Supplementary Material for A Benchmark Dataset for Event-Guided Human Pose Estimation and Tracking in Extreme Conditions

Hoonhee Cho*
KAIST
gnsgnsgml@kaist.ac.kr

Taewoo Kim*
KAIST
intelpro@kaist.ac.kr

Yuhwan Jeong
KAIST
jeongyh98@kaist.ac.kr

Kuk-Jin Yoon
KAIST
kjyoon@kaist.ac.kr

Abstract

We have included in the supplementary material the parts that we could not mention in the main paper. Section A covers the implementation details, Section B presents additional experiments, and Section C describes the detailed annotation process. Lastly, we have included a description of the license and ethical considerations in the Section D.

A Implementation Details

We conducted all experiments on models using the PyTorch framework. The training is conducted on an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz and an NVIDIA RTX A6000. We trained the model with a batch size of 12 and selected the best-performing model during training from 300 epochs to evaluate on the test set. The Adam [4] optimizer was used for training. The learning rate was initially set to 1×10^{-3} and was reduced by a factor of 0.1 at epochs 200 and 260.

B Additional Experiments

B.1 Baselines

We additionally provide experimental results for more baselines in Table B1. As shown in Table B1, the best performance is achieved in the multi-modal setting, regardless of changes to the baseline, proving that it is agnostic to the network.

B.2 Additional Splits

The train and test splits in the main paper are based on the random seed that ensures the most balanced distribution of the environment. We evenly distributed low-light environment scenes across the training and test sets and reviewed the corresponding experimental results.

Out of the total 158 sequences, there are 39 low-light scenes. These low-light scenes were distributed, with 29 in the training set and 10 in the test set, ensuring a balanced distribution with 121 sequences

*Equal contribution.

38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks.

Table B1: Multi-person pose estimation using recent baselines. We adopted the fusion method [6] for RGB+Event.

Modality	Method	mAP@0.5:0.95	mAP@0.5	mAP@0.75	mAR@0.5:0.95	mAR@0.5	mAR@0.75
RGB	GroupPose [5]	30.3	34.9	31.1	77.1	86.4	78.8
	ED-Pose [9]	33.1	39.2	34.0	75.5	83.7	76.4
Event	GroupPose [5]	32.2	44.6	32.8	64.8	78.9	68.1
	ED-Pose [9]	34.3	45.1	31.1	69.4	82.0	71.9
RGB+Event	GroupPose [5]	34.9	42.6	35.6	77.7	86.7	79.8
	ED-Pose [9]	38.1	48.4	38.8	74.8	85.3	76.3

in the training set and 37 sequences in the test set. We demonstrate the experimental results in this additional split for the multi-person pose estimation in Tab.B2 and multi-perspon pose tracking in Tab.B3 in this additional split. We observed consistent performance improvement when additional event modality was incorporated, similar to the results in the main paper.

Table B2: Multi-person pose estimation baselines evaluated on the split by light environment.

Modality	Method	mAP@0.5:0.95	mAP@0.5	mAP@0.75	mAR@0.5:0.95	mAR@0.5	mAR@0.75
RGB	HigherHRNet [2]	21.4	29.2	22.2	61.3	78.4	63.6
	DEKR [3]	23.0	31.2	23.6	58.7	78.8	59.8
	CID [7]	19.5	28.6	19.8	51.6	74.4	52.9
RGB + Event	HigherHRNet [2]	23.6 (+2.2)	31.6	24.3	64.6 (+3.3)	83.3	66.6
	DEKR [3]	27.1 (+4.1)	36.2	27.6	67.9 (+9.2)	89.0	69.8
	CID [7]	22.5 (+3.0)	32.3	22.7	60.1 (+8.5)	85.3	62.0

Table B3: Multi-person pose tracking baselines evaluated on the split by light environment.

Modality	Pose Estimation	Tracking	MOTA↑	IDF1↑	FP↓	IDSW↓	FN↓
RGB	DEKR [3]	ByteTrack [10]	24.52	13.99	278	290	6381
		UniTrack [8]	18.02	7.82	79	151	7381
		OC-SORT [1]	15.58	12.47	90	115	7568
RGB + Event	DEKR [3]	ByteTrack [10]	38.46 (+13.94)	18.05	395	433	4839
		UniTrack [8]	34.34 (+16.32)	7.04	172	406	5467
		OC-SORT [1]	27.42 (+11.84)	17.04	144	235	6303

C Annotation Details

For accurate annotation of degraded (motion blurred and/or low-light) images, annotations were made on sharp and well-lit reference images. A meticulous process was designed to ensure accurate annotation. Depending on the situation, bounding boxes may not always perfectly fit the external points of keypoints and can sometimes be considerably larger. Therefore, rather than calculating the bounding box from annotated keypoints, we annotated keypoints and bounding boxes separately. Furthermore, we connected annotations for the two tasks, which were performed separately, through the data association process. Additionally, during this process, we refined incorrectly set boundaries for boxes and keypoints that were not accurately annotated. The entire process are represented as follows:

(1) Annotating bounding boxes and track IDs for multi-object tracking.

(2) Annotating keypoints for multi-human pose estimation.

(3) Associating bounding boxes with keypoints, including track IDs, and reviewing each annotation.

Cross-checks between annotators were conducted at each step to enhance labeling quality. Fig. C1 shows the custom tool developed for our annotation toolbox. Special efforts were made in step (3) to refine annotations by addressing any missing or unmatched elements from steps (1) and (2), resulting in more precise annotations.

Figure C1: A custom-designed association tool.

D License & Ethical Impact

D.1 License

EHPT-XC dataset is designed solely for research purposes and is licensed under CC BY-NC 4.0, permitting non-commercial use only. Additionally, users granted access to the dataset must sign relevant usage agreements and provide information. These measures are in place to protect the privacy and security of individuals associated with the EHPT-XC dataset and to prevent data misuse.

D.2 Consent Form For Participant

As mentioned in the main paper, we voluntarily recruited experiment participants and provided them with ample time to understand the potential risks before obtaining their signatures on the consent form. As shown in Fig. D2, the consent form included notifications about the types of data collected and the purposes of collection. We have anonymized all data as thoroughly as possible to prevent the disclosure of any personal information of the individuals involved

D.3 Maintenance Plan & Usage Agreements

We recognize the critical role of data governance and the necessity to prevent misuse or unintended harm. We urge researchers and users to manage our dataset ethically and with respect for privacy. As shown in Fig. D3, prior to accessing our data, researchers must agree to adhere to our licensing terms. We are dedicated to continuous dialogue and cooperation with field experts to address concerns, aiming to benefit the research community while reducing any possible negative social impacts.

References

[1] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani. Observation-centric sort: Rethinking sort for robust multi-object tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9686–9696, 2023.

[2] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5386–5395, 2020.

[3] Z. Geng, K. Sun, B. Xiao, Z. Zhang, and J. Wang. Bottom-up human pose estimation via disentangled keypoint regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14676–14686, 2021.

[4] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[5] H. Liu, Q. Chen, Z. Tan, J.-J. Liu, J. Wang, X. Su, X. Li, K. Yao, J. Han, E. Ding, et al. Group pose: A simple baseline for end-to-end multi-person pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15029–15038, 2023.

Human Dataset Collection Agreement

Organizers and our affiliation, are referred to be *Producer*, ____________ (Name) is referred to be *Participant*. Producer promise to abide by the "Personal Information Protection Law" and other relevant laws on personal information, and the collected data will only be used for non-commercial use, including but not limited to academic research. Producer will anonymize the data before it is release. When providing collected data to the community, Producer will require users to sign and abide by the relevant data use agreement (See **Attachment 1**).

The subsequent information regarding the Participant will be gathered and provided to the appropriate academic community:

- Human body poses
- Human tracking
- Human video

Producer promise that only the above information will be released, and other personal information of Participant will not be disclosed. For violations of the data use agreement, Producer will terminate corresponding data use rights.

By signing this agreement, Participant understand and accept the statements regarding data collection and subsequence data use. We sincerely appreciate your cooperation and support in this matter.

(Date) (Name) (Signature)

Page 1 of 1

Attachment 1

Dataset Release Agreement

Before downloading and using our dataset, please agree to the following terms. Applicant, their employers and affiliation are referred to as *User*; Authors and their affiliation are referred to as *Producer*.

1. This dataset is provided solely for non-commercial research purposes.
2. The User agrees not to engage in the reproduction, duplication, sale, trade, resale, or exploitation of any portion of the images or any derived data for commercial endeavors.
3. Producer is NOT responsible for any further use in a defamatory, pornographic, or any other unlawful manner, or violation of any applicable regulations or laws.
4. User agrees NOT to further copy, publish, or distribute any portion of this dataset to any third party for any purpose.
5. This agreement effectively for any potential User of this dataset upon the date that User first access this data in any form.
6. Producer reserves the right to terminate User's access to this dataset at any time.

Figure D2: **Left:** Agreements for data collection involving participants. **Right:** Illustration of users' responsibilities and usage agreement for participants.

EHPT-XC Dataset Usage Agreement

In this agreement, applicant, their employers and affiliation are referred to as *User*; Authors and their affiliation are referred to as *Producer*. Before downloading and using our dataset, **user declares that User agrees to the following terms:**

- The dataset is restricted to **non-commercial uses**, including teaching, academic research, public demonstrations, and personal experimentation. User is prohibited from using this dataset or any derivative works for commercial purposes in any form. Examples of commercial purposes include business operations, licensing, leasing, selling the dataset, distributing it with commercial products, using it to create or utilize commercial products, or any activity aimed at generating commercial profit.
- The dataset **must not be copied, shared, distributed, resold, offered for resale, transferred, or sub-licensed**, in whole or in part, except that you may make one copy solely for archival purposes.
- Producer is NOT responsible for any further use in a defamatory, pornographic, or any other unlawful manner, or violation of any applicable regulations or laws.
- This agreement effectively for any potential User of this dataset upon the date that User first access this data in any form.
- Producer reserves the right to terminate User's access to this dataset at any time.

Name: ____________________

Affiliation: ____________________

Address: ____________________

____

Email: ____________________

Telephone: ____________________

(Date) (Name) (Signature)

Figure D3: Dataset usage agreement for users.

[6] L. Sun, C. Sakaridis, J. Liang, Q. Jiang, K. Yang, P. Sun, Y. Ye, K. Wang, and L. V. Gool. Event-based fusion for motion deblurring with cross-modal attention. In *European conference on computer vision*, pages 412–428. Springer, 2022.

[7] D. Wang and S. Zhang. Contextual instance decoupling for robust multi-person pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11060–11068, 2022.

[8] Z. Wang, H. Zhao, Y.-L. Li, S. Wang, P. Torr, and L. Bertinetto. Do different tracking tasks require different appearance models? *Advances in neural information processing systems*, 34:726–738, 2021.

[9] J. Yang, A. Zeng, S. Liu, F. Li, R. Zhang, and L. Zhang. Explicit box detection unifies end-to-end multi-person pose estimation. *arXiv preprint arXiv:2302.01593*, 2023.

[10] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*, pages 1–21. Springer, 2022.