1 Visualization of a Generated Sample in High-resolution

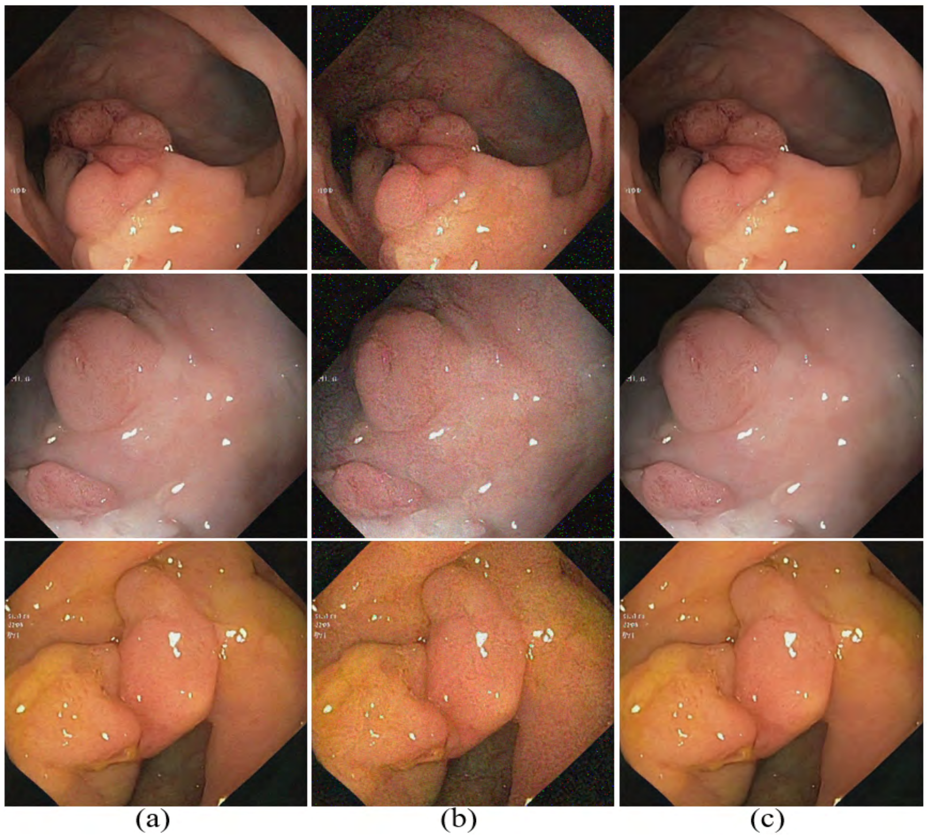

Fig. 1: High-resolution visualization of (a) sampled data $\tilde{\boldsymbol{x}}$, (b) the result of PGD-attack, and (c) the result of uncertainty considering PGD-attack with SDEdit on $\tilde{\boldsymbol{x}}$.

2 Ablation Studies on γ and K

Table 1: (Left) Ablation study on the PGD attack magnitude γ. (Right) Ablation study on the perturbation steps K. Kvasir-SEG dataset with U-Net was used.

γ	mIoU	mDice
0.1	92.71	96.22
0.5	92.90	96.32
1.0	**93.10**	**96.43**
1.5	92.97	96.35
2.0	92.91	96.32

K	mIoU	mDice
1	92.676	96.199
5	92.831	96.282
10	**93.10**	**96.43**
15	92.904	96.322

3 Pseudo Code

Algorithm 1 Hierarchical mask-to-image sampling scheme

Require: Bernoulli diffusion denoiser $\boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, t)$, conditional Gaussian diffusion denoiser $\boldsymbol{\epsilon}_\phi(\boldsymbol{x}_t, \boldsymbol{y}, t)$, diffusion timesteps T, T', Bernoulli noise scale β_t, Gaussian noise variance β'_t.

$\boldsymbol{y}_T \sim \mathcal{B}(\boldsymbol{y}_T; \frac{1}{2} \cdot \mathbf{1})$
for $t = T$ to 1 **do**
  $\hat{\boldsymbol{y}}_0 = \boldsymbol{y}_t \oplus \boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, t)$ ▷ Estimated $\boldsymbol{y}_0$ at timestep t
  $\bar{\beta}_t := \prod_{s=1}^{t}(1-\beta_s)$
  $\mathbf{Z} = [(1-\beta_t)\boldsymbol{y}_t + \beta_t/2] \odot \left[\bar{\beta}_{t-1}\hat{\boldsymbol{y}}_0 + (1-\bar{\beta}_{t-1})/2\right]$
    $+ [(1-\beta_t)(1-\boldsymbol{y}_t) + \beta_t/2] \odot \left[\bar{\beta}_{t-1}(1-\hat{\boldsymbol{y}}_0) + \left(1-\bar{\beta}_{t-1}\right)/2\right]$
  $\mu_\theta(\boldsymbol{y}_t, t) = \left[[(1-\beta_t)\boldsymbol{y}_t + \beta_t/2] \odot \left[\bar{\beta}_{t-1}\hat{\boldsymbol{y}}_0 + (1-\bar{\beta}_{t-1})/2\right]\right]/\mathbf{Z}$
  $\boldsymbol{y}_{t-1} \sim \mathcal{B}(\boldsymbol{y}_{t-1}; \mu_\theta(\boldsymbol{y}_t, t))$ ▷ Sample from $p_\theta^M\left(\boldsymbol{y}_{t-1} \mid \boldsymbol{y}_t\right)$
end for
$\boldsymbol{x}_{T'} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T'$ to 1 **do**
  $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\bar{\beta}'_t := \prod_{s=1}^{t}(1-\beta'_s)$
  $\mu_\phi(\boldsymbol{x}_t, \boldsymbol{y}_0, t) = \frac{1}{\sqrt{1-\beta'_t}}(\boldsymbol{x}_t - \frac{\beta'_t}{\sqrt{1-\bar{\beta}'_t}}\boldsymbol{\epsilon}_\phi(\boldsymbol{x}_t, \boldsymbol{y}_0, t))$
  $\sigma_t = \frac{1-\bar{\beta}'_{t-1}}{1-\bar{\beta}'_t}(1-\beta'_t)$
  $\boldsymbol{x}_{t-1} = \mu_\phi(\boldsymbol{x}_t, \boldsymbol{y}_0, t) + \sigma_t \boldsymbol{z}$ ▷ Sample from $p_\phi^I(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y})$
end for
return $\boldsymbol{x}_0, \boldsymbol{y}_0$

Algorithm 2 Adversarial attack in DASP

Require: generated image $\tilde{\boldsymbol{x}}$, pixel-wise label $\tilde{\boldsymbol{y}}$, segmentation network $f(\cdot)$, PGD steps K, PGD noise magnitude γ, SDEdit steps T^s.

$\tilde{\boldsymbol{x}}^0 = \tilde{\boldsymbol{x}}$
$\mathcal{W} = g(f(\tilde{\boldsymbol{x}}))$
for $k = 0$ **to** $K-1$ **do**
  $\tilde{\boldsymbol{x}}_0^k = \text{SDEdit}(\tilde{\boldsymbol{x}}^k, \tilde{\boldsymbol{y}}, T^s)$
  $\tilde{P}^k = \sigma(f(\tilde{\boldsymbol{x}}_0^k))$
  $\mathcal{L}_{adv}(\tilde{P}^k; \tilde{\boldsymbol{y}}, \mathcal{W}) = \frac{1}{H \times W}\sum_{i=1}^{H \times W}\left[\tilde{\boldsymbol{y}}_i \times \mathcal{W}_i \times \ell_{\text{bce}}\left((\tilde{P}^k)_i, \tilde{\boldsymbol{y}}_i\right)\right]$
  $\tau^k = \text{argmax}_\tau \mathcal{L}_{adv}(\tilde{P}^k; \tilde{\boldsymbol{y}}, \mathcal{W}) = \gamma \mathtt{sign}(\nabla_{\tilde{\boldsymbol{x}}^k}\mathcal{L}_{adv})$
  $\tilde{\boldsymbol{x}}^{k+1} = \tilde{\boldsymbol{x}}^k + \tau^k = \tilde{\boldsymbol{x}}^k + \gamma \mathtt{sign}(\nabla_{\tilde{\boldsymbol{x}}^k}\mathcal{L}_{adv})$
end for
$\tilde{\boldsymbol{x}}_0^K = \text{SDEdit}(\tilde{\boldsymbol{x}}^K, \tilde{\boldsymbol{y}}, T^s)$
return $\tilde{\boldsymbol{x}}_0^K$