

학습패치 크기와 ConvNeXt 적용이 CycleGAN 기반 위성영상 모의 정확도에 미치는 영향

The Effect of Training Patch Size and ConvNeXt application on the Accuracy of CycleGAN-based Satellite Image Simulation

원태연¹⁾ · 조수민²⁾ · 어양담³⁾

Won, Taeyeon · Jo, Su Min · Eo, Yang Dam

Abstract

A method of restoring the occluded area was proposed by referring to images taken with the same types of sensors on high-resolution optical satellite images through deep learning. For the natural continuity of the simulated image with the occlusion region and the surrounding image while maintaining the pixel distribution of the original image as much as possible in the patch segmentation image, CycleGAN (Cycle Generative Adversarial Network) method with ConvNeXt block applied was used to analyze three experimental regions. In addition, We compared the experimental results of a training patch size of 512*512 pixels and a 1024*1024 pixel size that was doubled. As a result of experimenting with three regions with different characteristics, the ConvNeXt CycleGAN methodology showed an improved R² value compared to the existing CycleGAN-applied image and histogram matching image. For the experiment by patch size used for training, an R² value of about 0.98 was generated for a patch of 1024*1024 pixels. Furthermore, As a result of comparing the pixel distribution for each image band, the simulation result trained with a large patch size showed a more similar histogram distribution to the original image. Therefore, by using ConvNeXt CycleGAN, which is more advanced than the image applied with the existing CycleGAN method and the histogram-matching image, it is possible to derive simulation results similar to the original image and perform a successful simulation.

Keywords : Occlusion, Image Restoration, ConvNeXt, CycleGAN, Training Patch

초 록

본 연구에서는 딥러닝을 통해 고해상도 광학 위성영상에 동종센서로 촬영한 영상을 참조하여 폐색 영역을 복원하는 방법을 제안하였다. 패치 단위로 분할된 영상에서 원본 영상의 화소 분포를 최대한 유지하며 폐색 영역을 모의한 영상과 주변 영상의 자연스러운 연속성을 위해 ConvNeXt 블록을 적용한 CycleGAN (Cycle Generative Adversarial Network) 방법을 사용하여 실험을 진행하였고 이를 3개의 실험지역에 대해 분석하였다. 또한, 학습 패치 크기를 512*512화소로 하는 경우와 2배 확장한 1024*1024화소 크기의 적용 결과도 비교하였다. 서로 특징이 다른 3개의 지역에 대하여 실험한 결과, ConvNeXt CycleGAN 방법론이 기존의 CycleGAN을 적용한 영상, Histogram matching 영상과 비교하여 개선된 R² 값을 보여줌을 확인하였다. 학습에 사용되는 패치 크기별 실험의 경우 1024*1024화소의 패치를 사용한 결과, 약 0.98의 R²값이 산출되었으며 영상밴드별 화소 분포를 비교한 결과에서도 큰 패치 크기로 학습한 모의 결과가 원본 영상과 더 유사한 히스토그램 분포를 나타내었다. 이를 통해, 기존의 CycleGAN을 적용한 영상 및 Histogram matching 영상보다 발전된 ConvNeXt CycleGAN을 사용할 때 원본 영상과 유사한 모의 결과를 도출할 수 있었고, 성공적인 모의를 수행할 수 있음을 확인하였다.

핵심어 : 폐색, 영상복원, ConvNeXt, CycleGAN, 학습패치

Received 2022. 06. 04, Revised 2022. 06. 14 Accepted 2022. 06. 25

1) Dept. of Advanced Technology Fusion, M.S. student, Konkuk University, Realtimevisual Inc. (E-mail: teadone@konkuk.ac.kr)

2) Dept. of Technology Fusion Engineering, M.S. student, Konkuk University (E-mail: jsm302@konkuk.ac.kr)

3) Corresponding Author, Member, Dept. of Civil and Environmental Engineering, Professor, Konkuk University (E-mail: eoandrew@konkuk.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

기후변화로 인한 구름 발생 및 대기오염 증가로 광학 영상 촬영 일수가 급격히 감소하여 적시에 영상을 획득하거나 시계열 분석에 필요한 주기적인 영상 획득이 어려워지고 있다(Lee *et al.*, 2017). 따라서 원격탐사 영상 사용자 입장에서는 결과의 신뢰도를 높이기 위해 사용 가능한 영상의 수를 늘리는 것이 현실적으로 매우 중요하다.

지금까지 악기상에 의한 광학 영상 촬영조건 제한을 극복하는 대안으로는 모의 영상 제작, 방사/대기보정 기법 등으로 접근하였으나 주로 중·저해상도 영상을 대상으로 하였고, 고해상도 영상에 대해서는 히스토그램 매칭 기법 정도가 비교적 보편적으로 사용하는 것으로 평가되고 있다. 기존의 기법들은 고해상도 영상의 경우 국소적 왜곡을 피할 수 없고(Lee *et al.*, 2017), 원본 영상의 일부가 어두운 그림자나 구름에 의해 완전히 가려지는 경우 대체 영상에서 해당 영역을 가져오거나 복원하는 새로운 방법을 고려해야 하므로 폐쇄된 영역을 복원하기 위한 연구가 필요하다.

Yoo and Lee (2010)는 항공영상과 LiDAR (Light Detection and Ranging) 자료에 대한 패치 기반 처리를 사용하여 폐쇄된 영역을 복원하는 연구를 수행하였다. 또한 Su *et al.* (2016)는 고해상도 전정색 영상에 대한 그림자 검출 및 그림자 폐색 복원에 관한 연구를 수행하였고, 바이모달(bimodal) 히스토그램 분할과 영상 매칭(matting) 기술을 결합하여 구름 탐지를 수행한 후 공간 적응형 비국소 희소(adaptive nonlocal sparse)로 복원하는 기술을 제안하였는데 이 방법은 그림자 영역에서 균일한 객체만 복원할 수 있고(Su *et al.*, 2016), 이론적으로는 관찰 시간과 카메라 앵글이 다르기에 복원을 위한 후처리 과정이 복잡할 수밖에 없다.

이러한 한계점을 극복하기 위해 최근에는 분류 및 탐지의 범위를 넘어 인공지능의 발달로 원격탐사 분야에서도 다양한 자료를 이용하여 학습한 뒤, 유사한 특성을 가진 출력값을 생성하는 GAN (Generative Adversarial Network) 모델을 적용한 연구가 진행되고 있다. 특히 고해상도 위성 영상의 경우 정밀한 기하학적 정합과 그림자로 인한 한계가 존재하기 때문에 특정한 영역만을 고려한 제한된 복원이 아니라 전체 영상 영역을 복원하는 방식이 유리하고 이를 위해서는 학습 기반 접근이 필수적이라는 주장을 뒷받침하는 기술 중 하나이다(Guo *et al.*, 2018; Zhang *et al.*, 2018). 특히, Won and Eo (2022)는 CycleGAN에 의해서 폐쇄된 영역을 복원하는 기술을 구현하였으나 실험기기의 성능한계로 영상을 분할하여 학습을 진행하였으며 영상 패치간 화소값 불연속이 존재하였

고 영상 배치크기를 2 이하로 설정하여 학습에 오랜 시간이 소요되었다.

본 연구에서는 학습환경의 한계에 따른 학습 패치 크기별 학습이 영상 복원 모의 결과에 어떤 영향을 미치는지 분석하였고, 기존의 Resnet 기반 CycleGAN에서는 입력 영상의 모든 채널에 영향을 받는 Convolution 필터를 사용함으로써 각 채널별 공간적 특징이 모호해지는 단점이 있으므로 채널별 특징을 추출하는 Depthwise Convolution 기반의 ConvNeXt CycleGAN을 적용하여 그 모의 정확도 향상 여부를 분석하였다.

2. 적대적 생성 신경망

GAN은 생산적 적대 신경망이라고 부르기도 하며, 기존 기계학습 분야에서 사용되던 적대적 기계학습(Adversarial Machine Learning)과 유사한 개념을 지니고 있다. GAN 알고리즘의 구조는 크게 두 모델로 이루어져 있는데, 생성을 담당하는 G(Generator)모델과 판독을 담당하는 D(Discriminator) 모델이 있다. 이 두 모델은 서로 적대적인 학습을 진행하며, G 모델은 D모델이 참(True)이라고 예측하도록 하는 결과물을 생성해내고, D모델은 실제 자료와 G모델이 생성해내는 결과물의 참, 거짓(False)을 구분하도록 학습시킨다. 결과적으로 Random Noise 혹은 Latent Variable이라고 불리는 랜덤 벡터 z 를 입력으로 사용해 참인 자료와 유사한 샘플링 분포를 지닌 자료를 생성하게 된다. 두 모델 G, D는 연결된 손실함수(loss function)를 통해 가중치가 갱신되어 학습이 서로 간에 이루어지게 된다(Goodfellow *et al.*, 2014). GAN 알고리즘의 발명은 무엇인가를 분류하고, 탐지하는 제한적인 딥러닝에서 학습 데이터를 모방한 결과물을 만들어내는 새로운 패러다임을 제시하였다.

이후 단순한 입력 데이터에 기반한 자료를 생성하는 방법론이 아닌 특정 데이터 세트 사이에 관계를 생성하여 영상을 만들어내는 영상 변환 방법론이 개발되었다. Conditional GAN의 일종인 Pix2Pix는 입력 영상과 출력 영상을 매핑하는 것이 목표이고, 두 데이터 세트 사이의 자료들을 짝을 지어 학습시킨다는 특징이 존재한다(Isola *et al.*, 2017).

CycleGAN은 Pix2pix 방법론에서 연장된 방법론이며 2017년도에 발표되었다. 이 방법론은 입력 데이터가 짝이 지어지지 (paired) 학습되어야만 했던 Pix2pix와는 다르게 짝이 지어지지 않은(unpaired) 학습 사이에도 매핑이 가능하다는 장점이 있다. 입력 영상이 모델을 거쳐 출력 영상으로 변하며, 이 출력 영상을 원래의 입력 영상에 매핑하는 언어 번역 등에 사용되

는 순환 일관성 개념을 채용하였다. 순환 일관성을 이용한 손실함수 덕분에 영상 간의 손실, 변화를 최소화하여 서로의 특징을 잘 변화시킬 수 있었다(Zhu *et al.*, 2017).

본 연구에서 제안하는 방법은 기하학적 정합 없이 두 영상 간의 관계를 학습함으로써 기하학적 일관성을 유지하기 어렵다는 문제를 극복하는 것을 목적으로 한다. 따라서 기하학적 정합 없이 고해상도 위성 영상 모의를 위해 데이터를 짝지어 학습해야 하는 Pix2Pix와 같은 paired-learning 방식 대신 데이터를 짝짓지 않고 학습을 진행하는 unpaired-learning 방식인 CycleGAN 방법론을 선택하여 수정 및 개선하였다.

3. 제안 방법론

3.1 Transformer, ConvNeXt

최근 2010년대의 말부터 합성곱 신경망 대신, 자연어 처리에서 사용되던 모델 구조를 차용하여 컴퓨터 비전 분야에 접목하려는 시도가 계속되어왔다. Vaswani *et al.* (2017)는 기존 RNN (Recurrent Neural Network)과 같은 구조에서 주로 사용되는 Attention 메커니즘을 사용하여 Transformer라는 모델을 제안하였고, 이는 자연어 처리 분야에서 우월한 성능을 보여주었다.

Alexey Dosovitskiy *et al.* (2021)는 자연어 처리에서만 사용되어왔던 이 Transformer를 컴퓨터 비전 분야에 접목함으로써 ViT (Vision Transformer)라는 방법론을 제안하였다. 이 ViT는 영상을 16*16화소의 패치로 분할한 후, Embedding 이라고 불리는 벡터 매핑 계층을 사용하였다. 이후 변환된 벡터를 Transformer 모델에 입력하였고 패치마다의 학습을 진행하여 기존 존재하는 CNN보다 우월한 성능을 도출하였다. 또한 제안한 ViT구조를 기존 CNN의 특징 추출계층인 Convolutional Layer와 융합하여 사용할 수 있다고 발표하였다. 하지만 Transformer 구조는 기존 CNN분야에서 사용되어왔던 객체탐지(Object Detection)나 세그멘테이션(Segmentation)에 적용하기 어렵다는 한계가 존재했고, 이를 극복하기 위한 다양한 연구가 진행되어왔다(Dosovitskiy *et al.*, 2021).

2022년 Liu *et al.* (2022)는 Transformer가 아닌 합성곱, 풀링과 같은 CNN의 특징을 사용한 새로운 딥러닝 방법론을 제안하였다. Liu *et al.* (2022)는 새로운 모델을 ConvNeXt라 칭하였으며, 이는 ResNet과 Transformer의 일부를 차용하여 개선한 모델이다(Liu *et al.*, 2022; Liu Z *et al.*, 2021). ResNet을 개량하여 일반적으로 CNN이 지닌 단순성과 효율성을 유지하면서도 이미 존재하는 많은 객체탐지와 세그멘테이션 알

고리즘에 기존 CNN을 적용하듯 자유롭게 사용이 가능하다는 점을 내세웠다. 기존의 ResNet과 가장 큰 차이점으로는 일반적인 convolution 연산이 아닌 Depthwise convolution을 사용한다는 것이며, 이 Depthwise convolution은 같은 채널끼리 convolution을 진행한다는 특징이 있다(Kim *et al.*, 2022). 일반적인 convolution은 모든 채널에 대한 공간적인 특징을 학습하므로 각 채널의 특정한 정보를 추출하기 어렵다. 반면에 Depthwise convolution은 각 채널의 Feature Map을 형성하여 각각의 공간적인 특성을 학습시킬 수 있다. 그러나 Depthwise convolution을 적용하면 시간이 오래 걸리며 그래픽 카드의 메모리를 더 많이 사용하게 된다. Liu *et al.* (2022)는 이러한 단점을 극복하기 위하여 Depthwise convolution layer를 Inverted bottleneck구조의 첫 단계로 진행하면서 채널 수를 줄였으며 일반적인 convolution 연산에 비하여 FLOPs(Floating point OPERations)를 효과적으로 감소시켜 결과적으로 매개변수의 수와 연산량이 적다는 장점을 확인하였다. 또한 ImageNet-22K, IFT-300M과 같은 대용량의 데이터 세트를 사전 훈련 시 향상된 성능을 보여주었다(Liu *et al.*, 2022). 본 연구에서는 CycleGAN 모델의 Discriminator의 경우 ResNet 구조에 기반한 모델을 사용하였고, Generator의 경우 기존 활용하였던 ResNet Generator의 구조 중, ResNet block 대신 기존의 CNN보다 전역적인 특징에 중점을 둔 ConvNeXt block을 적용하였다.

3.2 패치 크기 확장

일반적으로 그래픽카드의 메모리(VRAM) 내부에서는 주로 32비트(bit) 부동소수점 데이터 형식을 사용하여 연산을 진행한다. 하지만 이 32비트 부동소수점은 $1.175494e-38 \sim 3.402823e+38$ 로 엄청난 범위와 4바이트(byte)라는 용량을 지닌다. 그래픽카드 제조사 Nvidia의 연구팀은 딥러닝을 진행할 때 사용하는 모든 가중치가 담기는 노드는 이 32비트 부동소수점을 통해 할당되고 엄청난 용량과 넓은 범위로 불필요할 정도로 세밀한 연산을 진행한다는 연구 결과를 발표하였다. 따라서 이 노드에 할당되는 데이터의 범위를 축소 시킨다면, 그만큼 학습 속도의 향상과 정확도의 향상, 그리고 자원의 효율적인 분배가 가능할 것으로 추측하였다. 이 추측을 통해 개발된 AMP (Automatic Mixed Precision) (<https://pytorch.org/docs/stable/amp.html>) 방법론은 그들의 예상대로 학습에 사용하는 GPU의 내부 노드의 데이터 범위를 16비트 부동소수점 데이터 형식 $5.96e-8 \sim 65504$ 로 바꾸는 것만으로도 소모되는 그래픽카드 메모리의 양을 40퍼센트 이상을 줄일 수 있었다. 이뿐만 아니라 학습에 소모되는 시간도 30

퍼센트 이상 감소하였으며 미세하지만 정확도 향상도 이뤄낼 수 있었다(Micikevicius *et al.*, 2018). 이 AMP 방법론은 과거에는 APEX (PyTorch Extension)로 불리었으며 Pytorch의 외부 확장 모듈로 제공되었지만, 최근에는 AMP라는 모듈로써 Pytorch 프레임워크 내에서 사용할 수 있게 되었다.

본 논문에서는 패치 크기별 적용을 위해 AMP 방법론을 ConvNeXt CycleGAN의 G, D 모델 모두에 적용하여 학습 시간 및 그래픽카드 메모리 점유율을 낮추고, 낮은 그래픽카드 메모리 점유율을 이용하여 기존 512*512 화소로 분할하여 영상을 학습시키던 것에서, 1024*1024 화소로 영상을 분할하여 학습을 진행하였다.

4. 실험 및 분석

4.1 실험자료 및 지역

실험자료는 Kompsat-3A 위성영상으로서 수집 시간 등 Table 1에 정리되어 있다. 서로 다른 시기에 촬영한 영상들 간 겹치는 영역을 RoI (Region of Interest)로 설정하고 직사각형 모양으로 추출하였다. 그 후, 이미지는 ArcMap 프로그램의 Split Raster 도구를 사용하여 512*512화소의 패치로 분할시켰다. 따라서 3개의 다른 지역에서 이 과정을 수행하여 총 세 쌍의 데이터 세트를 구성하였고, 지역 1, 지역 2, 지역 3은 각각 데이터 A와 데이터 B 한 쌍으로 구성되어있다. 이후 약 80%의 영상이 훈련 자료로 사용되었으며, 그 외 20%는 실험자료로 사용하였다.

본 실험에 사용한 데이터는 Fig. 1과 같이 도심과 산림이 고루 분포한 지역, 도심만 존재하는 지역, 농촌과 산림 그리고 해안이 분포하는 지역들로 구성하였다. 따라서, 서로 다른 특성들을 가진 각각의 지역에 따라 동일한 학습 및 모의를 진행하였을 때 정확도의 변화를 비교하고자 하였다. 지역 1의 Data A는 맑은 날씨의 영상이고 Data B는 영상 전역에 걸쳐 구름과 에어로졸이 분포하고 있으며, 지역 2와 지역 3의 경우는 Data A와 B 모두 에어로졸이나 구름이 포함되어 있지 않은 맑은 날씨의 영상이다.

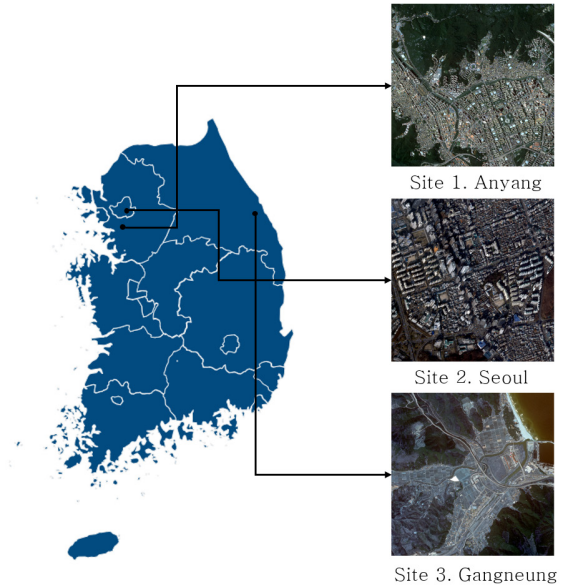


Fig. 1. Image locations used in the experiment

4.2 ConvNeXt 적용결과분석

Fig. 2에서 확인할 수 있듯이 ConvNeXt CycleGAN 방법론을 사용하여 2¹⁴(0~16383)의 값과 4개의 밴드를 갖는 Train_A 자료와 Train_B 자료를 사용하여 딥러닝 모델을 훈련시킨 후, 딥러닝 모델에 Test_A 자료를 입력하여 결과 Fake_B 자료를 얻었다. 이때 Train_A, Train_B는 서로 간에 특징이 변환되도록 학습시키는 자료이며, Test_A, Test_B는 학습이 완료된 모델에 입력하여 Test_A를 기반으로 B 영상과 같이 모의한 Fake_B와 Test_B를 기반으로 A 영상과 같이 모의한 영상인 Fake_A로 도출된다. 정확도 평가를 위해 Test_A를 Test_B의 히스토그램에 매칭시킨 영상은 ArcMap 도구를 사용하여 모의 자료와 같이 512*512 패치로 분할하였다. 그리고 모든 실험에서 원본 영상 데이터 세트를 Test_B 데이터 세트로 설정하였다. 이후 기존의 ResNet기반 CycleGAN 방법론과 널리 사용되는 히스토그램 매칭 기술을 사용하여 ConvNeXt CycleGAN 방법론과 비교하였다.

세 지역에 대한 학습 및 모의 결과는 Table 2와 같으며, ConvNeXt CycleGAN 방법론, 기존의 CycleGAN 방법론, Histogram matching을 통해 모의한 실험 결과의 일부를 확대한 영상인 Figs. 3 and 4 and 5를 원본 영상인 Fig. 6와 비교하면, 제안한 방법론의 결과가 기존의 CycleGAN 방법론의 결과에 비하여 더 뚜렷하게 보임을 확인하였다. 또한, 제안한 방법론의 결과가 기존의 CycleGAN 방법론의 결과에 비해 영상 내부의 갈라지는 증상과 화소가 깨지는 증상이 감소

Table 1. Kompsat-3A Images used in the experiment

	Site 1	Site 2	Site 3
Data A date	2019.09.18	2015.12.11	2019.03.08
Data B date	2019.09.17	2016.01.08	2019.04.12
Types of Land-covers	urban, forest	urban	rural, coast, forest

하였으며 기존의 방법론에 비해 영상의 색감이 시각적으로 원본 영상에 가까워 보임을 확인할 수 있었다. 산악 지역과 도시 지역이 혼합된 지역 1에서는 두 자료 세트 중 Data B에 구름과 에어로졸이 전반적으로 분포되어 있다. 따라서 정확도를 평가하기 위해 학습에 사용한 자료를 예측에도 사용하였고 구름이 존재하는 영역의 패치는 평가에서 배제하였다. 따라서 지역 1에 대한 평가는 다른 지역과 달리 과대평가되었을 가능성이 있다.

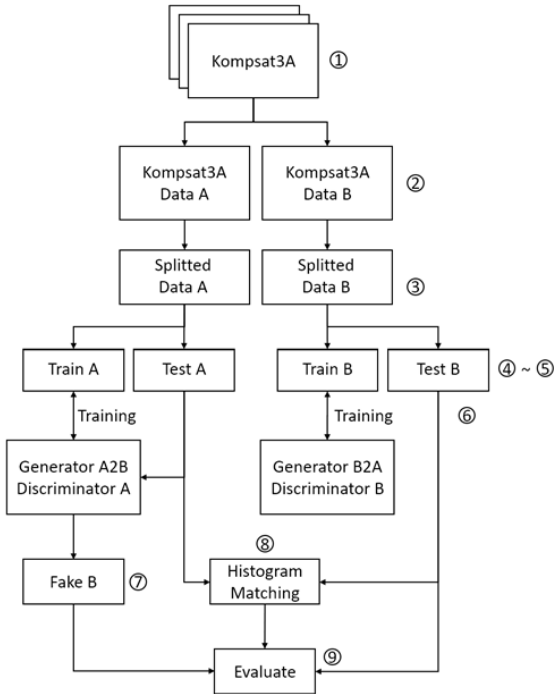


Fig. 2. Overall sequence flow of experimental



Fig. 3. ConvNeXt CycleGAN result image



Fig. 4. CycleGAN result image



Fig. 5. Histogram matching result image



Fig. 6. Original image

Table 2. Results of applying ConvNeXt CycleGAN

	ConvNeXt CycleGAN	CycleGAN	Histogram matching
Site 1			
RMSE	168.39	169.36	170.43
R ²	0.9930	0.9911	0.9896
Site 2			
RMSE	172.98	171.83	171.60
R ²	0.9884	0.9880	0.9853
Site 3			
RMSE	173.43	170.53	166.19
R ²	0.9891	0.9889	0.9869

도시 지역으로 구성된 지역 2의 경우 RMSE 172.98 및 R² 0.9884의 결과로 기존의 ResNet 기반 CycleGAN 방법론과 히스토그램 매칭 모의 결과와 비교하여 유사한 RMSE와 더 높은 R² 값을 도출하였다. 마찬가지로 지역 3의 경우 RMSE 173.43, R² 0.9891로 더 높은 R² 값을 도출하였다. 지역 3의 경우에는 바다가 포함된 지역이며, 해안 영역이 학습에 부정적인 영향을 미치기 때문에 히스토그램 매칭보다 RMSE 값이 더 크다고 판단하였다. 이러한 결과들은 단순히 화소값의 분포를 조정하는 히스토그램 매칭과 달리 ConvNeXt CycleGAN 방법론은 A자료와 B자료의 특징 간의 관계를 학습하고 원래의 영상으로 돌아갈 수 있도록 화소의 변화를 최소화하는 순환 개념을 사용하여 특징들을 서로 간의 변환과정을 통해 더 높은 상관관계를 보인 것으로 판단하였다.

하지만 학습 패치를 10개~15개를 사용한 것과 같은 학습 영상의 양이 충분하지 않은 경우에는 모의 영상에서 격자 형태의 화소 균열이 관찰되는 부정적인 결과가 도출되었다. 반

면에 학습에 20장 이상과 같은 충분한 자료를 사용한 경우에는 격자 형태의 화소 균열이 관찰되지 않았다. 따라서 폐색 영역의 시뮬레이션을 수행하려면 20개 이상의 패치를 이용하여 충분히 다양한 경우의 특성을 정확하게 반영할 수 있어야 한다고 판단하였다. 영상 상의 특성이 다른 영역을 시뮬레이션하기 위해서는 다시 학습을 진행해야 하며, 영상이 보유하고 있는 특징이 다양할수록 더 많은 epoch를 부여해 영상에 대한 특징을 고루 학습하여 낮은 오차를 얻을 수 있는 것으로 확인하였다. 또한, 실험기기의 한계가 있었지만 제안한 ConvNeXt CycleGAN 방법론은 기존의 모의 영상 기법이나 기존의 ResNet 기반의 CycleGAN 방법론에 비해 전역적인 특징을 더욱 잘 모의할 수 있으며, 추가적인 하이퍼 파라미터를 수정함으로써 Fig. 7과 같이 영상의 기하학적인 특징도 육안으로 식별이 가능할 정도로 변환이 가능한 것을 확인하였다.

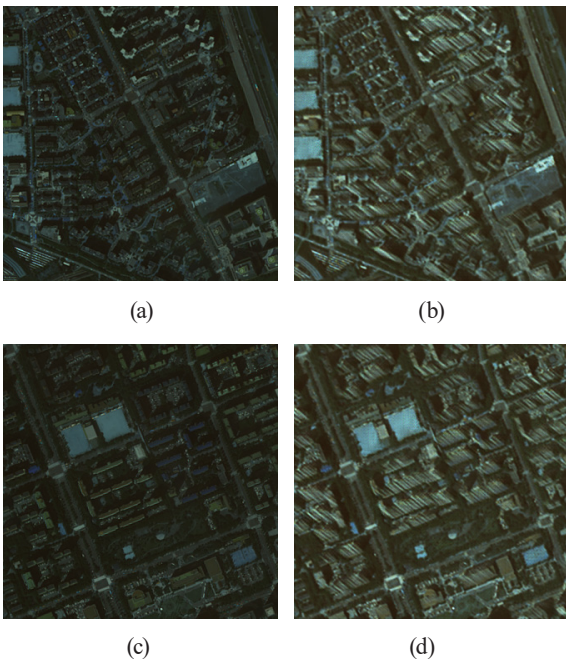


Fig. 7. Additional experimental results of changing hyper parameters for the ConvNeXt CycleGAN, result (b) is the result of inputting image (a) into the ConvNeXt CycleGAN, and result (d) is the result of inputting image (c) into the ConvNeXt CycleGAN.

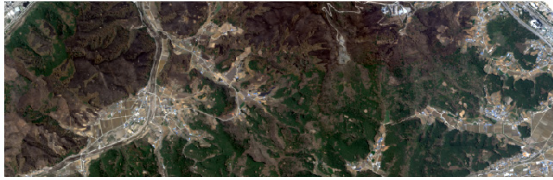
4.3 패치크기 변화에 따른 분석

지역 1, 2, 3의 영상을 512*512화소의 패치와 1024*1024화소의 패치로 분할하여 추가로 비교 실험을 진행하였다. 두 영

상의 크기 차이는 약 4배이며, Test와 Train 자료에 들어갈 영상의 개수도 512*512화소의 패치가 4배 많았다. 하이퍼 파라미터의 경우 지역 1, 2는 learning rate 0.0005, epoch 150, poolsize 16, 지역 3은 지역 1,2의 매개변수에서 poolsize만 20으로 변경하였으며, 512*512 패치와 1024*1024 패치 모두 동일하게 부여하여 실험하였다. 512*512화소로 분할하여 실험을 진행할 때는 데이터 셋을 구성하기에 충분하였으나 4배 정도 큰 1024*1024화소의 영상으로 실험을 진행할 때는 패치 크기를 증가시켰기 때문에 같은 영역이지만 데이터 셋의 양이 줄어든다. 이를 극복하기 위하여 두 가지 방법으로, 첫 번째는 양이 적은 데이터 셋을 그대로 사용하는 방법과 두 번째는 한 영상에서 화소의 개수를 다르게 조정하여 여러 영상을 만들었고 이를 분할하여 많은 학습 패치를 구성하도록 하였다. 그러나 실험 결과가 비슷한 값을 나타내었기에 첫 번째 방법을 사용하여 실험을 진행하였다. 실험 결과, 다음 Table 3과 같이 1024*1024 패치로 진행한 실험이 512*512 패치로 진행한 실험에 비하여 향상된 RMSE 값을 도출하였다. R²값은 낮은 값을 보였는데 이는 1024*1024 패치가 더 넓은 지역이며 학습 데이터 셋이 적음에도 불구하고 그 차이는 미미하였다. 다음 Fig. 8은 지역 3의 일부 영상이며 (a)는 원본 영상, (b), (c)는 학습 패치별 실험 결과이다. 실험 결과, 시각적으로 확인하였을 때 일부 경계가 보이며 패치 크기가 커질수록 이러한 단점을 보완되는 것을 확인하였다. 추가로 학습 패치별 실험 결과의 화소 분포가 원본 영상과 얼마나 유사한지 확인하기 위하여 지역 1, 2, 3의 원본 영상(1), 실험을 통해 얻은 1024*1024 크기의 영상(2)와 512*512 크기의 영상(3)의 대역별 표준편차를 Table 4에 비교하였다. (1)은 1024*1024 크기의 원본 영상, (2)는 1024*1024 크기의 패치 1장을 사용하였으며 (3)은 512*512 크기의 패치 4장의 영상을 모자이크하였다. 비교 결과, 지역 1, 2, 3 모두에서 1024*1024 패치의 실험 결과가 원본 영상과 더 유사한 히스토그램 표준편차를 도출하였다. 이는 영상 패치 크기가 클수록 학습되는 양이 많아지고 데이터 셋의 양은 부족하여도 결과적으로 원본과 유사한 영상을 모의해낼 수 있음을 확인하였다.

Table 3. Results by training patch size

patch size	512*512		1024*1024	
	RMSE	R ²	RMSE	R ²
Site 1	168.39	0.9930	167.70	0.9889
Site 2	172.98	0.9884	164.04	0.9826
Site 3	173.43	0.9891	154.01	0.9858



(a) Original image



(b) Fake_B image of 512*512 patch image



(c) Fake_B image of 1024*1024 patch image

Fig. 8. Mosaic image of ConvNeXt CycleGAN experiment result by training patch in Site 3

Table 4. Comparison of histogram standard deviations of sites 1,2 and 3 images

Site 1	Std.deviation			
	R	G	B	N
(1)	469.37	721.46	778.85	1545.96
(2)	608.58	944.12	974.39	1827.72
(3)	778.53	733.07	512.13	2156.35
Site 2	Std.deviation			
	R	G	B	N
(1)	563.12	969.10	1175.49	1625.19
(2)	682.03	1163.55	1366.22	1841.47
(3)	1126.11	1712.87	1872.72	2501.86
Site 3	Std.deviation			
	R	G	B	N
(1)	634.54	482.20	264.60	1344.77
(2)	649.57	521.24	287.23	1396.27
(3)	776.40	614.69	350.88	1583.64

4. 결론

본 연구에서는 ConvNeXt CycleGAN 모델을 적용하여 고

해상도 위성영상의 모의를 수행하고 그 정확도를 평가하였다. 제안한 방법은 기존 히스토그램 매칭, CycleGAN 방법론을 통해 얻은 영상 모의 결과와 비교하였을 때 실험지역에 대해서 결정계수 R^2 가 근소하게 향상되는 것으로 나타났다. 특히, 제안한 방법은 서로 간의 변환과정을 통해 ResNet 기반 CycleGAN 기법보다 영상이 지닌 건물의 기울기와 같은 기하학적인 특징을 우수하게 표현함을 확인하였다.

영상 딥러닝을 수행하는데 필요한 하드웨어 성능한계를 극복하는 방안으로 AMP 기술 적용하여 학습 패치 크기를 2배 확대하였으며 그 효용성을 분석하였다. 이때 1024*1024 패치로 실험한 값은 512*512 패치로 실험한 값보다 향상된 RMSE를 나타내었고, 결정계수 R^2 가 저조하였지만 차이는 미미하였다. 또한, 1024*1024 패치의 영상이 R, G, B, NIR 대역의 화소 분포도 원본과 더 유사하였다. 학습 패치별 실험의 경우, 패치가 클수록 데이터 셋을 구성하는 영상의 개수가 줄어들어도 개선된 결과를 도출하며 모의가 자연스럽게 진행됨을 확인할 수 있었다.

추후 연구로는 충분한 학습 패치를 포함하고 다양한 토지 피복 환경들로 구성된 위성영상들의 실험을 통해 제안방법의 보편적인 우수성을 검증할 필요가 있다. 또한, 패치 단위로 모의 영상을 모자이크(mosaic)하면 패치 간 경계 화소에 불연속성이 나타나므로 전체 영상을 학습 단위로 하기 전에는 불연속성에 대한 극복 문제점 해결방안을 연구하고, 모의 영상 정확도 향상을 위해 실제 정사영상(Ortho image) 및 다른 영상 자료와의 혼합, 전이 학습 도입 등의 방안을 고려한 연구가 필요할 것으로 사료한다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019R1A2C1085618).

References

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houshyby, N. (2021), An image is worth 16x16 words: transformers for image recognition at scale, *International Conference on Learning Representations 2021*, 3-7 May. <https://doi.org/10.48550/arXiv.2010.11929>

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014), Generative adversarial nets, *Neural Information Processing Systems-2014*, 8-13 December, Montreal, Canada, pp. 2672-2680.
- Guo, Q., He, M., and Li, A. (2018), High-resolution remote-sensing image registration based on angle matching of edge point features, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 11, No. 8, pp. 2881-2895.
<https://doi.org/10.1109/JSTARS.2018.2844295>
- Isola, P., Zhu, J.Y., Zhou, T., and Efros, and A.A. (2017), Image-to-image translation with conditional adversarial networks, *IEEE Conference on computer vision and pattern recognition-2017*, 21-26 July, Honolulu, USA, pp. 1125-1134.
- Kim, D., Wang, K., Sclaroff, S., and Saenko, K. (2022), A Broad Study of Pre-training for Domain Generalization and Adaptation, *arXiv preprint arXiv:2203.11819*.
<https://doi.org/10.48550/arXiv.2203.11819>
- LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D. (1989), Backpropagation applied to handwritten zip code recognition, *Neural computation*, Vol. 1, No. 4, pp. 541-551.<https://doi.org/10.1162/neco.1989.1.4.541>
- Lee, M.H., Lee, S.B., Eo, Y.D., Kim, S.W., Woo, J.H., and Han, S.H. (2017), A comparative study on generating simulated Landsat NDVI images using data fusion and regression method-the case of the Korean Peninsula, *Environmental Monitoring and Assessment*, Vol. 189, No. 7, pp. 1-13.
<https://doi.org/10.1007/s10661-017-6034-z>
- Lee, S.B., Park, W.Y., Eo, Y.D., Pyeon, M.W., Han, S., Yeon, S.H., and Lee, B.K. (2017), Analysis on the applicability of simulated image from SPOT 4 HRVIR image, *Korean Society of Civil Engineers*, Vol. 21, No. 4, pp. 1434-1442.
<https://doi.org/10.1007/s12205-016-0522-5>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021), Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, *arXiv preprint arXiv:2103.14020*.
<https://doi.org/10.48550/arXiv.2103.14030>
- Liu, M., Ma, J., Zheng, Q., Liu, Y., and Shi, G. (2022), 3D Object Detection Based on Attention and Multi-Scale Feature Fusion, *Sensor 2022*, Vol. 22, No. 10.
<https://doi.org/10.3390/s22103935>
- Liu Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., (2022), A ConvNet for the 2020s, *arXiv preprint arXiv:2201.03545*.
<https://doi.org/10.48550/arXiv.2201.03545>
- Liu, Y., Yano, T., Nishiyama, S., and Kimura, R. (2017), Radiometric correction for linear change-detection techniques: analysis in bi-temporal space, *International Journal of Remote Sensing*, Vol. 28, No. 22, pp. 5143-5157.
<https://doi.org/10.1080/01431160701268954>
- Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., and Wu, H. (2018), Mixed precision training, *International Conference on Learning Representations-2018*, 30-3 April-May, Vancouver, Canada.
- Su, N., Zhang, Y., Tian, S., Yan, Y., and Miao, X. (2016), Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images, *IEEE Applied Earth Observations and Remote Sensing*, Vol. 9, No. 6, pp. 2568-2582.
<https://doi.org/10.1109/JSTARS.2016.2570234>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017), *Attention is all you need*, *Neural Information Processing Systems-2017*, 4-9 December, Long Beach, USA, pp. 5998-6008.
- Won.T. and Eo, Y.D. (2022), An experiment on image restoration Applying the Cycle Generative adversarial network to partial occlusion Kompsat-3A image, *Korean Journal of Remote Sensing*, Vol. 38, No. 1, pp.33-43.
<https://doi.org/10.7780/kjrs.2022.38.1.3>
- Yoo, E.J. and Lee, D.C. (2010), Patch-based processing and occlusion area recovery for true orthoimage generation, *Journal of the Korean Society of Surveying*, Vol. 28, No. 1, pp. 83-92. (in Korean with English abstract)
- Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T.S. (2019), Free-form image inpainting with gated convolution, *International Conference on Computer Vision*, pp. 4471-

4480.

<https://doi.org/10.48550/arXiv.1806.03589>

Zhu, J.Y., Park, T., Isola, P., and Efros, A.A. (2017), Unpaired image-to-image translation using cycle-consistent adversarial networks, *International Conference on Computer Vision*, pp. 2223-2232.

<https://doi.org/10.48550/arXiv.1703.10593>

Zhang, Q., Yuan, Q., Zeng, C., Li, X., and Wei, Y. (2018), Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 8, pp. 4274-4288.

<https://doi.org/10.1109/TGRS.2018.2810208>