JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

香港中文大學
The Chinese University of Hong Kong

H2ONet: Hand-Occlusion-and-Orientation-aware Network for Real-time 3D Hand Mesh Reconstruction

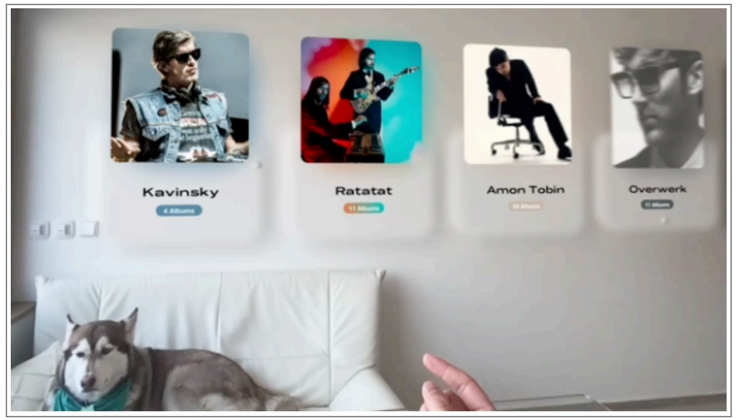
Hao Xu, Tianyu Wang, Xiao Tang, and Chi-Wing Fu

The Chinese University of Hong Kong

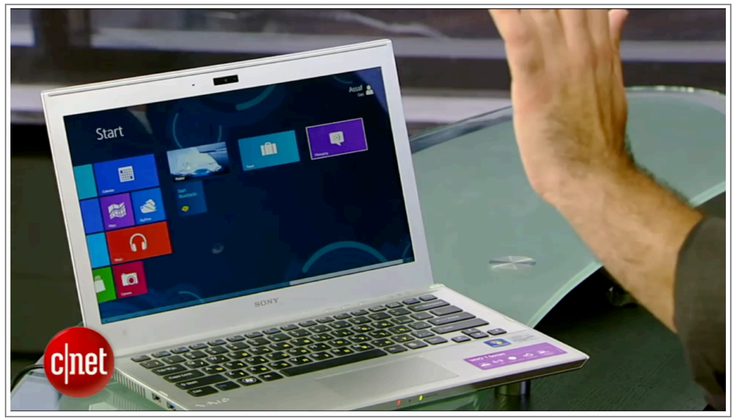
Paper tag: THU-AM-054

3D Hand Mesh Reconstruction

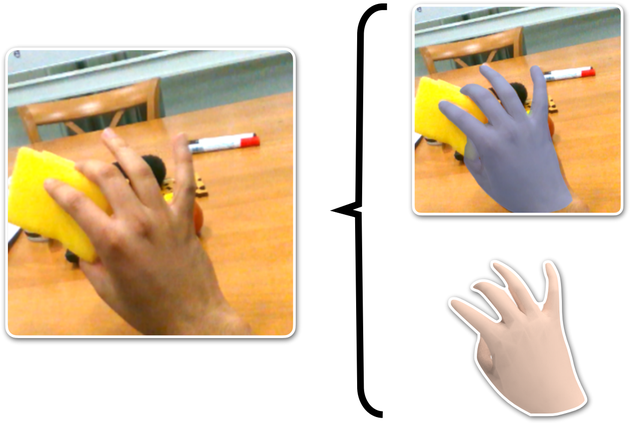
AR/VR



Behavior understanding

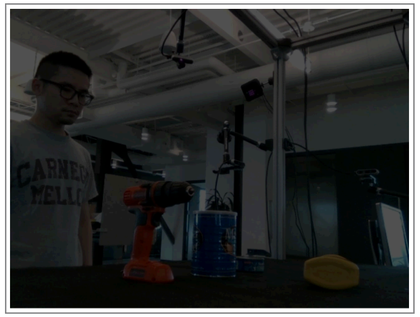


Human-computer interactions



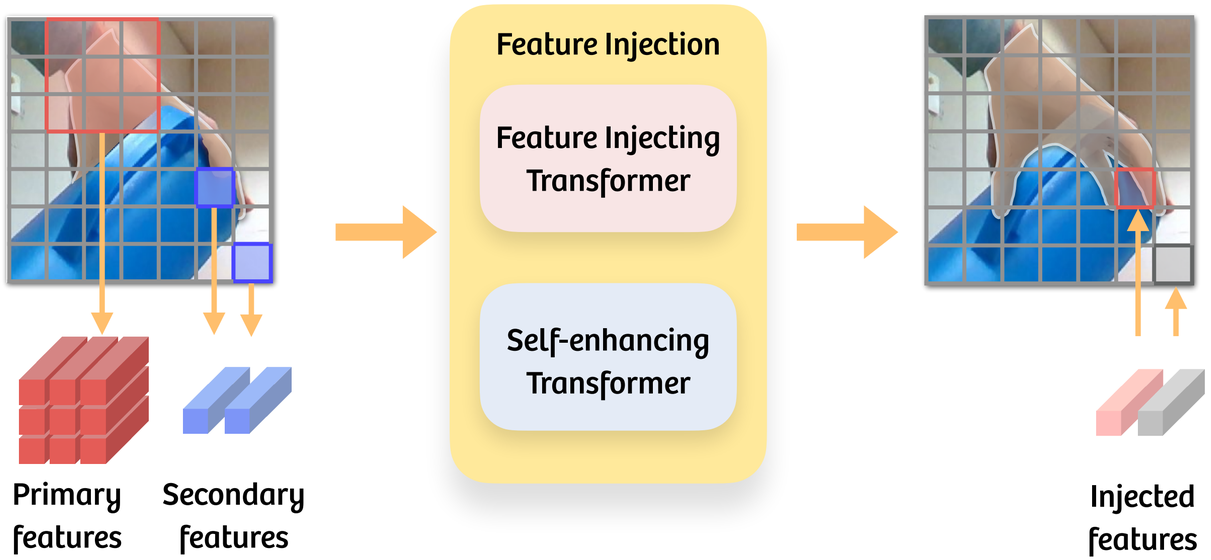
Accurate

Robust



Fast

Related Work: HandOccNet[1]



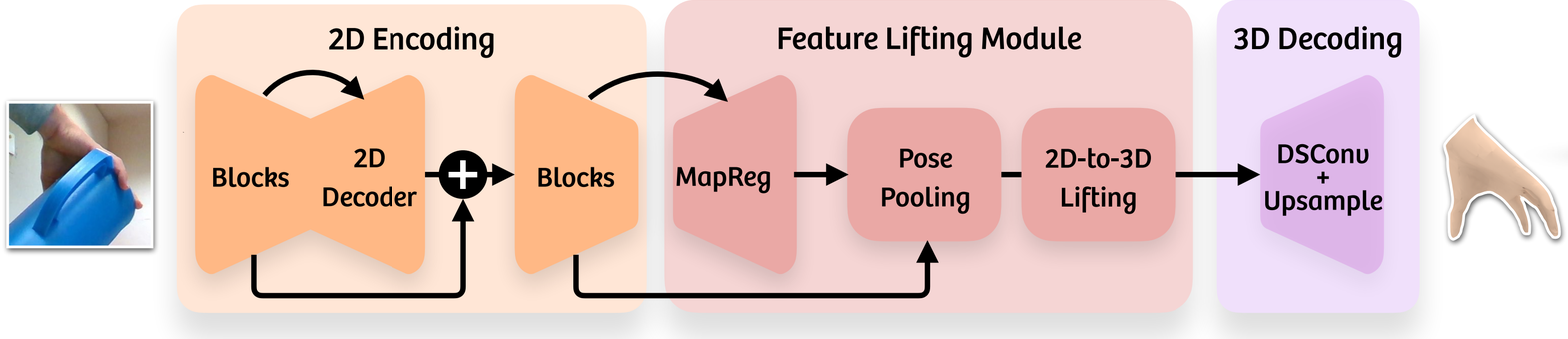
Occlusion formulation is implicit.

Predict occlusion probabilities as guidance.

Ill-posed issue still exists when occlusion is severe.

Multi-frame input can provide extra information.

Related Work: MobRecon[2]



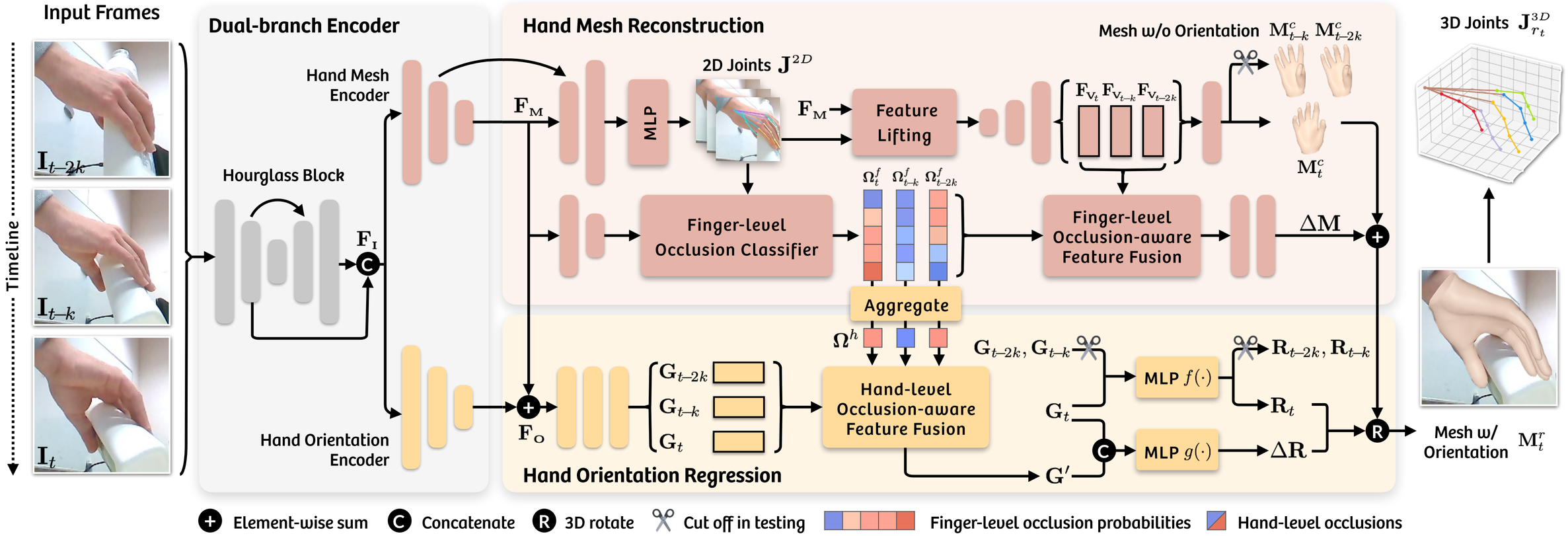
Run fast but no specific design for self and object caused occlusions.

Add modules and strategies to deal with occlusion while preserving fast speed.

[1] HandOccNet: Occlusion-Robust 3D Hand Mesh Estimation Network, CVPR, 2022.

[2] MobRecon: Mobile-Friendly Hand Mesh Reconstruction from Monocular Image, CVPR, 2022.

H2ONet: Hand-Occlusion-and-Orientation-aware Network



Our framework includes three stages:

- The dual-branch encoder extracts general and task-specific features;
- The hand mesh reconstruction module focuses on constructing hand meshes at canonical poses; and
- The hand orientation regression module predicts the global hand orientation using the hand-level visibility.

Experimental Results: Quantitative Comparison

Table 1. Results on the Dex-YCB dataset.

Methods	PA-J-PE	PA-J-AUC	PA-V-PE	PA-V-AUC	PA-F@5	PA-F@15
METRO	7.0	-	-	-	-	-
Spurr et al.	6.8	86.4	-	-	-	-
Liu et al.	6.6	-	-	-	-	-
HandOccNet	5.8	88.4	5.5	89.0	78.0	99.0
MobRecon	6.4	87.3	5.6	88.9	78.5	98.8
Our H2ONet	5.7	88.9	5.5	89.1	80.1	99.0
Our H2ONet	**5.3**	**89.4**	**5.2**	**89.6**	**80.5**	**99.3**

Methods	J-PE	J-AUC	V-PE	V-AUC	F@5	F@15
METRO	15.2	-	-	-	-	-
Spurr et al.	17.3	69.8	-	-	-	-
Liu et al.	15.3	-	-	-	-	-
HandOccNet	14.0	**74.8**	13.1	**76.6**	51.5	**92.4**
MobRecon	14.2	73.7	13.1	76.1	50.8	92.1
Our H2ONet	14.0	74.6	13.0	76.2	51.3	92.1
Our H2ONet	**13.7**	**74.8**	**12.7**	**76.6**	**52.1**	92.3

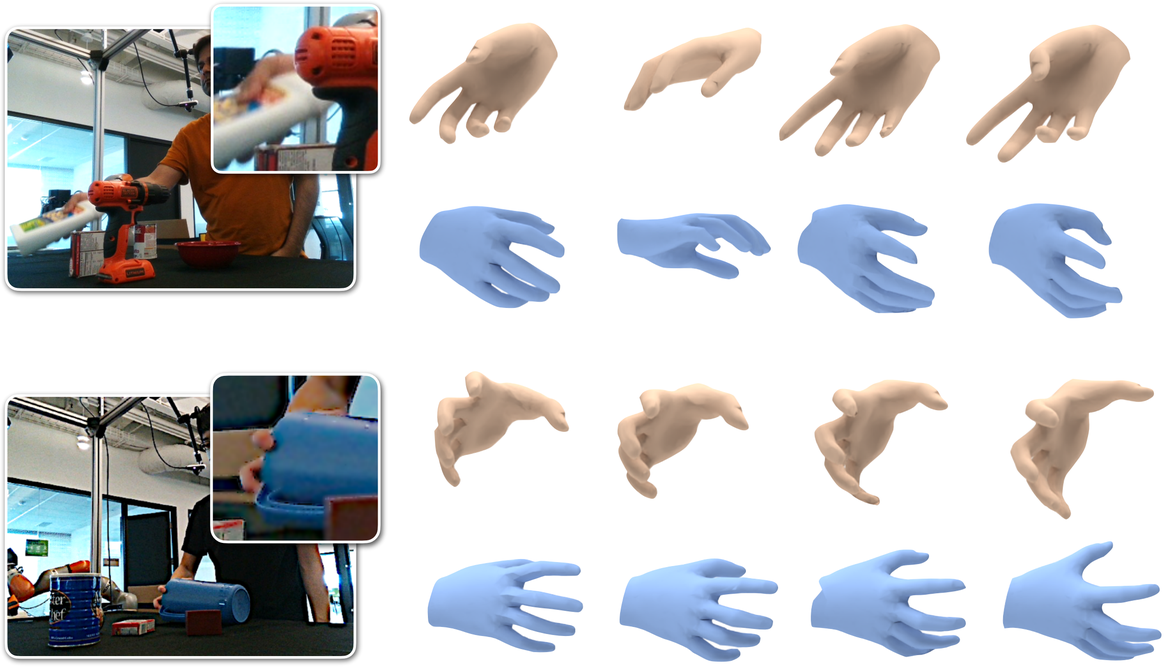
Table 2. Results on the HO3D-v2 dataset (after PA).

	Methods	J-PE	J-AUC	V-PE	V-AUC	F@5	F@15
Single-frame	Pose2Mesh	12.5	-	12.7	-	44.1	90.9
	I2L-MeshNet	11.2	-	13.9	-	40.9	93.2
	ObMan	11.1	-	11.0	77.8	46.0	93.0
	HO3D	10.7	78.8	10.6	79.0	50.6	94.2
	METRO	10.4	-	11.1	-	48.4	94.6
	Liu et al.	10.2	79.7	9.8	80.4	52.9	95.0
	I2UV-HandNet	9.9	80.4	10.1	79.9	50.0	94.3
	Tse et al.	-	-	10.9	-	48.5	94.3
	HandOccNet	9.1	81.9	9.0	81.9	56.1	96.2
	MobRecon	9.2	-	9.4	-	53.8	95.7
	MobRecon	9.4	81.3	9.5	81.0	53.3	95.5
	Our H2ONet	9.0	82.0	9.0	81.9	55.4	96.0
Multi-frame	Hasson et al.	11.4	77.3	11.4	77.3	42.8	93.2
	Hasson et al.	-	-	14.7	-	39.0	88.0
	Liu et al.	9.8	-	9.4	81.2	53.0	95.7
	Our H2ONet	**8.5**	**82.9**	**8.6**	**82.8**	**57.0**	**96.6**

Note: **bold** and underlined denote 1st and 2nd performance, respectively.

Experimental Results: Qualitative Comparison

Dex-YCB



HO3D-v2

HandOccNet MobRecon Ours GT

HandOccNet MobRecon Ours

Introduction

Introduction

Research problems

Single-hand reconstruction

Bimanual-hands reconstruction

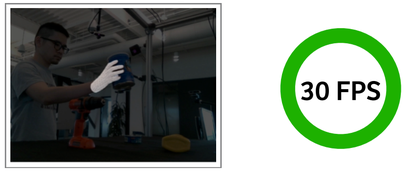
Hand-object reconstruction

Challenges

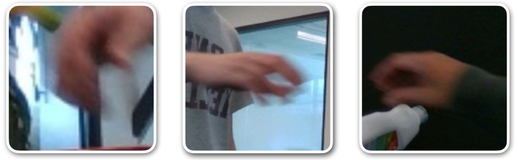
Accuracy

Occlusions

Real-time speed



Motion blur



Introduction

Research problems

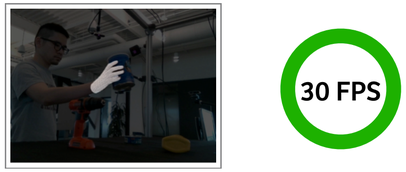
- Single-hand reconstruction

Challenges

- Accuracy

- Occlusions

- Real-time speed



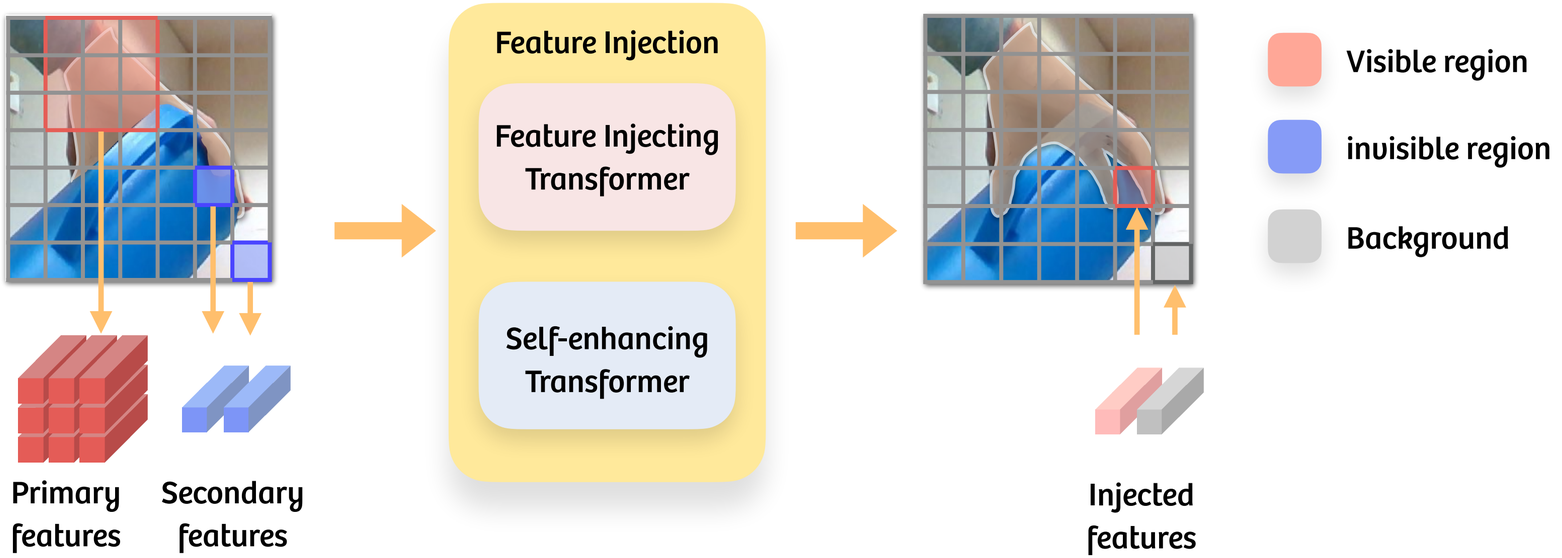
Related Work

HandOccNet[1]

[1] HandOccNet: Occlusion-Robust 3D Hand Mesh Estimation Network, CVPR, 2022.

Related Work

HandOccNet[1]



Occlusion formulation is implicit.

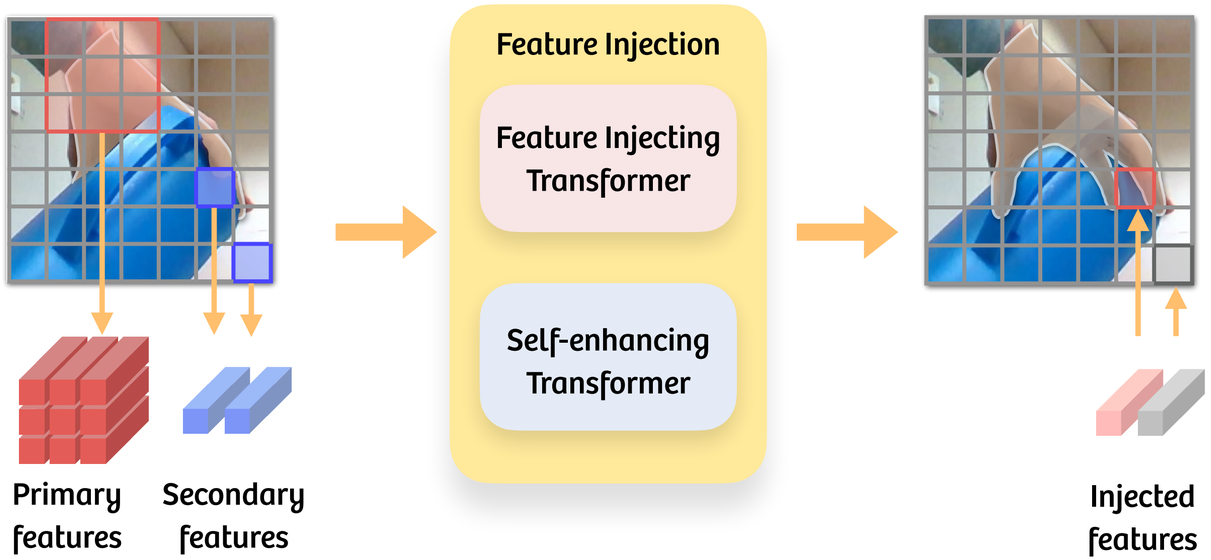
Predict occlusion probabilities as guidance.

Ill-posed issue still exists when occlusion is severe.

Multi-frame input can provide extra information.

[1] HandOccNet: Occlusion-Robust 3D Hand Mesh Estimation Network, CVPR, 2022.

Related Work: HandOccNet[1]



Occlusion formulation is implicit.

Predict occlusion probabilities as guidance.

Ill-posed issue still exists when occlusion is severe.

Multi-frame input can provide extra information.

[1] HandOccNet: Occlusion-Robust 3D Hand Mesh Estimation Network, CVPR, 2022.

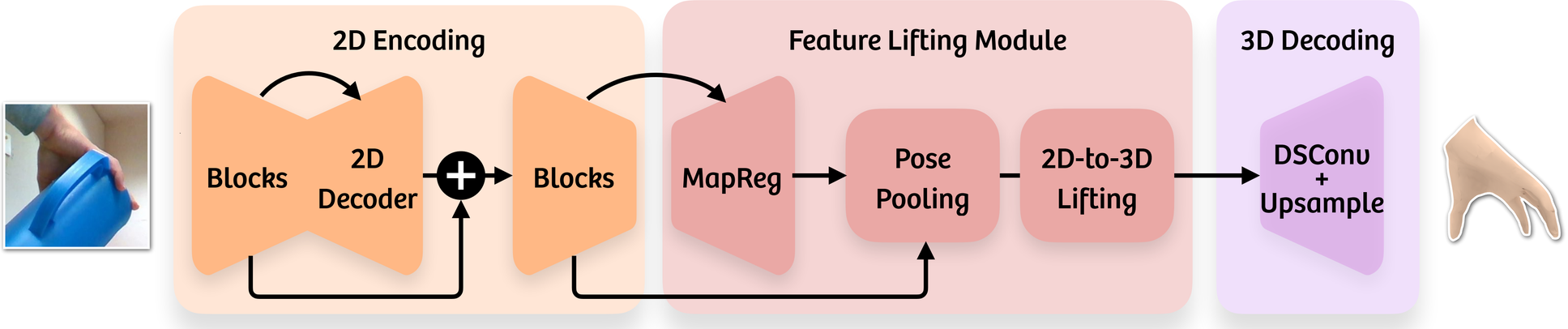
Related Work

MobRecon[2]

[2] MobRecon: Mobile-Friendly Hand Mesh Reconstruction from Monocular Image, CVPR, 2022.

Related Work

MobRecon[2]

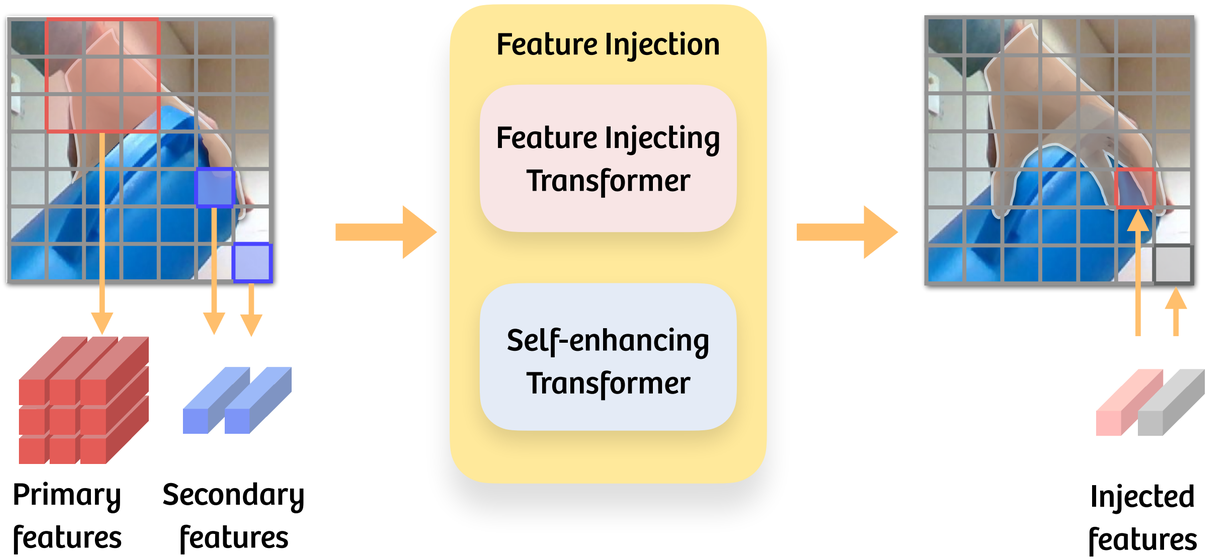


Run fast but no specific design for self and object caused occlusions.

Add modules and strategies to deal with occlusion while preserving fast speed.

[2] MobRecon: Mobile-Friendly Hand Mesh Reconstruction from Monocular Image, CVPR, 2022.

Related Work: HandOccNet[1]



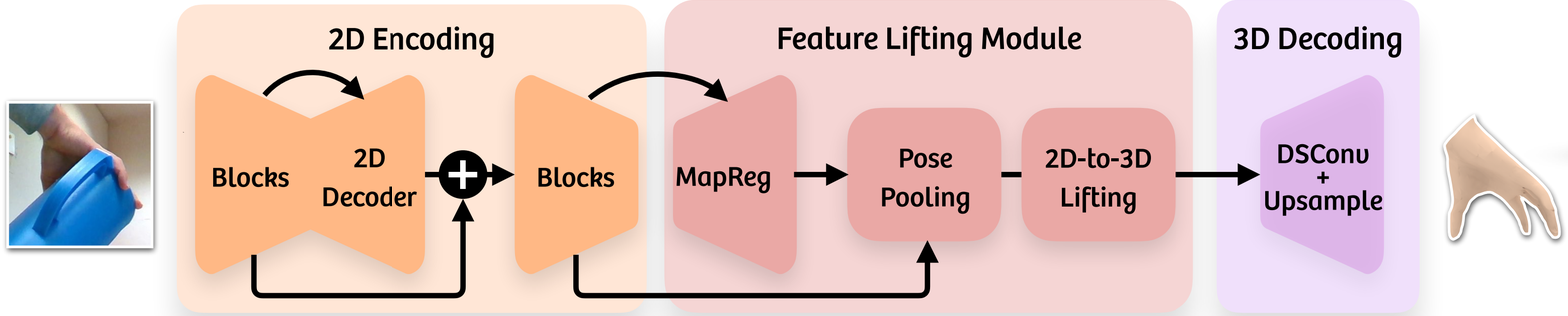
Occlusion formulation is implicit.

Predict occlusion probabilities as guidance.

Ill-posed issue still exists when occlusion is severe.

Multi-frame input can provide extra information.

Related Work: MobRecon[2]



Run fast but no specific design for self and object caused occlusions.

Add modules and strategies to deal with occlusion while preserving fast speed.

[1] HandOccNet: Occlusion-Robust 3D Hand Mesh Estimation Network, CVPR, 2022.

[2] MobRecon: Mobile-Friendly Hand Mesh Reconstruction from Monocular Image, CVPR, 2022.

Motivation

Motivation

58 dim

MANO coefficients

10 dim

MANO shape

48 dim

MANO pose

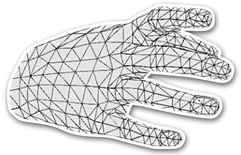
3 dim

Global orientation

45 dim

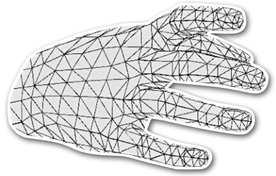
Fingers' orientations

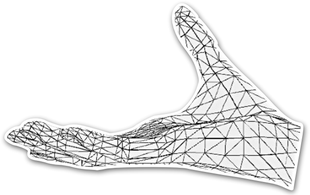
Hand shape and orientation are disentangled.



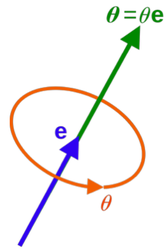
3D Vertices

3D vertices regression enables better alignment with input.





Hand shape at canonical pose



Hand orientation

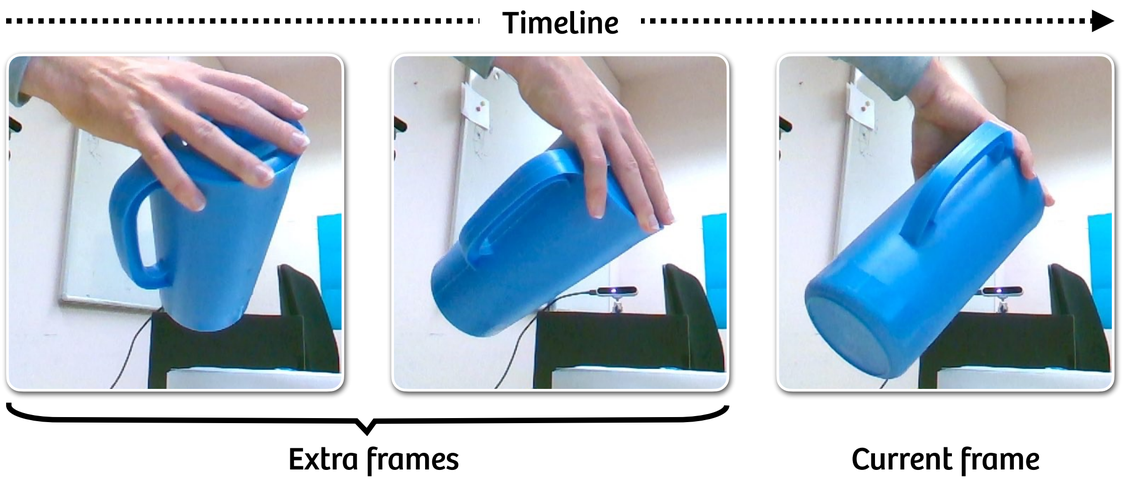
Disentangle hand shape and orientation when regressing 3D vertices.

Motivation

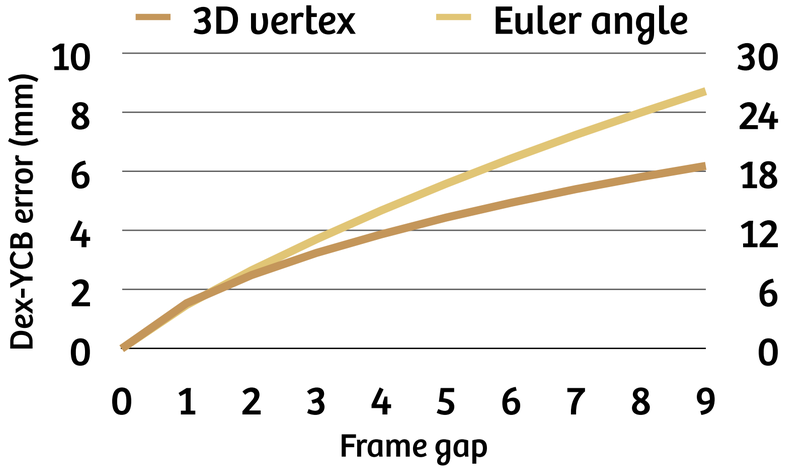
Single-frame input

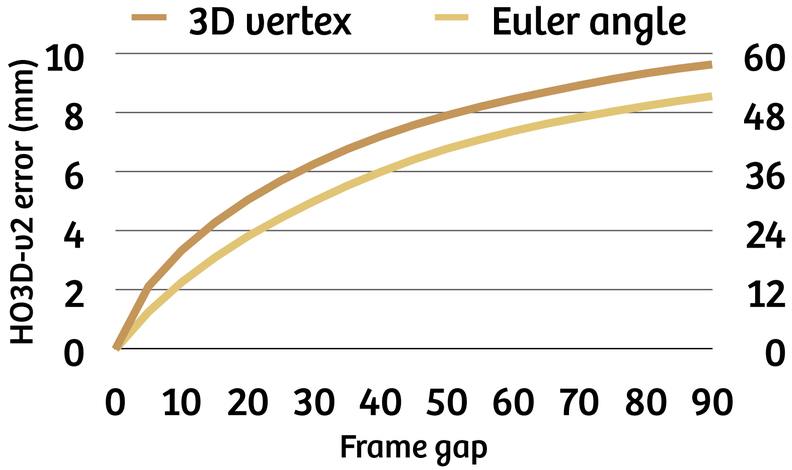
Segmentation Mask

Ill-posed issue occurs when occlusion is dominated in the input.



Multi-frame input helps to alleviate this issue.





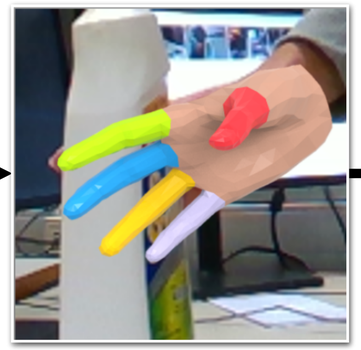
Rigid-motion assumption:

- Too near, useful information is limited;
- Too far, assumption is broken.

Our Method

Occlusion Label Preparation

(i) Input



(ii) Rendered hand

(iii) Input hand seg.

(iv) Masked hand

(v) Projected vertices

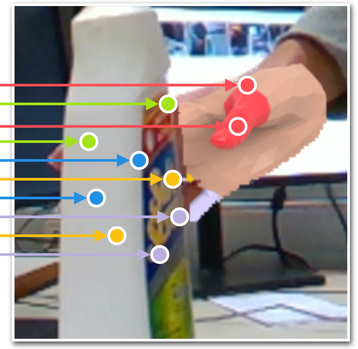
Multiply

Non-occluded

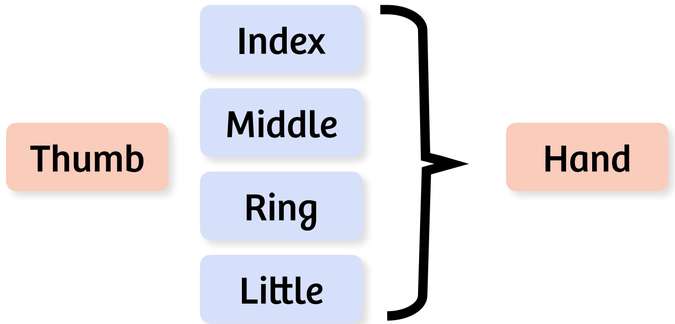
Occluded

Note: Two vertices per finger for clarity.

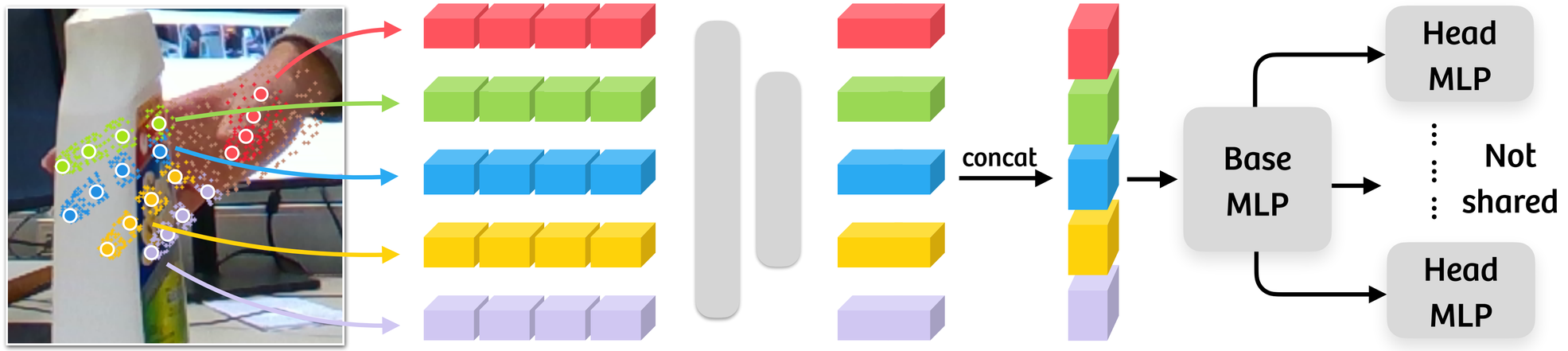
(v) Projected vertices



(iv) Masked hand



Occlusion Probabilities Prediction

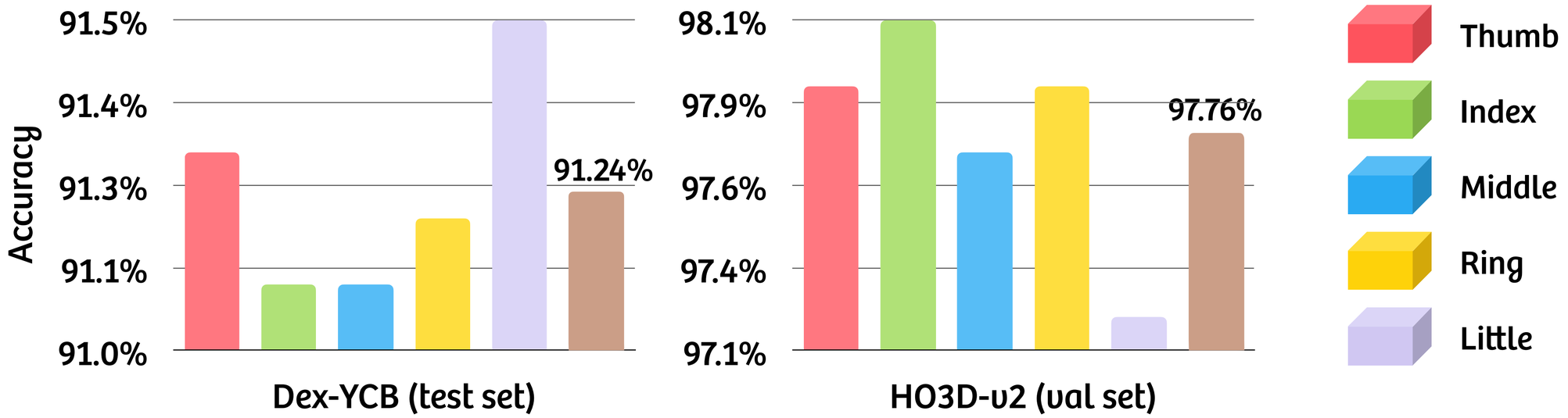


(i) Input

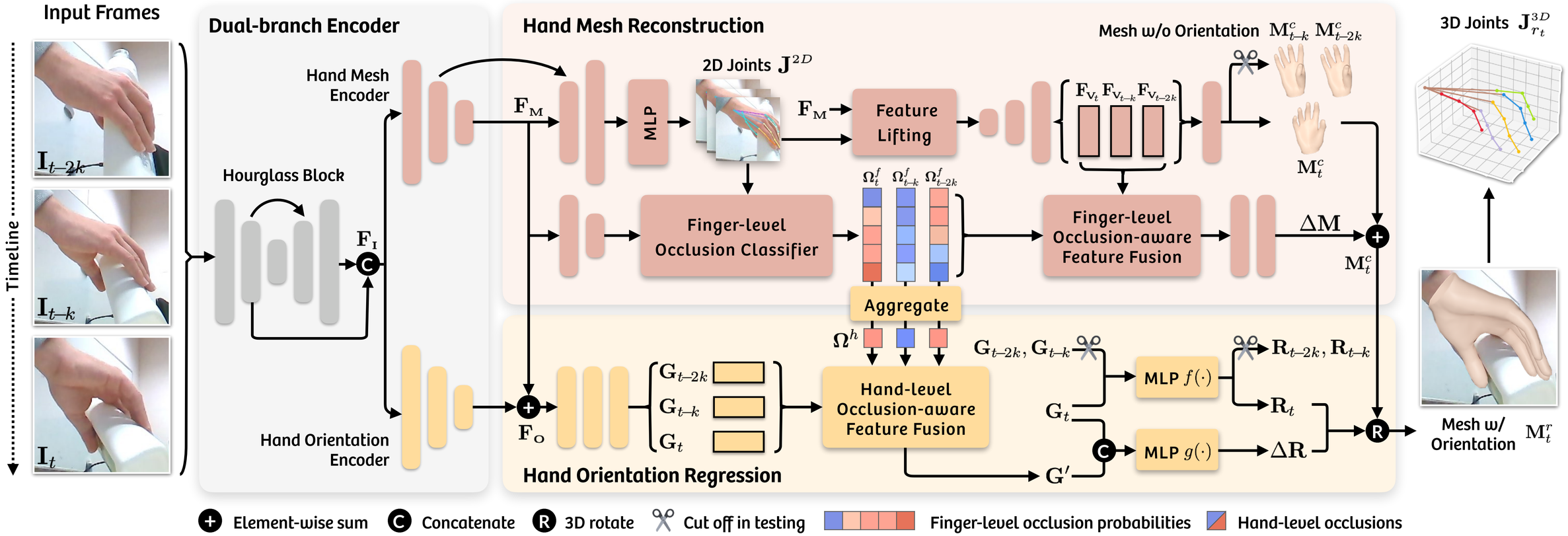
(ii) Joint-level feature

(iii) Finger-level feature

(iv) Multi-head Classifier



H2ONet: Hand-Occlusion-and-Orientation-aware Network

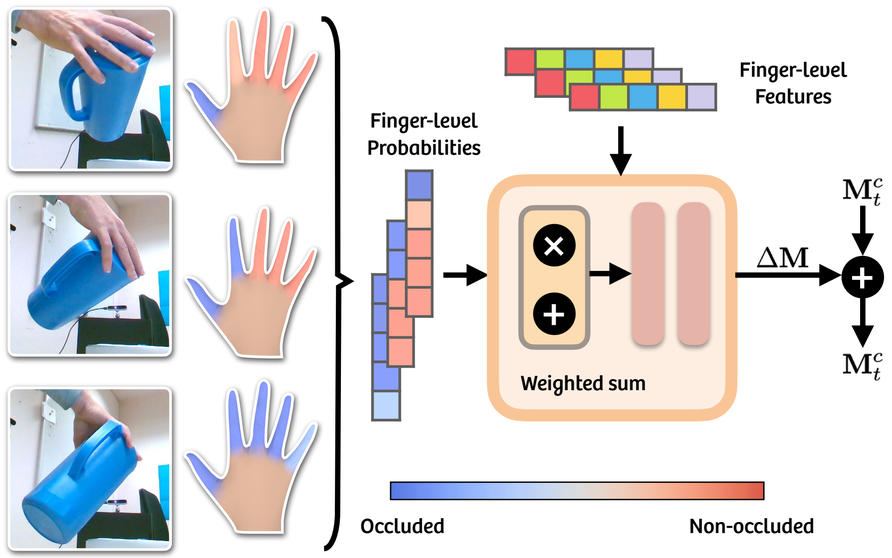


Our framework includes three stages:

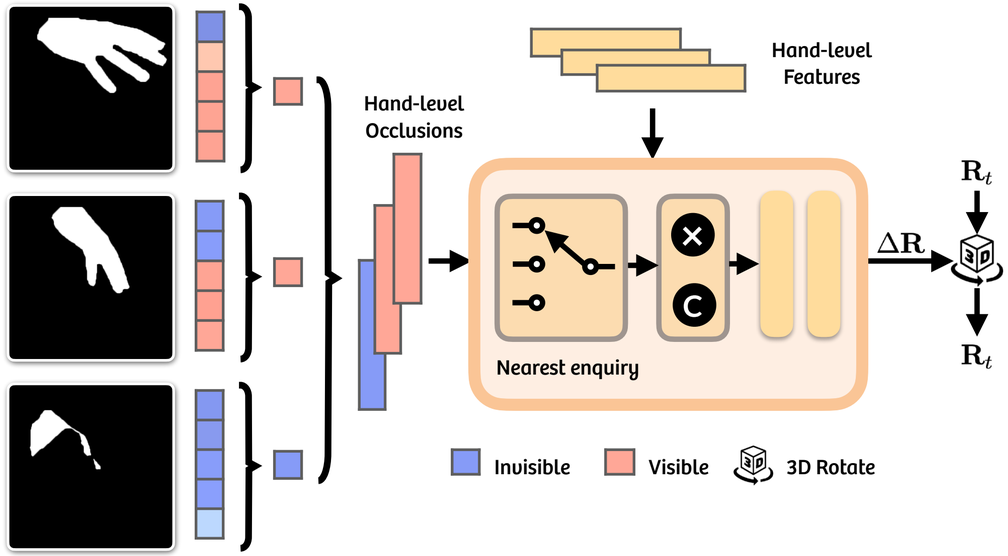
- The dual-branch encoder extracts general and task-specific features;
- The hand mesh reconstruction module focuses on constructing hand meshes at canonical poses; and
- The hand orientation regression module predicts the global hand orientation using the hand-level visibility.

Multi-frame Feature Fusion

Finger-level Feature Fusion



Hand-level Feature Fusion



Experimental Results

Experimental Results

Evaluation on the Dex-YCB dataset

Table 1. Results comparison after PA.

Methods	PA-J-PE	PA-J-AUC	PA-V-PE	PA-V-AUC	PA-F@5	PA-F@15
METRO	7.0	-	-	-	-	-
Spurr et al.	6.8	86.4	-	-	-	-
Liu et al.	6.6	-	-	-	-	-
HandOccNet	5.8	88.4	5.5	89.0	78.0	99.0
MobRecon	6.4	87.3	5.6	88.9	78.5	98.8
Our H2ONet	5.7	88.9	5.5	89.1	80.1	99.0
Our H2ONet	5.3	89.4	5.2	89.6	80.5	99.3

Table 2. Results comparison before PA.

Methods	J-PE	J-AUC	V-PE	V-AUC	F@5	F@15
METRO	15.2	-	-	-	-	-
Spurr et al.	17.3	69.8	-	-	-	-
Liu et al.	15.3	-	-	-	-	-
HandOccNet	14.0	74.8	13.1	76.6	51.5	92.4
MobRecon	14.2	73.7	13.1	76.1	50.8	92.1
Our H2ONet	14.0	74.6	13.0	76.2	51.3	92.1
Our H2ONet	13.7	74.8	12.7	76.6	52.1	92.3

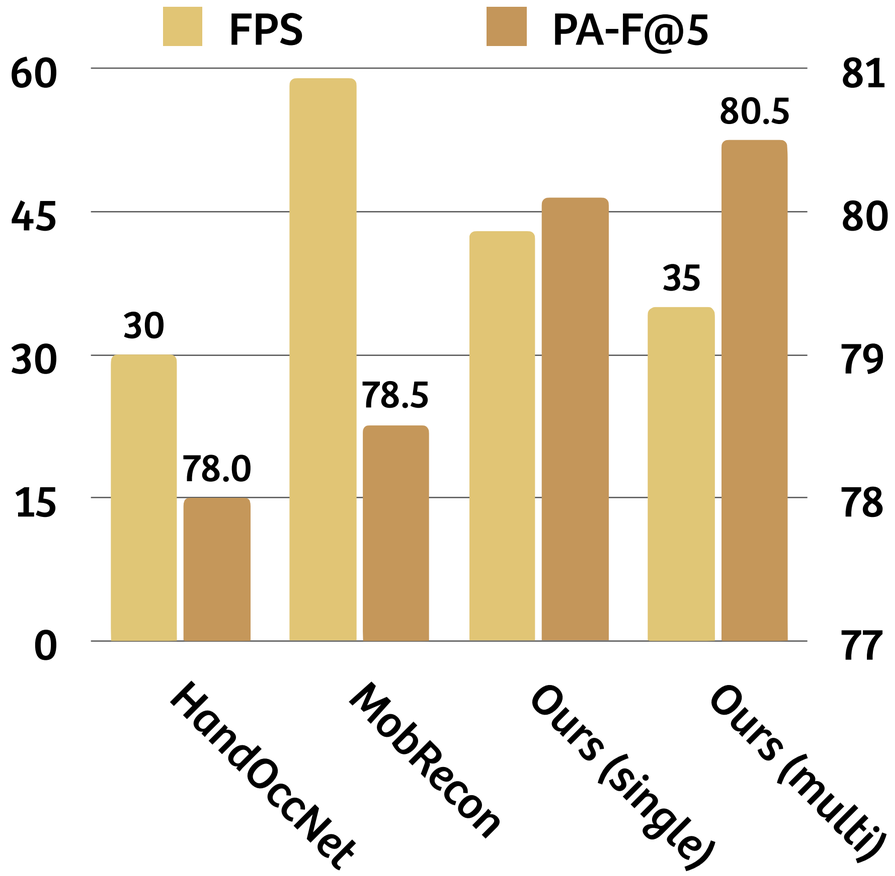
Quantitative Comparison

Evaluation on the HO3D-v2 dataset

Table 3. Results comparison after PA.

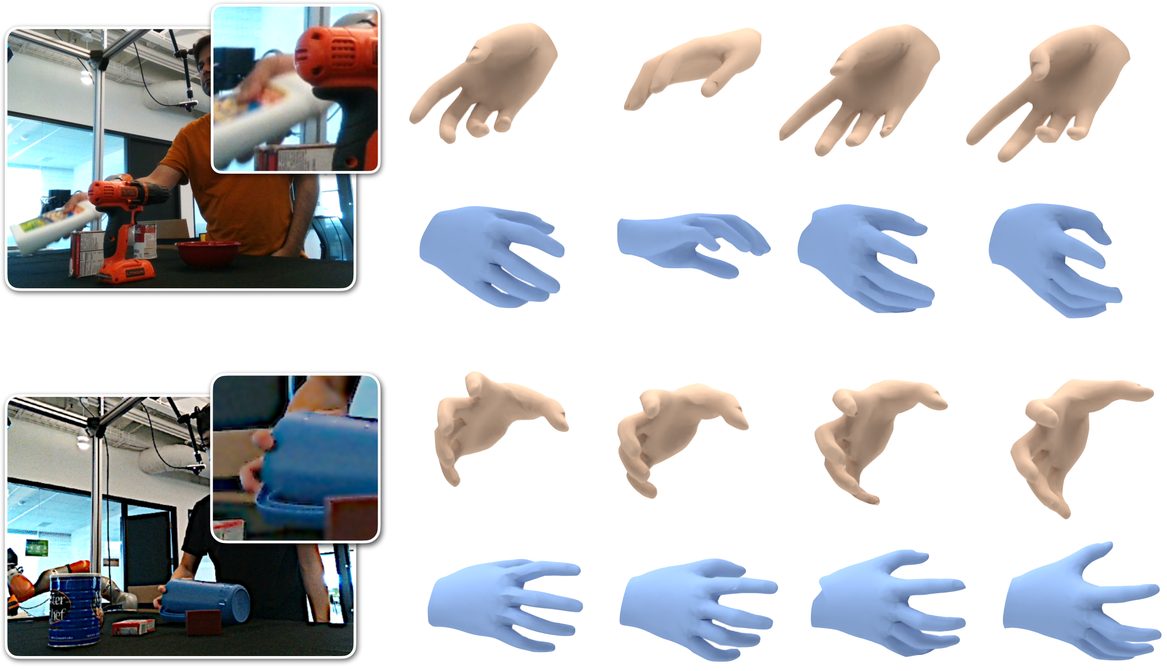
	Methods	PA-J-PE	PA-J-AUC	PA-V-PE	PA-V-AUC	PA-F@5	PA-F@15
Single-frame	Pose2Mesh	12.5	-	12.7	-	44.1	90.9
	I2L-MeshNet	11.2	-	13.9	-	40.9	93.2
	ObMan	11.1	-	11.0	77.8	46.0	93.0
	HO3D	10.7	78.8	10.6	79.0	50.6	94.2
	METRO	10.4	-	11.1	-	48.4	94.6
	Liu et al.	10.2	79.7	9.8	80.4	52.9	95.0
	I2UV-HandNet	9.9	80.4	10.1	79.9	50.0	94.3
	Tse et al.	-	-	10.9	-	48.5	94.3
	HandOccNet	9.1	81.9	9.0	81.9	56.1	96.2
	MobRecon	9.2	-	9.4	-	53.8	95.7
	MobRecon	9.4	81.3	9.5	81.0	53.3	95.5
	Our H2ONet	9.0	82.0	9.0	81.9	55.4	96.0
Multi-frame	Hasson et al.	11.4	77.3	11.4	77.3	42.8	93.2
	Hasson et al.	-	-	14.7	-	39.0	88.0
	Liu et al.	9.8	-	9.4	81.2	53.0	95.7
	Our H2ONet	8.5	82.9	8.6	82.8	57.0	96.6

Efficiency v.s. Effectiveness



Qualitative Comparison

Dex-YCB



HO3D-v2

HandOccNet MobRecon Ours GT

HandOccNet MobRecon Ours

More Results

香港中文大學
The Chinese University of Hong Kong

Thank you for watching!

Paper tag: THU-AM-054