

TB-CompletionFormer : Two-branch Backbone 기반

향상된 깊이 완성 연구

김석영¹⁾·김찬수²⁾

전남대학교 IoT인공지능융합전공¹⁾·전남대학교 지능형모빌리티융합학과²⁾

TB-CompletionFormer : Improved Depth Completion based on Two-branch Backbone

Seokyoung Kim¹⁾ · Chansoo Kim^{*2)}

¹⁾ Department of IoT AI Convergence, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 61186, Korea

²⁾ Department of Intelligent Mobility, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 61186, Korea

Abstract : Dense depth estimation about surrounding environments is one of the key components in autonomous driving applications such as object detection and SLAM. To predict accurate depth information, depth completion aims to predict a dense depth map from a sparse depth map by LiDAR and a RGB image by camera. In the depth completion, it is essential to effectively fuse both LiDAR and camera modalities to achieve good performance. In this paper, we propose a two-branch architecture based on Vision Transformer, which is a more powerful and accurate multi-modal backbone, to fuse different modalities from heterogeneous sensors. To fuse two modalities, the proposed model consists of two branch modules: a coarse-branch and a fine-branch. Using an RGB image and a sparse depth map, the coarse-branch module generates a dense depth map focused on color information. The fine-branch module estimates a final dense depth map focused on depth information through the color-dominant depth map and the sparse depth map. Additionally, residual connections to the Joint Convolutional Attention and Transformer (JCAT) is added to preserve the feature information. As a result, experiments on NYUv2 dataset demonstrate that the proposed method outperforms the previous models.

Key words : Depth completion(깊이 완성), Vision Transformer(비전 트랜스포머), Deep Learning (딥러닝), Multi-modal Learning(멀티모달 러닝), Deep Learning (딥러닝), Residual Connection(잔차 연결), Computer Vision (컴퓨터 비전)

1. 서론

최근 자율주행 기술이 고도화됨에 따라, 정확한 환경 인식을 위해 다양한 센서를 기반으로 한 인지 연구들이 진행 중이다. 주변 환경을 인식하고, 깊이 정보를 추출하기 위해 주로 사용되는 센서로는 카메라와 라이다(LiDAR)가 있다. 카메라 센서는 물체의 색상과 질감을 포함한 고밀도

의 풍부한 정보를 제공한다는 장점이 있지만, 정확한 거리 정보를 추정하는데 한계가 존재한다. 반면에 라이다(LiDAR) 센서는 레이저 신호를 기반으로 거리를 측정하여, 매우 정확한 깊이 정보를 제공한다. 이 때 라이다의 점군데이터(point cloud)를 카메라 영상 평면으로 투영시킨 깊이 지도(depth map)가 일반적으로 사용된다. 하지만 이 깊이 지도는 매우 희소(sparse)하며, 객체 경계

주변에 노이즈가 많이 생기는 단점이 존재한다. 각 센서들이 가진 문제점들을 해결하기 위한 대안으로 센서 퓨전 (sensor fusion) 기술이 도입되었다. 이는 Table 1과 같은 상호보완적인 특성을 활용하여, 카메라와 라이다의 장점을 최대한으로 활용할 수 있다는 장점이 있다. 대표적으로 라이다의 희소한 깊이 지도(sparse depth map)와 RGB 이미지를 융합하여 정확하고 고밀도의 깊이 지도(dense depth map)를 생성하는 깊이 완성(depth completion) 기술이 제안되었다. 이는 한 장면(scene) 내 구조적 깊이를 추정함으로써 자율주행, 증강현실, 3D 재구성과 같은 다양한 3D 비전 응용 작업에 도움을 줄 수 있다.

Table 1 Comparison of Camera and LiDAR.

	카메라	라이다
Raw Data 밀도	높음	낮음
깊이 추정 정확도	낮음	높음
채도 인식	높음	낮음
3D 인식 성능	낮음	높음
비용	낮음	높음

최근 딥러닝의 발전에 따라 CNN(Convolutional Neural Network)를 기반으로 한 깊이 완성 기법들이 좋은 성능을 보여주었다. 대부분의 이전 방법들은 모델을 설계할 때, CNN을 기반으로 한 single-branch 구조를 사용하였다. 그러나 이러한 구조는 깊이 완성을 수행하는데 크게 2가지의 문제점이 존재한다.

첫 번째로, 깊이 완성 분야의 특성상 양식이 다른 라이다와 카메라 데이터들을 어떻게 융합하느냐에 따라 깊이 완성 성능이 결정되기 때문에, 단순히 single-branch를 기반으로 두 데이터를 결합(concat)하여 사용하는 접근 방식은 성능 향상에 한계가 존재한다. 이러한 문제점을 해결하고, 상호보완적인 두 데이터를 효과적으로 활용하기 위해 two-branch 구조 기반 모델이 제안되었다. 예를 들어 Sparse-to-Dense¹⁾ 모델은 두 개의 인코더를 기반으로 각 센서로부터 특징을 추

출한 후, 하나의 디코더로 융합한다. 또한 PENet²⁾, FusionNet³⁾, DeepLiDAR⁴⁾에서도 two-branch 형태로 모델 네트워크를 설계하였다. PENet은 각 센서들의 데이터를 개별적으로 학습하기 위해 색상 중심(color-dominant) 브랜치와 깊이 중심(depth-dominant) 브랜치로 구성된다. 이후 깊이 정제(depth refinement) 모듈인 CSPN++⁵⁾을 추가하여 효율적으로 깊이 완성을 수행한다. FusionNet은 지역적(local)인 특징과 전역적(global)인 특징 정보를 각각 추출하기 위해 two-branch로 구성된다. DeepLiDAR 또한 색상과 법선 벡터(surface normal) 브랜치로 나누어 모델이 이루어져 있다. 각 브랜치는 두 센서 데이터를 한번에 입력으로 받아 네트워크 내 다중 융합(multi modality fusion) 방식을 적용하여 깊이 완성을 수행하였다. 이와 같이 two-branch 기반 특징 학습을 통해 양식이 다른 두 센서 데이터들을 완전히 활용함으로써, 유의미한 깊이 정보 특징들을 학습할 수 있다.

두 번째 문제점으로, CNN 기반 깊이 완성 모델들은 커널(kernel)을 이용하여 인접한 픽셀 간의 관계를 파악함으로써 지역적인 특징 추출에 우수한 성능을 보이지만, 전역적인 공간에서 특징 추출 시 한계가 존재한다. 최근 이 문제점을 해결하기 위해 전역적인 공간에서 특징 추출을 잘 해내는 비전 트랜스포머(vision transformer)를 CNN 대신 깊이 완성 분야에 적용하기 시작하였다. GuideFormer⁶⁾ 모델은 처음으로 비전 트랜스포머를 깊이 완성 분야에 적용함으로써 성능 향상의 가능성을 보여주었다. 하지만 비전 트랜스포머의 구조 특성상 이미지 패치(patch)들을 벡터에 투영하는 과정에서 지역적인 특징 손실이 발생한다는 한계점은 여전히 존재하였다. 이를 보완하고자 CNN 기반 지역 특징들과, 비전 트랜스포머 기반 전역 특징들을 결합한 모델인 CompletionFormer⁷⁾가 제안되었다. 이는 CNN과 비전 트랜스포머의 장점만을 살려 하나의 JCAT(Joint Convolutional Attention and Transformer) 블록으로 통합함으로써, 모든 공간에 대한 특징 추출을 효과적으로 수행한다. 이후 깊이 정제 모듈인 NLSPN⁷⁾을 적용하여 깊이 완성 성능을 향상하였다.

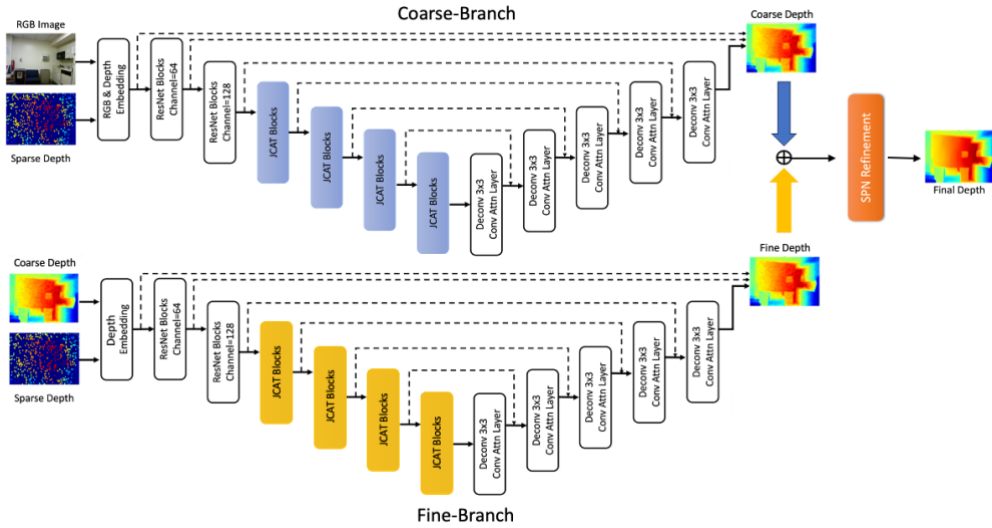


Fig. 1 Proposed Two-branch model architecture .

위에서 언급된 두 가지 문제점들을 동시에 해결하기 위해, 본 논문에서는 두 해결책을 적용한 네트워크를 제안한다. 최근 비전 트랜스포머를 이용하여 우수한 성능을 보여준 CompletionFormer 백본(backbone)을 베이스 모델로 삼아 two-branch 구조 기반 모델을 설계하였다. 기존에 single-branch로 구성된 베이스 모델은 RGB 이미지와 깊이 지도를 단순히 결합하여, 상호보완적인 두 센서 데이터들을 완전히 활용할 수 없다는 단점이 존재하였다. 이를 해결하고자, 재설계된 제안 모델에서는 색상에 집중하여 특징을 학습하는 coarse-branch와, 깊이에 집중하여 특징을 학습하는 fine-branch로 구성된 two-branch 백본을 설계한다. 제안 모델은 각각의 상보적인 브랜치를 통해 도출된 깊이 지도를 융합해줌으로써, 양식이 다른 두 데이터들을 완전히 이용한다. 더불어 긴 학습 과정 동안 특징이 손실되지 않도록 JCAT 블록 내 계층 간 잔차 연결을 추가하였다. NYUv2 데이터셋에 대한 실험 결과는 우리가 설계한 네트워크가 기존 모델 성능을 능가한다는 것을 보여준다.

2. 방법론

본 장에서는 제안하는 TB-CompletionFormer에 대하여 소개하고자 한다. Fig.1에서 알 수 있듯이 TB-CompletionFormer는 베이스 모델과 유사하지만,

기존 single-branch 모델의 단점을 보완한 two-branch 구조로 구성되어 있다.

2.1 Two-branch Model

Two-branch 구조의 모델은 각각의 브랜치에서 RGB이미지와 희소한 깊이 지도 정보를 잘 학습할 수 있도록 coarse-branch와 fine-branch로 구성된다. 기존 모델의 single-branch 구조는 희소 깊이 지도에서 객체 경계에 생기는 노이즈들을 줄일 수 없다는 한계가 존재하였다. 이러한 문제점을 해결하고자 우리는 두 브랜치에서 각각의 데이터 특성을 중심으로 학습을 진행한 후, 도출된 깊이 지도들을 융합하도록 모델을 설계한다. 이러한 구조를 통해 RGB이미지와 희소 깊이 지도의 특성을 완전히 이용하고, 서로의 장점들을 살려 정교한 특징 학습이 가능하도록 한다. coarse-branch와 fine-branch는 동일한 인코더-디코더 네트워크 구성으로 이루어져 있다.

먼저 coarse-branch는 RGB이미지와 희소한 깊이 지도를 입력으로 받아 색상을 중심으로 깊이 지도를 생성한다. 이는 RGB정보와 질감을 중점적으로 보기 때문에, 객체 경계 주변의 깊이를 효과적으로 학습할 수 있다는 장점이 존재한다. 더불어 고밀도의 이미지 내 구조 정보에 집중하므로 작은 물체에 대한 깊이 또한 잘 잡아낼 수 있다. 본 브랜치에서 사용된 백본은 CompletionFormer의 인코더-디코더 구조를 기반으로 설계된다. RGB이미지와 깊이 정보 임베딩 과정에서는 초기 단계(early stage)에서 카

메라와 라이다의 정보를 융합한다. 입력으로 받은 두 데이터를 인코딩하기 위해, 두 개의 개별 컨볼루션을 사용한다. 이후 도출된 각 특징을 결합하여 다음 네트워크에 넘겨준다. CNN과 트랜스포머 네트워크를 통합한 Joint Convolutional Attention and Transformer(JCAT) 블록에서는 효율적으로 깊이 정보들을 학습이 진행된다. 많은 파라미터와 연산량을 가진 트랜스포머의 오버헤드를 줄이기 위해, 병렬적으로 단일 경로를 기반으로 구성하였다. JCAT 모듈은 트랜스포머 레이어와 convolutional attention 레이어로 이루어진다. 트랜스포머 레이어는 멀티 헤드(multi-head) 메커니즘과 feed-forward layer(FNN)을 갖는 Spatial Reduction Attention(SRA)를 포함한 Pyramid Vision Transformer(PVT)⁸⁾ 모델이 사용된다. convolutional attention 레이어는 채널과 공간으로 나뉘 학습하는 CBAM⁹⁾ 모델을 사용한다. 이러한 JCAT 블록을 백본으로 사용함으로써, 네트워크의 수용 영역(receptive field)를 전역적으로 확장할 수 있다. 또한 RGB이미지와 깊이 정보가 하나의 토큰으로 포함되어 있기 때문에, self-attention 메커니즘은 전체 이미지 뿐만 아니라 깊이 별로 각 픽셀의 유사성을 비교하여 학습한다. 이는 유효한 깊이 값들이 이미지의 전역적인 공간에 영향을 줄 수 있다. 또한 본 논문에서 사용된 인코더-디코더 네트워크는 5단계로 구성되어, 서로 다른 스케일의 특징들 간 효과적으로 영향을 받을 수 있다.

fine-branch에서는 coarse-branch의 결과인 coarse depth map과 최소 깊이 지도를 입력으로 받아 깊이를 중심으로 깊이 지도를 예측하는 것을 목표로 한다. 이전 브랜치에서 생성된 깊이 지도를 가이드 삼아 더 정교하게 학습함으로써 기존보다 향상된 성능을 얻을 수 있다. 또한 본 브랜치에서는 색상을 중심으로 학습된 깊이 지도를 기반으로 하여, 기존 모델보다 깊이 정제 과정을 한번 더 거치기 때문에 훨씬 정교한 깊이 완성을 수행한다. 구체적인 과정은 위 브랜치와 유사하게 두 깊이 정보를 결합하여 임베딩한 후 다음 네트워크에 넘겨준다. 이후 인코더-디코더 과정은 coarse-branch와 동일하며, 마지막 디코더의 결과로 도출된 fine depth map과 앞선 coarse-branch 브랜치의 coarse depth map과 융합한다. 목적이 다른 각 브

랜치들을 통해 도출된 깊이 지도들은 서로 상보적인 특징을 가지고 있다. 이러한 특징을 완전히 이용하여 융합해줌으로써, 두 센서들 간 시너지 효과를 최대한으로 발휘하며 깊이 완성 결과를 얻을 수 있다.

Two-branch에 거쳐 긴 과정 동안 깊이 정보 특징들을 학습하면서, 예측된 깊이 지도의 픽셀에 누락 값이 발생하는 문제점이 발생한다. 이러한 누락 값을 채워 넣기 위해, 유효한 깊이 값들을 공간적으로 전파시켜 깊이를 정제하는 모듈인 Spatial Propagation Network(SPN)이 제안되었다. 기존 모델과 동일하게 본 논문에서도 깊이 정제 모듈인 NLSPN¹⁰⁾ 모델을 후처리 과정에 사용한다.

2.2 Residual Connection

ResNet¹¹⁾에 처음 도입된 잔차 연결(residual connection)은 입력 데이터에 모델이 예측한 특징을 추가해주는 방법이다. 이를 기반으로 네트워크의 가중치 소실 문제를 해결하며, 효과적인 성능 개선을 입증하였다. 본 모델에서도 Fig.2와 같이 백본의 JCAT 블록 내 잔차 연결을 추가하여, two-branch의 긴 학습 과정 중에도 특징 정보가 손실되지 않도록 설계되었다. Channel attention과 Spatial attention은 중요한 특징을 강조하고 지역적인 정보를 통합하는 도움을 주지만, 입력의 특징을 완전히 무시하지 않도록 잔차 연결을 통해 모델이 풍부한 정보를 활용하여 학습할 수 있다.

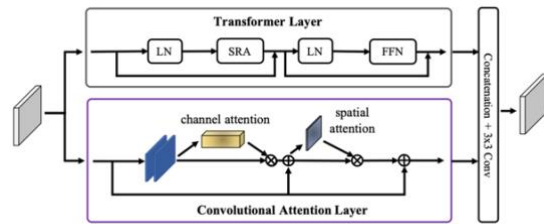


Fig. 2 Modified JCAT Block with Residual Connection.

3. 실험

3.1 데이터셋

제안 모델을 검증하기 위해 본 논문에서는 NYUv2¹²⁾ 데이터셋을 사용하였다. NYUv2 데이터셋은 마이크로소프트 Kinect 카메라를 기반으로 실내 장면에서 촬영한 464개의 RGB이미지와 깊이 이미지로 구성된다. 이전의 깊이 완성 모델들의 실험 환경과 유사하게 본 연구에서는 균일하게 샘플링된 50,000개의 이미지에 대해 학습을 진행한다. 이후 평가를 위해 654개의 레이블이 지정된 테스트셋을 기반으로 모델을 테스트한다. 학습 및 테스트 셋 모두에서 640x480 크기의 원본 프레임을 이중 선형 보간(Bilinear interpolation)을 이용하여 절반만 다운 샘플링(down-sampled) 한 후, 304x228 크기로 가운데를 잘라(center-crop) 사용한다.

3.2 평가 지표

NYUv2 데이터셋을 기반으로 깊이 완성 모델을 평가하기 위해 다음과 같은 두 가지 표준 지표를 사용한다.

- Root mean squared error (RMSE)

$$\sqrt{\frac{1}{|V|} \sum_{v \in V} |d_v^{gt} - d_v^{pred}|^2}$$

- Mean absolute relative error (REL)

$$\frac{1}{|V|} \sum_{v \in V} |(d_v^{gt} - d_v^{pred}) / d_v^{gt}|$$

특히 RMSE는 실제 값(ground truth)과 예측한 깊이 지도를 픽셀 별로 깊이 값을 비교하여 mm단위로 오차를 측정하므로 깊이 정확도와 직접적인 상관 관계가 있다. 이에 따라 KITTI Depth Completion 벤치마크에서도 RMSE 지표를 기준으로 모델의 성능 순위를 매긴다. 본 논문에서도 전체적으로 많이 사용하는 두 평가 지표를 이용하여 모델의 성능을 평가한다.

3.3 실험 결과

본 논문에서는 NYUv2 데이터셋에서 제안 모델을 평가하였다. 제안 모델은 종단간(end-to-end)으로

학습이 진행되며, 테스트셋을 기준으로 다른 모델들과 성능을 비교하였다.

Table 1은 NYUv2 리더보드를 기준으로 제안 모델과 다른 깊이 완성 모델들과의 정량적 비교를 나열한다. 본 연구에서 개발된 모델은 REL 평가 지표를 기준으로 다른 모든 모델을 능가하는 성능을 보여준다. 더불어 벤치마크의 주 평가 지표인 RMSE 또한 SOTA 모델과 동일한 결과를 나타낸다.

Table 1 Quantitative evaluation on NYUv2 dataset.

	RMSE (m)	REL (m)
CSPN++	0.117	0.016
DeepLiDAR	0.115	0.022
TWIS ¹³⁾	0.097	0.013
NLSPN	0.092	0.012
RigNet ¹⁴⁾	0.090	0.012
CompletionFormer	0.090	0.012
DYSPN ¹⁵⁾	0.090	0.012
BEV@DC ¹⁶⁾	0.089	0.012
Ours	0.089	0.011

특히 Table 2와 같이 베이스 모델인 CompletionFormer와 비교했을 때 RMSE 기준 17mm, REL 기준 5.4mm 이 개선됨을 확인할 수 있다. 즉 모든 평가 기준에서 전반적으로 더 향상된 성능을 보여주며 two-branch 백본의 우수함을 입증한다.

Table 2 Comparison of baseline and proposed model.

	RMSE (mm)	REL (mm)
CompletionFormer	907.1	121.3
Ours	890.1	115.9

Fig.3은 NYUv2 데이터셋의 (a)RGB 이미지와 (b)최소한 깊이 지도를 입력으로 받아 깊이 완성을 수행한 결과를 나타낸다. 결과 이미지는 깊이 정도를 수치화하여, 거리가 멀수록 빨간색, 가까울수록 파란색으로 시각화된다. 본 논문에서 제안한 기법과 비교하기 위해, 기준 모델인 CompletionFormer를 선정하였다. 두 모델 모두 전체적으로 좋은 성능을 보여주지만, 연두색 직사각형으로 표시된 상세 디테일에서 차이를 확인할 수 있다.

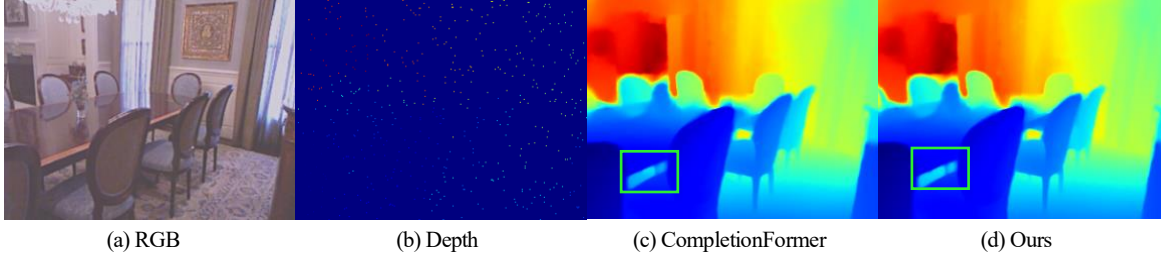


Fig. 3 Depth Completion results on the NYUv2 dataset

구체적인 차이를 확인하기 위해, Fig.4와 같이 결과 이미지 내 연두색 직사각형 부분을 확대하여 실제 값(Ground Truth)을 기준으로 정성적 비교를 진행한다. 베이스 모델인 CompletionFormer(b)의 경우 실제 값(a)보다 객체 주변에서 픽셀들이 뭉개지는 것을 확인하였다. 반면 제안 모델(c)은 기존 베이스 모델 대비 훨씬 픽셀들이 선명하며, 객체 주변에서도 노이즈들이 비교적 줄어든 것을 확인할 수 있다. 본 연구에서 제안한 two-branch 기반 백본을 통해, 카메라와 라이다 각각의 특성을 완전히 활용함으로써 정성, 정량적인 결과에서 모두 향상된 성능을 입증한다.

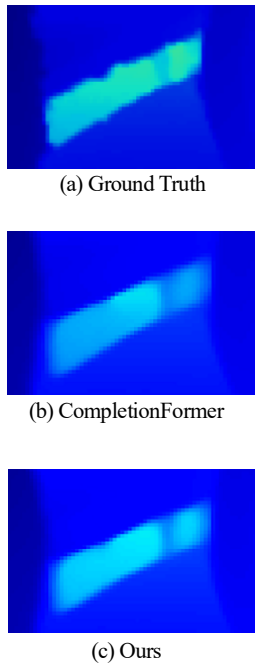


Fig. 4 Qualitative comparison with basemodel

4. 결론

본 연구에서는 기존의 CompletionFormer 모델을 two-branch 구조로 변경하여 효과적으로 깊이 완성을 수행하는 네트워크인 TB-CompletionFormer를 제안한다. single-branch로 구성된 기존 모델은 양식이 다른 두 센서 간 데이터들이 융합되는데 한계가 존재하였다. 이 문제점을 해결하기 위해 우리의 모델은 먼저 RGB이미지와 희소한 깊이 지도를 입력으로 받아 coarse한 깊이 정보를 가진 깊이 지도를 추출한다. 이후 coarse한 깊이 지도와 이전에 입력으로 제공한 희소 깊이 지도를 두 번째 브랜치로 전달하여, 더 정밀하고 안정적으로 깊이 완성을 수행한다. 이러한 모델 구조를 통해 양식이 다른 두 센서 데이터들이 효과적으로 융합할 수 있다. 더불어 베이스 모델에서 제안한 JCAT 블록 내 잔차 연결을 추가하여, two-branch의 긴 학습 과정 중에도 특징 정보가 손실되지 않도록 설계되었다. 결과적으로 본 논문에서 제안한 모델은 기존에 비해 모든 평가 지표에서 전반적으로 성능이 향상되었으며, SOTA 모델인 BEV@DC와 비교하여 유사하거나 더 우수한 성능을 보여준다.

다만 본 연구의 한계로는 비전 트랜스포머를 기반으로 한 모델 구조이기 때문에 많이 메모리와 연산량을 필요로 한다. 베이스 모델 대비 깊이 완성 성능을 향상시켰지만, 추론 속도는 기존의 10fps 대비 6fps로 저하되었다. 향후 연구 방향으로는 개선된 성능은 유지하면서, 더 빠른 추론 속도를 달성할 수 있는 효율적인 깊이 완성 모델을 개발하고자 한다.

5. 사 사

본 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2022-00166693).

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성사업의 연구 결과로 수행되었음(IITP-2023-RS-2022-00156287).

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음(IITP-2023-RS-2023-00256629).

References

- 1) F.Ma and S.Karaman. "Sparse-to-Dense: Depth Prediction from Sparse Depth Samples and a Single Image." in ICRA, 2018
- 2) M.Hu, S.Wang, B.Li, S.Ning, L.Fan and X.Gong. "PENet: Towards Precise and Efficient Image Guided Depth Completion." in ICRA, 2021
- 3) W.V.Gansbeke, D.Neven, B.D.Brabandere, and L.V.Gool. "Sparse and noisy lidar completion with rgb guidance and uncertainly." in MAV, 2019
- 4) J.Qiu, Z.Cui, Y.Zhang, X.Zhang, S.Liu, B.Zeng and M.pollefeys. "DeepLiDAR: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image." in CVPR, 2019
- 5) X.Cheng, P.Wnag, C.Guan and R.Yang. "CSPN++: Learning context and resource aware convolutional spatial propagation networks for depth completion." in AAAI, 2020
- 6) Rho.K, Ha.J and Kim.Y. "GuideFormer: Transformers for image guided depth completion" in CVPR, 2022
- 7) Zhang.H, Guo.X, Poggi.M, Zhu.Z and Mattoccia "CompletionFormer: Depth Completion with convolutions and vision transformers" in CVPR, 2023
- 8) W.Wang, E.Xie, X.Li, D.Fan, K.Song, D.Liang, T.Lu, P.Luo and L.Shao. "Pyramid Vision Transformer: A versatile backbone for dense prediction without convolutions." in ICCV, 2021
- 9) S.Woo, J.Park, J.Lee and I.Kweon. "Convolutonal block attention module." in ECCV, 2018
- 10) J.Park, K.Joo, Z.Hu, C.Liu and I.Kweon. "Non-local spatial propagation network for depth completion." in ECCV, 2020
- 11) K.He, X.Zhang, S.Ren and J.Sun."Identity mappings in deep residual networks." in ECCV, 2016
- 12) N.Silberman, D.Hoiem, P.Kohli and R.Fergus. "Indoor segmentation and support inference from rgbd images." in ECCV, 2012
- 13) S.Imran, X.Liu and D.Morris. "Depth completion with twin surface extrapolation at occlusion boundaries." in CVPR, 2021
- 14) Z.Yan, K.Wang, X.Li, Z.Zhang, B.Xu, J.Li and J.Yang. "RigNet: Repetitive image guided network for depth completopn." in ECCV, 2022
- 15) Y.Lin, T.Cheng, Q.Zhong, W.Zhou and H.Yang. "Dynamic spatial propagation network for depth completion." in AAAI, 2022
- 16) Zhou, W.Yan, X.Liao, Y.Lin, Y.Huang, J.Zhao, G.Cui and S.Li. "Bev@dc: Bird's eye view assisted training for depth completion." in CVPR, 2023