

# 구조적 유사도 기반 메모리 증강 오토인코더를 사용한 이상 탐지 기법 연구

박성준, 김대호, 김재일  
경북대학교

sjp611@gmail.com, kdaeho27@gmail.com, threeyears@gmail.com

## A Study on Anomaly Detection using Memory Augmented Autoencoders with Structural Similarity

Sungjoon Park, Daeho Kim, Jaeil Kim  
Kyungpook National University

### 요약

이상 영역 탐지를 위한 오토인코더 기반의 딥러닝 기법들이 제안되었다. 그 중 메모리 증강 오토인코더는 정상 데이터의 영상 특징을 기억하는 메모리 모듈을 사용하여, 비정상 데이터의 이상 영역을 이전에 기억해 둔 정상 패턴의 영상으로 복원함으로써 이상 영역을 탐지할 수 있도록 제안되었다. 본 논문은 구조적 유사도 기반의 메모리 증강 오토인코더를 사용한 향상된 이상 영역 탐지 기술을 소개한다. 제안 기법을 뇌 종양 및 정상인 뇌 MR 영상 데이터에 적용하여, 종양 위치 탐지에 대한 성능 평가를 수행하였으며, 기존 오토인코더와 변분 오토인코더 등과 비교해 제안 기술의 성능 향상을 확인하였다.

### I. 서론

최근 의료 영상을 사용한 지도 학습 기반의 머신 러닝 알고리즘은 질병 분할, 탐지 분야에서 전문의의 진단 수준에 견줄 정도로 발전하였다. 지도 학습 기반의 머신 러닝 기술은 뛰어난 성능에도 불구하고, 레이블링을 위해 전문의의 많은 시간과 노동이 요구되고 희귀 질병에 대한 데이터의 불균형으로 인해 개발에 어려움을 겪고 있다. 특히, 질환에 의한 병변의 특징은 매우 다양하므로 이를 포괄할 수 있는 데이터 셋을 얻는 것과 레이블링은 어려운 작업이다. 이에 따라 비지도 학습 기반 이상 탐지 기법이 많은 관심을 받고 있다.

의료 영상에서 이상 영역 탐지를 수행하기 위해, 딥러닝 모델은 질병이 없는 정상 영상 데이터를 학습하여 정상 매니폴드를 모델링한다. 병변을 포함한 환자 영상이 주어지면, 학습된 정상 매니폴드로부터 이상 영역의 복원 성능 하락에 따라 이상 영역이 탐지된다.

딥 오토인코더는 정상 매니폴드 학습을 위한 대표적인 딥러닝 기법이다. 최근 오토인코더를 기반으로 한 변형 모델이 제안되었다. Bergmann은 오토인코더의 손실 함수로 구조적 유사도(Structural Similarity, SSIM)를 사용하여 이상 탐지 성능을 향상시킨 결과를 소개하였고[1], Gong은 오토인코더에서 비정상 패턴의 복원을 제한하기 위해 메모리 모듈을 적용한 메모리 증강 오토인코더(Memory Augmented Autoencoder, MemAE)를 제안하였다[2].

본 논문은 MemAE의 성능 향상을 위해 MemAE의 복원 함수로 SSIM을 사용하여 복원 영상의 품질을 향상시킬 수 있는 학습 방법을 제안한다. MemAE의 메모리 모듈은 복원 손실 함수에 기반하여 정상 패턴을 저장하는데, SSIM은 희도, 대비 및 구조를 고려한 정상 패턴을 기록하도록 하고, 따라서 보다 정상 형태의 선명한 복원이 가능하다. 제안 기법의 성능 평가를 위해

뇌 종양 데이터 셋인 BRATS-2019[3]에서 이상 탐지를 실험하며 MemAE를 포함한 알려진 오토인코더 기반의 이상 탐지 기법을 정성적, 정량적으로 평가한다.

### II. 본론

#### 모델 구조

MemAE는 3 가지 모듈(인코더, 메모리 모듈, 디코더)로 구성된다.

**인코더, 디코더:** 인코더는 입력 데이터 샘플  $x$ 를 메모리 모듈의 가장 유사한 항목을 검색하기 위해 사용되는 쿼리 벡터  $q$ 로 변환하고, 디코더는 메모리 모듈로부터 정상 쿼리 벡터  $\hat{q}$ 를 입력 받아 입력 데이터 샘플  $\hat{x}$ 으로 복원한다:

$q \sim Encoder(x; \phi)$ ,  $\hat{x} \sim Decoder(q; \theta)$ , 여기서  $\phi$ 와  $\theta$ 는 각각 인코더, 디코더의 가중치이다.

**메모리 모듈:** 메모리 모듈은 가중치 행렬  $M \in R^{N \times C}$ 로 정의된다. 여기서  $N$ 은 메모리 모듈의 크기를 나타내는 초모수이고,  $C$ 는 입력 데이터  $x$ 의 채널 크기이다. 메모리 모듈은 인코더로부터 쿼리 벡터  $q$ 를 입력 받아 정상 쿼리 벡터  $\hat{q}$ 를 출력한다. 인코더로부터 쿼리 벡터  $q$ 가 주어지면 메모리 모듈은 각 메모리 항목  $m_i$ 과 유사도를 계산한다. 유사도 함수  $d(\cdot)$ 는 코사인 유사도 함수가 사용된다. 출력 할 메모리 항목을 결정하기 위해 각 메모리 항목 사이에 주의(attention) 기반의 주소 벡터  $w_i$ 를 계산한다:

$$w_i = \frac{\exp(d(q, m_i))}{\sum_{j=1}^N \exp(d(q, m_j))}$$

계산된 주소 벡터로부터 메모리 모듈에서 새로운 정상 쿼리 벡터  $\hat{q}$ 을 출력한다.

$$\hat{q} = wM = \sum_{i=1}^N w_i m_i$$

MemAE의 복원 함수로 사용되는 평균 제곱근 오차를 SSIM으로 대체하여 메모리 모듈에 단순한 정상 영상 차이의 특징 대신 정상 흐도, 대비, 구조를 고려한 특징을 기록하도록 한다.

쿼리 벡터가 다양한 메모리 항목을 참조하기 위해, 주소 벡터의 엔트로피  $E(\cdot)$ 를 최소화하여 주소 벡터의 희소성이 항상되도록 유도한다. 제안된 손실 함수는 다음과 같이 정의된다:

$$L(\phi, M, \theta) = \frac{1}{T} \sum_{t=1}^T ((SSIM(x^t, \hat{x}^t) + \alpha E(w^t)).$$

#### 데이터 셋 전처리 및 모델 학습 방법

본 연구는 뇌 MR 영상 공공 데이터 셋 CamCAN[5]과 BRATS-2019[4]의 coreg-t2-weighted 뇌 영상을 각각 훈련, 검증 및 테스트를 위해 사용하였다. CamCAN은 653 개의 질병을 갖지 않는 건강한 뇌 영상으로 구성되어 있다. 이 중 무작위로 선택된 600 개의 뇌 영상이 훈련 데이터로 사용되었고 나머지 53 개는 최적화된 모델 선택을 위한 검증 데이터로 사용되었다. BRATS-2019 데이터 셋은 335 개의 뇌 종양 영상으로 구성되며 모두 테스트 데이터로 사용되었다. 두 데이터 셋은 모두  $240 \times 240 \times 155$ 의 크기를 갖는다. 두 데이터 셋 사이의 모달리티를 동등하게 유지하기 위해 두개골 검출, 강체 변환, 최소-최대 정규화를 차례로 수행하였다.

실험 모델은 모두 패치 기반으로 학습되었다. 각 데이터 샘플은 z 축으로 보폭 1 쪽 이동하며  $240 \times 240 \times 3$  크기의 패치 추출 작업이 수행되었다. 모델은 패치의 중앙 슬라이스를 예측하도록 수행된다. 모든 모델은 배치 32, 학습률 0.0001의 Adam 옵티마이저로 150 에폭만큼 훈련되었다. 메모리 크기는 10으로 고정되어 실험하였다.

#### 실험 및 결과

이상 영역을 나타내는 이상 영상을 얻기 위해 다음의 단계를 진행한다. 1) 입력 영상과 복원 영상 사이의 차이로 이상 영상을 계산한다. 2) 모델 간 동등한 평가를 위해 이상 영상의 값을 [0,1] 사이로 정규화한다. 3) 이상 영상에 대해 구체적인 임계 값을 설정하여 임계 값을 이상을 이상 영역으로 구분하여 이진 영상을 얻는다.

		AE	VAE	SSIM-AE	MemAE	제안
임계 값	TPR	0.583	0.692	0.686	0.683	<b>0.718</b>
	TNR	0.953	0.947	0.950	0.945	0.941
임계 값	TPR	0.271	0.417	0.382	0.408	<b>0.432</b>
	TNR	0.982	0.978	0.981	0.977	0.973

표 1. 오토인코더 기반 모델 간 임계 값에 따른 TNR, TPR 결과.

본 실험에서 모델 별 TPR 과 TNR 을 계산하여 이상 영역이 종양을 포함하는지에 대한 정량적 평가를 수행한다. TPR 은 이상 영상이 종양을 얼마나 포함하는 수치로, 표 1 은 제안한 기법이 비교 모델보다 임계 값 기준과 무관하게 높은 TPR 을 나타냄을 보여준다.

그림 1 은 비교 모델들의 복원 결과와 이상 영역 탐지 결과를 나타낸다. 메모리 모듈을 사용하지 않은 AE 와 SSIM-AE 는 종양을 포함하여 영상 전역적으로 큰 오차를 나타낸다. VAE 는 입력 영상을 항상 사전 분포에서 샘플링하여 정상 영상을 복원하므로 종양 영역은 이상 영역으로 잘 탐지하지만, 복잡한 뇌의 겉 영역을 흐릿하게 복원하는 경향을 보인다. MemAE 또한 종양을 생성하지는 않지만 뇌의 구조를 고려하여 학습되지 않으므로 입력과 일치하지 않는 뇌실을

생성하는 경향을 보인다. 제안 기법은 정상 구조를 생성하여 이상 영상이 종양 영역을 포함할 뿐 아니라 완벽한 정상 영상 생성하여, 시각적으로 종양 위치를 쉽게 구별 가능하다.

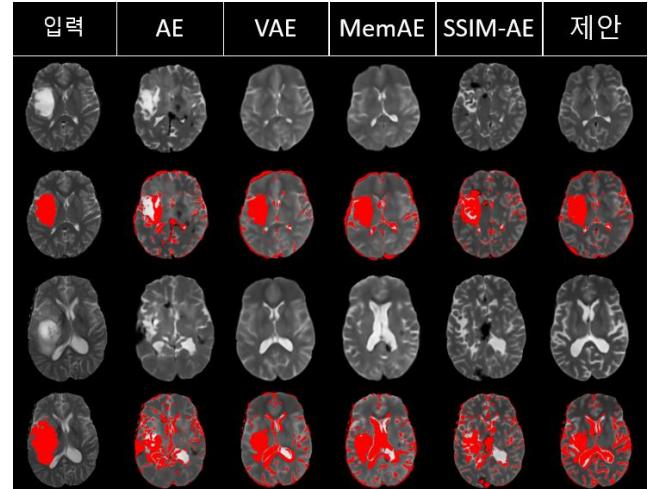


그림 1. 오토인코더 기반 모델 간 결과 비교 (테스트 데이터 셋) (홀수 행), 입력 영상 및 복원 영상 (짝수 행), 이상 영역 탐지 결과 (적색).

### III. 결론

본 논문은 메모리 모듈에 SSIM 을 결합한 향상된 오토인코더 기술을 제안했다. 뇌 MR 공공 데이터 셋에 대해 오토인코더 기반의 모델들과 비교하여 향상된 이상 탐지 성능을 얻었다.

### 참고 문헌

- [1] Bergmann, Paul, et al. "Improving unsupervised defect segmentation by applying structural similarity to autoencoders." *arXiv preprint arXiv:1807.02011* (2018).
- [2] Gong, Dong, et al. "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.
- [3] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing* 13.4 (2004): 600-612.
- [4] Menze, Bjoern H., et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." *IEEE transactions on medical imaging* 34.10 (2014): 1993-2024.
- [5] Taylor, Jason R., et al. "The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample." *Neuroimage* 144 (2017): 262-269.