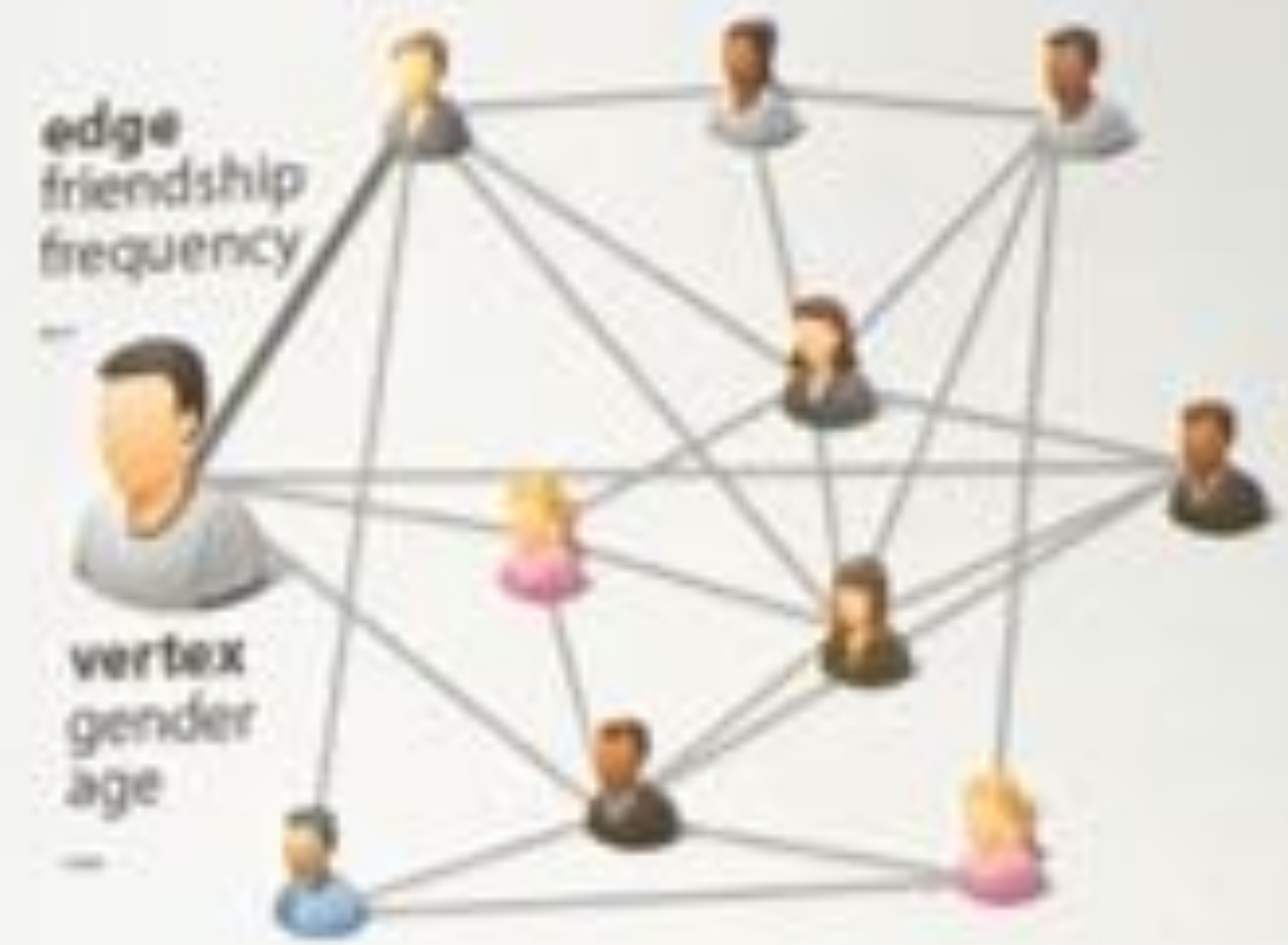
Prototypical non-Euclidean objects

Manifolds

Graphs

Domain structure vs Data on a domain

Domain structure



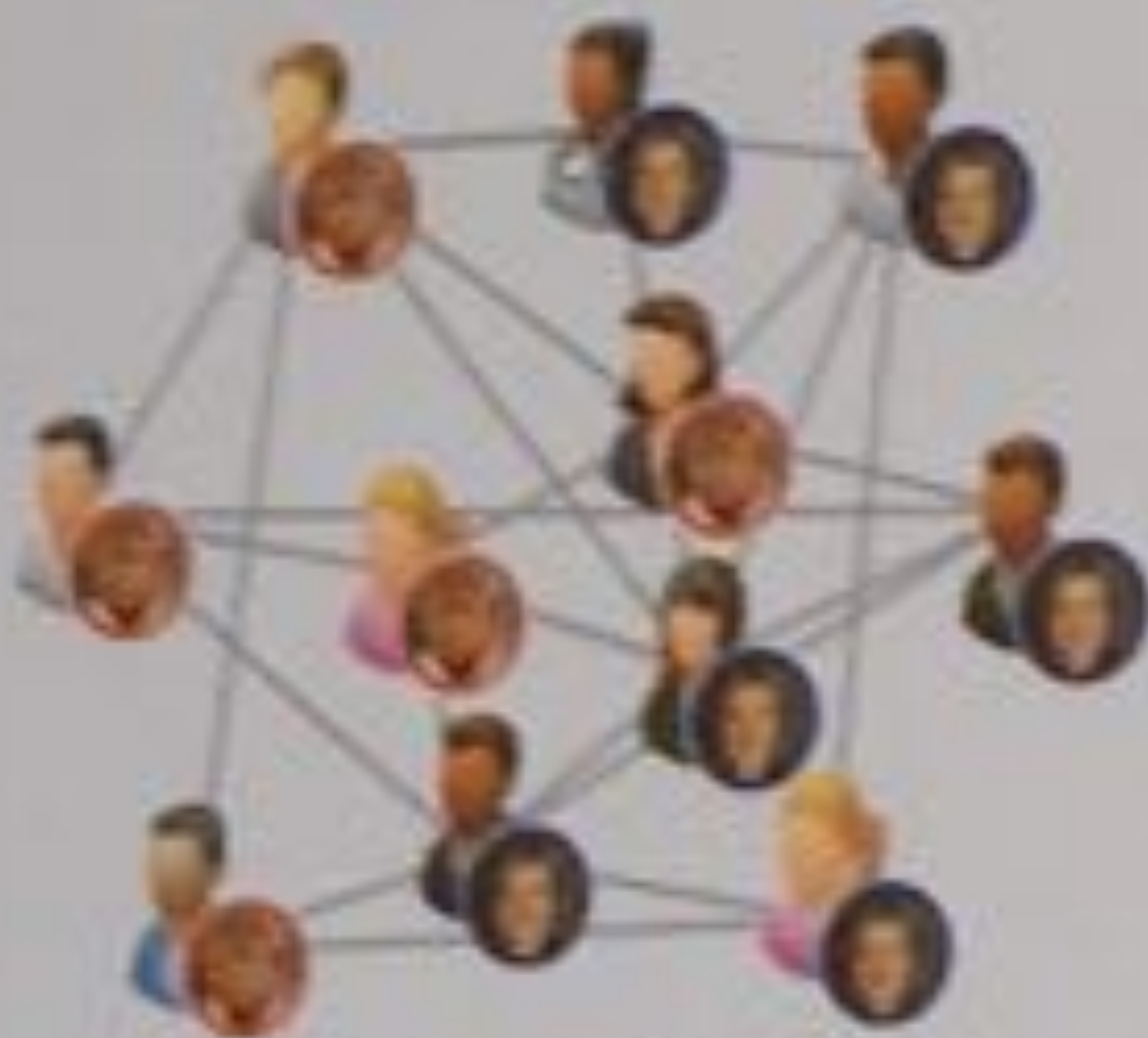
Data on a domain

Known vs Unknown domain

Given graph

Given point cloud in feature space
(graph metric has to be learnt)

Vertex-wise classification

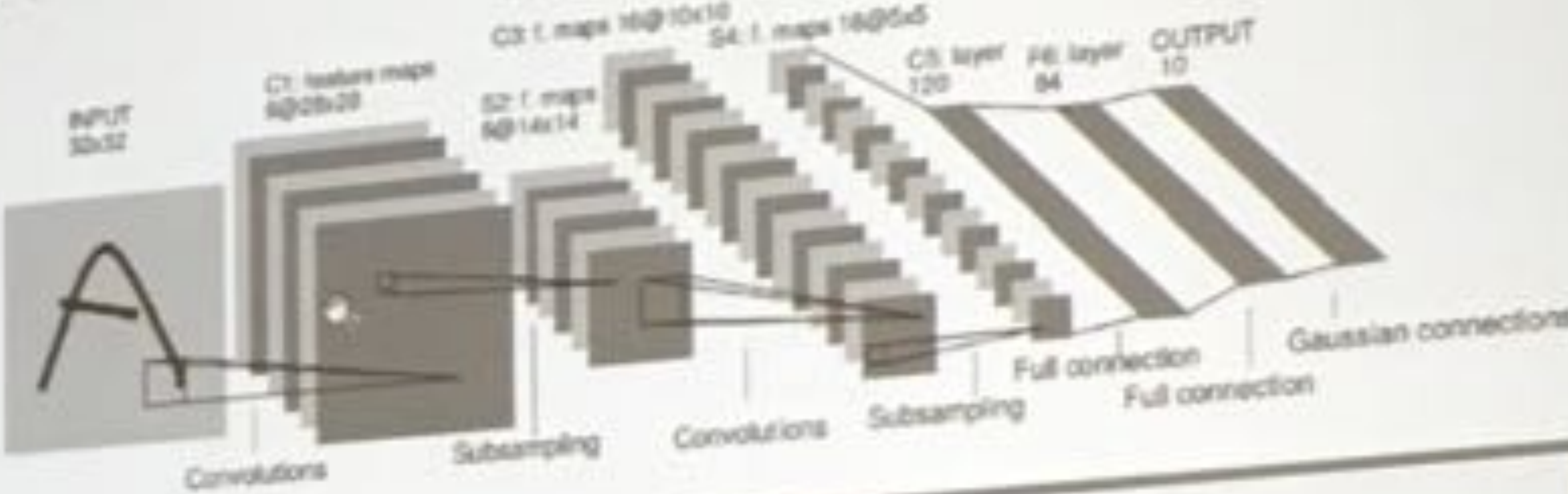


Community Detection

What this tutorial is about?

- Basics of Euclidean convolutional neural networks
- Basics of graph theory
- Fourier analysis on graphs
- Spectral-domain methods
- Spatial-domain methods
- Applications: network analysis, recommender systems, computer graphics, chemistry, high-energy physics, ...

Key properties of CNNs



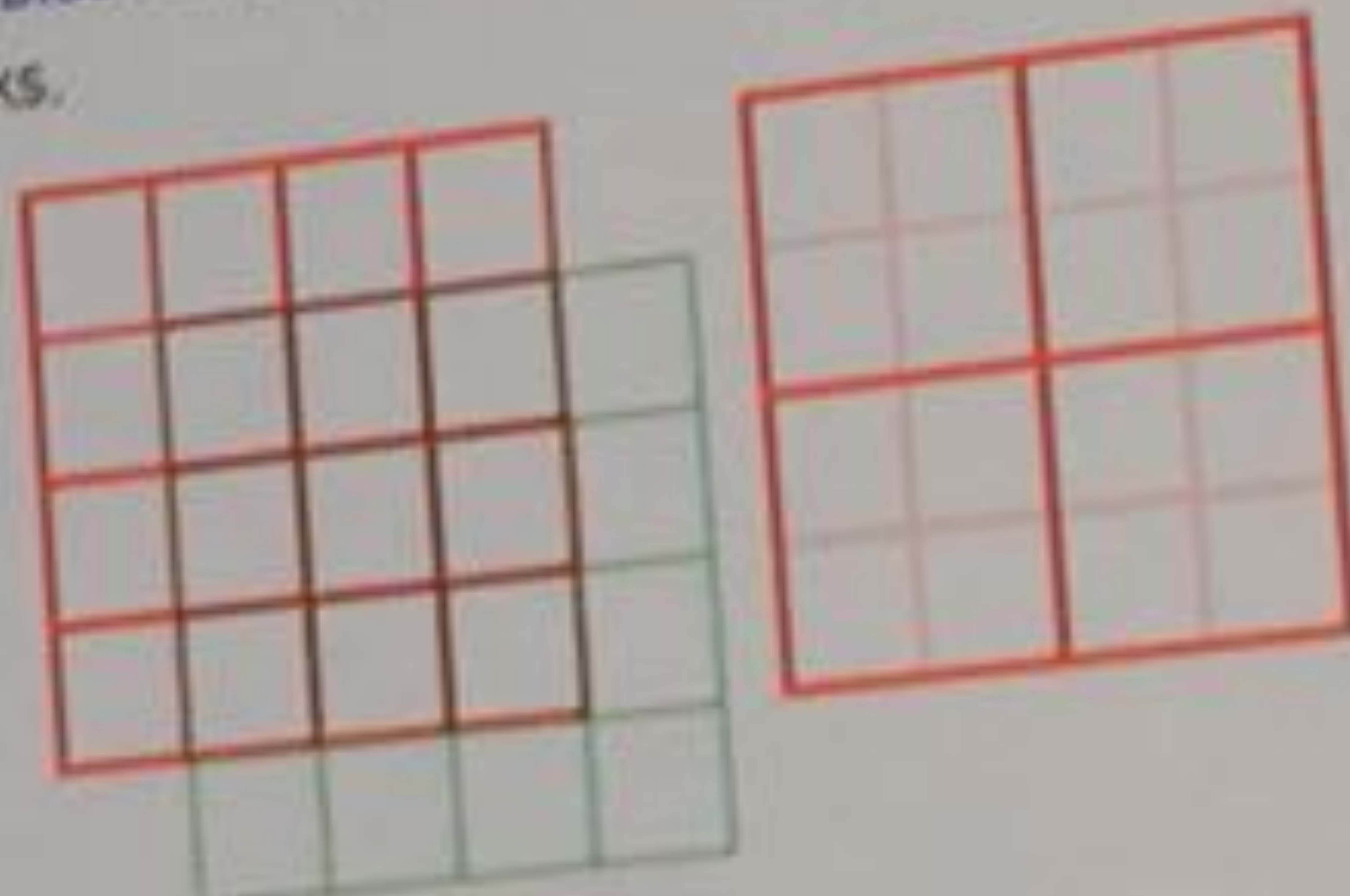
- ☺ Convolutional **(Translation invariance)**
- ☺ Scale Separation **(Compositionality)**
- ☺ Filters localized in space **(Deformation Stability)**
- ☺ $\mathcal{O}(1)$ parameters per filter (independent of input image size n)
- ☺ $\mathcal{O}(n)$ complexity per layer (filtering done in the spatial domain)
- ☺ $\mathcal{O}(\log n)$ layers in classification tasks

LeCun et al. 1989

CNNs and Euclidean Geometry

CNNs are defined over Euclidean domains or Grids Ω. Two fundamental properties:

- Translation Invariance (yielding convolutions).
- Multiscale structure (yielding downsampling).
- Inductive bias that exploits stationarity and deformation stability of many tasks.



Roadmap: extend CNNs to non-Euclidean geometries by replacing filtering and pooling by appropriate operators

Graph theory in one minute

- Weighted undirected graph $\mathcal{G}$ with vertices $\mathcal{V} = \{1, \ldots, n\}$, edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ and edge weights $w_{ij} \geq 0$ for $(i, j) \in \mathcal{E}$

- Functions over the vertices $L^2(\mathcal{V}) = \{f : \mathcal{V} \rightarrow \mathbb{R}\}$ represented as vectors $\mathbf{f} = (f_1, \ldots, f_n)$

- Hilbert space with inner product

$$\langle f, g \rangle_{L^2(\mathcal{V})} = \sum_{i \in \mathcal{V}} f_i g_i = \mathbf{f}^\top \mathbf{g}$$

Graph Laplacian

- Unnormalized Laplacian $\Delta : L^2(\mathcal{V}) \to L^2(\mathcal{V})$

$$(\Delta f)_i \;=\; \sum_{j:(i,j)\in\mathcal{E}} w_{ij}(f_i - f_j)$$

  (up to scale) difference between f and its local average

- Represented as a positive semi-definite $n \times n$ matrix $\mathbf{\Delta} = \mathbf{D} - \mathbf{W}$ where $\mathbf{W} = (w_{ij})$ and $\mathbf{D} = \mathrm{diag}(\sum_{j\neq i} w_{ij})$

- Dirichlet energy of f

$$\|f\|_{\mathcal{G}}^2 = \frac{1}{2}\sum_{ij=1}^{n} w_{ij}(f_i - f_j)^2 = \mathbf{f}^\top \mathbf{\Delta} \mathbf{f}$$

  measures the smoothness of f (how fast it changes locally)

Riemannian manifolds in one minute

- Manifold $\mathcal{X}$ = topological space
- Tangent plane $T_x\mathcal{X}$ = local Euclidean representation of manifold $\mathcal{X}$ around x
- Riemannian metric describes the local intrinsic structure at x

$$\langle \cdot, \cdot \rangle_{T_x\mathcal{X}} : T_x\mathcal{X} \times T_x\mathcal{X} \to \mathbb{R}$$

- Scalar fields $f : \mathcal{X} \to \mathbb{R}$ and vector fields $F : \mathcal{X} \to T\mathcal{X}$
- Hilbert spaces with inner products

$$\langle f, g \rangle_{L^2(\mathcal{X})} = \int_{\mathcal{X}} f(x) g(x) dx$$

$$\langle F, G \rangle_{L^2(T\mathcal{X})} = \int_{\mathcal{X}} \langle F(x), G(x) \rangle_{T_x\mathcal{X}} dx$$

Orthogonal bases on graphs

Find the smoothest orthogonal basis $\{\phi_1, \ldots, \phi_m\} \subseteq L^2(V)$

$$\min_{\mathbf{\Phi} \in \mathbb{R}^{n \times m}} \operatorname{trace}(\mathbf{\Phi}^\top \mathbf{\Delta} \mathbf{\Phi}) \quad \text{s.t.} \quad \mathbf{\Phi}^\top \mathbf{\Phi} = \mathbf{I}$$

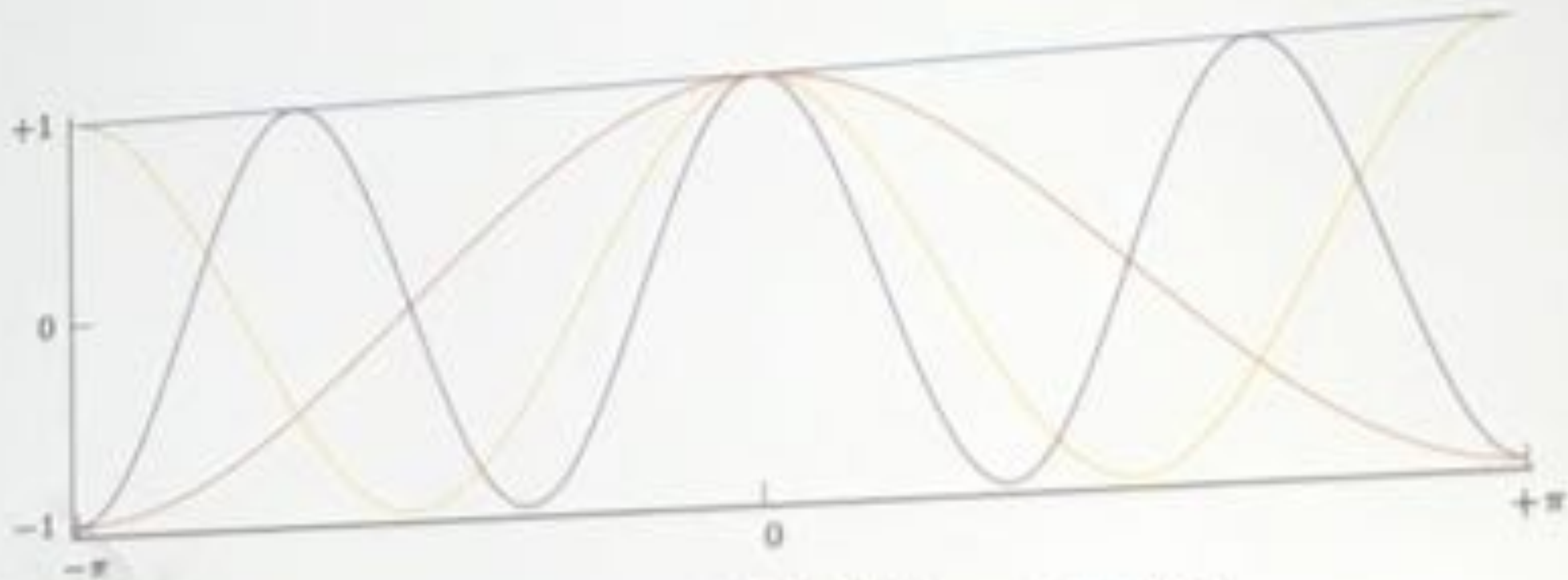
Solution: $\mathbf{\Phi}$ = Laplacian eigenvectors

Laplacian eigenvectors and eigenvalues

Eigendecomposition of a graph Laplacian

$$\boldsymbol{\Delta} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$$

where $\boldsymbol{\Phi} = (\phi_1, \ldots, \phi_n)$ are orthogonal eigenvectors ($\boldsymbol{\Phi}^\top\boldsymbol{\Phi} = \mathbf{I}$) and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the corresponding non-negative eigenvalues



First eigenfunctions of 1D Euclidean Laplacian

Warmup: sets as inputs

- how can we make a network that respects the notion of "set"?
  - invariant to permutations
  - dynamic resizing

Patch operators

Image

Manifold

Boscaini, Masci, Bron… …ndergheynst 2015

Spatial convolution on manifolds

- Geodesic polar coordinates

$$\mathbf{u}(x,y) = (\rho(x,y), \theta(x,y))$$

- Set of weighting functions

$$w_1(\mathbf{u}), \dots, w_J(\mathbf{u})$$

$u_1 = \rho$ $u_2 = \theta$

Spatial convolution

$$(f \star g)(x) = \sum_{j=1}^{J} g_j \underbrace{\int_X w_j(\mathbf{u}(x,x')) f(x') dx'}_{\text{patch operator } \mathcal{D}_j(x) f}$$

where $g_1, \dots, g_J$ are the spatial filter coefficients

Monti, Boscaini, Rodolà, Svoboda, Bronstein 2017

Mixture Model Network (MoNet)

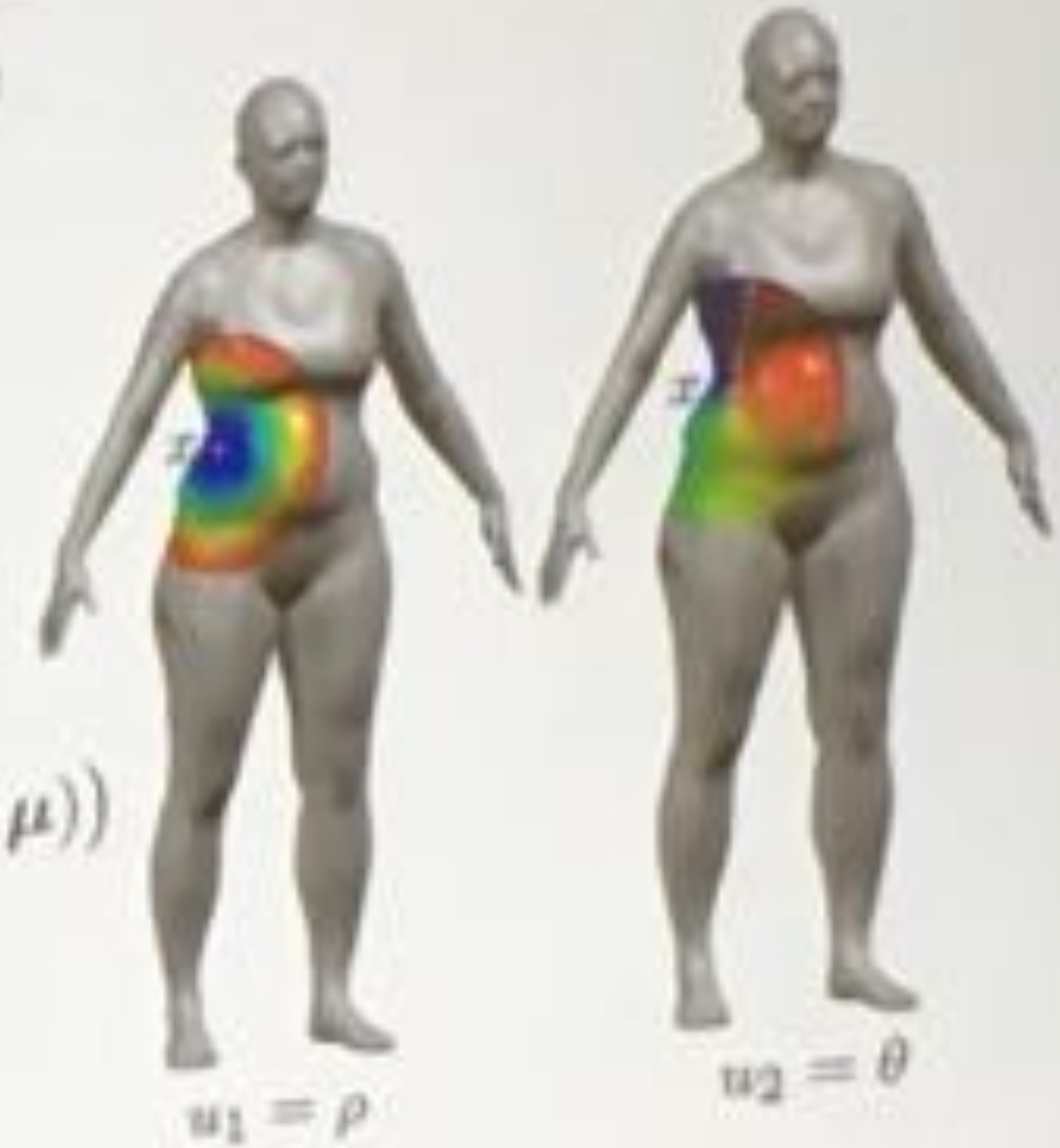
- Geodesic polar coordinates

$$\mathbf{u}(x,y) = (\rho(x,y), \theta(x,y))$$

- Gaussian weighting functions

$$w_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{u}) = \exp\left(-\tfrac{1}{2}(\mathbf{u}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{u}-\boldsymbol{\mu})\right)$$

with learnable covariance $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}$



Spatial convolution

$(f \star g)(x) =$ … $_{\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j}(\mathbf{u}(x,x'))\, f(x')dx'$

…sian mixture

where $g_1,\ldots,g_J$ are the … er coefficients and $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_J$ and $\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_J$ are patch oper… …eters

Monti, Boscaini, Rodolà, S…

Patch operator weighting functions on manifolds

GCNN[1]

ACNN[2]

MoNet[3]

[1]Masci, Boscaini, Bronstein, Vandergheynst 2015; [2]Boscaini, Masci, Rodolà, Bronstein 2016; [3]Monti, Boscaini, Rodolà, Svoboda, Bronstein 2017

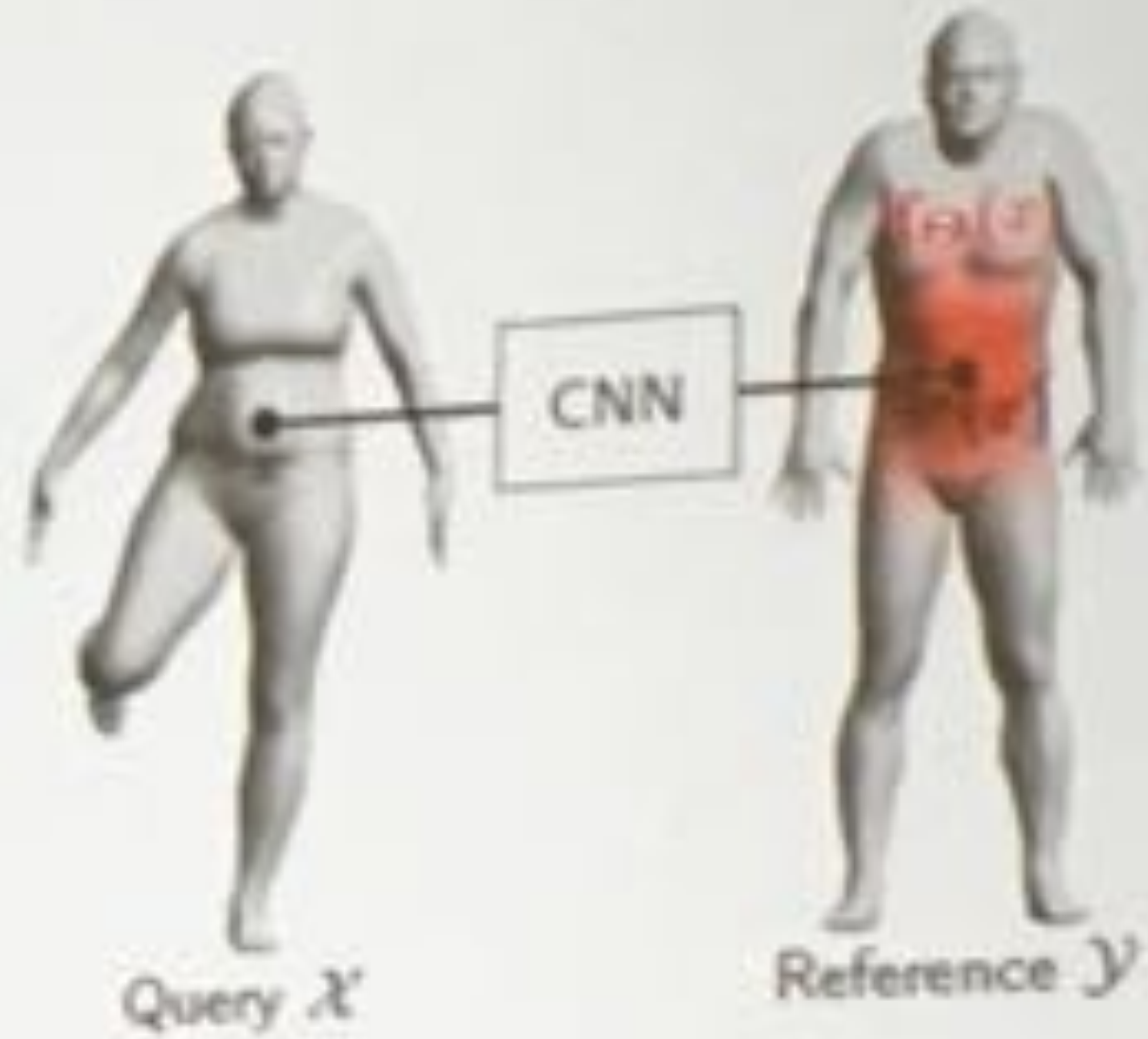

Deformable 3D correspondence
Isometric
Non-isometric
Partial

Correspondence II: Labeling

- Groundtruth correspondence $\pi^* : \mathcal{X} \to \mathcal{Y}$ from query shape $\mathcal{X}$ to some reference shape $\mathcal{Y}$ (discretized with n vertices)
- Correspondence = label each query vertex x as reference vertex y
- Net output at x after softmax layer

$$\mathbf{f}_\Theta(x) = (f_{\Theta,1}(x), \dots, f_{\Theta,n}(x))$$

= probability distribution on $\mathcal{Y}$

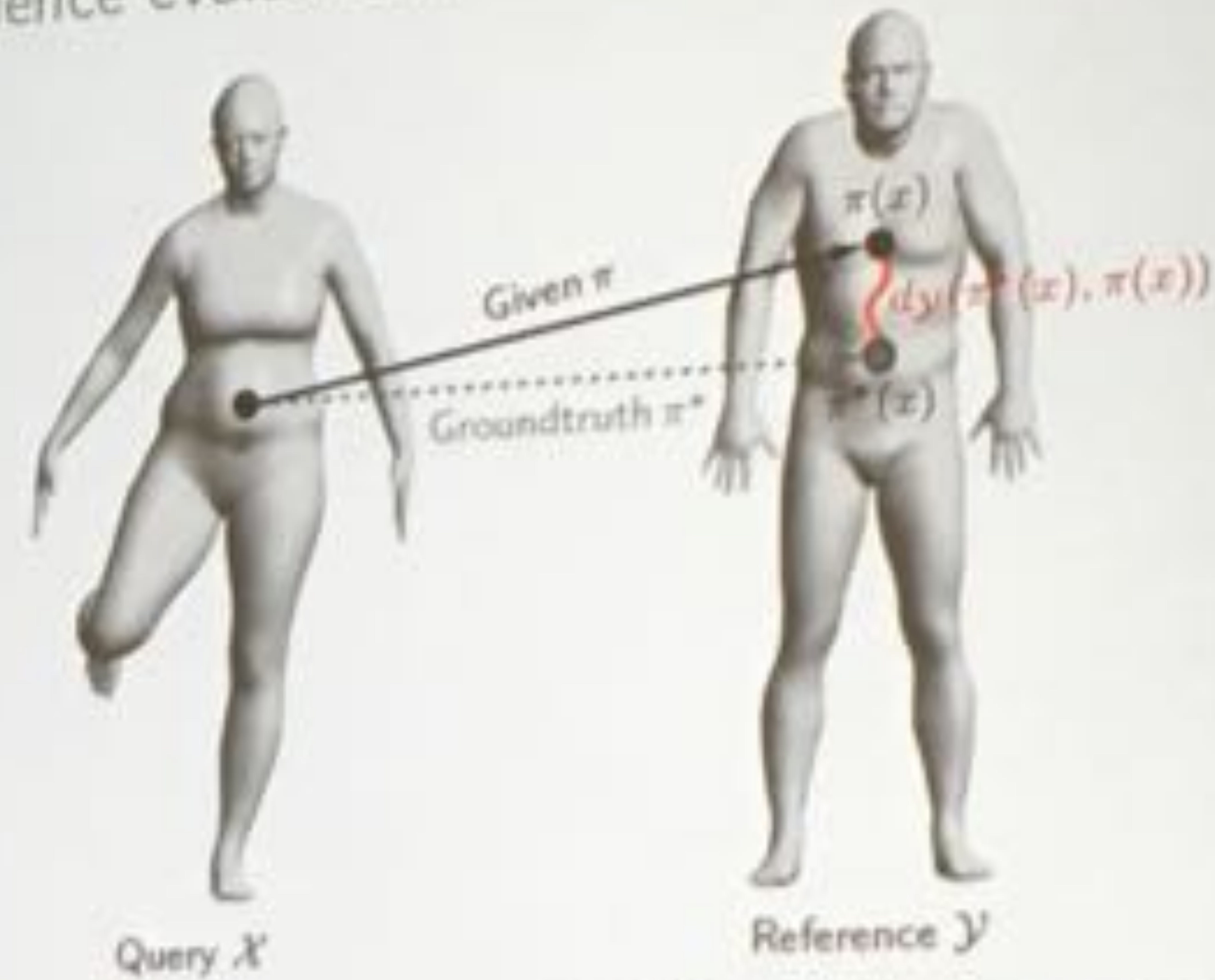


Minimize on training set the cross entropy between groundtruth correspondence and output probability distribution w.r.t. net parameters Θ

$$\min_\Theta \sum_x H(\delta_{\pi*(x)}, \mathbf{f}_\Theta(x))$$

Rodolà et al. 2014; Masci, Boscaini, Bronstein, Vandergheynst 2015

Correspondence evaluation: Princeton benchmark

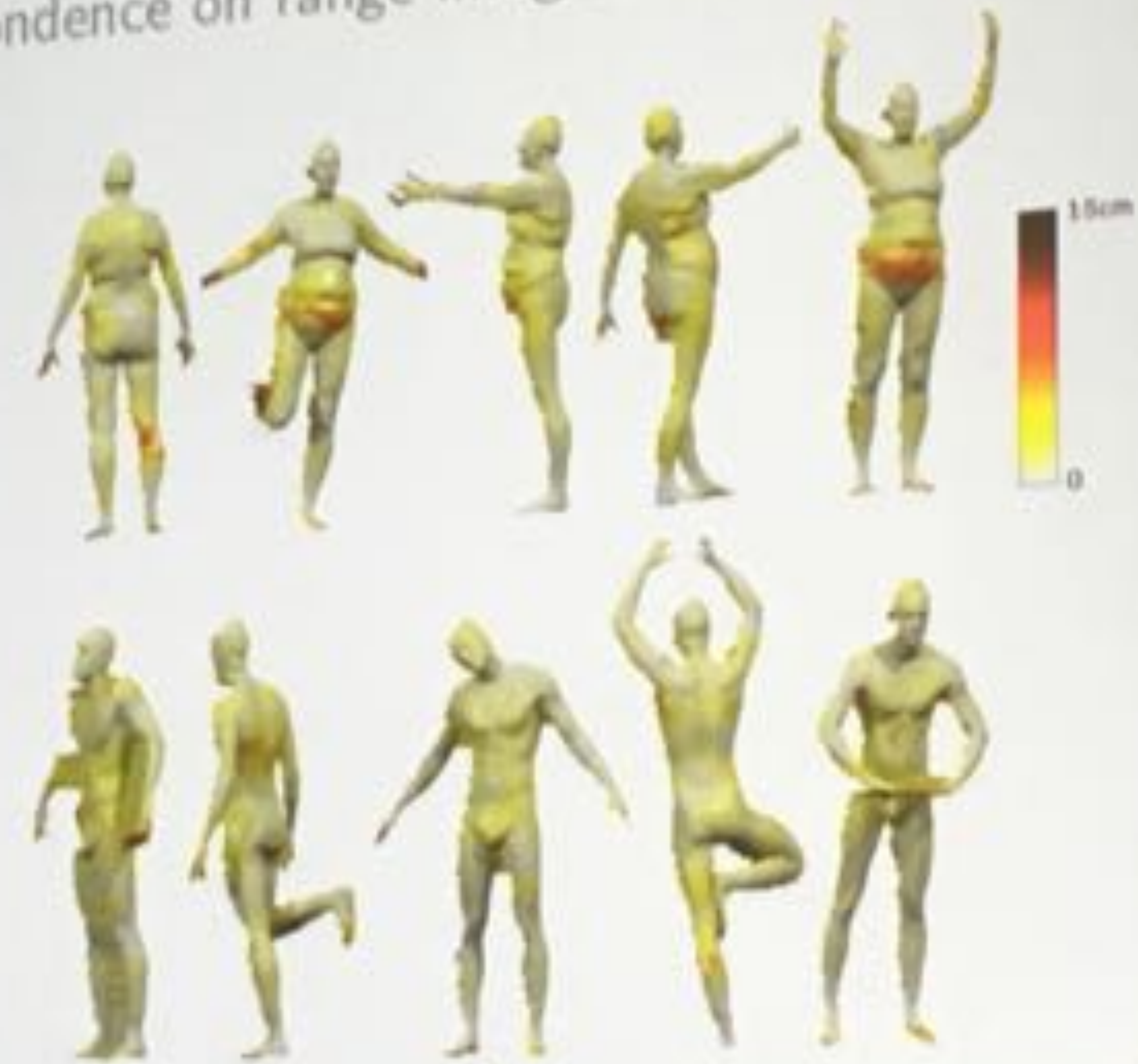


Pointwise correspondence error = geodesic distance from the groundtruth

$$\epsilon(x) = d_{\mathcal{Y}}(\pi^*(x), \pi(x))$$

Kim et al. 2011

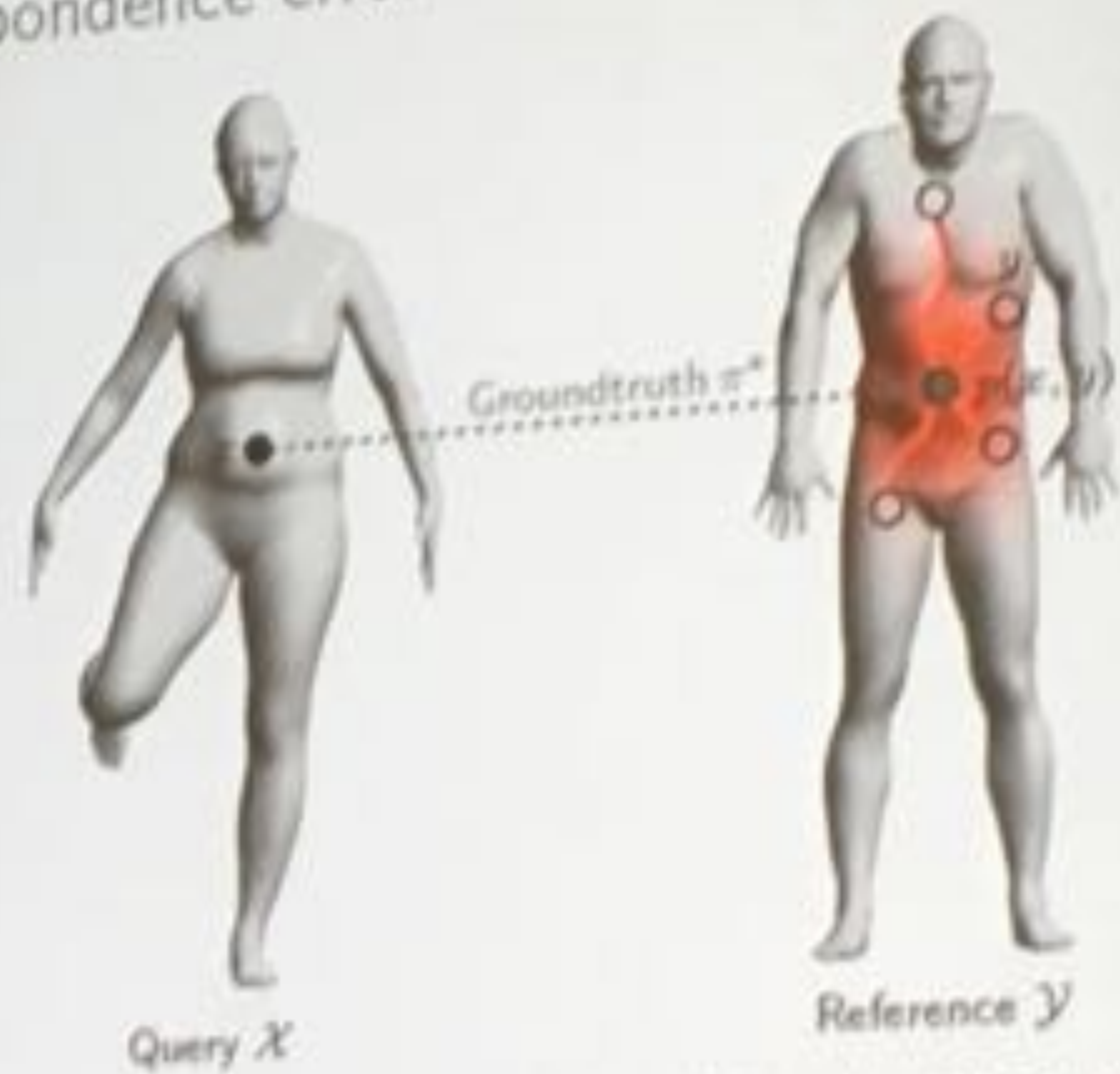
Correspondence on range images: MoNet



Pointwise correspondence error (geodesic distance from groundtruth)

Monti, Boscaini, Masci, Rodolà, Svoboda, Bronstein 2017

Soft correspondence error



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

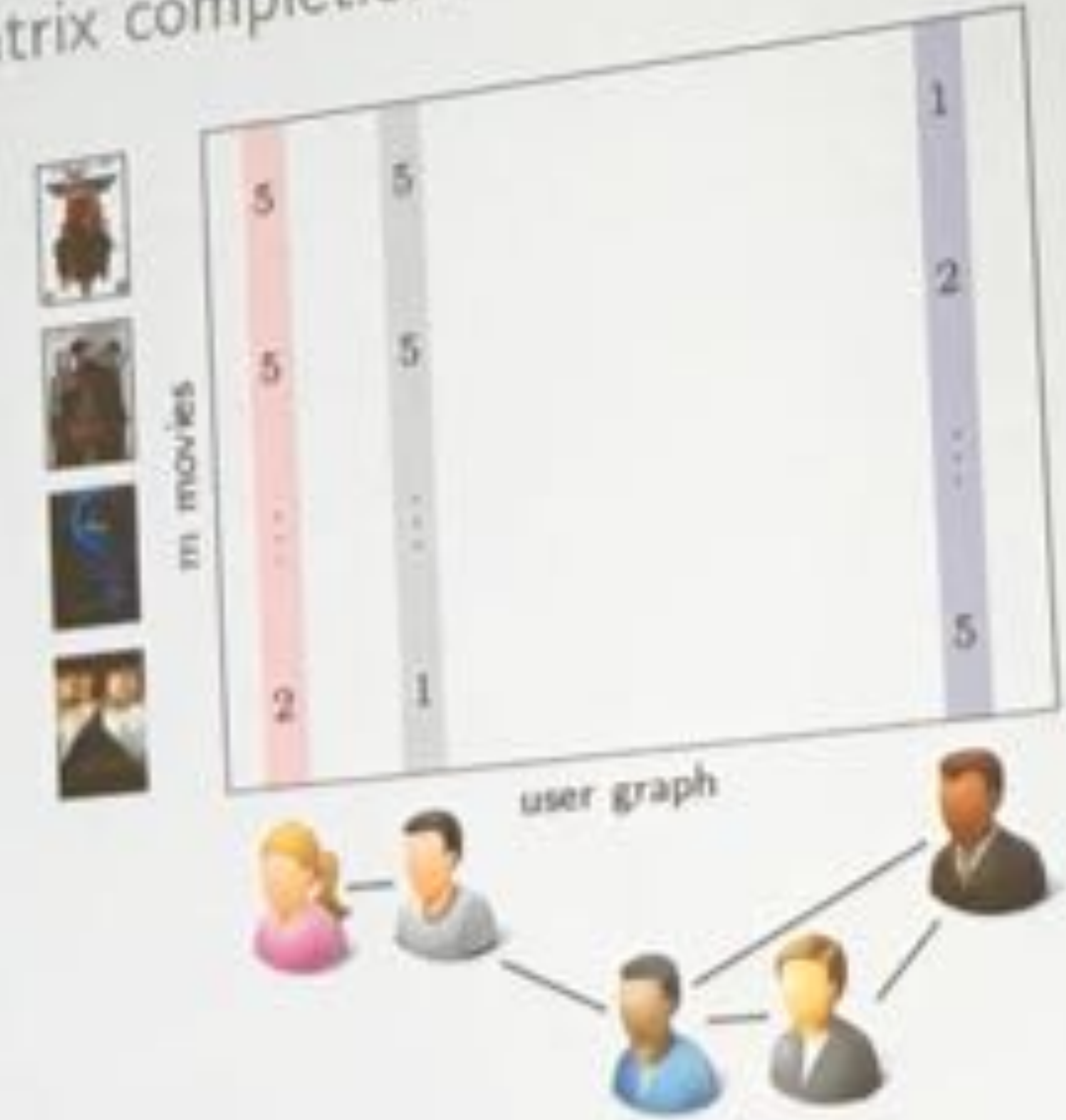
$$\tilde{\epsilon}(x) = \int_{\mathcal{Y}} p(x,y)\, d_{\mathcal{Y}}(\pi^*(x), y)\, dy$$

Kovnatsky, Bronstein, Bresson, Vandergheynst 2015

Matrix completion: 'Netflix challenge'

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \quad \operatorname{rank}(\mathbf{X}) \quad \text{s.t.} \quad x_{ij} = a_{ij} \;\; \forall ij \in \Omega$$

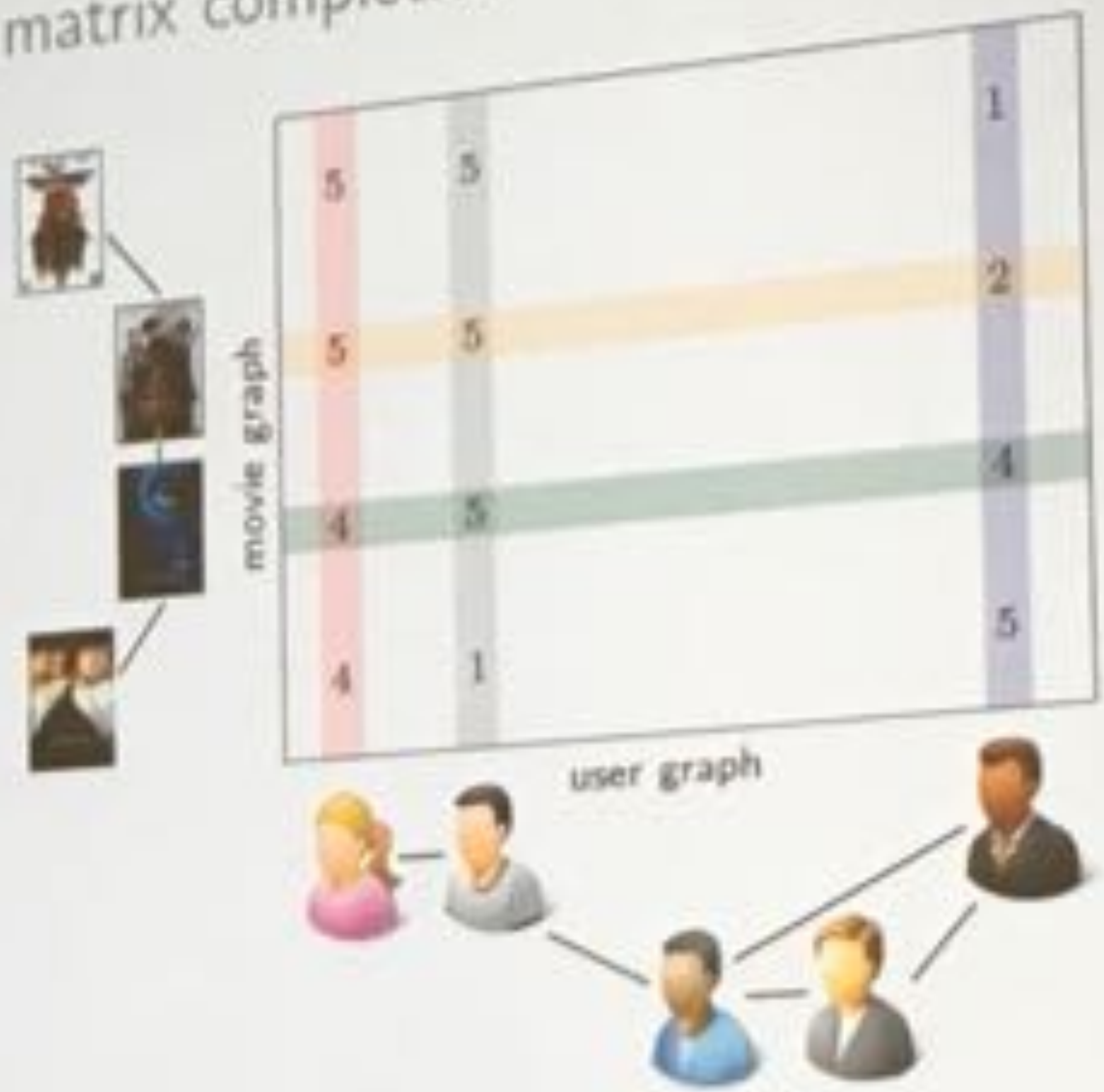
Geometric matrix completion



$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \mu\|\boldsymbol{\Omega}\circ(\mathbf{X}-\mathbf{A})\|_F^2 + \mu_c \operatorname{tr}(\mathbf{X}\boldsymbol{\Delta}_c\mathbf{X}^\top)$$

Kalofolias, Bresson, Bronstein, Vandergheynst 2014

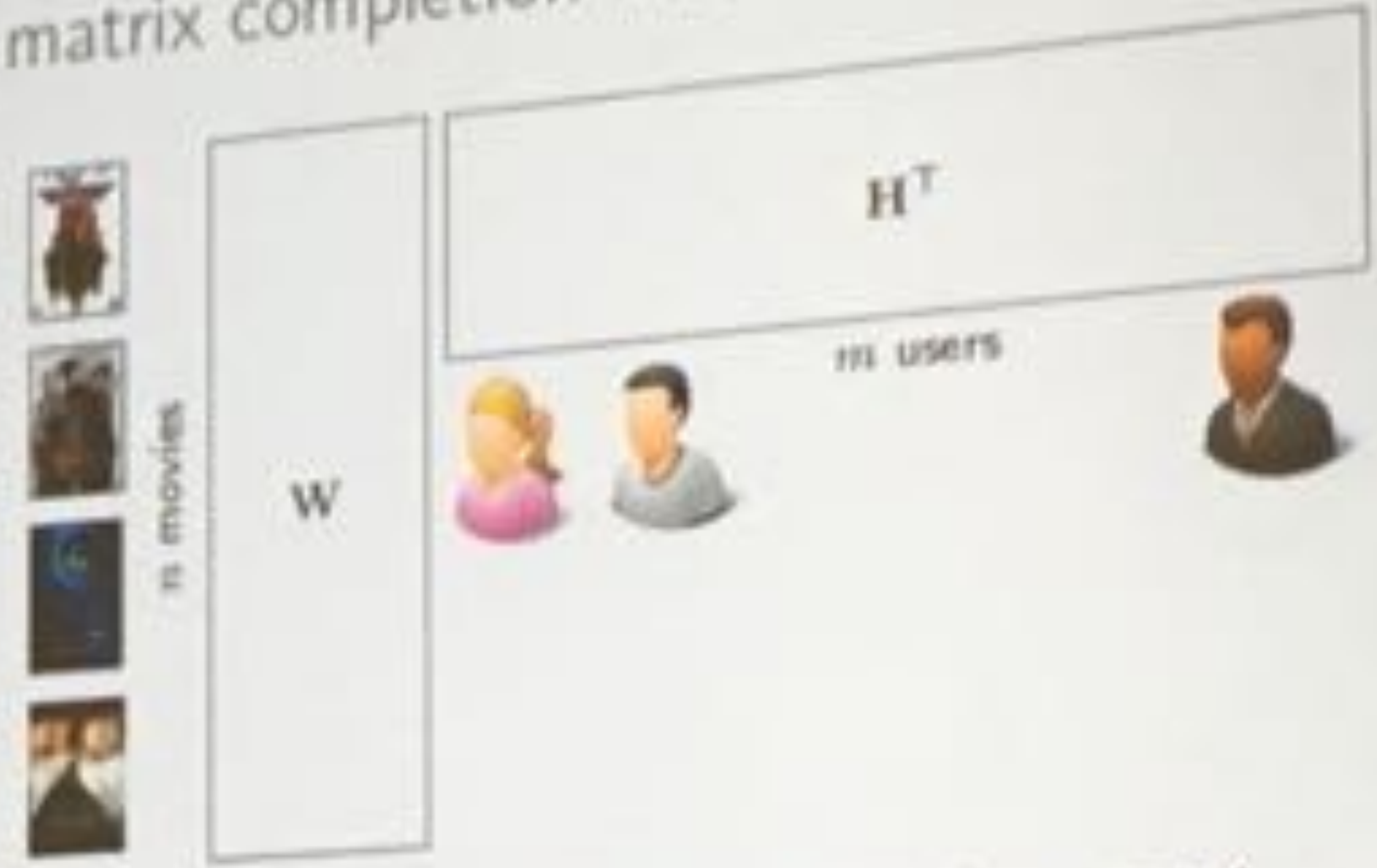
Geometric matrix completion



$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} \mu\|\boldsymbol{\Omega}\circ(\mathbf{X}-\mathbf{A})\|_F^2 + \mu_c \underbrace{\mathrm{tr}(\mathbf{X}\boldsymbol{\Delta}_c\mathbf{X}^\top)}_{\|\mathbf{X}\|^2_{\mathcal{G}_c}} + \mu_r \underbrace{\mathrm{tr}(\mathbf{X}^\top\boldsymbol{\Delta}_r\mathbf{X})}_{\|\mathbf{X}\|^2_{\mathcal{G}_r}}$$

Kalofolias, Bresson, Bronstein, Vandergheynst 2014

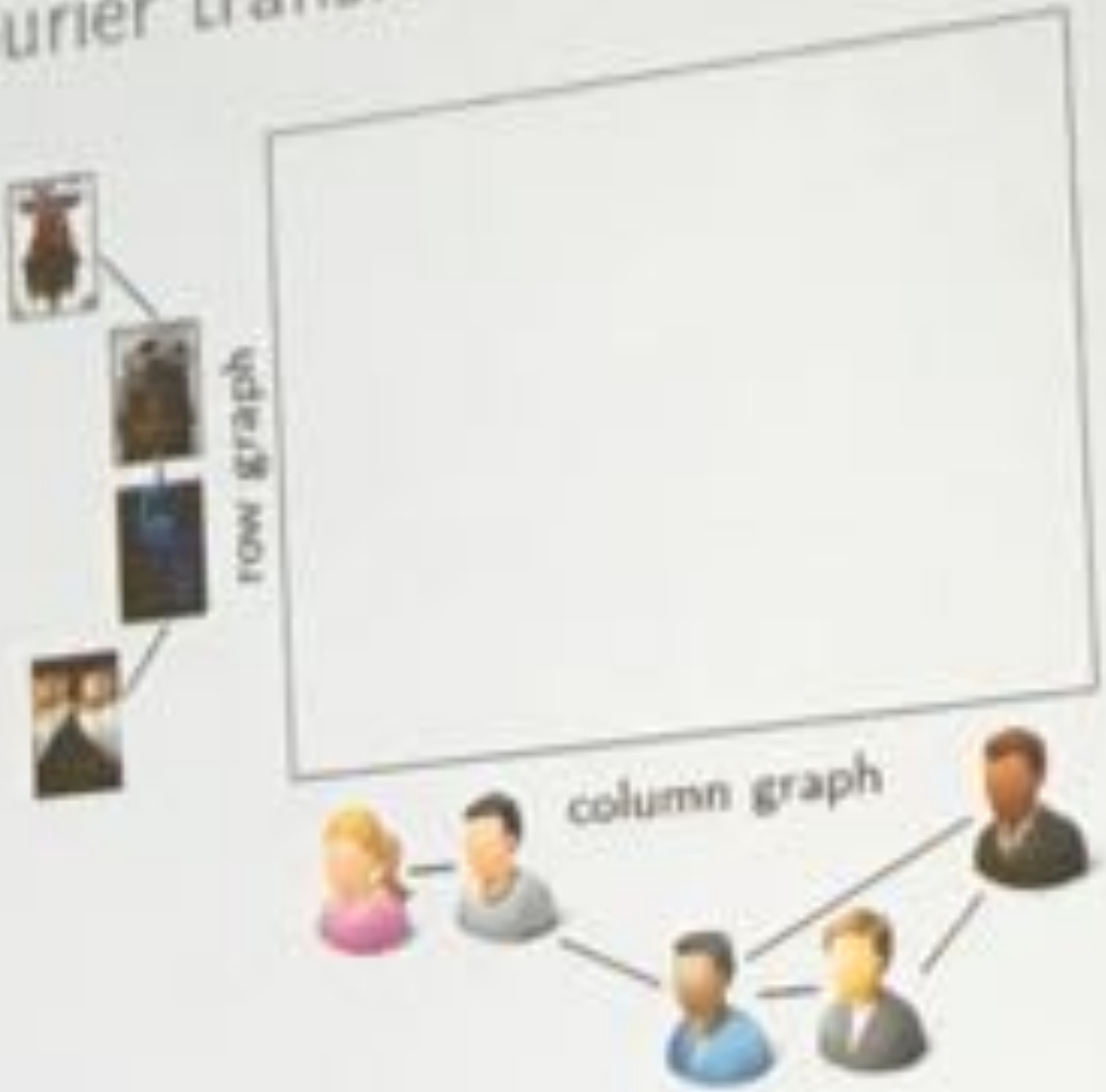
Factorized matrix completion models



$$\min_{\substack{\mathbf{W}\in\mathbb{R}^{m\times r}\\ \mathbf{H}\in\mathbb{R}^{n\times r}}} \mu\|\mathbf{\Omega}\circ(\mathbf{X}-\mathbf{A})\|_F^2 + \mu_c\|\mathbf{W}\|_F^2 + \mu_r\|\mathbf{H}\|_F^2$$

Srebro, Rennie, Jaakkola 2004

Multi-graph Fourier transform



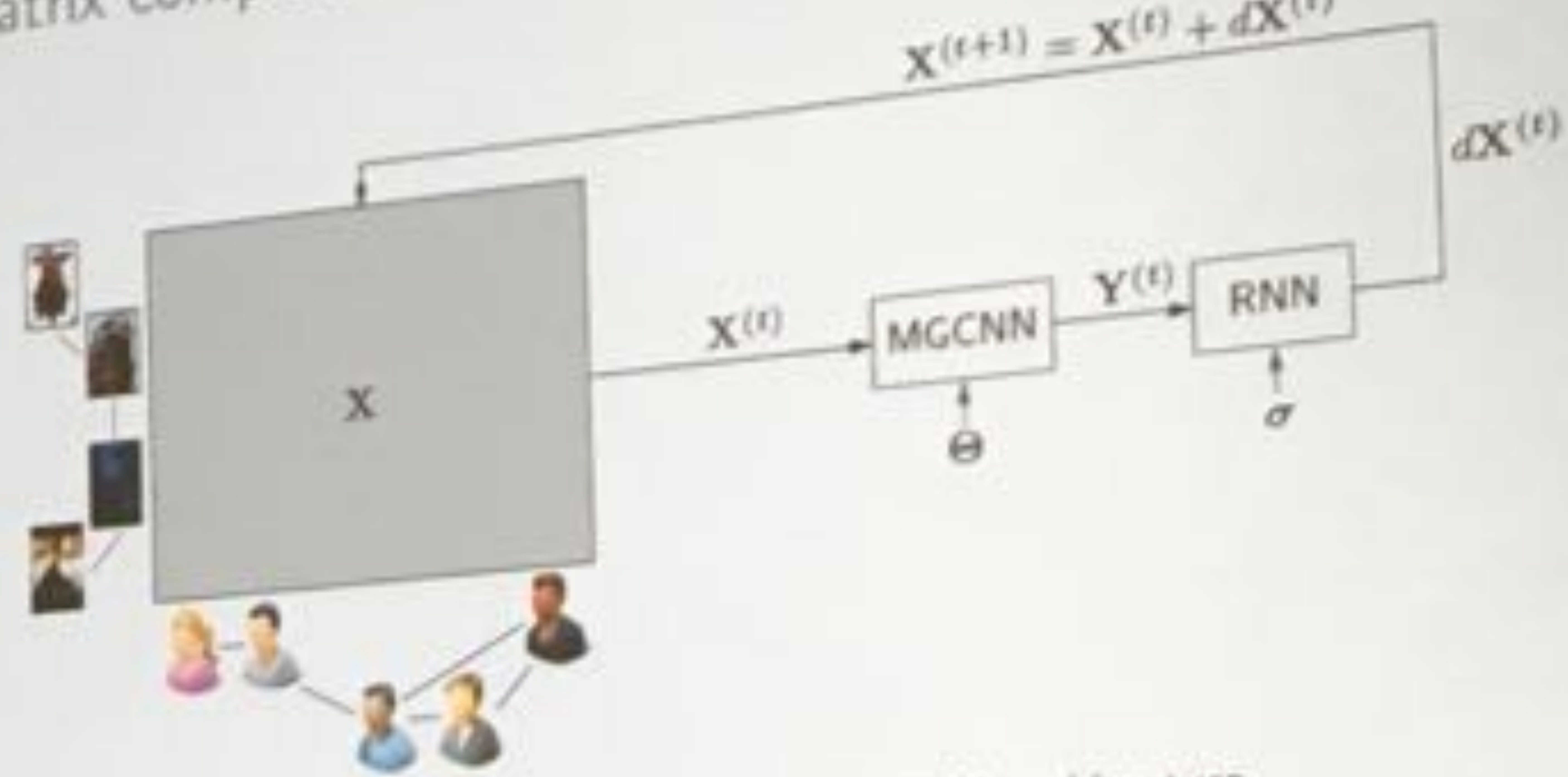
Multi-graph Fourier transform

$$\hat{X} = \Phi_r^\top X \Phi_c$$

where $\boldsymbol{\Phi}_c$ and $\boldsymbol{\Phi}_r$ are the eigenvectors of the column- and row-graph Laplacians $\boldsymbol{\Delta}_c$ and $\boldsymbol{\Delta}_r$, respectively

Monti, Bresson, Bronstein 2017

Matrix completion with Recurrent Multi-Graph CNN



Recurrent multigraph CNN (RMCNN) architecture
for matrix completion

$$\min_{\Theta,\sigma} \|\mathbf{X}^{(T)}_{\Theta,\sigma}\|^2_{\mathcal{G}_r} + \|\mathbf{X}^{(T)}_{\Theta,\sigma}\|^2_{\mathcal{G}_c} + \frac{\mu}{2}\|\boldsymbol{\Omega} \circ (\mathbf{X}^{(T)}_{\Theta,\sigma} - \mathbf{A})\|^2_{\mathrm{F}}$$

Monti, Bresson, Bronstein 2017

Matrix completion methods comparison

Method	MovieLens[1]	Flixster[2]	Douban[3]	Yahoo[4]
IMC[5]	1.653	–	–	–
GMC[6]	0.996	–	–	–
MC[7]	0.973	–	–	–
GRALS[8]	0.945	1.245	0.833	38.042
sRGCNN (Cheb)[9]	**0.929**	**0.926**	**0.801**	**22.415**
sRGCNN (Cayley)[10]	**0.922**	–	–	–

Accuracy (RMS error) of matrix completion methods on real data

Data: [1]Miller et al. 2003; [2]Jamali, Ester 2010; [3]Ma et al. 2011; [4]Dror et al. 2012
Methods: [5]Jain, Dhillon 2013; [6]Kalofolias, Bresson, Bronstein, Vandergheynst 2014; [7]Candes, Recht 2012; [8]Rao et al. 2015; [9]Monti, Bresson, Bronstein 2017; [10]Levie, Monti, Bresson, Bronstein 2017

Quadratic Assignment Problem

Find a node correspondence that minimizes a transportation cost between two graphs $\mathcal{G}_1$, $\mathcal{G}_2$

$$\min_{\mathbf{P}\in\Pi_n} \mathrm{tr}(\mathbf{W}_1\mathbf{P}\mathbf{W}_2\mathbf{P}^\top), \ \mathbf{W}_i = \text{adjacency matrix of graph } \mathcal{G}_i$$

$$\Pi_n = \text{space of } n \times n \text{ permutation matrices}$$


Why TAE?
1. Rapid experimentation
5 machines, 70,000+ experiments so far, once/10 minutes
3,000 experiments on Norman since July
2. Data so far suggests good scaling law
Scaling law: how difficult is it to get energy out from fusion?


Looking forward
In 2018: **Science**
Maximize electron temperature, verify scaling to high temperatures
More precise estimate of Q at commercial scale
In 2019 and beyond: **Engineering at scale**
Demonstrate Q > 1
Get estimated costs for fusion power