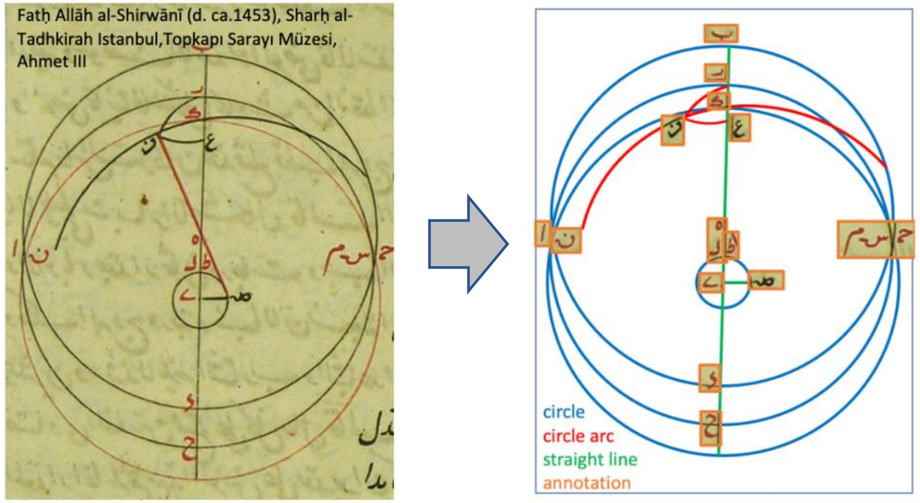
Deep Learning for 3D data

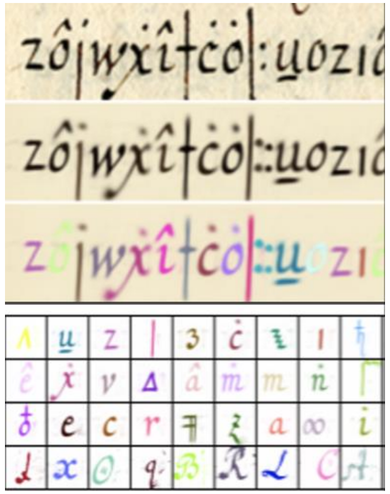
Mathieu Aubry

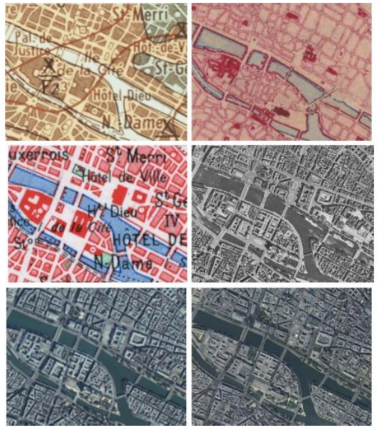
Imagine – LIGM, Ecole des Ponts ParisTech (ENPC)

A few words about my research

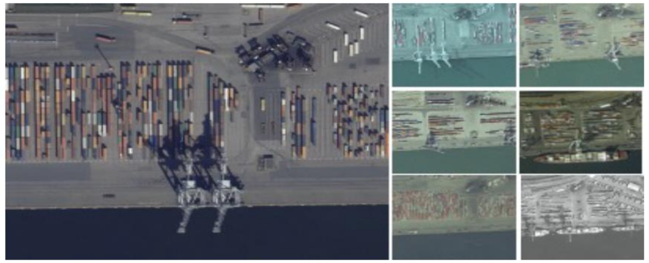
Current: Unsupervised image analysis, applications to historical data or Earth imagery







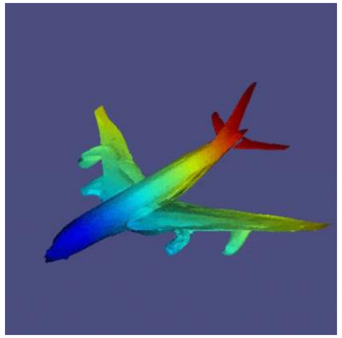
(b) Maps and aerial images [Nat]



(c) Harbor area (Le Havre) [Nat]

(d) Shale oil operations

Past: Deep 3D model generation/analysis.



Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

2015-2019

Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Mostly my students
2020-2024

Learning with synthetic data

Outline: Deep learning and 3D data

Important milestones:

1. **Classification and Segmentation**
2. Matching / Alignment
3. Generation and single view reconstruction

Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

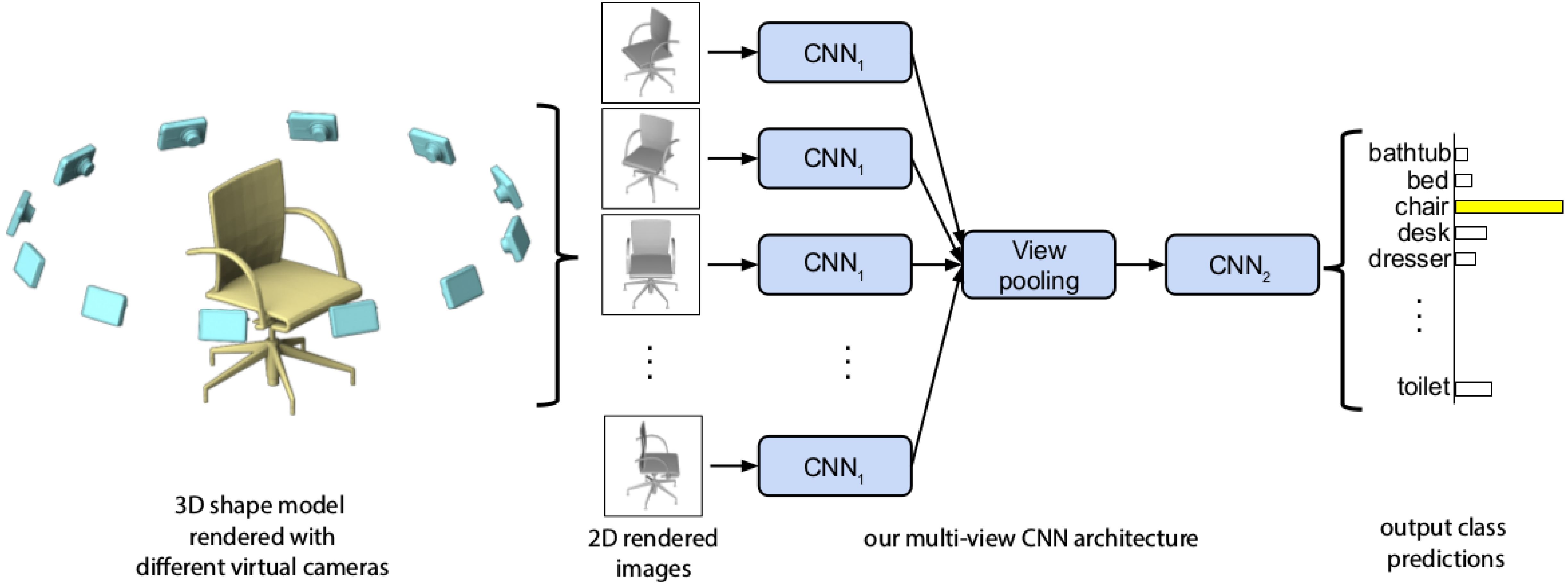
Key issue: 3D representation

- 2D views / Depth maps
- Voxels
- Points
- Meshes
- Parametric surface
- Implicit surface
- ”Procedural”

Key issue: 3D representation

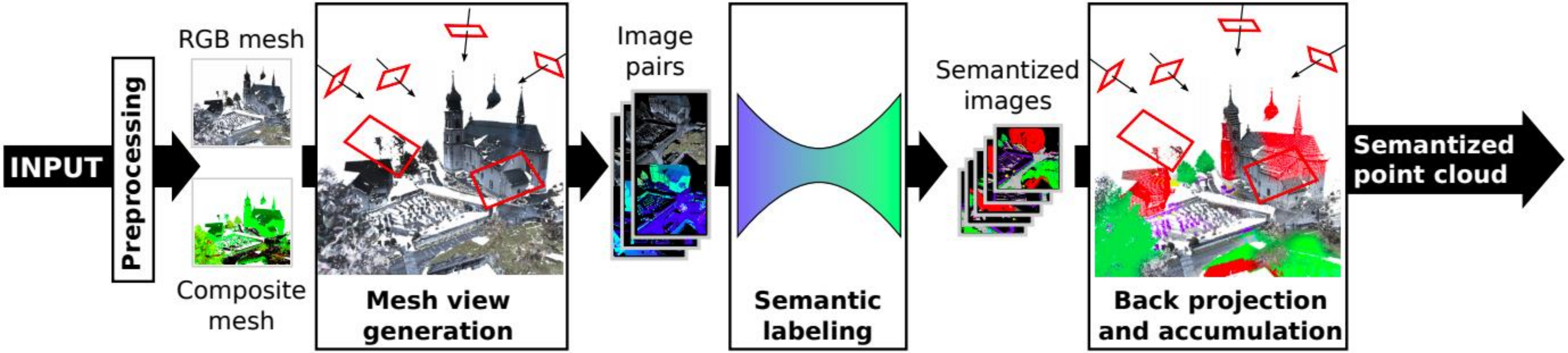
- **2D views / Depth maps**
- Voxels
- Points
- Meshes
- Parametric surface
- Implicit surface
- ”Procedural”

3D category recognition from rendered views



Su, H., Maji, S., Kalogerakis, E., & Learned-Miller, E. ICCV 2015
Multi-view Convolutional Neural Networks for 3D Shape Recognition.

Semantic segmentation from rendered views

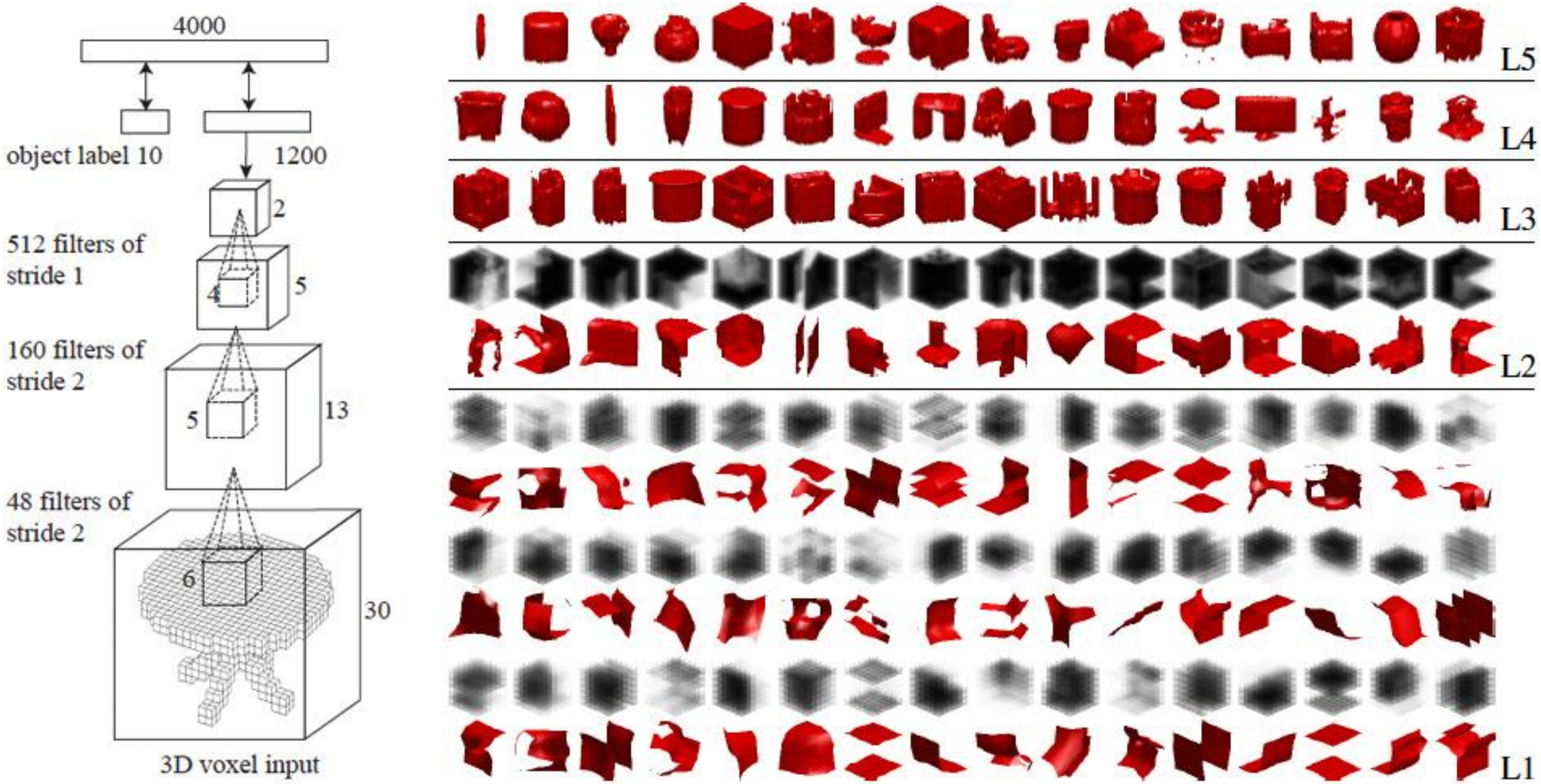


A. Boulch, B. L. Saux, and N. Audebert. Unstructured point cloud semantic labeling using deep segmentation networks. In Eurographics Workshop on 3D Object Retrieval 2017

Key issue: 3D representation

- 2D views / Depth maps
- **Voxels**
- Points
- Meshes
- Parametric surface
- Implicit surface
- ”Procedural”

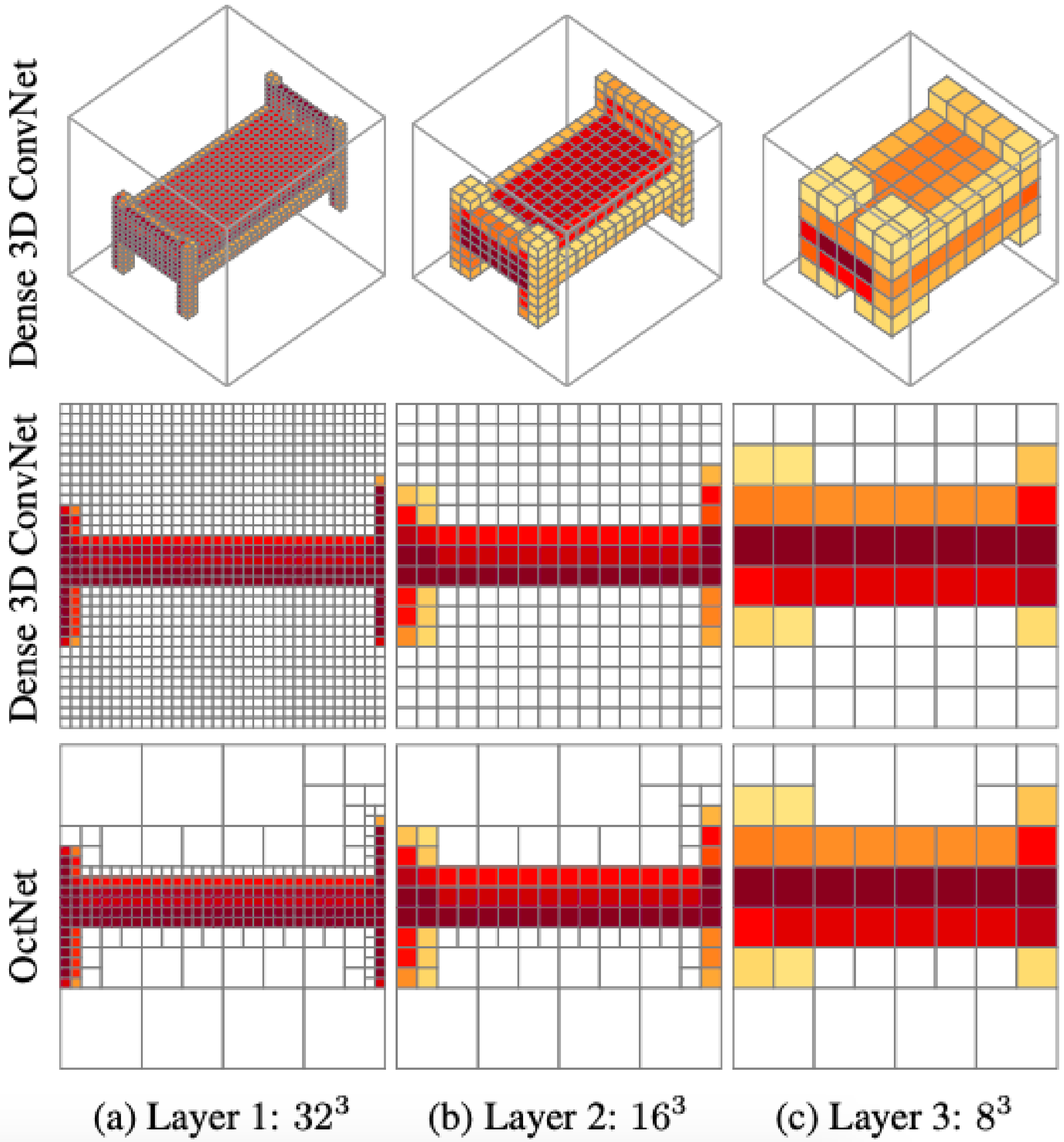
3D recognition from voxels



Wu, Z., Song, S., Khosla, A., Tang, X., & Xiao, J. CVPR 2015
3d shapenets: A deep representation for volumetric shapes.

OctNet

- Voxel representation tend to be costly:
  -> tree based representation

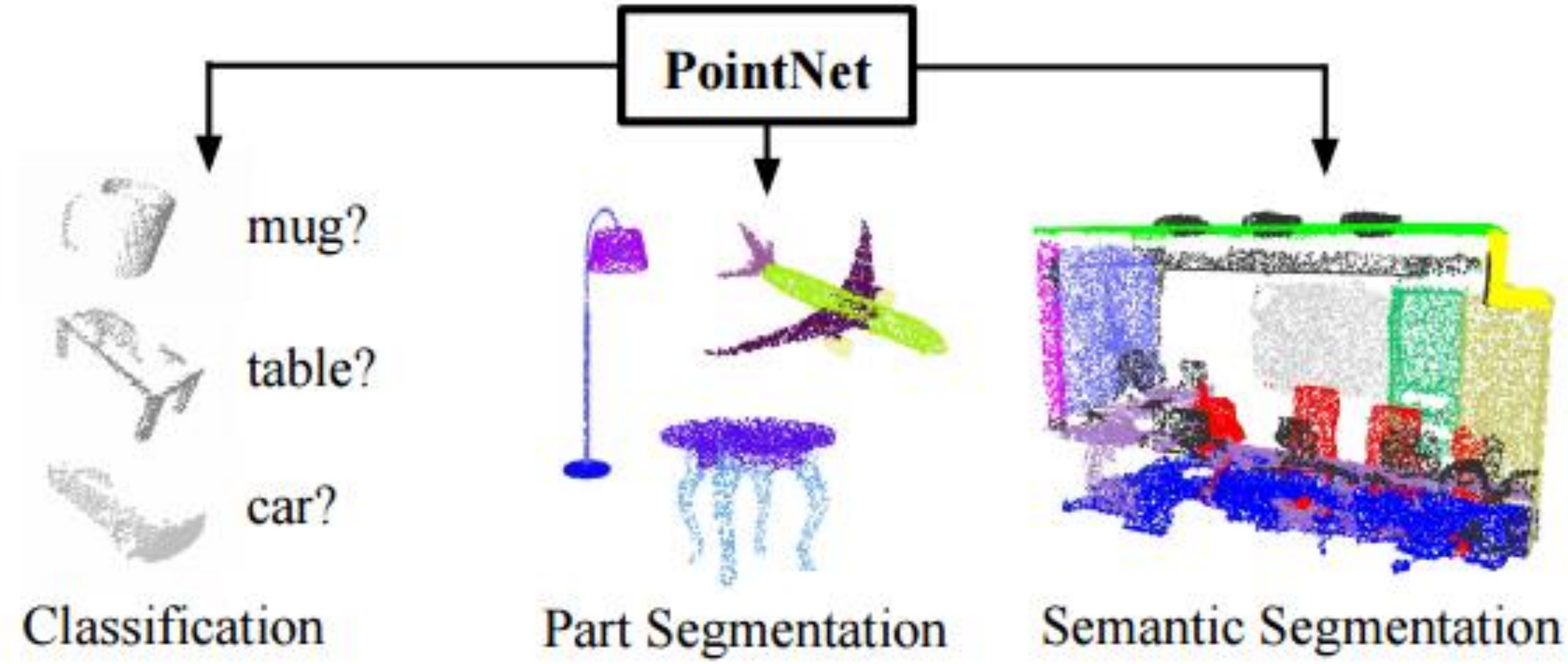


Riegler, G., Osman Ulusoy, A., & Geiger, A.
Octnet: Learning deep 3d representations at high resolutions.
CVPR 2017

Key issue: 3D representation

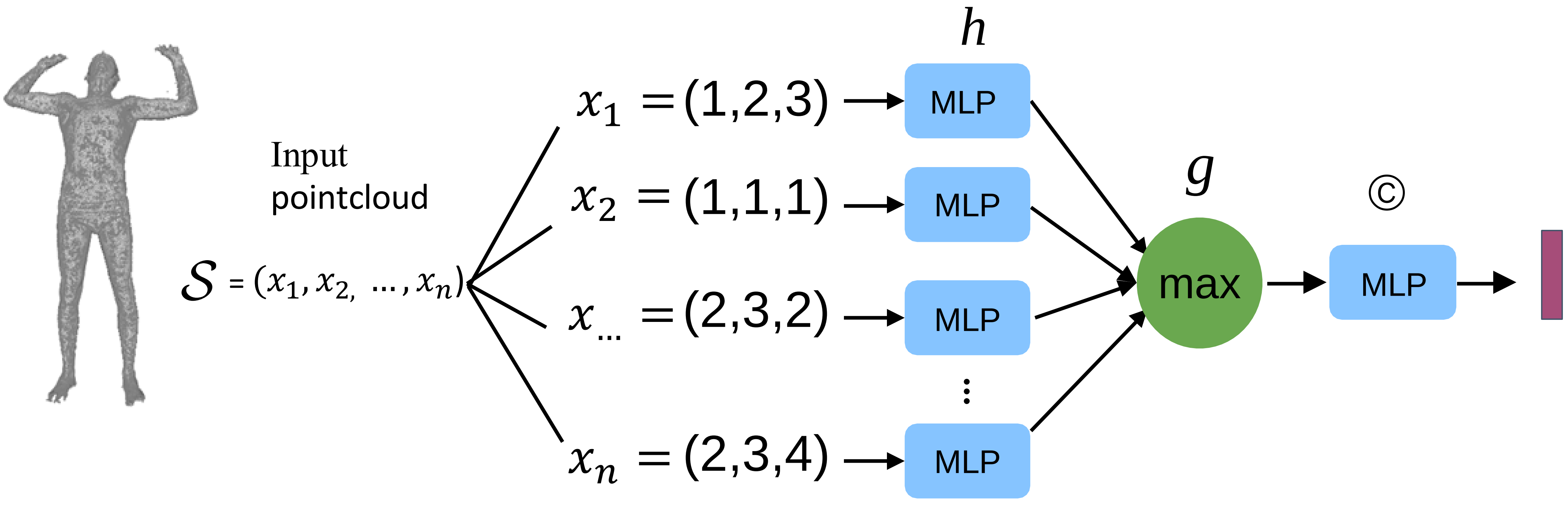
- 2D views / Depth maps
- Voxels
- **Points**
- Meshes
- Parametric surface
- Implicit surface
- ”Procedural”

3D recognition from point clouds



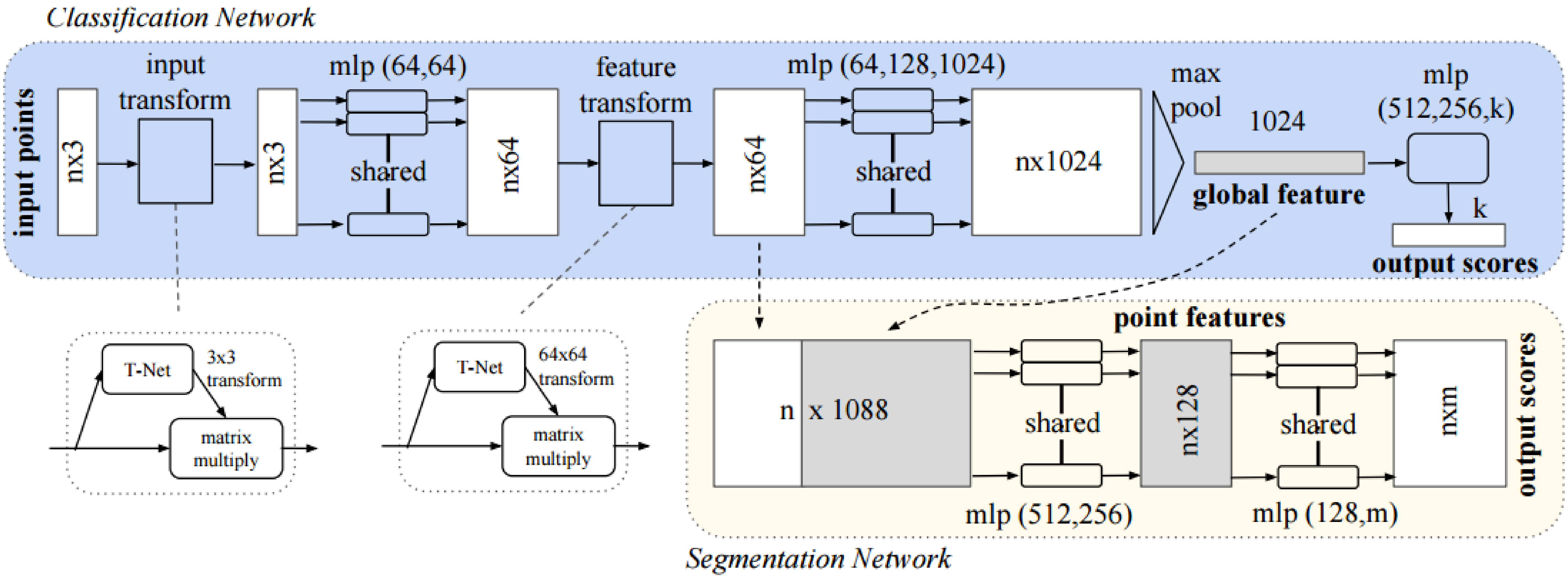
PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,
CR Qi, H Su, K Mo, LJ Guibas, CVPR 2017

PointNet



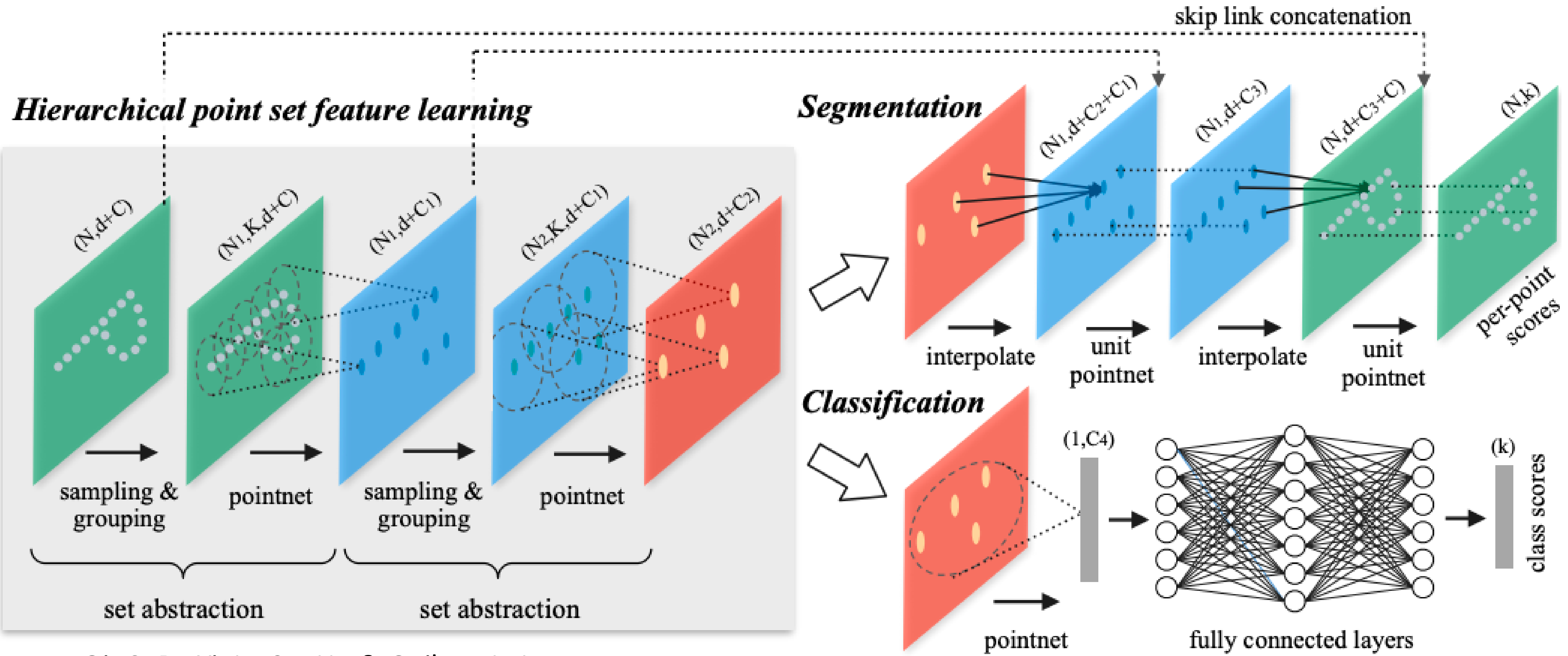
PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et *al*. CVPR (2017)

PointNet



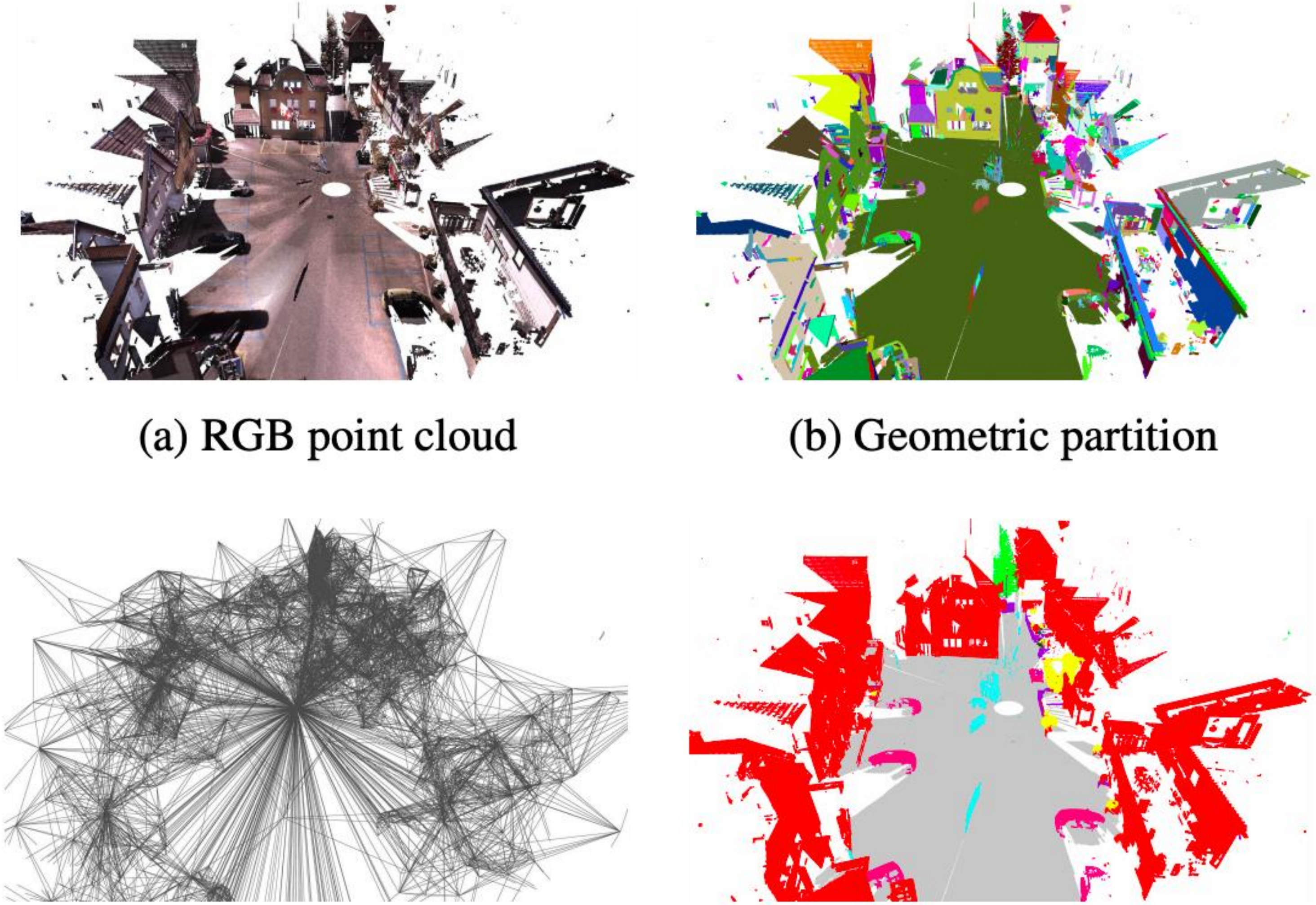
PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,
CR Qi, H Su, K Mo, LJ Guibas, CVPR 2017

PointNet++



Qi, C. R., Yi, L., Su, H., & Guibas, L. J.
Pointnet++: Deep hierarchical feature learning on point sets in a metric space. NeurIPS 2017

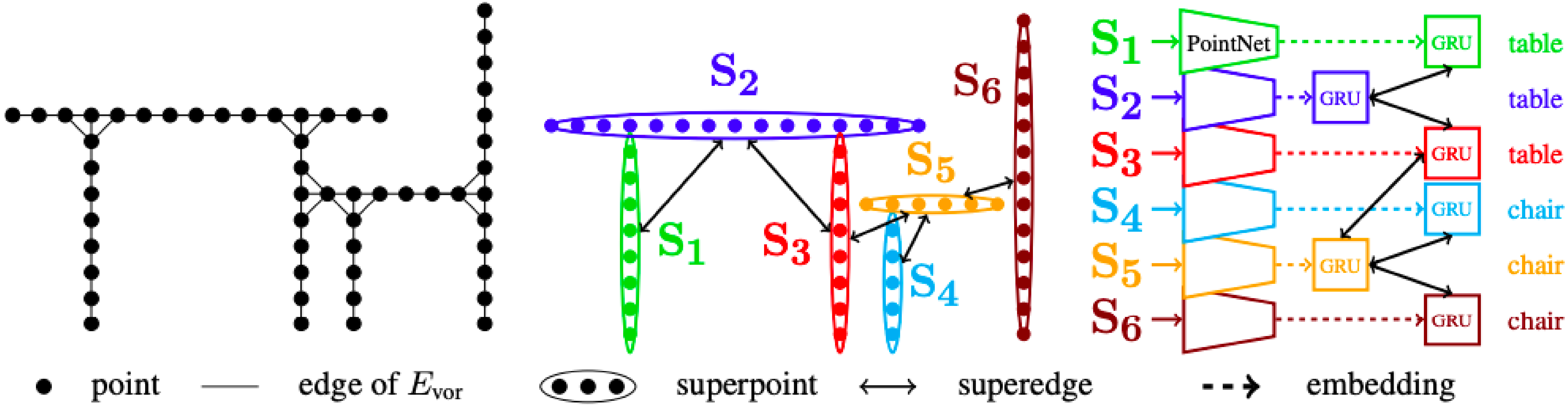
Superpoint Graphs



(a) RGB point cloud (b) Geometric partition

(c) Superpoint graph (d) Semantic segmentation

Landrieu, L., & Simonovsky, M.
Large-scale point cloud semantic segmentation with superpoint graphs
CVPR 2018



(a) Input point cloud (b) Superpoint graph (c) Network architecture

The GRU take as input the previous hidden state and a message computed as a weighted average of its neighbors hidden states.
The weights are computed from a small number of attributes using an MLP

M. Simonovsky and N. Komodakis. Dynamic edgeconditioned filters in convolutional neural networks on graphs. In CVPR, 2017

Key issue: 3D representation

- 2D views / Depth maps
- Voxels
- Points
- **Meshes**
- **Parametric surface**
- Implicit surface
- ”Procedural”

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. **Matching / Alignment**
3. Generation and single view reconstruction

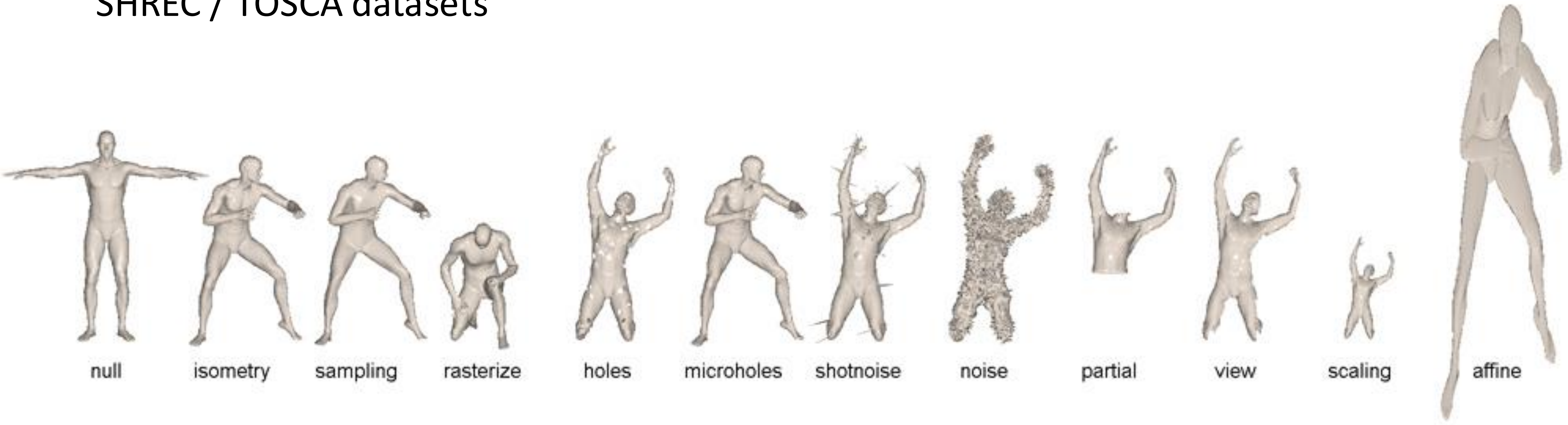
Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

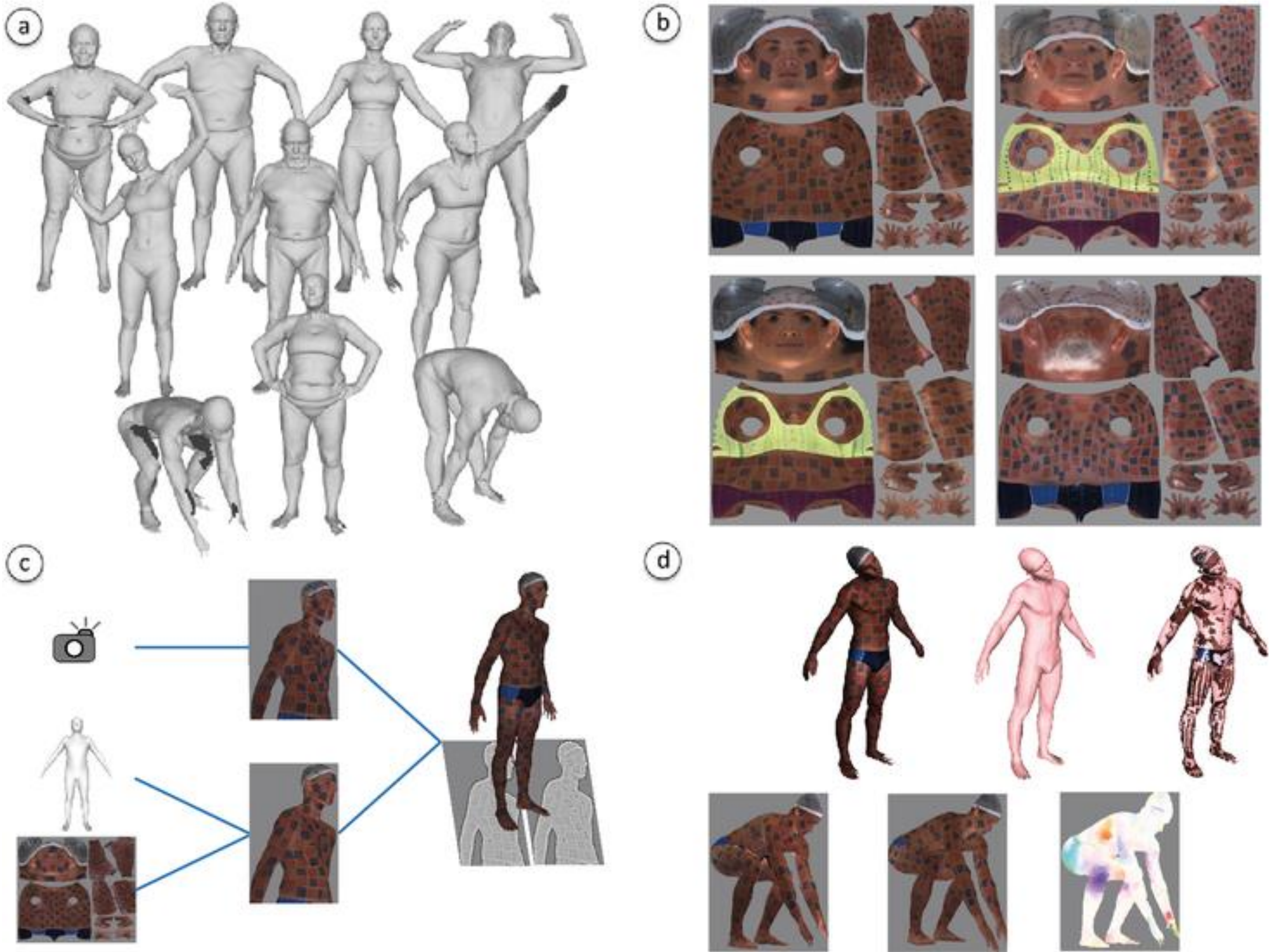
Non-rigid registration

- Evaluation?
  - Synthetic data:
    SHREC / TOSCA datasets

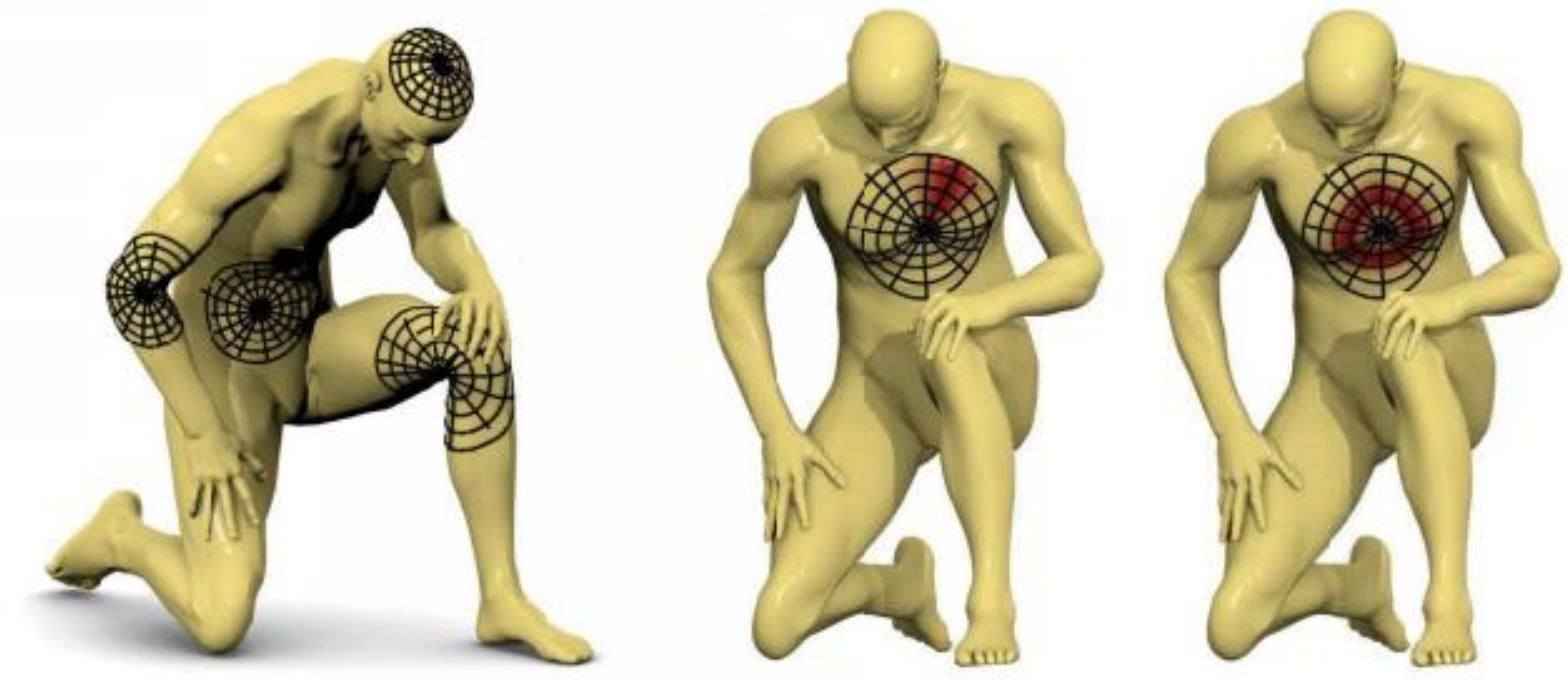


Non-rigid registration

- Evaluation?
  - Synthetic data:
    SHREC / TOSCA datasets
  - Real data:
    FAUST dataset

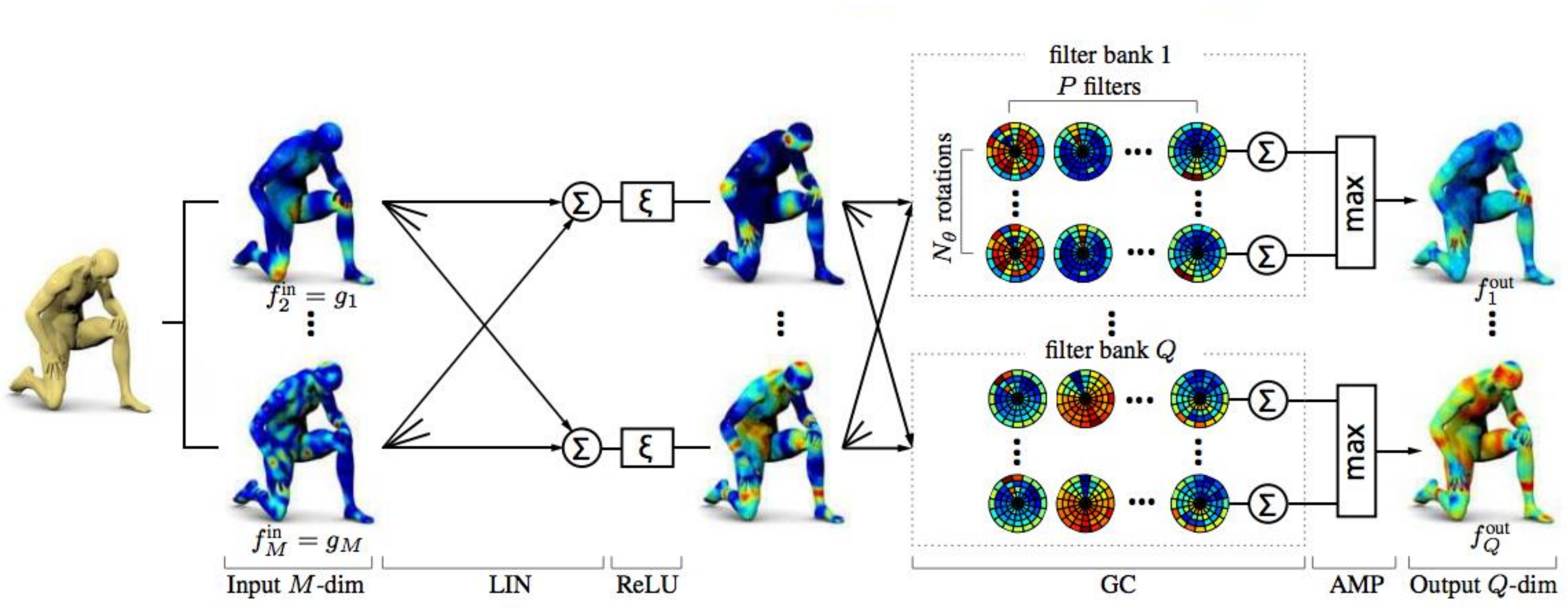


3D local descriptors with spectral CNNs



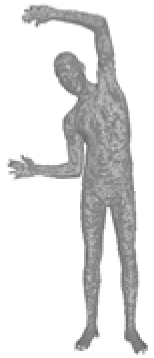
Geodesic convolutional neural networks on riemannian manifolds,
J. Masci, D. Boscaini, M. Bronstein, P. Vandergheynst, ICCV workshops 2015

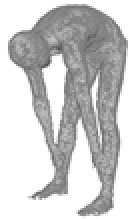
3D local descriptors with spectral CNNs



Geodesic convolutional neural networks on riemannian manifolds,
J. Masci, D. Boscaini, M. Bronstein, P. Vandergheynst, ICCV workshops 2015

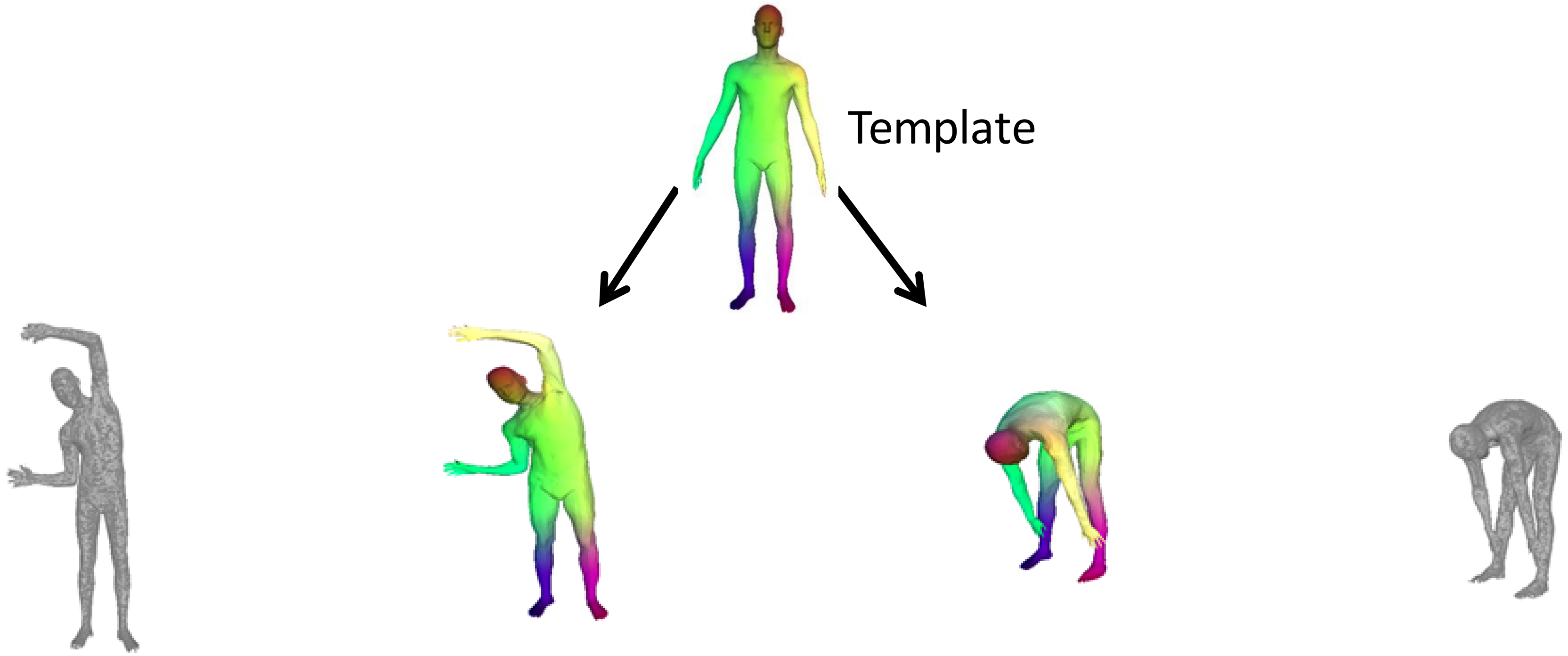
Correspondences through Deformation



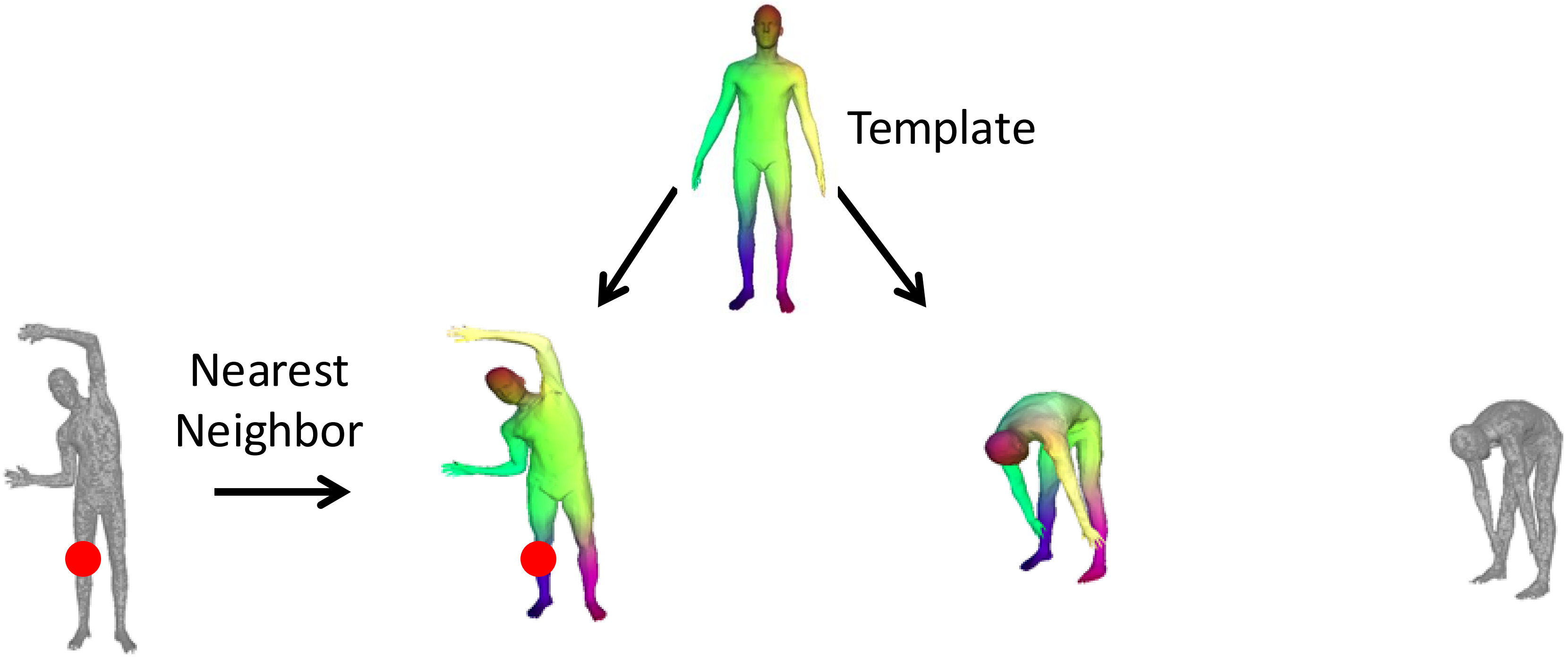


Groueix, T., Fisher, M., Kim, V. G., Russell, B. C., & Aubry, M.
3d-coded: 3d correspondences by deep deformation ECCV 2018

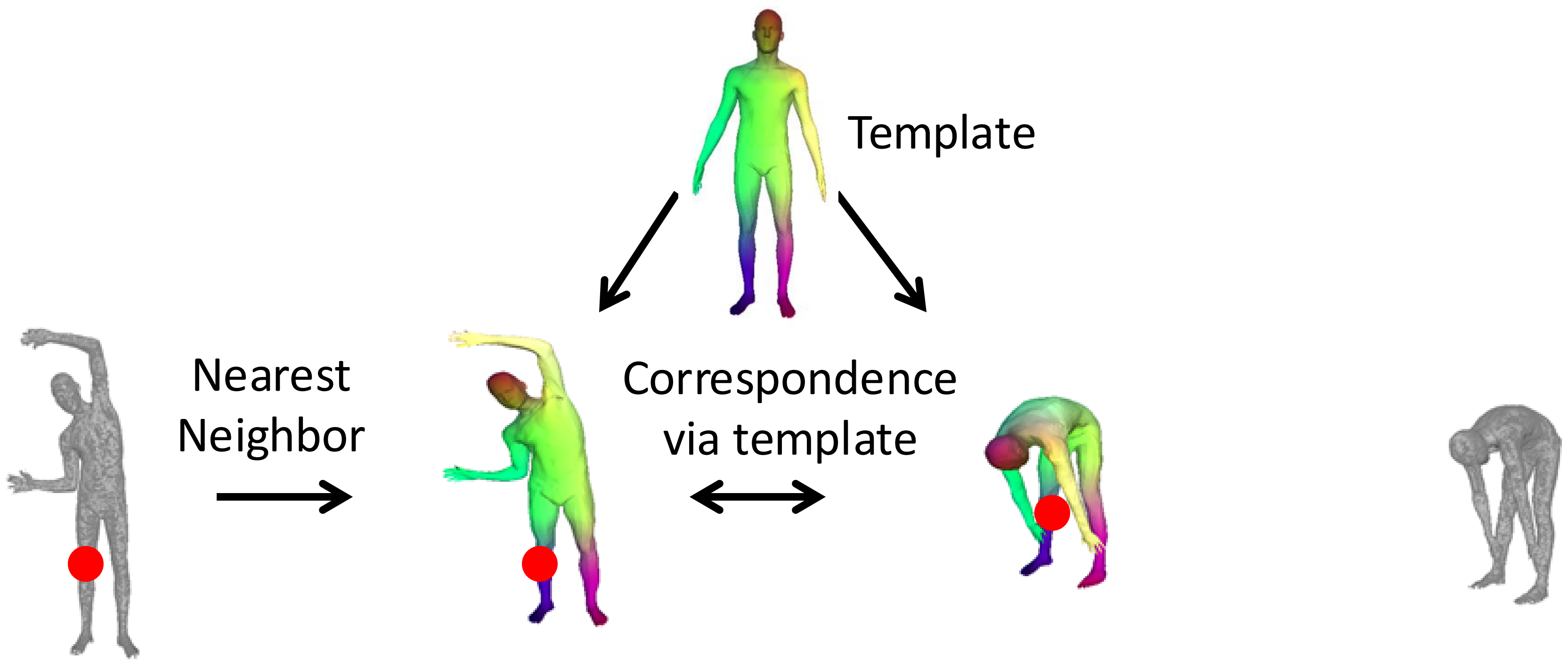
Correspondences through Deformation



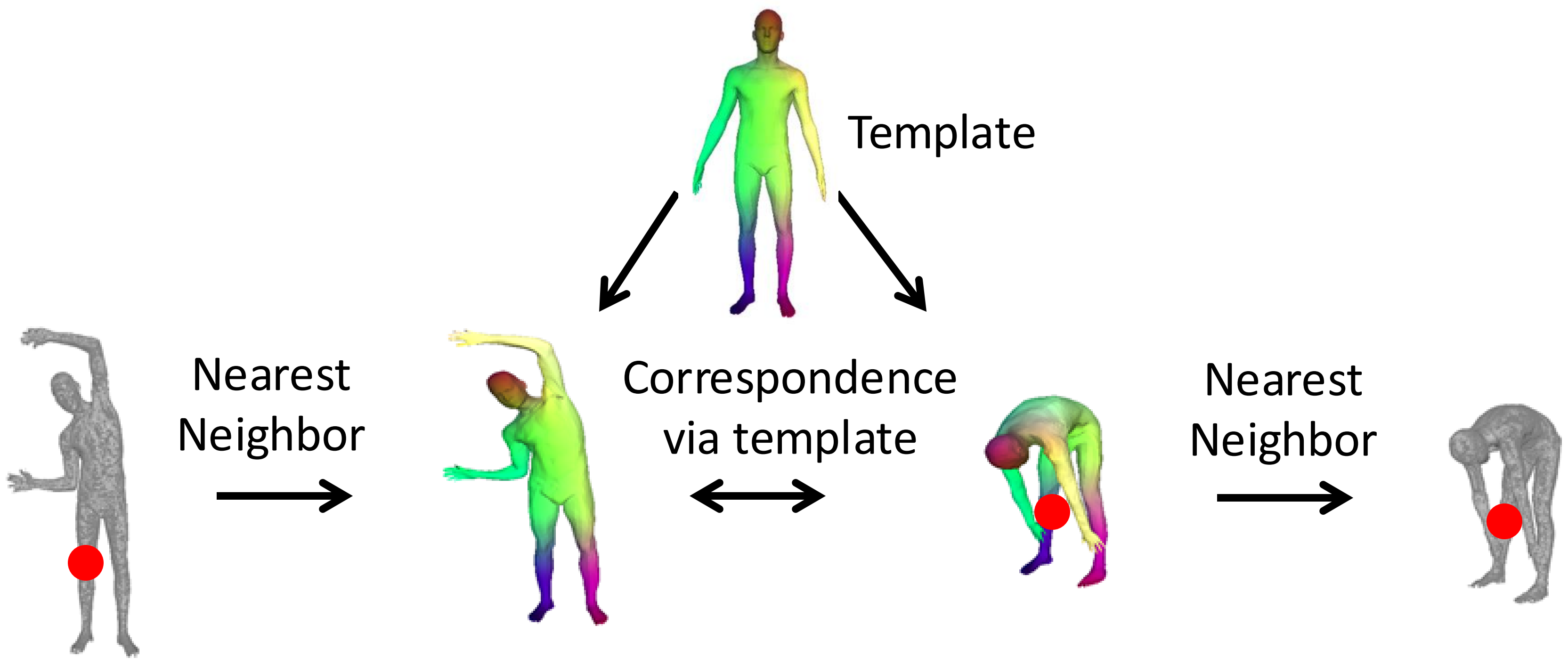
Correspondences through Deformation



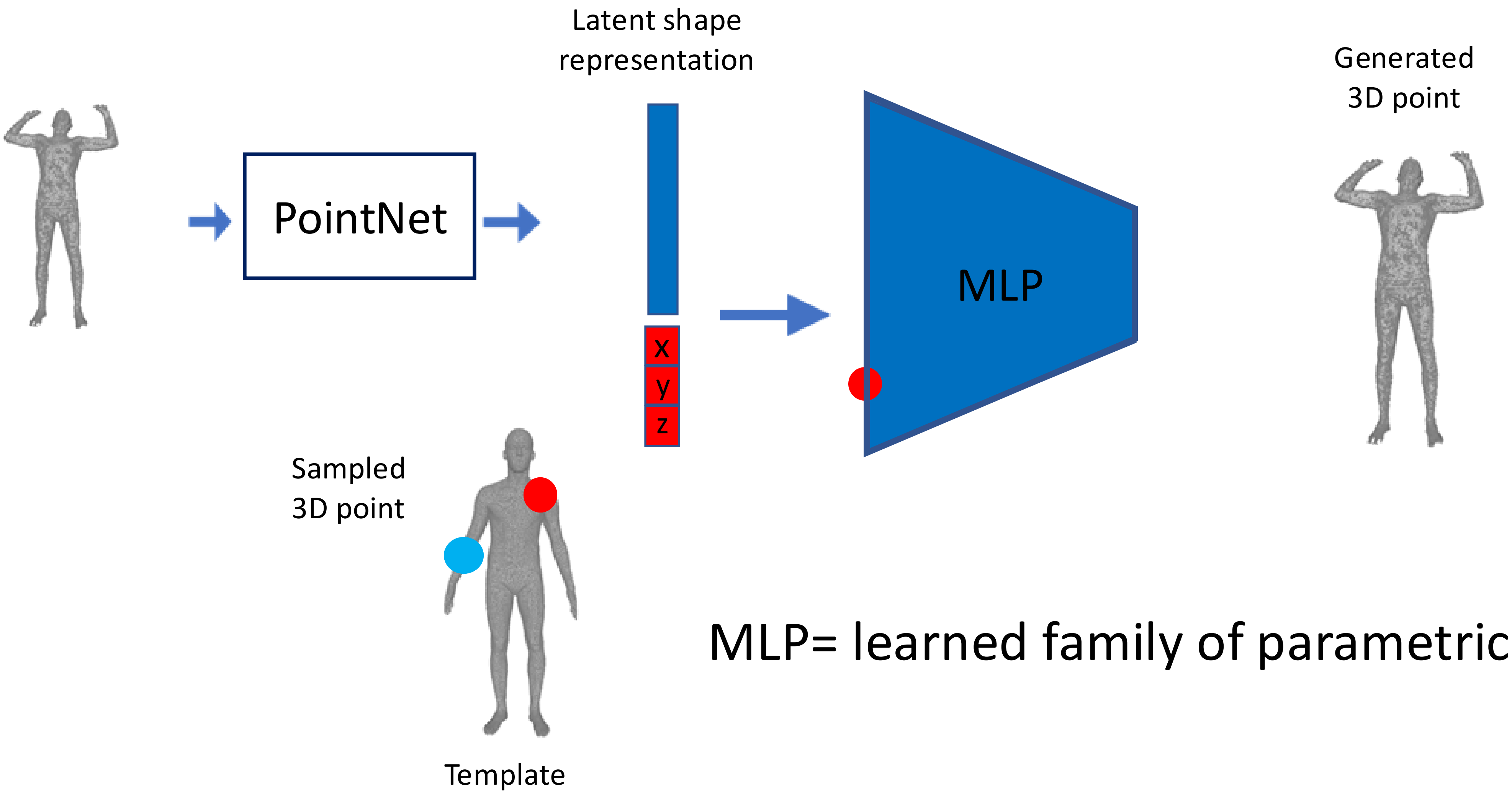
Correspondences through Deformation



Correspondences through Deformation



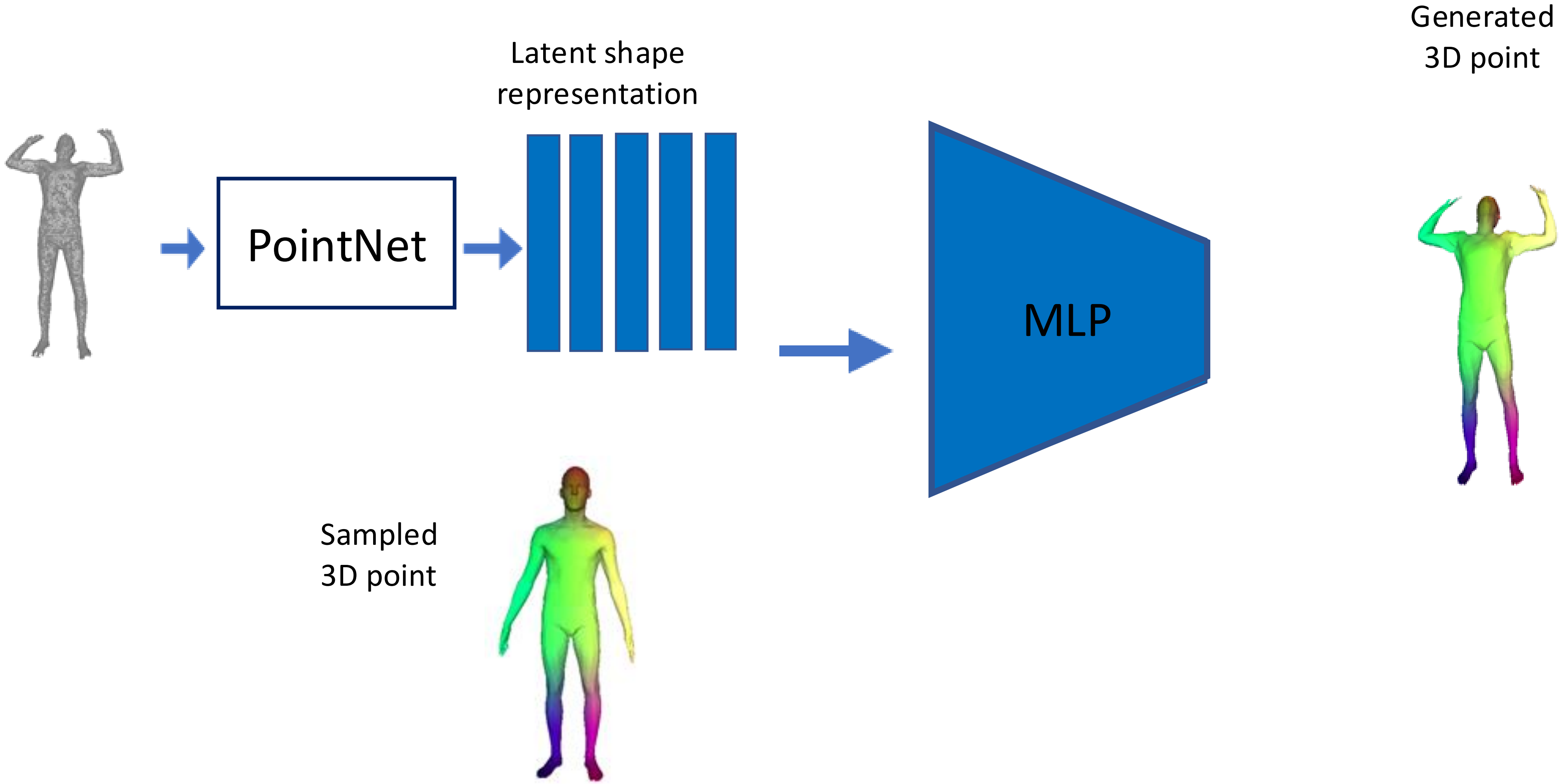
Key idea: deformation



MLP= learned family of parametric deformations

Key idea: deformation

The reconstructed shape is in dense correspondence with the template by design.



Losses

- Let's consider a source point cloud $\mathcal{X} = \{x_1, \ ..., \ x_n\}$ and a target point cloud $\mathcal{Y} = \{y_1, \ ..., \ y_n\}$
- Supervised case:

$$L(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^{n} \|x_i - y_i\|^2$$

- Unsupervised case:

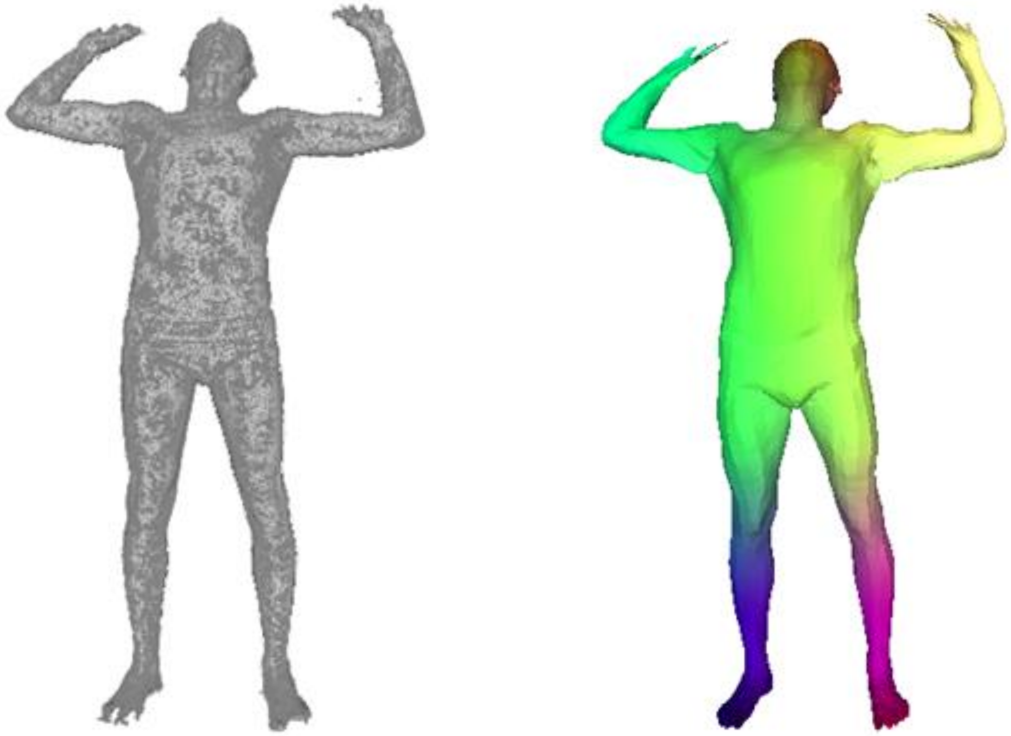
  Chamfer distance:

$$L(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^{n} \min_j \|x_i - y_j\|^2 + \sum_{j=1}^{n} \min_i \|x_i - y_j\|^2$$

  Earth mover distance:

$$L(\mathcal{X}, \mathcal{Y}) = \min_\pi \sum_{i=1}^{n} \|x_i - y_{\pi(i)}\|^2$$

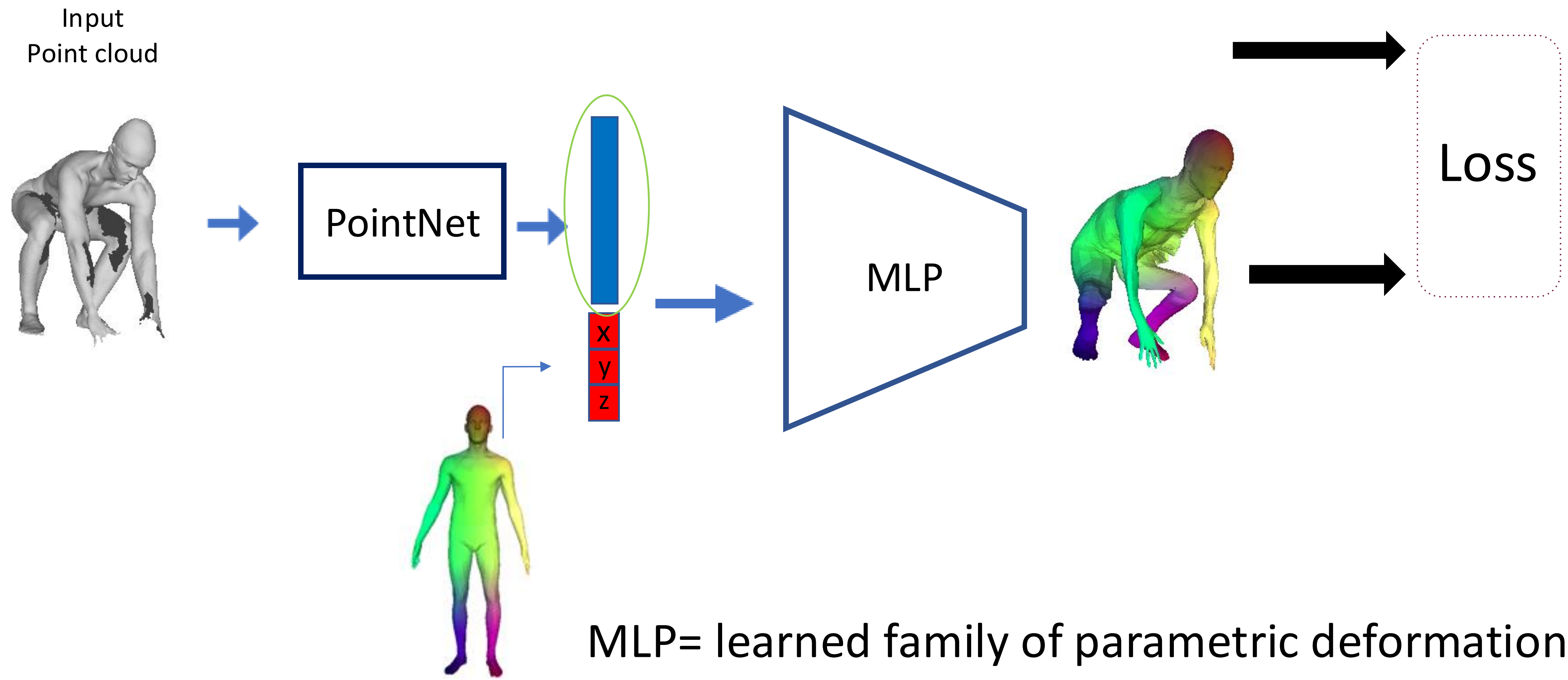
Results



Results

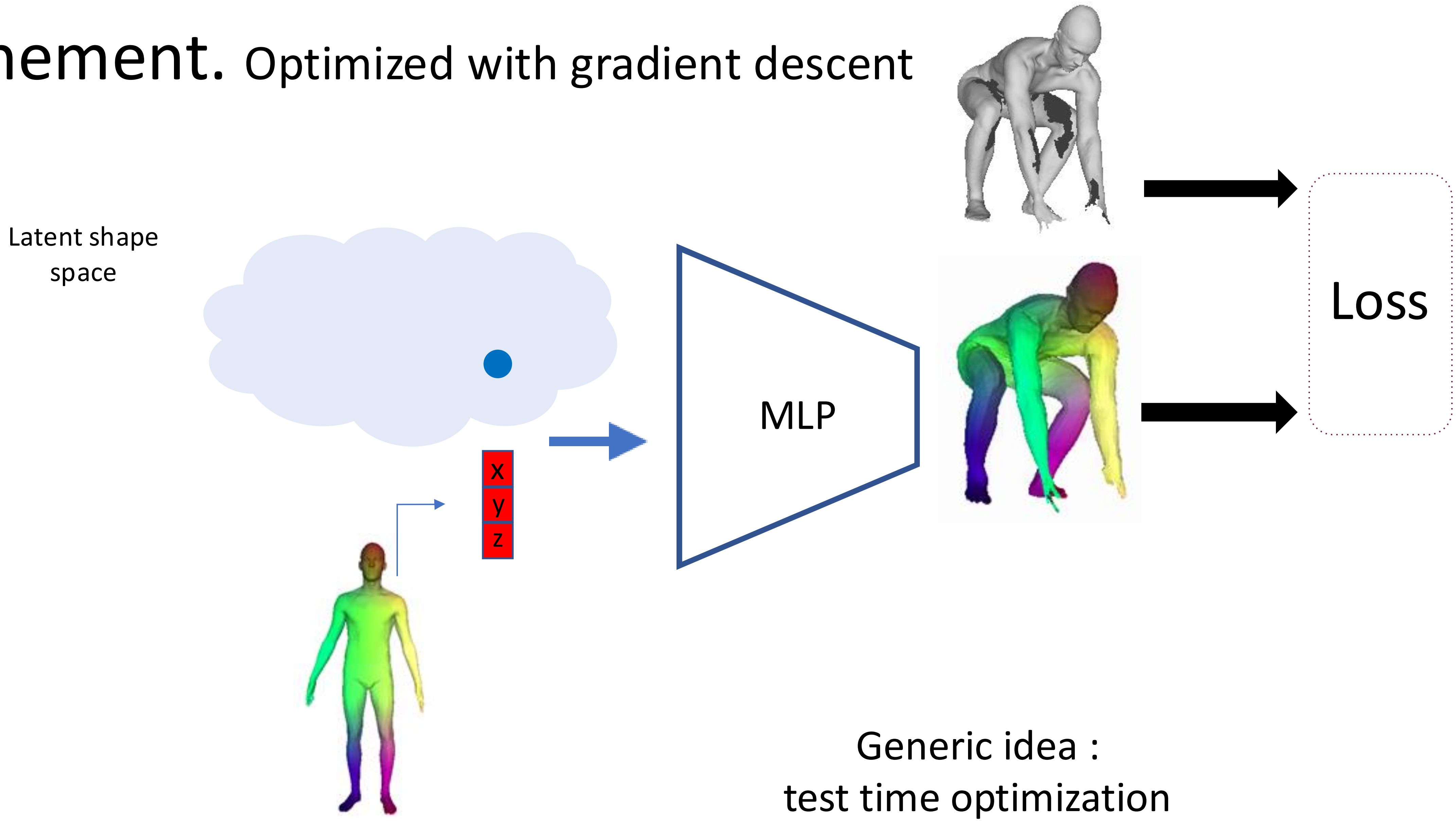
The nearest neighbors are likely to be poor

Refinement.



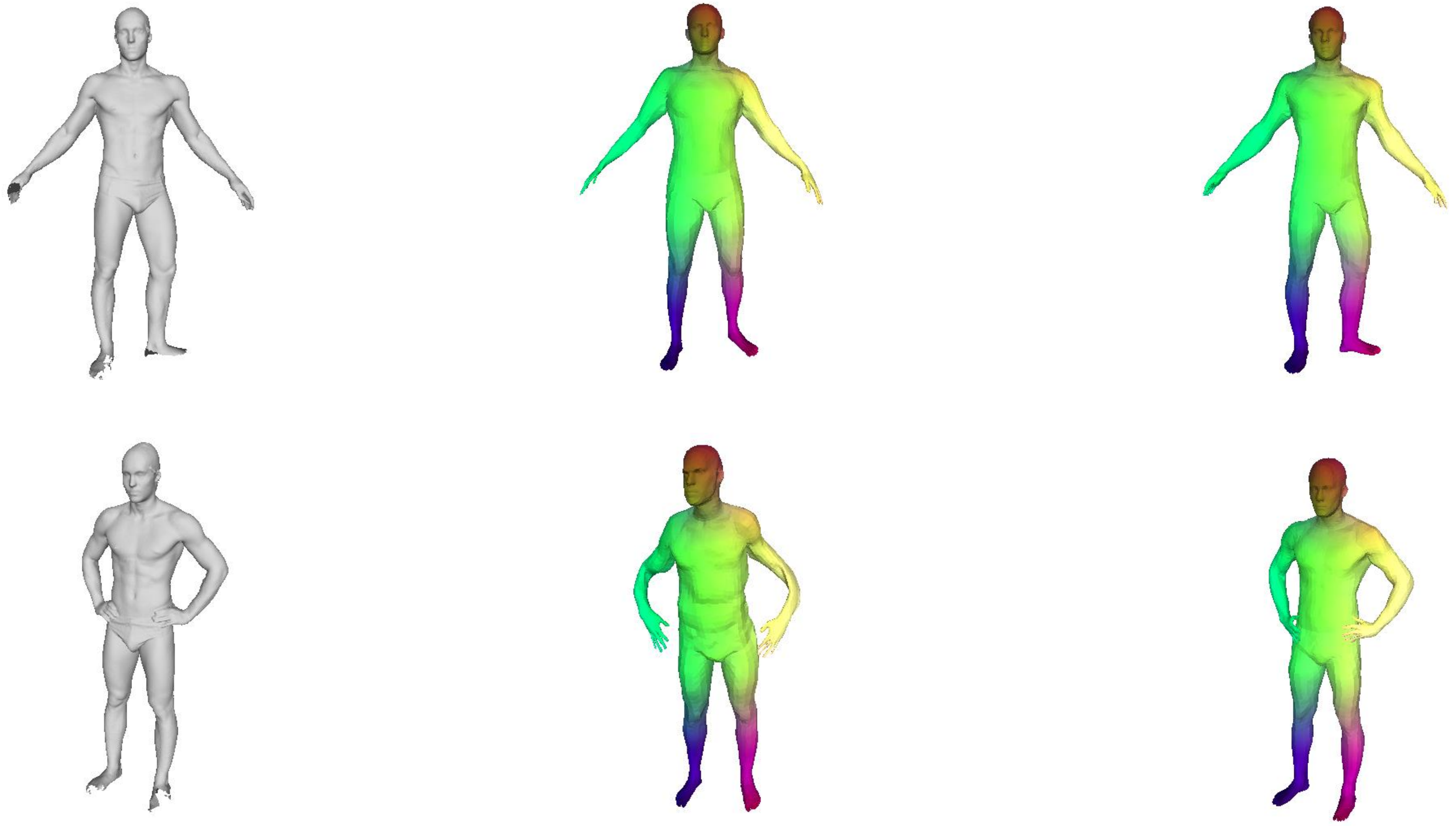
MLP= learned family of parametric deformations

Refinement. Optimized with gradient descent

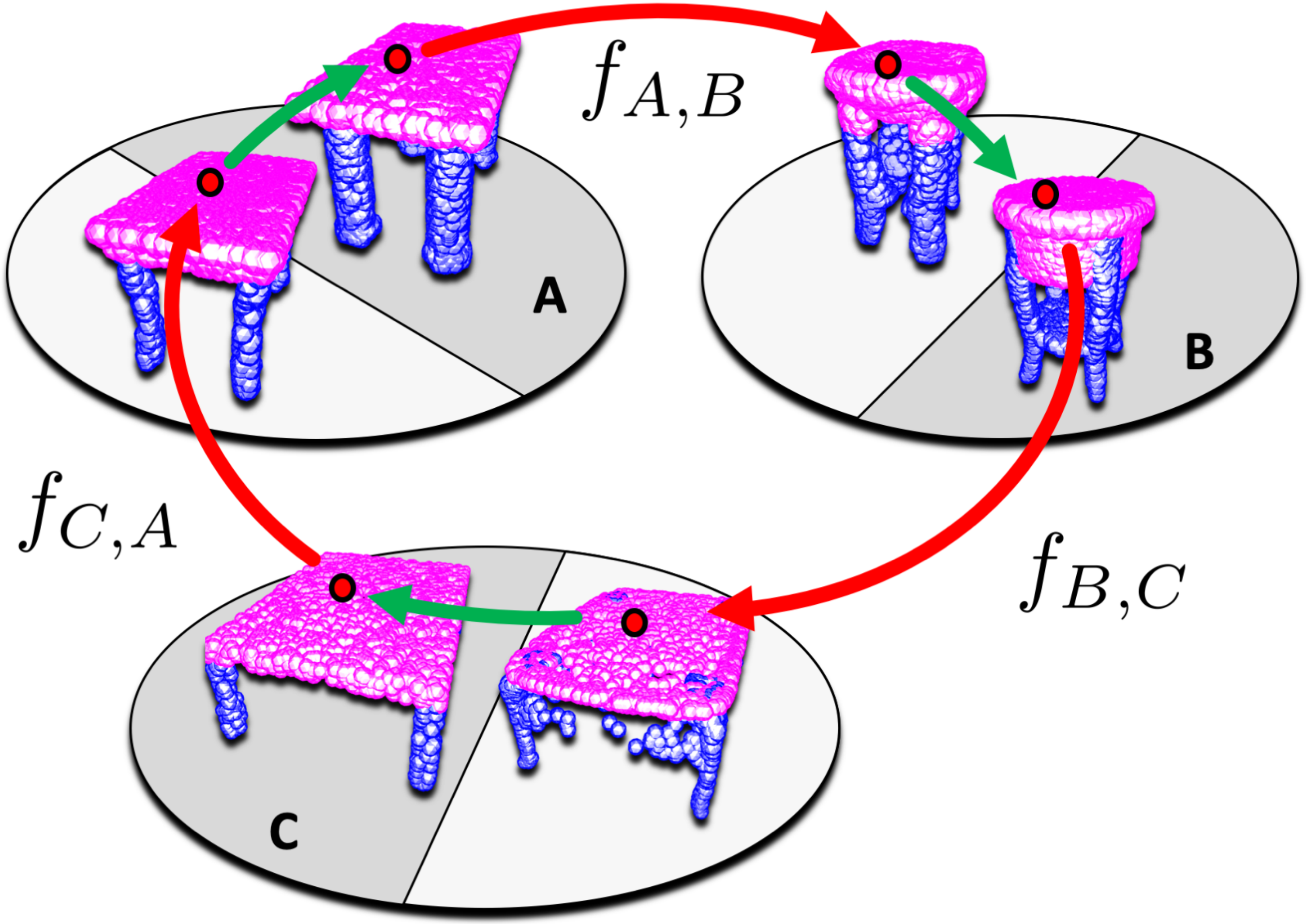


Generic idea :
test time optimization

Input Shape | Deformed Template | Optimized reconstruction



w/o template + w/ cycle consistency



T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, M. Aubry, Unsupervised cycle-consistent deformation for shape matching, SGP 2019

Key issue: 3D representation

- 2D views / Depth maps
- Voxels
- Points
- Meshes
- Parametric surface
- **Implicit surface**
- ”Procedural”

Outline: Deep learning and 3D data

Important milestones:

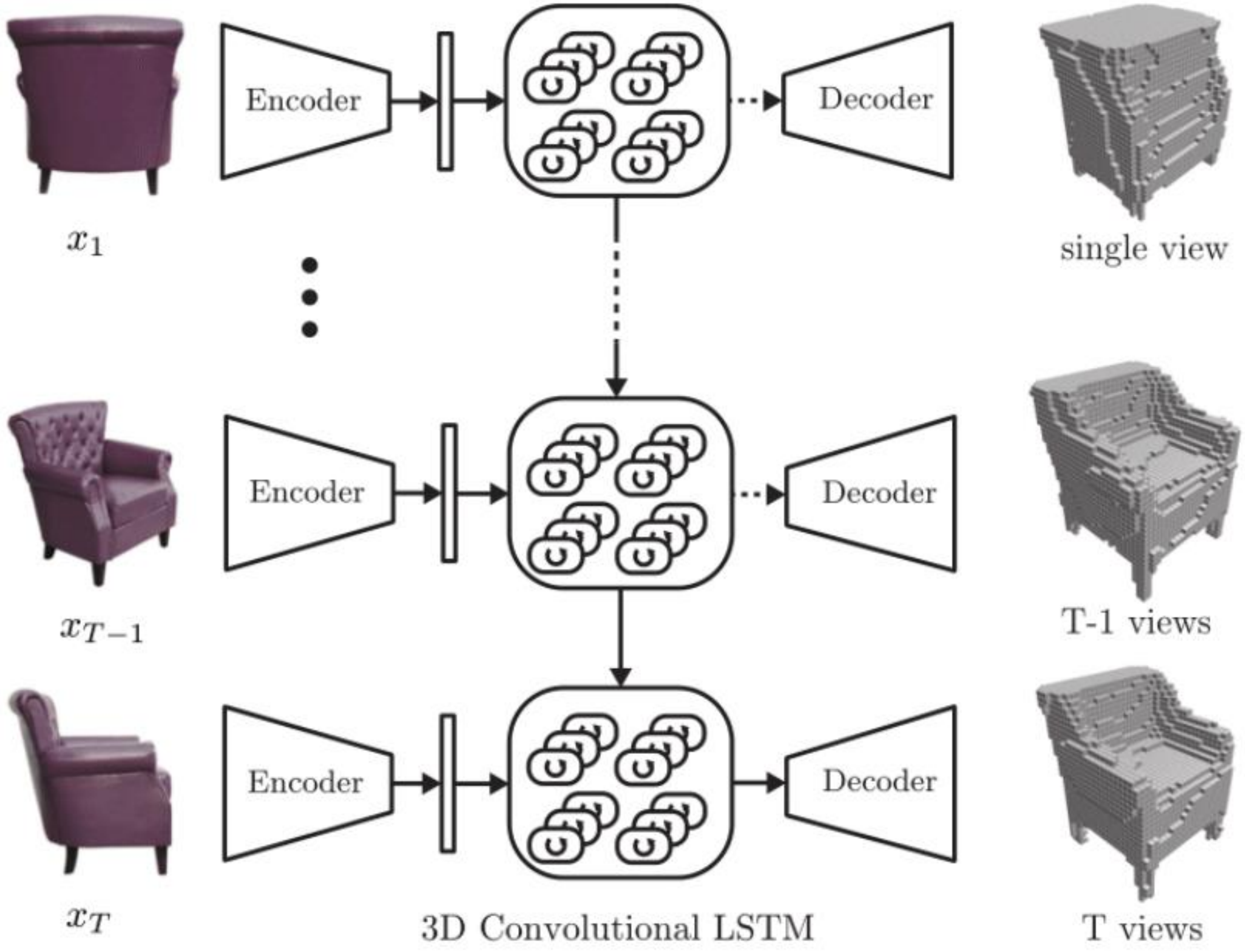
1. Classification and Segmentation
2. Matching / Alignment
3. **Generation and single view reconstruction**

Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

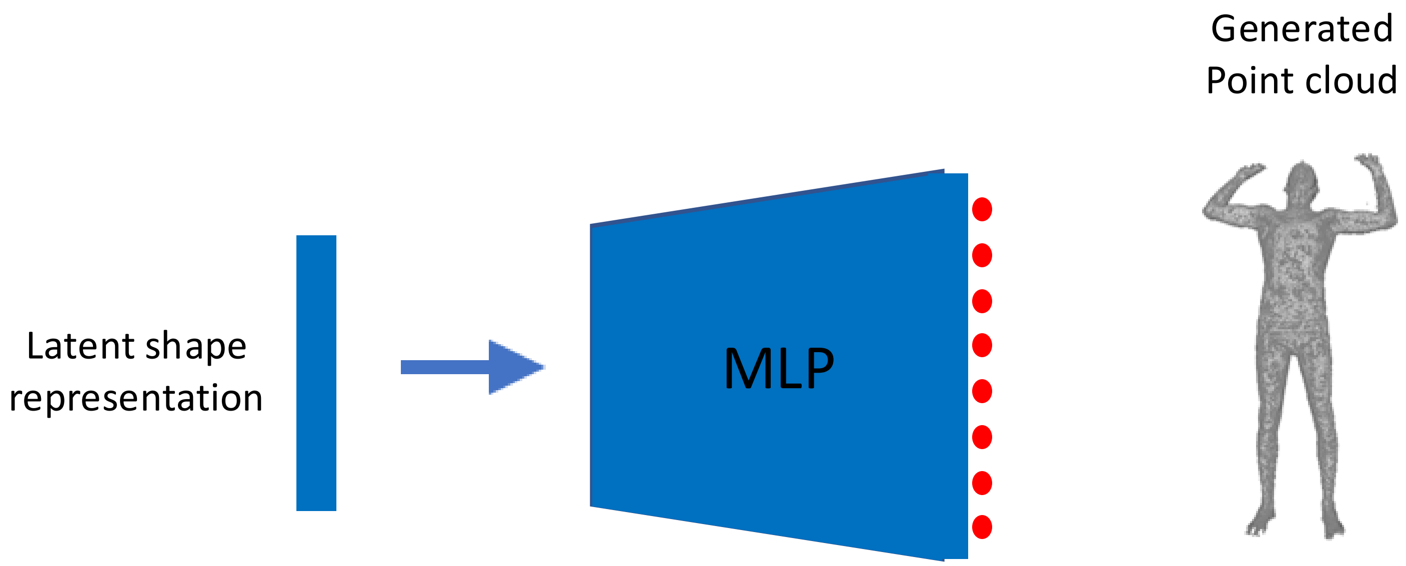
Learning with synthetic data

Voxels



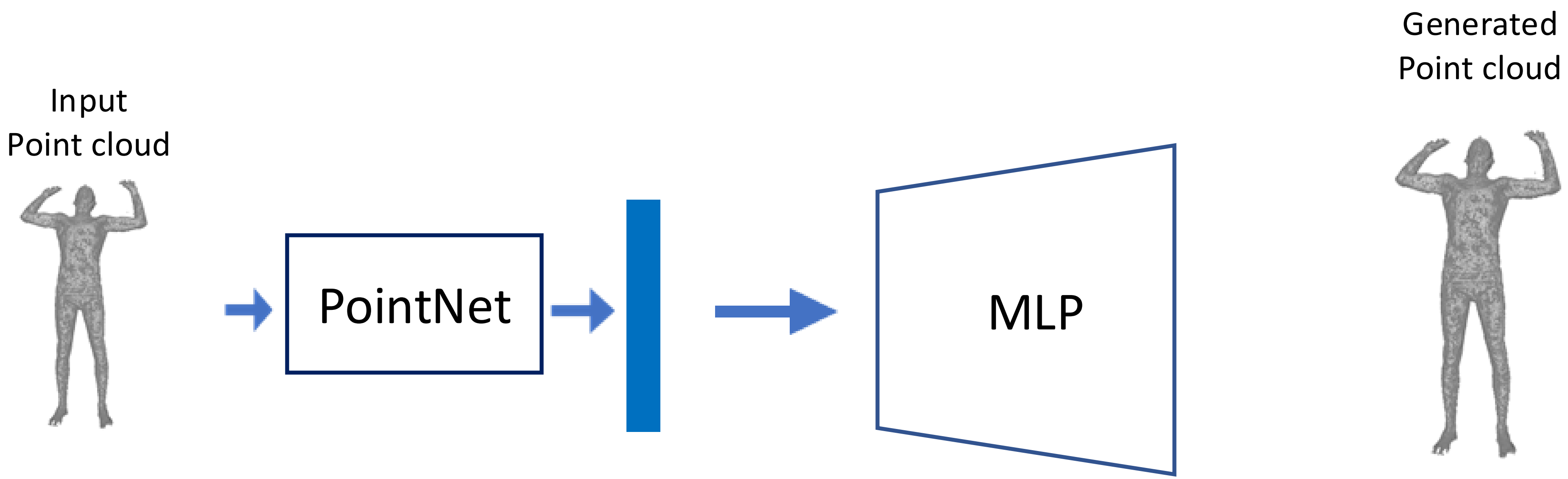
Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S.
3D-R2N2: A unified approach for single and multi-view 3d object reconstruction. ECCV 2016

Points



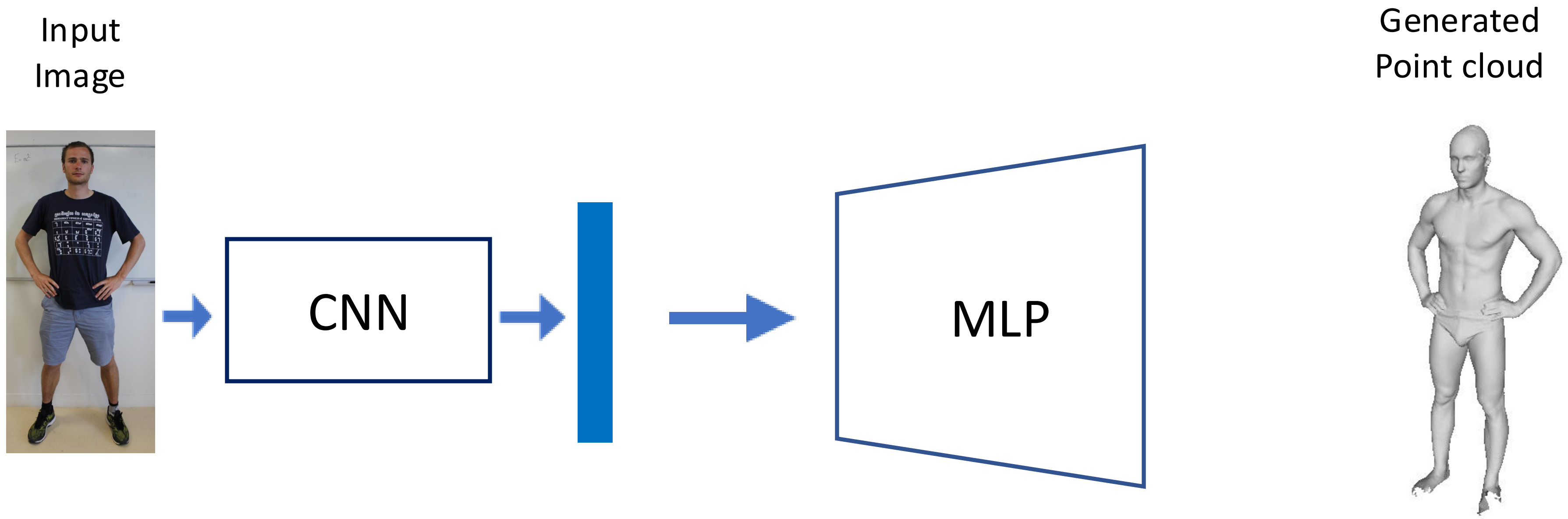
Fan, H., Su, H., & Guibas, L. J. A point set generation network for 3d object reconstruction from a single image, CVPR 2017

Points



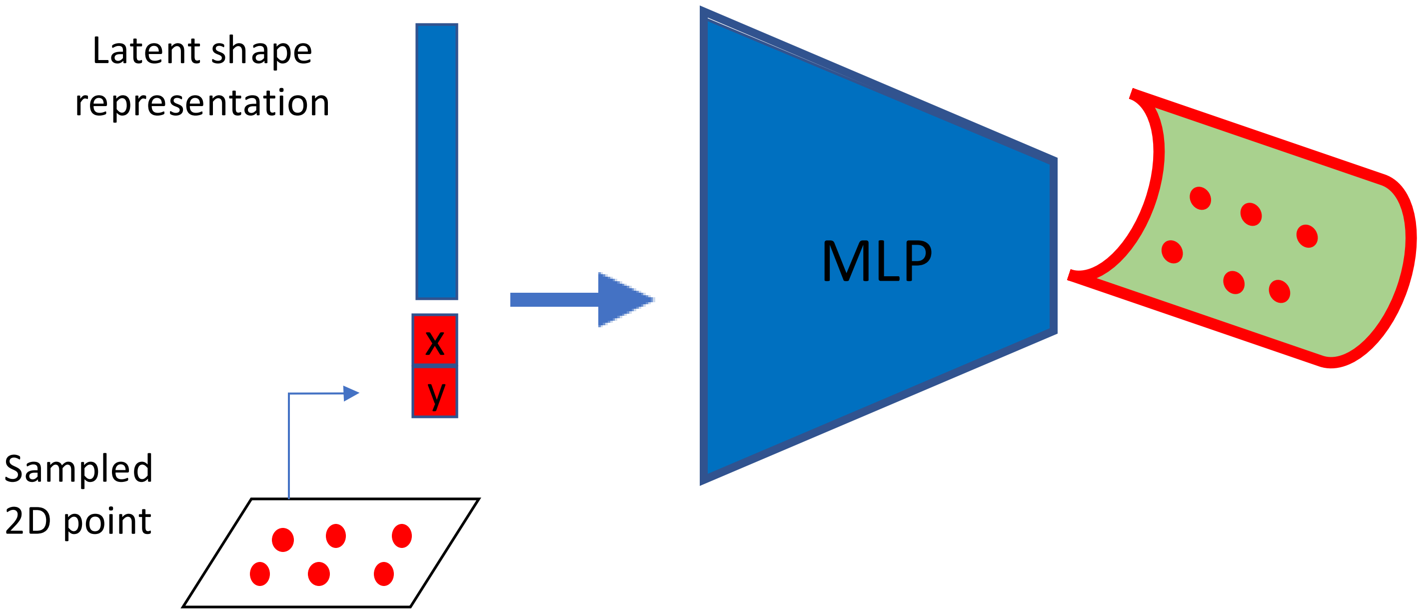
Fan, H., Su, H., & Guibas, L. J. A point set generation network for 3d object reconstruction from a single image, CVPR 2017

Points

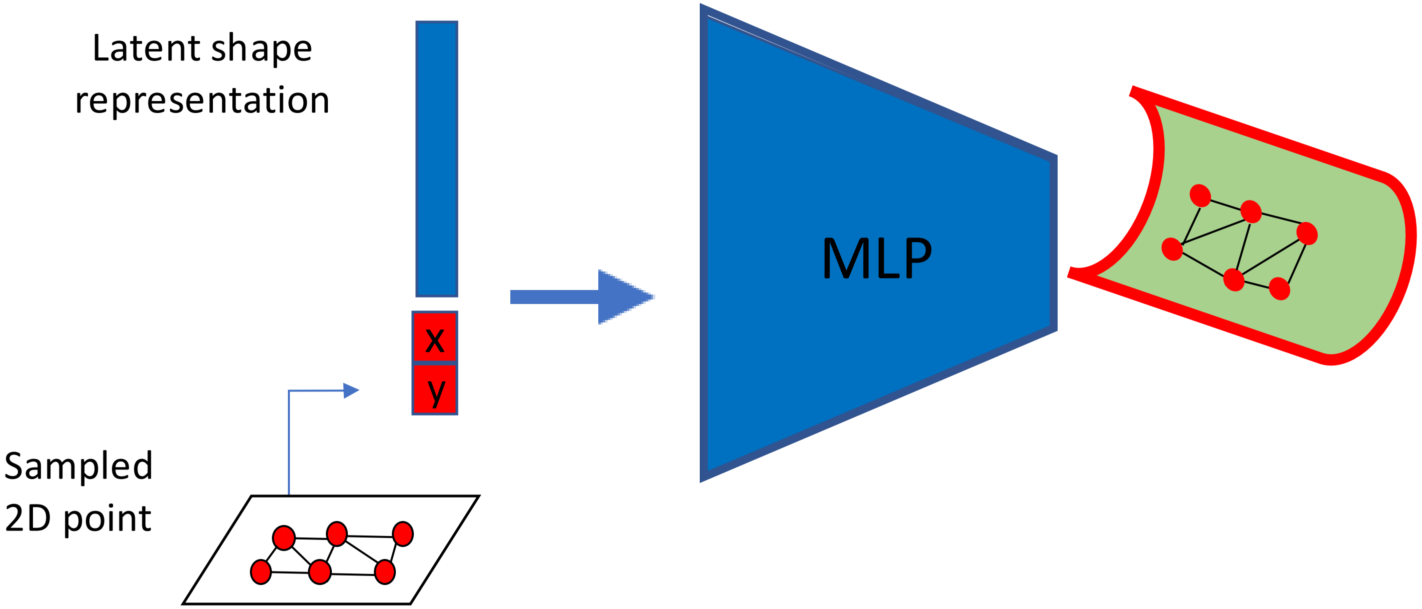


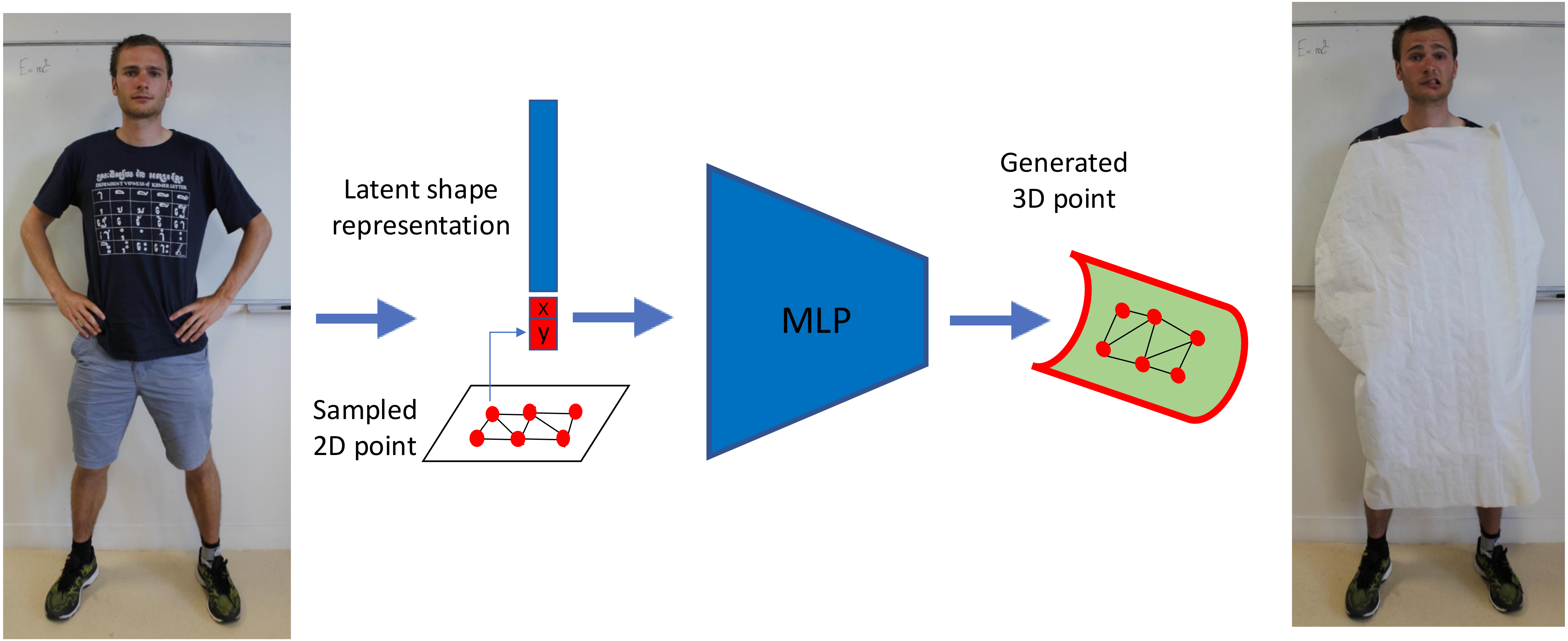
Fan, H., Su, H., & Guibas, L. J. A point set generation network for 3d object reconstruction from a single image, CVPR 2017

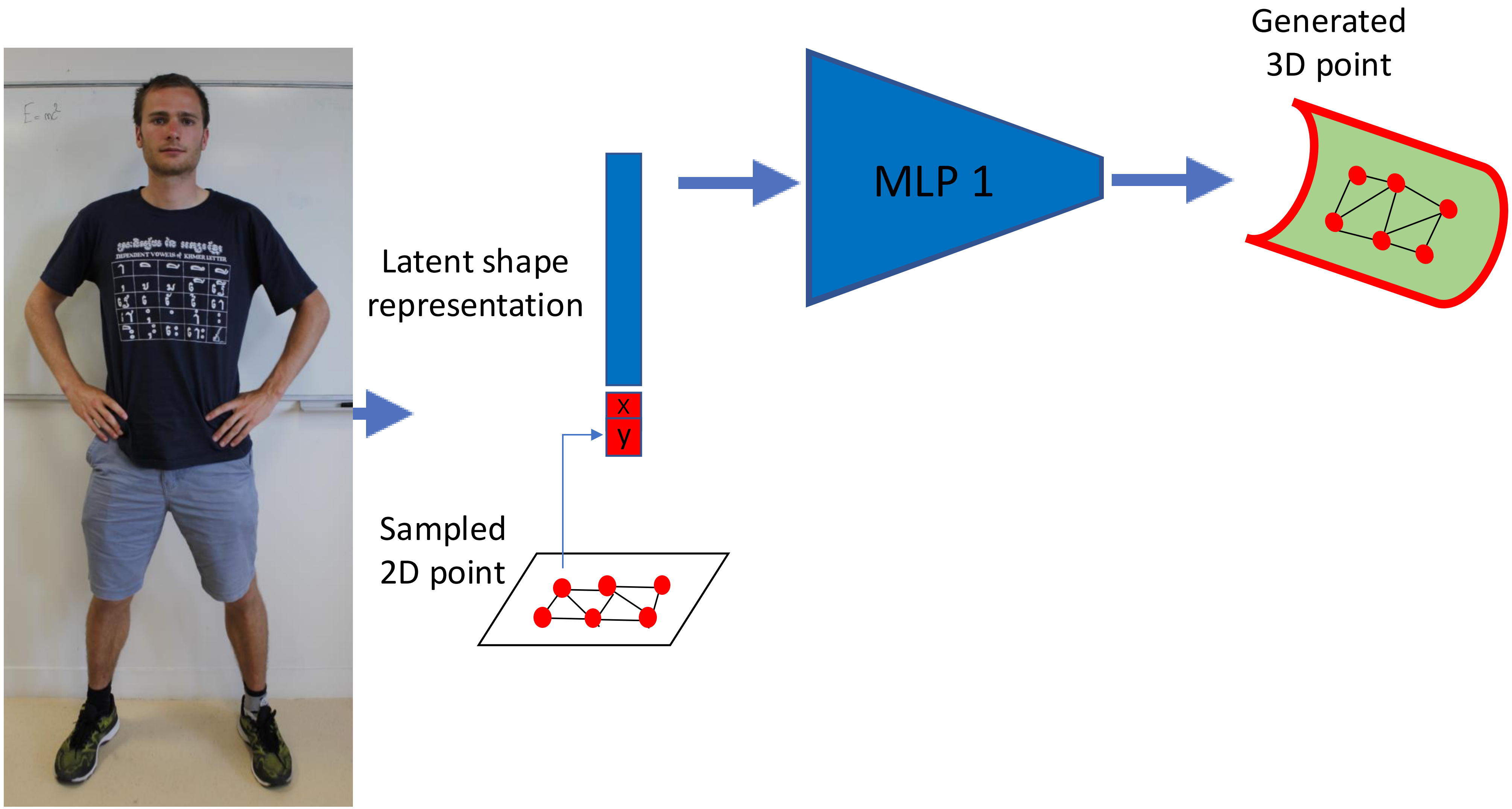
Parametric surface: Deform a unit square



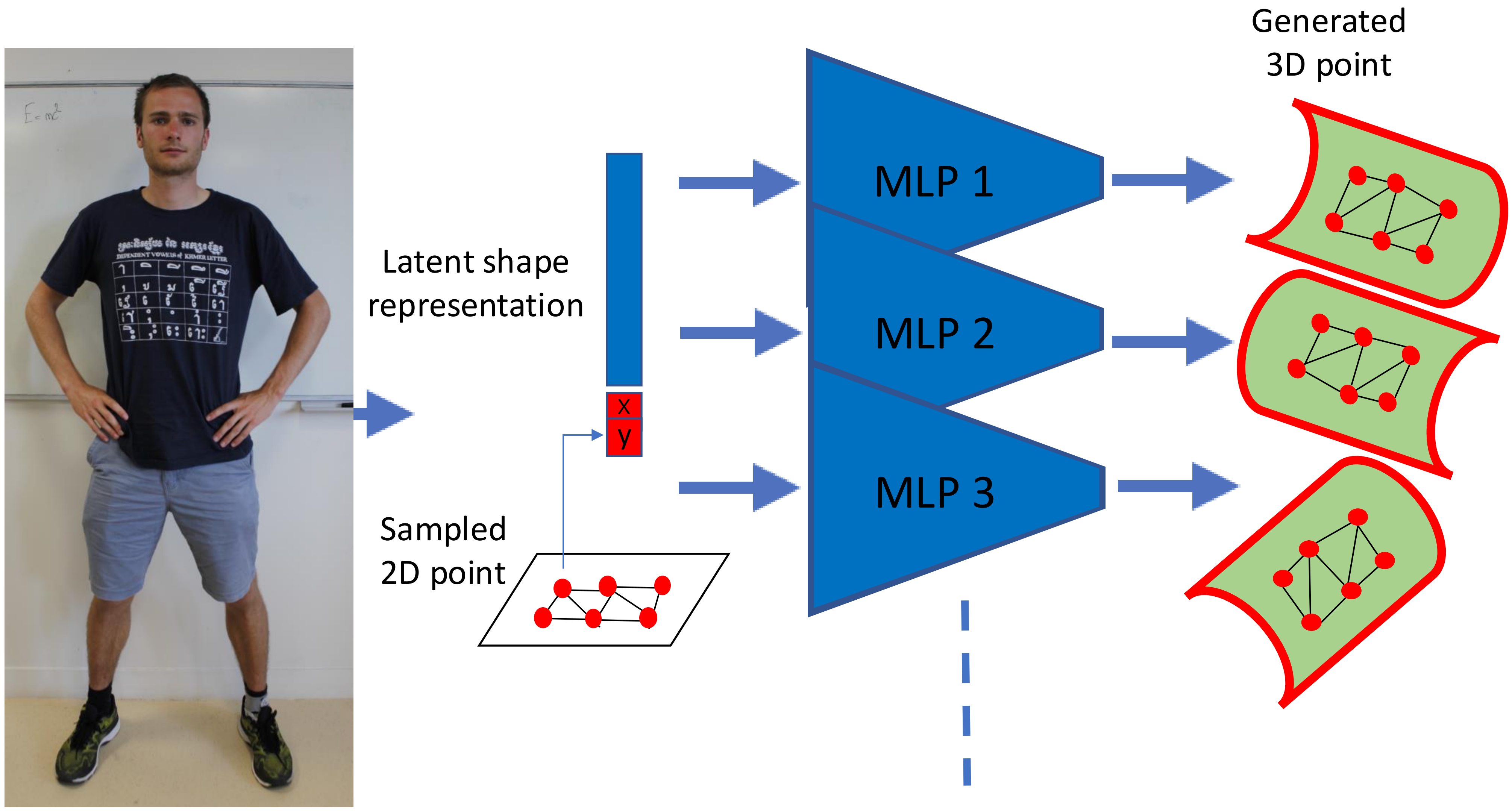
Parametric surface: Deform a unit square

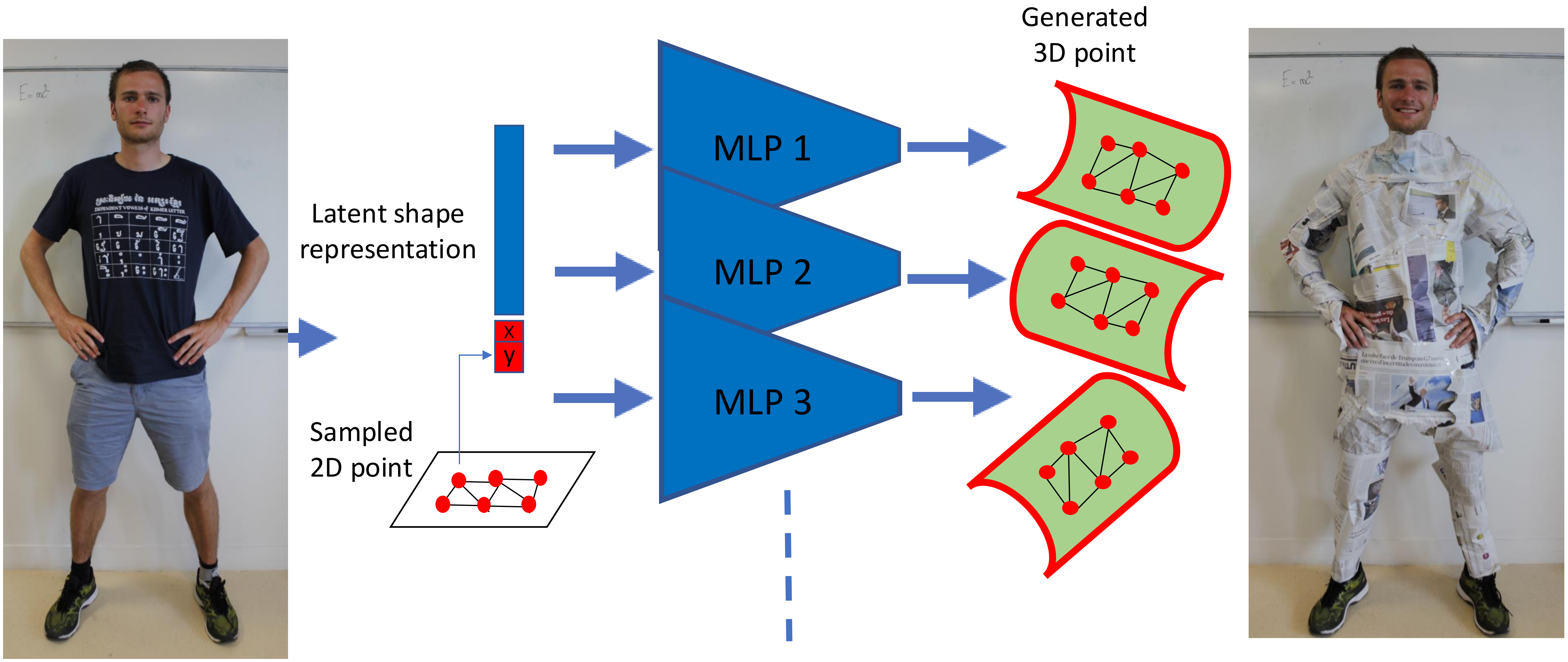



Latent shape
representation
x
y
Sampled
2D point
MLP
Generated
3D point



Latent shape representation
x
y
Sampled 2D point
MLP 1
Generated 3D point

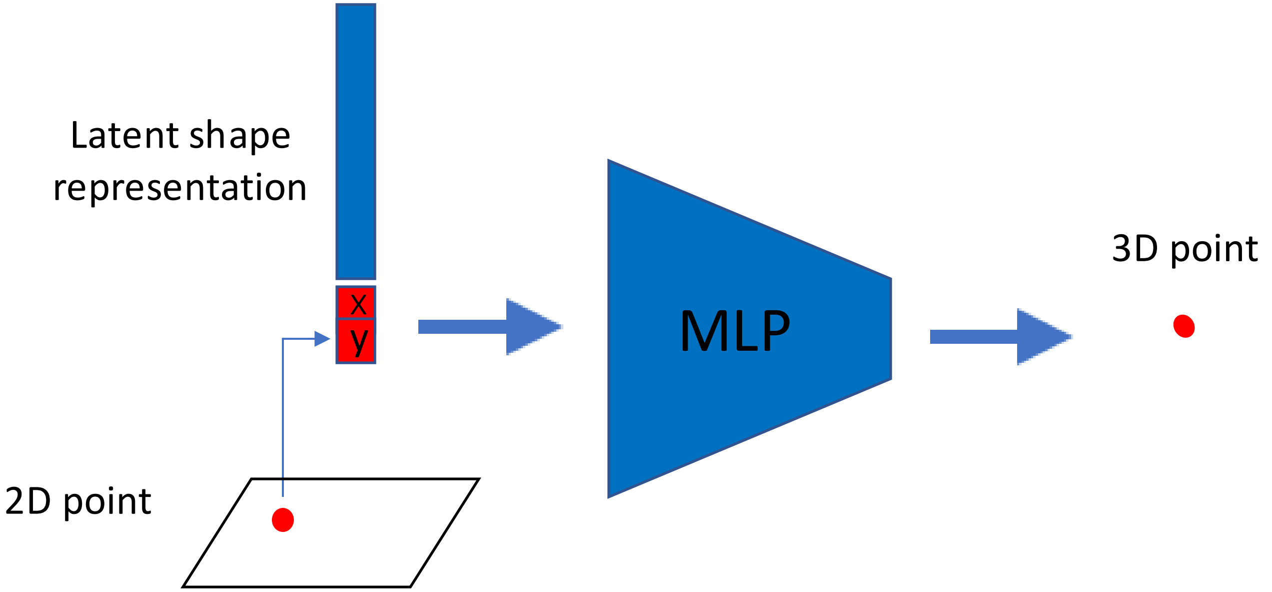

Generated
3D point
Latent shape
representation
MLP 1
MLP 2
MLP 3
x
y
Sampled
2D point



Learnt simply by sampling many points and minimizing Chamfer distance

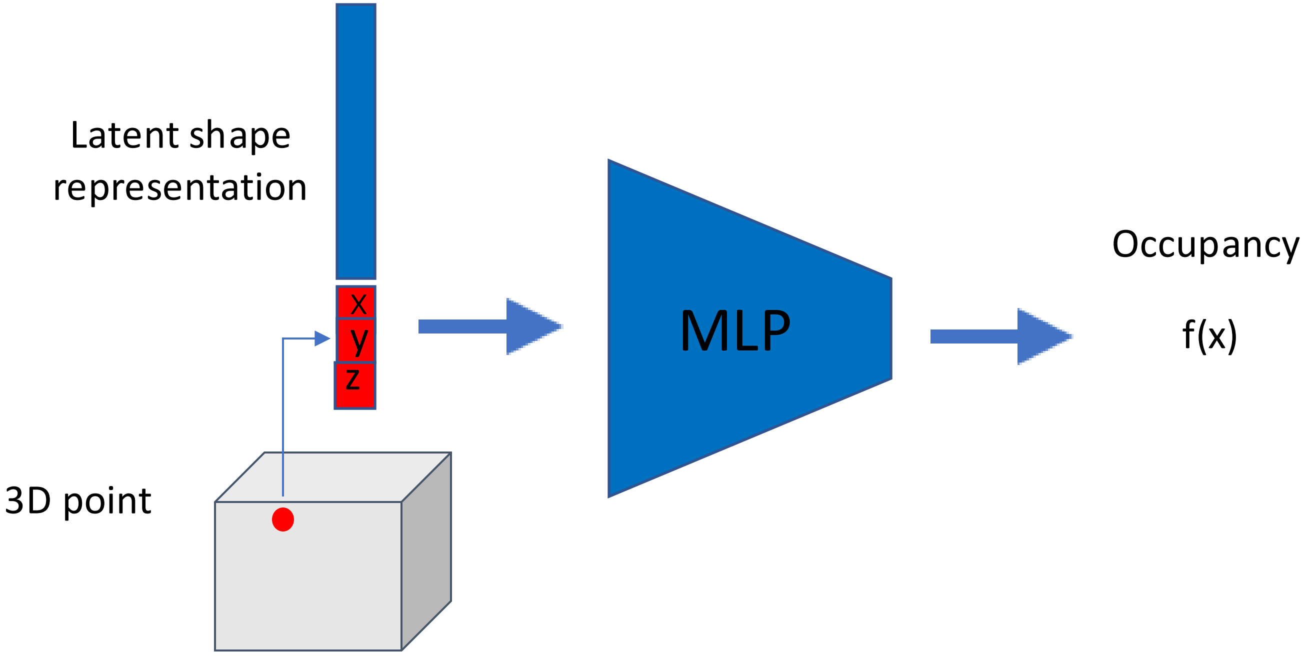
T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, M. Aubry, AtlasNet: A papier-mâché approach to learning 3d surface generation, CVPR. 2018

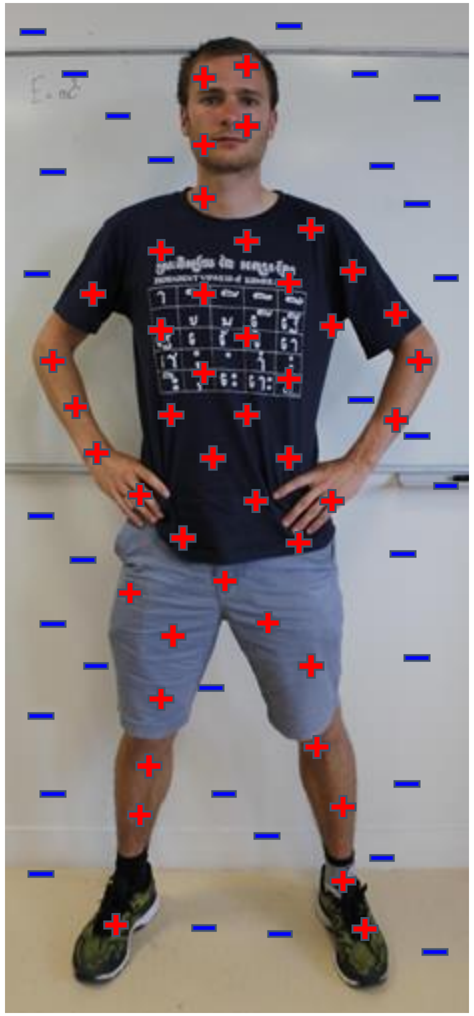
Parametric surface



Parametric volume [Mescheder2019, Park2019, Chen2019]

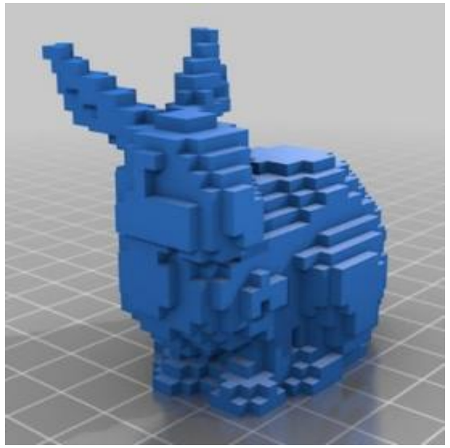
Park, J. J., Florence, P., Straub, J., Newcombe, R., & Lovegrove, S.
Deepsdf: Learning continuous signed distance functions for shape representation
Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A.
Occupancy networks: Learning 3d reconstruction in function space.
Chen, Z., & Zhang, H.
Learning **implicit fields** for generative shape modeling.



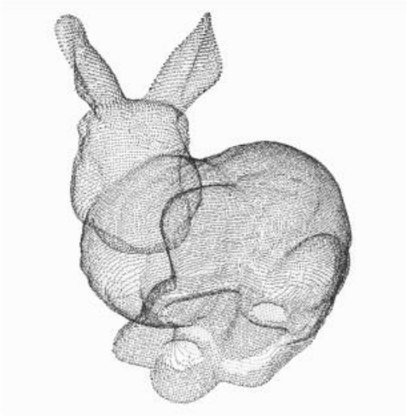


Summary: 3D shape representations for deep generation

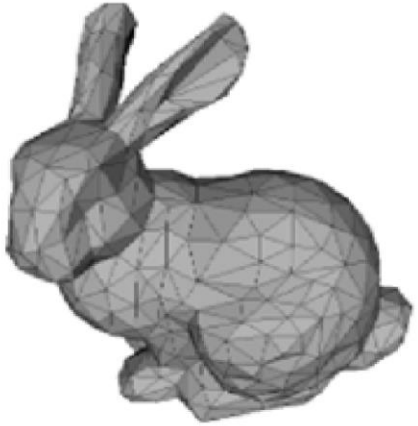
2016 | 2017 | 2018 | 2019



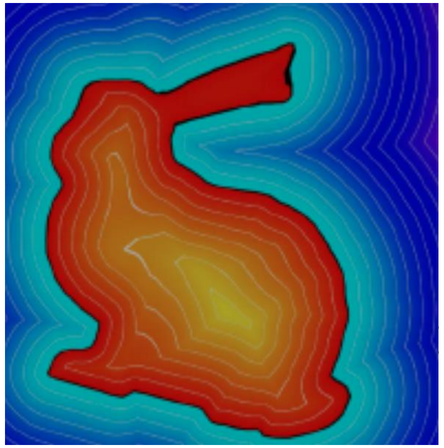
Voxels
Choy et al.



Point Clouds
Fan et al.



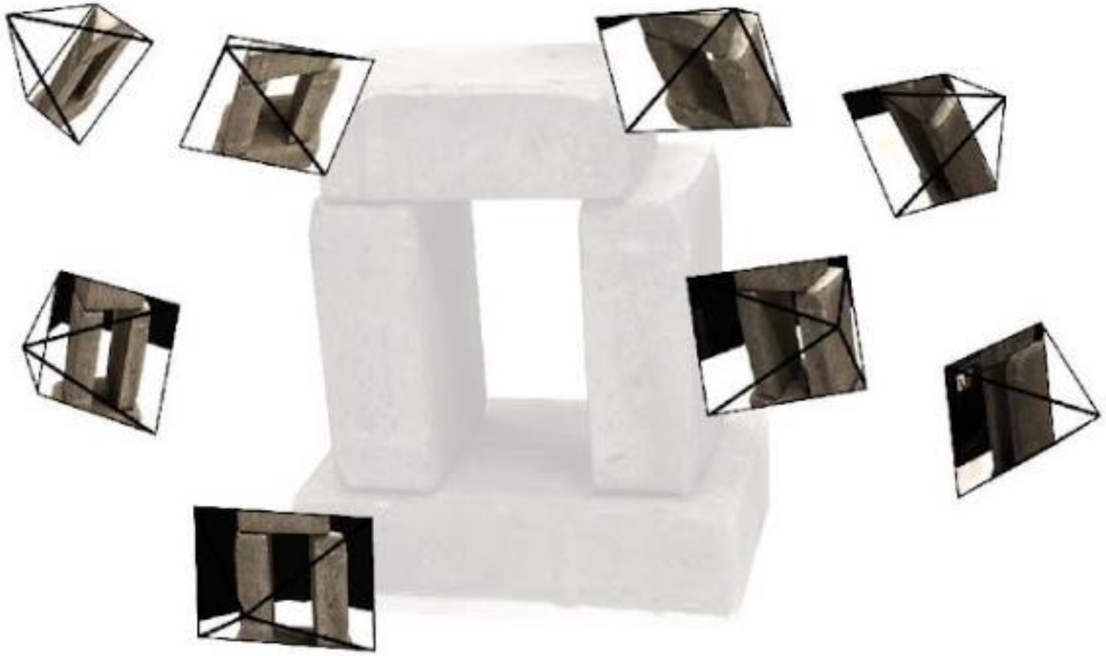
Mesh
Groueix et al.



Park et al. ; Chen et al.
Mescheder et al.

Slide from Thibault Groueix

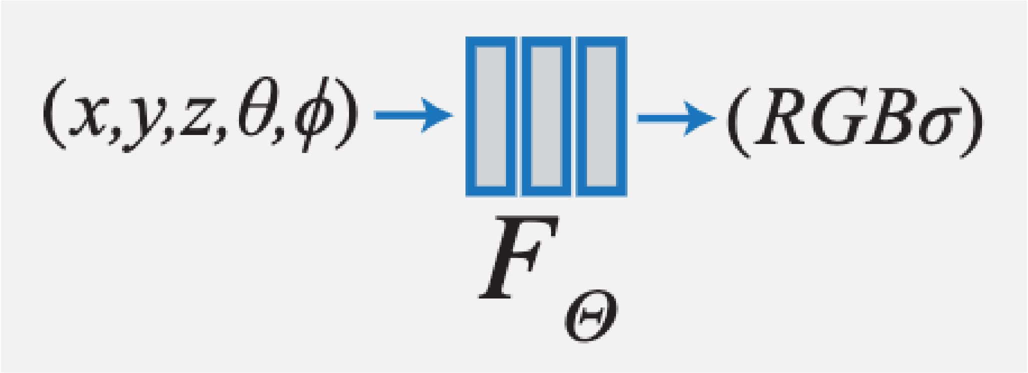
Parametric scene / Nerf [Mildenhall20]



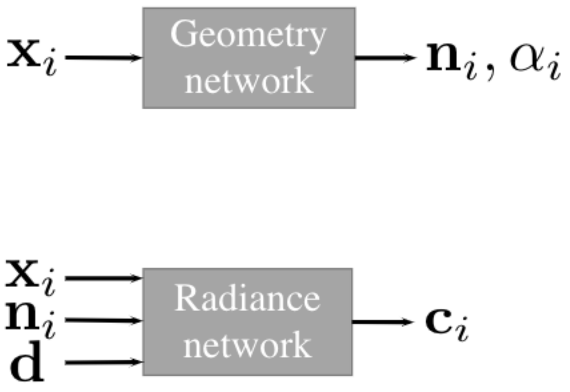
Input: a set of calibrated images

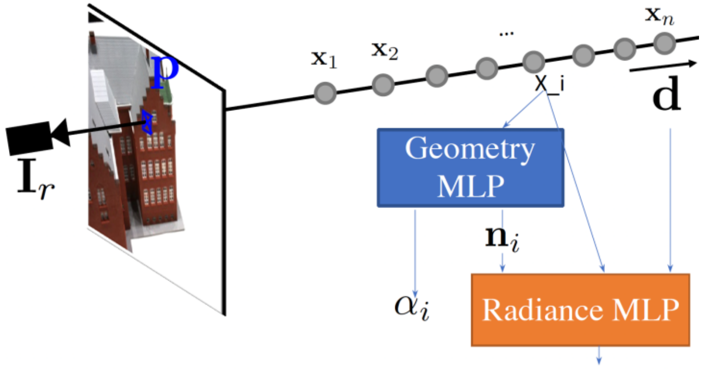
Output: rendering from any viewpoint
(from a scene model)

Parametric scene / Nerf [Mildenhall20]



More in details





$$\boxed{V(\mathbf{p}) = \sum_{i=1}^{n} \alpha_i \prod_{j<i} (1-\alpha_j)\mathbf{c}_i}$$

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

Recent works I am excited about:

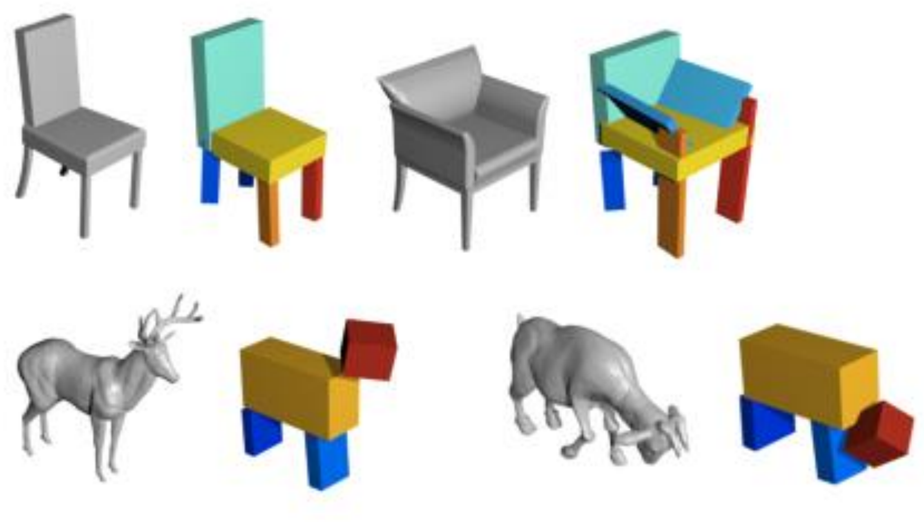
4. **Structured generation**
5. Unsupervised single view reconstruction

Learning with synthetic data

Key issue: 3D representation

- 2D views / Depth maps
- Voxels
- Points
- Meshes
- Parametric surface
- Implicit surface
- **”Procedural”**

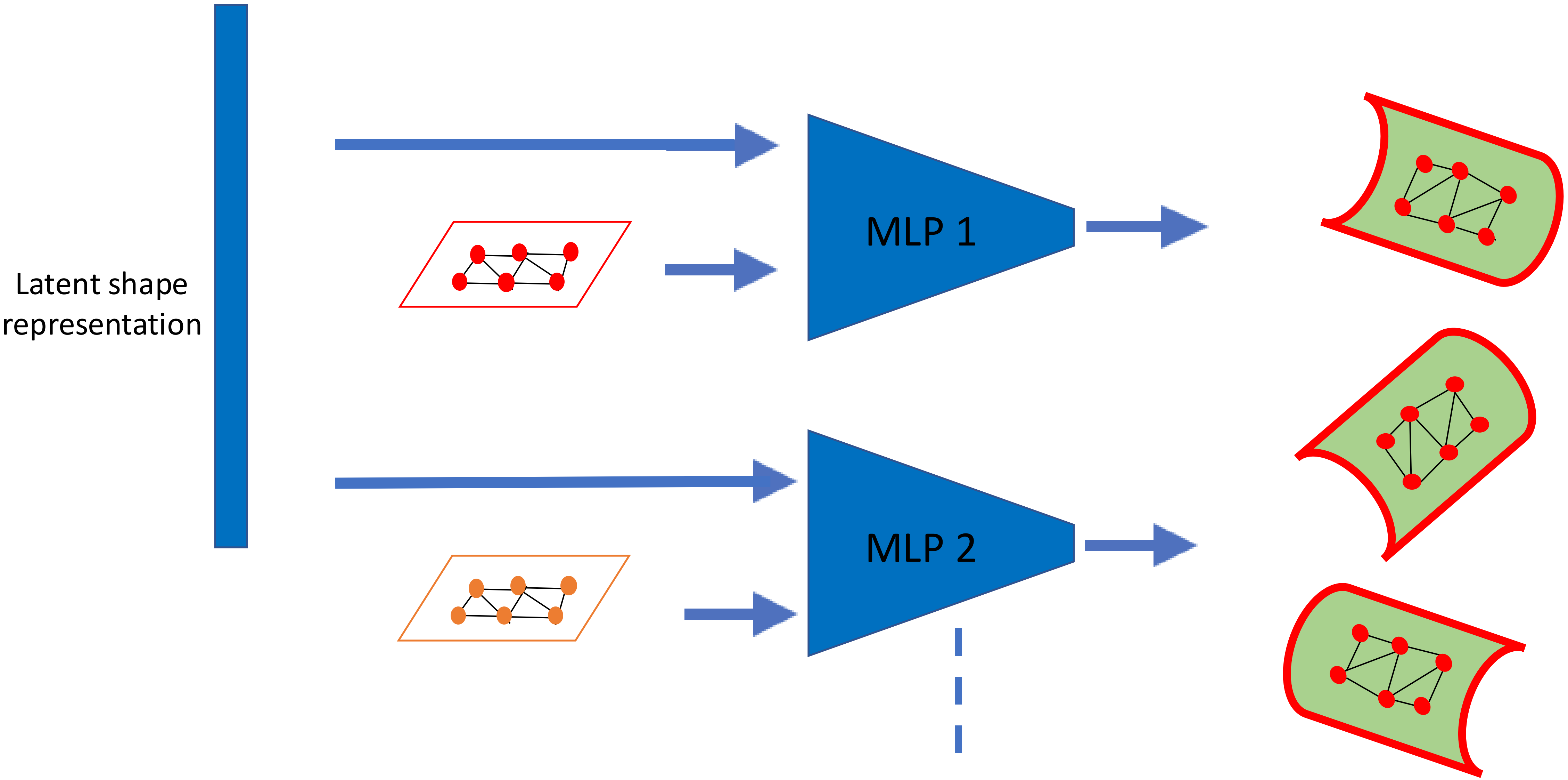
Learning to compose primitives



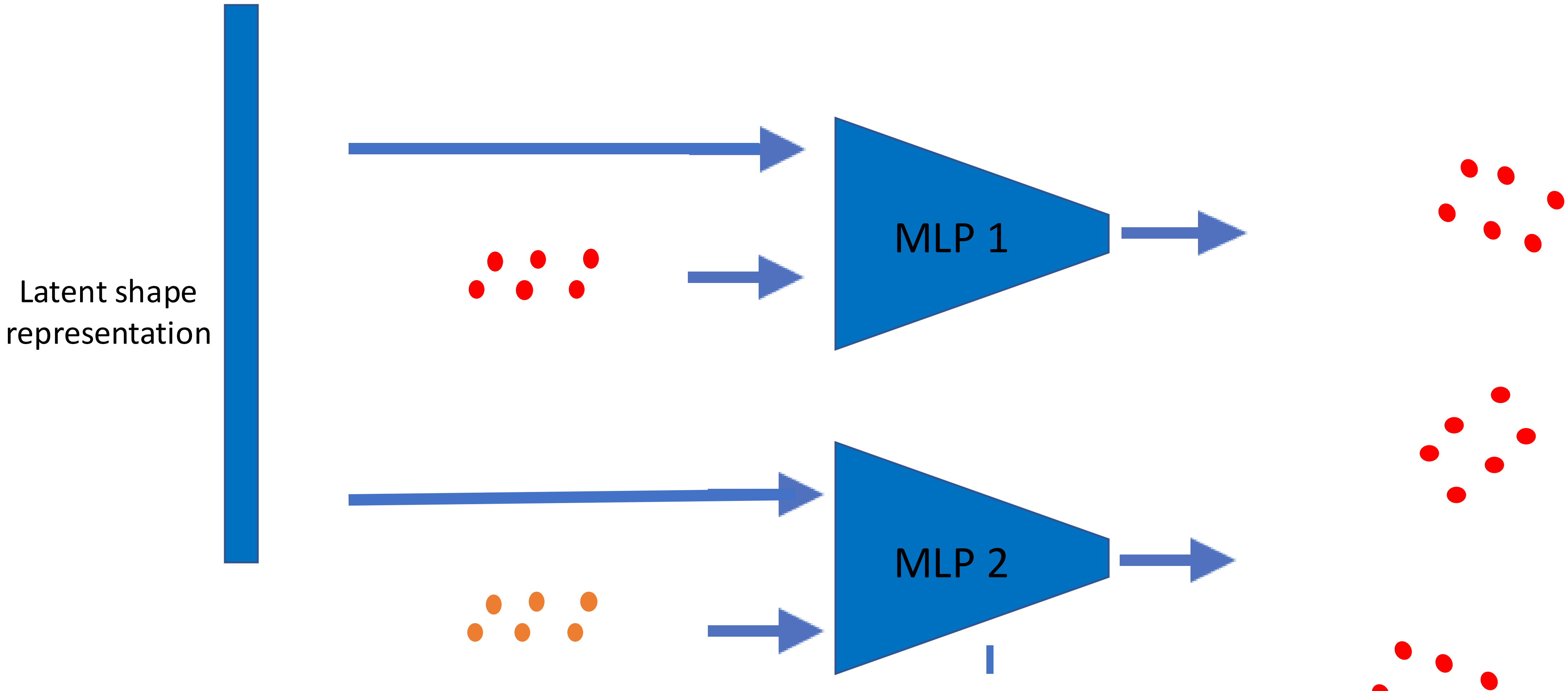
Learning Shape Abstractions by Assembling Volumetric Primitives, *Shubham Tulsiani , Hao Su, Leonidas J. Guibas, Alexei A. Efros, Jitendra Malik, CVPR 2017*

Superquadrics Revisited: Learning 3D Shape Parsing beyond Cuboids, *Despoina Paschalidou, Ali Osman Ulusoy, Andreas Geiger, CVPR 2018*

AtlasNet

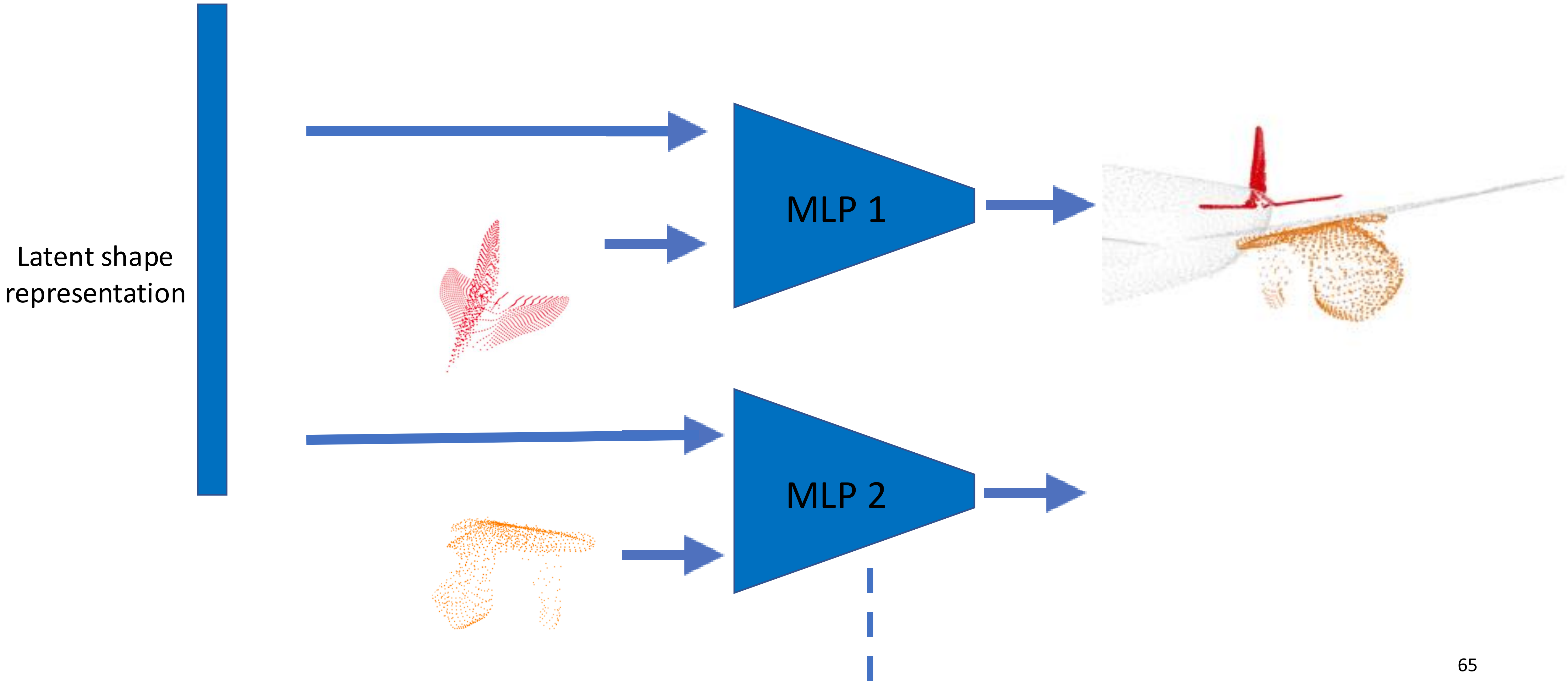


Learning elementary structures: Point Learning (AtlasNet v2)

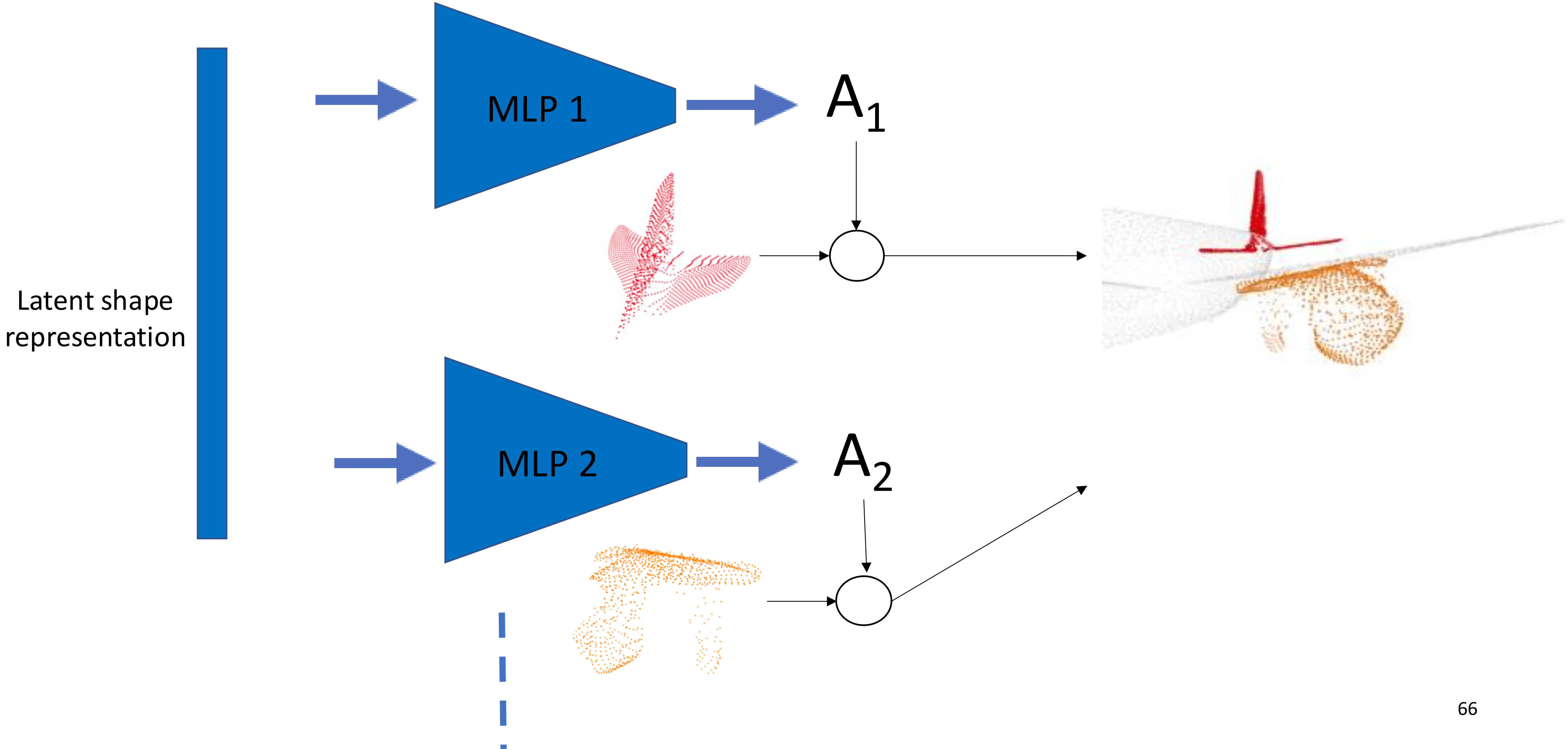


Learning elementary structures for 3D shape generation and matching
T. Deprelle, T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, M. Aubry
NeurIPS 2019

Learning elementary structures: Point Learning (AtlasNet v2)



Learning elementary structures

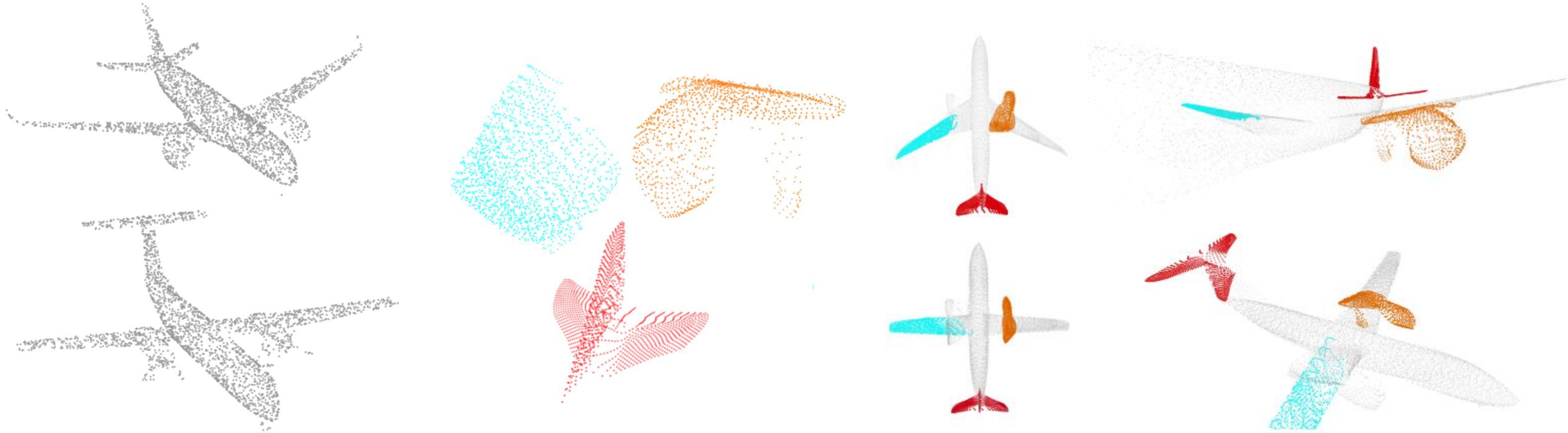


Results on Shapenet planes

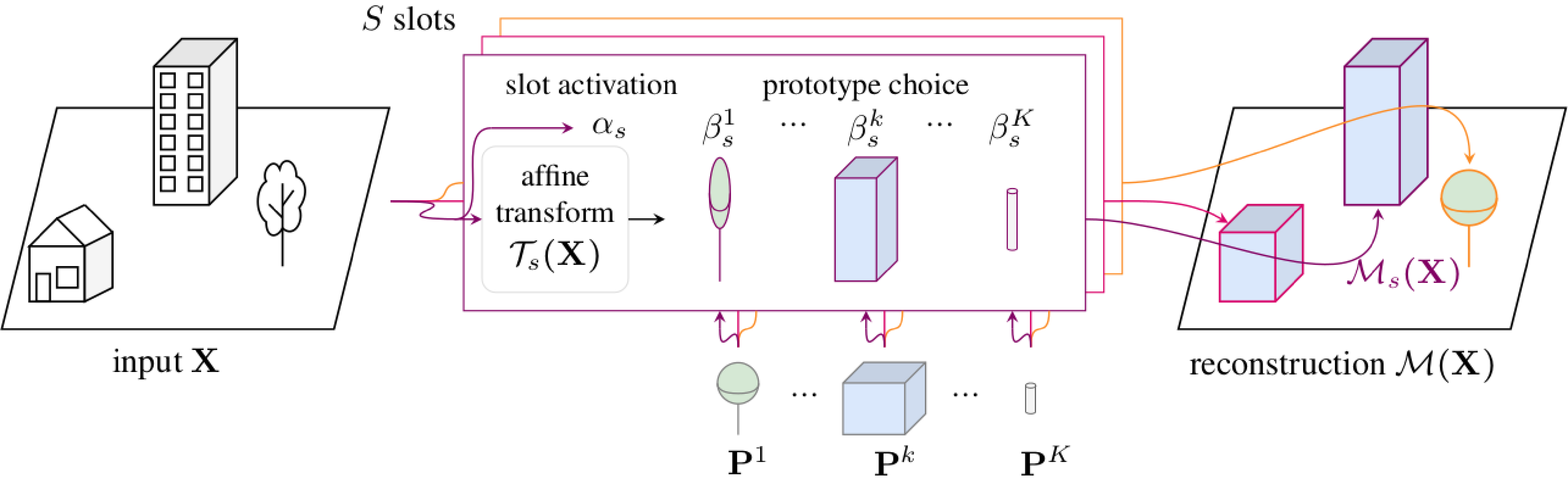
Input

Learned elementary structures

Reconstructions



Learnable Earth Parser

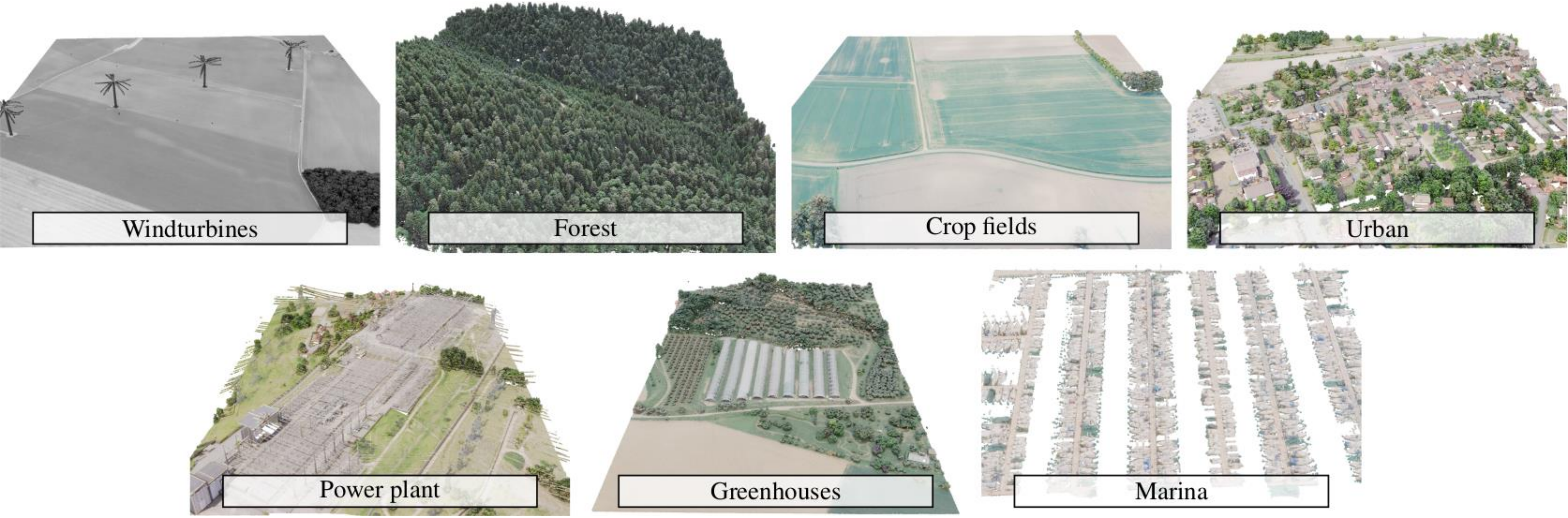


+ losses derived from a probabilistic scene model, developped in the paper

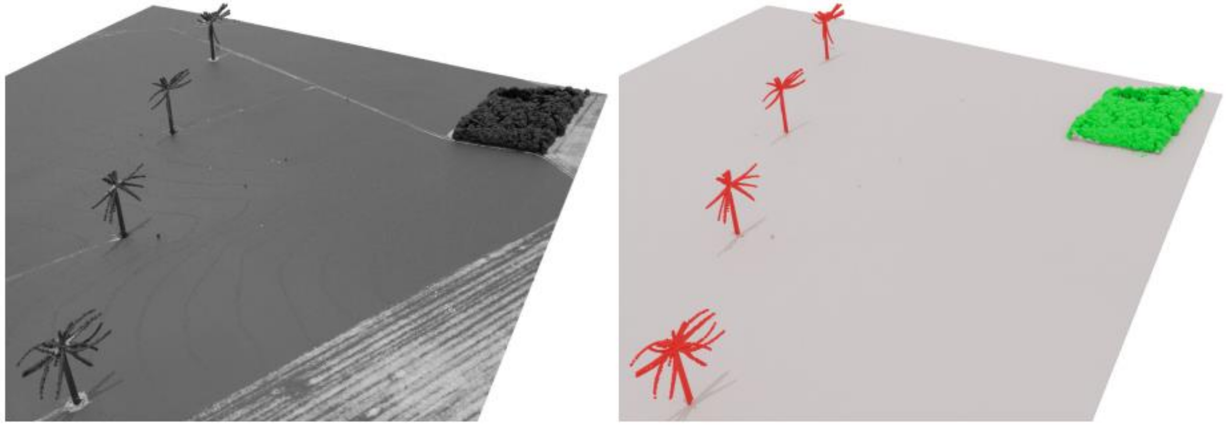
Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans R. Loiseau, E. Vincent, M. Aubry, L. Landrieu CVPR 2024

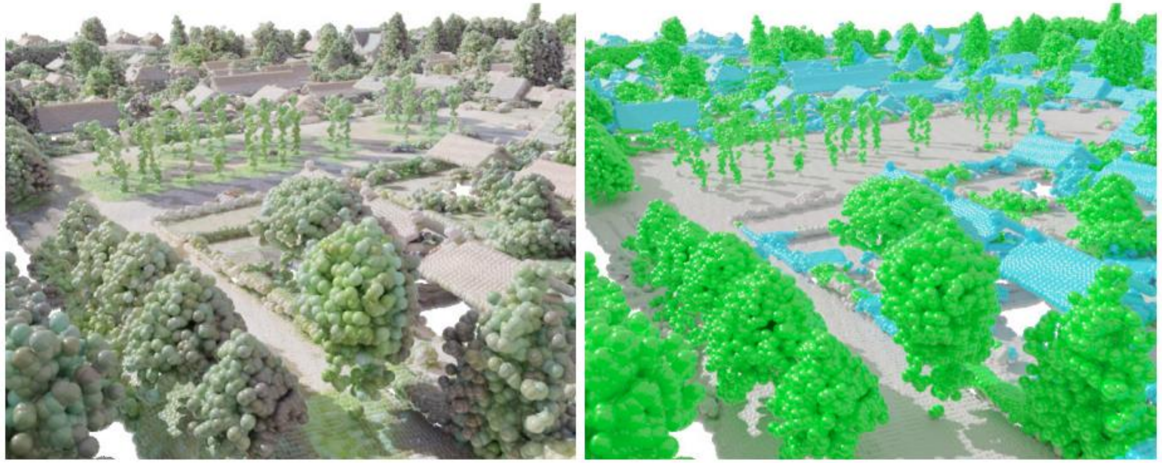
Data: LidarHD

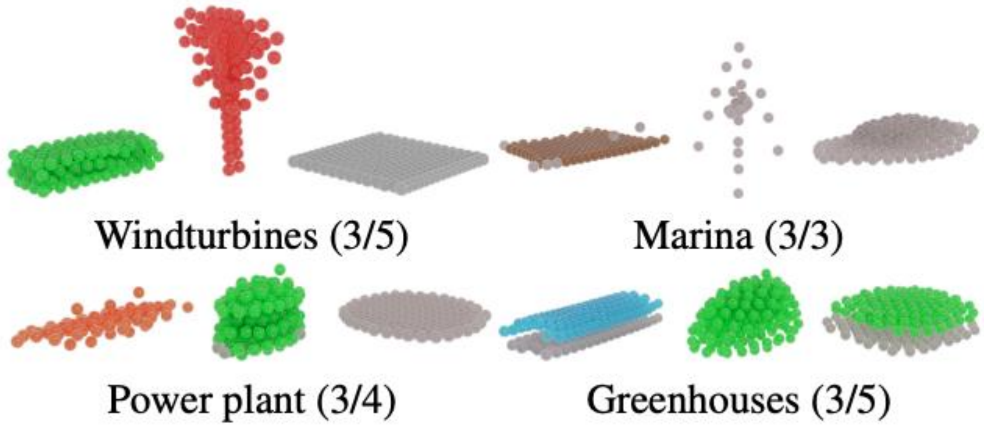
Name	Surface in km^2	# points $\times 10^6$	annotation ratio in %	num. of classes
Crop fields	1.1	19.7	77.4	2
Forest	1.1	46.7	97.8	2
Greenhouses	0.1	1.3	95.6	3
Marina	0.1	0.5	92.7	2
Power plant	0.2	8.6	78.4	4
Urban	1.1	15.7	95.9	3
Windturbines	4.2	5.6	—	—
Total	7.7	98.3	89.6	—

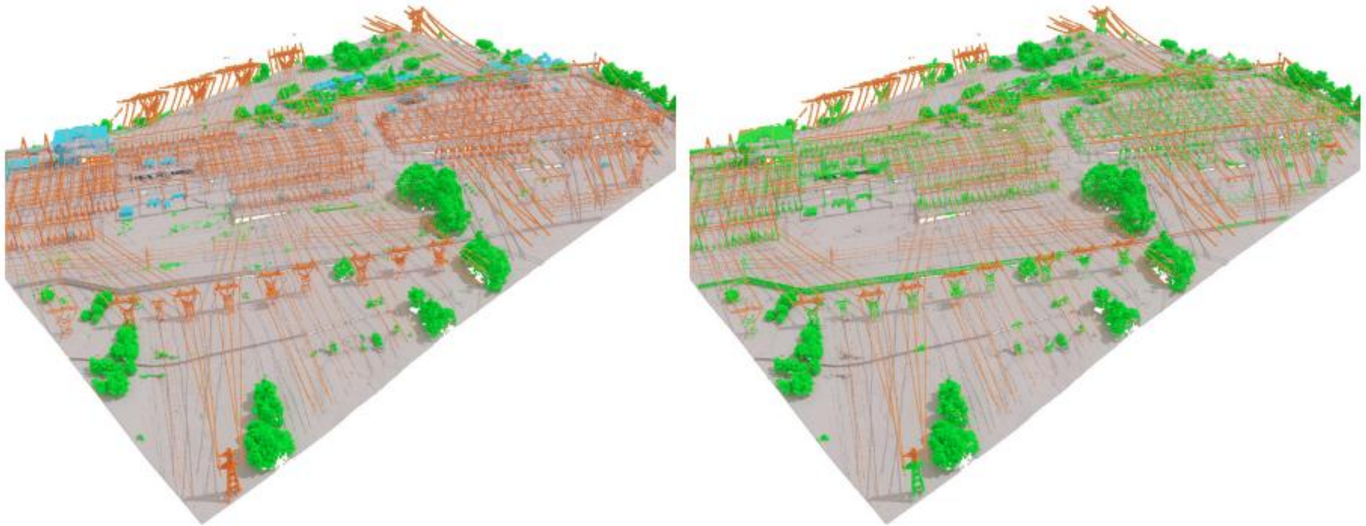


Semantic segmentation results



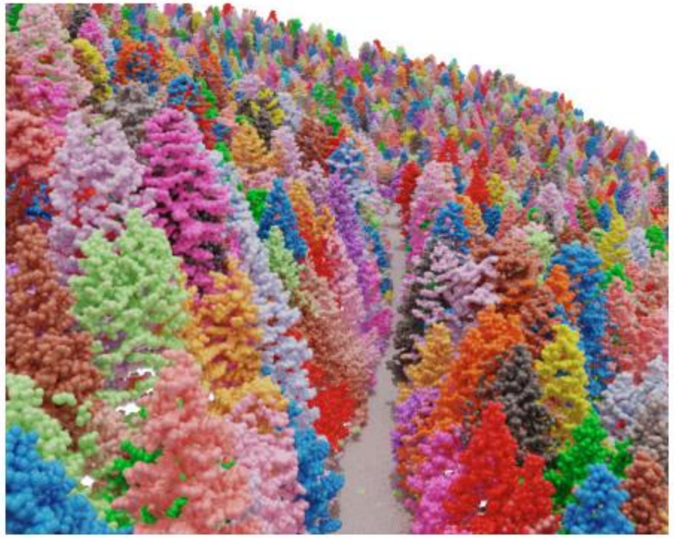


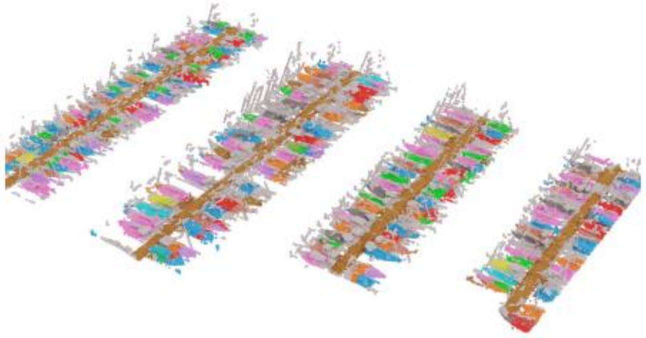
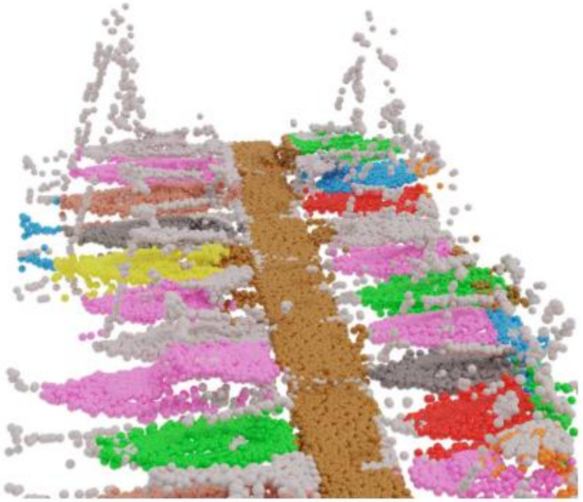




Instance segmenation results

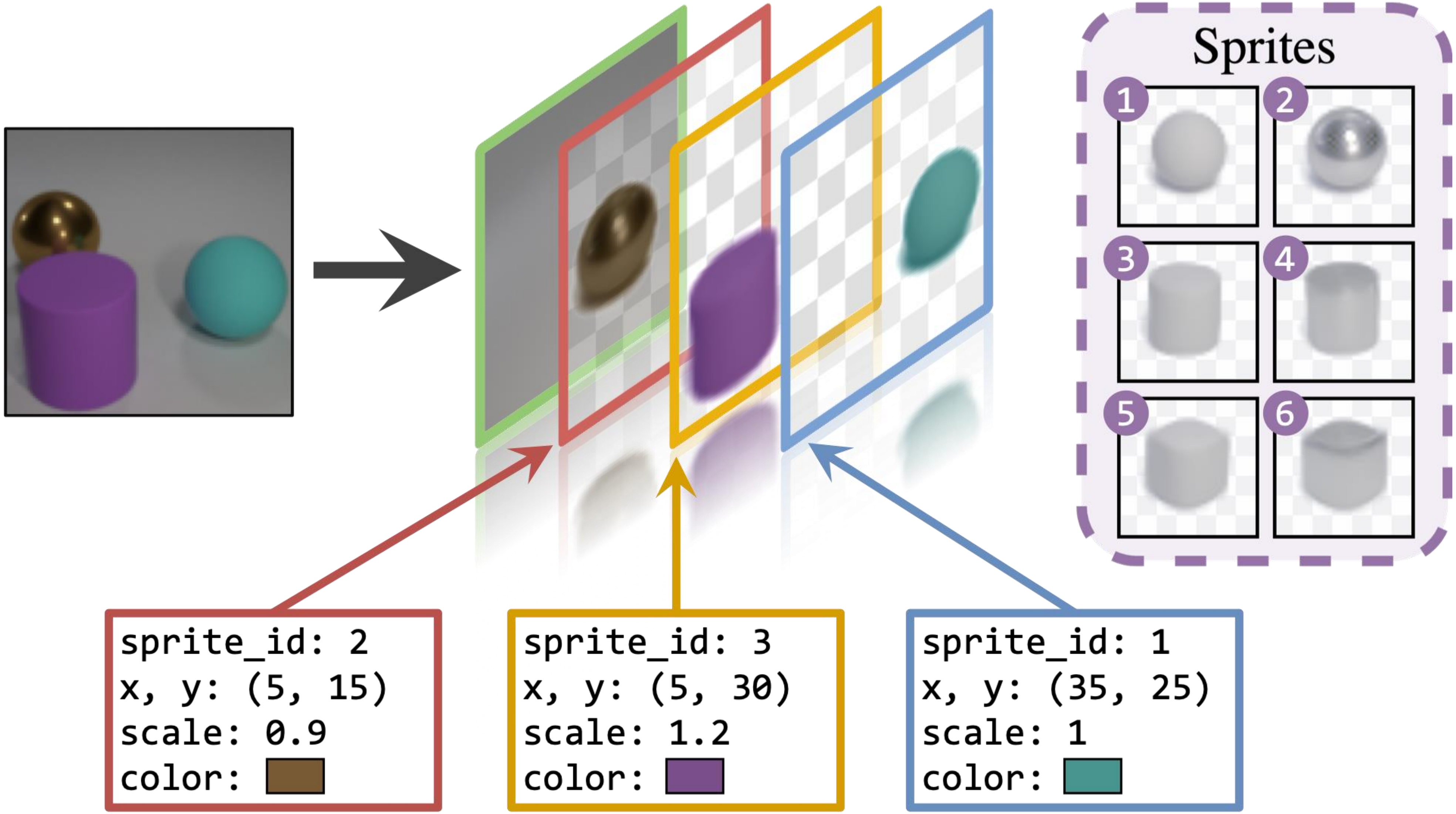




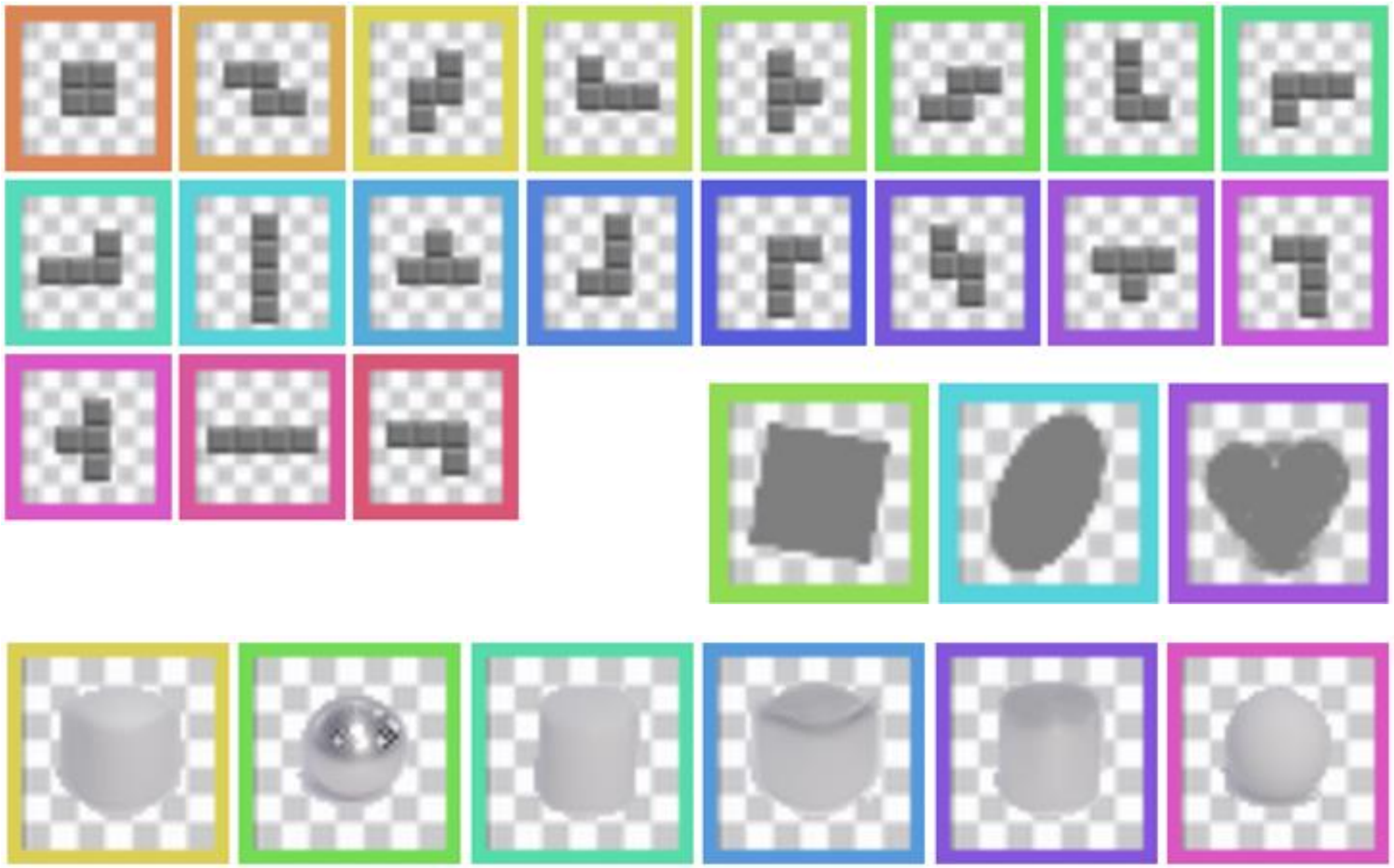
Structured generation for image analysis

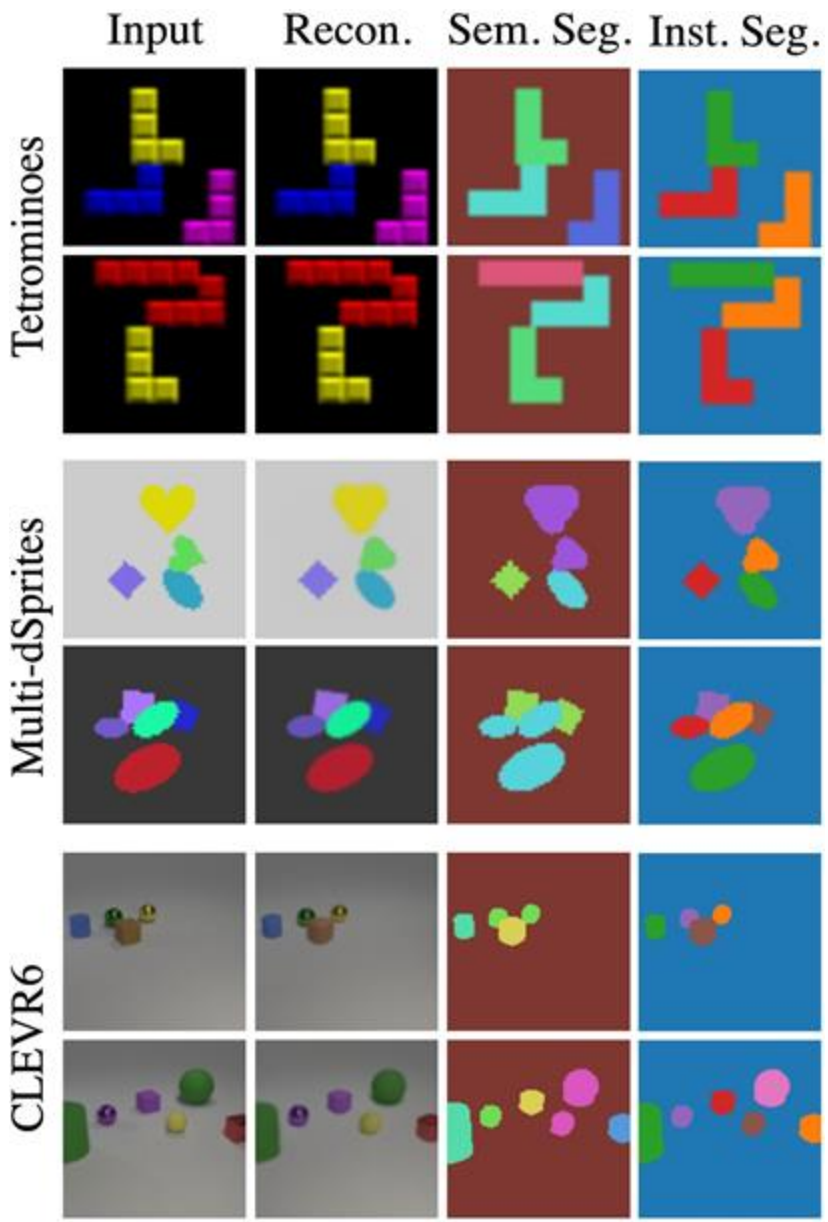
Unsupervised Layered Image Decomposition into Object Prototypes, T. Monnier, E. Vincent, J. Ponce, M. Aubry
ICCV 2021



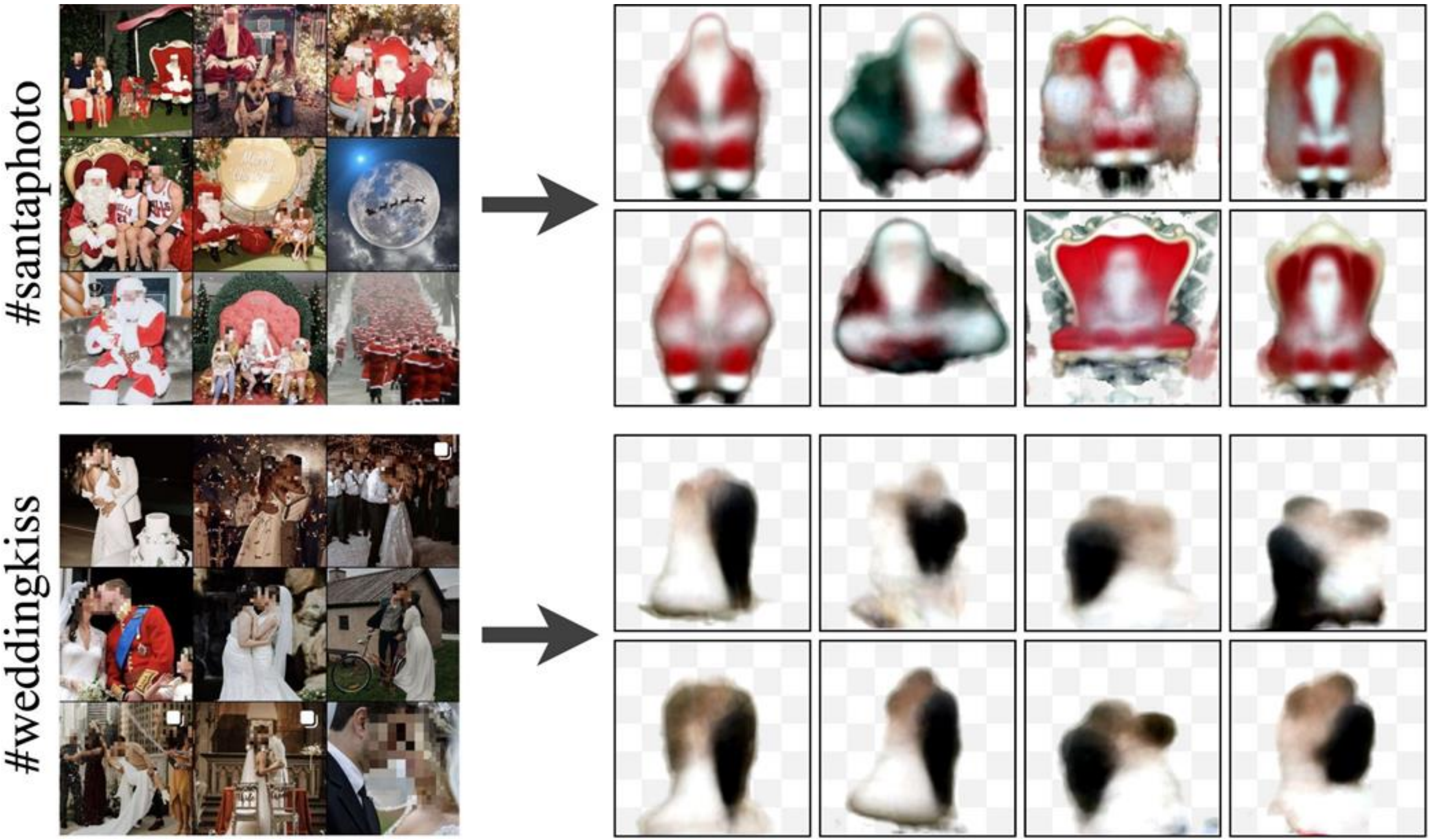
Multi-object discovery results

Discovered sprites





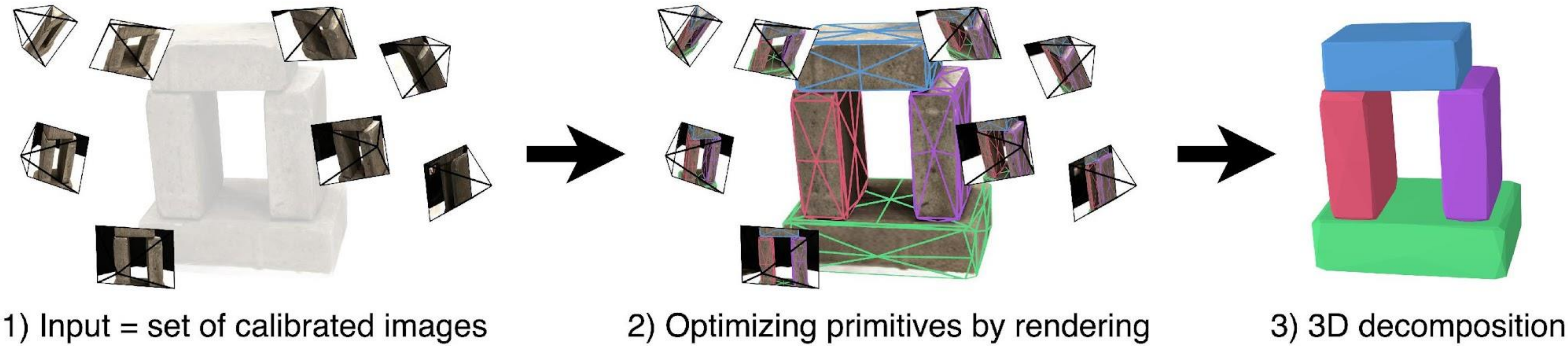
Object discovery on Instagram



Text lines, HTR and paleography

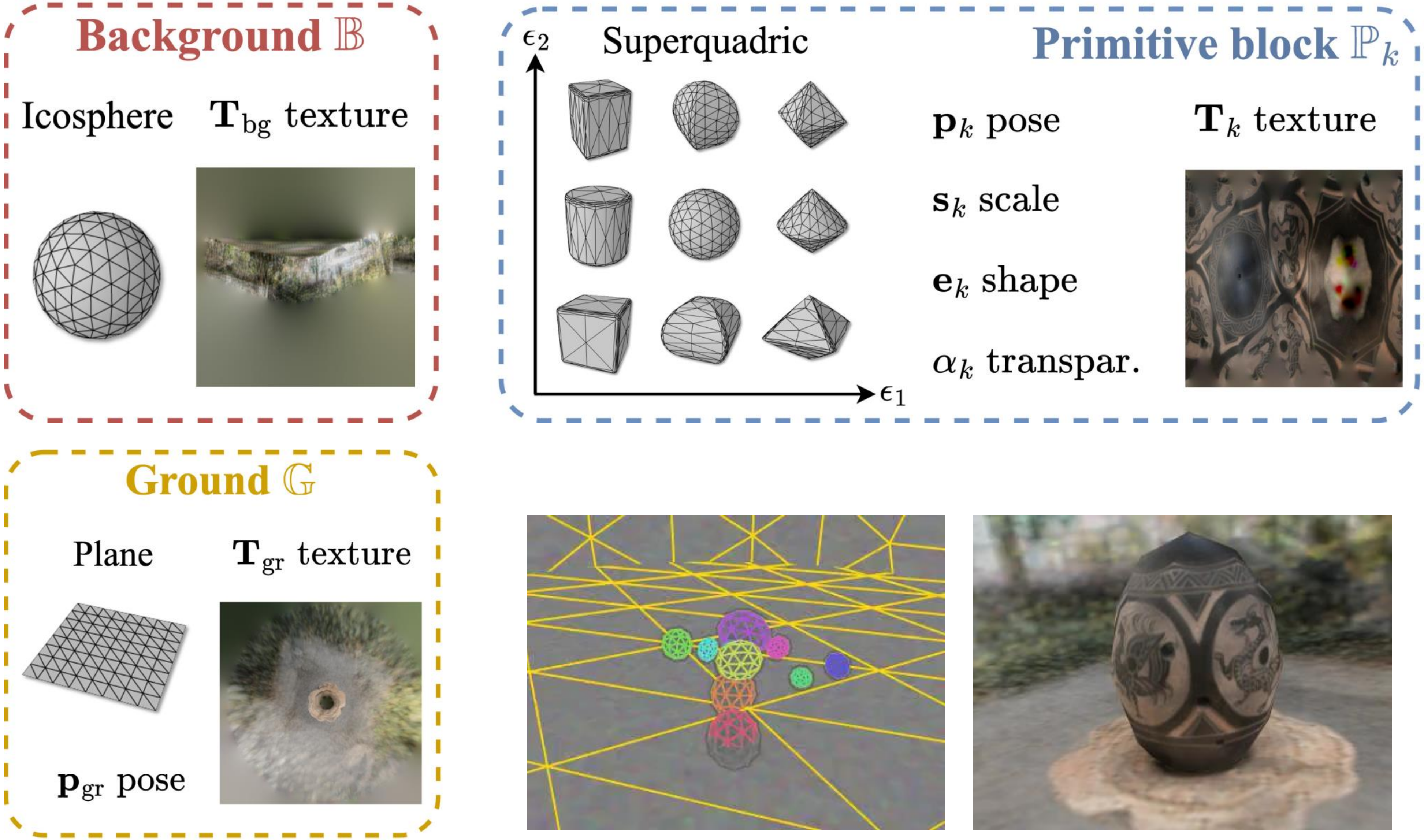
The Learnable Typewriter A Generative Approach to Text Line Analysis
Y. Siglidis, N. Gonthier, J. Gaubil, T. Monnier, M. Aubry, ICDAR 2024 (IAPR best paper award)

Differentiable Blocks World



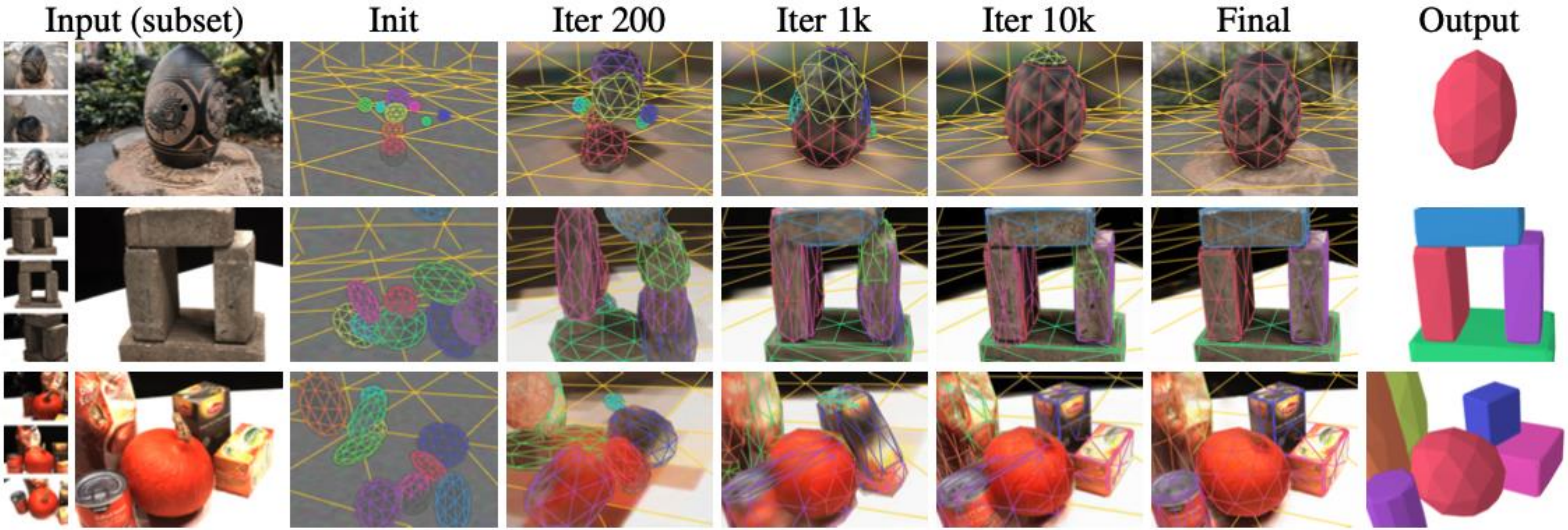
Differentiable Blocks World: Qualitative 3D Decomposition by Rendering Primitives
T. Monnier, J. Austin, A. Kanazawa, A. Efros, M. Aubry NeurIPS 2023

Approach

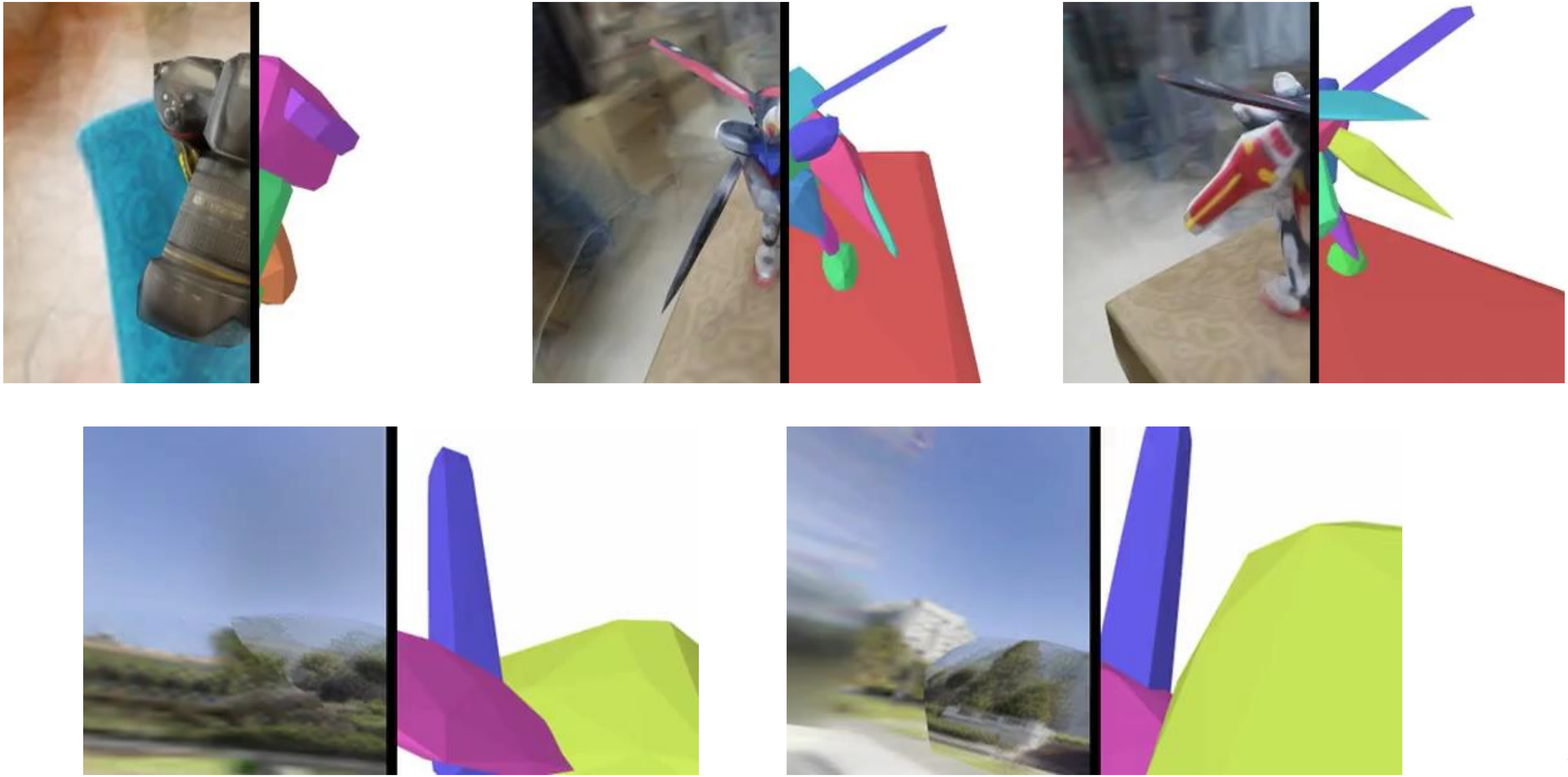


Background $\mathbb{B}$
Icosphere
$\mathbf{T}_{\text{bg}}$ texture
ϵ_2
Superquadric
ϵ_1
Primitive block $\mathbb{P}_k$
$\mathbf{p}_k$ pose
$\mathbf{s}_k$ scale
$\mathbf{e}_k$ shape
α_k transpar.
$\mathbf{T}_k$ texture
Ground $\mathbb{G}$
Plane
$\mathbf{T}_{\text{gr}}$ texture
$\mathbf{p}_{\text{gr}}$ pose

Optimization process

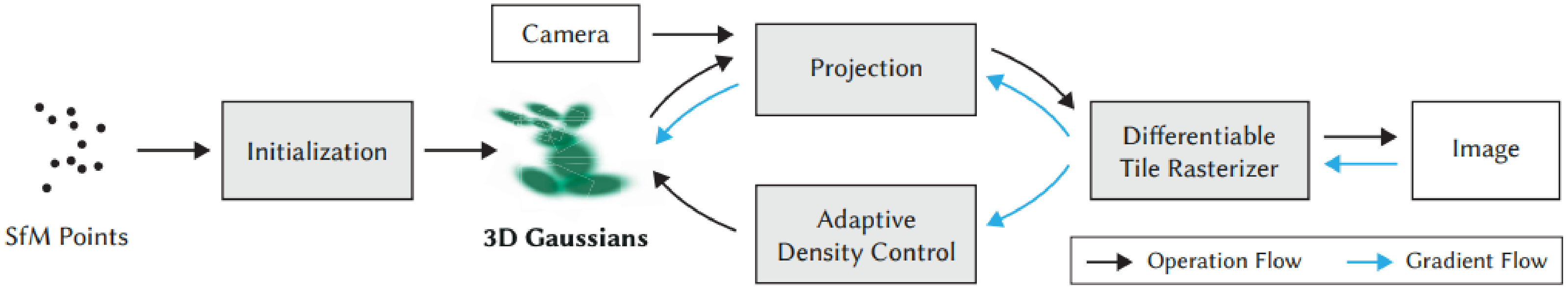


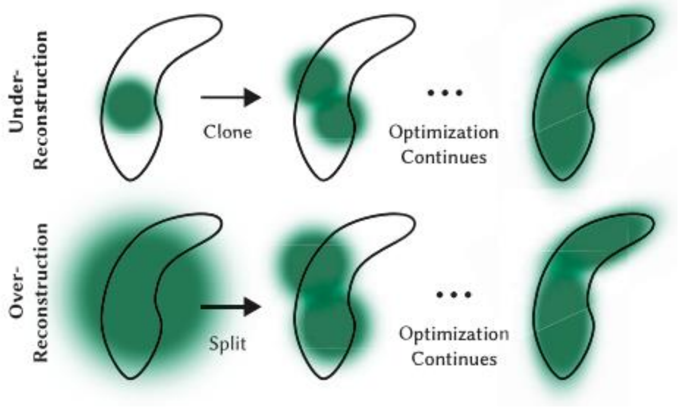
Qualitative results

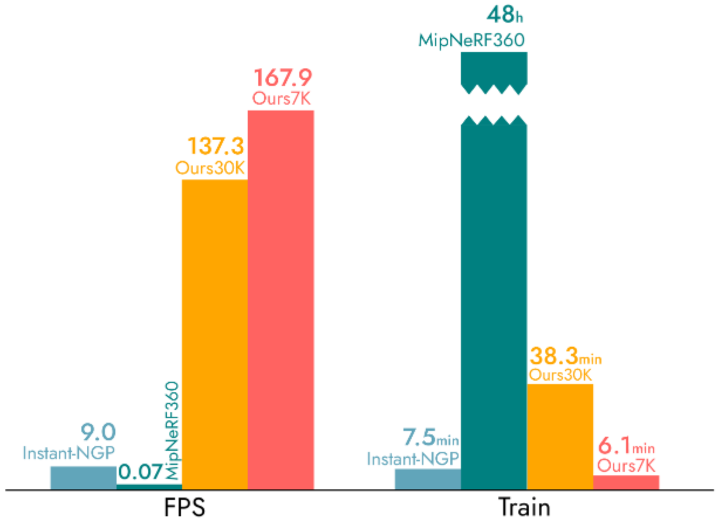


Applications

Gaussian splatting (Kerbl et al. 2023)







Outline: Deep learning and 3D data

Important milestones:

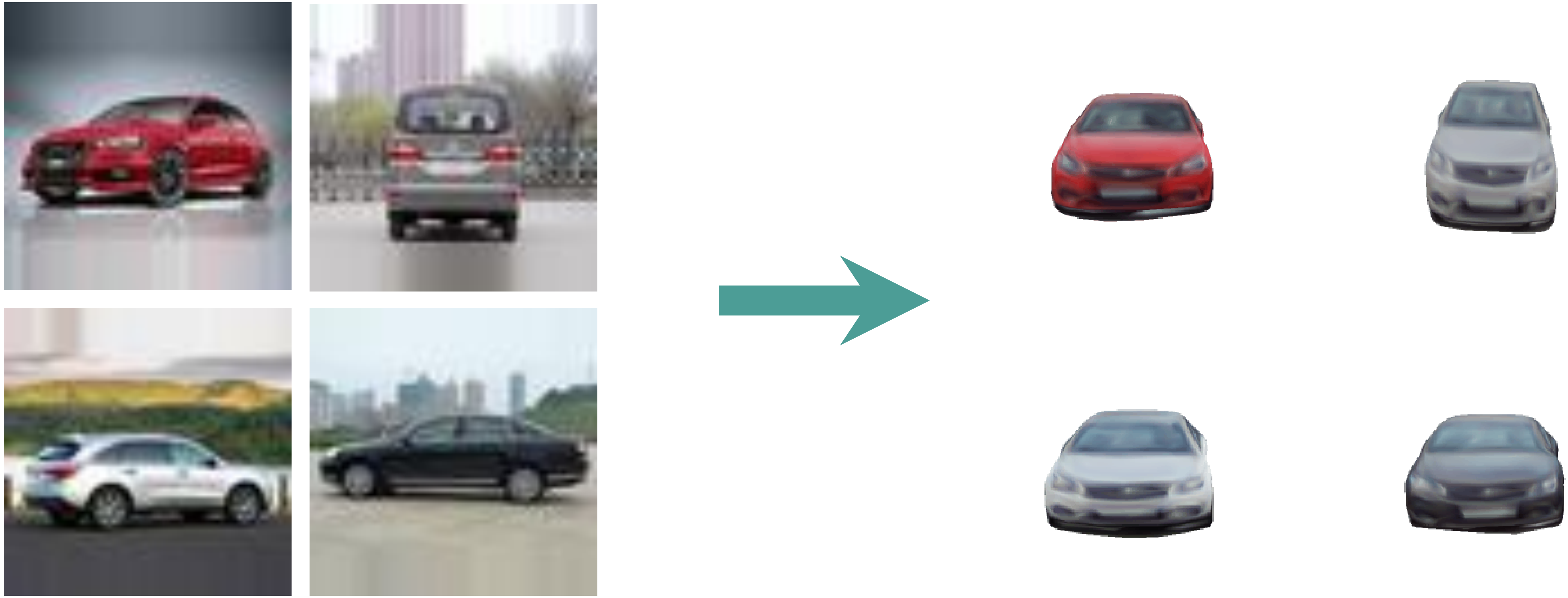
1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

Recent works I am excited about:

4. Structured generation
5. **Unsupervised single view reconstruction**

Learning with synthetic data

Goal → learn w/o supervision to reconstruct 3D objects from single views



Share With Thy Neighbors: Single-View Reconstruction by Cross-Instance Consistency
T. Monnier, M. Fisher, A. Efros, M. Aubry ECCV 2022

Single-View Reconstruction (SVR)

Method	Supervision	Synthetic data	Real data	Output
[6, 12, 30, 45]★	3D	ShapeNet	✗	3D
[26, 52]★	MV, C, S	ShapeNet	✗	3D
[5, 28, 36, 43]★	MV, C, S	ShapeNet	✗	3D, T
[57]	MV, C, S	✗	Bird, Car, Horse	3D, T
[20, 41]★	MV, S	ShapeNet	✗	3D, C
[23, 43, 44]★	CK, S	✗	Pascal	3D
[5, 22]	CK, S, P(†)	✗	Bird, Car, Plane	3D, T
[16]	CK, P(†)	ShapeNet	Bird, Car	3D, T
[10]	S, P(♢, †)	✗	Bird, Car, Moto, Shoe	3D, T, C
[42]	S, P(♢, †)	✗	Animal, Car, Plane	3D, T, C
[27]	S, P(↔, †)	✗	Animal, Car, Moto	3D, T, C
[48]	S, P(‡)	✗	Vase	3D, T, C
[49]	P(⊠, ⋖, †)	✗	Face	D, T, C
[15]	P(⊠, ∅)	Toy ShapeNet	✗	3D, C
Ours	None	ShapeNet	Animal, Car, Moto	3D, T, C

Legend: Multi-Views, Camera, Camera estimate or Keypoints, Silhouette, Prior (♢ template shape, † symmetry, ‡ solid of revolution, ↔ semantic consistency, ⊠ no/limited background, ⋖ frontal view, ∅ no texture), Depth, Texture.

Current trend → remove supervision from SVR pipelines

Why? → to learn 3D from raw 2D images « for free »

Our work

→ w/o hypotheses of prior works

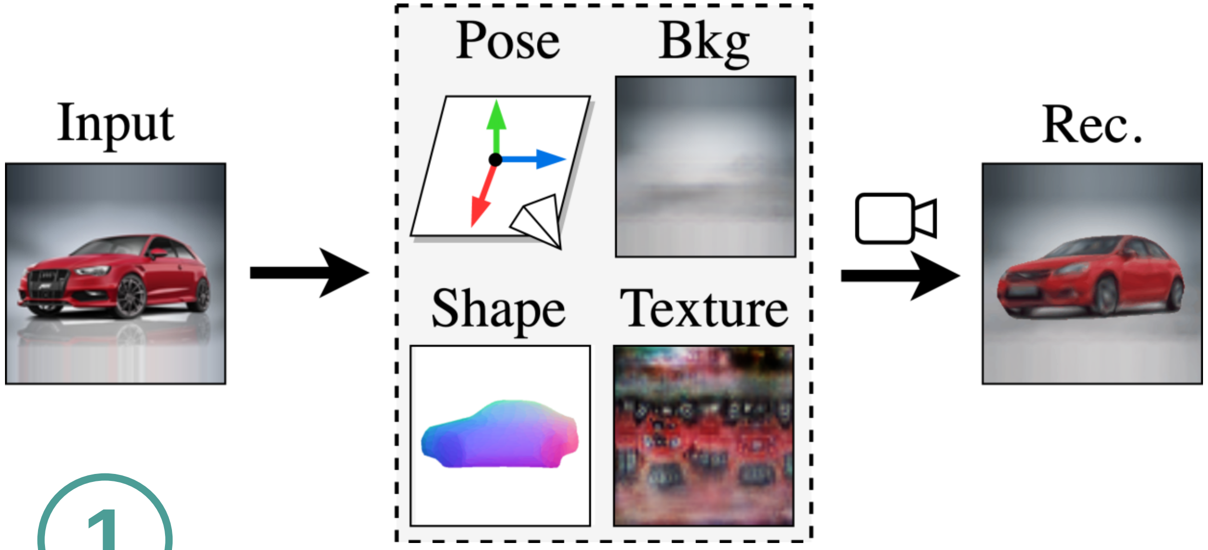
→ diverse shapes (ShapeNet)

→ high-quality results on real images

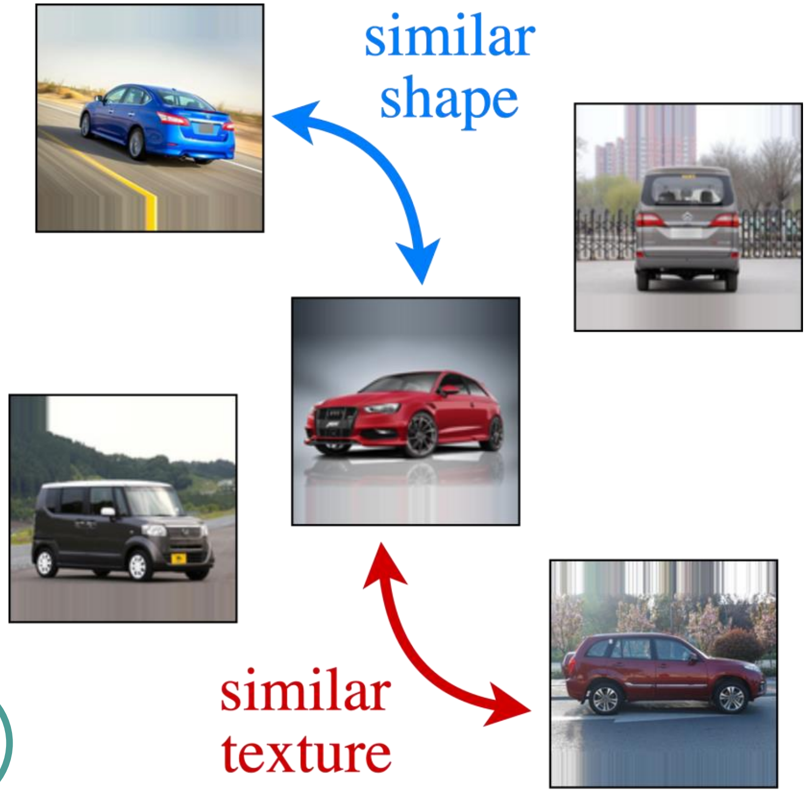
Disclaimer

→ we still use categorical images

Our approach



1



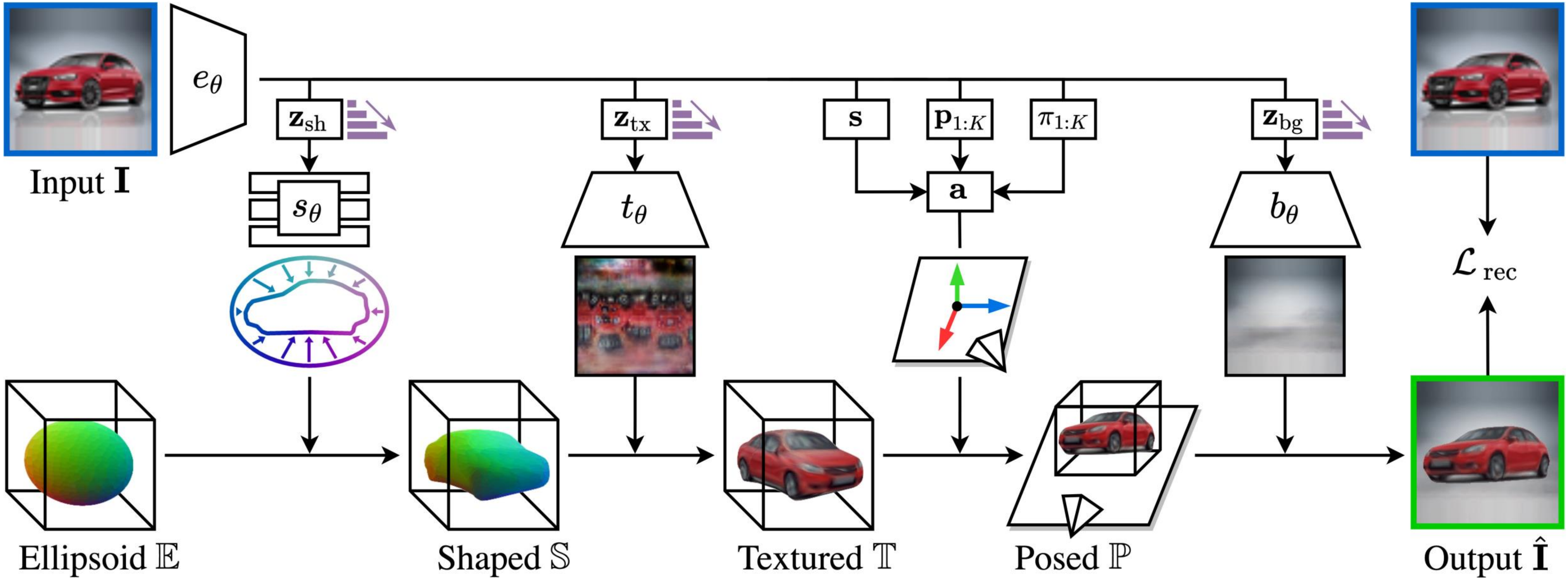
2

Structured autoencoding into explicit factors: shape, texture, pose, background

(analysis-by-synthesis fashion)

We leverage the **consistency across different instances** to remove supervision & priors

Structured autoencoding



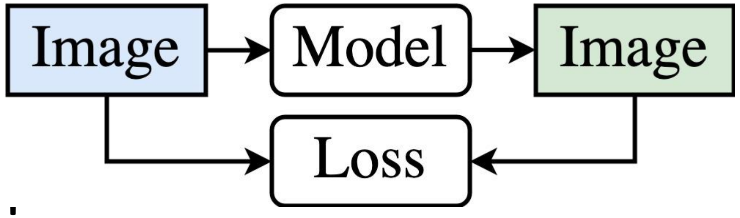
Shape deformation

Texturing

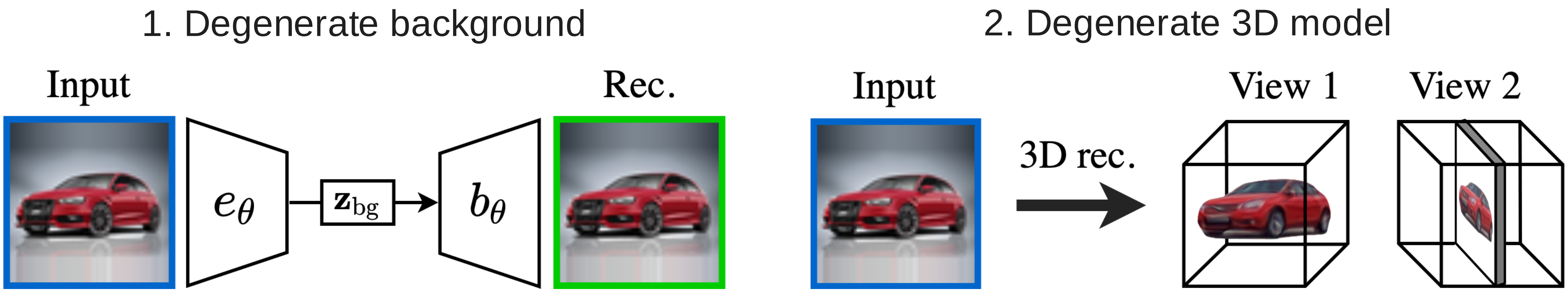
Affine transformation

Rendering with background

Structured autoencoding - issue



Task is highly unconstrained w/o supervision & priors：

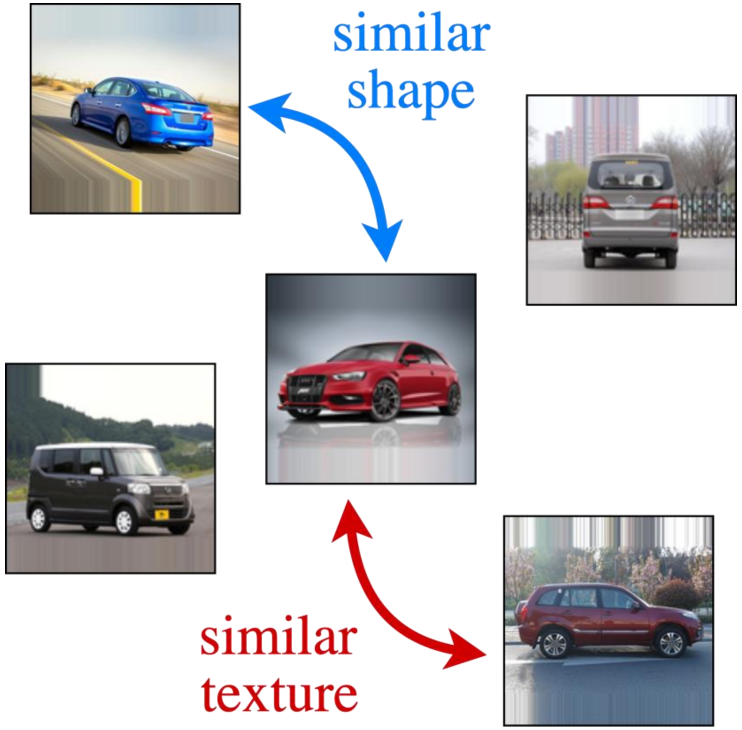


Two data-driven approaches leveraging **cross-instance consistency**:

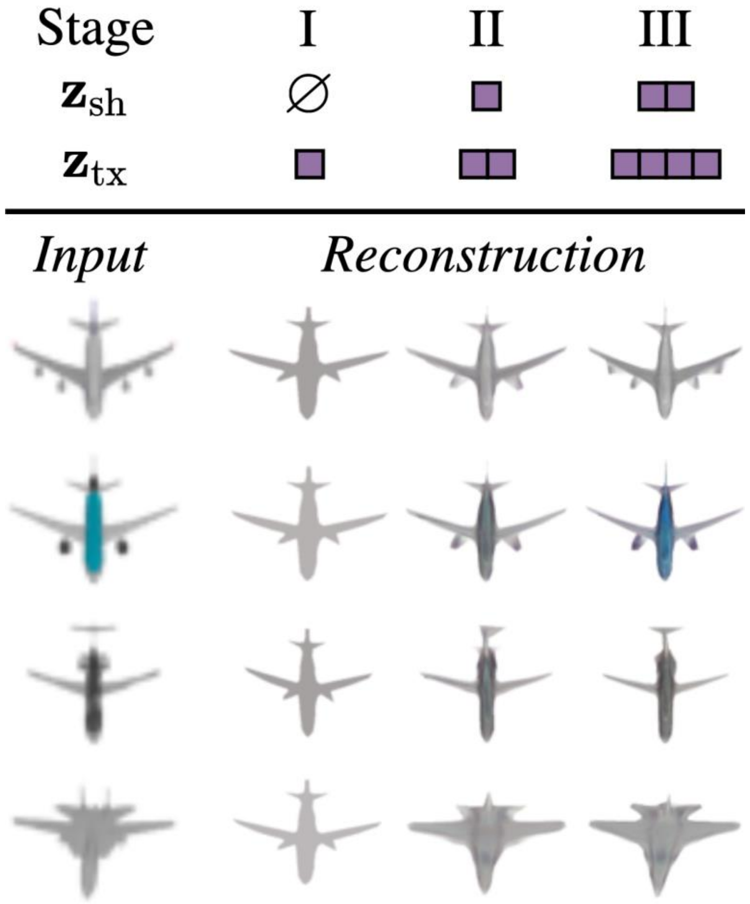
→ progressive conditioning (training procedure)

→ Neighbor reconstruction (training loss)

Progressive conditioning (PC)



Cross-instance consistency
→ instances with similar shapes and textures exist!



Progressive conditioning

→ gradually specialize from category to instances

→ progressively allow more variability by increasing the latent space dimension

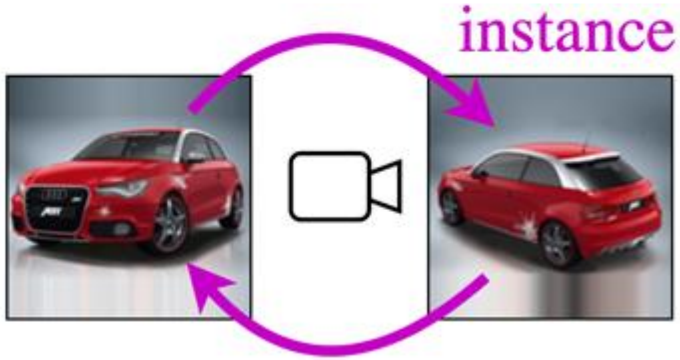
→ curriculum learning spirit

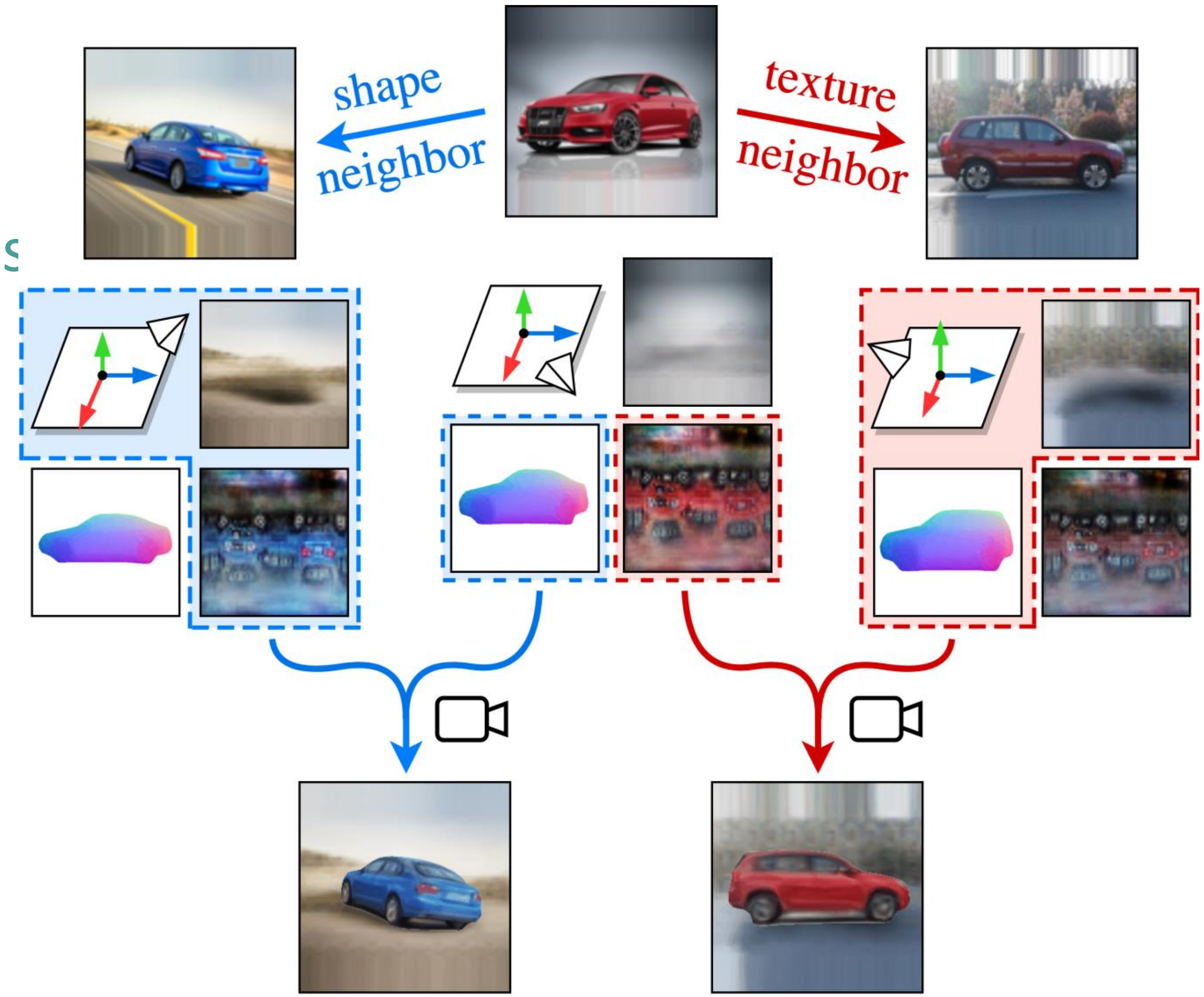
Neighbor reconstruction

Neighbor reconstruction loss
→ force consistency among instances w/ similar shapes & textures

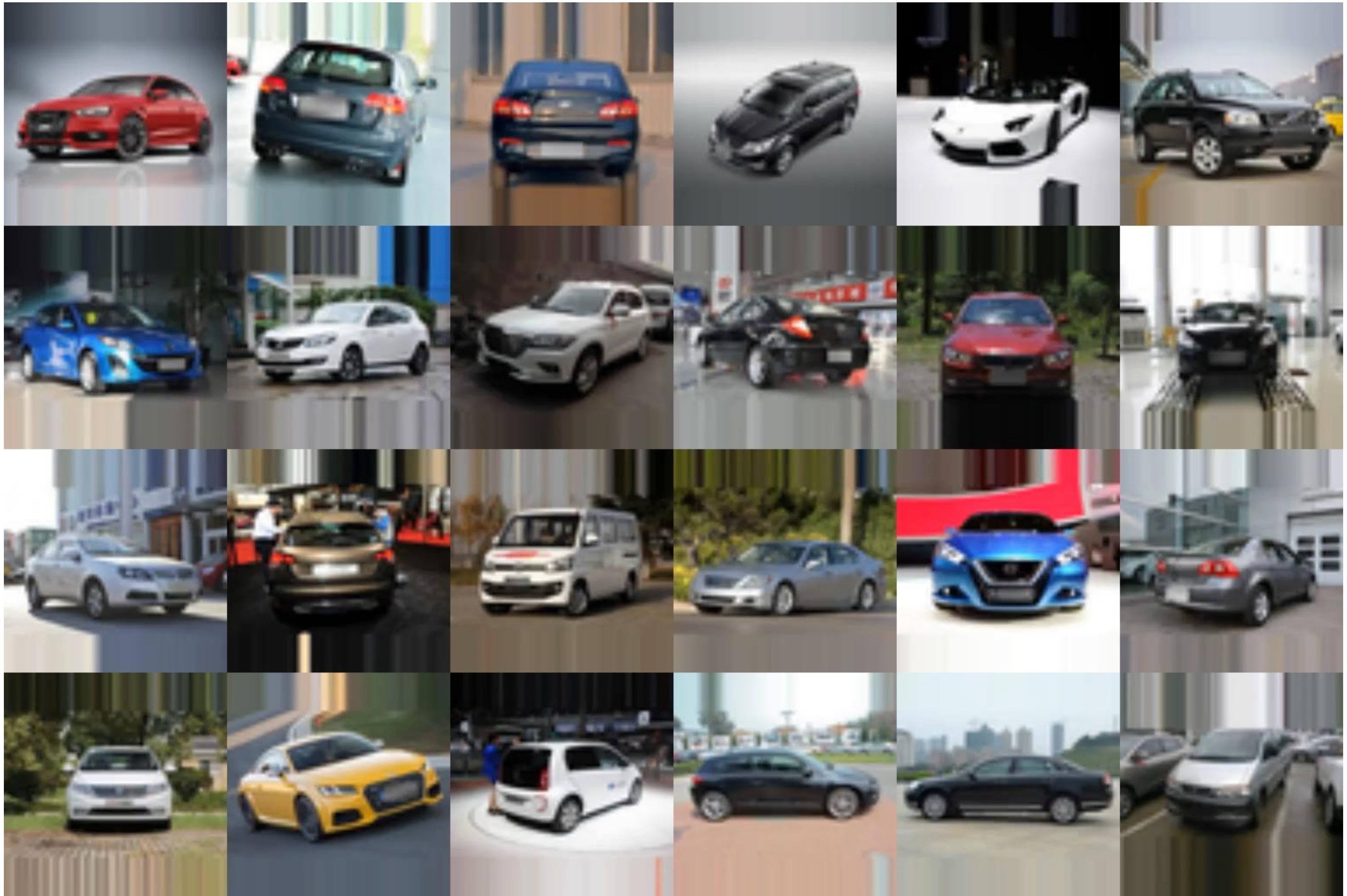
→ swapping characteritics should give similar reconstructions

→ like a multi-view supervision w/o having access to multi-views





Results - CompCars



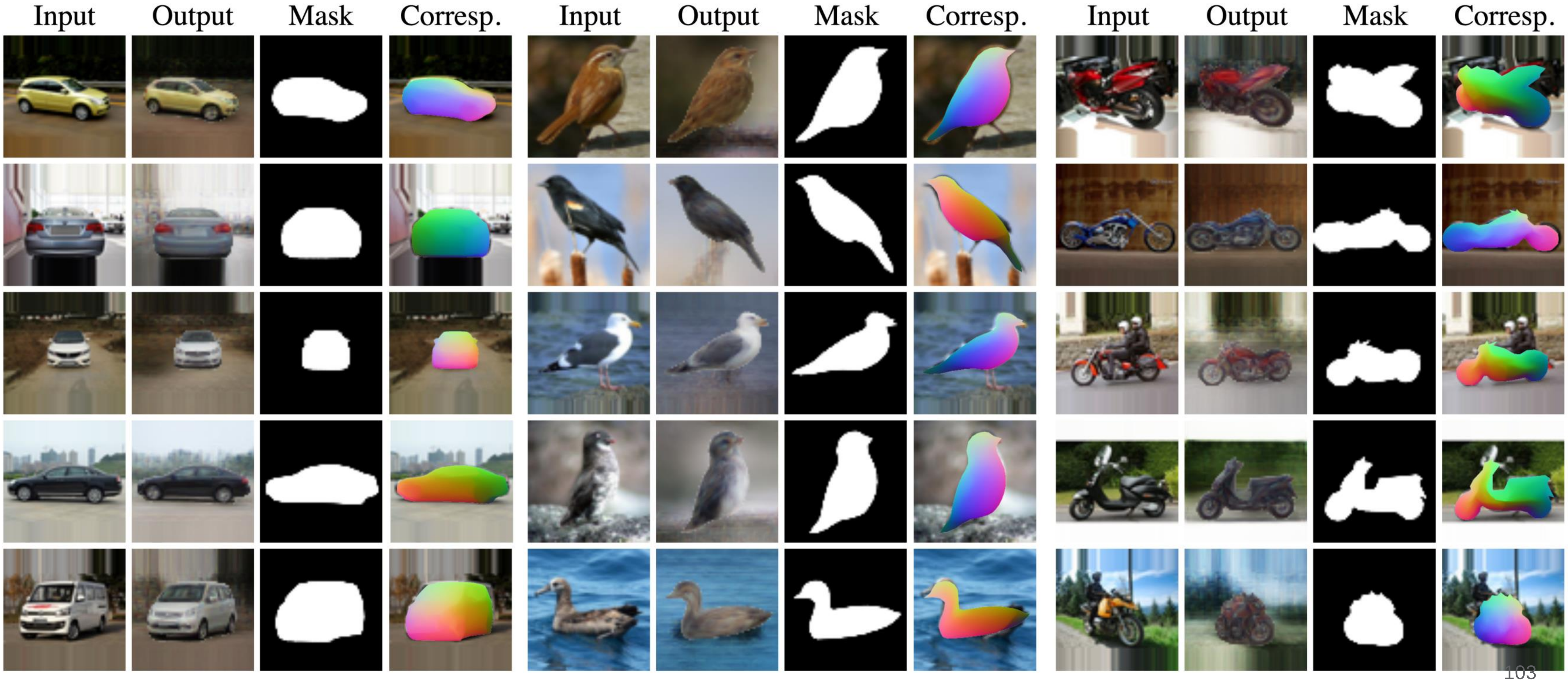
Ablation study

Input	Full model	w/o PC	w/o $\mathcal{L}_{\text{swap}}$

Results - ShapeNet

Results - Motorbikes

Free by-products – silhouettes & correspondences



Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

Learning from synthetic data

- Very appealing:
  - Annotations (almost) free
  - Can include things that are very hard to annotate (e.g. illumination, dense labels)
  - Can simulate rare situation (e.g. accidents)

- Challenge: domain gap - will the model trained on synthetic data work as well on real data?
- Strategies:
  - Realistic data
  - Domain adaptation
  - Domain randomization
  - Other

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

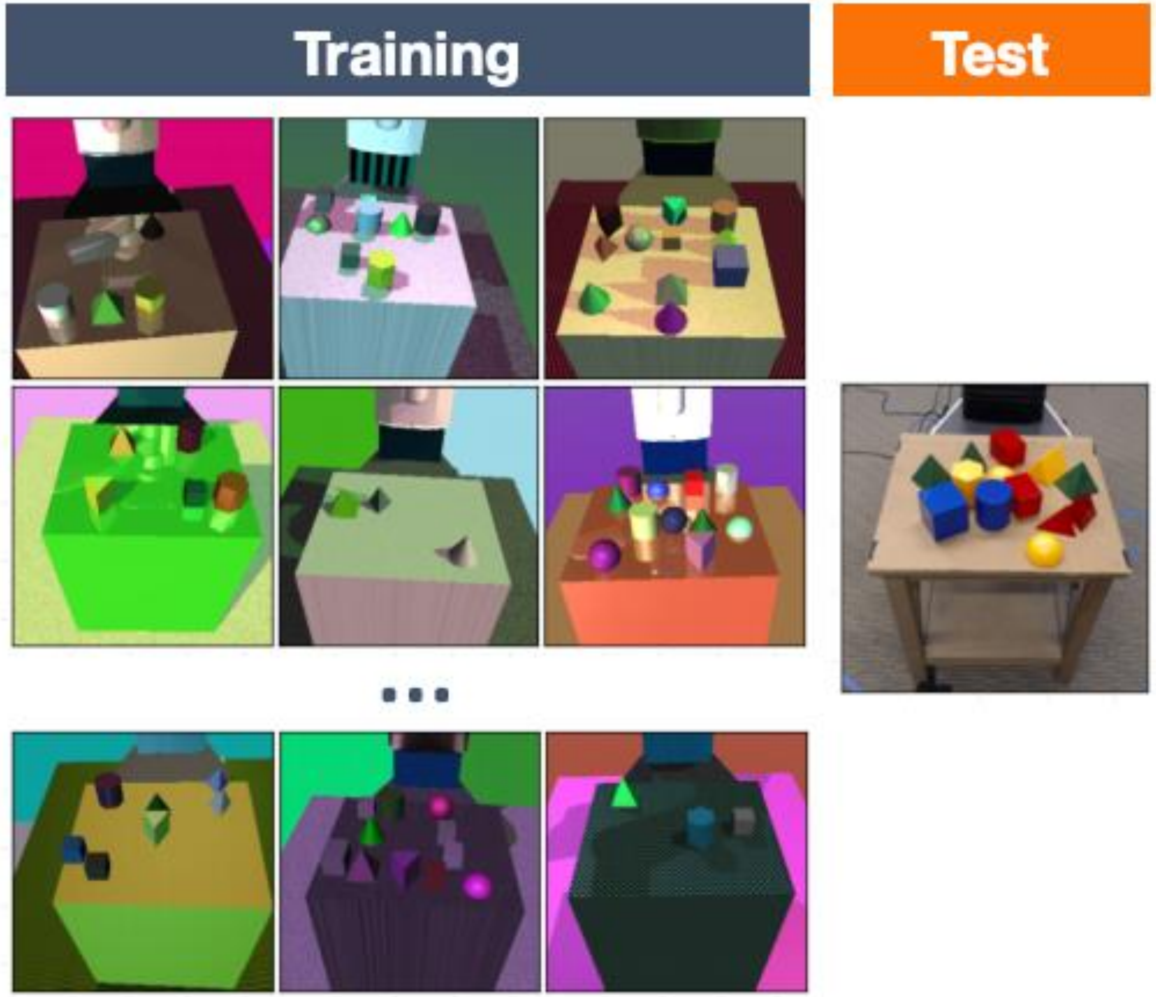
Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

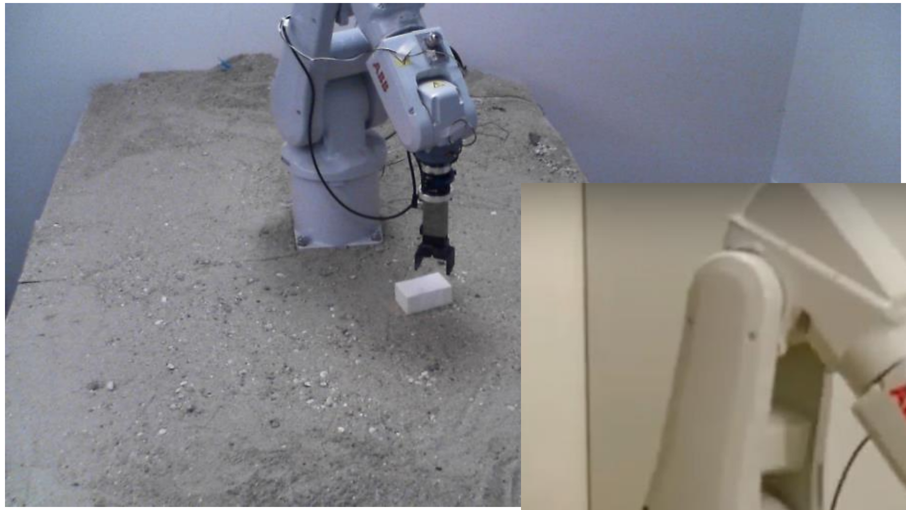
- **Domain randomization**
- Realistic data
- Domain adaptation

Domain randomization: predict 2D position



Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P.
Domain randomization for transferring deep neural networks from simulation to the real world
IROS 2017

Domain randomization: Learning relative position



Virtual Training for a Real Application: Accurate Object-Robot Relative Localization without Calibration
V. Loing, R. Marlet, M. Aubry, IJCV 2018

S. Zagoruyko, Y. Labbé, I. Kalevatykh, I. Laptev, J. Carpentier, M. Aubry and J. Sivic
RSS workshop 2019, ArXiv

Monte-Carlo Tree Search for Efficient Visually Guided Rearrangement Planning

Vision part extending

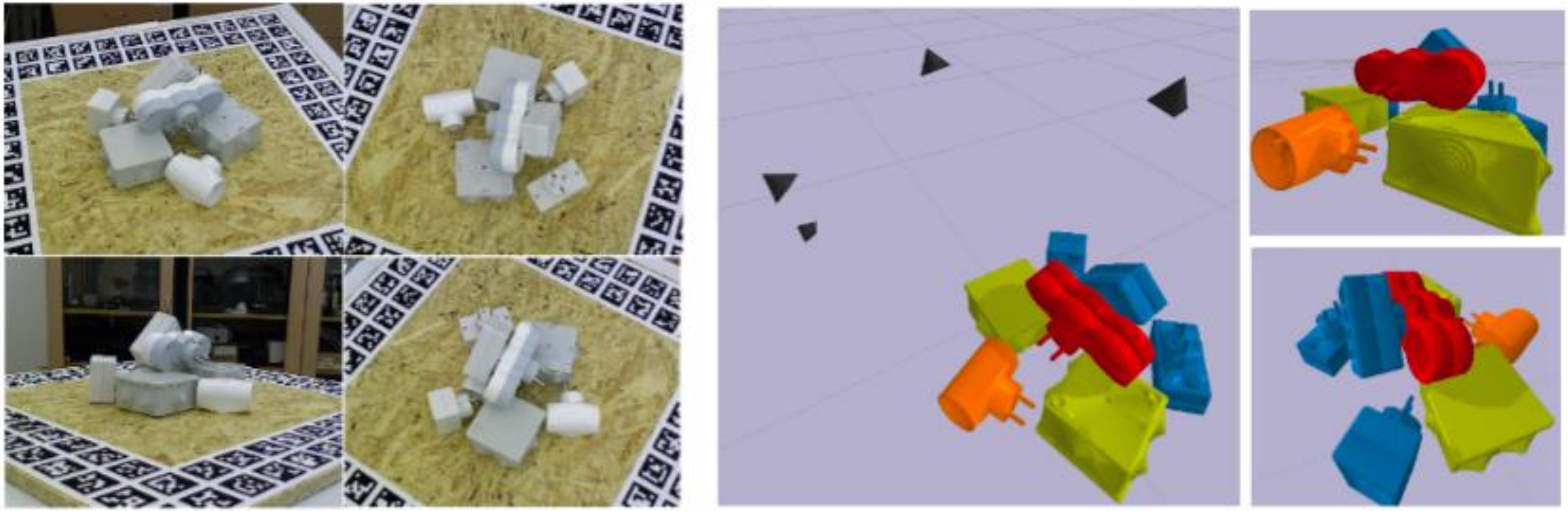
Virtual training for a real application: Accurate object-robot relative localization without calibration
V. Loing, R. Marlet, Mathieu AUbry
IJCV 2018

CosyPose:
Multi-views, multi-object

- Single view similar to deepIM (see later) with randomized training data

Multi-view multi-object 6D pose estimation via robust scene consistency optimization
Y. Labbé, J. Carpentier, M. Aubry, J.Sivic, ECCV 2020

CosyPose:
Multi-views, multi-object



Multi-view multi-object 6D pose estimation via robust scene consistency optimization
Y. Labbé, J. Carpentier, M. Aubry, J.Sivic, ECCV 2020

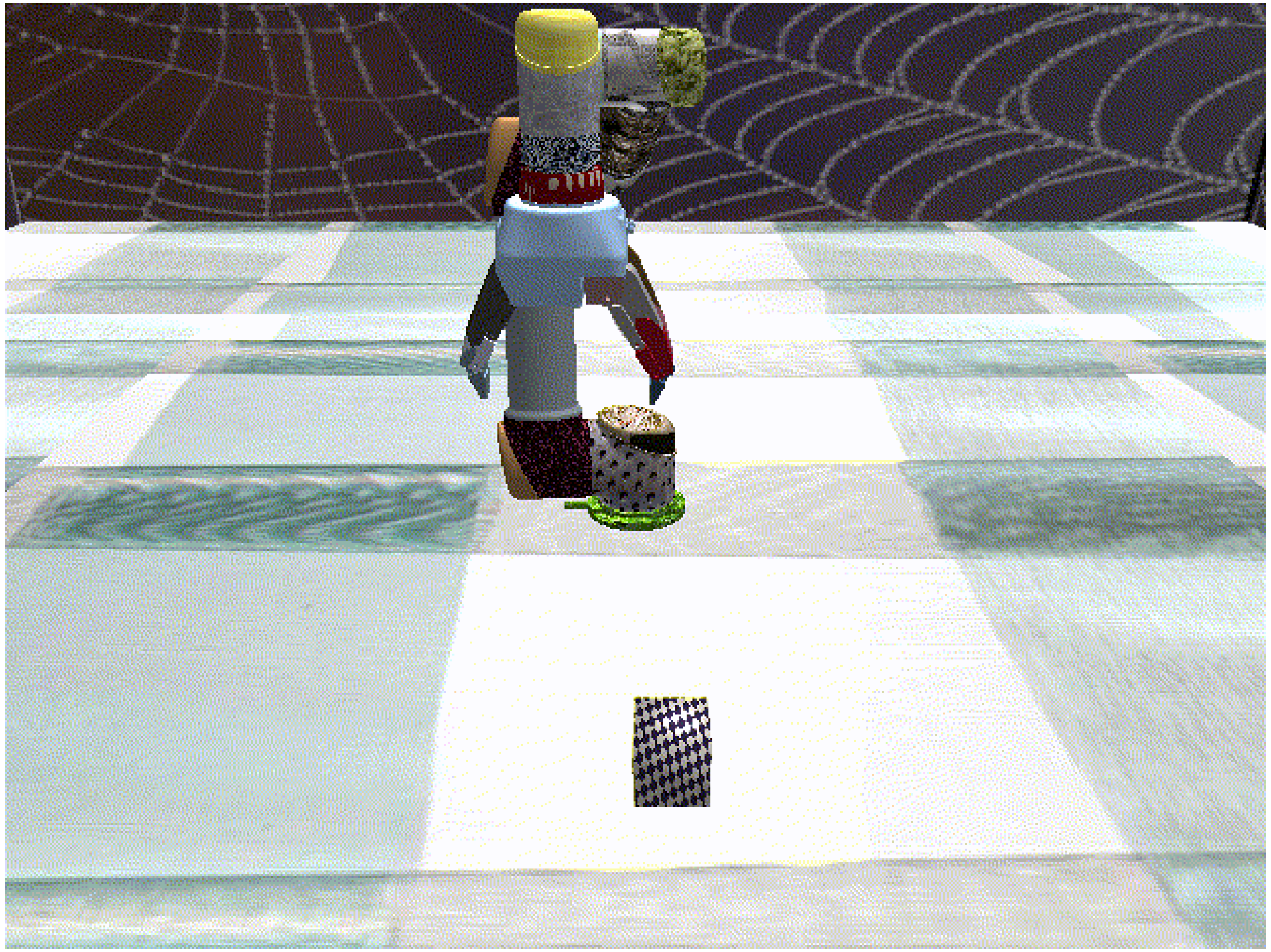
Single-view robot pose and joint angle estimation via render & compare

Extending the render and compare approach of
Multi-view multi-object 6D pose estimation via robust scene consistency optimization
Y. Labbé, J. Carpentier, M. Aubry, J.Sivic, ECCV 2020
to articulated objects

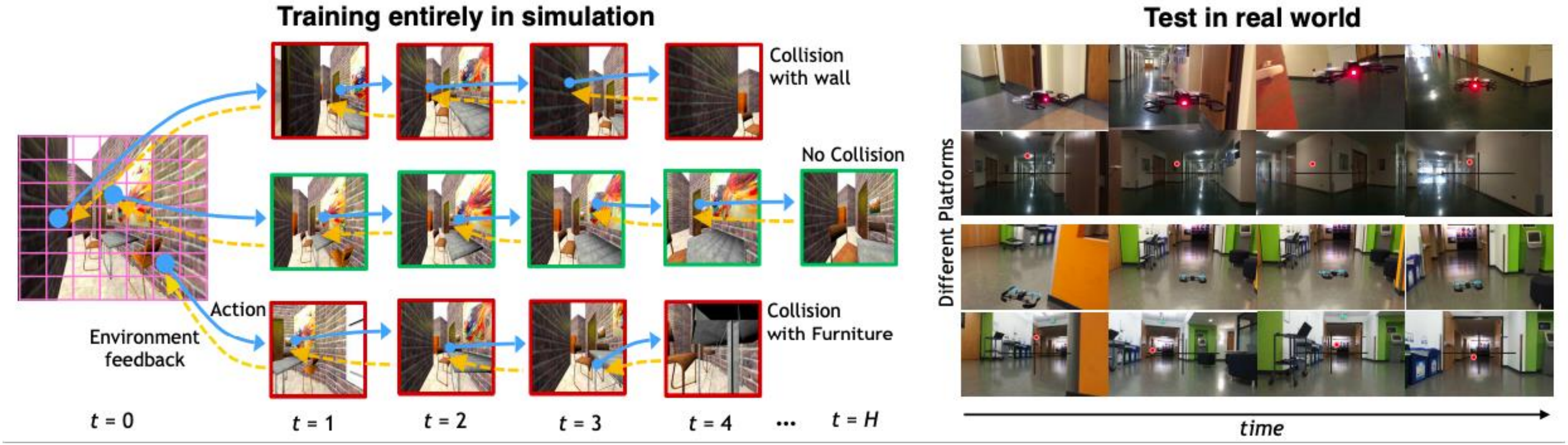
Domain randomization: Learning to act

Learning strategies:

- Imitation
- RL



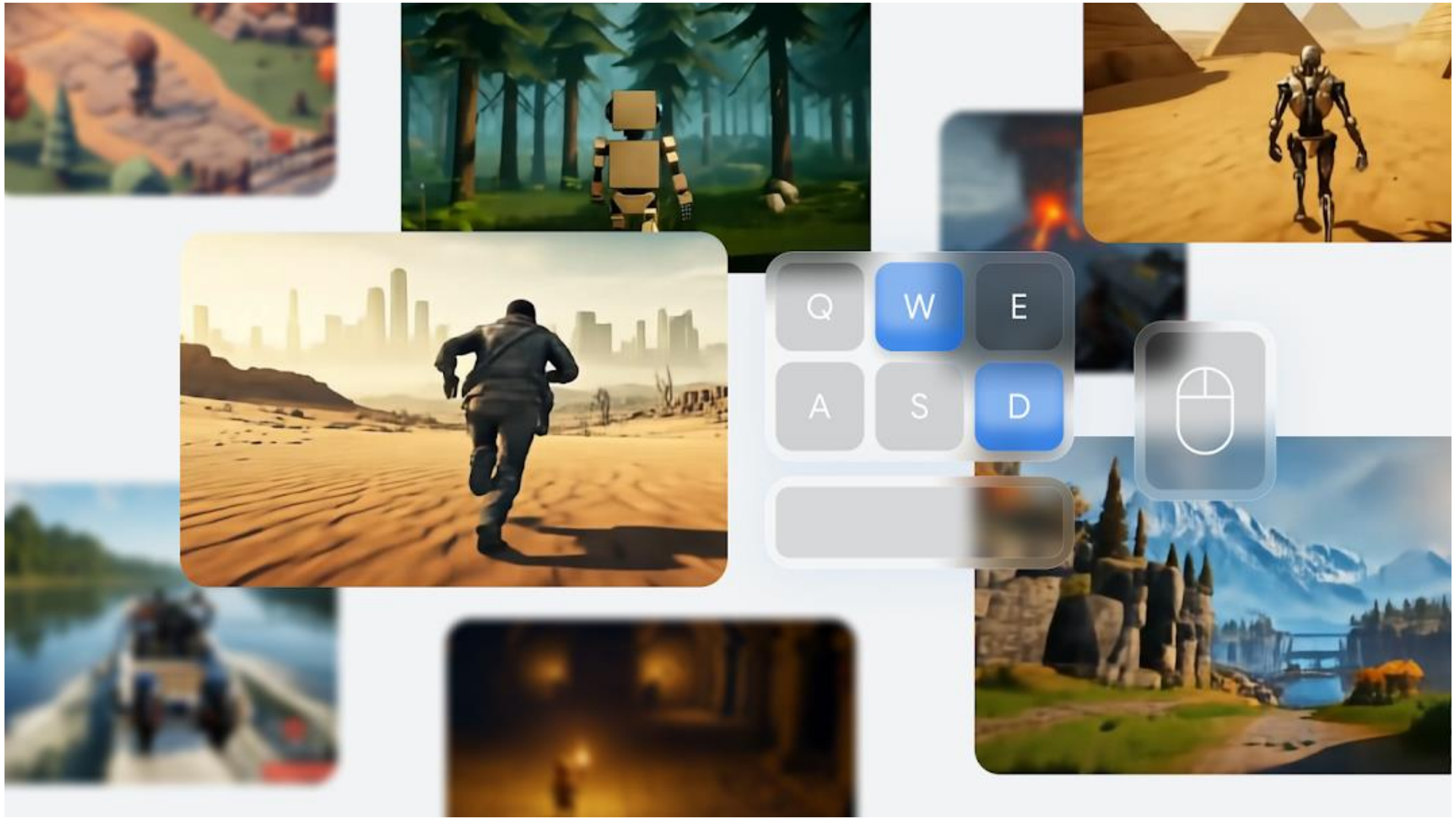
RL from synthetic data to real world



Sadeghi, F., & Levine, S. (2016). Cad2rl: Real single-image flight without a single real image.

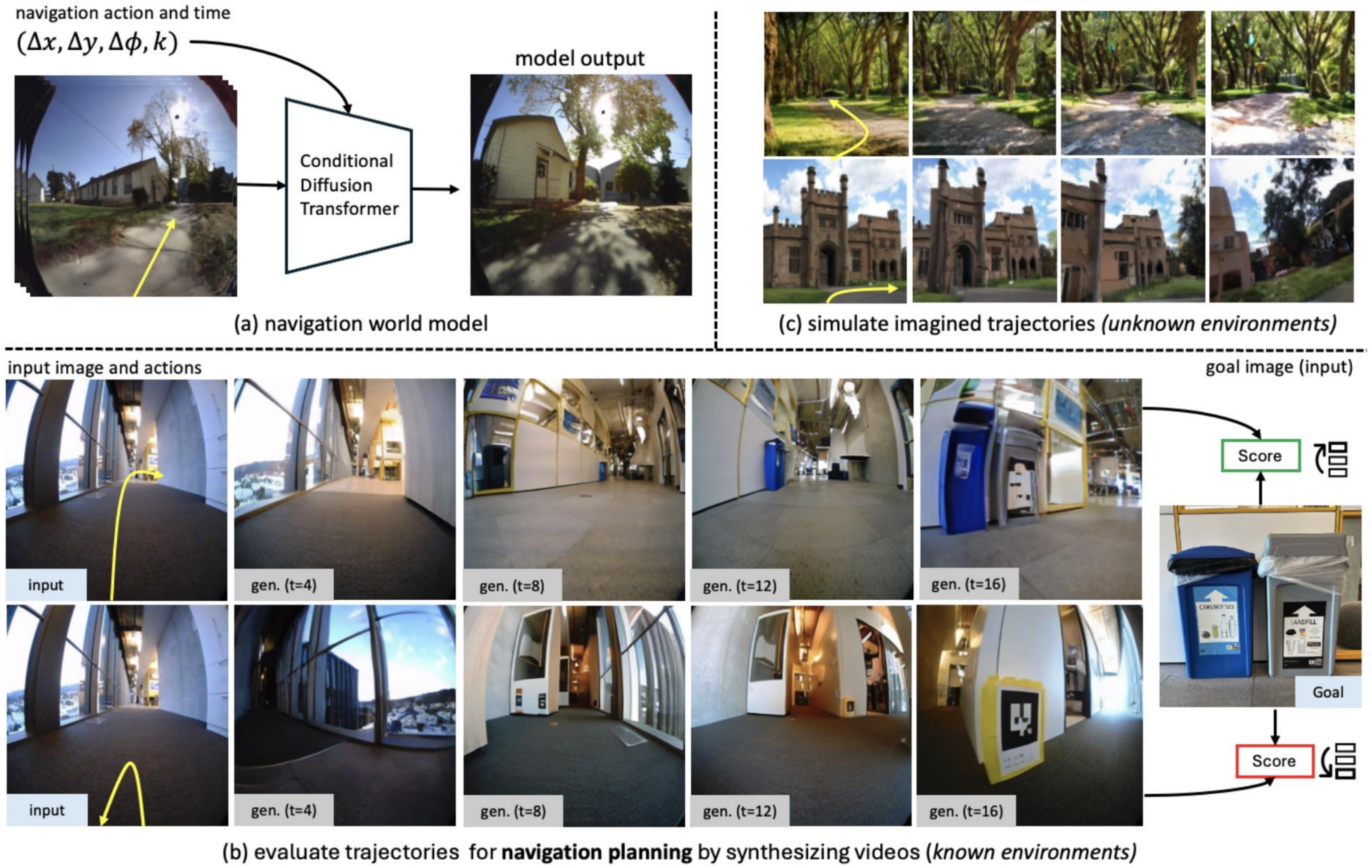
Genie 2

Realism and diversity, based on diffusion, aimed at training agents



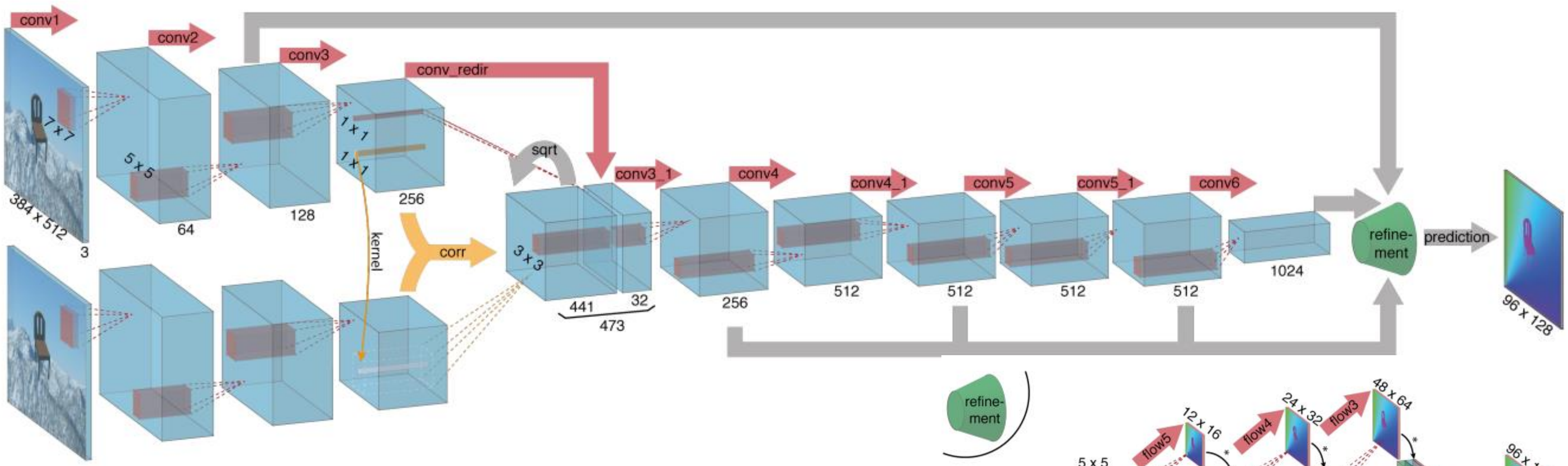
Last week, Google : https://deepmind.google/discover/blog/genie-2-a-large-scale-foundation-world-model/

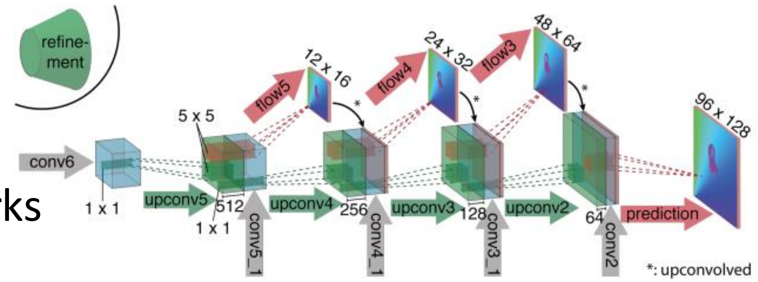
Navigation World Models



Bar, A., Zhou, G., Tran, D., Darrell, T., & LeCun, Y.
Navigation World Models. *arXiv last week, Meta*

Domain randomization: Optical flow





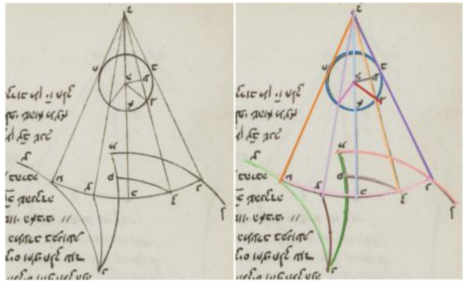
Flownet: Learning optical flow with convolutional networks
A Dosovitskiy et al., ICCV 2015

For historical data

- Illustration detection

docExtractor: An off-the-shelf historical document element extraction
T. Monnier, M. Aubry, *ICFHR 2020*

- Copy retrieval



Learning Co-segmentation by Segment Swapping for Retrieval and Discovery
X. Shen, A. Efros, A. Joulin, M. Aubry, *CVPR 2022 workshops*

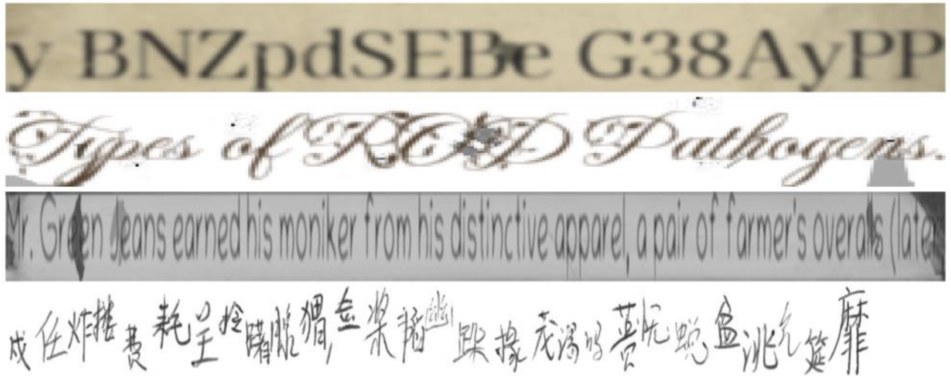
- Diagrams vectorization

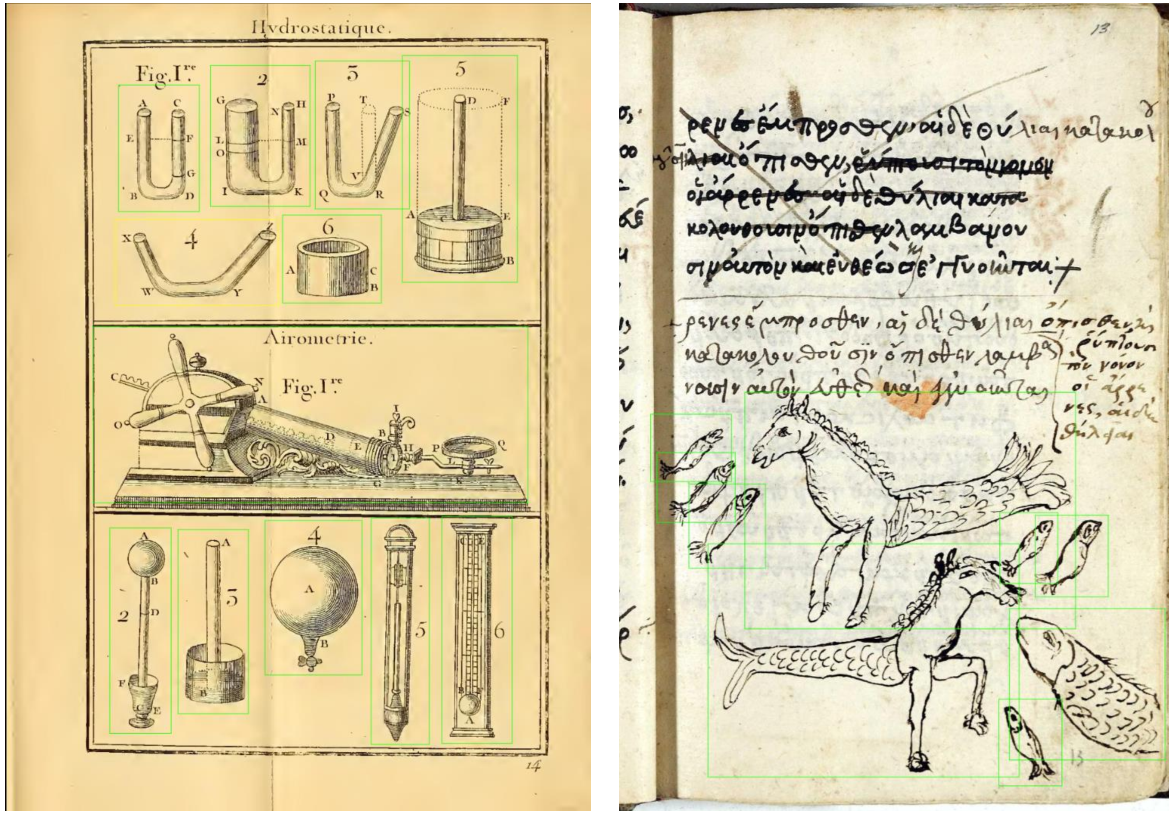
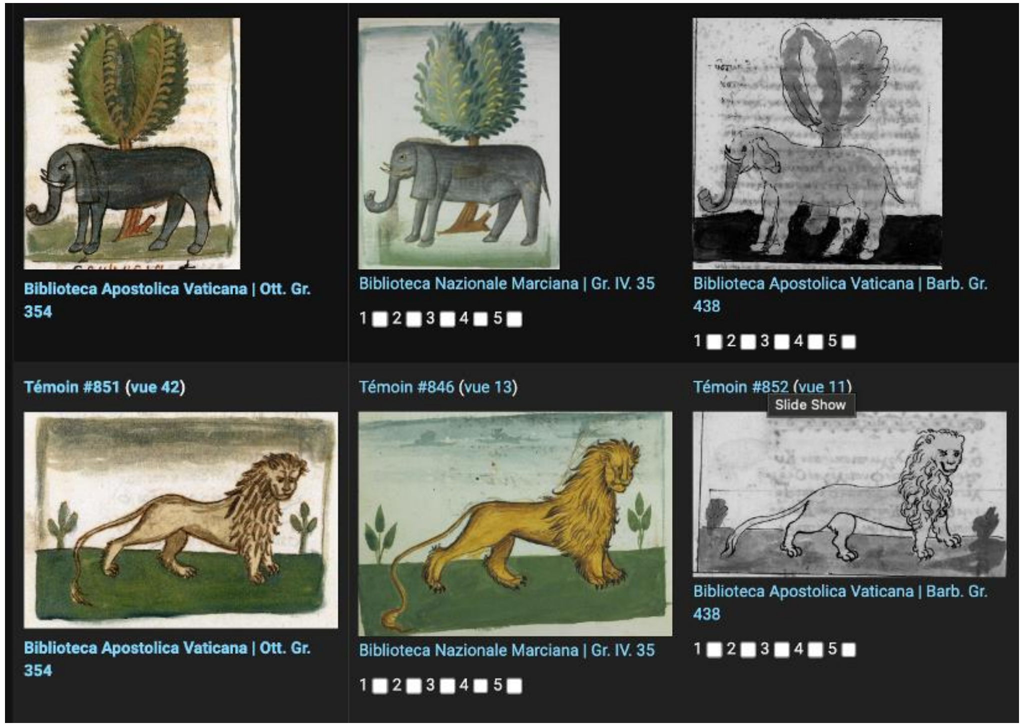
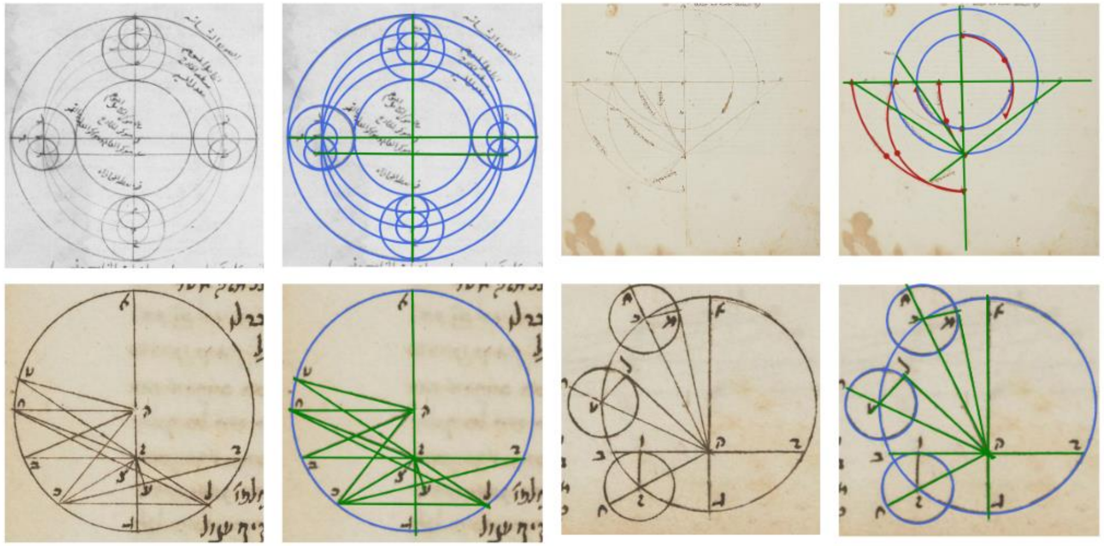
Historical Astronomical Diagrams Decomposition in Geometric Primitives
S. Kalleli, S. Trigg, S. Albouy, M. Husson, M Aubry, *ICDAR 2024*

- Text recognition (upcoming)

General Detection-based Text Line Recognition
R. Baena, S. Kalleli, M. Aubry, *NeurIPS 2024*

Synthetic data




Hydrostatique.
Airometrie.

Biblioteca Apostolica Vaticana | Ott. Gr. 354
Biblioteca Nazionale Marciana | Gr. IV. 35
Biblioteca Apostolica Vaticana | Barb. Gr. 438
Témoin #851 (vue 42)
Témoin #846 (vue 13)
Témoin #852 (vue 11)
Slide Show
Biblioteca Apostolica Vaticana | Ott. Gr. 354
Biblioteca Nazionale Marciana | Gr. IV. 35
Biblioteca Apostolica Vaticana | Barb. Gr. 438


country, and espied more
country, and espied more

Domain randomization: Co-segmentation

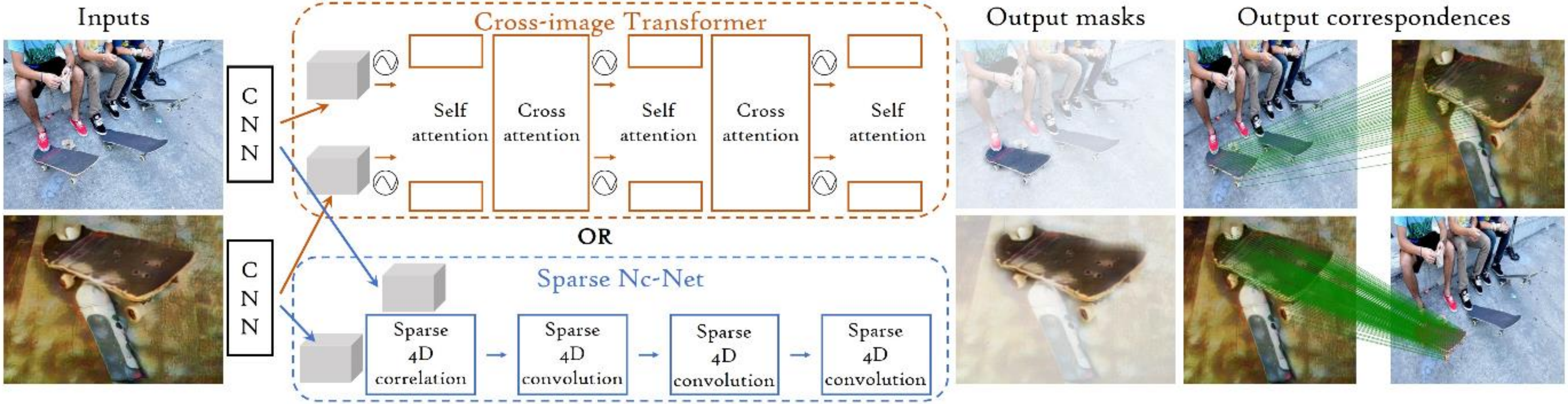
- Goal: identify reccurent objects and their correspondences

+

→

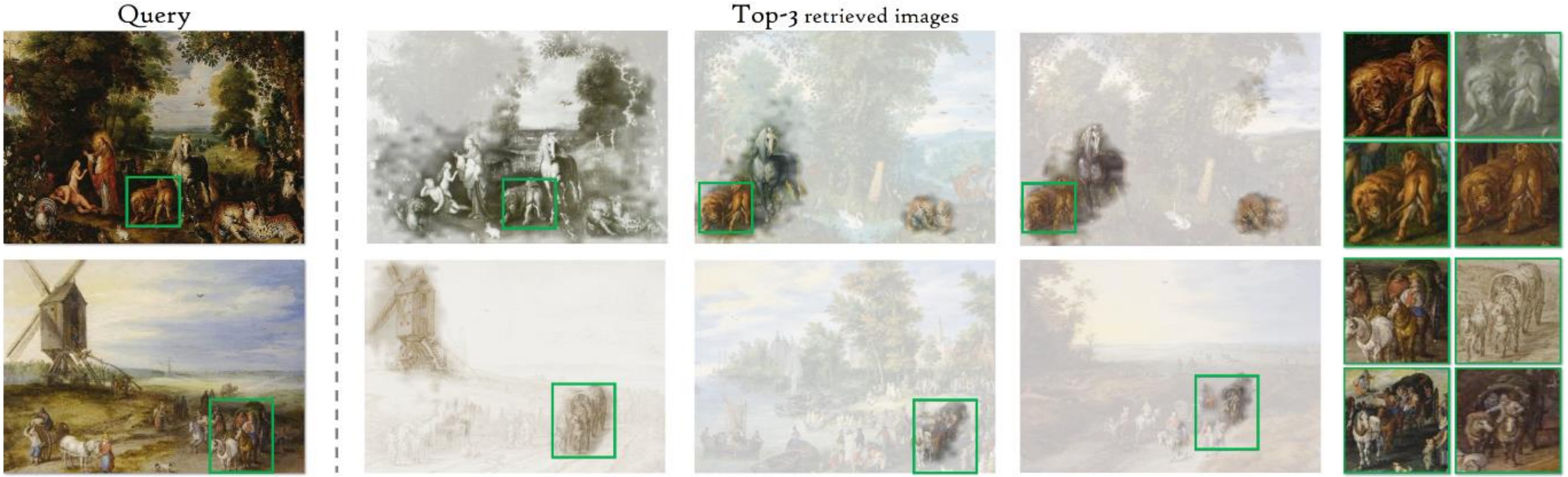
Learning Co-segmentation by Segment Swapping for Retrieval and Discovery
Xi Shen, Alexei Efros, Armand Joulin, Mathieu Aubry, CVPRw 2022

Architecture



Learning Co-segmentation by Segment Swapping for Retrieval and Discovery
Xi Shen, Alexei Efros, Armand Joulin, Mathieu Aubry, CVPRw 2022

Matching results



Learning Co-segmentation by Segment Swapping for Retrieval and Discovery
Xi Shen, Alexei Efros, Armand Joulin, Mathieu Aubry, CVPRw 2022

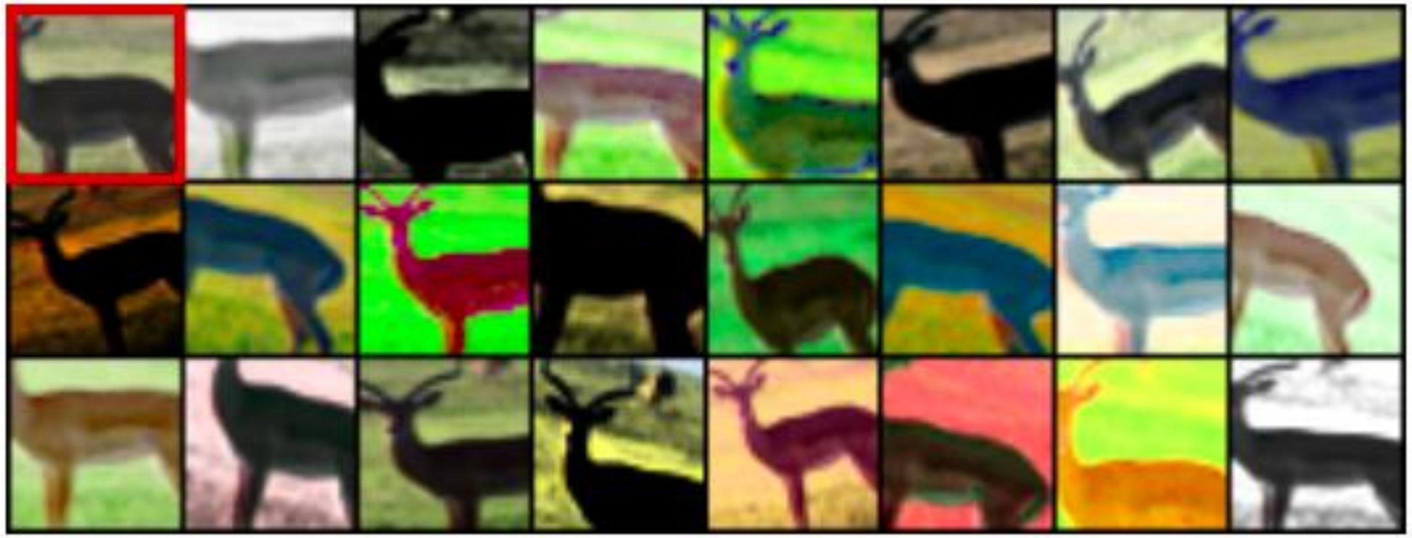
Goes beyond artwork analysis

localization

object discovery

Exemplar CNN

Idea:
1. learn feature with fake classes based on 1 image + augmentations
2. Use the features for another task



This type of extreme data augmentation is important in most self-supervised approaches

Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks
Alexey Dosovitskiy, Philipp Fischer, Jost Tobias Springenberg, Martin Riedmiller, Thomas Brox
NIPS 2014

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

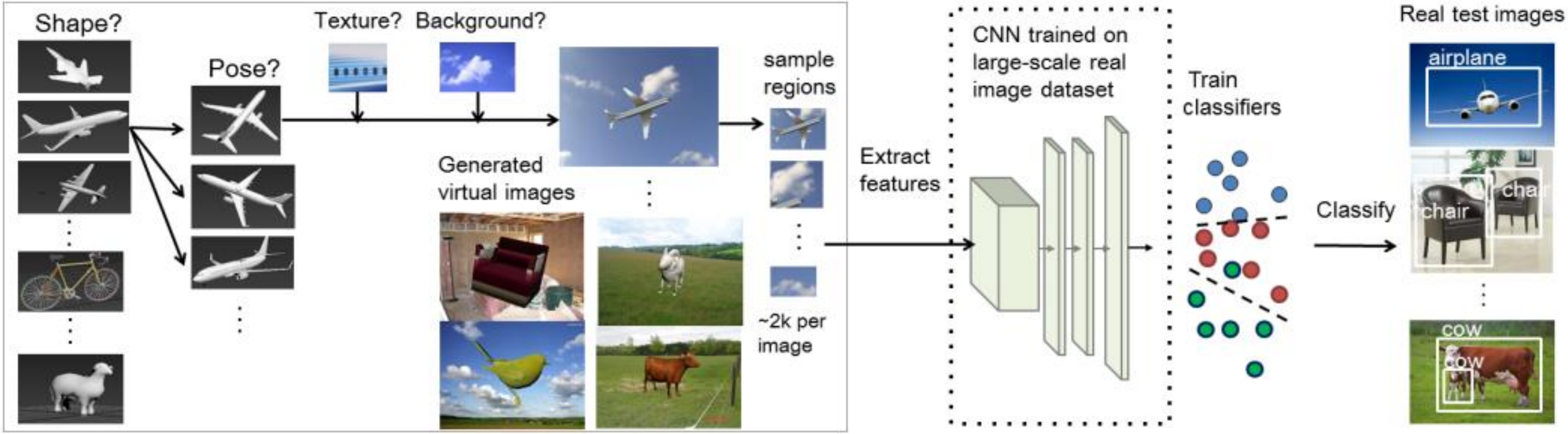
Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

- Domain randomization
- **Realistic data**
- Domain adaptation

Category detection



X. Peng, B. Sun, K. Ali, K. Saenko, ICCV 2015
Learning Deep Object Detectors from 3D Models

Pepik, B., Benenson, R., Ritschel, T., & Schiele, B. GCPR 2015
What Is Holding Back Convnets for Detection?.

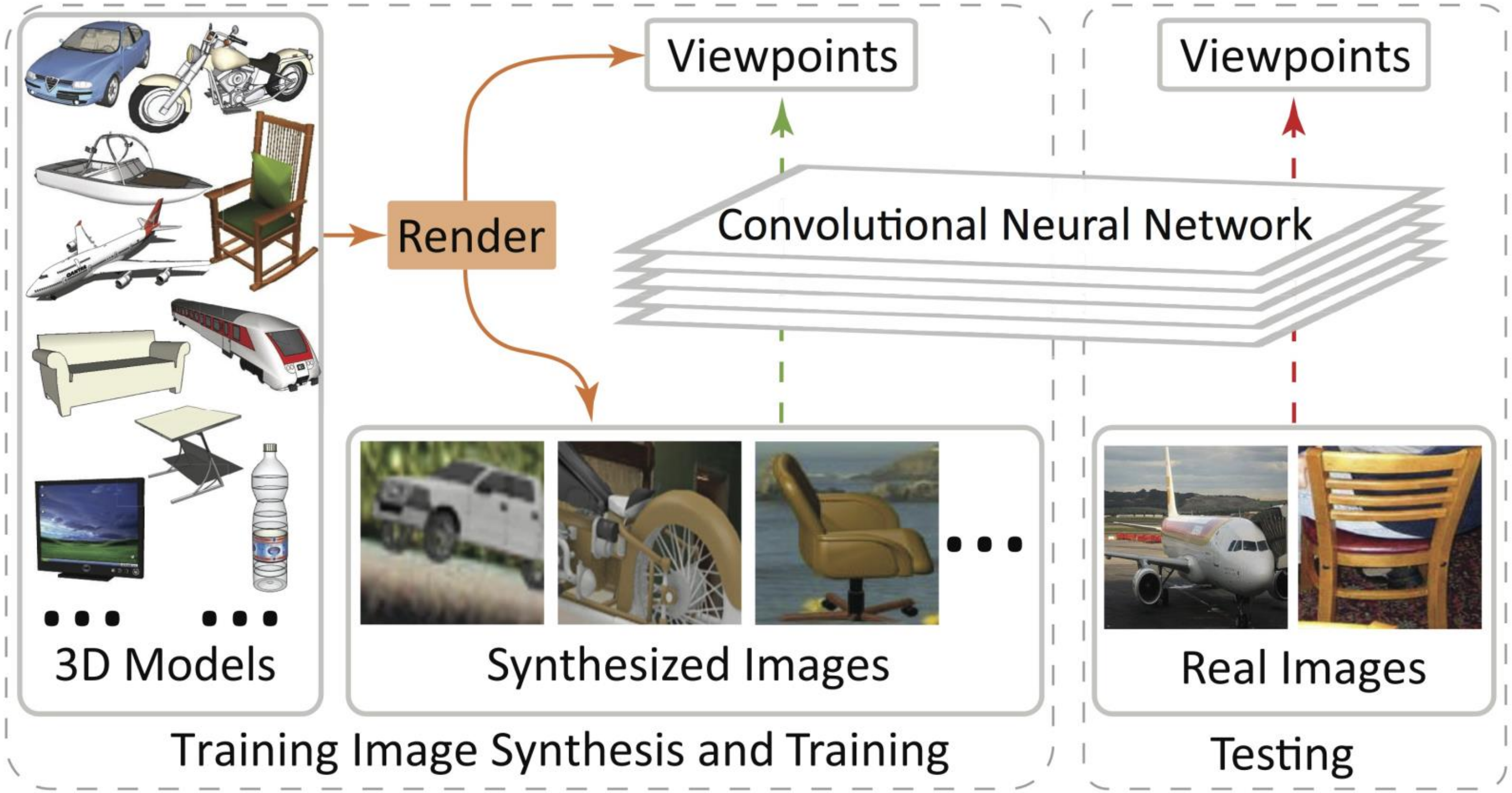
Importance of realism for category detection

	RR-RR	W-RR	W-UG	RR-UG	RG-UG	RG-RR
BG	Real RGB	White	White	Real RGB	Real Gray	Real Gray
TX	Real RGB	Real RGB	Unif. Gray	Unif. Gray	Unif. Gray	Real RGB

IMGNET	aero	bike	bird	boat	botl	bus	car	cat	chr	cow	tab	dog	hse	mbik	pers	plt	shp	sofa	trn	tv	mAP
RR-RR	34.3	34.6	19.9	17.1	10.8	30.0	33.0	18.4	9.7	13.7	1.4	17.6	17.7	34.7	13.9	11.8	15.2	12.7	6.3	26.0	18.9
W-RR	35.9	23.3	16.9	15.0	11.8	24.9	35.2	20.9	11.2	15.5	0.1	15.9	15.6	28.7	13.4	8.9	3.7	10.3	0.6	28.8	16.8
W-UG	38.6	32.5	18.7	14.1	9.7	21.2	36.0	9.9	11.3	13.6	0.9	15.7	15.5	32.3	15.9	9.9	9.7	19.9	0.1	17.4	17.1
RR-UG	26.4	36.3	9.5	9.6	9.4	5.8	24.9	0.4	1.2	12.8	4.7	14.4	9.2	28.8	11.7	9.6	0.7	4.9	0.1	12.2	11.6
RG-UG	32.7	34.5	20.2	14.6	9.4	7.5	30.1	12.1	2.3	14.6	9.3	15.2	11.2	30.2	12.3	11.4	2.2	9.9	0.5	13.1	14.7
RG-RR	26.4	38.2	21.0	15.4	12.1	26.7	34.5	18.0	8.8	16.4	0.4	17.0	20.9	32.1	11.0	14.7	18.4	14.8	6.7	32.0	19.3

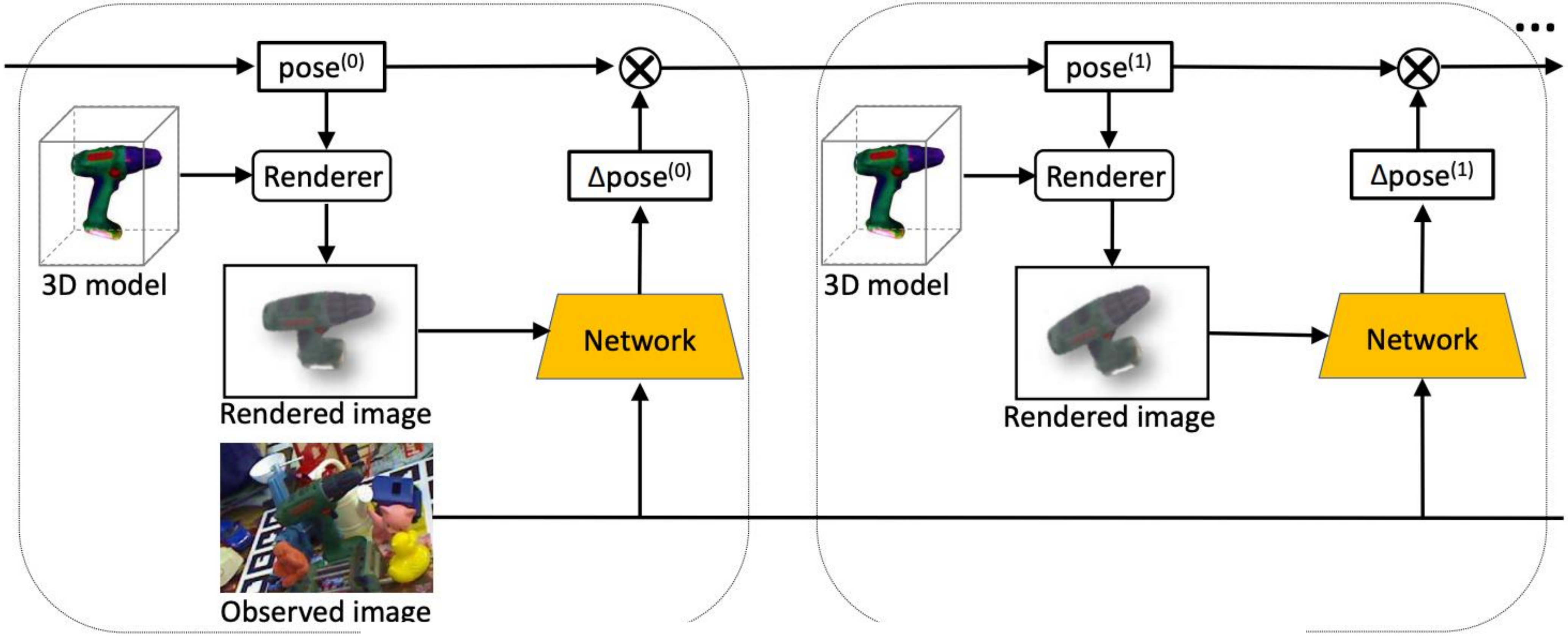
X. Peng, B. Sun, K. Ali, K. Saenko, ICCV 2015
Learning Deep Object Detectors from 3D Models

(1D) Pose estimation



Su, H., Qi, C. R., Li, Y., & Guibas, L. ICCV 2015
Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model

Render&compare for 6D pose estimation



$$L_{\text{pose}}(\mathbf{p}, \hat{\mathbf{p}}) = \frac{1}{n} \sum_{i=1}^{n} L_1\big((\mathbf{R}\mathbf{x}_i + \mathbf{t}) - (\hat{\mathbf{R}}\mathbf{x}_i + \hat{\mathbf{t}})\big)$$

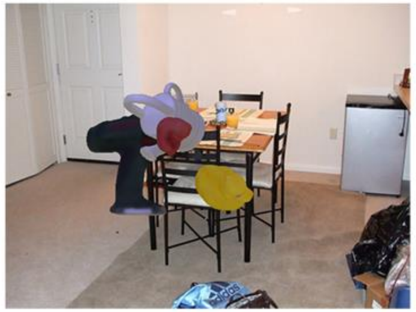
Li, Y., Wang, G., Ji, X., Xiang, Y., & Fox, D. . DeepIM: Deep iterative matching for 6d pose estimation. ECCV 2018

Training data

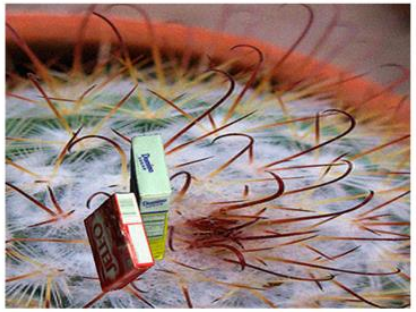
- DeepIM

- BOP challenge on 6D pose estimation 2020

(a) Synthetic Data for LINEMOD



(b) Synthetic Data for Occlusion LINEMOD

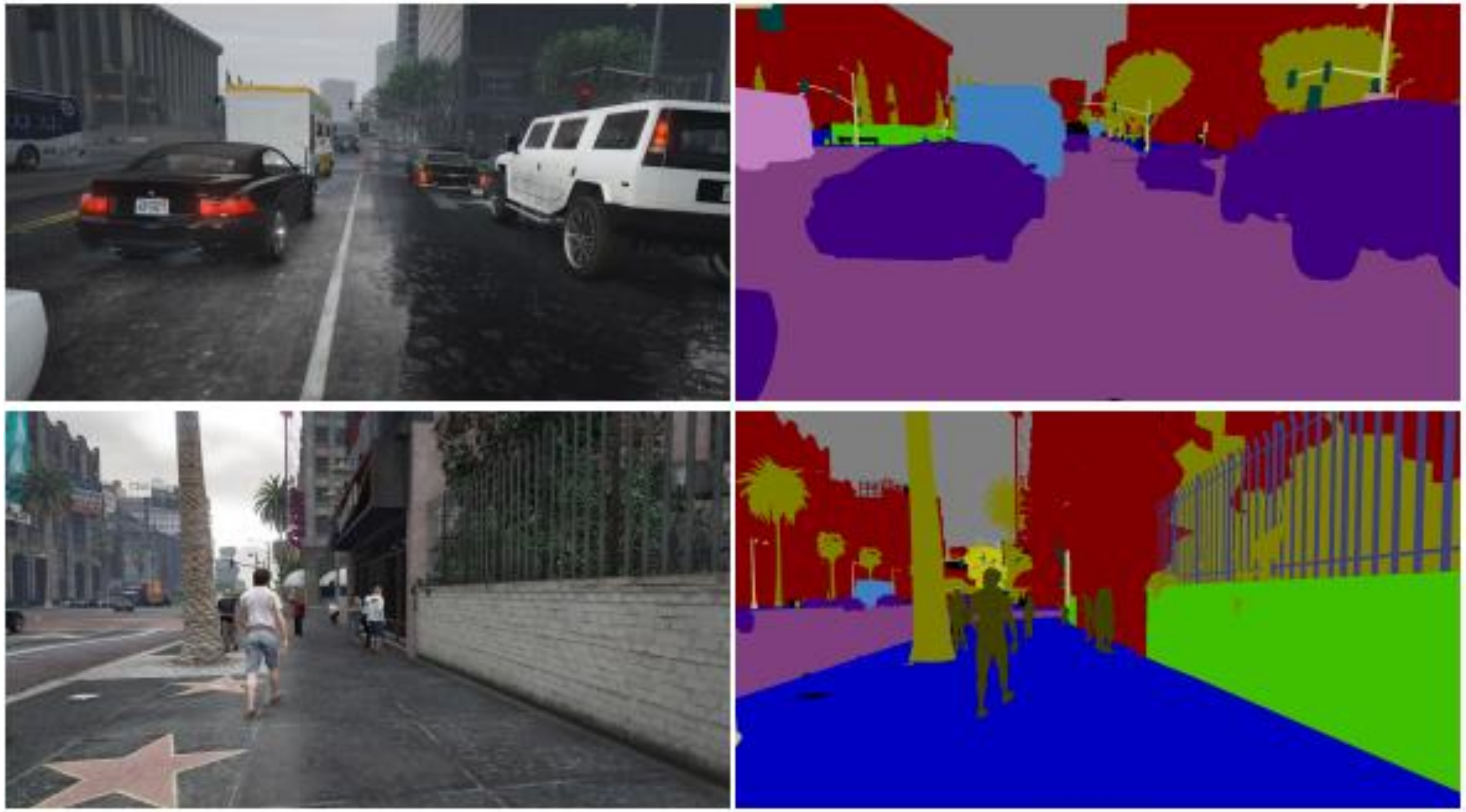


(c) Synthetic Data for YCB-Video

Commonly used “render & paste” synthetic training images

Photorealistic training images rendered by BlenderProc4BOP [7,6]

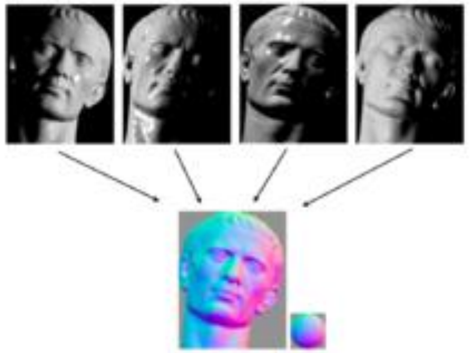
Using realistic game engines



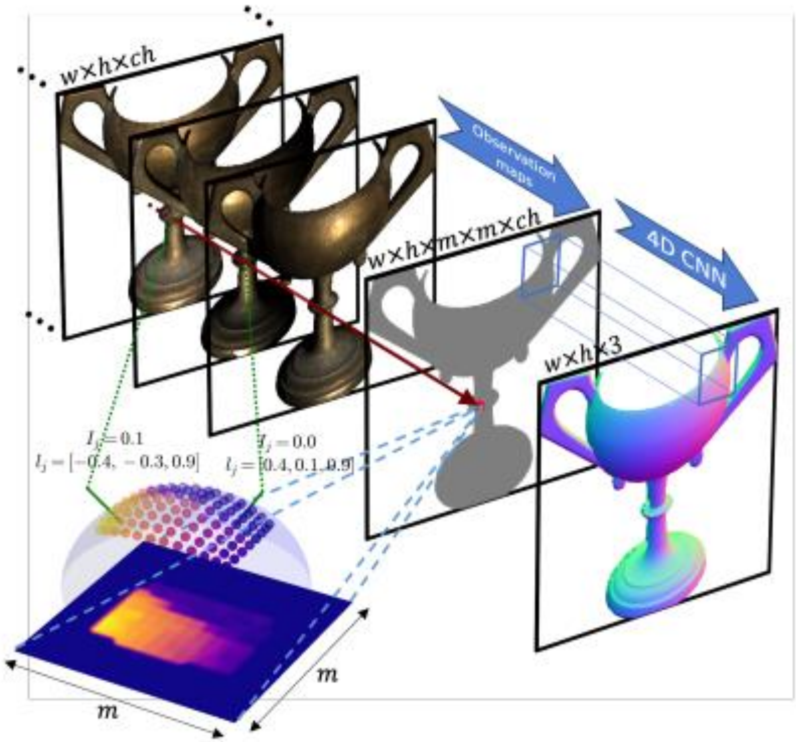
Playing for Data: Ground Truth from Computer Games
S. Richter, V. Vineet, S. Roth, V. Koltun, ECCV 2016

Photometric stereo

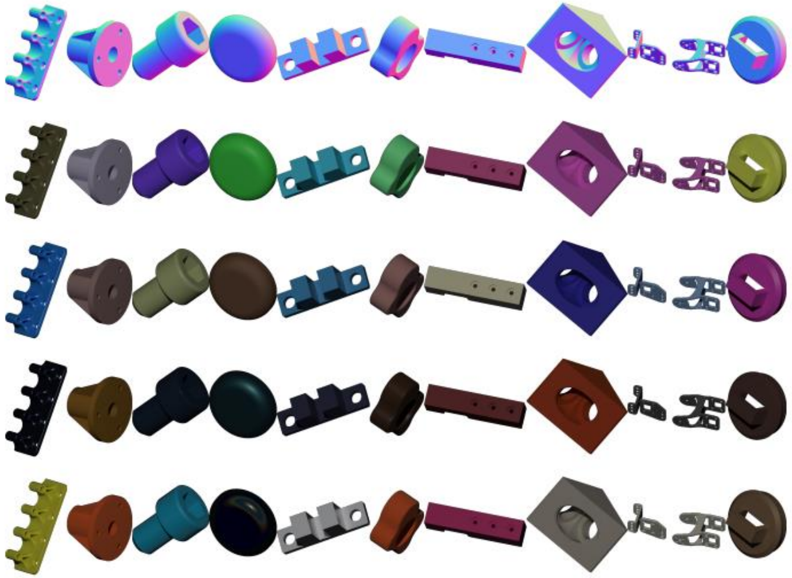
Setting



Approach



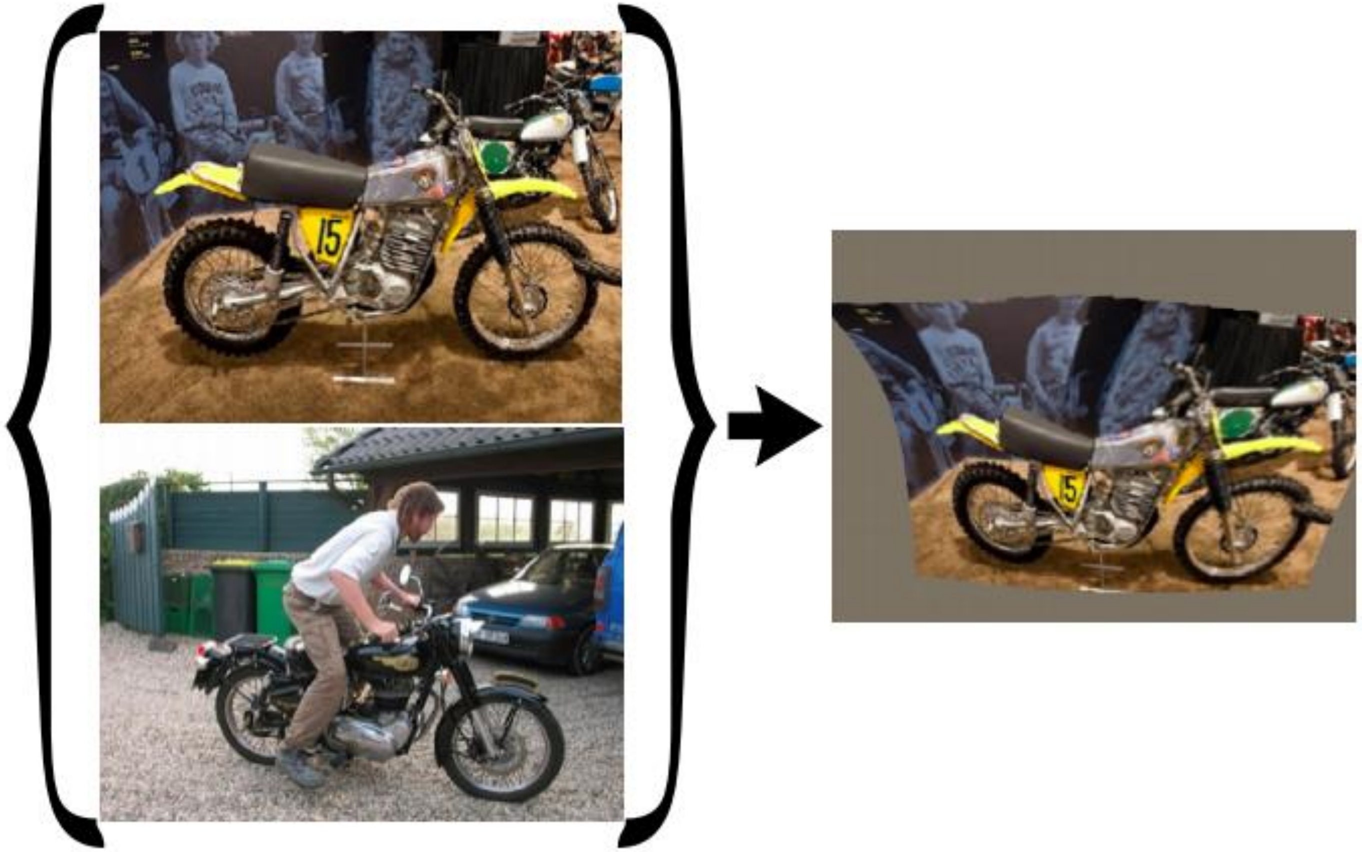
Data



Random shape, camera, material, illumination.
Rendered on the fly.

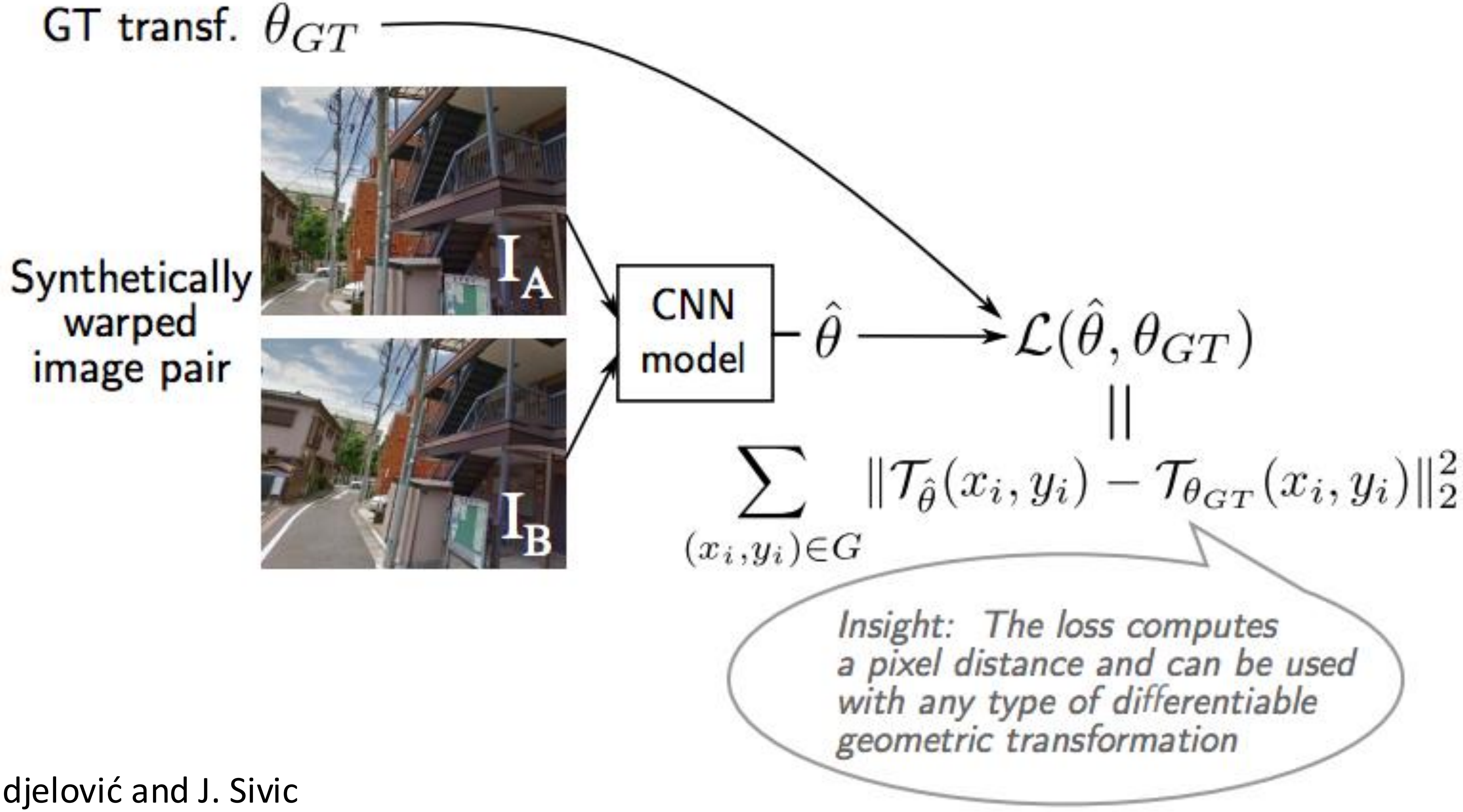
e.g. Aggregating Spatial and Photometric Context for Photometric Stereo, D. Honzátko, EPFL 2024

Category level correspondences

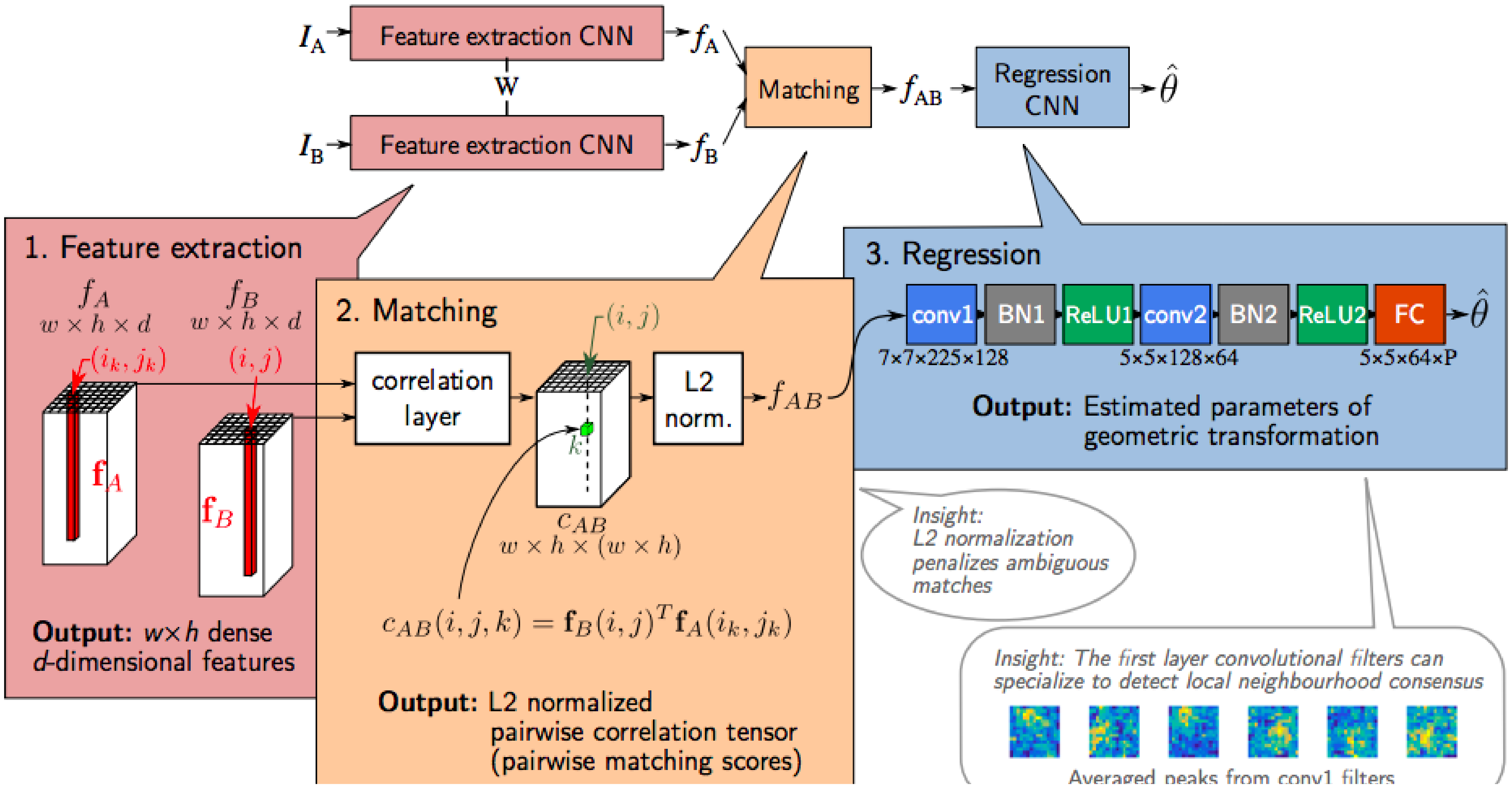


I. Rocco, R. Arandjelović and J. Sivic
Convolutional neural network architecture for geometric matching,
CVPR 2017

Hard annotations: category level correspondences

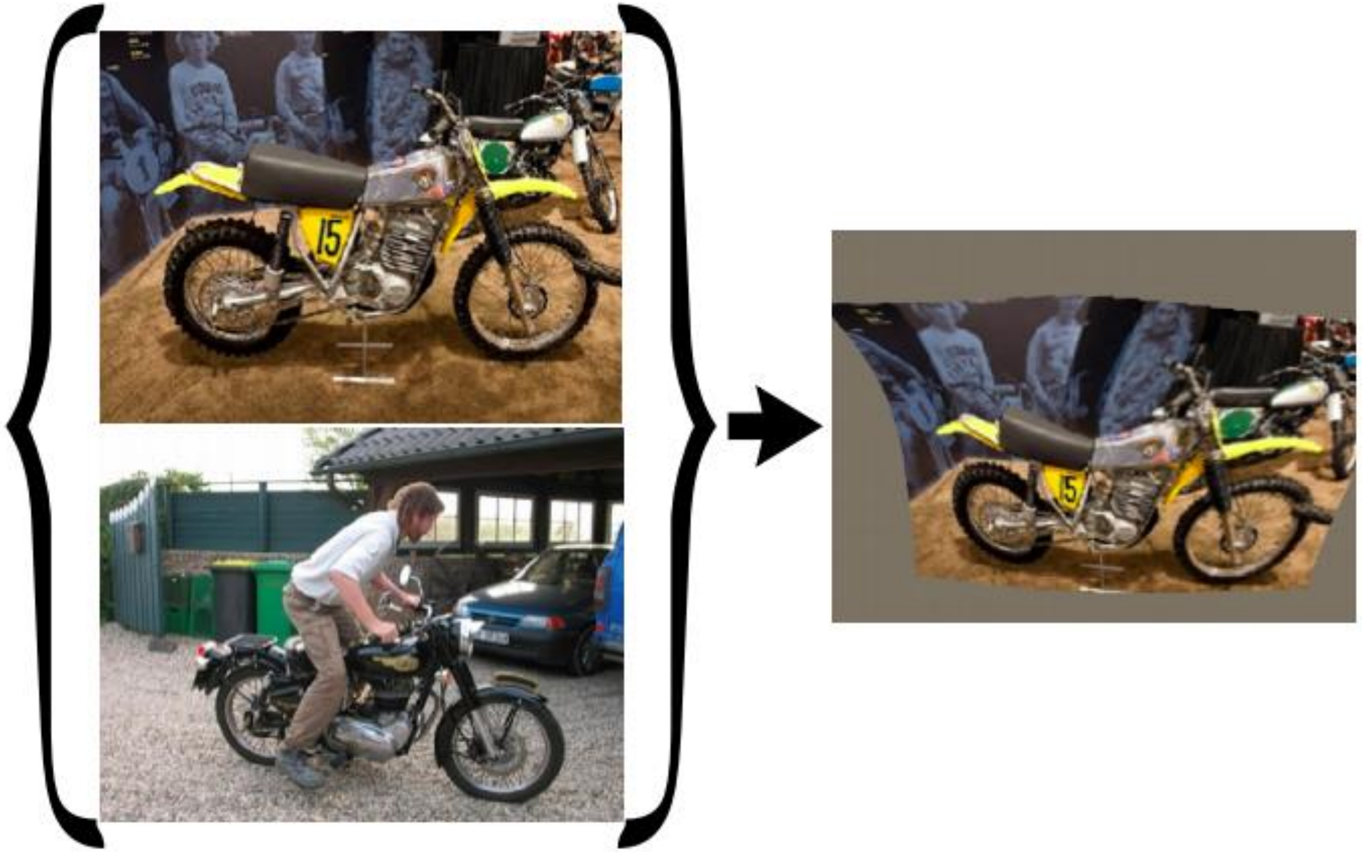


I. Rocco, R. Arandjelović and J. Sivic
Convolutional neural network architecture for geometric matching,
CVPR 2017



I. Rocco, R. Arandjelović and J. Sivic
Convolutional neural network architecture for geometric matching, CVPR 2017

Hard annotations: category level correspondences



I. Rocco, R. Arandjelović and J. Sivic
Convolutional neural network architecture for geometric matching,
CVPR 2017

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

- Domain randomization
- Realistic data
- **Domain adaptation**

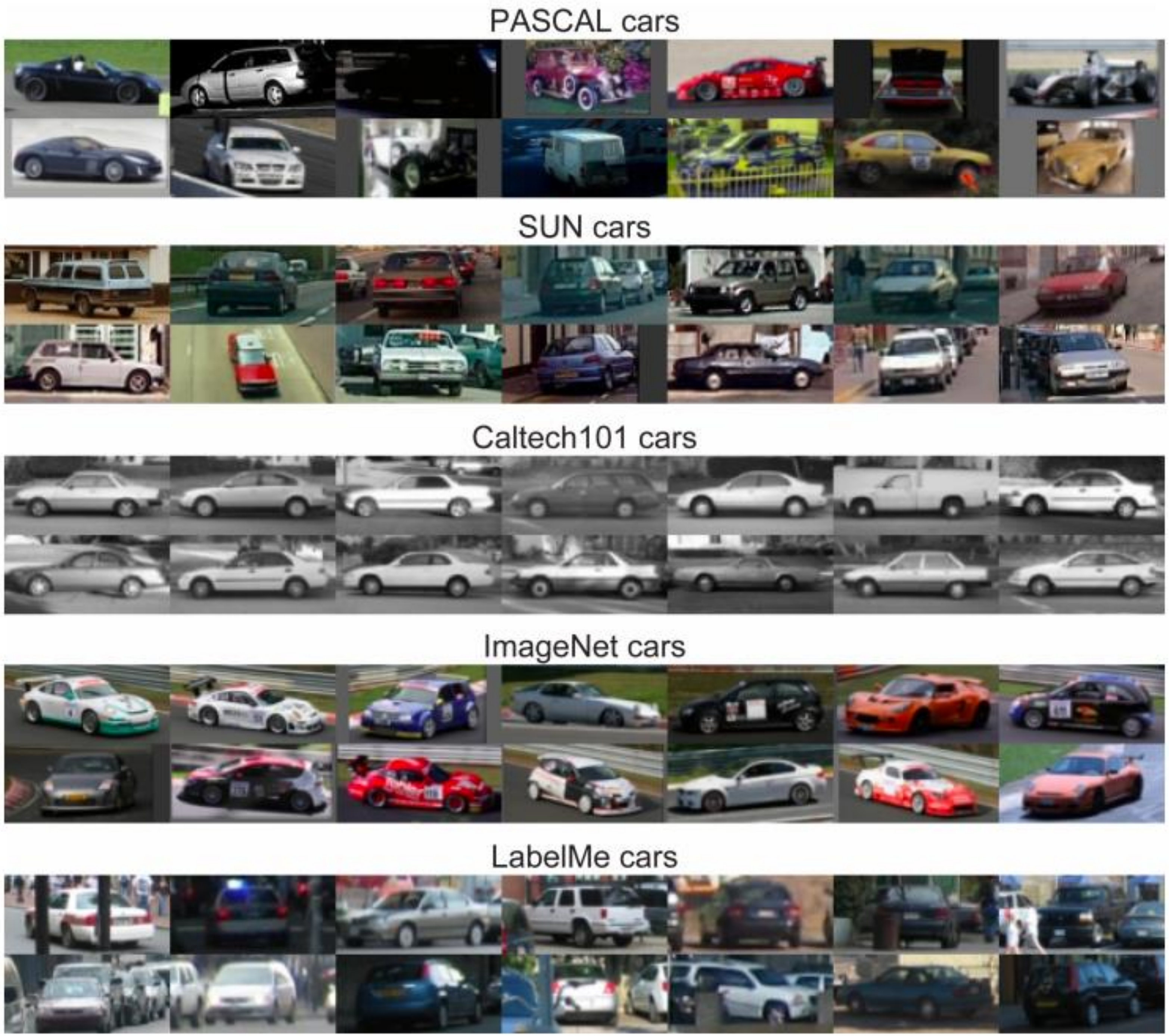
Domain gap / transfer

- Domain gap is a common and important issue, e.g. training on IN testing on Pascal, dataset biais
- Relation to overfitting/generalization/robustness
- Very clear when training data is synthetic

Domain adaptation

- Not specific to CNNs
- Supervised / unsupervised
- Find a mapping / find a common space

Dataset Biais



A. Torralba and A. A. Efros.
Unbiased look at dataset bias.
CVPR 2011

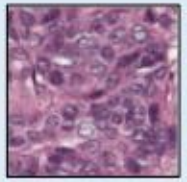
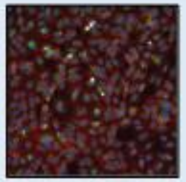
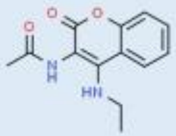
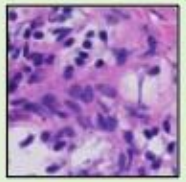
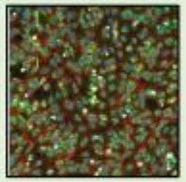
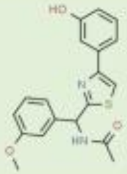
Domain adaptation

- Examples of standard datasets

(a) VisDA-C (b) Office-31 (c) Office-Home

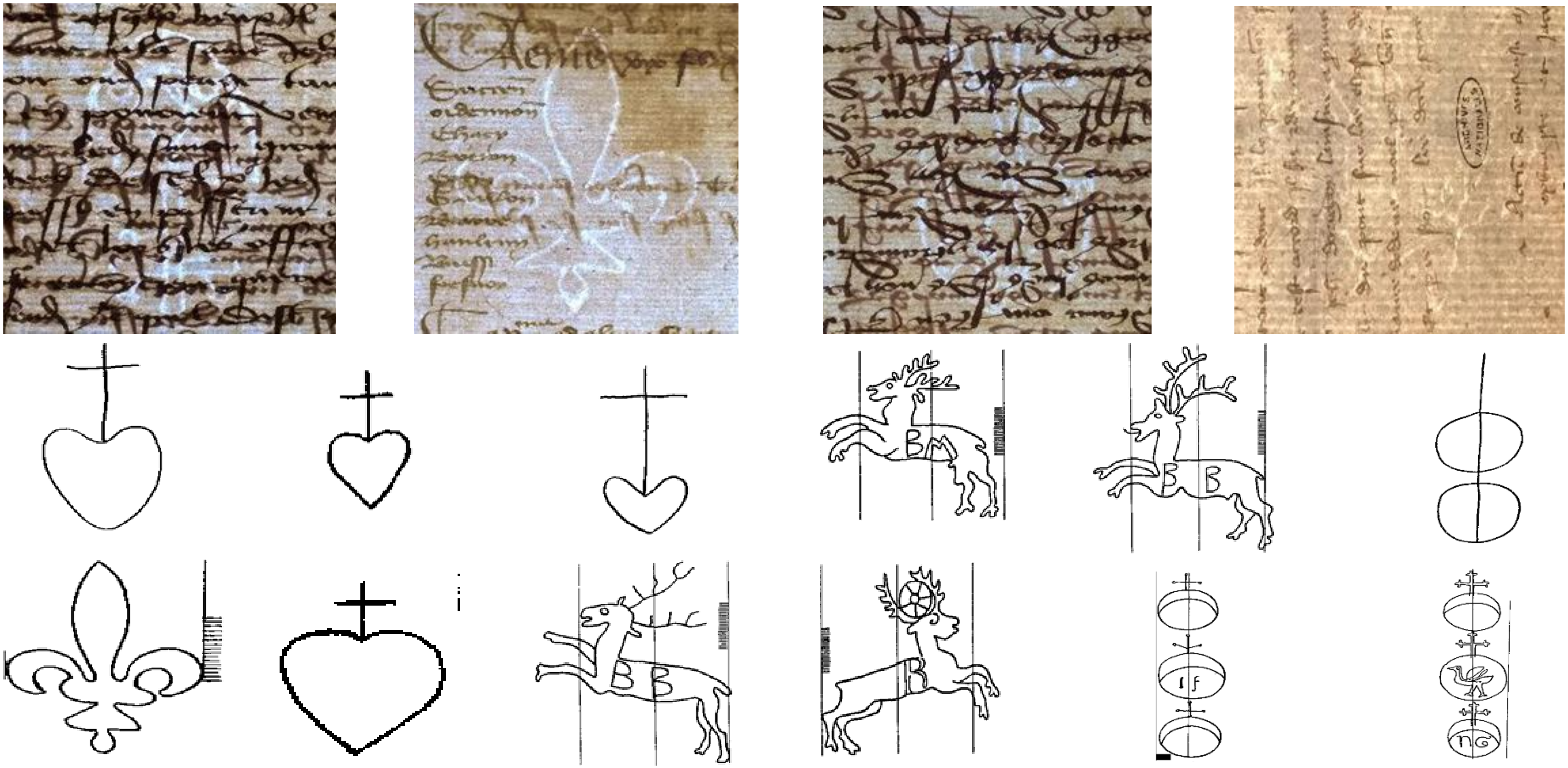
Image from : Chang, W. G., You, T., Seo, S., Kwak, S., & Han, B.
Domain-Specific Batch Normalization for Unsupervised Domain Adaptation.
CVPR 2019

Domain adaptation

	Domain generalization					Subpopulation shift	Domain generalization + subpopulation shift			
Dataset	**iWildCam**	**Camelyon17**	**RxRx1**	**OGB-MolPCBA**	**GlobalWheat**	**CivilComments**	**FMoW**	**PovertyMap**	**Amazon**	**Py150**
Input (x)	camera trap photo	tissue slide	cell image	molecular graph	wheat image	online comment	satellite image	satellite image	product review	code
Prediction (y)	animal species	tumor	perturbed gene	bioassays	wheat head bbox	toxicity	land use	asset wealth	sentiment	autocomplete
Domain (d)	camera	hospital	batch	scaffold	location, time	demographic	time, region	country, rural-urban	user	git repository
# domains	323	5	51	120,084	47	16	16 x 5	23 x 2	2,586	8,421
# examples	203,029	455,954	125,510	437,929	6,515	448,000	523,846	19,669	539,502	150,000
Train example						What do Black and LGBT people have to do with bicycle licensing?			Overall a solid package that has a good quality of construction for the price.	`import numpy as np … norm=np.___`
Test example						As a Christian, I will not be patronizing any of those businesses.			I *loved* my French press, it's so perfect and came with all this fun stuff!	`import subprocess as sp p=sp.Popen() stdout=p.___`
Adapted from	Beery et al. 2020	Bandi et al. 2018	Taylor et al. 2019	Hu et al. 2020	David et al. 2021	Borkan et al. 2019	Christie et al. 2018	Yeh et al. 2020	Ni et al. 2019	Raychev et al. 2016

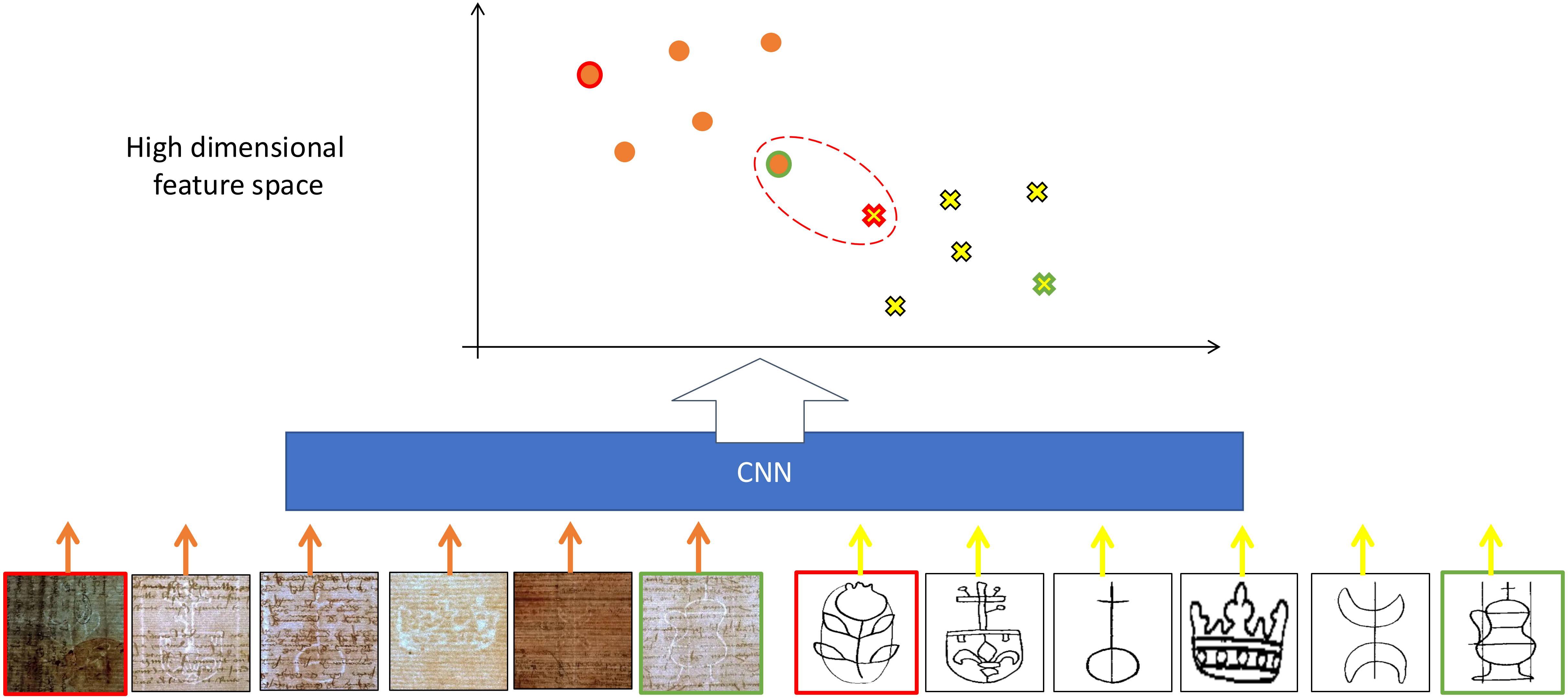
Koh, Pang Wei, et al. "Wilds: A benchmark of in-the-wild distribution shifts." *ICML* 2021.

Example: Watermark recognition



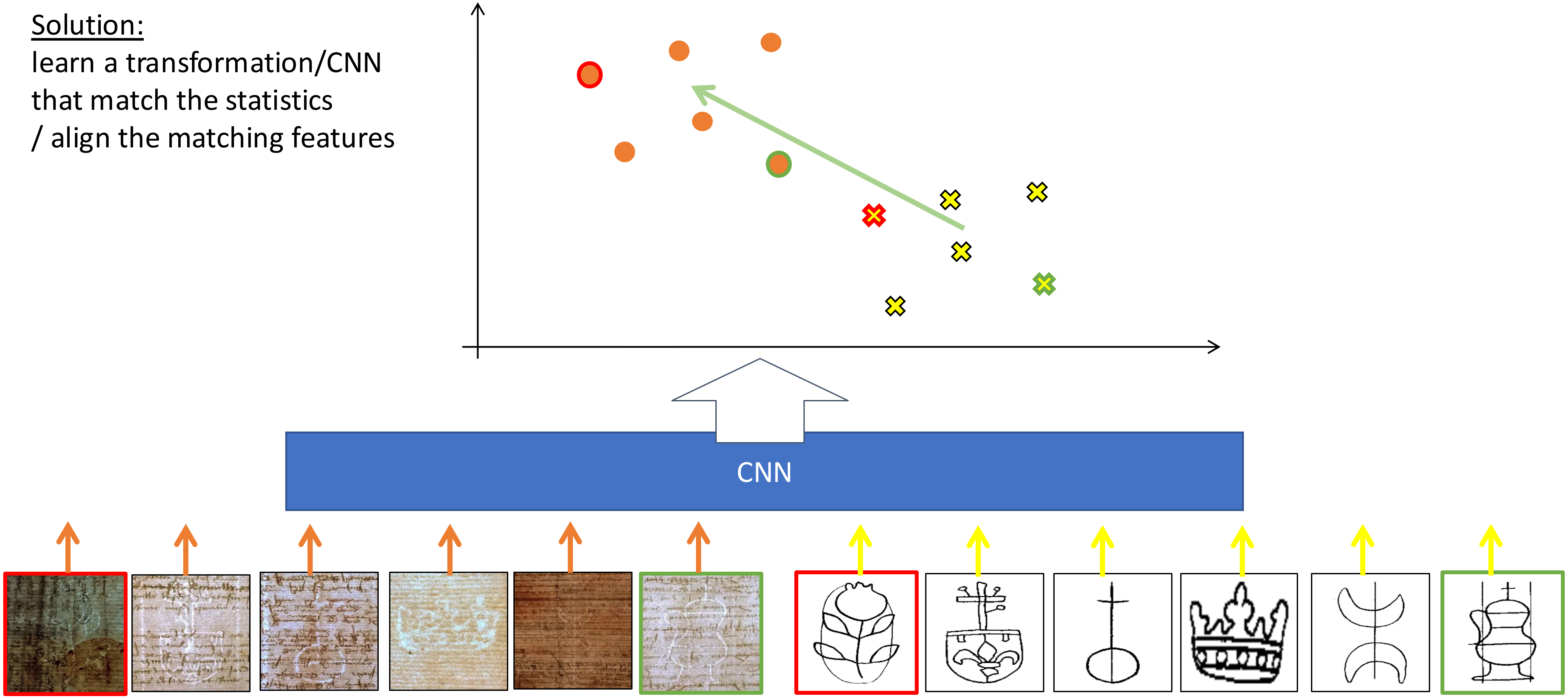
Domain Adaptation

High dimensional
feature space

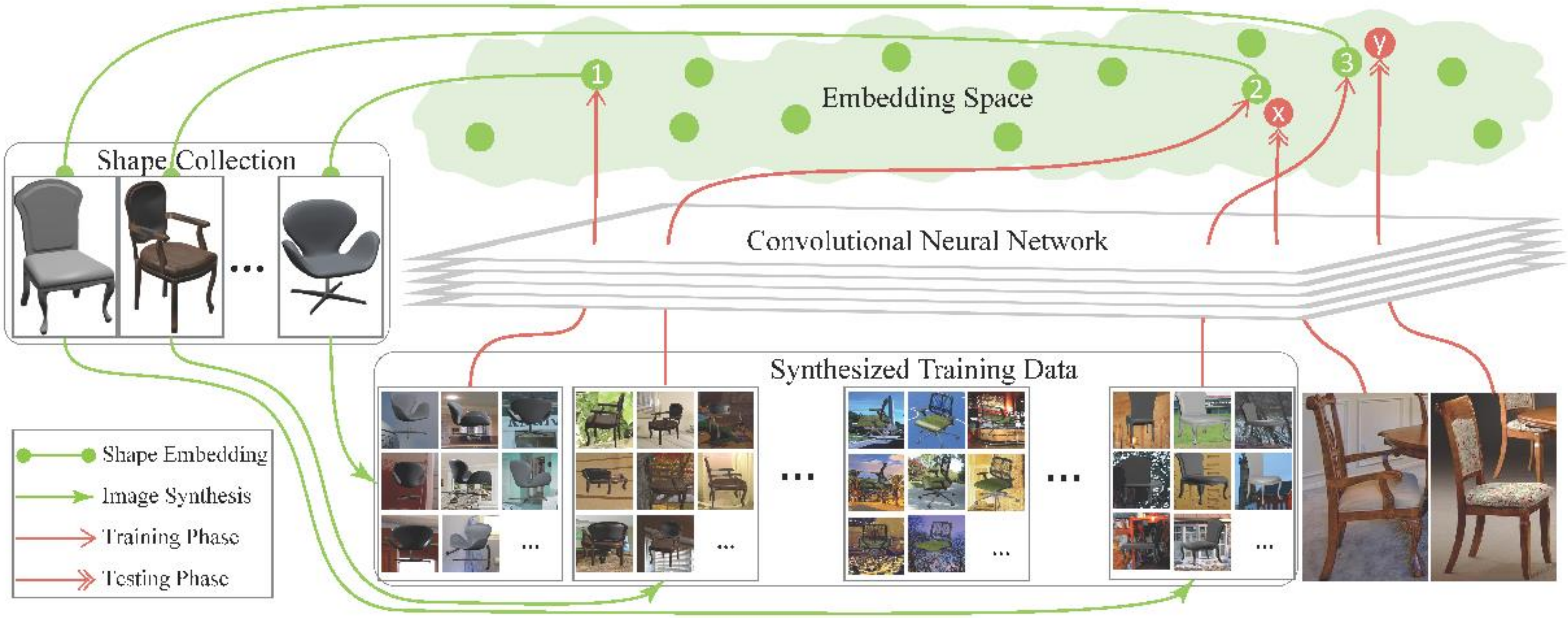


Domain Adaptation

<u>Solution:</u>
learn a transformation/CNN
that match the statistics
/ align the matching features

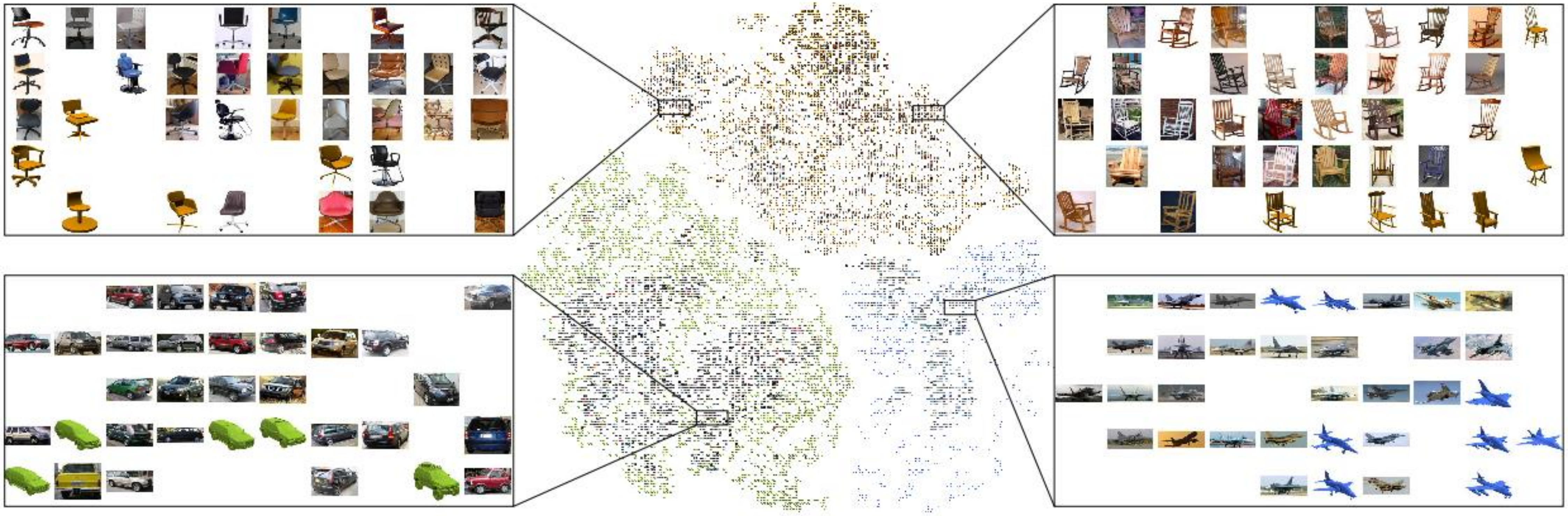


Learning joint embedding: example of 3D models and real images



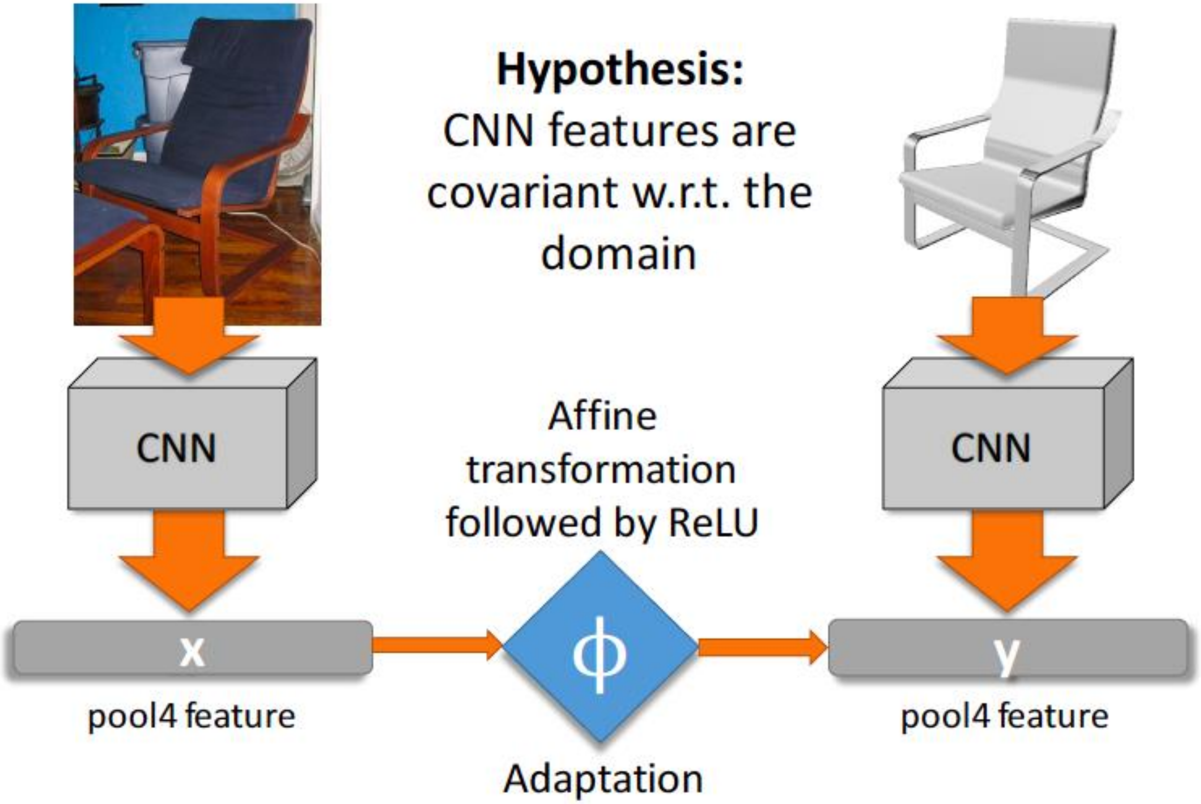
Li, Y., Su, H., Qi, C. R., Fish, N., Cohen-Or, D., & Guibas, L. J. TOG 2015
Joint embeddings of shapes and images via CNN image purification.

Learning joint embedding: example of 3D models and real images



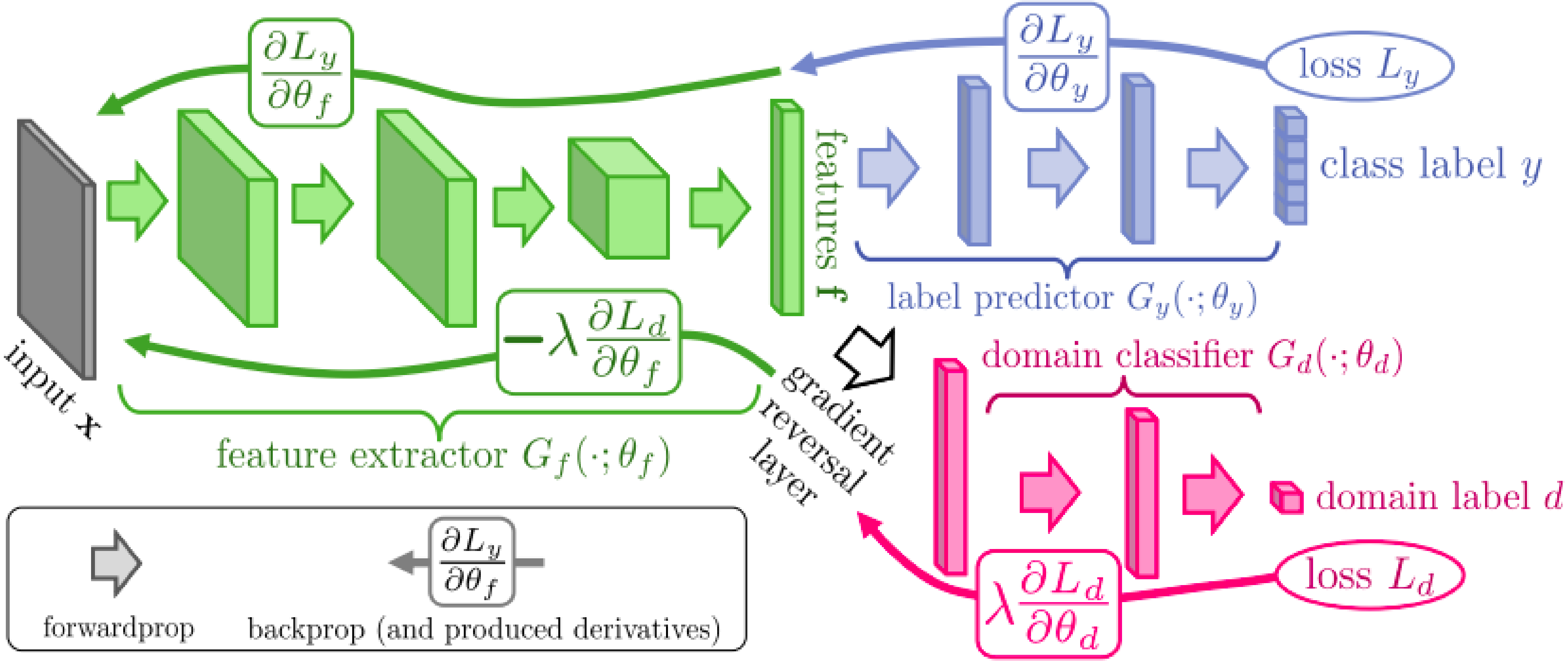
Li, Y., Su, H., Qi, C. R., Fish, N., Cohen-Or, D., & Guibas, L. J. TOG 2015
Joint embeddings of shapes and images via CNN image purification.

Learning adaptation: e.g. 3D instance detection



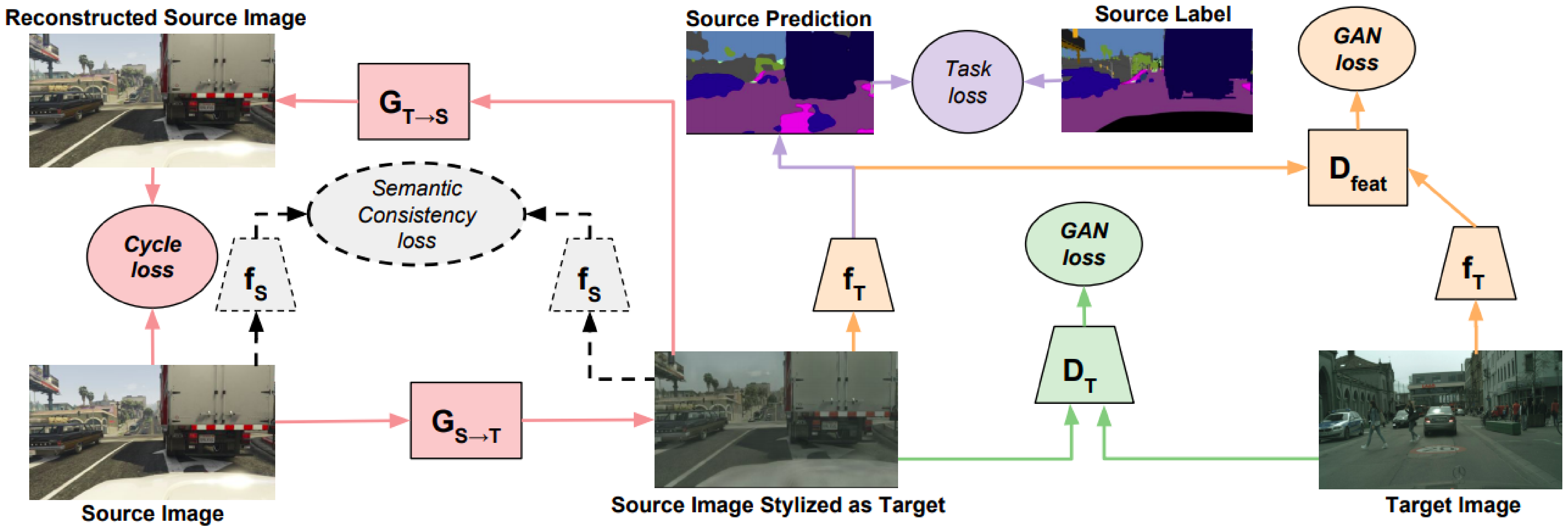
$$L(\phi) = -\sum_{i=1}^{N} \underbrace{\boxed{S(\phi(x_i), y_i)}}_{\text{Cosine Similarity}} + \underbrace{\boxed{R(\phi)}}_{\text{L2 Regularization}}$$

Adapting statistics using adversarial training



Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. Domain-adversarial training of neural networks.
JMLR 2016

Cycles for domain adaptation



Hoffman, J., Tzeng, E., Park, T., Zhu, J. Y., Isola, P., Saenko, K., ... & Darrell, T.
Cycada: Cycle-consistent adversarial domain adaptation.
ICLR 2018

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

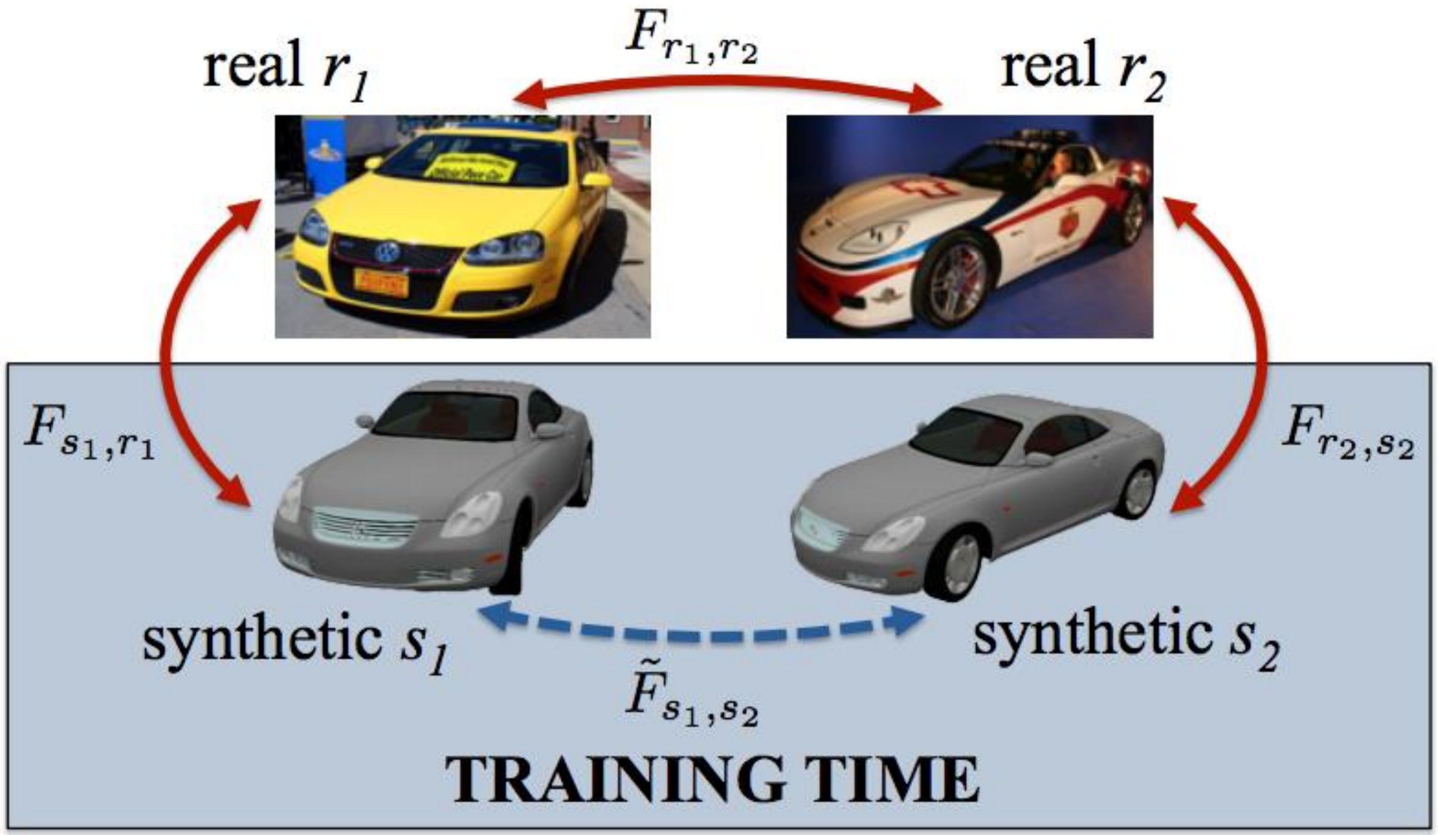
Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

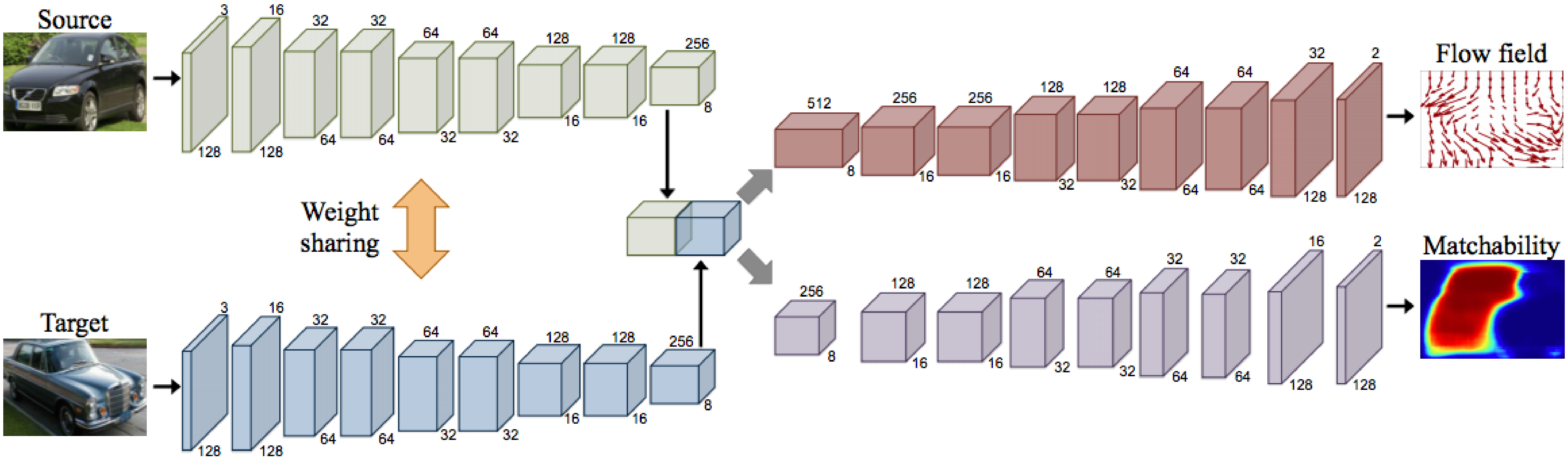
- Domain randomization
- Realistic data
- Domain adaptation
- **Other**

Cycle-consistency for dense category-level correspondences



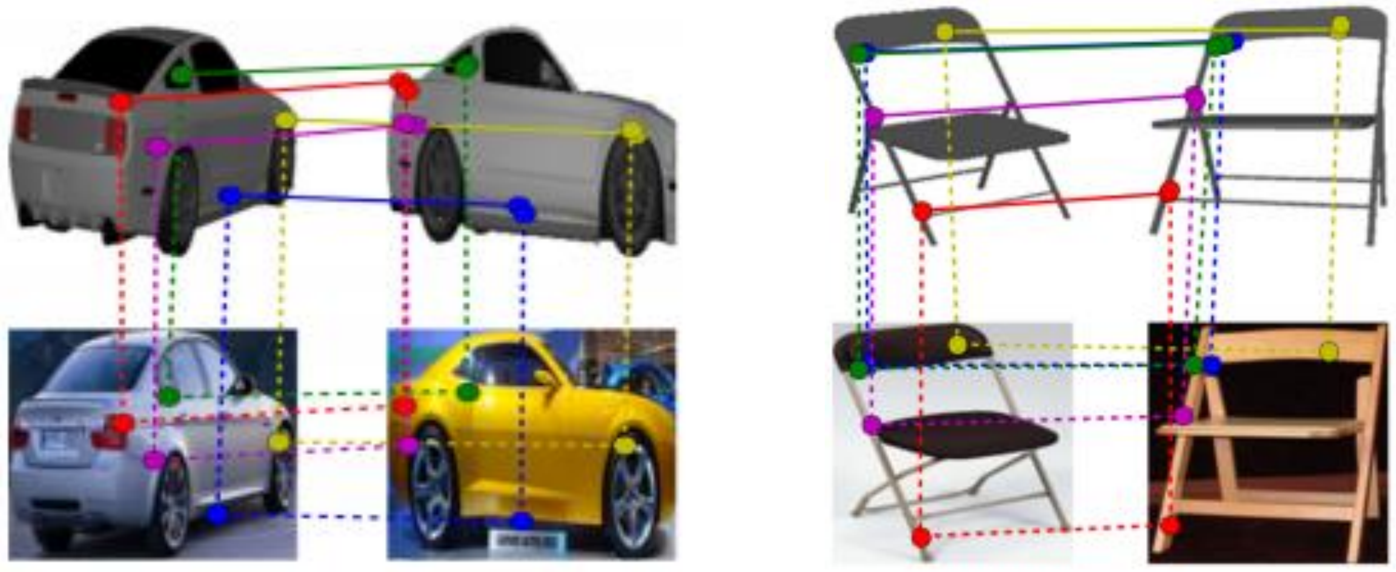
Learning Dense Correspondence via 3D-guided Cycle Consistency
T Zhou, P Krähenbühl, M Aubry, Q Huang, AA Efros, CVPR 2016

Dense category-level correspondences



Learning Dense Correspondence via 3D-guided Cycle Consistency
T Zhou, P Krähenbühl, M Aubry, Q Huang, AA Efros, CVPR 2016

Dense category-level correspondences



Learning Dense Correspondence via 3D-guided Cycle Consistency
T Zhou, P Krähenbühl, M Aubry, Q Huang, AA Efros, CVPR 2016

Outline: Deep learning and 3D data

Important milestones:

1. Classification and Segmentation
2. Matching / Alignment
3. Generation and single view reconstruction

Recent works I am excited about:

4. Structured generation
5. Unsupervised single view reconstruction

Learning with synthetic data

- Domain randomization
- Realistic data
- Domain adaptation