특집논문 (Special Paper)
방송공학회논문지 제28권 제4호, 2023년 7월 (JBE Vol.28, No.4, July 2023)
https://doi.org/10.5909/JBE.2023.28.4.410
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

저장 및 전송 시스템에서의 해상 영상을 위한 딥 러닝 기반 영상 개선 기술 분석

이 영 복[a)], 이 은 성[a)], 이 민 훈[a)], 변 주 형[a)], 안 현 모[b)], 심 동 규[a)‡]

Deep Learning-Based Image Enhancement Techniques for Maritime Video in Storage and Transmission Systems: A Research Study

Youngbok Lee[a)], EunSeong Lee[a)], Minhun Lee[a)], Joohyung Byeon[a)], Hyeonmo Ahn[b)], and Donggyu Sim[a‡]

요 약

최근 들어 높은 품질의 영상에 대한 소비가 증가함에 따라, 전송 및 저장이 필요한 응용 시스템에서는 영상 압축 기술을 활용하고 있다. 그러나, 대부분의 영상 압축 기술은 다양한 압축 열화를 수반하여 인간 시각 시스템에 불편함을 초래하므로, 압축 열화 제거 기술의 적용이 요구된다. 또한, 영상 데이터의 양을 최소화하기 위해 영상의 해상도와 프레임 율을 낮추는 시나리오에 대비하여, 영상 재생 단말에서는 초해상화 알고리즘과 영상 프레임 보간 알고리즘과 같은 후처리 기술을 적용할 필요가 있다. 본 논문에서는 딥 러닝 기반의 압축 열화 제거 기술과 초해상화 기술, 프레임 보간 기술의 최신 연구 동향을 소개하고, 적용 분야의 예시로 해상 영상에 대한 기술들의 상대적인 우수성을 확인하고자 한다. 또한, 각 기술의 적합성 평가를 위해서 복원 정확도 측면에서의 고품질 모드와 네트워크 복잡도 관점에서의 실시간 모드를 구별하여 선별하였다.

Abstract

With the increasing demand for high-quality video, video compression technologies have been widely used in various applications that require transmission and storage. However, most video compression techniques introduce various compression artifacts, causing discomfort to the Human Visual System (HVS). In addition, to minimize the amount of video data, scenarios manipulating resolution and frame rates must be considered. Therefore, it is necessary to adopt not only compression artifact removal techniques, but also super-resolution algorithms and frame interpolation algorithms in video playback devices. In this paper, we analyzed the latest research trends and demonstrated the relative superiority of the techniques for maritime videos as one of application examples. Furthermore, we evaluated the suitability of each technique and cherry-picked the best for a high-quality mode in terms of restoration accuracy and a real-time mode considering network complexity.

Keyword: Video Enhancement, Denoising, Super-resolution, Frame Interpolation, Maritime Video

Ⅰ. 서 론

최근 고성능 카메라의 보급 및 고해상도 디스플레이의 등장에 따라 저장 및 전송 과정에서 저하된 영상의 품질을 개선하기 위한 연구도 많아지고 있다[1]. 그리고 실제 응용 시스템 상에서는 저장 공간이 한정되거나 전송 채널의 대역폭이 제한되기 때문에, 영상 데이터의 크기를 최소화하기 위해 정답 영상을 HEVC (High efficiency video coding)[2], VVC (Versatile video coding)[3][4] 등의 영상 압축 표준 기술을 통해 압축하거나 해상도 및 프레임 율을 의도적으로 낮추는 경우가 있다. 그러나, 응용 시스템의 환경에 따른 요구사항을 만족시키기 위해 영상 데이터의 크기를 줄이는 일련의 과정은 불가피하게 영상의 품질을 손상시킬 수 있다[5]. 일반적으로 영상의 품질은 취득 과정에서 다양한 외부적 요인과 무작위 간섭에 의해 영향을 받을 수 있으며 압축 과정에서의 데이터 손실에 의해 훼손될 수 있다[6][7]. 이에 따라 영상 취득 과정부터 저장 및 전송 과정까지 수반되는 모든 품질 저하 문제를 해결하기 위한 영상 개선 기술들에 대한 연구가 활발히 진행되고 있다[5][8][9]. 영상 취득 과정에서는 움직임 열화, 저조도 환경에서의 잡음 등이 영상에 포함될 수 있으며 영상 압축 과정에서는 주로 압축 열화가 발생하거나 해상도 저하, 프레임 율 손실과 같은 문제들이 나타날 수 있다. 이러한 문제들을 해결하기 위해 활용할 수 있는 영상 개선 기술들로는 대표적으로 영상 내 잡음 및 압축 열화 제거 기술 (Denoising), 초해상화 기술 (Super-resolution), 프레임 보간 기술 (Frame interpolation)이 있다. 저장 및 전송 시스템 장치의 하드웨어 상의 제약으로 인해 저하된 영상의 품질을 영상 재생 장치에서 복원하기 위해 앞서 언급한 세 가지 기술들의 적용이 필수적이며, 영상 재생 단말에서의 처리를 용이하게 하기 위해 메모리 점유를 줄이고 연산량을 낮춰 작업 속도를 높일 수 있는 기술을 선택하여 사용하는 것이 바람직하다. 본 논문에서는 열화 제거, 초해상화, 프레임 보간 기술로 구성된 영상 개선 시스템을 그림 1과 같이 정의하고 각 기술들의 영상 개선 성능과 함께 실시간 처리 능력을 검증하고자 한다. 해당 영상 개선 시스템은 HEVC로 압축 및 복원된 영상을 입력으로 세 가지 영상 개선 기술을 통해 압축 열화 제거, 해상도 향상, 프레임 율 향상의 효과를 얻음으로써 저 품질의 입력 영상을 고품질의 영상으로 개선해주는 기능을 수행할 수 있다.

영상 내 열화는 영상 취득 및 압축 과정 모두에서 발생할 수 있다. 영상 취득 과정에서는 카메라의 움직임 또는 물체의 이동으로 인한 움직임 열화가 발생할 수 있으며, 영상 압축 과정에서는 영상 압축 열화가 발생할 수 있다[10]. 영상 내 열화 제거 기술은 이러한 움직임 열화 및 압축 열화 요인으로 인해 발생할 수 있는 영상 내 잡음을 줄이는 데 사용된다. 특히, 대용량 영상의 전송 및 저장에는 많은 시간과 비용이 들기 때문에 영상 압축 기술의 적용이 필수적이며, 영상을 압축하는 과정에서 영상 신호의 왜곡이 발생할 수 있다[4][11]. 예를 들어, 영상의 많은 고주파 구성 요소가 양자화로 인해 손실되어 링잉 (Ringing) 열화와 블록 경계에서 발생하는 차단 열화가 발생할 수 있다[3][12]. 이러한 압축 열화는 영상 품질을 저하시켜 인간 시각 시스템에 불편함을 초래하기 때문에, 픽셀들의 불 연속적인 특성을 제거하는 기술을 사용하여 영상 품질을 향상시키고 보다 정확한 정보 전달을 할 필요가 있다. 또한, 움직임 열화는 영상 내 물체의 선명도를 저하시켜 물체가 흐릿하게 보이는 문제를 초래할 수 있다. 따라서, 영상 데이터의 특징과 패턴을 학습하여 영상 내 열화를 제거하고, 선명한 영상을 재구성하기 위한 연구의 중요성이 대두되어 왔다[13]. 전통적인 신호처리 접근 방식은 영상 내 열화의 분포와 같은 선험적 특징과 함께 MAP (Maximum a posterior)을 사용하였다[14][15][16][17][18]. 그러나 이러한 알고리즘들은 합성 열화가 아닌 복잡한 특성을 갖는 실제 열화에 대해서는 효과적이지 않을 수 있다[19]. 딥 러닝 기술은 컴퓨터 비전을 비롯하여 음성 인식, 자연어 처리 등 다양한 분야에서 우수한 성능을 보여 널리

a) 광운대학교 컴퓨터공학과(Department of Computer Engineering, Kwangwoon University)
b) ㈜큐램(Quram Co., Ltd.)
‡ Corresponding Author : 심동규(Donggyu Sim)
E-mail: dgsim@kw.ac.kr
Tel: +82-2-940-5470
ORCID:https://orcid.org/0000-0002-2794-9932
※본 연구는 대한민국 정부 산업통상자원부 및 방위사업청 재원으로 민군협력진흥원에서 수행하는 민군기술협력사업(UM21411RD4)의 연구비 지원 및 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (NRF-2021R1A2 C2092848)의 연구결과로 수행되었음.
· Manuscript June 12, 2023; Revised July 21, 2023; Accepted July 21, 2023.

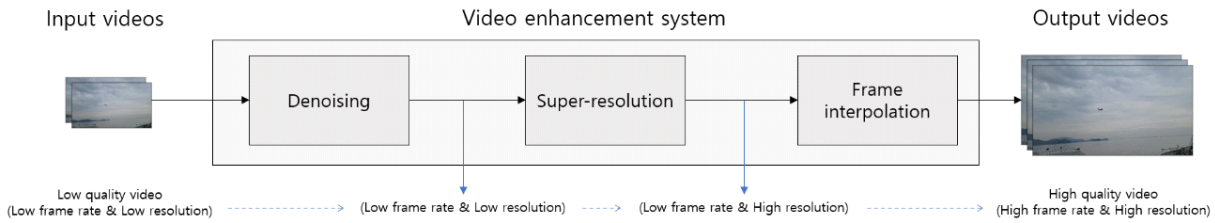

그림 1. 열화 제거 기술, 초해상화 기술, 프레임 보간 기술로 구성된 영상 개선 시스템의 블록 다이어그램
Fig. 1. A block diagram of video enhancement system consisting of denoising, super-resolution, and frame interpolation techniques

사용되고 있는데, 특히 커널의 수용 영역의 크기만큼 영상 내 특징을 추출할 수 있어 많은 정보를 참조하는 CNN (Convolutional neural network)을 통한 열화 제거 성능이 부각되고 있다[20][21][22]. 그러나 더 크고 복잡해지는 일부 알고리즘[23][24][25]은 열화 제거 결과에서 높은 성능을 얻을 수 있다고 하더라도, 에지 디바이스에 탑재하는 것과 같은 실제 시나리오에서는 적절하지 않을 수 있기 때문에 응용 시스템에 적용할 모델을 선택함에 있어 네트워크의 복잡도를 함께 고려해야 한다.

또 다른 영상 개선 기술 중 하나인 초해상화 기술은 영상의 해상도를 높여 영상 내 시각 정보의 정확하고 세밀한 전달을 가능하게 한다[26]. 열화가 포함된 저해상도의 영상에 초해상화 기술을 적용하면 영상의 해상도가 낮아질 때 발생하는 계단 현상을 완화시켜 더욱 선명하고 자연스러운 영상을 만들어낼 수 있다[27][28]. 또한, 낮은 조도에서 촬영된 영상에 대해서도 더 나은 시각적 품질을 가진 영상을 얻을 수 있다. 초해상화 기술은 입력 데이터에 따라, SISR (Single image super-resolution) 및 VSR (Video super-resolution)로 분류할 수 있다. SISR을 위한 전통적인 접근 방식으로는 희소 부호화 기반의 연구[26]와 스티어링 커널 회귀를 이용한 비 로컬 평균을 활용한 연구[27], 랜덤 포레스트 (Random forest) 기반의 연구 등이 수행되었다[28]. 그러나 이러한 알고리즘들은 외부 정보를 참조하므로 그에 맞는 최적화 알고리즘이 필요하며 매개 변수 조정이 필요하다는 단점이 있다[29][30]. 이를 해결하기 위해, 종단간 네트워크 아키텍처가 제안되었고[31][32][33], 인코더-디코더 네트워크[34][35], 경량화 모델[36][37]등이 개발되었다. VSR (Video super-resolution)을 위해서 시간 축으로 RNN (Recurrent neural network)을 사용하는 방법, 시공간 축의 정합 모듈과 함께 움직임 정보를 사용하는 방법도 지속적으로 연구되어 왔다[38][39][40][41][42]. 일반적으로 딥 러닝 기반 초해상화 기술은 전통적인 방식보다 우수한 성능을 보이지만, 실제 응용을 위해서는 계산 복잡도 및 메모리 접근 문제 등을 고려해야 한다. 따라서 실제 응용을 위한 대표적인 초해상화 기술들을 분석하고, 기술 적용 과정에서 당면할 수 있는 다양한 문제들을 고려해야 한다.

프레임 보간 기술은 초당 프레임 수를 늘려 객체의 움직임 및 픽셀의 변화를 부드럽게 만들거나 사용자 맞춤 기능 편의성을 위한 슬로우 모션 (Slow motion) 영상을 생성하기 위한 기술이다[43]. 연속된 있는 두 개의 프레임 사이의 중간 프레임을 생성함으로써 낮은 프레임 율을 가진 영상의 움직임 불연속성을 줄여 품질을 향상시킬 수 있다. 프레임 보간 기술은 일반적으로 커널 예측에 의존하고, 시공간 디코딩에 효과적일 수 있다[44][45][46][47]. 원하는 시간 단계에 대해 역방향 와핑 (Backward warping)을 적용하여 중간 프레임을 합성하는 연구가 이뤄지고 있으며[27][38][48], 광학 흐름 (Optical flow)을 통해 정방향 와핑 (Forward warping)을 수행하는 연구가 있다[49]. 또한, 정방향 와핑에서 다수 픽셀들이 같은 위치로 맵핑 되는 경우 발생하는 열화를 수정하기 위한 딥 러닝 기반의 프레임 보간 기술의 연구도 수행되었다[3][50][51]. 하지만, 아직도 양방향 동작의 추정이 어렵기 때문에 앞서 언급한 문제들을 해결할 수 있는 대표적인 프레임 보간 기술들을 분석하고, 고품질과 실시간성을 요구하는 응용 시스템에서의 적합성을 평가할 필요가 있다.

최근 딥 러닝의 발전으로 딥 러닝 기반 영상 개선 기술에 대한 연구가 활발히 이뤄지고 있다. 이러한 영상 개선 기술은 한정된 컴퓨팅 자원을 사용하는 해상 감지[52][53], 원격 탐지[54][55], 의료 영상 분석[56][57] 등 많은 분야에서 활용되고

있다. 특히 해양 감시 정찰은 무인수상정 및 선박에서의 대역폭 및 컴퓨팅 자원 문제로 인해 일반적으로 영상 압축 혹은 낮은 해상도로 전송이 이뤄지므로, 영상의 품질저하가 발생하여 영상 개선 기술 적용이 필요하다[58][59]. 현재 많은 영상 개선 기술 연구가 이뤄지고 있으나, 해상 감시 정찰의 해상 영상에 대한 딥 러닝 기반 솔루션의 포괄적인 조사는 아직 이뤄지지 않았다. 따라서, 본 논문에서는 해상 영상을 저장 및 전송하는 시스템에 대하여 앞서 언급한 세 가지 영상 개선 기술들의 기술 적합성을 평가하고자 한다. 이는 실험 결과에서 영상 내 열화 제거, 초해상화, 프레임 보간을 위한 기술로 구성되는 영상 개선 장치에 사용되기에 적합한 모델들을 선정하였다.

본 논문의 구성은 다음과 같다. 2장에서는 영상 개선 기술들 중 대표적인 기술로 영상 내 열화 제거 기술, 초해상화 기술, 프레임 보간 기술에 대한 딥 러닝 기반 기술을 분석하고, 3장에서는 적용 분야 중 하나의 예시로 해상 영상에 대한 실험을 수행하고, 각 기술에 대하여 고품질 모드와 실시간성 모드를 위한 네트워크를 선정한다. 마지막으로 4장에서 결론을 맺는다.

Ⅱ. 딥 러닝 기반 영상 개선 기술

최근 몇 년간 딥 러닝의 급속한 발전으로 딥 러닝 기반 영상 개선 기술들이 활발히 연구되어 왔다[60]. 초기 CNN 기반 방법에서 GAN (Generative adversarial network), 종단간 네트워크, ViT (Vision transformer)를 사용하는 방법에 이르기까지 영상의 품질을 개선하기 위해 다양한 딥 러닝 방법이 적용되어왔다[61]. 일반적으로 딥 러닝 기술을 사용하는 영상 개선 기술은 다양한 유형의 네트워크 아키텍처, 손실 함수, 학습 원리 및 전략의 세가지 측면에서 차이점이 존재한다[62]. 본 장에서는 각 영상 개선 기술들의 세 가지 측면에서 분석하도록 한다.

1. 영상 내 열화 제거 기술

Lee 등.[63]은 JPEG으로 압축이 수행된 복원 영상의 압축 열화를 제거하고자 광역 수용장 (Wide receptive field) 및 채널 주목 메커니즘을 이용한 네트워크인 WRCAN (Wide receptive residual block with channel attention network)을 제안하였다. WRCAN의 기본 구조는 U-Net 구조를 따르며, 스킵 연결 (Skip-connection)을 통해 고해상도 입력 영상의 컨텍스트를 캡처하여 출력 영상을 정밀하게 국소화 (Localization)하였다. 또한, 가장 깊은 계층에 WRCA (Wide receptive residual block with channel attention) 모듈을 사용한다. 이 모듈은 광역 수용장에서 특징을 추출하기 위해 병렬 아트러스 합성곱 (Parallel atrous convolution)을 갖는 잔차 블록을 사용하였다. 그리고 채널 주목 메커니즘을 통해 각 채널의 가중치를 추출 후 WRCA의 입력 특징 맵과 곱하여 특징 맵의 중요한 채널을 주목하여 사용하는 특징을 갖고 있다. WRCAN의 손실 함수는 식 (3)과 같으며, 네트워크의 최종 입출력 간의 손실 함수와 JPEG 압축 영상의 특성을 나타내는 JPEG 오토인코더 손실 함수로 구성되어 있다. 각 손실 함수는 식 (1)과 식 (2)에 나타나 있다. JPEG 오토인코더 손실 함수는 JPEG 압축 열화를 효과적으로 줄이기 위해 사용되며, 입력 영상과 WRCA 모듈을 제외한 U-Net기반의 오토인코더만을 통과한 출력 영상 간의 MSE (Mean square error)이다.

$$L_1 = \frac{1}{M}\sum_{i=1}^{M}||y' - y_{all}||_2 \quad (1)$$

$$L_2 = \frac{1}{M}\sum_{i=1}^{M}||y' - y_{AE}||_2 \quad (2)$$

$$Loss = \lambda_0 L_1 + \lambda_1 L_2 \quad (3)$$

여기서 M은 영상의 픽셀 수를 의미하며, y_{all}은 입력 영상을 모든 네트워크에 통과시킨 최종 출력이고, y'은 정답 영상이다. y_{AE}는 입력 영상을 오토인코더에 통과시킨 출력이고, λ_0, λ_1는 하이퍼 파라미터로 실험적으로 정해진다.

Qing 등.[12]은 HEVC로 압축이 수행된 복원 영상의 압축 열화를 제거하고자 공간 및 시간 정보를 모두 사용하고 압축된 패치를 적응적으로 활용하여 압축 열화를 제거하는 PSTQE (Patch-wise spatial-temporal quality enhancement)를 제안하였다. PSTQE는 공간 및 시간 정보를 효율적으로

활용하기 위해 공간적 특징 추출 서브넷, 시간적 특징 추출 서브넷을 통해 시공간적 특징을 추출한 후, 융합 서브넷을 통해 패치 별 융합을 수행한다. 공간적 특징 추출 서브넷은 공간 정보를 효과적으로 추출하기 위해 잔차 댄스 블록[6] 기반의 재귀 잔차 댄스 블록을 사용하며, 시간적 특징 추출 서브넷은 시간 축이 포함된 3차원 공간으로 증류하여 영상 작업에서 시간 정보를 처리하는데 적합한 3차원 합성곱을 수행한다. 융합 서브넷은 공간 및 시간적 특징 맵이 채널 차원으로 연결되고 CSAF (Channel and spatial-wise attention fusion) 블록을 통해 잔차를 더해주어 최종적으로 열화 제거를 위한 프레임 재구성을 수행한다. PSTQE의 손실 함수는 전체 네트워크의 입력과 출력 간의 MSE외에 각 서브넷의 MSE가 추가되며, 식 (7)과 같다.

$$L_{spa} = \frac{1}{M}\sum_{i=1}^{M}\|y' - y_{spa-\neq t}\|_2 \tag{4}$$

$$L_{tem} = \frac{1}{M}\sum_{i=1}^{M}\|y' - y_{tem-\neq t}\|_2 \tag{5}$$

$$L_{all} = \frac{1}{M}\sum_{i=1}^{M}\|y' - y_{all}\|_2 \tag{6}$$

$$Loss = \lambda_0 L_{spa} + \lambda_1 L_{tem} + \lambda_2 L_{all} \tag{7}$$

여기서 y'은 정답 영상이다. $y_{spa-net}$는 공간적 특징 추출 서브넷의 출력이며, $y_{tem-net}$은 시간적 특징 추출 서브넷의 출력이다. λ_0, λ_1, λ_2는 하이퍼 파라미터로 실험적으로 정해진다.

Yang 등.[5]은 잠재 표현 공간 (Latent representation)을 저해상도 정답 영상과 동일하게 만드는 가역 네트워크인 InvDN을 제안하였다. InvDN은 먼저 입력 영상을 하 웨이블릿 변환 (Haar wavelet transform)을 통해 열화가 포함된 세개의 잠재 표현 공간으로 변환한다. 그리고 열화가 포함된 잠재 표현 공간의 일부 채널을 사전 분포 (Prior distribution)에서 샘플링된 것으로 대체한 후 저해상도의 입력 영상과 결합하여 열화가 제거된 영상을 얻는다. 가역 네트워크는 기존의 심층 노이즈 제거 모델과 비교하여 다른 특징 추출 접근법을 활용한다. 기존의 기술들은 보통 특징을 추출하기 위해 패딩이 있는 합성곱 계층을 사용하지만, 해당 계층은 네트워크를 비가역적으로 만든다. 따라서 가역성을 보장하기 위해, 합성곱 계층을 사용하는 대신, InvDN은 스퀴즈 계층과 하 웨이블릿 (Haar wavelet) 변환을 사용하여 가역적 특징을 추출한다. 손실 함수의 경우 잠재 표현 공간의 일부 채널을 대체 과정은 아래와 같다.

$$L_{forw}(y, x_{LR}) = \frac{1}{M}\sum_{i=1}^{M}\| g(y)_{LR} - y'_{LR} \|_m \tag{8}$$

이때 y는 열화 영상이고, y'_{LR}은 저해상도의 정답 영상이며 정답 영상을 바이큐빅 (Bicubic) 다운샘플링을 통해 얻는다. 또한 $g(\cdot)$는 가역 변환을 나타내고, $\| \cdot \|_m$은 m-norm이며, m은 1 또는 2일 수 있다. 대체 과정 후 복원 시에는 역변환 $g^{-1}([g(y)_{LR}; z_{HF}])$를 사용한다. 역변환 시에 정규 분포 N(0,1)에서 추출한 랜덤 변수 z_{HF}를 사용한다. 복원 과정 시 손실 함수는 아래와 같다.

$$L_{back}(g(y)_{LR}, x) = \frac{1}{M}\sum_{i=1}^{M}\| g^{-1}([g(y)_{LR}; z_{HF}]) - y' \|_m \tag{9}$$

Tsai 등.[8]은 연산량을 줄이기 위해 주목 모듈을 intra/inter-SA (strip attention)로 분리하여 수평 및 수직 방향에 대한 주목 예측 모듈을 수행하는 Stripformer를 제안하였다. 입력 영상이 FEB (Feature embedding block)을 통해 패치의 인코딩 정보가 삽입된 특징 맵이 추출되고, intra/inter-SA를 통해 열화 정보를 얻고, 이를 FEB에서 추출된 특징 맵과 연결하여 열화 성분을 제거하기 위한 신호를 얻는다. 그리고 이 신호를 입력 영상과 합쳐 열화가 제거된 최종 영상을 출력한다. Intra-SA 블록에서는 먼저 1x1 합성곱 계층을 통해 계층 정규화를 수행하여 채널의 크기를 절반으로 줄인다. 이후 수평 및 수직 방향에 대해 멀티 헤드 주목 메커니즘 (Multi-head attention mechanism)과 소프트맥스 (Softmax)를 통해 특징 맵을 생성하고 연결 후 다중 계층 퍼셉트론 블록 (Multilayer perceptron block)을 거쳐

Intra-SA주목 맵을 추출한다. Inter-SA 블록은 Intra-SA 주목 맵을 통해 각 행 혹은 열 간 주목 맵을 계산한다. 세부적인 처리 과정은 Intra-SA 블록과 유사하다. 손실 함수의 경우 Charbonnier 손실 함수, 에지 (Edge) 손실 함수 그리고 대조 (Contrastive) 손실 함수로 구성된다. Charbonnier 손실 함수[64] 의 경우 정답 영상과 출력 영상 간의 픽셀 단위 오차를 측정하되, 패널티 구간을 추가하여 이상치에 강인한 특정을 갖는 손실 함수이다. 에지 손실 함수의 경우 정답 영상과 출력 영상 간 에지 공간에서의 MSE 다. 마지막으로 대조 손실 함수는 수식 (10)과 같으며, 최종 손실 함수는 수식 (11)와 같다.

$$L_{con} = \frac{\|(\psi(s)-\psi(y))\|_1}{L_1(\psi(x)-\psi(y))} \tag{10}$$

$$Loss = \lambda_0 L_{char} + \lambda_1 L_{edge} + \lambda_2 L_{con} \tag{11}$$

이때 $\| \cdot \|_1$은 L1-norm이며, 은 사전 학습된 VGG-19에서 추출한 특징 맵을 나타낸다. x는 입력 영상이고, y은 출력 영상, s는 샤프닝 (Sharpening) 처리된 정답 영상이다.

Wang 등.[10] 은 영상 복원을 위해 계층적 인코더-디코더 네트워크가 구성된 U자형 트랜스포머, Uformer를 제안하였다. Uformer는 오버랩되지 않는 윈도우를 이용한 자기 주목 모듈이 포함된 LeWin (Locally enhanced window) 트랜스포머 블록을 사용하였으며, 다운 및 업샘플링과 스킵 연결을 사용하여 결과적으로 컴퓨팅 자원을 줄이면서 성능을 향상시켰다. 인코더에서는 먼저 합성곱 계층을 통해 특징 삽입 과정을 수행한다. 이후 몇 차례에 걸쳐 LeWin 블록을 통한 주목 맵 추출과 다운샘플링을 반복한다. 이때 LeWin 블록은 모든 패치 간의 전역 자기 주목을 모두 계산하여 장기의존성 문제를 해결하면서 지역적인 컨텍스트를 캡처하도록 설계되었으며, LeWin 블록으로부터 추출된 주목 맵은 디코더로 스킵 연결된다. 그리고 보틀넥 (Bottleneck)에 위치한 LeWin 블록까지 거친 이후에 디코더가 수행된다. 디코더에서는 변조기 (Modulator)로부터 생성된 변조 정보와 함께 전치 합성곱을 이용한 업샘플링과 LeWin 블록을 통한 주목 맵 추출이 반복되어 정답 영상의 손상을 보상하기 위한 특징 맵이 생성되고, 해당 특징 맵을 입력 영상에 더하여 열화가 제거된 영상을 추출한다. 그리고 Uformer의 손실 함수는 다음과 같이 나타낼 수 있다.

$$Loss = \sqrt{\|y' - \hat{y}\|_2 + \epsilon^2} \tag{12}$$

이때 열화가 제거된 영상은 $\hat{y}$고, y'는 정답 영상이며, ϵ는 하이퍼 파라미터이다.

2. 초해상화 기술

Liang 등.[65]은 얕은 특징과 깊은 특징을 추출하여 낮은 주파수 정보를 보존하면서 초해상화를 수행하는 SwinIR을 제안하였다. SwinIR은 얕은 특징 추출 모듈, 깊은 특징 추출 모듈, 고해상도 영상 복원 모듈로 구성되어 있다. 얕은 특징 추출 모듈은 합성곱 계층을 사용하여 얕은 특징을 추출하며, 복원 모듈에 전달되어 낮은 주파수 정보가 보존되도록 한다. 깊은 특징 추출 모듈은 RSTB (Residual swin transformer block)로 구성되어 있으며, 이는 지역적 주목과 교차 윈도우 반복 방법이 적용된 여러 개의 스윈 트랜스포머와 이전 스윈 트랜스포머와 잔차 연결을 통해 깊은 특징을 추출한다. RSTB를 통해 연산량은 줄이면서 여러 레벨의 특징을 결합할 수 있다. 깊은 특징과 얕은 특징은 결합 후 복원 모듈의 업샘플링 계층을 통해 최종 고해상도 영상을 출력한다. 손실 함수는 L1 손실 함수를 사용하며 수식은 아래와 같다.

$$Loss = \frac{1}{M}\sum_{i=1}^{M}\|HR - SR\|_1 \tag{13}$$

여기서 HR은 원본 고해상도 영상이며, SR은 출력된 고해상도 영상이다.

Tian 등.[66]은 계층적인 저주파수의 특징을 추출하고 이를 고주파수의 특징으로 변환하여 고해상도 영상을 추출하는 LESRCNN을 제안하였다. LESRCNN은 IEEB (Information extraction and enhancement block), RB (Reconstruction block), IRB (Information refinement

block) 순서의 3개 블록으로 구성된다. IEEB는 적은 채널의 수를 유지하여 연산량을 줄이면서 잔차 연결을 통해 작은 수의 채널로도 정보를 보존하여 저주파 영역의 특징을 추출한다. RB는 전역적, 지역적 특징을 융합하여 저주파 영역의 특징을 고주파 영역의 특징으로 변환한다. 이 단계에서 업샘플링을 수행하며, 이때 타깃 해상도에 따른 부화소 합성곱 계층을 다르게 하여 연산량을 줄였다. 이후 IRB를 통해 더 정교한 고주파수 성분을 추출하여 업샘플링된 특징과 결합하여 최종 고해상도 영상을 추출한다. 손실 함수로는 MSE 손실 함수를 사용하며 수식은 식 (14)와 같다.

$$Loss = \frac{1}{M}\sum_{i=1}^{M}\|HR - SR\|_2 \tag{14}$$

Chadha 등.[67]은 생성기 (Generator)에서 다수개의 저해상도의 입력 프레임을 순차적으로 특징을 추출하고 다음 프레임의 특징에 투영하여 고해상도 영상을 예측하고, 판별기 (Discriminator)를 통해 예측된 고해상도 영상 여부를 판별하도록 학습하는 네트워크로 구성된 GAN 기반의 iSeeBetter를 제안하였다. 생성기는 다수개의 SISR을 수행하는 모듈과 각 프레임에서 SISR이 수행된 결과를 결합하는 투영 모듈로 구성되어 있다. 생성기와 판별기는 서로 적대적으로 학습하여 높은 성능을 나타낸다. 손실 함수는 MSE 손실 함수, 인식 손실 함수 (Perceptual loss function), 적대적 손실 함수 (Adversarial loss function), TV 손실 함수 (Total-variation loss function)로 구성되어 있다. MSE 손실 함수는 정답 고해상도 영상 과 네트워크를 통해 예측된 초해상화 영상 간의 평균 제곱 오차를 측정하는 손실 함수로 식 (15)와 같다. 인식 손실 함수는 HR과 SR을 VGG 특징 공간에서의 평균 제곱 오차를 측정하는 손실 함수이며 식 (16)에 나타나 있다. TV loss 함수는 식 (17)에 나타나 있으며, 예측된 고해상도 영상 공간에서 수직, 수평 축으로의 분산을 계산하며, 예측된 고해상도 영상이 공간 축으로 부드럽게 되도록 출력하여 열화가 제거되는 효과를 얻는다. 적대적 손실 함수는 생성기를 통해 출력되는 초해상화 영상이 판별기의 입력이 들어갔을 때 학습된 판별기의 입력 초해상화 영상을 정답이라고 판단하도록 손실 함수를 정의하는 것으로 식 (18)과 같다.

$$L_{mse} = \frac{1}{W}\frac{1}{H}\sum_{i=1}^{W}\sum_{j=1}^{H}\|HR_{i,j} - SR_{i,j}\|_2 \tag{15}$$

$$L_{per} = \frac{1}{W}\frac{1}{H}\sum_{i=1}^{W}\sum_{j=1}^{H}\|VGG(HR_{i,j}) - VGG(SR_{i,j})\|_2 \tag{16}$$

$$L_{TV} = \frac{1}{W}\frac{1}{H}\sum_{i=1}^{W}\sum_{j=1}^{H}\sqrt{\|SR_{i,j+1} - SR_{i,j}\|_2 - \|SR_{i+1,j} - SR} \tag{17}$$

$$L_{ad} = -\log\left(D_{\theta_D}(SR)\right) \tag{18}$$

$$Loss = \lambda_0 L_{mse} + \lambda_1 L_{per} + \lambda_2 L_{TV} + \lambda_3 L_{ad} \tag{19}$$

여기서 W와 H는 x축과 y축 픽셀 좌표를 의미하며, λ_0, λ_1, λ_2, λ_3는 각각 하이퍼 파라미터이다.

Behjati 등.[68]은 임의의 스케일 팩터 (Scale factor)에 대한 SISR을 수행하는 OverNet을 제안하였다. 네트워크는 저해상도의 입력 영상으로부터 유용한 특징을 추출하는 특징 추출 모듈과 타깃 해상도 N에 대해 추출한 특징을 N+1 크기로 오버 스케일링 (Overscaling)을 수행한 뒤 이를 사용하여 N 크기의 고해상도 영상을 생성하는 오버 스케일링 모듈로 구성되어 있다. 특징 추출 모듈은 특징 채널 간의 종속성을 명시적으로 모델링한 SE (Squeeze-and-excitation) 동작을 수행하여 유용한 특징을 추출한다. 추출된 특징은 바이큐빅 업, 다운샘플링을 수행하는 오버 스케일링 모듈을 통해 최종적으로 고해상도 영상을 얻게 된다. 손실 함수는 다중 스케일 손실 함수 (Multi-scale loss)를 사용하며, 이는 식 (20)과 같다.

$$\begin{aligned} &Loss = \sum_{s\in S}\|SR_s - bicubic\ (HR, s)\|_1, \\ &where\ S = \{s_1,\ s_2\ \cdots s_n\},\ s_i\ \ N \end{aligned} \tag{20}$$

여기서 S 는 스케일 팩터 집합이며, $bicubic_{\downarrow}$은 바이큐빅 다운샘플링을 의미한다. 다중 스케일 손실 함수를 사용하여 서로 다른 타깃 크기로부터 오는 부가적인 정보를 학습하여 성능을 향상시켰다.

Behjati 등.[69]은 저해상도의 입력 영상을 주파수 대역에 따라 저주파와 고주파로 분류하고, 주파수 대역을 기반으

로 낮은 연산량으로 높은 품질의 고해상도 영상을 추출하는 FENet을 제안하였다. FENet은 비선형 모듈과 복원 모듈로 구성된다. 비선형 모듈은 특징을 서로 다른 주파수의 특징으로 분해할 수 있다는 가정하에, 각 성분을 다르게 처리하기 위해 파라미터 수와 비선형 계층의 수를 다르게 처리하여 고주파 성분과 저주파 성분을 분해한 특징을 추출한다. 또한, 고주파 성분 추출 시 풀링과 바이큐빅 업샘플링을 사용하여 고주파 성분을 강조한 특징을 추출한다. 이후 복원 모듈에서 저주파 성분의 특징과 고주파 성분의 특징을 결합하여 최종 고해상도 영상을 추출한다. 손실 함수는 L1 손실 함수를 사용하며, 식 (21)에 나타난다.

$$Loss = \frac{1}{M}\sum_{i=1}^{M}\|HR - SR\|_1 \tag{21}$$

3. 프레임 보간 기술

Huang 등.[70]은 RIFE 네트워크에서 코어스 투 파인 전략 (Coarse-to-fine strategy)을 통해 점진적으로 프레임의 해상도를 높여 나가며 플로우를 예측하는 IFNet을 제안하였다. RIFE는 학습 시 더 많은 IFBlock을 갖는 티처 모델 (Teacher model)을 통한 프리빌리지드 지식 증류 (Privileged knowledge distillation)을 통해 성능을 향상시킨다. RIFE는 두 이미지를 퓨전 맵 (Fusion map)을 통해 아래의 수식으로 가중합하여 최종 추론된 이미지를 얻는다. 이는 식 (22)와 같이 표현할 수 있으며, 이렇게 생성된 $\hat{I}_t$를 입력으로 RefineNet을 통해 추가적인 고주파 성분을 복원하고 $\hat{I}_t$에 더해주어 스튜던트 모델 (Student model) 출력의 열화를 줄인다.

$$\hat{I}_t = M \odot \hat{I}_{t\ \ 0} + (1 - M) \odot \hat{I}_{t\leftarrow 0} \tag{22}$$

여기서 M은 $0 \leq M \leq 1$의 범위를 가지며, $\odot$는 원소 별 곱셈을 의미한다. 또한, IFNet에서 하나의 IFBlock을 통과할 때마다 특징 맵의 크기가 2배씩 커지게 된다. 각각의 IFBlock은 광학 흐름을 점진적으로 업데이트하며, 학습 시 사용되는 티처 모델의 경우 4개의 IFBlock을 사용하고, 스튜던트 모델의 경우 IFBlock의 개수가 3개이다. IFNet의 학습을 위한 손실 함수는 다음 식과 같이 정의된다.

$$TotalLoss = L_{rec} + L_{rec}^{Tea} + \lambda_d L_{dis} \tag{23}$$

$$L_{rec} = d(\hat{I}_t, I_t^{GT}), L_{rec}^{Tea} = d(\hat{I}_t^{\ Tea}, I_t^{GT}) \tag{24}$$

$$L_{dis} = \sum_{i\in\{0,1\}} \| F_{t\rightarrow i} - F_{t\rightarrow i}^{Tea} \|^2 \tag{25}$$

여기서 λ_d은 0.01을 사용하며, L_{rec}은 스튜던트 모델을 통해 추론된 최종 이미지와 정답 이미지를 각각 5개의 레벨의 라플라시안 피라미드 (Laplacian pyramid)로 구성했을 때의 L1 픽셀 손실을 의미하며, L_{rec}^{Tea}은 티처 모델에서 추론된 결과와 최종 이미지와의 5개의 레벨의 라플라시안 피라미드 공간에서의 L1 픽셀 손실을 의미한다. L_{dis}의 경우 티처 모델이 I_t^{GT}을 입력으로 추론한 광학 흐름과 스튜던트 모델에서 추론된 광학 흐름의 L2 손실을 측정하여 스튜던트 모델이 프리빌리지드 지식 증류를 수행할 수 있도록 한다.

Hu 등.[71] 은 프레임 쌍 간의 다양한 양방향 플로우를 추정 후 합성하여 보간을 수행하는 Many-to-many splatting 방법을 제안하였다. 스플래팅 (Splatting)은 다중 화소를 단일 화소에 매핑하는 방식으로 와핑 결과에 불필요한 구멍 (Hole)이 생기는 문제를 해결하였다. 입력 프레임 쌍이 주어지면, 기존 (Off-the-shelf) 방법으로 양방향 움직임 (Bidirectional motion)을 예측한다. 그리고 이 움직임 예측값을 움직임 보정 네트워크 (Motion refinement network)에 입력하여 다양한 모션 벡터 (Motion vector) 값을 예측하고, 입력 프레임의 각 픽셀 당 신뢰도 점수 (Reliability score)를 추정한 뒤 병합한다. 모션 특징 인코딩 (Motion feature encoding), 저차원 특징 변조 (Low-rank feature modulation), 출력 디코딩 (Output decoding)으로 구성되는 MRN (Motion refinement network) 파이프라인은 각 픽셀 당 다수 개의 모션 벡터를 예측하면서 기존 광학 흐름을 업샘플링 및 보정하는 역할을 수행한다. 손실 함수는 식 (26)과 같이 Charbonnier loss와 census loss의 합으로 사용한다.

$$L = L_{char} + L_{cen} \tag{26}$$

Sim 등.[72]은 큰 움직임이 포함된 4K 해상도의 영상을 다루기 위해 recursive multi-scale shared structure에 기반한 XVFI-Net을 제안하였다. XVFI-Net은 양방향 광학 흐름을 위한 두 개의 연속된 모듈 (Cascaded module)인 BiOF-I와 BiOF-T로 구성되며, 광학 흐름은 CFR (Complementary flow reversal)에 의해 근사화 된다. 추론 단계에서 BiOF-I 모듈은 입력의 다양한 스케일 (Scale)에서 시작할 수 있으며, BiOF-T 모듈은 프레임 보간의 정확성을 높이면서도 추론 과정이 가속화될 수 있도록 기본 입력 스케일에서 동작한다. 기존 PWC-Net과 같은 고정 숫자의 스케일 레벨의 구조는 다양한 공간적 해상도에 적응하는데 있어 증가한 스케일 깊이를 재학습하였다. 반면, XVFI-Net은 스케일 적응성을 갖추기 위해 어떠한 코어스 스케일 레벨 (Coarse scale level)에서도 광학 흐름을 추정하도록 설계되었다. 따라서 XVFI-Net은 다른 스케일 레벨끼리 파라미터를 공유한다. 또한, XVFI-Net은 큰 움직임을 잘 감지할 수 있도록 스트라이드 합성곱 (Strided convolution)을 사용하여 공간적 해상도가 줄어든 특징을 문맥 특징으로 변환하여 사용하며, 이 문맥 특징 맵은 스케일에 따라 재귀적으로 다운 스케일링이 가능하다. 손실 함수는 멀티 스케일 복원 손실 (Multi-scale reconstruction loss)과 1차 에지 감지 스무드니스 손실 (First-order edge-aware smoothness loss)의 가중치 합으로 구한다.

Kong 등.[73]은 중간 플로우를 추정한 다음 컨텍스트 기능을 추정하는 파이프라인의 문제를 해결한 IFRNet을 제안하였다. IFRNet은 소형화 및 빠른 추론을 위해 광학 흐름 및 문맥 특징을 별도의 인코더-디코더로 나누지 않고, 단일 인코더-디코더 기반 모델로 병합한다. 인코더에 의해 입력으로부터 피라미드 특징을 추출하고, 코어스 투 파인 디코더를 통해 중간 특징과 양방향 중간 플로우 필드 (Bilateral intermediate flow fields)를 공동으로 보정한다. 따라서, 물체의 움직임이 더 선명해지고 텍스처 세부 정보를 얻을 수 있다. 손실 함수는 다음과 같이 세 가지로 구성된다.

$$L_r = \rho(\hat{I}_t - I_t^{gt}) + L_{cen}(\hat{I}_t, I_t^{gt}) \tag{27}$$

$$L_d = \sum_{k=1}^{3}\sum_{l=0}^{1}\rho(u_{2^k}(F_t^{k}{}_1) - F_{t\to 1}^{p}) \tag{28}$$

$$L_g = \sum_{k=1}^{3} L_{cen}(\hat{\Phi}_t^{k}, \Phi_t^{k}) \tag{29}$$

IFRNet은 종단 간 훈련이 가능하며, 중간 프레임을 생성하기 위해 네트워크 출력 $\tilde{\Phi}_t$과 정답 영상 Φ_t^{gt} 사이에 영상 재구성 손실 함수 L_r을 사용한다. 또한, 프레임 합성 시 조도의 경우 로컬 미니멈 (Local minimum)에 최적화될 가능성이 있어 지식 증류 전략을 사용한다. 하지만, 무차별적인 증류 방식은 원하지 않는 영상 열화를 학습할 수 있기 때문에, 작업 중심 플로우 증류 손실 (Task-oriented flow distillation loss)을 제안하였다. 제안하는 손실 함수는 식 (28)과 같은 일반화된 Charbonnier 손실 함수이다. 여기서 u_s는 스케일 계수가 있는 이중 선형 보간 연산이다. 그 외에도, 개선된 중간 프레임 특징을 얻기 위해 식 (29)와 같이 특징 공간 기하학 일관성 손실 함수 (Feature space geometry consistency loss)를 추가한 특징을 갖고 있다.

Park 등.[74]은 비대칭 양방향 움직임 추정 (Asymmetric bilateral motion estimation) 모듈과 프레임 합성 모듈로 구성된 역방향 와핑을 기반으로 하는 프레임 보간 네트워크인 ABME를 제안하였다. ABME는 먼저 비대칭 양방향 움직임 추정 모듈에서 대칭 양방향 움직임 필드를 통해 임시 중간 프레임인 앵커 프레임 (Anchor frame)을 보간한다. 이때, 경계 영역에서 폐색으로 인한 오류를 방지하기 위해 마스크를 활용하여 폐색 인식 방식으로 앵커 프레임을 재구성한다. 이후 앵커 프레임에서 두 입력 프레임으로 비대칭 양방향 움직임 필드를 추정한다. 프레임 합성 모듈에서 입력 프레임은 양방향 움직임 필드를 통해 와핑한다. 이러한 와핑 프레임을 위해, FilterNet과 RefineNet으로 구성되어 있으며, FilterNet은 동적 필터를 생성하여 지역 정보를 활용하고, RefineNet은 전역 정보를 사용하여 잔차 프레임을 생성한다. 손실 함수의 경우 비대칭 양방향 움직임 추정 모듈을 위한 손실 함수 와 프레임 합성 모듈을 위한 손실 함수 두 가지로 구성되어 있다. 이는 각각 식 (30), 식 (31)과 같다.

$$L_{pho}=\rho\left(I_t^{GT}-\Phi_B\left(V_t^A{}_0,I_0\right)\right)+\rho\left(I_t^{GT}-\Phi_B\left(V_{t\to1}^A,I_1\right)\right)+ L_{cen}\left(I_t^{GT}-\Phi_B\left(V_{t\to1}^A,I_1\right)_{cen}\left(I_t^{GT}-\Phi_B\left(V_{t\to0}^A,I_0\right)\right)\right) \quad (30)$$

$$L_{sum}=\rho\left(I_t^{GT}-I_t\right)+ L_{cen}\left(I_t^{GT}-I_t\right) \quad (31)$$

여기서 $\rho(x)=(x^2+\epsilon^2)$는 Charbonnier 함수이고, L_{cen}은 Census 손실 함수이다. Φ_B는 역방향 와핑 함수이고, I_t는 추출된 보간 프레임이며, I_0과 I_1은 타깃 프레임이다. 그리고 $V_{t\to0}^A, V_{t\to1}^A$는 비대칭 양방향 움직임 필드이다.

Ⅲ. 실험 결과

본 장에서는 전송 및 저장 응용 시스템에서 영상 개선 기술들에 대한 실험을 수행하고 그 적합성을 평가하기 위해 표 1과 같이 실험 환경을 구축하고, 다양한 분야 중 하나의 예시로 해상 영상에 대한 객관적 및 주관적 화질 평가와 네트워크 복잡도에 대한 실험을 진행하여 최적의 모델을 선정하고자 한다. 해상 영상의 경우 1920×1080 크기인 총 5400장의 영상을 직접 취득하였고, 7:2:1의 비율로 나누어 학습 데이터 셋 3780장, 검증 데이터 셋으로 1080장, 실험 데이터 셋으로 540장을 구성하였다.

표 1. 실험 환경 구축을 위한 전송 및 저장 시스템 장치 구성
Table 1. Configuration of transmission and storage system devices for establishing the experimental environment

Components	Products
CPU	Intel® Core™ i9-10980XE Gold @ 3.00GHz
Memory	Samsung DDR4 25600 32GB (×8)
Storage	Samsung SSD 980PRO 2TB (×3)
GPU	Nvidia RTX 3090 (×4)

본 논문에서 진행한 실험에서는 사전 학습 모델에 대해 해상 영상을 이용하여 전이 학습을 수행하였다. 이 과정에서 손실 함수, 네트워크 구조, 배치 크기 등의 환경은 각 기술에서 사용된 환경과 동일하게 구성하였다. 영상 개선 기술 중, 열화 제거 및 초해상화 기술의 성능 평가를 위한 실험 영상은 1920×1080 해상도의 원본 영상 데이터에 대해 FFmpeg[75]을 사용하여 타깃 비트레이트를 200kbps, 300kbps, 500kbps로 설정하고, 영상의 해상도를 바이큐빅 다운샘플링을 통해 변경한 후, x.265을 이용하여 압축 및 복원을 수행하여 생성하였다. 보다 구체적으로, 타깃 비트레이트를 200kbps로 설정한 경우, 480×270 크기의 해상도를 가진 영상으로 다운샘플링을 수행하였고, 타깃 비트레이트를 300kbps와 500kbps로 설정한 경우, 960×540 크기의 해상도를 가진 영상으로 다운샘플링을 수행하였다. 또한, 각 기술의 평가를 위한 정답 영상을 위해 원본 영상에 바이큐빅 다운샘플링을 수행하여 각각 480×270 및 960×540 크기의 해상도를 가진 영상을 생성하였다. 프레임 보간 기술의 경우, 300 프레임의 원본 영상에 대해 홀수 번째 프레임과 짝수 번째 프레임만을 각각 추출한 150 프레임으로 구성된 2 세트의 영상들을 구성하였다. 따라서 홀수 번째 프레임으로만 구성된 15 fps 영상 및 짝수 번째 프레임으로만 구성된 15 fps 영상에 대한 실험을 수행하였다.

3.1절에서는 영상 내 열화 제거, 초해상화, 프레임 보간 기술들에 대하여 객관적 화질 평가를 하기 위해 각각의 기술들에 대해 수치적 지표인 PSNR (Peak signal to noise ratio)과 SSIM (Structural similarity index)을 비교하여 영상 품질의 복원 정확성을 평가하고, 각 모델들에 필요한 파라미터의 수와 메모리 사용량, 추론 시간을 측정하여 비교하였다. 3.2절에서는 각 기술들에 대한 주관적 화질 평가를 수행하여 인지 시각적인 측면에서 영상의 품질을 평가하였다. 3.3절에서는 실제 응용에 적용될 수 있는 낮은 네트워크 복잡성을 요구하는 실시간 환경 및 높은 복원 정확성을 위한 고품질 환경을 정의하고, 앞서 수행한 실험을 통해 비교 및 평가한 결과를 토대로 각 한경에 적합한 기술들을 선정하였다.

1. 객관적 화질 및 네트워크 복잡도 평가

표 2는 해상 영상에 대한 압축 열화 제거 기술들의 정량적 지표를 정리한 표이며, 조건 별 가장 좋은 성능을 나타내는 지표는 붉은색으로 표기하였다. 480270 해상도 영상의 실험 결과, PSNR은 WRCAN, InvDN, PSTQE, Stripformer, Uformer 모델 순으로 높았으며, SSIM은

WRCAN이 가장 높은 성능을 보였고, 나머지 기술은 비슷한 수준의 성능을 보인다. 추론 시간의 경우 PSTQE, InvDN, Stripformer가 0.03~0.04 (초/프레임)의 프레임 당 추론 시간을 보여주었으며, 메모리 사용량의 경우 InvDN, PSTQE, Stripformer, WRCAN, Uformer 순으로 적게 사용된 것을 확인하였다. 960×540 해상도 영상의 실험 결과, PSNR과 SSIM에 대하여 각각 WRCAN이 39.6 (dB)과 0.96으로 가장 높았고, PSTQE와 InvDN이 38.2~38.3 (dB)와 0.93 수준의 성능을 보여주었으며, Stripformer와 Uformer가 37.1~37.2 (dB) 및 0.92 수준의 성능을 나타냈다. 또한, 960×540 해상도 영상의 실험 결과는 WRCAN의 PSNR이 39.8 (dB)로 가장 좋은 성능을 보였으며, PSTQE와 InvDN은 38.7~38.9 (dB), Stripformer와 Uformer는 37.6~37.9 (dB)의 성능을 보여주었다. SSIM 지표로는 WRCAN이 0.97 수준으로 가장 성능이 높았고, 나머지 4개의 기술들은 0.94 및 0.95로 성능이 유사한 것을 확인하였다.

또한, 960×540 해상도 영상에 대하여 메모리 사용량은 PSTQE, InvDN, WRCAN, Stripformer, Uformer 순으로 많았고, 480×270 해상도 영상의 실험과 비교했을 때, 프레임 대비 Stripformer와 Uformer가 필요로 하는 메모리 사용량이 2배 이상 증가하였다. 추론 시간의 경우 Stripformer가 모든 조건에서 가장 짧은 시간이 소요되었고, 그 다음으로 InvDN과 PSTQE의 추론 시간이 짧았다. 또한, 각 조건에

표 2. 압축 열화 제거 모델 별 해상 영상에 대한 복원 정확성 및 네트워크 복잡도 측정 결과
Table 2. Reconstruction accuracy and network complexity of denoising model for maritime dataset

Video resolution	Model	PSNR (dB)	SSIM	Inference time (s/f)	Usage memory (GB)	Params (K)
480×270 (200kbps)	WRCAN	38.5	0.93	1.75	3,06	156974
	PSTQE	37.9	0.89	0.04	1.81	2174
	InvDN	38.0	0.89	0.03	1.72	2641
	Stripformer	37.0	0.91	0.03	2.58	19708
	Uformer	36.5	0.90	0.51	4.21	50880
960×540 (300kbps)	WRCAN	39.6	0.96	4.71	3.82	156974
	PSTQE	38.3	0.93	0.09	1.81	2174
	InvDN	38.2	0.93	0.07	1.92	2641
	Stripformer	37.1	0.92	0.04	5.69	19708
	Uformer	37.2	0.92	1.12	8.95	50880
960×540 (500kbps)	WRCAN	39.8	0.97	4.77	3.82	156974
	PSTQE	38.7	0.94	0.09	1.81	2174
	InvDN	38.9	0.95	0.07	1.92	2641
	Stripformer	37.9	0.94	0.04	5.69	19708
	Uformer	37.6	0.94	1.12	8.95	50880

표 3. 초해상화 모델 별 해상 영상에 대한 복원 정확성 및 네트워크 복잡도 측정 결과
Table 3. Reconstruction accuracy and network complexity of super-resolution model for maritime dataset

Video resolution	Model	PSNR (dB)	SSIM	Inference time (s/f)	Usage memory (GB)	Params (K)
480×270 (200kbps)	SwinIR	36.1	0.91	0.880	11.6	11752
	LESRCNN	37.1	0.89	0.004	6.4	774
	iSeeBetter	36.9	0.89	0.009	21.6	11900
	OverNet	36.4	0.89	0.067	4.8	1078
	FENet	36.6	0.89	0.015	3.7	675
960×540 (300kbps)	SwinIR	37.3	0.93	3.610	15.3	11752
	LESRCNN	37.2	0.91	0.004	8.1	626
	iSeeBetter	37.3	0.92	0.009	23.3	12438
	OverNet	37.1	0.91	0.334	9.2	1078
	FENet	37.2	0.92	0.126	8.2	675
960×540 (500kbps)	SwinIR	38.8	0.95	3.620	15.3	11752
	LESRCNN	37.4	0.90	0.004	8.1	626
	iSeeBetter	37.6	0.94	0.009	23.3	12438
	OverNet	38.5	0.94	0.334	9.2	1078
	FENet	38.3	0.95	0.126	8.2	675

대해 네트워크에 필요한 파라미터 수는 동일하였으며, PSTQE이 가장 적은 파라미터 수를 필요로 하였으며, 가장 많은 파라미터 수를 이용하는 WRCAN 대비 PSTQE가 약 1.38%의 파라미터 수만을 필요로 하였다.

표 3은 해상 영상에 대한 초해상화 기술들의 정량적 지표를 정리한 표이다. 480×270의 해상도를 가진 입력 프레임을 가로, 세로 방향 4배 크기의 해상도로 높이는 경우, PSNR은 LESRCNN이 37.1 (dB)로 가장 성능을 보였다. 또한, SSIM은 SwinIR이 0.91로 가장 높고 이외 기술은 모두 0.89를 기록하였다. 추론 시간은 LESRCNN이 0.004 (초/프레임)으로 가장 빠른 성능을 보였고, 다음으로는 iSeeBetter이 0.009 (초/프레임)의 성능을 보였다. 두 모델을 제외한 다른 기술은 모두 0.01 (초/프레임) 이상의 시간이 소요되었다. 메모리 사용량의 경우 FENet이 3.7 (GB)으로 가장 적은 메모리를 사용하였다. 또한, FENet, LESRCNN, OverNet 순으로 네트워크 파라미터 수가 가장 적었고, SwinIR과 iSeeBetter는 상대적으로 많은 네트워크 파라미터 수가 필요한 것을 확인하였다. 960×540의 해상도를 가진 입력 프레임을 가로, 세로 방향 2배 크기의 해상도로 높이는 경우, 타깃 비트레이트에 따라 약간의 성능 차이를 보였지만, 각 모델들의 경향은 유사한 것을 확인하였다. PSNR 및 SSIM 모두 SwinIR이 가장 높은 수치를 보였지만 추론 시간은 가장 오래 소요되었다. 메모리 사용량은 LESRCNN, FENet, OverNet, SwinIR, iSeeBetter 순으로, 네트워크 파라미터 수는 LESRCNN, FENet, OverNet, SwinIR, iSeeBetter 순으로 적은 것을 확인하였다. 종합적으로 비교해보았을 때, LESRCNN이 우수한 품질 향상을 기록함과 동시에 낮은 네트워크 복잡도를 갖는 것을 확인하였다.

표 4는 해상 영상에 대한 프레임 보간 기술들의 정량적 지표를 정리한 표이다. PSNR 지표로는 ABME가 41 (dB)를 기록하여 가장 높은 복원 정확도를 보여주었다. SSIM은 M2M_PWC가 0.96으로 가장 높았고, IFRNet과 ABME가 0.94, XVFI가 0.91, RIFE가 0.83을 기록하였다. 또한, 추론 시간은 RIFE가 0.018 (초/프레임), M2M_PWC가 0.023 (초/프레임), IFRNet이 0.105 (초/프레임), XVFI가 0.382 (초/프레임), ABME가 1.438 (초/프레임)으로 측정되었다. 메모리 사용량은 M2M_PWC가 3.0 (GB)로 가장 적게 사용되었고 XVFI가 19.4 (GB)의 메모리를 사용하여 메모리 사용량이 가장 많았으며, 네트워크 파라미터 수는 ABME, RIFE, IFRNet, XVFI, M2M_PWC 순으로 적은 수치를 보였다. 종합적으로 비교해보았을 때, M2M_PWC가 파라미터 수를 제외한 모든 수치에서 우수한 성능을 보였다.

2. 주관적 화질 평가

그림 2-4는 3.1절에서 비교한 각 영상 기술에 대한 모델들에 대해, 주관적 화질 비교를 수행한 그림이다. 표 1 및 표 2를 보면, 각 모델들에 대해 PSNR 및 SSIM이 유사한 결과를 얻었고 이에 따라 주관적인 화질도 거의 모든 기술이 유사한 것을 확인할 수 있다. 하지만, 표 3을 보면 프레임 보간 기술들을 비교했을 때, PSNR 관점에서 가장 좋은 성능을 보이는 ABME와 가장 낮은 성능을 보이는 RIFE의 PSNR 차이가 10 (dB)이상 나지만, 주관적 화질은 거의 유사한 것을 확인할 수 있다. 이는 RIFE의 경우, 생성한 프레임의 구조적인 특성은 원본 영상과 유사하지만, 일부 프레임에서 플리커링 (Flickering) 현상이 발생하여 낮은 PSNR을 갖는 것으로 분석된다.

표 4. 프레임 보간 모델 별 해상 영상에 대한 복원 정확성 및 네트워크 복잡도 측정 결과
Table 4. Reconstruction accuracy and network complexity of frame interpolation model for maritime dataset

Model	PSNR (dB)	SSIM	Inference time (s/f)	Usage memory (GB)	Params (K)
RIFE	30.4	0.83	0.018	5.5	3038
M2M_PWC	39.1	0.96	0.023	3.0	7611
XVFI	38.8	0.91	0.382	19.4	5661
IFRNet	34.4	0.94	0.105	5.2	4960
ABME	41.0	0.94	1.439	16.4	995

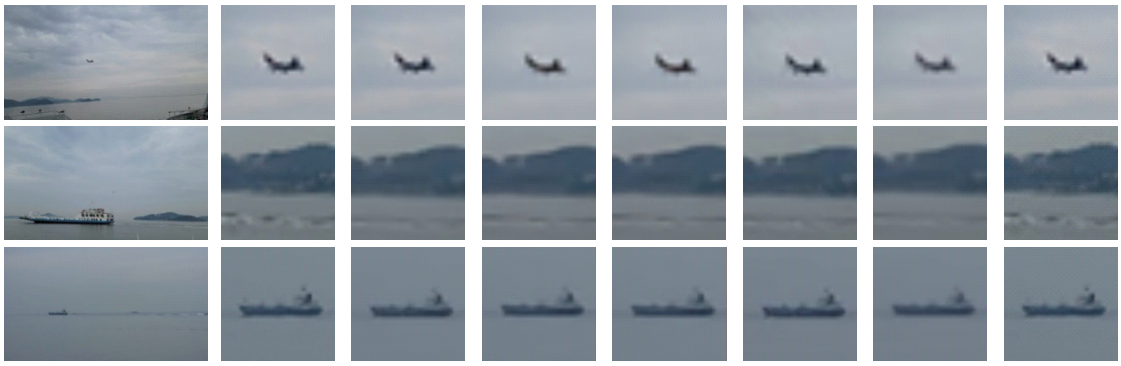

그림 2. 480×270 크기 해상도의 해상 영상에 대한 모델 별 열화 제거 주관적 화질 평가: (a) 입력 프레임, (b) 크롭된 입력 프레임 (c) WRCAN, (d) PSTQE, (e) InvDN, (f) Stripformer, (g) Uformer, (h) 크롭된 원본 프레임

Fig 2. Perceptual Assessment on denoising result of each model about maritime videos of 200kbps: (a) input frame, (b) cropped input frame, (c) WRCAN, (d) PSTQE, (e) InvDN, (f) Stripformer, (g) Uformer, (h) cropped original frame

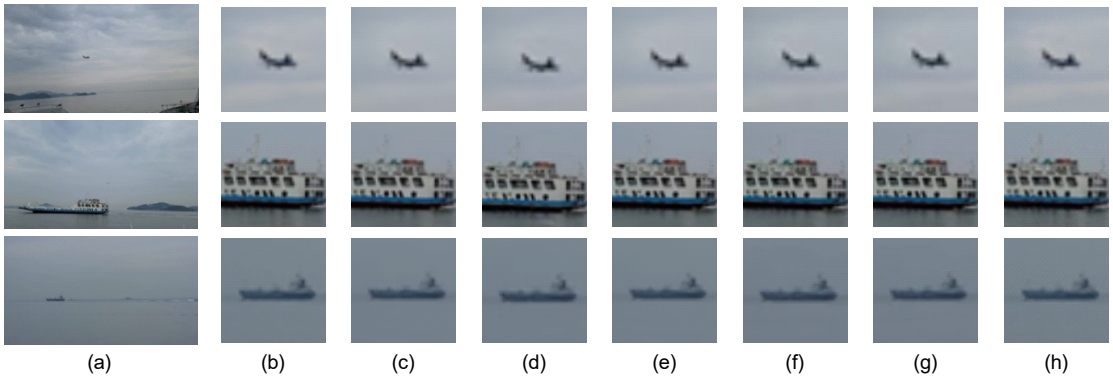

그림 3. 480×270 크기 해상도의 해상 영상에 대한 모델 별 초해상화 결과: (a) 입력 프레임, (b) 크롭된 입력 프레임 (c) SwinIR, (d) LESRCNN, (e) iSeeBetter, (f) OverNet, (g) FENet, (h) 크롭된 원본 프레임

Fig 3. Super-resolution result of each model about maritime videos of 200kbps: (a) input frame, (b) cropped input frame, SwinIR, (d) LESRCNN, (e) iSeeBetter, (f) OverNet, (g) FENet, (h) cropped original frame

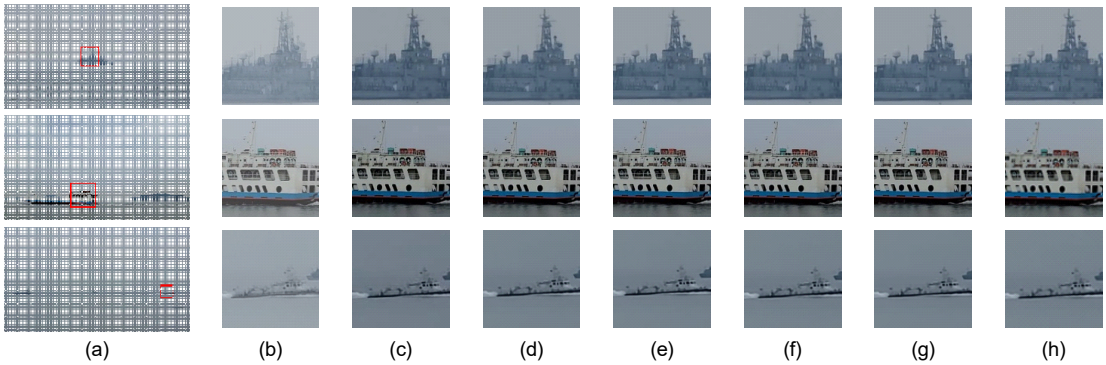

그림 4. 480×270 크기 해상도의 해상 영상에 대한 모델 별 프레임 보간 결과: (a) 겹쳐진 두 입력 프레임, (b) 크롭된 겹쳐진 두 입력 프레임 (c) RIFE, (d) M2M_PWC, (e) XVFI, (f) IFRNet, (g) ABME, (h) 크롭된 원본 프레임

Fig 4. Frame interpolation result of each model about maritime videos of 200kbps: (a) overlapped two input frames, (b) cropped input frames, (c) RIFE, (d) M2M_PWC, (e) XVFI, (f) IFRNet, (g) ABME, (h) cropped original frame

3. 기술 선정

본 절에서는 열화 제거, 초해상화, 프레임 보간 기술들에 대해 앞서 실험한 기술 별 다섯 가지 모델들의 실험 결과를 토대로 시스템이 요구하는 제한 사항에서 얼마나 높은 실시간성을 갖출 수 있는지를 평가하기 위해 실시간 환경과 고품질 환경을 정의하고, 환경 별 적합한 모델을 선정하기 위한 평가를 수행하였다. 실시간 환경의 경우, 15 fps의 1분 길이 영상이 입력되었을 때, 열화 제거 기술, 초해상화 기술, 프레임 보간 기술을 모두 수행하여 30 fps의 고해상도 영상이 1분 이내에 출력되어야 한다고 가정하였다. 이에 따라 실시간 모드의 경우 각 기술이 15 fps 이상으로 동작해야 하며, 이를 프레임당 처리속도로 환산한 경우 0.067초 이내로 처리를 수행해야 한다. 고품질 환경의 경우, 15 fps의 1분 길이 영상이 입력되었을 때, 열화 제거 기술, 초해상화 기술, 프레임 보간 기술을 수행하여 30 fps의 고해상도 영상이 12분 이내에 출력되어야 한다고 가정하였다. 이에 따라 각 기술은 1.25 fps 이상의 속도로 동작 시 해당 요구사항을 만족할 수 있으며 이를 만족하는 프레임 당 처리속도는 0.8초 이내를 만족해야 한다.

4. 실시간 모드

표 5는 세 가지 영상 개선 기술들 중 실시간 모드에 적합한 기술들을 선정하여 나타낸 표이다. 성능적인 면에서 다른 기술 대비 영상 품질이 떨어지지 않으면서 추론 시간이 상대적으로 빠르고 메모리 사용량 및 파라미터 수를 적게 요구하는 기술들을 선별하였다. 종합적인 처리 소요 시간이 최대 0.067초 내로 처리해야 하므로, 열화 제거 기술들 중 최대 0.04초의 처리속도를 내는 Stripformer가 실시간 모드에 적합하다. Stripformer는 객관적 성능 측면에서도 SSIM 0.91 이상의 성능을 내며, 주관적 화질 측면에서 육안으로 비행기나 배와 같은 타깃을 탐지하기에 문제가 없음을 확인하였다. 추론 시간 면에서 실시간 동작 조건을 만족하는 초해상화 네트워크는 LESRCNN, iSeeBetter이다. 둘 중, LESRCNN의 경우 파라미터의 수가 iSeeBetter 네트워크 대비 15배 이상 작으며 메모리 사용량이 1/3 수준으로 또 다른 영상 개선 네트워크와 동시에 GPU 메모리에 적재하여 사용할 경우 자원 효율성이 있어 실시간 모드로 LESRCNN이 적합하다. 프레임 보간 기술에서는 M2M_PWC 모델 또는 IFRNet 모델의 추론 시간이 각각 프레임 한 장 단위 평균 0.024초와 0.105초가 소요되어 대용량 영상을 처리하는데 적합하며, 사용하는 파라미터 수도 적어 모델 적재에 요구되는 메모리 사용이 적었고, latency 관점에서 볼 때 짧은 시간 안에 출력이 가능하기 때문에 M2M_PWC와 IFRNet을 실시간 모드를 위한 프레임 보간 기술에 대한 적합성이 우수하다고 평가하여 선정하였다.

5 고품질 모드

표 6은 세 가지 영상 개선 기술들 중 고품질 모드에 적합한 기술들을 선정하여 나타낸 표이다. 객관적 성능 지표인 PSNR과 SSIM 수치가 모두 높고 주관적 화질도 우수했던 기술들 중에서 추론 시간과 메모리 사용량, 파라미터 수를 전반적으로 고려하여 고품질 모드에 적합한 기술로 선정하였다. 12분 이내에 1분 길이의 영상을 처리하기 위해서 프레임 한 장을 처리하는데 초해상화와 프레임 보간에 소요되는 시간을 고려하여 0.8초보다 빠르게 처리해야 하며, 해당 기준을 만족하는 열화 제거 모델은 PSTQE, InvDN, Stripformer이다. Stripformer의 경우 PSTQE와 InvDN 대비 SSIM은 비슷한 수준이나, PSNR의 경우 InvDN 대비 세 가지 입력 모두 1 (dB)의 차이가 나고 PSTQE와 비교해도 떨어지는 성능을 나타낸다. PSTQE와 InvDN차이는 객관적인 측면에서는 InvDN이 PSTQE 대비 0.1 (dB)가 높을

표 5. 실시간 모드를 위한 영상 개선 기술 선정
Table 5. Video enhancement technique for real-time mode

	Denoising	Super-resolution	Frame interpolation
Selected model	Stripformer	LESRCNN	M2M_PWC, IFRNet

표 6. 고품질 모드를 위한 영상 개선 기술 선정
Table 6. Video enhancement technique for high-quality mode

	Denoising	Super-resolution	Frame interpolation
Selected model	InvDN	FENet	M2M_PWC

뿐만 아니라 SSIM 역시 0.1 높게 나타나는 것을 확인하였다. 이를 바탕으로 고품질 모드에 적합한 모델로 InvDN을 선정하였다. 또한, 고품질 모드의 제약 조건을 만족하는 초해상화 네트워크는 LESRCNN, iSeeBetter, OverNet, FENet이다. 이에 따라, 고품질 모드 초해상화 기술로 FENet이 적합하다. 프레임 보간 기술에 대해, SSIM의 수치는 M2M_PWC 모델이 0.97로 가장 높았으며, 그 다음으로 IFRNET, ABME 순으로 높은 결과를 보였고, PSNR 수치는 ABME, M2M_PWC, IFRNET 순으로 높은 것을 앞선 결과에서 확인하였다. IFRNet은 SSIM 수치가 높았지만, PSNR, 추론 시간, 메모리 사용량, 파라미터 수에서 상대적으로 낮은 성능을 보였으며, ABME는 PSNR과 SSIM 수치 모두 높았지만, 추론 시간이 길었다. 따라서, 객관적인 수치와 주관적인 화질 평가를 기반으로 M2M_PWC 모델을 고품질 모드를 위한 프레임 보간 기술로 선정하였다.

Ⅳ. 결 론

본 논문에서는 영상 압축 기술을 활용하는 전송 및 저장이 필요한 응용 시스템에서 필요한 영상 개선 기술에 대해, 대표적인 기술인 영상 개선 기술로 영상 내 열화 제거 기술, 초해상화 기술, 프레임 보간 기술에 대한 분석을 수행하고, 특히 높은 영상 품질저하가 발생할 수 있는 해양 감시 정찰 분야에 대해 전송된 해상 영상의 품질을 개선하는데 적합한 영상 개선 기술의 선별을 위한 실험을 수행하였다. 영상 개선의 복원 정확성을 판단하는 기준으로 PSNR과 SSIM 지표를 사용하고, 네트워크의 복잡성을 평가하는 기준으로 추론 시간과 메모리 사용량, 네트워크 파라미터를 사용하여 비교분석을 수행하였고, 실제로 저장 및 전송을 포함한 응용 시스템에 대해 대표적으로 적용될 수 있는 고품질 및 실시간 환경에 대해 정의하여 환경에 따른 기술 적합성을 평가하였다. 이에 따라 실시간성 모드로 각 카테고리 별 Stripformer, LESRCNN, M2M_PWC 및 IFRNet을 선정하였고, 고품질 모드로 각각 InvDN, FENet, M2M_PWC 모델을 선정하였다. 본 논문에서 제공하는 분석 결과는 향후 영상 개선 기술을 적용하는데 있어 조건에 따른 기술 선택에 도움을 줄 것으로 기대된다.

참 고 문 헌 (References)

[1] D. Ding, W. Wang, J. Tong, X. Gao, Z. Liu and Y. Fang, "Biprediction-Based Video Quality Enhancement via Learning," IEEE Transactions on Cybernetics, Vol.52, No.2, pp.1207-1220, Feb. 2022. doi: https://doi.org/10.1109/TCYB.2020.2998481

[2] G.J. Sullivan, J.R. Ohm, W.J. 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

[3] J. Lee, J. Park, H. Choi, J. Byeon, D. Sim, “Overview of VVC,” Broadcasting and Media Magazine, Vol.24, No.4, pp.10-25, Apr 2019.

[4] M. Lee, H. Song, J. Park, B. Jeon, J. Kang, J.-G. Kim, Y. Lee, J.-W. Kang, and D. Sim, “Overview of Versatile Video Coding (H. 266/VVC) and Its Coding Performance Analysis,” IEIE Transactions on Smart Processing & Computing, Vol.12, No.2, pp.122-154, Apr 2023. doi: https://doi.org/10.5573/IEIESPC.2023.12.2.122

[5] L. Yang, Z. Qin, S. Anwar, P. Ji, D. Kim, S. Caldwell, and T. Gedeon, "Invertible denoising network: A light solution for real noise removal," the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13365-13374, 2021. doi: https://doi.org/10.1109/cvpr46437.2021.01316

[6] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2472 – 2481, 2018. doi: https://doi.org/10.48550/arXiv.1802.08797

[7] R. Lee, SI Venieris, and ND Lane, “Deep neural network-based enhancement for image and video streaming systems: A survey and future direction,” ACM Computing Surveys, Vol.54, No. 169, pp 1-30, Oct 2021. doi: https://doi.org/10.1145/3469094

[8] FJ. Tsai, YT. Peng, YY. Lin, CC. Tsai, CW. Lin, “Stripformer: Strip Transformer for Fast Image Deblurring,” the European Conference on Computer Vision 2022, Vol 13679. 2022. doi: https://doi.org/10.1007/978-3-031-19800-7_9

[9] Y. Rao and L. Chen, “A Survey of Video Enhancement Techniques,” Journal of Information Hiding and Multimedia Signal Processing, Vol 3, No.1, Jan 2012.

[10] Z. Wang, W. Cun, J. Bao, W. Zhou, J Liu, and H. Li, “Uformer: A general u-shaped transformer for image restoration,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.17683-17693, 2022. doi: https://doi.org/10.1109/cvpr52688.2022.01716

[11] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, “Multi-level Wavelet-CNN for Image Restoration,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.773-782, 2018. doi: https://doi.org/10.1109/cvprw.2018.00121

[12] D. Qing, L. Shen, L. Yu, H. Yang, and M Xu, "Patch-wise spatial-temporal quality enhancement for HEVC compressed video," the IEEE Transactions on Image Processing, Vol.30, pp.6459-6472,

2021.
doi: https://doi.org/10.1109/TIP.2021.3092949

[13] D. Meng and F. D.Torre, “Robust matrix factorization with unknown noise,” the IEEE International Conference on Computer Vision, pp.1337 – 1344, 2013.
doi: https://doi.org/10.1109/iccv.2013.169

[14] F. Zhu, G. Chen, J. Hao, and P. Heng, “Blind image denoising via dependent dirichlet process tree,” the IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.39, No.8, pp.1518 – 1531, 2017.
doi: https://doi.org/10.1109/TPAMI.2016.2604816

[15] X. Cao, Y. Chen, Q. Zhao, D. Meng, Y. Wang, D. Wang, and Z. Xu, “Low-rank matrix factorization under general mixture noise distributions,” the IEEE/CVF International Conference on Computer Vision, pp.1493 – 1501, 2015.
doi: https://doi.org/10.1109/iccv.2015.175

[16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” the IEEE Transactions on Image Processing, Vol.16, No.8, pp.2080 – 2095, 2007.
doi: https://doi.org/10.1109/TIP.2007.901238

[17] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images,” the IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.34, No.11, pp.2233 – 2246, 2012.
doi: https://doi.org/10.1109/TPAMI.2011.282

[18] J. Xu and S. Osher, “Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising,” the IEEE Transactions on Image Processing, Vol.16, No.2, pp.534 – 544, 2007.
doi: https://doi.org/10.1109/TIP.2006.888335

[19] L. Liu, Y. Wang and W. Chi, "Image Recognition Technology Based on Machine Learning," IEEE Access, 2020.
doi: https://doi.org/10.1109/ACCESS.2020.3021590

[20] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, “Multi-level Wavelet-CNN for Image Restoration,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.773-782, 2018.
doi: https://doi.org/10.1109/cvprw.2018.00121

[21] X. Fu, Z. J. Zha, F. Wu, X. Ding, and J. Paisley, “JPEG Artifacts Reduction via Deep Convolutional Sparse Coding,” the IEEE/CVF International Conference on Computer Vision, pp.2501-2510, 2019.
doi: https://doi.org/10.1109/iccv.2019.00259

[22] S. Brehm, S. Scherer, and R. Lienhart, “High-resolution Dual-stage Multi-Level Feature Aggregation for Single Image and Video Deblurring,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.458 – 459, 2020.
doi: https://doi.org/10.1109/cvprw50498.2020.00237

[23] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, “Variational denoising network: Toward blind noise modeling and removal,” Neural Information Processing Systems, Vol.32, No.12, pp.1690 – 170, 2019.

[24] Z. Yue, Q. Zhao, L. Zhang, and D. Meng, “Dual adversarial network: Toward real-world noise removal and noise generation,” European Conference on Computer Vision 2020, pp.41 – 58, 2020.
doi: https://doi.org/10.1007/978-3-030-58607-2_3

[25] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. Yang, and L. Shao, “Learning enriched features for real image restoration and enhancement,” European Conference on Computer Vision 2000, UK, August, pp.492-511, 2020.
doi: https://doi.org/10.1007/978-3-030-58595-2_30

[26] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image superresolution via sparse representation,” IEEE Transactions on Image Processing, Vol.19, No.11, pp.2861-2873, 2010.
doi: https://doi.org/10.1109/TIP.2010.2050625

[27] K. Zhang, X. Gao, D. Tao, and X. Li, “Single image super-resolution with non-local means and steering kernel regression,” IEEE Transactions on Image Processing, Vol.21, No.11, pp.4544 – 4556, 2012.
doi: https://doi.org/10.1109/TIP.2012.2208977

[28] S. Schulter, C. Leistner, and H. Bischof, “Fast and accurate image upscaling with super-resolution forests,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.3791 – 3799, 2012.
doi: https://doi.org/10.1109/cvpr.2015.7299003

[29] X. Li, B. Du, C. Xu, Y. Zhang, L. Zhang, and D. Tao, “Robust learning with imperfect privileged information,” Artificial Intelligence, Vol.282, pp.103246, May 2020.
doi: https://doi.org/10.1016/j.artint.2020.103246

[30] D. Ren, W. Zuo, D. Zhang, L. Zhang, and M. Yang, “Simultaneous fidelity and regularization learning for image restoration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.43, No.1, pp.284-299, Jan 2019.
doi: https://doi.org/10.1109/TPAMI.2019.2926357

[31] C. Tian, Y. Xu, and W. Zuo, “Image denoising using deep cnn with batch renormalization,” Neural Networks, Vol.121, pp.461 – 473, 2020.
doi: https://doi.org/10.1016/j.neunet.2019.08.022

[32] C. Tian, Y. Xu, W. Zuo, B. Du, C. Lin, and D. Zhang, D., “Designing and training of a dual cnn for image denoising,” arXiv preprint, 2020, arXiv:2007.03951.
doi: https://doi.org/10.1016/j.knosys.2021.106949

[33] D. Yuan, N. Fan, and Z. He, “Learning target-focusing convolutional regression model for visual object tracking,” Knowledge-Based Systems, Vol.194, pp.105526, 2020.
doi: https://doi.org/10.1016/j.knosys.2020.105526

[34] X. Mao, C. Shen, and Y. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” Advances In neural Information Processing Systems. Vol.26, pp.2802 – 2810, 2016.

[35] N. Ahn, B. Kang, K. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” European Conference on Computer Vision 2018, pp. 252 – 268, 2018.
doi: https://doi.org/10.1007/978-3-030-01249-6_16

[36] N. Ahn, B. Kang, K. Sohn, “Photo-realistic image super-resolution with fast and lightweight cascading residual network,” arXiv preprint, 2019, arXiv:1903.02240.

[37] W. Bao, W. Lai, C. Ma, X. Zhang, Z. Gao, and M. Yang, “Depth-aware video frame interpolation,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
doi: https://doi.org/10.1109/cvpr.2019.00382

[38] Y. Jo, S. Oh, J. Kang, and S. Kim, “Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.3224 – 3232, 2018.
doi: https://doi.org/10.1109/cvpr.2018.00340

[39] Y. Huang, W. Wang, and L. Wang, “Bidirectional recurrent convolutional networks for multi-frame super-resolution,” Neural Information Processing Systems, pp.235 – 243, 2015,

[40] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi, “Real-time video super-resolution with spatio-temporal networks and motion compensation,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.4778 – 4787, 2017.
doi: https://doi.org/10.1109/cvpr.2017.304

[41] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, and T. Huang. “Robust video super-resolution with learned temporal dynamics,” the IEEE/CVF International Conference on Computer Vision, pp.2507 – 2515, 2017.
doi: https://doi.org/10.1109/iccv.2017.274

[42] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia. “Detail revealing deep video super-resolution,” I the IEEE/CVF International Conference on Computer Vision, pp.4472 – 4480, 2017.
doi: https://doi.org/10.1109/iccv.2017.479

[43] M. S. Sajjadi, R. Vemulapalli, and M. Brown, “Frame-recurrent video super-resolution,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.6626 – 6634, 2018.
doi: https://doi.org/10.1109/cvpr.2018.00693

[44] X. Cheng and X. Chen. “Video frame interpolation via deformable separable convolution”. Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, Vol.34, No.7, 2020.
doi: https://doi.org/10.1609/aaai.v34i07.6634

[45] S. Niklaus, L. Mai, and O. Wang, “Revisiting adaptive convolutions for video frame interpolation,” the IEEE/CVF Winter Conference on Applications of Computer Vision, pp.1099-1109, 2021.
doi: https://doi.org/10.1109/wacv48630.2021.00114

[46] M.Choi, H. Kim, B. Han, N. Xu, and K. Lee. “Channel attention is all you need for video frame interpolation,” Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, Vol.34, No.7, pp.10663-10671, 2020.
doi: https://doi.org/10.1609/aaai.v34i07.6693

[47] T. Kalluri, D. Pathak, M. Chandraker, and D. Tran, “Flavr: Flow-agnostic video representations for fast frame interpolation,” arXiv preprint, 2020, arXiv:2012.08512.
doi: https://doi.org/10.1109/wacv56688.2023.00211

[48] H. Jiang, D. Sun, V. Jampani, M. Yang, E. Learned-Miller, and J. Kautz, “Super slomo: High quality estimation of multiple intermediate frames for video interpolation,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9000-9008, 2018.
doi: https://doi.org/10.1109/cvpr.2018.00938

[49] S.Baker, D.Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, “A database and evaluation methodology for optical flow,” International journal of computer vision, Vol.92, No.1, pp.1 – 31, 2011.
doi: https://doi.org/10.1007/s11263-010-0390-2

[50] S. Niklaus, and F. Liu, “Context-aware synthesis for video frame interpolation,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.1701-1710, 2018.
doi: https://doi.org/10.1109/cvpr.2018.00183

[51] S. Niklaus, and F. Liu. “Softmax splatting for video frame interpolation,” the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
doi: https://doi.org/10.1109/cvpr42600.2020.00548

[52] Y. Qiu, Y. Yang, Z. Lin, P. Chen, Y. Luo, and W. Huang, “Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV,” China Communications, Vol.17, No.3, pp.46-57, 2020.
doi: https://doi.org/10.23919/jcc.2020.03.005

[53] H. Lu, A. Yamawaki, S. Serikawa, “Curvelet approach for deep-sea sonar image denoising, contrast enhancement and fusion,” Journal of International Council on Electrical Engineering, Vol.3, No.3, pp.250-256, 2013.
doi: https://doi.org/10.23919/JCC.2020.03.005

[54] P. Liu, M. Wang, L. Wang, and W. Han, “Remote-sensing image denoising with multi-sourced information,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol.12, No.2, pp.660-674, 2022.
doi: https://doi.org/10.1109/JSTARS.2019.2891566

[55] L. Han, Y. Zhao, H. Lv, Y. Zhang, H. Liu, and G. Bi, “Remote sensing image denoising based on deep and shallow feature fusion and attention mechanism,” Remote Sensing, Vol.14, No.5, pp.1243, 2022.
doi: https://doi.org/10.3390/rs14051243

[56] L. Gondara, “Medical image denoising using convolutional denoising autoencoders,” IEEE international conference on data mining workshops, pp. 241-246, 2016.
doi: https://doi.org/10.1109/icdmw.2016.0041

[57] D. Bhonsle, V. Chandra, and G. R. Sinha, “Medical image denoising using bilateral filter,” International Journal of Image, Graphics and Signal Processing, Vol.4, No.6, pp.36, 2012.
doi: https://doi.org/10.5815/ijigsp.2012.06.06

[58] M. A. M. Razif, M. Mokji, and M. M. A. Zabidi, "Low complexity maritime surveillance video using background subtraction on H.264," 2015 International Symposium on Technology Management and Emerging Technologies, Langkawi, Malaysia, pp. 364-368, 2015.
doi: https://doi.org/10.1109/istmet.2015.7359060

[59] C. Kwan, J. Larkin, B. Budavari, B. Chou, E. Shang, “Tran, T.D. A Comparison of Compression Codecs for Maritime and Sonar Images in Bandwidth Constrained Applications,” Computers 2019, Vol.8, No.32, 2019.
doi: https://doi.org/10.3390/computers8020032

[60] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature Vol.521, pp.436 – 444, 2015.
doi: https://doi.org/10.1038/nature14539

[61] C. Li, C. Guo, L. Han, J. Jiang, M. M. Cheng, J. Gu, and C. C. Loy "Low-light image and video enhancement using deep learning: A survey," IEEE transactions on pattern analysis and machine intelligence, Vol.44, No.12, pp.9396-416, 2021.
doi: https://doi.org/10.1109/TPAMI.2021.3126387

[62] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-Resolution: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.43, No. 10, pp. 3365-3387, 2021.
doi: https://doi.org/10.1109/TPAMI.2020.2982166

[63] D. Lee, C. Lee, and T. Kim, "Wide Receptive Field and Channel Attention Network for JPEG Compressed Image Deblurring," the IEEE/CVF Conference on Computer Vision and Pattern Recognition , Nashville, TN, USA, pp. 304-313, July 2021.
doi: https://doi.org/10.1109/cvprw53098.2021.00040

[64] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, and L. Shao, "Multi-stage progressive image restoration," the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp.14821-14831, 2021.
doi: https://doi.org/10.1109/cvpr46437.2021.01458

[65] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, R, "Swinir: Image restoration using swin transformer," the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.1833-1844, 2021.
doi: https://doi.org/10.1109/iccvw54120.2021.00210

[66] C. Tian, R. Zhuge, Z. Wu, Y. Xu, W. Zuo, C. Chen, and C. W. Lin, "Lightweight image super-resolution with enhanced CNN," Knowledge-Based Systems, Vol.205, pp.106235, 2020.
doi: https://doi.org/10.1016/j.knosys.2020.106235

[67] A. Chadha, J. Britto, and M. M. Roja, M. M. "iSeeBetter: Spatio-temporal video super-resolution using recurrent generative back-projection networks," Computational Visual Media, Vol.6, pp.307-317, 2020.
doi: https://doi.org/10.1007/s41095-020-0175-7

[68] P. Behjati, P. Rodriguez, A. Mehri, I. Hupont, C. F. Tena, and J. Gonzalez, "Overnet: Lightweight multi-scale super-resolution with overscaling network,". the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2694-2703, 2021.
doi: https://doi.org/10.1109/wacv48630.2021.00274

[69] P. Behjati, P. Rodriguez, C. F. Tena, A. Mehri, F. X. Roca, S. Ozawa, and J. Gonzàlez, "Frequency-Based Enhancement Network for Efficient Super-Resolution," IEEE Access, Vol.10, pp.57383-57397, 2022.
doi: https://doi.org/10.1109/ACCESS.2022.3176441

[70] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou, "Real-time intermediate flow estimation for video frame interpolation." European Conference on Computer Vision, Tel Aviv, Israel, pp. 624-642, 2022.
doi: https://doi.org/10.1007/978-3-031-19781-9_36

[71] P. Hu, S. Niklaus, S. Sclaroff, and K. Saenko, "Many-to-many splatting for efficient video frame interpolation," the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp.3553-3562. 2022.
doi: https://doi.org/10.1109/CVPR52688.2022.00354

[72] H. Sim, J. Oh, and M. Kim, "Xvfi: extreme video frame interpolation," the IEEE/CVF International Conference on Computer Vision, pp.14489-14498. 2021.
doi: https://doi.org/10.1109/iccv48922.2021.01422

[73] L. Kong, B. Jiang, D. Luo, W. Chu, X. Huang, Ying Tai, C. Wang, and J. Yang. "Ifrnet: Intermediate feature refine network for efficient frame interpolation." the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1969-1978. 2022.
doi: https://doi.org/10.1109/cvpr52688.2022.00201

[74] J. Park, C. Lee, and C. Kim, "Asymmetric bilateral motion estimation for video frame interpolation," the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.14539-14548. 2021.
doi: https://doi.org/10.1109/iccv48922.2021.01427

[75] FFmpeg, http://ffmpeg.org/

저 자 소 개

이 영 복

- 2022년 8월 : 광운대학교 컴퓨터공학과 학사
- 2022년 9월 ~ 현재 : 광운대학교 컴퓨터공학과 석사
- ORCID : https://orcid.org/0009-0002-2500-7228
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

저 자 소 개

이 은 성

- 2022년 2월 : 광운대학교 전자공학과 학사
- 2022년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 석사
- ORCID : https://orcid.org/0000-0001-6996-0809
- 주관심분야 : 영상신호처리, 컴퓨터비전, 영상압축

이 민 훈

- 2019년 2월 : 광운대학교 수학과, 전자공학과 (복수전공) 학사
- 2021년 2월 : 광운대학교 전자공학과 석사
- 2021년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0000-0001-8165-5380
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

변 주 형

- 2019년 2월 : 광운대학교 컴퓨터공학과 학사
- 2021년 2월 : 광운대학교 컴퓨터공학과 석사
- 2021년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0000-0002-6165-9189
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

안 현 모

- 2018년 : 수원대학교 정보통신공학과 공학사 졸업.
- 2020년 ~ 현재 : 주식회사 큐램 선임연구원.
- 2021년 ~ 현재 : 아주대학교 인공지능학과 석사 과정
- ORCID : https://orcid.org/0009-0004-0131-8309
- 주관심분야 : 컴퓨터비전, 이미지프로세싱, 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