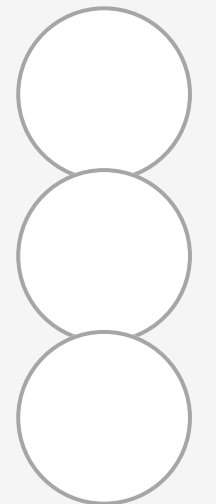


A MULTI-STAGE PROGRESSIVE NETWORK WITH FEATURE TRANSMISSION AND FUSION FOR MARINE SNOW REMOVAL

Lixin Liu, Yuyang Liao, Bo He, MDPI 2024



박정규

Deep Shark Lab 학부연구생

Contents

1

Introduction

2

Proposed
Method

3

Key Modules

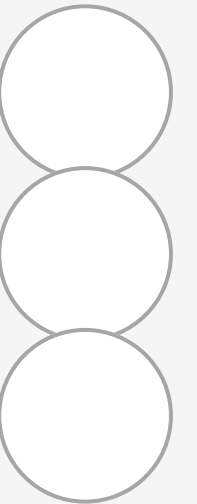
4

Experiments

5

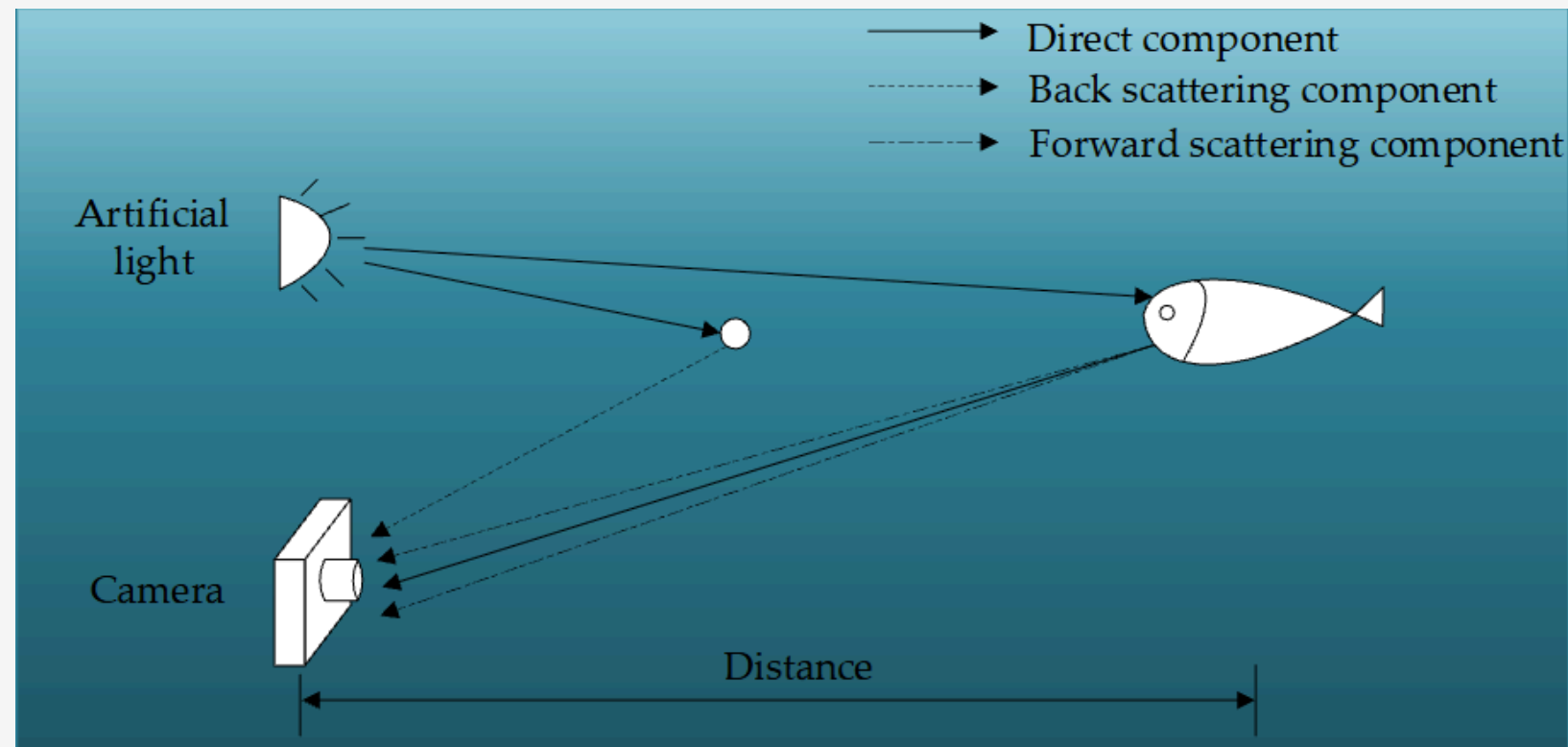
Conclusion

INTRODUCTION



Introduction

Research Background



Underwater Imaging

- 해양 탐사, 수중 관측 및 객체 인식에 활용된다.
- 정확한 분석을 위해 높은 영상 품질을 요구한다.

Image Degradation

- 빛의 흡수와 산란
- 색상 왜곡 및 선명도 저하
- 부유 입자에 의한 밝은 노이즈 발생

Impact on Vision Task

- 수중 장면의 시각적 인지 방해
- 물체의 특징과 경계 정보 왜곡
- 객체 탐지 및 세그멘테이션 성능 저하

Introduction

Marine Snow Definition



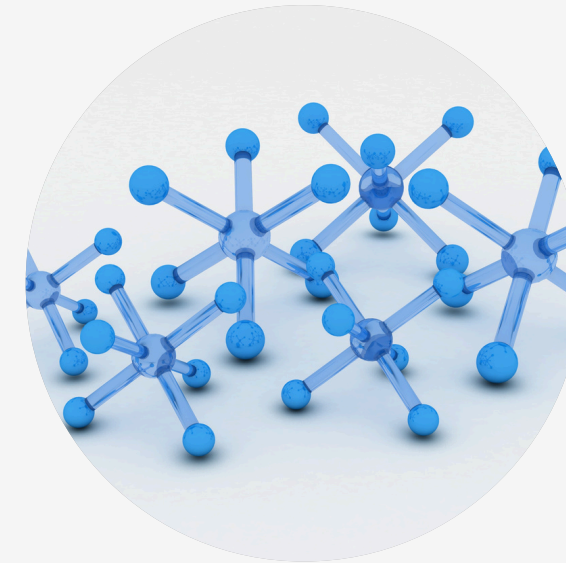
Definition

해양에 존재하며 해저로 가라앉는
미세 입자



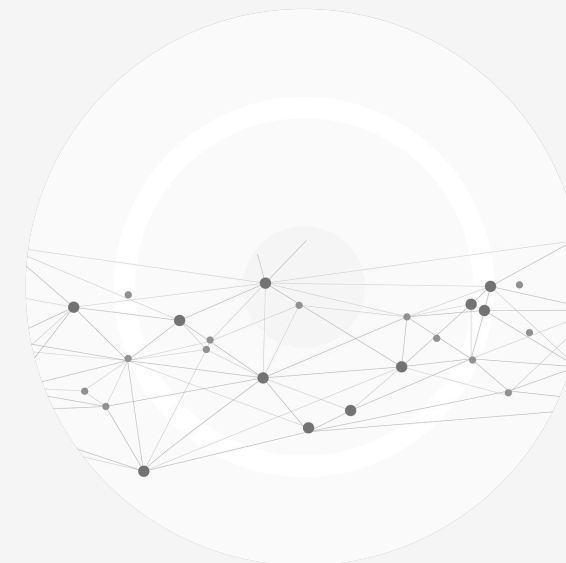
Characteristics

다양한 크기와 형태를 가지며 입자
마다 투명도가 다름



Composition

해양 생물의 잔해, 부유성 배설물,
부유 퇴적물, 기타 무기 물질



Formation

인공 조명이 수중의 부유물 입자에
의해 산란되고, 산란된 빛이 카메
라에 유입되어 이미지에서 밝은 흰
색 점으로 나타남

Introduction

Problem Definition

Problem Definition

Image Degradation

부유물에 의해 수중 이미지의 품질
과 장면 인지에 영향을 미침

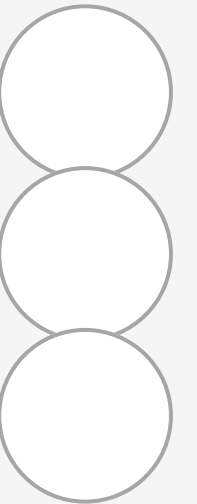
Misrecognition

밝기가 높은 부유물은 실제 특징으
로 잘못 인식될 수 있음

Downstream Tasks

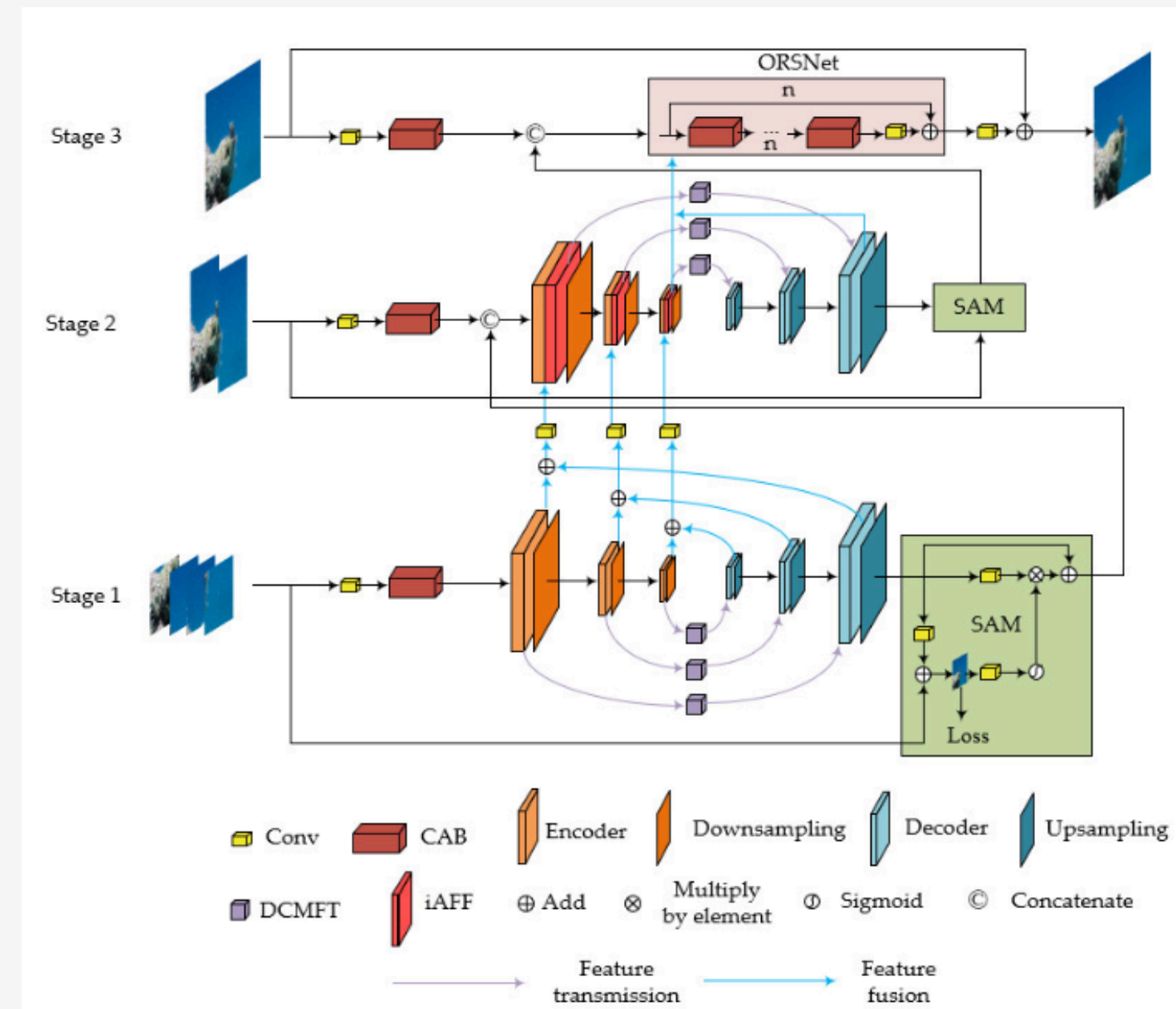
Computer Vision 작업의 성능이 저
하될 수 있음

PROPOSED METHOD



Proposed Method

Network Overview



Network Overview

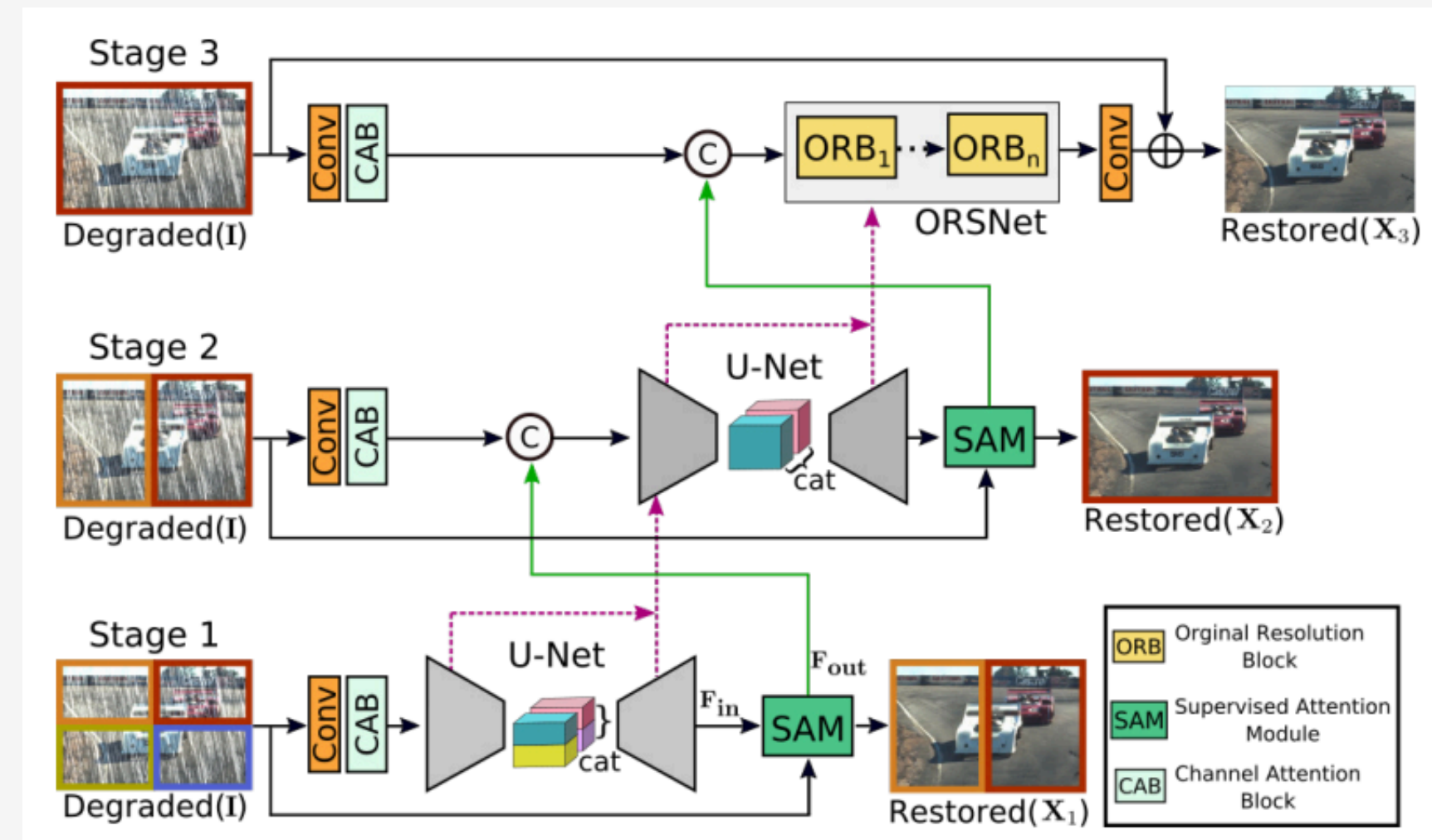
- 기존 MPRNet의 기본 구조를 기반으로 설계
- Marine Snow 제거를 위한 3단계 점진적 복원 네트워크
- 각 단계가 복원 결과와 특징 정보를 다음 단계로 전달

Feature Processing

- Stage 1과 Stage 2 에서 U자형 Encoder-Decoder 구조 사용
- 각 단계에서 추출한 특징을 다음 단계로 전달
- Stage 3 에서는 ORSNet을 사용해 최종 이미지의 세부 정보 보존

Proposed Method

What is MPRNet



MPRNet

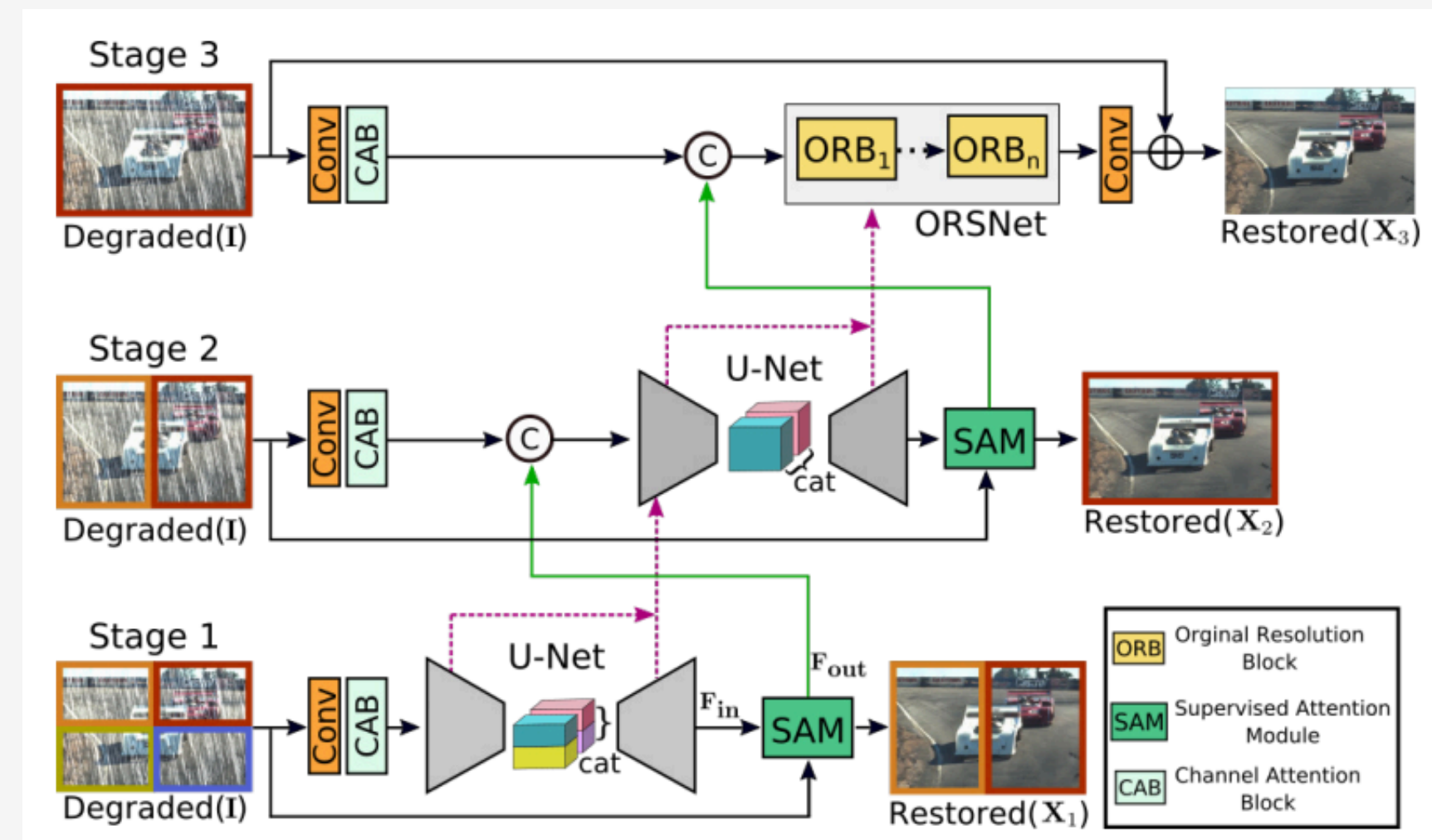
- 노이즈 제거, 빗줄기 제거, 흔들림 복원처럼 열화된 이미지를 깨끗한 이미지로 복원하는 작업을 3개의 Stage를 거치면서 점진적으로 복원시키도록 제안된 모델

Why Multi-Stage?

- 이미지 복원에서는 Contextual Information과 Spatial Details 두 종류의 정보가 모두 필요하다
- Stage 1, 2에서 U-Net 구조를 통해 문맥 정보를 학습
- Stage 3에서 원본 해상도의 전체적인 일관성과 세부 정보를 복원

Proposed Method

MPRNet Stage 1

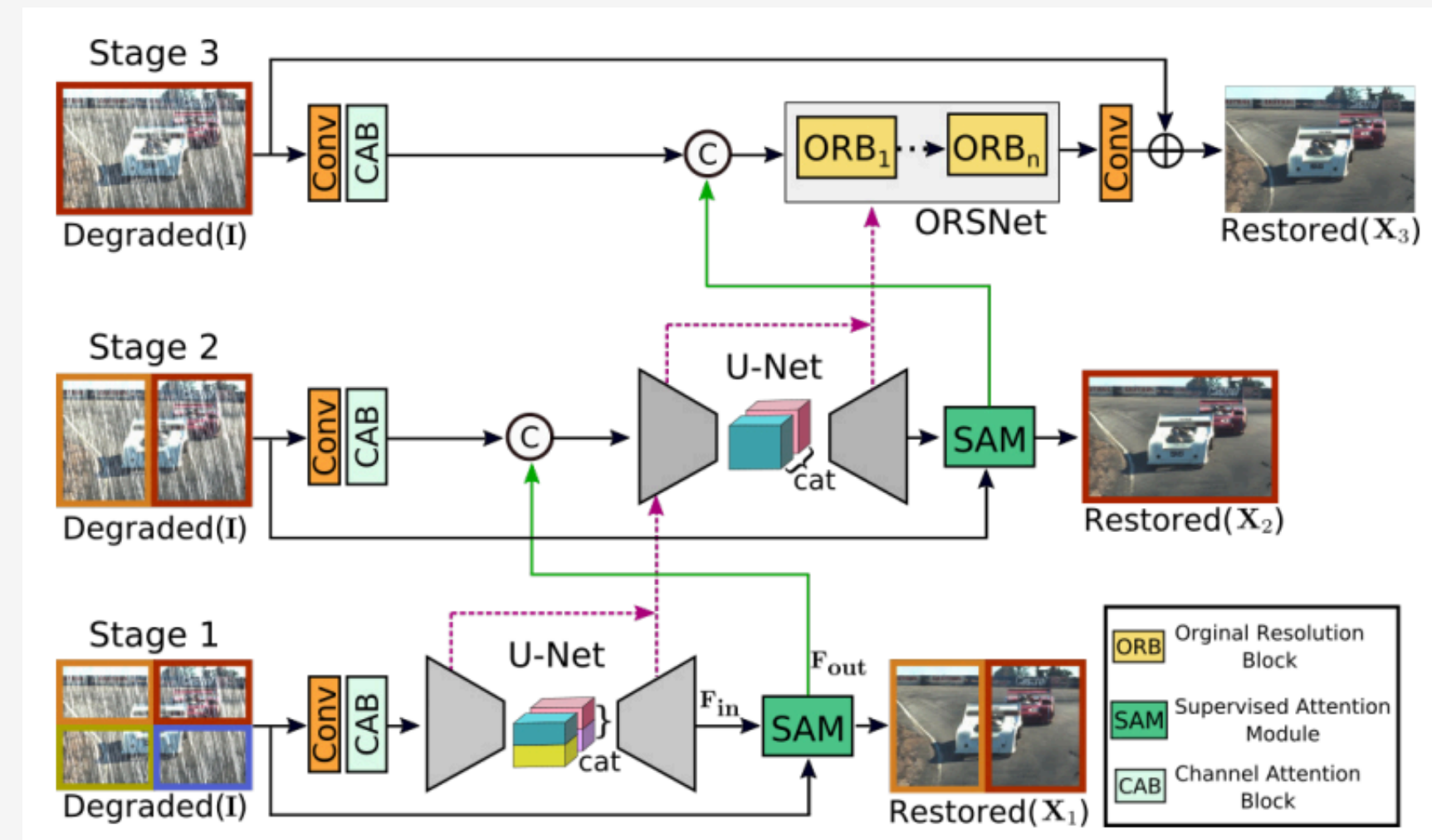


Stage 1

- 입력 이미지를 4개의 패치로 분할 후 각 패치를 Conv를 통과하여 $3 \times H \times W$ 이미지를 $C \times H \times W$ 로 변환 한다.
- CAB (Channel Attention Block)을 통과하여 각 채널 중 중요한 채널은 강조하고, 중요하지 않은 채널은 약화시킨다.
- Feature Channel은 U-Net을 지나 복원된 Feature Channel F_{in} 으로 출력된다.
- F_{in} 은 다시 SAM으로 입력되어 복원된 이미지 X_1 과 다음 Stage에 전달할 정제된 특징 F_{out} 을 출력한다.

Proposed Method

Convolution

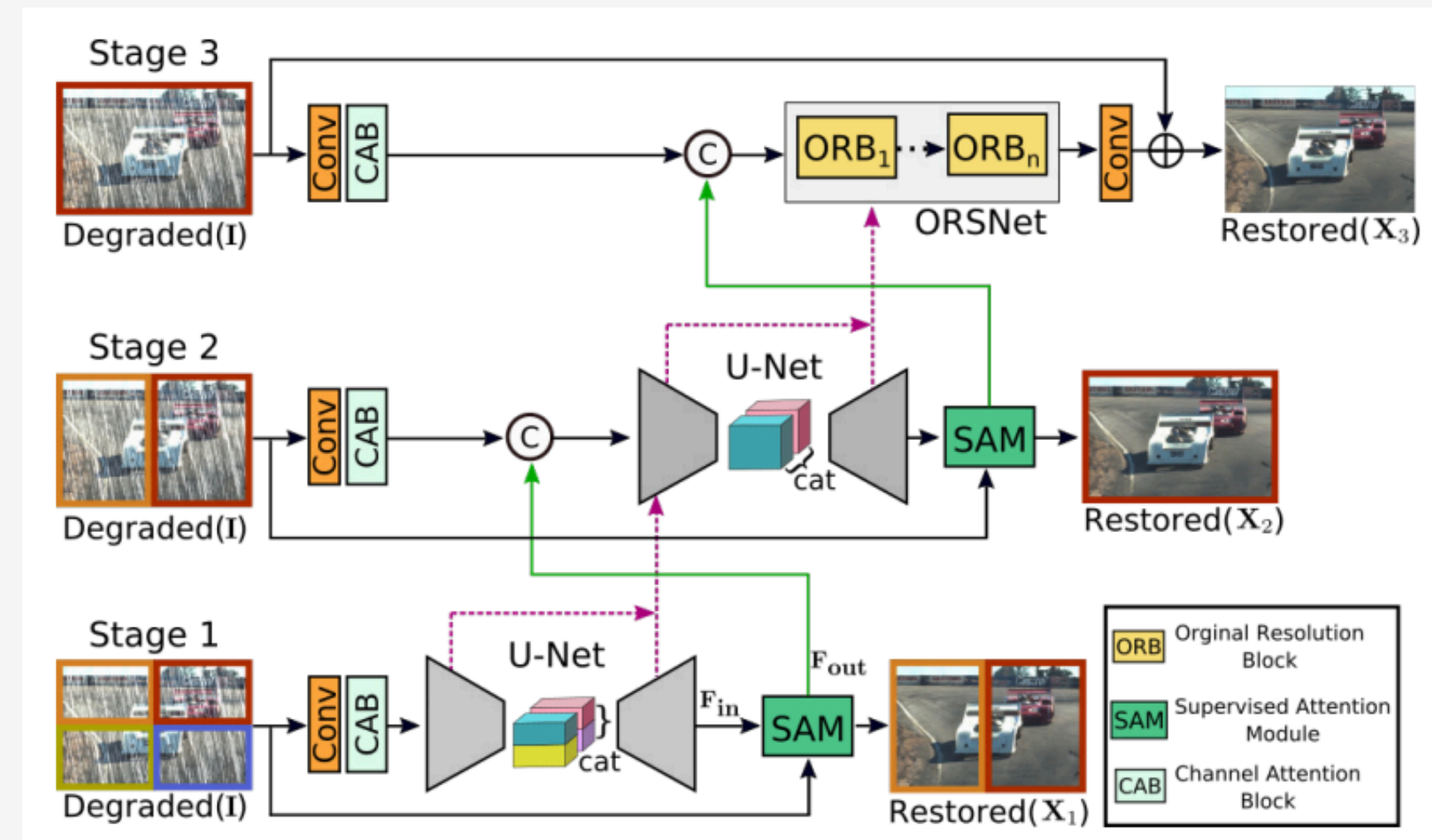


Convolution

- RGB 3채널의 Image를 그대로 U-Net에 입력으로 사용하지 않고, C개의 feature Channel로 변환하여 U-Net에 입력으로 사용
- 논문에서는 패치당 Channel 수를 40개로 나누었고, 단순히 RGB로써 패턴을 학습하지 않고 40개의 관점으로 노이즈를 바라보기 위해 사용

Proposed Method

Channel Attention Block



CAB

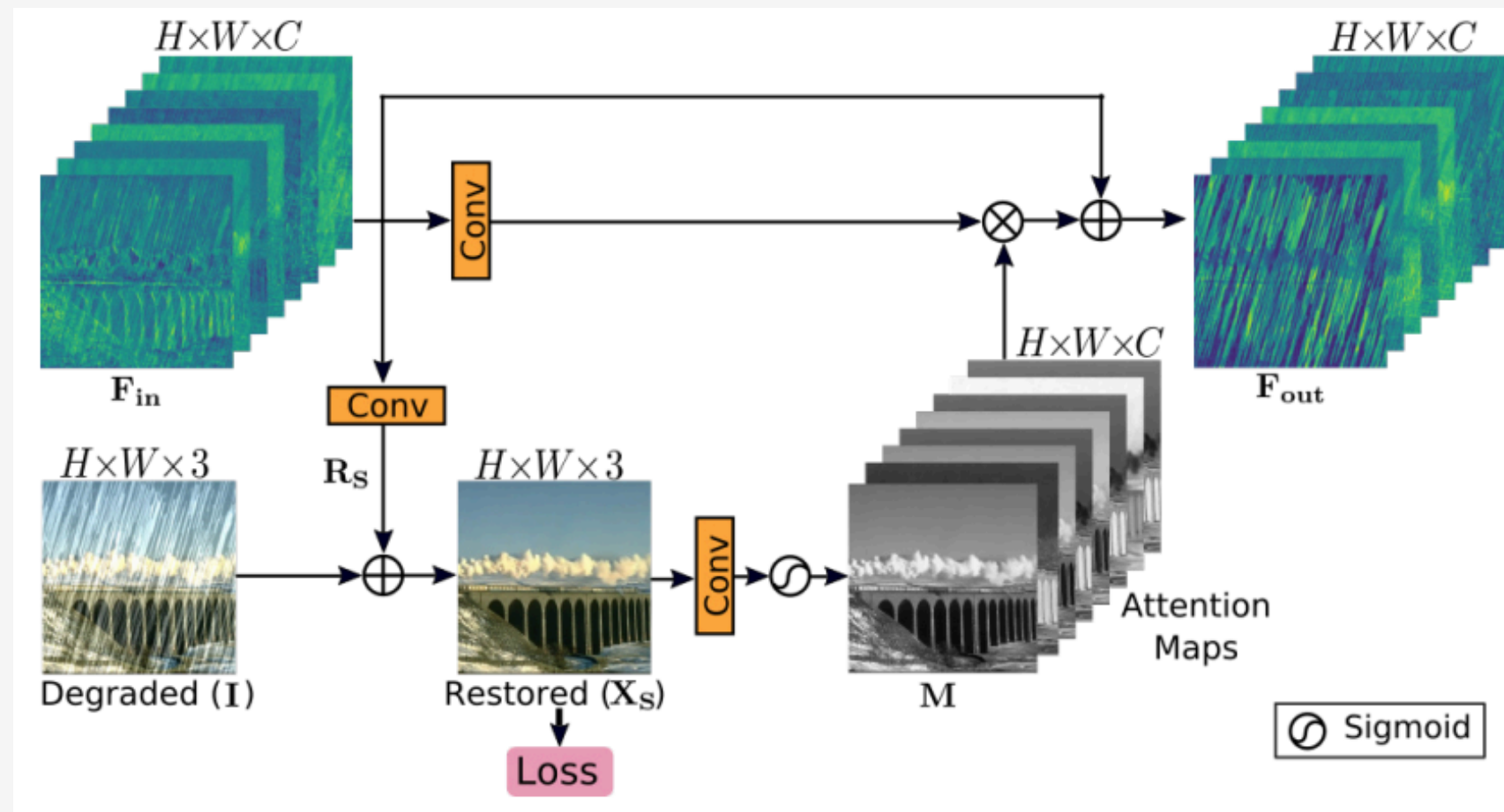
- 추출된 Feature Channel들중 어떤 Channel이 중요한지 구분하는 역할

$$[F_1, F_2, \dots, F_C] \xrightarrow{\text{CAB}} [w_1 F_1, w_2 F_2, \dots, w_C F_C]$$

- 구분된 $C \times H \times W$ 사이즈의 Feature Channel을 U-Net에 입력하고, downsampling 과정을 통해 이미지의 넓은 문맥을 파악한다. 그 후 디코더를 통과해 복원에 필요한 Feature Channel F_{in} 을 생성한다.

Proposed Method

Supervised attention module

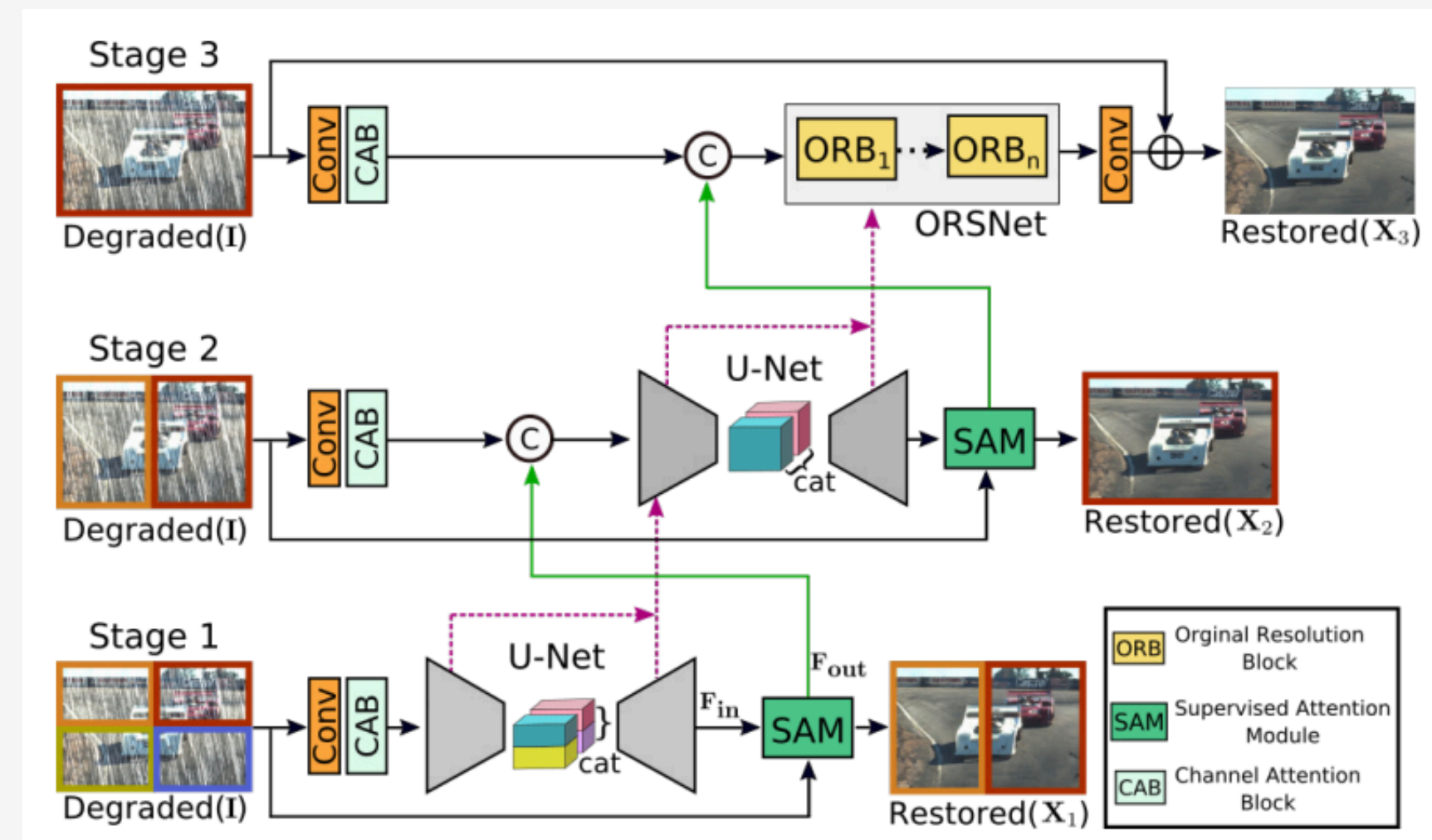


SAM

- U-Net에서 추출된 F_{in} 을 RGB 채널로 Conv를 진행해 보정값 R_s 를 계산하여 원본이미지에 더해 복원된 이미지 X 를 생성한다.
- 복원 이미지는 Ground Truth와 손실을 계산하여 각 stage의 역할에 맞는 복원을 하도록 유도한다.
- 복원 이미지를 이용해 다시 C개의 Attention Map을 생성하고 다음 Stage에 최종 전달할 Feature Channel을 계산하여 F_{out} 을 출력한다.

Proposed Method

MPRNet Stage 2

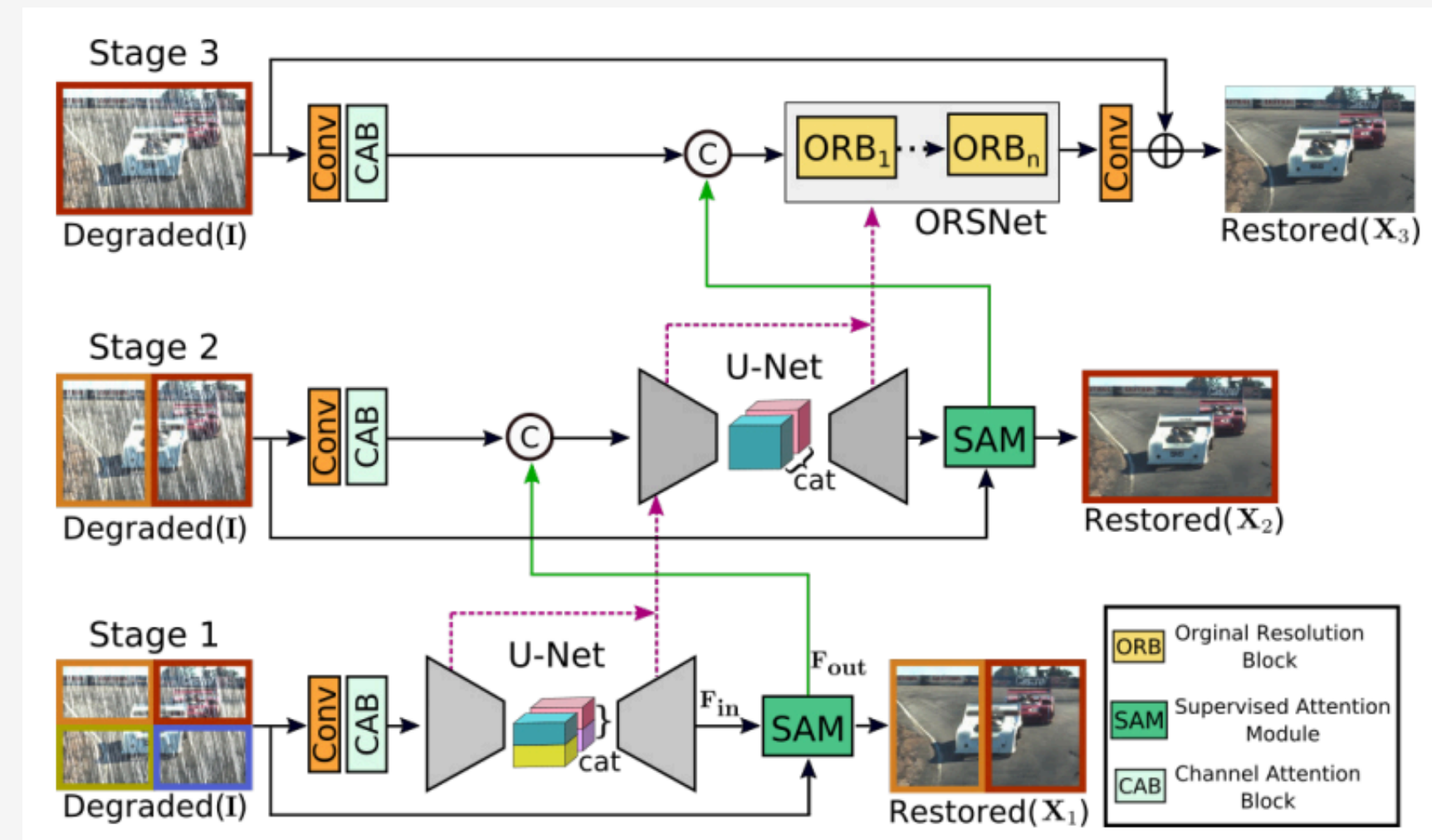


Stage 2

- 입력 이미지를 2개의 패치로 분할 후 각 패치를 Conv를 통과하여 $3 \times H \times W$ 이미지를 $C \times H \times W$ 로 변환 한다.
- Stage 1 보다 이미지를 조금 더 넓은 문맥에서 복원하는 단계
- Stage 1에서 정제된 국소적 Feature Channel F_{out} 과 Stage 2에서 추출된 Feature Channel을 Concat하여 U-Net에 입력
- U-Net의 Encoder에 Stage 1의 U-Net Feature Channel을 이용하여 Stage 2에서도 Stage 1의 정보를 이용할 수 있도록 유도

Proposed Method

MPRNet Stage 3

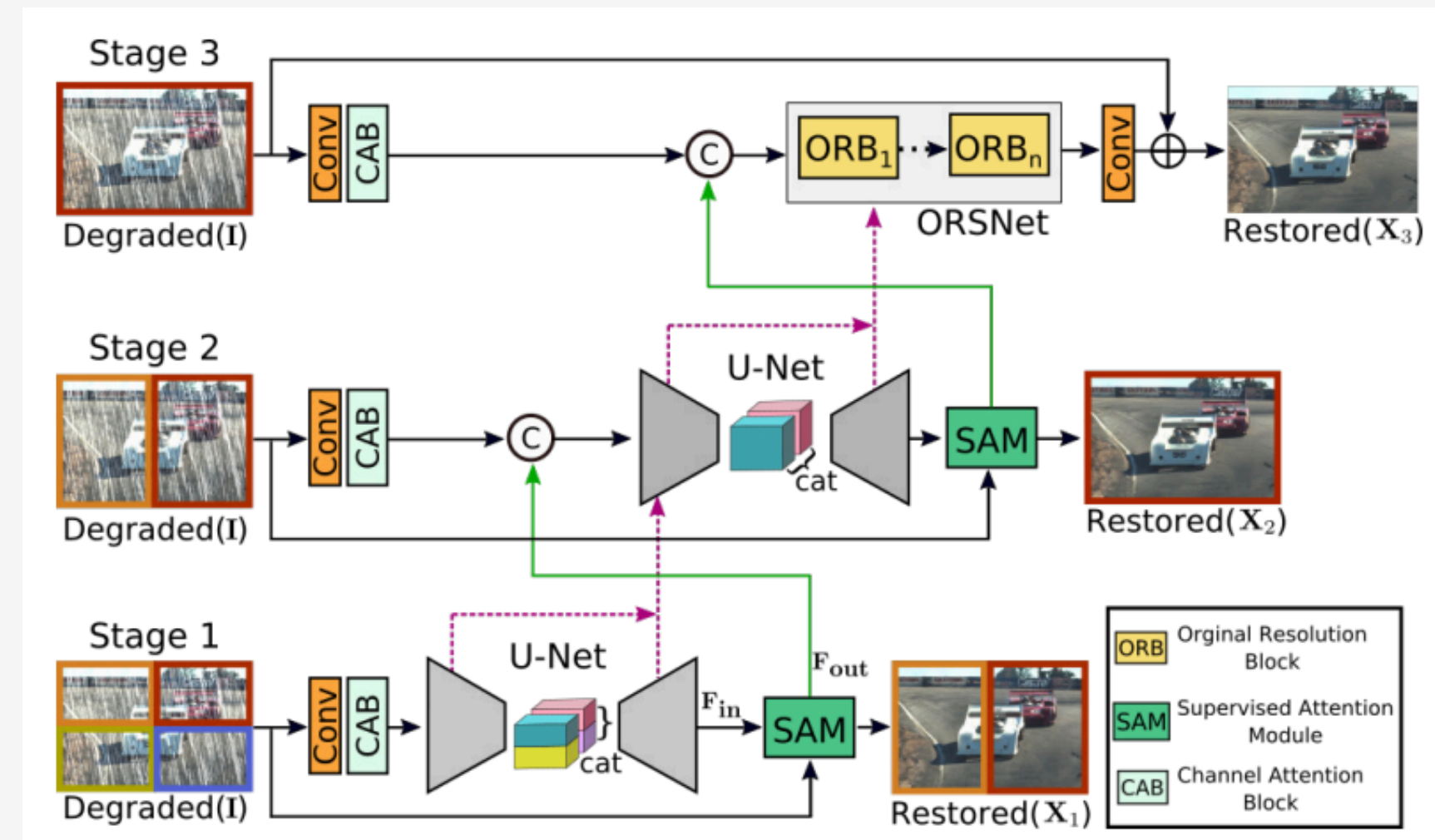


Stage 3

- Stage 1, 2에서 학습한 넓은 문맥 정보를 활용하면서, 원본 해상도의 공간적 세부 정보를 보존해 최종 이미지를 복원하는 단계
- 최종적으로 이미지를 생성하기 위해 ORSNet을 사용

Proposed Method

ORSNet



ORSNet

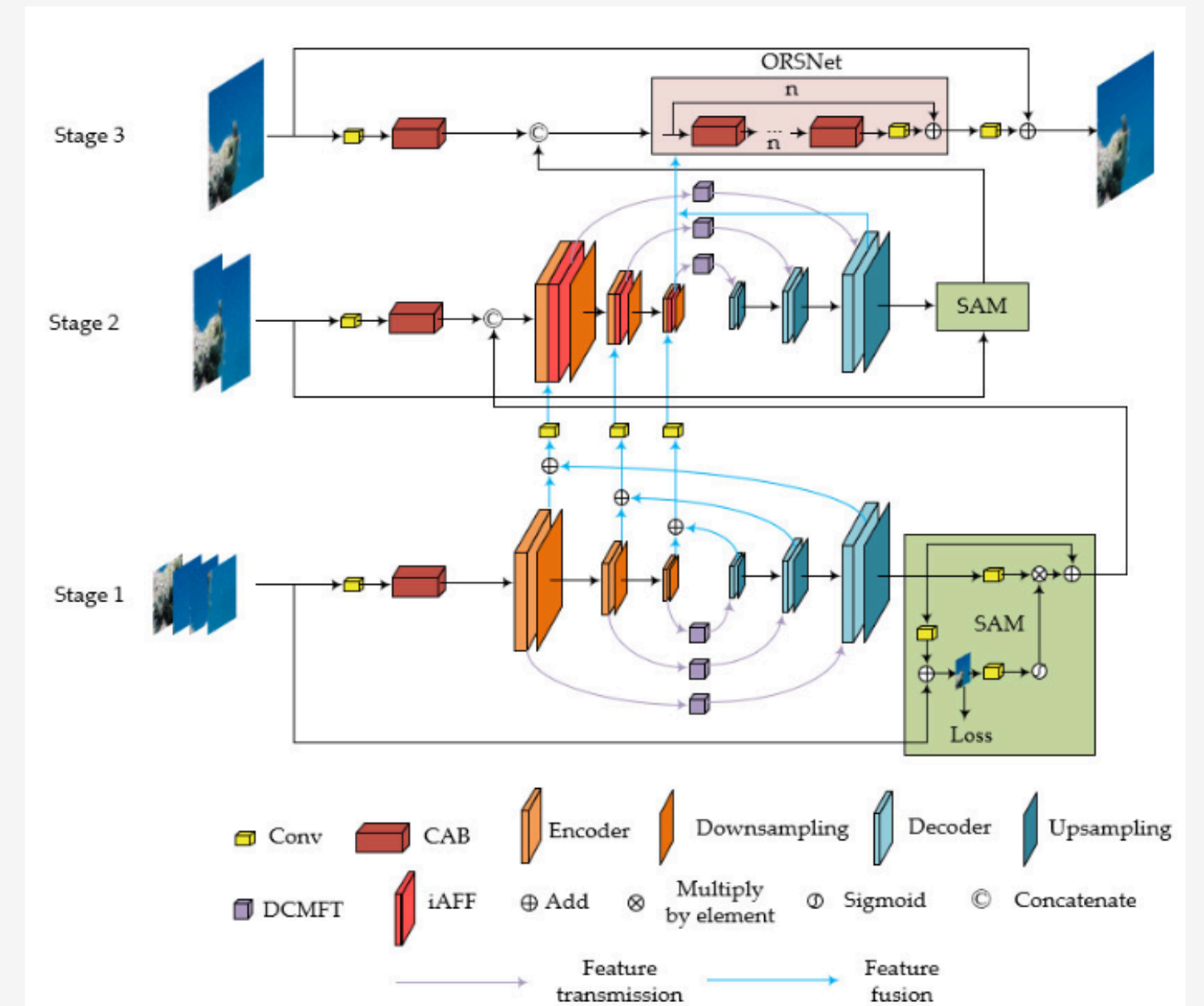
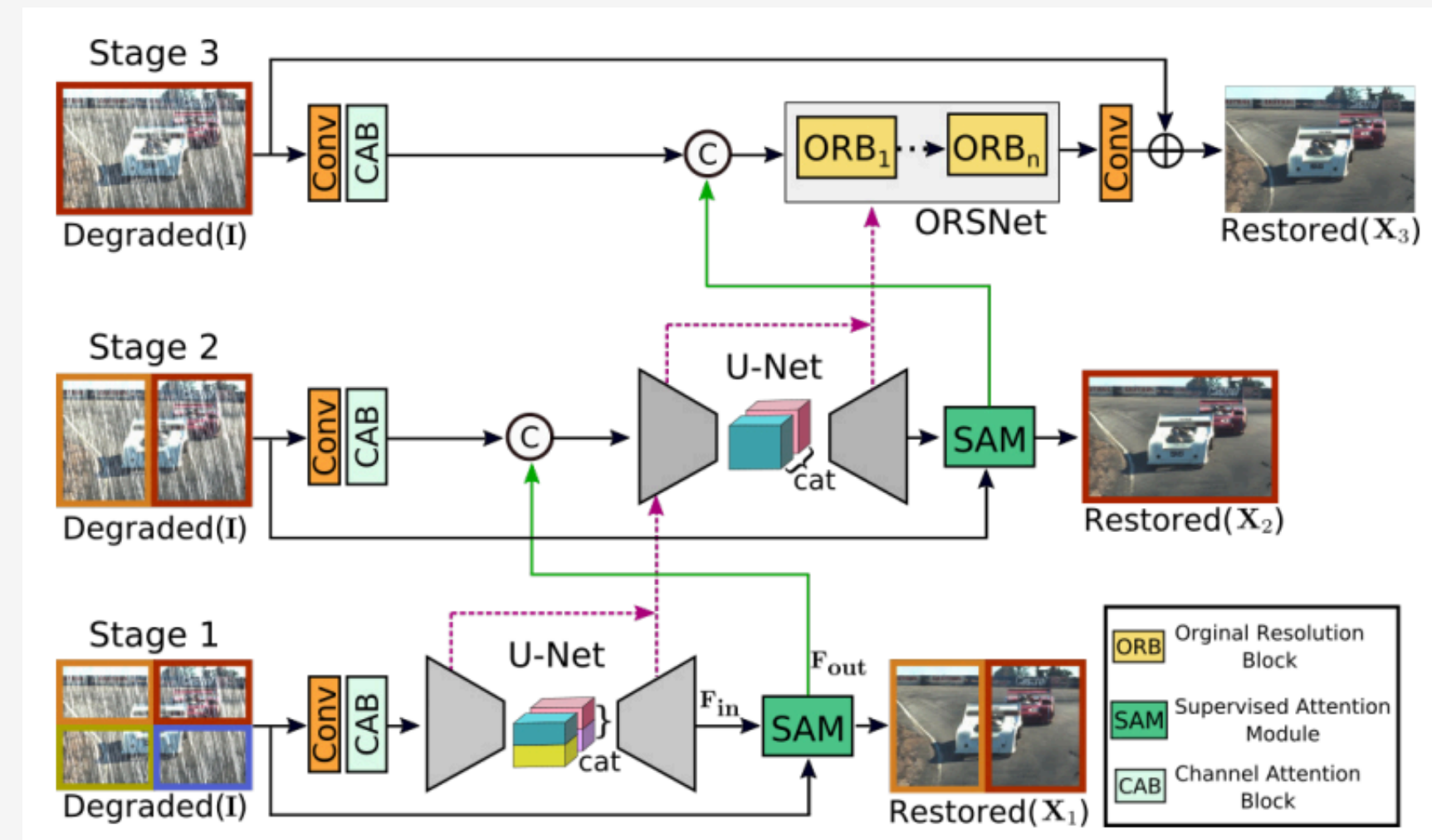
- 여러개의 ORB (Original Resolution Block)으로 구성된 네트워크

ORB

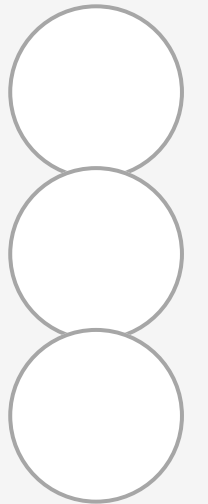
- 여러개의 CAB이 연속적으로 배치된 블록
- 입력으로 들어온 최종 Feature Channel들을 반복적으로 attention하여 정제하고, 최종 보정된 Feature Channel을 생성한다
- 이후 Convolution을 적용하여 RGB 3채널 residual image를 생성하고, 이를 원본과 더해 최종복원 이미지를 생성한다.

Proposed Method

Network Comparison



KEY MODULES



Key Modules

Overview

DCMFT

기존 U-Net의 Skip connection은 정보 전송 구조가 단순하여 정보 전송 과정에서 손실이 발생하기 때문에 DCMFT 사용

iAFF

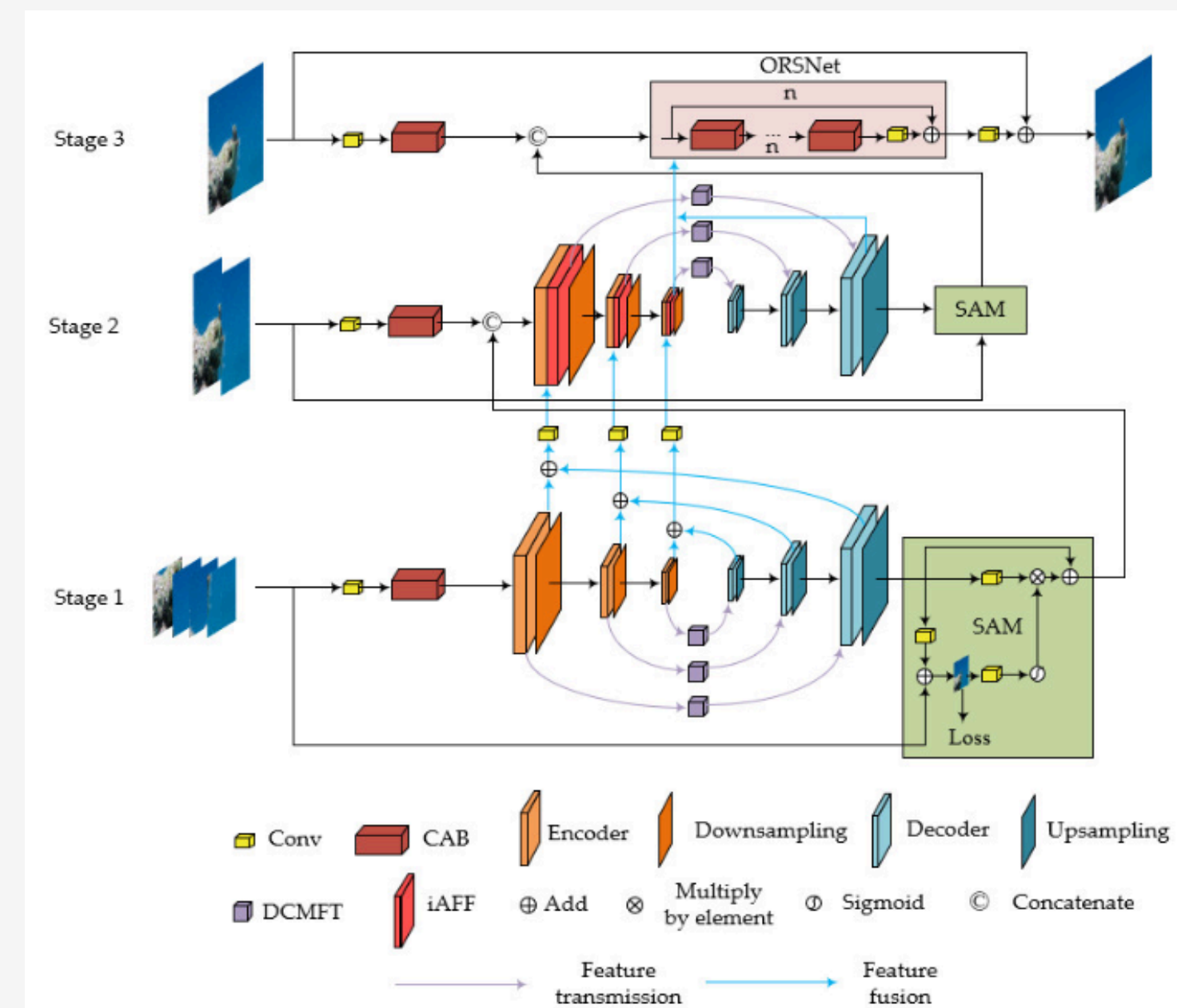
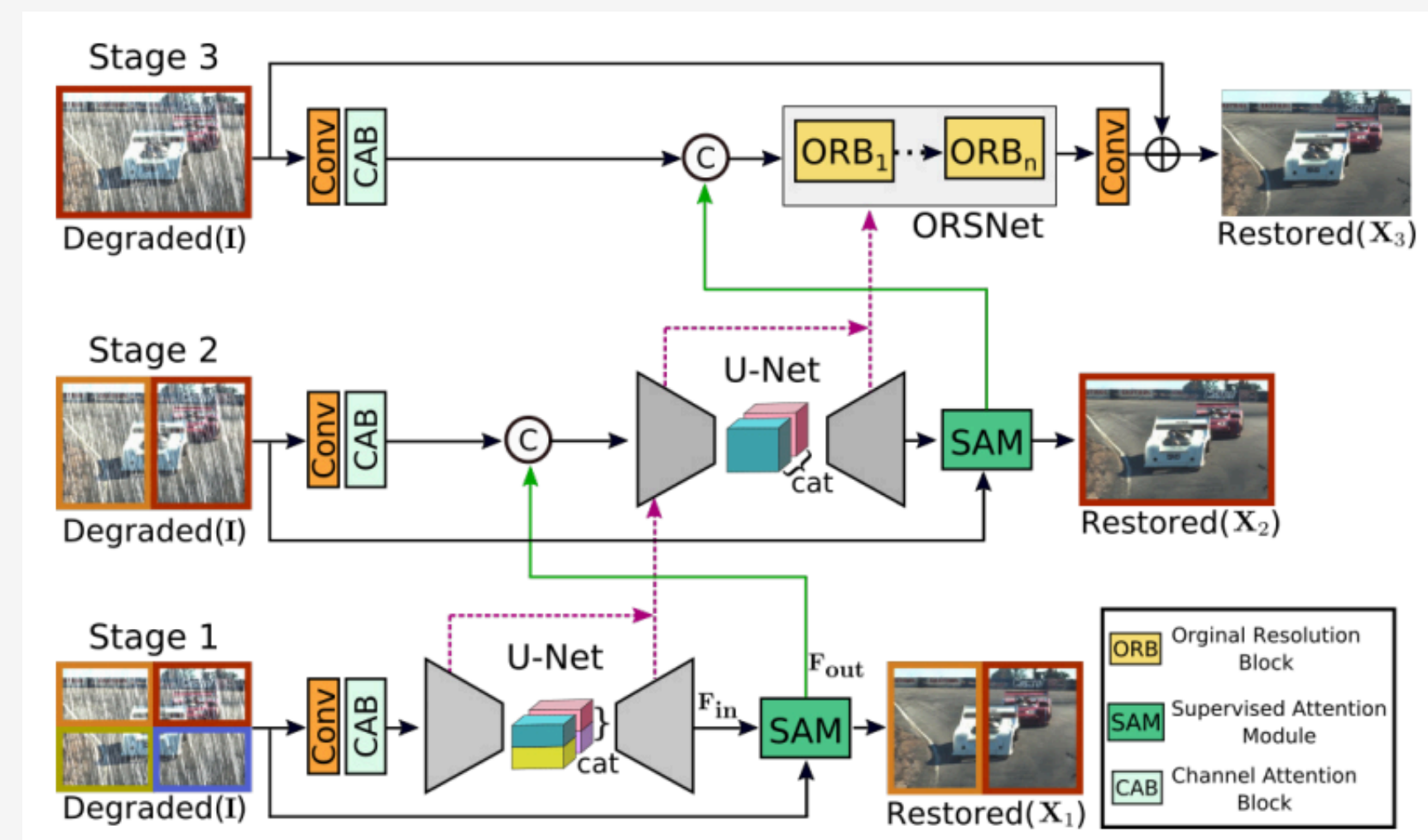
서로 다른 Stage의 Feature간 상관관계를 강화하기 위해 iAFF 사용. MPRNet은 이전 Stage와 현 Stage의 U-Net input Feature를 단순히 Concat 하여 사용하였지만, 두 Feature의 가중치를 동적으로 변경

Loss Function

MPRNet의 손실함수에 MS-SSIM 항을 추가 하여 밝기, 대비 및 구조적 유사성을 반영하고 학습 수렴과 복원 성능을 개선

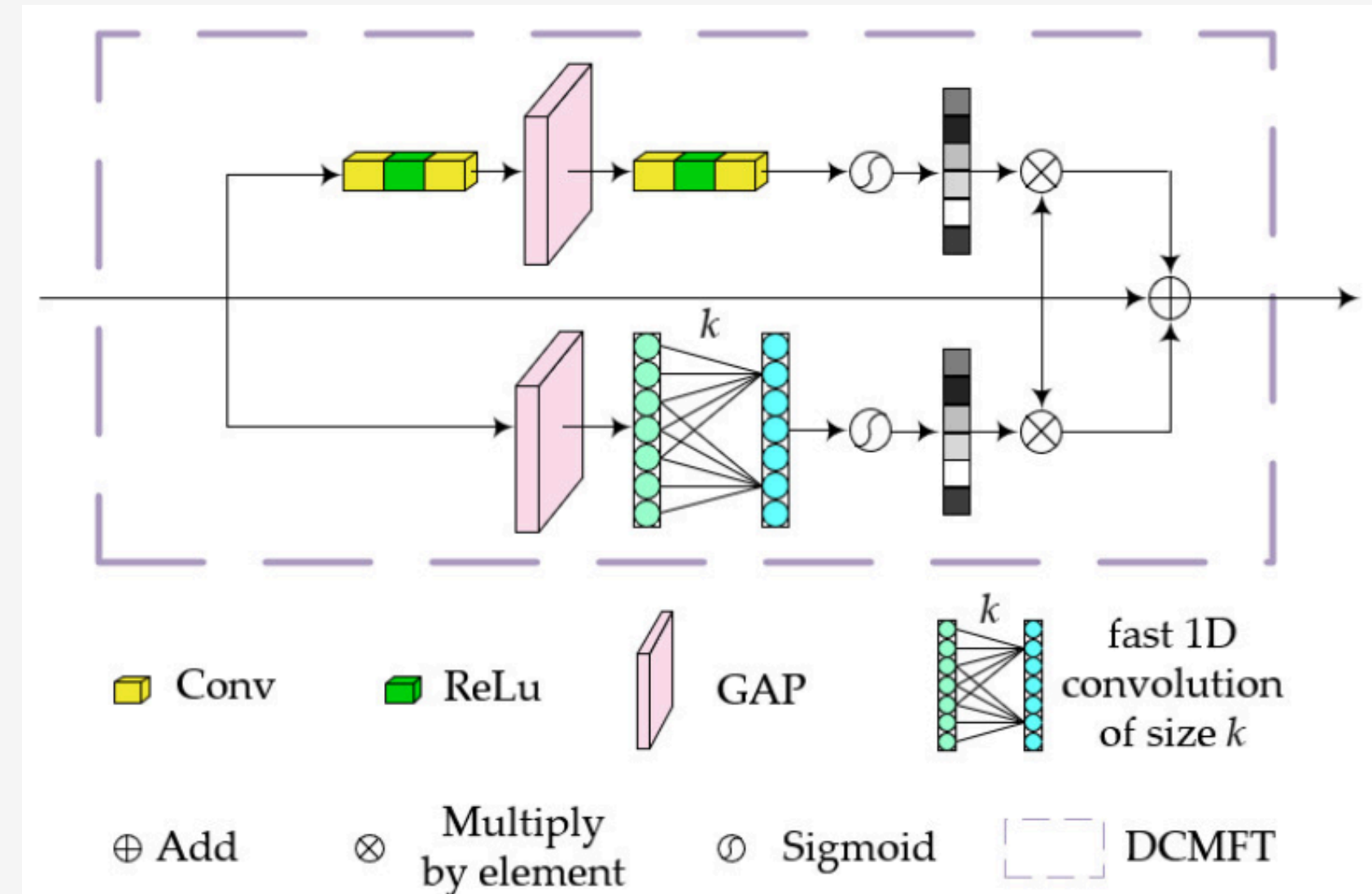
Key Modules

DCMFT



Key Modules

DCMFT

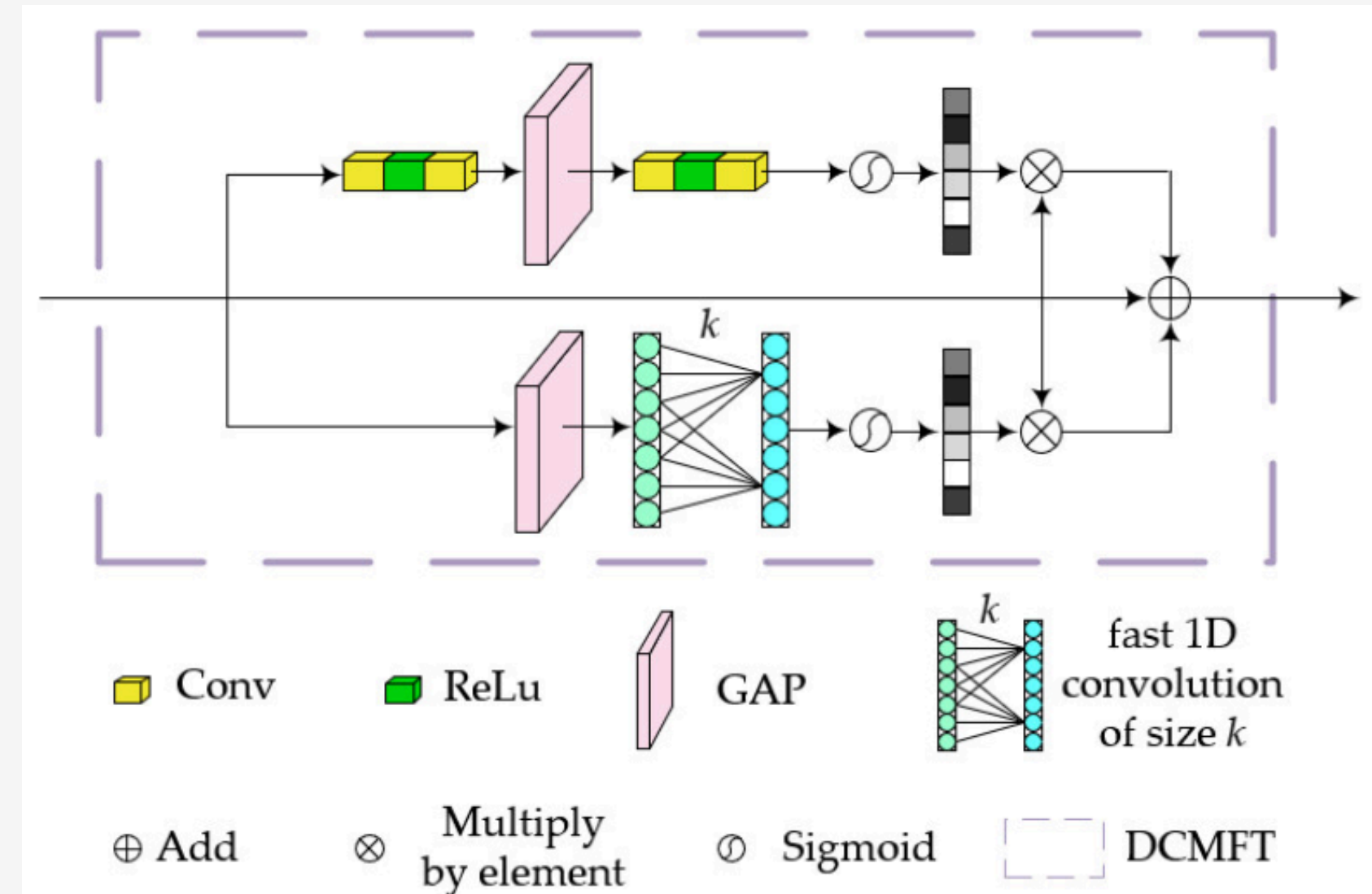


DCMFT

- Dual Channel Multi-Scale Feature Transmitter는 기존 U-Net의 skip connection을 개선하기 위해 새롭게 제안한 모듈이다.
- U-Net은 Encoder의 특징을 그대로 전달하는 반면, DCMFT는 각 특징 채널의 중요도를 계산하여 중요한 채널에 가중치를 더 부여하여 Decoder에 전달
- Branch 1은 전체적인 채널 중요도를 계산
- Branch 2는 채널방향의 1D Convolution을 이용해 주변 채널과의 관계를 계산
- 두 결과를 통합하여 최종적인 Channel Weight를 계산

Key Modules

DCMFT



GAP

- Global Average Pooling의 약자, Feature Map의 각 채널에 있는 모든 값을 평균내는 연산
- 각 채널이 얼마나 활성화 되었는지 확인

1D conv

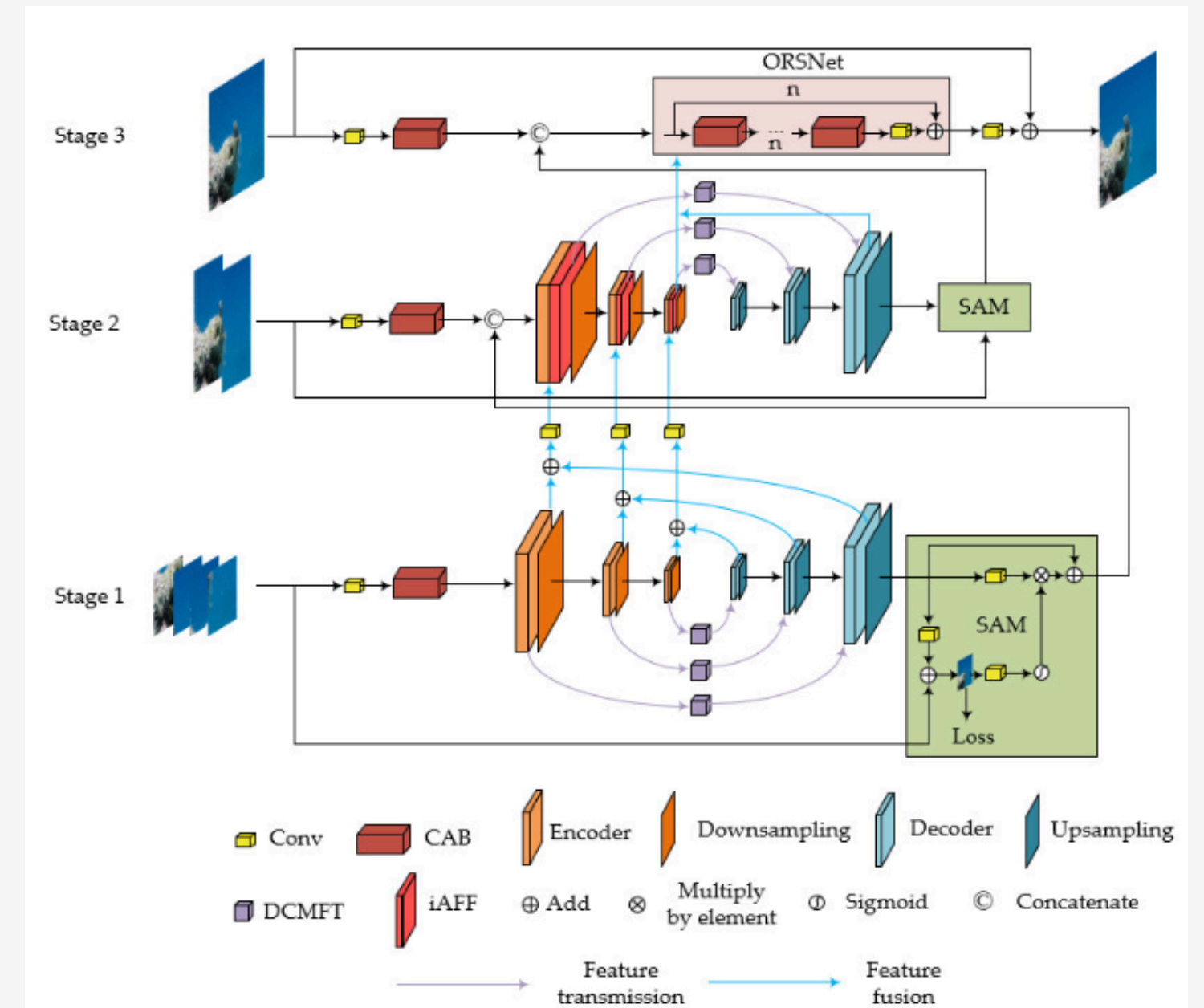
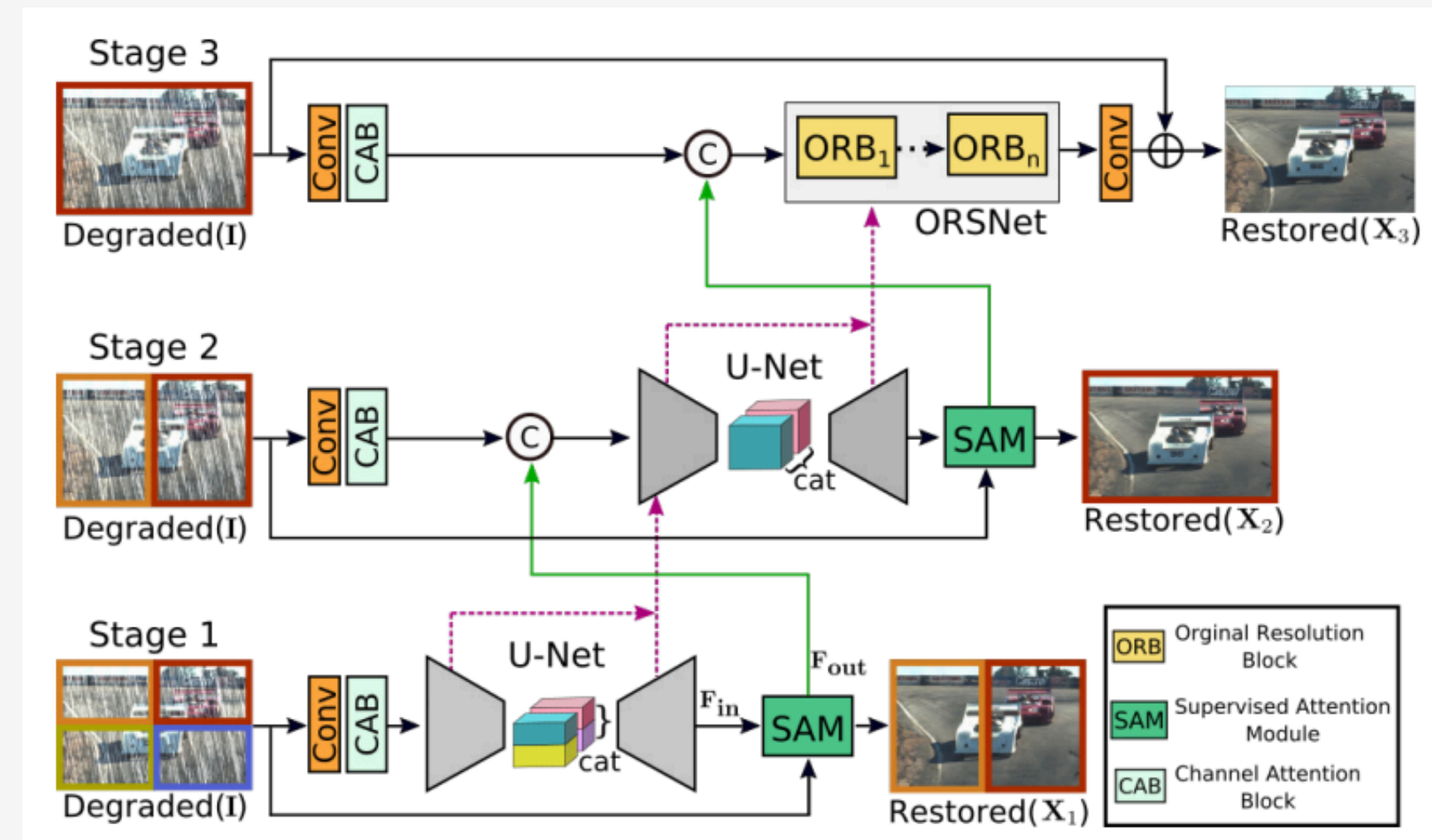
- 현재 채널의 앞뒤 k 개의 채널을 함께 연산에 사용해서 각 채널의 가중치를 정한다.

$$k = \varphi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}},$$

- 논문에서 $\gamma = 2$, $b = 1$ 사용

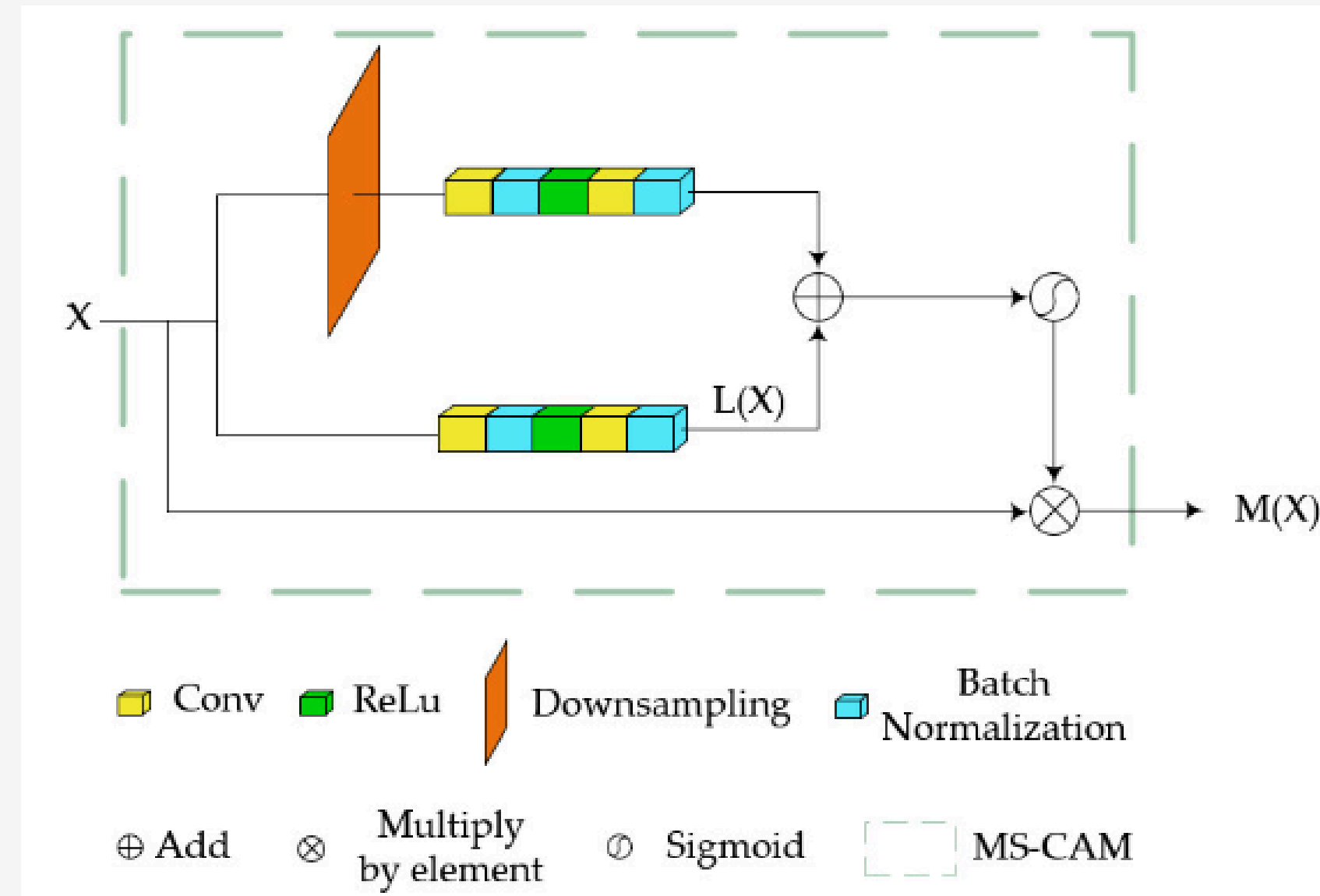
Key Modules

iAFF



Key Modules

iAFF



iAFF

- Iterative Attentional Feature Fusion은 각 Stage에서 이전 Stage의 Feature 정보를 활용할 때 더 중요한 Feature에 attention 하도록 유도하는 모듈

$$M(X) = X * \text{Sigmoid}(L(d(X)) + L(X))$$

$$L(X) = B(C_2(R(B(C_1(X)))))$$

- 입력 특징을 원래 해상도와 downsampling 한 해상도를 동시에 사용하여 국소적 특징과 넓은 영역의 문맥을 동시에 반영하도록 함

$$\psi(X, Y) = M(X + Y) * X + (1 - M(X + Y)) * Y$$

- 입력 특징의 포괄적인 융합을 보장하기 위해 두번 반복한 후 하위 Stage에서 상위 Stage로 특징 전달을 한다.

Key Modules

Loss Function

Charbonnier Loss

복원 이미지와 Ground Truth의 픽셀값 차이를 계산한다.

$$\sqrt{\|X_P - Y\|^2 + \varepsilon^2}$$

Edge Loss

Laplacian 연산을 사용하여 물체의 경계가 흐릿해지는 것을 방지한다.

$$\sqrt{\|\Delta(X_P) - \Delta(Y)\|^2 + \varepsilon^2}$$

MS-SSIM

Multi-Scale의 구조적 유사도를 계산한다.
평균 밝기, 명암 대비, 구조적 형태 등 구조가 뒤틀리는 것을 방지한다.

$$\mu S(X, Y)$$

Key Modules

SSIM

$$SSIM(X, Y) = l(X, Y)^\alpha * c(X, Y)^\beta * s(X, Y)^\gamma$$

$$l(X, Y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(X, Y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$$

$$s(X, Y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

- Luminance 향으로, 두 이미지의 평균 밝기가 비슷한지 비교
- Contrast 향으로, 두 이미지의 명암대비가 비슷한지 비교
- Structure 향으로, 두 이미지의 픽셀 변화 패턴이 비슷한지 비교

Key Modules

MS-SSIM

$$S(X, Y) = [l(X, Y)]^{\alpha_M} * \prod_{j=1}^M [c(X, Y)]^{\beta_j} * [s(X, Y)]^{\gamma_j}$$

MS-SSIM

- M개의 Scale 축소를 거치며 각 해상도별 SSIM을 구하는 식
- 명암(Contrast), 픽셀 변화(Structure)은 각 해상도별로 구해 곱연산을 진행하고, Luminance 항은 마지막 가장 축소된 Scale에서만 반영하여 최종 곱연산을 진행한다.

Key Modules

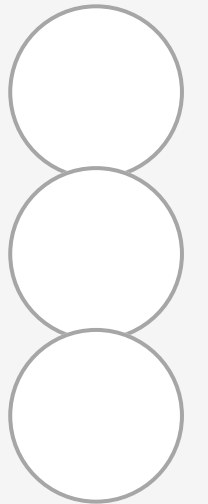
Loss Function

$$\mathcal{L} = \sum_{P=1}^3 [\sqrt{\|X_P - Y\|^2 + \epsilon^2} + \lambda \sqrt{\|\Delta(X_P) - \Delta(Y)\|^2 + \epsilon^2} + \mu S(X, Y)]$$

Loss Function

- 총 세개의 Stage에 대한 Loss의 합으로 Loss를 설정
- 최종 이미지를 생성하는 Stage는 Stage 3이지만, 하위 Stage들도 중간 이미지를 생성하면서 보다 정확한 Feature를 전달하도록 학습
- 실험적으로 $\lambda = 0.05$, $\mu = 0.01$ 으로 결정
- $\epsilon = 1e-3$ 으로 설정하여 기울기가 소실 되는 것을 방지

EXPERIMENTS



Experiments

Dataset configuration

Dataset



Experiments

Parameter Setting

Input size

384 X 384 pixel image

Optimizer

Adam Optimizer 사용

Parameter

Learning rate : 1e-6

Batch size : 4

Epoch : 100

Experiments

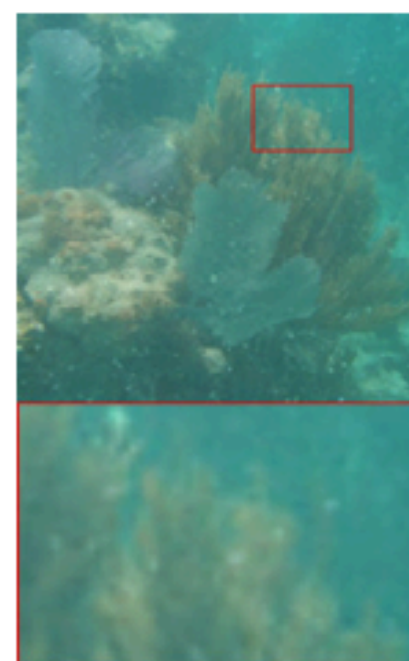
Output Comparison



Input



CycleGAN



CUT



BicycleGAN



pix2pix



MPRNet



DAD



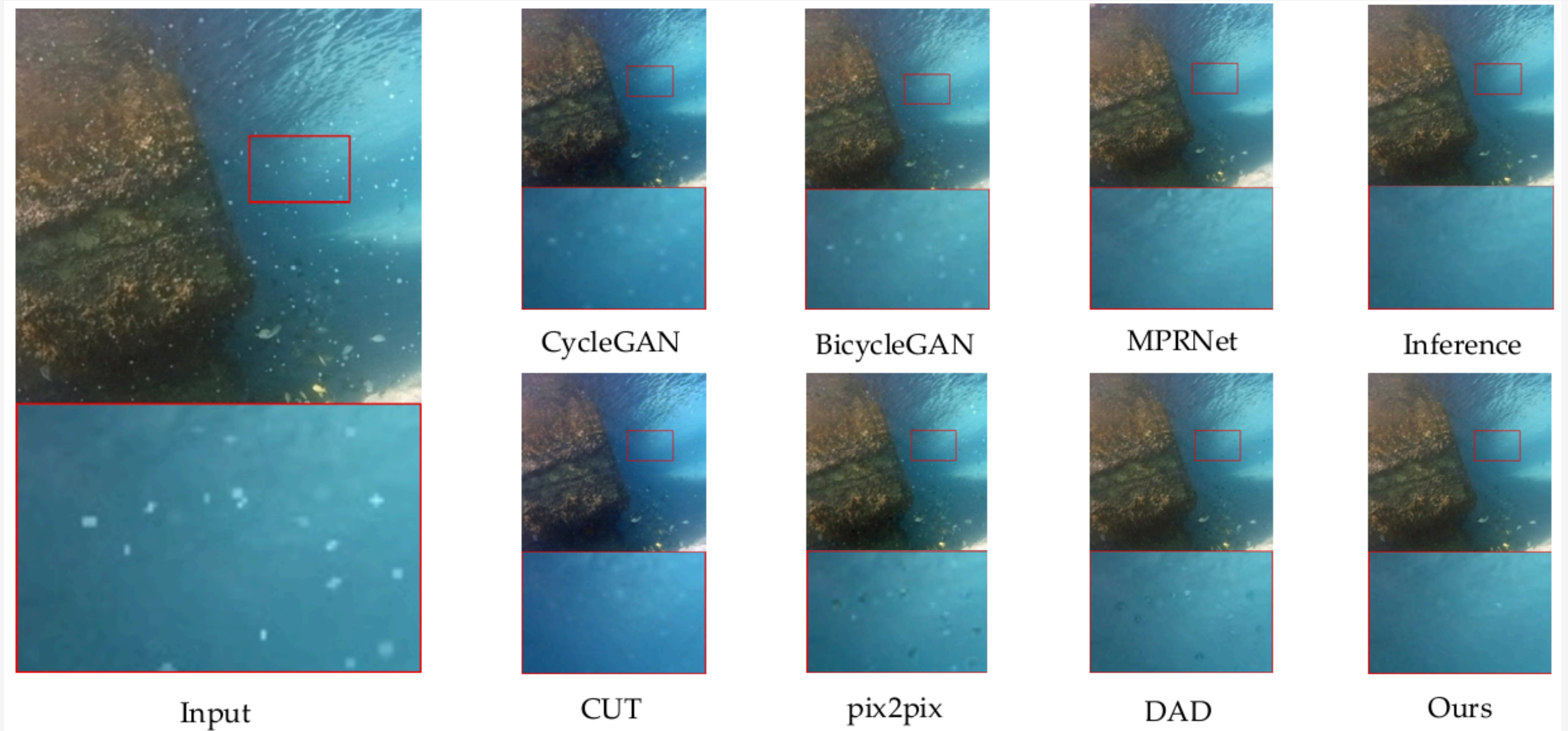
Inference



Ours

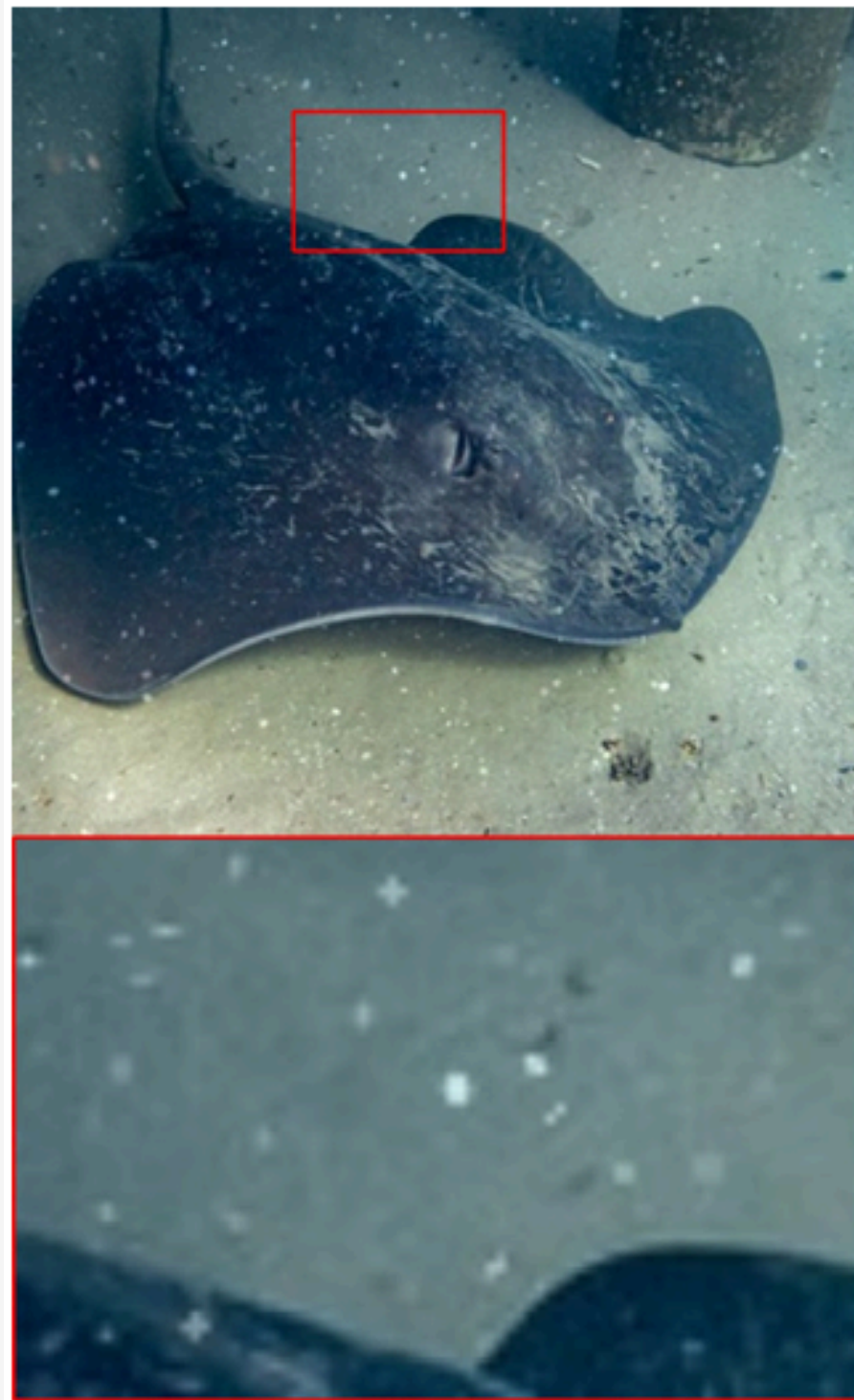
Experiments

Output Comparison

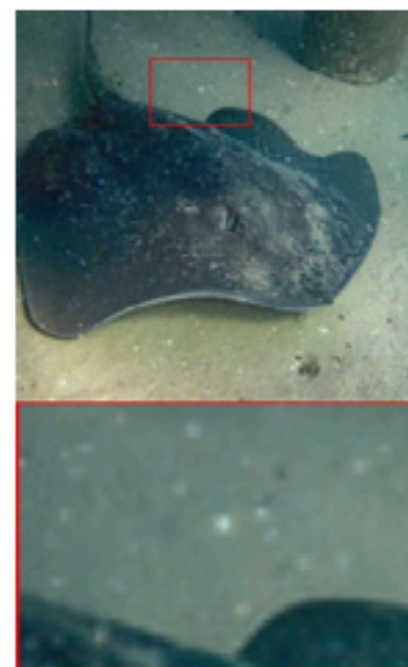


Experiments

Output Comparison



Input



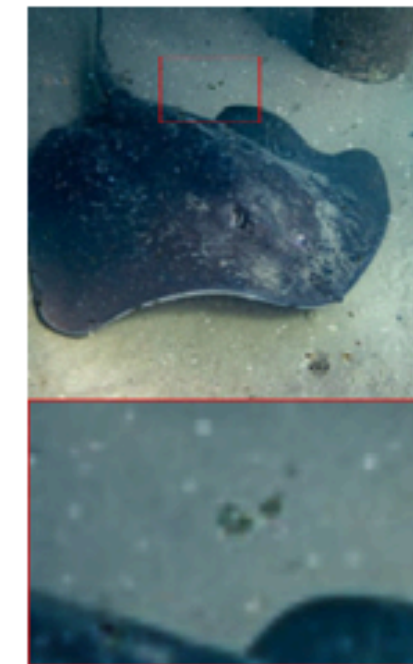
CycleGAN



CUT



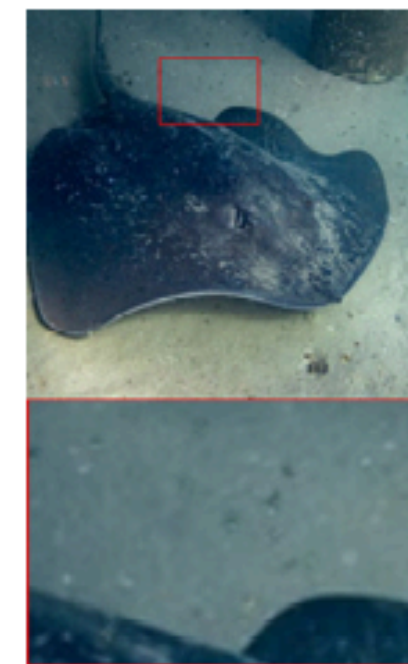
BicycleGAN



pix2pix



MPRNet



DAD



Inference



Ours

Experiments

Evaluation Metrics

PSNR

복원된 영상이 Ground Truth와
얼마나 유사한지를 측정하는 영상
품질 평가 지표

부유물 제거 영상이 원본 영상과
얼마나 비슷한지 측정

값이 높을수록 오차가 적다는 의미

SSIM

두 영상의 구조적 유사도를 평가하
는 지표

단순히 픽셀 차이가 아니라, 밝기
대비 구조를 함께 비교하여 전체적
인 유사도를 평가

0 ~ 1사이의 값을 가지며 높을수
록 원본과 유사하다는 의미

Experiments

Output Comparison

Method	PSNR (dB)	SSIM	MSE	Speed (FPS)	Model Size (MB)
CycleGAN [23]	25.3431	0.8382	348.3727	7.26	43.42
pix2pix [24]	30.9850	0.8846	68.8730	16.58	207.64
BicycleGAN [26]	30.4210	0.8653	159.4803	7.10	209.04
CUT [25]	26.2229	0.8440	287.0160	14.25	43.46
MPRNet [16]	38.5108	0.9747	14.2182	10.91	41.97
DAD [40]	34.2794	0.9257	41.4867	2.26	813.70
Ours	38.9251	0.9761	13.4415	10.44	45.23

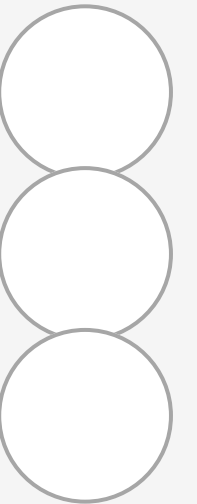
Experiments

Ablation Experiment

MPRNet에 iAFF와 MS-SSIM Loss, DCMFT 를 추가했을때 실제로 성능 향상에 기여했는지 확인하는 실험

Group	Module				Stage	PSNR (dB)	SSIM	MSE
	Baseline	iAFF	MS_SSIM	DCMFT				
1	✓	\	\	\	1	35.1214	0.9581	26.3924
		\	\	\	2	36.2208	0.9654	21.5644
		\	\	\	3	38.5108	0.9747	14.2182
2		✓	\	\	1	35.3099	0.9600	25.6271
		✓	\	\	2	36.6715	0.9676	19.5356
		\	\	\	3	38.6133	0.9753	14.1122
3		✓	✓	\	1	35.4822	0.9607	24.7388
		✓	✓	\	2	36.7619	0.9678	19.1418
		\	✓	\	3	38.7733	0.9756	13.8419
4		✓	✓	✓	1	35.7354	0.9621	23.6104
		✓	✓	✓	2	37.2048	0.9690	17.7378
		\	✓	\	3	38.9251	0.9761	13.4415

CONCLUSION



Conclusion

Network

- Marine Snow 제거를 위한 3단계 점진적 복원 네트워크 MP-MSRN 제안
- 새로운 Skip Connection 구조인 DCMFT를 통해 특징 정보 손실 감소 효과 확인
- 하위 Stage에서 Feature 전달을 위한 iAFF 구조를 통해 특징들을 선택적 융합, 활용하는 과정을 통해 각 Stage에서 추출된 Marine Snow 특징을 충분히 활용

Loss Function

- 기존 Charbonnier Loss와 Edge Loss에 MS-SSIM 항을 추가하여 여러 해상도에서 복원 이미지의 구조적 유사성 반영 성능을 확인
- 학습 수렴 측면에서도 성능 강화 확인

Future Direction

- 인접 프레임 정보를 활용하여 실시간 Marine Snow 제거로 확장하고, 영상 기반 실시간 복원으로의 확장 가능성 제시

A MULTI-STAGE PROGRESSIVE NETWORK WITH FEATURE TRANSMISSION AND FUSION FOR MARINE SNOW REMOVAL

발표를 마칩니다.

박정규

Deep Shark Lab 학부연구생

