

신경망 기반의 적응적 움직임 벡터 해상도 판별 기법

백한결, 박상효

경북대학교

qorgksruf123@gmail.com s.park@knu.ac.kr

Neural Network-Based Adaptive Motion Vector Resolution Discrimination Technique

Baek, Han-Gyul Park, Sang-Hyo

Kyungpook National University

요약

Versatile Video Coding(VVC)에서 동영상 압축 효율을 증가시키기 위한 다양한 화면 간 예측(inter prediction) 기법 중에 적응적 움직임 벡터 해상도(Adaptive motion vector resolution, 이하 AMVR) 기술이 채택되었다. 다만 AMVR을 위해서는 다양한 움직임 벡터 해상도를 테스트해야 하는 부호화 복잡도를 야기하였다. AMVR의 부호화 복잡도를 줄이기 위하여, 본 논문에서는 가벼운 신경망 모델 기반의 AMVR 조기 판별 기법을 제안한다. 이에 따라 불필요한 상황을 미리 조기에 인지하여 대응한다면 나머지 AMVR 과정을 생략할 수 있기에 부호화 복잡도의 향상을 볼 수 있다.

1. 서론

High Efficiency Video Coding(HEVC)가 발표된 이후로 현재 영상의 해상도는 높아졌다. 고해상도의 영상을 압축하기 위해서는 혁신적인 압축 기술이 필요하게 되었다. MPEG와 VCEG 표준화단체들은 공동으로 Joint Video Exploration Team(JVET)을 결성하여 Versatile Video Coding(VVC)이라는 차세대 동영상 표준을 발간하였는데, VVC에 채택된 다양한 기술 중 압축 효율을 증가시키는 화면 간 예측(inter prediction)의 적응적 움직임 벡터 해상도(Adaptive motion vector resolution, 이하 AMVR) 기술은 움직임 벡터(Motion vector, 이하 MV)의 차이값 부호화/복호화 비트를 효과적으로 줄일 수 있는 기법이다.[1] 하지만 움직임이 큰 동영상(예컨대 스포츠 동영상)의 경우, MV 값이 크게 나타날 수 있어 비트 수의 증가가 초래될 수 있기에 압축 효율의 저하가 나타날 수 있다. 이러한 비효율적인 MV를 적응적으로 부호화하는 방법이 AMVR이다.

실시간 동영상 촬영이나 스트리밍 서비스를 제공하는 기기에는 MV 값이 크기에 복잡도의 감소가 중요하다. 특히 VVC의 경우, 기존 동영상 표준인 HEVC의 부호화 복잡도 대비 3000% 더 복잡한 것으로 알려져 있다.[2] AMVR로 비트율을 효과적으로 줄이기 위해, 어떤 MV 해상도로 MV를 부호화해야 할지 정해서 매번 다양한 MV에 대한 테스트를 진행하게 된다. 따라서 율-왜곡 비용(rate-distortion cost)의 계산이 필요하기 때문에 부호화 복잡도가 증가하여 복잡도를 줄이는 것이 관건이다. Joint Video Exploration Team(JVET)에서 해결을 위해 여러 노력이 있었지만, AMVR의 복잡도 감소의 현재 연구는 아직까지 미비한 실

정이다. 최근에 AMVR의 부호화 복잡도를 통계적인 기법으로 경량 하려는 연구가 있었다.[3] 다만, VVC의 특성상 HEVC보다 훨씬 다양한 feature가 존재하지만, 이러한 feature와 AMVR과의 관계를 머신러닝 기법으로 심도 있게 연구한 바는 없다.

관련 연구 중 VVC에 채택된 MTT의 TT가 VVC 부호화기의 복잡도를 크게 증가시킨 사례에 대한 복잡도 감소 연구에서 LNN 모델을 사용하여 복잡도를 감소시키고 CNN 모델보다 높은 정확도를 증명한 논문이 있다.[4] 이 논문을 바탕으로 본 논문에서는 AMVR의 복잡도의 감소를 위해 Multi-layer perceptron(MLP)로 미리 다른 정보를 가지고 판단할 수 있는 가벼운 신경망 모델을 사용하였다. 데이터셋의 양이 많아 2개로 분리하여 사용하였으며, Accuracy는 평균 89%의 좋은 결과가 나왔음을 확인할 수 있었다.

2. AMVR Classification Dataset

AMVR Classification Dataset은 이전에는 존재하지 않는 새로운 Dataset이며, 이 Dataset을 가지고 MLP를 사용하여 복잡도를 줄였다. AMVR Classification Dataset은 양이 많아 Python에서는 하나의 csv 파일로 실행하기에는 처리가 어렵기 때문에 Class_D와 Class_E 두 개의 데이터로 양을 나누어 실행하였다. AMVR Classification Dataset의 데이터 총 사이즈는 278MB, 전체 Sample 수는 7,371,725개, train/test split 비율은 0.2로 train set 비율을 전체 데이터셋의 80% 사용했으며, test set 비율은 전체 데이터셋의 20%를 사용했다. Dataset의 총 feature 수는 15개로 14개의 속성과 1개의 class로 이루어져 있다. 따라서 Input 값은 14개이며, Output 값은 1개이다.

* 이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2020R111A3072227).

3. 제안하는 모델 기법과 실험결과

Model: "sequential_15"

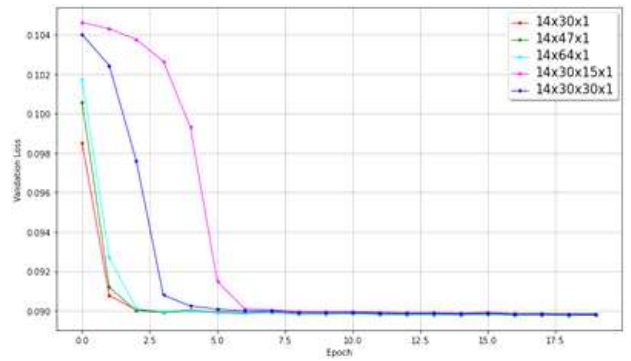
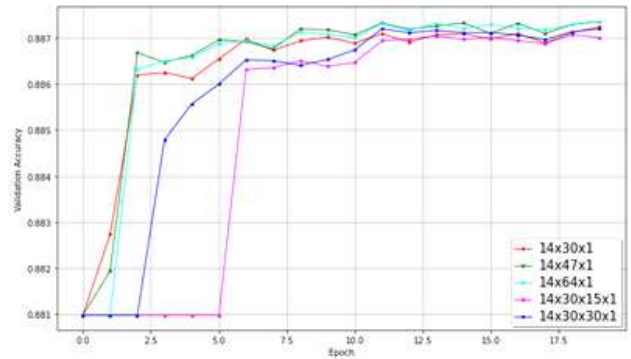
Layer (type)	Output Shape	Param #
dense_36 (Dense)	(None, 30)	450
dense_37 (Dense)	(None, 1)	31

Total params: 481
 Trainable params: 481
 Non-trainable params: 0

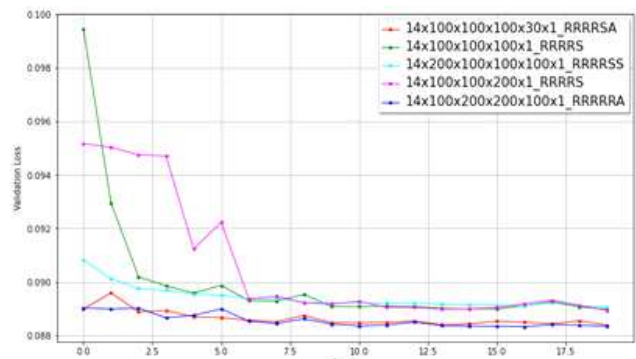
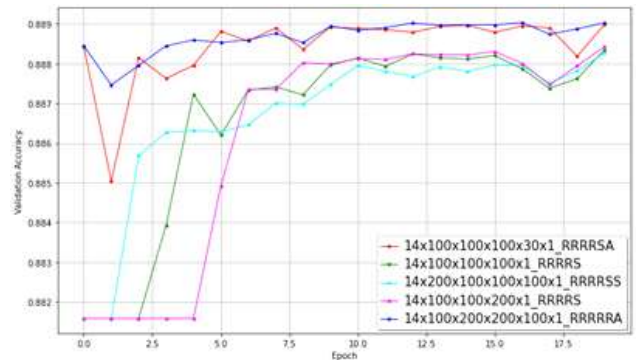
표 1. 모델 14x30x1의 모델 구조

본 논문의 핵심은 가벼운 신경망 모델로 AMVR 조기 판별 기법을 제안하는 것이기 때문에 가벼운 모델 5개와 Parameter를 변경한 모델 5개의 Test Accuracy를 비교하여 증명하였다. 실험 환경은 Python=3.7.10, keras=2.4.10, tensorflow=2.4.1, RAM 32GB, GeForce GTX 1660 SUPER에서 실행하였고 Epoch size=20, Batch size=100으로 설정하였다. 우리가 제안하는 모델의 model size는 14x30x1이다. 표 1과 같이 총 Parameter 수는 481개이며, 제안한 모델 Parameter의 상세정보는 Activation function으로 모두 Sigmoid, Loss function은 MSE, Optimizer로 SGD를 사용하였다. 따라서 우리는 14x30x1의 우수함을 증명하기 위해 간단하게 layer의 수와 노드의 수를 증가시킨 가벼운 모델들과 더 넓고 깊은 Parameter 변경 모델들을 함께 비교하였다. 가벼운 모델들의 model size는 14x30x1, 14x47x1, 14x64x1, 14x30x15x1, 14x30x30x1이며 Parameter 변경 model size는 14x100x100x100x30x1, 14x100x100x100x100x1, 14x200x100x100x100x1, 14x100x100x200x1, 14x100x200x200x100x1이다. 이 모델들은 다음과 같다.

그림 1의 (a), (b)는 Class_D Dataset에 대한 가벼운 모델들과 Parameter 변경 모델의 Epoch 별 Accuracy와 loss 그래프이다. (a)의 가벼운 모델 Accuracy의 최종 결과는 순차적으로 88.72%, 88.72%, 88.72%, 88.72%, 88.72%이며, (b)의 Parameter 변경 모델 Accuracy의 최종 결과는 88.9%, 88.9%, 88.9%, 88.9%, 88.9%이다. 그림 2의 (a), (b) Class_E Dataset에 대한 가벼운 모델들과 Parameter 변경 모델의 Epoch 별 Accuracy와 loss 그래프이다. Accuracy는 88.91%, 88.91%, 88.91%, 88.91%, 88.91% 와 Parameter 변경 모델은 90.14%, 90.14%, 90.14%, 90.14%, 90.14%이다. 따라서 AMVR Classification Dataset에 대한 가벼운 신경망 모델들과 Parameter 변경 모델의 Test Accuracy는 변경 모델이 조금 더 좋은 결과를 얻었지만, 매우 미세한 차이 때문에 가벼운 신경망 모델을 사용하는 것이 좋다. 최종적으로 결과를 종합해 보았을 때, 모델들 중 model size가 제일 작은 14x30x1이 가장 좋은 모델이라 할 수 있다.



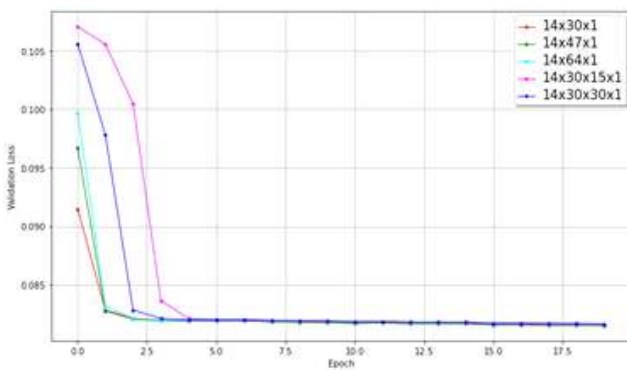
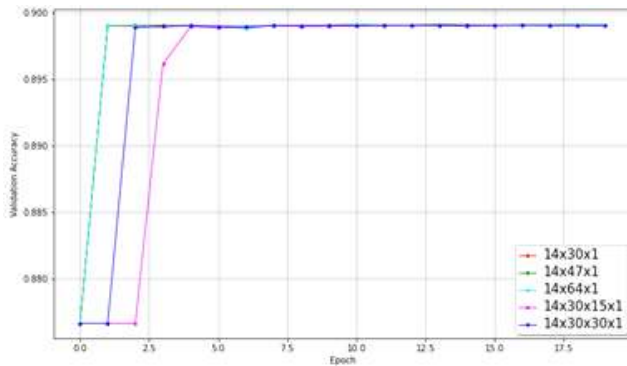
(a) Validation Accuracy/loss



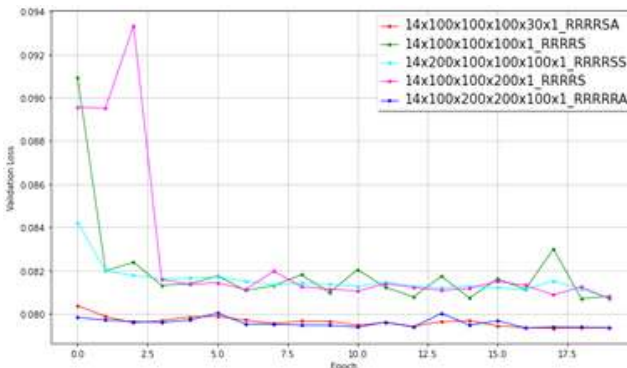
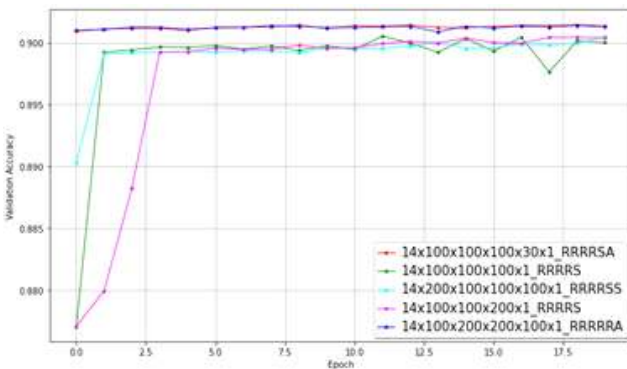
(a) Validation Accuracy/loss

그림 1. Class_D Dataset에 대한 가벼운 모델과 Parameter 변경

모델의 Epoch별 Accuracy와 loss 그래프



(a) Validation Accuracy/loss



(b) Validation Accuracy/loss

그림 2. Class_D Dataset에 대한 가벼운 모델과 Parameter 변경 모델의 Epoch별 Accuracy와 loss 그래프

참고문헌

[1] J. Chen, Y. Ye, and S. H. Kim, "Algorithm description for Versatile Video Coding and Test Model 6 (VTM 6)," Joint Video Experts Team (JVET) of ITU-T and ISO/IEC, Document JVET-O2002-v2, Jul. 2019. <http://phenix.it-sudparis.eu/jvet/> (accessed May 6, 2020)

[2] 한희지, 최재륜, 권대혁, 최해철, "VVC 화면내 예측 및 부호화 주요 기술", 방송과미디어 / v.24, no.4, 2019년 10월, pp.39-54

[3] Sang-hyo Park, "Fast Decision Method of Adaptive Motion Vector Resolution", 특집논문 (Special Paper) 방송공학회논문지 제 25권 제3호, 2020년 5월 (JBE Vol. 25, No. 3, May 2020). <https://doi.org/10.5909/JBE.2020.25.3.305>, ISSN 2287-9137 (Online) ISSN 1226-7953 (print)

[4] S. Park and J. Kang, "Fast Multi-type Tree Partitioning for Versatile Video Coding Using a Lightweight Neural Network," in ,inpress.DOI:10.1109/TMM.2020.3042062