

## SUPPLEMENTARY

### A. Proof of Theorem 1

The Lagrange function of sub-problem  $\mathbb{P}_\text{II}^\text{B}$  can be defined as:

$$\mathcal{L}(\mathbf{V}, \beta) = \sum_{m=1}^M \left( \frac{\alpha_m^{(1)} + \alpha_m^{(2)}}{v_m^\text{ES}} \right) + \beta \left( \sum_{m=1}^M v_m^\text{ES} - v_{\max}^\text{ES} \right), \quad (41)$$

where  $\beta$  is the Lagrange multiplier.

The Karush-Kuhn-Tucker (KKT) conditions are as follows:

- Stationarity:

$$\frac{\partial \mathcal{L}}{\partial v_m^\text{ES}} = -\frac{\alpha_m^{(1)} + \alpha_m^{(2)}}{(v_m^\text{ES})^2} + \beta = 0, \quad \forall m \in \mathcal{M}. \quad (42)$$

- Complementary Slackness:

$$\beta \left( \sum_{m=1}^M v_m^\text{ES} - v_{\max}^\text{ES} \right) = 0. \quad (43)$$

- Primal Feasibility:

$$\sum_{m=1}^M v_m^\text{ES} \leq v_{\max}^\text{ES}. \quad (44)$$

- Dual Feasibility:  $\beta \geq 0$ .

According to Eq. (42), we have:

$$v_m^\text{ES} = \sqrt{\frac{\alpha_m^{(1)} + \alpha_m^{(2)}}{\beta}}, \quad \forall m \in \mathcal{M}. \quad (45)$$

Considering Eq. (44), we have:

$$\beta = \left( \frac{\sum_{m=1}^M \sqrt{\alpha_m^{(1)} + \alpha_m^{(2)}}}{v_{\max}^\text{ES}} \right)^2. \quad (46)$$

Based on Eq. (45) and Eq. (46), we have:

$$(v_m^\text{ES})^* = v_{\max}^\text{ES} \frac{\sqrt{\alpha_m^{(1)} + \alpha_m^{(2)}}}{\sum_{m \in \mathcal{M}} \sqrt{\alpha_m^{(1)} + \alpha_m^{(2)}}}, \quad \forall m \in \mathcal{M}. \quad (47)$$

## B. Performance Evaluation (Case Study)

Fig. 18 presents sample images generated by three collaborative inference schemes: the HIS, PPGFIS, and CFIS. Among them, the HIS demonstrates superior performance by maintaining high perceptual quality even when the common inference step is set to a large value. This effectively overcomes a key limitation reported in prior work [15], where the common inference step is typically restricted to one-third of total number of inference steps (i.e., 8 out of 28). In contrast, both the PPGFIS and CFIS suffer from noticeable semantic drift, with significant content mismatches arising at step 8 and step 6, respectively. These inconsistencies suggest an inability to maintain semantic alignment across phases of the inference process.

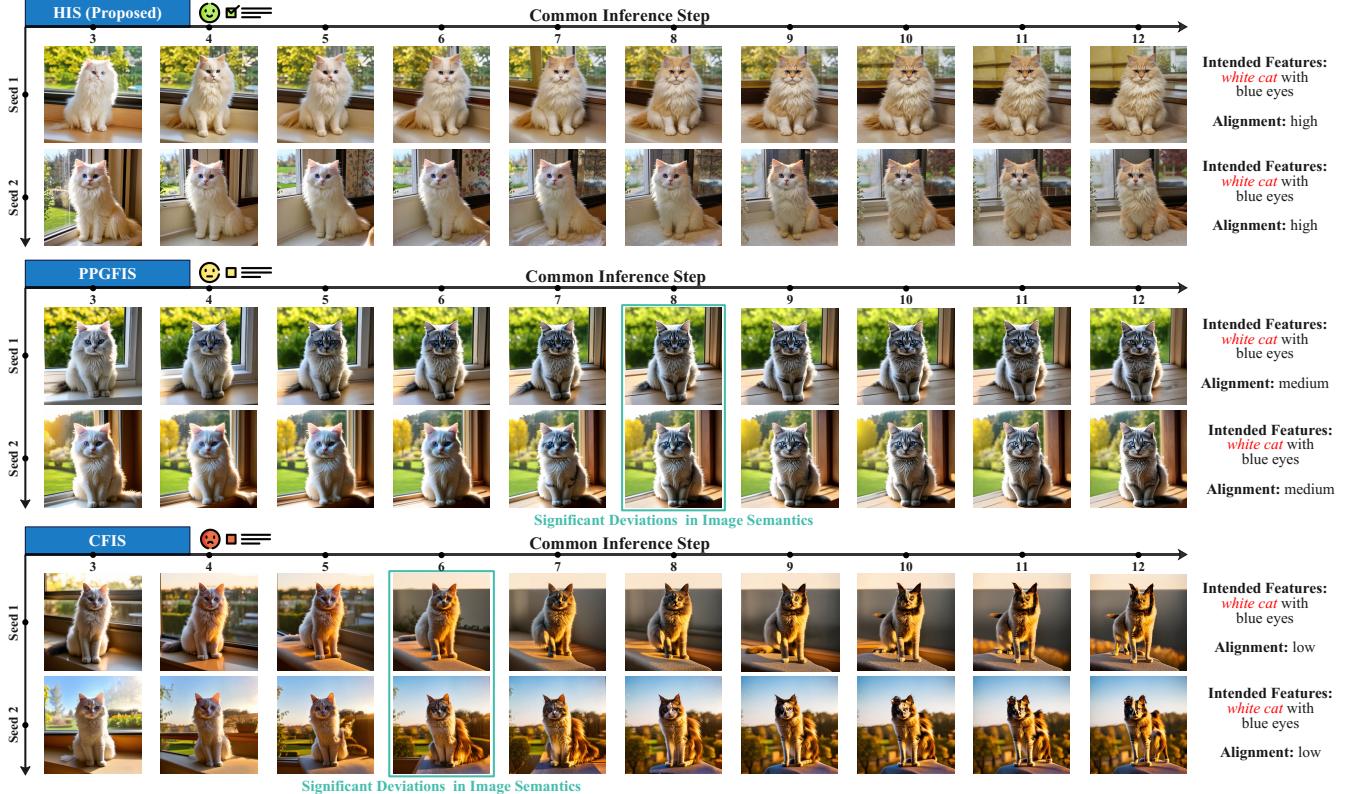


Fig. 18. Images generated by the HIS (**Top**), PPGFIS (**Middle**), and GFIS (**Bottom**) under different common inference step. All images are based on the same personal prompt: *A fluffy white cat with blue eyes sitting gracefully on a windowsill.*