

최신 AI 논문 3선(選)

데이터, 의료, 공학 분야의 최신 논문이 각각 1편이 이번 장에 포함되어 있습니다. 첫 번째로, "텍스트의 주요 핵심 내용을 1~2문장 길이로 요약해 주는 기술"을 담고 있는 논문이 소개됩니다. 의료 관련 논문으로는 메탈 아티팩트(metal artifact)를 딥러닝을 활용해 제거하는 방법을 담고 있습니다. 메탈 아티팩트는 CT 영상 촬영 시, 촬영 대상자의 몸에 삽입된 금속성 물질로 인해 영상 내 발생하는 인공 음영을 의미합니다. 마지막 논문은 깊은 강화학습으로 대상을 인식하는 방식을 담고 있는 연구 결과물입니다.

- 데이터** 고려대 데이터 사이언스&비즈니스 애널리틱 연구실: 빅데이터 및 사물인터넷 시대에 핵심적인 분석 방법론 개발 및 실제 사례 적용과 관련된 연구를 한다.
- 의료** 서울대 방사선의학물리연구실: 방사선과 물질의 반응을 통한 에너지 전달과정에서 발생하는 흡수선량 계측과 계산(Radiation Dosimetry)에 중점을 두고 있다.
- 공학** 서울대 인지지능연구실: 로보틱스(robotics)와 컴퓨터 비전(computer vision)을 연구한다.

글 | 김형석 hskim0263@korea.ac.kr

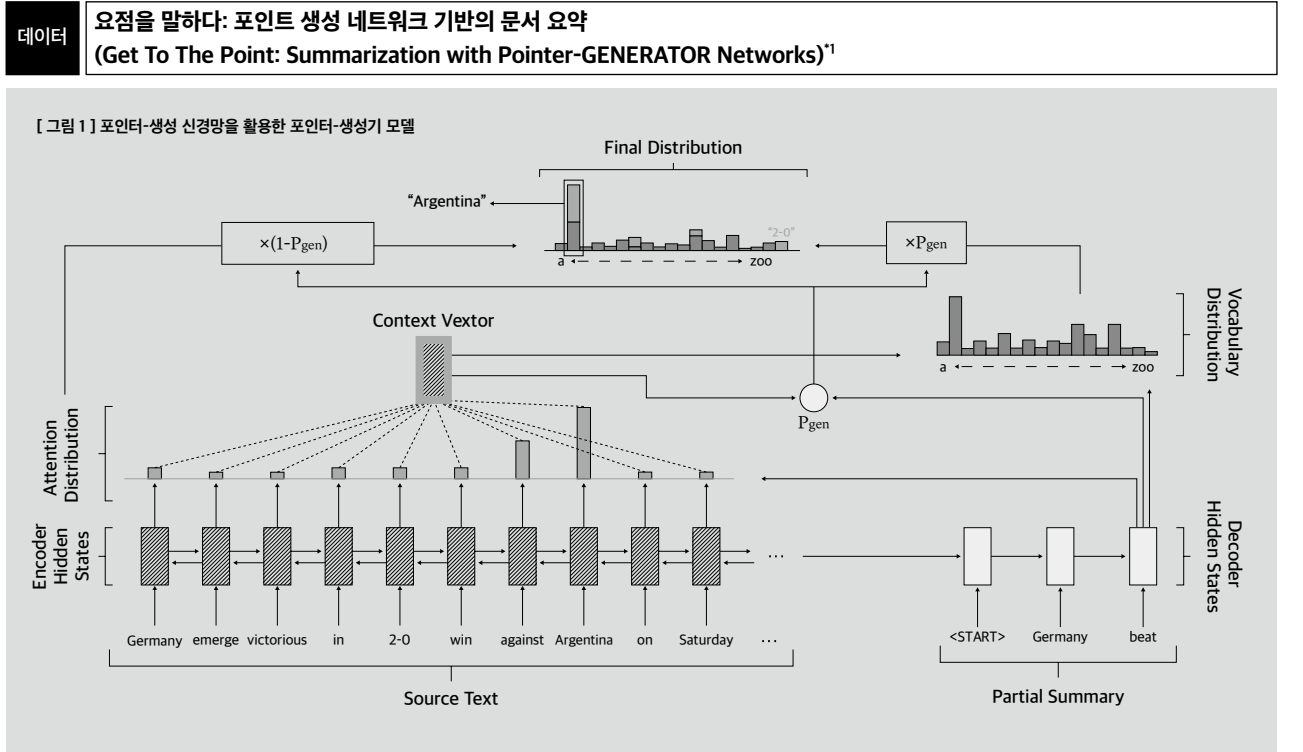
한때는 요리사가 꿈이었지만, 통계가 재미있어 결심한 대학원 석사과정 중 텍스트 마이닝 및 자연어 처리에 빠지게 되어 돌연 석박사 통합과정으로 전환하였다. 현재는 고려대학교 산업경영공학과 박사과정에 있으며, 주 연구 분야는 정보 검색(IR) 및 문서요약(Text Summarization)이다. 투박한 생김새와 다르게 디테일을 추구하는 섬세한 공대생. 언젠가는 매일 보는 논문을 잘 정리 요약 해주는 Tool을 만들어 대학원생들의 워너비스타를 꿈꾸기도 한다.

글 | 이지민 ljm861@gmail.com

늘 행복하고 싶은 대학원생. 학부 때는 딥러닝과 전혀 상관없는 원자력 및 양자공학을 전공하였고, 대학원에 진학하여 의학물리학을 공부하던 중 교수님의 추천으로 딥러닝 공부도 함께 시작하게 되었다. 학부 때 여러 행사를 준비했던 경험이 있어 운이 좋게도 TensorFlowKR 운영진으로 활동하게 되었으며, 지식의 교류가 활발한 이 분야에 엄청난 매력을 느끼고 있다. 현재 라이언의 무심한 귀여움에 반해 짝사랑 중이며, 핸드폰 케이스, 마우스 패드를 포함한 라이언 굿즈를 10개 넘게 보유하고 있다.

글 | 이경재 kyungjae.lee@cpslab.snu.ac.kr

서울대 전기 컴퓨터 공학부를 졸업한 뒤, 동 대학원 석박사 통합과정으로 입학하였다. 현재는 박사과정에 있으며, 주 연구 분야는 모방학습과 지능형 로보틱스이다. 좀 더 세부적으로는 강화 학습과 역강화 학습을 이용하여 로봇을 학습시키는 것이 목표이다. 한동안 군대 문제를 해결하기 위해 연구를 접고 영어공부에 매진했으나 영어실력은 그대로라고 한다. 최근 다시 연구를 시작하였으며 인공지능 및 로보틱스 분야의 많은 사람들과 교류하고 싶다. 석사과정 때는 이론 쪽 공부를 많이 하였으나 이제는 고속 주행 RC카나 매니퓰레이터와 같은 실제 로봇을 다뤄보고 한다.



본 논문 구글 리서치(Google Research)팀이 ACL2017에서 발표한 연구 결과이다. 텍스트(text)로부터 주요 핵심 내용을 1~2문장 길이로 요약해 주는 기술의 소개가 이 논문의 골자이다. 기존 추상적 문서 요약(abstractive text summarization) 방식에서 문제시되던 부정확한 재현성과 반복적인 생성 문제를 각각 'Point-Generator Network'와 'Coverage Mechanism'을 통해 해결했다는 것이 이 연구의 특징이다.

딥러닝 이전 문서 요약의 경우, 규칙 위주(rule based)의 혹은 통계적 기반(statistical based) 방법론이 주를 이루었지만 상용화할 수 있는 부분이 제한적이었고 그 성능 또한 사람에 의한 문서 요약보다 월등히 떨어졌다. 하지만, 현재 RNN(Recurrent Neural Network), CNN(Convolutional Neural Network)과 같은 다양한 딥러닝 기법을 활용하게 되면서 문서 요약(text summarization)은 비약적으로 그 성능을 향상시킬 수 있었다.

문서 요약이란, 주어진 문서로부터 특정 사용자나 작업에 적합한 축약된 형태의 문서로 재생성하는 작업을 말한다. 이를 통해서 복잡도를 줄이면서 필요한 정보를 유지하는 것이 문서 요약의 주목적이다. 일반적으로 문서 요약의 접근법은 크게 2가지 방법으로 나누어 볼 수가 있다. 첫 번째는 추출식 접근법(extractive approach)으로 말 그대로 본문에서 의미 있는 부분을 선택하고 추출하여 이를 재정렬하도록 요약하는 접근법이다. 쉽게 말하면, 일종의 하이라이터(highlighter)라고 생각할 수 있다. 두 번째는 추상적 접근법(abstractive approach)이다. 이는 RNN과 같은

자연어 생성 기술을 활용하여 기존 텍스트(text) 문장을 토대로 창의적인 새로운 문장을 작성하는 방법이다.

이 논문의 연구 결과는 기본적으로는 두 번째 추상적 접근법을 취하면서, 필요 시에는 앞선 추출식 접근법을 하이브리드(hybrid)로 동시에 사용하여 좀더 자연스럽고 함축적인 문서 요약을 가능하게 한다.

시퀀스-투-시퀀스 주목 모델²

라메쉬 날라파티(Ramesh Nallapati) 외 4명의 연구자가 2016년에 발표한 '시퀀스-투-시퀀스 순환신경망을 활용한 추상적 문서 요약'³ 논문의 주목 메커니즘(attention mechanism)을 통한 시퀀스 투 시퀀스 RNN 모델(seq2seq RNN model)⁴을 바탕에 두고 있다. attentional seq2seq RNN model은 크게 4가지 모듈로 구성되어 있다.

1) 인코더 RNN(Encoder RNN)

원문 텍스트로부터 word2Vec으로 표현된 단어들을 각 단어 (word-by-word) 단위로 읽는 모듈(module)이다. 일반적으로 인코더(encoder)는 양방향의 순서를 고려한 양방향 RNN(bidirectional RNN)을 활용한다.

2) 디코더 RNN(Decoder RNN)

요약을 구성하는 단어들의 시퀀스(sequence) 형태로 결과값

(output)을 반환하게 된다. 디코더 RNN(decoder RNN)은 인코더(encoder)와 다르게 한 방향 RNN(directional RNN)을 활용하여, 이전 요약 단계의 단어들을 입력값(input)으로 받게 된다. 원문 텍스트로부터 부분적인 문서 요약이 반환되는 것이다.

3) 주의 분포(Attention Distribution) 와 문맥 벡터(Context Vector)

원문 텍스트로부터 다음 단어를 생성할 때, 원문 텍스트(text)의 단어에 대한 확률인 주목 분포(attention distribution)는 인코더 RNN(encoder RNN)과 디코더 RNN(decoder RNN)의 은닉 상태(hidden state)를 입력값(input)으로 하여 아래와 같은 함수를 통해서 표현된다. 이는 직관적으로 해당 네트워크가 다음 단어를 생성 시 원문 텍스트 중에 어떠한 단어에 주목하여 봐야 할지를 표현해 주는 것이다.

$$e_i^t = v^T \tanh(W_h h_i + W_s s_t + b_{attn})$$

$$a^t = \text{softmax}(e^t)$$

Context Vector

$$h_i^* = \sum_i a_i^t h_i$$

[수식 1, 2]에서의 v^T, W_h, W_s, b_{attn} 등은 네트워크의 학습 과정을 통해서 학습하는 매개변수(parameter)이다. 이러한 주목 분포(attention distribution)를 활용하여 인코더(encoder)의 은닉 상태(hidden state)와 가중합을 통해서 문맥 벡터(context vector)를 생성한다. 이는 디코더(decoder)가 해당 원문으로부터 어떠한 단어를 읽어 오는지를 표현한다..

4) 어휘 분포(Vocabulary Distribution)

최종적으로 문맥 벡터(context vector)와 디코더 은닉 상태(decoder hidden state)의 출력값(output)을 결합하여 2개의 선형 결합 층(layer)을 통해서 단어의 분포(vocabulary distribution)를 표현한다. 이는 원문 텍스트가 아닌 전체 단어(일반적으로 그 사이즈를 제한)에 대한 확률을 표현한다. 이를 통해서 최종적으로 문서 요약을 통해 생성될 단어들이 순서대로 표현된다.

포인터 생성 네트워크

본 논문에서 제안하는 '포인트 생성 네트워크(pointer-generator network)'는 기존 모델(baseline model)로 언급한 상기 '주목 기반 순환신경망 모델(attentional seq2seq RNN model)과 2015년

비날스(Vinyals) 외 2명의 연구자가 제안한 '포인터 네트워크(pointer network)*5의 중간에 있는 하이브리드 네트워크(hybrid network)로 간주될 수 있다. 상기 2개 모델의 하이브리드 네트워크 기능을 수행하는 'Pointer-generator network'를 통해 요약(summarization)단계에서, 기존 텍스트 본문에서의 단어들을 (1-P_{gen})의 확률로 인용하거나 새로운 단어(novel words)들을 P_{gen}확률로 생성할 수 있다. 아래 그림을 통해서 '포인트 생성 네트워크'의 다이어그램(diagram)을 확인할 수 있다.

$$p_{gen} = \sigma(w_h^T h_i^* + w_s^T s_t + w_x^T x_t + b_{ptr})$$

$$p_{gen} \in [0, 1], \text{ soft switch}$$

$$P(w) = p_{gen} P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i=w} a_i^t$$

$$P_{vocab} \text{ is zero}$$

$$\sum_{i:w_i=w} a_i^t \text{ is zero}$$

직관적으로 '포인트 생성 신경망(pointer-generator network)'은 (1) 기존 baseline model의 출력값(vocabulary distribution)과 (2) 원문 텍스트로부터의 주목분포 사이에서 '포인터 생성 신경망' 기반의 학습된 생성 확률(P_{gen})을 반영하여 새로운 단어를 생성하여 문서를 요약(abstractive summarization)하거나, 기존 원문 텍스트를 재현하여 요약(extractive summarization)할지를 학습을 통해서 결정하는 역할을 수행하게 된다. 예를 들면, 생성 확률(P_{gen})이 0에 가까우면 기존 원문 단어로부터 재현된 단어들을 통해 문서 요약이 수행되고, 이와 반대로 생성 확률(P_{gen})이 1에 가까우면 encoder-decoder기반의 생성 단어들을 통해서 문서 요약이 수행되도록 하는 것이다.

Coverage mechanism

기존 주목 기반 순환신경망 모델의 문서 요약 단계에서 특정단어가 재활용되는 문제를 해결하기 위해서, 이 연구에서는 coverage vector를 통해서 일종의 패널티 조건(penalty term)을 손실(loss)에 반영하여 문제를 사전에 방지하고자 하였다. 이는 현재까지 사용되었던 단어의 주목 분포의 누적값을 coverage vector라 정의하고, 이를 'covLos'에 반영하여 요약 단계에서 같은 위치에 반복적으로 등장하는 것을 방지하는 역할을 한다. 이러한 'covLos'를 반영하여 기존 attention distribution과 학습을 위한 Loss는 기존 방법에서 아래와 같이 변화된다.

$$c^t = \sum_{t'=0}^{t-1} a^{t'}$$

$$C^t = \text{softmax}(e^t)$$

$$CO^t = \text{softmax}(e^t)$$

$$e_i^t = v^T \tanh(W_h h_i + W_s s_t + w_c C_i^t + b_{attn})$$

$$a^t = \text{softmax}(e^t)$$

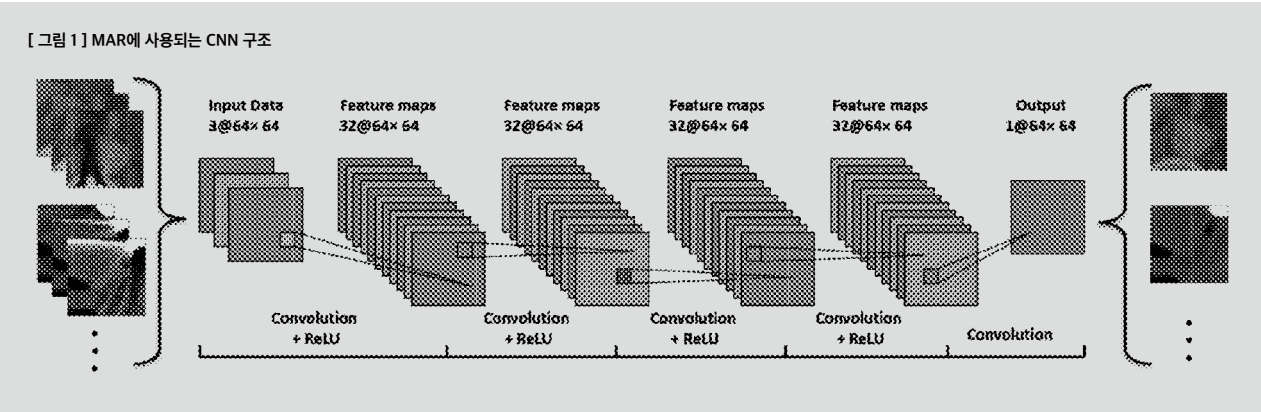
$$covloss_t = \sum_i \min(a_i^t, c_i^t) \quad loss_t = -\log P(w_i^*) + \gamma \sum_i \min(a_i^t, c_i^t)$$

아래는 기존 baseline-model*6과 본 논문의 'Summarization with Pointer-Generator Networks'의 문서 요약 예시이다.

baseline-model*6 과 본 논문의 'Summarization with Pointer-Generator Networks'의 문서 요약 예시
Original Text (truncated): lagos, nigeria (cnn) a day after winning niger- ria's presidency, muhammadu buhari told cnn's christiane amannpour that he plans to aggressively fight corruption that has long plagued nigeria and go after the root of the nation's unrest. buhari said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, he said his administration is confident it will be able to thwart criminals and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. buhari defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.
Baseline Seq2Seq + Attention: UNK UNK says his administration is confident it will be able to destabilize nigeria's economy. UNK says his administration is confident it will be able to thwart criminals and other nigerians. he says the country has long nigeria and nigeria's economy.
Pointer-Gen: muhammadu buhari says he plans to aggressively fight corruption in the northeast part of nigeria. he says he'll "rapidly give attention" to curbing violence in the northeast part of nigeria. he says his administration is confident it will be able to thwart criminals.
Pointer-Gen + Coverage: muhammadu buhari says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

두 번째로 제시된 baseline(seq2seq+attention)의 경우에서 'UNK'으로 표시된 비이상적인 단어들로 인해 부정확한 요약 결과가 생성되는 반면에, 해당 방법론을 사용한 경우 필요 시 원문 텍스트의 어휘들을 재현하면서도, 함축적인 새로운 어휘들을 통해서 문서 요약이 잘 수행된 것을 확인할 수 있다. 이로써 본 접근법이 'abstractive'와 'extractive'의 장점을 잘 반영한 문서 요약이 가능함을 알 수 있다.

의료 CNN 기반 엑스레이 CT 메탈아티팩트 제거 방법
(Convolutional Neural Network based Metal Artifact Reduction in X-ray Computed Tomography)⁷



영상(image) 기반 문제 해결에 CNN과 GAN과 같은 딥러닝 알고리즘이 높은 성능을 보이고 있는 요즘, 의료 AI 분야에서도 의료 영상 데이터를 이용한 연구가 활발히 진행되고 있다. 이번 호에서는 의료 영상 중 잘 알려진 CT (Computed Tomography) 영상의 여러 문제들을 딥러닝을 통해 해결해보고자 하는 논문에 대해 리뷰하고자 한다. 소개하는 논문은 CT 영상에서 발생하는 메탈 아티팩트(metal artifact)⁸를 딥러닝을 이용해 제거해보고자 하는 연구를 담고 있다.

논문 소개

본 논문은 직접 CT 영상 데이터베이스를 만들어 CNN과 기존의 MAR(Metal Artifact Reduction)방법을 적절히 차용하여 향상된 성능을 보여준다. 데이터 확보를 위해 필요한 CT 영상을 수치 시뮬레이션(numerical simulation)하여 직접 생성한 것과, CNN 학습 시 기존에 알려진 MAR 방법인 BHC (Beam Hardening Correction) 와 LI (Linear Interpolation)를 학습에 반영하고자 한 것이 흥미롭다.

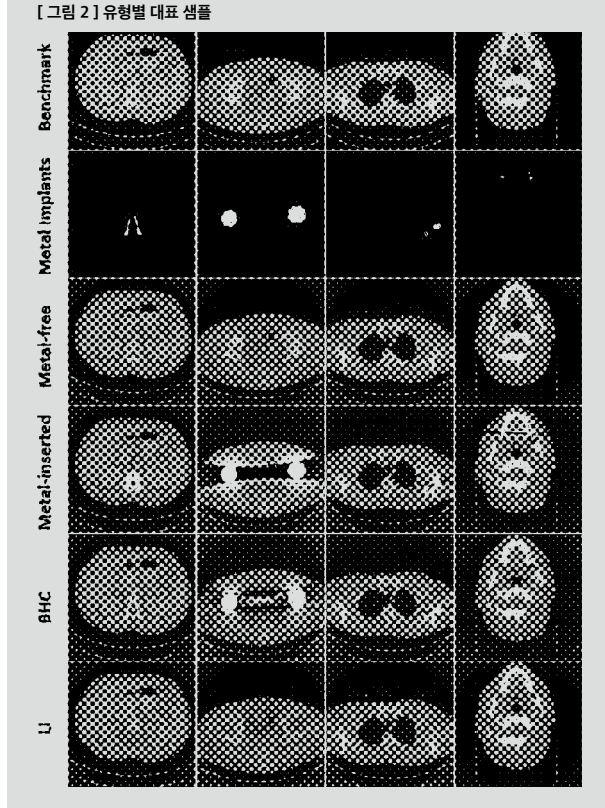
연구 방법

1) Metal Artifact Database 생성(numerical simulation 방법 사용)

- "The 2016 Low-dose CT Grand Challenge" 학습 데이터셋 (training dataset)을 사용하였다.
- 위 데이터 셋에서 메탈 아티팩트가 발생한 영상을 추출하여 금속 부분을 분리(segmentation)하고 바이너리 이미지(binary image)로 저장하는 방법을 통해 [그림 2]의 Metal Implants 에 해당하는 총 15개의 금속 형태(metal shape)를 추출하였다.
- [그림 2]의 Metal-free에 해당하는 74장의 금속이 없는(metal-free) CT에서 추출한 metal shape를 최대한 실제 임상

케이스(clinical case)에 맞도록 금속의 모양, 위치, 재질(티타늄, 철, 구리, 금) 등을 조절하고 삽입하여, 총 100개의 CT 를 합성하였다. 이것이 [그림 2]의 Metal-inserted에 해당한다.

· Metal-inserted 영상을 기존에 사용되던 MAR 방법인 BHC과 LI를 적용하여 각각 추가 영상 [그림 2]의 BHC와 LI 를 획득하였다.



2) CNN 학습

- 입력 : 64×64×3 image patch (Metal-inserted, BHC, LI 영상을 3 채널로 묶고, 작은 패치(patch)로 추출, 원본 이미지 크기 : 512×512)

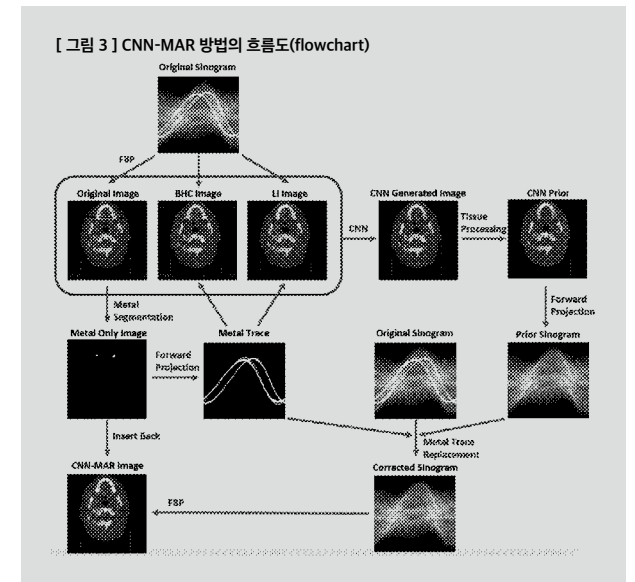
- 타겟 : 64×64 image patch (Input patch 와 동일 위치의 metal-free image 에서 추출한 patch)
- 총 10,000 patch samples 사용했으며 이 중 80%는 학습샘플(training samples)로, 20%는 검증샘플(validation samples)로 사용했다.
- 비용함수(cost function) : $H = argmin_H \frac{1}{R} \sum_{r=1}^R \|H(u_r) - v_r\|_F^2$ ($\| \cdot \|_F$: Frobenius norm, R : 학습샘플의 수, $H(u_r)$: 입력 데이터셋, V_r : 타겟 데이터셋)
- 학습 소요 시간 : 25.5 시간(2000 iterations with GeForce GTX 970 GPU)

3) CNN-MAR method

- [그림 3] 에서 CNN을 접목한 MAR 방법을 도식화하여 보여주고 있다.
- 먼저 원본 사이노그램(Original Sinogram) 으로부터 Original(Metal-inserted)⁹, BHC, LI 영상을 생성하여 3 채널로 묶는다.
- 2)에서 학습된 CNN을 통해 3 채널로 묶인 영상으로부터 메탈 아티팩트가 감소된 CNN 생성 이미지(CNN Generated Image)가 생성된다.
- CNN Generated Image 로부터 tissue processing을 하여 CNN Prior 영상을 생성하고, 전방 투시법(Forward Projection)을 통해 1차 사이노그램(Prior Sinogram)이 생성된다.

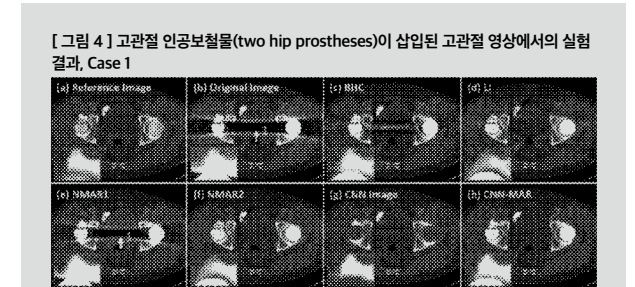
사이노그램이란?
CT는 방사선(x-ray)을 인체에 투과시켜 얻은 정보들로 영상을 재구성하여 만들어진다. x-ray가 어떤 물체를 통과하고 난 이후의 세기를 측정하면, p(s, φ)를 획득할 수 있게 되는데, 이를 조사각도 φ에서의 투사된 영상(projection)이라고 한다. 그리고 x-ray를 여러 각도에서 조사하여 투사된 영상을 얻으면, 각도에 따라 p(s, φ)도 달라진다.¹⁰ 이렇게 획득한 p(s, φ)를 s와 φ를 양 축으로 한 2차원 평면에 나타낸 것이 사이노그램이다.¹¹

- 맨 왼쪽의 Original (Metal-inserted) Image 로부터 금속 부분만 분리하여 금속부분만 나타난 이미지(Metal Only Image)를 생성하고 전방투시법을 하면 금속 모양(Metal Trace) 만 있는 사이노그램이 생성된다.
- Metal Trace Sinogram 과 Prior Sinogram을 이용해 Original Sinogram 으로부터 메탈의 흔적을 제거하면(Metal Trace Replacement)를 하면 보정된 사이노그램(Corrected Sinogram)이 된다.
- Corrected Sinogram을 FBP (Filtered Back Projection: Sinogram을 영상으로 변환해주는 과정인 FBP(Filtered Back Projection))를 하면 최종적으로 CNN-MAR 이미지가 생성되게 된다.

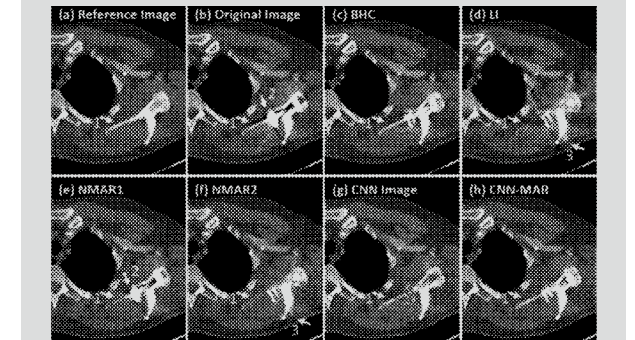


(4) Numerical Simulation을 통해 생성된 영상(Metal Artifact Database)에 대한 CNN-MAR 성능 평가

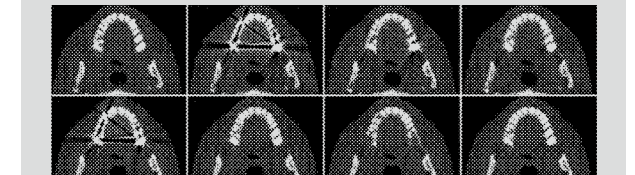
학습에 사용되지 않은 세 가지 케이스를 통해 제안된 CNN-MAR 방법의 성능을 평가해보았다.



[그림 5] 여러 보철물(two fixation screws and a round metal inserted in a bone) 이 삽입된 어깨 부위 영상에서의 실험 결과, Case 2



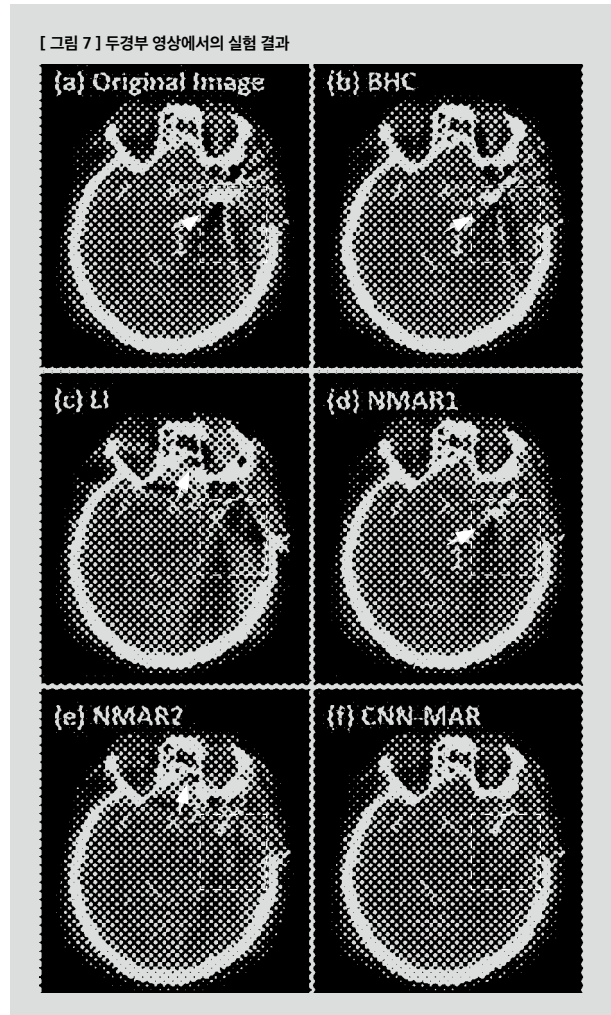
[그림 6] 충치 치료 시 사용하는 치과용 충전제(several dental fillings)가 삽입된 치아 부위 영상에서의 실험 결과, Case 3



본 논문에서 제안한 CNN-MAR 방법을 비롯하여 기존에 사용되던 MAR 방법인 BHC, LI, NMAR (Normalized Metal Artifact Reduction) 방법을 통해 생성된 영상들과 레퍼런스 영상을 비교하기 위해서, 평균 제곱근오차(Root Mean Squared Error, RMSE)와 구조유사도(Structural Similarity, SSIM) 인덱스(index)를 사용해 정량적으로 평가하였다.

5) 실제 환자 CT (Real Data) 에 대한 CNN-MAR 성능 평가

CNN-MAR 방법의 성능을 평가하기 위해 Numerical Simulation을 통해 생성된 영상이 아닌 실제 환자의 영상도 평가에 사용되었다.



환자는 머리에 외과용 클립(surgical clip) 이 삽입되어 있으며 [그림 7]을 보면, 촬영에는 지멘스(Siemens)의 소마톰(SOMATOM Sensation 16 CT scanner)이 사용되었다.(120 kVp, 496 mAs) 촬영을 통해 1160개의 투사된 영상(projection view)을 얻었으며, 외과용 클립으로 인해 발생한 메탈 아티팩트를 CNN-MAR 및 다른 방법으로 개선한 뒤 그 성능을 비교하였다.

실험 결과

1) Numerical Simulation을 통해 생성한 영상에 대한 결과

[그림 4], [그림 5], [그림 6]을 보면 세 케이스 모두 CNN-MAR 방법을 적용했을 때, 피부 조직 부분을 잘 보존하면서도 가장 레퍼런스 이미지(reference image)에 가까운 복원능력을 보여주었다.

[표 1] 수치 시뮬레이션 연구결과, 복원된 영상들의 평균 제곱근오차(RMSE). (단위 : HU)

	Original	BHC	LI	NMAR1	NMAR2	CNN	CNN-MAR
Case 1	155.0	86.3	46.2	121.2	35.4	33.1	29.1
Case 2	71.5	44.4	54.5	50.4	41.4	31.5	22.8
Case 3	320.3	183.5	107.3	234.9	82.3	83.4	58.4

[표 1]은 각각의 MAR 방법으로 복원된 영상들과 레퍼런스 이미지와의 RMSE(평균 제곱근오차)를 계산한 결과이다. CNN-MAR 방법을 통해 복원된 영상이 모든 케이스에서 가장 작은 RMSE 값을 가짐을 알 수 있다.

[표 2] 수치 시뮬레이션 연구 결과, 복원된 영상들의 구조 유사도(SSIM)

	Original	BHC	LI	NMAR1	NMAR2	CNN	CNN-MAR
Case 1	0.565	0.576	0.930	0.887	0.935	0.940	0.943
Case 2	0.883	0.854	0.931	0.955	0.950	0.965	0.977
Case 3	0.522	0.536	0.886	0.833	0.942	0.932	0.967

[표 2]는 각각의 MAR 방법으로 복원된 영상들의 구조 유사도를 비교한 결과이다. 구조 유사도 역시 CNN-MAR 방법을 통해 복원된 영상에서 가장 높은 결과를 보였다.

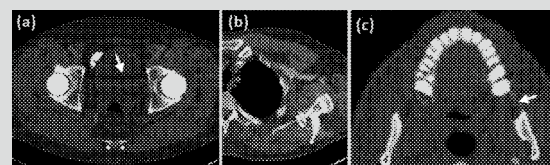
2) 임상 적용(clinical application)

[그림 7]을 보면, 실제 환자 CT 영상에 대해서도 CNN-MAR 기법이 가장 우수한 성능을 보인 것을 확인할 수 있다. CNN-MAR 방법을 적용할 경우, 메탈 아티팩트가 잘 제거되었고 (화살표 1로 표시된 검은 부분들이 잘 복원됨), 뼈 구조(Bony structures) 역시 변형시키지 않았다.

3) Properties of the Proposed CNN-MAR

CNN-MAR 방법의 특성을 알아보기 위해 다양한 실험을 추가적으로 진행하였다.

[그림 8] Tissue Processing 없이 CNN 만으로 실험한 결과



· 우선 tissue processing 의 효과를 확인하기 위해, tissue

processing 없이 CNN 만으로 prior image 를 생성해본 결과, 메탈 아티팩트가 완전히 사라지지 않고 일부 남아있는 모습을 보였다. ([그림 8]의 화살표 참고)

- 따라서 CNN 과 tissue processing 은 서로 상호보완적인 관계로 활용되고 있다는 것을 알 수 있다.
- 또한 CNN의 입력값(input)으로 원본 이미지(Original Image)와 함께 들어가는 기존의 다양한 MAR 적용 영상들에 대해서도 평가해보고자, 입력값(input)을 다양한 채널로 구성하여 추가 실험을 진행했다. (2 채널 : Original + LI / 3 채널 : Original + BHC + LI / 5 채널 : Original + BHC + LI + NMAR1 + NMAR2)

[표 3] 채널별 CNN과 CNN-MAR방식으로 복원된 영상들의 평균 제곱근오차 (RMSE)와 구조 유사도(SSIM)

		Two-channel input image		Three-channel input image		Five-channel input image	
		CNN image	CNN-MAR	CNN image	CNN-MAR	CNN image	CNN-MAR
Case 1	RMSE	40.0	31.0	33.1	29.1	27.3	27.7
	SSIM	0.931	0.938	0.940	0.943	0.942	0.942
Case 2	RMSE	43.4	35.9	31.5	22.8	26.9	22.4
	SSIM	0.956	0.971	0.965	0.977	0.968	0.975
Case 3	RMSE	97.3	66.0	83.4	58.4	79.5	59.3
	SSIM	0.912	0.956	0.932	0.967	0.954	0.971

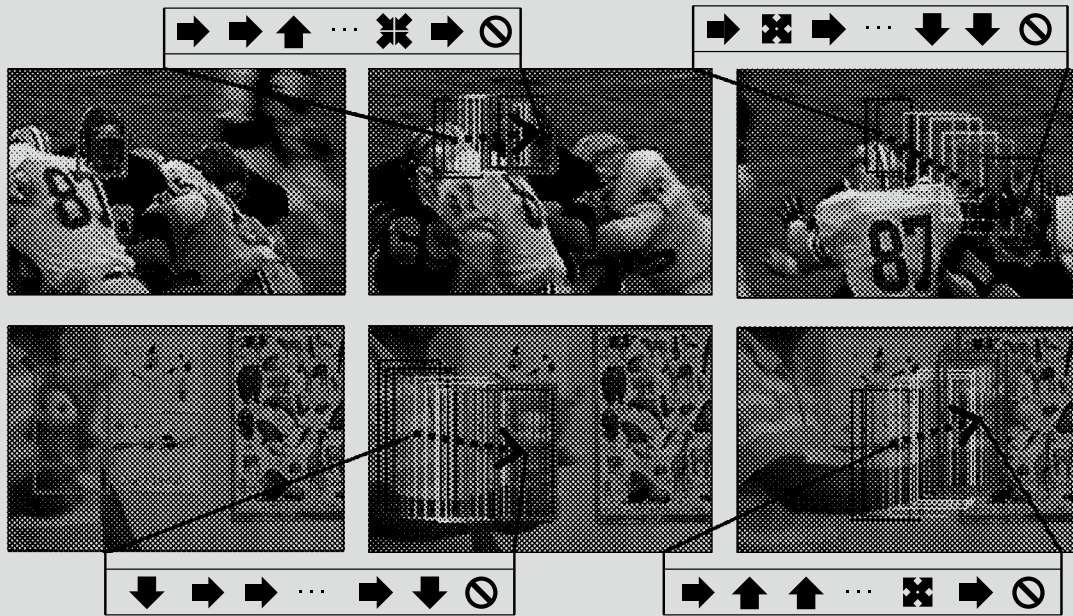
- [표 3]의 결과를 보면 5 채널 의 경우 성능이 가장 좋았으며 [표 3], 이는 CNN이 기존의 다양한 종류의 MAR 방법들을 통합할 수 있는 가능성을 지닌다고 저자들은 설명하였다.

결론

연구 결과, 제안된 CNN-MAR 방법에서 사용된 CNN 과 tissue processing 은 상호 보완적인 관계를 가지며, 함께 사용되어 메탈 아티팩트를 줄이고 주변의 세밀한 구조들을 복원하는 우수한 성능을 보여주었다. (CNN 은 서로 다른 MAR 방법으로부터 유용한 정보를 융합하였으며, tissue processing 은 대부분의 아티팩트를 제거하고 우수한 품질의 prior image를 생성하였다.) 이후 추가 연구를 통해 학습데이터를 늘리고 다양한 MAR 방법을 융합하여, CNN-MAR 방법의 성능을 향상시킬 예정이라고 한다.

공학 깊은 강화학습을 통한 시각적 추적을 위한 행동-결정 네트워크 (Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning)¹²

[그림 1] 연속적 액션을 통한 비주얼 트래킹의 개념



이 논문은 2017년 컴퓨터 비전 및 패턴인식학회(Conference on computer Vision and Pattern Recognition, CVPR)에 발표된 논문으로 컴퓨터 비전 분야에서 중요하게 다루는 객체 추적 문제(object tracking)에 깊은강화학습(deep reinforcement learning)을 다루고 있다. 현재 객체 추적 문제를 해결하는 최신 기술(state-of-the-art)은 합성곱인공신경망(convolutional neural network, CNN)을 기반으로 물체를 탐지(object detection)해서 목표로 하는 객체와 배경을 구별하는 방법을 사용하는 것이다¹³.

이러한 방법들은 다음과 같은 문제가 있었다.

- 객체 위치 후보를 전부 탐색하는 비효율성

- 객체 테두리상자 라벨(label)의 필요성

논문에서는 깊은강화학습을 이용한 ADNet(Action-Decision Network)을 도입하여 문제들을 해결하였다. 객체 추적 문제를 마르코프 결정 과정(Markov Decision Process, MDP)으로 생각하여 이전 프레임에서 객체의 테두리 상자(bounding box)가 액션을 통해 움직여서 현재 프레임의 테두리 상자의 위치와 모양을 결정하도록 했다.

- 상태(state)는 현재 프레임에서 테두리 상자의 위치와 이전

10번의 액션으로 결정됨

- 액션은 상하좌우 움직임, 그 2배의 움직임, 테두리 상자의 크기

확대와 축소, 그리고 멈춤으로 구성됨

- 보상(reward)은 멈춤 액션을 했을 때 테두리 상자가 정답(ground

truth)과 일치하는 정도가 높으면 1 아니면 -1을 가짐. ADNet은 현재의 상태가 입력값으로 들어오면 출력값으로 액션과 신뢰도(confidence)를 도출한다.

[그림 1]과 같이 이전 프레임의 객체 위치에서 시작하여 여러 번의 액션을 통해 현재 프레임에서 객체의 위치를 찾아간다.

ADNet의 학습은 다음과 같은 3단계로 나뉜다.

ADNet의 학습 3단계

지도학습(supervised learning)

- 상태를 입력으로 주고 액션과 신뢰도를 출력으로 하는 VGG-M¹⁴ 기반 CNN을 학습한다.
- 각 프레임의 정답을 기준으로 가우시안 노이즈(Gaussian noise)로 생성한 테두리상자들을 데이터로 사용하였다.
- 생성된 상자들이 정답에 가까워지기 위해 해야 하는 액션을 정하는 것이 네트워크 학습의 목표이다.
- 또한 이 상자의 내용물이 객체인지 아닌지 판별하는 것, 즉 신뢰도를 학습하는 것도 네트워크의 또 다른 목표이다.

깊은강화학습(deep reinforcement learning)

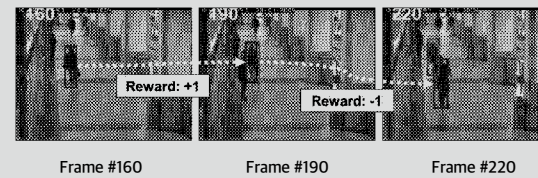
- Policy gradient 방법을 사용하여 네트워크에서 직접 정책(policy)을 얻는다.
- 지도학습에서 미리 학습된 네트워크를 정책 네트워크(policy network)의 시작점으로 활용한다.
- 한 프레임내의 마지막 예측지점이 객체와 맞나 안맞나를 통해 보상을 1 또는 -1로 하고 이를 바탕으로 정책 네트워크를 업데이트한다.

실시간 적응(online adaptation)

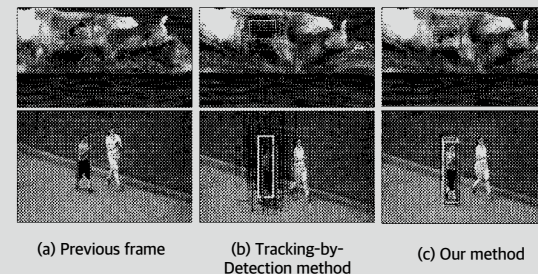
- 프레임에서 예측된 테두리상자를 정답의 기준으로 하는 샘플 데이터를 모아 실시간으로 지도학습을하여 성능을 향상시킨다.
- 신뢰도가 0.5 이하가 되어 객체를 놓쳤다고 판단되면 현재 위치에서 가우시안 노이즈로 생성한 주변 위치 중에서 가장 신뢰도가 높은 지점을 새로 위치로 지정하는 re-detection 과정이 있다.

강화학습을 사용하면 중간중간에 객체 위치 라벨이 없는 데이터에 대해서도 앞뒤 프레임의 정보를 활용하여 보상을 설정해 주는 식으로 학습할 수 있다[그림 2]. 따라서 ADNet은 데이터가 불완전한 준지도학습(semi-supervised learning)에도 대응할 수 있다.

[그림 2] 준지도학습(semi-supervised learning) 사물 추적 시뮬레이션



[그림 3] 탐색 전략 비교



[그림3]에서 ADNet을 활용한 객체 추적은 훨씬 적은 양의 탐색이 필요한 것을 알 수 있다. 이를 통해 최첨단기술 대비 약 3배 빠른 속도 향상을 보여 주었다.

- 최첨단기술인 (b)의 경우 물체 탐지(object detection)를 기반으로 하기 때문에 주변의 여러 후보 영역들을 탐색해야 한다. 현재 객체 추적 문제를 해결하는 CNN기반의 방법의 경우 프레임당 256개의 영역이다.

- ADNet의 경우 액션을 통해 이동하는 영역만 탐색하면 된다.

평균적으로 프레임당 28.26개의 영역이다.

[표 1] 사용자 반응 수집 지연에 따른 상태 CTR 변화

	Algorithm	Prec.(20px)	IOU(AUC)	FPS	GPU
Non-real-time	ADNet	88.0%	0.646	2.9	O
	ADNet-fast	85.1%	0.635	15.0	O
	MDNet [24]	90.9%	0.678	< 1	O
	C-COT [9]	90.3%	0.673	< 1	O
	DeepSRDCF [8]	85.1%	0.635	< 1	O
Real-time	HDT [25]	84.8%	0.564	5.8	O
	MUSTer [15]	76.7%	0.528	3.9	X
	MEEM [42]	77.1%	0.528	19.5	X
	SCT [5]	76.8%	0.533	40.0	X
	KCF [13]	69.7%	0.479	223	X
	DSST [7]	69.3%	0.520	25.4	X
	GOTURN [12]	56.5%	0.425	125	O

위 표는 ADNet의 결과와 다른 알고리즘들을 비교한 것이다. Prec.

(20px)는 정답과 예측한 테두리 상자의 중심간 거리가 20픽셀

이하인 비율이고 IOU(AUC)는 두 상자의 겹치는 비율을 뜻한다.

ADNet은 비실시간 최첨단 기술에 근접한 성능을 내면서도 속도는 더욱 빠른 결과를 보인다. 또한 실시간 적용 단계에서 학습에 필요한 샘플 데이터 수를 비교적 적게 뽑는 ADNet-fast의 경우 FPS 15.0으로 실시간으로 작동하는 알고리즘 중 최고 성능을 보인다.

¹¹ 출처 | See, A. (2017). Get to the point : Summarization with pointer-generator networks, doi : arXiv:1704.04368. ¹² 참고 | sequence-to-sequence attentional model ¹³ 참고 | abstractive text summarization using sequence-to-sequence RNNs and beyond ¹⁴ 참고 | 부정확한 재현성과 반복적 단어의 재활용 등이 문제점으로 지적된다. ¹⁵ 참고 | Vinyas, O., Fortunato, M., Jaitly, N. (2017). Pointer networks, doi : arXiv:1506.03134 ¹⁶ 논문 | Nallapati, R., Zhou, B., & Santos C. (2016). Abstractive text summarization using sequence-to-sequence RNNs and Beyond, doi : arXiv:1602.06023 ¹⁷ 논문 | Zhang, Y. & Yu, H. (2017). Convolutional neural network based metal artifact reduction in X-ray computed tomography, arXiv:1709.01581. ¹⁸ 참고 | CT 영상은 몸에 방사선(x-ray)을 조사하여 획득하게 되는데, 체내에 금속성 물질(예: 금니, 인공 관절, 인공심장박동기 등)을 삽입한 환자가 CT를 찍게 되면 Beam Hardening Effect, 산란(scatter), 푸아송 노이즈(Poisson noise), 환자의 움직임(motion) 등의 다양한 원인에 의해 영상에 인공음영이 생성된다. 이렇게 몸에 삽입된 금속성 물질로 인해 발생하는 인공음영을 메탈 아티팩트라고 일컫는다. 메탈 아티팩트를 감소시킬 목적으로 지멘스(SIEMENS)의 iMAR#, 필립스(PHILIPS)의 OMAR, GE헬스케어(GE Healthcare)의 Smart MAR 등의 상용화된 MAR 소프트웨어가 출시되었지만, 아직까지도 촬영 부위에 따라 효과적이지 못한 경우가 많이 발생하고 있다. 따라서 이 문제를 해결하고 영상의 품질을 개선하려는 연구들이 현재까지 계속되고 있다. 최근 답리닝을 활용한 영상연구가 괄목하게 발전하면서, 메탈 아티팩트 문제도 답리닝을 접목해서 해결하려는 연구가 활발하게 전개되고 있다. ¹⁹ 참고 | 같은 논문의 III. B. Tissue Processing 에 자세히 설명되어 있습니다. ²⁰ 참고 | http://terms.naver.com/entry.nhn?docId=3572221&cid=58944&categoryId=58970 ²¹ 참고 | https://patentimages.storage.googleapis.com/US8509504B2/US08509504-20130813-D00000.png ²² 참고 | http://openaccess.thecvf.com/content_cvpr_2017/papers/Yun_Action-Decision_Networks_for_CVPR_2017_paper.pdf ²³ 논문 | Nam, H. & Han, B. (2016). Learning multi-domain convolutional neural networks for visual tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. ²⁴ 논문 | Chatfield, Ken, et al. (2014). Return of the devil in the details : Delving deep into convolutional nets, doi : arXiv:1405.3531.