

CompressAI 를 활용한 객체 검출 네트워크 피쳐 맵 압축

도지훈, 이주영, 김연희, 최진수, 정세운

한국전자통신연구원

jhdo@etri.re.kr, leezy1003@etri.re.kr, kimyounhee@etri.re.kr, jschoi@etri.re.kr,

jsy@etri.re.kr

Object Detection Network Feature Map Compression using CompressAI

Jihoon Do, Jooyoung Lee, Younhee Kim, Jin Soo Choi, Se Yoon Jeong

Electronics and Telecommunications Research Institute

요 약

본 논문은 Detectron2 [1]에서 지원하는 객체 검출 임무 수행 네트워크의 과정 중에서 추출한 피쳐 맵을 신경망 기반으로 압축하는 방법을 제안한다. 이를 위해, 신경 망 기반 영상 압축을 지원하는 공개 소프트웨어인 CompressAI [2] 모델 중 하나인 bmshj2018-hyperprior 의 압축 네트워크를 활용하여 임무 수행 네트워크의 과정 중 스템 레이어(stem layer)에서 추출된 피쳐 맵을 압축하도록 학습시켰다. 또한, 압축 네트워크의 입력 피쳐 맵의 너비와 높이 크기가 64 의 배수가 되도록 객체 검출 네트워크의 입력 영상 보간 값을 조정하는 방법도 제안한다. 제안하는 신경망 기반 피쳐 맵 압축 방법은 피쳐 맵을 최근 표준이 완료된 차세대 압축 표준 방법인 VVC(Versatile Video Coding, [3])로 압축한 결과에 비해 큰 성능 향상을 보이고, VCM 앵커와 유사한 성능을 보인다.

1. 서론

최근 들어 사물인터넷, 스마트 시티, 자율 주행 차 등 다양한 응용 환경에서 수집되는 영상 데이터의 양은 기하급수적으로 증가하고 있으며, 이를 기반으로 기계가 영상의 객체나 이벤트를 인식하고 이를 분석하여 활용하는 서비스 요구가 증대되고 있다. 기존 영상 코덱의 경우, 사람의 인지 특성을 고려하여 설계되었기 때문에 기계의 임무 수행에 필요한 영상 정보가 아님에도 인간 시각 인지에 중요한 특성 정보는 유지하는 비효율성이 존재한다. 따라서, 기계에 임무 수행 성능을 유지하면서 데이터를 효율적으로 압축하기 위해서는 기계의 임무 수행에 초점을 맞춘 부호화 기술의 필요성이 대두되었다.

이에 따라, 기계의 임무 수행 성능을 유지하며 다중 임무

수행을 지원하는 압축 비트스트림 기술표준 제정을 목표로 ISO/IEC JTC1 SC29 WG11(MPEG)에서 2019년 7월 VCM(Video Coding for Machine) AHG(Ad-Hoc Group)이 결성되었고, 올해 1월 133차 MPEG 회의까지 각종 기술 및 제반 사항들에 대한 논의를 통해 CfE(Call for Evidence)를 발간하였다.

현재 MPEG VCM 의 파이프라인[4]은 부호화기 입력으로 영상 혹은 영상으로부터 추출된 피쳐 맵을 사용하는 2 가지 후보로 나뉜다. 추가로, 영상과 피쳐 맵 각각에 대해 표준이 완료된 VVC 와 같은 신호처리 기반 혹은 신경망 기반의 압축 및 복원을 통해 기계의 임무를 수행하는 그림 1 와 같이 4 가지 후보로 분류할 수 있다. 이 때, CfE 응답을 위한 평가기준(VCM 앵커)으로 입력 영상을 VVC 을 이용하여 압축 및 복원한 영상의 임무 수행 결과를 사용한다.

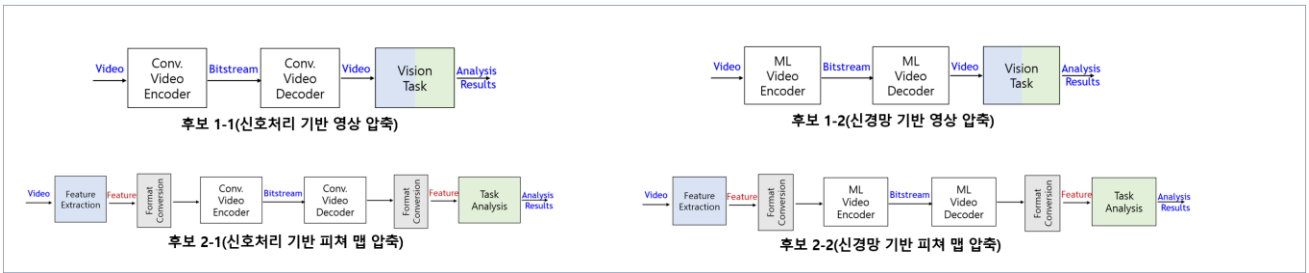


그림 1. VCM 후보 파이프라인

영상과 달리 높은 기계의 임무 수행 성능을 위해 추출된 피쳐 맵은 매우 많은 채널을 가지기 때문에 영상 대비 매우 큰 데이터 양을 갖고 자연 영상과는 다른 다양한 특성을 가진다. 이로 인해 피쳐 맵을 VVC 를 이용하여 압축한 경우, 그림 2 와 같이 회색 선으로 표시된 앵커와 비교하여 낮은 압축 성능을 보인다.

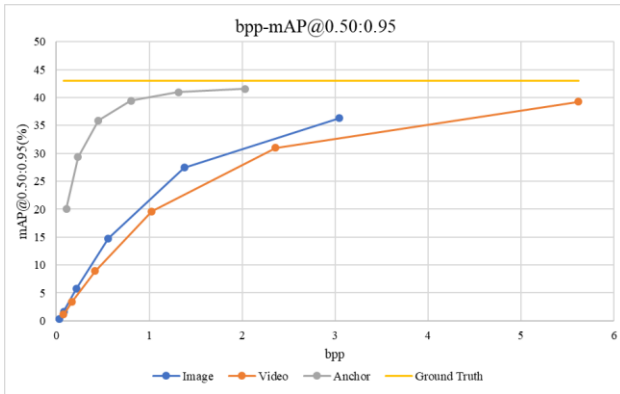


그림 2. 신호처리 기반 피쳐 맵 압축 성능[5]

본 논문에서는 보다 높은 압축 성능을 위해 객체 검출 임무 수행 네트워크 과정 중에 출력된 피쳐 맵을 신경망 기반으로 압축하는 방법을 제안한다. 추가로, 압축 성능을 높이기 위해 패딩 대신 압축 네트워크의 입력 단위의 배수가 되도록 객체 검출 네트워크의 영상 보간 값을 조정하는 방법을 제안한다[6].

본 논문의 구성은 다음과 같다. 2 절에서 본 논문에서 제안하는 신경망 기반 피쳐 맵 압축 방법을 설명하고, 3 절에서 제안 방법의 실험 결과를 얘기한다. 마지막으로 4 절에서 본 논문에 대한 결론을 맺는다.

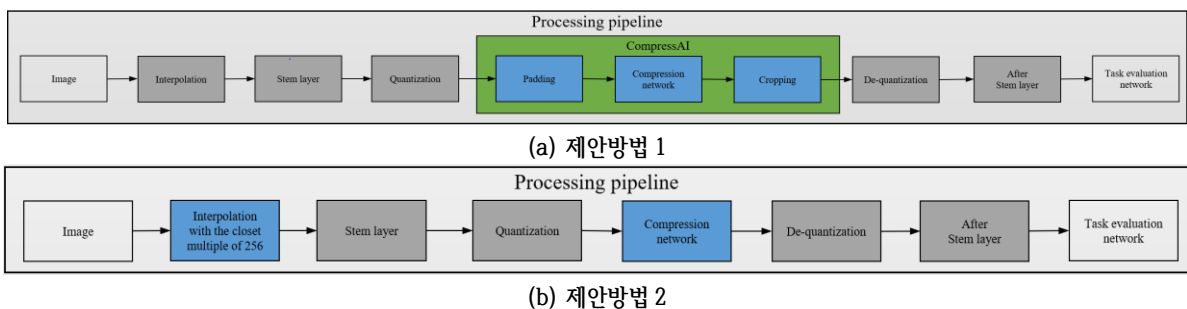


그림 3. 제안하는 신경망 기반 피쳐 맵 압축 방법

2. 제안하는 신경망 기반 피쳐 맵 압축 방법

제안하는 신경망 기반 피쳐 맵 압축 네트워크의 경우, 신경망 기반 영상 압축을 지원하는 공개 소프트웨어인 CompressAI 의 모델 중 하나인 bmshj2018-hyperprior 네트워크 구조를 사용한다. 또한, 3 채널로 구성되는 영상과 달리 객체 검출 네트워크 구조 중 스템 레이어를 통해 출력된 피쳐 맵은 64 채널을 가지기 때문에 압축 네트워크의 입력과 출력 채널을 64 로 변경하였다. 스템 레이어를 통해 출력된 피쳐 맵은 8-bit 균일 양자화 과정을 수행한 후 압축 네트워크의 입력으로 사용된다.

그림 3 과 같이 제안방법 1 은 객체 검출 네트워크를 고정시킨 상태에서 압축 네트워크의 입력 피쳐 맵의 너비와 높이를 64 의 배수가 되도록 패딩을 수행한 후 압축 네트워크를 통해 압축 및 복원된 피쳐 맵에 대해 크로핑을 수행한다.

제안방법 2 의 경우, 압축 네트워크에서 수행되는 패딩과 크로핑 대신 객체 검출 네트워크 과정에서 입력 영상에 대해 수행되는 보간 값을 조정하는 방법을 제안한다. 압축 네트워크의 입력 피쳐 맵의 경우, 보간 영상에 대해 너비와 높이가 2 번의 다운 샘플링이 수행되기 때문에, 피쳐 맵의 크기가 64 의 배수가 되도록 하기 위해서 객체 검출 네트워크의 영상 보간 과정에서 기존 보간 너비와 높이 값을 인접한 256 의 배수가 되도록 조정한다.

3. 실험 결과

제안하는 방법은 COCO2017 학습데이터셋 114000 장을 사용하여 여러 bpp(bit per pixel)에 대해 학습하였고, 학습된 모델 파라미터를 COCO2017 검증 데이터셋 5000 장에 대해 실험을 진행하였다.

그림 4는 VTM 8.2의 AI 부호화 모드에서 22, 27, 32, 37, 42, 47, 52의 QP(Quantization Parameter)로 부호화 한 앵커 결과[7]와 제안한 2 가지 방법의 실험 결과를 비교하여 나타낸다. 제안방법 1와 같이 입력 피쳐 맵에 대해 패딩을 수행하여 신경망 기반의 압축 네트워크를 통해 압축 및 복원하는 경우, 앵커에 비해 떨어지는 성능을 보이지만, 피쳐 맵을 VVC로 압축한 성능에 비해 나은 성능을 보인다. 반면에, 제안방법 2와 같이 기존의 영상 보간 방법을 피쳐 맵 압축 네트워크 구조에 맞춰 변경하는 경우, 앵커 성과 유사하고 bpp가 높아질수록 앵커보다 나은 성능을 보인다.

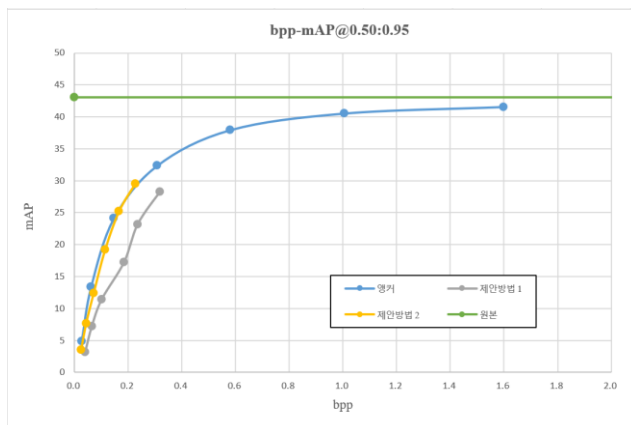


그림 4. 제안한 방법의 실험 결과

4. 결론

본 논문에서는 CompressAI의 압축 네트워크를 활용해 객체 검출 네트워크 과정 중에 추출되는 피쳐 맵을 압축하는 방법을 제안하였다. 제안 방법 모두 피쳐 맵을 VVC로 압축한 결과에 비해 나은 성능을 보인다. 하지만, 제안 방법 1의 경우, 피쳐 맵 영역에서 패딩으로 인해 발생하는 코딩 영역 증가와 엔트로피 코딩 성능 저하로 인해 앵커나 제안 방법 2에 비해 낮은 성능을 보인다. 제시된 실험결과를 바탕으로 입력 피쳐 맵을 분석하여 보다 높은 압축 성능을 제시하는 압축 네트워크 설계 혹은 네트워크 학습의 손실 함수에 대한 추가적인 연구가 필요하다.

감사의 글

이 논문은 2021년 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, (전문연구실)기계를 위한 영상부호화 기술)

참고 문헌

- [1] Detectron2, <https://github.com/facebookresearch/detectron2>
- [2] CompressAI, <https://github.com/InterDigitalInc/CompressAI>
- [3] VVC, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM
- [4] M. Rafie (GTI), Yuan Zhang (China Telecom), Shan Liu (Tencent), "[VCM] Evaluation Framework for Video Coding for Machines", ISO/IEC JTC 1/SC 29/WG 2, m56989, Online, Apr. 2021.
- [5] Shurun Wang, Zhao Wang, Yan Ye, Shiqi Wang, "[VCM] Image or video format of feature compression for object detection.", ISO/IEC JTC 1/SC 29/WG 2, m55786, Online, Jan. 2021.
- [6] Jihoon Do, Jooyoung Lee, Younhee Kim, Se Yoon Jeong, Jin Soo Choi, "[VCM] Experimental Results of Feature Compression using CompressAI", ISO/IEC JTC 1/SC 29/WG 2, m56716, Online, Apr. 2021.
- [7] Sheng-Po Wang, Erh-Chung Ke, Ching-Chieh Lin, Chun-Lung Lin, "[VCM] Anchor generation results for object detection on COCO dataset", ISO/IEC JTC 1/SC 29/WG 2, m54861, Online, Oct. 2020.