Self-Transfer Learning 을 통한 동영상 Super Resolution

김재연, 김재형, 김한빈

요약

4 차 산업혁명의 시대라고도 불리우는 현재, 인공지능 분야가 각광받고 있다. 머신러닝과 딥러닝은 인공지능 분야에서 필수적인 기술인 만큼, 우리 팀도 인공지능을 연마 및 활용하여 동영상의 화질을 높이는 기술을 연구해보고자 한다. 하지만, 기존에 동영상 초해상화 기술이 이미 많이 존재하기에, 동영상에 존재하는 부분을 활용한 동영상 초해상화 기법으로 차별화를 두고자 한다.

1. 서론

1.1. 연구 배경

2016 년 3 월, 이세돌 9 단과 알파고의 대국 이후, 세계는 인공지능 돌풍에 휩싸였다. 사람들은 머지않아 인공지능으로 기존에 하지 못하던 것을 이뤄내고, 새로운 기술을 접할 수 있다는 가능성을 염두하게 되었고, 점차 4 차 산업혁명 열풍이 일어남에 따라 소프트웨어 개발과 인공지능은 가장 큰 화두가 되었다. 2021 년 현재, 인공지능이 탑재된 전기차인 테슬라가 시중에서 많이 판매되고 있는것과 같이 인공지능을 탑재한 제품들이 점점 상용화되고 있는 가운데, 머신러닝과 딥러닝의 위상은 높아져가고 있으며, 이 기술들을 활용할 다른 분야에서 시장을 선점하려는 경쟁이 점점 심화되고있는 추세이다. 따라서 우리도 이러한 기술을 활용하여 연구를 해보고자 하는 것이 목표였고, 기존에 존재하던 인공지능을 활용한 화질이 낮은 동영상의 초해상화기술을 조금 더 발전시켜보자라는 것을 연구 주제로 삼게 되었다.

간략하게 설명하자면, 한 동영상 내에서 이미 존재하던 장면을, 화질이 낮은 다른 장면을 초해상화 시키는데 활용하는 기법을 연구해보고자 하는 것이다.

Super-Resolution 이란 저해상도의 이미지나 영상을 고해상도 이미지, 영상으로 변환시키는 방법으로 초해상화라고도 불린다. 기존의 고해상도 영상이나 이미지는 많은 용량을 차치하고, 따라서 전송하는데 시간이 많이 소요된다. 그렇기에 많은 비디오 스트리밍 플랫폼에서는 인터넷 사정에 맞춰 저해상도의 영상을 송출한다. 왜냐하면 해상도를 포기하고 전송속도를 높이는 것이 더 나은 사용자 경험을 제공하기 때문이다. 하지만 전달받은 저해상도의 영상이나 이미지를 사용자 측에서 빠르게

고해상도로 변환시킬수 있다면 속도와 해상도 문제를 모두 해결할 수 있다. 여기서 사용할 수 있는 기술이 바로 Super-Resolution 이다.

이러한 Super-Resolution 기술은 사실 꽤 오래전부터 존재했었다. Nearest Neighbor, Bilinear, Bicubic 등의 간단한 보간법(interpolation) 등을 사용하는 초해상화가 그 예인데, 이러한 초해상화 기법은 비교적 코드가 단순하지만, 썩 좋은 결과를 내지는 못했다. 하지만 최신 초해상화 기법은 딥러닝에 의해 발전했고, 더욱 질 좋은 결과물들을 만들어내는 중이다.

현재 영상 Super-Resolution 기술은 frame 단위로 Super-Resolution 을 적용한다. 영상의 자연스러운 연결을 위해 RNN 등을 이용해 인접한 frame 의 정보를 사용하는 경우도 있다. 하지만 인접한 frame 정보를 사용하는 것에 그치지 않고, 해당 영상에 대한 정보를 활용해서 Super-Resolution 할 수 있다면 더 좋은 결과를 얻을 수 있지 않을까라는 아이디어에서 본 연구를 시작하게 되었다.

1.2. 연구 목표

이 연구의 의의는 영상의 Super-Resolution 에 있어서 기존의 영상 Super-Resolution 의 방법과 비교해 얼마나 좋은 결과를 산출하는지 탐구하는 것이다. 영상길이, data augmentation 유무, fine tuning 레이어, 프레임 추출 등의 조건을 다르게 비교하면서 어떤 방법을 사용했을때 원본 영상을 더 나은 해상도로 만들 수 있는지 연구를 해볼 예정이다. 또한, 초해상화 모델을 적용시킬 영상의 주제와 산출되는 결과 사이의 유의미한 상관관계를 도출하는 것이 이 연구의 목표이다.

2. 관련 연구

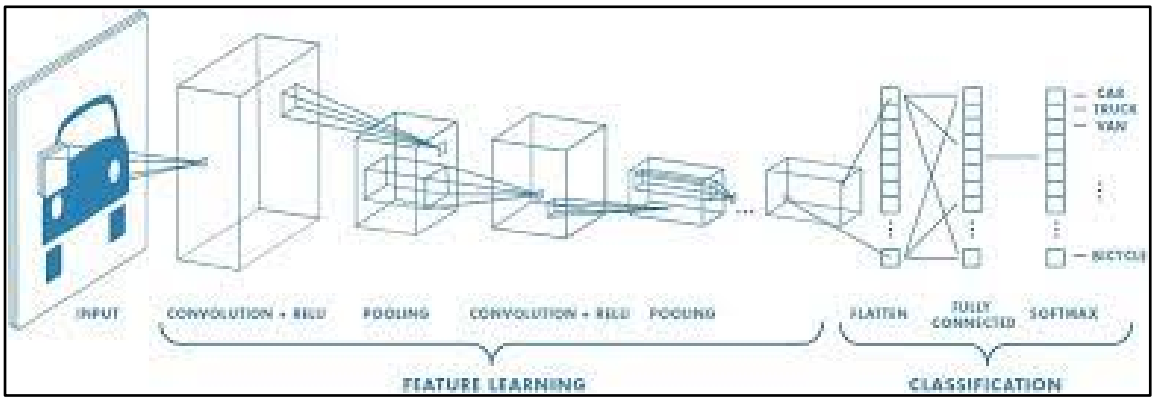
2.1. 딥러닝

딥 러닝은 여러 데이터간의 관계를 통해 학습시켜 높은 수준의 추상화를 시도하는 기계 학습 알고리즘의 집합으로 정의된다. 컴퓨팅 연산 속도가 발전함에 따라 빠른 연산과 높은 정확도로 인상적인 결과물들을 산출하고 있다.

2.2. 인공신경망 모델

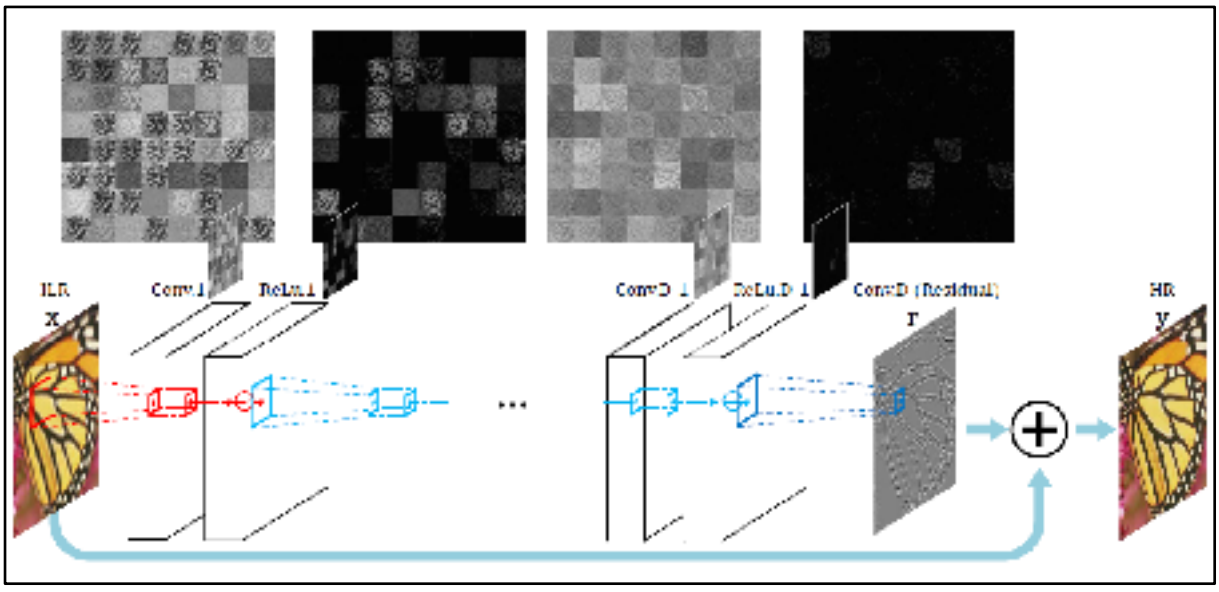
2.2.1 Convolutional Neural Network (CNN)

CNN(합성곱 신경망)은 딥러닝에서 주로 이미지나 영상 데이터를 처리할 때 쓰이며 Convolutional 이라는 전처리 작업이 들어가는 Neural Network 모델이다. 이미지 데이터를 가지고 신경망을 학습시켜야할 경우, 3 차원(공간 2 차원 + 색 1 차원) 데이터가 필요하다. 이미지의 3 차원 데이터를 손실하지 않고 학습이 가능한 모델이 CNN 이다.



CNN 의 신경망 계층은 Convolutional Layer 와 Pooling Layer 들을 활성화 함수 앞뒤에 배치하여 구성된다. Convolution Layer 에서는 이미지의 특성이 추출된 Feature Map 을 생성한다. Pooling Layer 에서는 Feature Map 에 있는 원소들을 최적화하는 연산을 수행한다. Convolution Layer 와 Pooling Layer 가 반복되며 이미지의 특징을 효과적으로 추출하고, Flatten, Fully-Connected Layer, Softmax function 을 적용해주면 최종 결과물을 출력하게 된다.

2.2.2 VDSR

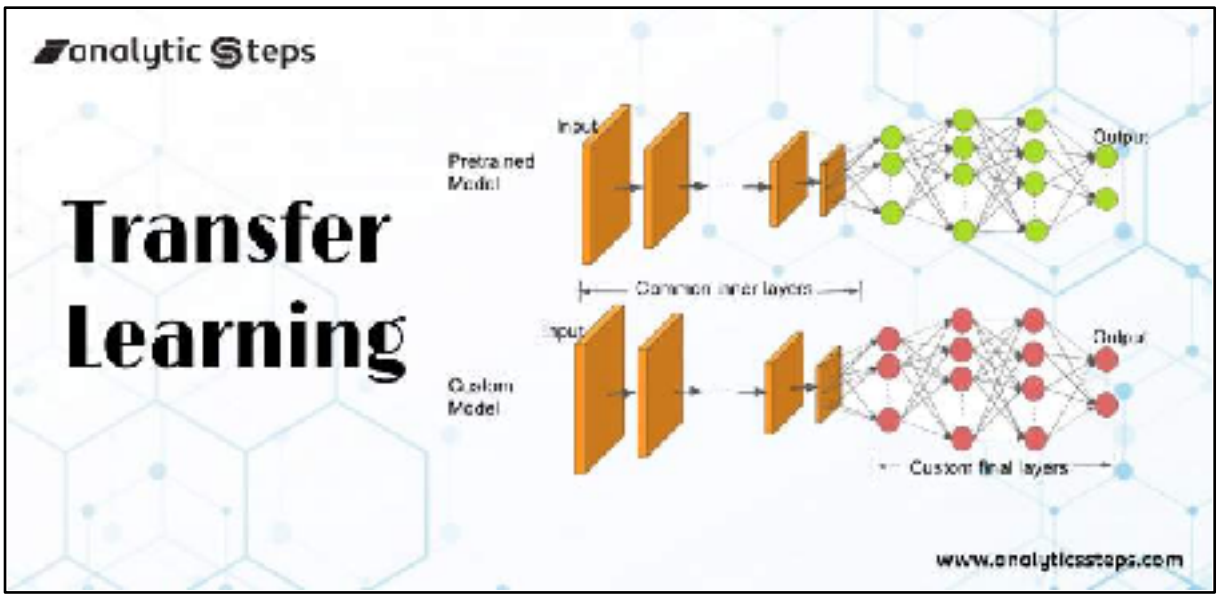


VDSR 은 Very Deep Super-Resolution 의 약자로, 보통 딥러닝을 시킬때 최대한 많은 데이터셋을 가지고 학습을 시킨 후 여러개의 신경망 곱을 통하여 가장 이상적인 결과를 도출해내는 방법을 사용하는 것과는 달리, 미리 저해상도 영상과 고해상도 영상의 매핑에 대해 훈련된 VDSR 신경망을 가지고 하나의 데이터에서 더 낳은 결과를 낳는 방법이다. 특히 단일 영상 초고해상화에 많이 사용되는 기술이다.

2.2.3 EDSR

기존의 SR 모델들은 PSNR 을 목표로 많은 성능 향상이 이루어졌다. 하지만 해당 모델들은 신경망 모델의 재구성 성능이 사소한 구조적인 변화에 민감하고, 다양한 scale 을 다루기 위하여 scale 에 따른 각각의 네트워크를 필요로 한다는 점에서 한계점이 존재했다. EDSR 모델은 위의 한계점을 제거하기 위하여 resnet 구조에서 불필요한 모듈을 분석하고 제거하여 최적화하고, 다른 scale 로 훈련된 모델로부터 knowledge 를 가져오는 방법을 조사하여 다른 scale 에 걸쳐 parameter 를 공유하는 새로운 multi-scale architecture 를 제안한다.

2.3. Transfer-Learning



일반적인 머신러닝, 딥러닝은 최대한 많은 데이터셋을 학습하여 많은 학습 모델을 만들고, 좀 더 보편적인 상황에 대한 해답을 주는 방법이라면, transfer-learning 은 특정한 상황에 대한 데이터셋만을 모아 따로 학습시켜 한 분야에 특화되도록 학습시키는 방법이다. 따라서 일반적인 딥러닝에 비해 자신의 전문 분야가 아닌 곳에서는 유효한 결과를 내지 못하지만, 학습한 분야에서는 훨씬 더

정교한 결과를 산출한다. 이러한 transfer-learning 의 특성을 바탕으로 하여, 동영상 초해상화작업 진행시 동영상을 장르별로 분류한 후 각각의 장르에 맞는 transfer-learning 모델을 사용한다면 더욱 해상도가 높은 동영상을 만들 수 있을 것이다.

2.4. NIQE

NIQE 란 Natural Image Quality Elevator 의 약자로서 기본적으로 이미지나 영상의 품질을 평가하기 위한 모델을 뜻한다. 기본적으로 이미지나 영상의 품질에 대한 평가를 하기 위해서는 품질을 비교평가하기위한 해상도가 높은 원본의 이미지를 필요로 하게 되는데, 이를 해결하기 위해 나온 방법이 NR(No-Reference) 방법이다. NR 은 비교할 원본 이미지 없이, 데이터셋을 학습하여 만들어진 자체 모델로 이미지를 평가한다. 하지만 이러한 방법은 사실 많이 불안정한데, 직관적인 사진을 제외한 애매모호한 사진에 대해(ex) 개의 머리, 고양이의 다리, 말의 몸통을 합친 사진을 넣으면 각각 불완전한 결과를 나타낸다)서는 일관적인 결과를 낼 수 없기 때문이다. 따라서 이러한 피사체 대상 라벨과, NR 의 대상 라벨에 대한 맞춤형 학습이 필요없이 평가를 할 수 있는 모델을 만들어 내는 것이 NIQE 기법이다. 즉, 특정 피사체에 대한 학습 없이 기본적으로 해상도가 좋은 모든 이미지를 학습한 후, 그 이미지들이 보편적으로 가지는 특성을 분석하여 모델을 만들고, 비교할 이미지의 특성이 해상도가 좋은 이미지들의 일반적 특성과 비슷한지 판단하여 비슷할수록 해상도가 좋은 이미지로 판결을 내리는 것이다.

3. 프로젝트 내용

기존의 동영상 Super Resolution 연구는 프레임 단위로 영상을 분할 한 뒤 각 프레임에 대해 Image Super Resolution 을 수행하고, SR 된 프레임을 다시 영상으로 합치는 방법을 주로 사용하였다. 인접한 프레임에 대한 초해상화 결과의 차이가 크다면 영상이 부자연스럽게 보일 수 있기 때문에 이를 개선하기 위해 인접한 5-7 프레임 가량을 RNN 등의 인공신경망을 통해 학습하여 초해상화에 사용하는 기법은 연구되어 왔으나, 인접하지 않은 한 동영상 내의 다른 프레임을 영상의 초해상화에 사용하는 연구는 이루어지지 않았다.

본 프로젝트는 하나의 영상은 하나의 주제에 대한 영상이며 그 안에서 다수의 중복된 장면이 확대되어 여러번 나타난다는 관찰에서 출발하여, 이러한 중복된 프레임이 영상의 초해상화에 사용될 수 있음을 입증하는 것을 목표로 한다.

영상의 프레임을 신경망에 접목시키기 위해서는 RNN, LSTM 등의 시계열 인공신경망을 사용하는 것이 일반적이다. 하지만 영상의 프레임이 많아지면 시계열 인공신경망에서 기억해야하는 정보의 양이 기하급수적으로 늘어나며, Patch 기반으로 초해상화를 진행할 경우 하나의 프레임 안에서도 수많은 Patch 들이 분리되기 때문에 이러한 시계열 인공신경망을 영상 전체 프레임으로 학습시켜 초해상화에 사용하기에는 큰 어려움이 있다. 따라서 본 프로젝트에서는 Pretrain 된 초해상화 신경망 모델을 초해상화 대상 영상의 프레임으로 다시 학습시키는 Self-Transfer Learning 기법을 사용하여 Super Resolution 을 수행한다.

먼저 DIV2K 등의 데이터셋으로 학습된 기존 Pretrained 모델로 동영상 초해상화를 수행해 Baseline 을 구축한다. 이후 Self-Transfer Learning 으로 해당 영상의 프레임으로 재학습된 Fine-tuned 모델로 동영상 초해상화를 수행하고, Ground-Truth, Baseline 과의 비교를 통해 제안하는 기법의 성능을 측정한다. 이미지 품질 측정에는 PSNR, SSIM 등의 객관적인 지표와 함께 사람이 체감하는 이미지 품질의 측정을 위해 MOS 를 사용한다.

최적의 고해상도 영상을 얻기 위해, 다양한 조건에 대해 실험을 수행한다. 영상의 길이가 길어질수록 Transfer Learning 에 사용되는 학습 데이터의 양이 늘어나기 때문에 이를 비교하기 위해 60 초 이내의 짧은 영상, 60 초에서 10 분 사이의 중간 길이 영상, 10 분 이상의 긴 영상으로 영상의 길이를 분류하여 다양한 길이의 영상에 대해 비교를 수행한다. 또한 영상의 분류가 해당 기법에 미치는 영향을 비교하기 위해 자연 영상, 인터뷰 영상, 도시 영상, 애니메이션, 화면 녹화 등 다양한 종류의 영상에 대해 비교 실험을 수행한다.

영상의 모든 프레임을 Transfer Learning 의 학습 데이터로 사용한다면 학습 데이터의 양은 매우 크게 늘어나지만, 대부분의 프레임이 중복되므로 학습 데이터 불균형 문제가 발생한다. 이러한 학습 데이터 불균형으로 인한 Over-fitting 의 영향을 분석하기 위해 프레임 추출 방식을 모든 프레임을 추출하는 것과 초당 n 개의 일부 프레임만 추출하는 것, 동영상의 I-Frame 만 추출하여 학습에 사용하는 것으로 구분하여 실험을 수행한다.

CNN 기반의 딥러닝 모델에서, 상위 레이어는 저수준의 특징을 추출하며 레이어의 단계가 깊어질수록 더 고수준의 특징을 추출한다는 것이 실험적으로 알려져 있다. Pretrain 된 모델의 저수준 특징은 초해상화 대상 영상의 프레임에서도 재사용할 수 있음을 경험적으로 알 수 있으므로, Fine-tuning 을 수행할 때 모델의 모든 레이어의 가중치를 조정한 모델과, 상위 레이어의 가중치는 고정한 채 고수준 특징을 추출하는 Last-K 레이어의 가중치만 조정한 모델의 성능을 비교하는 실험을 수행한다.

EDSR 의 경우 L2 Loss 를 사용하는데, 이는 PSNR 성능을 올리는데는 탁월하지만 사람이 느끼는 이미지 품질과 PSNR 수치와는 차이가 있음이 알려져있다. 따라서 사람이 느끼는 이미지 품질을 올리기 위해 L2 Loss 외에도 Perceptual Loss 등의 다양한 Loss 함수에 대한 실험을 수행한다. 또한 Self-Transfer Learning 에서는 LR 과 비교하여 학습하기 위한 HR 영상이 존재하지 않는데, 이 경우 LR 을 bicubic downsampling 하여 새 LR 을 만들고 기존의 LR 을 HR 로 대치하여 학습하는 방법과, HR 없이도 이미지 품질 측정이 가능한 NIQE 등의 No Reference Image Quality Metrics 를 활용하는 방법을 실험한다.

4. 결론 및 기대 효과

본 프로젝트에서 제안하는 기법으로 동영상 초해상화의 품질을 향상시키는데 성공한다면, 영상에 따라 초해상화 시켜야 하는 포인트를 찾아내어 맞춤형으로 더욱 세분화된 고화질 영상을 얻을 수 있다.

하나의 영상의 초해상화를 수행하기 위해서 모델 전체를 재학습하는 과정을 거쳐야 하므로, 본 프로젝트에서 제안하는 기법은 실시간 영상 초해상화 등 모델의 추론 성능이 중요한 활용 영역에는 적합하지 않다. 그러나 고해상도의 영상을 구할 수 없는 과거의 개인 영상이나 중요한 역사적 사료와 같은 경우, 초해상화에 수십~수백 시간의 시간이 걸리더라도 더 선명한 고화질 영상을 얻는 것이 중요할 것이다. 본 프로젝트에서 제안하는 기법은 이러한 분야에서 더욱 선명한 고화질 영상을 얻게 해줄 것이라 기대할 수 있다.

5. 참고 문헌

[1] “Image Super-Resolution Using Deep Convolutional Networks”, 2014 ECCV
[2] “Super Resolution Applications in Modern Digital Image Processing”, 2016 IJCAd
[3] “Loss Functions for Image Restoration with Neural Networks”, 2016 IEEE TCI
[4] “Deconvolution and Checkerboard Artifacts”, distill blog(https://distill.pub/2016/deconv-checkerboard/)
[5] “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”, 2016 CVPR

[6] “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network”, 2016 CVPR

[7] “Accelerating the Super-Resolution Convolutional Neural Network”, 2016 ECCV

[8] “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”, 2017 CVPR

[9] http://www.vision.ee.ethz.ch/~timofter/publications/NTIRE2017SRchallenge_factsheets.pdf

[10] “The Perception-Distortion Tradeoff”, 2018 CVPR

[11] “Deep Learning for Single Image Super-Resolution: A Brief Review”, 2018 IEEE Transactions on Multimedia (TMM)

[12] “A Deep Journey into Super-resolution: A survey”, 2019 arXiv

[13] CNN 소개 (http://taewan.kim/post/cnn/)

[14] “Accurate Image Super-Resolution Using Very Deep Convolutional Networks” CVPR 2016

[15] https://github.com/sanghyun-son/EDSR-PyTorch

[16] 안드로이드 플레이스토어 : Remini, Deepfake Studio

[17] “Learning with Privileged Information for Efficient Image Super-Resolution” ECCV 2020