WGAN-GP를 활용한 MRI 뇌 영상 이미지 Super Resolution 수행

2015103920 서범진

MRI Brain image Super Resolution with WGAN-GP

요 약

기존의 DenseNet 모델을 수정해 구성한 3D multi-Level Densely Connected Network 모델을 변형하여 MRI Brain image에 대해 super resolution을 적용해 저화질 이미지를 고화질로 개선함으로써, 의사 혹은 연구원에게 환자에 대한 더욱 정확한 정보를 제공할 수 있도록 한다.

1. 서 론

기존의 HR MRI는 정확한 anatomical information를 제공하면서 clinical application이나 quantitative image analysis 부문에서 중요한 역할을 하였다. 하지만 HR MRI는 몇몇가지 단점이 있는데, 1) 스캔 타임이 길고 2) spatial coverage가 작고 3) signal-to-noise ratio (SNR) 값이 작다는 단점이 있다. 따라서 최근 연구에서는 저화질의 이미지를 고화질로 개선하는 single image super-resolution (SISR) 기법 작업을 CNN을 통해 수행되는 연구들이 진행 중이다[1]. 한편 이 부분에 있어서도 3D image data를 다루는 만큼 메모리 소비를 줄이고 모델을 빠르고 가볍게 작동하도록 하는 것이 issue로 남아있다. 따라서 이번 프로젝트에서는 모델을 조금 더 가볍고 성능을 좋게 할 수 있는 방안에 대해 고려하는 관점에서 프로젝트를 진행한다.

기존의 논문에서는 3D multi-Level Densely Connected Network(mDCSRN) 이라는, DenseNet의 변형 모델을 구성하여 super resolution을 수행하였다[2]. 이 모델의 업데이트 버전으로, WGAN-GP를 함께 이용했다고 발표하였다. 모델의 성능이 개선됨을 확인할 수 있었지만 GAN을 함께 사용하며 학습해야 하는 파라미터가 기하급수적으로 늘어났고 그만큼 학습이 오래걸리고 모델이 커지는 단점이 있었다[3]

이번 프로젝트에서는 Data Augmentation, gradient penalty 적용, model compression을 통해 이런 단점을 개선해 보고자 한다.

2. 관련 연구

2.1 super Resolution

SR 작업은 저해상도 이미지를 고해상도 이미지로 복원하는 작업이다. SRCNN은 SR 작업을 최초로 deep learning에 적용해 수행한 모델이다.

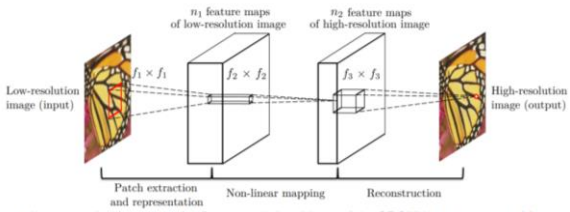

Fig. 2. Given a low-resolution image $\mathbf{Y}$, the first convolutional layer of the SRCNN extracts a set of feature maps. The second layer maps these feature maps nonlinearly to high-resolution patch representations. The last layer combines the predictions within a spatial neighbourhood to produce the final high-resolution image $F(\mathbf{Y})$.

그림 1. SRCNN 구조

위의 그림1을 보면, 입력 저해상도 이미지를 Y, 복원한 출력 고해상도 이미지는 F, ground truth 고해상도 이미지는 X로 표현하여 표시하고 있다[3]. 우선 Patch extraction and representation를 통해 저해상도 이미지
Y로부터 patch를 추출하고, 다차원 patch 벡터를 다른 다차원 patch 벡터로 mapping하는 Non-linear mapping 과정을 거친 뒤, 다차원 patch 벡터에서 최종 고해상도 이미지를 생성하는 reconstruction 과정을 수행한다.

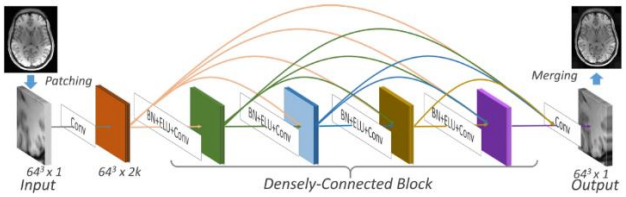

Figure 1. Framework of the proposed 3D Densely Connected Super-Resolution Networks (DCSRN).

그림 2. mDCSRN 구조

힌편, 이번 프로젝트에서 참고가 되었던 모델은 3D Multi-Level Densely Connected Super-Resolution Networks (mDCSRN)이다[2]. MRI Brain image를 patching 시켜, (64,64,64)의 patch를 구성하도록 한다. 모델의 내부에서는 BN-ELU-CONV가 하나의 Dense Block를 구성하고 있으며 총 4개의 Dense block들을 가진다. 해당 모델의 가장 두드러지는 특성은 모든 채널들이 concatenation 된다는 것이다. 이를 통해 Vanishing gradient 개선, Feature propagation 강화, Feature Reuse, Parameter number 절약 등의 효과를 거둘 수 있다.

2.2 WGAN-GP

현재 우리가 풀고자 하는 문제는 모델이 만들어 내고 있는 분포가 실제 데이터의 참분포와 유사한 정도가 크도록 구성되어 있다. 여기서 문제가 되는 점이 이 유사한 정도를 나타내는 '분포 사이의 거리'에 대해 어떤 조건들을 고려하며 거리를 정의해야 할지에 대한 부분들이다. 한편, WGAN을 사용하면 모델의 stability of learning이 개선되어 모델의 붕괴 문제를 해결할 수 있다[4]. 기존의 경우는 판별기가 학습이 앞서, 판별기의 함수가 step function으로 수렴하게 되어 gradient vanishing 문제가 발생하게 된다. WGAN을 사용하면, D는 G보다 먼저 최적화되어도 계속 의미 있는 gradient를 생성할 수 있어 model collapsing이 발생하지 않도록 한다. 하지만 WGAN을 사용하더라도 아직 문제가 남아있는데, weight를 clipping 해버린 점이다. Lipschitz constraint에 따라 개별 gradient를 제한하는 것은 의미가 없어 W 자체를 제한해 버린다[4]. 한편, 이러한 부분이 poor samples or fail to converge를 일으킨다는 것을 알아냈고 norm of gradient of critic을 penalize하는 것을 제안함으로써 이를 개선하고자 하였다[5].

$$L = \underbrace{\mathbb{E}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}\left[D(\tilde{\boldsymbol{x}})\right] - \mathbb{E}_{\boldsymbol{x}\sim\mathbb{P}_r}\left[D(\boldsymbol{x})\right]}_{\text{Original critic loss}} + \underbrace{\lambda \mathbb{E}_{\hat{\boldsymbol{x}}\sim\mathbb{P}_{\hat{\boldsymbol{x}}}}\left[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1)^2\right]}_{\text{Our gradient penalty}}.$$

그림 3. new objective of WGAN-GP

3. 제안 모델

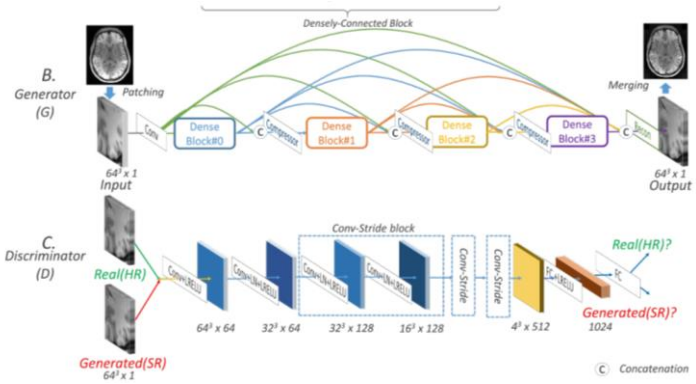

그림 4. mDCSRN-WGANGP 구조

3.1 기존 모델의 문제점

기존의 모델은 위의 그림4와 같은 구조로 구성되어 있다[4]. 기존의 mDCSRN model이 generator 역할을 하며 SR image를 생성해내고, Discriminator는 ground truth 이미지와 SR image를 입력으로 받아, 무엇이 진짜인지 구별해 나가는 관점에서 학습해 나간다. 기존 모델의 문제점을 파악해보면, 1) 파라미터를 계산해 보면, generator의 경우, 411648로 0.412M, discriminator의 경우, 31155777로 31.156M의 규모이다. 따라서 모델의 규모를 줄이면서 모델의 성능을 개선시키는 것을 이번 프로젝트의 목표로 한다.

3.2 제안 모델

3.1에서 제안한 문제점들로 근거해 크게 3가지 변화들을 수행한다.

1) 모델 파라미터의 큰 부분을 차지하는 Discriminator 모델을 모델 압축을 통하여 파라미터를 줄이는 작업을 수행한다.

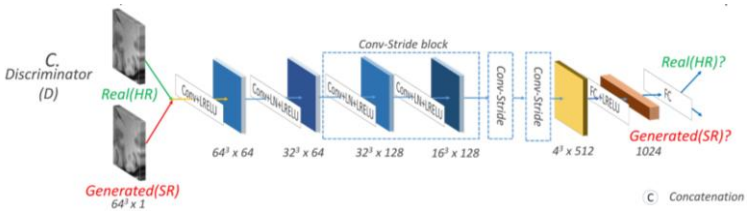

그림 5. Discriminator 구조

Discriminator의 구조를 확인해 보면, 그림5와 같다. 해당 구조에서 convolution의 kernel size는 3x3으로 구성되어 있으며, CONV-LN-LRELU 의 블록들로 구성되어 있다. 또한 마지막에는 Fully Connected를 수행하는 것으로 구성되어 있다.

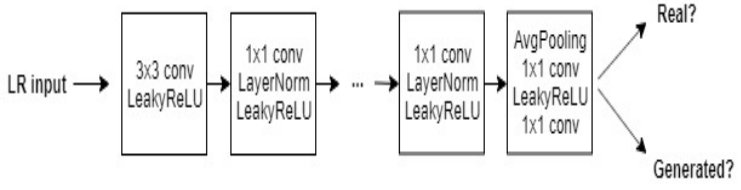

그림 6. 제안하는 Discriminator model 구조

그림6은 이번 프로젝트에서 제안하는 Discriminator의 모델이다. 기존 블록 내의 3x3 convolution을 모두 1x1convolution으로 교체해주고 Fully connected를 수행하는 대신, Global Average Pooling을 수행해준다.

2) 저자의 코드는 tensor flow로 구현된 mDCSRN의 모델만이 존재한다. pytorch로 구현된 다른 기존의 코드를 활용하였지만 코드를 직접 읽어보니 구현하는 데에 있어 코드가 문제가 많았다. (loss 함수가 잘못 정의되어 있거나 gradient penalty를 적용하였다고 기재하였으나 코드를 읽어보니 적용하지 않는 등.)

```
D_loss.backward()
self.optimizerD.step()

# weight clipping
for p in self.netD.parameters():
    p.data.clamp_(-0.01, 0.01)
return sr_patches, D_loss, G_loss, loss
```

그림 7. 기존 소스코드의 weight clipping으로 구현된 부분

최근 연구에 따른 부분으로 개선되어 있지 않은 부분 또한 있었는데, 그림7과 같이 WGAN에서 gradient penalty가 적용되지 않은 점 또한 그 중 하나였다. 따라서 gradient penalty를 적용하는 regularization 항을 더해준다. weight clipping을 대체함으로써 좀 더 효율적인 학습을 이끌 수 있다.

3) 이번 프로젝트를 수행하기 위해 Human Connectome Project (HCP) Data 중 MIR Brain image를 필요로 하였고 Laboratory Of Neuro Imaging(LONI) 기관에 데이터에 접근할 수 있는 권한을 받아 데이터를 다운로드할 수 있었다. 다운받은 데이터를 ground truth로 하여 low image를 생성하여 모델에서 사용할 이미지를 제작하여야 했다. 한편, 현재의 CNN-based SR network는 대다수가 합성하여 만들어진 LR image에 의존하는 경향이 있다[6,9]. 대부분의 기술들은 bicubic interpolation을 수행하여 LR image를 만든다. 이는 실제 세상에서 벌어지는 다양한 image blur의 원인들을 다 cover하며 만들지 못해, 해당 image로 만들어진 모델들이 학습 및 검증 과정에서는 실제로는 좋은 성능을 낸다고 하더라도, LR image를 타당하게 생성하지 못하는 경우 실제로 적용했을 경우 그러지 못한 경우 또한 존재한다[7,8]. 따라서 기존의 간단한 bicubic interpolation의 한계를 극복할 수 있을, 다양한 interpolation task들을 융합해 low resolution image를 생성할 수 있도록 Data Augmentation을 수행해 보도록 한다.

4. 실험 및 분석

기존의 mDCSRN model이 250000 step만큼 pretrained model을 baseline으로 활용하였으며 이를 기반으로 모델의 특정한 부분들을 수정하며 성능 개선을 시도하였다. Data는 Human Connectome Project (HCP) Data 중 MIR Brain image를 필요로 하였고 이에 따라 Laboratory Of Neuro Imaging(LONI) 기관에 데이터에 접근할 수 있는 권한을 받아 데이터를 다운로드할 수 있었다. 총 744장의 MRI Brain image data를 이용하였으며 (256,320,320)의 shape을 가지는 한 장의 image에서 (64,64,64)의 4 patches를 얻을 수 있었다. 이 중 training set으로 total data의 약 75%인 572장을 사용, validataion set으로 10%인 81장, evaluation set으로 10%인 81장, test set으로는 10장의 image를 설정해 수행하였다.

4.1 Data Augmentaion

1) 기존의 코드에서는 fft를 수행한 후, (256,320,320) shape의 data에서 각 중심을 기준으로 ± 20 pixel만큼 zero 값을 설정해 고주파 성분을 제거해, low resolution 이미지를 만들었다. 일차적으로 수행했던 Data Augmentaion은 fft만을 수행하는 대신, fftshift의 비율을 random하게 섞어 좀 더 모델에게 다양한 데이터를 공급할 수 있도록 하였다.

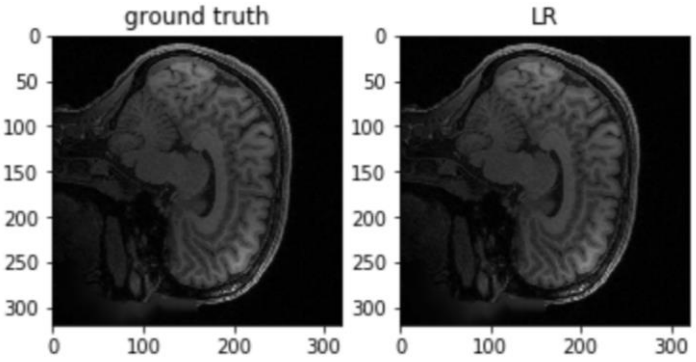

그림 8. fft와 fftshift를 난수의 비율로 섞어 생성한 LR image

첫 번째 방법으로 Data Augmentation을 수행한 결과, 그림8과 같이 육안으로 확인하기에 매우 차이가 미미했다. 따라서 좀 더 육안으로 확인이 가능한 Data Augmentaion 방안을 찾아보도록 했다.

2) 두번째로는, 기존의 방법대로 fft를 수행한 후, 추가로 Gausisan filter를 적용해 LR 이미지를 만드는 방법을 택했다. Gaussian filter는 고대역 부분은 약하게 통과시키고 고대역 부분은 강하게 통과시키는 저역통과필터의 역할로 사진의 윤곽선들을 뭉개주는 효과가 있다.

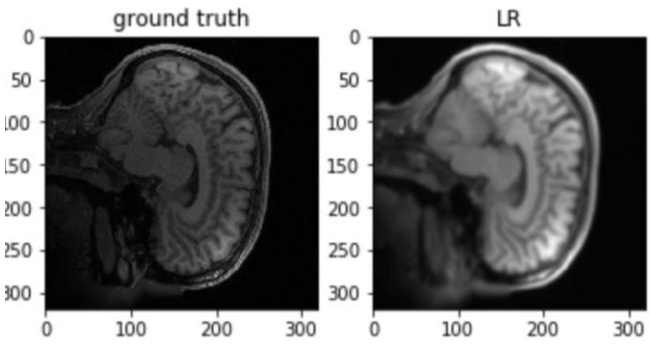

그림 9 . fft와 Gaussian filter를 통한 LR image

해당 방법을 사용한 결과 그림9와 같은 결과를 얻을 수 있었다. 육안으로 보기에도 효과적인 결과를 확인할 수 있었다. 한편, 실제로 모델을 학습 중에 개선되고 있는 이미지를 확인하기 위해 중간 결과를 살펴보았다.

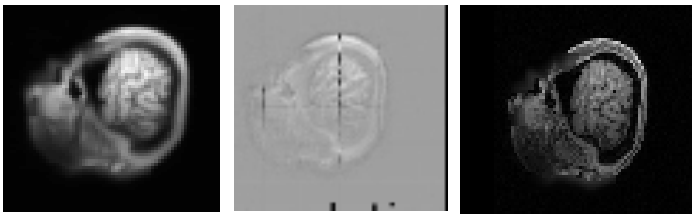

그림 10 . LR data from Gaussian filter로 학습중인 과정

그림10은 Gaussian filter를 활용하여 학습되고 있는 모델의 과정을 보여준다. 그림10의 왼쪽 사진이 Gaussian filter로 생성한 low image이고, 오른쪽 사진이 ground truth image 이며, 가운데 image가 학습 중에 중간 확인을 위해 저장한 resolution image이다. 어떤 문제인지 모르겠지만 Gaussian filter를 통한 LR image는 LR image로서 활용하기에 문제가 있다 판단되어 다른 Augmentation 방법을 사용해 보기로 하였다.

3) 효과적인 LR data 생성을 위해 여러 방안을 수행하다가 1)에서 data가 ground truth image와 큰 차이가 없는 것이 사진의 size에 비해 고주파 성분이 덜 제거되었기 때문이라는 결론이 섰다. 따라서 고주파 성분을 제거하는 zeroing 과정의 범위를 160으로 확장하여 진행하였다. 더불어 1)과 같은 방안으로 fftshift를 난수 비율로 적용하여 fft와 함께 적용하였다.

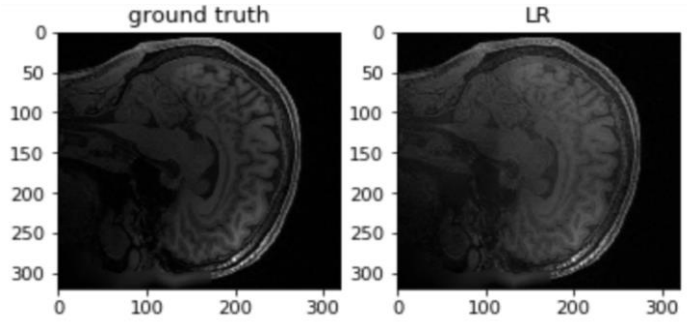

그림 11. zeroing range를 확장 수행한 fft + fftshift를 통한 LR image 생성

그림11을 통해 확인할 수 있듯이 육안으로 확인이 가능할 정도의 low resolution image가 생성됨을 확인할 수 있었다.

4) 다음은 fftn → fftshift → ifftshift → ifftn 을 통해 low resolution image를 생성한 결과이다.

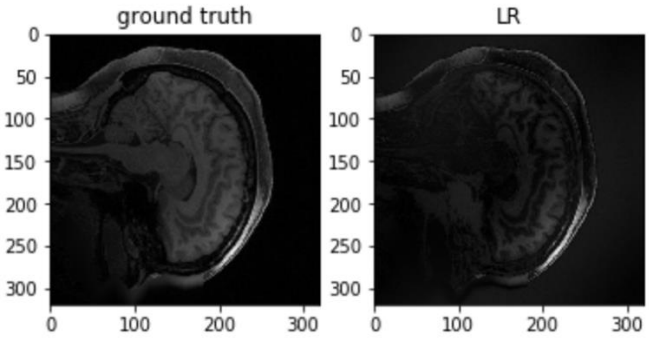

그림 12. fftn + fftshift를 활용한 LR image

육안으로 확인해 보기에는 각 픽셀의 색깔만 변하는 것으로 확인되어 보인다.

5) 다음은 fftshift → fftn → fftshift → ifftshift → ifftn → fftshift 을 통해 low resolution image를 생성한 결과이다.

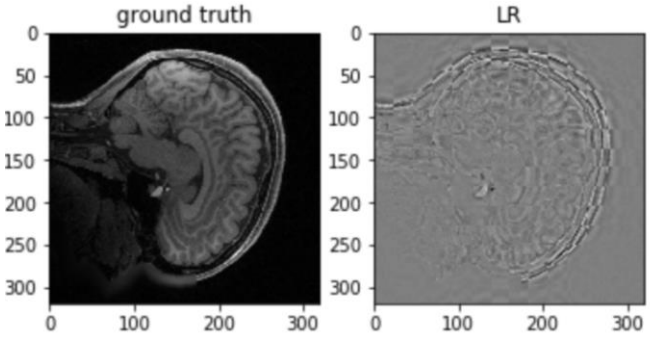

그림 13. fftshift + fftn + fftshift를 활용한 LR image

5)의 방법을 통해 얻은 결과는 그림13과 같았다. 이는 HR image와는 상당히 거리가 있어 학습 데이터로 적합하지 않다 판단되었다.

이와 같은 과정을 거치며 이 중 가장 학습 데이터로 적합하다고 판단된 3)의 방안으로 Data Augmentation을 수행하여 low resolution image를 생성하여 모델 학습을 진행하였다.

4.2 model size

앞에서 제안한 모델 압축 과정을 통해 모델의 사이즈는 아래와 같이 변함을 확인할 수 있었다.

	mDSCRN-WGAN	mDCSRN-WGANGP
#param of generator	0.412M	**0.412M**
#param of discriminator	31.156M	**17.629M**

표1. 모델 압축에 따른 파라미터 수 변화

표1은 위와 같이 이번 모델에서 제안하는 모델 압축에 따른 파라미터 수의 변화이다. 왼쪽이 기존 모델의 파라미터

수이고, 오른쪽이 프로젝트에서 제안한 모델의 파라미터 개수이다. 1x1 convolution을 수행함으로써 Discriminator의 파라미터의 개수를 기존의 절반 가량을 감소시킬 수 있었다.

4.3 Resolution

이번 프로젝트를 수행하면서 크고 작은 error들을 수정하면서 model을 끝없이 수정해왔다. 구글링을 하며 error들을 잡아나가는 데에 엄청난 시간을 투자하며 프로젝트를 진행해 온 중, 끝내 현재의 조건에서 해결하기 버거운 error를 접하게 되었다.

그림 14. CUDA out of memory RuntimeError

그림 14는 기존의 mDCSRN이 250000step 만큼 학습된 것을 기반으로, WGANGP와 model compression을 통해 255000 step을 거치던 중 발생한 error이다. 발생한 error를 살펴보면, CUDA out of memory. Tried to allocate 128.00 MiB (GPU 0; 15.90 GiB total capacity; 15.14 GiB already allocated; 37.88 MiB free; 24.10 MiB cached) 이다. error를 읽어보면 타당한 것으로 보이지만, Google colab의 RAM의 사용공간을 보면 여유롭게 남아있어 의문이었다. 이런 error가 일어나, 경희대학교 전자정보대학 B06 융합실습실 대여해 그 곳에서 모델 학습을 시도하였다.

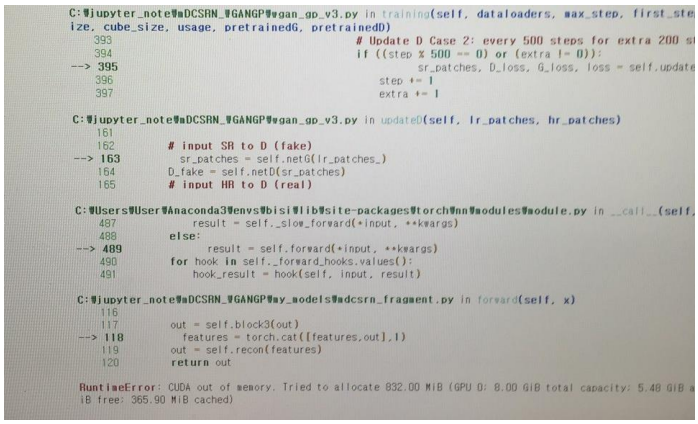

그림 15. CUDA out of memory RuntimeError

그러나 그림15와 같이 B06 융합실습실에서도 같은 error를 접해야 했다. 따라서 아쉬운대로 255000 step 학습 상태에서의 결과를 도출해 낸다.

1) image 결과

그림 16. step 255000 image 결과

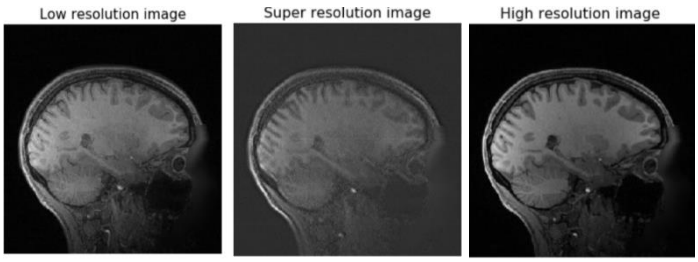

그림 17. step 255000 image 결과

2) SSIM, PSNR, NRMSE 결과

	mDCSRN (250000 step)			**mDCSRN-WGANGP (255000 step)**		
metric	SSIM	PSNR	NRMSE	**SSIM**	**PSNR**	**NRMSE**
MEAN	0.581	20.210	0.828	**0.620**	**20.730**	**0.788**
STD	0.133	2.496	0.155	**0.123**	**2.838**	**0.181**

표2. metric에 따른 결과

SSIM(Structural Similarity)은 두 이미지의 구조적 유사성을 비교하는 지표로 이미지 혹은 비디오의 객관적인 품질을 측정하는 지표로 활용되고 있고 PSNR(Peak Signal-to-Noise Ratio)은 최대신호대잡음비로, 또한 객관적인 품질을 측정하는 지표로 널리 활용되고 있다. NRMSE(Normalized Root Mean Squared Error)는 모델에 대한 평균적 오차를 나타내는 지표이다.SSIM과 PSNR은 값이 클수록 성능이 개선됨을 의미하며, NRMSE는 값이 작을수록 모델이 개선됨을 의미한다. 위의 그림16과 17을 보면, super resolution 과정의 image가 low resolution image에 비해 명확하지 않음이 시각적으로 확인된다. 해당 이미지는 아직 모델의 학습 중에 있는 이미지로 뚜렷한 결과를 판단하기에는 무리가 있다고 판단된다. (또는 Data preprocess 과정에서, 특히 low resolution 생성 과정에서 Data Augmentation 과정에서의 특이점으로 인한 것으로 추측된다.) 한편, 성능 상의 값을

확인한다면 SSIM, PSNR, NRMSE에서 보다 나은 값을 가짐을 확인할 수 있다. 그림16,17에 따른 결과는 학습 중에 저장된 이미지를 가져온 것이고, 표의 결과는 test set의 결과에 의한 지표들을 나타낸 값이다.

5. 결론

이번 프로젝트에서는 크게 Data Augmentation, model compression, gradient penalty 적용 등 크게 3가지의 model 개선을 수행하였는데, 결과적으로 최종 학습된 모델을 완성시키지 못해 많은 아쉬움이 남는다. 하지만 중간 결과의 지표상에서 나온 값을 확인해 볼 때, model이 개선되고 있음을 확인해 볼 수 있었고, 이에 따라 최종적인 model과 비교해봐도 개선된 성능을 예상해 볼 수 있을 것이라 생각된다.

참 고 문 헌

[1] E. Plenge, D. H. J. Poot, M. Bernsen, G. Kotek, G. Houston, P.Wielopolski, L. V. D. Weerd, W. J. Niessen, and E. Meijering,“Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?” Magnetic Resonance in Medicine, vol. 68, no. 6, pp. 1983–1993, Jan. 2012

[2] Chen, Yuhua, et al, "Brain MRI super resolution using 3D deep densely connected neural networks," Proceedings of 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 739-742.

[3] Chao Dong and Chen Change Loy and Kaiming He and Xiaoou Tang.(2014). Image Super-Resolution Using Deep Convolutional Networks.

[4] Chen, Yuhua, et al, “Efficient and Accurate MRI Super-Resolution using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network”

[5] https://github.com/hz2538/E6040-super-resolution-project

[6] Radu Timofte, Eirikur Agustsson, Luc Van Gool, MingHsuan Yang, Lei Zhang, Bee Lim, et al. NTIRE 2017 challenge on single image super-resolution: Methods and results.In The IEEE Conference on Computer Vision and Pattern
Recognition Workshops, July 2017.

[7] Netalee Efrat, Daniel Glasner, Alexander Apartsin, Boaz Nadler, and Anat Levin. Accurate blur models vs. image priors in single image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 2832–2839, 2013

[8] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 945–952, 2013.

[9] Ruofan Zhou and Sabine Susstrunk. Kernel Modeling Super-Resolution on Real Low-Resolution Images