OneNet: A Channel-Wise 1D Convolutional U-Net

VisionXP

Sanghyun Byun, Kayvan Shah, Ayushi Gang, Christopher Apton

USC Viterbi School of Engineering

Motivation

2D CNN→ expensive in U-Net

MobileNet → does not capture spatio-channel relations

Channel-Wise 1D → Captures all relations with pixel-unshuffle

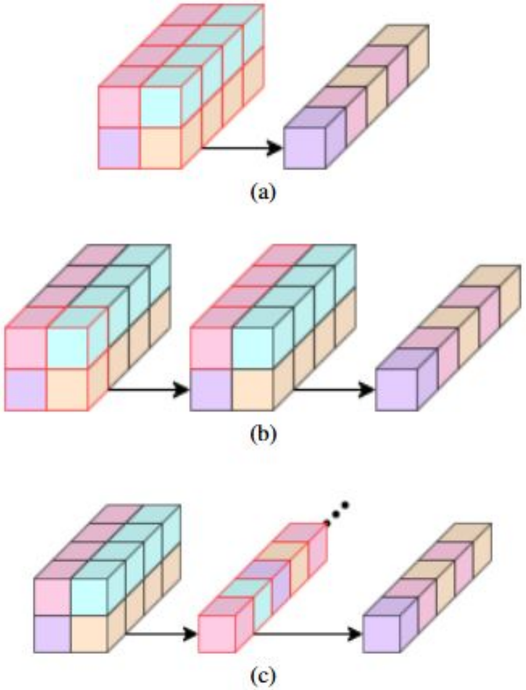
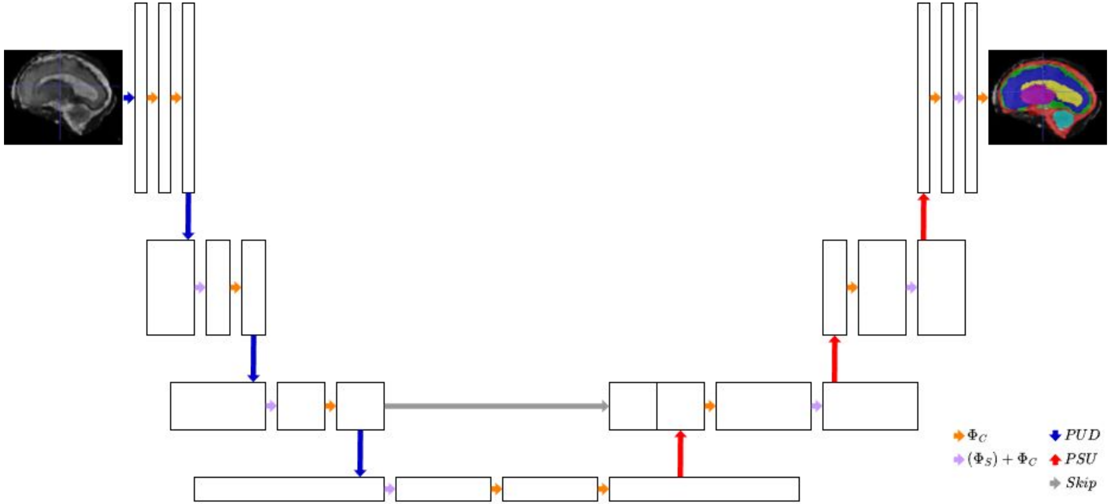


Figure 3. **Comparison of Convolutional Block** (a) Traditional 2D convolutional block with max pooling. (b) MobileNet [12] block with max pooling. (c) OneNet implementation with pixel-unshuffle downscaling followed by 1D convolution.

USC Viterbi School of Engineering

Method - Architecture

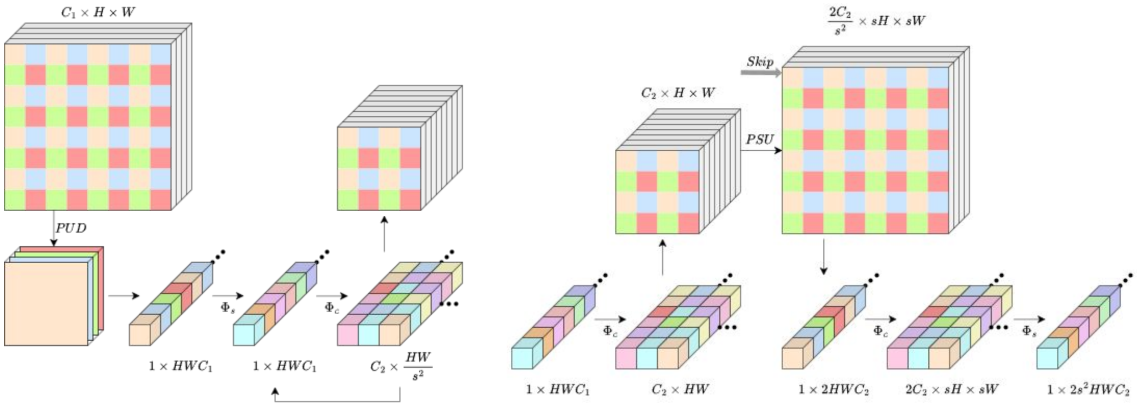
Max Pooling → Pixel-Unshuffle Downscaling
Upsampling → Pixel-Shuffle Upscaling
2D Convolution → 1D Channel-Wise Convolution



Method - Pixel-Shuffle / 1D Convolution

Encoder - pixel-unshuffle → flatten → 2X channel-wise 1D convolution → restore dimensions
Decoder - pixel-shuffle → flatten → 2X channel-wise 1D convolution → restore dimensions

Optimization - pixel-unshuffle in 1D (skip flatten/restore)



Algorithm 1 1D pixel-unshuffle downscaling

Input: X in shape $(B, C, H \times W)$, height H, width W
Output: pixel-unshuffled X in shape $(B, H \times W \times C)$

$I \leftarrow []$
for $i \leftarrow 0$ to $\frac{W}{2}$ **do**
  for $j \leftarrow 0$ to $\frac{H}{2}$ **do**
    $t \leftarrow 2i + 2j$
    $I \leftarrow [t, t+1, t+w, t+w+1]$
$X \leftarrow X[:, :, I].reshape(B, C, -1, 4)$
$X \leftarrow X.transpose(0, 2, 1, 3).flatten(dim = 1)$

Results - Accuracy

Missed accuracy with high mask count, scene-wise receptive field scenarios

- PASCAL VOC, Oxford PET - Full

Preserved accuracy in low mask count, local receptive field scenarios

- Oxford PET - Small, MSD

Method	VOC [9]			PET_F [20]			PET_S [20]			MSD Heart [1]			MSD Brain [1]			MSD Lung [1]		
	$\mathcal{L}_{CE}$	$\text{mAP}_{0.5}$	mIOU	$\mathcal{L}_{CE}$	$\text{mAP}_{0.5}$	mIOU	$\mathcal{L}_{CE}$	$\text{mAP}_{0.5}$	mIOU	$\mathcal{L}_{CE}$	$\text{mAP}_{0.5}$	mIOU	$\mathcal{L}_{CE}$	$\text{mAP}_{0.5}$	mIOU	$\mathcal{L}_{CE}$	$\text{mAP}_{0.5}$	mIOU
U-Net_4 [23]	1.985	0.0541	0.182	2.206	0.0329	0.316	0.243	0.7717	0.713	0.0223	0.4305	0.063	0.0455	0.2450	0.001	0.0010	0.5542	**0.009**
ResNet_{34} [11]	1.321	0.0806	0.332	**0.648**	0.0503	0.597	0.179	0.9098	0.801	0.0087	0.4375	0.065	0.1305	0.4188	0.079	0.0010	0.5529	0.009
ResNet_{50} [11]	**1.079**	**0.0921**	**0.372**	1.027	**0.0512**	**0.599**	**0.189**	**0.9303**	**0.815**	0.0086	0.4368	0.065	0.0496	0.5419	0.010	**0.0007**	0.5566	0.009
MobileNet [12]	2.007	0.0480	0.166	2.386	0.0329	0.252	0.262	0.6091	0.664	0.0288	0.3265	0.047	**0.0351**	0.5764	0.011	0.0010	**0.5787**	0.008
$\text{OneNet}_{e,4}$	2.144	0.0485	0.160	2.713	0.0279	0.216	0.309	0.5176	0.636	**0.0041**	**0.4396**	**0.066**	0.0363	**0.5789**	**0.105**	0.0007	0.5531	0.009
$\text{OneNet}_{ed,4}$	2.553	0.0363	0.149	3.080	0.0227	0.172	0.437	0.2179	0.535	0.0076	0.4187	0.062	0.0455	0.5415	0.099	0.0009	0.5510	0.008

Results - Reduction

Encoder Replacement — 47% reduction in model size, accuracy preserved

Encoder/Decoder Replacement — 71% reduction in model size, accuracy drop 5%~13%

Method	# Param (M)	Param (MB)	FLOPS (GB)	Memory (MB)
U-Net$_4$ [23]	31.04	124.03	104.72	509.61
U-Net$_5$ [23]	124.42	497.41	130.80	524.29
ResNet$_{34}$ [11]	25.05	98.07	29.40	**241.17**
ResNet$_{50}$ [11]	74.07	287.83	84.98	450.36
MobileNet [12]	14.40	57.47	83.96	671.09
OneNet$_{e,4}$	16.39	65.42	78.42	639.63
OneNet$_{e,5}$	65.73	262.63	98.82	656.41
OneNet$_{ed,4}$	**9.08**	**36.30**	**22.92**	799.01
OneNet$_{ed,5}$	36.38	145.47	39.00	885.00

Thank You