

MPEG-7 서술자 이진화를 이용한 CNN 추론 결과 압축

*진회용 *정민혁 *유도진 *김상균 **이진영 **이희경 **정원식

*명지대학교 **한국전자통신연구원

*skydesert6410@gmail.com, *jmh8900@gmail.com, *djyou128@gmail.com,

*goldmunt@gmail.com, **jinlee@etri.re.kr, **lhk95@etri.re.kr, **wscheong@etri.re.kr

Compression of CNN Inference Results Using MPEG-7 Descriptor Binarization

*Hoe-Yong Jin *Min Hyuk Jeong *Do-Jin Yoo *Sang-Kyun Kim **Jin Young Lee

**Hee Kyoung Lee **Won-Sik Cheong

*Myongji University **Electronics and Telecommunications Research Institute

요 약

본 논문은 물체 검출(Object Detection)과 물체영역분할(Object Segmentation)의 CNN 추론 결과를 MPEG-7 서술자 이진화를 통해 표현함으로써 원본과의 용량을 비교한다. 영상의 사용 목적에 따라 CNN 추론 결과를 압축하여 활용할 시 원본 영상 대비 용량을 측정하여 그 효율성을 판단하는 것이 목표이다. 물체 검출과 물체영역분할에 대한 추론 결과를 MPEG-7 서술자를 이용해 압축하였으며, 비교를 위해 원본 영상, CNN 추론 결과 파일, MPEG-7 서술자, MPEG-7 서술자 이진화 파일의 크기를 측정하였다. 실험 결과, MPEG-7 서술자를 이진화를 통한 표현 방식이 원본 영상 및 추론 결과 파일에 비해 효율적임을 알 수 있었다.

1. 서론

본 논문은 영상 CNN 추론 결과를 추가 정보로 사용하여 영상 압축 효율을 높이는 동시에 다른 여러 태스크를 동시에 해결할 수 있는 유스케이스 시나리오를 제안한다. 또한, 영상의 CNN 추론 결과를 MPEG-7 서술자 이진화를 통해 압축했을 때 줄어드는 크기를 파악하는 실험을 수행한다. CNN 추론 성능이 뛰어난 Detectron2 를 이용하여 두 가지 데이터셋에 대한 물체 검출(Object Detection)과 물체영역분할(Object Segmentation)의 결과를 서술자로 표현하여 이의 효율성을 판단하고자 한다.

본 논문의 구성은 다음과 같다. 2 장에서는 서술자를 이용한

CNN 추론 결과 활용 유스케이스 시나리오를 살펴본 후, 3 장에서는 실험 방법에 대해 설명하고, 4 장에서는 실험 데이터셋에 대한 MPEG-7 서술자 이진화 압축 결과를 설명한다. 마지막으로 5 장에서는 본 논문에 대한 결론을 맺는다.

2. 유스케이스 시나리오

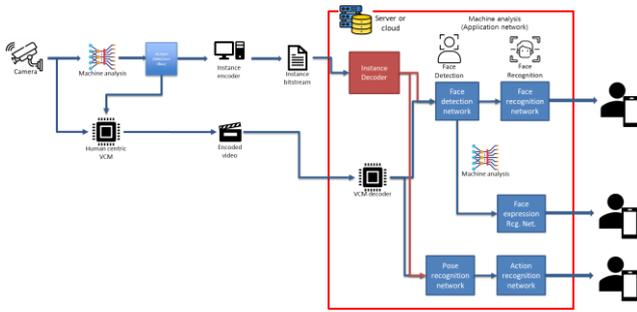


그림 1. CNN 추론 결과 활용 유스케이스

그림 1 은 카메라로 촬영된 영상에서 사람 검출을 수행하여 다양한 목적에 따라 활용하는 유스케이스다. 우선 카메라로 촬영한 영상에서 CNN 을 통해 사람 객체를 검출한다. 해당 정보를 이용하여 서술자로 표현한 후 이를 이진화 하여 압축한다. 그리고 영상에서 사람이 등장한 영역을 제외한 나머지 영역에 대해 화질을 낮추고, 사람 영역만 화질을 높여 영상 압축을 수행한다. 이후 서술자 이진화 파일과 압축한 영상을 서버/클라우드에 보낸다.

서버/클라우드에서는 이진화 결과를 서술자로 되돌리고 압축한 영상을 복원한다. 이진 영역 결과를 영상 내 사람 영역으로 복원하고, 압축한 영상을 복원하여 이를 이용해 사람과 관련된 다양한 태스크를 빠르게 해결할 수 있다. 예를 들어, 두 결과를 결합하여 얼굴 검출을 수행하고 이를 얼굴 인식과 표정 검출 등의 작업을 위해 그 결과를 활용할 수 있다. 또한 사람 영역 내 자세 인식을 수행하고, 사람의 동작 인식을 위한 작업을 동시에 수행할 수 있다.

3. 실험 방법

본 논문의 실험은 Facebook 에서 개발한 Detectron2 를 이용하여 CNN 추론 결과를 추출하고, MPEG-7 의 Region Locator 서술자를 이용하여 결과를 XML 로 표현한 후, MPEG-7 의 BiM(Binary MPEG format for XML) 참조 소프트웨어를 이용하여 이진화 하였다.

실험을 위한 데이터 셋으로는 물체 검출에 Open Images 데이터셋 5000 장을 사용했고 물체 영역 분할에 COCO 데이터셋 5000 장을 사용하였다. 물체 검출은 Faster R-CNN x101-FPN 모델을 이용하여 추론 결과를 도출하였고, 물체 영역 분할은 Mask R-CNN R50-FPN 모델을 이용하여 추론 결과를 도출하였다.

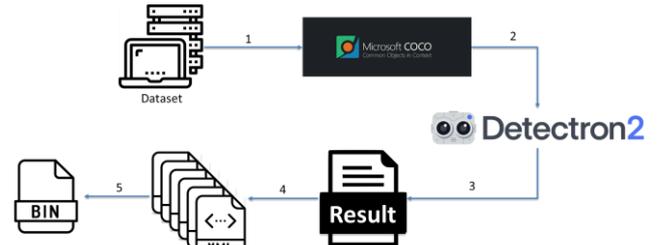


그림 2. 실험 프로세스

그림 2 는 실험의 과정을 보여준다. 가장 먼저 추론을 위해 Detectron2 에 데이터셋을 입력한다. Open images 데이터셋을 사용하는 물체 검출 프로세스의 경우 COCO 데이터셋의 형식으로 변환 후 입력한다(그림 2-1). 이후 CNN 을 통해 추론을 수행하고(그림 2-2), 추론 결과 파일을 추출한다(그림 2-3). 그리고 추론 결과 파일을 MPEG-7 서술자 형태로 변환하고(그림 2-4), BiM 참조 소프트웨어를 이용하여 MPEG-7 서술자 형태로 표현된 CNN 추론 결과를 이진화 하여 압축한다(그림 2-5).

```

XMin,XMax,YMin,YMax
0.001443387,0.132339701,0.00756371,0.417911997
0.903834403,1,0.218712608,0.545774384
0.016296538,0.986642897,0.070965157,0.97432187
0.897683382,0.999211609,0.136328885,0.235476833
0.015724119,0.589125395,0.023097706,0.442769714
0.034655105,0.973111629,0.046142694,0.871918064
0.706643224,1,0.026592559,0.450207438
0.03264619,0.445213735,0.00886589,0.446191695
0.703702152,1,0.025829773,0.477656625
0.898465574,1,0.18003506,0.541414472
0.018508079,0.699663904,0.016137467,0.560060757
0.048416894,0.133747652,0.007556114,0.260004604
0.237737072,0.988485535,0.228776231,0.988495529
0.002100744,0.278531634,0.525226533,0.989405036
0.091204413,0.998814506,0.24582164,0.983194709
0.028994444,0.98377001,0.229267672,0.984789252
0.092808038,0.990608807,0.239236996,0.988004506
0.11344673,0.985248527,0.237277448,0.977335453
0.352229327,0.884568214,0.314044,0.738466858
0.015827656,0.938850403,0.075441208,0.985226917
0,0.960397065,0.152948189,0.981517851
    
```

그림 3. 물체 검출 추론 결과

```

<mpeg7:Mpeg7 xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001" xmlns:xsi="http://www.w3.org/
<mpeg7:DescriptionUnit xsi:type="mpeg7:StillRegionType">
  <mpeg7:SpatialLocator>
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"130 12 831 342/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"941 268 37 32/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"891 275 51 24/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"112 152 194 88/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"815 311 82 34/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"198 256 34 18/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"879 279 13 18/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"112 153 180 79/>mpeg7:Box</
    <mpeg7:Box unlocatedRegion="false" mpeg7:dim="2 2"864 306 7 14/>mpeg7:Box</
    
```

그림 4. MPEG-7 서술자 표현된 물체 검출 추론 결과

그림 3 은 CNN 을 통해 추출한 물체 검출 추론 결과 파일에서 물체의 경계 상자(Bounding box) 정보만을 추출하여 정리한 파일이다. 그림 4 는 그림 3 의 물체의 경계 상자 정보를 MPEG-7 Region Locator 서술자로 표현한 XML 파일이다. 물체 분할 영역의 결과 또한 마찬가지로 물체의 경계 영역(polygon) 정보만 추출하여 XML 파일로 만들었다. 생성한 XML 파일을 BiM 소프트웨어를 통해 이진화 하였다.

4. 실험 결과

본 실험은 Open Images 데이터셋과 COCO 데이터셋의 원본 이미지 크기, 추론 결과 파일 크기, MPEG-7 서술자로 표현한 XML 파일의 크기, 그리고 BiM 소프트웨어를 사용해 이진화로 압축한 파일의 크기를 비교하였다. 압축 효율을 확인하기 위해 각 데이터 형식에 대한 BPP(Bits per pixel)를 계산하였다. 이를 아래 표 1 을 통해 결과를 정리하였다.

표 1. 데이터 형식 별 크기

Dataset	Data	Size	BPP
Open Images (Detection)	Image	1444MB	3.1566397
	Original result (text)	2.89MB	0.0063288
	XML	8.6MB	0.0187448
	BiM	0.8MB	0.0017493
COCO dataset (Segmentation)	Image	777MB	4.7620897
	Original result (JSON)	8.9MB	0.0545489
	XML	12MB	0.0735491
	BiM	2.8MB	0.0171614

실험 결과는 물체 검출의 경우 이진화 파일 크기가 원본 이미지 대비 99.95%, text 파일 대비 72.3% 감소하였고, 물체영역분할의 경우 이진화 파일 크기가 원본 이미지 대비 99.64%, JSON 파일 대비 68.6% 감소하였다.

표 2. mAP 측정 결과

Dataset	Original mAP	Restored mAP
Open Images (Detection)	58.182	58.133
COCO dataset (Segmentation)	37.168	37.006

표 2 는 mAP(mean average precision) 측정 결과로 IoU(Intersection over Union) 결과가 0.5 이상인 값들의 평균 AP 수치를 구한 결과이다. 원본 mAP 는 기존의 방식대로 추론 결과 파일을 통해 mAP 를 측정된 값이고, 복원 mAP 는 추론 결과 파일을 MPEG-7 서술자로 변환하고 이진화 한 후 다시 추론 결과 파일 형태로 복원하여 mAP 를 측정된 값이다. Open Images 데이터셋의 물체 검출 결과의 원본 mAP 는 58.182 이고 복원 mAP 는 58.133 이다. 그리고 COCO 데이터셋의 물체 영역 분할 결과의 원본 mAP 는 37.168 이고 복원 mAP 는

37.006 이다. 이를 통해 서술자 이진화 결과를 복원하여 사용하였을 경우 mAP 가 소폭 감소함을 알 수 있다. mAP 의 감소는 추론 결과 파일에서 물체의 영역을 나타낼 때 부동소수점 방식을 사용하는데, 이를 MPEG-7 서술자로 변환하는 과정에서 정수형으로 변환하면서 발생하는 것으로 보인다.

5. 결론

CNN 추론 결과를 MPEG-7 서술자로 표현하고 이진화 한 결과, 추론 결과 평가 점수인 mAP 는 소폭 감소하면서 용량은 원본 이미지 및 추론 결과 파일에 비해 크게 감소하였다. 경량화된 물체 영역 정보를 서버/클라우드로 전송하여 다양한 심화 태스크에 사용할 수 있을 것으로 보인다. 또한 동영상을 압축할 때 추론 결과를 활용하여 관심 영역은 고화질로, 배경 부분은 저화질로 압축하여 관심 영역에 대한 화질은 유지한 채 압축 효율을 증가시킬 수 있을 것으로 보인다.

감사의 글

본 연구 논문은 과학기술정보통신부 및 정보통신기획평가원의 출연금으로 수행되고 있는 “기계를 위한 영상 부호화 기술 개발”(2020-0-00011) 과제의 연구결과입니다.