

파편 탐지 성능 향상을 위한 딥러닝 초해상도화 효과 분석

이 유 석^{*,1)}

¹⁾ 국방과학연구소 국방시험연구원

Analysis of the Effect of Deep-learning Super-resolution for Fragments Detection Performance Enhancement

Yuseok Lee^{*,1)}

¹⁾ Defense System Test Center, Agency for Defense Development, Korea

(Received 12 February 2023 / Revised 4 April 2023 / Accepted 30 May 2023)

Abstract

The Arena Fragmentation Test(AFT) is designed to analyze warhead performance by measuring fragmentation data. In order to evaluate the results of the AFT, a set of AFT images are captured by high-speed cameras. To detect objects in the AFT image set, ResNet-50 based Faster R-CNN is used as a detection model. However, because of the low resolution of the AFT image set, a detection model has shown low performance. To enhance the performance of the detection model, Super-resolution(SR) methods are used to increase the AFT image set resolution. To this end, The Bicubic method and three SR models: ZSSR, EDSR, and SwinIR are used. The use of SR images results in an increase in the performance of the detection model. While the increase in the number of pixels representing a fragment flame in the AFT images improves the Recall performance of the detection model, the number of pixels representing noise also increases, leading to a slight decreases in Precision performance. Consequently, the F1 score is increased by up to 9 %, demonstrating the effectiveness of SR in enhancing the performance of the detection model.

Key Words : Arena Fragmentation Test(아레나 파편 시험), Deep-learning(딥러닝), Object Detection(물체 탐지), Super-resolution(초해상도화)

1. 서 론

탄두는 폭발물과 이를 감싼 금속으로 구성된다. 탄

두 내부 폭발물이 폭발하면서 발생하는 고온의 화염과 높은 폭발 압력, 그리고 파편이 복합적으로 작용하여 목표물을 파괴한다.

탄두의 성능을 평가하기 위해서는 폭발 압력과 파편의 특성, 그리고 파편의 분포를 측정해야 한다. 탄두 성능을 평가하기 위한 시험은 세 가지가 있다. 파

* Corresponding author, E-mail: ysl@add.re.kr
Copyright © The Korea Institute of Military Science and Technology

편의 중량분포 평가를 위한 PIT test, 과편 분산 위력 확인을 위한 Panel test, 그리고 폭발 압력과 과편 속도, 과편 공간분포를 종합적으로 평가하기 위한 Arena Fragmentation Test(AFT)가 있다¹¹⁾.

국방과학연구소 국방시험연구원(시험원)에서는 탄두 성능 평가를 위하여 AFT 시험을 수행하였다. 시험원에서 수행된 AFT 시험의 구성은 Fig. 1과 같다. 먼저 탄두를 중심으로 일정 거리에 철판을 ‘ㄷ’자 형태로 배치하고, 고속카메라를 철판 뒤에 설치하였다. 압력 센서를 탄두 옆에 설치하여 폭발 압력을 측정하였고, 철판을 통과하는 과편에 의해 발생하는 화염을 고속카메라로 측정하였다. 이때 철판을 통과하는 과편의 수가 수십 ~ 수백 개로, 사람이 분석하기에는 비용과 시간이 많이 소요된다. 이에 시험원에서는 딥러닝 기반 과편 탐지 기술을 활용하여 탄두 과편의 속도를 측정하고 과편 분포를 분석하였다¹²⁻³⁾.

AFT 시험 시 과편이 철판을 통과하는 순간의 화염을 포착하기 위해 고속 카메라의 셔터 속도는 초당 약 10,000 ~ 30,000번으로 설정하였다. 빠른 셔터 속도로 인해 해상도를 150 × 150 ~ 600 × 600 픽셀 수준으로 낮춰 촬영하였고, 이 때문에 처음 과편이 철판을 관통하는 순간 촬영되는 화염의 크기가 3 × 3 ~ 10 × 10 픽셀 정도로 매우 작았다. 초기 화염을 차지하는 픽셀의 수가 너무 적기 때문에 과편 탐지 모델이 이를 탐지하지 못하거나 과편의 화염이 크게 퍼질 때까지 탐지하지 못하는 등의 문제가 발생하였다. 낮은 해상도의 한계로 인해 발생하는 과편 탐지 모델의 성능 저하 문제를 해결하기 위하여 본 논문에서는 초해상도화 기법을 적용해보고자 하였다.

초해상도화란 저해상도 이미지를 고해상도 이미지로 바꿔주는 기법으로, 최근 딥러닝 기반 초해상도화 기법이 많이 연구되고 있다. 본 논문에서는 고전적인 초해상도화 기법인 Bicubic 보간법과 딥러닝 기반 초해상도화 모델인 ZSSR¹⁴⁾, EDSR¹⁵⁾, SwinIR¹⁶⁾을 이용하여 고속카메라에서 촬영된 이미지의 해상도를 높였다.

초해상도화 기법으로 이미지의 해상도를 높여 과편 탐지를 수행한 결과, 과편 탐지 모델의 성능 분석 지표인 정밀도(Precision), 재현율(Recall), 그리고 F1 score가 높게 나타났다. 특히 SwinIR 모델로 이미지의 해상도를 두 배 높인 경우 F1 score가 원본 이미지보다 9 % 높게 나타나 이미지의 해상도를 높이는 것이 과편 탐지 모델의 성능 향상으로 이어질 수 있음을 확인하였다.

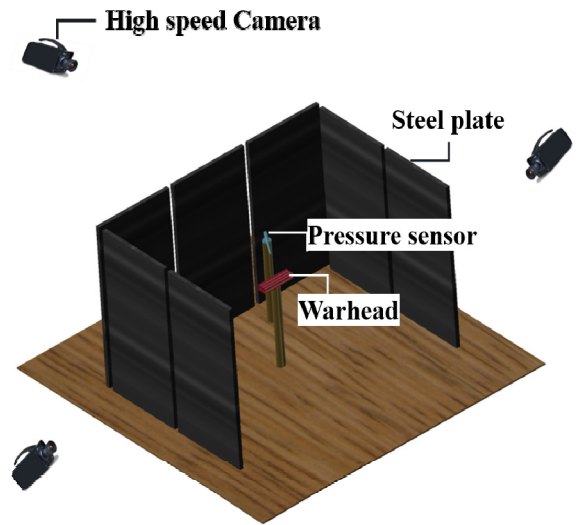


Fig. 1. AFT test environments

2장에서는 AFT 시험을 통해 얻은 이미지에 대한 설명을, 3장은 초해상도화 기법에 대한 설명을 기술하였다. 4장과 5장에서는 초해상도화에 따른 과편 탐지 모델의 성능 변화 결과와 결과 고찰을 기술하였고, 6장에서 결론을 맺는다.

2. AFT 이미지 특성

AFT 시험을 수행하며 시험마다 동일한 고속카메라를 사용할 수 없어, 고속카메라의 성능에 따라 서로 다른 해상도를 가진 시험 이미지를 확보하였다. 총 82 번의 탄두 시험을 수행하였고, 이를 통해 총 7244장의 AFT 이미지를 확보하였다.

시험별 AFT 이미지의 해상도를 Fig. 2에 히스토그램으로 나타냈다. 이미지의 넓이와 동일한 정사각형의 한 변의 길이를 특성 길이로 정의하였다. Fig. 2를 보면 시험별 AFT 이미지의 해상도가 150 × 150 ~ 600 × 600 픽셀로 낮은 것을 알 수 있다. 이로 인해 과편이 철판을 관통하는 순간, 화염을 나타내는 픽셀은 약 3 × 3 ~ 10 × 10 픽셀로 매우 작게 나타났다. AFT 이미지의 해상도가 낮고, 관심의 대상인 과편의 크기도 작기 때문에 AFT 이미지에 대해 딥러닝 기반 과편 탐지 수행 시, 과편 탐지 모델에서 실제 과편임에도 과편으로 탐지되지 않는 문제점이 존재하였다.

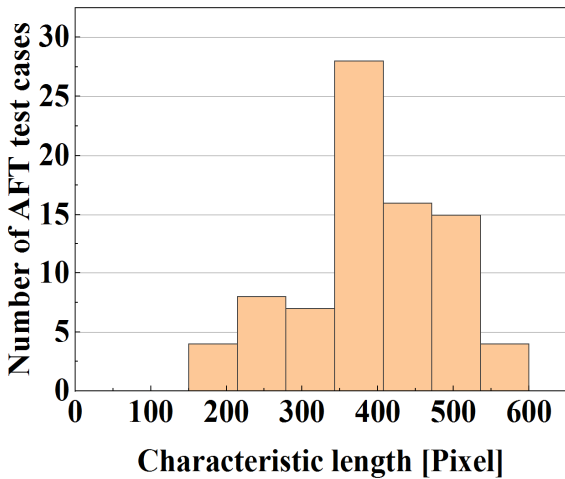


Fig. 2. Resolution histogram of the AFT test cases

3. 초해상도화 기법

초해상도화 기법이란 저해상도 이미지를 고해상도 이미지로 바꿔 해상도를 높여주는 기법이다. 딥러닝에 의한 초해상도화 기법이 적용되기 이전에는 Bilinear, Bicubic 등 보간법 기반의 초해상도화 기법이 사용되었다. 현재도 보간법 기반의 초해상도화 기법이 많이 사용되지만, 선형적인 방식의 한계로 이미지가 흐릿해지고, 복잡한 이미지의 초해상도화가 불가능하다는 단점이 있다. 반면 딥러닝 기반의 초해상도화는 ReLU 함수와 같은 비선형 활성화 함수를 이용해 선형성의 한계를 극복하였기 때문에 보간법 기반의 초해상도화 기법보다 초해상도화 성능이 좋게 나타난다.

Dong^[7]이 SRCNN(Super Resolution Convolution Neural Network) 모델을 처음 제안함으로써 딥러닝 기반의 초해상도화 기법 연구가 시작되었다. 많은 훈련 데이터를 필요로 하지 않는 ZSSR, 신경망 층을 깊게 만들어 성능을 높인 EDSR, 자연어 처리 기반으로부터 파생된 swin transformer 기반의 SwinIR 모델 등이 그것이다.

본 논문에서는 Bicubic 보간법, ZSSR, EDSR, 그리고 SwinIR을 이용하여 AFT 이미지에 대해 초해상도화를 수행하였다. EDSR과 SwinIR 모델은 Div2K 이미지 세트로 훈련된 모델을 이용하였고, AFT 결과 이미지 총 7244장 중 5962장을 이용하여 fine-tuning을 수행하였다. 테스트에는 다섯 개의 시험 케이스, 총 1282장의

이미지를 사용하였다. 모델 훈련 시, lr(learning rate)은 $1e-7$, optimizer는 Adam, batchsize는 32를 적용하였다. 훈련은 psnr이 더 이상 증가하지 않는 지점까지 수행하였고, EDSR은 epoch 5, SwinIR은 epoch 100까지 훈련시킨 모델을 사용하였다.

3.1 ZSSR model

ZSSR은 제로샷 기법을 활용한 초해상도화 모델이다. 한 개 이미지 내부의 정보를 이용하기 때문에 많은 데이터를 이용한 훈련이 필요 없는 모델이다. 일반적으로 초해상도화 딥러닝 모델은 훈련 방법에 의해 그 결과가 영향을 받는다. 예를 들어 Bicubic 보간법으로 축소된 이미지의 해상도를 높이도록 훈련된 경우 기타 노이즈에 의한 화질 저하는 개선하지 못하는 경우가 있다. 하지만 ZSSR은 한 개 이미지 내부의 정보를 활용하기 때문에 많은 훈련 데이터가 필요 없는 장점을 가질 뿐만 아니라 훈련 데이터에서부터 나오는 한계를 일부 극복한 모델이라고 할 수 있다.

ZSSR 모델의 구조는 Fig. 3과 같다. 먼저 하나의 이미지에서 내재된 정보를 충분히 활용하기 위해 이미지를 0, 90, 180, 270° 회전시키거나 좌우/상하 반전을 통해 데이터 증강 작업을 수행한다. 이런 전처리를 함으로써 ZSSR 모델은 통계학적으로 증강된 데이터 세트에 대한 훈련을 수행한다. 이렇게 만들어진 저해상도 이미지 $I_{LR} \in R^{H \times W \times C}$ (H: 이미지 높이, W: 이미지 폭, C: 이미지 채널)을 이용해 훈련을 수행한다. ZSSR 모델의 훈련은 먼저 원본 이미지의 해상도를 Bicubic 기법으로 줄여 I_{LR} 을 만들어 준다. 이 이미지는 식 (1)과 같이 Conv(Convolutional Layer)와 활성화 함수 ReLU로 구성된 H_{CR} 층을 통과한다. 본 논문에서는 8개의 H_{CR} 층을 통과하도록 하였다.

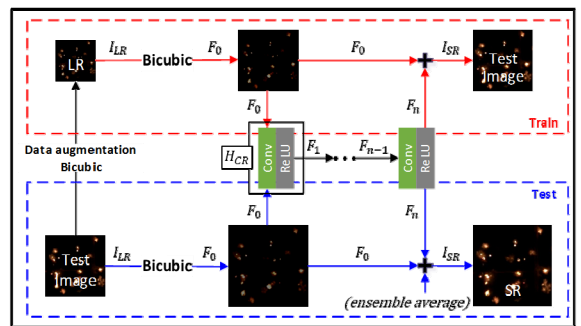


Fig. 3. ZSSR model structure

$$F_n = H_{CR}(F_{n-1}), n = 1, 2, \dots, 8 \quad (1)$$

H_{CR} 층을 통과한 특징맵 F_n 과 잔차 학습을 위해 F_0 가 마지막 단계에서 합쳐지며 초해상도화 이미지 I_{SR} 을 구하게 된다. 그리고 손실 함수 값을 줄여가며 훈련을 반복한다. 훈련이 마무리되는 기준은 lr을 기준으로 한다. 초기 lr을 1e-3으로 설정 후 훈련을 시작한 뒤, 훈련을 하면서 주기적으로 reconstruction error를 선형 피팅하여 오차의 표준편차가 선형 피팅의 기울기보다 클 때, lr을 10으로 나눠주어 훈련을 반복한다. 결과적으로 lr이 1e-6보다 낮아질 경우 반복을 멈추게 된다. 본 논문에서 AFT 이미지의 경우 약 1500 ~ 2000회 정도의 반복이 필요하였다. 반복 훈련이 끝난 뒤 geometric self-ensemble을 하여 초해상도화 이미지 $I_{SR} \in R^{x \times H \times x \times W \times C}$ (초해상도화에 따라 이미지의 높이와 폭이 \times 배만큼 증가)를 얻게 된다.

3.2 EDSR model

모델 내 네트워크층을 깊게 할수록 좋은 성능을 보이는 특징을 이용한 모델 중 하나가 EDSR이다. EDSR은 네트워크층이 깊어질 때 성능이 좋아지는 장점을 활용하고, 기울기 소실 등의 문제를 해결하기 위하여 ResNet^[8]에서 활용한 잔차 학습을 이용하였다.

EDSR 모델의 구조는 Fig. 4와 같다. 저해상도 이미지 $I_{LR} \in R^{H \times W \times C}$ 이 식 (2)의 Conv층을 통과하여 특징맵 F_0 가 출력 된다.

$$F_0 = Conv(I_{LR}) \quad (2)$$

여기서 나온 F_0 는 Conv층, ReLU층, 상수값을 곱해주는 Multi층으로 구성된 32개의 H_{CRCM} 층을 통과한다. 그리고 skip connection에 의한 잔차 학습이 반복되며 식 (3)과 같이 특징맵 F_n 이 추출된다. 잔차 학습층을 통해 구해진 F_n 은 식 (4)와 같이 한 번 더 Conv층을 통과한다. 마지막으로 식 (5)와 같이 해상도를 높여주는 pixel shuffle층이 포함된 H_{CPRC} 층을 통과한 뒤 초해상도화 이미지 $I_{SR} \in R^{x \times H \times x \times W \times C}$ 을 얻게 된다.

$$F_n = H_{CRCM}(F_{n-2}) + F_{n-2}, n = 2, 4, \dots, 64 \quad (3)$$

$$F_k = Conv(F_n), n = 64 \quad (4)$$

$$I_{SR} = H_{CPRC}(F_0 + F_k) \quad (5)$$

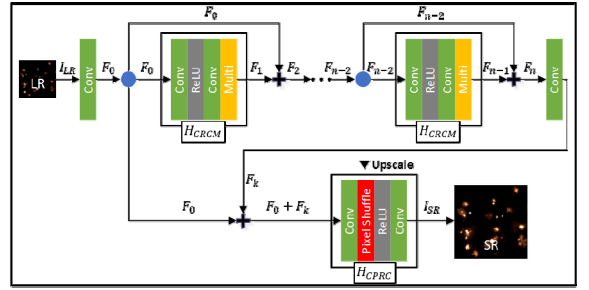


Fig. 4. EDSR model structure

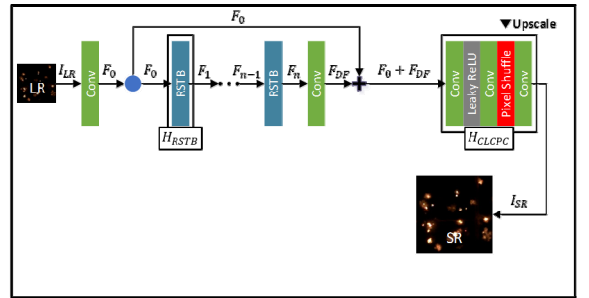


Fig. 5. SwinIR model structure

3.3 SwinIR model

2021년 발표된 SwinIR은 swin transformer를 기반으로 하는 초해상도화 모델이다. Swin transformer는 이미지를 패치별로 나눠 훈련하는 특징을 가지며, 공간적 정보를 획득하는 데 있어 Conv보다 더 유리한 특징을 가진다. 이러한 장점으로 초해상도화 분야에도 transformer 기반 연구가 수행되어 swinIR 모델이 발표되었다.

SwinIR 모델의 구조는 Fig. 5와 같다. SwinIR모델에서 훈련에 사용된 저해상도 이미지는 앞의 세 모델과 다르게 Bicubic 기법으로 해상도를 줄여 I_{LR} 을 만들지 않고 BSRGAN모델^[9]에서 사용된 Degradation기법을 사용하였다. Degradation 기법으로 만들어진 저해상도 이미지 $I_{LR} \in R^{H \times W \times C}$ 는 식 (6) ~ (8)과 같이 Conv층과 6개의 residual swin transformer block(RSTB)층, 그리고 Conv층을 통과하게 된다. SwinIR 모델에서 RSTB층이 가장 중요한 역할을 한다. RSTB층은 Swin transformer로 구성돼 local attention과 cross-window interaction으로 이미지의 특징맵을 추출하게 된다. 즉, 식 (6)의 Conv 층에서는 이미지의 shallow feature를 추출하고 식 (7) ~ (8)에서는 deep feature(F_{DF})를 추출한다. F_{DF} 는

F_0 에 더해진 뒤 식 (9)와 같이 Pixel shuffle이 포함된 H_{CLCPC} 층을 지나며 이미지를 재구성하고 해상도를 높여 결과적으로 초해상도화 이미지 $I_{SR} \in R^{xH \times xW \times C}$ 를 얻게 된다.

$$F_0 = Conv(I_{LR}) \tag{6}$$

$$F_n = H_{RSTB}(F_{n-1}), n = 1, 2, \dots, 6 \tag{7}$$

$$F_{DF} = Conv(F_n), n = 6 \tag{8}$$

$$F_{DF} = H_{CLCPC}(F_0 + F_{DF}) \tag{9}$$

4. 결과

4.1 초해상도화 결과

초해상도화를 통해 이미지의 해상도를 두 배, 네 배 증가시켰다. 원본 이미지의 해상도를 두 배 높여준 결과 이미지를 2x 이미지, 네 배로 높여준 결과 이미지를 4x 이미지로 명명하였다.

초해상도화 수행 결과는 Fig. 6과 같다. 두 개의 작은 파편이 철판을 통과한 순간을 확대하여 나타냈다. Original 이미지에서 볼 수 있듯이 두 개의 작은 파편은 5×5 픽셀 정도로 작은 파편이고, 두 파편 간 거리가 한 개의 픽셀 정도로 매우 가까이 나타났다.

Bicubic 보간법으로 해상도를 높인 경우 화염을 이루는 픽셀 수는 증가하였으나 흐릿해지는 것을 볼 수 있다. 또한 두 파편 간 간격도 명확하게 구분되지 않는다. 이에 비해 딥러닝 기반 모델의 결과는 해상도도 커지고, 이미지의 질도 높아진 것을 볼 수 있다. ZSSR은 이미지에 내재된 정보를 활용하는 모델인 만큼 ZSSR 이미지에는 초해상도화 과정에서 함께 커진 노이즈의 영향이 파편 주위에 관찰됐다. EDSR 이미지는 ZSSR 이미지와 유사해 보이지만 노이즈는 약간 더 작은 것으로 보인다. SwinIR 이미지는 육안으로 보기에 가장 좋은 결과를 나타냈다. SwinIR 이미지는 ZSSR과 EDSR 결과에 비해 노이즈가 가장 적었다. 다른 모델의 경우 화염이 옆으로 번지는 듯한 결과가 나타났지만 SwinIR 이미지에서는 파편간 경계가 명확하게 구분되어 보이는 등 네 가지 기법 중 가장 좋은 성능을 보였다.

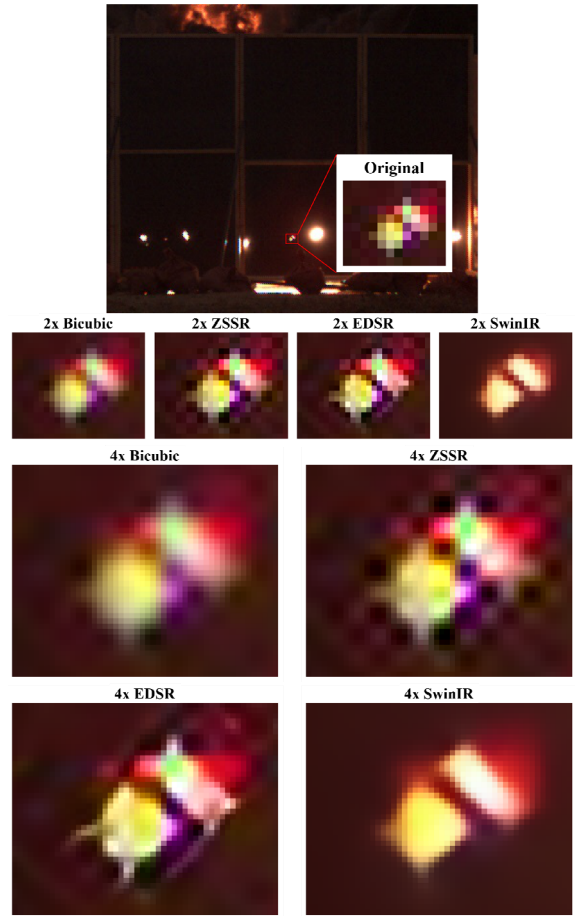


Fig. 6. An example of the result of the super-resolution

4.2 파편 탐지 결과

4.2.1 파편 탐지 모델 및 필터

초해상도화 이미지를 이용한 파편 탐지 성능 검증을 위해 파이토치 오픈 소스인 MMDetection^[10]을 사용하였다. Region of interest(ROI)를 기반으로 파편 탐지를 수행하기 위하여 탐지 알고리즘에는 Faster R-CNN 모델을^[11] 적용하였다. Faster R-CNN의 Backbone은 두 가지를 사용하였다. 먼저 초해상도화에 의한 파편 탐지 모델의 성능 향상 효과 분석을 위해 FPN이 적용된 ResNet-50을 사용하였고, 두 번째로 ResNet-50 기반 이외의 파편 탐지 모델에서의 초해상도화 효과를 보고자 Swin-T를 사용하였다. 탐지 모델의 Head는 RPN과 Classifier로 구성되어 객체 Localization과 객체 분류를 수행하였다. 고속영상 내에 배경을 제외하기 위해 철

관 부분을 ROI로 설정하여 ROI를 기준으로 파편을 탐지하도록 하였다. 7244개의 AFT 이미지 중 5962장의 이미지를 Faster R-CNN 탐지 모델 훈련에 사용하였고, 테스트에는 다섯 개의 시험 케이스, 총 1282장의 이미지를 사용하였다. 탐지 모델의 훈련 환경은 다음과 같다. lr은 1e-4, optimizer는 Adam, weight decay 0.05, batch size 32, epoch 36으로 설정하여 훈련하였다. 훈련 시 이미지의 해상도는 원본 이미지를 그대로 사용하였고, 탐지모델 내부 파이프라인에서 학습 이미지의 크기가 1333 × 800으로 수정된 후 학습이 이뤄졌다.

Faster R-CNN 모델이 파편을 탐지한 다음 파편 중복 탐지 방지를 위한 필터를 적용 한다. 파편 이미지의 특성상 첫 화염이 발생한 이후 프레임이 진행되어도 같은 위치에 화염이 유지된다. 시험 분석을 위해서는 최초 관통 시점에서의 화염만 필요하므로 탐지 모델이 최초 관통 시점 이후의 화염을 파편으로 인식하지 않도록 해야 한다. 이를 위해서 최초 관통 시점 화염 이후의 화염을 필터링 할 수 있도록 temporal filtering scheme을 사용하였다. 또한 화염이 시간에 따라 크기가 커지는 특성을 고려하여 초기 화염의 크기를 기준으로 비선형 최소 자승법인 Levenberg-Marquardt를 이용해 마스크 크기가 프레임마다 더 커지도록 하였다^[2]. Fig. 7과 같이 초기 파편을 기준으로 마스크를 씌워 처음 탐지된 파편의 화염이 이후의 프레임부터는 파편으로 탐지되지 않도록 하였고, 화염이 커짐에 따라 더 커진 새로운 마스크가 화염을 덮는 방식으로 필터링하였다.

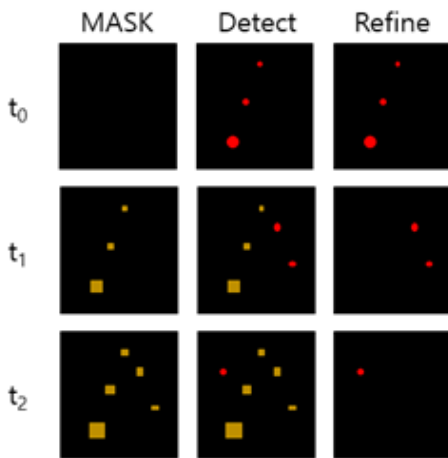


Fig. 7. Temporal filtering for detecting first penetrating fragment^[2]

4.2.2 파편 탐지 성능 지표

탐지 모델의 성능을 평가하기 위한 지표로 정밀도, 재현율, 그리고 F1 score를 사용하였다. 이 지표들은 True Positive(TP), False Positive(FP), 그리고 False Negative(FN)을 이용해 정의된다.

처음 화염이 이미지에 나타난 후 세 프레임 이내에 파편이라고 탐지해낸 경우 진짜 파편을 탐지했다고 정의하여 TP로 분류하였다. 노이즈를 파편으로 구분하거나 마스크를 벗어난 화염을 파편으로 구분한 경우, 가짜 파편을 진짜 파편으로 구분한 것으로 정의하여 FP로 분류하였다. 마지막으로 파편을 탐지하지 못하였거나 파편의 화염을 세 프레임 이내에 파편이라고 탐지하지 않은 경우, 진짜 파편을 가짜 파편으로 분류했다고 정의하였고 FN으로 분류하였다. 정밀도와 재현율은 식 (10) ~ (11)과 같다. 정밀도는 파편이라고 탐지한 것 중 진짜 파편의 비율을 나타내며, 재현율은 전체 파편 중 파편 탐지 모델이 파편을 찾아낸 비율을 나타낸다. F1 score는 정밀도와 재현율의 조화평균으로 식 (12)와 같다. 즉, F1 score는 정밀도와 재현율을 모두 고려하여 탐지 모델의 성능을 종합적으로 평가할 수 있는 평가지표다.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

4.2.3 파편 탐지 결과

다섯 개 시험 케이스의 해상도와 파편 개수 정보를 Table 1에 나타냈고, 2x 이미지를 이용한 파편 탐지 성능 지표인 정밀도, 재현율, 그리고 F1 score의 결과를 Table 2에 나타냈다. 그리고 4x 이미지를 이용한 파편 탐지 성능 지표인 정밀도, 재현율, 그리고 F1 score의 결과를 Table 3에 나타냈다.

Case 1은 464 × 640 해상도를 가진 AFT 이미지 181장으로 구성된다. 탐지해야 하는 파편의 개수는 215개로 해상도가 다른 케이스에 비해 크고 파편도 상대적으로 많다. Table 2의 Case 1에서 정밀도를 보면 Original(원본 이미지)의 값도 0.977로 높지만 초해상도화를 수행한 경우 ZSSR과 EDSR에서 더 높은 성능을

보였다. Case 1은 탐지해야 하는 파편의 개수가 많은 데, Original을 이용해 파편 탐지를 수행한 결과 재현율이 0.605로 낮게 나타났다. 하지만 2x 이미지를 보면 재현율이 5 ~ 33 % 증가한 것으로 나타났다. 특히 2x SwinIR의 재현율이 0.805로 가장 높게 나타났다. 정밀도와 재현율을 종합하여 판단할 수 있는 지표인 F1 score에서는 SwinIR이 0.878로 가장 높게 나타났다. 그러나 Case 1에 대해 4x 이미지로 파편 탐지를 수행한 결과 Table 3에서 볼 수 있듯이 정밀도에서는 초해상도화 전처리로 인해 오히려 성능이 떨어진 것으로 나타났다. 특히 SwinIR모델은 해상도가 네 배로 커지면서 파편 탐지 성능이 너무 민감해져, 노이즈를 파편으로 인식하거나 화염을 파편으로 인식하는 경우

가 늘어났다. 이로 인해 Original보다 정밀도가 18 % 감소했다. 재현율과 F1 score는 Original보다 더 높게 나타나긴 했으나 2x SwinIR에 비해서는 감소하였다.

Table 1. Resolution and number of fragments of the AFT test cases

AFT Test case	Resolution	# of fragments
Case 1	464 × 640	215
Case 2	360 × 208	28
Case 3	336 × 512	98
Case 4	352 × 512	99
Case 5	352 × 352	59

Table 2. Detection results for the 2x images in terms of precision, Recall, and F1 scores(ResNet-50 based Faster R-CNN)

Case	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Case 1	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Precision	0.977	0.946	0.993	0.993	0.966
Recall	0.605	0.647	0.660	0.633	0.805
F1 score	0.747	0.768	0.793	0.773	0.878
Case 2	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Precision	0.952	1.000	1.000	1.000	1.000
Recall	0.714	0.857	0.786	0.857	0.821
F1 score	0.816	0.923	0.880	0.923	0.902
Case 3	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Precision	0.973	0.900	0.986	0.958	0.922
Recall	0.735	0.735	0.725	0.714	0.857
F1 score	0.837	0.809	0.836	0.818	0.888
Case 4	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Precision	0.962	0.942	0.944	0.943	0.955
Recall	0.758	0.818	0.848	0.828	0.848
F1 score	0.847	0.876	0.894	0.882	0.898
Case 5	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Precision	0.933	0.900	0.894	0.896	0.922
Recall	0.750	0.804	0.750	0.768	0.839
F1 score	0.832	0.849	0.816	0.827	0.879
Average	Original	2x Bicubic	2x ZSSR	2x EDSR	2x SwinIR
Precision	0.959	0.934	0.963	0.958	0.953
Recall	0.712	0.772	0.754	0.760	0.834
F1 score	0.816	0.843	0.844	0.845	0.889

Case 2는 360 × 208 해상도를 가진 201개의 AFT 이미지로 구성된다. 탐지해야 하는 과편의 개수는 28개로 모든 케이스 중 가장 적다. Original의 정밀도와 재현율은 각각 0.952와 0.714로 나타났다. Table 2의 2x 이미지 정밀도는 모두 1.000으로 나타나 초해상도화로 인해 정밀도가 좋아졌다. 재현율 또한 2x 이미지에서 10 ~ 20 % 높아졌다. 하지만 Table 3의 4x 이미지에서는 Bicubic을 제외하고, Original보다 정밀도가 오히려 감소하였다. 이는 초해상도화에 의해 노이즈나 화염을 나타내는 픽셀 수가 증가하면서 노이즈나 마스크를 약간 벗어난 화염을 과편으로 탐지하는 횟수가 늘었기 때문이다. 재현율의 경우 4x EDSR을 제외하면 Original보다 증가하였고, 4x SwinIR의 경우 2x

SwinIR보다 재현율이 13 % 증가하였다. Case 2, 4x 이미지에서 F1 score는 Bicubic이 가장 좋게 나타났으며, 4x ZSSR도 F1 score가 소폭 상승하였다.

Case 3은 336 × 512 해상도를 가진 AFT 이미지 300장으로 구성된다. 탐지해야 하는 과편의 개수는 98개이다. Original의 정밀도와 재현율은 각각 0.973과 0.735로 나타났다. Table 2의 2x 이미지 정밀도는 2x ZSSR을 제외하면 초해상도화 이후 모두 감소하였다. 재현율도 2x SwinIR을 제외하면 Original보다 감소하였다. 2x SwinIR의 재현율은 0.857로 나타나 Original보다 17 % 향상됐다. F1 score도 2x SwinIR에서 0.888로 가장 높게 나타났다. Table 3의 4x 이미지에서는 4x SwinIR의 재현율을 제외하면 초해상도화에 의해

Table 3. Detection results for the 4x images in terms of precision, Recall, and F1 scores(ResNet-50 based Faster R-CNN)

Case 1	Original	4x Bicubic	4x ZSSR	4x EDSR	4x SwinIR
Precision	0.977	0.905	0.926	0.895	0.803
Recall	0.605	0.619	0.642	0.633	0.777
F1 score	0.747	0.735	0.758	0.741	0.790
Case 2	Original	4x Bicubic	4x ZSSR	4x EDSR	4x SwinIR
Precision	0.952	0.960	0.852	0.950	0.722
Recall	0.714	0.857	0.821	0.679	0.929
F1 score	0.816	0.906	0.836	0.792	0.813
Case 3	Original	4x Bicubic	4x ZSSR	4x EDSR	4x SwinIR
Precision	0.973	0.866	0.931	0.904	0.723
Recall	0.735	0.745	0.694	0.674	0.878
F1 score	0.837	0.802	0.795	0.772	0.785
Case 4	Original	4x Bicubic	4x ZSSR	4x EDSR	4x SwinIR
Precision	0.962	0.911	0.912	0.953	0.894
Recall	0.758	0.828	0.838	0.828	0.848
F1 score	0.847	0.868	0.874	0.886	0.870
Case 5	Original	4x Bicubic	4x ZSSR	4x EDSR	4x SwinIR
Precision	0.933	0.846	0.840	0.909	0.831
Recall	0.750	0.786	0.750	0.714	0.875
F1 score	0.832	0.815	0.792	0.800	0.852
Average	Original	4x Bicubic	4x ZSSR	4x EDSR	4x SwinIR
Precision	0.959	0.898	0.892	0.922	0.795
Recall	0.712	0.767	0.749	0.706	0.861
F1 score	0.816	0.825	0.811	0.798	0.822

정밀도, 재현율, 그리고 F1 score가 모두 감소하는 결과가 나타났다.

Case 4는 352 × 512 해상도를 가진 AFT 이미지 300장으로 구성된다. 탐지해야 하는 파편의 개수는 99개이다. Original의 정밀도가 Table 2 ~ 3의 결과를 통틀어 가장 높게 나타났다. 재현율은 2x ZSSR, 2x SwinIR, 4x SwinIR에서 0.848로 가장 높게 나타났다. F1 score는 2x SwinIR과 4x EDSR에서 가장 높게 나타났으며, 4x EDSR보다 2x SwinIR의 F1 score가 약간 더 높게 나타나 2x SwinIR의 성능이 가장 높게 나타났다.

Case 5는 Case 4와 동일한 352 × 352의 해상도를 가진 AFT 이미지 300장으로 구성되지만, 파편의 개수는 59개로 Case 4보다 적은 파편을 가진다. Case 5에서도 정밀도는 Table 2 ~ 3에서 볼 수 있듯이 Original에서 가장 높게 나타났다. 하지만 재현율의 경우 2x Bicubic, 2x EDSR, 2x SwinIR, 4x Bicubic, 4x SwinIR에서 Original보다 높게 나타났다. F1 score의 경우 2x SwinIR이 가장 높은 성능을 보였다.

Table 2에서 2x 이미지의 다섯 케이스 평균을 보면 정밀도는 2x ZSSR만 0.963으로 Original의 0.959보다 약간 높았고, 나머지 2x 이미지들은 Original보다 약간 낮은 성능을 보였다. 하지만 재현율과 F1 score에서는 초해상도화를 수행한 경우 전반적으로 성능이 향상되었다. 특히 2x SwinIR의 경우 재현율이 17 % 상승하였고, F1 score가 9 % 상승하였다.

Table 3에서 4x 이미지의 다섯 케이스 평균을 보면 정밀도는 Original 보다 모두 감소하였다. 재현율은 4x EDSR을 제외하고 소폭 상승하였다. 특히 4x SwinIR은 재현율이 Original에 비해 21 %가 향상되어 가장 좋은 성능을 보였으나 F1 score는 0.7 % 향상되는데 그쳤다.

Table 4. Performance verification result of Swin-T based detection model compared to ResNet-50 based detection model

Average	ResNet-50 based		Swin-T based	
	Original	2x SwinIR	Original	2x SwinIR
Precision	0.959	0.953	0.975	0.951
Recall	0.712	0.832	0.895	0.947
F1 score	0.816	0.889	0.933	0.949

초해상도화에 의해 파편 탐지 모델의 성능이 향상될 수 있음을 확인하였다. 추가로 이 성능 향상이 ResNet-50 기반 Faster R-CNN에 한정되는지 보기 위하여 Swin-T 기반의 Faster R-CNN을 이용해 파편 탐지를 수행하였다. 앞선 결과에서 가장 높은 성능을 보인 2x SwinIR을 이용하여, 다섯 개 시험 케이스의 파편 탐지 결과 평균을 Table 4에 나타냈다. Swin-T 기반의 탐지 모델로 파편 탐지를 수행한 결과 ResNet-50 기반의 탐지 모델보다 탐지 성능이 향상되는 것으로 나타났다. Original 기준 정밀도가 2 %, 재현율이 26 %, F1 score가 14 % 향상되었다. 이처럼 높은 성능을 보인 Swin-T 기반의 탐지 모델에 2x SwinIR 이미지로 파편 탐지를 수행한 결과 더 성능이 높아지는 것을 볼 수 있다. 정밀도는 ResNet-50 기반의 탐지 모델과 큰 차이가 없었지만, 재현율이 0.947로 나타나 0.895의 Original보다 6 % 성능이 향상되는 것으로 나타났다. F1 score도 초해상화 결과 2 % 성능이 향상되는 것으로 나타나, 초해상도화에 의한 성능 향상이 ResNet-50 기반의 Faster R-CNN 모델에 한정되지 않음을 알 수 있다.

5. 결과 고찰

AFT 시험을 통해 얻은 파편 이미지에는 파편이 철관을 관통하는 순간 발생하는 화염, 이 마스크를 벗어난 화염, 그리고 카메라의 노이즈 등이 담기게 된다. Original 이미지에서는 마스크를 벗어난 화염이나 노이즈 등을 나타내는 픽셀 수가 작아 파편으로 인식되지 않는 경우가 많았다. 하지만 이미지의 해상도를 높여줌에 따라 마스크를 벗어난 화염이나 노이즈를 구성하는 픽셀 수가 증가하게 되어 파편으로 탐지되는 케이스가 많아졌다. 이에 따라, 초해상도화 결과 전반적으로 Original 이미지 대비 탐지 모델의 정밀도는 낮아지고, 재현율은 높아지는 결과가 나타났다.

초해상도화로 인해 정밀도가 낮아지는 원인의 예를 Fig. 8과 Table 5에 나타냈다. Fig. 8은 해당 프레임에서 탐지되어야 하는 파편을 나타냈고, Table 5에 모델별 파편 탐지 결과를 나타냈다. 1번 파편은 모든 모델에서 탐지하였고, 2번 파편의 경우 원본 이미지와 4x ZSSR, 4x EDSR을 제외한 이미지에서 파편을 탐지하였다. 정밀도의 차이를 만든 것은 파란색 3번 박스이다. 3번 박스는 이전 탐지된 파편의 화염이 번지는 과

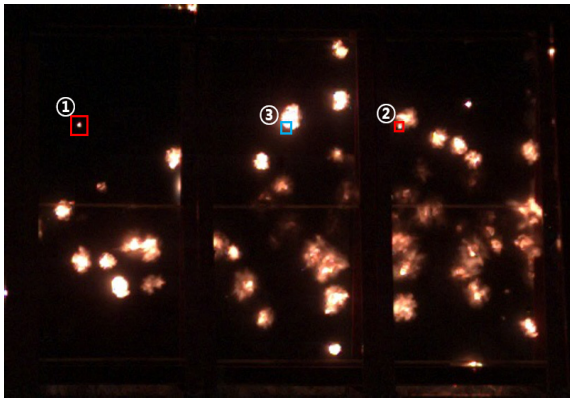


Fig. 8. An example of causes of differences in Precision performance. The Red boxes(①,②) mean real fragments, and the blue box(③) means fake fragment. Results are described in the Table 5. ResNet-50 based Faster R-CNN is used

Table 5. Precision performance results corresponding to Fig. 8.(TP : Detecting the true fragment, FN : Non-detecting the true fragment, FP : Detecting the false fragment)

Fragment #	Original	Bicubic		ZSSR		EDSR		SwinIR	
		2x	4x	2x	4x	2x	4x	2x	4x
1	TP	TP	TP	TP	TP	TP	TP	TP	TP
2	FN	TP	TP	TP	FN	TP	FN	TP	TP
3	-	-	-	-	-	-	-	FP	FP

정에서 SwinIR 모델로 해상도를 높인 이미지만 탐지 모델이 파편으로 인식하였다. Fig. 8에서 볼 수 있는 예처럼, 초해상도화 결과 이미지의 경우 마스킹을 벗어난 작은 화염 또는 노이즈 등을 탐지 모델이 파편으로 탐지하는 빈도가 늘어나는 경향이 확인되었다. 특히 4x SwinIR 이미지에서 이러한 현상이 두드러졌으며, 이로 인해 Table 3에서 볼 수 있듯이, 4x SwinIR의 정밀도는 0.794로 Original 대비 17 % 감소하였다.

초해상도화로 인해 재현율이 높아지는 원인의 예를 Fig. 9와 Table 6에 나타냈다. Fig. 9는 해당 프레임에서 탐지되어야 하는 파편을 나타내며, Table 6은 각 모델이 탐지한 파편 탐지 결과를 나타낸다. 이 결과를 보면, 원본 이미지를 이용해서는 탐지되지 않던 5~7번

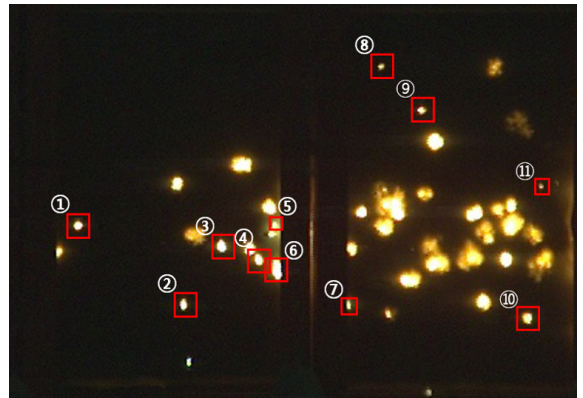


Fig. 9. An example of causes of differences in Recall performance. 11 fragments are showed. Results are described in the Table 6. ResNet-50 based Faster R-CNN is used.

Table 6. Recall performance results corresponding to Fig. 9.(TP : Detecting the true fragment, FN : Non-detecting the true fragment)

Fragment #	Original	Bicubic		ZSSR		EDSR		SwinIR	
		2x	4x	2x	4x	2x	4x	2x	4x
1	TP	TP	TP	TP	TP	TP	TP	TP	TP
2	TP	TP	TP	TP	TP	TP	TP	TP	TP
3	TP	TP	TP	TP	TP	TP	TP	TP	TP
4	TP	TP	TP	TP	TP	TP	TP	TP	TP
5	FN	FN	FN	FN	TP	FN	FN	TP	TP
6	FN	TP	FN	FN	FN	FN	TP	TP	TP
7	FN	TP	TP	TP	TP	TP	TP	TP	TP
8	TP	TP	TP	TP	TP	TP	TP	TP	TP
9	TP	TP	TP	TP	TP	TP	TP	TP	TP
10	TP	TP	TP	TP	TP	TP	TP	TP	TP
11	TP	TP	TP	TP	TP	TP	TP	TP	TP

파편이 초해상도화 이미지를 이용한 경우 탐지되는 것을 볼 수 있다. 특히, SwinIR 모델로 초해상도화를 수행한 경우 모든 파편을 탐지해 낸 것을 볼 수 있다. 따라서 해상도를 높인 경우 파편을 나타내는 픽셀 수가 증가하여 탐지 모델의 파편 인식률 높아지는 것을 볼 수 있다.

초해상도화에 의해 정밀도와 재현율이 변하는 원인

의 예를 보였다. 여기에 더해 Table 5 ~ 6을 보면 2x 이미지를 이용해 파편 탐지를 수행했을 때, 4x 이미지를 이용했을 때보다 더 높은 성능이 나타나는 것을 볼 수 있다. 이는 파편 탐지 모델 훈련 시, 원본 이미지만을 사용했기 때문으로 보인다. 탐지모델이 4x 이미지처럼 원본 이미지보다 해상도가 너무 커진 이미지에 대한 훈련이 돼 있지 않아 오히려 성능이 감소한 것으로 판단된다. 탐지 모델 훈련 시, 초해상도화된 이미지도 훈련에 사용한다면 이러한 성능 저하 현상이 완화될 수 있을 것으로 판단된다.

6. 결 론

본 논문에서는 딥러닝 기반 파편 탐지 성능을 향상시키기 위하여 초해상도화 기법을 적용하였다. 초해상도화 기법으로 Bicubic 보간법과 딥러닝 기반 ZSSR, EDSR, SwinIR 모델을 사용하였다. 초해상도화 수행 결과 Table 2 ~ 3에서 볼 수 있듯이 파편 탐지 모델의 성능이 향상되는 것을 볼 수 있다. 특히 SwinIR 모델로 이미지의 해상도를 두 배 증가시켰을 때 가장 높은 성능개선 효과가 나타났다. SwinIR 모델로 이미지를 두 배 증가시킨 경우 평균 정밀도는 원래 이미지에서보다 0.7 % 감소하였으나, 평균 재현율과 평균 F1 score는 각각 16 %와 8 % 향상되어 탐지 모델 성능이 크게 향상되었다.

4x 이미지를 파편 탐지에 사용할 경우 2x 이미지를 파편 탐지에 사용했을 때보다는 낮은 성능을 보였다. 향후 시험원에서 파편 탐지 모델 성능 개선 연구 시, 초해상도화된 이미지를 훈련에 포함하여 4x 이미지를 파편 탐지에 사용했을 때 성능이 저하되는 문제가 개선되는지 분석할 계획이다.

결과적으로 초해상도화 기법을 활용하여 AFT 이미지의 해상도를 높여줄 경우, 파편 탐지 모델의 성능이 향상될 수 있음을 확인하였다.

후 기

이 논문은 2023년 정부(방위사업청)의 재원으로 국방과학연구소에서 수행된 연구임(딥러닝 기술을 적용한 영상처리 기반 탄두 파편데이터 계측 및 정밀분석 기술).

References

- [1] Zecevic B., Terzic J., Catovic A. and Serdarevic-Kadic S., "Influencing Parameters on HE Projectiles with Natural Fragmentation. In International Conference on New Trends in Research of Energetic Materials," International Conference on New Trends in Research of Energetic Materials, pp. 780-795, April, 2006.
- [2] H. Lee, J. Kim, C. Jung, Y. Park, W. Park and J. Son, "A Deep Learning-based Fragment Detection Approach for the Arena Fragmentation Test," Applied Sciences, Vol. 10, No. 14, pp. 4744, 2020.
- [3] H. Lee, C. Jung, Y. Park, W. Park and J. Son, "A New Image Processing-Based Fragment Detection Approach for Arena Fragmentation Test," Journal of the Korea Institute of Military Science and Technology, Vol. 22, No. 5, pp. 599-606, 2019.
- [4] Shocher, Assaf, Nadav Cohen, and Michal Irani, "'Zero-Shot' Super-Resolution Using Deep Internal Learning," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [5] B. Lim, S. Son, H. Kim, S. Nah and K. Mu Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136-144, 2017.
- [6] Liang J., Cao J., Sun G., Zhang K., Van Gool L. and Timofte R, "Swinir: Image Restoration Using Swin Transformer," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833-1844, 2021.
- [7] Dong C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," European Conference on Computer Vision(ECCV), pp. 184-199, 2014.
- [8] He, K., Zhang, X., Ren, S. and Sun, J., "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
- [9] Wang X., Xie K., Dong C. and Shan Y., "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data Supplementary Material," Proceedings of the IEEE/CVF International

- Conference on Computer Vision, pp. 1905-1914, 2021.
- [10] Chen K., et al., “MMDetection: Open Mmlab Detection Toolbox and Benchmark,” arXiv preprint arXiv:1960.07155, 2019.
- [11] Ren S., He K., Girshick R. and Sun J., “Faster r-cnn: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1147, 2017.