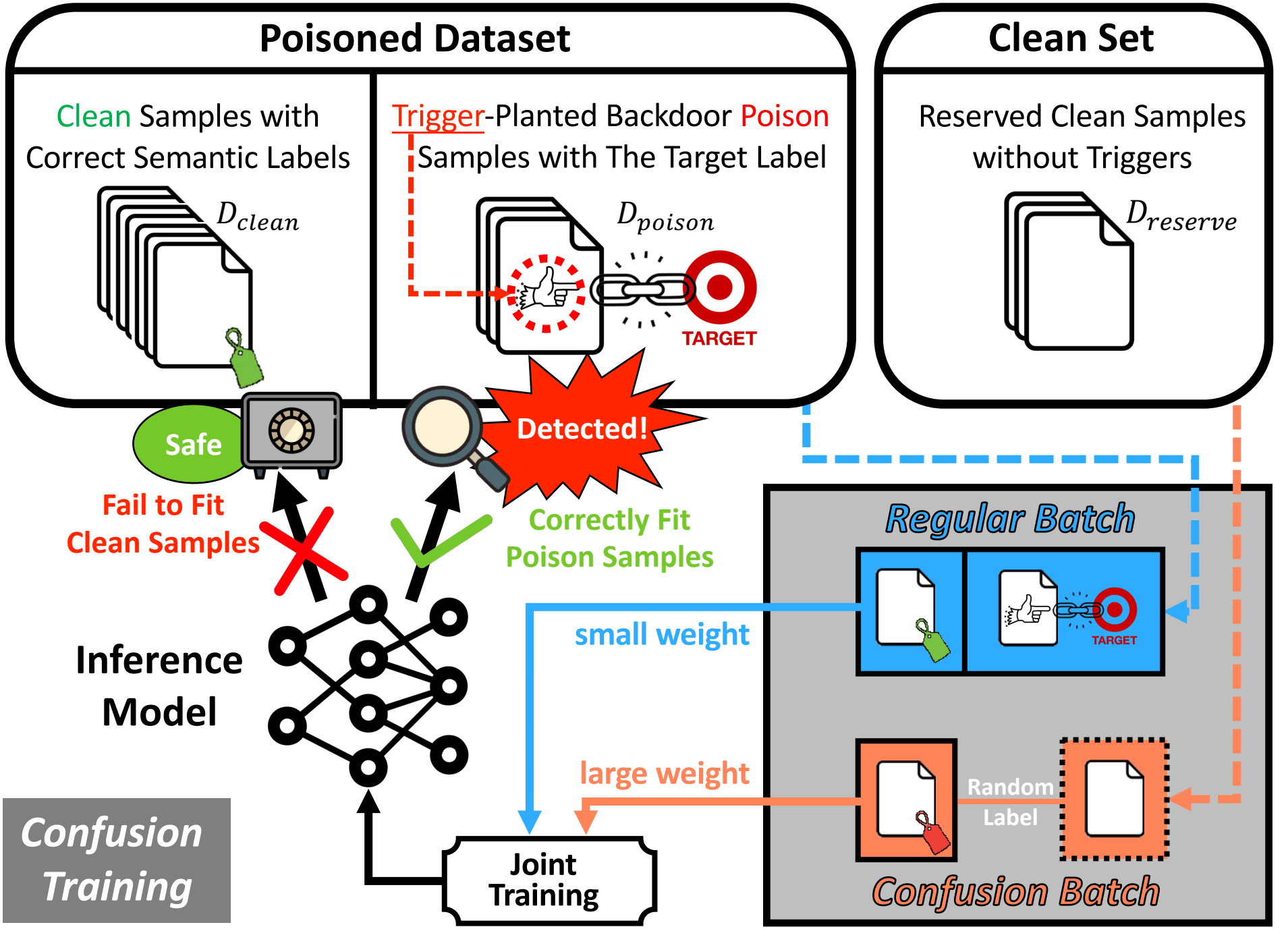
Poisoned Dataset
Clean Set
Clean Samples with Correct Semantic Labels
Trigger-Planted Backdoor Poison Samples with The Target Label
Reserved Clean Samples without Triggers
D_{clean}
D_{poison}
$D_{reserve}$
TARGET
Safe
Detected!
Fail to Fit Clean Samples
Correctly Fit Poison Samples
Inference Model
Regular Batch
small weight
large weight
Random Label
Confusion Batch
Joint Training
Confusion Training