Journal of the KIMST, Vol. 25, No. 5, pp. 476-486, 2022
ISSN 1598-9127(Print) · 2636-0640(Online)
DOI https://doi.org/10.9766/KIMST.2022.25.5.476

Research Paper 센서·신호처리 부문

# SSIM 목적 함수와 CycleGAN을 이용한 적외선 이미지 데이터셋 생성 기법 연구

이하늘[1] · 이현재[*,1]

[1] 조선대학교 항공우주공학과

# Synthetic Infra-Red Image Dataset Generation by CycleGAN based on SSIM Loss Function

Sky Lee[1] · Henzeh Leeghim[*,1]

[1] *Department of Aerospace Engineering, Chosun University, Korea*

(Received 5 July 2022 / Revised 20 September 2022 / Accepted 30 September 2022)

## Abstract

Synthetic dynamic infrared image generation from the given virtual environment is being the primary goal to simulate the output of the infra-red(IR) camera installed on a vehicle to evaluate the control algorithm for various search & reconnaissance missions. Due to the difficulty to obtain actual IR data in complex environments, Artificial intelligence(AI) has been used recently in the field of image data generation. In this paper, CycleGAN technique is applied to obtain a more realistic synthetic IR image. We added the Structural Similarity Index Measure(SSIM) loss function to the L1 loss function to generate a more realistic synthetic IR image when the CycleGAN image is generated. From the simulation, it is applicable to the guided-missile flight simulation tests by using the synthetic infrared image generated by the proposed technique.

Key Words : Generative Adversarial Networks(대립훈련신경망), Synthetic Image(합성이미지), Dataset(데이터셋), Structural Similarity Index Measure(구조적 유사도 지수)

## 1. 서 론

정밀 영상 탐색기 모델 포함 유도 조종 기법 시뮬레이션이나 모의 비행 시험 수행 시, 탐색기 인식과 추적 및 다양한 정찰 임무를 위한 알고리즘 동작을 평가하기 위하여 주어진 가상 환경에서의 동적인 합성(synthetic) 센서 영상이 요구된다. 이러한 가상 영상을 기반으로 하는 유도 조종 시스템은 상대적으로 적은 비용으로 성능을 검증 가능하기 때문에, 이에 대한 다양한 연구들이 지속되고 있다. 그럼에도 불구하고 현재 다양한 모델을 이용해 생성되는 합성 센서 영상은 실제 영상과는 차이가 분명히 존재하고,[1,2] 실제 복잡한 환경에서의 적외선 이미지 데이터를 획득하기

* Corresponding author, E-mail: h.leeghim@controla.re.kr
Copyright ⓒ The Korea Institute of Military Science and Technology

위한 시스템을 구축하는 데에는 많은 비용이 발생한다. 특히 다양한 시간대와 원하는 환경 조건에서 적외선 훈련 데이터셋을 구축하기 어렵기 때문에 최근 인공지능(AI)을 이용한 연구가 활발히 진행되고 있다.

이미지 생성 분야에서 2014년에 발표된 Generative Adversarial Network (GAN)[3] 기술을 활용하는 연구가 제시되었다. GAN이란 학습 데이터를 바탕으로 생성자(G)와 판별자(D)가 적대적으로 학습하며 유사 데이터를 산출하는 알고리즘으로, 훈련 데이터가 충분하지 않은 상황에서도 이미지 생성이 가능하다. 초반에 개발된 GAN 모델은 성능이 좋지 않았으나, 다양한 응용 기법들이 등장함으로써보다 사실적인 이미지 생성의 길이 열리게 되었다.

이 중 Pix2pix[4] 모델은 기존 노이즈 벡터였던 입력 이미지를 스케치 이미지로 대체해 원본과 비슷한 특성을 가지는 새 이미지를 생성해 내는 알고리즘이다. 하지만 실제 환경에서 짝지어진 이미지 데이터셋은 구축이 어렵기 때문에, 짝지어지지 않은 이미지 간의 유사 변형을 수행하는 CycleGAN[5] 모델이 연구되었다.

본 연구에서는 짝지어진 데이터셋 구축이 어려운 상황에서 보다 사실적인 합성 적외선 이미지를 생성하기 위해 CycleGAN 모델을 이용한다. 하지만 적외선 이미지는 색상 맵 분포가 한정적이기 때문에 모델이 해당 특성을 잘 잡아내지 못해 학습이 잘 이루어지지 않는 경향이 있다. 이러한 점을 해결하기 위해 기존 손실 함수를 수정해 데이터셋 분포가 한정적이라도 이미지가 손실되지 않고 기존 손실 함수에 비해 품질이 향상될 수 있도록 하였다. 최종적으로는 생성된 이미지의 평가 지표로 활용할 수 있는 기법들을 제안하여 이미지 분석에 이용될 수 있게 한다.

논문의 구성은 다음과 같다. 먼저 2장에서는 앞서 연구된 GAN과 평가 지표 관련 기술들에 대해 설명한다. 3장에서는 적외선 이미지를 생성 및 개선하기 위한 알고리즘에 대해 제시하고, 4장에서는 제시한 알고리즘 및 손실 함수의 성능을 검증하기 위하여 기존 손실 함수 사용 시와 이미지 결과를 비교하였다. 더해서 이미지 평가 지표로 활용될 수 있는 기법들에 대해 제안하고, 마지막 5장에서 결론 및 향후 연구에 대해 고찰한다.

## 2. 관련 연구

### 2.1 생성 모델

본 연구는 GAN을 이용하여 적외선 실감 영상을 만들어 내는 것을 목표로 한다. 일반적인 인공지능 모델은 어떠한 클래스에 대한 예측값이나 분류 등의 결과를 생성하지만, GAN은 데이터의 형태를 생성해 내는 생성 모델이다. 생성자의 목표는 판별자가 구분하지 못할 정도로 실제 데이터와 유사한 데이터를 만드는 것이고, 판별자의 목표는 진짜와 가짜 데이터를 잘 구분해 내는 것이 된다.

GAN은 크게 조건부 GAN(cGAN)과 비조건부 GAN (unconditional GAN)으로 나눌 수 있다. 비조건부 GAN에서는 점차 해상도를 높여 나가는 학습법을 사용하여 GAN에서 높은 해상도를 효율적으로 학습시키는 PGGAN[6]과, 이미지의 스타일을 부분적으로 제어할 수 있는 StyleGAN[7] 등 style transfer 관련 연구가 진행되었다. 조건부 GAN에서는 이미지 대 이미지 변환(image-to-image translation) 연구들이 진행되었다. Pix2pix는 일종의 지도학습(Supervised Learning)으로, 신경망은 라벨 이미지와 유사하게끔 입력을 바탕으로 이미지를 생성한다. 이미지 생성 품질을 높이기 위해 생성자의 모델 구조로 U-Net[8]을 사용하는데, U-Net은 기존 Encoder-Decoder 구조와 달리 Upsampling 시 Encoder의 특성 맵을 한 번 더 반영해 정확한 localization을 하는 역할을 한다. 또한 기존 cGAN의 손실 함수에 L1 손실을 더해 이미지 품질을 향상시킨다. CycleGAN은 기존 Pix2pix가 짝지어진 이미지로 학습을 수행해야 했던 점에서 벗어나, 짝지어지지 않은 이미지를 입력으로 한다. 이는 일종의 비지도학(Unsupervise Learning)으로 생성자는 정답을 보고 학습하는 게 아닌 상대 도메인의 이미지 특성을 학습하게 된다. 이렇게 되면 학습시키려는 모델이 실제 데이터 분포를 커버하지 못하고 다양성을 잃어버리는 Mode collapse 문제가 나타날 수 있다. CycleGAN은 이를 해결하기 위해 순환 일관성 손실(cycle consistency loss)을 도입하여 도메인 유지가 안정적으로 이루어질 수 있게 하였다. 또한 U-Net에 ResNet[9]을 추가한 구조를 사용하는데, 이는 층이 깊어질수록 기울기 소실이나 증폭 발생할 위험이 높은 기존 합성곱(Convolution) 구조에 전차(Residual) 개념을 적용하여 기울기 문제로부터 자유로울 수 있게 한다. 하지만 CycleGAN은 이미지의 전반적인 형태는 바꿀 수 없고, 데이터셋의 분포가 불안정하면 이미지를 제

대로 생성하지 못한다는 한계가 있다. 많은 연구자들이 CycleGAN의 성능을 향상시키고 이러한 한계를 극복하기 위해 네트워크 구조나 손실 함수를 변경하는 시도를 하였고, 실제로 좋은 결과를 도출하였다[10-12]. CycleGAN의 전체 구성도를 도식화한 것은 Fig. 1과 같다.

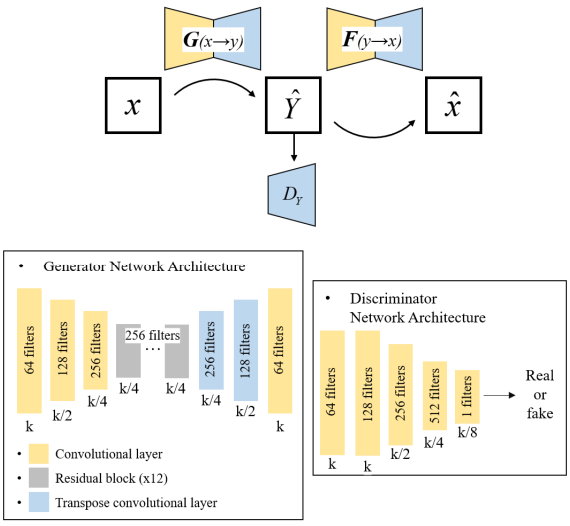


Fig. 1. CyclcGAN network architecture

### 2.2 손실 함수

현재 GAN에서 이미지 생성을 위해 사용하는 손실 함수의 대부분은 MSE를 최소화하는 방향, 즉 실제 값과 예측 값의 픽셀 값 차이에 의존한다. 가장 널리 사용되는 것은 L1 손실과 L2 손실로, 이들은 픽셀 단위로 오차가 역전파되기 때문에 네트워크가 수렴할 때까지 시간이 오래 걸리고, 이미지가 지나치게 평균화되기 때문에 고해상도 텍스처의 디테일을 잡아내지 못해 이미지 품질이 떨어진다는 한계가 있다.

Structural Similarity Index Measure(SSIM)[13]은 두 이미지 사이의 유사도를 측정하는 지표로 주로 사용된다. 이는 수치적 오차가 아닌 인간의 시각적인 차이를 평가하기 위해 고안된 방법으로, 사람의 시각이 이미지의 구조 정보를 도출하는 데에 특화되어 있기에 구조 정보의 왜곡이 지각에 큰 영향을 미친다는 것으로부터 파생된 이론이다. 이는 밝기(Luminance), 대비(Contrast), 구조(Structure)로 이루어진 3가지 요소를 곱하여 값을 산출하며, 픽셀의 평균($\mu$), 표준편차($\sigma$), 공분산($\sigma_{xy}$)을 계산해 밝기, 대비, 구조적 유사도에 반영하여 두 이미지 간의 유사도를 0-1 사이에서 산출한다. 각각의 수식은 다음과 같다.

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{1}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{2}$$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{3}$$

상수($C_1$, $C_2$, $C_3$)들은 분모가 0이 될 때의 안정성을 보장하기 위한 것이며 $C_1=(k_1L)^2$, $C_2=(k_2L)^2$으로 계산되고, 일반적으로 $K_1$ = 0.01, $K_2$ = 0.03, $L$ = 255를 사용한다[13]. 또한 각각의 평균($\mu$), 표준편차($\sigma$), 공분산($\sigma_{xy}$)은 다음과 같이 계산된다.

$$\mu_x=\frac{1}{N}\sum_{i=1}^{N}x_i \tag{4}$$

$$\sigma_x=\left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2\right)^{\frac{1}{2}} \tag{5}$$

$$\sigma_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y) \tag{6}$$

이를 일반화하기 위해 (1)-(3)에서 $C_3=\frac{C_2}{2}$라고 가정한 뒤 정리하면 다음과 같은 식을 얻을 수 있다.

$$SSIM(x,y)=l(x,y)c(x,y)s(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{7}$$

두 이미지를 비교하였을 때 SSIM 값이 1에 가까울수록 유사한 이미지가 된다. 많은 연구자들이 SSIM을 생성 모델의 손실 함수에 적용하려는 시도를 하였으며, 기존 모델의 학습 결과보다 더 나은 성능을 보였다[14-16]. 본 연구에서도 데이터셋 분포가 한정적인 경우에 학습 성능이 떨어지는 CycleGAN의 성능을 개선하기 위해 손실 함수에 SSIM 손실을 추가하여 학습한다.

### 2.3 이미지 품질 평가 지표

#### 2.3.1 T-SNE

SNE(Stochastic Neighbor Embeding)[17]는 비선형적인 방법을 사용하는 차원 축소 방법의 일종으로, 고차원 공간에서 점들의 유사성이 저차원에서도 보존될 수 있게끔 하는 기법이다. 이는 저차원에서 고차원 데이터 분포를 완벽하게 표현하기 위해 유사점 간 차이를 최소화하는 방향으로 학습하며, 고차원에서 비슷한 데이터 구조는 저차원에서도 가깝게 대응하게 된다. 이는 유클리드 거리를 유사성을 나타내는 조건부 확률로 변환하여 유사도를 계산한다. x가 고차원의 데이터이고 y가 저차원의 데이터라고 하면 각 조건부 확률은 다음과 같다.

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k \neq i}\exp(-\|x_i - x_k\|^2/2\sigma_i^2)} \tag{8}$$

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i}\exp(-\|y_i - y_k\|^2)} \tag{9}$$

SNE는 p와 q의 분포 차이를 줄이기 위하여 학습하게 되며, 이때 쿨백-라이블러 발산(Kullback-Leibler Divergence)을 통해 두 분포 사이의 값을 최소화하게 된다. 또한 SNE가 데이터 임베딩 시 고차원에서의 먼 거리를 저차원에서도 충분히 멀게 표현하지 못하는 과밀문제(crowding problem)를 해결하기 위하여 점 i와 j 사이의 거리를 대칭적으로 사용하였고,

$$p_{j|i} = \frac{p_{j|i} + p_{i|j}}{2}, q_{j|i} = \frac{q_{j|i} + q_{i|j}}{2} \tag{10}$$

정규분포보다 분산이 큰 T-분포를 사용하였다. 최종적으로 t-SNE[18]의 조건부 확률은 다음과 같이 표현된다.

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k \neq i}\exp(-\|x_i - x_k\|^2/2\sigma_i^2)} \tag{11}$$

$$q_{j|i} = \frac{(1+\|y_i - y_j\|^{-1})}{\sum_{k \neq i}(1+\|y_k - y_l\|^2)^{-1}} \tag{12}$$

#### 2.3.2 파워 스펙트럼 밀도

파워 스펙트럼(power spcetrum)[19,20]은 푸리에 변환(Fourier transform)을 기반으로 이미지가 가지는 주파수를 빈도에 따라 시각화한 것이다. 이는 하나의 전체 시계열을 짧은 구간으로 나눈 뒤 FFT(Fast Fourier Transform) 계산으로 나온 값의 평균값을 로그 스케일로 나타내는 것이 일반적이다. 이미지에 대한 푸리에 변환(F)은 다음과 같이 주어진다.

$$F(u,v) = \sum_{(x,y)} \frac{l(x,y) - \mu}{\mu} w(x,y) e^{2\pi i (ux/width/height)} \tag{13}$$

이때 u와 v는 공간 주파수의 좌표에 해당한다. 여기에서 $\mu$는 가중 평균 강도(weighted mean intensity)로, 다음과 같이 정의된다.

$$\mu = \frac{\sum_{(x,y)} l(x,y) w(x,y)}{\sum_{(x,y)} w(x,y)} \tag{14}$$

촬영한 임의의 자연 이미지를 분석한 스펙트럼 결과는 Fig. 1과 같다. 스펙트럼 그래프가 선형성을 보이는 것을 확인하였고, 이를 통해 알고리즘으로 생성한 가짜 이미지와 실제 이미지의 FFT를 비교하여 실제와 유사하게 잘 생성되었는지 평가 기준으로 활용할 수 있다.

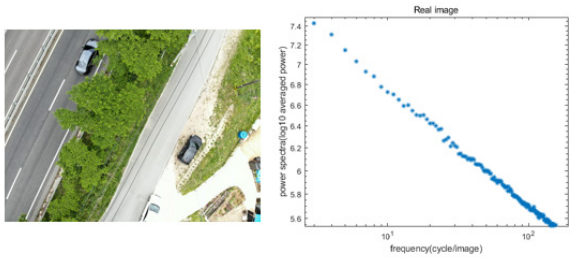


Fig. 2. FFT analysis of Natural Images, if the graph is linear, it can be judged to be similar to natural(real) images

## 3. 제안된 알고리즘

제안하는 알고리즘은 A 도메인을 가상 환경, B 도메인을 적외선 환경으로 설정했을 때, CycleGAN을 통

해 가상 환경과 적외선 환경으로부터의 가상 이미지 데이터셋을 생성하되 기존 CycleGAN의 손실 함수에 SSIM 손실을 추가하여 학습한다. 이때 SSIM의 이미지 비교 영역을 다양하게 설정하여 영역 크기에 따른 이미지 품질을 비교한다. L1 손실이 픽셀 하나하나를 비교하는 것과 다르게 SSIM은 영역 단위로 비교하기 때문에 이를 손실 함수에 적용하면 생성된 이미지의 개선이 있을 것이라 판단하였다.

### 3.1 적외선 이미지 생성

CycleGAN은 두 도메인 간을 오갈 수 있는 두 개의 생성자(G, F)와 판별자($D_x$ $D_y$)를 학습시킨다. 생성자는 짝지어지지 않은 이미지의 한 도메인을 입력으로 받아서 다른 도메인의 분포로 매핑하는 작업을 하며, 이때 다른 도메인에서 원래 도메인으로 돌아왔을 때의 차이를 적게 하기 위해 순환 일관성 손실 함수를 기존 GAN의 적대적 손실에 추가한다. 이렇게 되면 생성자는 도메인이 변경될 때 이미지가 생성되었다가 재구성되는 매핑까지 고려하며 학습하게 된다. 또한 입력과 출력 이미지의 색감이 과하게 달라지는 것을 방지하기 위해 idenitiy 손실 함수가 도입되었다. 이는 타겟 도메인이 입력으로 들어왔을 때, 타겟과 동일한 도메인으로 매핑하는 경우의 차이가 적도록 한다. 각 손실 함수를 식으로 표현하면 다음과 같다.

$$L_{GAN} = E_{x \sim P_{data}(x)}[\log(1 - D_Y(G(x)))] + E_{y \sim P_{data}(y)}[\log D_Y(y)] \quad (15)$$

$$L_{cyc} = E_{x \sim P_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)}[\|G(F(y)) - y\|_1] \quad (16)$$

$$L_{identity} = E_{x \sim P_{data}(x)}[\|F(x) - x\|_1] + E_{y \sim P_{data}(y)}[\|G(y) - y\|_1] \quad (17)$$

총 손실 함수는 적대적 손실에 순환 일관성 손실, identity 손실을 더해서 사용한다. 식은 아래와 같다.

$$L_{CycleGAN} = L_{GAN}(G, F, D_X, D_Y) + \lambda L_{cyc}(G, F) + \alpha L_{identity}(G, F) \quad (18)$$

생성자는 인코더(Encoder), 변환기(Transformer), 디코더(Decoder)로 구성되며 인코더와 디코더는 U-Net 구조를 따른다. 인코더는 세 개의 합성곱 연산을 사용하고, 변환기는 ResNet 9-block을 기반으로 한다. 판별자는 [5]에서 제안된 70×70 Patch GAN을 사용하였다. 이는 이미지 전체 대신 특정 크기의 Patch로 지역적인 비교를 수행한다.

### 3.2 이미지 품질 향상을 위한 SSIM 손실 적용

색상 맵이 단일 도메인으로 구성된 해당 데이터 세트의 경우 L1 손실을 사용하므로 생성자가 가상 이미지에서 적외선 이미지를 생성할 때 원본의 구체적 특성을 복구하지 못하고 흐려 버리거나, 비슷한 패턴을 반복하여 사용하는 문제점이 나타날 수 있다. 원본 이미지의 도메인만 변경하며 구체적인 특성들은 최대한 보존되는 것이 중요하므로 기존 CycleGAN의 손실 함수에 SSIM 손실을 더해 함께 사용하였다. SSIM은 순환 일관성 손실과 동일하게 생성자 부분에서 입력 이미지와 재구성된 이미지의 차이를 줄이도록 학습하게 된다. 따라서 재구성된 CycleGAN의 손실 함수는 다음과 같다.

$$L_{CycleGAN} = L_{GAN}(G, F, D_X, D_Y) + \lambda L_{cyc}(G, F) + \alpha L_{identity}(G, F) + \beta L'_{SSIM}(G, F) \quad (19)$$

이때 SSIM 값은 1에 가까울수록 이미지의 유사도가 높다는 의미가 되는데, 생성 모델의 손실 함수는 값이 작을수록 모델의 성능이 좋아지므로 L_SSIM을 다음과 같이 정의하여 사용하였다.

$$L'_{SSIM}(G, F) = (1 - L_{SSIM}(G, F)) \quad (20)$$

또한 L1 손실 대비 SSIM 손실의 효과를 확인하기 위한 학습을 진행하였다. L1 손실은 순환 일관성 손실에 $\lambda = 10$[5] 가중치를 곱해 적용되며, 이를 기준으로 $\lambda$와 $\beta$의 가중치를 변화시키면서 이미지 생성 결과를 확인하였다. SSIM은 영역 단위 비교를 수행할 때 해당 영역의 정의가 필요한데, 이를 윈도우(window)라 칭한다. 이는 표준편차가 1.5인[21,22] 가우시안 분포를 이용하여 생성된다. 생성된 윈도우는 평균($\mu$), 표준편차($\sigma$), 공분산($\sigma_{xy}$) 계산 시 이용되는 합성곱 계층의 스트라이드(stride)와 패딩(padding)에 사용되며, 픽셀 대 픽셀이 아닌 윈도우 크기를 기준으로 이미지를 비교한다. 본 연구에서는 윈도우 크기에 따라 이미지 품질에 차이가 있을 것이라고 판단하여 멀티 윈도우 개념을 적용하였다.

$$L'_{SSIM} = \sum_i w_i L_{SSIM_p} \quad (21)$$

$L'_{SSIM}$은 pxp 윈도우와 해당 윈도우에 가중치 $w_i$를 곱해서 계산되며, 본 연구에서는 22×22와 11×11 윈도우를 이용하여 학습을 진행하였다. 손실 함수에 사용된 각 가중치의 값은 $\lambda$ = 10, $\alpha$ = 0.1으로 설정하였다[5].

## 4. 시뮬레이션 결과

### 4.1 학습 데이터셋 구성

GAN 학습을 위한 데이터 세트는 두 개의 서로 다른 두 도메인을 요구한다. 우리는 합성 이미지의 도메인을 적외선 이미지로 변경하는 것이 목표이기 때문에 다음과 같이 두 도메인을 설정하였다. 먼저 IR 이미지 도메인을 위해 적외선 카메라가 장착된 드론 시스템을 사용하여 항공 적외선 이미지를 수집한다. 이미지 수집은 드론에 부착된 열화상 카메라를 사용하였다. 드론 및 카메라의 스펙은 Table 1, 2에 나타내었다. 가상 이미지 도메인을 위해서는 이미지 매핑 및 모델링 소프트웨어인 RealityCapure 소프트웨어를 사용하여 3D 모델링을 수행하였다. 이때 텍스처링을 거치지 않은 메쉬 상태의 모델로부터 이미지를 추출하여 사용하였고, 이를 메쉬 이미지라고 칭한다. 해당 이미지 데이터는 특성이 흐려져 학습이 잘 이루어지지 않을 경우에 대비하여 대비(Contrast)를 증가시켜 사용하였다. 또한 촬영된 RGB IR 이미지로 학습을 진행할 경우 모델이 색상 특성을 잘 포착하지 못해 학습이 잘 이루어지지 않는 문제가 발생하였다. 따라서 RGB 이미지를 모두 회색조(Grayscale)로 변환하여 사용하였다. 학습에 사용될 데이터는 모두 256×256으로 변환하고, 데이터 증강(Data Argumentation)을 위해 회전시킨 이미지를 데이터셋에 포함시켰다. 따라서 90도, 180도, 270도 회전시킨 이미지를 학습에 함께 사용하였다. 알고리즘에 사용한 데이터셋 및 가공 과정은 Fig. 3에 나타내었다.

이렇게 가상 모델로부터 추출해 데이터 증강을 거친 이미지 1,840장과 촬영으로 얻은 적외선 이미지 3,800장으로 총 5,640장을 사용하였고, 960장은 테스트 데이터로 사용하였다.

Table 1. UAV specification

| Attribute | Value |
|---|---|
| Description | DJI MATRICE 300 RTK |
| Weight | 6.3 kg<br>(Including two TB60 batteries) |
| Diagonal length | 895 mm |

Table 2. IR Camera specification

| Attribute | Camera Specification |
|---|---|
| Description | DJI ZENMUSE H20T |
| Sensor | Vanadium Oxide (VOx)<br>microwave bolometer<br>DFOV: 40.6° |
| Lens | Focal length: 13.5 mm<br>Aperture: f/1.0<br>Focus: 5 m ~ ∞ |

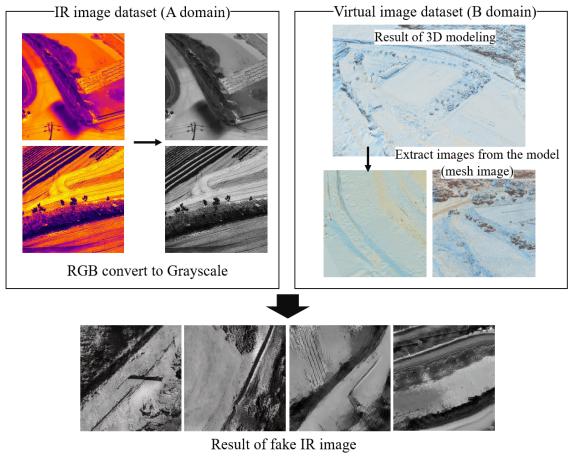


Fig. 3. Dataset and processing process used in the algorithm

### 4.2 시뮬레이션 결과 및 FID 분석

mesh 이미지와 IR 이미지로 CycleGAN을 학습시킨 결과는 Fig. 4과 같다. 이때 L1 손실과 SSIM 손실의 가중치를 다르게 하며 학습을 진행하였다. 각 이미지 세트에 대한 SSIM 가중치 파라미터 값은 Table 3과

같다. L1 손실의 가중치는 [5]에서 제안된 10을 고정해 두고 SSIM의 가중치를 다르게 하며 학습한 결과, CaseB와 CaseC가 이미지 변환을 가장 잘 수행하였다. 이때 각 데이터셋의 FID(Frechet Inception Distance)[23], [24] 점수는 Table 4와 같다. FID 값이 낮을수록 두 이미지 세트가 유사하다는 지표로, 점수가 낮을수록 성능이 좋은 모델로 평가한다. 해당 데이터셋의 FID 값을 계산한 결과 점수가 가장 낮은 CaseB의 가중치 파라미터로 윈도우 크기의 가중치 파라미터에 따른 학습 분석을 진행하였다.

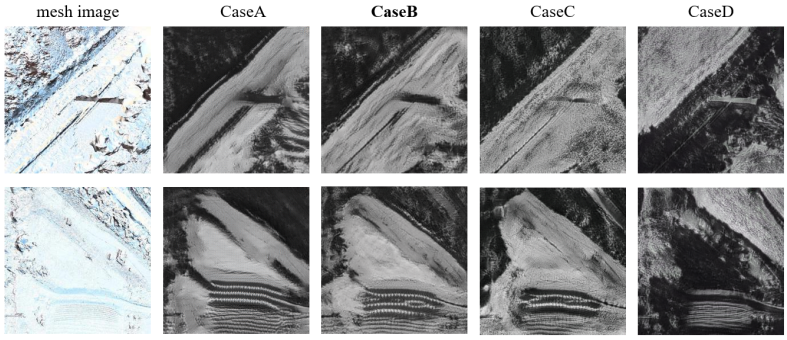


Fig. 4. Constructed image comparison due to the L1 loss between SSIM loss

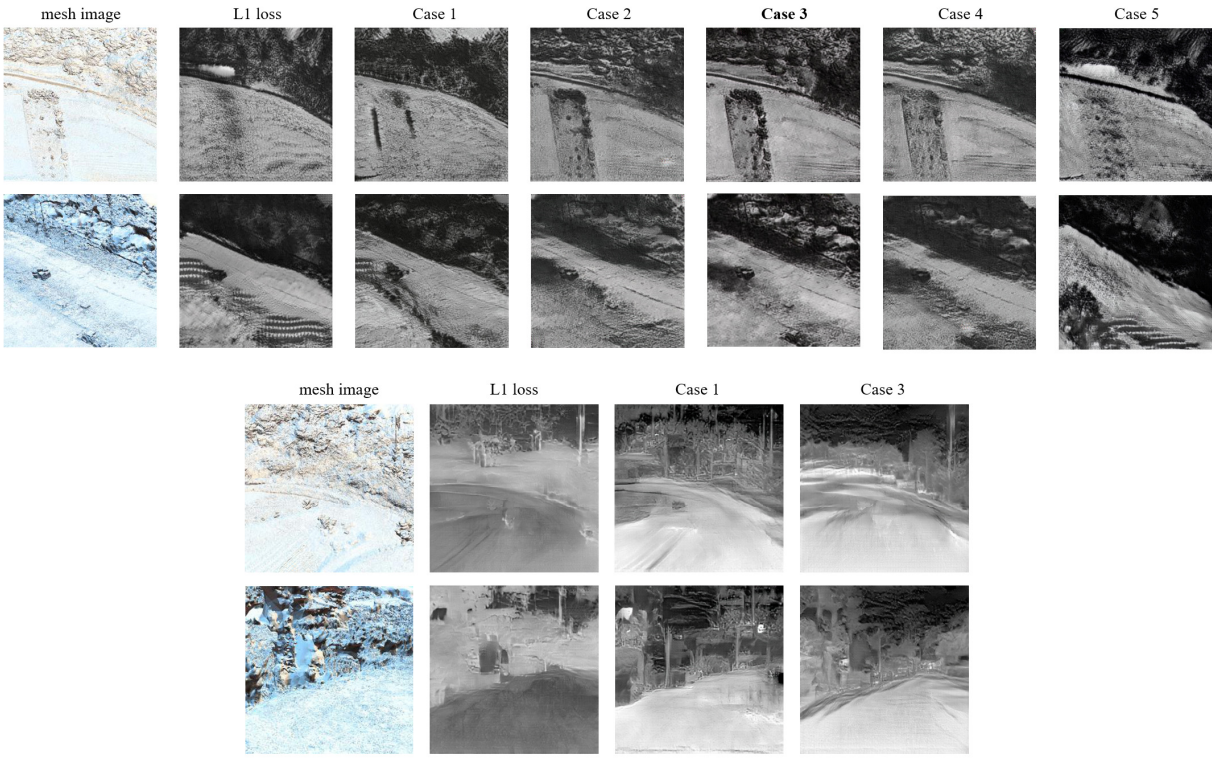


Fig. 5. Constructed image comparison due to the multi-window weighting parameters, top is custom infrared data, bottom is FLIR data

Table 3. Weight parameters in different cases of L1 loss between SSIM loss

| weight parameter | CaseA | CaseB | CaseC | CaseD |
|---|---|---|---|---|
| L1 loss | 10 | 10 | 10 | 10 |
| SSIM loss | 0.1 | 1 | 10 | 100 |

Table 4. FID socre in different cases of L1 loss between SSIM loss

| weight parameter | FID scroe |
|---|---|
| CaseA | 242.863 |
| CaseB | **201.806** |
| CaseC | 202.380 |
| CaseD | 221.065 |

Fig. 4는 좌측의 가상 이미지로부터 적외선 도메인으로 변환된 이미지를 손실 함수에 따라 나타낸 결과이다. L1 손실과 SSIM 손실의 가중치는 Table 5에서 제시한 CaseB를 따르되, SSIM 손실의 윈도우 크기에 따른 가중치를 다시 부여하였다. 본 논문에서 제안한 SSIM 손실의 성능을 검증하기 위하여 Case1과 기존 CycleGAN의 L1 손실과 비교하였을 때, L1 손실만 사용하는 기존 CycleGAN은 이미지의 구체적인 특성을 살리지 못하고 이미지를 평균화시키거나 특정 물체를 훼손시키는 경향을 보였다. 하지만 SSIM 손실이 추가된 Case1 이미지에서는 해당 특성들이 상대적으로 복구되는 것을 확인하였다. SSIM 손실의 윈도우 크기를 가중치 파라미터에 따라 학습을 진행한 결과는 Case1부터 Case5에 해당한다. 이때 각 이미지의 SSIM 윈도우 가중치 파라미터 값은 Table 5와 같다. 기본 값인 11에 22 윈도우를 추가한 뒤 해당 윈도우의 가중치를 달리하여 비교 분석을 진행한 결과 기존 11 윈도우 단독으로 학습한 Case1보다는 윈도우를 혼합하여 사용한 Case2, 3, 4가 이미지 생성 측면에서 향상된 결과를 보였다. 두 윈도우 모두 가중치를 0.5로 놓고 학습한 Case3의 경우에는 원본 이미지의 특성을 거의 흐리지 않고 도메인만 변경하는 정도의 이미지 변환을 수행함을 확인하였다. 또한 제안하는 SSIM 손실 및 베스트 케이스인 Case3의 성능을 검증하기 위하여

Table 5. Weight parameters in different cases of window size

| weight parameter | Case1 | Case2 | Case3 | Case4 | Case5 |
|---|---|---|---|---|---|
| $w_1$(11×11) | 1.0 | 0.6 | 0.5 | 0.4 | 0.0 |
| $w_2$(22×22) | 0.0 | 0.4 | 0.5 | 0.6 | 1.0 |

Table 6. FID socre in different cases of window size

| loss function \ dataset | Custom dataset | FLIR dataset |
|---|---|---|
| L1 loss | 208.748 | 232.017 |
| Case1(CaseB) | 201.806 | **228.008** |
| Case3(best) | **198.506** | 232.487 |

public 데이터셋인 FLIR[25]에 대해서도 학습을 진행하였다. 이미지를 적외선 도메인으로 변경하는 과정에서 윈도우에 따른 가중치 파라미터가 있는 SSIM 손실이 보다 향상된 성능을 보이는 것을 확인하였다.

Table 6은 해당 데이터 세트와 실제 적외선 이미지 데이터셋에 대한 FID 값이다. SSIM 손실을 추가한 데이터셋과 육안으로 가장 잘 생성되었다고 판단된 Case3, L1 손실만 적용된 기존 CycleGAN의 FID 값을 계산한 결과 SSIM 손실을 추가한 모델이 더 나은 성능을 보이는 것을 확인하였다.

### 4.3 t-SNE를 이용한 이미지 분석

학습에 사용한 실제 적외선 데이터, 가상 모델로부터 추출한 mesh 데이터, SSIM 손실을 적용해 CycleGAN으로 생성한 가짜 적외선 데이터를 이용해 t-SNE로 분석한 결과는 Fig. 6과 같다. t-SNE 알고리즘은 혼잡도(Perplexity)를 이용하여 데이터 간 거리를 정의할 수 있으며, 혼잡도를 조절하여 점들 간 간격을 균등하게 설정할 수 있다. 40의 혼잡도에 따라 이미지들을 시각화한 결과, 직접 수집한 IR 데이터셋과 FLIR 데이터셋 모두 생성한 가짜 이미지가 실제 적외선 이미지와 군집을 이루며 분포하는 것을 확인하였다. t-SNE 기법은 고차원 공간에서의 유사성을 2차원으로 시각화해서 나타내므로 가짜 이미지가 실제 적외선 데이터와 군집화되었다는 것은 데이터 분포가 유사하게 이미지가 생성되었다고 판단할 수 있다.

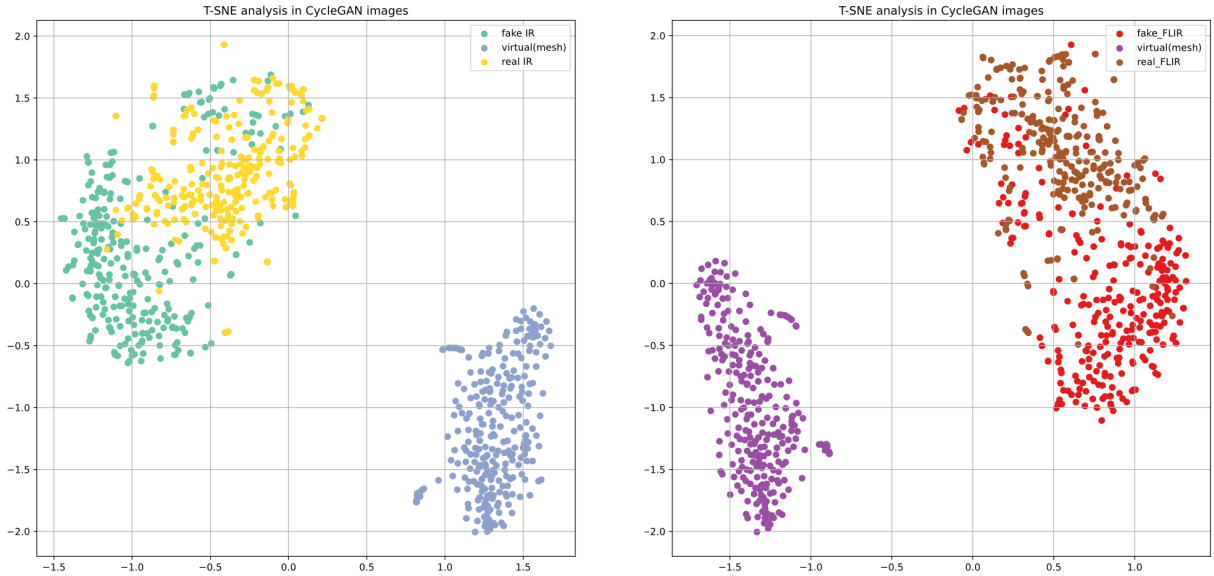


Fig. 6. (a) t-SNE analysis with 40 perplexity plotting of IR images including the images generated by CycleGAN proposed, The left is the custom IR dataset, and the right is the FLIR dataset

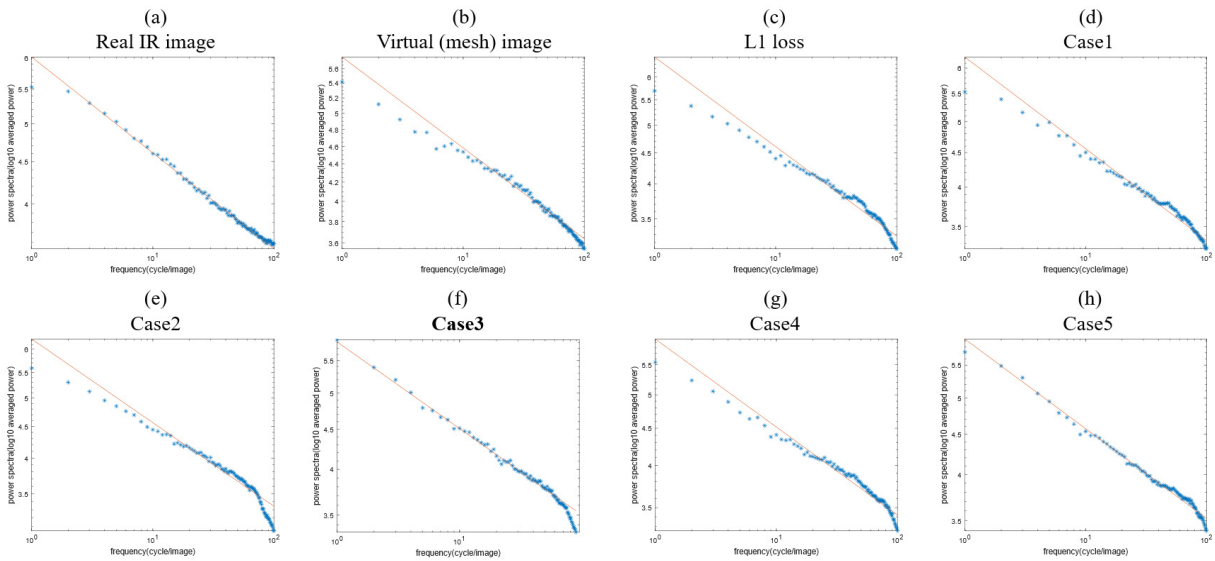


Fig. 7. Power spectrum analysis for the constructed image by CycleGAN proposed and a natural IR image

4.4 Power spectrum density를 이용한 이미지 분석

Fig. 6은 실영상과 생성한 적외선 이미지의 가중치에 따른 파워 스펙트럼 분석 결과이다. (a)는 실제 적외선 이미지의 스펙트럼으로, 주파수 분포가 대체로 선형성을 유지하였다. (b)는 가상 모델로부터 추출한 이미지로, 저주파 영역이 떨어지고 스펙트럼 분포가 곡선을 그리는 것을 확인하였다. 스펙트럼이 (a)를 따르면 실제 이미지, (b)를 따르면 가짜 이미지와 유사하다고 했을 때 L1 손실을 단독 사용한 (c)의 스펙트럼은 (b)에 가까운 그래프를 그리는 것을 확인할 수 있었는데,

SSIM 손실을 함께 사용한 Case들의 주파수 스펙트럼은 상대적으로 저주파 영역을 많이 회복하였으며 (c)보다 선형성이 추가된 것을 확인할 수 있었다. 이를 통해 SSIM 손실을 추가해서 학습한 이미지가 기존 L1 손실을 사용하는 이미지보다 실제 이미지의 스펙트럼 분포를 따르게끔 생성된 것을 확인할 수 있다.

## 5. 결론 및 향후 연구

본 연구에서는 짝지어진 데이터셋이 없는 경우에 가상 도메인으로부터 CycleGAN 알고리즘을 이용하여 적외선 이미지를 생성하는 모델에 대해 제안하였다. 이때 적외선 영상의 특성상 도메인이 단일화되어 학습이 잘 이루어지지 않는 점을 고려하여 기존 L1 손실에 SSIM 손실을 추가함으로써 이미지 생성 측면에서 향상된 성능을 보였다. 또한 생성된 이미지의 평가 지표로 활용될 수 있는 기법들을 제안하여 이미지 평가에 이용될 수 있게 하였다. GAN 관련 기술이 많이 발전했음에도 불구하고 GAN을 이용하여 생성한 이미지들은 실제 이미지와 이질감이 들거나 원본 도메인을 잘 유지하지 못하고 특성을 잃어버리는 등의 한계가 있다. 이때 본 연구에서 제안한 SSIM 손실을 이용해 학습한다면 원본 이미지의 특성을 최대한 유지하며 도메인을 변경할 수 있으므로 다양한 적외선 비행 환경 구축 관련 이미지 생성 연구 분야에 활용될 수 있을 것으로 보인다.

향후 연구에서는 합성 IR 이미지를 생성할 시에 날씨에 대한 인자를 추가하여 대기에 따른 이미지 변화를 생성할 수 있는 연구와 함께, 다양한 도메인 및 데이터셋에 대해 SSIM의 윈도우별로 이미지를 다양하게 분석하여 성능에 대한 검증을 진행할 예정이다. 또한 가짜 이미지의 주파수 스펙트럼을 끌어올리기 위하여 초해상화(Super resolution) 등의 기법을 이용하여 보다 실제 이미지와 유사한 데이터를 생성해 낼 수 있을 것으로 기대된다.

## 후 기

이 연구는 인공지능 비행제어 특화연구실 프로그램의 일환으로 국방과학연구소와 방위사업청의 지원으로 수행되었음.

## References

[1] Zhang, Ruiheng and Mu, Chengpo and Xu, Min and Xu, Lixin and Shi, Qiaolin and Wang, Junbo, “Synthetic IR Image Refinement Using Adversarial Learning with Bidirectional Mappings,” IEEE Access, Vol. 7, pp. 153734-153750, 2019.

[2] Kniaz, VV and Gorbatsevich, VS and Mizginov, VA, “Thermalnet: a Deep Convolutional Network for Synthetic Thermal Image Generation,” The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 42, p. 41, 2017,

[3] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza and Bing Xu, “Generative Adversarial Networks,” Neural Information Processing Systems(NIPS), Motreal, 2014.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” CVPR, 2017.

[5] Jun-Yan Zhu, Taesung Park, and Alexei A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” Proceedings of the IEEE International Conference on Computer Vision, pp. 2223-2232, 2017.

[6] Karras, Tero and Aila, Timo and Laine, Samuli and Lehtinen, Jaakko, “Progressive Growing of Gans for Improved Quality, Stability, and Variation,” arXiv preprint arXiv:1710.10196, 2017.

[7] Karras, Tero and Laine, Samuli and Aila, Timo, “A Style-based Generator Architecture for Generative Adversarial Networks,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401-4410, 2019.

[8] Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234-241, 2015.

[9] Kaiming He, Xiangyu Zhang, and Jian Sun, “Deep Residual Learning for Image Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[10] Li, Wenda and Wang, Jian, “Residual Learning of Cycle-GAN for Seismic Data Denoising,” IEEE Access, Vol. 9, pp. 11585-11597, 2021.

[11] Maniyath, Shima Ramesh and Vijayakumar, K and Singh, Laxman and Sharma, Sudhir Kumar and Olabiyisi, Tunde, “Learningbased Approach to Underwater Image Dehazing Using CycleGAN,” IEEE Access, Vol. 14, No. 18, pp. 1-11, 2021.

[12] Teng, Long and Fu, Zhongliang and Yao, Yu, “Interactive Translation in Echocardiography Training System with Enhanced Cycle-GAN," IEEE Access, Vol. 8, pp. 106147-106156, 2020.

[13] Hang Zhao, Orazio Gallo, and Jan Kautz, “Loss Functions for Image Restoration With Neural Networks,” IEEE Transactions on Computational Imaging, pp. 47-57, 2017.

[14] Hwang, Jieon and Yu, Chushi and Shin, Yoan, “SAR-to-Optical Image Translation Using SSIM and Perceptual Loss based Cycleconsistent GAN,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 191-194. 2020.

[15] Tao, Li and Zhu, Chuang and Xiang, Guoqing and Li, Yuan and Jia, Huizhu and Xie, Xiaodong, “LLCNN: A Convolutional Neural Network for Low-Light Image Enhancement,” 2017 IEEE Visual Communications and Image Processing(VCIP), pp. 1-4, 2017.

[16] Shi, Haoyue and Wang, Le and Zheng, Nanning and Hua, Gang and Tang, Wei, “Loss Functions for Pose Guided Person Image Generation,” Pattern Recognition, Vol. 122, p. 108351, 2022.

[17] G. E. Hinton and S. T. Roweis, “Stochastic Neighbor Embedding,” Proceedings of Advances in Neural Information Processing Systems(NIPS), pp. 833-840, 2002.

[18] L. J. P. van der Maaten and G. E. Hinton, “Visualizing High-Dimensional Data Using t-SNE,” Journal of Machine Learning Research, Vol. 9, pp. 2579-2695, 2008.

[19] A. van der SCHAAF, J. H. van HATEREN “Modelling the Power Spectra of Natural Images: Statistics and Information,” Vision Research, Volume 36, Issue 17, pp. 2759-2770, 1996.

[20] Koch, Michael and Denzler, Joachim and Redies, Christoph, “1/f2 Characteristics and Isotropy in the Fourier Power Spectra of Visual Art, Cartoons, Comics, Mangas, and Different Categories of Photographs,” PLoS One, Vol. 5, No. 8, p. e12268, 2010.

[21] Venkataramanan, Abhinau K and Wu, Chengyang and Bovik, Alan C and Katsavounidis “A Hitchhiker’s Guide to Structural Similarity,” IEEE Access, Vol. 9, pp. 28872-28896, 2021.

[22] Moorthy, Anush K and Bovik, Alan C, “Perceptually Significant Spatial Pooling Techniques for Image Quality Assessment,” Human Vision and Electronic Imaging XIV, Vol. 7240, pp. 339-349, 2009.

[23] Heusel, Martin and Ramsauer, Hubert and Unterthiner, Thomas and Nessler, Bernhard and Hochreiter, Sepp, “Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” Advances in Neural Information Processing Systems, Vol. 30, 2017.

[24] Obukhov, Artem and Krasnyanskiy, Mikhail, “Quality Assessment Method for GAN based on Modified Metrics Inception Score and Fréchet Inception Distance,” Proceedings of the Computational Methods in Systems and Software, pp. 102-114, 2020.

[25] FLIR Systems, 2019. FREE FLIR Thermal Dataset for Algorithm Training. https://flir.com/oem/adas/adas-dataset-form