딥러닝 응용현황

딥러닝 실제 - 3주차 과제

목차

[객체 탐지 (Object Detection) 3](#_Toc67233512)

[Segmentation 4](#_Toc67233513)

[자연어처리 (Text) 5](#_Toc67233514)

[강화학습 (Reinforcement Learning) 6](#_Toc67233515)

[Generative Adversarial Networks (GAN) 7](#_Toc67233516)

[Style Transfer 8](#_Toc67233517)

[Deep Photo Style Transfer 9](#_Toc67233518)

[StarGAN 10](#_Toc67233519)

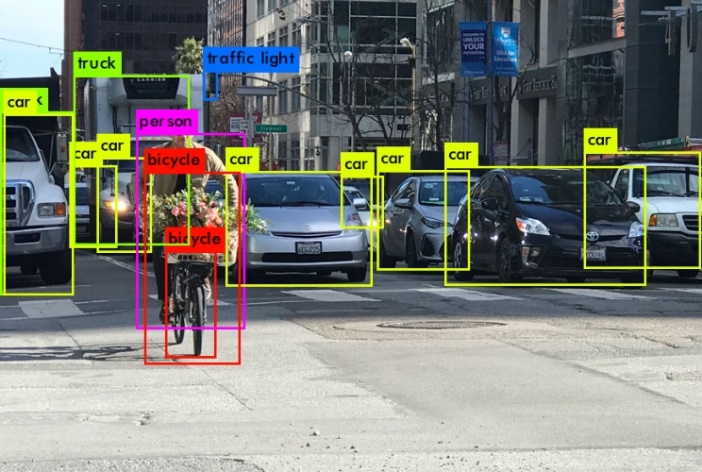
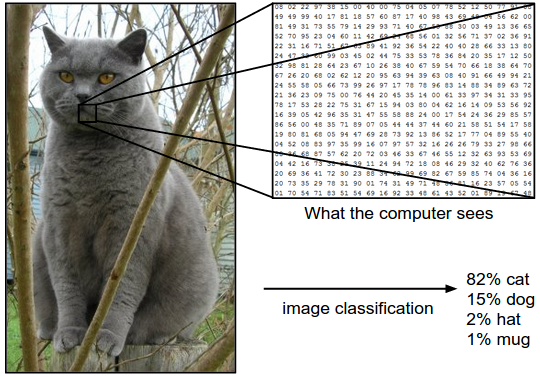
[Creative Adversarial Networks 11](#_Toc67233520)

[SRGAN 12](#_Toc67233521)

[Image Completion 13](#_Toc67233522)

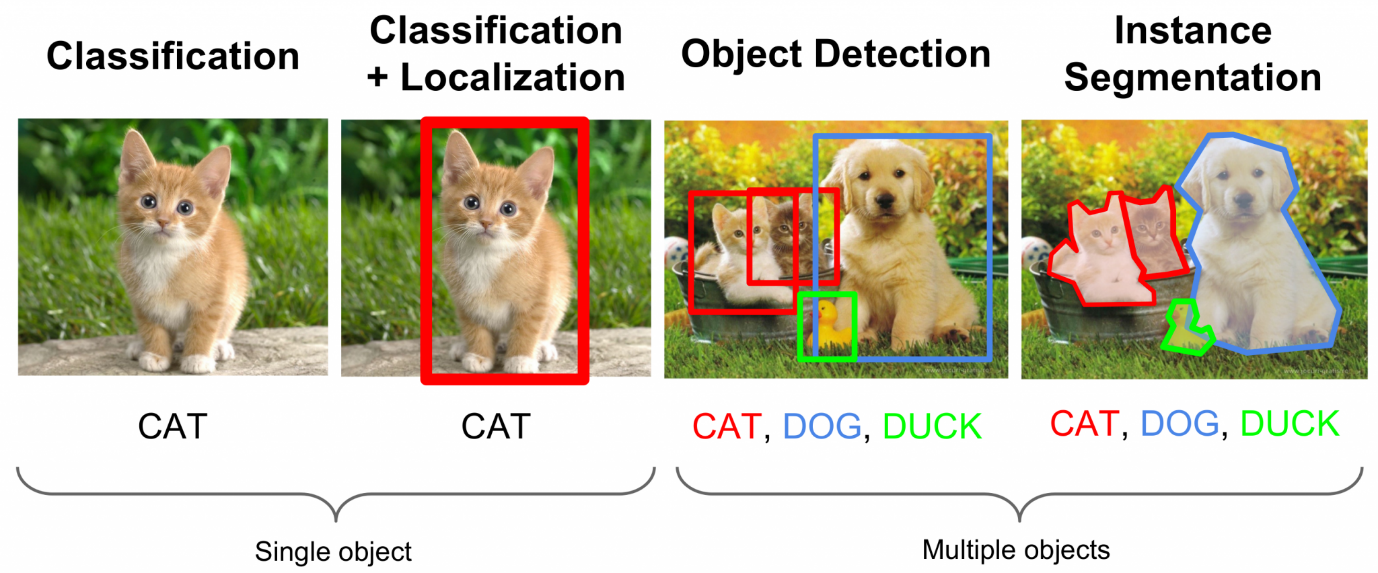
[Deepfake 14](#_Toc67233523)

# 객체 탐지 (Object Detection)



객체 탐지는 어떠한 이미지 및 비디오 속에 포함 되어있는 물체들에 대해서 어떤 물체인지 분류해내는 이미지 분류를 하는 것과 동시에 해당 물체가 이미지 및 비디오 속에 어디에 위치하였는지 찾아내는 일입니다. 다량의 이미지 및 비디오 데이터를 활용하며, 이미지 및 비디오 내 특정 물체의 위치 정보를 X, Y 좌표 값과 해당 물체의 크기인 Width, Height 값을 레이블 정보로 이용하여 딥러닝 모델이 학습합니다. 연구자들 사이에서는 (X, Y, W, H) 정보를 보통 Bounding Box라고 많이 표현하고 있습니다. 객체 탐지 기술은 자율주행자동차, CCTV 등 카메라 기술을 바탕으로 이루어지고 있는 제품 및 서비스에 최근 도입되고 있으며, 현재 다양한 연구가 진행되고 있습니다.

# Segmentation

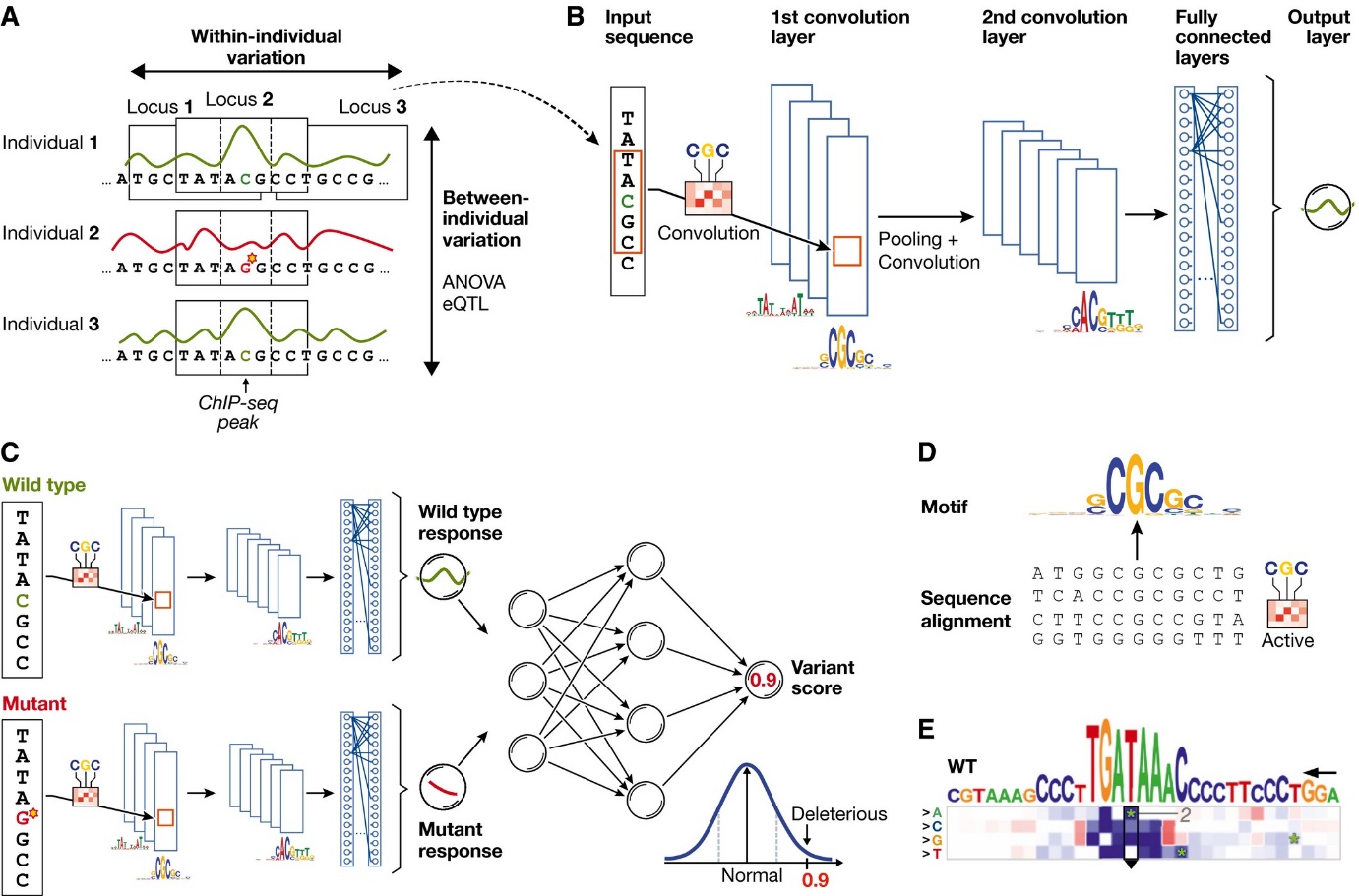


Segmentation은 앞서 Object Detection보다 정교한 탐지를 요구하는 연구 분야입니다. 앞서 제시한 Object Detection 연구 분야는 특정 물체를 사전에 정의된 특정 클래스로 분류하며 입력 값으로 이용된 이미지 혹은 비디오 화면 내 어느 위치에 존재하는지 파악하는 연구 분야입니다. 여기서, 어느 위치에 존재하는지 파악할 때, (x, y, w, h) 정보를 이용하여 Boundary Box를 그리게 되는데, 이 Boundary Box 내 물체가 꽉 차 있을 순 있지만, 해당 물체가 아닌 영역이 존재할 수 있습니다. Segmentation은 특정 위치에 Boundary Box로 물체 존재 유무로 표현하는 것의 한계를 극복하기 위해, 이미지 및 비디오 내 존재하는 모든 픽셀에 대해 특정 클래스로 예측하는 방식으로 진행합니다. 이미지 및 비디오를 표현하는 최소 단위인 픽셀 수준으로 접근하여, 각 픽셀별로 특정 클래스를 예측한다면, 예측된 정보를 바탕으로 픽셀에 색을 다르게 표현하여 이미지로 표현한다면, 이미지 및 비디오 내 모든 영역이 특정 클래스로 표현됩니다.

특정 클래스로 분류된 물체에 대해 (x, y, w, h) 로 위치 정보를 표현하여 Boundary Box를 그린 Object Detection 분야는 그림을 통해 확인할 수 있듯이 사람이 아닌 다른 영역도 Boundary Box가 포함하고 있습니다. 하지만, 이미지 내 존재하는 모든 픽셀에 대해 클래스로 분류한 후 클래스 별로 색을 다르게 표현하여 이미지로 형상화한 Semantic Segmentation을 보면, 사람의 위치에 해당하는 픽셀들은 분홍색, 사람이 아닌 위치의 픽셀들은 검정색으로 표현하여 이미지에 대해 물체의 클래스 및 위치 정보를 Object Detection에 비해 정확히 추출할 수 있는 장점을 갖고 있습니다.

하지만, 동일한 클래스로 분류된 복수의 물체들의 픽셀 값들이 바로 옆에 존재하여 연결되어 있는 경우, 여러 물체가 하나의 물체로 인식될 수 있는 문제가 존재합니다. 따라서, 각 물체 별로 해당되는 픽셀 값들을 구분하고, 구분된 픽셀 값들에 대해서 클래스를 예측하는 Instance Segmentation 방식의 연구분야도 존재합니다.

# 자연어처리 (Text)

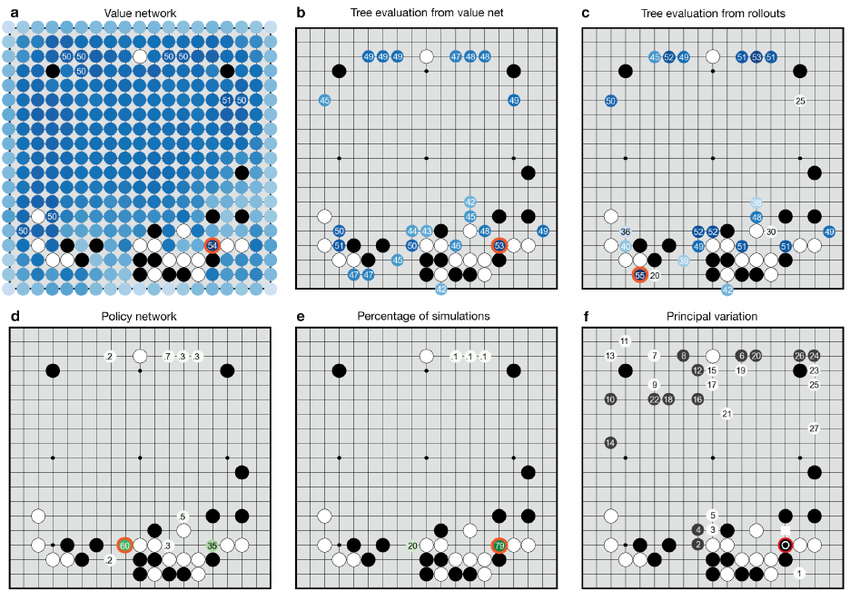


텍스트 분야에서의 딥러닝 적용 또한 꾸준히 연구되었습니다. 텍스트 분야는 세부 Task로 나뉘어 연구가 되었는데 대표적으로 다음과 같은 것들이 있습니다.

* 언어 번역 (Machine Translation)
* 문장(혹은 문서) 분류 (Sentence Classification)
* 질의 응답 시스템 (Question & Answer System: Q/A)
* 개체명 인식 (Named Entity Recognition: NER)

이미지 영역에서는 딥러닝 모델이 인간보다 더 좋은 성능을 보이는 모습을 보인 반면, 텍스트 분야에서는 인간의 성능을 따라잡기 어려웠습니다. 이미지에 비해 텍스트 Task는 배경지식이 요구된다는 점도 어려운 점이었고, 주로 사용한 Recurrent Neural Network (RNN) 계열의 모델의 한계 역시 해결해야 할 문제 중 하나였습니다. 하지만 2017년 구글이 발표한 Attention Is All You Need 라는 논문의 Transformer 모듈 연구를 시작으로 이와 관련된 모델들이 활발히 연구되기 시작했고, 인간의 성능을 넘어서는 Language Model이 개발되기 시작했습니다. 그리고 최근에는 이를 이용하여 학계에서는 다양한 분야의 추가 연구가, 산업계에서는 이와 관련된 서비스나 제품 연구가 활발히 이루어지고 있습니다.

# 강화학습 (Reinforcement Learning)



인공지능 바둑 프로그램 알파고는 구글 (Google)의 딥마인드 (DeepMind)가 개발한 인공지능 바둑기사로 2016년 한국의 이세돌 기사와 대국해 4승 1패로 승리하며 세상을 놀라게 하였습니다. 이 알파고의 기본 원리는 강화학습으로 현재 상태 (바둑판)에서 어떠한 행동 (수)을 취해야 먼 미래에 보상이 최대 (승리)가 될 것인지에 대해서 학습하는 알고리즘입니다. 이는 수많은 시뮬레이션을 해야 하지만 가능합니다. 하지만 바둑의 경우의 수는 무한대에 가깝기 때문에, 많은 전문가들이 인공지능이 바둑을 두기는 어려울 것이라 의견을 냈습니다. 그러나 딥마인드는 이 강화학습 알고리즘과 딥러닝을 통해 문제를 해결하기에 이르렀습니다. 처음에는 학습해야 할 경우의 수를 줄이기 위해 바둑의 기보를 학습하였으나, 2018년에는 알파고 제로 버전을 발표하면서 기보 없이 스스로 학습해 나가는 인공지능 바둑기사를 개발하였습니다. 알파고의 등장을 계기로 강화학습과 딥러닝을 결합한 심층 강화학습 (Deep Reinforcement Learning)의 연구가 활발히 진행되기 시작하였습니다.

# Generative Adversarial Networks (GAN)



딥러닝은 기본적으로 이미지와 텍스트 분류 쪽에 초점을 맞추어 발전해왔습니다. 그러면서 점점 더 다양한 분야, 다양한 방식으로 발전을 해왔습니다. 그러던 와중에 인공지능 분야에서 획기적인 생성 모델인 GAN이 등장했습니다. 대중들에게 알파고가 엄청난 각인을 끼쳤다면 GAN은 연구자들에게 엄청난 각인을 끼쳤습니다. GAN은 2014년 Ian Goodfellow의 박사논문으로 처음 제안되었고, 우리나라에서 본격적으로 알려 진 건 2016년 즈음입니다. 무언가 '예측'하는 걸 넘어서 '생성' 해내는 모델을 제안을 한 것입니다. 이미지 분류 모델을 예로 들면 Input데이터가 이미지이고 딥러닝 모델의 Output은 Label입니다. 이 이미지가 고양이인지 강아지 인지 분류하는 모델인 것이죠. 그러나 이 GAN은 input이 Random Noise이고 Output이 임의의 이미지입니다. 학습하는 대상에 따라서 숫자 이미지가 될 수도 있고, 사람 이미지가 될 수도 있습니다. 즉, 내가 숫자 이미지를 만들어내는 딥러닝 모델을 만들고자 한다면, Input Noise, Output을 숫자 이미지로 하여 GAN을 학습하면 되는 것입니다.

딥러닝 모델로 분류 외에 생성을 한다는 것은 많은 이들에게 충격으로 다가왔습니다. GAN의 등장은 알파고의 등장과 비슷하게 인공지능 연구의 새로운 패러다임을 제시하였습니다. 처음에는 일부 사람들이 생성되는 데이터의 질이 그렇게 좋지 않으므로 사용할 분야가 마땅치 않을 것이라 지적하였으나, 최근에 연구가 되고 있는 GAN의 성능은 이미 인간의 눈으로 구분하지 못 할 정도의 고품질의 이미지 (또는 텍스트)를 생성해냅니다. 아래 그림은 2018년 발표 BigGAN (Large Scale GAN Training for High Fidelity Natural Image Synthesis, Brock et al., 2018)이 생성한 이미지입니다. 즉, 이 세상에 실제로 존재하지 않는 객체에 대한 이미지라는 것이죠. 단순히 데이터를 생성해내는 것뿐만 아니라, GAN이 가지는 학습 알고리즘의 특성을 이용하여 다양한 분야로 발전되어 오고 있습니다.

# Style Transfer



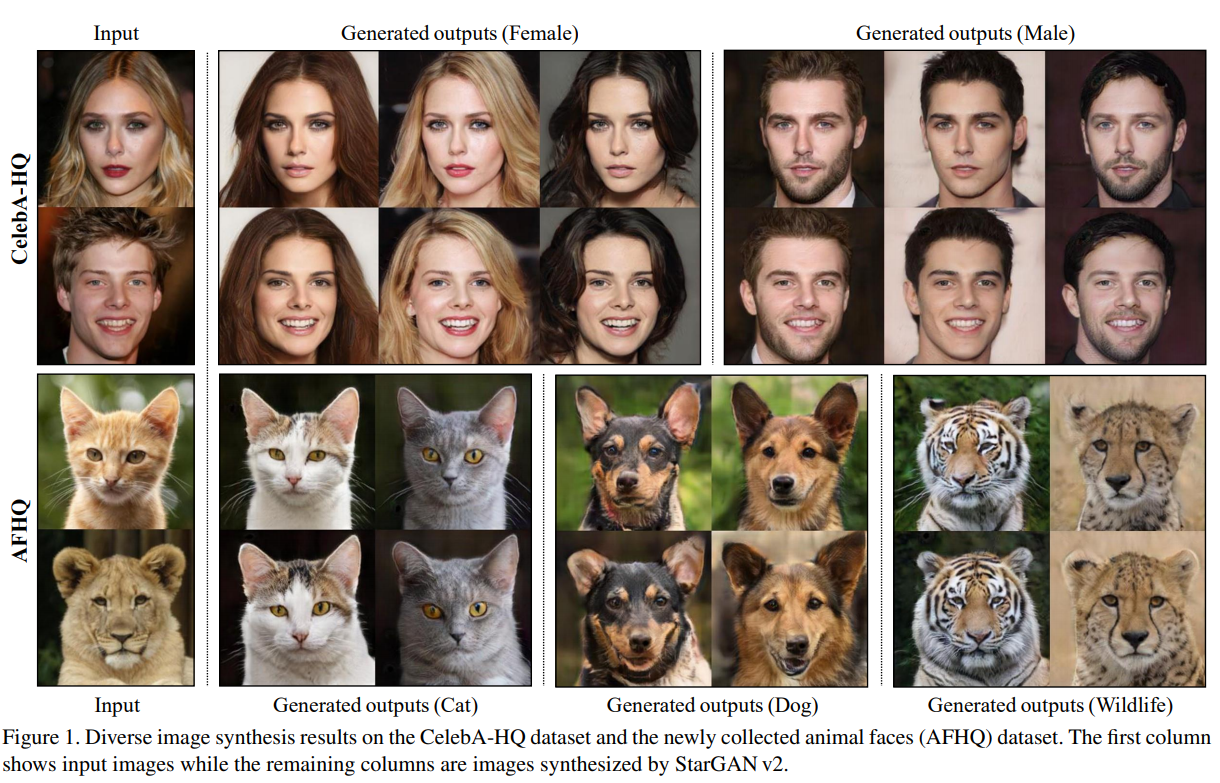
딥러닝이 발전하면서, 나의 사진을 고흐풍으로 바꿔주거나, 아래 그림처럼 낮 풍경의 사진을 밤 풍경의 사진으로 바꾸거나, 여름 풍경을 겨울 풍경으로 바꿔주는 것이 가능하게 되었습니다. 이러한 연구 분야를 Style Transfer라고 부릅니다. Style Transfer를 이용하면 굳이 포토샵을 이용할 필요가 없어지게 됩니다. GAN이 발전하게 되면서 Style Transfer에 GAN을 적용하기 시작하게 되었습니다 (CycleGAN: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Zhu et al., 2017). 여름 사진을 겨울 사진으로 바꾸고 싶을 때 CycleGAN의 Input은 여름사진, Output은 겨울사진이 됩니다. 물론, 반대로 겨울사진을 여름사진으로 바꿀 수도 있습니다. 이렇게 GAN의 등장은 기존의 학습 구조 (Input을 이미지로 Label을 Output으로 설정)를 넘어 다양한 학습구조를 가능하게 하였습니다.

# Deep Photo Style Transfer



CycleGAN은 Style Transfer 분야에서 가장 기본적으로 많이 쓰이는 모델인데 CycleGAN을 기반으로 하여 다양한 모델이 나오기 시작했습니다. Deep Photo Style Transfer (Luan et al., 2017) 는 아래 그림처럼 고해상도의 사진에 대하여 Transfer시키는 모델입니다.

# StarGAN



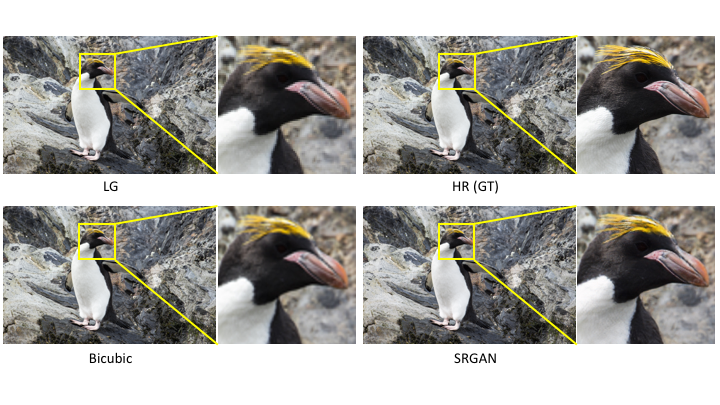
StarGAN (Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation, Choi et al., 2018) 은 아래 그림처럼 Input 이미지에 대하여 참조 이미지의 피부 표정 등을 Transfer시키는 GAN 모형입니다

# Creative Adversarial Networks



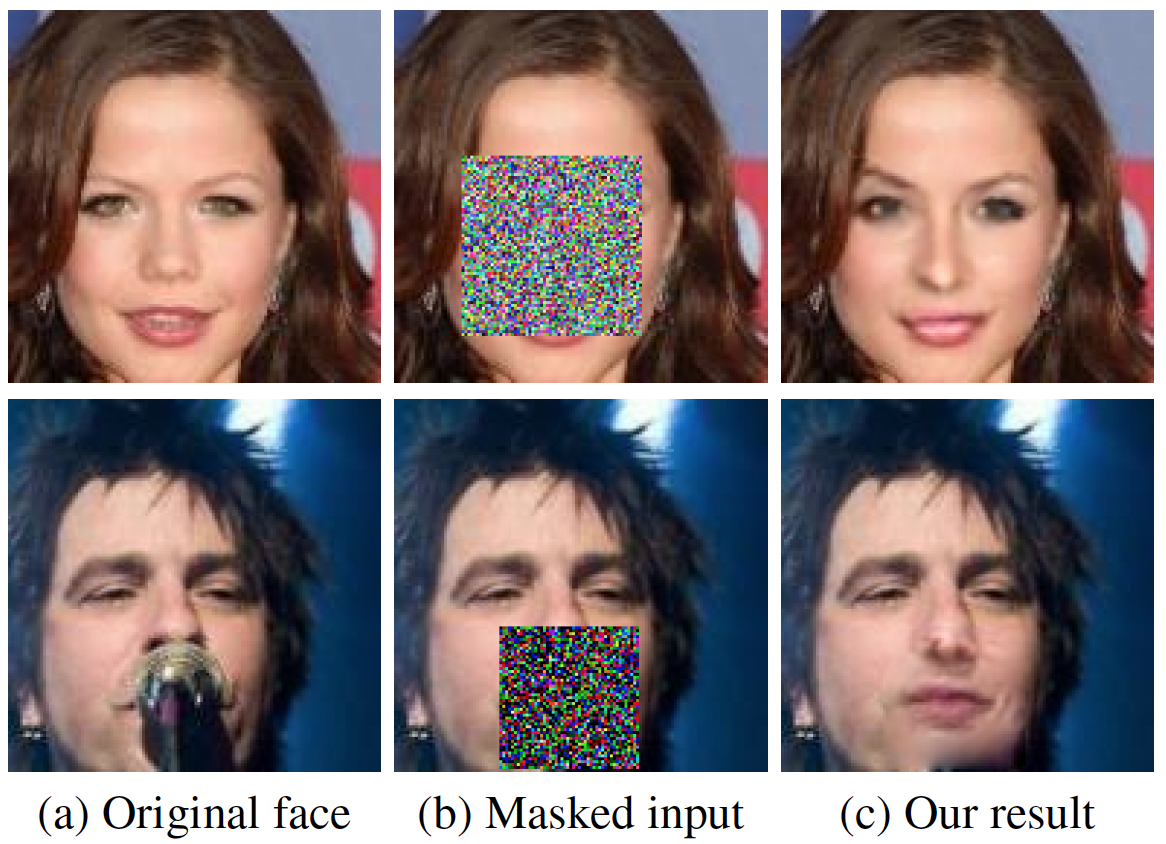
CAN (Creative Adversarial Networks, Elgammal et al., 2017) 은 예술품을 생성해내는 GAN 모형입니다. Generator는 가짜 예술품을 생성하고 Discriminator는 생성해낸 가짜 예술품이 진짜인지 가짜인지 판단하게 됩니다. 아래 그림이 CAN이 만들어낸 그림인데 많은 사람들에게 설문조사를 하였을 때 실제 예술품과 비슷한 점수를 받았다고 합니다.

# SRGAN



SRGAN (Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, Ledig et al, 2017) 은 저해상도 이미지에서 고해상도 이미지로 복원하는 GAN모형입니다. Generator는 이미지를 가짜 고해상도 이미지를 만들고 discriminator는 진짜 고해상도 이미지인지 가짜 고해상도 이미지인지 판단하게 됩니다.

# Image Completion



사진속을 Random하게 지워버렸을 때 채워주는 GAN 모델입니다. Generator는 빈 부분을 채워 넣으려고 하고 Discriminator는 가짜로 채워진 이미지인지 원본 이미지인지 구분하려고 노력합니다.

# Deepfake



딥러닝, 그 중에서도 GAN의 발전은 이미지의 다양한 Task에 대하여 엄청난 발전을 이끌었습니다. 위에서 언급한 CycleGAN을 비롯하여, 합성관련 Task에 대해서도 엄청난 발전이 있었습니다. Deepfake (딥러닝 + fake)라는 기술을 이용하여 만든 합성 영상입니다. 어색한 부분이 없진 않지만, 얼핏 보기에 진짜라고 느낄 수 있을 정도입니다. 매우 신기한 기술이기도 하고 유용하게 쓰일 수도 있다고 생각이 들 수도 있지만, 사실 악용될 여지가 많은 기술입니다.

한 포르노 사이트에서는 유명 연예인 여성의 얼굴을 합성해 포르노 동영상을 유포하기도 하였습니다. 또한 사기 등 다양한 범죄에 활용될 여지가 많습니다. 이처럼 인공지능의 발전은 인간의 삶을 조금 더 편리하게 바꿀 뿐만 아니라, 악영향을 끼치기도 합니다.