

일반논문 (Regular Paper)

방송공학회논문지 제30권 제4호, 2025년 7월 (JBE Vol.30, No.4, July 2025)

<https://doi.org/10.5909/JBE.2025.30.4.620>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

위성 영상의 탐지 및 분할 성능 향상을 위한 작업 특화 초고해상화 방법

서 예 원^{a)}, 한 승 석^{a)}, 최 해 철^{a)*}

Task-Specific Super-Resolution Training Method for Enhancing Detection and Segmentation Performance in Satellite Imagery

Yewon Seo^{a)}, Seungseok Han^{a)}, and Haechul Choi^{a)*}

요 약

위성 영상에서 높은 탐지 및 분할 성능을 달성하기 위해서는 고해상도 영상이 요구되지만, 이를 확보하는 데는 높은 비용이 소요된다. 본 논문에서는 저해상도 위성 영상을 기계 작업에 최적화된 고해상도로 변환하여 탐지 및 분할 성능을 향상시키는 방법을 제안한다. 제안 방법은 기존 초고해상화 신경망에 재구성 손실, 지각 손실, 작업 손실을 결합한 새로운 손실함수를 적용하는 작업 특화 초고해상화 신경망 학습 구조를 갖는다. 이러한 결합 손실함수는 기존의 재구성 화질에 특화된 초고해상화 신경망을 작업에 특화된 초고해상화 신경망으로 학습시켜 탐지 혹은 분할 작업의 성능을 향상시킬 수 있다. 실험 결과는 재구성 손실만을 활용한 기존 초고해상화 신경망 대비 제안 방법이 탐지와 분할 성능을 일관되게 향상시킴을 보이며, 이 결과는 제안 방법이 작업 특화 고해상도 영상 생성에 효과적임을 입증한다.

Abstract

High-resolution images are essential for achieving high performance in detection and segmentation tasks for satellite images, but their acquisition involves significant costs. This paper introduces a method that transforms low-resolution satellite images into task-optimized high-resolution ones to enhance detection and segmentation performance. The proposed approach employs a task-specific super-resolution network training framework that integrates reconstruction, perceptual, and task losses into a unified loss function. This combined loss function enables the super-resolution network to better adapt to detection and segmentation tasks, thereby improving task-specific performance. Experimental results demonstrate that the proposed method consistently outperforms conventional super-resolution networks relying solely on reconstruction loss, proving its effectiveness in generating task-optimized high-resolution images.

Keyword : Super resolution, Satellite image, Detection, Segmentation, Task loss, Perceptual loss

I. 서론

딥러닝 신경망을 활용한 탐지 및 분할 작업에서 입력 영상의 해상도는 작업 성능에 많은 영향을 미친다. 고해상도(High Resolution, HR) 영상은 객체의 형태와 경계를 정밀하게 표현할 수 있어, 저해상도(Low Resolution, LR) 영상보다 탐지 및 분할 작업에서 우수한 성능을 발휘한다. 이러한 특징은 특히 객체와의 거리가 먼 곳에서 촬영되어, 객체가 작게 표현되는 위성영상에서 더욱 두드러진다. 위성영상을 활용한 객체 탐지 및 분할 작업에서 높은 성능 달성하기 위해서는 고해상도 영상이 매우 유리하지만, 고해상도 위성영상을 획득하는 데는 높은 비용이 요구된다. 이에 따라, 제한된 해상도의 영상으로도 높은 작업 성능을 확보할 수 있게 하는 기술에 대한 관심이 증가하고 있다.

최근에는 저해상도 위성영상을 초고해상도(Super-Resolution) 신경망을 통해 고해상도의 영상으로 변환한 뒤, 이를 객체 탐지 및 분할과 같은 작업에 활용하려는 연구들이 활발히 진행되고 있다^{[1][2]}. 그러나 대부분의 초고해상도 신경망은 시각적 품질 개선에 중점을 두고 설계되었기 때문에, 복원 과정에서 작업에 중요한 시각적 또는 의미적 특징이 왜곡되거나 손실될 가능성이 존재하며, 이는 작업 성능에 부정적인 영향을 미칠 수 있다. 따라서 초고해상도 과정에서 작업에 중요한 특징이 온전히 복원될 수 있도록 설계된 학습 방법이 필요하다.

본 논문에서는 초고해상도 신경망의 학습을 위해 기존에 활용되던 재구성 손실(reconstruction loss)에 추가하여 시각 손실(perceptual loss)과 작업 손실(task loss)을 결합한 새로운 손실함수를 제안한다. 이 손실함수로 학습된 초고해상도 신경망으로 생성한 고해상도 영상을 탐지 혹은 분

할 작업의 입력으로 활용함으로써 작업 정확도를 향상시키고자 한다. 제안된 방법에서 시각 손실은 재구성 과정에서 발생할 수 있는 의미적 정보 손실을 완화하며, 작업 손실은 작업에 특화된 특징을 강화함으로써 작업 친화적인 결과를 도출하는 데 기여할 것으로 기대된다.

본 논문의 구성은 다음과 같다. II장에서는 관련 연구에 대해 설명하고, III장에서는 제안하는 방법의 구조와 학습 방식을 소개한다. IV장에서는 실험을 통해 제안 방법의 성능을 검증하고, V장에서는 결론 및 향후 연구 방향에 대해 논의한다.

II. 관련 연구

최근 초고해상도 신경망은 Convolution Neural Network(CNN) 기반 신경망과 Transformer^[3] 기반 신경망으로 활발히 연구되고 있다. CNN 기반 신경망은 영상 내 지역적인 특징을 추출하는 데 최적화된 구조이다. 일반적으로 합성곱 레이어와 서브샘플링(subsampling) 레이어를 반복적으로 구성한 후, 완전 연결(Fully Connected) 레이어를 통해 최종 출력을 산출한다. 합성곱 레이어는 학습 가능한 필터를 사용해 입력 영상으로부터 지역적인 특징 맵(feature maps)을 생성하며, 서브 샘플링 레이어는 공간 해상도를 줄여 연산량을 감소시키고 주요 정보를 보존하는 역할을 수행한다. 그림 1은 CNN의 대표적인 초기 구조인 LeNet-5^[4]를 보여준다. 입력 영상은 먼저 C1 합성곱 레이어를 통과하며 특징 맵을 생성하고, 이어서 S2 서브샘플링 레이어를 통해 공간 크기를 줄이고 주요 특징만을 유지한다. 이후에도 C3 합성곱, S4 서브샘플링, C5 완전 연결 레이어 등 특징 추출과 압축을 반복하면서 최종적으로 F6 완전 연결 레이어와 출력 레이어(Output layer)를 통해 분류 작업이 수행된다. 이러한 CNN 기반 신경망의 대표적인 예로는 Residual Feature Distillation Network(RFDN)^[5]과 Residual Channel Attention Networks(RCAN)^[6] 등이 있다. RFDN은 기존의 CNN 기반 초고해상도 신경망에 특징 증류 블록을 추가하여 각 레이어에서 추출된 특징 중 중요한 특징만 반복적으로 정제한다, 또한, 계층적 스킵 연결 구조를 활용하여 정보 손실을 줄이면서도 연산 효율을 유지하

a) 국립한밭대학교 소프트웨어융합대학원 지능미디어공학과(Department of Intelligent Media Engineering, Graduate School of Software Convergence, Hanbat National University)

‡ Corresponding Author : 최해철(Haechul Choi)

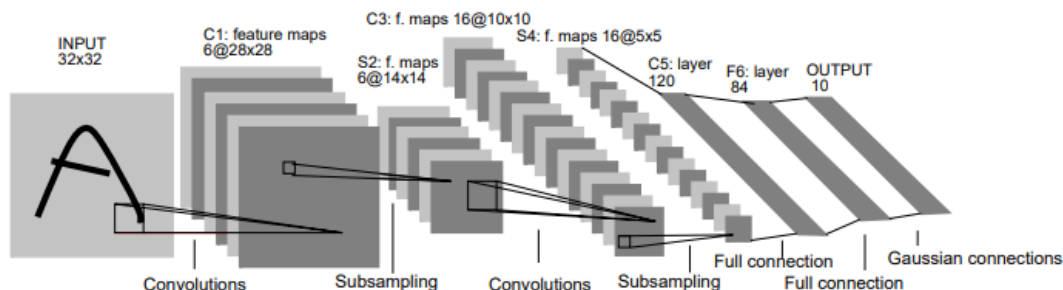
E-mail: choihc@hanbat.ac.kr

Tel: +82-42-825-1196

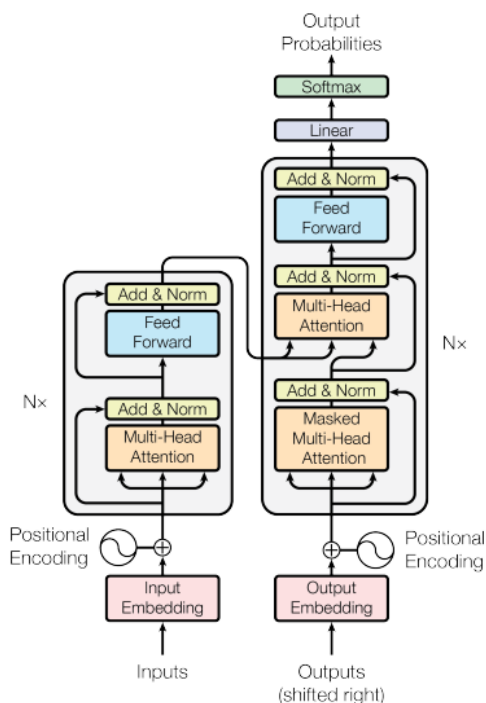
ORCID: <https://orcid.org/0000-0002-7594-0828/>

※ 이 논문은 2022년도 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임 (KRIT-CT-22-047, 우주계층 지능통신망 특화연구실)

· Manuscript May 13, 2025; Revised June 20, 2025; Accepted June 26, 2025.

그림 1. LeNet-5의 구조^[4]Fig. 1. Architecture of LeNet-5^[4]

는 경량 초고해상화 모델이다. RCAN은 CNN 기반 초고해상화 구조 위에 Residual-in-Residual 블록을 쌓아 신경망의 깊이를 증가시키면서도 안정적인 학습이 가능하도록 하였으며, 여기에 채널 어텐션 기술을 적용해 각 채널의 중요도를 학습함으로써 의미 있는 특징을 강조하고 복원 품질을 향상시킨다.

그림 2. Transformer 모델의 구조^[3]Fig. 2. Architecture of Transformer^[3]

Transformer 기반 신경망은 셀프 어텐션을 통해 입력 영상의 관계를 동시에 학습할 수 있는 구조로, CNN 기반 신경망과 달리 전역적인 문맥 정보를 효과적으로 처리할 수 있다. 인코더와 디코더 모두에서 셀프 어텐션과 위치별 전방향(feed-forward) 레이어를 반복적으로 쌓아 구성된 구조로, 입력 간 전역적인 상호 관계를 동시에 학습할 수 있도록 설계되었다. 그림 2는 이러한 Transformer 구조의 기본 구조를 보여주며, 왼쪽 블록은 인코더, 오른쪽 블록은 디코더이다. Transformer 기반 초고해상화 신경망의 대표적인 예로는 SwinIR^[7]과 HAT-L^[8] 등이 있다. SwinIR은 Swin Transformer 구조를 기반으로 하며, 이동 윈도우 기반 셀프 어텐션 기술을 통해 인접 윈도우 간의 연산 경계를 넘나드는 정보 흐름을 가능하게 하여 전역 문맥 정보를 효과적으로 통합한다. HAT-L은 하이브리드 어텐션 블록을 통해 채널 및 공간 어텐션을 동시에 결합하며 레이어 스케일 및 지역 강화 모듈(Local Enhancement Module)을 도입해 계산 효율성과 성능을 동시에 확보한 경량 Transformer 구조로, State-of-the-Art 성능을 달성하였다.

객체 탐지 분야에서는 다양한 방향으로 회전된 객체를 효과적으로 탐지하기 위한 신경망 구조가 연구되어 왔으며, 이러한 접근을 대표하는 신경망 중 하나로 Oriented R-CNN^[9]이 있다. Oriented R-CNN은 ARC-ResNet50을 기반으로 기본 특징을 추출하고, 이후 회전된 관심 영역 정렬과 각도 분류 모듈을 통해 객체의 회전 각도에 정렬된 특징 맵을 생성한다. 이 과정은 회전 박스에 맞춰 샘플링 위치를 재배치함으로써 회전된 객체 정보를 보다 정확히 반영한다. 이를 통해 위성 영상과 같이 객체의 방향이 다양하게 분포하는 복잡한 장면에서도 높은 탐지 성능을 보인다. 객체 분

할 분야에서는 U-Net^[10]이 널리 사용되고 있다. U-Net은 EfficientNet-B0에 기반으로, 인코더-디코더 대칭 구조를 기반으로 하며, 각 인코더 레이어에서 추출된 특징 맵을 디코더 레이어에 스킵 커넥션을 통해 맥락을 효과적으로 이해하고 정확한 위치 복원을 가능하게 하는 특징을 갖는다. VGG-16^[11]은 영상 인식 분야에서 널리 사용되는 사전 학습된 신경망으로, 13개의 합성곱 레이어와 3개의 완전 연결 레이어로 구성되어 있으며, 모든 합성곱은 3×3 필터를 사용하고, 각 블록 끝에 최대 풀링 연산이 포함되어 있다. 이러한 레이어 구조는 시각적 특징을 효율적으로 추출이 가능하다.

III. 제안 방법

본 연구에서는 객체 탐지 및 분할 작업에 최적화된 고해상도 영상을 생성하기 위해, 작업 특화 초고해상도 신경망 학습 구조를 제안한다. 기존의 초고해상도 기법은 고해상도 영상 복원 성능은 우수하지만, 복원된 영상이 실제 객체 탐지나 분할과 같은 작업 성능에 얼마나 기여하는지에 대한 반영은 구조적으로 제한되어 있다. 특히 학습 과정에서 작업 신경망을 포함하지 않는 경우, 복원 영상이 후속 작업에 최적화되지 않아 실질적인 성능 향상에 제약이 발생한다. 제안 방법은 이러한 한계를 해결하기 위해, 초고해상도 신경망 학습 과정에 작업 신경망과 VGG-16을 구조적으로 통합하고, 각각의 출력에서 도출한 작업 손실, 지각 손실,

재구성 손실을 결합한 가중합 손실함수로 초고해상도 신경망 학습을 수행한다. 제안 방법의 학습 시스템 블록도는 그림 3과 같다. 제안 학습 시스템 블록도는 학습 과정에서 작업 손실과 지각 손실을 초고해상도 신경망 학습에 활용하기 위해 작업 신경망과 지각 정보를 추출하기 위한 VGG-16을 초고해상도 신경망에 연동한 구조로 설계되었다. 학습에 활용되는 최종 손실 값은 재구성 손실, 지각 손실, 작업 손실의 가중 합을 계산하여 학습에 활용한다. 이를 통해 작업에 최적화된 고해상도 영상을 생성하며, 저해상도 영상을 재구성하는 과정에서 발생할 수 있는 의미적 정보 손실과 시각적 품질 저하를 효과적으로 보완할 수 있다.

구체적으로 학습 과정에서 전방향의 흐름은 그림 3의 검은색 화살표를 따라 진행된다. 먼저, LR 영상이 초고해상도 신경망에 입력되며, 신경망을 통해 사전에 설정된 스케일로 해상도가 확장되어 SR 영상이 생성된다. 만들어진 SR 영상은 작업 손실과 지각 손실 계산을 위해 작업 신경망과 VGG-16에 입력된다.

본 방법에서는 앞서 설명하였듯, 재구성 손실, 지각 손실, 작업 손실의 세 가지 손실 값을 결합한 수식 (1)의 최종 손실함수를 학습에 활용한다.

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \alpha\mathcal{L}_{task} + \beta\mathcal{L}_{per} \quad (1)$$

수식 (1)에서 \mathcal{L}_{rec} , \mathcal{L}_{task} , \mathcal{L}_{per} 는 각각 재구성 손실, 작업 손실, 지각 손실을 의미하며, α 와 β 는 작업 손실과 지각 손실이 최종 손실 값에 미치는 영향을 조절하는 하이퍼

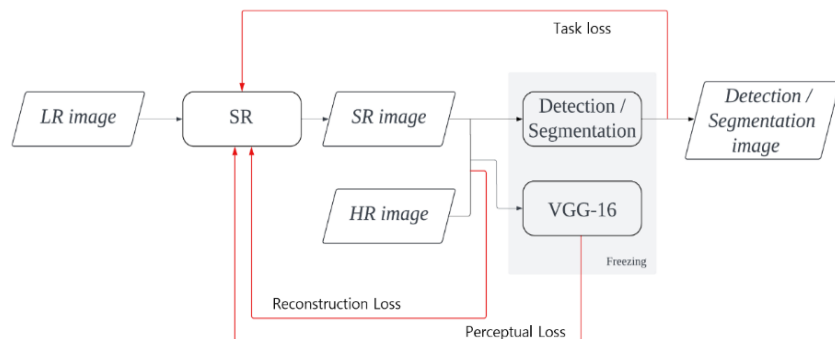


그림 3. 제안한 시스템 블록도
Fig. 3. Proposed system block diagram

파라미터이다. 세 손실 값을 결합한 \mathcal{L}_{total} 이 최종 초고해상화 신경망 학습에 활용된다.

먼저, 재구성 손실의 경우, SR 영상과 원본 HR 영상 간의 Mean Square Error(MSE) 손실을 사용하며, LR 영상의 세부 정보를 복원하는 데 초점을 맞춰 SR 영상이 HR 영상과 최대한 유사하도록 유도하는 역할을 한다.

작업 손실은 객체 탐지의 경우 Smooth L1 Loss, 객체 분할의 경우 교차 엔트로피(Cross-Entropy)를 사용한다. 이러한 작업 손실을 활용하면, SR 신경망이 작업의 성능을 최적화하기 위한 방향으로 학습이 수행되어, 작업에 특화된 고해상도 영상을 생성하도록 유도할 수 있다.

지각 손실은 수식 (2)와 같이 SR 영상과 원본 HR 영상을 VGG-16 신경망에 입력하여 신경망의 중간 특징 맵을 추출해 특징 맵 간의 MSE 손실을 계산해 활용한다.

$$\mathcal{L}_{per} = \frac{1}{N} \sum_{i=1}^N (\phi(X_{SR})^i - \phi(X_{HR})^i)^2 \quad (2)$$

수식 (2)에서 $\phi(X_{SR})^i$ 와 $\phi(X_{HR})^i$ 는 각각 SR 영상(X_{SR})과 원본 HR 영상(X_{HR})을 VGG-16 신경망(ϕ)의 입력으로 넣었을 때, i 번째 계층에서 추출한 특징 맵을 의미한다. 추출된 SR과 HR의 특징 맵 간 차이는 MSE를 이용해 지각 손실 값으로 활용된다. 지각 손실은 SR 영상과 원본 HR 영상 간의 시각적 품질과 의미적 정보를 보존하기 위해 사용된다.

본 논문에서 제안하는 작업 특화 초고해상화 신경망 학습 구조는 수식 (1)에서 정의된 최종 손실함수를 기반으로 학습이 수행된다. 이 과정에서 작업 신경망과 VGG-16 신경망은 학습되지 않으며, 이는 손실함수 계산 시 일관된 특징 추출과 작업 능력을 바탕으로 안정적인 손실 값을 얻기 위함이다.

IV. 실험 결과

제안 방법이 탐지 및 분할 작업 성능을 개선함을 증명하기 위해, 다양한 초고해상화 신경망에 제안하는 학습 구조

를 적용하여 실험을 진행하였다. 실험에서는 기존의 초고해상화 영상만을 작업 신경망에 입력한 경우와, 제안한 구조로 학습된 초고해상도 영상을 입력한 경우를 비교함으로써, 제안 방식이 탐지 및 분할 성능에 미치는 영향을 정량적으로 분석하였다. 초고해상화 신경망으로는 Convolution Neural Network 기반인 RCAN과 RFDN, Transformer 기반인 SwinIR과 HAT-L을 사용하여 저해상도 영상에 대해 4배 초고해상화하였으며, 객체 탐지 신경망으로는 Oriented R-CNN, 객체 분할 신경망으로는 U-Net을 사용하였다.

모든 실험은 Ubuntu 20.04 환경에서 수행되었으며, 하드웨어로는 GeForce RTX 3060 GPU를 사용하였다. 실험 데이터셋은 1024×1024 크기로 구축되었으며, 탐지 작업을 위해 DOTA dataset^[12] 2,045장, 분할 작업을 위해 Inria dataset^[13] 1,945장을 사용하였다. 저해상도 영상은 고해상도 영상에 Gaussian Blur를 적용하여 텍스처와 세부 정보를 제거한 후, Bicubic Interpolation을 사용하여 4배 다운샘플링하여 생성하였다. 모델 학습에는 Batch size는 3, Epoch는 100, 학습률은 1e-4, Optimizer는 Radam, Learning rate scheduler는 PolyLR로 지정하였다. 성능 평가는 초고해상화된 영상의 품질을 평가하기 위한 지표로 Peak Signal-to-Noise Ratio(PSNR)^[14]를 활용하였으며, 또한 작업에 대한 성능 향상을 확인하기 위해 탐지 작업에 대해 mean Average Precision(mAP)^[15]를, 분할 작업의 성능 측정을 위해 Accuracy(ACC)^[16]와 mean Intersection over Union(mIoU)^[17]를 기준으로 사용하였다.

표 1은 저해상도 영상(LR), 기존 방식의 초고해상화 영상(SR), 제안된 방법의 초고해상화 영상(Ours)에 대해 Oriented R-CNN으로 탐지한 실험 결과를 보여준다. 실험 결과, 재구성 손실만을 활용하는 기존의 초고해상화 신경망보다 제안된 결합 손실함수의 학습 구조를 적용한 경우에 일관되게 높은 객체 탐지 성능을 달성하였다. 또한 HAT-L을 제외한 모든 초고해상화 신경망에서 PSNR의 향상됨을 확인할 수 있다. 특히, RFDN 신경망에서 제안 방법을 적용했을 때 mAP가 기존 대비 5.8% 향상되어 가장 높은 탐지 성능인 mAP 85.6%를 기록하였다. 그림 4의 상단은 각 초고해상화 방법의 객체 탐지 작업의 결과를 시각화한 것이다. 저해상도 영상과 기존 초고해상화 신경망보다 제안 방법을 적용했을 때, 더 많은 객체를 탐지한 것을 확인

할 수 있다. 이 결과는 제안한 신경망 구조가 초고해상화 과정에서 영상의 시각적 품질을 개선함과 동시에 탐지 작업에 필요한 의미적 정보를 효과적으로 반영했음을 보여준다.

표 1. 저해상도 영상(LR), 기존 모델로 4배 초고해상화한 영상(SR), 제안된 방법으로 4배 초고해상화한 영상(Ours)에 대해 Oriented R-CNN으로 탐지한 결과

Table 1. Detection results using Oriented R-CNN for low-resolution images (LR), 4 × super-resolved images by conventional models (SR), and 4 × super-resolved images by the proposed method (Ours)

Model	Method	PSNR	mAP(%)
RFDN	LR	-	38.8
	SR	27.038	79.8
	Ours	29.523	85.6
RCAN	SR	27.161	79.3
	Ours	29.252	83.5
SwinIR	SR	26.98	74.7
	Ours	29.204	81.5
HAT-L	SR	30.574	84.2
	Ours	30.053	84.7

표 2는 저해상도 영상(LR), 기존 방식의 초고해상화 영

표 2. 저해상도 영상(LR), 기존 모델로 4배 초고해상화한 영상(SR), 제안된 방법으로 4배 초고해상화한 영상(Ours)에 대해 UNet으로 분할한 결과
Table 2. Segmentation results using U-Net for low-resolution images (LR), 4 × super-resolved images by conventional models (SR), and 4 × super-resolved images by the proposed method (Ours)

Model	Method	PSNR	ACC(%)	mIoU(%)
RFDN	LR	-	76.5	41.9
	SR	24.4	82.2	42.1
	Ours	24.587	87	59.7
RCAN	SR	24.465	82.2	42.1
	Ours	24.452	89.3	69.8
SWINIR	SR	25.05	81.8	42.2
	Ours	24.177	90.7	71.4
HAT-L	SR	26.693	81.7	42.5
	Ours	23.108	93.3	78.7

상(SR), 제안된 방법의 초고해상화 영상(Ours)에 대해 UNet으로 분할한 실험 결과를 보여준다. 실험 결과, PSNR 평가에서는 RFDN을 제외한 모든 초고해상화 신경망에서 제안한 신경망 구조가 재구성 손실만 사용하는 초고해상화 신경망보다 다소 낮은 값을 기록하였으나, ACC와 mIoU 지표에서는 모든 초고해상화 신경망에서 일관되게 성능이



그림 4. DOTA-v1.0 데이터셋의 p0128 이미지에 대한 탐지 결과 시각화 (상), Inria 데이터셋의 vienna2 이미지에 대한 분할 결과 시각화 (하)

Fig. 4. Visualization of detection results on the p0128 image from the DOTA-v1.0 dataset (top) and segmentation results on the vienna2 image from the Inria dataset (bottom)

향상되었다. 특히, HAT-L 신경망에서 제안 방법을 적용했을 때 기존 대비 ACC가 11.6%, mIoU가 36.2% 향상되어 가장 높은 분할 성능인 ACC 93.3%, mIoU 78.7%를 기록하였다. 그림 4의 하단은 각 초고해상화 방법의 객체 분할 작업의 결과를 시각화한 것이다. 저해상도 영상과 기존 초고해상화 신경망보다 제안 방법을 적용했을 때, 객체를 정확하게 분할한 것을 확인할 수 있다. 이는 제안한 신경망 구조가 초고해상화를 통해 저해상도 영상을 분할 작업에 특화된 고해상도 영상으로 재구성하였음을 시사한다.

V. 결 론

본 연구에서는 작업에 특화된 초고해상화 신경망의 학습을 위해 재구성 손실, 작업 손실, 지각 손실을 결합한 새로운 손실함수를 학습에 활용하는 작업 특화 초고해상화 신경망 학습 구조를 제안하였다. 이를 위해 작업 신경망과 VGG-16을 초고해상화 신경망에 연동하여 추가적인 손실을 측정하여 초고해상화 학습에 적용하였다. 실험 결과, 제안한 방법으로 학습한 초고해상화 신경망으로 생성한 고해상도 영상이 기존 초고해상화 신경망으로 생성한 고해상도 영상을 입력으로 사용할 때보다 모든 작업 성능 지표에서 향상된 결과를 보였으며, 이를 통해 제안한 방법이 탐지 및 분할 작업에 적합한 고해상도 영상을 효율적으로 생성할 수 있음을 보여준다. 향후 연구에서는 현재 실험적으로 설정된 결합 손실함수의 가중치를 자동화하여 최적의 조합을 탐색하는 방법을 개발하는 데 초점을 맞추어 예정이다. 이를 통해 모델 학습의 효율성을 더욱 높이고, 성능 개선 가능성을 극대화할 수 있을 것으로 기대된다.

참 고 문 헌 (References)

- [1] Haris, M., Shakhnarovich, G., and Ukita, N., "Task-Driven Super Resolution: Object Detection in Low-resolution Images," arXiv preprint arXiv:1803.11316, submitted Mar. 30, 2018. doi: <https://doi.org/10.48550/arXiv.1803.11316>
- [2] Wang, B., Lu, T., and Zhang, Y., "Feature-Driven Super-Resolution for Object Detection," arXiv preprint arXiv:2004.00554, submitted Apr. 1, 2020. doi: <https://doi.org/10.48550/arXiv.2004.00554>
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," arXiv preprint arXiv:1706.03762, 2017. <https://arxiv.org/abs/1706.03762>
- [4] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner "GradientBased Learning Applied to Document Recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278 - 2324, Nov. 1998. <https://ieeexplore.ieee.org/document/726791>
- [5] Liu, J., Tang, J., & Wu, G. "Residual Feature Distillation Network for Lightweight Image Super-Resolution," Proceedings of the European Conference on Computer Vision Workshops (ECCVW), 2020, pp. 1-16, September 2020. <https://arxiv.org/abs/2009.11551>
- [6] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," Proceedings of the European Conference on Computer Vision (ECCV), vol. 11211, pp. 286-301, September 2018. <https://arxiv.org/abs/1807.02758>
- [7] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. "SwinIR: Image Restoration Using Swin Transformer," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2021, pp. 1833-1844, October 2021. <https://arxiv.org/abs/2108.10257>
- [8] Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., & Liu, Z. "Activating More Pixels in Image Super-Resolution Transformer," arXiv preprint arXiv:2205.04437, May 2022. <https://arxiv.org/abs/2205.04437>
- [9] Xie, X., Che, G., Wang, J., Yao, X., & Han, J. "Oriented R-CNN for Object Detection," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 3520-3529, October 2021. <https://arxiv.org/abs/2108.05699>
- [10] Ronneberger, O., Fischer, P., & Brox, T. "U-Net: Convolutional Networks for Biomedical Image Segmentation," Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 9351, pp. 234-241, October 2015. <https://arxiv.org/abs/1505.04597>
- [11] Simonyan, K., & Zisserman, A. "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, submitted Sep. 4, 2014, last revised Apr. 10, 2015. doi: <https://doi.org/10.48550/arXiv.1409.1556>
- [12] Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M.-Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7778-7796, Nov. 2022. doi: <https://doi.org/10.1109/TPAMI.2021.3117983>
- [13] Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. "Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, pp. 3226 - 3229. doi: <http://doi.org/10.1109/IGARSS.2017.8127684>
- [14] Huynh-Thu, Q., & Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment. Electronics Letters, 44(13), 800 - 801.

doi: <http://doi.org/10.1049/el:20080522>

- [15] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303 - 338.
doi: <http://doi.org/10.1007/s11263-009-0275-4>
- [16] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*.

Springer. <https://oa.mg/work/10.1007/978-0-387-45528-0>

- [17] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2011). The PASCAL Visual Object Classes (VOC) Challenge 2011 (VOC2011) Results. <https://link.springer.com/article/10.1007/s11263-009-0275-4>

저 자 소 개



서 예 원

- 2025년 : 국립한밭대학교 정보기술대학 정보통신공학과 학사
- 2025년 ~ 현재 : 국립한밭대학교 소프트웨어융합대학원 지능미디어공학과 석사과정
- ORCID : <https://orcid.org/0009-0002-3443-4774>
- 주관심분야 : 특징 압축, 기계 학습, 딥러닝



한 승 석

- 2024년 : 국립한밭대학교 정보기술대학 정보통신공학과 학사
- 2024년 ~ 현재 : 국립한밭대학교 소프트웨어융합대학원 지능미디어공학과 석사과정
- ORCID : <https://orcid.org/0009-0008-3443-0322>
- 주관심분야 : 컴퓨터 비전, 기계 학습, 딥러닝



최 해 철

- 1977년 : 경북대학교 전자공학과 학사
- 1999년 : 한국과학기술원 전기및전자공학과 석사
- 2004년 : 한국과학기술원 전기및전자공학과 박사
- 2004년 ~ 2010년 : 한국전자통신연구원 방송미디어연구부 선임연구원
- 2010년 ~ 현재 : 국립한밭대학교 정보기술대학 지능미디어공학과 교수
- ORCID : <https://orcid.org/0000-0002-7594-0828>
- 주관심분야 : 비디오 부호화, 컴퓨터 비전, 딥러닝