

Image Colorization

최진영
이OO
오OO
이OO

1. 서론

1.1. 프로젝트 목표

흑백 이미지는 각자의 추억을 담고 있거나 역사적 가치를 가지고 있다. 하지만 흑백 이미지들은 컬러가 없어 그 당시의 분위기나 감정을 고스란히 전하지 못한다. 생동감이 떨어지고, 해당 이미지가 담고 있는 정보의 전달력 또한 제한적이다. 이러한 흑백 이미지를 컬러로 복원한다면 개인적 측면에서는 사람들이 각자의 추억을 생생하게 되살릴 수 있고, 사회적 측면에서는 역사적으로 의미있는 자료들을 현대적으로 재조명함으로써 새로운 사회적 가치를 창출할 수 있다.

우리는 흑백 이미지를 입력 받으면 컬러 이미지로 변환해 주는 모델을 만드는 것을 이 프로젝트의 주제로 한다. 이를 통해 기존 흑백 이미지의 수동 컬러화가 효율성과 일관성이 부족해 대규모 작업이 어려운 점을 극복하는 것을 목표로 한다. 자동으로 컬러화를 수행해 더 빠르고 일관된 결과를 사용자에게 제공함으로써 단순히 색을

입히는 것을 넘어 당시의 맥락과 감정을 보다 사실적으로 복원해 사용자가 더욱 풍부한 경험을 할 수 있도록 돋는 것이 이 프로젝트의 궁극적인 목표이다.

1.2. 프로젝트 진행 방식

수집한 데이터는 흑백이미지와 그에 대응되는 컬러이미지의 쌍으로 구성되어 있다. 흑백이미지를 모델의 입력으로, 컬러이미지를 정답이 되는 label 값으로 지정하여 지도학습을 진행한다. 지도학습이란, 입력값과 그에 대응되는 target(label) 값이 존재하며 target을 잘 맞추는 것을 목표로 하는 인공지능 학습을 의미한다.

1.3. 데이터셋

프로젝트 시작 전에 수집한 데이터셋과 모델 최적화 과정에서 추가로 수집한 데이터를 모두 포함하여 최종적으로 3가지의 데이터셋을 이용하였다. 첫 번째 데이터셋은 이미지 파일(.png)로 이루어진 데이터셋이다. 또, 두 번째는 numpy 배열로 이루어진 LAB 색공간으로 정의되어 있는 흑백사진과 컬러사진 데이터셋이며, 세 번째는 Landscape

이미지 파일(jpg)로 되어 있는 흑백사진과 컬러사진 데이터셋이다.

1.4. 컬러사진 표현 방법

1.4.1. LAB

LAB 색공간은 밝기(L채널, Lightness)와 색상정보(A채널: 빨간색-초록색 축, B채널: 파란색-노란색 축)로 분리된 구조이다. 흑백 이미지를 컬러화할 때 밝기 정보를 유지하며 자연스러운 색상 정보를 추가하는데 유리하다.

1.4.2. RGB

RGB는 색을 빨강, 초록, 파랑의 세 가지 색상의 혼합으로 표현한다. 각 색상은 0~255 사이의 값으로 표현된다.

1.4.3. BGR

BGR은 RGB와 비슷하지만, 색상 순서가 반대이다. OpenCV 같은 이미지 처리 라이브러리에서 기본적으로 사용하는 형식이다.

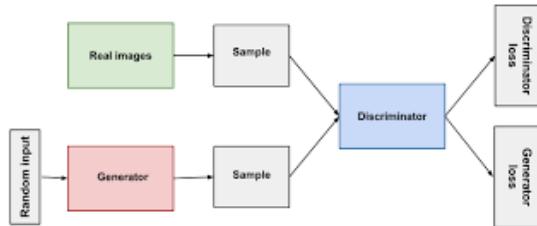
2. 모델링

2.1. 모델 소개

2.1.1. GAN

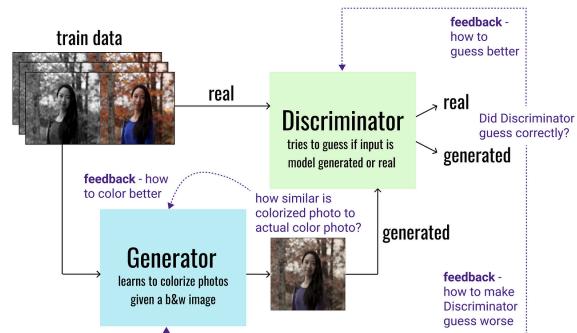
GAN 모델은 Generative Adversarial Network의 약자로, 두 개의 신경망인 Generator와 Discriminator(판별기)로 구성된 모델이다. 생성기는 입력으로부터 가짜 데이터를

생성하고 판별기는 주어진 데이터가 진짜인지 가짜인지 구분한다. 두 네트워크가 서로 경쟁하며 학습하기에 더 target에 가까운 데이터를 만들어낼 수 있다.



2.1.2. Pix2Pix

Pix2Pix 모델은 GAN 모델의 변형으로, 이미지-이미지 변환을 실행하기 위해 설계된 모델이다. Pix2Pix는 조건부 GAN(Condition GAN)을 기반으로 한다. 주로 사용되는 손실 함수로는 GAN 손실(BCE loss)과 L1 손실이 있다.



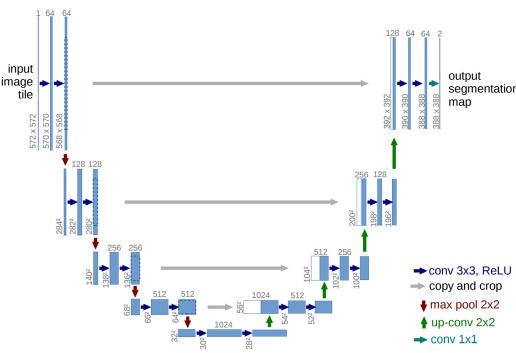
2.1.3. CWGAN

CWGAN이란, Conditional Wasserstein GAN의 약자로, 조건부 정보를 이용하여 데이터를 생성할 수 있는 모델을 뜻한다. GAN모델을 기반으로 하지만, Generator와 Critic으로 이루어져 있다. CWGAN에서 Generator는 조건부 정보에 따라 생성된 데이터가 가져야 하는 특성을 결정하고 랜덤 벡터를 입력 받아 가짜 데이터를 생성한다. Critic은 두 분포 간의 Wasserstein 거리를 계산하고 해당 조건에 맞는 데이터를 구별하는 역할을 한다.

2.1.4. U-Net

CNN(Convolutional Neural Network) 기반의 모델로, 인코더(Contracting Path)와 디코더(Expanding Path)로 구성되어 있고 두 경로가 연결되면서 U자 형태를 이룬다.

U-Net의 인코더는 입력 이미지를 작은 특성맵으로 변환하여 특징을 추출하는 목적을 가지고 있다. 여러 개의 Convolution Layer와 Max Pooling 2D로 이루어져 있다. 디코더의 경우 인코더에서 추출한 특징을 바탕으로, 해상도를 점진적으로 복원하여 원래 이미지를 재구성한다. 디코더에는 Up-sampling(Deconvolution)과 Skip Connections가 존재한다.



2.2. 1차 모델링

2.2.1. 1차 데이터셋

첫 번째 데이터셋은 이미지 파일(.png)로 이루어진 파일이며 256 x 256 크기의 10개의 쌍이 있는 데이터셋이다.

2.2.2. 모델 구조

1차 모델링에 사용된 모델은 Pix2Pix이다. Pix2Pix 모델은 Generator 그리고 Discriminator 클래스로 구성되어 있다.

먼저, Generator 클래스는 인코더와 디코더, 즉 Convolution과 Deconvolution 과정이 모두 존재하는 U-Net 구조를 따른다. 인코더에서는 LeakyReLU¹ 활성화 함수를 이용하였고 디코더에서는 ReLU 활성화 함수를 이용하였다. 인코더를 통해 입력 이미지의 고차원 특징을 추출한 후 디코더를 사용하여 원본 이미지 크기로 복원하는 구조이다.

¹ 기존 ReLU함수는 음수 입력에 대해 항상 0을 출력하지만 LeakyReLU의 경우 음수 입력에서도 작은 기울기를 유지하기 때문에 뉴런이 완전히 죽는 상황을 피할 수 있다.

Discriminator 클래스는 PatchGAN의 구조를 가지고 있는데, 이미지를 작은 패치 단위로 평가한다는 의미이다. 각 패치의 진짜 또는 가짜 여부를 구분하고 결과를 종합하여 전체 이미지를 평가한다.

Pix2Pix 모델의 실행과정을 다시 요약하면, 256 x 256 크기의 흑백 이미지를 입력으로 받는다. 입력된 이미지는 Generator 클래스에서 컬러 이미지로 변환된다. 그 이후 Discriminator 클래스를 거쳐 출력 이미지의 로컬 패치를 평가받게 된다.

2.2.3. 모델 평가



위 사진은 Pix2Pix 모델의 target이었던 컬러 이미지와 모델이 생성한 이미지이다. 흑백에서 컬러화가 특정 부분 잘 진행된 것으로 보인다. 그러나, 사진의 화질이 너무 낮고 파란색이나 초록색 부분은 컬러화가 제대로 진행되는 않는다는 문제점이 있다.

2.2.4. 문제 분석

1차 데이터셋이 너무 부족하였고 기존 target이 되는 이미지들의 화질도 너무 낮았다. 작은 양의 데이터셋으로 컬러화가 진행될 수 있었던 이유는 U-Net 구조를 가진

Generator 클래스 덕분이었으나 성능을 더 높이기 위해서 데이터 양을 늘리는 Data Augmentation이 필요하다.

2.3. 2차 모델링

2.3.1. 2차 데이터셋

2차 모델링에서는 numpy 배열로 이루어진 LAB 색공간으로 정의되어 있는 흑백 사진과 컬러사진 데이터셋을 이용하였다. 데이터는 224 x 224 크기의 흑백이미지와 컬러이미지 각각 25,000장을 사용하였고 1차 데이터셋 이후 Data Augmentation을 진행하였다.

numpy 배열로 이루어져 있어 흑백이미지에 해당하는 데이터의 크기는 (224, 224, 1)이며 컬러 이미지에 해당하는 데이터의 크기는 (224, 224, 2)이다.

2.3.2. 모델 구조

2차 모델링에 사용한 모델은 CWGAN이다.

CWGAN의 Generator 클래스는 U-Net 기반으로, ResBlock을 이용하여 3x3 합성곱을 진행하고 배치를 정규화하며 ReLU 함수로 활성화한다. 이후 인코더에 해당하는 DownSampleConv가 있으며 디코더에 해당하는 UpSampleConv가 있다. Dropout(0.2)를 적용하여 과적합이 되는 것을 방지하였다.

CWGAN에는 Critic 클래스가 존재한다. Critic Class는 PatchGAN 기반이며, 4x4 합성곱을 진행하고 Instance Normalization하며

LeakyReLU 함수로 활성화하도록 구성되어 있다.

CWGAN의 손실함수는 2가지이다. 첫 번째 손실함수는 Generator 과정에서 사용되는 L1 loss이다. 생성된 컬러 정보와 실제 컬러 정보 간 차이를 측정한다. 두 번째 손실함수는 Critic 과정에서 사용되는 Wasserstein 거리²이다. 생성된 데이터와 실제 데이터의 분포 차이를 측정한다. 또, 이 과정에서 Gradient Penalty를 이용하여 학습의 안정성을 보장하게 하였다.

Optimizer로는 Adam을 사용하였고, StempLR scheduler를 이용하여 학습률을 조절하였다. 모델은 Critic과 Generator가 교대로 학습하게 구성하였고, Critic은 Generator보다 5배 더 자주 업데이트되도록 설정하였다.

Epoch는 150으로 설정하여 10번마다 이미지를 출력하게 하였다.

2.3.3. 1차 모델링과의 차이

Pix2Pix의 경우 Generator가 일반적인 Conv2D와 TransposeConv기반의 U-Net으로 구성되어 있었지만, 2차 모델링 CWGAN에서는 ResBlock 기반의 U-Net으로 변경하였다.

1차 모델링에서 사용한 손실 함수는 간단한 GAN(Binary CrossEntropy) loss와 L1 loss였지만

2차 모델링에서는 Wasserstein loss와 Gradient Penalty 그리고 L1 loss를 사용하였다.

CWGAN에서는 Generator보다 Critic을 더 자주 업데이트하는 학습 방식을 사용하지만, Pix2Pix에서는 Generator와 Discriminator를 동일한 빈도로 업데이트하여 학습한다. 또한, 2차 모델링에서는 Dropout을 적용하고 학습률 스케줄러를 포함하고 있지만 1차 모델링에서는 포함하고 있지 않다.

2.3.4. 모델 평가



위 사진은 CWGAN을 학습시킨 이후 각각 Epoch 20과 Epoch 100일 때의 결과 사진이다. 눈으로 봤을 때 생성된 사진이 target 사진과 차이가 있다. 생성된 사진은 완벽하게 컬러화가 되지 않고 있는 것으로 보인다.

또한, 모델을 평가할 때 쓰이는 2개의 손실함수 중 critic_loss의 값이 음수가 나오는 경우가 발생했다.

² Wasserstein 거리는 분포 간의 연속적인 차이를 측정하는 지표로, 학습 안정성과 품질을 높이는데 기여한다.

2.3.5. 모델 개선

Wasserstein 거리를 정의할 때 절댓값을 해주지 않아 critic_loss의 값이 음수가 되었음을 파악하고 수정하였다. 또한, 컬러화가 제대로 되고 있지 않는 것 같아 새로운 loss를 추가하여 학습에 반영하게 하였다.

새로운 loss는 Perceptual loss와 color loss이다. VGG³16 모델을 이용하여 Perceptual loss를 계산하였다. VGG 특성을 target 이미지와 생성된 이미지에 대해서 추출하고 MSE loss를 이용하여 Perceptual loss를 계산하였다. Color loss는 생성된 이미지(AB 채널)와 target 이미지(AB 채널) 간의 색 차이를 최소화하는 것을 목적으로 한다. 생성된 A, B 채널과 실제 A, B 채널의 차이를 L1 loss를 이용하여 Color loss를 계산한다. 컬러화가 제대로 진행되는 것 같지 않아 색 정확도에 대한 성능을 향상시키기 위해 사용하였다. 새롭게 추가한 Perceptual loss와 Color loss는 각각 0.01, 0.0005의 가중치를 부여 받아 기존의 L1 loss와 결합되어 재정의된 Generator loss가 된다.

학습 결과, 여전히 컬러화가 제대로 진행되지 않아 Generator loss를 계산할 때의 Perceptual loss, Color loss에 대한 가중치를 0.005, 0.1로 변경하였다. 결과는 아래와 같다.



생성된 이미지 중 컬러화가 된 부분은 target³이 되는 이미지와 동일하다.

2.3.6. 한계

데이터셋에 target 이미지 중 흑백으로 되어 있는 이미지가 섞여 있었다.



target 이미지가 흑백 이미지인데도 불구하고 근소하게 컬러화가 된 것에 의의가 있으나, 데이터셋을 변경할 필요가 있다.

2.4. 3차 모델링

2.4.1. 3차 데이터셋

320 x 320 크기의 흑백 이미지와 컬러 이미지 각각 7129장이 있는 Landscape 데이터셋을 이용하였다. 사람 얼굴이 있는 데이터를 제대로 컬러화를 진행하지 않는 것 같아 풍경 데이터셋을 사용하였다.

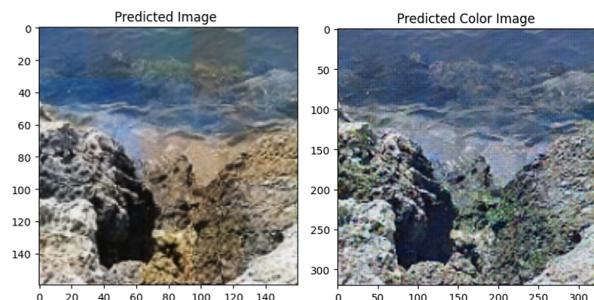
³ VGG는 사전 학습된 이미지 분류 네트워크로, 다양한 수준의 시각적 특징 추출 가능한 모델.

2.4.2. 모델 구조

3차 모델링에서 사용한 모델은 U-Net 구조를 기반으로 한 이미지 컬러화 모델이다.

모델은 입력 이미지를 down-sampling하고 다시 up-sampling을 통해 원래 크기로 복원하는 구조를 가지고 있다. down-sampling는 Conv2D를 통해 특징을 추출하고 BatchNormalization을 통해 학습 안정화를 진행하는 과정이다. Dropout을 적용하고 LeakyReLU 활성화 함수를 이용하였다. up-sampling은 전치 합성곱으로 원래 크기로 복원하는 과정이다.

이 과정을 통해 기존보다 더 세분화된 픽셀 단위의 학습을 가능하게 했다. 데이터셋의 이미지를 보면 이미지가 깨져있는 것을 확인 후 픽셀을 더 세분화하여 화질 개선을 시도하였다. 입력 데이터의 해상도가 클수록 이미지의 화질이 개선되는 것을 확인 후, 160x160 픽셀에서 320x320 픽셀로 확장하였다. 또, U-Net 구조 기반 모델을 학습시킨 초기에는 Perceptual loss를 MAE로 정의하여 사용하였으나 VGG19 네트워크를 사용하여 세부적인 구조를 더 잘 학습하도록 하였다.



VGG19 기반 Perceptual Loss는 단순 픽셀 값 손실을 줄이는 것뿐만 아니라, 시각적으로 더욱 자연스러운 색 복원이 이루어질 수 있도록 한다. VGG19의 중간 레이어에서 'block5_conv4' 출력을 활용하여 학습 품질을 높였다. 마지막으로, down sampling에서 크기를 (3,3)에서 (5,5)로 확장하여 더 넓은 영역에서 정보를 추출하였으며 batch size를 16에서 8로 조정하였다. 이에 따라 이미지의 세부 패턴을 더욱 효과적으로 학습하도록 개선하였다. 다양한 필터 크기와 레이어 구조를 실험적인 방법으로 많이 실행을 해보며 최적화를 시도하였다.

U-Net 모델을 사용 초기에는 여러 Optimizer 중 Adam을 사용하였고 손실 함수는 CrossEntropy loss만 사용하였다. 그러나 학습이 제대로 이루어지지 않아 learning rate scheduler를 epoch이 20이 넘으면 감소되도록 설계하여 학습률이 학습 중에 변경되도록 수정하였다.

다양한 하이퍼파라미터를 시도해보고자 Optimizer를 RAdam을 선정하여 1차 모멘텀과 2차 모멘텀을 각각 0.9, 0.999로 설정하여 모델을 학습시켜 보았다. 그러나 결과는 Optimizer Adam을 사용했을 때와 확연한 차이가 나지 않았다.

2.4.3. 이전 모델링과의 차이

1차와 2차 모델링에서 모두 U-Net 구조를 가진 Generator 클래스를 가지고 있었으나 3차 모델링에서는 U-Net 구조 기반 CNN 모델을 이용하였다. CNN 모델이라 말할 수 있는

이유는 합성곱과 전치 합성곱 층을 기반으로 설계되어 있기 때문이다.

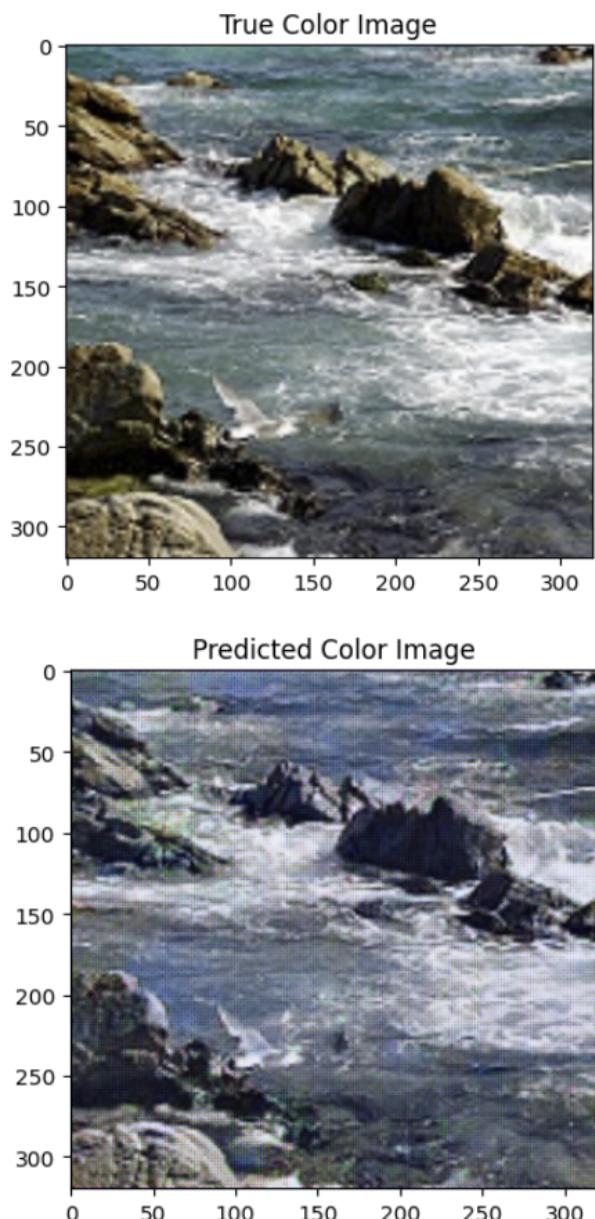
2차 모델링에서 Perceptual loss를 연산할 때 VGG16 모델을 사용했는데, 3차 모델링 U-Net 기반 이미지 컬러화 모델에서는 VGG19 모델을 사용하여 Perceptual loss를 연산한다. VGG16과 VGG19는 모두 3개의 완전 연결 레이어가 있지만 Convolution layer의 개수가 다르다. VGG19는 VGG16보다 Convolution layer가 3개가 더 많아 깊은 네트워크라 볼 수 있다.

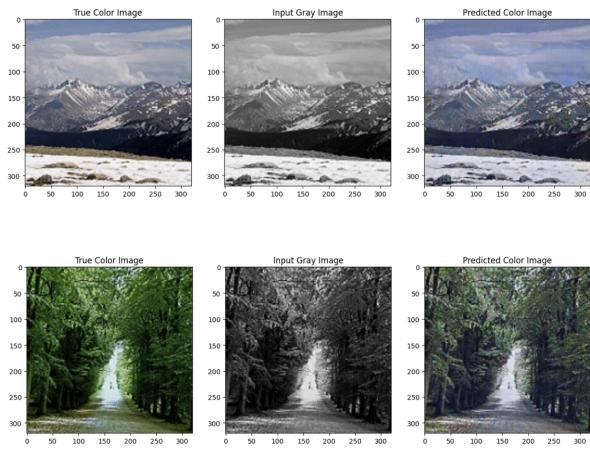
2.4.4. 모델 평가

모델을 평가하기 위해 2가지 지표를 도입하였다. 첫 번째 지표는 PSNR이다. PSNR은 Peak Signal-to-Noise Ratio로, 두 이미지 간의 픽셀 값 차이를 수학적으로 계산하여 복원된 이미지가 원본과 얼마나 유사한지 평가한다. PSNR의 평가 방식은 MSE로 계산한 값을 log 스케일로 변환하여 계산한다. 30 dB 이상의 점수가 나오면 높은 품질이며 원본과 매우 유사하다고 판단된다. PSNR는 픽셀 간 차이만을 평가하기에, 사람들이 느끼는 시각적 품질을 반영하지는 못한다는 한계가 있다. 두 번째 지표는 SSIM이다. SSIM은 Structural Similarity Index Measure의 약자로, 이미지의 구조적 유사성을 평가한다. 인간의 시각 시스템을 고려하여 밝기, 대비, 구조 3가지 요소를 비교한다. SSIM 지표의 결과값은 0에서 1사이이며, 1에 가까울 수록 원본과 복원 이미지가 구조적으로 유사하다고

판단된다. 이는 PSNR에 비해 시각적 품질을 더 반영한다.

학습 이후에 PSNR로 지표를 출력시켜본 결과 최대 18.37 dB이며 평균적으로는 15.63 dB의 점수를 얻었다. SSIM 지표를 출력시켜본 결과 최대 0.4604이며 평균적으로는 0.2181의 점수를 얻었다.

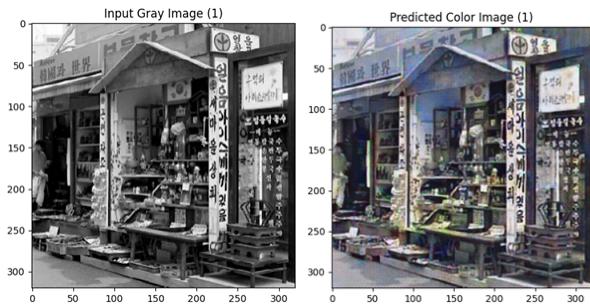




위 사진들은 최종 모델링의 결과이며, 시각적으로 보기에 상당히 많이 컬러화가 진행되었으며 여러 모델링을 거쳐본 결과 가장 좋은 성능을 가지게 되었다.

3. 테스트

이 프로젝트의 목적은 과거의 사진을 컬러화하여 사회적 가치를 창출하는 것이다. 이를 위해 과거의 이미지를 320x320 px로 조정한 후 최종 모델에 입력해 보았다.



그 결과, 생성된 이미지는 흑백 사진을 상당히 높은 성능으로 컬러화한 것으로 나타났다. 이는 훈련 세트에 포함되지 않은 이미지도 효과적으로 컬러화할 수 있다는 점에서 큰 의미가 있다.

4. 한계 및 개선할 점

최종 모델링은 사람에 대해 컬러화가 제대로 진행되지 않아 사람이 없는 Landscape 데이터셋을 이용했다. 풍경이 아닌 사람이 있는 사진들은 성능이 떨어지고 완성도 높은 컬러화가 진행되지 않는다는 한계점이 존재한다.

추후에, 사람 위주의 흑백 이미지와 그에 대응되는 컬러이미지 데이터셋을 이용하여 사람이 있는 사진도 성공적인 컬러화를 진행할 수 있도록 모델을 다시 개발하고 관련 내용을 연구할 것이다.

또, 컬러화 모델을 바탕으로 흑백 동영상은 컬러 동영상으로 변환하는 모델까지 구축해보는 것이 목표이다.

4. 참고문헌

Shravan Kumar. *Image Colorization*. Kaggle, 2022.

<https://www.kaggle.com/datasets/shravankumar9892/image-colorization>.

salimhammadi07. "Pix2Pix : Image Colorization with Conditional WGAN." Kaggle, May 14, 2023.
<https://www.kaggle.com/code/salimhammadi07/pix2pix-image-colorization-with-conditional-wgan>.

Effectivemadness.

"Effectivemadness/Colorization_pix2pix: Image Colorization Using Pix2pix." GitHub. Accessed December 14, 2024.
https://github.com/effectivemadness/colorization_pix2pix.

TheBlackMamba31. *Landscape Image Colorization*. Kaggle, 2023.
<https://www.kaggle.com/datasets/theblackmamba31/landscape-image-colorization>.

GAN(Generative Adversarial Network) , githurb , MIKE CHU, <https://mike7chu.github.io/GAN/>.

Reimagining History in Color with Pix2Pix as an Assistant Creatives and AI collaborating: how a trained pix2pix model can help with colorization , Medium, Tina Tang,
<https://towardsdatascience.com/reimagining-history-in-color-with-pix2pix-as-an-assistant-23d6e4a64e75>.

U-Net 논문 리뷰 — U-Net: Convolutional Networks for Biomedical Image Segmentation

딥러닝 기반 OCR 스터디 — U-Net 논문 리뷰, Medium, 강준영,
<https://medium.com/@msmapark2/u-net-%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%BD-u-net-convolutional-networks-for-biomedical-image-segmentation-456d6901b28a>.

"Korea Art TV." 디지털화된 고전: 사진 컬러화의 예술적 가치. Accessed June 16, 2024.
<http://www.koreaarttv.com/detail.php?number=769>.