2023 졸업프로젝트 진행 보고서

주제: 3D Object Detection

2023. 05. 24. ~ 2023. 06. 22.

신상윤, 윤성우

공부 내용

* SECOND 논문 요약
* SECOND에서 이룬 Voxel Net으로부터의 개선점

1. Introduction

* 현재(2018) 3D detector는 image data와 point cloud data를 모두 활용하는 융합 방법을 사용하고 있다.
* Voxel Net의 computational cost(3D convolution) 때문에 real time application에서 활용하기 어렵다.
* Voxel Net의 단점 1: slow inference speed (느린 추론 속도)
* Voxel Net에서는 ground truth와 prediction의 예측 방향차이가 pi일 때 큰 손실이 발생한다.
* Voxel Net의 단점 2: low orientation estimation performance (낮은 방향 추정성능)

3. Second detector

3.1 network architecture

**3.1.1 point cloud grouping**

Voxel Net와 방법은 같고 수치만 조정

차량감지에는 T = 35, 보행자와 자전거는 T = 45로 설정

**3.1.2 Voxelwise Feature Extractor**

Voxelwise features를 추출하기 위해 voxel feature encoding (VFE)

layer 사용 최종 Voxelwise feature는 sparse 4D tensor로 나타난다.

크기는 C \* D' \* H' \* W'

**3.1.3 sparse convolutional middle extractor (핵심 내용)**

- GEMM 계산

일반적인 딥러닝 연산 out = in \* W + b은 CNN에서 convolution 계산을 할 때 병렬처리가 힘들다. (7중 for문 필요)

따라서 im2col 방법을 이용하여 convolution 연산을 행렬 2개의 곱으로 나타낸다.

일반화된 식은 C = alpha AB + beta C이다.

*(im2col 참고 reference: https://amber-chaeeunk.tistory.com/31)*

1) 일반적인 2d dense convolution algorithm

2) GEMM-based algorithm

3) 2)에서 sparse data를 적용

4) 그중 0아닌 data만 적용

이때 Rk,j를 Rule이라고 한다. (k: kernel offset j : output index)

input-output index rule matrix를 통하여 sparse한 original data에서 바로 data를 만든다.

- Sparse Convolutional Middle Extractor

z축에 대한 정보를 학습하고 희소한 3D 데이터를 2D BEV(조감도) 이미지로 변환하는 데 사용

(논문 참고 🡪 노란색 sparse convolution, 흰색 submanifold convolution, 빨간색 sparse to dense layer)

각 단계에는 z축에서 다운샘플링을 수행하기 위해 여러 개의 하위 매니폴드 컨볼루션 레이어와 하나의 일반 희소 컨볼루션이 포함된다.

z 차원이 하나 또는 두 개로 다운샘플링된 후 희소 데이터가 조밀한 피쳐 맵으로 변환된다.

그런 다음 데이터를 이미지와 같은 2D 데이터로 재구성한다.

**3.1.4 RPN**

Voxel Net과 같음

**3.1.5 Anchors and Targets**

검출되는 물체는 대략적으로 고정된 크기를 가지기 때문에, KITTI training set에서 0도와 90도 회전된 GT의 중심 위치와 크기의 평균을 기반으로 고정된 크기의 anchor를 사용

각각의 anchor는 one-hot vector of classification target과 7-vector box regression target, one-hot vector of direction classification target에 할당된다.

서로 다른 Class는 서로 다른 threshold 사용

3.2 Training and Inference

**3.2.1 Loss**

Voxel Net에서 방향이 0과 π로 예측된 경우 같은 Box를 예측하지만 Angle regression에서 큰 Loss가 발생하는 문제가 있음

따라서 새로운 Angle loss regression을 제안하며 두가지 이점이 있다.

(1) 0과 π의 방향 사이의 문제를 해결하고, (2) 각도 오프셋 함수에 대해 IoU를 자연스럽게 모델링한다.

반대 방향의 상자를 동일한 것으로 취급하는 문제를 해결하기 위해 RPN의 output에 간단한 direction classifier를 추가하였다.

classifier를 학습시키기 위해서 direction classifier target을 만들어야 한다.

Ground Truth의 z 축 기준 yaw의 회전이 0보다 크다면 positive, 그렇지 않다면 negative로 target을 생성하여 학습을 진행한다.

Classification은 Focal Loss를 사용한다. 생성되는 anchor들이 많은데 대부분 negative anchor이기 때문이다.

최종 Loss function에서 네트워크가 물체의 방향을 인식하기 어려운 경우를 대비해 β3를 상대적으로 작은 값으로 사용한다.

Conclusion

Car와는 달리 Pedestrian, Cyclist detection의 경우 더 많은 false positive와 false negative 결과 생성

image에서의 instance density가 Car보다 높고 각 instance 마다 더 적은 point를 가지고 있다.

Pedestrian과 Cyclist는 적은 부피를 가지고 있기 때문에 CNN이 가지는 효과가 한계적으로 적용된다.