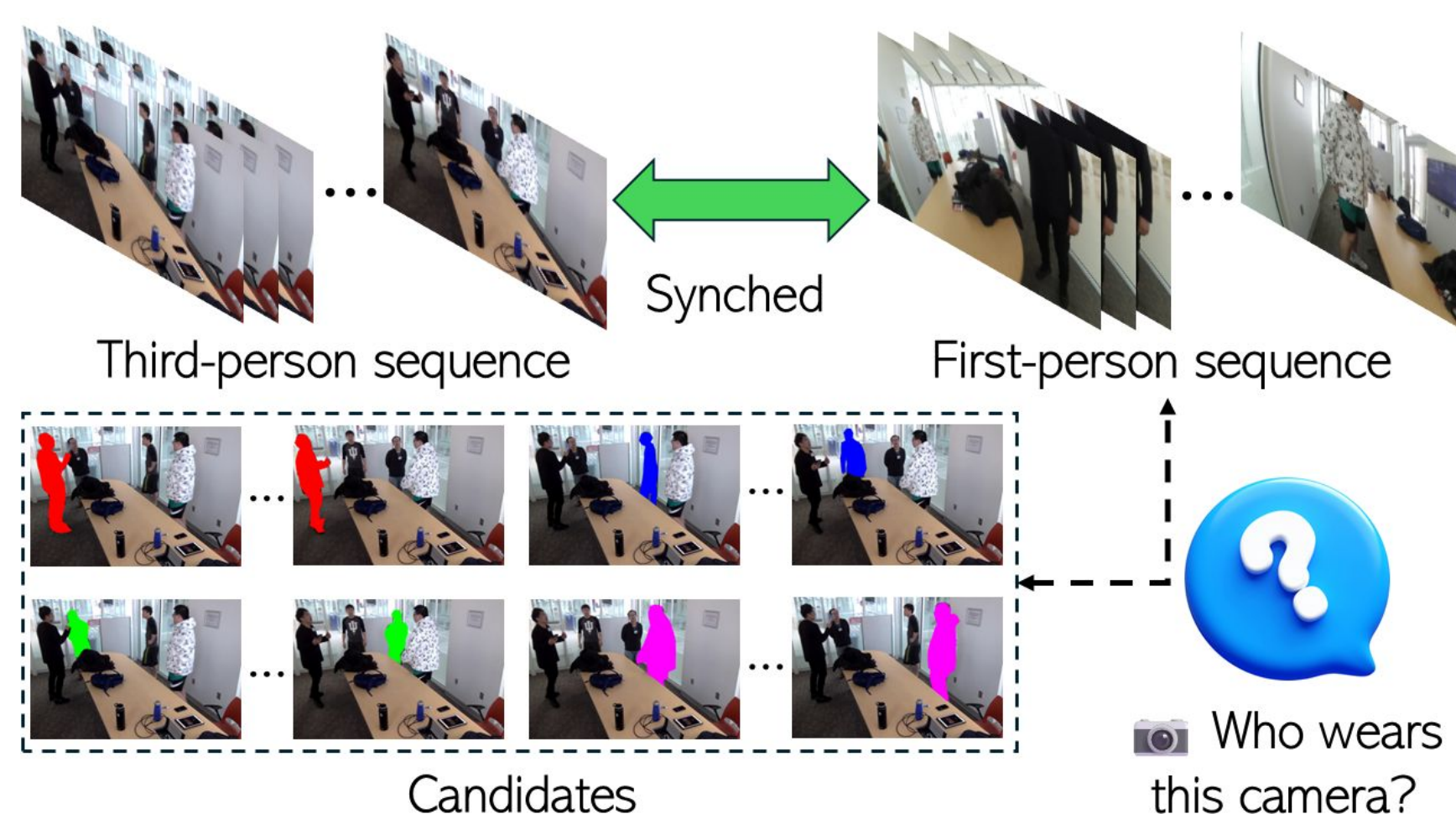




Introduction



Goal: Identify a camera wearer using only their first-person view as clue.

Challenges:

- Appearance of camera wearer not available.
- Differences between first- and third-person perspectives in field-of-view, object details, and motion blur.

Prior work [1,2,3]:

- Contrastive learning based on images and optical flow.
- Frame-based heuristics, such as geometric matching and appearance.

Data Collection

We created a dataset from two sources:

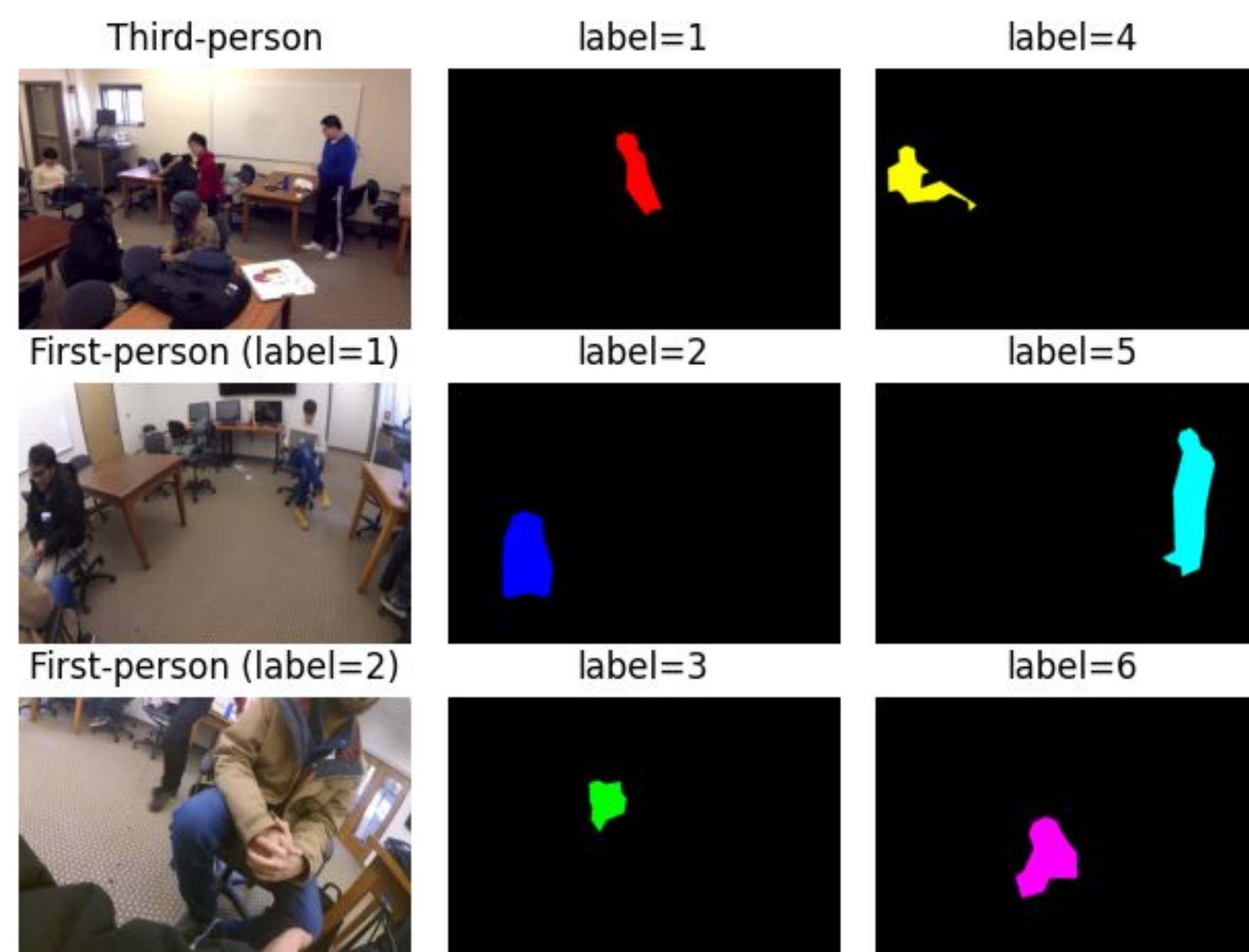
- Recorded and annotated by ourselves.
- From Ego4d, with our annotations.

Dataset Configuration:

Each frame includes a third-person view, two first-person views synchronized with third-person, a segmentation mask for each individual, with labels marking their identities.

Three different train/test splits:

Seen wearer, Unseen wearer, Cross-dataset



Data Samples

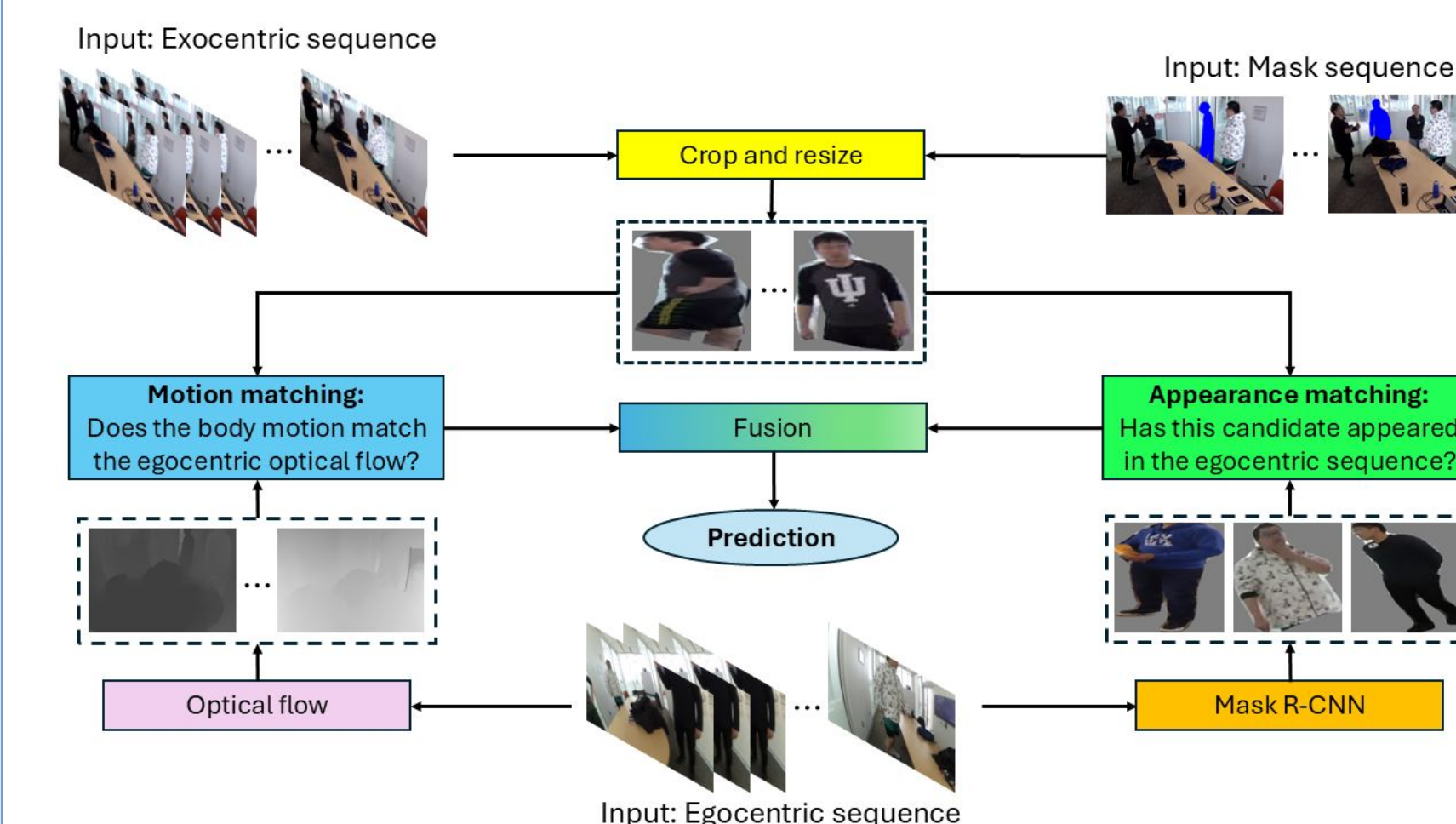


Methodology and Results

Our method consists of two branches:

1. **Motion matching:** We match the optical flow in the first-person view to the body motion of the camera wearer in the third-person view.
2. **Appearance matching:** We filter out people that can be seen in the first-person view.

The results from the two branches are fused to make the final prediction.



Accuracy↑	Seen	Unseen	Cross
Random	28.7%	33.1%	46.4%
SOTA [3]	85.9%	58.8%	52.9%
Ours	93.8%	84.4%	75.7%

Contact

Ziwei Zhao
Indiana University Bloomington
zz47@iu.edu

Acknowledgement

This work was supported in part by the National Science Foundation under award DRL-2112635 to the AI Institute for Engaged Learning. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Identifying first-person camera wearers in third-person videos. Fan et al., CVPR 2017
2. Joint Person Segmentation and Identification in Synchronized First- and Third-person Videos, Xu et al., ECCV 2018
3. Fusing personal and environmental cues for identification and segmentation of first-person camera wearers in third-person views. Zhao et al., CVPR 2024