

TEMPORAL COHERENT TEST-TIME OPTIMIZATION FOR ROBUST VIDEO CLASSIFICATION - SUPPLEMEN- TARY

Anonymous authors

Paper under double-blind review

1 MINI KINETICS-C AND MINI SSV2-C SAMPLES

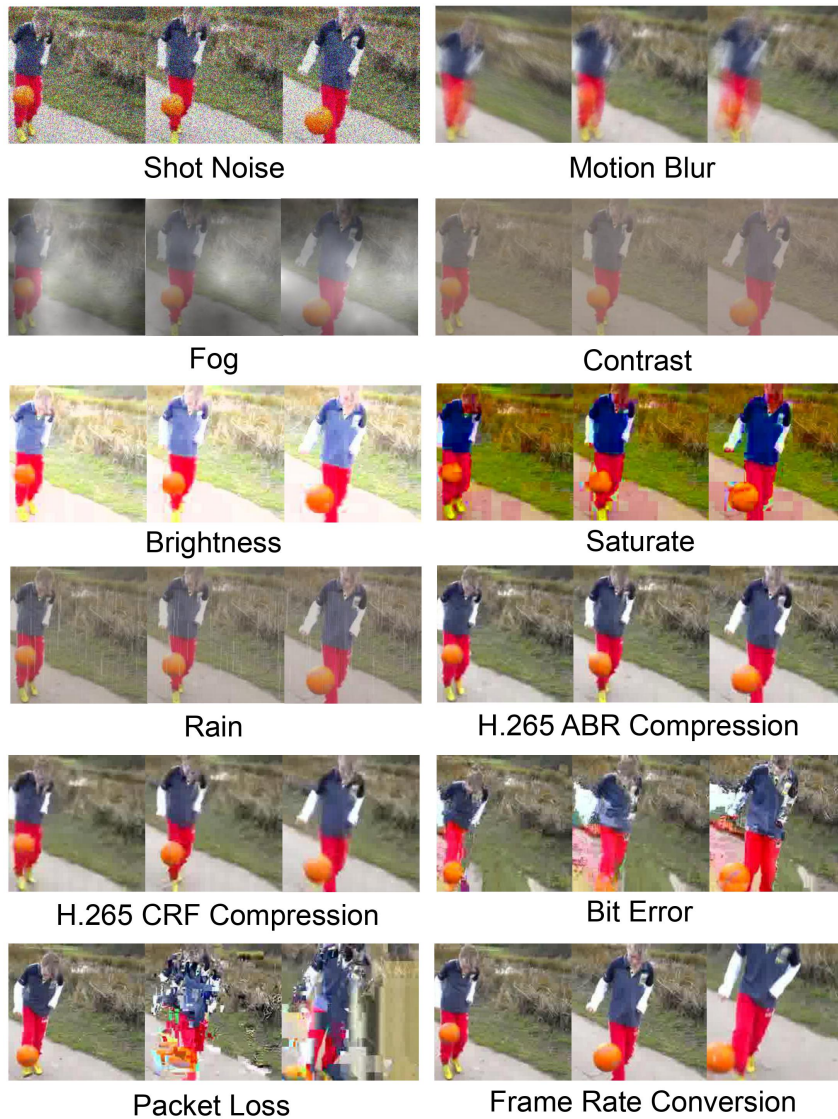


Figure 1: We show 12 types of corruptions at severity level 3 on Mini Kinetics-C. The sample images are created by following the setting in Yi et al. (2021).

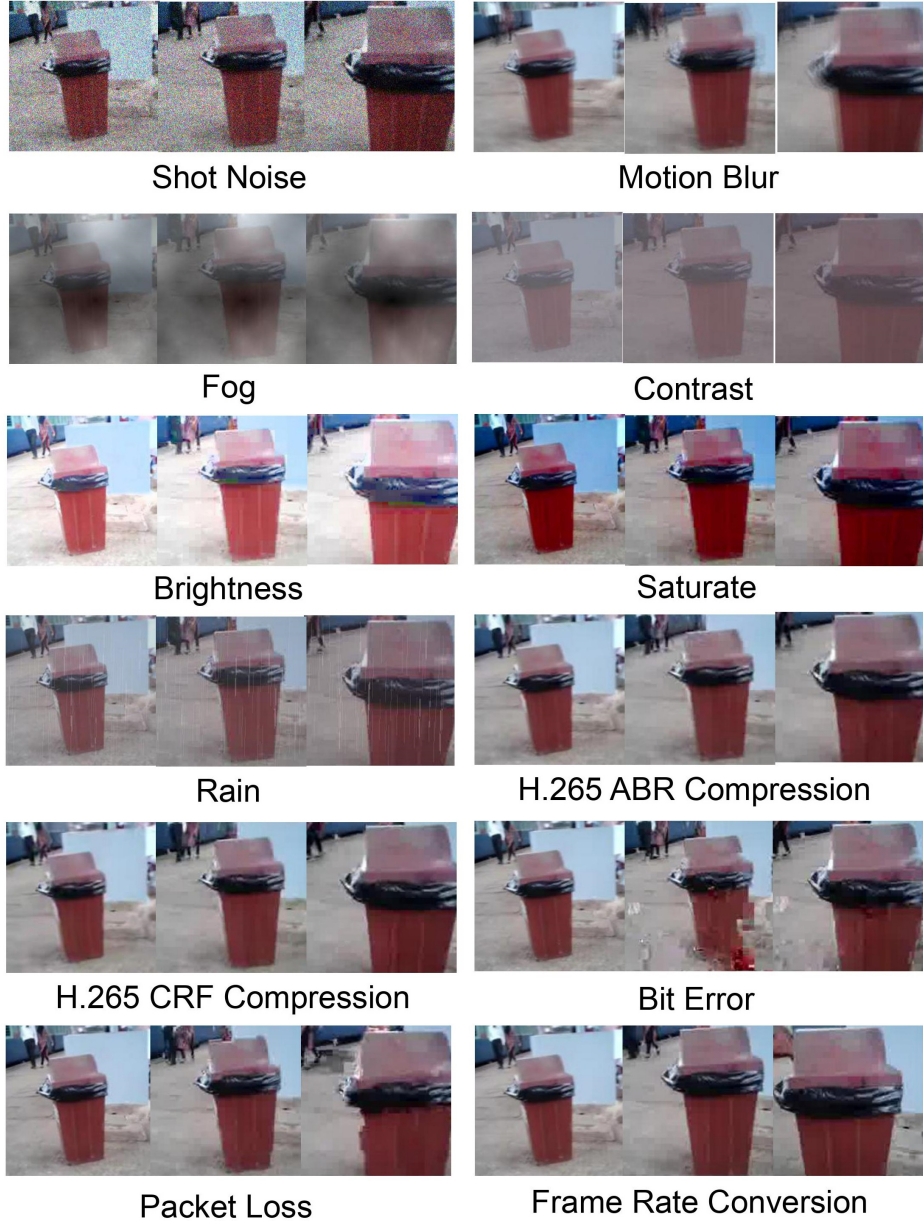


Figure 2: Similarly, we show the corrupted video frames in Mini SSV2-C.

2 FULL EXPERIMENTAL RESULTS ACROSS BACKBONES

We show the accuracy of various models on 12 types of corruptions at 5 levels of severities in Figure 3, Figure 4, Figure 5 and Figure 6. TeCo outperforms other baseline methods in most cases. Figure 3 and Figure 5 show that SHOT performs well when encountering corruptions like noise, rain, and fog on both datasets. Its performance deteriorates when facing compression, bit error, packet loss, and frame rate conversion. These corruptions disturb the temporal information and require the model to be generalizable as well (Yi et al., 2021). We hypothesize that SHOT may fail when the model is overfitting to the pseudo label. Similarly, Liu et al. (2021) found that models in test-time training can yield overfitting to the auxiliary task and their performance will deteriorate. In Figure 4, we find the performance of Tent drops significantly when the corruption becomes more severe. Figure 6 shows that test-time optimizations which update network parameters fail on videos

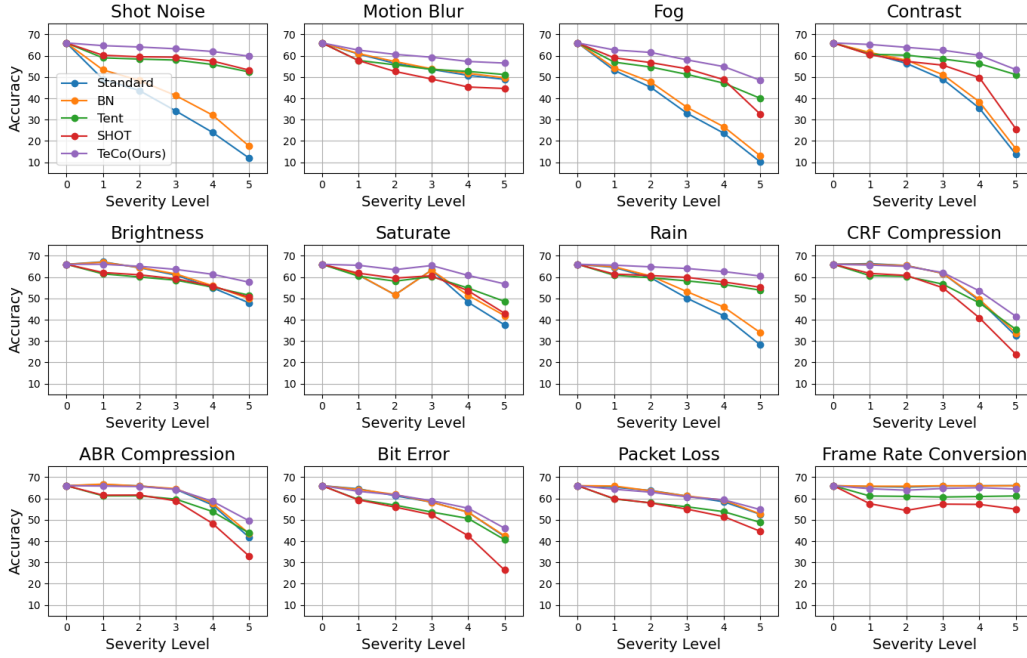


Figure 3: Full results on Mini Kinetics-C, with a backbone of ResNet18.

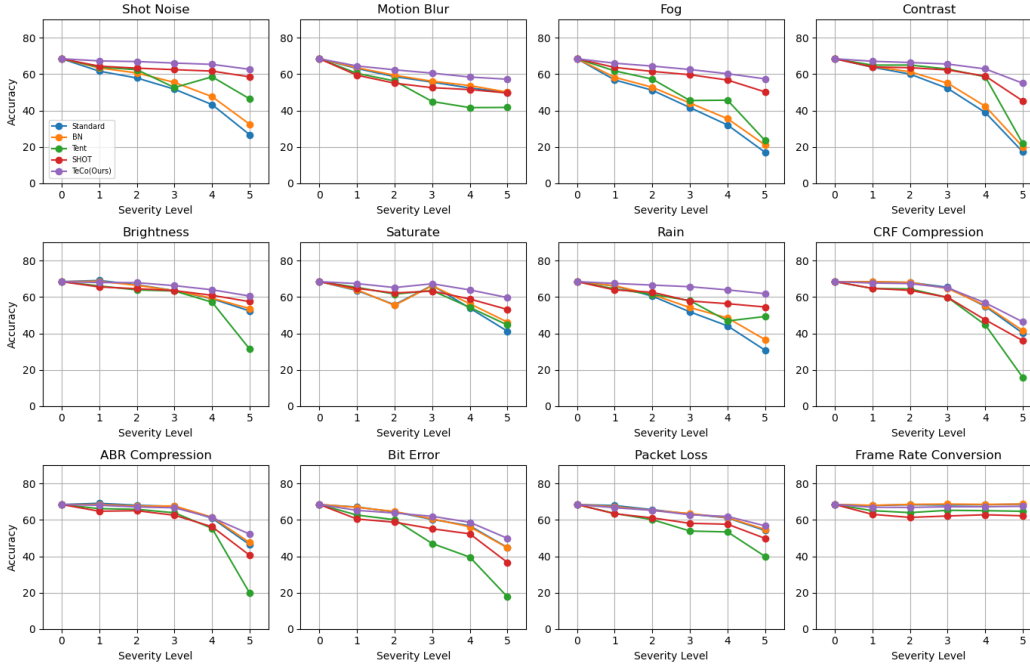


Figure 4: Full results on Mini Kinetics-C, with a backbone of TAM-ResNet18.

corrupted by motion blur and bit error, while standard and BN methods obtain better performance. It implies the limitation of the test-time optimization methods. There remains an open question that when will the methods thrive and fail.

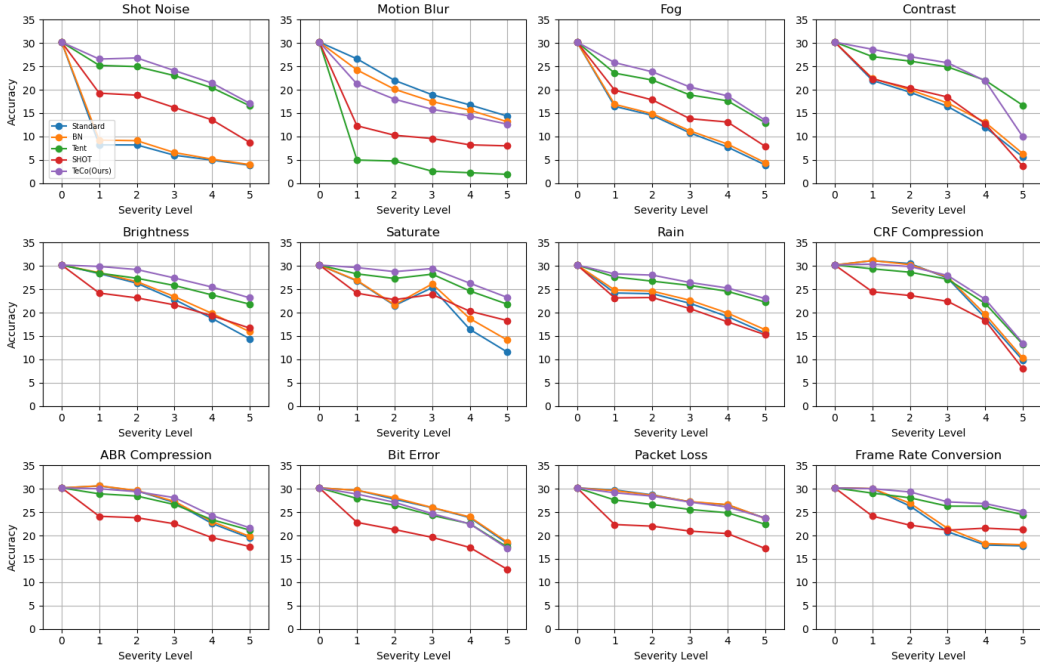


Figure 5: Full results on Mini SSV2-C, with a backbone of ResNet18.

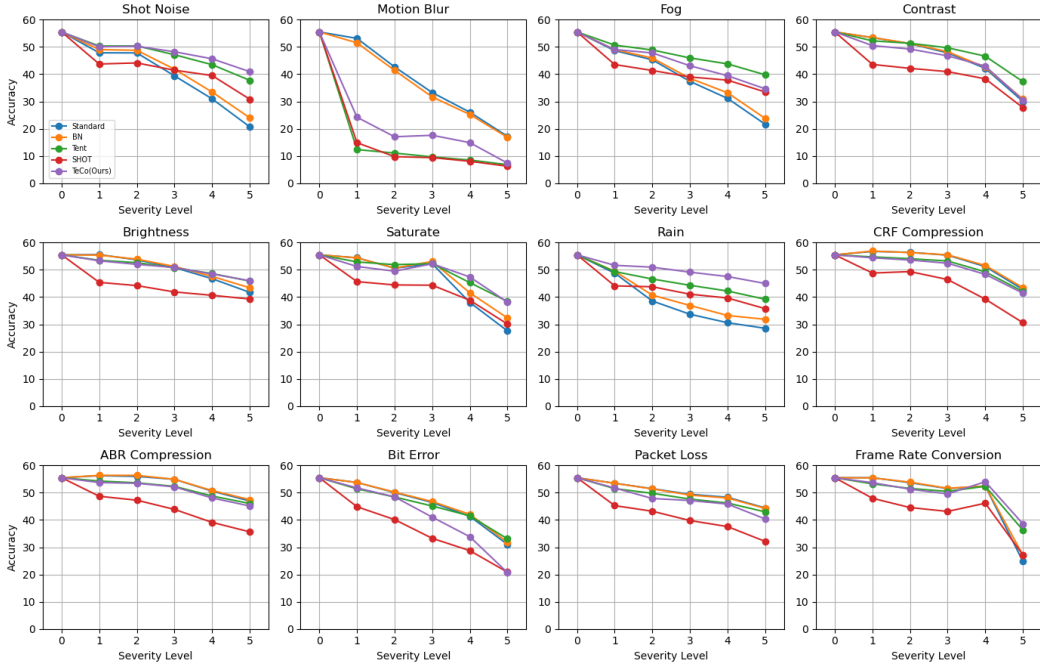


Figure 6: Full results on Mini SSV2-C, with a backbone of TAM-ResNet18.

REFERENCES

Yuejiang Liu, Parth Kothari, Bastien van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt+: When does self-supervised test-time training fail or thrive? In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Ad-*

vances in Neural Information Processing Systems, volume 34, pp. 21808–21820. Curran Associates, Inc., 2021. URL <https://proceedings.neurips.cc/paper/2021/file/b618c3210e934362ac261db280128c22-Paper.pdf>.

Chenyu Yi, Siyuan Yang, Haoliang Li, Yap-peng Tan, and Alex Kot. Benchmarking the robustness of spatial-temporal models against corruptions. In *Advance in Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.