3D-Assisted Image Feature Synthesis for Novel Views of an Object

Hao Su*
Fan Wang*
Li Yi
Leonidas Guibas

* Equal contribution

Stanford University

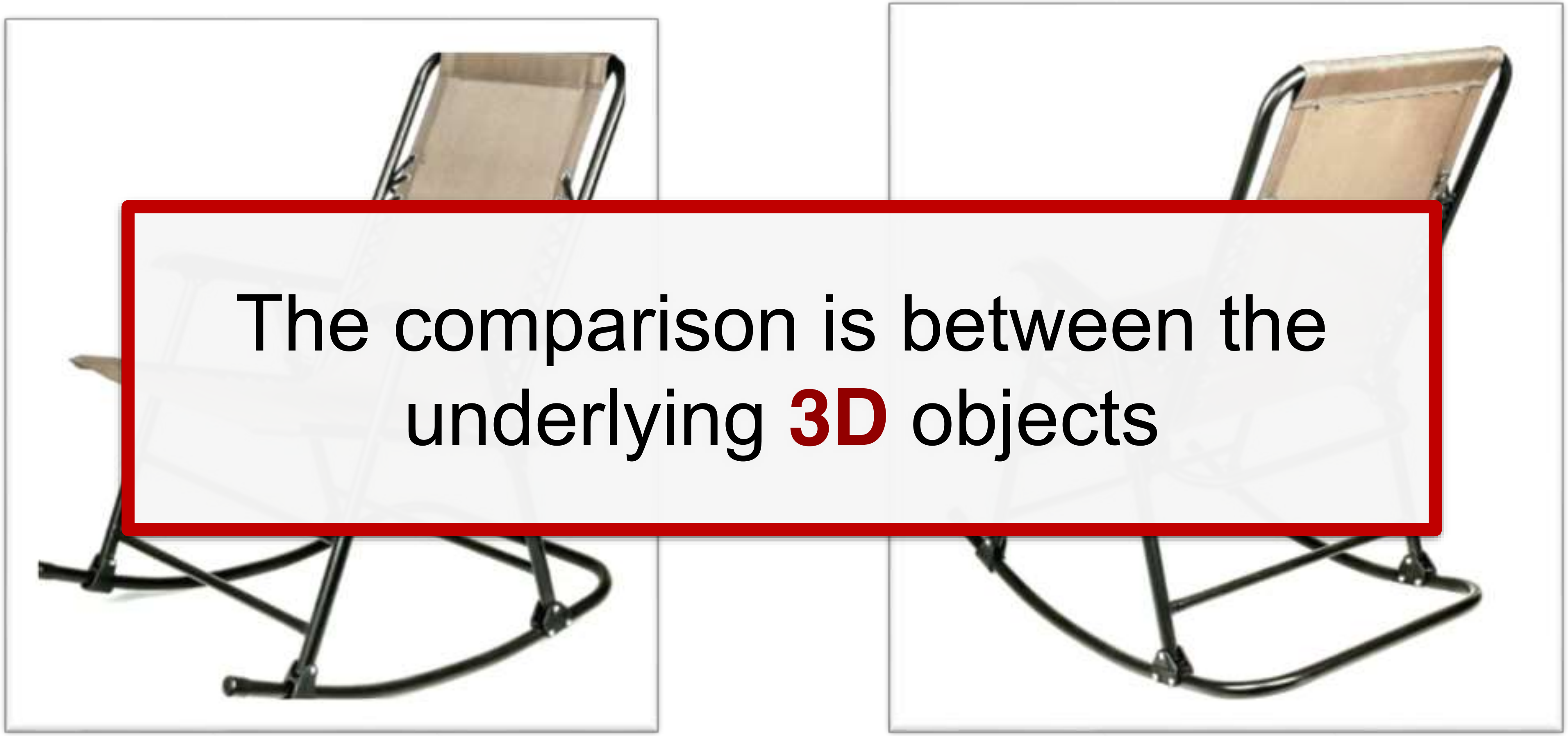
View-agnostic Image Retrieval

Retrieval using AlexNet features

Query

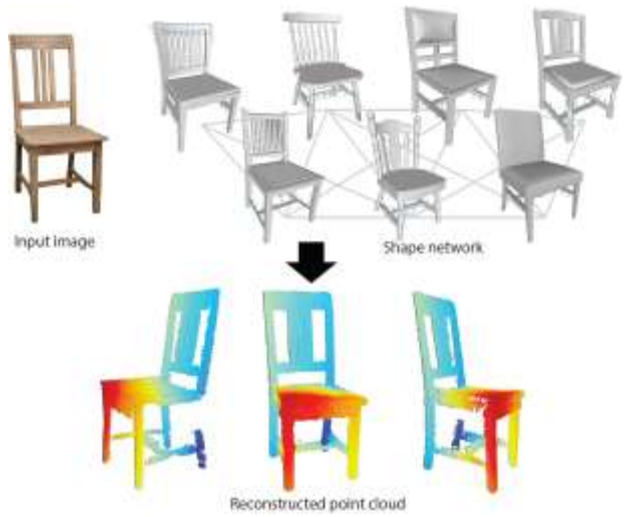
Cross-view Image Comparison

Cross-view Image Comparison

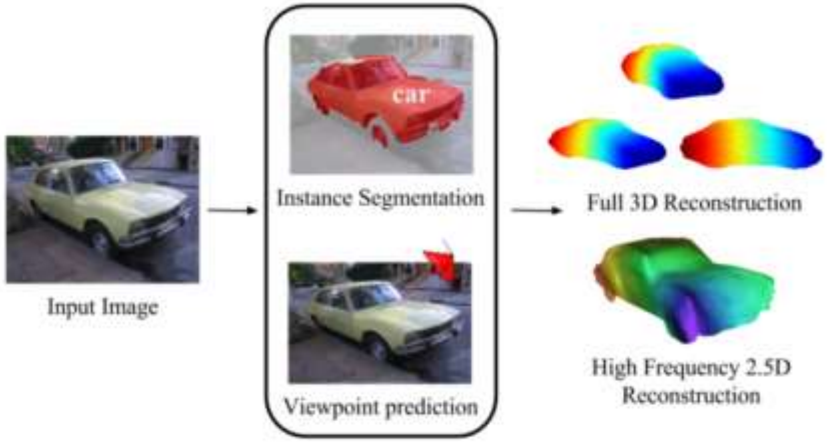


The comparison is between the underlying **3D** objects

Reconstruct 3D and then compare?



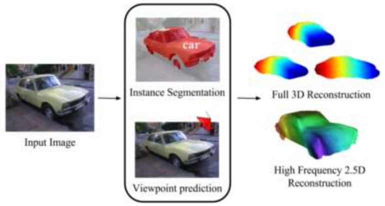
Su et al, SIGGRAPH'14



Kar et al, CVPR'15

Huang et al, SIGGRAPH'15

Single-image based 3D Reconstruction is hard

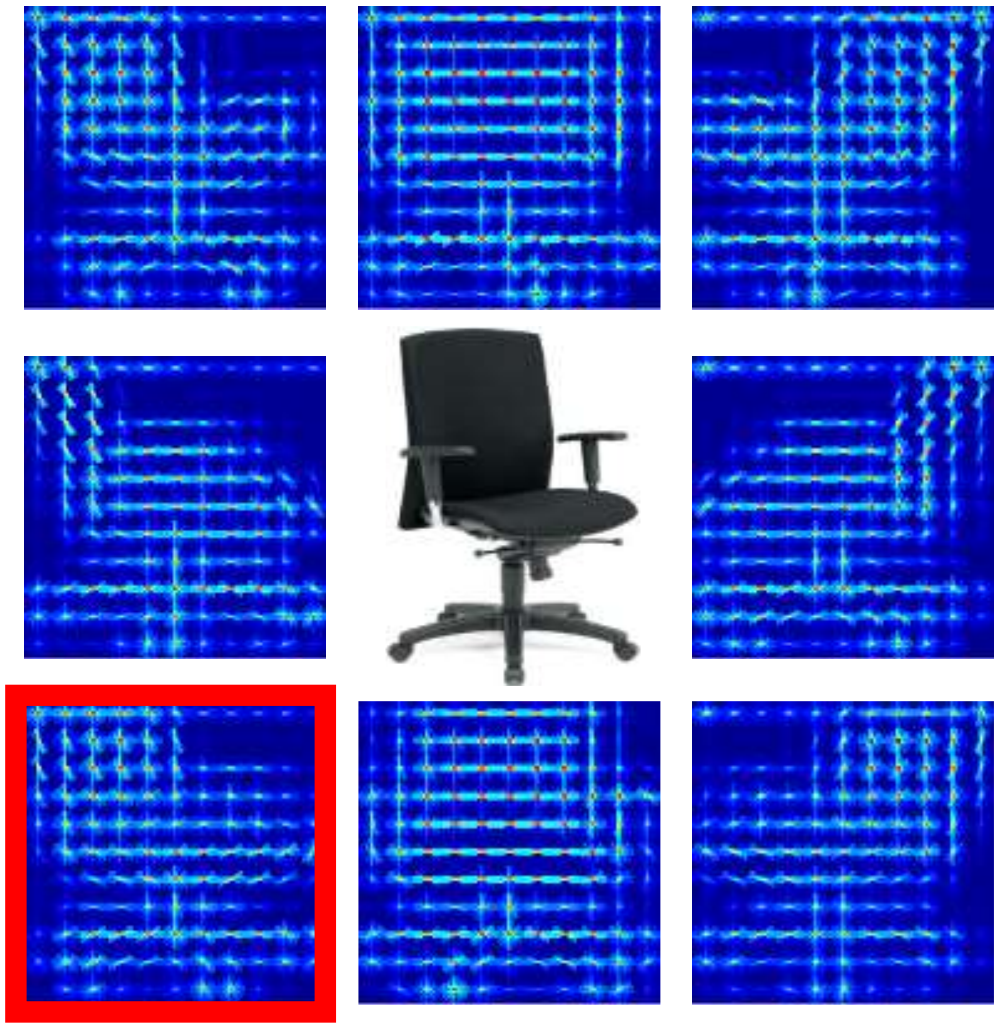


Common dependencies:

Many dependencies
Not Robust
Slow

Fg/bg segmentation	*Keypoint detection*
2D image part segmentation	*2D-3D Correspondence*
3D shape part segmentation	*Non-convex iterative optimization*

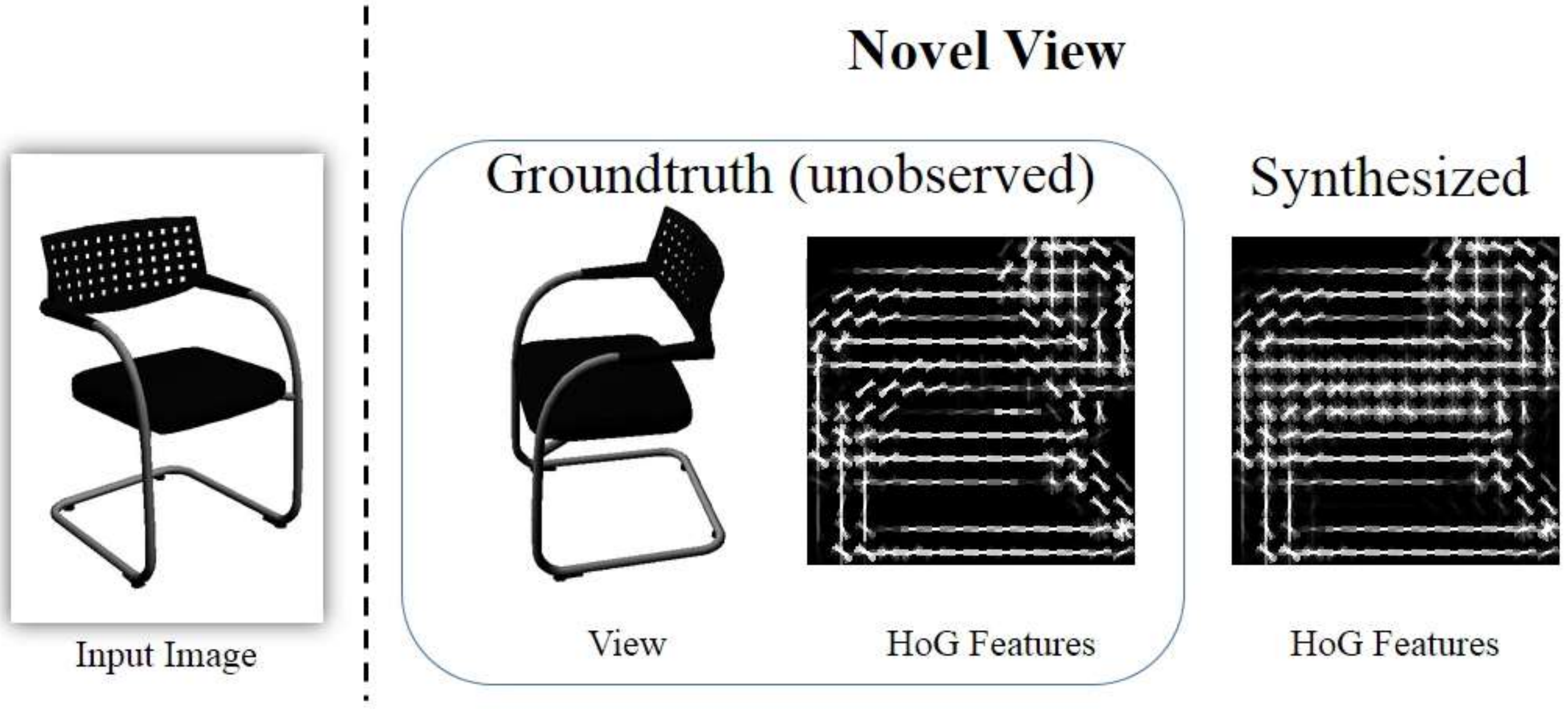
Our Formulation: Novel View Feature Synthesis



Observed view

(HoG feature as an example)

Our Novel View Feature Synthesis Results



(HoG feature as an example)

Outline

Motivation

Approach

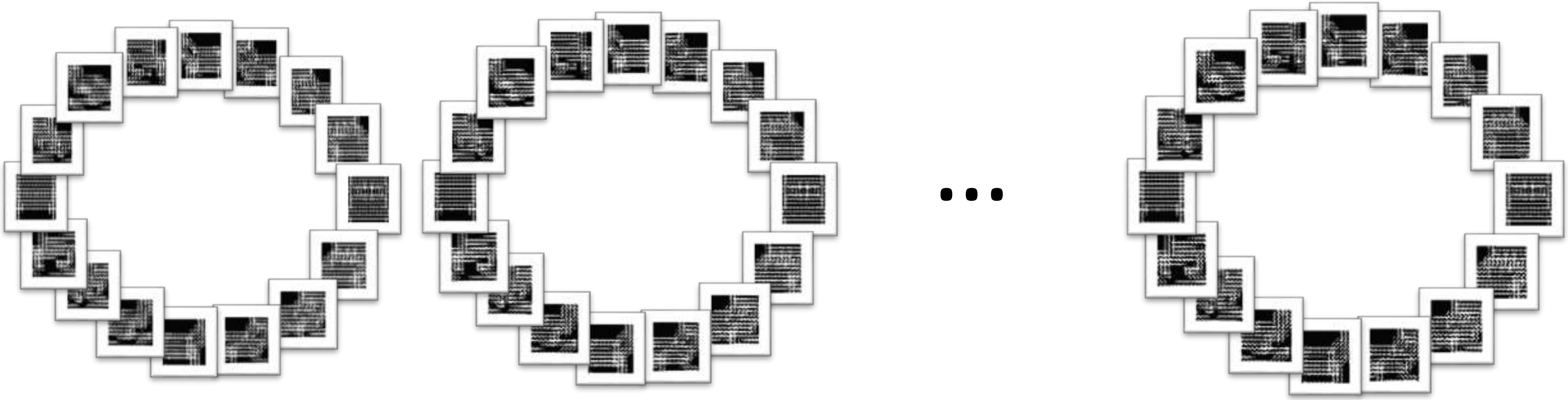
Applications

Method Diagnosis

Conclusion

Key idea

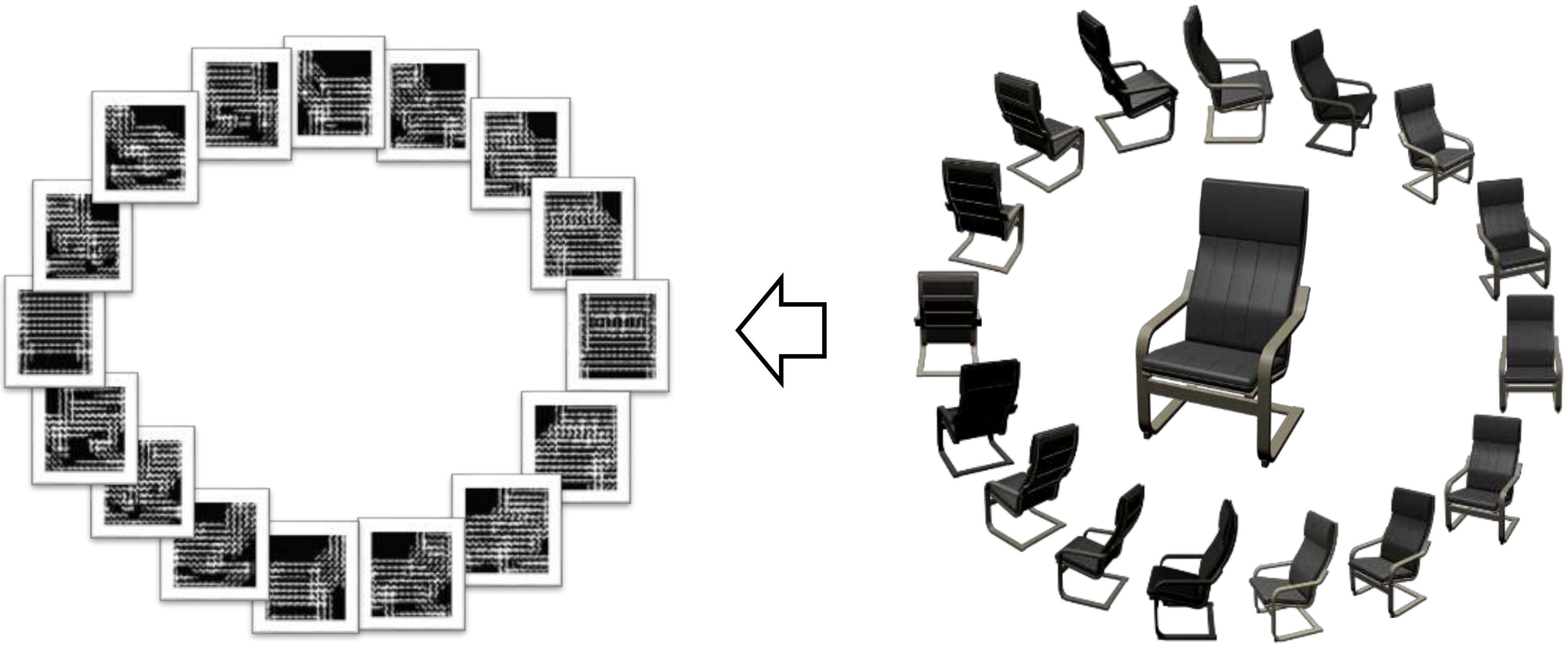
Learn from a dataset of many objects with multi-view features



Key idea

Learn from a dataset of multi-view features

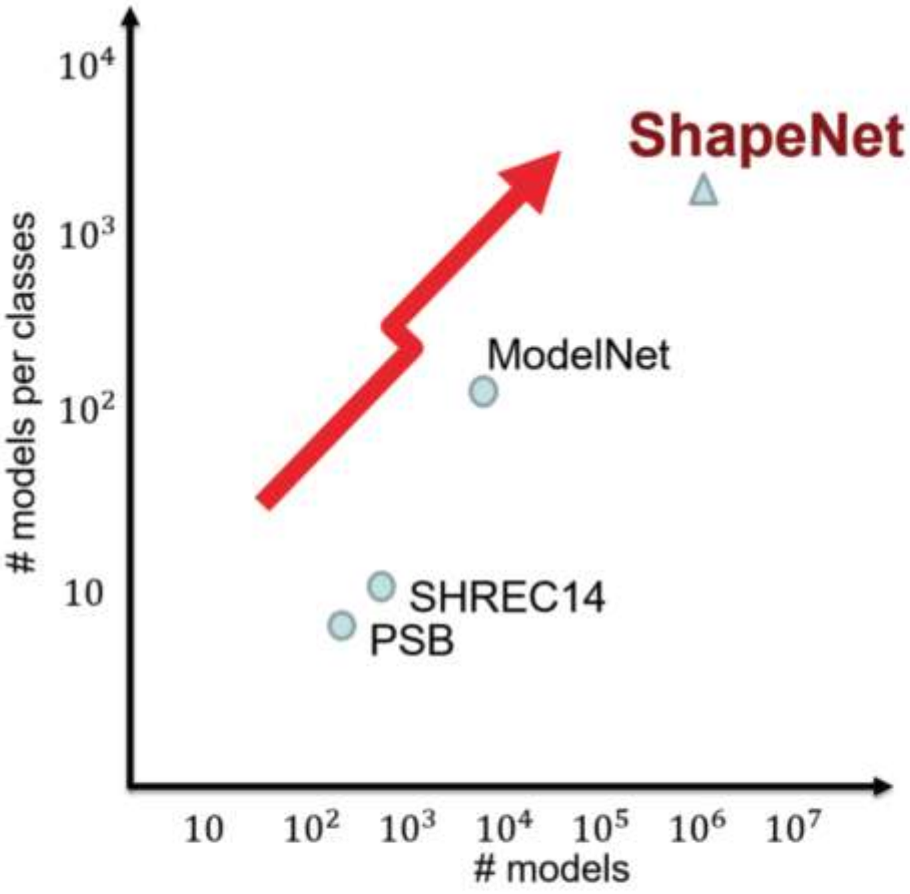
The dataset is generated by rendering **3D models**



Key idea

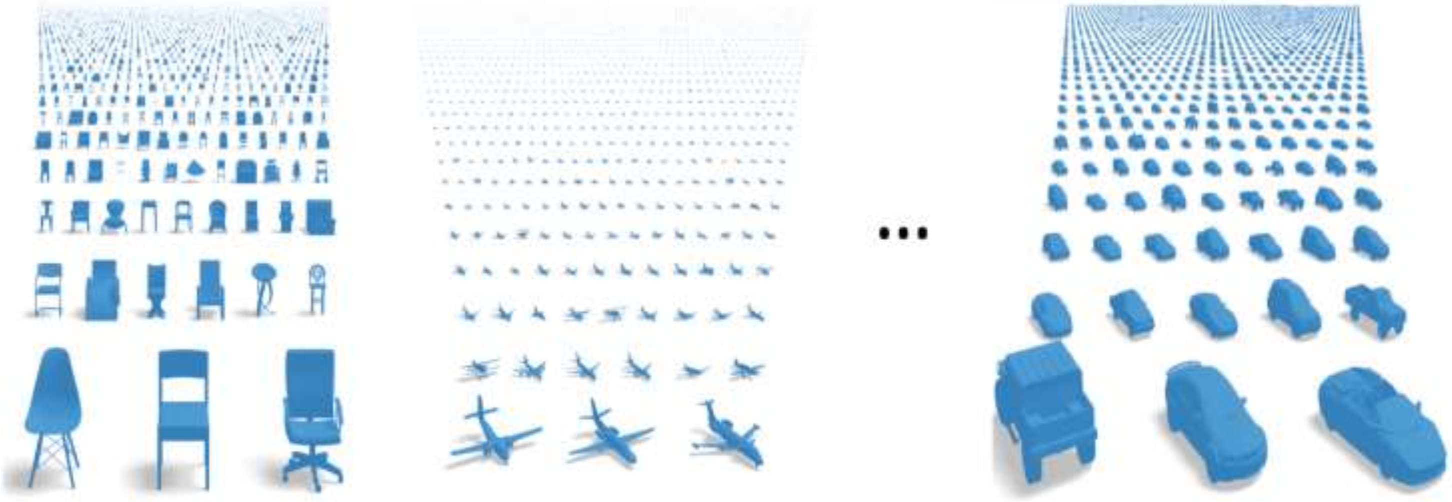
Learn from a dataset of multi-view features

The dataset is generated by rendering **large-scale 3D models**

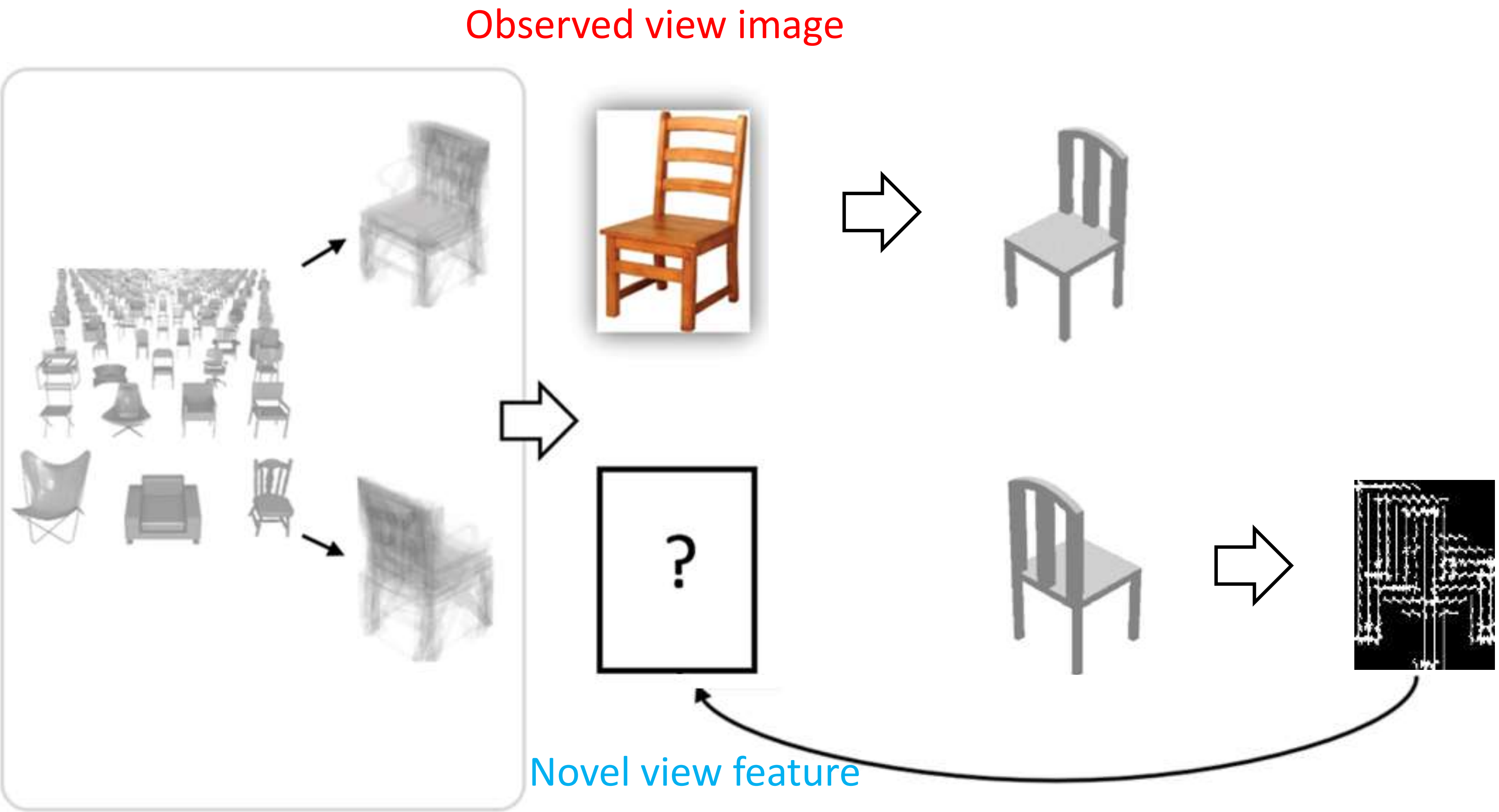


SHAPENET

http://shapenet.cs.stanford.edu

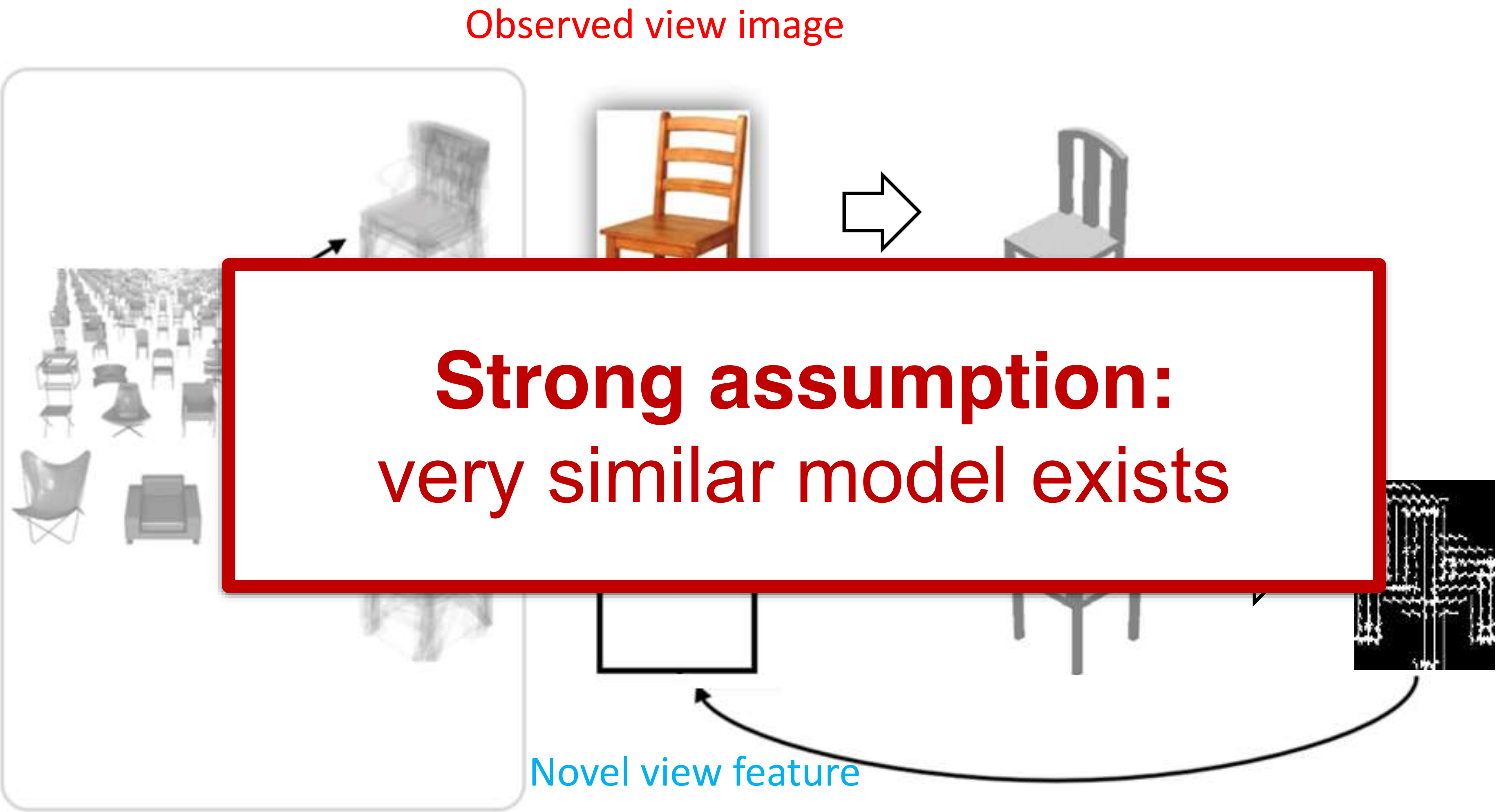


3D-assisted Feature Synthesis: Nearest Neighbour



(HoG feature as an example)

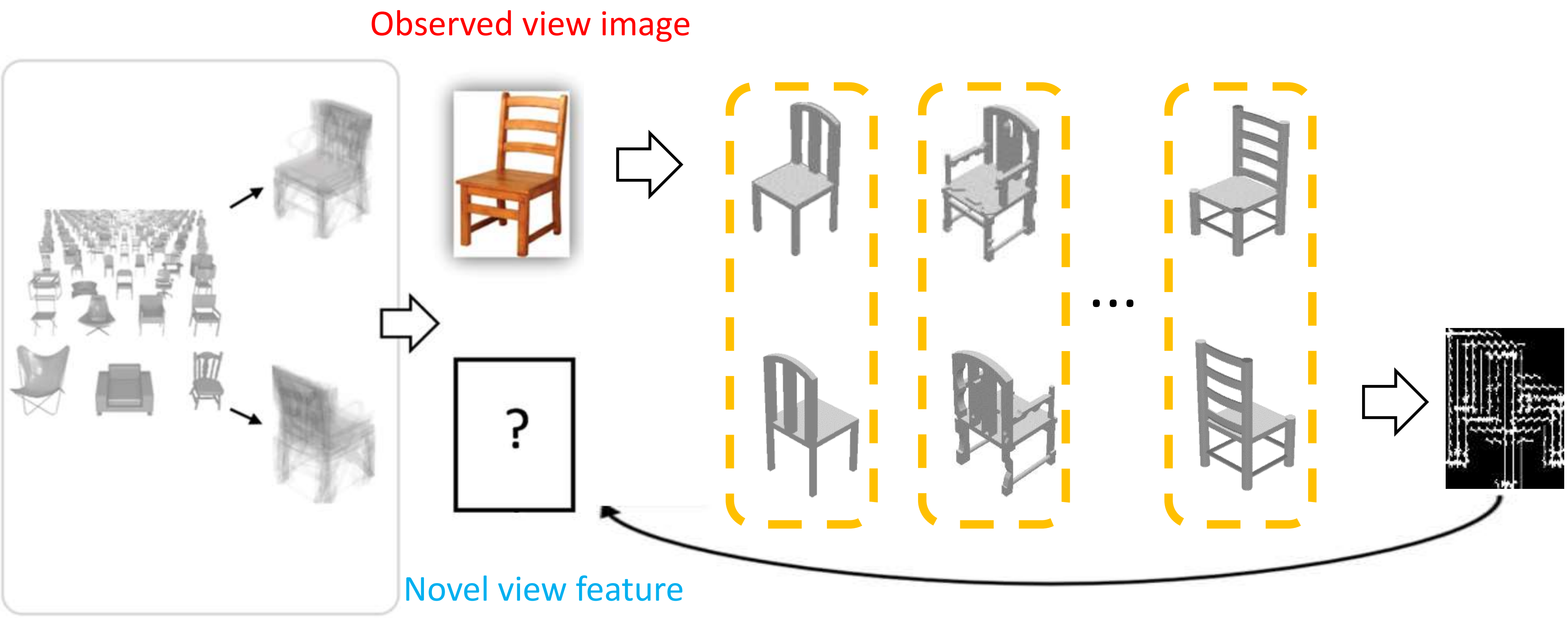
3D-assisted Feature Synthesis: Nearest Neighbour



Strong assumption:
very similar model exists

(HoG feature as an example)

3D-assisted Feature Synthesis: Multiple Shapes



(HoG feature as an example)

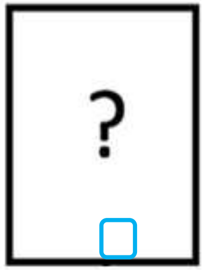
3D-assisted Feature Synthesis: Multiple Shapes

Attention:

Brain games start!

Pipeline

Observed view image

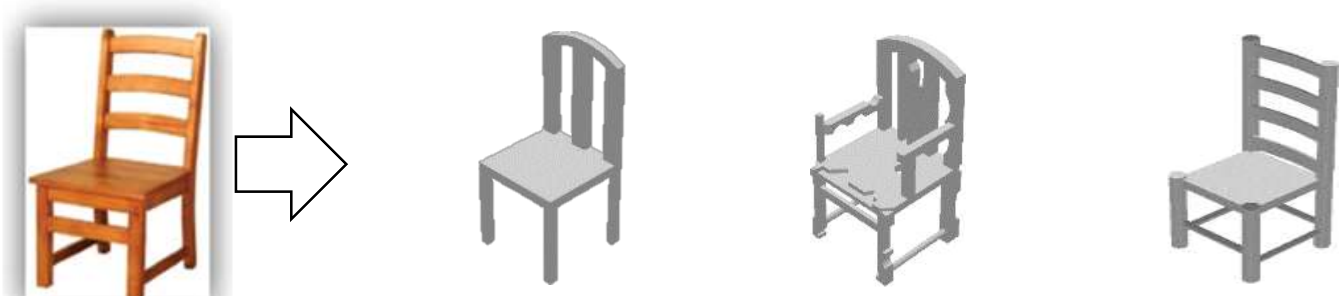


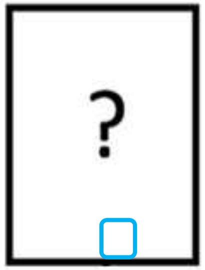
Novel view feature

(HoG feature as an example)

Pipeline

Observed view image



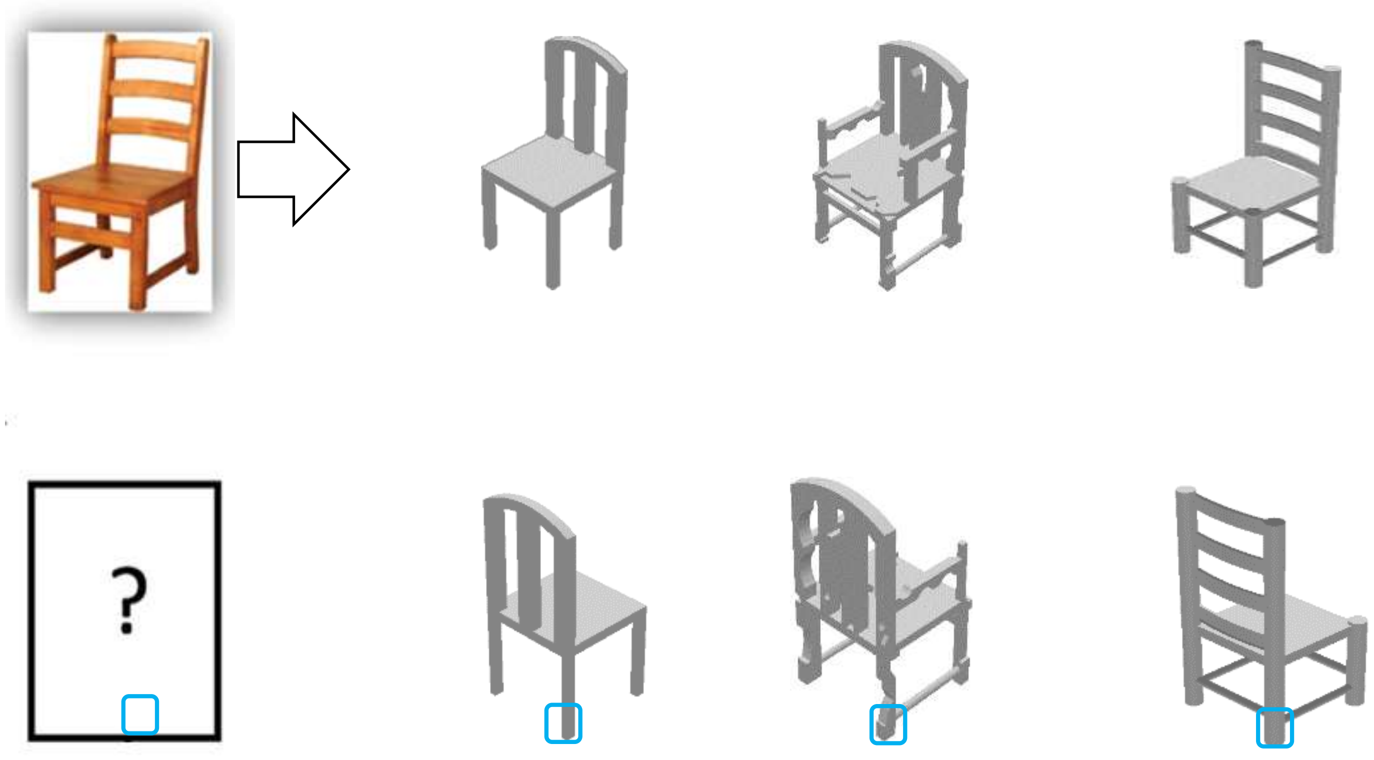


Novel view feature

(HoG feature as an example)

Pipeline

Observed view image

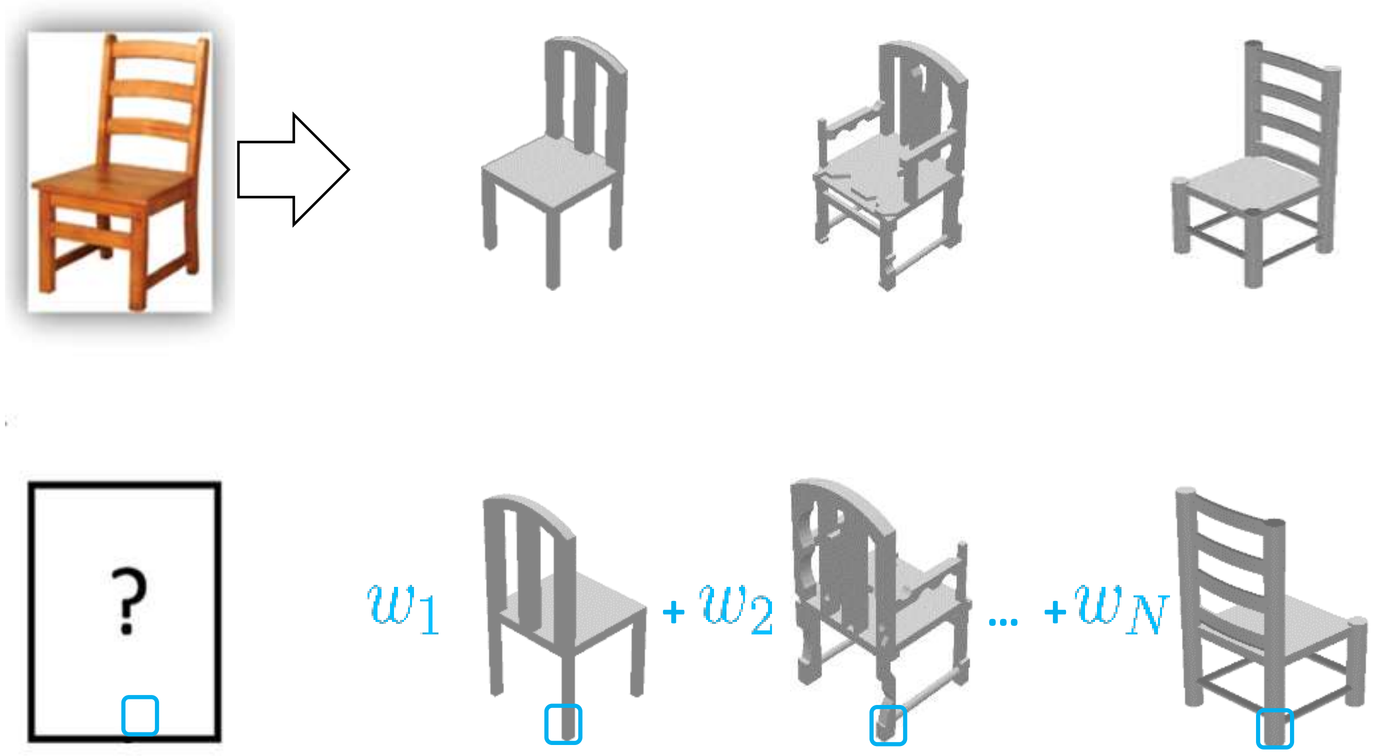


Novel view feature

(HoG feature as an example)

Pipeline

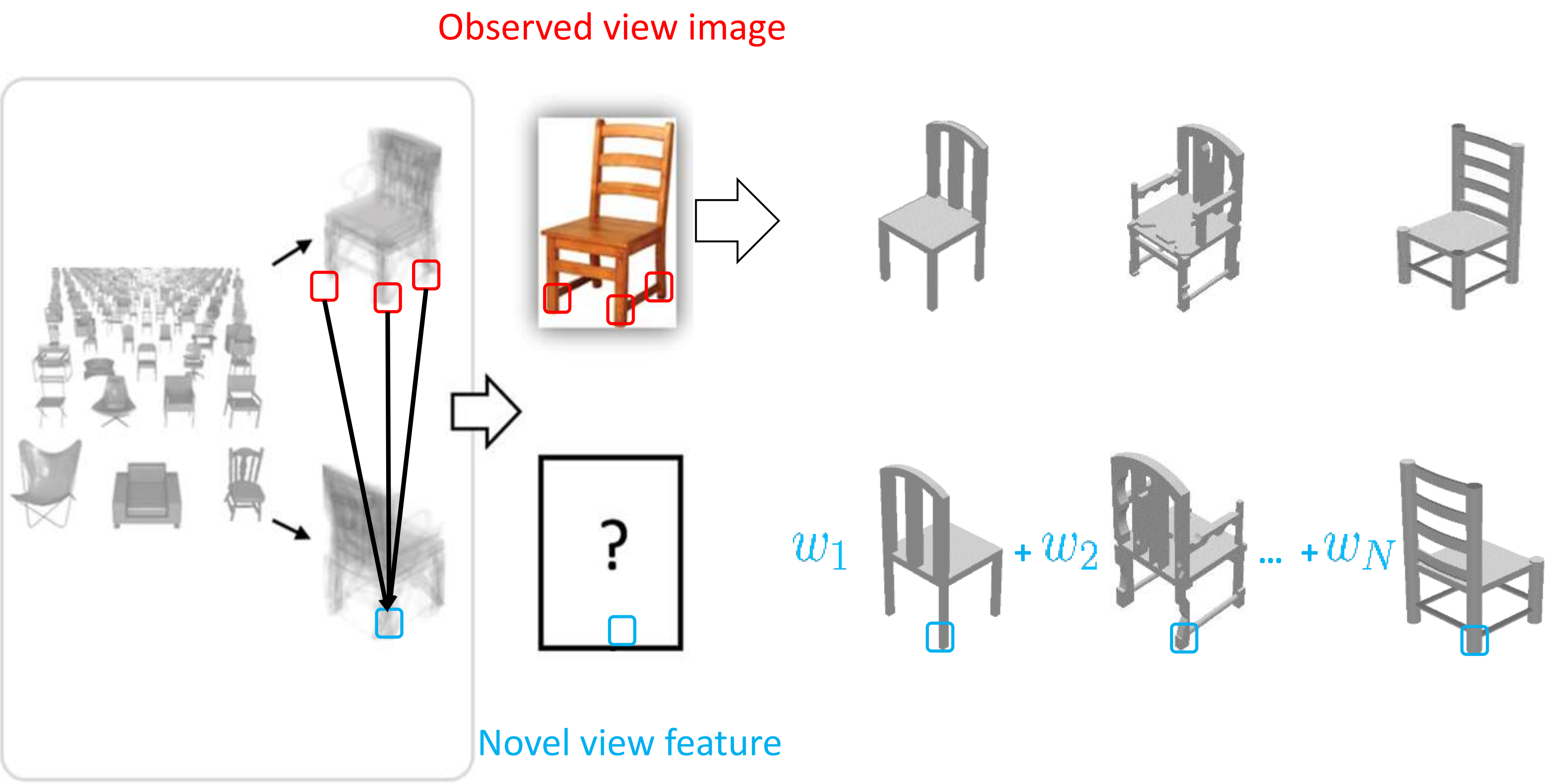
Observed view image



Novel view feature

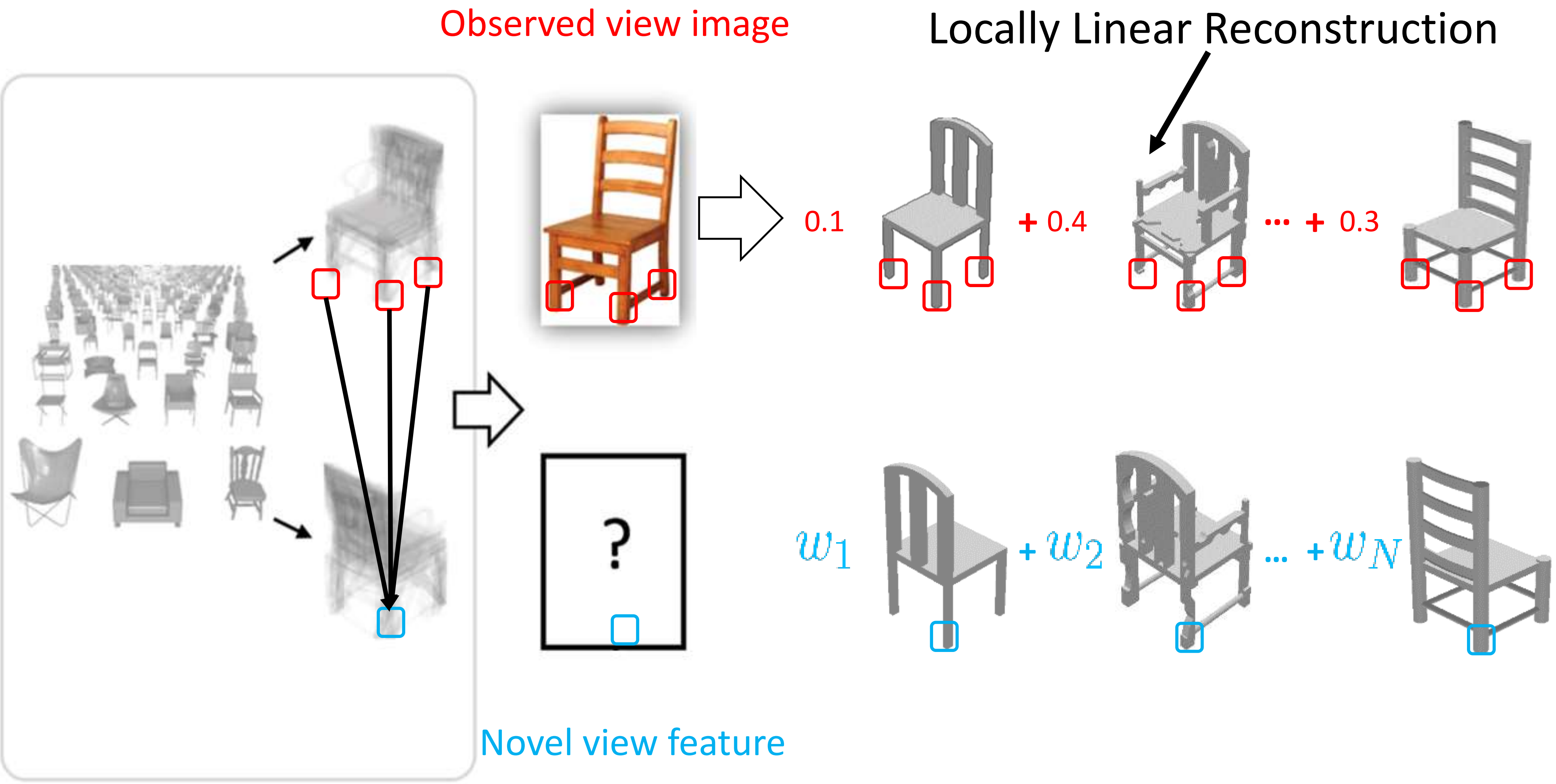
(HoG feature as an example)

Pipeline



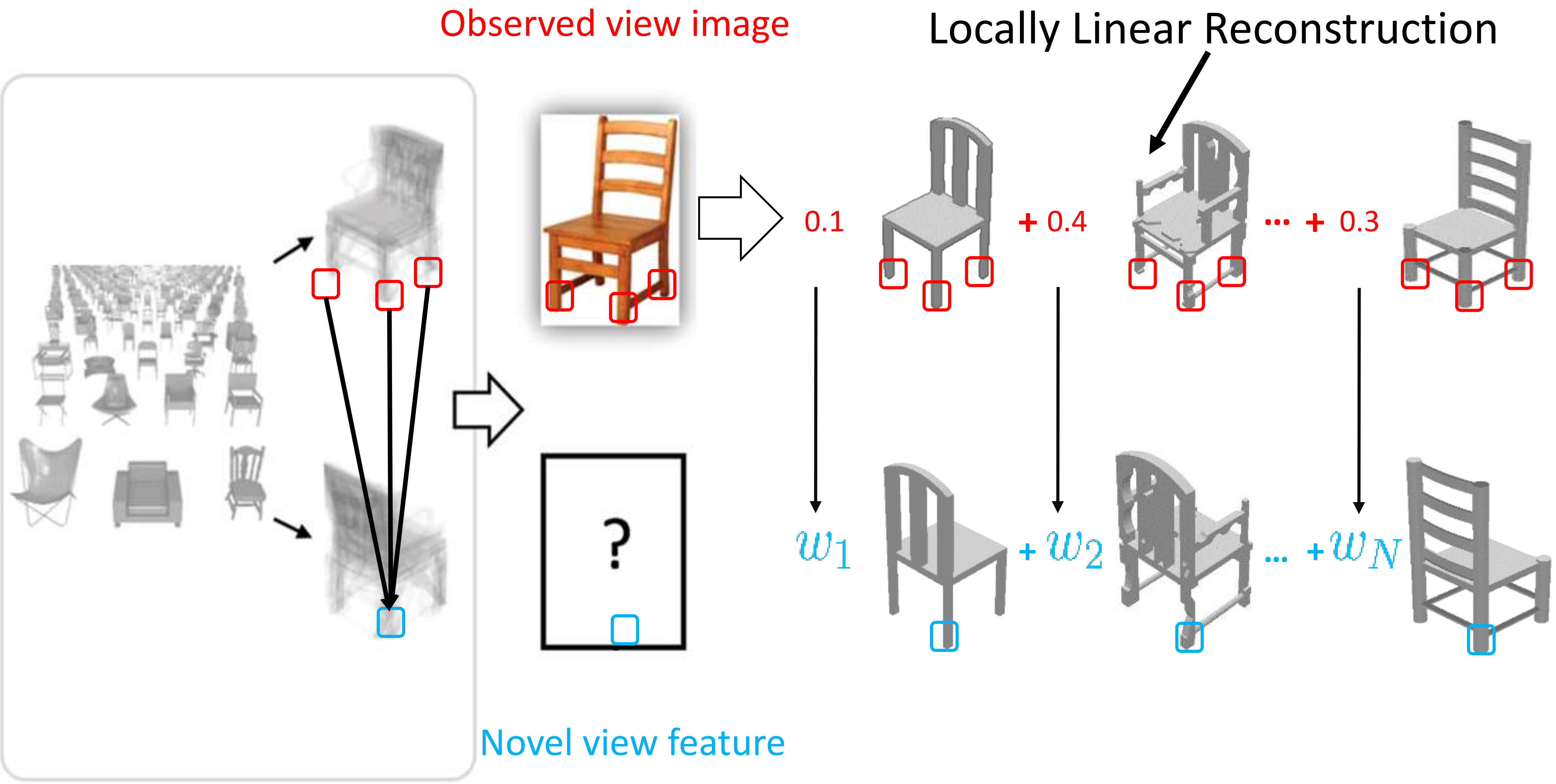
(HoG feature as an example)

Pipeline



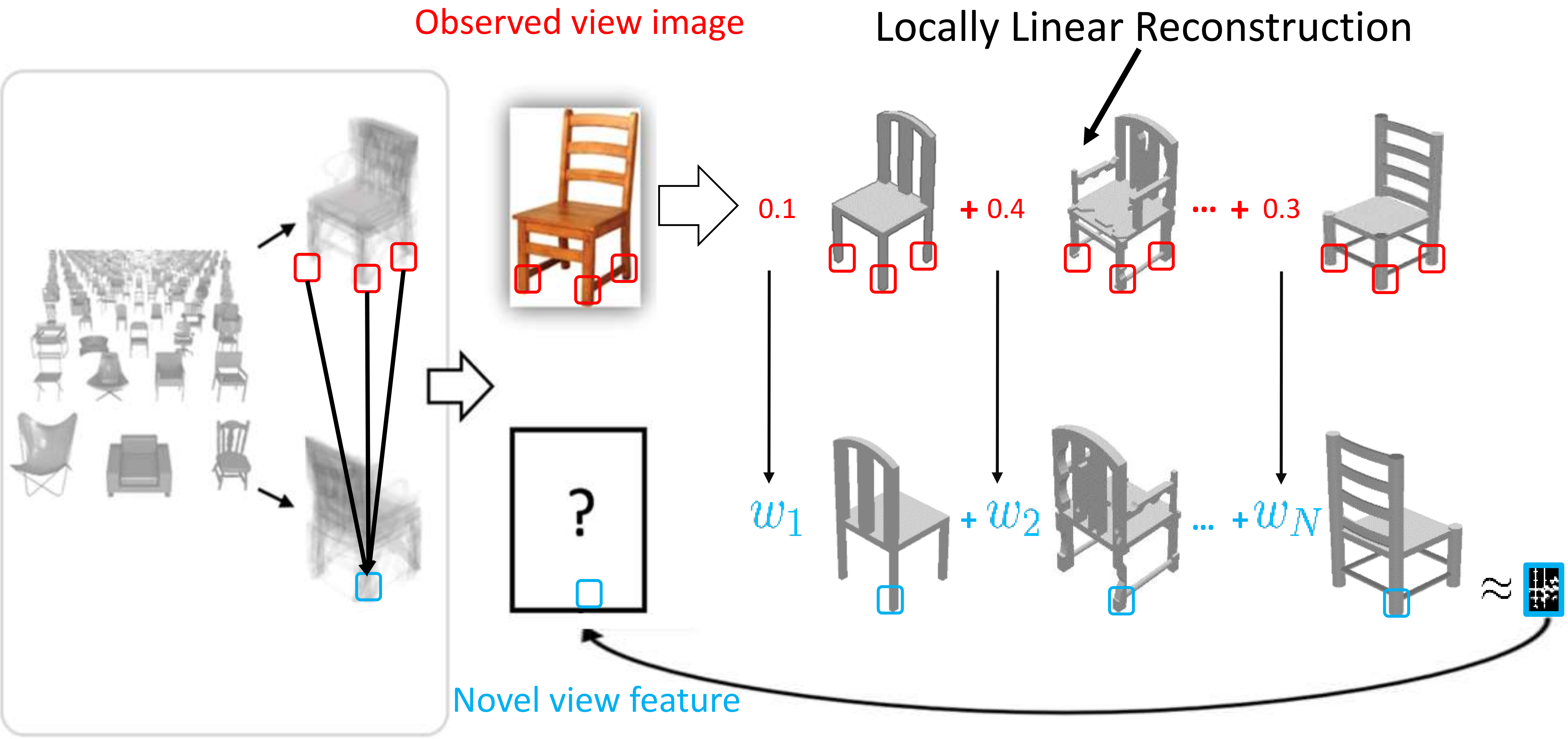
(HoG feature as an example)

Pipeline



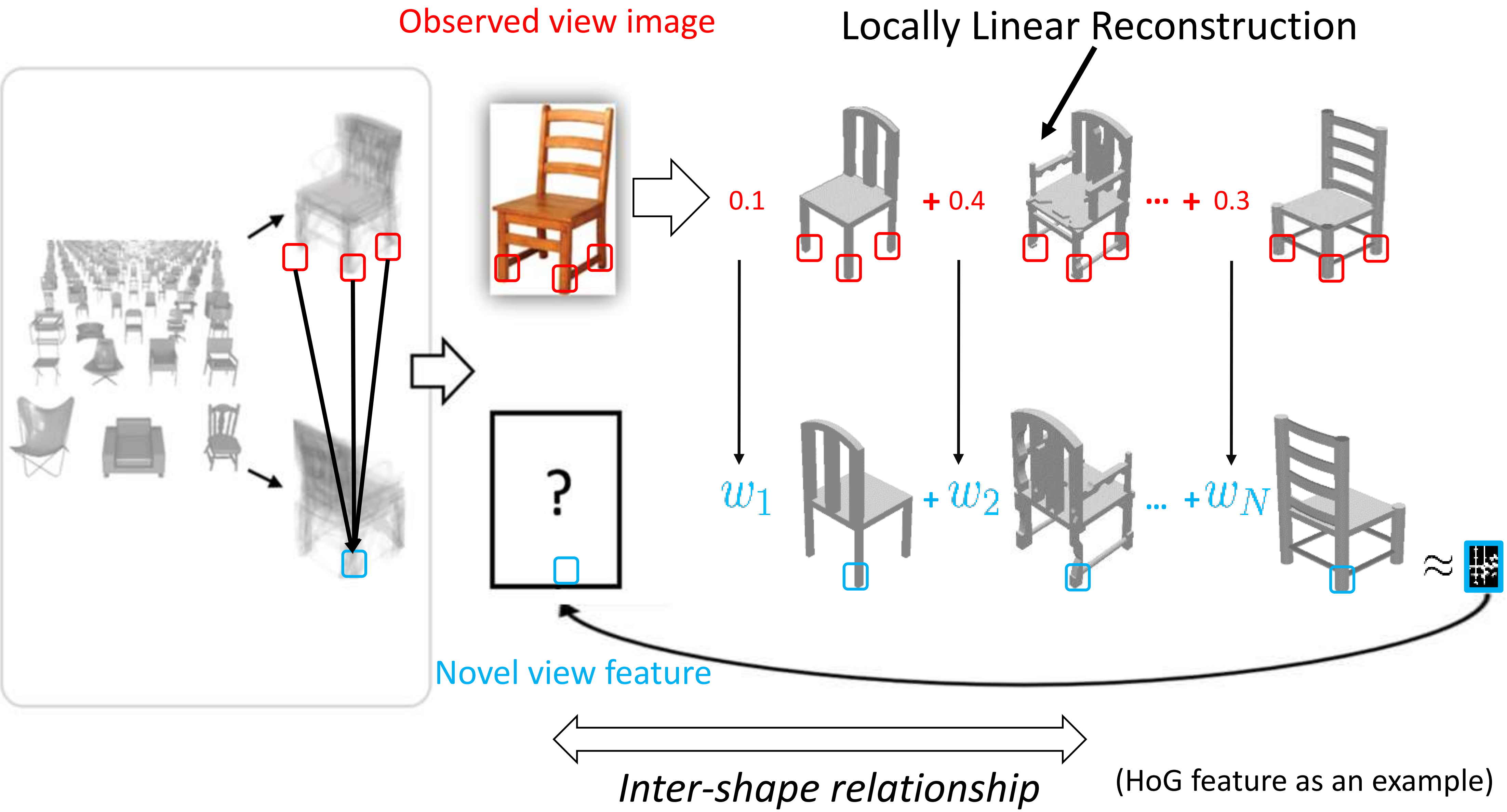
(HoG feature as an example)

Pipeline

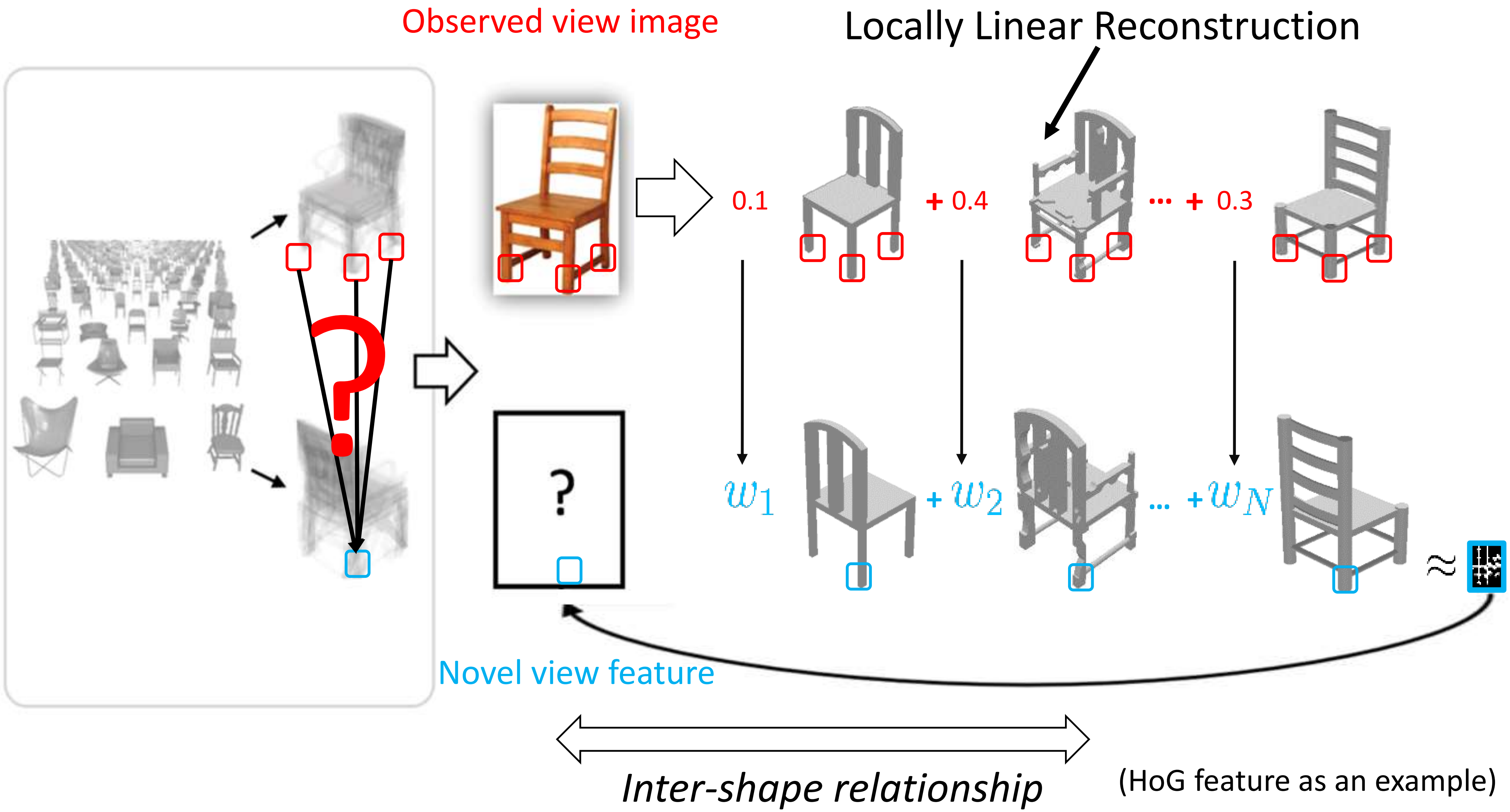


(HoG feature as an example)

Pipeline


Observed view image
Locally Linear Reconstruction
0.1
+ 0.4
... + 0.3
w_1
+ w_2
... + w_N
≈
?
Novel view feature
Inter-shape relationship
(HoG feature as an example)

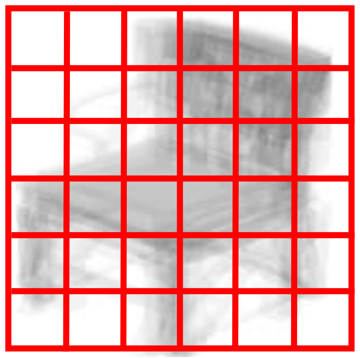
Surrogate Relationship Discovery



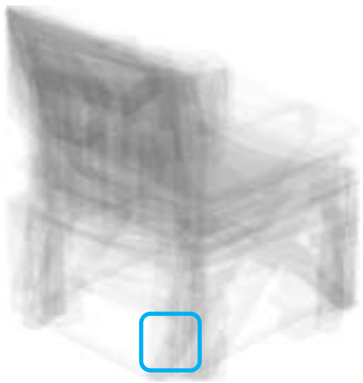
Surrogate Relationship Discovery

Shape Collection

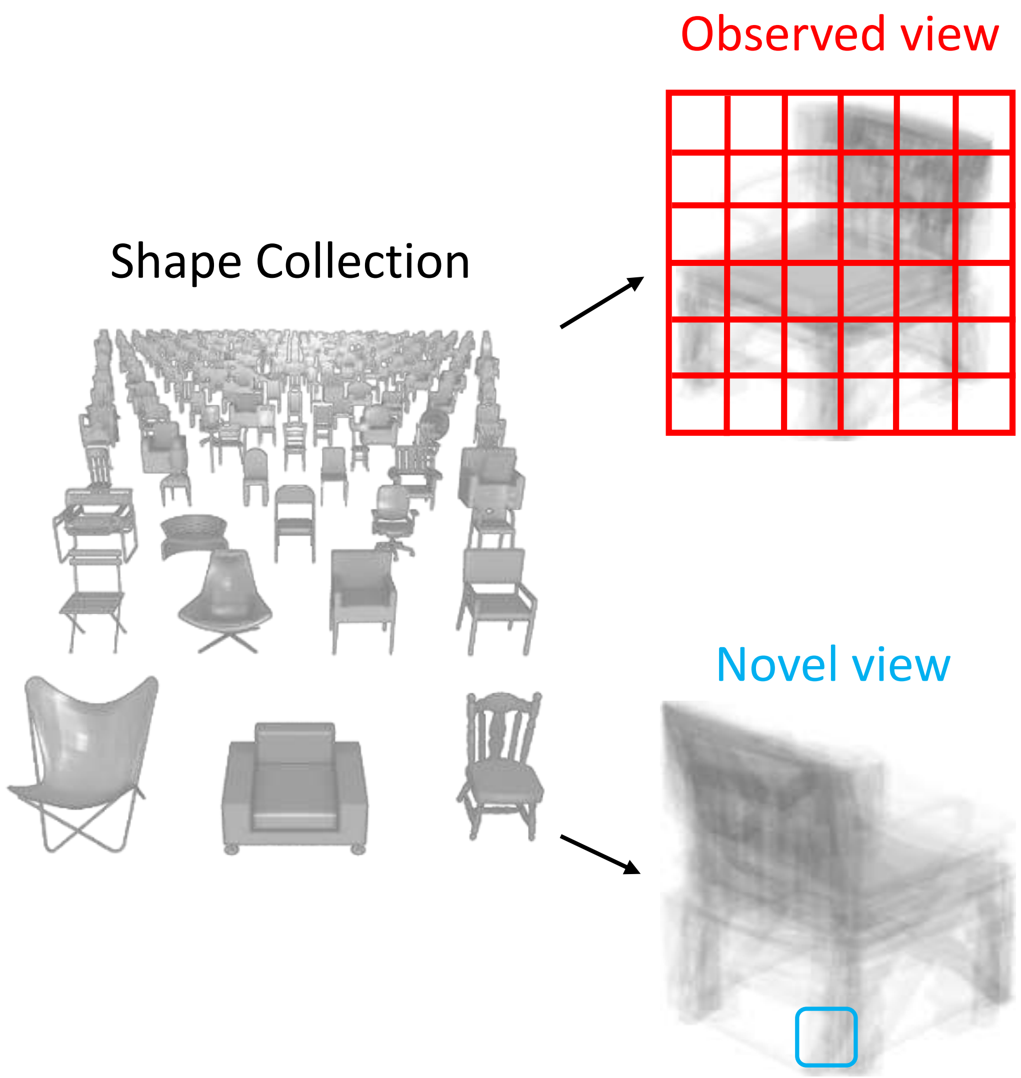
Observed view

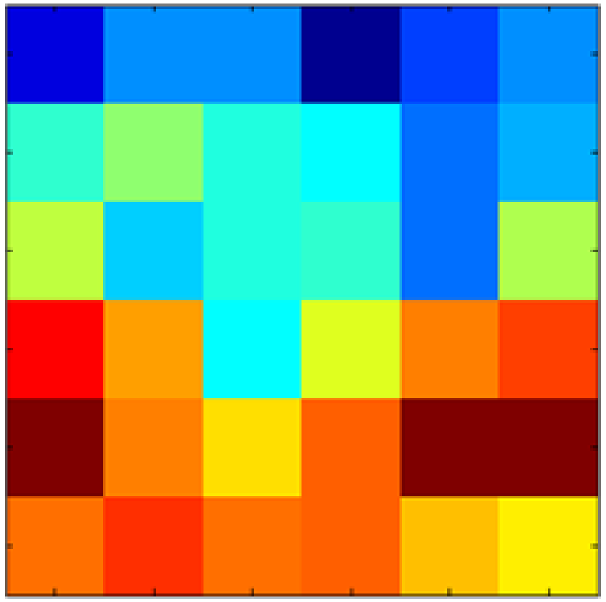


Novel view



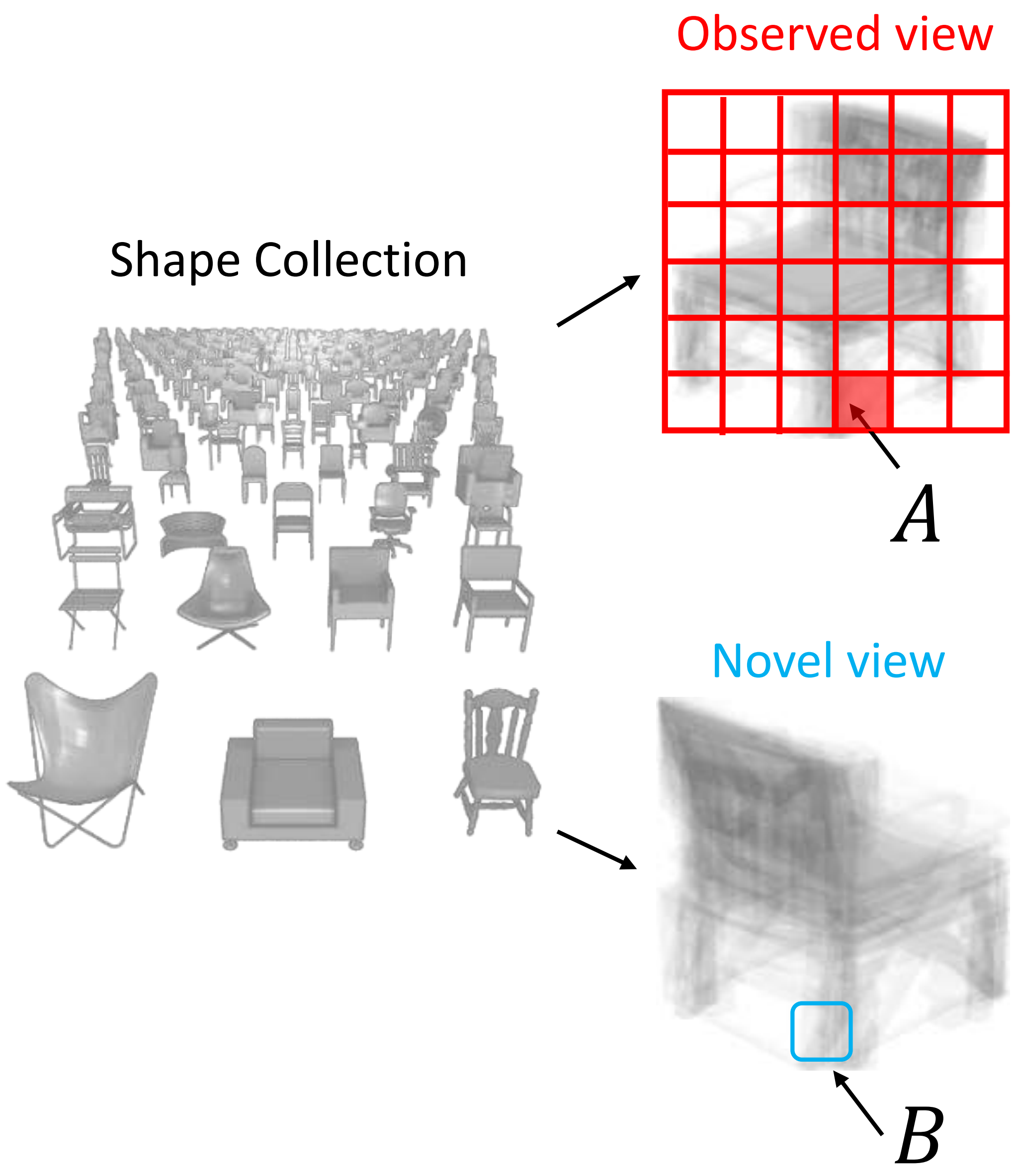
Surrogate Relationship Discovery





Surrogate suitability matrix

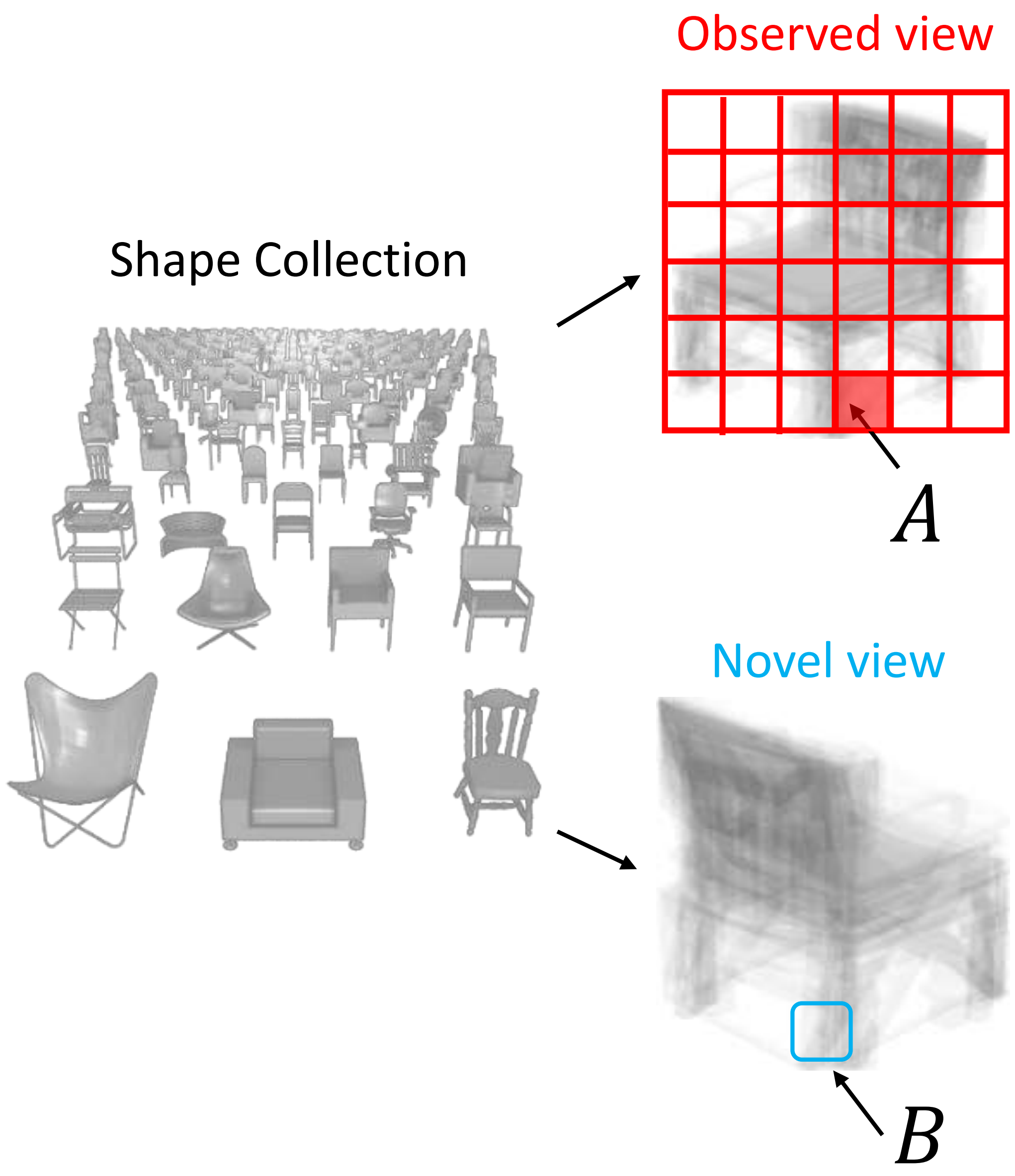
Formal Definition of *Surrogate Suitability*



Assume

A, B are discrete random variables

Formal Definition of *Surrogate Suitability*

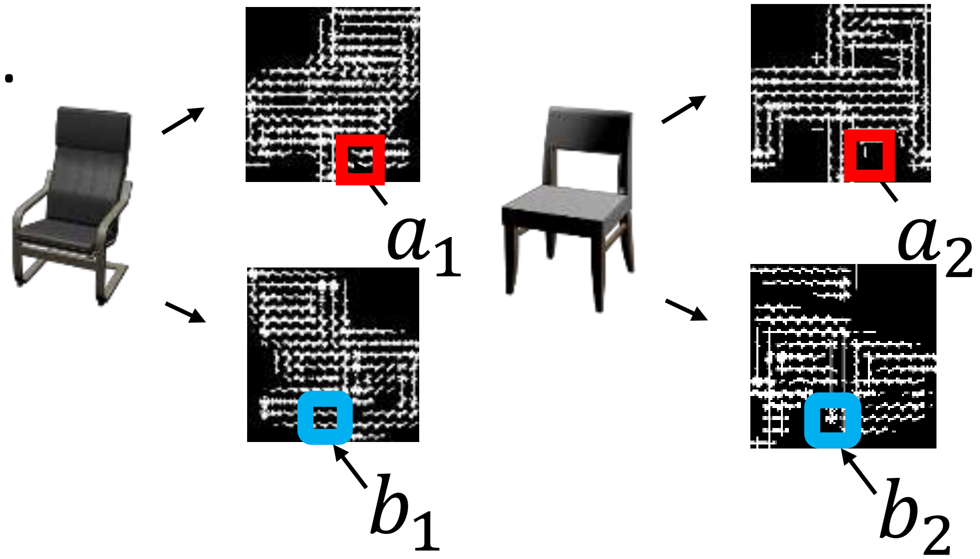


Assume

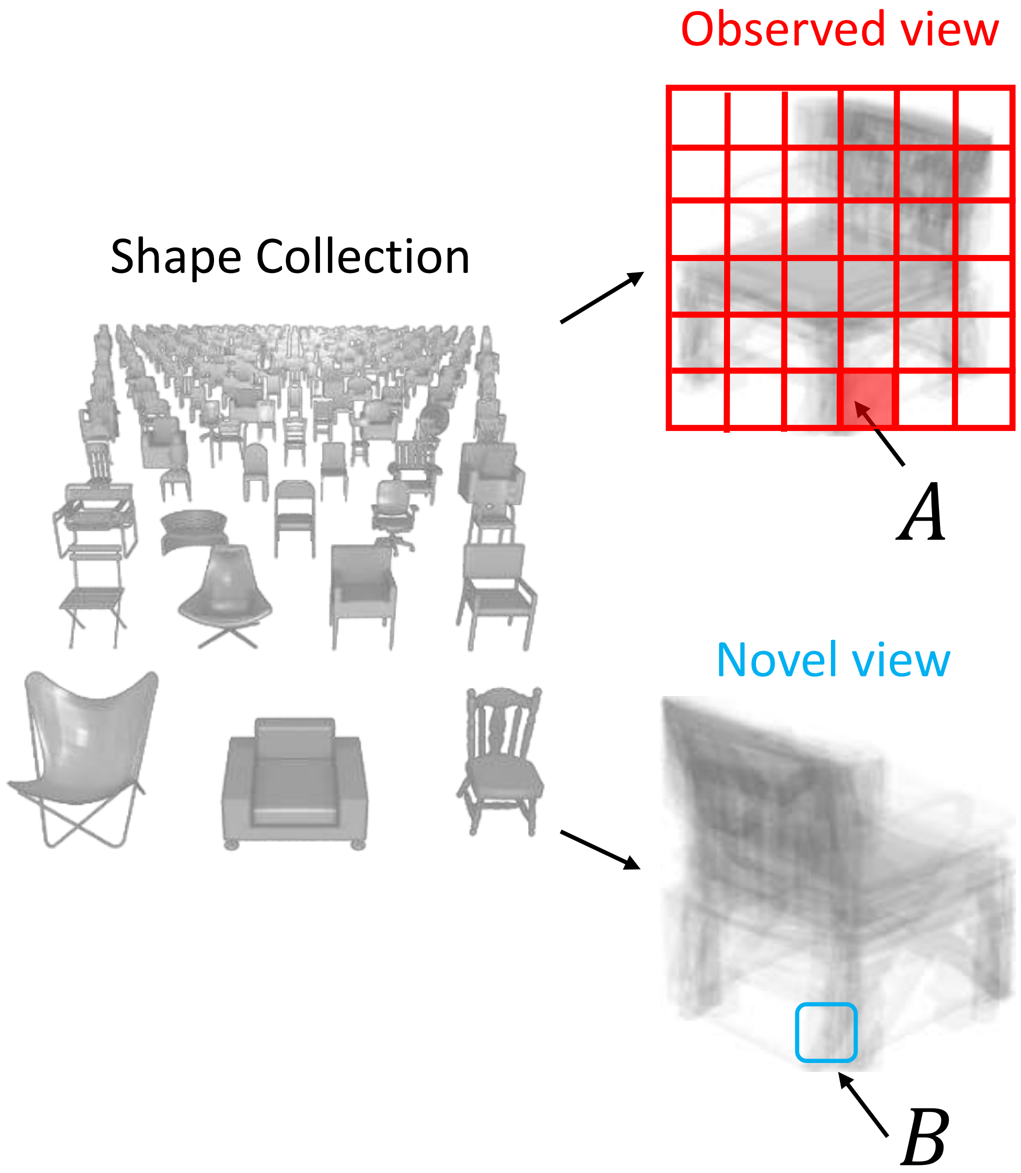
A, B are discrete random variables

$(a_1, b_1), (a_2, b_2)$, are i.i.d samples of (A, B)

e.g.



Formal Definition of *Surrogate Suitability*

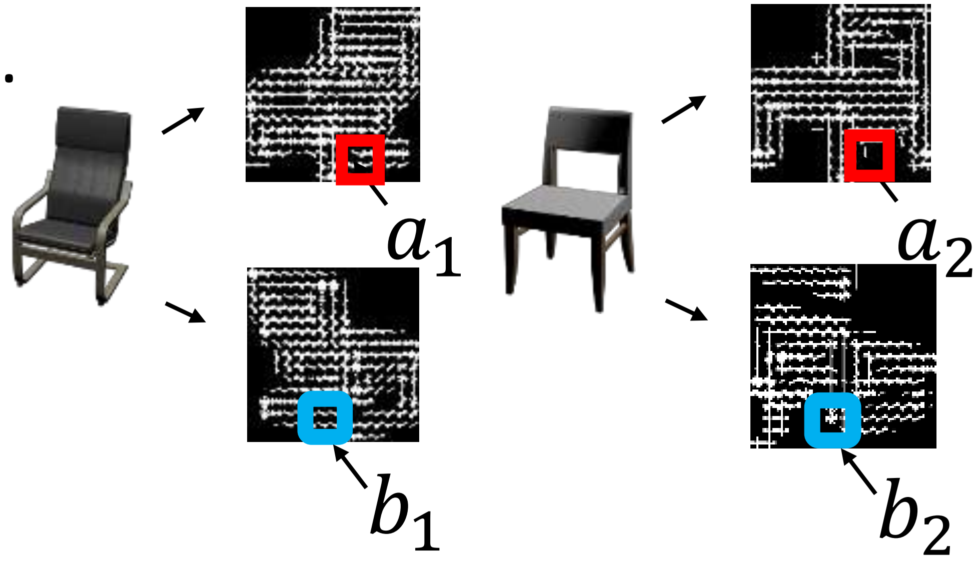


Assume

A, B are discrete random variables

$(a_1, b_1), (a_2, b_2)$, are i.i.d samples of (A, B)

e.g.

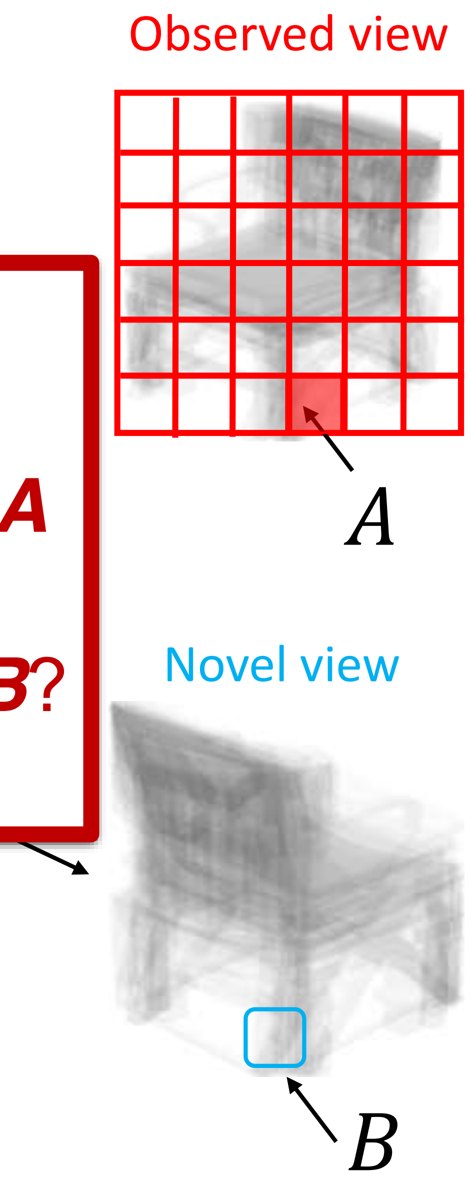


Surrogate suitability:

$$\gamma(A; B) = \log P(b_1 = b_2 | a_1 = a_2)$$

Formal Definition of *Surrogate Suitability*

How well can
the sameness at *A*
predict
the sameness at *B*?

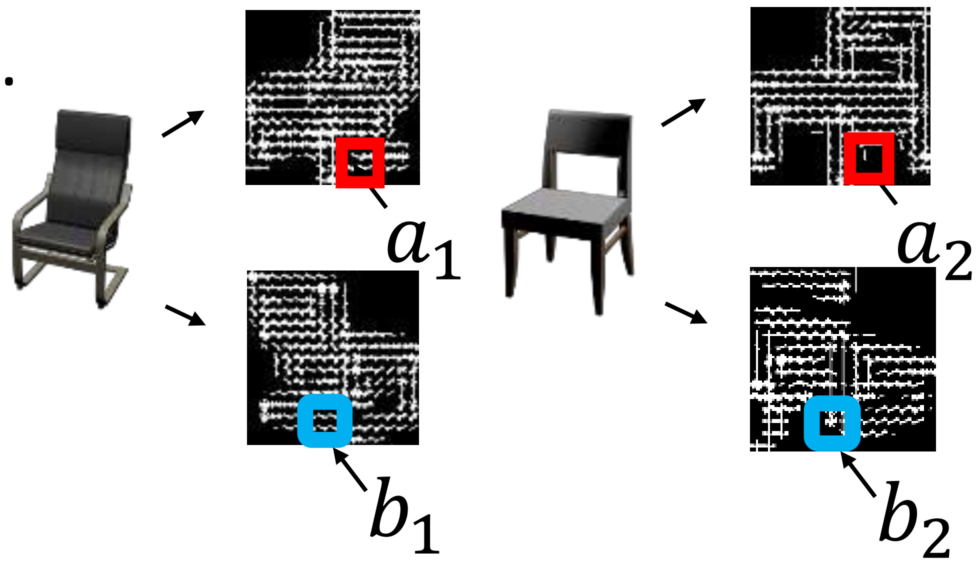


Assume

A, B are discrete random variables

$(a_1, b_1), (a_2, b_2)$, are i.i.d samples of (A, B)

e.g.



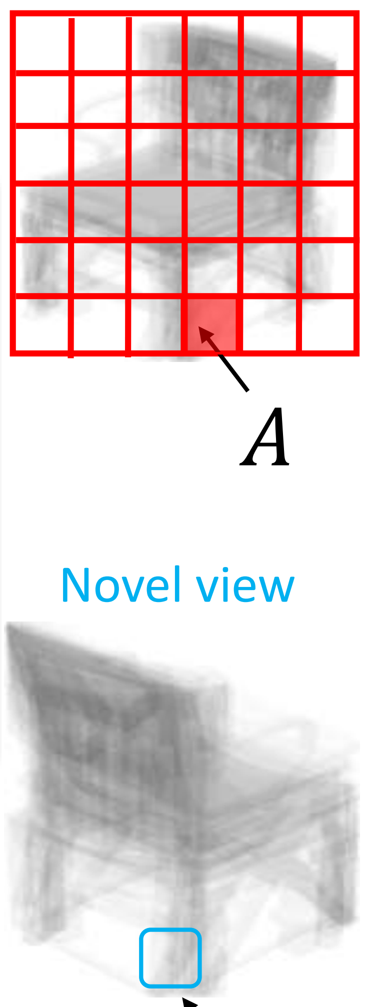
Surrogate suitability:

$$\gamma(A; B) = \log P(b_1 = b_2 | a_1 = a_2)$$

Formal Definition of *Surrogate Suitability*

How well can
the sameness at *A*
predict
the sameness at *B*?

Cross-view transfer
of **relationships**

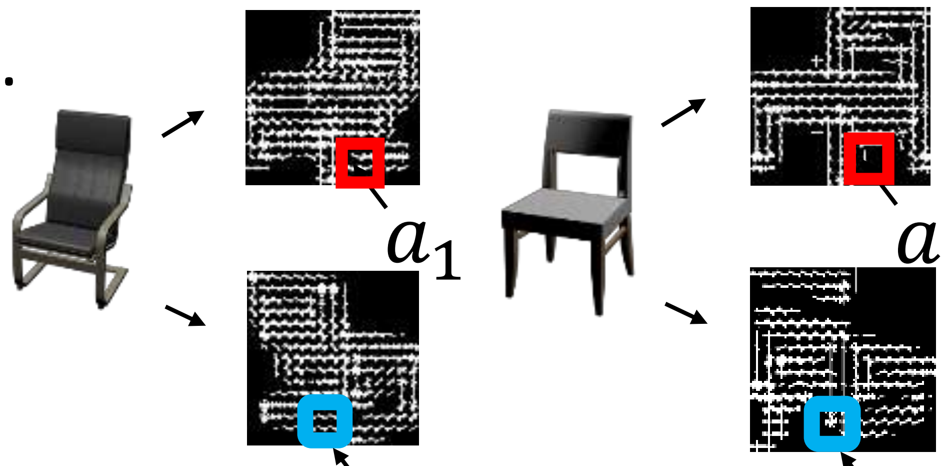


Assume

A, B are discrete random variables

$(a_1, b_1), (a_2, b_2)$, are i.i.d samples of (A, B)

e.g.



Surrogate suitability:

$$\gamma(A; B) = \log P(b_1 = b_2 | a_1 = a_2)$$

Estimation of Surrogate Suitability

Derivation shows

$$\begin{aligned}\gamma(A;B) &= \log \sum P^2(A,B) - \log \sum P^2(B) \\ &= -H_R(A,B) + H_R(B)\end{aligned}$$

H_R: *Renyi-entropy*

Estimation of Surrogate Suitability

Derivation shows

Sample complexity: tight bound $\Theta(V_A + V_B)$

Sample complexity: tight bound $\Theta(V_A + V_B)$

where V_A and V_B are vocabulary size of A and B

Estimation of Surrogate Suitability

Derivation shows

Sample complexity: tight bound $\Theta(V_A + V_B)$

Sample complexity: tight bound $\Theta(V_A + V_B)$

where V_A and V_B are vocabulary size of A and B

Theoretically optimal algorithm is proposed that reaches the bound

Estimation of Surrogate Suitability

Derivation shows

Sample complexity: tight bound $\Theta(V_A + V_B)$

Sample complexity: tight bound $\Theta(V_A + V_B)$

where V_A and V_B are vocabulary size of A and B

Theoretically optimal algorithm is proposed that reaches the bound

Strong connection with Mutual Information

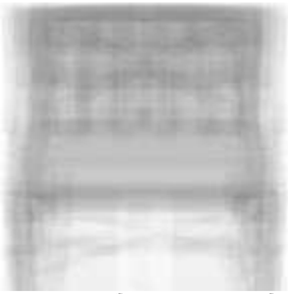
$$MI(A, B) = -H(A, B) + H(A) + H(B)$$

More Visualization of Surrogate Suitability Matrix

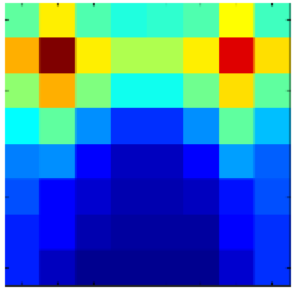
Novel view

Observed view

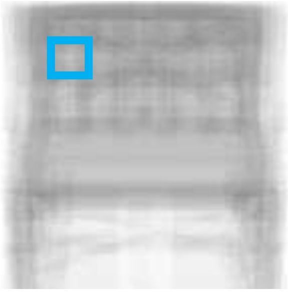
$\gamma(A; B)$



$0°$ (front)



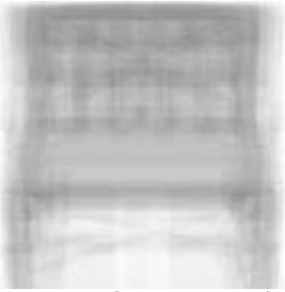
B



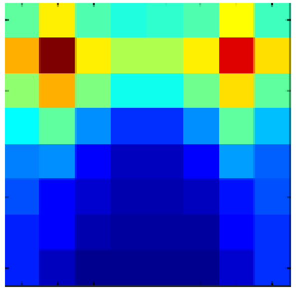
$0°$ (front)

More Visualization of Surrogate Suitability Matrix

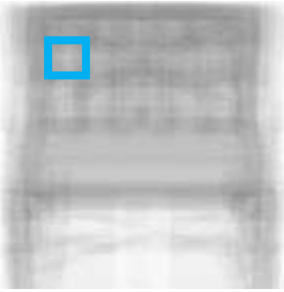
Novel view | Observed view | $\gamma(A; B)$



0° (front)

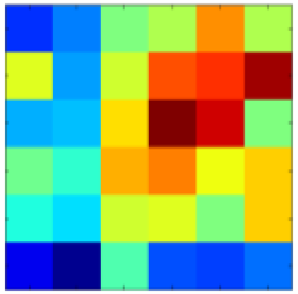


B

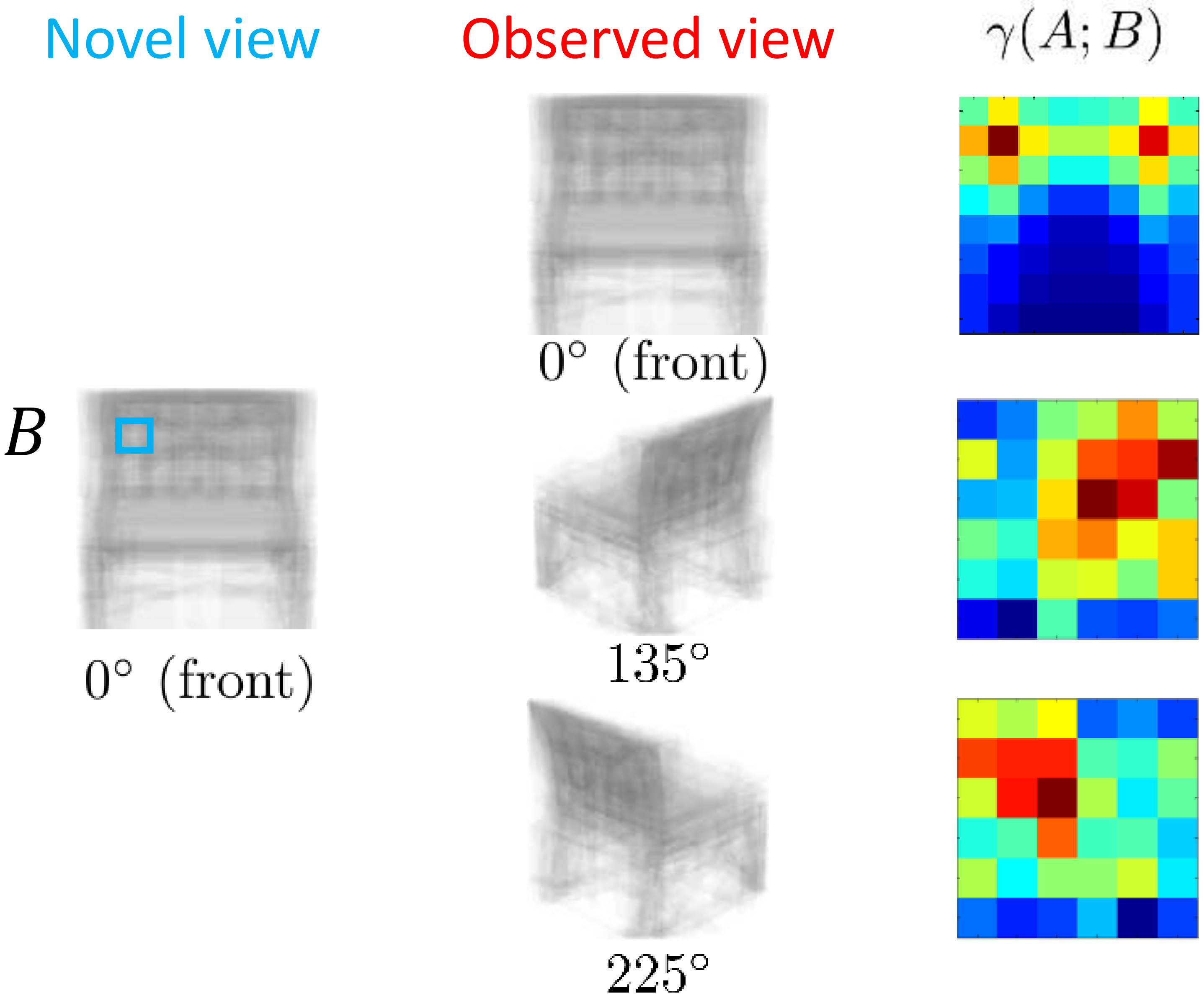


0° (front)

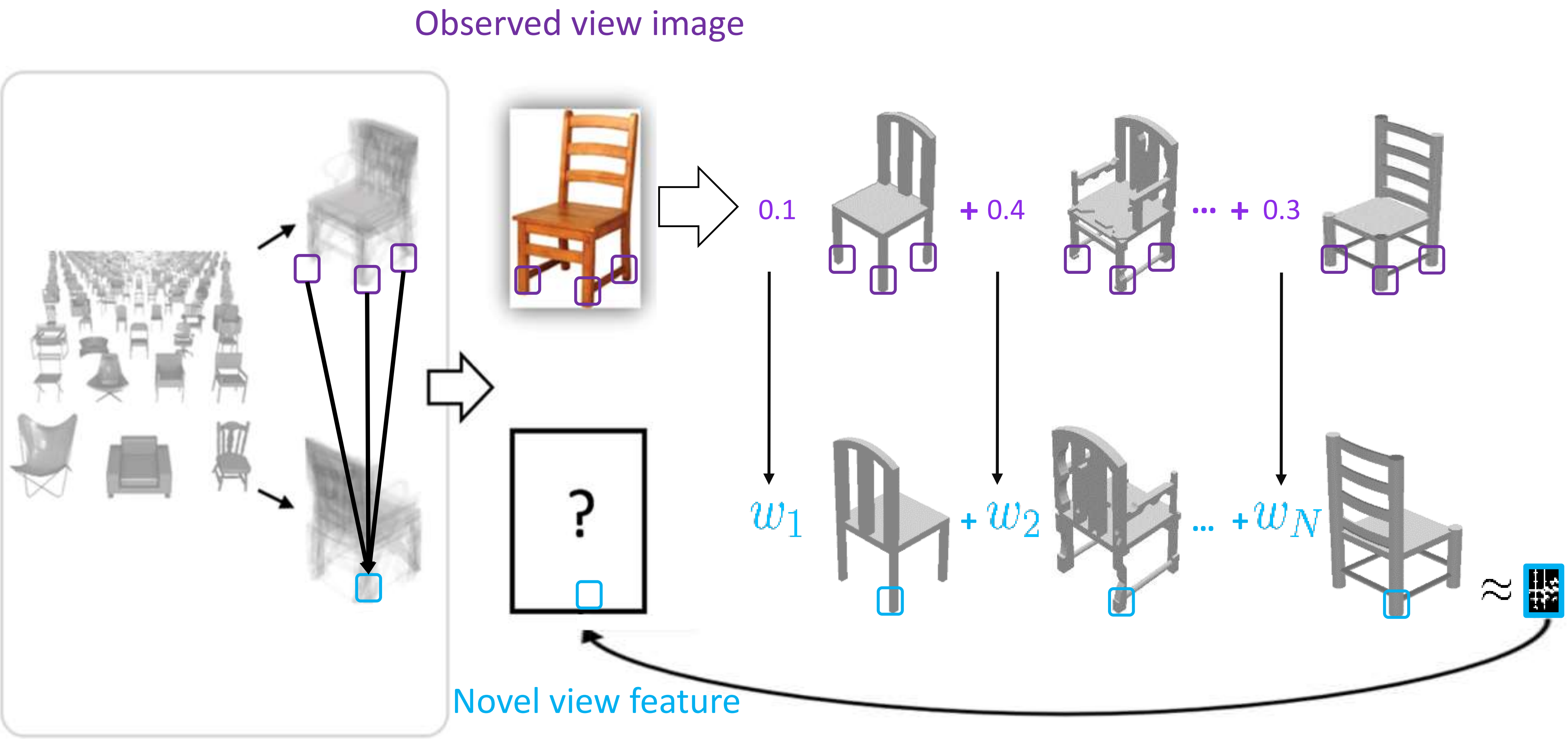
135°



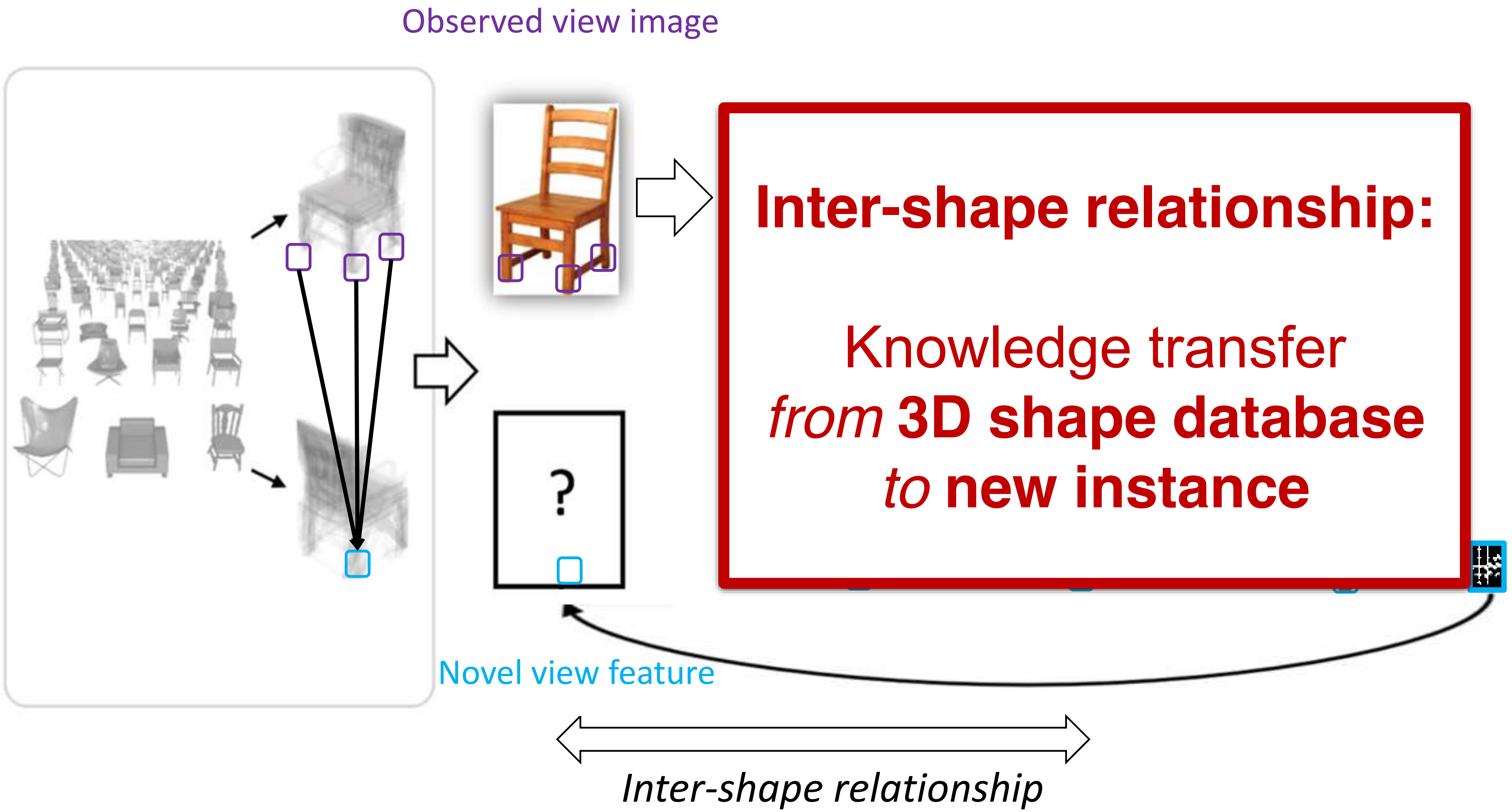
More Visualization of Surrogate Suitability Matrix



Review of Pipeline



Review of Pipeline



Review of Pipeline

Observed view image

Intra-shape relationship

Intra-shape relationship:

Knowledge transfer
from **observed view**
to **novel view**

Inter-shape relationship:

Knowledge transfer
from **3D shape database**
to **new instance**

Novel view feature

Inter-shape relationship

Outline

Motivation

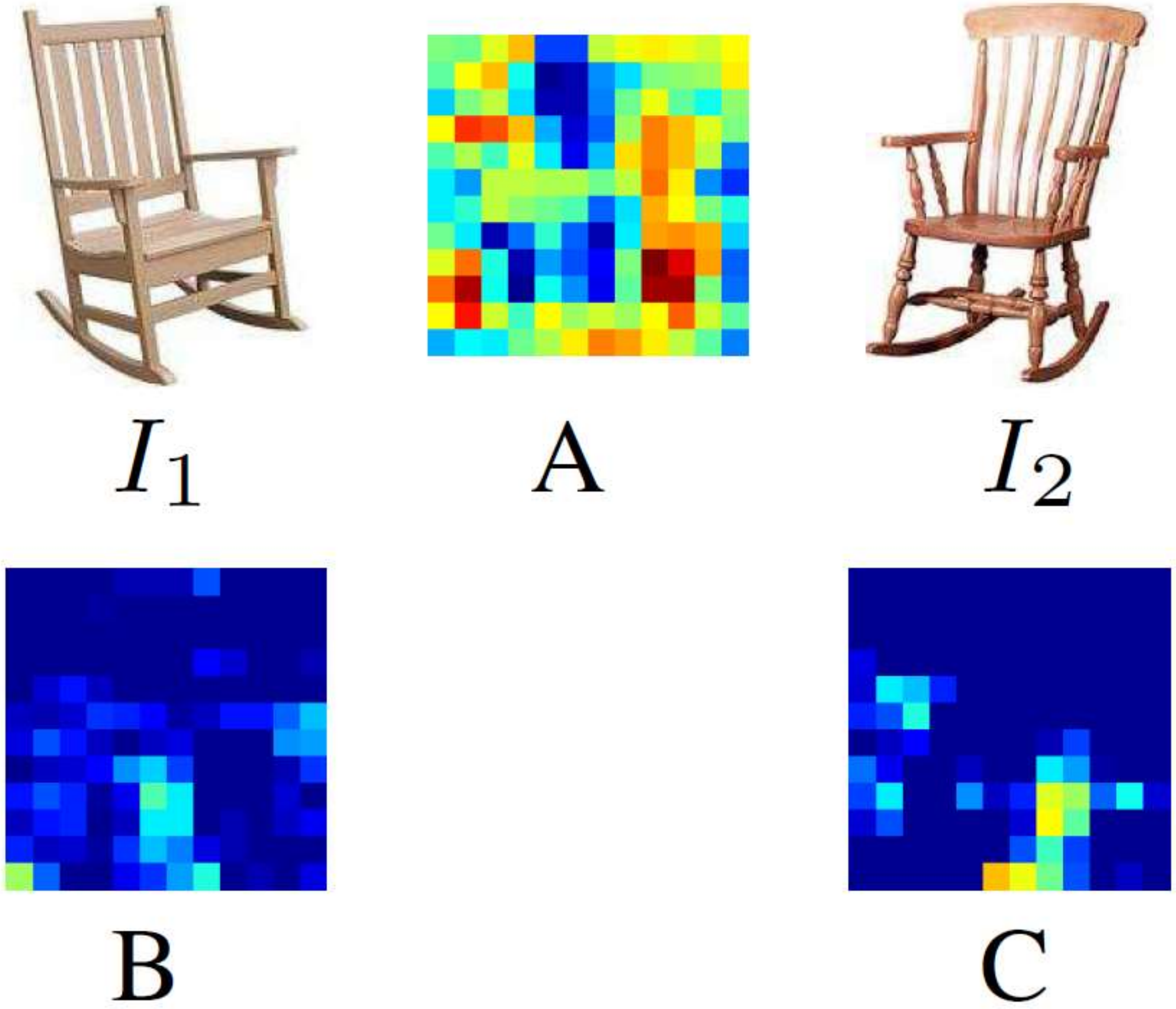
Approach

Applications

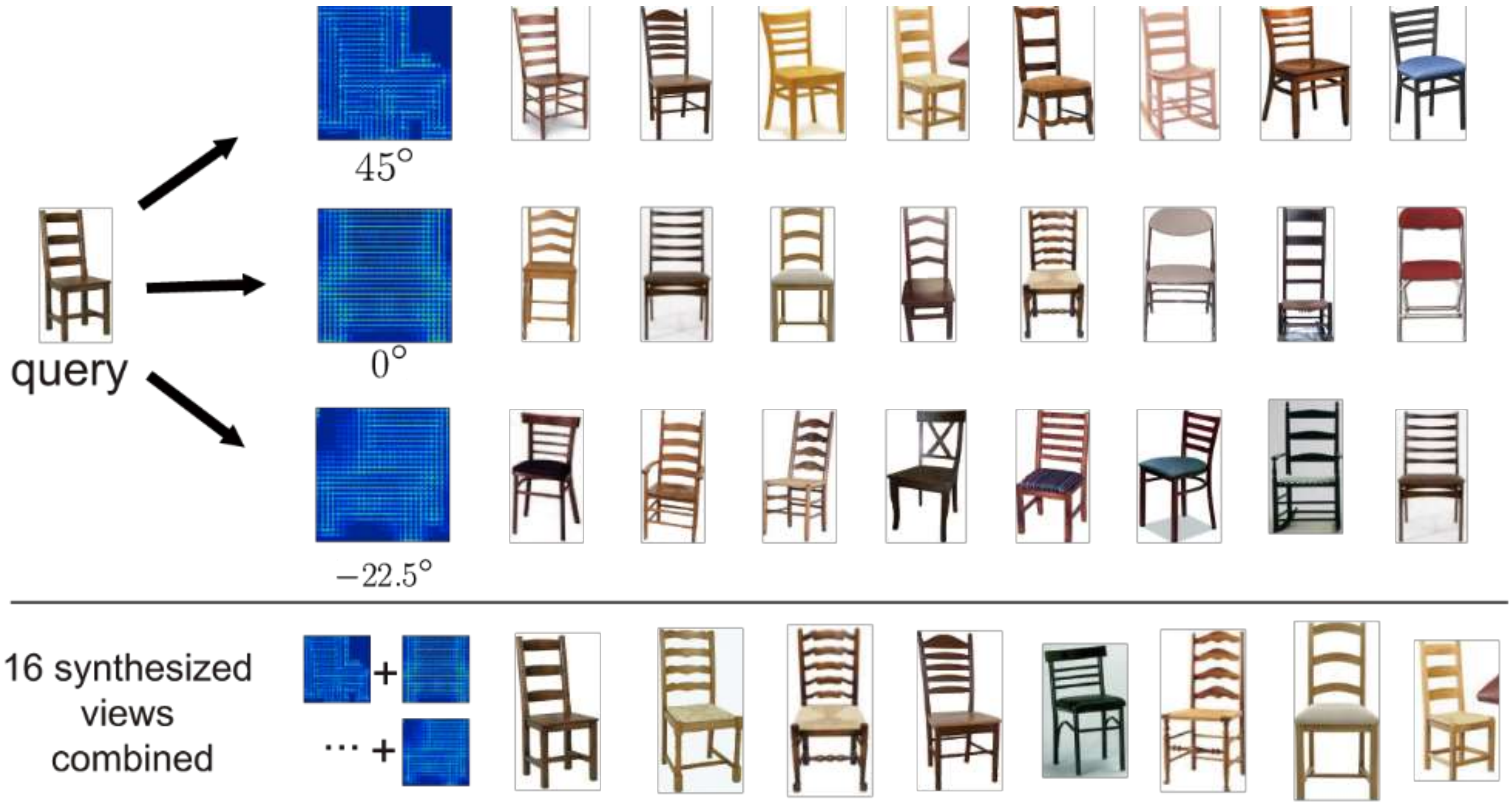
Method Diagnosis

Conclusion

Application: Cross-view localized image comparison



Cross-view Image Retrieval



Application: View-agnostic Image Retrieval

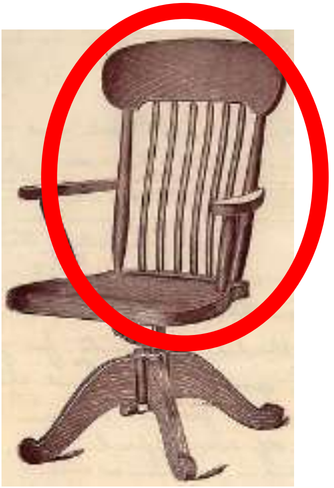
vertical bars

swivel base

HoG L2

Ours (combined HoG)

Application: View-agnostic Image Retrieval



vertical bars

swivel base

HoG L2

Ours (combined HoG)

Application: View-agnostic Image Retrieval

vertical bars

swivel base

HoG L2

Ours (combined HoG)

Part-based View-agnostic Image Retrieval

Generalizability to Many Feature Types

Feature	Method	Chair	Car	Bus	Motorbike	Train	Avg	
HoG	original	71.0	27.8	37.4	40.7	52.1	45.8	+7.5
	augmented	80.1	32.0	43.0	48.0	63.6	53.3	
BoVW	original	67.8	28.0	38.0	40.2	52.1	45.2	+4.4
	augmented	70.2	30.9	41.7	44.1	61.0	49.6	
Fisher	original	67.5	27.0	35.3	42.1	48.1	44.0	+5.3
	augmented	70.2	30.7	38.4	46.9	60.2	49.3	
LLC	original	71.7	28.3	35.4	40.6	55.9	46.4	+5.7
	augmented	74.9	34.8	44.9	45.7	60.2	52.1	
Caffe Pool5	original	69.0	26.7	39.1	42.1	55.3	46.4	+3.2
	augmented	74.6	31.0	42.0	44.8	55.7	49.6	
Caffe FC7	original	74.4	28.7	38.6	45.6	58.2	49.1	+4.3
	augmented	78.5	34.8	42.5	49.8	61.3	53.4	

- Task: fine-grained retrieval (images and annotations are from ImageNet)
- Metric: Average Precision

Outline

Motivation

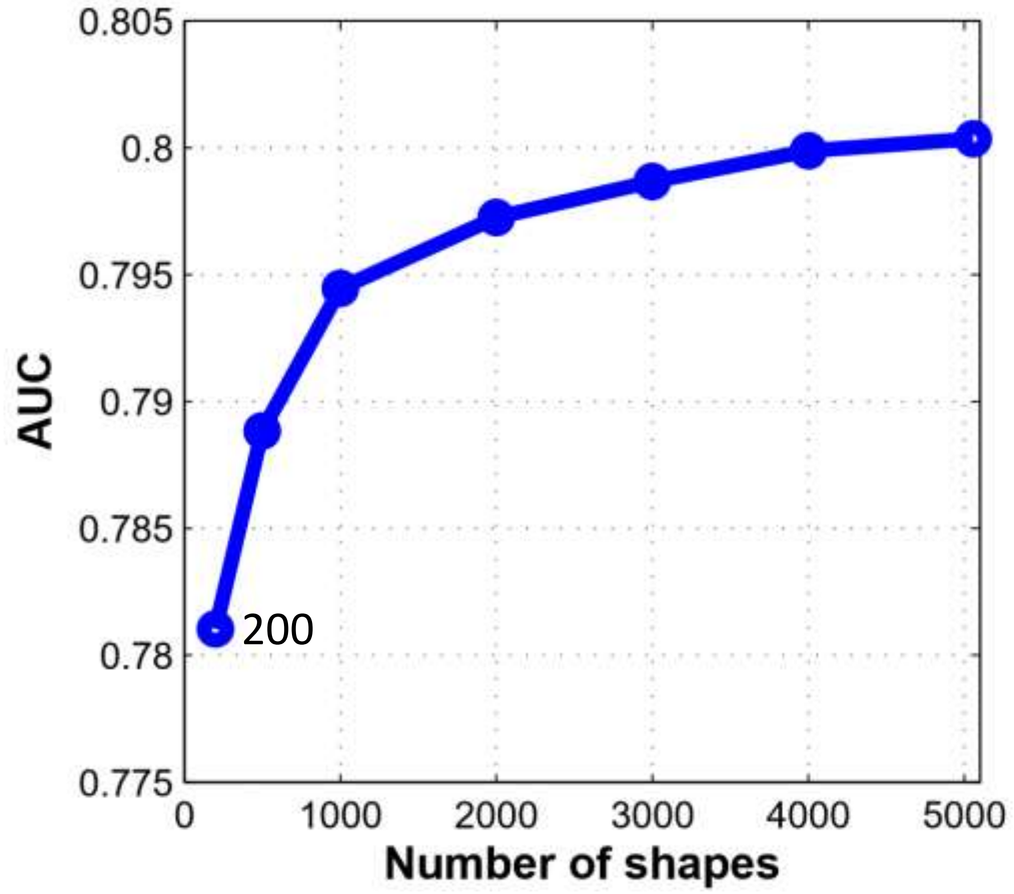
Approach

Applications

Method Diagnosis

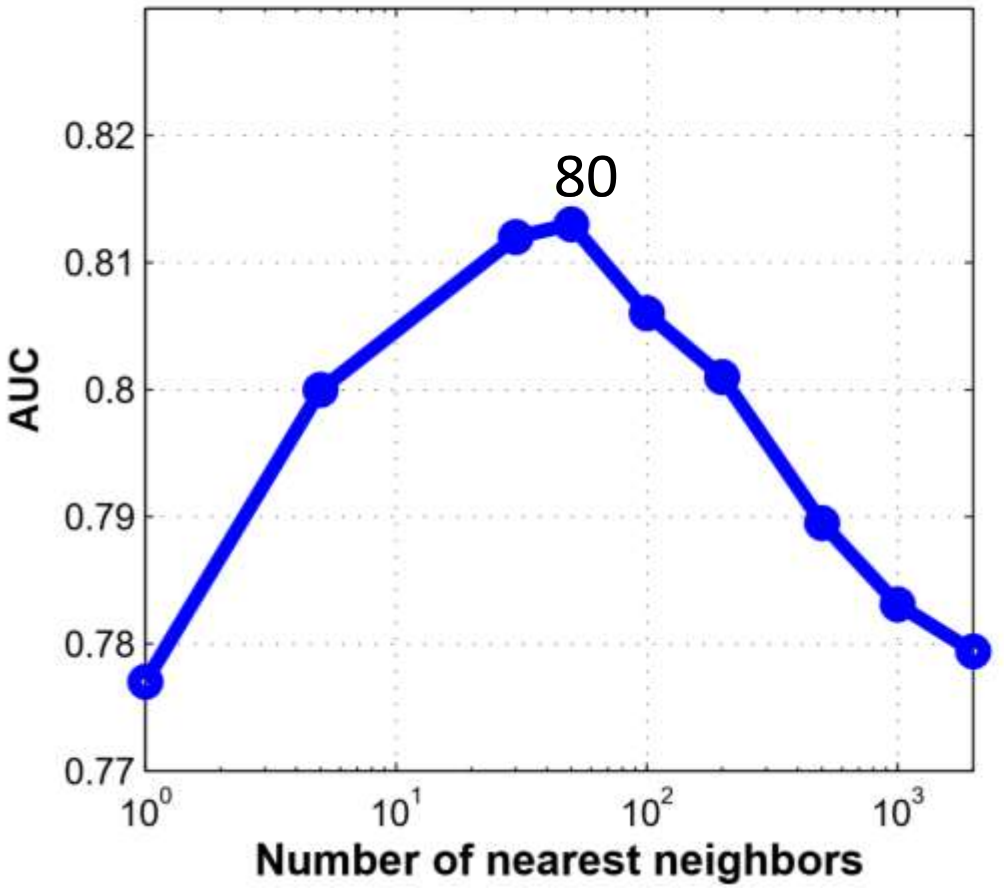
Conclusion

How many shapes are sufficient?



(Measured by Average Precision on Fine-grained retrieval for Chairs)

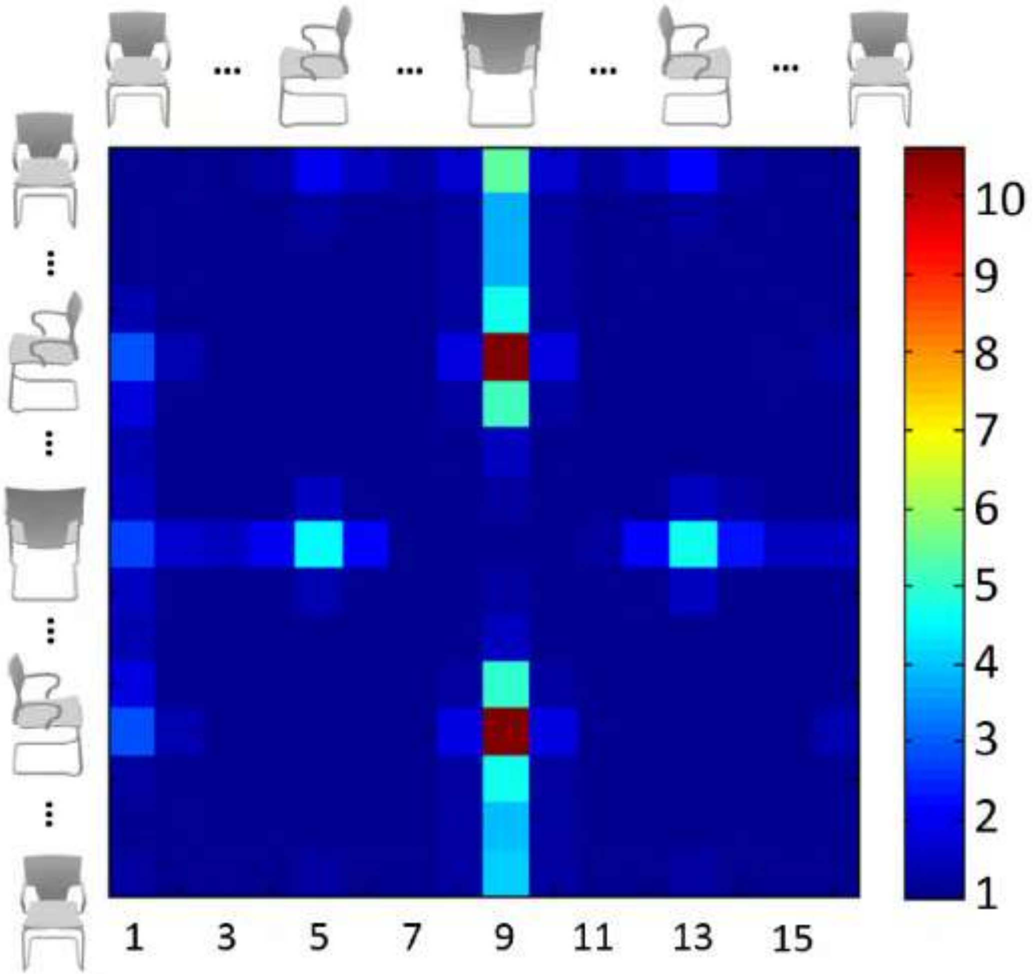
How many neighboring shapes for interpolation?



(Measured by Average Precision on Fine-grained retrieval for Chairs)

How well can one view predict another view?

Controlled diagnosis on renderings



Cross-view retrieval rank

Outline

Motivation

Approach

Applications

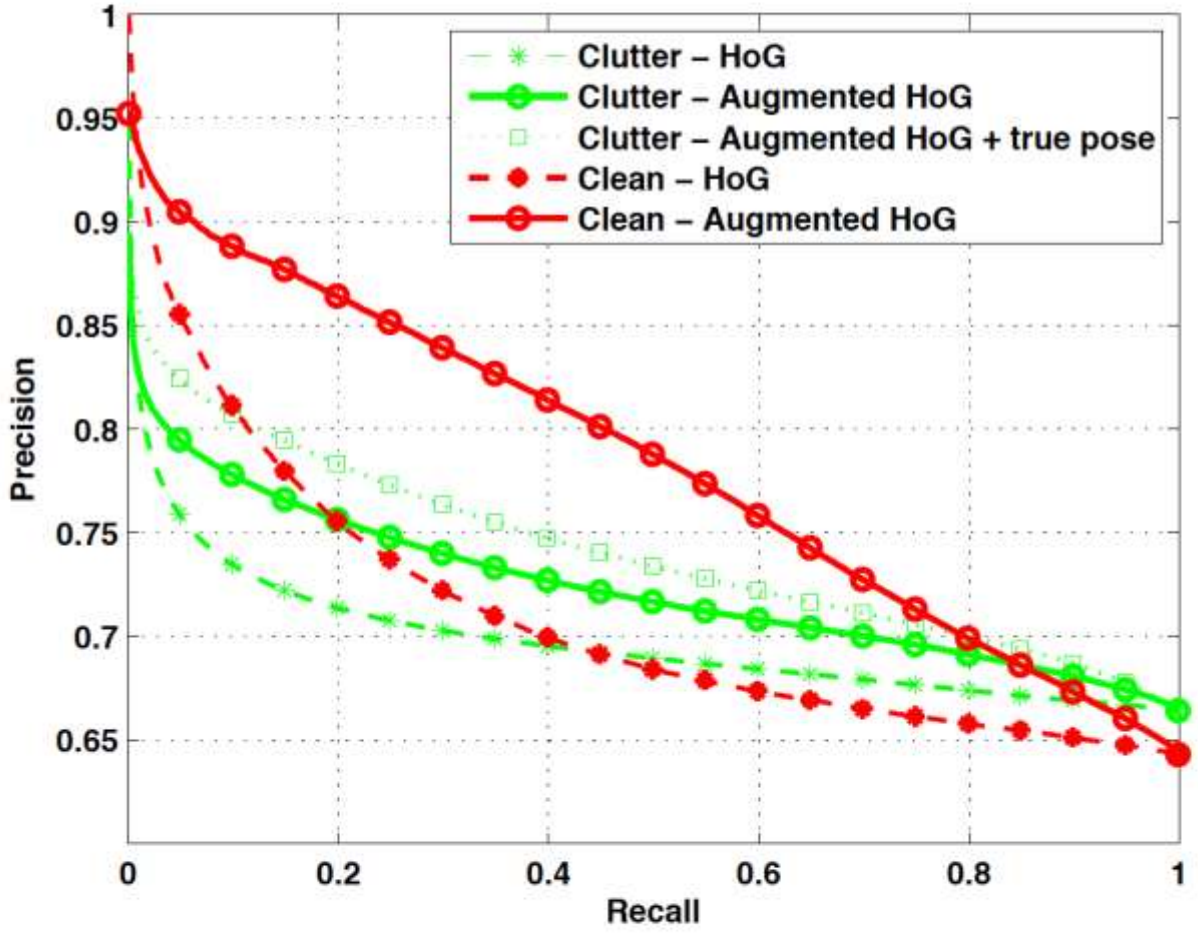
Method Diagnosis

Conclusion

Conclusion

- A novel framework for synthesizing object features at novel views
- 3D shape database provides the knowledge of feature synthesis
- For relationship transfer, surrogate suitability is defined, which is a type of "predictability" between random variables.
- A theoretically optimal estimator is proposed

Thank you!



(b) Clean vs. cluttered