Learning with Context Feedback Loop for Robust Medical Image Segmentation Supplementary Material

Kibrom Berihu Girum, Gilles Créhange, and Alain Lalande

1 Integration and training strategy of two networks through context feedback loop scheme

The forward and feedback system's integration and training strategy through the contextual feedback looping is shown in Algorithm 1. We trained the forward system (S) and the feedback system (F) alternatively for several times until it meets the early stop criteria or reaches the maximum number of epochs. We used the early stop of 100 for the convergence criteria. The number of training epochs was 500. All functions and symbols are defined in the main paper, training strategy section. For example, h_f and h_s are the latent space of the feedback system and forward system, respectively.

Algorithm 1 Pseudo code for training the LFB-Net

Input : Input image X_{data}, output label Y_{data} , maximum number of training epochs, convergence early stop
Output: Updated weights of forward system (S) w^s, and feedback system (F) w^f
repeat
  for *Number of batches over a training epoch* **do**
    Initialize h_{f_i} to zero tensors
    sample a mini-batch of training images $X_{batch} \sim X_{data}$
    sample a mini-batch of output labels $Y_{batch} \sim Y_{data}$
    predict $\hat{y}$ for X_{batch} and h_{f_i} with $S((X_{batch}, h_{f_i}); w_i^s)$
    $w_i^s \leftarrow$ propagate back the stochastic gradient $\nabla L_{total}(Y_{batch}, \hat{y})$
    predict $\hat{y}$ for X_{batch} and h_{f_i} with $S((X_{batch}, h_{f_i}); w_i^s)$
    predict $\hat{\hat{y}}$ for $\hat{y}$ with $F(\hat{y}; w_i^f)$
    $w_i^f \leftarrow$ propagate back the stochastic gradient $\nabla L_{total}(Y_{batch}, \hat{\hat{y}})$
  end
  for *Number of batches over a training epoch* **do**
    sample a mini-batch of training images $X_{batch} \sim X_{data}$
    sample a mini-batch of output labels $Y_{batch} \sim Y_{data}$
    predict $\hat{y}$ for X_{batch} with $S(X_{batch}; w_i^s)$
    predict h_{s_i} for X_{batch} with $S_e(X_{batch}; w_{e_i}^s)$
    predict h_{f_i} for $\hat{y}$ with $F_e(\hat{y}; w_{e_i}^f)$
    predict $\hat{y}_{i+1}$ for h_{s_i} and h_{f_i} with $S_d((h_{s_i}, h_{f_i}); w_{d_{i+1}})$
    $w_{d_{i+1}}^s \leftarrow$ propagate back the stochastic gradient $\nabla L_{total}(Y_{batch}, \hat{y}_{i+1})$
  end
  $w_i^s \leftarrow w_{i+1}^s$
until *Convergence or reached maximum number of training epochs*;