Supplementary Materials for Edit-A-Video: Single Video Editing with Object-Aware Consistency

Chaehun Shin* CHAEHUNY@SNU.AC.KR
Heeseung Kim* GMLTMD789@SNU.AC.KR
Che Hyun Lee SAGA1214@SNU.AC.KR
Sang-gil Lee TKDRLF9202@SNU.AC.KR
Data Science and AI Lab, ECE, Seoul National University, Seoul 08826, Korea
Sungroh Yoon† SRYOON@SNU.AC.KR
Data Science and AI Lab, ECE and Interdisciplinary Program in AI, Seoul National University, Seoul 08826, Korea

Editors: Berrin Yanıkoğlu and Wray Buntine

Appendix A. Human Preference Study

We conduct two distinct human preference studies. The first one involves evaluating the overall video editing quality of our model and baselines. We collect feedback from 62 participants who are asked to rate scores on a 5-point scale from 1-5 by taking into account three key factors as follows:

- **Background Preservation** Edited video preserves unedited details of the original video.
- **Text Alignment** Edited video matches the target edit description provided.
- **Video Realism** The overall visual quality and smoothness of the edited video.

In the second study, we measure the effectiveness of our proposed TC Blending technique in terms of the *background inconsistency problem*. We check this by assessing the degree of preservation and consistency of the background in the presence or absence of the TC Blending technique. Once again, we ask 49 participants to provide feedback, with scores related to background preservation.

Appendix B. Additional Samples

We show several additional samples in this section. First, various examples of our model are in Fig. 1, 2. We also compare the generated samples of our model and several baselines in Fig. 3. In addition, samples according to the adjustment of various hyperparameters of attention injection are shown in Fig. 4.

Figure 1: **Qualitative Results** Additional selected samples for our model.

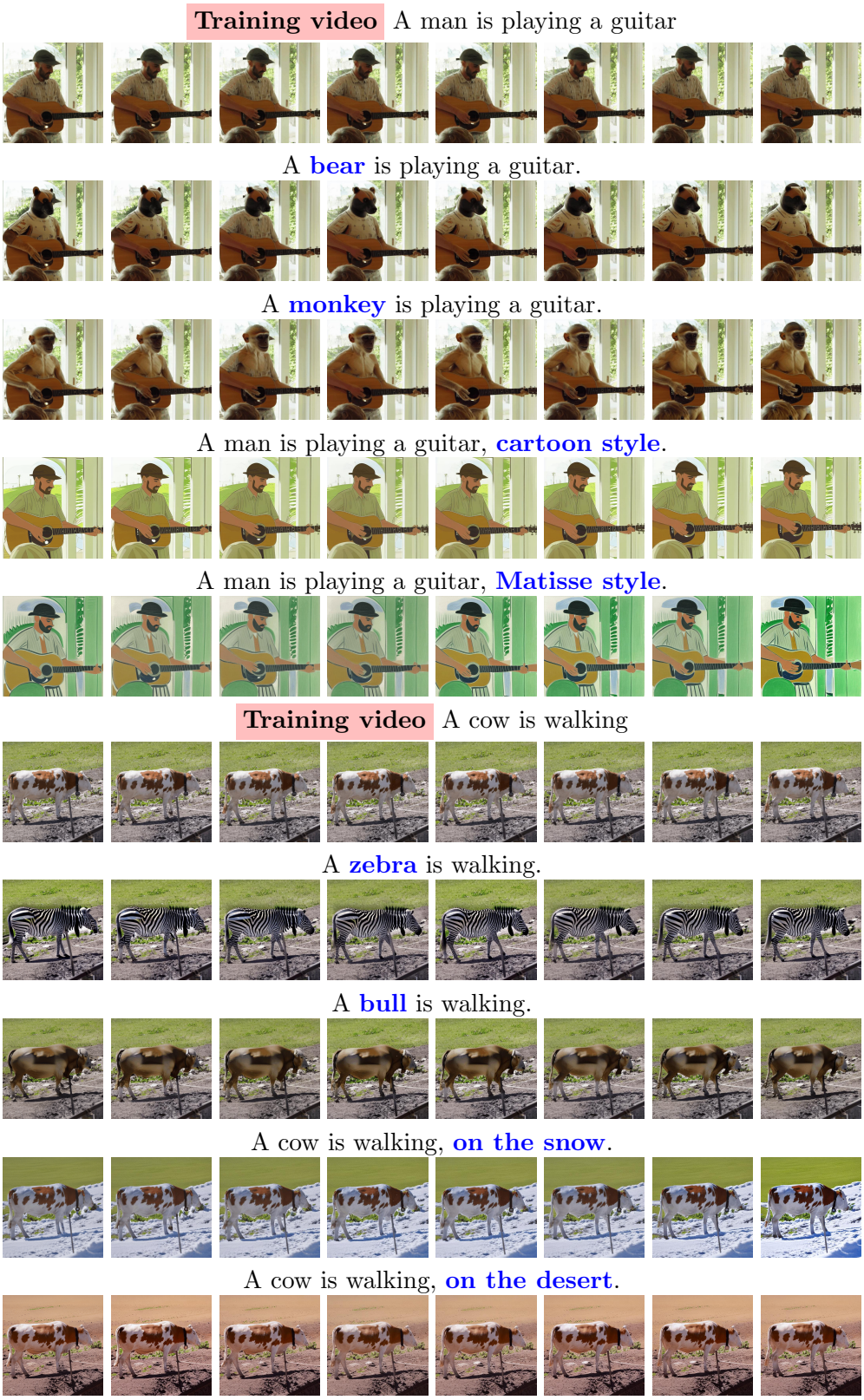

Figure 2: **Qualitative Results** Additional selected samples for our model.

Figure 3: **Baseline Comparison** Additional samples for our model and baselines.

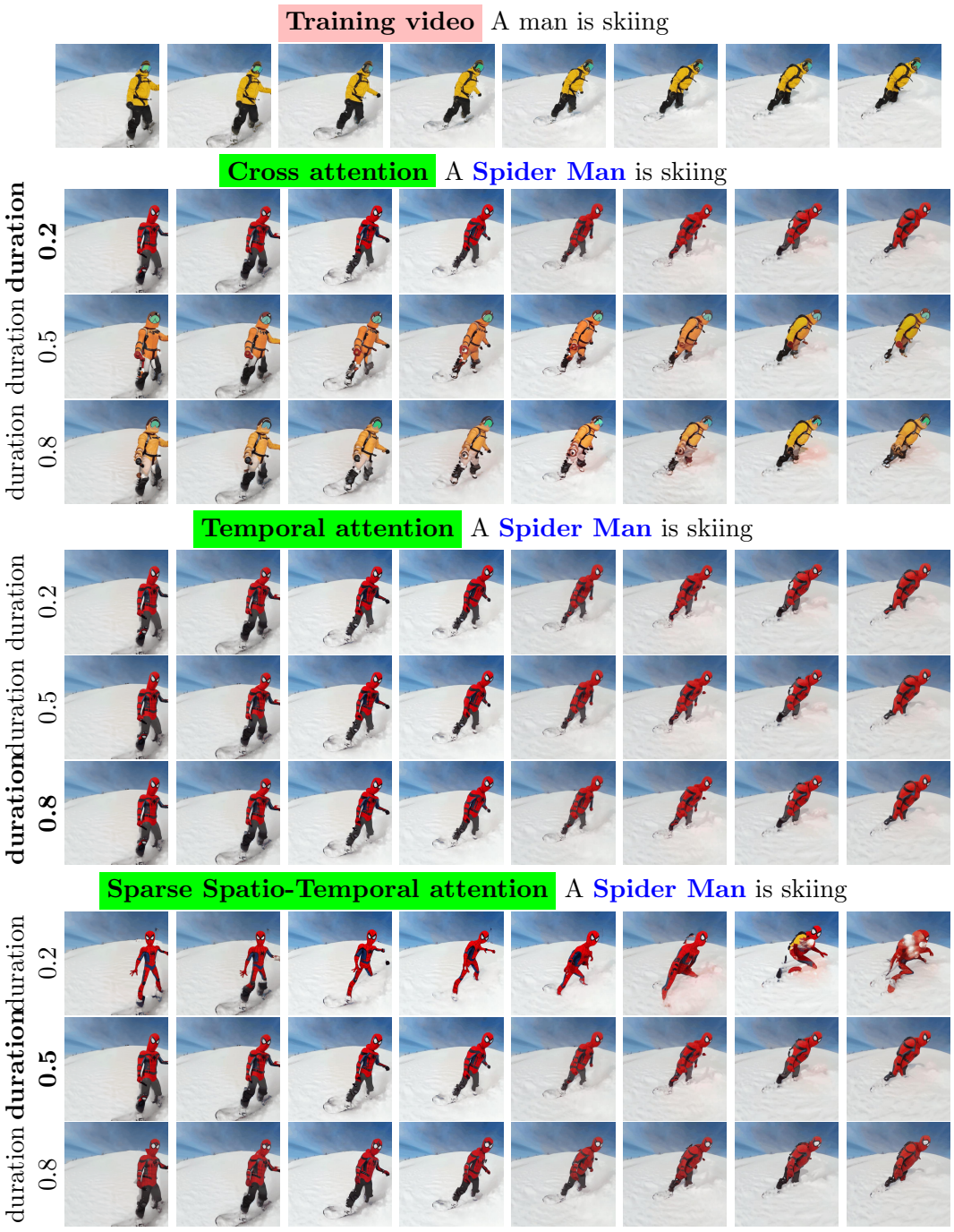

Figure 4: **Attention Injection Analysis** Samples according to attention injection hyper-parameters.