

# Bodily Behaviors in Social Interaction: Novel Annotations and State-of-the-Art Evaluation

Supplementary Material

Michal Balazia\*  
michal.balazia@inria.fr  
INRIA Sophia Antipolis  
Sophia Antipolis, France

Philipp Müller\*  
philipp.mueller@dfki.de  
DFKI Saarbrücken  
Saarbrücken, Germany

Ákos Levente Tanczos  
akos.tanczos@inria.fr  
INRIA Sophia Antipolis  
Sophia Antipolis, France

August von Liechtenstein  
august.liechtenstein@dfki.de  
DFKI Saarbrücken  
Saarbrücken, Germany

François Brémond  
francois.bremond@inria.fr  
INRIA Sophia Antipolis  
Sophia Antipolis, France

## CCS CONCEPTS

• **Computing methodologies** → **Image and video acquisition; Activity recognition and understanding**; • **Human-centered computing** → *Collaborative and social computing*; • **Applied computing** → *Psychology*.

## KEYWORDS

dataset, body pose, gesture, social signals, behavior detection

### ACM Reference Format:

Michal Balazia, Philipp Müller, Ákos Levente Tanczos, August von Liechtenstein, and François Brémond. 2022. Bodily Behaviors in Social Interaction: Novel Annotations and State-of-the-Art Evaluation: Supplementary Material. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*, October 10–14, 2022, Lisboa, Portugal. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3503161.3548363>

## 1 CONFIDENCE MATRICES OF BEHAVIOR RECOGNITION METHODS

Evaluation of all methods can be visualized by aggregating all confidence vectors into a confidence matrix. See Figure 1 and Figure 2 for the confidence matrices of all four behavior recognition algorithms.

## 2 ILLUSTRATIVE EXAMPLES OF PDAN PREDICTIONS

We analyzed four major sources of impact on performance: additional person in the scene, occlusion, low class agreement, distinction between visually similar classes. Figure 3 illustrates these examples in the context of ground truth and prediction.

\*Authors contributed equally.

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

*MM '22, October 10–14, 2022, Lisboa, Portugal*

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-9203-7/22/10...\$15.00  
<https://doi.org/10.1145/3503161.3548363>

## REFERENCES

- [1] João Carreira and Andrew Zisserman. 2017. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 4724–4733. <https://doi.org/10.1109/CVPR.2017.502>
- [2] Ji Lin, Chuang Gan, and Song Han. 2019. TSM: Temporal Shift Module for Efficient Video Understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 7082–7092. <https://doi.org/10.1109/ICCV.2019.00718>
- [3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 10012–10022.
- [4] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. 2016. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. In *ECCV*.

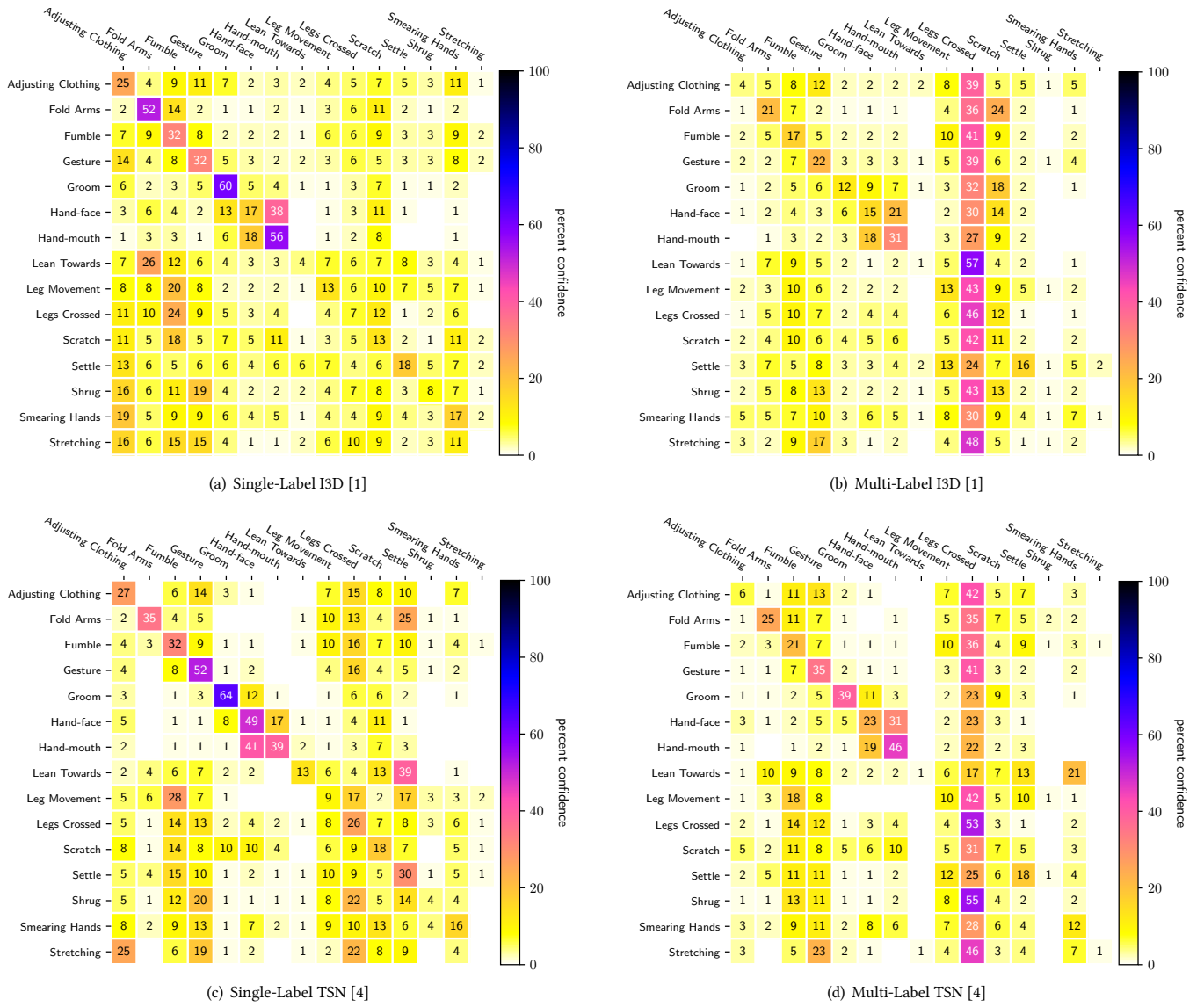


Figure 1: Confidence matrices of behavior recognition by I3D and TSN, trained in both labeling settings.

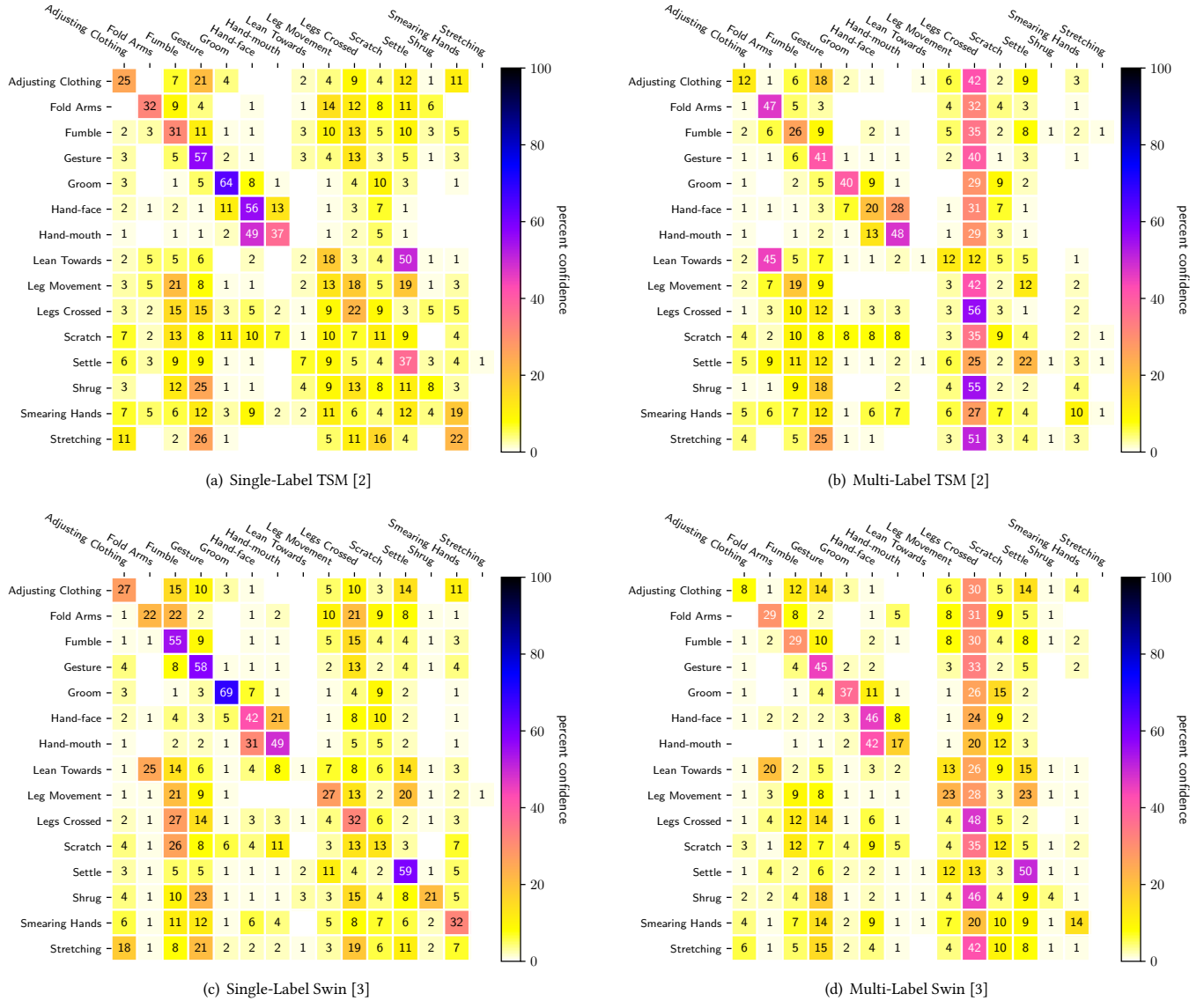
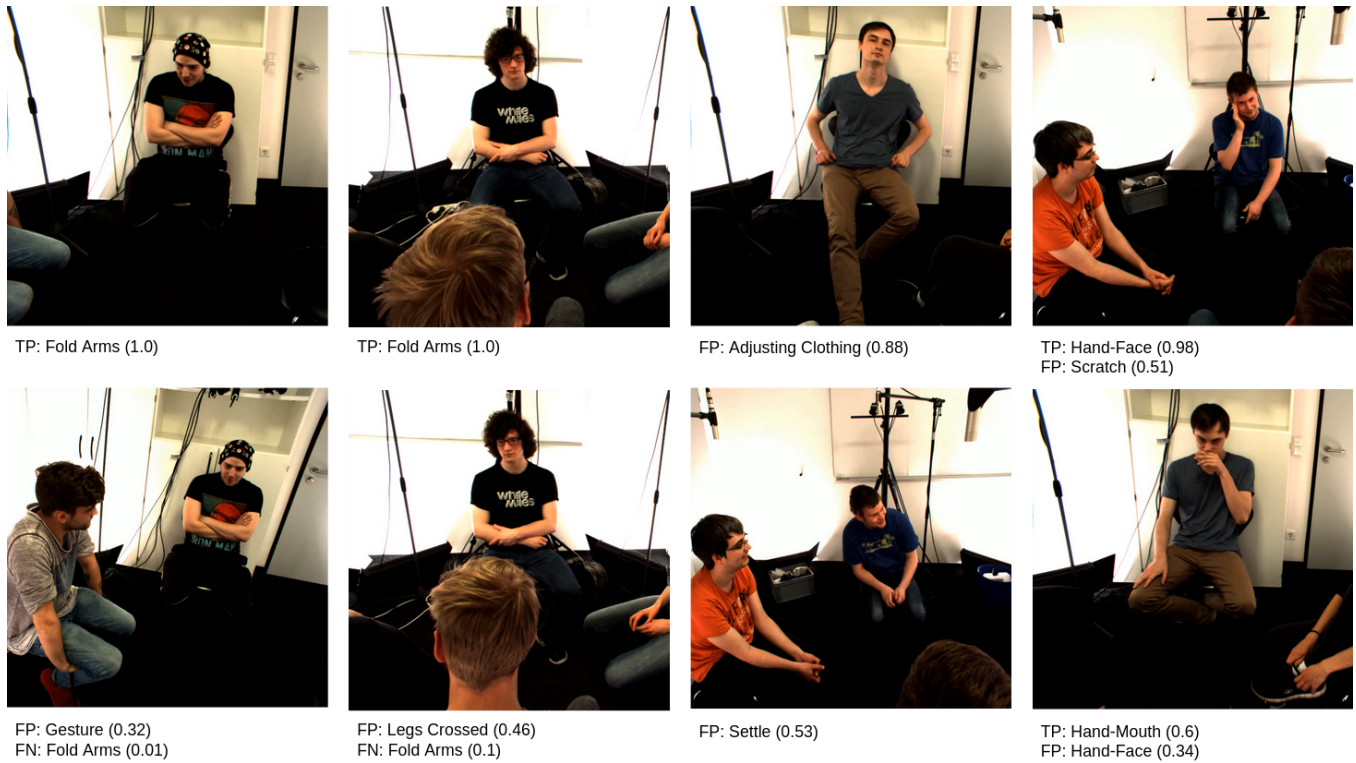


Figure 2: Confidence matrices of behavior recognition by TSM and Swin, trained in both labeling settings.



**Figure 3: Illustrative examples of PDAN predictions on the frame level. Ground truth and acceptance threshold of 0.3 determine one of the three statuses: True Positive (TP), False Positive (FN) and False Negative (FN). First column shows a decrease in performance caused by an additional person present in the scene. Second column shows the effect of occlusion. Third column indicates that classes with lower level of agreement have a generally worse performance. Fourth column highlights the great challenge of distinguishing between classes of very subtle differences.**