

의료영상 분할 모델의 도메인 일반화 성능 향상을 위한 자기 지도 학습의 활용

이예진¹ · 이상우² · 황상흠^{2*}

¹서울과학기술대학교 데이터사이언스학과 / ²서울과학기술대학교 산업공학과

Improving Domain Generalization Performance for Medical Image Segmentation by Self-Supervised Learning

Yejin Lee¹ · Sangwoo Lee² · Sangheum Hwang^{2*}

¹Department of Data Science, Seoul National University of Science and Technology

²Department of Industrial Engineering, Seoul National University of Science and Technology

In recent years, deep learning technology has been widely used for medical image analysis. However, deep neural networks tend to produce lower generalization performance for data in novel domains, which is a frequent scenario in the field of medical imaging since the domain can be easily shifted by a patient's physical characteristics and image acquisition equipment. Meanwhile, self-supervised learning is recently known not only to further enhance the performance of a model, but also to improve the robustness of it. Based on this finding, we empirically demonstrated that a model's domain generalization performance can be improved by using self-supervised pre-training in this study. Moreover, we additionally found that data augmentation applied to the pretext task can significantly impact on domain generalization performance of a model.

Keywords: Self-Supervised Learning, Domain Generalization, Segmentation, Medical Image

1. 서론

의료영상 분석 분야는 인공지능 기술이 가장 활발하게 도입되고 있는 분야 중 하나이다. 특히 딥러닝 기반의 의료영상 분석 방법은 CT, X-ray뿐만 아니라 초음파, MRI까지 다양한 의료 영상에 대해서 병변의 탐지 및 분류 등을 목적으로 많은 진단 영역에 적용되고 있으며, 이와 관련된 연구 논문의 수도 빠르게 증가하고 있다(Litjens *et al.*, 2017). 이러한 가운데 최근 몇 년간 이루어진 컴퓨터 하드웨어의 비약적인 발전과 데이터양의 폭발적인 증가로 고성능의 딥러닝 모델들이 개발되었고, 이에 기인하여 의사들의 진단 및 판독 능력을 넘어서는 결과들이 등장하고 있다(Gulshan *et al.*, 2016; Esteva *et al.*, 2017;

Nam *et al.*, 2019; Kim *et al.*, 2020; McKinney *et al.*, 2020).

그러나 딥러닝 기반의 의료영상 분석 분야에도 기술적 난제가 존재한다. 바로 해부학적 구조를 모델에 반영하기 힘들다는 점이다. 해부학적 구조란, 인간의 신체 내부에 존재하는 진단 영역간의 상대적인 위치 관계를 의미한다. 직관적으로, 의료진의 입장에서는 의료 영상 데이터의 도메인이 달라져도 인간의 본질적인 해부학적 구조는 변하지 않기 때문에 진단 영역을 파악하는 데에 큰 어려움이 없다. 예를 들어, 서로 다른 환자에 대해 서로 다른 영상 획득 장비를 사용하더라도 갈비뼈 내부에 폐가 존재하며 흉부 중앙에 심장이 존재한다는 사실은 변함이 없다. 이와 같이 의료진은 진단을 위해서 해부학적 지식에 근거한 사전 지식을 활용하지만 인공지능 모델은 단순히 학습에 사용

이 논문은 서울과학기술대학교 교내연구비(2020-0575)의 지원을 받아 수행되었음.

* 연락저자 : 황상흠 교수, 01811 서울시 노원구 공릉로 232 서울과학기술대학교 프론티어관 618호, Tel : 02-970-6462, Fax : 02-974-2849,

E-mail : shwang@seoultech.ac.kr

2020년 12월 14일 접수; 2021년 1월 28일 수정본 접수; 2021년 1월 28일 게재 확정.

된 데이터만 보기 때문에 학습 데이터의 도메인과 상이한 도메인을 만났을 경우 성능이 저하되는 문제를 가진다(Long *et al.*, 2018). 학습된 데이터 셋에 편향돼 있고 데이터 셋이 포함하지 못하는 도메인의 데이터에 대해서 일반화 성능이 낮은 모델은 범용적으로 활용하기 어려울 뿐 아니라 높은 신뢰도를 얻기 어렵다. 특히 의료 영상 분야에서는 환자 개인의 특성과 영상 획득 장비에 따라 영상 데이터의 도메인이 달라지는 경우가 빈번하게 발생하기 때문에 도메인의 변화에도 강건한 모델을 구축할 필요가 있다. 이를 해결하기 위해선 의료진의 진단 영역 파악에 근거가 되는 해부학적 구조에 대한 지식을 모델에 충분히 반영시켜야 한다. 모델이 학습 데이터로부터 해부학적 구조를 이해한다면, 의료진과 마찬가지로 다양한 도메인에도 변하지 않는 진단 영역 간의 상대적인 위치 관계에 근거한 판단을 내릴 것이다. 예를 들어, 폐 영역 분할 문제에서 해부학적 구조를 학습한 모델의 경우 특정 영역을 분류할 때 해당 영역과 주변이 같비뼈 구조를 이루고 있다면 이를 인지하고 올바르게 폐 영역으로 분류해 낼 것이다. 한편, 도메인 일반화를 달성하기 위한 가장 직관적인 방법은 모델을 학습할 때 다양한 도메인으로부터 수집된 학습 데이터를 사용하는 것이다(Cho *et al.*, 2015). 즉, 충분한 양의 학습 데이터를 활용하는 것인데, 양질의 데이터를 새롭게 확보하는 것은 쉽지 않은 일이다. 의료 영상 데이터 특성상 숙련된 전문 인력만이 영상을 판독할 수 있으므로 라벨 정보를 포함하고 있는 데이터를 확보하기 위해선 비용과 시간이 많이 소요된다. 또한 환자의 의료영상 활용에 대한 법적 이슈도 존재하기 때문에 충분한 영상 데이터를 활용하기가 어려운 실정이다. 따라서 데이터를 마련하는데 필요한 시간과 비용을 최소화 하면서도 다양한 도메인에 동시적으로 성능이 좋은 방법론의 개발이 중요하다.

위와 같은 문제를 해결하기 위해서 다양한 선행연구가 이루어져 왔다. 기존의 선행연구는 크게 두 가지로 나눌 수 있다. 하나는 도메인 적응(domain adaptation)이며 다른 하나는 도메인 일반화(domain generalization)이다. 도메인 적응은 도메인 변화에 대한 강건성을 달성하기 위해 학습 시 타겟 도메인의 일부 데이터에 대해 접근이 가능한 상황을 가정한다. 그러나 실제 응용 분야에서는 위와 같은 방법이 현실적이지 못하다. 접근이 허락된 특정 도메인에 대해서는 좋은 성능을 낼 수 있으나, 또 다른 새로운 도메인을 만났을 때 도메인 변화에 대한 문제가 여전히 존재하기 때문이다. 반면 도메인 일반화는 여러 도메인으로 구성되어 있는 소스 데이터만을 사용해 학습 시 보지 못한 타겟 도메인에 대한 일반화 성능을 높임으로써 도메인 강건성을 달성하고자 한다. 이는 학습 시 타겟 도메인의 데이터를 활용하지 않기 때문에 보다 현실적이다. 하지만 여전히 다양한 도메인으로 구성된 소스 데이터가 요구된다는 한계를 가지고 있다. 의료 영상과 같이 정형화된 구조를 가지고 있는 데이터의 경우 그 구조를 모델이 이해할 수 있다면 하나의 도메인으로 구성된 학습 데이터만으로도 도메인 일반화 성능의 향상이 가능하다. 본 연구에서는 이와 같이 하나의 도메인에서 취득된 소스 데이터를

활용하여 학습한 모델이 다른 도메인들에 대하여 강건한 일반화 성능을 가질 수 있도록 하는 방법론 개발에 초점을 맞춘다.

자기 지도 학습(self-supervised learning)은 컴퓨터 비전과 자연어 처리 분야에서 괄목할 만한 성과를 보이며 활발하게 연구되고 있는 학습 방법론이다. 최근에는 이미지 분류 문제에서 지도 학습(supervised learning)보다 높은 수준의 성능을 달성할 수 있는 방법론이 개발되기도 했다(Chen *et al.*, 2020). 기존의 지도 학습은 데이터로부터 입력과 라벨 간의 관계를 찾는 것에 집중한다. 그러나 자기 지도 학습은 데이터 자체의 특징을 학습하는 알고리즘으로, 데이터에 대한 전반적인 특징을 모델이 학습하기 때문에 의료영상 영역 분할 문제에서 해부학적 구조를 모델에 반영하지 못해 발생하는 문제를 해결할 수 있는 방안이 된다. 실제로 자기 지도 학습은 모델의 성능을 추가적으로 향상시킬 뿐만 아니라 모델의 강건성을 향상시키는 데 효과적이라는 사실은 이전에 연구된 바가 있다(Hendrycks *et al.*, 2019). 이에 근거해 자기 지도 학습을 이용한 도메인 적응 방법론이 개발되었지만 여전히 타겟 데이터를 필요로 하는 한계점이 존재한다(Xu *et al.*, 2019).

본 연구에서는 의료영상 분석에서 가장 많이 사용되는 알고리즘인 영역 분할 문제를 다루며 다양한 진단 영역 중 폐 영역 분할 문제에 초점을 맞춘다. 또한 하나의 소스 데이터만을 사용해 자기 지도 학습으로 사전에 모델을 학습하고 이를 미세조정(fine-tuning)하여 영역 분할 모델을 학습한다면 도메인의 변화에 대한 영역 분할 성능의 강건성을 향상시킬 수 있음을 실험적으로 확인하였다. 더 나아가 자기 지도 학습으로 사용할 사전과제(pretext task)에 적용되는 데이터 증강 기법(data augmentation)이 도메인 일반화 성능에 큰 영향을 미칠 수 있음을 보였다.

2. 선행 연구

2.1 도메인 일반화

다양한 도메인에 대한 일반화 성능을 향상시키기 위해 최근까지 많은 연구가 이루어져왔다. 비교적 간단한 방법으로 데이터 증강 기법을 활용할 수 있다. 데이터 증강 기법은 학습 데이터에 인위적인 변화를 주어 모델이 더 넓은 영역의 학습 데이터를 볼 수 있도록 하는 방법이다. Augmix(Hendrycks *et al.*, 2019)는 다양한 데이터 증강 기법을 적절히 융합해서 모델의 성능을 향상시킬 수 있음을 보였다. 이를 통해 원본 이미지의 일관성을 해치지 않으면서도 데이터의 다양성을 유지하였으며, 도메인 강건성에 효과적임을 보였다. DST(Zhang *et al.*, 2019)는 3D 의료영상 분할 모델의 일반화 성능을 향상시키기 위해 다양한 데이터 증강 기법을 순차적으로 데이터에 적용한 방법으로, 단일 데이터 증강 기법을 사용했을 때 보다 여러 증강 기법을 적용했을 때 모델의 일반화 성능이 더욱 향상되는 것을 실험을 통해 확인했다. 또한 소스도메인 내에 존재하는 다양한 도메인 정보를 바탕

으로 만들어진 데이터 증강 기법을 사용하기 위해, 입력 데이터를 다른 소스 도메인의 스타일로 변화시킨 뒤 모델을 학습하는 방법론도 제시되었다(Somavarapu *et al.*, 2020).

이 외에도 새로운 학습 알고리즘을 제안하여 보다 나은 도메인 일반화 성능을 달성하기 위한 시도가 있었다. Episodic training (Li *et al.*, 2019)은 서로 다른 도메인에서 학습되는 특징 추출기(feature extractor)와 분류기(classifier)를 교차하여 학습함으로써 새로운 도메인에서의 이미지 분류 성능을 우수하게 유지하는 방법을 제안했다. 또한 도메인 적응과 도메인 일반화 문제에 모두 적용 가능한 새로운 손실 함수를 제안한 연구도 찾아볼 수 있다 (Motiian *et al.*, 2017). 서로 같은 클래스이지만 도메인이 다른 데이터 쌍은 임베딩 공간에서의 거리를 최소화하고, 서로 다른 클래스 이면서 동시에 도메인이 다른 데이터 쌍의 유사도는 최소화 되도록 만드는 손실 함수를 사용하여 모델의 도메인 일반화 성능을 향상시켰다. 기존의 방법들은 대부분 다양한 소스 도메인이 주어지는 경우를 가정한다. 그러나 본 연구에서는 학습 데이터가 하나의 소스 도메인을 가지는 보다 제한적인 상황만을 고려한다.

2.2 자기 지도 학습

지도 학습 기반의 방법론은 충분한 양의 라벨 정보를 담고 있는 데이터가 주어져 있어야 좋은 성능을 달성할 수 있다는 단점을 가진다. 이러한 한계점을 극복하기 위해 라벨이 없는 데이터만을 활용하여 데이터 자체의 특징(feature)을 효과적으로 학습하고자 하는 비지도 학습의 접근이 다양한 방식으로 연구되고 있다. 자기 지도 학습은 비지도 학습 분야 중 하나로, 데이터로부터 라벨을 생성하여 지도 학습에 활용하는 방법을 의미한다. 즉, 데이터 자체에 담겨있는 정보만을 이용하여 사전과제를 해결하도록 함으로써 보다 나은 데이터 표현(representation)을 학습하는 방법이다.

지난 몇 년간 여러 종류의 사전과제가 제안되었다. 원본 이미지로부터 회전된 각도를 예측하는 task (Gidaris *et al.*, 2018), 흑백 이미지를 색칠하는 task (Zhang *et al.*, 2016), 마스킹 된 영역을 복원하는 task (Pathak *et al.*, 2016) 등을 예시로 들 수 있다. 특히 의료영상 분야에서 제안된 사전과제로는 context restoration (Chen *et al.*, 2019)이 있다. Context restoration은 무작위로 뽑은 작은 패치 쌍 간의 위치를 바꾸는 과정을 여러 번 반복한 이미지를 입력으로 받아 원본 이미지를 복원하는 task다. 본 연구에서는 앞서 언급한 context restoration과 inpainting task (Pathak *et al.*, 2016)를 사용하여 모델이 해부학적 지식을 습득할 수 있도록 사전학습을 진행한다.

3. 방법론

3.1 연구 프레임워크

본 논문에서 제안하는 연구 프레임워크는 크게 두 단계로

나뉜다. <Figure 1>과 같이 먼저 사전과제를 모델에 충분히 학습시킨다. 사전과제들은 학습 과정에서 모델이 이미지의 국지적인 영역에 집중하기 보단 전체적인 특징을 학습하도록 유도된다. 주어진 사전과제를 잘 수행하기 위해선 해부학적 구조의 상대적인 위치 관계를 학습해야 하기 때문에 우리는 이 과정을 통해 모델이 해부학적 구조를 이해한다고 가정한다. 이후 학습된 모델을 영역 분할 문제로 전이학습(transfer learning) 하며 학습 과정을 마무리한다. 이러한 모든 과정에서 타겟 도메인의 데이터는 사용되지 않는다. 또한 본 연구에서는 사전과제로 inpainting과 context restoration을 적용하였다.

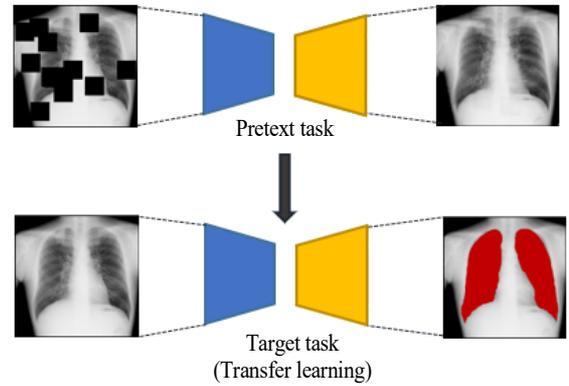


Figure 1. A Framework of the Proposed Method

3.2 Inpainting Task

Inpainting task는 마스킹 된 이미지를 원본 이미지와 비슷하게 복원하는 문제를 의미한다. 원본 이미지에 마스킹하는 방법은 여러 가지가 있다. 단순한 방법으로는 이미지의 가운데 영역을 마스킹하는 방법이 있다. 그러나 이 방법은 모든 입력 이미지가 동일한 영역에 마스킹 되기 때문에 모델이 특정한 영역에만 집중하는 경향성을 가질 수 있다(Chen *et al.*, 2019). 따라서 본 연구에서는 마스킹 된 영역이 전체 이미지의 1/4면적이 되도록 작은 영역을 무작위로 여러 개 추출하여 마스킹하는 방식을 적용하였다(<Figure 2> 참고).

Inpainting task를 모델에 학습시키기 위해 사용된 손실 함수는 식 (1)과 같다. $L_{adv}(x)$ 와 $L_{rec}(x)$ 은 높이가 h 이고 넓이가 w 인 입력 이미지 $x \in R^{h \times w}$ 에 대해 계산된 adversarial loss와 reconstruction loss를 의미한다. 계산된 두 손실 함수를 가중합하여 최종 손실 함수로써 사용하고, λ_{adv} 는 adversarial loss에 대한 가중치이다.

$$L_{inp}(x) = \lambda_{adv}L_{adv}(x) + (1 - \lambda_{adv})L_{rec}(x) \quad (1)$$

L_{rec} 은 입력 이미지를 최대한 원본 이미지와 비슷하게 복원시키기 위해 사용되는 손실 함수(reconstruction loss)로 식 (2)와 같이 표현된다.

$$L_{rec}(x, M) = \|M \odot (x - G((1 - M) \odot x))\|_1 \quad (2)$$

이때, 폐 영역 분할에서 사용될 모델(예를 들어, U-Net(Ronneberger *et al.*, 2015))을 복원된 이미지를 생성하는 generator ($G: R^{h \times w} \rightarrow R^{h \times w}$)로써 사용하고, generator는 마스킹 된 폐 사진을 입력으로 받아 원본 이미지 x 와 비슷하게 복원하도록 학습한다. $M \in R^{h \times w}$ 은 마스킹 할 영역은 1, 그렇지 않은 영역은 모두 0으로 이뤄진 마스크 행렬이다. Reconstruction loss의 종류는 성능에 큰 영향을 미치지 않으나, L_1 loss를 사용했을 때 가장 안정적임을 실험을 통해 확인하였다.

한편, L_{rec} (즉, L_1 reconstruction loss)만을 사용하여 모델을 학습하게 되면, 모델은 복원 영역을 평균적으로 비슷하게 만들어 흐릿한 이미지를 생성 해내는 경향을 가진다(Pathak *et al.*, 2016). 따라서 보다 선명한 이미지를 생성하기 위해 적대적 학습(adversarial training)을 병행한다. 식 (3)의 discriminator ($D: R^{h \times w} \rightarrow R$)는 입력 이미지가 원본 이미지인지, generator가 생성해낸 가짜 이미지인지 구분하는 역할을 하는 CNN 모델이다.

$$L_{adv}(x) = (1 - D(G(x)))^2 \quad (3)$$

$D(x)$ 는 입력 이미지 x 가 실제 이미지일 확률을 나타낸다. 따라서 generator는 discriminator가 구분하기 어렵도록 원본 이미지와 비슷한 이미지를 복원하고, 동시에 discriminator는 가짜 이미지를 잘 구분하도록 서로 경쟁적으로 학습한다. 모델은 inpainting task를 수행하기 위해서 입력 이미지에 대해 어느 곳이 마스킹 된 영역인지, 해당 영역이 어떤 값으로 채워져야 하는지 알아야 한다.

3.3 Context Restoration Task

Context restoration은 의료영상 분석의 성능 향상을 위해 제안된 자기 지도 학습 방법이다(Chen *et al.*, 2019). Inpainting처럼 입력 이미지에 마스킹하는 방식은 이미지의 context를 제거하게 되며, 이는 이미지의 강도 분포를 원래 분포와 다르게 변경하는 결과를 초래한다. 변경된 분포의 이미지로부터 학습된 특징은 원래 분포의 이미지에 대해 유용하지 않을 수 있다. Context restoration은 이러한 문제를 완화하기 위해 이미지에서 무작위로 선택한 두 영역의 위치를 교환하는 과정을 반복함으로써 입력 이미지를 생성한다. 각 task에서의 입력 이미지를 나타낸 <Figure 2>에서 볼 수 있듯이, context restoration에서의 입력 이미지는 특정 픽셀의 위치는 변경되었지만 픽셀 값의 정보는 유지하고 있다. 모델은 생성된 이미지를 입력으로 받아서 원본 이미지와 최대한 유사하게 복원하도록 학습된다. 학습과정은 inpainting task와 유사하지만, 입력 이미지를 생성하는 방법과 별도의 discriminator를 사용하지 않는다는 점이 다르다. 또한 context restoration task에서는 식 (4)와 같이 reconstruction loss로 L_2 loss를 사용하였다.

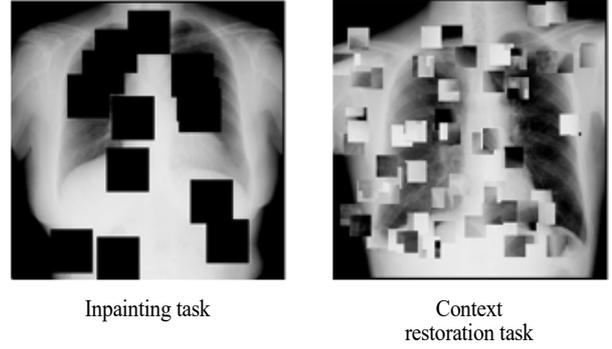


Figure 2. Examples of Input Images for Each Method

$$L_{con}(x) = \|x - G(\hat{x})\|_2 \quad (4)$$

$x \in R^{h \times w}$ 와 $\hat{x} \in R^{h \times w}$ 은 각각 원본 이미지와 생성된 입력 이미지를 의미하며, $G: R^{h \times w} \rightarrow R^{h \times w}$ 는 입력 이미지를 원본 이미지와 유사하게 복원하는 역할을 하는 generator이다. 모델은 context restoration task를 수행하기 위해서 어떤 영역이 변경된 영역인지, 해당 영역의 원래 위치는 어느 곳인지 알아야 한다.

3.4 Segmentation

영역 분할(Segmentation)의 목적은 이미지의 모든 픽셀을 해당 클래스로 분류하는 것이다. 즉, 이미지의 각 픽셀 단위로 해당 픽셀이 폐의 영역에 속하는 픽셀인지 그렇지 않은 영역인지 구분하는 것을 목표로 한다. 이때 모델을 학습시키기 위해 사용된 손실 함수는 하나의 입력 이미지 $x \in R^{h \times w}$ 에 대해 식 (5)와 같이 계산된다.

$$L_{seg}(x) = -\frac{1}{N} \sum_i^N (y_i \log P(y_i | x_i; \theta) - (1 - y_i) \log (1 - P(y_i | x_i; \theta))) \quad (5)$$

y_i 는 총 $h \times w = N$ 개의 픽셀을 가지는 입력 이미지의 i 번째 픽셀 x_i 의 정답 라벨이며, 폐에 속하는 픽셀인 경우 1, 그렇지 않은 경우 0의 값을 가진다. θ 은 영역 분할 모델 $G: R^{h \times w} \rightarrow R^{h \times w}$ 의 학습 가능한 매개변수(parameter)를 의미하고, $P(y_i | x_i; \theta)$ 는 영역 분할 모델에 의해 추정된 x_i 가 폐에 속할 확률이다. 따라서 모델은 해당 영역의 정답 라벨과 가까운 $P(y_i | x_i; \theta)$ 값을 출력할 수 있도록 매개변수를 업데이트하는 과정을 가진다.

본 연구에서는 inpainting task 혹은 context restoration task로 학습된 모델의 최종 매개변수가 해부학적 지식을 담고 있을 것으로 가정하며, 이를 폐 영역 분할 task에서의 모델의 초깃값으로 설정한다. 이후, L_{seg} 를 최소화 하도록 매개변수를 다시 업데이트 한다.

4. 연구 결과

4.1 데이터 셋 및 실험 환경

본 연구에서 사용하는 데이터 셋은 Japanese Society of Radiological Technology(이하 JSRT)(Shiraishi *et al.*, 2000), Montgomery County(이하 MC)(Jaeger *et al.*, 2014) 그리고 Shenzhen(이하 SZ)(Jaeger *et al.*, 2014)으로 총 3개의 데이터 셋을 사용한다. JSRT 데이터 셋은 크기가 2,048×2,048인 총 247개의 Chest X-ray 이미지로 구성되어 있다. 그 중 154개는 폐 결핵 환자의 사진이며, 나머지 93개는 정상 이미지이다. MC 데이터 셋은 Montgomery County's Tuberculosis screening program으로부터 받은 138개의 Chest X-ray 이미지 데이터 셋이다. 이 중 58개는 결핵 환자의 이미지이고, 나머지 80개는 정상인의 이미지이다. 이미지의 크기는 4,020×4,892 혹은 4,892×4,020으로 이루어져 있다. SZ 데이터 셋은 총 662개의 Chest X-ray 이미지로 구성되어 있고, 그 중 336개는 결핵 환자의 이미지이다. 나머지 326개는 정상 이미지이다. 이미지의 크기는 샘플마다 다르나, 대부분이 3,000×3,000 정도의 크기를 가진다. SZ 데이터 셋은 영역 분할 라벨이 제공되지 않기 때문에 kaggle에 공개되어 있는 영역 분할 라벨을 사용했다(<https://www.kaggle.com/yoctoman/shcxr-lung-mask>). 이때 영역 분할 라벨이 존재하지 않거나 잘못 라벨링 된 데이터를 제외한 521개의 데이터를 사용하였다. 잘못 라벨링 된 데이터는 다른 장기에 가려진 영역까지 폐의 영역으로 구분한 경우와, 라벨이 실제 폐와 정확하게 일치하지 않는 경우로 구성되어 있다. 이에 대한 예시는 <Figure 3>에 나타내었다. 실험 과정에서 사용되는 모든 데이터는 256×256로 사이즈를 재조정하였고, 소스 도메인 데이터 셋의 경우 70%를 훈련 데이터 셋, 나머지 30%를 테스트 데이터 셋으로 사용하였다.

실험에서 고려한 세 가지 데이터 셋은 영상 획득 장비와 환자의 상태 등이 다르기 때문에 각기 다른 도메인으로 정의한다. 본 연구에서는 도메인 일반화 성능 측정을 위해 하나의 데이터 셋(소스 도메인)으로 모델을 학습한 후, 학습 과정에서 보지 못한 나머지 두 데이터 셋(타겟 도메인)에 대해 폐 영역 분할 성능을 평가한다. 소스 도메인에서의 영역 분할 성능과 타겟 도메인에서의 영역 분할 성능의 차이가 적을수록 도메인 일반화의 성능이 좋은 것으로 해석할 수 있다. 본 연구에서는

사전학습을 하지 않고, 무작위 초깃값으로 소스 데이터에 대해 영역 분할 task를 학습한 모델을 baseline으로 삼는다. 자기 지도 학습으로 사전 학습한 모델과 baseline 모델의 각 도메인에서의 영역 분할 성능을 비교함으로써 본 논문에서 제안한 방법이 효과가 있는지 실험을 통해 검증하였다.

본 연구에서는 의료영상 분석 분야에서 널리 사용되고 있는 U-Net을 영역 분할 모델로 사용하였다. U-Net은 인코더와 디코더로 구성된 네트워크이며 인코더의 각 block은 2개의 3×3 convolution과 2×2 maxpooling으로 이루어져 있다. 또한 디코더는 2×2 up-convolution과 3×3 convolution으로 이뤄진 block을 가진다. Inpainting task에 대해 사전학습을 진행한 경우에는 손실 함수의 수렴을 위해 500 에폭 동안 배치 사이즈를 32로 설정하여 학습을 진행하였다. Generator에 대해서는 영역 분할 task에서 사용되는 optimizer와 동일한 설정을 사용했고, discriminator에 대해서는 $\beta_1 = 0.5$, $\beta_2 = 0.999$ 인 Adam optimizer (Kingma *et al.*, 2014)를 사용하였다. Learning rate는 generator의 경우 0.001, discriminator는 0.0001로 설정하여 학습하였으며 별도의 decay schedule은 적용하지 않았다. 또한 λ_{adv} 는 0.001로 설정했다. 이러한 하이퍼 파라미터는 다양한 실험을 통해 손실 함수의 커브가 가장 잘 수렴하고 복원 능력이 시각적으로 우수한 수치를 선택하였다. Context restoration으로 사전학습을 진행할 때에는 입력 이미지를 생성하기 위해 20×20 크기의 패치를 한 쌍 만들어 서로 자리를 바꾸었고, 이를 30회 반복하였다. 이렇게 생성된 이미지로 200 에폭 동안 모델을 학습하였다. 이때 사전학습에 사용되는 데이터 셋은 폐 영역 분할에서 소스 도메인으로 사용하고자 하는 데이터 셋을 사용했다. 이후 폐 영역 분할 task로 전이학습을 진행할 때에는 배치 사이즈를 32로 설정하여 120 에폭 동안 모델을 학습 시켰다. Optimizer는 Adam을 사용하였으며, weight decay는 0.0001, $\beta_1 = 0.5$, $\beta_2 = 0.999$, $\epsilon = 0.001$ 로 설정했다. Learning rate는 0.001을 사용하였다.

본 연구에서는 확률적으로 밝기와 대조를 조정하는 데이터 증강 기법을 적용하였다. 밝기와 대조를 조절하는 요소는 매 입력 이미지마다 Uniform(0.6, 1.4) 분포로부터 무작위로 추출된 값을 사용하였다. 데이터 증강 기법은 영역 분할 task에서 항상 적용하였으며, 사전과제에서는 동일한 데이터 증강 기법을 적용한 것과 적용하지 않은 것으로 나눠 비교실험 하였다. 이를 통해 본 연구에서는 사전과제에 적용되는 데이터 증강 기법이 도메인 일반화 성능에 영향을 미칠 수 있음을 확인하였다.

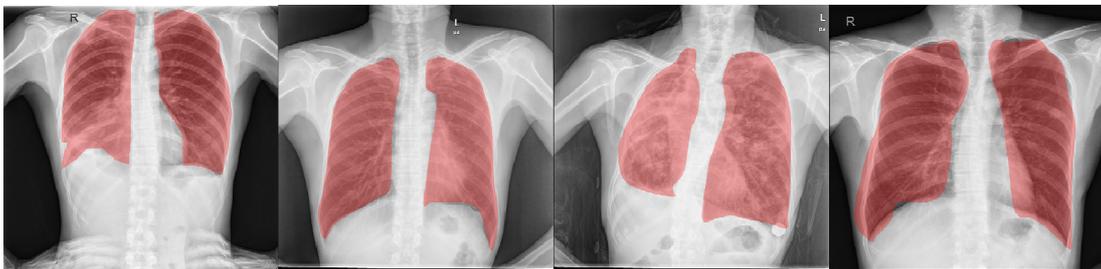


Figure 3. Examples of Mislabeled Data in SZ Dataset

4.2 성능 평가 지표

(1) Jaccard Similarity Coefficient(JSC)

JSC는 모델이 예측한 영역과 정답 이미지의 영역이 얼마나 겹치는지를 측정하기 위한 지표이다. 아래의 식 (6)처럼 두 영역의 교차영역 넓이를 합 영역의 넓이로 나눈 값을 뜻한다. 이때 X는 모델이 예측한 폐의 영역을 의미하고, Y는 실제 정답 이미지에서 폐의 영역을 의미한다.

$$JSC = \frac{|X \cap Y|}{|X \cup Y|} \quad (6)$$

(2) Dice Similarity Coefficient(DSC)

DSC는 영역 분할 모델의 성능 평가 척도 중 하나이며, 앞서 언급한 JSC와 계산 방식이 유사하다. JSC와 마찬가지로 예측 영역과 정답 영역이 많이 겹칠수록 DSC 점수는 높아진다. 단, JSC는 잘못된 분류가 발생했을 때 더 많은 페널티를 부여하는 지표임에 반해, DSC는 평균 성과에 가까운 값을 측정하는 경향이 있다. 따라서 두 성능 지표를 모두 사용하여 모델을 적절하게 평가할 필요가 있다.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (7)$$

(3) Average Contour Distance(ACD) & Average Surface Distance(ASD)

ACD와 ASD는 거리 기반의 지표로, 예측된 영역의 위치가 정답 영역과 얼마나 멀리 떨어져 있는가에 집중을 한다. 모델이 예측한 영역의 경계와 실제 경계 간의 거리가 클수록 페널티를 부여하기 때문에 JSC와 DSC가 고려하지 못하는 점을 보완할 수 있다. s_i 를 영역 분할된 경계 S의 i 번째 픽셀이라고 하고, g_j 를 정답 영역의 경계 G의 j 번째 픽셀이라고 하자. 이때 G와 s_i 간의 최소 거리를 $d(s_i, G) = \min \|g_j - s_i\|$ 라고 정의하면 ACD와 ASD는 아래의 식 (8), 식 (9)에 따라 계산된다. 이때 n_s 와 n_g 는 S, G의 픽셀 수를 의미한다.

$$ACD(S, G) = \frac{1}{2} \left(\frac{\sum_i d(s_i, G)}{n_s} + \frac{\sum_j d(g_j, S)}{n_g} \right) \quad (8)$$

$$ASD(S, G) = \frac{1}{2} \left(\sum_i d(s_i, G) + \sum_j d(g_j, S) \right) \quad (9)$$

4.3 실험 결과

<Figure 4>는 각각의 사전과제로 학습이 완료된 모델에 테스트 이미지를 입력해 나온 결과이다. 모델이 원본 이미지와 비슷한 이미지를 올바르게 복원시키기 위해서는 갈비뼈와 폐의 상대적인 위치 등과 같은 전체적인 구조에 대한 이해가 필수적이다. 우리는 이 과정을 통해서 학습된 모델이 해부학적 지식을 습득했을 것이라 가정한다.

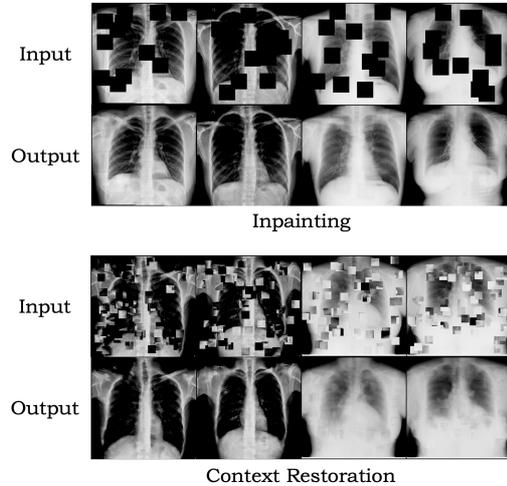


Figure 4. Results from Models Trained with Pretext Tasks

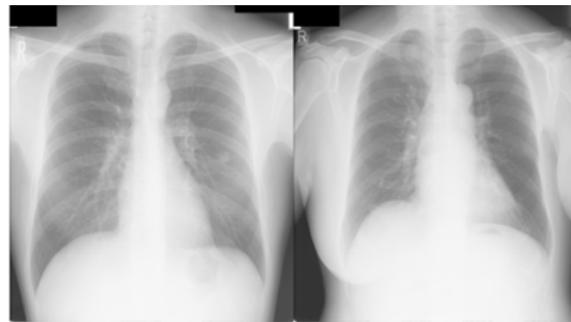
<Table 1>은 사전학습 단계에도 데이터 증강 기법을 적용했을 때에 대한 영역 분할 성능을 5번 반복 실험하여 기록한 값이다. 유의수준이 0.1일 때 baseline에 대비하여 성능이 유의하게 향상된 경우는 *를 표시했다. 사전과제로 사전학습 한 경우의 대부분이 baseline보다 소스와 타겟에 대해서 더욱 좋은 성능을 기록하였다. 특히 타겟 도메인에 대한 영역 분할 성능이 상대적으로 많이 향상된 것을 확인할 수 있었으며, 그 중에서도 거리기 반 지표의 성능 향상이 돋보였다. 소스 도메인이 JSRT인 경우 논문에서 제안하는 방법이 baseline보다 모두 좋은 성능을 보였고 전체적인 결과로 미루어 보아 context restoration task보다 inpainting task가 도메인 일반화 성능 향상에 더 효과적인 것으로 보인다. 특히 타겟 도메인이 SZ인 경우, context restoration과 inpainting 모두 baseline에 비해 모든 부분의 성능이 큰 폭으로 향상되었다. 한편 소스 데이터 셋이 MC이고 타겟 데이터 셋이 JSRT일 때는 논문에서 제안하는 방법이 baseline보다 더 낮은 성능을 보였다. 소스 도메인이 SZ이고 타겟 데이터 셋이 JSRT인 경우도 마찬가지로 기존의 baseline이 더 좋은 영역 분할 성능을 기록했다. 반면에 소스 데이터 셋이 MC이고 타겟이 SZ인 경우와 반대의 경우에는 논문에서 제안하는 방법이 모든 지표에서 더 좋은 결과를 보였다. 따라서 결절환자의 이미지로 모델을 사전학습 한 경우는 결핵환자에 대해 일반화 성능을 향상시키는데 도움이 되지만, 결핵환자 이미지로 사전학습을 한 경우는 결절환자의 도메인에 대해 상대적으로 일반화 성능이 떨어진다는 결론을 내릴 수 있다. 본 연구에서는 이러한 결과가 결절환자와 결핵환자 간의 병변 영역의 크기 차이로 인해 발생한다고 추측한다(<Figure 5> 참고). 사전학습을 하는 동안 모델은 X-ray 이미지의 국지적인 영역이 아닌 넓은 영역을 보고 이미지의 전체적인 구조를 학습하게 된다. 결핵환자가 포함된 도메인으로 사전학습 하는 경우, 결핵환자의 병변 영역은 폐 전반에 퍼져 있기 때문에 전체적인 구조를 학습하는 방법이 오히려 모델이 폐의 경계가 모호한 구조를 학습하는 결과를 초래하게 된다. 이로 인해 모델이 다른 도메인인 결절환자의 X-ray 이미지를 만났을 때 올바르게 폐의 경계를 분할하는데 어려움을 겪는 것으로 예상

Table 1. Segmentation Performance with Data Augmentation at Pretraining Phase

			Baseline	Context restoration	Inpainting
Source domain	JSRT	JSC ↑	0.958±0.004	0.957±0.003	0.959±0.002
		DSC ↑	0.978±0.002	0.978±0.002	0.979±0.001
		ACD ↓	0.997±0.141	0.983±0.063	0.943±0.045
		ASD ↓	0.698±0.036	0.715±0.037	0.688±0.022
Target domain 1	MC	JSC ↑	0.901±0.015	0.912±0.013	0.912±0.011
		DSC ↑	0.946±0.009	0.952±0.007	0.952±0.008
		ACD ↓	3.028±1.345	1.976±0.284	2.009±0.233
		ASD ↓	1.226±0.207	1.051±0.093	1.037±0.055
Target domain 2	SZ	JSC ↑	0.875±0.025	0.904±0.012*	0.906±0.005*
		DSC ↑	0.932±0.015	0.949±0.007*	0.950±0.003*
		ACD ↓	4.101±1.075	2.655±0.839*	2.485±0.21*
		ASD ↓	1.481±0.215	1.183±0.1348*	1.173±0.037*
			Baseline	Context restoration	Inpainting
Source domain	MC	JSC ↑	0.956±0.003	0.954±0.006	0.957±0.003
		DSC ↑	0.977±0.002	0.976±0.003	0.978±0.002
		ACD ↓	0.970±0.127	1.103±0.205	0.983±0.127
		ASD ↓	0.656±0.022	0.673±0.037	0.663±0.021
Target domain 1	JSRT	JSC ↑	0.880±0.03	0.821±0.086	0.855±0.043
		DSC ↑	0.934±0.019	0.890±0.066	0.917±0.029
		ACD ↓	2.836±0.6	4.272±1.914	3.473±0.931
		ASD ↓	1.291±0.123	1.503±0.34	1.414±0.176
Target domain 2	SZ	JSC ↑	0.867±0.026	0.893±0.018	0.894±0.005*
		DSC ↑	0.926±0.016	0.942±0.011	0.943±0.003*
		ACD ↓	5.107±1.147	3.728±0.894*	3.558±0.611*
		ASD ↓	1.608±0.221	1.337±0.145*	1.345±0.0688*
			Baseline	Context restoration	Inpainting
Source domain	SZ	JSC ↑	0.929±0.002	0.930±0.003	0.931±0.002
		DSC ↑	0.963±0.001	0.963±0.002	0.964±0.002
		ACD ↓	1.437±0.05	1.422±0.066	1.448±0.069
		ASD ↓	0.936±0.021	0.921±0.016	0.925±0.022
Target domain 1	MC	JSC ↑	0.931±0.004	0.935±0.008	0.936±0.004
		DSC ↑	0.964±0.002	0.966±0.004	0.967±0.002*
		ACD ↓	1.534±0.083	1.445±0.148	1.401±0.069*
		ASD ↓	0.976±0.048	0.934±0.088	0.911±0.046*
Target domain 2	JSRT	JSC ↑	0.935±0.004	0.923±0.023	0.931±0.01
		DSC ↑	0.966±0.002	0.959±0.013	0.964±0.005
		ACD ↓	1.474±0.09	1.789±0.530	1.609±0.244
		ASD ↓	0.951±0.033	1.020±0.142	0.977±0.078



Tuberculosis patient's X-ray images in SZ



Solitary Pulmonary Nodule (SPN) patient's X-ray images in JSRT

Figure 5. Examples of Tuberculosis and SPN Patient's X-ray Image

한다. 한편 결절환자가 포함된 도메인으로 사전학습 하는 경우, 결절환자의 병변 영역은 폐 내부의 일부분에 위치하기 때문에 올바른 해부학적 구조를 학습하는 과정에 영향을 미치지 않는다.

따라서 결절환자 도메인으로 사전학습을 거친 모델은 올바른 폐의 구조를 학습할 수 있게 되고, 결핵환자의 X-ray 이미지를 만났을 때에도 정답에 가까운 폐의 경계를 찾을 수 있게 된다.

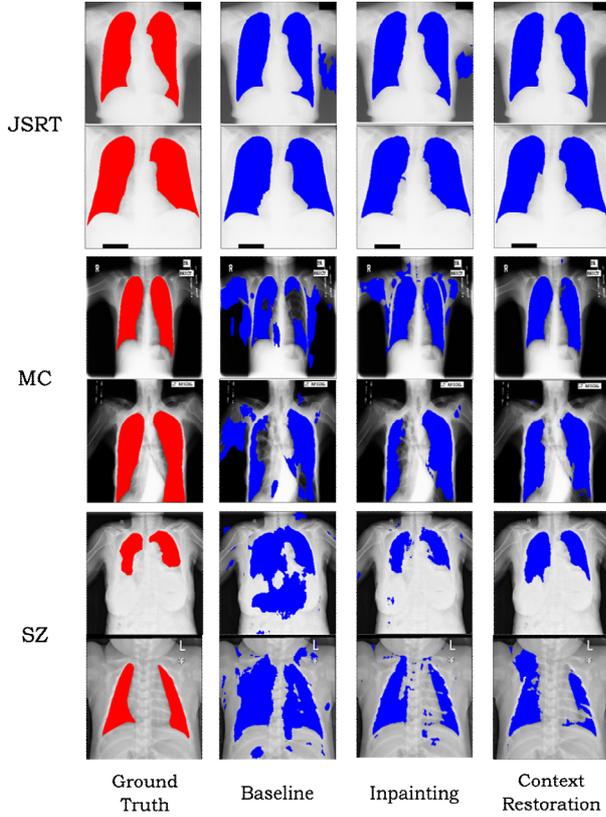


Figure 6. Results of Lung Segmentation

<Figure 6>은 JSRT 데이터 셋을 소스 데이터 셋으로 학습한 각 모델의 폐 영역 분할 결과를 시각화 한 예시이다. 맨 왼쪽 열부터 정답 라벨, 사전학습을 하지 않은 baseline 모델의 예측 결과, inpainting task로 사전 학습한 모델의 예측 결과, 그리고 context restoration task로 사전 학습한 모델의 예측 결과 순으로 시각화 하였다. 소스 도메인인 JSRT 데이터 셋에서는 첫 번째와 두 번째 행에서 볼 수 있듯이 사전 학습한 모델들에서는 baseline이 폐로 잘못 예측한 영역이 작아지거나 없어진 것을 볼 수 있다. 타겟 도메인인 MC와 SZ 데이터 셋에서는 baseline이 좋지 못한 예측 결과를 보였으며, 도메인 변화에 취약한 모습을 드러냈다. 그러나 사전학습을 한 모델에서는 폐에 해당

하는 영역의 윤곽선을 보다 세밀하게 잡거나, baseline에서 잘못 폐로 예측한 영역을 올바르게 예측하는 모습을 보여주고 있다. 특히 SZ 데이터 셋의 경우 폐의 영역이 전체 이미지의 크기에 비해 작은 영역이거나 영유아의 폐와 같이 일반적인 폐와 상이하여 난이도가 어려운 데이터가 포함되어 있다. 이에 대해 baseline은 영역 분할을 제대로 해내지 못하였으나, 사전 과제로 학습한 모델들은 폐 영역을 대략적으로 분할해내는 결과를 볼 수 있다.

<Table 2>는 baseline 대비 평균 성능 상승률을 나타낸 것이다. 위의 영역은 사전과제에도 동일한 데이터 증강 기법 적용한 경우이고, 아래 영역은 영역 분할 task에만 증강 기법을 사용한 경우이다. 데이터 증강 기법을 모두 사용한 경우를 보면, 전반적으로 타겟 도메인에 대한 영역 분할 성능이 상승한 것을 알 수 있다. 이를 통해 사전학습이 도메인 일반화 성능에 미치는 긍정적인 영향을 확인하였다. 또한 inpainting의 거리 기반 성능(ACD, ASD)이 10.22%, 5.73% 정도의 큰 성능 향상을 이룬 것으로 보아 데이터 증강 기법이 거리 기반 지표에 많은 영향을 미치는 것을 알 수 있다. 반면에 사전학습 시 데이터 증강 기법을 사용하지 않은 경우는 모든 부문에서 baseline보다 낮은 성능을 기록했다. 특히 inpainting의 경우 데이터 증강 기법의 적용 여부에 따라 성능 차이가 극심하다. 소스 도메인에서는 JSC 기준으로 1.73% 하락하였고, 타겟 도메인에서는 훨씬 큰 수치인 9.48% 하락하였다. Context restoration의 경우엔 성능 차이가 inpainting보다는 작지만, 여전히 데이터 증강 기법을 모두 적용한 경우에 비해 성능이 크게 떨어진다. 소스 도메인에서는 JSC 기준으로 0.51% 하락하였고, 타겟 도메인에서는 3.9% 하락하였다. 이 결과로부터 사전과제에 적용되는 데이터 증강 기법이 도메인 일반화의 성능에 큰 영향을 미치는 것을 알 수 있으며, 사전과제의 종류에 따라 그 영향력이 다르게 적용되는 것을 알 수 있다.

앞선 실험을 통해 사전과제를 학습하는 과정이 모델의 일반화 성능을 향상 시키는 데에 도움이 될 수 있다는 사실을 보였다. 그러나 이러한 효과성이 모델의 해부학적 구조에 대한 이해로부터 기인된 것인지 검증할 필요가 있다. 따라서 본 연구에서는 사전학습을 수행한 모델이 의도한 대로 해부학적

Table 2. Average Improvement of Segmentation Performance(%)

	With Data Augmentation							
	Source domain				Target domain			
	JSC	DSC	ACD	ASD	JSC	DSC	ACD	ASD
Context restoration	-0.1	-0.06	-3.76	-1.52	0.01	-0.18	2.87	3.19
Inpainting	0.15	0.08	1.06	0.19	0.86	0.42	10.22	5.73
	Without Data Augmentation							
	Source domain				Target domain			
	JSC	DSC	ACD	ASD	JSC	DSC	ACD	ASD
Context restoration	-0.61	-0.34	-12.92	-2.64	-3.89	-2.65	-27.28	-8.34
Inpainting	-1.58	-0.85	-48.78	-18.94	-8.62	-6.02	-63.82	-29.43

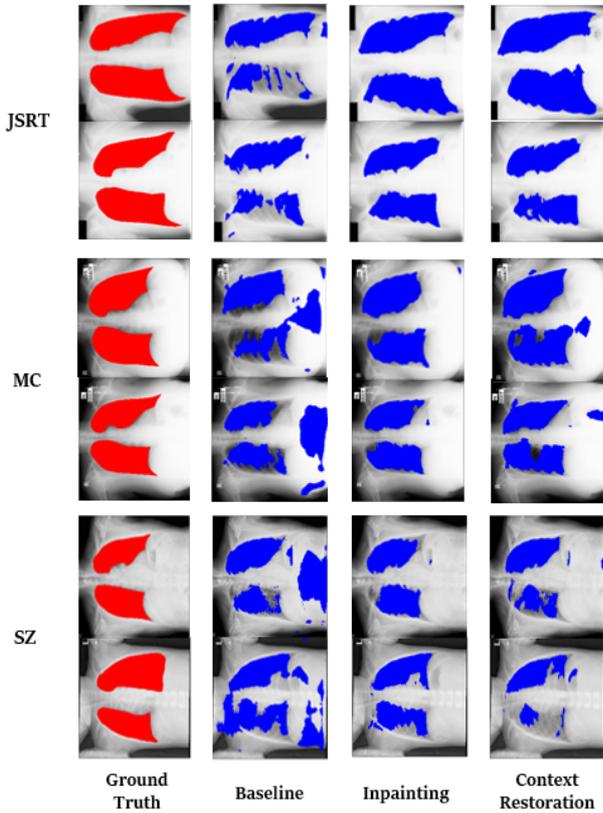


Figure 7. Results of Lung Segmentation for 90 Degree Rotated Test Images

구조를 이해하고 있는지 알아보기 위해 추가적인 실험을 진행하였다. 사전학습을 통해 모델이 해부학적 구조를 이해했다면 모델은 신체 내부 진단 영역의 상대적인 위치를 파악하고 있는 것이기 때문에 의료진과 마찬가지로 회전된 X-ray 이미지를 보고도 폐의 영역을 파악할 수 있어야 한다. 이를 확인하기 위해 사전학습을 진행하지 않은 baseline 모델과 사전학습을 진행한 모델들을 대상으로 90도 회전된 테스트 데이터 셋에 대한 영역 분할 성능을 비교하였다. 소스 데이터 셋으로는 JSRT 데이터 셋을 사용하였고, <Table 3>에 그 결과를 나타내었다. 사전학습을 한 경우가 baseline보다 소스와 타겟에 대해서 더욱 좋은 성능을 기록하였으며, 테스트 이미지를 회전시키지 않은 경우에 대비해 성능이 상대적으로 많이 향상된 것을 확인할 수 있었다(<Table 1> 참고). <Figure 7>은 90도 회전된 이미지에 대한 영역 분할 결과를 시각화 한 예시이다. Baseline은 소스와 타겟 도메인 모두에 대해 회전된 X-ray 이미

지의 폐 영역 분할을 제대로 해내지 못하였으나, 사전 학습한 모델의 경우 잘못 폐로 예측한 영역이 상대적으로 줄어든 것을 알 수 있다. 이를 통해 사전 학습한 모델은 이미지의 변화에 강건하다는 사실을 알 수 있고, 해부학적 구조를 파악하고 있다는 결론을 내릴 수 있다.

5. 결론

본 연구에서는 인간이 가지고 있는 해부학적 지식을 폐 영역 분할 문제를 수행하는 심층 신경망에 반영하는 것을 목표로 한다. 이를 위해 자기 지도 학습으로 모델을 사전학습 하는 방법을 제안했으며, 해당 방법이 하나의 도메인으로 구성된 학습 데이터만으로도 도메인 변화에 대한 강건성을 지닐 수 있음을 실험을 통해 확인하였다. 따라서 본 논문에서 제안하는 방법은 데이터를 마련하는데 필요한 시간과 비용을 최소화하면서도 환자 개개인의 특성과 다양한 영상 데이터의 도메인에 강건한 방법이라 할 수 있다. 그러나 도메인 일반화 성능은 사전과제에 적용되는 데이터 증강 기법에 민감하기 때문에 실험 과정에서 적절한 사전과제의 종류와 데이터 증강 기법을 찾아 줄 필요가 있다. 또한 본 논문에서 제안하는 방법은 영역 분할 task를 수행하기 전에 사전학습 단계를 거쳐야하기 때문에 시간적인 추가 비용이 발생한다. 하지만 새로운 도메인의 데이터 셋을 필요로 하지 않기 때문에 데이터 셋이 제한적인 경우에는 보다 실용적인 방법이 된다.

제안하는 접근 방법의 효과를 확인하기 위해 폐 영역 분할 문제에 대해서 도메인 일반화 성능 향상을 검증하였다. 그러나 분류 문제 및 병변 탐지 등과 같은 다양한 알고리즘과 척추 X-ray, 뇌 MRI 이미지 등의 다양한 데이터 셋에서도 쉽게 적용할 수 있는 방법이기 때문에 적절한 사전과제를 적용한다면 동일한 효과를 낼 것으로 기대한다. 또한 본 연구에서는 하나의 사전과제를 이용해 모델에 해부학적 지식을 학습시키고자 하였으나, 더 나아가 다양한 사전과제를 모델에 동시에 학습시키는 과정이 도메인 일반화 성능에 어떠한 영향을 미치는지에 대한 경향성을 살펴볼 필요가 있다. 또한 이미지의 전체적인 특징을 학습하도록 제안된 구조를 사용함으로써 모델의 구조 관점에서 해부학적 지식을 반영하거나, 사전학습을 사용하는 방법과 모델의 구조를 활용하는 방법을 결합하는 연구 또한 의미 있는 연구 방향으로 제시한다.

Table 3. Segmentation Performance for 90 Degree Rotated Test Images

	Source domain(JSRT)				Target domain 1+Target domain 2(MC+SZ)			
	JSC ↑	DSC ↑	ACD ↓	ASD ↓	JSC ↑	DSC ↑	ACD ↓	ASD ↓
Baseline	0.615	0.757	9.603	2.655	0.506	0.665	12.546	3.214
Context restoration	0.642	0.776	8.878	2.503	0.511	0.668	11.453	2.917
Inpainting	0.652	0.785	8.350	2.407	0.541	0.690	10.334	2.790

참고문헌

- Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., and Rueckert, D. (2019), Self-Supervised Learning for Medical Image Analysis Using Image Context Restoration, *Medical Image Analysis*, **58**, 101539.
- Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020), A Simple Framework for Contrastive Learning of Visual Representations, *arXiv preprint arXiv:2002.05709*.
- Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. (2015), How Much Data is Needed to Train a Medical Image Deep Learning System to Achieve Necessary High Accuracy?, *arXiv preprint arXiv:1511.06348*.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017), Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks, *Nature*, **542**(7639), 115-118.
- Gidaris, S., Singh, P., and Komodakis, N. (2018), Unsupervised Representation Learning by Predicting Image Rotations, *arXiv preprint arXiv:1803.07728*.
- Gulshan et al. (2016), Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, *Jama*, **316**(22), 2402-2410.
- Hendrycks, D., Mazeika, M., Kadavath, S., and Song, D. (2019), Using Self-Supervised Learning can Improve Model Robustness and Uncertainty, *In Advances in Neural Information Processing Systems*, 15663-15674.
- Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., and Lakshminarayanan, B. (2019), Augmix : A Simple Data Processing Method to Improve Robustness and Uncertainty, *arXiv preprint arXiv:1912.02781*.
- Jaeger, S., Candemir, S., Antani, S., Wang, Y. X. J., Lu, P. X., and Thoma, G. (2014), Two Public Chest X-ray Datasets for Computer-Aided Screening of Pulmonary Diseases, *Quantitative Imaging in Medicine and Surgery*, **4**(6), 475-477.
- Kaggle, <https://www.kaggle.com/yoctoman/shcxr-lung-mask>
- Kim et al. (2020), Changes in Cancer Detection and False-Positive Recall in Mammography Using Artificial Intelligence : A Retrospective, Multireader Study, *The Lancet Digital Health*, **2**(3), e138-e148.
- Kingma, D. P. and Ba, J. (2014), Adam : A Method for Stochastic Optimization, *arXiv preprint arXiv:1412.6980*.
- Li, D., Zhang, J., Yang, Y., Liu, C., Song, Y. Z., and Hospedales, T. M. (2019), Episodic Training for Domain Generalization, *In Proceedings of the IEEE International Conference on Computer Vision*, 1446-1455.
- Litjens et al. (2017), A Survey on Deep Learning in Medical Image Analysis, *Medical Image Analysis*, **42**, 60-88.
- Long, M., Cao, Z., Wang, J., and Jordan, M. I. (2018), Conditional Adversarial Domain Adaptation, *In Advances in Neural Information Processing Systems*, 1640-1650.
- McKinney et al. (2020), International Evaluation of an AI System for Breast Cancer Screening, *Nature*, **577**(7788), 89-94.
- Motiani, S., Piccirilli, M., Adjeroh, D. A., and Doretto, G. (2017), Unified Deep Supervised Domain Adaptation and Generalization, *In Proceedings of the IEEE International Conference on Computer Vision*, 5715-5725.
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A. (2016), Context Encoders : Feature Learning by Inpainting, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2536-2544.
- Ronneberger, O., Fischer, P., and Brox, T. (2015), U-net : Convolutional Networks for Biomedical Image Segmentation, *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 234-241.
- Shiraishi et al. (2000), Development of a Digital Image Database for Chest Radiographs with and without a Lung Nodule : Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules, *American Journal of Roentgenology*, **174**(1), 71-74.
- Somavarapu, N., Ma, C. Y., and Kira, Z. (2020), Frustratingly Simple Domain Generalization via Image Stylization, *arXiv preprint arXiv:2006.11207*.
- Volpi, R., Namkoong, H., Sener, O., Duchi, J. C., Murino, V., and Savarese, S. (2018), Generalizing to Unseen Domains Via Adversarial Data Augmentation, *In Advances in Neural Information Processing Systems*, 5334-5344.
- Xu, J., Xiao, L., and López, A. M. (2019), Self-Supervised Domain Adaptation for Computer Vision Tasks, *IEEE Access*, **7**, 156694-156706.
- Zhang et al. (2019), When Unseen Domain Generalization is Unnecessary? Rethinking Data Augmentation, *arXiv preprint arXiv:1906.03347*.
- Zhang, R., Isola, P., and Efros, A. A. (2016), Colorful Image Colorization, *In European Conference on Computer Vision*, Springer, Cham, 649-666.

저자소개

이예진 : 서울과학기술대학교 산업공학과에서 2020년 학사학위를 취득하고 서울과학기술대학교에서 데이터사이언스학과 학석사 연계과정에 재학 중이다. 연구 분야는 딥러닝 방법론 개발 및 응용이다.

이상우 : 서울과학기술대학교 산업공학과 학석사 연계과정에 재학 중이다. 연구분야는 딥러닝 방법론 개발 및 응용이다.

황상훈 : KAIST 산업및시스템공학과에서 2005년 학사, 2012년 박사학위를 취득하였다. 삼성전자 종합기술원과 루닛에서 연구원으로 재직했고 2018년부터 서울과학기술대학교 산업공학과 조교수로 재직 중이다. 주요 연구 분야는 기계학습/딥러닝 방법론 개발 및 응용 등이다.