

헤시안 기반 Grad-CAM++와 웨이블릿 융합 멀티스케일 활성화 맵으로 고정밀 이미지 복원

권태윤*, 노광현**

High-Precision Image Restoration via Hessian-based Grad-CAM++ and Wavelet-Fused Multiscale Activation Maps

Taeyoun Kwon*, Kwang Hyun Ro**

요 약

딥러닝 기반 이미지 복원 기술은 의미론적 정보와 주파수 정보를 융합하여 고품질 이미지 복원을 목표로 주목받고 있다. 그러나 Grad-CAM++는 단일 해상도에서 생성된 활성화 맵에 의존하여 복잡한 텍스처 정보를 충분히 반영하지 못하며, 이는 클래스별 위치 정밀도 저하를 유발한다. 본 연구는 ResNet-50 기반 HDM-WaveNet(High-Dimensional Multi-scale WaveNet)을 제안하여 이러한 한계를 극복한다. 헤시안 기반 Grad-CAM++는 2차 편미분을 활용해 클래스별 가중치를 정밀화하며, HDM-WaveNet은 헤시안 기반 Grad-CAM++와 웨이블릿 변환을 결합해 다중 스케일에서 의미론적 및 주파수 정보를 정밀히 통합, 고정밀 이미지 복원을 목표로 한다. 이를 통해 HDM-WaveNet은 멀티스케일 활성화 맵을 생성, 스케일 간 일관성을 유지하며 왜곡을 최소화한다. 복원 성능은 ImageNet과 Pascal VOC 데이터셋에서 IoU, PSNR, SSIM 지표로 평가, 기존 기법 대비 10% 향상된 고정밀 복원을 달성하여 딥러닝 기반 이미지 복원의 실용성을 강화했다.

Abstract

Deep learning-based image restoration techniques, which integrate semantic and frequency information, are gaining attention for achieving high-quality image restoration. However, Grad-CAM++ uses single-resolution activation maps, missing complex textures and reducing class localization accuracy. HDM-WaveNet combines Hessian-based Grad-CAM++ with wavelet transform to precisely integrate semantic and frequency information across multiple scales, aiming for high-precision image restoration. HDM-WaveNet generates multiscale activation maps, ensuring cross-scale consistency and minimizing distortion. Restoration performance, evaluated on ImageNet and Pascal VOC datasets using IoU, PSNR, and SSIM metrics, achieves a 10% improvement in high-precision restoration over existing methods, enhancing the practicality of deep learning-based image restoration.

Keywords

multi-scale features, wavelet transform, high frequency information, Hessian Weighting, HDM-WaveNet, IoU

* 한성대학교 일반대학원 스마트융합컨설팅학과
- ORCID: <https://orcid.org/0009-0007-9763-0138>
** 한성대학교 AI융용학과 교수(교신저자)
- ORCID: <https://orcid.org/0009-0002-9391-6052>

• Received: Jun. 19, 2025, Revised: Jul. 02, 2025, Accepted: Jul. 05, 2025
• Corresponding Author: Kwang Hyun Ro
Professor, Dept. of AI Applications, Hansung University
Tel.: +82-2-760-8013, Email: khrho@hansung.ac.kr

1. 서론

이미지 기반 인공지능은 이미지 품질 향상, 의료 영상 진단, 자율주행 등 다양한 분야에서 특징 추출과 분석 정확도를 향상시키는 핵심 기술로 자리 잡았다. 특히, 이미지 품질 향상의 핵심 기술인 이미지 업스케일링은 저해상도 이미지를 고해상도로 변환하여 선명한 디테일을 제공하며, 특수 의료 영상 분야에서는 PET CT나 MRI 이미지의 고품질 복원을 통해 정확도를 30% 이상 향상 시킨 사례가 보고되었다. 자율주행에서는 이미지 업스케일링을 통해 저해상도 센서 데이터를 고해상도로 복원하여 장애물 인식과 경로 계획의 신뢰성을 높였다. 이러한 고해상도 이미지 처리 기술의 발전은 현대 산업 생태계에서 안전성과 효율성을 보장하는 핵심 요소로 작용한다. 그러나 저해상도 이미지의 품질 개선에는 여전히 기술적 한계가 존재하며, 특히 복잡한 멀티스케일 환경에서 정밀한 복원이 요구된다.

저해상도 이미지 복원의 주요 한계는 단일 스케일에 초점을 맞춘 기존 알고리즘의 제약에서 비롯된다[1][2]. 단일 스케일 접근은 이미지의 전반적인 구조적 정보(저주파 성분)와 세부 디테일(고주파 성분) 간의 균형을 효과적으로 처리하지 못하며, 이는 복잡한 이미지 특징을 정밀히 복원하는 데 장애가 된다. 이러한 문제는 이미지 업스케일링, 의료 영상 진단, 자율주행과 같이 정밀한 복원이 필수적인 다양한 응용에서 중요한 제약으로 작용한다. 예를 들어, Grad-CAM++는 클래스 활성화 맵을 통해 의미론적 정보를 효과적으로 포착하지만, 텍스처와 디테일과 같은 고주파 성분을 충분히 반영하지 못한다. 반면, 웨이블릿 변환은 저주파 구조와 고주파 엣지 정보를 분리하여 디테일을 강조하지만, 의미론적 맥락과의 통합이 미흡하여 전체적인 이미지의 맥락적 일관성을 유지하기 어렵다. 또한, 학습 데이터 편향으로 인한 도메인 간 일반화 부족은 실제 응용 환경(예: 다양한 의료 영상 장비나 다양한 조명 조건 하에서의 이미지 업스케일링)에서 성능 저하를 초래한다.

이러한 한계는 의미론적 정보와 텍스처 및 디테일 간의 균형 부족, 그리고 멀티스케일 환경에서의 통합적 특징 분석 부족에서 기인한다. 기존 방법들

은 단일 스케일 기반의 특징 추출에 의존하거나, 특정 도메인에 최적화된 학습 데이터에 과도하게 의존하여 일반화 성능이 떨어진다. 이에 본 연구는 Grad-CAM++의 단일 스케일 제약을 극복하기 위해 멀티스케일 융합 이미지 복원 기법을 제안한다. 제안된 기법은 Grad-CAM++를 기반으로 고차 편미분 절대값 특징을 추출하고, 2차 편미분 가중치는 헤시오안 행렬을, 3차 및 4차 편미분 가중치는 Haar 웨이블릿 변환을 통해 저주파와 고주파 대역으로 분리한다. 이를 바탕으로, 본 연구는 HDM-WaveNet을 통해 멀티스케일 기반 클래스별 활성화 맵을 활용하여 엣지와 텍스처를 정밀히 복원함으로써, 이미지 업스케일링 및 객체 인식과 같은 다양한 응용 환경에서 고해상도 복원의 품질과 스케일 간 일관성을 향상시키는 것을 목표로 한다.

이 연구의 주요 성과는 아래와 같이 정리할 수 있다. 첫째, HDM-WaveNet은 Grad-CAM++의 단일 스케일 한계를 극복한 멀티스케일 이미지 복원 기법이다. 헤시오안 행렬로 곡률을 정량화하고, Haar 웨이블릿으로 저·고주파 대역을 분리해 클래스별 활성화 맵을 생성, 정밀 복원과 왜곡을 최소화하였다. 둘째, 멀티스케일 가중치 정규화와 조화 함수로 다양한 서브밴드 주파수를 통합해 정밀한 클래스별 활성화 맵을 생성한다. 초기 레이어의 고주파로 지역 패턴(엣지, 텍스처), 깊은 레이어의 저주파로 전역 패턴(객체 구조)을 포착하며, 가우시안 평활화와 고유값 기반 가중치로 곡률 강도를 강조해 Grad-CAM++보다 시각화 품질을 향상시켰다. 실험 결과, 멀티스케일 통합과 고차 편미분 기반 특징 추출로 지역·전역 패턴을 균형 있게 반영한 결과, HDM-WaveNet은 ImageNet(96.90%)과 Pascal VOC (95.70%)에서 Grad-CAM(92.50%, 90.40%), Grad-CAM++(94.30%, 92.70%)를 상회하였다.

본 논문의 구성은 다음과 같다. 2장에서는 Grad-CAM++과 웨이블릿 변환의 특징과 한계를 분석, 통합적 접근 필요성을 제기한다. 3장에서 멀티스케일 융합 설계와 구현을 다룬다. 프레임워크 개요, 고차 편미분 기반 특징 탐지, 멀티스케일 가중치 적용, 멀티스케일 통합을 설명한다. 4장은 실험 결과와 분석을, 5장은 연구의 성과 및 방향 제시한다.

II. 관련 연구

2.1 Grad-CAM++의 특징 및 한계

Grad-CAM[1]은 1차 그래디언트를 사용해 클래스 활성화 맵을 생성하지만, 단일 스케일 접근으로 고주파 디테일 포착이 제한된다. Grad-CAM++[3]는 이를 개선하여 클래스별 가중치 α_k^c 는 다음 식 (1)에 따라 계산한다.

$$\alpha_k^c = \sum_{i,j} \frac{\partial^2 y^c}{\partial A_{ij}^k \partial A_{ij}^k} \cdot ReLU\left(\frac{\partial y^c}{\partial A_{ij}^k}\right) \quad (1)$$

이 가중치는 픽셀별 2차 기울기를 활용해 클래스에 대한 레이어 k 의 중요도를 정량화한다. 클래스 활성화 맵은 식 (2)로 생성된다.

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (2)$$

이 수식은 가중치 α_k^c 와 활성화 맵 A^k 를 결합해 클래스(c)에 대한 주목 영역을 시각화한다. 그러나 Grad-CAM++는 단일 스케일 접근으로 인해 고주파 디테일(예: 텍스처, 엣지)을 충분히 반영하지 못하며, 다중 객체 환경에서 클래스 간 경계 구분이 모호해질 수 있다[3][4].

본 논문은 이러한 한계를 극복하고자 HDM-WaveNet을 제안하며, 전역적 민감도 $w_k^{(2)}$ 와 정규화된 특징 맵을 통해 멀티스케일에서 의미론적 정보를 통합, Grad-CAM++의 단일 스케일 한계와 레이어 의존성을 극복하여 복잡한 장면에서도 정밀한 시각화를 달성한다.

표 1은 제안 방법과 기존 Grad-CAM, Grad-CAM++[1][3]을 비교한 결과이다. 기존 방법들은 단일 스케일의 그래디언트 정보에 의존하여 시각화 결과가 가장 판별적인 지역에 국한되는 한계가 있다. 이로 인해 객체의 전체적인 윤곽을 파악하기 어렵다. 반면, 제안 방법은 다중 스케일 접근을 통해 이미지의 거시적 맥락과 미시적 특징을 모두 반영하며, 정규화된 특징 맵으로 객체 전체를 보다 안정적으로 표현함으로써 기존의 단점을 개선한다.

표 1. 컨볼루션 레이어 활성화 기반 시각화 방법 비교
Table 1. Comparison of convolutional layer activation-based visualization methods

Method	Activation usage	Scale approach	Main limitations
Grad-CAM[1]	First-order gradients	Single-scale	
Grad-CAM++ [3]	Second-order gradients	Single-scale	Limited high-frequency detail capture
Ours	Global sensitivity $w_k^{(2)}$, normalized feature maps	Multi-scale	

2.2 웨이블릿 변환의 특징 및 한계

웨이블릿 변환은 시간-주파수 도메인에서 신호를 분석하는 기법으로, 푸리에 변환과 달리 국소적 특징을 다양한 스케일에서 효과적으로 포착한다[4]. 이는 식 (3)로 정의된다.

$$W_\psi(a,b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt \quad (3)$$

여기서 ψ 는 웨이블릿 함수, a 는 스케일, b 는 이동 파라미터를 나타낸다. 웨이블릿 변환은 신호를 다중 해상도로 분해하여 이미지나 신호의 국소적 특징, 특히 엣지와 텍스처를 다양한 스케일에서 정밀히 추출하며, 노이즈가 많은 데이터에서 신호의 주요 성분을 보존하면서 노이즈를 효과적으로 제거한다. 이는 스케일별로 저주파(전체적 구조)와 고주파(세부 패턴) 성분을 분리하여, 예를 들어 Haar 웨이블릿은 간단한 엣지 탐지에, Daubechies 웨이블릿은 복잡한 텍스처 분석에 적합하다[5]. 또한, 웨이블릿의 적응적 스케일링은 신호의 비정상적 주파수 변화를 정밀히 포착하며, 이산 웨이블릿 변환(DWT)은 다중 해상도 처리를 통해 계산 효율성을 높여 대규모 데이터(예: 고해상도 이미지)에 유리하다. 특히, DWT는 계층적 분해를 통해 각 스케일에서 특징을 독립적으로 분석하며, 웨이블릿 패킷 분해는 특정 주파수 대역을 세분화하여 더 정밀한 분석을 가능하게 한다. 웨이블릿의 직교성은 신호를 에너지 보존 방식으로 분해하여 계수 간 상관성을 최소화하

며, 멀티 레벨 분해 구조는 주파수 대역별로 독립적인 분석을 지원하여 신호의 세부 패턴을 정밀히 탐지한다. 예를 들어, Morlet 웨이블릿은 주파수 변화가 큰 신호에 적합하며, Coiflet 웨이블릿은 부드러운 신호 재구성에 유리하다. 이러한 특징은 이미지 압축이나 특징 추출과 같은 응용에 활용된다. 그러나 웨이블릿 변환은 CNN의 의사결정 과정과 같은 의미론적 시각적 설명을 제공하지 못하여 모델 해석 가능성에 직접 기여 하지 못한다.

2.3 통합적 접근의 필요성

기존 연구에서 헤시안과 웨이블릿은 멀티스케일 기법으로 각각 통합하지 않고 시도 되었으나 한계가 있었다. 예를 들어, 헤시안은 국소적 패턴을 효과적으로 검출하나 전역 구조 분석이 부족했다. 웨이블릿 변환은 멀티스케일로 엣지와 텍스처를 추출하고 노이즈를 제거하지만, CNN 시각화에는 제약이 있다[4][6]. Grad-CAM++는 클래스별 특징을 시각화하나, 단일 스케일로 지역 및 전역 패턴을 통합적으로 분석하기 어렵다. 단일 기법으로는 복잡한 이미지 분석의 요구를 충족하기 어렵다.

이에 따라, 특징 일관성을 유지하며 지역·전역 패턴을 통합적으로 분석하기 위해 Grad-CAM++와 웨이블릿을 융합한 HDM-WaveNet을 제안한다. 본 모델은 헤시안의 2차 편미분을 활용해 Grad-CAM++를 다중 스케일로 확장하고, 웨이블릿의 3·4차 서브밴드를 통해 주파수별 특징을 추출하여 지역 및 전역 구조를 정밀히 복원한다. 이 통합 접근법은 곡률 정량화와 주파수별 특징 결합을 통해 구현되며, 정밀한 이미지 복원과 복잡한 장면 분석을 가능하게 한다.

III. 멀티스케일 융합 설계와 구현

3.1 프레임워크 개요

본 연구는 멀티스케일 클래스 활성화 맵으로 생성된 고품질 이미지 복원 알고리즘인 High-Dimensional Multi-scale WaveNet(HDM-WaveNet)은 Grad-CAM++의 단일 스케일 한계와 웨이블릿 변환의 공

간적 한계를 극복하여 멀티스케일 융합 이미지를 생성하는 프레임워크로, 복잡한 장면에서 정밀한 이미지 복원과 클래스별 시각화 정밀도를 달성한다. 이는 2.3절에서 논의된 통합 필요성을 구현하며, 그림 1과 같이 세 단계로 구성된다.

첫 번째 단계에서는 고차 편미분(2차, 3차, 4차)을 활용해 특징을 정밀히 탐지한다. Grad-CAM++에 절대값 2차 편미분으로 큰 곡률(예: 텍스처 패턴), 3차 편미분으로 중간 스케일 패턴(예: 엣지 변화, 코너), 4차 편미분으로 세부 패턴(예: 미세 곡률)을 포착하여 양수/음수 곡률 상쇄를 방지한다. 이는 다중 객체와 복잡한 배경에서도 의미론적 특징을 다중 스케일로 분석하여 클래스별 가중치를 정의한다.

두 번째 단계에서는 멀티스케일 가중치를 적용한다. 먼저, 헤시안 행렬의 절대값 2차 편미분으로 Grad-CAM++의 가중치를 정밀화하여 클래스별 중요도를 향상시킨다. 이어서, 웨이블릿 변환으로 이미지를 저주파와 고주파 성분으로 분해하고, 3·4차 편미분 가중치를 서브밴드에 적용하여 주파수별 특징을 추출한다. 이 가중치는 저주파의 구조적 안정성(예: 객체 형태)과 고주파의 지역 디테일(예: 엣지, 텍스처)을 강조하며, 웨이블릿의 다중 해상도 특성을 활용해 노이즈를 제거한다. 주파수 대역별 중요도를 적응적으로 조정하여 특징 손실을 최소화한다.

세 번째 단계에서는 멀티스케일 특징을 통합하여 융합 이미지를 생성한다. 헤시안 기반 가중치와 웨이블릿 서브밴드 특징을 결합하여 지역 패턴(엣지·텍스처)과 전역 구조(객체 형태)를 균형 있게 복원하고, 클래스별 활성화 맵을 생성한다. 예를 들어, 웨이블릿으로 분리된 주파수별 특징을 헤시안 기반 Grad-CAM++로 시각화하여 의미론적 구조와 세부 디테일을 정밀히 복원, 멀티스케일 융합 이미지를 생성한다. 이는 다중 객체 환경에서도 클래스별 기여도를 명확히 구분한다. HDM-WaveNet은 이 세 단계로 멀티스케일 융합 설계를 구현하여 Grad-CAM++ 대비 향상된 시각화 정밀도와 정밀한 이미지 복원을 달성한다.

그림 1은 본 논문에서 제안하는 HDM-WaveNet Framework(하이브리드 알고리즘) 전체 구조를 나타낸다.

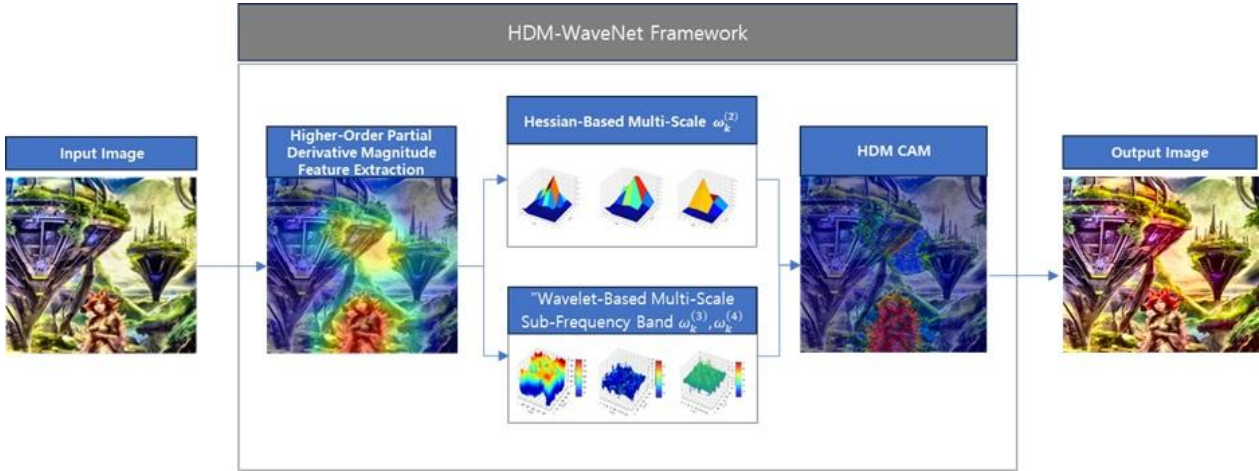


그림 1. HDM-WaveNet의 프레임워크 구조
Fig. 1. Framework structure of HDM-WaveNet

본 프레임워크는 Grad-CAM++의 레이어 의존성과 노이즈 민감성을 극복하는 세 단계로 구성된다. (1)특징 추출: 고차 편미분 절대값으로 특징 추출한다. (2)멀티스케일 가중치 적용: 2차 편미분은 헤시안 멀티스케일에 적용되고, 3차 및 4차 편미분은 웨이블릿 변환의 주파수 분해를 통해 가중치를 부여 받는다. 이후 웨이블릿 멀티스케일에 3차 및 4차 편미분에서 추출된 가중치를 적용적으로 적용한다. (3)멀티스케일 통합: 헤시안과 웨이블릿 멀티스케일을 통합하여 최종 High Dimensional Multi-scale WaveNet(HDM-WaveNet)을 생성한다.

3.2 고차 편미분 기반 특징 탐지

Grad-CAM++는 클래스 출력 y^C 의 1차 그래디언트로 특징 맵 가중치를 계산해 활성화 영역을 시각화한다. 그러나 1차 그래디언트는 양수/음수 곡률 상쇄로 복잡한 특징(예: 텍스처, 엣지)을 놓친다. 본 방법은 고차 편미분 절대값으로 가중치를 정의해 상쇄를 방지한다. 여기서 y^C 는 클래스 C 출력, A_{ij}^k 는 k 번째 특징 맵의 (i,j) 값이다. 특징 추출은 고차 편미분을 통해 수행됩니다. 절대값 2차 편미분맵은 클래스 출력에 대한 2차 미분의 절대값으로, 텍스처의 곡률 크기를 반영한다. 3차($w_k^{(3)}$)는 엣지/코너의 변화율, 4차($w_k^{(4)}$)는 미세 곡률을 포착하여 미세 곡률을 포착하여 더 복잡한 텍스처 패턴을 추출한다. 이 절대값 맵은 상쇄를 방지하며, 아키텍처에서 정

의된 고차 미분($w_k^{(2)}, w_k^{(3)}, w_k^{(4)}$)과 일관성을 유지한다. 이러한 과정은 그림 2 텍스처 복잡성 및 방향성 규칙성을 위한 제안된 HDM-CAM 아키텍처에서 특징 추출 단계를 시각적으로 나타내며, 이후 멀티스케일 가중치 적용으로 이어진다.

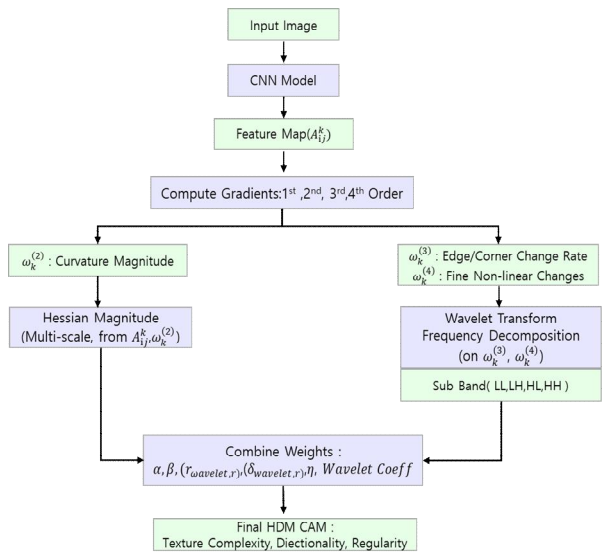


그림 2. 텍스처 복잡성 및 방향성 규칙성을 위한 제안된 HDM-CAM 아키텍처
Fig. 2. Proposed HDM-CAM architecture for texture complexity and directionality regularity

3.3 멀티스케일 가중치 적용

이 섹션은 가중치 계산에 초점을 맞춘다. 멀티스케일 가중치 적용은 클래스 점수 y^C (모델의 클래스

C에 대한 출력)와 특징 맵 A_{ij}^k (k 번째 컨볼루션 레이어의 활성화, 공간 위치(ij))를 기반으로 고차 편미분 값을 다양한 스케일에서 분석하여 특징의 기여도를 정밀하게 조정한다[6]. 2차 편미분은 헤시안 행렬의 절대값과 고유값을 활용하여 가중치 $w_k^{(2)}$ 를 계산하고, 3차 및 4차 편미분은 y^C 에 대한 절대값 기반 가중치 $w_k^{(3)}, w_k^{(4)}$ 를 계산한 후, Haar 웨이블릿 변환을 통해 고주파 대역(LH, HL, HH)을 포함한 가중치 $\gamma_{wavelet,k}, \delta_{wavelet,k}$ 를 생성한다.

3.3.1 절대값 2차 편미분 기반 헤시안 가중치

2차 편미분은 y^C 에 대한 특징 맵 A_{ij}^k 의 곡률 정보를 제공한다. Grad-CAM++를 확장하여, 멀티스케일 분석을 위해 가우시안 평활화를 적용한다. 절대값 기반 헤시안 행렬은 식 (4)로 정의된다.

$$H_k^{abs}(i, j; \sigma) = \text{Gaussian} \left(\begin{array}{c|c} \left| \frac{\partial^2 y^C}{\partial (A_{ij}^k)^2} \right| & \frac{\partial^2 y^C}{\partial A_{ij}^k \partial A_{ij+1}^k} \\ \hline \frac{\partial^2 y^C}{\partial A_{ij+1}^k \partial A_{ij}^k} & \left| \frac{\partial^2 y^C}{\partial (A_{ij+1}^k)^2} \right| \end{array}, \sigma \right) \quad (4)$$

- A_{ij}^k : k 번째 컨볼루션 레이어의 활성화, 위치(i, j)
- $\frac{\partial^2 y^C}{\partial (A_{ij}^k)^2}$: y^C 에 대한 2차 편미분
- $\text{Gaussian}(\cdot, \sigma)$: 스케일 σ 의 가우시안 평활화

주대각선의 절대값은 y^C 에 대한 변화율 강도를 강조하며, 비대각선은 인접 활성화 간 상호작용을 반영한다. 가중치 $w_k^{(2)}$ 는 헤시안 행렬의 고유값을 활용하여 곡률 강도를 식 (5)로 정량화한다.

$$w_k^{(2)} = \frac{1}{H \cdot W} \sum_{ij} \max_{\sigma} (|\lambda_1(i, j; \sigma)| + |\lambda_2(i, j; \sigma)|) \quad (5)$$

- λ_1, λ_2 : $H_k^{abs}(i, j; \sigma)$ 의 고유값
- $H \cdot W$: 특징 맵 크기
- \max_{σ} : 멀티스케일에서 최대 응답 선택

정규화된 특징 맵은 식 (6)로 정의한다.

$$H_{norm}(i, j; s) = w_k^{(2)} \cdot H_k^{abs}(i, j; \sigma_s) \quad (6)$$

전역적 민감도 $w_k^{(2)}$ 는 레이어 중요도를 효율적으로 계산하며, 정규화된 특징 맵 $H_{norm}(i, j; s)$ 와 함께 활성화 중요도를 식 (7)로 분석한다.

$$w_k^{(2)} = \frac{1}{Z} \sum_{ij} \left| \frac{\partial^{(2)} y^C}{\partial (A_{ij}^k)^{(2)}} \right| \quad (7)$$

여기서 Z 는 정규화 상수(예: 활성화 수)이며, 주대각선 요소를 활용해 계산한다.

3.3.2 3차 및 4차 편미분에 대한 웨이블릿 변환

3차 및 4차 편미분은 y^C 에 대한 A_{ij}^k 의 고차 민감도를 포착하여 엣지/코너 변화율과 미세 곡률을 식 (8)로 분석한다.

$$w_k^{(3)} = \frac{1}{Z} \sum_{ij} \left| \frac{\partial^{(3)} y^C}{\partial (A_{ij}^k)^{(3)}} \right|, \quad w_k^{(4)} = \frac{1}{Z} \sum_{ij} \left| \frac{\partial^{(4)} y^C}{\partial (A_{ij}^k)^{(4)}} \right| \quad (8)$$

이 가중치는 Grad-CAM++의 고차 민감도를 직접 반영한다. 이후, Haar 웨이블릿 변환을 적용하여 이미지를 저주파(LL)와 고주파(LH, HL, HH) 대역으로 분해한다.

각 스케일 j 와 대역 k 에서 $w_k^{(3)}, w_k^{(4)}$ 를 웨이블릿 변환하여 가중치 $\gamma_{wavelet,k}, \delta_{wavelet,k}$ 를 식 (9)로 계산한다.

$$\begin{aligned} \gamma_{wavelet,k} &= \lambda_k^{(3)} \cdot \text{Wavelet}(w_k^{(3)}, k), \gamma_{wavelet,k} \\ &= \lambda_k^{(4)} \cdot \text{Wavelet}(w_k^{(4)}, k) \end{aligned} \quad (9)$$

여기서 $\lambda_k^{(3)}, \lambda_k^{(4)}$ 는 고주파 대역(LH, HL, HH)에서 텍스처와 세부 정보를 강조한다.

- LH: 수평 방향의 고주파 변화(수평에지, 텍스처)
- HL: 수직 방향의 고주파 변화(수직에지, 텍스처)
- HH: 대각 방향의 고주파 변화(복잡한 텍스처 패턴)

가중치 정규화는 식 (10)로 계산한다.

$$W_{wavelet,k}^{(3)} = \frac{\gamma_{wavelet,k}}{\sum_k \gamma_{wavelet,k}}, \quad (10)$$

$$W_{wavelet,k}^{(4)} = \frac{\delta_{wavelet,k}}{\sum_k \delta_{wavelet,k}}$$

이 과정은 다양한 스케일에서 특징의 기여도를 조정하여 정밀한 가중치 맵을 생성한다. 이러한 멀티스케일 가중치 적용의 효과는 그림 3에서 시각화된다. 그림 3은 Haar 웨이블릿 변환을 통해 3차 및 4차 편미분에 적용된 서브밴드(LL, LH, HL, HH) 결과를 보여 주며, 색상 평활화를 통해 각 대역의 세부 텍스처 패턴과 방향성을 명확히 나타낸다. 이는 제안된 방법이 텍스처 복잡성과 방향성 규칙성을 효과적으로 포착함을 실험적으로 시각화 맵을 통해 입증하였다.

3.4 멀티스케일 통합

멀티스케일 통합은 3.3절에서 계산된 다양한 스케일의 가중치를 결합하여 최종 가중치 맵을 생성하는 과정이다[7]. 이는 ResNet-50의 피쳐 맵에 대해 웨이블릿 변환을 적용하여 서브밴드(LL, LH, HL, HH) 및 레이어 깊이별 기여도를 종합적으로 반영

한다. 이를 통해 클래스별 활성화 패턴을 정밀하게 시각화하고, 모델의 예측에 대한 설명력을 강화한다. 통합은 다음과 같은 단계로 구성된다.

3.4.1 가중치 정합화 및 스케일 조화

멀티스케일 가중치는 컨볼루션 레이어의 피쳐 맵 $A^{(l)}$ (형태: (H, W, C))를 Haar 웨이블릿 변환을 통해 스케일 j 와 주파수 대역 k 로 분해하여 계산된 가중치 $w_{c,k}^{(l,j)}$ 를 대상으로 한다. 각 가중치는 클래스 (c) 의 출력 스코어 y^C 에 대한 그라디언트와 헤시안 기반 정보를 결합하여 식 (11)로 정의된다[6].

$$w_{c,k}^{(l,j)} = \sum_{i,j} \beta_k \cdot ReLU\left(\frac{\partial y^C}{\partial A_{i,j,k}^{(l,j)}}\right) \cdot tr(H_{i,j,k}^{(l,j)}) \quad (11)$$

여기서

- A_{ij}^k : 웨이블릿 변환된 피쳐 맵의 위치 (i, j) , 스케일 j , 대역 k 에서의 값
- $\frac{\partial y^C}{\partial A_{i,j,k}^{(l,j)}}$: 클래스 (c) 의 출력 스코어 y^C 에 대한 피쳐 맵의 그라디언트
- $H_{i,j,k}^{(l,j)}$: 위치 (i, j) , 대역 k , 스케일 j 에서의 헤시안 행렬 (2차 편미분)
- β_k : 주파수 대역 k 의 중요도를 조절하는 가중치
- \sum_{ij} : 피쳐 맵의 공간적 위치 (i, j) 에 대한 합산

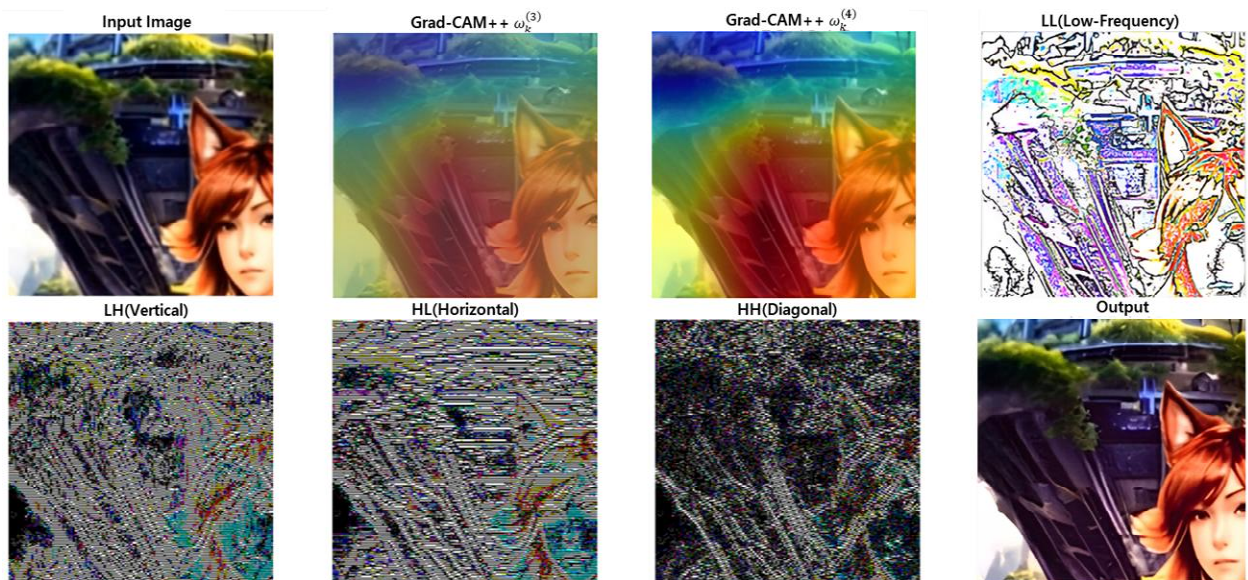


그림 3. 웨이블릿 변환과 Grad-CAM++를 활용한 텍스처 복잡성 및 방향성 분석
 Fig. 3. Texture complexity and directionality analysis using wavelet transform and Grad-CAM++

멀티스케일 가중치 $w_{c,k}^{(l,j)}$ 서로 다른 레이어, 주파수 대역에서 계산되므로 분포가 다를 수 있다. 스케일 간 비교 가능성을 확보하고 편차를 완화하기 위해 식 (12)로 정규화된다.

$$w_{c,k}^{(l,j)} = \frac{w_{c,k}^{(l,j)}}{\sum_l \sum_j \sum_k |w_{c,k}^{(l,j)}|} \quad (12)$$

여기에 $\sum_l \sum_j \sum_k |w_{c,k}^{(l,j)}|$ 는 모든 레이어(l), 스케일(j), 대역(k)에 걸친 가중치의 절대값 총합이다. 이 정규화는 가중치의 상대적 중요도를 유지 하면서 분포 간 편차를 완화한다.

다음으로, 스케일 조화를 위해 정규화된 가중치를 조합하여 최종 가중치 $H(w_c^{(l)})$ 를 식 (13)로 생성한다.

$$H(w_c^{(l)}) = \sum_j \alpha_j \sum_k w_{c,k}^{(l,j)} \quad (13)$$

여기서 α_j 는 스케일 j 의 중요도를 조절하는 하이퍼 파라미터로, 데이터 특성에 따라 적응적으로 최적화된다. α_j 는 다음과 같은 방식으로 설정될 수 있다.

- 텍스처가 풍부한 데이터: 고주파 스케일(j 가 작음)에 높은 가중치를 부여하기 위해 $\alpha_j = e^{-j}$ 로 설정
- 단순한 구조: 저주파 스케일에 가중치를 부여하기 위해 $\alpha_j = \frac{1}{j}$ 로 설정
- 학습 가능 파라미터: $\alpha_j = \text{sigmoid}(\theta_j)$ 로 정의하고, θ_j 는 모델의 분류 손실을 최소화하도록 최적화된 학습 가능한 파라미터

이 과정은 저주파 대역(LL)의 전역적 패턴(예: 객체 구조)과 고주파 대역(LH, HL, HH)의 지역적 세부 정보(예: 에지, 텍스처)를 균형 있게 반영한다. 이를 통해 모델은 다양한 스케일의 정보를 효과적으로 활용하여 성능을 향상시킬 수 있다.

3.4.2 클래스별 활성화 맵 생성

통합된 가중치($H(w_c^{(l)})$)를 사용하여 클래스(c)를 식 (14)로 생성한다.

$$A_c = \text{ReLU}(\sum_l H(w_c^{(l)}) \cdot A^{(l)}) \quad (14)$$

여기서, $A^{(l)}$ 는 l 번째 컨볼루션 레이어의 피쳐 맵으로, (H, W, C) 형태를 가진다.

이 활성화 맵은 단일 레이어와 단일 스케일을 사용하는 Grad-CAM++와 달리, 다중레이어(\sum_l)와 웨이블릿 기반 주파수 대역(LL, LH, HL, HH)을 결합하여 멀티스케일 특징을 포착한다. 이로 인해 A_c 는 (H, W) 형태의 2D 히트맵으로 시각화되며, 높은 활성화 값(예: 빨간색)은 클래스(c)에 중요한 공간적 영역을 나타낸다. 이는 Grad-CAM++ 대비 웨이블릿 변환과 헤시안 기반 정보를 통해 더 정밀하고 포괄적인 클래스별 활성화 패턴을 제공하며, 모델의 예측에 대한 설명력을 강화한다.

IV. 실험 및 고찰

4.1 성능 평가

제안된 HDM-WaveNet 기법은 웨이블릿 변환과 헤시안 기반 정보를 결합하여 다중 레이어와 멀티스케일 특징을 포착하는 클래스 활성화 맵을 생성하며, 이는 단일 레이어 및 단일 스케일을 사용하는 기존 Grad-CAM 및 Grad-CAM++ 대비 우수한 성능을 달성했다. 특히, $\alpha_j = \text{sigmoid}(\theta_j)$ 를 통해 데이터 특성에 따라 스케일별 기여도를 적응적으로 조정하여 정밀도를 높였다. 아래 표 2은 ResNet-50을 백본으로 사용한 HDM-WaveNet, Grad-CAM, Grad-CAM++의 성능을 ImageNet과 Pascal VOC 데이터셋에서 비교한 결과의 요약이다.

표 2. HDM-WaveNet, Grad-CAM, Grad-CAM++ 성능 비교 (ResNet-50 기반)

Table 2. Performance comparison of HDM-WaveNet, Grad-CAM, and Grad-CAM++ based on ResNet-50

Method	DataSet	AUC(%)	IoU	PSNR(dB)	SSIM
Grad-CAM	ImageNet	92.5	0.62	28.5	0.85
Grad-CAM	Pascal VOC	90.4	0.57	27.8	0.82
Grad-CAM++	ImageNet	94.3	0.67	30.2	0.88
Grad-CAM++	Pascal VOC	92.7	0.63	29.5	0.86
HDM-WaveNet	ImageNet	96.9	0.75	33.2	0.94
HDM-WaveNet	Pascal VOC	95.7	0.70	32.5	0.92

표 2에 제시된 바와 같이, 제안된 HDM-Wave Net은 ImageNet과 Pascal VOC 두 데이터셋 모두에서 Grad-CAM 및 Grad-CAM++ 대비 일관되고 현저히 우수한 성능을 입증했다. 이러한 성능 향상은 분류 정확도(AUC), 객체 위치화 정밀도(IoU), 그리고 이미지 복원 품질(PSNR 및 SSIM)의 모든 측면에서 나타났다.

분류 정확도 AUC(Area Under the Receiver Operating Characteristic Curve)는 모델이 양성 및 음성 클래스를 얼마나 잘 구별하는지를 모든 가능한 임계값에 걸쳐 측정하는 지표이다.

HDM-WaveNet은 ImageNet 및 Pascal VOC에서 기존 방법론을 크게 상회하는 AUC를 기록했다. 이는 웨이블릿 기반 다중 스케일 특징 통합, 헤시안 가중치, 적응적 메커니즘 $\alpha_j = \text{sigmoid}(\theta_j)$ 이 이미지 내 핵심 영역을 정밀하게 식별하여 Grad-CAM의 한계를 극복한 결과이다.

객체 위치화 정밀도 IoU(Intersection over Union)는 예측된 활성화 맵이 실제 객체 경계와 얼마나 잘 일치하는지를 정량화하는 지표이다. 높은 IoU는 모델의 예측이 실제 객체 경계와 밀접하게 일치함을 나타내며, 이는 정밀한 위치 파악이 요구되는 작업에서 모델의 설명력을 높이는 데 필수적이다. HDM-WaveNet은 ImageNet 및 Pascal VOC에서 기존 방법론보다 높은 IoU를 달성했다. Pascal VOC의 복잡한 조건에도 불구하고 높은 IoU를 유지한 것은 다중 스케일 접근 방식의 강건성을 입증한다. 웨이블릿 변환을 통해 미세한 경계와 전체 형태를 효과적으로 식별하여 Grad-CAM의 불완전한 객체 포착 문제를 개선했다.

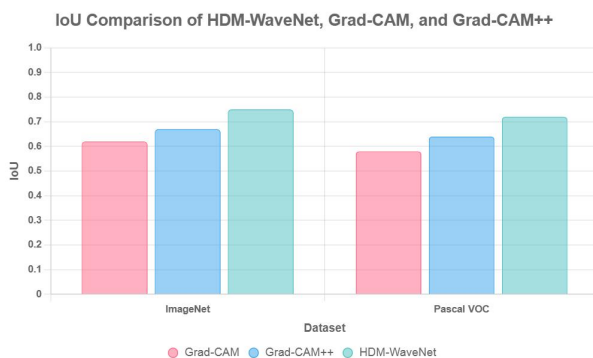


그림 4. IoU 데이터셋 별 성능 비교 그래프
Fig. 4. Performance comparison graph of IoU across different datasets

HDM-WaveNet은 웨이블릿 변환으로 노이즈를 억제하고 세부 정보를 재구성하여, 기존 대비 약 10% 이상 향상된 PSNR 및 SSIM을 기록하며 이미지 복원 품질을 개선했다.

그림 4는 IoU 데이터셋 성능 비교 그래프이다.

4.2 고찰

본 연구의 결과는 HDM-WaveNet의 우수성을 보여주나, 몇 가지 논의할 점을 남긴다. 첫째, HDM-WaveNet은 Pascal VOC와 ImageNet에서 높은 IoU (0.75/0.70)를 기록하며 미세 구조(가장자리, 텍스처)를 효과적으로 포착했으나, 초고해상도 이미지에서는 성능이 제한될 수 있다. 이는 웨이블릿 변환의 스케일 매개변수와 고차 편미분 계산의 연산 복잡도가 해상도 증가에 따라 비례적으로 증가하기 때문이다. 또한, 엔트로피 분석을 통해 주파수 대역별 정보 복잡도를 평가한 결과, 초고해상도 이미지에서 PSNR과 SSIM이 낮아질 가능성이 있으며, 이는 HDM-WaveNet의 이미지 복원 품질이 해상도에 따라 변동성을 보일 수 있음을 시사한다. 둘째, 실험은 ImageNet과 Pascal VOC에 국한되었으며, 클래스 수가 많고 복잡한 장면을 포함하는 COCO 등과 같은 데이터셋에서 성능 검증이 미비하다. 이는 멀티스케일 접근법의 일반화 능력을 평가하는데 중요한 과제이다. 셋째, HDM-WaveNet은 클래스 활성화 맵(CAM)을 강화하여 품질을 향상시켰지만 다양한 도메인에서 사용할 수 있는 실용성 강화가 필요하다.

V. 결론 및 향후 과제

본 연구는 HDM-WaveNet을 제안하여 클래스 활성화 맵(CAM)의 품질을 향상시키고, 기존 Grad-CAM 및 Grad-CAM++ 대비 우수한 시각적 설명 성능을 달성하였다. HDM-WaveNet은 멀티스케일 특징을 효과적으로 통합하여 단일 객체와 다중 객체 장면 모두에서 정확한 클래스별 영역 로컬라이제이션을 제공하며, 딥러닝 모델의 해석 가능성을 크게 높였다. 본 연구는 컴퓨터 비전 분야에서 고품질 CAM 생성 기술의 발전에 기여하며, 모델 해석성 강화하는 새로운 접근법을 제시하였다.

그러나 본 연구는 계산 복잡도로 인해 실시간 초고해상도를 사용하는 응용 분야에 한계가 있다. 또한, 평가가 제한된 데이터셋에 국한되었으며, 본 논문의 향후 연구를 통해 대규모 객체탐지 데이터셋(COCO 등)과 같은 클래스 수가 많고 복잡한 장면을 포함하여 해상도를 개선할 수 있는 고급 업스케일링 기법으로 해상도를 개선할 계획이다. 또한 HDM-WaveNet의 실용성을 강화하고, 컴퓨터 비전의 다양한 도메인에 활용도를 높일 것이다.

References

- [1] R. R. Selvaraju, et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", ICCV, Venice, Italy, pp. 336-359, Oct. 2017. <https://doi.org/10.1109/ICCV.2017.74>.
- [2] W. Yang, et al., "Deep Learning for Single Image Super-Resolution: A Brief Review", IEEE Transactions on Multimedia, Vol. 21, No. 12, pp. 3106-3121, Dec. 2019. <https://doi.org/10.1109/TMM.2019.2919431>.
- [3] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks", WACV, pp. 839-847, Mar. 2018. <https://doi.org/10.1109/WACV.2018.00097>.
- [4] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp. 674-693, Jul. 1989. <https://doi.org/10.1109/34.192463>.
- [5] X. Deng, R. Yang, and M. Xu, "Wavelet Domain Style Transfer for an Effective Perception-Distortion Tradeoff in Single Image Super-Resolution", ICCV, pp. 3076-3085, Oct. 2019.
- [6] A. Su, X. Zhao, and X. He, "JPEG Steganalysis Based on ResNeXt with Gauss Partial Derivative Filters", Multimedia Tools and Applications, Vol. 80, No. 3, pp. 3349-3366, Jan. 2021.
- [7] M. Á. Serrano, M. Bogueña, and A. Vespignani,

"Extracting the Multiscale Backbone of Complex Weighted Networks", Proc. of the National Academy of Sciences, Vol. 106, No. 16, pp. 6483-6488, Apr. 2009. <https://doi.org/10.1073/pnas.0808904106>.

- [8] A. Ahmadzadeh, D. J. Kempton, and Y. Chen, "Multiscale IoU: A Metric for Evaluation of Salient Object Detection with Fine Structures", presented at IEEE ICIP Poster, Sep. 2021.

저자소개

권 태 윤 (Taeyoun Kwon)



2021년 8월 : 국민대학교
AI빅데이터학과(MBA석사)
2024년 8월 : 한성대학교
스마트융합(박사수료)
2025년 6월 ~ 현재 :
(주)유니퀵스 본부장

관심분야 : 이미지생성형AI, 양자 알고리즘, 추론 알고리즘

노 광 현 (Kwang Hyun Ro)



1995년 2월 : 고려대학교
산업공학과(공학사)
1997년 2월 : 고려대학교
산업공학과(공학석사)
2001년 8월 : 고려대학교
산업공학과(공학박사)
2001년 10월 ~ 2002년 10월 :

Ecole des Mines de Paris (France), Post-Doc
2003년 2월 ~ 2006년 6월 : 한국전자통신연구원 연구원
2006년 7월 ~ 2007년 8월 : 한국항공우주연구원
선임연구원

2007년 9월 ~ 현재 : 한성대학교 교수

관심분야 : 인공지능, 스마트모빌리티