

# 다중 스케일 이산 웨이블릿 변환을 통한 잠재 확산 모델 기반 초해상화의 세부 품질 향상

(Enhancing Detail Quality in Latent Diffusion Model-Based  
Super-Resolution via Multi-Scale Discrete Wavelet  
Transform)

서 준 혁 <sup>†</sup>      이 동 규 <sup>‡‡</sup>

(Jun-Hyeok Seo)    (Dong-Gyu Lee)

**요약** 본 논문은 잠재 확산 모델 기반 초해상화 기법의 세부 품질을 향상하기 위한 다중 스케일에서 이산 웨이블릿 변환 프레임워크를 제안한다. 기존의 잠재 확산 모델 기반 초해상화 기법들은 저해상도 영상의 초해상화 과정에서 영상에 존재하는 미세한 윤곽이나 질감과 같은 세부 정보를 잘 보존하지 못한다는 한계가 존재했다. 본 논문에서는 모델이 추론한 초해상화 영상과 실제 고해상도 영상을 다양한 스케일로 증강하고 두 영상 사이의 오차를 이산 웨이블릿 변환을 통해 고주파 영역에서 추출함으로써 모델이 영상의 세부 정보를 더 정확하게 보존할 수 있도록 한다. 제안하는 프레임워크는 벤치마크 데이터셋을 통한 기존 기법들과의 성능 비교에서 우수한 성능을 보이며 초해상화 영상의 세부 품질을 향상하는데 효과적임을 보여준다.

**키워드:** 컴퓨터비전, 초해상화, 확산모델, 잠재확산모델, 다중스케일, 이산웨이블릿변환

**Abstract** This paper proposes a multi-scale discrete wavelet transform framework designed to enhance the detail quality of super-resolution (SR) methods based on latent diffusion models. Existing SR approaches that utilize latent diffusion models have struggled to preserve fine details, such as subtle contours and textures, during the super-resolution process of low-resolution images. In this study, both the super-resolved images generated by the model and the corresponding high-resolution images are augmented across various scales. The errors between them are then extracted in the high-frequency domain using discrete wavelet transform. This approach enables the model to retain detailed information more accurately. Experimental results on benchmark datasets demonstrate that the proposed framework outperforms previous methods, effectively improving the detail quality of super-resolved images.

**Keywords:** computer vision, super resolution, diffusion model, latent diffusion model, multi scale, discrete wavelet transform

· 이 성과는 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학 ICT연구센터의 지원(IITP-2025-RS-2020-II201808, 50%)과 정부(과학기술 정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2025-02214941, 50%).

· 이 논문은 2024 한국소프트웨어종합학술대회에서 '다중 스케일 고주파 오류 맵을 활용한 확산 모델 기반 이미지 초해상화'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 경북대학교 인공지능학과 석박통합과정 학생  
ssam2s@knu.ac.kr

‡‡ 정회원 : 경북대학교 인공지능학과 교수(Kyungpook Nat'l Univ.)  
dglee@knu.ac.kr  
(Corresponding author)

논문접수 : 2025년 4월 17일

(Received 17 April 2025)

논문수정 : 2025년 6월 5일

(Revised 5 June 2025)

심사완료 : 2025년 7월 11일

(Accepted 11 July 2025)

## 1. 서 론

영상 초해상화(Super-Resolution, SR)는 저해상도의 영상을 고해상도의 영상으로 변환함으로써 영상 복원 및 품질 개선 등 다양한 목적[1]을 위해 연구되어 왔다. 기존에는 합성곱 신경망(Convolutional Neural Networks, CNN)에 기반한 접근[2]이 주를 이루었으며, 이후 적대적 생성 신경망(Generative Adversarial Networks, GAN)을 도입하여 생성자와 판별자가 경쟁적으로 사설적인 영상을 만들어내는 연구들이 큰 주목을 받았다. 더 나아가 트랜스포머(Transformer)를 기반의 ViT(Vision Transformer)[3]가 등장하면서, 영상을 패치 단위로 나누어 전체 맵력을 이해해 자연스러운 영상을 생성하는 방식[4]이 괄목한만한 성과를 거두었다.

최근에는 영상 생성 분야에서 확산 모델(Diffusion model)이 뛰어난 성능을 보임에 따라, 이를 SR에도 적용하려는 시도가 활발히 이어지고 있다[5]. 확산 모델은 영상을 연속적으로 노이즈화·제거하는 과정을 반복 학습하면서, 세밀한 윤곽과 질감 정보를 정교하게 복원한다는 장점을 지닌다. 그러나 영상 전체에 대해 반복적인 확산 과정을 거치며 처리해야 하므로, 학습 및 추론 과정에서 상대적으로 많은 연산 자원이 요구된다는 한계가 있다. 이를 보완하기 위해 제안된 잠재 확산 모델(Latent Diffusion Model, LDM)[6]은 영상을 잠재 공간(Latent space)으로 압축한 뒤 확산 과정을 수행함으로써, 기존 확산 모델 수준의 성능을 유지하면서도 연산 복잡도를 줄였다. 최근에는 이러한 LDM 기반 구조를 활용한 SR 연구도 활발히 진행되며 다양한 접근이 제안되고 있다[7,8]. 다만 LDM은 추론 과정에서 영상에 존재하는 윤곽선이나 질감과 같은 고주파 성분이 손실되며, 역확산 단계에서 모델이 학습 데이터의 전반적 분포에 의존해 세부 정보를 재구성할 때 원본과 다르게 표현될 수 있다. 그림 1의 빨간 박스는 LDM이 왜곡된 세부 정보를 인위적으로 생성하는 문제의 예시를 나타낸다. 이처럼 잠재 공간은 영상 정보를 추상화한 형태로 다루기 때문에 미세한 질감이나 윤곽보다는 전체적 구조나 색감이 우선 보존되어 왜곡 문제가 발생할 수 있다.

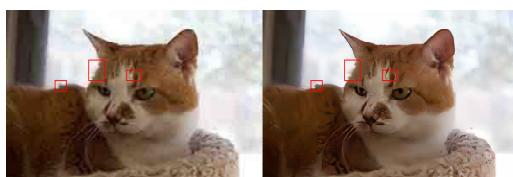


그림 1 입력 영상(좌) 및 LDM의 추론 영상(우)  
Fig. 1 Input image (left) and result from LDM (right)

본 논문에서는 기존 LDM을 활용한 연구들에서 발생하는 왜곡 현상을 최소화하면서 연산 효율성을 유지할 수 있는 새로운 프레임워크를 제안한다. 제안하는 프레임워크는 이산 웨이블릿 변환(Discrete Wavelet Transform, DWT)을 활용하여 원본 영상에 대한 고주파 성분을 다양한 스케일에서 추출하고 각 성분에 대해 상대적 가중치를 부여한다. 이를 통해 LDM이 가지는 경량화 이점 을 유지하면서 영상의 세부 정보를 더 정확하게 복원하도록 유도하며, 세 가지 벤치마크 데이터셋에 대한 성능 비교 실험을 통해 제안하는 방법의 효율성을 증명한다.

## 2. 본 론

본 논문에서는 영상에 존재하는 고주파 성분을 효과적으로 보존할 수 있는 새로운 프레임워크를 제안한다. 기본적으로 LDM을 그대로 활용하면서, 그림 2와 같이 DWT를 활용해 영상의 세부 정보를 포함한 고주파 성분을 보다 효과적으로 복원할 수 있도록 설계되었다.

DWT는 2차원 영상을 크게 네 가지 대역으로 분해한다. LL(Low-Low) 대역은 가로, 세로 방향이 모두 낮은 주파수를 가진 저주파 성분을, LH(Low-High)와 HL(High-Low) 대역은 가로, 세로 방향 중 하나만 고주파인 중간 주파수 성분을, 그리고 HH(High-High) 대역은 가로, 세로 방향 모두 고주파인 성분을 포착한다. 여기서 고주파 성분을 포함하고 있는 대역은 LH, HL, HH로, 본 프레임워크는 해당하는 세 가지 성분을 활용한다.

그림 2는 제안하는 프레임워크의 구체적인 흐름을 보여준다. 먼저 저해상도(Low Resolution, LR) 영상과 고해상도(High Resolution, HR) 영상을 함께 입력받아, 모델이 생성한 SR 영상과 HR 영상 간의 손실을 학습 과정에서 최소화한다. 이때 단순 픽셀 단위의 차이만 고려하는 것이 아니라, 다양한 크기로 증강된 LR 및 HR 영상에 대해 주파수 대역별로 DWT를 수행함으로써 세부 정보 복원에 중요한 고주파 성분을 별도로 추출한다. 추출된 고주파 영역 간의 차이를 오류 맵으로 산출하고,

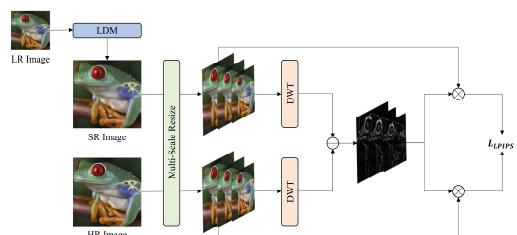


그림 2 제안하는 프레임워크의 개요  
Fig. 2 Overview of the proposed framework

해당 스케일의 SR 및 HR 영상에 곱해주어, 추론 과정에서 세부 정보가 제대로 재현되지 못한 영역을 강조한다.

고주파 성분이 강조된 HR 영상과 SR 영상을 활용하여 LPIPS(Learned Perceptual Image Patch Similarity) [9] 손실 함수를 계산한다. LPIPS는 대규모 영상 데이터로 학습된 신경망의 중간 특징 공간이 사람의 인지 특성과 유사하다는데에서 착안한 평가 지표로, 단순한 픽셀 단위의 차이뿐 아니라 사람의 인지 과정에서 느껴지는 왜곡을 더 잘 반영하기 때문에 자연스러운 SR 결과를 유도하면서도 고주파 성분의 왜곡을 최소화할 수 있다. 또한 본 프레임워크에서는 LPIPS 손실을 단일 스케일이 아닌 다중 스케일에서 계산함으로써 각 수준에서 나타나는 세부 정보들을 더 정교하게 반영하도록 설계하였다.

본 논문에서는 4가지 스케일로 증강된 영상에서 LPIPS를 각각 계산한 뒤, 큰 스케일에서 얻어진 값일수록 더 많은 세부 정보를 반영한다고 가정하여 가중치를 그에 비례하게 부여한다. 큰 스케일의 영상은 비교적 높은 해상도로 구성되어 있어 미세한 질감, 에지와 같은 세부 정보를 더 많이 포함하므로 여기서 계산된 손실 값이 더 반영되도록 가중치를 높게 설정했다. 이에 반해 작은 스케일에서는 영상의 전반적인 구조가 강조되는 반면 세부 정보는 사라진다. 그럼에도 작은 스케일에서의 손실을 동시에 고려하여 전역적인 구조 및 거시적 특성을 함께 학습할 수 있어 전체적인 일관성을 확보할 수 있다.

#### 알고리즘 1 전체 손실 함수 계산 알고리즘

Algorithm 1 An algorithm to compute total loss function

<b>Algorithm 1</b> Computing Multi-Scale LPIPS Loss	
1:	<b>Input:</b> $x_i, \hat{x}_i, S_i, W_i$
2:	<b>Output:</b> $L_{total}$
3:	Initialize $L_{total}$ to zero.
4:	for $i = 1$ to $4$ do
5:	Resize $x_i, \hat{x}_i$ to scale $S_i$ .
6:	Get $LH, HL, HH$ with DWT from $x_i$ .
7:	Get $\widehat{LH}, \widehat{HL}, \widehat{HH}$ with DWT from $\hat{x}_i$ .
8:	High Frequency Map $HF \leftarrow LH + HL + HH$
9:	High Frequency Map $\widehat{HF} \leftarrow \widehat{LH} + \widehat{HL} + \widehat{HH}$
10:	Error Map $E \leftarrow  HF - \widehat{HF} $
11:	$x_i \leftarrow E \cdot x_i$
12:	$\hat{x}_i \leftarrow E \cdot \hat{x}_i$
13:	Compute $L_{LPIPS}$ with $x_i, \hat{x}_i$ .
13:	$L_{total} \leftarrow L_{total} + (W_i \cdot L_{LPIPS})$
14:	<b>end for</b>
15:	<b>Output:</b> The summation of $L_{LPIPS}$ from each scale.

알고리즘 1은 본 프레임워크에서 제안하는 최종 손실 함수 계산 과정을 나타낸다. 전체 스케일에 대한 손실 함수는 각 스케일별 LPIPS 손실에 설정된 가중치의 가중합으로 계산된다. 이는 높은 스케일에서 낮은 스케일 까지 다양한 수준의 정보를 균형 있게 반영하여, 인간의 시각적 인지와 유사한 왜곡 문제를 해소함과 동시에 영상의 전체적인 구조까지 효과적으로 고려한다. 본 논문에서는 스케일  $S_i \in \{1, 0.5, 0.25, 0.125\}$ 에 대응되는 가중치  $W_i \in \{0.4, 0.3, 0.2, 0.1\}$ 를 설정하였다. 각 스케일별 LPIPS 손실은 모두 더해져 최종 손실 함수  $L_{total}$ 로 도출된다. 최종 손실 함수는 수식 (1)로 정의된다.

$$L_{total} = \sum_{i=1}^4 W_i \cdot L_{LPIPS}(S_i \cdot x_i, S_i \cdot \hat{x}_i). \quad (1)$$

### 3. 실험 결과

본 논문에서 제안한 프레임워크를 객관적으로 평가하기 위해, CNN, GAN, Transformer, LDM 기반의 기존 연구들과 세 가지 벤치마크 데이터셋에 대한 비교 실험을 수행하였다. 기존 연구와의 공정한 비교를 위해 SinSR[7]을 베이스라인으로 선정하여, 학습 데이터셋과 검증 데이터셋을 동일하게 설정하여 성능을 검증하였다. 실험에 사용된 ImageNet-Test는 ImageNet의 검증 데이터에서 얻은 3,000개의 영상으로 이루어져 있으며 RealSR과 RealSet65는 각각 2가지의 서로 다른 카메라로 촬영한 100개의 영상, 온라인에서 수집한 65개의 영상으로 이루어져 있다.

최근 연구들[10,11]에서, 픽셀 수준의 평가 지표인 참조 기반 방식의 지표를 사용하면 지표상으로 더 좋은 성능을 보이더라도 실제 사람의 시각에서 인지 품질이 낮아질 수 있음을 보이며, 인지 수준의 평가 지표인 비참조 기반 방식의 지표들이 도입되고 있다. 따라서 본 논문에서는 비참조 기반 평가 지표로 CLIP 기반의 영상 품질 측정 방식인 CLIPQA[12], 그리고 Multi-Scale Transformer 기반의 방식인 MUSIQ[13]을 사용하였다. 이 두 지표는 사람의 시각적 인지 또는 학습된 분산 표현을 기반으로 영상의 품질을 측정한다. 구체적으로 CLIPQA는 사전 학습된 CLIP 모델을 활용해 추출된 고차원 특징 벡터 간의 유사도를 품질 척도로 삼아 시각적으로 사실적인 영상일수록 높은 점수를 부여한다. MUSIQ는 다중 스케일로 영상의 특징 벡터를 추출해 전역적, 국소적 특성을 고루 평가하여 더 좋은 품질의 영상에 높은 점수를 부여한다.

표 1은 벤치마크 데이터셋에 대한 기존 연구들과의 정량적 성능 비교 결과를 보여준다. 본 논문에서 제안하는 프레임워크는 기존 연구 대비 전반적으로 높은 성능

표 1 세 가지 벤치마크 데이터셋에 대한 성능 비교  
Table 1 Performance comparison on three benchmark datasets

Methods	Architectures	ImageNet-Test		RealSR		Realset65	
		CLIPQA	MUSIQ	CLIPQA	MUSIQ	CLIPQA	MUSIQ
SwinIR [14]	Transformer	0.564	53.790	0.465	59.636	0.578	<b>63.822</b>
RealESRGAN [15]	GAN	0.523	52.538	0.489	59.678	0.599	<u>63.220</u>
DASR [16]	CNN	0.536	48.337	0.362	45.825	0.496	55.708
LDM [6]	LDM	0.572	50.895	0.383	49.317	0.427	47.488
ResShift [8]	LDM	0.603	<u>53.897</u>	0.595	59.873	0.653	61.330
SinSR [7]	LDM	<u>0.611</u>	53.357	<u>0.688</u>	<b>61.582</b>	<u>0.715</u>	62.169
<b>Ours</b>	LDM	<b>0.648</b>	<b>54.212</b>	<b>0.695</b>	<u>61.465</u>	<b>0.718</b>	63.216

을 보였으며 특히 CLIPQA에 대해 모든 데이터셋에서 가장 높은 성능을 보였다. 다만 RealSet65 데이터셋에서 MUSIQ가 다소 낮은 성능을 보였다. 이는 제안하는 방법이 사용하는 LDM의 성능에 영향을 크게 받기 때문이다.

본 논문에서는 정확한 고주파 성분 추출을 위해 DWT를 활용한 프레임워크를 제안하였다. 제안 기법의 타당성을 더 명확히 입증하기 위해, DWT와 유사한 주파수 성분 추출 알고리즘인 고속 푸리에 변환(Fast Fourier Transform, FFT)의 성능을 확인하는 비교 실험을 진행하였다. 비교 실험에서는 프레임워크의 모든 구성 요소를 동일하게 설정하고 고주파 성분 추출 알고리즘만을 대체하여 세 가지 벤치마크 데이터셋 중 하나인 ImageNet-Test에 대한 성능 평가를 진행하였다.

고주파 성분 추출 방식을 변경했을 때의 성능 비교 결과는 표 2에서 보여준다. FFT를 고주파 성분 추출에 사용했을 때 제안하는 방식보다 전반적인 성능이 하락하는 모습을 보였다. 이는 각 주파수 대역별로 가변적인 윈도우를 사용하는 DWT가 전체 신호에 대한 주파수 스펙트럼을 구하는 FFT보다 더 정밀한 고주파 성분 추출이 가능함을 보여준다. 또한 DWT의 적용 과정에서 추가적인 윤곽 정보를 제공해 더 정확한 고주파 성분을 추출할 수 있는지에 대한 검증을 위해 대표적인 윤곽

검출 방법인 Canny Edge를 각 스케일에서 추가하여 실험을 진행하였다. 해당 실험에서 CLIPQA는 0.35, MUSIQ는 1.127이 하락하며 성능이 떨어지는 모습을 보였다. 이는 픽셀 수준에서 추출한 고주파 성분이 주파수 수준에서 DWT를 통해 얻은 고주파 성분에 상호 간섭을 발생시키는 잡음에 가까운 중복 정보가 될 수 있다는 것을 나타낸다.

표 3은 ImageNet-Test 데이터셋에 대해 참조 기반 지표를 사용한 기존 연구들과의 정량적 비교 결과를 보여준다. 본 논문에서 제안하는 프레임워크는 참조 기반 지표에서 전반적으로 낮은 성능을 보였다. 하지만 베이스라인과의 결과 영상 비교를 통한 정성적 평가를 그림 3과 같이 수행하였을 때, 실제 인지 측면에서 느껴지는 품질에서는 시각적으로 더 우수함을 보였다. 이를 통해 본 논문에서 제안하는 프레임워크가 영상에 존재하는 세부 정보와 질감을 더 선명하고 풍부하게 표현하는 것을 확인할 수 있었다.

표 3 ImageNet-Test에 대한 참조 기반 지표의 성능 비교  
Table 3 Performance comparison on ImageNet-Test using a reference-based metric

Methods	ImageNet-Test		
	PSNR	SSIM	LPIPS
SwinIR [11]	23.99	0.667	0.238
RealESRGAN [12]	24.04	0.665	0.254
DASR [13]	<u>24.75</u>	<b>0.675</b>	0.250
LDM [6]	24.49	0.651	0.248
ResShift [14]	<b>24.90</b>	<u>0.673</u>	0.228
SinSR [8]	24.56	0.657	<b>0.221</b>
<b>Ours</b>	24.13	0.641	<u>0.224</u>

표 2 고주파 성분 추출 방식에 따른 성능 비교  
Table 2 Performance comparison of high-frequency component extraction methods

Methods	CLIPQA	MUSIQ
<b>DWT</b>	<b>0.648</b>	<b>54.212</b>
FFT	0.618	53.584
DWT + Edge	0.613	53.085

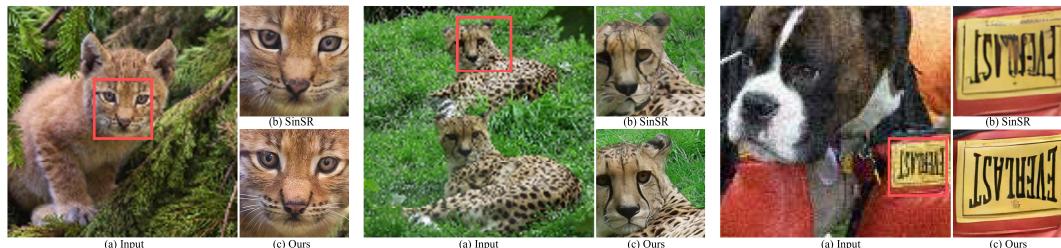


그림 3 입력 영상(a)에 대한 SinSR(b) 및 제안 방법(c)의 정성적 비교

Fig. 3 Qualitative comparison between SinSR (b) and the proposed framework (c) for the input image (a)

#### 4. 결 론

본 논문에서는 기존의 잠재 확산 모델 기반 영상 초해상화 기법들에 존재하던 세부 정보 왜곡 문제를 해결하기 위해 다중 스케일 이산 웨이블릿 변환을 활용한 새로운 프레임워크를 제안하였다. 제안한 방법은 저해상도와 고해상도 영상 간의 차이를 다양한 스케일에서 DWT로 고주파 성분을 추출하여 손실 계산에 활용했고, LPIPS 손실을 스케일에 따라 상대적으로 부여하여 가중합을 수행함으로써 영상의 세부 정보와 질감을 효과적으로 보존할 수 있도록 하였다.

다양한 벤치마크 데이터셋을 통한 실험 결과, 제안된 프레임워크는 기존의 LDM, GAN, 트랜스포머 기반 기법들과 비교하여 전반적으로 뛰어난 성능을 보였다. 특히 비참조 기반 평가 지표에서 가장 높은 성능을 보이며 고주파 성분의 표현력 향상 측면에서 우수성을 입증하였다. 또한, FFT 및 DWT와 Canny Edge 기반 고주파 추출 방법과의 비교 실험을 통해, 본 연구에서 사용한 DWT 기반 접근법이 영상의 세부 품질을 개선하는 데 효과적임을 확인하였다. 본 연구는 잠재 확산 모델의 효율성을 유지하면서도 영상 초해상화 분야에서 흔히 나타나는 세부 정보의 손실 문제를 효과적으로 극복할 수 있는 가능성을 제시하였으며, 향후 초해상화 연구 및 응용 분야에 기여할 수 있을 것으로 기대된다.

#### References

- [1] Wang, Zhihao, Jian Chen, and Steven CH Hoi. "Deep learning for image super-resolution: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 10, 2020, pp. 3365–3387.
- [2] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 2, 2015, pp. 295–307.
- [3] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).
- [4] Liang, Jingyun, et al. "Swinir: Image restoration using swin transformer." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
- [5] Saharia, Chitwan, et al. "Image super-resolution via iterative refinement." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 4, 2022, pp. 4713–4726.
- [6] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [7] Wang, Yufei, et al. "SinSR: diffusion-based image super-resolution in a single step." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
- [8] Yue, Zongsheng, Jianyi Wang, and Chen Change Loy. "Resshift: Efficient diffusion model for image super-resolution by residual shifting." *Advances in Neural Information Processing Systems*, Vol. 36, 2023, pp. 13294–13307.
- [9] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
- [10] Blau, Yochai, and Tomer Michaeli. "The perception-distortion tradeoff" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
- [11] Yu, Fanghua, et al. "Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
- [12] Wang, Jianyi, Kelvin CK Chan, and Chen Change Loy. "Exploring clip for assessing the look and feel of images." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. No. 2. 2023.
- [13] Ke, Junjie, et al. "Musiq: Multi-scale image quality transformer." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

- [14] Liang, Jingyun, et al. "Swinir: Image restoration using swin transformer." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
- [15] Wang, Xintao, et al. "Real-esrgan: Training real-world blind super-resolution with pure synthetic data." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
- [16] Liang, Jie, Hui Zeng, and Lei Zhang. "Efficient and degradation-adaptive network for real-world image super-resolution." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.



서 준 혁

2024년 국립금오공과대학교 컴퓨터공학  
과(학사). 2024년~현재 경북대학교 인공  
지능학과 석박통합과정. 관심분야는 딥러  
닝, 컴퓨터비전 등



이 동 규

2011년 광운대학교 컴퓨터공학과 졸업(학사)  
2019년 고려대학교 컴퓨터학과 졸업(박사)  
2020년~2024년 경북대학교 인공지능학  
과 조교수. 2024년~현재 경북대학교 인  
공지능학과 부교수. 관심분야는 딥러닝,  
컴퓨터비전, 멀티모달 학습