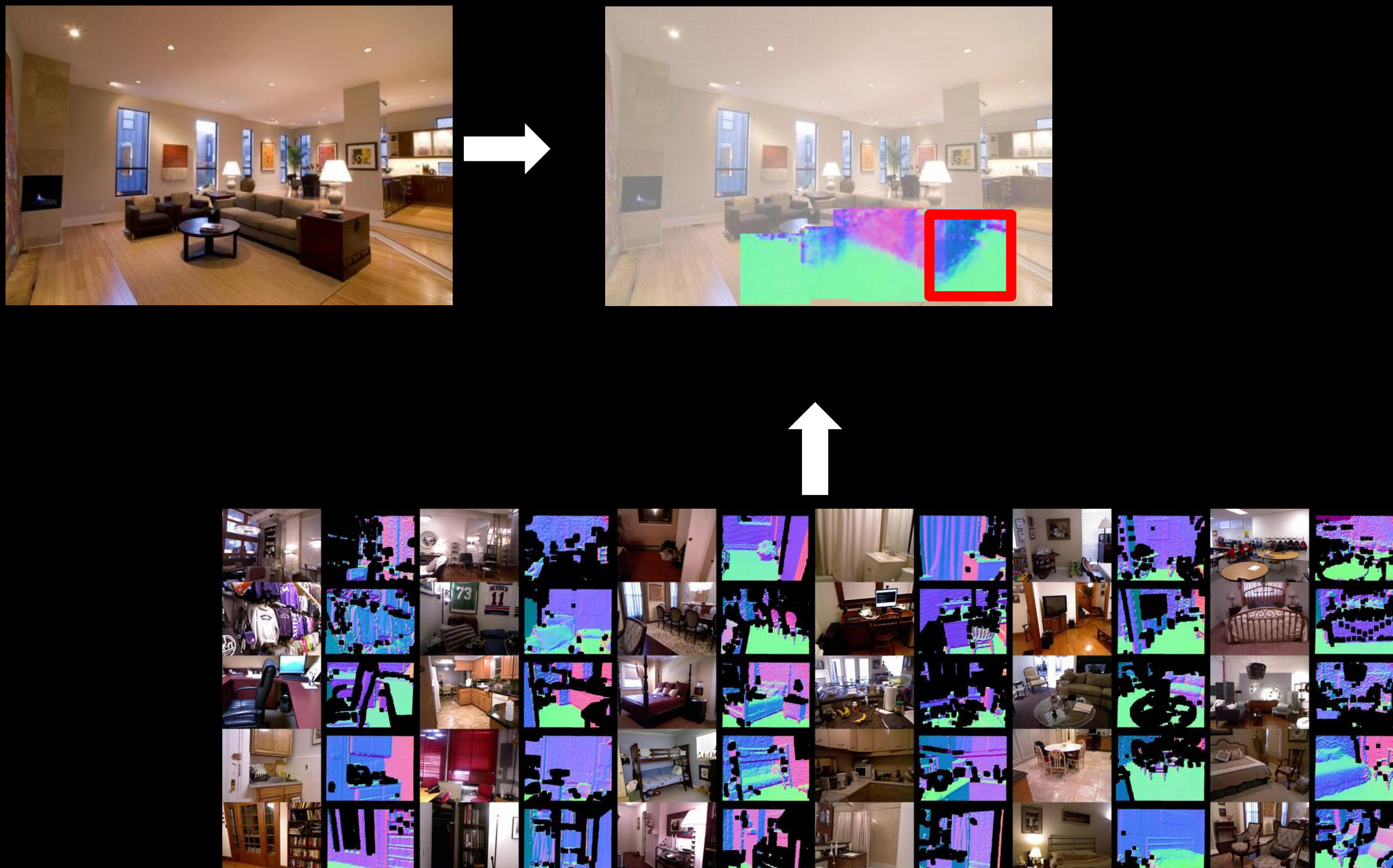
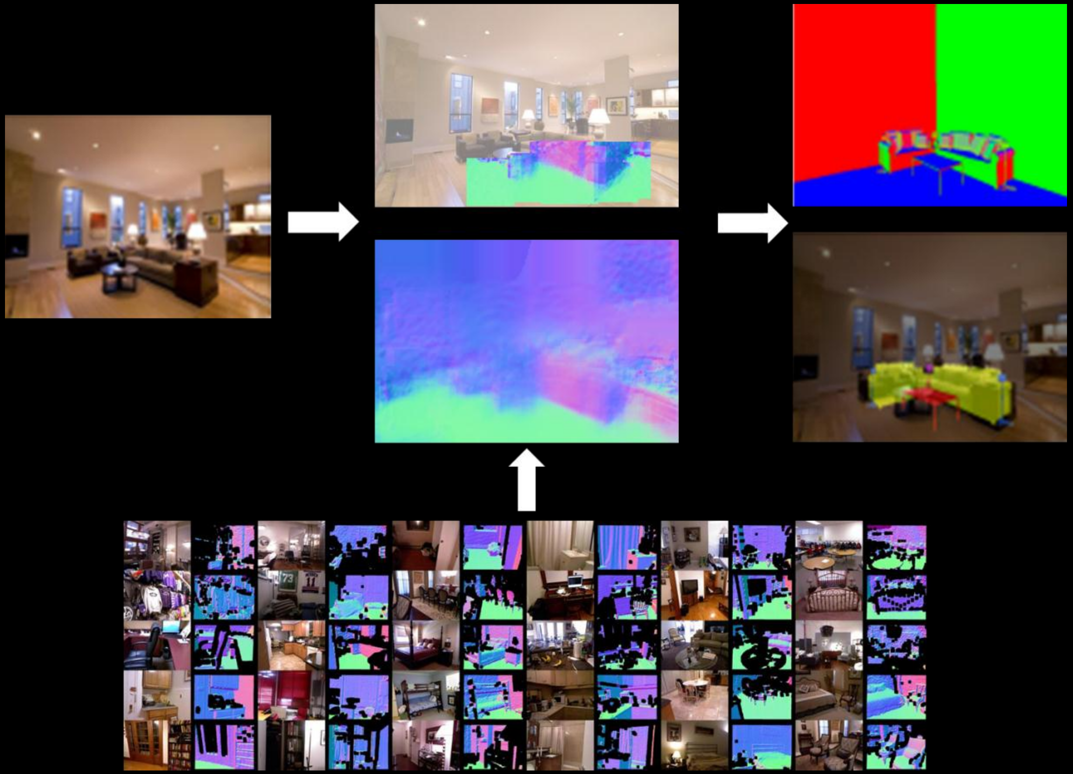
Learning from 3D Data for Image Interpretation

Martial Hebert
Abhinav Gupta
David Fouhey, Adrien Matricon, Wajahat Hussain

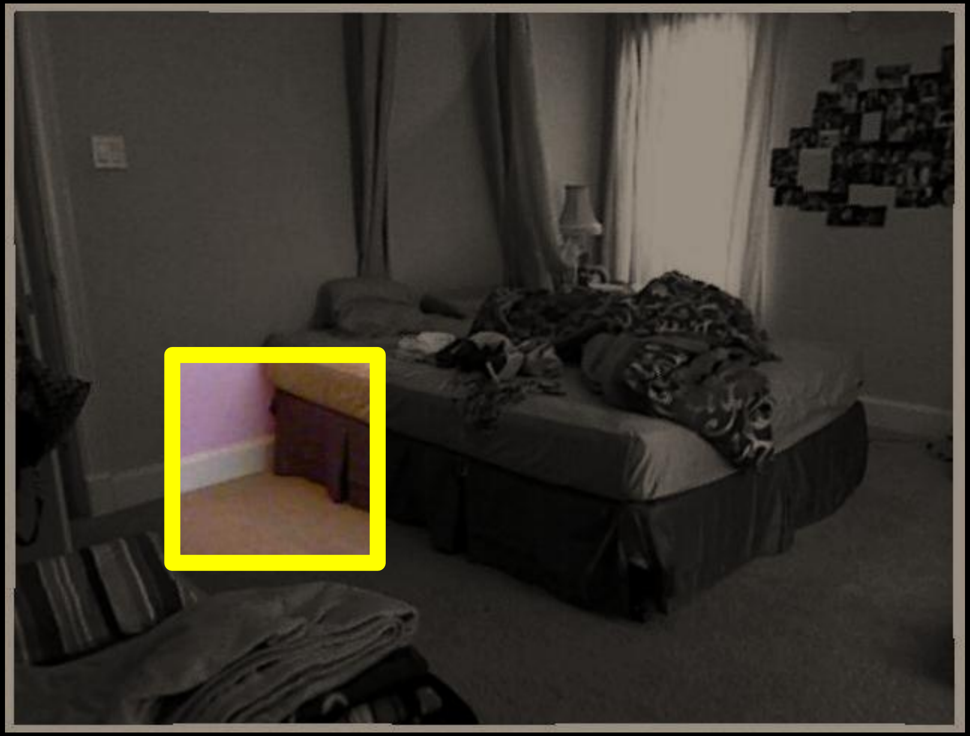


Slides adapted from David Fouhey

- Mid-level primitives learned from image+3D can be used to transfer geometric information?
- Geometric reasoning can use this local evidence to produce a consistent geometric interpretation?

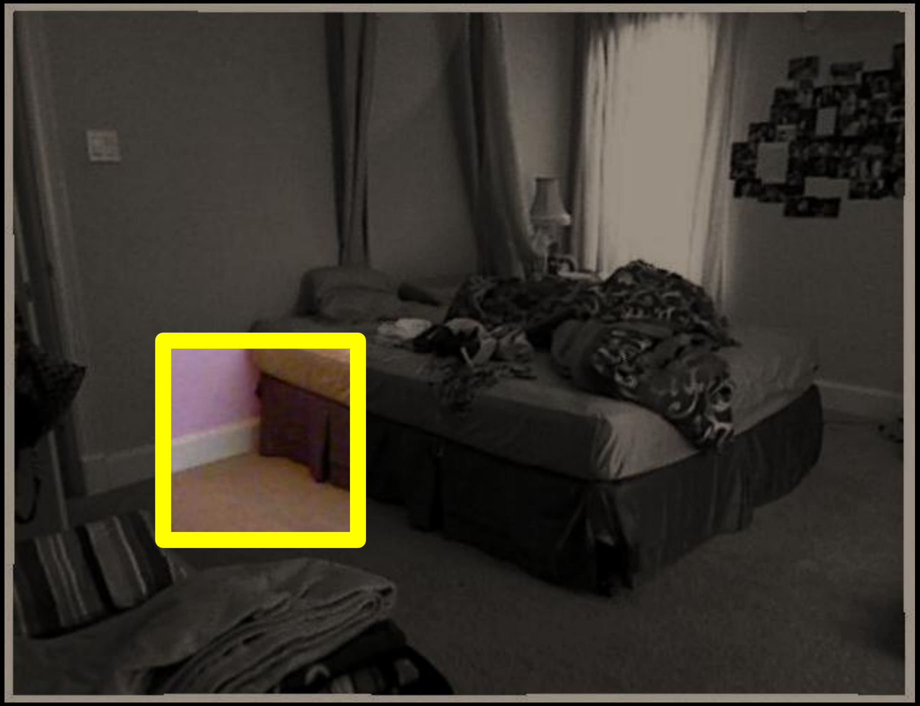


Pattern Repetition



Common patterns correspond to common geometric configurations

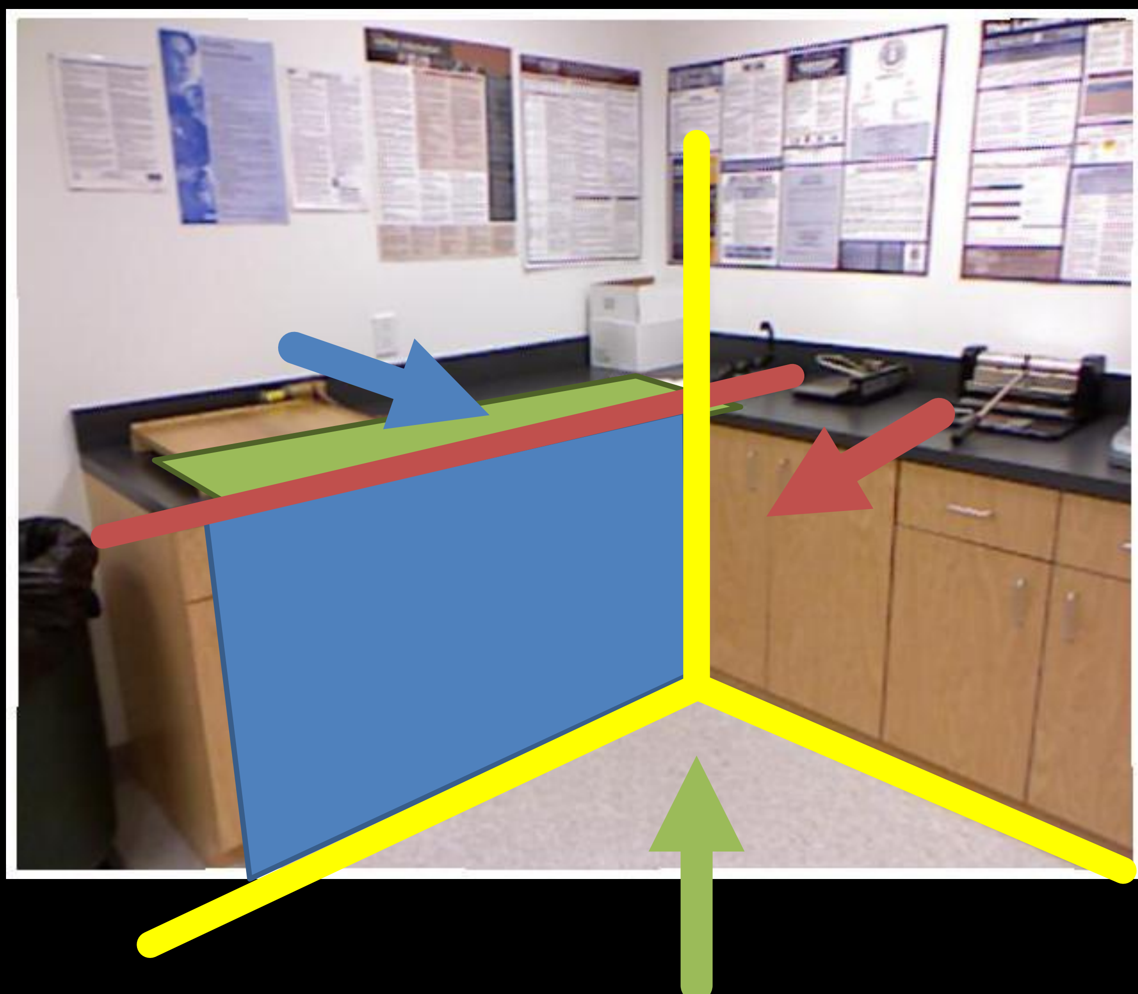
Pattern Repetition



Pattern Repetition

...

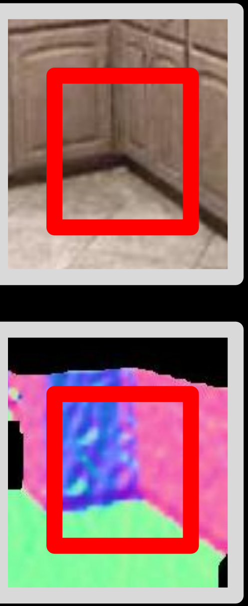
Physical/Geometric Constraints



Primitives

Visually Discriminative

Geometrically Informative

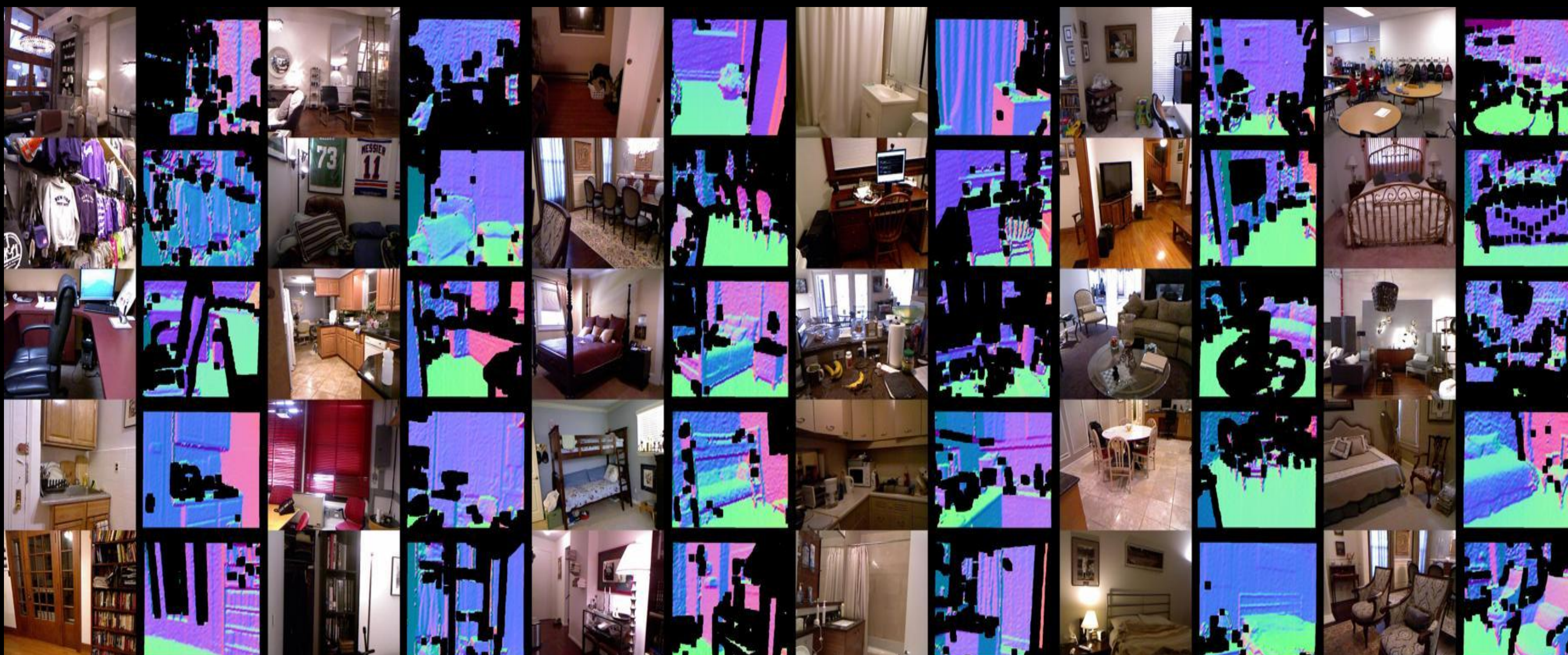


Image

Surface Normals

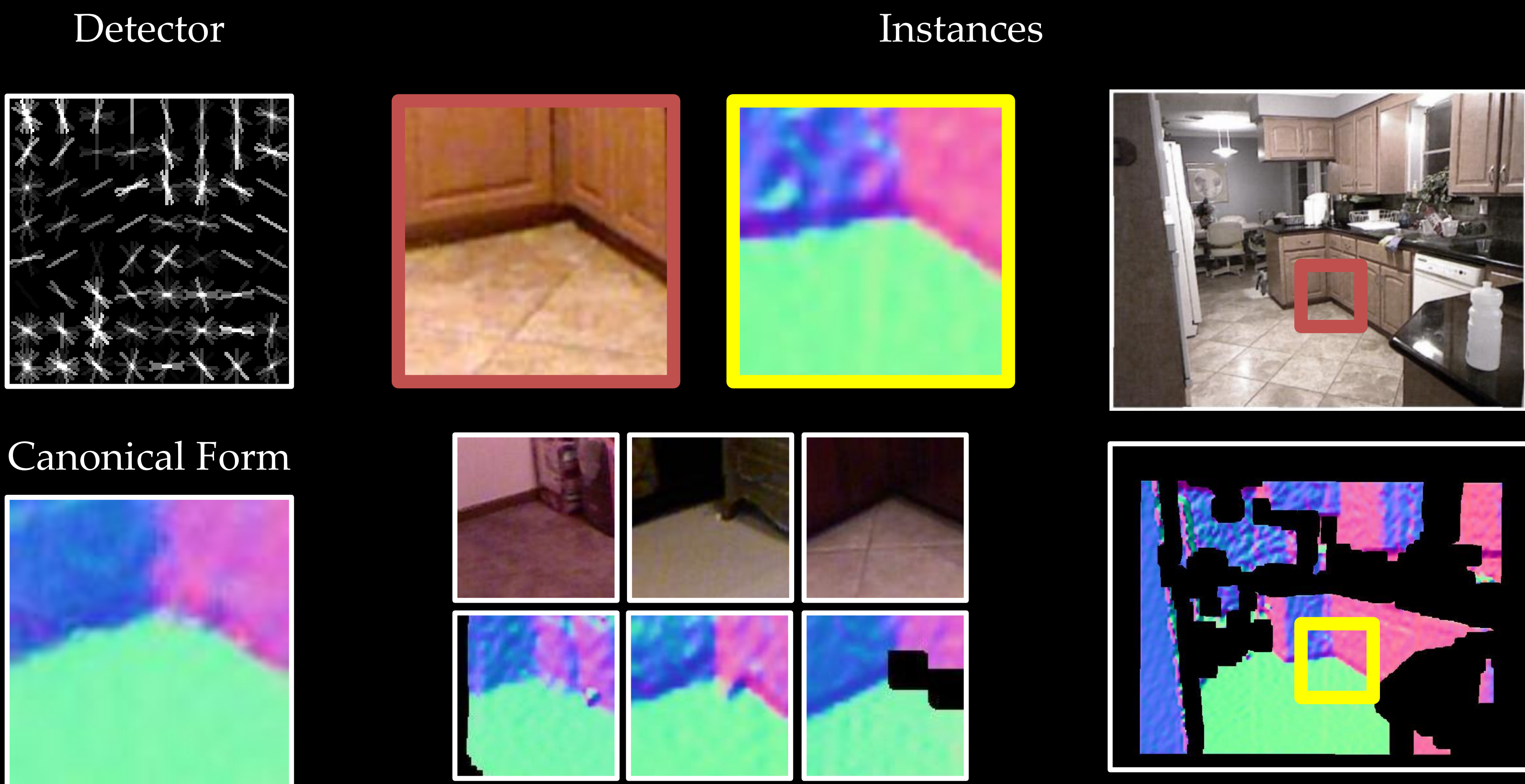
Saurabh Singh et al. *Discriminative Mid-Level Patches*

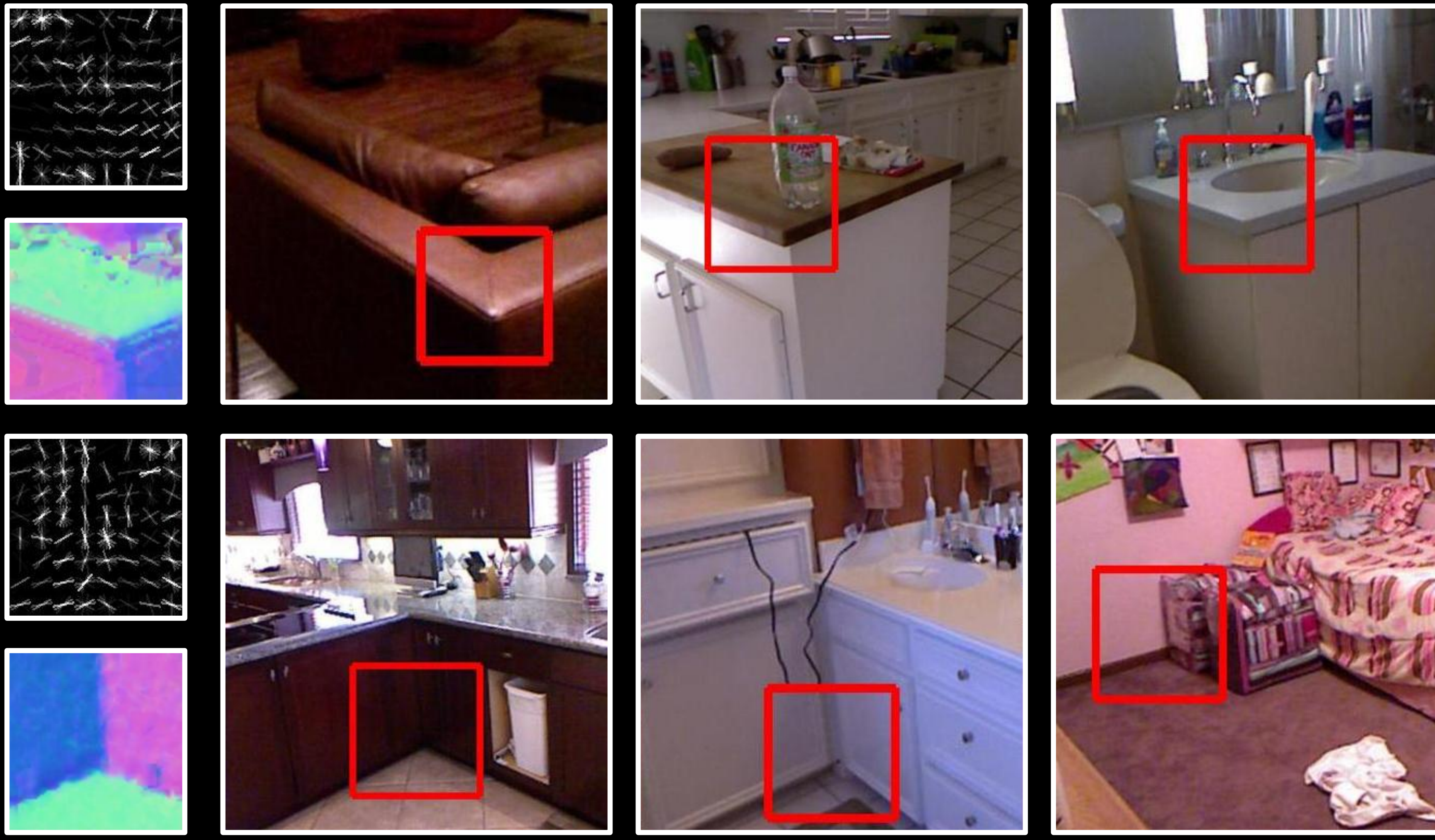
Geometric configurations from large-scale RGBD data.

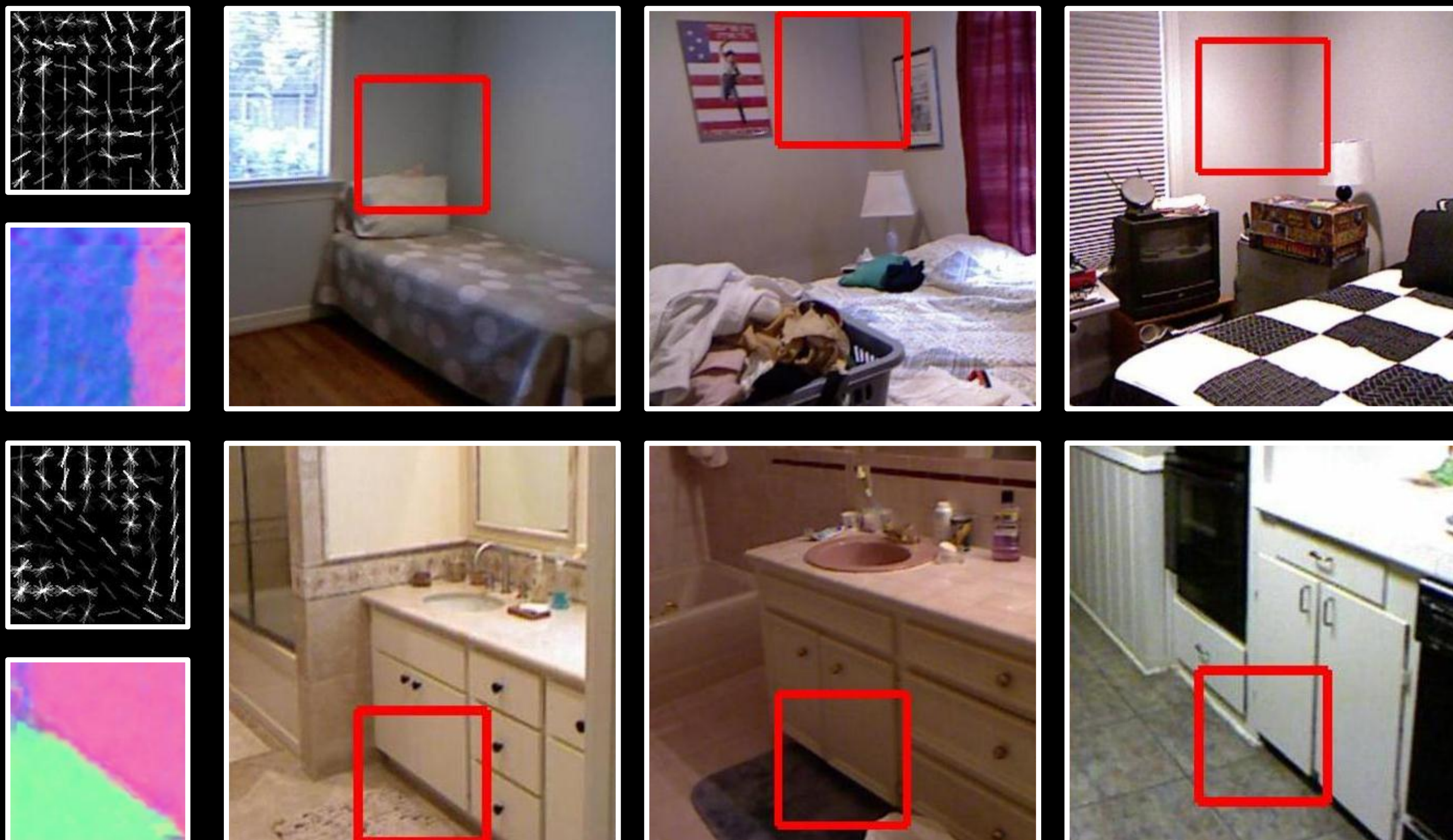


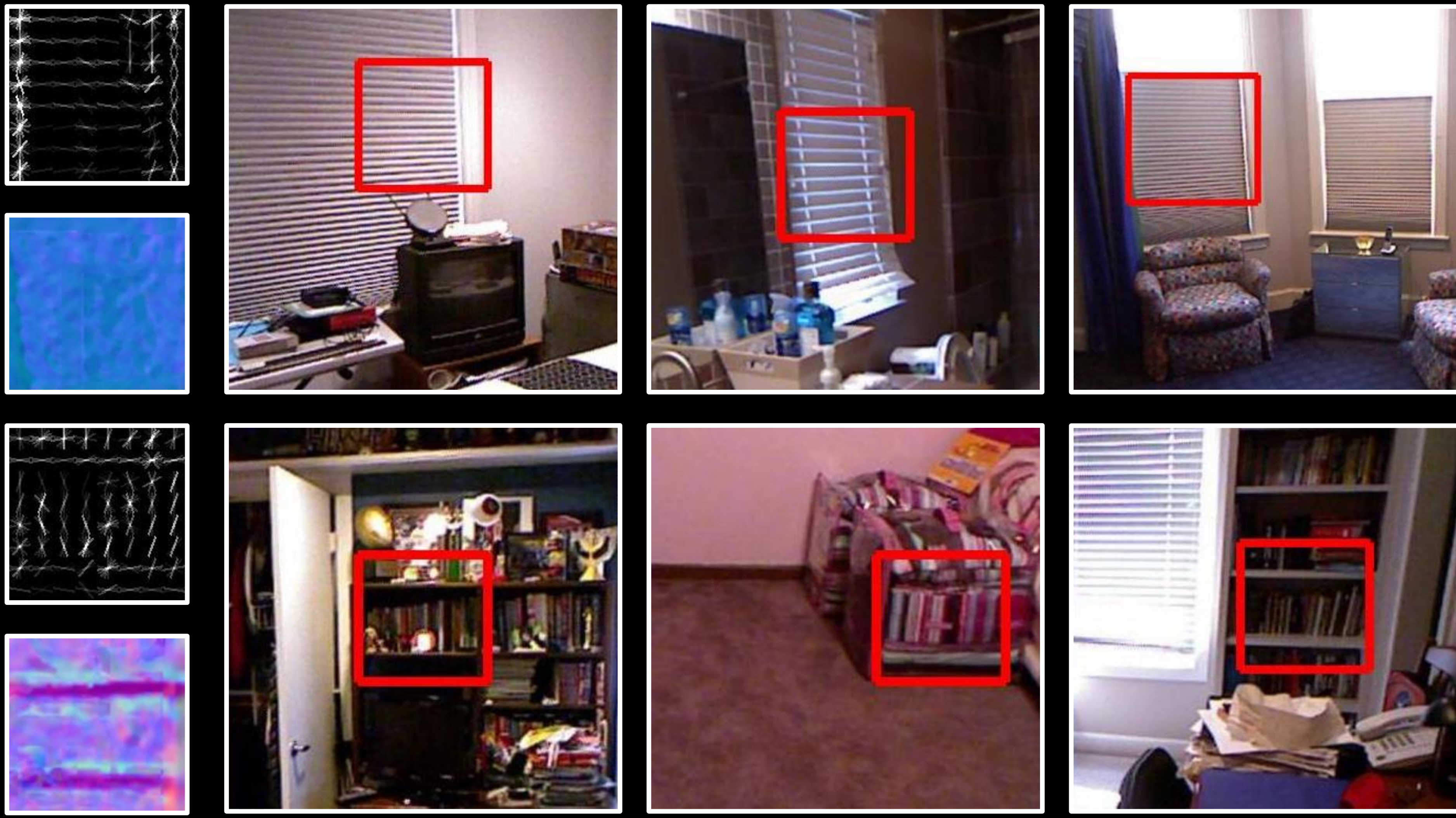
NYU v2 Dataset (Silberman et al., 2012)

Representation



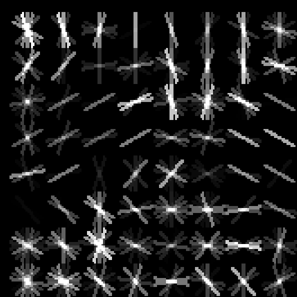






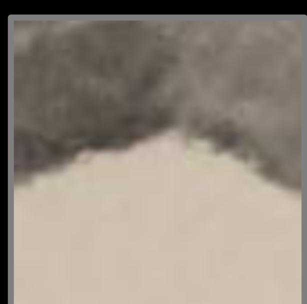
Representation

Detector

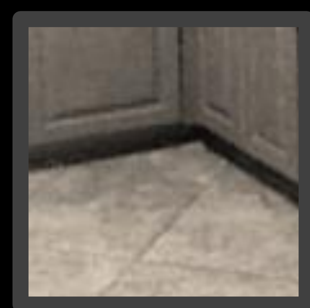


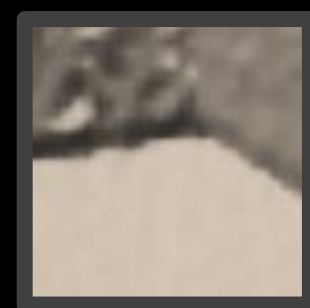
W

Canonical Form



Instances



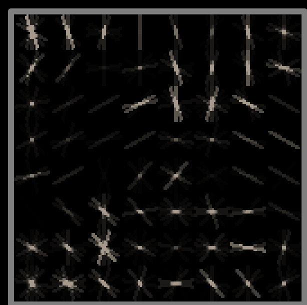


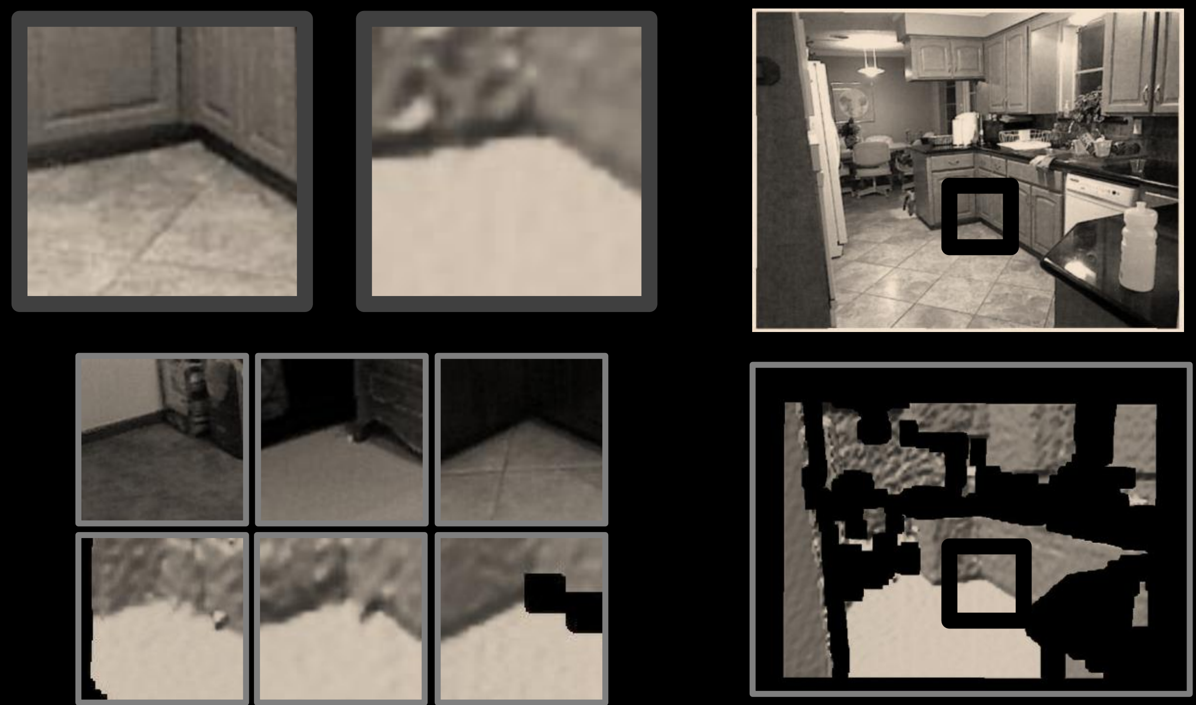
8x8

Representation

Detector

Instances





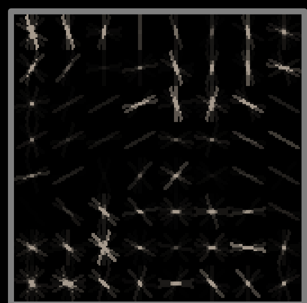
N

Canonical Form

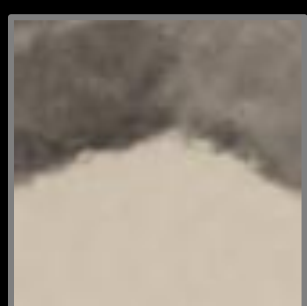
Representation

y

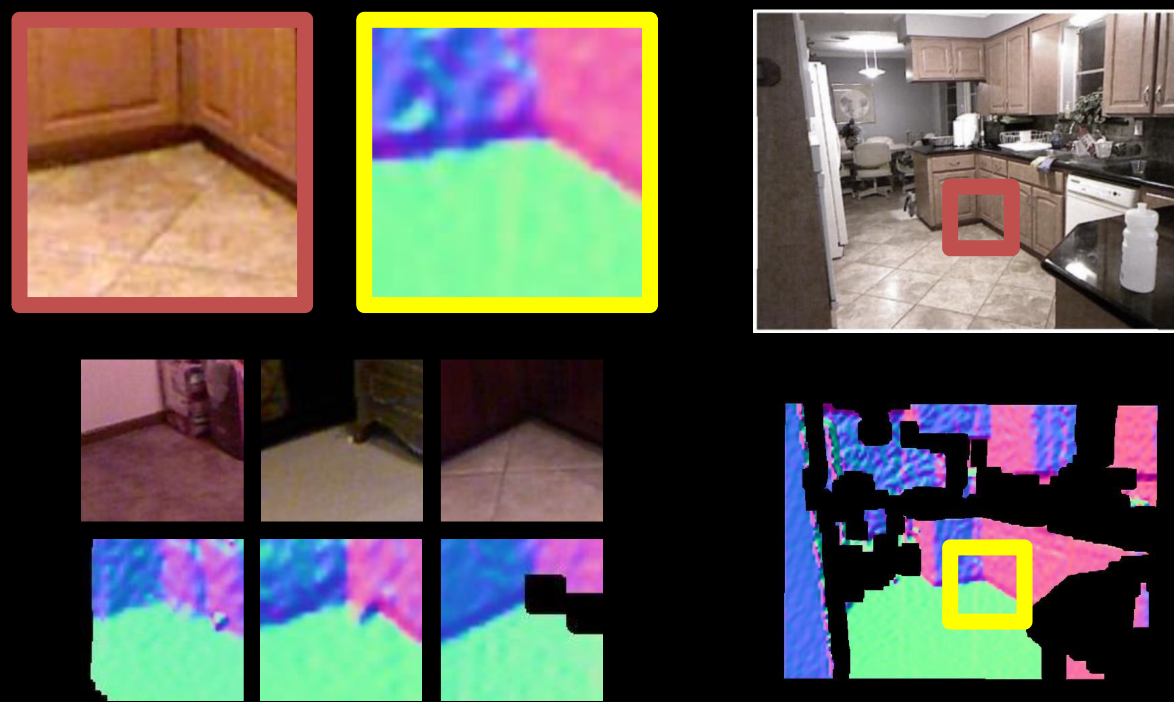
Detector



Canonical Form



Instances

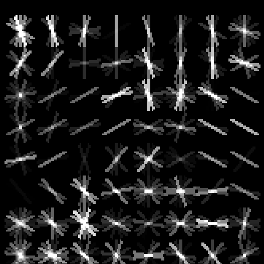


Learning Primitives

$$\min_{\mathrm{y,w,N}} R(w) + \sum_i c_1 y_i \Delta\left(\mathrm{N}, \mathrm{x}_i^G\right) + c_2 L(\mathrm{w}, \mathrm{N}, \mathrm{x}_i^A, y_i)$$

Primitive

$\mathbf{w}$

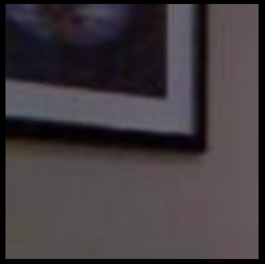


$\mathbf{N}$

$\mathbf{y}$

Patch

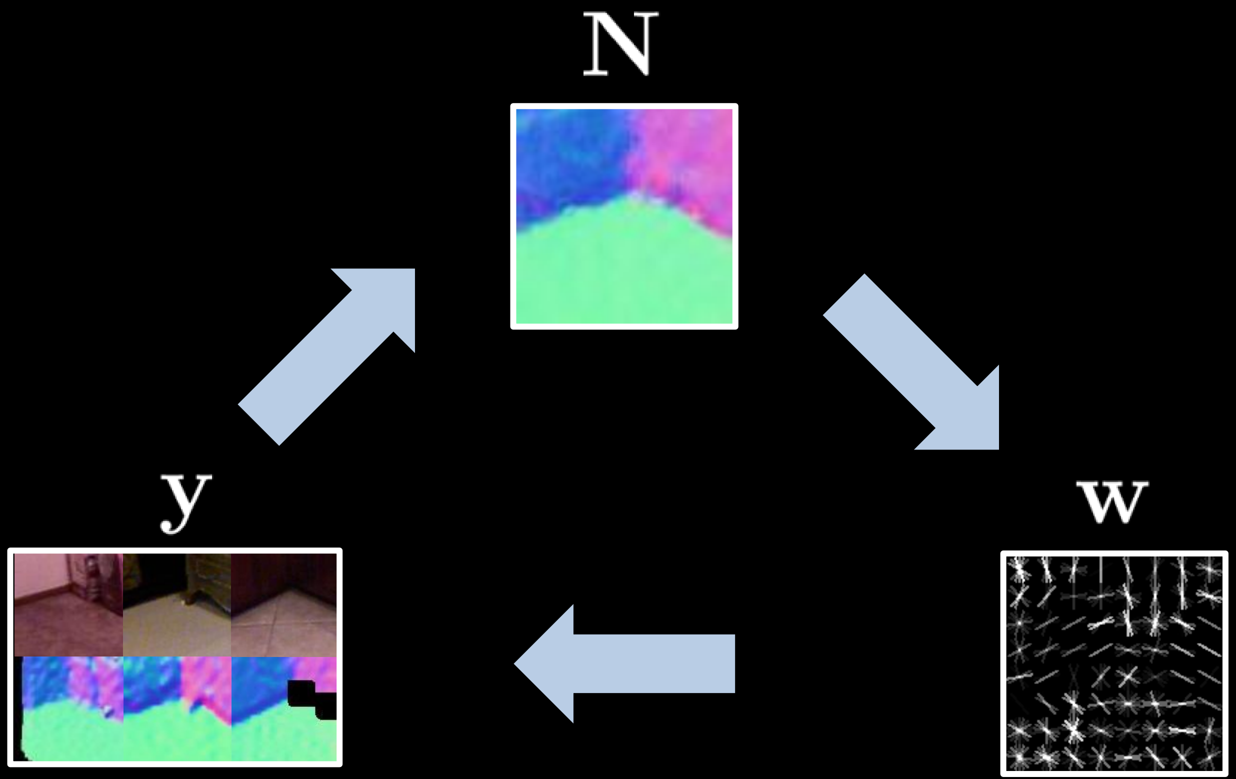
$\mathbf{x}_i^A$



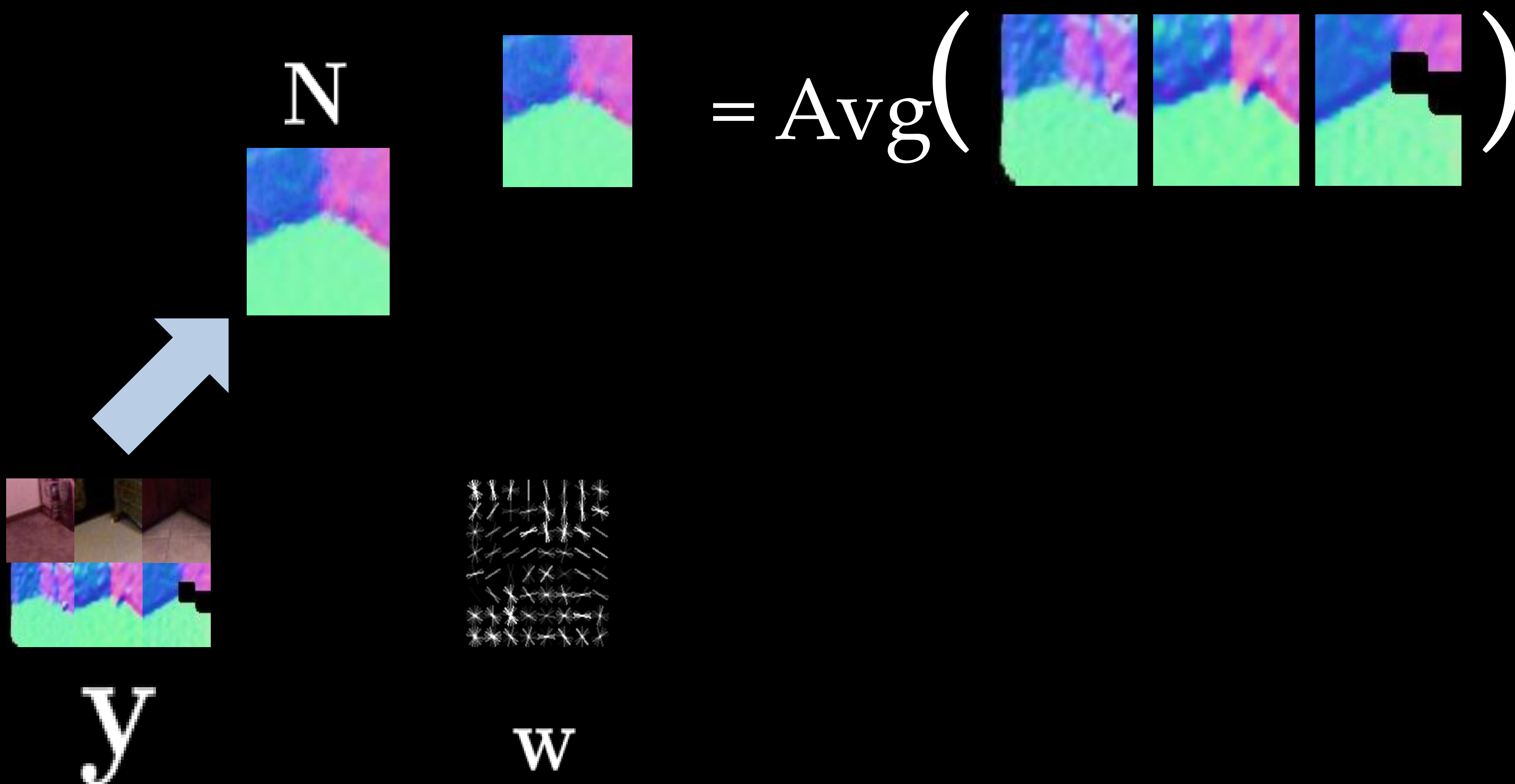
$\mathbf{x}_i^G$

Learning Primitives

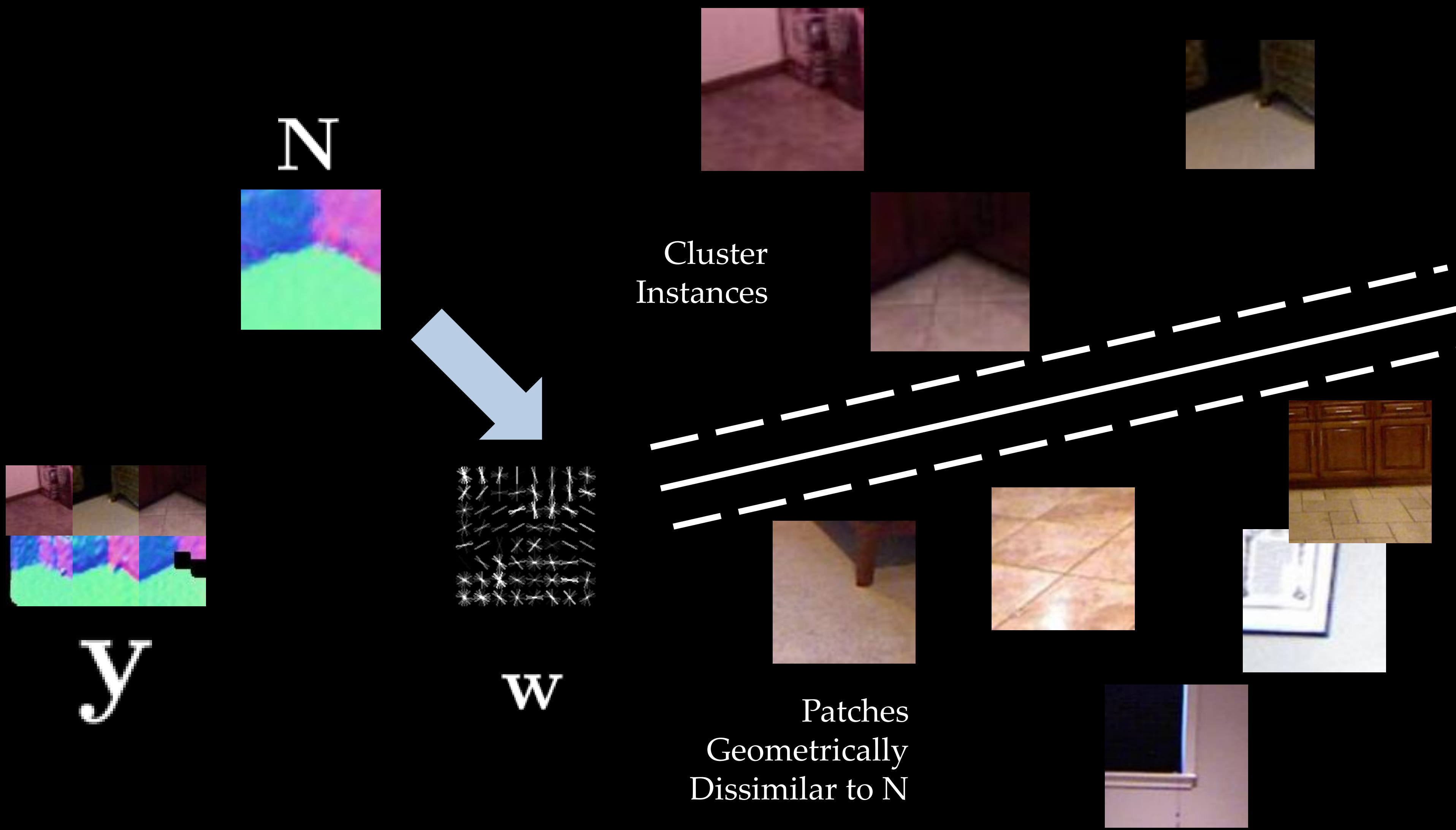
Approach: iterative procedure



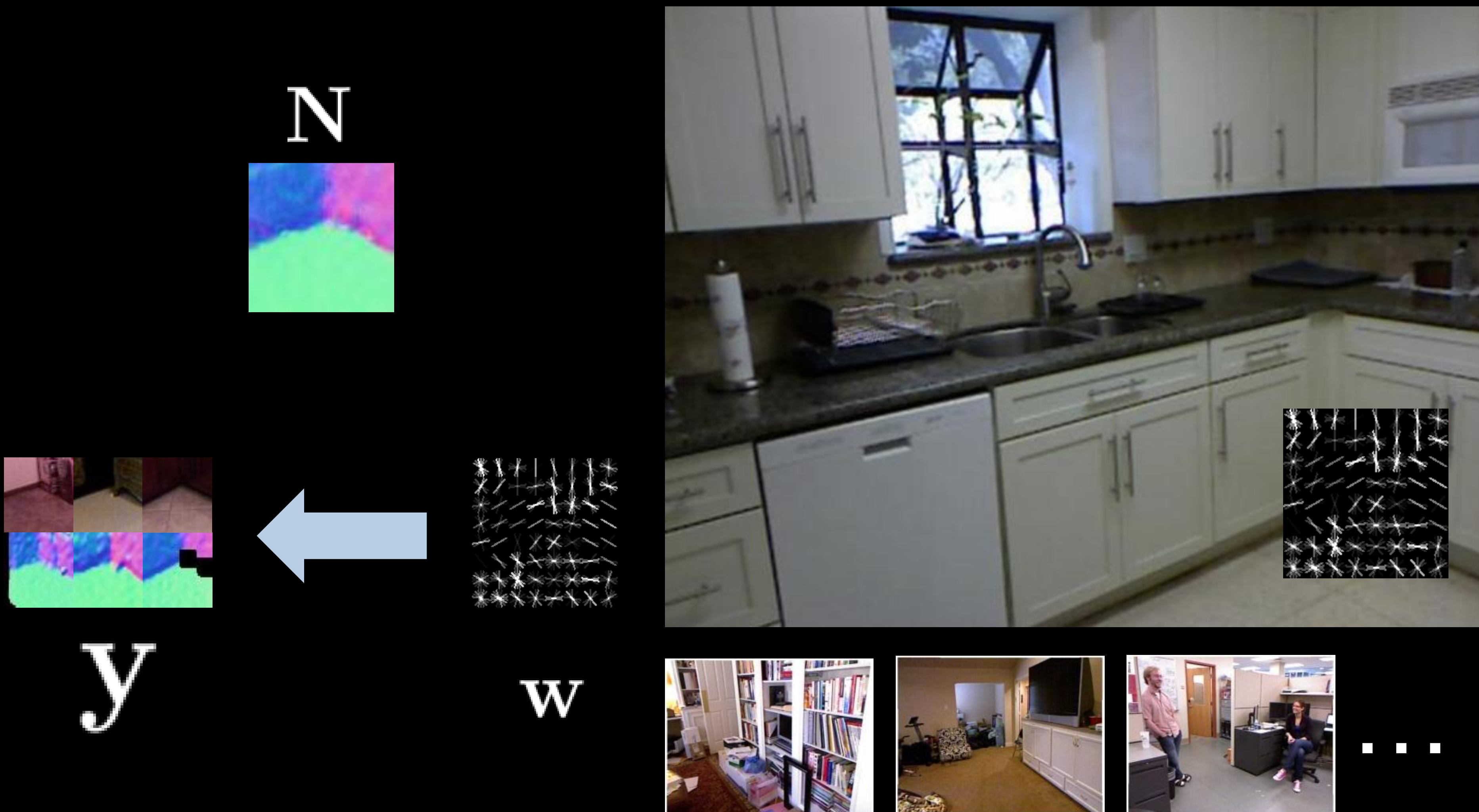
Learning Primitives



Learning Primitives

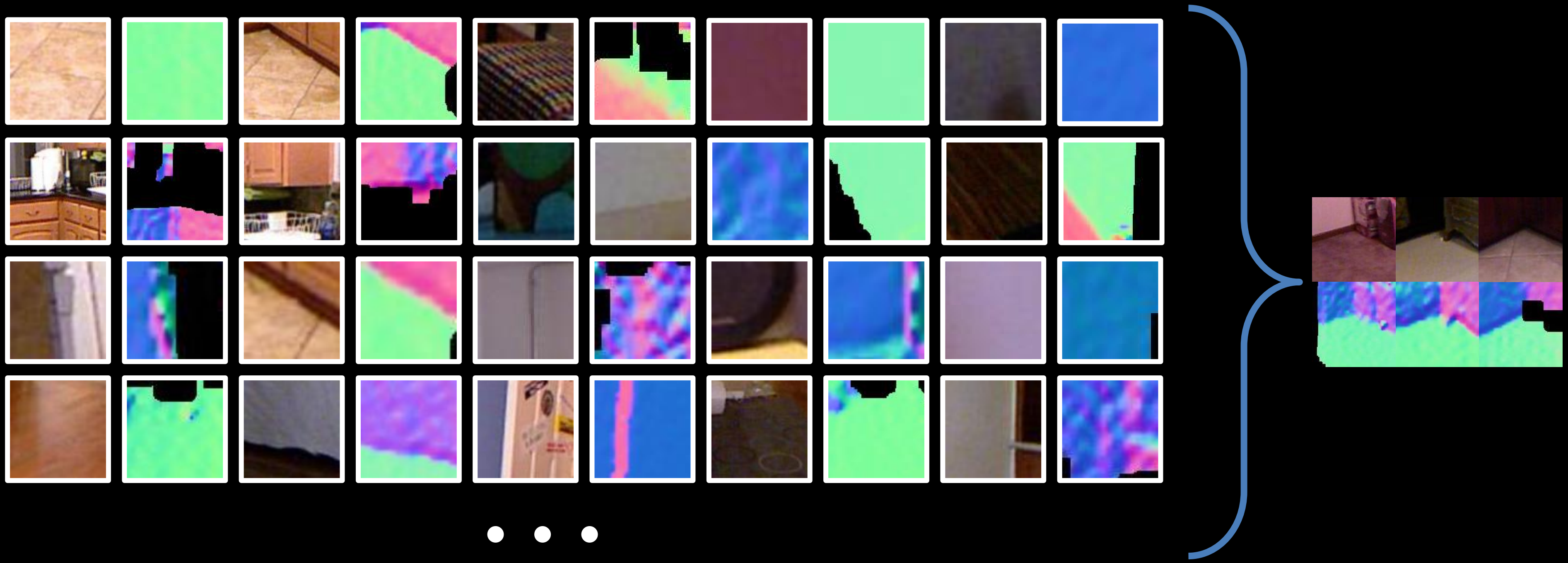


Learning Primitives



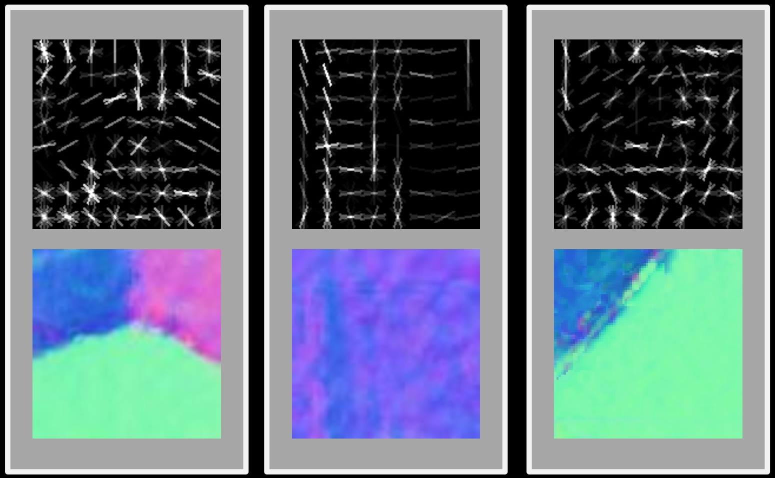
Learning Primitives

Initialize y by clustering sampled patches



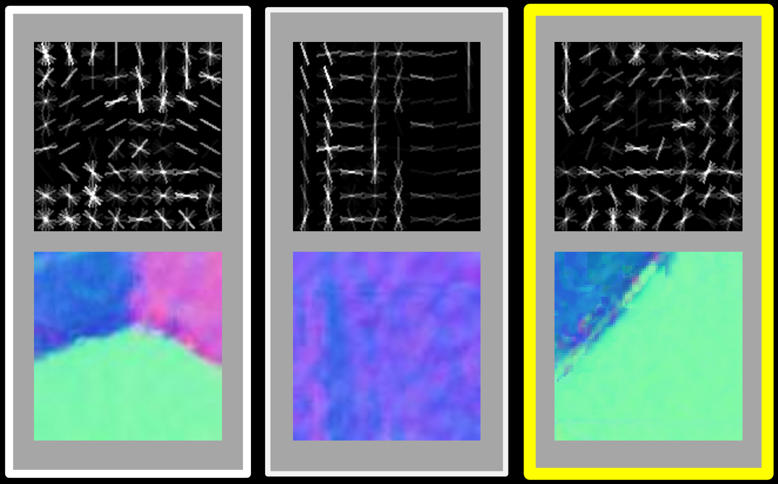
Inference

Sparse Transfer



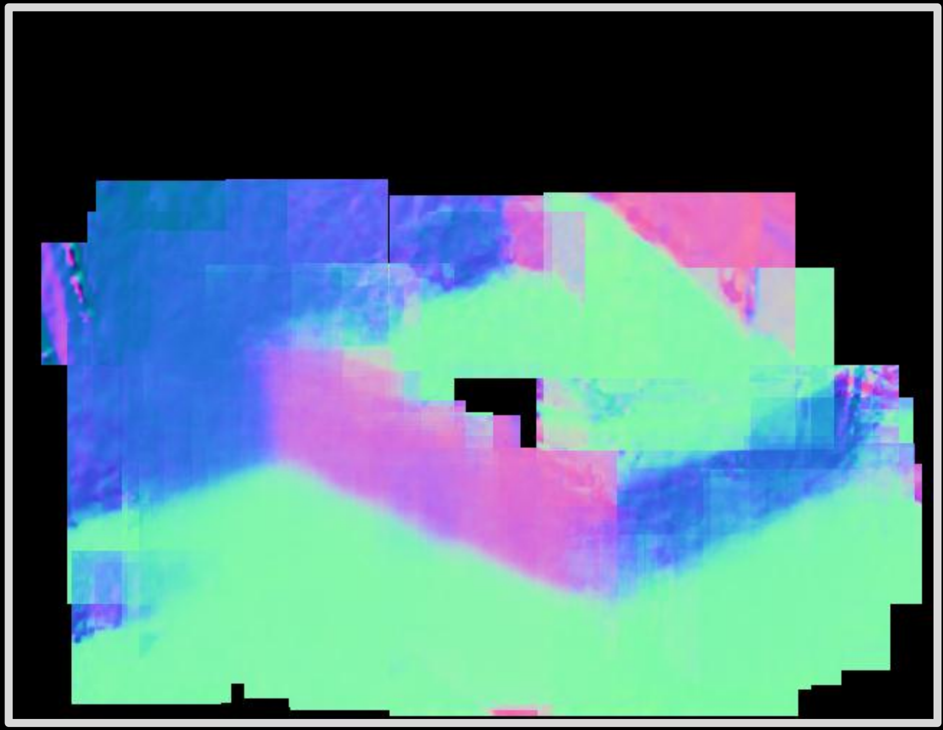
Inference

Sparse Transfer



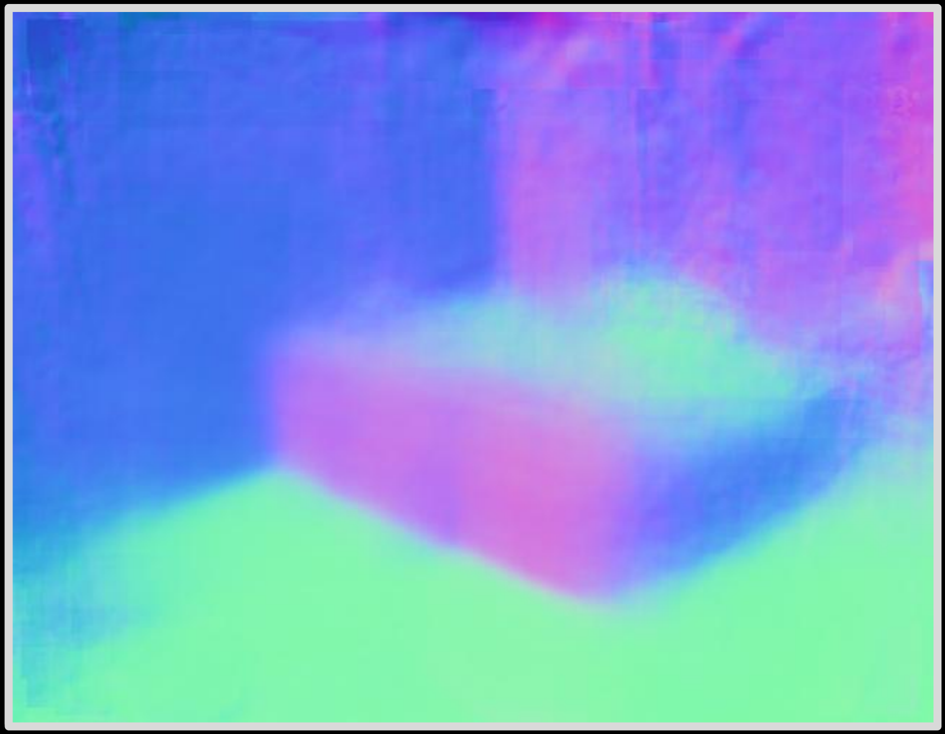
Inference

Sparse Transfer

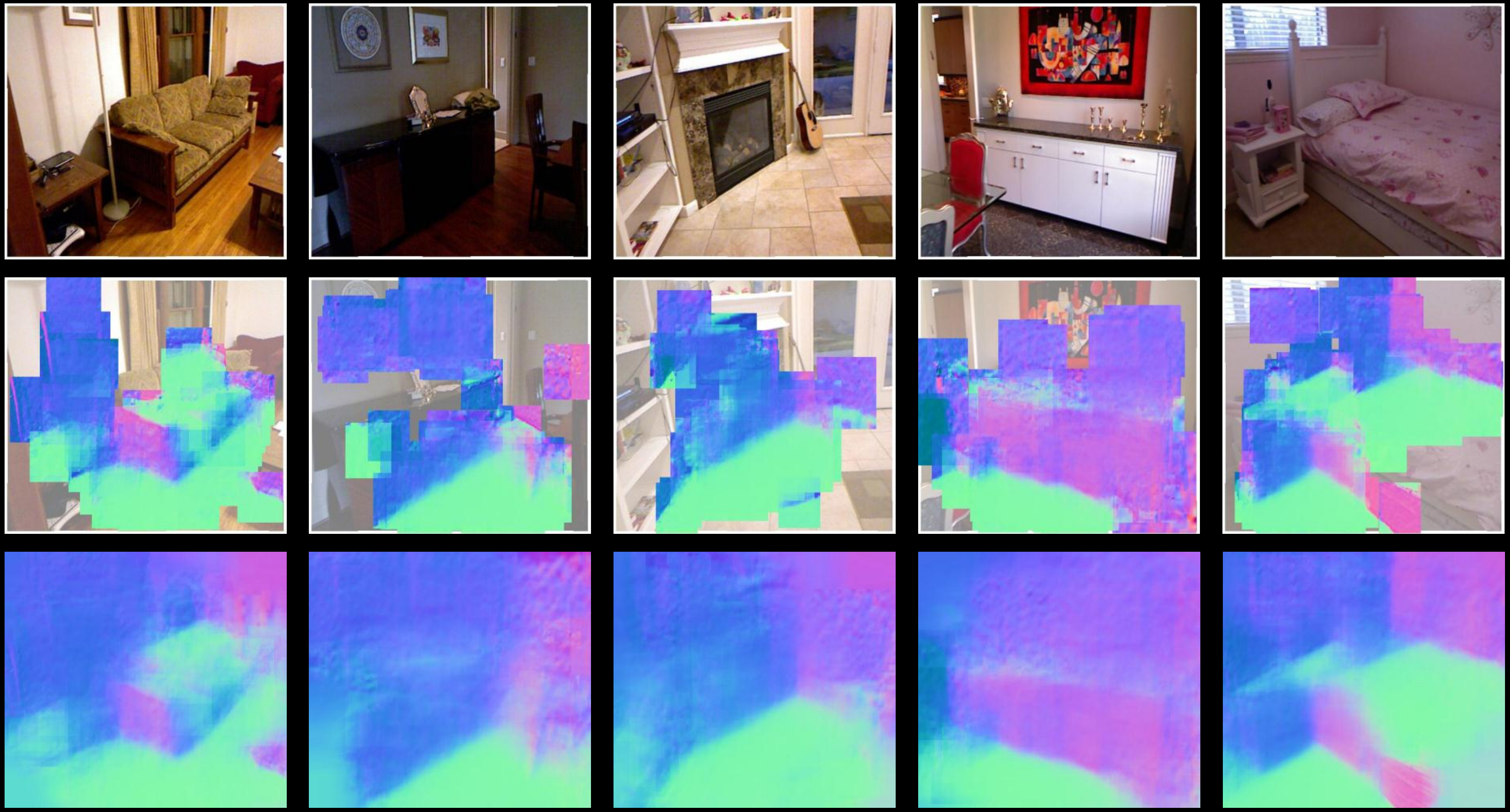


Inference

Dense Transfer

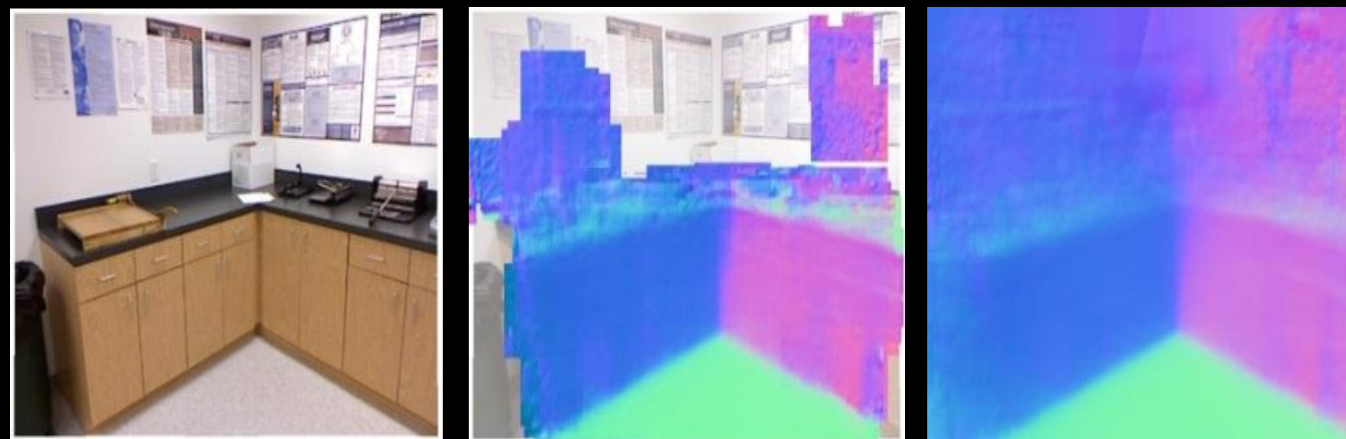


Sample Results – Qualitative



Confidences

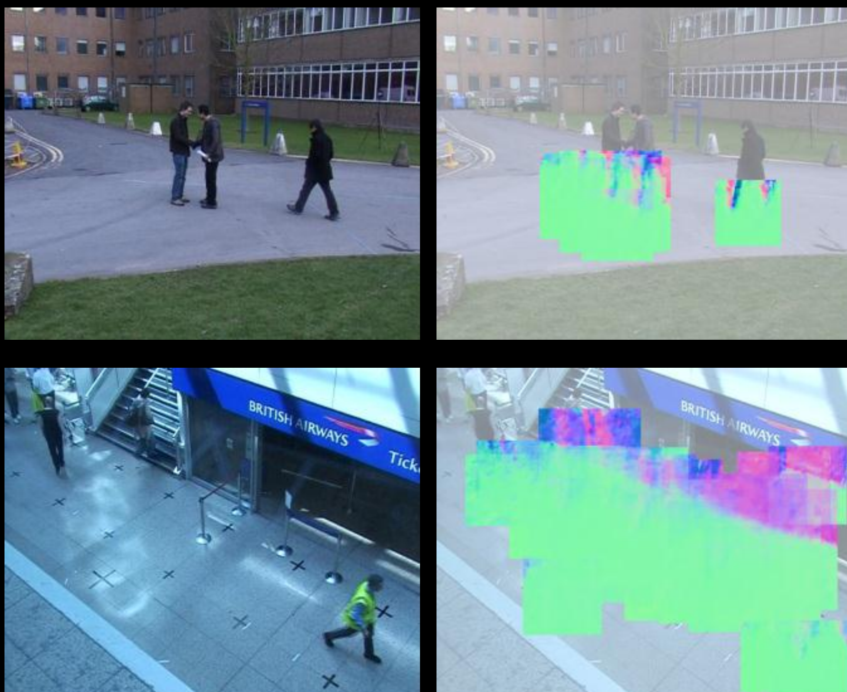
Most Confident Result



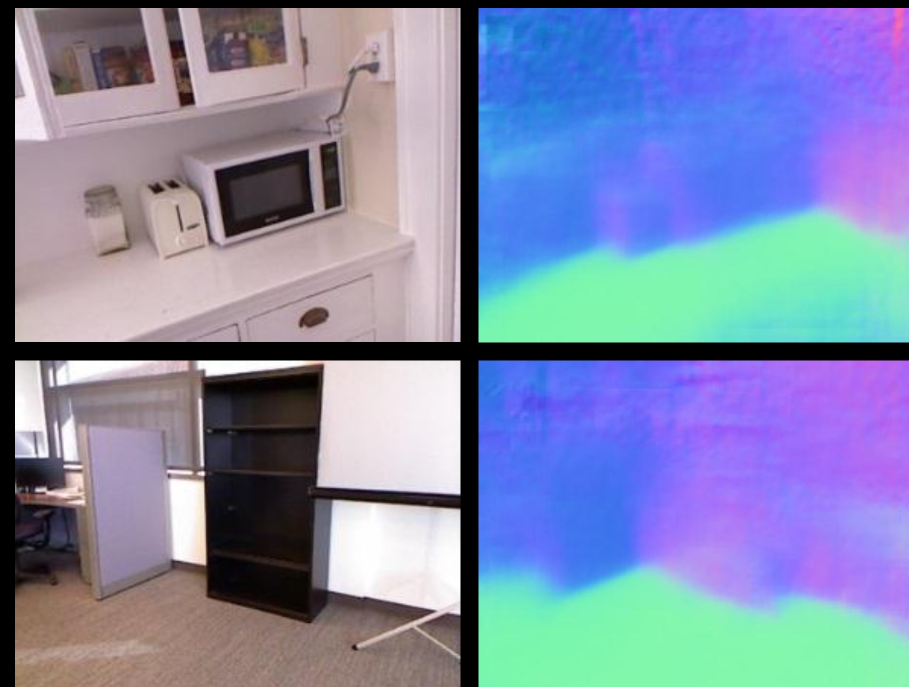
Least Confident Result

rank

Cross-dataset

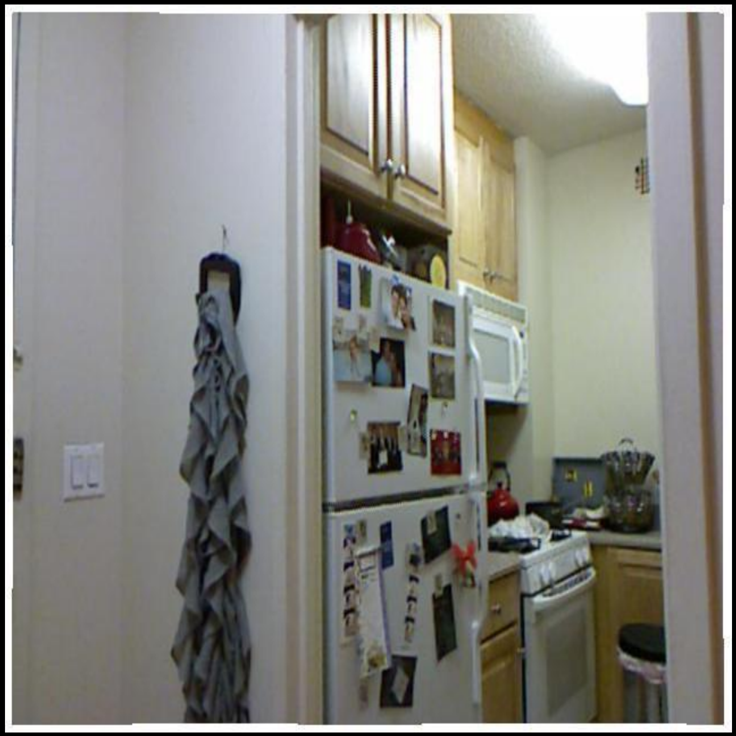


PETS



B3DO

Failures

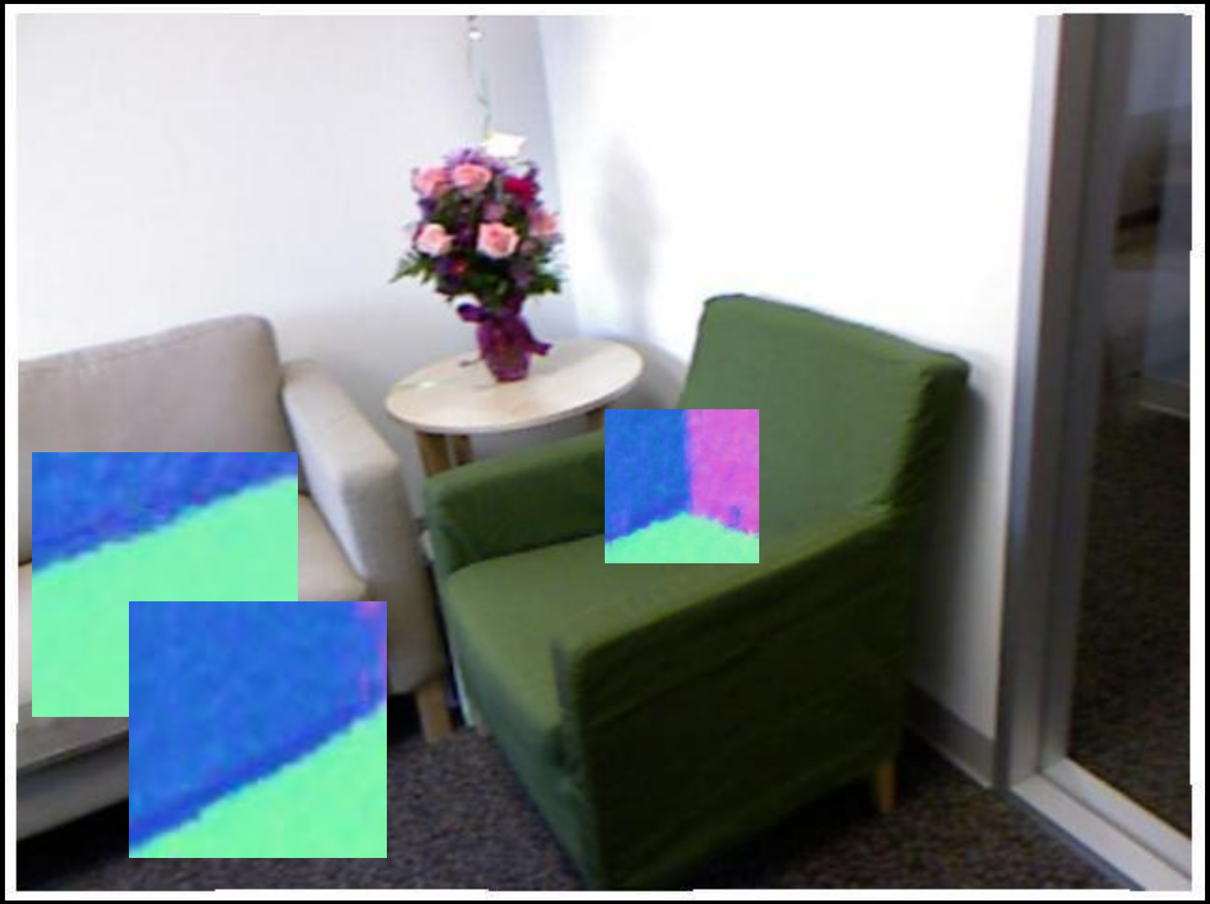


	Summary Stats (^{0}) (Lower Better)			% Good Pixels (Higher Better)		
	Mean	Median	RMSE	11.25^0	22.5^0	30^0
3D Primitives	**33.0**	**28.3**	**40.0**	**18.8**	**40.7**	**52.4**
Singh et al.	35.0	32.4	40.6	11.2	32.1	45.8
Karsch et al.	40.8	37.8	46.9	7.9	25.8	38.2
Hoiem et al.	41.2	34.8	49.3	9.0	31.7	43.9
Saxena et al.	47.1	42.3	56.3	11.2	28.0	37.4
RF + Dense SIFT	36.0	33.4	41.7	11.4	31.1	44.2

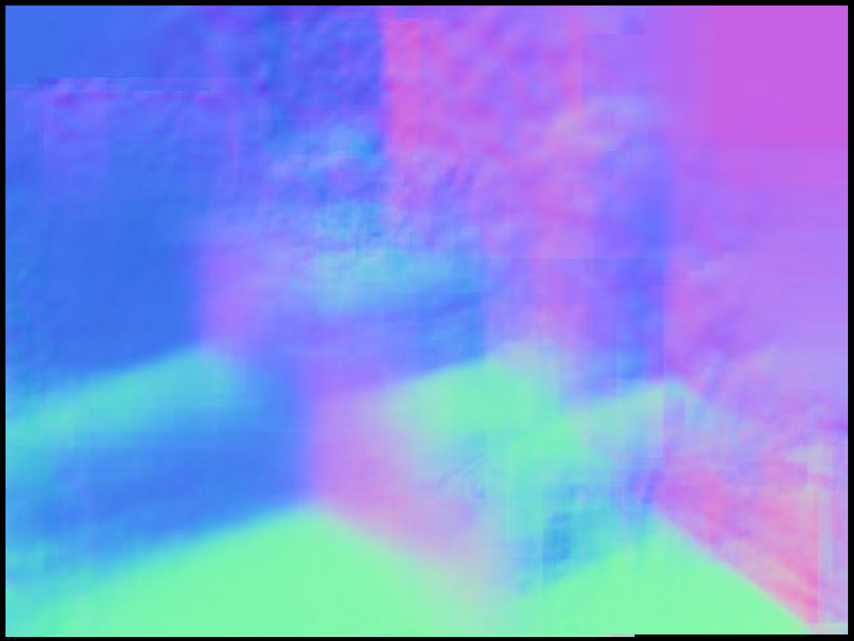
RMSE

Using geometric and physical constraints

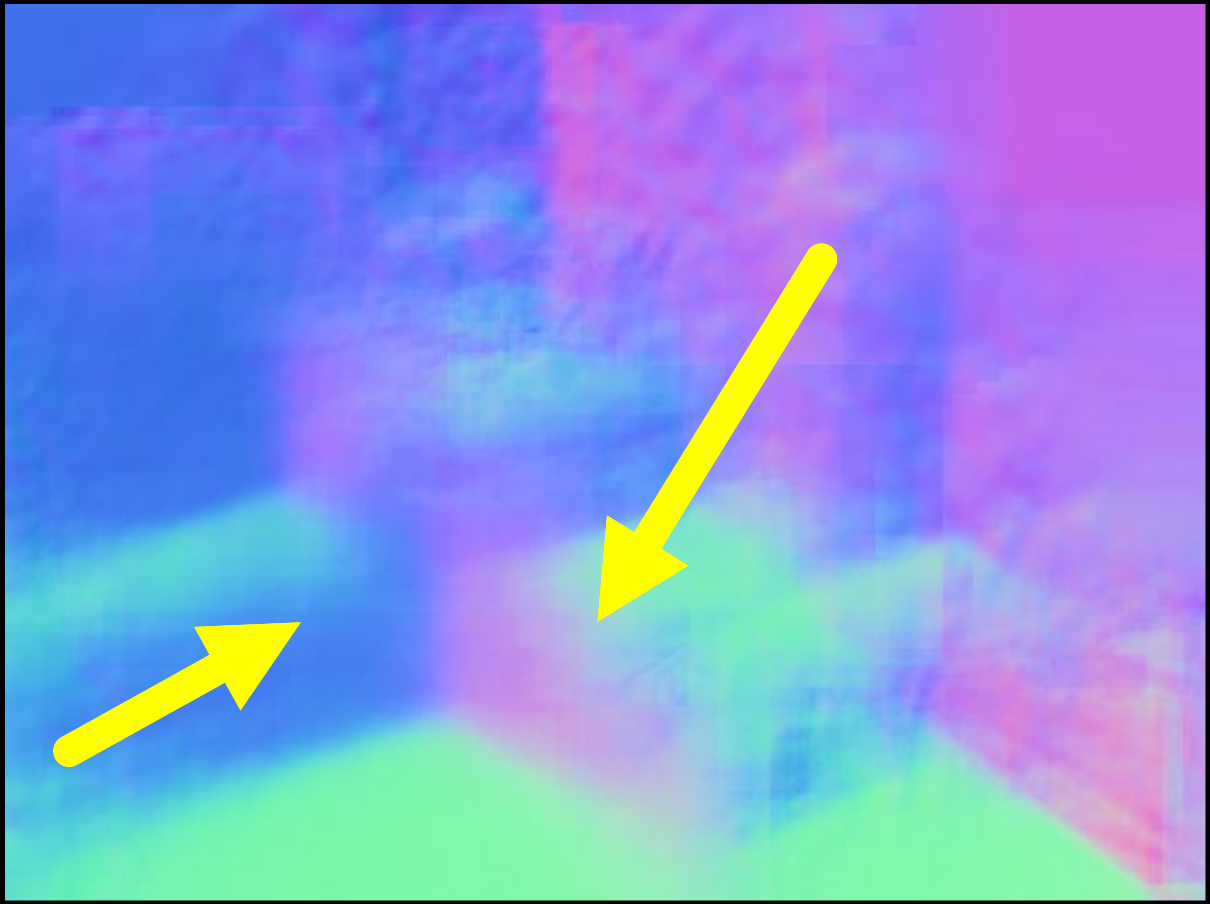
The Story So Far (Sparse)



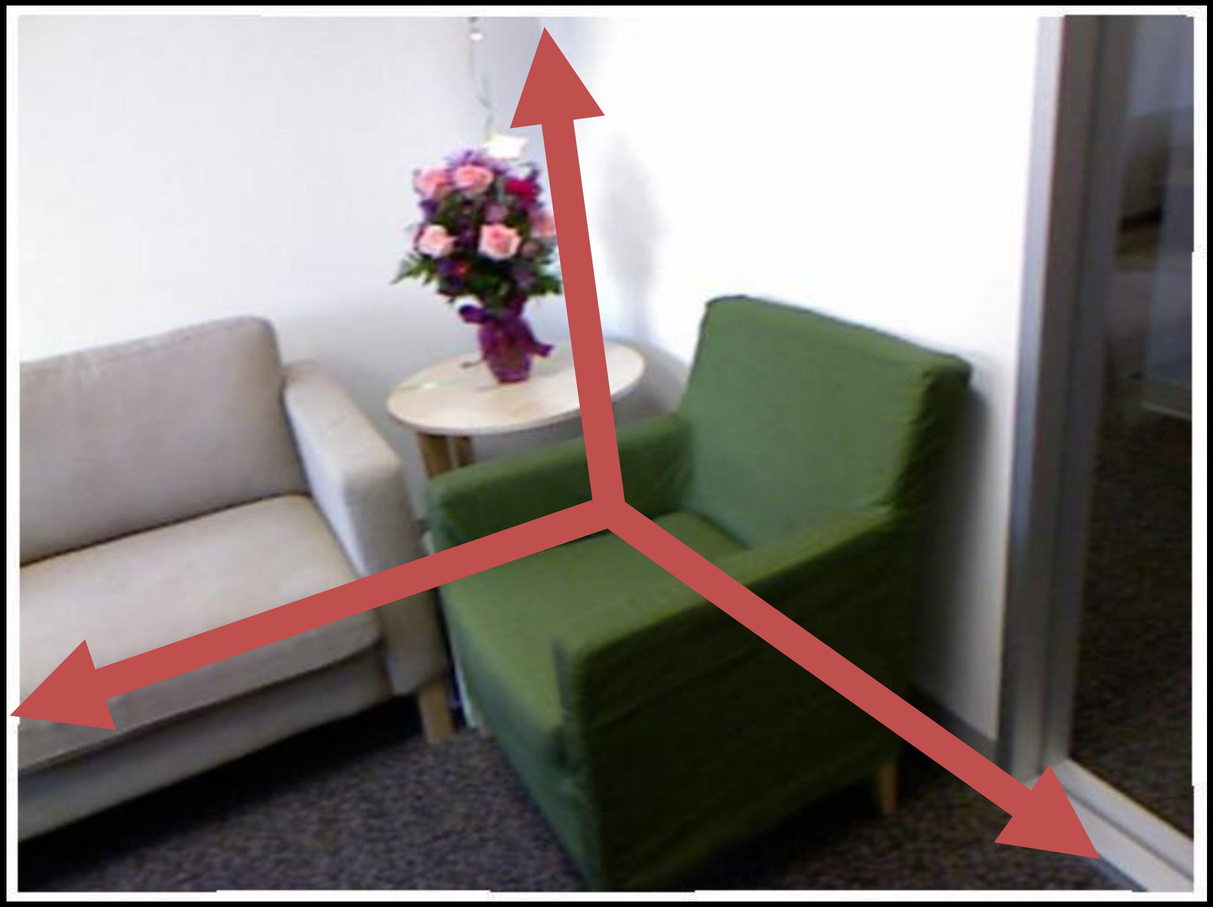
The Story So Far (Dense)



The Story So Far



Adding Physical/Geometric Constraints

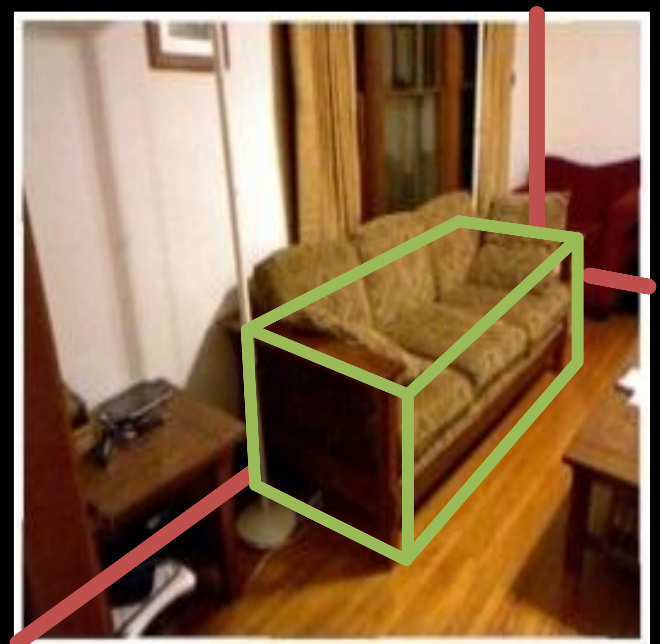


Adding Physical/Geometric Constraints

Past Physical Constraints

Camera-in-a-box

Hedau et al. 2009, Flint et al. 2011, Satkin et al. 2012, Schwing et al. 2012, etc.

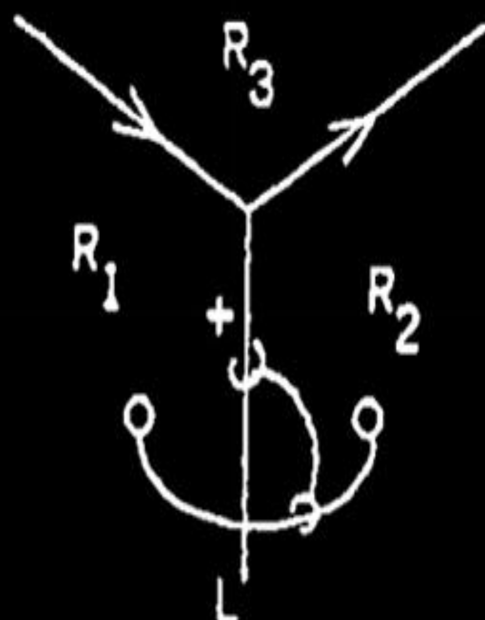


Top-down Cuboid

Lee et al. 2010, Gupta et al. 2010, Xiao et al. 2012, etc.

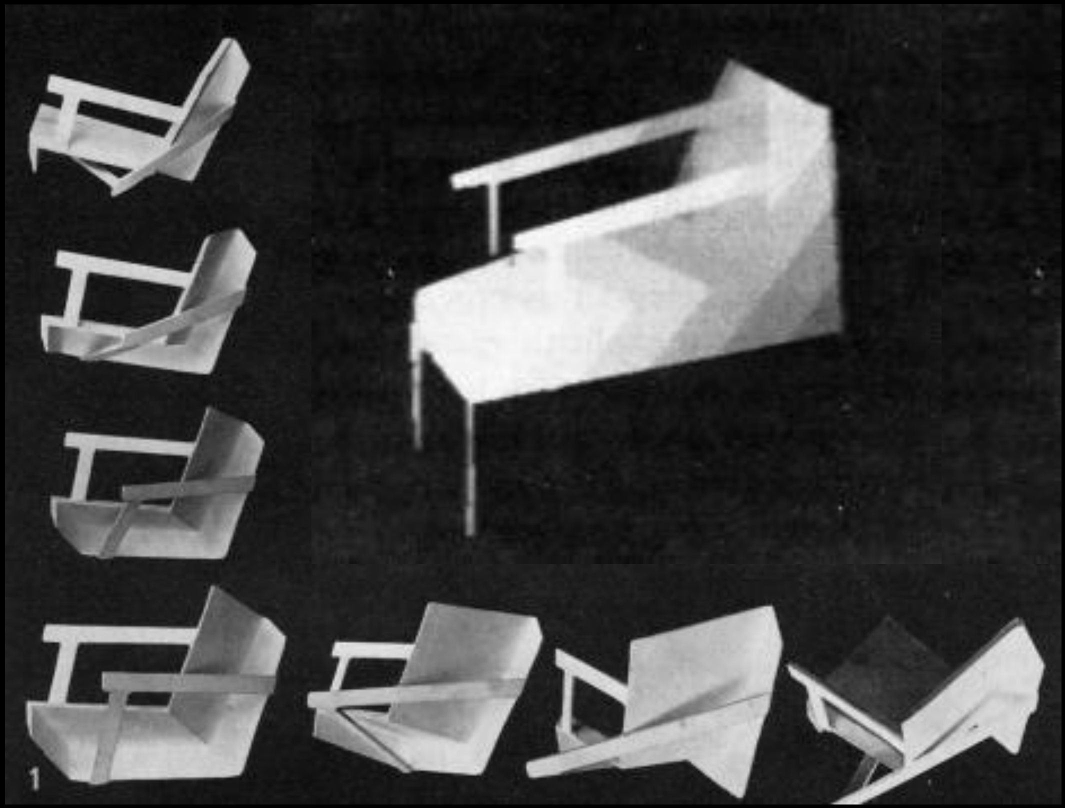
Digression: Inspiration from the past….

Kanade's Origami World, 1978



From the past….

- Kanade's chair… (Artificial Intelligence, 1981)

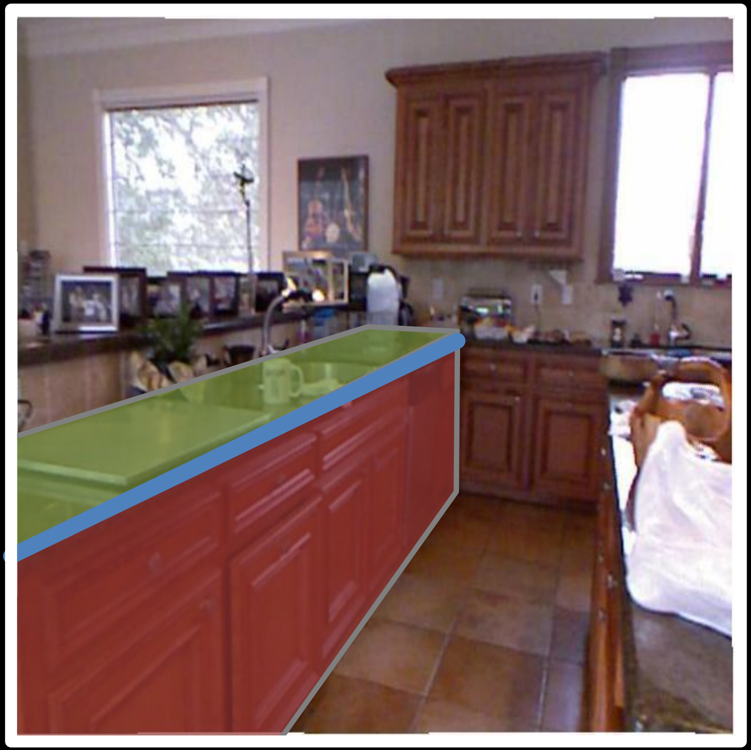


Edges between surfaces

Concave
(-)

Convex
(+)

Edges between surfaces

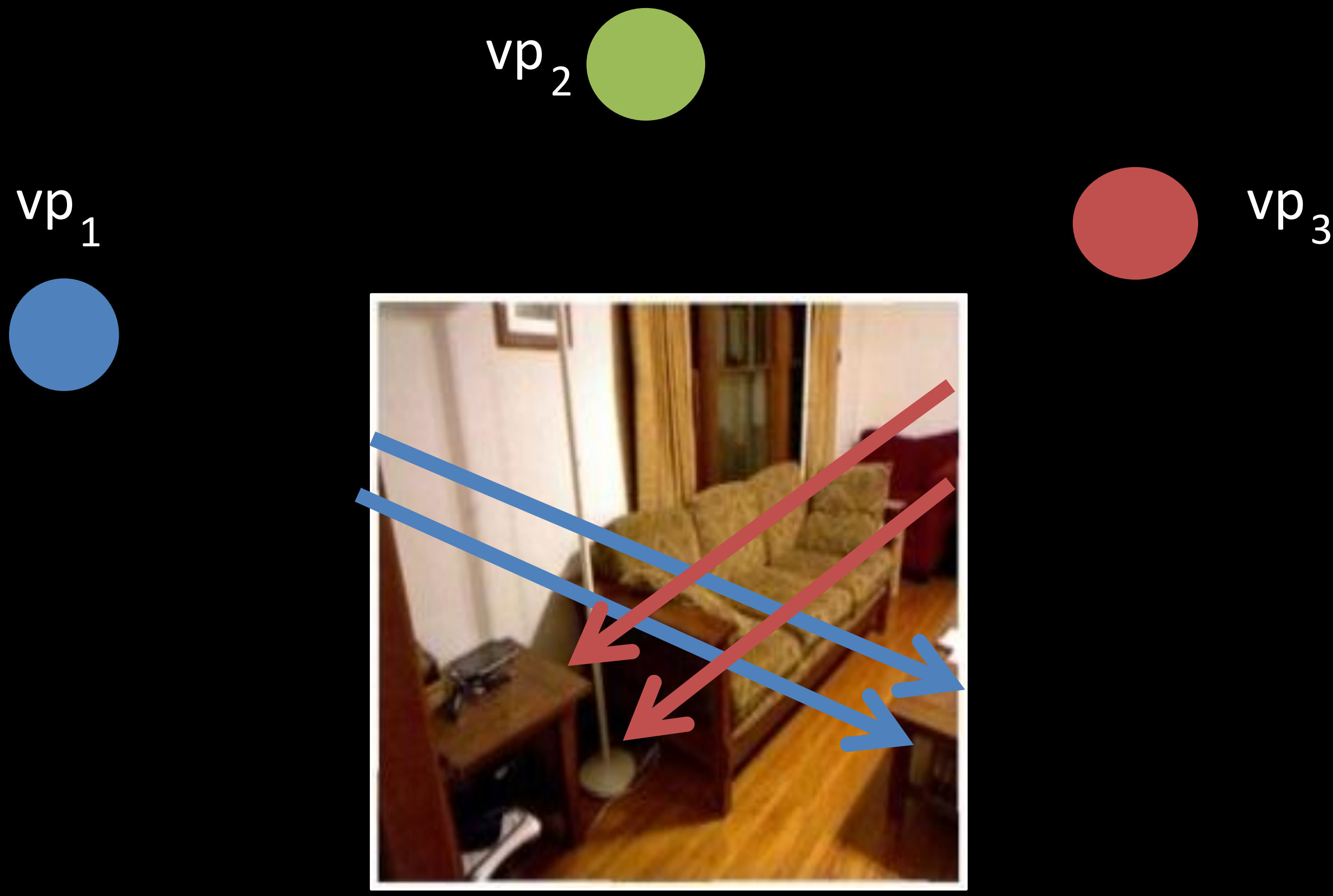


Concave
(-)

Convex
(+)

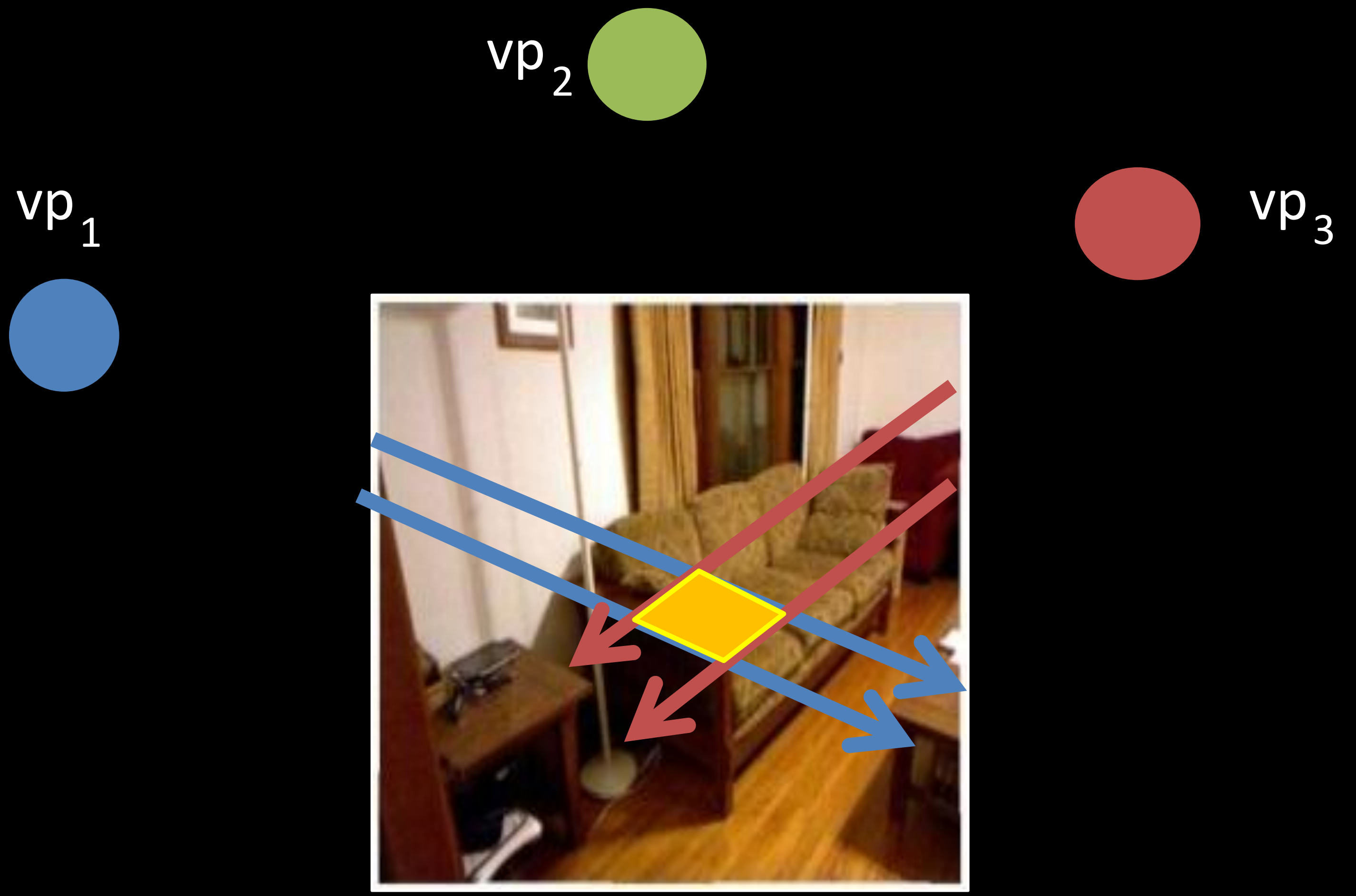
Parameterization

Parameterization

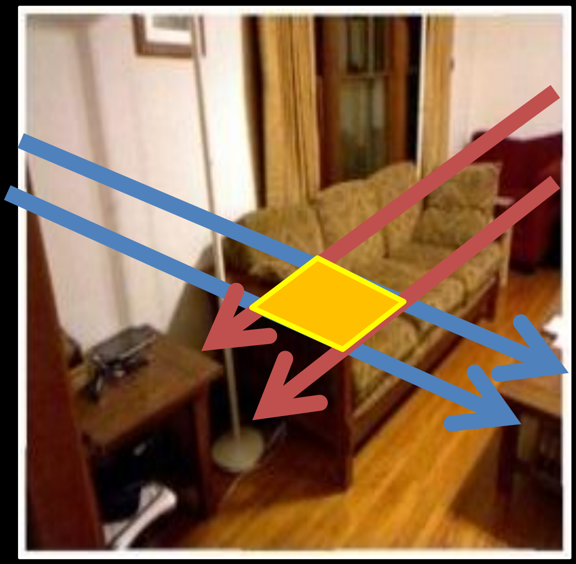


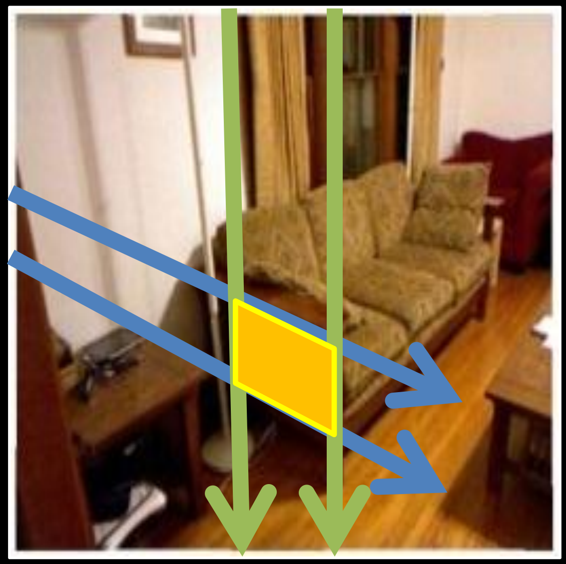
Schwing 2013, Hedau 2010

Parameterization

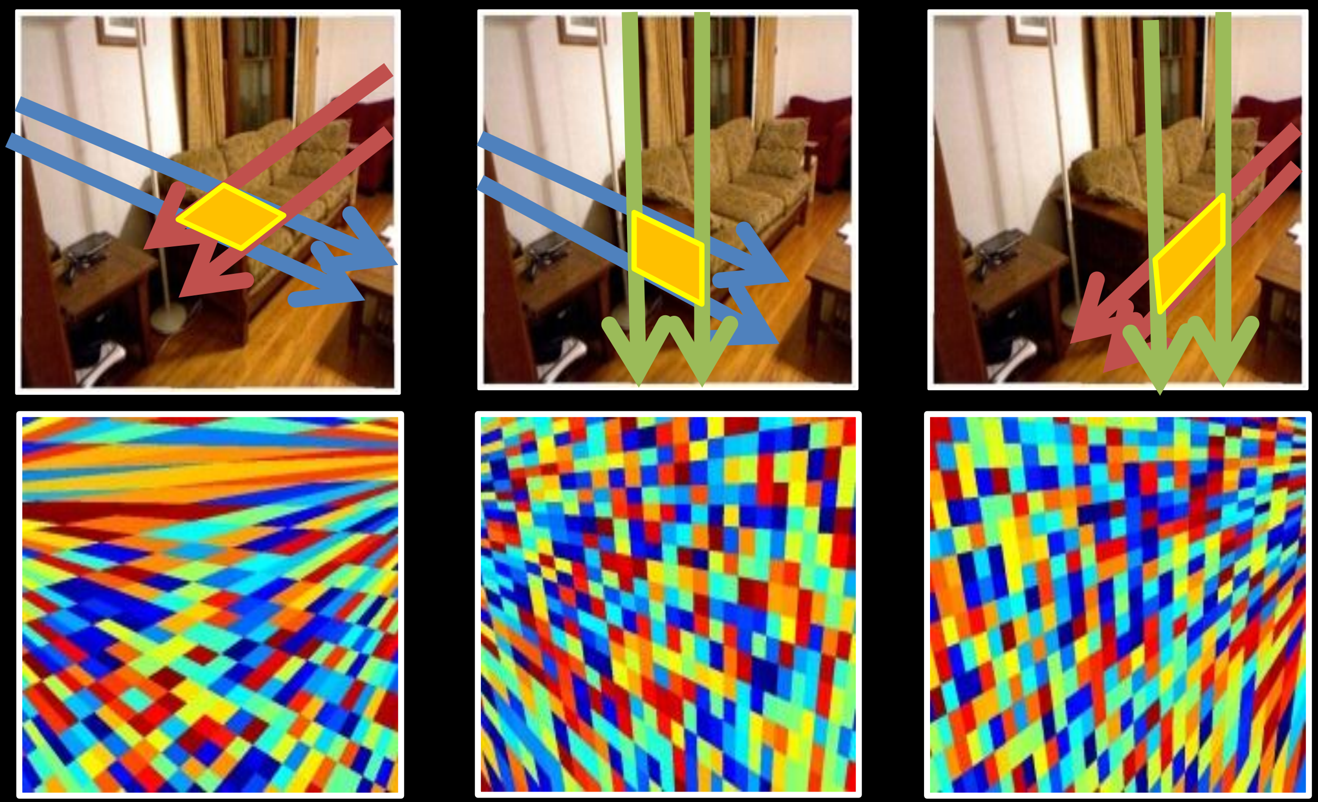


Parameterization

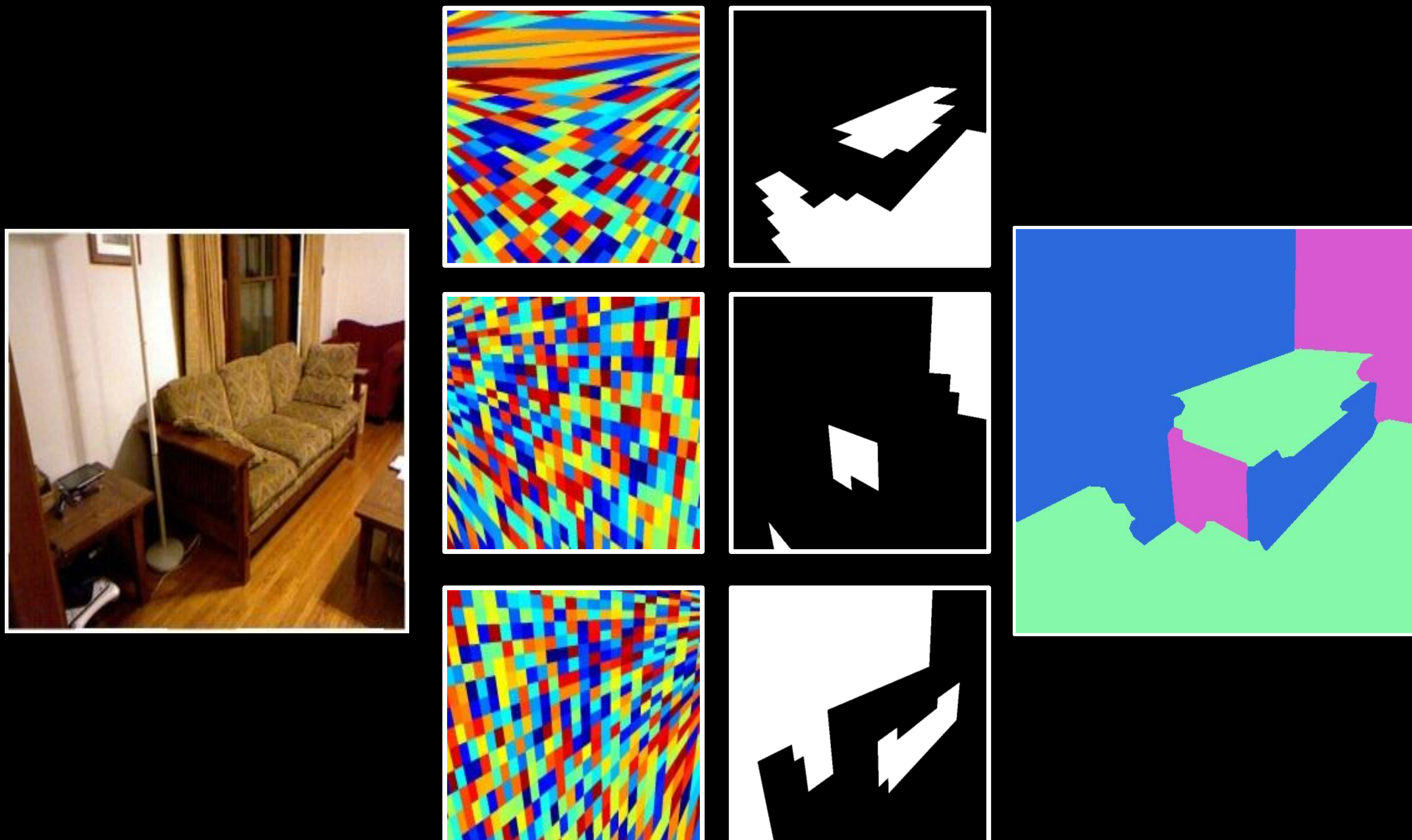




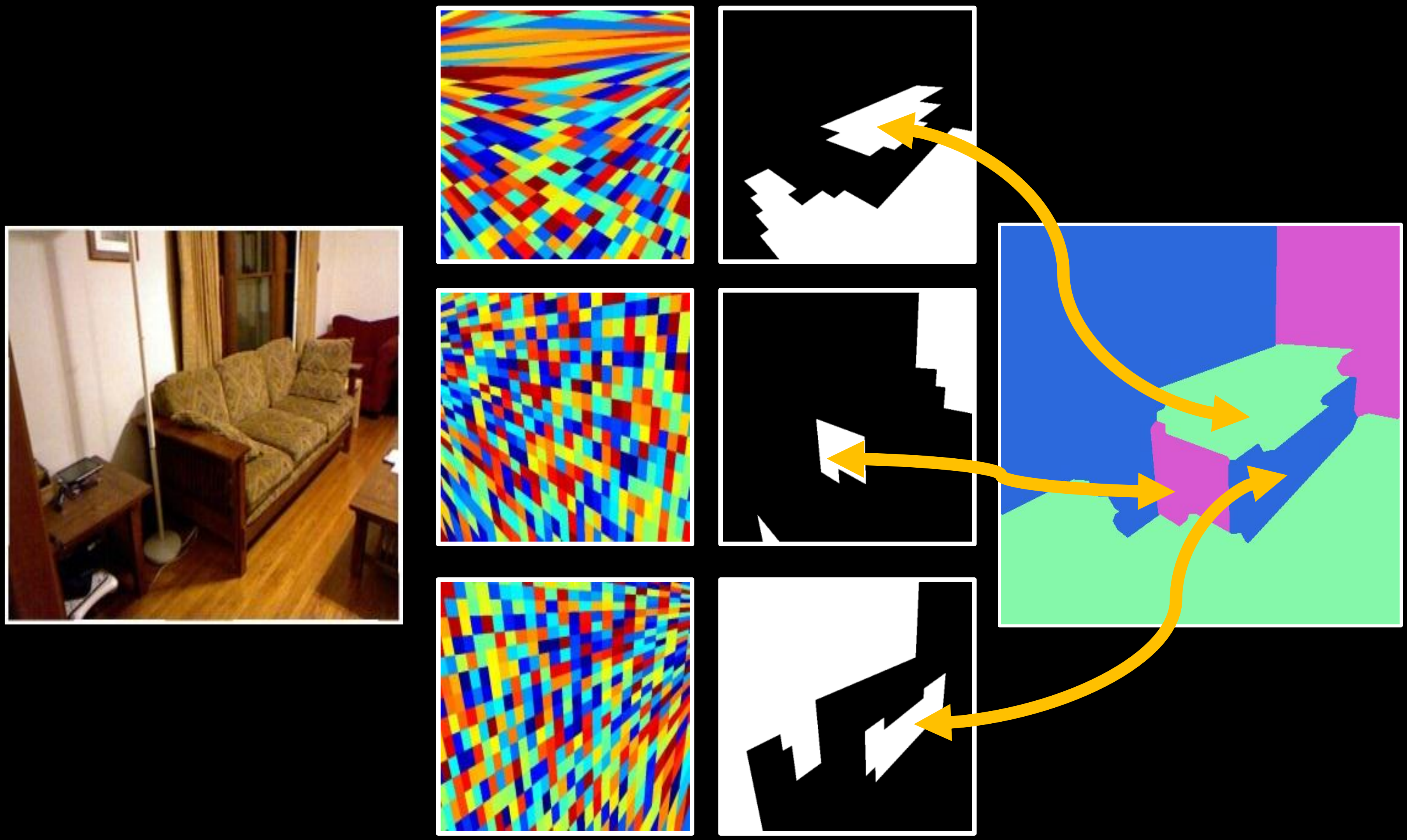
Parameterization



Parameterization

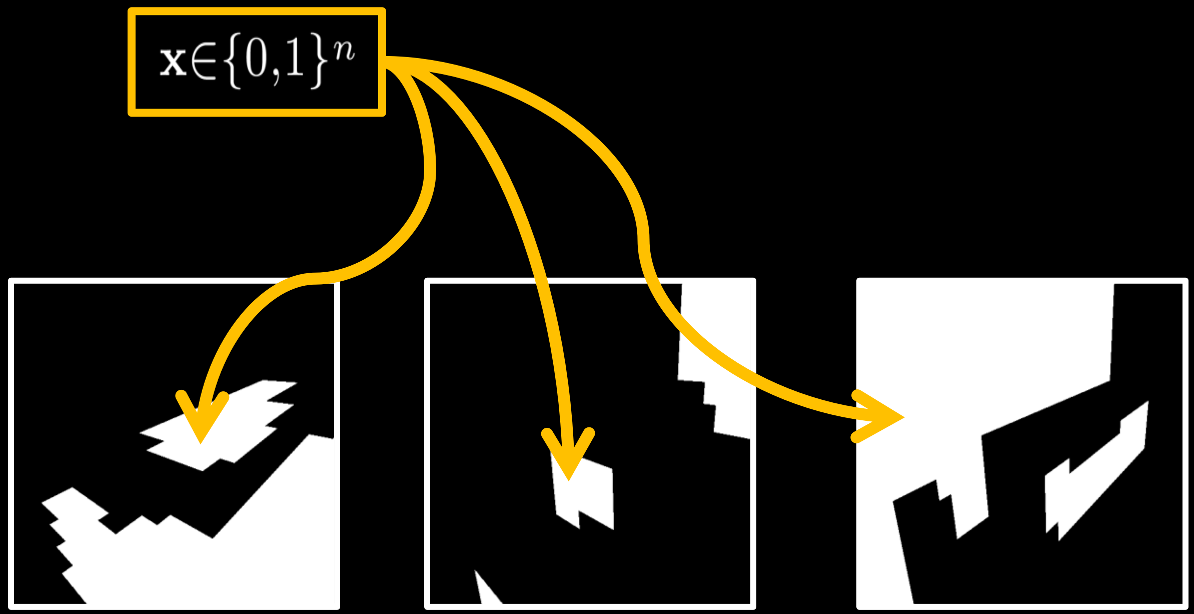


Parameterization



Labeling

x_i: is cell i on?



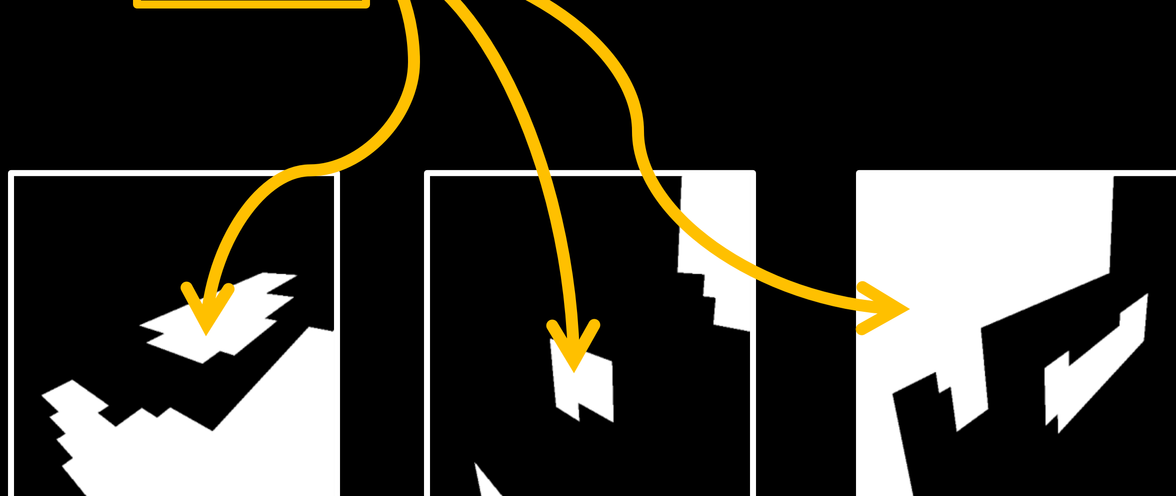
Formulation

$$\underset{\mathbf{x}\in\{0,1\}^n}{\arg\max}\, \mathbf{c}^T\mathbf{x} + \mathbf{x}^T\mathbf{H}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{1}$$

Variable

x_i: is cell i on?

$$\underset{\boxed{\mathbf{x}\in\{0,1\}^n}}{\arg\max}\, \mathbf{c}^T\mathbf{x}+\mathbf{x}^T\mathbf{H}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x}\le \mathbf{1}$$



Unary Potentials

c_i: should cell i be on?

$$\arg\max_{\mathbf{x}\in\{0,1\}^n} \boxed{\mathbf{c}^T\mathbf{x}} + \mathbf{x}^T\mathbf{H}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{1}$$

Binary Potentials

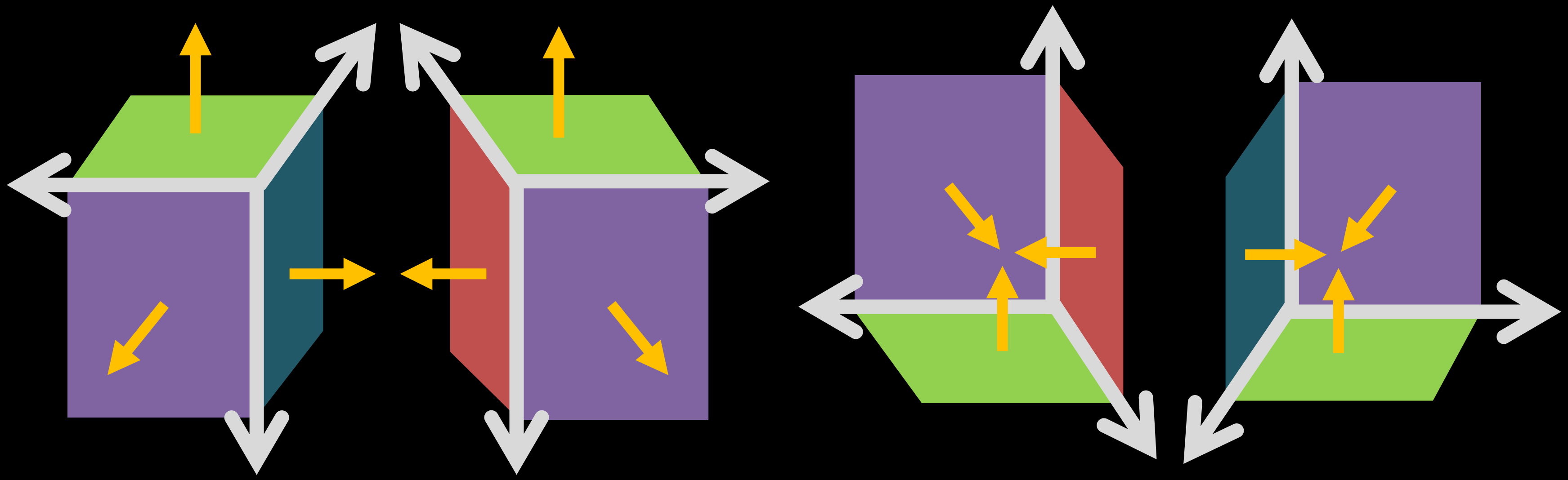
$H_{i,j}$: should cells *i* and *j* both be on?

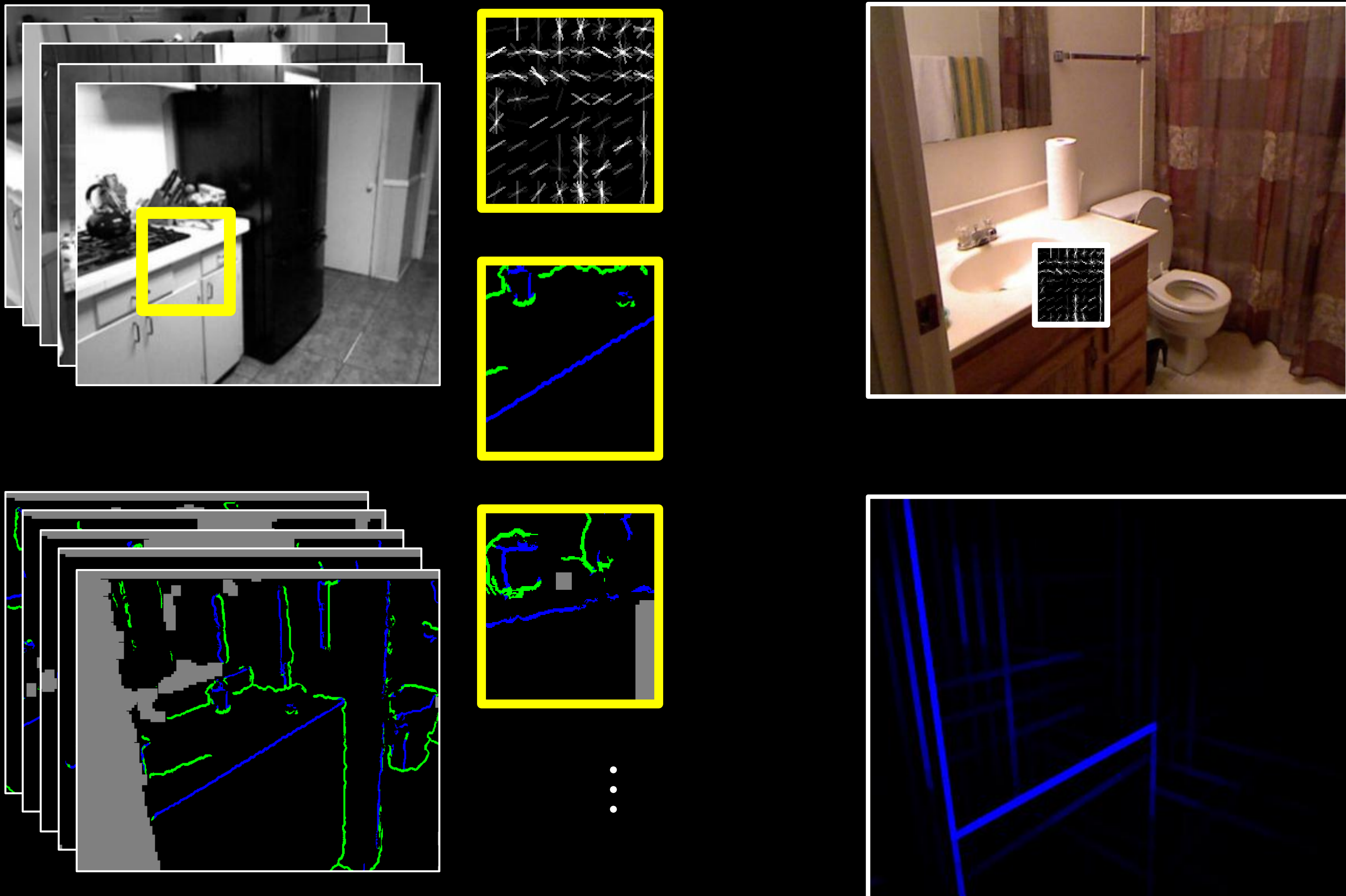
$$\arg\max_{\mathbf{x}\in\{0,1\}^n} \mathbf{c}^T\mathbf{x} + \boxed{\mathbf{x}^T\mathbf{H}\mathbf{x}} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{1}$$

Binary Potentials

Convex (+)

Concave (-)

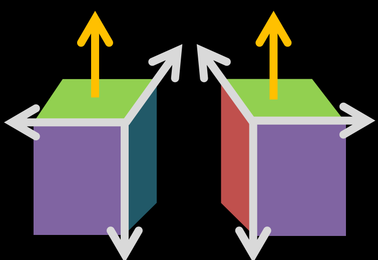




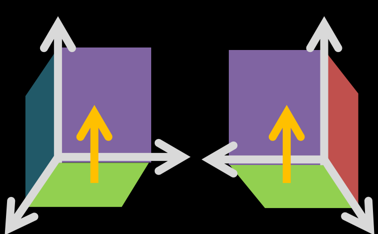
8o7s+UCM

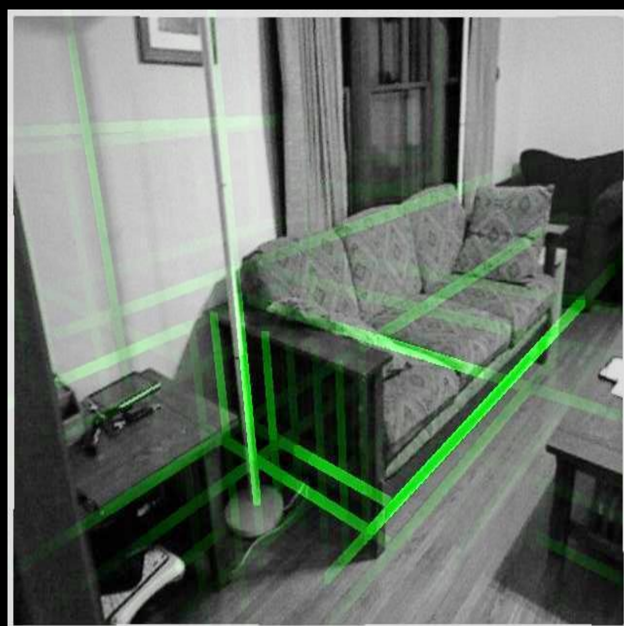
Binary Potentials

Convex (+)



Concave (-)

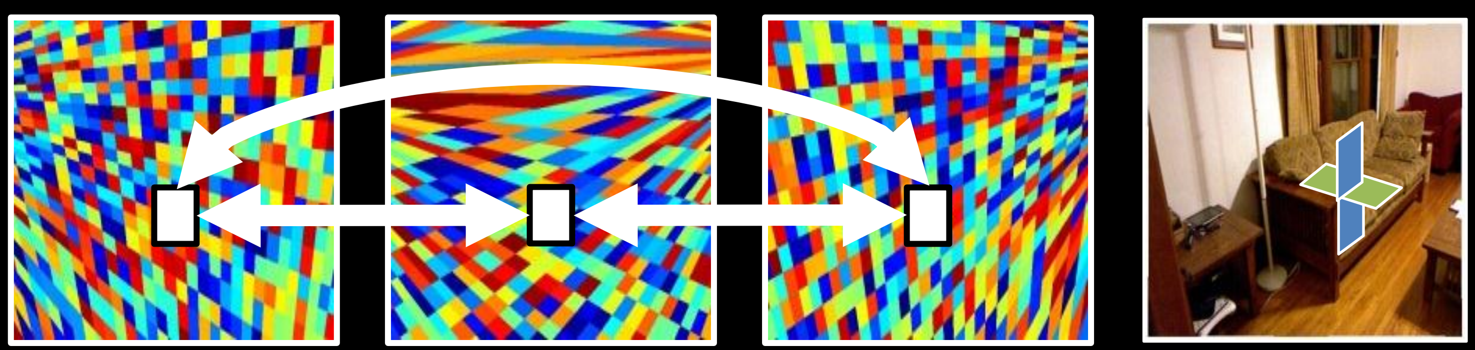




Constraints

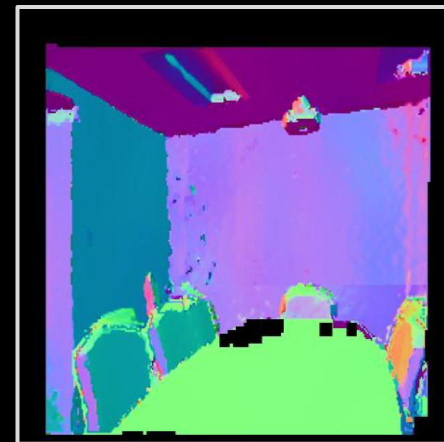
What configurations are forbidden?

$$\arg\max_{\mathbf{x}\in\{0,1\}^n} \mathbf{c}^T\mathbf{x} + \mathbf{x}^T\mathbf{H}\mathbf{x} \quad \text{s.t.} \quad \boxed{\mathbf{A}\mathbf{x} \leq \mathbf{1}}$$



Gurobi BB

Input



Ground Truth

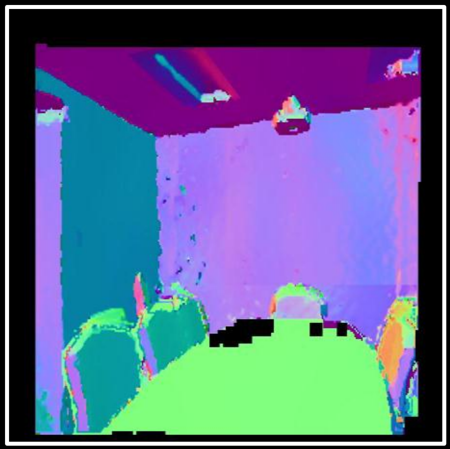
3D Primitives

Projected 3D Primitives

Proposed

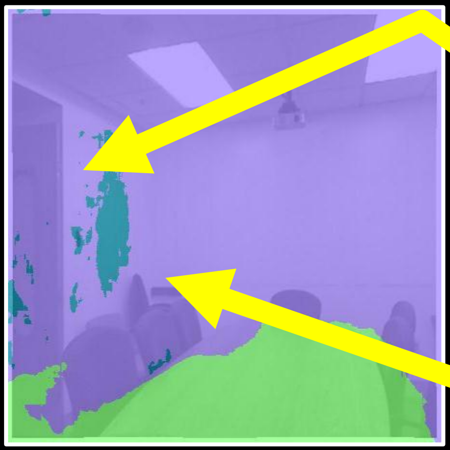
Qualitative Results

Input

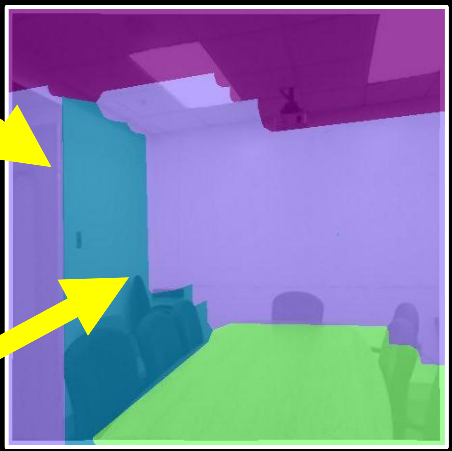


Ground Truth

3D Primitives



Projected 3D Primitives



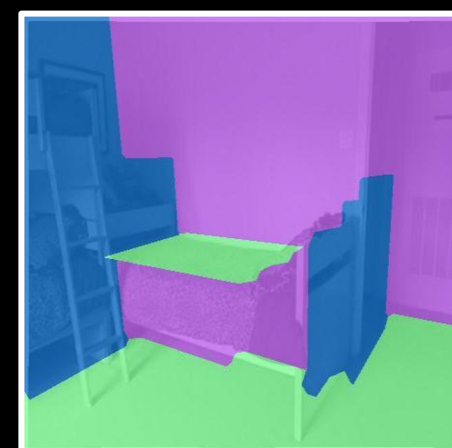
Proposed

Input

Ground Truth

3D Primitives

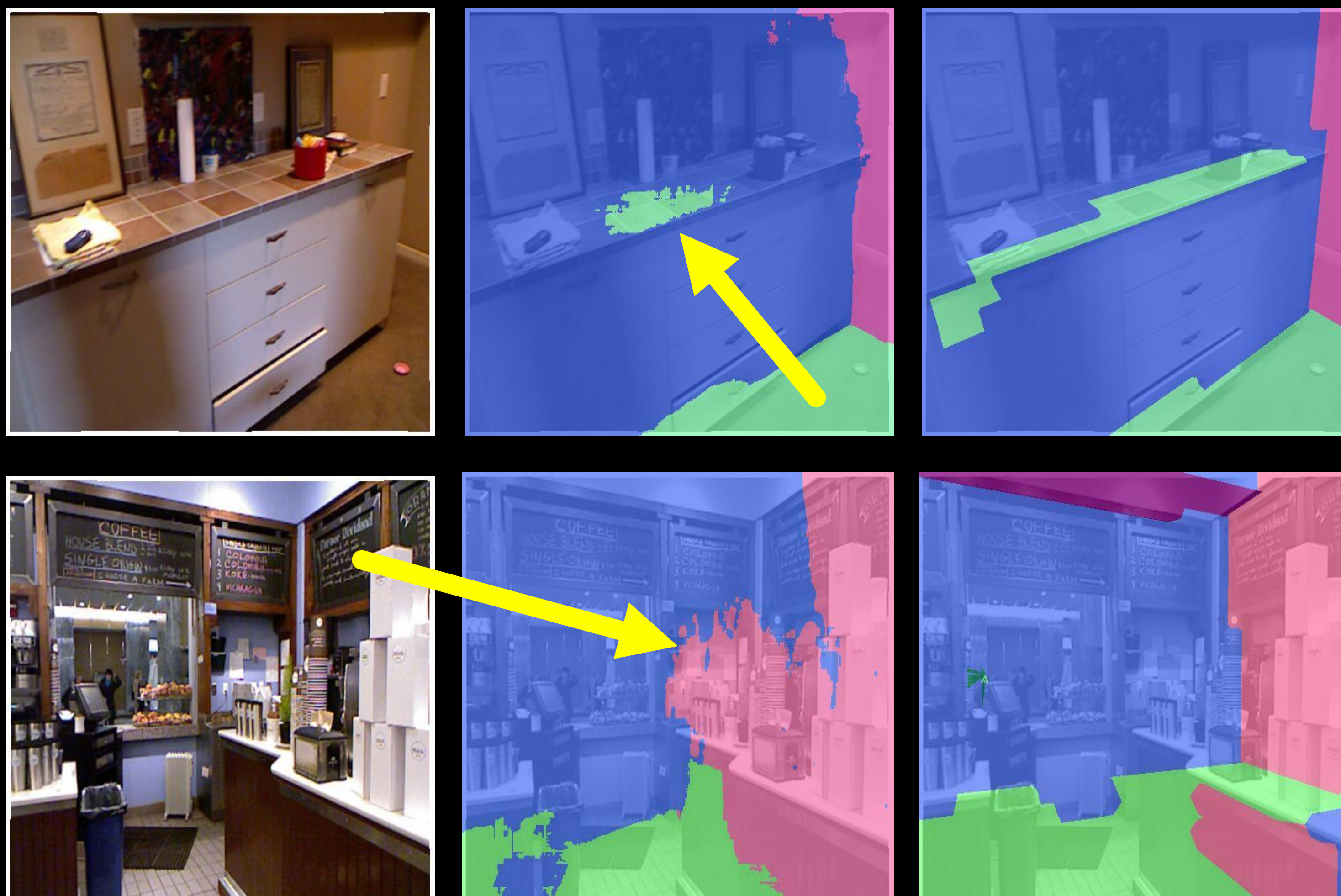
Projected 3D Primitives



Proposed

Random Qualitative Results

3D Primitives | Proposed



Quantitative Results

	Summary Stats (^{0}) (Lower Better)			% Good Pixels (Higher Better)		
	Mean	Median	RMSE	11.25^0	22.5^0	30^0
Proposed	**37.5**	**17.2**	**53.2**	**41.9**	**53.9**	**58.0**
3D Primitives	38.5	19.0	54.2	41.7	52.4	56.3
Hedau et al.	43.2	24.8	59.4	39.1	48.8	52.3
Lee et al.	47.6	43.4	60.6	28.1	39.7	43.9
Karsch et al.	46.6	43.0	53.6	5.4	19.9	31.5
Hoiem et al.	45.6	38.2	55.1	8.6	30.5	41.0

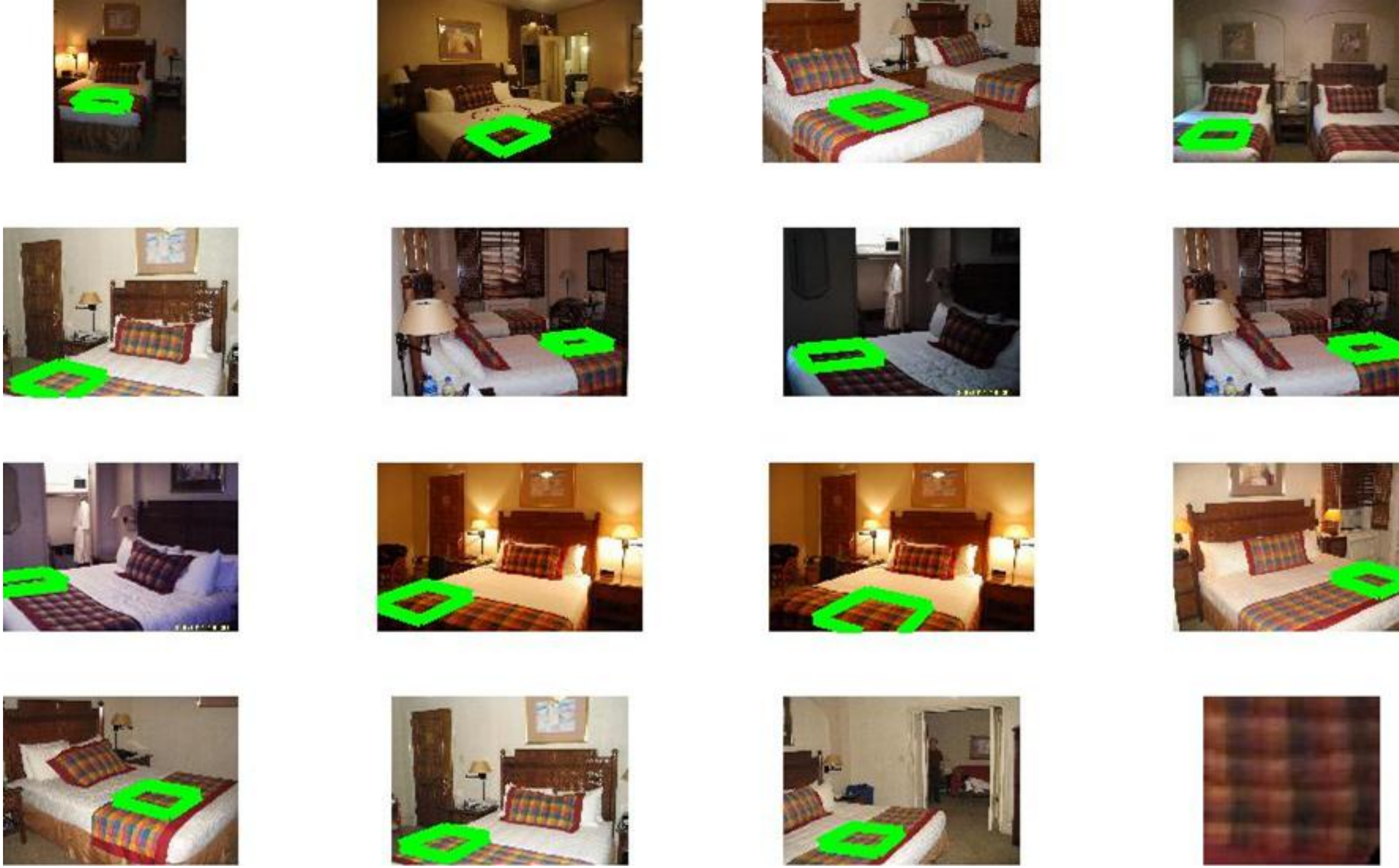
rank

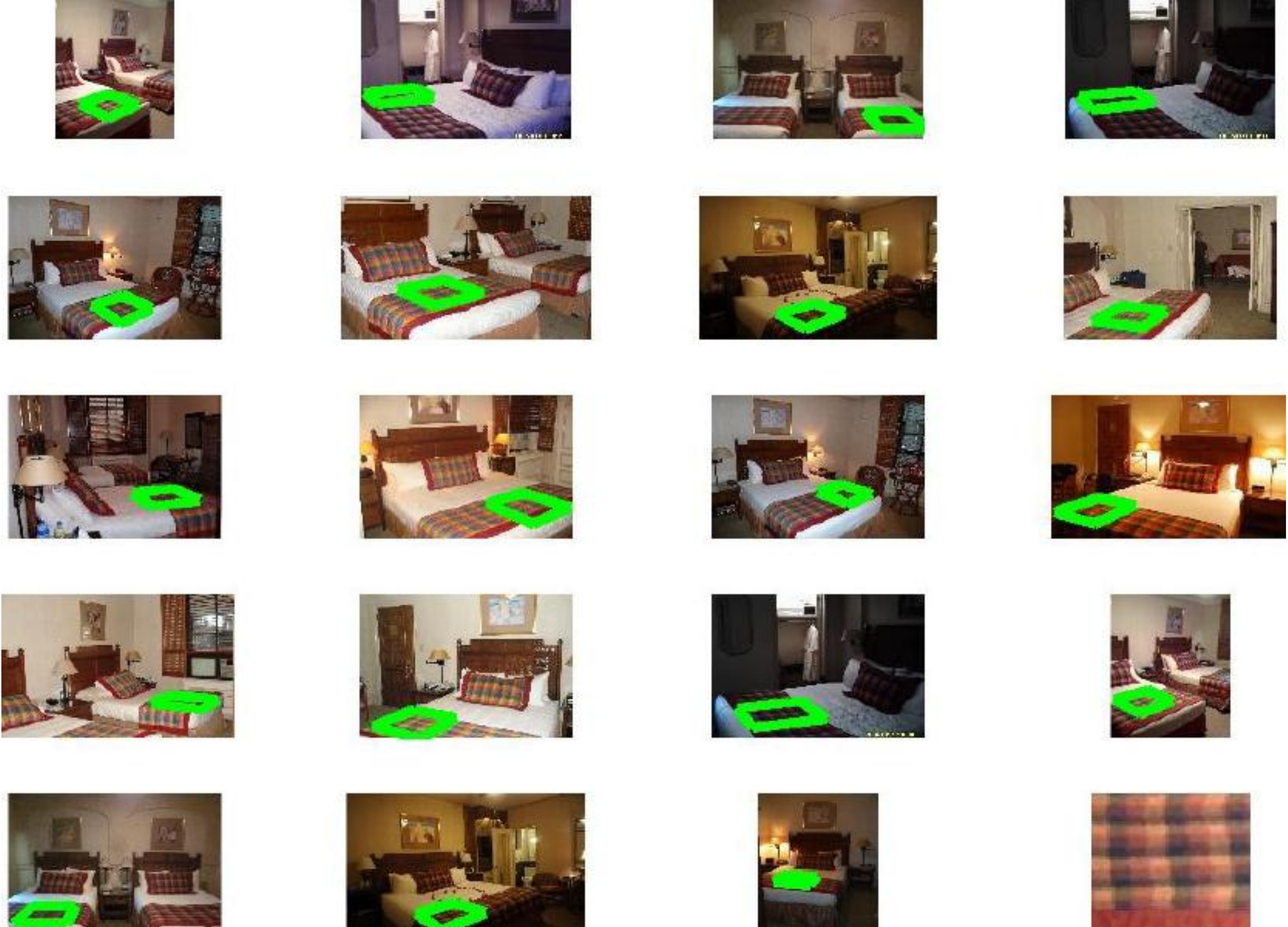



Style vs. structure?

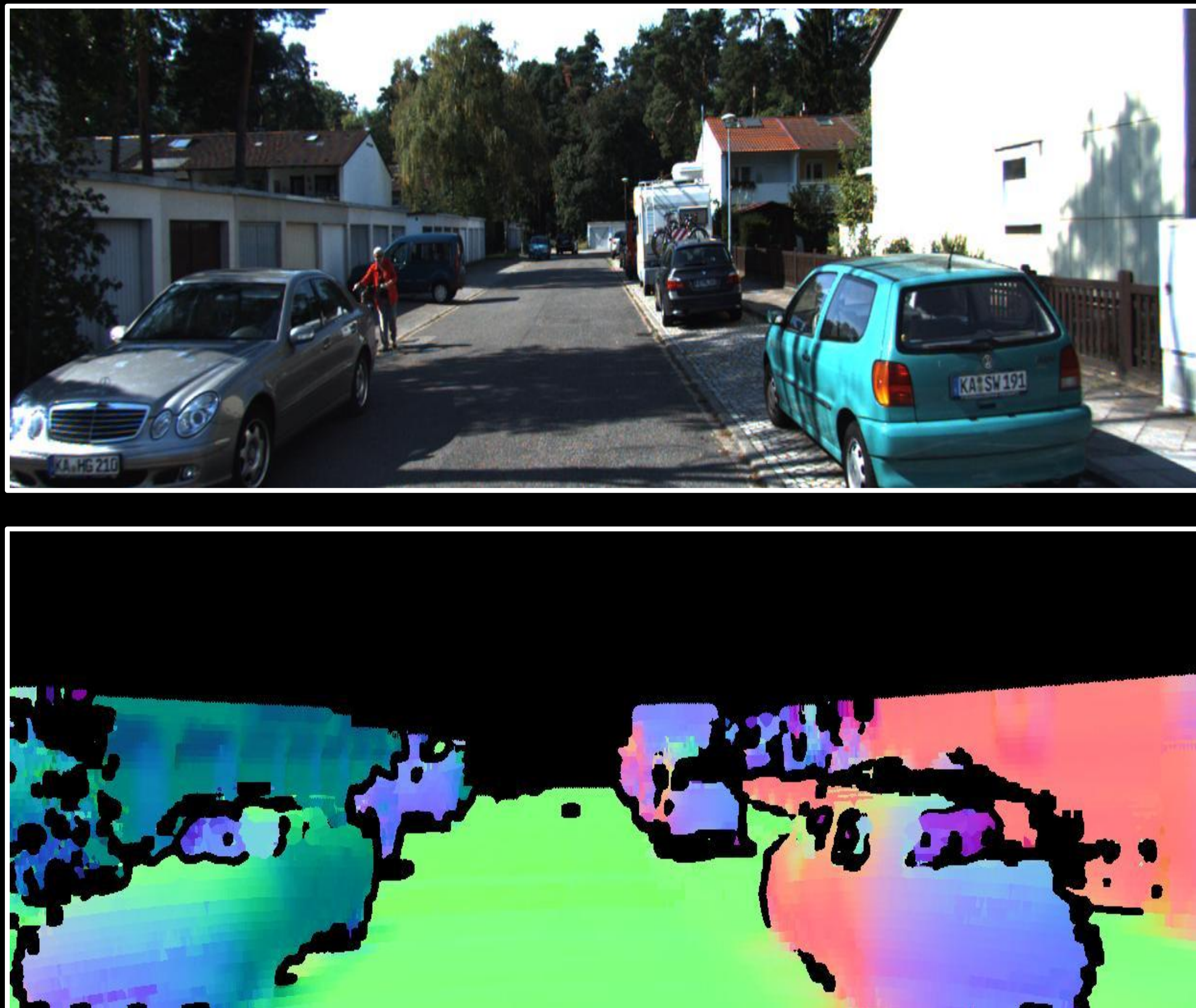

Tenenbaum & Freeman. Separating Style and Content with Bilinear Models. Neural Computation. 2000.

Casablanca Hotel, New York



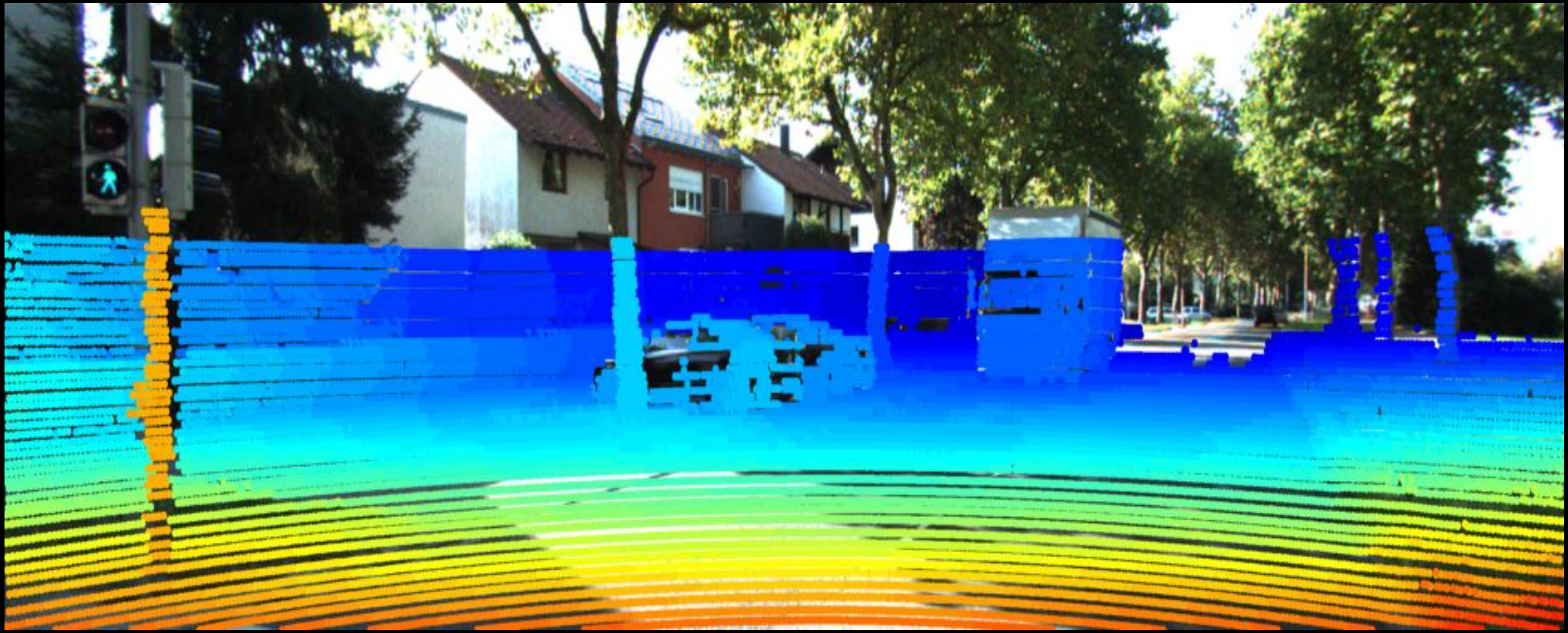


More general environments?

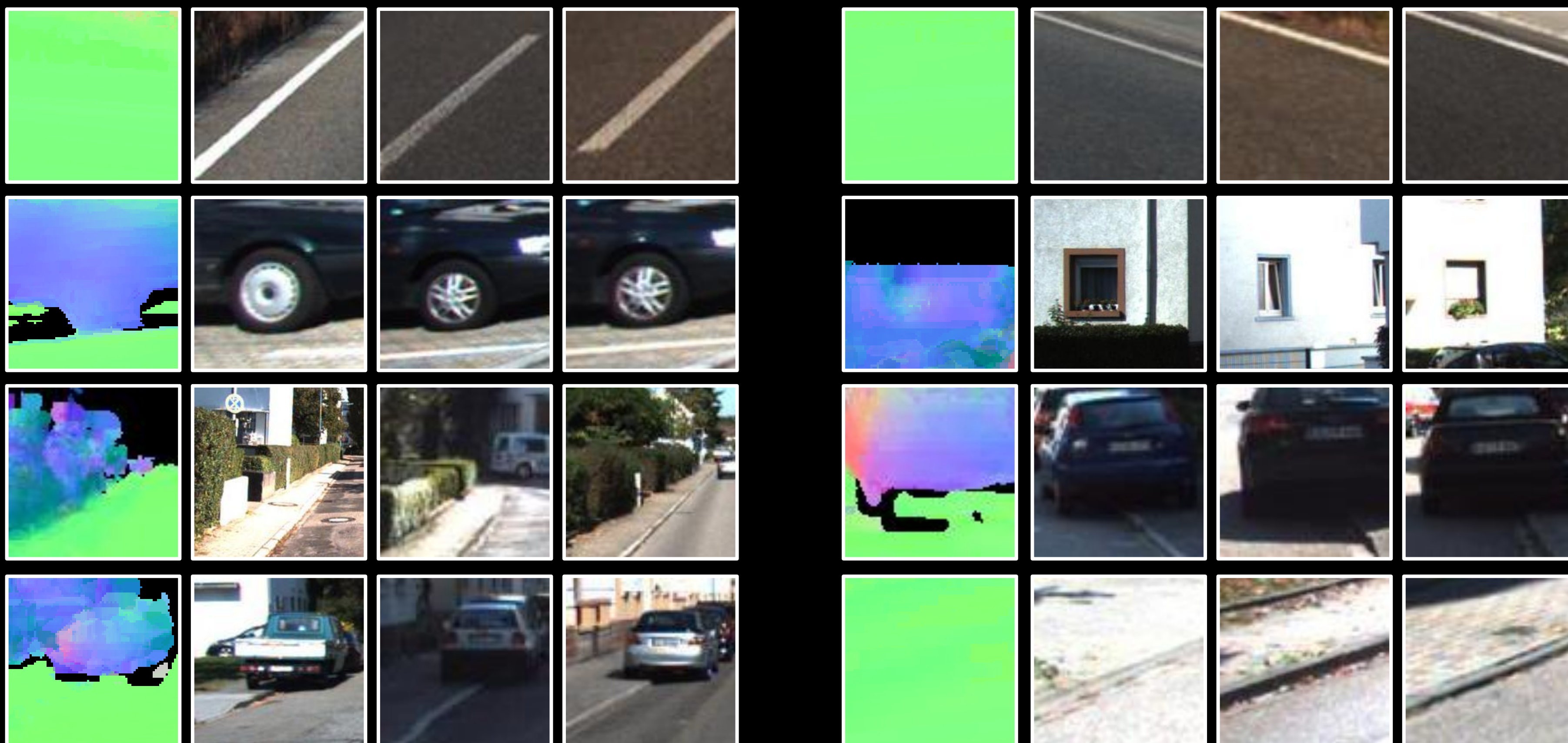


KITTI Dataset: Geiger, Lenz, Urtasun, ‘12

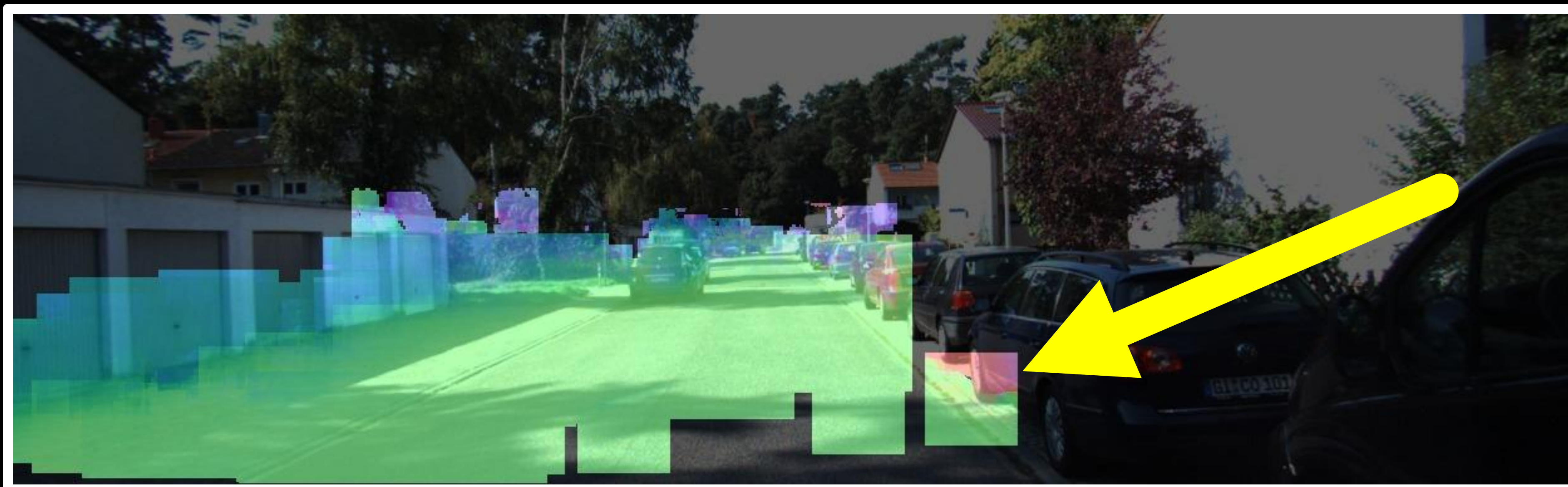
- Large regions without surface interpretation
- Fewer linear/planar structures to anchor
- Irregular distribution of 3D training data



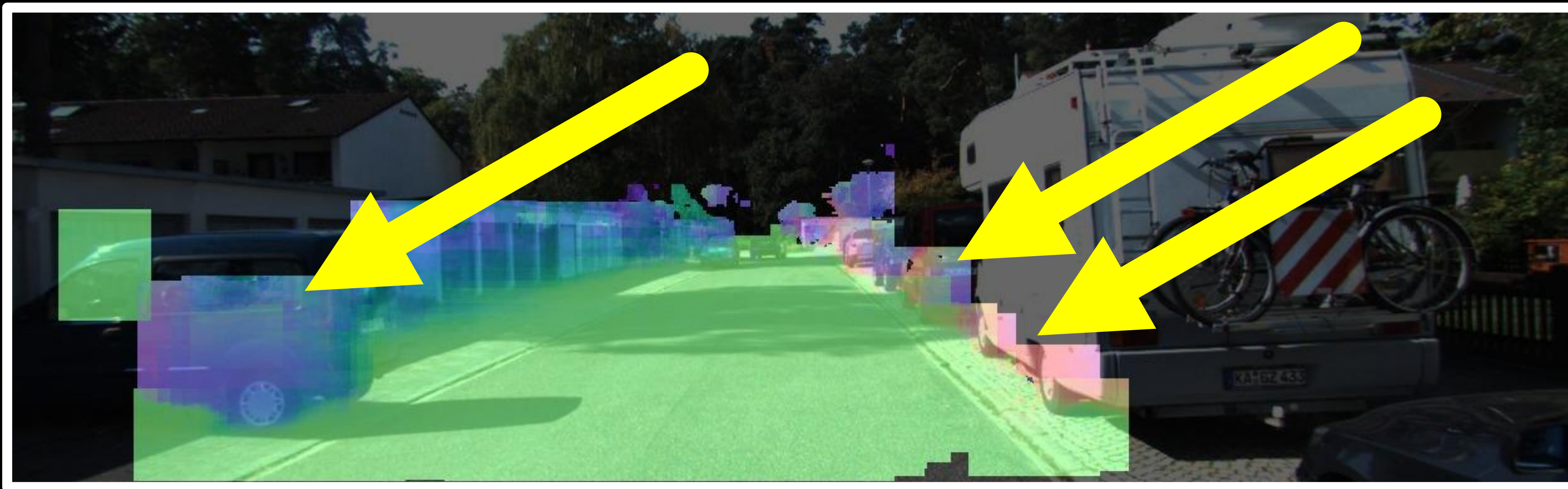

Discovered Primitives (Examples)

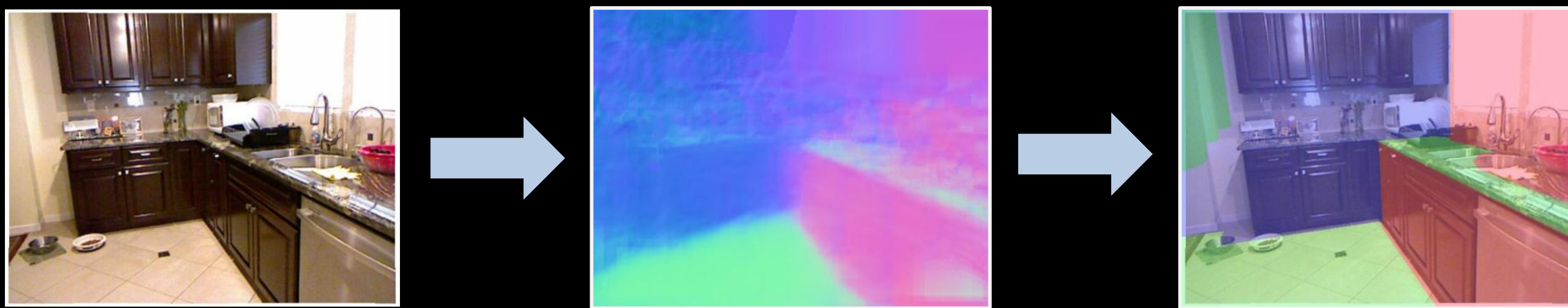


Contact points



Object surfaces + Contact points





Next:

- Better reasoning
- Semantic information
- Less structured environments
- Evaluation
- Applications

Data-Driven 3D Primitives For Single-Image Understanding, Fouhey, Gupta, Hebert, In ICCV 2013.

Unfolding an Indoor Origami World, Fouhey, Gupta, Hebert, In ECCV 2014.

- Harvested from tripadvisor.com

Countries	8	USA, Japan, London, Germany, Canada, Australia, Thailand, Indonesia
Cities	> 10	New York, London, Berlin, Sydney, Tokyo, Las Vegas, San Francisco etc.
Chains	~ 5	Hilton, Marriott, Intercontinental, Sheraton, Best Western etc.

Sheraton Los Angeles

Meritan Apartments Sydney

Le Champlain Quebec

Project digression…..

CENTER FOR MISSING & EXPLOITED CHILDREN

YEARS OF HOPE

WHO WE ARE | DONATE | WHAT WE DO | HOW YOU CAN HELP | RESOURCES

What We Do • Child Sexual Exploitation • CyberTipline

- Missing Children
- Child Sexual Exploitation
  - CyberTipline
  - Child Victim Identification
  - Sex Offender Tracking
  - Child Sex Trafficking
  - Voluntary Industry Initiatives
  - International Collaboration
  - Success Stories
  - FAQ
- Child Safety & Prevention
- Law Enforcement Training
- Victim & Family Support

Safety starts with NetSmartz

CyberTipline

The CyberTipline® receives leads and tips regarding suspected crimes of sexual exploitation committed against children. More than 2.3 million reports of suspected child sexual exploitation have been made to the CyberTipline between 1998 and March 2014.

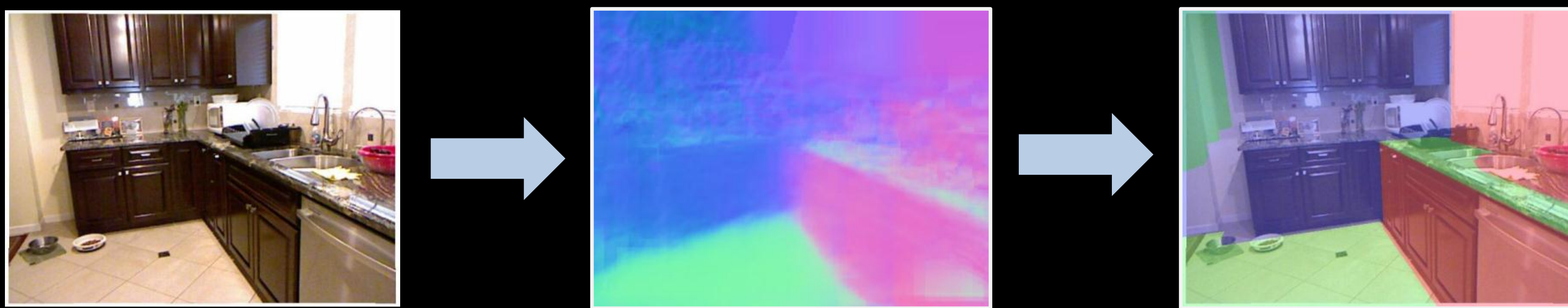
If you have information regarding possible child sexual exploitation, report it to the CyberTipline.

MAKE A CYBERTIPLINE REPORT

Purpose and function

CYBER TIPLINE www.cybertipline.com 1-800-843-5678

The CyberTipline is operated in partnership with the FBI, Immigration and Customs Enforcement, U.S. Postal Inspection Service, U.S. Secret Service, military criminal investigative organizations, U.S. Department of Justice, Internet Crimes Against Children Task Force program, as well as other state and local law enforcement agencies. Reports to the CyberTipline are made by the



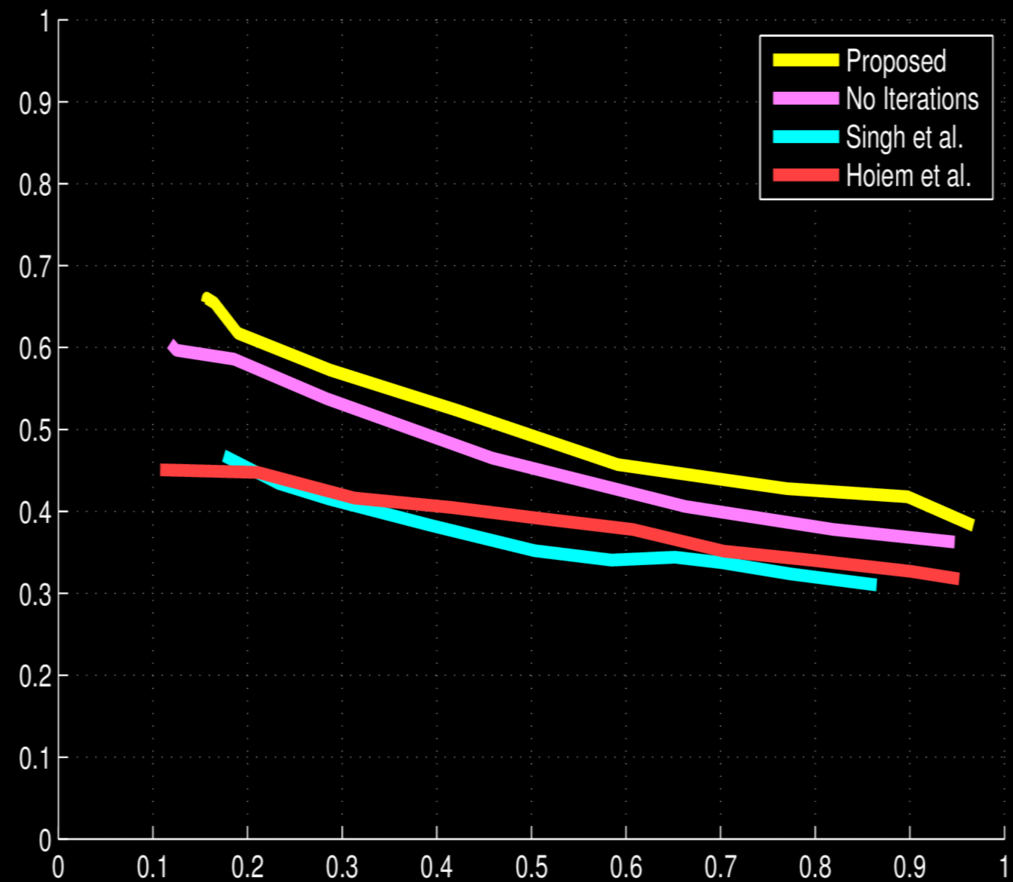
Next:

- Better reasoning
- Semantic information
- Less structured environments
- Evaluation
- Applications

Data-Driven 3D Primitives For Single-Image Understanding, Fouhey, Gupta, Hebert, In ICCV 2013.

Unfolding an Indoor Origami World, Fouhey, Gupta, Hebert, In ECCV 2014.

Results – Quantitative



Recall