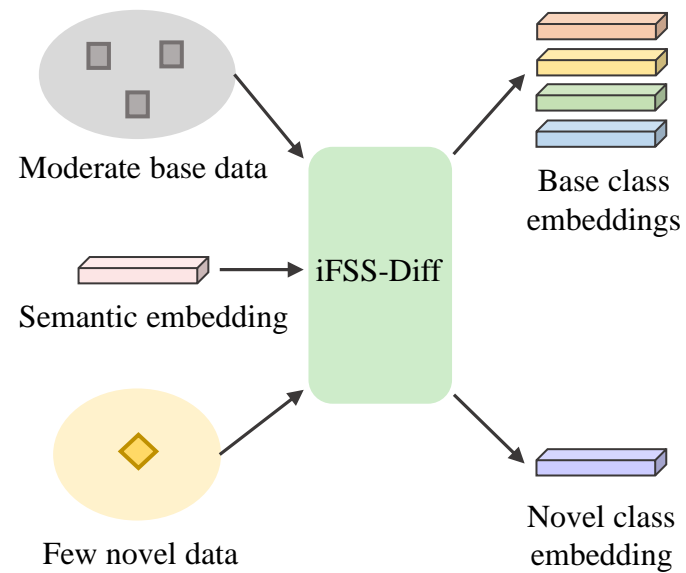
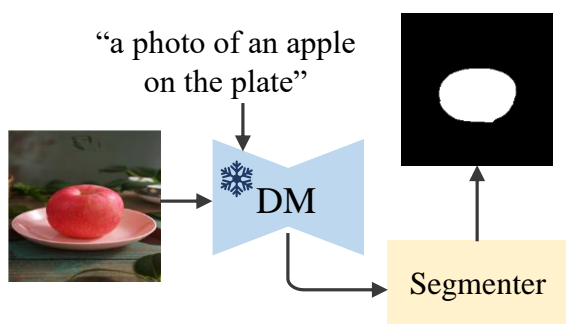


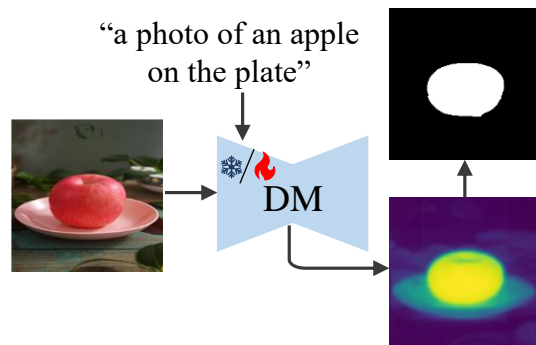
(a) Previous method



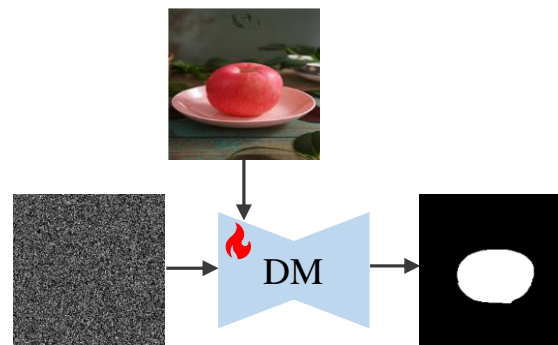
(b) iFSS-Diff (ours)



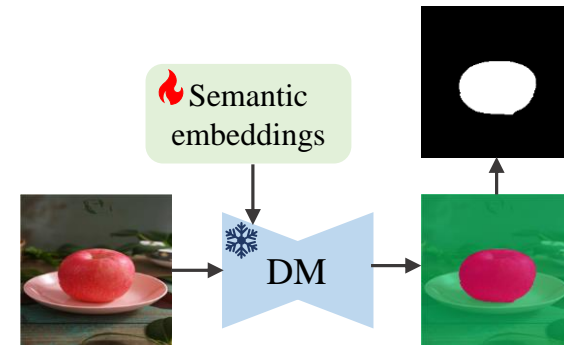
(a) Diffusion model as feature extractor



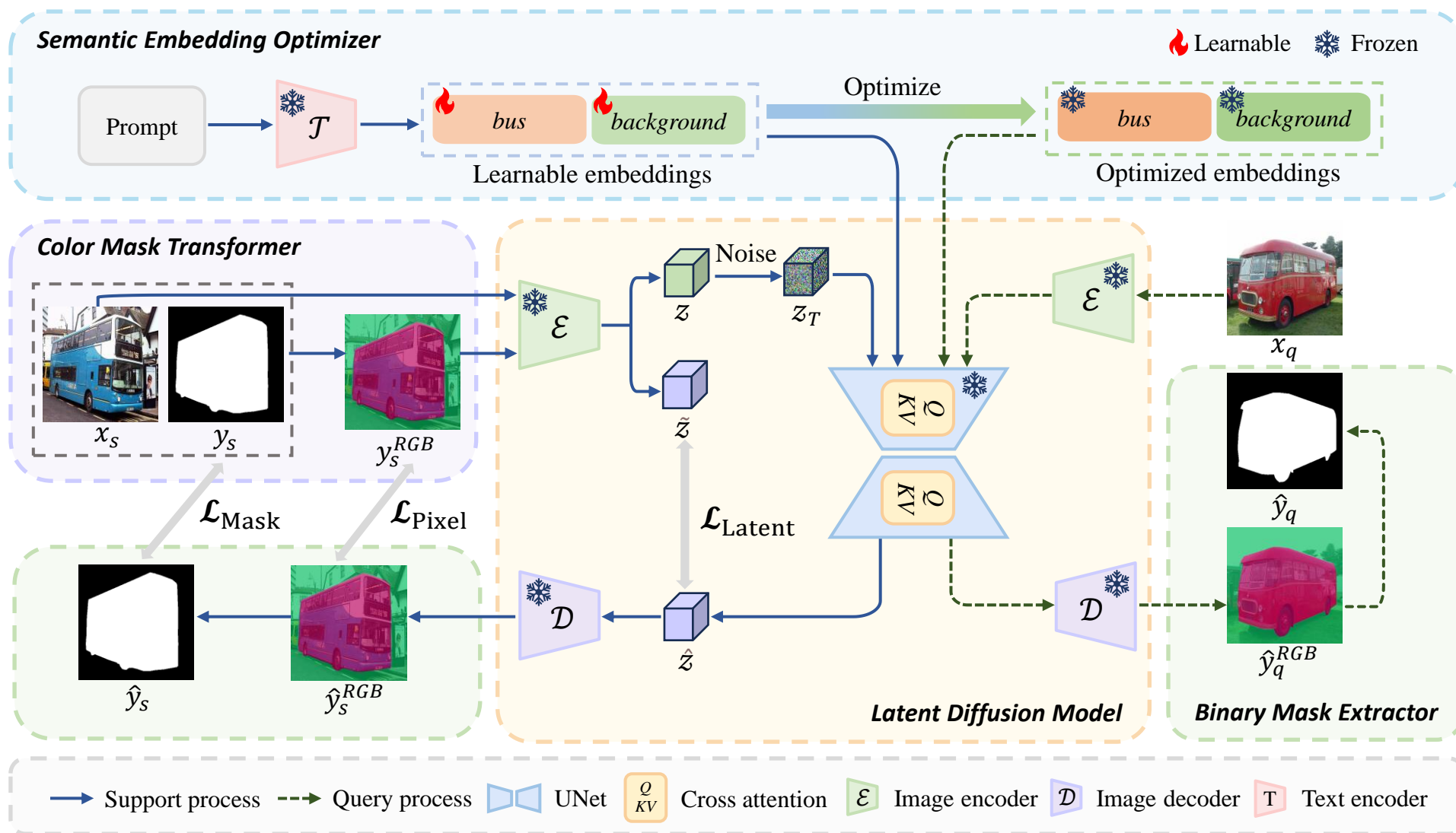
(b) Based on attention map

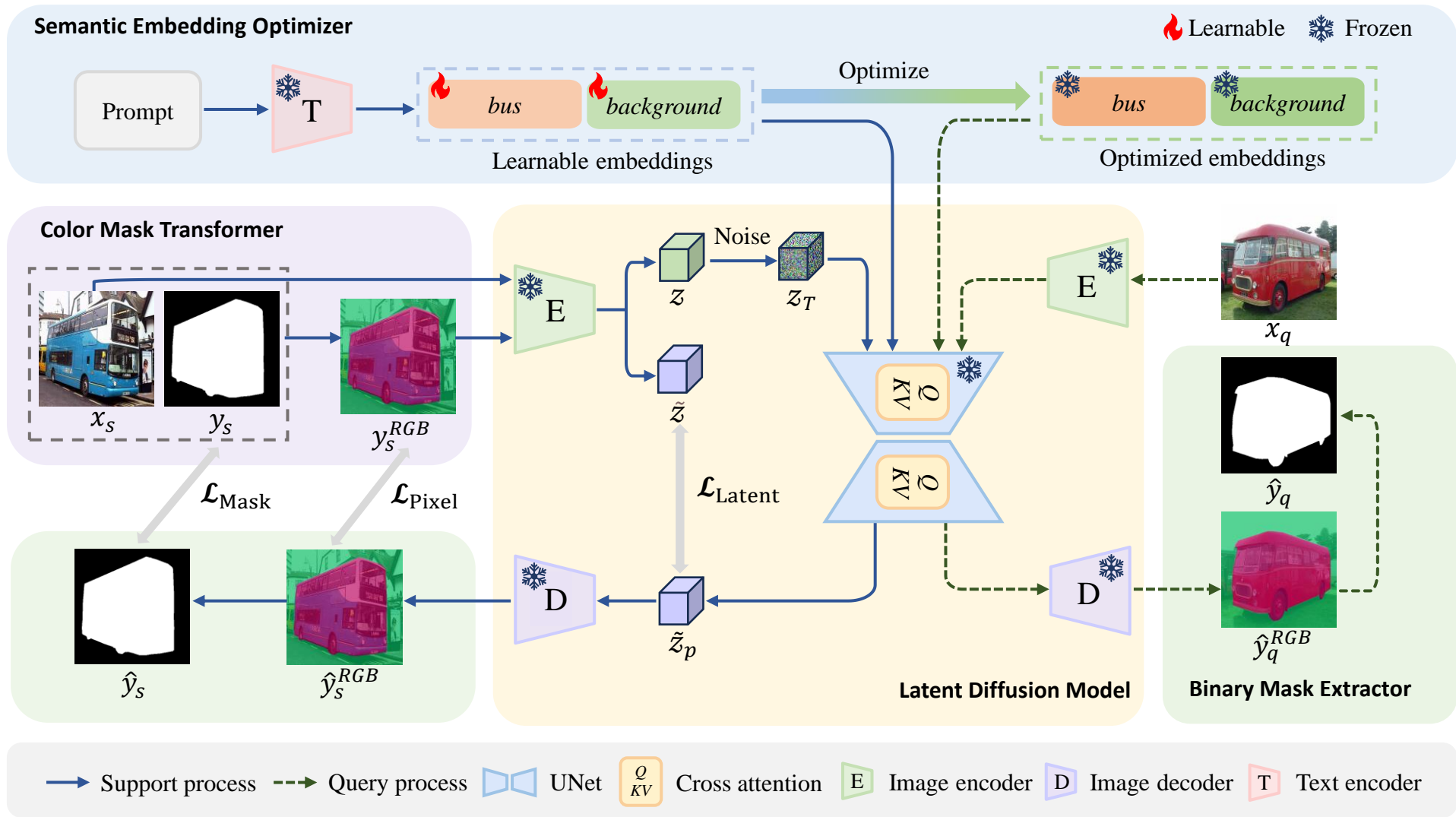


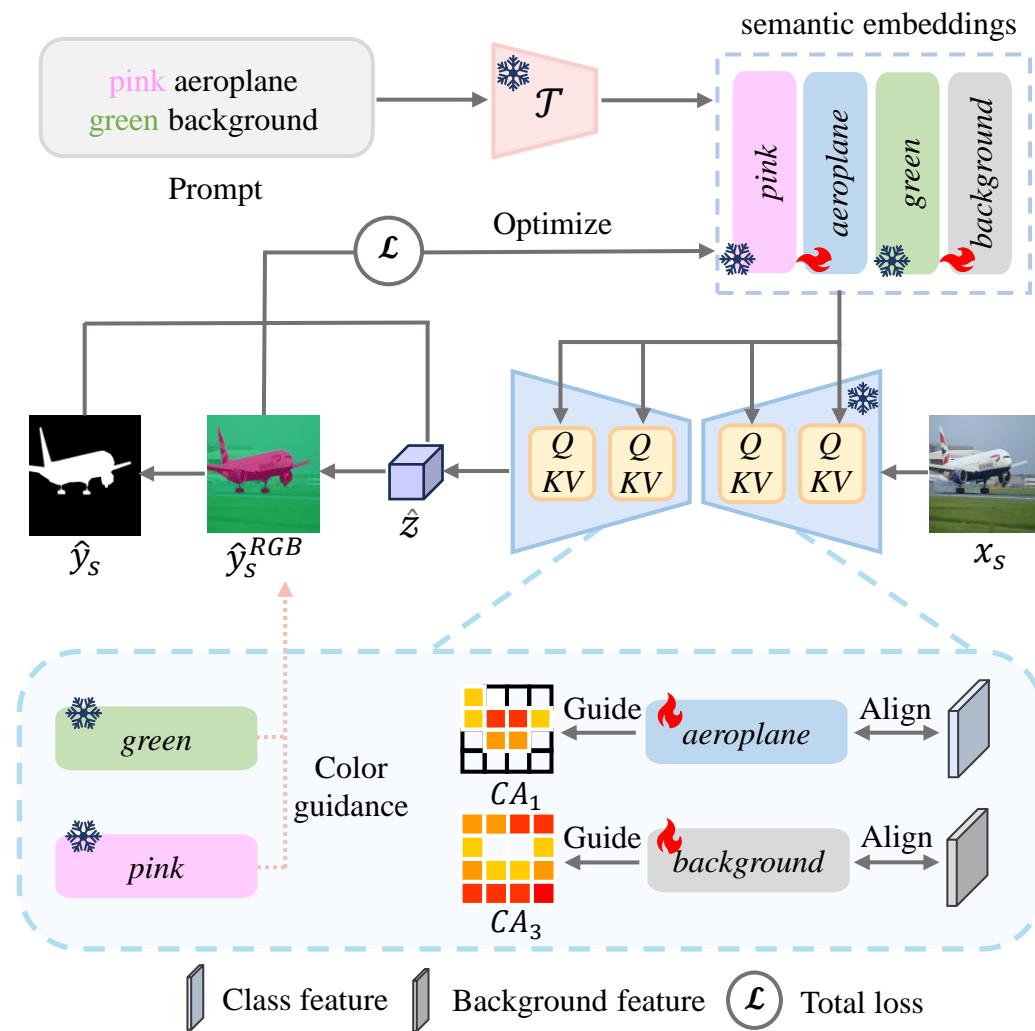
(c) Model mask distribution

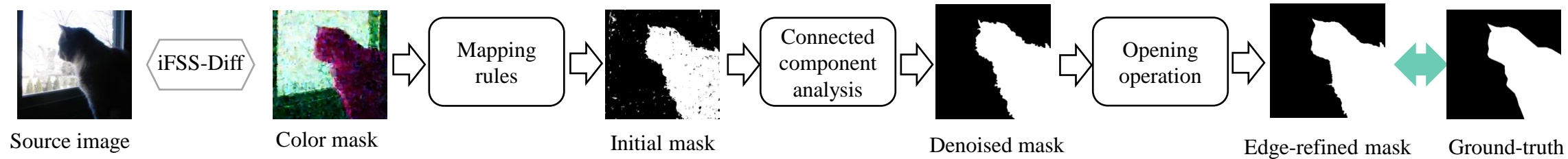


(d) iFSS-Diff (**ours**)

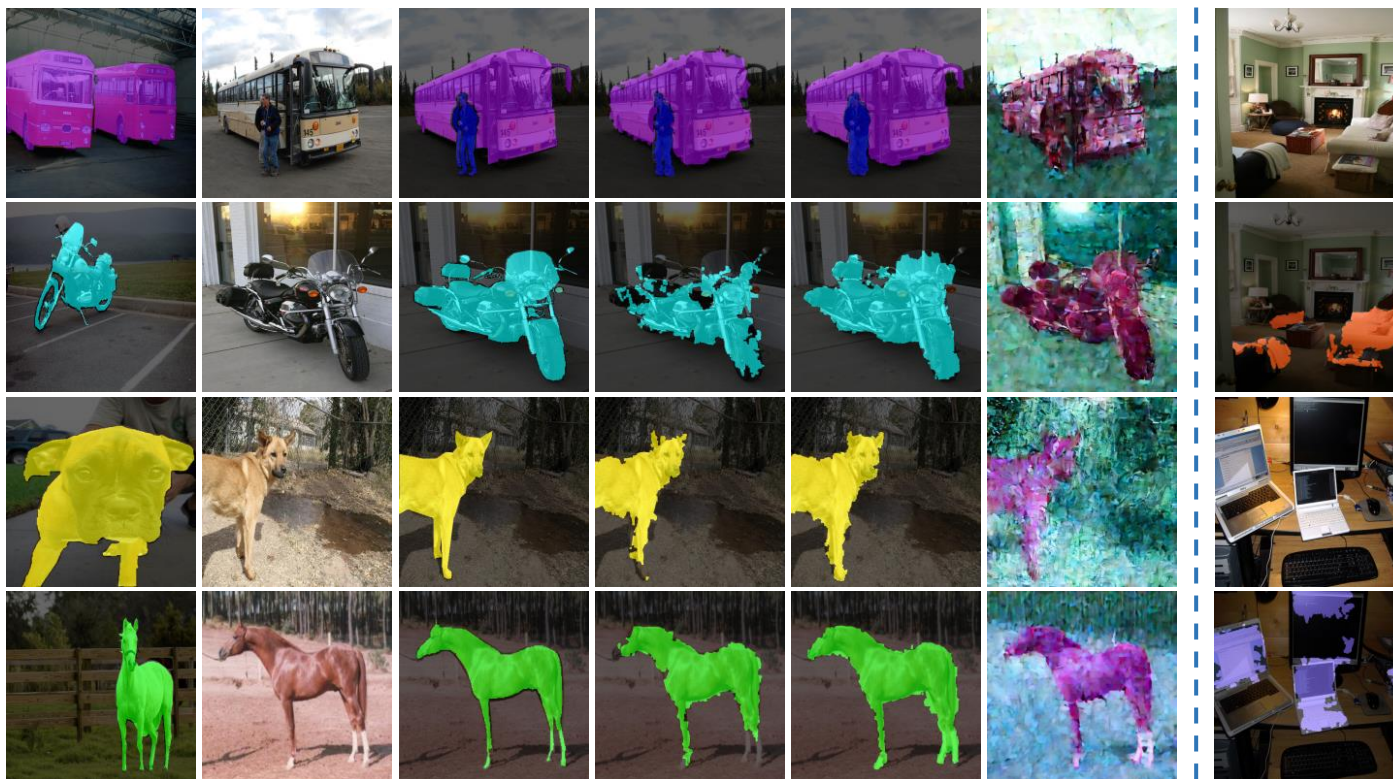








■ bus
 ■ person
 ■ motorbike
 ■ dog
 ■ horse
 ■ sofa
 ■ tv-monitor



Support (train)

Query (test)

Ground-truth

Baseline

iFSS-Diff (**ours**)

RGB mask

Failure cases

