

불확실성으로 유도된 Sparse Sampling Attention 을 활용한 UDAformer 의 수중 이미지 향상 성능 개선

김민성¹, 김태현²

¹ 한양대학교 AI 응용학과 석사과정

² 한양대학교 컴퓨터소프트웨어학부 교수

idecotu@naver.com, taehyunkim@hanyang.ac.kr

Boosting UDAformer for Underwater Image Enhancement via Uncertainty Guided Sparse Sampling Attention

Min-Seong Kim¹, Tae-Hyun Kim²

¹Dept. of Artificial Intelligence Application, Hanyang University

²Dept. of Computer Science, Hanyang University

요 약

본 논문은 UDAformer[1] 기반 수중 이미지 향상(Underwater Image Enhancement, UIE)에 불확실성 기반 랭킹에서 유도된 Sparse Sampling Attention Pre Adapter(SSA)를 삽입하여 성능을 높이는 방법을 제안한다. UDAformer[1]의 원래 구조는 변경하지 않고, Encoder-1/Decoder-1/Encoder-2/Decoder-2 진입부에 불확실성을 알은 잔차 형태로 주입하는 pre-adapter 를 추가하여 입력의 특징을 정제하였다. 불확실성 랭킹 σ 는 1*1 convnet 인 U-Head 로 추정하고 학습 과정에서 Uncertainty Driven Loss(UDL)[2]에 의해 학습되어 이미지에서 개선이 어려워지는 지점들에 연산을 집중시킨다. 본 논문에서 제시한 구조로 UIEB-890[3] 데이터셋을 train = 800/test = 90 으로 분할한 뒤, 500 epoch 로 학습한 결과, 베이스라인(UDAformer[1]) PSNR 23.48dB/SSIM 0.92 대비, PSNR 24.1713dB/SSIM 0.9237 로 각각 + 0.6913dB/+ 0.0037 의 성능 향상을 달성하였다. 또한 Encoder1 의 불확실성 랭킹인 σ 에서 높은 값을 가진 부분들이 Decoder1 에서 실제 보정량의 증가로 이어짐을 시각적으로 확인할 수 있었다.

1. 서론

수중 이미지는 매체의 산란(scattering)과 흡수(absorption)로 인해 대비 저하, 색 왜곡, 탁도 증가가 동반되며, 이는 수중 데이터에 대한 인식·탐지 성능의 저하로 이어진다. 최근 Transformer 계열 복원 모델은 전역 속성을 활용해 품질을 개선했으나, 연산이 전역에 균일 분배되고 난이도(복원 불확실성)에 따른 선택적 집중이 부족하다는 한계가 있다. 특히 UDAformer[1]는 DATB(Dual Attention Transformer Block, DATB)/CSAT(Channel Self-Attention Transformer, CSAT) 블록을 조합하여 수중 이미지를 향상하는데 강점을 보이지만, 입력 조건의 변화나 지역적으로 어려운 경우에는(부유물, 조도 급변 등)에 대해서는 불필요한 연산 낭비나 혹은 과보정/미보정이 발생할 수 있다.

본 연구는 UDAformer[1]의 블록을 변경하지 않고, 각 스테이지 진입부에 불확실성 맵 σ 으로 가이드 되는 Sparse Sampling Attention pre-adapter 를 삽입하는 방법을 제안한다. 제안 모듈은 DATB/CSAT 구조 앞에서 입력 특징을 가볍게 정제하며, U-Head 가 추정한 불확

실성 맵을 UDR(Uncertainty Driven Ranking) 기반 attention 에 주입해 어려운 영역 중심의 계산을 유도한다. 동작 개요는 그림 1 에, 수식적 정의와 기호 설명은 식 (1) 및 3 장에서 제시한다.

불확실성 학습은 안정성과 해석 가능성을 함께 제공한다. 우리는 학습 단계에서 불확실성 맵을 UDL 로 감독하여 난이도가 높은 위치에서 값이 커지도록 유도하고, 추론 단계에서는 이 맵이 SSA 의 가이드로 작동해 연산을 어려운 영역에 집중시키도록 한다.

본 연구의 기여는 다음과 같다.

1. Pre-Adapter 설계: UDAformer[1]의 DATB/CSAT 는 그대로 유지한 채, Encoder/Decoder L1·L2 진입부 앞에 불확실성 유도 Sparse Sampling Attention 을 삽입하여 안정성·호환성을 확보하였다.
2. 불확실성 직접 감독(UDL): U-Head 가 예측한 σ 를 UDL[2]로 학습해 어려운 영역에 대한 가이드를 명시화하고, 추론 시에도 유효한 집중 효과를 제공한다.

3. 정량·정성 검증: UIEB-890[3]을 train = 800/ test = 90 으로 분할, 500 epoch 학습에서 베이스라인 (UDAformer[1]) PSNR 23.48dB/SSIM 0.9200 대비, 본 논문의 제안방법은 PSNR 24.1713dB/SSIM 0.9237 로 +0.6913dB/+0.0037 의 성능향상을 달성하였다. 또한 시각적인 결과물을 통해서 불확실성이 추가된 이미지와 이로 인해 해당 영역의 보정량이 증가함을 시각화 하여 확인하였다.

2. 관련 연구

2.1 수중 이미지 향상(UIE)과 UDAformer

수중 이미지 향상은 산란·흡수로 인한 대비 저하/색 왜곡을 복원하는 문제로, 물리식에 기반한 복원부터 최근의 딥러닝 기반 접근까지 폭넓게 연구되었다. UDAformer[1]는 DATB(Dual Attention Transformer Block) 속에 CSAT(Channel Self-Attention Transformer)와 SW-PSAT(Shifted Window Pixel Self-Attention Transformer)를 결합하여 채널 문맥을 파악하면서 Shifted Window 를 통해 전역 정보를 동시에 활용하는 구조로, 다양한 수중 조건에서 안정적 성능을 보여주었다. 다만 연산이 전역적으로 균일하게 분배된다는 특성 때문에, 어려운 영역(부유물, 급격한 조도 변화 등)에 집중적으로 자원을 배분하는 것에는 한계가 있다. 본 논문은 이 한계를 보완하기 위해, 원 구조(DATB/CSAT)는 보존하되 각 블록의 앞 부분에 불확실성-유도 Sparse Sampling Attention Pre-Adapter(SSA-UDR)를 삽입하는 방법을 제안하고자 한다.

2.2 이미지 복원을 위한 Transformer

최근 Transformer 계열 복원 모델은 전역 문맥 모델링과 윈도우 주의로 뛰어난 성능을 보여왔다. IPT[5]는 다양한 low-level 과제들을 아우르는 초기 Transformer 기반 복원 모델이고, SwinIR[6]은 shifted window 구조로 Super Resolution/Denoising/JPEG 복원에서 강력한 성능을 보였다. Uformer[7]는 U-shape 와 local window attention 을 결합해 효율과 성능의 균형을 잡았고, Restormer[8]는 channel attention 강화로 고해상도 복원에 특화되었다. 이러한 모델들은 주로 균일한 attention 전개를 가정하며, 난이도에 따른 선택적 계산이나 불확실성과의 직접적인 결합은 상대적으로 부족하다. 본 논문은 이 지점을 파고들어, σ (불확실성)를 SSA+UDR 모듈에 주입하여 어려운 영역 중심의 Sparse Attention 을 실현한다.

2.3 불확실성(σ) 모델링과 UDL

딥러닝 복원에서의 불확실성 모델링은 음의 로그우도(Negative Log-Likelihood) 형태의 손실로 σ 를 직접 학습하는 방식이 일반적이다[2]. 본 연구에서도 이를 UDL(Uncertainty-Driven Loss)로 채택하여, 오차가 큰 픽셀에서 σ 가 커지도록 σ 를 직접 감독한다. 이렇게 학습된 σ 는 추론 단계에서도 SSA+UDR 의 가이드 신

호로 재사용된다. 결과적으로 σ 는 해석 가능성과 연산 집중을 동시에 제공한다.

2.4 Sparse Sampling Attention 과 불확실성 기반 랭킹

복원에서의 SSA 는 중요 영역에 계산을 집중시켜 효율을 높이는 방향으로 발전해 왔다. UDR-S2Former[4]는 UDR 을 사용하여 토큰 선택 및 attention 을 유도해 희소 샘플링을 구현했고, Deraining 과제에서 효과를 보였다. 본 연구는 UDR 의 핵심 아이디어를 UIE 도메인에 pre-adapter 형태로 이식하되, 원 블록(DATB/CSAT)을 대체하지 않고 앞 부분에 얇게 결합한다는 점에서 구조적 차별성을 갖는다. 또한 σ 를 UDL[2]로 직접 감독하여, 추론 시에도 어려운 영역에 대한 attention 이 유지되도록 설계하였다.

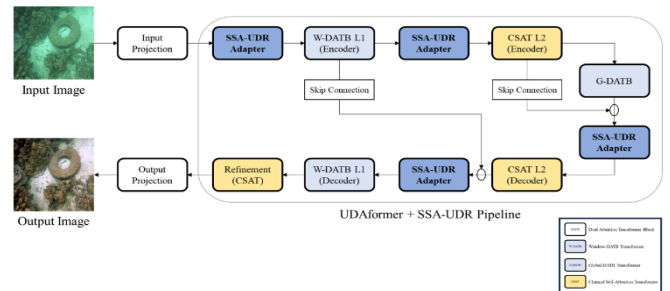
3. 제안 방법

3.1 전체 파이프라인 개요

본 연구의 핵심은 UDAformer[1]의 원 구조 (DATB/CSAT)를 보존하면서 각 스테이지 진입부에 SSA-UDR 로 얻은 불확실성을 얇은 잔차(residual) 형태로 삽입하는 것이다. 입력 특징 x 에 대해, pre-adapter 는 U-Head 가 추정한 불확실성 맵 σ 를 UDR(Uncertainty-Driven Ranking)기반 Sparse Sampling Attention 에 주입해 어려운 영역 중심의 보정 y 를 산출하고, 이를 스칼라 계수 α 로 축소해 원 특징에 더한다:

$$x_{pre} = x + \alpha \cdot UDR(x, \sigma) \quad (1)$$

여기서 $x \in \mathbb{R}^{B \times C \times H \times W}$ 는 스테이지 첫 블록으로 입력되는 특징이고, x_{pre} 는 pre-adapter 출력을 잔차 형태로 합성한 특징이다. $\alpha > 0$ 는 pre-adapter 출력을 주입하는 잔차 계수이며, $\sigma \in \mathbb{R}^{B \times 1 \times H \times W}$ 는 U-Head 가 예측하는 불확실성 맵이다. $UDR(\cdot)$ 은 σ 를 가이드로 사용하는 sparse attention 모듈로, 난이도가 높은 위치에 더 큰 보정이 이루어지도록 동작한다. 여기서 B 는 배치 크기, C 는 채널 수, H, W 는 공간 해상도를 의미한다.” 이렇게 얻은 x_{pre} 가 각 스테이지의 첫 DATB/CSAT 블록에 입력된다. α 는 작게 설정하여 원 경로의 표현을 보존하면서 제한적으로 개입하도록 하였다. 전체 파이프라인을 도식화하면 그림 1 과 같다.



(그림 1) UDAformer[1] + SSA-UDR 파이프라인

3.2 SSA-UDR Pre-Adapter

U-HEAD (Uncertainty Head)

U-Head 는 1×1 컨볼루션 스택으로 구성된 경량 모듈로, 입력 $x \in \mathbb{R}^{B \times C \times H \times W}$ 에서 불확실성 맵 $\sigma \in \mathbb{R}^{B \times 1 \times H \times W}$ 을 추정한다. 마지막 활성 함수는 Softplus를 사용해 $\sigma > 0$ 을 보장한다.

UDR 기반 Sparse Sampling Attention

UDR 모듈은 윈도우 크기 $W_s = 8$ 인 윈도우 기반 self-attention 내부에서 σ 를 토큰 선택 및 가중의 가이드로 사용한다. 구체적으로, 각 윈도우에서 불확실성 랭킹, 즉 σ 큰 순으로 상위 토큰에 주의를 더 강하게 배분하거나 후보를 희소화 하여, 어려운 영역에 주의를 집중시킨다.

잔차 주입(Residual injection)

UDR 출력 $y = UDR(x, \sigma)$ 는 스칼라 α 로 스케일링 되어 x 에 더해진다(식 (1)). α 는 학습하지 않고 고정하며, 스테이지 및 해상도에 따라 소폭 차등 설정하여 안정성을 확보한다. 본 실험에서는 $L1 \alpha = 0.025, L2 \alpha = 0.0175$ 로 설정하였다. 이를 통해 기존 UDAformer[1] 경로는 그대로 유지되고, pre-adapter 는 입력 정제를 위한 전처리용으로 동작한다.

3.3 손실 함수

최종 손실은 복원 손실 과 UDL 의 합으로 구성한다.

$$\mathcal{L} = \|I_{pred} - I_{gt}\|_2^2 + \lambda_u^{(L1)} \mathcal{L}_{UDL}^{(L1)} + \lambda_u^{(L2)} \mathcal{L}_{UDL}^{(L2)} \quad (2)$$

I_{pred} 는 네트워크의 복원 결과 영상이고 I_{gt} 는 정답 영상이다. $\lambda_u^{(L1)}$ 과 $\lambda_u^{(L2)}$ 는 각각 L1 스테이지와 L2 스테이지에서의 불확실성 손실 가중치이며, $\mathcal{L}_{UDL}^{(L1)}$ 과 $\mathcal{L}_{UDL}^{(L2)}$ 은 각 스테이지의 불확실성 맵을 사용해 계산한 UDL 항을 뜻한다. 본 실험에서는 $\lambda_u^{(L1)} = 0.30$, $\lambda_u^{(L2)} = 0.21$ 을 사용하였다.

UDL 은 아래와 같이 정의한다.

$$\mathcal{L}_{UDL} = \frac{1}{2} e^{-s} err^2 + \frac{1}{2} s, err = I_{pred} - I_{gt} \quad (3)$$

$s = \log \sigma^2$ 로 예측변수이고 $err = I_{pred} - I_{gt}$ 는 복원 오차이다. σ 는 양수이며, 오차가 큰 위치에서 σ 가 증가하도록 학습되기 때문에 불확실성 맵은 어려운 영역을 지칭하도록 정렬된다. 이렇게 정렬된 σ 는 추론 단계에서 UDR 의 토큰 선택과 가중을 이끌어 난이도가 높은 영역에 주의를 집중시키는 역할을 한다.

3.4 스테이지 구성 및 구현 세부

Pre-adapter 는 Encoder/Decoder L1 · L2 에 배치한다. L1 은 다운샘플 이전(Enc1)과 업샘플 직후(Dec1)의 고 해상도 특징을 다루므로 작은 α 로도 효과가 나타난다. L2 는 skip-connection 합류 지점(CSAT/G-DATB) 직전/직 후에 두어, 전역 문맥과 상호작용하기 전에 어려운

영역 정제를 선행한다. 학습은 500 epoch 로 진행하며, 추론 시에도 학습과 동일하게 σ 를 UDR 에 주입하고 추가 테스트-타임 튜닝은 적용하지 않는다.

4. 실험

4.1 데이터

UIEB-890[3] 데이터셋을 train = 800/test = 90 으로 나누어 학습 및 검증에 사용하였다. UDAformer[1]의 저자가 공개한 train/test 목록과 동일하게 설정하였고 batch = 4, epoch = 500 을 사용하였다.

4.2 평가 지표 및 비교 대상

정량 지표로 PSNR 과 SSIM 을 사용하였다. 비교 대상은 원 구조를 변경하지 않은 UDAformer (Baseline)[1] 와, 본 논문에서 제안하는 UDAformer + SSA-UDR 이다. 두 방법은 동일한 데이터 분할, 학습 스케줄, 평가 프로토콜을 공유한다.

Methods	Test-UIEB890	
	PSNR ↑	SSIM ↑
Water-Net[3]	20.28	0.85
Ucolor[9]	21.29	0.88
U-shape[10]	20.77	0.83
UDAformer[1]	23.48	0.92
Ours	24.1713	0.9237

<표 1> 기존 UIE 방법들과 성능 비교

4.3 실험 결과

정량 비교 결과는 표 1 을 통해 요약하여 제시하였다. UDAformer(Baseline)[1]는 PSNR 23.48dB/SSIM 0.92 를 기록하였고, 본 논문의 제안법은 PSNR 24.1713dB/SSIM 0.9237 을 달성하였다. 이는 기존 대비 +0.6913dB/+0.0037 로 유의미하게 향상되었다는 것을 확인하였다.

Methods	Test-UIEB890	
	PSNR ↑	SSIM ↑
Replace: SW-PSAT → SSA-UDR	21.2743	0.8727
Ours w/o UDL	23.9141	0.9230
Ours	24.1713	0.9237

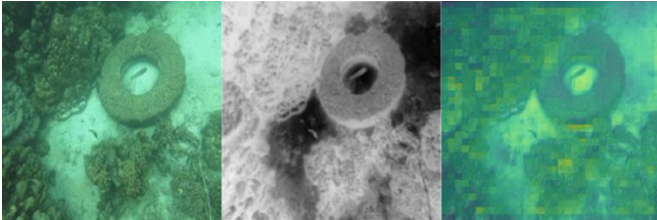
<표 2> Ablation study on UIEB-Test

표 2 는 본 논문의 제거 연구 결과를 보여준다. UDAformer[1] DATB 의 두 Transformer 중 SW-PSAT 을 SSA-UDR 로 교체한 Replace 의 경우는 기존 모델의 성능보다 오히려 낮게 나왔다. 본 논문의 제안법에서 UDL 을 비활성화하면 제안법의 성능보다 수치가 떨어지는 결과를 얻었다.



(그림 2) 원본 수중 이미지(좌) 이미지 향상 결과(우)

그림 2는 수중 이미지가 입력되었을 때 논문의 제안법에 따라 이미지를 향상한 결과이다. 채도 편차 완화, 명암 대비 회복, 디테일 선명도가 많이 개선된 것을 확인할 수 있다. 특히 배경 대비가 약한 영역이나 녹색/청색 캐스트가 심한 영역에서 색 균형이 안정적으로 복원되었다.



(그림 3) 원본(좌) 불확실성 맵(중앙) SSA-UDR 시각화(우)

4.5 불확실성 유도과 보정 집중에 대한 해석

그림 3은 원본과 함께 불확실성 맵(σ), 그리고 SSA-UDR 보정 히트맵(Δ) 오버레이를 제시한다. σ 는 U-Head가 예측한 위치별 복원 난이도를 나타내며 흰 부분이 높은 값으로 표현된다. Δ 는 모델 파이프라인을 거쳐 최종 Decoder L1 단계에서 SSA-UDR이 실제로 가한 보정의 크기를 공간적으로 나타낸다. 시각화 결과, σ 가 높은 영역에서 Δ 가 함께 증가하는 경향이 뚜렷하게 관찰되며, 이는 학습 단계에서의 불확실성 감득(UDL)과 추론 단계에서의 SSA-UDR이 동일한 공간적 초점을 공유한다는 것을 뒷받침한다. 다시 말해, 본 논문의 제안법은 어려운 영역에 보정을 가하여 attention을 선별적으로 배분하도록 설계·학습되었고, 실제 복원 과정에서도 그 의도가 구현된다.

5. 결론 및 향후과제

본 논문은 UDAformer[1]의 원 구조(DATB/CSAT)를 변경하지 않고, 각 스테이지 앞 부분(Encoder/Decoder L1-L2)에 불확실성(σ)으로 유도된 SSA-UDR Pre-Adapter를 잔차(α)로 삽입하는 간단하면서도 효과적인 방법을 제안하였다. σ 는 경량 U-Head로 예측되고 UDL[2]로 직접 감득되어, 추론 단계에서 어려운 영역 중심의 sparse sampling attention을 자연스럽게 유도할 수 있다. UIEB-890[3]에서 500 epoch 학습 결과, UDAformer[1] 대비 PSNR +0.6913dB/SSIM +0.0037의 향상을 확인했으며, σ - Δ 시각화로 불확실성 고영역이 실제 보정 증가로 이어진다는 것을 확인하였다.

향후에는 LSUI(Large Scale Underwater Image Dataset)[10], Seathru[11]와 같은 다중 데이터셋으로 확장하고 멀티 스케일에서의 개별 σ 측정, 학습 데이터에 따른 적응형 $\alpha \cdot \lambda_u$ 를 도입하여 성능과 효율을 끌어올릴 계획이다.

참고문헌

- [1] Z. Shen et al., "UDAformer: Underwater image enhancement based on dual attention transformer," *Computers & Graphics*, 111(C), pp. 77–88, 2023.
- [2] A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?," in *NeurIPS*, Long Beach, USA, 2017.
- [3] C. Li et al., "An Underwater Image Enhancement Benchmark Dataset and Beyond," *IEEE Trans. Image Process.*, 29, 2020.
- [4] S. Chen et al., "Sparse Sampling Transformer with Uncertainty-Driven Ranking for Unified Removal of Raindrops and Rain Streaks," in *ICCV*, Paris, France, 2023.
- [5] H. Chen et al., "Pre-Trained Image Processing Transformer," in *CVPR*, Nashville, USA, 2021, pp. 12299–12310.
- [6] J. Liang et al., "SwinIR: Image Restoration Using Swin Transformer," in *ICCV*, Montreal, Canada, 2021, pp. 1833–1844.
- [7] Z. Wang et al., "Uformer: A General U-Shaped Transformer for Image Restoration," in *CVPR*, New Orleans, USA, 2022.
- [8] S. W. Zamir et al., "Restormer: Efficient Transformer for High-Resolution Image Restoration," in *CVPR (Oral)*, New Orleans, USA, 2022.
- [9] C. Li et al., "Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding," *IEEE Trans. Image Process.*, 30, pp. 4985–5000, 2021.
- [10] L. Peng et al., "U-shape Transformer for Underwater Image Enhancement," *IEEE Trans. Image Process.*, 32, pp. 3066–3079, 2023.
- [11] D. Akkaynak and T. Treibitz, "Sea-Thru: A Method for Removing Water From Underwater Images," in *CVPR*, Long Beach, USA, 2019.