

Bi-PointFlowNet

6차 보고

Experiment – Half Channel Student

0. Scratch Student

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.052	0.798	0.942	0.286	2.903	0.846
KITTI	0.073	0.816	0.899	0.228	2.525	0.8541

1. Basic Knowledge Distillation Student

Gamma: 0.3

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.046	0.833	0.953	0.246	2.602	0.871
KITTI	0.072	0.833	0.907	0.206	2.393	0.868

Experiment – DecreasedLinear

Architecture

1. DepthwiseConv + PointwiseConv + Dec-Linear

	Baseline	Student
FLOPs	13.3G	→ 12.6G
Param	7.98M	→ 4.79M

```

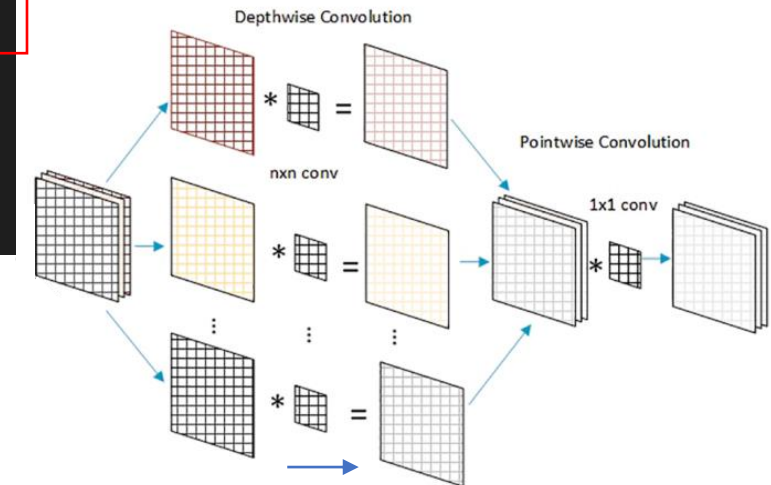
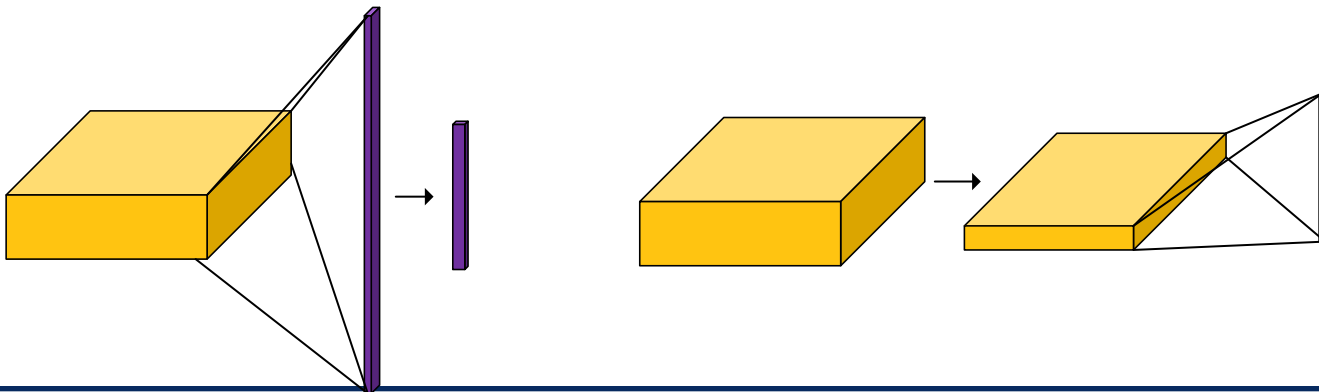
8 level4.weightnet.mlp_bns.0.weight torch.Size([8])
8 level4.weightnet.mlp_bns.0.bias torch.Size([8])
8 level4.weightnet.mlp_bns.1.weight torch.Size([8])
8 level4.weightnet.mlp_bns.1.bias torch.Size([8])
16 level4.weightnet.mlp_bns.2.weight torch.Size([16])
16 level4.weightnet.mlp_bns.2.bias torch.Size([16])
2109440 level4.linear.weight torch.Size([256, 8240])
256 level4.linear.bias torch.Size([256])
16384 deconv4_3.composed_module.0.weight torch.Size([64, 256, 1])
64 deconv4_3.composed_module.0.bias torch.Size([64])
16384 deconv3_2.composed_module.0.weight torch.Size([64, 256, 1])
64 deconv3_2.composed_module.0.bias torch.Size([64])
4096 deconv2_1.composed_module.0.weight torch.Size([32, 128, 1])
32 deconv2_1.composed_module.0.bias torch.Size([32])
2048 deconv1_0.composed_module.0.weight torch.Size([32, 64, 1])
32 deconv1_0.composed_module.0.bias torch.Size([32])

```

```

8 level4.weightnet.mlp_bns.0.weight torch.Size([8])
8 level4.weightnet.mlp_bns.0.bias torch.Size([8])
8 level4.weightnet.mlp_bns.1.weight torch.Size([8])
8 level4.weightnet.mlp_bns.1.bias torch.Size([8])
16 level4.weightnet.mlp_bns.2.weight torch.Size([16])
16 level4.weightnet.mlp_bns.2.bias torch.Size([16])
515 level4.conv.weight torch.Size([515, 1, 1, 1])
515 level4.conv.bias torch.Size([515])
8240 level4.pointwise.weight torch.Size([16, 515, 1, 1])
16 level4.pointwise.bias torch.Size([16])
16384 deconv4_3.composed_module.0.weight torch.Size([64, 256, 1])
64 deconv4_3.composed_module.0.bias torch.Size([64])
16384 deconv3_2.composed_module.0.weight torch.Size([64, 256, 1])
64 deconv3_2.composed_module.0.bias torch.Size([64])
4096 deconv2_1.composed_module.0.weight torch.Size([32, 128, 1])
32 deconv2_1.composed_module.0.bias torch.Size([32])
2048 deconv1_0.composed_module.0.weight torch.Size([32, 64, 1])
32 deconv1_0.composed_module.0.bias torch.Size([32])

```



Experiment

Scratch Student (Epoch: 484)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.045	0.840	0.955	0.238	2.495	0.882
KITTI	0.047	0.873	0.938	0.170	1.670	0.911

Hint Training Knowledge Distillation(Epoch: 470)

Gamma: 0.3

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.043	0.852	0.960	0.223	2.397	0.892
KITTI	0.049	0.878	0.938	0.167	0.1675	0.905

Experiment – WeightNet4 ($C_{mid} = 4, K = 16$)

Architecture

- Decreasing Weight Net Channel $16 \rightarrow 4$

	Baseline	Weightnet=8	Weightnet=4
FLOPs	13.3G	9.36G	7.37G
Param	7.98M	4.47M	2.73M

FlyingThings3D Dataset

Scratch(7일)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.0325	0.902	0.974	0.163	1.833	0.915
KITTI	0.0319	0.923	0.964	0.140	1.118	0.949

Hint Training Loss(9일)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.0313	0.905	0.975	0.159	1.755	0.917
KITTI	0.0325	0.929	0.961	0.137	1.136	0.946

Experiment – WeightNet4

FlyingThings3D Dataset

Attentive Imitation Loss(측정 예정)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings						
KITTI						

Bi-Directional Hint Training Loss(현재 측정 중: 50 Epoch)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.0492					
KITTI						

Experiment – WeightNet48 (C_{mid} = 4,8, K = 16)

Architecture

- Decreasing Level 1, 2 Weight Net Channel 16 → 8
- Decreasing Level 3, 4 Weight Net Channel 16 → 4

	Baseline	Weightnet=8	Weightnet=48
FLOPs	13.3G	9.36G	7.65G
Param	7.98M	4.47M	3.85M

FlyingThings3D Dataset

Scratch(현재 측정 중: 10 Epoch)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.0814					
KITTI						

Hint Training Loss (측정 예정)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings						
KITTI						

Experiment – WeightNet48

FlyingThings3D Dataset

Attentive Imitation Loss (측정 예정)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings						
KITTI						

Bi-Directional Hint Training Loss (측정 예정)

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings						
KITTI						

Experiment – Fast Distillation

중반 Validation 측정 과정 축소

- 초반: 학습을 잘하는지 확인용
- 후반: Best 성능을 내는 Checkpoints 저장

Epoch 20 ~ 420 → 5의 배수마다 측정(25, 30, ...)

→ 약 28시간 축소...

Further Study

1. Meta Knowledge Distillation

- Teacher도 Student 모델이 학습하는 특성에 맞춰서 학습
- W. Zhou, **BERT Learns to Teach: Knowledge Distillation with Meta Learning**, Association for Computational Linguistics, 2022

2. Student 성능 > Teacher 성능

- Knowledge distillation 방법이 효과적일지
- Early Stop Knowledge Distillation 방법..?

Teacher(baseline): 0.0288

Scratch Student(c_mid=8, k=16): 0.0280

Experiment

Minimized FlyingThings3D Dataset

- 기존 Dataset에서 1/5 축소(115GB -> 22GB)
- 경향성, hyperparam 확인을 위해서 개인컴에서 활용(3일 소요)
- 연구실컴에서는 Full Dataset 활용

Baseline Result

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.045	0.819	0.955	0.286	2.446	0.871
KITTI	0.037	0.900	0.957	0.163	1.403	0.929

Weightnet=8

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.047	0.802	0.950	0.298	2.450	0.864
KITTI	0.035	0.886	0.961	0.166	1.462	0.922

Weightnet=4

Dataset	EPE3D(↓)	ACC3DS(↑)	ACC3DR(↑)	Outliers3D(↓)	EPE2D(↓)	ACC2D(↑)
FlyingThings	0.048	0.789	0.948	0.310	2.547	0.856
KITTI	0.040	0.887	0.949	0.174	1.575	0.920