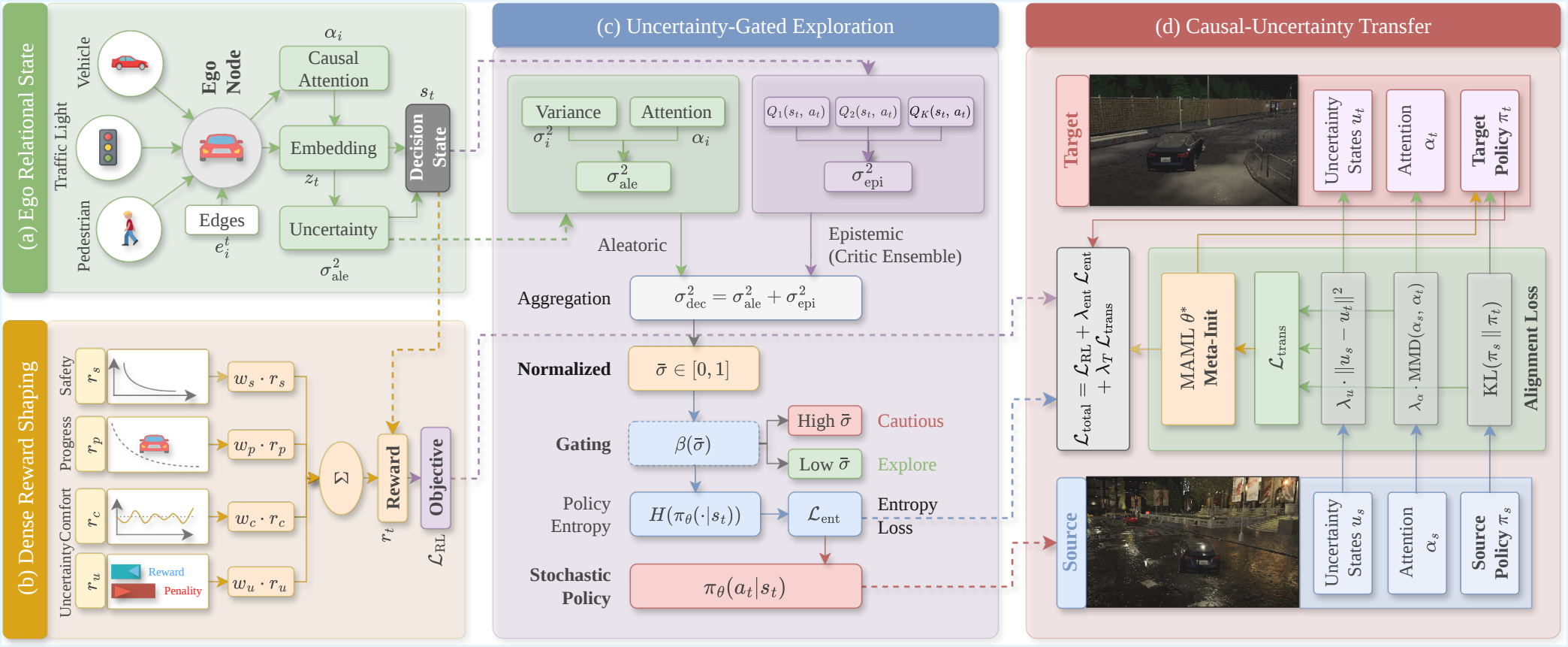
(a) Ego Relational State
Vehicle
Traffic Light
Pedestrian
Ego Node
Edges
e_i^t
α_i
Causal Attention
Embedding
z_t
Uncertainty
σ^2_{ale}
s_t
Decision State
(b) Dense Reward Shaping
Safety
Progress
Comfort
Uncertainty
r_s
r_p
r_c
r_u
Reward
Penalty
$w_s \cdot r_s$
$w_p \cdot r_p$
$w_c \cdot r_c$
$w_u \cdot r_u$
Σ
Reward
r_t
Objective
$\mathcal{L}_{\text{RL}}$
(c) Uncertainty-Gated Exploration
Variance
Attention
σ_i^2
α_i
σ^2_{ale}
$Q_1(s_t, a_t)$
$Q_2(s_t, a_t)$
$Q_K(s_t, a_t)$
σ^2_{epi}
Aleatoric
Epistemic
(Critic Ensemble)
Aggregation
$\sigma^2_{\text{dec}} = \sigma^2_{\text{ale}} + \sigma^2_{\text{epi}}$
Normalized
$\bar{\sigma} \in [0, 1]$
Gating
$\beta(\bar{\sigma})$
High $\bar{\sigma}$
Cautious
Low $\bar{\sigma}$
Explore
Policy Entropy
$H(\pi_\theta(\cdot|s_t))$
$\mathcal{L}_{\text{ent}}$
Entropy Loss
Stochastic Policy
$\pi_\theta(a_t|s_t)$
(d) Causal-Uncertainty Transfer
Target
Uncertainty States u_t
Attention α_t
Target Policy π_t
$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{RL}} + \lambda_{\text{ent}} \mathcal{L}_{\text{ent}} + \lambda_T \mathcal{L}_{\text{trans}}$
MAML θ^* Meta-Init
$\mathcal{L}_{\text{trans}}$
$\lambda_u \cdot \|u_s - u_t\|^2$
$\lambda_\alpha \cdot \text{MMD}(\alpha_s, \alpha_t)$
$\text{KL}(\pi_s \| \pi_t)$
Alignment Loss
Source
Uncertainty States u_s
Attention α_s
Source Policy π_s