일반논문 (Regular Paper)
방송공학회논문지 제25권 제1호, 2020년 1월 (JBE Vol. 25, No. 1, January 2020)
https://doi.org/10.5909/JBE.2020.25.1.1
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

주목 메커니즘 기반의 멀티 스케일 조건부 적대적 생성 신경망을 활용한 고해상도 흉부 X선 영상 생성 기법

안 경 진[a),b)], 장 영 걸[a),b)‡], 하 성 민[a),c)], 전 병 환[a),b)], 홍 영 택[a),b)], 심 학 준[a)], 장 혁 재[d)]

Generation of High-Resolution Chest X-rays using Multi-scale Conditional Generative Adversarial Network with Attention

Kyeongjin Ann[a),b)], Yeonggul Jang[a),b)‡], Seongmin Ha[a),c)], Byunghwan Jeon[a),b)], Youngtaek Hong[a),b)], Hackjoon Shim[a)], and Hyuk-Jae Chang[d)]

요 약

의료분야에서 질환별 유병률 차이로 인한 데이터 수적 불균형은 흔하게 발생되는 문제로 인공지능 학습 성능을 저하시켜 개발의 어려움을 초래한다. 최근 이러한 데이터 수적 불균형문제를 해결하기 위한 한 방법으로 적대적 생성 신경망(GAN) 기술이 도입되었고 다양한 분야에 성공적으로 적용되어왔다. 그러나 수적 불균형에 의해 저하된 성능 문제를 해결하는데 있어서 기존 연구들의 영상 해상도가 아직 충분하지 않고 영상 내 구조가 전역적으로 일관성 있게 모델링 되지 않아 좋은 결과를 얻기 어렵다. 본 논문에서는, 흉부 X선 영상 데이터의 수적 불균형문제를 해결하기 위하여 고해상도 영상을 생성할 수 있는 주목 메커니즘 기반 멀티 스케일 조건부 적대적 생성 네트워크를 제안한다. 해당 네트워크는 질환제어 조건변수에 의해 하나의 네트워크만으로 다양한 질환 영상을 생성할 수 있어 각 클래스별로 학습을 하는 비효율성을 줄였고, 자기 주목 메커니즘을 통해 영상 내 장거리 종속성 문제를 해결하였다.

Abstract

In the medical field, numerical imbalance of data due to differences in disease prevalence is a common problem. It reduces the performance of a artificial intelligence network, leading to difficulties in learning a network with good performance. Recently, generative adversarial network (GAN) technology has been introduced as a way to address this problem, and its ability has been demonstrated by successful applications in various fields. However, it is still difficult to achieve good results in solving problems with performance degraded by numerical imbalances because the image resolution of the previous studies is not yet good enough and the structure in the image is modeled locally. In this paper, we propose a multi-scale conditional generative adversarial network based on attention mechanism, which can produce high resolution images to solve the numerical imbalance problem of chest X-ray image data. The network was able to produce images for various diseases by controlling condition variables with only one network. It's efficient and effective in that the network don't need to be learned independently for all disease classes and solves the problem of long distance dependency in image generation with self-attention mechanism.

Keyword : Medical, Multi-scale cGAN, Chest X-rays, Self Attention, High-Resolution

Ⅰ. 서 론

1. 연구 배경

최근 인공지능 기술의 발전으로 다양한 분야에서 이를 접목해 비약적인 성능 향상을 이루고 있다[1]. 그 중 의료를 도메인으로 한 인공지능 기반의 많은 연구들이 활발히 진행되고 있는데, 이는 전 세계적인 고령화에 따른 인간의 기대수명 증가로 의료 수요가 확대되어 의료산업의 경제적 가치가 높아졌기 때문이다. 특히, 다양한 의료 융합연구들 중 심장, 폐를 대상으로 한 연구들이 많이 있다. 흉부 관련 질환은 환경적인 요인으로 인해 일반인들도 빈번하게 접할 수 있는 흔한 질환이지만 조기발견 후 적절한 조치가 이뤄지지 않으면 치명적일 수 있다. 때문에 타 기관에서 발생한 여러 질환들과 비교해봤을 때 상대적인 중요도가 커 흉부 기관 대상으로 한 다양한 연구가 이루어지고 있다.

심·폐질환을 조기에 발견하기 위한 일반적인 검사로는 흉부 X선이 있으며, X선을 흉곽 부위에 투과하여 심장과 폐 관련 질환 유무를 판단한다. 흉부 X선 검사는 전산단층촬영술(Computed Tomography, CT)이나 자기공명영상(Magnetic Resonance Imaging, MRI) 등 다른 고가검사와 비교하여 상대적으로 저렴하고 단시간에 촬영 가능하며 피폭 선량이 적다는 장점이 있다.

하지만 바쁜 의료현장에서 임상의들이 수많은 환자들의 영상을 정량적으로 정확하게 판독하는 것은 시간소모적인 작업이며 노동비용 또한 크다. 이를 보완하기 위해 최근에는 인공지능 기술을 활용하여 빠르고 정확하게 영상판독이 가능한 모듈에 대한 연구결과가 나오고 있다.

한편, 현재의 지도학습기반 인공지능 모델을 잘 학습하는데 있어 균형 있는(balanced) 데이터 셋은 학습 수렴에 있어 매우 중요하다. 그러나 질환 유병률 차이에 의해 수집된 의료 데이터는 대부분 수적으로 불균형하며, 전처리과정 없이 바로 학습 데이터로 사용할 경우 특정 질환에 과적합(overfitting)되어 원하지 않는 수렴 결과를 얻을 수 있다. 흉부 X선 영상의 경우도 마찬가지로 빈번히 발생하는 질환들(무기폐, 흉수, 침윤 등)은 학습이 잘 진행되는 반면, 그렇지 않은 질환(폐렴, 심장 비대 등)들의 경우 학습 성능이 상대적으로 좋지 못하다.

이러한 데이터 수적 불균형[2]으로 인한 학습 성능 저하 문제를 해결하고자 최근 다양한 데이터 증강(data augmentation) 기술들이 제안되었고, 성능의 향상이 있었다. 영상 데이터의 대표적인 데이터 증강방법은 상하좌우 반전, 밝기조절 등이 있으나, 이렇게 재구성된 영상 데이터는 표준 데이터를 기반으로 생성되기 때문에 적은 데이터에서는 성능 향상이 높지 않을 수 있다. 따라서 훨씬 더 광범위한 데이터 집합을 생성하기 위한 근본적인 해결책으로 표준데이터 수의 증가가 필요하다.

최근, 이러한 문제의 새로운 해결 방안으로 표준데이터 분포의 추정을 학습하는 적대적 생성 신경망(Generative Adversarial Network, GAN)[3]기반의 데이터 증강 기법이 소개되었다. 이 신경망은 다양한 사례에 적용되어 실제와 거의 유사한 거짓 데이터를 만들어 데이터의 수적 불균형 문제를 해결함으로써 분류 성능이 향상됨을 입증하였다. 그러나 목표 클래스가 다수 개일 경우 각 클래스별로 적대적 생성 신경망을 학습해야 하기 때문에 비효율적이며 영상 수가 극단적으로 제한적일 경우 학습이 거의 불가능하다는 문제가 있다.

2. 정의

본 논문에서는 14개의 질환에 대한 흉부 X선 영상 데이터 셋의 수적 불균형문제를 해결을 목적으로 단 하나의 네

a) 연세대학교 심장·혈관 ICT기술연구센터(Cardio-vascular ICT Research Center, Yonsei University)
b) 연세대학교 의과학과(Brain Korea 21 Project for Medical Science, Yonsei University College of Medicine)
c) 연세대학교 의과대학 생체공학협동과정(Graduate School of Biomedical Engineering, Yonsei University, College of Medicine)
d) 연세대학교 의과대학 세브란스병원 심장내과(Department of Cardiology, Yonsei University College of Medicine)
‡ Corresponding Author : 장영걸(Yeonggul Jang)
E-mail: jyg1722@gmail.com
Tel: +82-2-2227-9551
ORCID:https://orcid.org/0000-0002-5805-7494
※ This work was supported by Institute of Information & communications Technology Planning & evaluation(IITP) grant funded by the Korea government(MSIT) (No.2018-0-00861, Intelligent SW Technology Development for Medical Data Analysis).
※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (No.2018-0-00861, 의료데이터 분석 지능형 SW 기술개발)
· Manuscript received May 7, 2019; Revised September 17, 2019; Accepted November 6, 2019.

트워크만으로도 질환 별 특징이 잘 반영된 고해상도 영상을 생성할 수 있는 기법을 제안한다. 이전 연구와 달리, 제안된 네트워크는 입력으로 조건 변수를 추가하여 생성되는 영상의 목표 질환을 제어할 수 있어 학습의 비효율성을 제거하였을 뿐만 아니라 극단적으로 불균형한 경우 학습이 불가능한 문제를 해결하였다. 또한, 주목 메커니즘(attention mechanism)의 적용을 통해 국소적인 부분만이 아니라 영상 전체적으로 형태의 일관성이 보존되어 더 사실과 가까운 영상을 생성할 수 있도록 하였다.

Ⅱ. 관련 연구

Ian Goodfellow et al.,[3]에 의해 처음 제시된 적대적 생성 신경망은 기존 지도학습 중심의 학습 패러다임을 비지도 학습으로 바뀌도록 초석을 다진 연구로 큰 파급력을 가지며 관련 연구들을 활성화시켰다. 하지만 적대적 생성 신경망은 생성자와 판별자가 대립하며 학습하는 구도로 학습이 불안정하다는 단점이 있기 때문에 적대적 생성 신경망의 훈련 안전성을 높이기 위한 많은 연구가 진행되고 있다.

대표적인 연구로는 심층 합성곱 적대적 생성 신경망(Deep Convolutional GAN, DCGAN)[4]이 있는데, 이 신경망은 완전연결 신경망(Fully-Connected Network, FCN)을 심층 합성곱 신경망(Deep Convolution Neural Network)으로 대체하여 기존의 적대적 생성 신경망이 가지고 있던 학습이 불안정한 점과 그로 인해 고해상도 영상(256x256)을 생성하지 못했던 한계를 개선하는데 핵심 역할을 해주었다.

그러나 위와 같은 연구들은 학습에 사용된 실제 영상과 유사한 거짓 영상을 생성해 낼 수는 있지만, 원하는 영상 생성까지 제어할 수 없다는 한계가 있다. 따라서 이를 제어하기 위해서 조건부 적대적 생성 신경망(Conditional GAN, cGAN)[5]이 제안되었다. 조건부 적대적 생성 신경망은 생성 조건을 제어해 줄 수 있는 조건 변수를 생성자와 판별자에 추가한 형태의 네트워크로, 조건 변수는 클래스나 속성 등 다양한 형태를 가질 수 있어 그 활용도가 높다. 조건부 적대적 생성 신경망 기반 관련된 연구로는 Pix2Pix[6]과 CycleGAN[7], DiscoGAN[8], StarGAN[9] 등이 있으며, 조건 제어를 통해 기존 영상의 스타일을 원하는 형태로 변형 가능하다는 특징이 있다.

한편, 최근 의료분야에서는 의료 영상 불균형 문제를 해결하고자 DCGAN에 기반을 둔 흉부 X선 영상 생성 연구를 진행하였다[10][11]. 그러나 기존 연구 방법을 이용하여 영상 불균형 문제를 해결하기에는 다음과 같은 한계가 있다.

첫째, 학습이 불안정하다.

GAN의 불안정성은 고질적인 문제로 인공지능 학습을 어렵게 만든다. DCGAN이 학습을 안정시키는데 큰 기여를 했지만 그건 기존에 비해 나아졌다는 것이지 아직까지도 개선의 여지가 많다. 또한, 높은 해상도의 영상을 생성할수록 혹은 다양한 변화를 위해 조건변수를 추가하여 다양성을 추구할수록 학습의 불안정성은 커진다. 이는 모드붕괴(mode-collapse)라는 문제를 수반하여 영상 생성을 더 어렵게 만든다.

둘째, 생성 영상의 해상도가 낮다.

DCGAN을 통해 생성된 의료 영상을 활용하기 위해서는 기본적으로 육안으로 구분이 가능할 정도의 높은 해상도가 보장되어야한다. 하지만 기존의 연구들의 경우 상당히 낮은 해상도를 갖기 때문에 정확한 식별이 어렵다.

셋째, 변화를 다양하게 주기 어렵다.

DCGAN[10]은 오직 한 종류에 해당하는 영상생성이 가능하다. 여러 종류의 영상을 생성하기 위해 [11]에서는 다수의 DCGAN을 두어 개별적으로 학습을 하였다. 이는 학습 시간이 몇 배로 걸릴 뿐만 아니라 노동, 자원 비용 또한 크다. 이러한 수고를 막기 위해 기존 모델에 조건변수를 추가한 cDCGAN을 사용할 수 있는데, cDCGAN의 경우 최소한의 이미지 종류에 한해서만 높은 품질의 영상생성이 가능하다는 한계가 있다. 즉, 다양한 영상을 생성할 때는 좋은 결과를 얻지 못한다. 이는 다양성을 추구할수록 학습의 안정성은 감소하다는 점에서 첫 번째 한계로 지적한 학습 불안정성 문제와 같은 맥락으로 해석된다.

따라서 본 논문에서는 안전성을 보장하면서 다양한 종류의 고해상도 영상 생성이 가능하도록 StackGAN++[12]기법을 확장하여 상기의 문제점을 모두 개선한 새로운 네트워크를 제안하고자 한다.

III. 제안 기법

본 논문에서는 흉부 X선 영상 내 늑골, 횡격막, 폐, 심장 등의 세부적인 특징들을 잘 반영하여 고해상도 영상을 생성할 수 있는 네트워크를 제안한다. 해당 네트워크는 질환 조건 제어인자를 추가해 단 하나의 네트워크만으로 다수 질환에 대한 영상을 생성할 수 있도록 하였으며, 주목 메커니즘을 적용하여 생성 영상 내 장기 종속성 문제를 해결함으로써 흉부 X선 영상 생성 성능을 향상시켰다.

본 절에서는 제안된 네트워크를 1) 주목 메커니즘과 2) 질환 조건부 멀티 스케일 영상 생성 두 부분으로 크게 나누어 자세한 내용을 설명한다.

1. 주목 메커니즘

주목 메커니즘[13]은 중요도가 높은 특정 벡터에 더 집중하도록 하는 기법으로 기계번역을 위한 RNN[14]의 'sequence-to-sequence'[15]에 처음 도입되어 시퀀스(sequence)가 긴 경우 발생하는 장기 의존성(long-term dependency) 문제를 효과적으로 해결하였다. 이후 많은 후속 연구들이 발표되어 텍스트 외 영상 등 다양한 문제에 적용되었다.

특히, 영상 도메인으로 확장시 장거리 종속성 문제(long-range dependency)를 고려해주어야하는 이슈가 있었다. 이를 해결하기 위해 적대적 생성 신경망에서는 합성곱의 수용영역(receptive field)의 범위가 지역적으로 제한되어 영상 내 멀리 떨어진 위치들 간의 형태를 일관성 있게 모델링하기 어려웠던 문제를 자기 주목 메커니즘(self-attention mechanism)[16]을 적용해 해결하였다. 흉부 X선 영상의 경우 환자마다 골격이나 기관의 모습이 다르고(patient-specific) 혈관과 같은 주변기관과 잡음으로 인해 명암도 레벨의 분포가 불명확하여 지역적 특성만 고려해서는 질환이 무엇인지 판단하기 어렵다. 또한, 질환 영역이 영상 전체에 걸쳐 있거나 여러 개가 멀리 떨어져 분포할 수 있기 때문에 영상 내 장거리 종속성 문제 해결이 필수적으로 요구된다.

본 논문에서 제안하는 네트워크는 생성자에 자기 주목 메커니즘을 추가함으로써 영상 내 장거리 종속성 문제를 해결해 흉부 X선 영상 생성 성능을 향상시켰다.

2. 적대적 생성 신경망

흉부 X선 영상 내 기관들의 세부적인 특징들을 잘 반영하기 위해서는 고해상도 영상이 필수적으로 요구되기에 본 논문에서는 StackGAN++과 LSGAN[17]을 확장하여 저해상도 영상에서 고해상도 영상까지 멀티 스케일 영상 분포를 학습하는 네트워크를 제안한다 (그림 1).

우선, 제안한 네트워크에 대해 소개하기 전에 해당 손실

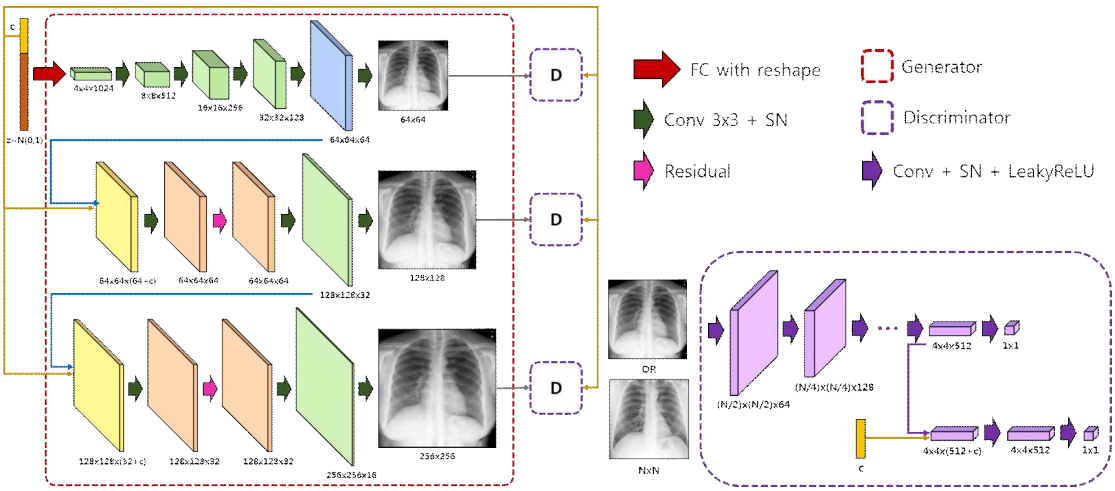


그림 1. 제안하는 주목 메커니즘 기반의 멀티 스케일 조건부 적대적 생성 신경망의 네트워크 구조
Fig. 1. Network Structure of Multi-scale Conditional Generative Adversarial Network Based on Attention Mechanisms

함수의 기반이 되는 StackGAN++과 LSGAN에 대한 이해가 선행되어야한다.

2.1 인용한 적대적 생성 신경망 1: StackGAN++

StackGAN++은 한번에 고해상도 이미지를 생성하기 어려운 문제를 해결하고자 분기별로 업샘플링(up-sampling)을 통해 얻어진 영상을 판별자를 통해 판단함으로써 해상도가 낮은 영상부터 단계적으로 생성하여 해상도를 보완하는 방식으로 이루어진 네트워크이다. 뿐만 아니라, 조건변수(c)를 추가하여 하나의 네트워크만으로 다양한 종류의 영상을 생성할 수 있다는 특징이 있다.

StackGAN++의 생성자(2)와 판별자(1)의 손실함수는 아래와 같은데, 기본 GAN의 생성자와 판별자에 i를 사용하여 분기를 표현하였고 영상 생성 제어를 위해 조건변수를 추가하였다. (1)과 (2) 손실함수가 수렴한다는 것은 Han Zhang et al.이 발표한 [12]의 5.2를 통해 증명되었다. (구체적인 수식은 3.2.3에서 설명)

$$\mathcal{L}_{\mathcal{D}_i} = -\mathbb{E}_{x_i\sim\mathcal{P}_{data_i}}[\log \mathcal{D}_i(x_i)] - \mathbb{E}_{s_i\sim\mathcal{P}_{G_i}}[\log(1-\mathcal{D}_i(s_i))] + \\ -\mathbb{E}_{x_i\sim\mathcal{P}_{data_i}}[\log \mathcal{D}_i(x_i,c)] - \mathbb{E}_{s_i\sim\mathcal{P}_{G_i}}[\log(1-\mathcal{D}_i(s_i,c))] \quad (1)$$

$$\mathcal{L}_{G_i} = -\mathbb{E}_{s_i\sim\mathcal{P}_{G_i}}[\log \mathcal{D}_i(s_i)] - \mathbb{E}_{s_i\sim\mathcal{P}_{G_i}}[\log \mathcal{D}_i(s_i,c)] \quad (2)$$

하지만, [12]는 제안한 모델을 최적화하기 위해 교차 엔트로피 손실 함수(cross entropy loss function)를 사용하고 있어 생성자 업데이트시 기울기 소실(gradient vanishing)이 일어난다는 한계가 있다.

2.2 인용한 적대적 생성 신경망 2: LSGAN

LSGAN[17]은 기존의 GAN이 가지고 있는 문제들을 해결하기 위해 변형된 손실함수를 제안한다. 손실함수 관점에서 기존의 GAN이 가지는 한계를 아래와 같이 요약해볼 수 있다.

첫째, 기존 GAN에서는 생성자가 만든 샘플이 실제 데이터 분포와 거리가 멀어도 시그모이드 교차 엔트로피 손실함수(sigmoid cross entropy loss function)는 판별자를 속이는 역할만 하기 때문에 적절한 피드백을 주지 못한다. 이는 생성자가 더 이상 학습을 진행하지 못하게 만든다.

둘째, 시그모이드 교차 엔트로피 손실 함수를 사용하면 x값에 따라 saturation 될 수 있다.

LSGAN은 위 두 문제를 해결하기 위해 시그모이드 교차 엔트로피 손실 함수대신 최소 제곱 손실함수(least square loss function) (3)(4)를 적용하였다. 이렇게 변형된 손실함수를 사용하면 샘플이 결정 경계(decision boundary)로부터 멀리 떨어져 있는 경우 거리에 따라 패널티를 받기 때문에 생성자가 좀 더 결정 경계에 가까운 샘플을 만들 수 있다.

또한, 최소 제곱 손실함수는 오직 한 점에서 최솟값을 갖기 때문에 x값에 따라 saturation되는 문제를 완화시켜 조금 더 안정적인 학습이 가능하다. (b는 실제 데이터, a, c는 생성 데이터)

$$\min_{\mathcal{D}} V_{LSGAN}(\mathcal{D}) = \frac{1}{2}\mathbb{E}_{x\sim\mathcal{P}_{data}(x)}[(\mathcal{D}(x)-b)^2] + \mathbb{E}_{z\sim\mathcal{P}_z(z)}[(\mathcal{D}(G(z))-a)^2] \quad (3)$$

$$\max_{G} V_{LSGAN}(G) = \frac{1}{2}\mathbb{E}_{z\sim\mathcal{P}_z(z)}[(\mathcal{D}(G(z))-c)^2] \quad (4)$$

따라서 본 논문에서는 StackGAN++이 가지고 있는 기울시 소실문제를 해결하기 위해 LSGAN 손실함수 적용하여 새로운 손실함수를 제안하였다.

2.3 제안한 적대적 생성 신경망: 멀티 스케일 조건부 적대적 생성 신경망

네트워크는 하나의 생성자와 다수개의 판별자로 구성되어 마치 트리 구조의 모습을 보인다. 크게 영상 생성부분과 판별부분으로 나뉘는데, 영상생성 부분에서는 생성자 내 각 분기(branch)별로 저해상도 영상에서 고해상도 영상까지 멀티 스케일 영상분포를 학습해 점진적으로 고화질 흉부 X선 영상을 생성하고, 판별부분에서는 생성자의 각 분기별마다 있는 판별자가 생성된 멀티 스케일 영상이 잘 만들어졌는지 평가함으로써 생성자가 최적화 될 수 있도록 인도한다.

먼저, 생성자는 잠재변수 z (z ~ P_{noise})와 함께 조건변수 c를 입력으로 받는데 조건변수는 잠재변수의 일부를 대신해 생성할 영상을 질환별로 제어할 수 있도록 한다. 본 논문

에서는 8개의 대표 흉부 질환에 해당하는 영상 생성을 목표로 하였기 때문에 조건변수에는 8개의 클래스에 대한 one-hot encoding 값이 할당된다. 즉, 표준 정규 분포로 표현된 초기 분포 P_{noise}는 조건변수와 함께 생성자의 여러 은닉 층(hidden layer)을 거치면서 각 분기별마다 해당 스케일의 표준 데이터 분포 P_{data}내 데이터로 근사화된다.

판별자의 경우 기존 GAN의 손실함수는 입력으로 들어온 영상이 참인지 거짓인지 구별하는 역할만 하였으나, 조건 변수를 추가한 제안한 네트워크의 손실함수는 입력으로 들어온 영상과 조건변수가 일치하는지 판단하는 역할도 추가된다. 구체적으로 생성자의 총 m개의 분기에 대하여 순차적으로 저해상도 영상 판별자(i=1)와 고해상도 영상 판별자(i=m)가 있으며, i번째 생성자 분기의 판별자 손실함수 L은 (5)와 같이 정의된다. (a, b 값은 각각 0, 1로 설정됨.)

$$\mathcal{L}_{\mathcal{D}_i} = \frac{1}{2}(\mathbb{E}_{x_i \sim \mathcal{P}_{data_i}}[(\mathcal{D}_i(x_i) - b)^2] + \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}[(\mathcal{D}_i(s_i) - a)^2] + \mathbb{E}_{x_i \sim \mathcal{P}_{data_i}}[(\mathcal{D}_i(x_i|c) - b)^2] + \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}[(\mathcal{D}_i(s_i|c) - a)^2]) \quad (5)$$

위 식에서 s_i 는 생성자로부터 생성된 크기별 샘플을 의미한다. 즉, $s_i = G_i(h_i)$ 이고 i는 0, 1, ..., m-1의 범위를 갖는다. 또한 $h_i = F_i(h_{i-1}, z)$이며, 여기서 h_i는 G_i의 은닉 피처(hidden feature)이다.

한편, 생성자의 경우는 각 분기별로 판별자 손실함수들의 합으로 표현 된다(6). 즉, 생성자는 저해상도에서 고해상도의 영상들에 대하여 비조건부 영상과 조건부 영상의 분포를 근사화하는 방향으로 학습된다. (d 값은 1로 설정됨.)

이처럼 멀티 스케일에서 영상 분포를 모델링하게 되면 각 스케일별 경사(gradient)가 발생되어 전달되기 때문에 초기 층까지 경사를 잘 전달한다는 장점을 갖고 있다. 결국 이러한 특징은 네트워크 학습을 안정화시키는데 핵심적인 역할을 해주어 고해상도 영상 생성이 가능하게 된다.

3. 모델 세부 사항

(그림 1)과 같이 생성자는 위에서부터 총 3개의 분기로 나눠져 저해상도에서 고해상도의 영상을 생성한다. (그림 1)의 가장 상위 분기는 저해상도 영상(64x64)을 위한 서브-생성자로 4개의 up-block 층과 주목(attention) 층 그리고 3x3 합성곱을 거쳐 영상을 생성한다. 특히, up-block 층에는 체커보드 인공물(checkerboard-artifact) 발생을 완화시키기 위해 up-sampling으로 최근접 이웃(nearest-neighbor) 방법을 사용하였다.

(그림 1)의 중간 분기와 마지막 분기의 서브-생성자는 결합(joining) 층과 두 번의 Residual 층 그리고 up-block 층으로 구성되어있다. 이 두 분기에서는 4단계의 층을 통과한 후 3x3 합성곱을 거쳐 각각 128x128 영상과 256x256영상을 출력한다.

반면, 각 분기별 판별자는 down-sampling 층으로 구성되며 조건과 비조건 손실함수 계산을 위해 마지막 층 전단계의 입력부분을 두 부분으로 나누어 한 곳에만 조건변수를 결합시켜준다.

학습을 위해 배치 사이즈는 16, 학습 에폭(epoch)은 20으로 수행하였다. 최적화방법으로는 Adam을 사용하였고 학습율은 2x10-5, 모멘텀은 0.5로 하였다. 또한, 주목 층에서는 사용하는 γ는 0에서 시작하여 학습 에폭 5마다 0.1씩(최대 1) 증가하게 해주었다. 생성자와 판별자의 업데이트 비율은 1:5이고, 안정적인 학습을 위해 스펙트럼 정규화(Spectral Normalization, SN)를 사용하였다.

Ⅳ. 실험 및 결과

1. 데이터

본 논문에서는 흉부 X선 데이터 증강 실험을 위하여 NIH Clinical Center에서 제공한 14개 심·폐질환에 대한 112,120건의 흉부 X선 영상 데이터를 사용하였다. 그 중 8개의 대표 질환(무기폐, 심장 비대, 흉수, 침윤, 종괴, 결절, 폐렴, 기흉)만을 사용하여 해당 질환들에 대한 데이터 수적 불균형 문제를 해결하고자 하였다.

$$\mathcal{L}_G = \frac{1}{2}\sum_{i=1}^{m}(\mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}[(\mathcal{D}_i(s_i) - d)^2] + \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}[(\mathcal{D}_i(s_i|c) - d)^2]) \quad (6)$$

표 1. 조건부 심층 합성곱 적대적 생성 신경망(cDCGAN)과 본 논문에서 제안한 방법을 통해 생성된 질환별 FDD 결과
Table 1. Comparison of FDD results with conditional Deep Convolution Generative Adversarial Network and Proposed method

Model	Fréchet DenseNet Distance (FDD)							
	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax
cDCGAN	4.05E-05	2.87E-05	**1.34E-04**	3.27E-04	1.49E-04	3.38E-04	**3.62E-05**	2.17E-04
Proposed method	**2.15E-05**	**6.65E-06**	3.48E-04	**2.29E-04**	**5.97E-05**	**9.34E-05**	4.52E-04	**3.30E-05**

2. 측정

본 논문에서는 성능비교를 위해 8개의 흉부 대표 질환에 대해 2가지 실험을 진행하였다.

첫 번째 실험은 1)조건부 심층 합성곱 적대적 생성 신경망[10]과 2)본 논문에서 제안하는 네트워크인 주목 메커니즘 기반의 멀티 스케일 조건부 적대적 생성 신경망으로 고해상도 영상(256x256)을 질환별로 생성(그림 2)한 후 생성자료의 품질을 평가하기 위해 새롭게 제안한 지표인 프레쳇 덴스넷 거리(Fréchet DenseNet distance, FDD)로 네트워크 성능을 비교한 것이다 (표 1).

여기서, (그림 2)는 학습단계에서 생성자의 손실 그래프(좌측)와 판별자의 손실 그래프(우측)를 나타낸다 (x축은 iteration, y축은 loss). 두 신경망은 서로 경쟁하는 관계로 각 신경망 입장에서는 손실 값이 커야 올바르게 학습되고 있음을 의미한다.

실험에서 사용한 영상의 개수는 대략 10만장으로 batch size를 16으로 하였을 때 한 에폭당 7,000번의 iteration이 이루어진다. (그림 2)의 그래프를 보면 판별자는 5 에폭 이후 점차 안정화 되었고 생성자는 20 에폭(iteration 140,000) 부근에서 함수 변동(fluctuation)이 안정화 되었다. 그 이상 학습을 진행했을 시 두 모델이 균형을 잃어 mode collapse가 발생하여 학습을 중단하였다.

두 번째 실험은 DCGAN과 제안한 모델을 통해 나온 실험결과를 실제 영상과 정성적 비교한 것으로, 제안한 모델 결과에 대해서는 질환별로 질환위치를 찾아 표시하였다 (표 2).

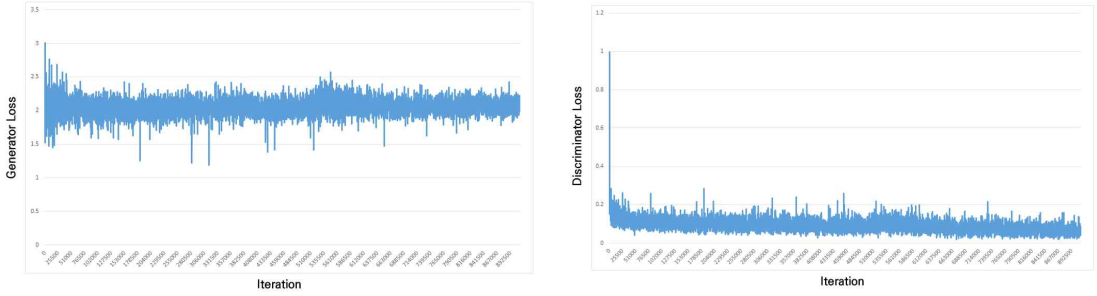
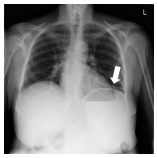
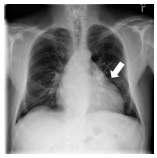
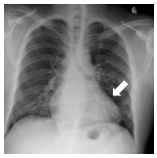
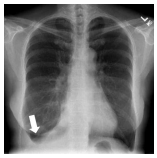
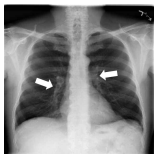
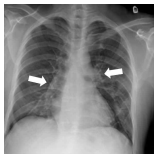
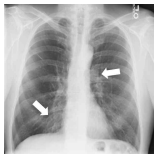
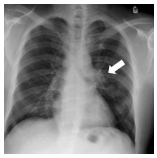
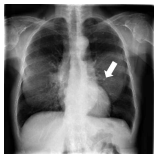
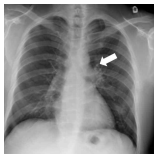
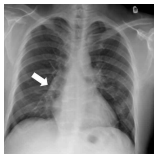
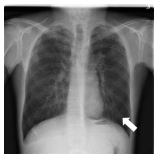


그림 2. 생성자 손실 그래프(좌측)와 판별자 손실 그래프(우측)
Fig. 2. Generator Loss Graph(Left) and Discriminator Loss Graph(Right)

표 2. 조건부 심층 합성곱 적대적 생성 신경망(cDCGAN)과 본 논문에서 제안한 방법과의 결과 비교
Table 2. Comparison of results with conditional Deep Convolution Generative Adversarial Network (cDCGAN) and the proposed method

	실제 영상 (Real image)	cDCGAN	제안하는 방법(Proposed method)
정상 (Normal)			
무기폐 (Atelectasis)			
심장 비대 (Cardiomegaly)			
흉수 (Effusion)			
침윤 (Infiltration)			
종괴 (Mass)			
결절 (Nodule)			
폐렴 (Pneumonia)			
기흉 (Pneumothorax)			

2.1. 프레쳇 덴스넷 거리 (Fréchet DenseNet Distance, FDD)

프레쳇 인셉션 거리(Fréchet Inception Distance, FID)[18]는 두 정규 분포의 차이를 측정한 것으로 인셉션 점수(Inception Score, IS)(7)[19]가 실제 자료의 분포를 사용하지 않는 단점을 보완하기 위해서 제안되었다. FID는 (8)와 같이 계산되는데 값이 작을수록 좋은 품질을 의미한다. (여기서, (m, C)와 (m_w, C_w)는 생성자료와 실제 자료의 평균과 공분산이다.)

$$IS(G) = \exp(\mathbb{E}_{\hat{x}\sim P_g} KL(p(y|\hat{x})||p(y))) \quad (7)$$

$$FID = \|m - m_w\|_2^2 + Tr(C + C_w - 2(CC_w)^{1/2}) \quad (8)$$

하지만, 1,000개의 클래스와 120만개로 구성된 자연 이미지인 ImageNet을 사전 학습한 인셉션 모델을 기반으로 각 정규분포를 구하게 되면 분포를 구할 때 사용되는 이미지가 사전 학습된 모델을 통해 1,000개의 클래스에 속할 확률 벡터를 출력한다는 점에서 문제가 된다. 그 이유는 인셉션 모델이 자연 이미지를 기반으로 학습되었기 때문에 흉부 X선과 같은 의료 영상을 클래스로 포함하지 않기 때문이다. 즉, 흉부 X선 영상은 자연 이미지가 아니기 때문에 이렇게 특화된 특징을 해석하기에 적합하지 않다. 따라서 의료 영상의 분포를 구하고 분포간 거리를 통해 품질을 평가하기 위해서는 의료 도메인에 최적화되어진 모델이 필요하다. 따라서 의료영상을 학습데이터로 하여 DenseNet을 학습시킨 후 이 모델을 기반으로 두 의료 영상의 정규분포를 구해 거리를 측정하는 프레쳇 덴스넷 거리(Fréchet DenseNet Distance, FDD)를 새로운 지표로 제안하여 실험을 수행하였다.

3. 실험 결과

3.1. 정량적 결과

우선, 첫 번째 실험을 진행하기 위해 학습데이터 78,468개, 검증데이터 11,219개로 DenseNet-121을 학습하였다. 그리고 학습된 모델을 기반으로 실제 영상과 cDCGAN 그리고 실제 영상과 제안한 모델에 대한 FDD값을 구하여 모델의 성능을 정량적으로 평가하였다. 여기서 실제 영상은 DenseNet-121 학습에 사용되지 않은 나머지 부분을 의미한다. 하지만 의료영상 특성상 데이터 불균형이 심해 각 클래스마다 영상 데이터의 개수가 고르지 못하다. 따라서 클래스별 실제영상과 생성된 영상의 개수는 Atelectasis 1,000개, Cardiomegaly 575개, Effusion 1,000개, Infiltration 1,000개, Mass 729개, Nodule 774개, Pneumonia 242개, Pneumothorax 539개로 총 5,859개의 영상을 동일하게 맞추어 테스트를 수행하였다.

그 결과 (표 1)에서 보이듯이 실제 영상과 제안한 모델에서 생성한 영상의 거리를 계산한 FDD값이 실제 영상과 cDCGAN에서 생성한 영상의 거리를 계산한 FDD값보다 대체로 낮으므로 제안한 모델의 성능이 더 우수하다는 것을 확인할 수 있다.

3.2 정성적 결과

(표 2)는 실제 데이터와 1) 조건부 심층 합성곱 적대적 생성 신경망의 결과 2) 제안한 모델의 결과를 나타낸 것이다. 해당 질환에 대한 구체적인 구별방법은 다음과 같다.

무기폐(Atelectasis)는 폐 전체 혹은 일부의 공기 감소를 일으키며, 일반 적으로 폐 용적 감소를 동반하고 호흡길(기도)이 치우치거나 폐의 백색화가 일어난다. 심장 비대(Cardiomegaly)는 심실벽(근육)이 두꺼워져 심근의 무게가 증가한 상태로 흉곽음영의 내부 길이에 비하여 심장음영의 길이가 절반이상 차지한 경우로 정의한다. 흉수(Effusion)는 흉막강 내 이상으로 고인 액체로 좌우 흉강의 불균형한 음영증가가 보인다. 침윤(Infiltration)은 정상 조직에 염증세포가 모여 있는 모양으로 폐포성 음영증가를 보이며 폐 주변부위에 잘 나타난다. 종괴(Mass)는 직경 3cm이상의 큰 덩어리 모양으로 증가된 음영. 폐, 흉막, 종격(동), 흉벽 등 흉부 모든 곳에서 기술 가능하다. 결절(Nodule)은 3cm이하의 경계가 그려지는 둥근 폐 음영이다. 폐렴(Pneumonia)은 폐에 염증이 일어나는 반응으로 주로 세균의 감염을 통해 일어나고 영상 내 그물모양이나 벌집 모양 같은 음영이 발견된다. 마지막으로, 기흉(Pneumothorax)은 흉막강 내에 구멍이 생겨 공기가 고이는 상태로 큰 공기주머니가 보인다.

위 질환 설명을 기반으로 1) 조건부 심층 합성곱 적대적

생성 신경망의 결과들은 실제 데이터와 비교해 영상의 해상도가 확연히 떨어짐을 볼 수 있다. 즉, 늑골의 모양이나 폐 기관지 모습이 선명하게 보이지 않아 주요 특징들이 제대로 학습되었다고 보기 어렵다. 반면, 2) 제안한 모델의 결과는 실제 영상과 구별되지 않을 정도의 영상특징과 해상도를 보였으며, 각 질환별 특징도 실제 영상과 큰 차이가 없을 정도로 잘 학습하였다. 한 예로 심장 비대의 경우 다른 질환과 비교하였을 때 심장의 모습이 확연히 커진 것을 볼 수 있는데, 이는 '심장 크기 증가'라는 해당 질환의 특징을 잘 학습한 결과이다.

V. 결 론

본 논문은 흉부 X선 영상의 질환별 데이터 수적 불균형 문제를 해결하고자 질환별 특징을 잘 반영한 고해상도 영상 생성이 가능한 네트워크를 제안하였다. 네트워크는 해상도가 낮은 영상부터 단계적으로 생성하여 해상도를 보완하는 방식으로 이루어져 있으며, 조건변수(c)를 추가하여 하나의 네트워크만으로 다양한 종류의 영상을 생성할 수 있어 학습 효율을 높일 수 있다. 그 결과 단 하나의 네트워크만으로 훨씬 더 선명하고, 질환 별로 특징이 잘 반영된 영상을 생성할 수 있다.

또한, 네트워크를 통해 생성된 영상의 품질을 평가하기 위하여 정량적 평가 지표인 FDD를 새롭게 제안하였다. 이는 영상품질을 평가할 때 흔히 사용하는 지표인 FID가 자연이미지 기반으로 학습되어 의료영상에 대한 정량적 평가가 어렵다는 한계를 극복하고자 실제 흉부 X선 영상데이터 기반의 분류 모델을 활용해 고안한 방법이다. 해당 지표를 사용하여 품질 측정을 해본 결과 제안한 모델의 성능이 타 연구보다 우수하다는 것을 수치로 입증할 수 있었다.

하지만 제안한 모델을 통해 생성된 영상의 신뢰도를 높이기 위해서는 다양한 평가방법을 통한 추가적인 실험이 필요하다. 크게 정량적인 평가와 정성적인 평가방법을 들 수 있는데, 추후 정량적인 평가를 위해서 실제 영상데이터와 생성된 영상데이터가 추가된 데이터 셋을 구성하고 분류(classification), 검출(Detection), 분할(Segmentation)과 같은 문제에 적용해 성능 향상 정도를 비교함으로써 본 논문에서 제안한 네트워크의 효용성을 추가로 입증할 계획이다.

그리고 정성적인 평가를 위해서는 결과물에 대한 임상의의 검증을 진행하여 생성된 영상의 시각적 분석과 함께 임상적용 유효성을 평가할 것이다.

참 고 문 헌 (References)

[1] Gyeongwan Kug, Application of Artificial Intelligence Technology and Industry, IITP, pp.22-26, March, 2019.

[2] F Provost, "Machine learning from imbalanced data sets 101," Proceedings of the AAAI'2000 workshop on imbalanced data sets, Vol. 68, No. 2000, AAAI Press, 2000.

[3] Goodfellow, Ian, et al., "Generative adversarial nets," Advances in neural information processing systems, pp. 2672-2680, 2014.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in International Conference on Learning Representations (ICLR), 2015.

[5] Mirza, Mehdi, and Simon Osindero, "Conditional generative adversarial nets," arXiv preprint, arXiv:1411.1784, 2014.

[6] Isola, Phillip, et al., "Image-to-image translation with conditional adversarial networks," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134, 2017.

[7] Zhu, Jun-Yan, et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proceedings of the IEEE International Conference on Computer Vision, pp. 2223-2232, 2017.

[8] Kim, Taeksoo, et al., "Learning to discover cross-domain relations with generative adversarial networks," Proceedings of the 34th International Conference on Machine Learning, Volume 70, JMLR. org, 2017.

[9] Choi, Yunjey, et al., "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[10] Salehinejad, Hojjat, et al., "Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 990-994, April, 2018.

[11] Salehinejad, Hojjat, et al., "Synthesizing Chest X-Ray Pathology for Training Deep Convolutional Neural Networks," IEEE transactions on medical imaging, 38.5: 1197-1206, 2018.

[12] Zhang, Han, et al., "Stackgan++: Realistic image synthesis with stacked generative adversarial networks," arXiv preprint, arXiv: 1710.10916, 2017.

[13] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint, arXiv:1409.0473, 2014.

[14] Hochreiter, Sepp, and Jürgen Schmidhuber, "Long short-term memory," Neural computation, 9.8: 1735-1780, 1997.

[15] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le, "Sequence to sequence learning with neural networks," Advances in neural information proc-

essing systems, 2014.
[16] Zhang, Han, et al., “Self-Attention Generative Adversarial Networks,” arXiv preprint, arXiv:1805.08318, 2018.
[17] Mao, Xudong, et al., "Least squares generative adversarial networks," Proceedings of the IEEE International Conference on Computer Vision, 2017.
[18] Heusel, Martin, et al., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," Advances in Neural Information Processing Systems, 2017.
[19] Salimans, Tim, et al., "Improved techniques for training gans," Advances in neural information processing systems, 2016.

저 자 소 개

안 경 진

- 2014년 3월 ~ 2019년 2월 : 한국외국어대학교 컴퓨터공학과 학사
- 2019년 3월 ~ 현재 : 연세대학교 의과학과 석사과정
- ORCID : https://orcid.org/0000-0001-5962-1635
- 주관심분야 : 의료영상처리, 인공지능

장 영 걸

- 2009년 3월 ~ 2013년 8월 : 한국외국어대학교 컴퓨터공학과 학사
- 2014년 9월 ~ 현재 : 연세대학교 의과학과 석박통합과정
- ORCID : https://orcid.org/0000-0002-5805-7494
- 주관심분야 : 의료영상처리, 인공지능

하 성 민

- 2007년 3월 ~ 2014년 2월 : 한국외국어대학교 디지털정보공학과 학사
- 2016년 9월 ~ 현재 : 연세대학교 생체공학협동과정 석박통합과정
- ORCID : https://orcid.org/0000-0002-0731-2301
- 주관심분야 : 의료영상처리, 인공지능

전 병 환

- 2009년 3월 ~ 2013년 8월 : 한국외국어대학교 컴퓨터공학과 학사
- 2014년 3월 ~ 2019년 8월 : 연세대학교 의과대학 이학박사 취득
- ORCID : https://orcid.org/0000-0002-0414-1762
- 주관심분야 : 의료영상처리, 인공지능

저 자 소 개

홍 영 택

- 2007년 3월 ~ 2012년 8월 : 한국외국어대학교 디지털정보공학과 학사
- 2012년 9월 ~ 2018년 8월 : 연세대학교 의과대학 이학박사 취득
- ORCID : https://orcid.org/0000-0003-2104-5905
- 주관심분야 : 의료영상처리, 인공지능

심 학 준

- 1993년 : 서울대학교 전기공학부 공학사
- 1995년 : 서울대학교 전기공학부 공학석사
- 2007년 : 서울대학교 전기컴퓨터공학부 공학박사
- 2008년 ~ 2010년 : 서울대학교 전기공학부 BK21 연구교수
- 2011년 ~ 현재 : 연세대학교 세브란스 병원 연구교수
- 주관심분야 : 컴퓨터 비전, 패턴인식, 영상처리

장 혁 재

- 1994년 : 연세대학교 의학과 의학사
- 1999년 : 연세대학교 의학과 의학석사
- 2003년 : 아주대학교 의학과 의학박사
- 2013년 ~ 현재 : 연세대학교 의과대학 심장내과 정교수
- https://orcid.org/0000-0002-6139-7545
- 주관심분야 : 심장관막질환, 심부전, 심근질환