KIDELTA LEARNING

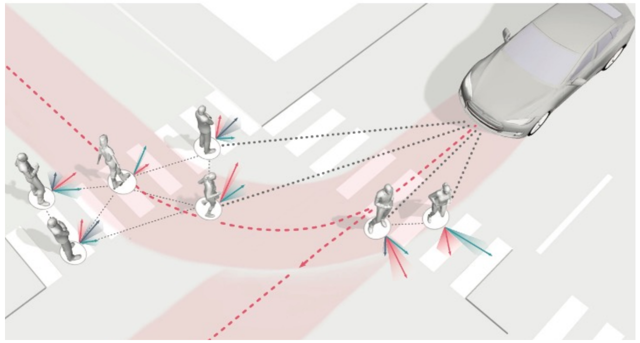
Scalable AI for Automated Driving

Final Event | March 9, 2023

Self-Supervised 3D Human Pose Estimation for Autonomous Driving

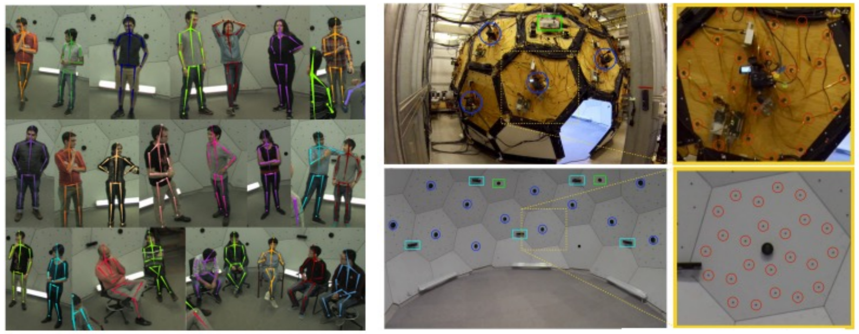
Arij Bouazizi

Motivation

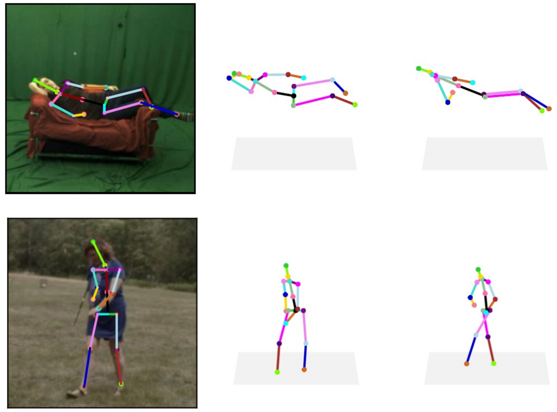


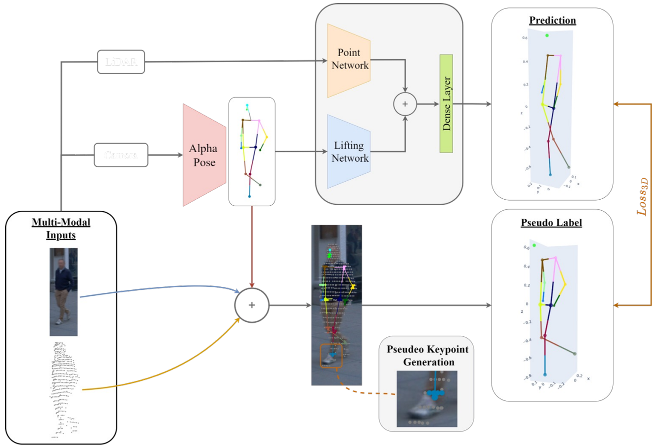
3D Trajectory Prediction
[Ivanovic et al., ICRA2020]

Action & Gesture Anticipation
[Wiederer et al., IROS2020]

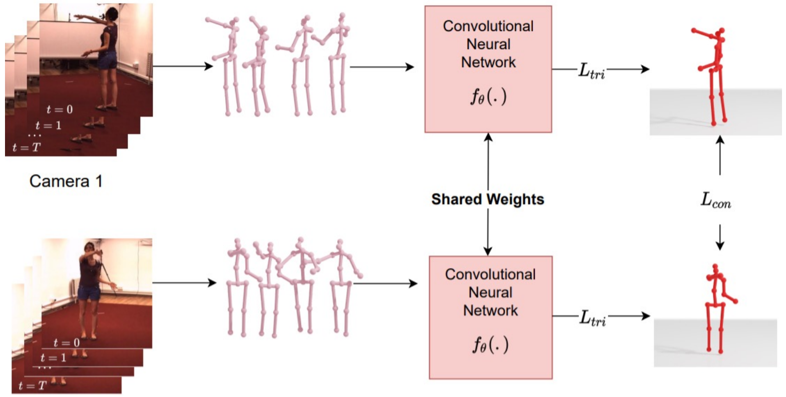


Motion Capture Systems
[Joo et al., TPAMI2016]





LIDAR Supervision



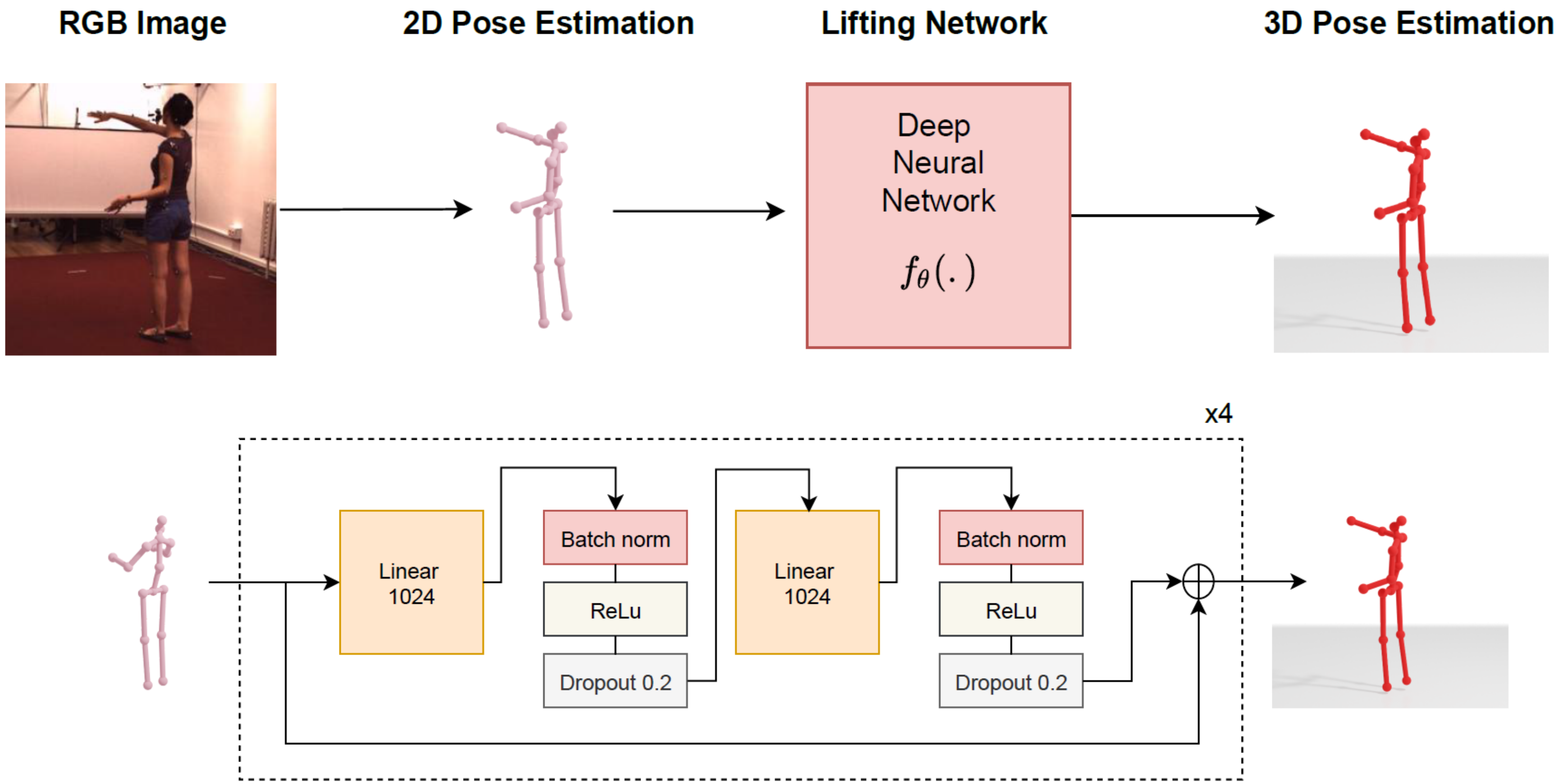
Multi-view Supervision

KIDELTA LEARNING
Scalable AI for Automated Driving

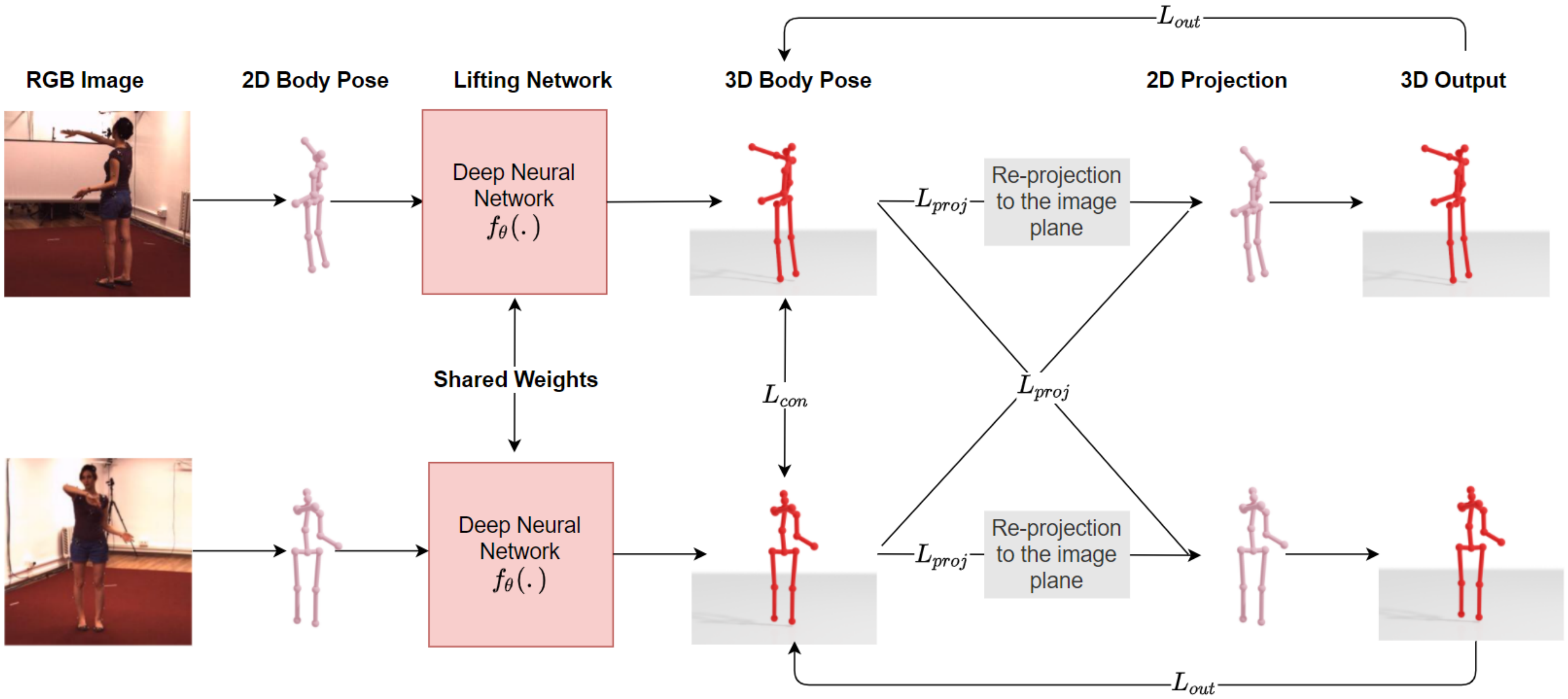
Self-supervised 3D Human Pose Estimation for Indoor Environments

Internal

Self-supervised Single-Frame 3D Human Pose Estimation



Self-supervised Single-Frame 3D Human Pose Estimation



Multi-view Single Frame 3D Human Pose Estimation for Indoor Environments

Self-supervised Single-Frame 3D Human Pose Estimation

Input Triangulation Loss:

$$\mathcal{L}_{in} = \sum_{s=1}^{S}\sum_{c=1}^{C} \| \rho_{w\to c}(\hat{\mathbf{Y}}^{s}_{in}) - f_\theta(\hat{\mathbf{y}}^{s}_{c}) \|^2$$

Projection Loss:

$$\mathcal{L}_{proj} = \sum_{s=1}^{S}\sum_{c=1}^{C}\sum_{c\prime=1}^{C} \| \hat{\mathbf{y}}^{s}_{c} - \tau_c(f_\theta(\hat{\mathbf{y}}^{s}_{c\prime})) \|$$

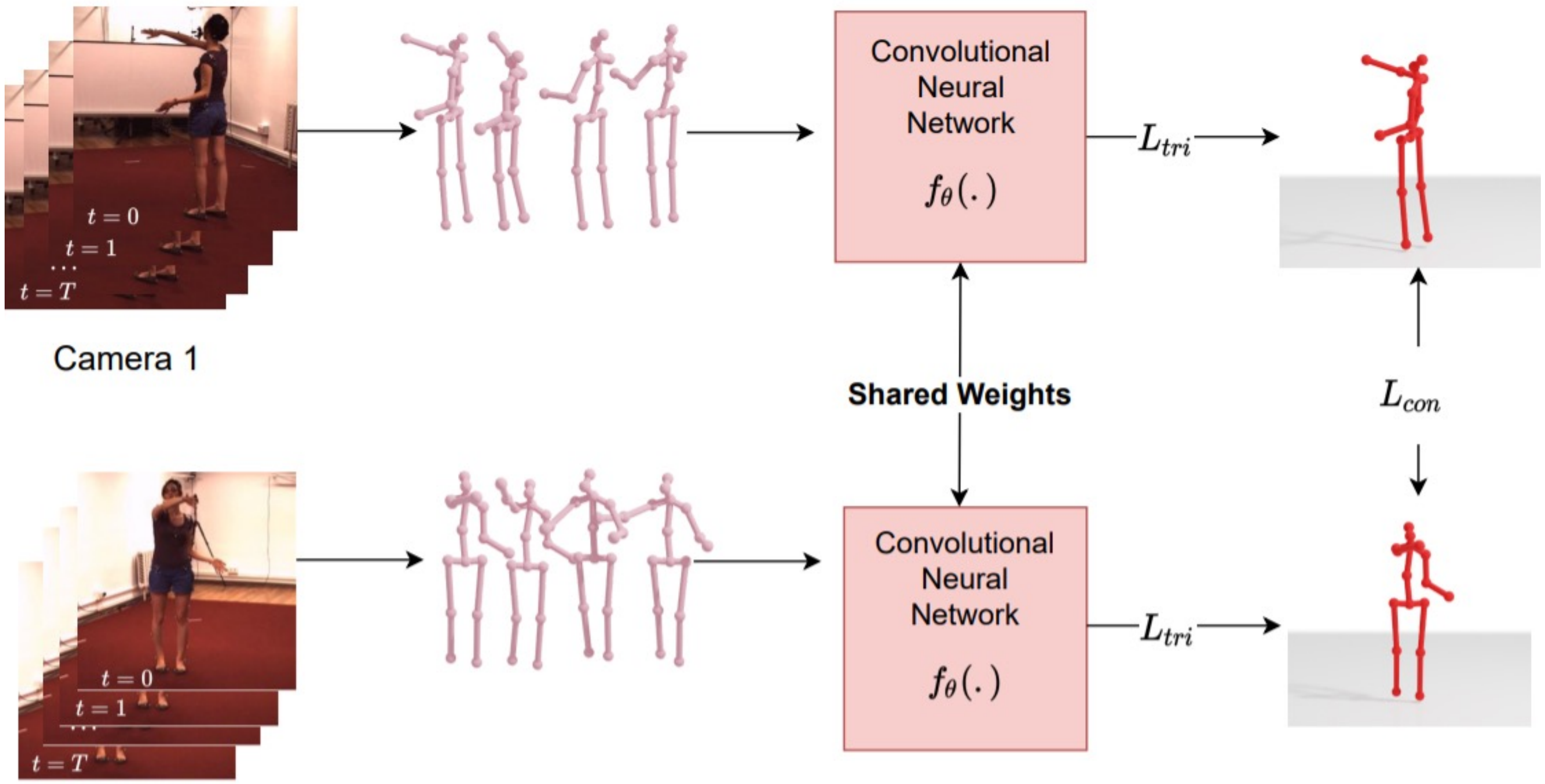
Consistency Loss:

$$\mathcal{L}_{con} = \sum_{s=1}^{S}\sum_{c=1}^{C}\sum_{\substack{c\prime=1\\ c\neq c\prime}}^{C} \| f_\theta(\hat{\mathbf{y}}^{s}_{c}) - \rho_{c\prime\to c}(f_\theta(\hat{\mathbf{y}}^{s}_{c\prime})) \|$$

Output Triangulation Loss:

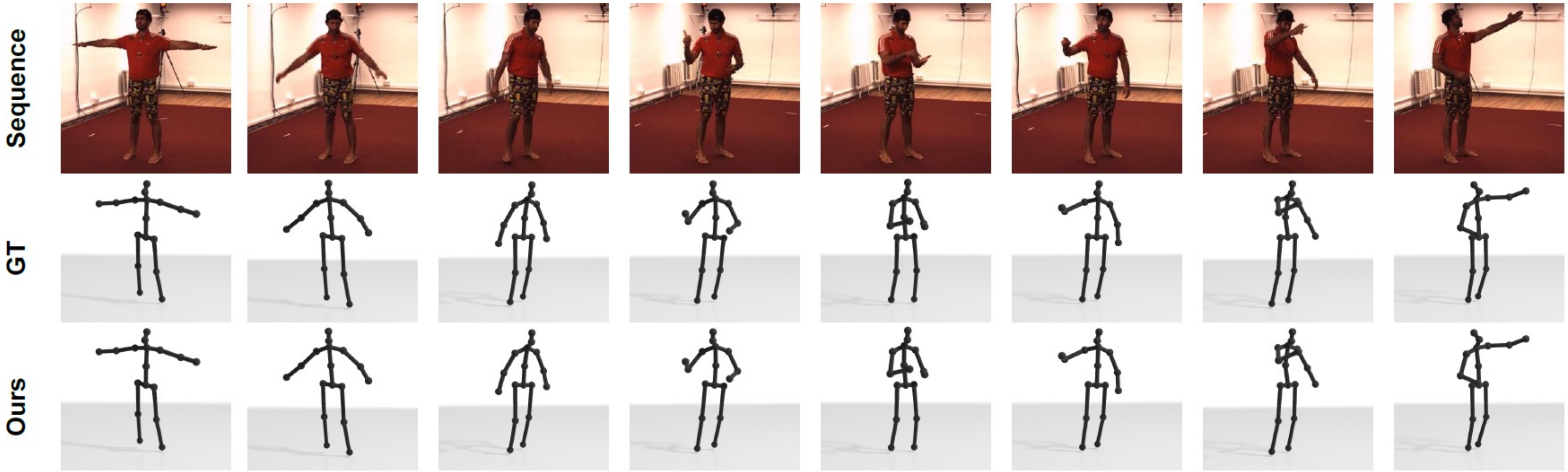
$$\mathcal{L}_{out} = \sum_{s=1}^{S}\sum_{c=1}^{C} \| \rho_{w\to c}(\widetilde{\mathbf{Y}}^{s}_{out}) - f_\theta(\hat{\mathbf{y}}^{s}_{c}) \|^2$$

Self-supervised Temporal 3D Human Pose Estimation

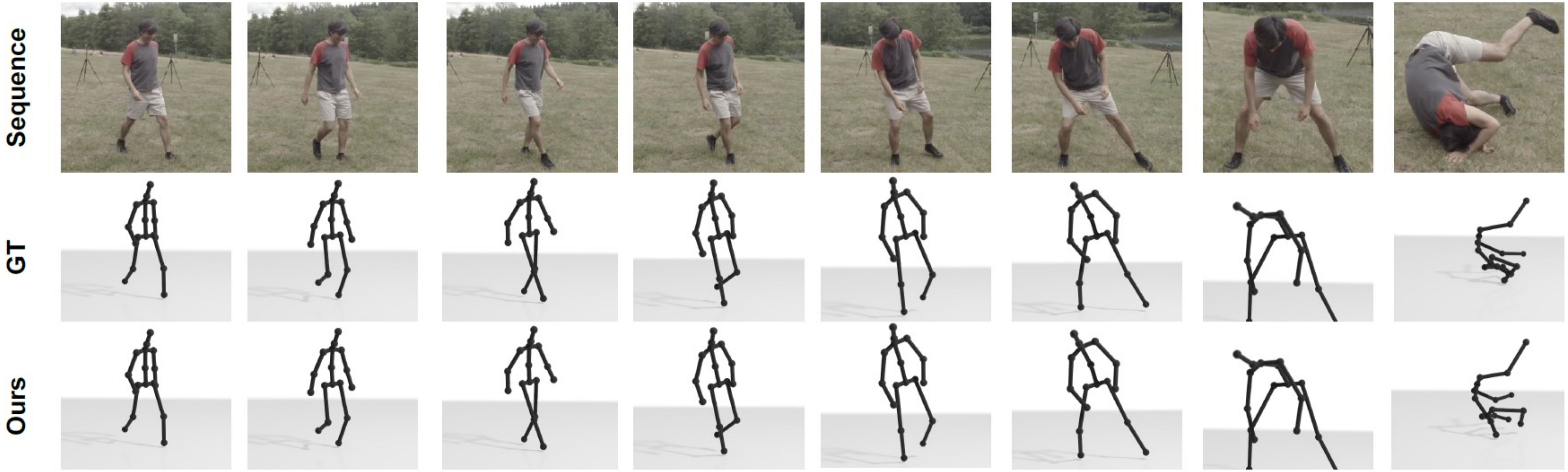


Multi-view Temporal 3D Human Pose Estimation for indoor Environments

Qualitative Results

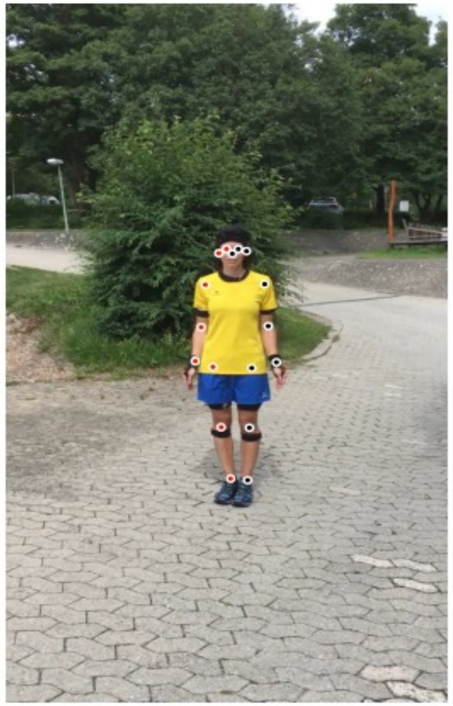


Qualitative Results

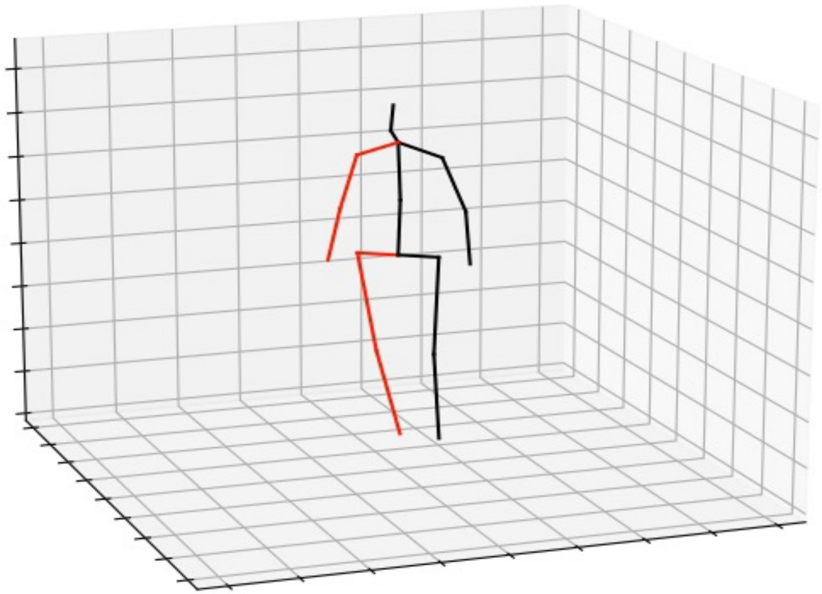


Qualitative Results

Input



Reconstruction



KIDELTA LEARNING

Scalable AI for Automated Driving

Weakly-supervised 3D Human Pose Estimation for Autonomous Driving

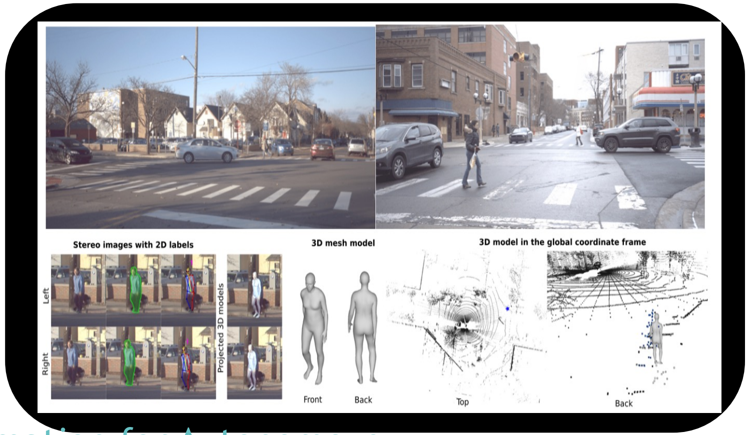
Related Work

Cao et al. *2018* [1]

- Open Pose
- First multi-person realtime 2D pose detection system

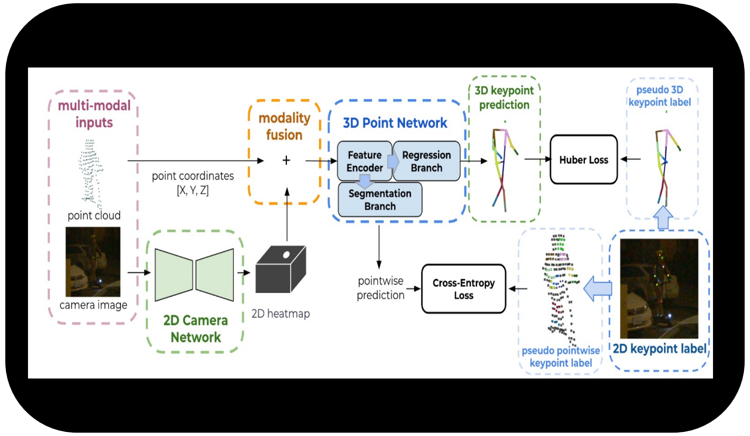
Kim *et al. 2018* [2]

- Energy term minimization
- Strong dependency on labels and sensors
- Evaluation on 3D MOCAP data

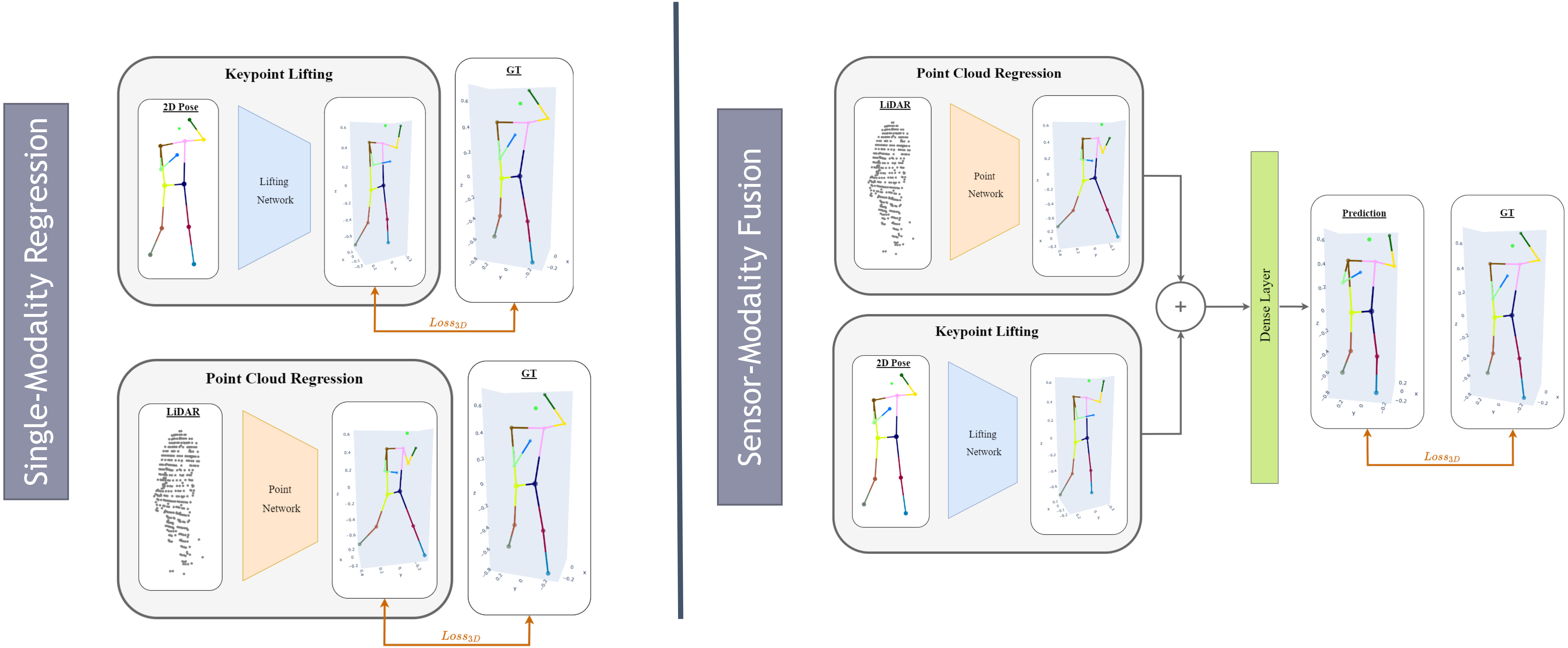


Zheng *et al. 2022* [3]

- Deep learning based
- Multi-modal approach
- Moving vehicle
- 3D pose estimation via weak supervision

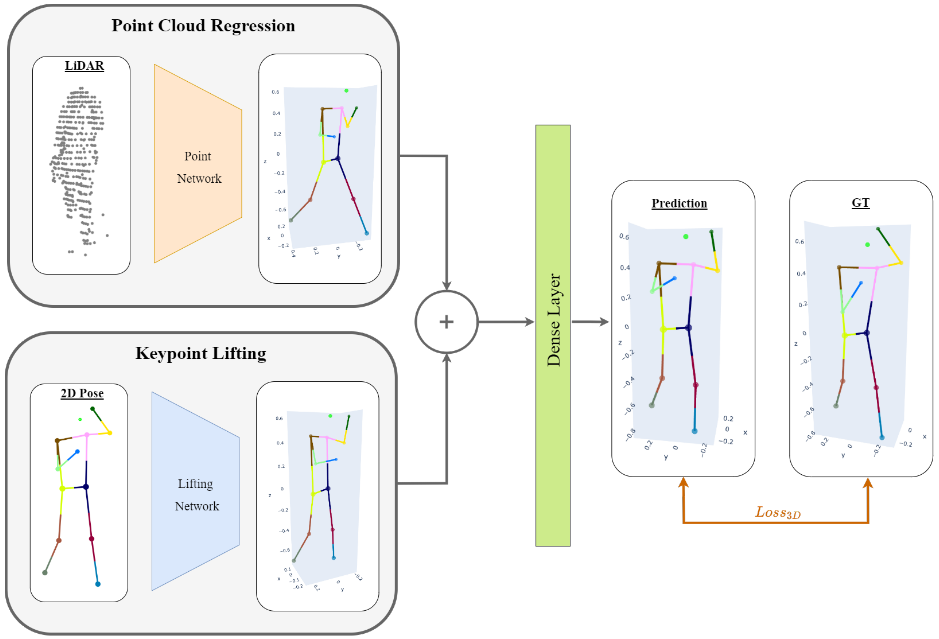


3D Human Pose Estimation for Autonomous Driving

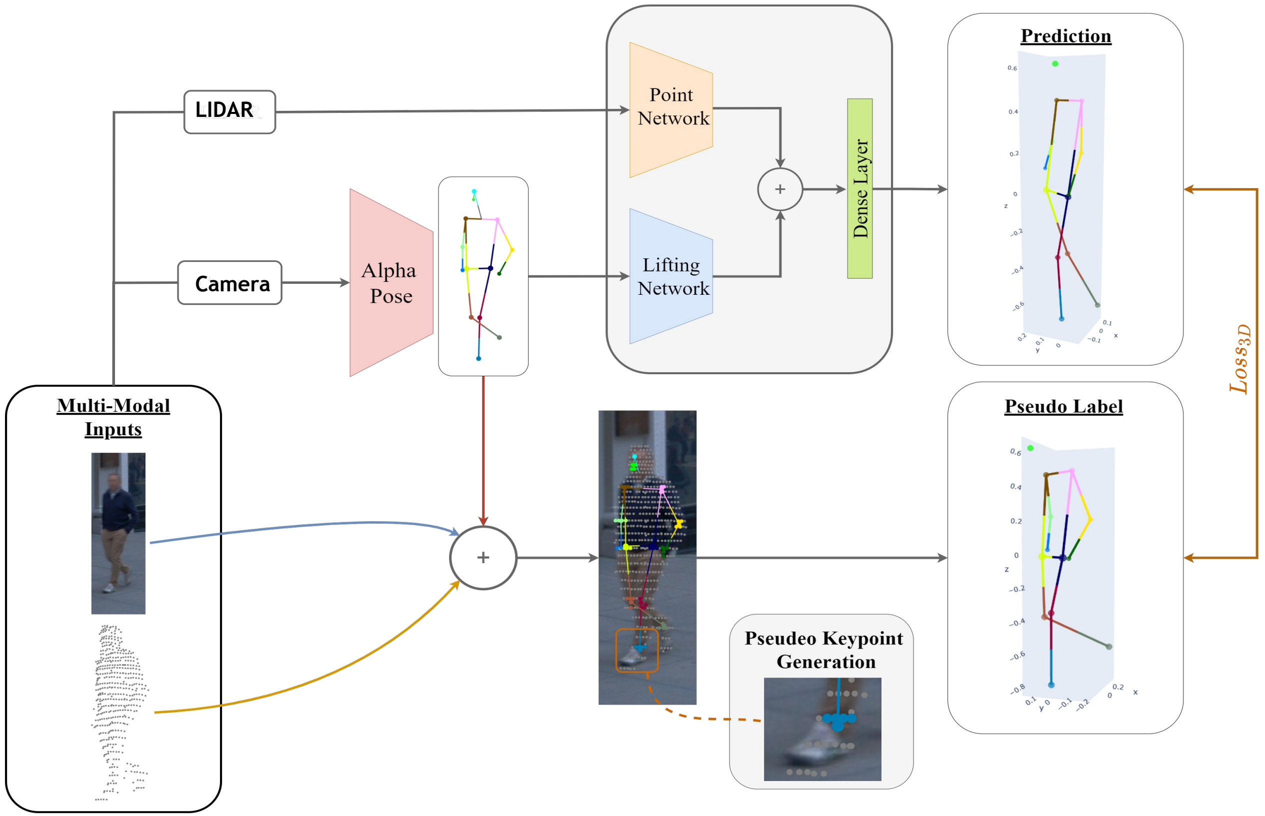


Weakly-supervised Multimodal 3D Human Pose Estimation

Supervised Approach

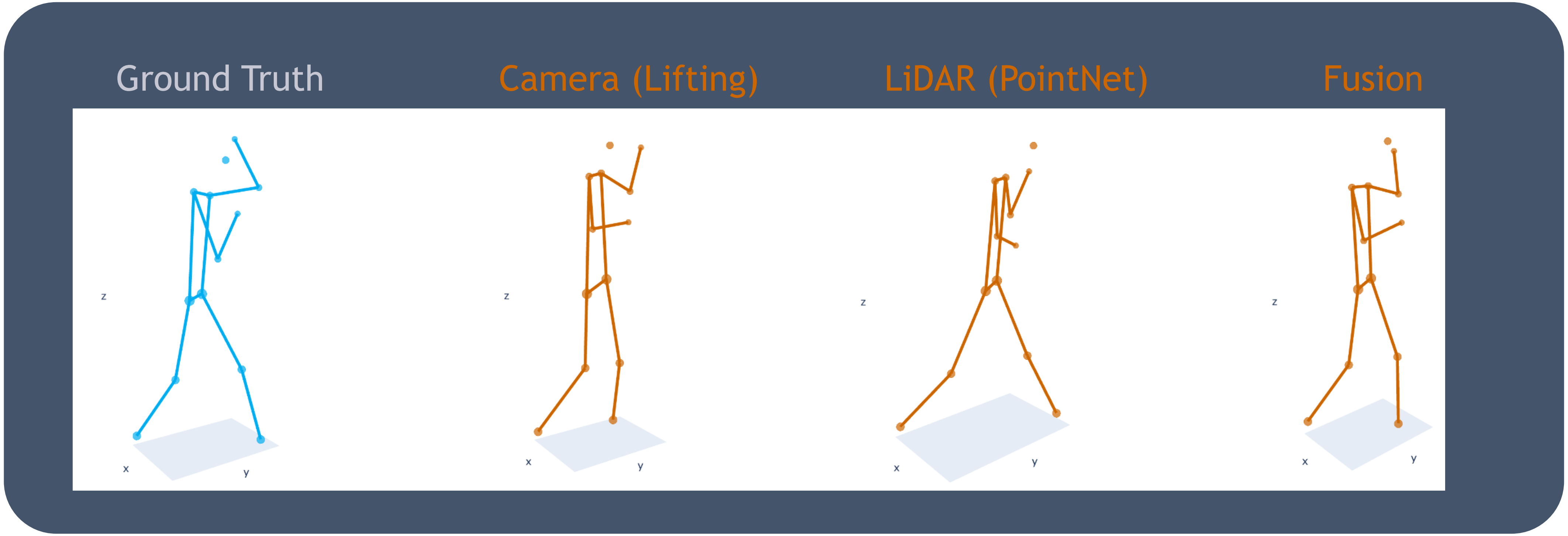


Weakly-supervised Approach



Weakly-supervised Multimodal 3D Human Pose Estimation

- Qualitative Results of the weakly-supervised Approach
  - Comparison between keypoint lifting, LIDAR-based regression and sensor-modality fusion



Conclusion

Self-supervised Training Strategy 3D Human Pose Estimation in Indoor Environments

- Multiple view Supervision without 3D Ground-Truth.
- Single-Frame and Temporal Approaches
- State-of-the-art results on public benchmarks
- Competitive performance to fully-supervised approaches and generalization in the wild.

Self-supervised Training Strategy 3D Human Pose Estimation for Autonomous Driving

- Weakly-supervised Approach
- Multimodal Approach (Camera, LIDAR)

- Future work:

References

[1] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2016). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*, 77–85. https://doi.org/10.48550/arxiv.1612.00593

[2] Kim, W., Ramanagopal, M. S., Barto, C., Yu, M.-Y., Rosaen, K., Goumas, N., Vasudevan, R., & Johnson-Roberson, M. (2018). *PedX: Benchmark Dataset for Metric 3D Pose Estimation of Pedestrians in Complex Urban Intersections*. http://arxiv.org/abs/1809.03605

[3] Zheng, J., Shi, X., Gorban, A., Mao, J., Song, Y., Qi, C. R., Liu, T., Chari, V., Cornman, A., Zhou, Y., Li, C., & Anguelov, D. (2021). Multi-modal 3D Human Pose Estimation with 2D Weak Supervision in Autonomous Driving. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2022-June*, 4477–4486. https://doi.org/10.48550/arxiv.2112.12141

KI DELTA LEARNING

Scalable AI for Automated Driving

Arij Bouazizi | Mercedes-Benz AG

arij.bouazizi@mercedes-benz.com

KI Delta Learning is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

KI FAMILIE

Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag

www.ki-deltalearning.de @KI_Familie KI Familie