VirtualHome Action Genome: A Simulated Spatio-Temporal Scene Graph Dataset with Consistent Relationship Labels (Supplementary Material)

This supplementary material provides additional details of the proposed dataset VirtualHAG, including scripts for video generation, co-occurrence statistics for objects and relationships, and dataset examples. We also provide more experimental results on the proposed dataset VirtualHAG.

A. Additional Details on VirtualHAG Dataset

Scripts for Video Generation. We manually designed 108 unique scripts to generate videos in the VirtualHome simulator. We show eight script examples in Table 6. Each script consists of a sequence of actions, and each action contains an action type and related objects. During the execution of scripts, object instances are randomly instantiated.

Co-occurrence Statistics for Objects and Relationships. We show the co-occurrence statistics for objects and relationships of the VirtualHAG dataset in Figure 8. Relationship distribution tends to show relevance with object affordance. Such as, for small objects (*e.g*. "cupcake" class), the "touching", "holding", and "in" relationships dominate, while for "chair" and "bench", the "above" relationship is relatively more dominant than other relationships.

Dataset Examples. We show six dataset examples sampled from the VirtualHAG dataset in Figure 9 and Figure 10 ((a) to (f)). In each example, we show eight frames sampled from two different observation viewpoints.

B. Additional Experimental Results on VirtualHAG Dataset

We show three example results for frames and viewpoints sampled from the VirtualHAG dataset in Figure 11. The SGTracker exhibited high performance in localizing objects and humans and determining human-object relationships in these three examples. However, we also found that SGTracker is less accurate in predicting correct relationships around the action change moment (*e.g*. Figure 11 example (a) Frames 250 and 384), and there is room for SGTracker to improve in detecting small objects (*e.g*. Figure 11 example (c) Frames 142, 152, 395 and 420).

Script abstract	Scripts
Sit and stand up from a chair	[WALK] $\langle chair \rangle$ [SIT] $\langle chair \rangle$ [STANDINGUP] $\langle chair \rangle$ [WALK] $\langle bed \rangle$
Sit and stand up from a sofa	[WALK] $\langle sofa \rangle$ [SIT] $\langle sofa \rangle$ [STANDINGUP] $\langle sofa \rangle$ [WALK] $\langle bench \rangle$
Switch on and switch off a light switch	[WALK] $\langle lightswitch \rangle$ [SWITCHON] $\langle lightswitch \rangle$ [SWITCHOFF] $\langle lightswitch \rangle$ [WALK] $\langle chair \rangle$
Switch on and switch off a tv	[WALK] $\langle tv \rangle$ [SWITCHON] $\langle tv \rangle$ [SWITCHOFF] $\langle tv \rangle$ [WALK] $\langle table \rangle$
Open and close curtains	[WALK] $\langle curtains \rangle$ [OPEN] $\langle curtains \rangle$ [CLOSE] $\langle curtains \rangle$ [WALK] $\langle toilet \rangle$
Grab, drink, and put a cup of milk back	[WALK] $\langle milk \rangle$ [GRAB] $\langle milk \rangle$ [DRINK] $\langle milk \rangle$ [PUTOBJBACK] $\langle milk \rangle$ [WALK] $\langle bed \rangle$
Grab and put a cupcake in a fridge, and close the fridge	[WALK] $\langle cupcake \rangle$ [GRAB] $\langle cupcake \rangle$ [WALK] $\langle fridge \rangle$ [OPEN] $\langle fridge \rangle$ [PUTIN] $\langle cupcake \rangle$ $\langle fridge \rangle$ [CLOSE] $\langle fridge \rangle$ [WALK] $\langle table \rangle$
Grab a breadslice, heat the breadslice, and put the breadslice in a plate	[WALK] $\langle breadslice \rangle$ [GRAB] $\langle breadslice \rangle$ [WALK] $\langle toaster \rangle$ [PUTIN] $\langle breadslice \rangle$ $\langle toaster \rangle$ [SWITCHON] $\langle toaster \rangle$ [SWITCHOFF] $\langle toaster \rangle$ [GRAB] $\langle breadslice \rangle$ [WALK] $\langle plate \rangle$ [PUTIN] $\langle breadslice \rangle$ $\langle plate \rangle$

Table 6. Script examples used in the VirtualHAG dataset. The action types and objects are circle in "[]", "$\langle\rangle$", separately.

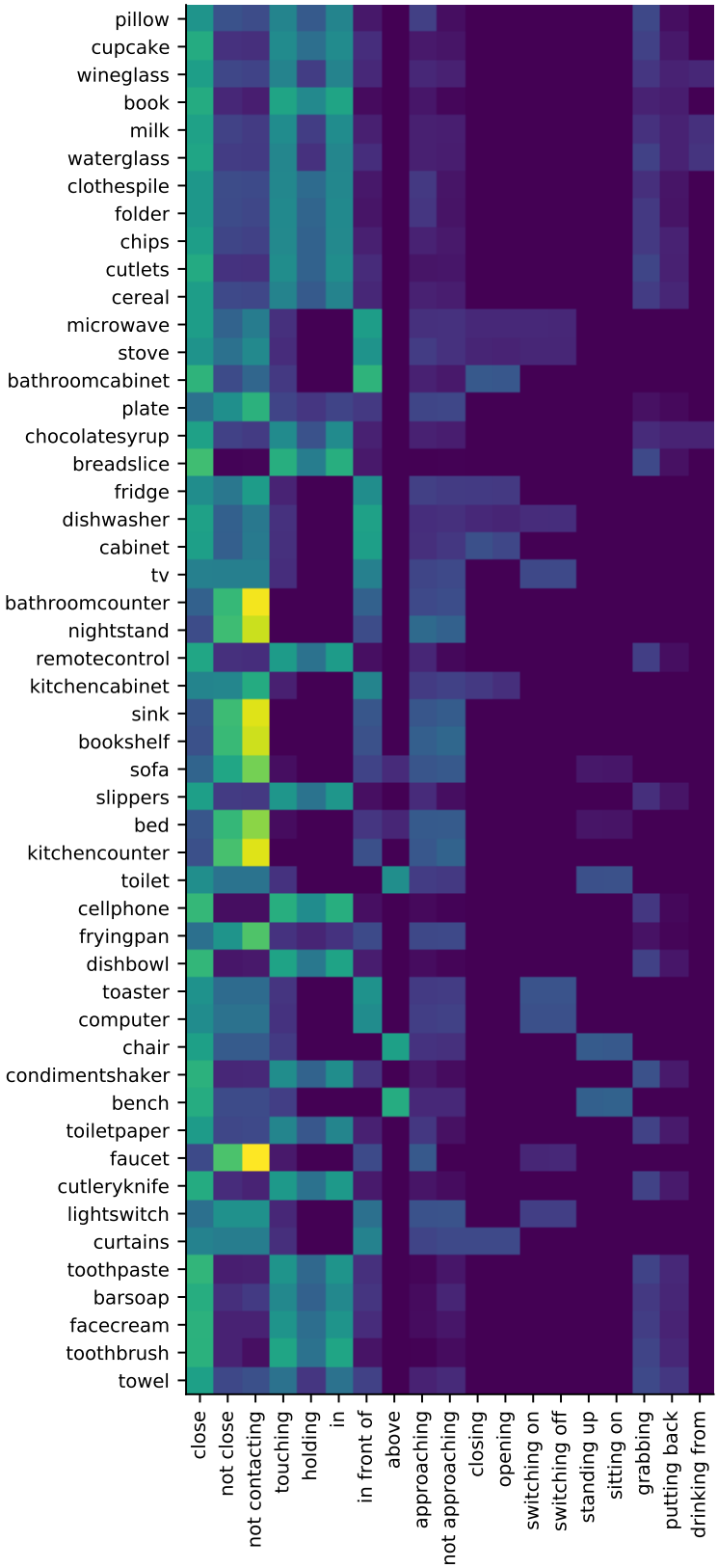

Figure 8. The co-occurence statistics for objects and relationships in the VirtualHAG dataset.

Figure 9. Three examples sampled from the VirtualHAG dataset.

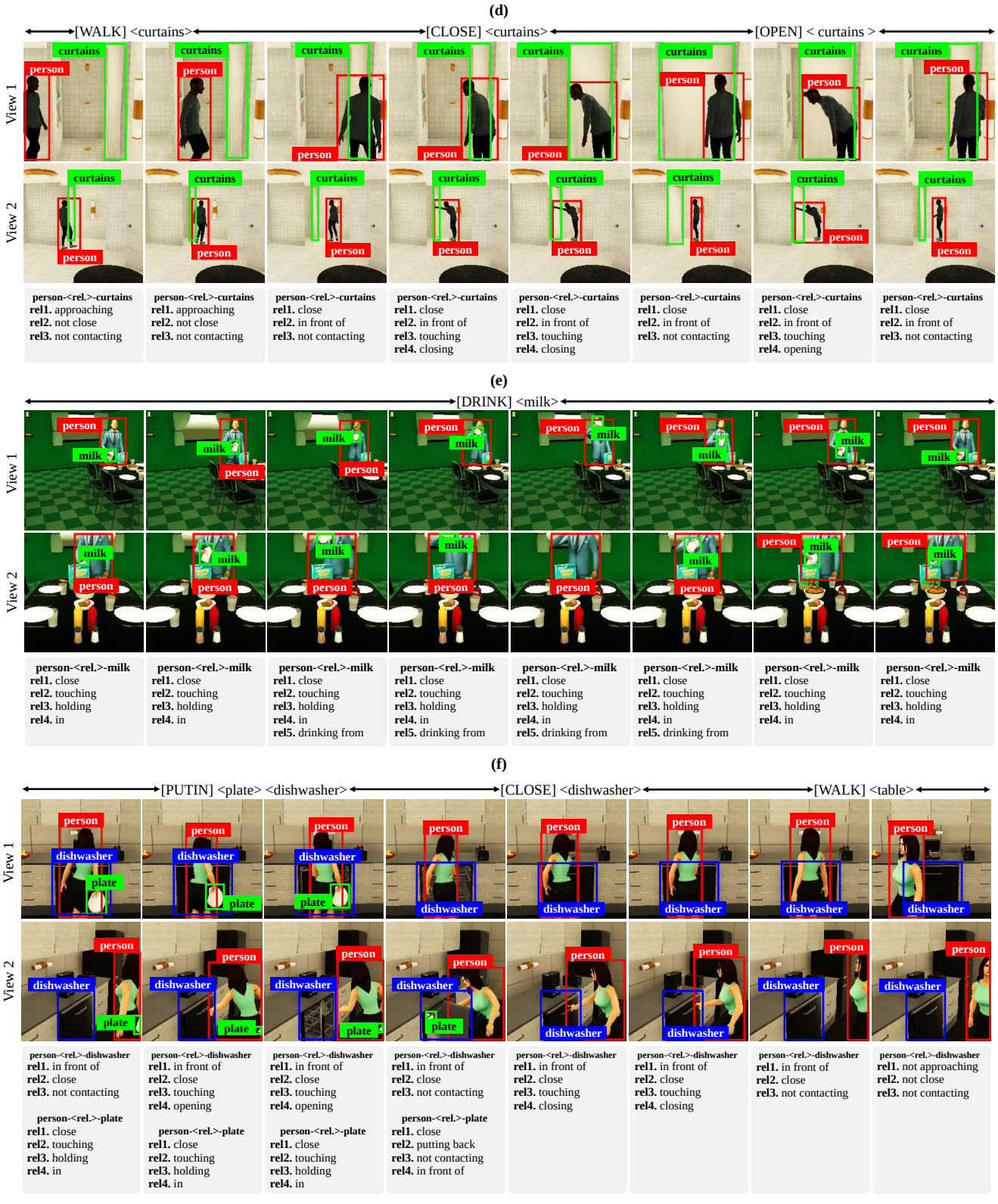

Figure 10. Three examples sampled from the VirtualHAG dataset.

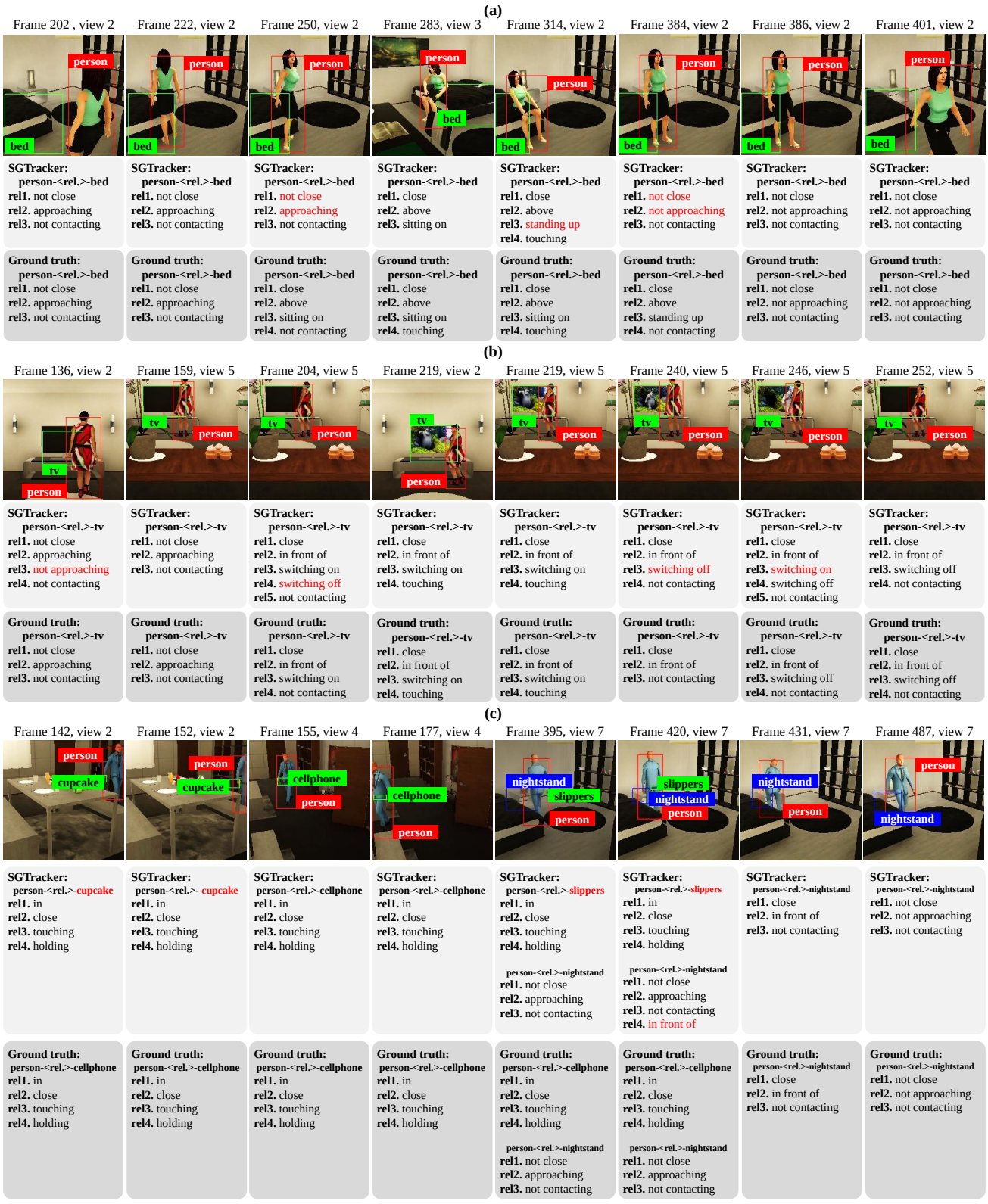

Figure 11. Example results on the VirtualHAG dataset. The incorrect predictions are highlighted in red.