

잠재된 특징 지식 전이 기반의 해충 영상 초해상화

정재민*, 손창환**

Pest Image Superresolution based on Latent Feature Knowledge Transfer

Jae-Min Jeong*, Chang-Hwan Son**

본 성과물은 농촌진흥청 연구사업(과제번호: PJ016303)의 지원에 의해 이루어진 것임

요 약

최근 인지 주도 방식의 초해상화 기법은 적대적 생성망 프레임워크에서 고수준 특징 추출과 인지적 화질 손실의 도입으로 우수한 성능을 거두고 있다. 그러나 인지 주도 기반 초해상화 기법은 저해상도 입력 영상에서 고수준의 특징을 학습하기 때문에 원본 고해상도 영상이 갖는 특징 분포를 생성하기가 어렵다. 이로 인해 복원 영상에서 에지의 선명도가 떨어지거나 텍스처의 표현 능력이 부족한 단점이 발생한다. 따라서 원본 고해상도 영상의 잠재된 특징 정보를 학습할 수 있는 오토인코더 기반의 교사 네트워크를 제시하고자 한다. 특히 에지 영상의 잠재된 특징 사전 정보를 초해상화 네트워크에 전이할 수 있는 방법을 소개하고자 한다. 실험 결과를 통해, 제안한 잠재된 특징 지식 전이 기반의 초해상화 모델이 기존의 인지 주도 방식보다 에지의 선명도를 증가하고 디테일의 표현을 향상할 수 있음을 보이고자 한다.

Abstract

Recently, perceptual-driven super-resolution methods have achieved excellent performance by introducing high-level feature extraction and perceptual image quality loss in an adversarial generative network framework. However, since the perceptually guided super-resolution method learns high-level features from low-resolution input images, it is difficult to generate the feature distribution of the original high-resolution images. As a result, the sharpness of the edge in the reconstructed image is lowered or the representation ability of the texture is insufficient. Therefore, we propose an autoencoder-based teacher network that can learn the latent feature information of the original high-resolution image. In particular, a method that can transfer latent feature prior information of edge images to a super-resolution network is introduced. Through the experimental results, it is shown that the proposed super-resolution model based on latent feature knowledge transfer can increase the sharpness of edges and improve the expression of details compared to the existing perceptual-driven methods.

Keywords

pest superresolution, generative adversarial network, laplacian filter, teacher network

* 군산대학교 소프트웨어학과 학사과정
- ORCID: <http://orcid.org/0009-0004-1134-3305>
** 군산대학교 소프트웨어학과 교수(교신저자)
- ORCID: <http://orcid.org/0000-0001-7077-3074>

• Received: Mar. 27, 2023, Revised: Apr. 19, 2023, Accepted: Apr. 22, 2023
• Corresponding Author: Chang-Hwan Son
Dept. of Software Science & Engineering, Republic of Korea
Tel.: +82-63-469-8915, Email: cson@kunsan.ac.kr

I. 서 론

최근 기후변화와 농산물 교역량 증가로 인해 새로운 병해충 발생이 증가하고 있다. 특히 외래·돌발 해충의 경우 예찰 및 진단법, 확산경로, 방제약제에 관한 정보가 부족하다. 조기 대응에 실패할 경우, 농작물 피해 확산, 생태계 교란 및 농가의 경제적 피해가 가중될 수 있다. 따라서 적기 방제를 위한 실시간 모니터링 시스템 구축이 필수적이다. 특히 온·습도를 포함한 기상환경 및 시기별 해충 발생량과 종류를 파악하기 위해, 최근 딥러닝 기반의 객체 검출 및 카운팅 기법이 모니터링 시스템에 활발히 도입되고 있다. 그러나 노지, 시설하우스 또는 트랩에서 촬영된 영상의 경우, 촬영거리 및 카메라 센서 성능으로 인해 객체 검출 및 분류를 위한 고해상도의 영상 취득이 어려울 수 있다. 이런 문제를 해결하기 위해, 촬영 영상의 해상도를 증가하면서 해충의 디테일 정보를 복구할 수 있는 초해상화(Superresolution) 기술[1]이 필요하다.

초해상화 기법은 이웃 화소 및 양선형 보간법에서 시작해서 스파스 코딩[2]과 같은 기계학습을 접목하다가 현재는 딥러닝 기반 기법이 대세를 이루고 있다. 최근 딥러닝 기반의 초해상화 기법[3]은 인지 화질 평가를 적대적 생성망 프레임워크에 적용한 인지 주도 방식(Perceptual-driven)[4]이 주류를 이루고 있다. 인지 주도 방식의 초해상화 기법은 복구 영상에 대한 인지 손실을 평가한다. 인지 손실은 원본 영상과 복원 영상을 VGG[5]와 같은 사전 학습된 신경망을 통과한 후, 생성된 특징 맵의 오차를 계산하여 측정된다. 사전 학습된 신경망은 인간 시각 시스템을 모델링한 것으로 간주할 수 있다. 따라서 인지 손실은 기존의 픽셀 단위의 화질 평가가 아닌, 인간 시각이 인지하는 화질 수준을 모델링할 수 있다. 이런 인지 손실을 도입할 경우, 화질 평가에 큰 영향을 주는 텍스처 표현을 사실감 있게 재현할 수 있다.

인지 주도 방식의 초해상화 기법은 기존의 VDSR[6], RCAN[7]과 같은 픽셀 주도 방식의 신경망 모델에 비해 텍스처의 생성 능력을 향상할 수 있다. 그러나 복원된 텍스처의 모양이 왜곡되거나 텍스처의 선명도가 원본에 비해 여전히 떨어지는

문제점을 갖고 있다. 이러한 이유는 텍스처 정보가 손실된 저해상도의 입력 영상에서 원본 고해상도가 갖는 고수준의 특징을 추정하기 때문이다. 따라서 본 논문에서는 기존의 인지 주도 방식의 고해상도의 특징 추출 능력을 강화하기 위해, 잠재된 특징 지식 전이(Latent feature knowledge transfer) 기반의 초해상화 모델을 제안하고자 한다. 특히, 원본 에지 영상의 잠재된 고수준의 특징을 추출하기 위해, 오토인코더 기반의 교사 네트워크(Teacher network)를 제시하고자 한다. 그리고 이 사전 학습된 교사 네트워크를 기존의 생성적 적대 네트워크(GAN, Generative Adversarial Network)와 결합하여, 원본 에지 영상의 잠재된 특징 분포와 고해상도 영상의 특징 분포가 유사하도록 전이하고자 한다. 이를 통해, 복원 영상의 해상도를 강화하고 텍스처의 선명도를 개선할 수 있음을 확인하고자 한다.

II. 관련 연구

2.1 인지 주도 초해상화 모델

초기 딥러닝 기반 초해상화 기법은 복원 영상과 원본 영상 간의 픽셀 오차의 합을 손실로 정의했다. 대표적인 픽셀 기반 손실함수로는 평균제곱오차(MSE, Mean Squared Error)와 구조 유사도(SSIM, Structure Similarity)[8]가 있다. 이러한 픽셀 기반 손실함수는 다수의 해를 평균하기 때문에 최종 복원 영상에서 텍스처가 흐릿해지는 효과를 초래한다. 이런 문제를 해결하기 위해, 인간 시각 시스템을 모델링한 인지 손실(Perceptual loss)이 개발되었다. 인지 손실이란 VGG[5]와 Resnet(Residual Neural Network)[9] 같은 사전 학습된 시각 시스템을 활용하여 복원 영상과 원본 영상 간의 오차를 계산하는 새로운 화질 평가 방식이다.

$$L_p = \frac{1}{HW} \sum_i^H \sum_j^W (\phi(\text{HR})_{i,j} - \phi(\text{SR})_{i,j})^2 \quad (1)$$

여기서 HR과 SR은 각각 원본 영상과 복원 영상을 의미한다.

그리고 ϕ 는 VGG 네트워크이고 고수준의 특징 맵을 출력한다. i 와 j 는 특징 맵의 공간 좌표에 해당한다. 따라서 식 (1)은 원본 영상과 복원 영상에 대응하는 특징 맵 간의 오차의 평균이다. VGG[5]는 이미지넷에 학습된 사전 모델로써, 인간 시각 시스템을 모델링한 것으로 볼 수 있다. 픽셀 기반 손실과는 달리, 인지 손실은 인간 시각 시스템이 인지하는 화질 평가를 수행할 수 있다. 이러한 인지 손실은 결과 영상에서 텍스처의 생성 능력을 제고할 수 있다. 대표적인 인지 주도 초해상화 모델로는 SRGAN[4], ESRGAN[10], SPSR[11], EDSR[12] 등이 있다.

2.2 SRGAN 모델

SRGAN은 인지 손실을 채택한 최초의 인지 주도 초해상화 모델이다. 기존의 생성적 적대 신경망 프레임워크를 따르면서 인지 손실을 추가하여 사실감 있는 텍스처 생성 능력을 확인하였다. SRGAN 모델은 그림 1에서 보듯이 크게 생성망(Generator)과 구별자(Discriminator)로 구성된다. 생성자는 저해상도 입력 영상을 고해상도 출력 영상으로 변환하는 역할을 담당하며 특징 추출기와 업샘플링 변환기로 이루어져 있다. 특징 추출기는 잔차 블록(Residual

block)과 롱-스킵 연결(Long skip connection)로 구현되었고 업샘플링 변환기는 합성곱(Convolution)과 픽셀 셔플러(Pixel shuffler)의 조합으로 설계되었다. 구별자는 VGG[5]와 같은 사전 학습된 신경망을 사용해서 원본과 복구 영상의 진위 여부를 결정한다. 잔차 블록과 밀집 블록(Dense block) 등과 같은 레이어 정보는 참고문헌에 자세하게 기록되어 있다.

III. 제안한 잠재된 특징 지식 전이 기반의 해충 영상 초해상화

SRGAN와 같은 인지 주도 방식의 초해상화 모델은 텍스처의 사실적 표현 능력을 강화할 수 있지만 여전히 개선 사항이 존재한다. 인지 손실의 사용은 사실감 있는 텍스처를 생성하지만, 텍스처의 왜곡을 초래하기도 한다. 또한 복구 영상의 텍스처의 선명도가 원본 고해상도의 수준에는 미치지 못한다. 본 연구에서는 기존 인지 주도 방식의 초해상화 모델의 텍스처 생성 능력을 강화하기 위한 잠재된 특징 지식 전이 기반의 초해상화 모델을 제안하고자 한다. 그리고 제안한 모델을 해충 영상에 적용하여 기존 인지 주도 모델에 비해 텍스처의 선명도와 디테일을 증가할 수 있음을 확인하고자 한다.

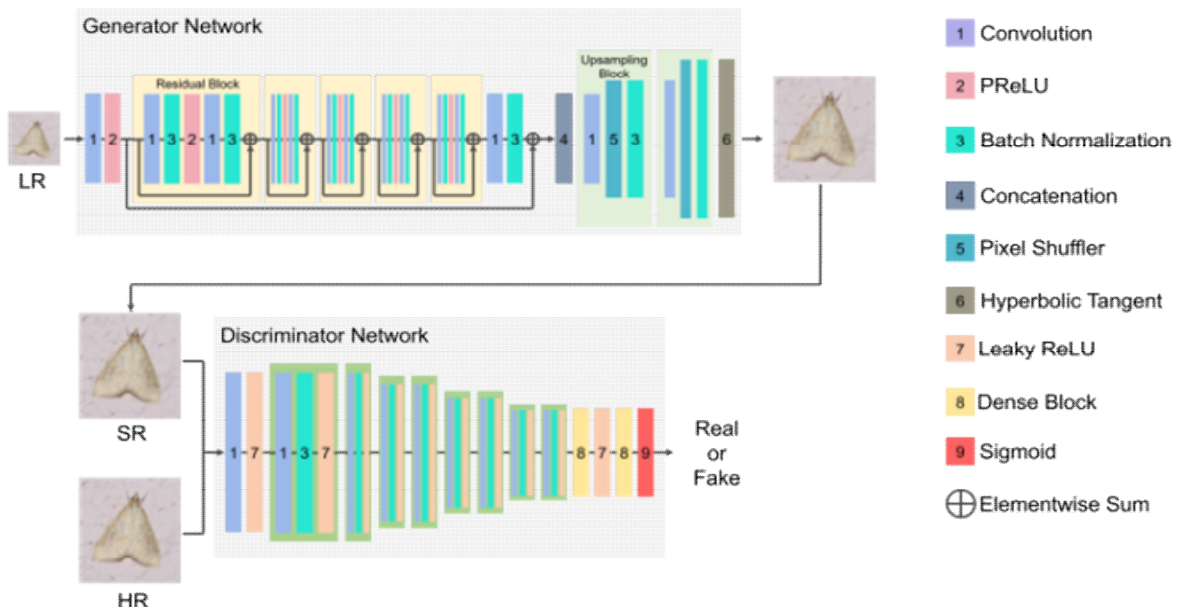


그림 1. SRGAN 구조

Fig. 1. Architecture of SRGAN

3.1 제안한 접근 방법

기존의 인지 주도 기반의 초해상화 모델은 저해상도 입력 영상에서 원본 고해상도 영상의 특징을 예측해야 한다. 그러나 저해상도 영상은 이미 디테일한 텍스처 정보가 손실되었기 때문에 고해상도의 텍스처를 복구하기에 한계가 있다. 이에 본 연구에서는 원본 고해상도 영상의 특징 분포에 대한 사전 정보(Prior information)를 활용하는 방안을 제시하고자 한다. 원본 고해상도 영상의 잠재된 특징은 오토인코더 기반의 교사 네트워크를 학습하여 추출할 수 있다. 따라서 학습된 교사 네트워크의 잠재된 특징 분포를 SRGAN[4]의 생성망에 전이하고자 한다.

그림 2는 제안한 잠재된 특징 지식 전이의 개념이다. 그림에서 보듯이, 교사 네트워크는 오토인코더로 구현되고 입출력이 원본 고해상도 예지 영상으로 동일하다. 즉, 오토인코더는 자기 자신을 학습하기 때문에 원본 예지 영상이 지닌 잠재된 고수준의 특징 분포를 모델링할 수 있다. 이렇게 사전 학습된 교사 네트워크의 특징에 대한 사전 정보를 SRGAN에 전이함으로써 생성망의 특징 추출 능력을 강화할 수 있다. 참고로 교사 네트워크는 학습 단계에서만 사용하고 테스트 단계에서는 필요하지 않기 때문에 SRGAN 모델의 용량은 증가하지 않는다. 그리고 그림 2에서 초해상화 모델로 SRGAN 외에 ESRGAN[10], SPSR[11] 등 다른 초해상화 모델도 적용할 수 있다.

3.2 교사 네트워크의 학습 영상 생성

고해상도 영상의 잠재된 특징 분포를 추정하기 위해 교사 네트워크의 입출력으로 원본 고해상도 영상을 그대로 사용할 수 있다. 그러나 실험을 통해, 원본 고해상도 영상 자체보다는 예지 영상을 학습하는 것이 더 좋은 결과를 유도할 수 있었다. 따라서 본 연구에서는 교사 네트워크를 학습하기 위해, 그림 3과 같이 예지 영상을 사용할 것이다.

예지 영상 생성은 고전적인 영상처리 기술로써, 1차 또는 2차 미분 필터를 적용하면 된다. 본 연구에서는 2차 미분 필터에 해당하는 라플라시안 필터를 적용했다. 물론 다양한 필터들의 결과를 입출력으로 사용해도 되지만 성능 결과에 큰 차이가 없음을 확인했다.

$$\mathcal{L}(f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{2}$$

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y) \tag{3}$$

$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y) \tag{4}$$

식 (2)의 \mathcal{L} 는 라플라시안 방정식을 의미하고 영상 밝기 f 에 대한 x 의 방향의 2차 미분과 y 방향의 2차 미분을 더한 스칼라값으로 표현된다.

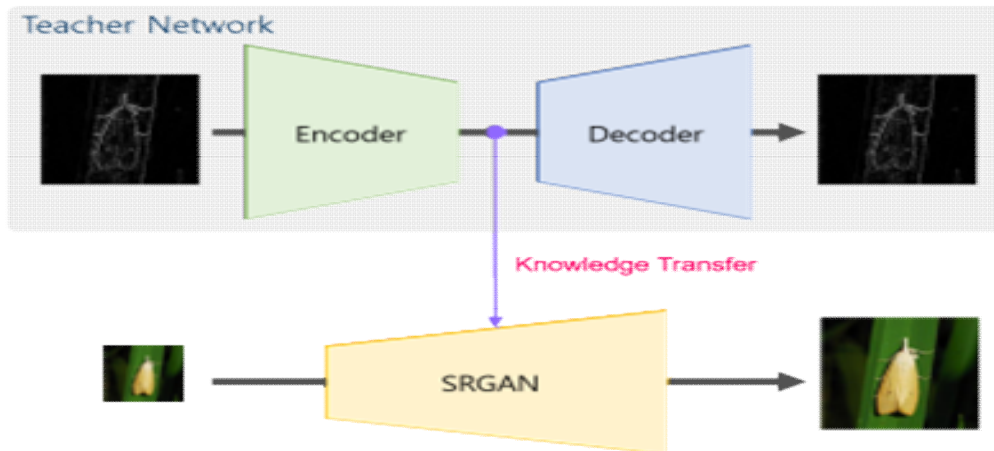


그림 2. 잠재된 특징 지식의 전이 개념
Fig. 2. Transfer concept of latent feature knowledge

식 (3)과 (4)는 각각 x 방향과 y 방향에 대한 2차 편미분 계산식이다. 즉, x 방향과 y 방향으로의 밝기의 변화량에 해당한다. 그림 3은 라플라시안 필터링을 적용해서 에지 영상을 생성하는 과정을 보여준다. 먼저, RGB 칼라 영상을 그레이 영상으로 변환한 후, 라플라시안 필터링을 통해 에지 영상을 구하게 된다. 결과 영상에서 원본 고해상도 영상의 에지 특성을 잘 추출할 수 있음을 알 수 있다.

3.3 제안한 잠재적 특징 지식 전이 기반의 초해상화 모델

그림 4는 기존 SRGAN[4] 모델에 교사 네트워크를 결합한 제안한 잠재된 특징 지식 전이 기반의 초해상화 모델의 아키텍처이다. 제안한 모델은 크게

교사 네트워크(상단)와 SRGAN 네트워크(하단)로 구성된다. 교사 네트워크는 다시 특징을 추출하기 위한 인코더(Encoder)와 추출된 특징으로부터 입력 영상을 재구성하는 디코더(Decoder)로 구성된다. 오토 인코더의 아키텍처로 U-net[13]이 많이 활용되지만, 본 연구에서는 특징 구별력을 강화하기 위해 어텐션 기능을 추가한 인코더-디코더 기반의 새로운 아키텍처를 적용했다. 교사 네트워크의 인코더는 이미지넷 데이터에 사전 학습된 Resnet[9]을 활용하여 버틀넥 블록(Bottleneck Block)을 쌓아서 고수준의 특징을 추출한다. 버틀넥 블록은 그림 5와 같이, 스킵 커넥션(Skip connection), 합성곱(Convolution), ReLU, 배치 정규화(Batch normalization) 계층으로 구성되어 있다.



그림 3. 교사 네트워크 학습을 위한 라플라시안 필터링 기반의 에지 영상 생성
Fig. 3. Edge image generation based on laplacian filtering for teach network learning

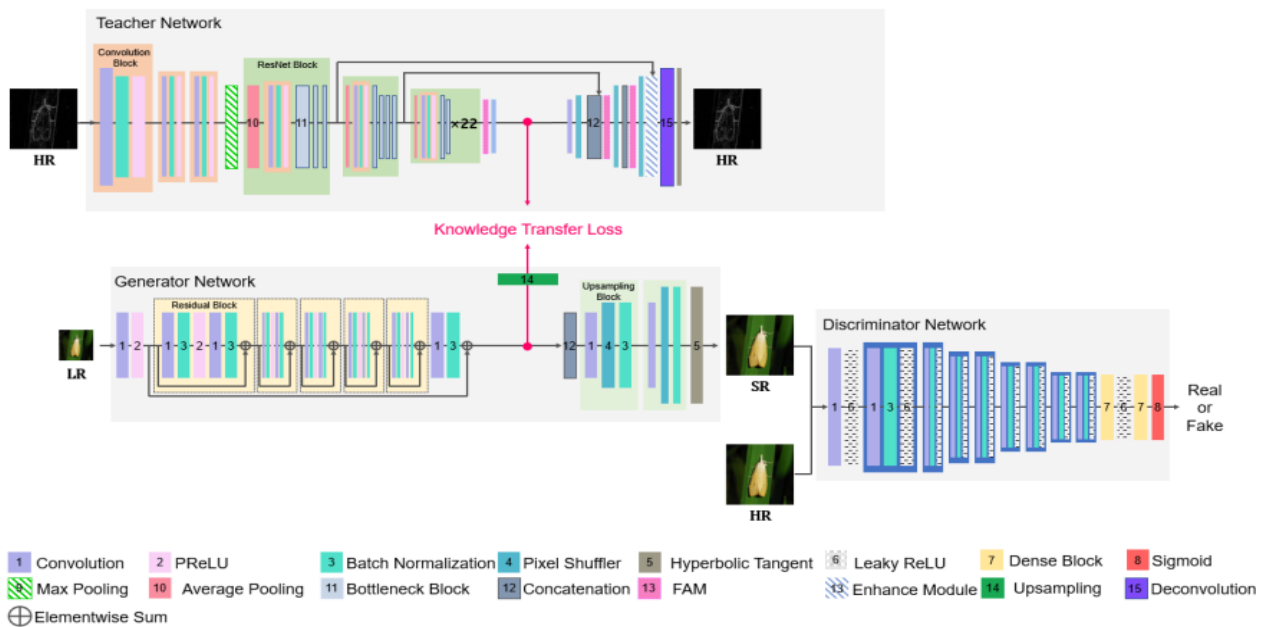


그림 4. 제안한 잠재된 특징 지식 전이 기반의 초해상화 모델
Fig. 4. Architecture of the proposed superresolution model based on latent feature knowledge transfer

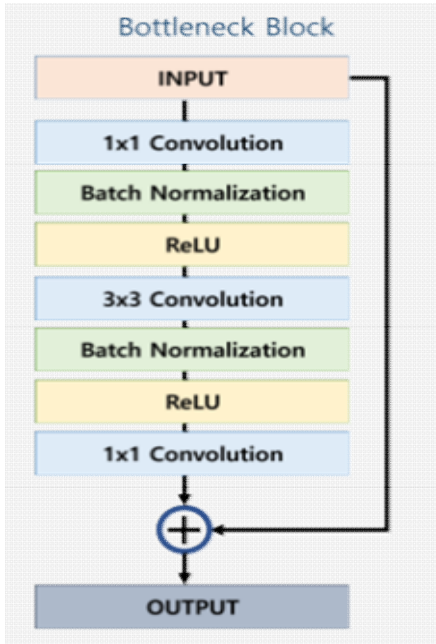


그림 5. 버틀넥 블록의 구조
Fig. 5. Architecture of bottleneck block

버틀넥 블록은 1×1 합성곱을 활용하여 입력 차원을 줄여 연산량을 감소시킬 수 있으며, 작은 크기의 필터를 사용해서 메모리 사용량을 줄일 수 있다. 또한 스킵 커넥션을 통해 입력값을 출력값에 추가하여, 높은 정확도를 유지할 수 있다.

교사 네트워크의 디코더는 원본 고해상도 영상의 스케일로 변환하기 위해 업샘플링 레이어를 적용한다. 이 논문에서는 픽셀 셔플러와 디컨볼루션(Deconvolution)을 사용해서 업샘플링을 구현했다. 일반적인 업샘플링 작업은 모든 픽셀을 한 번에 처리하기 때문에 계산 비용이 크게 들어가지만, 픽셀 셔플러는 작은 필터의 합성곱을 사용하여 각각의 픽셀을 따로 처리하기 때문에 계산 비용이 적게 들어간다. 디컨볼루션은 일종의 역학습 알고리즘으로써, 작은 크기의 패턴을 인식하는데 효과적인 커널을 사용하여 특징맵의 정보 손실을 최소화할 수 있다. 그리고 특징 추출과 멀티스케일 영상 복원 능력을 강화하기 위해, 특징 어텐션 모듈(Feature attention module)과 강화 모듈(Enhance module)을 추가했다.

그림 6은 특징 어텐션 모듈을 보여주고 있다. 그림에서 보듯이, 특징 어텐션 모듈은 채널 어텐션 블록(Channel attention block)과 픽셀 어텐션 블록(Pixel Attention block)으로 구성된다. 채널 어텐션은 평균

풀링(Average pooling)을 통해서 특징 맵을 1×1 크기로 변환하여 각 채널의 중요도를 학습하고 픽셀 어텐션은 컨볼루션과 최대 풀링(Max pooling)을 통해서 각 픽셀의 중요도를 모델링하여 어텐션 기반의 특징 기능을 실현한다.

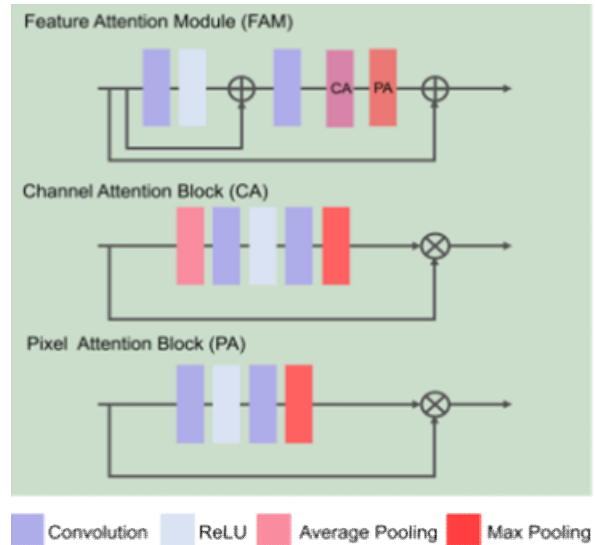


그림 6. 특징 어텐션 모듈
Fig. 6. Feature attention module

그림 7은 강화 모듈로써, 영상 재구성 능력을 강화하기 위해 평균 풀링을 사용해서 멀티스케일 퓨전 기능을 모델링했다. 그림과 같이, 입력 특징을 2×2, 4×4, 8×8, 16×16 평균 풀링을 적용해서 멀티스케일 특징을 추출하였다. 그리고 멀티스케일 특징을 퓨전 하기 위해 연결층과 합성곱을 사용했다.

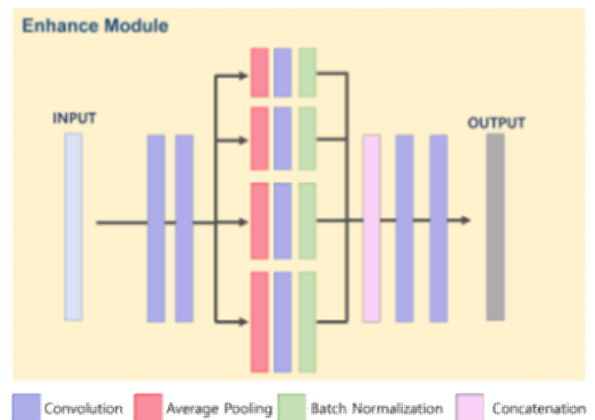


그림 7. 강화 모듈
Fig. 7. Enhancement module

그림 4의 교사 네트워크에서 빨간색 원으로 표시된 부분이 인코더의 마지막 부분에 해당하며 여기서 잠재된 고수준의 특징을 추출한다. 그리고 지식 전이 손실(KTL, Knowledge Transfer Loss)을 도입해서 잠재된 사전 특징 분포를 SRGAN 특징 추출기에 전이하고자 한다.

3.4 지식 전이 손실

SRGAN[4]은 텍스처 정보가 누락된 저해상도 입력 영상에서 특징을 학습하기 때문에 원본 고해상도 영상이 지닌 고수준의 특징 분포를 그대로 재현하기 어렵다. 이런 단점을 보완하기 위해, 사전 학습된 교사 네트워크의 특징 분포를 SRGAN에 전이하고자 한다. 이를 위해, 본 연구에서는 지식 전이 손실을 적용하고자 한다.

$$KTL = \frac{1}{HW} \sum_i^H \sum_j^W (F_{SRGAN}^{(i)} - UP(F_{TN}^{(i)}))^2 \quad (5)$$

$F_{TN}^{(i)}$ 과 $F_{SRGAN}^{(i)}$ 는 각각 교사 네트워크에서 추출된 잠재 특징과 SRGAN 생성자 네트워크에서 추출된 특징을 가리킨다. 하지만 두 특징 맵의 크기가 다르기 때문에 식 (5)와 같이 업샘플링을 적용하여 입력 크기를 정합하였다. 식 (5)는 SRGAN와 교사 네트워크에서 추출된 특징 맵의 오차를 최소화한다. 즉, 사전 학습된 교사 네트워크의 잠재된 고수준의 특징 분포를 SRGAN 생성망의 추출기에 전이할 수 있다.

IV. 실험 및 결과

4.1 실험 환경

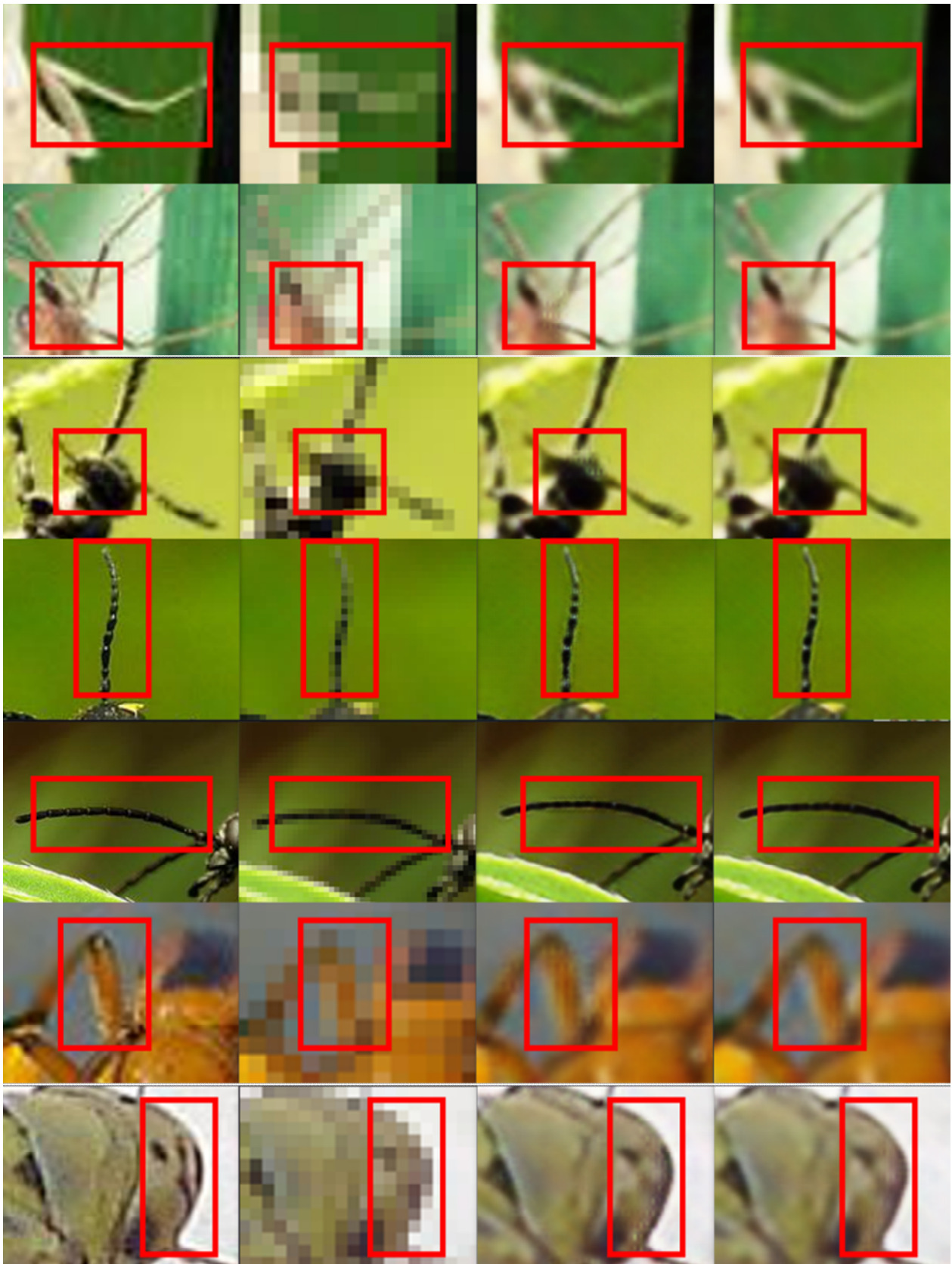
제안한 초해상도 모델을 학습하기 위해, IP102 공개 해충 데이터셋[14]을 사용했다. IP102 데이터셋은 총 102개의 해충으로 분류되어 있고 총 7,558장의 영상을 포함한다. 본 연구에서는 배경만 포함하거나 워터마크와 같은 불필요한 영상은 제거했다. 학습과 테스트 데이터의 비율은 7:3으로 정했고 과

적합 방지를 위해 뒤집기(Flipping)와 잘림(Cropping) 변환과 같은 데이터 강화(Data augmentation)을 적용했다. 저해상도 다운샘플링 비율은 4, 학습을 위한 배치 크기는 10, 에폭은 200, 학습률은 0.001로 설정했다. 사용된 최적화 기법은 아담 옵티마이저(Adam Optimizer)[15]이다.

4.2 정성적 화질 평가

제안한 잠재된 특징 지식 기반의 초해상도 기법의 화질 성능을 평가하기 위해, SRGAN 기법과 비교했다. 즉, 사전 학습된 교사 네트워크의 효과를 검증하기 위해 제안한 모델에서 초해상도 네트워크로 사용된 SRGAN 모델과 비교할 필요가 있다. 그림 8은 초해상도 실험 결과 영상이다. 그림에서 첫 번째 열은 고해상도 원본 영상이고 두 번째 열은 입력 저해상도 영상이다. 그리고 세 번째 열과 마지막 열은 각각 SRGAN 모델과 제안한 초해상도 기법을 적용한 복원 영상이다. 먼저 SRGAN의 경우, 전반적으로 에지 부분이 왜곡되거나 평탄 영역에서 잡음이 발생하는 경향이 있다. 예를 들어, 첫 번째 행의 SRGAN 결과 영상에서 해충의 다리 영역이 블러링 되어 선명도가 떨어지고 두 번째 행에서는 해충의 몸통 부분에서 텍스처 왜곡으로 잡음이 발생한 것처럼 보인다.

반면 제안한 모델은 사전 학습된 고해상도 영상의 잠재된 특징 지식 전이의 결과로 전반적으로 SRGAN에 비해 에지의 선명도가 강화되고 텍스처 생성 능력과 잡음제거 기능이 개선됨을 알 수 있다. 예를 들어, 세 번째 행의 해충 머리 부분에서 잡음이 제거되었고 다섯 번째 행에서는 해충 더듬이의 선명도가 강화되었다. 또한 네 번째와 여섯 번째 행에서도 각각 에지 영역의 표현 능력과 평탄 영역의 잡음제거 능력이 더 강화된 것을 알 수 있다. 이를 통해, 사전 학습된 교사 네트워크의 사용은 원본 고해상도 영상의 잠재된 특징 정보를 전이함으로써, 최종 복원 영상의 텍스처의 선명도를 강화하고 잡음 제거를 유도하는 효과가 있음을 확인할 수 있다.



(a) (b) (c) (d)

그림 8. 실험 결과 (a) 원본 영상, (b) 저해상도 영상, (c) SRGAN[4], (d) 제안한 기법
Fig. 8. Experimental results (a) Original image, (b) LR image, (c) SRGAN[4] and (d) Proposed method

4.3 정량적 화질 평가

기존 SRGAN 기법과 제안한 초해상화 기법의 정량적 화질 평가를 위해, 최대 신호 대 잡음비 (PSNR, Peak Signal-to-Noise Ratio)와 구조적 유사성 (SSIM, Structural Similarity)[8]을 평가 척도로 사용했다. 최대 신호 대 잡음비는 원본 영상과 복원 영상 간의 픽셀 오차를 계산하여 영상 전체의 화질이 얼마나 유사한지를 판별한다.

$$PSNR(I,J) = 10 \log_{10} \left(\frac{R^2}{MSE(I,J)} \right) \quad (6)$$

$$MSE(I,J) = \frac{1}{HW} \sum_i^H \sum_j^W (I_{i,j} - J_{i,j})^2 \quad (7)$$

$$SSIM(I,J) = [l(I,J)]^\alpha \times [c(I,J)]^\beta \times [s(I,J)]^\gamma \quad (8)$$

식 (6)은 PSNR 척도로써, R 은 최대 밝기 값에 해당하고 I 와 J 는 원본 영상과 결과 영상을 각각 의미한다. MSE는 평균제곱오차를 의미한다. 식 (6)의 PSNR은 최종 로그 함수를 적용하여 인간 시각 특성을 반영하여 식 (7)과 같이 픽셀 간 오차를 계산한다. 식 (8)은 SSIM 척도로써, 3가지의 항목으로 구성된다. 즉, 밝기 유사도(l), 대비 유사도(c), 구조 유사도(s) 항목으로 구성된다. 그리고 α , β , γ 는 각 항목의 중요도를 반영하기 위한 가중치 값이다. 각 항목은 0~1 사이의 값을 가지며 수치가 높을수록 두 영상의 유사도가 높은 것을 의미한다. 표 1은 테스트 데이터에 대한 PSNR과 SSIM의 결과이다. 표의 결과와 같이, 제안한 기법은 SRGAN 모델에 비해 PSNR 수치가 약 0.2dB 정도 개선되었으며 SSIM 수치는 약 0.02 정도 향상되었다. 이는 제안한 기법이 기존 SRGAN 모델에 비해 해상도 성능이 개선되었고 고 에지 표현 능력이 향상되었음을 뒷받침해준다.

표 1. 정량적 화질 평가

Table 1. Quantitative image quality evaluation

Methods	PSNR	SSIM
SRGAN[4]	30.263	0.8401
Proposed mthod	30.497	0.8420

V. 결 론

본 연구에서는 해상도 강화와 텍스처의 선명도 개선을 위한 고해상도 영상의 잠재된 특징 지식을 전이 가능한 초해상화 기법을 제안하였다. 특히, 어텐션 특징과 강화 모듈을 탑재한 인코더-디코더 기반의 교차 네트워크를 도입하여 원본 에지 영상의 잠재된 특징 분포를 학습하였다. 그리고 이 사전 학습된 특징 정보를 초해상화 SRGAN의 특징 추출기에 전이함으로써 특징 추출 능력과 영상 재구성 능력을 개선하였다. 실험 결과를 통해, 제안한 초해상화 모델이 교차 네트워크의 사전 정보를 활용함으로써 텍스처 선명도와 해상도 증가 효과가 있음을 확인하였다. 즉, 기존의 SRGAN와 비교하여, SSIM 0.02 상승과 PSNR 0.2dB 화질 개선을 달성할 수 있었다.

References

- [1] J. Y. Jeong and C. H. Son, "Class-specific perceptual loss modeling for pest superresolution", Proc. of JKIIIT, Vol. 20, No. 8, pp. 123-132, Aug. 2022. <https://doi.org/10.14801/jkiit.2022.20.8.123>.
- [2] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation", In Proc. IEEE Transactions on Image Processing, Vol. 19, No. 11, pp. 2861-2873, May 2010. <https://doi.org/10.1109/tip.2010.2050625>.
- [3] J. Sun, Z. Xu, and H. Y. Shum, "Image super-resolution using gradient profile prior", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Anchorage, USA, pp. 1-8, Jun. 2018. <https://doi.org/10.1109/CVPR.2008.4587659>.
- [4] C. Leidg, et al., "Photo-realistic single image super-resolution using a generative adversarial network", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, pp. 4681-4690, Jul. 2017. <https://doi.org/10.1109/CVPR.2017.19>.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", In Proc. International Conference on Learning Representations, San Diego, CA, USA, May 2015.

[6] J. Kim, J. L. Kwon, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 1646-1654, Jun. 2016. <https://doi.org/10.1109/CVPR.2016.182>.

[7] Y. Zhang, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks", In Proc. European Conference on Computer Vision, Munich, Germany, pp. 286-301, Sep. 2018.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600-612, Apr. 2004. <https://doi.org/10.1109/tip.2003.819861>.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 770-778, Jun. 2016. <https://doi.org/10.48550/arXiv.1512.03385>.

[10] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks", In Proc. European Conference on Computer Vision, Munich, Germany, pp. 63-79, Sep. 2018. <https://doi.org/10.48550/arXiv.1809.00219>.

[11] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super-resolution with gradient guidance", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 7769-7778, Jun. 2020. <https://doi.org/10.48550/arXiv.2003.13081>.

[12] B. Lim, S. Son, He. Kim, S. Nah, and K. M.

Lee, "Enhanced deep residual networks for single image super-resolution", Jul. 2017. <https://doi.org/10.48550/arXiv.1707.02921>.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", In Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, pp. 234-241, Oct. 2015. https://doi.org/10.1007/978-3-319-24574-4_28.

[14] X. Wu, C. Zhan, Y. K. Lai, M. M. Cheng, and J. Yang, "IP102: A large-scale benchmark dataset for insect pest recognition", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Long Beach, USA, pp. 8787-8796, Jun. 2019. <https://doi.org/10.1109/CVPR.2019.00899>.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", In Proc. International Conference on Learning Representation, San Diego, USA, May 2015. <https://doi.org/10.48550/arXiv.1412.6980>.

저자소개

정재민 (Jae-Min Jeong)



2018년 3월 ~ 현재 : 군산대학교
소프트웨어학과 학사과정
관심분야 : 컴퓨터 비전, 영상처리,
기계학습, 딥 러닝

손창환 (Chang-Hwan Son)



2017년 4월 ~ 현재 : 군산대학교
소프트웨어학과 부교수
관심분야 : 컴퓨터 비전, 영상처리,
기계학습, 딥 러닝