특집논문 (Special Paper)
방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)
https://doi.org/10.5909/JBE.2022.27.3.273
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

인공지능 기반 멀티태스크를 위한 비디오 코덱의 성능평가 방법

김 신[a], 이 예 지[a], 윤 경 로[a)‡], 추 현 곤[b], 임 한 신[b], 서 정 일[b)]

Evaluation of Video Codec AI-based Multiple tasks

Shin Kim[a)], Yegi Lee[a)], Kyoungro Yoon[a)‡], Hyon-Gon Choo[b)], Hanshin Lim[b)], and Jeongil Seo[b)]

요 약

MPEG 내 VCM 그룹은 머신을 위한 비디오 코덱을 표준화하는 것으로 목표로 하고 있다. VCM 그룹은 객체 탐지, 객체 분할, 객체 추적 등 3가지의 머신비전 태스크를 포함한 데이터 세트와 데이터 세트 별 기준 데이터인 Anchor를 제공하고 있으며, 평가 템플릿을 이용하여 후보 기술군과 Anchor의 압축 대비 머신비전 성능을 비교할 수 있다. 하지만 성능 비교는 머신비전 태스크 별로 분리하여 수행되고 있으며, 다수의 머신비전 태스크에 대한 성능 평가를 수행할 수 있는 비트스트림을 생성할 수 있는 데이터는 별도로 제공하고 있지 않다. 본 논문에서는 인공 지능 기반 멀티 태스크를 위한 비디오 코덱의 성능 평가 방안에 대해 제안한다. 하나의 비트스트림의 크기 척도인 픽셀 당 비트수(BPP, Bits Per Pixel) 와 각 태스크의 정확도 결과인 Mean Average Precision(mAP)를 기반으로 산술 평균, 가중 평균, 조화 평균 등 총 3가지의 멀티 태스크 성능 평가 지표를 제안하며 mAP 결과를 기반으로 성능 결과를 비교하고자 한다. 멀티 태스크에서 태스크 별 mAP 결과 값의 범위의 차이가 있을 수 있으며 차이로 인해 생길 수 있는 성능 평가와 관련된 문제를 방지하고자 정규화한 mAP 기반 멀티 태스크 성능 결과를 산출하고 평가하고자 한다.

Abstract

MPEG-VCM(Video Coding for Machine) aims to standardize video codec for machines. VCM provides data sets and anchors, which provide reference data for comparison, for several machine vision tasks including object detection, object segmentation, and object tracking. The evaluation template can be used to compare compression and machine vision task performance between anchor data and various proposed video codecs. However, performance comparison is carried out separately for each machine vision task, and information related to performance evaluation of multiple machine vision tasks on a single bitstream is not provided currently. In this paper, we propose a performance evaluation method of a video codec for AI-based multi-tasks. Based on bits per pixel (BPP), which is the measure of a single bitstream size, and mean average precision(mAP), which is the accuracy measure of each task, we define three criteria for multi-task performance evaluation such as arithmetic average, weighted average, and harmonic average, and to calculate the multi-tasks performance results based on the mAP values. In addition, as the dynamic range of mAP may very different from task to task, performance results for multi-tasks are calculated and evaluated based on the normalized mAP in order to prevent a problem that would be happened because of the dynamic range.

Keyword : Video Coding for Machine, Machine Vision, Multiple Machine Vision Tasks, Multi-tasks Performance evaluation

Ⅰ. 서 론

인공지능 알고리즘의 발전은 현재 다양한 분야의 산업에 큰 영향을 주고 있고, 시장에서 인공지능을 탑재한 전자 제품은 이제 어렵지 않게 접할 수 있게 되었다. 하지만, 인공지능 기술 중 특히 머신비전 기술은 사람의 시각 체계를 근간으로 하여 개발되었으며, 사람의 시각 체계는 또한 전통적인 비디오 코덱의 근간이 되고 있다. 전통적인 비디오 코덱은 사람의 시각적 인식 체계에서 차이의 인지가 힘든 한도 내에서 비디오를 압축하는 데 중점을 두고 있다. 하지만, 인공지능 알고리즘을 기반으로 한 제품이 기하급수적으로 증가하면서 미디어와 같은 무거운 데이터를 통신하거나 저장하는 데 있어 과도한 네트워크 사용 또는 과도한 저장 비용 문제 등 다른 문제를 유발하게 되었고, 이는 사람의 시각을 위한 것이 아닌 기계를 위한 비디오 압축의 당위성을 강조하게 되는 동기가 되었다.

MPEG(Moving Picture Experts Group) 내 VCM(Video Coding for Machine) 그룹은 사람의 시각이 아닌 기계의 시각에 초점을 맞춰 이미지와 비디오를 압축한 비트스트림을 정의하는 것을 표준의 범위로 잡고 있으며, 다양한 종류의 머신비전 태스크에 사용될 수 있는 비트스트림의 비트레이트 관점에서 효율적으로 압축하는 것을 목표로 하고 있다. 하지만 VCM 그룹은 현재까지 비교군에 해당하는 VVC[1] 기반의 비트스트림인 Anchor 데이터와 새로운 코덱의 다양한 후보들과의 비교를 위한 템플릿만을 제공하고 있으며, 새로운 비디오 코덱에 대한 멀티 태스크와 관련하여 추가적인 성능 평가 척도에 대하여 고려하고 있지 않다. 따라서 본 논문에서는 하나의 비디오 코덱을 기반으로 생성한 비트스트림을 기반으로 다수의 머신비전 태스크에 대한 성능을 평가하는 방법에 대하여 제시하고자 한다. 단일 비트스트림에 대해 여러 머신비전 태스크를 수행하고 태스크별 결과를 활용하여 멀티 태스크에 대한 단일 결괏값을 산출하여 제안하는 성능 평가 결과 방법에 대해 분석하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서 본 논문의 배경지식인 VCM의 현황 및 후보 기술군에 대해 요약하며, 제 3장에서는 VCM 비디오 코덱의 멀티 태스크 성능 평가를 위한 방안에 대하여 제안하며, 제 4장에서는 3장에서 제안한 방법을 토대로 멀티 태스크에 대한 성능을 평가하고, 평가 결괏값을 분석한다. 마지막으로 제5장에서는 본 논문의 결론을 짓는다.

Ⅱ. 배경지식

1. Video Coding for Machine

VCM 그룹은 사람의 시각 인지가 아닌 기계를 위한 비디오 코덱의 필요성에 의해 구성된 그룹이다. 인공지능은 전통적인 비디오 코덱 기반의 이미지를 기반으로 수행되나, 감지기의 양이 많아지면서 데이터의 양도 그에 따라 증가하여 지연시간 등에 대한 문제가 야기되었으며, 또한 전송이나 저장 등에 대한 추가적인 고려도 필요하게 되었다. VCM 표준 그룹의 목표는 머신비전에 활용하기 위한 비트스트림 포맷을 정의하는 것으로, 인간의 시각 체계만을 고려하여 비디오를 압축하는 전통적인 비디오 코덱과 달리 VCM 코덱의 주된 대상은 머신비전이다.

VCM 그룹은 VCM 코덱이 다수의 머신비전 태스크를 지원하는 것을 목표로 하고 있으며, 현재 VCM 그룹 내 공통 실험 조건[2]에 해당하는 머신비전 태스크는 객체 탐지(Object Detection), 객체 분할(Instance Segmentation), 객체 추적(Object Tracking) 등 크게 3가지가 있으며, 이를 실험할 수 있는 데이터 세트로 FLIR[3], TVD(Tencent Video Dataset)[4], OpenImageV6[5], SFU-HW-Objects-v1[6] 등이 있다. 제공되는 데이터 세트 중 하나의 비트스트림에 대해 다수의 머신비전 태스크를 수행할 수 있는 데이터 세트는

a) 건국대학교 컴퓨터공학과(Dept. of Computer Science and Engineering, Konkuk University)
b) 한국전자통신연구원 실감미디어연구실(Immersive Media Research Laboratory, Electronics and Telecommunications Research Institute)
‡ Corresponding Author : 윤경로(Kyoungro Yoon)
E-mail: yoonk@konkuk.ac.kr
Tel: +82-2-450-4129
ORCID: https://orcid.org/0000-0002-1153-4038
※ 본 연구 논문은 과학기술정보통신부 및 정보통신기획평가원의 출연금으로 수행되고 있는 한국전자통신연구원 “기계를 위한 영상 부호화 기술 개발”(2020-0-00011)의 연구결과입니다.
※ This work was supported by Institute of Information & Communication Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT).
· Manuscript April 12, 2022; Revised May 6, 2022; Accepted May 6, 2022.

현재 TVD 데이터 세트만 해당하며, TVD 데이터 세트는 객체 탐지와 객체 분할에 대한 검증 자료를 제공하고 있다.

VCM 그룹은 새로운 기술의 머신을 위한 비디오 코덱에 대한 평가 템플릿[6]을 제공하고 있으나, 현재 각 머신비전 태스크 별로 후보 코덱에 대한 영상 압축 성능과 해당 머신비전 태스크에 대한 정확도만을 평가하고 있다. 즉, 단일 태스크에 대하여 영상 압축 성능 및 머신비전 태스크의 퍼포먼스를 측정하고 있기 때문에 하나의 비트스트림 또는 하나의 코덱에 기반을 둔 멀티 태스크 성능에 대한 새로운 종합적 평가 방법에 관한 연구가 필요한 상태이다.

2. Evidence

133차 MPEG 국제 표준화 회의에서 VCM 그룹은 Call for Evidence[7]를 공표하였으며, 134차 MPEG 회의 때 총 4건의 Response for Evidence가 기고되었고 이 중 2건이 의미 있는 Evidence로 채택되었다. 후보 기술 중 하나는 중국 저장대에서 기고한 이미지 코덱[8]으로 재훈련한 Cheng2020 네트워크로 end-to-end 압축하여 Anchor 대비 약 22.80%의 BD-rate 효율을 얻어냈다. 또한 다른 후보 기술[9]은 건국대와 ETRI에서 공동 기고한 이미지 코덱으로 객체 탐지 네트워크를 이용하여 객체와 배경을 분리하여 상이한 양자화 파라미터를 이용하여 별도로 압축하는 방식으로 Anchor 대비 약 30.76%의 BD-rate 효율을 성취했다.

Ⅲ. 멀티 태스크 평가 방법 제안

1. 멀티 태스크 척도 정의

VCM 그룹에서는 단일 태스크에 대한 성능 평가 기본 척도로 BPP(Bits per pixel)과 mAP(mean Average Precision)를 선정하였으나, 현재 멀티 태스크에 대한 성능 지표는 지정되어 있지 않다. 따라서 머신을 위한 비디오 코덱을 이용하여 압축한 하나의 비트스트림을 기반으로 한 멀티 태스크에 대한 성능을 평가하기 위하여 새로운 성능 평가 방법이 필요하다.

다수의 머신비전 태스크에 대해 단일 성능 점수를 표현할 수 있도록 평균을 도입하며, 모든 태스크에 대한 mAP 기반으로 산술 평균(Arithmetic Average), 가중 평균(Weighted Average), 조화 평균(Harmonic Average) 등 3가지 평균 점수를 통해 n개의 멀티 태스크에 대한 종합적 성능을 표현하고자 한다. 각 평균에 대한 수식은 수식 (1)과 같다.

이때 mAP_{task_n}는 n번 태스크에 대한 정확도이다.

현재 VCM 내 하나의 비트스트림으로 멀티 태스크에 대한 성능 평가가 가능한 데이터 세트는 TVD 이미지 데이터 세트로 객체 탐지와 객체 분할 등 2가지의 머신비전 태스크에 대한 Ground Truth를 지원하며 이를 위한 산술 평균, 가중 평균, 조화 평균에 대한 수식은 수식 (2)와 같다.

$$
\begin{aligned}
&Arithmetic\,Average = \frac{(mAP_{task_1} + mAP_{task_2} + \cdots)}{n} \\
&Weighted\,Average = (mAP_{task_1} \times w_{task_1} + mAP_{task_2} \times w_{task_2} + \cdots + mAP_{task_n} \times w_{task_n}) \\
&\text{단, } w_{task_1} + w_{task_2} + \cdots + w_{task_n} = 1 \\
&Harmonic\,Average = \frac{n}{(\frac{1}{mAP_{task_1}} + \frac{1}{mAP_{task_2}} + \cdots + \frac{1}{mAP_{task_n}})}
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&Arithmetic\,Average = \frac{(mAP_{ObjDet} + mAP_{\in sSeg})}{2} \\
&Weighted\,Average = (mAP_{ObjDet} \times w_{ObjDet} + mAP_{\in \mathrm{sSeg}} \times w_{\in \mathrm{sSeg}}), \\
&\text{단, } w_{ObjDet} + w_{\in \mathrm{sSeg}} = 1 \\
&Harmonic\,Average = \frac{2}{(\frac{1}{mAP_{ObjDet}} + \frac{1}{mAP_{\in \mathrm{sSeg}}})}
\end{aligned}
\tag{2}
$$

이때 mAP_{ObjDet}는 객체 탐지에 대한 정확도, $mAP_{\in sSeg}$는 객체 분할에 대한 정확도, w_{ObjDet}는 객체 탐지 정확도에 대한 가중치, $w_{\in sSeg}$는 객체 분할 정확도에 대한 가중치이다.

하지만 머신비전 태스크에 따라 동일한 비트스트림이라 하더라도 태스크별 네트워크 차이, Ground Truth 차이 등으로 인해 각 머신비전 태스크 별 mAP 는 차이가 날 가능성이 있다. 가령, VCM 그룹에서 제공하는 TVD 데이터 세트의 Anchor는 동일한 비트스트림에 객체 탐지 및 객체 분할을 수행하였지만, 이미지 스케일 100% 및 양자화 파라미터를 22로 설정하여 VTM[10]으로 압축한 이미지에 대한 객체 탐지 mAP는 55.748%, 객체 분할 mAP는 45.004%로 태스크 간 10%의 mAP 값의 차이를 보여준다. 태스크별 mAP에 대해 데이터 가공을 하지 않은 상태로 평균을 구하는 경우, 예를 들어 한 태스크의 mAP 결괏값이 특출나게 높으며 나머지 태스크들의 mAP가 상대적으로 적을 때 산술 평균을 구하는 경우 평균 점수가 더 높게 나올 가능성이 있으며 이는 적절한 멀티 태스크 성능 평가 척도라고 보기 어려울 수 있다. 따라서 태스크별 mAP 결괏값을 정규화하는 과정이 필요하다. 각 머신비전 태스크의 mAP 는 0에서 1 사이의 값으로 산출된다. 하지만 각 태스크마다 산출되는 값의 범위가 다를 수 있으며 태스크별 mAP의 차이는 모든 태스크에 대한 평균 mAP를 구할 때 평균값에 영향을 끼칠 수 있다. 따라서 멀티 태스크 성능 평가 시 한 태스크에 대한 결과가 전체 결과에 주도적 영향을 미치는 것을 방지하기 위하여 각 태스크의 결괏값을 정규화할 필요성이 있다. 제안하는 각 태스크의 결괏값을 정규화하는 방법은 다음과 같다.

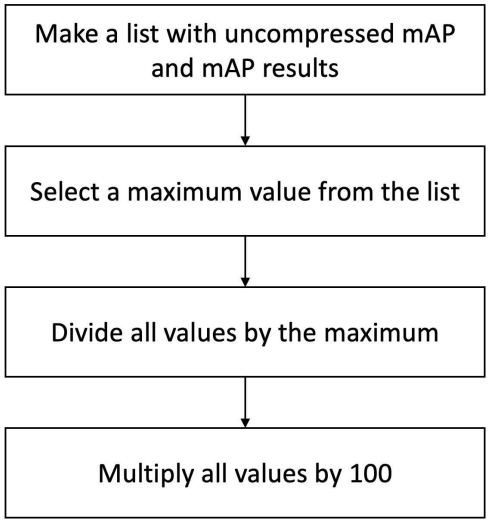

그림 1. mAP 결괏값 정규화 방법
Fig. 1. mAP results normalization method

점수를 정규화하기 위해, 먼저 압축하지 않은 원 영상에 대한 태스크 수행 결과 mAP 와 압축을 수행한 후의 영상에 대한 태스크 mAP 결괏값들로 리스트를 생성한다. 리스트 중에서 가장 큰 값을 최댓값으로 선정하여, 태스크 mAP 결괏값들을 최댓값으로 나누어 0과 1 사이의 값으로 정규화를 하며 나눈 값에 100을 곱하여 0에서 100 사이의 값으로 mAP의 범위를 치환하며, 치환된 값으로 산술 평균, 가중 평균, 조화 평균을 구하여 멀티 태스크의 성능을 측정한다.

Ⅳ. 성능 평가 결과

1. 멀티 태스크 성능 평가 환경

멀티 태스크 성능 평가를 위해 표 1 및 표 2와 같은 네트워크 및 환경에서 실험을 진행하였다.

표 1. 태스크별 사용 네트워크
Table 1. Network used for each task

Task	Network
Object detection	Faster R-CNN X101-FPN[11][12]
Instance Segmentation	Mask R-CNN X101-FPN[13]

표 2. 실험 환경 설정
Table 2. Experimental environmental setting

Software/Framework/Language	version
CUDA	11.1
Pytorch	1.9
Python	3.7/3.8
Detectron2	0.5
Tensorflow	2.7
FFMPEG	4.2.2

2. 멀티 태스크 성능 평가 결과

표 3은 TVD anchor[14,15] 중 스케일 100% 및 75%에 대한

표 3. TVD anchor[14,15]에 대한 머신비전 태스크별 실험 결과
Table 3. Machine vision task results on TVD anchor[14,15]

Image Dataset			Test Task 1 Object Detection				Test Task 2 Instance Segmentation			
Scale	Dataset	QP	BPP	mAP	Normalized mAP	Weight	BPP	mAP	Normalized mAP	Weight
100%	TVD	22	0.475	55.748	99.404	0.6	0.475	45.005	99.645	0.4
		27	0.266	53.752	95.845		0.266	43.822	97.026	
		32	0.145	50.632	90.281		0.145	40.746	90.215	
		37	0.075	45.311	80.793		0.075	36.796	81.470	
		42	0.037	38.586	68.802		0.037	29.091	64.410	
		47	0.017	20.155	35.938		0.017	18.346	40.620	
75%	TVD	22	0.310	55.913	99.698	0.6	0.310	44.163	97.781	0.4
		27	0.179	52.074	92.853		0.179	41.945	92.870	
		32	0.098	49.988	89.133		0.098	37.950	84.025	
		37	0.051	42.668	76.081		0.051	33.226	73.565	
		42	0.025	28.295	50.452		0.025	21.454	47.501	
		47	0.012	14.061	25.072		0.012	13.167	29.153	

멀티 태스크 성능 평가를 위한 BPP, COCO[16] 기반의 mAP, 정규화한 mAP, 태스크 별 가중치를 보여준다. 태스크 별 가중치는 객체 탐지 0.6 대 객체 분할 0.4로서, 임의로 설정된 값이며 머신비전 태스크의 중요도에 따라 변경될 수 있다.

표 4는 2장에서 언급한 후보 기술 중 하나로서 VCM의 Evidence로 채택된 기술[9]인 객체 및 배경 분할 기반 비디오 코덱을 활용한 멀티 태스크 결과를 보여주며, TVD 데이터 세트 중 스케일 100% 및 75%에 대한 BPP, 태스크 결과 mAP, 정규화한 mAP 및 태스크 별 가중치를 보여준다.

표 5는 표 3을 기반으로 하여 TVD anchor의 멀티 태스크 성능을 산출한 결과를 나타낸다. Multiple Tasks Results의 경우 정규화하지 않은 본래의 mAP 값으로 산출하였으며, 앞 절에서 정의한 산술 평균, 가중 평균, 조화 평균 등 총 3가지의 성능 평갓값을 보여준다. 가중 평균의 경우, 표 3에서 설정한 객체 탐지에 대한 가중치 0.6과 객체 설정에 대한 가중치 0.4로 설정하여 계산하였다. Normalized Multiple Tasks Results의 경우 정규화한 mAP 값을 이용하여 산출하였으며, 동일하게 세 가지의 평균값을 보여준다.

표 4. TVD에 대한 Encoder 1[9]의 머신비전 태스크별 실험 결과
Table 4. Machine vision task results of Encoder 1[9] on TVD dataset

Image Dataset			Test Task 1 Object Detection				Test Task 2 Instance Segmentation			
Scale	Dataset	QP	BPP	mAP	Normalized mAP	Weight	BPP	mAP	Normalized mAP	Weight
100%	TVD	22	0.209	54.585	97.329	0.6	0.209	44.291	98.065	0.4
		27	0.118	52.023	92.762		0.118	43.094	95.413	
		32	0.064	47.667	84.994		0.064	37.916	83.950	
		37	0.035	41.579	74.140		0.035	33.470	74.106	
		42	0.019	30.302	54.030		0.019	23.620	52.297	
		47	0.010	17.853	31.834		0.010	15.368	34.026	
75%	TVD	22	0.166	54.211	96.664	0.6	0.166	43.656	96.658	0.4
		27	0.094	50.376	89.825		0.094	40.366	89.374	
		32	0.051	43.745	78.001		0.051	36.087	79.900	
		37	0.027	37.740	67.293		0.027	29.854	66.100	
		42	0.015	24.280	43.293		0.015	18.727	41.463	
		47	0.008	13.113	23.382		0.008	11.370	25.175	

표 5. TVD anchor의 멀티 태스크 성능 산출 결과
Table 5. Multi-tasks performance results on TVD anchor

TVD Image Dataset			Multiple Task Results			Normalized Multiple Tasks Results		
Scaling Factor	QP	BPP	Arithmetic Average	Weighted Average	Harmonic Average	Arithmetic Average	Weighted Average	Harmonic Average
100%	22	0.475	50.377	51.451	49.804	99.524	99.500	99.524
	27	0.266	48.787	49.780	48.282	96.435	96.317	96.432
	32	0.145	45.689	46.678	45.154	90.248	90.255	90.248
	37	0.075	41.054	41.905	40.612	81.132	81.064	81.130
	42	0.037	33.839	34.788	33.172	66.606	67.045	66.534
	47	0.017	19.251	19.431	19.208	38.279	37.811	38.136
75%	22	0.310	50.038	51.213	49.348	98.739	98.931	98.730
	27	0.179	47.010	48.022	46.464	92.861	92.860	92.861
	32	0.098	43.969	45.173	43.145	86.579	87.090	86.504
	37	0.051	37.947	38.891	37.360	74.823	75.075	74.802
	42	0.025	24.875	25.559	24.404	48.977	49.272	48.932
	47	0.012	13.614	13.703	13.599	27.112	26.704	26.959

표 6은 표 4를 근간으로 하여 Encoder 1[9]의 TVD 데이터 세트에 대한 멀티 태스크 성능을 계산한 결과를 보여준다. Multiple Tasks Results의 경우, 표 5와 동일하게 산술 평균, 가중 평균, 조화 평균 등 총 세 가지의 점수를 보여주며, 가중 평균의 태스크 별 가중치는 6:4로 설정하였다. Normalized Multiple Tasks Results의 경우, 정규화한 mAP 값을 이용하여 3가지 지표의 값을 산출하였다.

그림 2는 표 3, 표 5를 근간으로 하여 그린 TVD anchor에 대한 멀티 태스크 결과 Pareto-Front 커브를 보여준다. 그림 2의 하단의 커브는 본래 값을 기반으로 한 커브이며, 상단의 커브는 정규화한 평균을 이용하여 그린 커브이다. 본래 값을 이용한 커브의 경우, 커브 간 미세한 차이를 보이는 것을 알 수 있으며 그 중 노란색 커브인 가중 평균 값이 가장 높은 값을 보여주는데, 그 이유는 TVD 데이터 세트에 대한 객체 탐지의 mAP의 결과가 객체 분할의 mAP보다 비교적 높으며, 이에 대한 가중치를 더 주었기 때문이다. 정규화한 값을 기반으로 그린 커브의 경우 세 가지 척도 간 차이가 육안으로 보이지 않는 것을 확인할 수 있다.

표 6. Encoder 1[9]의 멀티 태스크 성능 산출 결과
Table 6. Multi-tasks performance results of Encoder 1[9] on TVD dataset

TVD Image Dataset			Multiple Task Results/Precision			Multiple Tasks Results/Precision (Normalized)		
Scaling Factor	QP	BPP	Arithmetic Average	Weighted Average	Harmonic Average	Arithmetic Average	Weighted Average	Harmonic Average
100%	22	0.209	49.438	50.467	48.902	97.697	97.623	97.696
	27	0.118	47.558	48.451	47.139	94.088	93.823	94.069
	32	0.064	42.791	43.767	42.236	84.472	84.576	84.469
	37	0.035	37.525	38.336	37.087	74.123	74.126	74.123
	42	0.019	26.961	27.629	26.547	53.164	53.337	53.150
	47	0.010	16.610	16.859	16.517	32.930	32.710	32.893
75%	22	0.166	48.934	49.989	48.364	96.661	96.661	96.661
	27	0.094	45.371	46.372	44.819	89.600	89.645	89.599
	32	0.051	39.916	40.682	39.549	78.950	78.761	78.939
	37	0.027	33.797	34.586	33.337	66.697	66.816	66.691
	42	0.015	21.503	22.059	21.145	42.378	42.561	42.358
	47	0.008	12.242	12.416	12.180	24.279	24.100	24.246

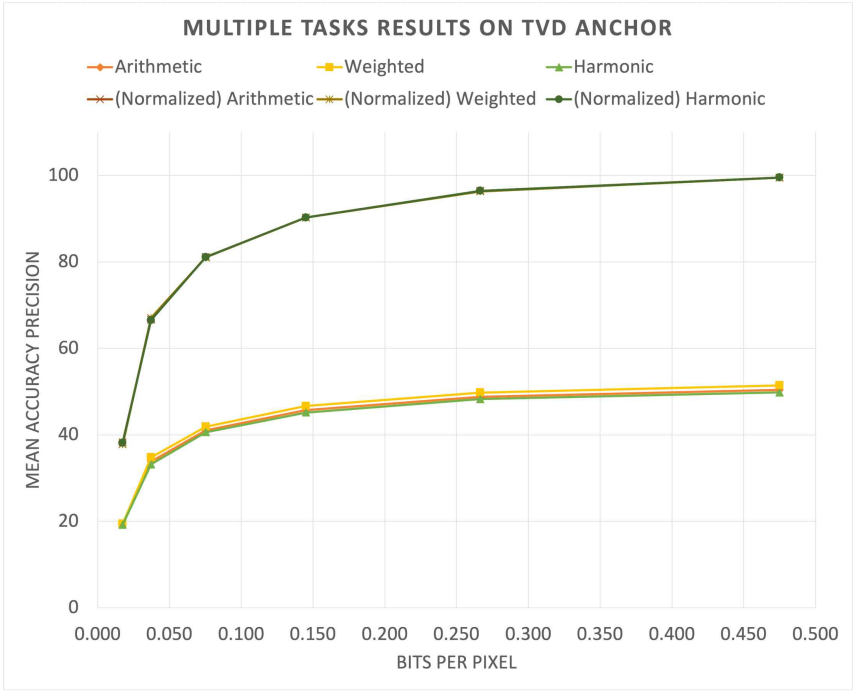

그림 2. TVD anchor[14]에 대한 멀티 태스크 결과 Pareto-Front 커브
Fig. 2. Multi-tasks results Pareto-Front curves on TVD anchor[14]

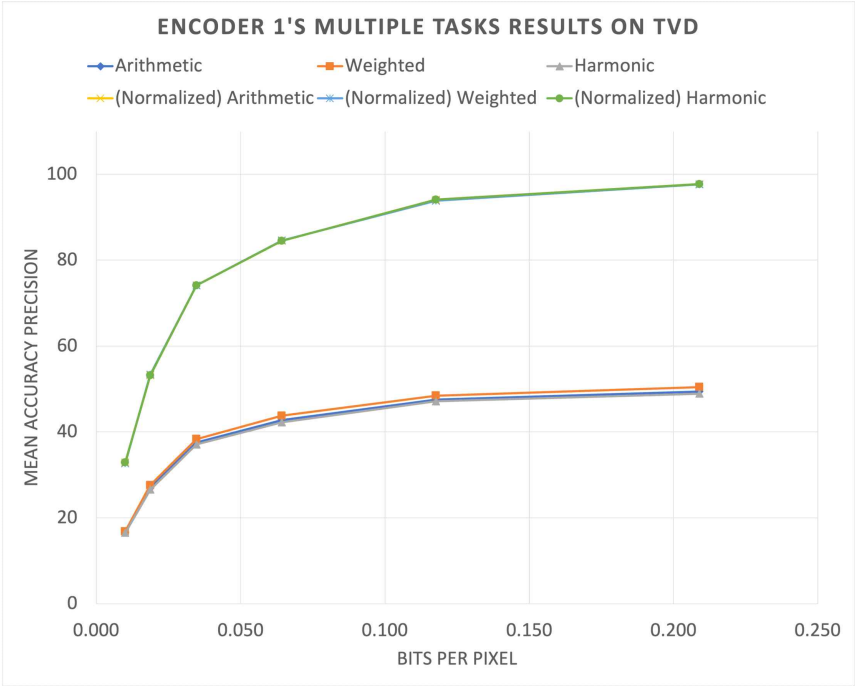

그림 3. Encoder 1[9]의 TVD에 대한 멀티 태스크 결과 Pareto-Front 커브
Fig. 3. Multi-tasks results Pareto-Front curves of Encoder 1[9] on TVD

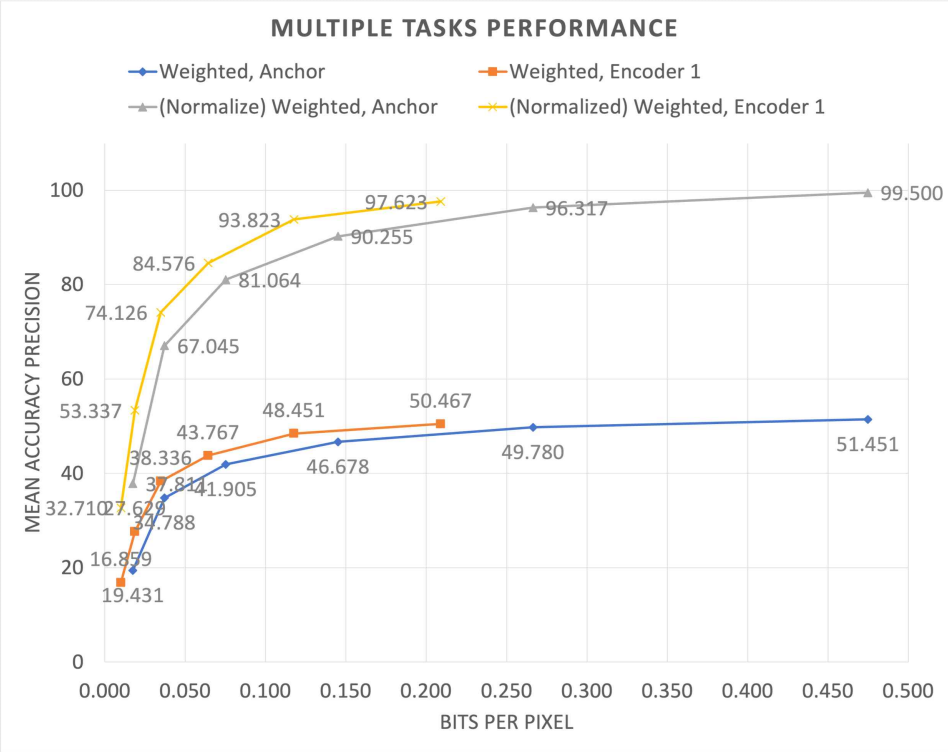

그림 4. TVD Anchor와 Encoder 1의 Arithmetic Average 기반 멀티 태스크 성능 비교
Fig. 4. Multi-tasks performance comparison based on arithmetic average of TVD Anchor and Encoder 1

그림 3은 표 3, 표 5를 근간으로 하여 그린 Encoder 1의 Pareto-Front 커브를 보여준다. 그림 3의 하단에 있는 커브는 본래 mAP 값을 이용하여 그린 커브이며, 상단에 있는 커브는 정규화한 평균 mAP 값을 이용하여 나타낸 커브이다. 그림 2와 유사하게, 본래 mAP 값을 기반으로 산출한 경우, 상당히 작긴 하지만 세 가지 척도의 커브 간 격차가 존재하며, 가중 평균치의 값이 그 중 가장 높은 값을 보이는 것을 확인할 수 있다. 정규화한 mAP 값을 기반으로 한 커브는 그림 2와 마찬가지로 모든 커브가 육안으로 구분이 불가할 정도로 동일한 선상에 위치하고 있는 것을 확인할 수 있다.

그림 4는 TVD anchor와 Encoder 1에 대한 산술 평균과 정규화한 산술 평균의 Pareto-Front 커브를 보여준다. 산술 평균 및 정규화한 산술 평균 모두 Encoder 1의 결과가 Anchor에 비해 동일한 BPP 에서 더 높은 mAP 값을 보여주며, 정규화한 산술 평균의 경우 커브 간 격차가 정규화하지 않은 경우에 비해 조금 더 벌어지는 것을 확인할 수 있다.

Ⅴ. 결 론

본 논문에서는 머신비전을 위한 비디오 코덱의 멀티 태스크 성능을 평가하는 방법에 대해 제안하고 이를 기반으로 머신비전 멀티 태스크의 성능을 평가하는 지표에 대해 분석하였다. 멀티 태스크 성능은 적절한 수식을 이용하여 하나의 척도로 표현될 수 있으며, 하나의 비디오 코덱에 대해 멀티 태스크 성능을 평가하고자 할 때, 본래 값을 이용하여 수치화하는 경우 수식에 따라 지표별로 사소한 차이는 있으나 거의 유사한 값이 산출되고 있으며, 정규화한 값을 이용하여 멀티 태스크에 대한 성능 점수를 나타내고자 하는 경우 Pareto-Front 커브 상 차이가 거의 없는 것을 확인하였다. 따라서, 제안하는 방법은 다수의 머신비전 태스크에 대한 압축 성능 별 mAP를 표현하고자 할 때 단일 값으로 표현할 수 있고 Anchor와의 비교를 통해 후보 기술군의 멀티 태스크에 대한 성능을 용이하게 평가할 수 있을 것으로 예상한다.

현재 VCM 그룹에는 TVD뿐만 아니라 FLIR, OpenImage, SFU 등 비디오 코덱의 평가를 위한 다양한 데이터 세트가 존재한다. 그러나 본 논문에서는 하나의 비트스트림에서 멀티 태스크 성능을 평가할 수 있는 TVD 데이터 세트에 한정하여 성능을 평가하였으며, 태스크 별 이질적인 데이터세트를 사용하는 경우를 고려한 멀티 태스크의 성능 평가 방법에 대해 추가로 고려해야 할 것으로 예상된다.

참 고 문 헌 (References)

[1] ISO/IEC 23090-3, "2021 Information technology - Coded representation of immersive media - Part 3:Versatile video Coding" https://www.iso.org/standard/73022.html

[2] ISO/IEC JTC 1/SC 29/WG 2, "Common Test Conditions and Evaluation Methodology for Video Coding for Machines," the 137th MPEG meeting, January 2022. https://dms.mpeg.expert/doc_end_user/documents/137_OnLine/wg11/MDS21288_WG02_N00163.zip

[3] Free FLIR Thermal dataset, https://www.flir.com/oem/adas/dataset/ (accessed Jan, 8, 2020).

[4] 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, May 2021. doi: https://doi.org/10.48550/arXiv.2105.05961

[5] Open Images V6, https://storage.googleapis.com/openimages/web/index. html (accessed Mar, 1, 2020)

[6] T. Takehiro, H. Choi, and I. V. Bajić. "SFU-HW-Tracks-v1: Object Tracking Dataset on Raw Video Sequences." arXiv preprint arXiv:2112.14934 , 2021. doi: https://doi.org/10.48550/arXiv.2112.14934

[7] ISO/IEC JTC 1/SC 29/WG 2, "Call for Evidence for Video Coding for Machines", the 133rd MPEG meeting, January 2021. https://dms.mpeg.expert/doc_end_user/documents/133_OnLine/wg11/MDS20126_WG02_N00042.zip

[8] B. Zhu, L. Yu, D. Li and Y. Pan, "[VCM] ZJU response to cfe: deep learning-based compression for machine vision ", the 134th MPEG meeting, April 2021. https://dms.mpeg.expert/doc_end_user/documents/134_OnLine/wg11/m56445-v3-m56445[VCM]ZJUresponsetocfe.zip

[9] Y. Lee, S. Kim, K. Yoon, H. Lim, H. Choo, W. Cheong and J. Seo, "[VCM] Response to CfE: Object detection results with the FLIR dataset," the 134th MPEG meeting, April 2021. https://dms.mpeg.expert/doc_end_user/documents/134_OnLine/wg11/m56572-v1-m56572_v2.zip

[10] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, 28, 2015. https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

[11] VTM 12.0/VVCSoftware_VTM, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-12.0 (accessed April, 1, 2021) https://openaccess.thecvf.com/content_cvpr_2017/papers/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.pdf

[12] T. Y. Lin, P. Dollár, R. Girshick, He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.2117-2125, 2017. https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf

[13] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask r-cnn," In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017. https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48

[14] W. Gao, X. Xu, S. Liu and M. Qin, "[VCM] TVD dataset for Object Segmentation", the 135th MPEG meeting, July 2021.

[15] W. Gao, X. Xu and S. Liu "[VCM] Updated anchor results for object detection using TVD dataset", the 135th MPEG meeting, July 2021.

[16] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D and Zitnick, C. L, "Microsoft coco: Common objects in context," In European conference on computer vision, Springer, Cham, pp. 740-755, 2014.

저 자 소 개

김 신

- 2015년 8월 : 건국대학교 컴퓨터공학과 졸업(학사)
- 2017년 2월 : 건국대학교 컴퓨터공학과 졸업(석사)
- 2017년 3월 ~ 현재 : 건국대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0000-0001-8492-3758
- 주관심분야 : 영상처리, 인공지능, 컴퓨터 비전

저 자 소 개

이 예 지

- 2018년 2월 : 극동대학교 스마트모바일학과 졸업(학사)
- 2020년 2월 : 건국대학교 스마트ICT융합과 졸업(석사)
- 2020년 3월 ~ 현재 : 건국대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0000-0002-0292-160X
- 주관심분야 : 영상처리, 인공지능, 컴퓨터비전

윤 경 로

- 1987년 2월 : 연세대학교 전자전산기공학과 졸업(학사)
- 1989년 12월 : University of Michigan, Ann Arbor, 전기공학과 졸업(석사)
- 1999년 5월 : Syracuse University, 전산과학과 졸업(박사)
- 1999년 6월 ~ 2003년 8월 : LG전자기술원 책임연구원/그룹장
- 2003년 9월 ~ 현재 : 건국대학교 컴퓨터공학과/스마트ICT융합공학과 교수
- ORCID : https://orcid.org/0000-0002-1153-4038
- 주관심분야 : 스마트미디어시스템, 멀티미디어검색, 영상처리, 멀티미디어/메타데이터 처리

추 현 곤

- 1998년 2월 : 한양대학교 전자공학과 (공학사)
- 2000년 2월 : 한양대학교 전자공학과 (공학석사)
- 2005년 2월 : 한양대학교 전자통신전파공학과 (공학박사)
- 2005년 2월 ~ 현재 : 한국전자통신연구원 선임연구원
- 2015년 1월 ~ 2017년 1월 : 한국전자통신연구원 디지털홀로그래피연구실장
- 2017년 9월 ~ 2018년 8월 : Warsaw University of Technology, Poland, 방문 연구원
- ORCID : https://orcid.org/0000-0002-0742-5429
- 주관심분야 : Computer vision, 3D image and holography, 3D depth imaging, 3D broadcasting system

임 한 신

- 2004년 2월 : 연세대학교 전기전자공학부 (수학 부전공)(공학사)
- 2006년 2월 : 한국과학기술원 전기전자공학부 (공학석사)
- 2007년 9월 ~ 2007년 12월 : TU Berlin 방문연구원
- 2014년 2월 : 한국과학기술원 전기전자공학부(공학박사)
- 2004년 3월 ~ 현재 : 한국전자통신연구원 선임연구원
- ORCID : https://orcid.org/0000-0003-4829-2893
- 주관심분야 : 2D/3D Image Processing, Computer Vision, 3D Reconstruction and Modeling, VR/AR Technology

서 정 일

- 1994년 2월 : 경북대학교 전자공학과 (공학사)
- 1996년 2월 : 경북대학교 대학원 전자공학과 (공학석사)
- 1995년 8월 : 경북대학교 대학원 전자공학과 (공학박사)
- 1998년 2월 ~ 2000년 10월 : LG반도체 주임연구원
- 2010년 8월 ~ 2011년 7월 : 영국 Southampton University, ISVR 방문연구원
- 2000년 11월 ~ 현재 : 한국전자통신연구원 실감미디어연구실 실장
- ORCID : https://orcid.org/0000-0001-5131-0939
- 주관심분야 : 오디오 신호처리, 실감은향, 디지털방송, 멀티미디어 표준화