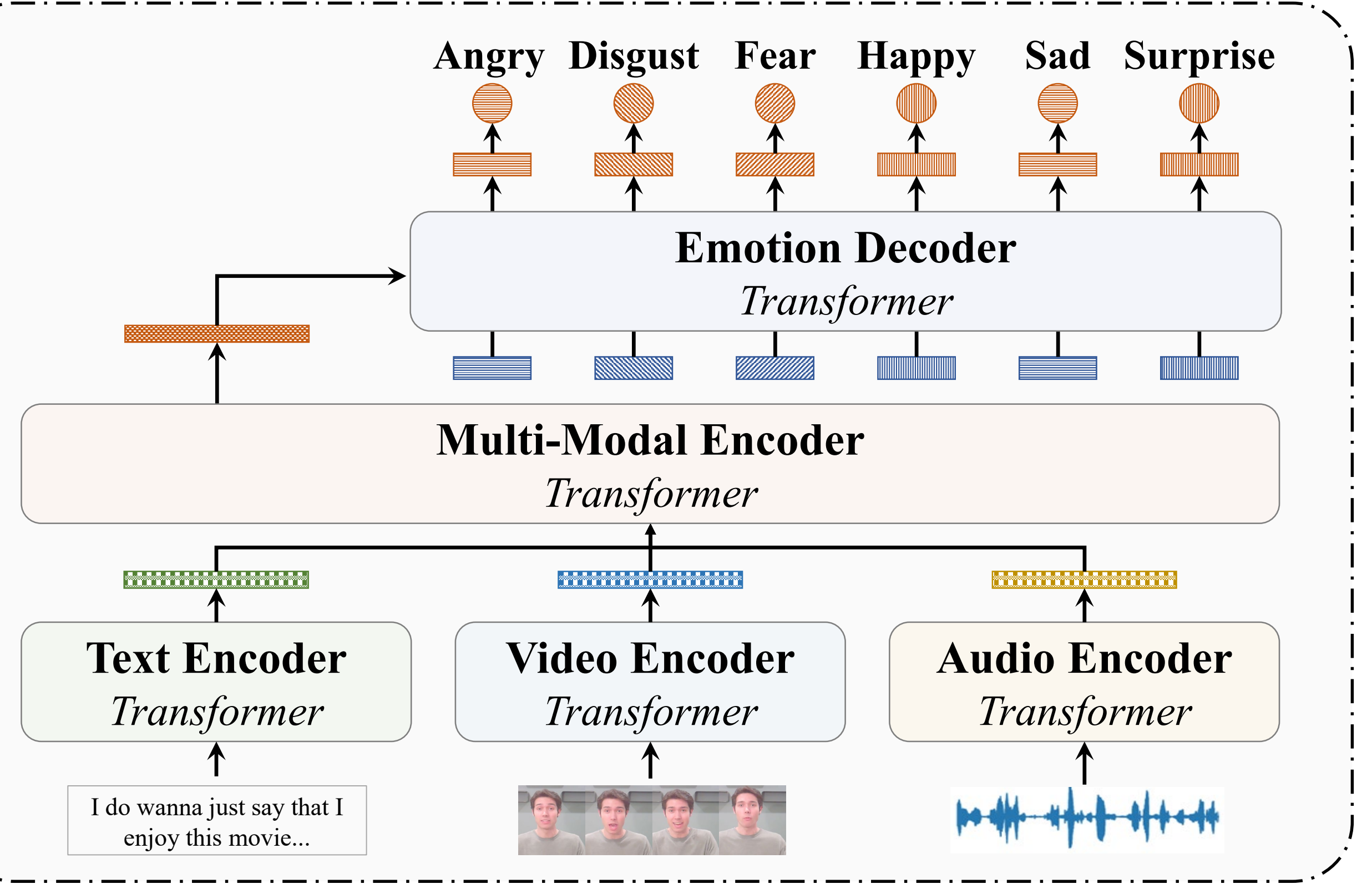
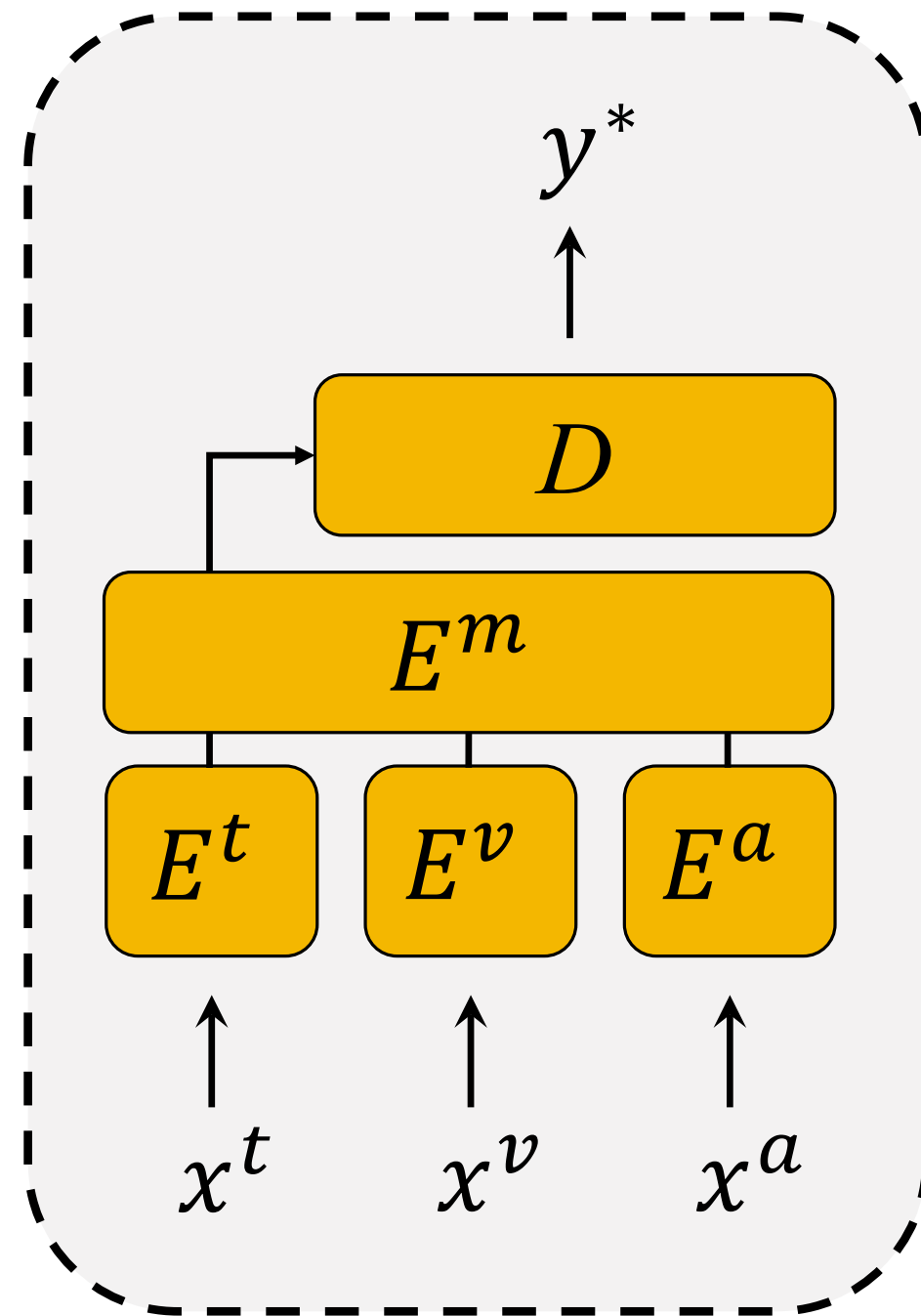
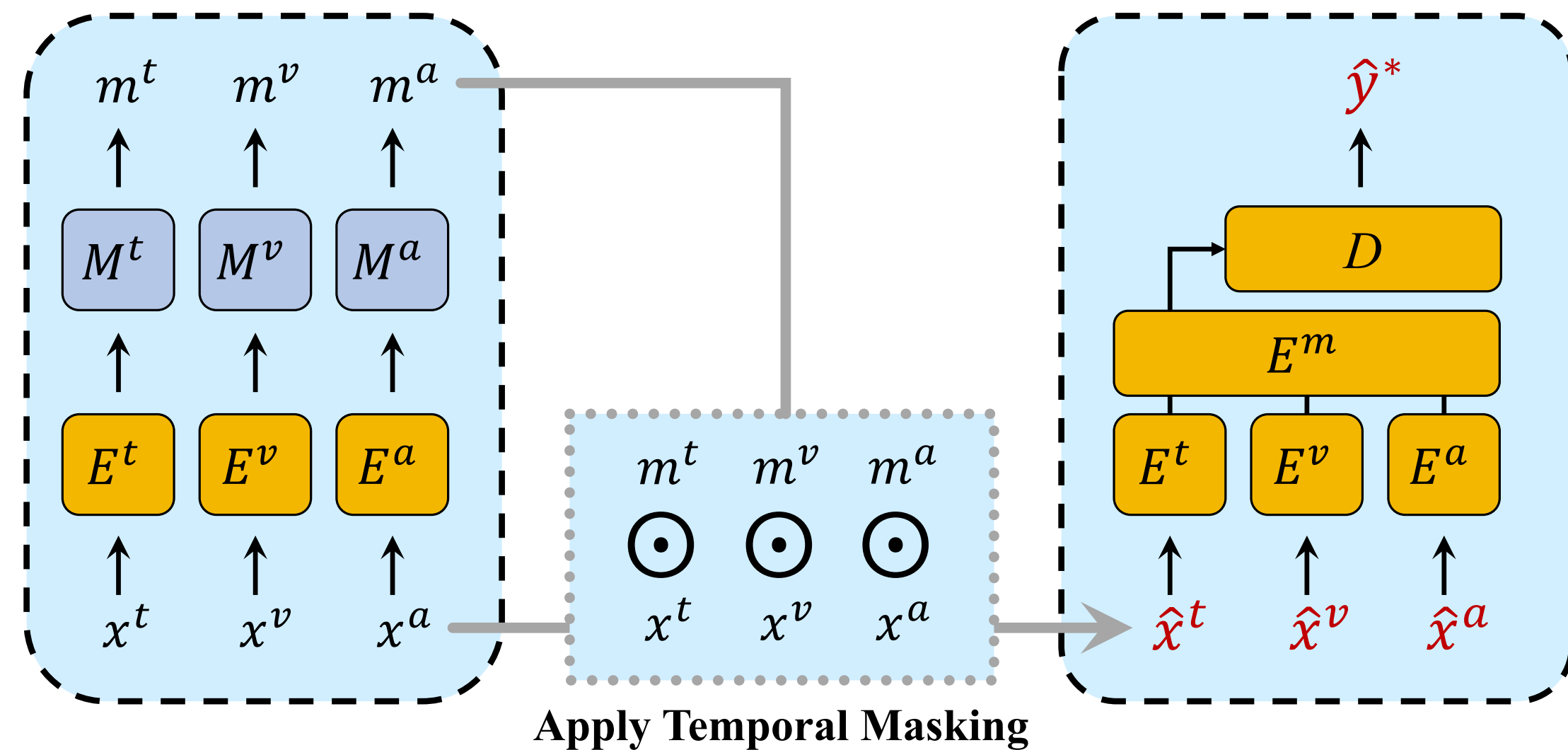


$$\min_{E^t, E^v, E^a, E^m, D} \mathbb{E}_{(x^t, x^v, x^a, y) \sim \mathcal{D}} [\mathcal{L}_r(y^*, y)]$$



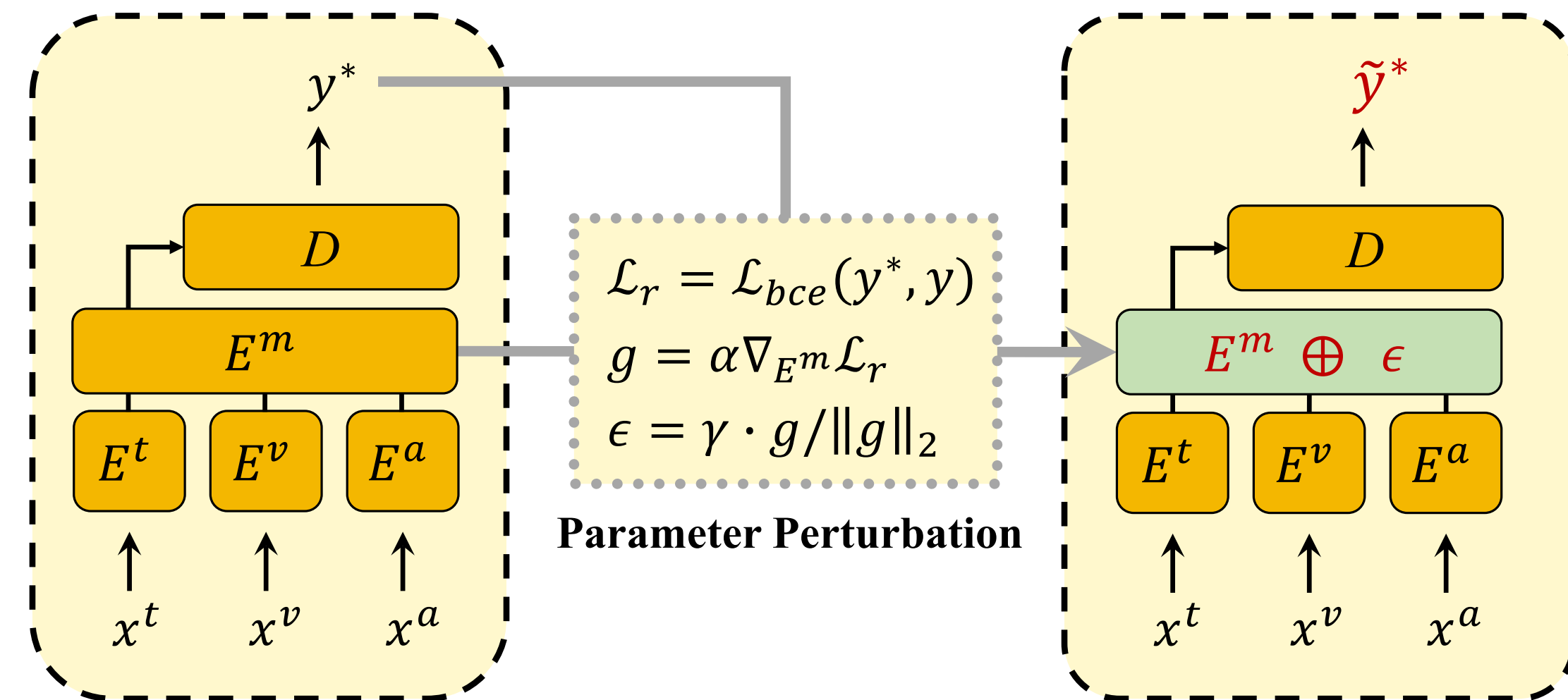
(a) Multi-Modal Multi-Label Emotion Recognition

$$\min_{E^t, E^v, E^a, E^m, D} \mathbb{E}_{(x^t, x^v, x^a, y) \sim \mathcal{D}} \left[\max_{M^t, M^v, M^a} \mathcal{L}_m(\hat{y}^*, y) \right]$$



(b) Adversarial Temporal Masking

$$\min_{E^t, E^v, E^a, E^m, D} \mathbb{E}_{(x^t, x^v, x^a, y) \sim \mathcal{D}} \left[\max_{\epsilon} \mathcal{L}_p(\tilde{y}^*, y^*) \right]$$



(c) Adversarial Parameter Perturbation