RFNet-4D: Joint Object Reconstruction and Flow Estimation from 4D Point Clouds

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY · DEAKIN UNIVERSITY AUSTRALIA · VinAI RESEARCH · woven planet

ECCV TEL AVIV 2022

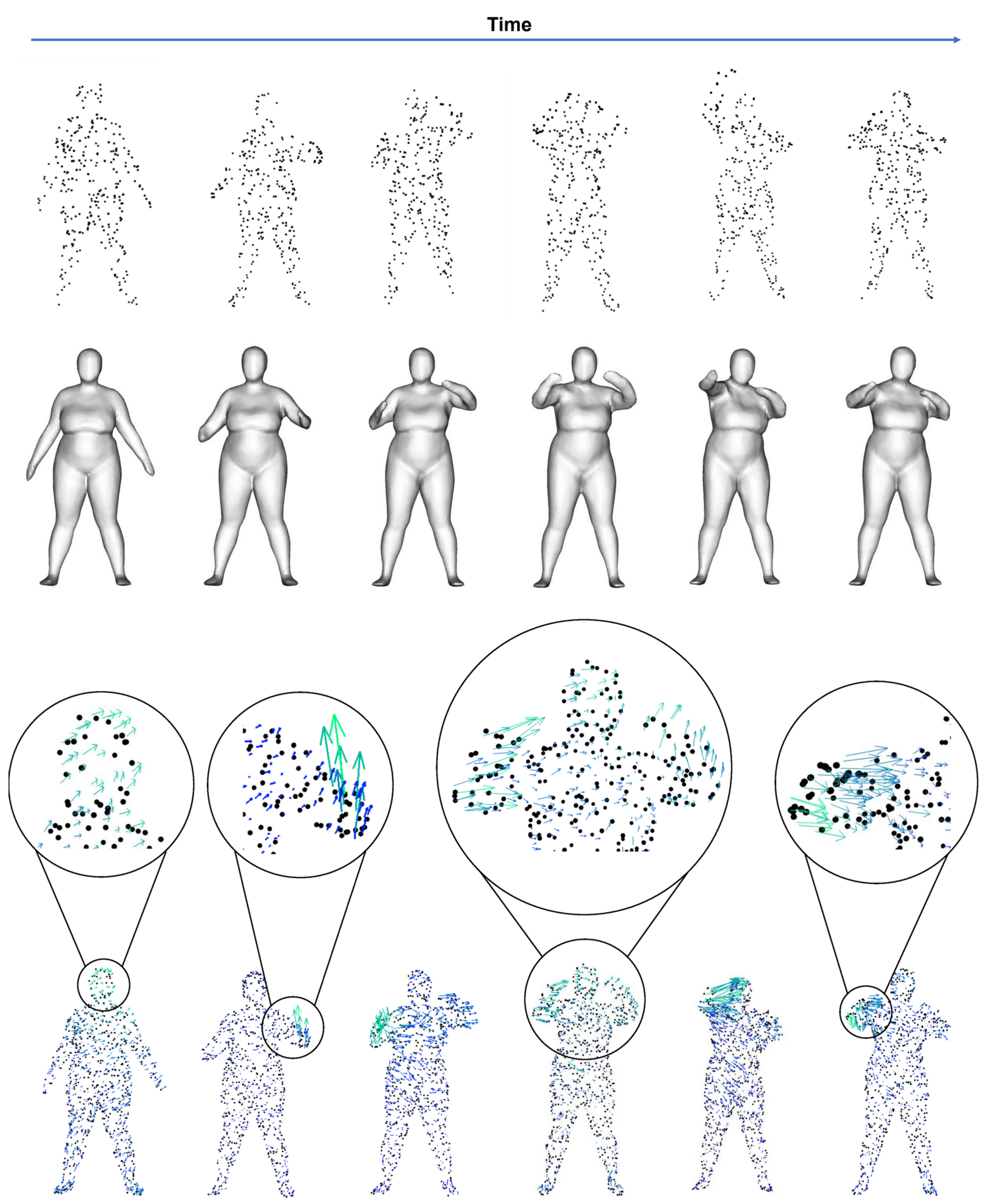
Tuan-Anh Vu   Duc-Thanh Nguyen   Binh-Son Hua   Quang-Hieu Pham   Sai-Kit Yeung

MOTIVATION

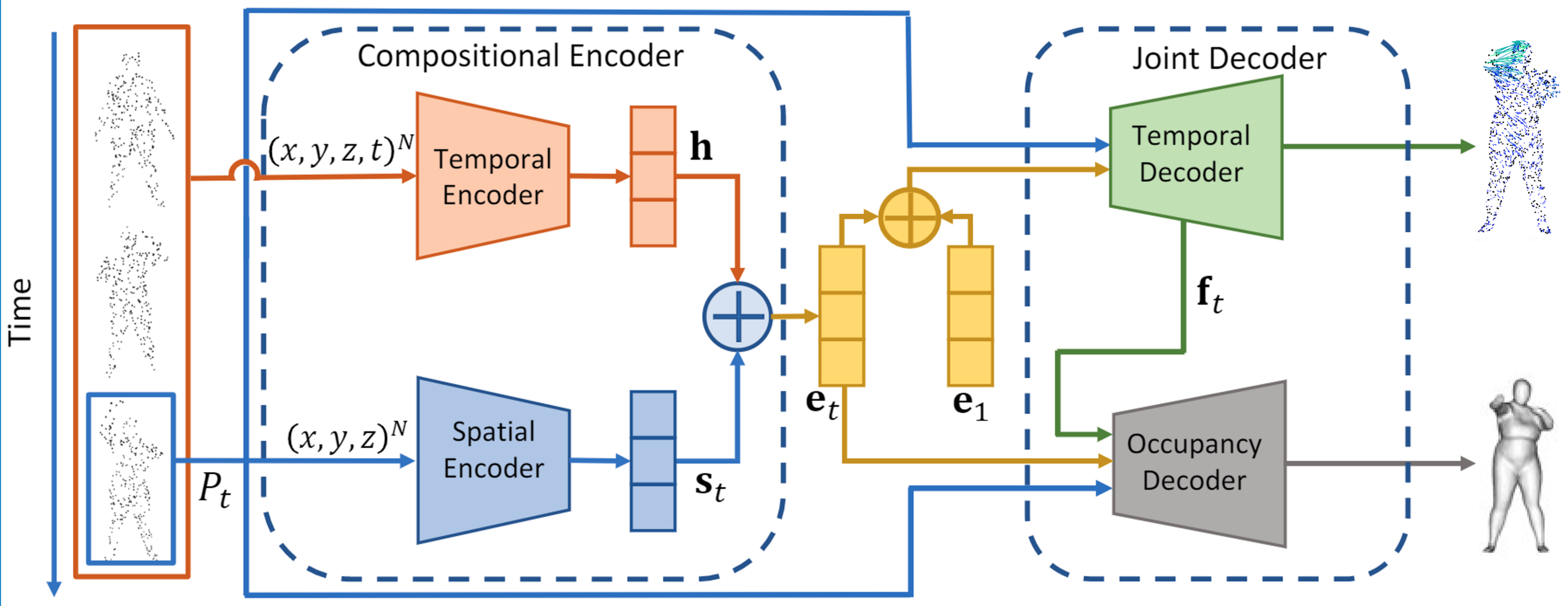
- Object reconstruction from 3D point clouds has achieved impressive progress in the computer vision and computer graphics research field. However, reconstruction from time-varying point clouds (a.k.a. 4D point clouds) is generally overlooked.
- Recent works still have many drawbacks and exist many challenges to solve, such as:
  - Ignore the aggregation of shape properties from multiple frames.
  - Inefficiently capture temporal dynamics.
  - Low computational efficiency (due to solving neural ordinary differential equations).
  - Large amount of annotated data (for supervision training).

OUR REPRESENTATION

- Represent **motion** by a temporally and spatially **continuous vector field.**
- Represent **3D shape** as the continuous decision boundary of a binary classifier:
  - Spatially and temporally continuous
  - Implicit correspondences over time
  - Fast inference



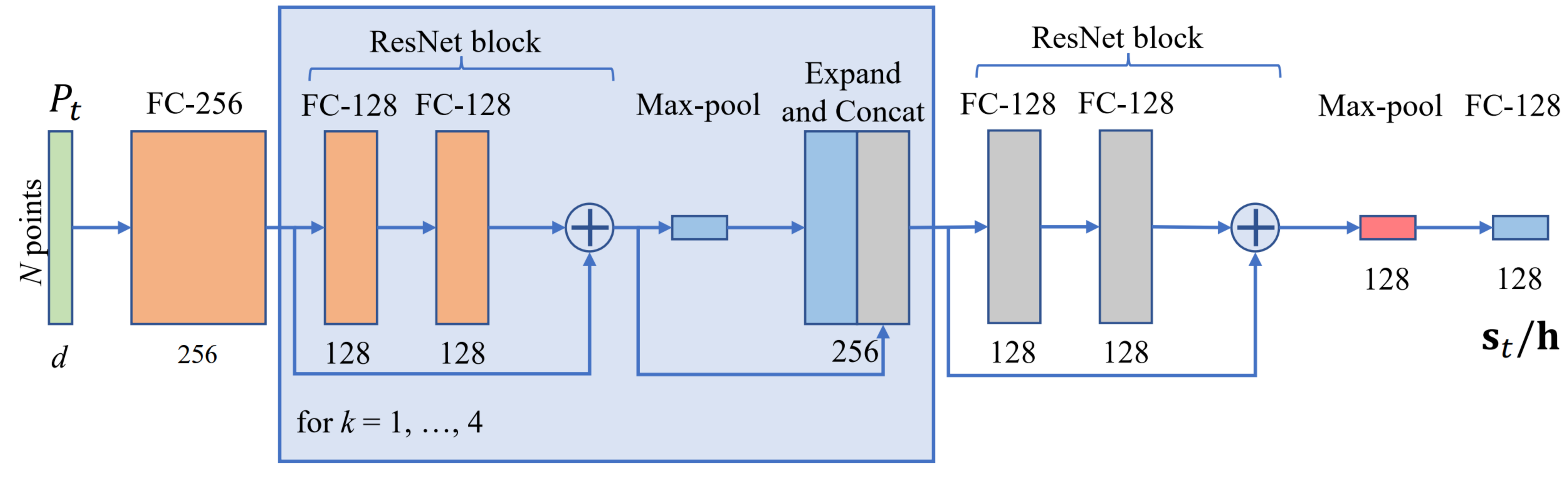
OUR RFNET-4D



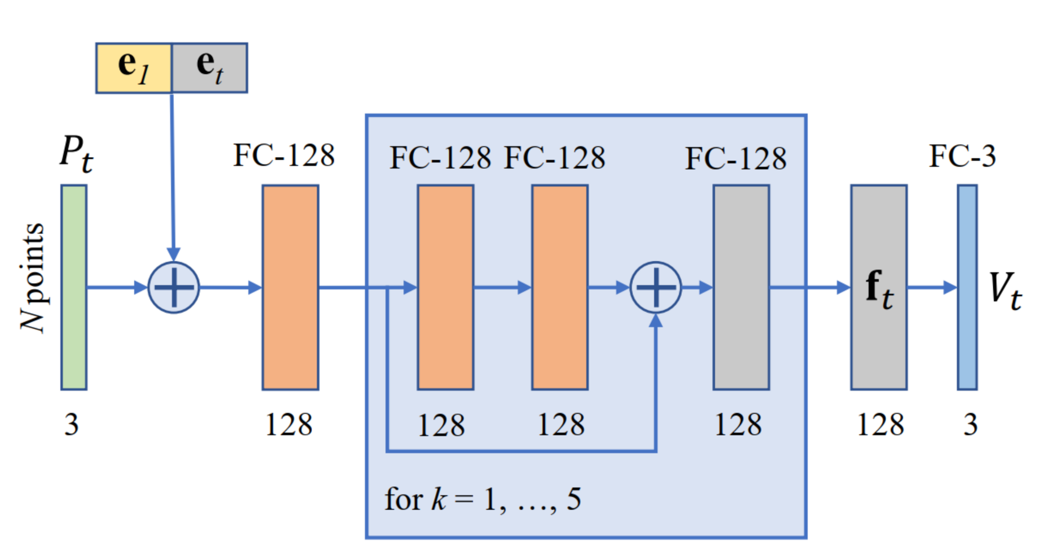
Our RFNet-4D is trained by jointly performing two optimisation processes: unsupervision for flow estimation and supervision for object reconstruction.

$$\mathcal{L}_{reconstruction} = \sum_{t} \sum_{\mathbf{p} \in P_t} \mathcal{L}_{\text{BCE}}\left(O_i(\mathbf{p}), O_i^{gt}(\mathbf{p})\right) \qquad \mathcal{L}_{flow} = \sum_{t} \max\left\{ \frac{1}{|P_t|} \sum_{\mathbf{p} \in P_t + V_t} \min_{\mathbf{p}' \in P_{t+1}} \|\mathbf{p} - \mathbf{p}'\|_2, \frac{1}{|P_{t+1}|} \sum_{\mathbf{p}' \in P_{t+1}} \min_{\mathbf{p} \in P_t + V_t} \|\mathbf{p}' - \mathbf{p}\|_2 \right\}$$

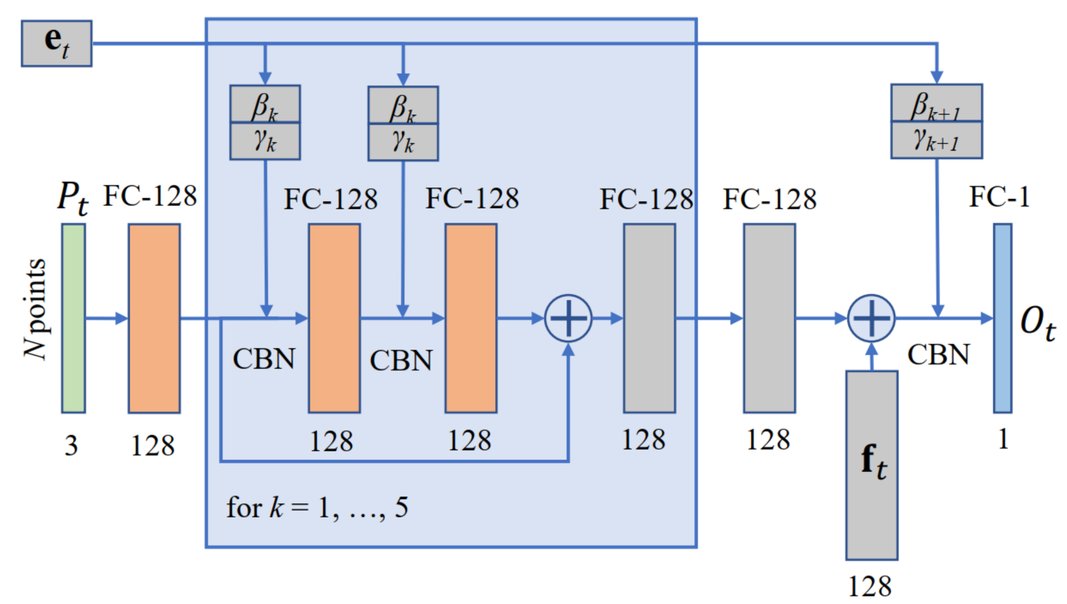
$$\mathcal{L} = \mathcal{L}_{flow} + \lambda \mathcal{L}_{reconstruction} \qquad \text{where } \lambda \text{ is a hyper-parameter.}$$



Compositional Encoder

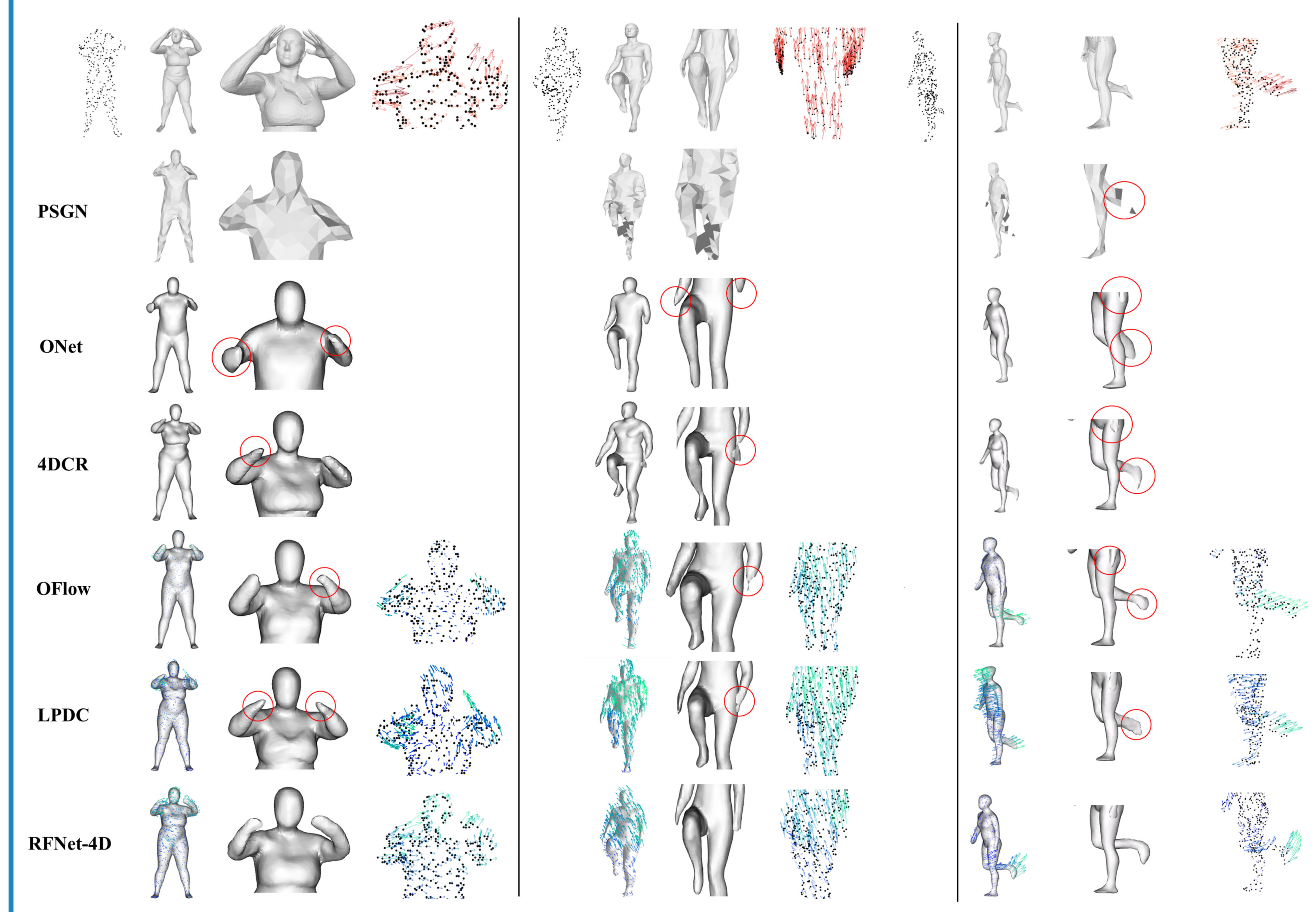


Temporal Decoder



Occupancy Decoder

RESULTS



Each following row represents corresponding reconstruction and flow estimation results. Severe errors are highlighted.

Methods	Seen Individuals Unseen Motions			Unseen Individuals Seen Motions		
	IoU↑	Chamfer↓ ($\times 10^{-3}$)	Corres.↓ ($\times 10^{-2}$)	IoU↑	Chamfer↓ ($\times 10^{-3}$)	Corres.↓ ($\times 10^{-2}$)
PSGN-4D [10]	-	0.6189	1.1083	-	0.6877	1.3289
ONet-4D [27]	0.7712	0.5921	-	0.6827	0.7007	-
OFlow [31]	0.8172	0.1773	0.8699	0.7361	0.2741	1.0842
LPDC [38]	0.8511	0.1526	**0.7803**	0.7619	0.2188	0.9872
4DCR [15]	0.8171	0.1667	-	0.6973	0.2220	-
RFNet-4D	**0.8547**	**0.1504**	0.8831	**0.8157**	**0.1594**	**0.9155**

Variant	IoU ↑	Chamfer ($\times 10^{-3}$) ↓	Corr. ($\times 10^{-2}$) ↓
RFNet-4D (only temporal flows)	-	-	1.5519
RFNet-4D (only spatial points)	0.7712	0.5921	-
RFNet-4D (only FW motion)	0.4988	2.4887	3.5868
RFNet-4D (SWD loss)	0.4305	4.4621	4.0711
RFNet-4D (HD loss)	0.7953	0.2103	1.3017
RFNet-4D (supervised)	**0.8656**	**0.0927**	**0.8125**
RFNet-4D (unsupervised)	0.8547	0.1504	0.8831

Method	Memory	Training (sec/iter)	Inference (sec/seq)
OFlow [31]	3.96GB	4.65s	0.95s
LPDC [38]	11.90GB	2.09s	0.44s
RFNet-4D	14.20GB	1.33s	0.24s

CONCLUSION

Summary:

- We presented 4D reconstruction method (RFNet-4D) that simultaneously performs flow estimation and reconstruction across time.
- Experimental results showed RFNet-4D's advance in comparison to the state-of-the-art, as well as the effectiveness of the joint learning approach.

Future works:

- Improve accuracy of unsupervised flow estimation.
- Extend experiment on different types of objects and data input (RGB, LiDAR, so on).
- Extend the pipeline to fully unsupervised for both reconstruction and flow estimation

ACKNOWLEDGEMENT

This paper was partially supported by an internal grant from HKUST (R9429)

Visit Us