

스펙트럼-정합 확산 샘플링을 통한 이미지 초해상도 연구

전승재, 권진경, 전윤희, 박천음*

국립한밭대학교

{sjjeon, 20221043}@edu.hanbat.ac.kr, {yhjeon, parkce}@hanbat.ac.kr

Spectral-Aligned Diffusion Sampling for Image Super-Resolution

Seungjae Jeon, Jinkyung Kwon, Yunho Jeon, Cheoneum Park*

Hanbat National University

요약

확산 모델은 이미지 생성과 초해상도에서 뛰어난 세부 묘사를 보이지만, over-smooth와 grainy artifact가 여전히 관찰된다. 본 논문은 저해상도 이미지의 주파수 정보를 추출하여 역확산 단계에 스펙트럼-정합 가이드선으로 주입함으로써 이러한 문제를 완화하는 방법을 제안한다. 제안 방식은 외부 모델·추가 학습이 없는 플러그-인 방법으로, 어떤 확산 기반 SR 샘플러에도 적용 가능하다. SRDiff 기반 평가에서 표준 벤치마크 기준 PSNR이 평균 +0.9 dB, WSNR이 +1.1 dB 향상되었고, 주파수 오차가 감소하는 경향을 보였다.

I. 서론

이미지 초해상도(Super-Resolution, SR)는 저해상도(Low-Resolution, LR) 이미지로부터 고해상도 영상을 복원하는 기술로 의료, 위성, 보안 분야에서 중요한 역할을 한다. 최근 확산 모델이 SR에 도입되어 우수한 성능을 보이며 활발히 연구되는 중이다[1, 2]. 또한 사전 학습된 확산 모델을 활용한 접근법들이 주목받고 있는데, 먼저 DDRM [3]은 역확산 과정에서 열화 연산자의 null space 성분만을 복원하는 방법이다. DDNM [4]은 입력에 없는 고주파 성분을 확산 과정에서 보충하는 방식으로 추가 학습 없이도 높은 성능을 달성한다.

이처럼 확산 모델 기반 SR은 유의미한 결과를 보였지만, 여전히 over-smooth와 grainy noise 같은 고질적 한계가 있다. 이를 극복하기 위해 Spectral Diffusion [5], Wavelet Diffusion GAN [6]와 같이 주파수 영역 정보를 활용하는 연구들이 진행된다: Spectral Diffusion은 역확산 과정에서 고주파 성분이 나중에 생성된다는 관찰을 기반으로 U-Net에 wave gating 모듈을 추가하여 고주파 성분을 강조한다. Wavelet Diffusion GAN은 Discrete Wavelet Transform을 통해 영상을 주파수 성분으로 분해한 후 저차원 공간에서 확산을 수행하여 적은 step으로도 고주파를 보존한다.

본 논문은 확산 모델의 low-frequency bias 문제를 해결하기 위해 LR 영상에서 추출한 주파수 정보를 역확산 단계에 spectral-aligned 가이드선(guidance)로 주입하는 추론 전용 플러그인을 제안한다. 제안 방법은 기존 샘플러(DDPM, DDIM, DPM-Solver++)에 추가 학습 없이 적용 가능하며, SRDiff[2] 모델을 백본으로 한 표준 벤치마크 실험(x4, 100 steps)에서 PSNR +0.9 dB, WSNR +1.1 dB의 성능 향상을 달성하고 주파수 오차 감소를 확인한다.

II. 제안 방법

확산 기반 초해상도는 [수식 1]와 같이 고해상도(High Resolution, HR) 이미지 $\mathbf{x}_0$ 를 점진적으로 가우시안 노이즈로 변형하는 확산 과정과, 이를 저해상도 조건 $\mathbf{y}$ 에서 복원하는 역확산 과정으로 정의된다.

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}) \quad (1)$$

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_\uparrow) = \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow), \sigma_t^2\mathbf{I})$$

여기서 $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, $\alpha_s \in (0, 1)$ 는 노이즈 분산 스케줄이다. $\boldsymbol{\mu}_\theta$ 는 [수식 2]과 같이 ϵ 을 예측하는 신경망 ϵ_θ 로 매개변수화된다.

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow) \right) \quad (2)$$

본 논문에서는 사전학습된 확산 초해상도 모델(SR3, SRDiff 등)에 대해 샘플링 단계에서만 가이드선을 주입하는 방법을 제안한다. 가이드선이 포함된 역확산 업데이트는 [수식 3]과 같다.

$$\mathbf{x}_{t-1} \leftarrow \boldsymbol{\mu}_\theta(\mathbf{x}_t, t, \mathbf{y}_\uparrow) + \sigma_t \mathbf{z} - \eta_t \nabla_{\mathbf{x}} \mathcal{E}(\mathbf{x}_{t-1}) \quad (3)$$

여기서 $\eta_t > 0$ 는 가이드선 스케일, $\mathbf{y}$ 는 LR 이미지, $\mathbf{y}_\uparrow$ 는 LR 이미지를 bicubic 업샘플링한 결과이다. $\mathcal{E}(\mathbf{x})$ 는 본 논문에서 제안하는 결합 가이드선이다. 이는 [수식 4]와 같이 데이터 일관성 항과 주파수 가이드선 항의 가중합으로 정의된다.

$$\mathcal{E}(\mathbf{x}) = \lambda_{dc} \mathcal{L}_{dc}(\mathbf{x}; \mathbf{y}) + \lambda_{freq} \mathcal{L}_{freq}(\mathbf{x}; \mathbf{y}) \quad (4)$$

데이터 일관성 항과 그래디언트(gradient)는 [수식 5]와 같이 계산된다. 다운샘플링 연산자 D 에 의해 HR 이미지가 LR 관측값과 일치하도록 강제한다.

$$\mathcal{L}_{dc}(\mathbf{x}; \mathbf{y}) = \|D(\mathbf{x}) - \mathbf{y}\|_2^2, \quad \nabla_{\mathbf{x}} \mathcal{L}_{dc} = 2D^\top (D(\mathbf{x}) - \mathbf{y}) \quad (5)$$

이 방법은 $\mathbf{y}$ 의 스펙트럼을 참조하여, 고주파 영역을 선택적으로 보강한다. [수식 6]에서 $\mathcal{F}$ 는 2D FFT (Fast Fourier Transform), $|\cdot|$ 는 스펙트럼 크기이다. 먼저 LR의 정규화된 스펙트럼 분포를 정의한다. 이를 기반으로 고주파를 강조하는 가중 마스크 $W(\omega)$ 를 구성한다. 단조 증가 함수 $\phi(S_y)$ 는 $\exp(S_y/\tau) - 1$ 으로 설정했다. 여기서 τ 는 고주파 영역을 결정하는 임계값이다.

Dataset	Method	PSNR [↑]	SSIM [↑]	MS-SSIM [↑]	LPIPS [↓]	FID [↓]	WSNR [↑]	Freq Error [↓]	LR PSNR [↑]
Set5	DDPM	<u>26.926</u>	0.754	<u>0.958</u>	0.161	<u>45.564</u>	<u>30.793</u>	0.090	30.733
	DDIM	19.551	0.318	0.823	0.630	110.964	23.184	0.719	22.811
	DPM++	19.471	0.314	0.821	0.632	115.044	23.096	0.733	22.757
	Ours(DDPM)	27.837	0.819	0.969	<u>0.167</u>	43.245	31.750	0.068	<u>34.009</u>
	Ours(DDIM)	26.497	0.743	0.952	0.197	51.141	30.461	0.080	34.646
	Ours(DPM++)	26.327	<u>0.759</u>	0.956	0.223	57.384	30.413	<u>0.075</u>	33.293
BSD100	DDPM	<u>24.665</u>	0.624	<u>0.919</u>	0.332	79.989	<u>27.996</u>	0.193	30.812
	DDIM	18.761	0.243	0.764	0.721	144.827	22.355	0.875	22.542
	DPM++	18.714	0.241	0.762	0.722	145.080	22.308	0.886	22.501
	Ours(DDPM)	25.226	0.685	0.929	<u>0.358</u>	<u>81.170</u>	28.539	0.170	<u>33.956</u>
	Ours(DDIM)	24.152	0.601	0.901	0.391	90.563	27.554	0.191	34.492
	Ours(DPM++)	24.216	<u>0.634</u>	0.910	0.414	87.610	27.619	<u>0.184</u>	33.729
Urban100	DDPM	<u>22.591</u>	<u>0.651</u>	<u>0.929</u>	<u>0.312</u>	10.059	<u>26.017</u>	0.331	28.535
	DDIM	17.985	0.293	0.793	0.711	34.102	21.440	0.994	22.382
	DPM++	17.944	0.291	0.792	0.713	35.837	21.396	1.006	22.345
	Ours(DDPM)	22.952	0.704	0.937	0.300	<u>11.879</u>	26.368	0.307	31.490
	Ours(DDIM)	22.118	0.621	0.912	0.368	24.398	25.596	0.327	33.824
	Ours(DPM++)	22.072	0.645	0.918	0.366	23.211	25.610	<u>0.320</u>	<u>33.008</u>

표 1: 다양한 샘플러에 가이드를 적용한 결과이다. 가장 우수한 수치는 **볼드체**, 그다음 우수한 수치는 밑줄로 표기했다.

$$S_y(\omega) = \frac{|\mathcal{F}\{y\}(\omega)|^2}{\|\mathcal{F}\{y\}\|_2^2}, W(\omega) = 1 + \gamma \cdot \phi(S_y(\omega)) \quad (6)$$

아래 [수식 7]은 주파수 가이드와 그라디언트이다.

$$\begin{aligned} \mathcal{L}_{\text{freq}}(\mathbf{x}; \mathbf{y}) &= \|W(\omega) \odot (|\mathcal{F}\{\mathbf{x}\}| - \alpha_t \cdot |\mathcal{F}\{\mathbf{y}\}|)\|_2^2 \\ \nabla_{\mathbf{x}} \mathcal{L}_{\text{freq}} &= 2 \Re \left[\mathcal{F}^{-1} \left(W(\omega)^2 \odot (|\mathcal{F}\{\mathbf{x}\}| - \alpha_t |\mathcal{F}\{\mathbf{y}\}|) \cdot \frac{\mathcal{F}\{\mathbf{x}\}}{|\mathcal{F}\{\mathbf{x}\}|} \right) \right] \end{aligned} \quad (7)$$

여기서 $\Re$ 는 실수부를 취하는 연산이며, α_t 는 t 에 따라 고주파를 증가시키는 스케줄 파라미터이다.

III. 실험

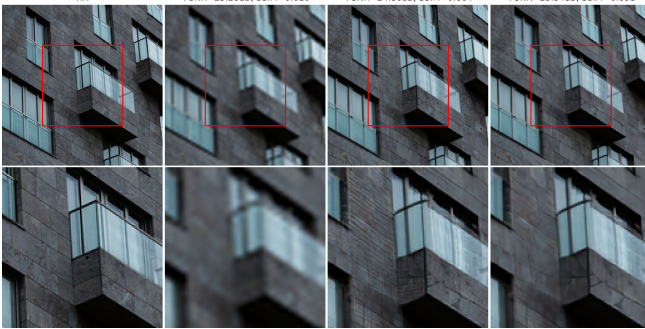


그림 1: 제안한 가이드의 아티팩트 제거 효과를 보여준다. 아래 사진은 빨간 박스 영역을 2배 확대한 것이다.

제안 방법의 유효성 검증을 위해, DIV2K 데이터셋으로 사전 학습된 SRDiff를 고정하고 샘플러를 교체하며 실험한다. 모든 실험은 4 초해상도와 100-step으로 설정한다. 샘플러 비교는 DDPM, DDIM, DPM-Solver++와 수행하며, 각 샘플러에 제안한 스펙트럼-정합 가이드를 플러그인 형태로 결합하여 평가한다. 평가는 Set5, Set14, BSD100, Urban100 데이터셋으로 수행하고, 평가 지표는 왜곡 지표 PSNR/SSIM/MS-SSIM, 지각 지표 LPIPS/FID, 주파수 특화 지표 WSNR/Freq Error, 다운샘플 재투영 일관성 LR PSNR을 사용한다.

[표 1]에 따라 제안 방법이 왜곡, 지각, 주파수 지표를 동시에 향상시킬 수 있다. Ours (DDPM)은 기본 DDPM 대비 PSNR과 WSNR이 안정적으로 상승하고 Freq Error가 감소한다. 지각 지표는 데이터셋에 따라 등락이 있으나, Urban100에서는 LPIPS가 0.312→0.300로 감소하는 등 고주파 데이터 셋에서 강인한 특성을 보인다. 한편 결정론 샘플러(DDIM/DPM++)에 가이드를

결합했을 때 효과가 더 크다. 이는 결정론적 샘플러가 계산 효율성을 위해 확률적 노이즈를 제거하면서 손실한 고주파 복원 능력을, 제안한 스펙트럼-정합 가이드가 보완했기 때문으로 사료된다. 추가적으로, 모든 데이터셋에서 LR PSNR이 유의미하게 증가했는데, 이는 제안한 가이드가 LR 일관성을 강화하고 [그림 1]과 같이 grain artifact 생성을 억제했다고 해석할 수 있다.

IV. 결론

본 연구는 LR에서 주파수 정보를 추출하여 역확산 단계에 스펙트럼-정합 가이드로 주입하는 플러그인을 제안하였다. 실험 결과, SRDiff x4 100-step 설정에서 기존 샘플러 대비 PSNR +0.9 dB, WSNR +1.1 dB의 향상을 보였으며, 특히 DDIM/DPM++ 같은 결정론적 샘플러에서는 성능이 더 크게 개선되었다. 이는 주파수 정합이 저주파 편향에 따른 평활화와 고주파 디테일 손실/잡음을 완화하는 경향 해법임을 보였다. 다만 가이드 스케일, FFT 실험 설계의 민감성과 LR에서 정보를 추출한 한계로 평활화를 완전히 해결하진 못했다. 향후 확산 모델의 ϵ_θ 에 대한 주파수 분석 연구를 진행하여 샘플러 적용 가능성을 탐구할 예정이다.

참고 문헌

- [1] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, “Image super-resolution via iterative refinement,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, 2023.
- [2] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, “SRdiff: Single image super-resolution with diffusion probabilistic models,” *Neurocomputing*, vol. 479, pp. 47–59, 2022.
- [3] B. Kavar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [4] Y. Wang, J. Yu, and J. Zhang, “Zero-shot image restoration using denoising diffusion null-space model,” in *International Conference on Learning Representations (ICLR)*, 2023.
- [5] X. Yang, D. Zhou, J. Feng, and X. Wang, “Diffusion probabilistic model made slim,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22552–22561, 2023.
- [6] L. Aloisi, L. Sigillo, A. Uncini, and D. Communiello, “A wavelet diffusion GAN for image super-resolution,” in *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2024.