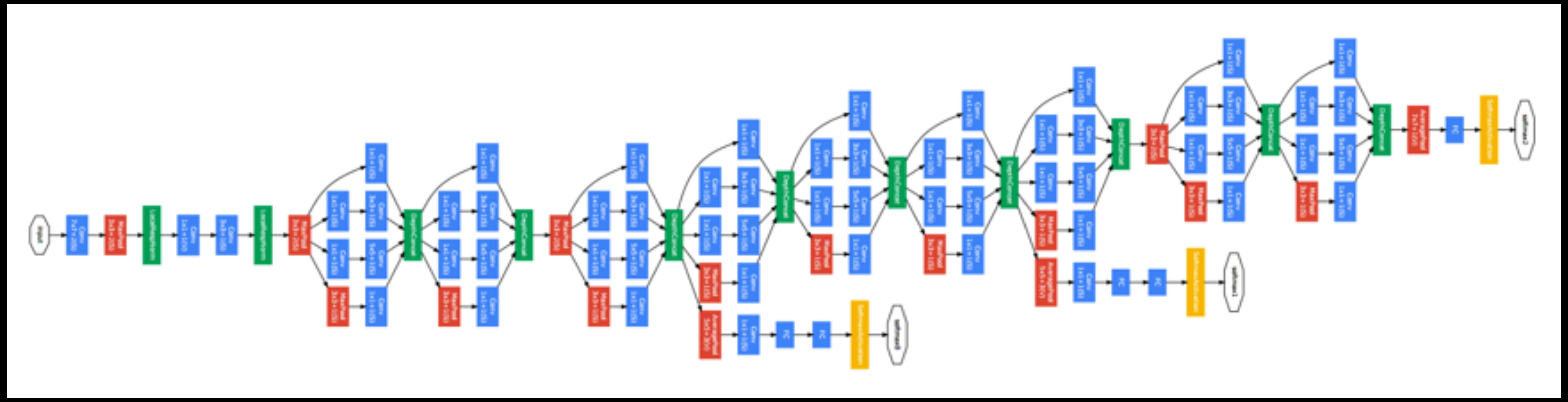
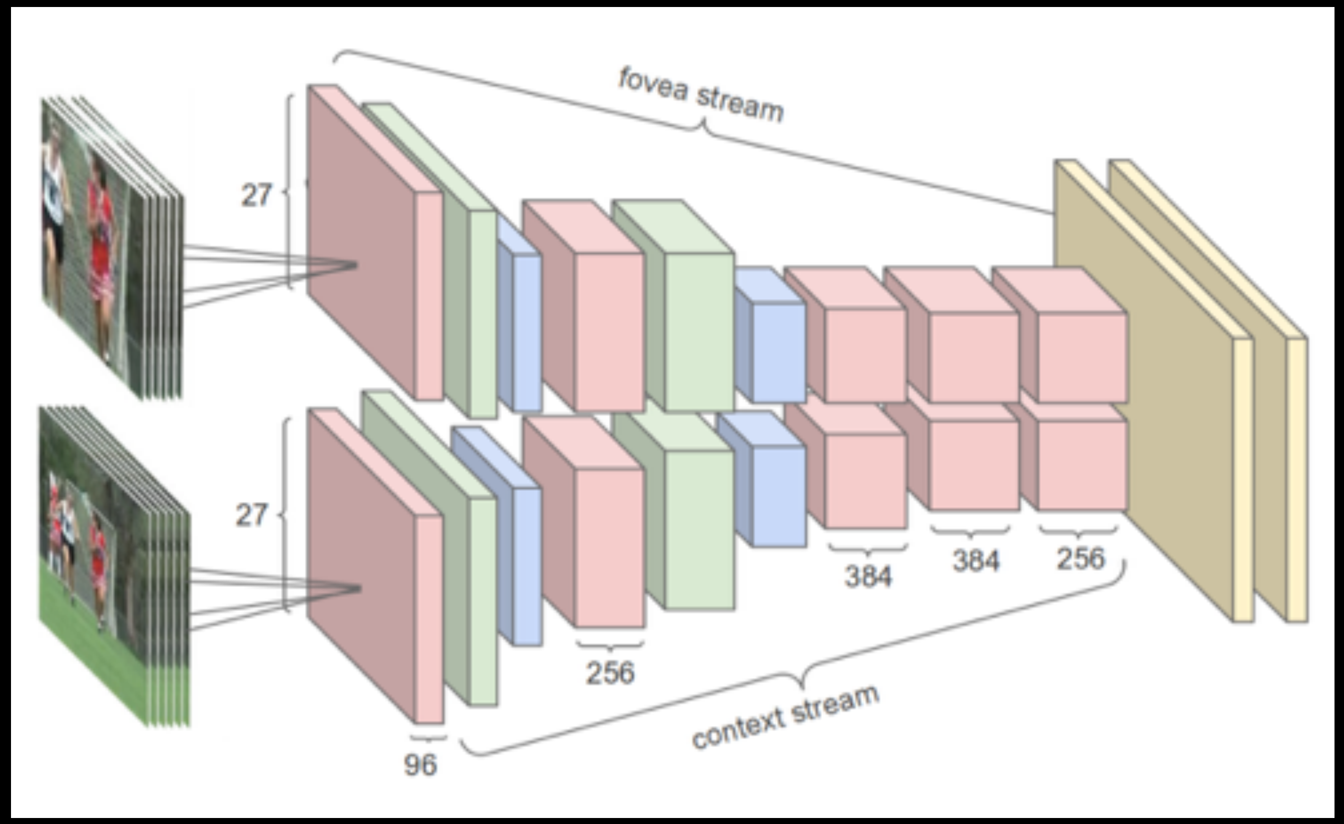
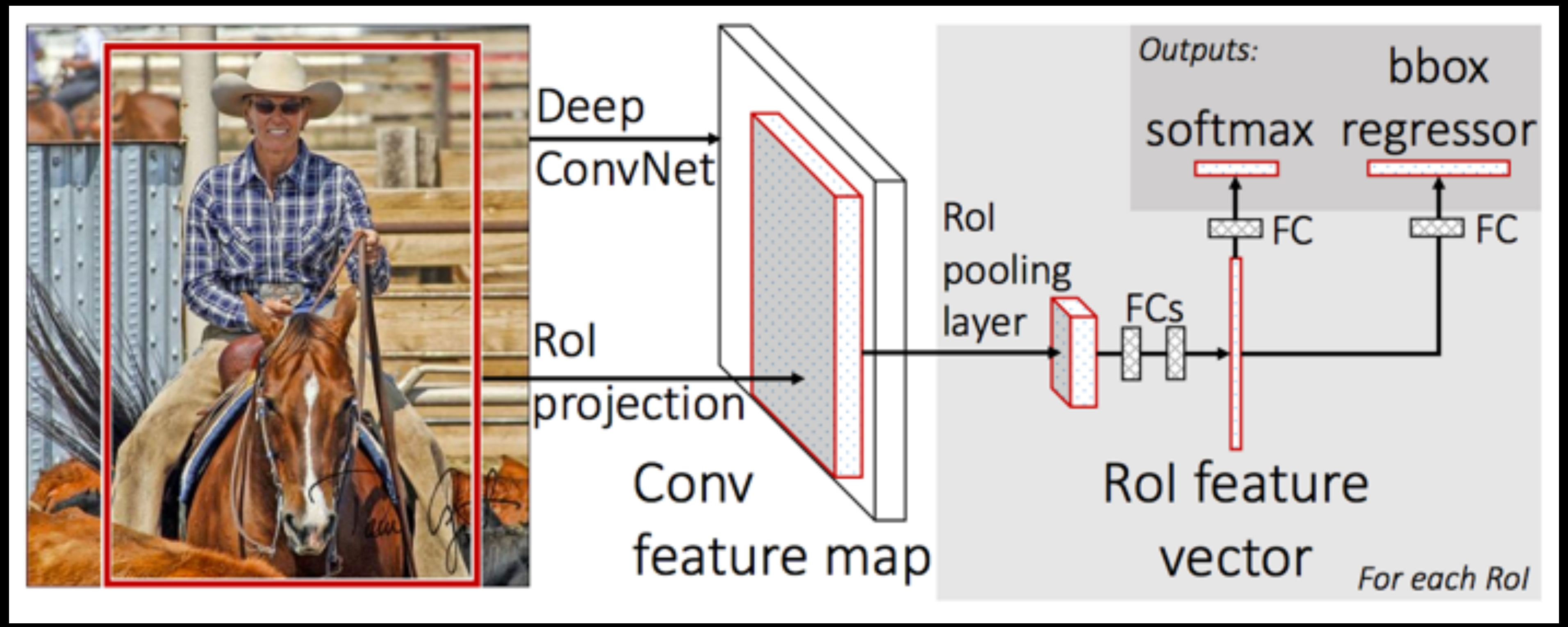


Image Classification [Szegedy 2014]



Video Classification [Karpathy 2014]



Object Detection [Girshick 2015]

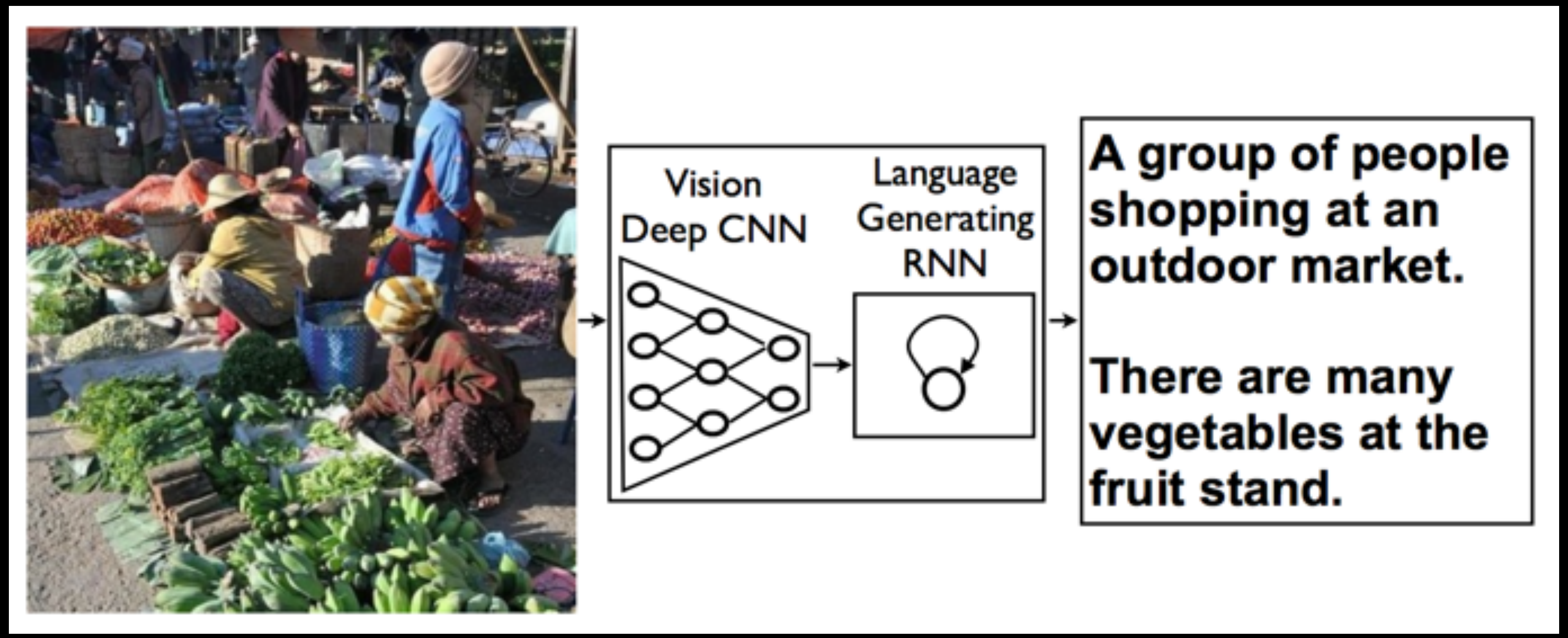
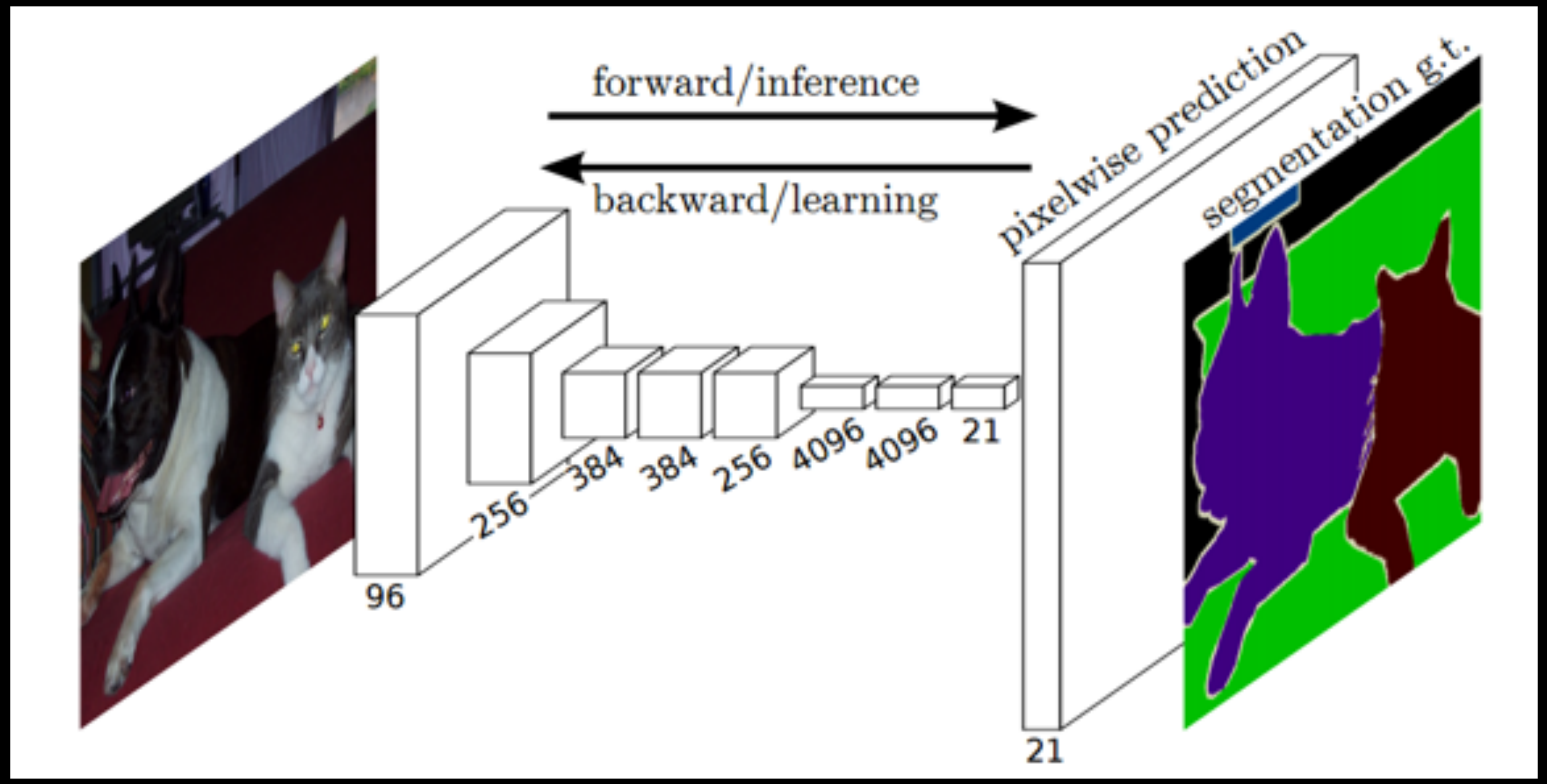
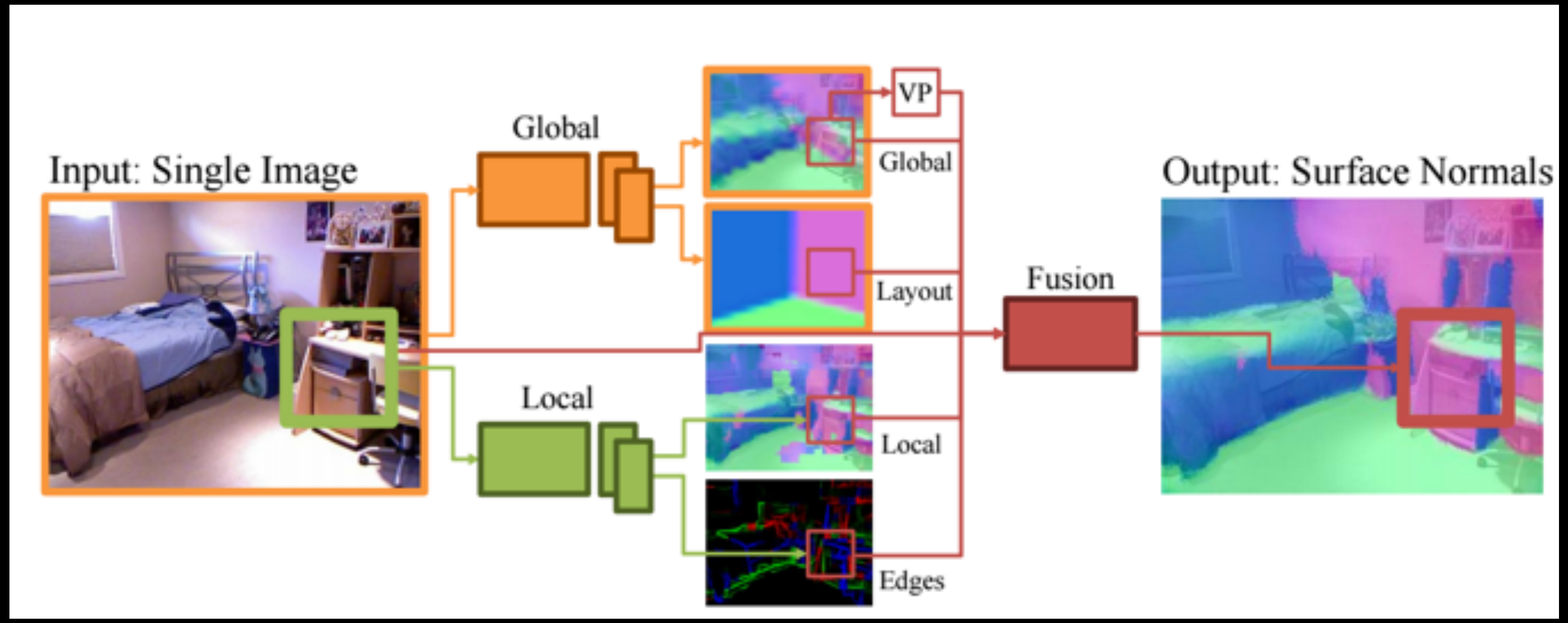


Image Captioning [Vinyals 2015]

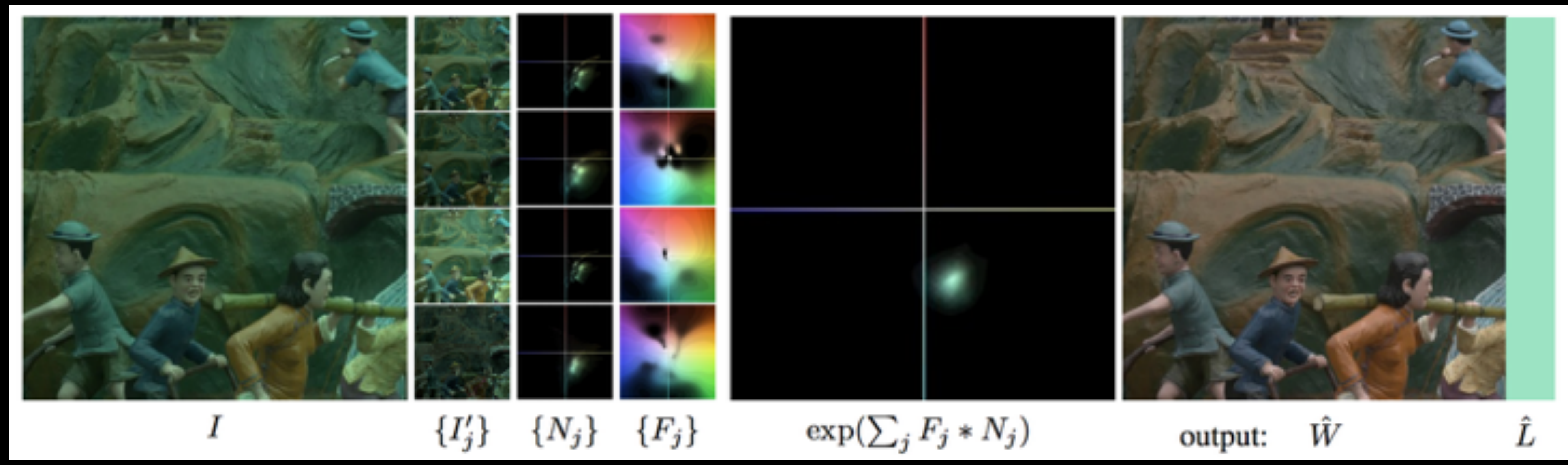
DEEP LEARNING HAS TAKEN OVER



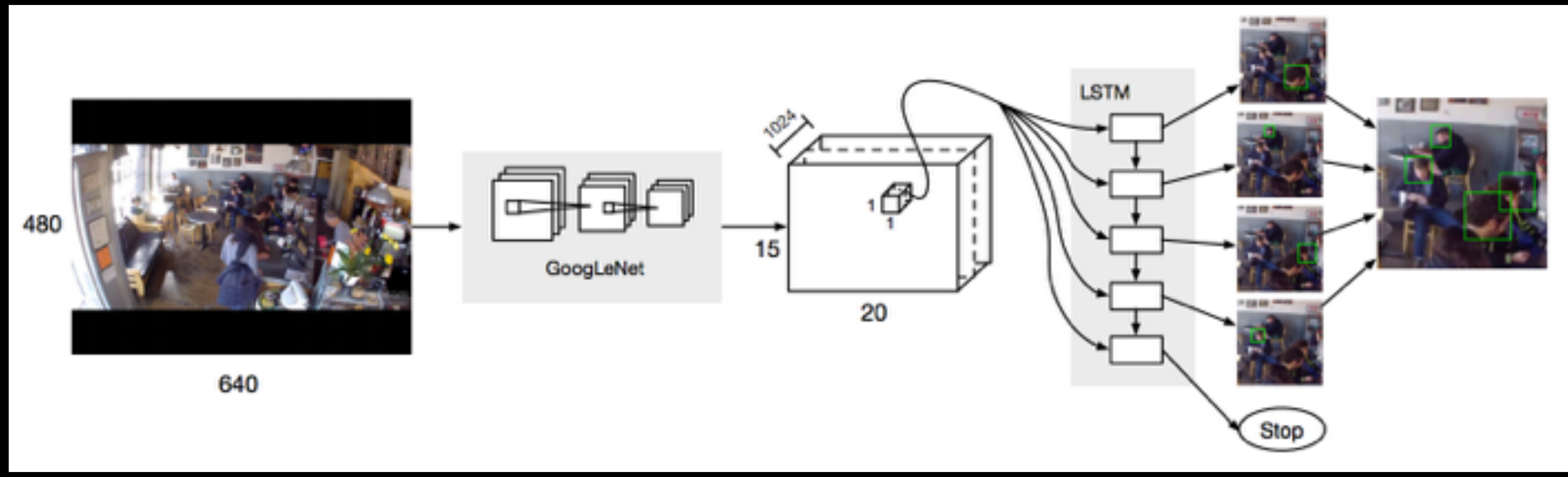
Semantic Segmentation [Long 2014]



Surface Normals [Wang 2014]

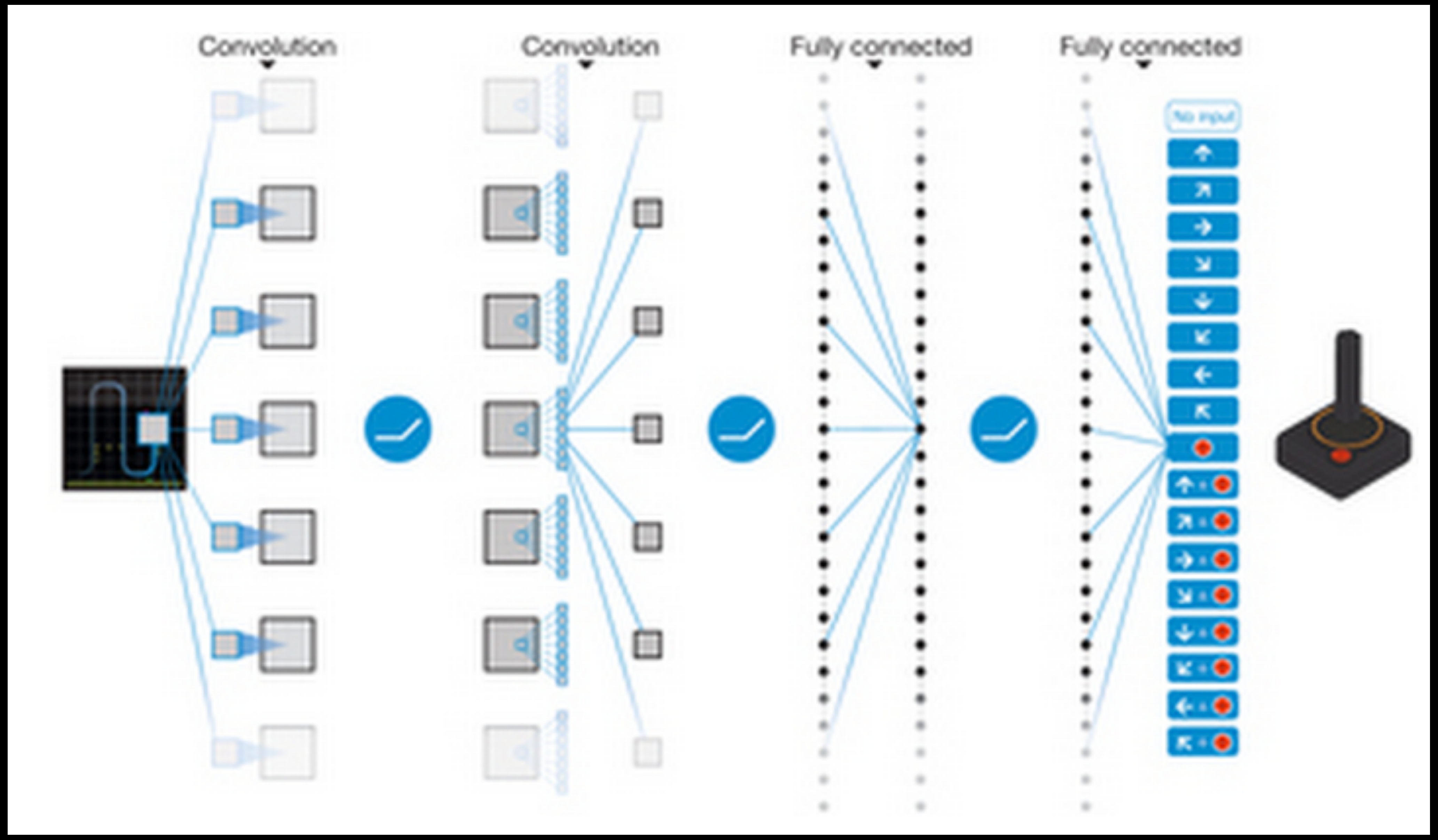


Color Constancy [Barron 2015]



People Counting [Stewart 2015]

DEEP LEARNING HAS TAKEN OVER

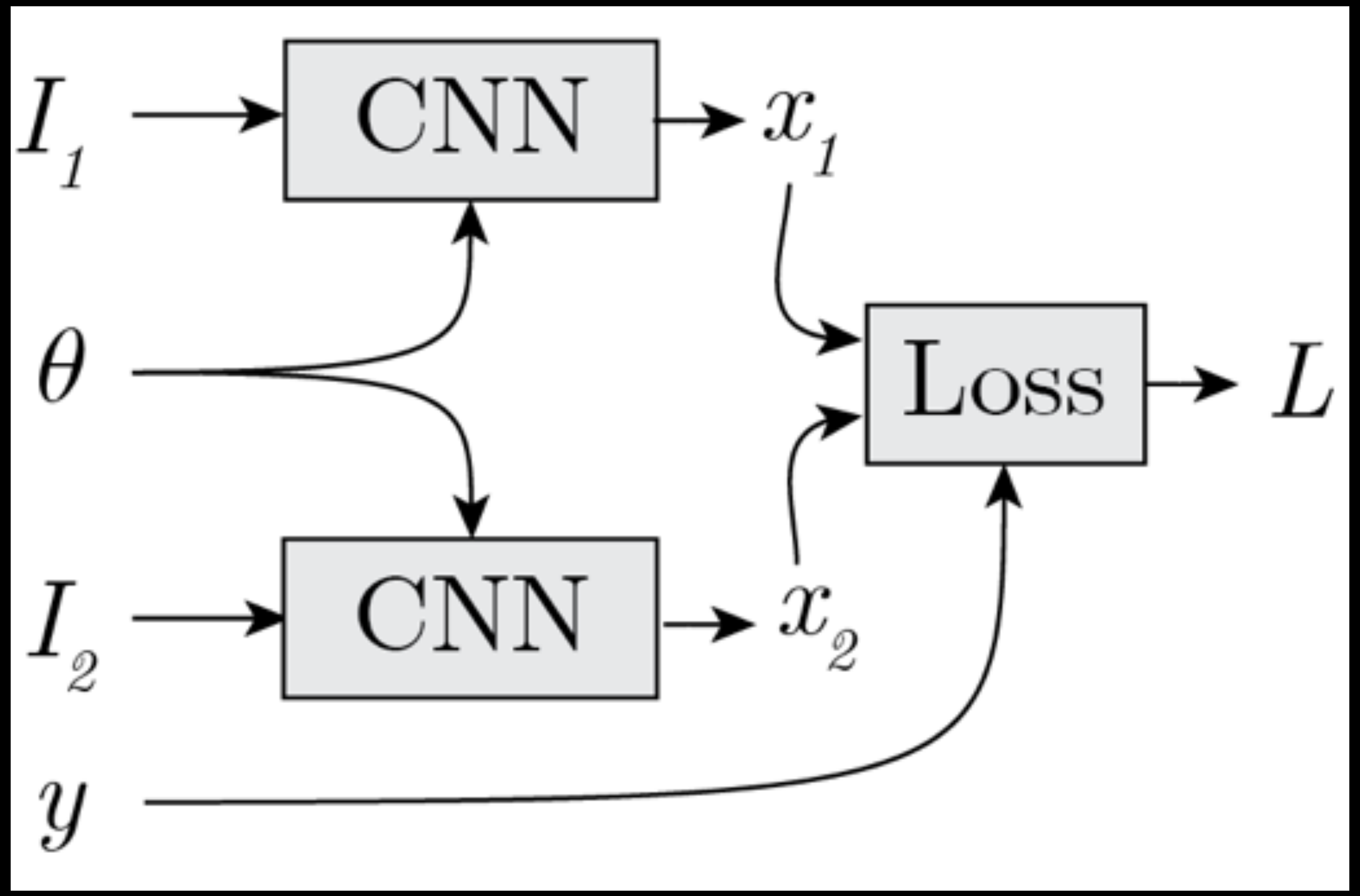


Playing video games [Minh 2015]

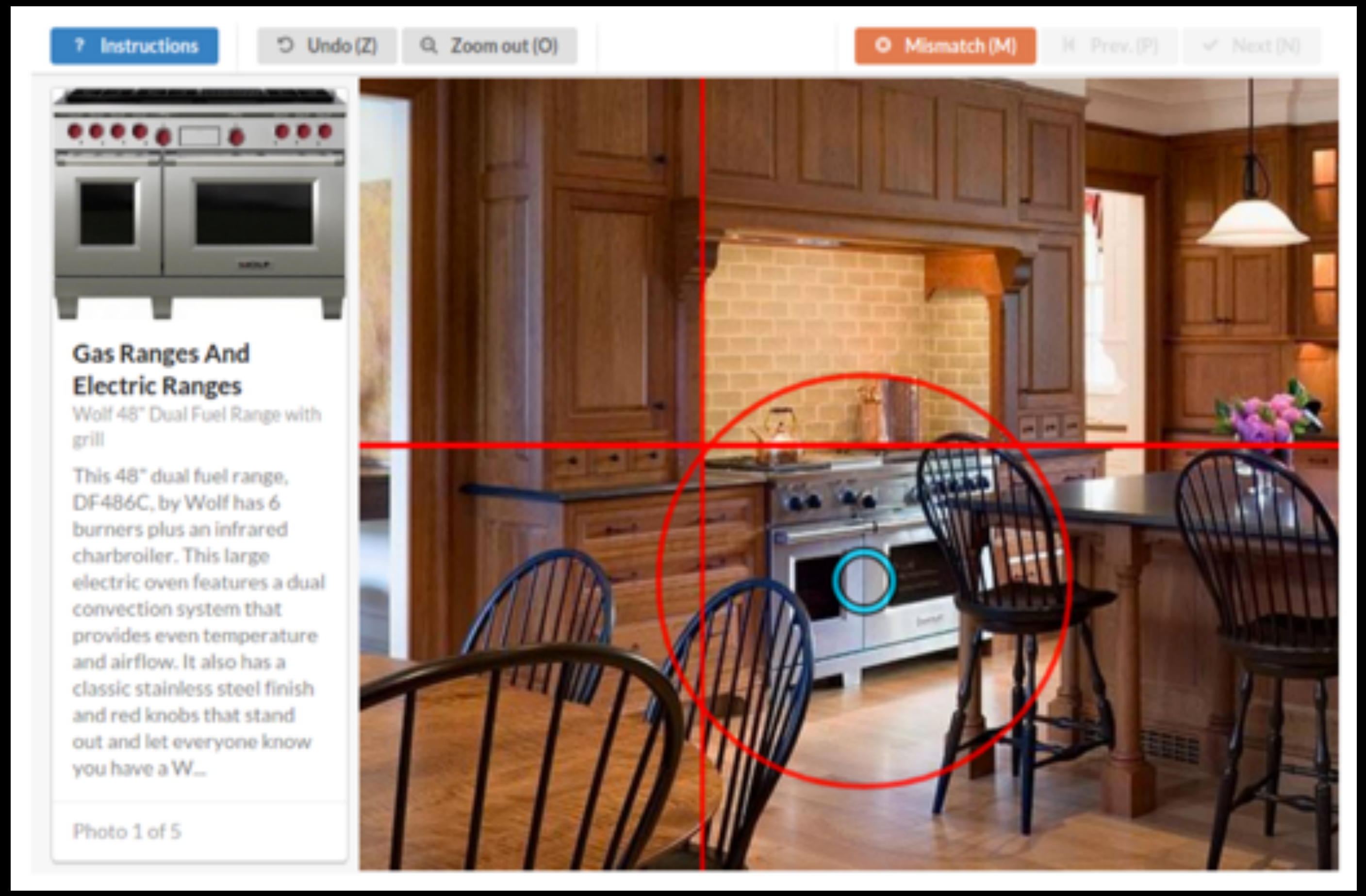
Generating house numbers [Gregor 2015]

…

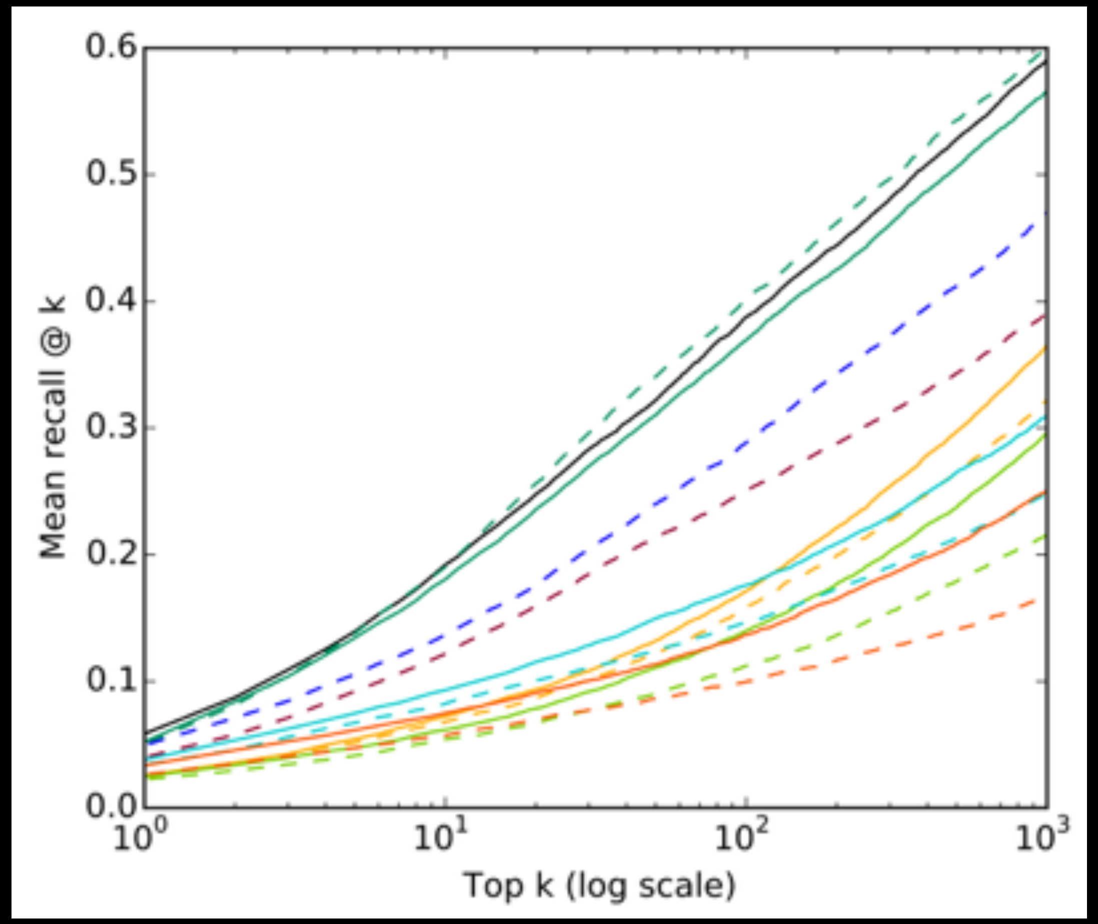
OVERVIEW



(1) Training the CNN

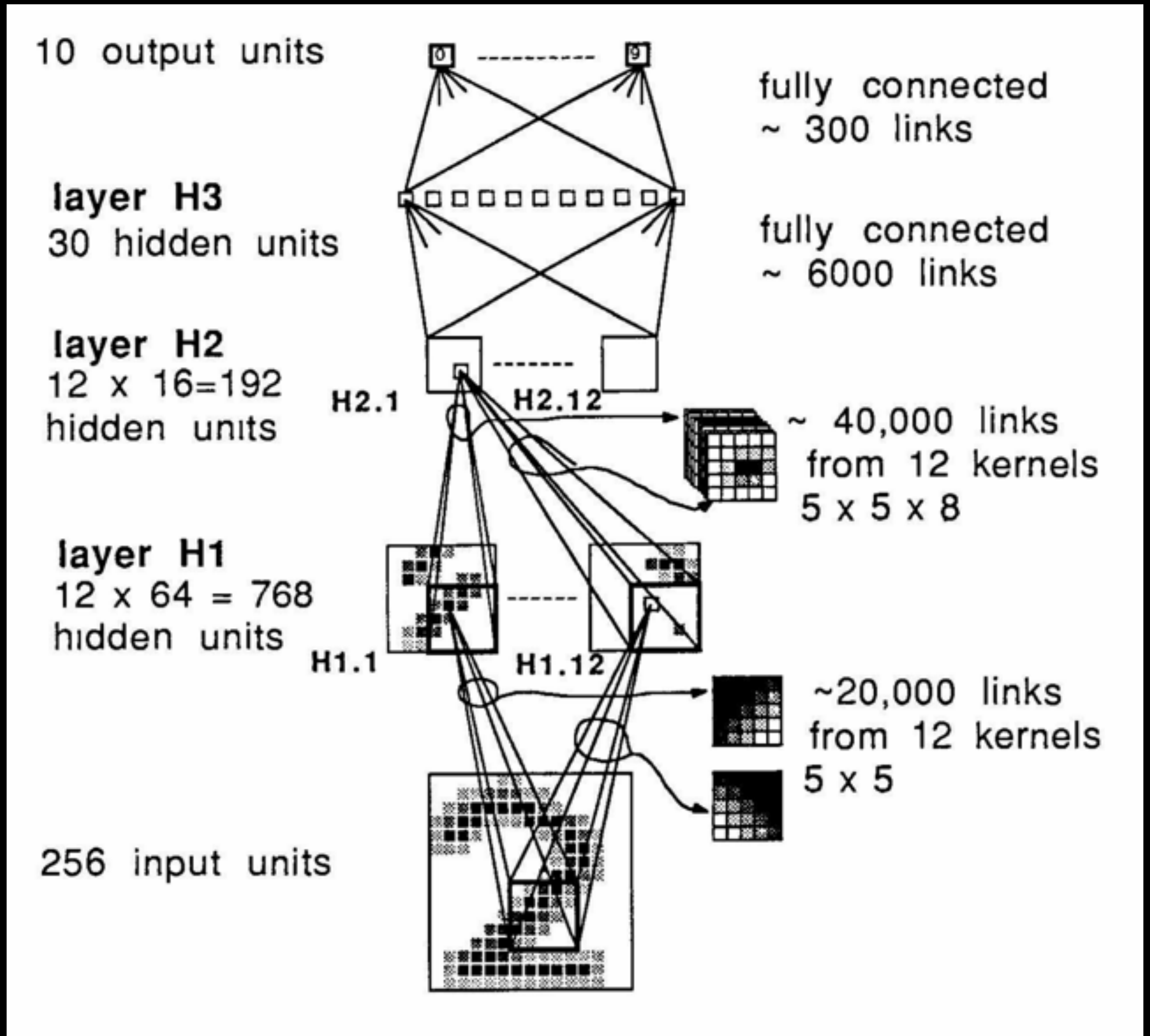


(2) Gathering training data



(3) Search Results

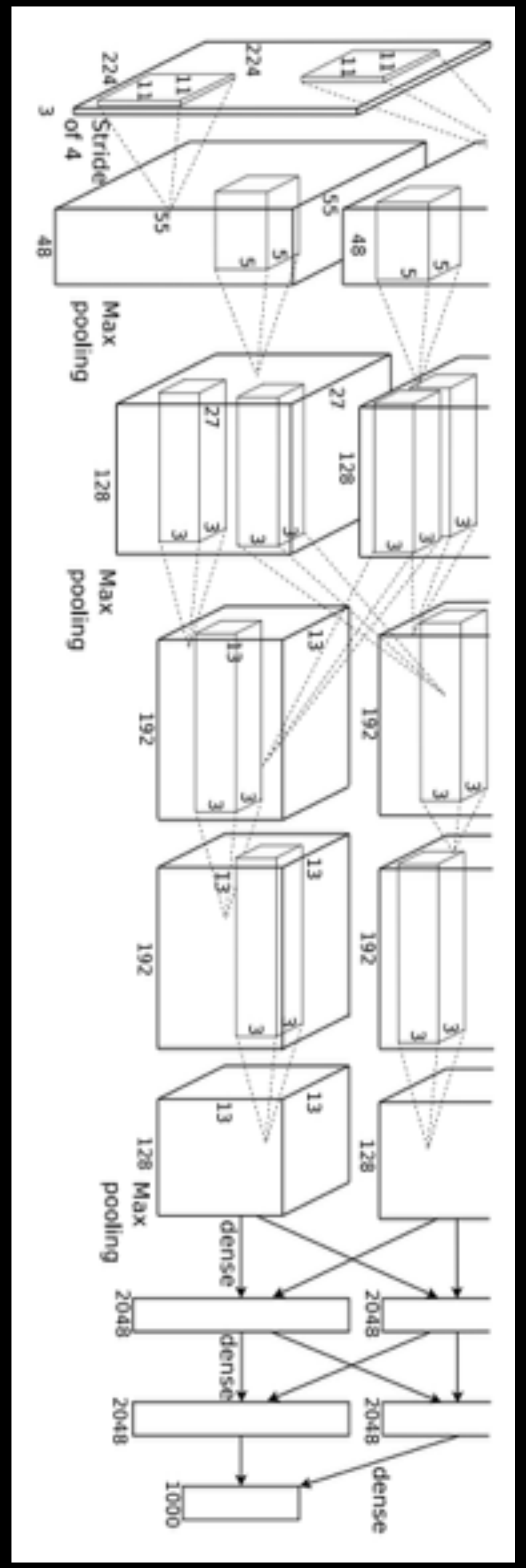
CONVOLUTIONAL NEURAL NETWORKS (CNN)



"LeNet"

Backpropagation applied to handwritten zip code recognition

[Lecun **1989**]



"AlexNet"

ImageNet Classification with Deep Convolutional Neural Networks

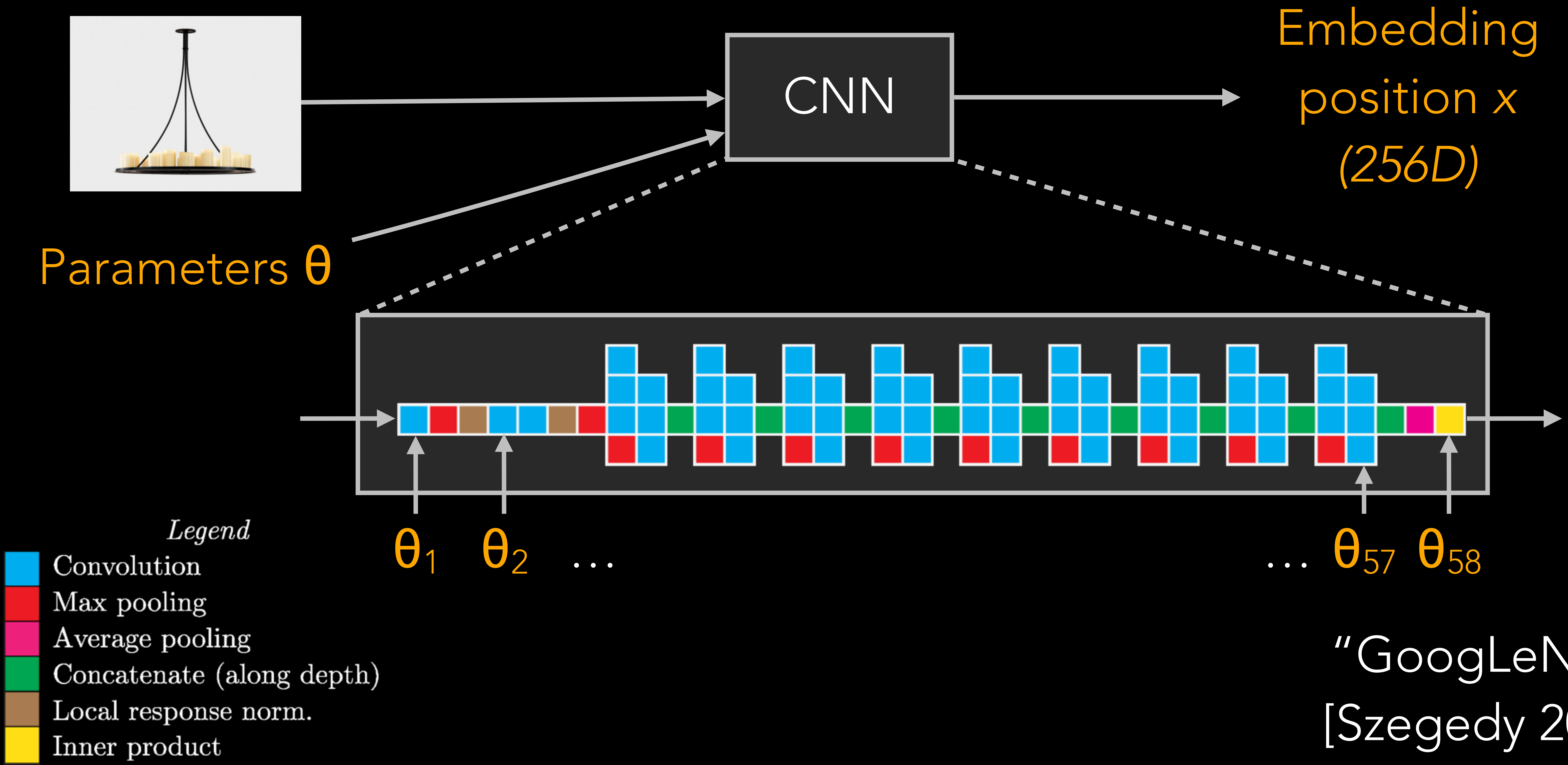
[Krizhevsky **2012**]

"GoogLeNet"

Going deeper with convolutions

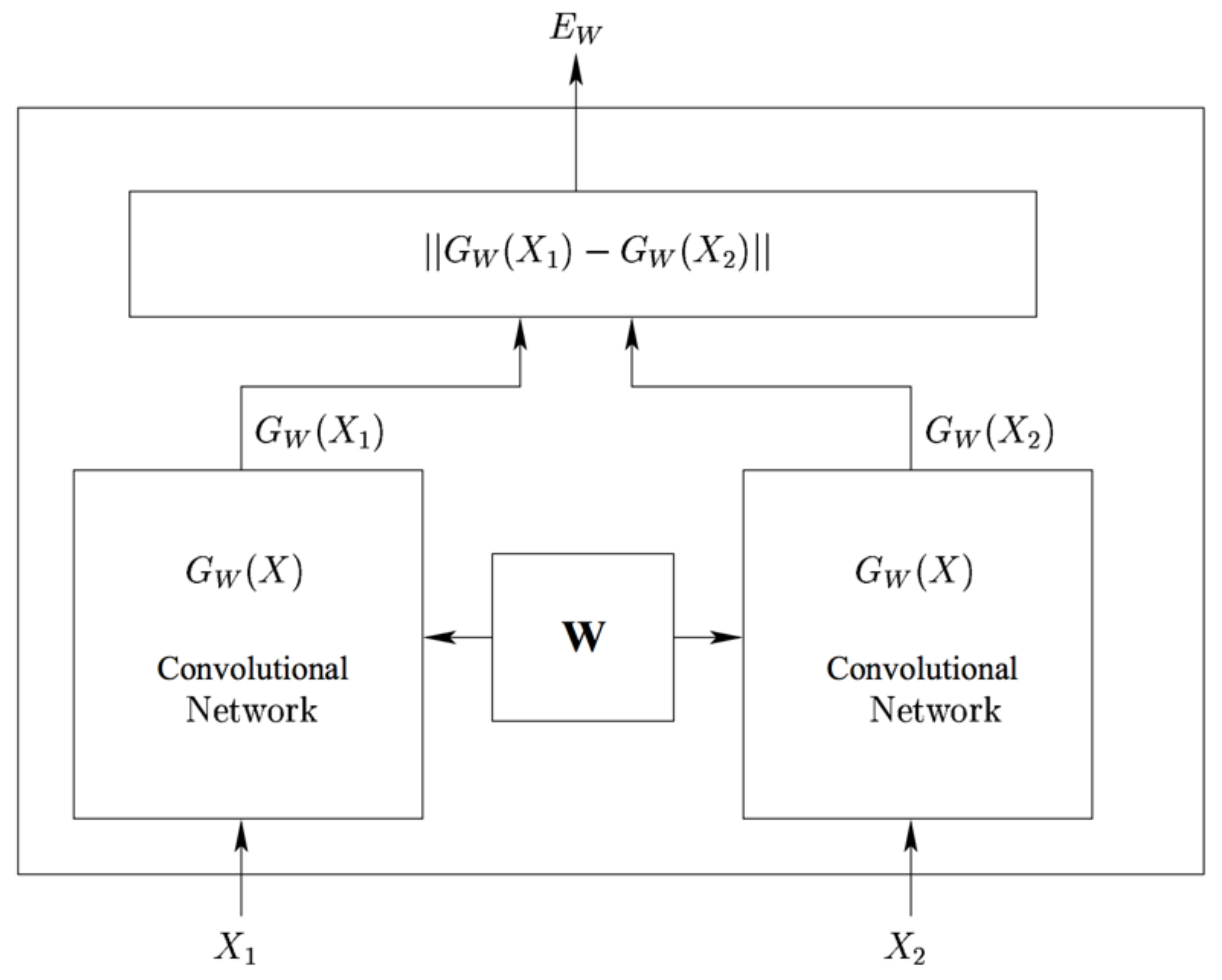
[Szegedy **2014**]

CONVOLUTIONAL NEURAL NETWORKS (CNN)

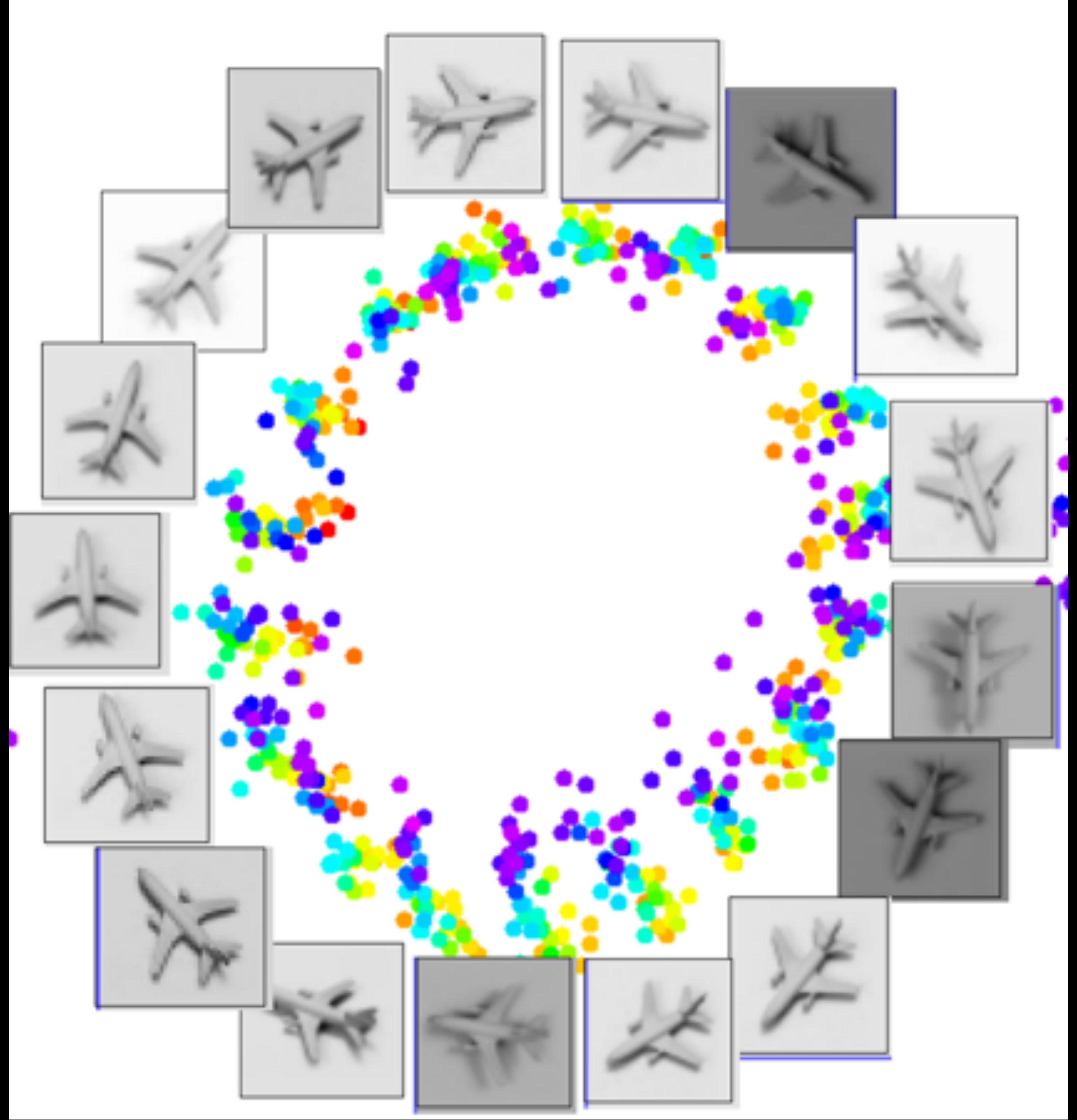


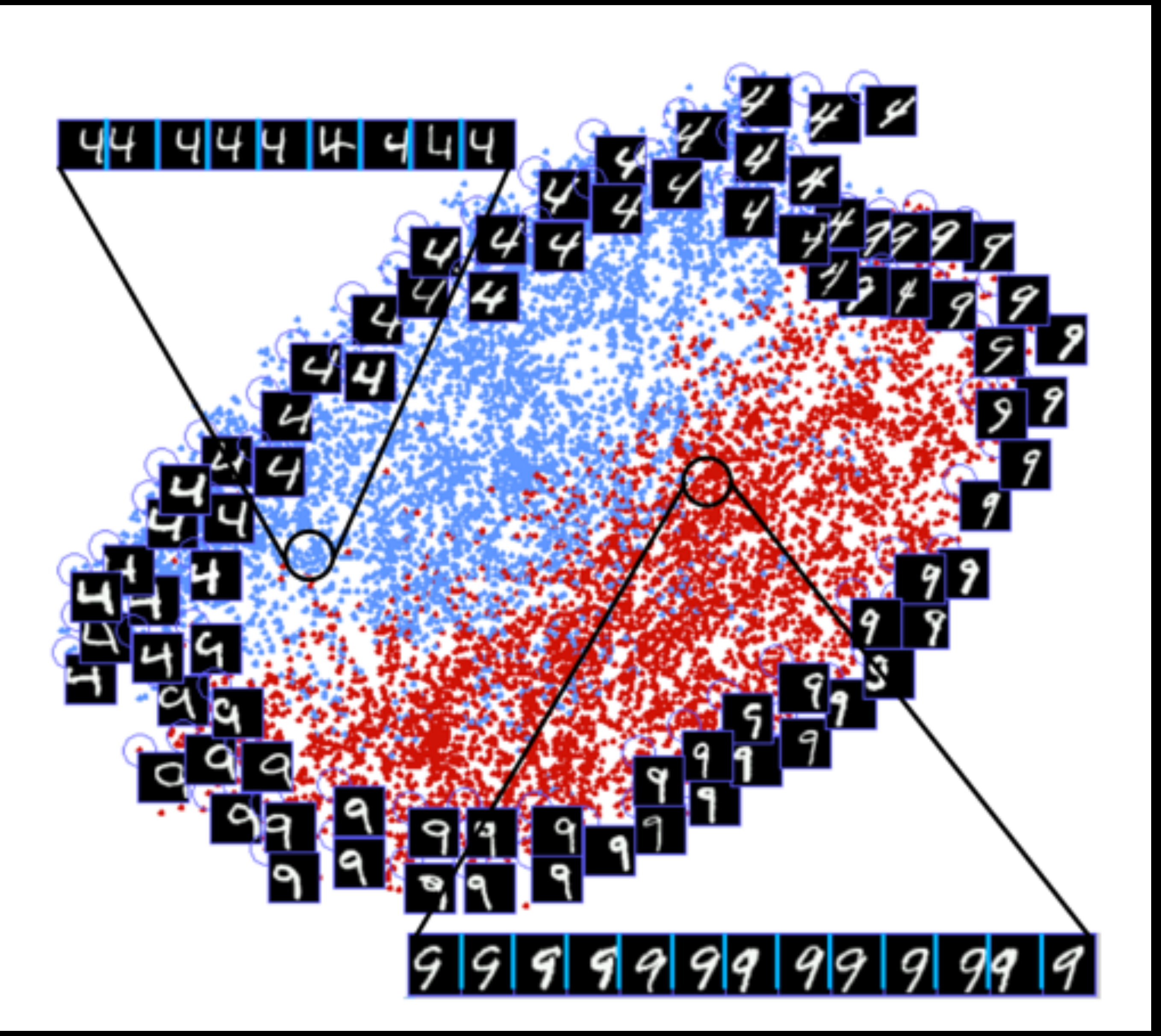
"GoogLeNet"
[Szegedy 2014]

SIAMESE NETWORK AND CONTRASTIVE LOSS

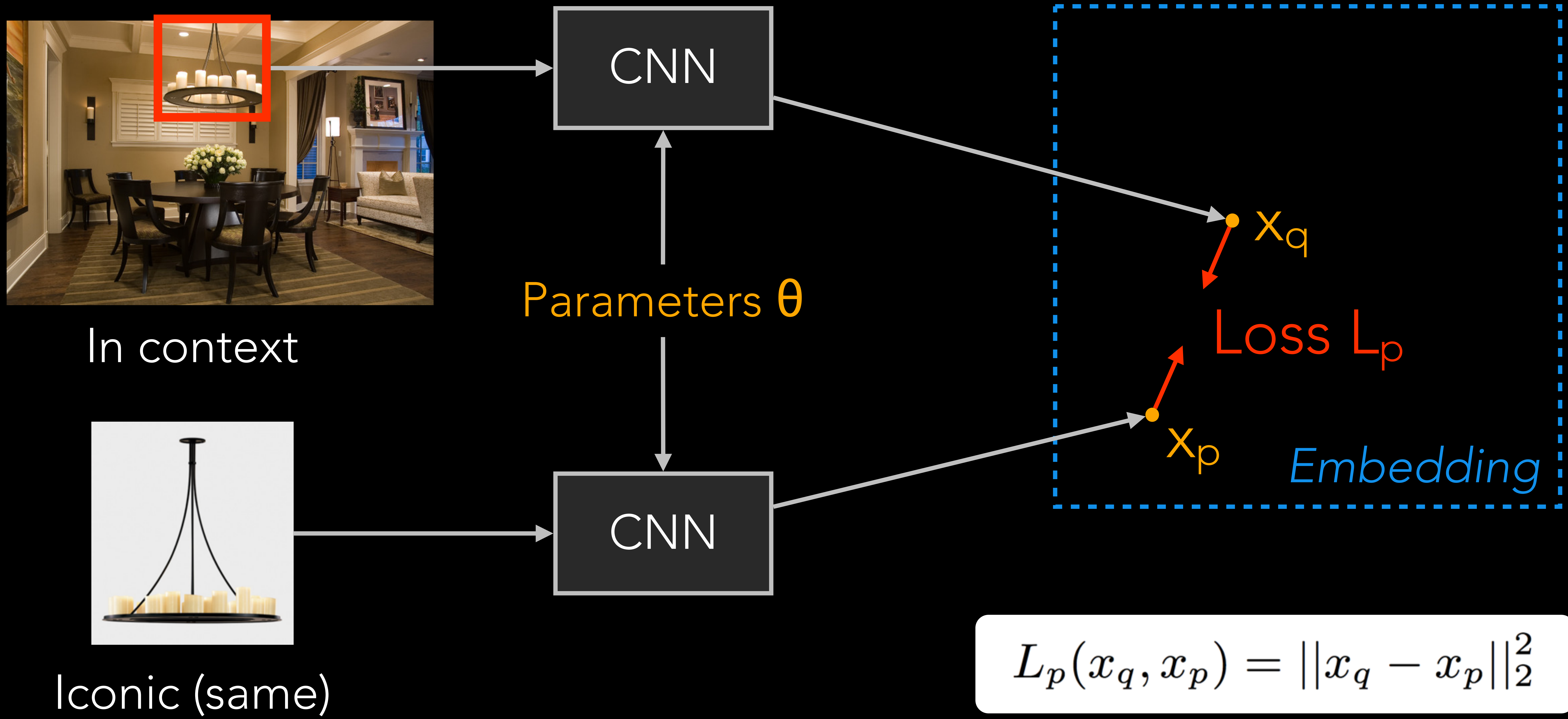


Siamese Architecture
[Chopra 2005, Hadsell 2006]



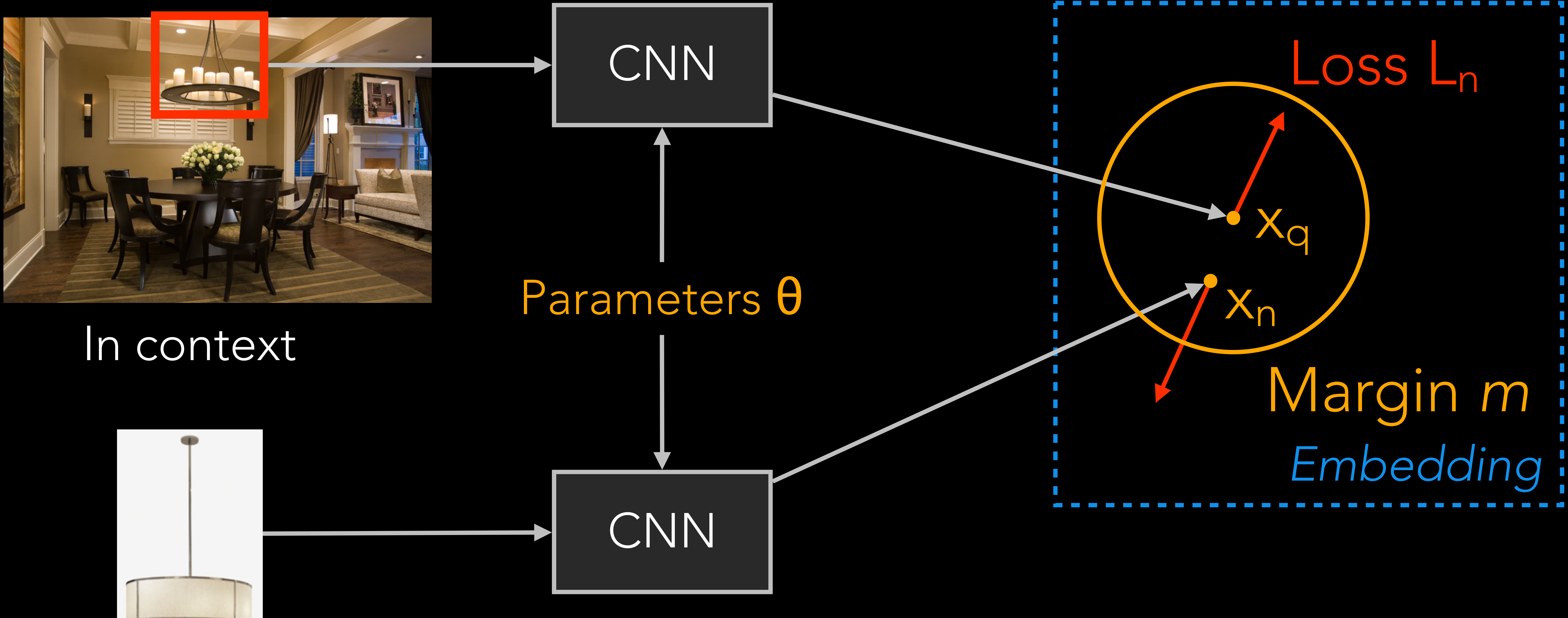


CONTRASTIVE LOSS: POSITIVE EXAMPLE



$$L_p(x_q, x_p) = ||x_q - x_p||_2^2$$

CONTRASTIVE LOSS: NEGATIVE EXAMPLE



$$L_n(x_q, x_n) = \max\left(0, m^2 - ||x_q - x_n||_2^2\right)$$

CONTRASTIVE LOSS: ALL TOGETHER

$$L(\theta) = \underbrace{\sum_{(x_q, x_p)} L_p(x_q, x_p)}_{\text{Penalty for similar images that are far away}} + \underbrace{\sum_{(x_q, x_n)} L_n(x_q, x_n)}_{\text{Penalty for dissimilar images that are nearby}}$$

$$L_p(x_q, x_p) = ||x_q - x_p||_2^2$$

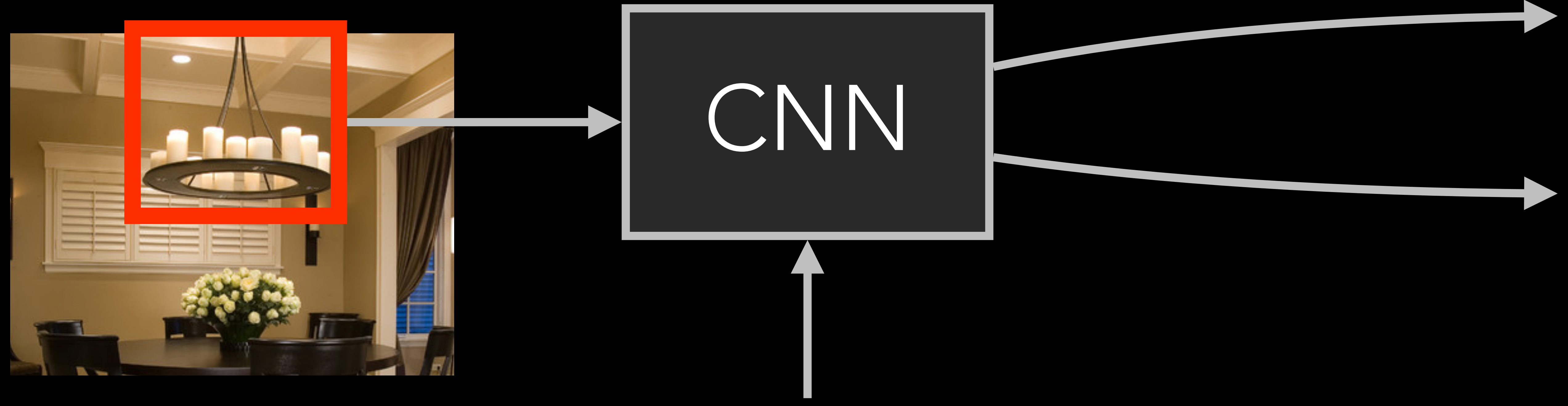
Margin

$$L_n(x_q, x_n) = \max\left(0, m^2 - ||x_q - x_n||_2^2\right)$$

Minimize L(θ) with stochastic gradient descent and momentum

[Chopra 2005, Hadsell 2006]

TRAINING WITH OBJECT CATEGORY



Category C_q

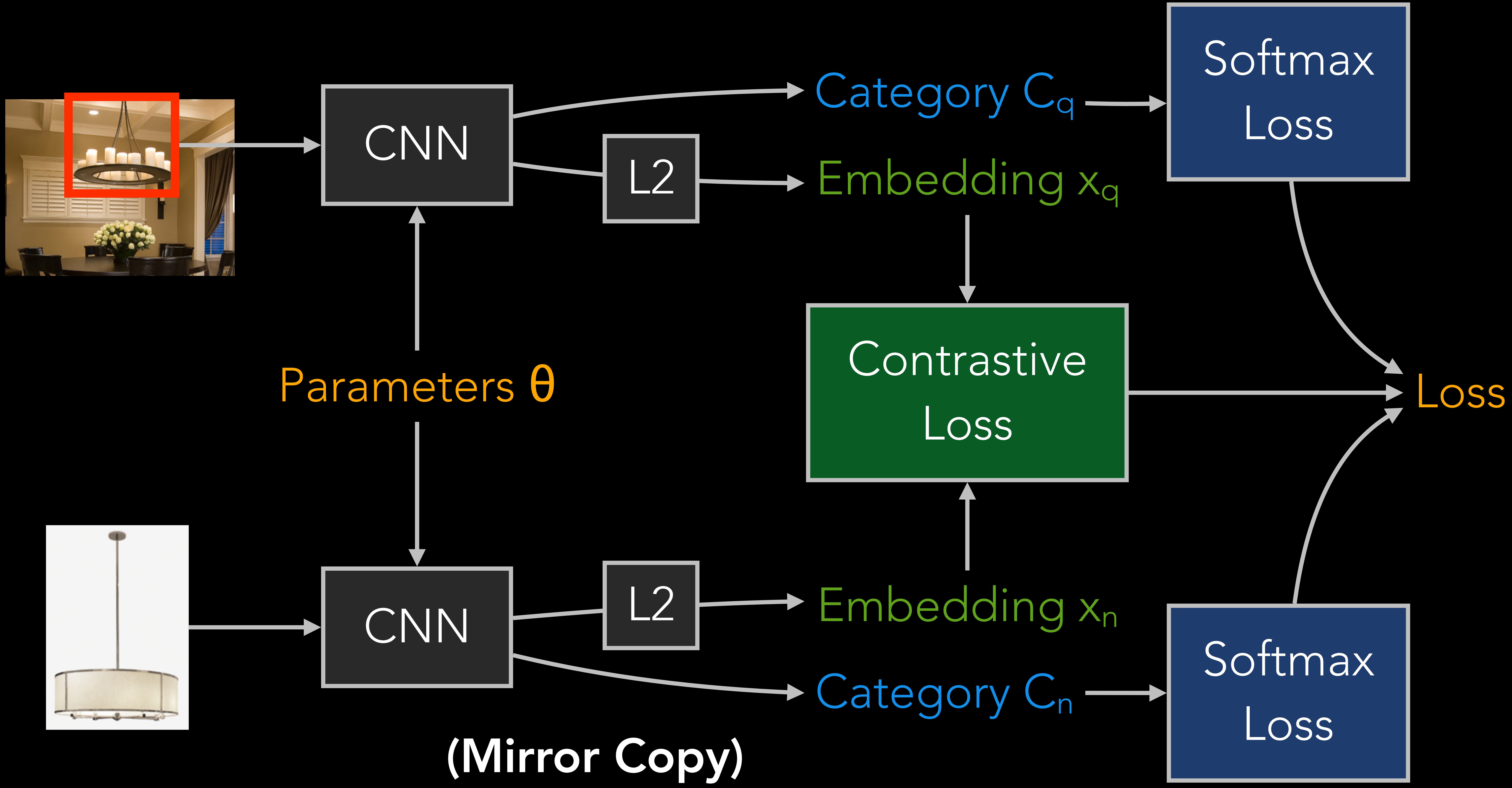
Embedding x_q

(839 categories, e.g. "Accent chair", "Pendant Light", …)

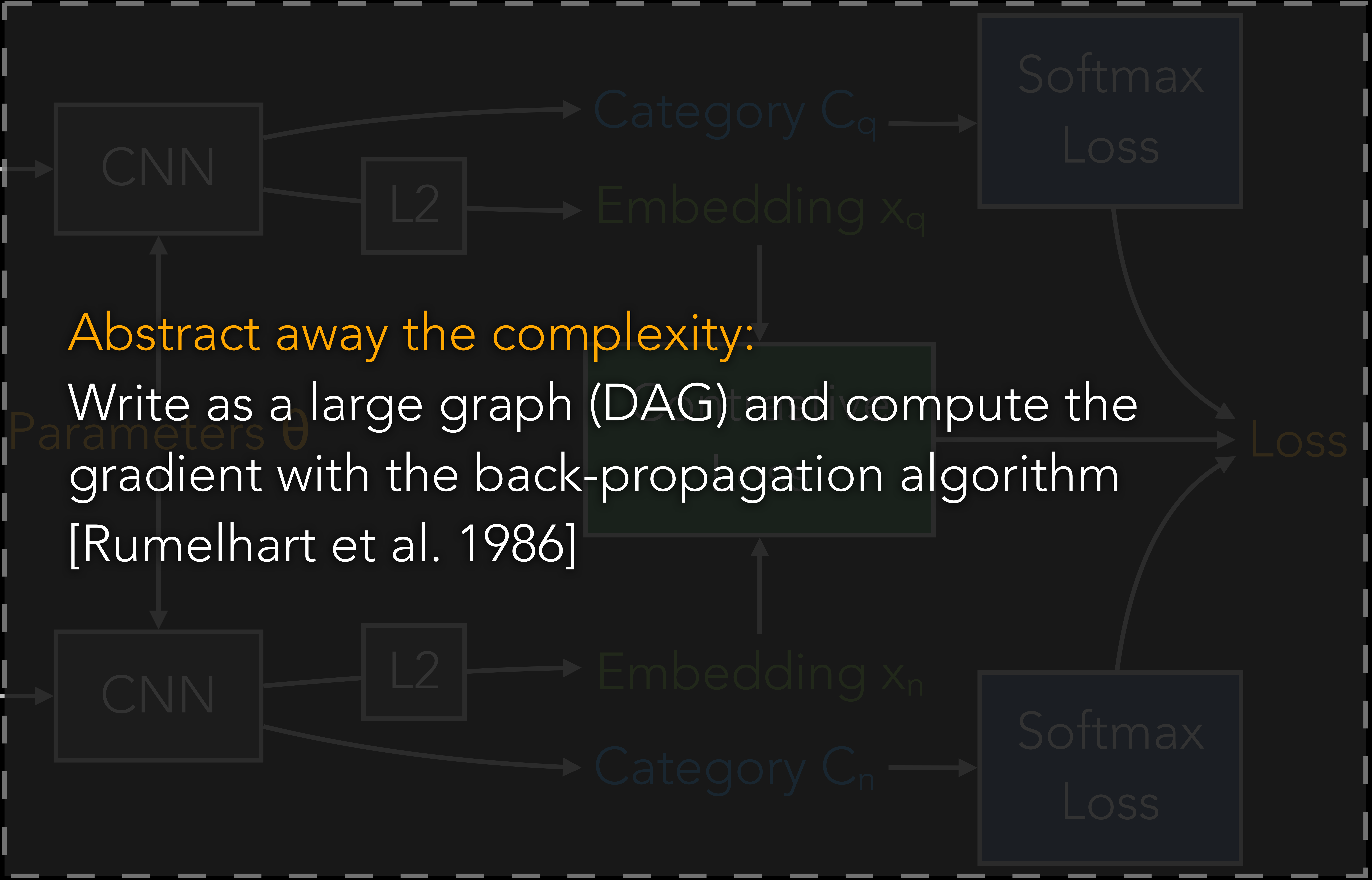
Parameters θ

- Use object categories to train CNN: helps regularize the CNN representation
  - Why? We have 3M iconic images but only 100k products in context
  - Categories give an additional training signal to the remaining 2.9M iconic images

TRAINING: FULL SYSTEM



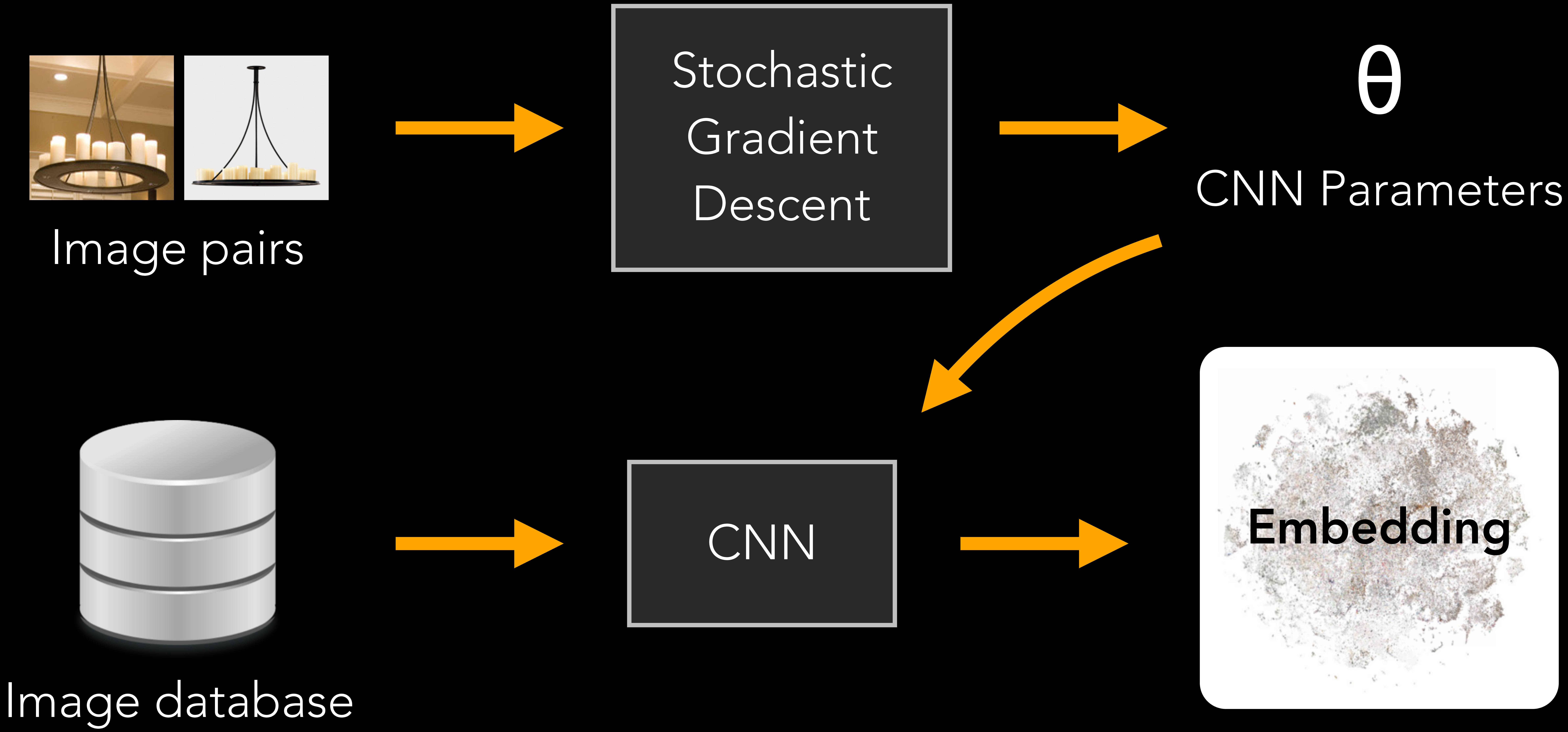
TRAINING: FULL SYSTEM



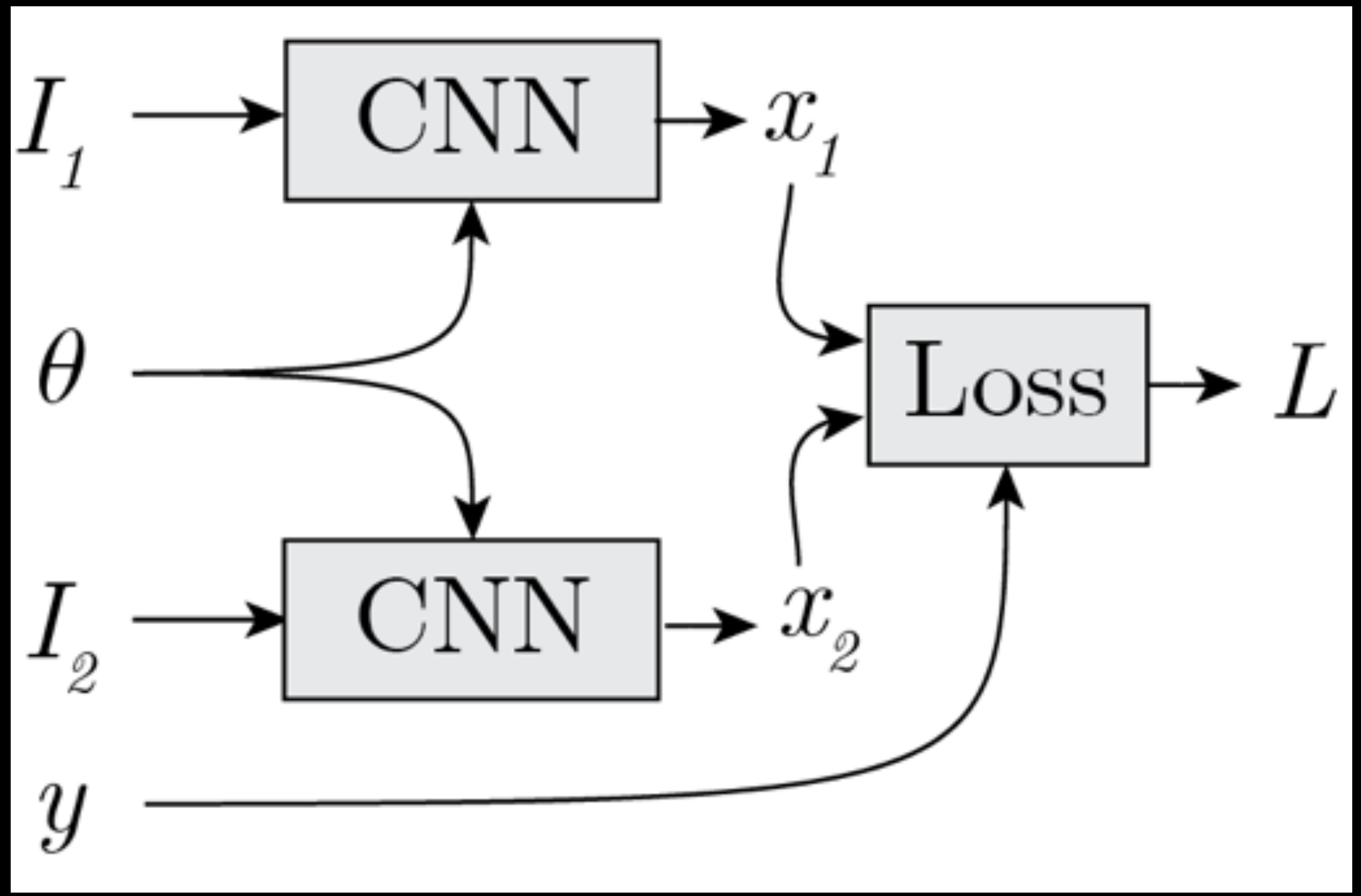
Abstract away the complexity:

Write as a large graph (DAG) and compute the gradient with the back-propagation algorithm [Rumelhart et al. 1986]

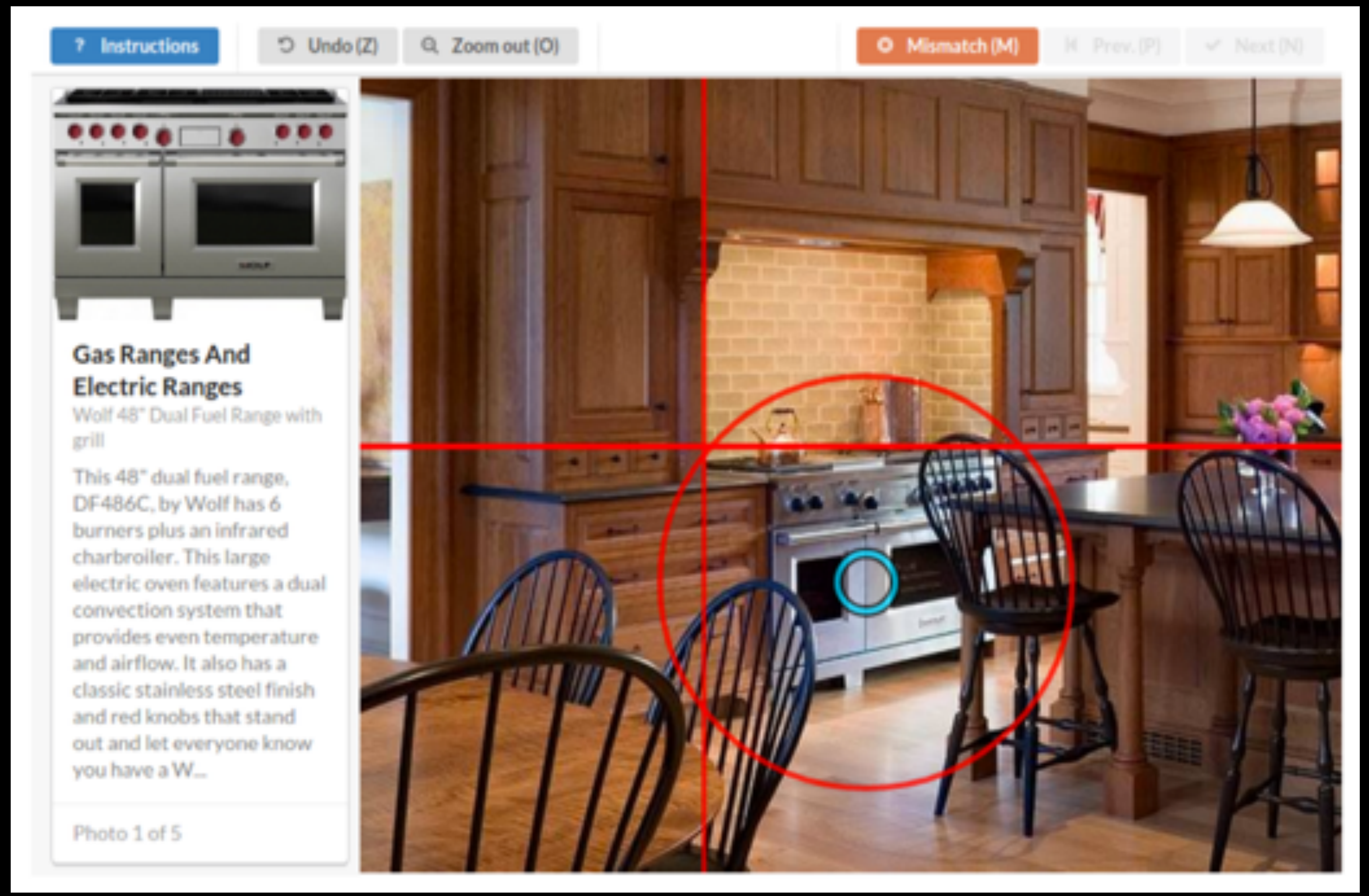
TRAINING PIPELINE



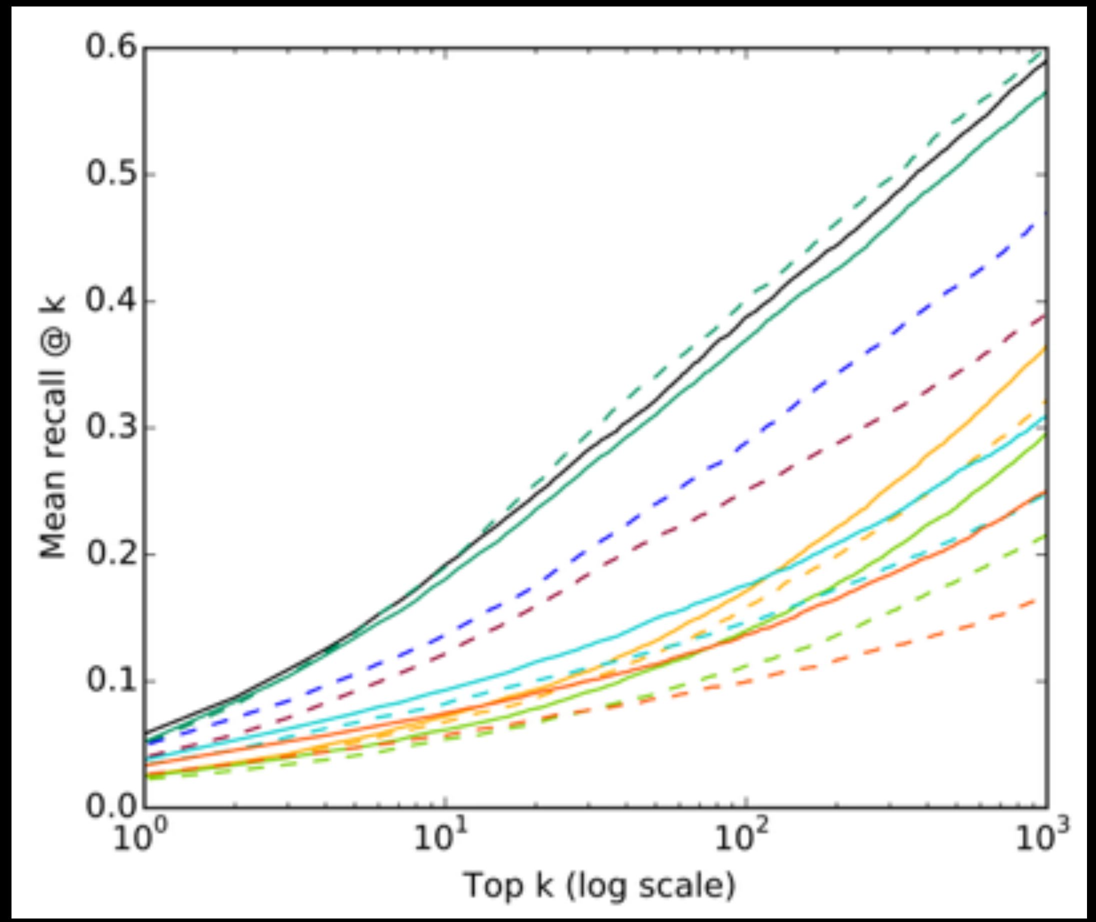
OVERVIEW



(1) Training the CNN

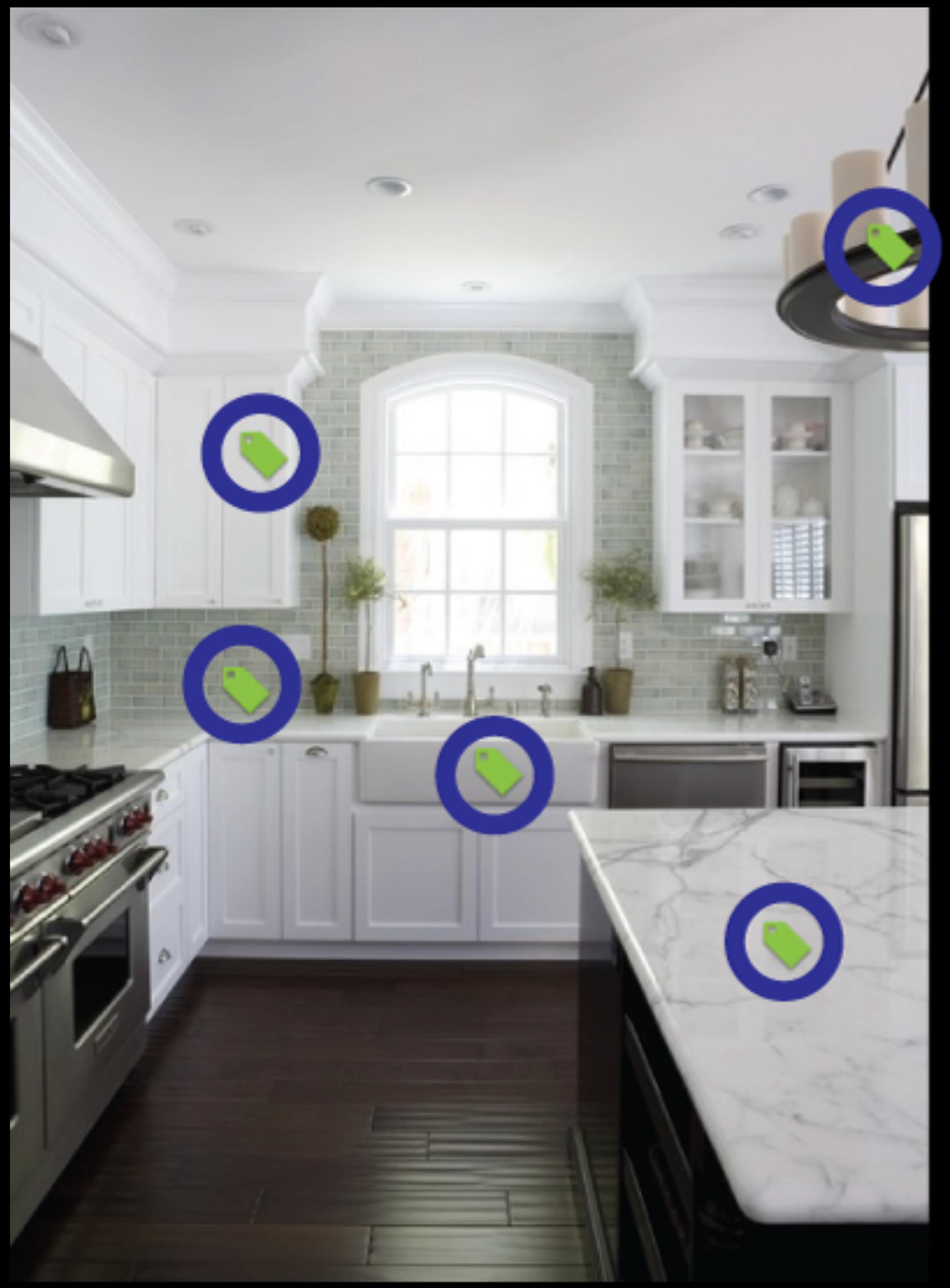


(2) Gathering training data



(3) Search Results

HOUZZ.COM PRODUCT TAGS



(a) Full scene

Some scenes (~2%) have product tags

HOUZZ.COM PRODUCT TAGS

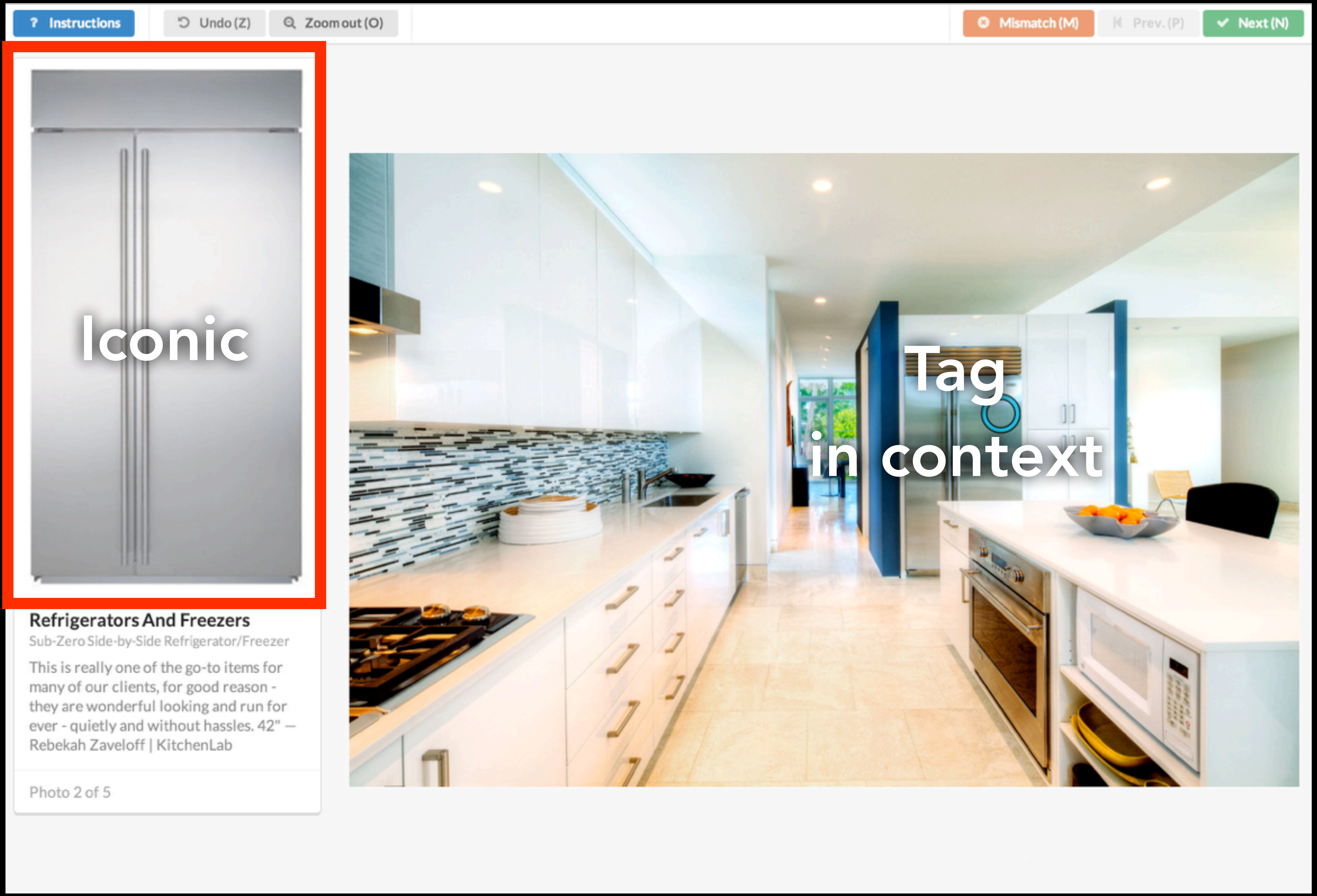
(a) Full scene

(b) Iconic product images

Problem: what is the spatial extent of each tag?

SPATIAL EXTENT: CROWDSOURCING ON MECHANICAL TURK

- Users are shown the iconic image and tag in context
- Users draw a bounding circle to set the zoom level (1 click)
- Users draw a bounding box (4 clicks)
- Can label "mismatch"



SPATIAL EXTENT: QUALITY CONTROL

Duplication:

- Multiple workers draw boxes
- Until any 2 agree with IOU > 0.7

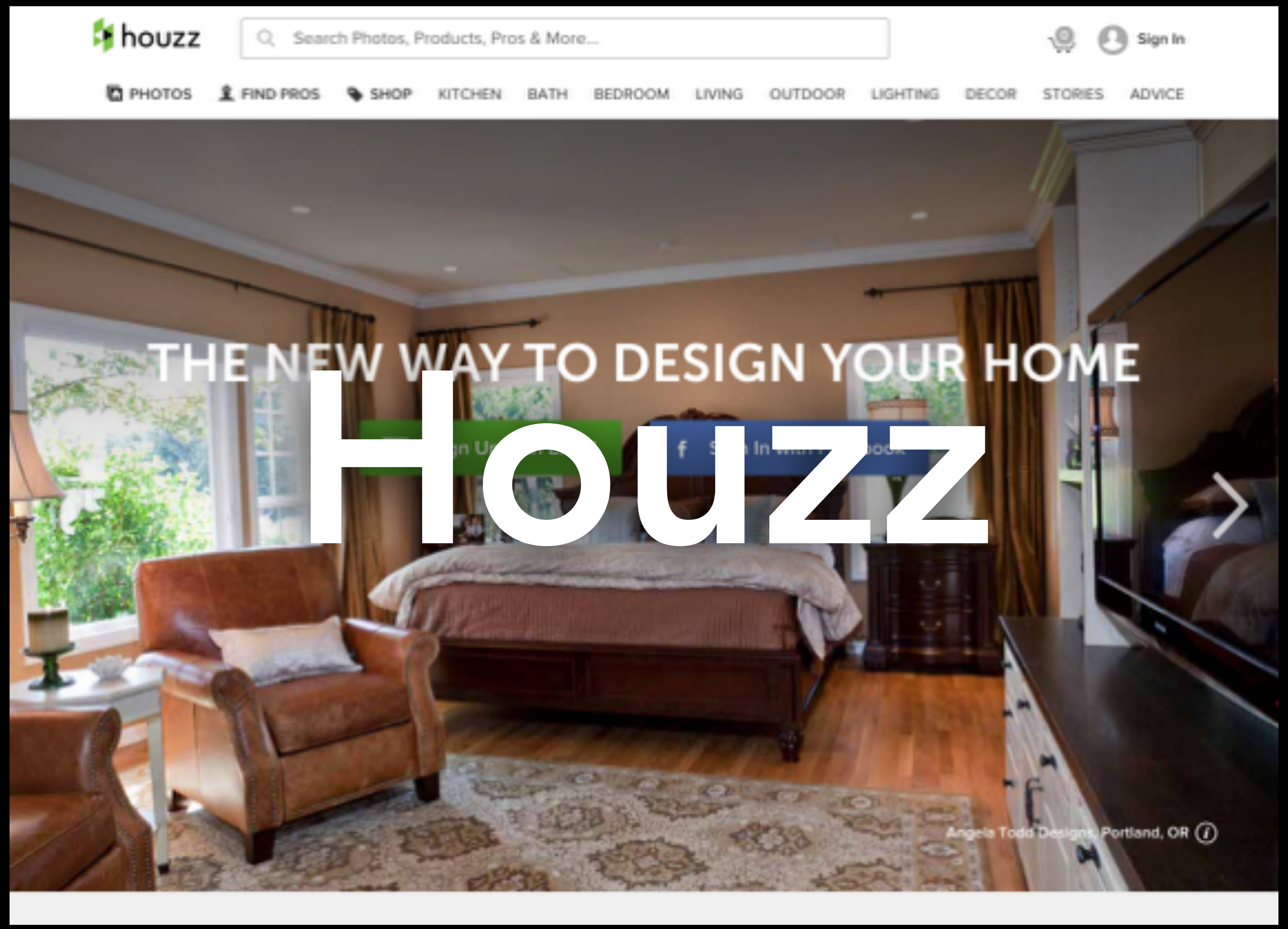
Sentinels:

- Secret test items (1 out of 7)
- Block low accuracy workers

$$\mathrm{IOU}(A, B) = \frac{\mathrm{Intersection}(A,B)}{\mathrm{Union}(A, B)}$$

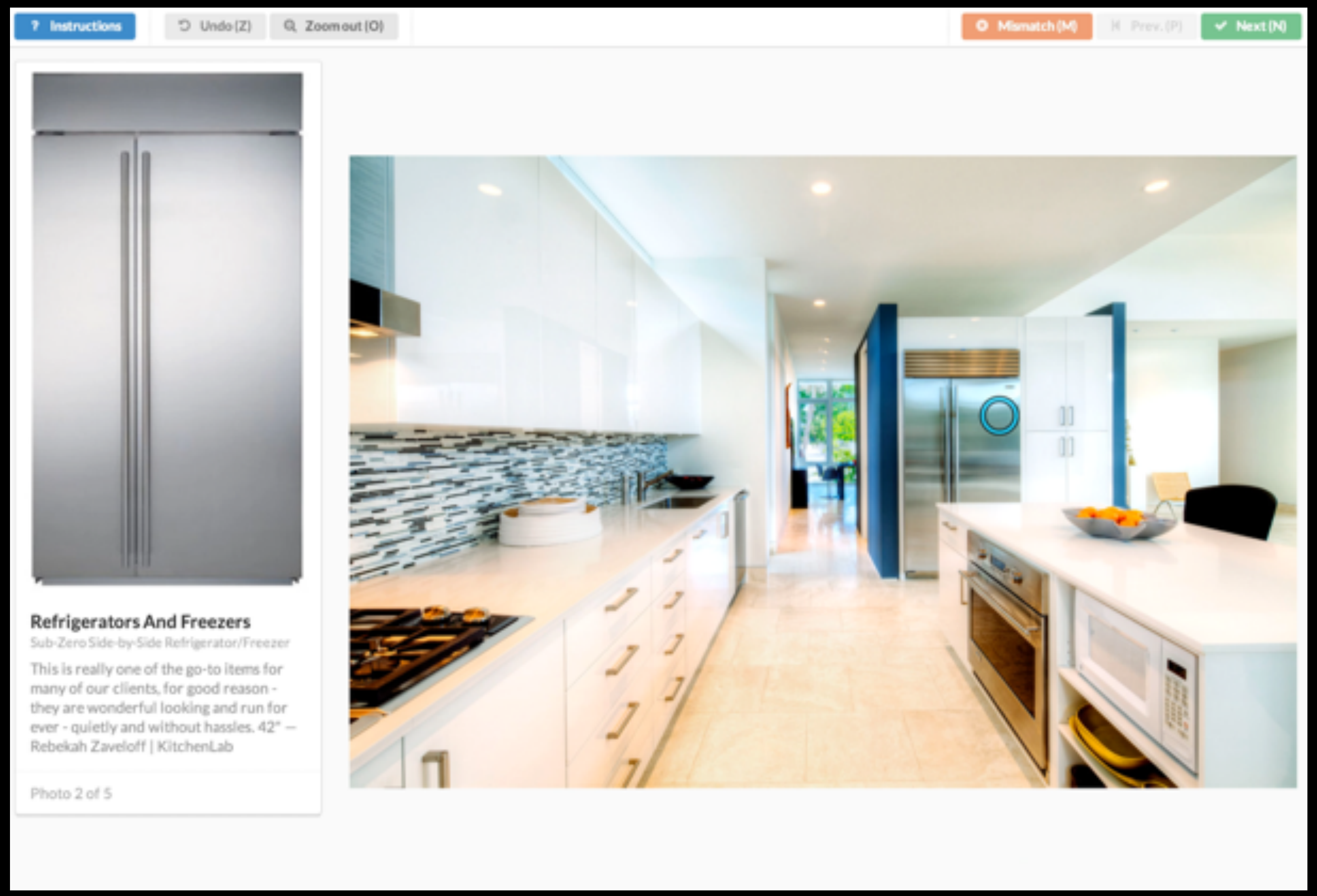
CROWDSOURCING PIPELINE

(1) Browse houzz.com, remove near-duplicates:



3M products images
180k product tags

(2) Crowdsourcing on Mechanical Turk:



100k boxes
with high agreement

(3) Sample image pairs for training:

SAMPLING IMAGE PAIRS

Boxes

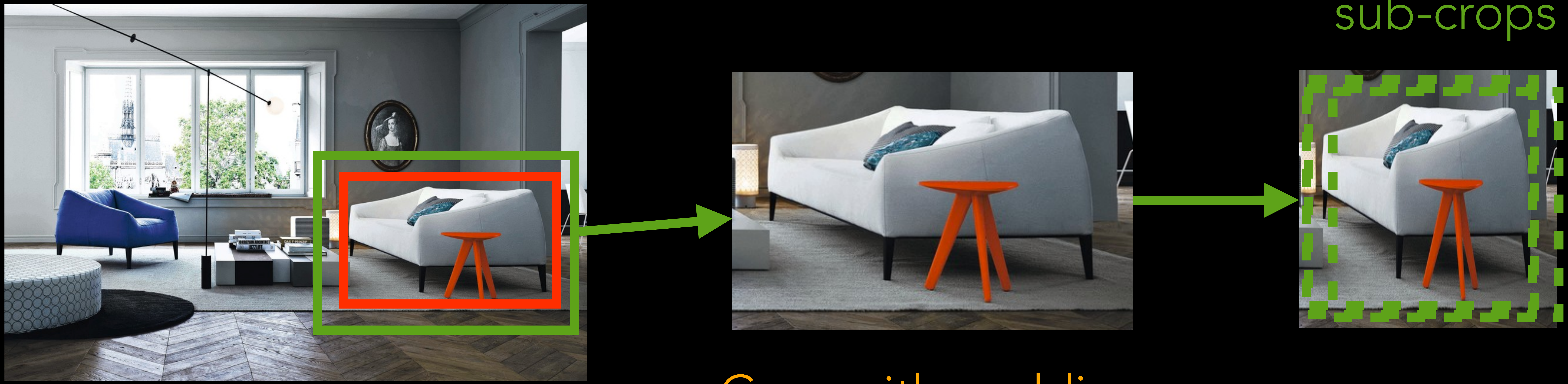
Sampled image pairs

For each box:

- Add 1 positive pair: (in-context box, tagged iconic image)
- Add 20 negative pairs: (in-context box, random other iconic image)
  80% from same category, 20% from different category

CROPPING IMAGES FOR CNN INPUT

Random 224x224 sub-crops

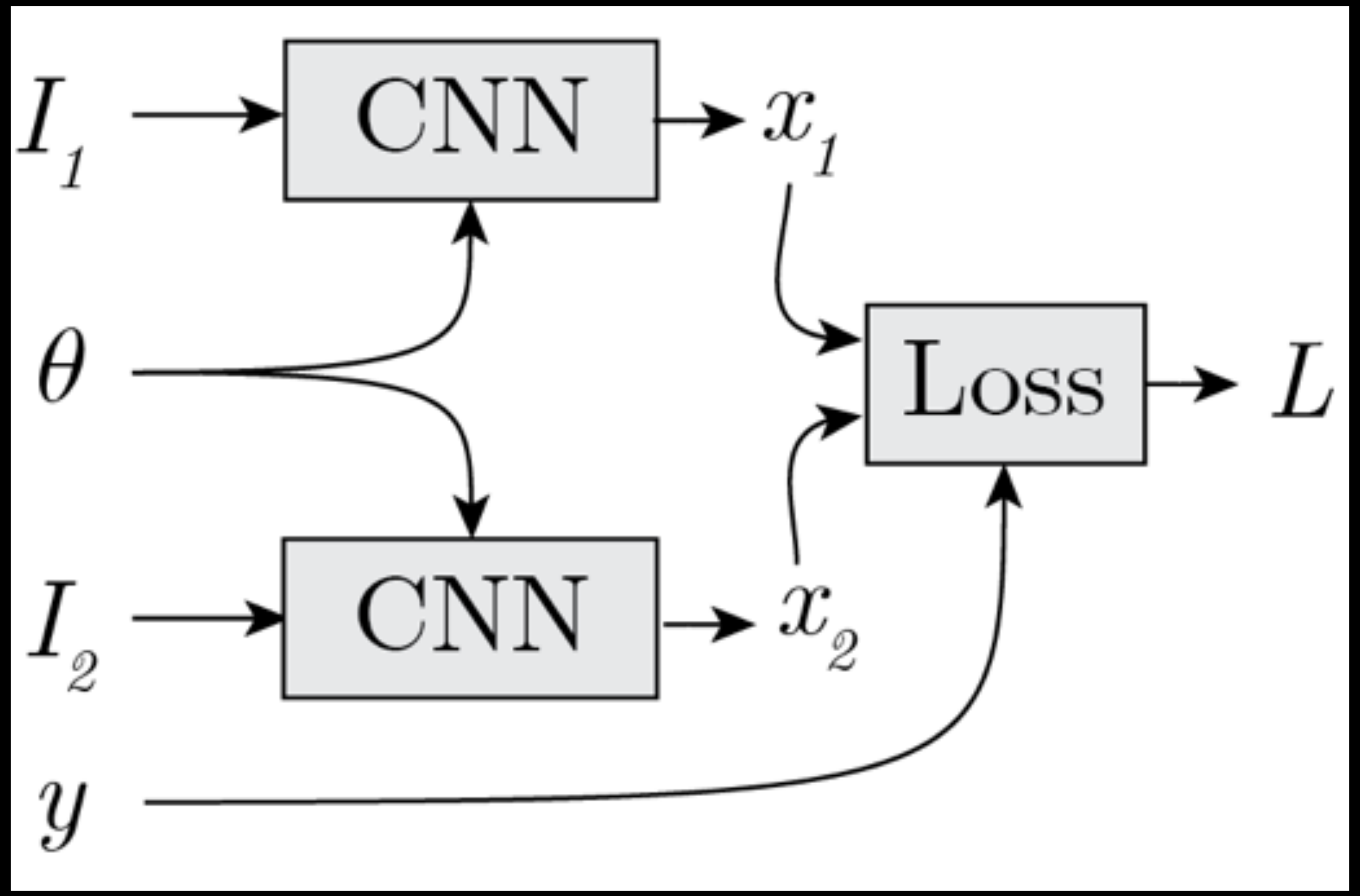


Original image

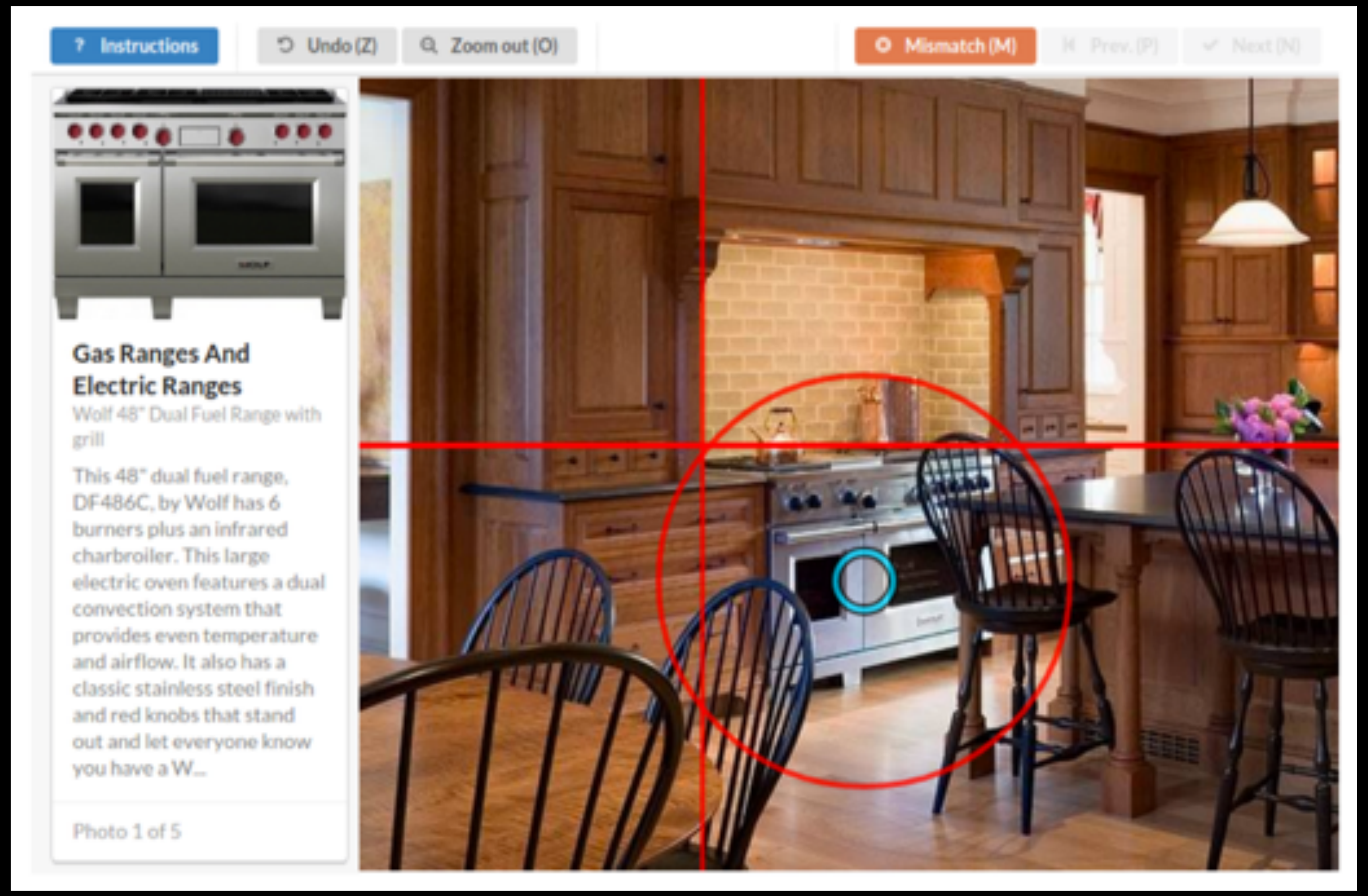
Crop with padding (repeat with different amounts)

Warp to square (256x256)

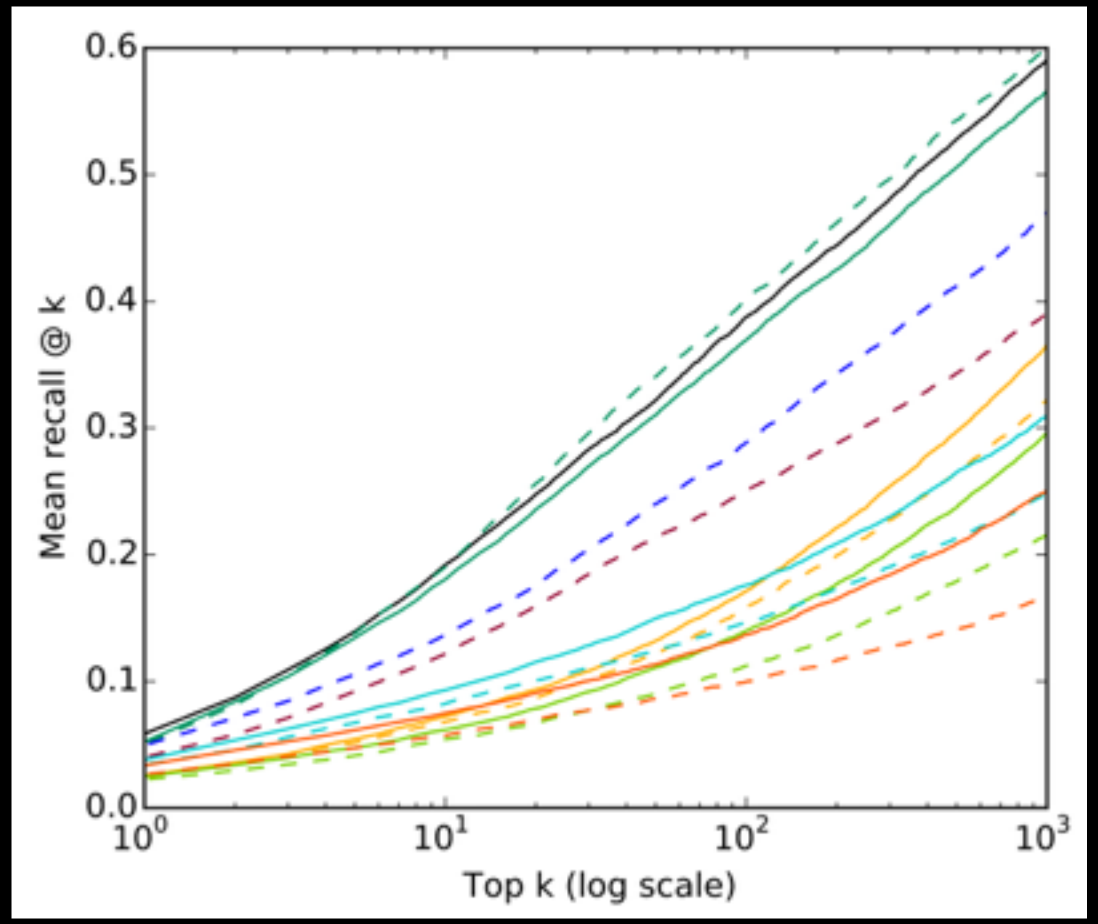
OVERVIEW



(1) Training the CNN



(2) Gathering training data



(3) Search Results

Train the CNN for 200,000 mini-batches…

(go for a *very* long walk and come back)

Runtime: milliseconds

EMBEDDING (PROJECTED, TO 2D)

EMBEDDING VISUALIZED IN 2D: **CHAIRS**

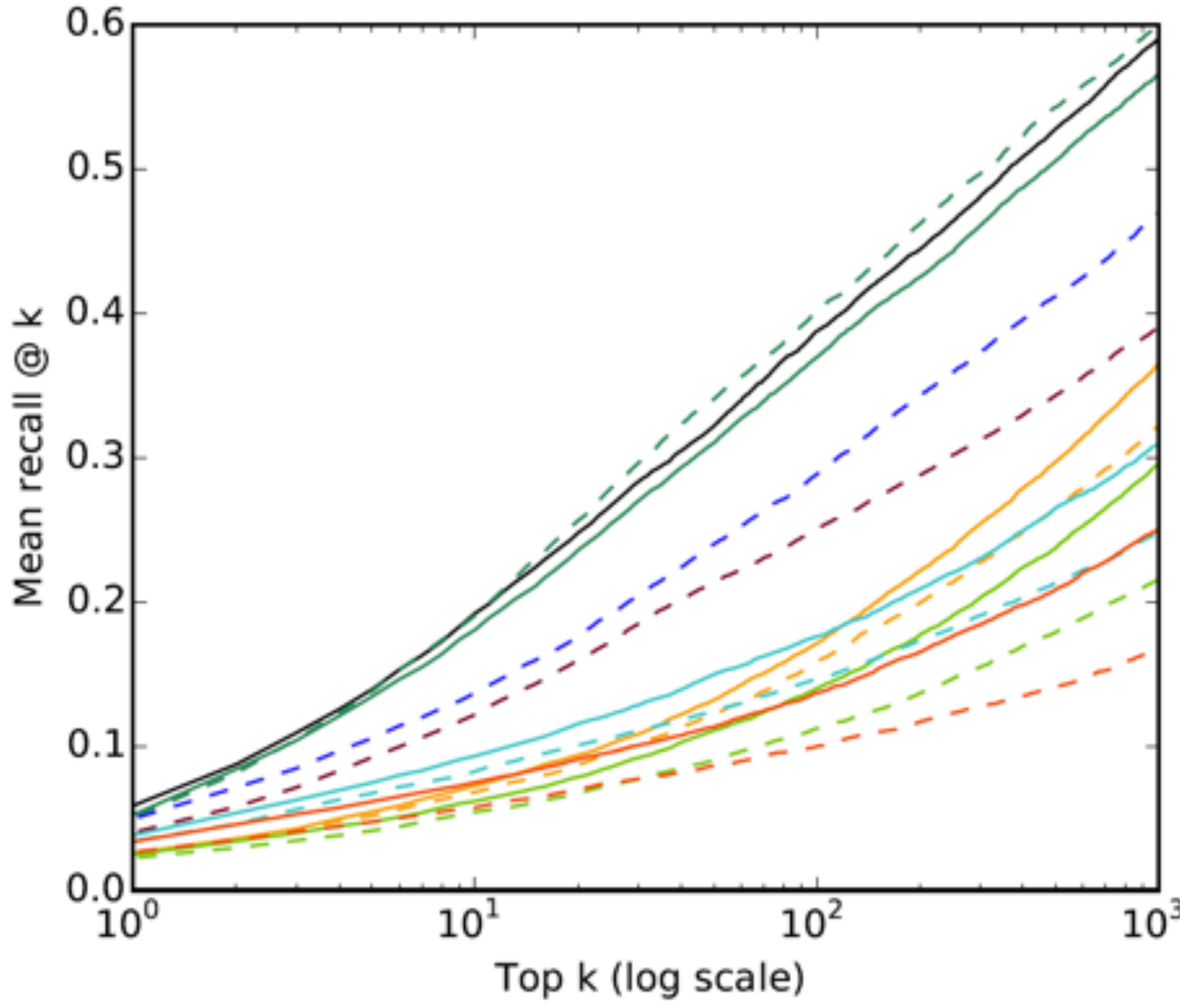
RESULTS

(1) "What is it?"

(2) "Where is it used?"

(3) Cross-category search

(4) Quantitative Evaluation



RESULTS: "WHAT IS IT?"

In context

RESULTS: "WHAT IS IT?"

In context

Iconic

Louis Ghost Chair with Arms

Living Room | Contemporary | 1 | View jpg

Outdoor Lounge Chairs

Eclectic | 44 | View jpg

2 boxes

Top 4 results:

Victoria Ghost Chair by Philippe Starck for Kartell | Design Within Reach

Dining Chairs | Modern

15 | View jpg | View detail

0.019

Kartell | Victoria Ghost Chair, Set of 4

Dining Chairs | Modern

View jpg | View detail

0.023

Kartell | Louis Ghost Chair

Dining Chairs | Contemporary

View jpg | View detail

0.024

Chaises

Office Chairs | Unknown style

View jpg | View detail

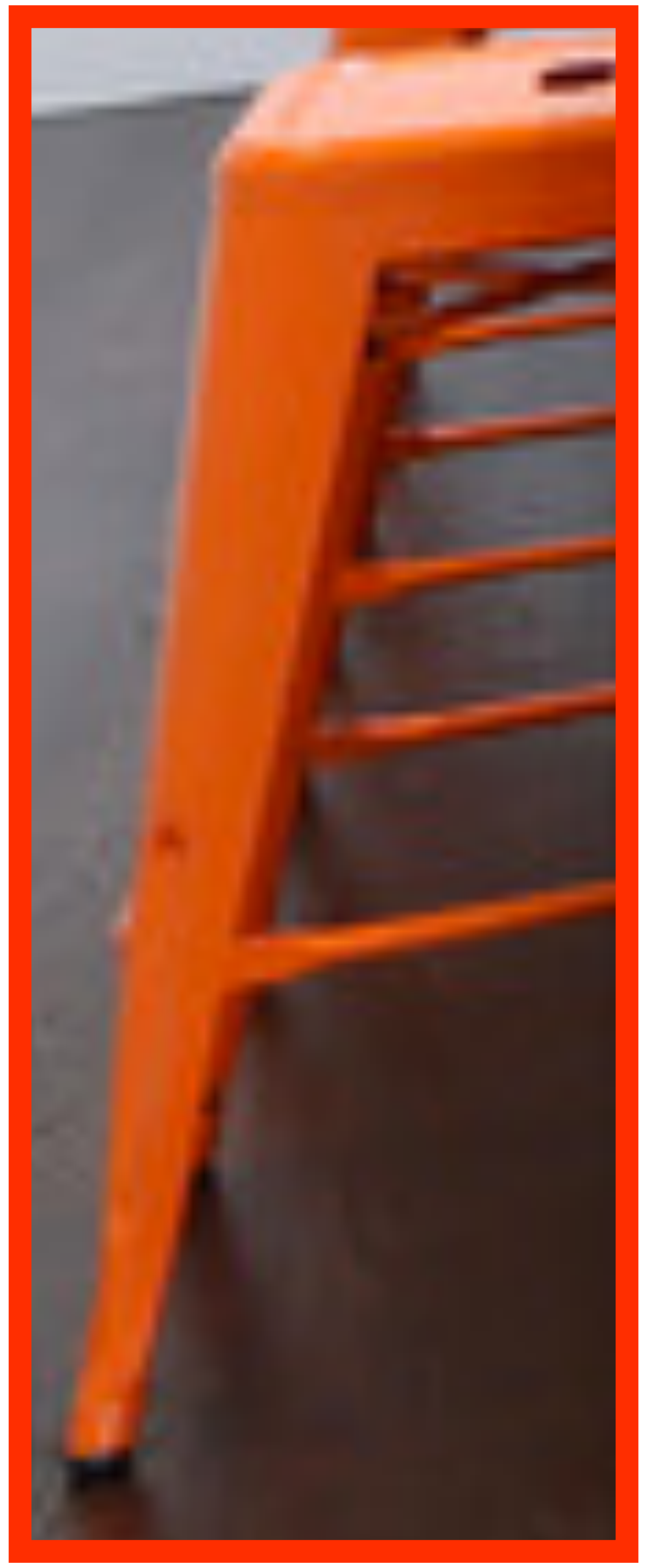
0.025

RESULTS: "WHAT IS IT?"

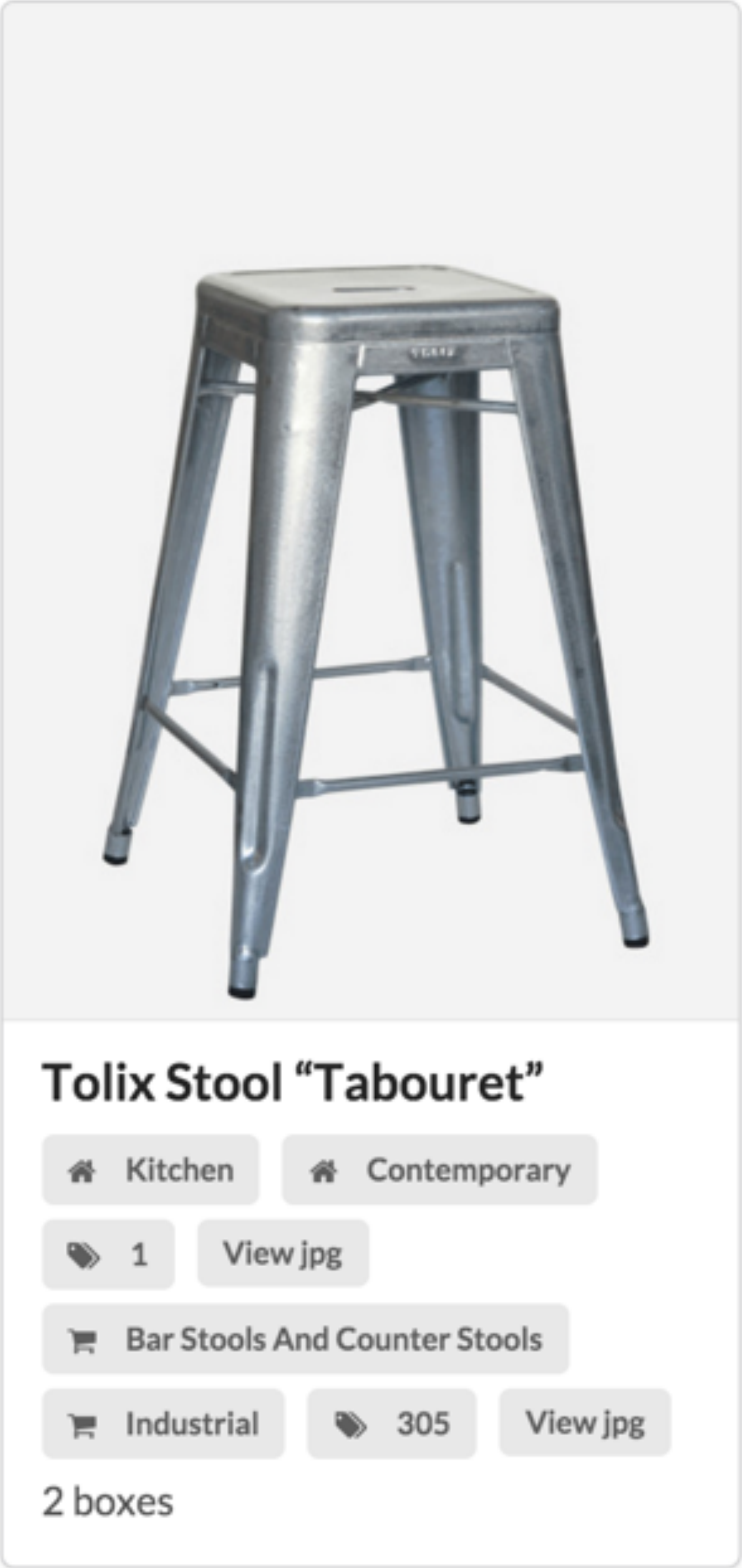
In context

RESULTS: "WHAT IS IT?"

In context



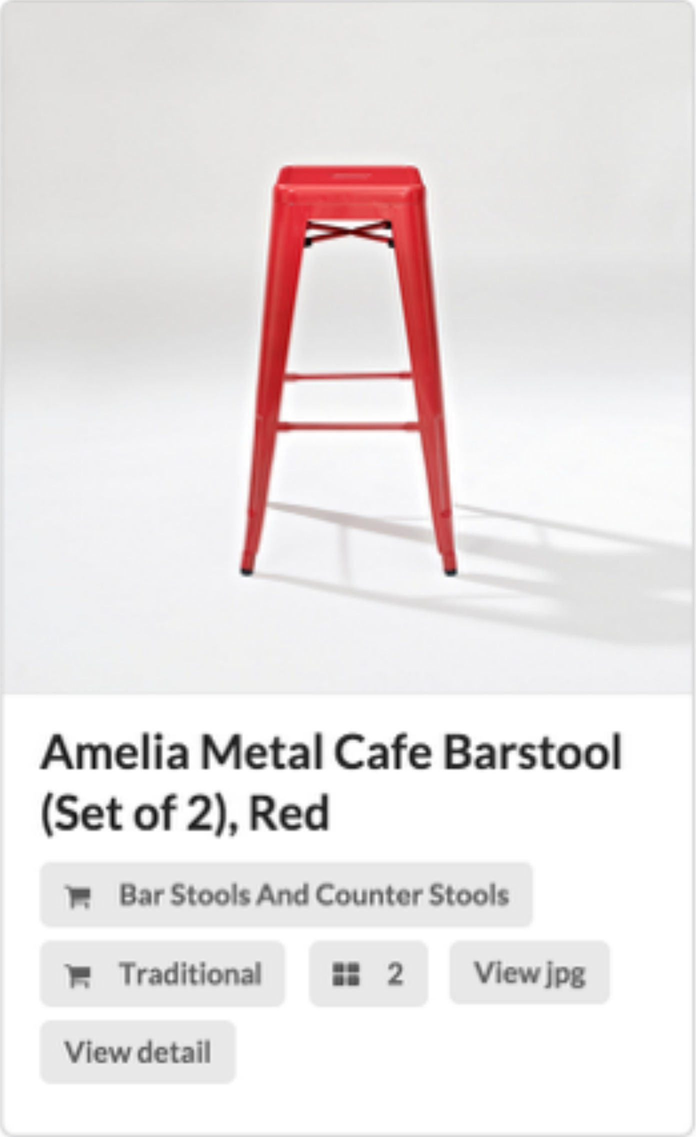
Iconic



Top 4 results:

0.046

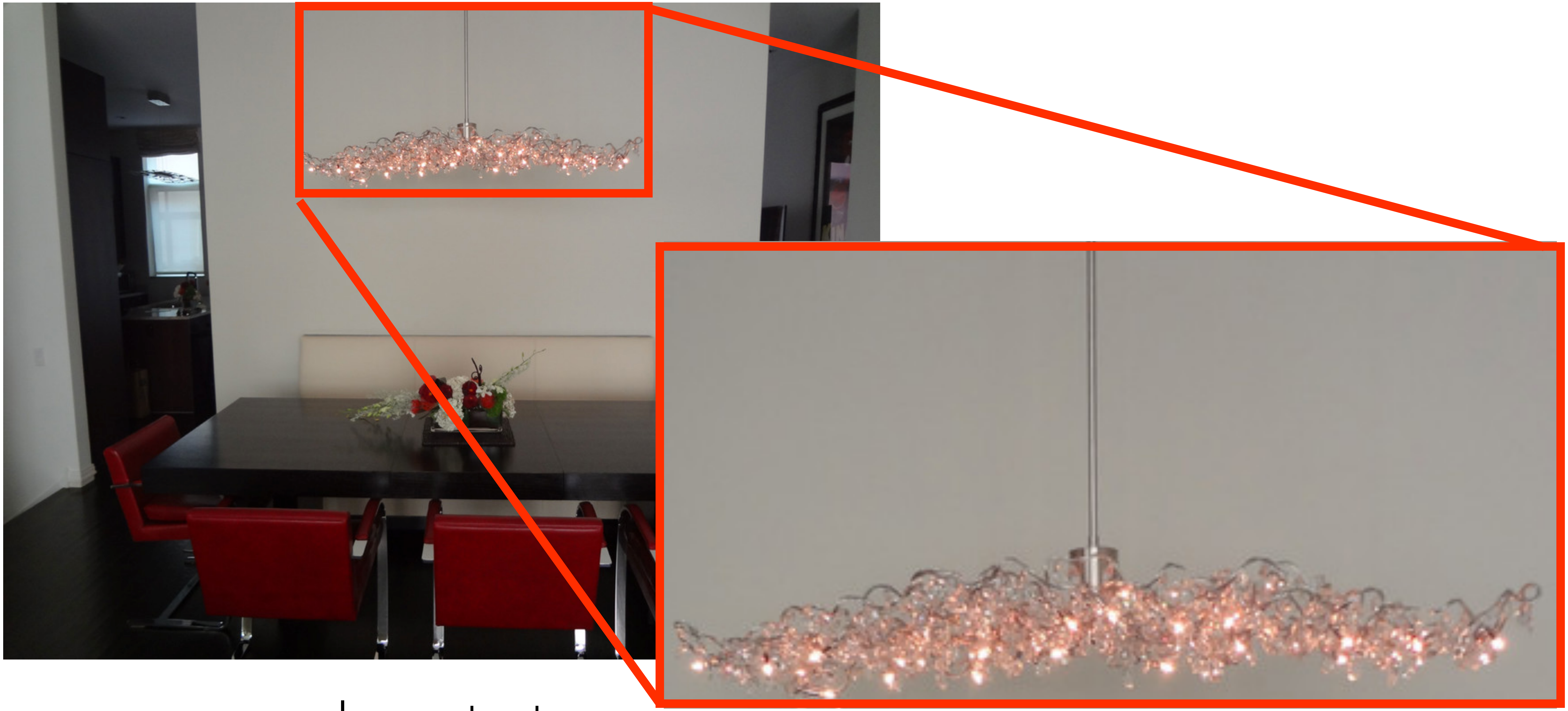
0.049



0.050

0.050

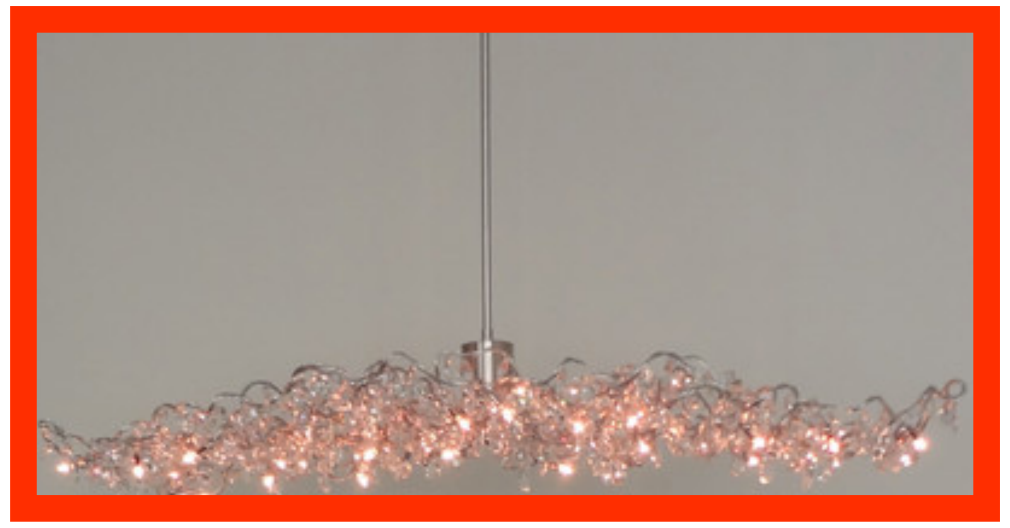
RESULTS: "WHAT IS IT?"



In context

RESULTS: "WHAT IS IT?"

In context



Iconic

Tiara Oval Suspension by Harco Loor

Dining Room | Contemporary | 1 | View jpg | Wall Sconces | Unknown style | 1 | View jpg

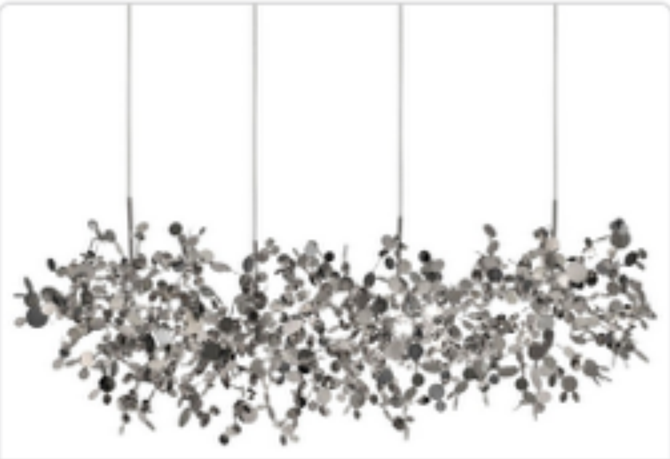
2 boxes

Top 4 results:

HARCO LOOR Tiara Chandelier

Chandeliers | Contemporary | 1 | View jpg | View detail

0.035



Argent N92S Suspension Light

Bathroom Vanity Lighting | Modern | View jpg | View detail

0.038

Eurofase 25620-016 Divo 9 Light Pendant in Nickel 25620-016

Pendant Lighting | Modern | View jpg | View detail

0.041

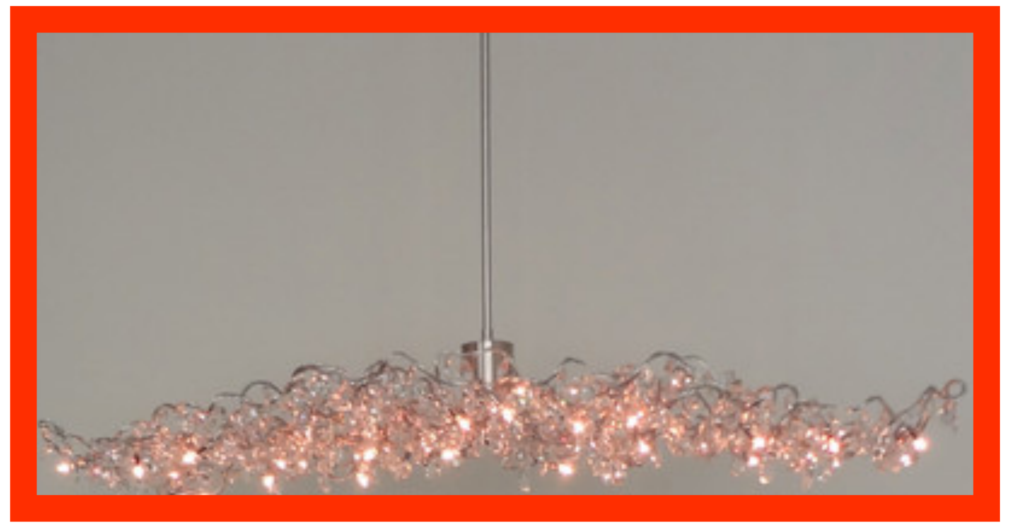
Eurofase 25618-013 Divo 6 Light Pendant in Nickel 25618-013

Pendant Lighting | Modern | View jpg | View detail

0.047

COMPARISON: TRAINED ONLY ON CATEGORIES

In context



Iconic

Tiara Oval Suspension by Harco Loor

Dining Room | Contemporary

1 | View jpg

Wall Sconces | Unknown style

1 | View jpg

2 boxes

Top 4 results:

'Apollon' Black Shaded 6-light Crystal Chandelier

Chandeliers

Contemporary

View jpg | View detail

1792.880

Authentic Deer and Elk Antler Banquet Table Chandelier

Chandeliers

Unknown style

View jpg | View detail

1895.330

Harco Loor | Tiara Oval HL 15 Suspension Light

Pendant Lighting

Modern | View jpg

View detail

1911.520

Vermeer Hexagonal Pendant

Pendant Lighting

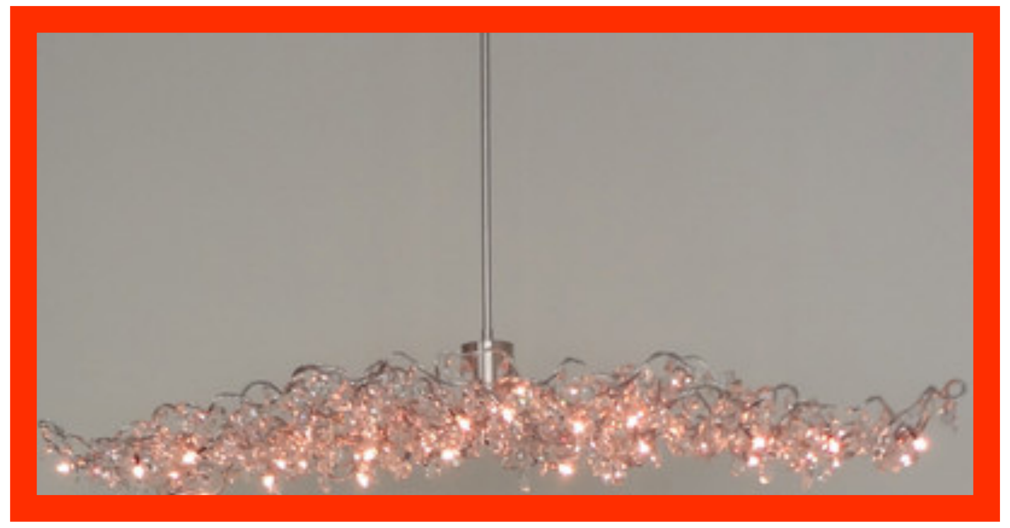
Modern | View jpg

View detail

2035.000

COMPARISON: TRAINED ONLY ON IMAGENET

In context



Iconic

Tiara Oval Suspension by Harco Loor

Dining Room | Contemporary

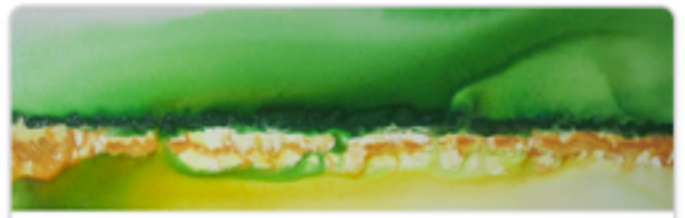
1 | View jpg

Wall Sconces | Unknown style

1 | View jpg

2 boxes

Top 4 results:



MERGING TOGETHER

Artwork

Contemporary

View jpg | View detail

498.493

"Elephant View" Artwork

Paintings

Contemporary

View jpg | View detail

499.135

Terzani Argent N92S Chandelier

Chandeliers

Modern | 2

View jpg | View detail

503.232

Mensa Hanging Light Fixture

Chandeliers

Contemporary

View jpg | View detail

514.867

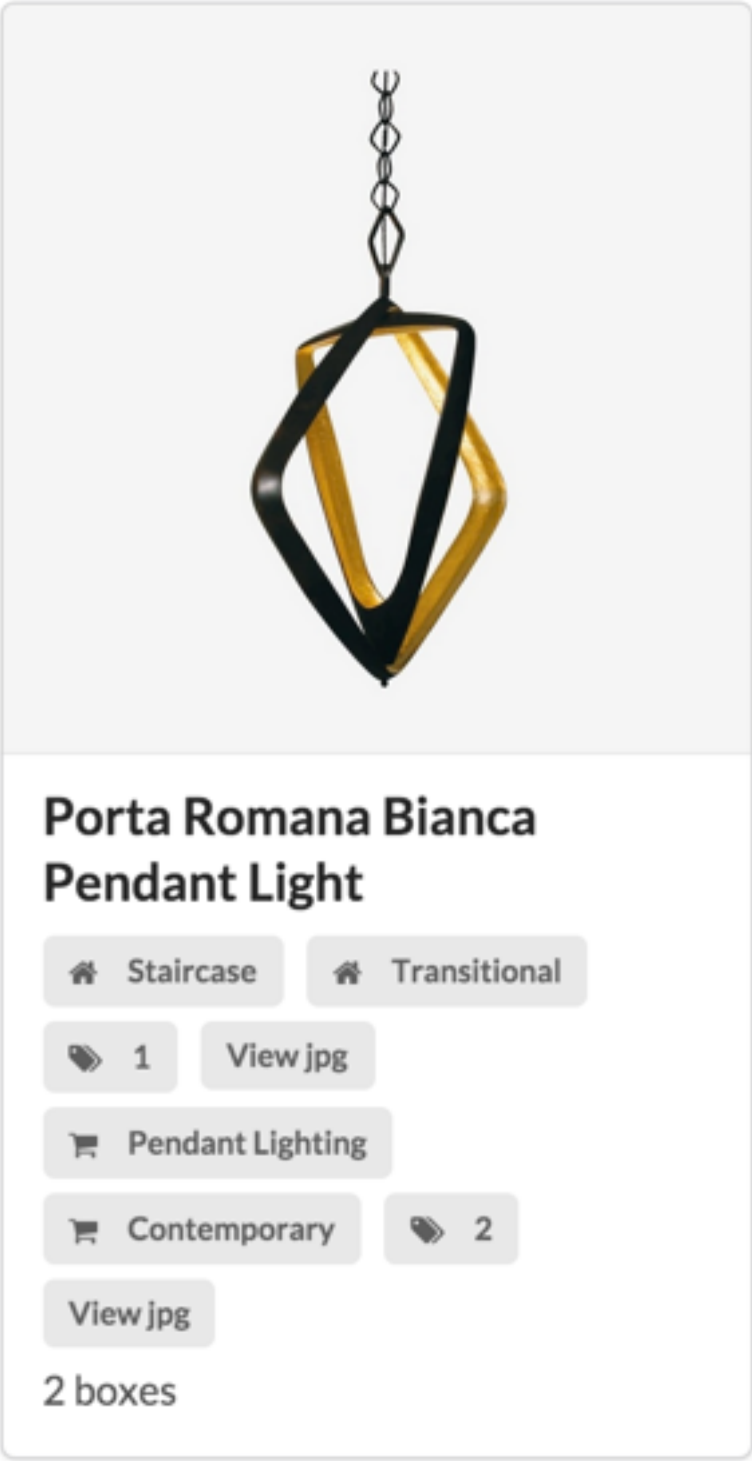
RESULTS: FAILURE CASE

In context

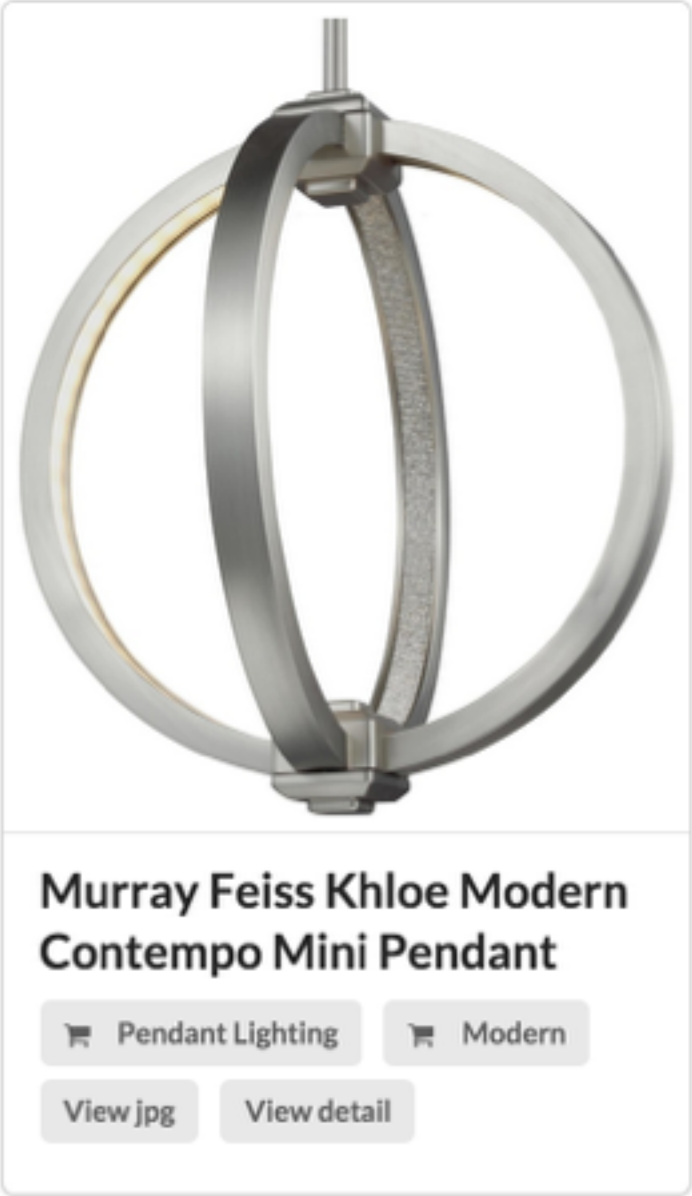
RESULTS: FAILURE CASE

In context

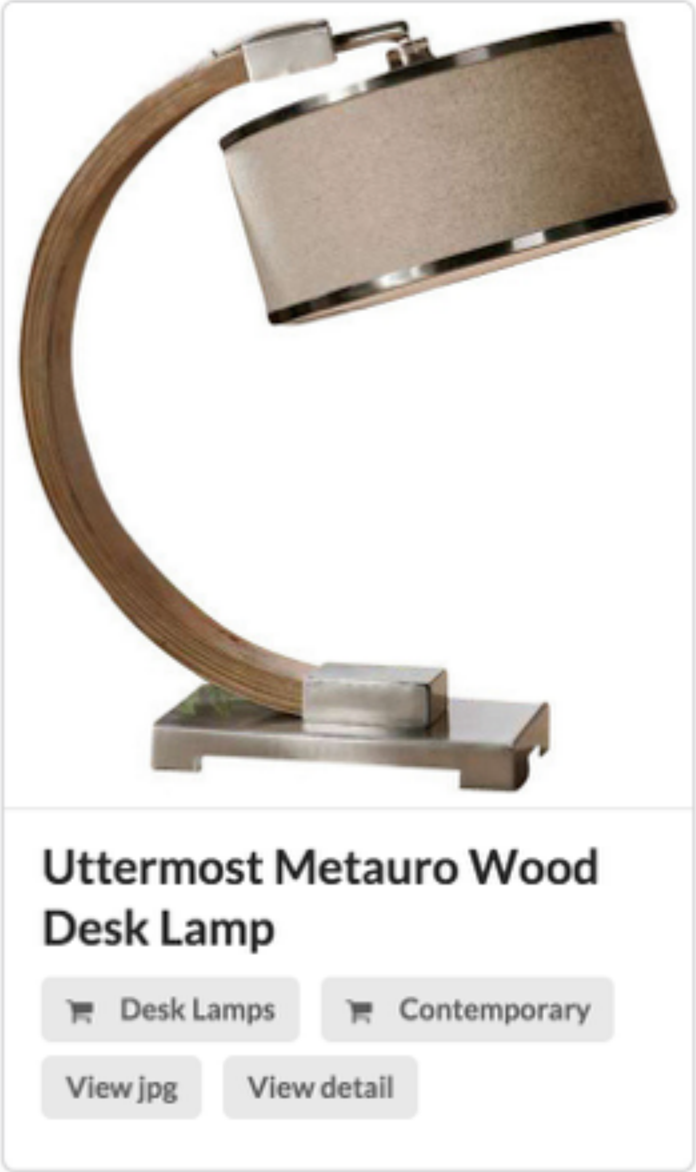
Iconic



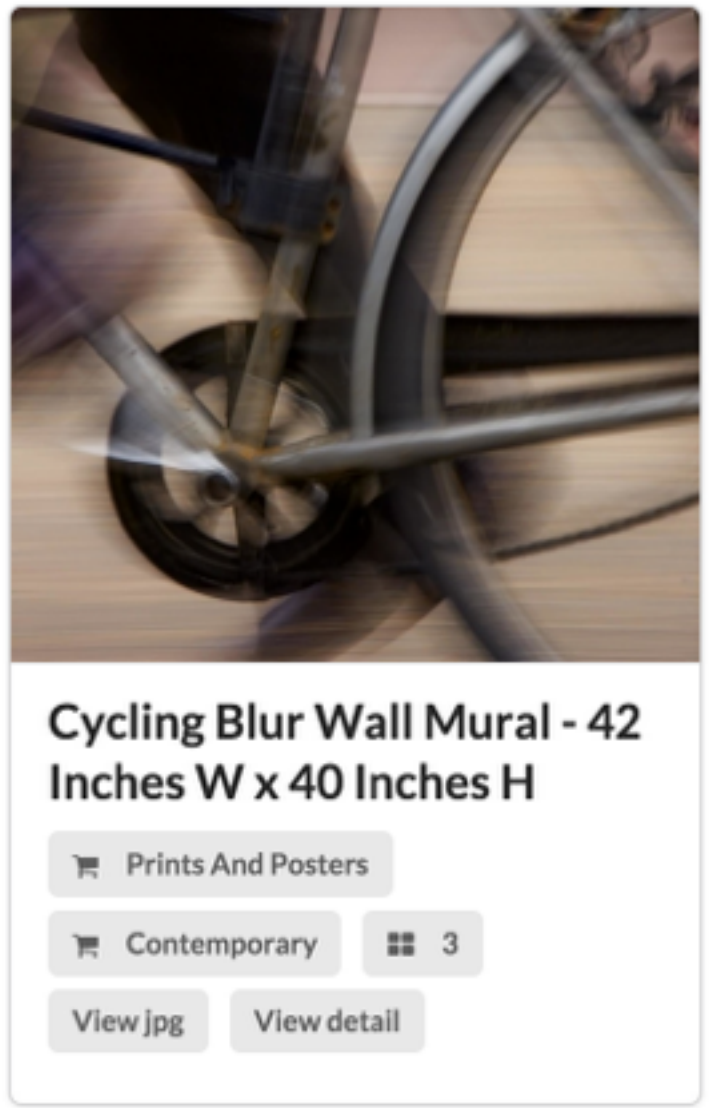
Top 4 results:



0.066



0.067



0.069

0.069

RESULTS: "WHERE IS IT USED?"

"Maskros Pendant Lamp"

RESULTS: "WHERE IS IT USED?"

"LEM Piston Stool | Design Within Reach"

SEARCHING ACROSS CATEGORIES

HOW DO WE EVALUATE THE RESULTS?

In the list of 3M products,
what is the **rank** of the iconic image?

In context

Iconic

HOW DO WE EVALUATE THE RESULTS?

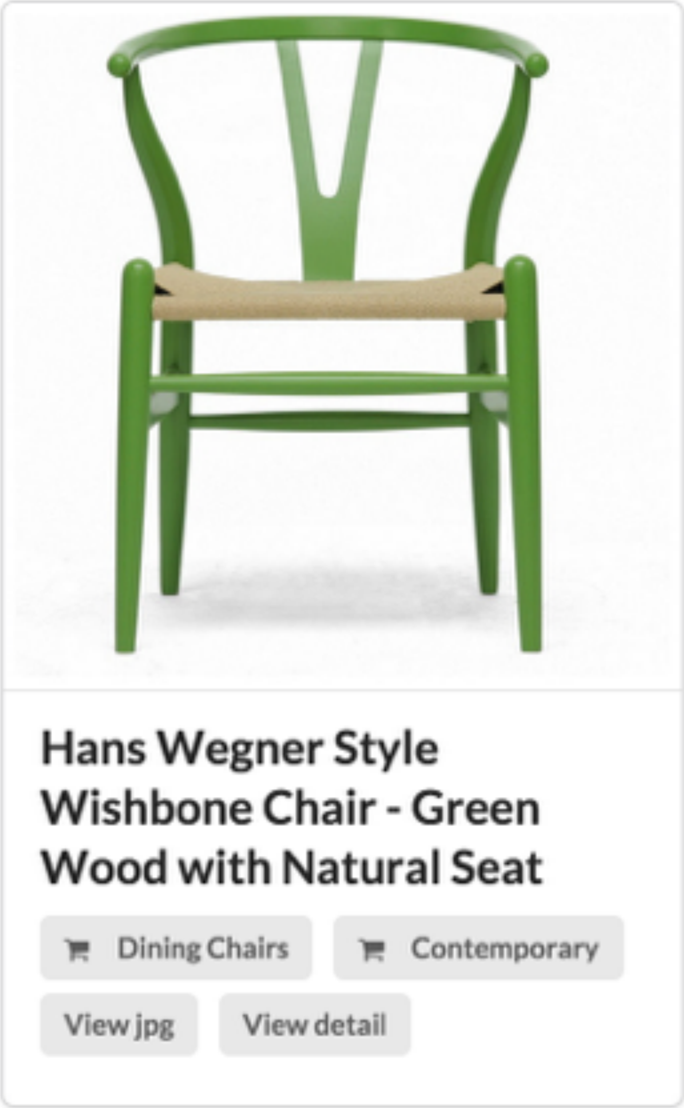
Query

Iconic

Top 4 results:

0.030

0.034



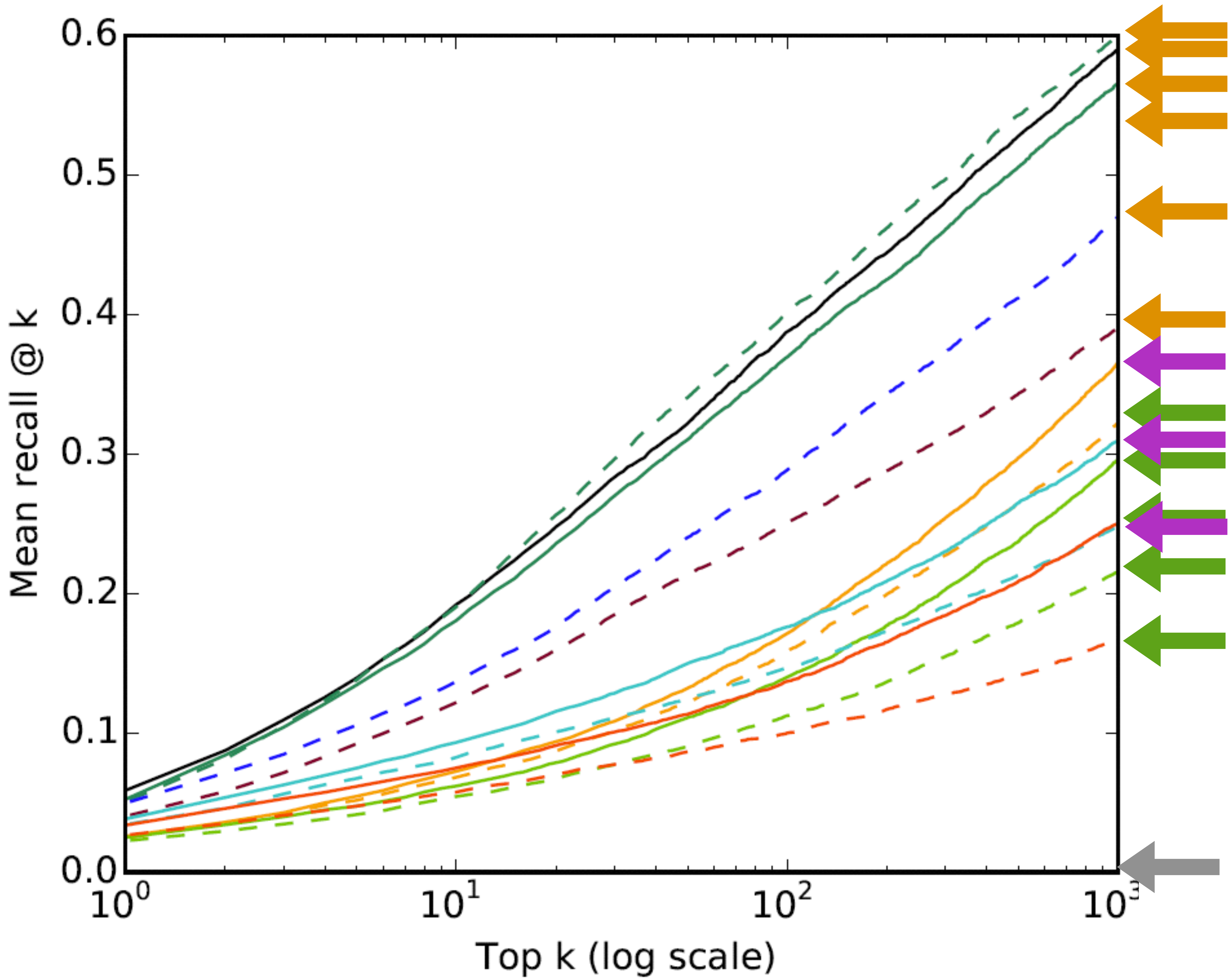
0.036

0.037

Rank: 440 / 3,387,555 (top 0.01%)

Even when the top product is correct, the rank may be large

QUANTITATIVE EVALUATION



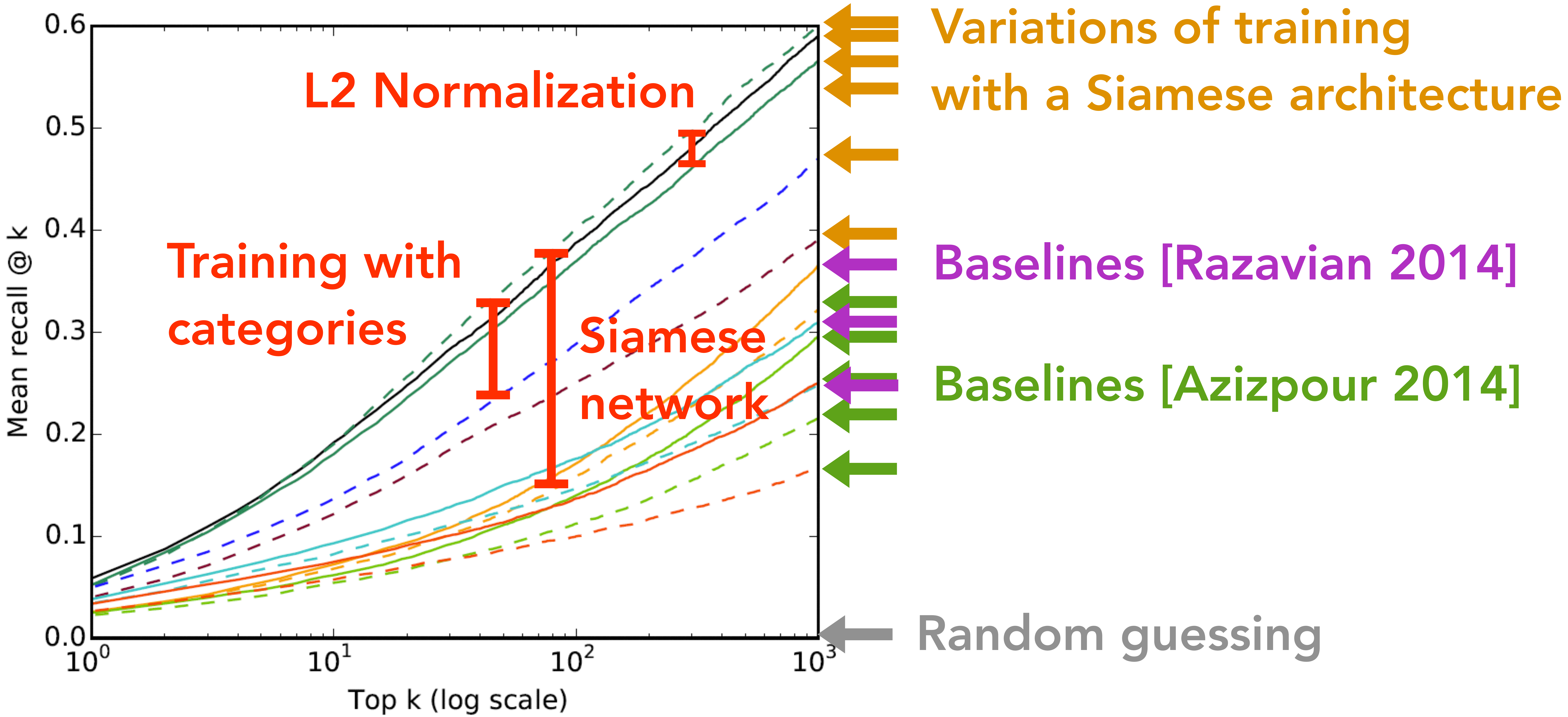
Variations of training with a Siamese architecture

Baselines [Razavian 2014]

Baselines [Azizpour 2014]

Random guessing

ABLATION STUDY



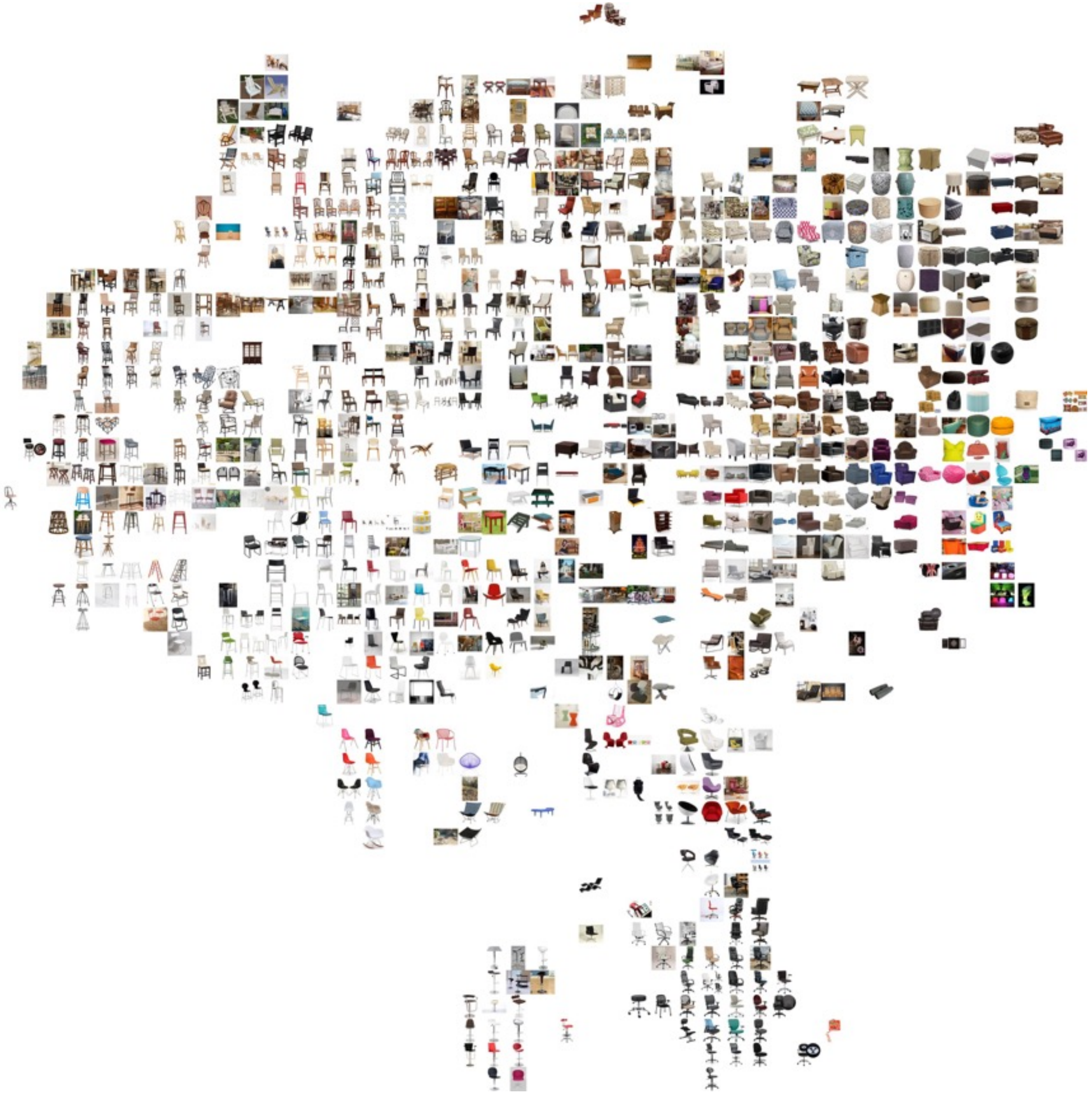
LIMITATIONS AND FUTURE WORK

- Better metrics to evaluate performance (e.g. measure precision)
- Efficiently collecting more training data (we are limited by what is tagged)
- Reduce crowdsourcing costs (less duplication)
- Explicitly model and train for style similarity/compatibility

CONTRIBUTIONS

Visual search for products:
"what is it?" and "where is it used?"

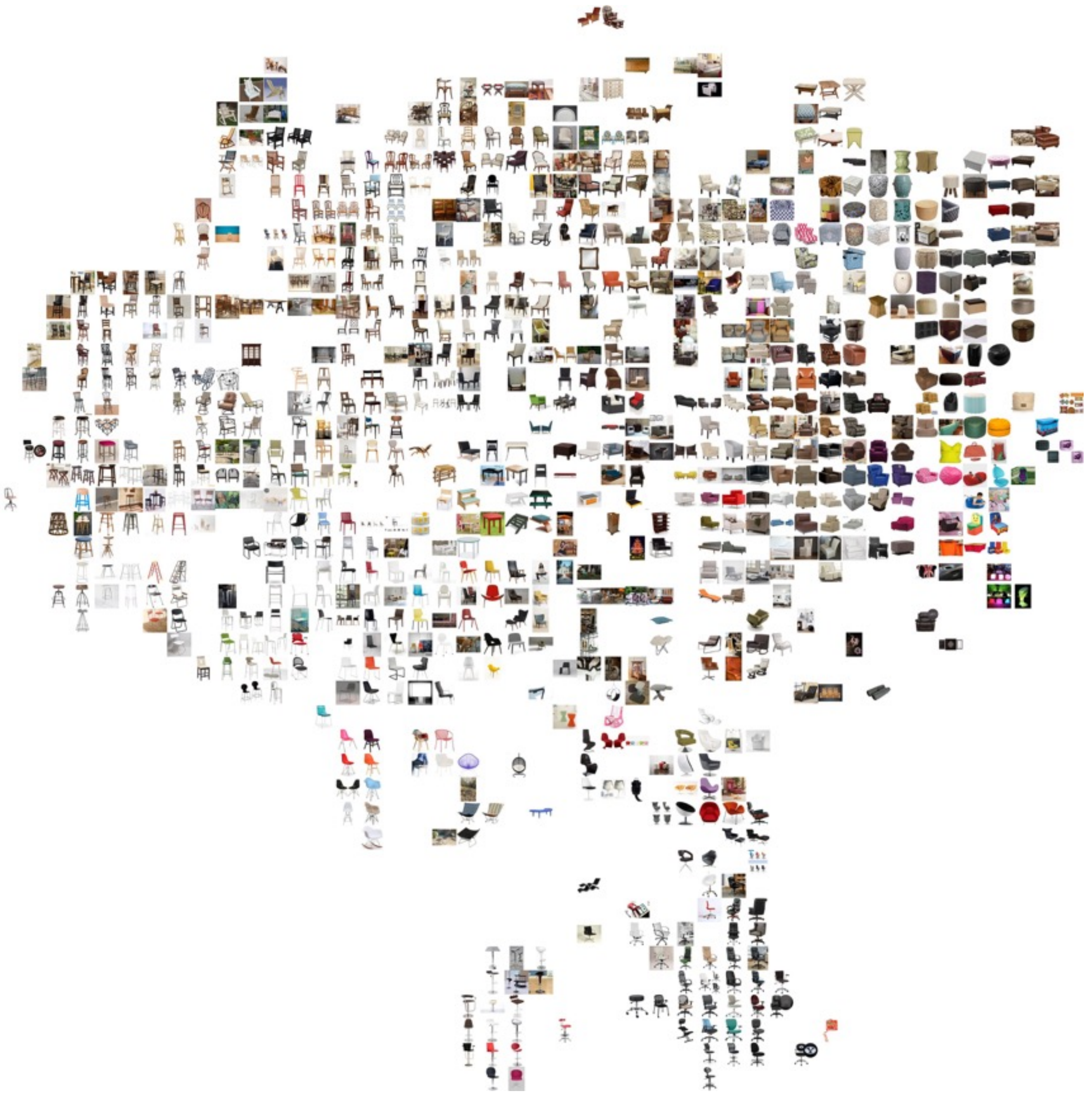
- Crowdsourced pipeline to collect image pairs across domains
- Use deep learning to learn a high quality embedding
- Applications in interior design and product search



THANK YOU

Acknowledgements:

- NSERC PGS-D
- National Science Foundation
- NVIDIA
- Intel Science and Technology Center for Visual Computing
- Amazon AWS for Education
- Houzz users

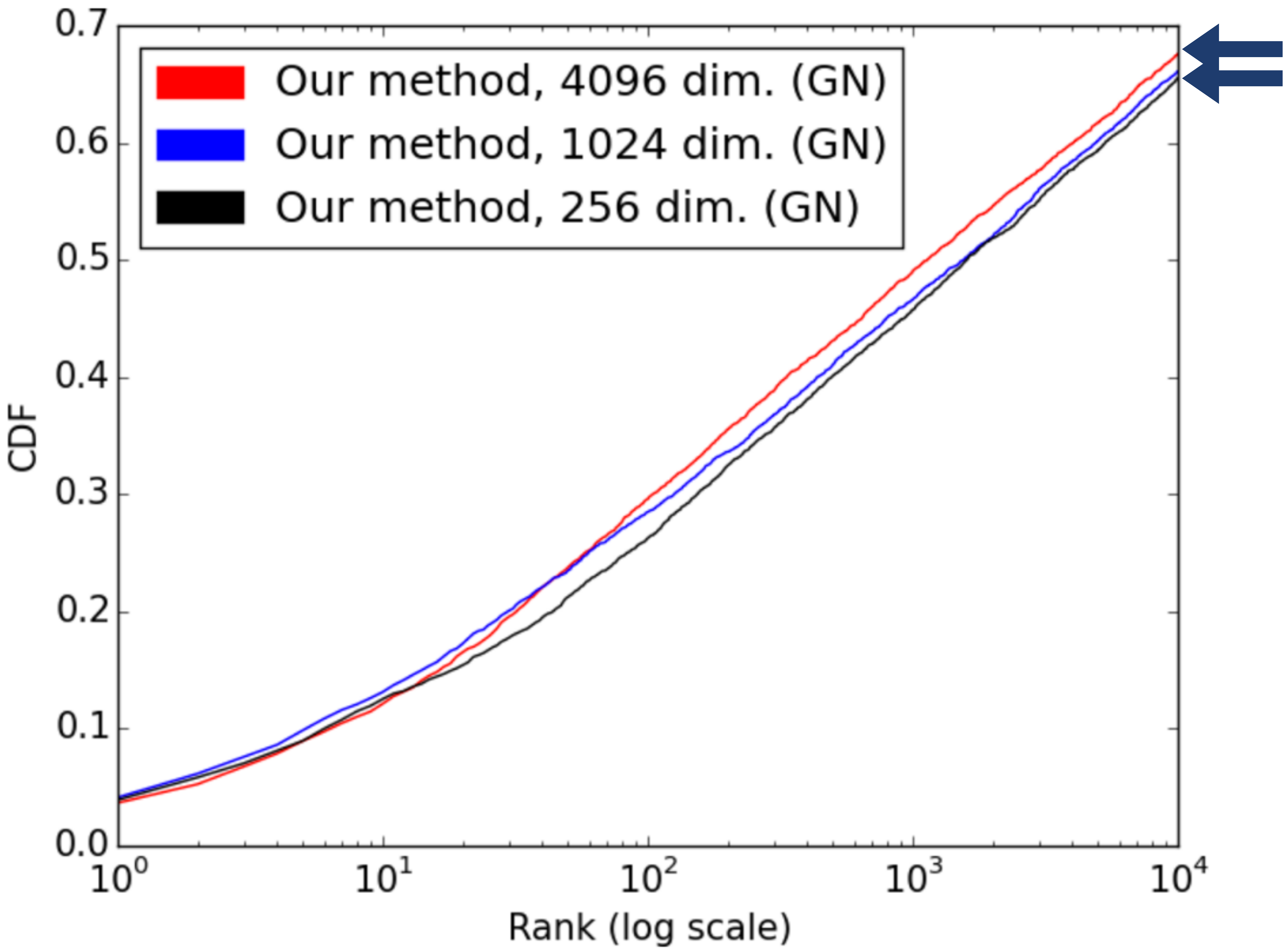


GROKSTYLE

We're looking for interns!

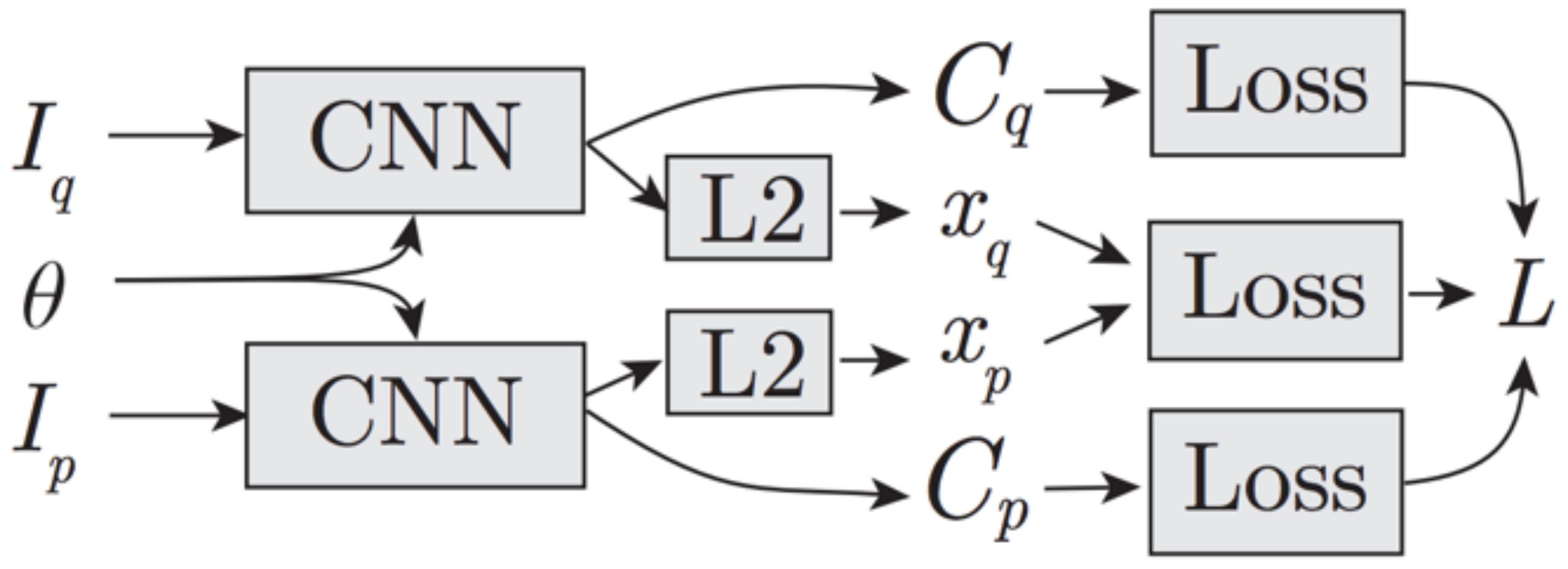
sbell@grokstyle.com

USING MORE DIMENSIONS PERFORMS THE SAME

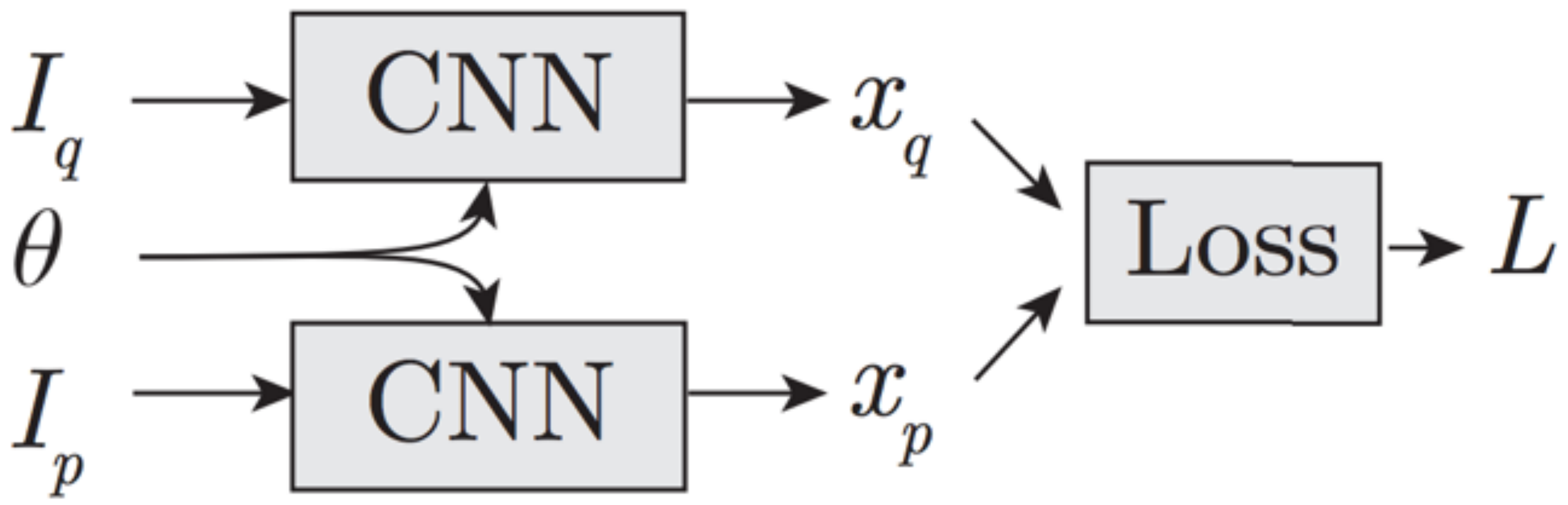


4096 dimensions
256 dimensions
(16x faster for almost the same performance)

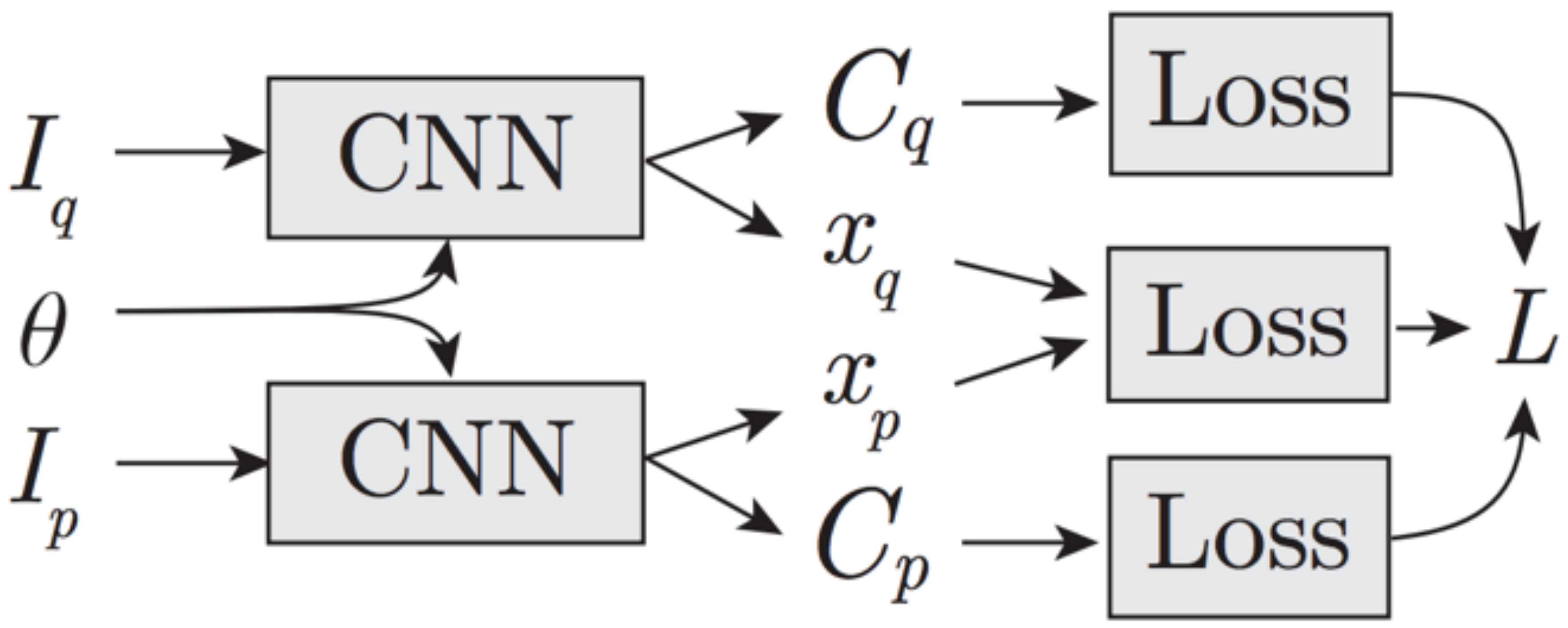
ARCHITECTURE VARIATIONS



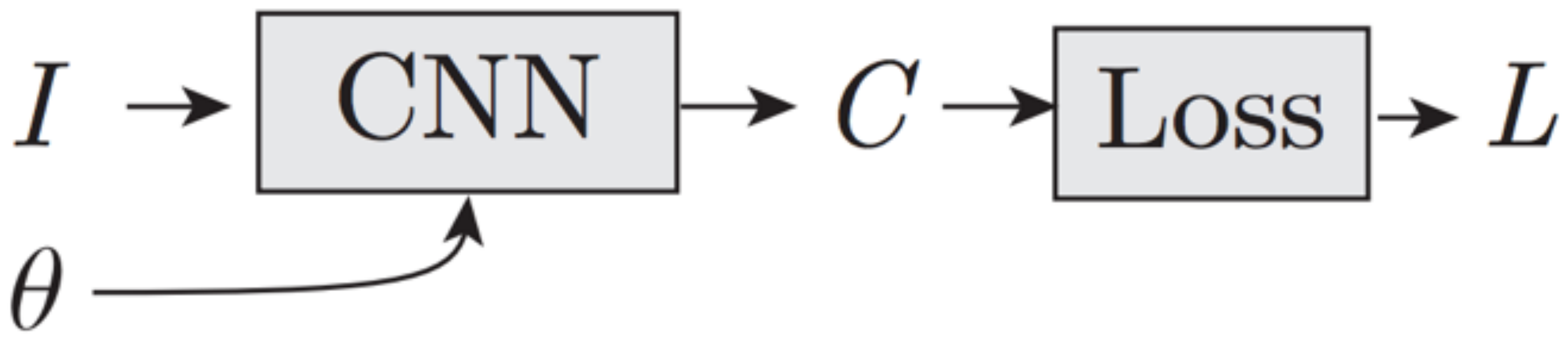
(Our system)



[Chropa 2005, Hadsell 2006, …]



[Weston 2008]



[Razavian 2014, …]