Discovering Multi-Label Actor-Action Association in a Weakly Supervised Setting

Sovan Biswas[1] and Juergen Gall[1]

University of Bonn, 53115 Bonn, Germany
{biswas,gall}@iai.uni-bonn.de

1 Impact of HGRNN for Actor-Association:

Table 1. Results of various actor-action assignment approaches using MIML logits and MIML+HGRNN logits for label assignment. The HGRNN improves the label assignment.

Actor-action association	Backbone	Val-mAP
MIML	Slowfast	21.8
Proposed Approach w/o HGRNN	Slowfast	22.5
MIML+HGRNN	Slowfast	23.1
Proposed Approach	Slowfast	25.1

In Table 1, we analyze the impact of HGRNN for weakly supervised action detection. In case of "Proposed Approach w/o HGRNN", actor-action assignment is performed with MIML logits and HGRNN is not used at any stage. The proposed actor-action assignment improves the result in both cases. In case of HGRNN, the improvement is more due to HGRNN's ability to incorporate spatial and temporal information as seen in Table 1.

2 Comparison of MIML with proposed method:

Fig. 1 shows the per class mAP for all 60 classes on the validation set. It is an extension of Figure 4 of the main manuscript. The proposed approach improves the accuracy for the majority of the classes with major gains for, e.g., "play musical instrument" or "close".

3 Effect of α:

α from Eq. 7 of the original manuscript seems to have minor effect on performance of the proposed approach as seen in Fig. 2. Here, $\alpha = 0$ corresponds to MIML+HGRNN when actor-association is not used.
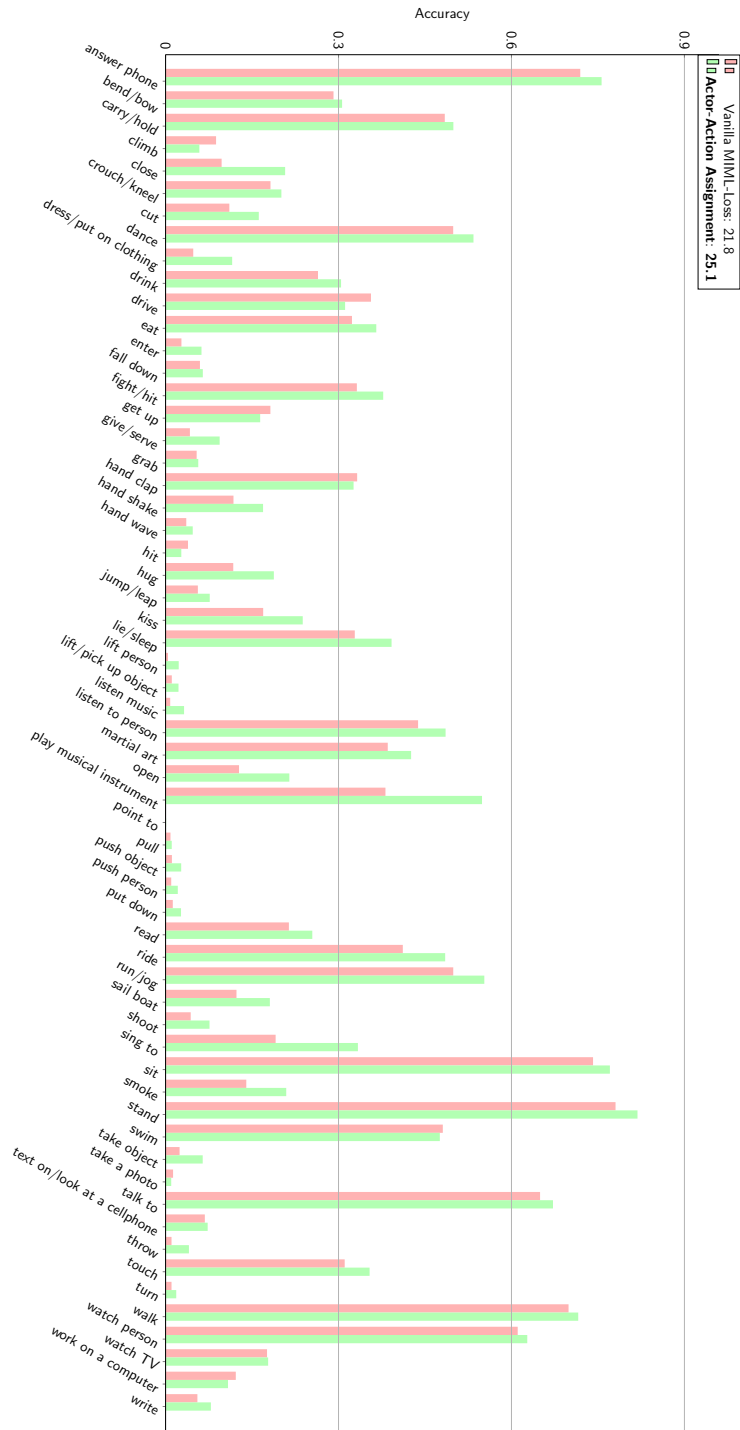
Fig. 1. Comparison of MIML with proposed method. The plot shows the per class mAP for all 60 classes on the validation set. The actions are sorted by alphabetical order from left to right.
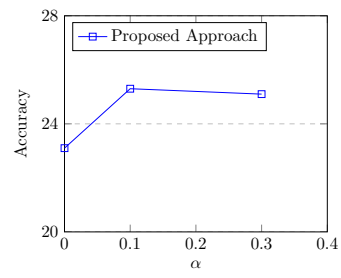
Fig. 2. Effect of varying α of Eq. 7 in the main manuscript