MSP-DISK: Naturalistic and Diverse In-Vehicle Database for Joint Pose and Seat Belt Detection

Isaac Brooks	**isaac.brooks@utdallas.edu**
Susmitha Gogineni	sxg155930@utdallas.edu
Sumit Jha	sumit.jha@utdallas.edu
Soumitry Jagadev Ray	soumitryjray@edgetensor.com
Rajesh Narasimha	rajeshnarasimha@edgetensor.com
Naofal Al-Dhahir	aldhahir@utdallas.edu
Carlos Busso	busso@utdallas.edu

IEEE ITSC 2023 BILBAO

NSF

THE UNIVERSITY OF TEXAS AT DALLAS

UT Dallas MSP Multimodal Signal Processing Laboratory

Introduction

- **Dataset**
  - MSP-DISK
    - Diverse in-vehicle seat belt and driver keypoint dataset
- **Light-weight detection model**
  - Two simultaneous tasks
  - Keypoint detection
    - F1-score of 0.773
  - Seat belt segmentation
    - Intersection over union of 27.8%

Outline

1. Significance
2. Related Works
3. Dataset
4. Model Implementation
5. Results
6. *Inspiration* and Future Work

Significance

- **Driver accountability and safety**
  - 3,000 deaths, 300,000 injuries from distracted driving in U.S. in 2020 [1]
- **Driver monitoring system (DMS)**
  - Monitor driver pose/attention/gaze/distraction
  - Safety of driver, passengers, other drivers, pedestrians
- **Dataset**
  - Existing datasets collected in controlled, simulator settings
    - Safety, human subjects
  - Datasets should reflect real-world driving
  - Collect data from driving videos
    - Diverse interiors, subjects, camera position, lighting, etc.

[1] NHTSA, "Distracted driving 2020," National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT), Washington, DC, Technical Report DOT HS 813 309, May 2022. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/81330

Related Works

Database	Context	Variety	Pose	Seat belt
MPII [2]	Out-of-vehicle	-	Yes	No
DriPE [3]	In-vehicle	Multi-vehicle	Yes	No
NADS-Net [4]	In-vehicle	Dedicated vehicle	Yes	Yes
Kim *et. al.* [5]	In-vehicle	Dedicated vehicle	Yes	Yes
Drive&Act [6]	In-vehicle	Dedicated vehicle	Yes	Yes
MSP-DISK (ours)	In-vehicle	Multi-vehicle	Yes	Yes

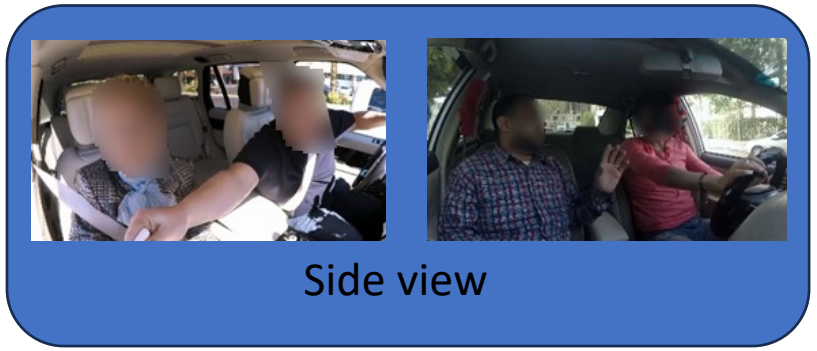
- **Limited in-vehicle datasets**
  - Want human pose and seat belt
  - Occlusion, illumination variations in vehicles
- **Ours expands on diversity**
  - Over 80 unique vehicle interiors
  - Many videos have multiple unique subjects
  - Camera angles

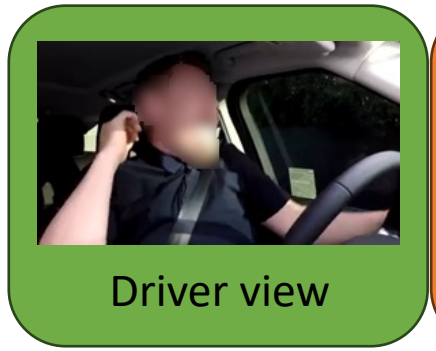
*References on final slide

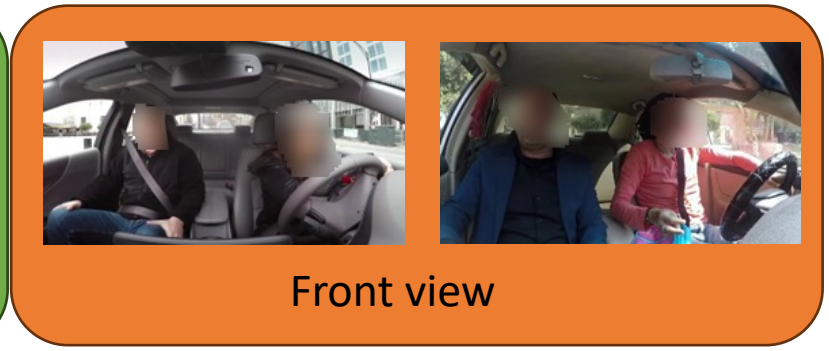
MSP-DISK Dataset – Collection

- **In-vehicle videos**
  - Vlogs, entertainment shows, etc.
- **Cluster by camera angle**
  - Ensure diversity in available angles/scenes
  - Reject frames from scenes outside the vehicle

Partition	Videos	Frames
Train	290	5,038
Development	81	1,095
Test	171	1,571
Total	542	7,704







MSP-DISK Dataset – Keypoint Annotation

- **Improve upon state-of-the-art model performance**
  - Use OpenPose (Cao 2017) [7] for initial keypoint locations
  - Manually adjust incorrect predictions
    - Missed/extra predictions, incorrect locations
  - Reject hidden lower-body keypoints
  - Split into 8 channels
  - Apply Gaussian heatmap centered on each keypoint

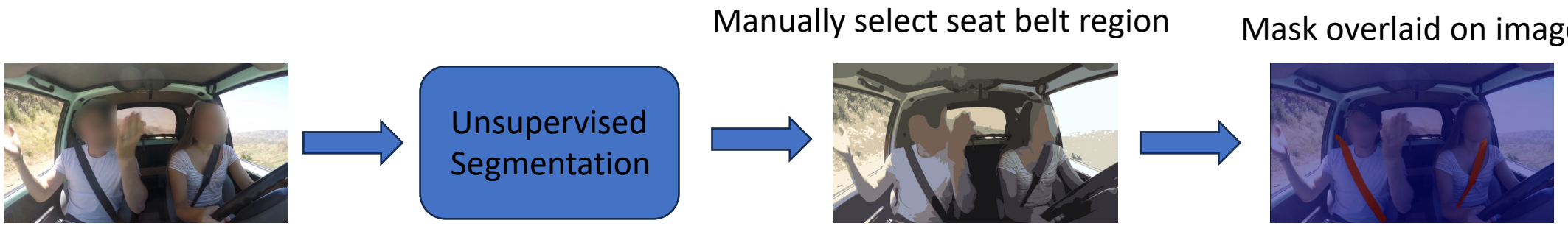
[7] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, January 2021.

Heatmaps overlaid on image

MSP-DISK Dataset – Seat Belt Annotation

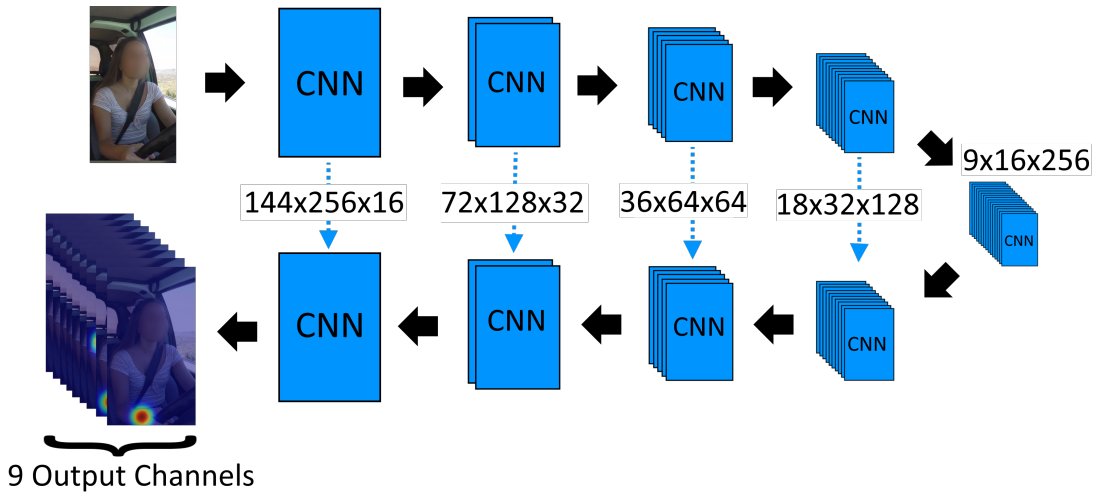
- **Low-cost seat belt segmentation**
  - Perform unsupervised image segmentation (Kanezaki 2018) [8]
    - Group similar colors within image
    - Group shades of seat belt into contiguous region
  - Manually select seat belt region
    - Extract pixel coordinates
  - Use binary mask of seat belt locations

[8] A. Kanezaki, "Unsupervised image segmentation by backpropagation,"
in IEEE International Conference on Acoustics, Speech and Signal
Processing (ICASSP 2018), Calgary, AB, Canada, 2018, April 2018,
pp. 1543–1547.



Lightweight Detection Model

- **U-Net model architecture**
  - Input: 144x256 image cropped per subject
  - Output: 8 keypoint channels + 1 seat belt channel
- **Combination loss function**
  - Weighted L2 loss on keypoint channels
  - Focal loss [9] on seat belt channel
  - Intersection over union loss on all 9 channels



[9] H. Law and J. Deng, "CornerNet: Detecting objects as paired key-points," in European Conference on Computer Vision (ECCV 2018), ser. Lecture Notes in Computer Science, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Munich, Germany: Springer Berlin Heidelberg, September 2018, vol. 11218, pp. 765–781.

Results

- **F1 metrics**
  - Harmonic mean of precision and recall
  - “Correct” prediction within 20px of ground truth
- **Seat belt**
  - Intersection over union
- **Model parameters**
  - Ours: < 2M parameters
  - OpenPose: < 22M parameters

F1 Prediction Scores

Keypoint	OpenPose [7]	Ours	Ours – with flip augmentation
R Wrist	0.443	0.358	0.459
R Elbow	0.670	0.652	0.743
R Shoulder	0.932	0.911	0.928
L Shoulder	0.937	0.905	0.922
L Elbow	0.688	0.682	0.722
L Wrist	0.431	0.299	0.418
Neck	0.965	0.927	0.938
Head	0.985	0.932	0.937
Average	0.757	0.709	0.773
Seat belt [IoU]		21.7%	27.8%

Conclusion

- **MSP-DISK Dataset**
  - Naturalistic images
  - Diverse, challenging images
  - Keypoint and seat belt detection tasks
- **Lightweight detection model**
  - Trained on MSP-DISK
  - Detects upper-body keypoints and seat belt

Thanks

- **NSF grant IIP-1950249**
- **EdgeTensor**
- **Multimodal Signal Processing lab at UT Dallas**

Contact:
Isaac Brooks (Isaac.Brooks@utdallas.edu)
Carlos Busso (busso@utdallas.edu)

MSP-DISK Dataset GitHub

MSP-DISK Code GitHub

References

[1] NHTSA, "Distracted driving 2020," National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT), Washington, DC, Technical Report DOT HS 813 309, May 2022. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/81330

[2] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014), Columbus, OH, USA, June 2014, pp. 3686–369

[3] R. Guesdon, C. Crispim-Junior, and L. Tougne, "DriPE: A dataset for human pose estimation in real-world driving settings," in IEEE/CVF International Conference on Computer Vision Workshops (ICCVW 2021), Montreal, BC, Canada, April 2021, pp. 2865–28

[4] S. Chun, N. Ghalehjegh, J. Choi, C. Schwarz, J. Gaspar, D. McGehee, and S. Baek, "NADS-Net: A nimble architecture for driver and seat belt detection via convolutional neural networks," in IEEE/CVF International Conference on Computer Vision Workshop (ICCVW 2019), Seoul, Republic of Korea, October 2019, pp. 2413–242

[5] G. Kim, H. Kim, J. Kihoon Kim, S.-S. Cho, Y.-H. Park, and S.-J. Kang, "Integrated in-vehicle monitoring system using 3D human pose estimation and seat belt segmentation," in AAAI 2022 workshop AI for Transportation, Vancouver, BC, Canada, February-March 2022, pp. 1–8.

[6] M. Martin, A. Roitberg, M. Haurilet, M. Horne, S. Reiß, M. Voit, and R. Stiefelhagen, "Drive&act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles," in IEEE/CVF International Conference on Computer Vision (ICCV 2019), Seoul, Republic of Korea, October - November 2019, pp. 2801–28

[7] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, January 2021.

[8] A. Kanezaki, "Unsupervised image segmentation by backpropagation," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018), Calgary, AB, Canada, 2018, April 2018, pp. 1543–1547.

[9] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in European Conference on Computer Vision (ECCV 2018), ser. Lecture Notes in Computer Science, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Munich, Germany: Springer Berlin Heidelberg, September 2018, vol. 11218, pp. 765–781.