# Multi-granularity Generator for Temporal Action Proposal –Supplementary Material

Yuan Liu[♯*] Lin Ma[♮†] Yifeng Zhang[♯†] Wei Liu[♮] Shih-Fu Chang[§]
[♮]Tencent AI Lab [♯]Southeast University [§]Columbia University
{lhy19930911,forest.linma}@gmail.com yfz@seu.edu.cn {wl2223,sc250}@columbia.edu

This supplementary material includes additional experiments that are not presented in the main paper and more qualitative results to demonstrate the performances of our proposed MGG.

**Recall Rates.** Ground-truth proposals with short temporal spans are hard to capture, which mainly dues to that short proposals are of less semantic information. In the main paper, we illustrate the improvement of recall rates in short proposals with the U-shape architecture. Here, we demonstrate the short proposal recall rates of different methods including DAP [2], TURN [4], CTAP [3], BSN [6] and our proposed MGG on the testing set of THUMOS-14. The temporal spans ranges from 1 frame to 60 frames, and the recall rates are computed with AN and tIoU set to 100 and 0.75, respectively. As shown in Table 1, the recall rates of MGG outperform the other competitor methods. One reason is that the temporal boundary adjustment (TBA) module is helpful for the proposal to be accurate in boundaries. Thus the generated proposals will have high overlap with ground-truths. Another reason is the U-shape architecture, which provides high-level semantic information for lower layers and helpful for the capture of proposals with short temporal durations.

**Qualitative Results.** More qualitative results are illustrated in Fig. 1. The first four rows are videos from the validation set of ActivityNet-1.3 [1] and the last two rows are from the testing set of THUMOS-14 [5]. It can be observed that the refined proposals are of higher accuracy, which demonstrates the effectiveness of the proposed MGG. Some failure cases are shown in Fig. 2. For ground-truth proposals with short temporal durations, false negatives are produced. Moreover, if the videos are of low quality, it will be hard to capture the corresponding semantic meanings and thereby result in wrong proposals.

Table 1: Recall rates of different methods on generated proposals with short temporal extents on testing set of THUMOS-14, where AN and tIoU thresholds are set to 100 and 0.75, respectively.

| Method | 1-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| DAP [2] | 0.000 | 0.000 | 0.025 | 0.047 | 0.097 | 0.106 |
| TURN [4] | 0.000 | 0.000 | 0.043 | 0.117 | 0.174 | 0.370 |
| CTAP [3] | 0.000 | 0.000 | 0.043 | 0.126 | 0.267 | 0.357 |
| BSN+NMS [6] | 0.000 | 0.048 | 0.168 | 0.237 | 0.339 | 0.400 |
| MGG | 0.000 | **0.081** | **0.183** | **0.296** | **0.364** | **0.431** |

*This work was done while Yuan Liu was a Research Intern with Tencent AI Lab.
†Corresponding authors.

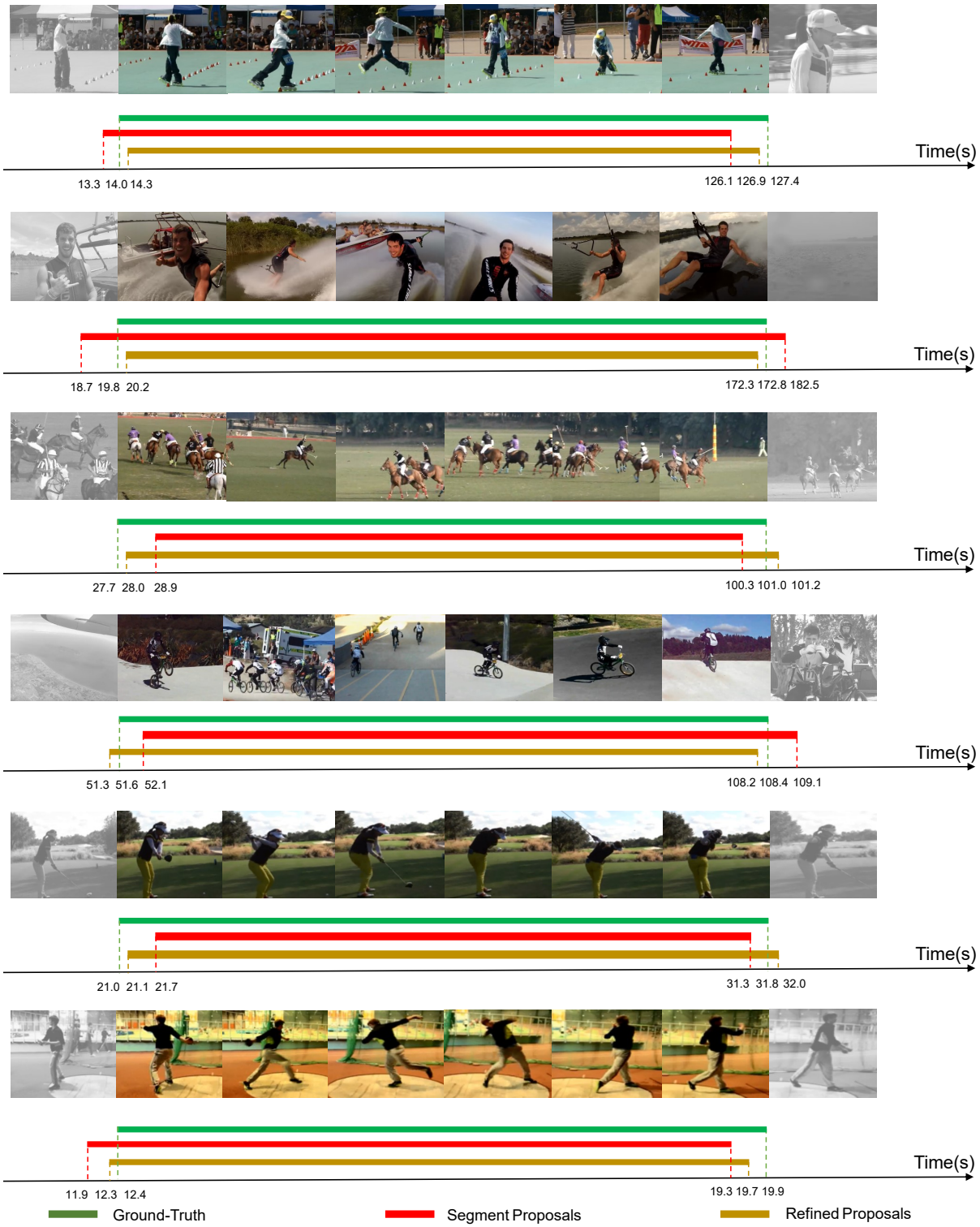


Figure 1: Qualitative results of proposals generated by MGG. First four rows represent temporal proposals on ActivityNet-1.3. Last two rows represent temporal proposals on THUMOS-14. After TBA adopted to adjust proposal boundaries generated by segment proposal generator (SPG), the refined proposals will have high overlap with the ground-truth proposals.

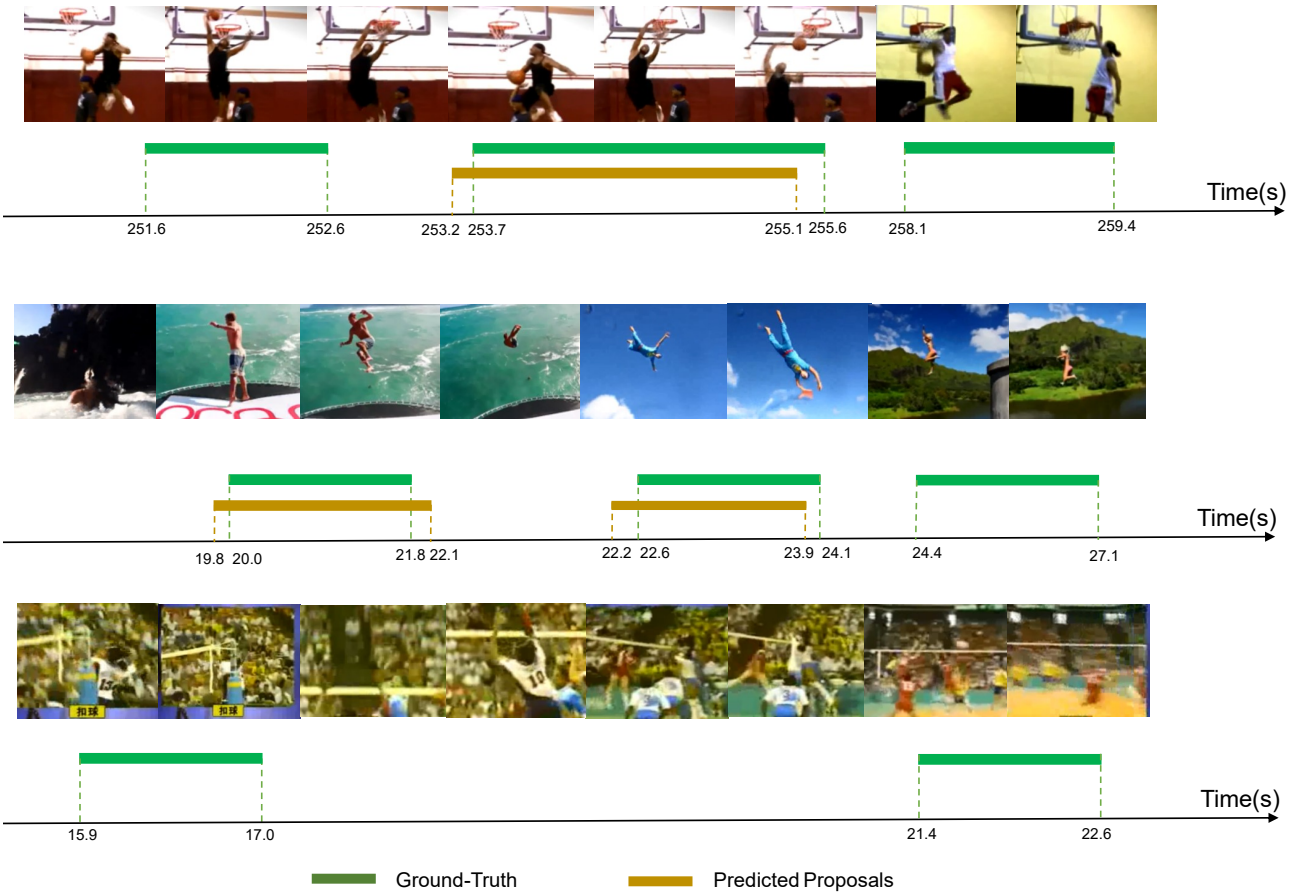


Figure 2: Failure cases generated by MGG on THUMOS-14. For ground-truths with short temporal spans (first two rows), it is challenging for MGG to locate them. While quality of video frames is poorer (the last row), the performance will be reduced further.

## References


[1] F. Caba Heilbron, V. Escorcia, B. Ghanem, and J. Carlos Niebles. Activitynet: A large-scale video benchmark for human activity understanding. In *CVPR*, pages 961–970, 2015.

[2] V. Escorcia, F. C. Heilbron, J. C. Niebles, and B. Ghanem. Daps: Deep action proposals for action understanding. In *ECCV*, pages 768–784, 2016.

[3] J. Gao, K. Chen, and R. Nevatia. Ctap: Complementary temporal action proposal generation. *arXiv preprint arXiv:1807.04821*, 2018.

[4] J. Gao, Z. Yang, C. Sun, K. Chen, and R. Nevatia. Turn tap: Temporal unit regression network for temporal action proposals. In *ICCV*, pages 3648–3656, 2017.

[5] Y.-G. Jiang, J. Liu, A. Roshan Zamir, G. Toderici, I. Laptev, M. Shah, and R. Sukthankar. THUMOS challenge: Action recognition with a large number of classes. In *ECCV Workshop*, 2014.

[6] T. Lin, X. Zhao, H. Su, C. Wang, and M. Yang. Bsn: Boundary sensitive network for temporal action proposal generation. *arXiv preprint arXiv:1806.02964*, 2018.