Representation learning and fusion of multi-modal data for the statistical analysis of medical imaging populations

Nicolas Duchateau

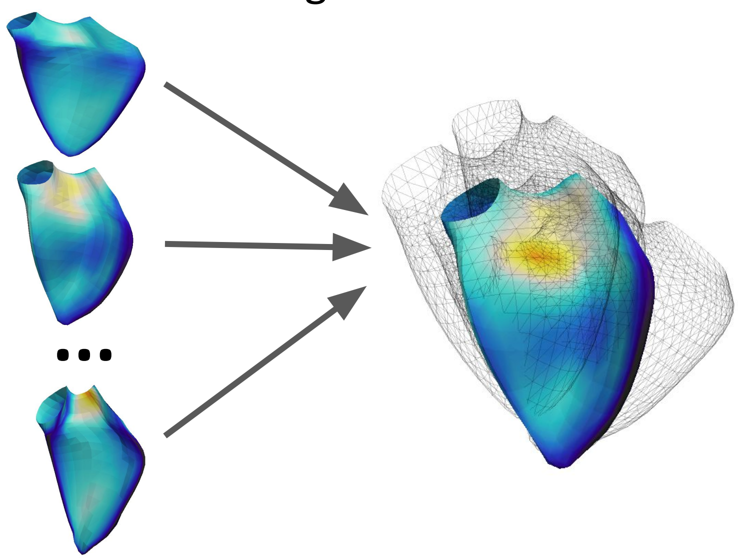
CREATIS lab, Université Lyon 1, France
Institut Universitaire de France

GdR ISIS - Apprentissage automatique multimodal et fusion d'informations (2ième édition) - 19/01/2022

About (ND)

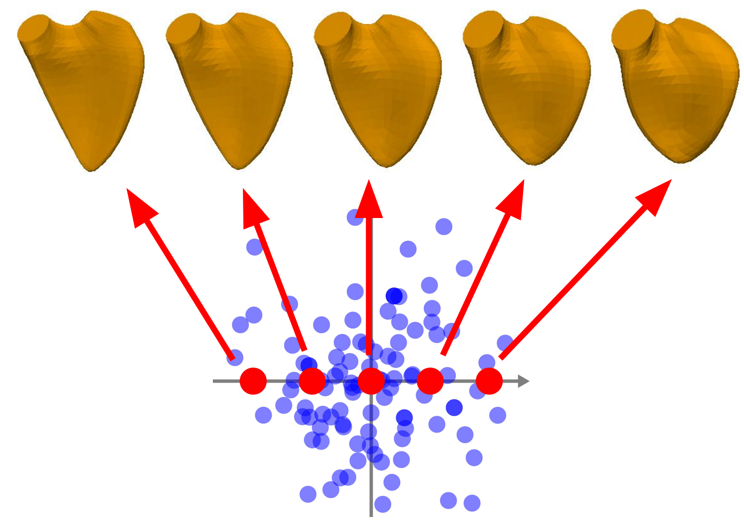
Statistical atlases

= “aligned” data



Representation learning

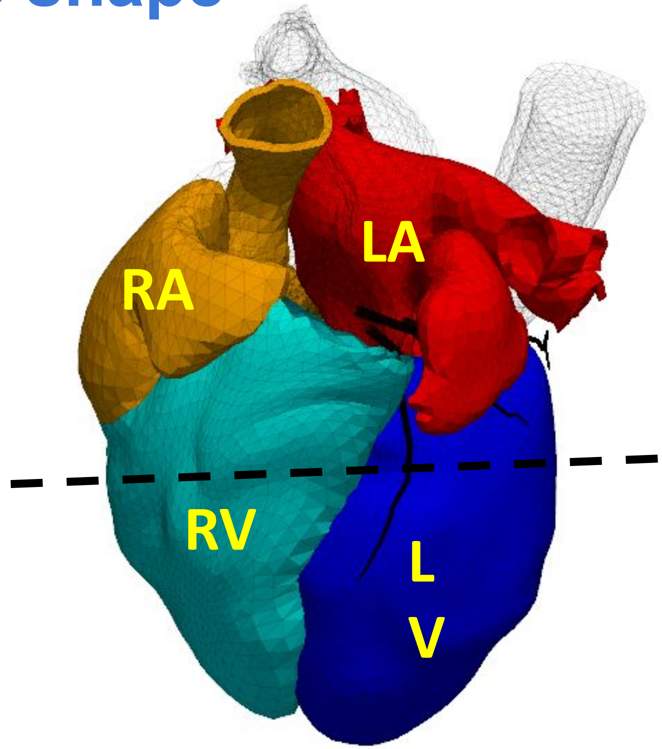
= “simplified” view of a population



Clinical (cardiac) questions

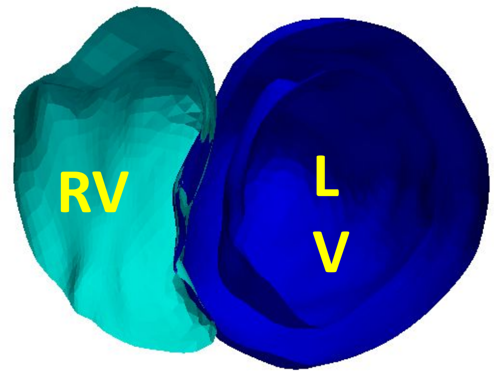
- Abnormality quantification
- Phenotyping
- Longitudinal follow-up

Cardiac shape



3D atlas from:

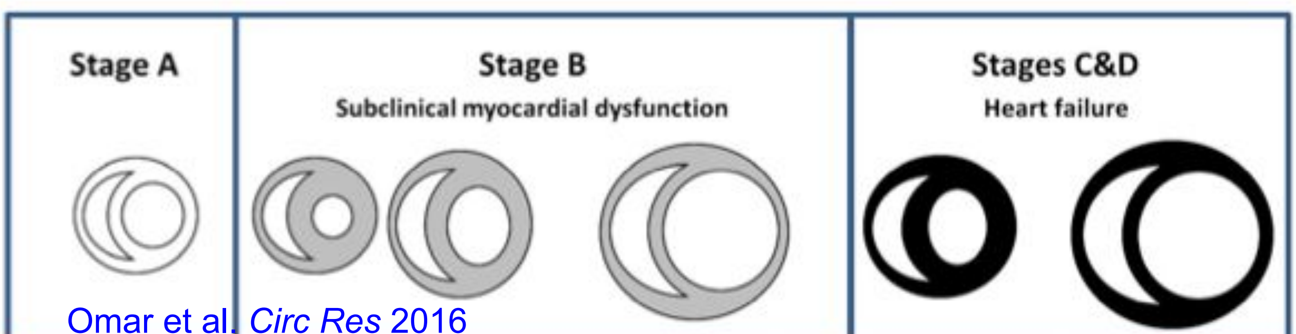
Hoogendoorn et al. *IEEE T Med Imaging* 2013



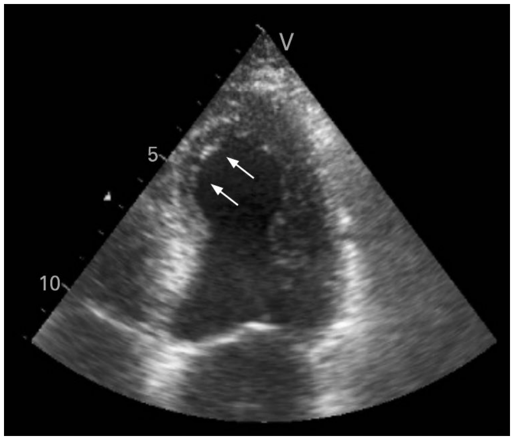
local remodelling

ex: infarct

Disease: global remodelling



Omar et al. *Circ Res* 2016



Cardiac dynamics

Displacement	$0 \rightarrow t$	vector
Velocity	$t \rightarrow t+1$	vector
Strain	$0 \rightarrow t$	tensor
Strain rate	$t \rightarrow t+1$	tensor

Deformation:

- ≠ motion
- partially reflects contractility
- tensor… often simplified

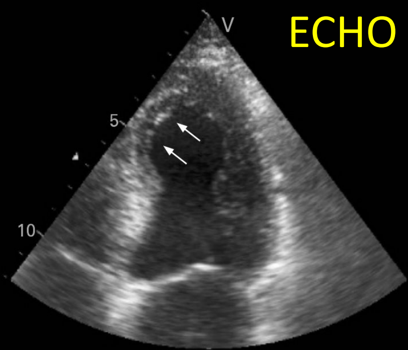
Eject enough blood... = ejection fraction $\frac{V_{ED} - V_{ES}}{V_{ED}}$

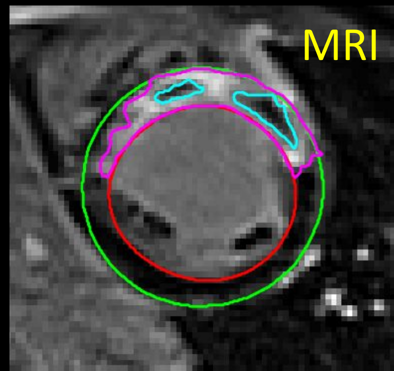
+ complementary factors:

- Interactions between shape + fibers vs. deformation
- Valves / flow
- Electrical activation
- Patient characteristics (age, sex, etc.)
- External factors, effort, etc.

Cardiac diseases

Imaging **is useful…**



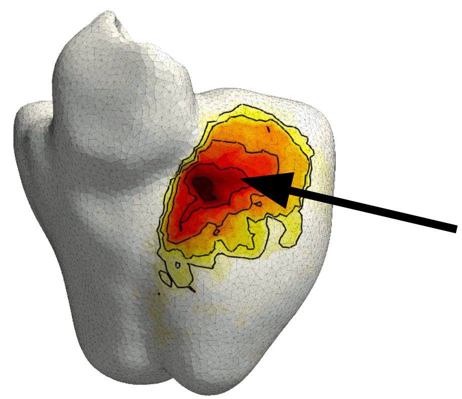


… but **“truncated”** analysis in the clinic !

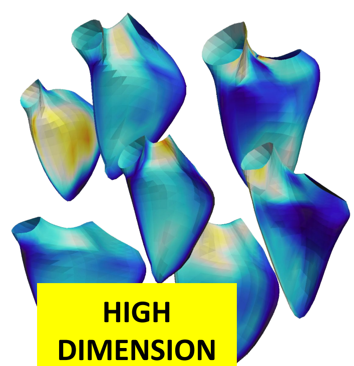
Scalars values:
size, transmurality

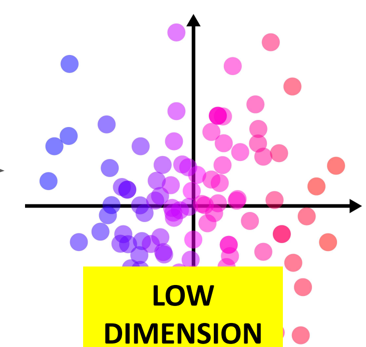
Where can we contribute ?

→ Need for **precise AND reliable biomarkers**



→ And a **statistically-relevant framework** to analyze them

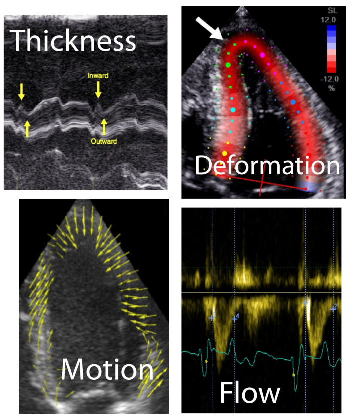




Representation learning

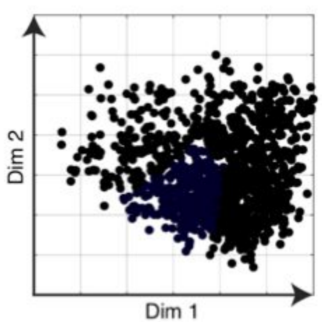
CREATIS UNIVERSITÉ DE LYON

Idea = better represent the data space
- lower dimensional space
- unsupervised (?)



Inputs =
- **Single** **high-dimensional** descriptors
- **Multiple** scalars

...or **Multiple** **high-dimensional** descriptors

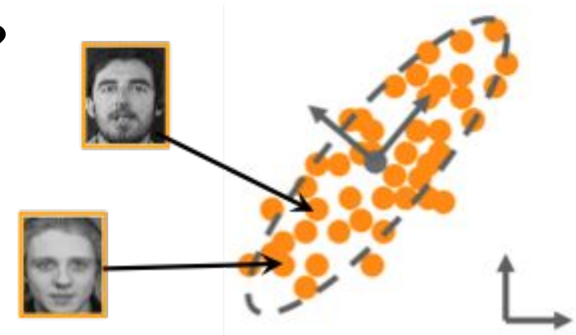


Output =
- **Low-dimensional** representation

Representation learning

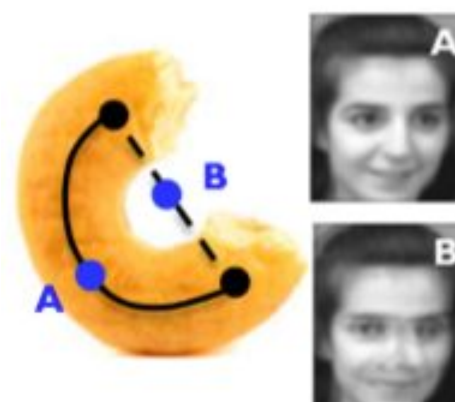
<u>Idea</u> = better represent the data space
- lower dimensional space
- unsupervised (?)

Linear ?



ex: <u>average face</u> for a given person = ?

Non-linear !!!

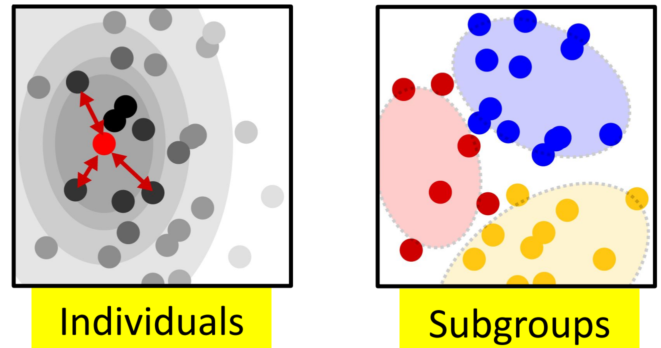


ORL database

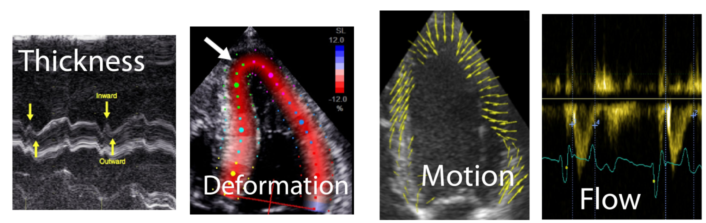
Representation learning

Medical images/descriptors = warnings !!!

- **Non linear** data space
- **Exploitable** latent space
  - Distances
  - Reconstruction (interpretation)



- **Unsupervised**: validation = ?
- **Heterogeneous** types of data



- Not millions of samples
  - No deep?
  - Pre-normalization?
  - Better metrics?

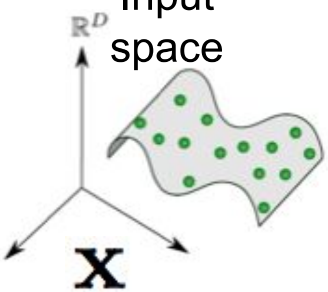
Representation learning pipeline

CREATIS UNIVERSITÉ DE LYON

Dimensionality reduction: $D \gg d$

Input space   Latent space

Input space

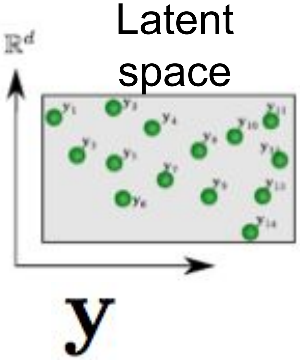


Inputs =
- Single **high-dimensional** descriptors
- **Multiple** scalars
  ...or **Multiple high-dimensional** descriptors

1) Embedding / encoding

3) Reconstruction / decoding

Latent space



Output =
- **Low-dimensional** representation

2) Manifold / latent space

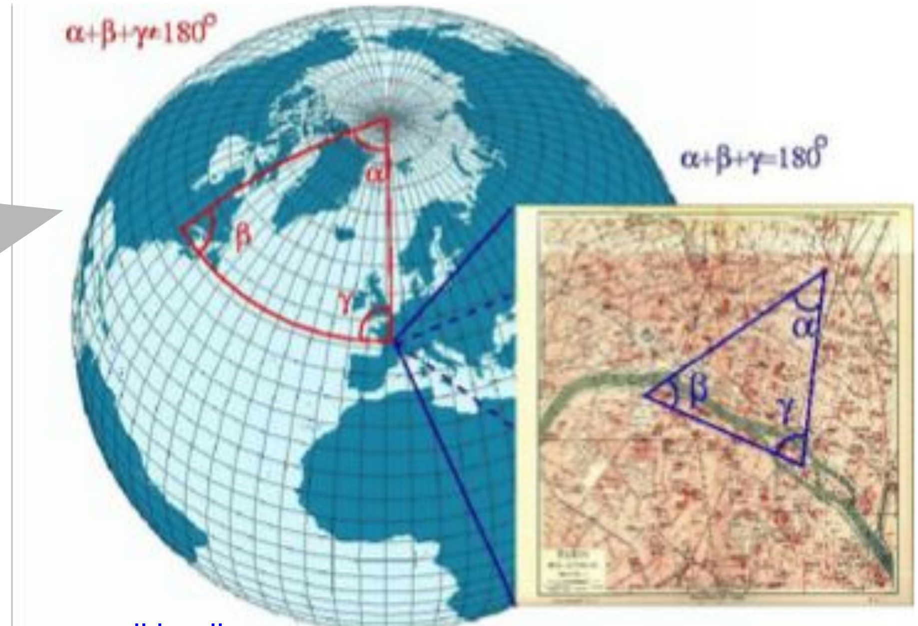
Manifold

Manifold of dimension *N* = topological space that near each sample resembles (is homeomorphic to) a *N*-dimensional Euclidean space

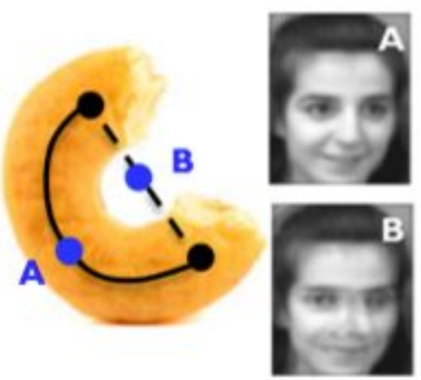
ex: **- lines and circles (*N*=1)**

- plane, sphere, surfaces (*N*=2)

- brain images, cardiac shapes (*N*=?)



www.wikipedia.org



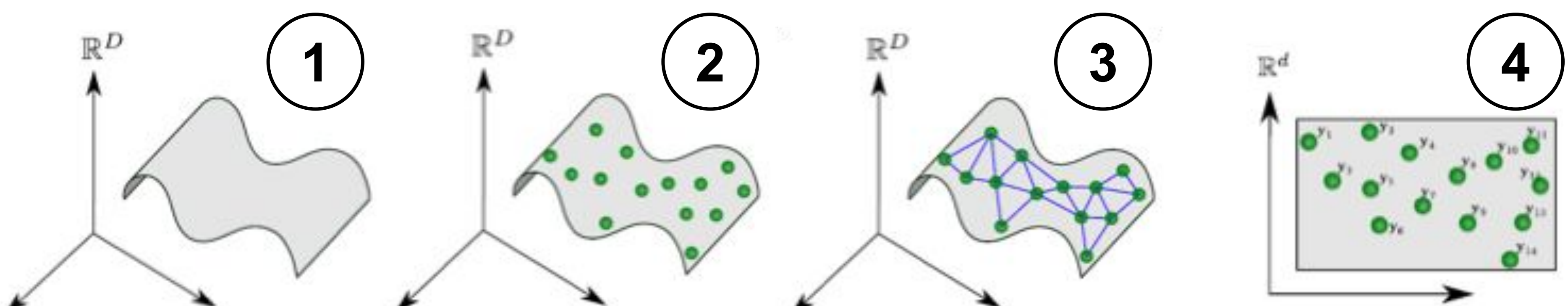
Interesting for medical data:
(specific) distance in the high-dimensional space
= Euclidean distance in the low-dimensional space

Manifold learning

<u>Manifold of dimension *N*</u>

- In some cases = **known structure**
- **Otherwise = <u>learn</u> it from data !**

1. Assumption = data **lies on / close to a manifold**
2. **Few samples** (on the manifold) available
3. **Neighborhood graph** = approximation of the manifold
4. Dimensionality reduction = **spectral decomposition** of…?

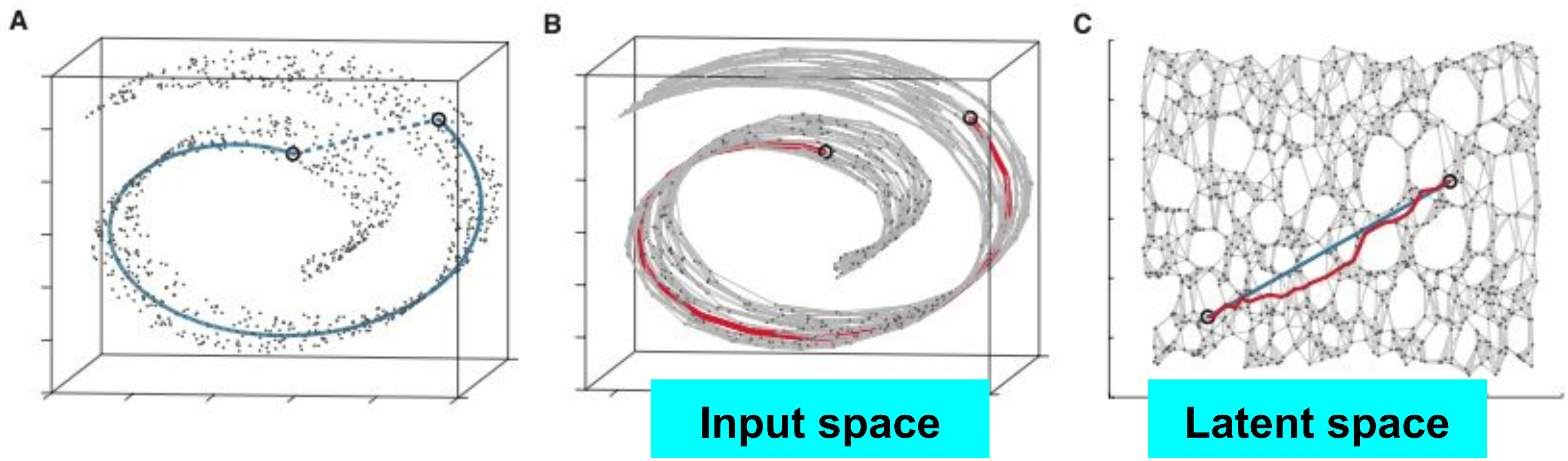


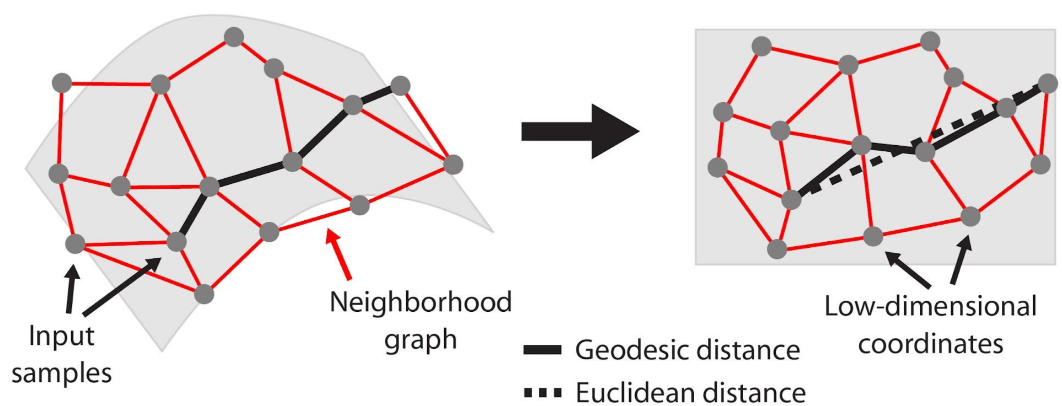
Atasoy & Mateus *MICCAI tutorial* 2011

Isomap

Tenenbaum et al. *Science* 2002

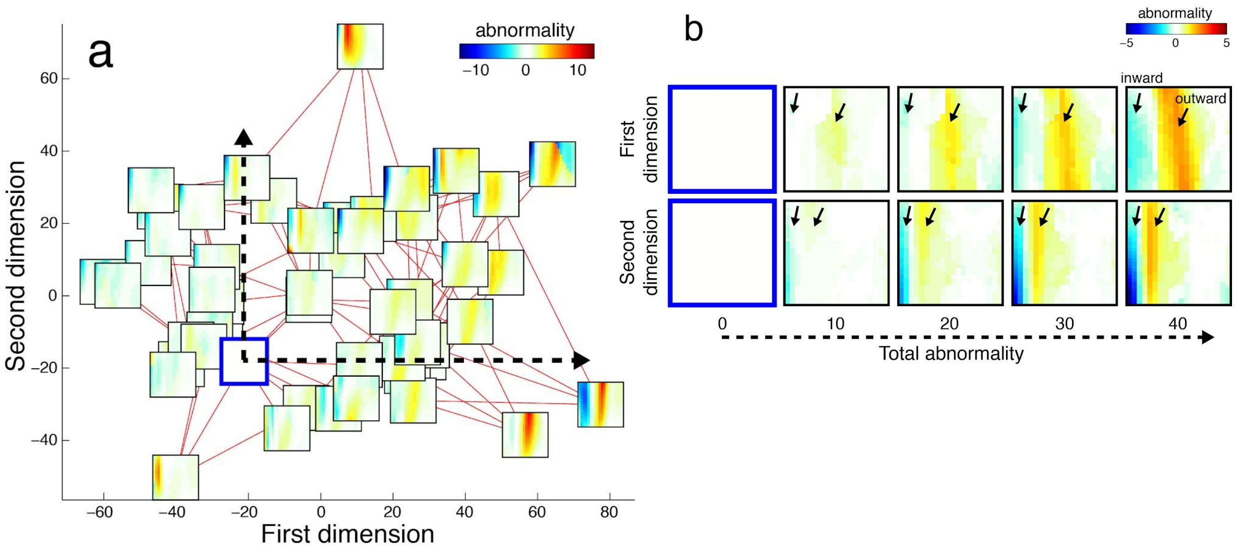
Idea = approximate **geodesic distances** / shortest path along the graph





Isomap

Idea = approximate **geodesic distances** / shortest path along the graph

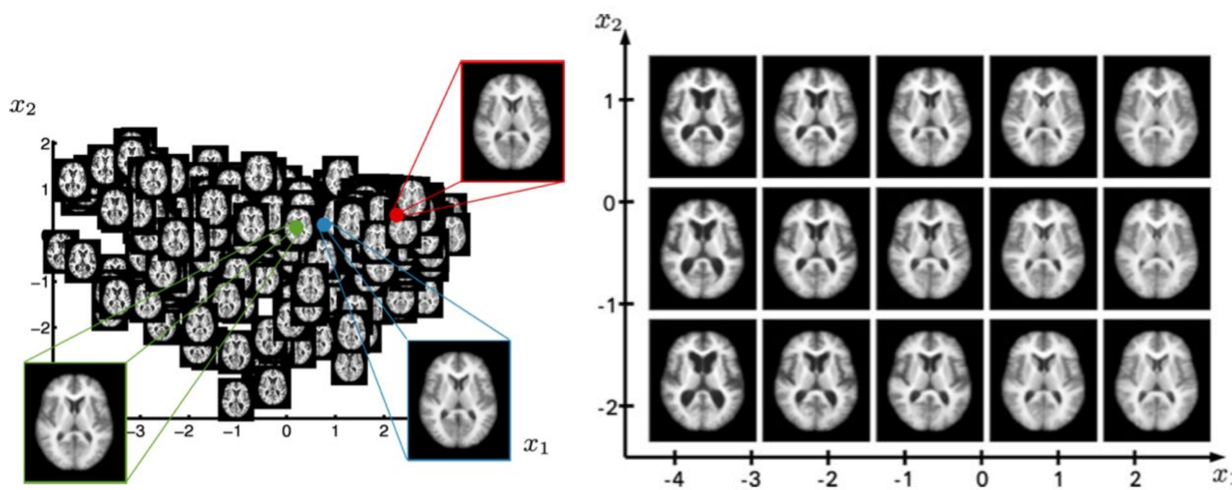


ex: disease evolution
(cardiac velocity patterns)

Duchateau et al. *Med Image Anal* 2012

ex: preprocessing for regression (brain images)

Gerber et al. *Med Image Anal* 2010



Laplacian eigenmaps

Belkin & Niyogi *Neur Comput* 2003

<u>Idea</u> = diagonalize the **graph Laplacian**

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\mathrm{argmin}} \sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \underset{\mathbf{Y}}{\mathrm{argmin}}\, \mathbf{Y}^T \mathbf{L} \mathbf{Y}$$

Constraint: $\mathbf{Y}^T \mathbf{D} \mathbf{Y} = 1$

Graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$

<u>Parameter</u> = kernel bandwidth

$$w_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$$

- Close inputs ➔ $w_{ij} \approx 1$ ➔ close outputs
- Far inputs ➔ $w_{ij} \approx 0$ ➔ minor influence

Laplacian eigenmaps

CREATIS UNIVERSITÉ DE LYON

Belkin & Niyogi *Neur Comput* 2003

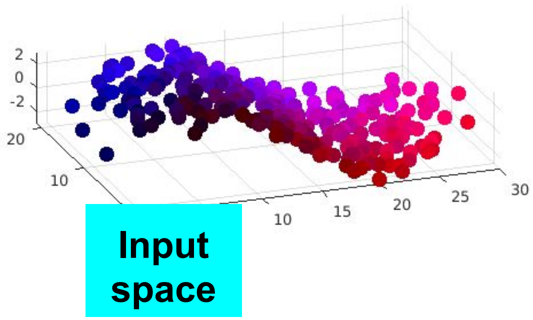
<u>Idea</u> = diagonalize the **graph Laplacian**

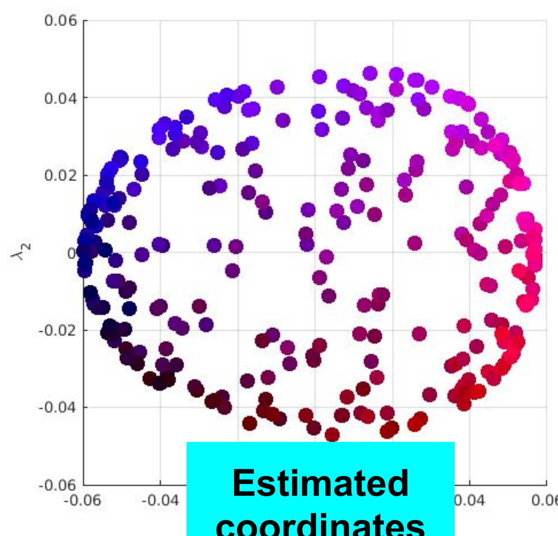
$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\operatorname{argmin}} \sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \underset{\mathbf{Y}}{\operatorname{argmin}} \mathbf{Y}^T \mathbf{L} \mathbf{Y}$$

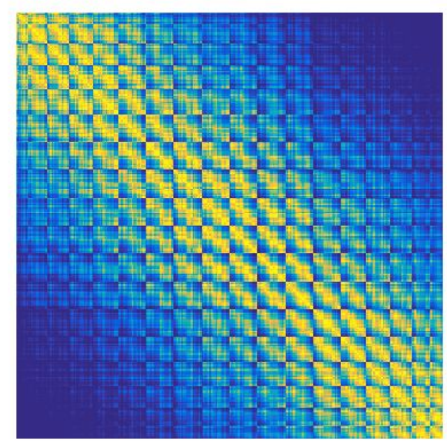
<u>Eigendecomposition problem</u>:

$$\mathbf{L}\mathbf{v} = \lambda \mathbf{D}\mathbf{v}$$

$$\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}\mathbf{v} = (1-\lambda)\mathbf{v}$$





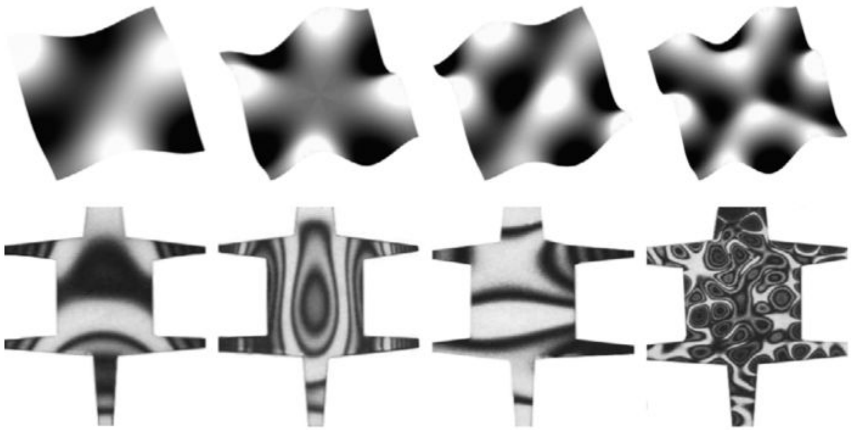


W

Laplacian eigenmaps

Imaged vibrations (interferometry)

Xu et al. *Appl Optics* 1983

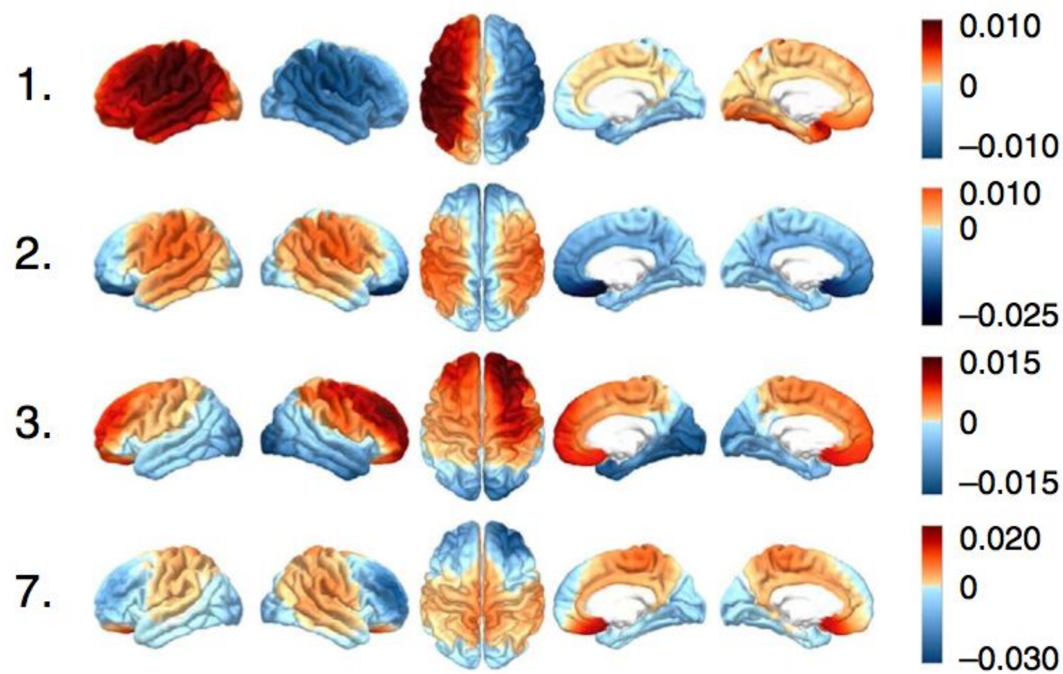


Rectangular metal plates

Shaped metal plates

Learnt embedding:

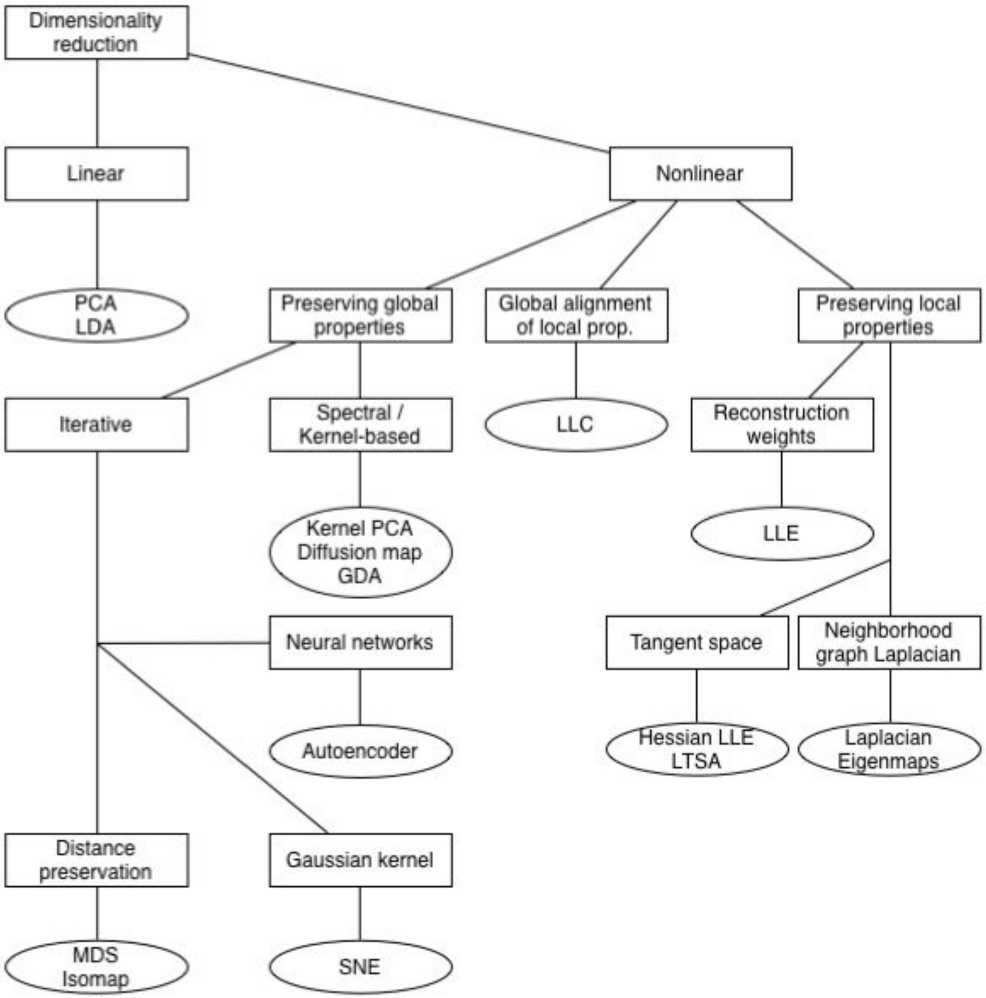
ex: connectome harmonics



Atasoy et al. *Nat Comm* 2016

… and many other algorithms !!!

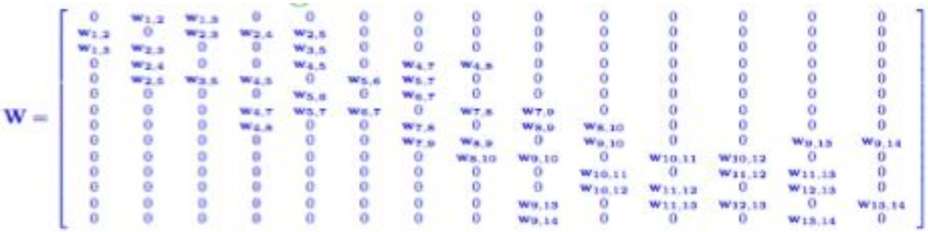
van der Maaten et al. 2009

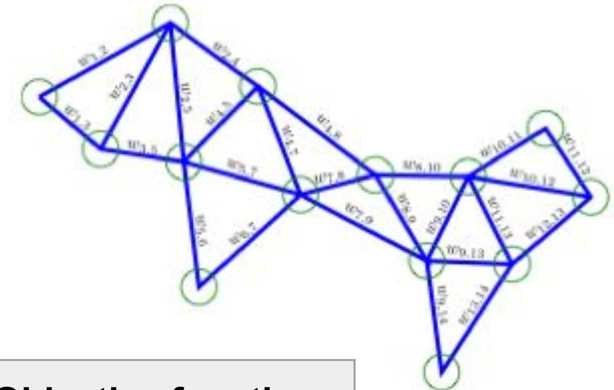


Depends on:

- Your knowledge on data
- Your objectives (distance=?)

Unified framework?





Method	**Operator / Matrix**	**Preserved**	**Objective function**
PCA	Covariance matrix	Variance of the dataset / Euclidean distances between data points	$\mathbf{u}^T \Sigma \mathbf{u}$
Laplacian eigenmaps	Graph Laplacian	Distances within the local neighborhood of each data point	$\mathbf{u}^T L \mathbf{u}$
Isomap	Geodesic distance matrix	Geodesic distances between data points	$\mathbf{u}^T D_G \mathbf{u}$
LLE	Reconstruction weighs	Reconstruction weights within the local neighborhood of each data point	$\mathbf{u}^T W \mathbf{u}$

Atasoy & Mateus *MICCAI tutorial* 2011

Unified framework?

CREATIS UNIVERSITÉ DE LYON

Yan et al. *IEEE PAMI* 2007

(cf. Laplacian eigenmaps…)

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\operatorname{argmin}} \sum_{i,j} \boxed{w_{ij}} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \underset{\mathbf{Y}}{\operatorname{argmin}} \mathbf{Y}^T \mathbf{L} \mathbf{Y}$$

Unified framework?

CREATIS UNIVERSITÉ DE LYON

Yan et al. *IEEE PAMI* 2007

(cf. Laplacian eigenmaps…)

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\operatorname{argmin}} \sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \underset{\mathbf{Y}}{\operatorname{argmin}} \mathbf{Y}^T \mathbf{L} \mathbf{Y}$$

Under the constraint:

Supervised

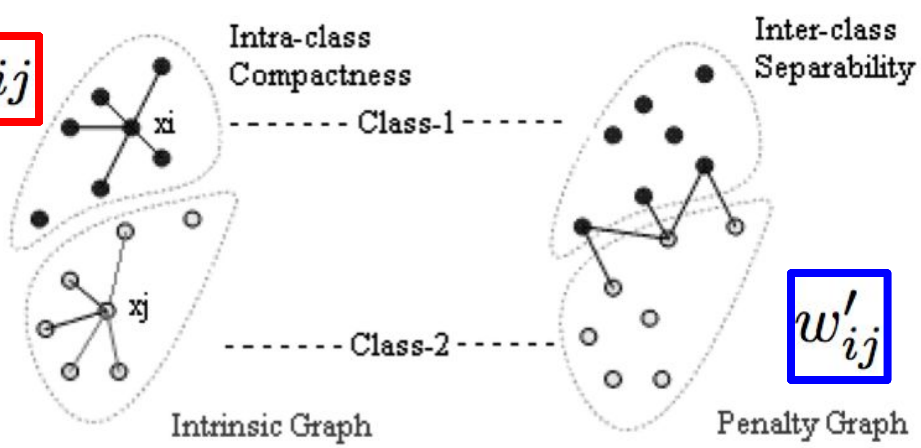
$$\sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 w'_{ij} = 1$$

Unsupervised

$$\sum_i \|\mathbf{y}_i\|^2 d_{ii} = 1$$

$$d_{ii} = \sum_j w_{ij}$$

w_{ij}



w'_{ij}

Unified framework?

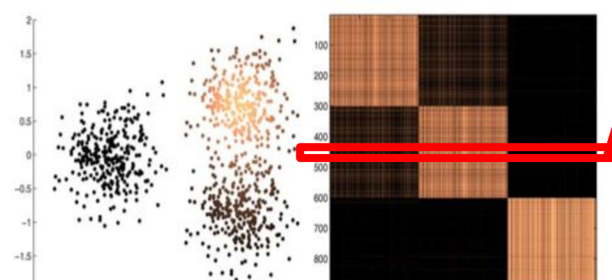
Yan et al. *IEEE PAMI* 2007

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}}{\operatorname{argmin}} \sum_{i,j} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \underset{\mathbf{Y}}{\operatorname{argmin}} \mathbf{Y}^T \mathbf{L} \mathbf{Y}$$

Under the constraint:

Supervised	Unsupervised
$\sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 w'_{ij} = 1$	$\sum_i \|\mathbf{y}_i\|^2 d_{ii} = 1$

$$d_{ii} = \sum_j w_{ij}$$

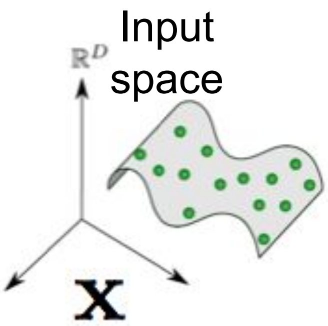


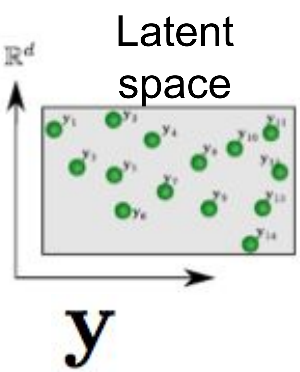
Coifman & Lafon *Appl Comput Harm* 2006

Representation learning pipeline

Dimensionality reduction: $D \gg d$

Input space — Latent space





Inputs =
- Single **high-dimensional** descriptors
- **Multiple** scalars

...or **Multiple high-dimensional** descriptors

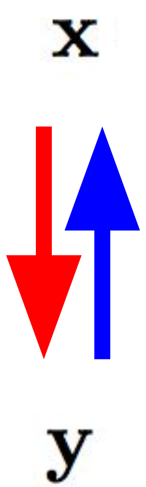
1) Embedding / encoding

3) Reconstruction / decoding = ?

Output =
- **Low-dimensional** representation

2) Manifold / latent space

Reconstruction (a-posteriori)



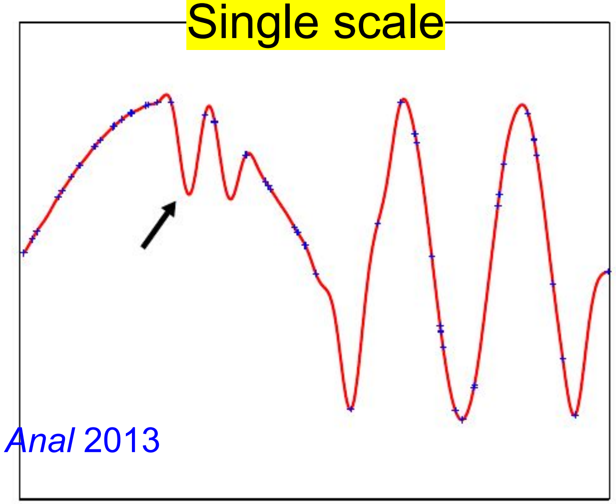
In many cases, no analytical formula !
= out-of-sample extension / interpolation

ex: **kernel** regression:
- Close to existing samples
- Smooth enough

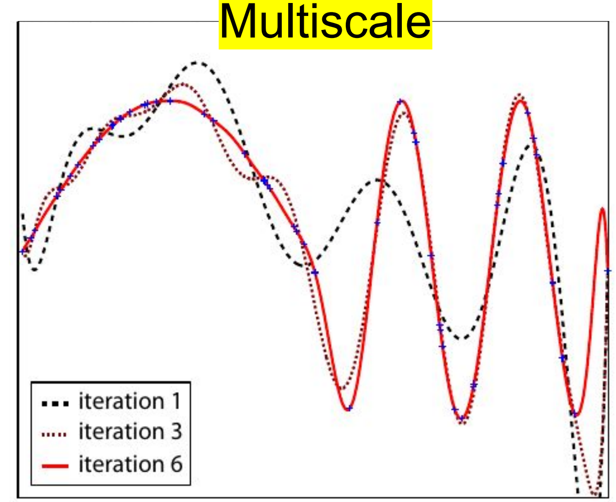
$$\underset{f}{\operatorname{argmin}} \quad \|f\|^2_{\mathcal{F}} + \gamma \sum_{i=1}^{N} \|f(\mathbf{y}) - \mathbf{x}_i\|^2$$

Multiscale version = more robust to density of samples

Single scale

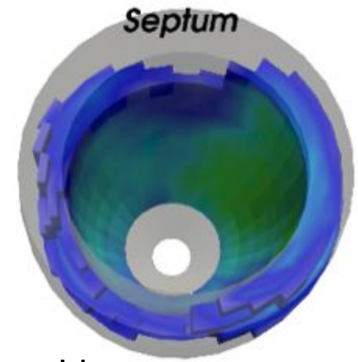


Multiscale



Bermanis et al. *Appl Comput Harm Anal* 2013
Duchateau et al. *GSI* 2013

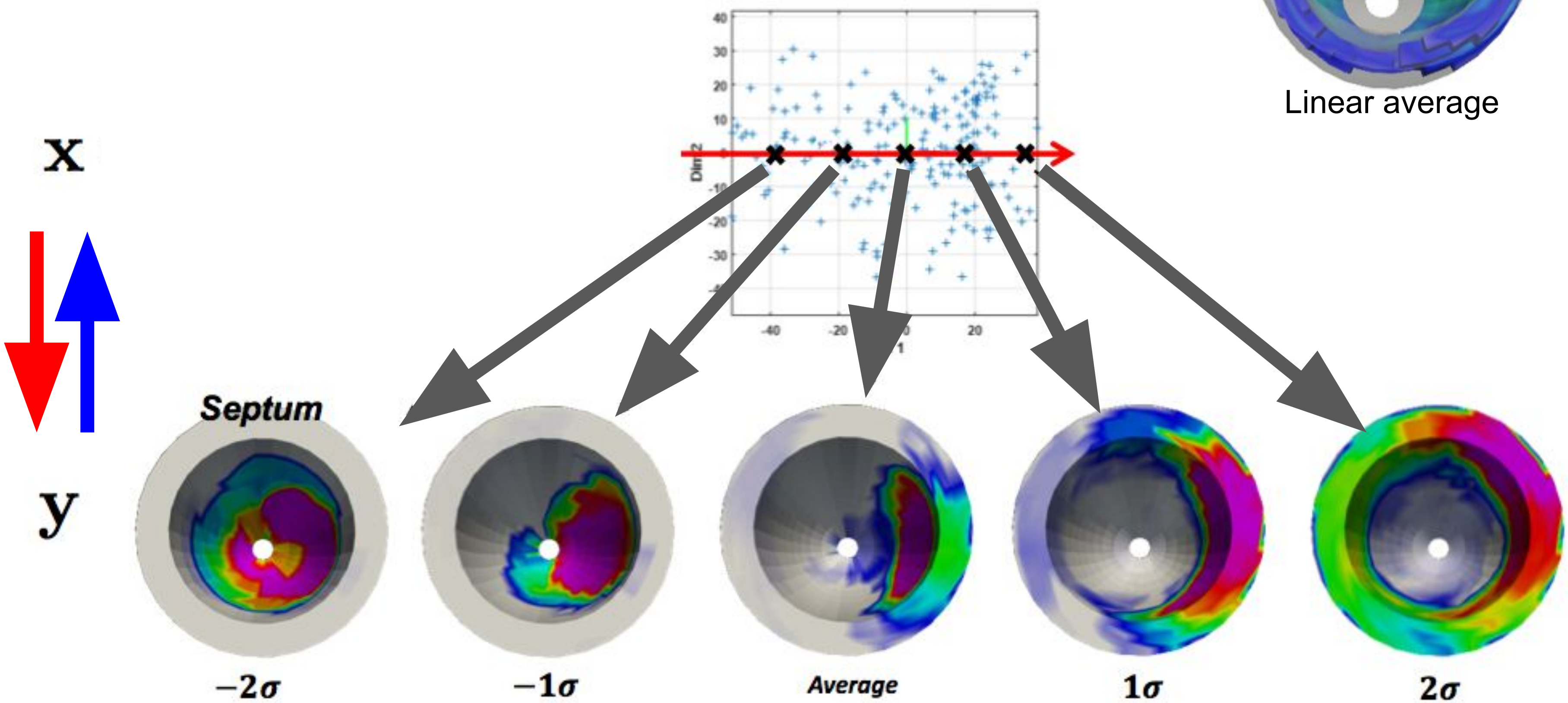
Reconstruction (a-posteriori)



Linear average

ex: variability in myocardial infarcts

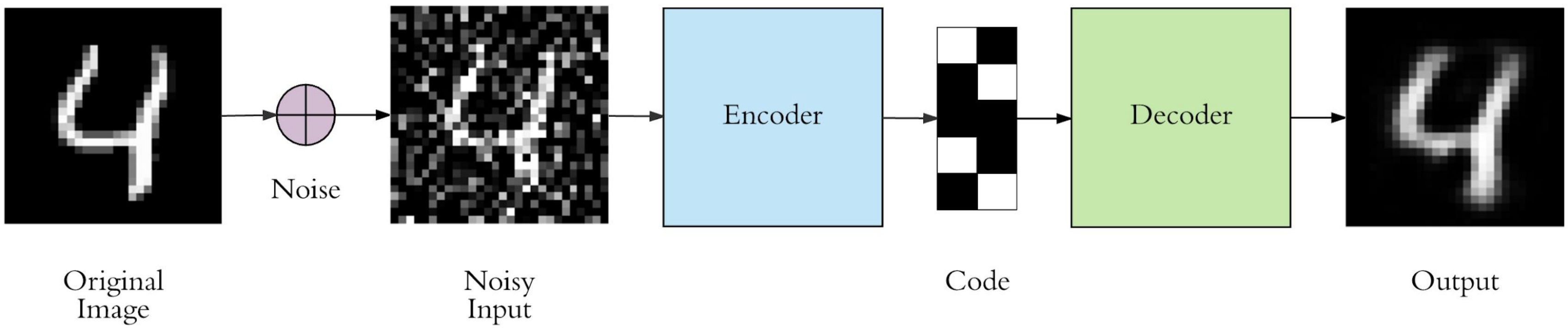
Di Folco et al. *CNIV* 2019



Reconstruction (intrinsic)

→ Autoencoders

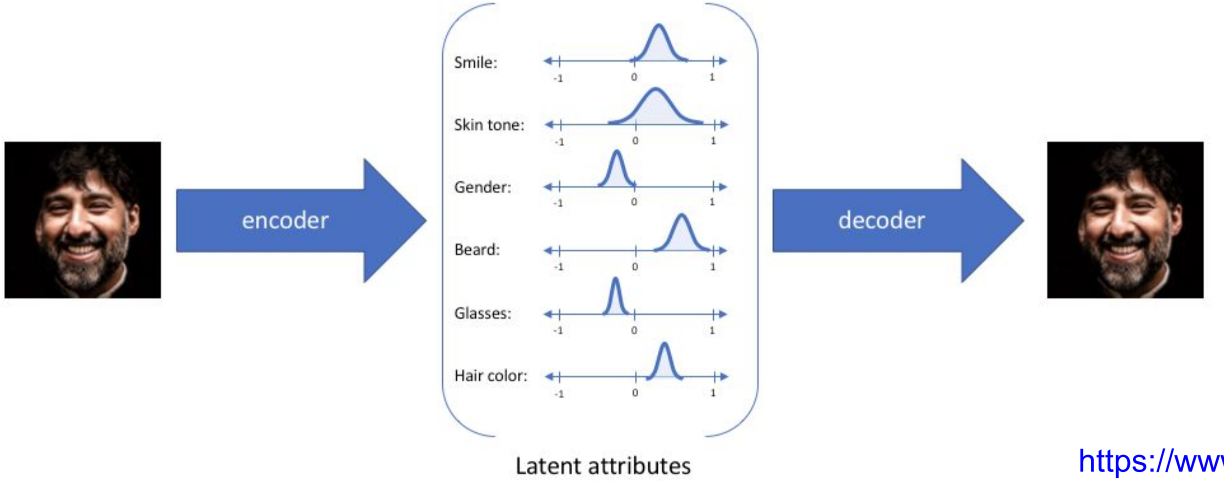
LeCun et al. *PhD* 1987
Hinton & Zemel *NIPS* 1994



https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

→ Variational autoencoders

Kingma & Welling *ICLR* 2014



Latent attributes in probabilistic terms

https://www.jeremyjordan.me/variational-autoencoders/

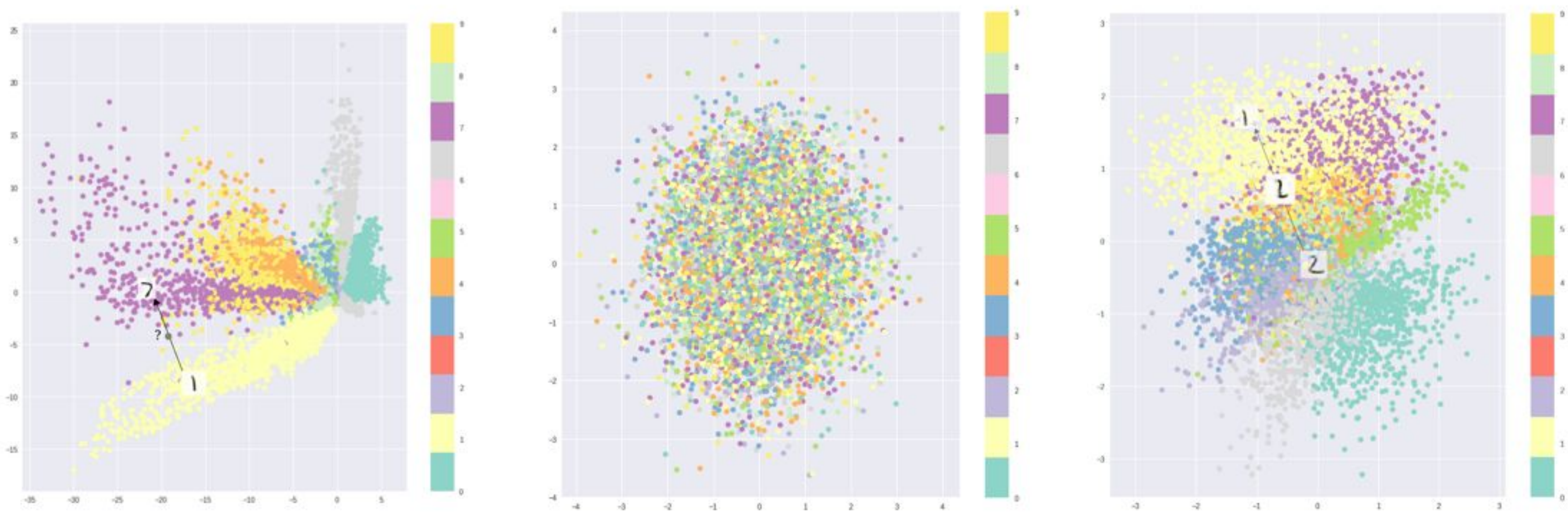
Reconstruction (intrinsic)

Kingma & Welling *ICLR* 2014

→ Variational autoencoders

$$\mathcal{L}(x, \hat{x}) + \sum_j KL\left(q_j(z|x) \,||\, p(z)\right)$$

Only reconstruction loss | Only KL divergence | Combination

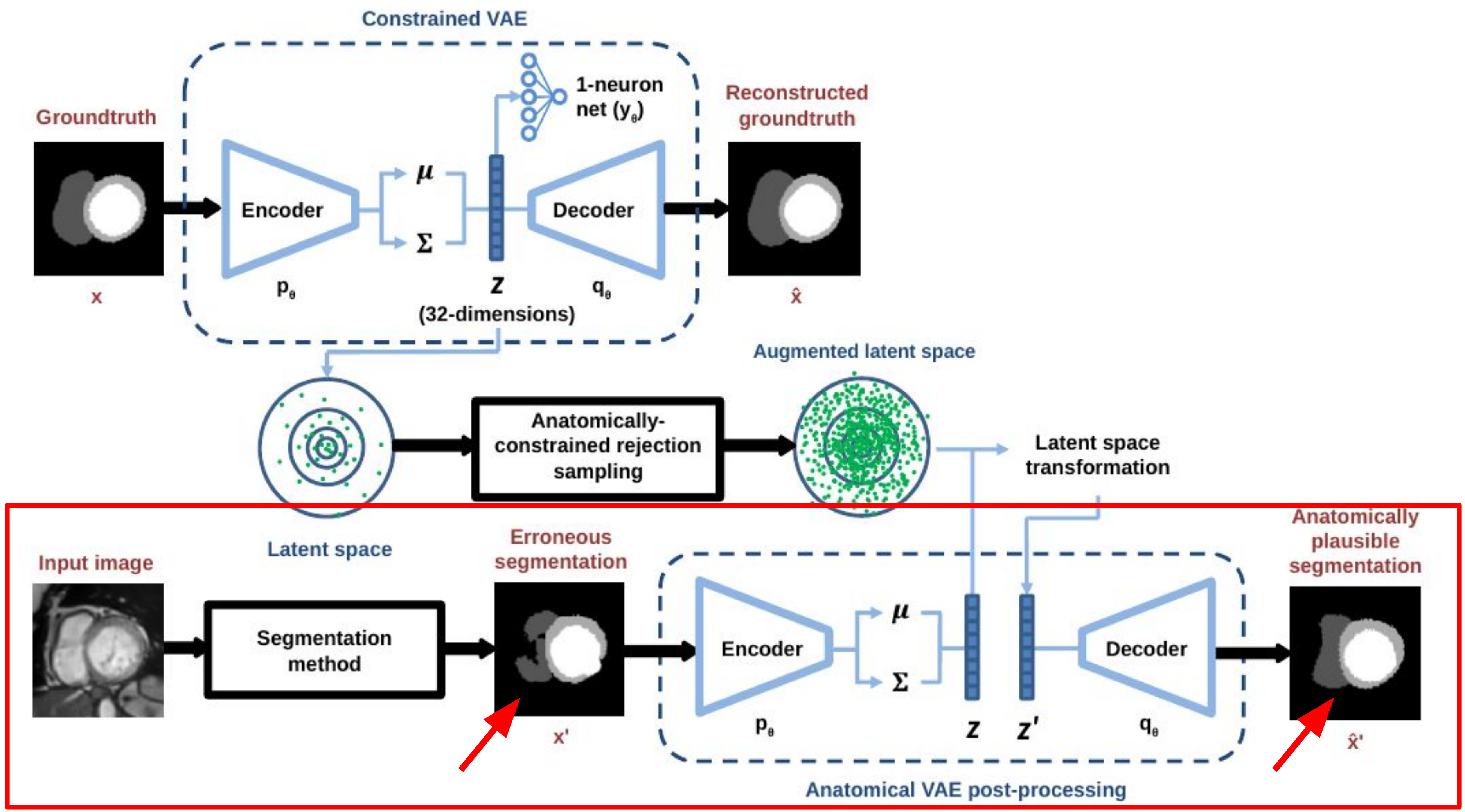


https://www.jeremyjordan.me/variational-autoencoders/

Reconstruction (intrinsic)

ex: variational autoencoders

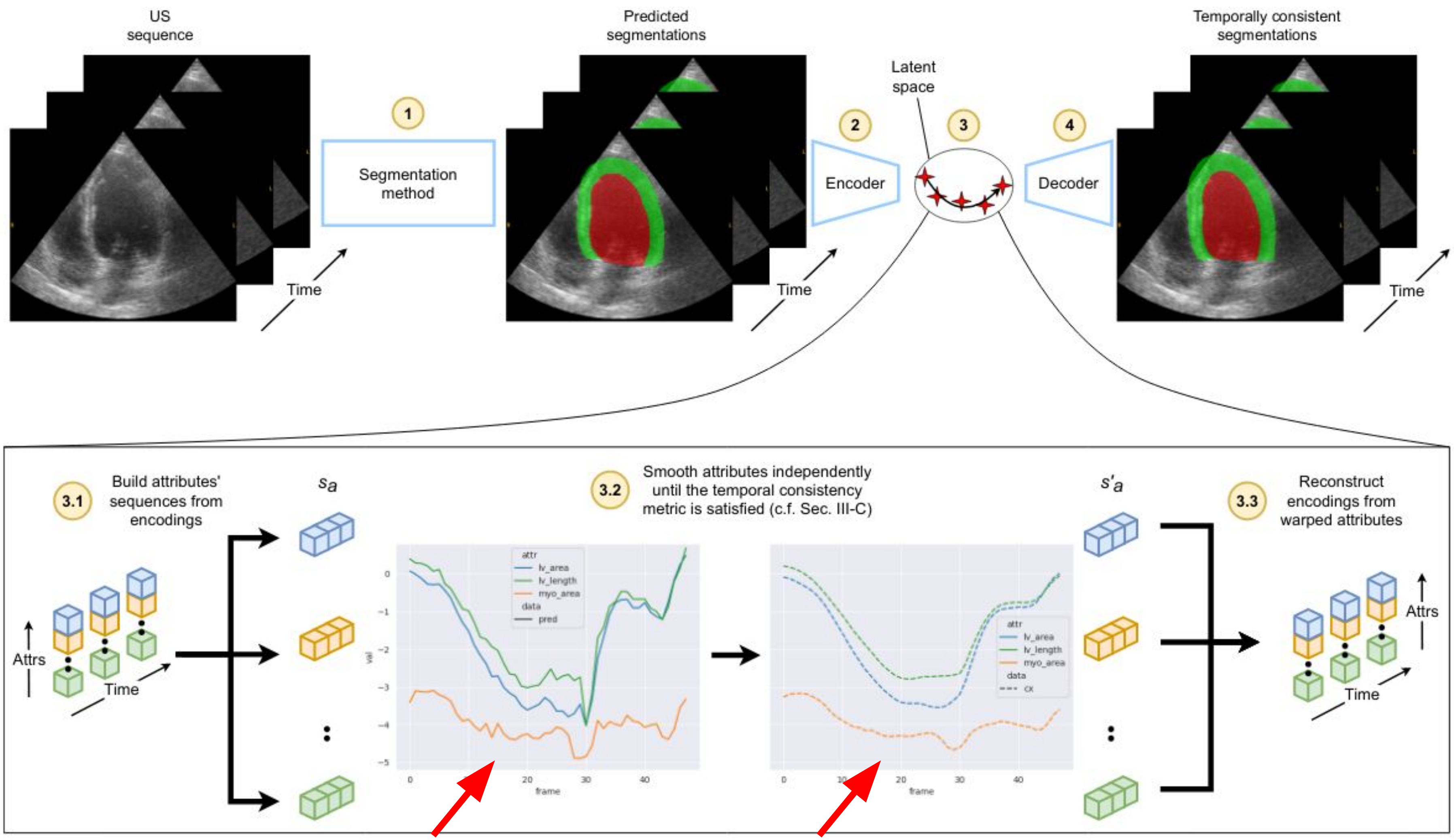
Anatomical consistency ? → Painchaud et al. *IEEE TMI* 2020



Reconstruction (intrinsic)

ex: variational autoencoders

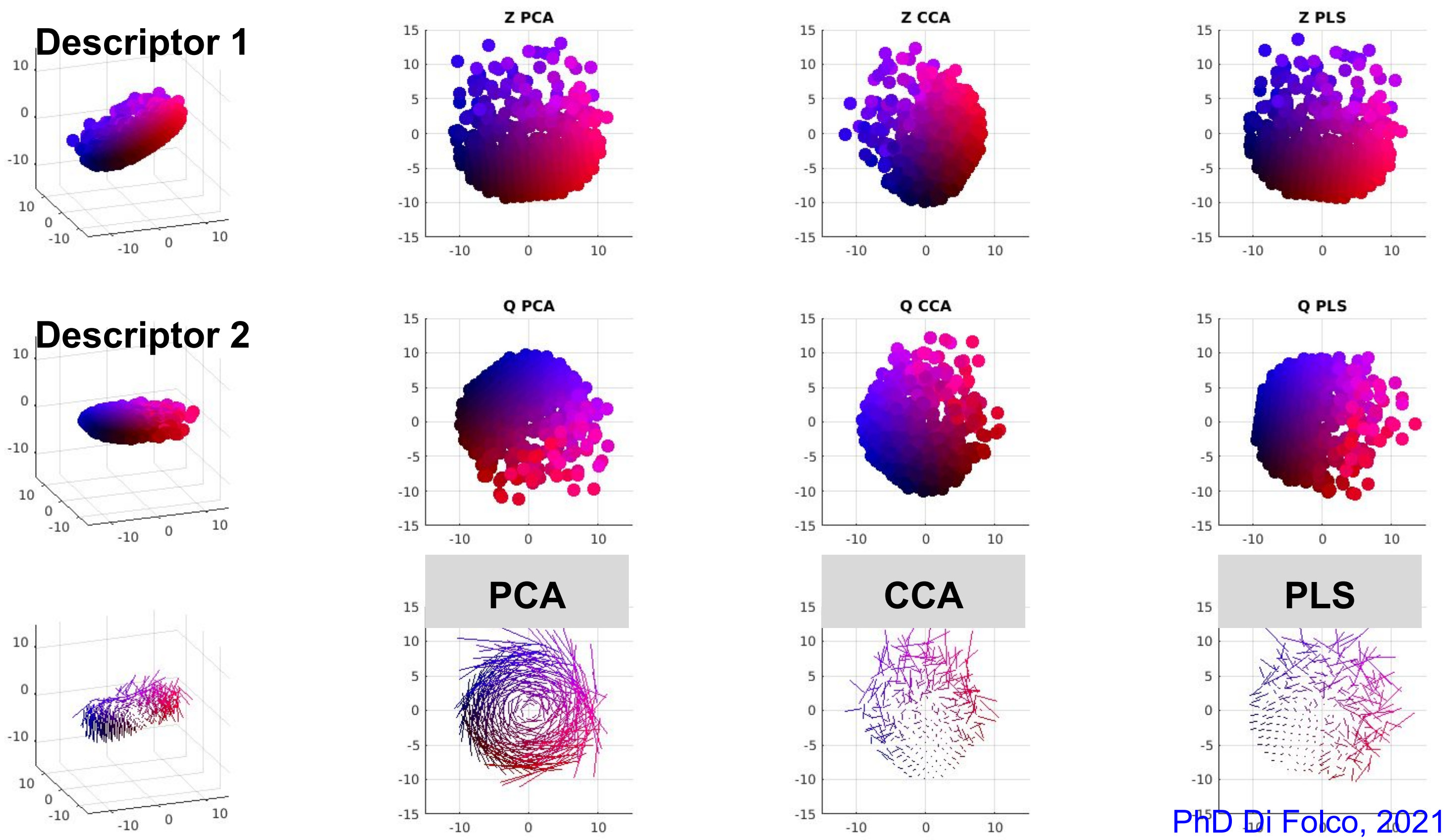
Temporal consistency ? → Painchaud et al. *arXiv* 2021



Multiple descriptors

Multiple descriptors: linear

		Nb descriptors	Maximizes...	
Principal Component Analysis	**PCA**	1	Variance	
Partial Least Squares	**PLS**	2+	Covariance	Wold et al. *Chemo* 1984
Canonical Correlation Analysis	**CCA**	2+	Correlation	Hotelling *Biometr* 1936

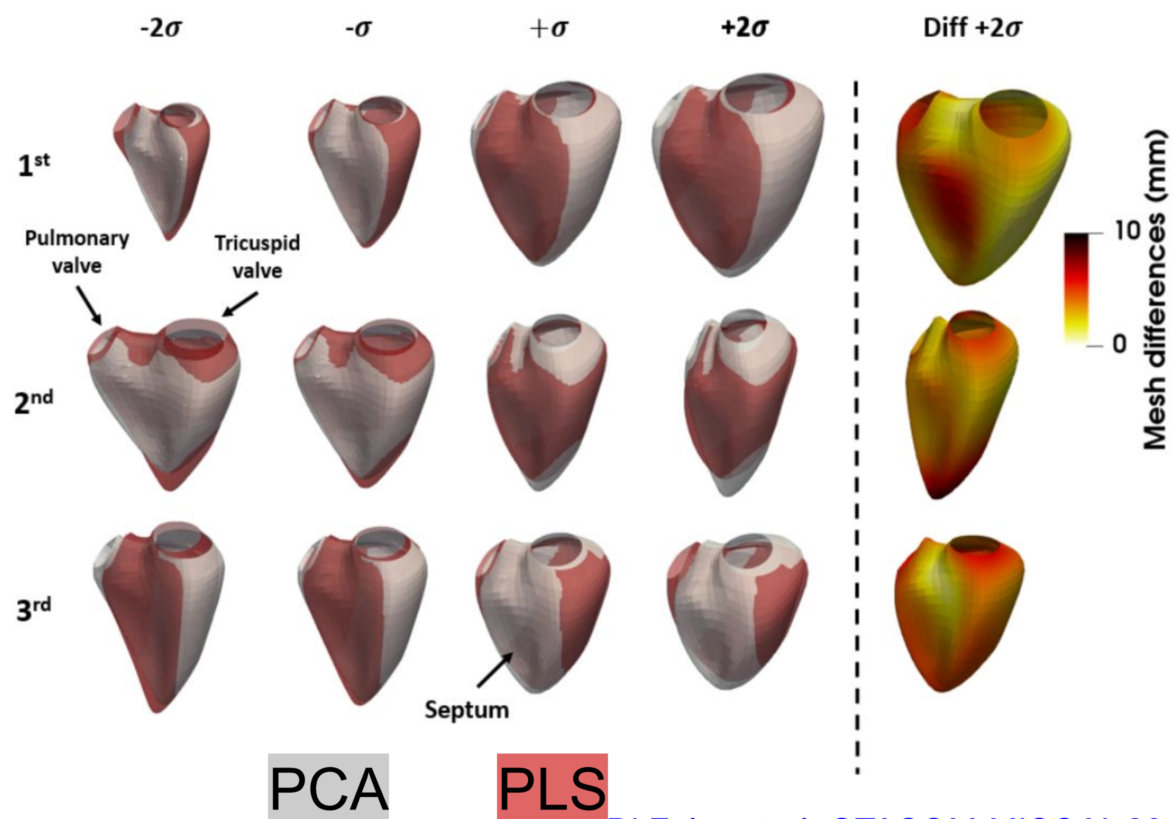


PhD Di Folco, 2021

Multiple descriptors: linear

		Nb descriptors	Maximizes...	
Principal Component Analysis	**PCA**	1	Variance	
Partial Least Squares	**PLS**	2+	Covariance	Wold et al. *Chemo* 1984
Canonical Correlation Analysis	**CCA**	2+	Correlation	Hotelling *Biometr* 1936

Strong differences in disease-relevant zones !



Di Folco et al. *STACOM-MICCAI*, 2019

Multiple descriptors: <u>non-linear</u>

- Interesting extensions of "standard" manifold learning:

  - **<u>Multiple kernel learning (MKL)</u>** Lin et al. *PAMI* 2011
    with Sergio Sanchez + Gemma Piella (UPF Barcelona)

  - **<u>Manifold alignment</u>** Ham et al. *AISTATS* 2005 + Clough et al. *PAMI* 2020
  - **<u>Multiple manifold learning (MML)</u>** Valencia et al. *CIARP* 2011 + Lee et al. *Patt Rec* 2016
    with Maxime Di Folco (CREATIS)

  - **<u>Similarity network fusion (SNF)</u>** Wang et al. *CVPR* 2012 + *Nat Meth* 2014

- Interesting extensions of variational auto-encoders:

  - **<u>Multi-channel VAE</u>** Antelmi et al. *PMLR* 2019

Multiple kernel learning (fusion)

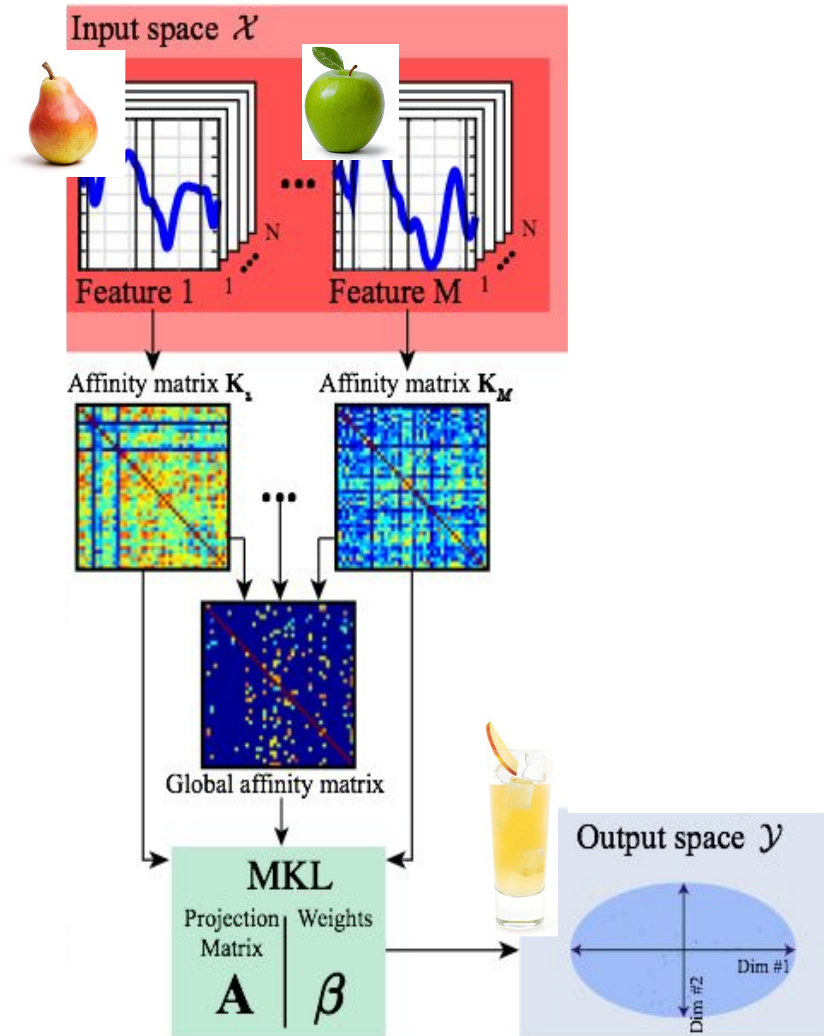
Idea = mixing kernels

Lin et al. *IEEE PAMI* 2011
Sanchez et al. *Med Image Anal* 2017

How to mix
heterogeneous data ?

Normalization
+ concatenation

Kernels



Multiple kernel learning (fusion)

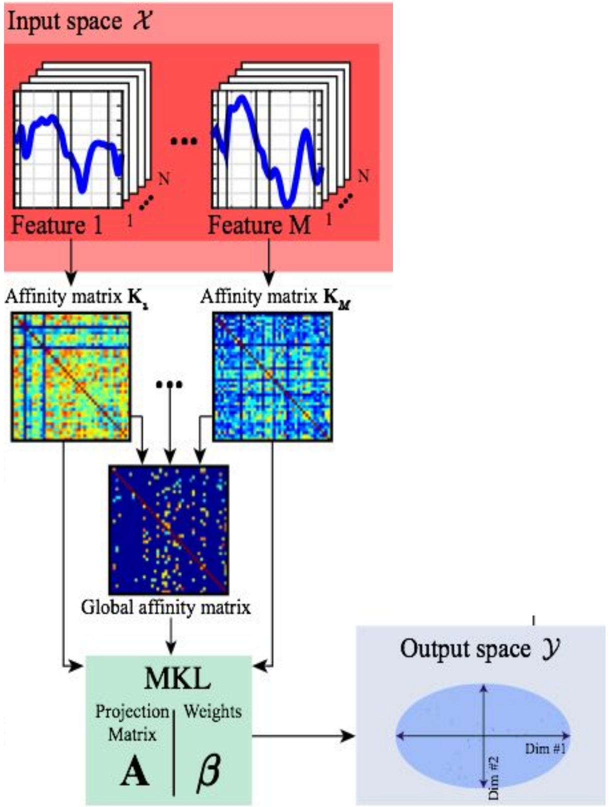
<u>Linear, 1 descriptor:</u> $\mathbf{y}_i = \mathbf{v}^t \mathbf{x}_i$

$$\hat{\mathbf{v}} = \underset{\mathbf{Y}^t\mathbf{B}\mathbf{Y}=\delta}{\operatorname{argmin}} \sum_{i,j} \|\mathbf{v}^t\mathbf{x}_i - \mathbf{v}^t\mathbf{x}_j\|^2 w_{ij} = \underset{\mathbf{Y}^t\mathbf{B}\mathbf{Y}=\delta}{\operatorname{argmin}} \; \mathbf{v}^t\mathbf{X}\mathbf{L}\mathbf{X}^t\mathbf{v}$$

<u>Non-linear, multiple descriptors:</u>

$$\underset{\mathbf{A},\beta}{\operatorname{argmin}} \sum_{i,j=1}^{N} \|\mathbf{A}^t\mathbb{K}^{(i)}\beta - \mathbf{A}^t\mathbb{K}^{(j)}\beta\|^2 w_{ij}$$

Lin et al. *IEEE PAMI* 2011
Sanchez et al. *Med Image Anal* 2017



Multiple kernel learning (fusion)

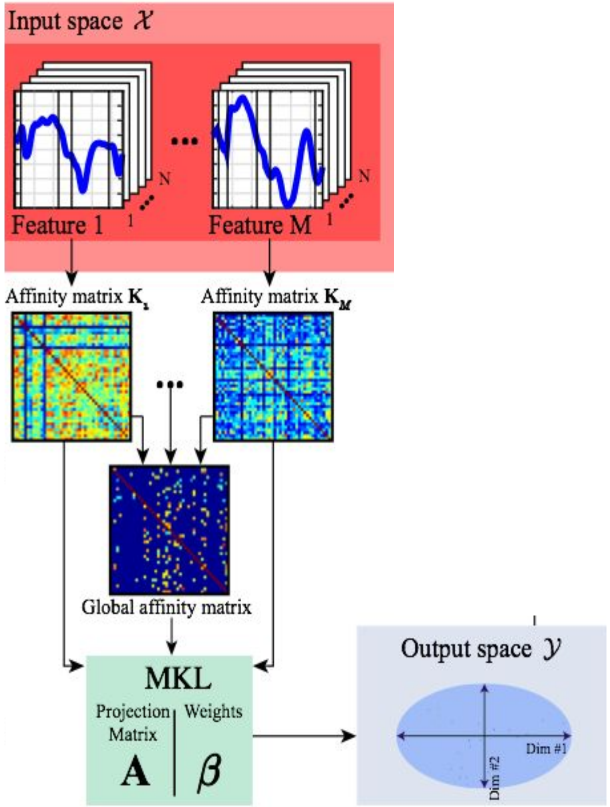
Lin et al. *IEEE PAMI* 2011
Sanchez et al. *Med Image Anal* 2017

Linear, 1 descriptor: $\mathbf{y}_i = \mathbf{v}^t\mathbf{x}_i$

$$\hat{\mathbf{v}} = \underset{\mathbf{Y}^t\mathbf{B}\mathbf{Y}=\delta}{\operatorname{argmin}} \sum_{i,j} \|\mathbf{v}^t\mathbf{x}_i - \mathbf{v}^t\mathbf{x}_j\|^2 w_{ij} = \underset{\mathbf{Y}^t\mathbf{B}\mathbf{Y}=\delta}{\operatorname{argmin}} \; \mathbf{v}^t\mathbf{X}\mathbf{L}\mathbf{X}^t\mathbf{v}$$

Non-linear, multiple descriptors:

$$\underset{\mathbf{A},\beta}{\operatorname{argmin}} \sum_{i,j=1}^{N} \|\mathbf{A}^t\mathbb{K}^{(i)}\beta - \mathbf{A}^t\mathbb{K}^{(j)}\beta\|^2 w_{ij}$$



In practice: alternate optimization

1. Fix $\mathbf{A}$ and optimize β

   = solving a trace-ratio problem

2. Fix β and optimize $\mathbf{A}$

   = not anymore a generalized eigenvalue problem due to $\beta_m \geq 0$

   Problem: non-convex, hard to solve

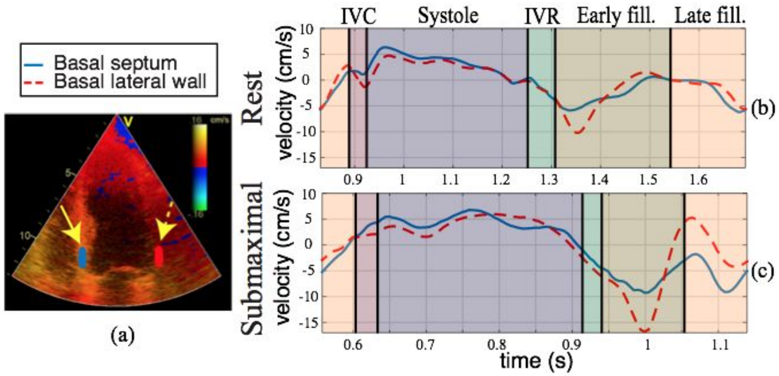
   > Can be relaxed and solved by semidefinite programming

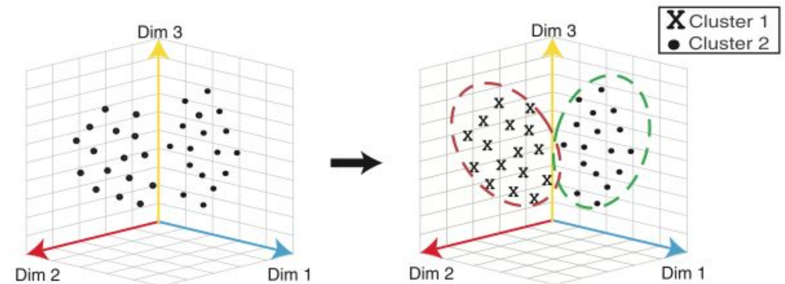
Multiple kernel learning (fusion)

CREATIS — UNIVERSITÉ DE LYON

<u>ex:</u> Heart Failure with Preserved Ejection Fraction (HFpEF)

Sanchez et al. *Med Image Anal* 2017
Sanchez et al. *Circ-Imaging* 2018



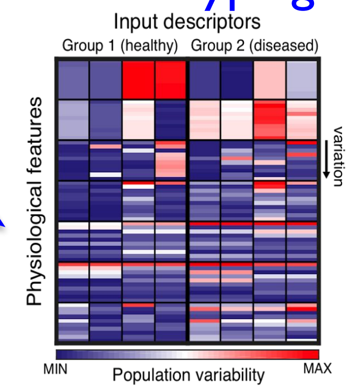


<u>Why unsupervised</u> ?

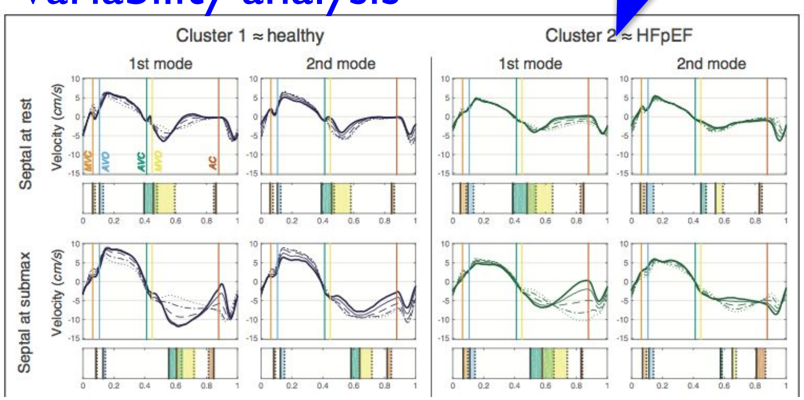
Continuum from normal to diseased

/ no trust in the [normal/HFPEF] labels

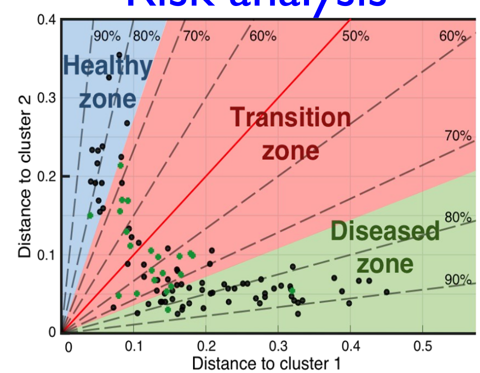
Phenotyping



Variability analysis



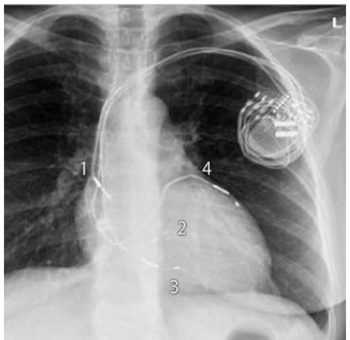
Risk analysis



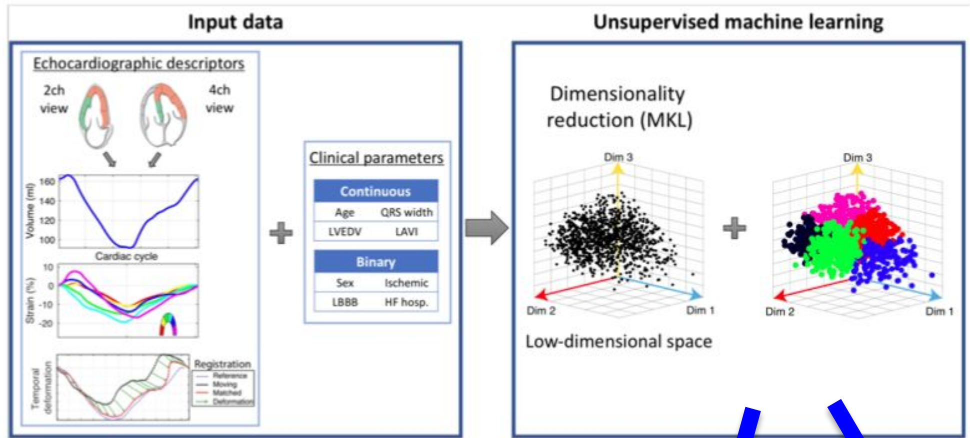
Multiple kernel learning (fusion)

ex: Cardiac Resynchronization Therapy (CRT)

Cikes et al. *Eur J Heart Fail* 2019



Scheffer & van Gelder 2012

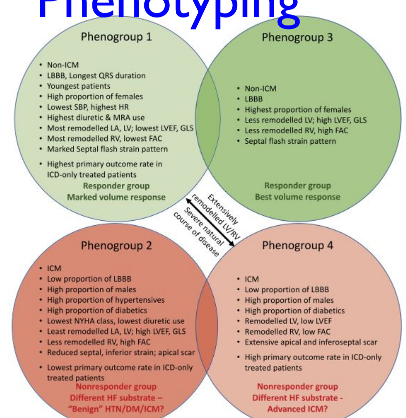


Why unsupervised ?

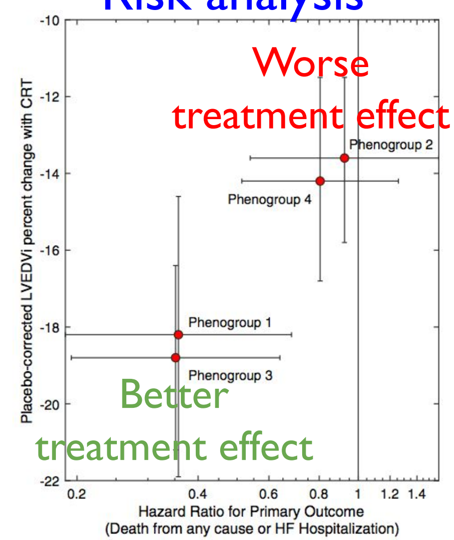
30% of non-responders with current selection

+ Limits of supervised formulation

Phenotyping

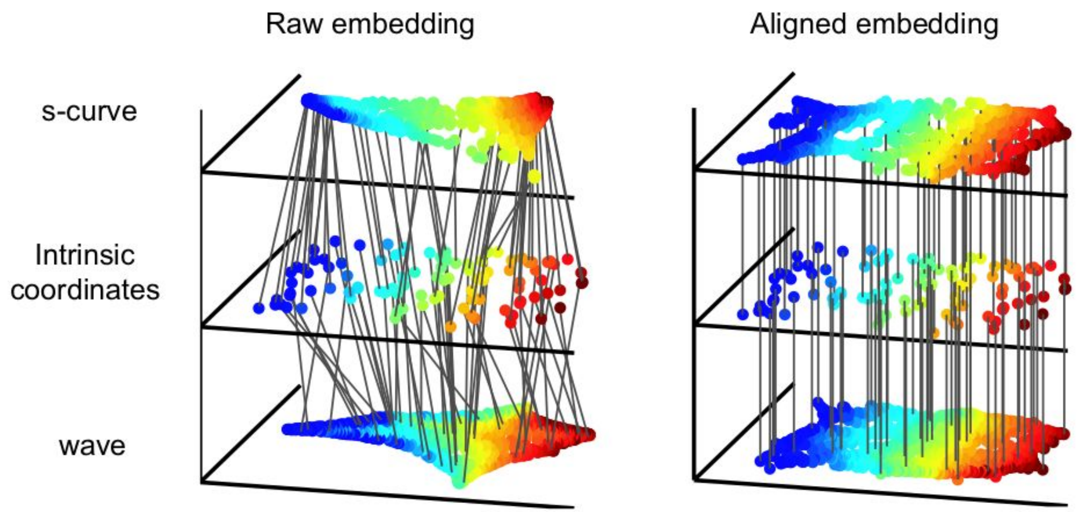


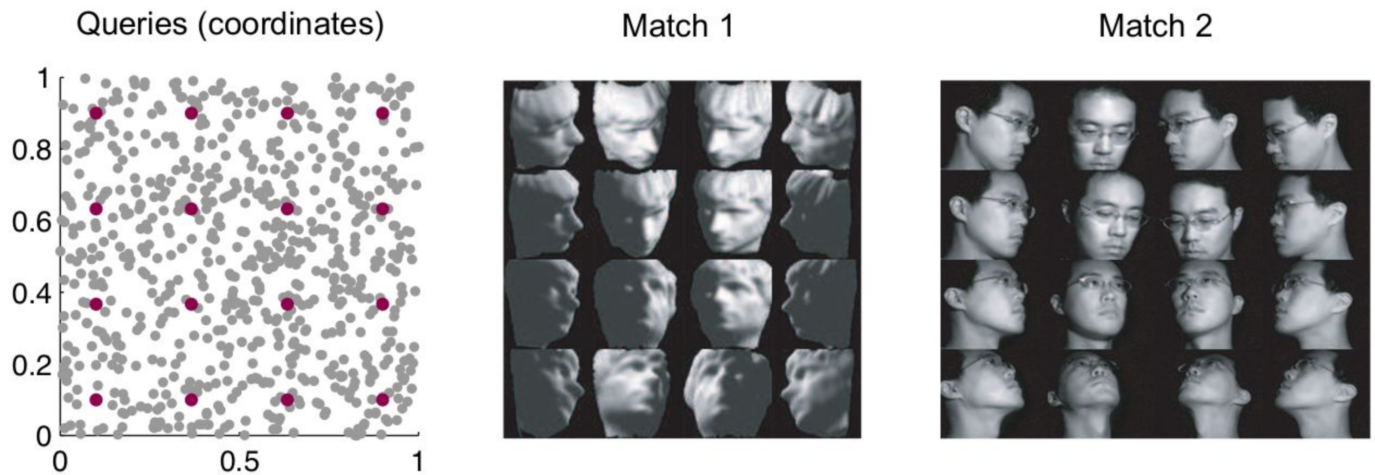
Risk analysis



Multiple manifold learning (alignment)

Ham et al. *AISTATS* 2005
Clough et al. *IEEE PAMI* 2020





Multiple manifold learning (alignment)

Ham et al. *AISTATS* 2005
Clough et al. *IEEE PAMI* 2020

$$\mathrm{tr}(\mathbf{Y}^T\mathbf{L}_A\mathbf{Y}) = \sum_{i,j=1}^{K} \|\mathbf{y}_i^{(1)} - \mathbf{y}_j^{(1)}\|^2 W_{ij}^{(1)} + \sum_{i,j=1}^{K} \|\mathbf{y}_i^{(2)} - \mathbf{y}_j^{(2)}\|^2 W_{ij}^{(2)} + \mu \sum_{i,j=1}^{K} \|\mathbf{y}_i^{(1)} - \mathbf{y}_j^{(2)}\|^2 M_{ij}$$

$$\mathbf{L}_A = \mathbf{D}_A - \mathbf{A} \quad \text{with} \quad \mathbf{A} = \left[\begin{array}{c|c} \mathbf{W}^{(1)} & \mu\mathbf{M} \\ \hline \mu\mathbf{M}^T & \mathbf{W}^{(2)} \end{array}\right]$$

$$\text{and} \quad \mathbf{D}_A = \begin{bmatrix} \ddots & & 0 \\ & \sum_{j=1}^{K} A_{ij} & \\ 0 & & \ddots \end{bmatrix}$$

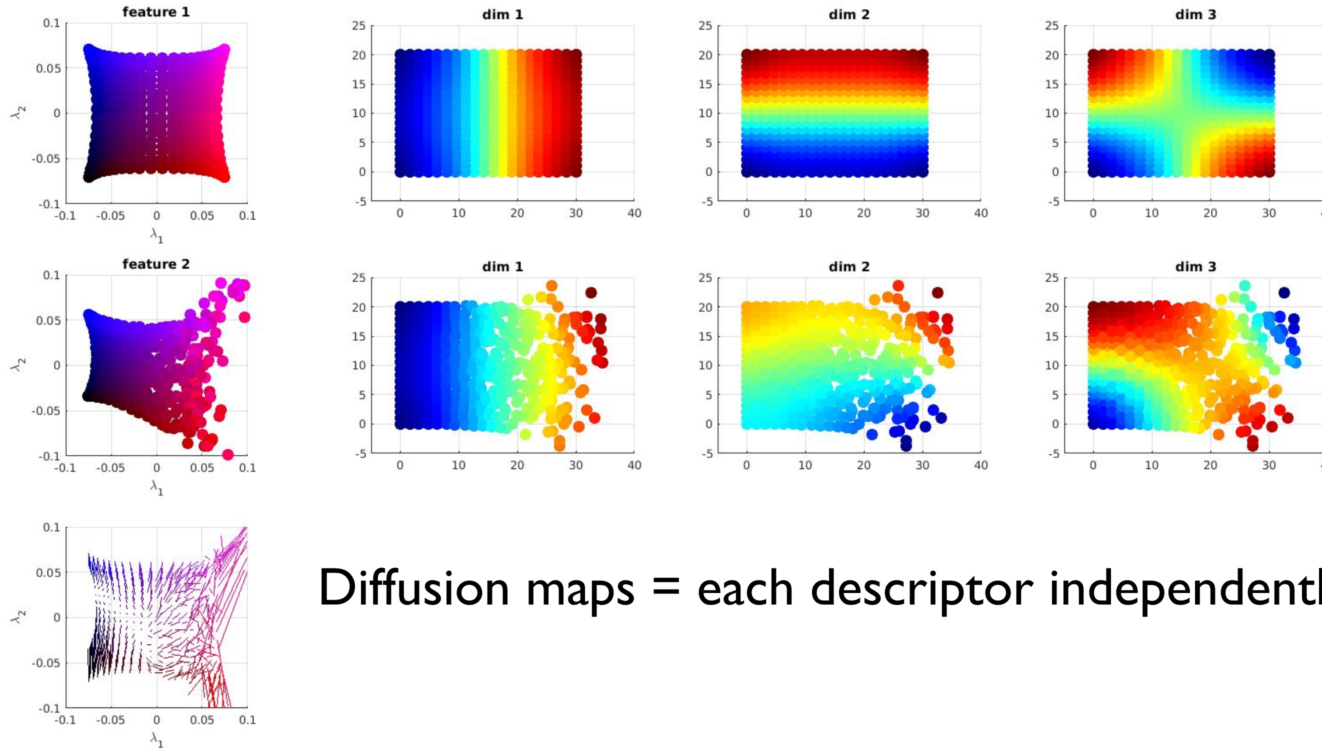
<u>In practice:</u> $\mathbf{M}$ sparse

<u>Specific case:</u> $\mathbf{M} = \mathbf{I}$
= pairwise alignment

Multiple manifold learning (alignment)

CREATIS UNIVERSITÉ DE LYON

Output (low dimensional) coords

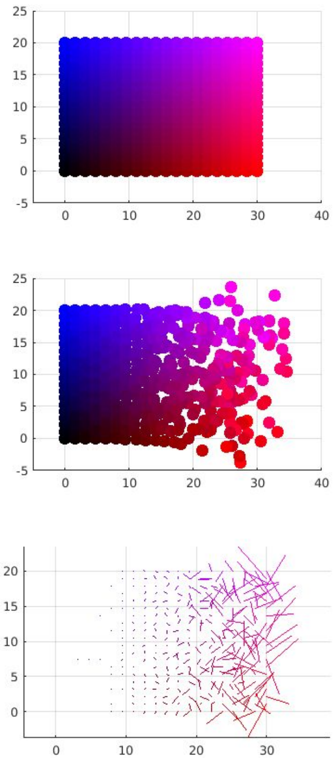


Input (high dimensional) data

Descriptor #1

Descriptor #2

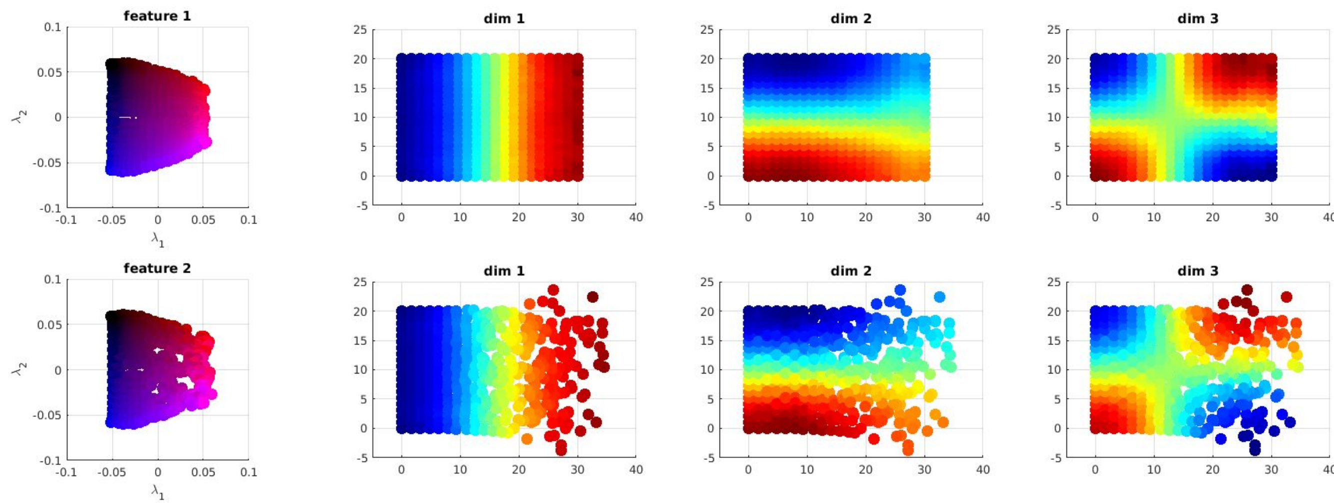
Difference



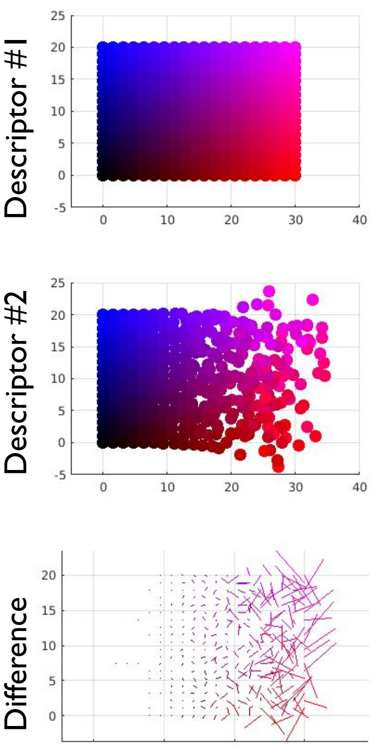
Diffusion maps = each descriptor independently

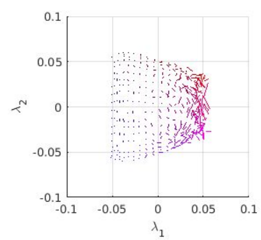
Multiple manifold learning (alignment)

Output (low dimensional) coords



Input (high dimensional) data





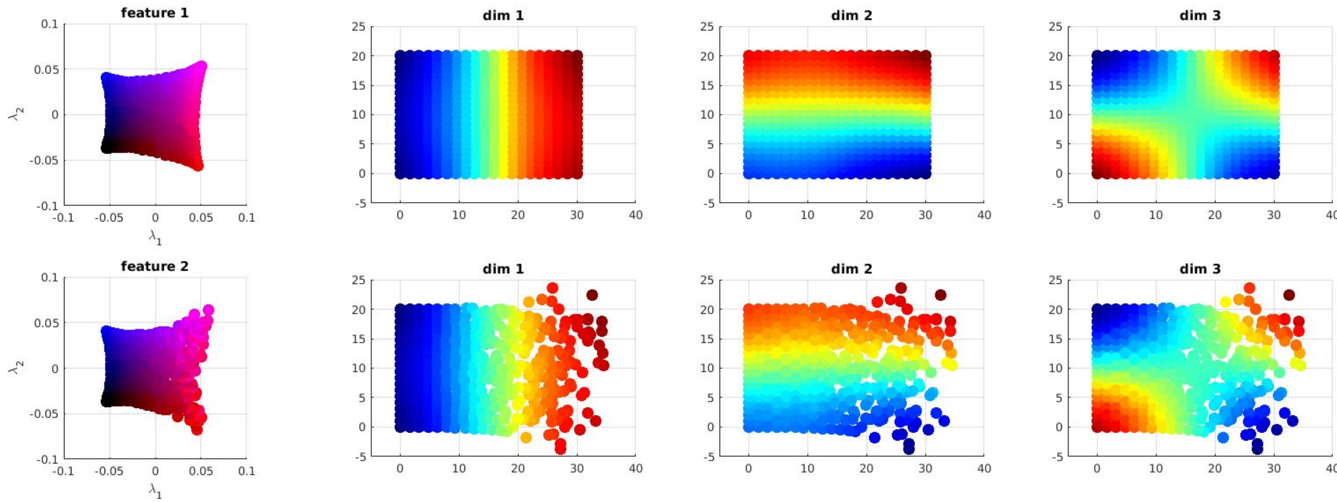
~~Diffusion maps = each descriptor independently~~

Multiple manifold learning

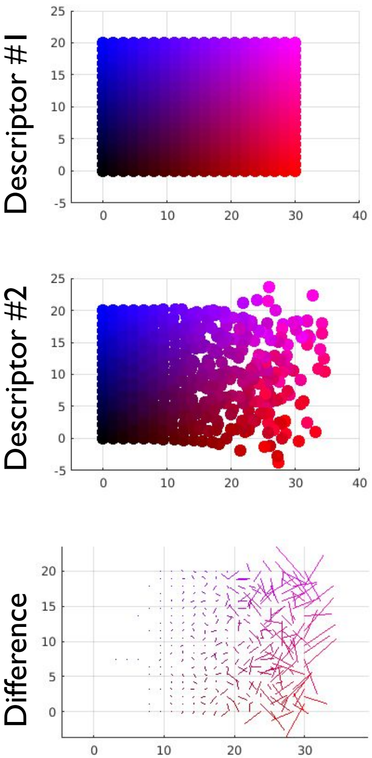
Multiple manifold learning (alignment)

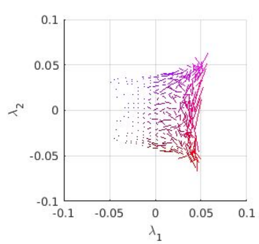
CREATIS UNIVERSITÉ DE LYON

Output (low dimensional) coords



Input (high dimensional) data





~~Diffusion maps = each descriptor independently~~

~~Multiple manifold learning~~

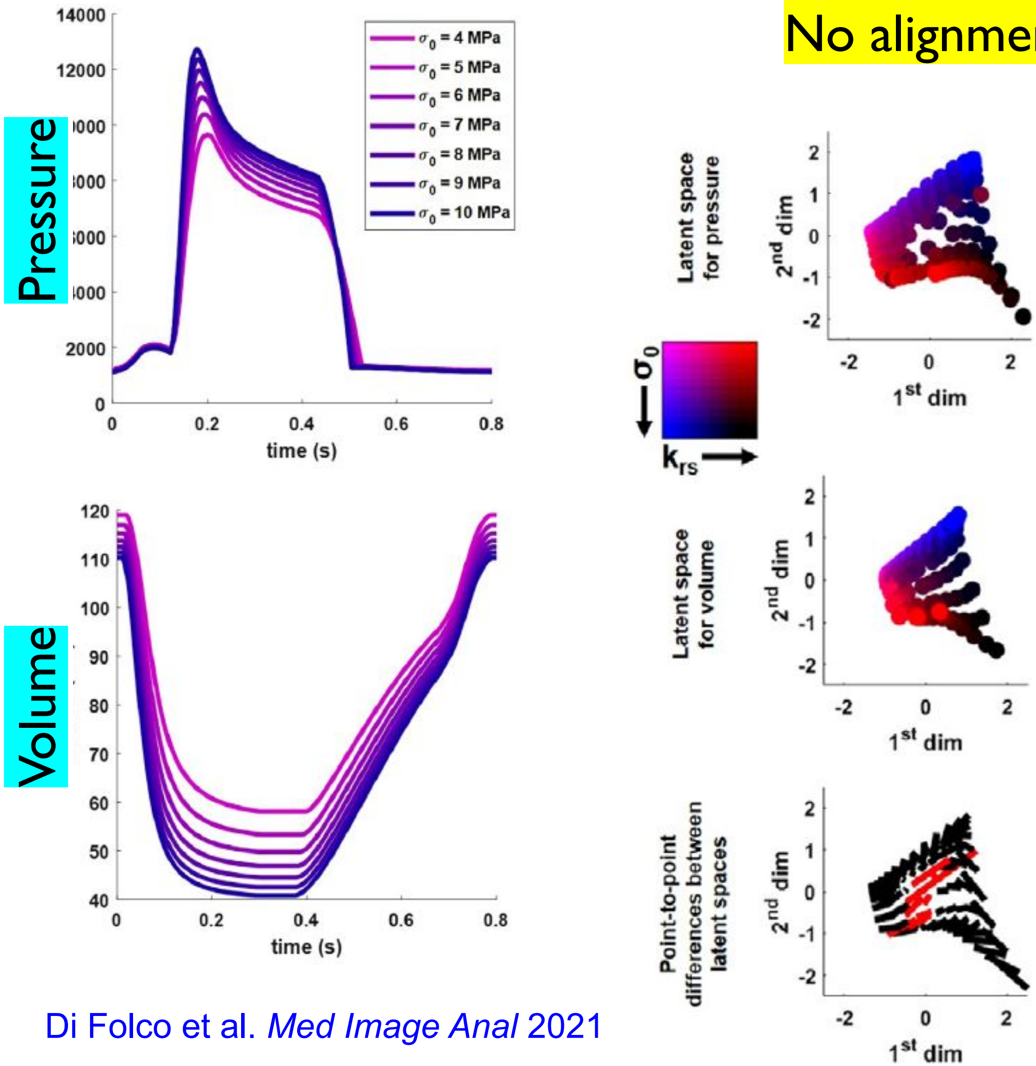
Concatenated descriptors (= ground truth)

Multiple manifold learning (alignment)

ex: synthetic data

- 0D model = pressure + volume (Mollero et al. Biomech Model Mechanobiol 2018)
- deteriorated link on some samples

No alignment

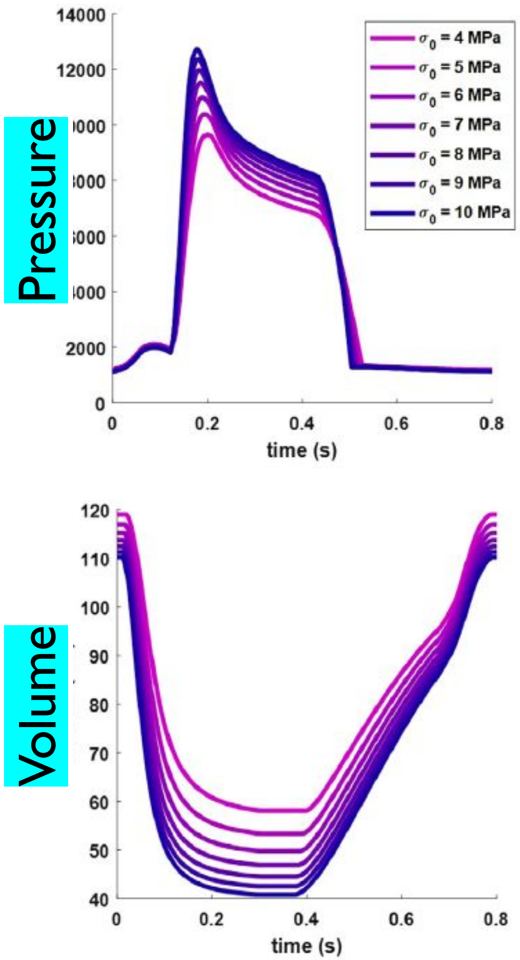


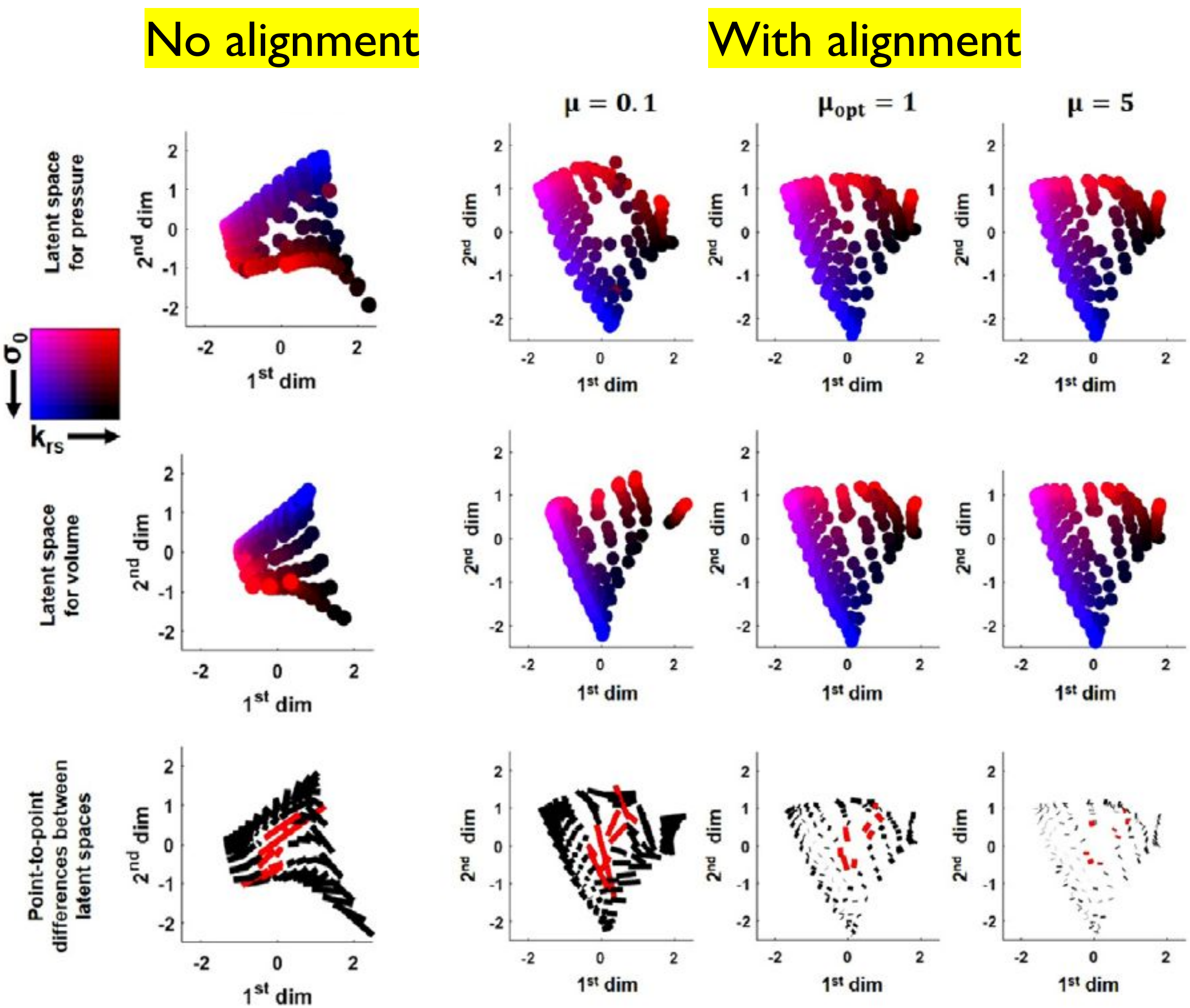
Di Folco et al. *Med Image Anal* 2021

Multiple manifold learning (alignment)

ex: synthetic data

- 0D model = pressure + volume (Mollero et al. Biomech Model Mechanobiol 2018)
- deteriorated link on some samples





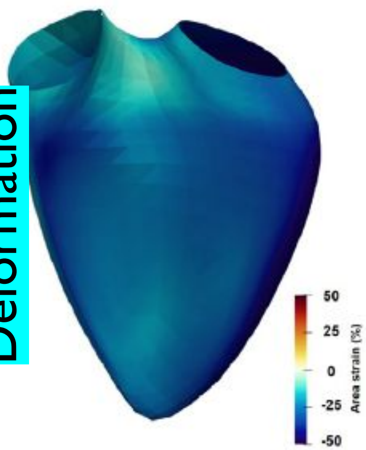
Di Folco et al. *Med Image Anal* 2021

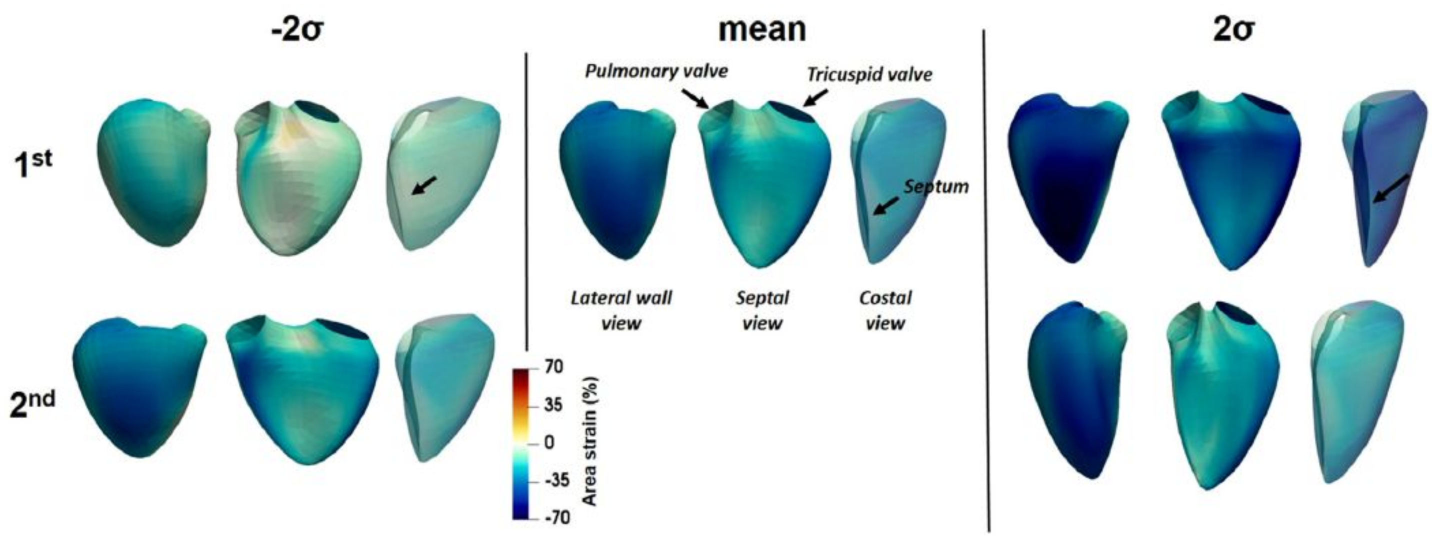
Multiple manifold learning (alignment)

ex: real data

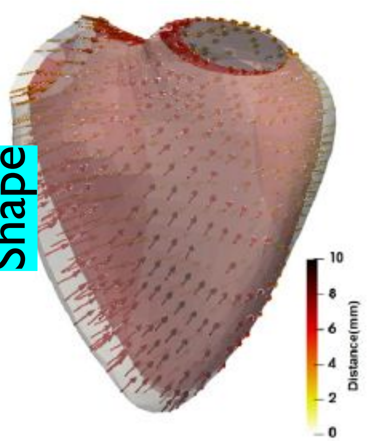
- 3D RV meshes (Moceri et al. *Eur Heart J Cardiovasc Imaging* 2018, 2020)
- Varying shape/deformation interactions depending on disease

Deformation





Shape



➔ Consistent shape/deformation patterns
➔ Fusion vs. alignment
➔ “Pairwise” vs. “relaxed” alignment

Di Folco et al. *Med Image Anal* 2021

Similarity network fusion

Idea = non-linear fusion (iterative cross-diffusion until…)

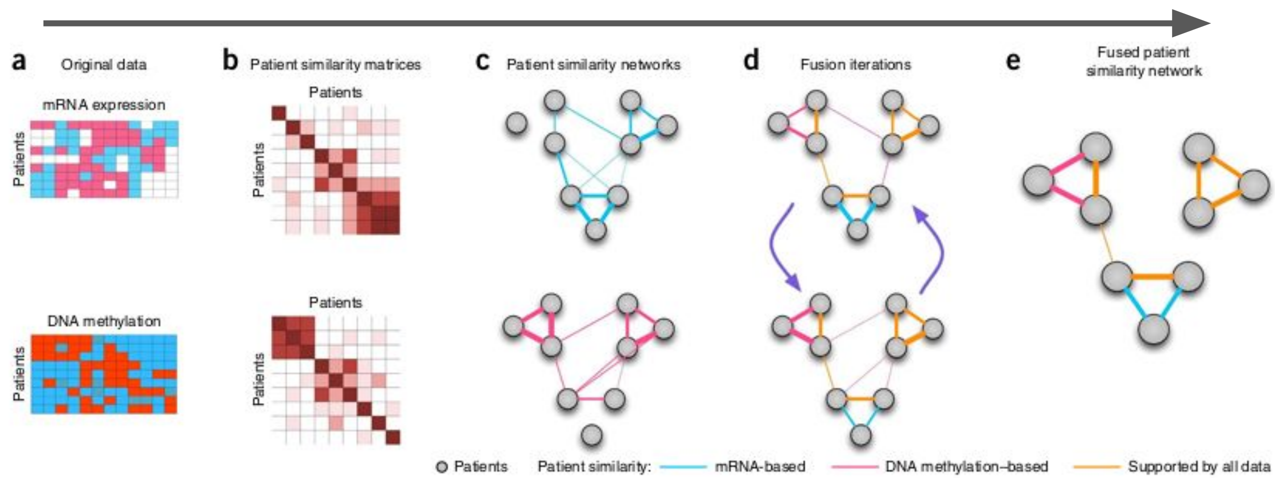
Wang et al. *CVPR* 2012
Wang et al. *Nat Meth* 2014

$$\mathbf{P}_{t+1}^{(1)} = \mathbf{S}^{(1)} \times \mathbf{P}_t^{(2)} \times (\mathbf{S}^{(1)})^T$$

$$\mathbf{P}_{t+1}^{(2)} = \mathbf{S}^{(2)} \times \mathbf{P}_t^{(1)} \times (\mathbf{S}^{(2)})^T$$

$$\mathbf{P}(i,j) = \begin{cases} \dfrac{\mathbf{W}(i,j)}{2\Sigma_{k \neq i}\mathbf{W}(i,k)}, & j \neq i \\ 1/2, & j = i \end{cases}$$

$$\mathbf{S}(i,j) = \begin{cases} \dfrac{\mathbf{W}(i,j)}{\Sigma_{k \in N_i}\mathbf{W}(i,k)}, & j \in N_i \\ 0 & \text{otherwise} \end{cases}$$



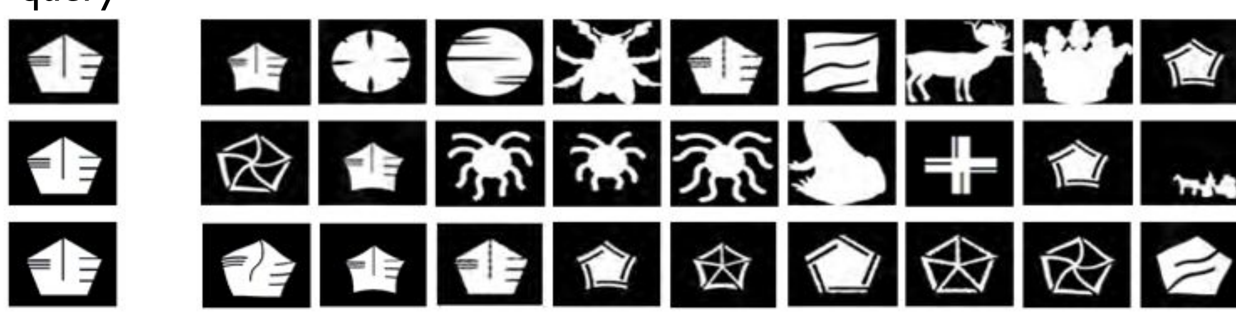
Input query

Closest images

Dataset #1

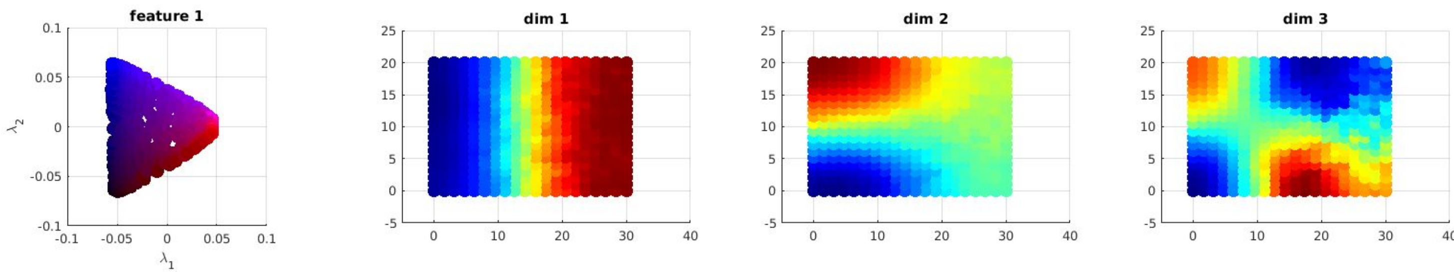
Dataset #2

Merged datasets



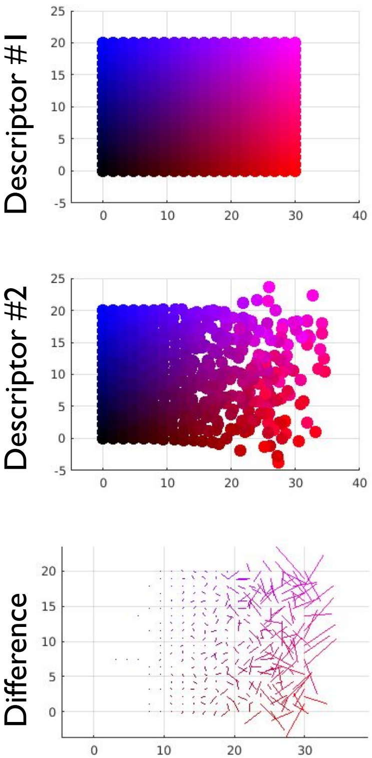
Similarity network fusion

Output (low dimensional) coords



Similarity network fusion

Input (high dimensional) data



Multi-channel VAE

$$\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_C\}$$

Multiple channels

Latent space:

$$\mathbf{z} \sim p(\mathbf{z}),$$

$$\mathbf{x}_c \sim p(\mathbf{x}_c | \mathbf{z}, \boldsymbol{\theta}_c),$$

Antelmi et al. *PMLR* 2019

Fit to distribution

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) = \mathbb{E}_c \left[L_c - \mathcal{D}_{KL}\left(q(\mathbf{z}|\mathbf{x}_c, \boldsymbol{\phi}_c) \,||\, p(\mathbf{z}) \right) \right]$$

Reconstruction error

$$\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c, \boldsymbol{\phi}_c)} \sum_{i=1}^{C} \ln p(\mathbf{x}_i | \mathbf{z}, \boldsymbol{\theta}_i)$$

Multi-channel VAE

Latent space:

$$\mathbf{z} \sim p(\mathbf{z}),$$

$$\mathbf{x}_c \sim p(\mathbf{x}_c|\mathbf{z}, \boldsymbol{\theta}_c),$$

$$\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_C\}$$

Multiple channels

Antelmi et al. *PMLR* 2019

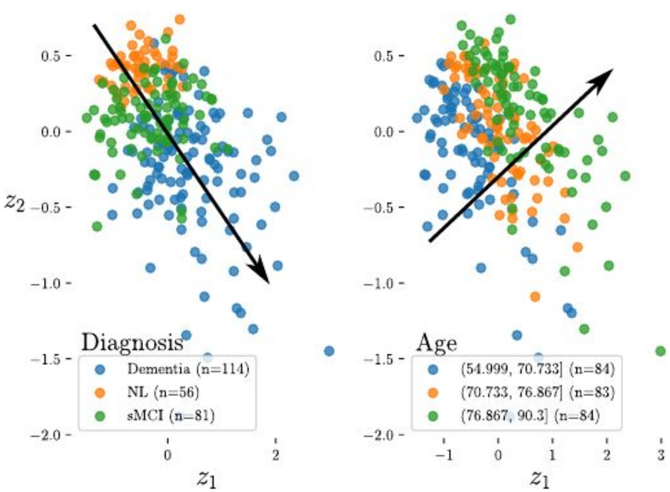
Fit to distribution

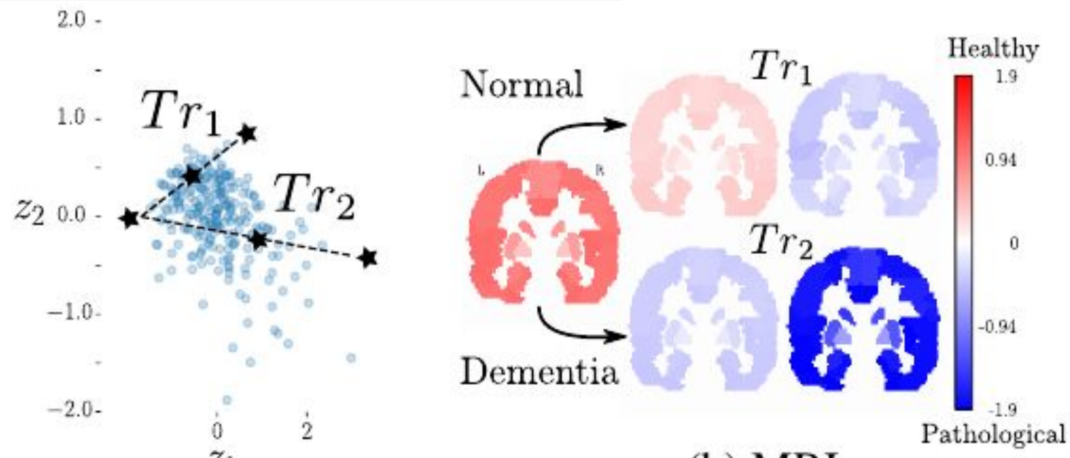
$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) = \mathbb{E}_c\left[L_c - \mathcal{D}_{KL}\left(q(\mathbf{z}|\mathbf{x}_c, \boldsymbol{\phi}_c)\,||\,p(\mathbf{z})\right)\right]$$

Reconstruction error

$$\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c, \boldsymbol{\phi}_c)} \sum_{i=1}^{C} \ln p(\mathbf{x}_i|\mathbf{z}, \boldsymbol{\theta}_i)$$

1 clinical channel (6 scalar variables)
+ 3 imaging channels (structural = MRI, functional = FDG-PET + Amyloid-PET)





(a) Latent sub-space

(b) MRI

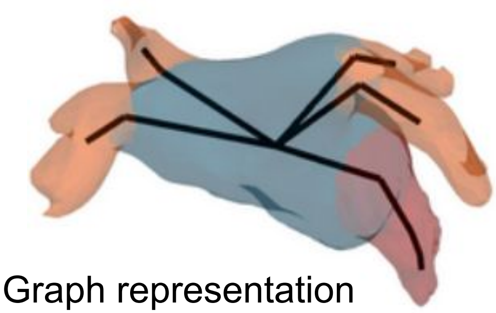
Multi-channel VAE

Interesting extension: meta-learning for heterogeneous datasets

Harrison et al. *MICCAI* 2021

Reconstruction error = on labels !

$$\mathcal{L}(x,y,\theta,\phi) = \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}_{z\sim q_{\phi_k}(z|x_k,y_k)}\left[\sum_{k=1}^{N} \ln p_\theta(y_k|z,x_k)\right] - \mathcal{D}_{KL}[q_{\phi_k}(z|x_k,y_k)||p_\theta(z|x))]$$



Graph representation of the left atrium

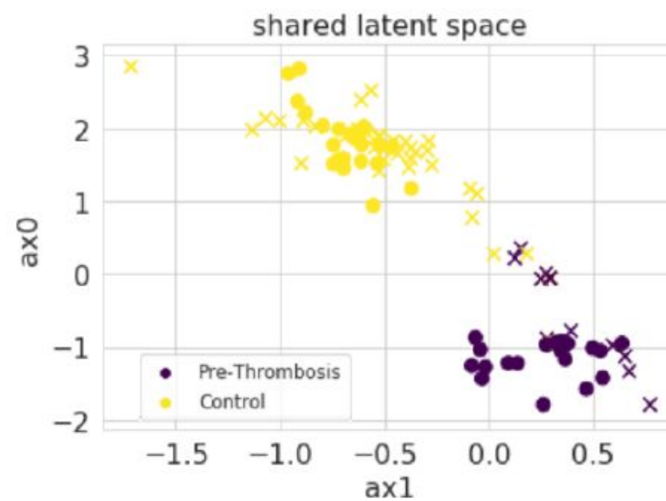
Datasets:

$$D = \{D_k\}_{k=1}^{N}$$

Single latent space:

$$p(z|x_k, y_k)$$

observation label



x = systole

● = diastole

Multiple descriptors: non-linear

- Interesting extensions of "standard" manifold learning:

- **Multiple kernel learning (MKL)** — Lin et al. *PAMI* 2011
  with Sergio Sanchez + Gemma Piella (UPF Barcelona)

- **Manifold alignment** — Ham et al. *AISTATS* 2005 + Clough et al. *PAMI* 2020
- **Multiple manifold learning (MML)** — Valencia et al. *CIARP* 2011 + Lee et al. *Patt Rec* 2016
  with Maxime Di Folco (CREATIS)

- **Similarity network fusion (SNF)** — Wang et al. *CVPR* 2012 + *Nat Meth* 2014

- Interesting extensions of variational auto-encoders:

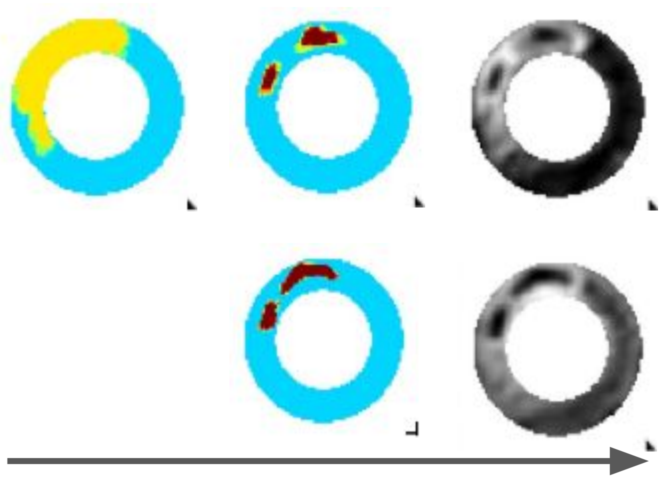
- **Multi-channel VAE** — Antelmi et al. *PMLR* 2019

➔ **What about not merging all data at once ?**

Imposed data hierarchy

LATE Gadolinium

EARLY Gadolinium



Level 1 infarct

Level 2 Microvascular obstruction

Level 3 Images

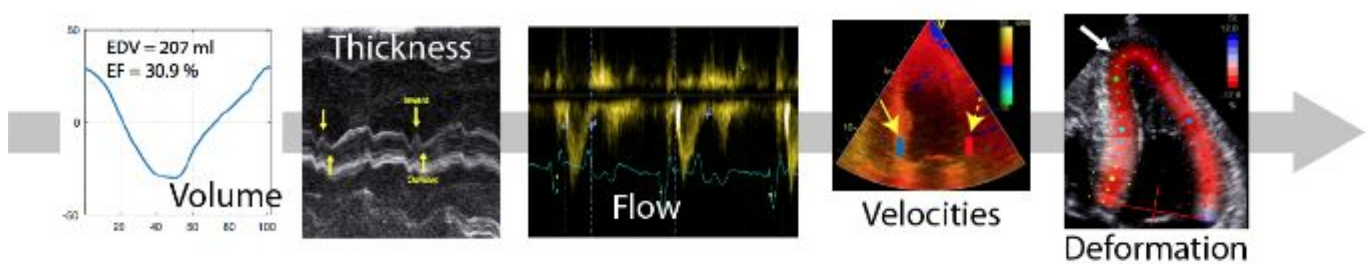
Ischemia-reperfusion with MRI:

➔ Value of non-truncated images ?

PhD of B. Freiche (CREATIS)

Imaging protocol:

➔ Which acquisition is recommended next ?



Postdoc of G. Bernardino (CREATIS)

Incremental integration

Zhou et al. *Knowl Based Syst* 2019

$$\min_{\mathbf{K}_b^{(1)},\ldots,\mathbf{K}_b^{(m)},\mathbf{K}^{(t)}} \sum_{i=1}^{m} \|\mathbf{K}^{(t)} - \mathbf{K}_b^{(i)}\|_F^2 + \|\mathbf{K}^{(t)} - \mathbf{K}_c^{(t)}\|_F^2$$

$$+ \lambda_1 \sum_{i=1}^{m} \|\mathbf{K}_b^{(i)} - \hat{\mathbf{K}}_b^{(i)}\|_F^2$$

$$+ \lambda_2 \left(rank(\mathbf{K}^{(t)}) + \sum_{i=1}^{m} rank(\mathbf{K}_b^{(i)}) \right),$$

Evolution of consensus kernel

Evolution of base kernels

Regularization

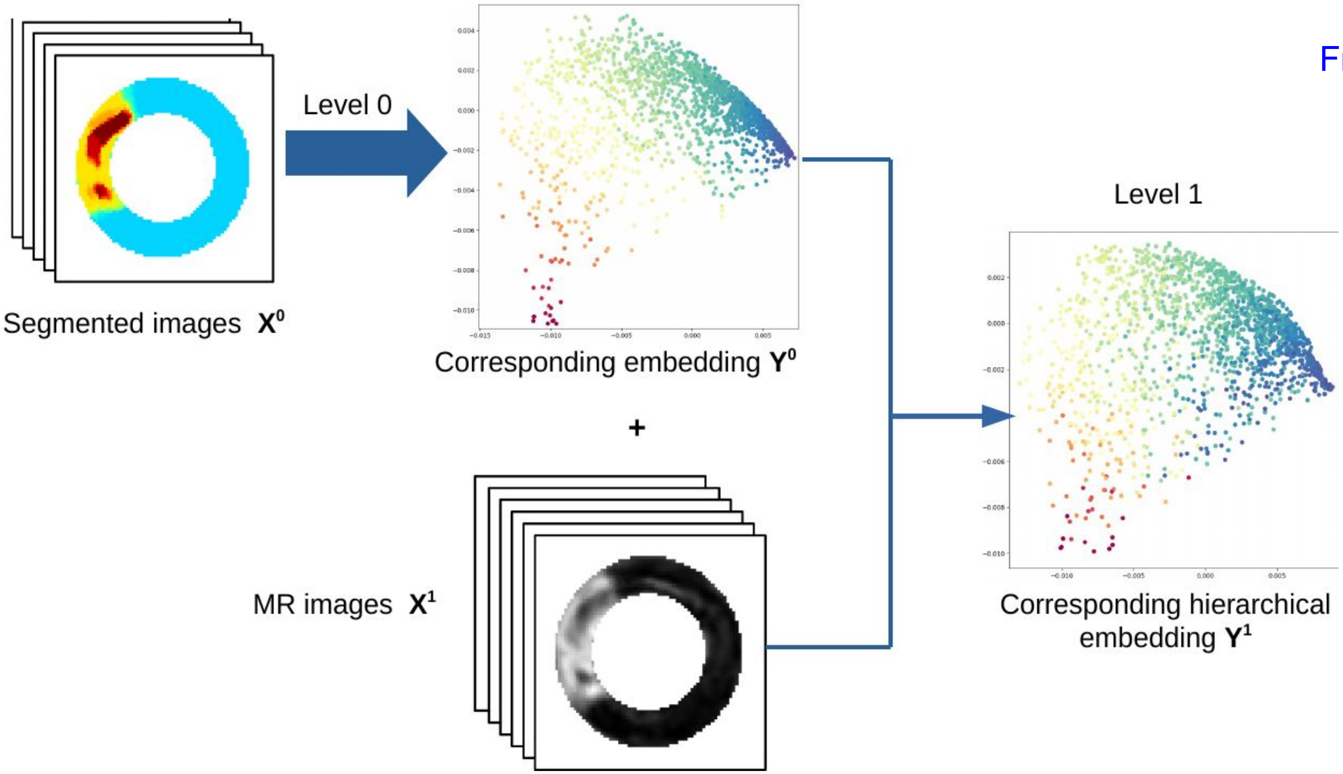
Yin et al. *Neural Networks* 2021

$$\min_{\widetilde{\boldsymbol{W}}_{t+1}} \underbrace{\|\widetilde{\boldsymbol{W}}_{t+1} - \widetilde{\boldsymbol{W}}_t\|_F^2 + \|\widetilde{\boldsymbol{W}}_{t+1} - \boldsymbol{W}^{(t+1)}\|_F^2}_{\text{Evolution of consensus kernel}} + \underbrace{\lambda \|\widetilde{\boldsymbol{W}}_{t+1}\|_1}_{\text{Regularization}}$$

Imposed data hierarchy

CREATIS UNIVERSITÉ DE LYON

Bhatia et al. *IEEE TMI* 2014
Freiche et al. *STACOM-MICCAI* 2021



Hierarchical cost function :

$$\arg\min_{\mathbf{Y}^1} \; (1-\mu)\sum_i \sum_j \|\mathbf{y}_i^1 - \mathbf{y}_j^1\|^2 W_{ij}^1 + \mu \sum_i \|\mathbf{y}_i^1 - \mathbf{y}_i^0\|^2$$

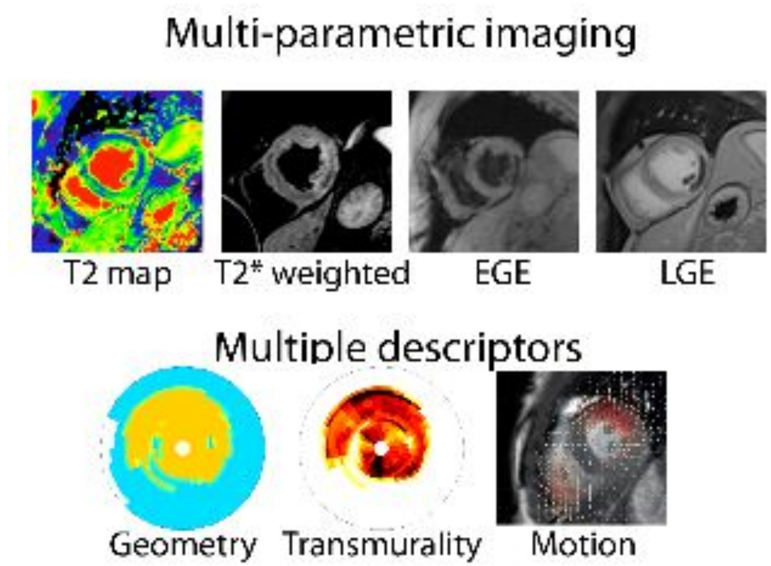
Analytic solution :

$$\mathbf{Y}^1 = (\mu\mathbf{I} + 2(1-\mu)\mathbf{L}^1)^{-1}\mu\mathbf{Y}^{\mathbf{0}}$$

Imposed data hierarchy

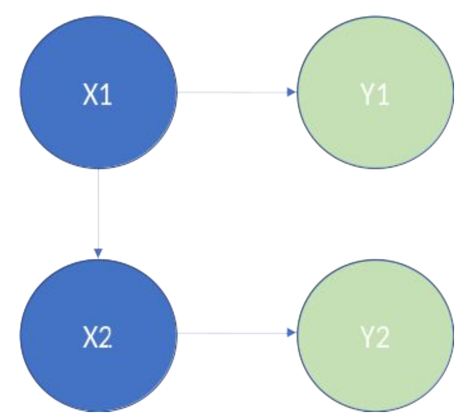
Going further:

- <u>More descriptors ?</u>
  Postdoc of F. Zheng



- <u>More generic / probabilistic formulation ?</u>
  Bernardino & Freiche *Work in progress*

$$\sum_{i} d(\mathbf{y}_i^1, f(\mathbf{y}_j^2))$$



Learning data hierarchy

CREATIS UNIVERSITÉ DE LYON

Bernardino et al. *STACOM-MICCAI* 2021

<u>Estimating the need of acquiring new data:</u>

$$P(Y|\mathbf{X_0},\mathbf{X_1}) = (\prod_i (w^i P(Y^i|X_0^i,\theta_0) + (1-w^i)P(Y^i|X_1^i,\theta_1)))P(\mathbf{w}|\mathbf{X_0})$$

weights

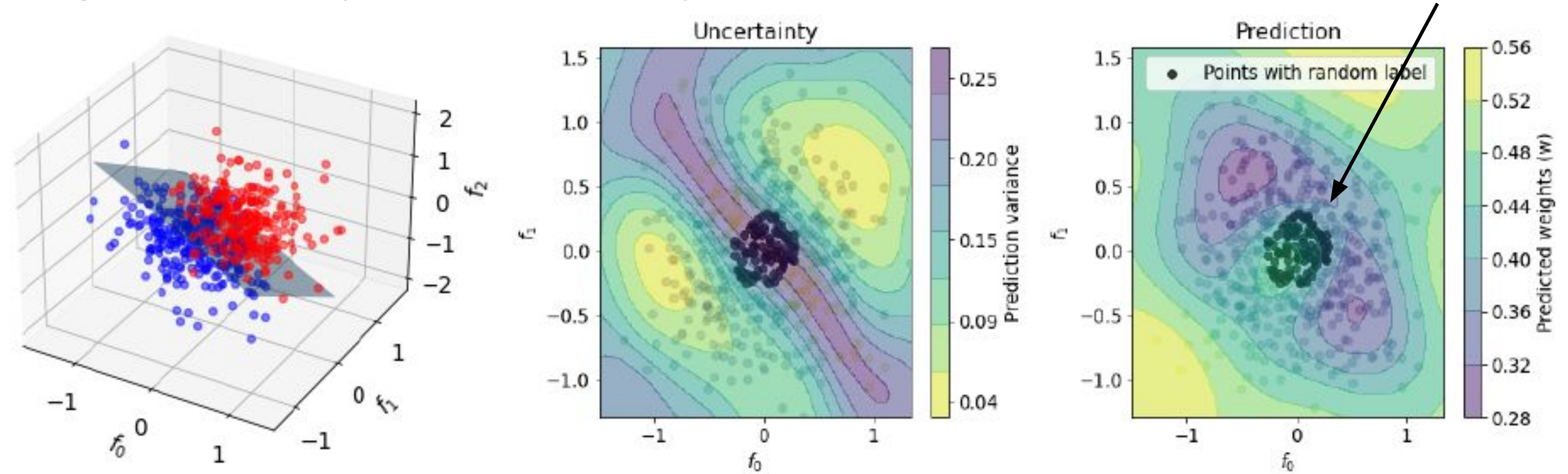
single view classification

<u>Prediction:</u>

$$P(Y^i|X_0^i,X_1^i) = \begin{cases} P(Y^i|X_0^i) & w^i \geq k, \\ w^i P(Y^i|X_0^i) + (1-w^i)P(Y^i|X_1^i) & w^i < k. \end{cases}$$

<u>ex:</u> synthetic data (label = X1 + noise)

No further acquisition required

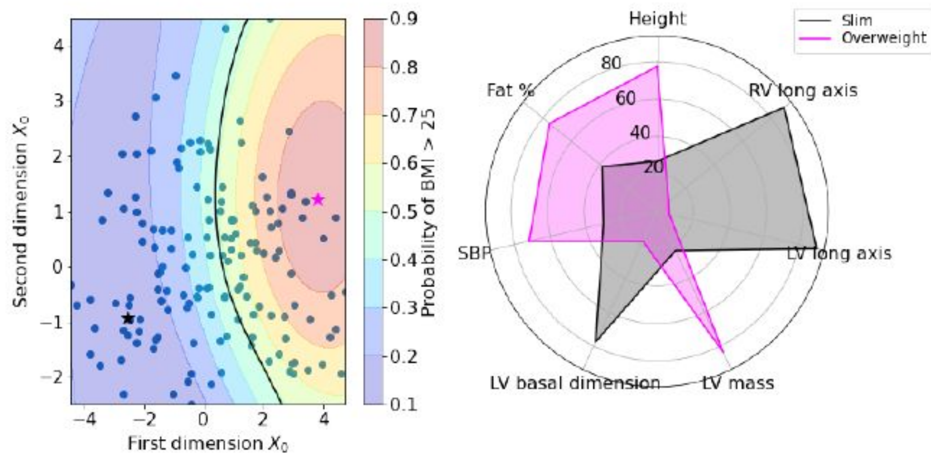


Learning data hierarchy

CREATIS UNIVERSITÉ DE LYON

Bernardino et al. *STACOM-MICCAI* 2021

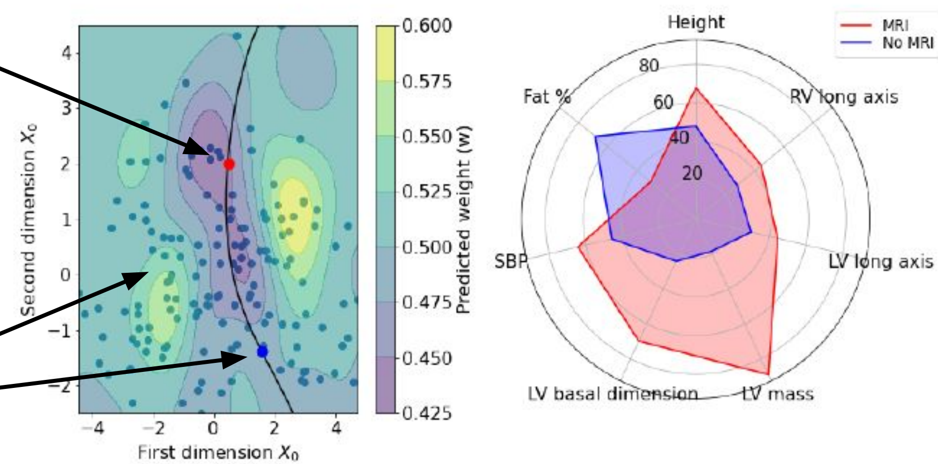
ex: real data = ECHO measurements (X_0), N=480
MRI cardiac shape (X_1), N=159



(a) Overweight

MRI more valuable

MRI less valuable



(b) Predicted weights

Learning data hierarchy

Reinforcement learning:

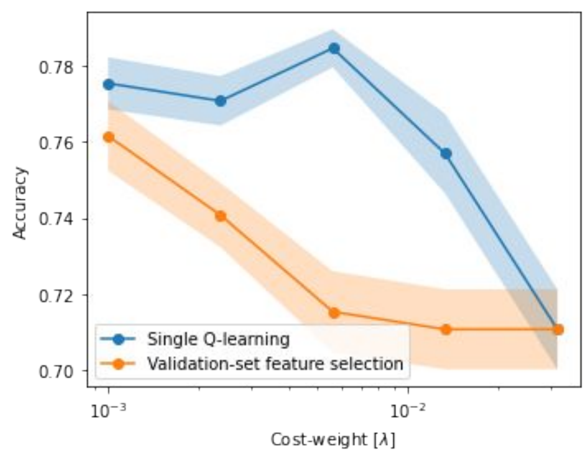
Bernardino et al. *Work in progress*
Wang et al. *NIPS* 2015

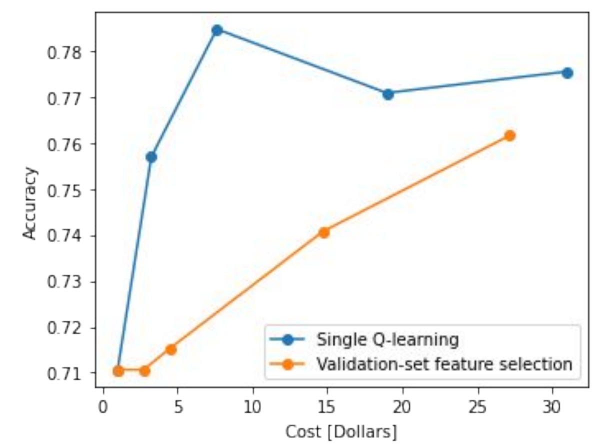
Learning a policy, an optimal sequences of actions done by an actor to maximise a reward.
➔ RL estimates the value of taking actions (long term instantaneous reward).

Actor: clinician
Action: acquiring new modality / finish and do diagnosis
Reward: accuracy + cost
States: Which modalities have been acquired + its measurements.

ex: Cleveland ICU dataset

- Predict CAD in patients admitted to ICU. Prevalence ~60%
- 11 variables, in 4 groups (demographic, admission data, exercise test, laboratory). Each has a different cost.



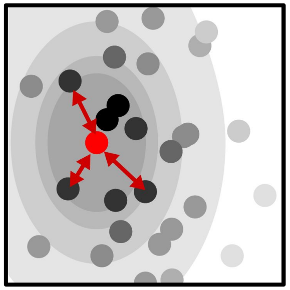


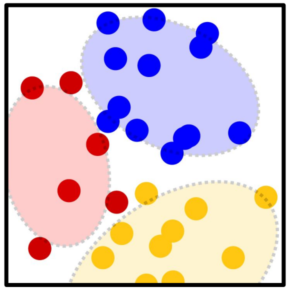
Summary

	Embedding space	Reconstruction
Manifold learning	**Meaningful** distances between *input* samples ~ Euclidean distance between *output* coordinates	Interpolation ?
Auto-encoders	Limited statistical meaning Limited for “non convolutionable” data ?	Optimized encoding/decoding

Nice statistical framework for analysis + extensions...

- Exploitable latent space
- Heterogeneous types of data
- Better metrics





… keeping in mind the clinical question

Thanks… questions?

nicolas.duchateau@creatis.insa-lyon.fr