Multi-label affordance mapping from egocentric vision

Lorenzo Mur-Labadia   Jose J. Guerrero   Ruben Martinez-Cantin
I3A - Universidad de Zaragoza
lmur, jguerrer, rmcantin@unizar.es

1. EPIC-Aff dataset

We build the EPIC-Aff dataset by joining the EPIC-100 Kitchens narrations with the EPIC-VISOR masks, producing two versions of the dataset shown in Tables 1 and 2. We use the authors provided remapping to work with the 304 different semantic classes. We use Colmap with a mask of the dynamic objects for extracting the camera poses. We perform guided epipolar matching and an exhaustive extraction of matches. For the depth extraction, we run at inference the *large* version of MiDAS, a deep neural network trained on 12 datasets with multi-objective optimization.

We show multiple samples of the EPIC-Aff ground truth and the DeepLab-Asym predictions in Figures 2 and 3. Our multi-label approach shows the relevance of considering several interactions per object: the pepper on the chopping board of the third row in Figure 2 is *cuttable, removable, takeable, moveable* and *throwable*. Furthermore, since the labels are interaction grounded, they are associated with a specific region of the region, e.g (on the last row of Figure 2) *turn-on* or *adjust* are only at the hub controllers, not in all the objects.

Table 1: List of easy EPIC-Aff 20 classes

'take', 'put', 'open', 'close', 'wash', 'cut', 'mix', 'pour', 'throw', 'move', remove', 'dry', 'turn-on', 'turn', 'shake', 'turn-off', 'peel', 'adjust', 'empty', 'scoop'

Table 2: List of complex EPIC-Aff 43 classes

'put', 'take', 'wash', 'open', 'close', 'cut', 'mix', 'pour', 'remove', 'move', 'turn-on', 'turn-off', 'throw', 'dry', 'turn', 'insert', 'peel', 'shake', 'adjust', 'squeeze', 'check', 'empty', 'fill', 'flip', 'scrape', 'wrap', 'add', 'break', 'spray', 'roll', 'search', 'fold', 'sprinkle', 'press', 'scoop', 'cook', 'smell', 'pull', 'set', 'apply', 'drink', 'eat', 'sort'

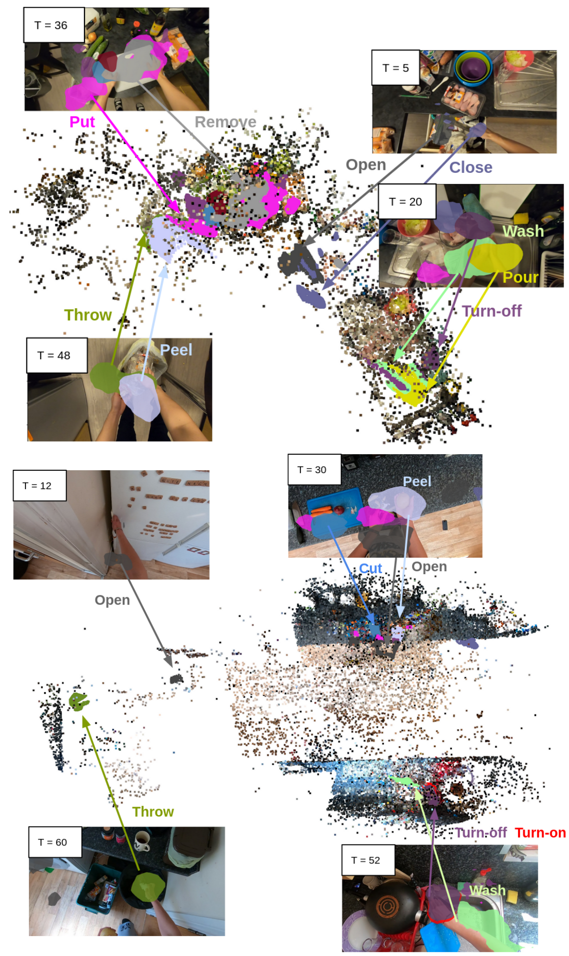

Figure 1. Sparse multi-label affordance mapping during the preparation of different recipes. We show the Colmap key-points and the projection of the 2D-affordances labels on the map for different time-steps,

2. Mapping of affordances

On Figure 1, we show two more examples of the affordance mapping, illustrating the spatial distribution of certain actions in the environment.

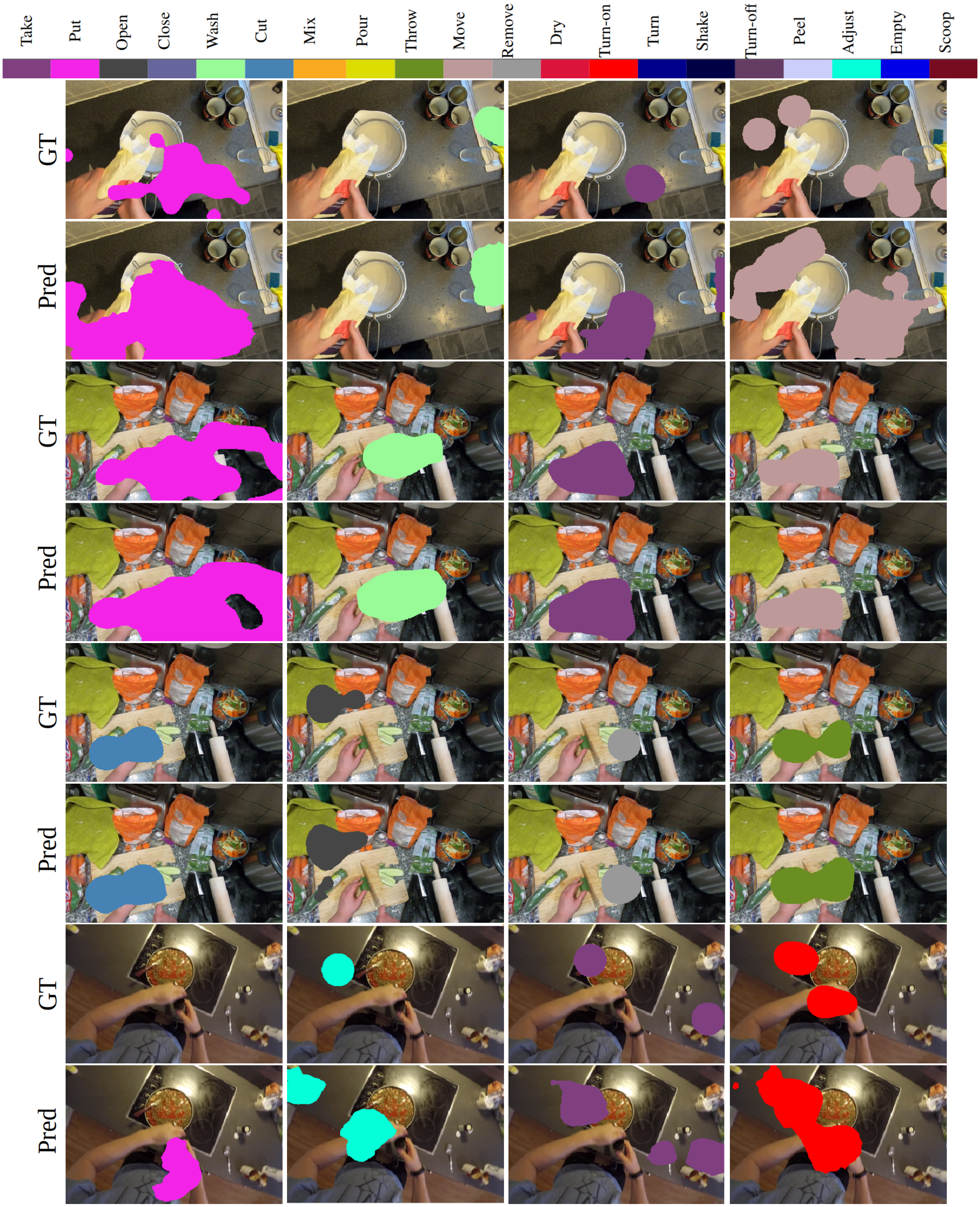

Figure 2. Multi-label affordances samples on the EPIC-Aff with 20 classes

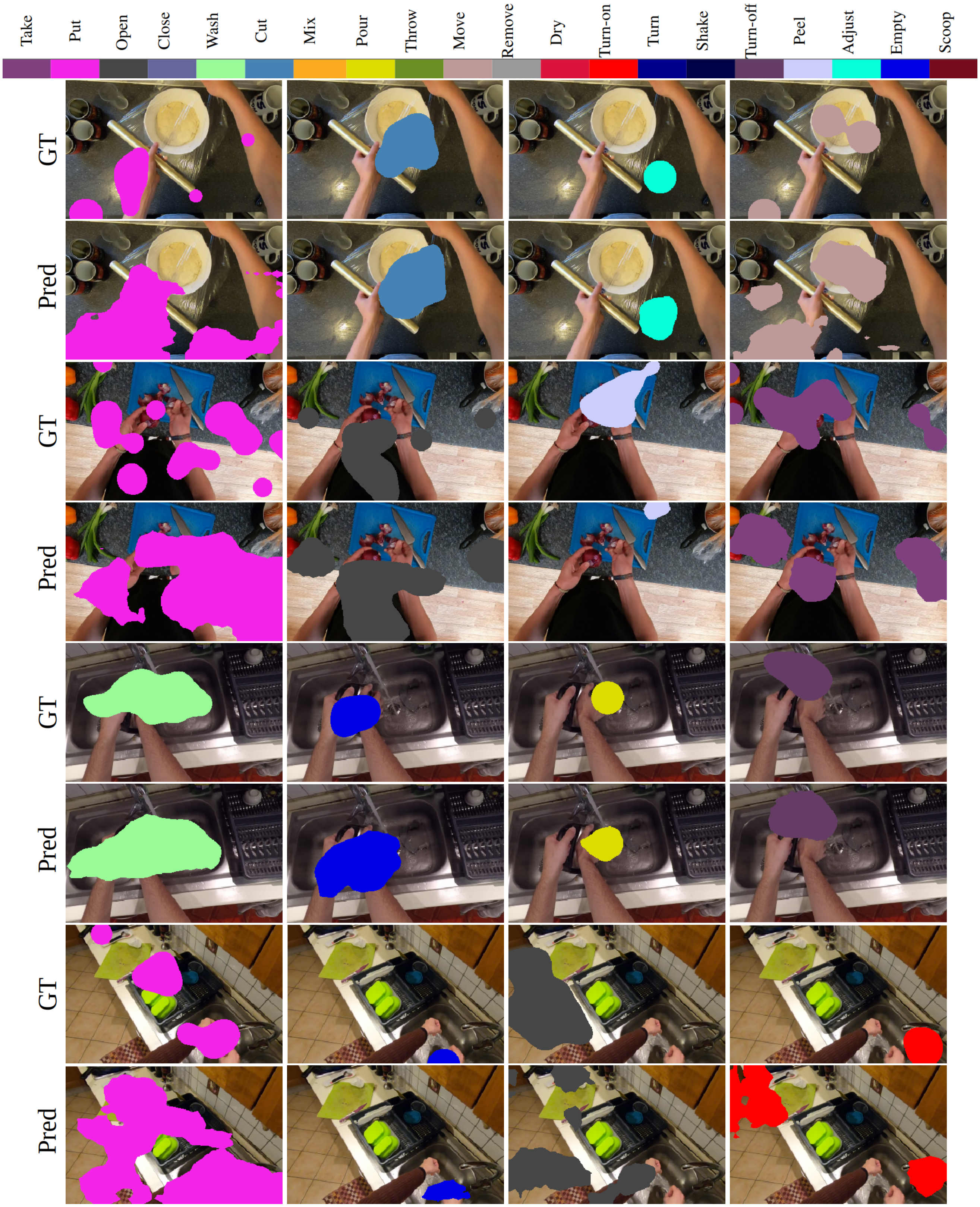

Figure 3. Multi-label affordances samples on the EPIC-Aff with 20 classes