### [구글 문서 링크](https://docs.google.com/document/d/11-5Toy66cl1p9Ezk0X_aVnUSgmZsBv7pXfTqd7UNV4g/edit)

### 참조 이미지 기반 super resolution 연구

[CrossNet: An End-to-end Reference-based Super Resolution Network using Cross-scale Warping](https://openaccess.thecvf.com/content_ECCV_2018/html/Haitian_Zheng_CrossNet_An_End-to-end_ECCV_2018_paper.html)

* Flower dataset, Lytro ILLUM 카메라로 촬영한 3343개의 꽃, 식물 라이트필드 이미지(크기: 376 \* 541)

라이트필드 이미지

* 일반적인 2D 이미지와는 달리, spatial samples, angular samples를 포함하며 이미지를 일련의 방향에서 촬영한 특별한 유형의 이미지. 깊이와 방향 정보를 포함

논문에서의 spatial samples : 376 \* 541, angular samples 14\*14개

* LFVideo dataset, Lytro ILLUM 카메라로 촬영된 실내 및 실외 2D 라이트필드 이미지 (크기: 376 \* 541)

2D 라이트필드 이미지: 라이트필드 이미지를 평면에 투영한 것. 깊이 정보를 소실하여 라이트필드 이미지의 일부 정보만을 포함함.

논문에서의 spatial samples : 376 \* 541, angular samples 8\*8개

* RefSR(Reference-based Super-Resolution)알고리즘 사용.
  + HR 참조 이미지를 이용하여 LR 이미지를 HR 이미지로 변환하는 기술.
  + LR 이미지와 HR 참조 이미지간의 이미지 대응 문제, LR 이미지의 고해상도 합성 문제 해결해야함.-> 각각 Optical Flow estimation, Encoder-Decoder architecture를 사용함
* CrossNet 네트워크 사용.
  + LR 이미지 인코더, 참조 이미지 인코더, 그리고 디코더로 구성
  + 다음과 같은 단계로 이루어짐

LR 이미지 인코더: LR 이미지에서 다양한 scale의 feature map 추출

참조 이미지 인코더: HR 참조 이미지에서 LR 이미지와 대응되는 다양한 scale의 feature map 추출

추출된 feature map: cross-scale warping진행, 대응되는 realationship을 파악

->이를 이용하여 디코더가 HR 이미지로 합성

* 논문에서는 라이트필드 이미지의 (0,0), 가장 중심 위치에서 취득한 이미지(LR)이미지를 HR 참조 이미지로 사용하였음. 이는 이 이미지가 다른 이미지와의 시점 차이가 가장 적기 때문.

### super resolution 관련된 연구

### [Accurate image super-resolution using very deep convolutional networks](https://openaccess.thecvf.com/content_cvpr_2016/html/Kim_Accurate_Image_Super-Resolution_CVPR_2016_paper.html)(2016) by Jiwon Kim / 6,493회

* very deep CNN 사용.
* 20개의 convolution layer과 1개의 pooling layer 사용. resnet 구조를 사용. 입력 이미지와 출력 이미지 사이의 residual를 이용하여 고해상도 이미지 생성.

convolutional layer

* 입력된 이미지와 필터(kernel)을 합성곱하여 feature map 생성. 이때 필터는 입력 이미지를 슬라이딩하며 각 위치에서의 곱셈 연산 수행. 이 결과를 다음 레이어로 전달.

pooling layer

* featuring map의 크기를 줄이는데에 사용됨. featuring map을 일정 크기의 영역으로 나누어 각 영역에서 가장 큰 값을 선택하여 새로운 featuring map을 만듦.
* 모델 방식 : 기존의 사진(사진 A)을 저해상화하여(사진 B) 사진 A와 B 사이의 mapping을 만든 후, 이를 사진 A에 적용하여 고해상화 사진(사진C)를 생성한다.
* 기존의 super resolution은 픽셀을 기반으로 하여 이를 직접 조작하지만, 여기서 사용한 구조는 입력 이미지의 저해상도 버젼을 이용하여 고해상도 이미지를 생성함 -> 더욱 자연스러움

([Image Super-Resolution Using Deep Convolutional Networks](https://ieeexplore.ieee.org/abstract/document/7115171?casa_token=QzjMPQyAxHwAAAAA:X6qSNTOqlZZsp11nbSCXirn0pwakX-HNpzcYqeT-_V1BLyLY_BXtPowRXG-nK_D9GoV1h_I) (2015, IEEE) by Chao Dong / 8,000회 / [pdf](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7115171&casa_token=T3aUghbknngAAAAA:Oi69lCChT3vF65UFowhMshkEc6cbVwADKmvBylnB6hus3MPXFGVJ-RbMpUw15l1DP-4FEYY&tag=1))

* very deep CNN 사용.
* 20개의 convolution layer과 1개의 pooling layer 사용. resnet 구조를 사용. 입력 이미지와 출력 이미지 사이의 residual를 이용하여 고해상도 이미지 생성.

convolutional layer

* 입력된 이미지와 필터(kernel)을 합성곱하여 feature map 생성. 이때 필터는 입력 이미지를 슬라이딩하며 각 위치에서의 곱셈 연산 수행. 이 결과를 다음 레이어로 전달.

pooling layer

* featuring map의 크기를 줄이는데에 사용됨. featuring map을 일정 크기의 영역으로 나누어 각 영역에서 가장 큰 값을 선택하여 새로운 featuring map을 만듦.
* 모델 방식 : 기존의 사진(사진 A)을 저해상화하여(사진 B) 사진 A와 B 사이의 mapping을 만든 후, 이를 사진 A에 적용하여 고해상화 사진(사진C)를 생성한다.
* 기존의 super resolution은 픽셀을 기반으로 하여 이를 직접 조작하지만, 여기서 사용한 구조는 입력 이미지의 저해상도 버젼을 이용하여 고해상도 이미지를 생성함 -> 더욱 자연스러움
* [code](https://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html)(matlab)

[Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network](https://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html) (2017) by Christian Ledig / 10,682회

* SRGAN(super resolution generative adversarial network) 프레임워크 사용.

perceptual loss function(지각 손실 함수)란?

* 이미지 생성 및 변환 작엄에서 사용되는 loss function.
* 인간의 perception(지각)을 모방함.
* 이미지의 feature를 추출하는 딥러닝 모델을 사용하여 이미지간 차이 계산.

SRGAN의 perceptual loss function

* adversarial loss와 content loss로 구성됨.

adversarial loss

* discriminator network를 사용하여, SRGAN이 생성한 HR 이미지와 실제 HR이미지(자연스러운 이미지)와 구분되지 않도록 학습

content loss

* vgg19 network를 사용하여, SRGAN이 생성한 HR 이미지와 실제 HR 이미지간의 feature 차이 계산 -> 유사한 feature를 가지도록 학습.

=> 둘 다 자연스러운 이미지 생성에 기여

* generator(생성자)는 16개의 residual block, 2개의 sub-pixel convolution layer

discriminator(판별자)는 8개의 convolutional layer, 2개의 fully connected layer로 이루어짐.

residual block

* layer 아님.
* gradient vanishing을 해결하기 위해 사용.
* 모델 방식 : 기존의 사진(사진 A)을 저해상화하여(사진 B) 사진 A와 B 사이의 mapping을 만든 후, 이를 사진 A에 적용하여 고해상화 사진(사진C)를 생성한다.
* 논문 코드 x, SRGAN에서 제공하는 [코드](https://github.com/tensorlayer/srgan)

[ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks](https://openaccess.thecvf.com/content_eccv_2018_workshops/w25/html/Wang_ESRGAN_Enhanced_Super-Resolution_Generative_Adversarial_Networks_ECCVW_2018_paper.html) (2018) by Xintao Wang

/ 3,018회

* ESRGAN(Enhanced SRGAN) 사용.

[딥러닝 기반 Super Resolution 기술의 현황 및 최신 동향](http://www.kibme.org/resources/journal/20200504094149078.pdf) (2020) by 서강대학교 서유림, 강석주

* SRCNN, FSRCNN, VDSR, EDSR 등의 모델이 사용되고 있음.
* 대용량 학습을 통한 효과적인 화질 개선 가능.

SRCNN

* 최초의 딥러닝 활용 초해상화 모델.
* end - to - end 방식으로, 모든 단계를 하나의 네트워크에서 처리함.
* 바이큐빅 보간법, 디컨볼루션 연산 사용.
* 3층의 FCN(Fully Convolutional Network) 구조.

FCN

* 합성곱 계층(Convolutional Layer)만으로 이루어진 신경망 구조
* 입력 이미지를 합성곱 계층을 통하여 Feature map으로 변환 후 다시 합성곱 계층을 통하여 원하는 출력값으로 변환.
* 이미지의 크기에 상관 없이 동작 가능.

FSRCNN(Fast Super Resolution Convolutional Neural Network)

* 바이큐빅 보간법과 디컨볼루션 연산 대신 1개의 합성곱 계층과 1개의 업스케일링 계층 사용.
* 더 경량화되어, 더 높은 연산 속도를 보이며 성능도 더 좋음.

VDSR(Very Deep Super Resolution)

* 바이큐빅 보간법, 디컨볼루션 연산 사용.
* SRCNN과는 달리, 잔여학습(Residual Learning)을 사용하여 더 deep한 네트워크 구성.
* 20개의 합성곱 계층 구조.

EDSR(Enhanced Deep Super Resolution)

* 잔여학습과 Skip Connection을 사용함.
* 32개 이상의 레이어 사용, 채널 수도 4배 이상 늘려 파라미터가 매우 많음.
* 성능이 매우 높음.

### (위성)영상과 관련된 super resolution 연구

[딥러닝 기반 위성영상 복원에 관한 연구](https://journal-home.s3.ap-northeast-2.amazonaws.com/site/2020f/presentation/0234.pdf)(2018,한국항공우주학회지) by 한국항공우주연구원 임지현, 이재열, 김민아, 정대원

* 영상 인페이팅 기술을 활용한, 위성 영상 속 손상된 영역 복구 연구.
* GLCIL(Globally and Locally Consistent Image Completion) 모델 사용. CNN과 GAN(Generative adversarial networks)을 결합한 이중구조.
* 단순한 위성 영상에서는 (시각적으로) 자연스럽게 복원하였으나, 건물 밀집지영 및 도로와 같이 복잡도가 높은 경우에는 완성도가 떨어짐.

[Real-time Single Image and Video Super-Resolution using an Efficient Sub-pixel Convolutional Neural Network](https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Shi_Real-Time_Single_Image_CVPR_2016_paper.html) (2016, IEEE) by Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang / 5,448회

* 720p, 1080p 해상도의 이미지를 실시간으로(초당 30프레임 이상) 4k 해상도로 변환하는 기술을 연구
* Sub-pixel CNN 사용. 입력된 이미지를 저해상도 이미지로 받아들인 후, 여러 개의 CNN layer을 거쳐 고해상도의 이미지로 변환. 각 레이어에서 feature map을 생성하고 다음 레이어로 넘긴 후, 마지막에는 sub-pixel layer로 넘김. sub- pixel layer는 feature map을 일정하게 재배치하여 고해상도 이미지를 생성해냄.

### image recognization에 관한 연구(초해상화에 큰 관련 x)

### 

[Deep Residual Learning for Image Recognition](https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html) (2016) by Kaiming He / 172,159회

* ResNet 사용.

ResNet

* gradient vanishing을 해결하기 위해 사용

gradient vanishing

* backpropagation algorithm(역전파 알고리즘)에서 발생하는 문제

backpropagation algorithm

* weight 학습에 사용되는 알고리즘
* 출력값과 실제값 사이의 residual을 이용하여 각 layer의 weight를 역방향으로 조정 -> 입력값과 출력값 사이의 복잡한 relation 학습.

* 네트워크가 깊어질수록 gradient가 작아지거나 커져서 학습이 어려워짐
* direct mapping에서 심해짐. residual mapping에서는 gradient가 안정적으로 유지됨 -> 더 깊은 네트워크 학습 -> 높은 accuracy 달성
* skip connection 사용으로 입력값이 출력값에 직접적으로 전달되어 문제 해결
* direct mapping이 아닌 residual mapping 사용

direct mapping

* 입력값을 출력값으로 바로 매핑
* 네트워크가 입력값과 출력값 사이의 복잡한 relation을 학습해야함

residual mapping

* 입력값과 출력값 사이의 차이를 출력값에 더해줌
* 차이를 학습하므로 네트워크가 학습해야하는 relation이 비교적 단순함.