

논문 2024-61-2-11

의사 참조 이미지를 활용한 무참조 이미지 화질 평가

(Blind Image Quality Evaluation using Pseudo Reference Images)

용 윤 정*, 오 희 석**

(Yunjeong Yong and Heeseok Oh[©])

요 약

무참조 이미지 화질 평가(no-reference image quality assessment; NR-IQA)는 참조 영상에 대한 정보가 주어지지 않은 상태에서 인간 시각 체계를 반영하여 이미지 화질 저하 수준을 객관적으로 정량화 하는 것을 목표로 한다. 기존 NR-IQA 기술은 특정 왜곡 특성에 대해 높은 민감성을 갖고 있으며, 의미론적 화질 인지 및 화질 저하 수준을 판단하기에는 한계를 지니고 있다. 본 논문에서는 기존 NR-IQA 접근 방식의 한계를 극복하고 예측 정확도를 향상시키기 위해 지역적 왜곡 패턴 추출과 고수준 의미 정보 추출에 보다 능숙한 완전 참조(full-reference IQA; FR-IQA) 방식의 프레임워크를 따르는 NR-IQA 모델인 multi-scale pseudo image quality assessor(MPIQ)를 제안한다. 제안하는 MPIQ는 하이브리드 구조에 기반해 합성곱 네트워크를 통한 지역적 왜곡 패턴 추출과 트랜스포머 기반의 전역적 화질 수준 이해를 도모하며, 의사 참조 이미지 reconstructor와 quasi FR-IQA regressor라는 두 가지 모듈로 구성된다. 의사 참조 이미지 reconstructor는 FR-IQA 방식과 유사하게 화질 추론 모델이 동작할 수 있도록 인코더-디코더 구조를 차용하여 의사 참조 이미지를 재구성하고 이미지 왜곡 정보를 학습한다. 이 과정에서 크기 불변 왜곡 패턴의 추출을 위해 다중 스케일 구조를 반영하였다. Quasi FR-IQA regressor는 왜곡 이미지에서 추출된 특징과 의사 참조 이미지 특징의 차이를 통해 전역적 왜곡 수준을 도출함으로써 이미지 화질 수준을 최종적으로 예측한다. MPIQ는 주관적 평가를 통해 획득된 mean opinion score에 end-to-end로 학습되었으며, 실험 결과 기존 NR-IQA 기술 대비 20%의 성능 향상을 달성하였다.

Abstract

No-reference image quality assessment (NR-IQA) aims to objectively quantify the level of image quality degradation by reflecting the human visual system in the absence of information about the pristine image. Existing NR-IQA techniques have high sensitivity to specific distortion types, but have limitations in determining semantic quality information or global image quality degradation. In this paper, to resolve the limitations of existing NR-IQA approaches and improve predictive power, we propose a multi-scale pseudo image quality assessor (MPIQ). MPIQ is an NR-IQA model that follows the framework of full-reference IQA (FR-IQA), which is more proficient in extracting local distortion patterns and aggregating higher-level perceptual quality information. The proposed MPIQ employs a hybrid scheme that seeks to understand local distortion patterns through the convolutional neural networks and global level of image quality based on transformers, and consists of two modules: a pseudo-reference image reconstructor and a quasi FR-IQA regressor. Similar to the FR-IQA approach, the pseudo-reference image reconstructor utilizes an encoder-decoder structure to reconstruct the pseudo-reference image and learn image degradation information compared to a distorted one. Here, a multi-scale structure is reflected to extract scale-invariant distortion patterns. Quasi FR-IQA regressor predicts the image quality score by deriving the global distortion level through the difference between the features extracted from the distorted and the pseudo-reference images. MPIQ was supervised onto the mean opinion score obtained through subjective evaluation in an end-to-end manner, and experimental results showed a 20% performance improvement compared to the existing NR-IQA.

Keywords : Image quality assessment, Perceptual quality, Pseudo-reference image, Scale-invariant distortion, Quality score regressor

*비회원, **정회원, 한성대학교 AI응용학과(Department of Applied AI, Hansung University)

© Corresponding Author(E-mail : ohhs@hansung.ac.kr)

※ 본 연구는 한성대학교 교내학술연구비 지원과제임.

Received ; October 4, 2023

Revised ; November 12, 2023

Accepted ; November 29, 2023

I. 서 론

이미지 화질 평가(image quality assessment; IQA)는 인간 시각 체계 기반 화질 인지 추론 과정과 유사하도록 주어진 이미지에 대한 객관적 화질 점수를 예측하는 영상처리 및 비전 분야의 고난이도 task이다. IQA의 핵심 기술은 주어진 이미지의 지역적 화질 저하 요소를 주관적 의견과 상관성이 높도록 전역적인 품질로 종합해 정량화하는 것이다. 일반적으로 이미지는 획득, 압축, 처리, 저장 및 전송^[1]과 같은 과정에 의해 화질 저하가 필연적으로 발생한다. 이로 인해 본래 원본 이미지가 내포하는 시각적 정보가 손실되며 이는 해당 이미지를 기반으로 한 응용 분야에서 고수준의 인지 및 분석 시나리오에 부정적인 영향을 미칠 수 있으므로, 객관적으로 정량화된 화질 예측 점수를 통해 이미지 품질 관리에서의 기준을 유지함과 동시에 저하 요소를 파악하고 이를 최소화함으로써 중요한 시각적 정보를 보존하고 신뢰성을 보장할 수 있다. 더불어 최근 방대한 디지털 영상 정보가 범람하는 환경에서는 해당 과정을 인력을 동원해 일일이 평가하는 것이 아닌, 객관적 지표를 통해 자동적인 이미징 기술 개선 및 사용자 경험 향상에 효율적으로 기여할 수 있다.

IQA를 위한 주관적 방법은 다수의 피험자를 활용한 크라우드 소싱을 바탕으로 한 이미지 화질의 평균 의견 점수(mean opinion score; MOS)에 의존한다. 이는 방대한 이미지 평가에 있어 물적/시간적인 어려움을 야기하므로, 정상 이미지와 왜곡된 이미지 간의 시각적 속성에 대한 차이를 정량적으로 측정하기 위해 객관적인 메트릭 및 평가 알고리즘이 요구된다. 객관적 IQA 방법은 참조 이미지의 사용 가능 여부에 따라 full-reference IQA(FR-IQA)^[2], reduced-reference IQA(RR-IQA)^[3], no-reference IQA(NR-IQA)^[4]의 세 가지 접근으로 크게 구분할 수 있다. FR-IQA는 참조 이미지(pristine 이미지)와 왜곡 이미지 간의 차이를 측정해 수치화하는 방식이며, RR-IQA는 참조 이미지의 일부 정보(메타데이터, 왜곡 타입 등)를 활용하는 방법이다. 반면 NR-IQA는 참조 이미지에 대한 정보가 전혀 없는 상태에서 왜곡 영상에 대한 화질 저하 수준을 예측해 수치화하는 방식이다. NR-IQA는 참조 이미지를 요구하지 않으므로 실용성/효율성 측면에 장점을 가지며 실제 환경 적용에 범용성이 높다. 하지만 일반적으로 NR-IQA는 FR-IQA와 동일한 수준의 정확도 달성의 어려움 및 인간 시각 인지 측면에서의 화질 수준

과 상관관계가 부족한 경향이 있다.

본 연구에서는 그림 1에서 나타내는 바와 같이 현실에서 참조 이미지와 왜곡 이미지가 쌍으로 존재하는 경우가 희소함에 따라 NR-IQA를 통한 정량적 화질점수를 추론하되, 참조 이미지 없이도 왜곡 이미지로부터 의사 참조(pseudo reference) 이미지를 재구성함으로써 FR-IQA와 유사하게 화질 저하로 비롯되는 시각 인지 요소를 캡처할 수 있도록 NR-/FR-IQA 양쪽의 장점을 모두 취하고자 모델을 설계하였다.

기존 NR-IQA 연구에서는 지역적/전역적 영역 모두를 살펴보고 영상 화질을 판단하기 위해 인간 시각 체계를 모델링하는 연구가 다수 수행되었다^[5, 6]. 이러한 인지 모델링 기반의 IQA는 이미지의 특정 왜곡 특성에 대해 높은 민감성을 지니나, 복합적 왜곡 패턴과 의미론적 이미지 화질 저하를 판단하기에는 한계가 있다. 따라서 기존에는 주관적 점수인 MOS가 제공된 이미지 데이터로부터 화질 점수로의 매핑을 학습하기 위한 convolutional neural network (CNN)을 기반으로 한 방식이 주를 이루었다. CNN 기반 IQA는 왜곡 이미지에서의 계층적 특징과 복잡한 지역적 왜곡 패턴을 포착하며 높은 성능을 보였다^[7-10]. 하지만 CNN기반의 IQA는 저수준 및 중간 수준의 이미지 왜곡 특징을 추출에는 효과적이거나 고수준의 의미론적 특징과 전역적 영상 context를 이해에는 한계가 존재한다. 최근에는 CNN의 지역적 화질 저하 추출 위주의 단점을 보완하기 위해 transformer 기반의 IQA 모델이 도입되었으나, 대부분 이미지 내 구조적 특성에 집중하여 활용되었으며 이미지 화질을 이해하기 위한 시각적 특징 도출에 초점을 맞추지는 못하였다^[11, 12].

NR-IQA 특성상 지역-전역적 특성을 고루 고려해야 하는 필요가 있으며, 본 논문에서는 이의 달성을 위해 지역적 정보 취합을 통한 전역적 화질 이해라는 인간 시각 인지 체계를 반영하고자 전체적인 아키텍처를 CNN과 transformer의 hybrid 구조로 구현하였으며, CNN을 통해 도출한 지역적 화질 저해 요소를 transformer를 통한 장거리 종속성 이해 바탕으로 종합함으로써 전역적인 화질 수준으로 변환한다^[13].

또한 의사 참조 이미지 reconstructor를 도입함으로써 IQA를 단순한 MOS로의 회귀 문제로 해결하는 것이 아니라, 기존 은닉 공간에서의 화질 저하 수준 차이를 학습하는 FR-IQA의 프레임워크를 그대로 적용 가능한 방법을 고안하였다(그림 1). Reconstructor 뒷단의 quasi FR-IQA regressor는 임의의 FR-IQA 모델을

차용해도 NR-IQA가 가능하도록 구현하였다. 즉, MOS에 상관관계가 높은 예측 점수 추론을 위해 지역적/전역적 화질 중요도가 반영되어 재구성된 의사 참조 이미지로부터 왜곡 이미지와의 차이를 회귀하는 방식을 취한다. 여기서 ViT^[13] 형태의 regressor를 이용하였으므로 참조 이미지와 왜곡 이미지 간 차이를 attention을 통해 화질 점수로 예측하며 해당 local-interaction을 가시화함으로써 화질 정량점수 추론에 대한 분석이 가능하다는 장점이 있다.

추가적으로 의사 참조 이미지와 왜곡 이미지간의 공간적 왜곡 차이를 반영 시, 단순히 특징 맵의 차이 값 자체만을 계산하는 것이 아닌, 채널 별 공간 구조 정보를 반영해 다중 스케일에 대한 왜곡을 고려할 수 있도록 화질 저하의 크기 불변한 모델을 구축하였다. 상기 기술을 통해 제안하는 방법이 기존 NR-IQA 기술 대비 실제 의견과의 상관도에서 20%의 성능 향상을 달성하였다.

- 화 인지 특성, 공간 및 채널의 구조적인 정보까지 크기 불변하게 고려할 수 있도록 구현하였다.
- 제안하는 기술을 통해 기존 NR-IQA기술 대비 20%의 성능 향상을 달성하였다.

II. 관련 연구

기존 NR-IQA를 위한 기술적 접근은 크게 NSS(natural scene statistics) 기반의 통계량 추정과 실제 피험자를 대상으로 한 주관적 화질 점수인 MOS로의 학습인 data-driven 기반 방법으로 나눌 수 있으며, CNN과 transformer와 같은 신경망을 regressor로써 차용한 모델들이 소개되어왔다.

1. NSS기반 NR-IQA

MOS학습을 활용하지 않는 NR-IQA 접근은 대부분 자연 영상에 대한 일반적인 장면의 통계량인 NSS를 적극적으로 활용하였다. NSS를 바탕으로 하는 NR-IQA 방법은 특정 도메인에서의 계수를 정해진 확률 분포로 모델링하여 왜곡 영상과 해당 확률 분포와의 Euclidean 거리 및 mahalanobis 거리 계산을 통해 화질의 수준을 정량화한다. 자연 이미지를 이산 코사인 변환 및 웨이블릿 변환을 통해 특정 도메인으로 옮기게 될 경우, 해당 계수들은 asymmetric generalized Gaussian 분포와 같은 특정 확률 분포를 따르는 규칙적인 통계 특성을 가지고 있다. 하지만 압축, 노이즈, 블러 등에 의해 발생한 다양한 왜곡패턴은 이러한 자연 영상이 내포해야 할 통계적 특성에서 멀어지게 한다. Mittal 등이 제안한 NIQE^[14]는 자연 이미지에서 블록 단위 지역적 특징 집합을 추출한 후, 특징 벡터를 활용해 다변량 가우시안 분포의 모수를 추정하고 fitting한다. 이미지의 화질은 해당 NSS 분포와의 거리를 계산함으로써 예측되며, 거리가 멀수록 NSS와 차이가 큰 이유로 왜곡이 크다고 판단한다. Mittal 등이 제안한 BRISQUE^[15]는 평균 차분과 대비 정규화를 통해 구성된 지역 정규화 도메인에서 일반화된 가우시안 분포로 NSS를 모델링한 후, 각 영상별 모수를 특징으로 추출하여 MOS와의 support vector regression을 학습해 이미지의 품질을 예측한다. DIIVINE^[16]은 NSS의 웨이블릿 계수 모델에서 파생된 요약 통계를 활용하고 품질 평가를 위해 두 단계 프레임워크를 사용한다. 첫 번째 단계에서 DIIVINE은 이미지가 여러 왜곡 범주의 영향을 받을 확률을 식별하여 왜곡 식별 추정치를 얻는다. 그런 다음

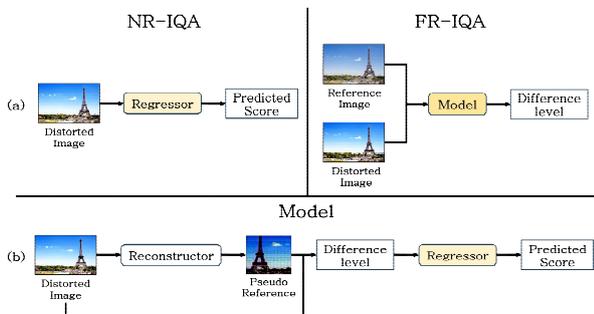


그림 1. (a) 기존 NR-IQA와 FR-IQA 프레임워크; (b) 본 논문에서 제안하는 FR-IQA와 같이 왜곡 이미지와 의사 참조 이미지의 차이를 반영하는 NR-IQA 프레임워크

Fig. 1. (a) Previous NR-IQA and FR-IQA frameworks; (b) Proposed NR-IQA framework which employed the difference between pseudo reference image and the distorted one similar to FR-IQA.

본 연구의 주요 기여는 아래와 같다.

- 단순한 MOS 회귀가 아닌 의사 참조 이미지를 재구성하여 FR-IQA의 프레임워크를 따라 NR-IQA가 가능하도록 구현하였다.
- Transformer 기반 의사 참조 이미지와 왜곡 이미지 간의 지역적 왜곡 추출과 이들 간 상호작용을 수행하는 NR-IQA로써 attention map 분석을 통한 화질 저해 패턴 분석이 가능하다.
- 의사 참조 이미지와 왜곡 이미지 간 차이를 도출함에 있어 다중 스케일 구조를 적용함으로써, 화질 열

두 번째 단계에서 특징 벡터를 각 왜곡 범주의 품질 점수에 매핑하여 왜곡별 품질 점수를 얻는다. 최종적으로 이 두 가지 정보를 결합하여 이미지의 품질 점수를 예측한다.

2. CNN 기반 NR-IQA

NSS 기반 방법론은 도메인에 특화된 특징의 설계가 필요하며, 매 번 직접 특징을 추출해야 하므로 연산량이 많아지기에 시간 복잡도가 증가한다. 또한, 장면의 통계적 특성은 영상의 전체적인 이미지 화질을 나타내므로 실제 왜곡된 이미지에서 일반적으로 나타나는 지역적 왜곡에 의한 화질을 측정하기 어렵다. 더불어 유형별로 정해진 왜곡과는 달리 실제 시나리오에서는 더욱 다양하고 복잡한 형태의 왜곡이 존재한다. 결정적으로, feature-engineering 관점에서의 특징 추출 접근은 실제 왜곡을 모델링하는 능력이 떨어진다. 하지만 학습 기반 접근 방식은 왜곡 이미지에서 품질 수준에 차별적인 특징을 추출 가능하여 실제 왜곡 이미지 방면에서 전통적인 방법과 달리 좋은 성능을 보인다.

Le Kang^[17]은 각 이미지를 로컬 패치 단위로 CNN에 통과시켜 특징맵을 산출하고 해당 특징맵의 최소값과 최대값을 벡터화하여 이미지 전체에 대한 품질을 예측한다. Talebiet^[18] 등은 MOS나 DMOS(differential MOS)와 같은 의견의 평균값 대신 주관적 품질 점수의 지각적 분포를 이용하여 모델을 훈련시켰다. Varga^[19] 등은 ResNet-101을 백본으로 사용하여 특징을 추출하고, 적응형 공간 피라미드 풀링 계층을 추가하여 백본 네트워크가 서로 다른 이미지의 크기를 고정된 크기로 변환하여 추출할 수 있도록 한다. Zhu^[11] 등은 사전 지식을 공유함으로써 적은 수의 이미지만을 가지고 기존에 접해보지 않은 왜곡에 대한 NR-IQA 모델의 학습 성능이 향상될 수 있도록 모델을 구성하였다. Kede^[20] 등은 왜곡의 유형과 이미지 품질을 동시에 예측해내는 더 깊은 네트워크 기반의 IQA 모델을 제안하였다.

3. Transformer 기반 NR-IQA

CNN은 합성곱 필터를 사용해 지역적인 특성을 파악하며, 학습 후에는 고정된 가중치를 사용한다. 따라서 CNN이 최종 예측에 대해 강한 지역성 편향과 더불어 가중치의 공유를 통한 공간적 불변 편향을 보이기 때문에 더 복잡한 고수준 특징들 간 관계를 학습하기 어렵다는 단점이 있다. 반면 transformer는 attention을 통해 지역적 정보 간 상호 의존성을 파악하고 전역적인

정보를 도출함으로써 지역-전역 관계 파악의 자유도를 높이게 된다. 지역적 왜곡과 인간 시각 체계의 전역적 특성을 반영하여 화질점수를 추론해야하는 IQA 특성상, 지역적인 정보와 전역적인 정보가 동시에 요구됨에 따라 최근 몇 년간 transformer를 적용한 NR-IQA 접근이 크게 증가하였다. You와 Korhonen^[11]은 transformer 인코더를 CNN의 특징 맵과 결합하는 구조를 활용하여 이미지의 화질을 효과적으로 평가할 수 있음을 보였다. 입력 이미지의 해상도에 따라 맥스 풀링을 적용한 특징에 학습 가능한 위치 임베딩을 사용하였으며, transformer 인코더로 추출된 최종 특징으로부터 regression 헤드를 통해 이미지 화질을 예측하였다. Golestaneh^[21] 등은 transformer를 사용하여 화질 최대값과 차순위 품질 점수 사이의 거리에 대한 triplet loss를 활용한 대비학습을 통해 상대적 화질 정보 활용을 바탕으로 오검출 민감도를 감소시켰다. Ke^[22] 등은 다양한 입력 이미지 해상도를 처리하기 위해 헤시 기반 2D 공간 임베딩 및 스케일 임베딩을 통해 다중 스케일을 정보를 표현한 후, 트랜스포머 인코더를 통해 이미지 품질 점수를 예측하였다.

하지만 대부분의 기존 접근은 레이블로 주어지는 MOS로의 모델 학습에 의존적이며, CNN에 의한 지역적 왜곡 패턴 학습과 transformer를 이용한 왜곡 패턴 간 전역적 품질 예측으로의 연결성이 부족하거나 각 단계별로만 동작할 수 있다는 한계를 갖는다.

반면 본 논문에서는 의사 참조 이미지를 활용하여 왜곡 이미지와의 공간적 왜곡 차이를 명시적으로 구분할 수 있도록 모델의 설계를 도모하였다. 특히 다중 스케일을 적용한 의사 참조 및 왜곡 이미지 간의 차이를 transformer 입력으로 활용해 화질 열화 인지 특성과 공간 및 채널 구조적인 정보를 얻을 수 있어 효과적인 화질 예측의 성능을 달성하였다.

III. MPIQ (Multi-Scaled Pseudo Image Quality Assessor)

일반적으로 FR-IQA가 NR-IQA 대비 화질 열화 측정 성능에 대한 강건성과 신뢰도가 높으나 이미지 압축, 비디오 스트리밍 등과 같이 실제 시나리오에서는 화질 평가를 원하는 수신측에서의 참조 이미지가 존재하지 않는 경우가 대다수이며 이에 FR-IQA의 활용이 불가하므로, 제안 방법은 FR-IQA 성능에 준함과 동시에 NR-IQA의 활용성을 모두 갖는 모델인

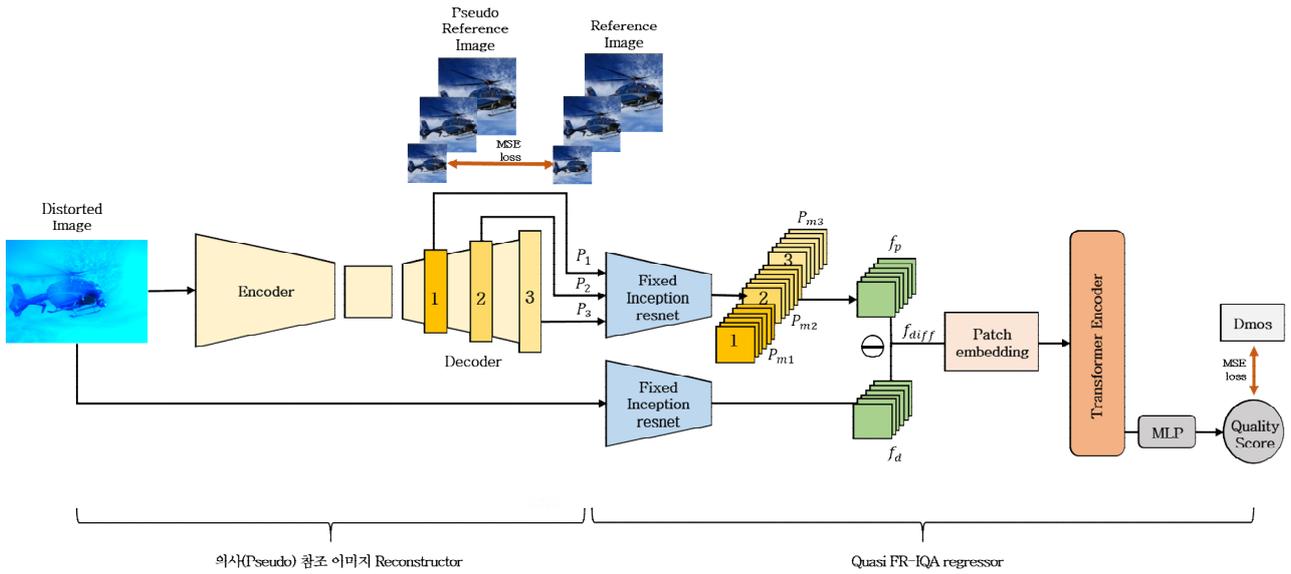


그림 2. MPIQ의 전체 구조: 의사 참조 이미지 restructor, 재구성된 의사 참조 이미지와 왜곡 이미지 특징 차이 추출, 그리고 transformer 기반의 화질 평가 점수 예측을 위한 quasi FR-IQA regressor

Fig. 2. Overall framework of MPIQ: pseudo reference image restructor, extraction of the difference between pseudo reference image and distorted image, and transformer-based quasi FR-IQA regressor aiming for prediction of quality score.

MPIQ(multi-scale pseudo image quality assessor)를 설계하였으며 전체적인 모델의 구조를 그림 2에 도식화하였다. MPIQ는 의사 참조 이미지로의 복원을 수행하는 restructor와 복원된 의사 참조 이미지와 왜곡 영상 간 차이를 바탕으로 화질 점수를 예측하는 quasi FR-IQA regressor의 두 부분으로 구성된다. 특히 이 과정에서 multi-scale distortion을 고려함으로써 이미지 크기에 따른 화질 열화 인지 수준의 차이를 반영할 수 있도록 하였다.

1. 의사 참조 이미지 Reconstructor

첫 번째로 FR-IQA 접근법과 유사하게 NR-IQA에서의 참조 이미지 역할을 수행하는 의사 참조 이미지 추출을 목적으로 인코더-디코더 구조를 갖는 restructor $F_\phi(\cdot)$ 를 구축하였다. 의사 참조 이미지 restructor는 기본적으로 왜곡 이미지의 참조 영상으로 재복원을 목표로 학습되나, 단순한 왜곡 제거뿐만이 아닌 주관적 화질 점수인 MOS 지도로 학습됨에 따라 왜곡 영상과의 공간적 왜곡 차이를 구분할 수 있도록 내부 특징이 학습되고 이를 반영한 차이 정보를 regressor에 전달될 수 있도록 설계하였다. $F_\phi(\cdot)$ 는 입력되는 256×256 왜곡 이미지 I_d 의 공간적 차원을 32×32 까지 축소함으로써 저차원의 매니폴드를 구성하고 이를 다시 256×256 으로 확장하며, 이를 위한 신

경망 구조는 VQGAN^[23]에서의 생성망 구조를 활용하였다. $F_\phi(\cdot)$ 의 세부 레이어 구조는 표 2에 나타내었다. 특히 인코더와 디코더를 거쳐 복원된 단일 크기의 이미지만을 화질 평가의 의사 참조로 활용하는 것이 아닌, 디코더의 2-4번째 residual block에서 다중 스케일을 갖는 3개($P_1, P_2, P_3 = F_\phi(I_d)$)의 특징들을 모두 추출해 활용한다. 단일 스케일에서의 화질 열화 특징 추출 보다 P_1, P_2, P_3 을 모두 활용함으로써 인간 시각에서의 화질 이해와 유사하게 지역적 왜곡을 다양한 수용장으로 관찰함으로써 전역적 왜곡 수준 역시 다중 스케일로 이해하고, 화질 점수를 예측하기 위한 왜곡 특징 추출의 품질을 향상시킬 수 있으며, 추가적으로 디코딩 과정에서의 공간 정보 손실을 최소화 할 수 있다는 장점이 있다.

또한 이 restructor는 의사 이미지를 추출할 수 있으므로 어떠한 FR 모델을 활용해도 NR-IQA 모델처럼 사용 가능하다는 장점도 있다.

본 연구에서는 $P_1 \in \mathbf{R}^{32 \times 32 \times 512}$, $P_2 \in \mathbf{R}^{64 \times 64 \times 128}$, $P_3 \in \mathbf{R}^{512 \times 512 \times 164}$ 의 총 3개 스케일 특징을 이용하였으며, 추출된 특징의 공간 차원을 이미지 크기와 동일하게 256×256 으로 변경하고 합성곱 연산을 통해 채널 크기 역시 3으로 변경한 $\hat{P}_1, \hat{P}_2, \hat{P}_3$ 로써 학습에 활용하였다. 여기서 추출된 3개의 각 스케일의 특징은

디코딩 되는 의사 참조 이미지가 정상 이미지인 I_r 에 가깝게 유도하기 위해 I_r 과의 reconstruction loss를 적용하였다.

표 1. 의사 참조 이미지 reconstructor 구조
Table 1. Network architecture of pseudo reference image reconstructor.

Encoder	Decoder
1×1 convolution $\rightarrow \mathbf{R}^{h \times w \times C}$	1×1 convolution $\rightarrow \mathbf{R}^{\frac{h}{8} \times \frac{w}{8} \times d}$
Residual Block $\rightarrow \mathbf{R}^{h \times w \times C}$	GroupNorm
Downsample $\rightarrow \mathbf{R}^{\frac{h}{2} \times \frac{w}{2} \times C}$	Residual Block $\rightarrow \mathbf{R}^{\frac{h}{8} \times \frac{w}{8} \times 4C}$
Residual Block $\rightarrow \mathbf{R}^{\frac{h}{2} \times \frac{w}{2} \times 2C}$	Upsample $\rightarrow \mathbf{R}^{\frac{h}{8} \times \frac{w}{8} \times 4C}$
Downsample $\rightarrow \mathbf{R}^{\frac{h}{4} \times \frac{w}{4} \times 2C}$	Residual Block $\rightarrow \mathbf{R}^{\frac{h}{4} \times \frac{w}{4} \times 4C}$
Residual Block $\rightarrow \mathbf{R}^{\frac{h}{4} \times \frac{w}{4} \times 4C}$	Upsample $\rightarrow \mathbf{R}^{\frac{h}{4} \times \frac{w}{4} \times 2C}$
Downsample $\rightarrow \mathbf{R}^{\frac{h}{8} \times \frac{w}{8} \times 4C}$	Residual Block $\rightarrow \mathbf{R}^{\frac{h}{2} \times \frac{w}{2} \times 2C}$
Residual Block $\rightarrow \mathbf{R}^{\frac{h}{8} \times \frac{w}{8} \times 4C}$	Upsample $\rightarrow \mathbf{R}^{\frac{h}{2} \times \frac{w}{2} \times C}$
GroupNorm	Residual Block $\rightarrow \mathbf{R}^{h \times w \times C}$
1×1 convolution $\rightarrow \mathbf{R}^{\frac{h}{8} \times \frac{w}{8} \times d}$	1×1 convolution $\rightarrow \mathbf{R}^{h \times w \times C}$

2. Quasi FR-IQA regressor

본 논문에서는 재구성된 의사 참조 이미지를 도출하여 기존 제안되었던 FR-IQA 접근법^[24]을 활용할 수 있는 프레임워크를 구성하였다. 의사 참조 이미지 reconstructor를 통해 재구성된 의사 참조 이미지와 왜곡 이미지 간 차이를 입력으로 MOS에 학습되어 화질 점수를 예측하도록 한다. 특히, 해당 과정에서 CNN을 통해 추출된 지역적 특징을 transformer를 통해 화질 열화 인지에 있어 인간 시각 특징과 유사하게 전역 공간에서의 전체적인 왜곡 수준으로 이해할 수 있도록 신경망을 구성하였다. 자세한 구조를 살펴보면 $F_\theta(\cdot)$ 에서 추출한 P_1, P_2, P_3 와 왜곡 이미지인 I_d 가 가중치를 공유하는 동일한 CNN에 입력됨으로써 지역적 왜곡을 이해하고 특징을 추출하도록 한다. 여기서 지역적 왜곡을 이해하고 특징을 추출하기 위한 CNN은 ImageNet 데이터셋으로 사전 학습된 Inception-Resnet-V2^[25]을 활용하였으며, 학습 과정에서 CNN의 가중치는 고정하였다. 특징 맵의 추출은 기존 IQT^[24]와 동일하게 6개 중간층(mixed_5b, block35_2, block35_4, block35_6, block35_8, block35_10)을 활용하였으며, 본 논문에서는 CNN의 입력으로 3개의 다중 스케일 특징 맵 P_1, P_2, P_3 을 활용함에 따라 결과적으로 각 스케일 별 6개

의 특징 맵을 출력한다. 스케일 별 출력된 각각의 특징 맵을 채널 단위로 연결하여 P_{m1}, P_{m2}, P_{m3} 를 구성하고, 이들을 또 다시 연결하고 단일 합성곱 레이어를 통과시킴으로써 하나의 특징 맵인 f_p 로 변환한다. 즉, f_p 는 의사 참조 이미지에 해당하는 특징맵으로써 활용되며, 왜곡 이미지 I_d 역시 CNN을 통과함으로써 지역적 특성을 반영한 특징 맵 f_d 로 가공된다.

본 논문에서는 추출된 두 특징 맵 f_p, f_d 의 차이 $f_{diff} = f_p - f_d$ 로써 전반적인 영상 화질 저하 수준을 추론할 수 있도록 하며, 이 과정에서 의사 참조 이미지 및 왜곡 이미지에 대한 전역적 영상 왜곡 수준 이해와 이를 반영하는 고수준 특징 추출을 통해 화질 점수를 추론함이 목적이다. 즉, 의사 참조 이미지와 왜곡된 이미지 간의 특징 공간 상 거리를 멀게 하는 방향으로 모델이 학습되어 f_p 가 추출된다.

f_{diff} 는 transformer 입력으로 들어가 전역 왜곡 수준을 이해할 수 있도록 한다. 입력된 $f_d \in \mathbf{R}^{32 \times 32 \times 1024}$ 는 1×1 합성곱을 통해 보다 저차원인 512 채널로 매핑되며, 픽셀별 토근화를 통해 $32^2 = 1024$ 길이의 토큰 시퀀스 \hat{f}_{diff} 으로써 transformer에 입력된다. 토큰 시퀀스의 위치 인코딩을 위해 학습 가능한 임베딩 벡터를 추가하였으며 이는 이후 지역 왜곡 간 장기 의존성 파악에 따른 전체적인 화질 수준 표현을 나타내기 위함이다.

화질 인코딩을 위한 transformer 신경망은 ViT^[13]와 유사한 구조로 구성했으며 2개의 레이어를 활용하였다. 본 논문에서는 transformer 입력이 CNN에서 추출되는 특징 맵을 기반으로 한 토큰 시퀀스가 적용되기 때문에 hybrid transformer 형태로 $Q_\theta(\cdot)$ 를 구성하였다. 이로써 CNN의 특징인 지역적 왜곡 특성 이해와 transformer의 전역적 특성을 함께 반영할 수 있어 화질 점수 예측에 강인한 성능을 갖는다.

3. 모델 학습

제안하는 MPIQ는 의사 참조 이미지 reconstructor $F_\theta(\cdot)$ 부터 quasi FR-IQA regressor $Q_\theta(\cdot)$ 까지 end-to-end로 학습함으로써 NR-IQA로 동작하며, quasi FR-IQA regressor 이전 특징 추출을 위한 Inception-Resnet-V2은 학습 과정에서 고정한다.

$F_\theta(\cdot)$ 에서의 multi-scale 왜곡 특징 반영

표 2. 기존 영상 화질 평가 기술과의 성능 비교

Table 2. Performance comparison with previous IQA methods.

Method	LIVE		CSIQ		KADID		TID2013	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
BRISQUE	0.925	0.923	0.744	0.812	0.538	0.527	0.654	0.626
DIIVINE	0.906	0.872	0.756	0.792	0.475	0.410	0.644	0.591
ILNIQE	0.904	0.865	0.806	0.808	0.558	0.534	0.678	0.521
BIECON	0.961	0.960	0.823	0.815	0.597	0.586	0.790	0.725
WaDIQaM	0.954	0.955	0.796	0.781	0.662	0.656	0.772	0.716
DBCNN	0.968	0.971	0.884	0.879	0.774	0.763	0.797	0.764
HyperIQA	0.956	0.952	0.871	0.865	0.761	0.752	0.796	0.760
HOSA	0.902	0.896	0.742	0.741	0.567	0.530	0.695	0.683
MPIQ-S	0.895	0.898	0.937	0.935	0.825	0.806	0.715	0.735
MPIQ (Ours)	0.956	0.927	0.954	0.946	0.889	0.881	0.885	0.867

reconstruction loss는 디코더에서 추출되는 $h \times w$ 크기의 특징 맵 $\hat{P}_1, \hat{P}_2, \hat{P}_3$ 와 정상 이미지 I_r 간 MSE(mean squared error)를 이용하였다.

$$L_{rec} = MSE(F_\phi(I_d) - I_r) = \frac{1}{hw} (\|\hat{P}_1 - I_r\|_2^2 + \|\hat{P}_2 - I_r\|_2^2 + \|\hat{P}_3 - I_r\|_2^2). \quad (1)$$

Quasi FR-IQA regressor $Q_\theta(\cdot)$ 는 데이터셋 내 i 번째 이미지에 대해 의사 참조 이미지 특징과 왜곡 이미지 특징 차이를 반영한 토큰 시퀀스 $\hat{f}_{diff,i}$ 를 입력으로 MOS에 학습됨으로써 실제 화질 수준을 이해하고 정량적 화질 점수를 예측하기 위함이며, L1 loss를 활용하였다.

$$L_{reg} = \sum_i |Q_\theta(\hat{f}_{diff,i}) - MOS_i|. \quad (2)$$

최종적으로 제안하는 MPIQ의 $F_\phi(\cdot)$ 과 $Q_\theta(\cdot)$ 최적화를 위한 전체 목적함수는 다음과 같다.

$$\min_{\phi, \theta} (L_{rec} + \lambda L_{reg}). \quad (3)$$

여기서 λ 는 이중의 두 loss 간 균형을 위한 페널티이며 실험적으로 1을 이용하였다.

IV. 실험

1. 데이터셋

본 논문에서는 평가를 위해 KADID-10k^[26], LIVE^[27], CSIQ^[28] 및 TID2013^[29] 4개의 실제 왜곡된 이미지가 포함된 데이터베이스를 사용했다.

- KADID-10k^[26]는 25가지 유형의 왜곡을 가지며

10,125개의 왜곡 이미지, 81개의 참조 이미지 및 주관적인 품질 점수를 제공하는 대규모 이미지 품질 평가 데이터셋이다. 왜곡 유형으로는 JPEG 압축, 가우시안 블러, 노이즈 및 공간 왜곡 등이며, 크라우즈소싱을 통해 2,209명의 평가자는 왜곡된 이미지의 품질을 0에서 100까지의 등급으로 나눠 평가하였다.

- LIVE^[27]는 JPEG2000 압축, JPEG 압축, 가우시안 블러, 화이트 노이즈 및 fast fading과 같이 총 5가지 유형들로 열화된 779장의 왜곡 이미지와 29장의 참조 이미지를 포함한다. 각 이미지에 대해 161명의 피실험자에 의해 평가된 0에서 100 사이의 점수를 제공한다.
- CSIQ^[28]는 30개의 참조 이미지로부터 생성된 866장의 왜곡 이미지들과 그에 따른 품질 점수를 제공한다. 왜곡 유형은 JPEG, JPEG2000, 대비 감소, 가우시안 노이즈 및 가우시안 블러와 같이 6가지로 분류되며 3-5단계의 열화 등급을 가지고 있다. 35명의 피실험자들에게 왜곡 이미지를 보여주고 각 이미지에 대한 품질 점수를 0에서 100 사이로 평가했다.
- TID2013^[29]는 25개의 참조 이미지와 24가지 유형의 왜곡으로 구성된 3000개의 왜곡 이미지를 제공한다. 총 985명의 피실험자는 왜곡 이미지의 시각적 품질을 평가했으며, 각 이미지에 대해 0에서 9 사이의 점수 범위를 제공한다.

2. 구현 상세

Quasi FR-IQA regressor의 transformer 차원을

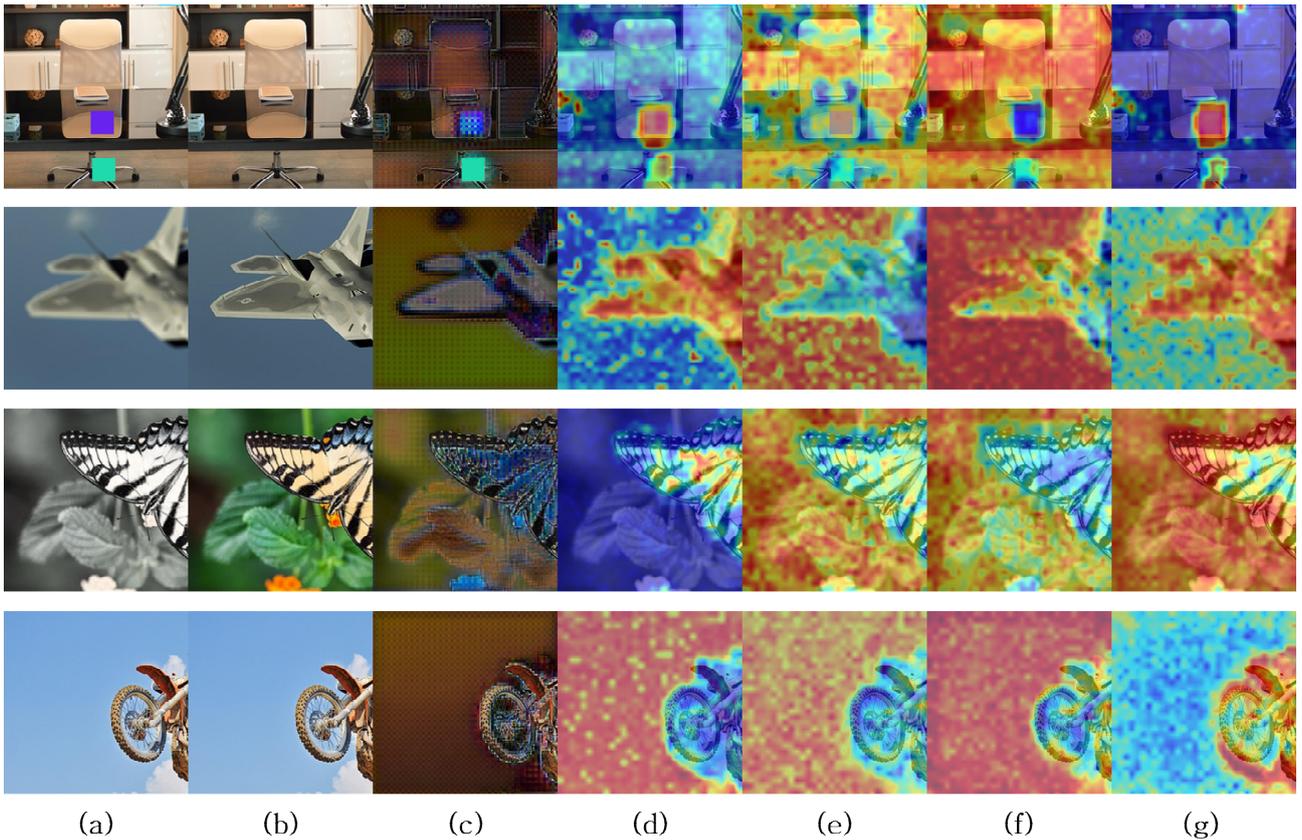


그림 3. 왜곡 타입 별 attention 맵의 시각화. (행 순서대로 color block, Gaussian blur, change of color saturation, darken): (a) 왜곡 영상; (b) 참조 영상; (c) f_{diff} ; (d)-(g) 각 헤드 별 attention 맵

Fig. 3. Visualization of attention map in accordance with the distortion types (color block, Gaussian blur, change of color saturation, and darken, in row order): (a) Distorted image; (b) reference image; (c) f_{diff} ; (d)-(g) attention map of each head.

512로 설정하였고, 헤드 개수는 4개, 인코더에서의 position-wise feed-forward 레이어의 차원은 512, 인코더 레이어의 수는 2개, 예측 헤드의 은닉 공간 차원은 2048로 설정하였다. quasi FR-IQA regressor의 인코더에서 특징맵의 토큰 개수는 32×32 로 1024개로 주어진다. 본 논문은 초기 학습률 10^{-5} 으로 코사인 냉각을 적용한 스케줄러를 적용했으며, 미니 배치 크기를 8로 설정하고 Adam optimizer를 사용하여 MPIQ를 학습하였다. NVIDIA RTX3090 GPU를 통해 모델 학습을 진행하였다.

3. 성능 비교

이미지 화질 평가 모델의 전반적인 성능을 정량적으로 평가하기 위해, 예측 화질 점수와 MOS와의 상관 계수를 계산하였다. 상관성 지표로는 피어슨 선형 상관 계수(Pearson linear correlation coefficient; PLCC)^[30], 스피어만 순위 상관 계수(Spearman

rank-order correlation coefficient; SRCC)^[31]를 사용하였다. 본 논문에서 제안하는 MPIQ 모델의 성능 검증을 위해 기존 NR-IQA 모델들과 네 가지 데이터셋(LIVE, CSIQ, KADID-10k, TID2013)에 대해 PLCC와 SRCC를 검증하였다.

표 2는 평가 결과를 보여주며 각 데이터베이스에 대해 80%는 훈련 데이터셋으로 사용하고 나머지 20%는 테스트 데이터셋으로 사용했으며, 다른 IQA 모델보다 세 가지 데이터셋(CSIQ, KADID-10k, TID2013)에서 더 우수한 성능을 보인다. 표 2에서 MPIQ는 그림 2에서와 같이 다중 스케일 특징 3개(P_1 , P_2 , P_3)를 추출해서 학습시킨 프레임워크이며, MPIQ-S는 의사 참조 reconstructor에서 출력할 때 다중 스케일 특징을 적용하지 않고 마지막 레이어의 단일 스케일 특징 맵에서 추출된 P_3 과 왜곡 이미지 간 차이를 quasi FR-IQA regressor에 적용한 모델이다. 두 모델의 성능 차이를 비교해보았을 때, MPIQ의 예측 성능이 더 높은 것을

수치적으로 확인 가능하며, 이는 디코더에서 다중 스케일 특징맵 추출을 적용함으로써 다중 스케일 왜곡 특징을 효과적으로 활용한 것이 이미지 화질 평가의 성능 향상에 도움을 준다는 것을 방증한다.

그림 3은 KADID-10k 데이터셋의 이미지로 지역적 왜곡에 대한 전역적 attention 맵을 시각화한 것이다. 왼쪽에서부터 차례로 왜곡 이미지, 원본 이미지, f_{diff} 시각화한 이미지, 헤드1, 헤드2, 헤드3, 헤드4에 대한 이미지를 나타낸다. MPIQ가 집중하는 영역을 시각화함으로써 확인할 수 있듯, 헤드별 attention 맵이 화질 추론을 위해 강조하는 공간 영역이 일부 overlap 되기는 하나, 각각 헤드의 attention은 화질 추론에 있어 서로 다른 영역을 관찰함으로써 보다 정확한 전역적 화질 예측을 수행하는 것으로 판단할 수 있다. 즉, 지역적 왜곡과 전역적 화질 이해 모두에 초점을 맞추도록 설계한 제안 방법이 평가 성능을 높이는 데 있어 적절히 작용함을 증명할 수 있다. 더불어 지역적으로 집중된 왜곡(예: color block)이 존재하는 경우에는 해당 영역에 대한 attention 또한 집중되고 있음을 보여주고 있다. 눈여겨 볼 사항은 이미지 왜곡 뿐 아니라 화질 예측을 하기 위해 이미지 내 의미론적 이해도 수행하는 것처럼 attention이 동작하며, 모델에 어떠한 의미론적 요소도 추가하지 않았음에도 불구하고 인간 시각과 유사하게 가장 중요한 객체 영역에 주의를 기울이고 있다는 결과를 나타낸다. 즉, 제안 모델은 저수준 왜곡 특징에만 집중하고 있는 것이 아닌, 이들의 전역적 의존성과 의미론적 측면 모두를 활용함을 그림 3의 가시화 결과로부터 확인할 수 있다.

4. 교차 검증 테스트

표 3. 세 가지 테스트 데이터셋에 대한 MPIQ-S 및 MPIQ 모델의 교차 검증 성능 비교
Table 3. Cross-validation performance tests on MPIQ-S and MPIQ models for three datasets.

Train on	KADID-10					
	LIVE		TID2013		CSIQ	
Test on	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
MPIQ-S	0.691	0.690	0.517	0.496	0.677	0.594
MPIQ (Ours)	0.863	0.851	0.645	0.619	0.836	0.789

의사 참조 이미지 reconstructor의 디코더에서 다중 스케일의 왜곡을 통해 3개 크기의 의사 참조 이미지를 추출하여 MPIQ 모델을 학습하였다. MPIQ-S는 의사

참조 이미지 reconstructor의 디코더에서 다중 스케일을 적용하지 않고 입력 이미지인 왜곡 이미지와 같은 크기의 출력 이미지를 통해 학습한 단일 스케일 모델이다. 표 4에서는 여러 데이터셋으로 교차 검증을 시행하였으며, MPIQ-S와 MPIQ 모두 KADID-10k 데이터셋으로 학습하고 나머지 데이터셋으로 테스트한 결과를 나타내었다. 결과에서 확인 가능하듯 다중 스케일 왜곡을 반영한 MPIQ 모델 성능이 더 우수하며, 이는 곧 해당 모델이 크기 불변한 왜곡 패턴을 훌륭히 캡처함을 방증한다. 예를 들어, 고해상도 이미지에서는 덜 시각적인 압축 아티팩트가 존재할 수 있으나 다중 스케일의 왜곡 분석을 통해 MPIQ 모델은 단일 특징을 추출하는 MPIQ-S 모델보다 더 강인한 화질을 예측한다. 또한, 의사 참조 reconstructor 역할 자체가 단순히 참조 이미지 재구성을 넘어 Quasi FR-IQA regressor에 입력되는 왜곡 차이를 강조하는 방향으로 학습되었음을 알 수 있다. 따라서, 제안하는 MPIQ는 다양한 해상도에 대해 보다 일관된 예측 화질 점수를 제공한다.

V. 결론

본 논문에서는 의사 참조 이미지 reconstructor를 도입해 왜곡 이미지와의 화질 저하 수준 차이를 학습하고 이를 FR-IQA의 프레임워크를 차용해 MOS에 회귀하는 NR-IQA 방법을 제안하였다. 다중 스케일 왜곡까지 고려하여 지역적/전역적 왜곡 수준을 객관화 가능함을 보였으며, 모델 내부에서 화질 이해를 위해 도출하는 attention 맵을 분석함으로써 열화 패턴에 대한 인간 시각 인지와 유사하게 MPIQ 모델이 동작함을 확인하였다. 본 논문과 유사한 접근을 바탕으로 이미지뿐만이 아닌 비디오 화질 평가로의 확장이 가능하며, 추후 3D 장면과 같이 아직 정량화된 메트릭이 규정되지 않은 다양한 도메인에 대해 객관화된 지표를 마련할 수 있을 것으로 예상된다. 더불어, 고전적인 영상 왜곡 패턴을 넘어 PIPAL^[32]과 같이 생성모델을 통해 복원된 영상에 대한 화질 평가에도 적용 가능하다.

REFERENCES

[1] H. Zhu, L. Li, J. Wu, W. Dong and G. Shi, "MetaIQA: deep meta-learning for no-reference image quality assessment," in *Proc. IEEE/CVF Conf. on Computer Vision*

- and *Pattern Recognition (CVPR)*, pp. 14131–14140, Seattle, WA, USA, August 2020.
- [2] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1969–1977, Honolulu, HI, USA, November 2017.
- [3] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DWT coefficients of locally weighted gradient magnitudes," *IEEE Trans. Image Processing*, vol. 25, no. 11, pp. 5293–5303, November 2016.
- [4] P. Ye, J. Kumar, L. Kang and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1098–1105, Providence, RI, USA, July 2012.
- [5] Saad, Michele A., Alan C. Bovik and Christophe Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Processing*, vol. 21, no. 8, pp. 3339–3352, March 2012.
- [6] L. Zhang, L. Zhang and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Processing*, vol. 24, no. 8, pp. 2579–2591, April 2015.
- [7] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Processing*, vol. 27, no. 1, pp. 206–219, October 2017.
- [8] J. Kim, A.-D. Nguyen and S. Lee, "Deep CNN-based blind image quality predictor," *IEEE Trans. Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 11–24, January 2019.
- [9] K. Ding, K. Ma, S. Wang and E. P. Simoncelli, "Image quality assessment: unifying structure and texture similarity," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2567–2581, December 2022.
- [10] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 3667–3676, Seattle, WA, USA, June 2020.
- [11] J. You and J. Korhonen, "Transformer for image quality assessment," in *Proc. IEEE International Conf. on Image Processing (ICIP)*, pp. 1389–1393, Anchorage, AK, USA, August 2021.
- [12] A. Chubarau and J. Clark, "Vtamiq: transformers for attention modulated image quality assessment," *arXiv:2110.01655*, October 2021.
- [13] A. Dosovitskiy, L. Beyer, K. Alexander, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, J. Uszkoreit and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. International Conf. on Learning Representations (ICLR)*, May 2021.
- [14] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," in *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, March 2013.
- [15] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Trans. Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [16] A. K. Moorthy and A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Trans. Image Processing*, vol. 20, no. 12, pp. 3350–3364, December 2011.
- [17] L. Kang, P. Ye, Y. Li and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1733–1740, Columbus, OH, USA, June 2014.
- [18] H. Talebi and P. Milanfar, "Nima: Neural image assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, August 2018.
- [19] D. Varga, D. Saupe and T. Szirányi, "Deepn: A Content Preserving Deep Architecture for Blind Image Quality Assessment," in *Proc. IEEE International Conf. on Multimedia and Expo (ICME)*, pp. 1–6, San Diego, CA, USA, 2018.
- [20] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z.

- Wang and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Processing*, vol. 27, no. 3, pp. 1202-1213, March 2018.
- [21] S. A. Golestaneh, S. Dadsetan and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV)*, pp. 3989-3999, Waikoloa, HI, USA, 2022.
- [22] J. Ke, Q. Wang, Y. Wang, P. Milanfar and F. Yang, "MUSIQ: multi-scale image quality transformer," in *Proc. IEEE/CVF International Conf. on Computer Vision (ICCV)*, pp. 5128-5137, Montreal, QC, Canada, 2021.
- [23] P. Esser, R. Rombach and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 12873-12883, 2021.
- [24] C. Manri, S. Yoon, B. Kang, and J. Lee, "Perceptual image quality assessment with transformers," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 433-442, 2021.
- [25] C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI conf. on artificial intelligence*, vol. 31, no. 1, 2017.
- [26] H. Lin, V. Hosu and D. Saupe, "KADID-10k: A Large-scale Artificially Distorted IQA Database," in *Proc. Eleventh International Conf. on Quality of Multimedia Experience*, 2019.
- [27] H. R. Sheikh, M. F. Sabir and A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440-3451, November 2006.
- [28] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, January 2010.
- [29] N. Ponomarenko et al., "Color image database TID2013: Peculiarities and preliminary results," *European Workshop on Visual Information Processing (EUVIP)*, pp. 106-111, Paris, France, 2013.
- [30] K. Pearson, "Mathematical contributions to the theory of evolution.—iii. regression, heredity and panmixia," *Philos. Trans. R. Soc. London*, vol. 187, pp. 253-318, January 1896.
- [31] C. Spearman, "The proof and measurement of association between two things," *American Journal of Psychology*, vol. 15, pp. 72-101, January 1904.
- [32] J. Gu, H. Cai, H. Chen, X. Ye, J. S. Ren and C. Dong, "Pipal: a large-scale image quality assessment dataset for perceptual image restoration," in *Proc. of European Conf. Computer Vision (ECCV)*, 2020.

— 저 자 소 개 —



윤 윤 정(비회원)
2021년 한성대학교 IT융합공학부 졸업
2021년~한성대학교 AI응용학과 석사과정

<주관심분야: 영상처리, 컴퓨터비전>



오 희 석(정회원)
2017년 연세대학교 전기전자공학과 박사 졸업
2017년~2017년 삼성전자 DMC 연구소 책임연구원
2017년~2020년 한국전자통신연구원 선임연구원

2020년~현재 한성대학교 AI응용학과 조교수
<주관심분야: 영상처리, 컴퓨터비전, 혼합현실, 심층생성모델 등>