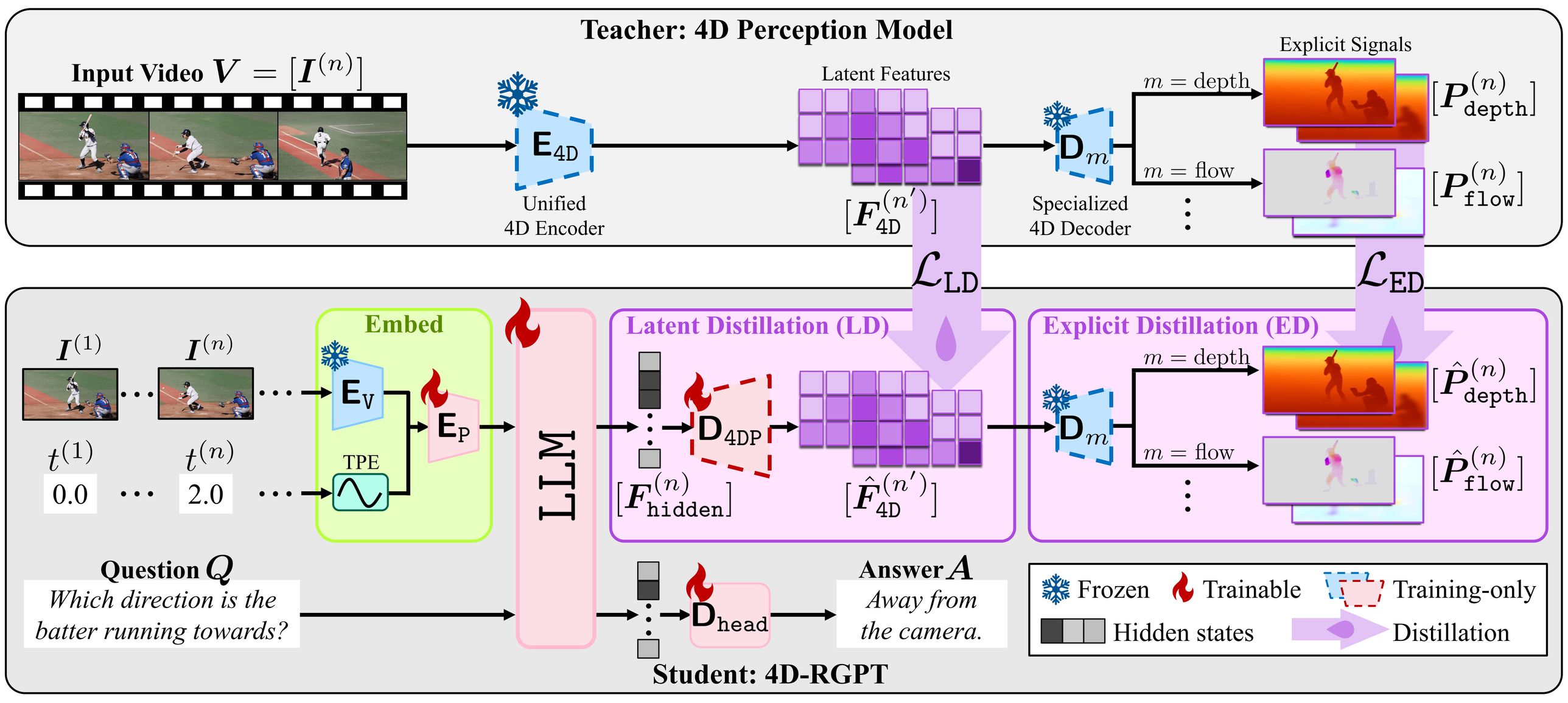
Teacher: 4D Perception Model
Input Video $V = [I^{(n)}]$
E_{4D}
Unified 4D Encoder
Latent Features
$[F_{4D}^{(n')}]$
D_m
Specialized 4D Decoder
m = depth
m = flow
Explicit Signals
$[P_{depth}^{(n)}]$
$[P_{flow}^{(n)}]$
$\mathcal{L}_{LD}$
$\mathcal{L}_{ED}$
$I^{(1)}$
$I^{(n)}$
$t^{(1)}$
$t^{(n)}$
0.0
2.0
Embed
E_V
E_P
TPE
LLM
Latent Distillation (LD)
D_{4DP}
$[F_{hidden}^{(n)}]$
$[\hat{F}_{4D}^{(n')}]$
Explicit Distillation (ED)
$[\hat{P}_{depth}^{(n)}]$
$[\hat{P}_{flow}^{(n)}]$
Question Q
Which direction is the batter running towards?
D_{head}
Answer A
Away from the camera.
Frozen
Trainable
Training-only
Hidden states
Distillation
Student: 4D-RGPT