

고해상도 텍스처 맵 생성을 위한 딥러닝 기반 초해상도 기법들의 비교 분석 연구

김혜주^o 나재호*

상명대학교

hyeju724@naver.com jaeho.nah@smu.ac.kr

Comparative analysis of the deep-learning-based super-resolution methods for generating high-resolution texture maps

Hyeju Kim^o Jah-Ho Nah*

Sangmyung University

요약

디스플레이의 해상도의 증가에 따라 고해상도 텍스처 맵을 내장한 앱들도 함께 증가하는 추세에 있다. 최근 딥러닝 기반 이미지 초해상도 기법들의 발전은 이러한 고해상도 텍스처 생성 작업을 자동화할 수 있는 가능성을 만들고 있다. 하지만 이러한 적용 사례에 대해 심층적으로 분석한 연구는 아직 부족한 상태이다. 그래서 본 논문에서는 최신 초해상도 기법들 중 BSRGAN, Real-ESRGAN, SwinIR(classical 및 real-world image SR)을 택하여 텍스처 맵의 업스케일링(upsampling)에 적용한 후 그 결과를 정량적, 정성적으로 비교, 분석하였다. 실험 결과 업스케일링 후 나타나는 다양한 아티팩트(artifact)들을 발견할 수 있었으며, 이를 통해 기존 초해상도 기법들을 텍스처 맵 업스케일링에 곧바로 적용하기에는 일부 미흡한 부분이 존재한다는 점을 확인하였다.

Abstract

As display resolution increases, many apps also tend to include high-resolution texture maps. Recent advancements in deep-learning-based image super-resolution techniques make it possible to automate high-resolution texture generation. However, there is still a lack of comprehensive analysis of the application of these techniques to texture maps. In this paper, we selected three recent super-resolution techniques, namely BSRGAN, Real-ESRGAN, and SwinIR (classical and real-world image SR), and applied them to upscale texture maps. We then conducted a quantitative and qualitative analysis of the experimental results. The findings revealed various artifacts after upscaling, which indicates that there are still limitations in directly applying super-resolution techniques to texture-map upscaling.

키워드: 딥러닝, 초해상도 기법, 텍스처 매핑, 이미지 프로세싱

Keywords: Deep Learning, Super-resolution, Texture mapping, Image Processing

1. 서론

데스크탑 모니터, TV, 스마트폰, 태블릿에 탑재되는 디스플레이의 해상도는 계속해서 증가하는 추세에 있다. 그 결과, 모바일에서는 WQHD+ 해상도가, 모니터 및 TV에서는 4K UHD 해상도가

대중화되었다. 이러한 고해상도 디스플레이에서 3차원 그래픽스로 생성된 영상을 선명하게 즐기기 위해서는 렌더링 해상도뿐 아니라 오브젝트에 매핑되는 텍스처의 해상도도 함께 증가하여야 한다. 이에 따라 주요 게임 회사들은 고해상도의 HD 텍스처 팩을 별도로 배포하여 이러한 추세에 대응하고 있다. 하지만 기존에

*corresponding author: Jah-Ho Nah /Sangmyung University (jaeho.nah@smu.ac.kr)

만들어진 앱에 적용할 수 있는 HD 텍스처 팩을 별도로 제작하기 위해서는 디자이너의 많은 노고가 요구된다.

인공지능 기반의 초해상도 (super-resolution, SR) 기법은 이러한 단점을 해결하기 위한 방안 중 하나로 주목받고 있다. 이미 몇몇 게임의 경우에는 이용자가 직접 인공지능 기반의 업스케일링(upsampling) 기법을 적용하여 HD 텍스처팩을 제작, 배포하는 시도가 이루어지고 있다 [1, 2]. 모델을 바꾸지 않는 경우 메시 매개변수화(mesh parameterization)를 통해 텍스처 좌표를 다시 생성할 필요 역시 없어지기 때문에, 텍스처의 해상도를 높이는 것만으로도 화질 향상을 이끌어낼 수 있다. 보통 텍스처는 mip매핑(mip-mapping)[3]을 통해 적절한 해상도가 선택되어 물체에 입혀지는데, 레벨 0에 존재하는 가장 큰 텍스처의 해상도를 높이면 근거리에서 보다 높은 해상도의 mip맵(mip-map)이 선택될 수 있기 때문이다. 하지만 인공지능 기반의 초해상도 기법의 검증은 주로 사진 위주로만 이루어져 왔고, 3차원 그래픽스에 사용되는 텍스처 맵에 적합한 초해상도 기법에 대해서는 극히 드문 연구 결과만 존재한다 [4, 5].

본 논문에서는 최신 초해상도 기법들을 오프라인 상의 텍스처 업스케일링에 적용시 어떠한 결과를 나타내는지 비교, 분석하고자 한다. 그 대상으로 2021년에 발표된 대표적인 세 가지 초해상도 기법 BSRGAN [6], Real-ESRGAN [7], SwinIR [8]을 선택하였다. 그리고 사진, 게임 텍스처, GIS 데이터, 합성 이미지, 노멀 맵 등 다양한 형태의 텍스처에 대해 실험을 수행하여, 각 기법의 장단점을 파악하고자 하였다. 이미지 품질의 비교를 위해서는 FID [9] 알고리즘을 사용하였고, 이에 더하여 각 기법이 발생시키는 아티팩트(artifact)들을 분류하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 기존 초해상도 기법들 및 이미지 품질 측정 방법들에 대해 간략히 요약한다. 3장에서는 실험 방법 및 환경을 소개하고, 실험 결과의 분석을 통해 각 초해상도 기법들을 정량적, 정성적으로 비교한다. 또한 렌더러 상에서 초해상도 기법이 적용된 텍스처를 실제로 물체에 매핑(mapping)한 결과 및 여기에 엔비디아의 DLSS[10]를 함께 적용한 결과를 비교 분석한다. 4장에서는 결론과 함께 향후 연구 방향을 제시하며 본 논문을 마무리한다.

2. 기존 연구

2.1 초해상도 기법

최초의 딥러닝 기반 초해상도 기법인 SRCNN [11]은 합성곱 신경망(CNN)을 초해상도 알고리즘에 적용하였다. 이후 Christian 등 [12]은 비지도 학습에서 사용되는 생성적 적대 신경망(GAN)을 적용한 SRGAN을 제안하였다. 이러한 CNN과 GAN을 바탕으로 한 다양한 초해상도 기법들에 대한 연구가 활발히 진행되어 왔는데, ESRGAN [13]은 그 중 대표적인 기법 중 하나이다. 이 방법은 SRGAN의 배치 정규화(batch normalization)를 제거한 Residual in Residual Dense Block(RRDB)의 사용과 함께, 상대적

판별자(relativistic discriminator)와 인지적 손실(perceptual loss)을 적용하였다. 이를 통해 ESRGAN은 SRGAN 보다 더 현실감 있고 자연스러운 초해상도 결과를 보여주었다. 최근 ESRGAN보다 초해상도 품질을 더욱 개선한 RealSR [14], BSRGAN [6], Real-ESRGAN [7], SwinIR [8] 등이 발표되었다. 이 기법들은 공통적으로 해상도의 증가와 더불어 실제 사진에서 나타나는 여러가지 품질 열화 요소들(노이즈 등)을 최대한 제거하여 원본 이미지를 고품질의 이미지로 복원하려는 특징을 가지고 있다.

2020년 발표된 RealSR [14]은 현실 세계 이미지에 적합하도록 구현된 초해상도 기법이다. 이 방법은 저해상도 이미지 생성을 위해 다양한 블러(blur) 커널과 실제 노이즈 분포에 중점을 둔 새로운 열화 프레임워크를 사용하고, ESRGAN의 판별자를 패치 판별자(patch discriminator)로 변경하여 여러가지 아티팩트들을 방지하였다. 그 결과 노이즈가 많은 저해상도 이미지를 보다 효과적으로 복원할 수 있었다.

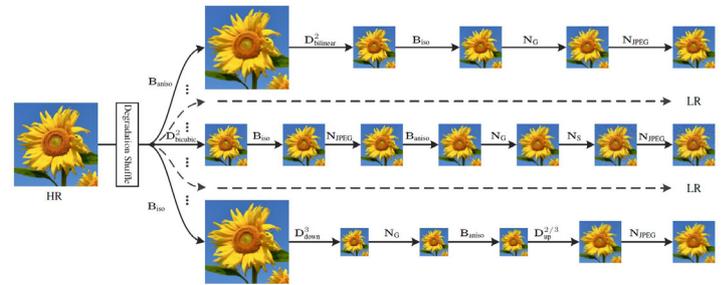


Figure 1: Degradation model of BSRGAN [6]

2021년 발표된 BSRGAN [6]은 합성 또는 실제 이미지에서 우수한 복원 성능을 보이기 위해 여러가지 다양한 열화 기법들을 기존 네트워크의 입력에 적용하였다. 먼저 블러(blur) 적용시 등방성 또는 비등방성 가우시안 커널을 무작위로 적용하고, 다운 샘플링시에도 최근방 보간법(nearest interpolation), 선형 보간법(bilinear interpolation), 3차 보간법(bicubic interpolation) 중 하나를 무작위로 선택한다. 또한 노이즈의 경우에도 각각 다른 노이즈 레벨을 적용한 가우시안 노이즈, 다른 품질 요소를 적용한 JPEG 압축 노이즈, 역방향 카메라 ISP(이미지 신호처리) 파이프라인 모델과 RAW 이미지 노이즈 모델을 통한 카메라 센서 노이즈 등을 생성, 사용한다. BSRGAN은 앞서 소개한 여러가지 기법들을 임의적으로 섞어 열화된 저해상도 이미지를 생성, 이를 학습에 이용한다.

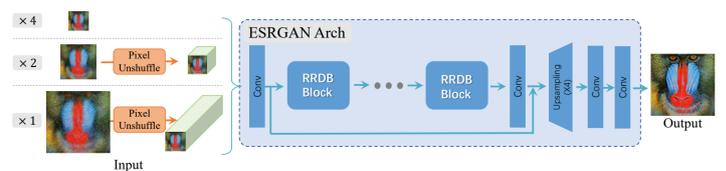


Figure 2: Degradation model of Real-ESRGAN [7]

BSRGAN과 유사한 시기에 발표된 Real-ESRGAN [7] 역시

ESRGAN에 다양한 전처리 열화 기법 단계를 추가하였다. 이 방법의 가장 큰 특징은 고차원(high-order) 열화 단계로, 블러, 다운 샘플링, 노이즈, JPEG 압축 등의 열화 기법들을 각기 다른 파라미터를 사용하여 여러 번 적용함으로써 학습에 필요한 저해상도 이미지를 생성한다. 또한 고주파 링잉(ringing) 및 오버슈트(overshoot) 아티팩트를 고려하여 sinc 필터를 블러 커널의 마지막 부분 및 JPEG 압축 과정에 적용하였다. 마지막으로 이러한 다양한 범주의 열화 현상을 다루기 위해 스펙트럼 정규화(spectral normalization)가 추가된 U-Net 판별자를 사용하였다. 그 결과, 이미지의 선명도 및 노이즈 측면에서 ESRGAN에 비해 많은 개선이 있었다.

역시 같은 해 발표된 SwinIR은 CNN과 Transformer [15]의 장점을 합친 스윈 트랜스포머(Swin Transformer)[16]를 기반으로 이미지를 복원하는 방법이다. SwinIR의 네트워크 구조는 얇은 특징 추출(shallow feature extraction), 깊은 특징 추출(deep feature extraction), 고품질 이미지 재구성 (high-quality image reconstruction)의 3개 모듈로 구성된다. 먼저 얇은 특징 추출 모듈은 컨볼루션 층(convolution layer)을 사용하여 얇은 특징들을 추출하고, 저주파 정보의 보존을 위해 이 특징들은 재구성 모듈로 직접 전송된다. 깊은 특징 추출 모듈은 여러 개의 스윈 트랜스포머 층, 컨볼루션 층, 잔차 연결로 이루어진 RSTB(Resident Swin Transformer Block)로 구성되어 있으며, 이를 통해 깊은 특징들을 추출한다. 마지막으로 고품질 이미지 재구성 모듈은 얇은 특징들과 깊은 특징들을 합쳐 고품질 이미지를 추출한다. 이러한 과정을 통해 SwinIR은 이전 기법보다 적은 파라미터로 초해상도 뿐만 아니라 이미지 압축 아티팩트 및 노이즈 제거 부문에서도 좋은 성능을 보여주었다. SwinIR은 여러가지 다른 방식의 이미지 초해상도 옵션을 제공하는데, 그 중 classical image SR은 훈련시 L_1 pixel loss 만 사용한 버전이고, real-world image SR은 이미지 향상을 위해 여기에 GAN loss와 perceptual loss를 더하여 훈련시킨 버전이다.

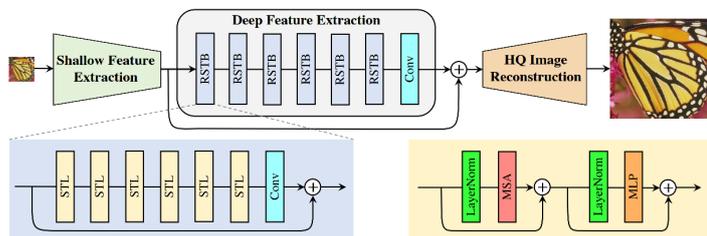


Figure 3: Architecture of SwinIR [8]

텍스처 맵의 해상도를 높이기 위한 초해상도 기법에 대해서도 몇 가지 연구가 확장된 요약문(extended abstract) 형태로 발표되었다. 2018년 Takemura [4]는 압축되지 않거나 ETC2 또는 ASTC 로 압축된 텍스처 맵의 해상도를 높이기 위한 4층 CNN 구조를 제안하였고, 기존 CNN 기반의 방식보다 나은 품질을 보여주었다. 최근 Nah와 Kim [5]은 훈련 파라미터 튜닝과 색상 보정 후처리를

통해 기존 Real-ESRGAN보다 업스케일링된 텍스처의 품질을 높였고, PSNR기반의 ASTC 블록 크기 결정 방법 [17]을 적용하여 업스케일링으로 인한 용량 증가를 감소시켰다.

최근 초해상도 기법은 텍스처 맵 뿐만 아니라 다른 분야의 이미지 해상도 증가를 위해서도 활발히 이용되고 있다. 2019년 조우영 외[18]는 초해상도 기법을 이용하여 작은 글씨체에 대한 인식 정확도를 비교연구한 논문을 발표하였고, 2020년 성상하 외[19]는 딥 러닝 기법을 활용한 이미지 내 한글 텍스트 인식에 관한 연구를 발표하였다. 텍스트 이외 의학 분야에서 또한 초해상도 기법에 대한 연구가 이루어지고 있다. Shi 외[20]는 의료영상으로의 적용에 초점을 두고 딥러닝 기반 초해상도 기법들을 조사 및 정리하였다. 지리학 분야 등에서도 관련 연구가 진행되고 있는데, 일례로 Jia 외[21]는 실제 항공 이미지를 초해상도 기법으로 재구성하기 위한 생성적 적대 네트워크(MA-GAN)를 제안하였다.

2.2 이미지 품질 측정 방법

대표적인 이미지 품질 측정 방법으로는 PSNR(Peak Signal to Noise Ratio)과 SSIM(Structural Similarity) [22]이 있다. 최근 발표된 FLIP [9]은 이 두 가지 방법의 단점을 보완하여, 참조 이미지와 대상 이미지 간의 차이를 좀 더 직관적이고 정확하게 보여준다.

PSNR은 이미지가 가질 수 있는 최대 신호에 대한 잡음의 비율을 나타내며 데시벨(dB) 단위로 표시된다. PSNR 수치가 높을수록 참조 이미지 대비 손실이 적다는 것을 의미하므로, 일반적으로 PSNR 수치가 높을수록 참조 이미지와 유사한 결과라고 볼 수 있다. 그러나 PSNR은 참조 이미지와의 수치적 차이만으로 품질을 평가하여, 인간이 실제로 보고 느끼는 품질의 정도와 다른 결과를 보여주기도 하는 한계를 갖고 있다.

SSIM은 구조적 유사도 지수(Structural Similarity Index Measure)의 약자로, 인간이 느끼는 시각적 품질 차이를 평가하기 위해 고안되었다. SSIM은 두 이미지의 유사도를 휘도(luminance), 대비(contrast), 구조(structure) 3가지 요소를 이용하여 비교한다. 값의 범위는 [0,1] 사이의 값을 가질 수 있으며, 1에 가까울수록 두 이미지가 유사하다는 것을 의미한다.

FLIP은 2020년 NVIDIA에서 발표한 논문으로, 이 논문의 방법은 서로 다른 두 이미지를 번갈아 뒤집을 때 보이는 차이점의 측정을 목표로 둔다. 이 방법은 참조 이미지와 대상 이미지 간의 색상과 특징을 독립적으로 비교하는 파이프라인을 사용하고, 여기에 다양한 기존 색상 및 지각(perception) 연구에서 나온 개념들을 적용함으로써, 렌더링 및 사진에서 보여지는 다양한 아티팩트들, 예를 들면 노이즈, 색상 왜곡, 에일리어싱(aliasing), 블러링(blurring)을 기존 방법들보다 정확히 판별해 낸다. 비교 결과는 오류 지도(error map), 히스토그램 및 평균값, 가중 중앙값(weighted median) 등으로 출력되며, 평균값 및 가중 중앙값은 SSIM과 마찬가지로 [0, 1] 사이의 범위를 가지지만, SSIM과 반대로 0에 가까울수록 원본과의 차이가 적음을 의미한다.

3. 실험 및 분석

3.1 실험 환경 및 방법

먼저 실험에 사용한 데이터 셋은 다음과 같다. 기본적으로 Quick-ETC2 [23, 24] 논문에서 사용된 64개의 텍스처 셋을 고려하였는데, 이는 다양한 형태 및 다양한 해상도의 텍스처를 포함하고 있고 텍스처 압축 품질 측정을 위해 여러 논문에서 사용되어 왔기 때문이다. 다만, 8K의 초고해상도로 카메라로부터 캡처된 7개의 이미지는 추가 업스케일링이 필요가 없기 때문에 실험에서 제외하였다. 그리고 노멀맵도 현실감을 증대시키기 위해 널리 쓰이는 텍스처형태중 하나이므로 추가로 WavererNormalMap Set[25,26]을 사용하였다.

기존 이미지 초해상도 논문들에서는 DIV2K [27] 등과 같이 2K 이상의 고해상도 이미지들 및 이를 여러 단계로 다운스케일링(downscaling)한 이미지들을 주로 사용한다. 하지만, 본 논문에서 사용할 텍스처 셋은 저해상도 텍스처와 고해상도 텍스처가 혼재되어 있기 때문에, 다음과 같은 두 가지 방법으로 실험을 진행하였다. 첫번째 방법은 원본 이미지의 가로 및 세로를 각각 1/2배로 다운스케일링한 다음, 이를 다시 2배로 업스케일링하여 원본과 비교하는 방식이다. 이 방식은 각각의 기법들이 디테일이 소실된 저해상도의 이미지로부터 얼마나 원본 이미지와 유사한 결과를 재구성해낼 수 있는지에 대한 능력을 확인하는 데 목적이 있다. 두 번째 방법은 원본 이미지를 2배로 업스케일링 한 다음 이를 다시 원본 크기로 다운스케일링하여, 이를 원본과 비교하는 방식이다. 이 방식에서는 앞선 방식과 달리 충분한 해상도가 제공되는 상태에서 업스케일링이 수행되기 때문에, 각 초해상도 기법들이 얼마나 원본과 다른 아티팩트를 만들어내는지 확인하는데 목적이 있다.

본 실험에서는 다운스케일링을 하기 위해 QualityScaler[28]를 사용하였다. 다운스케일링시 사용하는 보간법으로 INTER_AREA를 적용하였는데, 이는 여러 가지 다운스케일링 방식 중 높은 품질을 보여 상용 소프트웨어에서도 널리 사용되기 때문이다 [29].

실험을 통해 비교한 기법은 2장에서 소개한 이미지 초해상도 기법들 중 최신 기법인 BSRGAN, Real-ESRGAN, SwinIR(classical 및 real-world image SR)이다. SwinIR의 두 가지 옵션은 앞으로 편의상 SwinIR-classical 및 SwinIR-real로 지칭하고자 한다. 이 네 방식 모두 파라미터 수정 및 추가 훈련 없이 기본값을 이용하였다. 다만 SwinIR의 경우 RGBA 텍스처를 업스케일링시 알파값을 제외하고 RGB 형태로 출력하였기 때문에, 알파값이 포함된 9개 이미지에 대한 결과는 수치 산출시 제외하였다. 그리고 이러한 이유로 카테고리별 수치 산출시 알파 채널을 가지는 텍스처와 이를 가지지 않는 텍스처를 분리하여 계산하였다.

정량적 분석을 위해서는 2장에서 기술된 이미지 품질 측정 방

법 중 PSNR 및 FLIP을 사용하였다. FLIP의 경우 최신 버전인 1.2를 사용하였고, 이 프로그램에서 제공하는 여러 메트릭(metric) 중 가중 중앙값을 사용하였으며, PPD (pixels per degree) 등 설정 가능한 파라미터들은 모두 기본값을 사용하였다. PSNR, FLIP과 함께 많이 쓰이는 SSIM의 경우 합성 이미지나 사진 이미지를 평가할 때 잘못된 결과를 보여주는 경우가 존재하여[30], 본 논문에서는 이미지 품질 측정 방법에서 제외하였다.

추가적으로 낮은 해상도의 프레임을 입력으로 받아 더 높은 해상도로 출력해주는 DLSS[10]와의 관계를 알아보는 실험을 수행하였다. 이 실험에서는 DLSS 3 SDK[31]의 Sponza scene을 이용하여, 원래 해상도로 렌더링한 경우, DLSS만 적용한 경우, 이미지 초해상도 기법을 적용한 텍스처만 사용한 경우, 그리고 이 두 가지를 함께 사용한 경우를 비교하였다.

본 실험을 위해서 사용한 기기는 Intel Core i7-1260P, 16GB DDR4 RAM, Geforce RTX 2050 4GB 그래픽 카드를 포함하고, 윈도우11이 설치된 노트북이다. 텍스처 셋의 업스케일링 및 다운스케일링, PSNR 및 FLIP값 산출 등을 일괄적으로 처리하기 위해, Python 3.9으로 간단한 배치 프로그램을 개발하여 실험을 수행하였다. PSNR의 수치 산출을 위해서는 파이썬용 OpenCV 패키지를 사용하였다.

3.2 결과 분석

본 논문에서는 실험 결과 분석을 위해 텍스처 데이터 셋을 사진, 디퓨즈 맵(diffuse map) 위주로 선택된 게임 텍스처, GIS 데이터, 합성 이미지, 노멀 맵의 5개 카테고리 분류하였다. 이를 토대로 실험에 사용된 초해상도 기법들을 정량적, 정성적으로 분석하고자 한다.

3.2.1 정량적 결과 분석

먼저 FLIP 및 PSNR 값을 정량적으로 분석한 결과는 다음과 같다. 그림 4는 카테고리별로 각 초해상도 기법 간의 품질 차이를 수치적으로 비교한다. 참고로 FLIP은 낮을수록, PSNR은 높을수록 원본에 가까움을 의미한다. 대체로 수치상으로는 SwinIR-classical이 모든 카테고리에서 가장 좋은 결과를 보였다. 또한 일반 텍스처보다는 노멀 맵에서 타 기법에 비해 훨씬 더 왜곡이 적게 나타났다. 노멀 맵에서의 RGB 값은 각 텍셀이 저장하는 법선 벡터(normal vector)의 XYZ 축 방향을 의미하기 때문에, 노멀 맵의 색상 왜곡은 법선 곡률의 변형으로 이어진다. 그 다음으로 좋은 결과를 낸 기법을 살펴보면, 사진, 게임 텍스처, 합성 이미지에서는 BSRGAN이, GIS 지도 데이터에서는 BSRGAN과 Real-ESRGAN이 실험 방법(업스케일링과 다운스케일링 순서) 및 컬러 채널(알파값 유무)에 따라 혼재된 결과를 보였으며, 노멀 맵에서는 Real-ESRGAN이 상대적으로 나은 결과를 나타냈다.

각 초해상도 기법별로 가장 좋은 수치를 나타낸 카테고리를

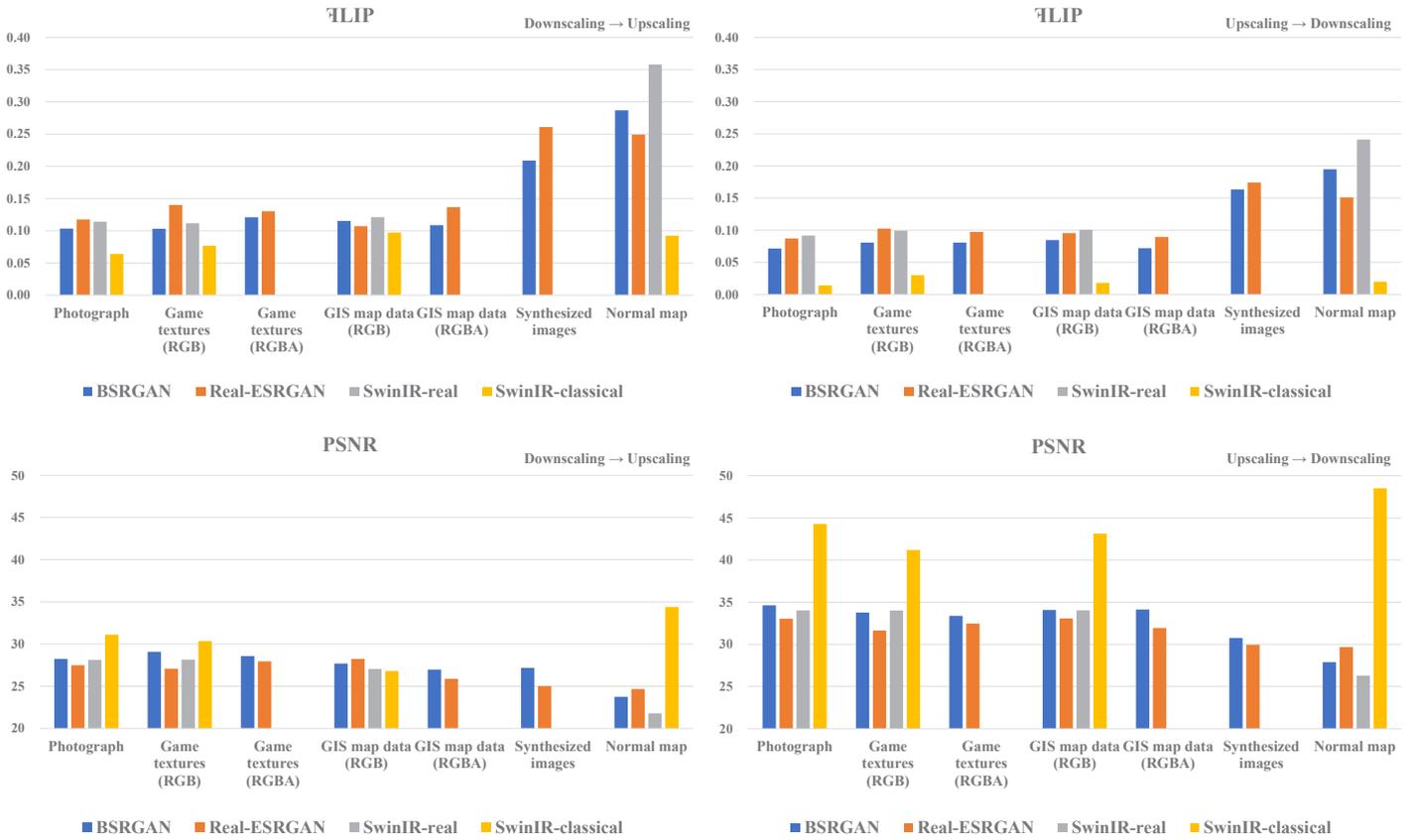


Figure 4: Average FID and PSNR comparison by category. Lower FID and higher PSNR values are better, respectively. Due to SwinIR ignoring the alpha channel, we excluded the RGBA results (including synthesized images) from SwinIR (grey and yellow bars).

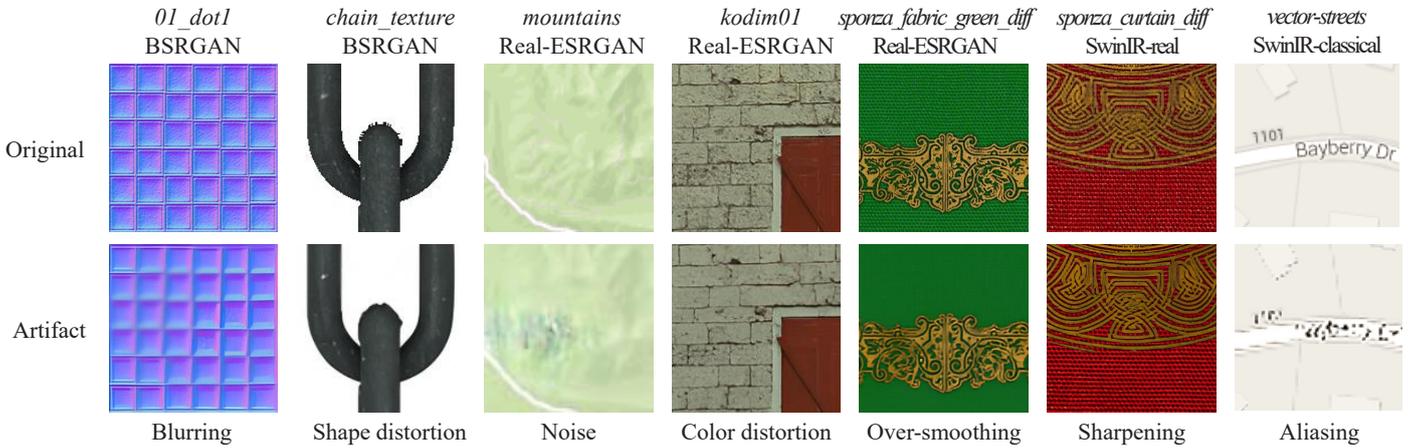


Figure 5: Representative artifact patterns appeared in output images after super-resolution. All images were upscaled after downscaling except *mountains* which was downsampled after upscaling.

살펴보면 다음과 같다. BSRGAN과 SwinIR-real은 대체로 사진 또는 게임 텍스처에서, Real-ESRGAN은 실험 방법에 따라 사진 또는 GIS 지도 데이터에서 가장 좋은 결과를 보여주었다. 반면 SwinIR-classical은 카테고리에 크게 영향을 받지 않는 것을 확인할 수 있었다. 이는 일반적인 사진 이미지 및 열화 기법을 이용하여 훈련된 초해상도 기법들이 사진으로부터 만들어진 텍스처에 잘 적용될 수 있지만, 다른 형태의 텍스처에서는 상대적으로 더

많은 아티팩트를 만들어낼 수 있음을 의미한다.

다운스케일링 후 업스케일링을 수행한 첫번째 실험 결과와 업스케일링 후 다운스케일링을 수행한 두번째 실험의 결과를 카테고리별로 살펴보면 다음과 같다. SwinIR-classical을 제외하고, 사진이나 게임 텍스처, GIS 지도 데이터에서는 FID 값의 차이가 아주 크지 않은 반면, 합성 이미지 및 노멀 맵에서는 이 차이가 크게 벌어지는 것이 확인된다. 이는 초해상도 기법 적용 후 특정

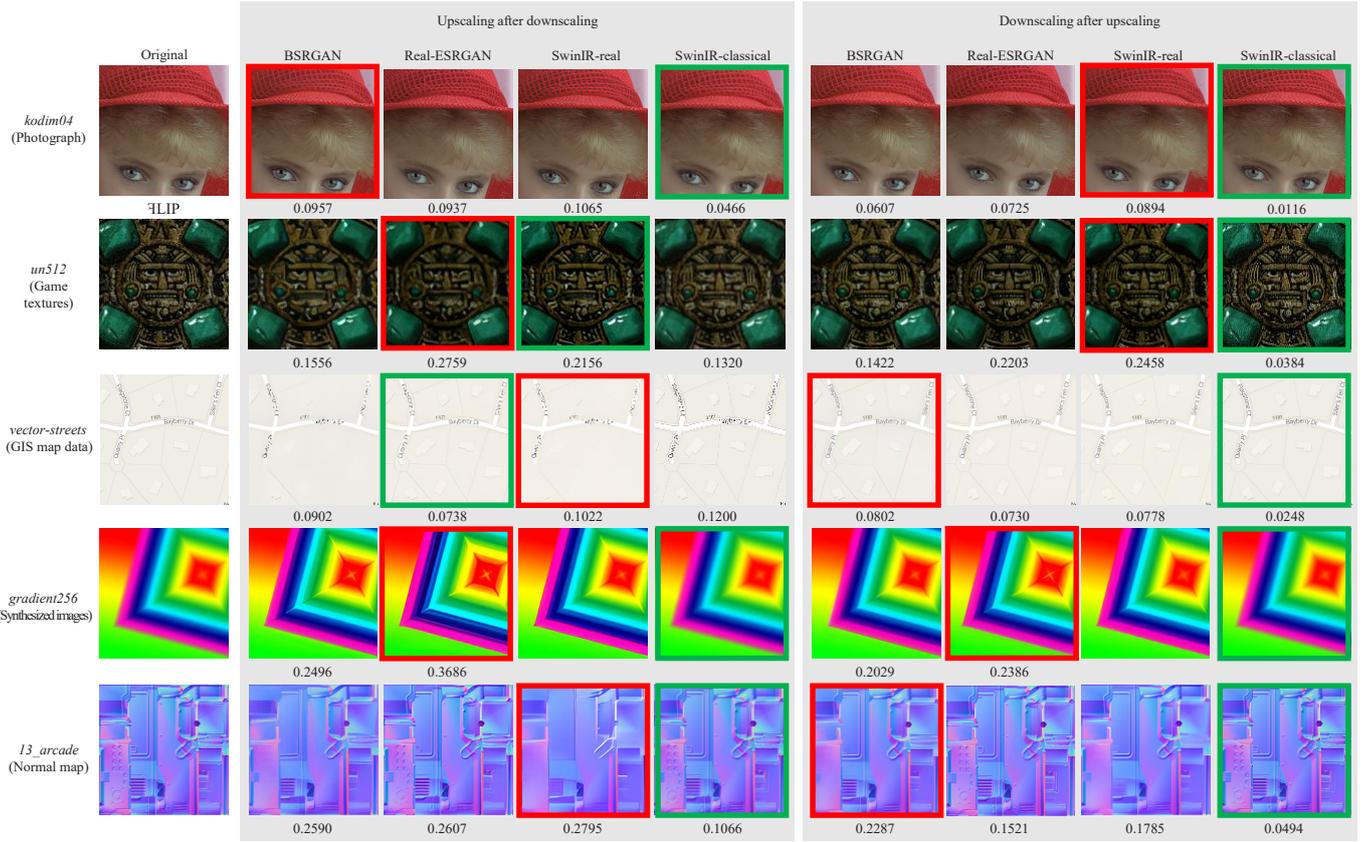


Figure 6: Qualitative comparisons of output images after super-resolution by category. The green and red squares represent the best and worst quality based on our visual comparison, respectively. The number below each image is the F1IP value (lower is better), and our best and worst choice may differ from the F1IP values. The F1IP value of SwinIR in *Gradient256* was not inserted into the figure because its original (four channels) and output images (three channels) have a different number of color channels.

한 아티팩트가 발생한 경우, 이러한 아티팩트가 원본 이미지의 해상도가 낮을 때 더 증폭될 수 있다고 해석할 수 있다.

3.2.2 정성적 결과 분석

정량적 분석의 결과, 텍스처 초해상도 적용시 SwinIR-classical 이 대체적으로 가장 좋은 품질을 냈다. 하지만 업스케일링 결과물은 텍스처의 형태 또는 초해상도 기법에 따라 매우 다른 양상을 보이므로, 시각적으로 보여지는 아티팩트의 형태에 따른 정성적 분석을 추가로 하고자 한다.

그림 5는 업스케일링 후 나타나는 여러가지 아티팩트의 형태를 보여준다. 먼저 첫번째 아티팩트로 *01_dot1*의 블러링 현상을 들 수 있다. 다운스케일링 후 BSRGAN으로 업스케일링을 실행한 결과, 좌측 상단 부분이 원본과 달리 커다랗게 블러 처리가 된 것을 확인할 수 있다. 또한 격자 무늬 안의 울퉁불퉁한 모양도 소실되었다. 두번째로 *chain_texture*에서는 모양 왜곡 현상이 나타났다. 원본 이미지는 체인의 경계가 매끄럽지 않으나, BSRGAN은 업스케일링 후 강제로 체인의 경계를 부드럽게 만든 것을 확인할 수 있다.

세번째 아티팩트는 노이즈 현상이다. *Mountains*를 Real-ESRGAN으로 업스케일링 후 다운스케일링 한 결과, 원본 이미지

에서 존재하지 않던 여러가지 색상의 노이즈가 추가되었다. 네번째 아티팩트는 색상 왜곡이다. 다운스케일링한 *kodim01*을 Real-ESRGAN을 업스케일링한 결과, 원본 이미지의 색상과 다르게 노란 색감이 줄어들고 밝기도 밝아졌다. Real-ESRGAN 뿐만 아니라 BSRGAN, SwinIR-real 역시 자주 원본 이미지의 색감을 변형하는 경향을 지닌다. 다섯번째 때때로 지나치게 매끄러워지는(over-smoothing) 현상이 나타난다. *sponza_fabric_green_diff*를 Real-ESRGAN으로 업스케일링한 결과, 원본 이미지에 있던 천의 직조된 질감이 색종이처럼 매끄러운 질감으로 바뀌어 표현되었다. 여섯번째 아티팩트는 샤프닝(sharpening) 현상이다. *sponza_surtain_diff*는 SwinIR-real의 업스케일링 이후 금색 무늬 및 빨간색 천 모두 질감이 과도하게 선명해졌으며, 전체적으로 채도 역시 증가하였다. 마지막 아티팩트는 에일리어싱이다. *vetor-streets*를 SwinIR-classical로 다운스케일링 후 업스케일링한 결과, 선과 글자 부분에서 에일리어싱 현상이 두드러지게 나타났다.

그림 6은 각 카테고리별로 주요 아티팩트가 발생하는 대표 이미지를 1개씩 뽑아 비교한 결과이다. 전반적으로 SwinIR-classical을 제외한 초해상도 기법들은 색상을 왜곡시키는 경향을 보이지만, 각기 다른 양상의 아티팩트를 발생시키기도 한다.

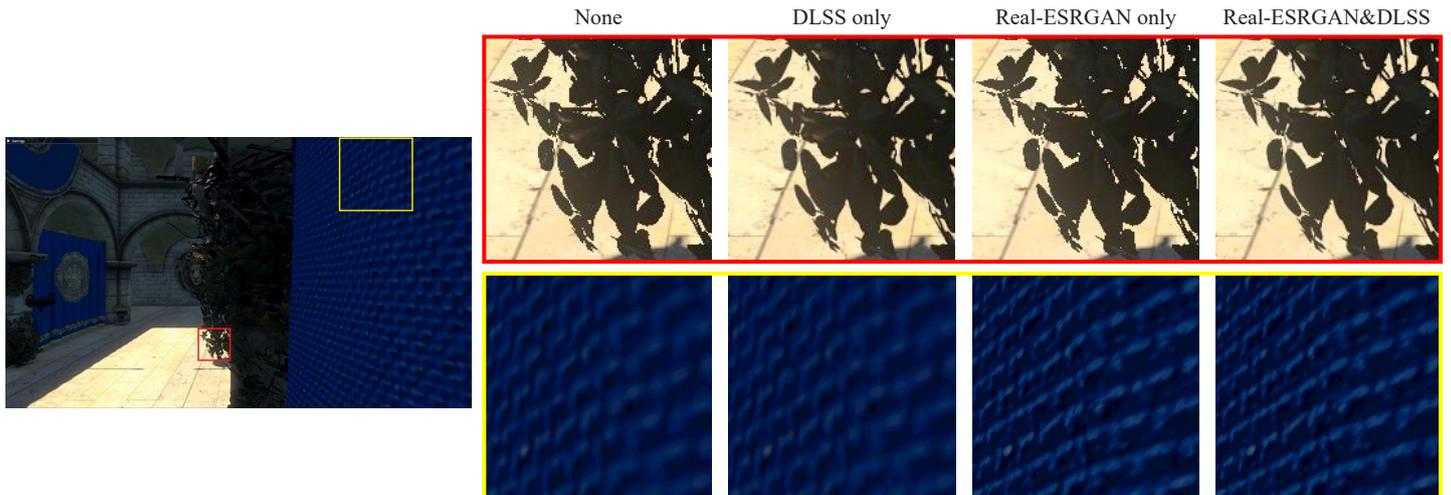


Figure 7: Comparison of results applying texture super-resolution and DLSS to the Sponza scene. As depicted in the figure, the combination of Real-ESRGAN and DLSS provided the clearest image quality.

이러한 아티팩트를 각 기법별로 분석해 보면 다음과 같다. BSRGAN에서는 원본 이미지의 해상도가 낮을 때 블러링 현상 및 모양 왜곡을 주로 나타낸다. 하지만 노멀 맵 같은 경우 해상도가 높아도 모양을 왜곡시키기도 하고, 원본 이미지가 매우 부드러운 계조를 띠는 경우에는 이를 날카롭게 만드는 샤프닝을 발생시키기도 한다. Real-ESRGAN에서는 이러한 모양 왜곡과 색상 왜곡이 대체로 BSRGAN보다 좀 더 심하게 나타나지만, 노멀 맵이나 저해상도의 지도 등 종종 원본 이미지의 디테일을 훨씬 더 잘 보존하는 경우도 존재한다. SwinIR-real은 대체로 샤프닝과 대비가 타 기법에 비해 커지는 모습을 보였으며, 그 덕분에 FLIP 수치와는 별개로 저해상도의 사진이나 게임 텍스처의 업스케일링시 타 기법들보다 선명한 결과를 도출한다. 반면 노멀 맵이나 저해상도의 지도 같은 경우 디테일을 과도하게 뭉개는 블러링을 보이기도 한다. SwinIR-classical은 원본이미지에 다운스케일링을 먼저 적용하여 낮은 해상도로 만든 경우, 대체로 지나치게 초해상도 이미지를 흐릿하게 만드는 경향이 강하며, 앞서 언급한 것처럼 에일리어싱을 만들어 내기도 한다. 하지만 원본이미지의 해상도가 충분한 경우 (SwinIR-classical을 먼저 적용한 후 다운스케일링한 경우) 이러한 문제들은 대부분 훨씬 완화되었다.

앞에서 서술한 카테고리 별 정량적 분석 결과와 마찬가지로, SwinIR-classical을 제외하면 노멀 맵에 초해상도 기법을 적용할 경우 다른 카테고리의 텍스처 맵보다 훨씬 더 다양한 아티팩트 현상이 두드러지게 나타났다. 초해상도 기법을 적용한 몇몇의 노멀 맵의 경우, 실제 텍스처 맵으로 사용하기 어려울 정도로 왜곡 현상이 일어나는 경우도 존재하였다.

3.2.3 실제 렌더링 결과

마지막으로, 이미지 초해상도 기법을 실제 렌더링 시에 사용하였을 경우에 어떠한 효과를 나타내는지 살펴보고자 한다. 이 때 DLSS[10]를 함께 적용했을 때의 효과를 살펴보았는데, 그 이유

는 이 기법은 프레임의 해상도를 높이기 위해 널리 사용되고 있기 때문이다. 그림 7은 Sponza scene의 결과를 확대한 것으로, 이 장면 내에 포함된 텍스처의 해상도를 높이기 위해 Real-ESRGAN을 선택하였다. 그 이유는 BSRGAN은 노멀 맵에서 다소 왜곡이 심하고 SwinIR은 RGBA 채널을 가진 입력을 RGB 형태로 출력함으로 인해 반투명 텍스처에 적용 불가하기 때문이다.

그림 7의 첫번째 행은 DLSS가 효과적인 부분을 보여준다. DLSS는 샘플로 사용되는 픽셀의 개수를 늘려준다. 이를 통해, 이 기법은 낮은 해상도로 인해 나뭇잎의 가장자리에 보이는 계단 현상을 상당 부분 완화시켜 주었다. 반면 이 경우에는 텍스처의 해상도 증가가 화질에 큰 영향을 끼치지 않았는데, 그 이유는 맵 사용시 이미 충분한 해상도의 텍스처 이미지가 선택되었기 때문으로 보인다.

반면, 두번째 행은 텍스처 해상도의 증가가 효과적인 부분을 보여준다. 파란 커튼은 시점 가까이에서 매우 크게 보이는 물체이기 때문에, 레벨 0의 가장 큰 텍스처 이미지가 필터링되어 흐릿하게 보여진다. DLSS가 이러한 커튼 내부의 선명도를 높이지 못한 반면, Real-ESRGAN으로 텍스처의 해상도를 높였을 경우 이 부분의 선명도가 눈에 띄게 높아졌다. 이처럼 DLSS와 텍스처 초해상도 기법은 각기 다른 영역(각각 프레임버퍼 및 텍스처)의 해상도를 높이기 때문에, 이 두 기법을 함께 사용할 경우 (가장 오른쪽 열) 근거리, 원거리 물체 모두에 대해 효과적인 것을 확인 가능하다.

4. 결론 및 향후 연구

본 논문에서는 최신 딥러닝 기반의 초해상도 기법들 중 BSRGAN, Real-ESRGAN, SwinIR을 다양한 형태의 텍스처의 해상도를 높이는 데 적용하는 실험을 수행하였다. 실험 결과를 정량적 및 정성적으로 분석한 결과, 텍스처 맵의 형태 및 크기에 따라 이 방법들은 혼재된 결과를 나타내었으며, 업스케일링 후의 아티

팩트 역시 서로 다른 형태로 나타났다. 다만 네 방법 중 별도의 열화 적용을 하지 않는 SwinIR-classical이 정량, 정성적으로 다른 기법들 대비 우위를 보여 줬다. 그 이유는 텍스처의 경우 일반 스마트폰으로 찍은 사진에서 보여지는 흔들림, 광량 부족, 작은 센서 크기, 과도한 JPEG 압축 등과 같은 품질 저하 요소가 적용되지 않은 반면, 다른 기법들에서는 이러한 품질 저하를 복구하기 위해 사용한 열화 기반 훈련이 오히려 추론시 부정적인 결과를 만들어내기 때문으로 보인다. 하지만 SwinIR-classical도 다운스케일링을 먼저 하여 원본 이미지의 해상도를 낮춘 경우에는 원본의 디테일을 크게 훼손하는 경우를 발생시켰다. 또한 SwinIR은 RGBA 이미지의 alpha 채널을 지원하지 않는다. 이러한 한계점들은 기존 이미지 초해상도 기법들이 텍스처 맵이 아닌 일상 사진의 복원을 목표로 연구되었기 때문에 자연스럽게 나타난 결과로 보인다.

이러한 이유로 고해상도 텍스처의 자동 생성에 딥러닝 기반의 초해상도 기법을 적용하기 위해서는 향후 다음과 같은 측면에서 연구가 계속 진행되어야 할 필요가 있다. 먼저, 네트워크의 훈련에 사용 가능한 대규모 텍스처 셋이 필요하다. 본 논문에서 사용한 텍스처 셋은 훈련에 사용하기에는 개수가 적고, 저해상도와 고해상도 이미지가 혼재되어 있다. 그러므로 실제 앱들에 내장되는 수천 수만개의 고해상도 이미지들을 밍맵 형태로 구성하고, 이를 네트워크 훈련에 사용한다면 추론 결과의 품질이 훨씬 더 높아질 것이다. 다음으로, 텍스처 종류에 따라 다르게 추론이 가능한 네트워크도 설계 가능할 것이다. 텍스처는 디자이너가 직접 그리기도 하지만 사진 편집을 통해 생성되거나 프로시저 텍스처(procedural texture)와 같이 수학적 알고리즘을 통해서도 만들어지는 등 다양한 방법으로 생성 가능하다. 또한 디퓨즈 맵, 노멀 맵, 스펙큘러 맵(specular map), 반사 맵 등과 같이 텍스처 내에는 완전히 다른 형태의 정보가 저장될 수도 있다. 이러한 특성을 설계시에 반영한 네트워크는 일부 텍스처 형식에서 업스케일링시 나타나는 아티팩트를 크게 줄일 수 있을 것이다. 마지막으로, 이러한 아티팩트는 TexSR[5]과 같이 후처리를 통해 어느 정도 제거 가능하므로, 추론 후 수행 가능한 후처리 기법들에 대해서도 다양한 연구가 진행 가능할 것이라고 생각한다.

또한, 본 논문에서 소개한 최신 단일 이미지 초해상도(single-image super-resolution) 방식들은 이미지당 수~수십 초의 추론 시간을 필요로 한다. 본 논문에서 실험한 바와 같이 오프라인 상에서 텍스처의 해상도를 높인 후 실시간 렌더링 시에 이를 적용할 때에는 이러한 긴 추론 시간이 큰 문제가 되지 않는다. 하지만 앱 로딩시마다 앱 내에 저장된 텍스처의 해상도를 GPU 드라이버 단계에서 실시간으로 조정하려 하거나[32], render-to-texture 기법들을 통해 만들어지는 텍스처(deferred rendering을 위한 G-buffer 등)의 해상도를 높여야 할 때에는, BSRGAN, Real-ESRGAN, SwinIR과 같은 방법들은 느린 속도로 인해 사용이 불가능하다. 따라서 실시간 초해상화를 위해서는 어느 정도의 품질을 희생하더라도 속도가 수천~수만 배 빠른 새로운 해법이 요구되며, 이러한 실시간 처리를 위해서는 알고리즘의 경량화 뿐

아니라 TV의 화질 향상 칩[33]에 구현된 것과 같은 별도의 하드웨어 가속 구조가 GPU 내에 추가될 필요가 있다. 또한 학습된 필터링을 추가로 사용하는 새로운 텍스처 압축 코덱[34]을 확장하여, 압축된 텍스처가 입력 텍스처보다 더 높은 해상도를 갖도록 하는 방법도 고려해 볼 수 있을 것이다.

마지막으로 본 논문에서는 가로, 세로 길이를 2배로 늘려 업스케일링한 결과들에 대해서만 분석을 수행하였다. 이를 4배 또는 8배까지 늘려 밍맵 레벨을 2~3단계 추가할 경우, 어떠한 초해상도 알고리즘을 사용하는 것이 더 나은지에 대해서는 더 많은 실험과 분석이 필요하다. 그리고 이와 같이 4배~8배 업스케일링을 수행할 경우 텍스처의 크기는 각각 16배 및 64배로 늘어나게 된다. 그러므로 이로 인해 생기는 메모리 및 스토리지의 오버헤드를 효과적으로 처리하기 위해서는 많은 고민이 필요할 것이다. 보다 높은 압축률의 새로운 코덱을 사용하거나[34], 기존 ASTC와 같은 압축 코덱 사용시 적응적으로 블록 크기를 변경[17, 35]하는 방법을 밍맵 레벨별로 다르게 적용하는 것이 해결 방안이 될 수 있을 것이라 생각하며, 이는 향후 연구로 남겨 두고자 한다.

감사의 글

본 연구는 2023학년도 상명대학교 교내연구비를 지원받아 수행하였음 (2023-A000-0318). 실험 데이터는 Kodak, Crytek, UNC GAMMA Lab, Spiral Graphics, Ignacio Castaño가 공개한 이미지 및 장면을 포함함.

References

- [1] J. Papadopoulos. (2023) New Cyberpunk 2077 HD texture packs improve the textures of NPCs faces, hair & beards. [Online]. Available: <https://www.dsogaming.com/mods/new-cyberpunk-2077-hd-texture-packs-improve-the-textures-of-npcs-faces-hair-beards/>
- [2] GPUUnity. (2023) Arkham city redux. [Online]. Available: <https://www.nexusmods.com/batmanarkhamcity/mods/407/>
- [3] L. Williams, "Pyramidal parametrics," in *Proceedings of the 10th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 1983, pp. 1–11.
- [4] S. Takemura, "Optimize deep super-resolution and denoising for compressed textures," in *ACM SIGGRAPH Asia 2018 Posters*, 2018.
- [5] J.-H. Nah and H. Kim, "TexSR: Image super-resolution for high-quality texture mapping," in *ACM SIGGRAPH Asia 2022 Posters*, 2022.

- [6] K. Zhang, J. Liang, L. V. Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021, pp. 4791–4800.
- [7] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1905–1914.
- [8] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1833–1844.
- [9] P. Andersson, J. Nilsson, T. Akenine-Möller, M. Oskarsson, K. Åström, and M. D. Fairchild, "FLIP: A difference evaluator for alternating images," *Proceedings of the ACM on Computer Graphics and Interactive Techniques (HPG)*, vol. 3, no. 2, pp. 1–23, 2020.
- [10] NVIDIA Developer. (2023) Deep learning super sampling (DLSS). [Online]. Available: <https://www.nvidia.com/en-us/geforce/technologies/dlss/>
- [11] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," in *Proceedings of European Conference on Computer Vision (ECCV) 2014*, pp. 184–199, 2014.
- [12] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.
- [13] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Computer Vision – ECCV 2018 Workshops*, 2019, pp. 63–79.
- [14] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang, "Real-world super-resolution via kernel estimation and noise injection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 1914–1923.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.
- [16] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9992–10 002.
- [17] J.-H. Nah, "ASTC block-size determination method based on PSNR values," *Journal of the Korea Computer Graphics Society*, vol. 28, no. 2, pp. 21–28, 2022.
- [18] W. Cho, J. Kwon, S. Kwon, and J. Yoo, "A comparative study on OCR using super-resolution for small fonts," *The International Journal of Advanced Smart Convergence*, vol. 8, no. 3, pp. 95–101, 2019.
- [19] S.-H. Sung, K.-B. Lee, and S.-H. Park, "Research on Korea text recognition in images using deep learning," *Journal of the Korea Convergence Society*, vol. 11, no. 6, pp. 1–6, 2020.
- [20] Y. Li, B. Sixou, and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems," *IRBM*, vol. 42, no. 2, pp. 120–133, 2021.
- [21] S. Jia, Z. Wang, Q. Li, X. Jia, and M. Xu, "Multiattention generative adversarial network for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [22] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [23] J.-H. Nah, "QuickETC2: How to finish ETC2 compression within 1 ms," in *ACM SIGGRAPH 2020 Talks*, 2020.
- [24] J.-H. Nah, "QuickETC2: Fast ETC2 texture compression using luma differences," *ACM Transactions on Graphics (SIGGRAPH Asia 2020)*, vol. 39, no. 6, p. 10, nov 2020.
- [25] J. M. P. van Waveren and I. Castaño, "Real-time normal map DXT compression," *Technical Report, Id Software*, 2008. [Online]. Available: <https://developer.download.nvidia.com/whitepapers/2008/real-time-normal-map-dxt-compression.pdf>
- [26] I. Castaño. (2021) Image datasets - a repository of public image data sets. [Online]. Available: <https://github.com/castano/image-datasets>
- [27] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *2017*

IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1122–1131.

- [28] A. Gianluca. (2023) QualityScaler - image/video AI upscaler app (BSRGAN). [Online]. Available: <https://github.com/Djdefrag/QualityScaler/>
- [29] S. Yuk and Y. Cho, “A study on novel steganography communication technique based on thumbnail images in sns messenger environment,” *Journal of Internet Computing and Services*, vol. 22, no. 6, pp. 151–162, 2021.
- [30] J. Nilsson and T. Akenine-Möller, “Understanding SSIM,” *ArXiv e-prints*, 2020, arXiv:2006.13846.
- [31] NVIDIA Developer. (2023) DLSS - download and get started. [Online]. Available: <https://developer.nvidia.com/rtx/dlss/get-started>
- [32] J.-H. Nah, B. Choi, and Y. Lim, “Classified texture resizing for mobile devices,” in *ACM SIGGRAPH 2018 Talks*, New York, NY, USA, 2018. [Online]. Available: <https://doi.org/10.1145/3214745.3214763>
- [33] LG Electronics. (2022) α 9 Gen5 AI Processor. [Online]. Available: https://www.lg.com/levant_en/tvs/2022/alpha9
- [34] K. Vaidyanathan, M. Salvi, B. Wronski, T. Akenine-Moller, P. Ebelin, and A. Lefohn, “Random-access neural compression of material textures,” *ACM Transaction on Graphics*, vol. 42, no. 4, jul 2023. [Online]. Available: <https://doi.org/10.1145/3592407>
- [35] J.-H. Nah, “Addition of an adaptive block-size determination feature to astcenc, the reference ASTC encoder,” *Software Impacts*, vol. 17, 2023.

< 저자 소개 >



김혜주

- 2023 상명대학교 식품영양학전공/컴퓨터과학 전공/외식경영학연계전공 학사
- 2023~현재 상명대학교 컴퓨터과학 석사과정
- <https://orcid.org/0000-0003-1934-9053>



나재호

- 2005 연세대학교 컴퓨터산업공학부 학사
- 2007 연세대학교 컴퓨터과학과 석사
- 2012 연세대학교 컴퓨터과학과 박사
- 2012 세종대학교 컴퓨터공학과 방문연구원
- 2012~2013 University of North Carolina at Chapel Hill Visiting Researcher
- 2014~2021 LG전자 선임연구원, 책임연구원
- 2021~현재 상명대학교 컴퓨터과학과 조교수
- 관심분야: 광선 추적법, 렌더링 알고리즘, GPU 구조 등
- <https://orcid.org/0000-0001-7805-5333>