Segmentation의 Pseudo-Labeling 성능 분석

전우민, 오태근, 이성진
동서울대학교 신산업특화단
2104023@du.ac.kr, tgoh@du.ac.kr, sungjinlee@du.ac.kr

Analysis of the Pseudo-Labeling Performance of Segmentation

Woomin Jun, Taegeun Oh, Seongjin Lee
DongSeoul Univ.

요 약

본 논문은 자율주행 기술의 핵심 부분인 Drivable Area Segmentation에 대규모 데이터세트를 활용하는 과정에서 발생 발생하는 높은 Labeling 비용 문제를 해결하기 위해 Pseudo-Labeling 기술을 탐구한다. 이 기술은 통해 모델이 스스로 레이블을 생성하고 학습하여 Labeling 비용을 줄이는 동시에 성능 향상을 목표로 연구를 진행했다. 본 연구로 통해 같은 모델로 Labeling하고 학습하면 성능이 0.18% 안 좋아지므로 다른 모델을 활용해서 Pseudo-Labeling 기술을 활용하면 좋고, Pseudo-Labeling은 Labeling 하는 모델의 성능에 비례해서 학습하는 모델의 성능이 항상 한다.

Ⅰ. 서 론

자율 주행 기술의 발전은 지난 몇 년간 비약적인 진보를 이루었다. 이러한 진보의 핵심에는 대규모 데이터세트의 활용과 이를 통한 정밀한 알고리즘 학습이 자리 잡고 있다. 특히, BDD100K[1] 데이터세트는 다양한 주행 환경과 상황을 포괄하는 방대한 데이터를 제공함으로써, 이 분야의 연구에 있어 중요한 자원 되고 있다. 본 연구는 BDD100K의 Drivable Area Segmentation 데이터세트를 활용하여, Pseudo-Labeling[2] 방식이 Drivable Area Segmentation 학습에 성능 향상에 어떻게 기여될 수 있는지 탐구한다. Pseudo-Labeling이란 기술을 사용하여 레이블이 지정되지 않은 데이터에 대해, 모델이 스스로가 레이블을 생성하고 학습 활용하는 방식 즉 semi supervised learning[3]을 말한다. 이러한 방법은 대규모 데이터세트에서 엄청난 Labeling 비용을 감소시키기 위해서 사용한다.

Ⅱ. 본론

본 논문에 사용하는 모델로는 Efficient-Net-V2M[4]-FPN(Feature Pyramid Networks) [5] 기반으로 학습하였고 ResNet50[6] , Swin-Transformer-L [7]도 사용해서 Pseudo-Labeling을 추가해서 성능을 비교했다. Vision 분야에서 Transfomer 기술을 적용한 ViT(Vision Transformer)[8] 제안되었다. 원래 자연어 처리에 사용되던 Transformer 기술을 사용하여, 기존의 CNN(Convolutional Neural Network)의 기술을 뛰어넘는 새로운 대안으로 자리 잡았다.

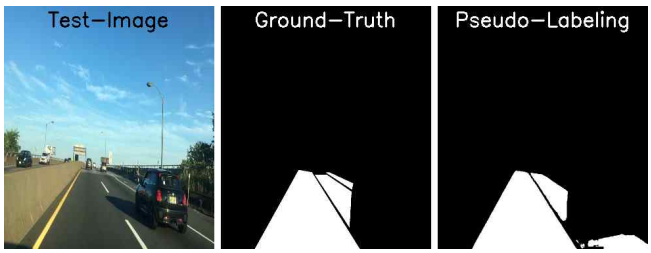


그림 1. Pseudo-Labeling 예제

1. 실험

본 논문에서는 BDD100K의 데이터 7만 장을 절반으로 나눈 후 A, B로 지정하고 실험을 진행하였다. 데이터세트 A를 사용하여 초기 모델 학습을 수행한 후, 이를 기반으로 데이터세트 B에 대해 두 가지 다른 접근 방식을 이용하여 실험을 진행했습니다.

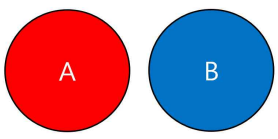
첫 번째 접근 방식에서는 데이터세트 B를 그림 1과 같이 Labeling 하여 'B-Pseudo_Data'라는 새로운 데이터세트를 생성합니다. 이후, 데이터세트 A로 사전 훈련된 모델에 B-Pseudo_Data와 B-Ground-Truth-Data를 사용하여 추가 학습을 진행합니다.

두 번째 접근 방식에서는 A의 사전학습 없이 오직 데이터세트 B만을 사용하여 모델을 학습시킵니다. 이 실험은 A 데이터세트의 사전학습이 없을 때 B 데이터세트에 대한 모델의 성능을 변화의 초점을 맞췄다.

또한, 별도의 실험에서는 ResNet, Swin-Transformer, EfficientNetV2M 모델을 사용하여 데이터세트 A에 대해 학습을 수행하고, 이 모델을 사용하여 데이터세트 B에 대한 'Pseudo_Labeling'을 진행합니다. 이렇게 생성된 데이터세트는 EfficientNetV2M-FPN에 학습을 수행한다. 이 실험은 모델의 성능에 따른 성능에 영향을 알기 위해서 실험하였다.

2. 실험 환경

이용한 실험을 위한 환경으로는 Ubuntu-20.04를 사용하며, 2-way RTX 4090 GPU에 Tenorflow, pytorch를 사용하였다. 훈련으로는 50 epoch 중 최고성능을 IoU(Intersection Over Union)로 도출하였고 손실함수로는 Dice Loss[9]를 사용하였다. 최적화 알고리즘으로는 AdamW[10] 사용하였다.



$$IoU = \frac{A \cap B}{A \cup B}$$

$$\text{Dice Loss} = \frac{2 * |A \cap B|}{|A| + |B|}$$

∪은 합집합
∩은 교집합

그림 2. Dice Loss, IoU 예제

	IoU
Efficient-Net	49.47%
ResNet	49.42%
Swin-Transformer	50.12

표 1. A-데이터 학습 Intersection Over Union 성능 표

	IoU
A to B(Ground-Truth)	50.05%
A to B(Pseudo)	49.61%
B(Ground-Truth)	49.28%
B(Pseudo_EfficientNetV2)	48.97%
B(Pseudo_ResNet)	49.07%
B(Pseudo_Swin)	49.15%

표 2. Intersection Over Union 성능 표

Ⅲ. 결론

본 연구에서는 표 2에서 사전학습하고 Pseudo-Labeling 한 성능은 A to B에서 GT와 Pseudo 차이는 0.44%가 나고 B에서는 0.31% 가 차이가 난다. 사전학습을 하면 더욱 성능 차이가 나는데 동일한 모델로 만든 Pseudo-Data로 학습을 진행하면 과적합이 되어서 성능이 떨어지는 것이라고 다음 결과로 생각할 수 있다. 표 2에서 다른 Pseudo_Data로 했을 때 성능이 ResNet으로 Pseudo-Labeling 했을 때 성능 차이는 0.21% 이고 Swin-Trausformer로 봤을 때 성능은 0.13% 이다. Efficient-Net과 ResNet의 표1에 결과를 보면 ResNet이 성능이 안 좋다. 근데 Pseudo_ResNet의 성능 좋은 걸 보면 앞서 언급한 과적합이 문제에 보인다. 그러므로 동일한 모델로 Pseudo-Labeling 기술을 사용하면 과적합의 위험이 있다는 걸 알 수 있다.

표 1의 결과에서 Swin-Transformer , Efficient-Net ,ResNet 순서대로 성능이 좋다. 표 2에서 앞서 제일 좋은 Swin-Transformer 가 좋은 걸로 보면 Pseudo-Labeling은 Labeling 하는 모델의 성능이 비례하는 걸 볼 수 있다. 따라서 Pseudo-Labeling 기술을 사용할 때는 높은 성능의 모델을 사용하는 게 좋다.

ACKNOWLEDGMENT

이 논문은 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2023 신산업분야 특화 선도전문대학 지원사업)

참 고 문 헌

[1] F. Yu, “BDD100K: A large-scale diverse driving video database,”BAIR, 2018.(https://bair.berkeley.edu/blog/2018/05/30/bdd/)

[2] E. Arazo, D. Ortego, P. Albert, N. O'Connor and K. McGuinness,“Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning”, *arXiv preprint axXiv:1908.02983v1*, 2019

[3] A. Oliver, A. Odena, C. Raffel, E. Cubuk, and I. Goodfellow, “Realistic Evaluation of Deep Semi-Supervised Learning Algorithms,” In Advancesin Neural Information Processing Systems(NeurIPS), 2018

[4] M. Tan and Q. Le,“EfficientNetV2: Smaller Models and Faster Training”, *arXiv preprint axXiv:2104.00298*, 2021

[5] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramidnetworks for object detection,” 2017 IEEEConf. CVPR, pp. 936-944, Honolulu, HI,USA, 2017

[6] K. He, X. Zhang, S. Ren and J. Sun,“Deep Residual Learning for Image Recognition”, *arXiv preprint axXiv:1512.03385*, 2015

[7] Z. Liu, Y. Liu, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin and B. Guo, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”, *arXiv preprint axXiv:2103.14030*, 2021

[8] A. Dossvitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby,“An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, *arXiv preprint axXiv:2010.11929*, 2020

[9] C. Sudre, W. Li, T. Vercauteren, S. Ourselin and M. Cardoso,“Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations”, *arXiv preprint axXiv:1707.03237*, 2017

[10] I. Loshcilov and F. Hutter,“Decoupled Weight Decay Regularization”, *arXiv preprint axXiv:1711.05101*, 2017