PG-GAN을 이용한 패션이미지 데이터 자동 생성

김양희, 이찬희, 황태선, 김경민, 임희석*
고려대학교 정보대학 컴퓨터학과

Automaitc Generation of Fashion Image Dataset by Using Progressive Growing GAN

Yanghee Kim, Chanhee Lee, Taesun Whang, Gyeongmin Kim, Heuiseok Lim*
Department of Computer Science Engineering, College of Informatics, Korea University

요 약 이미지와 같은 고차원 데이터로부터 새로운 샘플 데이터를 생성하는 기술은 음성 합성, 이미지 변환 및 이미지 복원 등에 다양하게 활용되고 있다. 본 논문은 고해상도의 이미지들을 생성하는 것과 생성한 이미지들의 variation을 높이기 위한 방안으로 Progressive Growing of Generative Adversarial Networks(PG-GANs)을 구현 모델로 채택하였고, 이를 패션 이미지 데이터에 적용하였다. PG-GANs은 생성자(Generator)와 판별자(discriminator)를 동시에 점진적으로 학습하도록 하는데, 저해상도의 이미지에서부터 계속해서 새로운 레이어들을 추가하여 결과적으로 고해상도의 이미지를 생성할 수 있게끔 하는 방식이다. 또한 생성 데이터의 다양성을 높이기 위하여 미니배치 표준편차 방법을 제안하였고 GAN 모델을 평가하기 위한 기존의 MS-SSIM이 아닌 Sliced Wasserstein Distance(SWD) 평가 방법을 제안하였다.

주제어 : 딥러닝, 이미지, 생성 모델, 패션 기술, GAN

Abstract Techniques for generating new sample data from higher dimensional data such as images have been utilized variously for speech synthesis, image conversion and image restoration. This paper adopts Progressive Growing of Generative Adversarial Networks(PG-GANs) as an implementation model to generate high-resolution images and to enhance variation of the generated images, and applied it to fashion image data. PG-GANs allows the generator and discriminator to progressively learn at the same time, continuously adding new layers from low-resolution images to result high-resolution images. We also proposed a Mini-batch Discrimination method to increase the diversity of generated data, and proposed a Sliced Wasserstein Distance(SWD) evaluation method instead of the existing MS-SSIM to evaluate the GAN model.

Key Words : Deep Learning, Image, Generative Model, Fashion Technology, Generative Adversarial Networks

1. 서론

이미지와 같은 고차원 데이터로부터 새로운 샘플 데이터를 생성하는 기술은 음성 합성, 이미지 변환 및 이미지 복원 등에 다양하게 활용되고 있다. 이러한 이미지 생성 기술은 원본 데이터로부터 적절한 특징을 추출하고, 추출된 데이터의 특징을 공유하는 새로운 이미지를 생성하는 데에 주안점을 둔다. 현재 가장 보편적인 접근법으

*교신저자 : 임희석(limhseok@korea.ac.kr)
접수일 2018년 06월 24일 수정일 2018년 08월 28일 심사완료일 2018년 09월 16일

로는 VAE (Variation Autoencoders)[1] 와 GAN (Generative Adversarial Networks)[2]을 들 수 있는데, 본 연구에서는 원본 데이터의 특징과 해상도를 유지하면서 다양한 결과물을 효과적으로 생성할 수 있는 접근법을 모색하였다. 이를 위해 기존 GAN을 변형 및 발전시킨 PG-GANs (Progressive Growing of GANs)을 기존 PG-GANs모델의 학습 데이터가 아닌, 패션 데이터를 이용하여 구현하였다. PG-GANs[3]의 핵심 통찰력은 모델의 생성자와 판별자를 저해상도부터 고해상도까지 점차적으로 성장시킬 수 있다는 점에 있다. 훈련이 진행됨에 따라 고해상도의 세부 정보를 도입하기 위한 새로운 레이어가 추가되는 방식을 채택하였는데, 이러한 방식은 훈련 과정이 가속화되고 고해상도에서의 안정성이 향상된다는 점에서 기존 이미지 생성 모델들의 단점을 보완할 방안으로써 의의가 있다. 이에 본 논문에서는 PG-GANs의 이미지 생성 및 판별 방법을 소개하고, 패션 데이터를 이용한 PG-GANs 구현 과정을 밝힌 후, 구현 실험의 평가 결과와 의의를 제시하고자 한다.

2. 관련 연구

2.1 PG-GANs과 기존 생성 모델 비교

PG-GANs(Progressive Growing of GANs)은 준지도 학습(Semi-supervised Learning) 방식을 채택하며, 기본적으로 GAN(Generative Adversarial Networks) 모델과 같이 생성자(Generator)와 판별자(Discriminator)로 구성되어 학습을 진행하는 생성 모델이다.

기존의 생성 모델 접근법 중 널리 쓰이는 모델로 VAE(Variation Autoencoders)와 GAN(Generative Adversarial Networks)을 들 수 있다. VAE의 경우 훈련하기는 쉽지만 모델의 제한으로 인해 흐릿한 이미지를 생성하는 경향을 보인다[4]. GAN은 VAE보다 선명한 이미지를 만들어 내지만, 최근의 기술적 발전에도 불구하고 훈련이 계속 불안정하다는 단점이 있다[5]. 여러 생성 모델의 방식을 결합한 하이브리드 기법은 각 모델의 장점을 결합하였음에도 이미지 품질면에서는 GAN보다 뒤떨어진다는 한계를 보인다[6].

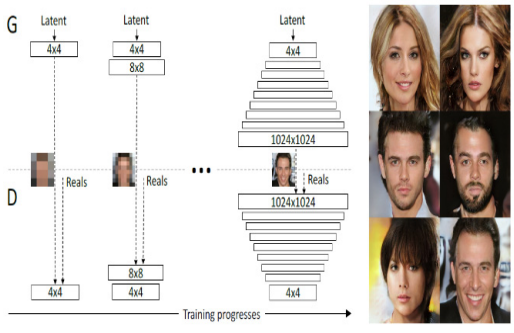
PG-GANs은 생성자와 판별자의 두 네트워크로 이루어진 GAN의 구성을 따른다. 생성자는 잠재 코드로부터 원본 샘플과 같은 이미지를 생성하며, 이들 이미지의 분포는 이상적으로는 훈련 데이터의 분포와 구별될 수 없을 정도로 동일해야 한다. 판별자 네트워크는 이렇게 생성된 이미지와 원본 이미지 간의 차이를 구별하도록 훈련된다.

이러한 구조에는 여러가지 잠재적인 문제점이 있다. 훈련 데이터의 분포와 생성된 데이터의 분포 사이의 거리를 측정할 때, 분포가 실질적으로 겹치지 않으면, 즉, 구별하기가 쉬운 경우, 그래디언트가 적절한 방향을 가질 수 있다. 초기에는 Jensen-Shannon divergence가 거리 측정 방법으로 사용되었지만[2], 최근에 그 공식화가 개선되었고, 최소 제곱[7], 마진과 절대 편차[8], Wasserstein distance[9] 등이 우수한 성능을 보였다.

PG-GANs은 이러한 접근 방법과는 상당 부분 다른 방식을 채택한다. 해상도가 높을수록 생성된 이미지를 훈련 이미지 와 구분하는 것이 쉬워지므로 그래디언트를 대폭 증폭하는 문제가 생기고 결과적으로 고해상도 이미지 생성이 어려워진다. 또한 큰 해상도는 메모리 제약으로 인해 더 작은 배치 크기를 사용해야 하므로 훈련 안정성이 저하될 수 있다.

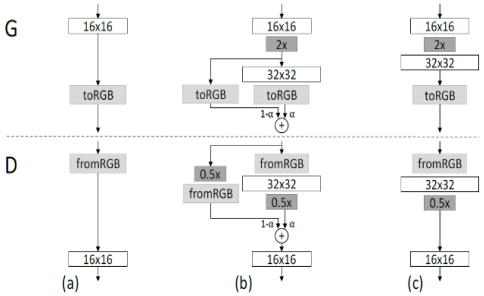
2.2 PG-GANs의 구조 및 특징

이러한 단점을 보완하면서 등장한 PG-GANs은 저해상도의 이미지에서 시작하여 생성자와 판별자를 점진적으로 성장시킬 수 있다는 특징을 지닌다. [Fig. 1]에서 볼 수 있듯이 PG-GANs 모델은 훈련이 진행됨에 따라 고해상도의 세부 정보를 도입하는 새로운 레이어를 추가되는데, 이것은 전체 훈련을 크게 가속화하고 고해상도에서 안정성을 향상시킨다는 장점이 있다.



[Fig. 1] Model structure of the Progressive Growing of GANs

PG-GANs에서는 생성자 네트워크와 판별자 네트워크가 항상 동기화되어 성장한다. 두 네트워크의 기존 계층은 모두 훈련 과정 전반에 걸쳐 점진적 훈련이 가능하다. 학습 과정에서 낮은 해상도의 이미 잘 훈련된 레이어가 갑작스러운 충격을 받지 않도록 하기 위해 [Fig. 2]와 같이 새로운 레이어가 네트워크에 추가되면 부드럽게 페이드인 한다.



[Fig. 2] Transition of the Generator's and the Discriminator's layers

이와 같은 구조로 저해상도에서부터 점진적인 훈련이 가능해지면, 초기 클래스 정보가 적고 모드가 적다는 점에서 작은 이미지를 생성하는 데에도 훨씬 안정적이다. 또 다른 이점은 훈련 시간의 감축이다. PG-GANs의 경우 학습 과정의 반복이 대부분 저해상도에서 이루어지며, 비슷한 품질의 결과가 최종 출력 해상도에 따라 2-6배 빠르게 출력된다.

기존 Generative Adversarial Networks (GAN)기반 모델들의 경우, 생성자가 일부 모드에 한정된 샘플만을 포착하기 때문에 생성되는 데이터의 다양성(variation)에 한계가 생긴다. Salimans은 이에 대한 해결책으로 "미니 배치 판별(Mini-batch Discrimination)"을 제안했다[5]. 판별자의 끝단에 미니 배치 레이어를 모두 추가함으로써 판별자가 미니 배치 전체에 대한 통계를 낼 수 있고, 생성자는 이를 바탕으로 보다 다양한 샘플을 생성해낼 수 있게 된다.

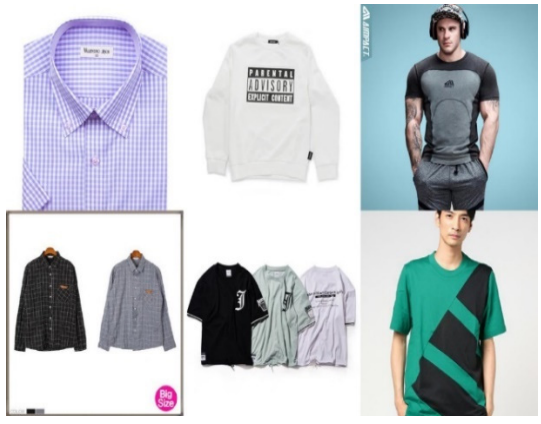
PG-GANs은 미니 배치에 대한 표준편차를 이용하여 이 접근 방식을 크게 단순화하는 동시에 다양성을 개선하였다. 먼저 미니 배치를 통해 각 공간 위치의 각 특성(feature)에 대한 표준편차를 계산하였고, 모든 기능 및 공간 위치에 대해 이러한 예상치를 평균하여 단일 값에 도달하도록 하였다. 그 후 이러한 값을 모두 복사하고 모든 공간 위치와 미니 배치를 연결하여 또 하나의 feature map을 구성하였다.

3. 패션 데이터를 이용한 PG-GANs 구현

3.1 모델 학습

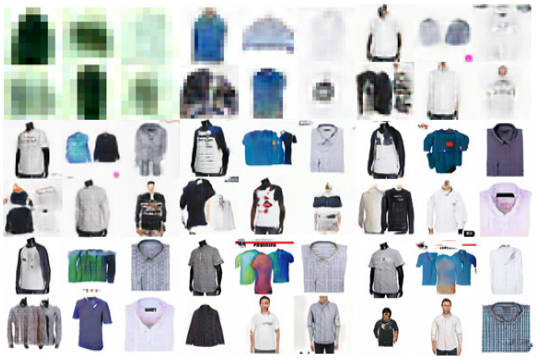
기존 Progressive Growing of GANs 모델의 경우 CelebA-HQ 데이터 셋을 사용하여 Tesla V100 GPU 8개를 동반한 Nvidia DGX-1를 이용해 2일간 학습을 진행하였고 1024 x 1024의 해상도를 가지는 총 30,000개의 연예인 얼굴 사진들을 학습 데이터로 사용하였다.

본 실험에서는 네이버를 이용해 크롤링을 진행하여 모은 약 60,000장의 남자 상의 의류 이미지를 새로운 데이터셋으로 사용하였다. 약 60,000장 중 512 x 512 이상의 해상도를 가지는 이미지들을 샘플링하여 총 29,000장의 512 x 512 해상도의 정사각형 형태의 데이터셋을 구축하였다. 구축한 데이터셋의 예시는 [Fig. 3]과 같다.



[Fig. 3] Example of 512 x 512 men's clothing-tops images

본 실험에서는 Nvidia Geforce GTX 1080 Ti GPU 4대를 이용하여 약 2일 3시간 가량 학습을 진행하였고, 모델에게 총 8,861,800장의 이미지를 보여주었다. [Fig. 4]는 모델이 학습을 진행하면서 생성한 가짜 이미지들을 나타낸 것이다.



[Fig. 4] Fake images generated through the model training

3.2 평가 방법

PG-GANs은 기존의 모델 평가 방법이었던 MS-SSIM[10]의 한계를 보완하고자 다중 표준 통계 유사도(Multi-Scale Statistical Similarity)를 고려한 Sliced Wasserstein Distance(SWD) 기법을 평가 방식으로 채택하였다. 이에 본 실험 역시 PG-GANs의 방식과 동일한 SWD 기법을 모델 평가 방법으로 이용하였다.

기존의 GAN 모델들을 서로 비교 평가하기 위해서는 Multi-Scale Statistical Similarity 방법이 사용되었다. 한 GAN의 결과를 다른 GAN의 결과와 비교하려면 많은 수의 이미지를 조사해야 하는데, 그 과정은 주관적일 수 있으며 그 자체로 어려운 일이다. 따라서 객관적인 평가를 위해서는 이미지를 대량 수집하여 일부 지표를 계산하는 자동화된 방법에 의존하는 것이 바람직하다. MS-SSIM과 같은 기존의 방법은 대규모 모드가 안정적으로 축소되지만 색상이나 텍스처의 다양성 손실과 같은 세부적인 효과에 반응하지 않으며 이미지를 직접 평가하지도 않는다는 사실을 발견하였다.

PG-GANs은 라플라시안 피라미드[11]에서 생성된 로컬 이미지 패치의 생성과 대상 이미지의 분포 사이의 다중 표준 통계 유사도(Multi-Scale Statistical Similarity)를 고려하여 16 x 16 픽셀의 저해상도에서 시작하여 이를 연구할 것을 제안하였다. 표준 실습에 따라 피라미드는 전체 해상도에 도달할 때까지 점차적으로 두 배에 이르렀으며, 각 연속 레벨은 이전 레벨의 업 샘플링 된 버전과의 차이를 인코딩하였다.

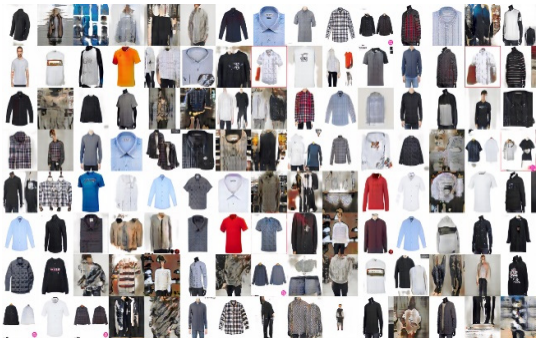
단일 라플라시안 피라미드 레벨은 특정 공간 주파수 대역에 해당한다. PG-GANs 평가 연구에서는 16384개의 이미지를 무작위로 샘플링하고 라플라스 피라미드의 각 레벨에서 128개의 디스크립터를 추출하여 레벨당 2.1M개의 디스크립터를 제공하였다. 각각의 디스크립터는 $x \in 7\times7\times3 = R147$으로 표시되는 3개의 컬러 채널을 가진 7 × 7 픽셀이다. 각 색상 채널의 평균 및 표준 편차를 계산한 다음 512 투영을 사용하여 효율적으로 계산 가능한 무작위 근사 대 지구 이동 거리 인 슬라이더 Wasserstein 거리 SWD({xli}, {yli})를 계산하여 통계적 유사성을 추정하는 방식을 채택하였다[12].

직관적으로 작은 Wasserstein 거리는 패치의 분포가 유사함을 나타낸다. 즉, 트레이닝 이미지와 생성자 샘플은 이 공간 해상도에서 모양과 다양성 모두 유사하게 나타난다. 특히, 가장 낮은 해상도의 16 x 16 이미지에서 추출한 패치 세트 사이의 거리는 대규모 이미지 구조에서 유사성을 나타내지만 가장 미세한 패치는 가장자리 및 노이즈의 선명도와 같은 픽셀 수준 속성에 대한 정보를 인코딩한다.

3.3 성능 평가

남자 상의 의류를 사용한 본 구현에서는 약 860만장의 이미지를 모델에게 학습시킨 후에, 평가를 진행하였다.

평가 방법으로는 Sliced Wasserstein Distance (SWD)를 채택하였으며, 모델의 해상도별 레이어마다 평가를 진행하였다. 아래의 [Fig. 5]는 최종 모델이 학습을 완료한 후 랜덤으로 생성한 이미지를 보여준다.



[Fig. 5] Fake images randomly generated by the model

기존의 CelebA-HQ 데이터 셋을 이용한 모델의 경우, 모델이 수렴할 때를 기준으로 SWD 2에서 4 정도의 값을 도출할 수 있었다. 하지만 남성 상의 의류 이미지를 사용하여 학습한 모델의 경우 [Table. 1]과 같이 대부분의 레이어에서 SWD 약 12에서 13의 성능이 나왔고, 이미지

32 레이어와 64 레이어의 경우 약 5~6의 성능을 보여주었다.

본 구현 모델의 학습 데이터로 패션 데이터를 이용할 때와 CelebA-HQ를 이용할 때 성능 면에서 상당한 차이가 있음을 확인하였다.

이는 남성 상의 의류 이미지의 경우 연예인 얼굴 사진보다 다양한 형태의 이미지들이 존재하기 때문에 모델 학습에 한계가 존재하는 것으로 보인다.

같은 데이터셋의 데이터라 하더라도 상의 의류를 착용 중인 사람의 사진, 마네킹 사진, 한 사진에 여러 벌의 의상이 있는 사진이 포함되는 등 각 데이터의 분포가 다양한 것이 성능 차이의 원인이 되는 것으로 보인다.

[Table. 1] Evaluation results of the model trained using the men's clothing-tops image dataset

	SWDx1e3 _512	SWDx1e3 _256	SWDx1e3 _128	SWDx1e3 _64	SWDx1e3 _32	SWDx1e3 _16	SWDx1e3 _avg
8,801k	16.8281	16.0269	16.1470	13.2497	9.1864	17.9320	14.8950
8,601k	**13.4479**	**12.9347**	**13.7200**	10.3544	7.1607	21.5755	13.1988
8,401k	18.2637	19.6419	15.3039	10.8330	7.5277	19.4553	15.1709
8,200k	19.6092	18.3970	11.1272	**6.7871**	**5.1076**	15.7997	**12.8046**
8,000k	25.4830	13.5645	11.0334	9.2138	6.9981	23.5873	14.9800
7,800k	26.4956	15.1836	13.0970	10.0377	6.8637	**14.5639**	14.3736
7,400k	23.6944	14.1429	8.7995	6.8641	6.4266	26.1033	14.3385
7,000k	36.2444	16.9220	10.8440	6.7701	5.9924	21.8302	16.4339

4. 결론 및 향후 연구 방향

PG-GANs은 일반적으로 초기 GAN 연구에 비해 높은 성능을 보였고, 훈련 과정이 안정적이라는 장점이 있지만 사실적인 이미지를 구현하는 데에는 여전히 한계를 지닌다. 이를 극복하기 위해서는 모델의 훈련이 곡선이 아닌 직선 형태의 객체에 특화되는 등, 주어진 데이터셋에만 한정된다는 점을 이해하는 것이 중요하다. 또한 생성 이미지의 세부 구조에 대해서도 개선할 여지가 있다. 즉, PG-GANs의 모델 개선 방향은 이미지의 리얼리즘을 향상시키는 데에 있다.

이러한 한계는 남성 상의 이미지를 이용한 구현 과정에서도 드러났다. 생성 이미지 중 일부는 훈련 데이터의 속성을 잘 추출하였음에도 전체 텍스처가 뭉개지는 등 사실적이지 않은 결과를 도출하기도 하였다. [Table. 1]에서도 볼 수 있듯이 특정 레이어에서 성능이 급격하게 떨어지는 경향을 보이는데, 모델의 현실적인 이미지 생성을 위해 이러한 resolution transition 과정을 세밀하게 조정하는 방안을 마련한다면 보다 향상된 결과를 얻을 수 있을 것으로 기대한다.

REFERENCES

[1] Durk.P. Kingma, Shakir Mohamad, Danilo Jimenez Rezende and Max Welling, "Semi-supervised Learning with Deep Generative Models," Advances in Neural Information Processing Systems (NIPS) 27, 2014.

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, "Generative Adversarial Nets," Advances in Neural Information Processing Systems (NIPS) 27, 2014.

[3] Tero Karras, Timo Aila, Samuli Laine and Jaakko Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," ICLR, 2018.

[4] Xi Chen, Diederik P. Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever and Pieter Abbeel, "Variational Lossy Autoencoder," arXiv.org, 2016.

[5] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford and Xi Chen, "Improved Techniques for Training GANs," Advances in Neural Information Processing Systems (NIPS) 29, 2016.

[6] Alireza Makhzani and Brendan Frey, "PixelGAN Autoencoders," University of Toronto, 2017.

[7] Song Han, Xingyu Liu, Huizi Mao, Jing Pu, Ardavan Pedram, Mark A. Horowitz and William J. Dally, "EIE: Efficient Inference Engine on Compressed Deep Neural Network," 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA), 2016.

[8] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang and Jiaya Jia, "Pyramid Scene Parsing Network," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2881-2890, 2017.

[9] Martin Arjovsky, Soumith Chintala and Léon Bottou, "Wasserstein Generative Adversarial Networks," Proceedings of the 34th International Conference on Machine Learning (PMLR), Vol.70, pp.214-223, 2017.

[10] Augustus Odena, Christopher Olah1 and Jonathon Shlens1, "Conditional Image Synthesis with Auxiliary

Classifier GANs," Proceedings of the 34th International Conference on Machine Learning (ICML'17), Vol.70, pp.2642-2651, 2017.
[11] Peter J. Burt and Edward H. Adelson, "Method for compensating for void-defects in images," US Patent, 1987.
[12] Julien Rabin, Gabriel Peyré, Julie Delon and Marc Bernot, "Wasserstein Barycenter and Its Application to Texture Mixing," International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), pp.435-446, 2011.

김 양 희(Kim, Yanghee) [정회원]

- 2018년 2월 : 한국외국어대학교 이탈리아어학과(문학사)
- 2018년 4월 ~ 현재 : 고려대학교 Human Inspired AI & Computing 연구센터 소속 연구원

<관심분야>
Deep Learning, NLP, 뇌신경 언어 정보처리

이 찬 희(Chanhee Lee) [정회원]

- 2016 서강대학교 컴퓨터공학심화(학사)
- 2016~현재 고려대학교 컴퓨터학과 석박사 통합 과정

<관심분야>
인공지능, 자연어처리, 딥러닝

황 태 선(Whang, Taesun) [정회원]

- 2016년 : 인하대학교 산업경영공학과(공학사)
- 2018년 3월 ~ 현재 : 고려대학교 컴퓨터학과 석사과정

<관심분야>
Deep Learning, NLP, Dialogue System

김 경 민(Kim, Gyeongmin) [정회원]

- 2017년 8월 : 백석대학교 정보통신학부 (공학사)
- 2018년 3월 ~ 현재 : 고려대학교 컴퓨터학과 석사과정

<관심분야>
Deep Learning, NLP

임 희 석(Lim, Heuiseok) [정회원]

- 1992년 : 고려대학교 컴퓨터학과(이학학사)
- 1994년 : 고려대학교 컴퓨터학과(이학석사)
- 1997년 : 고려대학교 컴퓨터학과(이학박사)
- 2008년 ~ 현재 : 고려대학교 컴퓨터학과 교수

<관심분야>
Deep Learning, NLP, 뇌신경 언어 정보처리