# CycleGAN과 Mask R-CNN을 활용한 수채화 스타일 이모티콘 생성

#박소영, #한지안, 최지원, 김도연, *황효석
가천대학교 소프트웨어학과
e-mail : soyoung98, hjw705, gstar1106, sso06034@ gc.gachon.ac.kr,
hshwang@gachon.ac.kr
#contributed equally, *corresponding author

Watercolor Emoticon Generation using CycleGAN and Mask R-CNN

#Soyoung Park, #Jian Han, Jiwon Kim, Doyeon Kim, *Hyoseok Hwang
Department of Software
Gachon University

## Abstract

In this paper, we discuss the process which makes watercolor style emoticon from real dog image using CycleGAN and Mask R-CNN. Our method is proposed into image-to-image translation step and image segmentation step. In image-to-image translation step, we translate the real dog images into watercolor style images by a trained CycleGAN model using our datasets. In image segmentation step, using Detectron2 which is based on Mask R-CNN, we removed the background of translated image and only the target object is extracted from the image. For training our model, we build our dataset, which is an unpaired dataset. The dataset contains 684 real dog images and 419 watercolor images. To evaluate the performance of CycleGAN in image generation, we used 'Fréchet Inception Distance'(FID).

## I. 서론

반려동물을 양육하는 인구가 늘어나면서 반려동물 관련 사업 또한 점차 커져 가고 있는 추세이다. 반려동물 중에서도 강아지가 가장 많은 비율을 차지함으로 본인의 강아지에 맞춤화 된 이미지를 생성하는 기술을 구현하였다. 기존의 인공지능 기술은 GAN(Generative Adversarial Network)[1] 을 활용해 사람의 이미지를 맞춤화하여 다양한 버전의 일러스트나 이모티콘으로 생성하는 기술이 이미 구현된 적이 있다[2]. 실제 사람 사진에 만화(Cartoon) 스타일을 적용하여 이미지를 변환하는 경우가 있으며, 자연 풍경을 다양한 명화 스타일에 적용하는 경우가 있다[3]. 기존의 이미지 변환 연구는 대부분 사람을 대상으로 하였기 때문에 반려동물에 직접 적용시키기는 어렵다.

본 논문에서는 GAN을 활용하여 강아지 이미지의 특징을 살려 이미지를 변환한 후 세그멘테이션하여 이모티콘으로 만드는 프로세스를 제안하고자 한다.

## II. 관련 연구

### 2.1 이미지 변환

최근 이미지 변환은 Generative Adversarial Networks (GAN)[4]를 기반으로 다양하게 발전해왔다. Pix2Pix[5]는 이미지 생성 네트워크로, pair한 input 이미지, output 이미지의 데이터를 사용한다. 생성된 이미지가 label이라 할 수 있는 output 이미지와의 차이를 측정하여 이미지를 변환하는데 중점을 두었다. 그러나 Pix2Pix는 pair한 이미지를 대량으로 구하기 어

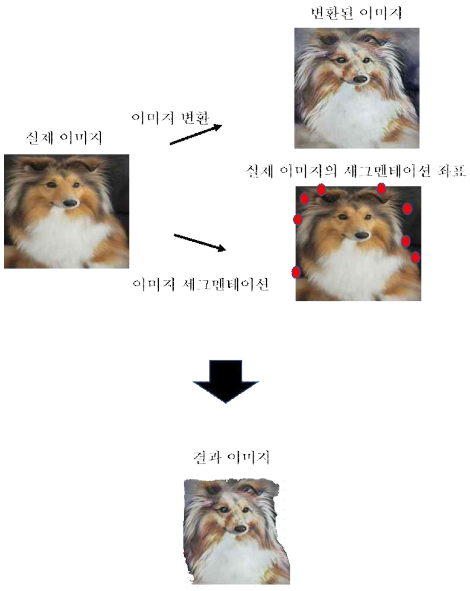


그림 1. 이미지 변환 단계: 실제 강아지 이미지를 수채화풍 이미지로 변환, 이미지 세그멘테이션: 변환된 이미지에서 객체의 영역을 제외하고 배경을 지움

렵다는 단점이 존재하였다.

CycleGAN[3]은 두 개의 생성자와 판별자를 사용하여, X에서 Y로 가는 것과 Y에서 X로 가는 양방향의 이미지 생성이 가능한 모델이다. 따라서, CycleGAN은 mode collapse를 해결할 수 있다. Mode collapse[3]는 input과 output이 의미 있게 짝지어지지 않고 서로 별개의 성향을 띄는 것을 의미한다. 어떠한 input 이미지이던 모두 같은 output 이미지로 매핑하면서 최적화에 실패하는 것이다. CycleGAN은 생성된 이미지의 복원을 이용한 학습을 제안하였으며, 해당 방법을 통해 pair한 이미지만을 트레이닝 데이터로 사용할 수 있는 Pix2Pix와 다르게 unpair한 이미지를 트레이닝 데이터로 사용할 수 있다.

### 2.2 이미지 세그멘테이션

이미지 세그멘테이션 기법은 의료, 보안, 자동차 관련 분야와 같은 다양한 분야에서 영상에서의 특정 객체를 추출하는데 사용된다. 이미지 세그멘테이션을 위해 여러 알고리즘들이 적용이 되어왔다.

Watershed[6] 알고리즘은 이미지를 회색조로 변환한 후 경계선을 만들어 이미지의 배경과 추출하고자 하는 객체 영역을 구분하여 세그멘테이션한다. 영역 사이의 정확한 경계선을 찾아 세그멘테이션의 성능을 높이는 것은 Watershed 알고리즘의 지속적인 과제였다.

Graph Cut[7]은 사용자가 직접 이미지에서 직사각형

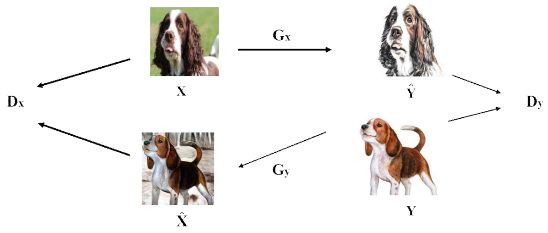


그림 2. CycleGAN 실행의 전체구조

으로 객체의 범위를 선정하여 세그멘테이션을 진행할 수 있도록 해주는 알고리즘이다. 반복적으로 앞 전경과 뒤 전경을 분리를 하는 작업을 하며 세그멘테이션의 성능을 높인다. 하지만 그래프 컷 알고리즘은 반복적으로 앞선 작업을 수행한다는 점에서 시간이 오래 걸려 비효율적이다.

Mask R-CNN[8] 알고리즘은 기존의 Faster R-CNN[9]의 객체를 인지하는 기능에 세그멘테이션을 수행할 수 있는 기능이 추가가 되어 나온 알고리즘이다. 자동으로 객체를 인지한 후, 객체가 같은 클래스여도 개별적으로 객체를 인지한다. 객체의 영역은 인스턴스 세그멘테이션을 통해 분할을 진행한다. 우리는 효율성과 정확성을 위해 Mask R-CNN 알고리즘을 이용하여 세그멘테이션을 진행하였다.

## III. 본론

본 논문에서는 이미지 변환과 이미지 세그멘테이션을 통하여, 사용자의 강아지 사진을 수채화풍으로 변환하는 프로세스를 제시하고자 한다. 그림 1에서 볼 수 있듯이, 실제 이미지에서 이미지 세그멘테이션을 진행하여 감지된 객체 중 강아지 객체를 추출해 내기 위한 픽셀 정보를 추출한다. 이 후 이미지 변환을 통해서, 실제 이미지를 수채화풍으로 변환한다. 마지막으로 이미지 세그멘테이션 단계에서 추출한 픽셀 정보를 이용해 변환된 이미지에서 강아지 객체의 영역을 추출한다.

### 3.1 이미지 변환 단계

이미지 변환 단계에서는 CycleGAN을 사용한다. CycleGAN은 두 개의 생성자와 판별자를 사용하여, 양방향의 이미지 생성 모델을 만든다. 본 논문에서는 실제 강아지 이미지를 X, 수채화풍의 강아지 일러스트 이미지를 Y로 하여 CycleGAN 모델을 학습한다.

그림 2에서 볼 수 있듯이, 우리의 모델은 X → Y, Y → X 두 개의 생성자와 판별자를 갖는다. Dy는 Gx로부터 학습된 이미지와 Y를 구분하기 어렵게 학습되도

록 한다. 반대로, Dx는 Gy로부터 학습된 이미지와 X를 구분하기 어렵게 학습되도록 한다. 이러한 학습 방법은 궁극적으로 Gx와 Gy의 성능을 향상시킨다.

양방향 이미지 학습을 위해 CycleGAN은 Cycle consistency loss[3]를 사용한다. Cycle consistency loss는 생성된 이미지인 G(x)를 반대 방향으로 생성한 이미지인 F(G(x))에 원본 이미지인 x와의 차이를 최소화하는 방향으로 학습되며 다음의 수식을 따른다:

$$L_{cyc}(G,F) = E_{x \sim Pdata(x)}[\| F(G(x)) - x \|_1] + E_{y \sim Pdata(y)}[\| G(F(y) - y \|_1]. \quad (1)$$

즉, Cycle consistency loss를 사용하면 복원된 이미지가 실제 원본 이미지와 비슷하도록 학습되어 output 이미지가 input 이미지에 관련된 결과로 나오도록 유도된다.

3.2 이미지 세그멘테이션 단계

실제 강아지 데이터셋에서 강아지 객체를 추출하기 위해 이미지 세그멘테이션 단계를 진행한다. 이미지 세그멘테이션을 위해 실제 강아지 사진에 Mask R-CNN을 적용하여 강아지 객체의 위치 좌표를 저장한다. 마지막으로 저장한 좌표들을 이용해 이미지 변환 단계에서 생성한 이미지의 배경을 제거하여 강아지 객체를 추출한다.

Mask R-CNN은 객체의 영역을 지정해주는 bounding box에 각각의 픽셀이 객체인지 아닌지 마스킹하는 mask branch가 존재한다. 이를 통해 정밀하게 이미지 세그멘테이션을 진행한다. 추출된 mask 영역은 ROI Align[8] 기법이 사용되어 추출된 위치 정보가 왜곡되지 않도록 해준다. 모델 구조는 FPN(Feature Pyramid Network)[10]과 ResNet101로 구성되어 있다. 해당 구조는 낮은 수준에서 높은 수준까지의 semantic feature를 모두 가지기 때문에 더 효율적이고 정확하게 feature를 추출할 수 있다.

## Ⅳ. 실험 및 향후 연구 방향

본 논문에서는 직접 수집한 데이터셋을 이용하여 CycleGAN을 학습한다. 실험에는 Intel i7-5960X CPU, Nvidia GTX 1080 GPU가 사용되었고 epoch은 400으로 설정하였다. 데이터셋은 실험을 위해 직접 생성하였으며, unpair한 데이터로 실제 강아지 이미지가 684개, 수채화풍 강아지 일러스트 이미지가 419개이다. CycleGAN과 Mask R-CNN 알고리즘 기반의 이미지 세그멘테이션 시스템인 Detectron2[12]를 통해서 결과를 내었고 그림 3에서 확인할 수 있다. 그림에서 볼 수 있는 바와 같이 실제 강아지 이미지를 수채화풍 이미지로 변환하고 배경을 제거하여 이모티콘을 완성하였다.

정량적인 성능 평가를 위해서는 Fréchet Inception Distance(FID)[11]를 추가로 사용하였다. FID는 실제 이미지와 생성된 이미지의 유사도를 측정하는 방법이다. FID는 Inception Network v3를 이용하여 추출한 각 이미지의 feaure를 통해 유사도를 측정한 값이다. FID는 이미지 feature의 mu와 sigma를 사용하여 이미지간 거리를 측정한다. 각 이미지에서 Inception Network v3를 통해 얻어낸 feature에서 평균과 공분산 값을 사용하여 연산한 값이며 다음의 수식을 따른다:

$$d^2((m, C), (m_w, C_w)) = \| m - m_w \|_2^2 + Tr(C + C_w - 2(CC_w)^{1/2}). \quad (2)$$

실험의 순서는 두 가지 경우를 시도해 보았다. 이미지 변환 후 배경을 제거한 경우가 FID값이 약 66.113으로 76.210인 배경 제거 후 이미지를 변환한 경우보다 값이 작아 최종 모델로 선정하였다. 실험 결과는 표1에 요약되어 있다.

| Process | FID |
|---|---|
| 이미지 변환 → 배경 제거 | 66.113 |
| 배경 제거 → 이미지 변환 | 76.210 |

표 1. 실험의 순서에 따른 FID 값

본 논문에서는 직접 생성한 데이터셋을 사용하여 실제 강아지 이미지를 기반으로 수채화풍 이모티콘을 생성하여 프로세스를 완성했다. 본 연구에서는 CycleGAN과 Detectron2를 활용하여 이미지를 변환하고, 세그멘테이션을 진행하였다.

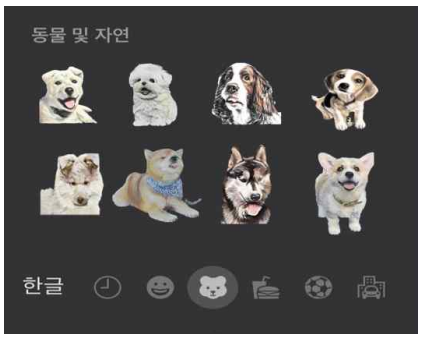


그림 4. 이모티콘 생성 및 사용 예시

본 실험을 통해서 우리가 기대하는 효과는 다음과 같

다. 기존의 모바일 서비스에서 이모티콘은 정해진 형태로 나와 있어 사용자가 그 중에서 고르고, 사용하는 방식이었다. 그러나 우리는 사용자가 직접 이모티콘화 하고자 하는 대상을 고르고, 이를 반영할 수 있다는 점에서 맞춤형 서비스로 의의가 있다. 그림 4와 같이 모바일 등 다양한 플랫폼에 적용하여 많은 사용자를 끌어들일 수 있는 서비스로 발전시킬 예정이다.

## Ⅴ. 결론

이전의 CycleGAN 연구들은 풍경이나 사람의 이미지를 변환하는 것에 집중되어 있었다. 본 논문에서는 강아지의 이미지를 변환하는 것에 최적화된 모델을 만들고자 하였다. 또한 이모티콘을 생성한다는 목적에 맞게 이미지 변환에 세그멘테이션을 접목하여 실제 이모티콘과 같은 결과 이미지를 생성하였다. 이번 연구는 강아지, 고양이와 같은 동물과 나아가 이외의 다른 여러 가지 객체에도 적용된다면 사용자가 원하는 이미지를 이모티콘화 하는 플랫폼을 만들 수 있을 것으로 전망된다.

## 참고문헌


[1] I. Goodfellow et al, Generative adversarial nets, Advances in neural information processing systems. 2014.

[2] R. Wu et al. Landmark Assisted CycleGAN for Cartoon Face Generation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul, 2019.

[3] J.Y. Zhu et al, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Oct, 2017.

[4] I. Goodfellow et al, Generative Adversarial Nets, Curran Associates, Inc., 2014.

[5] P. Isola, et al, Image-To-Image Translation With Conditional Adversarial Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[6] L. Vincent, and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations.", IEEE Transactions on Pattern Analysis & Machine Intelligence 6, 1991.

[7] C. Rother et al, "GrabCut":interactive foreground extraction using iterated graph cuts, ACM transactions on graphics (TOG), August, 2004.

[8] K. He et al, Mask R-CNN, Proceedings of the IEEE International Conference on Computer Vision (ICCV), October, 2017.

[9] S. Ren et al, Faster r-cnn: Towards real-time object detection with region proposal networks, In Advances in neural information processing systems, 2015.

[10] T.-Y. Lin et al, Feature pyramid networks for object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[11] M. Heusel et al, Gans trained by a two time-scale update rule converge to a local nash equilibrium, Advances in neural information processing systems, 2017.

[12] https://research.fb.com/downloads/detectron/