

고주파에 적합한 교차 엔트로피 손실함수에 대한 초해상도

오윤주¹, 김태현²

¹한양대학교 인공지능학과 박사과정

²한양대학교 인공지능학과 교수

oyj0813@hanyang.ac.kr, taehyunkim@hanyang.ac.kr

Super-Resolution with Cross-Entropy Loss Adapted to High Frequencies

Oh Yoon Ju¹, Kim Tae Hyun²

¹Dept. of Artificial Intelligence, Hanyang University

²Dept. of Artificial Intelligence, Hanyang University

요약

Super resolution에서 High-frequency Details를 개선하는 것이 최근 문제이다. 기존에는 Super resolution을 Regression task로 접근하므로써 L2 Loss를 사용하여 이미지가 흐릿하게 되었다. 이를 해결하기 위해, Classification task로 바꾸므로써 Cross Entropy Loss를 적용하여 Cross-entropy Super-resolution (CS)를 설계한다. CS를 통해 선명도와 Details이 개선되지만, 저주파의 CE Loss 학습으로 인한 Black Artifacts가 발생한다. 그래서, L2 Loss는 저주파와 같이 큰 신호에 더 초점을 맞추므로, 성능 개선을 위해 저주파를 L2 Loss에서, 고주파를 CE Loss에서 학습시킨 Frequency-specific Cross-entropy Super-resolution (FCS)을 제안한다. 우리는 왜곡에 강하며 Human의 인식과 유사한 측정지표인 Learned Perceptual Image Patch Similarity (LPIPS)로 평가한다. 실험한 모든 데이터 셋에서 우리의 FCS는 Baseline보다 LPIPS가 약 1.7배 정도 개선되었다.

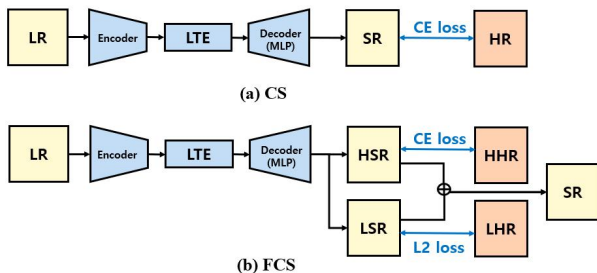
1. 서론

Super resolution (SR)은 Low resolution (LR)에서 소실된 정보를 복원하며, 최근에는 Implicit한 방법을 사용하여 이미지를 Arbitrary resolution으로 표현하는 연구가 등장하였다. LTE [1]는 Continuous representation을 학습하므로 High-frequency Details을 표현한다. 그런데, L1 loss를 사용하기 때문에 Blurry SR을 만드는 문제가 있다. Edge나 Texture와 같은 High-frequency Information은 Blur 효과가 없는 것이 Details를 더 잘 살릴 수 있으므로, Cross Entropy (CE) Loss 사

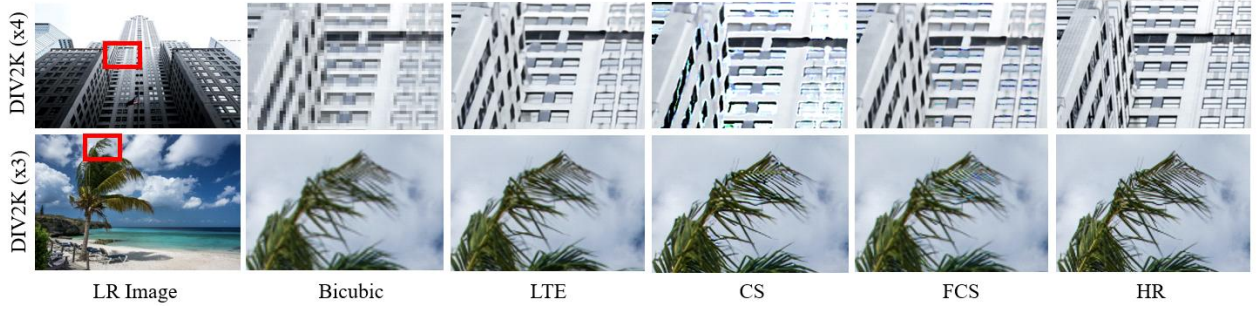
용한다. CE Loss는 픽셀 간의 상대적인 차이에 민감하게 반응하므로, 고주파에 대한 예측 오차를 줄이는 데 더 많은 가중치를 부여할 수 있다. 그래서, 고주파가 더 정확하게 복원된다. 따라서, 네트워크를 CE Loss로만 학습시키는 Cross-entropy Super-resolution (CS)를 설계한다. 또한, CE Loss를 사용하기 위해 Super resolution을 Regression task에서 Classification task로 변환한다.

CS는 선명하고 Detail이 개선되지만, 저주파를 CE Loss로 학습하여 Black Artifacts가 발생하며, [2]에서 L1, L2 Loss는 저주파와 같이 큰 신호에 더 초점을 맞춘다고 언급하였다. 따라서, 저주파 이미지는 L2 Loss로, 고주파 이미지는 CE Loss로 학습시키는 Frequency-specific Cross-entropy Super-resolution (FCS)를 제안한다.

실험한 모든 데이터셋에서 CS와 FCS는 Baseline보다 LPIPS가 개선되었다. FCS는 CS의 Black Artifacts를 완화시켰지만, LPIPS는 CS보다 높으므로 Artifacts 제거와 고주파에 적합한 네트워크 설계가 필요하다.



(그림 1) CS 및 FCS 아키텍처



(그림 2) SR 결과에 대한 기존 연구와 우리 관측 실험의 정성적 비교

2. 방 법

2.1 Frequency-specific Loss

그림 1에서 (a)는 CS 모델로, CE Loss의 특성을 가진 SR 이미지를 생성한다. Output의 Channel 수는 256개로 구성하여 0에서 255의 pixel 값을 정수형으로 표현할 수 있도록 한다. 다음은 CS 모델의 손실함수로 사용되는 CE Loss에 대한 Formulation이다.

$$O_1 = \sum_{i=1}^M (1 [\hat{y}_i]^T \log y_i) \quad (1)$$

그림 1에서 (b)는 FCS 모델로, HHR은 CE Loss로 학습시켜 HSR 이미지를 얻고, LHR는 L2 Loss로 학습시켜 LSR 이미지를 생성한다. 이 생성된 두 이미지를 결합하여 최종적으로 SR 이미지를 얻게 된다. 다음과 같이 FCS 모델의 Frequency-specific Loss는 Jointly-training을 적용한다.

$$O_2 = \sum_{i=1}^M (1 [g(\hat{y}_i)]^T \log y_{iH} + \|y_{iL} - f(\hat{y}_i)\|_2^2) \quad (2)$$

LHR는 $f(\hat{y})$ 으로 정의한다. $f()$ 는 LPF이고, \hat{y} 은 HR이다. HHR는 $g(\hat{y}) = \hat{y} - f(\hat{y})$ 으로 정의한다. SR은 $y_k = y_{kL} + y_{kH}$ 으로 표현한다. y_{kL} 은 LSR이고, y_{kH} 은 HSR이다.

3. 실 험

3.1 정량적 실험 결과

표 1는 CS와 FCS에서 LPIPS는 Baseline을 능가한다. 결과적으로, 고주파와 저주파로 분리해서 학습할 때, LPIPS는 Baseline을 약 1.7배 정도 능가한다.

3.2 정성적 실험 결과

그림 2에서, CS는 가장 선명하고, Detail한 부분에서 Black & Bright Artifacts가 심하다. 반면, FCS는 Edge의 Black Artifacts가 개선되었다. 여전히 존재하는 Bright Artifacts 제거에 대한 추가 연구가 필요하다.

		x2		x3		x4	
		PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
DIV2K	Baseline	34.72	0.0591	31.02	0.1407	29.04	0.2021
	CS	34.28	0.0319	30.24	0.0663	28.08	0.0964
	FCS	34.16	0.0331	30.50	0.0695	28.59	0.1011
Set5	Baseline	38.04	0.0394	34.43	0.0925	32.24	0.1356
	CS	37.83	0.0213	34.14	0.0495	31.85	0.0722
	FCS	37.84	0.0223	34.26	0.0525	31.92	0.0794
Set14	Baseline	33.72	0.0700	30.37	0.1617	28.65	0.2292
	CS	33.30	0.0393	29.62	0.0830	27.88	0.1199
	FCS	33.52	0.0400	30.10	0.0862	28.39	0.1254
B100	Baseline	32.21	0.1117	29.14	0.2291	27.62	0.3085
	CS	32.00	0.0549	28.71	0.0977	27.02	0.1325
	FCS	32.06	0.0554	28.99	0.1001	27.49	0.1374
Urban 100	Baseline	32.29	0.0476	28.32	0.1209	26.24	0.1827
	CS	31.59	0.0407	27.34	0.0911	24.95	0.1366
	FCS	31.73	0.0403	27.75	0.0939	25.72	0.1427

<표 1> Benchmark datasets에 대한 정량적 측정 결과 (빨간색: 가장 좋은 결과, 파란색: 두 번째로 좋은 결과)

4. 결 론

Super resolution을 Classification task로써 CE Loss를 사용하여 Blurry한 이미지를 최소화하도록 하는 CS를 설계한다. Black Artifacts가 발생으로 저주파를 L2 Loss로 설계하고, 고주파를 CE Loss에 적용하는 FCS를 제안한다. 실험을 통해 LPIPS에서 CS와 FCS가 Baseline을 능가함을 보였다.

참고문헌

- [1] Lee, Jaewon, and Kyong Hwan Jin. "Local texture estimator for implicit representation function." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- [2] Schwarz, Katja, Yiyi Liao, and Andreas Geiger. "On the frequency bias of generative models." Advances in Neural Information Processing Systems. 2021.