

# MLP 기반의 GAN 을 사용한 흑백 사진 채색 기법

왕철, 조인휘\*

한양대학교 컴퓨터소프트웨어학과

\*한양대학교 컴퓨터소프트웨어학과

e-mail : wangzhe10101@hanyang.ac.kr, \*iwjoe@hanyang.ac.kr

## Colorful Image Colorization using GAN with MLP

Zhe Wang, Inwhhee Joe\*

Dept. of Computer software, HanYang University

\*Dept. of Computer software, HanYang University

### 요 약

본 논문에서 grayscale 이미지를 그럴듯한 컬러 이미지로의 전환을 다루고자 한다. 기존의 CNN Network 를 통해 실제 Image 를 만들어내려는 기법들은 모든 Pixel 의 Error 를 Loss 로 사용한다. 각 픽셀별로 가장 완벽한 답을 찾으려고 하기보다는, 전체 픽셀의 관점에서의 Loss 를 줄이려고 하기 때문에, 픽셀 값이 정확한 값대신 안전한 값으로 넘어간다는 단점이 있다. 이 문제를 해결하기 위해 본 논문에서 GAN 기반의 Image-to-Image Translation 기법에 NIN(Network in Network) 적용해 이 문제를 해결할 수 있음을 보인다. 전통 CNN 기법보다 더 Photo-realistic 한 이미지를 생성할 수 있게 된다.

### 1. 서론

Auto colorization 과 같이 기존의 입력 영상으로부터 새로운 영상을 생성해내는 Image-to-Image Translation, Style Transfer 등 연구에는 CNN(Convolution Neural Network)기반의 구조를 사용한 딥러닝 기반의 기법들이 사용되고 있다. CNN 은 컴퓨터 비전의 다양한 분야에 사용되어 높은 성능 향상을 보여주고 있다. 특히 GAN(Generative Adversarial Network)의 개념이 등장으로 생성 영상의 성능이 큰 폭으로 향상되었다. 본 논문에서 GAN 기반의 Image-to-Image Translation 기술을 적용해 좋은 성능의 자동 colorization 결과를 확인하였다. 실험에서는 학습을 위한 베이스라인 신경망으로 Pix2PixGAN 네트워크를 사용한다. Colorization 문제에 맞추어 최적화하기 위해 추가적으로 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks 제안되는 Identity Loss 와 PatchGAN 그리고 LSGAN 의 핵심 아이디어로 베이스라인 GAN 신경망의 구조를 개선했다. 더 정확한 feedback 을 얻기 위해 GAN 에 적합한 MLP(Multi-Layer Perceptron)를 Discriminator 에 적용해 Discriminator 의 구조를 구축했다. 이를 통해 보다 안정적인 좋은 성능의 colorization translation 결과를 얻었다.

### 2. 관련 연구

#### 2.1 Pix2PixGAN:

Pix2PixGAN 는 대칭 데이터의 가공이며 영상의 스타일 변환을 위한 cGAN 기반의 일반화된 구조이다. 기존의 GAN (또는 DCGAN)에서는 Noise Distribution 으로부터 Data Distribution 을 뽑아내는 Learning 을 하게 된다면, 여기서는 한 Image Domain (ex. Laebl Image)로부터 또 다른 Image Domain (ex. Synthetic Image)로부터의 Mapping Function 을 Learning 을 하게 되고, 여기서부터 Discriminator 를 통해서 진짜인지 아닌지를 검사를 받게 되는 것이다. CNN 을 통한 Loss Function 외에도, GAN Loss 를 추가하게 되면 어느 정도 Photo-realistic 한 Image 를 만들어 낼 수 있게 되었다. L1(reconstruction loss)와 cGAN(adversarial)를 같이 사용하는 것이 Pix2PixGAN 의 주요 아이디어라고 할 수 있다.

#### 2.2 Identity loss

비대칭 학습 데이터를 기반으로 한 CycleGAN 에서 Identity loss 라는 개념을 도입했다. X 에서 Y 로 Transfer 를 할 때 Loss 에다가 Y 에서 Y 로 가는 Mapping Function 의 Loss 를 추가한 것이다. 진짜 Y Domain 의 Image 가 들어 왔을 경우에도 아무것도 하지 않아서 Y Domain 의 특징을 유지하도록 만들면, X 에서 Y 로 가는 Mapping Function 이 단순히 Transfer 를 생각 없이 하는 것이 아니라 X 의 형태를 좀 더 깊게 보게 된다는 것이다. 즉, X 일 때는 Y 로 바꾸고 Y 일 때는 냅뒀 라는 것을 배움으로서 Mapping 의 특

정을 좀 더 유의미하게 만드는 것이다.

2.3 PatchGAN

GAN의 Discriminator가 전체 영역이 아니라, 특정 크기의 patch 단위로 진짜/가짜를 판별하고, 그 결과에 평균을 취하는 방식이 PatchGAN이다. 픽셀들 간의 연관성은 거리에 비례하여 작아지는 경향이 있으며, 특정 크기의 patch에 대하여 진짜 같은 이미지를 생성할 수 있다면, 그리고 그런 patch의 수가 많아지는 방향으로 학습을 하게 된다면, generator의 성능은 더 올라갈 수 있다.

2.4 LSGAN

LSGAN은 Discriminator에 sigmoid cross entropy loss 대신 least square loss를 사용해서 decision boundary에서 멀리 있는 sample들에게 penalty를 주는 것입니다. sample이 좀 더 real data에 가깝게 나오도록 generator를 학습할 수 있게 됩니다.

2.5 MLP(Multi-Layer Perceptron)

일반적인 CNN 구조는 local receptive field에서 feature를 추출해내는 능력은 우수하지만 non-linear한 feature를 추출하기엔 어려움이 있다. 이 부분을 극복하기 위해 feature-map의 개수를 늘리면서 연산량이 늘어나지 않은 구조가 필요하다. 그래서 Convolution을 수행하기 위한 filter 대신에 MLP(Multi-Layer-Perceptron)를 사용하여 feature를 추출하도록 하였다. Fully-connected NN 대신에 최종단에 global average pooling을 사용하였다. 연산량이 늘어나지 않으면서 non-linear한 성질을 잘 활용할 수 있게 되었다.

3. 제안 방법

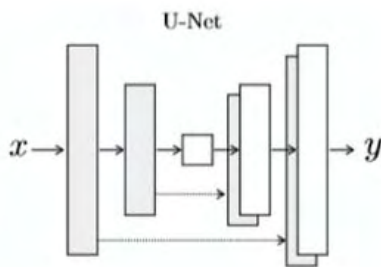
3.1 실험 데이터

학습에 앞서, gray-color대칭된 충분한 데이터를 위해 데이터 가공을 자동화했다. Place365 데이터셋을 사용해 128\*128의 gray-color RGB 훈련 이미지를 생성했다.

3.2 네트워크

3.2.1 Generator 구조

Generator에는 U-net 구조를 사용했다. U-net은 encoder-decoder 구조에서 영상크기를 줄였다가 다시 키우는 과정에서 detail이 사라지면서 영상이 blur해지는 문제를 피하기 위해서 위 그림처럼 skip-connection을 갖는다. encoder에서 decoder쪽으로 직접 정보를 넘기기 때문에 훨씬 선명한 결과를 얻을 수 있다.

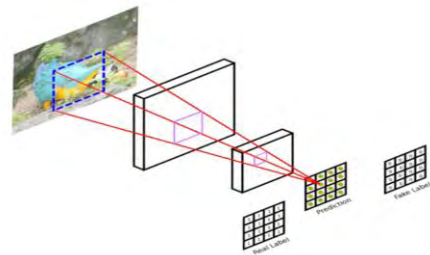


(그림 1) U-net은 2015년 Olaf Ronneberger 팀이

“U-net: Convolution Network for Biomedical Image Segmentation” 논문으로 발표하였다.

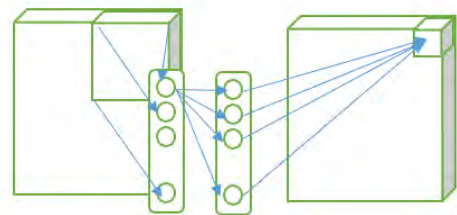
3.2.2 Discriminator 구조

Discriminator에서 PatchGAN을 적용했다. 기존의 GAN에서 Discriminator의 역할은 Image 전체를 보고 진짜인지 가짜인지를 판별하게 되는데 이것을 Image의 Overlap되는 Patch 단위로 해보자는 것이다. 그렇게 되면 Patch 단위로 Loss가 Back-propagate되어서 좀 더 Detail한 부분에서 Generator가 Feedback을 받을 수 있게 되었다.



(그림 2) PatchGAN

Convolution layer와 Batch normalization layer를 번갈아 사용하는 일반적인 GAN의 구조는 filter의 특징이 linear하기 때문에 non-linear한 성질을 가진 feature를 추출하기 어렵다. 이 문제를 해결하기 위해 convolution layer에서 MLP(Multi-Layer-Perceptron)를 사용하여 feature를 추출하도록 하였다, 그 구조는 아래 그림과 같다.



(그림 3) MLP(Multi-Layer-Perceptron) 구조

1\*1 convolution을 사용하여 feature-map을 줄일 수 있도록 하였다. 망을 깊게 만들기 위해, MLP layer를 여러 개를 쌓아 사용을 했다. MLP layer마다 뒤에 Batch normalization layer, ReLU function, dropout layer를 추가했다.

3.3 loss 함수

더 안정적인 학습과 빠른 수렴을 위해 LSGAN의 아이디어를 선택했다. 기존의 GAN은 아래와 같은 minimax problem을 푸는 반면

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

LSGAN은 다음과 같이 변형된 문제를 풀게 된다.

$$\min_D V_{LSGAN}(D) = \frac{1}{2} \mathbb{E}_{x \sim p_{data}(x)} [(D(x) - b)^2] + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2] \quad (1)$$

$$\min_G V_{LSGAN}(G) = \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - c)^2] \quad (2)$$

여기서  $b$ 는 real label  $a$ 는 fake label  $c$ 는  $G$  입장에서  $D$ 가 믿도록 하고 싶은 값입니다. Generator 더 정확한 목표 이미지를 생성하도록 두 가지 loss 항을 추가했습니다. Image-to-Image Translation with Conditional Adversarial Networks에서 제안한 Reconstruction Loss, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks에서 제안한 Identity Loss가 있다. 이 논문에서 GAN의 학습과정에서 L1 loss function은 가이드의 역할을 해서 L1 loss가 많으면 많을수록 좋다는 실험결과가 나왔다.

### 3.4 학습

총 1000 회의 대칭 데이터를 학습에 사용했으며 이 중 학습에 포함되지 않은 100 장의 이미지를 검증용 데이터로 사용했다. 영상을 128\*128으로 리사이즈 후 grayscale 해서 데이터셋을 만들었다.

최적화 알고리즘으로 Adam Optimizer를 사용했다. 학습률은 2e-4로 설정했으며, Generator과 Discriminator의 convolution layer에서 LeakyReLU activation functions, Batch Normalization, dropout을 적용했다. 100 epoch를 학습시켰다.

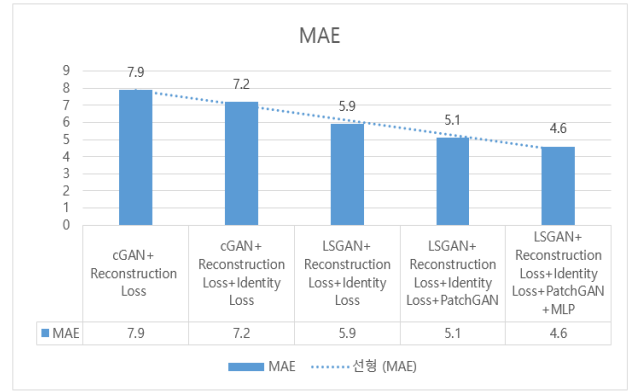
## 4. 실험 결과

### 4.1 평가방법

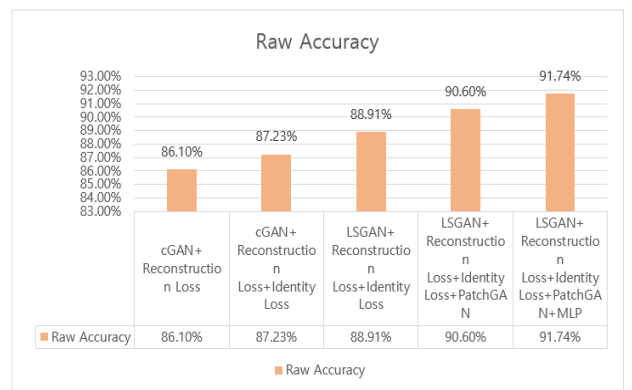
성능을 측정하기 위해 Image Colorization using Generative Adversarial Networks 논문에서 제안한 mean absolute error(MAE)과 Raw accuracy(AuC)를 채택했다. MAE는 각 색상 채널에 대해 픽셀 수준에서 절대 값의 평균을 취하여 계산된다. Raw accuracy(AuC)는 생성된 이미지와 원본 이미지의 픽셀 중에 같은 컬러 정보를 가진 픽셀의 비율을 계산한다. 두 color channel 간에 거리가 설정한 threshold distance보다 적으면 같은 정보를 가졌다고 간주한다.

### 4.2 결과

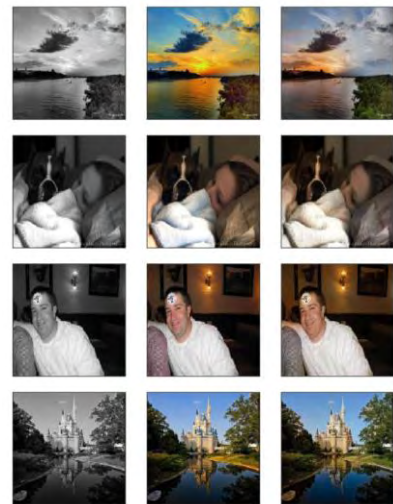
평가에서 같은 데이터셋으로 각 network의 성능을 비교했다. 본 논문에서 사용한 구조를 가장 좋은 결과를 얻은 것으로 보인다. LSGAN의 stable한 특성상, 빠른 수렴, Reconstruction Loss가 GAN을 원하는 목표 이미지를 생성하도록 하고, Identity Loss가 Mapping의 특징을 좀 더 의미미하게 만들고, MLP가 이미지의 non-linear한 feature를 추출해 Generator에게 더 정확한 feedback을 전달함으로써 colorization 문제에서 이 조합을 가장 좋은 성능을 얻을 수 있음을 실험을 통해 증명했다.



(그림 4) 결과 비교 1



(그림 5) 결과 비교 2



(그림 6) 실험결과 맨 오른쪽 LSGAN+ Reconstruction Loss+Identity Loss+PatchGAN+MLP으로 생성 결과 이미지

### 참고문헌

- [1] Image-to-Image Translation with Conditional Adversarial Networks, Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, CVPR 2017
- [2] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros, ICCV 2017
- [3] Image Colorization using Generative Adversarial Networks Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi
- [4] Least Squares Generative Adversarial Networks Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, and Zhen Wang
- [5] U-Net: Convolutional Networks for Biomedical Image Segmentation Olaf Ronneberger, Philipp Fischer, and Thomas Brox
- [6] Network In Network Min Lin, Qiang Chen, Shuicheng Yan