SARM

Key Take-aways

1. A Dual Reward Model Architecture

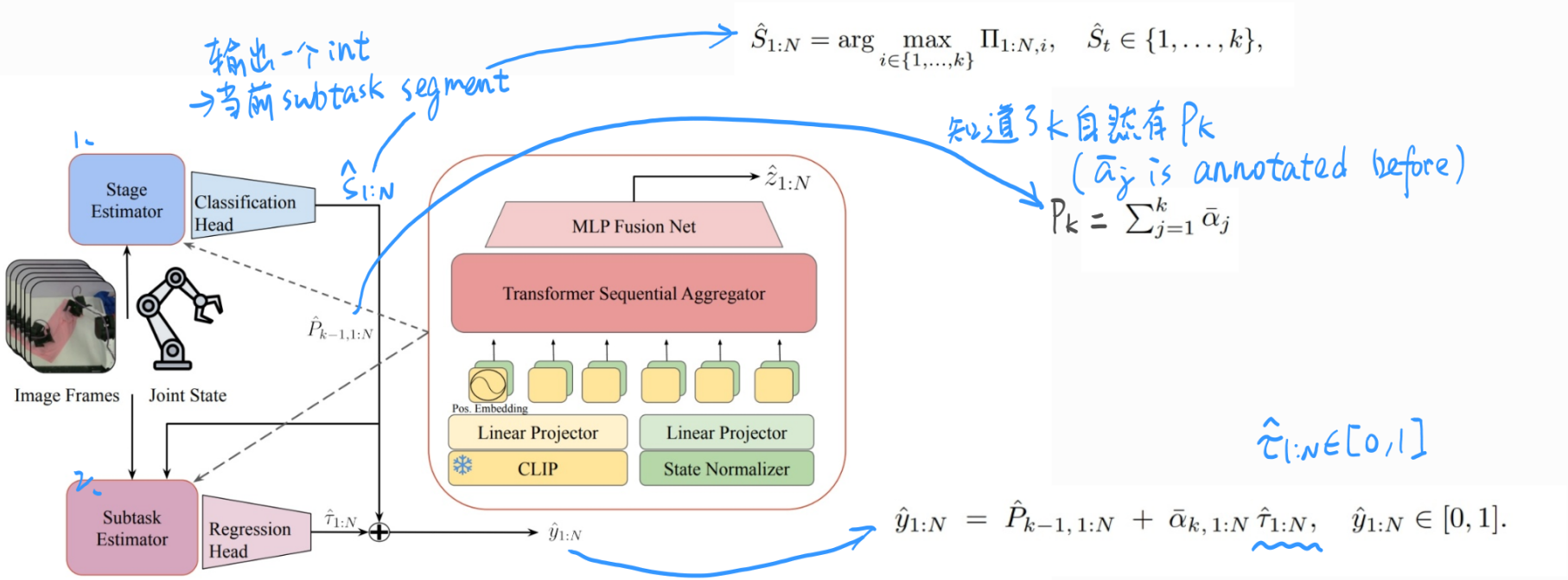

Figure 2: Overview of **SARM,** stage-aware reward modeling. **Left:** SARM overview, which includes both a stage estimator and subtask estimator. First the task stage is predicted from the observations. This prediction is additionally passed into the subtask estimator which predicts a scale value of the progress within the stage. **Right:** An overview of the estimator architecture which is replicated for both the stage estimator and the subtask estimator.

* Note,
1. N, a sequence of N images
2. give $\bar{\alpha}_k$ from labeling (i.e., annotation)

Labeling by subtask priors: Let trajectory i have total length T_i and be segmented into K subtasks with lengths $\{L_{i,k}\}_{k=1}^{K}$. We estimate a dataset-level prior proportion for each subtask

$$\bar{\alpha}_k = \frac{1}{M}\sum_{i=1}^{M}\frac{L_{i,k}}{T_i}, \qquad \bar{\alpha}_k \geq 0, \quad \sum_{k=1}^{K}\bar{\alpha}_k = 1, \tag{1}$$

where M is the number of trajectories.