아주대학교
세상의 A+가 되겠습니다

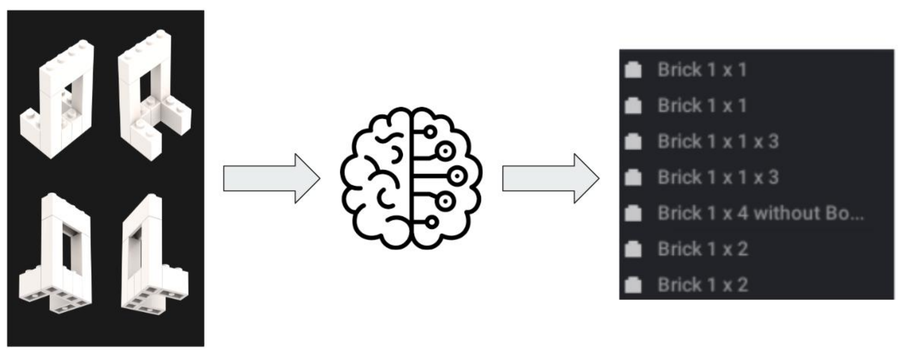
2023-2
AJOU SOFTCON

Lego-Disassembler : MLM 메커니즘을 이용한 3D 입체 구조 학습

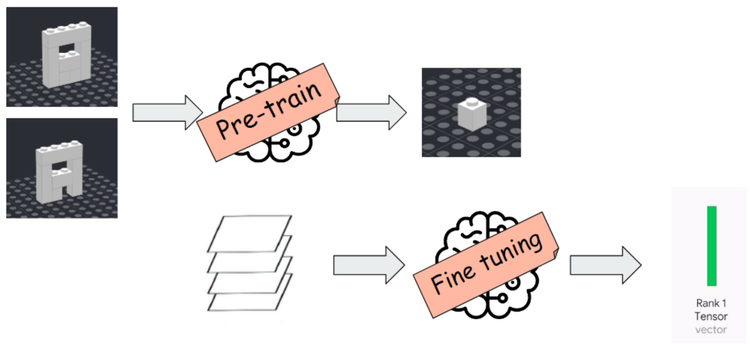
팀 명 LEGOND **팀 원** 성민석, 양태규, 이선우 **지도교수** 유종빈 교수님 **멘 토**

개발 동기 및 목적

- 인공신경망이 평면적인 이미지를 넘어 입체적인 구조를 이해할 수 있는지 확인하고자 함
- 현실의 이미지는 데이터의 비용과 문제의 복잡도가 높음
  → 단순화된 Lego 블록을 사용
- Lego의 입체 구조를 이해 → Lego 구조물의 구성을 이해
- 입력된 Lego 구조물 이미지들을 기반하여 Lego 구조물을 구성하는 블록의 종류와 개수를 예측하는 문제를 통해 인공신경망이 입체 구조를 이해할 수 있는지 실험

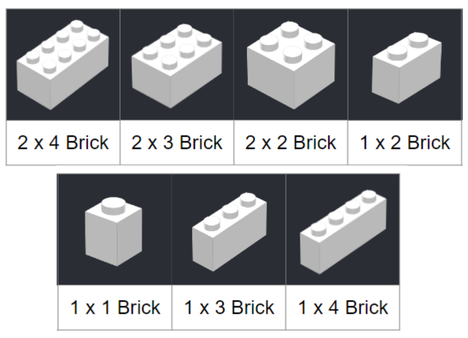


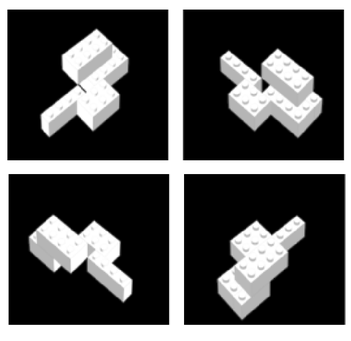
- 자연어처리에서 널리 이용되는 BERT 모델은 MLM, NSP를 통해 높은 문맥에 대한 이해를 보여줌
  → 이미지에 대해서도 MLM (Masked Language Model)을 적용하여 성능을 향상시켜 보려고 함



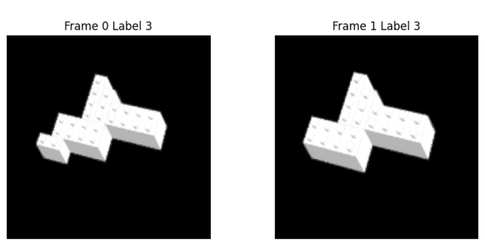
개발 내용

- 모델 입력 데이터
  - 7종류의 블록 5~10개로 구성된 Lego 구조물
  - Stud.io 프로그램을 통해 모델링 → 4가지 각도 이미지 캡처





- Pretrained 데이터
  - 5~6개로 구성된 Lego 구조물
  - 원본 이미지와 1개의 블록을 제거한 이미지로 구성



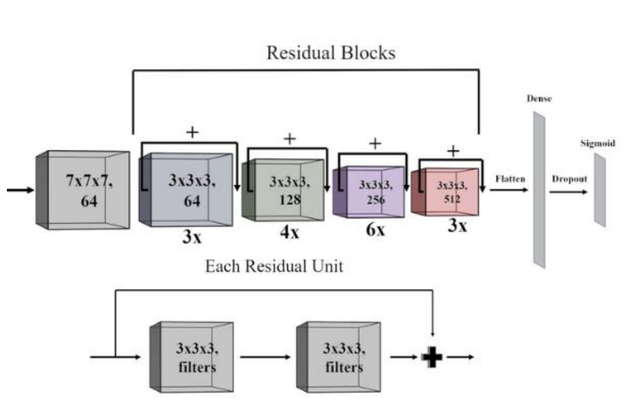
- 모델 출력
  - Pretrained모델: 원본과 비교하여 제거된 블록에 대한 예측
  - 일반 모델: Lego 구조물을 구성하는 블록의 종류별 개수

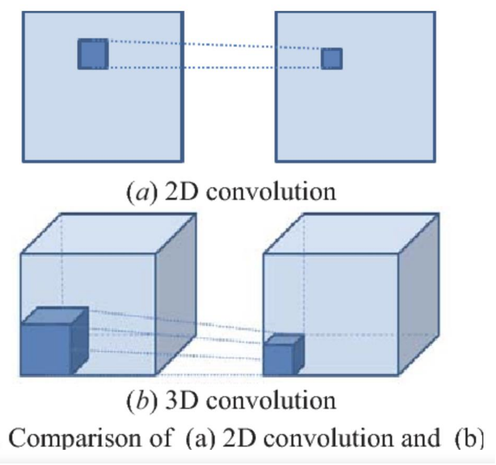
오픈소스 URL

- https://github.com/kenshohara/3D-ResNets-PyTorch (ResNet-3D)
- https://github.com/Sung-Minsoek/Lego-Disassembler Lego-Disassembler)

주요기술

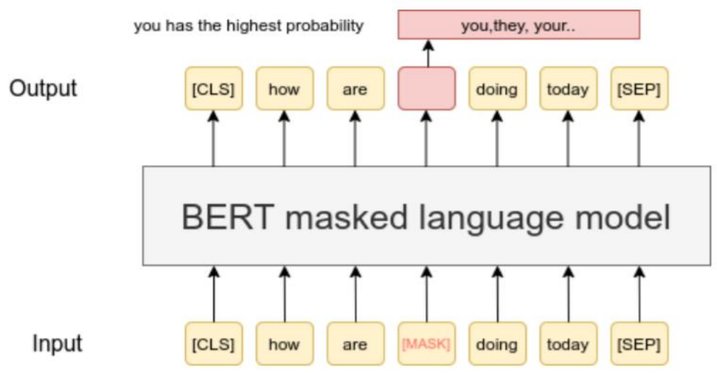
- 사용 모델: ResNet-3D
  - 4가지 각도로 캡처한 이미지를 하나의 텐서로 사용
    → 4 프레임의 영상 데이터
  - 비디오 데이터 기반의 ResNet 아키텍처
  - 3D convolution을 사용하여 프레임 간 관계성도 학습





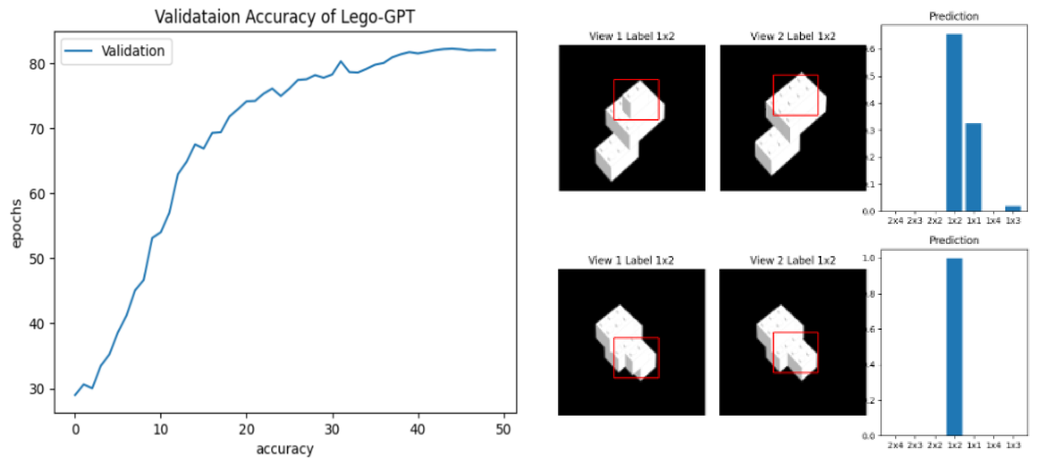
3. Comparison of (a) 2D convolution and (b)

- 성능 향상: MLM (Masked Language Model) 메커니즘
  - BERT에서 사용한 Pretrained 방식
  - 입력 sequence 중 하나의 토큰을 mask하고 모델이 mask된 토큰을 예측하며 모델 학습
  - 학습을 통해 sequence 전체에 대한 높은 이해도를 보임



결과 및 분석

- Pretrain 결과
  - Top accuracy: 82.20% (50 epoch)



- 일반 모델 결과
  - MSE loss로 성능 측정
  - Pretrain 없이 학습 시 학습이 되지 않고 loss 수렴(회색)
  - Pretrain 모델 기반으로 fine-tuning 시 최저 loss 달성 (0.5127, 청색)
  - MLM 메커니즘이 모델 성능을 향상시킴 → 이미지 데이터도 MLM 전략이 유효함