



특집논문 (Special Paper)

방송공학회논문지 제30권 제6호, 2025년 11월 (JBE Vol.30, No.6, November 2025)

<https://doi.org/10.5909/JBE.2025.30.6.942>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

VCM을 위한 학습기반 영상 전·후처리 필터 네트워크

곽 상 운^{a)†}, 정 성 문^{b)}, 고 종 환^{b)}, 정 순 흥^{a)}, 추 현 곤^{a)}

Learned Pre-filter and Post-filter Network for VCM

Sangwoon Kwak^{a)†}, Seongmoon Jeong^{b)}, Jong Hwan Ko^{b)}, Soon-heung Jung^{a)}, and Hyon-Gon Choo^{a)}

요 약

본 논문에서는 기계를 위한 영상 부호화를 위해, 머신비전 성능 대비 압축 효율을 최적화할 수 있는 학습기반의 영상 전·후처리 필터 네트워크를 제안한다. 제안된 필터 네트워크는 VCM 환경에 특화된 구조로 설계되었으며, 내부 인코딩 과정에서 발생하는 영상의 머신비전 품질 저하를 최소화하면서 압축 성능을 개선하기 위한 전처리 필터와 내부 디코딩된 영상의 머신비전 품질을 개선하기 위한 후처리 필터로 구성된다. 일반적으로 미분이 불가능한 상용 코덱을 대신하여, 미분이 가능하면서도 압축 왜곡을 효과적으로 모사할 수 있는 모사 코덱을 활용하여 필터 네트워크를 중단간 학습할 수 있도록 학습 구조를 설계하였다. 또한 MPEG VCM 표준화의 참조 소프트웨어인 VCM-RS에 효과적으로 적용하기 위해, VTM으로 압축된 영상 및 BDT 모듈을 포함하여 파인 튜닝을 수행하였다. 학습된 모델을 VCM-RS v0.11에 적용한 결과, 전·후처리 필터 각각 17%, 13%의 BD-rate 성능 개선을 확인하였다.

Abstract

In this paper, we propose a learning-based video pre- and post-processing filter network that optimizes compression efficiency with respect to machine vision performance for video coding for machines. The proposed filter network is designed with a structure specialized for the VCM (Video Coding for Machines) environment, and consists of a preprocessing filter to enhance compression efficiency while minimizing degradation of machine vision quality during the internal encoding process, and a postprocessing filter to improve the machine vision quality of internally decoded videos. To enable end-to-end training of the filter network, we adopt a differentiable surrogate codec that effectively simulates compression distortion, instead of using a conventional non-differentiable commercial codec. Furthermore, to apply the network effectively to VCM-RS, the reference software of MPEG VCM standardization, fine-tuning was performed using VTM-compressed videos and the BDT module. Experimental results on VCM-RS v0.11 demonstrate BD-rate performance improvements of 17% and 13% for the pre-filter and post-filter, respectively.

Keyword : Video Coding for Machines, Machine Visions, MPEG, Filter Network

I. 서론

영상 콘텐츠의 사용 환경이 인간 중심에서 기계 중심으로 전환되며, 영상 부호화 기술에도 새로운 요구가 대두되고 있다^[1]. 특히 자율주행, 스마트시티, 산업 감시 시스템 등에서는 영상 데이터를 인간이 시청하는 것이 아니라, 기계가 분석하고 이해하는 것이 중요하고, 이러한 환경에서는 시각적 품질에 초점을 맞추고 있는 전통적인 비디오 부호화 방식이 아닌 머신 비전(Machine Vision) 관점에서 최적화된 새로운 부호화 기술이 필요하다. 이에 따라, 국제 표준화 기구인 MPEG(Moving Picture Experts Group)에서는 기계 중심의 효율적인 영상 부호화를 목표로 하는 새로운 표준화 활동인 VCM(Video Coding for Machines)을 제안하여 활발히 표준화를 진행하고 있다^[2]. MPEG VCM에서는 비디오 객체 검출 및 추적 임무를 중심으로, 머신 비전 관점에서 관심 영역을 강조하거나 불필요한 정보를 제거하고, 또한 시공간 방향으로 적응적인 리샘플링 및 복원을 수행하는 등의 다양한 영상 전·후처리 알고리즘이 제안되고 있으며, 참조 소프트웨어인 VCM-RS를 통해 채택 기술을 취합하고 성능을 평가하고 있다^[3].

한편, 관련 학계에서는 인공 신경망 기반의 전후처리 기술을 통해 영상 내 의미 정보를 강조하거나 왜곡을 제거하는 연구 또한 활발히 진행되고 있다^[4-6]. 이러한 연구 방향 중 하나로, 영상을 상용 코덱으로 압축하면서 동시에 머신 비전 관점에서 최적화하기 위해 학습기반의 필터 네트워크를 최적화하는 기법이 사용되고 있다. 이때 일반적으로 미분이 불가능한 상용 코덱을 대체하기 위해 미분이 가능하면서도 압축에

의한 왜곡을 효과적으로 모사할 수 있는 Surrogate 코덱^[4] 또는 Proxy 코덱^[5-6]을 도입하여 중단간 학습을 통해 필터 네트워크를 최적화하는 프레임워크들이 활용되고 있다.

본 논문에서는 이러한 구조에 기반하여, 압축 왜곡을 반영하여 사전 학습된 영상 전·후처리 필터 네트워크를 MPEG VCM 프레임워크에 적용하고 성능을 평가한다. LIC(Learned Image Compression) 모델을 모사 코덱으로 하는 학습 구조를 기반으로, 입력 영상을 머신비전 성능 대비 압축 효율 관점에서 최적화하는 전·후처리 필터 네트워크를 학습시키고, 이를 VCM-RS^[3]에 적용한 결과에 대해 소개한다. 또한 학습 과정에서 기존 VCM-RS 모듈과의 호환성과 성능 최적화를 위해 적용된 2단계 학습 과정에 대해 설명한다. 이를 통해, VCM-RS 대비 전처리 및 후처리 필터의 적용 결과가 각각 17%, 13%의 BD-rate 성능 개선이 있음을 확인하였다.

II. 본론

1. 필터 네트워크 학습 구조

본 논문에서 제안하는 필터 네트워크는 그림 1과 같은 U-Net 기반의 구조로, 전처리 필터와 후처리 필터는 모두 동일한 구조를 가진다. 네트워크는 두 개의 브랜치(branch)로 구성되어 있으며, 그 중 U-Net 형태의 브랜치는 영상으로부터 의미론적인 정보를 추출하는 역할을 수행하고, 다른 하나는 1x1 컨벌루션(convolution)을 통해 픽셀별 특징을 추출할 수 있도록 설계하였다.

먼저 전처리 필터 네트워크의 학습 구조는 그림 2와 같다. 학습 데이터셋은 MPEG VCM에서 권장하고 있는 데이터셋인 OpenImages-V6^[7]가 사용되었으며, 압축 왜곡 모사를 위한 LIC 모델로는 mbt2018^[8]이 사용되었다. 손실함수는 세 가지의 Loss가 조합된다. LIC 모델로부터 출력되는 ‘예측 비트율’이 압축에 의한 Rate-Distortion을 최적화하기 위한 Rate Loss로 사용되고, 머신 비전 성능의 최적화를 위해서는 ResNext-101^[9] 기반의 Perceptual loss가 사용된다. 또한 여기에 추가적인 규제 항으로, 머신 비전을 위해 최적화 된다고 하더라도, 출력 영상이 학습 영상의 도메인에서

a) 한국전자통신연구원(Electronics and Telecommunications Research Institute)

b) 성균관대학교 인공지능학과(Artificial Intelligence, Sungkyunkwan University)

‡ Corresponding Author : 곽상운(Sangwoon Kwak)
E-mail: s.kwak@etri.re.kr
Tel: +82-42-860-4962

ORCID: <https://orcid.org/0000-0002-9295-8966>

※ 이 논문의 결과 중 일부는 한국방송·미디어공학회 2025년 하계학술대회에서 발표한 바 있음

※ This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00011, Video Coding for Machine)

· Manuscript September 15, 2025; Revised October 27, 2025; Accepted October 28, 2025.

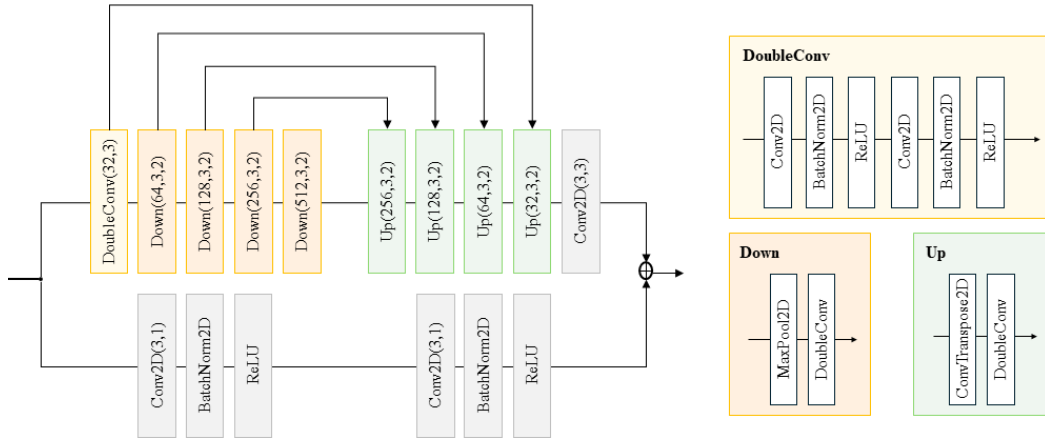


그림 1. 필터 네트워크 구조
Fig. 1. Filter network architecture

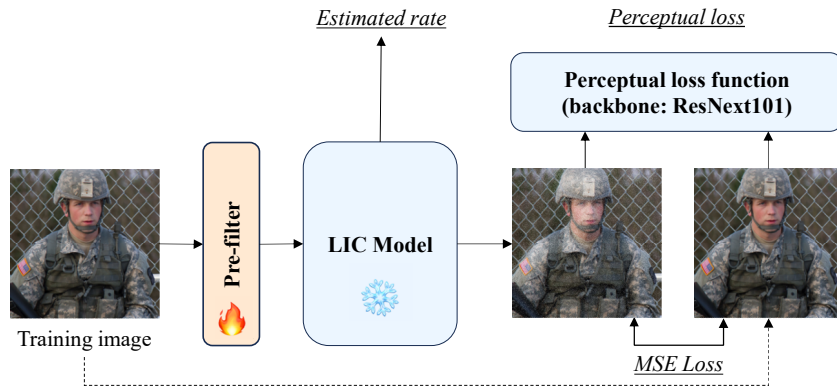


그림 2. 전처리 필터 네트워크의 학습 구조
Fig. 2. Training pipeline of the pre-filter network

지나치게 크게 벗어나는 것을 방지하기 위해 MSE Loss도 결합되어 학습에 사용된다. 손실함수의 구성은 수식 (1)과 같으며, λ 의 조절을 통해 필터 네트워크가 타겟하는 비트율 영역을 조정할 수 있다. 본 논문의 학습 과정에서 λ_1 은 다양한 타겟 비트율 범위에 적응하기 위해 [1.0, 1.5, 2.5, 4.0] 중 랜덤하게 샘플링하여 사용하였다.

$$L = L_{per} + \lambda_1 \cdot L_{rate} + \lambda_2 \cdot L_{MSE} \quad (1)$$

후처리 필터 네트워크의 학습 구조는 총 두 단계로 구성되며 이는 그림 3과 같다. 첫 번째 단계에서는 LIC를 기반으로 전반적으로 전처리 필터 네트워크와 유사하게 학습되

지만, 여기에 BDT(Bit Depth Truncation) 모듈의 효과를 반영하기 위해 모사 코덱 앞 단계 해당 모듈이 추가된다. BDT 모듈은 VCM에 채택되어 현재 VCM-RS 내에 포함되어 있는 모듈로, 객체의 검출 및 추적 성능 대비 압축 효율을 증대시키기 위해, 인코더 단에서 입력 영상에 대해 Luma 성분을 1 bit만큼 절단하고, 디코더 단에서 이를 다시 복원하는 모듈이다. 이러한 BDT를 거치면 영상의 구조적인 특징은 남아있지만 밝기가 어두워지게 되는데, 필터 네트워크는 일반적인 밝기를 가지는 영상 도메인에 대해 학습이 되었으므로, BDT가 적용된 영상이 입력되면 정상적으로 동작하지 않을 수 있다. 이러한 문제를 해결하기 위해 후처리 필터의 경우 학습 구조에 BDT 모듈을 반영함으

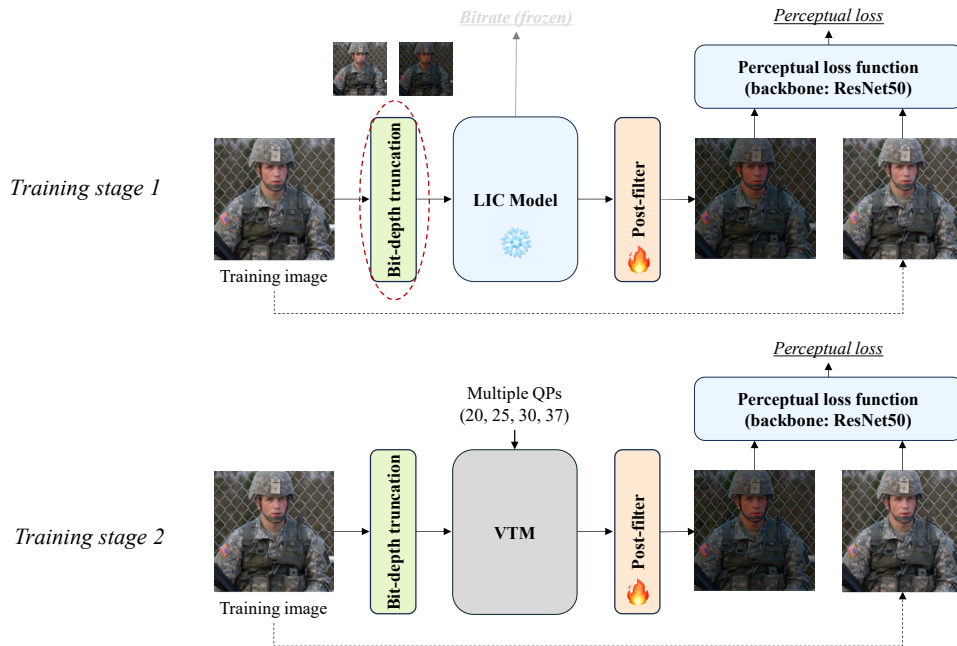


그림 3. 후처리 필터 네트워크의 학습 구조
 Fig. 3. Training pipeline of the post-filter network

로써 어두워진 영상에 대해서도 효과적으로 필터링을 수행할 수 있도록 설계하였다. 그리고 후처리 필터의 경우, 비트율에는 영향을 주지 않으므로 **Rate Loss**는 사용되지 않는다. 다만 다양한 수준의 압축 왜곡을 반영하기 위해 LIC의 품질(Quality factor)을 랜덤 값으로 조합하여 사용하였다. 후처리 필터를 통과한 출력 이미지는 전처리 필터와 마찬가지로 **Perceptual Loss**가 계산되는데, 이 때 **BDT**가 적용되지 않은 원래의 학습 이미지와 비교가 수행된다. 즉, 후처리 필터 네트워크는 **BDT**에 의해 어두워진 영상을 입력으로 하지만, 원본 영상과 가까워지도록 학습된다. 이는 후처리 필터 네트워크가 일종의 학습기반 비트 깊이 복원(Bit Depth Restoration, BDR) 모듈의 역할도 동시에 수행하도록 학습된다는 것을 의미한다.

학습의 두 번째 단계에서는 실제 사용될 상용 코덱에 의한 압축 왜곡을 추가로 반영하기 위해 **VTM**이 사용된다. 이 경우 학습 데이터셋으로부터 중단간 학습되는 것이 아니고, **BDT** 및 **VTM**에 의한 압축이 적용된 영상을 사전에 생성한 뒤, 후처리 필터 네트워크 이후부터 학습 과정이 수행된다. 여기서도 마찬가지로 **Rate loss**는 사용되지 않으나, 여

러 수준의 압축 왜곡을 반영하기 위해 **VTM**에 입력되는 **QP**를 [20, 25, 30, 37]의 값 중 랜덤하게 선택하여 사용한다.

2. VCM-RS 적용 구조

본 절에서는 사전 학습된 전처리 및 후처리 필터 네트워크가 **VCM** 참조 소프트웨어에 적용되는 구조에 대해서 설명한다. 학습된 필터 네트워크를 **VCM-RS v0.11**에 적용한 구조는 그림 4와 같다. 먼저 전처리 필터 네트워크의 경우 **RoI-based Processing**의 앞 단계, 후처리 필터 네트워크는 **Inner Codec(VTM)**의 바로 뒤에 위치한다. 이렇게 배치된 이유는 해당 위치가 성능 측면에서 가장 좋은 것으로 관찰되었기 때문인데, 그 이유를 추론해 보면 다음과 같다. 먼저 전처리 필터의 경우, 가능한 **BDT** 모듈과의 호응성 문제를 피하기 위해 **BDT** 적용 이전에 위치해야 한다. 그리고 **RoI-based Processing** 역시 머신 비전 관점에서의 관심 객체를 강조하기 위해 배경 영역을 제거하거나 **RoI**의 크기를 변형하는 등의 처리를 수행한다. 이러한 영상 변형을 거치는 것 역시 일반적인 영상과 구조적인 측면에서 크게 달라

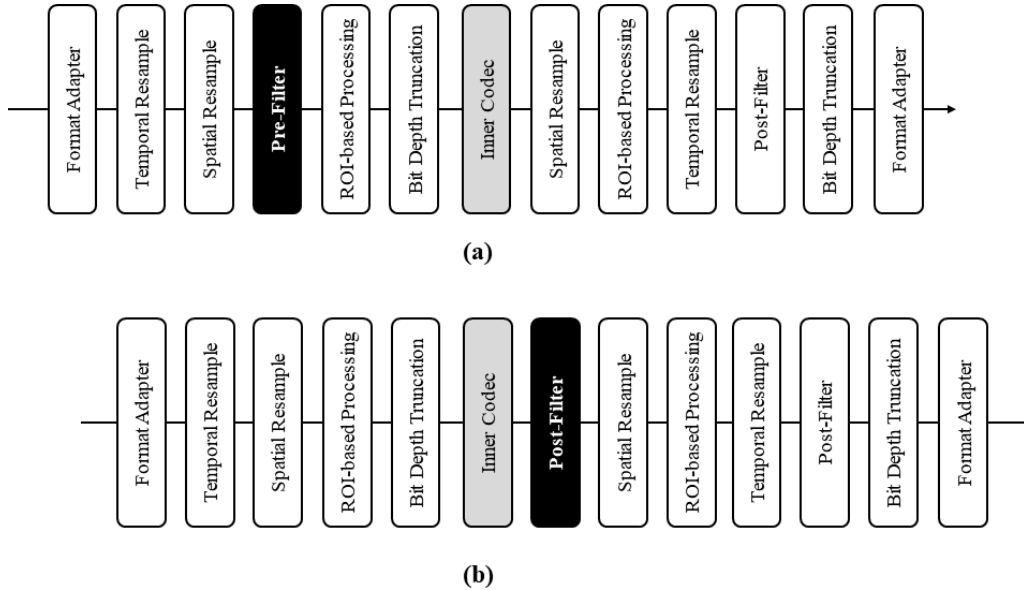


그림 4. 필터 네트워크의 VCM-RS v0.11 적용 구조 (a) 전처리 필터 네트워크 (b) 후처리 필터 네트워크
 Fig. 4. Applying of filter networks on VCM-RS v0.11 (a) pre-filter network (b) post-filter network

지기 때문에, 이 이후에 적용하는 것보다 이보다 앞서 적용하는 것이 좋은 성능을 보인 것으로 판단된다. 한편, 후처리 필터의 경우, 압축 왜곡을 포함하는 영상을 머신비전 관점에서 효과적으로 복원하도록 학습되었기 때문에, 가능한 Inner Codec과 가까워야 좋은 성능을 보이는 것으로 예상된다. 특히 비트 깊이를 복원하는 모듈은 VCM-RS 구조상 가장 마지막에 위치하기 때문에, 만약 이 모듈보다 후처리 필터를 뒤에 위치시키게 되면 입력되는 영상이 압축 왜곡 뿐만 아니라 다양한 후처리 과정에 의한 영상 변형을 모두 포함한 상태로 입력되어, 학습 과정에서 관찰한 영상과 특성이 크게 달라지게 된다.

또한 후처리 필터의 경우 앞서 기술하였듯이 비트 깊이 복원 모듈의 역할도 동시에 수행하도록 학습되었기 때문에, 기존의 BDT 모듈의 역할을 대체한다. 따라서 VCM 디코더가 비트 깊이를 복원하는 조건(비트 깊이 복원 플래그)이 활성화될 경우에만 후처리 필터가 동작하며, 기존의 비트 쉬프트 기반의 비트 깊이 복원 모듈은 동작하지 않는다.

3. 실험 결과 및 분석

전처리 필터 네트워크의 VCM-RS v0.11 적용 결과는 표

1, 표 2와 같다. VCM CTC^[3]에 따라, 성능 평가에 사용된 데이터셋 및 머신 비전 임무는 SFU-HW 데이터셋에 대한 객체 검출(object detection) 임무와 TVD 데이터셋에 대한 객체 추적(object tracking) 임무로 구성되며, 인코딩 모드는 총 6개로, RA, LD, AI 모드에 대해 전체 VCM 인-디코더 범위에 대해 해당 모드의 조건을 적용하는 E2E와, Inner Codec 범위에 대해서만 조건을 한정하는 Inner 모드로 정의되어 있다. 표 1은 SFU-HW 데이터셋에 대한 결과를, 표 2는 TVD 데이터셋에 대한 결과를 나타낸다.

전처리 필터 네트워크의 적용 성능을 보면 전체적으로 평균 17%가량의 BD-rate 성능 개선을 보이고 있음을 확인할 수 있다. 여기서 BD-rate이 “#####”으로 표시되는 부분은 두 그래프의 y축에서 겹치는 영역이 없어 발생하는 계산 오류로, 해당 경우에 대해 BD-mAP 또는 BD-MOTA 값을 보면 성능 개선이 발생했음을 알 수 있다. SFU 데이터셋의 결과를 보면, 클래스에 따라 성능 개선 폭에 차이가 존재하는데, 이는 학습 영상 데이터셋의 해상도와 유사할수록 높은 성능 개선 폭을 가지기 때문이다.

후처리 필터의 적용 성능은 표 3과 같다. 후처리 필터의 경우 전술한 것처럼 VCM 디코더가 비트 깊이를 복원하는 조건일 때만 활성화가 되는데, 데이터셋 범위로

표 1. 전처리 필터 네트워크의 VCM-RS v0.11 적용 성능 (SFU-HW 데이터셋)

Table 1. Performance of the pre-filter network applied to VCM-RS v0.11 on SFU-HW dataset

| Object | SFU-HW RA mode | | | |
|-----------|----------------|--------|-------------|--------|
| | Inner | | E2E | |
| Detection | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | -1.66% | 0.47 | -3.51% | 0.41 |
| Class B | -13.71% | 1.90 | -7.54% | 0.96 |
| Class C | -25.44% | 2.91 | -21.48% | 2.70 |
| Class D | -15.08% | 0.97 | -15.24% | 0.99 |
| Average | -17.07% | 1.81 | -14.42% | 1.50 |
| Object | SFU-HW LD mode | | | |
| | Inner | | E2E | |
| Detection | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | -0.27% | 0.43 | -11.41% | 2.14 |
| Class B | -2.94% | 0.72 | -23.57% | 3.54 |
| Class C | -19.69% | 2.28 | -25.34% | 3.35 |
| Class D | -24.87% | 2.70 | -21.34% | 1.88 |
| Average | -15.61% | 1.87 | -22.40% | 2.81 |
| Object | SFU-HW LD mode | | | |
| | Inner | | E2E | |
| Detection | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | -6.64% | 0.59 | -2.18% | 0.19 |
| Class B | -18.68% | 2.31 | ##### | 2.29 |
| Class C | -25.66% | 3.36 | -16.80% | 2.35 |
| Class D | -15.07% | 0.92 | 0.70% | -0.18 |
| Average | -18.80% | 2.05 | ##### | 1.31 |

표 2. 전처리 필터 네트워크의 VCM-RS v0.11 적용 성능 (TVD 데이터셋)

Table 2. Performance of the pre-filter network applied to VCM-RS v0.11 on TVD dataset

| Object | TVD RA mode | | | |
|----------|-------------|---------|-------------|---------|
| | Inner | | E2E | |
| Tracking | BD-Rate [%] | BD-MOTA | BD-Rate [%] | BD-MOTA |
| TVD-01-1 | -18.28% | 2.33 | -11.65% | 0.84 |
| TVD-01-2 | 43.70% | -1.79 | 14.30% | -1.25 |
| TVD-01-3 | -34.40% | 2.84 | -10.26% | 1.37 |
| TVD-02-1 | -27.91% | 2.03 | -12.97% | 1.18 |
| TVD-03-1 | -34.65% | 4.57 | -29.01% | 4.26 |
| TVD-03-2 | -24.77% | 3.76 | -19.52% | 3.55 |
| TVD-03-3 | -38.57% | 5.61 | -24.02% | 5.08 |
| Average | -19.27% | 2.76 | -13.30% | 2.15 |
| Object | TVD LD mode | | | |
| | Inner | | E2E | |
| Tracking | BD-Rate [%] | BD-MOTA | BD-Rate [%] | BD-MOTA |
| TVD-01-1 | -15.94% | 2.16 | -0.73% | -0.01 |
| TVD-01-2 | -22.55% | 1.88 | -18.61% | 3.56 |
| TVD-01-3 | -21.07% | 2.33 | -16.31% | 3.14 |
| TVD-02-1 | 13.80% | -0.61 | -6.31% | 0.55 |
| TVD-03-1 | -25.86% | 4.59 | -26.08% | 4.61 |
| TVD-03-2 | -23.19% | 4.27 | -27.63% | 5.88 |
| TVD-03-3 | -29.82% | 4.81 | -31.64% | 5.97 |
| Average | -17.80% | 2.77 | -18.19% | 3.39 |
| Object | TVD LD mode | | | |
| | Inner | | E2E | |
| Tracking | BD-Rate [%] | BD-MOTA | BD-Rate [%] | BD-MOTA |
| TVD-01-1 | -29.09% | 2.61 | #### | #### |
| TVD-01-2 | 13.94% | -0.83 | #### | #### |
| TVD-01-3 | -24.27% | 2.57 | #### | 1.15 |
| TVD-02-1 | 0.93% | -0.06 | -7.78% | 0.36 |
| TVD-03-1 | -17.70% | 3.49 | -45.01% | 3.01 |
| TVD-03-2 | -17.77% | 4.14 | -23.81% | 1.74 |
| TVD-03-3 | -20.84% | 3.21 | -33.04% | 0.97 |
| Average | -13.54% | 2.16 | #### | #### |

표 3. 후처리 필터 네트워크의 VCM-RS v0.11 적용 성능 (SFU-HW 데이터셋)

Table 3. Performance of the post-filter network applied to VCM-RS v0.11 on SFU-HW dataset

| SFU-HW RA mode | | | | |
|------------------|-------------|--------|-------------|--------|
| Object Detection | Inner | | E2E | |
| | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | 0.00% | 0.00 | 0.00% | 0.00 |
| Class B | 0.00% | 0.00 | 0.00% | 0.00 |
| Class C | -13.92% | 1.18 | -17.54% | 2.46 |
| Class D | -19.89% | 2.01 | -21.46% | 2.20 |
| Average | -11.27% | 1.06 | -13.00% | 1.55 |
| SFU-HW LD mode | | | | |
| Object Detection | Inner | | E2E | |
| | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | 0.00% | 0.00 | 0.00% | 0.00 |
| Class B | 0.00% | 0.00 | 0.00% | 0.00 |
| Class C | -23.11% | 2.18 | -23.10% | 2.56 |
| Class D | -15.45% | 1.76 | -27.08% | 2.75 |
| Average | -12.85% | 1.31 | -16.73% | 1.77 |
| SFU-HW LD mode | | | | |
| Object Detection | Inner | | E2E | |
| | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | 0.00% | 0.00 | 0.00% | 0.00 |
| Class B | 0.00% | 0.00 | 0.00% | 0.00 |
| Class C | -10.55% | 1.24 | -21.21% | 2.51 |
| Class D | -20.16% | 2.15 | -24.81% | 2.61 |
| Average | -10.24% | 1.13 | -15.34% | 1.71 |

표 4. 전처리 및 후처리 필터 네트워크의 VCM-RS v0.11 적용 성능 (SFU-HW 데이터셋)

Table 4. Performance of the pre-and post-filter network applied to VCM-RS v0.11 on SFU-HW dataset

| SFU-HW RA mode | | | | |
|------------------|-------------|--------|-------------|--------|
| Object Detection | Inner | | E2E | |
| | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | -1.66% | 0.47 | -3.51% | 0.41 |
| Class B | -13.71% | 1.90 | -7.54% | 0.96 |
| Class C | -30.23% | 3.57 | -29.03% | 4.25 |
| Class D | -28.94% | 2.43 | -30.17% | 2.65 |
| Average | -23.29% | 2.52 | -21.91% | 2.58 |
| SFU-HW LD mode | | | | |
| Object Detection | Inner | | E2E | |
| | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | -0.27% | 0.43 | -11.41% | 2.14 |
| Class B | -2.94% | 0.72 | -23.57% | 3.54 |
| Class C | -27.88% | 3.31 | -38.15% | 4.81 |
| Class D | -33.73% | 3.84 | -36.55% | 3.72 |
| Average | -21.29% | 2.60 | -31.74% | 3.91 |
| SFU-HW LD mode | | | | |
| Object Detection | Inner | | E2E | |
| | BD-Rate [%] | BD-mAP | BD-Rate [%] | BD-mAP |
| Class A | -6.64% | 0.59 | -2.18% | 0.19 |
| Class B | -18.68% | 2.31 | ##### | 2.29 |
| Class C | -35.53% | 4.20 | -34.52% | 4.27 |
| Class D | -29.69% | 2.57 | -22.85% | 2.15 |
| Average | -26.96% | 2.88 | ##### | 2.73 |



그림 5. 전처리 필터 네트워크의 적용 결과
 Fig. 5. Results of applying pre-filter network

는 SFU-HW의 Class C, D에 해당한다. 따라서 이 경우에 대해서만 결과를 표기하였다. 표의 결과를 확인하면, 후처리 필터 네트워크 역시 모든 경우에 대해 평균 성능 개선이 확인되며, 전체적으로는 약 13%의 BD-rate 개선이 도출됨을 확인할 수 있다. 또한 전처리 및 후처리 필터 네트워크가 모두 적용된 경우에 대한 결과는 표 4와 같다. 후처리 필터가 적용되는 경우인 SFU 데이터셋에 대한 결과만 결과를 나타내었으며, 이 경우 약 25%의 BD-rate 성능 개선이 얻어짐을 확인할 수 있다.

필터 네트워크가 적용되었을 때 영상의 변화는 그림 5와 6에서 확인할 수 있다. 먼저 그림 5는 TVD 데이터셋에 대해 전처리 필터 네트워크가 적용된 결과를 나타낸다. 결과를 보면, 전처리 필터 네트워크가 적용됨으로써 객체 인지에 중요한 정보인 객체의 경계 영역 등은 강조가 되지만, 배경 영역에 해당하는 바닥의 낙엽 등은 오히려 블러(blur)해지는 것을 확인할 수 있다. 전처리 필터 네트워크의 이러한 효과를 통해 객체의 인지 성능은 보존되거나 향상되더라도, 비트율은 오히려 절감하거나 최소한만 증가함으로써 BD-rate 성능 개선을 얻을 수 있게 된다. 또한 그림 6은 SFU 데이터셋에 대해 후처리 필터 네트워크가 적용된 결과를 나타낸다. 후처리 필터의 경우 비트 깊이까지 복원된 최종 영상으로 비교하였다. 결과를 보면, 기존 앵커(anchor)에서 사람의 팔과 같이 구조적 특징이 무너지는 부분이 후처리 필터가 적용된 경우 일부 복원되는 것을 확인할 수

있으며 또한 객체 인근의 아티팩트(artifact)도 제거됨을 확인할 수 있다.



그림 6. 후처리 필터 네트워크의 적용 결과
 Fig. 6. Results of applying post-filter network

III. 결론

본 논문에서는 머신 비전 성능 대비 압축 효율을 최적화하기 위해 VCM 환경에 특화된 학습기반 전·후처리 필터 네트워크를 제안하였다. 제안된 네트워크는 모사 코덱을 활용한 중단간 학습 구조와 BDT 모듈을 고려한 학습 전략을 통해 압축 왜곡 및 구조적 손실을 효과적으로 보정할

수 있도록 설계되었다. 실험 결과, 전처리 필터와 후처리 필터는 각각 VCM-RS v0.11 대비 평균 17% 및 13%의 BD-rate 성능 개선을 달성하였으며, 두 필터를 결합하여 적용한 경우 약 25%의 성능 향상을 확인할 수 있었다. 이러한 결과는 제안된 필터 네트워크가 VCM 프레임워크의 성능을 유의미하게 개선할 수 있음을 보여주며, VCM 표준화 과정에서 기계 중심 영상 부호화의 실질적인 활용 가능성이 있음을 제시한다. 향후 표준화 및 VCM 시스템에 효과적으로 적용되기 위해서는 신경망 기반의 모듈이 이종 머신에서도 동일한 결과를 산출할 수 있도록 보장해야 하며, 또한 객체 검출 및 추적 외에도 다양한 머신 비전 임무에 대해 일반화 성능 검증이 필요하며, 이러한 연구를 통해 제안 기법의 실질적인 적용 가능성을 높이는 것이 중요한 과제로 남아있다.

참 고 문 헌 (References)

- [1] L.-Y. Duan, J. Liu, W. Yang, T. Huang, and W. Gao, "Video coding for machines: A paradigm of collaborative compression and intelligent analytics," *IEEE Transactions on Image Processing*, vol. 29, pp. 8680 - 8695, 2020.
- [2] H. Choo, W.-S. Cheong, and J. Seo, "Standardization trends of video coding for machines," *Broadcasting and Media*, vol. 28, no. 1, pp. 38 - 52, 2023.
- [3] ISO/IEC JTC1/SC29/WG4, "Common test conditions for video coding for machines," N00638, Jan. 2025.
- [4] J. P. Klopp, K.-C. Liu, L.-G. Chen, and S.-Y. Chien, "How to exploit the transferability of learned image compression to conventional codecs," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 16165 - 16174, 2021.
- [5] O. G. Guleryuz, P. A. Chou, H. Hoppe, D. Tang, R. Du, P. Davidson, and S. Fanello, "Sandwiched image compression: Wrapping neural networks around a standard codec," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pp. 3757 - 3761, 2021.
- [6] G. Lu, X. Ge, T. Zhong, Q. Hu, and J. Geng, "Preprocessing enhanced image compression for machine vision," *IEEE Transactions on Circuits and Systems for Video Technology*, early access, 2024.
- [7] OpenImages V6, <https://storage.googleapis.com/openimages/web>
- [8] J. Bégin, F. Racapé, A. Pushparaja, and M. Akbari, "CompressAI: A PyTorch library and evaluation platform for end-to-end compression research," GitHub repository, <https://github.com/InterDigitalInc/CompressAI>
- [9] Detectron2, GitHub repository, <https://github.com/facebookresearch/detectron2>

저 자 소 개



곽 상 운

- 2012년 2월 : 연세대학교 전기전자공학 학사
- 2014년 2월 : 한국과학기술원 전기및전자공학 석사
- 2014년 ~ 현재 : 한국전자통신연구원 미디어연구본부 선임연구원
- ORCID : <https://orcid.org/0000-0002-9295-8966>
- 주관심분야 : 실감미디어, AI기반 영상 처리



정 성 문

- 2020년 : 성균관대학교 전자전기공학 학사
- 2020년 ~ 현재 : 성균관대학교 인공지능학과 석박통합과정
- ORCID : <https://orcid.org/0009-0002-4859-5093>
- 주관심분야 : 신경망 기반 데이터 압축, 비지도 학습, 생성형 모델

저 자 소 개



고 종 환

- 2004년 : 서울대학교 기계항공공학/컴퓨터공학(복수전공) 학사
- 2006년 : 서울대학교 전기·컴퓨터공학 석사
- 2006년 ~ 2019년 : 국방과학연구소
- 2017년 : Microsoft AI and Research, 리서치 인턴
- 2018년 : Georgia Tech 전기·컴퓨터공학 박사
- 2019년 ~ 2025년 : 성균관대학교 조교수
- 2025년 ~ 현재 : 성균관대학교 부교수
- ORCID : <https://orcid.org/0000-0003-4434-4318>
- 주관심분야 : 딥러닝 가속, 멀티미디어 처리, IoT, 디지털 시스템



정 순 흥

- 2001년 : 부산대학교 전자공학 (학사)
- 2003년 : 한국과학기술원 전기및전자공학 (석사)
- 2016년 : 한국과학기술원 전기및전자공학 (박사)
- 2005년 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0003-2041-5222>
- 주관심분야 : 실감미디어, 컴퓨터 비전, 머신러닝, 영상 부호화, 영상처리



추 현 곤

- 1998년 2월 : 한양대학교 전자공학과(공학사)
- 2000년 2월 : 한양대학교 전자공학과(공학석사)
- 2005년 2월 : 한양대학교 전자통신전파공학과(공학박사)
- 2015년 ~ 2017년 : 한국전자통신연구원 디지털홀로그래피연구실장
- 2017년 ~ 2018년 : Warsaw University of Technology, Poland 방문연구원
- 2023년 ~ 현재 : 한국전자통신연구원 실감미디어연구실장
- ORCID : <https://orcid.org/0000-0002-0742-5429>
- 주관심분야 : Computer Vision, 3D imaging and holography, 3D depth imaging, 3D broadcasting system