High-Resolution Image Segmentation with a U-Net-Based Segmentation CNN on Multiple GPUs

C. Verburg,[@][*] A. Heinlein,[*] E.C. Cyr[†]

[@]c.verburg@tudelft.nl, [*]Delft Institute of Applied Mathematics, TU Delft, [†]Sandia National Laboratories, US

Introduction

Most semantic image segmentation models focus on low-resolution images, neglecting the challenges posed by high-resolution datasets due to extremely high GPU memory constraints. Conventional approaches to processing high-resolution images, such as down-sapling or patch cropping, often lead to the loss of either fine-grained details or global contextual information. In this work, we address the challenge of high-resolution image segmentation by combining strategies from Domain Decomposition Methods (DDMs) with machine learning, aiming to optimize memory usage while conserving accurate segmentation results.

Our proposed approach (see Figures 1, 2, and 3) builds upon the U-Net architecture, a well-established CNN for image segmentation tasks. However, the main challenge with existing U-Net variants lies in their large memory requirements, making them unsuitable for high-resolution applications. In our novel approach, we integrate the U-Net architecture with a divide-and-conquer spatial domain decomposition strategy. Our approach enables the memory-efficient segmentation of high-resolution images while minimizing communication overhead.

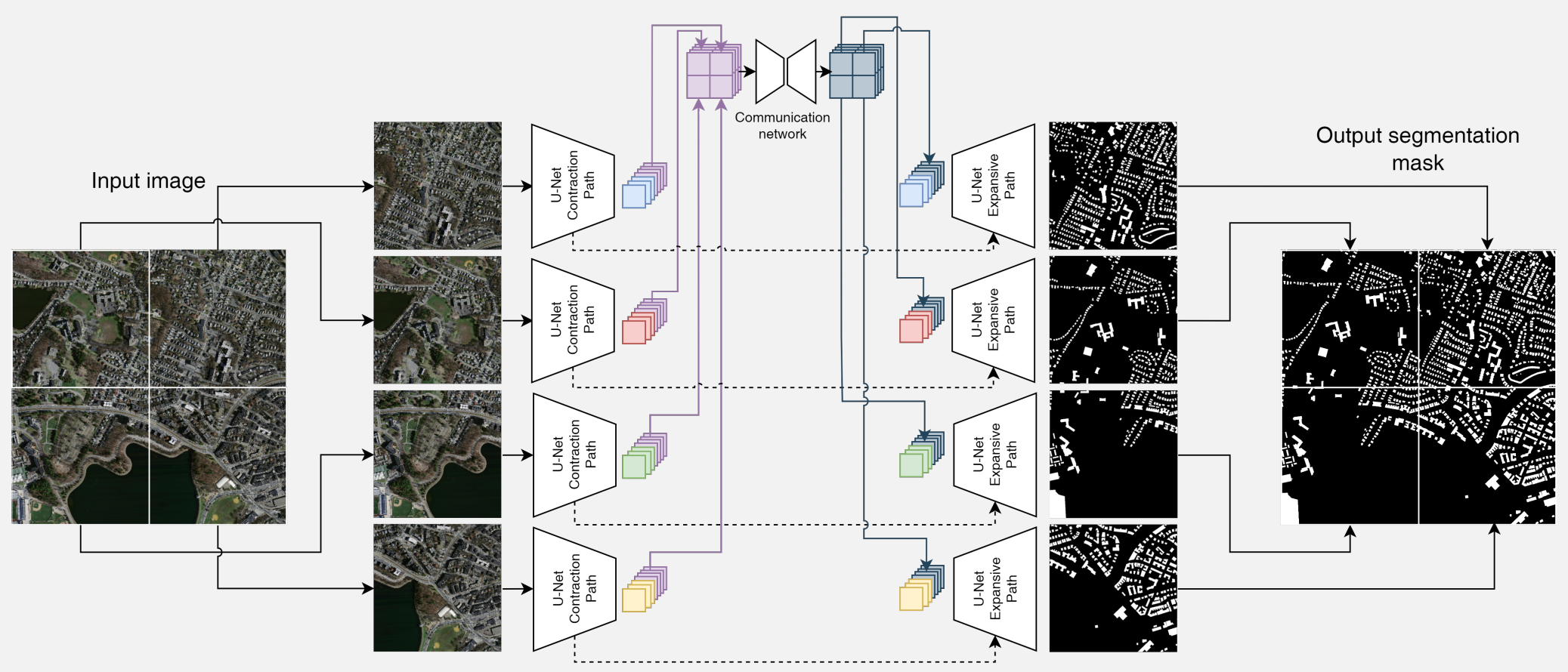

Figure 1: Global overview of the proposed model for four subdomains.

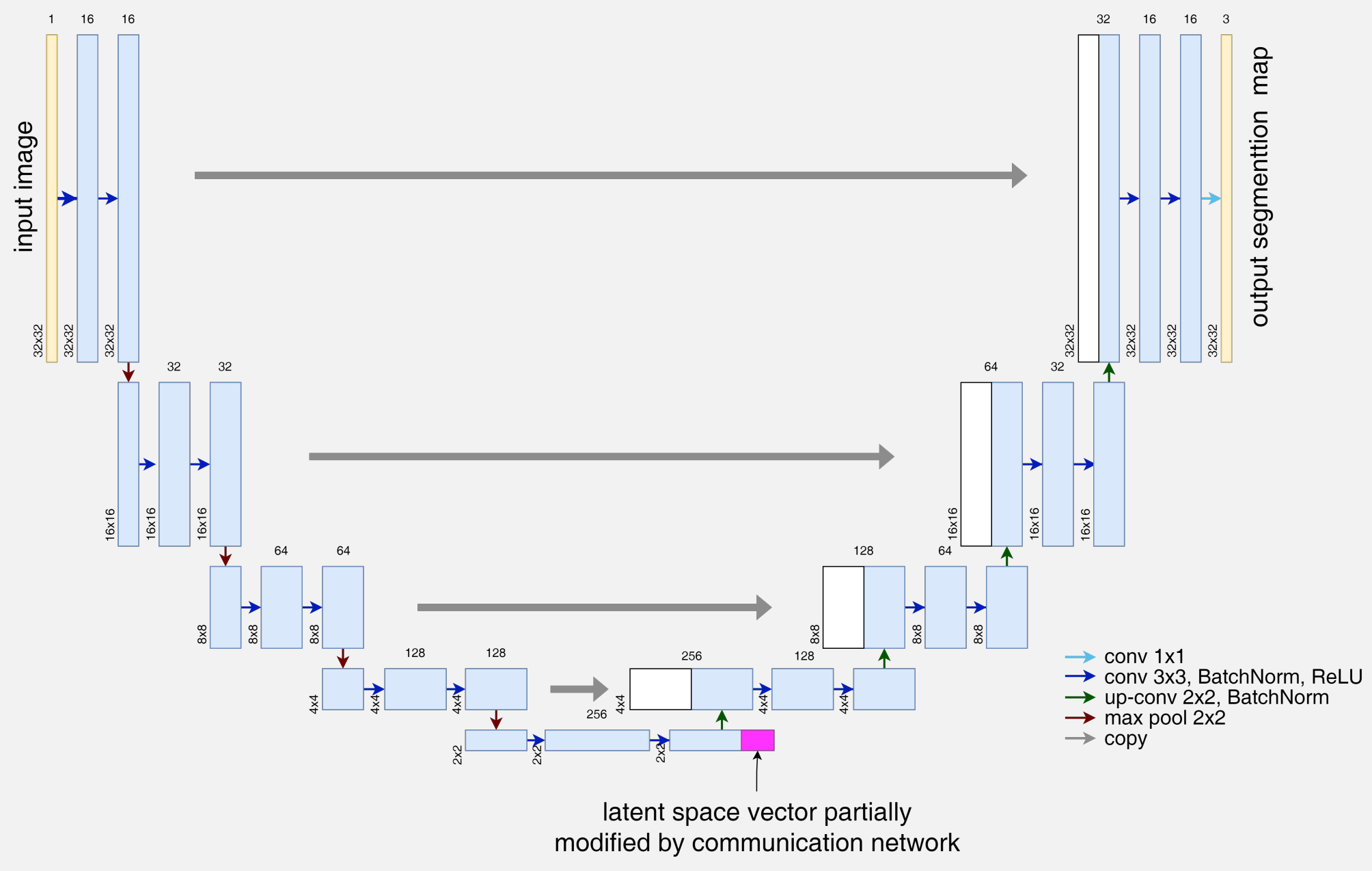

Figure 2: Proposed architecture for the encoder-decoder architecture.

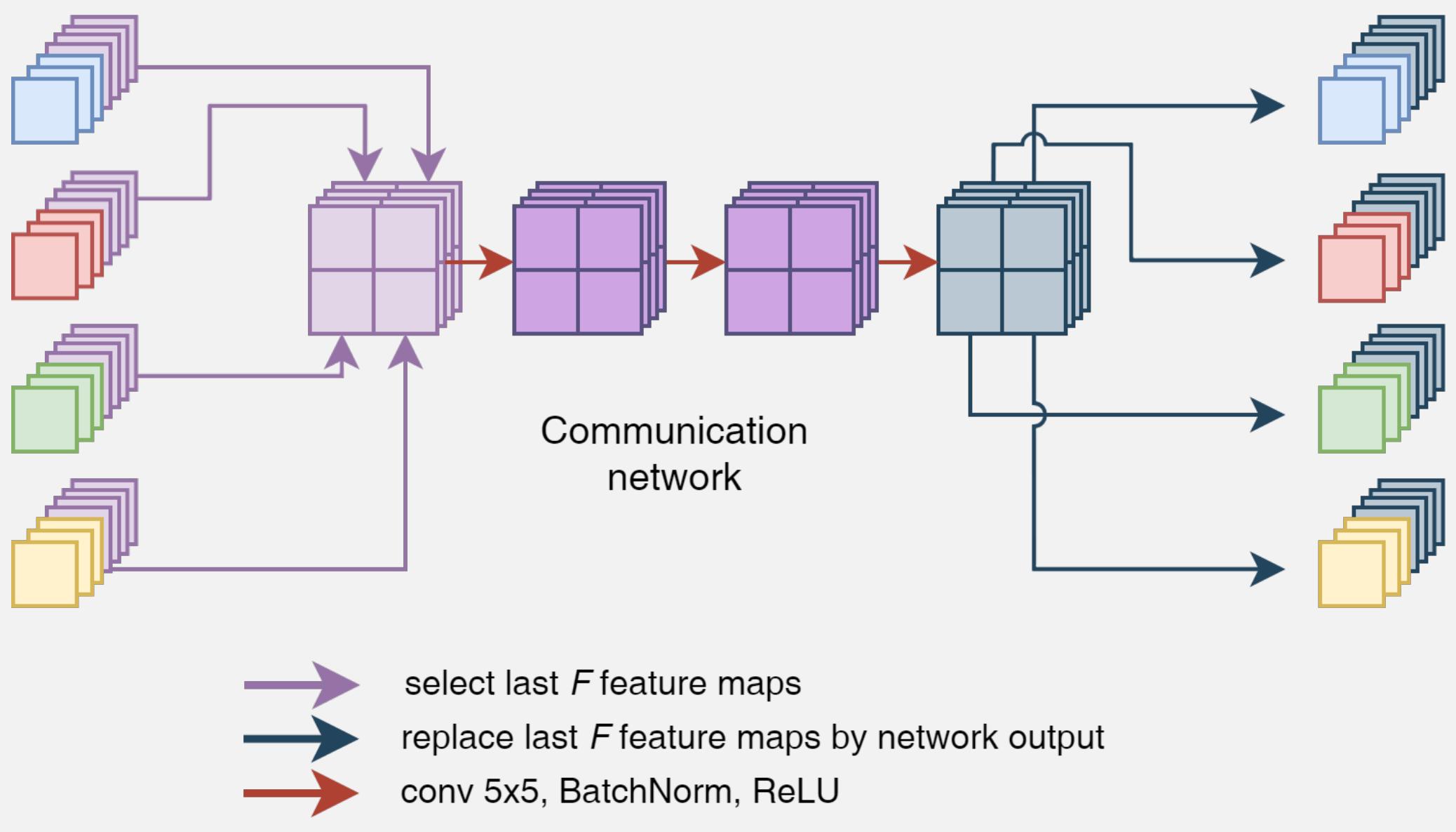
Communication network
select last F feature maps
replace last F feature maps by network output
conv 5x5, BatchNorm, ReLU

Figure 3: Proposed communication network architecture.

Experimental setup

We evaluate the proposed model architecture on a synthetic dataset (see Figure 5) and a realistic dataset (see Figure 6) to answer the following questions:

- **Memory efficiency**: What are the memory requirements for the proposed architecture compared to the U-Net architecture? (See Figure 4)
- **Communication network**: What is the role of the communication network? (See Figure 7)
- **Accuracy**: What is the accuracy of the generated segmentations, quantitatively and qualitatively? (See Figures 5, 6 and Table 1)

Memory efficiency and accuracy

Our study indicates that our model architecture requires slightly more memory per device compared to a baseline U-Net. However, in contrast to the baseline U-Net, our method incorporates a communication module, facilitating the transfer of contextual information among devices (see Figure 4).

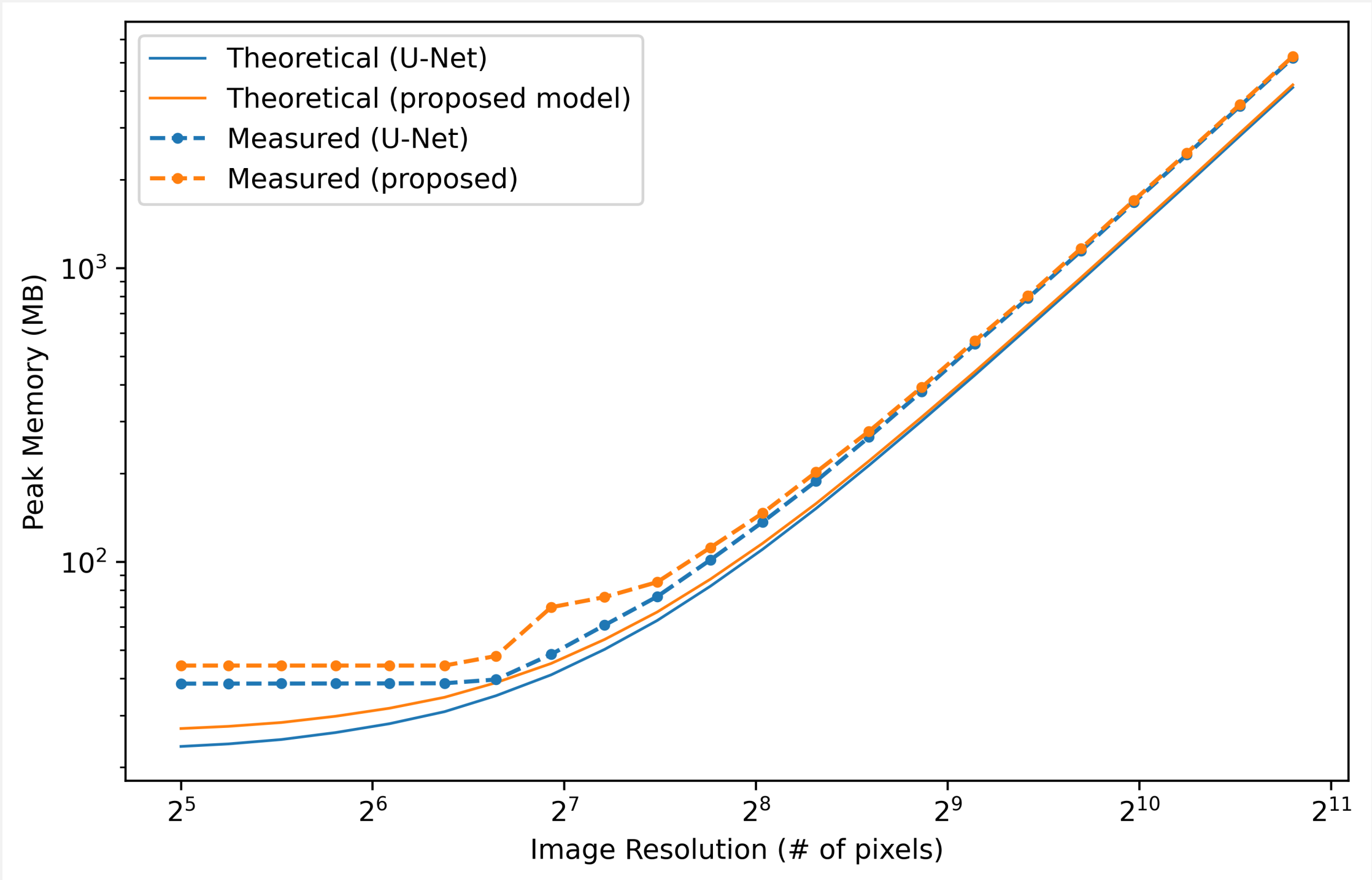

Figure 4: Peak memory usage per GPU during training for the proposed model and the U-Net.

The proposed model, trained on 256×256 pixels subdomains coupled by the communication module, achieves comparable or even superior accuracy compared to the baseline U-Net model trained on the full-resolution image (refer to Table 1).

model	mean IoU		
dataset	*train*	*test*	*val*
baseline U-Net	**0.7239**	**0.6723**	0.6858
proposed w/ communication	0.7229	0.6680	**0.6907**
proposed w/o communication	0.7025	0.6474	0.6631

Table 1: Mean IoU score for the DeepGlobe Satellite Segmentation Dataset

Qualitative results

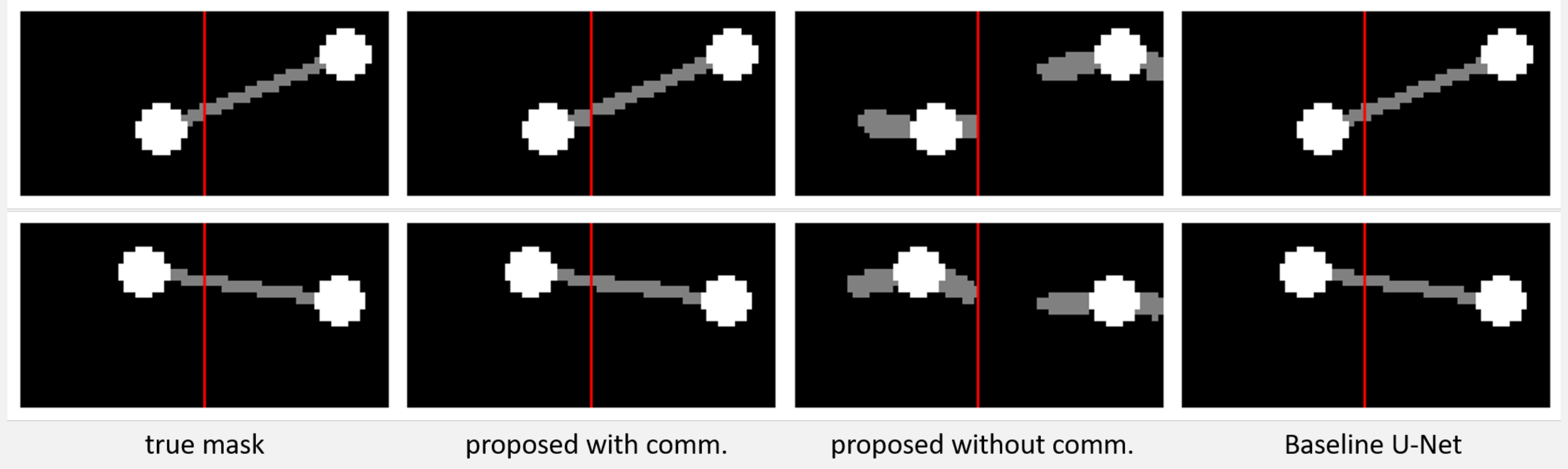

Figure 5: From left to right: true mask, mask predicted w/ communication module, mask predicted w/o communication module, and mask predicted by baseline U-Net.

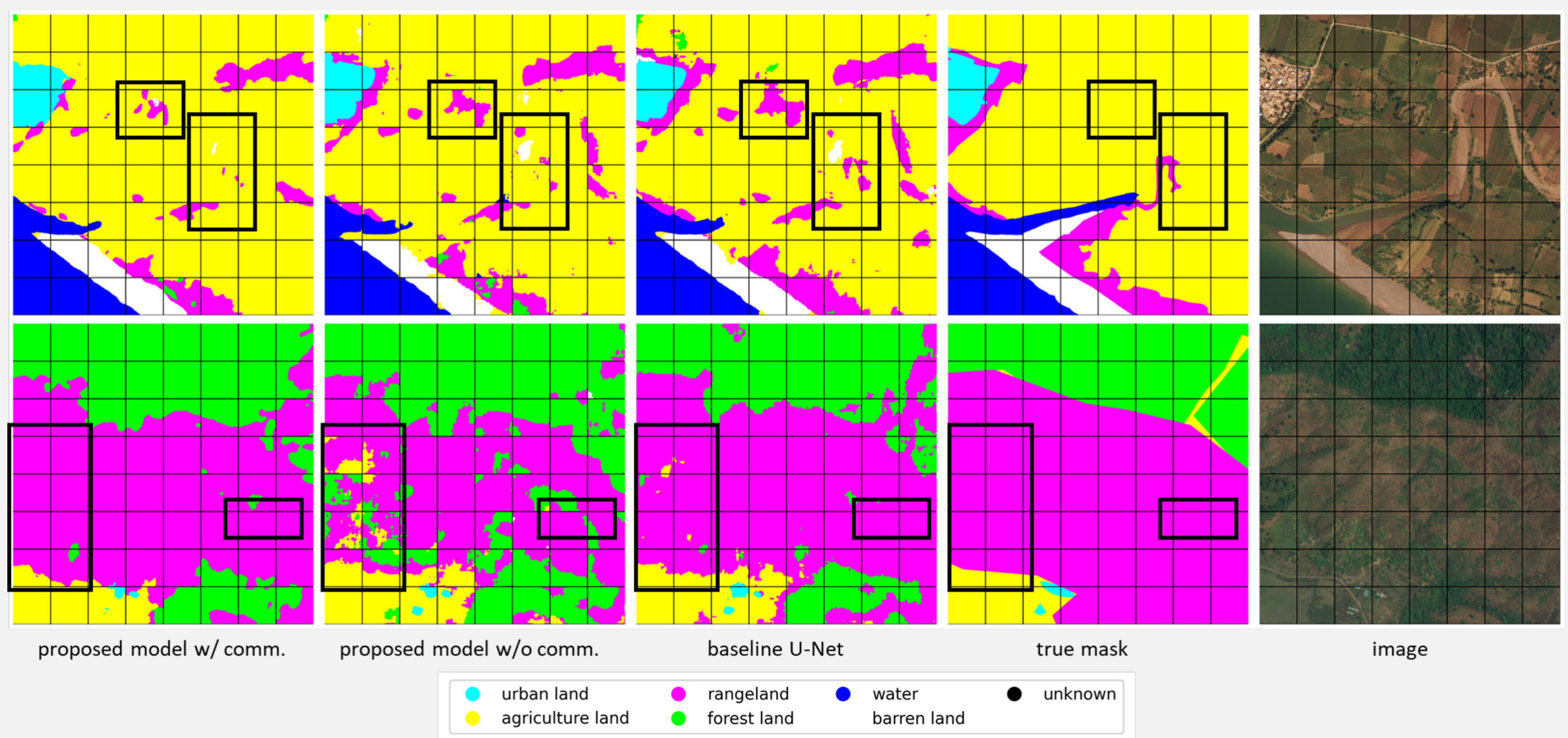

Figure 6: From left to right: true mask, mask predicted w/ communication module, mask predicted w/o communication module, and mask predicted by baseline U-Net.

Impact of the communication network size

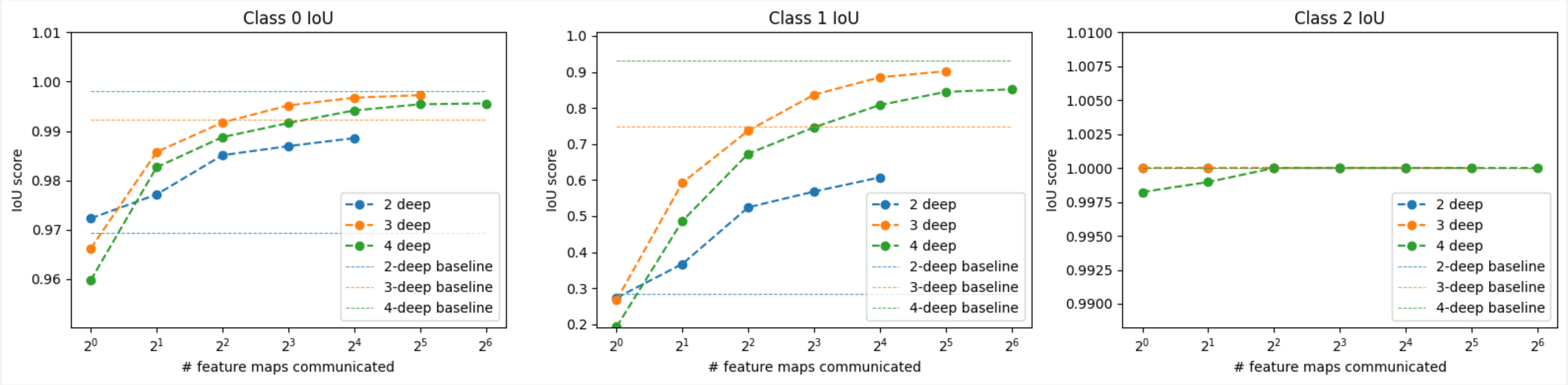

Figure 7: IoU score as a function of the number of feature maps communicated for different U-Net depths for the synthetic dataset. Class 0: background, Class 1: line segment, Class 2: circle.

TU Delft Delft University of Technology

https://github.com/corne00/HiRes-Seg-CNN

Delft Institute of Applied Mathematics