# Supplementary Material for Binarizing Sparse Convolutional Networks for Efficient Point Cloud Analysis

Xiuwei Xu<sup>1,2</sup>, Ziwei Wang<sup>1,2</sup>, Jie Zhou<sup>1,2</sup>, Jiwen Lu<sup>1,2\*</sup> <sup>1</sup>Department of Automation, Tsinghua University, China <sup>2</sup>Beijing National Research Center for Information Science and Technology, China {xxw21, wang-zw18}@mails.tsinghua.edu.cn; {jzhou, lujiwen}@tsinghua.edu.cn



Figure 1. The overall frameworks of FCN (a) and UNET (b) which are constructed with the basic blocks in (c).

## Abstract

In this supplementary material, we detail the network architectures and training hyperparamters used in our experiments. Section A shows how to construct the whole network with basic blocks. Section B details the training hyperparamters used in our experiments.

#### A. Network Architecture

#### A.1. Overall Framework

We illustrate the architectures of FCN and UNET in Figure 1. Only two levels of downsampling/upsampling are shown in the figure for simplicity. Note that for each level only one SSC block is drawn, while in fact there may be one or two blocks according to the size of networks.

FCN-S has 16 filters in the input layer, and one SSC block per level. FCN-H has 24 filters in the input layer, and two SSC blocks per level. Both networks use eight levels of downsampling and upsampling. We increase the number of filters in the networks when downsampling: in particular, we add 16 (S) or 24 (H) filters every time we reduce the scale. UNET-S has 16 initial filters and one SSC block per level. UNET-H has 32 initial filters and two SSC blocks per level. Both networks use six levels of downsampling and upsampling. Each downsampling operation adds 16 (S) or 32 (H) filters, while upsampling operation subtracts the same numbers of filters correspondingly.

When binarizing the network, the first SSC layer and the



Figure 2. Details of the basic blocks. (a) SSC layer, (b) SSC block, (c) Conv layer, (d) DeConv layer.

linear layers are kept real-valued following previous methods [1–3]. We adopt our SFSC module in the SSC block for BSC-Manual and BSC-Net, which is detailed in next subsection.

## A.2. Block Detail

We detail the structure of basic blocks contained in the binary FCN and UNET in Figure 2. For SSC layer, the projection indicates identity mapping when the input and output channel are equal, otherwise it is a 1-bit  $1 \times 1$  convolution operation. For Conv and DeConv blocks, we keep the  $1 \times 1$  convolution operation real-valued, which is proved to be essential for binarizing performance [2, 3].

## **B.** Training Hyperparamters

There are 2 training stages when training BSC-Baseline and BSC-Manual while 3 stages when training BSC-Net. For all stages, we set max epoch as 128, weight decay as 0 and adopt Adam optimizer with a stepwise scheduler which steps at 60 and 100 epoch (reduce the learning rate by a factor of 10). The initial learning rates for the first and second stage when training BSC-Baseline and BSC-Manual are set to 0.001 and 0.0002. While for training BSC-Net, the initial learning rates for the three stages are set to 0.001, 0.001 and 0.0002. The weight of the confidence loss is set to 0.1 for FCN and 0.01 for UNET. Note that for training UNET-H on 2cm voxel, we double the max epoch and the stepping epochs as the huge network is hard to converge.

### References

 Zechun Liu, Zhiqiang Shen, Marios Savvides, and Kwang-Ting Cheng. Reactnet: Towards precise binary neural network with generalized activation functions. In *ECCV*, page 143–159, 2020. 2

- [2] Zechun Liu, Baoyuan Wu, Wenhan Luo, Xin Yang, Wei Liu, and Kwang-Ting Cheng. Bi-real net: Enhancing the performance of 1-bit cnns with improved representational capability and advanced training algorithm. In *ECCV*, pages 722–737, 2018. 2
- Brais Martinez, Jing Yang, Adrian Bulat, and Georgios Tzimiropoulos. Training binary neural networks with real-tobinary convolutions. *arXiv preprint arXiv:2003.11535*, 2020.
  2