A THEORETICAL INSIGHT INTO THE GRADIENTS

As discussed in Section 5, the gradient of SRL with respect to a given pixel in the predicted mask remains constant throughout the training process, which is given by:

$$\frac{\partial \mathscr{L}_{SRL}}{\partial s_\theta^{jk}} = -\frac{1}{|K|} \frac{y^{jk}}{\sum_{i \in \Omega} y^{ik}}$$

Since the masks do not change over the epochs, the value of y^{ik}, which is the value of i^{th} pixel in tubed skeleton mask for k^{th} class, remains constant for a given sample. The gradient value, which backpropagates in the neural network, is independent of predicted pixel values s_θ^{jk}. Now, each pixel in tubed skeleton mask, y^{ik} can either belong to foreground or to background. So it can assume only two values: 0 and 1. Hence, the gradient of SRL boils down to a simple function:

$$\frac{\partial \mathscr{L}_{SRL}}{\partial s_\theta^{jk}} = \begin{cases} \frac{1}{\sum_{i \in \Omega} y^{ik}}, & \text{if } y^{jk} = 1, \\ 0, & \text{if } y^{jk} = 0. \end{cases}$$

So, irrespective of the segmentation prediction, SRL loss backpropagates a constant value through i^{th} pixel, if its corresponding pixel in tubed skeleton mask is part of the skeletonized region (foreground) else not. The loss does not include information about the current state of model weights or predictions which is the root cause of inefficiency of SRL.

We further note that each predicted pixel can only belong to one of the four categories: True Positive (TP), True Negative (TN), False Positive (FP) or False Negative (FN). Note that here we refer to a pixel as belonging to a certain category based on its comparison with the original ground truth mask, and not the skeletonized mask. We present an analysis of the gradient of SRL for each category of predicted pixel.

1. *True positive(TP)*
   In case the predicted pixel was a TP, we can subdivide the discussion into further two cases. The skeleton is a thinner version of the ground mask, and therefore a pixel aligning with the positives in the original ground truth mask may or may not align with the positives of the skeletonized mask. Now two cases are possible:
   - **The pixel aligns with the positives in the skeletonized mask.** In that case $y^{jk} = 1$. Therefore, a constant gradient keeps flowing throughout the training process, which eventually keeps pushing the parameters θ for this pixel further away from the already achieved optimal values leading to an increase in FNR.
   - **The pixel does not align with the positives in the skeletonized mask.** In this case, $y^{jk} = 0$. Therefore, the parameters θ affecting this pixel value would only be trained by the generic loss functions and would not experience any alteration due to indulgence of SRL.
2. *True negative(TN)*
   In case the predicted pixel was a TN, it would always align with the negative pixels of the skeletonized masks. Therefore, $y^{jk} = 0$ always, implying that the gradient flowing through the parameters of this pixel are unaffected by SRL.
3. *False positive(FP)*
   If the predicted pixel is a FP, then the pixel definitely aligns with the negative region of the original ground truth mask as well as the skeletonized mask. Similar to case 2, $y^{jk} = 0$ here and hence gradient flowing due to SRL does not affect the parameters of this pixel.
4. *False negative(FN)*
   Similar to case 1, if a predicted pixel turns out to be a FN, it's effect can be studied as two separate cases.
   - **The pixel aligns with positives of the skeletonized mask.** In that case, $y^{jk} = 1$, and hence a constant gradient would keep affecting the parameters associated with the chosen pixel. So, even when the generic losses try to alter the parameters in the right direction, SRL keeps pushing the gradient in some absurd direction, hence reducing

the efficiency of the generic loss functions, and therefore contributing in enhancing the FNR.

- **The pixel does not align with the positives of the skeletonized mask.** Here, $y^{jk} = 0$ which implies that the SRL does not affect the overall gradient of the loss function.

Note that predicting a pixel to be a positive either does not affect the SRL in case of a FP, or it reduces the value of the overall loss heavily in case of a TP due to combined effects of multiple loss functions. Therefore, SRL does not penalize for a FP but rewards for a TP. These incentives push the model to predicting more positives. Also, in case of a positive prediction, the gradient of SRL affects the parameters only in one of the three possible cases (2 parts of case 1 and case 3), implying that SRL does not take care of the FP predictions. These factors lead to an increase in the FPR.

B Additional Results

We tried to assess the effect of SRL on additional datasets. Two benchmark multi-class classification based datasets were employed for this purpose. We utilized the Automatic Cardiac Diagnosis Challenge (ACDC) (Bernard et al., 2018), which is a popular non-tubular 3D dataset for research in fields related to cardiac diagnosis and includes 150 multi-equipments CMRI recordings with task of segmenting the Left Ventricle, Right Ventricle and Myocardium dataset. BoMBR (Raina et al., 2024), a tubular, 2D dataset, consists of histology images for segmenting different components of a bone marrow biopsy report. Both datasets involve structures of absurd shapes and sizes, testing the effect of SRL in alternate domains.

Table 2: Model performances using different loss functions on the BoMBR and ACDC datasets.

Method	**DSC** $\uparrow$	**clDice** $\uparrow$	**JSI** $\uparrow$	**FNR** $\downarrow$	**FPR** $\downarrow$
BoMBR Dataset (Raina et al., 2024)					
Dice + CCE	**78.99**	**76.22**	**72.52**	**0.16**	**0.10**
Dice + CCE + SRL	67.54	59.98	57.81	0.24	0.16
ACDC Dataset (Bernard et al., 2018)					
Dice + CCE	**92.15**	**95.99**	**85.76**	**0.08**	**0.0008**
Dice + CCE + SRL	91.99	94.89	85.54	**0.08**	**0.0008**

Table 2 illustrates that within a dataset composed of simple shapes—specifically, the blobs in the ACDC dataset—the application of SRL does not yield any significant performance improvements. Rather, its usage is counter productive as it hinders the functioning of the base model. While, in case of a dataset with a large variety of complex segmentation mask shapes, the topology based loss fails miserably. It is also a proof of the limited scope of SRL in use cases as it can only be used in dataset comprised purely of tubular segments.

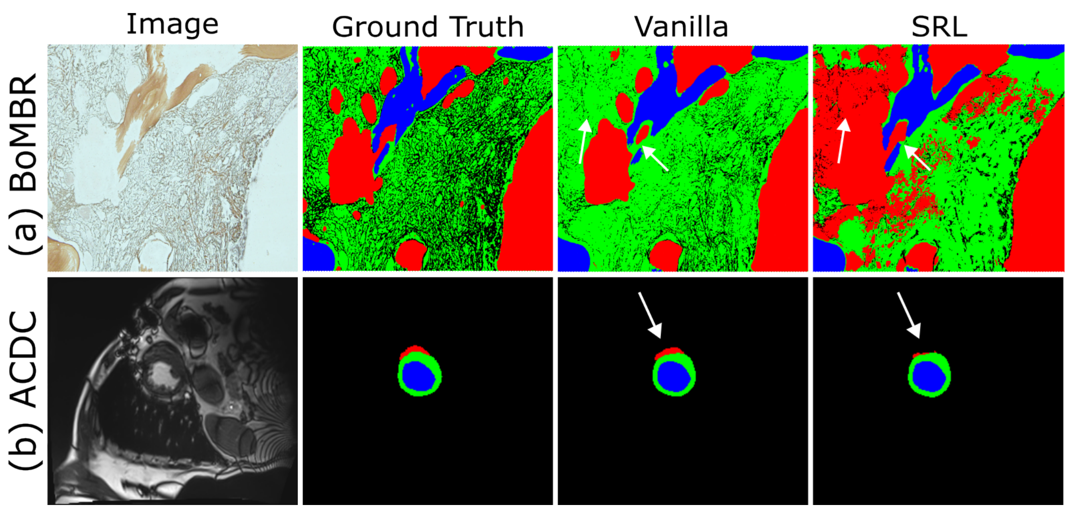

Figure 3: Results of proposed method over 2 other datasets of alternate domains, namely- (a) BoMBR (Raina et al., 2024)d and (b) ACDC (Bernard et al., 2018). A similar challenge is encountered here as with the tubular structure datasets (2). Specifically, in the ACDC dataset, the nnUNet model fails to detect smaller regions, as demonstrated in our example. In contrast, for the BoMBR dataset, the SRL model significantly deviates from expected performance, and substantially over-predicts the region corresponding to a single class.