

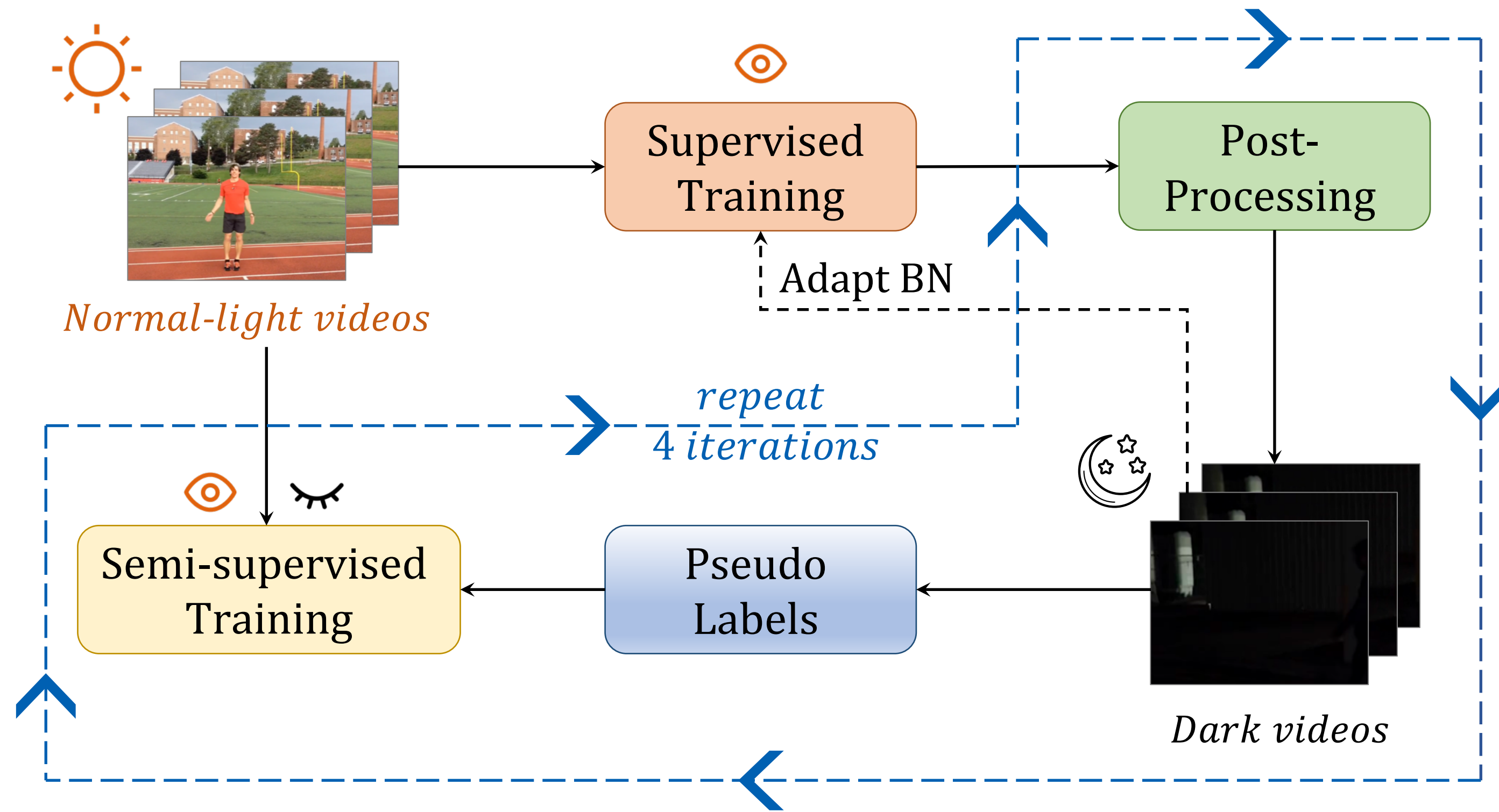
The 2nd Place Solution for UG2+ Challenge 2022: Semi-Supervised Action Recognition in the Dark

Kunchang Li^{1,2*}, Kaibin Tian^{1*}, Jingjie Shang^{1*}, Haisheng Su^{1†}, Yangguang Li¹

1 SenseTime Research 2 Shanghai AI Laboratory

(*Equally-contributed first authors, †Corresponding author)

Introduction



The training process can roughly be divided into two stages:

- **Supervised Training** for training the basic models;
- **Pseudo-label Iterations** for generating accurate pseudo labels progressively.

Besides, we propose effective **post-processing** for model ensemble.

We finally achieve **90.7%** top-1 accuracy for extreme dark video classification in UG2+ Challenge 2022.

Supervised Training

In supervised training, we fine-tune four different models:

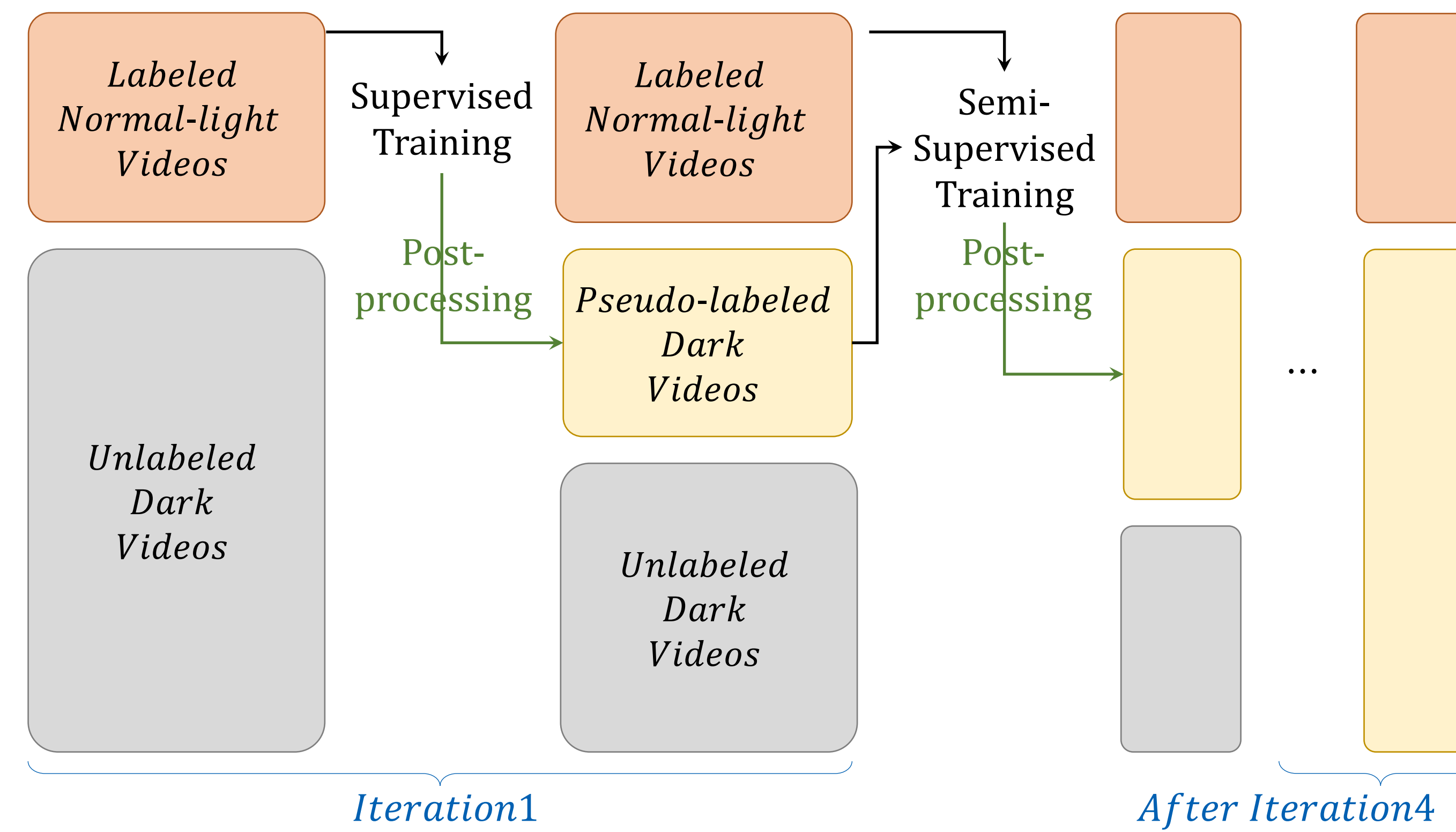
- **UniFormer-B32** pre-trained with Kinetics-600 [1];
- **UniFormer-B32** pre-trained with Something-Something V2 [1];
- **MViT-B32** pre-trained with Kinetics-600 [2];
- **SlowFast-R101** pre-trained with Kinetics-700 [3].

All the models with **32 frames** sampled uniformly from the videos by default, because **uniform sampling** is better than dense sampling.

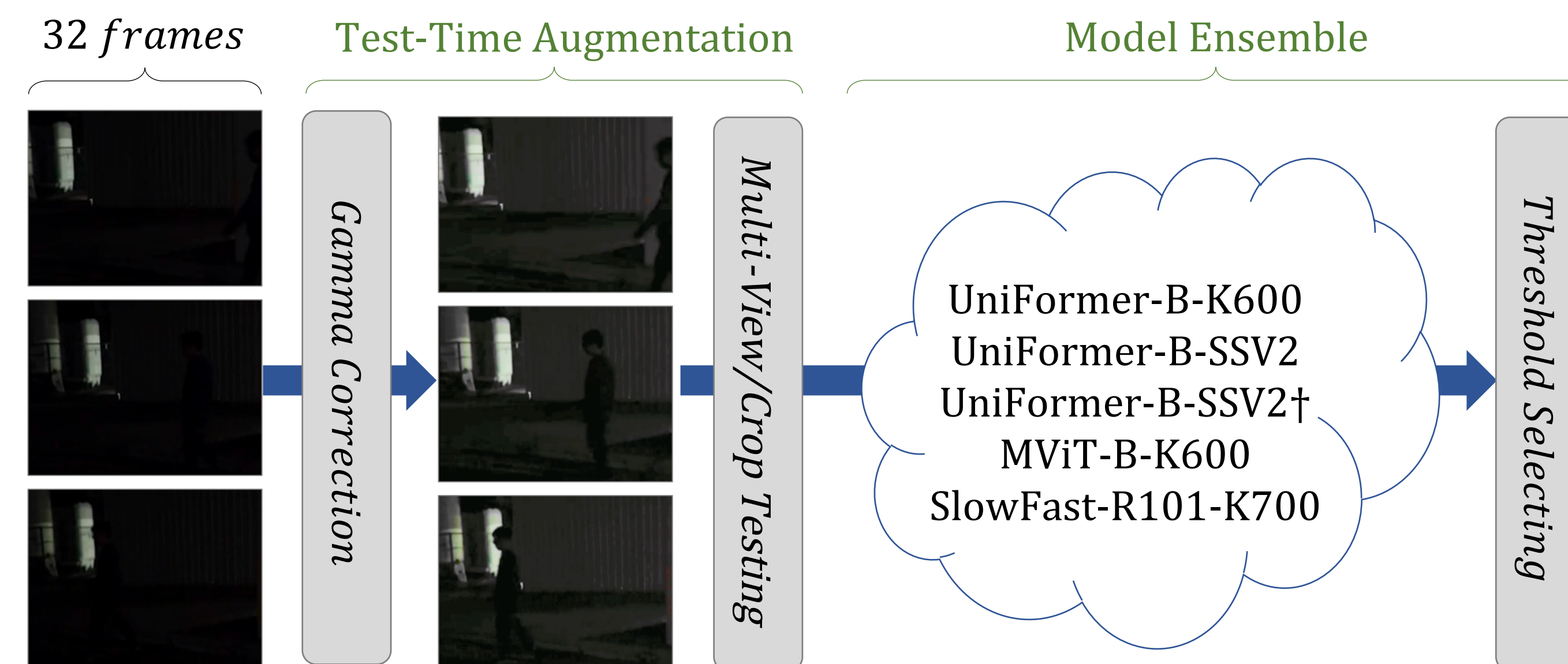
For better recognition in the dark, we use the unlabeled dark videos to **adapt** the parameters in **BatchNorm** layers.

For the following pseudo-label iterations, we simply adopt the same hyperparameters as used in supervised training.

Pseudo-label Iterations

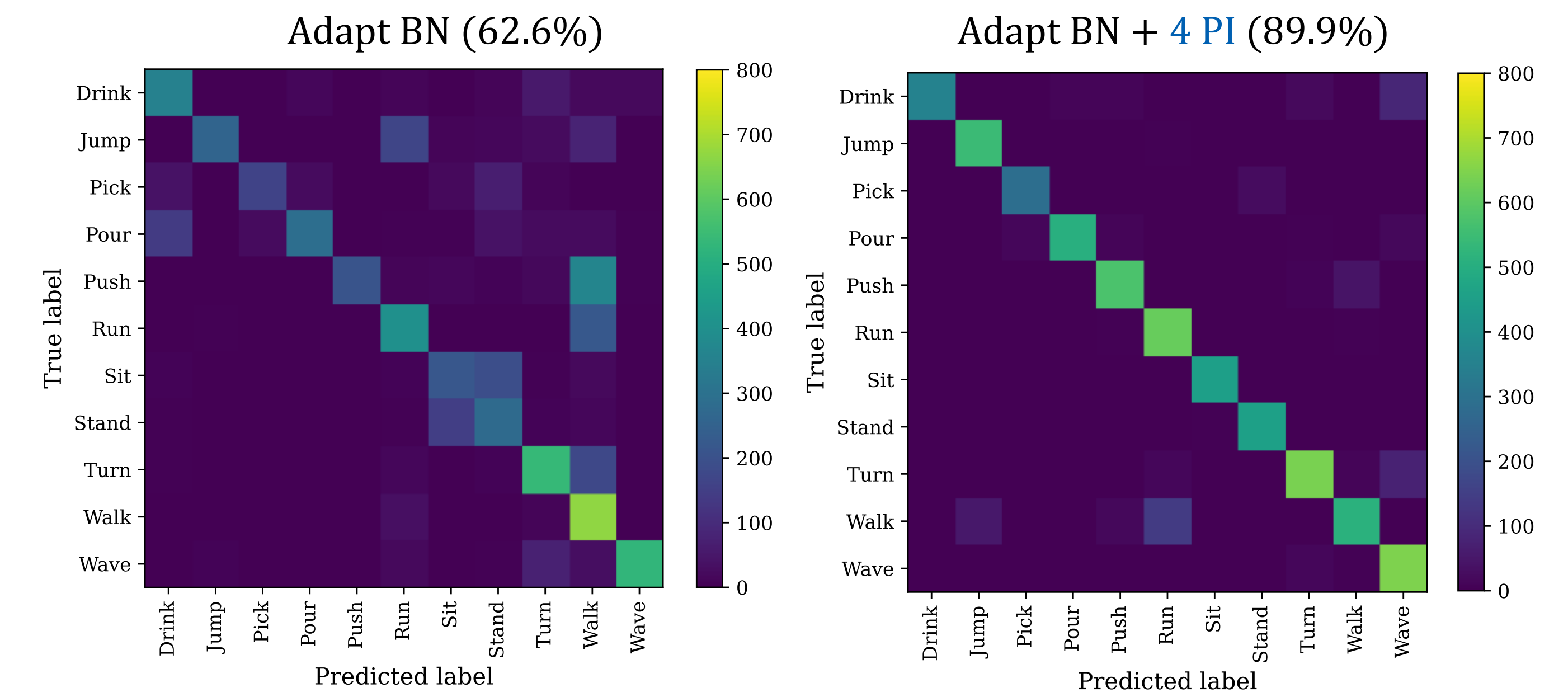


Post-processing



Method	Gamma	#view	#crop	ARID Top-1
UniFormer B32-K600 + Adapt BN	×	1	1	60.5%
	√	1	1	62.6% (+2.1%)
	√	1	3	62.8% (+2.3%)
	√	3	1	62.7% (+2.2%)
	√	3	3	62.9% (+2.4%)

Results



Method	ARID Top-1
UniFormer-B32-K600	39.3%
+ Adapt BN	62.6% (+23.3%)
+ Adapt BN + 1 Pseudo-label Iteration	83.5% (+42.2%)
+ Adapt BN + 2 Pseudo-label Iterations	87.8% (+48.5%)
+ Adapt BN + 3 Pseudo-label Iterations	89.5% (+50.2%)
+ Adapt BN + 4 Pseudo-label Iterations	89.9% (+50.6%)
+ Adapt BN + 4 Pseudo-label Iteration + Other Models Ensemble	91.1% (+51.8%)

Code & References

Code is available at <https://github.com/Andy1621/Nightcrawler>. ★

- [1] Li K, Wang Y, Gao P, et al. Uniformer: Unified Transformer for Efficient Spatiotemporal Representation Learning[C]//International Conference on Learning Representations (ICLR). 2021.
- [2] Fan H, Xiong B, Mangalam K, et al. Multiscale vision transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021.
- [3] Feichtenhofer C, Fan H, Malik J, et al. Slowfast networks for video recognition[C]//Proceedings of the IEEE/CVF international conference on computer vision (ICCV). 2019.
- [4] Xu Y, Yang J, Cao H, et al. Arid: A new dataset for recognizing action in the dark[C]//International Workshop on Deep Learning for Human Activity Recognition, 2021.