# DispVoxNets: Non-Rigid Point Set Alignment with Supervised Learning Proxies


**Soshi Shimada**[1,2] **Vladislav Golyanik**[3] **Edgar Tretschk**[3]

**Didier Stricker**[1,2] **Christian Theobalt**[3]

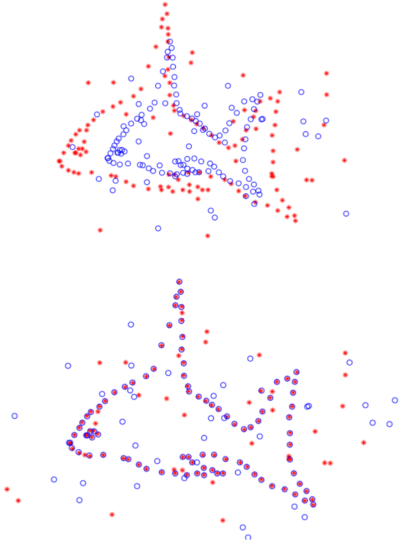
[1]University of Kaiserslautern [2]DFKI [3]MPI for Informatics, SIC



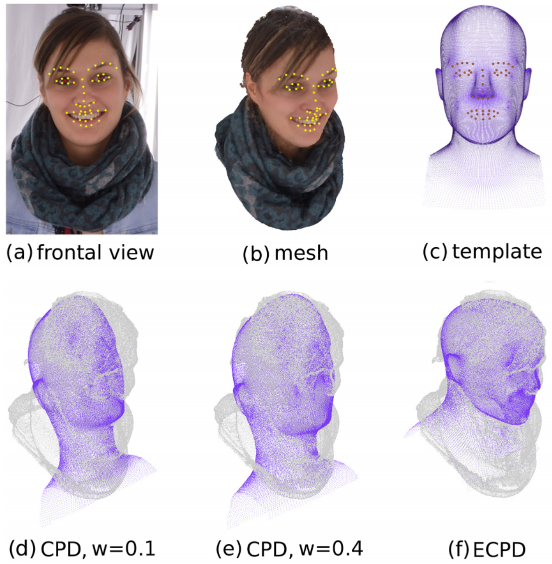



# Non-rigid Point Set Registration (NRPSR)

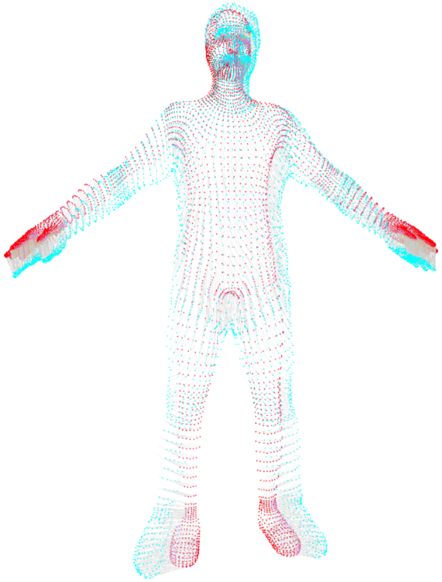
Objective: given two point sets, find displacements (or correspondences) between the point sets.



*2D point set registration*
[Myronenko and Song 2010]



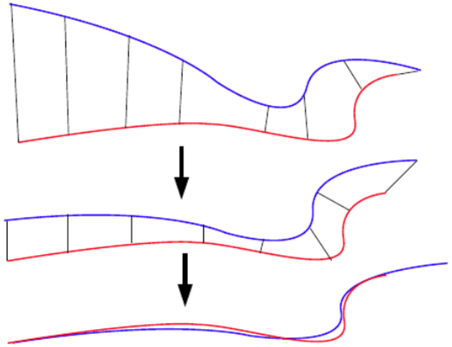

*3D face registration*
[Taetz *et al.* 2016]



*3D pose registration*
[Golyanik *et al.* 2017]

# Related Works,  General-Purpose Methods

# Related Works, General-Purpose Methods

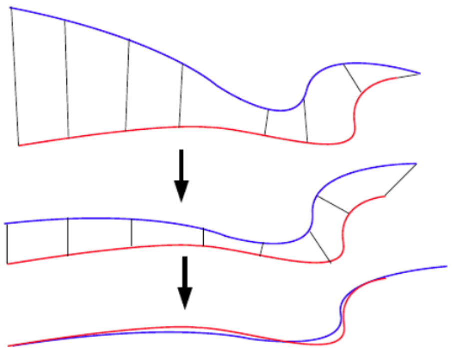


Iterative Closest Point (ICP)

[Besl and McKay 1992]

the image is taken from [Smistad *et al*. 2015]

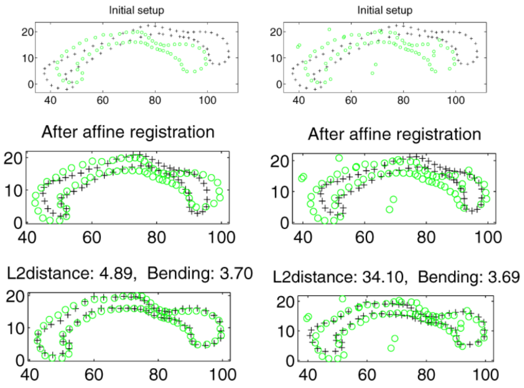
# Related Works,  General-Purpose Methods



Iterative Closest Point (ICP)

[Besl and McKay 1992]

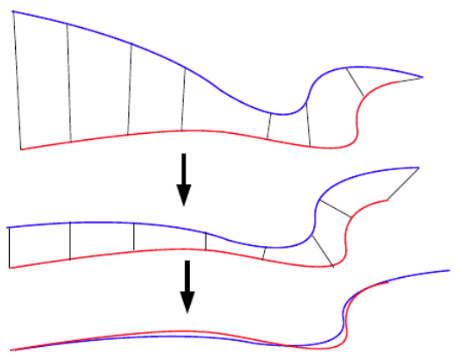
the image is taken from [Smistad *et al*. 2015]



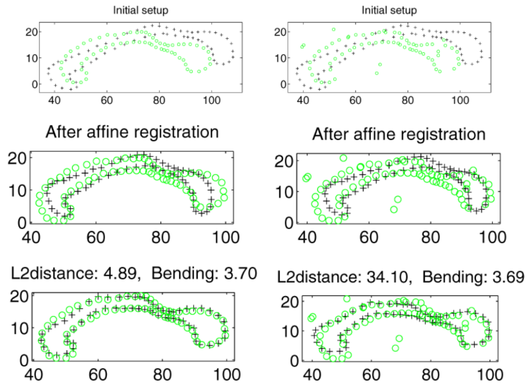

Gaussian Mixture Model Registration (GMR)

[Jian *et al*. 2005 ]

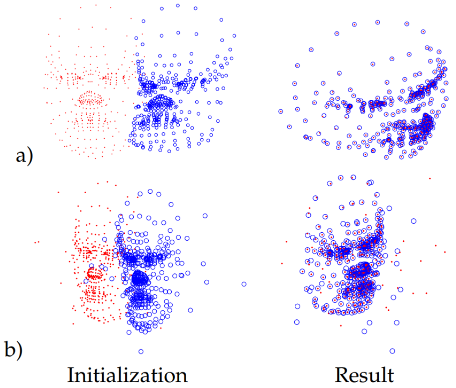
# Related Works, General-Purpose Methods



Iterative Closest Point (ICP)
[Besl and McKay 1992]
the image is taken from [Smistad *et al*. 2015]



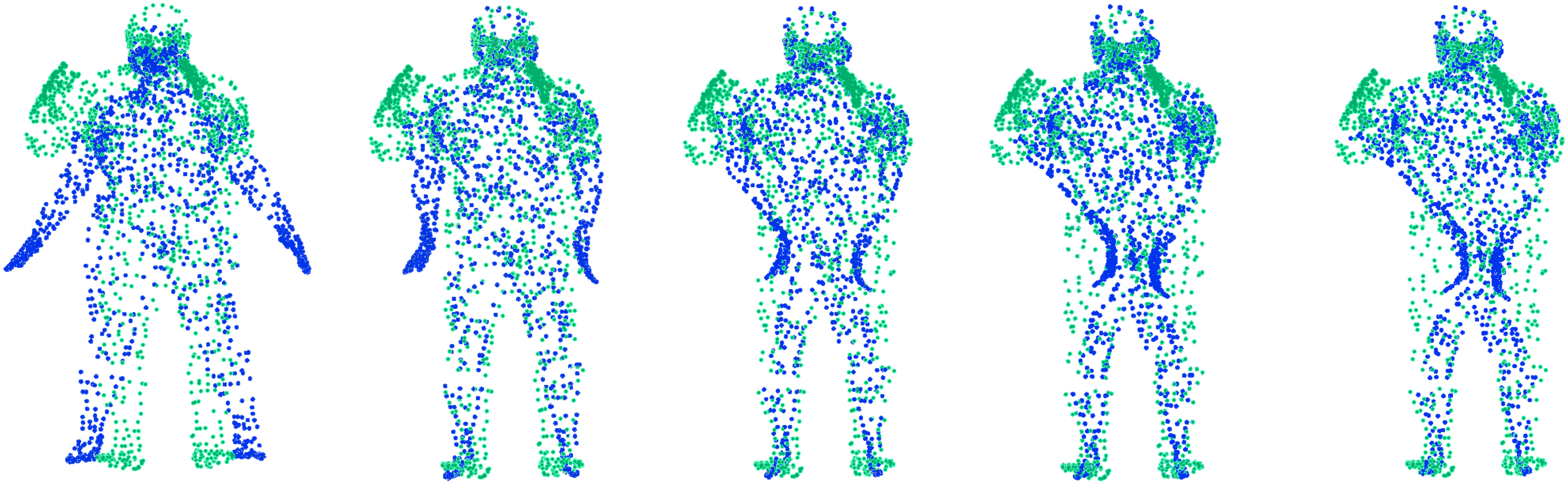

Gaussian Mixture Model Registration (GMR)
[Jian *et al*. 2005 ]




Coherent Point Drift (CPD)
[Myronenko and Song 2010 ]

# Related Works, General-Purpose Methods

Green : reference
Blue : template

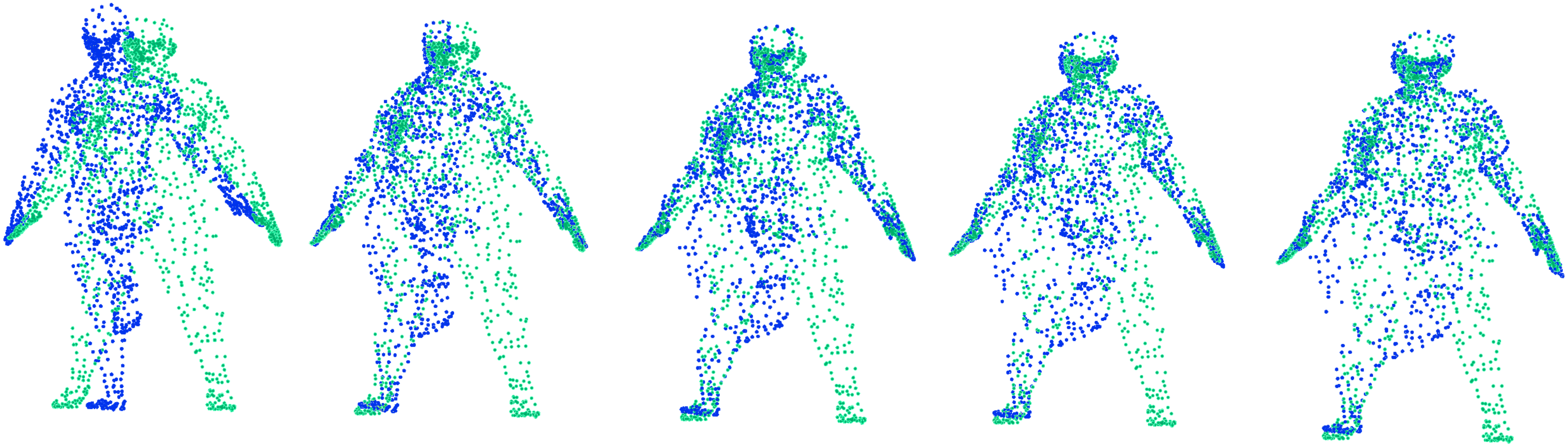


Initialisation Iteration 25 Iteration 50 Iteration 75 Iteration 100

Gravitational Approach for NRPSR
[Ali *et al.* 2018]

# Related Works, General-Purpose Methods

Green : reference
Blue : template

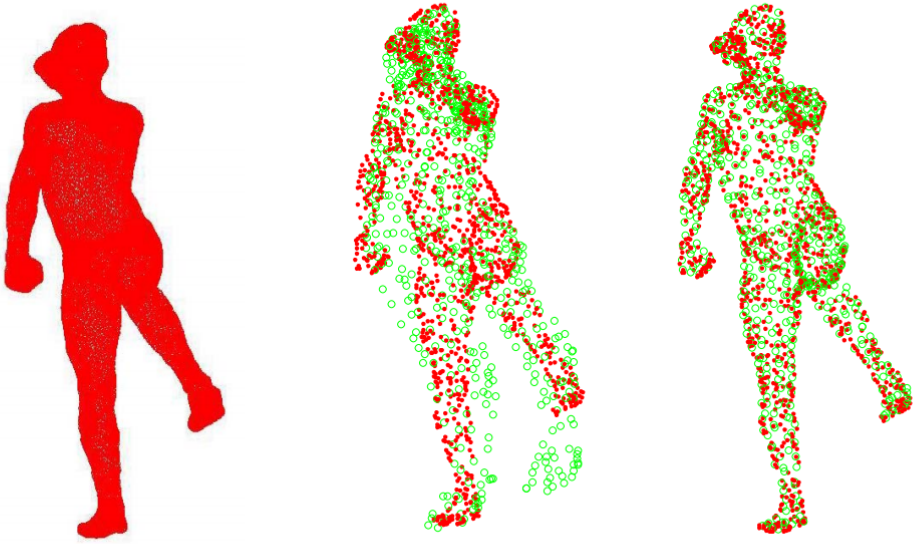


Initialisation Iteration 25 Iteration 50 Iteration 75 Iteration 100

Gravitational Approach for NRPSR
[Ali *et al.* 2018]

➡ Often fails with large deformations and articulated motions between the point sets.

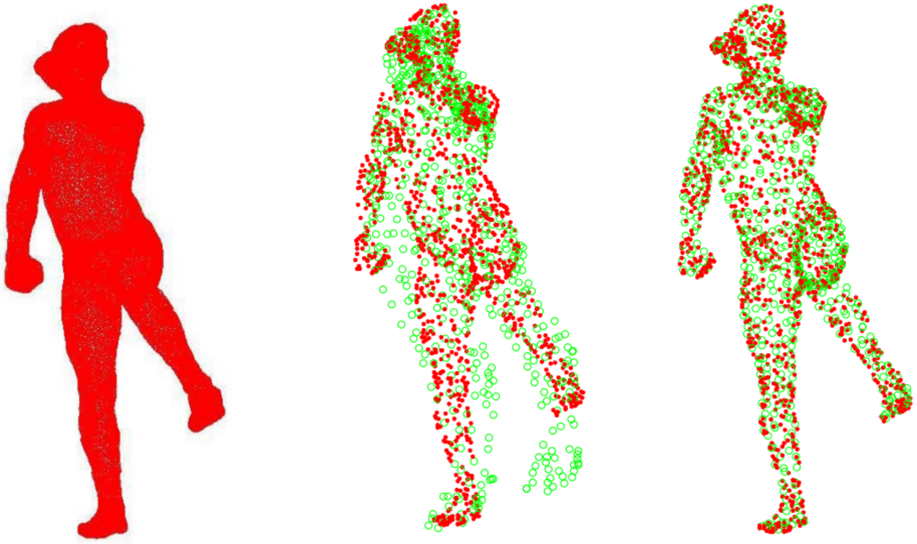
# Related Works, General-Purpose Methods



Target CPD GLTP

[Ge *et al*. 2014 ]

# Related Works, General-Purpose Methods

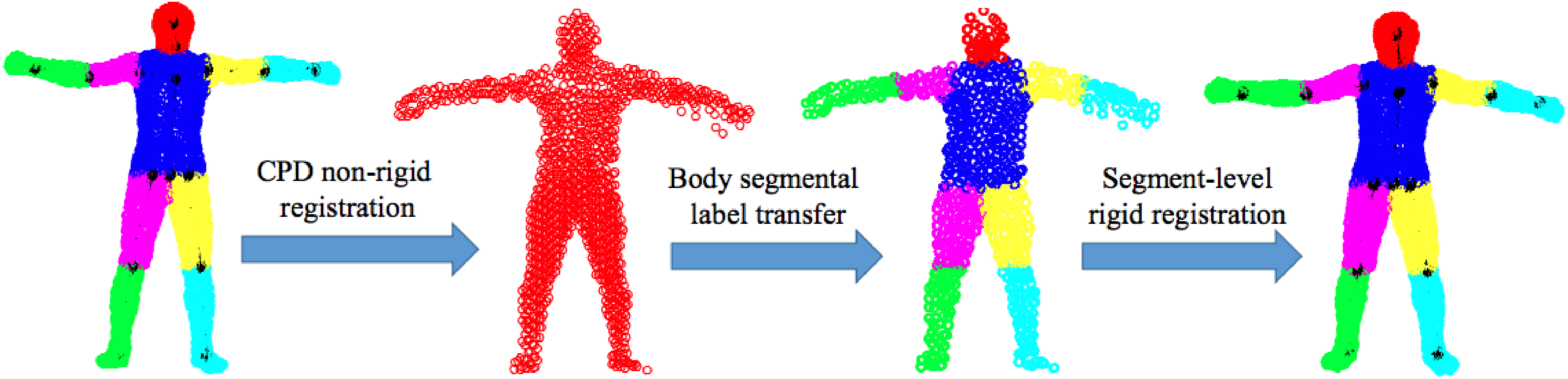


Target CPD GLTP

[Ge *et al.* 2014 ]

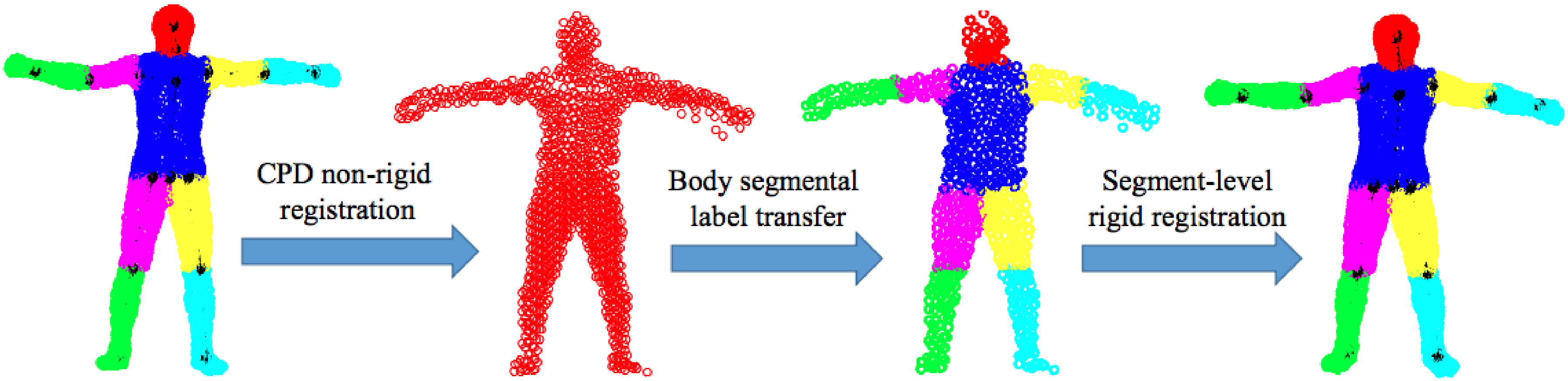
Relatively accurate however sensitive to noises

# Related Works,  Class-Specific Methods




[Ge and Fan 2015 ]

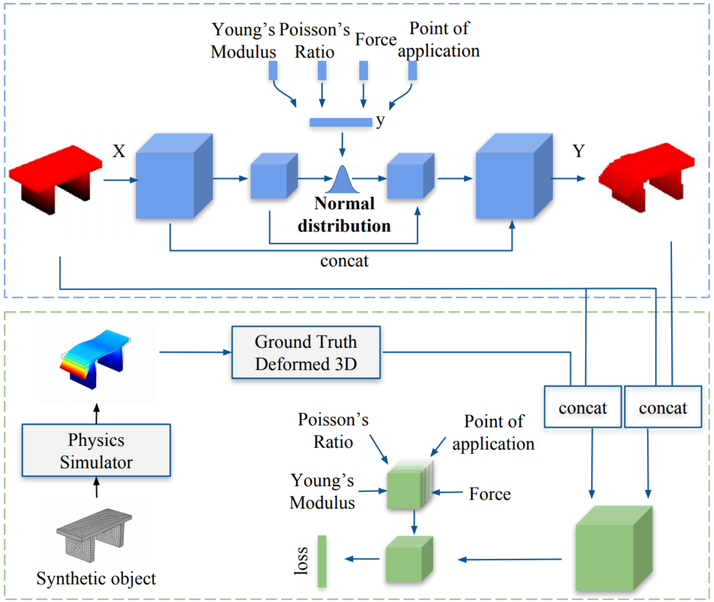
# Related Works, Class-Specific Methods




[Ge and Fan 2015 ]

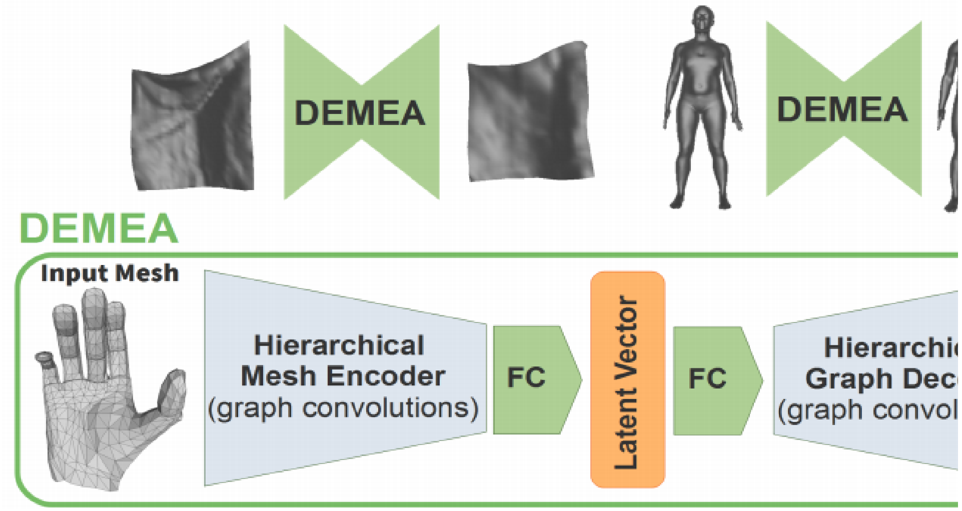
Perform well with large deformations and articulated motions between the point sets. However, the generalisability is limited.

# Related Works,  Neural Network Based Approaches (Other Fields)




3D-PhysNet

[Wang *et al.* 2018 ]



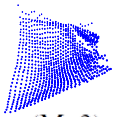

…

DEMEA

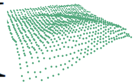
[Tretschk *et al.* 2019 ]

# Pipeline

# Pipeline

Y 

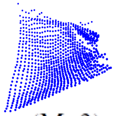
(Mx3)

X 

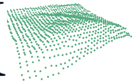
(Nx3)

- Y: template point set,  X: reference point set

# Pipeline

Y 

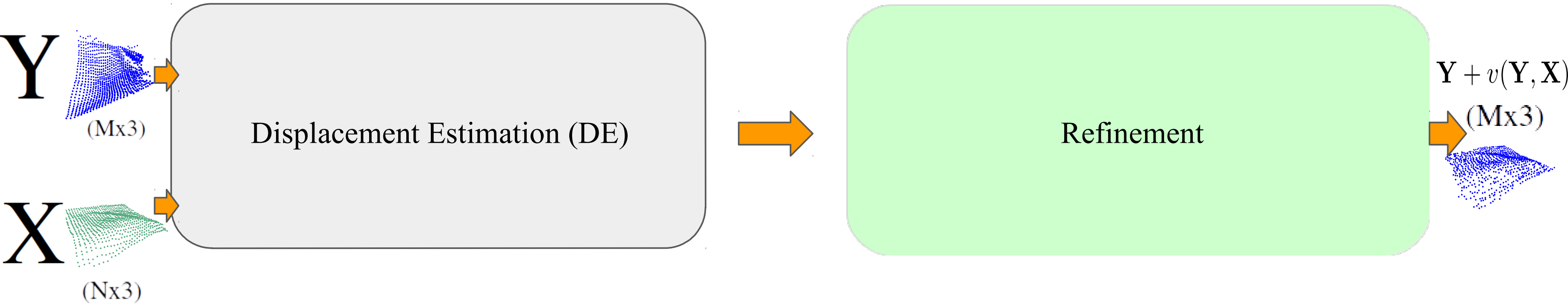
(Mx3)

X 

(Nx3)

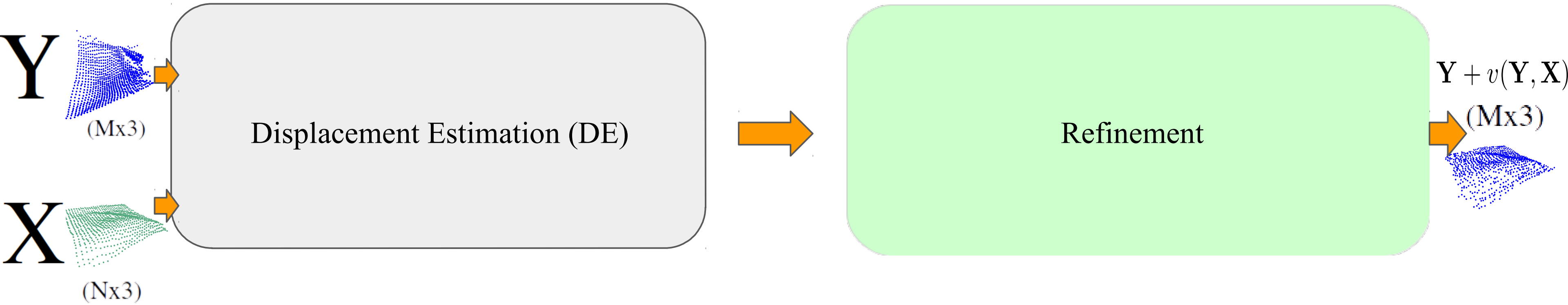
- Y: template point set,  X: reference point set
- Assume M is not equal to N in general

# Pipeline



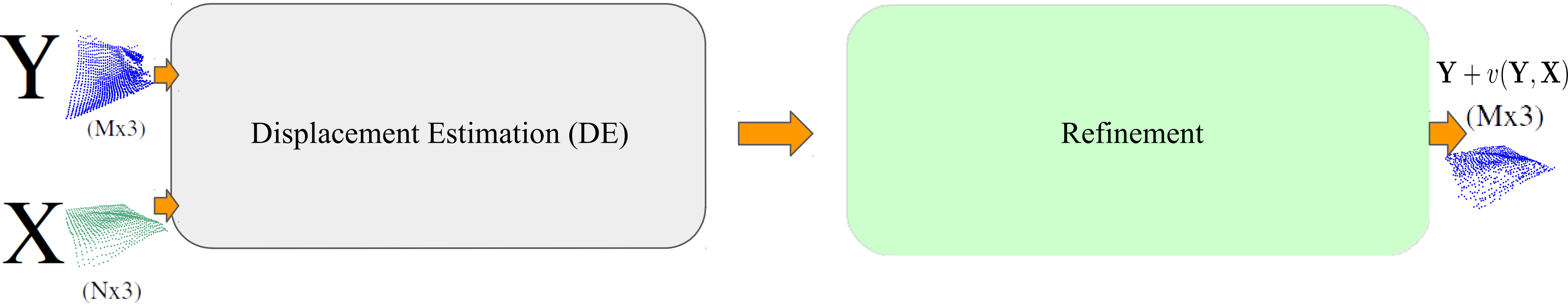

- Y: template point set, X: reference point set
- Assume M is not equal to N in general

# Pipeline



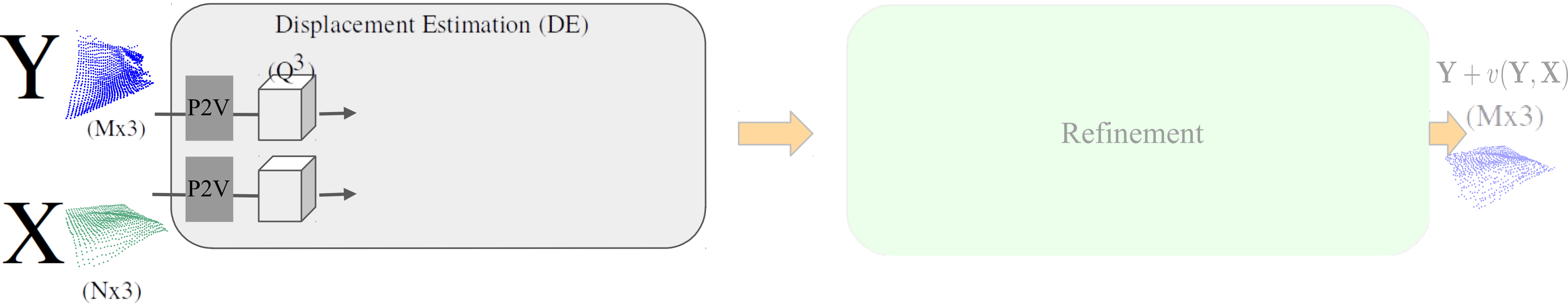

- Y: template point set,  X: reference point set
- Assume M is not equal to N in general
- DE stage regresses global displacements between Y and X

# Pipeline



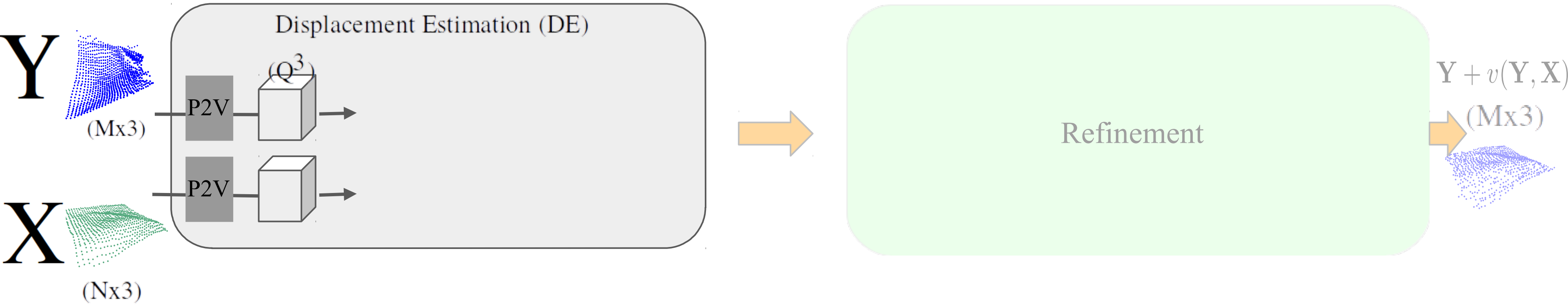

- Y: template point set,  X: reference point set
- Assume M is not equal to N in general
- DE stage regresses global displacements between Y and X
- Refinement stage improves the initial displacements

# Pipeline



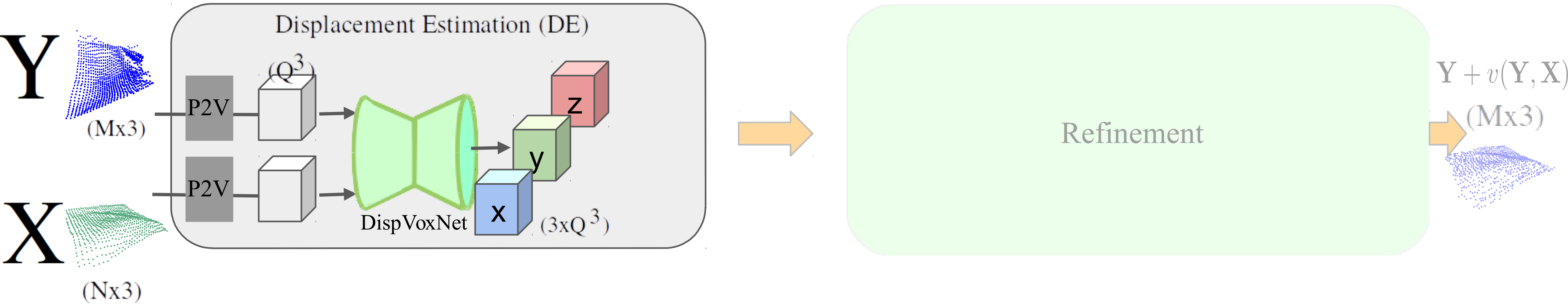

- Y and X are firstly converted into voxel representation (P2V)

# Pipeline



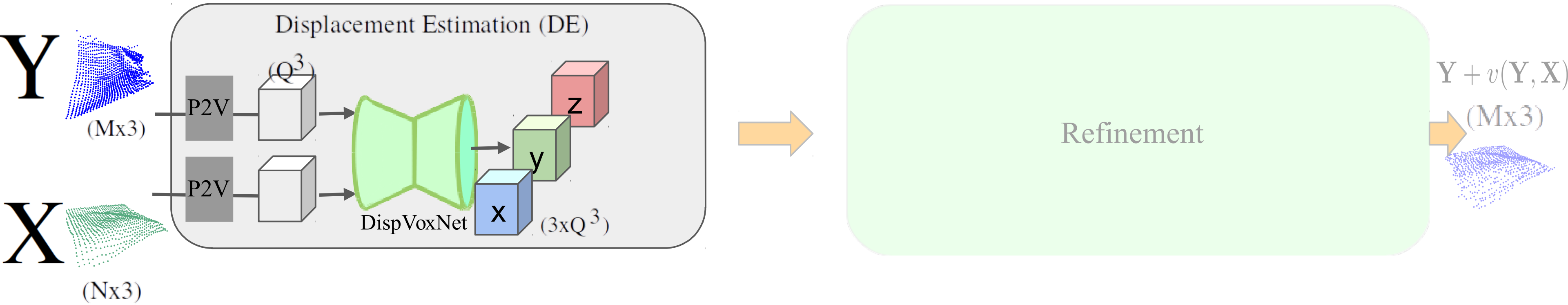

- Y and X are firstly converted into voxel representation (P2V)
- During the conversion, point-voxel correspondence information is stored in an **affinity table**

# Pipeline



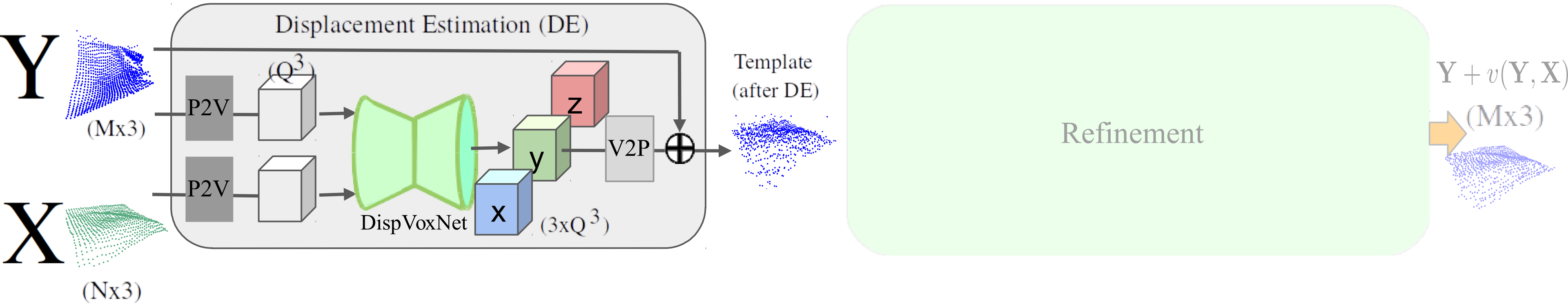

- Y and X are firstly converted into voxel representation (P2V)
- During the conversion, point-voxel correspondence information is stored in an **affinity table**

# Pipeline



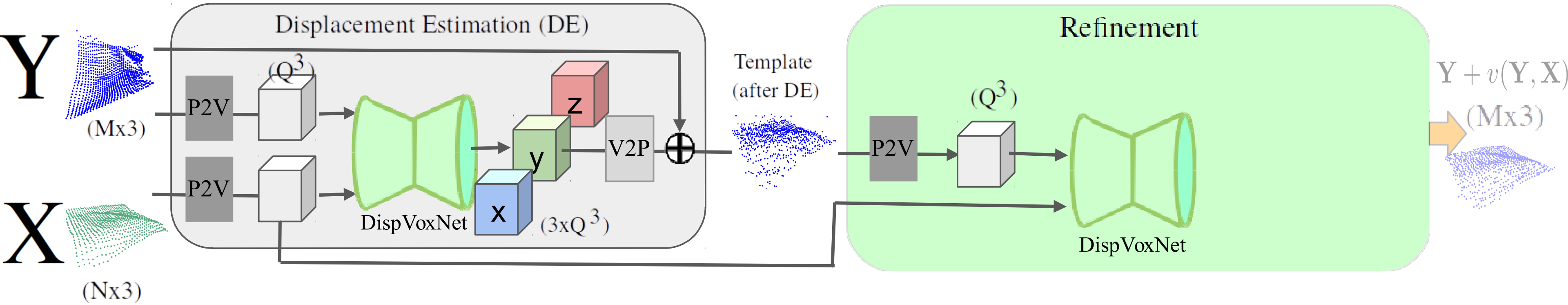

- Y and X are firstly converted into voxel representation (P2V)
- During the conversion, point-voxel correspondence information is stored in an **affinity table**
- DispVoxNet accepts two voxel grids and returns voxel displacements

# Pipeline



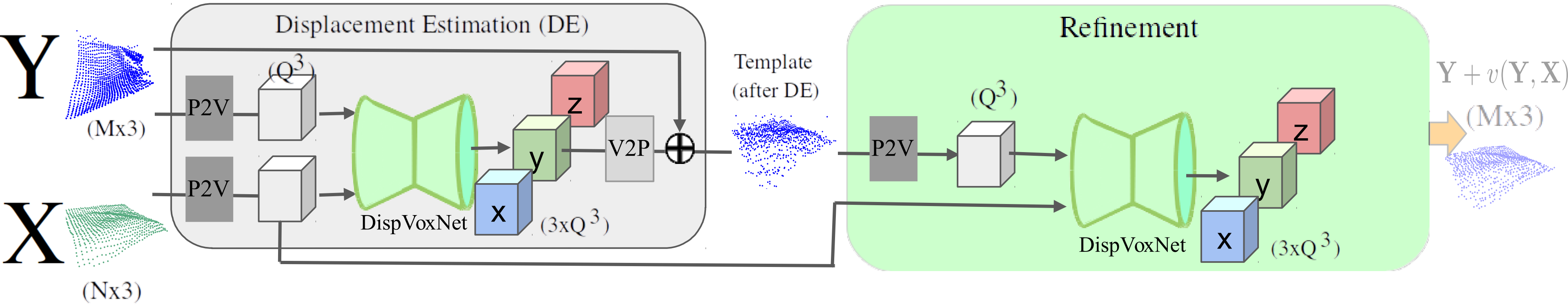

- Y and X are firstly converted into voxel representation (P2V)
- During the conversion, point-voxel correspondence information is stored in an **affinity table**
- DispVoxNet accepts two voxel grids and returns voxel displacements
- The displacements are applied using the affinity table at the end of DE stage

# Pipeline



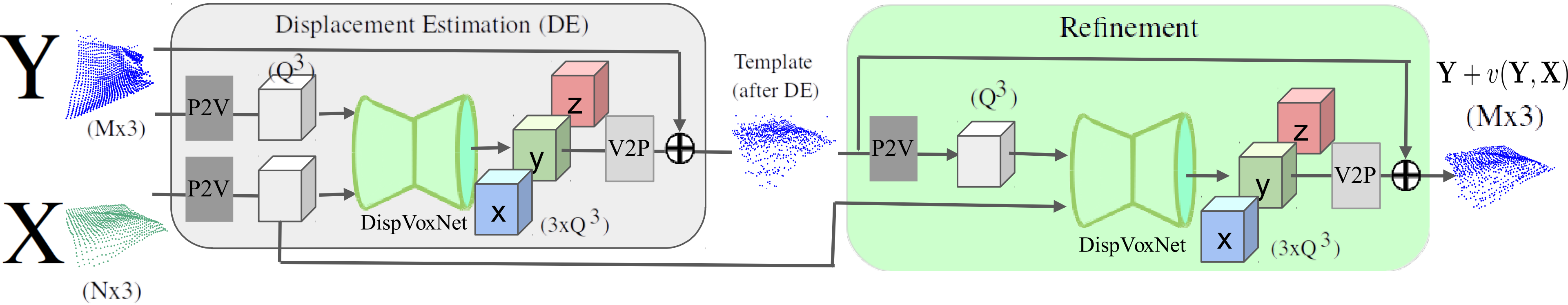

- The outputs from the DE stage are further sent to the Refinement stage after P2V

# Pipeline



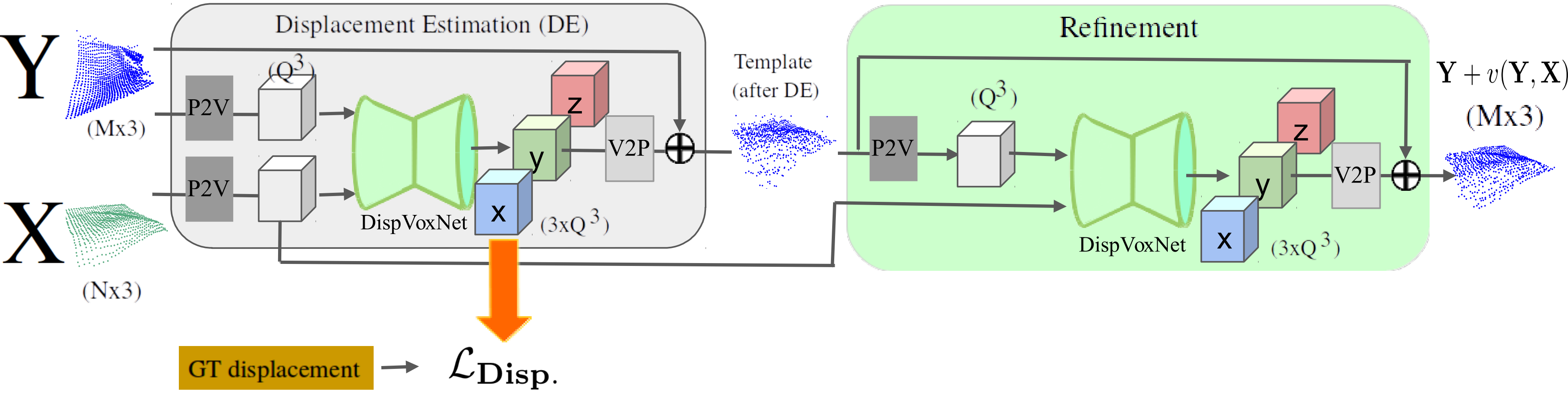

- The outputs from the DE stage are further sent to the Refinement stage after P2V
- The new instance of DispVoxNet returns small displacements for refinement

# Pipeline



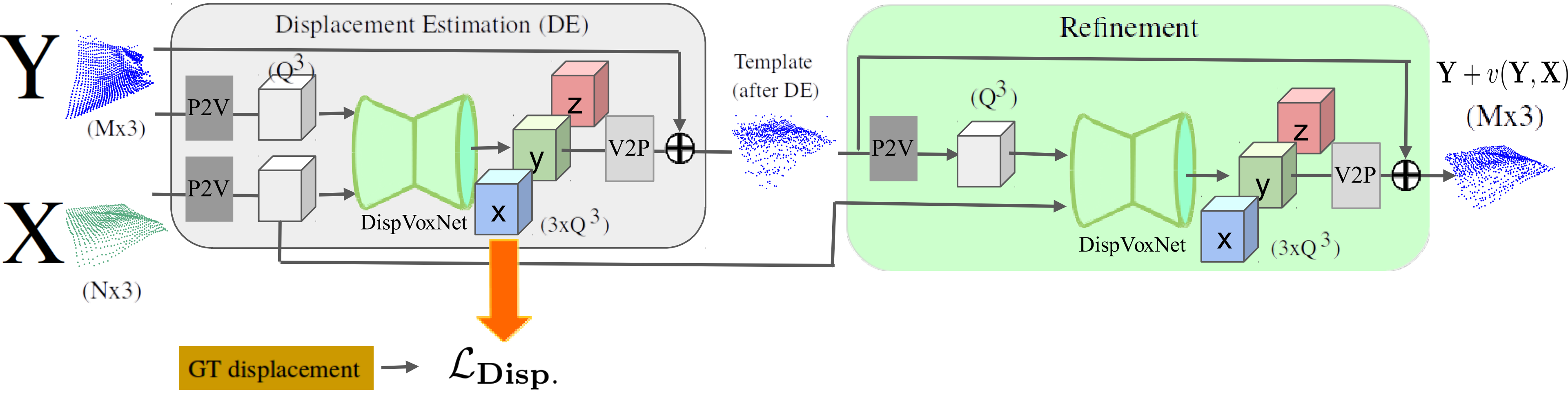

- The outputs from the DE stage are further sent to the Refinement stage after P2V
- The new instance of DispVoxNet returns small displacements for refinement
- The inferred displacements are added to the template points

# Pipeline



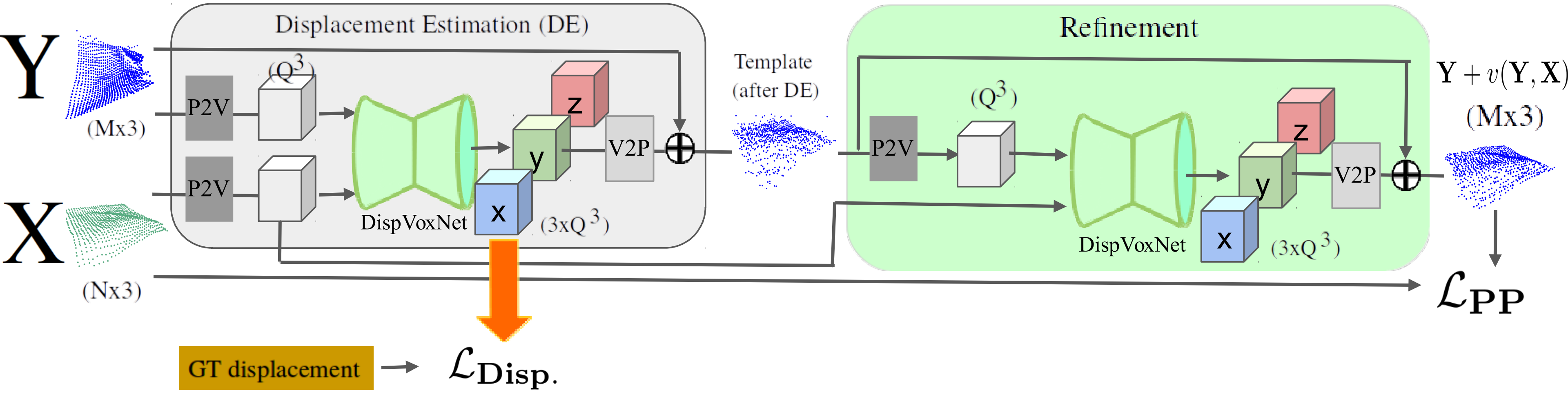
Displacement Estimation (DE)
Y
(Mx3)
X
(Nx3)
P2V
P2V
($Q^3$)
DispVoxNet
z
y
x
($3xQ^3$)
V2P
Template
(after DE)
Refinement
P2V
($Q^3$)
DispVoxNet
z
y
x
($3xQ^3$)
V2P
$\mathbf{Y} + v(\mathbf{Y}, \mathbf{X})$
(Mx3)
GT displacement
$\mathcal{L}_{\mathbf{Disp.}}$

# Pipeline




- The network in the DE stage is trained in a supervised manner (displacement loss)

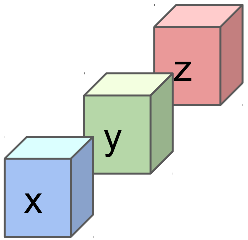
# Pipeline



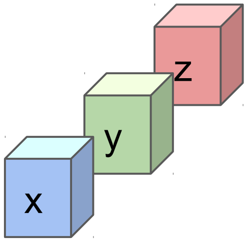

- The network in the DE stage is trained in a supervised manner (displacement loss)
- The network in the Refinement stage is trained in an unsupervised manner (point projection loss)

# Loss Functions - Displacement Loss

$$\mathcal{L}_{\mathbf{Disp.}} = \Bigg\| \quad \text{Network output} \quad - \quad \text{GT displacement} \quad \Bigg\|_2^2$$

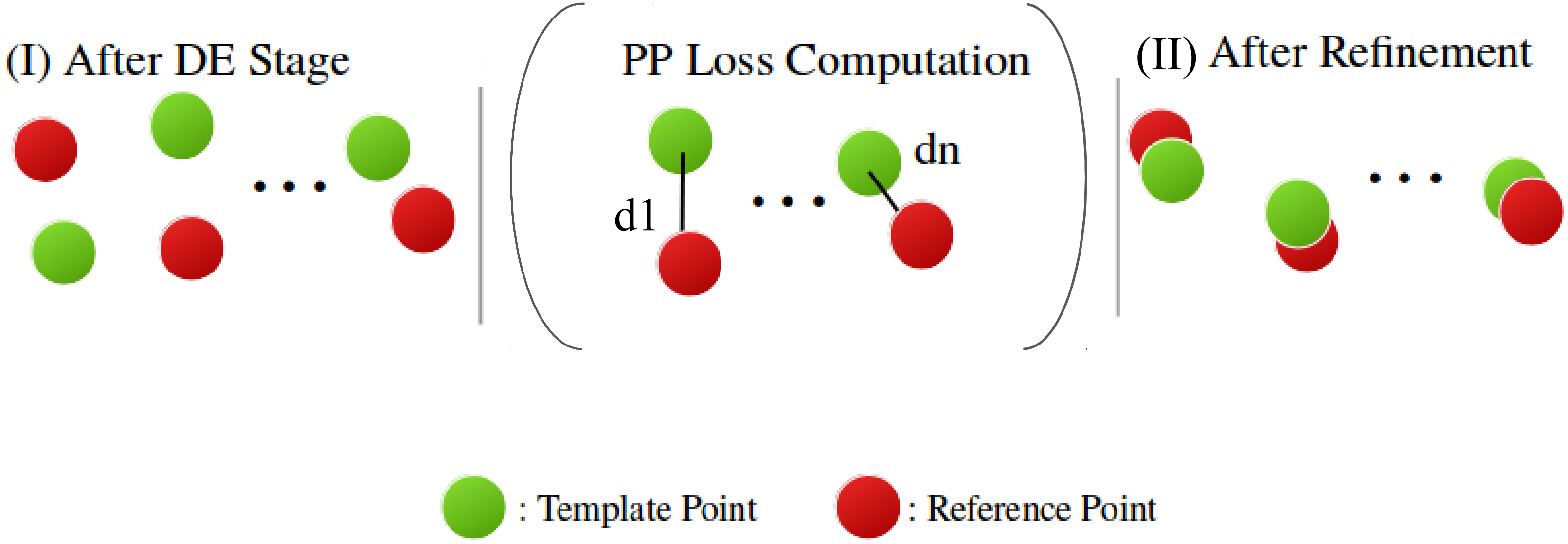



Network output

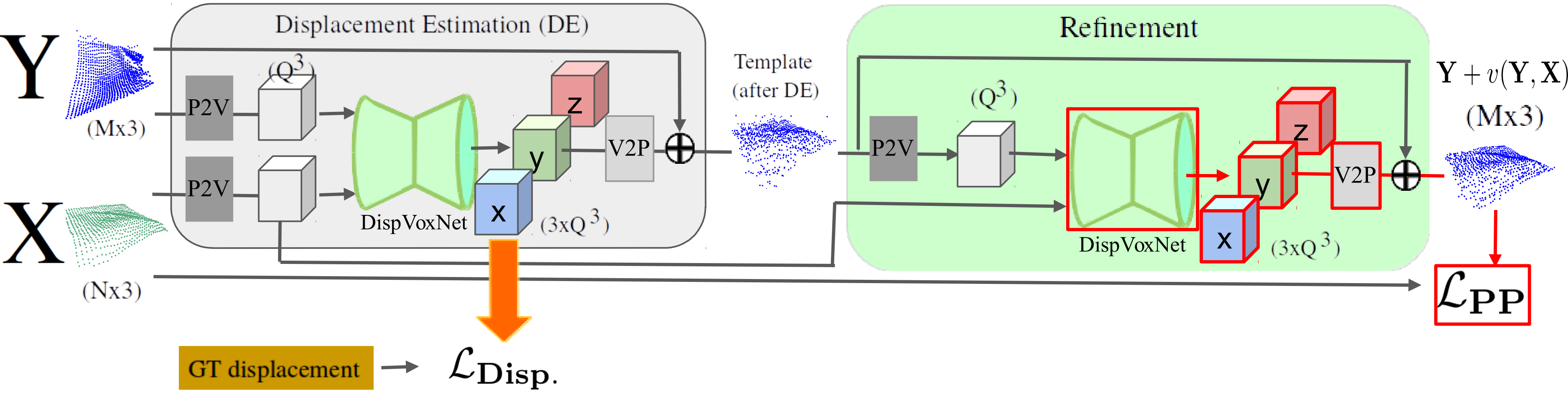



GT displacement

# Loss Functions - Point Projection Loss



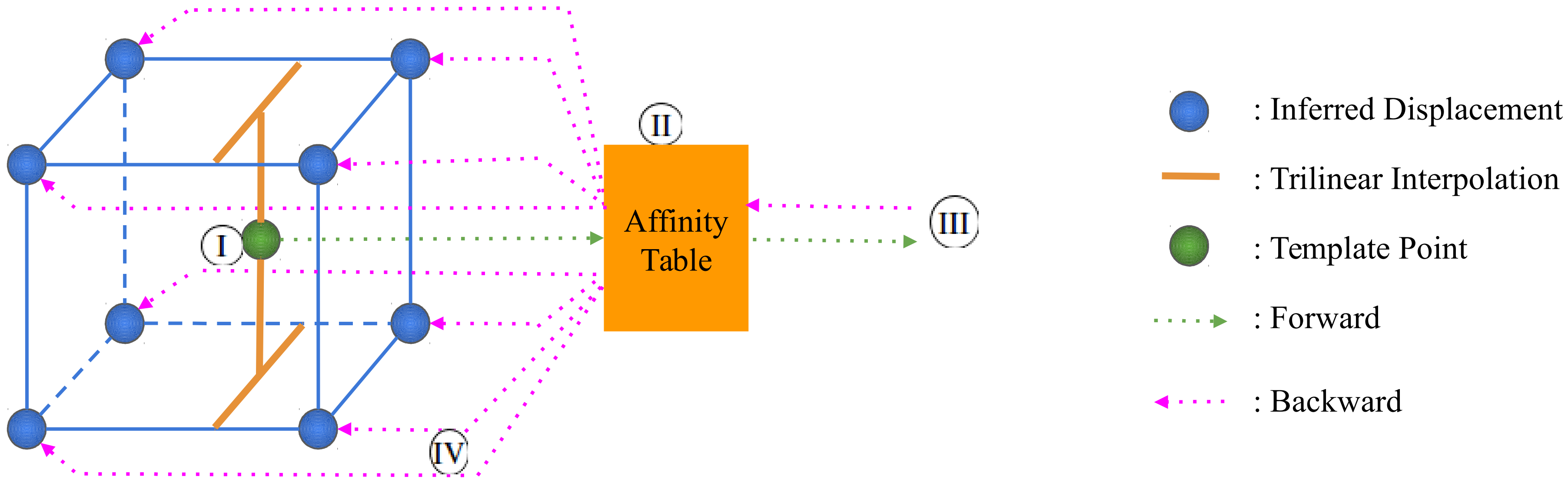
# Pipeline




Problem 1: Discretisation effect due to the nature of voxel grids

Problem 2: Indifferentiability problem

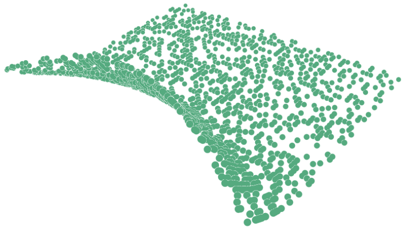
# Solution for Discretisation and Indifferentiability Problems




I. Compute trilinear weights for each template point using its 8 nearest inferred displacements
II. Record the weights and indices of the 8 nearest displacements in the affinity table
III. Compute the point projection loss
IV. Distribute gradients following the IDs and weights information recorded in the affinity table in II.

# Datasets

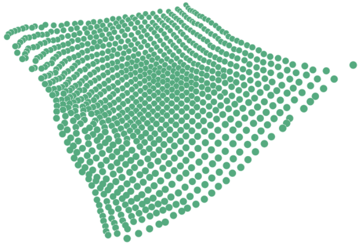
# Datasets



*thin plate*
[Golyanik *et al*. 2018]

*FLAME*
[Li *et al*. 2017]

*DFAUST*
[Bogo *et al*. 2017]

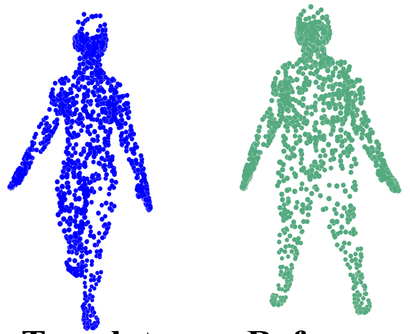


*cloth*
[Bednařík *et al*. 2018]

# Evaluation

# Quantitative Results - Baseline and Outliers

# Quantitative Results - Baseline and Outliers



**Template** **Reference**

| | | Ours | NR-ICP [9] | CPD [38] | GMR [29] |
|---|---|---|---|---|---|
| ***thin plate*[17]** | $e$ | 0.0103 | 0.0402 | **0.0083** / 0.0192 | 0.2189 |
| | $\sigma$ | **0.0059** | 0.0273 | 0.0102 / 0.0083 | 1.0121 |
| ***FLAME*[33]** | $e$ | 0.0063 | 0.0588 | **0.0043** / 0.0094 | 0.0056 |
| | $\sigma$ | 0.0009 | 0.0454 | 0.0008 / **0.0005** | 0.0007 |
| ***DFAUST*[5]** | $e$ | **0.0166** | 0.0585 | 0.0683 / 0.0721 | 0.2357 |
| | $\sigma$ | **0.0020** | 0.0215 | 0.0314 / 0.0258 | 0.8944 |
| ***cloth*[2]** | $e$ | **0.0080** | 0.0225 | 0.0149 / 0.0138 | 0.2189 |
| | $\sigma$ | **0.0021** | 0.0075 | 0.0066 / 0.0033 | 1.0121 |

Baseline Comparison

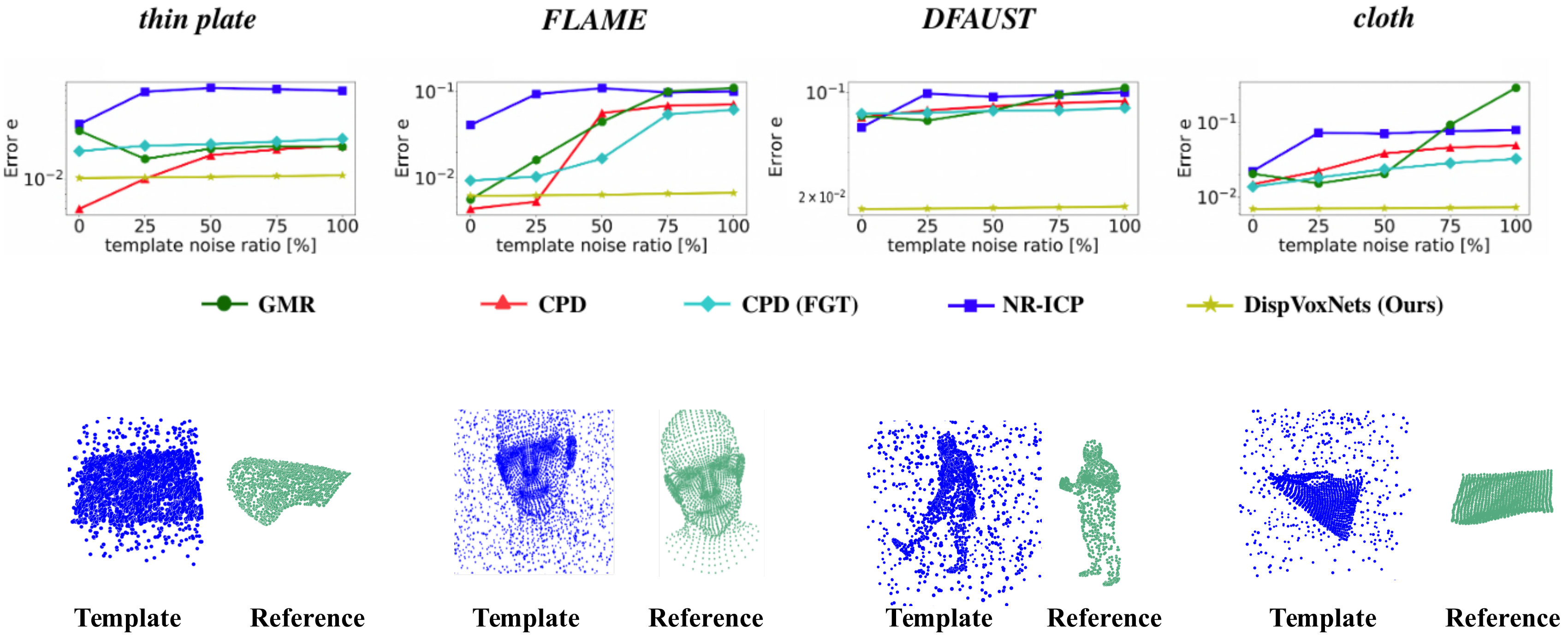
**Template** **Reference**

| | | | Ours | NR-ICP [9] | CPD [38] | GMR [29] |
|---|---|---|---|---|---|---|
| ***thin plate*[17]** | ref. | $e$ | **0.0107** | 0.0668 | 0.0218 / 0.0386 | 0.4415 |
| | | $\sigma$ | **0.0061** | 0.0352 | 0.0148 / 0.0067 | 1.4632 |
| | temp. | $e$ | **0.0108** | 0.0334 | 0.0479 / 0.0471 | 0.4287 |
| | | $\sigma$ | 0.0062 | 0.0281 | 0.0101 / **0.0038** | 1.3832 |
| ***FLAME*[33]** | ref. | $e$ | 0.0084 | 0.0519 | **0.0046** / 0.0140 | 0.0193 |
| | | $\sigma$ | 0.0010 | 0.0451 | 0.0009 / **0.0006** | 0.0008 |
| | temp. | $e$ | 0.0088 | 0.0215 | **0.0076** / 0.0201 | 0.0274 |
| | | $\sigma$ | **0.0010** | 0.0219 | **0.0010** / 0.0016 | 0.0019 |
| ***DFAUST*[5]** | ref. | $e$ | **0.0167** | 0.0463 | 0.0562 / 0.0636 | 0.0714 |
| | | $\sigma$ | **0.0029** | 0.0195 | 0.0308 / 0.0216 | 0.0282 |
| | temp. | $e$ | **0.0169** | 0.0426 | 0.0672 / 0.0710 | 0.0737 |
| | | $\sigma$ | **0.0033** | 0.0194 | 0.0291 / 0.0229 | 0.0243 |
| ***cloth*[2]** | ref. | $e$ | **0.0090** | 0.0455 | 0.0248 / 0.0315 | 0.0288 |
| | | $\sigma$ | **0.0018** | 0.0061 | 0.0056 / 0.0027 | 0.0087 |
| | temp. | $e$ | **0.0132** | 0.0208 | 0.0486 / 0.0347 | 0.0397 |
| | | $\sigma$ | 0.0019 | 0.0087 | 0.0077 / **0.0014** | 0.0092 |

Outlier

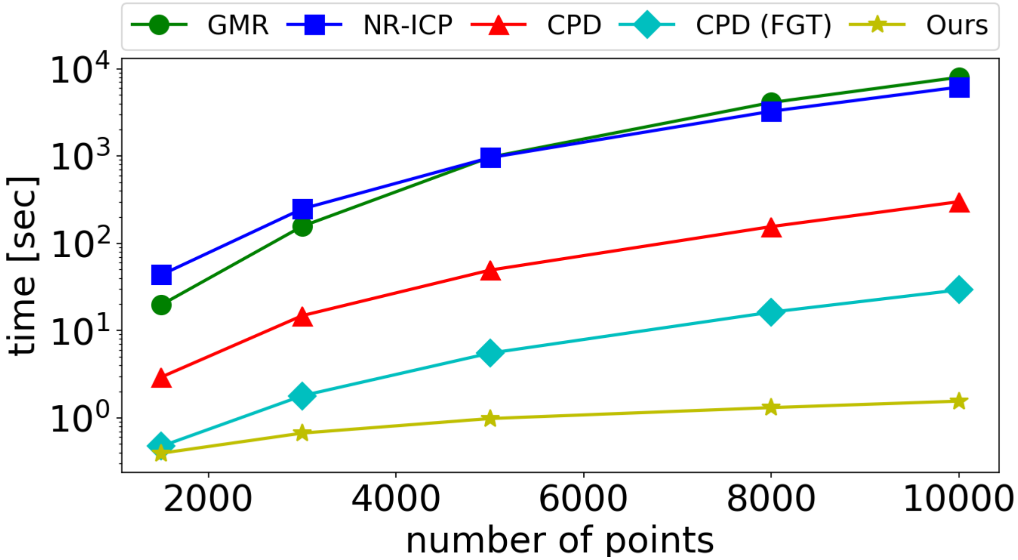
# Quantitative Results - Uniform Noises

# Quantitative Results - Uniform Noises



# Quantitative Results - Runtime

# Quantitative Results - Runtime

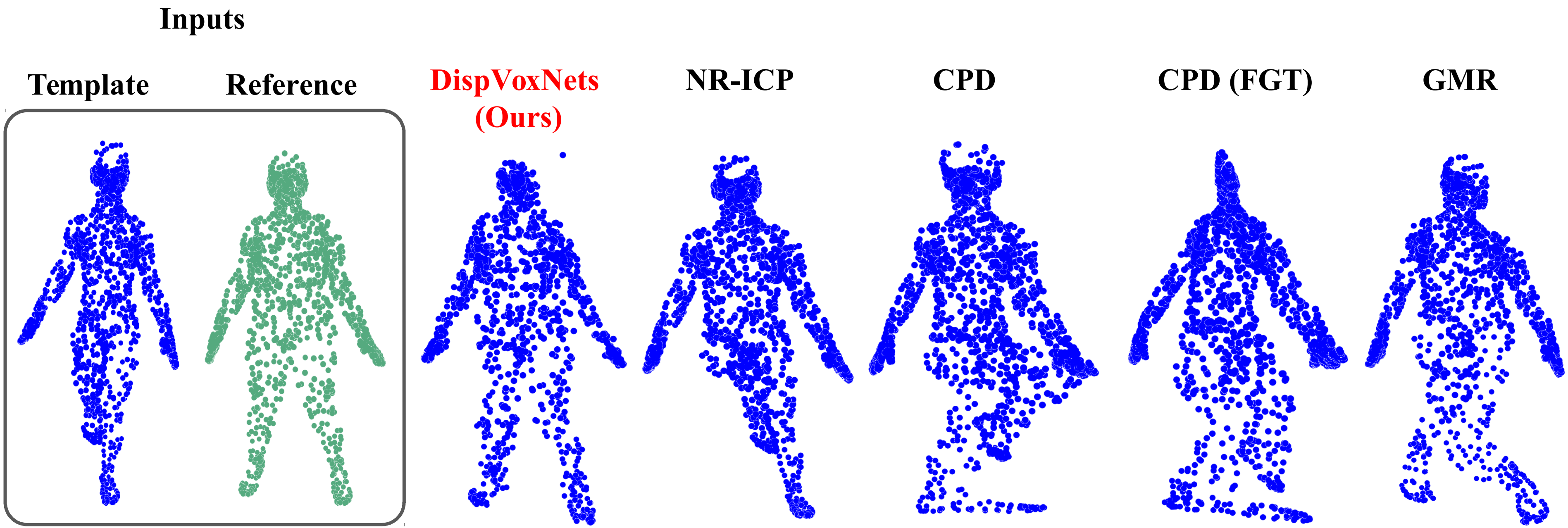



- With 10K points, our approach requires only a second per registration whereas others require around 2 hours - 15 seconds

# Qualitative Results

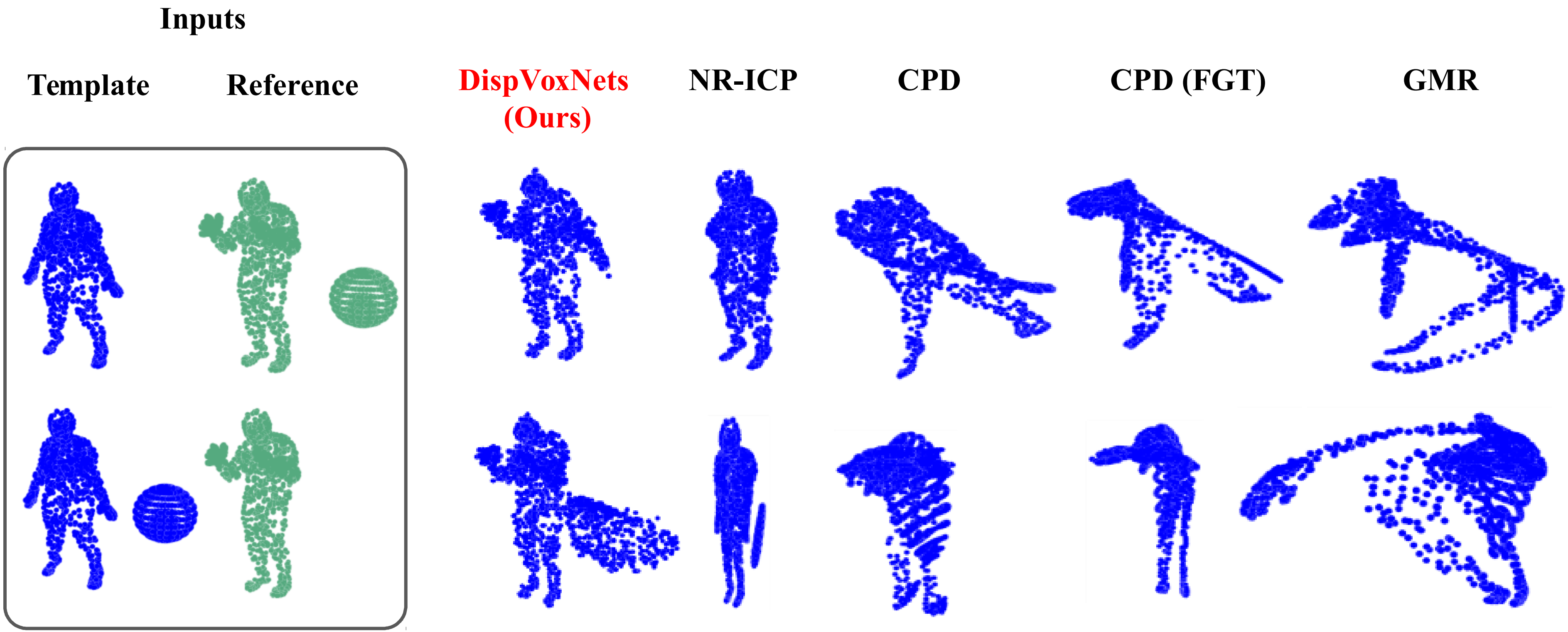
# Baseline Comparison

# Baseline Comparison



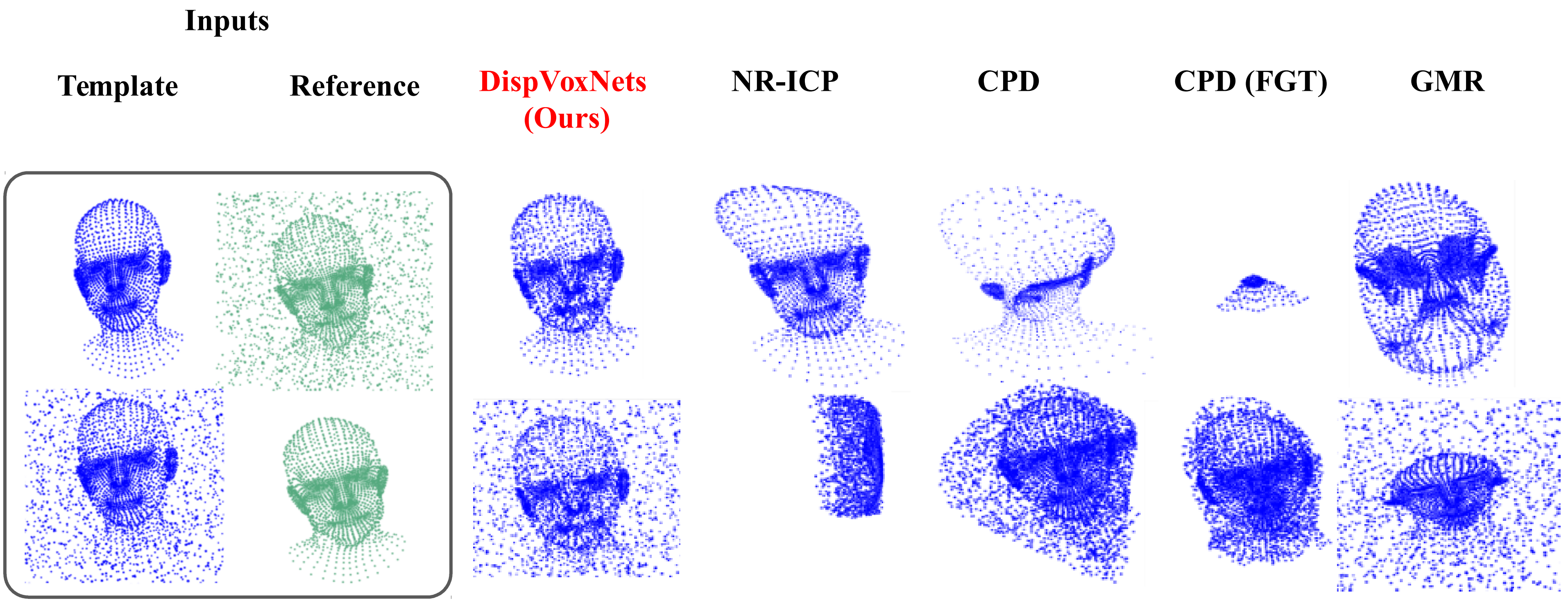
# Outliers

# Outliers


Inputs
Template
Reference
DispVoxNets (Ours)
NR-ICP
CPD
CPD (FGT)
GMR

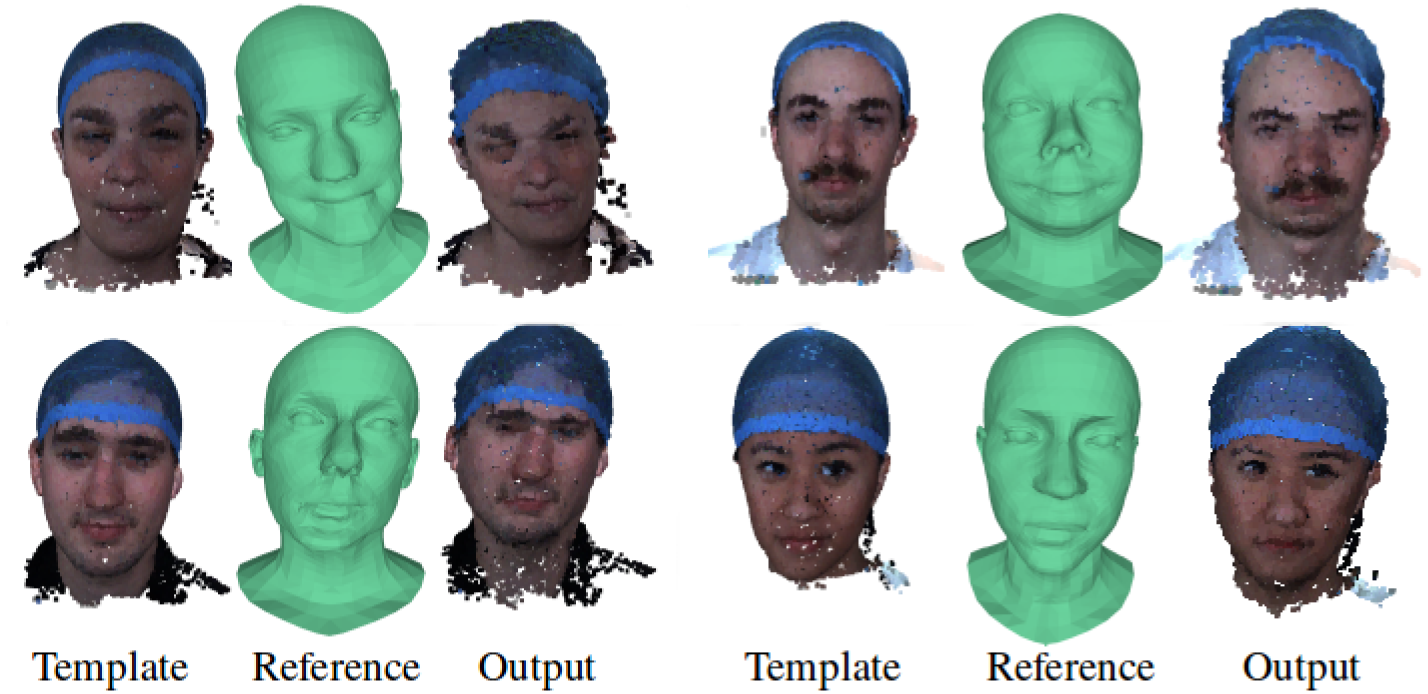
# Uniform Noises

# Uniform Noises



# Real Face Dataset

# Real Face Dataset




Datasets: [Dai et al. 2017], [Li *et al.* 2017]

# Summary

# Summary

# Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

# Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

# Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

- Runs orders of magnitude faster than previous techniques.

# Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

- Runs orders of magnitude faster than previous techniques.

- Limitation: topology preserving is not yet fully satisfying.

# Summary

- To the best of our knowledge, this is the first neural network based approach for NRPSR that is invariant to the number and order of points.

- Our approach outperforms other existing general-purpose methods in the presence of large deformations, articulated motion, noise, outliers and missing data.

- Runs orders of magnitude faster than previous techniques.

- Limitation: topology preserving is not yet fully satisfying.

Project Page: http://gvv.mpi-inf.mpg.de/projects/DispVoxNets/

# References


1. J. Bednařík, P. Fua, and M. Salzmann. Learning to reconstruct texture-less deformable surfaces from a single view. In International Conference on 3D Vision (3DV), 2018.
2. F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black. Dynamic FAUST: Registering human bodies in motion. In Computer Vision and Pattern Recognition (CVPR), 2017.
3. Besl, Paul J., and Neil D. McKay. Method for registration of 3-D shapes. Sensor fusion IV: control paradigms and data structures. Vol. 1611. International Society for Optics and Photonics, 1992.
4. H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. In Computer Vision and Image Understanding (CVIU), 2003.
5. S. Ge, G. Fan, and M. Ding. Non-rigid point set registration with global-local topology preservation. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2014.
6. S. Ge, and G. Fan. Articulated non-rigid point set registration for human pose estimation from 3D sensors. Sensors, 15(7), 15218-15245.
7. V. Golyanik *et al.* Extended coherent point drift algorithm with correspondence priors and optimal subsampling. In IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.
8. V. Golyanik *et al.* A framework for an accurate point cloud based registration of full 3D human body scans. Fifteenth IAPR International Conference on Machine Vision Applications (MVA), 2017.
9. V. Golyanik, S. Shimada, K. Varanasi, and D. Stricker. Hdm-net: Monocular non-rigid 3d reconstruction with learned deformation model. In Virtual Reality and Augmented Reality (EuroVR), 2018.
10. B. Jian and B. C. Vemuri. A robust algorithm for point set registration using mixture of gaussians. In International Conference for Computer Vision (ICCV), 2005.
11. T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero. Learning a model of facial shape and expression from 4D scans. SIGGRAPH Asia, 2017.
12. A. Myronenko and X. Song. Point-set registration: Coherent point drift. Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2010.
13. E. Smistad, et al. Medical image segmentation on GPUs–A comprehensive review. In Medical image analysis, 2015.
14. Z. Wang, S. Rosa, and A. Markham. Learning the intuitive physics of non-rigid object deformations. In *Neural Information Processing Systems (NIPS) Workshops*, 2018.
15. E. Tretschk, A. Tewari, M. Zollhofer, V. Golyanik, and ¨ C. Theobalt. DEMEA: Deep Mesh Autoencoders for NonRigidly Deforming Objects. *arXiv e-prints*, 2019.
16. H. Dai, N. Pears, W. A. P. Smith, and C. Duncan. A 3d morphable model of craniofacial shape and texture variation. In ICCV, 2017.

# Questions?

# DispVoxNets: Non-Rigid Point Set Alignment with Supervised Learning Proxies

**Soshi Shimada**[1,2] **Vladislav Golyanik**[3] **Edgar Tretschk**[3]

**Didier Stricker**[1,2] **Christian Theobalt**[3]

[1]University of Kaiserslautern [2]DFKI [3]MPI for Informatics, SIC





# Thank you