

# 수정된 구조의 AE 모델을 이용한 X-ray 산란선 보정 기법

서효경, 정지훈, 이동규, 한승화, 김호준  
한동대학교 전산전자공학부

shk9809@naver.com, jihoojo03@naver.com, asas0409@nate.com, hans0525@naver.com, hjkim@handong.edu

## Scattered X-ray Correction Using a Modified Auto-Encoder

Hyogyeeong Seo, Jihoon Jeong, Donggyu Lee, Seunghwa Han, Hojoon Kim  
School of Computer Science and Electrical Engineering, Handong Global University

### 요 약

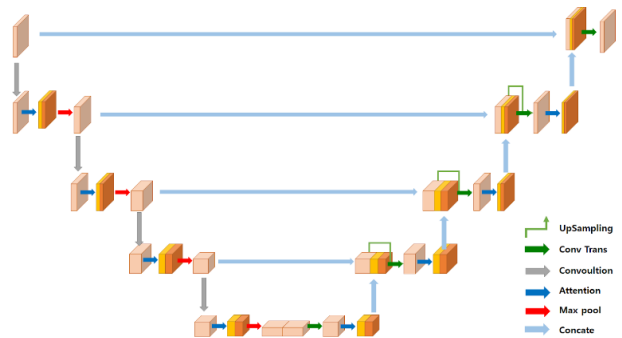
본 논문에서는 X-ray 진단에서 산란선으로 인한 영상의 왜곡을 보정하는 방법으로서 수정된 구조의 AE(Auto-Encoder) 모델에 기반한 방법론을 제안한다. 기존 AE 모델의 계층에 따라 특징지도의 크기가 축소되고 팽창되는 과정에서 영상 복원에 필요한 정보가 소실될 가능성을 보완하기 위하여 동일레벨 계층 간에 스킵 연결을 추가하였다. 또한 X-ray 영상에서 피사체 세부 부위의 두께와 밀도에 따라 산란선의 영향이 서로 다른 형태로 나타난다는 특성을 학습 과정에 효과적으로 반영하기 위하여 어텐션 모듈을 추가한 네트워크 구조를 도입하였다. 총 80 쌍의 흉부 X-ray 영상 데이터에 대하여 기존의 AE 모델을 사용한 방법 및 U-Net 과 FFA-Net 모델을 사용한 영상 복원 기법의 실험 결과를 상호 비교함으로써 제안된 방법의 타당성을 평가하였다.

### 1. 서론

X-ray 진단에서 신호의 굴절현상으로 인한 산란선(scattered radiation)은 영상의 선명도를 감소시킬 뿐만 아니라 다양한 형태의 왜곡을 발생시킬 수 있다. 이러한 산란선을 제거하기 위하여 일반적으로 하드웨어 그리드(Grid) 장치를 사용하고 있으나, 이로 인해 진단과정에서 피폭되는 방사선량이 다소 증가할 수 있고 영상 내에 그리드의 영향으로 인한 아티팩트가 발생할 수 있다는 점이 단점으로 지적된다. 이에 최근에는 산란선의 영향을 보정하는 과정을 소프트웨어 기법으로 구현하고자 하는 연구가 수행되고 있다. 최근의 연구에서 영상의 노이즈와 왜곡을 보정하기 위한 영상복원 기법으로서 다양한 CNN 기반의 심층 신경망 모델이 제시된 바 있다[1-3]. U-Net 기반 모델을 사용하여 불완전한 학습 데이터만으로 원영상을 복원하는 학습방법이 제안되었고[1], 수축 단계와 팽창 단계를 통과하며 각 계층 구조를 학습하는 AE(Auto-Encoder) 모델이 제안되었다[2]. 또한 입력 영상에서 상대적으로 중요한 채널과 픽셀을 판단하여 서로 다른 가중치를 적용하는 FFA-Net 이 제안되었다[3].

본 연구에서는 X-ray 산란선 보정을 위하여 수정된 구조의 AE 모델을 제안하고 그 유용성을 실험적으로 평가한다. (그림 1)에 보인 바와 같이 기존의 AE 모델

의 형태에서 수축 단계와 팽창 단계의 동일 레벨 계층에 스킵 연결을 추가하였으며, 각 계층에 어텐션(Attention) 모듈[4]을 추가하였다. 일반적인 흉부 X-ray 영상에서 산란선은 세부 조직의 밀도에 따라 굴절률이 달라지며 특정 조직과 피사체 영상의 영역에 따라 서로 다른 형태로 나타난다. 제안된 모델에서 어텐션 모듈은 영상의 위치와 조직의 특성에 따라 상이하게 나타나는 산란선의 특성을 효과적으로 학습할 수 있게 한다. 본 연구에서는 실험을 통하여 이러한 방법이 동일한 계층구조를 가진 기존의 다른 영상 복원 모델에 비하여 산란선 보정 성능을 개선할 수 있음을 보인다. 또한 기존의 AE 모델을 적용한 경우와 비교함으로써 어텐션 모듈의 효과를 실험적으로 평가한다.



(그림 1) 수정된 구조의 AE 모델 구조

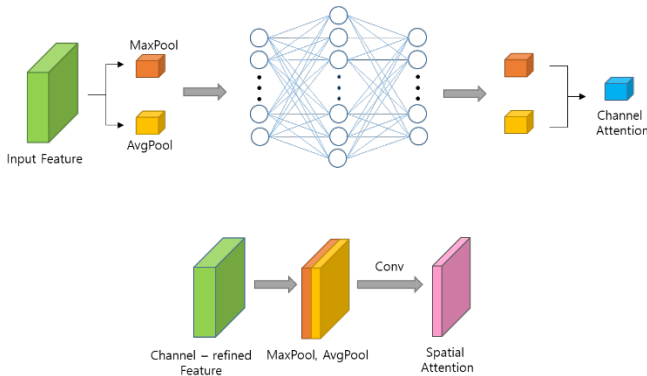
## 2. 수정된 구조의 AE 모델

수정된 구조의 AE 모델은 CNN 을 기반으로 하는 모델로서, 수축 단계의 특징지도 생성 과정과 팽창 단계의 영상 복원 과정에서 정보의 소실을 방지하고 학습의 효과를 개선하기 위하여 스킵 연결과 어텐션 모듈을 추가한 모델이다. 팽창 단계의 정보 확장 과정 중에 수축 단계에서 생성한 특징 정보를 추가하는 스킵 연결은 영상 복원에 필요한 정보를 제공한다. 어텐션 모듈은 특징 정보의 중요도를 선별적으로 반영하는 채널 어텐션(Channel Attention) 모듈과 픽셀에 대한 상대적 중요도를 반영하는 위치 어텐션(Spatial Attention) 모듈이 결합된 형태로 구성된다. 어텐션 모듈을 구성하는 각각의 세부 모듈은 학습단계에서 ‘무엇’과 ‘어디’에 초점을 맞춰 특징지도에서 강조하거나 억제할 대상과 영역을 판별한다.

채널 어텐션 과정은 식 (1)과 같이 이루어진다. 입력값에 대해 평균 풀링과 최대 풀링을 각각 수행하여 채널 어텐션 지도를 생성하고, 이를 동일한 다층 퍼셉트론을 사용하여 학습한다. 위치 어텐션은 식 (2)와 같이 구성되며 풀링 연산으로 채널의 수를 압축한 2 차원 공간 특징지도에 컨볼루션을 수행한다. 각 어텐션에서 활성화 함수는 시그모이드 함수를 사용한다.

$$M_c(X) = \sigma(MLP(Avg(X)) + MLP(Max(X))) \quad (1)$$

$$M_s(X) = \sigma(f_{7 \times 7}([Avg(X); Max(X)])) \quad (2)$$



(그림 2) 어텐션 모듈의 세부 구조

제안된 모델은 앞 절의 (그림 1)에 보인 바와 같이 수축 단계에서는 컨볼루션 연산을 수행하고 어텐션 모듈을 통과하여 특징 정보를 추출한 뒤 최대 풀링 연산을 수행하여 특징지도를 축소시킨다. 팽창 단계에서는 각 층마다 수축 단계의 해당 층에서 추출된 레이어와 이전 팽창 단계의 필터를 전치시킨 레이어를 더하여 특징 정보를 복원한 뒤 어텐션 모듈을 통과한다. 각 층에서 활성화 함수는 ReLU 함수를 사용한다. 수축과 팽창 단계를 수식으로 나타내면 다음과 같다.

$$Enc_0 = Input \quad (1)$$

$$Enc_n = ReLU(Enc_{n-1} W_{n-1}^{enc} + b_n^{enc}) \quad (0 < n \leq N) \quad (2)$$

$$Enc_n = Max(Enc_n + M_s(Enc_n + M_c(Enc_n))) \quad (0 < n \leq N) \quad (3)$$

$$Dec_N = ReLU(Enc_N(W_N^{enc})^T + b_N^{dec}) \quad (4)$$

$$Dec_n = ReLU(Dec_{n+1}(W_{n-1}^{enc})^T + Enc_n W_n^{dec} + b_n^{dec}) \quad (0 < n < N) \quad (5)$$

$$Dec_n = Dec_n + M_s(Dec_n + M_c(Dec_n)) \quad (0 < n < N) \quad (6)$$

$$Output = Dec_0 = Dec_1 W + Input W_0^{dec} + b_0^{dec} \quad (7)$$

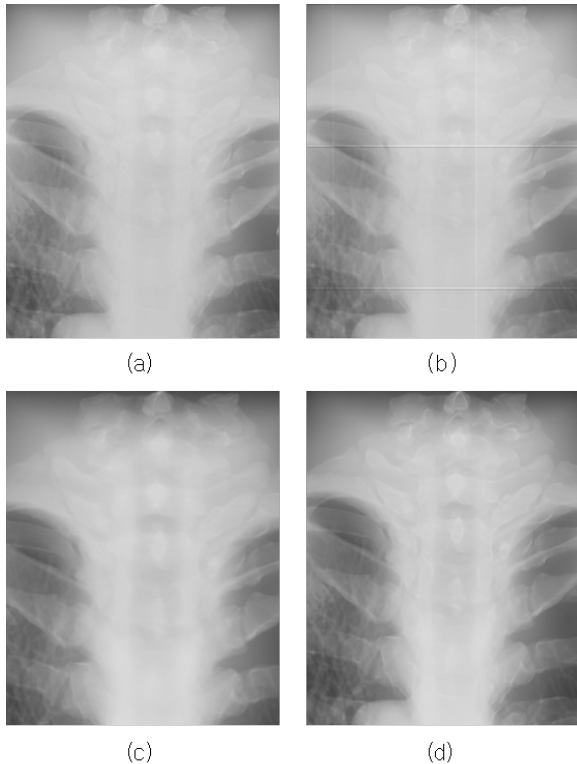
N 개의 층으로 이루어진 모델에서  $Enc_n$ 과  $Dec_n$ 은 각각 n 번째 수축과 팽창 층을 의미한다. 식에서  $W_n^{enc}$ ,  $W_n^{dec}$ 는 n 번째 수축과 팽창 층을 출력으로 하는 컨볼루션 연산에 사용된 가중치이며  $b$ 는 이 때의 편향(bias)이다. 어텐션 모듈은 식 (3), (6)과 같이 잔차 학습을 활용하여 학습 효율을 향상시켰다.

## 3. 실험 결과 및 고찰

본 연구에서는 제안한 모델의 성능을 평가하기 위한 실험을 수행하였다. 흉부 팬텀 모형에 대하여 촬영한 X-ray 영상과, 하드웨어 그리드를 사용한 영상의 쌍으로 총 80 개의 12bit 단일 채널, DICOM 포맷 형태의 실험 데이터를 생성하고 이들 중 학습데이터 65 쌍과 테스트 데이터 15 쌍을 무작위로 선별하였다. 첫 번째 실험에서는 수정된 구조의 AE 모델을 기존의 영상 복원 심층 신경망과 비교하여 성능을 평가하였다. (표 1)은 FFA-Net, U-Net 및 수정된 구조의 AE의 산란선 제거 효과를 정량적으로 비교한 것이다. 영상 복원 지표인 PSNR 과 구조적 유사도 평가 지표인 SSIM 값으로 비교하였을 때 수정된 구조의 AE 를 사용한 산란선 보정 영상이 우수한 영상 복원 성능을 보인다. 또한 기존에 제시된 심층 신경망의 파라미터 수와 학습에 소요되는 시간을 제안된 모델의 것과 비교하였을 때 수정된 구조의 AE 모델이 효과적인 학습을 하고 있음을 알 수 있다.

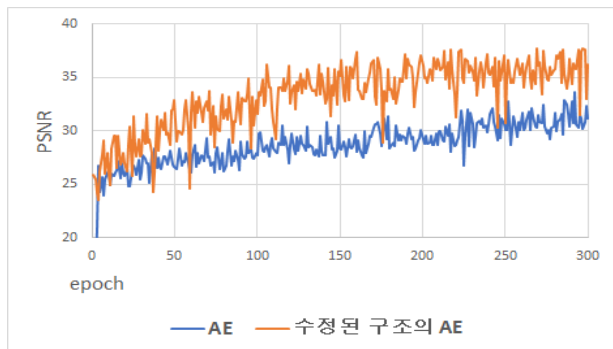
<표 1> 산란선 보정 실험 결과의 정량적 비교

Method	Image Quality Evaluation		# Param	min/epoch
	PSNR	SSIM		
U-Net	29.34	0.9654	2.90M	22.63
FFA-Net	26.14	0.9522	4.46M	178.22
Att. AE	34.78	0.9719	2.97M	31.03



(그림 3) 산란선 보정 실험 결과의 예  
(a) 입력영상, (b) FFA-Net 결과,  
(c) U-Net 결과, (d) 수정된 구조의 AE 결과

(그림 3)은 수정된 구조의 AE 모델을 적용한 결과를 FFA-Net 모델의 결과 및 U-Net 모델의 결과와 상호 비교한 것이다. (그림 3)에 보인 바와 같이 세 가지 모델 모두에서 산란선이 제거됨으로써 영상의 선명도가 상당부분 개선되었다. 그러나 FFA-Net의 경우 폐치의 경계선 영역에서 근접 영역에 대한 정보 부족으로 인한 블로킹 이펙트가 일부 발견되었고, U-Net 모델은 수정된 구조의 AE 모델에 비하여 상대적으로 평균 성능이 다소 미흡함을 보였다.



(그림 4) epoch에 따른 모델별 영상 지표 추이

두 번째 실험에서는 기존의 AE 모델과 수정된 구조의 AE 모델을 비교하여 어텐션 모듈의 유용성을 평가하였다. (그림 4)는 기존 AE 모델과 수정된 구조의 AE에 대하여 epoch 별 학습 효과 변화를 비교한 것이다. 이는 기존 AE 모델에 어텐션 모듈을 추가함으로써 학습 과정에서 유의미한 성능 향상을 보일 수 있

음을 시사한다. 이와 같은 실험 결과는 AE 모델의 수축과 팽창 과정이 일반적인 조직 영역의 특징 값을 입력 영상과 동일한 분포가 되도록 복원할 때에 어텐션 모듈이 특정 조직 영역에 집중적으로 분포되어 있는 산란선을 효과적으로 찾아내어 이를 보정할 수 있음을 보인다.

#### 4. 결론

영상 복원 문제를 위한 딥러닝 기법은 기존의 전형적인 패턴분류 문제와는 달리 입력과 출력 값이 동일한 크기의 데이터로 매핑되는 학습 문제이다. CNN 기반의 AE 모델에서 특징지도의 생성 과정과 영상의 복원 과정 중에 나타나는 특징정보의 소실 가능성을 보완하기 위하여 동일 레벨의 계층 간에 연결구조를 추가함으로써 수축 단계의 각 계층에서 생성한 특징지도 데이터를 팽창 단계의 동일 레벨에서 사용하도록 하였다. 또한 이 과정에서 각 채널 및 픽셀의 위치에 따라 서로 다르게 나타나는 상대적 중요도를 반영하기 위하여 어텐션 모듈을 추가한 구조를 적용하였다. 이는 X-ray 산란선 개선 문제에서 피사체의 조직 영역에 따라 서로 다르게 나타나는 산란선 분포의 특성을 학습 과정에 효과적으로 반영할 수 있게 한다. 이와 같은 연구는 X-ray 영상의 질을 개선하여 진단의 정확도를 향상시킬 뿐만 아니라, 추가 아티팩트를 발생시키지 않는 산란선 억제 방법론으로서 기존의 하드웨어 그리드 장치의 사용을 보완할 수 있게 할 것이다. 향후 연구에서는 학습 과정에서의 파라미터를 최적화하는 연구와 다양한 형태의 X-ray 영상에 적용하여 일반적인 성능을 향상시키는 연구가 필요한 것으로 판단된다.

※ 본 연구는 과학기술정보통신부와 정보통신기술진흥 센터의 소프트웨어중심대학지원사업 (2017-0-00130)의 지원을 받아 수행되었음.

#### 참고문헌

- [1] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, Timo Aila., "Noise2Noise: Learning Image Restoration without Clean Data." arXiv:1803.04189 [cs.CV], 2018.
- [2] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber, "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction", Artificial Neural Networks and Machine Learning – ICANN 2011, 2011, pp.52-59.
- [3] Xu Qin, Zhilin Wnag, Yuanchao Bai, Xiaodong Xie, Huizhu Jia, "FFA-Net: Feature Fusion Network for Single Image Dehazing," arXiv:1911.07559v2 [cs.CV], 2019.
- [4] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S., "Cbam: Convolutional block attention module.", In Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.