

신경망 특징맵 부호화를 위한 특징맵 재배열 및 압축 방법

*한희지 **곽상운 **윤정일 **정원식 **서정일 *최해철†

*한밭대학교 정보통신전문대학원 멀티미디어공학과

**한국전자통신연구원

choihc@hanbat.ac.kr†

Feature map channel reordering and compression
for Neural Network feature map coding

Heeji Han, Sangwoon Kwak, Joungil Yun, Won-Sik Cheong, Jeongil Seo and Haechul Choi

Hanbat National University, Graduate School of Information and Communications Multimedia Engineering

Electronics and Telecommunications Research Institute

요 약

최근 영상 혹은 비디오를 이용한 신경망 기반 기술들이 활발히 응용되고 있으며, 신경망이 처리하는 임무도 다양하고 복잡해지고 있다. 이러한 신경망 임무의 다양성과 복잡성은 더욱 많은 비디오 데이터를 요구하기 때문에 비디오 데이터를 효과적으로 전송할 방법이 필요하다. 이에 따라 국제 표준화 단체인 MPEG에서는 신경망 기계 소비에 적합한 비디오 부호화 표준 개발을 위해서 Video Coding for Machines 표준화를 진행하고 있다. 본 논문에서는 신경망의 특징 맵 부호화 효율을 개선하기 위해 특징 맵 채널 간의 유사도가 높도록 특징맵 채널을 재배열하여 압축하는 방법을 제안한다. 제안 방법으로 VCM의 OpenImages 데이터셋의 5000개 검증 영상 중 임의 선택된 360개 영상에 대해 부호화 효율을 평가한 결과, 객체 검출 임무의 정확도가 유지되면서 모든 양자화 값에 대해 화소당 비트수가 감소했으며, BD-rate 측면에서 2.07%의 부호화 이득을 얻었다.

1. 서론

최근 고품질 비디오 서비스의 증가와 커넥티드 카, 감시 시스템과 같은 인공 신경망 기반의 임무의 다양화 및 대중화할 수 많은 객체 또는 이벤트에 대한 데이터를 발생시켰고 임무 수행에 대한 효과적인 분석을 위한 비디오 데이터의 급증에 기여하였다[1]. 이에 국제 표준화 단체인 MPEG은 2019년 7월 스웨덴 예테보리에서 열린 제 127차 MPEG 회의에서 이전 세대의 비디오 부호화 표준을 개선할 새로운 표준을 수립하기 위해 Video Coding for Machines (VCM) 그룹을 구성하였다[2][3]. 가장 최근에 진행된 회의인 제 134차 MPEG 회의에서는 본격적인 표준화에 앞서 기술조사요청서(Call for

Evidence, CfE)의 응답에 대한 평가를 진행하였고 현재 가장 논의가 활발하게 이루어지고 있는 객체 검출임무와 관련된 기술들을 Exploration Experiment로 지정하여 차기 회의에서 추가적인 논의가 이루어질 예정이다[4].

VCM은 압축률이 높고 계산 복잡도가 낮은 머신 비전 기반 분석 작업에 대한 딥 러닝 네트워크 특징 맵의 간결한 표현을 달성하는 것을 목표로 한다. 지난 회의에서 평가방법과 절차가 업데이트되었고 총 3개의 파이프라인이 정의되었다[5]. 그중 파이프라인 2는 머신 비전을 위한 파이프라인이다. 특징은 주어진 입력 이미지에 대해 중간 특징 맵으로 부터 추출되며 파이프라인 2의 비디오 인코더 입력은 이미지 또는 비디오 신호가 아닌 특징 맵으로 대체될 수 있다. 또한 특징 맵은 여러 개의

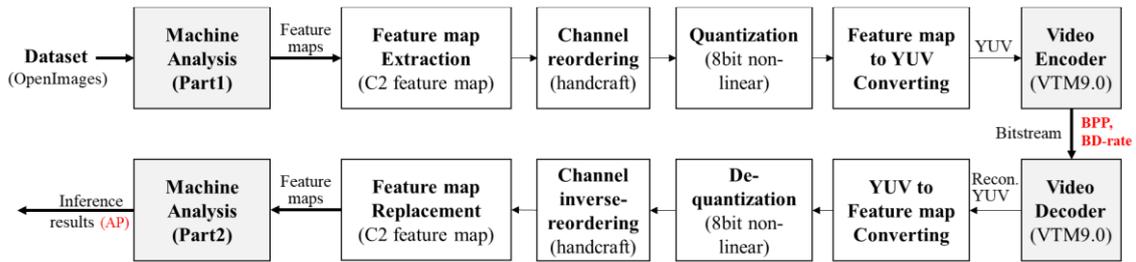


그림 1. 제안 특징맵 채널 재배열 및 압축 방법의 파이프라인

채널로 구성되어 있기 때문에 특징 맵 채널은 시간적으로 배열될 수 있다. 즉, 동일한 해상도의 비디오 시퀀스로 간주할 수 있다.

기존 비디오 코덱의 예측 기술은 공간적 또는 시간적으로 인접한 화소 간의 중복성을 제거하여 잔차 신호의 크기를 줄임으로써 압축 효율을 높였다. 따라서 본 논문에서는 특징 맵 채널을 유사도에 따라 시간적으로 재정렬하고 압축하는 특징 맵 부호화 방법을 제안한다.

2. 제안 방법

제안 방법의 전반적인 절차는 그림 1 과 같다. 위 절차에서 인코딩 프로세스는 먼저 입력된 이미지 데이터에 대하여 머신 비전 임무 수행을 위한 네트워크의 중간 단계에서 특징 맵을 추출한다. 추출된 특징 맵은 유사도에 따라 재배열되고 양자화 되고 이후 비디오 인코더에 입력하기 위해 특징 맵을 YUV 형태로 변환하고 인코딩된다. 디코딩 이후의 과정은 인코딩 과정의 역순이다.

2.1 특징 맵 추출

제안 특징 맵 부호화는 객체 검출을 위해 평가된 네트워크인 X101-FPN 백본 네트워크를 기반으로 제안되었다. X101-FPN 은 C2~C5(또는 P2~P6)의 특징 맵이 있는 피라미드 구조이며 각 특징 맵의 데이터 크기는 표 1 과 같다. 모든 특징 맵 데이터의 총 합은 원본 이미지 데이터 보다 아주 크다. X101-FPN 은 피라미드 구조이기 때문에 C3 부터 C5 까지의 특징 맵은

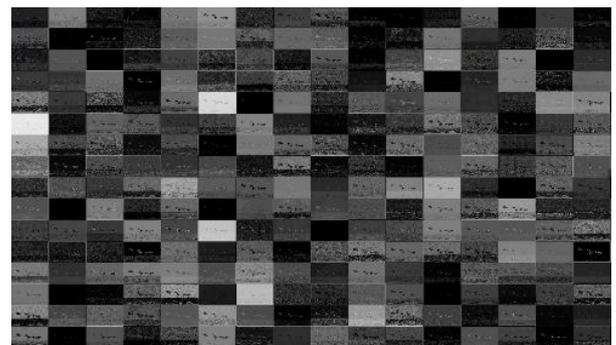
표 1 각 특징 맵의 크기와 데이터 양

Feature	Size (W×H×C)	Raw data size
Input image	1024×768×3	2,304 KB
C2	272×200×256	13,600 KB
C3	136×100×512	6,800 KB
C4	68×50×1024	3,400 KB
C5	34×25×2048	1,700 KB
P2	272×200×256	13,600 KB
P3	136×100×256	3,400 KB
P4	68×50×256	750 KB
P5	34×25×256	212.5 KB
P6	17×13×256	55.25 KB

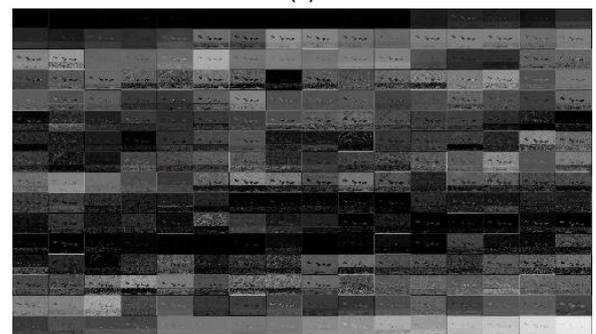
C2 로부터 얻을 수 있다. 따라서 제안 방법은 C2 특징 맵 만을 추출하여 사용하여 수행된다.

2.2 특징 맵 채널의 재배열

대부분의 비디오는 많은 시간적 상관성에 의해 부호화 효율을 얻는다. 즉, 유사한 프레임이 시간적으로 가까우면 시간적 중복성을 줄임으로써 부호화 효율을 얻을 수 있다. 제안 방법은 시간적 상관 관계를 높이기 위해 특징 맵 채널을 재정렬한다. 사전 조사를 위해 재배열 순서는 주관적으로 결정되었다. 그림 2 는 C2 특징 맵의 모든 채널에 대한 재정렬 결과를 공간적으로 시각화한 것이다. 서로 다른 이미지에서 특징 맵을 추출하더라도 동일한 네트워크를 기반으로 추출되었다면 같은 인덱스의 특징 맵 채널은 유사한 특성을 가진다. 따라서, 특징 맵 채널은 고정된 순서로 재정렬하였다.



(a)



(b)

그림 2 “f6e43b2a482d9a4e” 이미지의 특징 맵 재정렬

(a)전, (b)후 비교 결과

2.3 추출된 특징 맵의 양자화

비디오 인코더는 정수를 기반으로 연산하도록 설계되었다. 하지만, 네트워크의 출력인 특징 맵은 32 비트 부동 소수점을 기반으로 연산되고 결과를 출력한다. 따라서, 특징 맵의 값을 8 비트 정수 형태로 양자화 하는 과정이 필요하다. C2 특징 맵은 특징이 가지는 거의 대부분의 값이 0 에 가까운 영역에 분포하기 때문에, 균등 양자화보다는 비균등 양자화를 수행하는 것이 값의 범위를 표현하는데 더 효과 적일 것이다. 이에 제안 방법은 특징

$$N = 1 \ll bits$$

$$interval[N - idx] = \left(1 - \frac{\log_2(idx)}{\log_2(N)}\right) \times (max - min)$$

$$representative\ value = 0 \quad (if\ cur_value == 0)$$

$$representative\ value = idx \quad (if\ cur_value \geq interval[idx]\ and\ cur_value < interval[idx+1])$$

맵을 비선형적으로 양자화 했다. 양자화 프로세스는 아래와 같다.

2.4 YUV 변환 및 VVC 인코딩

양자화된 특징 맵 채널은 크로마 성분을 128 로 채움으로써 YUV 420 형태로 생성되며 VTM-9.0 을 기반으로 아래 매개변수와 함께 부호화 된다[10].

- Common parameters
 - Configuration: -c encoder_random_access_vtm.cfg
 - Quantization Parameters (QPs): -q {22, 27, 32, 37}
 - Bit depths: -InputBitDepth=8, -OutputBitDepth=8
- Sequence Parameters
 - Resolution: -wdt W(width of feature map channel) -hgt H(height of feature map channel)
 - Frame information: -FrameRate 1 -FrameSkip 0 -FramesToBecoded num. of feature map channels
 - Format: -InputChromaFormat 420 (default)

3. 실험 결과

최근 VCM 회의에서 논의된 OpenImages 데이터 셋 기반의 VCM 앵커는 기존 데이터셋에서 선택된 5000 개의 이미지를 기반으로 평가되었다. 제안 특징 맵 재정렬 방법은 5000 개의 이미지 중 360 개의 이미지를 임의 선택하여 평가되었다[6][7]. 표 2, 표 3 은 특징 맵 재정렬의 결과를 나타낸 것이다. 여기서 BPP 는 원본 이미지 크기에 대한 비트스트림의 크기 비율로 계산된다. 실험 결과, 특징 맵 채널 재정렬이 특징 맵 채널 재정렬을 하지 않은 부호화 방법과 비교하여 평균 2.07%의 BD-rate 감소한 것을 표 2 에서 확인할 수 있다[9]. 또한 표 3 은 부호화 효율과 객체 검출 임무의 정확도를 BPP 와 mAP 수치로 나타낸 것이다. 여기서 Uncompressed 는 비디오 압축을 하지 않은 원본 네트워크의 성능을 의미한다. Uncompressed 와

표 2 360 개 이미지로부터 추출된 특징맵에 대한 부호화 효율

Feature	Size (W×H×C)	Raw data size
Input image	1024×768×3	2,304 KB
C2	272×200×256	13,600 KB
C3	136×100×512	6,800 KB

표 3 360 개 이미지에 대한 실험 결과 (mAP, BPP)

Feature map	QP	mAP	BPP
Uncompressed	-	0.90068398	-
Compressed (Original feature map)	22	0.90049769	42.6246175
	27	0.90068357	26.5709715
	32	0.90022980	15.0271201
	37	0.86249207	6.5804821
Compressed (Reordered feature map)	22	0.89768669	42.2714206
	27	0.89907881	26.2025631
	32	0.89856748	14.6628847
	37	0.87458157	6.2219070

비교하여 제안 방법은 최악의 경우 QP 37 에서 약 0.04 의 mAP 가 감소한다. 따라서 특징 맵 채널 재정렬 및 압축 방법은 모든 QP 에서 정확도를 거의 유지했다.

4. 결론

본 논문은 VCM 의 파이프라인 2 를 기반으로 특징맵 채널 재정렬 및 로그 스케일 양자화를 적용하여 압축하는 방법을 제안했다. 실험 결과 특징 맵 채널 재정렬은 재정렬 하지 않는 것과 비교해 평균 BD-rate 를 2.07% 감소시켰다. 기존 비디오 부호화 표준을 사용해 특징 맵을 압축할 때 특징 맵 재정렬 방법이 주요 부호화 기술이 될 수 있음을 확인할 수 있었다.

감사의 글

본 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, 기계를 위한 영상 부호화 기술 개발)

참고 문헌

- [1] Cisco, "Cisco Annual Internet Report (2018-2023) White Paper", Mar. 2020.
- [2] ISO/IEC JTC1/SC29/WG11, "Versatile Video Coding (Draft 9)", JVET-R2001, Apr. 2020.
- [3] Y. Zhang, "Video Coding for Machines", ITU Workshop on "The future of media", Oct. 2019.
- [4] ISO/IEC JCT1/SC29/WG2, "Call for Evidence for Video Coding for Machines", m55605, Oct. 2020.

- [5] ISO/IEC JCT1/SC29/WG2, "Evaluation Framework for Video coding for Machines", w20125, Jan. 2021.
- [6] ISO/IEC JCT1/SC29/WG2, "VCM anchors on OpenImages datasets", w55745, Jan. 2021.
- [7] ISO/IEC JCT1/SC29/WG2, "[VCM] Anchor crosscheck for object segmentation on OpenImages dataset (crosscheck of m56189)", m56237, Jan. 2021.
- [8] ISO/IEC JCT1/SC29/WG2, m55786, "[VCM] Image or video format of feature map compression for object detection", Jan. 2021.
- [9] G. Bjøntgaard, "Calculation of Average PSNR Differences between RD-Curves", ITU-T SG16 Q.6 VCEG, VCEG-M33, April 2001.
- [10] Joint Video Experts Team of ITU-T VCEG and ISO/IEC MPEG, "VTM software 9.0", https://vcgit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM, 2020.