레이더 포인트 클라우드 밀도에 따른 FPGA 동적 부분 재구성 기반 적응형 PointNet 가속기 설계

유주하, 이성주*
세종대학교 반도체시스템공학과 및 지능형드론융합전공 *세종대학교 AI융합전자공학과 및 지능형드론융합전공

juha@itsoc.sejong.ac.kr, *seongjoo@sejong.ac.kr

Design of FPGA Dynamic Partial Reconfiguration-based Adaptive PointNet Accelerator according to Radar Point Cloud Density

Juha Yoo, Seongjoo Lee*
Dept. of Semiconductor Systems Engineering and Convergence Engineering for Intelligent Drone, Sejong Univ., *Dept. of AI Convergence Electronic Engineering and Convergence Engineering for Intelligent Drone, Sejong Univ.

요 약

본 논문에서는 레이더 기반 3차원 포인트 클라우드(PCD)의 밀도 변화에 따라 최적의 신경망 모델을 동적으로 선택하는 DPR(Dynamic Partial Reconfiguration) 기반 처리 구조를 제안한다. 제안 구조는 PointNet을 응용한 SharedMLP 신경망을 프레임별 포인트 수에 따라 Low, Medium, High로 나누고, 실험적으로 도출한 N_1 /N_2 임계값을 통해 모델을 선택한다. PyTorch 기반 AutoEncoder 실험 결과, 평균 지연 시간이 30% 감소하였으며 MSE는 수용 가능한 수준으로 유지되었다. Vitis HLS 기반 하드웨어 분석에서는 모델 규모가 증가할수록 DSP, LUT 자원 사용량과 Iteration Latency가 선형적으로 증가하였으며, DPR 구조의 자원 효율성을 입증하였다.

키워드: 포인트 클라우드, 동적 부분 재구성, PointNet, SharedMLP, 자원 효율성, 고수준 합성

Ⅰ. 서론

자율주행 및 무인 시스템의 발전에 따라 레이더 기반 3차원 PCD를 활용한 객체 탐지가 주목받고 있다. 하지만 기존 시스템은 다양한 밀도의 데이터를 단일 규모의 신경망으로 처리해 자원 낭비·속도 저하 문제가 있다. 이를 해결하기 위해 본 논문은 프레임별 포인트 수에 따라 세 가지 규모(Low, Medium, High)의 PointNet 기반 Shared MLP 모델을 동적으로 선택하는 방식을 제안한다. 해당 구조는 Vitis HLS를 기반으로 FPGA에서의 Partial Reconfiguration(DPR)을 고려해 설계되었으며, 자원 사용량과 성능은 Vitis HLS 분석을 통해 평가하였다.

Ⅱ. 본론

본 연구에서는 레이더 기반 3차원 PCD를 프레임별 포인트 밀도에 따라 스케일을 조절하는 DPR(Partial Reconfiguration) 기반 신경망 모델을 구현하였다.

A. 데이터 전처리 및 포인트 밀도 측정

본 논문에서는 자율주행 데이터셋인 nuScenes의 레이더 데이터를 활용하여 5개 채널을 퓨전한 후, Open3D 기반의 통계적 이상치 제거(SOR: Statistical Outlier Removal)를 통해 3차원 PCD를 전처리하였다. SOR은 각 포인트 pi의 위치에서 주변 k개 포인트(본 연구에서 k = 20)와의 평균 거리 di를 계산하고, 전체 평균 거리 μd 및 표준편차 σd를 기준으로 di가 μd +2 σd를 초과하는 포인트를 이상치로 간주하여 제거하는 방식이다. 해당 과정은 아래 수식으로 요약된다.

$$d_i = \frac{1}{k}\sum_{j=1}^{k}\|p_i - p_j\|,\ \ \mu_d = \frac{1}{N}\sum_{i=1}^{N} d_i,\ \ \sigma_d = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \mu_d)^2} \quad (1)$$

B. 스케일별 SharedMLP 네트워크 설계

데이터 밀도의 변화에 따라 최적의 연산 효율을 달성할 수 있도록 세 가지 규모(Low, Medium, High)의 SharedMLP 네트워크를 설계하였다. 전처리된 데이터셋을, Bottom-up 방식으로 히든 유닛 수를 16부터 1024까지 증가시키며 Autoencoder의 재구성 성능(MSE)을 분석하였다. 그 결과, 구간별 성능 향상을 기준으로 Low(3→32→64), Medium(3→64→128), High(3→128→256) 모델 구조를 결정하였다.

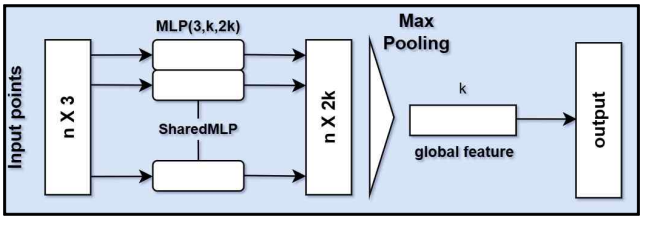

그림 1. 제안하는 SharedMLP 기반 스케일 가변 신경망 구조
(※ Low/Medium/High 모델은 각각 k=32, 64, 128로 설정됨)

C. 스케일별 SharedMLP 네트워크 설계

프레임별 포인트 수에 따라 적절한 모델을 선택하기 위해 두 개의 임계값 N_1 (Low↔Medium), N_2 (Medium↔High)를 설정하였다. 이를 위해 다양한 N_1 /N_2 조합에 대해 AutoEncoder 기반 복원 연산을 수행하고, 평균 MSE와 지연 시간(Latency)을 측정하였다. 정적 High 모델과 비교하여 아래의 스코어 식을 기반으로 최적 임계값을 탐색하였다.

* 교신저자 : 이성주

$$Score = \alpha \cdot (Static\_Latency - DPR\_Latency) - \beta \cdot (DPR\_MSE - Static\_MSE) \quad (2)$$

그 결과, N_1 = 134, N_2 = 197이 최적 조합으로 선정되었으며, 해당 기준으로 각 프레임에서 동적 모델 선택이 이루어졌다.

D. 소프트웨어 단계 DPR 성능 평가

하드웨어 구현에 앞서, 소프트웨어 수준에서 DPR 방식의 성능 효율성을 검증하였다. DPR과 고정 모델 방식(STATIC_HIGH)을 비교하여 평균 MSE와 처리 지연(Latency)을 측정하였다. 여기서 프레임별 포인트 수에 따라 모델을 동적으로 선택한 DPR은 평균 Latency를 약 30% 줄였으며, STATIC_HIGH 모델에 비해 MSE는 소폭 증가하였지만, 수용 가능한 수준이었다.

표 1. DPR vs Static-High 전체 성능 비교

Mode	Avg MSE	Avg Latency(ms)
DPR	0.0471	0.14
Static-High	0.0140	0.20

또한 DPR 방식에서 각 모델의 사용 비율 및 성능은 표 2와 같다. 이로써 DPR 구조가 소프트웨어 수준에서도 자원-성능 간 트레이드오프를 유연하게 달성할 수 있음을 확인하였다.

표 2. 구간별 성능 및 사용량

Model	Frames	Resource occupation rate (%)	Avg MSE	Avg Latency (ms)
Low	42	10.4%	0.0956	0.14
Medium	173	42.8%	0.0739	0.17
High	189	46.8%	0.0118	0.24

Ⅲ. 실험

A. 실험 환경

본 연구는 Python 기반 전처리와 PyTorch 기반 모델 학습·성능 평가를 수행하고, 하드웨어 분석은 Xilinx Vitis HLS 2021.2를 사용하였다. 입력 데이터는 nuScenes의 레이더 PCD이며, Open3D를 이용해 통계적 이상치 제거 및 필터링 기반 전처리를 수행하였다. 이후 DPR 방식을 적용하여 Low, Medium, High 모델을 동적으로 선택하도록 설계하였다. 모델 학습은 NVIDIA RTX GPU에서, HLS 합성은 Zynq-7000 SoC를 대상으로 진행하였다.

B. HLS 기반 SharedMLP 모델의 자원 사용량 및 성능 분석

본 실험에서는 제안한 Low, Medium, High 규모의 SharedMLP 모델에 대해 Vitis HLS 2021.2를 사용하여 고수준 합성을 수행하였다. Python 기반 구현을 C++ 소스 및 테스트벤치 코드로 변환한 뒤, Simulation과 Synthesis 단계를 통해 BRAM, DSP, FF, LUT 사용량과 Iteration Latency를 측정하였으며, BRAM은 비교 일관성을 위해 모두 4로 고정하였다. 합성 결과는 표 3에 정리하였다.

표 3. 모델별 Vitis HLS 자원 사용량 및 연산 지연 시간

Model	BRAM	DSP	FF	LUT	Iteration Latency
Low	4	160	16,649	33,151	313
Medium	4	320	32,222	90,832	585
High	4	640	63,363	269,236	1129

분석 결과, 모델 규모가 커질수록 연산 자원(DSP, FF, LUT)과 지연 시간이 선형적으로 증가하였으며, 이는 연산 계층의 복잡도 차이에서 비롯되었다. 이러한 결과는 입력 밀도에 따라 경량 모델을 선택하는 DPR 구조가 자원 효율성 측면에서 효과적임을 시사한다. 또한, HLS 분석만으로도 제안한 구조가 정적 방식 대비 유연한 자원-성능 트레이드오프를 달성할 수 있음을 확인하였다.

Ⅳ. 결론

본 논문은 레이더 PCD의 밀도 변화에 대응해 세 가지 규모의 SharedMLP 모델을 동적으로 선택하는 DPR 기반 처리 구조를 제안하였다. 실험 결과, DPR 방식은 정적 모델 대비 평균 지연 시간을 약 30% 단축하면서도 성능(MSE)을 실용 수준으로 유지하였다. 또한 Vitis HLS 기반 분석을 통해 모델 규모에 따른 자원 사용량과 연산 지연 증가를 확인하였으며, DPR 방식의 자원-성능 효율성을 입증하였다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원, 한국연구재단의 지원(No. 2023R1A2C1006340) 및 정부(교육부)의 재원으로 한국연구재단의 이공분야 대학중점연구소지원사업의 지원(No. 2020R1A6A1A03038540)을 받아 수행된 연구이며 검증을 위한 EDA관련 툴은 IDEC의 지원을 받았음.

참 고 문 헌

[1] J. Meng, L. Zou, and S. Choi, "Development of 3D Object Detection Algorithm Based on PointNet," in Proc. Korea Information and Communications Society Summer Conf. 2019, pp. 1083-1084, Jeju Island, Korea, June 2019.

[2] C. Ding, L. Zhang, H. Chen, H. Hong, X. Zhu, and F. Fioranelli, "Sparsity-Based Human Activity Recognition With PointNet Using a Portable FMCW Radar," IEEE Internet Things J., vol. 10, no. 11, pp. 10024-10037, June 2023.

[3] H. Wang, W. Li, D. Li, and Z. Guo, "An Improved PointLSTM Gesture Recognition Method Based on Millimeter-Wave Radar 3D Point Cloud," in Proc. Int. Conf. Commun., Image Signal Process. (CCISP) 2024, pp. 72-77, Gold Coast, Australia, Nov. 2024.

[4] N. Charaf, A. Kamaleldin, M. Thümmler, and D. Göhringer, "RV-CAP: Enabling Dynamic Partial Reconfiguration for FPGA-Based RISC-V System-on-Chip," in Proc. IEEE Int. Parallel Distrib. Process. Symp. Workshops (IPDPSW) 2021, pp. 172-179, Portland, OR, USA, May 2021.

[5] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A Multimodal Dataset for Autonomous Driving," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) 2020, pp. 11618-11628, June 2020. [Online]. Available: https://registry.opendata.aws/motional-nuscenes, accessed May 7, 2025