

# 초해상도 이미지 생성을 위한 SE Block과 Inception 모듈 기반 ESRGAN 판별자

김예은<sup>○</sup>      남아영      양단아\*

한국성서대학교 컴퓨터소프트웨어학과

{hydrogen, ejr3389, dana1112}@bible.ac.kr

## Discriminator of ESRGAN based on Squeeze-and-Excitation Block and Inception Module for Qualitative Image

Yeeun Kim<sup>○</sup>      Ayoung Nam      Dana Yang\*

Dept. of Computer Software, Korean Bible University

### 요 약

생성적 적대 신경망(Generative Adversarial Network, GAN)을 활용한 초해상도 기술은 다양한 분야에서 주목받고 있다. 특히 저해상도 이미지를 고해상도로 변환할 때, 보다 자연스럽게 사실적인 텍스처 복원 능력을 보여준다. 하지만 기존의 초해상도 GAN 모델들은 여전히 고해상도 이미지 생성 시 질감, 가장자리, 패턴 등의 세부적인 특성을 정밀하게 복원하는데 한계를 가지고 있다. 본 연구에서는 초해상도 생성 모델, 특히 Enhanced Super-Resolution Generative Adversarial Network(ESRGAN)의 성능을 개선하기 위해 Inception 모듈과 SE (Squeeze-and-Excitation) Block을 병합한 판별자 구조를 제안한다. 이러한 접근 방법을 통해 모델은 다양한 스케일의 이미지 특징을 보다 효과적으로 학습하고 중요한 정보에 집중할 수 있게 된다. 제안하는 판별자 기반 ESRGAN 모델은 이미지 품질 평가지표에서 PSNR은 29.07dB, SSIM은 0.82, LPIPS는 0.27, NIQE는 5.85를 달성하여 기존 모델 대비 개선되었음을 보인다.

### Abstract

GAN models in super-resolution technology have garnered significant attention across various fields. The restoration of more natural and realistic textures is remarkably demonstrated, when converting low-resolution images to high-resolution counterparts. However, the traditional GAN models for super-resolution still face limitations in precisely reconstructing detailed characteristics such as textures, edges, and patterns when generating high-resolution images. This paper proposes a novel discriminator architecture that integrates the Inception module and SE Block to enhance the performance of super-resolution generative models, particularly the ESRGAN. This model is able to learn multi-scale image features and focus on critical information more effectively. The proposed model achieves performance improvements over the existing ESRGAN model, with a PSNR of 29.07 dB, SSIM of 0.82, LPIPS of 0.27, and NIQE of 5.85 on image quality evaluation metrics.

**키워드:** 초해상도, ESRGAN, 판별자, Inception 모듈, SE 블록, 생성적 적대 신경망

**Keywords:** Super Resolution, Discriminator, ESRGAN, Inception Module, SE Block, GAN

\*corresponding author: Dana Yang/Korean Bible University(dana1112@bible.ac.kr)

## 1. 서론

초해상도 기술은 저해상도 이미지를 고해상도로 변환하는 과정에서 이미지의 세부적인 정보를 복원 해내는데 핵심적인 역할을 한다. 이 기술은 이미지의 화질을 크게 향상하여 의료 영상 분석, 위성 이미지의 품질 개선, 감시 시스템과 같은 다양한 분야에서 필수적인 도구로 자리 잡고 있다. 예를 들어 보안 감시 시스템에서는 높은 해상도의 이미지가 사건의 식별 및 분석에 큰 도움이 된다[1]. 최근 생성적 적대 신경망(Generative Adversarial Network, GAN)[2]을 활용한 초해상도 기술이 주목받고 있다. 이 중에서도 Enhanced Super-Resolution Generative Adversarial Network(ESRGAN)와 같은 모델은 저해상도 이미지에서 고해상도 이미지로의 변환 시 현실적이고 자연스러운 질감을 매우 효과적으로 복원해 낼 수 있다[3].

복잡한 Residual in Residual Dense Block(RRDB)의 결합으로 이루어진 ESRGAN의 생성자는 매우 세밀한 이미지의 디테일을 복원할 수 있는 능력을 지니며, 이에 상응하는 판별자는 모델 훈련 과정에서 중요한 역할을 한다. 판별자는 생성된 고해상도 이미지와 실제 고해상도 이미지 사이의 미세한 차이를 효과적으로 구별 해내야 하며, 이는 전체 네트워크의 학습 성능에 기여한다. 이러한 생성자와 판별자 간의 상호 작용은 학습 과정에서 균형을 요구하며, 이는 결국 네트워크가 보다 정교하고 현실적인 이미지를 생성하는데 필수적이다. 하지만 ESRGAN 모델은 동일한 필터 크기를 가진 판별자(Discriminator)를 활용하여 고해상도 이미지의 판별을 수행한다. 이러한 경우 판별자가 생성자(Generator)가 생성한 디테일한 고화질 이미지의 세부적인 특성을 충분히 인식하지 못하는 문제가 발생할 수 있다. 따라서 복잡한 장면이나 미세한 디테일이 중요한 이미지에서 질감, 가장자리, 패턴 등의 세부적인 특성을 완벽하게 복원하지 못하게 되며, 충분히 모델 훈련이 이루어지지 않아 성능저하로 인한 이미지 열화현상이 발생할 수 있다. 이러한 한계는 특히 고해상도 이미지가 필수적인 응용 분야에서 전반적인 품질 저하를 초래할 수 있다.

따라서 본 논문에서는 초해상도 이미지 생성 모델인 ESRGAN의 성능을 개선하기 위해 복잡하고 높은 성능의 생성자에 맞는 강인한 판별자 구조를 제안함으로써 초해상도 이미지 생성의 질을 더욱 향상시키는데 기여한다. 본 논문은 판별자의 구조에 Inception 모듈과 SE Block(Squeeze-and-Excitation Block)을 병합하여 다중 스케일 필터와 채널 중요도 조절을 통한 새로운 접근 방식을 제안한다. 이러한 구조적 개선은 판별자가 고해상도 이미지를 더 정밀하게 판별하도록 도우며, 전체 모델의 성능 향상에 기여한다. 이러한 구조적 개선을 통해 제안하는 모델이 기존 모델 대비 이미지의 세부 정보를 더 잘 학습하고 이미지 내 중요한 정보

에 더 효과적으로 집중함으로써 고해상도 이미지 생성의 질을 향상시키는 것을 목표로 한다.

본 논문에서는 제안하는 모델의 구조적 개선을 피쳐 맵 분석과 이미지 성능 평가지표를 통해 평가한다. 피쳐 맵 분석을 통해 판별자가 처리한 피쳐 맵을 분석함으로써 모델이 이미지 내 중요 정보를 얼마나 잘 포착하고 있는지 확인한다. 또한 이미지 평가지표를 사용하여 제안하는 모델과 기존 모델의 성능 차이를 분석함으로써 생성된 이미지의 품질을 비교한다.

본 논문은 먼저 2장에서 관련 연구로 초해상도 기술과 관련된 ESRGAN에 대한 기존 연구를 검토하며, 3장의 제안 모델에서 본 연구의 핵심인 Inception 모듈과 SE Block을 병합한 판별자 기반의 새로운 접근법을 제안한다. 이후 4장의 평가 부분에서 제안하는 모델의 성능을 검증하기 위한 실험 결과를 제시하고 마지막 5장의 결론 부분에서 연구 결과 및 요약과 함께 향후 연구 방향에 대해 논의한다.

## 2. 관련 연구

초해상도 기술은 저해상도 이미지에서 고해상도 이미지로 변환하는 과정을 통해 이미지의 세부 정보를 복원하는 것을 목표로 한다. 초해상도 분야는 컴퓨터 비전과 이미지 처리의 중요한 분야 중 하나로 초기의 전통적인 방법론에서부터 최근의 딥러닝 기반 기술까지 꾸준히 발전해 왔다. 전통적인 방법론은 주로 보간법과 같은 수학적 모델을 기반으로 하였으나 이러한 방법들은 고해상도 이미지에서 자연스러운 질감이나 세밀한 디테일을 복원하는데 한계가 있었다[4]. 이후 딥러닝의 등장은 초해상도 기술에 혁신을 가져왔다. 컴퓨터 비전에 핵심이 되는 심층 합성곱 신경망은 초해상도 연구에서 중요한 역할을 하였으며, 이를 통해 더 깊은 네트워크와 새로운 구조를 도입한 VDSR[5], LapSRN[6], MemNet[7] 등과 같은 모델이 제안되었다. 이러한 모델 중 SRGAN(Super Resolution Generative Adversarial Network)은 GAN의 개념을 적용하여 사실적인 초해상도 이미지를 생성한다. SRGAN은 생성자와 판별자 두 주요 구성 요소로 이루어져 있다. 생성자는 저해상도 이미지를 고해상도 이미지로 변환하는 기능을 수행하며, 판별자는 생성된 이미지가 실제 고해상도 이미지와 얼마나 유사한지를 평가하는 역할을 한다. 이 과정을 통해 이미지의 품질을 지속적으로 개선함으로써 높은 해상도에서도 세밀한 디테일과 자연스러운 질감을 유지하고, 시각적 품질을 높인다[8].

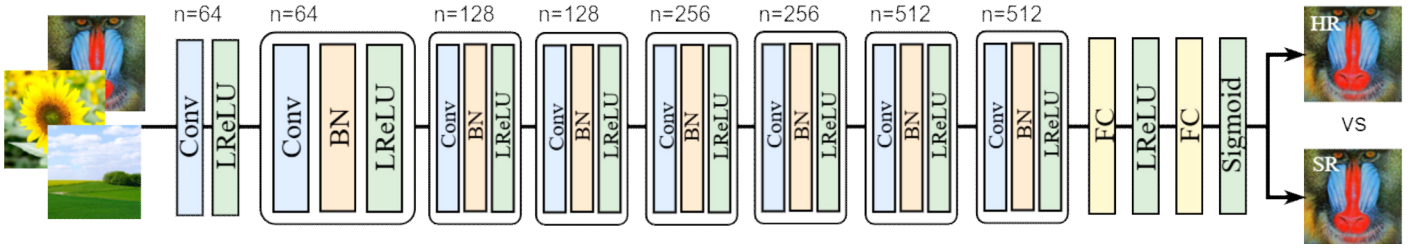


Figure 1. Discriminator of ESRGAN

## 2.1 ESRGAN (Enhanced Super-Resolution Generative Adversarial Network)

ESRGAN은 GAN 기반의 초해상도 모델로 기존 SRGAN의 아키텍처를 발전시킨 네트워크 구조와 정교화된 손실 함수를 도입함으로써 더욱 향상된 고해상도 이미지 생성 능력을 제공한다. ESRGAN은 SRGAN의 기본적인 개념, 즉 생성자와 판별자 사이의 경쟁적 학습을 계승하면서도 보다 정교한 적대적 네트워크 구조를 통해 시각적으로 만족스럽고 현실적인 텍스처를 생성한다. 이러한 향상은 특히 이미지의 자연스러운 질감과 세밀한 디테일을 복원하는데 중요한 역할을 하며, 이를 통해 초해상도의 주된 목표인 저화질 이미지를 효과적으로 복원할 수 있다[9].

### 2.1.1 Generator

ESRGAN의 생성자는 초해상도 이미지를 생성하여 이미지 화질을 개선하는데 핵심적인 역할을 수행한다. ESRGAN에 생성자의 핵심적인 구성 요소는 RRDB로 이는 Multi-level Residual Network와 Dense Connection을 결합하여 생성자의 표현력을 크게 향상시킨다. RRDB는 기존의 Dense Connection에 Residual Connection을 추가하여 피쳐 전달과 재사용을 최적화함으로써 깊은 네트워크에서도 정보의 손실을 최소화하고 효과적인 학습을 가능하게 한다[10]. 이러한 구조적 개선 외에도 ESRGAN은 훈련의 안정성과 계산 효율성을 높이기 위해 배치 정규화(Batch Normalization, BN) 계층을 제거하였다. BN 제거는 내부 공변량 변화를 줄여 모델이 다양한 입력 데이터에 대해 보다 일관된 성능을 나타낼 수 있도록 한다. 이는 네트워크의 계산 복잡성을 줄이고 특히 화질 개선 작업에서 원하는 세부적인 텍스처와 이미지의 질감을 더 잘 재현할 수 있도록 돕는다[11, 12]. ESRGAN의 생성자는 또한 적대적 손실 외에도 퍼셉추얼 손실을 사용하여 생성된 이미지의 질적인 측면을 더욱 향상시킨다. 이는 VGG 네트워크를 활용하여 고해상도 이미지의 세부적인 텍스처와 색상을 더욱 정확하게 모방하고 최종적으로 더욱 실제와 유사한 이미지를 생성하는데 기여한다[1

3]. 본 논문에서는 이러한 생성자를 활용하여 초해상도 모델의 성능을 향상하고자 한다.

### 2.1.2 Discriminator

ESRGAN의 판별자는 생성된 초해상도 이미지를 평가하여 모델 훈련 과정에서 생성자의 성능 향상에 중요한 역할을 한다. 이러한 ESRGAN의 판별자는 기존 SRGAN의 판별자 구조를 사용한다. SRGAN의 판별자는 다층의 컨볼루션 네트워크로 구성되어 있으며, 각 계층은 3x3 크기의 필터를 사용하여 특징을 추출한다. 이 필터들은 이미지에서 세부적인 질감과 패턴을 포착하며, 각 컨볼루션 계층 후에는 배치 정규화와 Leaky ReLU 활성화 함수가 적용된다. 이는 모델의 비선형 학습 능력을 강화하고 훈련 과정 중 안정성을 높이는 역할을 한다.

Figure 1에 ESRGAN 판별자의 피쳐 맵(n)의 수는 네트워크가 깊어질수록 2배씩 증가하며, 64(n=64)개에서 시작하여 네트워크의 끝부분에서는 512(n=512)개까지 증가한다. 이러한 방식으로 피쳐의 차원을 점진적으로 확장하는 것은 모델이 이미지의 복잡한 특징을 점차 학습하게 하며, 더 깊은 계층에서는 이미지의 전반적인 구조와 컨텍스트를 분석하는데 필요한 정보를 제공하게 된다. 모델의 마지막에는 완전 연결 계층(Fully Connected Layer, FC)을 통해 모든 학습된 피쳐를 종합하고 Sigmoid 활성화 함수를 사용하여 최종적으로 이미지가 진짜 고해상도 이미지일 확률을 출력하게 된다. 이 단계는 판별자가 생성된 초해상도 이미지와 실제 고화질 이미지 사이를 효과적으로 구별할 수 있도록 한다. 이러한 구조는 전체 GAN 시스템의 성능을 결정하는 중요한 요소로 작용하며, 생성자가 더 정교하고 현실적인 이미지를 생성하도록 돕는다[14]. 따라서 판별자의 구조적 개선은 복잡한 이미지 특징과 다양한 스케일에서 디테일을 보다 효과적으로 처리하여 초해상도 이미지의 질을 향상시킬 뿐만 아니라 기존 초해상도 모델의 한계를 극복하고 이미지의 자연스러움과 세밀한 디테일을 더욱 정밀하게 복원할 수 있는 능력을 향상하게 된다.

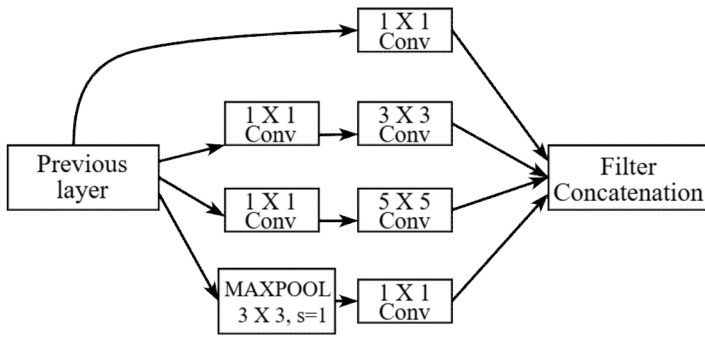


Figure 2. Inception Module

## 2.2 Inception 모듈

Inception 모듈은 여러 스케일의 컨볼루션 필터를 통해 다양한 공간 해상도의 피처를 추출한다[15]. 이미지 데이터 세트가 고해상도 이미지인 경우, 작은 크기의 필터보다는 크기가 다양한 필터를 사용하는 것이 고해상도 이미지의 세밀하고 정교한 특징을 추출하는데 도움이 된다[16]. 따라서 기존 ESRGAN에서 사용하고 있는 단일 3x3 필터보다는 다중 스케일의 필터를 사용하는 것이 더 효과적인 성능을 발휘할 수 있다.

Figure 2를 보면 먼저 1x1 컨볼루션은 주로 채널 간 정보를 효율적으로 통합하여 계산 비용을 최소화하면서 피처를 추출한다. 3x3 컨볼루션은 중간 범위의 피처를 포착하여 공간적 맥락을 더 깊이 파악하며, 5x5 컨볼루션은 더 넓은 범위의 공간적 정보를 수집하여 복잡한 패턴을 인식하게 된다. 또한 3x3 MAX Pooling을 통해 다양한 공간 특성을 최대한 활용하여 정보의 다양성을 확보하게 된다. 이러한 구조를 통해 각각의 특징들이 서로 보완적으로 작용하여 고해상도 이미지 분석에 필수적인 다양한 정보의 포착을 돕는다[17]. 따라서 Inception 모듈의 도입은 이미지의 복잡한 특징과 다양한 스케일의 디테일을 더욱 효과적으로 처리할 수 있도록 하여 전체적인 이미지 품질을 개선하는데 기여한다.

## 2.3 SE Block (Squeeze-and-Excitation Block)

SE Block은 이미지 인식 및 분류 작업에서 모델 성능을 향상시키기 위해 Channel-wise Feature Responses를 동적으로 조정한다[18]. Figure 3에서 먼저 컨볼루션 피처 맵에서 Channel Descriptor를 추출하기 위해 글로벌 평균 풀링(Global Average Pooling, GAP)을 사용한다. 이 Channel Descriptor들은 두 개의 FC를 통해 처리되며, 첫 번째는 차원을 축소하고 두 번째는 차원을 복원한다. 이어서 ReLU와 Sigmoid 활성화 함수가 각각 적용된다. 이 과정은 시그모이

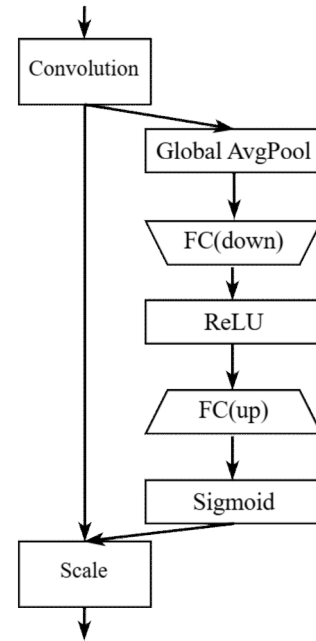


Figure 3. SE Block

드 함수의 출력을 기반으로 채널별 피처 중요도를 재조정하고 이 스케일 요인들을 원래의 피처 맵에 다시 적용한다. 이 메커니즘은 중요한 특징을 강조할 뿐만 아니라 덜 유용한 정보를 억제함으로써 모델의 판별력을 극대화할 수 있다[19]. 따라서 SE Block의 적용은 이미지 내 세밀한 디테일과 구조적 특징을 더 잘 이해하고 중요한 부분에 집중하도록 함으로써 판별자의 성능을 향상하는데 기여한다. 이는 SE Block이 각 채널의 중요도를 동적으로 조정함으로써 판별자가 이미지의 중요 특성에 더욱 집중하고 세밀한 차이를 효과적으로 구별할 수 있도록 만들기 때문이다[20].

## 3. 제안 모델

Figure 4의 제안하는 판별자 구조는 Inception 모듈을 사용하여 다양한 크기의 커널로 이미지의 여러 스케일에서 피처를 추출한다. 이는 네트워크가 더 넓은 컨텍스트를 고려하면서도 세밀한 디테일에 주의를 기울일 수 있도록 한다. 또한 SE Block을 병합하여 피처 맵의 채널별 중요도를 조정한다. 이를 통해 네트워크는 이미지의 중요한 부분에 더 많은 연산 자원을 할당하고 중요도가 낮은 정보는 억제함으로써 효율적인 학습이 가능해진다[21]. 이러한 구조는 Multi-scale Receptive Fields와 Channel-wise Feature Recalibration을 통합하여 초해상도 이미지의 생성 품질을 높게 된다[22, 23].



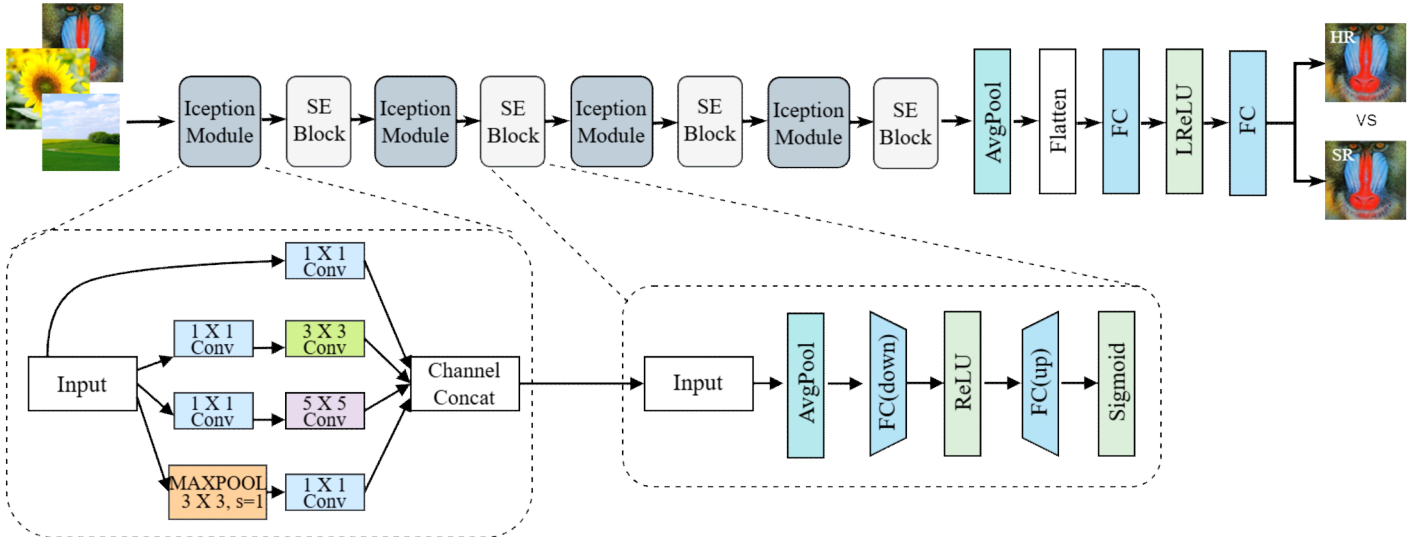


Figure 4. The structure of Discriminator that merges the Inception Module and SE Block

### 3.1 Inception 모듈 구조

Inception 모듈은 네트워크 내에서 다양한 크기의 커널을 병렬로 사용하여 이미지의 다양한 스케일에서 피처를 동시에 추출할 수 있도록 설계되었다. ESRGAN의 판별자는 생성자가 생성한 고해상도 이미지를 판별하게 되는데 이미지의 화질이 고화질인 경우에는 3x3 단일 필터 구조를 사용하는 것보다는 다양한 크기의 필터의 구조를 사용하는 것이 더 세부적인 특징을 추출하는데 유리하다. 따라서 Figure 4에서 제시된 바와 같이 각각의 필터 크기가 특정 기능을 수행하도록 설계하였다.

1x1 컨볼루션은 차원 축소를 목적으로 사용되어 계산 부하를 감소시키는 동시에 중요한 정보의 보존이 가능하다. 이는 정보의 효율적인 압축과 처리 속도의 향상에 기여한다. 1x1 컨볼루션 이후 3x3 컨볼루션은 중간 크기의 패턴과 텍스처를 효과적으로 포착하여 계산의 효율성을 높인다. 1x1 컨볼루션 이후 5x5 컨볼루션은 더 큰 이미지 영역을 포착하여 네트워크가 더 넓은 범위의 정보를 처리할 수 있게 하며, 복잡한 이미지 내에서 다양한 크기의 특징을 동시에 포착할 수 있다. 3x3 MAX Pooling 이후 1x1 컨볼루션은 피처를 강화하고 차원 축소를 통해 계산 효율성을 높이며, 모델의 과적합을 방지한다. 이러한 다양한 크기의 필터 조합은 최적의 지역 최소 구조를 근사화하고 커버하는데 효과적이며, 이러한 특징의 Inception 모듈은 네트워크의 전반적인 인식 능력을 향상시키는데 중요한 역할을 한다. 각각의 커널 크기에서 생성된 피처 맵은 하나의 통합된 피처 맵으로 병합되어 출력되며, 이 과정은 정보의 손실을 최소화하고 네트워크의 효율성을 증가시키는데 기여한다. 이 구조는 기존의 동일하고 작은 필터만을 사용하는 접근보다 더 광범위하고 정교한 특징을 학습할 수 있도록 돕는다[12].

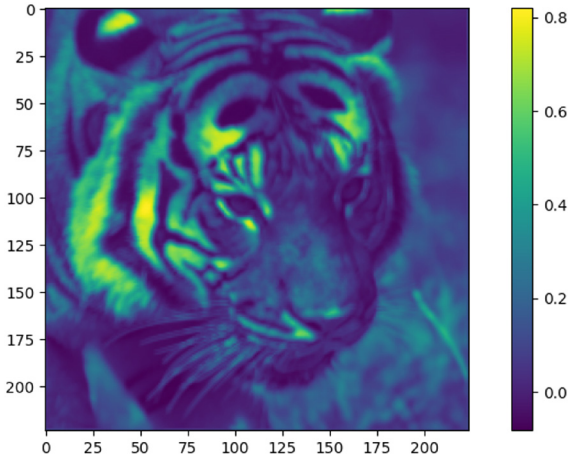
이러한 특징들은 본 논문에서 제안된 Inception 모듈의 설계와 구현을 통해 SR 모델의 성능을 크게 향상시킨다.

### 3.2 SE Block 구조

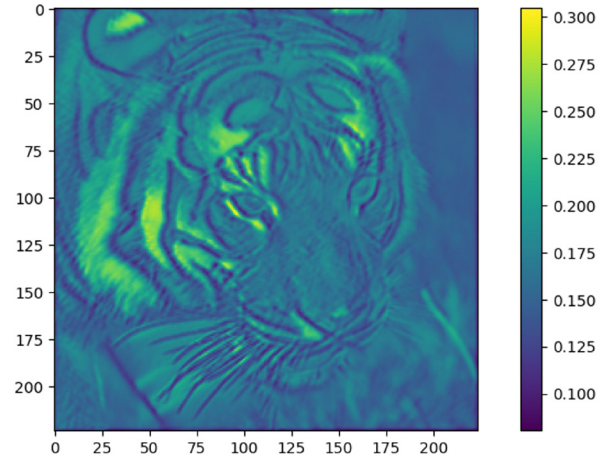
각 Inception 모듈에 통합된 SE Block은 이미지 특성을 학습하는데 중요한 역할을 한다. SE Block은 각 채널의 글로벌 정보를 압축하여 중요도를 평가하고 이를 통해 특정 채널의 활성화 정도를 조정함으로써 중요한 특성을 강조하고 상대적으로 중요도가 낮은 특성을 억제하여 정보의 선택적 처리를 가능하게 한다. 이러한 채널별 가중치 조절은 모델의 효율성을 높이고 이미지의 세밀한 특성을 보다 정확하게 재현할 수 있도록 돕는다[24]. 따라서 SE Block은 네트워크의 Convolutional Features들 사이의 상호 의존성을 명시적으로 모델링함으로써 네트워크의 표현력을 향상시킨다.

Figure 4에서 제시된 바와 같이 이 구조는 두 가지 주요 연산으로 구성되어 있으며, 첫 번째 연산인 Squeeze 연산(down)에서는 먼저 GAP을 활용하여 각 피처 맵의 전반적인 정보를 요약하고 Channel Descriptor로 압축한다. 이 과정에서 특히 각 채널에서 가장 중요한 정보만을 추출하는데 중점을 둔다. 두 번째 연산인 Excitation 연산(up)에서는 압축된 정보를 바탕으로 채널 간 의존성을 계산하고 FC를 통해 각 채널의 활성화 가중치를 조정한다. 이러한 재조정 과정을 통해 네트워크는 선택적으로 중요한 특성에 집중하고 결과적으로 고해상도 이미지 처리의 정확성을 높이며, 관련성이 높은 정보 처리를 최적화한다[25].

이후 Inception 모듈과 SE Block을 모두 통과한 피처들은 GAP을 거쳐 공간적 차원이 감소된 후 Flatten Layer를 통해 일차원 벡터로 변환된다. 이후 Leaky ReLU 활성화



**Figure 5.** Feature map  
: Existing Discriminator



**Figure 6.** Feature map  
: Proposed Discriminator

합수와 연속된 FC를 통해 출력을 생성하며, 최종적으로 어떤 이미지가 더 사실적인지를 판단하여 손실을 구하게 된다. 이러한 과정을 통해 복잡한 초해상도 이미지에서 중요한 특징을 정확히 추출하고 분류할 수 있도록 돕는다.

## 4. 평가

본 연구에서는 ESRGAN의 새로운 판별자 구조를 제안하며, 이를 기존의 ESRGAN 판별자와 비교하여 그 성능을 평가하였다. 제안하는 판별자 구조는 Inception 모듈과 SE Block을 병합한 구조를 도입함으로써 특징 추출 능력과 네트워크의 판별 정확도를 향상하였다. 이 구조는 다양한 크기의 Receptive field를 가진 Inception 모듈과 각 채널의 중요도를 동적으로 조절할 수 있는 SE Block의 조합을 통해 더욱 정교하고 세밀한 피처를 추출할 수 있도록 설계되었다.

### 4.1 실험 데이터 및 설정

본 실험에서는 고해상도 이미지 모델 훈련을 위한 표준 데이터 세트인 Div2K[26]를 사용하였으며, 4배 업스케일 초해상도 복원을 진행하였다. 이 데이터 세트는 고해상도와 다양한 장르와 주제를 포함하는 이미지를 제공하여 초해상도 모델의 학습과 성능평가에 적합하다. 학습과 평가를 위해 데이터 세트는 학습, 검증, 테스트 세트로 분할되었으며, Bicubic 보간법을 사용하여 저해상도 이미지를 생성하였다.

**Table 1.** Comparison of Feature Map  
of Discriminator

	Existing Discriminator	Proposed Discriminator
Mean	0.1273	0.1647
Std Dev	0.2752	0.2398
Max	2.6276	2.8624
Min	-0.4468	-0.3957

### 4.2 Feature Map 분석

피처 맵 분석은 딥러닝 모델의 중간 출력값을 시각화하여 모델이 이미지의 어떤 특성을 인식하고 있으며, 각 층을 통과하면서 어떠한 정보가 추출되는지를 파악하는 방법이다. 이 방법을 통해 본 연구에서 제안한 구조의 효과를 기존 모델과 비교하여 시각적으로 검증할 수 있다[27]. 이러한 접근은 모델이 처리하는 이미지의 세밀한 특성을 더욱 명확하게 이해할 수 있게 하며, 구조적 개선이 실제 성능 향상에 어떻게 기여하는지를 명시적으로 알 수 있다. Figure 5, 6은 기존 판별자 모델과 본 연구에서 제안하는 판별자 모델이 생성한 초기 피처 맵의 시각적 비교이다. 네트워크의 초기 레이어는 이미지의 세밀한 특징과 같은 저수준의 피처를 파악하고, 네트워크가 깊어질수록 구조적인 고수준의 피처를 파악하게 된다. Figure 5는 기존 판별자에 의해 처리된 결과를 시각화한 것이며, Figure 6은 제안하는 판별자에 의한 결과이다. 이러한 시각적 분석은 모델 간의 성능 차이를 직관적으로 이해할 수 있도록 돕는다.

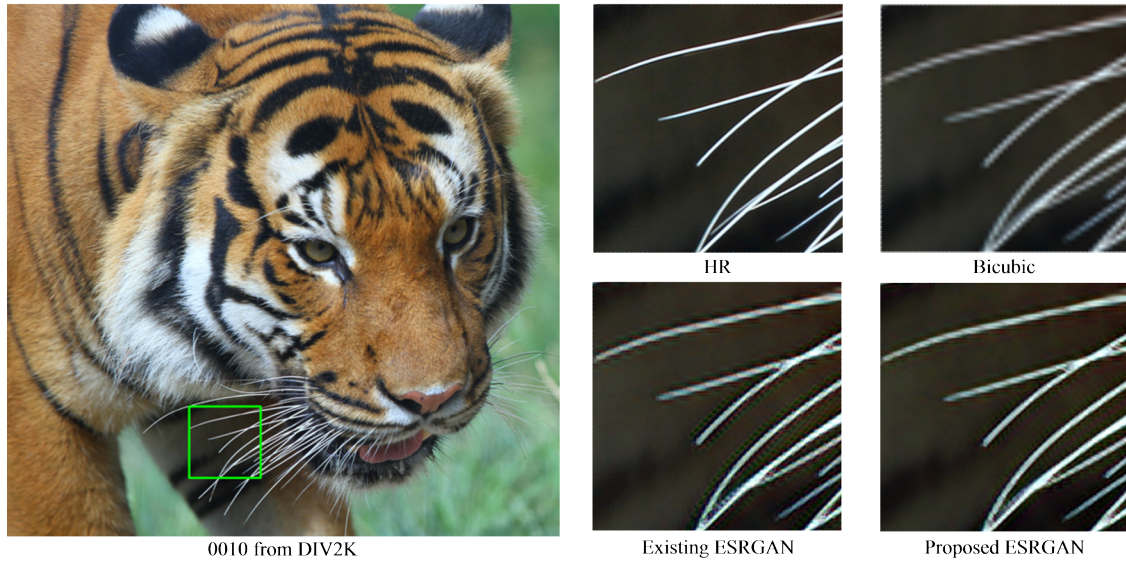


Figure 7. Qualitative result of Existing and Proposed ESRGAN

Table 1은 각 판별자 모델에 초기 레이어의 활성화 정도를 시각화한 결과이다. 각 수치는 Figure 5, 6의 피쳐 맵에 평균 활성화 값(Mean), 표준편차(Std Dev), 최대 활성화 값(Max), 최소 활성화 값(Min)을 계산 결과이다.

제안하는 판별자 모델을 적용한 Figure 6의 평균 활성화 값(Mean)은 0.16으로, 0.12인 기존 판별자 모델을 적용한 Figure 5에 비해 높다. 이는 활성화 된 픽셀의 강도가 더 높다는 것을 의미하며, 해당 레이어가 이미지의 특징을 더 강하게 감지하고 있을 수 있음을 시사한다. 표준편차(Std Dev)는 Figure 5가 0.27로, Figure 6보다 높은 수치를 보인다. 이는 기존 판별자 모델의 픽셀 간 활성화 된 정도의 차이가 더 크다는 것을 나타내며, 상대적으로 픽셀 간의 미세한 차이를 효과적으로 감지하지 못하고 있음을 의미한다. 최대 활성화 값(Max)은 Figure 6가 2.86으로 Figure 5보다 높다. 이는 중요한 특징 부분에서 강한 활성화를 보인다는 것을 의미한다. 최소 활성화 값(Min)은 Figure 5가 -0.44로, Figure 6보다 음의 활성화 값이 크게 나타난다. 이는 중요도가 낮은 특징을 강하게 억제하는 경향이 있음을 나타낸다.

결과적으로, 제안하는 판별자 모델의 피쳐 맵은 기존 판별자 모델에 비해 더욱 세밀하고 구조화된 감지 능력을 보여준다. 색상의 변화와 텍스처의 세밀함에서 뚜렷한 차이를 관찰할 수 있으며, 이는 개선된 모델이 이미지의 복잡한 디테일과 세밀한 특징을 더 잘 포착하고 있음을 시사한다. 따라서 Figure 6의 피쳐 맵에서는 특히 색상 대비와 객체의 경계가 더 명확하게 표현되어 있어 이미지의 세밀한 특징을 보다 잘 파악하고 있음을 알 수 있다. 이러한 시각적 비교를 통해 두 모델의 판별 능력에 있어서 제안하는 판별자 모델이 기존 판별자 모델보다 보다 향상되었음을 알 수 있다.

Table 2. Evaluation of Existing ESRGAN and Proposed ESRGAN

	Existing ESRGAN	Proposed ESRGAN
PSNR	27.0428	29.0720
SSIM	0.8065	0.8230
LPIPS	0.2960	0.2782
NIQE	7.0558	5.8521

이는 ESRGAN이 더 효과적으로 고해상도 이미지를 생성하며, 더 정확하게 고해상도와 저해상도 이미지 간의 차이를 구분할 수 있음을 의미한다.

### 4.3 이미지 품질 비교

본 연구에서는 기존의 ESRGAN 모델과 제안하는 ESRGAN 모델의 성능을 평가하기 위해 이미지 품질 평가지표인 Peak Signal-to-Noise Ratio(PSNR), Structural Similarity Index Measure(SSIM), Learned Perceptual Image Patch Similarity(LPIPS) 및 Natural Image Quality Evaluator(NIQE)를 사용하여 각 초해상도 이미지를 비교 분석하였다. 이러한 지표들은 초해상도 모델이 생성한 이미지 품질의 개선 정도를 객관적으로 측정할 수 있도록 도와 모델의 정량적인 성능 향상 정도를 알 수 있다. 또한 생성된 초해상도 이미지의 질적인 평가를 통해 시각적으로 개선된 부분을 확인하였다.



먼저 Table 2의 PSNR에서는 제안하는 모델이 29.07dB로 기존 모델의 27.04dB보다 높은 성능 향상을 보였다. 이는 제안하는 판별자 모델이 저화질 이미지의 노이즈를 보다 효과적으로 개선하며, 이미지의 세부 정보를 더 잘 보존하고 있음을 의미한다. SSIM은 0.80에서 0.82로 개선되어 제안하는 판별자 모델이 이미지의 구조적 유사성을 보다 정확하게 재현할 수 있음을 보인다. 이는 향상된 모델이 원본 이미지와의 구조적 및 시각적 유사성을 더 잘 보존하고 있다는 것을 의미한다. LPIPS는 기본 모델의 0.29에서 0.27로 감소하였다. 이는 제안하는 모델이 인간의 시각적 지각에 더 부합하는 이미지를 생성하고 있음을 나타내며, 시각적 품질이 개선되었음을 의미한다. NIQE 수치는 7.05에서 5.85로 감소하였다. 이는 제안하는 모델이 생성한 이미지의 자연스러움과 질감이 기존 모델에 대비 더욱 향상되었음을 의미한다.

Figure 7의 질적 평가에서는 기존 모델과 제안하는 모델의 이미지 복원 성능을 비교하였다. 각 이미지는 원본 고해상도 이미지(HR), Bicubic 보간법으로 생성한 저해상도 이미지(Bicubic), 기존 판별자 모델로 훈련된 ESRGAN(Existing ESRGAN), 그리고 제안하는 판별자 모델로 훈련된 ESRGAN(Proposed ESRGAN)으로 구성된다. 모든 이미지를 약 70배 확대하여 차이를 비교한 결과, 제안하는 판별자 모델(Proposed ESRGAN)을 적용한 경우 기존 판별자 모델(Existing ESRGAN)에 비해 노이즈와 왜곡이 현저히 줄어든 것을 확인할 수 있었다. 이는 제안하는 판별자 모델이 생성자가 복원한 이미지의 디테일을 더 효과적으로 포착하여, 생성자가 보다 자연스러운 해상도로 이미지를 복원하도록 훈련되었음을 의미한다.

결과적으로 제안하는 판별자를 적용한 모델은 기존 판별자 적용 모델보다 전반적인 이미지 품질, 구조적 및 시각적 유사성, 그리고 자연스러움 면에서 유의미한 개선을 이루었음을 보여준다. 특히 세밀한 텍스처와 이미지의 전반적인 품질 면에서 뚜렷한 개선이 관찰되었으며, 이러한 결과는 Inception 모듈과 SE Block을 병합한 구조가 이미지의 다양한 스케일과 채널별 중요도를 보다 효과적으로 처리할 수 있음을 시사한다.

## 5. 결론

본 연구는 초해상도 모델의 성능 개선을 위해 Inception 모듈과 SE Block을 통합한 새로운 판별자 구조를 제안하였다. Inception 모듈의 사용은 다양한 크기의 커널을 병렬로 적용함으로써 네트워크가 여러 스케일의 세밀한 특징을 포착할 수 있는 능력을 크게 향상시킨다. 이러한 구조는 모델의 표현력을 증대시키면서도 추가적인 계산 비용을 최소화한다. SE Block은 적은 계산 비용으로 채널별 중요도를 조

정할 수 있으며, 이는 중요한 세부 사항을 강조하고 불필요한 정보를 억제함으로써 전체 네트워크의 계산 효율성을 개선한다.

본 연구에서는 ESRGAN 모델의 판별자 구조를 개선하여 다중 스케일 필터 및 채널 활성화 조절이 모델 성능에 미치는 영향을 실험적으로 검증하였다. 또한 판별자의 구조적 개선이 세밀한 디테일의 포착 능력과 계산 효율성 향상에 기여 정도에 대해 분석하였다. 이러한 개선을 통해 제안하는 모델은 기존 ESRGAN이 가지고 있는 한계를 극복하고, 더욱 높은 품질의 고해상도 이미지를 생성할 수 있었다. 결과적으로 모델은 다양한 스케일의 이미지 특징을 보다 효과적으로 학습하고, 세밀한 차이에 집중함으로써 고해상도 이미지 생성의 질을 향상시켰다. 제안하는 모델은 이미지 품질 평가지표에서 PSNR은 29.07dB, SSIM은 0.82, LPIPS는 0.27, NIQE는 5.85를 달성하였다. 이는 기존 모델 대비 우수한 결과로 본 연구의 기여도를 입증한다.

향후 연구 방향으로서는 계산 효율성에 대해 연구해 볼 필요가 있다. Inception 모듈과 SE Block의 병합은 네트워크의 성능을 개선하는데 도움을 주었지만, 이러한 구조적 개선이 모델의 계산 복잡도와 자원 사용량에 미치는 영향에 대해 아직 탐구가 필요하다. 특히 대용량 이미지나 실시간 이미지 처리 요구 사항을 가진 응용 프로그램에서 모델의 실용성을 평가하기 위해서는 이러한 측면에서의 추가적인 최적화가 요구된다. 또한 제안하는 모델의 구조적 개선을 더욱 발전시켜 다양한 스케일과 복잡도를 가진 이미지에 대한 처리 능력을 강화하는 것을 고려할 수 있다. 이와 함께 계산 효율성과 모델의 표현력을 동시에 높일 수 있는 새로운 기술의 탐색도 중요한 연구 주제가 될 것이다[28]. 본 연구는 딥러닝 기반 화질 개선 분야에 있어 중요한 기술적 진보를 제시하며, 고해상도 이미지 개선 기술의 발전에 기여할 것으로 기대된다. 제안하는 모델의 개선 사항은 향후 이미지 처리, 컴퓨터 비전, 의료 영상 분석 등 다양한 분야에서 고화질 이미지가 요구되는 응용에 중요한 영향을 미칠 것이다. 이를 통해 본 연구는 SR 기술의 발전뿐만 아니라 관련 응용 분야의 발전에도 기여할 것으로 전망된다.

## 감사의 글

이 논문은 2024학년도 한국성서대학교 대학혁신지원사업의 지원을 받아 수행된 연구임.

## References

- [1] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365-338



7, 2020.

- [2] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53-65, 2018.
- [3] T.-Y. Lim, Y.-J. Cho, S.-H. Heo, and J.-K. Yoo, "Development of compound eye image quality improvement based on ESRGAN," *Journal of the Korea Computer Graphics Society*, vol. 30, no. 2, pp. 11-19, 2024.
- [4] N. Z. F. N. Azam, H. Yazid, and S. A. Rahim, "Super resolution with interpolation-based method: a review," *IJRAR-International Journal of Research and Analytical Reviews*, vol. 9, no. 2, pp. 168-174, 2022.
- [5] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1-34, 2020.
- [6] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 624-632, 2017.
- [7] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4539-4547, 2017.
- [8] B. Liu and J. Chen, "A super resolution algorithm based on attention mechanism and SRGAN network," *IEEE Access*, vol. 9, pp. 139138-139145, 2021.
- [9] 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," *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 1-15, 2018.
- [10] H. Chen, J. Gu, and Z. Zhang, "Attention in attention network for image super-resolution," *arXiv preprint arXiv:2104.09497*, 2021.
- [11] Y. Rao and O.-S. Kwon, "Deep residual dense network for single image super-resolution," *Electronics*, vol. 10, no. 5, pp. 555, 2021.
- [12] N. Rakotonirina, N. Carraz, and A. Rasoanaivo, "Further improving enhanced super-resolution generative adversarial network," *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.
- [13] J. Liang, H. Zeng, and L. Zhang, "Details or Artifacts: A Locally Discriminative Learning Approach to Realistic Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5657-5666, 2022.
- [14] J. Song, H. Yi, W. Xu, X. Li, B. Li, and Y. Liu, "Dual Perceptual Loss for Single Image Super-Resolution Using ESRGAN," *arXiv preprint arXiv:2201.06383*, 2022.
- [15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.
- [16] X. Ding, X. Zhang, J. Han, and G. Ding, "Diverse Branch Block: Building a Convolution as an Inception-like Unit," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10886-10895, 2021.
- [17] A. Song and J. Choi, "Fully Convolutional Networks with Multiscale 3D Filters and Transfer Learning for Change Detection in High Spatial Resolution Satellite Images," *Remote Sensing*, vol. 12, no. 5, pp. 799, 2020.
- [18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132-7141, 2018.
- [19] X. Cheng, X. Li, J. Yang, and Y. Tai, "SESR: Single image super resolution with recursive squeeze and excitation networks," in *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 147-152, IEEE, August 2018.
- [20] J. He, W. Shi, K. Chen, L. Fu, and C. Dong, "GCFSR: A Generative and Controllable Face Super Resolution Method Without Facial and GAN Priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1889-1898, 2022.
- [21] Y. Zhang, K. Li, K. Li, and Y. Fu, "MR Image Super-Resolution With Squeeze and Excitation Reasoning Attention Network," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13425-13434, 2021.
- [22] F. Hao, J. Wu, H. Lu, J. Du, and J. Xu, "Large coordinate kernel attention network for lightweight image super-resolution," *arXiv preprint arXiv:2405.09353*, 2024.
- [23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 1-15, February 2017.
- [24] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent Spatial and Channel 'Squeeze & Excitation' in Fully Convolutional Networks," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, Springer International Publishing, pp. 421-429, 2018.
- [25] C. Ma, "Uncertainty-Aware GAN for Single Image Super Resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1000-1007, 2020.

ial Intelligence, pp. 4071-4079, 2024.

[26] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1122-1131, 2017.

[27] Y. Liao, Y. Gao, and W. Zhang, "Feature activation map: Visual explanation of deep learning models for image classification," arXiv preprint arXiv:2307.05017, 2023.

[28] A. Bulat, J. Yang, and G. Tzimiropoulos, "To Learn Image Super-Resolution, Use a GAN to Learn How to Do Image Degradation First," in Proceedings of the European Conference on Computer Vision (ECCV), pp. 185-200, 2018.

## 〈 저자 소개 〉

### 김 예 은

- 2020년 3월 ~ 현재 한국성서대학교  
컴퓨터소프트웨어학과 (학사)
- 관심 분야 : 인공지능, 딥러닝
- <https://orcid.org/0009-0006-6354-7478>



### 남 아 영

- 2021년 3월 ~ 현재 한국성서대학교  
컴퓨터소프트웨어학과 (학사)
- 관심 분야 : 인공지능, 딥러닝
- <https://orcid.org/0009-0005-4867-0809>



### 양 단 아

- 2014년 2월 : 한국성서대학교  
컴퓨터소프트웨어학과 (공학사)
- 2023년 2월 : 이화여자대학교  
컴퓨터공학(석박통합)
- 관심 분야 : 블록체인, 머신러닝, 딥러닝,  
네트워크 보안
- <https://orcid.org/0000-0001-6463-9784>

