Supplemental Materials

Behzad Hejrati[1] ⋆, Soumyanil Banerjee[1] ⋆, Carri Glide-Hurst[2], and Ming Dong[1] ⋆⋆

[1] Department of Computer Science, Wayne State University, MI, USA
{b.hejrati,s.banerjee,mdong}@wayne.edu
[2] Department of Human Oncology, University of Wisconsin-Madison, WI, USA
glidehurst@humonc.wisc.edu

Algorithm 1 Training algorithm for cDAL

Definitions:
x_θ: Diffusion Model (generator), D: Discriminator, x_0: GT-Label, I: Image, T: Number of time steps, A_D: Attention Map
for batch of (I, x_0) in Dataset **do**
Sample $\varepsilon \sim \mathcal{N}(0, \mathbb{I}), z \sim \mathcal{N}(0, \mathbb{I})$
$\beta_{\min} = 0.1$, $\beta_{\max} = 20$; $\beta_t = 1 - e^{-\beta_{\min}(\frac{1}{T}) - \frac{1}{2}(\beta_{\max} - \beta_{\min})\frac{2t-1}{T^2}}$
$\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$
$t \sim \text{Uniform}(\{1, 2, \ldots, T\})$; $x_t = q(x_0, t, \epsilon)$
Take gradient step on $D(x_t, x_{t-1}, t)$ where, $x_{t-1} = q(x_0, t-1, \epsilon)$ & x_θ frozen
$\hat{x}_0 = x_\theta(x_t, t, z, I)$
Take gradient step on $D(x_t, \hat{x}_{t-1}, t)$ where, $\hat{x}_{t-1} = q(\hat{x}_0, t-1, \epsilon)$ & x_θ frozen
$A_D = (\Sigma_{i=1}^{C} F_i)/C$
$x_0^{att} = x_0 \odot A_D$
Take gradient step $\nabla_\theta \|x_0 - x_\theta(x_t^{att}, t, z, I)\|^2$; $x_t^{att} = q(x_0^{att}, t, \epsilon)$ & D frozen
end for

Algorithm 2 Sampling/Inference algorithm for cDAL

Input T: number of timestemps, I: Images
$x_T \sim \mathcal{N}(0, \mathbb{I})$
for t ← T to 1 **do**
$\beta_{\min} = 0.1$, $\beta_{\max} = 20$; $\beta_t = 1 - e^{-\beta_{\min}(\frac{1}{T}) - \frac{1}{2}(\beta_{\max} - \beta_{\min})\frac{2t-1}{T^2}}$
$\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$, $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$
$x_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_\theta(x_t, t, z, I)$, where $z \sim \mathcal{N}(0, \mathbb{I})$
end for
return x_0

⋆ Equal contribution
⋆⋆ Corresponding author

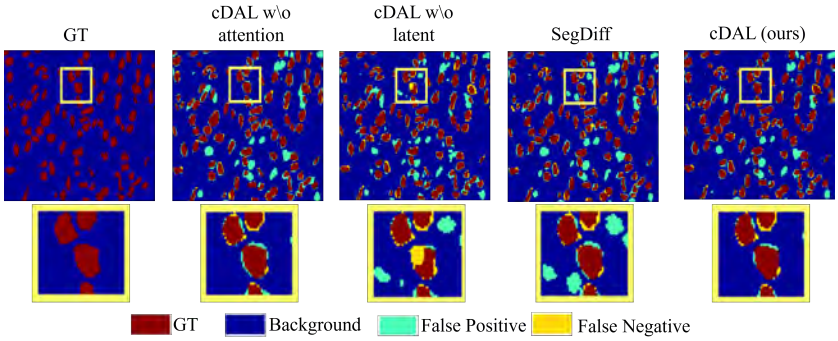

Fig. 1: Visualization of the Image, ground-truth (GT) and predictions with different models for MoNuSeg dataset, where the zoomed figures demonstrate the mispredictions of SegDiff, cDAL without the attention map A_D and cDAL without the random latent z.

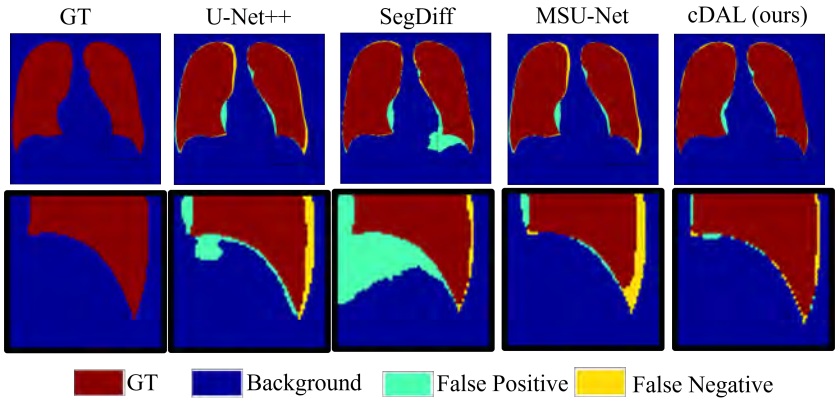

Fig. 2: Visualization of the Image, ground-truth (GT) and predictions with different models for the CXR dataset, where the zoomed figures demonstrate the mispredictions with other models.

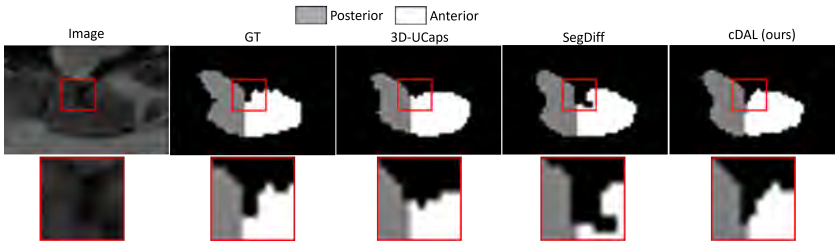

Fig. 3: Visualization of the Image, ground-truth (GT) and predictions with different models for Hippocampus dataset, where the zoomed figures demonstrate the mispredictions with other models.