website

WED-AM-074

BKinD-3D: Self-Supervised 3D Keypoint Discovery from Multi-View Videos

Jennifer J. Sun*[1], Lili Karashchuk*[2], Amil Dravid*[3], Serim Ryou[4], Sonia Fereidooni[2], John Tuthill[2], Aggelos Katsaggelos[3], Bingni W. Brunton[2], Georgia Gkioxari[1], Ann Kennedy[3], Yisong Yue[1], Pietro Perona[1]

[1]Caltech, [2]University of Washington, [3]Northwestern, [4]SAIT

Caltech

UNIVERSITY of WASHINGTON

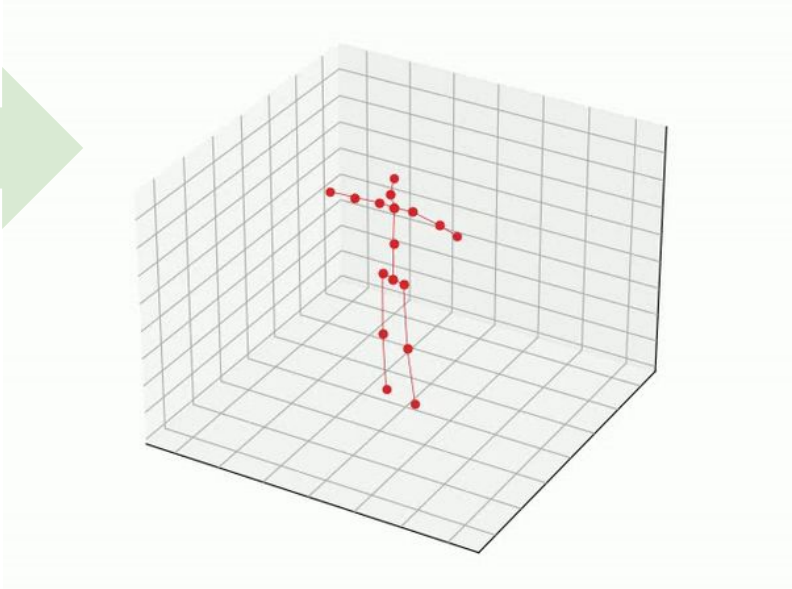
Northwestern University

SAMSUNG

Our goal: 3D keypoint discovery

No 2D or 3D annotations

BKinD-3D



Multi-view video

3D keypoints

Our goal: 3D keypoint discovery

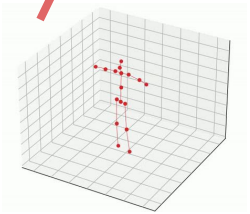
Multi-view video

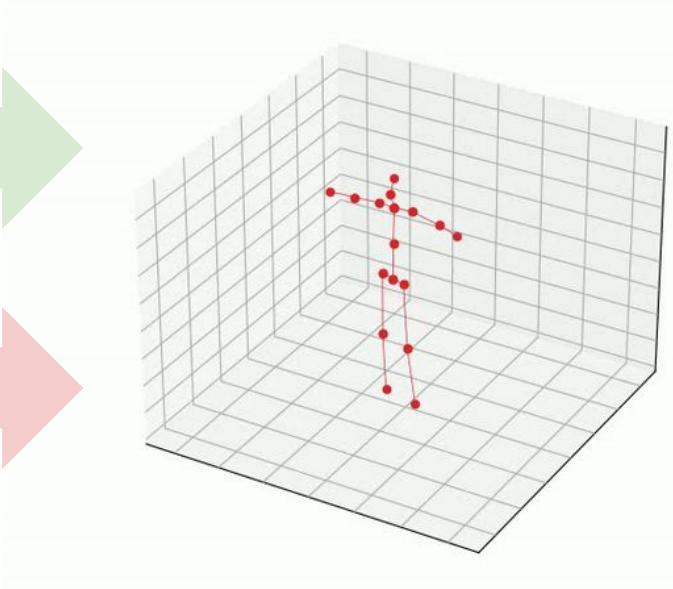
No 2D or 3D annotations

BKinD-3D

Previous methods

Require 2D or 3D annotations to train

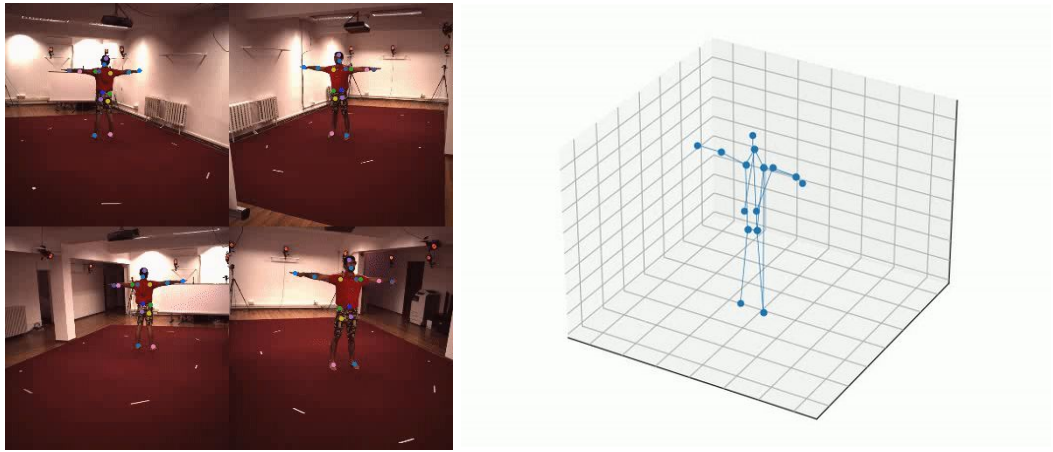




3D keypoints

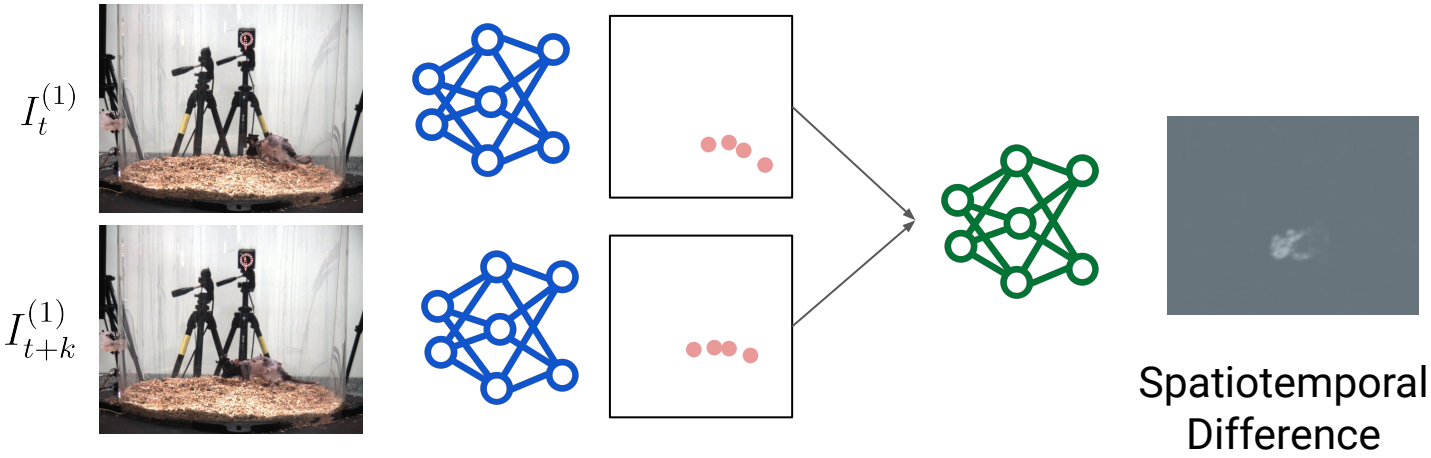
BKinD-3D for keypoint discovery

Human3.6M



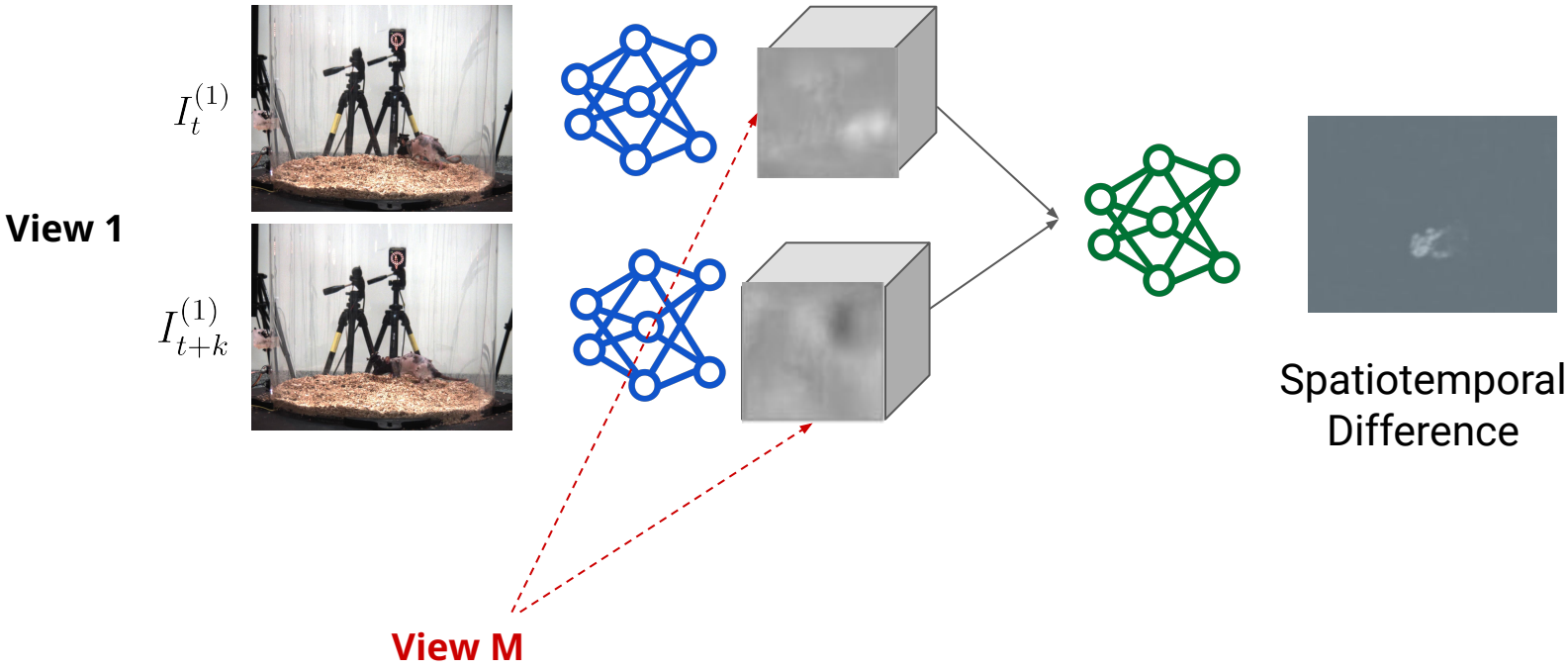
Rat7M

BKinD-3D for keypoint discovery



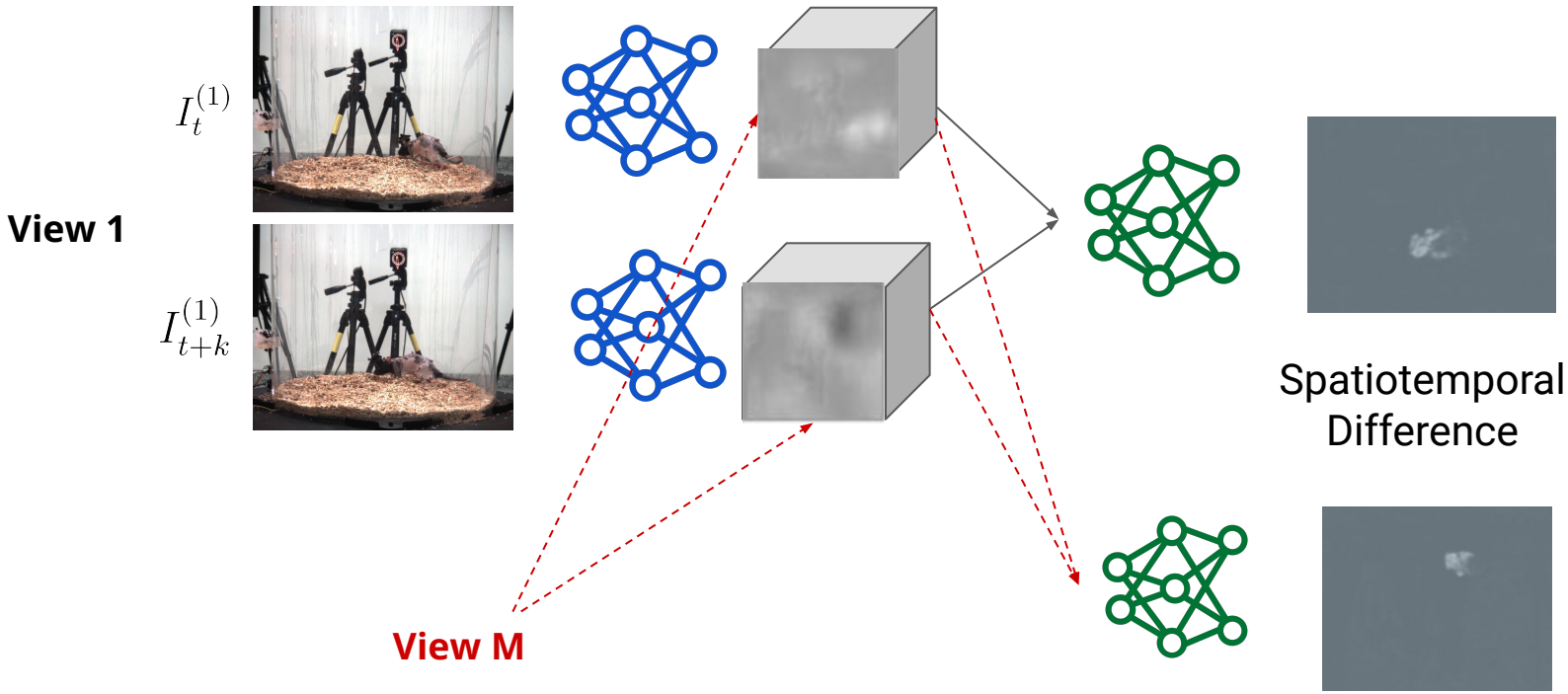
Discover keypoints that reconstructs agent movement

BKinD-3D for keypoint discovery



Discover keypoints that reconstructs agent movement
across views

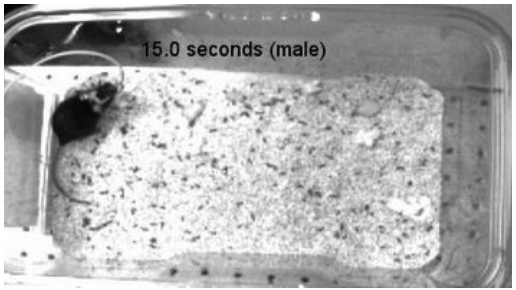
BKinD-3D for keypoint discovery



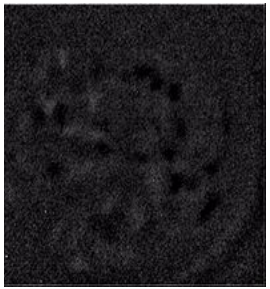
Discover keypoints that reconstructs agent movement
across views

Behavior Analysis in 3D

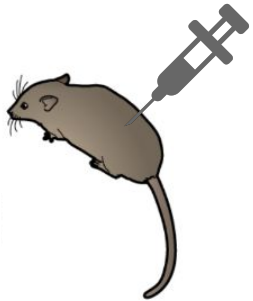
Scientific Experiments



Neural Activity

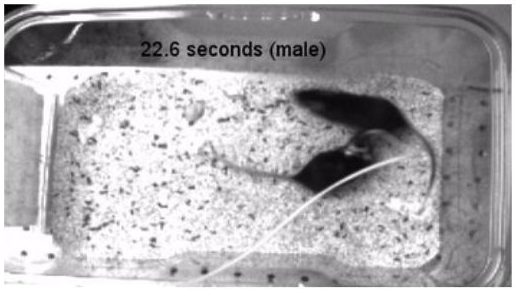


Pharmacological Evaluation



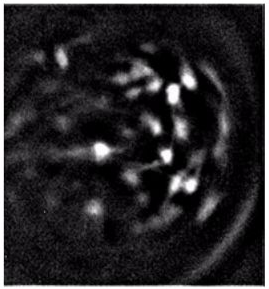
Behavior Analysis in 3D

Scientific Experiments

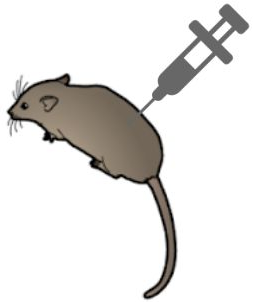


User Applications

Neural Activity



Pharmacological Evaluation



Human-Robot Interactions

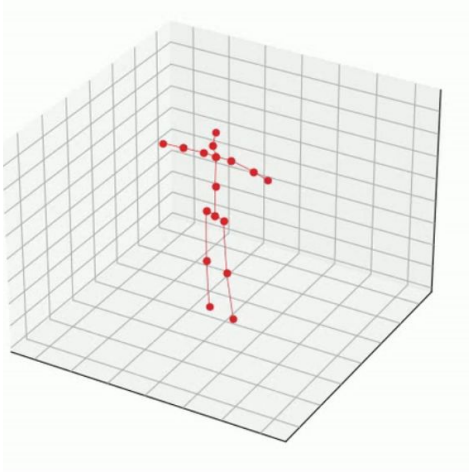
Healthcare Monitoring

Related Areas

Supervised 3D Pose estimation

3D annotations required

Visual data



3D keypoints

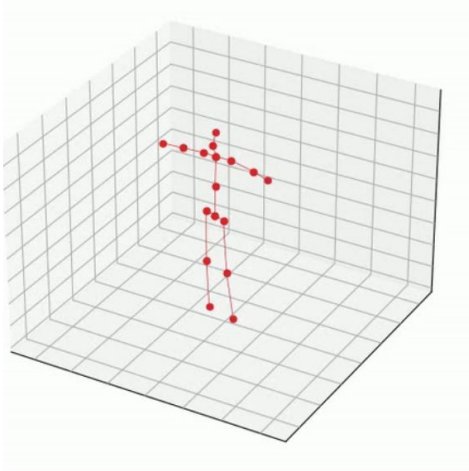
Related Areas

Visual data

Supervised 3D Pose estimation

3D from 2D (ex: lifting)

2D or 3D annotations required



3D keypoints

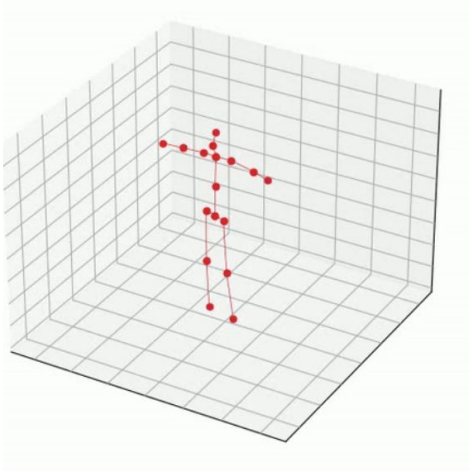
Related Areas

Visual data

Supervised 3D Pose estimation

3D from 2D (ex: lifting)

2D Keypoint Discovery



3D keypoints

Related Areas

Visual data

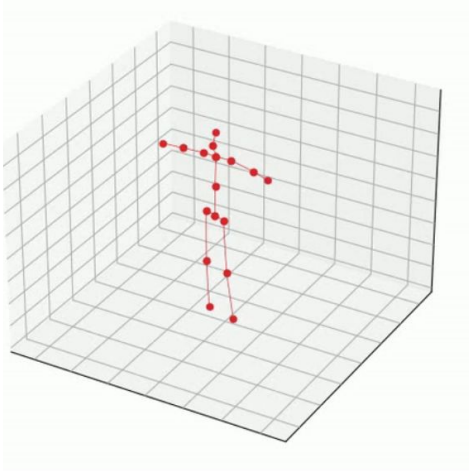
Supervised 3D Pose estimation

3D from 2D (ex: lifting)

2D Keypoint Discovery

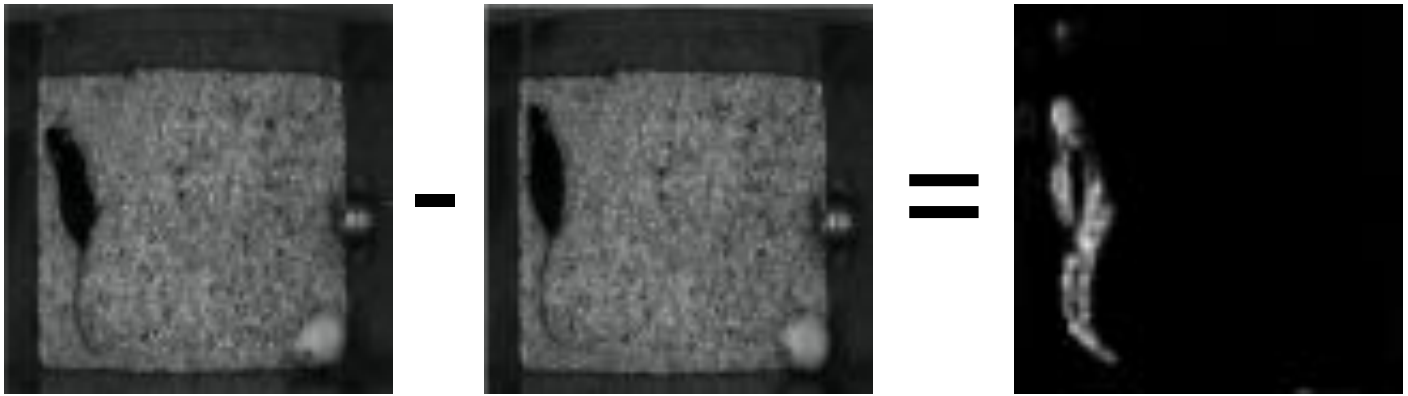
3D Keypoint Discovery (ours)

No 2D or 3D annotations required

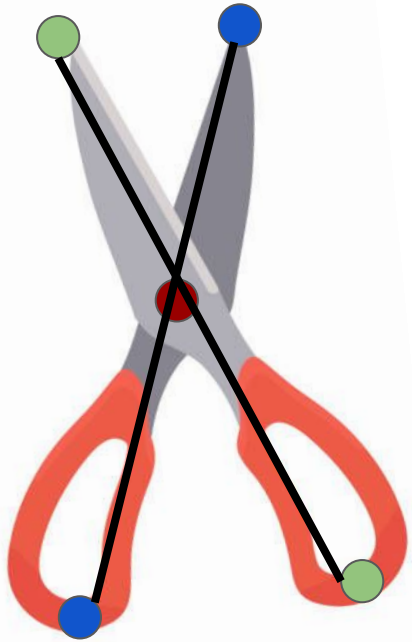


3D keypoints

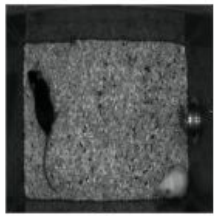
Method Intuition



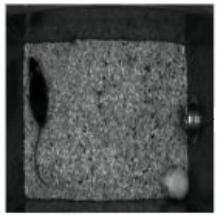
$I_t^{(1)}$ $I_{t+k}^{(1)}$ Spatiotemporal Difference



Behavioral **K**eypo**in**t **D**iscovery in **3D**
(BKinD-3D)

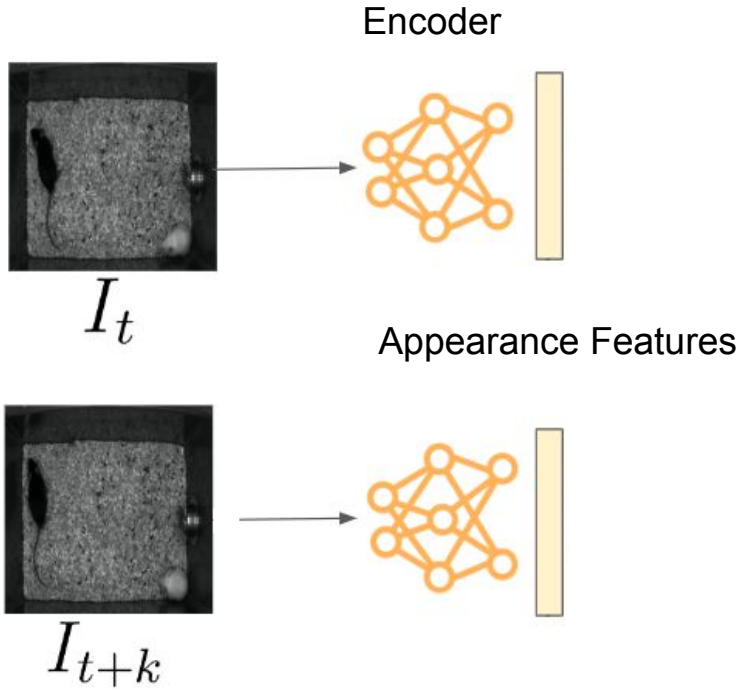


I_t

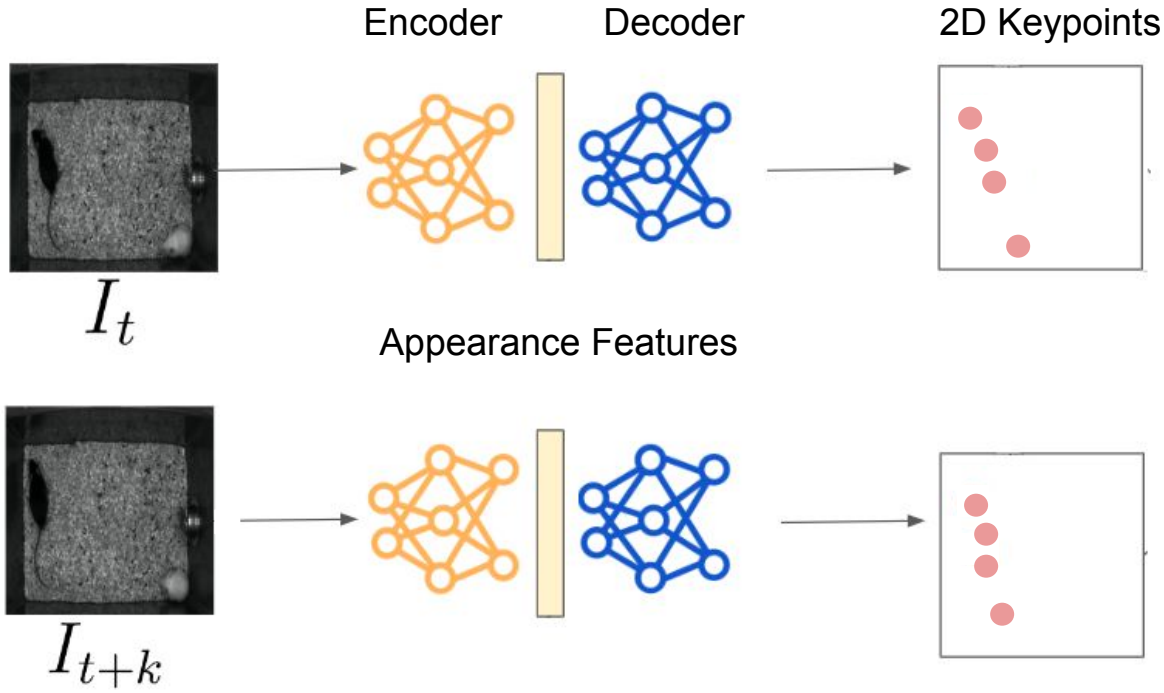


I_{t+k}

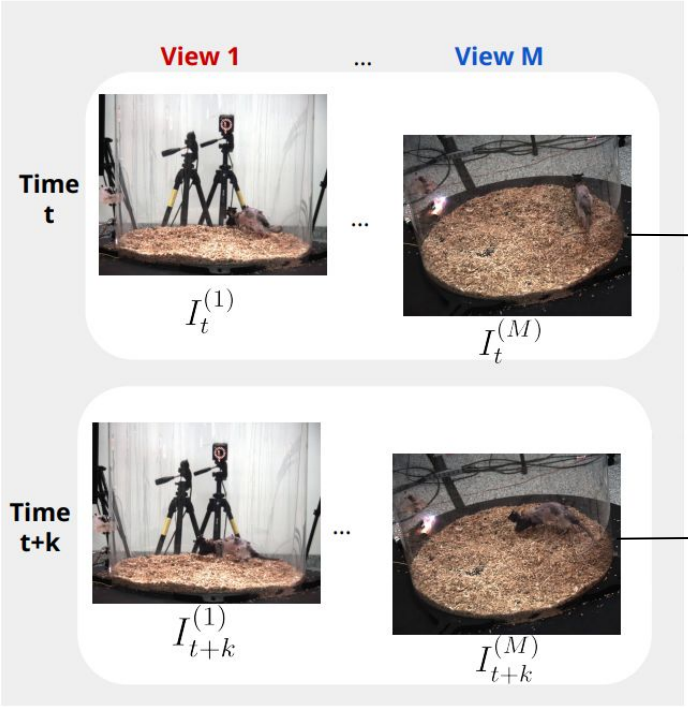
Behavioral **K**eypo**in**t **D**iscovery in **3D**
(BKinD-3D)



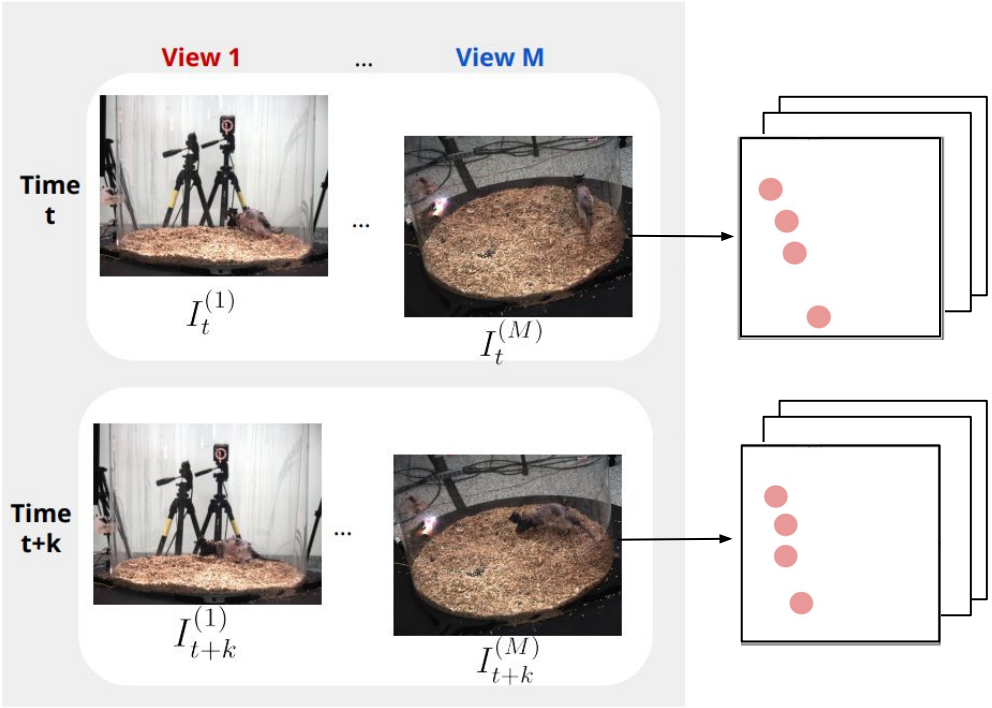
Behavioral **K**eypo**in**t **D**iscovery in **3D**
(BKinD-3D)



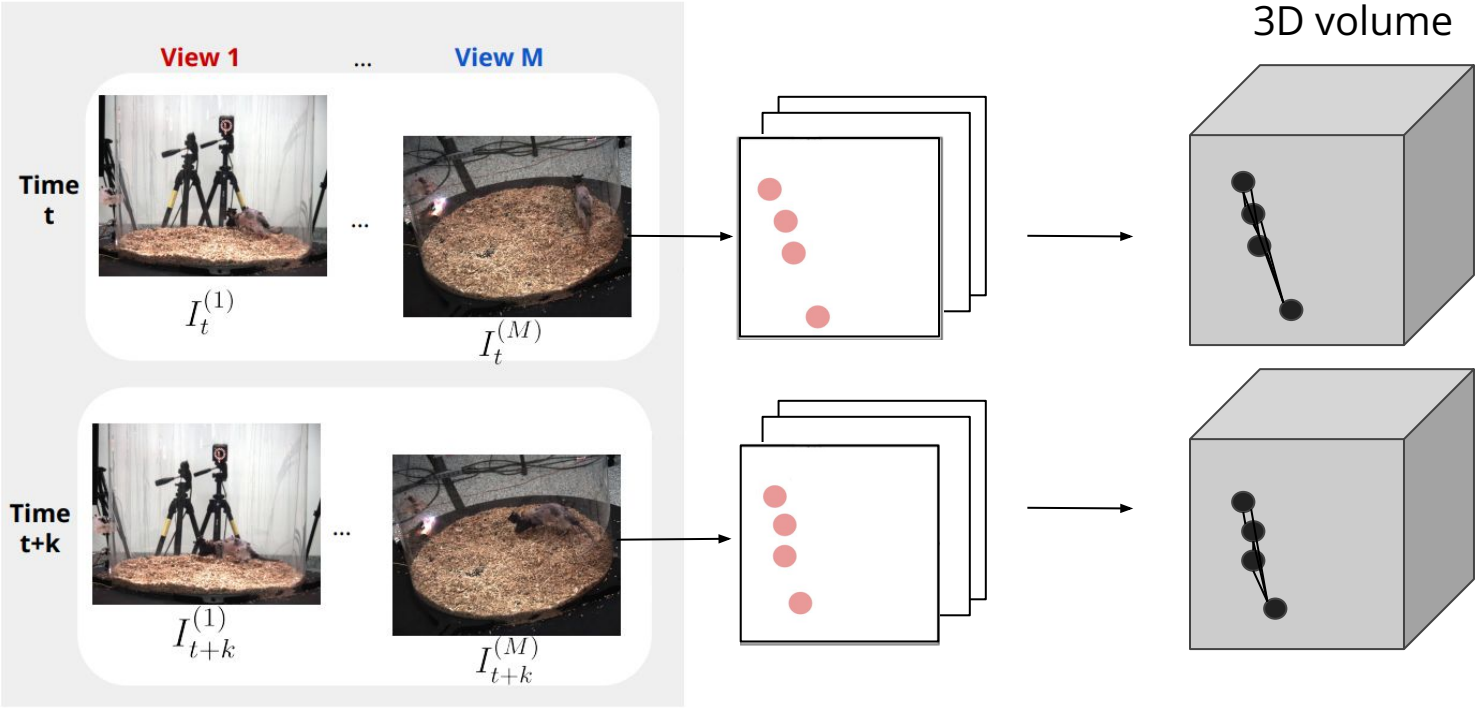
Repeat Across Views



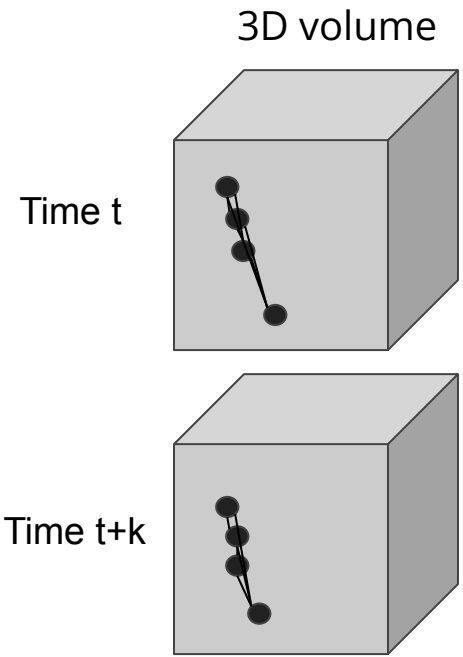
Repeat Across Views



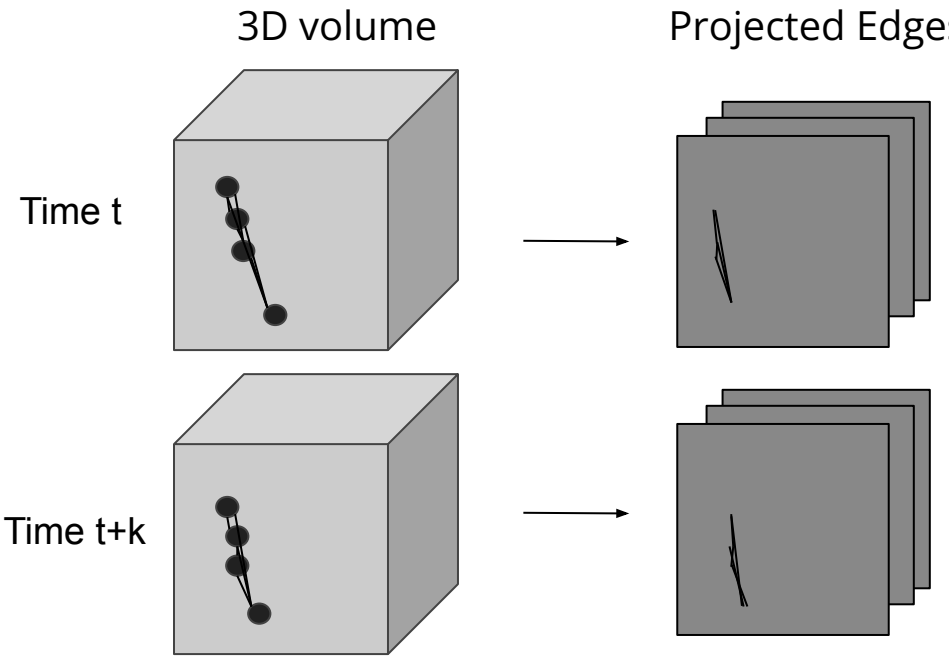
Repeat Across Views



Reprojection



Reprojection



Reprojection

View M, Time t Appearance Features

3D volume

Projected Edges

Time t

Time t+k

View M
Reconstruction

Reconstruction
Decoder

$$D(I_t, I_{t+k})$$

Reprojection

View M, Time t Appearance Features

3D volume

Projected Edges

Time t

Time t+k

Repeat for all views

Reconstruction Decoder

$$D(I_t, I_{t+k})$$

Reprojection

View M, Time t Appearance Features

3D volume

Projected Edges

Time t

Time t+k

Repeat for all views

Reconstruction Decoder

$$D(I_t, I_{t+k})$$

Reprojection

View M, Time t Appearance Features

+ Learned Length Constraint
+ Separation Loss

3D volume

Projected Edges

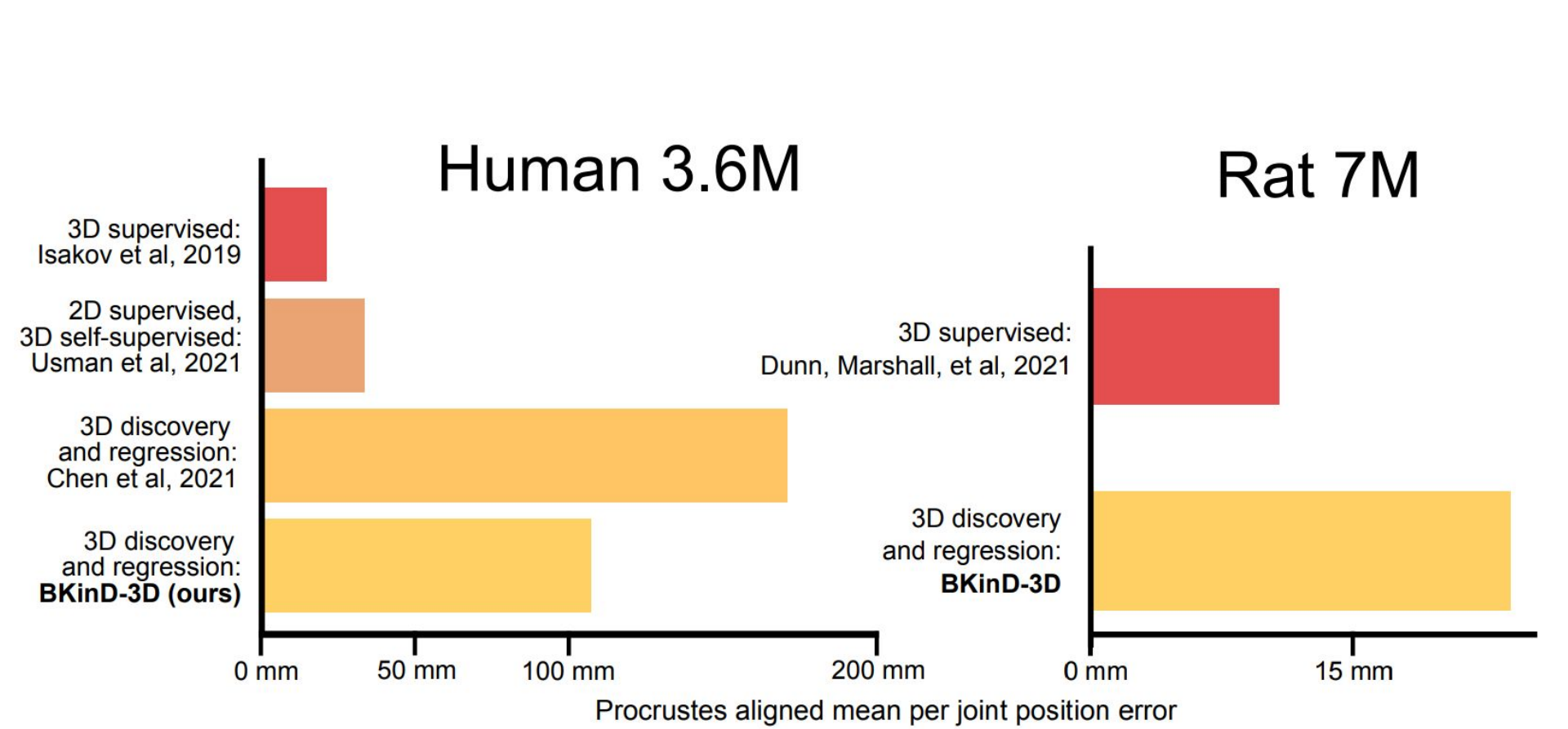
Time t

Time t+k

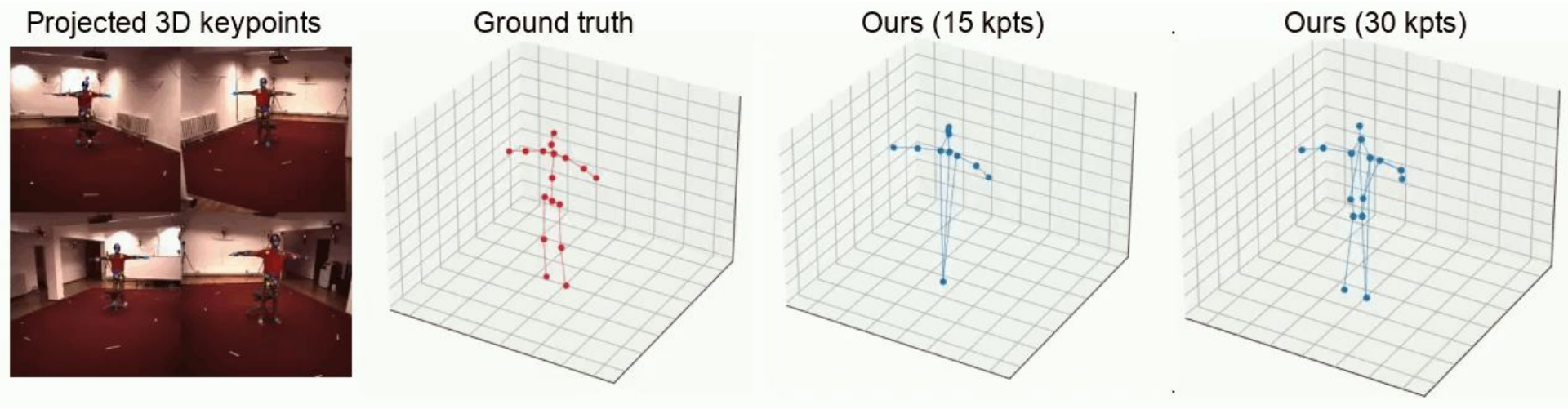
Repeat for all views

Reconstruction Decoder

$$D(I_t, I_{t+k})$$

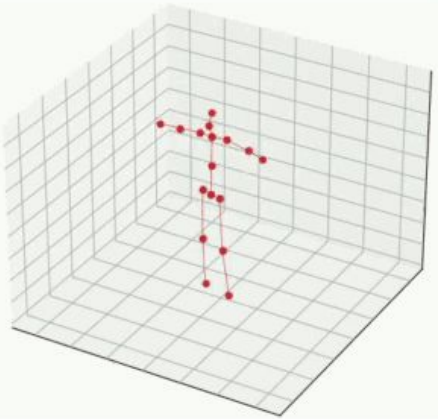

Human 3.6M
Rat 7M
3D supervised:
Isakov et al, 2019
2D supervised,
3D self-supervised:
Usman et al, 2021
3D discovery
and regression:
Chen et al, 2021
3D discovery
and regression:
BKinD-3D (ours)
3D supervised:
Dunn, Marshall, et al, 2021
3D discovery
and regression:
BKinD-3D
0 mm
50 mm
100 mm
200 mm
0 mm
15 mm
Procrustes aligned mean per joint position error

Qualitative Visualizations

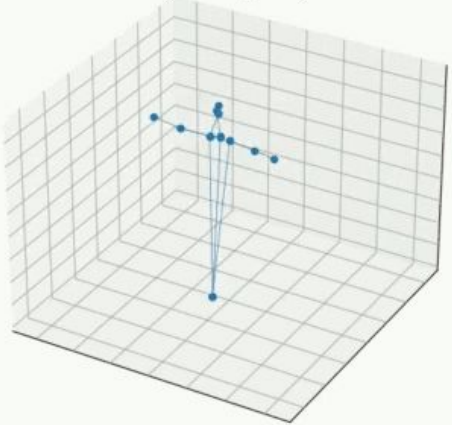


Projected 3D Keypoints

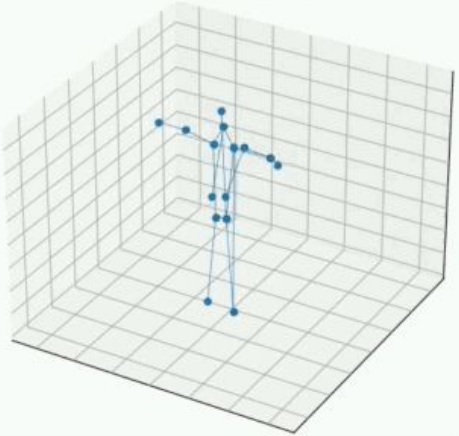
Ground truth



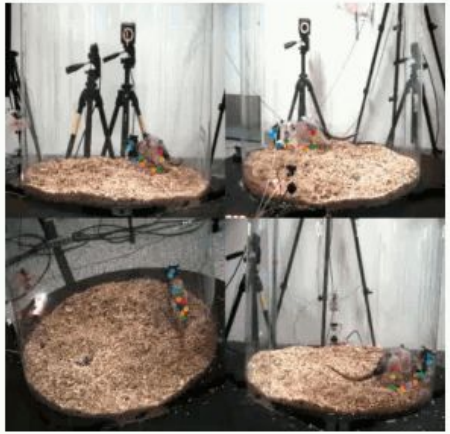
Ours (15 kpts)



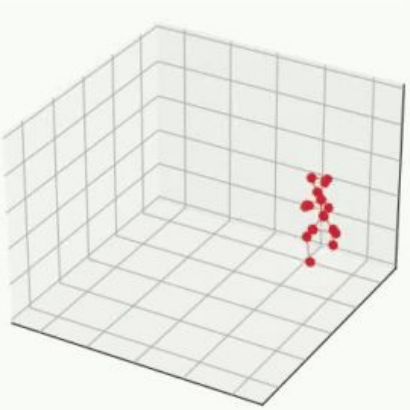
Ours (30 kpts)



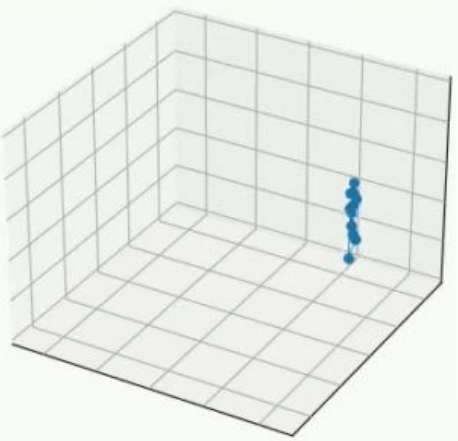
Projected 3D Keypoints



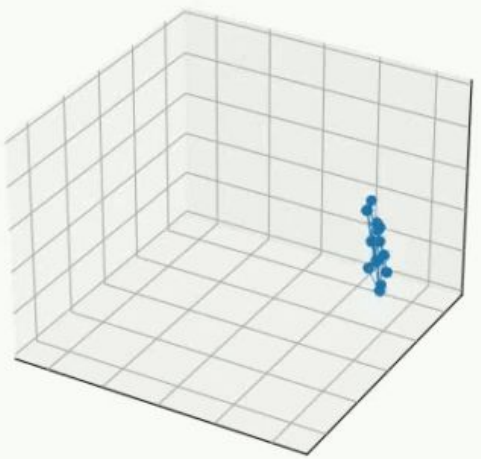
Ground truth

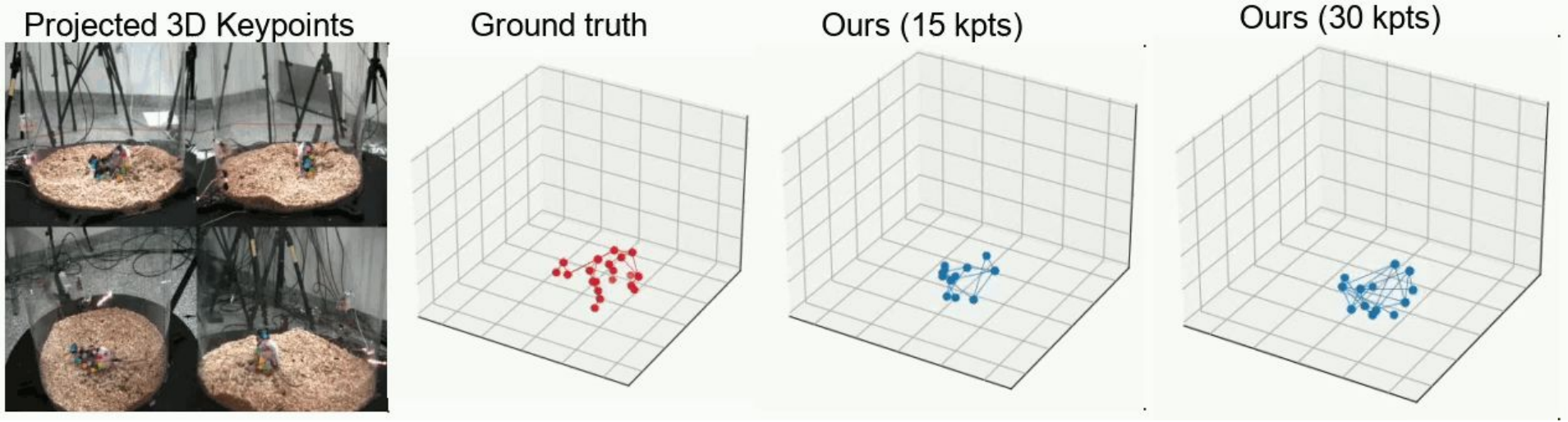


Ours (15 kpts)



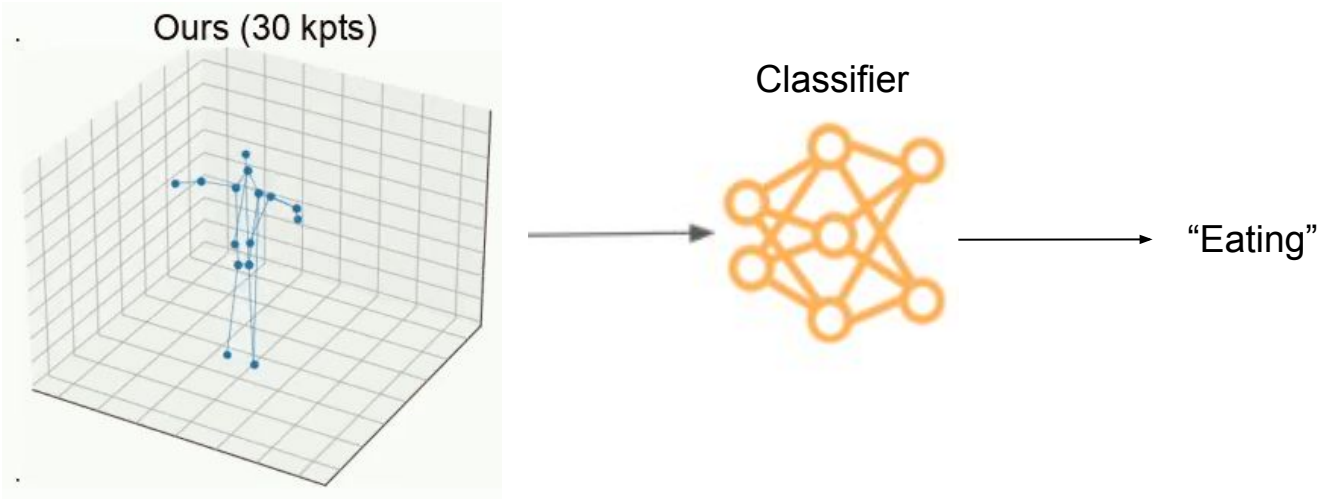
Ours (30 kpts)




Projected 3D Keypoints
Ground truth
Ours (15 kpts)
Ours (30 kpts)

Applications

- Human action recognition

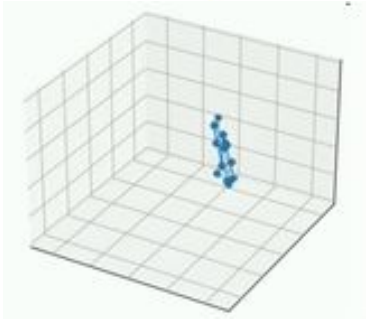


Applications

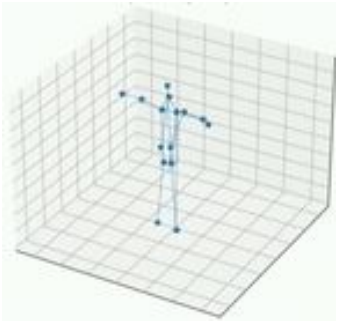
- 16 actions
- Human annotated keypoints: 64.8%
- BKinD-3D keypoints: 64.9%

Other Applications

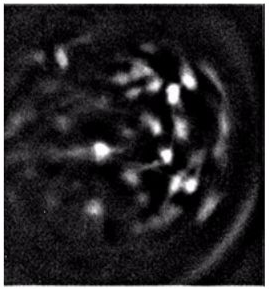
Scientific Experiments



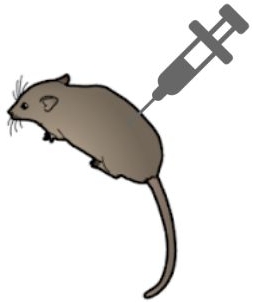
User Applications



Neural Activity



Pharmacological Evaluation



Human-Robot Interactions

Healthcare Monitoring

Project Website:
https://sites.google.com/view/b-kind/3d

Paper Code

Questions? Email: jjsun@caltech.edu

BKinD-3D: Self-Supervised 3D Keypoint Discovery from Multi-View Videos

Caltech

UNIVERSITY of WASHINGTON

Northwestern University

SAMSUNG