Input representations X_1 & X_2 Task-relevant information
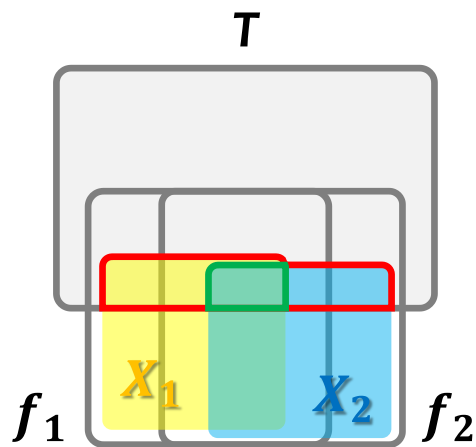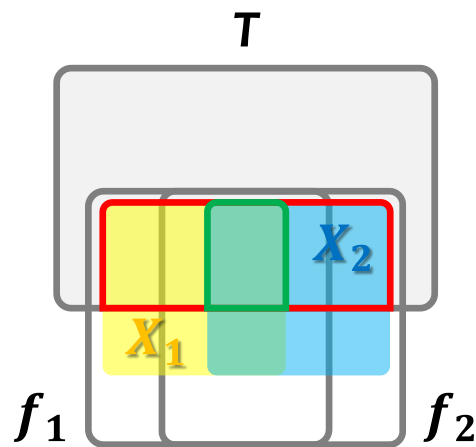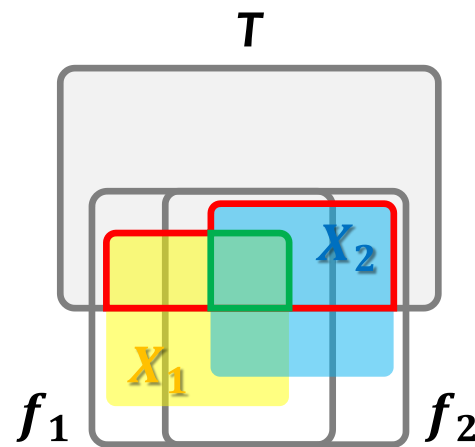
Task & representations shared information

Representations X_1 & X_2 shared information

(a) Factorized (2D+1D)
Spatial-temporal Operation
Sufficiency ▼ Efficiency ▲

(b) Alignment-guided (Ours)
Spatial-temporal Operation
Sufficiency ▲ Efficiency ▲

(c) Joint (3D)
Spatial-temporal Operation
Sufficiency ▲ Efficiency ▼