Hyper-Convolution Networks for Biomedical Image Segmentation Supplemental Material

Tianyu Ma
Cornell University
Cornell Tech
tm478@cornell.edu

Adrian V. Dalca
Massachusetts Institute of Technology
Massachusetts General Hospital
Harvard Medical School
adalca@mit.edu

Mert R. Sabuncu
Cornell University
Cornell Tech
msabuncu@cornell.edu

Kernel Size	N_L	No. Channels	Params (M)
5×5	8	4,8,16,32	0.04
5×5	2	8,16,32,64	0.05
5×5	4	8,16,32,64	0.08
5×5	8	8,16,32,64	0.14
5×5	2	16,32,64,128	0.19
5×5	4	16,32,64,128	0.31
5×5	8	16,32,64,128	0.54
5×5	2	32,64,128,256	0.73
5×5	4	32,64,128,256	1.2
5×5	8	32,64,128,256	2.2
5×5	24	32,64,128,256	5.3

Table 1. Hyper-UNet architecture details for the models that produce the results of Figure 4 and 6 in the paper. N_L is the number of nodes in the last hidden layer of the Hyper-network. "No. Channels" are the number of channels used in the successive convolutional layers in the encoder (i.e., contraction half) of the network. The total number of learnable parameters for each model is listed in the right-most column.

Abstract

In this supplemental material, we provide further details for the network architectures of UNet and Hyper-UNet used in Figures 4 and 6 of the paper (Section 1). Section 2 further discusses the smoothness of convolutional kernels in different models. We present a detailed comparison of UNet and Hyper-UNet for both experiments in Section 3. Furthermore, we provide some representative visualizations of the ground truth and predicted segmentations for different models, in the liver lesion and MS-lesion tasks (Section 4).

Kernel Size	No. Channels	Params (M)
3×3	4,8,16,32	0.04
3×3	8,16,32,32	0.08
3×3	8,16,32,64	0.14
3×3	16,32,64,64	0.35
3×3	16,32,64,128	0.54
3×3	32,64,128,128	1.4
3×3	32,64,128,256	2.1
5×5	32,64,128,256	5.3

Table 2. UNet architecture details for the models that produce the results of Figure 4 and 6 in the paper. N_L is the number of nodes in the last hidden layer of the Hyper-network. "No. Channels" are the number of channels used in the successive convolutional layers in the encoder (i.e., contraction half) of the network. The total number of learnable parameters for each model is listed in the right-most column.

1. UNet and Hyper-UNet Architectures

The main paper text shows the Dice scores for UNet and Hyper-UNet with different numbers of parameters in Figures 4 and 6. For the UNet architecture, we change the kernel size and the number of channels to obtain models with different numbers of parameters. For Hyper-UNet, we vary N_L – the number of nodes in the last hidden layer of the Hyper-network – and the number of channels. In Tables 1 and 2, we provide details of the different architectures used in the experiments.

2. Kernel Smoothness

Figure 1 plots the smoothness (measured by Laplacian) of convolutional kernels in eight layers of the UNet (Orange) and Hyper-UNet (Blue). Each subfigure includes 4 models from a segmentation task. We observe that Hyper-UNet models have lower Laplacian values than the UNet models in both experiments. Kernels from the deeper lay-

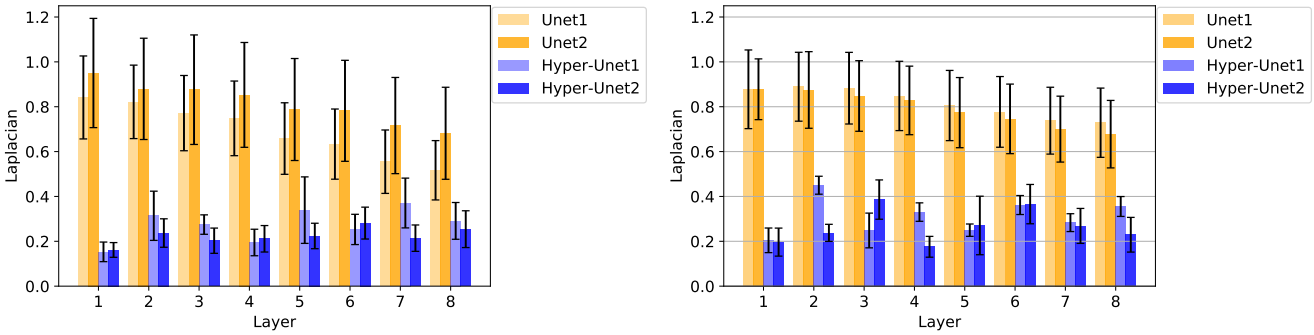

Figure 1. Average Laplacian of convolutional kernels at different layers of 4 models (2 UNet models and 2 Hyper-UNet models of the same task) for liver lesion segmentation (left) and MS-lesion (right) The Lapalacian value for a model is the average of all the kernels in a layer, and the error bar is the standard deviation. Hyper-UNet models always have lower Laplacian values than the UNet models.

ers of the Hyper-UNet models tend to be less smooth than the first layer. Beyond the first layer, we do not observe any particular trend of smoothness in the Hyper-UNet models with respect to depth. On the other hand, the baseline UNet models have smoother kernels in deeper layers, which is a pattern we found across all models and datasets.

3. Test Results

In the main paper we show the average Dice scores achieved by the UNet and Hyper-UNet models in Tables 1 and 3. In Figure 2 , we present the Dice scores for all samples in the test datasets of both experiments. There are two Dice scores for every scan in the MS-lesion segmentation task, because there are two independent expert annotations. We observe that for a vast majority of cases (87% and 80% in liver lesion and MS-lesion segmentation), the Hyper-UNet achieves a higher Dice score than the UNet baseline. The paired t-test for liver lesion and MS-lesion segmentation yields highly significant p-values of 10^{-5} and 10^{-10}.

4. Visualizations

Figure 3 and 4 show example images and segmentation contours for both experiments. We use the test samples for the liver segmentation and validation samples for the MS-lesion segmentation because the ground-truth segmentation for the ms-lesion task are not made available. We visualize the segmentation obtained by three methods: UNet, non-local UNet, and Hyper-UNet along with their corresponding Dice scores. In these examples, we observe that the Hyper-UNet performs better at the edges possibly due to its dense kernels that can aggregate both short-range and long-range information with details.

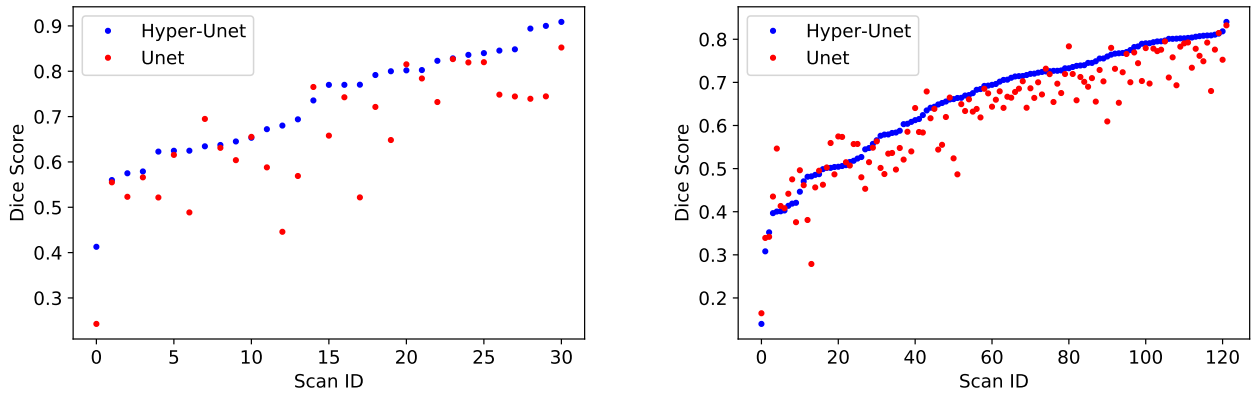

Figure 2. Dice scores for 5×5 Hyper-UNet (blue) and standard UNet (red) of all test scans in liver lesion segmentation (left) and MS-lesion segmentation (right). The scan IDs are sorted by Hyper-UNet Dice score. The testing Dice scores are higher for Hyper-UNet in most scans.

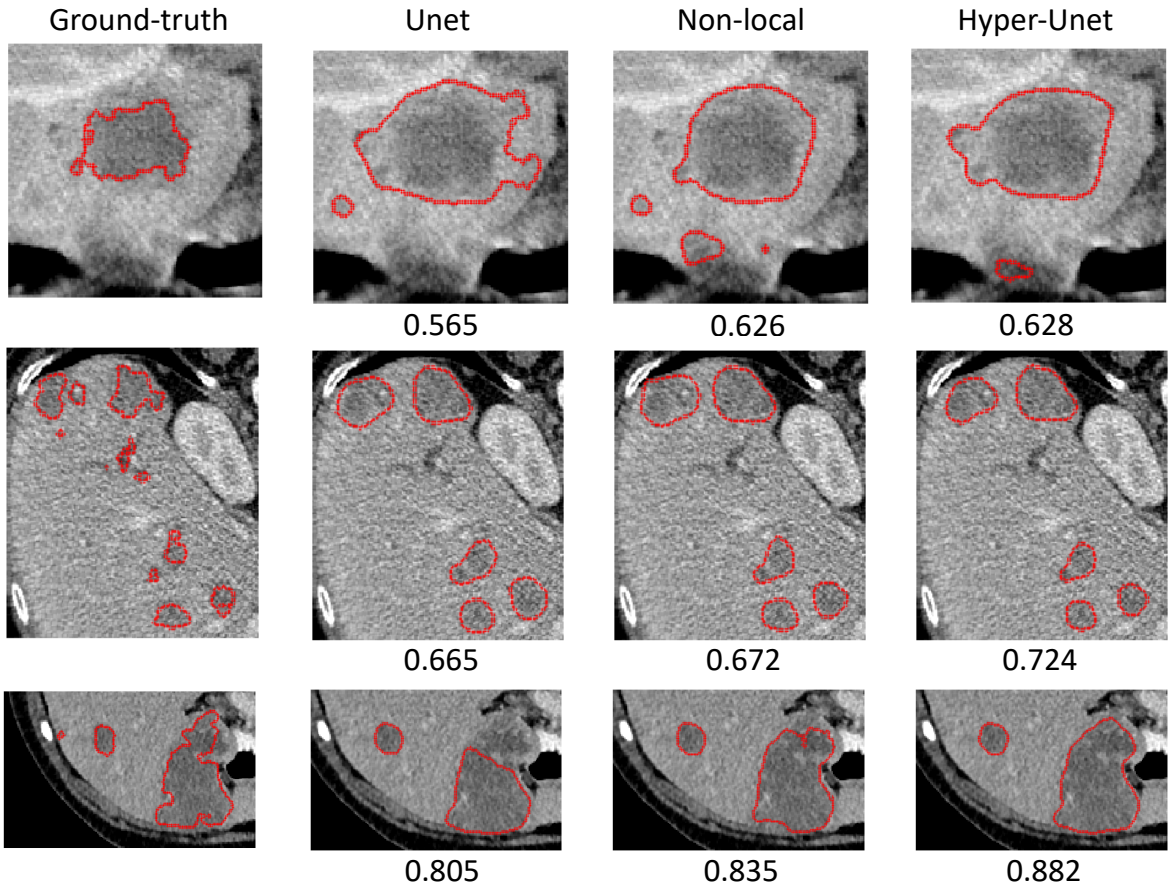

Figure 3. Visualizations of three representative liver lesion images and segmentation contours from different models: manual ground truth, UNet, Non-local UNet, and Hyper-UNet with Dice score listed below.

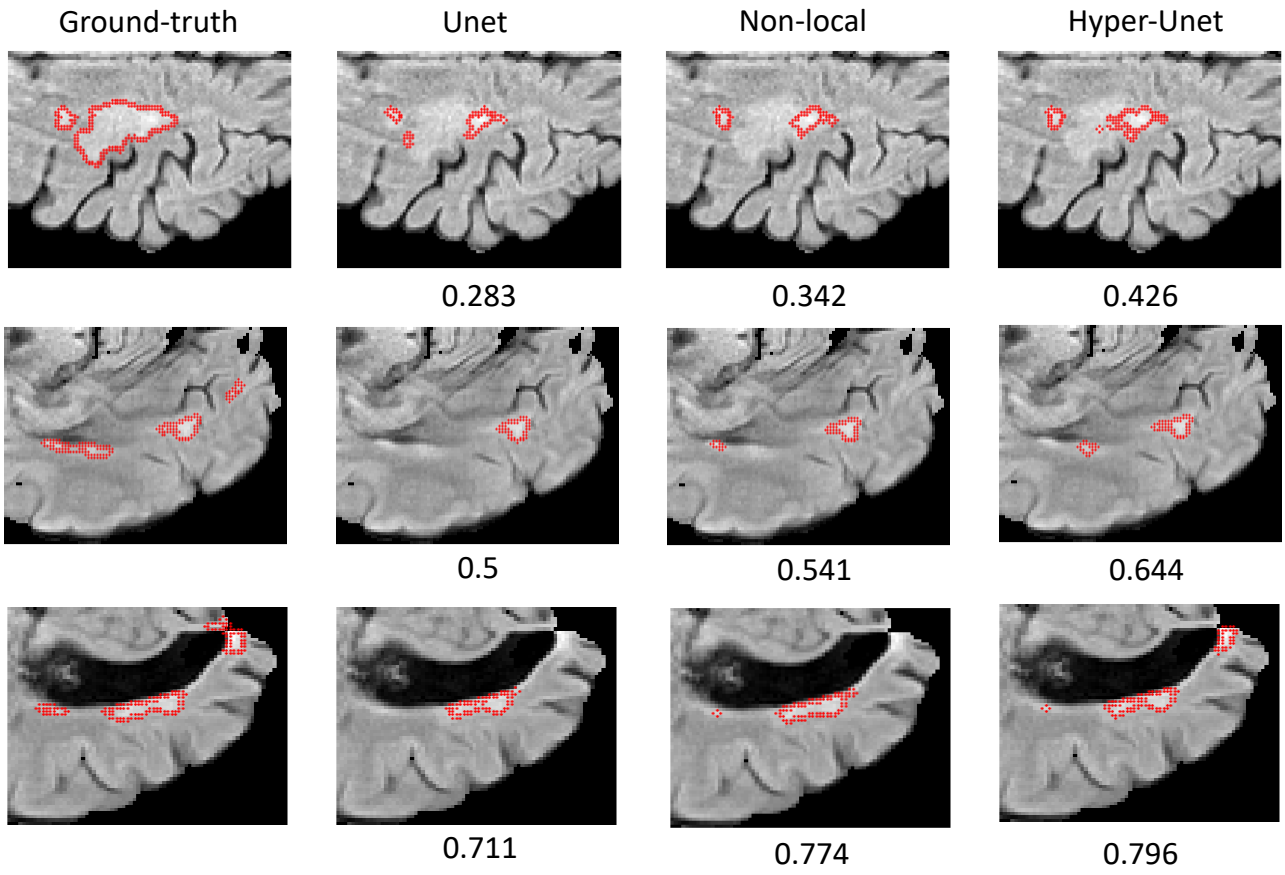

Figure 4. Visualizations of three representative MS-lesion images and segmentation contours by different models: manual ground truth, UNet, Non-local UNet, and Hyper-UNet with Dice score listed below.