Classifier-Free Guidance

Part 4/7: Sampling

Innovation: No separate classifier needed

Training Strategy

Randomly drop condition c with probability p (e.g., 10%)

Unconditional
$\varepsilon_\theta(x_t, t, \emptyset)$

+

Weight
w

×

Difference
$(\varepsilon_\theta(x_t, t, c) - \varepsilon_\theta(x_t, t, \emptyset))$

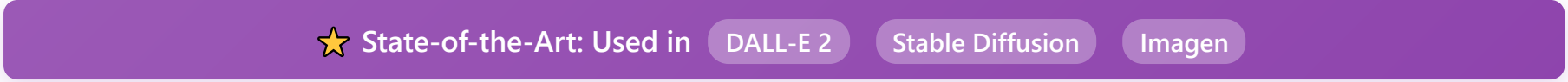
Guided Prediction Formula

$$\tilde{\varepsilon} = \varepsilon_\theta(x_t, t, \emptyset) + w \cdot (\varepsilon_\theta(x_t, t, c) - \varepsilon_\theta(x_t, t, \emptyset))$$

Guidance Scale
w > 1 increases strength

Trade-off
Higher w = better align, less diverse

Flexibility
Single model, multiple conditions


State-of-the-Art: Used in
DALL-E 2
Stable Diffusion
Imagen