Transtional Asymmetric Non-local Neural Network for Real-Time Dirt Road Segmentation

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

ICPR 2020

CONTENTS

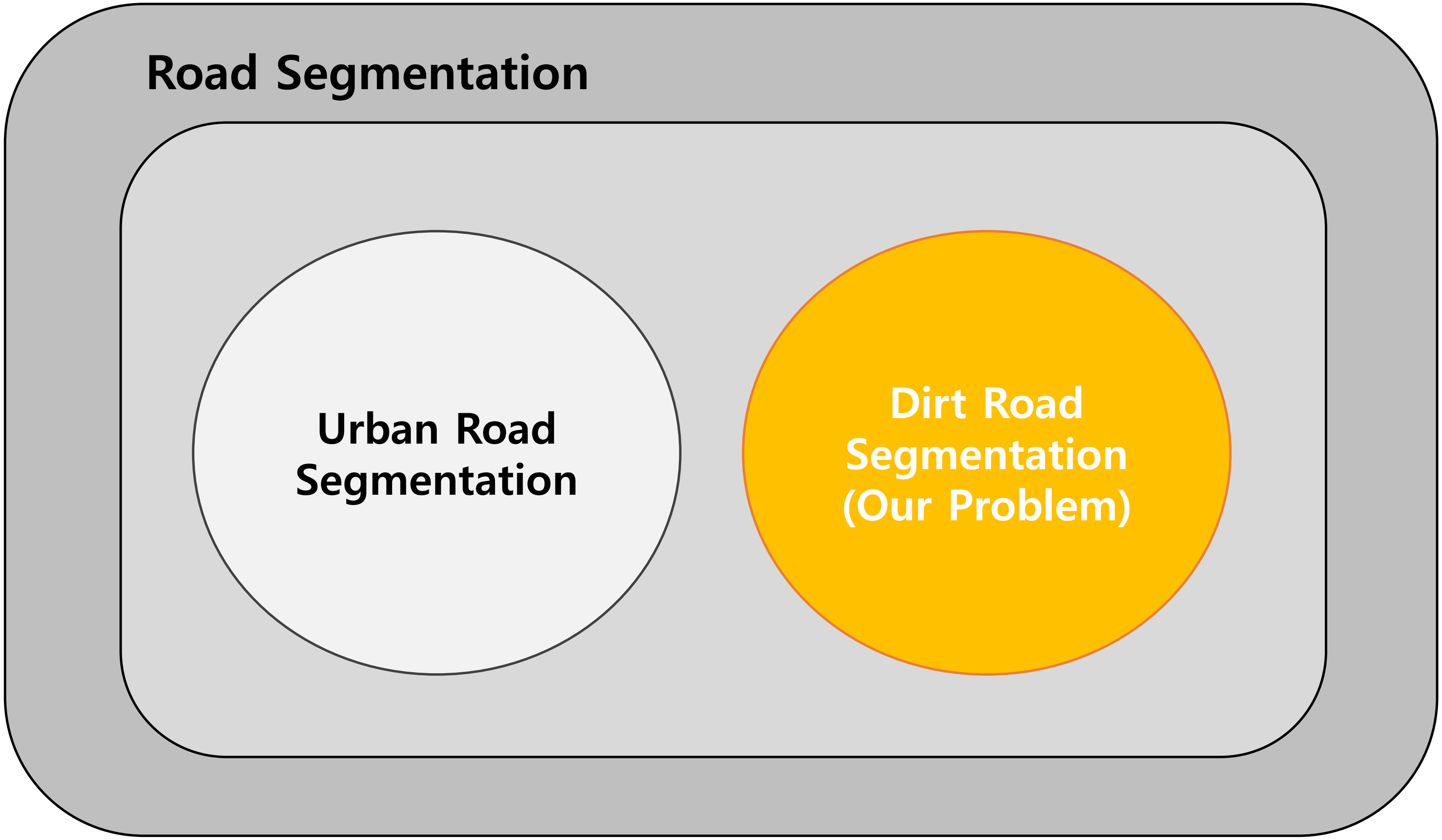
01

Problem Definition

- Dirt Road Semantic Segmentation
- **Real-Time** Dirt Road Semantic Segmentation

01. Problem Definition – Dirt Road Semantic Segmentation

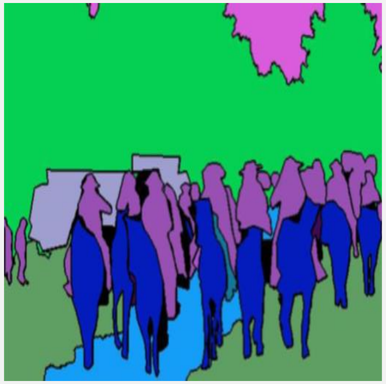
Semantic Segmentation classifies every pixel values in an input image into semantic classes



01. Problem Definition – Examples

Semantic Segmentation (Example from COCO)

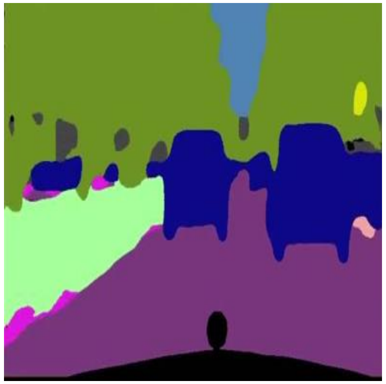
Input



GT

Urban Road Segmentation (Example from Cityscapes)

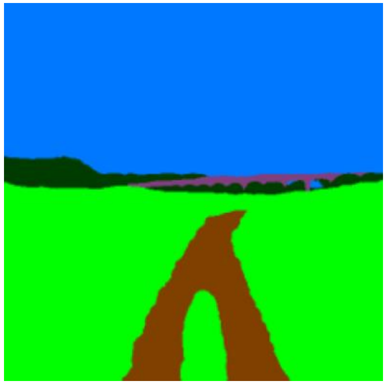
Input



GT

Dirt Road Segmentation

Input



GT

Why Challenging?

-Vague boundary between drivable and non-drivable areas.

Why Crucial?

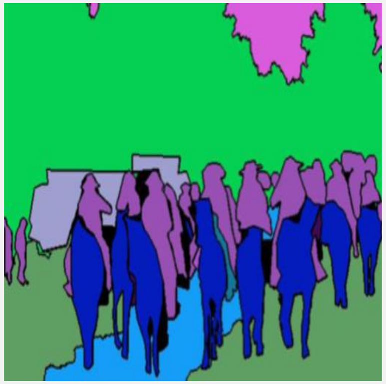
-Significant portion of the road in the world is not paved yet.

-Advanced dirt road segmentation method are required to realization of fully autonomous vehicles.

01. Problem Definition – Examples

Semantic Segmentation (Example from COCO)

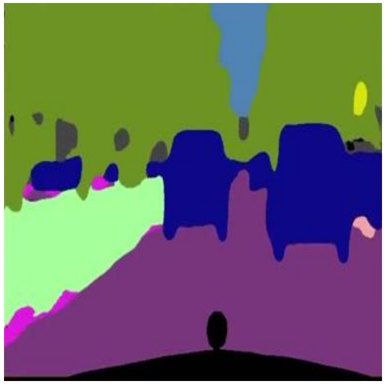
Input



GT

Urban Road Segmentation (Example from Cityscapes)

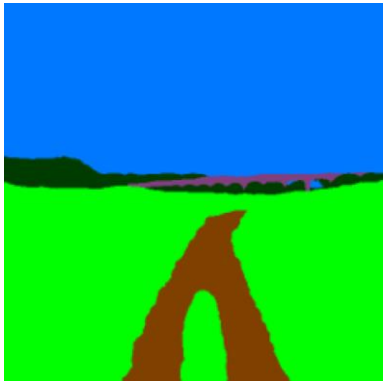
Input



GT

Real-Time Dirt Road Segmentation

Input



GT

Why Crucial?

-Real-time speed for road segmentation is required to fully autonomous driving

02

Method

- Motivation
- Asymmetric Non-local Block (ANB)
- Proposed TAN-Net

02. Method - Motivation

Recent Studies 1
Attention Modules (=Blocks)

1.Nonlocal Block (NB) (Kaiming He, CVPR 2018)

2.Asymmetric Pyramid Nonlocal Block (APNB)

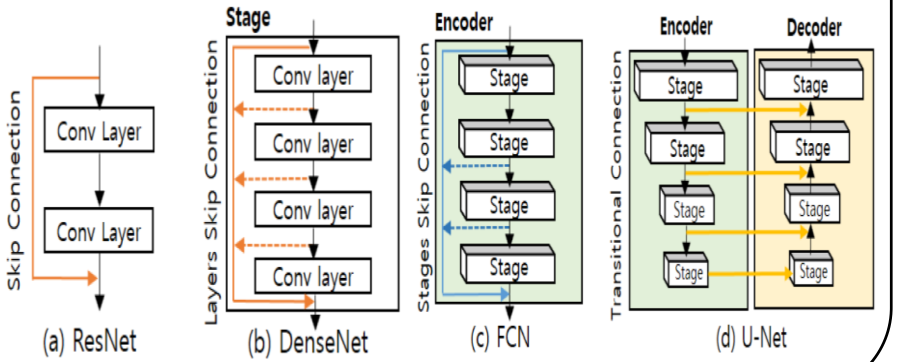
3.Compact Generalized Nonlocal Block (CGNB)

4.Convolutional Bottleneck Attention Module (CBAM)

5.Bottleneck Attention Module (BAM)

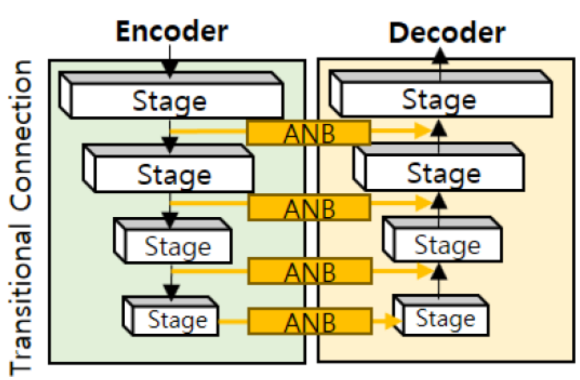
…

Recent Studies 2
Enlarging distance of skip connections



+

FC-HarDNet (Urban Road Segmentation SOTA model, Encoder-Decoder Architecture)



TAN-Net includes ANBs into the transitional connections (Encoder-Decoder level skip connections) to capture global context between distant encoder-decoder layers.

Asymmetric Nonlocal Block (ANB)*

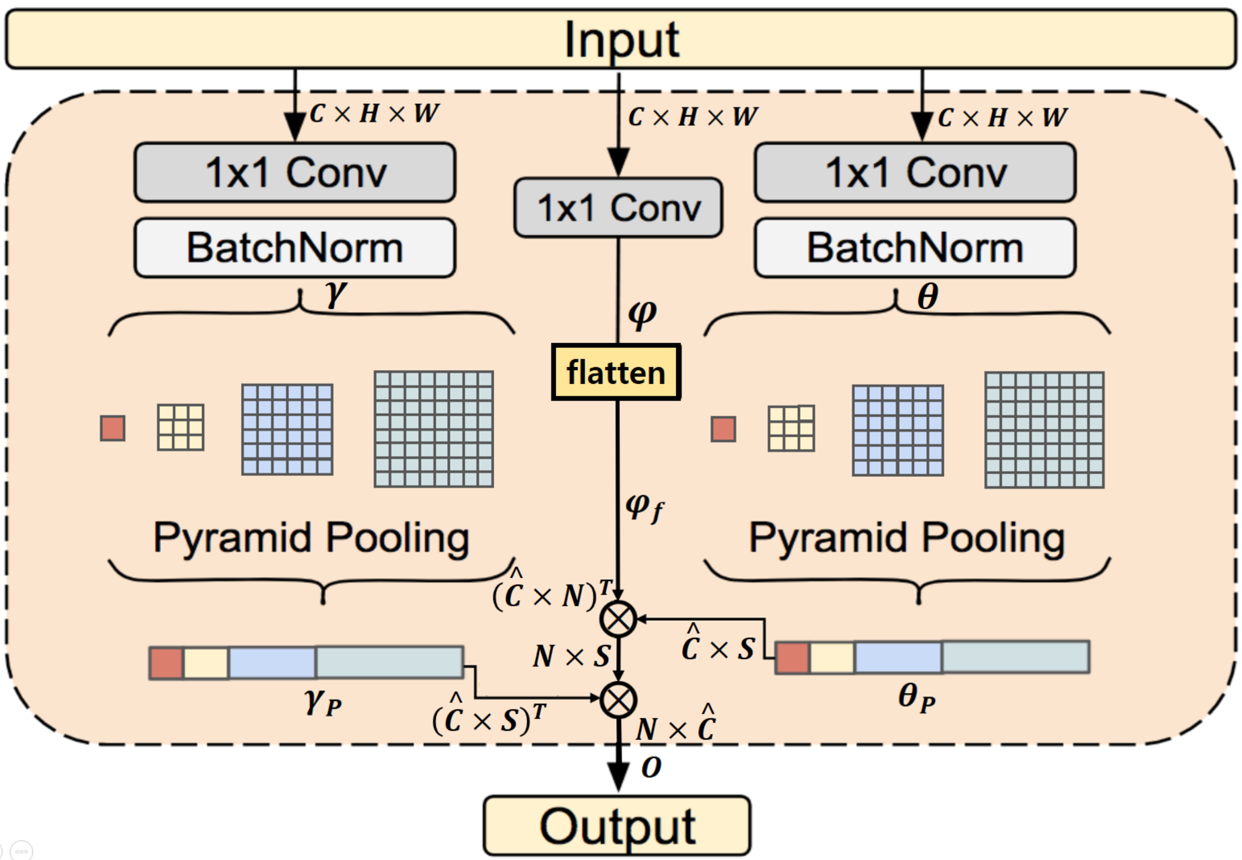


Figure. Asymmetric Nonlocal Block

Steps

1. Input ($\in R^{CxHxW}$) is divided into the three parts of the convolution module to get $(\gamma, \varphi, \theta)$

2. γ_P and $\theta_P \in R^{C\hat{}xS}$ are computed by pyramid pooling operations of γ *and* θ with pooling ratios of (1,2,3,6) where $S = 1^2 + 2^2 + 3^2 + 6^2$

φ is flatten to $\varphi_f \in R^{C\hat{}xN}$

3. Three outputs φ_f γ_P and θ_P are multiplied to get final output ***O***.

*Multiplication implies capturing pixel-level global context of input feature.

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

03. Method

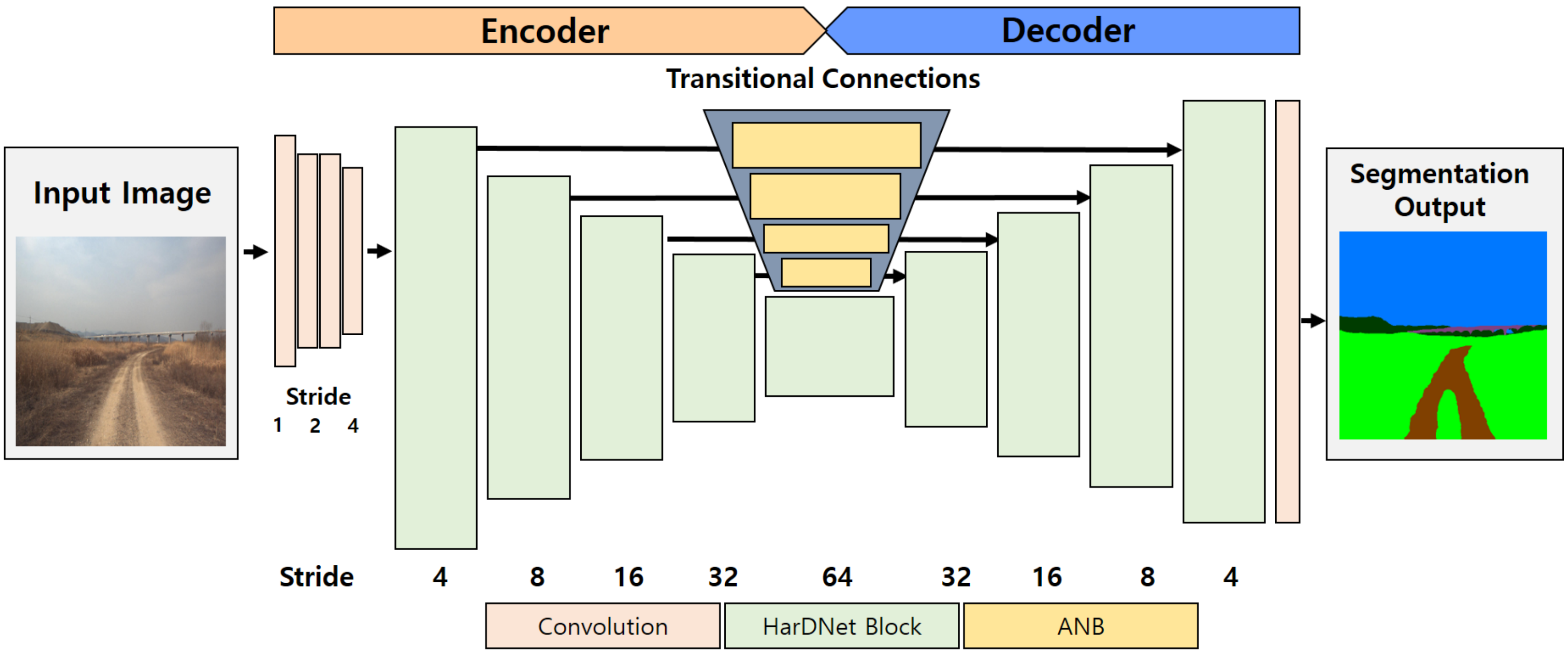


Figure. Overall Architecture of our proposed TAN-Net

-TAN-Net includes transitional ANB connections with four ANBs in FCHarDNet backbone network.

04

Study Results

- Comparisons with the SOTA
- FCHarDNet vs TAN-Net
- Transitional vs Bottleneck
- Ablation studies – Modules / Stages

Experiment Setting

Real-World Dirt Roads Dataset

- Few objects such as people or cars
- Road boundary is ambiguous
- We have total 7 classes (unpaved roads, paved roads, grass, forest, mountains, sky, objects, and voids)

To evaluate the proposed approach, we conducted experiments on a real-world dirt road dataset of eight different locations.

Comparisons with the state-of-the-art methods of urban semantic segmentation

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

Our model : TAN-Net62,70,78,86

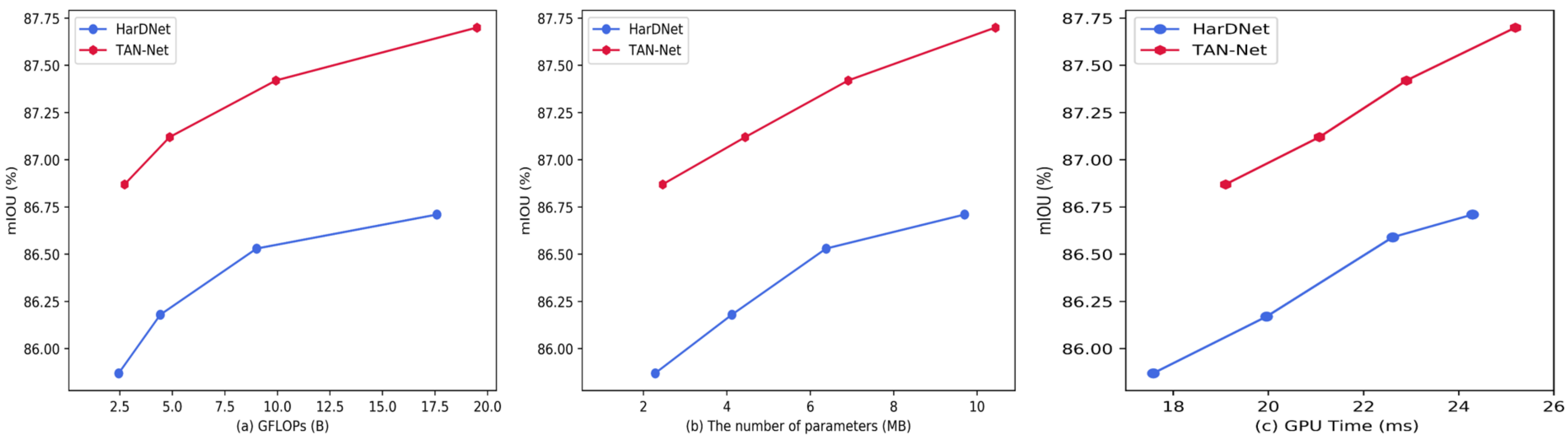
Comparatively heavy models
- ResNeXt-50,101
- ResNet-50,101,152
- U-Net

State-of-the-art models
- FCHarDNet
- ICNet
- SegNet

Our models outperform all the other published methods.

FCHarDNet vs TAN-Net : Effectiveness Analysis

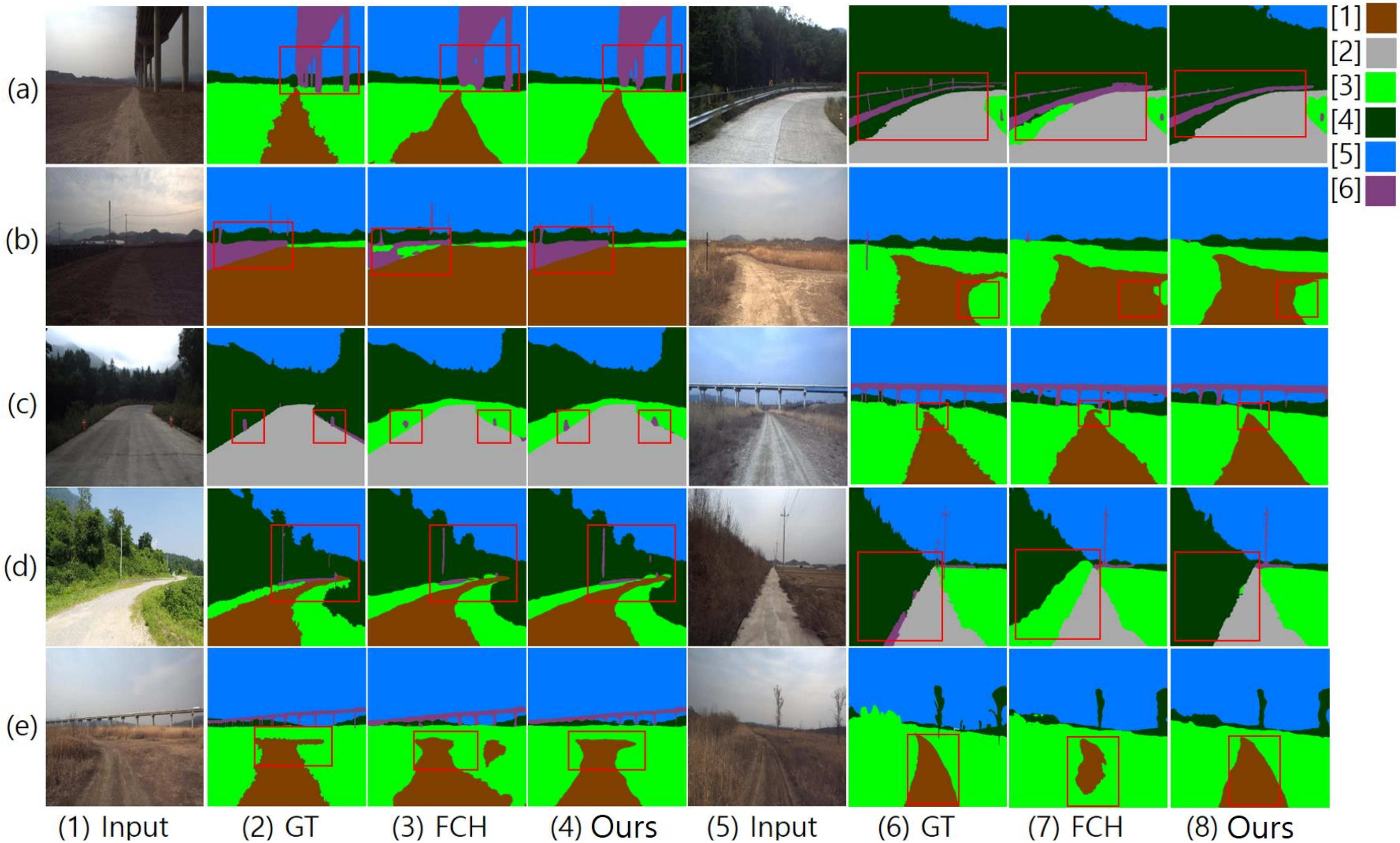
in GFLOPs, the number of parameters, GPU Time



TAN-Net (with Less GFLOPS, small parameter size, and faster inference time)

significantly **achieves higher mIOU** than FCHarDNet

FCHarDNet vs TAN-Net : Qualitative Results



Our TAN-Net accurately predicts dirt roads [1] and artifacts [6] Better than FCHarDNet

Transitional and Bottleneck Usage of ANBs

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**

Transitional : Adding ANBs **in Encoder-Decoder skip connections.**
Bottleneck : Adding 2*ANBs (shared weights) **between layers in Encoder and Decoder**

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

Ablation study – Different Attention Modules

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

*For fair comparison, we put modules into the every transitional connections in FCHarDNet.

Our re-designed ANB on TAN-Net achieves the highest accuracy outperforming other attention modules.

Ablation study – Different Stages

# Blocks	Stage 1	Stage 2	Stage 3	Stage 4	mIOU	Flops
0	-	-	-	-	85.87	2.45
1	-	-	-	O	86.47	2.49
	-	-	O	-	86.51	2.50
	-	O	-	-	**86.53**	2.53
	O	-	-	-	86.52	2.56
2	-	-	O	O	86.54	2.54
	-	O	-	O	**86.73**	2.57
	-	O	O	-	86.59	2.58
	O	-	-	O	86.55	2.60
	O	-	O	-	86.53	2.61
	O	O	-	-	86.58	2.64
3	-	O	O	O	86.55	2.62
	O	-	O	O	86.55	2.65
	O	O	-	O	86.57	2.68
	O	O	O	-	**86.62**	2.69
4	O	O	O	O	**86.87**	2.73

The mIOU gradually increase
as the number of ANB are added into the network.

Conclusion

- Real-time dirt road segmentation is crucial for fully autonomous driving

- We propose TAN-Net with ANBs in transitional connections with lower computational cost compared to HarDNet.

- Our ablation study indicate the effectiveness of transitional usages of attention modules.

THANK YOU