특집논문 (Special Paper)
방송공학회논문지 제30권 제3호, 2025년 5월 (JBE Vol.30, No.3, May 2025)
https://doi.org/10.5909/JBE.2025.30.3.366
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

선택적 학습기법을 통한 FCM 시험모델 성능 개선

한 규 웅[a], 유 인 근[a], 이 주 영[b], 정 세 윤[b], 김 재 곤[a]‡

Enhancement of FCM Test Model with Selective Learning Strategy

Gyu-Woong Han[a], In-Keun Yoo[a], Jooyoung Lee[b], Se Yoon Jeong[b], and Jae-Gon Kim[a]‡

요 약

MPEG(Moving Picture Experts Group)에서는 머신비전(machine vision) 응용을 위한 새로운 비디오 부호화 표준으로 영상/비디오로부터 추출되는 특징(feature)을 효율적으로 압축하기 위한 FCM(Feature Coding for Machines) 표준을 개발하고 있다. 본 논문에서는 선택적 학습기법(SLS: Selective Learning Strategy)과 QP-적응적 채널절삭(QACT: QP-daptive Channel Truncation) 기법을 통한 FCM 시험모델인 FCTM(Feature Compression Test Model)의 압축성능 개선기법을 제시한다. SLS는 특징 채널들을 그 중요도 순서로 특징맵에 정렬하여 특징맵의 에너지를 공간적으로 집중시키면서 공간적 상관도를 높인다. 이를 바탕으로 QACT는 압축률에 따라 중요도가 떨어지는 특징 채널들을 절삭함으로써 특징맵의 해상도를 QP 적응적으로 조절하여 단일 학습모델로 넓은 비트율 범위에서 FCTM의 성능을 개선한다. 제안기법은 FCTM v5.0 대비 최대 63.32% BD-rate, 평균적으로 22.9% BD-rate 비트율 절감의 부호화 성능향상을 보여준다.

Abstract

The Moving Picture Experts Group (MPEG) is developing a new video coding standard called Feature Coding for Machines (FCM), designed to efficiently compress features extracted from images/videos in machine vision applications. This paper proposes compression enhancement methods for the Feature Compression Test Model (FCTM), the test model of FCM, by applying a Selective Learning Strategy (SLS) and QP-Adaptive Channel Truncation (QACT). SLS rearranges feature channels in a feature map in the order of their importances, thereby spatially concentrating feature energy and enhancing spatial correlation. Based on this, QACT adaptively adjusts the resolution of the feature map by truncating less important channels depending on the compression rate, improving the performance of FCTM across a wide range of bitrates. The proposed method achieves up to a 63.32% reduction in BD-rate and an overall 22.9% improvement in BD-rate compared to FCTM v5.0.

Keyword : SLS(Selective Learning Strategy), QACT(QP-Adaptive Channel Truncation), FCM(Feature Coding for Machines), FCTM(Feature Compression Test Model), Channel removal

a) 한국항공대학교 항공전자정보공학과(School of Electronics and Information Engineering, Korea Aerospace University)
b) 한국전자통신연구원(Electronics and Telecommunications Research Institute)
‡ Corresponding Author : 김재곤(Jae-Gon Kim)
E-mail: jgkim@kau.ac.kr
Tel: +82-2-300-0414
ORCID: https://orcid.org/0000-0003-3686-4786

※ This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00011, Video Coding for Machines).
· Manuscript April 7, 2025; Revised May 6, 2025; Accepted May 7, 2025.

Ⅰ. 서 론

최근 자율주행, 스마트 시티, 감시 비디오, 사물인터넷(IoT) 등 다양한 분야에서 신경망 기반 머신비전(machine vision) 기술의 활용이 빠르게 확산되면서, 기계가 소비하고 처리하는 비디오 데이터의 양도 급격히 증가하고 있다. 동시에 높은 임무(task) 수행 성능이 요구됨에 따라, 신경망 모델의 구조와 추론 과정의 복잡도도 함께 증가하고 있다. 그러나 모바일 기기는 계산 자원의 제약으로 인해 고성능 신경망 기반 비전 임무를 실시간으로 처리하는 데 한계가 있다. 이를 해결하기 위해 최근에는 머신비전 신경망을 분할하고 중간 계층에서 추출된 중간 특징(intermediate feature)을 서버로 전송하여 고성능 연산 자원을 활용하는 분할 추론(split inference) 방식이 주목받고 있다. 이때 특징을 효과적으로 압축하여 전송 비트율과 통신 지연을 최소화하고 서버 측에서 높은 성능으로 복잡한 머신비전 임무를 수행하는 구조가 고려되고 있다. 이러한 기술적 흐름에 대응하여, MPEG(Moving Picture Experts Group)은 기계 소비 환경에 최적화된 새로운 비디오 부호화 표준으로 VCM(Video Coding for Machines)과 FCM(Feature Coding for Machines)을 개발 중이다[1]. VCM은 입력 비디오 자체를 비전 임무를 고려하여 압축 전송하는 반면, FCM은 분석 네트워크를 통하여 입력 비디오로부터 추출된 특징을 압축 전송하고 서버에서 비전 임무를 수행하는 환경을 가정한다.

현재, FCM은 기술제안요청서(CfP: Call for Proposals)의 응답기술 평가를 바탕으로 시험모델인 FCTM(Feature Compression Test Model)[2]을 개발하면서 기술검증 및 표준기술 개발을 위한 핵심실험(CE: Core Experiments)을 진행하고 있으며[3][4], 오는 152차 MPEG 회의에서 위원회 초안(CD: Committee Draft)이 발간될 예정이다.

본 논문에서는 선택도(selectivity)를 조절하여 개선된 선택적 학습전략(SLS: Selective Learning Strategy)[5][6]과 QP-적응적 채널절삭 기법(QACT: QP-adaptive Channel Truncation)을 적용한 FCM의 시험모델인 FCTM의 압축성능 개선기법을 제안하고 그 성능을 분석한다. 제안기법은 SLS는 특징 채널들을 그 중요도 순서대로 특징맵(feature map)에 정렬하여 특징맵의 에너지를 공간적으로 집중시키면서 공간적 상관성을 증가시킨다. QACT는 SLS가 적용된 특징맵에서 주어진 비트율(QP)에 따라 중요도가 떨어지는 특징 채널부터 순차적으로 절삭하여 해상도를 조정함으로써 보다 효율적인 특징 압축을 가능하게 한다.

본 제안기법은 핵심실험 CE 4(Core Experiment 4)의 일부[8]로 진행되었으며 핵심실험 결과 앵커(anchor)인 FCTM v5.0 대비 최대 63.32% BD-rate, 평균적으로 22.90% BD-rate 비트율 절감의 부호화 성능향상을 보였으며, FCTM의 학습기법으로 FCM의 공통실험조건(CTTC: Common Training and Test Conditions)에 채택되었다[9].

본 논문은 2장에서 FCM의 전체구조와 3장에서 SLS에 대해 소개하고, 4장에서는 SLS를 통해 정렬된 특징맵을 주어진 QP에 따라 적응적으로 절단하는 QACT에 대해 소개한다. 5장에서는 SLS와 QACT를 적용한 FCTM의 실험결과와 성능분석을 제시하고, 마지막으로 6장에서 결론을 맺는다.

Ⅱ. FCM 개요

ISO/IEC JTC 1/SC 29의 MPEG Video 그룹(WG 4)은 머신비전 환경에서 사용되는 특징 데이터를 효율적으로 부호화하기 위한 새로운 표준으로 FCM을 개발 중이다. FCM은 분할 추론 구조에서 생성되는 중간 특징을 효율적으로 압축하고 복원하는 것을 목표로 하며, 머신비전 신경망의 Part 1(분석 네트워크)에서 추출된 특징을 부호화하여 Part 2에서 임무 수행을 위해서 사용하는 방식으로 구성된다. 이를 통해 네트워크 분할 환경에서 에지(edge)와 서버를 통한 실용적 활용이 가능하며, 전송 효율과 연산 효율을 동시에 고려할 수 있다.

그림 1은 FCTM의 전체 파이프라인을 나타낸 것으로, 특징 축소(feature reduction), 특징 변환(feature conversion), 내부 부호화(inner coding), 역변환(inverse conversion), 그리고 복원(restoration) 단계로 구성된다. FENet은 다중스케일 특징 텐서(tensor)를 단일스케일 표현으로 축소하고, DRNet은 이를 원래 차원으로 복원하는 역할을 수행한다. 축소된 특징은 2D 프레임 형태로 변환된 후 양자화(quantization) 및 정규화(normalization)를 통해 내부 코덱(codec)으로 입력된다. 내부 코덱은 VVC(Versiatile Video

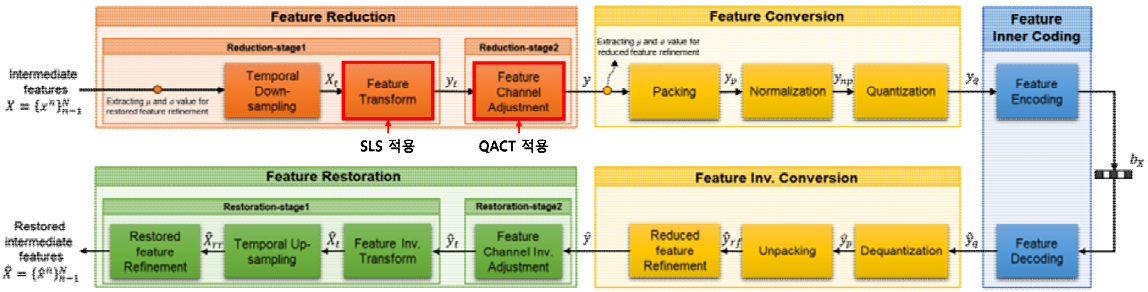


그림 1. FCTM의 특징 압축 파이프라인
Fig. 1. Feature compression pipeline of FCTM

Coding), HEVC(High Efficiency Video Coding), 또는 AVC(Advanced Video Coding) 등 표준 2D 비디오 코덱이 활용되며, 입력된 특징맵을 비트스트림(bitstream)으로 압축한다. 복호화 과정에서는 비트스트림이 다시 특징맵으로 복원되며, 이후 역변환과 복원을 거쳐 신경망의 입력 형식으로 재구성된다. 이 과정에서는 특징 언패킹(unpacking), 역양자화(dequantization), 특징 보정(feature refinement)이 순차적으로 수행된다. 전체 시스템은 환경 설정을 통해 다양한 신경망 구조와 임무 시나리오에 맞게 유연하게 구성할 수 있다. 또한, FCM은 평가의 일관성을 위해 학습 및 테스트 조건을 CTTC로 정의하고 있다. CompressAI-Vision 프레임워크는 FCM의 실험 및 성능측정을 위해 활용되고 있으며, 객체 검출(object detection), 인스턴스 분할(instance segmentation), 다중 객체 추적(multiple object tracking) 등 다양한 비전 임무에 대해 CTTC가 마련되어 있다. FCM의 구조는 영상 기반 전송 환경에서도 기계학습 성능을 유지하면서 전송 효율을 극대화할 수 있도록 설계되었다. 이러한 구조적 설계를 바탕으로, FCM은 기계 소비 중심의 다양한 응용 환경에서도 높은 임무 수행성능을 유지하며 효율적인 압축을 실현할 수 있는 표준기술을 제공한다.

III. 선택적 학습기법

선택적 학습기법(SLS)은 특징 채널별 중요도를 반영한 특징맵 생성을 가능하게 하는 종단간(end-to-end) 학습기법이다. SLS는 학습 단계마다 무작위로 연속된 일부 채널만을 선택적으로 활성화하여 학습에 참여시키며, 선택되지 않은 나머지 채널은 0으로 마스킹(masking)되어 학습에서 비활성화된다. 이러한 방식은 각 특징 채널이 학습에 포함되는 빈도에 따라 그 중요도가 자연스럽게 반영되도록 설계되어 있으며, 별도의 후처리 없이 채널 중요도를 내재적으로 학습할 수 있는 학습 구조를 제공한다.

그림 2와 같이 FENet의 출력인 단일스케일 특징은 균등분포를 따르는 무작위 변수 n에 의해 마스킹되며, 0부터 $(n-1)$까지의 특징 채널만이 학습에 활성화되고 나머지는 제거된 형태로 특징맵으로 변환된다. 변환된 특징맵은 내부 코덱으로 압축 복원 과정을 거쳐 DRNet으로 전달되어 복원된다. FENet은 다중스케일의 특징 텐서를 단일스케일로 축소(reduction)하는 학습 가능한 특징 축소 네트워크이며, DRNet은 압축된 특징으로부터 원래 차원의 특징을 복원(restoration)하는 학습 기반 특징 복원 네트워크이다. 이와 같이 일부 특징 채널을 마스킹하는 학습 구조는 FCTM에 네트워크의 구조적 변경 없이 적용될 수 있다. 학습 중 반복적으로 활성화되는 하위 인덱스의 특징 채널들이 상대적으로 더 많은 정보를 포함하도록 학습하게 됨으로써 최종 출력 특징맵은 그림 5의 예와 같이 채널별 중요도에 따라 특징맵의 좌상단부터 순서대로 정렬된 형태를 갖는다.

그림 3은 학습과정에서 무작위 변수 n에 따라 채널이 활성화되는 예시를 보여준다. SLS는 FENet의 출력 단일스케일 특징에 적용되며, 1 ~ 320의 특징 채널 범위 중 무작위로 결정된 마스킹 변수 n에 따라 n부터 320까지의 특징 채널을 마스킹함으로써 적용된다. 그림 3은 FENet의 출력이 8개의 특징 채널로 구성된다고 가정한 경우의 SLS의 채널

마스킹 과정을 보여준 것이다. 예를 들어 (n = 2)일 경우, 전체 8개의 채널 중 하위 2개의 채널만 활성화되고 나머지 상위 6개의 채널은 0으로 마스킹되어 비활성화 된다. (n = 8)인 경우 전체 8개의 모든 채널이 활성화 된다. 이와 같이 하위 인덱스의 채널이 활성화될 확률이 높아지며 상위 인덱스의 채널이 활성화될 확률은 상대적으로 낮아진다.

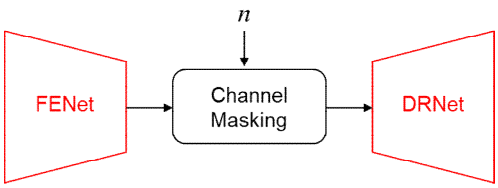


그림 2. FCTM에 적용된 SLS의 채널 마스킹 과정
Fig. 2. Channel masking process of SLS in FCTM

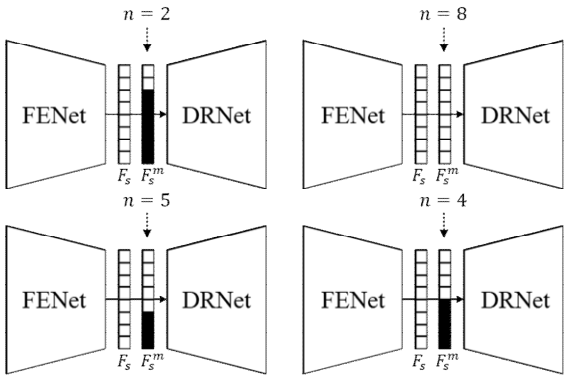
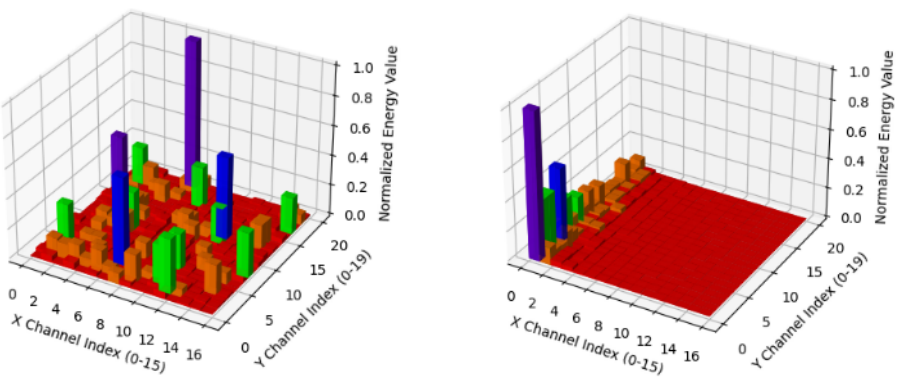


그림 3. SLS의 채널 마스킹 과정 예시
Fig. 3. Example of the channel masking process of SLS

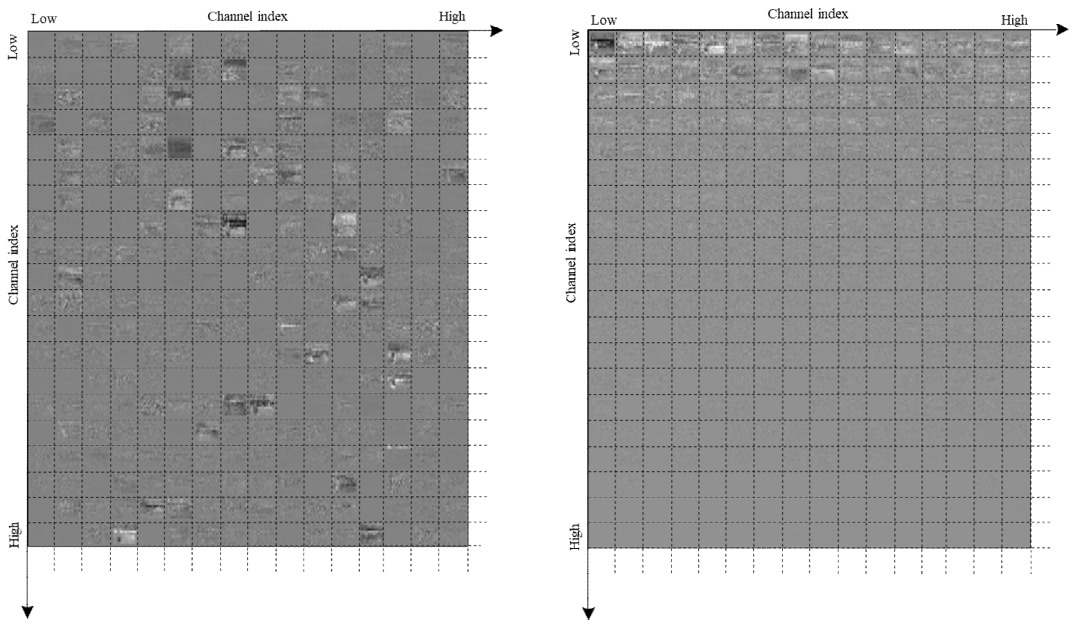
그림 4는 SLS의 적용 여부에 따른 FCTM에서 생성된 2D 프레임 형태로 변환된 특징맵의 각 채널별 에너지 분포를 나타낸 것이다. 그림 4와 같이 SLS를 적용하지 않고 기존 기법으로 생성된 특징맵의 경우 각 채널의 에너지는 무작위로 분포한다. 반면, SLS를 적용할 경우 하위 인덱스의 특징 채널이 더 많은 에너지를 포함하고 인덱스가 증가함에 따라 상위 인덱스 채널에는 에너지가 희박하게 분포하게 된다. 즉, 그림 4를 통하여 SLS에 의해서 특징맵의 에너지가 채널의 인덱스에 따라 순서대로 집중됨을 보여준다. 또한, 특징맵의 좌상단부터 에너지가 감소되는 분포를 보임으로써 특징맵의 공간적 상관도가 SLS에 의해서 증가됨을 확인할 수 있다. 이러한 공간적 상관도 증가는 특징맵을 VVC로 압축할 때 압축 효율을 개선함으로써 SLS가 FCTM의 성능을 개선하는 결정적인 역할을 하게 된다.

그림 5는 SLS를 적용한 FCTM의 2D 프레임 형태로 변환된 특징맵을 나타낸 것이다. FENet에서 출력된 단일스케일 특징맵은 SLS의 적용으로 채널별 중요도 순으로 정렬되며, 래스터 스캔(raster scan) 방식으로 첫 번째 인덱스부터 마지막 인덱스까지 순서대로 2D 프레임 형태로 변환된다. 이로 인해 좌상단에는 중요한 정보를 담은 하위 인덱스의 채널이 우선 배치된다. 그림 5와 같이 하위 인덱스에 중요한 정보를 담은 채널이 재배치되어 특징맵의 에너지가 집중되도록 하고, 상위 인덱스에 상대적으로 덜 중요한 채널이 배치되어 정렬된 것을 확인할 수 있다. 이러한 특징맵은



(a) Without SLS (b) With SLS

그림 4. SLS 적용에 따른 FCTM의 특징맵 에너지 분포
Fig. 4. Energy distribution of the feature map in FCTM with SLS



(a) Without SLS (b) With SLS

그림 5. SLS 적용에 따른 FCTM의 단일스케일 특징맵
Fig. 5. Single-scale feature map of FCTM

VVC로 부호화되는 특징맵의 공간적 상관성을 증가시켜 보다 효율적인 특징 압축이 가능하며, QACT를 통하여 적은 에너지를 포함하는 중요도가 떨어지는 특징 채널을 제거하여 특징맵의 크기를 조절하여 단일 학습모델로 임무 성능을 유지하면서 주어진 다양한 압축 비트율로 압축 가능하도록 한다.

한편, 특징맵 재정렬 기법으로 제안된 FCR(Feature Channel Rearrangement)[16]은 학습과정에서 채널 중요도를 내재화하는 본 논문의 SLS와 달리 Gain Unit을 기반으로 추론 과정에서 채널 재배열을 수행한다. 즉, FCR은 수치화된 Gain 값을 바탕으로 중요도 순서를 고정적으로 구성하여 인코더와 디코더가 동일하게 재정렬한다.

Ⅳ. SLS와 QACT를 적용한 FCTM

본 논문에서 제시한 SLS와 QACT를 적용한 FCTM의 성능 개선기법은 기존 FCTM 구조를 변경하지 않고 SLS로 미리 학습된 FENet와 DRNet 기존의 네트워크 모델을 대체하고 테스트 과정에서 QACT를 적용함으로써 실현 가능하다.

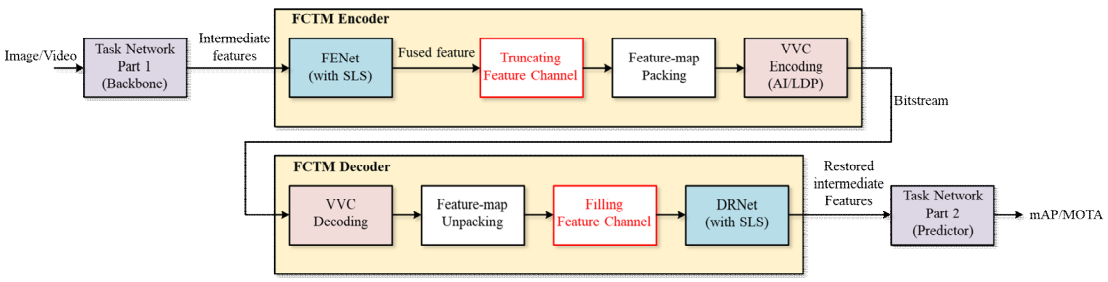


그림 6. SLS와 QACT를 적용한 FCTM 처리 파이프라인
Fig. 6. FCTM processing pipeline with SLS and QACT

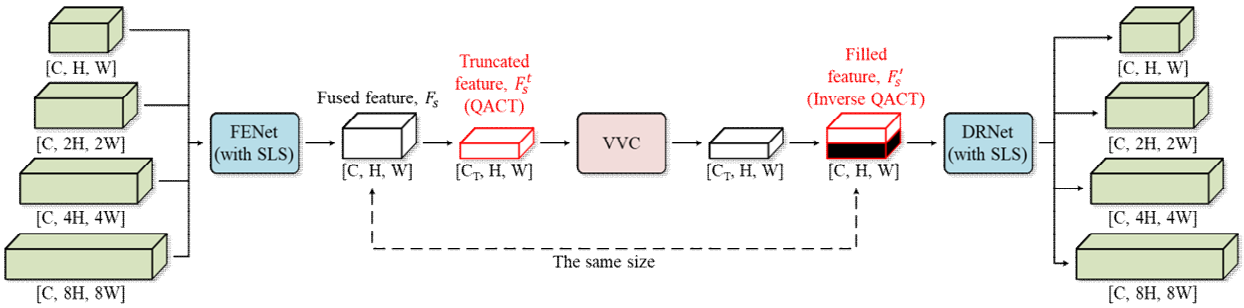


그림 7. SLS와 QACT를 적용한 FCTM 프레임워크
Fig. 7. FCTM framework with SLS and QACT

그림 6과 그림 7은 그림 1의 FCTM 압축 파이프라인에 SLS와 QACT를 구현한 FCTM의 전체 처리 파이프라인과 프레임워크를 보인 것이다. 먼저, 머신비전 네트워크로부터 추출된 다중스케일 특징은 FENet을 통해 단일스케일 특징으로 축소되며, 이 과정에서 각 단일스케일 특징 채널은 SLS를 통하여 채널별 중요도에 따라 특징맵에 정렬된다. 정렬된 단일스케일 특징맵은 QACT에 의해 주어진 QP 값에 따라 인덱스가 높은 중요도가 상대적으로 낮은 특징 채널부터 적응적으로 절삭된다. 그리고 VVC 기반 부호화를 위한 2D 프레임 형태의 특징맵으로 변환된다. 변환된 특징맵은 VVC 압축 복원과정을 거치고, 복호화된 특징맵은 단일스케일 특징으로 역변환되며, 이때 절삭으로 인해 제거된 채널을 영 값으로 패딩하여 원래의 특징 채널 수를 복원한다. 마지막으로, DRNet을 통해 복원 과정을 수행하여 다시 다중스케일 특징으로 재구성된다.

QACT는 SLS에 의해 중요도에 따라 재배열된 특징맵의 특징 채널을 주어진 QP에 따라 중요도가 낮은 특징 채널부터 제거함으로써 VVC로 부호화될 특징맵의 해상도를 조정한다. 따라서, 압축 비트율이 낮을 경우 압축 전송될 특징맵의 크기를 줄이면서도 정보 손실을 최소화할 수 있도록 한다. 결국, SLS 기반의 QACT는 QP에 적응적으로 특징맵 크기를 조정함으로써 단일 학습모델로 넓은 범위의 다양한 압축 비트율에서 높은 특징 압축성능을 유지할 수 있도록 한다. 주어진 QP에 따라 QACT를 적용한 후 특징맵에 패킹되는 적절한 특징 채널 수는 실험적으로 구할 수 있다. 표 1은 FCTM에서 QP에 따른 각 성능 포인트(PP: Performance Point)에 대한 각 데이터셋에서의 특징 채널 수 예이다.

표 1. QACT 적용 이후 특징맵에 패킹되는 특징 채널의 수
Table 1. Number of packed feature channels after QACT

Dataset	PP	Number of feature channels
OpenImage	1 (Low QP)	320
	2 (Medium-low QP)	160
	3 (Medium-high QP)	80
	4 (High QP)	40
SFU,TVD, HiEve	1 (Low QP)	320
	2 (Medium-low QP)	240
	3 (Medium-high QP)	180
	4 (High QP)	135

SLS로 정렬된 특징맵은 가장 높은 비트율(Low QP)에 해당하는 PP1의 특징맵에 특징 채널 제거 없이 모두 포함되고, 다른 PP의 경우 특징 채널 수는 절반 또는 1/4로 감소한다. 그림 8은 SLS 기반의 QACT 적용 후 각각 320, 160, 80, 40 채널로 구성된 특징맵들을 보여준다. 그림 8에서, 가장 높은 비트율(Low QP)로 압축한 PP1의 경우, 특징 채널 제거 없이 FENet의 출력인 단일스케일 특징 채널을 모두 특징맵에 포함한다. 비트율이 낮아짐에 따라 낮은 비트율에서도 정보 손실을 최소화하면서 주어진 비트율 조건을 만족하기 위해서 채널절삭이 적용되어 특징맵의 크기가 축소되었음을 확인할 수 있다.

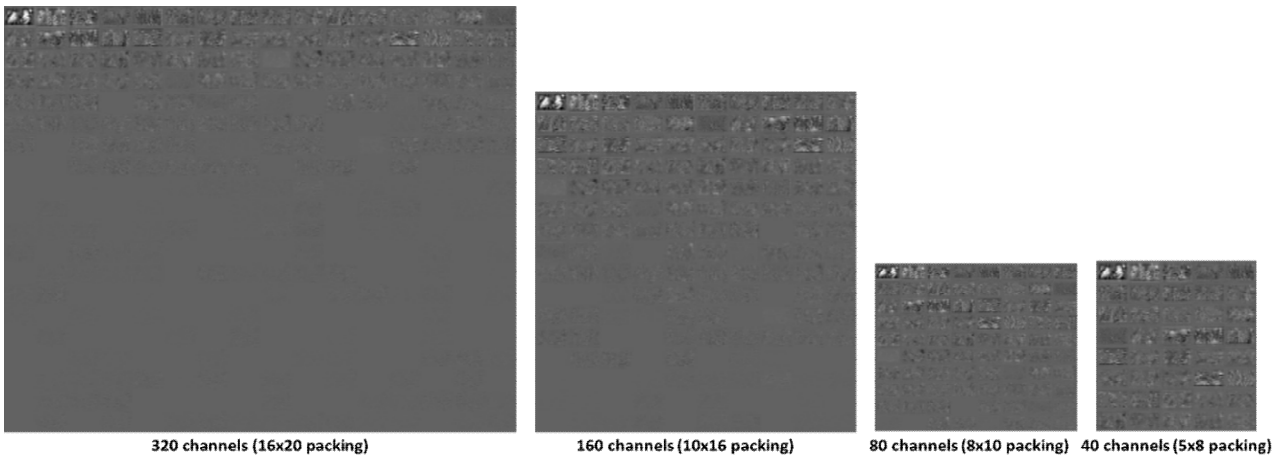


그림 8. SLS 기반의 QACT가 적용된 FCTM의 특징맵
Fig. 8. Feature map of FCTM with SLS based QACT

Ⅴ. 실험결과

FCM에서는 표 2와 같이 제안기술들의 성능평가를 위해 사용하는 머신비전 네트워크와 학습 및 평가 데이터셋(dataset) 등을 CTTC[9]으로 정의하고 있다. 본 논문의 제안 기법의 성능평가를 위한 모든 실험은 FCM의 CTTC에 따라 학습과 추론을 진행하였다.

학습 단계에서는 인스턴스 분할 및 객체 탐지 임무에서는 OpenImage V6 학습 데이터셋을 사용하였고 객체 추적 임무에서는 OpenImage V6와 PedtrackPP 데이터셋을 함께 활용하여 학습을 진행하였다. 손실함수는 학습 가능한 내부 코덱(learnable inner codec)에서 계산된 비트율과 평균 제곱오차(MSE: Mean Squared Error)로 측정된 왜곡(distortion)을 기반으로 한 율-왜곡(rate-distortion) 함수로

표 2. 분할지점 모델별 학습 정보
Table 2. Training information by model at split points

Task	Instance segmentation	Object detection	Object tracking
Batch size	1	1	1
Training time	About 5 days	About 5 days	About 5 days
Training sets	Pre-trained Mask R-CNN model Detectron2[10] (model id: 139653917) Dataset: OpenImage V6 train dataset	Pre-trained Faster R-CNN model Detectron2 (model id: 39173657) Dataset: OpenImage V6 train dataset	Pre-trained jde-1088x608 model[11] (jde.1088x608. uncertainty.pt) Dataset: OpenImage V6 and PedtrackPPtrain dataset
Number of epoch	21 epochs	30 epochs	21 epochs
Learning rate	Initial learning rate: 1e-4	Initial learning rate: 1e-4	Initial learning rate: 1e-4
Optimizer	Adam	Adam	Adam
Loss function	Rate-Distortion(MSE) loss	Rate-Distortion(MSE) loss	Rate-Distortion(MSE) loss

표 3. FCM의 공통실험조건(CTTC)
Table 3. FCM CTTC

Task	Train Dataset	Test Dataset	Network Architecture
Instance segmentation	OpenImage-V6 (train)	OpenImage-V6[12] (5k)	Mask R-CNN X101-FPN[10]
Object detection		OpenImage-V6 (5k)	Faster R-CNN X101-FPN[10]
		SFU-HW[13] (video)	Faster R-CNN X101-FPN
Object tracking	OpenImage-V6 (train), Pexel & Pixabay datasets	TVD[14] (video)	JDE-1088x608[11]
		HiEve[15] (video)	JDE-1088x608

정의된다. FCTM의 학습은 학습 가능한 엔트로피 모델을 내부 코덱으로 사용하여 종단간 방식으로 FENet과 DRNet을 학습하는 1차 학습과 VVC 압축 손실을 보상하기 위해서 DRNet을 미세조정하는 2차 학습으로 구성된다. 본 논문에서 제안방법은 1차 학습만으로 진행하였다. 추론 과정에서는 SLS와 QACT를 FCTM 구조에 통합하여 CTTC에 따라 실험을 수행하였으며, 전체 실험은 CTTC에 따라 진행되었다.

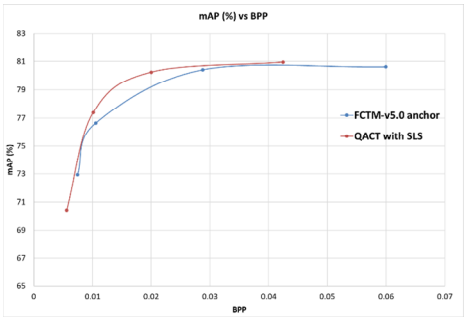
표 4는 제안기법의 성능을 FCTM v5.0와 비교한 성능을 나타낸 것이다. 표 4와 같이 제안기법은 기존의 FCTM v5.0 보다 모든 임무의 모든 데이터셋에서 성능 개선을 보였고, 전체적으로 앵커 대비 평균 22.90%의 BD-rate 이득이 있음을 보였다. 개별 기술에 대한 성능향상을 확인한 것으로 제안기법 중 SLS만 적용한 성능은 표 5와 같다. 표 5와 표 4를 비교하였을 때, 제안기법 중 SLS에 의한 성능향상이 큰 것을 확인할 수 있다. 그림 9는 각 임무 및 테스트 데이터셋에 대한 표 4에 보인 압축성능을 율-성능(RP: Rate-Performance) 곡선으로 나타낸 것이다. 일부 데이터셋에서 RP 곡선의 교차가 있지만 전반적으로 기존 FCTM의 성능을 확연히 개선함을 확인할 수 있다.

표 4. 제안기법의 FCTM v5 대비 압축성능(BD-rate)
Table 4. BD-rate of the proposed method compared to FCTM v5

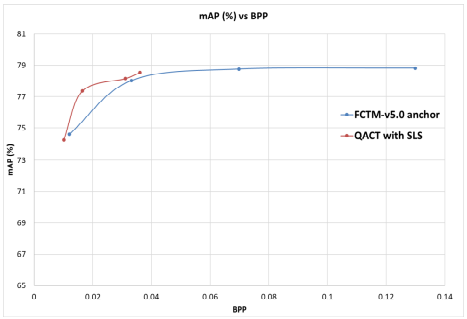
Task	Dataset	BD-rate
Object segmentation	OIV6	-14.45%
Object detection	OIV6	-20.67%
	SFU Class A/B	-63.32%
	SFU Class C	-27.95%
	SFU Class D	-37.16%
Object tracking	TVD	-8.85%
	HiEve 1080p	-5.59%
	HiEve 720p	-5.20%
	Overall	-22.90%

표 5. 제안기법 중 SLS만의 FCTM v5 대비 압축성능(BD-rate)
Table 5. BD-rate of the proposed SLS-only compared to FCTM v5

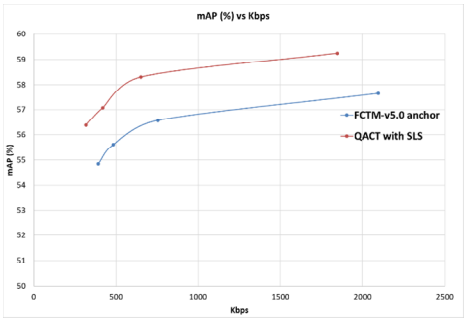
Task	Dataset	BD-rate
Object segmentation	OIV6	-10.75%
Object detection	OIV6	-24.91%
	SFU Class A/B	-67.48%
	SFU Class C	-26.61%
	SFU Class D	-33.15%
Object tracking	TVD	-4.72%
	HiEve 1080p	-2.60%
	HiEve 720p	-4.20%
	Overall	-21.80%



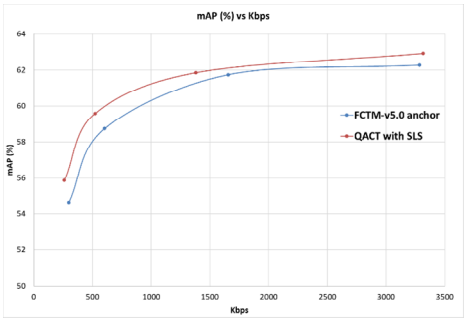
(a) OpenImage Segmentation



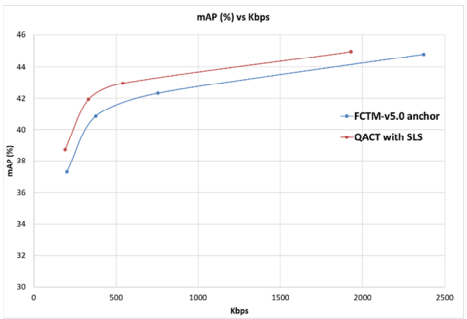
(b) OpenImage Detection



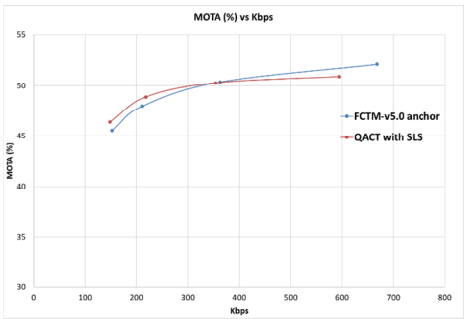
(c) SFU Detection Class A/B



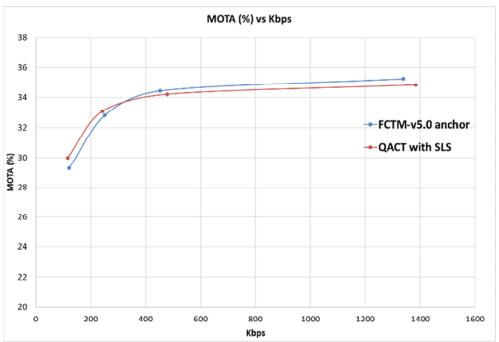
(d) SFU Detection Class C



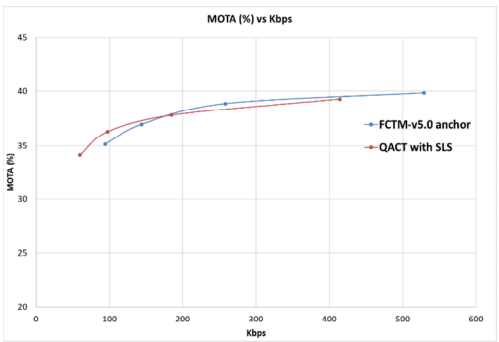
(e) SFU Detection Class D



(f) TVD Tracking overall



(g) HiEve Tracking 1080p



(h) HiEve Tracking 720p

그림 9. 제안기법의 FCTM v5.0 대비 압축성능

Fig. 9. Compression performance of the proposed method compared to FCTM v5.0

Ⅵ. 결 론

본 논문에서는 선택적 학습기법(SLS)과 QP-적응적 채널 절삭(QACT) 기법을 적용한 FCTM의 압축성능 개선기법을 제시한다. SLS를 적용하여 VVC로 부호화되는 특징맵을 특징 채널별 중요도 순서대로 정렬함으로써 에너지를 공간적으로 집중시키면서 공간적 상관성을 높여 특징맵 압축 효율을 개선한다. 또한, QACT를 통하여 SLS로 정렬된 특징 채널을 중요도가 떨어지는 채널부터 절삭함으로써 주어진 비트율에 최적의 크기로 특징맵 해상도를 조절 가능하게 한다. 이를 통하여 SLS와 QACT 기법은 단일 학습모델로 넓은 범위의 비트율에서 FCTM의 압축성능을 확연히 개선하였다. 제안방법은 기존의 FCTM v5.0 대비 최대 63.32%, 전체적으로 22.90%의 BD-rate 비트율 절감의 부호화 성능향상을 보였다. 제안방법은 FCM의 핵심실험(CE) 검증을 통하여 SLS는 FCM의 학습기법으로 CTTC에 채택되었다. 향후, QACT에서 절삭되는 특징 채널 수와 비트율의 관계를 모델링하여 좀 더 적절한 특징 채널 수를 보다 체계적으로 구하는 방법 등의 추가 개선을 통하여 SLS 기반의 채널 제거기법으로 향후 표준기술로 고려될 수 있을 것으로 기대된다.

참 고 문 헌 (References)

[1] C. Rosewarne and Y. Zhang, “AHG on Feature Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 4, m70669, Jan. 2025.

[2] “Algorithm description of FCTM,” ISO/IEC JTC 1/SC 29/WG 04 N00623, Jan. 2025.

[3] “FCM CE 3 NN-based inner coding,” ISO/IEC JTC 1/SC 29/WG 04 N00629, Jan. 2025.

[4] “FCM CE 4 Feature conversion,” ISO/IEC JTC 1/SC 29/WG 04 N00630, Jan. 2025.

[5] Y. -U. Yoon, D. Kim, J. Lee, B. T. Oh, and J. -G. Kim, “An Efficient Multi-Scale Feature Compression With QP-Adaptive Feature Channel Truncation for Video Coding for Machines,” IEEE Access, vol. 11, pp. 92443-92458, 2023.
doi: https://doi.org/10.1109/ACCESS.2023.3307404

[6] Y. -U. Yoon, G. -W. Han, J. Lee, S. Y. Jeong and J. -G. Kim, “An Advanced Multi-Scale Feature Compression using Selective Learning Strategy for Video Coding for Machines,” in Proc. IEEE Int. Conf. Visual Comm. Image Process. (VCIP 2023), Jeju, Korea, 2023, pp. 1-5.
doi: https://doi.org/10.1109/VCIP59821.2023.10402770

[7] Versatile Video Coding (VVC), ISO/IEC 23090-3, 2020.

[8] G.-W. Han, I.-K. Yoo, J.-G. Kim, J. Lee, Y. Kim, S. Jeong, “[FCM] CE4.2.1: QP-Adaptive Channel Truncation with Selective Learning Strategy,” ISO/IEC JTC 1/SC 29/WG 4, m71202, Jan. 2025.

[9] “Common test and training conditions for FCM,” ISO/IEC JTC 1/SC 29/WG 04 N00625, Jan. 2025.

[10] Facebook, 2019, “Detectron2,”. https://github. com/facebookresearch/detectron2

[11] Wang, Z., Zheng, L., Liu, Y., Li, Y., Wang, S. (2020). Towards Real-Time Multi-Object Tracking. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12356. Springer, Cham.
doi: https://doi.org/10.1007/978-3-030-58621-8_7

[12] A. Kuznesova, et al, “The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale,” Int. J. Comput. Vis., vol. 128, no. 7, pp. 1956-1981, Nov. 2020.

[13] T. Tanaka, H. Choi, and I. V. Bajic, “SFU-HW-Tracks-v1: Object Tracking Dataset on Raw Video Sequences,” 2021. arXiv 2112.14934. https://arxiv.org/abs/2112.14934

[14] X. Xu, S. Liu, and Z. Li, “Tencent Video Dataset (TVD): A Video Dataset for Learning-based Visual Data Compression and Analysis,” arXiv: 2105.05961. https://arxiv.org/abs/2105.05961

[15] Lin, W., Liu, H., Liu, S. et al. HiEve: A Large-Scale Benchmark for Human-Centric Video Analysis in Complex Events. Int J Comput Vis 131, 2994–3018(2023).
doi: https://doi.org/10.1007/s11263-023-01842-6

[16] H. Han, H. Choi, S. -H. Jung, J. Y. Lee, S. Kwak, W. -S. Cheong, and H. -G. Choo “[FCM] CE4.3.5a report” ISO/IEC JTC 1/SC 29/WG 4, m71342, Jan. 2025.

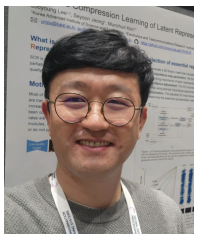
저 자 소 개

한 규 웅

- 2023년 2월 : 한국항공대학교 항공전자정보공학부 학사
- 2025년 2월 : 한국항공대학교 항공전자정보공학과 석사
- ORCID : https://orcid.org/0009-0009-5711-781X
- 주관심분야 : 기계를 위한 비디오 부호화, 비디오 부호화, 딥러닝, 영상처리

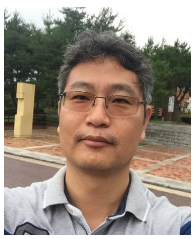
유 인 근

- 2023년 2월 : 한국항공대학교 정보통신학과 학사
- 2024년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학과 석사과정
- ORCID : https://orcid.org/0009-0003-4541-555X
- 주관심분야 : 기계를 위한 부호화, 비디오 부호화, 이미지 처리



이 주 영

- 2003년 : 아주대학교 미디어학부 학사
- 2006년 : KAIST 전산학부 석사
- 2024년 : KAIST 전기및전자공학부 박사
- 2006년 ~ 현재 : 한국전자통신연국원 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0003-0753-0699
- 주관심분야 : 인공지능, 컴퓨터 비전, 생성 모델, 이미지/비디오 압축



정 세 윤

- 1995년 : 인하대학교 전자공학과 학사
- 1997년 : 인하대학교 전자공학과 석사
- 2014년 : KAIST 전기및전자공학과 박사
- 1996년 ~ 현재 : ETRI 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0002-1675-4814
- 주관심분야 : 실감 방송, 비디오 코딩, 컴퓨터 비전

김 재 곤

- 1990년 : 경북대학교 전자공학과 학사
- 1992년 : KAIST 전기 및 전자공학과 석사
- 2005년 : KAIST 전자전산학과 박사
- 1992년 ~ 2007년 : 한국전자통신연구원(ETRI) 선임연구원/팀장
- 2001년 ~ 2002년 : Columbia University, NY, 연구원
- 2015년 ~ 2016년 : UC San Diego, Visiting Scholar
- 2007년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
- ORCID : https://orcid.org/0000-0003-3686-4786
- 주관심분야 : 비디오 신호처리, 비디오 부호화, 이머시브 비디오