Transitional Asymmetric Non-local Neural Network For Real-World Dirt Road Segmentation

Yooseung Wang and Jihun Park*
Agency for Defense Development
{yswang,jhpark_a}@add.re.kr

Introduction

Problem

Understanding dirt roads in real-time by classifying pixel-level semantic classes is essential technique for autonomous driving.

Motivation

- Many approaches (See Fig 1.) have been proposed to reinforce instant context information by capturing long-range dependencies by ***skip connections*** in CNNs.
- Recent studies have proposed ***non-local attention modules*** as a bottleneck between stages to enlarge receptive-field size of previous convolutional operations.

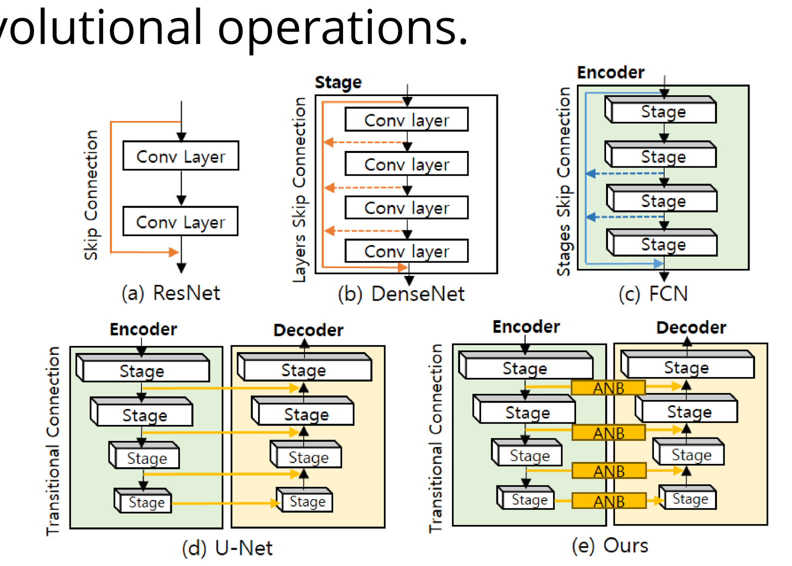


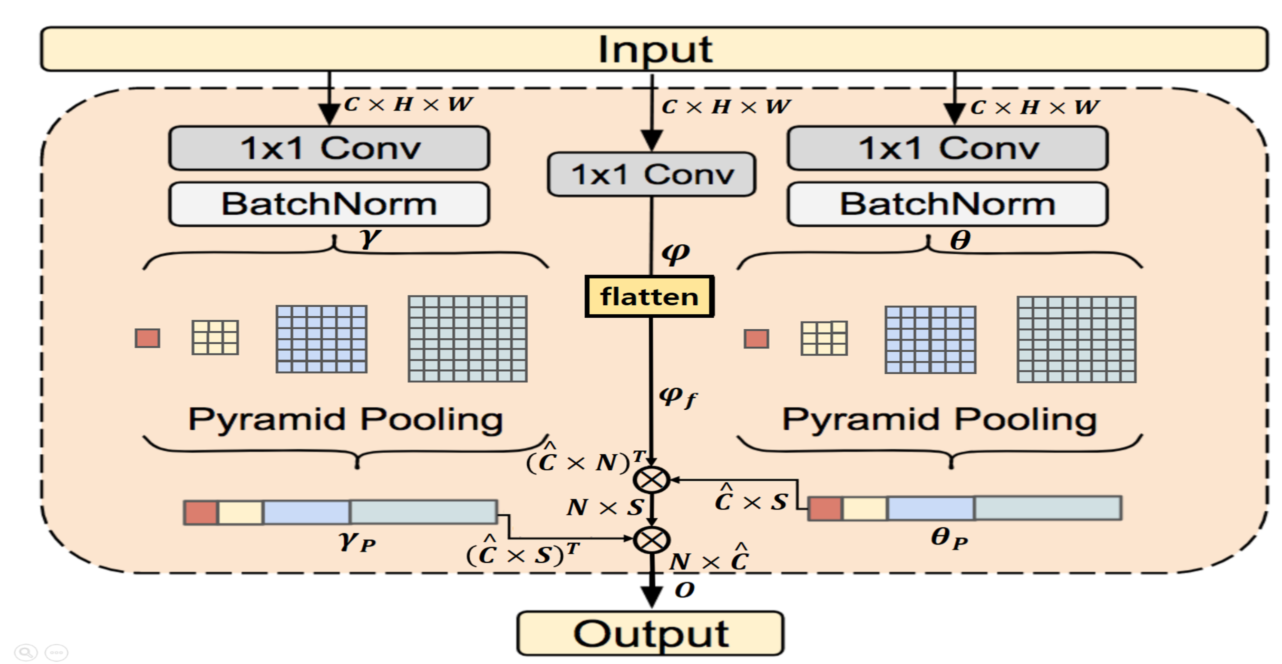
Figure 1. Recent proposed skip connections and proposed transitional connection with ANBs

Contributions

- We present TAN-Net that incorporates ANBs in transitional manner on SOTA semantic segmentation model (HarDNet) and TAN-Net outperforms previous SOTA model.
- We examined applicability of the semantic segmentations methods on dirt road dataset. Our dataset consists of 3k images collected from eight different locations of unpaved roads.
- Our ablation studies show that applying attention modules in transitional manner improve accuracies and our adopted ANB is effective than others.

Methodology

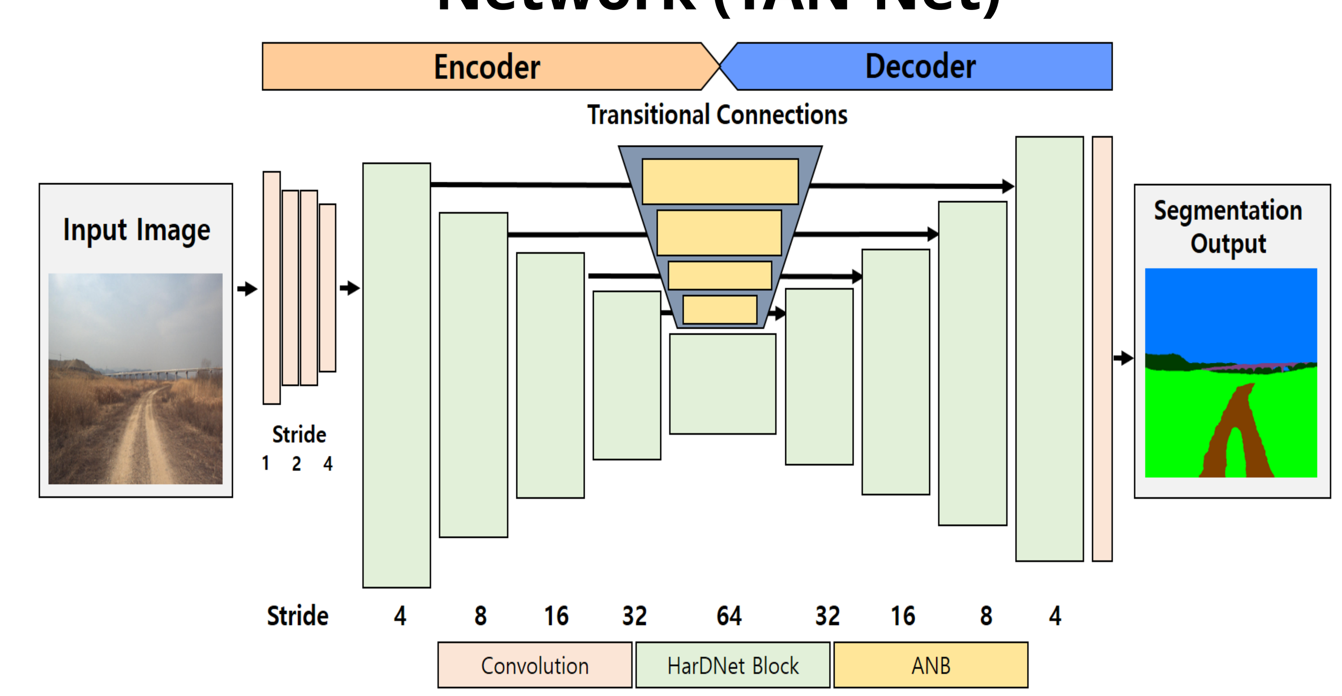
Asymmetric Nonlocal Block (ANB)*



ANB efficiently reduce the computational cost of multiplications by pyramid pooling of given two input.

* Asymmetric Non-local Neural Networks for Semantic Segmentation, Zhu et al., (ICCV2019)

Transitional Asymmetric Nonlocal Network (TAN-Net)



TAN-Net contains four ANBs into the all transitional connections to enhance pixel-level global context understanding.

Study Results

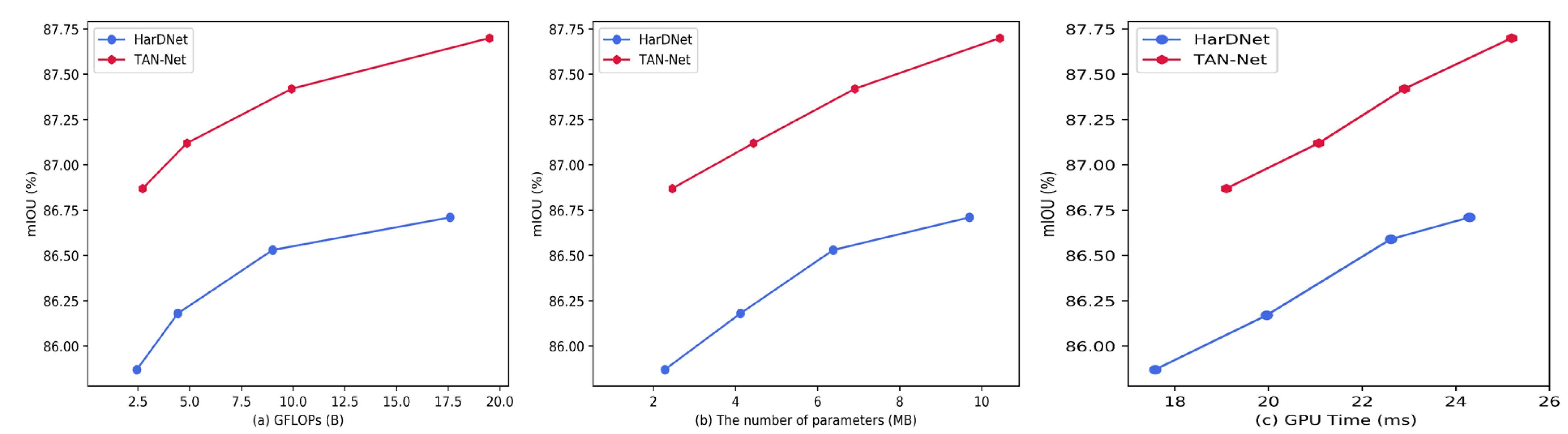
Comparisons with SOTA Real-Time SemSeg Models

Models	mIOU(%)	GFLOPs(B)	Params(MB)
ResNext50 [28]	79.85	29.50	23.12
ResNext101	81.25	49.15	42.27
ResNet50 [11]	85.83	32.63	23.52
ResNet101	85.90	52.11	42.51
ResNet152	86.07	71.63	58.16
UNet [22]	85.31	7.19	4.56
SegNet [1]	80.39	161.32	29.45
ICNet [33]	85.56	3.68	7.75
FCHarDNet62 [2]	85.87	2.45	2.28
FCHarDNet70	86.17	4.43	4.12
FCHarDNet78	86.59	9.01	6.38
FCHarDNet86	86.71	17.59	9.70
TAN-Net62(ours)	86.87	2.73	2.46
TAN-Net70	87.12	4.87	4.44
TAN-Net78	87.42	9.93	6.91
TAN-Net86	**87.70**	19.49	10.44

All TAN-Net(62,70,78,86) with comparable small parameter size and less GFLOPs outperform the SOTA models of urban road scene segmentation.

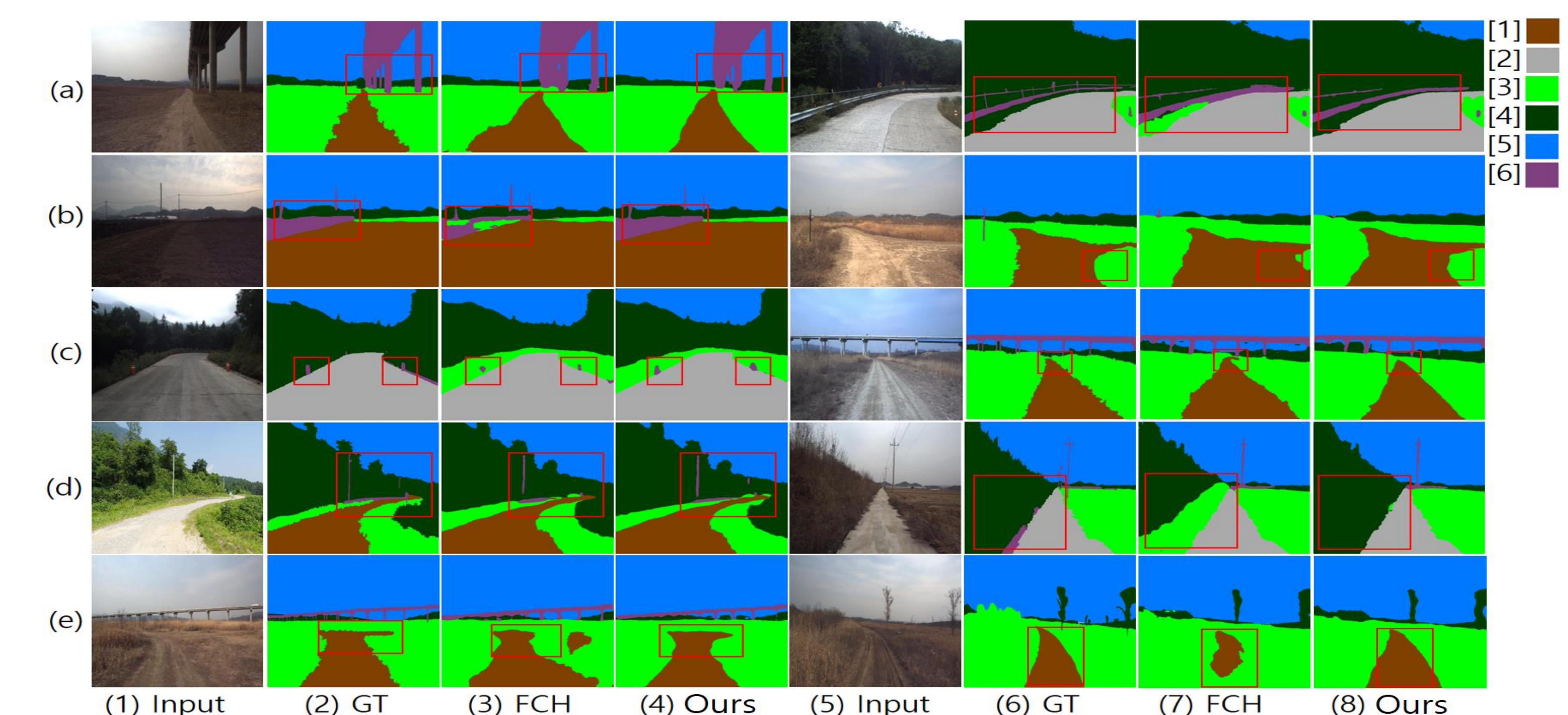
Study Results

HarDNet (SOTA, blue) vs TAN-Net(Ours, red)



TAN-Net are more faster (GFLOPs and GPU inference time) and lighter (parameter size) than FCHarDNet.

Qualitative Results



Our TAN-Net achieves higher accuracy qualitatively, especially for drivable regions [1] and artifacts [6].

Ablation Studies

Bottleneck and transitional usage of attention modules

Usage Type	Stage 1	Stage 2	stage 3	Stage 4	mIOU
Bottleneck	-	-	-	O	83.01
Transitional	-	-	-	O	**86.47**
Bottleneck	-	-	O	O	83.70
Transitional	-	-	O	O	**86.54**
Bottleneck	-	O	O	O	83.82
Transitional	-	O	O	O	**86.62**
Bottleneck	O	O	O	O	83.93
Transitional	O	O	O	O	**86.87**

Even in the same parameter size and GFLOPs, transitional connections of ANB in TAN-Net are better than bottleneck connections.

Different attention modules in TAN-Net

Attention Module	mIOU (%)	GFLOPs (B)	Params (MB)
Baseline (None)	85.87	2.45	2.28
CBAM [27]	86.12	2.59	2.35
CGNLB [30]	86.15	2.66	2.37
NLB [26]	86.17	3.28	2.38
AFNLB [36]	86.25	2.73	2.44
APNLB [36]	86.28	2.73	2.46
ANB(Ours)	**86.87**	2.73	2.46

Our re-designed ANB on TAN-Net achieves the best mIOU among attention modules. (blocks)