LDW-Pooling: Learnable Discrete Wavelet Pooling for Convolutional Networks

Bor-Shiun Wang[1]
eddiewang.ai08@nycu.edu.tw
Jun-Wei Hsieh[1]
jwhsieh@nctu.edu.tw
Ping-Yang Chen[1]
pingyang.cs08g@nctu.edu.tw
Ming-Ching Chang[3]
mchang2@albany.edu
Lipeng Ke[4]
lipengke@buffalo.edu
Siwei Lyu[4]
siweilyu@buffalo.edu

[1] National Yang Ming Chiao Tung University, Taiwan
[2] University at Albany, State University of New York, USA
[3] University at Buffalo, State University of New York, USA

1 Efficiency comparisons among Max, Average, and LDW-Pooling

To avoid unfair comparisons due to special hardware designs for Max pooling and average pooling, we re-implemented all pooling methods by C++. The parameters for kernel size, channel size, input width, and height are 3, 256, 28, and 28, respectively. Table 1 shows the efficiency comparisons among Max, Average, and LDW-Pooling. Clearly, our proposed LDW-pooling method outperforms the other two commonly used pooling techniques.

Table 1: Efficiency comparisons among different pooling technique

Type	Time (ms)
MaxPool	0.001555
AveragePool	0.001879
LDWPool	0.**001253**

2 Model size comparisons

We also proposes a new LDW-Pooling scheme by decomposing the input signal into four discrete-wavelet subands (LL, LH, HL, HH) without information loss. For each decomposition, only C/4 channels are used to generate feature maps which are further decomposed to C channels with the LDW-pooling technique. The LDW-Pooling performs 2-D convolution by decomposing it into two 1-D convolutions. The parameter size for a 2-D convolution kernel is K×K but becomes 2K due to the 1-D convolution design. Table 2 shows the model with our LDW-Pooling is smaller than the original design. Specifically, our model is almost half of the size of the ResNet18.

Table 2: Model parameter size

Model	LDW	Param.(M)
ResNet50		23.53
ResNet50	√	**22.54**
ResNet18		11.18
ResNet18	√	**6.93**