특집논문 (Special Paper)
방송공학회논문지 제30권 제4호, 2025년 7월 (JBE Vol.30, No.4, July 2025)
https://doi.org/10.5909/JBE.2025.30.4.561
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

주파수 표현 기반 상태 공간 학습을 활용한 임의 배율 초해상도 모델

권 상 윤[a)], 최 민 규[a)], 진 경 환[a)‡]

IMF: Implicit MambaIR via Fourier Representation for Arbitrary-Scale Super-Resolution

Sangyoun Kwon[a)], Mingyu Choi[a)], and Kyong Hwan Jin[a)‡]

요 약

최근 CNN과 Transformer 기반 구조는 이미지 초해상도(SR) 성능을 크게 향상시켰지만, CNN은 수용 영역이 제한되고, Transformer는 연산 복잡도가 높아 고해상도 입력에 비효율적이다. 이에 대한 대안으로 제안된 Mamba는 Structured State-Space Model(SSM)을 기반으로 선형 복잡도로 장거리 의존성을 모델링할 수 있으며, 이를 이미지 복원에 확장한 MambaIR은 Transformer 기반 SR 모델보다 연산 효율성과 복원 성능 측면에서 우수한 결과를 보였다. 그러나 MambaIR은 고주파 질감 복원이 제한되고, 고정된 배율만 지원하는 한계를 지닌다. 본 논문에서는 Fourier 기반 주파수 표현과 좌표 기반 복원을 결합한 IMF(Implicit MambaIR via Fourier Representation)를 제안한다. 제안된 모델은 상태 공간 연산에 주파수 표현을 통합하고, Implicit Neural Representation(INR) 디코더를 통해 임의 배율 복원을 실현한다. 이를 통해 단일 학습으로 다양한 배율에 대응 가능하며, SR 성능과 연산 효율성 모두에서 우수한 결과를 보였다.

Abstract

CNNs and Transformers have advanced super-resolution (SR), but suffer from limited receptive fields and high computational cost, respectively. Mamba, based on Structured State-Space Models (SSM), enables efficient long-range modeling, and its extension, MambaIR, improves efficiency over Transformer-based SR models. However, MambaIR struggles with high-frequency texture restoration and scale flexibility. We propose IMF (Implicit MambaIR via Fourier Representation), which incorporates Fourier-based frequency features into state-space computation and employs a coordinate-based Implicit Neural Representation (INR) decoder. IMF supports arbitrary-scale SR with a single training process and achieves both efficient and high-quality reconstruction.

Keyword : Arbitrary Scale Super-Resolution, Structured State-Space Model, Fourier Domain Learning, High-Frequency Texture Reconstruction, Implicit Neural Representation

Ⅰ. 서 론

초해상도(super-resolution, SR)는 저해상도 영상을 고해상도로 복원함으로써 위성 영상 분석, 의료 진단, 제조 공정 검사 등 다양한 분야에서 세밀한 정보 확보와 정밀한 의사결정에 핵심적인 역할을 수행한다. SR은 본질적으로 손실된 고주파 정보를 복원해야 하는 ill-posed 문제로, 입력 영상으로부터 시각적으로 자연스러운 세부 정보를 복원하는 정교한 표현력이 요구된다. 초기에는 bilinear나 bicubic 보간(interpolation) 기법이 널리 사용되었으나, 이러한 방식은 구조 보존과 텍스처 복원 측면에서 제약이 존재했다. 이를 극복하기 위해 SRCNN[1]과 같은 CNN 기반 학습[2] 방식이 제안되었으며, 이후 EDSR(Enhanced Deep Super-Resolution Network)[3]과 같이 깊은 잔차 구조를 도입한 모델들이 등장하면서 SR의 성능은 크게 향상되었다.

최근에는 Transformer 기반 아키텍처[4]가 전역적 문맥 정보를 효과적으로 포착할 수 있다는 장점으로 다양한 컴퓨터 비전 과제에 활발히 도입되고 있으며, SR 분야에서도 SwinIR[5]과 같은 모델을 통해 그 성능이 입증되었다. 이러한 모델들은 자기 어텐션(Self-Attention)을 통해 전역적인 특징 상호작용을 유도할 수 있으나, 연산 복잡도가 입력 해상도의 제곱에 비례하는 이차적(quadratic) 특성으로 고해상도 입력 처리 시 계산 비용과 메모리 사용량이 급격히 증가하는 구조적 한계를 지닌다. 따라서 Transformer[6] 기반 SR 모델은 뛰어난 복원 성능에도 불구하고, 연산 효율성의 한계로 실제 응용에 어려움이 있다.

이러한 한계를 극복하기 위한 대안으로 Structured State-Space Model(SSM)[7,8,9] 기반 아키텍처가 주목받고 있으며, 그중 Mamba[10]는 긴 시퀀스를 처리할 수 있도록 설계된 효율적인 순환 구조로 attention 없이도 장거리 의존성을 효과적으로 학습할 수 있는 특징을 지닌다. MambaIR[11]은 이 Mamba 구조를 이미지 복원 과제에 맞춰 확장한 모델로, 원래 1D 시퀀스를 처리하는 Mamba 블록을 2차원 영상 입력에 적용할 수 있도록 구조화하였다. 이를 통해 영상 내 전역 정보를 효과적으로 포착하면서도, self-attention 기반 모델보다 낮은 연산 복잡도를 유지할 수 있다. 특히 고해상도 입력에서도 높은 연산 효율과 복원 성능을 동시에 달성할 수 있으며, 실제로 평균 PSNR 기준 SwinIR 대비 +0.45 dB 향상을 보이는 등 정량적 성능 또한 확인되었다.

그럼에도 불구하고 기존 MambaIR은 전역 구조 복원에는 강점을 보이는 반면, 미세한 고주파 텍스처나 에지 복원에는 상대적으로 취약하며, 고정된 업스케일링 배율에 맞춰 학습되기 때문에 다양한 스케일 복원에 유연하게 대응하지 못한다. 이러한 제약을 보완하기 위한 방안으로, 최근에는 좌표 기반의 연속적 복원 방식인 Implicit Neural Representation(INR)[12]이 주목받고 있으며, 단일 구조로 다양한 스케일 복원을 효과적으로 수행할 수 있음을 확인하였다.

이와 더불어, Local Texture Estimator(LTE)[13]는 전역 정보 학습에 중점을 둔 기존 모델들과는 다른 방향에서, 주파수 도메인 기반의 국소 텍스처 표현을 강화함으로써 고주파 복원 측면에서 의미 있는 성능 향상을 이끌어낸 바 있다. LTE는 특징 맵의 로컬 패치를 푸리에 변환한 뒤, 지배적인 주파수 성분(진폭 및 위상)을 추출하고 이를 사인 및 코사인 기반의 좌표 신호로 변환하여 MLP 디코더에 입력함으로써, 고해상도 영역에서의 텍스처 정밀도를 향상시킨다. 이러한 방식은 INR 기반 복원 구조와도 자연스럽게 호환되며, 고주파 보존과 스케일 일반화를 동시에 달성할 수 있다는 점에서 주목된다.

본 연구에서는 이러한 LTE의 주파수 기반 학습 아이디어를 MambaIR의 연산 구조 내 주요 단위인 state-space 모듈 내부에 통합함으로써, 기존 SSM 기반 구조가 포착하기 어려운 미세한 고주파 정보를 보완하고자 한다. 아울러, 출력부에 좌표 기반 Implicit Neural Representation(INR) 디코더를 결합하여, 기존의 고정된 배율 중심 복원 구조를 확장하고 임의 배율의 연속 복원을 가능하게 하였다. 제안하는 구조는 기존 MambaIR의 한계를 보완하여 전역 표현력과 고주파 복원력을 모두 향상시키는 동시에, 하나의 모델로 임의 배율 복원을 실현하며 연산 효율성까지 개선하였다.

a) 고려대학교 전기전자공학부(Korea University)
‡ Corresponding Author : 진경환(Kyong Hwan Jin)
E-mail: kyong_jin@korea.ac.kr
Tel: +82-2-3290-3259
ORCID: https://orcid.org/0000-0001-7885-4792
· Manuscript May 19, 2025; Revised June 26, 2025; Accepted June 30, 2025.

Ⅱ. 제안하는 방법

본 연구에서는 최근 주목받고 있는 효율적인 시퀀스 모델링 구조인 Mamba[10]를 이미지 복원 과제에 적용한 MambaIR[11]을 기반으로, 고주파 텍스처 복원과 임의 배율 초해상도 복원을 동시에 수행하는 IMF(Implicit MambaIR via Fourier Representation)를 제안한다. 이를 위해 핵심 연산 모듈인 Vision State-Space Module(VSSM)[14]을 주파수 기반 학습이 가능한 Frequency State-Space Block(FSSB)으로 대체하고, 복원 단계에 Implicit Neural Representation (INR)[9] 구조를 도입하였다.

1. MambaIR 구조

MambaIR은 자연어 처리 분야에서 제안된 Mamba 시퀀스 모델을 이미지 복원 과제에 맞게 확장한 구조로, 전역 순차 의존성 학습을 선형 복잡도 내에서 수행할 수 있다는 구조적 효율성을 바탕으로 한다. 네트워크는 입력 이미지로부터 특징을 추출한 후, 반복적으로 적용되는 Residual State-Space Group(RSSG)을 통해 깊은 표현을 학습하고, 마지막 디코더를 통해 고해상도 이미지를 복원하는 방식으로 구성된다.

RSSG는 여러 개의 Residual State-Space Block(RSSB)으로 구성되며, 각 RSSB에는 핵심 연산 모듈인 VSSM이 포함된다. VSSM은 Mamba에서 제안된 1D Mamba block을 2D 이미지 도메인에 맞게 확장한 구조로, 다음과 같은 연속 시간 시스템을 기반으로 한다.

$$\dot{h}(t) = Ah(t) + Bx(t), \quad y(t) = Ch(t) + Dx(t) \tag{1}$$

여기서 $h(t) \in \mathbb{R}^N$는 은닉 상태, $x(t)$, $y(t) \in \mathbb{R}$는 각각 입력과 출력이며, A, B, C, D는 각각 상태 전이, 입력, 출력, 직접 연결 행렬을 의미한다. 이를 실제 신경망 연산에 적용 가능한 형태로 이산화하면 다음과 같은 recurrence 구조로 정리된다.

$$h_k = \overline{A} h_{k-1} + \overline{B} x_k, \quad y_k = C h_k + D x_k \tag{2}$$

이 수식은 VSSM[14] 내부에서 은닉 상태를 시간적으로 갱신하고 출력 특징을 생성하는 핵심 연산으로 사용된다.

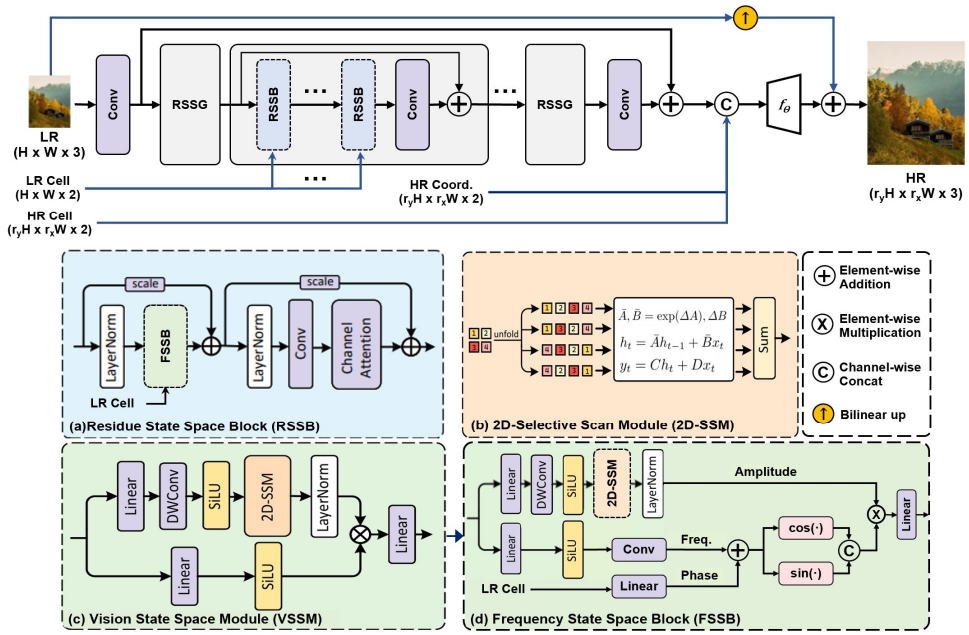

그림 1. IMF 구조 개요
Fig. 1. Overview of the IMF architecture

다만, Mamba의 원래 구조는 1D 시퀀스를 처리하도록 설계되었기 때문에, MambaIR에서는 이를 2D 이미지에 맞게 확장한 2D Selective Scan Module을 도입하였다. 이 모듈은 입력 특징 맵을 네 방향(↘, ↖, ↙, ↗)으로 펼쳐 각각 1D 시퀀스로 변환한 뒤, 방향별로 recurrence 기반 SSM 연산을 수행하고 그 결과를 평균적으로 통합함으로써, 2차원 공간 구조를 유지하면서도 전역 순차 정보를 효과적으로 학습할 수 있도록 설계되었다(그림 1 (b) 참조).

또한, 각 RSSB는 VSSM 외에도 channel attention[15]과 residual connection을 포함하고 있어, 지역 정보 손실과 채널 중복 문제를 보완한다(그림 1 (a) 참조). 이러한 RSSB가 다수 누적된 RSSG 구조는 네트워크의 수용 영역을 계층적으로 확장하며, 전역 문맥 정보와 정교한 지역 정보를 동시에 포착할 수 있는 표현력을 제공한다.

2. 주파수 기반 상태 공간 블록(FSSB) 설계

기존의 VSSM은 전역 순차 구조를 효과적으로 학습할 수 있는 구조이나, 미세한 고주파 질감 표현에는 한계가 존재한다. 이를 보완하기 위해 IMF는 RSSB 내의 VSSM을 주파수 기반 구조인 FSSB(Frequency State-Space Block)로 대체하였다(그림 1 (d) 참조). FSSB는 이전 은닉 상태 Ah_{k-1}를 입력으로 받아 푸리에 변환을 적용한 후, dominant frequency 성분을 추정한다. 이는 LTE[13]에서 제안된 주파수 해석 기법에 기반하며, 본 연구에서는 이를 상태 공간 구조 내에 통합함으로써 고주파 정보를 명시적으로 강화하였다. 이 과정은 다음 수식과 같이 표현된다.

$$F(Ah_{k-1}) = Amp_{k-1} \circ \begin{bmatrix} \cos(\pi \cdot Fr_{k-1} + Ph(c)) \\ \sin(\pi \cdot Fr_{k-1} + Ph(c)) \end{bmatrix} \quad (3)$$

여기서 Amp_{k-1}, Fr_{k-1}, $Ph(c)$, c는 각각 amplitude, frequency, phase 정보 및 LR 이미지의 셀 크기를 나타낸다. 이후 수식 (4)와 같이 역푸리에 기반한 신호 변환 구조를 통해 주파수 표현을 시공간 표현으로 복원하며, 입력 Bx_k와 결합하여 최종 은닉 상태 h_k를 얻는다.

$$h_k = F^{-1}(Amp_{k-1} \circ \begin{bmatrix} \cos(\pi \cdot Fr_{k-1} + Ph(c)) \\ \sin(\pi \cdot Fr_{k-1} + Ph(c)) \end{bmatrix}) + Bx_k \quad (4)$$

FSSB는 기존 VSSM과 동일한 위치에 삽입되며, 이를 통해 전역 순차 모델링 구조를 유지하면서도 고주파 정보를 명시적으로 주입하여 은닉 공간 내 텍스처 표현력을 강화한다.

3. INR 기반 연속 도메인 이미지 초해상도 복원

IMF는 복원 단계에서 기존 MambaIR[11]의 고정 배율 업샘플링 구조를 제거하고, 좌표 기반 continuous decoder인 Implicit Neural Representation(INR)[12]을 도입하였다. 이 디코더는 연속적인 고해상도 공간상에서 쿼리 좌표 x_q에 대해 해당 위치의 RGB 값을 직접 예측하며, 다음과 같은 함수로 정의된다.

$$I(x_q) = f_\theta(y, [x_q - v, c]) \quad (5)$$

여기서 y는 해당 위치에 가장 가까운 latent feature, v는 그 feature의 기준 anchor 좌표, c는 복원 목표 스케일에 따라 결정된 HR 이미지의 셀 크기를 나타내며, f_θ는 학습된 MLP 기반 디코딩 함수이다. 이 구조는 좌표 기반 정보를 이용하여 연속 도메인에서의 임의 배율 복원을 구현한다.

4. 학습 손실 함수

IMF는 복원 성능 향상을 위해 L_1 손실 함수를 사용한다. 이는 복원 이미지와 정답 이미지 간의 절대 차이를 최소화하는 방식으로, 다음과 같이 정의된다.

$$L = \|I_{HQ} - I_{GT}\|_1 \quad (6)$$

5. 구조 요약

IMF는 다음의 두 가지 구조적 개선을 통해 기존 MambaIR 기반 SR 구조의 한계를 효과적으로 확장한다.

1) VSSM → FSSB 치환: 주파수 기반 hidden state 보강을 통해 고주파 텍스처 표현력 강화
2) INR 기반 복원 구조: 연속 쿼리 기반으로 임의 배율 복원 가능

이러한 설계는 MambaIR의 전역 표현 학습 능력을 유지하면서도, 고주파 복원력과 스케일 일반화를 동시에 달성할 수 있는 초해상도 프레임워크를 구현한다.

Ⅲ. 실험 및 분석

1. 학습 설정

제안하는 IMF 모델은 초해상도 복원 과제에 대해 DIV2K 데이터셋을 사용하여 학습되었으며, 성능 평가는 대표적인 다섯 개 벤치마크 데이터셋(Set5[16], Set14[17], BSDS100[18], Urban100[19], Manga109[20])에서 수행되었다. 데이터 증강을 위해 수평 반전과 90°, 180°, 270° 회전이 적용되었으며, 학습 시에는 64×64 크기의 패치 단위로 이미지가 분할되어 사용되었다. MambaIR은 먼저 2배 모델을 50만 회 학습한 후, 해당 가중치를 초기 값으로 활용하여 3배 및 4배 모델을 각각 50만 회 추가 파인튜닝하는 방식으로 학습되었다. 반면, IMF는 좌표 기반 임의 스케일링 학습을 적용하여, 1배~4배 사이의 다양한 배율 패치를 무작위로 추출하고 단일 50만 회 반복만으로 모든 스케일을 학습할 수 있는 구조를 채택하였다.

2. 정량적 성능 평가(Quantitative Results)

표 1은 PSNR 및 SSIM 기준 IMF와 기존 방식들의 성능을 비교한 결과로, 모든 스케일(2배, 3배, 4배) 및 데이터셋에서 IMF가 일관되게 MambaIR[11]을 상회하였다. 특히 Urban100 데이터셋의 4배율에서, IMF는 PSNR이 최대 0.35dB, SSIM 기준 0.0097의 성능 향상을 기록하며 고주파 텍스처 재현 측면에서의 강점을 확인하였다.

그림 2는 이러한 정량적 결과를 시각화한 것으로, (a)에서는 배율 증가에 따른 SSIM 차이를 보여주며, IMF가 모든 배율에서 MambaIR보다 높은 품질을 유지하고, 배율이 커질수록 성능 격차가 더욱 두드러짐을 확인할 수 있다. (b)는 PSNR 성능을 다양한 배율에 따라 비교한 결과로,

표 1. 다양한 배율에서의 정량적 성능 비교(PSNR/SSIM)
Table 1. Quantitative comparison across scales and datasets (PSNR/SSIM)

Method	scale	#param	Set5		Set14		BSDS100		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CARN[21]	×2	1,592K	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
IMDN[22]	×2	694K	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283	38.88	0.9774
LAPAR-A[23]	×2	548K	38.01	0.9605	33.62	0.9183	32.19	0.8999	32.10	0.9283	38.67	0.9772
SwinIR-light	×2	910K	38.14	0.9611	33.86	0.9206	32.31	0.9012	32.76	0.9340	39.12	0.9783
MambaIR v1-light	×2	905K	38.13	0.9610	33.95	0.9208	32.31	0.9013	32.85	0.9349	39.20	0.9782
IMF(Ours)	×2	1,635K	38.21	0.9612	34.00	0.9212	32.36	0.9016	33.15	0.9369	39.32	0.9785
CARN[21]	×3	1,592K	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.50	0.9445
IMDN[22]	×3	703K	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9465
LAPAR-A[23]	×3	544K	34.36	0.9267	30.34	0.8421	29.11	0.8054	28.15	0.8523	33.51	0.9441
SwinIR-light	×3	918K	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
MambaIR v1-light	×3	913K	34.63	0.9288	30.54	0.8459	29.23	0.8084	28.70	0.8631	34.12	0.9479
IMF(Ours)	×3	1,635K	34.69	0.9296	30.62	0.8481	29.30	0.8107	29.03	0.8687	34.28	0.9495
CARN[21]	×4	1,592K	32.13	0.8943	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.47	0.9084
IMDN[22]	×4	715K	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838	30.45	0.9154
LAPAR-A[23]	×4	659K	32.15	0.8944	28.61	0.7818	27.61	0.7366	26.14	0.7871	30.42	0.9074
SwinIR-light	×4	930K	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
MambaIR v1-light	×4	924K	32.42	0.8977	28.74	0.7847	27.68	0.7400	26.52	0.7983	30.94	0.9135
IMF(Ours)	×4	1,635K	32.56	0.8996	28.87	0.7887	27.77	0.7433	26.87	0.8080	31.25	0.9183

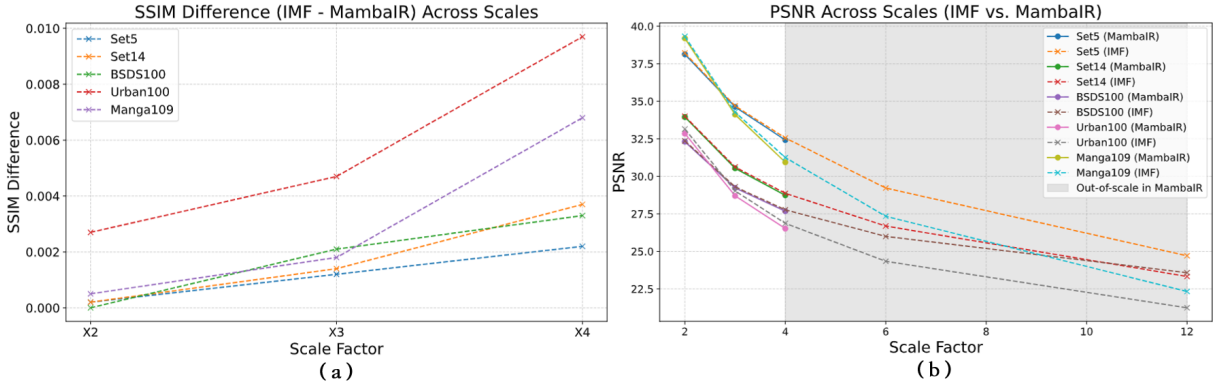

그림 2. (a) 스케일 증가에 따른 SSIM 차이(IMF – MambaIR) (b) 다양한 배율에서의 PSNR 비교(IMF vs MambaIR)
Fig. 2. (a) SSIM difference(IMF – MambaIR) across scales (b) PSNR comparison across multiple scales(IMF vs MambaIR)

MambaIR이 고정된 스케일에만 대응 가능한 반면, IMF는 학습에 포함되지 않은 고배율(out-of-scale) 조건에서도 안정적인 복원 성능을 유지함을 보여준다.

3. 정성적 성능 평가(Qualitative Results)

그림 3은 2배율 초해상도 복원에서의 정성적 비교 결과를 보여준다. BSDS100의 얼룩말 이미지(img_253027)에서는 IMF가 SwinIR-light 및 MambaIR-light보다 더 선명한 줄무늬와 경계를 복원하였으며, Urban100의 사선 구조(img_059)와 가로, 세로 줄무늬 텍스처(img_011, img_024)에서도 IMF는 더 뚜렷하고 정돈된 선형 패턴을 재현해냈다. 이러한 결과는 IMF가 다양한 구조와 데이터셋에 대해 고주파 질감 복원 능력을 일관되게 유지함을 입증한다.

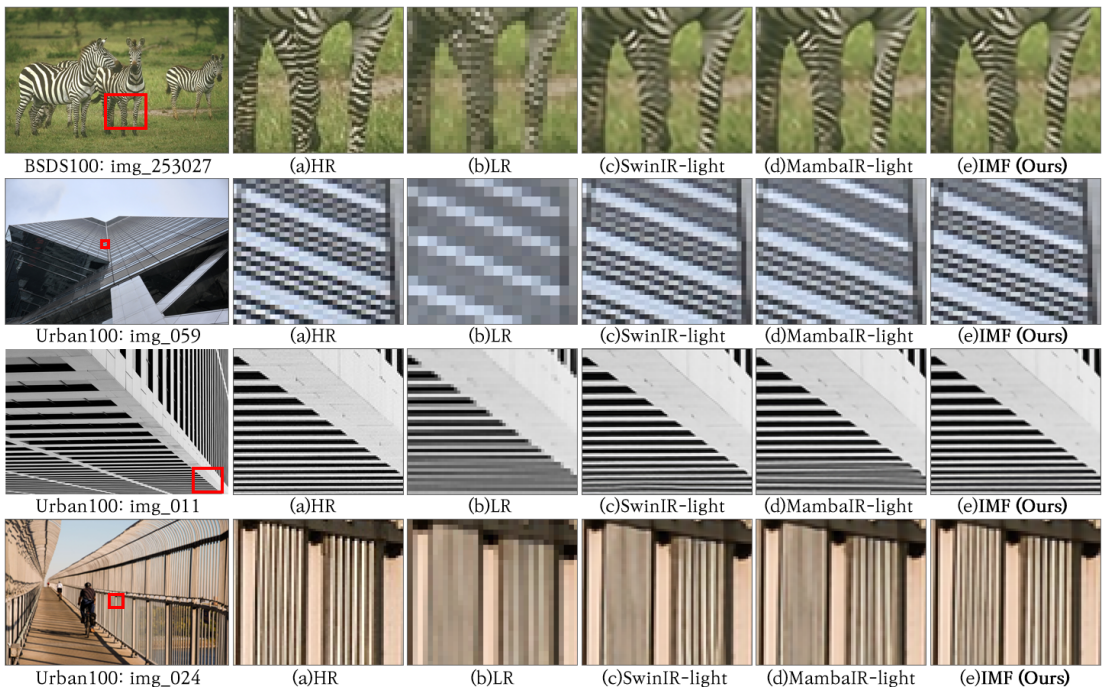

그림 3. 2배율 초해상도 복원 결과의 정성적 비교 (a) HR (b) LR (c) SwinIR-light (d) MambaIR-light (e) IMF(Ours)
Fig. 3. Qualitative comparison for x2 super-resolution (a) HR (b) LR (c) SwinIR-light (d) MambaIR-light (e) IMF(Ours)

4. 연산 효율성 비교

IMF는 FSSB 및 INR 구조를 포함함에 따라 총 파라미터 수는 약 78.9% 증가한 1,635K이지만, 모든 배율(2배, 3배, 4배)을 단일 학습으로 처리할 수 있어, 기존 MambaIR이 각 배율마다 50만 회 반복(총 1.5M)을 요구하던 것과 달리 총 50만 회만으로 전체 스케일을 학습한다. 이에 따라 학습 반복 횟수 기준 66.7%의 절감 효과가 발생하며, 결과적으로 전체 학습 비용은 오히려 감소하였다(표 2 참조). 이는 학습 시간과 계산 자원 측면에서의 효율성과 실용성 측면에서 매우 유리한 구조임을 나타낸다.

5. Ablation Study

FSSB와 INR 모듈의 개별 기여도를 정량적으로 검증하기 위해, 2배율 조건에서 15만 회 동안 학습한 후 Urban100 데이터셋을 활용해 Ablation 평가를 수행하였다. 표 3의 결과에 따르면, FSSB 모듈만을 적용한 경우 PSNR이 baseline 대비 +0.12dB 향상되었고, INR만 적용한 경우는 +0.06dB 향상에 그쳤다. 이를 통해 FSSB 모듈이 주파수 정보를 반영함으로써 복원 성능 향상에 보다 지배적으로 기여했음을 확인할 수 있었다. 또한, 두 모듈을 동시에 적용한 IMF(FSSB+INR) 모델은 baseline 대비 +0.20dB 향상을 달성하여, 두 모듈의 결합이 상호 보완적으로 작용하며 시너지 효과를 내는 것으로 분석된다.

추가로, 제안된 IMF 모델의 학습 안정성을 확인하기 위해 2배율 조건에서 50만 회 동안의 학습 과정 중 validation PSNR의 변화를 관찰하였다(그림 4 참조). 그 결과, IMF 모델이 전 구간에 걸쳐 baseline 대비 일관되게 높은 PSNR 값을 유지하며 학습이 진행되었음을 보여준다. 이는 제안된 모델이 구조적으로 안정적인 학습 특성을 갖고 있음을 의미한다.

표 2. IMF와 MambaIR의 연산 효율성 비교
Table 2. Comparison of efficiency between IMF and MambaIR

Method	Scale	# Parameters	# Training Iterations
MambaIR	X2	905K	0.5M
	X3	913K	0.5M
	X4	924K	0.5M
	Total	**914K (Avg.)**	**1.5M (Sum.)**
IMF (Ours)	X2	1,635K	0.5M
	X3		
	X4		
	Total	**1,635K (▲78.9%)**	**0.5M (▼66.7%)**

표 3. FSSB 및 INR 모듈에 대한 Ablation 실험 결과
Table 3. Ablation Results for FSSB and INR Modules

Variant	w/ FSSB	w/ INR	Scale	iter	Urban100
(a) MambaIR (baseline)			X2	150K	32.48
(b) MambaIR + FSSB	✓				32.60
(c) MambaIR + INR		✓			32.54
(d) IMF (FSSB +INR)	✓	✓			32.68

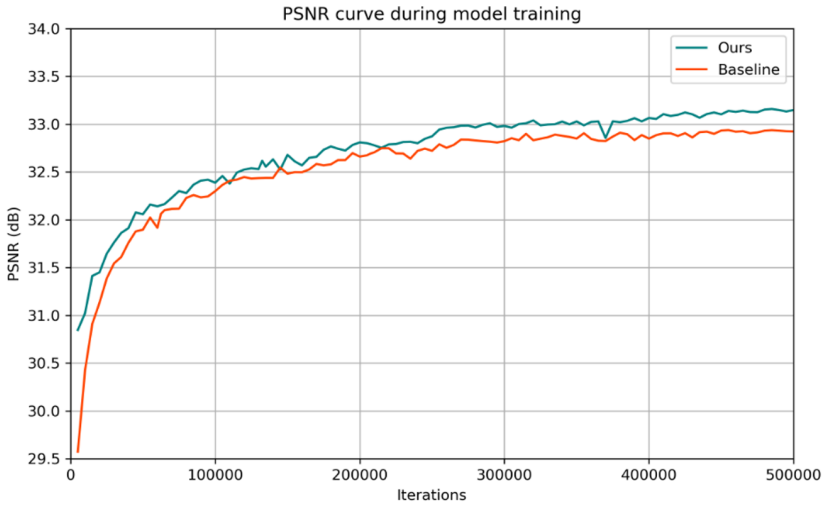

그림 4. 2배율 학습 과정에서의 검증 PSNR 변화 추이
Fig. 4. Validation PSNR Trend during ×2 Scale Training

6. 종합 요약

IMF는 다음과 같은 측면에서 기존 MambaIR 대비 실질적 우위를 가진다.

- 복원 성능 향상: 모든 배율 및 벤치마크에서 PSNR/SSIM 우위 확보
- 고주파 표현력 강화: 텍스처 중심 복원에서 더 높은 정밀도 확보
- 범용 배율 대응력 확보: arbitrary scale 복원이 가능한 구조
- 학습 효율성 확보: 단일 학습으로 다양한 배율 대응 가능

Ⅳ. 결 론

본 연구에서는 전역 구조 복원에 강점을 가진 MambaIR의 Structured State-Space Model(SSM) 기반 아키텍처를 확장하여, 고주파 질감 표현과 다양한 배율 복원 성능을 동시에 향상시킨 초해상도 프레임워크 IMF를 제안하였다. IMF는 기존 MambaIR의 VSSM을 주파수 기반 연산 모듈인 FSSB로 대체하여 은닉 상태 공간에서 지배 주파수 성분(dominant frequency)을 직접적으로 추정·활용할 수 있도록 설계되었으며, 좌표 기반 INR 디코더를 통해 임의 배율이 가능한 유연한 복원 구조를 구현하였다.

실험 결과, IMF는 PSNR 기준 최대 0.35dB의 향상을 달성하였고, 고주파 텍스처가 풍부한 Urban100 및 Manga109 데이터셋에서 특히 뛰어난 재현 성능을 보였다. 또한 단일 학습으로 다양한 배율에 대응 가능하여, 연산 효율성과 실용성 측면에서도 우수한 성과를 확인하였다.

참 고 문 헌 (References)

[1] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. European Conference on Computer Vision (ECCV), Zürich, Switzerland, pp. 184 – 199, Sep. 2014.
doi: https://doi.org/10.1007/978-3-319-10593-2_13

[2] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2472 – 2481, Jun. 2018.
doi: https://doi.org/10.1109/CVPR.2018.00262

[3] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 136 – 144, Jul. 2017.
doi: https://doi.org/10.1109/CVPRW.2017.151

[4] X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, “Activating more pixels in image super-resolution transformer,” in Proc. IEEE/CVF

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22367 – 22377, Jun. 2023.
doi: https://doi.org/10.1109/CVPR52729.2023.02142

[5] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in Proc. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1833 – 1844, Oct. 2021.
doi: https://doi.org/10.1109/ICCVW54120.2021.00210

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998 – 6008, Dec. 2017.
doi: https://doi.org/10.48550/arXiv.1706.03762

[7] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," arXiv preprintarXiv:2111.00396, Nov. 2021.
doi: https://doi.org/10.48550/arXiv.2111.00396

[8] A. Gu, I. Johnson, K. Goel, K. Saab, T. Dao, A. Rudra, and C. Ré, "Combining recurrent, convolutional, and continuous-time models with linear state space layers," in Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 572 – 585, Dec. 2021.
doi: http://doi.org/10.48550/arXiv.2110.13985

[9] J. T. H. Smith, A. Warrington, and S. W. Linderman, "Simplified state space layers for sequence modeling," arXiv preprint, arXiv: 2208.04933, Aug. 2022.
doi: https://doi.org/10.48550/arXiv.2208.04933

[10] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," arXiv preprintarXiv:2312.00752, Dec. 2023.
doi: https://doi.org/10.48550/arXiv.2312.00752

[11] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia, "MambaIR: A simple baseline for image restoration with state-space model," in Proc. European Conference on Computer Vision (ECCV), pp. 222 – 241, Sep. 2024.
doi: https://doi.org/10.1007/978-3-031-72649-1_13

[12] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8628 – 8638, Jun. 2021.
doi: https://doi.org/10.1109/CVPR46437.2021.00852

[13] J. Lee and K. H. Jin, "Local Texture Estimator for Implicit Representation Function," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1929 – 1938, Jun. 2022.
doi: https://doi.org/10.1109/CVPR52688.2022.00197

[14] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, "VMamba: Visual state space model," arXiv preprintarXiv: 2401.10166, Jan. 2024.
doi: https://doi.org/10.48550/arXiv.2401.10166

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132 – 7141, Jun. 2018.
doi: https://doi.org/10.1109/CVPR.2018.00745

[16] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in Proc. British Machine Vision Conference (BMVC), pp. 135.1 – 135.10, Sep. 2012.
doi: https://doi.org/10.5244/C.26.135

[17] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse representations," in Curves and Surfaces: 7th International Conference, Avignon, France, Jun. 2010, pp. 711 – 730.
doi: https://doi.org/10.1007/978-3-642-27413-8_47

[18] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in Proc. Eighth IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 416 – 423, Jul. 2001.
doi: https://doi.org/10.1109/ICCV.2001.937655

[19] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5197 – 5206, Jun. 2015.
doi: https://doi.org/10.1109/CVPR.2015.7299156

[20] Y. Matsui, K. Ito, Y. Aramaki, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using Manga109 dataset," Multimedia Tools and Applications, vol. 76, no. 20, pp. 21811 – 21838, Oct. 2017.
doi: https://doi.org/10.1007/s11042-016-4020-z

[21] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in Proc. European Conference on Computer Vision (ECCV), pp. 252 – 268, Sep. 2018.
doi: https://doi.org/10.1007/978-3-030-01249-6_16

[22] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in Proc. 27th ACM International Conference on Multimedia, pp. 2024 – 2032, Oct. 2019.
doi: https://doi.org/10.1145/3343031.3351084

[23] W. Li, K. Zhou, L. Qi, N. Jiang, J. Lu, and J. Jia, "LAPAR: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 20343 – 20355, Dec. 2020.
doi: https://doi.org/10.48550/arXiv.2105.10422

저 자 소 개

권 상 윤

- 2014년 1월 ~ 2024년 2월 : 삼성전자 반도체연구소
- 2024년 3월 ~ 현재 : 고려대학교 전기전자공학부 석사과정
- ORCID : https://orcid.org/0009-0006-9711-2253
- 주관심분야 : 초해상도, 이상감지

최 민 규

- 2012년 1월 ~ 2024년 8월 : LG디스플레이
- 2024년 9월 ~ 현재 : 고려대학교 전기전자공학부 박사과정
- ORCID : https://orcid.org/0009-0009-1600-611X
- 주관심분야 : 초해상도, 역문제

진 경 환

- 2008년 ~ 2015년 : KAIST 바이오및뇌공학과 박사
- 2016년 6월 ~ 2019년 8월 : EPFL Biomedical Imaging Group Post-doc
- 2019년 9월 ~ 2021년 2월 : 삼성리서치 Camera T/F – 글로벌 AI 센터 연구원
- 2021년 2월 ~ 2023년 8월 : DGIST 전기전자컴퓨터공학과 조교수
- 2023년 9월 ~ 현재 : 고려대학교 전기전자공학부 부교수
- ORCID : https://orcid.org/0000-0001-7885-4792
- 주관심분야 : 딥러닝, 신호처리, 역문제