특집논문 (Special Paper)
방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)
https://doi.org/10.5909/JBE.2022.27.3.283
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

계층 간 특징 복원-예측 네트워크를 통한 피라미드 특징 압축

김 민 섭[a)], 심 동 규[a)‡]

Pyramid Feature Compression with Inter-Level Feature Restoration-Prediction Network

Minsub Kim[a)] and Donggyu Sim[a)‡]

요 약

딥 러닝 네트워크에서 사용되는 특징 맵은 일반적으로 영상보다 데이터가 크며 특징 맵을 전송하기 위해서는 영상의 압축률보다 더 높은 압축률이 요구된다. 본 논문은 딥러닝 기반의 영상처리에서 객체의 크기에 대한 강인성을 가지는 FPN 구조의 네트워크에서 사용되는 피라미드 특징 맵을 높은 압축률로 전송하기 위해 제안한 복원-예측 네트워크를 통해 전송된 일부 계층의 피라미드 특징 맵으로 전송하지 않은 계층의 피라미드 특징 맵을 예측하며, 압축으로 인한 손상을 복원하는 구조를 제안한다. 제안한 방법의 COCO 데이터 셋 2017 Train images에 대한 객체 탐지의 성능은 rate-precision 그래프에서 VTM12.0을 통해 특징 맵을 압축한 결과 대비 BD-rate 31.25%의 성능향상을 보였고, PCA와 DeepCABAC을 통한 압축을 수행한 방법 대비 BD-rate 57.79%의 성능향상을 보였다.

Abstract

The feature map used in the network for deep learning generally has larger data than the image and a higher compression rate than the image compression rate is required to transmit the feature map. This paper proposes a method for transmitting a pyramid feature map with high compression rate, which is used in a network with an FPN structure that has robustness to object size in deep learning-based image processing. In order to efficiently compress the pyramid feature map, this paper proposes a structure that predicts a pyramid feature map of a level that is not transmitted with pyramid feature map of some levels that transmitted through the proposed prediction network to efficiently compress the pyramid feature map and restores compression damage through the proposed reconstruction network. Suggested mAP, the performance of object detection for the COCO data set 2017 Train images of the proposed method, showed a performance improvement of 31.25% in BD-rate compared to the result of compressing the feature map through VTM12.0 in the rate-precision graph, and compared to the method of performing compression through PCA and DeepCABAC, the BD-rate improved by 57.79%.

Keyword : Feature map compression, Feature pyramid network, Video coding for machine, Deep learning network, Principal component analysis

Ⅰ. 서 론

최근 딥 러닝[1]을 활용한 심층신경망(deep neural network)은 의료, 게임, 공장, 자동차, CCTV, 모바일에 이르기까지 다양한 분야에 여러 가지 형태로 사용되고 있다. 심층신경망의 여러 가지 형태와 응용으로 확장이 진행되면서 자동화를 위한 기계 간 통신으로의 확장 역시 빠르게 진행되었다[2]. 원활한 기계 간 통신을 위해서는 기계 간에 취득한 영상이나 정보를 공유해야 하며 이를 위해서는, 효율적인 데이터 전송을 위한 높은 효율의 압축 방법이 요구되게 된다. 대표적인 영상 압축 표준인 HEVC(High Efficiency Video Coding)[3]나 VVC(Versatile Video Coding)[4]를 통해 영상을 압축을 수행하는 경우, 인간의 시각적 특징(feature)만을 고려하고, 기계에서 수행되는 심층신경망에서 중요하게 작용하는 특징에 대한 고려가 되어있지 않기 때문에, 기계에서 수행되는 심층신경망을 통한 객체 탐지(object detection)의 정확도가 낮아진다[5]. 객체 탐지의 정확도가 낮아지는 문제점을 해결하기 위해서 인간의 시각적 특성만이 아닌, 기계 내부에서 수행되는 심층신경망에 중요하게 작용하는 특징을 고려한 압축 방법이 연구되기 시작했다. 이에 ISO/IEC JCT1/SC29 MPEG(Moving Picture Experts Group)은 WG2에서 VCM(Video Coding for Machine)[6]이라는 새로운 표준화 활동을 2019년 7월 제127차 회의부터 진행하였다. 주된 요구사항으로 고려된 임무(task)는 객체 탐지, 객체 분할, 객체 추적, 활동 인식, 자세 추정으로 총 5가지로 정해졌으며, 이를 위한 방법으로 영상으로부터 추출된 특징 맵(feature map)을 압축하는 트랙(track)1과 영상을 압축하는 트랙 2가 VCM의 트랙으로 생성되었다. 트랙 1의 경우 특징 맵을 VVC의 참조 소프트웨어인 VTM(VVC Test Model)12.0[7]을 통해 압축을 수행한 뒤, 복원된 특징 맵을 사용하여 기계 임무를 수행하는 방법을 anchor로 사용하고 있으며, 트랙 2의 경우 영상을 VTM12.0으로 압축을 수행한 뒤, 복원된 영상을 사용하여 기계 임무를 수행하는 방법을 anchor로 사용하고 있다.

영상을 압축하는 방법으로 제안된 방법은 주로 딥 러닝을 활용하여 압축을 수행하는 방식이 제시되었다. 딥 러닝을 활용한 압축 방법은 손실함수(cost function)를 정의하여 손실함수를 최소화하는 학습방식을 통해 각 기계 임무 수행에 대한 네트워크의 최적화를 수행한 뒤, 네트워크를 통해 압축을 수행한다. 최적화를 통해 트랙2의 anchor보다 높은 압축 성능을 보인다[8][9]. 특징 맵을 압축하는 방법으로 제안된 방법은 주로 주성분 분석 방법인 PCA (Principal Component Analysis)를 활용한 방법이 좋은 성능을 보이며, 트랙 1의 anchor는 물론 트랙 2의 anchor와도 비교할 만한 성능을 보인다[10][11]. 영상을 압축하는 방법의 경우 수신부에서 복원된 영상으로부터 객체 탐지와 같은 임무를 위한 영상의 특징을 추출해야 한다는 단점이 있으며, 임무를 수행하는 과정에서 서버와 사용자 간 영상을 전송하는 과정이 요구되기 때문에 사용자 개인의 정보가 유출될 수 있는 환경에 노출된다는 점에서 사용자의 거부감을 야기한다. 전송이 완료된 후 심층신경망에 입력하기 위한 영상이 서버에서 복원되기 때문에 전송 과정 이후인 영상처리 과정에서도 사용자의 개인 정보가 유출될 수 있는 환경에 노출된다는 문제점이 발생하게 된다. 이러한 문제점은 전송하는 데이터를 영상이 아닌 심층신경망에서 사용하는 특징 맵을 전송하는 방법으로 해결할 수 있다. 그러나, 영상의 특징 맵을 전송하는 방법은 특징 맵이 일반적으로 영상보다 큰 데이터 크기를 가지고 있기 때문에 높은 압축률을 달성해야 한다는 단점이 있다. 이 단점은 특히, 객체 탐지, 객체 추적, 객체 영역 분할과 같은 분야에서 객체와의 거리 및 객체의 크기에 대한 강인성(scale invariant)을 이유로 FPN(Feature Pyramid Network)[12] 구조를 기반으로 한 네트워크를 주로 사용하고, FPN구조를 기반으로 한 네트워크는 여러 개의 특징 맵을 계층(level)별로 출력하기 때문에 크게 문제가 된다. 큰 데이터양은 전송 시 통신 네트워크 환경에 부하를 유발할 수 있으며, 큰 데이터양을 줄이기 위한 높은 압축률은 데이터의 손상을 야기하여 심층신경망에서의 결과가 저하된다.

a) 광운대학교 컴퓨터공학과(Department of Computer Engineering, Kwangwoon University)
‡ Corresponding Author : 심동규(Donggy Sim)
E-mail: dgsim@kw.ac.kr
Tel: +82-2-940-5470
ORCID: https://orcid.org/0000-0002-2794-9932
※이 논문은 2022년도 광운대학교 교내학술연구비 지원 및 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2021R1A2C2092848)의 지원을 받아 작성되었습니다.
· Manuscript April 12, 2022; Revised May 3, 2022; Accepted May 3, 2022.

특징 맵을 압축하는 방법으로 제안된 방법 중 한 가지는 PCA를 통해 특징 맵을 기저 벡터와 변환 계수로 변환하여 각각을 VTM12.0과 DeepCABAC[13]을 통해서 압축하여 전송하는 방법(KW_anchor)[10] 이다. 특징 맵의 주성분들을 순서대로 유도하는 PCA를 통해 많은 energy가 포함된 상위 주성분만을 전송하는 방식으로, 높은 데이터 압축률을 보이지만 VTM12.0으로 인한 기저 벡터의 손상과 DeepCABAC으로 압축을 수행한 변환 계수의 손상이 동시에 존재하기 때문에 압축률에 따른 성능 저하가 크게 나타난다. 다른 방법 중 한 가지는 앞서 언급한 기술과 유사한 기술로, PCA를 활용하였지만 변환 계수를 전송하지 않고 송신부와 수신부가 대량의 데이터셋을 통해 유도한 변환 계수를 공유하고 있으며, 해당 변환 계수를 통해 기저 벡터를 유도하여 상위 기저 벡터만을 전송하는 방법[11]이다. 이는 앞서 KW_anchor에서의 기저 벡터와 변환 계수에 손상이 모두 존재한다는 문제점을 해결하였으나, 데이터셋을 통해 유도한 평균적인 변환 계수를 사용하여 유도된 기저 벡터에 포함된 energy는 분포가 커지기 때문에, 보다 많은 기저 벡터를 전송해야 하는 단점이 있다.

기본적으로 특징 맵의 데이터의 양이 영상보다 크다는 단점과 기저 벡터와 변환 계수에 존재하는 손상으로 인한 손상된 특징 맵을 사용함으로써 생기는 성능 저하의 문제를 해결하기 위해서 본 논문에서는 피라미드 특징 맵(pyramid feature map)의 손상을 복원하기 위한 피라미드 특징 복원 네트워크와 특징 맵의 많은 데이터양에 대한 단점을 해결하기 위한 피라미드 특징 예측 네트워크를 제안한다. 제안하는 방법은 다음과 같다. 낮은 계층의 피라미드 특징 맵을 제외한 피라미드 특징 맵을 PCA와 DeepCABAC을 통해 압축을 수행한 뒤 전송한다. 전송된 피라미드 특징 맵은 피라미드 특징 복원 네트워크로 입력되어 압축으로 인한 손상이 복원된다. 복원된 피라미드 특징 맵은 피라미드 특징 예측 네트워크에 입력되어 전송하지 않은 계층의 피라미드 특징 맵을 예측하여 생성한다. 제안한 방법의 COCO data set 2017 Train images[14]에 대한 객체 탐지의 성능인 mAP(mean average precision)는 모든 피라미드 특징 맵을 VTM12.0으로 압축하여 전송하는 방법 대비 rate-precision 그래프에서 BD-rate[15] 35.77%의 성능향상을 보였으며, PCA와 DeepCABAC을 통해 압축하여 전송하는 방법인 KW_anchor 대비 BD-rate 57.48%의 성능향상을 보였다.

본 논문의 구성은 다음과 같다. 2장에서 관련 이론을 소개하며, 3장에서는 본 논문에서 제안한 방법에 대해서 자세히 설명하고, 4장에서는 제안하는 방법의 성능을 평가하고 5장에서 결론을 맺는다.

Ⅱ. 제안한 방법

영상으로부터 추출한 특징 맵을 압축하는 방법은 특징 맵이 일반적으로 영상보다 데이터양이 많기 때문에 영상을 압축하여 전송하는 방법보다 높은 압축 효율을 요한다. 서론에서 언급한 KW_anchor 방법은 FPN으로부터 출력된 피라미드 특징 맵을 PCA를 통해 기저 벡터와 변환 계수를 유도한 뒤, 유도된 기저 벡터와 변환 계수를 전송한다. 해당 방법은 VTM12.0으로 인한 기저 벡터의 손상과 DeepCABAC으로 압축을 수행한 변환 계수의 손상이 동시

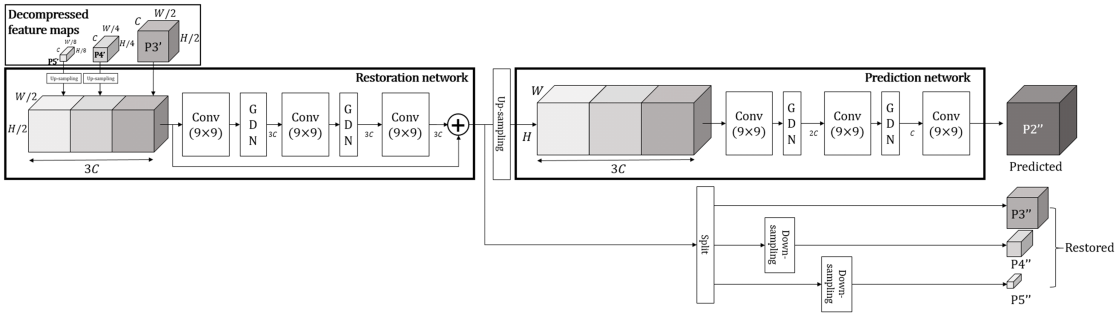


그림 1. 특징 복원-예측 네트워크의 구조도
Fig. 1. Block diagram of feature restoration-prediction network

에 존재하기 때문에 압축률에 따른 성능 저하가 크게 나타난다. 이를 해결하기 위해 본 논문에서는 피라미드 특징 맵의 큰 데이터양과 압축으로 인한 손상을 동시에 해결하기 위한 방법을 제안한다. 제안하는 네트워크의 전체적인 구조도는 그림 1과 같다.

인코더(encoder)에서 FPN 기반 구조를 가진 네트워크로부터 추출된 피라미드 특징 맵 중에 가장 낮은 계층의 피라미드 특징 맵인 P2를 제외한 나머지 피라미드 특징 맵인 P3, P4 그리고 P5를 인코딩(encoding)을 진행한 뒤 디코더(decoder)로 전송한다. 디코더는 전송된 데이터를 디코딩(decoding)한 후, 복호화된 피라미드 특징 맵인 P3', P4' 그리고 P5'을 피라미드 특징 맵 복원 네트워크의 입력으로 넣는다. 입력된 특징 맵들은 P3'의 높이와 너비를 W와 H로 이중 선형 보간법(bilinear interpolation)을 통해 업 샘플링(up sampling)한다. 이후 업 샘플링된 압축 손실을 복원하기 위해 학습된 복원 네트워크를 거치게 된다. 복원 네트워크로부터 출력된 tensor에 분할과 다운 샘플링(down sampling)과정을 거쳐 각각 복원된 P3, P4 그리고 P5인 P3'', P4'' 그리고 P5''가 재생성 되며, 복원 네트워크로부터 출력된 tensor는 예측 네트워크의 입력으로도 사용된다. 예측 네트워크는 입력된 데이터를 통해 가장 낮은 계층의 특징 맵을 예측한 P2''을 생성한다.

1. 피라미드 특징 맵의 계층 간 상관성

컴퓨터 비전(computer vision)에서 객체 인식 및 객체 분류의 경우, 영상 내에서 객체의 크기가 변하더라도, 같은 결과가 나와야 하므로, 크기에 대한 강인성이 요구된다. 크기에 대한 강인성을 가지기 위해 주로 FPN구조를 기반으로 한 네트워크가 사용되며, FPN구조를 기반으로 한 심층 신경망의 특징 추출기에서는 계층별 피라미드 특징 맵이 추출된다. FPN구조를 기반으로 한 네트워크의 구조도는 그림 2와 같다.

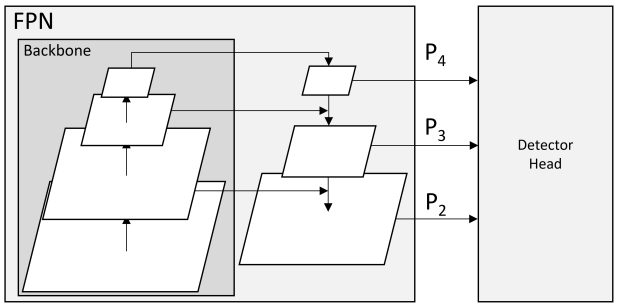


그림 2. FPN 기반의 객체 탐지 네트워크 구조도
Fig. 2. Block diagram of object detection network based on FPN structure

특징 맵 추출기는 상향식 접근방식으로 계층별 중간 피라미드 특징 맵을 추출하는 백본(backbone)과 추출된 중간 피라미드 특징 맵을 측면 연결과 하향식 접근방식으로 취합하여 최종적으로 피라미드 특징 맵을 출력하는 넥(neck)으로 이루어져 있다. 백본의 네트워크 구조는 고정되어 있지 않으며, ResNet[16]이나 ResNeXt[17]과 같은 네트워크를 백본 네트워크로 사용할 수 있다. FPN 구조를 기반으로 한 심층 신경망에서 출력된 계층별로 이루어진 피라미드 특징 맵은 출력되는 과정에 의해서 계층 간의 높은 상관성을 가지게 된다. 본 논문에서 제안한 네트워크에 입력되는 피라

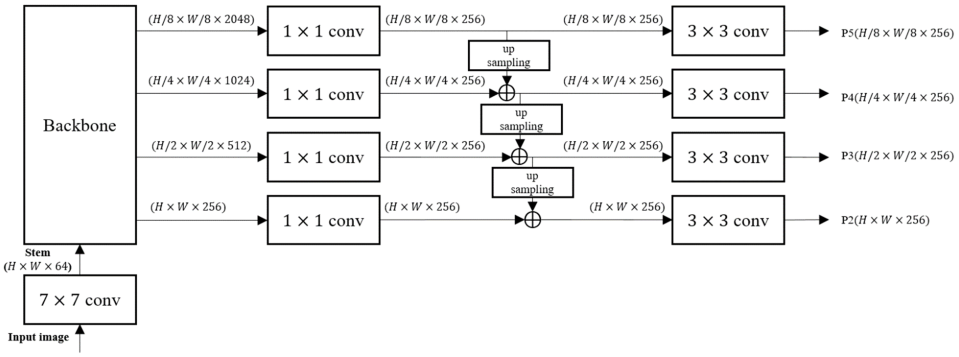


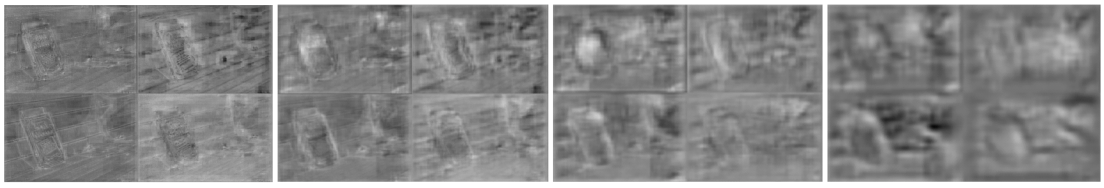
그림 3. FPN기반 네트워크의 특징 맵 추출부 구조도
Fig. 3. Feature map extraction part of network based on FPN structure

미드 특징 맵을 추출하는 과정으로 사용한 네트워크는 Detectron2[18]의 faster_rcnn_X_101_32x8d_FPN으로 구조는 그림 3과 같다.

입력된 영상으로부터 필터(filter)의 크기가 7 × 7이고 간격(stride)가 2인 합성곱(convolution) 연산을 수행한 후, 활성화 함수(activation function)인 Rectified Linear Unit (ReLU)[19] 을 거쳐, 커널(kernel)의 크기가 3 × 3이고 간격이 2인 최대 풀링(max pooling)을 통해 크기가 H × W × 64인 줄기 특징 맵(stem)을 추출한다. 추출된 줄기 특징 맵은 백본 네트워크의 입력으로 사용되며, 백본 네트워크로부터 출력된 피라미드 특징 맵은 그림 3에서 백본 네트워크의 이후 과정을 거쳐 최종적인 피라미드 특징 맵이 추출된다. 추출된 피라미드 특징 맵은 P2, P3, P4, P5, P6로 이루어져 있으며, P6는 P5의 서브 샘플링(sub sampling)을 통해서 생성한다. 피라미드 특징 맵을 추출하는 과정에서 낮은 계층의 피라미드 특징 맵을 생성하기 위해 중간에 생성된 특징 맵과 높은 계층의 특징 맵과 합 연산이 존재한다.

그림 4는 각 계층의 특징 맵을 P2의 해상도로 시각화 한 것으로, 합 연산 과정으로 인하여 낮은 계층의 피라미드 특징 맵은 상위 계층의 피라미드 특징 맵과 정보의 상관성이 존재하는 것을 확인할 수 있다. 본 논문에서는 FPN으로부터 출력되는 피라미드 특징 맵에서 계층 간의 정보의 상관성을 활용하여 효과적으로 압축을 수행한다. 정보의 유사성은 피라미드 특징 맵의 압축 효율을 상승시키기 위해 효과적으로 활용될 수 있다. 피라미드 특징 맵의 계층 간 정보의 상관성을 확인하기 위해서 다음과 같은 실험을 수행했다.

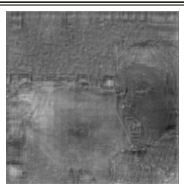
표 1은 그림 3의 과정을 통해서 출력된 특징 맵을 계층별로 시각화 한 것으로 피라미드 특징 맵의 계층별 차원(dimension)은 P2가 전체의 75%가량을 차지하고 있다. 이는 전송 데이터의 측면으로 생각할 경우, 계층 간 정보의 상관성을 통해 압축 효율을 올리는 과정에 있어서 P2가 가장 효과적인 대상일 가능성이 높다는 것을 의미한다. 가장 하위 계층의 피라미드 특징 맵인 P2와 그 바로 상위 계층의 피라미드 특징 맵인 P3와의 상관성을 확인하기 위해서, P3를 특징 맵 추출과정에서 수행하는 업 샘플링 과정과 동일한 업 샘플링 방법인 최단 입점 보간법(nearest neighbor interpolation)을 통해서 업 샘플링을 수행하고 P2 대신 사용하여 객체 탐지 성능의 변화를 확인했다. 원본 P2를 사용했을 때와 최단 입점 보간법을 통해서 업 샘플링된 P3를 P2로



(a) P2 (b) P3 (c) P4 (d) P5

그림 4. 피라미드 특징 맵의 계층별 시각화 (a) P2; (b) P3; (c) P4; (d) P5
Fig. 4. Visualization of each level of pyramid feature map (a) P2; (b) P3; (c) P4; (d) P5

표 1. 피라미드 특징 맵의 계층 간 차원 비율
Table 1. Dimensional ratio between layers in FPN feature map

Level	P2	P3	P4	P5
				
Dimension	200×200	100×100	50×50	25×25
Percentage	75.2	18.8	4.7	1.1

사용했을 때의 성능 변화는 표 2와 같다.

표 2. 사용된 P2 데이터에 따른 객체 탐지 성능
Table 2. Object detection performance according to the data used as P2

CASE	mAP	AP_L	AP_M	AP_S
Original P2	42.525	55.204	44.032	26.085
W/O P2	30.464	55.257	35.845	1.168
Up-sampled P3	37.881	55.204	43.775	14.964
W/O lateral P2	42.878	55.298	44.408	28.069

표 2는 COCO 2017 Val images에서 100장을 랜덤으로 추출하여 객체 탐지 성능을 P2의 변형을 통해 확인한 결과이다. Original P2는 그림 3에서 추출된 P2를 사용한 결과이며, W/O P2는 P2가 성능에 어느 정도 영향을 미치는지 확인하기 위해 원본 피라미드 특징 맵인 P2 대신 P2와 동일한 차원을 가진 0으로 채워진 특징 맵을 사용했을 때의 객체 탐지 성능이고, Up-sampled P3는 P3와 P2 간의 정보의 상관성을 확인하기 위해서 P3를 최단 입점 보간법을 통해 업 샘플링을 수행하여 P2로 사용했을 경우의 성능이다. 마지막으로, W/O lateral P2는 각 피라미드 계층의 피라미드 특징 맵의 마지막에 존재하는 3×3 합성곱 층의 영향을 확인하기 위해 그림 상의 왼쪽에 있는 1×1 합성곱 층으로부터 전달되는 tensor 대신 0으로 이루어진 동일한 차원의 tensor를 사용한 경우이다. W/O P2에서 주로 작은 크기의 객체를 탐지하는 성능 지표인 AP_S의 성능이 크게 내려간 것을 보아, P2는 작은 크기의 객체 탐지에 큰 영향을 미친다는 것을 확인할 수 있다. 또한, Up-sampled P3를 통해, 하락한 AP_M과 AP_S가 큰 향상을 보이는 것을 보아, P3가 P2와 정보의 상관성이 크다는 것을 확인할 수 있다. 마지막으로 W/O lateral P2의 성능이 원본 P2를 사용한 Original P2와 오차 범위 내의 성능을 보이는 것을 통해 왼쪽으로부터 전달되는 정보가 결과에 미치는 비중이 크지 않음을 알 수 있다. 또한, Up-sampled P3와 W/O lateral P2의 성능 차이를 통해, P3를 P2로 업 샘플링하는 방법보다 각 피라미드 계층별로 마지막에 위치한 3×3 합성곱 층 이전에 시점에서 업 샘플링을 수행하고 3×3 합성곱을 수행하는 방법이 좋은 성능을 보이는 것을 확인할 수 있다. 이는 각 피라미드 계층별로 마지막에 위치한 3×3 합성곱 층이 성능에 미치는 영향이 높다고 유추할 수 있고, 3×3 합성곱 층 이전에 위치에서 데이터에 오차가 발생할 경우 오차가 증폭될 가능성이 높다고 해석할 수 있다. 따라서, 본 논문은 3×3 합성곱 층 이후에 위치한 피라미드 특징 맵을 피라미드 계층 간 피라미드 특징 맵 복원 및 예측 네트워크를 통하여 압축 및 전송하는 방법을 제안한다.

2. Generalized Divisive Normalization

본 논문에서 GDN(Generalized Divisive Normalization)[20]은 심층신경망에서 여러 노드 간에 전달되는 데이터의 가중 합을 출력하는 활성화 함수로 사용된다. 그 외에 많이 사용되는 활성화 함수로는 Sigmoid, ReLU, Leaky ReLU등이 있다. Sigmoid의 경우 입력되는 값에 따른 결과값의 기울기 값이 0에 수렴하는 구간에 의해 심층신경망의 깊이가 깊어질수록 역 전파(back propagation) 과정에서 연쇄 법칙(chain rule)으로 인한 기울기 소실(gradient vanishing)에 의한 문제가 생긴다. ReLU의 경우 이러한 기울기 소실 문제가 완화되기 때문에 많은 연구에서 sigmoid가 아닌 ReLU를 활성화 함수로 사용한다. 하지만 ReLU의 경우 0보다 작은 값들이 입력되면 0을 출력하기 때문에, 정보의 손실이 생길 가능성이 있다. 또한 값을 출력하는 과정에서 0 이상의 값에 대해 그대로 출력하기 때문에 높은 비선형성을 위해서는 많은 수의 ReLU가 필요하게 된다. ReLU의 특성과 다르게 GDN의 경우 0보다 작은 값에 대해 0으로 출력하지 않고 모든 값에 대해서 변환을 수행하여 값을 출력하기 때문에 ReLU보다 높은 비선형성을 가지고 적은 정보의 손실이 일어나기 때문에 최근 연구에서 널리 사용되고 있다[21][22][23][24]. 본 논문은 ReLU를 활성화 함수로 사용한 네트워크 구조보다 적은 수의 합성곱 층을 사용하기 위해서 GDN을 활성화 함수로 사용한다.

3. 압축 손실 복원

피라미드 특징 맵의 효율적인 송수신 과정을 위해 인코딩 및 디코딩 과정이 수행되는 과정에서 특징 맵에는 이로

인한 손상이 생기게 된다. PCA를 기반으로 압축을 수행하는 방법에서 손상은 크게 3가지가 있다. PCA를 통해서 유도된 기저 벡터에서 변환 계수의 분산이 큰 MBV(major basis vector)만을 사용하여 각 계층의 피라미드 특징 맵을 복원하는 과정에서 생기는 손상과 유도된 MBV를 VTM 12.0을 통해 압축하는 과정에서 생기는 손상 그리고 MBV의 계수를 DeepCABAC으로 압축하는 과정에서 생기는 손실이다. 이러한 손상을 포함한 피라미드 특징 맵으로 객체 탐지를 수행하는 경우 성능 하락이 나타난다. 피라미드 특징 맵에 생긴 압축 손상을 복원하여 감소한 객체 탐지의 성능을 향상시키기 위해서 본 논문이 제안하는 네트워크인 피라미드 특징 맵 복원 네트워크의 전체적 구조도는 그림 5와 같다.

제안한 피라미드 특징 맵 복원 네트워크에 입력으로 들어오는 피라미드 특징 맵은 그림 3에서의 P3, P4, P5 피라미드 특징 맵을 부복호화 과정에서 압축 손상 추가된 P3', P4', P5'이다. 그림 4를 통해 피라미드 특징 맵의 각 계층 간의 상관성이 존재하는 것을 확인할 수 있으며, 이는 각 계층을 복원하는 과정에 있어서 서로 간의 정보를 활용할 수 있음을 의미한다. 각 피라미드 특징 맵에 생긴 압축 손상을 복원하는 과정에서 계층 간의 상관성을 활용하기 위해서 P4'와 P5'를 P3'의 너비(W/2)와 높이(H/2)로 각각 이중 선형 기반 업 샘플링을 수행한 후 채널(C) 방향으로 붙여 최종적인 입력으로 사용하였다. 따라서, 최종적으로 입력되는 특징 맵은 W/2 × H/2 × 3C의 크기를 가지며, 크기가 9인 필터와의 2-D 합성곱과 활성화 함수인 GDN를 거치며 복원된다. 제안한 피라미드 특징 맵 복원 네트워크의 학습 과정은 다음과 같다.

압축 손상이 포함된 피라미드 특징 맵을 입력으로 하며, 압축 손상이 포함되지 않은 원본 피라미드 특징 맵을 정답(ground truth)으로 학습한다. 네트워크를 거쳐 출력되는 특징 맵을 각 계층의 크기로 다운 샘플링한 후, 원본 특징 맵과의 MSE(Mean Squared Error) 값을 손실함수로 사용하여 최적화를 수행한다. 이를 통해 네트워크는 압축으로 인해

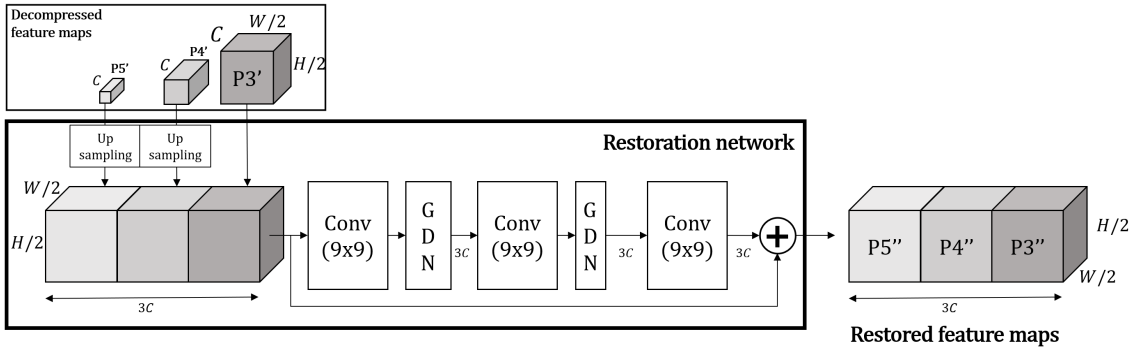


그림 5. 제안한 피라미드 특징 맵 복원 네트워크의 구조도
Fig. 5. Proposed block diagram of pyramid feature map restoration network

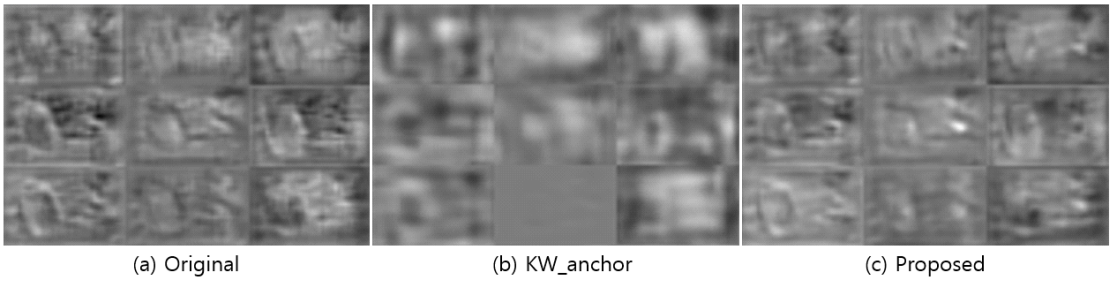


그림 6. 복원된 P5 피라미드 특징 맵 (a) Original; (b) KW_anchor; (c) Proposed (QP 47)
Fig. 6. Restored P5 pyramid feature map (a) Original; (b) KW_anchor; (c) Proposed (QP 47)

생긴 손상을 학습하여 손상이 생긴 특징 맵에 해당 손상을 보상해줌으로써 복원이 이루어진다. 그림 6은 왼쪽부터 원본 특징 맵, 복원하기 전인 KW_anchor 압축 방법의 특징 맵 그리고 복원 네트워크로부터 출력된 복원된 특징 맵이다. 압축 과정에 의해서 생긴 손상으로 손실된 특징이 크게 복원된 것을 눈으로 확인할 수 있다.

4. 피라미드 특징 맵 계층 간 예측

그림 3의 구조에서 추출되는 피라미드 특징 맵은 P2, P3, P4, P5, P6로 이루어져 있으며 그 중 P2는 전체 데이터양의 75%에 해당하는 크기를 가진다. 본 논문에서 제안하는 피라미드 특징 맵 예측 네트워크는 피라미드 특징 맵을 추출하는 과정에 의해 생기는 정보의 상관성을 활용하여, 가장 큰 데이터양을 차지하는 P2 피라미드 특징 맵을 전송하지 않고 예측하여 생성해내는 방법이다. 본 논문에서 제안한 피라미드 특징 맵 예측 네트워크의 전체적 구조도는 그림 7과 같다.

본 논문에서 제안한 피라미드 특징 맵 예측 네트워크에는 앞에서 설명한 피라미드 특징 맵 복원 네트워크에서 출력된 결과를 P2의 너비(W)와 높이(H)로 최단 입점 보간법 기반 업 샘플링을 수행한 결과가 입력된다. 따라서, 최종적으로 입력되는 특징 맵은 (W, H, 3C)의 크기를 가지며, 크기가 9인 필터와의 2-D 합성곱과 활성화 함수인 GDN을 거쳐 P2 피라미드 특징 맵을 예측하여 출력한다. 제안한 피라미드 특징 맵 예측 네트워크의 학습 과정은 다음과 같다.

압축 손상이 복원된 특징 맵인 P3'', P4'', P5''를 입력으로 하며, 예측 네트워크를 거쳐 출력되는 예측된 P2 특징 맵인 P2''(Proposed)와 압축 손상이 없는 원본 P2와의 MSE 값을 손실 함수로 사용하여 최적화를 수행한다. 그림 8은 왼쪽부터 원본 특징 맵, KW_anchor 압축 방법을 통해 전송했을 때의 특징 맵 그리고 제안한 예측 네트워크로부터 출력된 예측된 P2 특징 맵이다. KW_anchor 압축 방법으로 실제로 전송된 P2와 전체 데이터의 75%를 차지하는 P2 데이터를 전송하지 않고 예측한 P2''(Proposed)가 유사한 것을 확인할 수 있다.

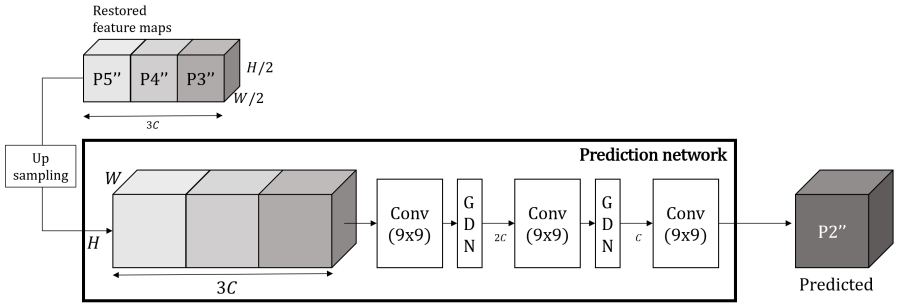


그림 7. 제안한 피라미드 특징 맵 예측 네트워크의 구조도
Fig. 7. Proposed block diagram of pyramid feature map prediction network

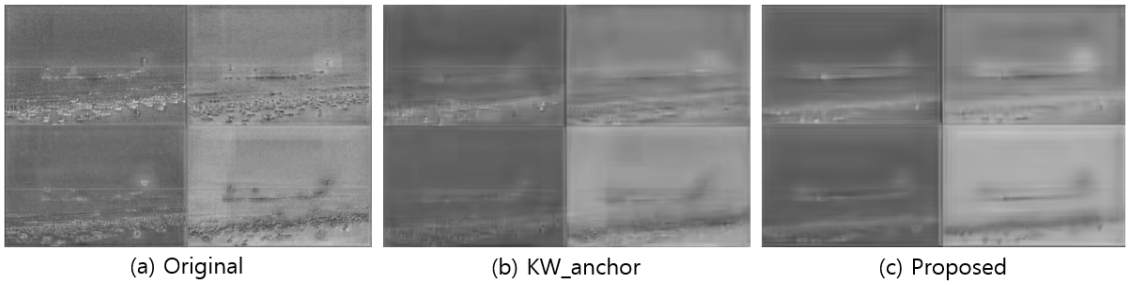


그림 8. 예측된 P2 피라미드 특징 맵 (a) Original; (b) KW_anchor; (c) Proposed (QP 47)
Fig. 8. Predicted P2 pyramid feature map (a) Original; (b) KW_anchor; (c) Proposed (QP 47)

Ⅲ. 실험 결과

1. 실험 환경

본 논문의 구현 및 실험은 Python 3.7.9, Pytorch 1.7.1, Detectron2 0.3+cu102의 환경에서 이루어졌으며, CPU 및 GPU는 각각 Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz, 4개의 GeForce RTX 3090을 사용하였다. 실험에 사용한 데이터는 학습(training)을 하기 위한 데이터로 COCO data set의 2017 Train images 중에서 40,000장을 사용하였고, 검증(validation)하기 위한 데이터로 COCO data set의 2017 Val images 5,000장을, 테스트(test)로 사용한 데이터는 학습에 사용한 데이터를 제외한 2017 Train images 중에서 5,000장을 사용하였다.

2. 실험 결과 분석

본 논문에서는 평가를 위한 객체 탐지 모델로 detectron2의 faster rcnn[25]기반의 faster_rcnn_X_101_32x8d_FPN을 사용하였으며, COCO 데이터셋에 대한 객체 크기 별 객체 탐지의 성능(mAP, AP_L, AP_M, AP_S)을 모든 피라미드 특징 맵을 압축하여 전송하는 방법, 계층 간 예측 방법으로 최단 입점 보간법을 적용한 방법과 R-P (rate-precision) 그래프에서 BD-rate로 비교하였다.

본 논문에서 제안한 방법은 FPN기반 네트워크에서 출력되는 피라미드 특징 맵 간의 상관성을 활용하여 압축으로 인한 손상을 복원하고, 전송하지 않은 피라미드 특징 맵을 생성하는 네트워크를 제안하였다. FPN기반 네트워크에서 출력된 피라미드 특징 맵들은 특징 맵을 추출하는 과정에 의해 정보의 상관성을 가지고 있음을 확인하였다. 피라미드 특징 맵에서 75% 정도의 데이터양을 차지하는 P2를 제외한 나머지 피라미드 특징 맵을 전송하여, 전송한 각 피라미드 특징 맵 계층 간의 정보의 상관성을 제안한 피라미드 특징 맵 복원 네트워크를 통해 효과적으로 압축으로 인해 생긴 손상을 복원하였다. 전송 및 복원된 피라미드 특징 맵들은 복원된 정보와 피라미드 특징 맵의 계층 간 정보의 상관성을 이용하는 피라미드 특징 맵 예측 네트워크를 통해 전송하지 않은 특징 맵 예측을 수행하였다. 실험에 대한 결과 비교를 위해 COCO data set 2017 Train images에 대한 MPEG-VCM 트랙 2의 anchor인 MPEG-VCM anchor와 KW_anchor의 객체 탐지 성능을 측정하였다.

표 3을 통해 QP에 따른 객체 탐지 성능을 확인할 수 있다. AP_S, AP_M, AP_L은 COCO 데이터셋의 기준에 의해 각각 영상에서 해당 객체가 차지하는 화소의 따라 결과를 small, medium, large로 나누어 측정한 결과이다. 제안한 방법은 데이터의 75% 정도의 데이터를 차지하는 P2를 전송하지 않았음에도 불구하고, AP_S를 제외한 모든 결과에 대해 우수한 성능을 보였다. AP_S의 경우 small로 판별된 객체에 대한 객체 탐지 성능으로, P2의 영향을 가장 많이 받게 된다. 전송하지 않고 예측을 통해 P2를 생성했지만 다른 압축 방법에 준하는 객체 탐지 성능을 보이는 것을 확인할 수 있다.

표 3. 제안하는 방법의 QP 별 COCO 데이터셋 2017 Train images 5000장에 대한 bpp 및 mAP
Table 3. bpp and mAP of the proposed method according to QPs on COCO data set 2017 Train 5000 images

QP	MPEG-VCM anchor			KW_anchor			Proposed		
	AP_S(%)	AP_M(%)	AP_L(%)	AP_S(%)	AP_M(%)	AP_L(%)	AP_S(%)	AP_M(%)	AP_L(%)
22	35.6	61.16	70.01	-	-	-	-	-	-
27	33.82	59.61	68.72	-	-	-	-	-	-
32	31.26	55.87	64.87	28.35	61.43	71.97	27.39	62.70	72.40
37	24.39	47.53	57.85	25.67	61.20	71.93	25.47	61.25	69.71
42	14.31	34.77	46.44	20.53	59.35	70.81	20.23	60.32	69.67
47	4.65	17.86	31.18	10.28	50.74	66.37	11.23	56.94	70.14
52	-	-	-	1.97	20.42	43.06	-	-	-
57	-	-	-	0.29	1.43	4.76	-	-	-

표 4. 제안하는 방법의 QP 별 COCO 데이터셋 2017 Train images 5000장에 대한 객체 크기에 따른 bpp 및 mAP
Table 4. bpp and precision of the proposed method according to QPs on COCO data set 2017 Train 5000 images

QP	MPEG-VCM anchor		KW_anchor		Proposed	
	bpp	mAP (%)	bpp	mAP (%)	bpp	mAP (%)
22	1.58	55.06	-	-	-	-
27	0.99	53.63	-	-	-	-
32	0.58	50.31	2.55	51.93	0.69	51.62
37	0.31	42.88	1.31	50.83	0.36	49.33
42	0.15	31.61	0.67	47.67	0.28	47.06
47	0.06	17.87	0.38	40.26	0.19	43.57
52	-	-	0.27	20.57	-	-
57	-	-	0.22	2.00	-	-

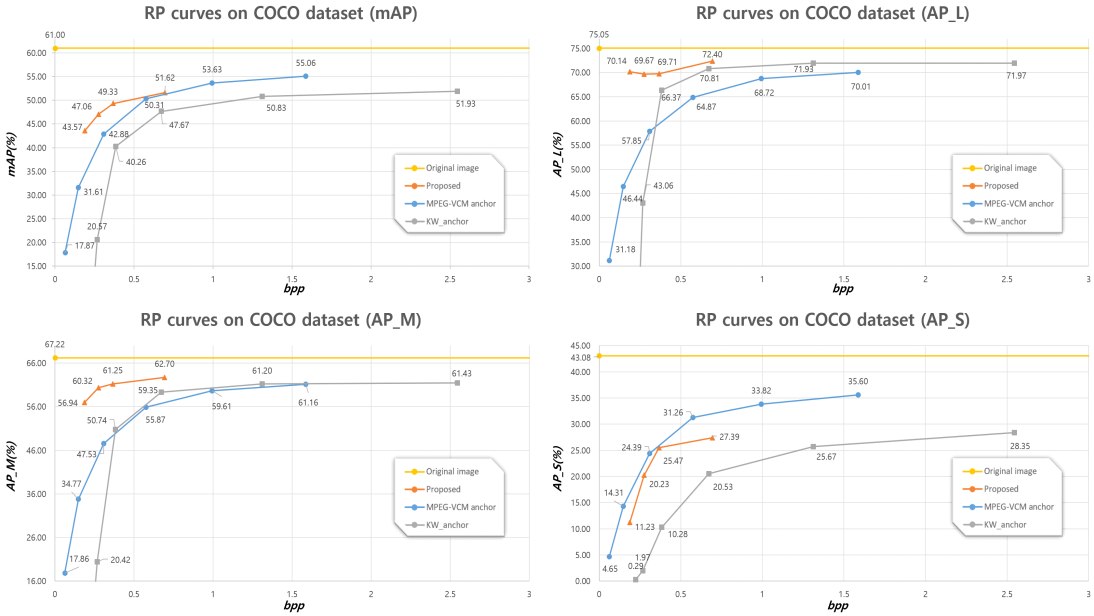


그림 9. COCO 데이터셋 2017 Train images 5000장에 대한 rate-precision 그래프
Fig. 9. Rate-precision graph on COCO data set 2017 Train 5000 images

표 4와 그림 9를 통해 제안하는 방법의 bpp(bit per pixel)에 따른 객체 탐지 성능을 확인할 수 있다. 제안한 방법은 AP_S를 제외한 모든 결과에서 다른 방법보다 높은 성능을 보였다. 그림 9의 결과 중에서 높은 bpp 영역에서의 결과는 KW_anchor의 결과가 상한치로 보이는 경향이 있음을 확인할 수 있다. 이는 제안한 네트워크가 기저 벡터에 생긴 손상인 VTM12.0에 의한 손상을 복원함에 있어서 효과적으로 작용했지만 전송하지 않은 기저 벡터에 의해 생긴 손상을 복원하는 것에는 큰 효과를 보이지 못했음을 이유로 예상할 수 있다. 이는 PCA에 의해서 유도된 기저 벡터들의 정보의 직교성(orthogonality)에 의해서 전송한 기저 벡터를 통해 정보에 있어 직교관계에 있는 전송하지 않은 기저 벡터를 예측하여 생성해내는 것에 있어 어려움이 있기 때문이라고 실험 결과를 분석할 수 있다.

표 5는 COCO data set 2017 Train images에 대해 R-P (Rate-Precision)그래프에서 MPEG-VCM anchor와 KW_anchor의 객체 탐지의 성능(mAP)에 대한 제안한 방법의 BD-rate를 측정한 결과이다. 영상을 압축하여 전송하는

표 5. 제안하는 방법의 COCO 데이터셋 2017 Train images 5000장에 대한 BD-rate
Table 5. BD-rates of the proposed method on COCO data set 2017 Train 5000 images

Target anchor	BD-rate on mAP (%)			
	mAP (%)	AP_S (%)	AP_M (%)	AP_L (%)
MPEG-VCM anchor	-31.25	28.8	-74.99	-86.35
KW_anchor	-57.79	-60.11	-66.83	-41.87

방법인 MPEG-VCM anchor에 대해서는 약 31.25%의 향상을 보였으며, PCA와 DeepCABAC을 통한 압축을 수행한 KW_anchor에 대해서는 약 57.79%의 향상을 보였다.

Ⅳ. 결 론

본 논문에서는 피라미드 특징 맵의 효과적인 압축 및 복원 방법을 위한 피라미드 특징 복원 네트워크와 피라미드 특징 예측 네트워크를 제안한다. 특징 맵의 경우, 영상보다 데이터양이 많기 때문에, 보다 높은 압축 효율이 요구된다. 높은 압축 효율을 달성하기 위해 피라미드 특징 맵의 분석을 통해 계층 간 정보 상관성을 확인하였으며, 계층 간 정보 상관성을 활용한 방법을 제안했다. 제안한 피라미드 특징 맵 예측 네트워크를 통해 전송해야 할 특징 맵의 데이터양을 감소시키고, 제안한 피라미드 특징 맵 복원 네트워크를 통해 압축률에 의해 생긴 손상을 복원하여 효율적인 압축률을 보였다. 제안한 피라미드 특징 맵 복원 예측 네트워크를 활용한 압축 방법은 다음과 같다. 송신부에서 낮은 계층의 특징 맵을 제외한 특징 맵만을 압축하여 전송하고, 수신부에서 계층 간 정보의 상관성을 활용한 복원 네트워크를 통해서 전송된 특징 맵들의 압축률에 따른 손실을 복원한 뒤, 계층 간 정보의 상관성을 이용하여 전송하지 않은 낮은 계층의 특징 맵을 예측한다. 제안한 방법의 COCO data set 2017 Train images 5000장에 대한 객체 탐지의 성능(mAP)은 R-P(rate-precision) 그래프에서 영상을 압축하여 전송하는 방법인 MPEG-VCM anchor에 대해서는 약 31.25%의 향상을 보였으며, PCA와 DeepCABAC을 통한 압축을 수행한 KW_anchor에 대해서는 약 57.79%의 향상을 보였다.

참 고 문 헌 (References)

[1] Y. LeCun, Y. Bengio, G. E. Hinton, "Deep learning," Nature, vol. 512, pp. 436-444, 2015.
doi: https://doi.org/10.1038/nature14539

[2] M. F. Mahmood, N. Hussin, "Information in conversion era: Impact and influence from 4th industrial revolution," International Journal of Academic Research in Business and Social Sciences, Vol.8, No.9, pp. 320-328, 2018.
doi: https://doi.org/10.6007/IJARBSS/v8-i9/4594

[3] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.
doi: https://doi.org/10.1109/TCSVT.2012.2221191

[4] B. Bross, Y. K. Wang, Y. Ye, S. Liu, J. Chen, "Overview of the versatile video coding (VVC) standard and its applications," IEEE Transactions on Circuits and Systems for Video Technology, Vol 31, No 10, pp. 3736-3764, 2021.
doi: https://doi.org/10.1109/TCSVT.2021.3101953

[5] S. Wang, Z. Wang, Y. Ye, S. Wang, "[VCM] Investigation on feature map layer selection for object detection and compression," ISO/IEC JTC 1/SC 29/WG 2, m55787, Online, Dec. 2020.

[6] Vedeo Coding for Machines, https://mpeg.chiariglione.org/standards/exploration/video-coding-machines (accessed July. 2019).

[7] VTM12.0, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-12.0 (accessed Nov. 26, 2021).

[8] Y. Lee, S., K. Yoon, H. Lim, H. Choo, W. Cheong, J. Seo, "[VCM] Updated FLIR Anchor results for object detection," ISO/IEC JTC 1/SC 29/WG 2, m57375, Online, Jul. 2021.

[9] S. Wang, Z. Wang, Y. Ye, S. Wang, "[VCM] End-to-end image compression towards machine vision for object detection," ISO/IEC JTC 1/SC 29/WG 2, m57500, Online, Jul. 2021.

[10] M. Lee, H. Choi, S. Park, M. Kim, "[VCM] A feature map compression based on optimal transformation with VVC and DeepCABAC for VCM," ISO/IEC JTC 1/SC 29/WG 2, m58022, Online, October. 2021.

[11] D. Gwak, C. Kim, J. Lim, "[VCM track 1] Feature data compression based on generalized PCA for object detection," ISO/IEC JTC 1/SC 29/WG 2, m58785, Online, Jan. 2022.

[12] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, "Feature pyramid networks for object detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117-2125, July. 2017.
doi: https://doi.org/10.48550/arXiv.1612.03144

[13] S. Wiedemann et al., "DeepCABAC: A universal compression algorithm for deep neural networks," IEEE J. Sel. Topics Signal Process., Vol. 14, No. 4, pp. 700-714, May 2020.
doi: https://doi.org/10.1109/JSTSP.2020.2969554

[14] COCO2017 validation set, https://cocodataset.org/#download (accessed Nov. 26, 2021).

[15] G. Bjøntegaard, "Calculation of average PSNR differences between RDcurves," Tech. Rep. VCEGM33, Video Coding Experts Group

(VCEG), 2001.
doi: https://doi.org/10.3169/itej.67.529
[16] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, June. 2016.
doi: https://doi.org/10.1109/cvpr.2016.90
[17] S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He, "Aggregated Residual Transformations for Deep Neural Networks," arXiv, 2017.
doi: https://doi.org/10.1109/cvpr.2017.634
[18] Detectron2, https://github.com/facebookresearch/detectron2 (accessed 2019).
[19] V. Nair, G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," International Conference on Machine Learning, June. 2010.
doi:https://dl.acm.org/doi/10.5555/3104322.3104425
[20] J. Ballé, V. Laparra, E. P. Simoncelli, "Density modeling of images using a generalized normalization transformation," In 4th International Conference on Learning Representations, May. 2016.
doi: https://doi.org/10.48550/arXiv.1511.06281
[21] J. Ballé, V. Laparra, E. P. Simoncelli, "End-to-end optimized image compression," In 5th International Conference on Learning Representations, May. 2017.
doi: https://doi.org/10.48550/arXiv.1611.01704
[22] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, "End-to-end blind image quality assessment using deep neural networks," IEEE Transactions on Image Processing, Vol.27, No.3, pp. 1202-1213, 2017.
doi: https://doi.org/10.1109/tip.2017.2774045
[23] J. Lee, S. Cho, H. Y. Kim, J. S. Choi, "A study on nonlinear transform layers in neural networks for image compression," In Proceedings of the Korean Society of Broadcast Engineers Conference, The Korean Institute of Broadcast and Media Engineers, pp. 267-269, 2018.
doi: https://www.koreascience.or.kr/article/CFKO201815540966800
[24] J. Ballé, P. A. Chou, D. Minnen, S. Singh, N. Johnston, E. Agustsson, G. Toderici, "Nonlinear transform coding," IEEE Journal of Selected Topics in Signal Processing, Vol.15, No.2, pp. 339-353, 2021.
doi: https://doi.org/10.1109/JSTSP.2020.3034501
[25] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems, pp. 91-99, 2015.
doi: https://doi.org/10.1109/tpami.2016.2577031

저 자 소 개

김 민 섭

- 2020년 2월 : 광운대학교 컴퓨터공학과 학사
- 2022년 2월 : 광운대학교 컴퓨터공학과 석사
- 2022년 3월 ~ 현재 : 디지털인사이트(주) 연구원
- ORCID : https://orcid.org/0000-0001-6837-4736
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전, 딥러닝

심 동 규

- 1993년 2월 : 서강대학교 전자공학과 공학사
- 1995년 2월 : 서강대학교 전자공학과 공학석사
- 1999년 2월 : 서강대학교 전자공학과 공학박사
- 1999년 3월 ~ 2000년 8월 : 현대전자 선임연구원
- 2000년 9월 ~ 2002년 3월 : 바로비젼 선임연구원
- 2002년 4월 ~ 2005년 2월 : University of Washington Senior research engineer
- 2005년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 교수
- ORCID : https://orcid.org/0000-0002-2794-9932
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전