Semantic Segmentation with Spreading Scribbles

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Hand-annotating medical images with segmentation masks requires an immense amount of time and effort from clinical experts. Replacing full masks with a simpler annotating gesture can mitigate annotation costs. This can come in the form of a scribble, and leads to weakly supervised training scenarios. Scribble-supervised segmentation typically utilises advanced neural architectures to compensate for the limited training data. Instead of just relying strictly on the pixels from each scribble, we also enhance each scribble by spreading, i.e. propagating, annotation labels through the image. We use a hierarchical partitioning of the image, produced with watershed/waterfall transforms, and propagate the individual pixel labels through the waterfall regions. We propose that a semantic label can be propagated to all other pixels in the same waterfall region. This increases the number of pixels that can be used for training supervision. We show experimentally that this technique greatly boosts the performance of established neural architectures on public semantic segmentation datasets like ACDC and MSCMRseg.

Keywords: Scribble-supervised segmentation · Waterfall transform · Spreading scribbles.

1 Introduction

Training deep learning (DL) models for medical image segmentation typically requires large-scale, pixel-wise annotated datasets, which are expensive and time-consuming to obtain. To mitigate the annotation burden, various learning paradigms have been explored. Weakly supervised learning (WSL) [12, 22, 41] aims to train models using sparse annotations, such as image-level labels [23], bounding boxes [20], key points [1], and scribbles [17]. Among these weak annotation types, scribbles provide a compelling balance between annotation efficiency and segmentation quality [17, 28]. As a result, scribble-based supervision has gained traction in medical image segmentation, offering a practical alternative to dense manual labeling [12, 14].

Scribble-Supervised Segmentation. Scribble annotations provide a cost-effective alternative to pixel-wise labeling, reducing the annotation burden. Established techniques such as GraphCuts [3], GrabCut [25], Random Walker [9], GrowCut [30], ITK-SNAP [38], Slic-Seg [39], propagate sparse annotations into dense segmentation masks. More recently, DL-based methods have used scribble

annotations to train segmentation models. Lin et al. [17] propose a graphical-model-based method to expand scribbles, while Tang et al. [27] integrate conditional random field (CRF) regularisation as a loss function to refine segmentation predictions. Can et al. [4] propose to iteratively refine predictions with CRF post-processing to enhance segmentation accuracy from scribbles.

Although these methods are effective, they often require iterative relabeling or additional constraints to improve the quality of supervision. Alternative strategies include leveraging adversarial training [28] or uncertainty-aware mean teachers [18] to enforce structural consistency in segmentation masks. These strategies primarily focus on refining network predictions rather than directly enhancing the scribble annotations themselves. Consequently, the quality of the initial annotations remains a limiting factor.

Data Augmentation for Semantic Segmentation. Data augmentation plays a crucial role in mitigating overfitting and improving generalisation in DL models. Conventional augmentation techniques such as rotation, flipping, and elastic deformations have been widely employed in medical image segmentation. More advanced augmentation methods, such as mixup strategies [39], involve blending multiple images and their labels to generate new training samples. Manifold MixUp [29] extends this approach to intermediate feature representations, while CutMix [37] replaces image regions with patches from other samples to enhance diversity. In the context of weakly supervised learning, augmentation techniques have been adapted to improve performance. Chaitanya et al. [6] demonstrate that mixup augmentation can enhance semi-supervised segmentation by enforcing smooth transitions in label space. All these methods primarily focus on image-level augmentations rather than directly improving weak annotations such as scribbles.

Enhancement of Weak Annotations. While prior works have explored augmentation techniques for image data, fewer studies have focused on the enhancement of weak annotations. Most existing methods either attempt to refine scribble-based annotations through graphical propagation [17] or employ DL models to generate pseudo-labels [15, 18]. However, these approaches often require additional supervision or iterative training, limiting their efficiency.

Our Contribution. Our algorithm directly enhances an initial set of scribble annotations before using them for model training. Instead of relying on DL-based relabelling, our method spreads scribble annotations systematically across the image, improving their spatial coverage and effectiveness as weak labels. By generating scribble-based training labels in this way, we demonstrate that deep segmentation models trained on our enhanced labels mostly outperform those using raw scribble annotations. Our approach provides a novel alternative to existing weak annotation techniques by enhancing the initial annotation quality rather than relying on iterative refinement or additional network components.

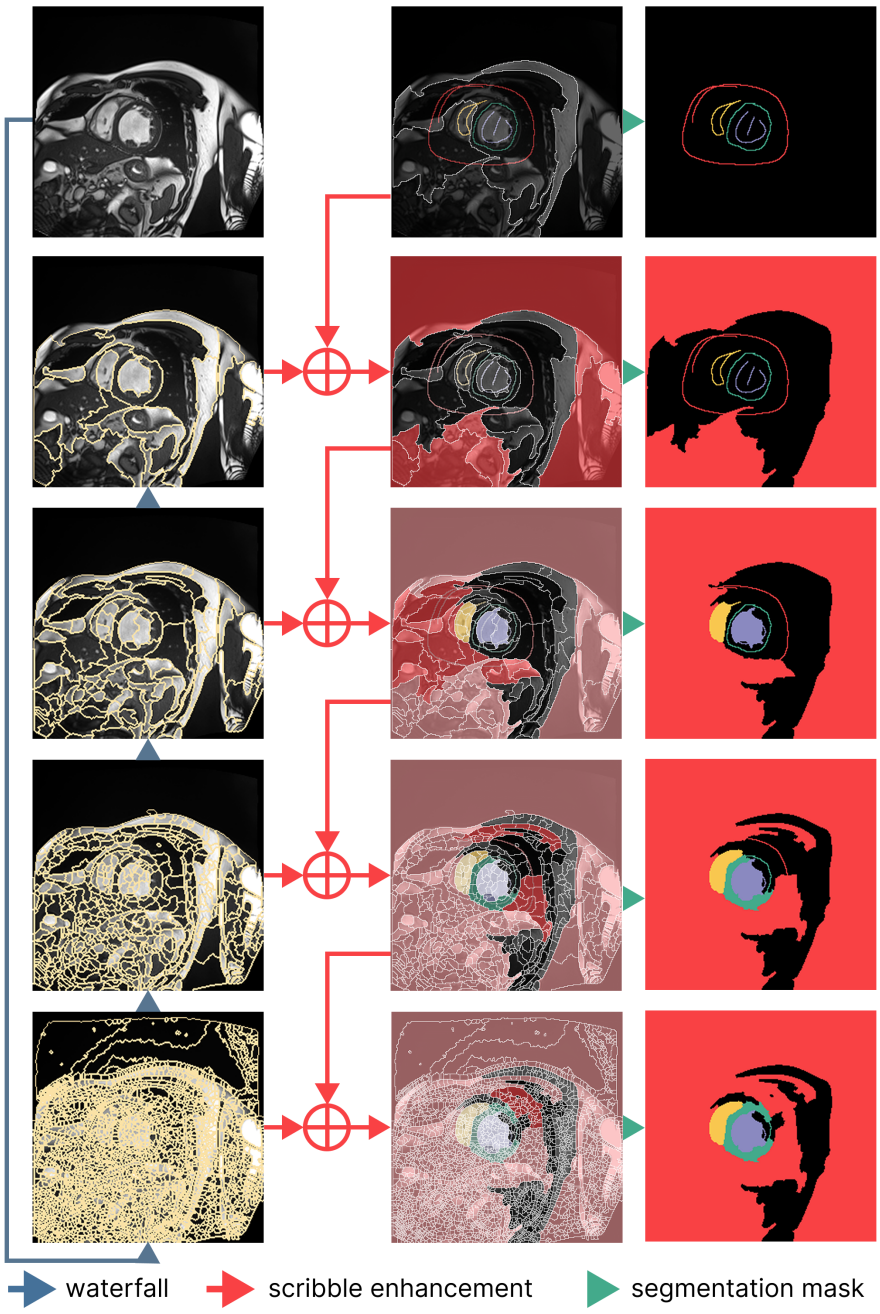

Fig. 1. Visualisation of the scribble spreading algorithm. Waterfall layers are constructed for the image. From the coarsest partitioning layer down to the finest, scribble pixel labels are propagated to all pixels in the same region, if no conflicting label is present. The bright overlay shows new labels; the pale overlay—the previous iteration.

2 Spreading Scribbles Using Hierarchical Waterfall Partitioning

We propose a novel technique for spreading scribbles across an image to construct a spatially enhanced set of labels. The scribble spreading process is visualised in Fig. 1. Our approach relies on hierarchical partitioning of the image into semantically meaningful regions using the waterfall transform. The waterfall transform [2, 8] is a hierarchical image partitioning approach that is based on the watershed transform [2, 24]. These transforms utilise the variations in the gradient magnitude values across the image to partition it into regions of similar-intensity pixels. There are fast parallel algorithms for GPU execution of the watershed and waterfall transforms [36, 7, 34].

Algorithm 1 Scribble Spreading Using Waterfall Transform

Input: Image I with pixel domain Ω and scribble mask S
Output: Enhanced mask E
Let value *unknown* indicate unlabeled pixels in S
Initialise $E = S$
Construct hierarchical waterfall partitioning $WF = \{L_0, L_1, L_2, L_3, L_4\}$
Reverse layer sequence $\overline{WF} = \{L_4, L_3, L_2, L_1, L_0\}$
for each $L \in \overline{WF}$ **do**
    Initialize an empty region-to-label map Map
    for each pixel $p \in \Omega$ **do**
        if $S(p) \neq unknown$ **then**
            if $L(p) \notin Map$ **then**
                $Map(L(p)) = S(p)$
            else if $Map(L(p)) \neq S(p)$ **then**
                $Map(L(p)) = unknown$
            end if
        end if
    end for
    for each pixel $p \in \Omega$ **do**
        if $E(p) = unknown$ **and** $L(p) \in Map$ **and** $Map(L(p)) \neq unknown$ **then**
            $E(p) = Map(L(p))$
        end if
    end for
end for
return E

The proposed approach for constructing enhanced labels by spreading scribbles is presented in Alg. 1. For each 2D slice in a 3D medical image volume, our approach constructs a hierarchical partitioning of the image into regions of decreasing granularity by applying the waterfall transform for five layers. The layers are then considered in reverse order—from coarsest to finest. Per partitioning layer L, for each pixel p with a scribble label $S(p)$, its label is considered for spreading throughout the whole partitioning region. If there is another pixel

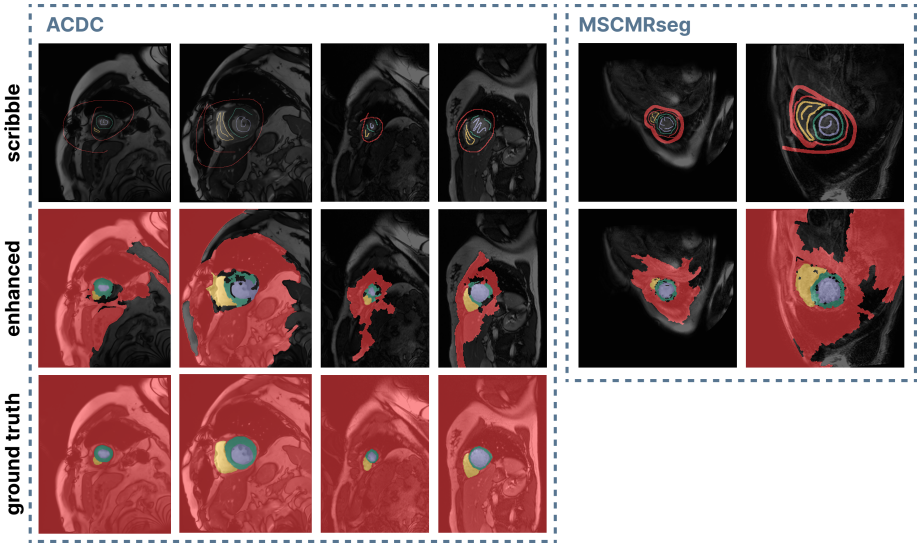

Fig. 2. Scribble spreading results from sample images from the ACDC and MSCMRseg datasets, compared against the ground truth where possible. Ground truth masks are not available for the training images in MSCMRseg. Enhanced labels are only approximations of the expected masks. They may contain inaccurate labels on a portion of pixels.

in the same region with a *different* scribble label, then none of the labels is spread. An efficient implementation of this logic is achieved by maintaining a mapping from regions into corresponding label candidates. Eventually, regions with a *unique* label candidate are 'flooded' with that label. By considering the layers in reverse order, we spread scribbles across larger regions first, then complement those with further incremental spreading in smaller regions. Figure 2 illustrates several examples of scribble spreading on slices from the ACDC and MSCMRseg datasets.

2.1 Limitations

The proposed scribble spreading algorithm has two main limitations.

- When no labels are present or multiple conflicting labels are present in the same partitioning region, no label propagation is performed. This may result in swathes of the image left without labels. This effect can be seen especially in the third, fourth, and fifth columns of Fig. 2.
- Since partitioning region boundaries do not necessarily align with semantic boundaries, enhanced labels only approximate the expected masks and may contain inaccurate labels on a portion of pixels. In Fig. 2, inaccuracies around region boundaries can be easily traced when comparing the enhanced labels and ground truth masks for ACDC samples.

3 Experimental Setup

3.1 Datasets

We evaluate the performance of our label enhancement framework on the two scribble-based datasets below.

ACDC dataset consists of cine-MRI scans of 150 patients. Scribble annotations are provided for 100 patients and used to generate the training and validation sets. We use a five-fold cross-validation method, allocating 80 scans for training and 20 for validation. The remaining 50 cases, which do not have scribble annotations, are used as the test set.

MSCMRseg dataset consists of late gadolinium enhancement MRI scans, from 45 patients diagnosed with cardiomyopathy. The dataset is split into 25 scans for training, 5 scans for validation, and 15 scans for testing. We aggregate and report results from five different runs.

Both datasets are annotated for three classes: left ventricle (LV), right ventricle (RV), and myocardium (MYO). All reported results are averages across the three classes.

3.2 Scribble-Supervised Segmentation Algorithms

To demonstrate the effectiveness of our scribble-spreading method, we conduct a comprehensive comparison of the performance of different state-of-the-art (SOTA) methods trained on the original scribbles versus those trained on the enhanced labels produced by spreading scribbles with our algorithm. In this experiment we include the following methods.

A Baseline fully-supervised segmentation methods including ENet [21], CCT-UNet [33], DS-UNet [11], Efficient UNet [26], PNet [31] and SwinNet [5]. All of these methods are created for the task of medical image segmentation and use either UNet-like architectures or transformers.

B Several scribble-supervised loss strategies for UNet from the literature, such as partial cross entropy (UNet_{pce}) [17], uncertainty-aware self-ensembling and transformation-consistent loss (UNet_{ustr}) [18], or Mumford–Shah loss (UNet_{mloss}) [13], entropy minimization (UNet_{em}) [10], and gated conditional random field (UNet_{crf}) [40].

C Pseudo-label-based methods such as ScribbleToLabel (S2L) [15], dynamically mixed pseudo labels supervision (DMPLS) [19], dense combinations of dense pseudo-labels (DCDPL) [32], and ScribbleVC [16].

3.3 Evaluation Measures

For evaluation purposes during the testing phase, the 2D slice-by-slice predictions are stacked into a 3D volume. The model used for inference is the one that achieved the best validation score during training, ensuring optimal performance.

To ensure a fair comparison, no post-processing methods are applied. We use the 3D Dice similarity coefficient (DSC), the 95^{th} percentile of the Hausdorff distance (HD95), and Average surface distance (ASD) [35] to evaluate the similarity and differences between the predicted mask and the ground truth. HD95 and ASD are measured in mm. Good performance is reflected in large DSC and small HD95 and ASD.

3.4 Implementation Details

For the experimental setup, we followed the approach proposed in [19]. As a pre-processing step, the intensities of each slice were re-scaled to the range $[0, 1]$. Subsequently, data augmentation techniques such as random rotation, random flipping, and random noise were applied to enhance the training set. The batch size was set to 12, and the total number of iterations was 60k. All methods were implemented using PyTorch and executed on machines equipped with GeForce RTX 3060 and GTX 1080 GPUs.

4 Results and Discussion

4.1 Qualitative Results Analysis

A qualitative comparative analysis of the performance of six segmentation methods (UNet_{crf}, DCDPL, S2L, UNet_{em}, ENet, CCT-UNet) on the ACDC dataset is presented in Fig. 3. For each image and each method, results for both original scribbles and enhanced labels are shown. Clear false-positive mistakes with the original setup, but not the enhanced setup, can be seen for S2L, UNet_{em}, ENet, CCT-UNet. Across most architectures, enhanced labels lead to improvement in boundary delineation, evident by the improved HD95 score. However, the degree of improvement significantly varies depending on the model.

Models like CCT-UNet and ENet exhibit the most substantial improvements when using enhanced labels, reducing ASD and HD95 by over 100 mm for all three example volumes and increasing DSC by at least 30%. This indicates that these architectures are highly dependent on the amount of provided scribbles to achieve high-quality segmentation. Similar performance improvements are observed with the UNet_{em} and S2L models, especially in examples (b) and (c). Although for volume (a) enhanced labels show slightly worse performance, the reduction is marginal.

For models UNet_{crf} and DCDPL, enhancement of labels leads to degradation of scores, in particular DSC. This implies that these models are more robust to the amount of scribble labelling provided. However, UNet_{crf} still generally benefits from spreading scribbles, due to improvements in boundary precision (lower HD95). In case of DCDPL, scribble spreading worsens performance on all three volumes, possibly due to the introduction of inaccurate labels during the enhancement process.

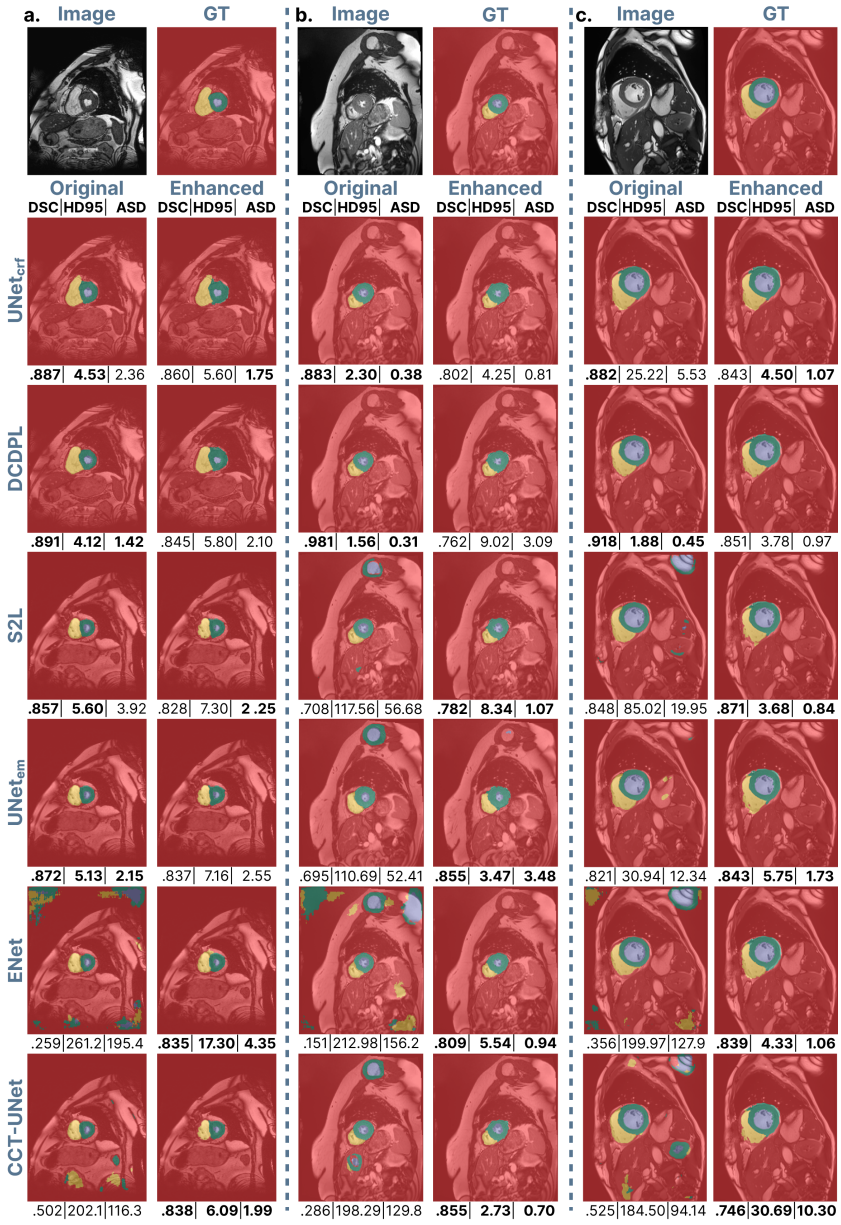

Fig. 3. Comparison of segmentation performance using original scribbles and enhanced labels on three scans (a, b, c) from ACDC. The first row displays the original MRI slice and GT, while subsequent rows show segmentation results from different models: UNet$_{crf}$, DCDPL, S2L, UNet$_{em}$, ENet, and CCT-UNet. DSC, ASD, and HD95 scores for the whole 3D scan are reported for both original and enhanced training setups. Clear false-positive mistakes with the original setup, but not the enhanced setup, can be seen for S2L, UNet$_{em}$, ENet, CCT-UNet.

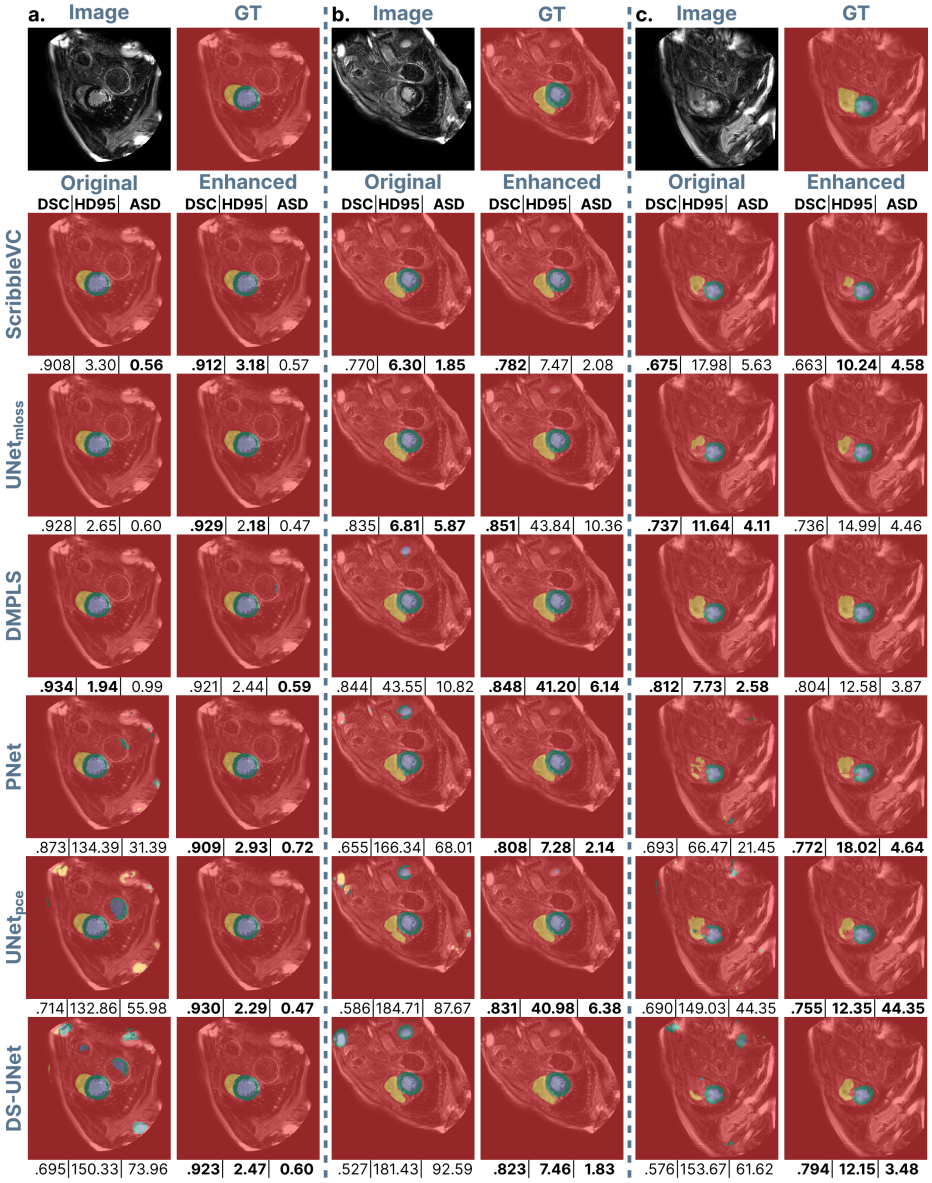

Fig. 4. Comparison of segmentation performance using original scribbles and enhanced labels on three scans (a, b, c) from MSCMRseg. The first row displays the original MRI slice and GT, while subsequent rows show segmentation results from different models: ScribbleVC, UNet$_{mloss}$, DMPLS, PNet, UNet$_{pce}$, and DS-UNet. DSC, ASD, and HD95 scores for the whole 3D scan are reported for both original and enhanced training setups. Clear false-positive mistakes with the original setup, but not the enhanced setup, can be seen for PNet, UNet$_{mloss}$, DS-UNet.

The qualitative results on the MSCMRseg dataset (Fig. 4) show six more methods (ScribbleVC, UNet$_{mloss}$, DMPLS, PNet, UNet$_{pce}$, DS-UNet), complementary to those in Fig. 3 on ACDC. Notably, for most image–method combinations the performance in the enhanced setup shows a clear performance improvement over the original scribble setup. Compared to ACDC, fewer cases show performance degradation for MSCMRseg. This can be potentially explained by the MSCMRseg training set consisting of around three times fewer MRI volumes than the ACDC training set. The enhancement of labels through spreading scribbles boosts the amount of supervision data, even if not completely reliably.

While UNet$_{mloss}$ does exhibit some regression of boundary measures (HD95 and ASD) in volume (b), the effects are not severe but are isolated. This may suggest that the relationship between scribble quality and model performance varies across different medical imaging modalities and anatomical targets.

4.2 Quantitative Results Analysis

The full performance of all segmentation methods, with the original and enhanced setups, is reported in Table 1 for ACDC and Table 2 for MSCMRseg. It can be seen that using the enhanced setup improves performance according to at least two of the three performance measures for twelve and ten methods for ACDC and MSCMRseg, respectively. In some cases, the DSC improvement reaches more than 45%; HD95 and ASD go down from three-digit to single-digit numbers.

In general, in the enhanced setup, the performance scores of the different algorithms are very close: most DSC are above 0.8, ASD is around 6–9 mm and HD95 is around 1–3 mm.

Using the enhanced labels mostly degrades the performance of UNet$_{crf}$, DCDPL, and ScribbleVC. The latter two are SOTA pseudo-label-based methods. We can hypothesise that the training guidance provided by the pseudo-labels does not align with the enhanced labels in these particular cases. More effort is required to identify how these two techniques can be combined to improve performance. In case of UNet$_{crf}$, the assumption is that the potential presence of mislabelled pixels in the enhanced setup mislead the network when used with the *crf* loss.

For ACDC, the best DSC score of .883 is achieved with the original setup of DMPLS; ASD of 5.48 mm and HD95 of 1.39 mm with the original setup of DCDPL. For MSCMRseg, the original setup of DMPLS gets the best DSC of .886; the best ASD of 5.88 mm and the best HD95 of 1.67 mm is achieved with the *enhanced* setup of PNet. Thus, baseline methods originally proposed for fully supervised segmentation can outperform pseudo-label-based methods specifically designed for scribble-supervised learning, if enhanced labels, produced by spreading scribbles with our algorithm, are used for training.

	Original			Enhanced		
	DSC	**ASD**	**HD95**	**DSC**	**ASD**	**HD95**
ENet [21]	.508 ±.152	178.16 ±31.39	91.31 ±33.56	**.818** ±.028	**7.49** ±1.91	**2.17** ±0.72
CCT-UNet [33]	.602 ±.046	168.83 ±6.27	77.48 ±7.90	**.835** ±.005	**8.55** ±2.01	**2.54** ±0.67
DS-UNet [11]	.628 ±.041	169.80 ±8.09	75.07 ±10.25	**.832** ±.013	**11.09** ±6.30	**2.95** ±1.68
Efficient UNet [26]	.780 ±.003	85.01 ±11.27	27.86 ±3.24	**.839** ±.007	**6.78** ±0.92	**1.94** ±0.30
PNet [31]	.813 ±.010	79.87 ±10.64	22.27 ±4.09	**.827** ±.008	**8.43** ±1.91	**2.57** ±0.55
SwinUNet [5]	**.851** ±.000	35.18 ±0.00	9.30 ±0.00	.827 ±.000	**7.38** ±0.00	**2.12** ±0.00
Different scribble-supervised strategies on UNet						
UNet_{pce} [17]	.631 ±.003	165.13 ±5.14	72.35 ±5.56	**.837** ±.006	**9.58** ±2.04	**2.62** ±0.56
UNet_{ustr} [18]	.815 ±.020	77.68 ±21.29	23.48 ±6.52	**.826** ±.009	**10.41** ±3.42	**2.72** ±0.87
UNet_{mloss} [13]	**.864** ±.010	13.33 ±4.55	3.61 ±1.12	.834 ±.009	**8.00** ±1.45	**2.03** ±0.30
UNet_{em} [10]	**.853** ±.010	37.36 ±5.03	10.52 ±1.27	.837 ±.006	**9.58** ±1.17	**2.62** ±0.37
UNet_{crf} [40]	**.869** ±.001	**6.99** ±1.25	**1.92** ±0.31	.829 ±.006	8.39 ±2.86	2.34 ±0.83
Pseudo-label-based methods						
S2L [15]	**.846** ±.001	44.16 ±8.4	12.81 ±2.43	.838 ±.011	**7.30** ±1.53	**1.97** ±0.49
DMPLS [19]	**.883** ±.001	9.19 ±1.40	2.41 ±0.21	.849 ±.000	**8.01** ±0.00	**2.10** ±0.00
DCDPL [32]	**.874** ±.020	**5.48** ±1.07	**1.39** ±0.33	.824 ±.014	6.56 ±1.50	1.88 ±0.45
ScribbleVC [16]	**.860** ±.009	**5.94** ±0.57	**1.42** ±0.21	.817 ±.010	6.49 ±0.66	1.58 ±1.58

Table 1. Comparison of various segmentation methods on the ACDC dataset using original scribbles and enhanced labels. The performance of the models is evaluated using DSC, ASD, and HD95 scores. The models are categorised into three groups: general segmentation models, scribble-supervised UNet variants, and models utilising pseudo-labeling to enhance weak supervision. In each row, the bold number indicates the best performance. *Note: some reported results are based on less than five runs for now. These will be finalised in the camera-ready version.*

	Original			Enhanced		
	DSC	**ASD**	**HD95**	**DSC**	**ASD**	**HD95**
ENet [21]	.473 ±.030	185.76 ±1.89	110.82 ±0.98	**.836** ±.034	**6.26** ±1.41	**1.95** ±0.63
CCT-UNet [33]	.682 ±.019	138.44 ±11.93	58.73 ±7.78	**.858** ±.016	**10.04** ±3.71	**2.71** ±1.05
DS-UNet [11]	.666 ±.016	143.57 ±3.55	64.57 ±4.12	**.871** ±.005	**9.89** ±3.70	**2.81** ±1.02
Efficient UNet [26]	.705 ±.000	135.86 ±0.00	53.02 ±0.00	**.863** ±.000	**6.70** ±0.00	**1.83** ±0.00
PNet [31]	.794 ±.000	111.89 ±0.00	33.45 ±0.00	**.865** ±.000	**5.88** ±0.00	**1.67** ±0.00
SwinUNet [5]	**.720** ±.000	31.99 ±0.00	10.18 ±0.00	.700 ±.000	**10.79** ±0.00	**3.47** ±0.00
Different scribble-supervised strategies on UNet						
UNet_{pce} [17]	.670 ±.0026	143.67 ±4.47	65.00 ±8.87	**.869** ±.006	**8.30** ±3.91	**2.39** ±0.90
UNet_{ustr} [18]	.758 ±.004	130.84 ±5.95	50.49 ±1.42	**.866** ±.003	**7.30** ±1.52	**2.40** ±0.69
UNet_{mloss} [13]	**.870** ±.004	**9.01** ±3.88	**2.47** ±0.62	.865 ±.004	9.96 ±3.35	2.61 ±0.89
UNet_{em} [10]	.859 ±.005	34.38 ±6.73	9.57 ±2.15	**.864** ±.004	**8.45** ±1.98	**2.58** ±0.42
UNet_{crf} [40]	**.877** ±.004	**6.72** ±2.05	**1.85** ±0.31	.864 ±.003	7.10 ±1.54	1.98 ±0.37
Pseudo-label-based methods						
S2L [15]	.840 ±.015	67.98 ±14.19	21.08 ±5.97	**.846** ±.008	**59.61** ±12.46	**17.95** ±3.61
DMPLS [19]	**.886** ±.002	**8.81** ±0.99	**2.69** ±0.60	.877 ±.005	10.54 ±2.56	3.05 ±1.00
DCDPL [32]	**.854** ±.022	**8.63** ±0.95	2.51 ±0.20	.777 ±.014	9.24 ±1.88	**2.38** ±0.38
ScribbleVC [16]	**.851** ±.022	**6.58** ±0.95	**1.69** ±0.20	.820 ±.014	7.19 ±1.88	1.85 ±0.38

Table 2. Comparison of various segmentation methods on the MSCMRseg dataset using original scribbles and enhanced labels. The performance of the models is evaluated using DSC, ASD, and HD95 scores. The models are categorised into three groups: general segmentation models, scribble-supervised UNet variants, and models utilising pseudo-labeling to enhance weak supervision. In each row, the bold number indicates the best performance. *Note: some reported results are based on less than five runs for now. These will be finalised in the camera-ready version.*

5 Conclusion

Since gesture-based scribble annotations are more lax and less time-consuming than contour drawing, these make semantic segmentation tasks more easily achievable with less effort. The additional cost is that of pre-processing the images into waterfall regions, which is a fully automated task requiring minimal (parallel) compute. Scribble-based segmentation is best suited to tasks where the boundary of the feature is clear and without further aleatoric uncertainty.

Future work will explore how our scribble-spreading method can be better utilised in pseudo-label-based models to further improve performance.

References

1. Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L.: What's the point: Semantic segmentation with point supervision. In: Computer Vision – ECCV 2016. pp. 549–565. Springer (2016)
2. Beucher, S.: Watershed, hierarchical segmentation and waterfall algorithm. In: Mathematical Morphology and Its Applications to Image Processing, pp. 69–76. Springer (1994)
3. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 1, pp. 105–112 (2001)
4. Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., Baumgartner, C.F.: Learning to segment medical images with scribble-supervision alone. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018. pp. 236–244. Springer (2018)
5. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Computer Vision – ECCV 2022 Workshops. pp. 205–218. Springer (2023)
6. Chaitanya, K., Karani, N., Baumgartner, C.F., Becker, A., Donati, O., Konukoglu, E.: Semi-supervised and task-driven data augmentation. In: Information Processing in Medical Imaging. IPMI 2019. pp. 29–41. Springer (2019)
7. Gabrielyan, Y., Yeghiazaryan, V., Voiculescu, I.: Parallel partitioning: Path reducing and union–find based watershed for the GPU. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 1501–1505 (2022)
8. Golodetz, S.M., Nicholls, C., Voiculescu, I.D., Cameron, S.A.: Two tree-based methods for the waterfall. Pattern Recognition **47**(10), 3276–3292 (2014)
9. Grady, L.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(11), 1768–1783 (2006)
10. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: Advances in Neural Information Processing Systems. vol. 17. MIT Press (2004)
11. Huang, Y., Bian, S., Li, H., Wang, C., Li, K.: DS-UNet: A dual streams UNet for refined image forgery localization. Information Sciences **610**, 73–89 (2022)
12. Khoreva, A., Benenson, R., Hosang, J., Hein, M., Schiele, B.: Simple does it: Weakly supervised instance and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 876–885 (2017)
13. Kim, B., Ye, J.C.: Mumford–Shah loss functional for image segmentation with deep learning. IEEE Transactions on Image Processing **29**, 1856–1866 (2019)

14. Koch, L.M., Rajchl, M., Bai, W., Baumgartner, C.F., Tong, T., Passerat-Palmbach, J., Aljabar, P., Rueckert, D.: Multi-atlas segmentation using partially annotated data: methods and annotation strategies. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(7), 1683–1696 (2017)
15. Lee, H., Jeong, W.K.: Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020. pp. 14–23. Springer (2020)
16. Li, Z., Zheng, Y., Luo, X., Shan, D., Hong, Q.: ScribbleVC: Scribble-supervised medical image segmentation with vision-class embedding. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 3384–3393. ACM (2023)
17. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3159–3167 (2016)
18. Liu, X., Yuan, Q., Gao, Y., He, K., Wang, S., Tang, X., Tang, J., Shen, D.: Weakly supervised segmentation of COVID19 infection with scribble annotation on CT images. Pattern Recognition **122**, 108341 (2022)
19. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 528–538. Springer (2022)
20. Papandreou, G., Chen, L.C., Murphy, K.P., Yuille, A.L.: Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1742–1750 (2015)
21. Paszke, A., Chaurasia, A., Kim, S., Culurciello, E.: ENet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147 (2016)
22. Pathak, D., Krahenbuhl, P., Darrell, T.: Constrained convolutional neural networks for weakly supervised segmentation. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1796–1804 (2015)
23. Pathak, D., Shelhamer, E., Long, J., Darrell, T.: Fully convolutional multi-class multiple instance learning. arXiv preprint arXiv:1412.7144 (2014)
24. Roerdink, J.B., Meijster, A.: The watershed transform: Definitions, algorithms and parallelization strategies. Fundamenta Informaticae **41**(1-2), 187–228 (2000)
25. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics **23**(3), 309–314 (2004)
26. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. vol. 97, pp. 6105–6114. PMLR (2019)
27. Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised CNN segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 507–522 (2018)
28. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Transactions on Medical Imaging **40**(8), 1990–2001 (2021)
29. Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., Bengio, Y.: Manifold mixup: Better representations by interpolating hidden states. In: Proceedings of the 36th International Conference on Machine Learning. vol. 97, pp. 6438–6447. PMLR (2019)

30. Vezhnevets, V., Konouchine, V.: GrowCut: Interactive multi-label ND image segmentation by cellular automata. In: Proceedings of GraphiCon. pp. 150–156 (2005)
31. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(7), 1559–1572 (2019)
32. Wang, Z., Voiculescu, I.: Weakly supervised medical image segmentation through dense combinations of dense pseudo-labels. In: MICCAI Workshop on Data Engineering in Medical Imaging (DEMI 2023). pp. 1–10. Springer (2023)
33. Yan, Y., Liu, R., Chen, H., Zhang, L., Zhang, Q.: CCT-Unet: A U-shaped network based on convolution coupled transformer for segmentation of peripheral and transition zones in prostate MRI. IEEE Journal of Biomedical and Health Informatics **27**(9), 4341–4351 (2023)
34. Yeghiazaryan, V., Gabrielyan, Y., Voiculescu, I.: Parallel watershed partitioning: GPU-based hierarchical image segmentation. arXiv preprint arXiv:2410.08946 (2024)
35. Yeghiazaryan, V., Voiculescu, I.: Family of boundary overlap metrics for the evaluation of medical image segmentation. Journal of Medical Imaging **5**(1), 015006 (2018)
36. Yeghiazaryan, V., Voiculescu, I.: Path reducing watershed for the GPU. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 577–585 (2018)
37. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6023–6032 (2019)
38. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage **31**(3), 1116–1128 (2006)
39. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
40. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1529–1537 (2015)
41. Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. pp. 581–588. Springer (2016)