# Conv-MCD: A Plug-and-Play Multi-task Module for Medical Image Segmentation

Balamurali Murugesan[1,2], Kaushik Sarveswaran[2], Sharath M Shankaranarayana[3], Keerthi Ram[2], Jayaraj Joseph[2] and Mohanasankar Sivaprakasam[1,2]

[1]Indian Institute of Technology Madras (IITM), India
[2]Healthcare Technology Innovation Centre (HTIC), India
[3]Zasti, India

## Introduction

### Motivation

1. Encoder-decoder networks suffer from structural information loss due to downsampling operations performed via max-pooling layers.
2. Cross-entropy loss does not address the foreground-background class imbalance problem.
3. In the case of medical imaging, networks should be capable of producing smooth boundaries and is expected to work for multi-instance object segmentation.

### Related Work

1. U-Net [2]: Single encoder and single decoder architecture with cross entropy loss for mask.
2. DCAN [3]: Single encoder and two decoder architecture with cross entropy loss for mask and contour.
3. DMTN [4]: Single encoder and two decoder architecture with cross entropy loss for mask and mean square error for distance map.

| | [2] | [3] | [4] | Ours |
|---|---|---|---|---|
| Shape information | x | ✓ | ✓ | ✓ |
| Class imbalance | x | x | ✓ | ✓ |
| Smooth boundary | x | x | ✓ | ✓ |
| Multiple object instances | ✓ | ✓ | x | ✓ |

### Contributions

1. We propose a novel module Conv-MCD (**M**ask prediction, **C**ontour extraction and **D**istance Map estimation). The module consists of three parallel convolutional filters to learn the three related tasks simultaneously [1].
2. The proposed module can be added to any state-of-the-art network with minimal effort. We observed that the networks with our module showed improvement in performance across various metrics in relation to the base networks.

## Methodology

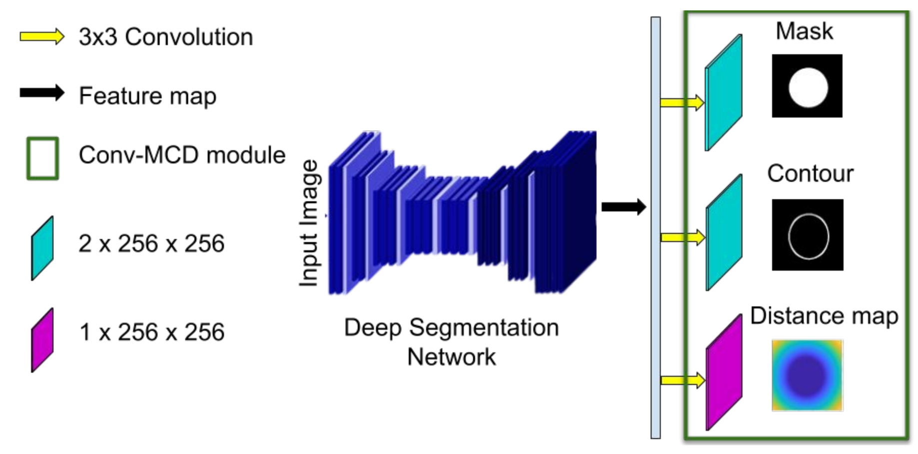


**Figure 1:** Block diagram illustrating the proposed module Conv-MCD. Proposed module could be included at the end of a typical deep segmentation network.

### Proposed Module

- The proposed module Conv-MCD (Figure 1) takes the feature maps from the deep learning networks as input and outputs mask, contour and distance map.
- This module helps the network to learn the multiple related tasks in parallel, enabling the network to generalize well.
- Mask prediction and contour extraction are classification tasks while the distance map estimation is a regression task.
- All these outputs are obtained by parallel convolution layers.
- The parameters of the filters are: kernel size is 3 × 3 with stride 1 and padding 1, the number of channels of the kernel is decided by the number of output channels of the feature maps. The number of filters for classification task is 2, which denotes the number of classes considered. Similarly, the number of filters for regression is 1.

### Capturing Structural Information

- Contour map is obtained by extracting the boundaries of connected components based on the ground truth segmentation maps which are subsequently dilated using a disk filter of radius 5.
- The distance maps are obtained by applying euclidean signed distance transform to the contour.

### Loss Function

The loss function consists of three components - Negative Log Likelihood (NLL) loss for mask and contour, Mean Square Error (MSE) loss for the distance. The total loss is given by :

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{mask} + \lambda_2\mathcal{L}_{contour} + \lambda_3\mathcal{L}_{distance}$$

## Experiments and Results

### Dataset

We use the Polyp segmentation dataset from MICCAI 2018 Gastrointestinal Image ANAlysis (GIANA) [5]. We deemed this dataset to be ideal for our experiments since it exhibits the following characteristics:

- large variations in shape.
- multi-instance object occurrences.
- foreground-background imbalance.
- difficulty in extracting the boundary of smooth, blobby objects.

The dataset consists of 912 images with ground truth masks. The dataset is randomly split into 70% for training and 30% for testing. The images are center cropped and resized to 256 × 256.

### Evaluation Metrics

- *Segmentation evaluation:* Jaccard index and Dice similarity score.
- *Shape Similarity:* Hausdorff Distance (HD)
- *Segmentation around boundaries:* The relative number of misclassified pixels within a narrow band (trimap) surrounding actual object boundaries
- *Boundary smoothness:* Maximum F-score (MF) between the boundaries of predicted and ground truth mask.

**Table 1**: Comparison of the base network (U-Net) with and without adding our proposed module

| Architecture | Dice | Jaccard | HD | MF | Time (ms) | # parameters |
|---|---|---|---|---|---|---|
| U-Net | 0.8125 | 0.7323 | 24.13 | 0.6144 | 1.3131 | 7844256 |
| DCAN | 0.8151 | 0.7391 | 22.74 | 0.616 | 1.8677 | 10978272 |
| DMTN | 0.8283 | 0.7482 | 22.68 | 0.5681 | 1.8531 | 10977984 |
| U-Net + Conv-MC (Ours) | 0.8149 | 0.7389 | 22.86 | 0.6083 | 1.3384 | 7844832 |
| U-Net + Conv-MD (Ours) | 0.8286 | 0.7489 | 22.54 | 0.5844 | 1.3235 | 7844544 |
| U-Net + Conv-MCD (Ours) | **0.8426** | **0.7692** | **22.27** | **0.6552** | 1.3501 | 7845120 |

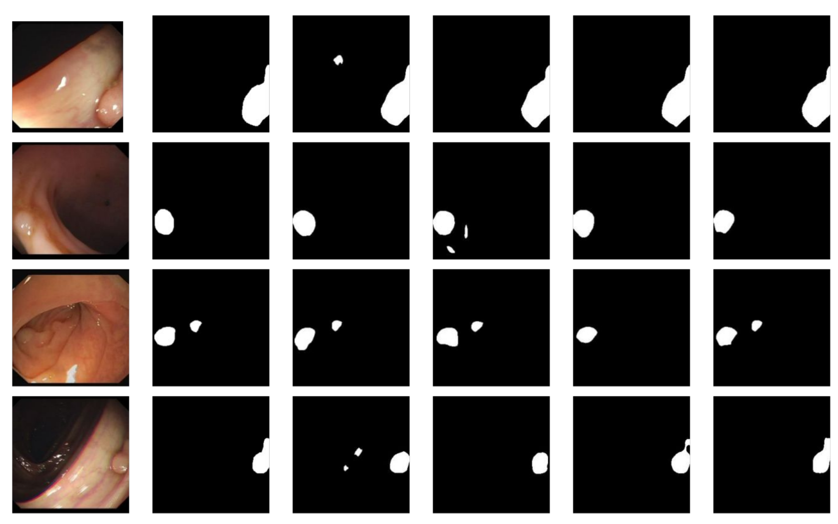

**Figure 2:** Four sample cases, from left to right: Image, Ground truth, U-Net, DCAN, DMTN and U-Net + Conv-MCD.

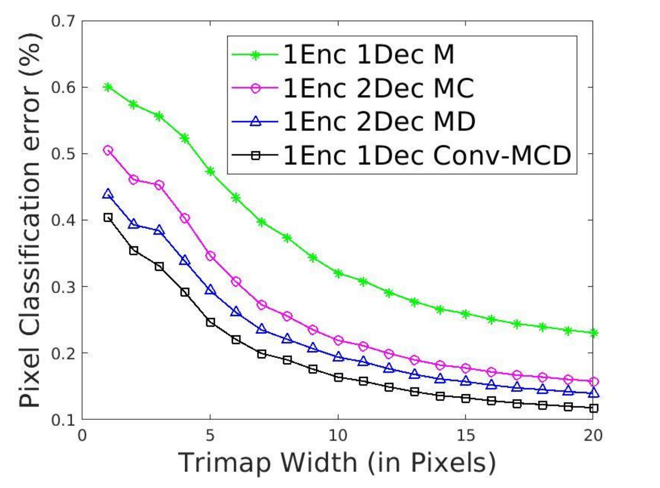


**Figure 3:** Pixel classification error vs trimap width for U-Net (Green), DCAN (Pink), DMTN (Blue) and U-Net + Conv-MCD (Black).

**Table 2**: Comparison of evaluation metrics for various state-of-the-art networks

| Architecture | Module type | Dice | Jaccard | HD | MF |
|---|---|---|---|---|---|
| SegNet | - | 0.6515 | 0.5497 | 41.31 | 0.4127 |
| | Conv-MCD | **0.7194** | **0.6314** | **35.61** | **0.4923** |
| UNet | - | 0.8249 | 0.7468 | 23.27 | 0.6144 |
| | Conv-MCD | **0.838** | **0.7652** | **21.05** | **0.6161** |
| UNet16 | - | 0.8441 | 0.7676 | 15.24 | 0.7514 |
| | Conv-MCD | **0.9124** | **0.8559** | **13.17** | **0.7753** |
| LinkNet34 | - | 0.8835 | 0.8206 | 16.02 | 0.7358 |
| | Conv-MCD | **0.8979** | **0.8383** | **14.28** | **0.7474** |

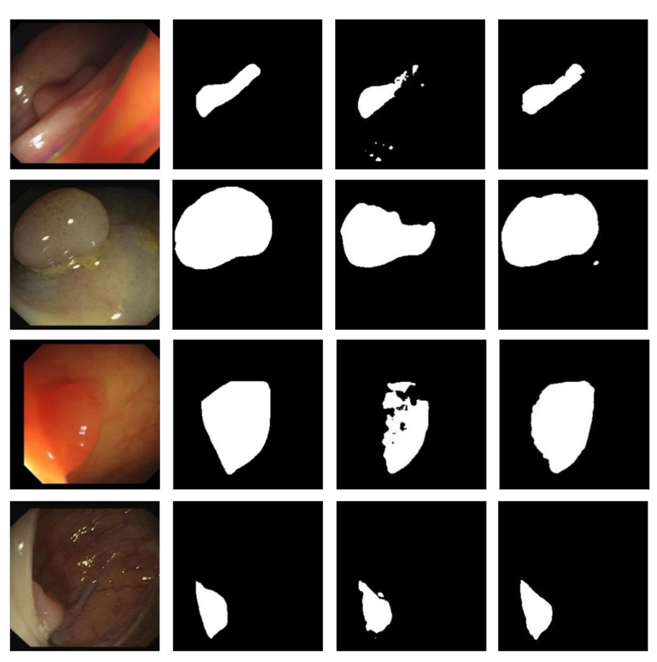

**Figure 4:** From row 1 to 4: SegNet, UNet, UNet16 and LinkNet34. In each row from left to right: Image, Ground Truth, network without Conv-MCD and network with Conv-MCD

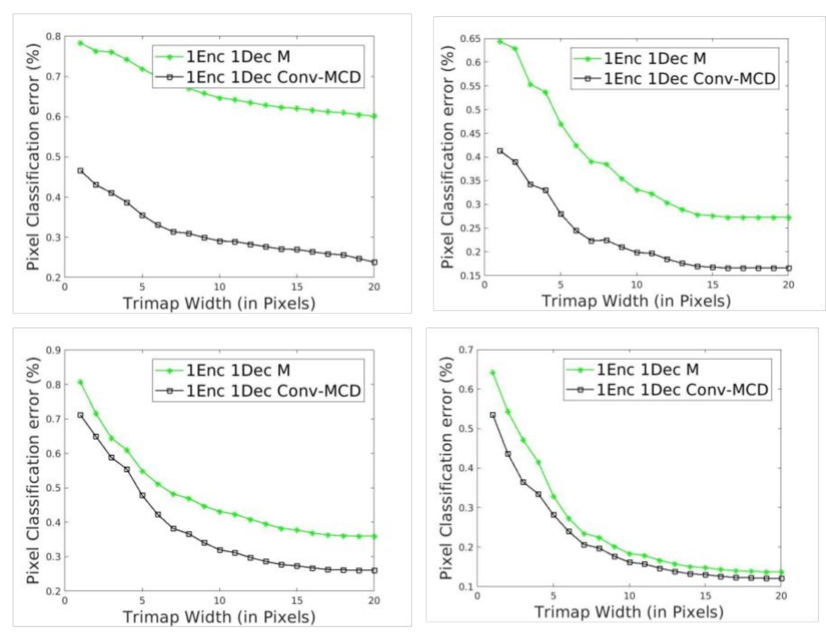


**Figure 5:** Row-wise order: Pixel classification error vs trimap width for SegNet, UNet, UNet16 and LinkNet34. In the graph, green line represents the base network and black line represents network with Conv-MCD.

## References

1. Caruana, R.: Multitask learning: A knowledge-based source of inductive bias. In: ICML (1993)
2. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241 (2015)
3. Chen, H., Qi, X., Yu, L., Heng, P.: DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2487–2496 (2016)
4. Tan, C., Zhao, L., Yan, Z., et al.: Deep multi-task and task-specific feature learning network for robust shape preserved organ segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1221–1224 (2018)
5. V´azquez, D., Bernal, J., Sanchez, F.J., et al.: A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of healthcare engineering 2017 (2017)

Code available at https://github.com/Bala93/Multi-task-deep-network
Contact : balamurali@htic.iitm.ac.in

MICCAI 2019 SHENZHEN