VCM 을 위한 다중 스케일 특징 압축 방법

*한희지 *최민석 **정순흥 **곽상운 **추현곤 **정원식 **서정일
*최해철†
*한밭대학교
**한국전자통신연구원
choihc@hanbat.ac.kr†

multi-scale feature compression for VCM

Heeji Han, Minseok Choi, Soon-heung Jung, Sangwoon Kwak, Hyon-Gon Choo, Won-Sik Cheong, Jeongil Seo,

and Haechul Choi

Hanbat National University

Electronics and Telecommunications Research Institute

요 약

최근 신경망 기반 기술들의 발달에 따라, 신경망 기술들은 충분히 높은 임무 수행 성능을 달성하고 있으며 사물인터넷, 스마트시티, 자율주행 등 다양한 환경을 고려한 응용 역시 활발히 연구되고 있다. 하지만 이러한 신경망의 임무 다양성과 복잡성은 더욱 많은 비디오 데이터가 요구되며 대역폭이 제한된 환경을 고려한 응용에서 이러한 비디오 데이터를 효과적으로 전송할 방법이 필요하다. 이에 따라 국제 표준화 단체인 MPEG 에서는 신경망 기계 소비에 적합한 비디오 부호화 표준 개발을 위해 Video Coding for Machines (VCM) 표준화를 진행하고 있다. 본 논문에서는 신경망의 특징 부호화 효율을 개선하기 위하여 VCM 을 위한 다중 스케일 특징 압축 방법을 제안한다. COCO2017 데이터셋의 검증 영상을 기반으로 제안방법을 평가한 결과, 압축된 특징의 크기는 원본 이미지의 0.03 배이며 6.8% 미만의 임무 정확도 손실을 보였다.

1. 서론

최근 인공 신경망 기반 기술들은 충분히 높은 임무 수행 성능을 달성하고 있고 이를 기반으로 다양한 환경을 고려한 응용이 활발히 연구되고 있다. 커넥티드 카, 감시 시스템과 같은 인공신경망 기반 임무는 많은 객체 또는 이벤트에 대한 수 많은 데이터를 발생시킴으로써 임무 수행에 대한 효과적인 분석을 위한 비디오 데이터 급증에 기여하고있다. 이와 같은 임무들이 대량의 데이터를 발생시키는 것에 반해 실제 수행 환경은 굉장히 제한적이다[1]. 이에 국제 표준화 단체인 MPEG 은 이전 세대의 비디오 부호화 표준을 개선할 새로운 AI 기반 표준을 수립하기 위해 Video Coding for Machines (VCM) 그룹을 구성하였다[2][3]. VCM 은 머신 비전 기반 분석 작업에 대한 딥러닝 네트워크 특징의 간결한 표현을 달성하는 것을 목표로 하고 있으며, 기계소비, 인간소비, 두 가지 모두를 고려한 3 가지 파이프라인을 정의하고 있다[4][5]. 그중 파이프라인 2 는 특징압축에 대한 것으로, 특징 압축의 과업은 부호화 할 특징 데이터의 양이 원본 이미지를 부호화 하는 것 보다 훨씬 많다는 것이다. 이를 해결하기 위해 특징 압축에 대한 다양한 연구가 진행되었다.

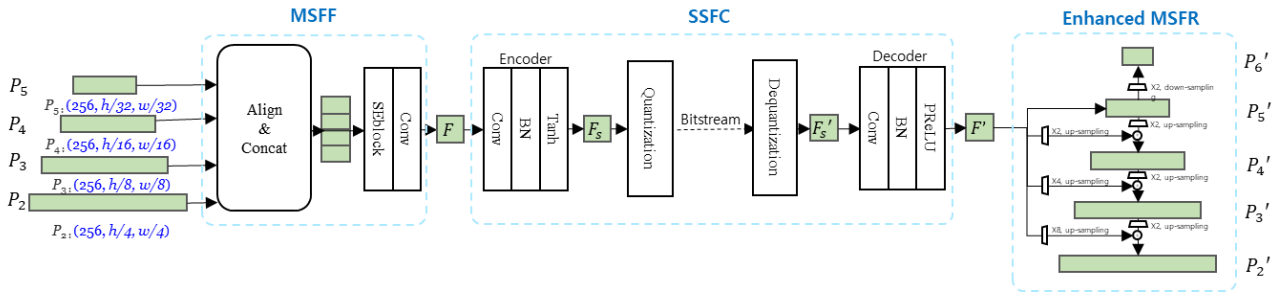

그림 1. 제안 특징맵 압축 및 재구성 과정

Multi-scale feature compression (MSFC) [10] 구조는 다중 임무 학습 네트워크를 위한 데이터 압축의 가능성을 보여주었다. 구체적으로는, 딥러닝 네트워크를 두 개의 부분으로 분할하기 위한 분할 점을 선택하며 특징을 추출한 뒤, 다중 스케일 특징을 단일 스케일 특징으로 변환한 다음 추가적인 채널 축소를 수행하여 특징을 압축한다. 압축된 특징은 복원 과정을 통해 압축 해제되며 다시 다중 스케일 특징으로 변환된다.

VCM 에서는 어느 위치에서 특징을 압축할 것인지 정의하기 위해 분할점을 정의하였다. MSFC 구조에서는 P2, P3, P4, P5 레이어에 대한 모든 정보를 전송하기 위해 P 레이어를 분할점으로 정의하였으며 이는 다중 스케일에 해당한다. [10]의 실험에 의하면 FPN 구조에서 P5 레이어가 가장 중요하다. 하지만, P 레이어의 특징을 복원할 때 MSFC 는 FPN 의 bottom-up 구조 그대로 P5 레이어를 재구성하기 때문에 P5 재구성 시 성능 저하가 발생할 수 있으며 이는 전체 네트워크의 임무 정확도 성능의 저하를 야기한다. 이에 따라 본 논문에서는 다중임무 수행을 위한 MSFC 네트워크의 특징 복원 과정에서 P5 레이어의 특징을 더 잘 복원하기 위한 방법을 제안한다.

2. 제안 방법

그림 1 은 제안 특징 압축 방법의 전반적인 과정을 나타낸 것이다. 기존 MSFC 는 분할점을 기준으로 추출된 다중 스케일 특징을 MSFF 모듈을 통해 특징의 해상도와 채널 수를 줄인다. SSFC 모듈을 통해 추가적으로 특징의 채널 수를 감소시킬 수 있다. 이후 과정에서는 줄어든 채널의 수를 복원시키고 MSFR 에서는 최종적으로 P 레이어의 특징 맵을 복원한다. 제안 방법은 MSFR 을 복원할 때 P2 부터 P5 를 만들어내도록 top-down 구조로 변경함으로써 보다 P5 레이어의 원본 특징과 가까운 특징으로 복원한다.

2.1 특징 맵 추출

제안 특징 맵 부호화는 Mask R-CNN R50-FPN 백본 네트워크를 기반으로 제안되었다. R50-FPN 은 다중 스케일 특징 맵을 기반으로 하는 피라미드 구조이며 각 특징 맵의 데이터 크기는 **표 1** 과 같다. 모든 특징 맵 데이터의 총 크기는 원본 이미지 데이터와 비교하여 7.7 배가량 크다.

표 1 각 특징 맵의 크기와 데이터 양

Feature	Size (W×H×C)	Raw data size
Input image	1024×768×3	2,304 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

2.2 개선된 P 레이어 특징맵의 복원 과정

[10]는 통계적인 방법으로 P5 가 정확도 성능에 가장 많은 영향을 미치는 레이어임을 증명하였다. MSFC 의 특징 맵 복원 과정은 P2 레이어로부터 P5 레이어를 복원하는 bottom-up 구조로 구성되어 있는 반면, 제안 특징 맵 복원 과정은 P5 레이어로부터 P2 레이어를 복원해 내는 top-down 구조로 설계되었다. 생성하는 순서를 변경함으로써 P5 레이어로부터 P2 레이어를 생성할 때 발생할 수 있는 에러 전파의 가능성을 크게 줄일 수 있다.

3. 실험 결과

제안 특징 압축 방법은 Detectron2[11]에서 제공하는 사전학습 된 Mask R-CNN 모델을 기반으로 학습되었고 이는 VCM 에서 객체 검출 임무와 인스턴스 분할 임무에 대한 평가 기준으로 사용되는 네트워크이다. 학습과정에서는 COCO 2017 학습 데이터셋을 8:2 비율로 나누어 학습과 학습도중의 평가를 위해 사용하였다. 또한 COCO 2017 검증 데이터셋 기반으로 추론을 진행하였다. **표 2** 는 제안 특징 압축 방법에 양자화 방법을 적용하였을 때의 추론 성능을 각 임무별로 나타낸 것이다. 양자화 하지 않은 경우 두 임무 모두 기존 Mask R-CNN 네트워크보다는 성능이 감소하지만, MSFC 보다는 3%의 정확도가 향상되었다. 반면, 양자화를 수행한 경우 저 비트 양자화에 대해 정확도 유지 성능이 훨씬 우수함을 확인할 수 있다. **표 3** 은 제안 특징 압축 방법을 사용할 때 감소되는 특징 데이터의 총 량을 나타낸다. 분할 지점에서 추출된 P 레이어들의 데이터 총량은 원본 이미지와 비교하여 80.278 배 크다. 반면에, 제안 구조에서 8 비트 양자화를 적용한 경우 약 16.94 배 압축률을 보였으며 2 비트 양자화에서는 최대 67.76 배 압축됨을 보였다. 따라서 제안

표 2 제안 방법의 양자화 적용/미적용시 추론 정확도 결과

Object Detection				
model	**w/o quant.**	**8-bit**	**4-bit**	**2-bit**
Mask R-CNN	61.000	-	-	-
MSFC[9]	54.781	54.762	52.733	14.411
Ours	57.403	57.388	56.831	44.940
Instance Segmentation				
Model	**w/o quant.**	**8-bit**	**4-bit**	**2-bit**
Mask R-CNN	57.993	-	-	-
MSFC[9]	50.525	50.492	48.452	12.305
Ours	53.097	53.004	52.346	41.063

표 3 제안 구조에 의한 특징의 크기 비교

Step	**Data Type**	**Data size (bits×C×W×H)**	**Data size ratio**
Original image	uint8	8×3×640×480	1.0
sum of P2 to P5	fp32	32×18496000	80.278
SSFC encoder output	fp32	32×64×25×34	0.236 (4.23×)
Quantized feature	uint8	8×64×25×34 (8-bit)	0.059 (16.94×)
	uint8	4×64×25×34 (4-bit)	0.030 (33.88×)
	uint8	2×64×25×34 (2-bit)	0.015 (67.76×)

방법을 사용한 경우 저 비트 양자화를 적용할 때 비교적 적은 임무 정확도 손실로 높은 압축률을 얻을 수 있다는 가능성을 보였다.

4. 결론

본 논문은 VCM 의 파이프라인 2 를 기반으로 분할 점에서 추출한 특징 맵을 전송을 위해 더 압축하기 위한 네트워크를 제안한다. 실험 결과 특징 맵의 크기는 원본 이미지와 비교하여 제안방법에 4 비트 양자화를 적용했을 때 원본 이미지의 0.03 배 크기를 가지며 이때 임무 정확도 손실은 6.8% 미만임을 보였다. 제안 방법은 높은 압축률을 필요로 하는 환경에서 주요 부호화 기술이 될 수 있음을 확인했다.

감사의 글

본 논문은 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2020-0-00011, 기계를 위한 영상 부호화 기술).

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”, w21287, Jan. 2022.

[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] ISO/IEC JCT1/SC29/WG2, m58772, “[VCM] Investigation on deep feature compression framework for multi-task”, Jan. 2022.

[10] Z. Zhang, M. Wang, M. Ma, J. Li and X. Fan, "MSFC: Deep Feature Compression in Multi-Task Network," 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021

[11] Detectron2, “https://github.com/facebookresearch/detectron2”