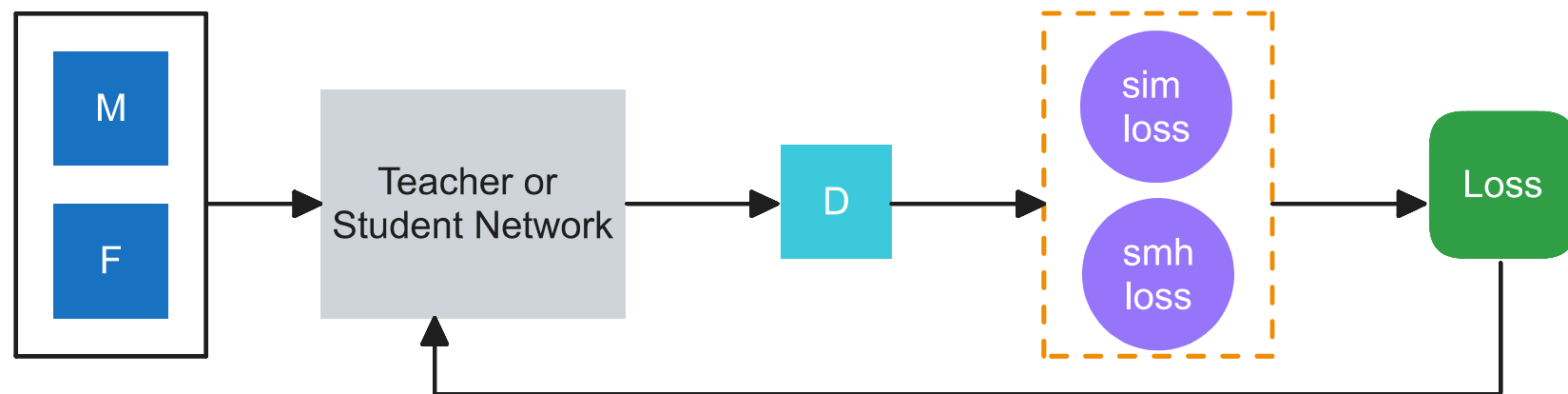
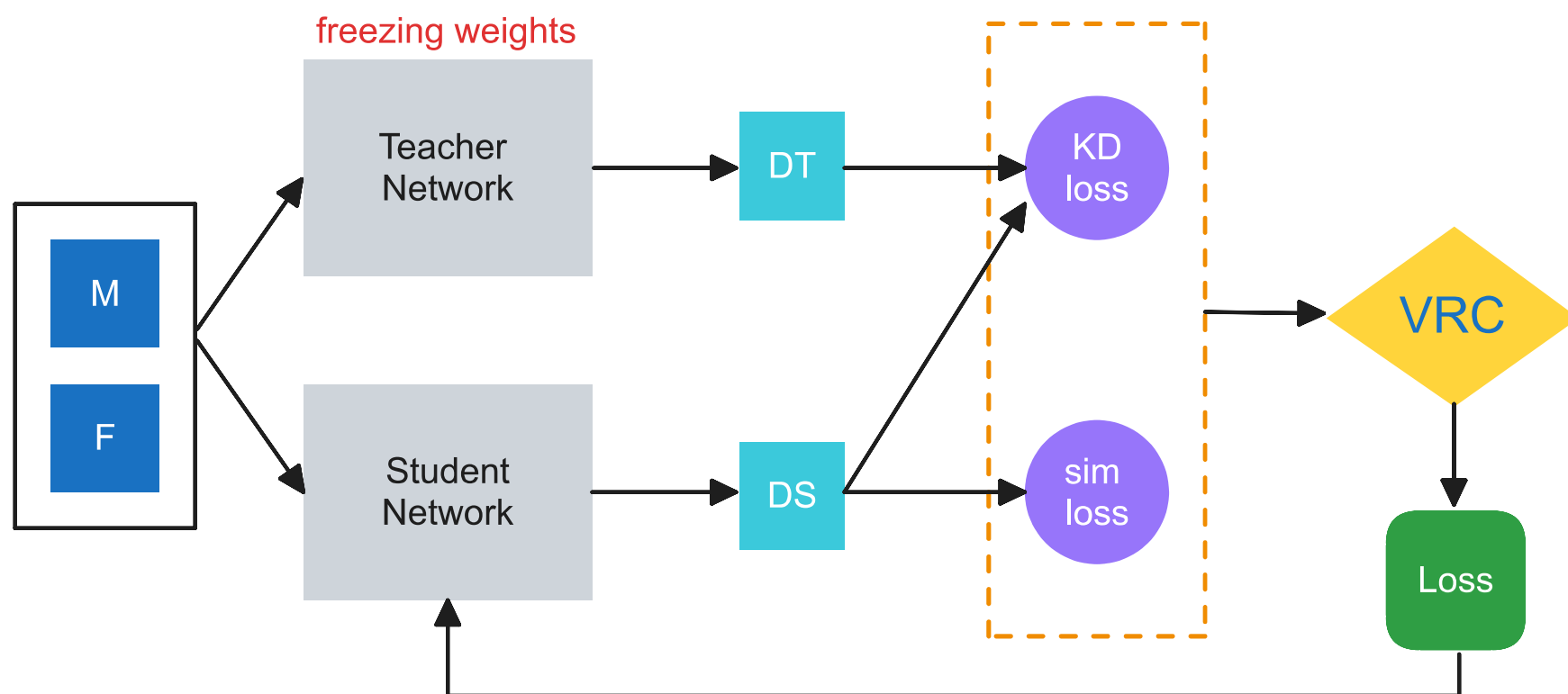


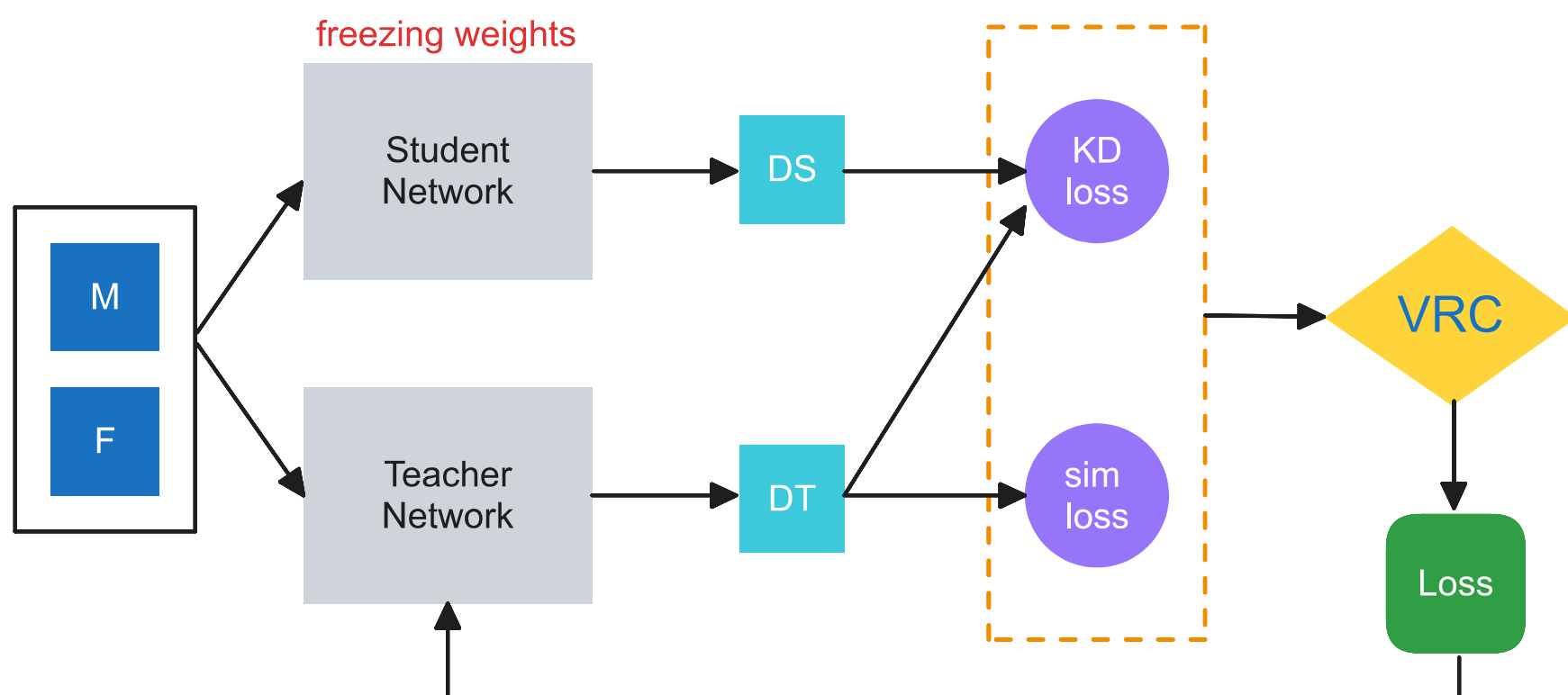
Stage 1: UnSupervised pre-training for the teacher and student networks respectively with similarity and smooth loss



Stage 2: Fine-tune the student network with similarity and KD loss



Stage 3: Fine-tune the teacher network with similarity and KD loss



Recursive training