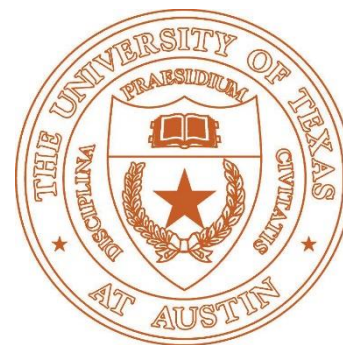


GAMES

Geometric Deep Learning III

Qixing Huang
Oct. 14th 2021

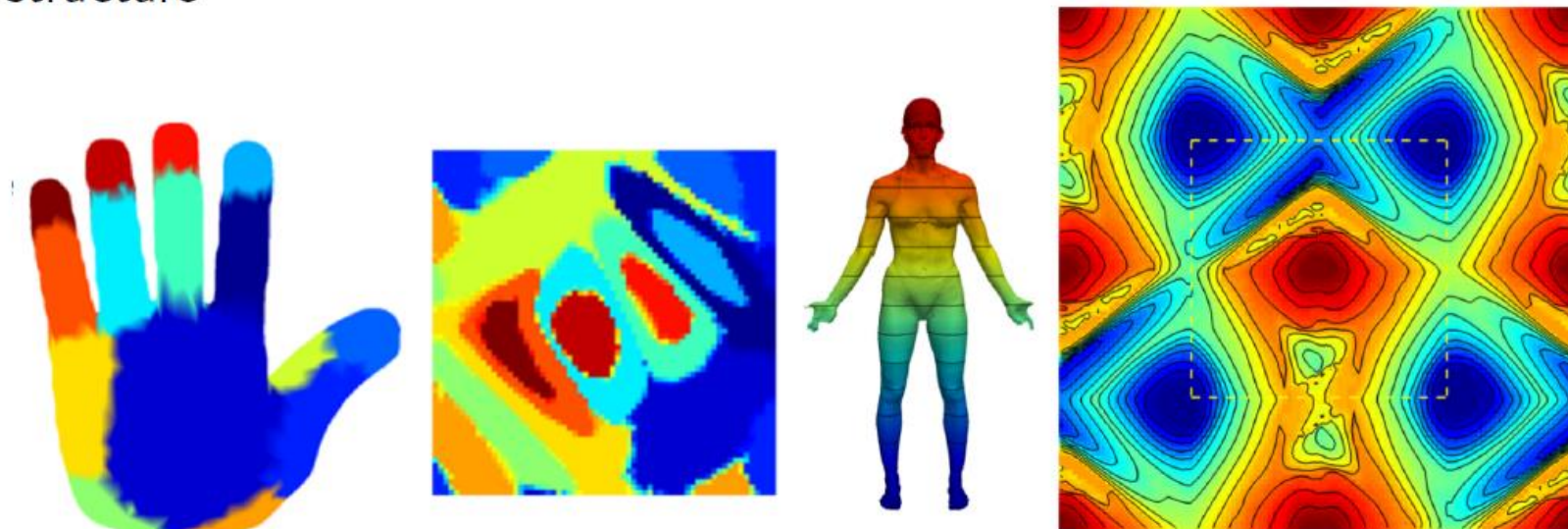


Slide credit: Michael Bronstein

Parametric domain
geometric deep learning methods

Global parametrization

Map the input surface to some **parametric domain** with shift-invariant structure



- ☺ Allows to use standard CNNs (pull back convolution from the parametric space)
- ☺ Guaranteed invariance to some classes of transformations
- ☹ Parametrization may not be unique
- ☹ Embedding may introduce distortion

Translation invariance on manifolds

Translation on manifold = locally Euclidean translation

Translation invariance on manifolds

Translation on manifold = locally Euclidean translation = flow along a non-vanishing vector field

Translation invariance on manifolds

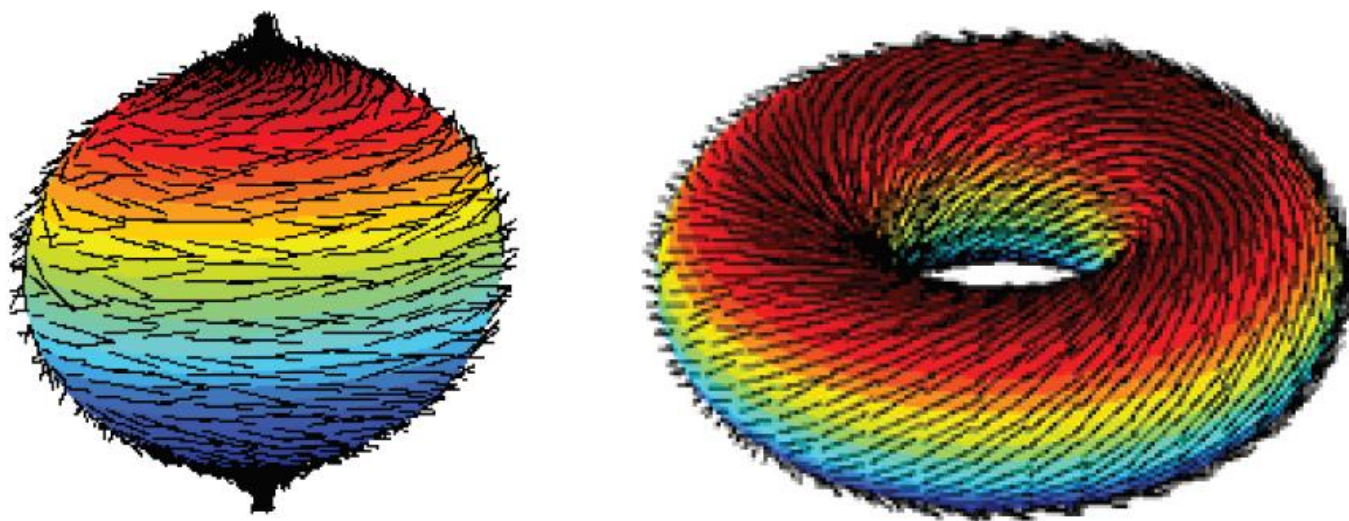
Translation on manifold = locally Euclidean translation = flow along a non-vanishing vector field

Poincaré-Hopf Theorem Non-vanishing vector field on a closed orientable compact 2-manifold implies manifold of genus 1 (torus)

Translation invariance on manifolds

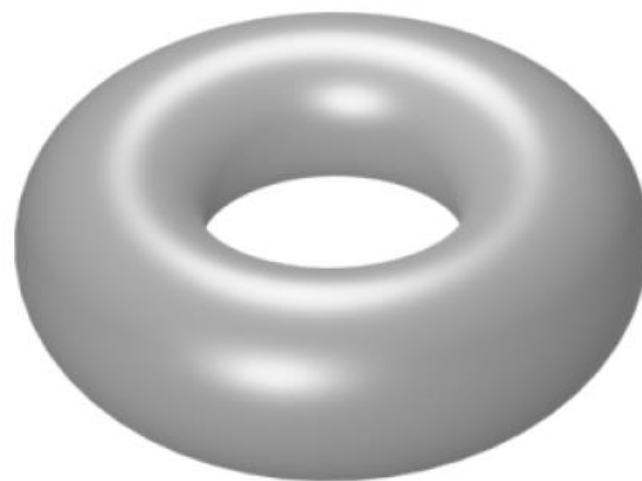
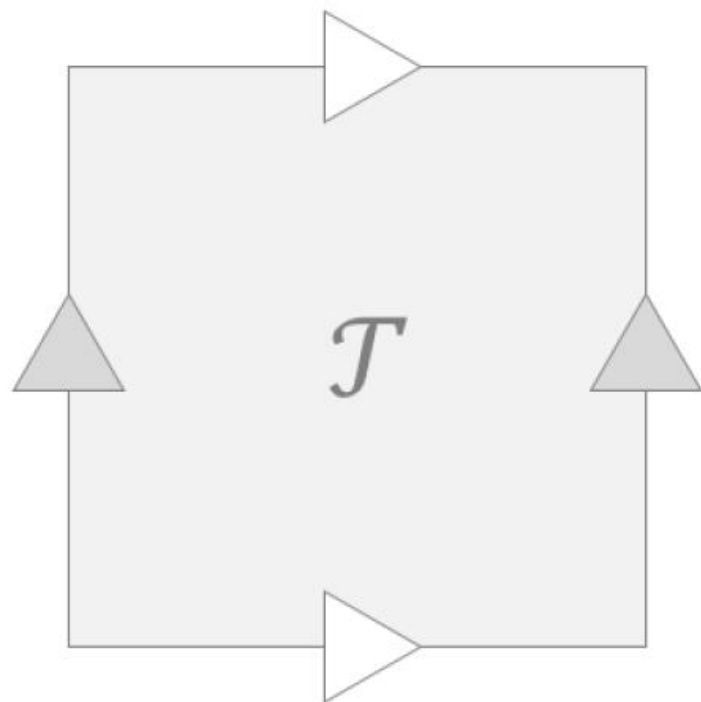
Translation on manifold = locally Euclidean translation = flow along a non-vanishing vector field

Poincaré-Hopf Theorem Non-vanishing vector field on a closed orientable compact 2-manifold implies manifold of genus 1 (torus)



'Hairy Ball Theorem' states that a sphere cannot be combed

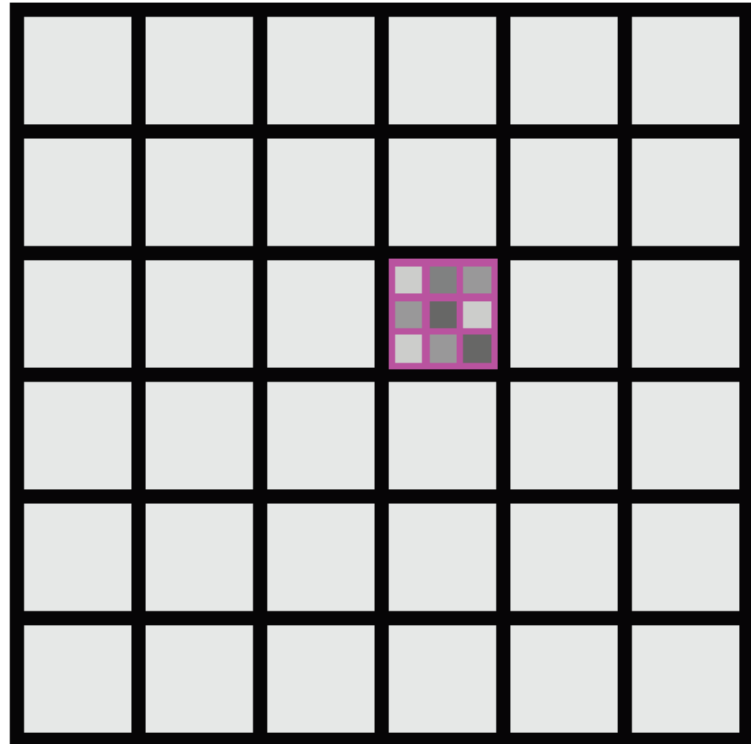
Translation invariance on the torus



Torus is the only closed orientable surface admitting a translation group

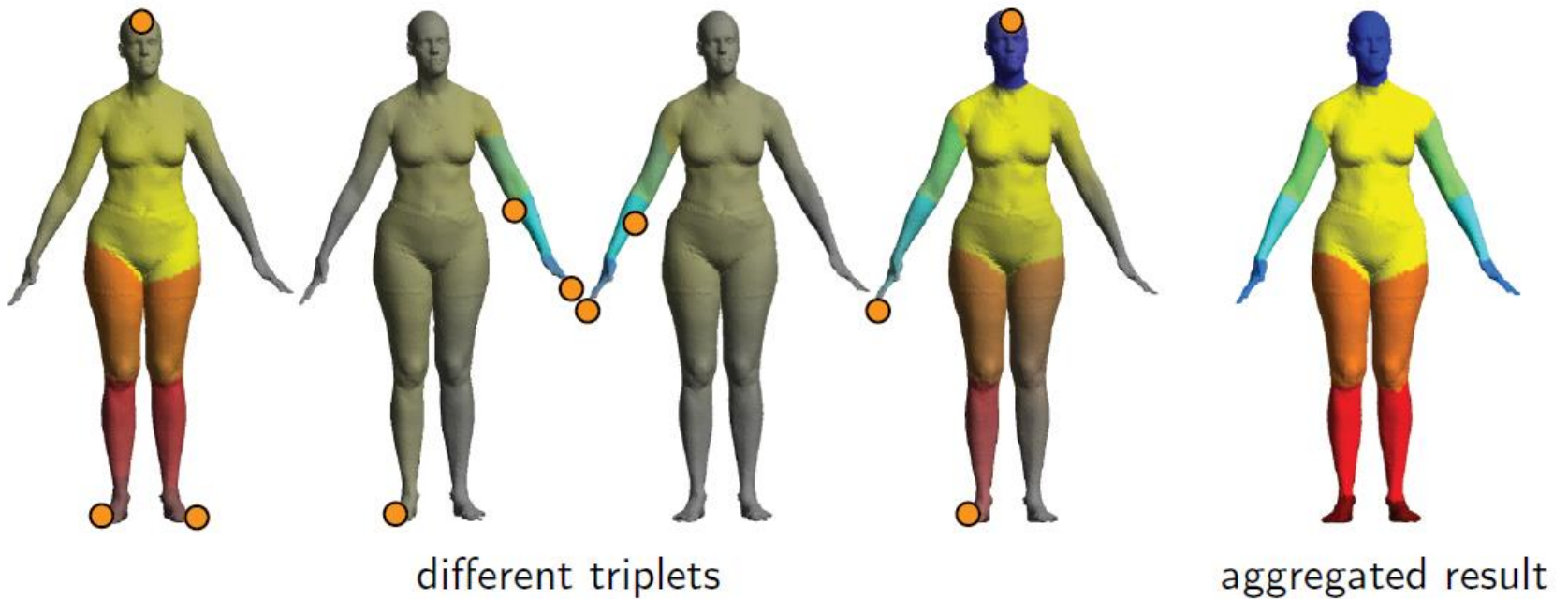
Convolution on torus

For any triplet of points on \mathcal{X} , construct [conformal homeomorphism](#) from the 4-cover \mathcal{X}^4 to \mathcal{T} using [orbifold-Tutte](#) method

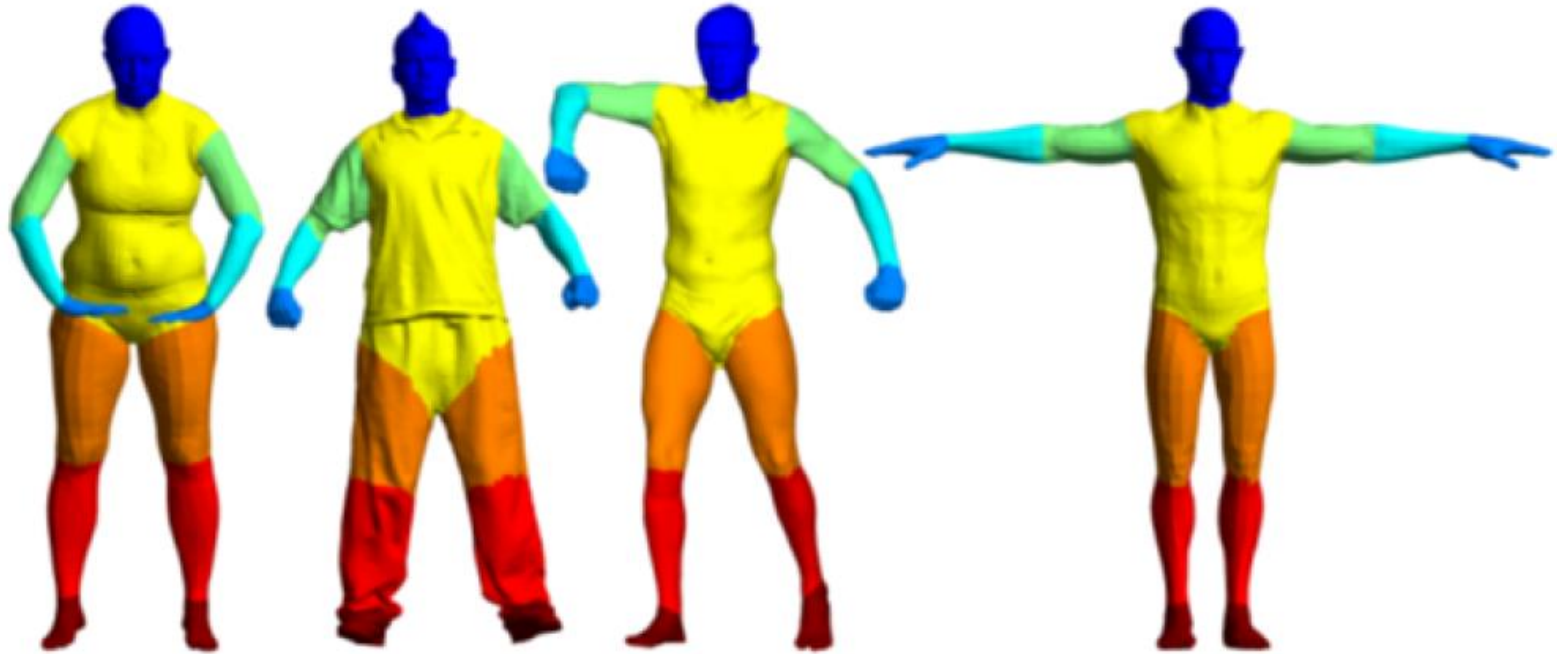


Conformal zoom

- Embedding depends on the choice of the triplets of points
- 'Conformal zoom' effect
- Choose multiple triples and aggregate results in training / test phase

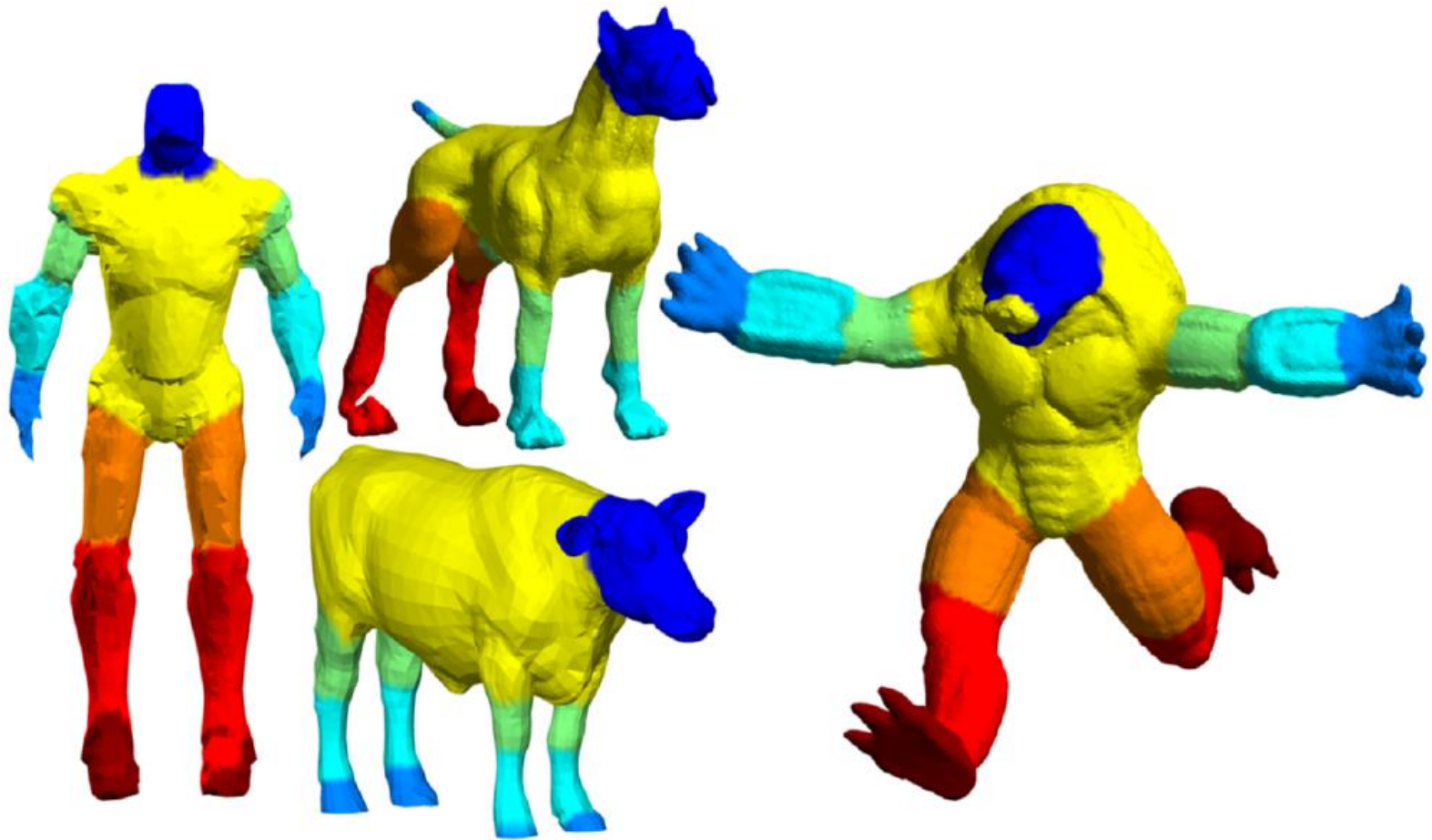


Example: shape segmentation with Toric CNN



Examples of shape segmentation obtained with Toric CNN

Example: shape segmentation with Toric CNN



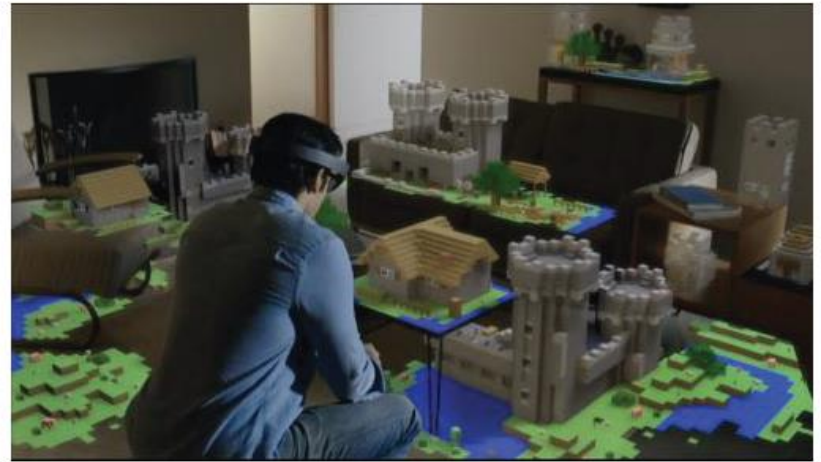
Examples of shape segmentation obtained with Toric CNN

Application in Computer Graphics and 3D Vision

Application dealing with 3D data



Computer graphics



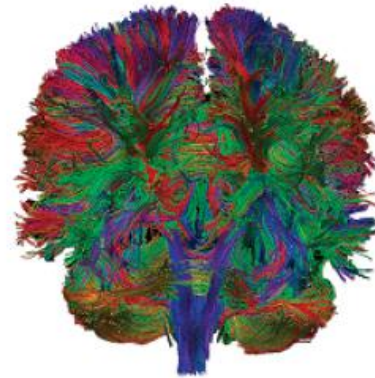
Virtual/augmented reality



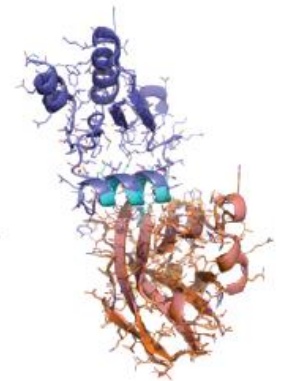
Robotics



Autonomous driving

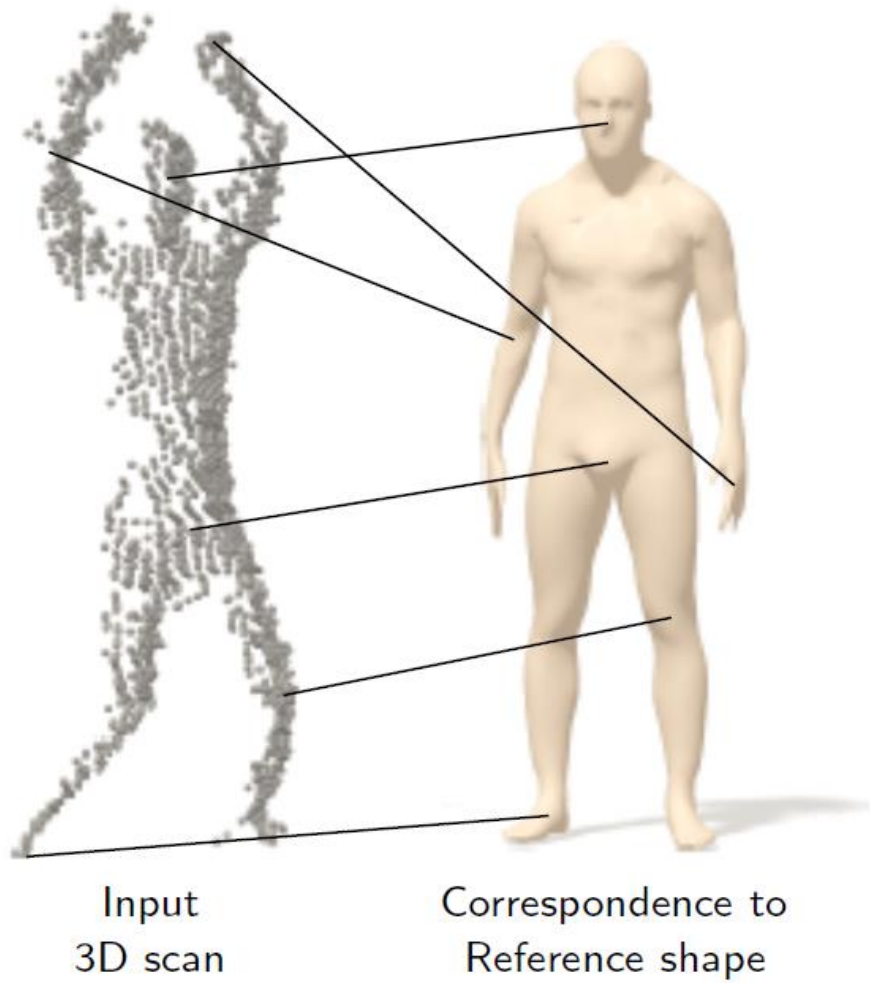


Medicine

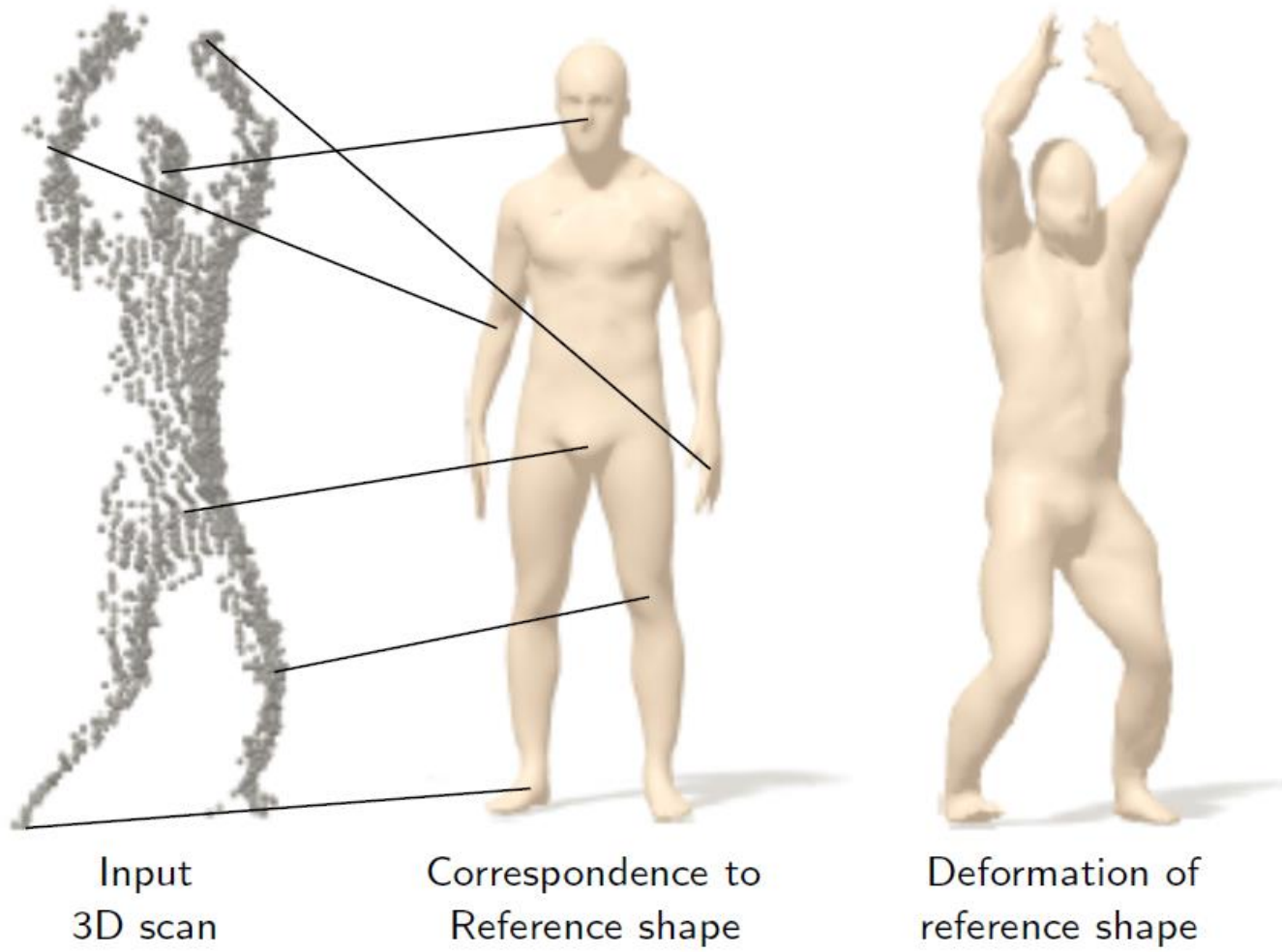


Drug design

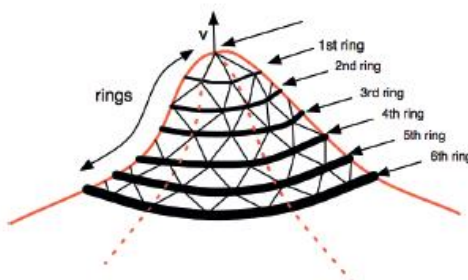
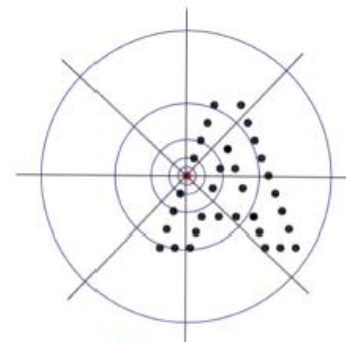
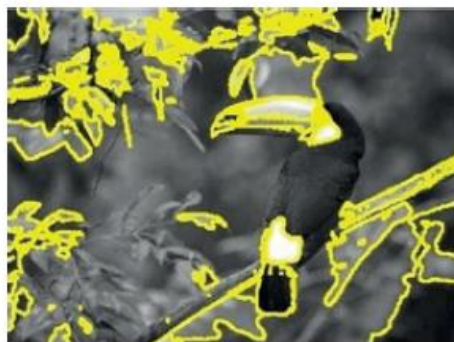
Analysis and synthesis



Analysis and synthesis



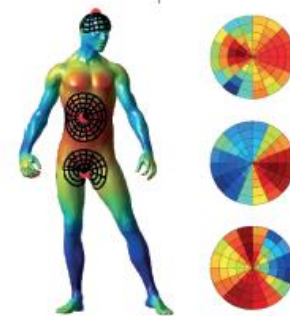
3D feature descriptors



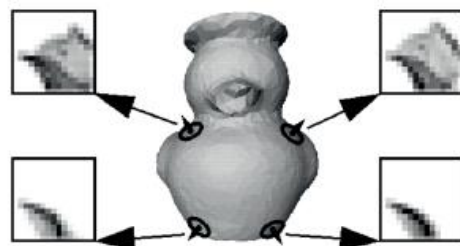
SIFT¹ / MeshHOG²



MSER³ / ShapeMSER⁴



(Intrinsic⁶) Shape context⁵



Spin image⁷



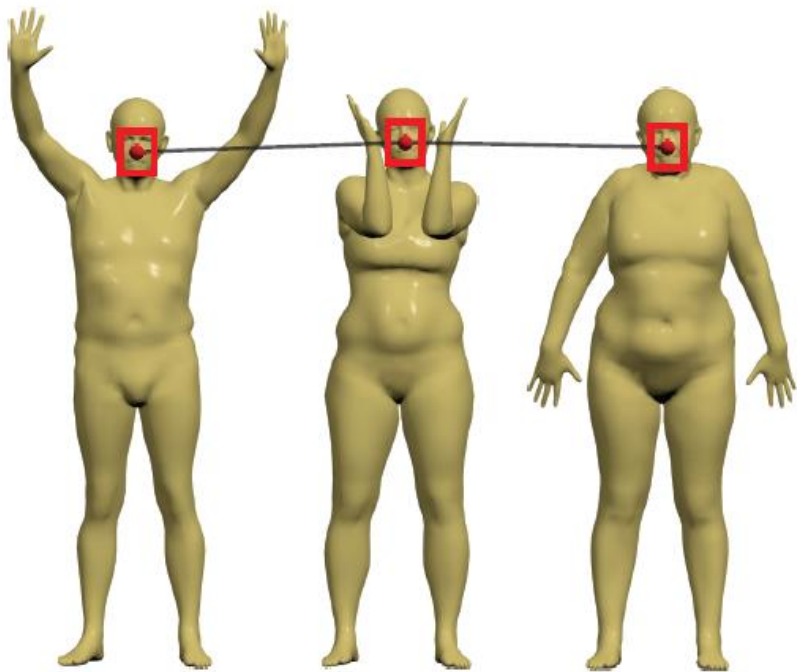
Heat kernel signature⁸

¹Lowe 2004; ²Zaharescu et al. 2009; ³Matas et al. 2002; ⁴Litman et al. 2010;

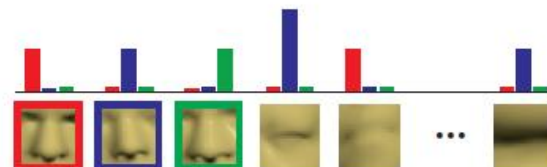
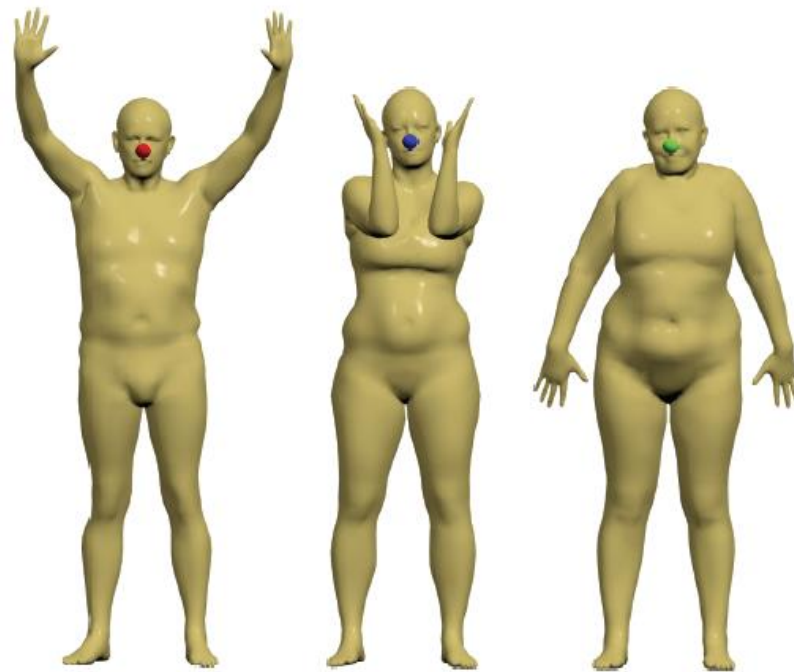
⁵Belongie et al. 2000; ⁶Kokkinos et al. 2012; ⁷Johnson et al. 1999; ⁸Sun et al. 2009

Task-specific features

Correspondence



Similarity



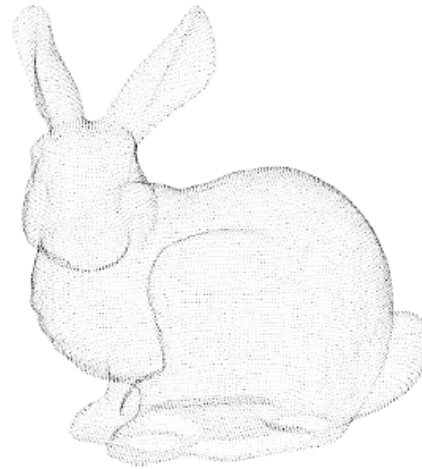
Shape representation



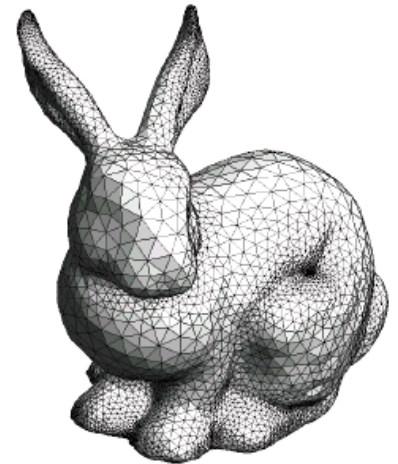
Image-based



Volumetric



Point-based



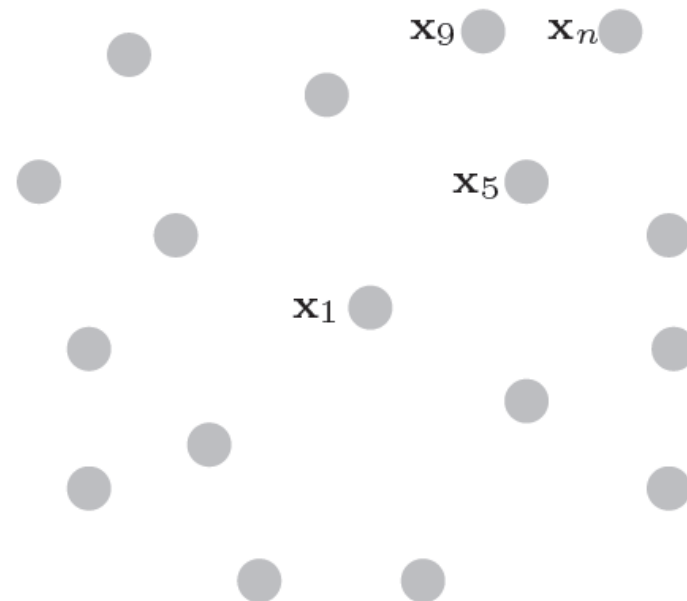
Surface-based

PointNet: learning on sets

- **Permutation-invariant** function

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = f(\mathbf{x}_{\pi_1}, \dots, \mathbf{x}_{\pi_n})$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is feature at vertex i



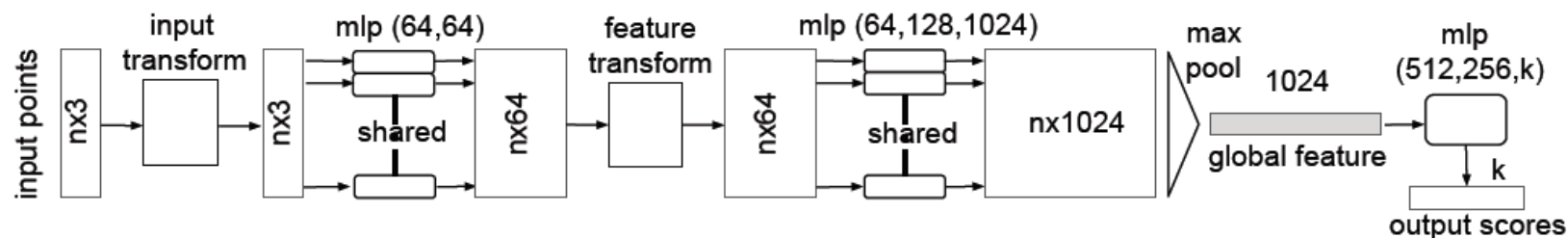
PointNet: learning on sets

- **Permutation-invariant** function

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = f(\mathbf{x}_{\pi_1}, \dots, \mathbf{x}_{\pi_n})$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is feature at vertex i

- **Shared function** $h_{\Theta}(\cdot)$ applied to each point + permutation-invariant aggregation (max or \sum)



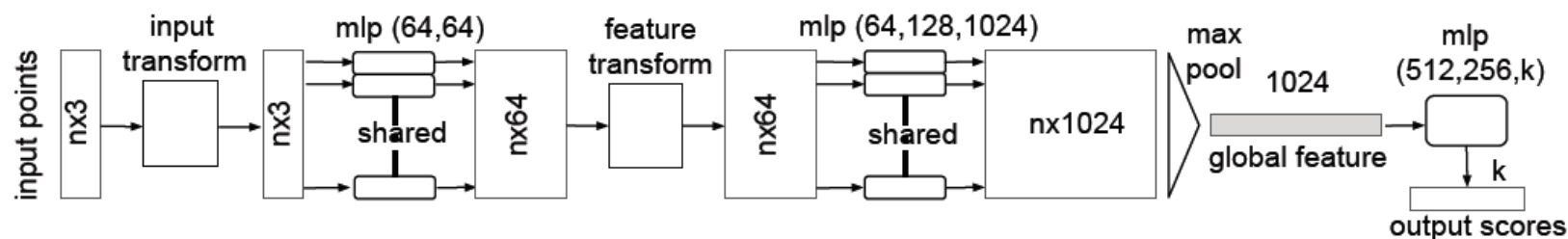
PointNet: learning on sets

- **Permutation-invariant** function

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = f(\mathbf{x}_{\pi_1}, \dots, \mathbf{x}_{\pi_n})$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is feature at vertex i

- **Shared function** $h_{\Theta}(\cdot)$ applied to each point + permutation-invariant aggregation (max or \sum)
- Spatial transformer units
- Local grouping (PointNet++, PCPNet)



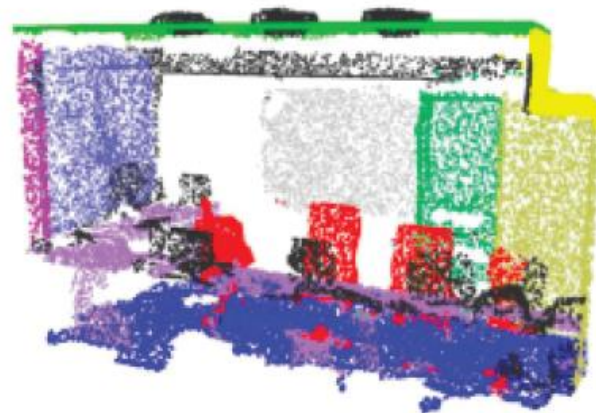
PointNet applications



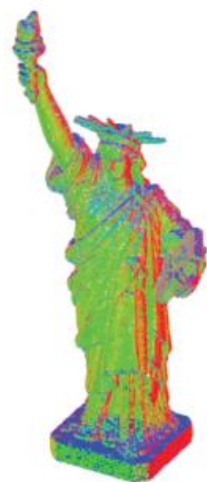
Object recognition



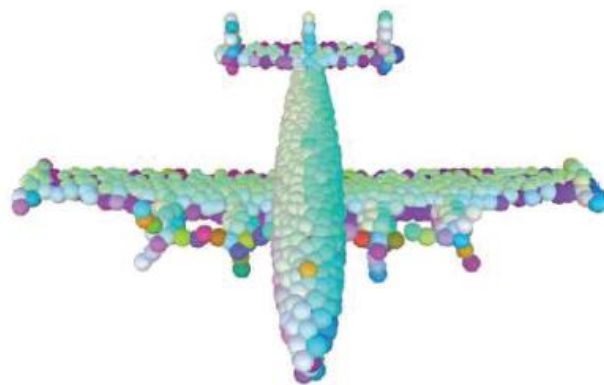
Part segmentation



Semantic segmentation



Curvature



Normals

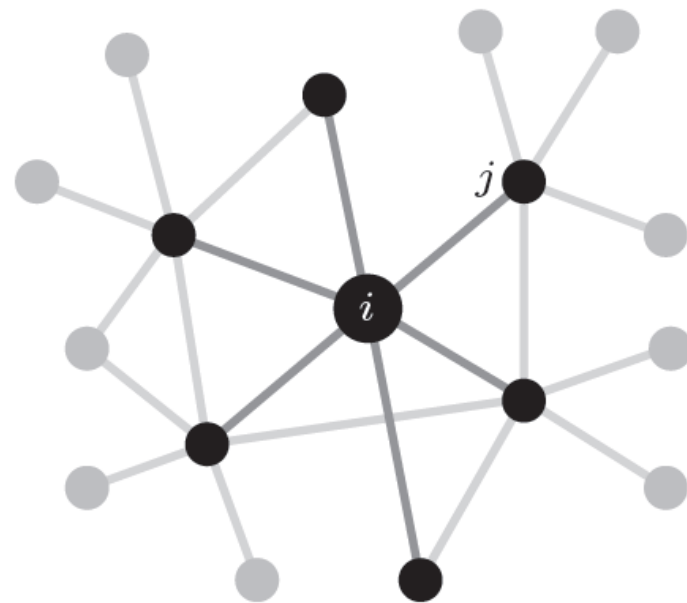


Point cloud generation

Graph-based edge convolution

- Local neighborhood structure modeled as a **graph**
- **Edge feature** function $h_{\Theta}(\cdot, \cdot)$ parametrized by Θ
- Permutation-invariant **aggregation operator** \square (e.g. \sum or \max) on the neighborhood of i
- **Edge convolution** (EdgeConv)

$$\mathbf{x}'_i = \square_j h_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)$$



Learnable local (nonlinear) operator

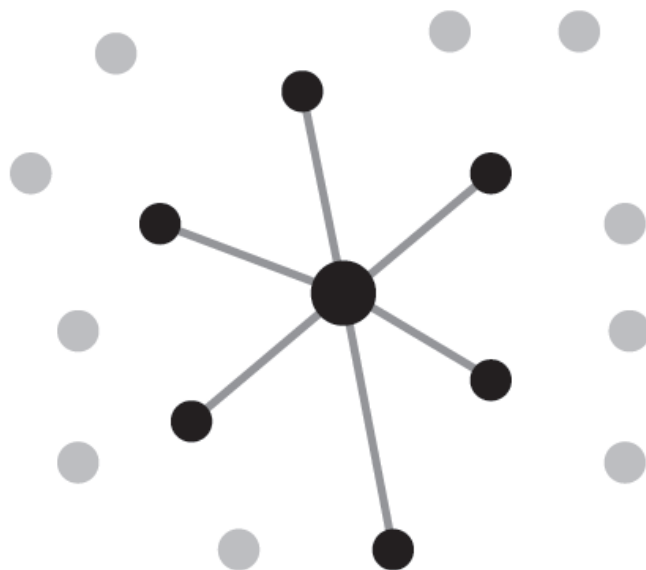
Particular cases

Method	Aggregation \square	Edge feature $h(\mathbf{x}_i, \mathbf{x}_j)$
Laplacian	\sum	$w_{ij}(\mathbf{x}_j - \mathbf{x}_i)$
PointNet ¹	–	$h(\mathbf{x}_i)$
PointNet++ ²	max	$h(\mathbf{x}_i)$
MoNet ³	\sum	$\sum_{\ell} g_{\ell} w_{\ell}(\mathbf{u}_{ij}) \mathbf{x}_j$
PCNN ⁴	\sum	$\sum_{\ell m} c(\mathbf{x}_i \cdot \mathbf{k}_{\ell m}) w_i q_{\Theta_{\ell}}(\mathbf{x}_i, \mathbf{x}_j)$

Wang et al. 2018; ¹Qi et al. 2017; ²Qi, Su et al. 2017; ³Monti et al. 2017; ⁴Atzmon et al. 2018

Dynamic Graph CNN (DynGCNN)

Construct k -NN graph in feature space and **update it after each layer**

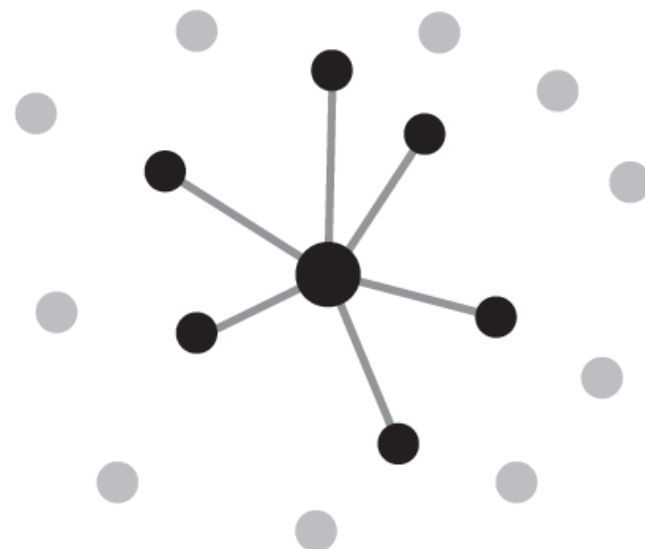


Layer l

Features $\mathbf{x}_1^{(l)}, \dots, \mathbf{x}_n^{(l)} \in \mathbb{R}^{d_l}$

k -NN graph $\mathcal{G}^{(l)}$

$h^{(l)} : \mathbb{R}^{d_l} \times \mathbb{R}^{d_l} \rightarrow \mathbb{R}^{d_{l+1}}$



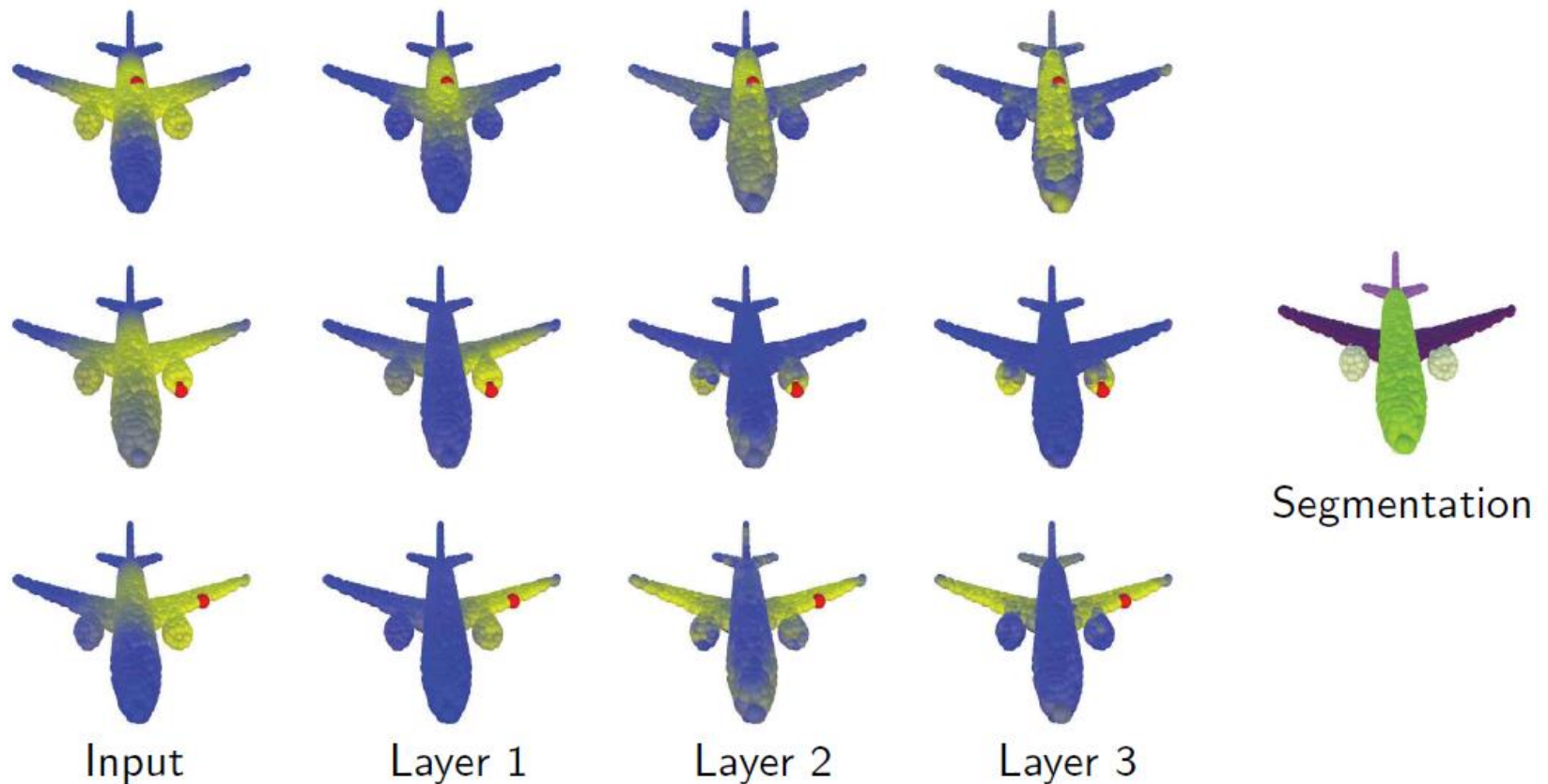
Layer $l + 1$

Features $\mathbf{x}_1^{(l+1)}, \dots, \mathbf{x}_n^{(l+1)} \in \mathbb{R}^{d_{l+1}}$

k -NN graph $\mathcal{G}^{(l+1)}$

$h^{(l+1)} : \mathbb{R}^{d_{l+1}} \times \mathbb{R}^{d_{l+1}} \rightarrow \mathbb{R}^{d_{l+2}}$

Learning semantic features



Left: Distance from red point in the feature space of different DynGCNN layers
Right: semantic segmentation results

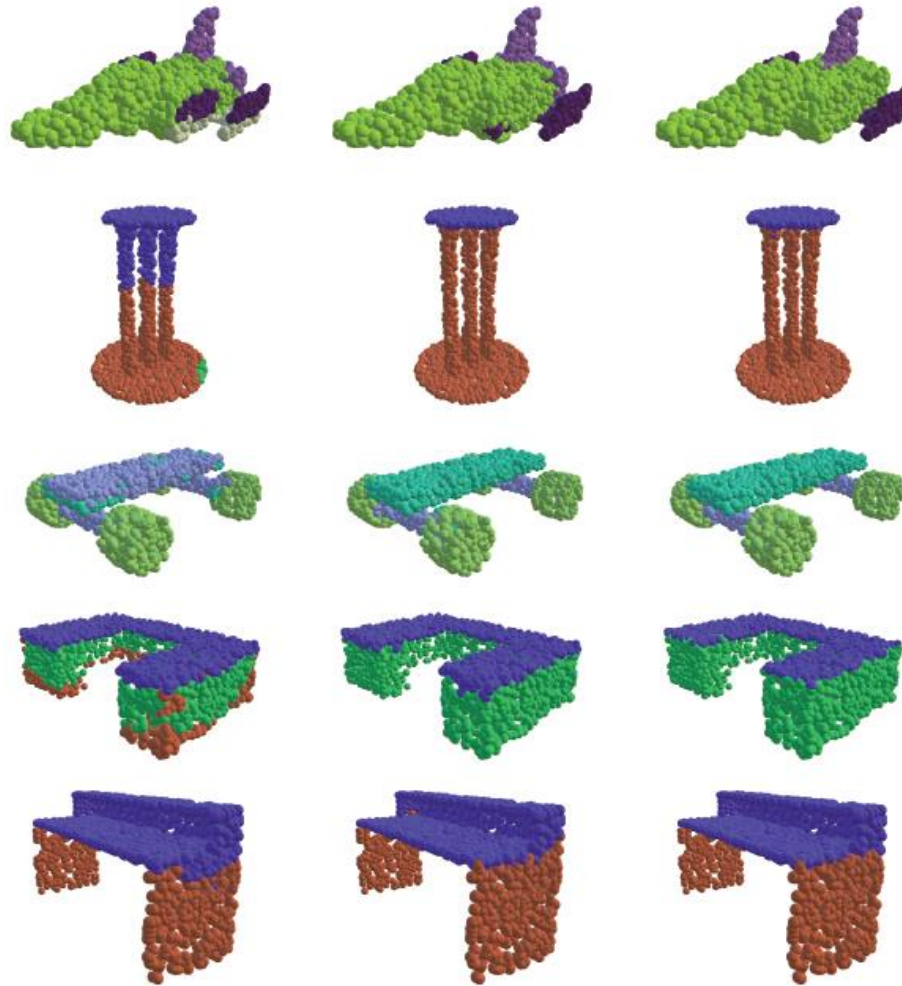
Shape classification (ModelNet40)

Method	Mean class accuracy	Overall accuracy
3DShapeNet ¹	77.3%	84.7%
VoxNet ²	83.0%	85.9%
Subvolume ³	86.0%	89.2%
ECC ⁴	83.2%	87.4%
PointNet ⁵	86.0%	89.2%
PointNet++ ⁶	—	90.7%
Kd-Net ⁷	—	91.8%
DynGCNN (baseline) ⁸	88.8%	91.2%
DynGCNN⁸	90.2%	92.2%

Classification accuracy of different methods on ModelNet40

Methods: ¹Wu et al. 2015; ²Maturana et al. 2015; Qi et al. 2016; ⁴Simonovsky, Komodakis 2017; ⁵Qi et al. 2017; ⁶Qi, Su et al. 2017; ⁷Klokov, Lempitsky 2017; ⁸Wang et al. 2018; data: Wu et al. 2015 (ModelNet)

Semantic segmentation: synthetic (ShapeNet)



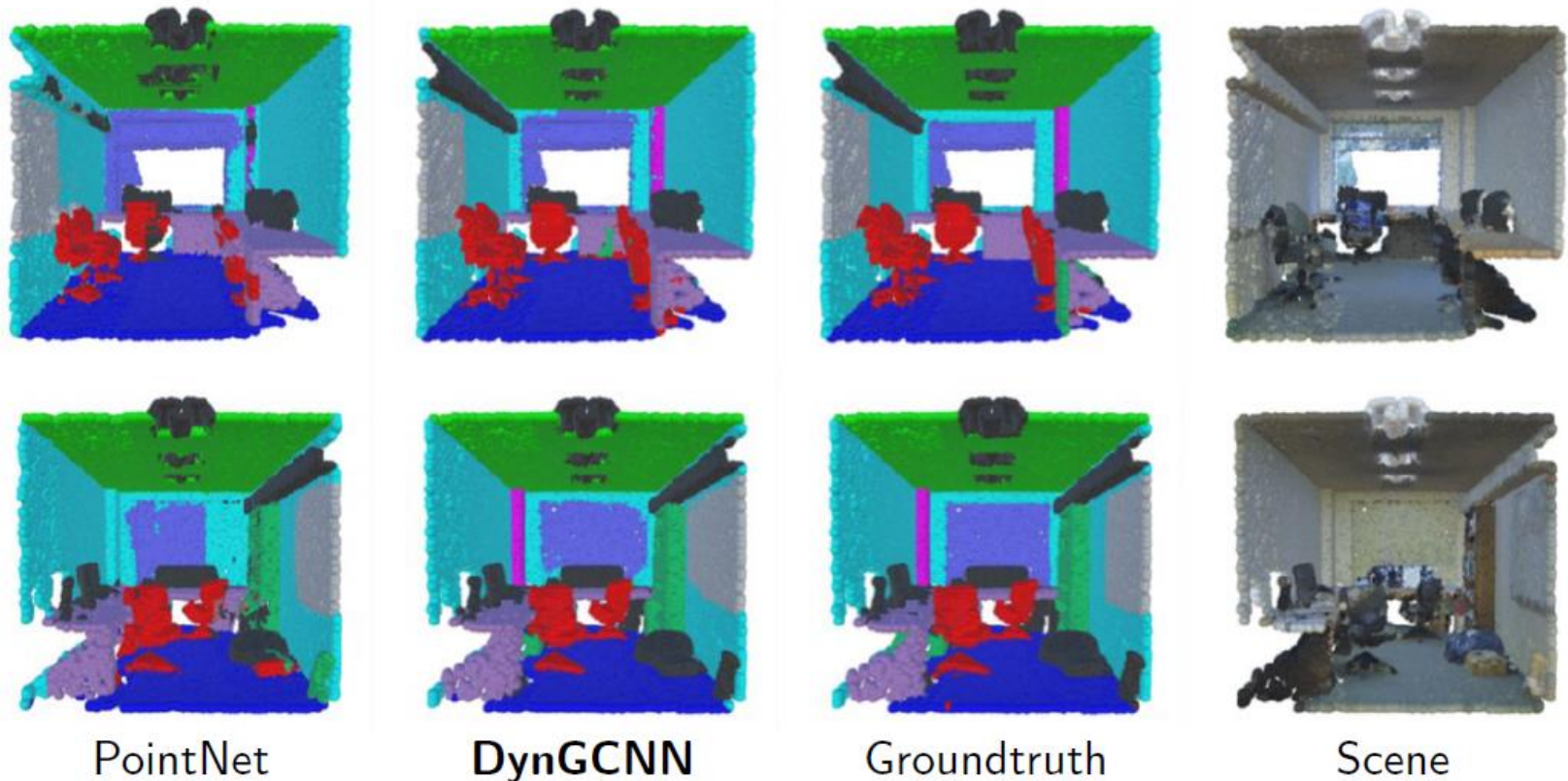
PointNet

DynGCNN

Groundtruth

Methods: Qi et al. 2017 (PointNet); Wang et al. 2018 (DynGCNN); data: Yi et al. 2016 (ShapeNet)

Semantic segmentation: indoor scans (S3DIS)



Results of semantic segmentation of point cloud+RGB data
using different architectures

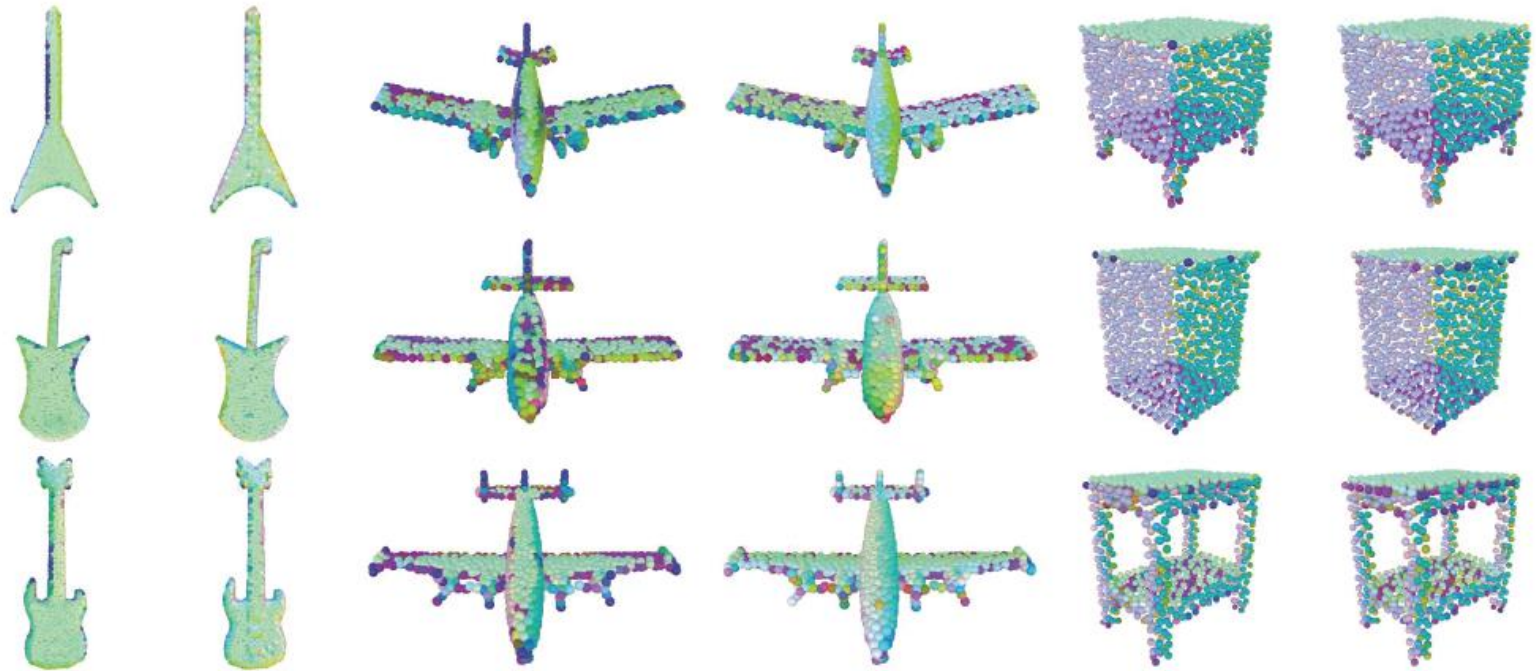
Methods: Qi et al. 2017 (PointNet); Wang et al. 2018 (DynGCNN); data: Armeni et al. 2016 (S3DIS)

Method	Mean IoU	Overall accuracy
PointNet (Baseline) ¹	20.1%	53.2%
PointNet ¹	47.6%	78.5%
MS + CU(2) ²	47.8%	79.2%
G + RCU ²	49.7%	81.1%
DynGCNN³	56.1%	84.1%

S3DIS indoor scene semantic segmentation accuracy

Methods: ¹Qi et al. 2017; ²Engelmann et al. 2017 ³Wang et al. 2018; data: Armeni et al. 2016 (S3DIS)

Surface normal prediction



Surface normal predicted using DynGCNN (odd columns) and groundtruth (even columns). Normal direction is color-coded

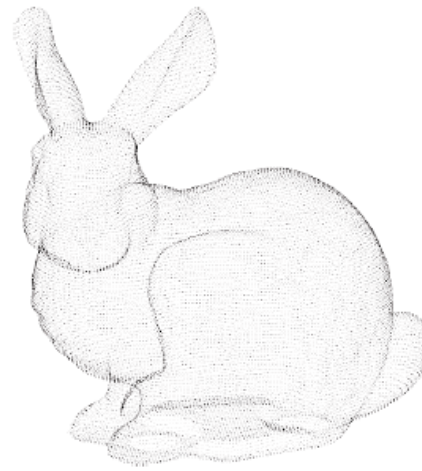
Shape representation



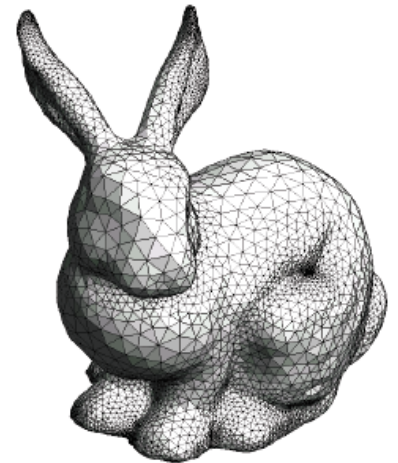
Image-based



Volumetric



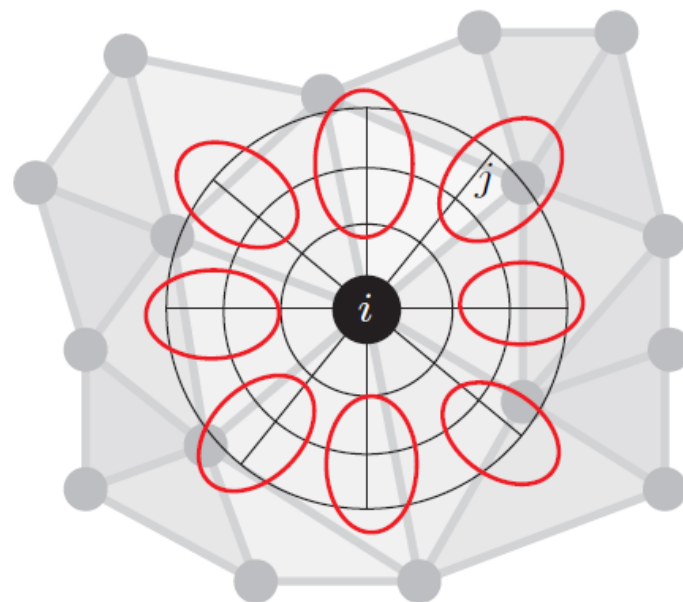
Point-based



Surface-based

Convolution on meshes

- Local system of coordinates \mathbf{u}_{ij} around i (e.g. geodesic polar)
- Local weights $w_1(\mathbf{u}), \dots, w_L(\mathbf{u})$ w.r.t. \mathbf{u} , e.g. Gaussians
 $w_\ell = \exp\left(-(\mathbf{u} - \boldsymbol{\mu}_\ell)^\top \boldsymbol{\Sigma}_\ell^{-1}(\mathbf{u} - \boldsymbol{\mu}_\ell)\right)$

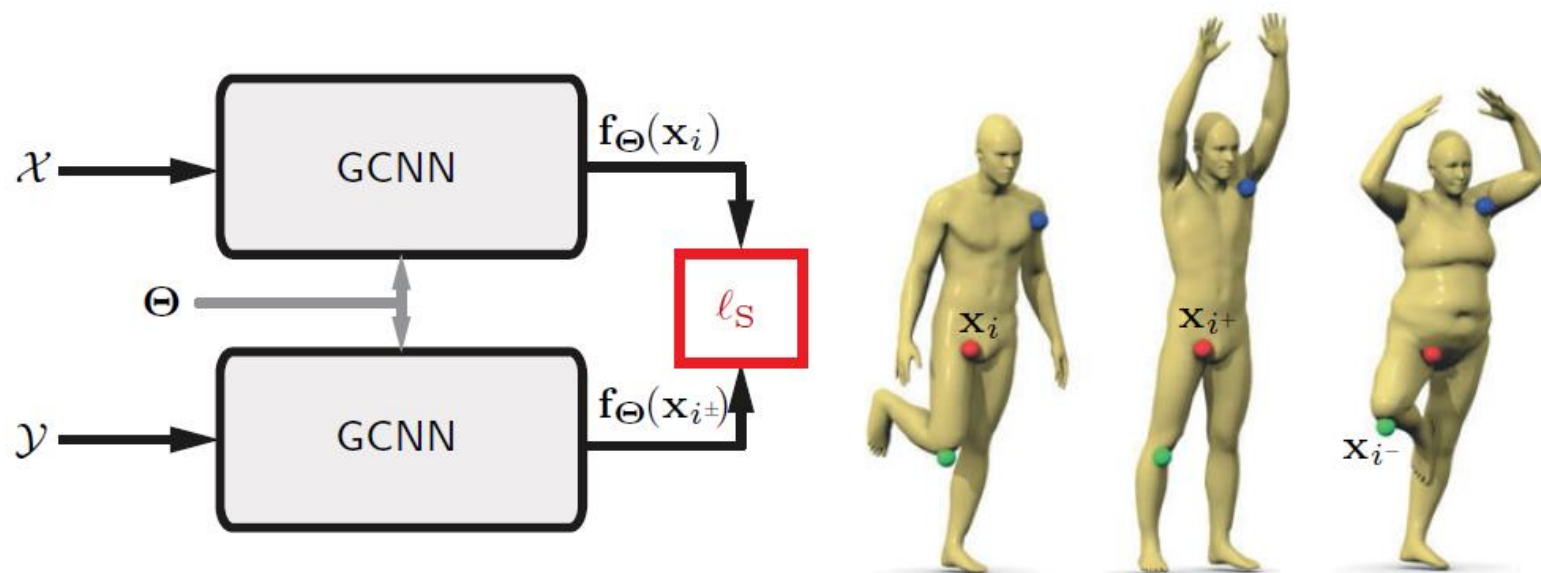


- Spatial convolution with filter g

$$\mathbf{x}'_i \propto \sum_{\ell=1}^L g_\ell \underbrace{\sum_{j=1}^n w_\ell(\mathbf{u}_{ij}) \mathbf{x}_j}_{\text{patch operator}}$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is feature at vertex i

Learning local descriptors with intrinsic CNN



Training set

Siamese net

Pointwise feature cost

positive (i, i^+) and negative (i, i^-) pairs of points

two net instances with shared parameters Θ

$$\begin{aligned} \ell_S(\Theta) = & \gamma \sum_{i, i^+} \|\mathbf{f}_{\Theta}(\mathbf{x}_i) - \mathbf{f}_{\Theta}(\mathbf{x}_{i^+})\|_2^2 \\ & + (1 - \gamma) \sum_{i, i^-} [\mu - \|\mathbf{f}_{\Theta}(\mathbf{x}_i) - \mathbf{f}_{\Theta}(\mathbf{x}_{i^-})\|_2^2]_+ \end{aligned}$$

HKS descriptor



Distance in the space of local Heat Kernel Signature (HKS) features
(shown is distance from a point on the shoulder marked in white)

Descriptor: Sun, Ovsjanikov, Guibas 2009 (HKS); data: Bronstein, Bronstein, Kimmel 2008 (TOSCA); Anguelov et al. 2005 (SCAPE); Bogo et al. 2014 (FAUST)

WKS descriptor



Distance in the space of local Wave Kernel Signature (WKS) features
(shown is distance from a point on the shoulder marked in white)

Descriptor: Aubry, Schlickewei, Cremers 2011 (WKS); data: Bronstein, Bronstein, Kimmel 2008 (TOSCA); Anguelov et al. 2005 (SCAPE); Bogo et al. 2014 (FAUST)

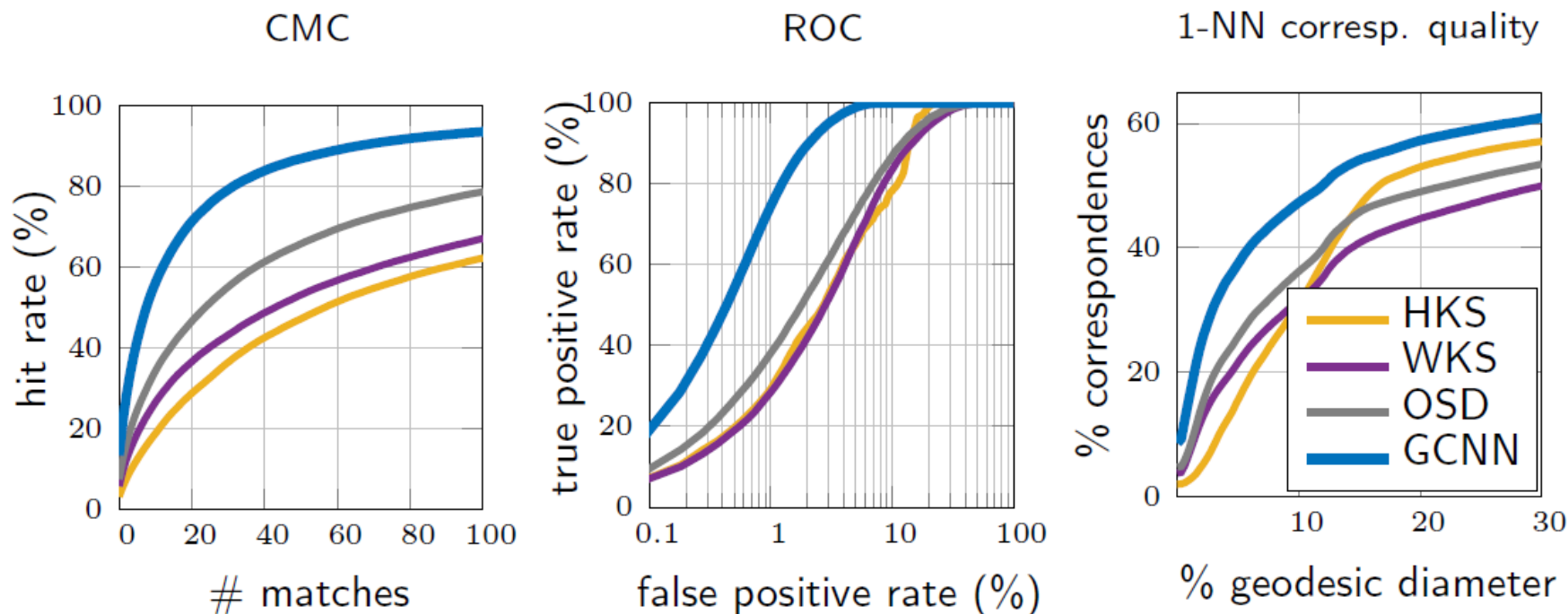
Descriptor learned with GCNN



Distance in the space of local GCNN features
(shown is distance from a point on the shoulder marked in white)

Descriptor: Masci et al. 2015 (GCNN); data: Bronstein, Bronstein, Kimmel 2008 (TOSCA); Anguelov et al. 2005 (SCAPE); Bogo et al. 2014 (FAUST)

Descriptor quality comparison



Descriptor performance using symmetric Princeton benchmark
(training and testing: disjoint subsets of FAUST)

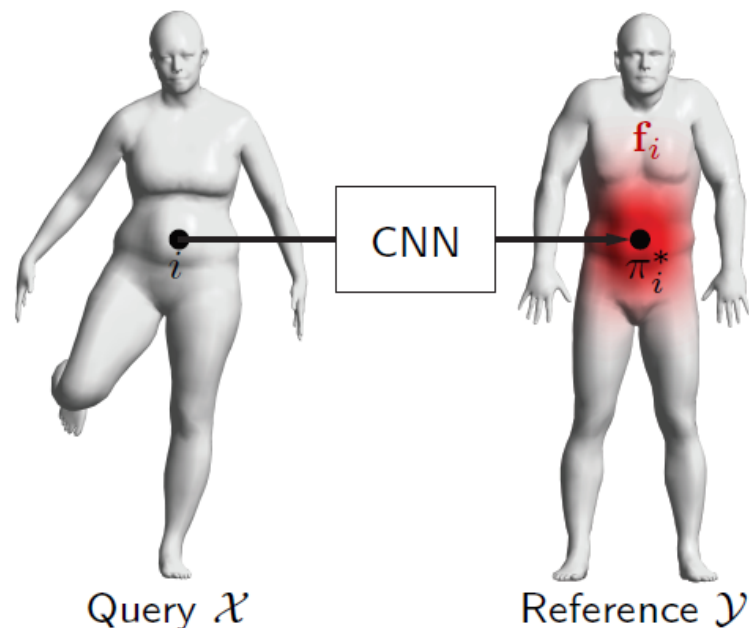
Methods: Sun et al. 2009 (HKS); Aubry et al. 2011 (WKS); Litman, B 2014 (OSD); Masci et al. 2015 (GCNN); data: Bogo et al. 2014 (FAUST); benchmark: Kim et al. 2011

Learning deformation-invariant correspondence

- Groundtruth correspondence
 $\pi^* : \mathcal{X} \rightarrow \mathcal{Y}$ from query shape \mathcal{X}
to some **reference shape** \mathcal{Y}
- Correspondence = **label** each query
vertex $i \in \{1, \dots, n\}$ as reference
vertex $\pi_i \in \{1, \dots, m\}$
- Net output at i after softmax layer

$$\mathbf{f}_{\Theta}(\mathbf{x}_i) = (f_{i1}, \dots, f_{im})$$

= probability distribution on \mathcal{Y}



Minimize on training set the **cross entropy** between groundtruth correspondence and output probability distribution w.r.t. net parameters Θ

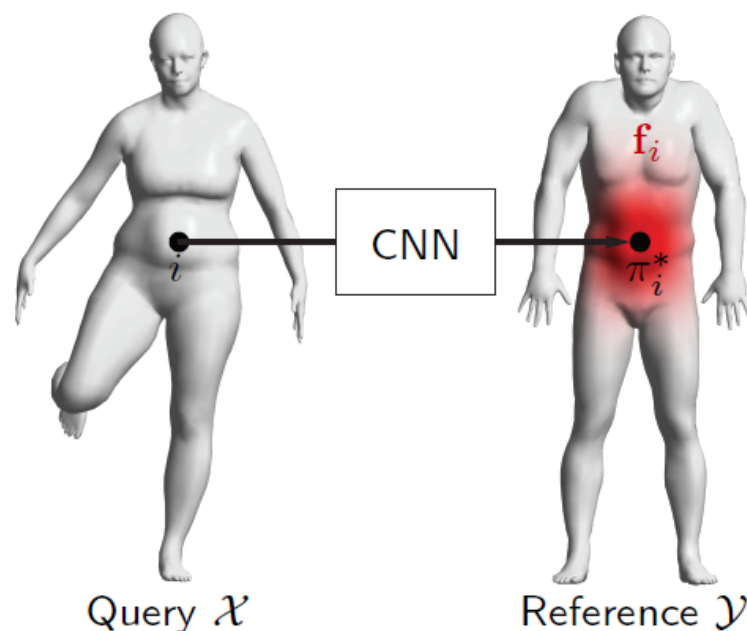
$$\min_{\Theta} \sum_{i=1}^n H(\delta_{\pi_i^*}, \mathbf{f}_{\Theta}(\mathbf{x}_i))$$

Learning deformation-invariant correspondence

- Groundtruth correspondence
 $\pi^* : \mathcal{X} \rightarrow \mathcal{Y}$ from query shape \mathcal{X}
to some **reference shape** \mathcal{Y}
- Correspondence = **label** each query
vertex $i \in \{1, \dots, n\}$ as reference
vertex $\pi_i \in \{1, \dots, m\}$
- Net output at i after softmax layer

$$\mathbf{f}_{\Theta}(\mathbf{x}_i) = (f_{i1}, \dots, f_{im})$$

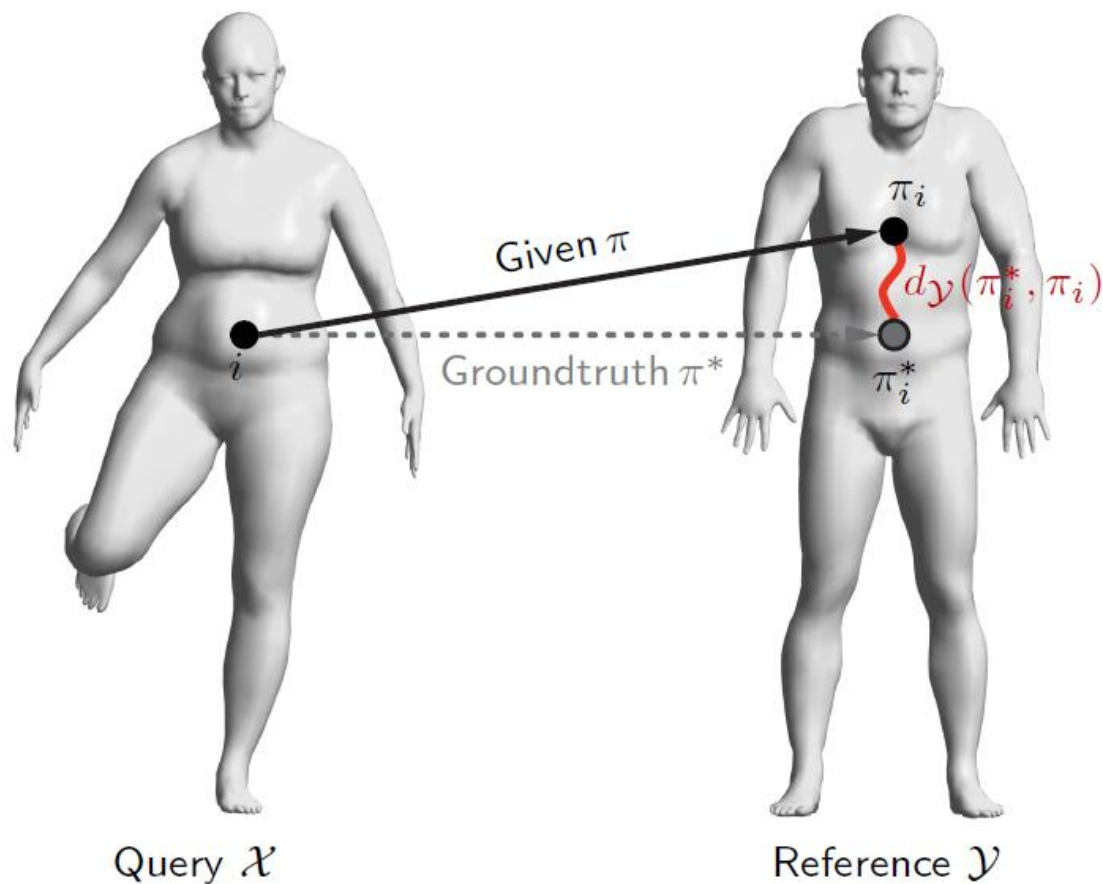
= probability distribution on \mathcal{Y}



Minimize on training set the **cross entropy** between groundtruth correspondence and output probability distribution w.r.t. net parameters Θ

$$\min_{\Theta} \sum_{i=1}^n H(\delta_{\pi_i^*}, \mathbf{f}_{\Theta}(\mathbf{x}_i))$$

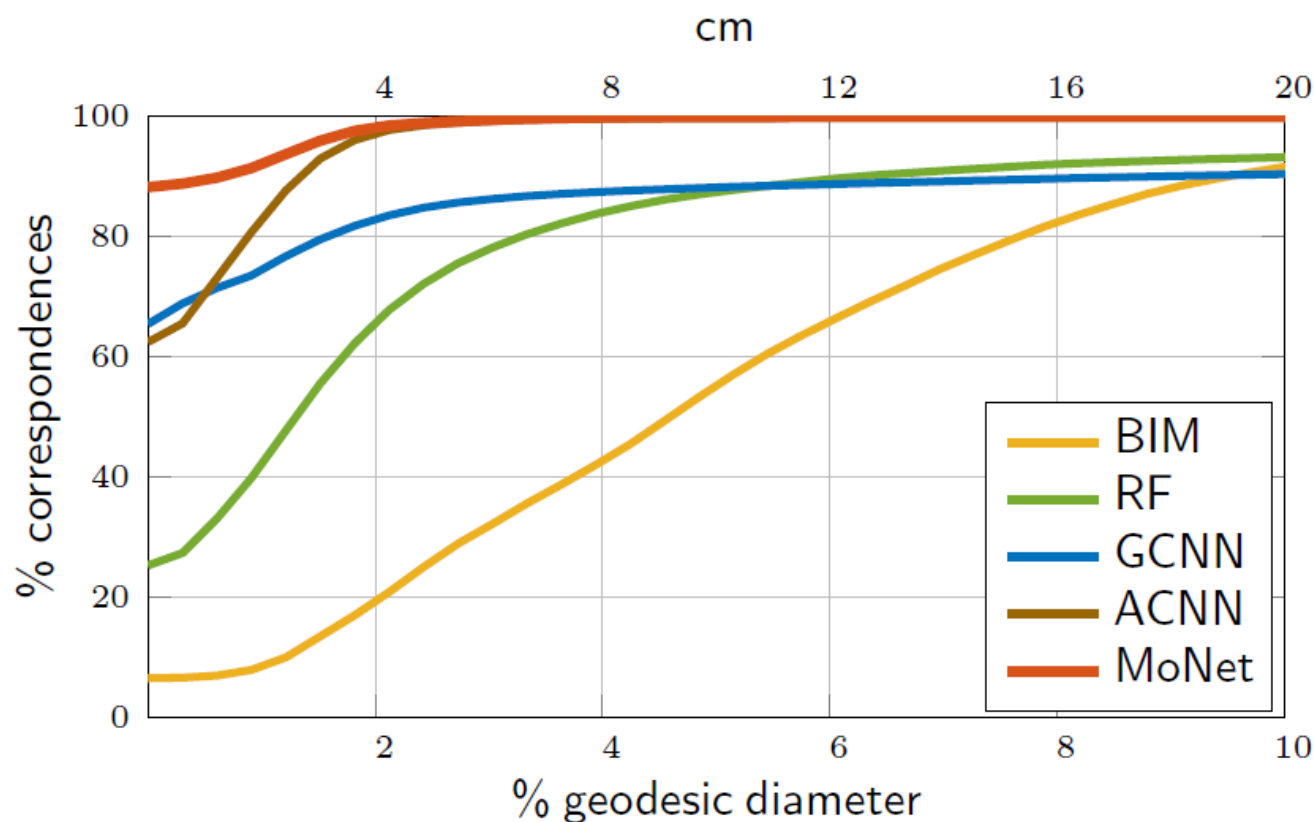
Correspondence evaluation: Princeton benchmark



Pointwise correspondence error = geodesic distance from the groundtruth

$$\epsilon_i = d_{\mathcal{Y}}(\pi_i^*, \pi_i)$$

Correspondence quality comparison



Correspondence evaluated using asymmetric Princeton benchmark
(training and testing: disjoint subsets of FAUST)

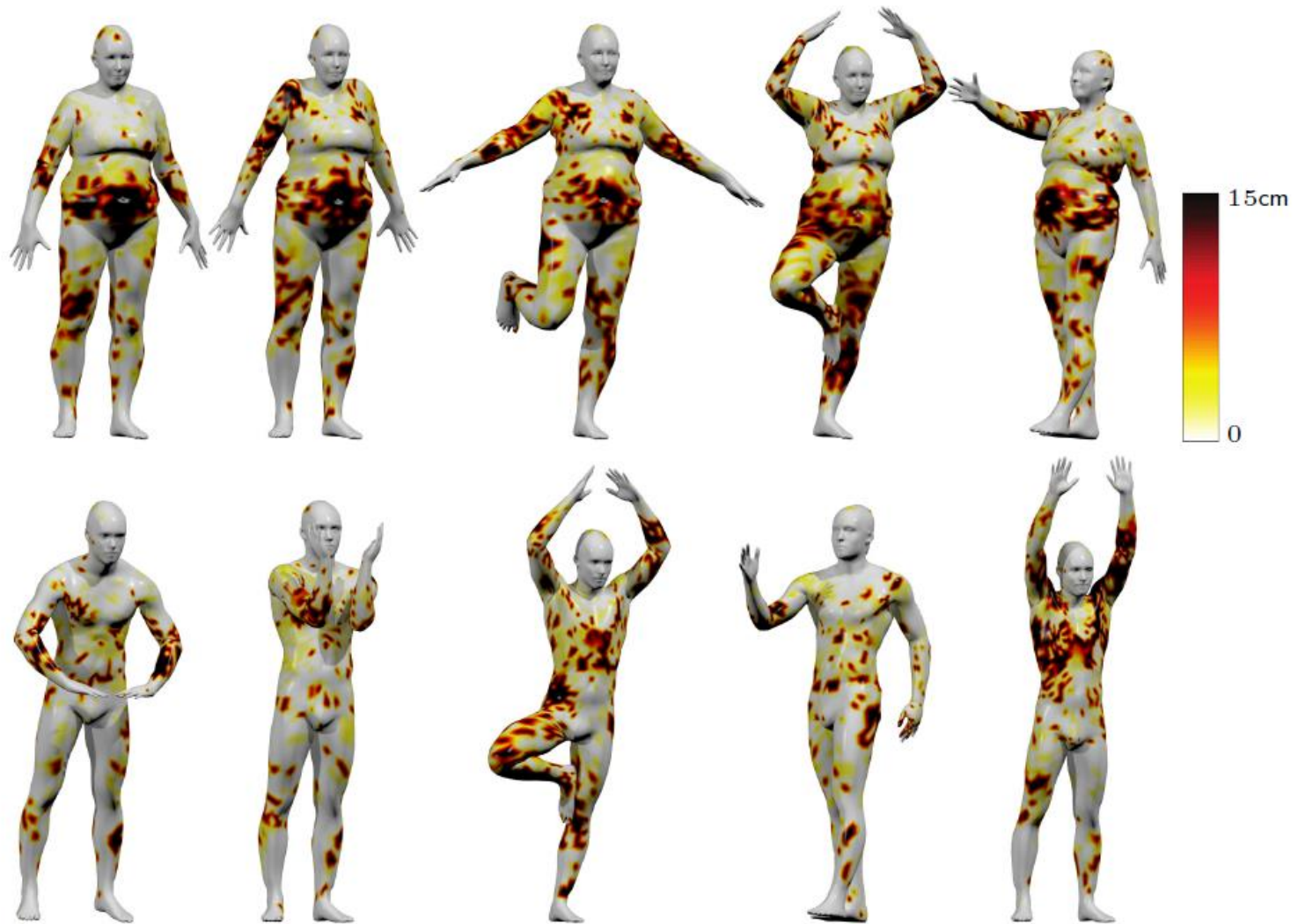
Methods: Kim et al. 2011 (BIM); Rodolà et al. 2014 (RF); Boscaini et al. 2015 (ADD); Masci et al. 2015 (GCNN); Boscaini et al. 2016 (ACNN); Monti et al. 2016 (MoNet); data: Bogo et al. 2014 (FAUST); benchmark: Kim et al. 2011

Shape correspondence error: Blended Intrinsic Map



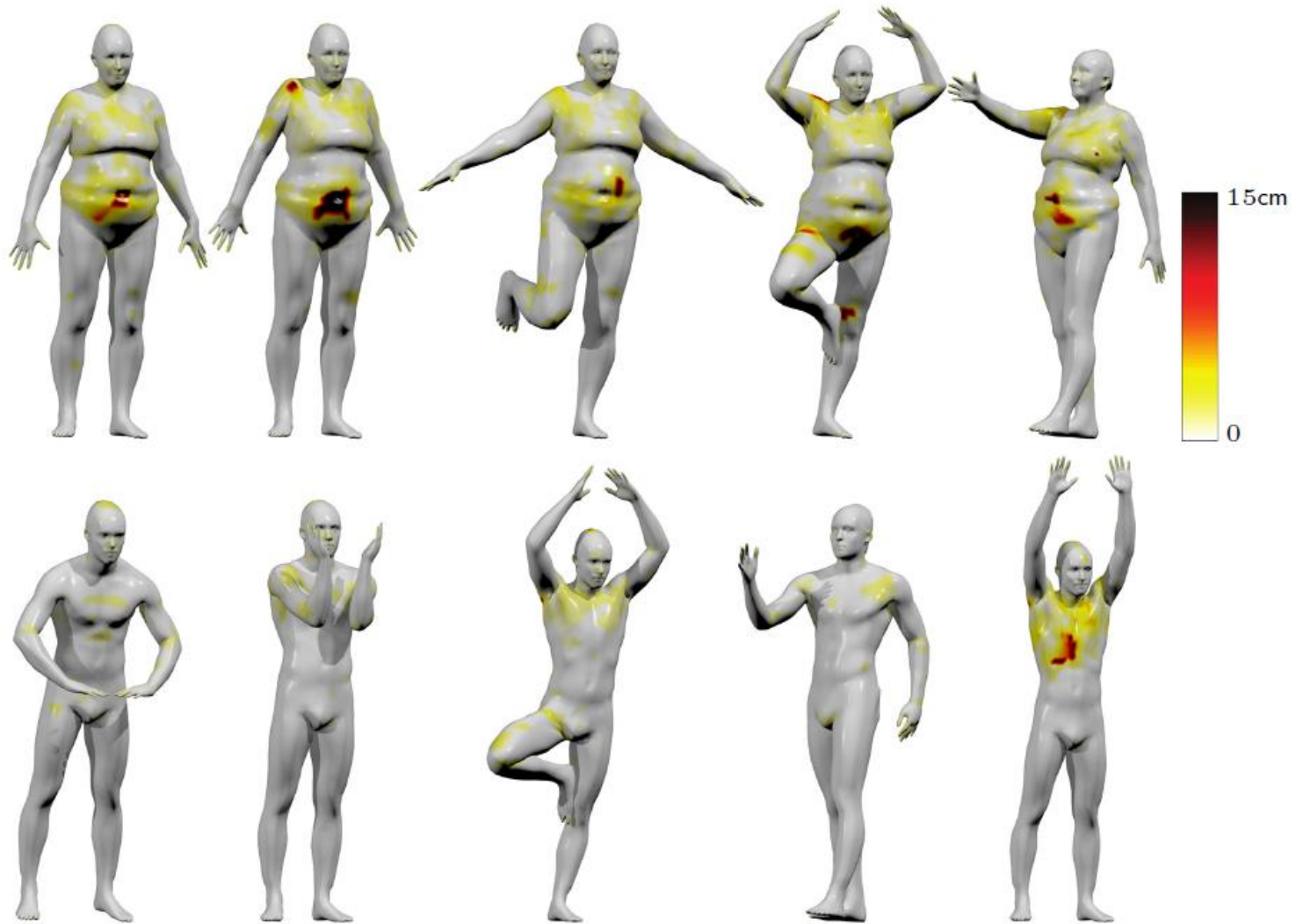
Pointwise correspondence error (geodesic distance from groundtruth)

Shape correspondence error: Geodesic CNN



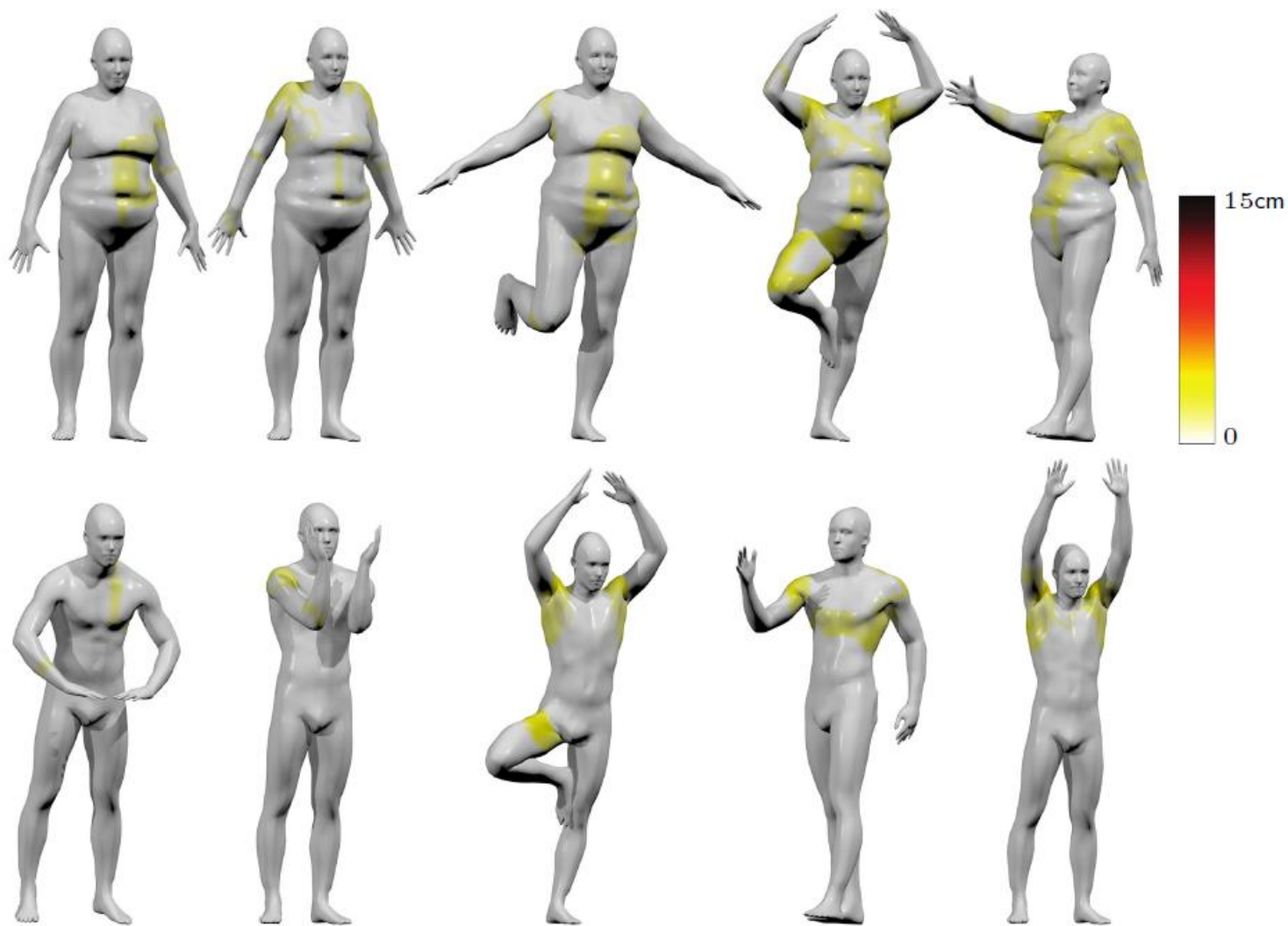
Pointwise correspondence error (geodesic distance from groundtruth)

Shape correspondence error: Anisotropic CNN



Pointwise correspondence error (geodesic distance from groundtruth)

Shape correspondence error: MoNet



Pointwise correspondence error (geodesic distance from groundtruth)

Shape correspondence visualization: MoNet

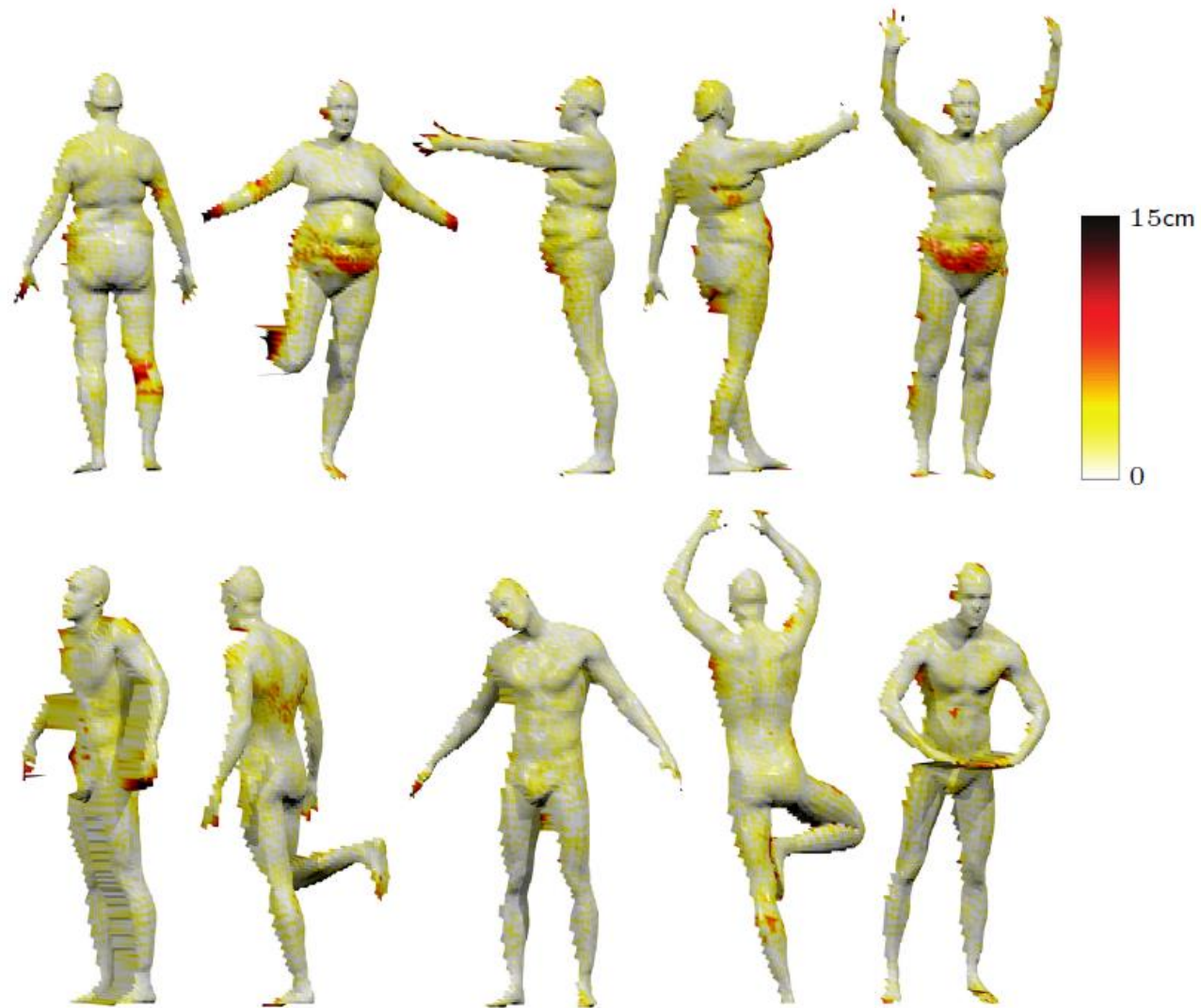


Reference



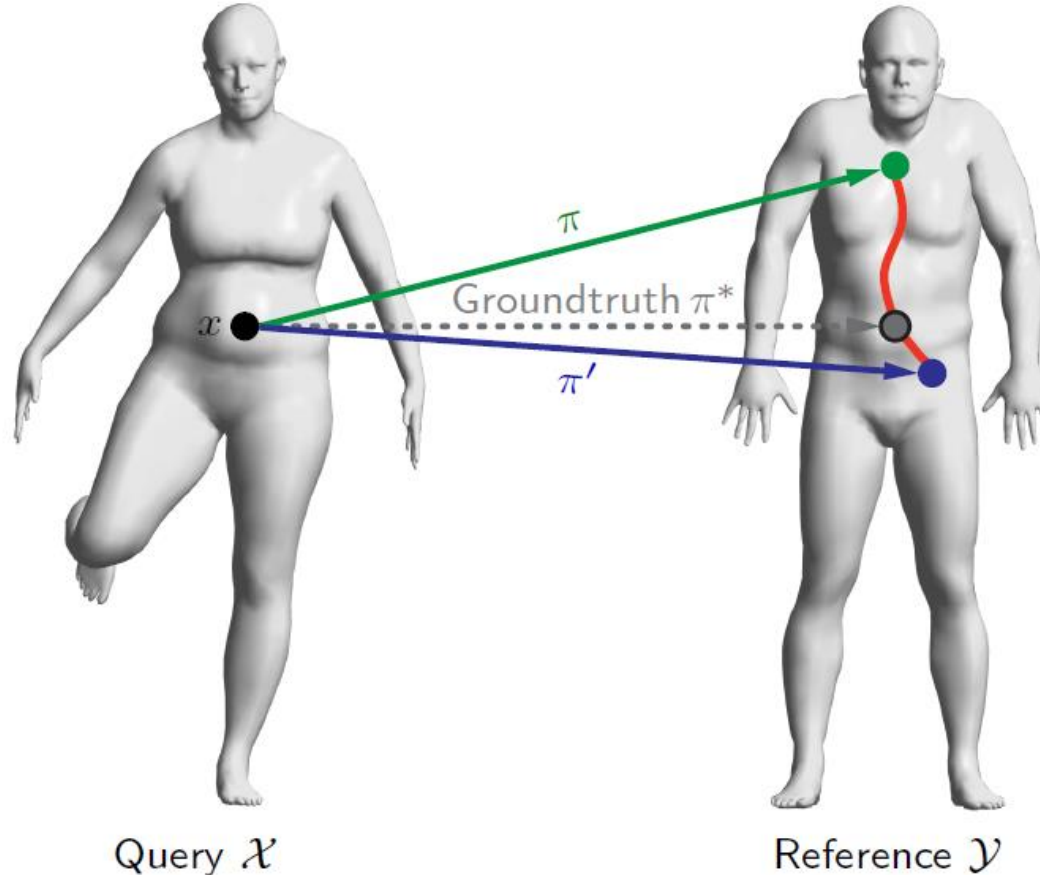
Texture transferred from reference to query shapes

Correspondence on range images: MoNet



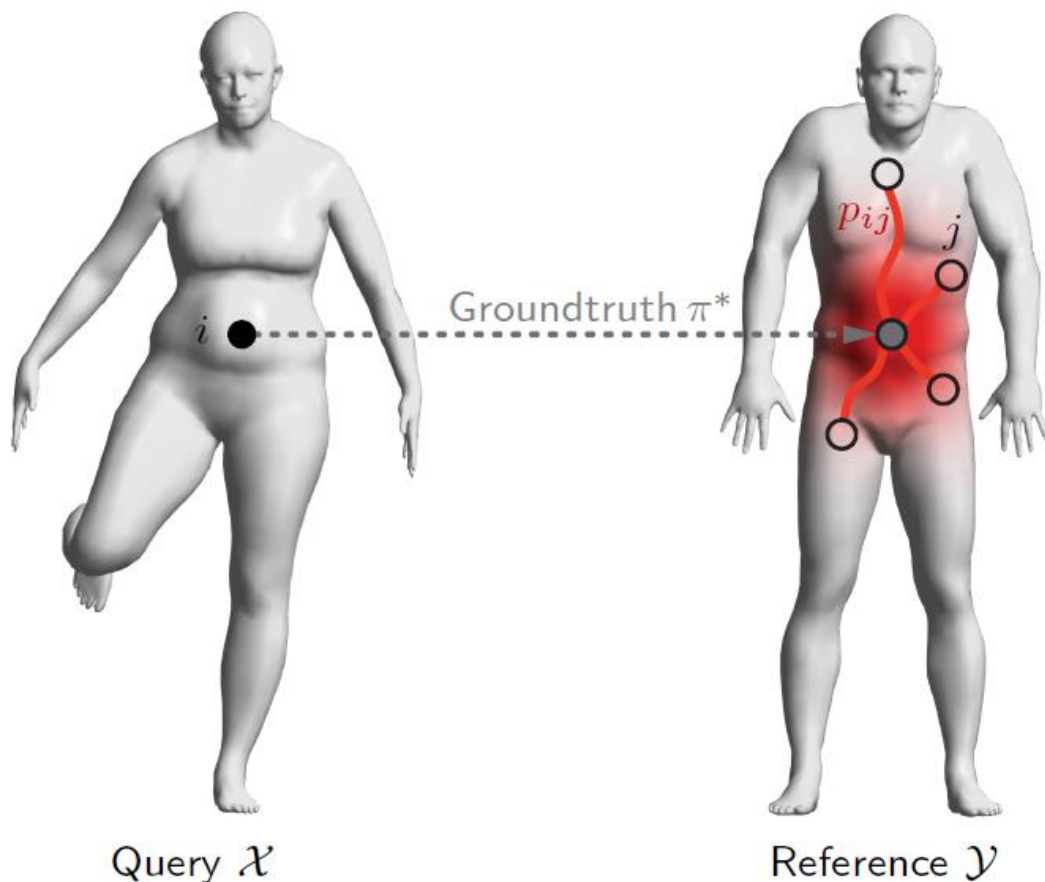
Pointwise correspondence error (geodesic distance from groundtruth)

Correspondence as classification problem, revisited



Classification cost considers equally correspondences that deviate from the groundtruth (no matter how far)

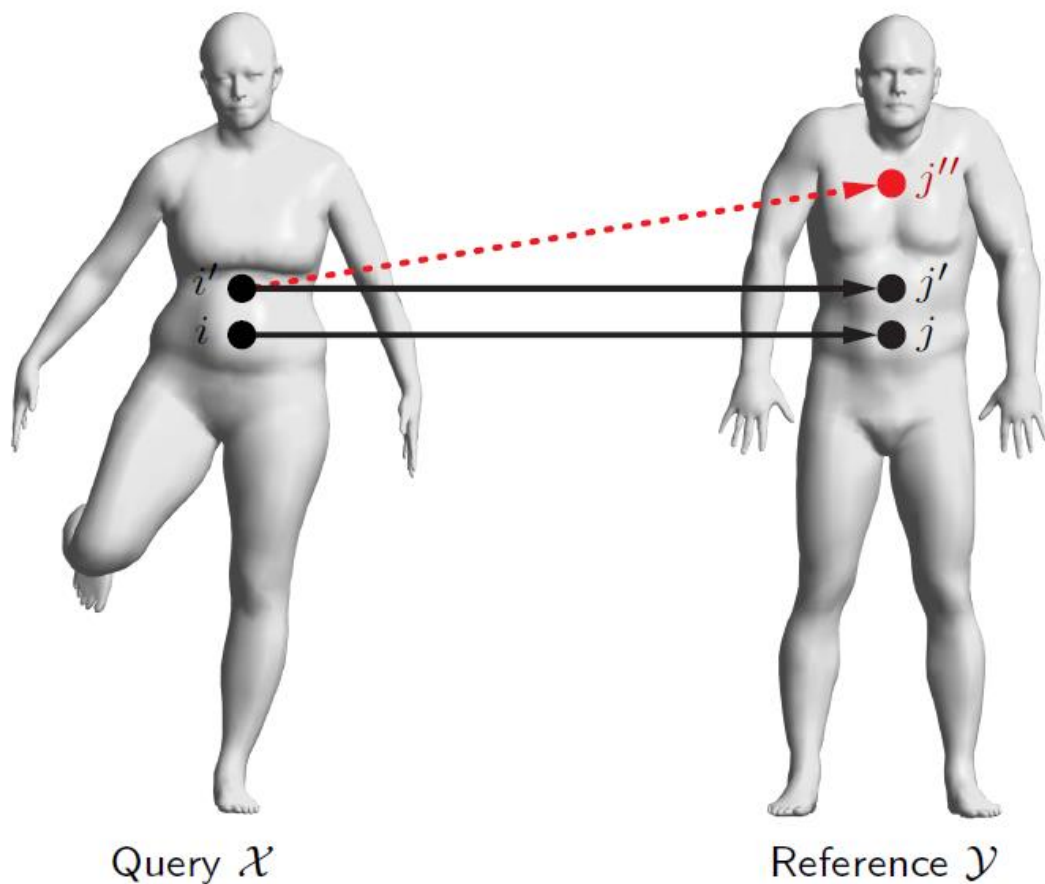
Soft correspondence error



Soft correspondence error = probability-weighted geodesic distance from the groundtruth

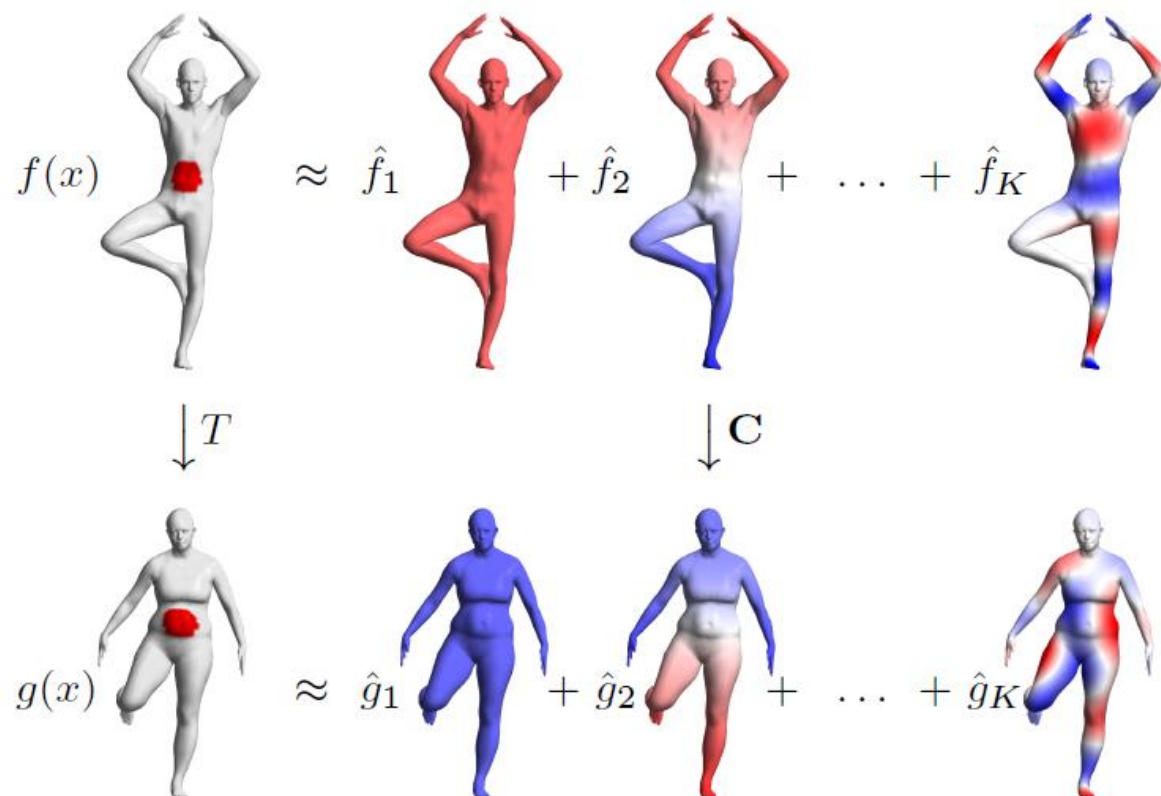
$$\bar{\epsilon}_i = \sum_{j=1}^m p_{ij} d_{\mathcal{Y}}(\pi_i^*, j)$$

Pointwise vs Structured learning



Nearby points i, i' on query shape are **not guaranteed** to map to nearby points j, j' on reference shape at **test time**

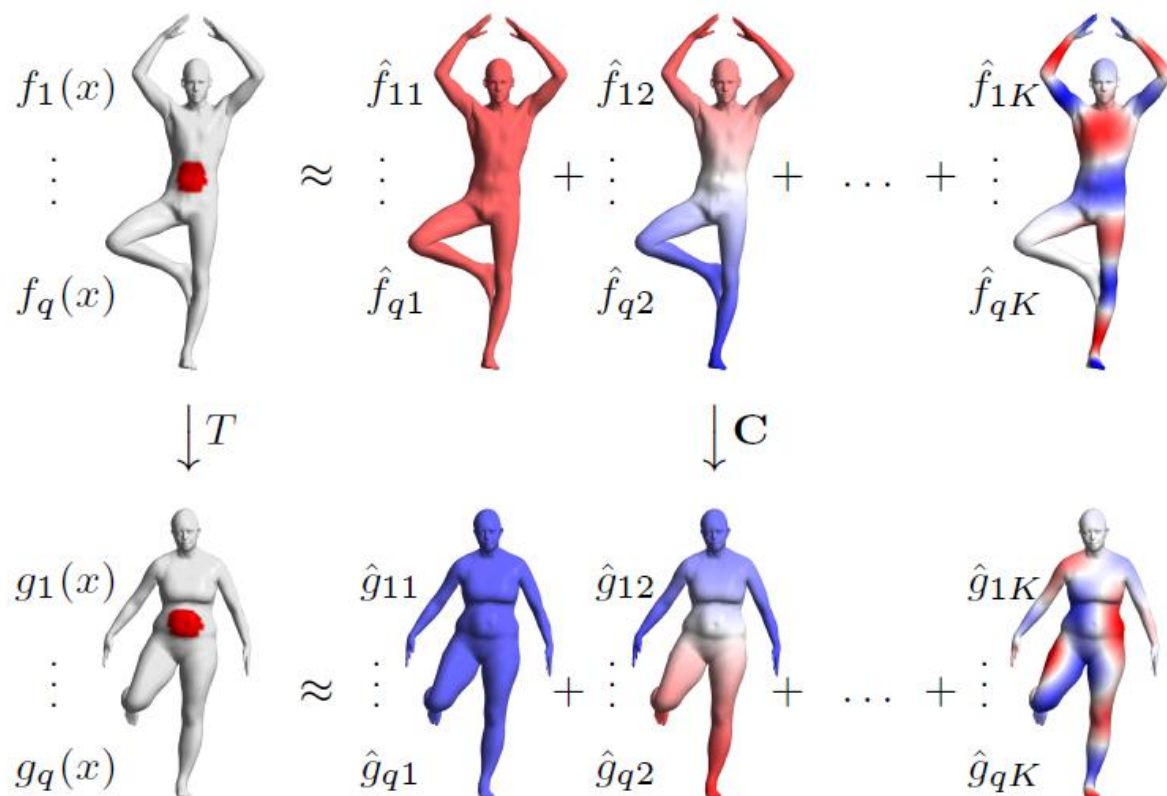
Functional maps: spectral domain



Functional correspondence $T =$ linear map C between Fourier coefficients

$$\hat{\mathbf{g}}^\top = \hat{\mathbf{f}}^\top \mathbf{C}$$

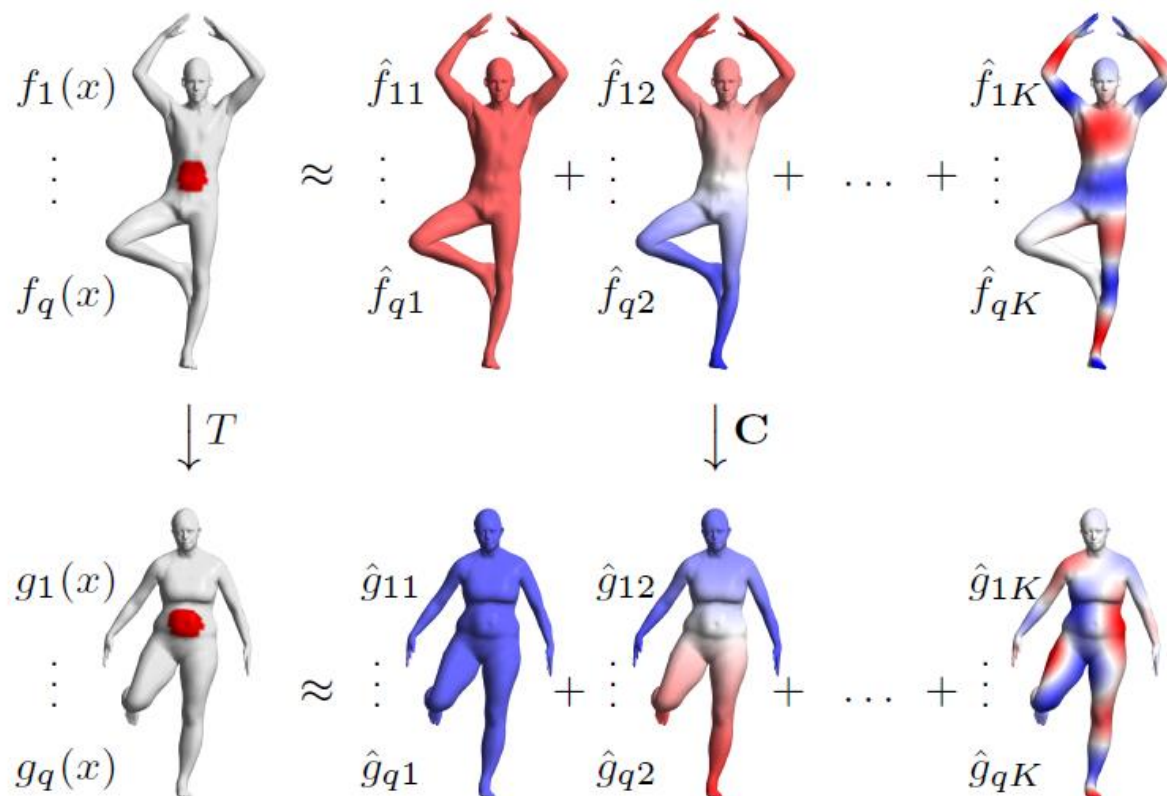
Functional maps: spectral domain



Recover correspondence from $q \geq k$ dimensional pointwise features

$$\begin{pmatrix} \hat{g}_{11} & \hat{g}_{12} & \dots & \hat{g}_{1K} \\ \vdots & \vdots & & \vdots \\ \hat{g}_{q1} & \hat{g}_{q2} & \dots & \hat{g}_{qK} \end{pmatrix} = \begin{pmatrix} \hat{f}_{11} & \hat{f}_{12} & \dots & \hat{f}_{1K} \\ \vdots & \vdots & & \vdots \\ \hat{f}_{q1} & \hat{f}_{q2} & \dots & \hat{f}_{qK} \end{pmatrix} \mathbf{C}$$

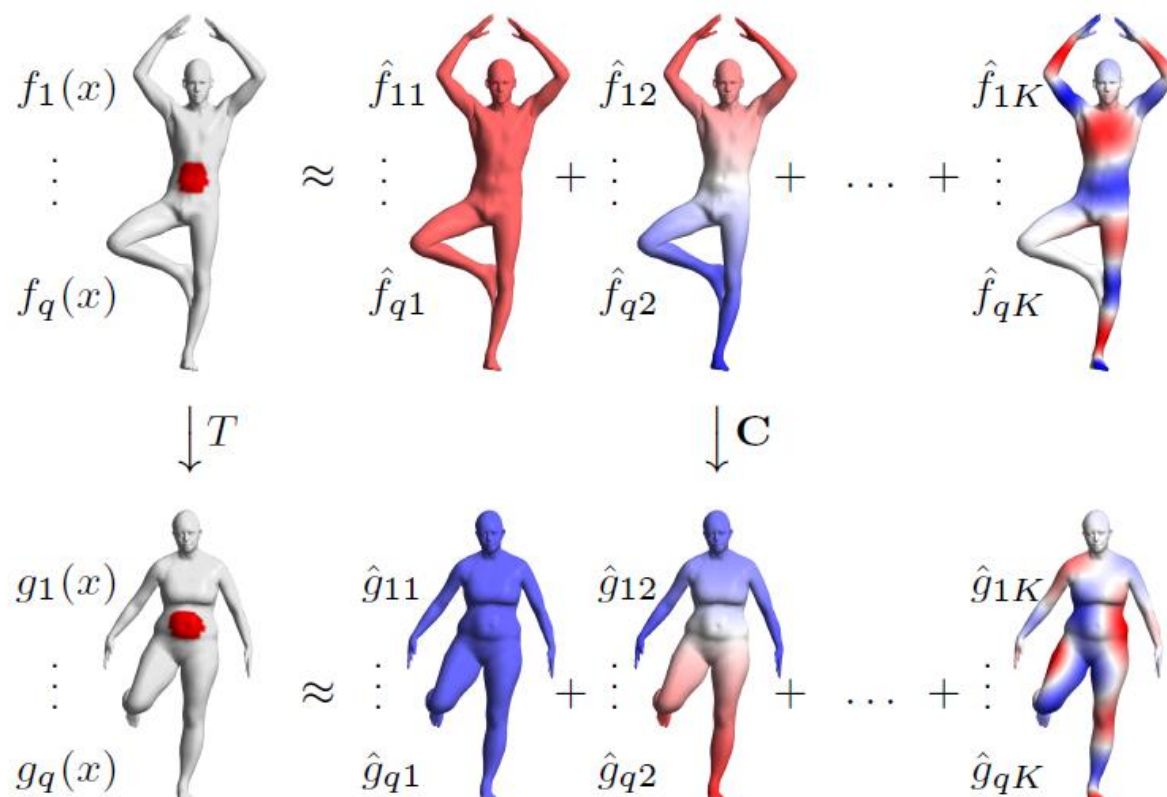
Functional maps: spectral domain



Recover correspondence from $q \geq k$ dimensional pointwise features

$$\hat{\mathbf{G}} = \hat{\mathbf{F}}\mathbf{C}$$

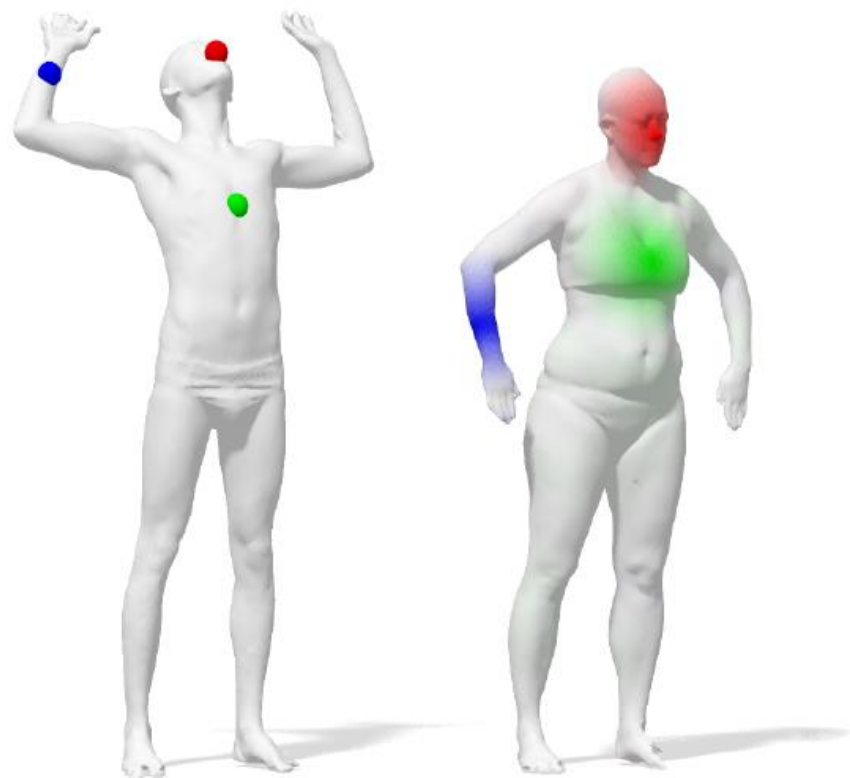
Functional maps: spectral domain



Recover correspondence from $q \geq k$ dimensional pointwise features

$$\mathbf{C}^* = \underset{\mathbf{C}}{\operatorname{argmin}} \|\hat{\mathbf{F}}\mathbf{C} - \hat{\mathbf{G}}\|_{\mathbf{F}}^2$$

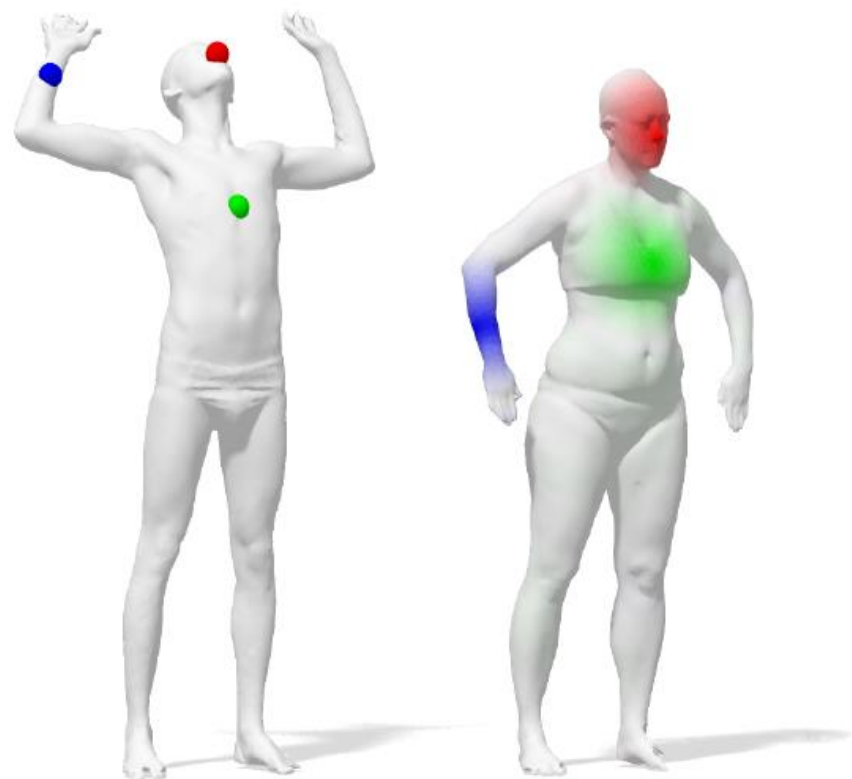
Functional maps: spatial domain



Rank- K approximation of spatial correspondence

$$\mathbf{T} \approx \Psi \mathbf{C} \Phi^T$$

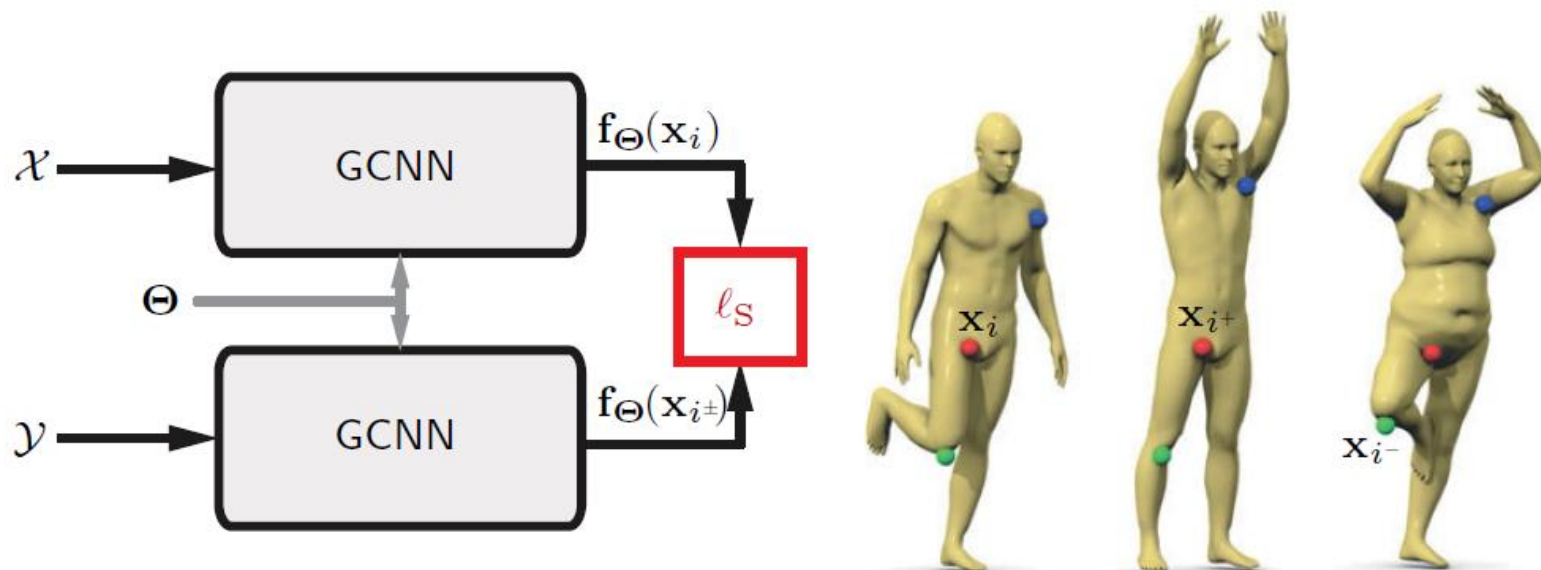
Functional maps: spatial domain



Probability p_{ij} of point j mapping to i

$$\mathbf{P} \approx |\Psi \mathbf{C} \Phi^T|_{\|\cdot\|}$$

Siamese metric learning



Training set

positive (i, i^+) and negative (i, i^-) pairs of points

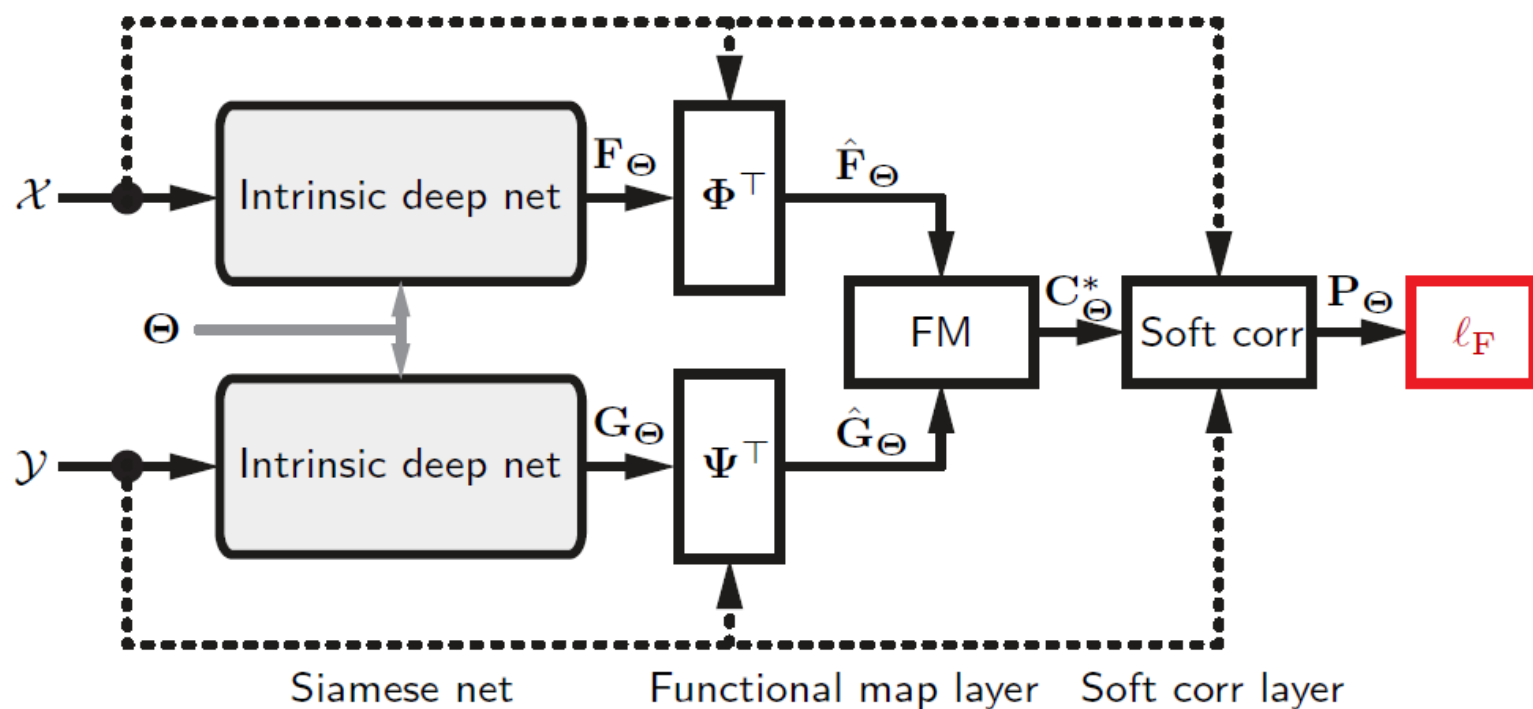
Siamese net

two net instances with shared parameters Θ

Poitwise feature cost

$$\begin{aligned} \ell_S(\Theta) = & \gamma \sum_{i, i^+} \|\mathbf{f}_{\Theta}(\mathbf{x}_i) - \mathbf{f}_{\Theta}(\mathbf{x}_{i^+})\|_2^2 \\ & + (1 - \gamma) \sum_{i, i^-} [\mu - \|\mathbf{f}_{\Theta}(\mathbf{x}_i) - \mathbf{f}_{\Theta}(\mathbf{x}_{i^-})\|_2^2]_+ \end{aligned}$$

Structured correspondence with FMNet



Siamese net

two net instances with shared parameters Θ

Functional map layer

$$\mathbf{C}_\Theta^* = \underset{\mathbf{C}}{\operatorname{argmin}} \|\hat{\mathbf{F}}_\Theta \mathbf{C} - \hat{\mathbf{G}}_\Theta\|_F^2$$

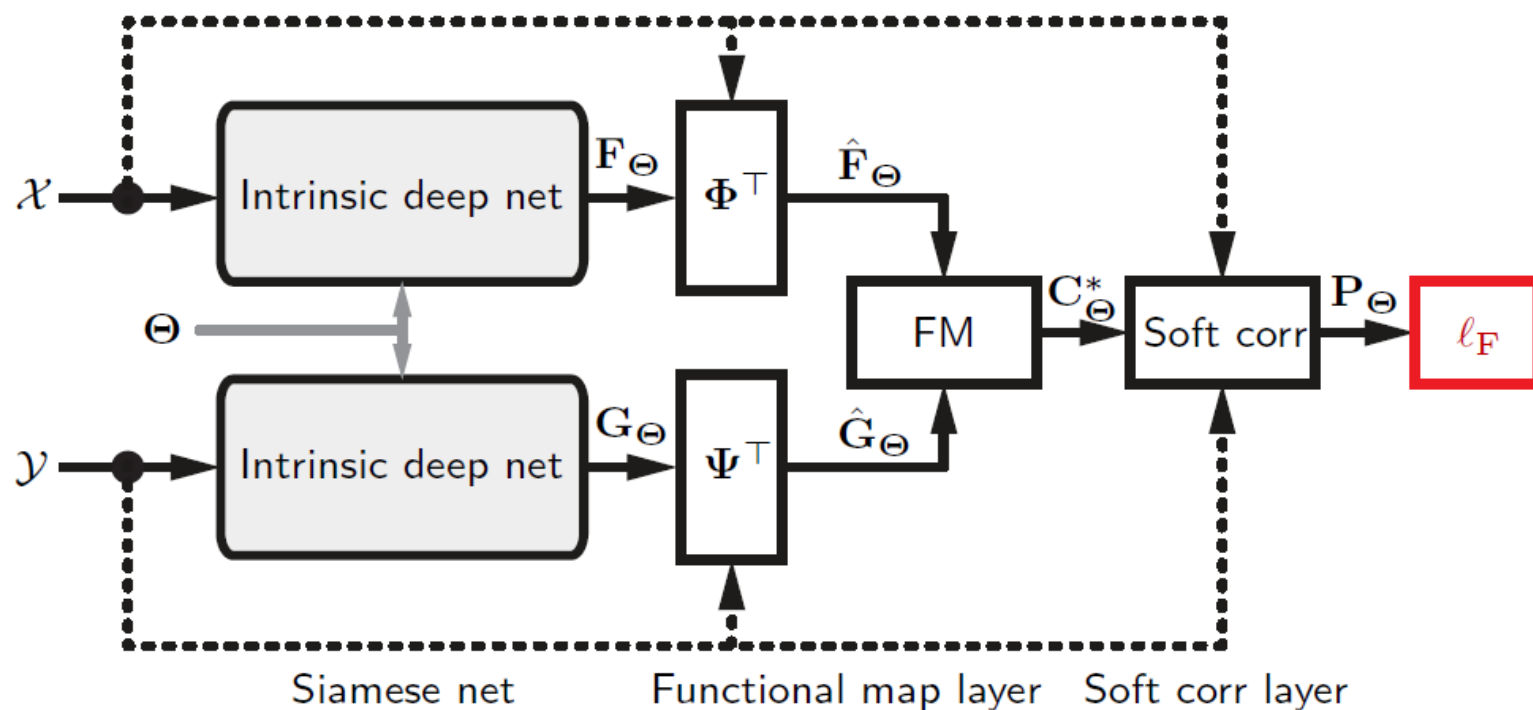
Soft correspondence layer

$$\mathbf{P}_\Theta = |\Psi \mathbf{C}_\Theta \Phi^\top|_{\|\cdot\|}$$

Soft error cost

$$\ell_F(\Theta) = \sum_{i=1}^n \sum_{j=1}^m p_{\Theta,ij} d_{\mathcal{Y}}(\pi_i^*, j)$$

Structured correspondence with FMNet



Siamese net

two net instances with shared parameters Θ

Functional map layer

$$C_{\Theta}^* = \hat{F}_{\Theta}^{\dagger} \hat{G}_{\Theta}$$

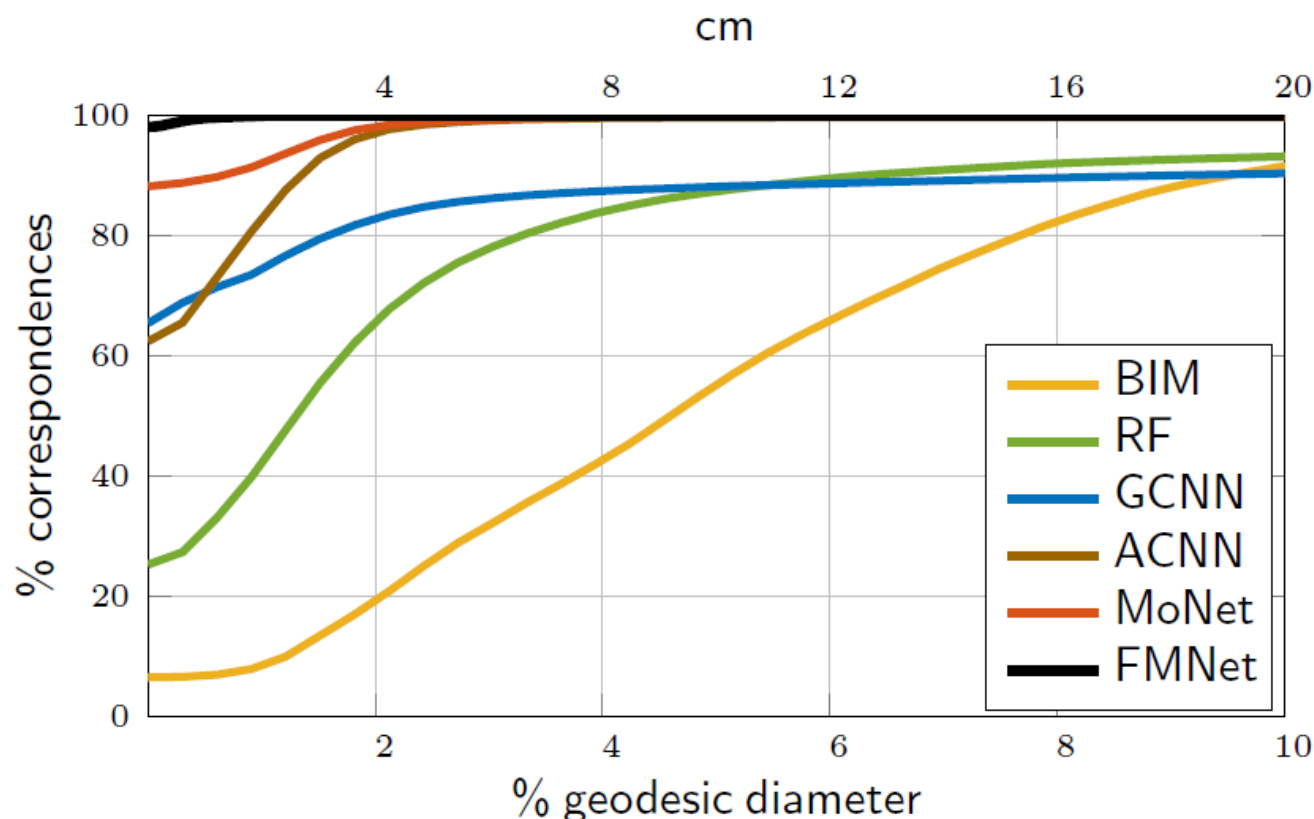
Soft correspondence layer

$$P_{\Theta} = |\Psi C_{\Theta} \Phi^{\top}|_{\|\cdot\|}$$

Soft error cost

$$\ell_F(\Theta) = \|P_{\Theta} \circ D_Y\|$$

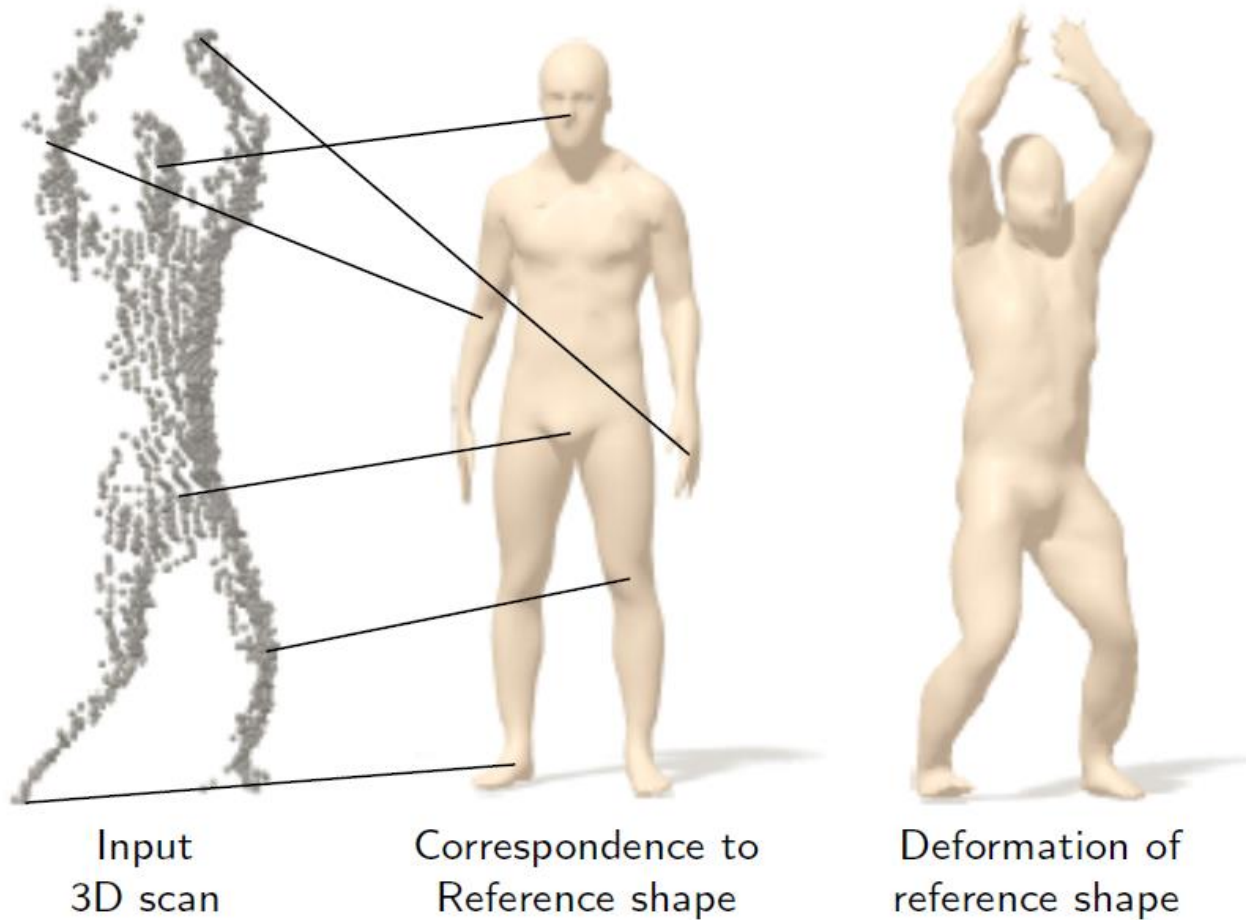
Correspondence quality comparison



Correspondence evaluated using asymmetric Princeton benchmark
(training and testing: disjoint subsets of FAUST)

Methods: Kim et al. 2011 (BIM); Rodolà et al. 2014 (RF); Boscaini et al. 2015 (ADD); Masci et al. 2015 (GCNN); Boscaini et al. 2016 (ACNN); Monti et al. 2016 (MoNet); Litany et al. 2017 (FMNet); data: Bogo et al. 2014 (FAUST); benchmark: Kim et al. 2011

3D shape analysis and synthesis

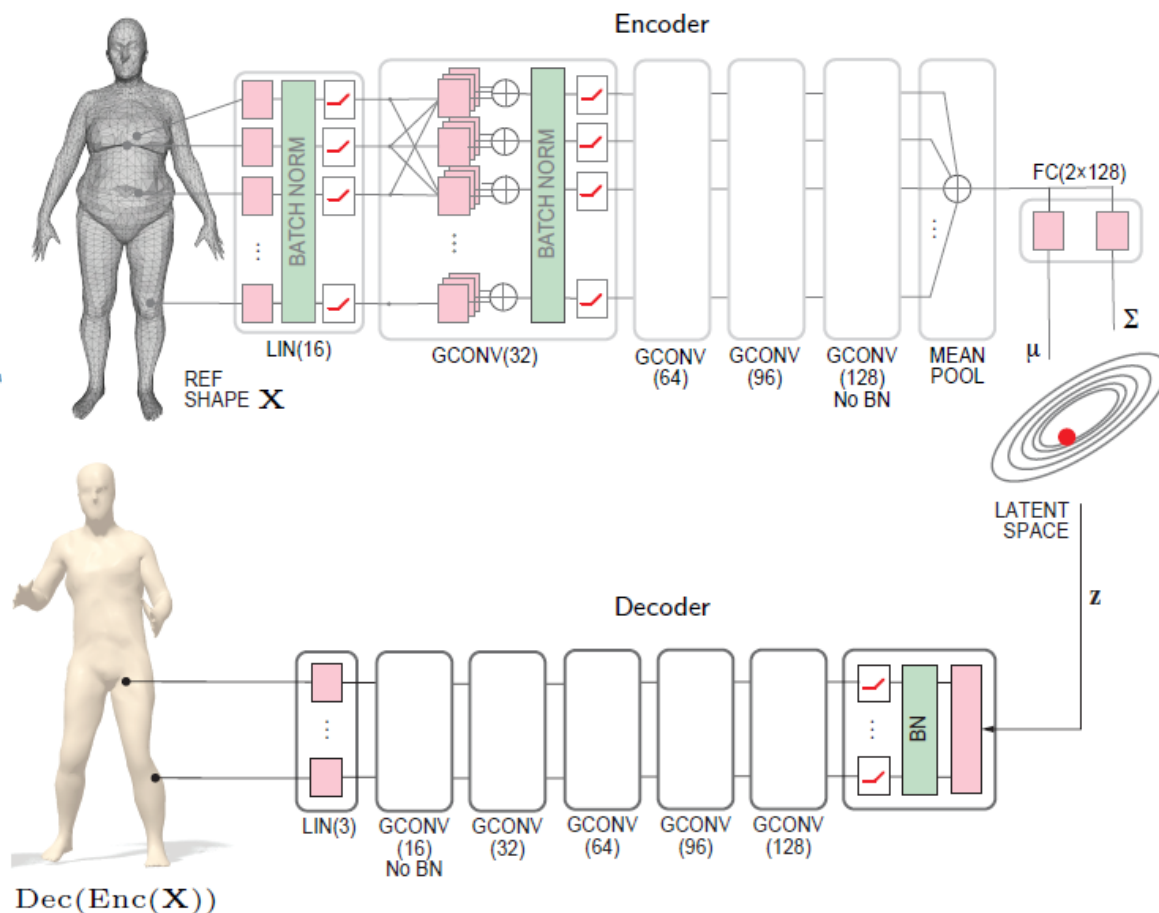


Intrinsic Variational Autoencoder (VAE)

Minimize

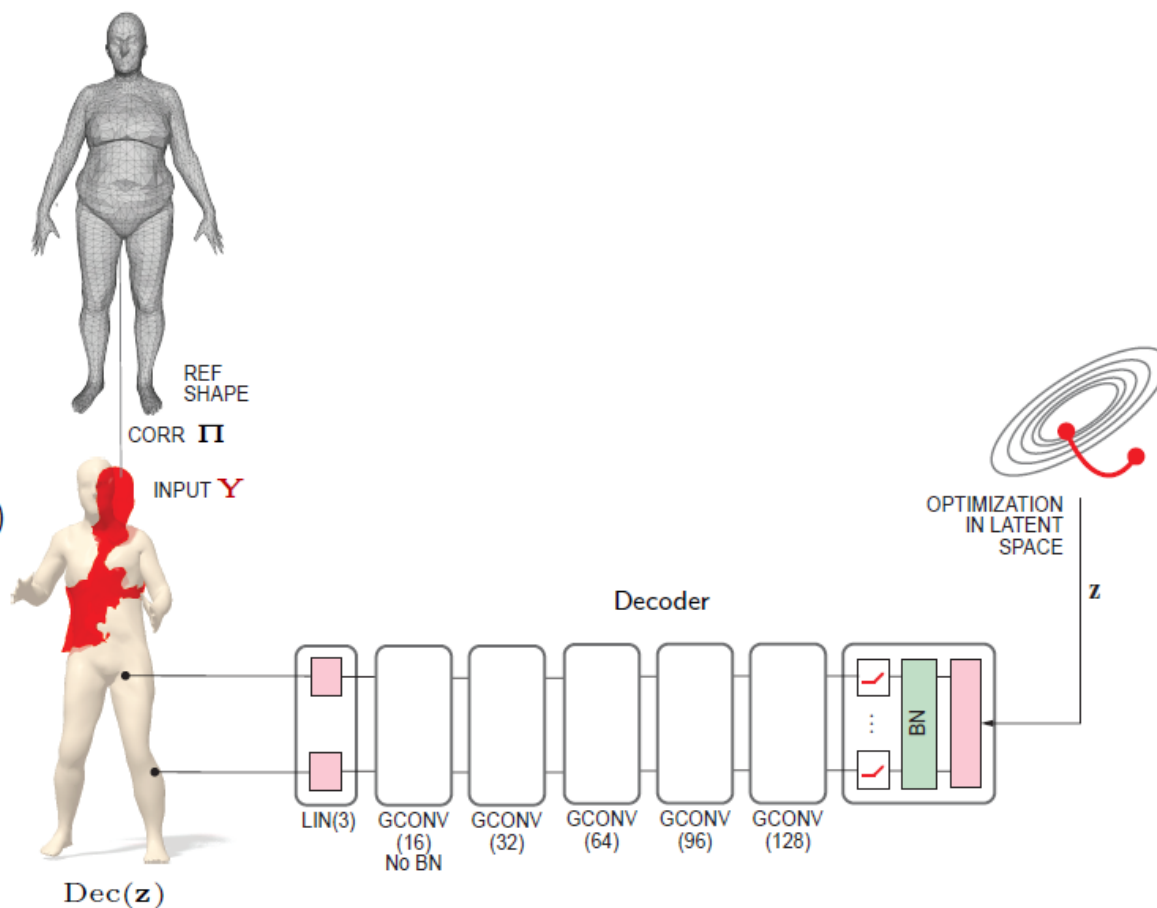
$$\| \text{Dec}(\text{Enc}(\mathbf{X})) - \mathbf{X} \|_F + \lambda D_{\text{KL}}(q(\mathbf{z}|\mathbf{X})||p(\mathbf{z}))$$

w.r.t. net parameters

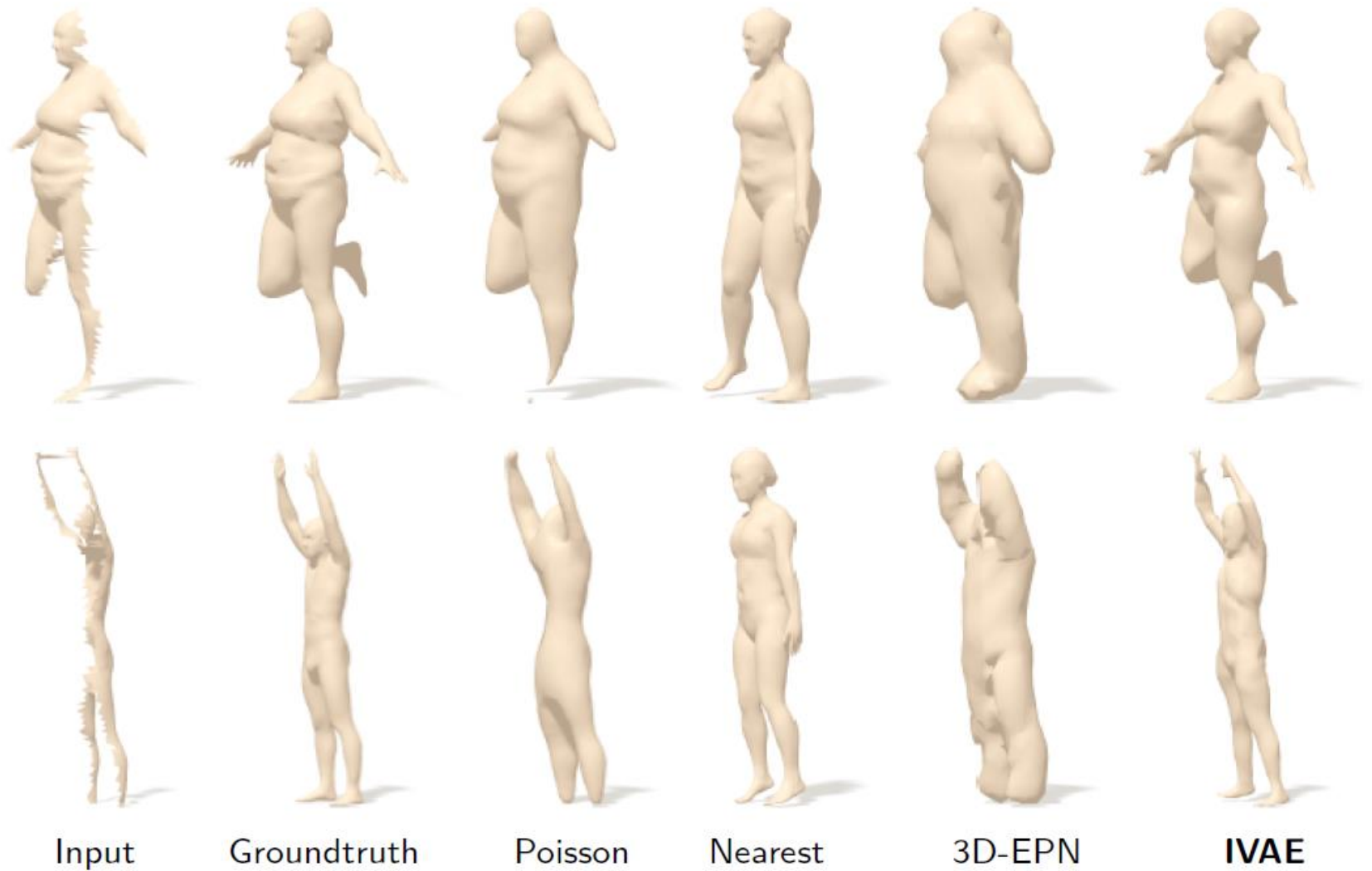


Shape completion

Minimize
 $\|\text{Dec}(\mathbf{z})\mathbf{\Pi} - \mathbf{T}\mathbf{Y}\|_F$
in alternating manner
w.r.t. \mathbf{z} and $\mathbf{T} \in \text{SO}(3)$



Shape completion comparison

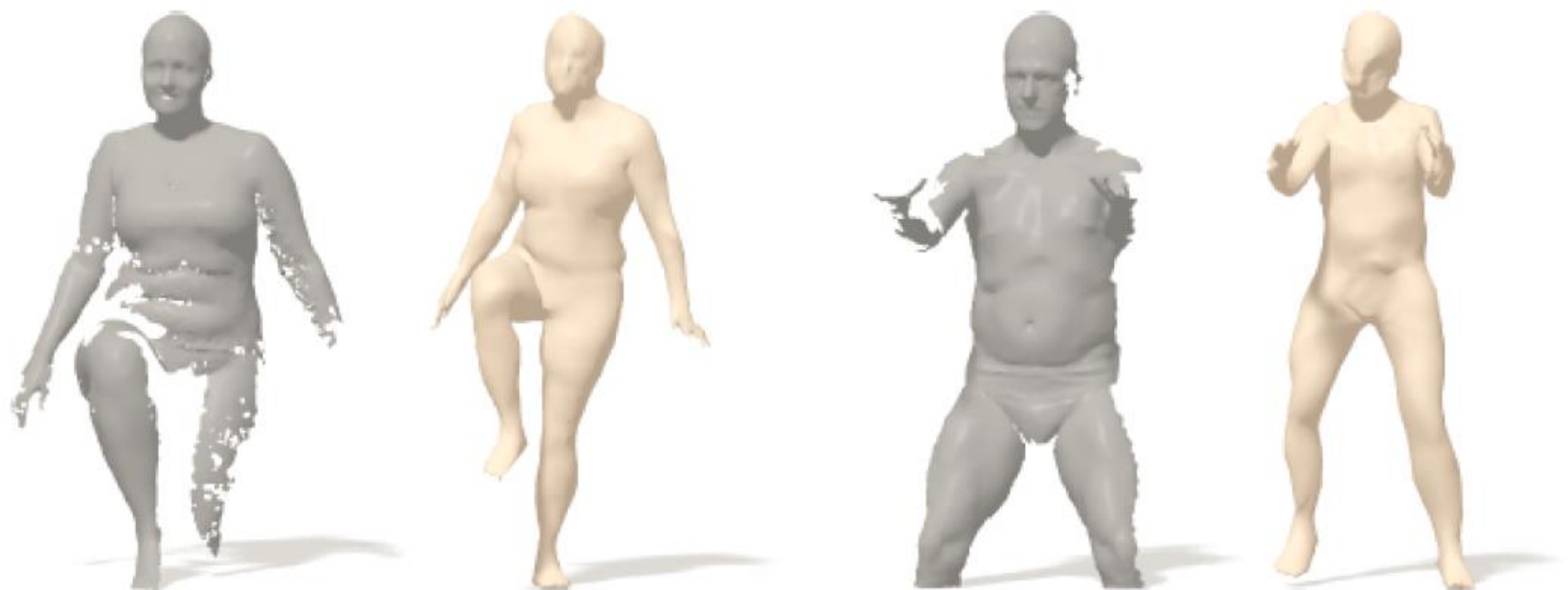


Methods: Litany et al. 2017; Dai et al. 2016 (3D-EPN); Kazhdan et al. 2013 (Poisson)

Shape completion examples



Shape completion examples



Litany et al. 2017; data: Bogo et al. 2014 (FAUST)