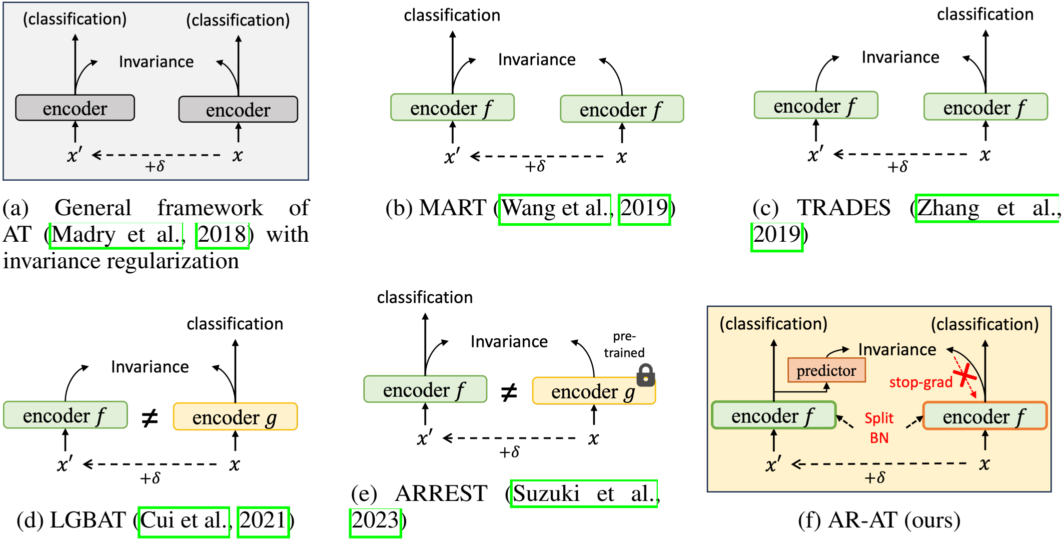

(a) General framework of AT (Madry et al., 2018) with invariance regularization

(b) MART (Wang et al., 2019)

(c) TRADES (Zhang et al., 2019)

(d) LGBAT (Cui et al., 2021)

(e) ARREST (Suzuki et al., 2023)

(f) AR-AT (ours)

Figure 1: Comparison of invariance regularization-based adversarial defense methods. Our approach employs an asymmetric structure for invariance regularization with a stop-gradient and predictor, and a split-BatchNorm (BN) to maintain consistent batch statistics during training.