

Using a generative adversarialnetwork to generate synthetic MRIimages for multi-class automaticsegmentation of brain tumors

≡ 분류	뇌데이터처리 모델
≡ 링크	https://www.frontiersin.org/journals/radiology/articles/10.3389/fradi.2023.1336902/full
≡ Summary	GAN 기반의 Pix2PixNifTI를 통해 합성 이미지를 생성하여 학습을 했다는 점, 4개의 채널로 이루어진 이미지 채널의 3가지씩의 조합이 눈여겨볼점
👤 발표자	유림 유림 오

1 Introduction

- 영상의 중요성:** MRI는 교모세포종의 진행 상황을 추적하는 데 선호되는 방법으로, T1 가중(T1w), T2 가중(T2w), 조영증강 T1 가중(T1CE), FLAIR(액체 억제 회복)과 같은 다양한 영상 시퀀스를 사용합니다. 이러한 시퀀스는 종양의 생물학적 상태를 평가하고 종양 미세 환경의 미세한 변화를 감지하는 데 도움을 줍니다.
- 수동 분할의 도전 과제:** 뇌 병변의 수동 분할은 노동 집약적이고 시간이 많이 소요되며, 수행하는 개인의 전문성에 따라 변동성이 클 수 있습니다. 이로 인해 일관성이 떨어지고 편향이 발생할 수 있어 임상에서 신뢰성이 낮아 집니다.
- 자동화의 필요성:** 수동 방법의 한계는 정확하고 재현 가능한 결과를 제공할 수 있는 자동화된 분할 도구의 필요성을 강조합니다. 딥러닝(DL) 기술, 특히 3차원 합성곱 신경망(3D CNN)이 이 과정을 자동화하는 해결책으로 제안됩니다.
- 다양한 데이터셋으로 훈련:** 3D CNN이 정확한 결과를 내기 위해서는 대규모의 다양한 데이터셋으로 훈련되어야 합니다. 이러한 훈련을 통해 모델은 다양한 영상 모달리티에 적응하고 수동 분석보다 성능을 향상시킬 수 있습니다.
- 임상적 관련성과 입력 요구 사항:** 이러한 알고리즘의 효과는 임상 영상 관행의 변동성에 의해 도전받습니다. 모든 필요한 이미지가 분석에 사용되지 않을 수 있기 때문에, 이는 실제 환경에서 AI 모델의 적용을 방해할 수 있습니다.
- 누락된 데이터 해결:** 연구에서는 생성적 적대 신경망(GAN)을 사용하여 특정 입력 이미지가 누락된 경우 이를 보완할 수 있는 합성 이미지를 생성하는 방법을 제안합니다. 이 접근 방식은 누락된 데이터의 근사치를 제공하여 분할 모델의 강건성을 향상시키는 것을 목표로 합니다.
- 연구 목표:** 이 연구의 주요 목표는 일부 입력 채널이 누락된 경우에도 뇌종양의 정확한 다지역 분할을 위한 CNN 기반 모델을 개발하는 것입니다. 두 번째 목표는 기존 모델을 지원하기 위해 합성 데이터를 사용하는 가능성을 평가하여 분할이 신뢰할 수 있고 임상적으로 적용 가능하도록 하는 것입니다.
- 가설:** 연구자들은 GAN에 의해 생성된 합성 데이터를 원본 이미지와 결합하면 분할 모델의 성능을 효과적으로 향상시킬 수 있으며, 이를 통해 뇌종양의 정확한 다중 레이블 분할이 가능하다고 가설을 세우고 있습니다.

요약하자면, 소개 부분은 교모세포종의 진단 및 치료를 개선하기 위해 고급 영상 기술과 자동화된 알고리즘의 필요성을 강조하며, 수동 분할의 도전 과제와 임상 영상 관행의 변동성 문제를 다루고 있습니다.

2.1 Materials

이 연구에서는 뇌종양 분할(BraTS) 2021 데이터셋을 사용하여 알고리즘의 훈련, 검증 및 테스트를 수행했습니다. 이 데이터셋은 복미 방사선학회, 미국 신경방사선학회, 의료 영상 컴퓨팅 및 컴퓨터 보조 개입 학회가 공동으로 공개한 자료로, 1,251명의 환자에 대한 다기관 수술 전 기초 다중 매개변수 MRI(mpMRI) 스캔을 포함하고 있습니다. 이 스캔은 T1w, T2w, T1CE, FLAIR와 같은 3D 시퀀스를 포함하며, Neuroimaging Informatics Technology Initiative(NIfTI) 형식으로 제공됩니다.

각 환자에 대해 부종(ED), 괴사 종양(NC), 조영증강 종양(CET)으로 구성된 다중 레이블 분할이 제공되어 자동 분석의 기준 진실(ground truth, GT)로 사용되었습니다. 모든 이미지는 강직하게 정렬되어 있으며, 1 mm^3 의 등방성 해상도로 재샘플링되고 두개골이 제거된 상태로 제공됩니다. 모델 훈련 전에 이미지는 여러 전처리 단계를 거쳤습니다. 여기에는 관심 영역(ROI) 마스크 및 전체 종양(WT) 마스크 생성, GT 마스크의 하위 영역을 자동 알고리즘에서 기대하는 표준화된 복셀 값으로 재레이블링, 모든 입력 모달리티, GT 및 ROI 마스크를 동일한 해부학적 표현과 동일한 이미지 크기($240 \times 240 \times 155$)로 공동 등록, 그리고 평균 0 및 분산 1로 표준화하는 과정이 포함됩니다. 이후 데이터셋은 무작위로 1,000명과 251명의 환자로 나누어 알고리즘의 훈련 및 검증, 테스트에 사용되었습니다.

이 연구에서는 mpMRI BraTS2021 데이터셋을 두 가지 딥러닝(DL) 기반 방법에 적용했습니다. 첫 번째는 뇌 병변의 다중 클래스 자동 분할을 위한 것이고, 두 번째는 이미지 간 변환을 위한 것입니다. 다중 레이블 분할의 기준으로 DeepMedic 알고리즘이 사용되었고, 이미지 간 변환을 위해 Pix2PixNIfTI 도구가 구현되었습니다. Pix2PixNIfTI는 조건부 GAN 아키텍처의 3D 단일 입력 단일 출력 변형으로, 두 MRI 시퀀스 간의 매핑을 학습하여 목표 이미지의 근사 맵을 생성합니다. 1,000명의 환자 하위 집합은 각각 훈련 및 검증을 위해 800명과 200명으로 나누어졌습니다. 훈련 및 검증 데이터의 분할은 5배 교차 검증을 따랐으며, 이는 데이터셋을 무작위로 80%와 20%로 나누는 과정을 포함합니다. 모든 CNN은 먼저 5배 교차 검증 하위 집합을 사용하여 평가된 후, 251명의 환자 집단을 사용하여 개별적으로 테스트되었습니다. 모든 알고리즘은 이미지를 빠르게 처리할 수 있도록 그래픽 처리 장치가 장착된 리눅스 워크스테이션에서 실행되었습니다.

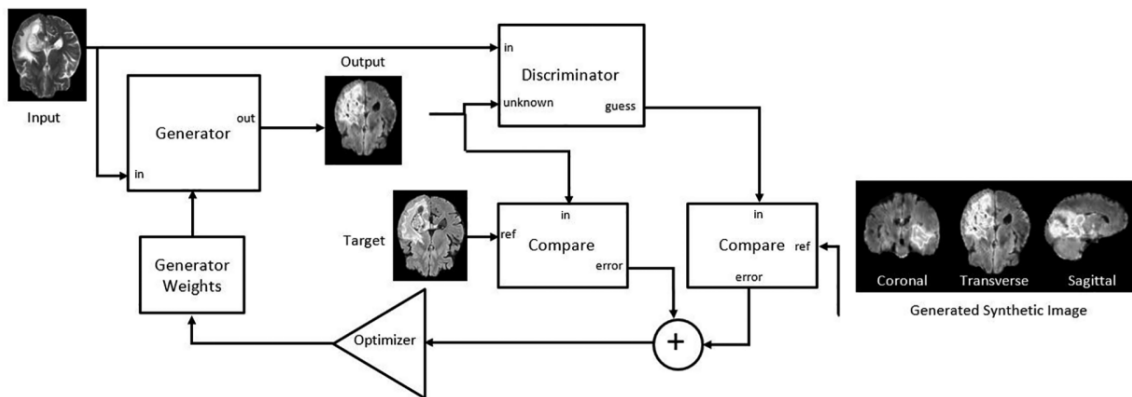
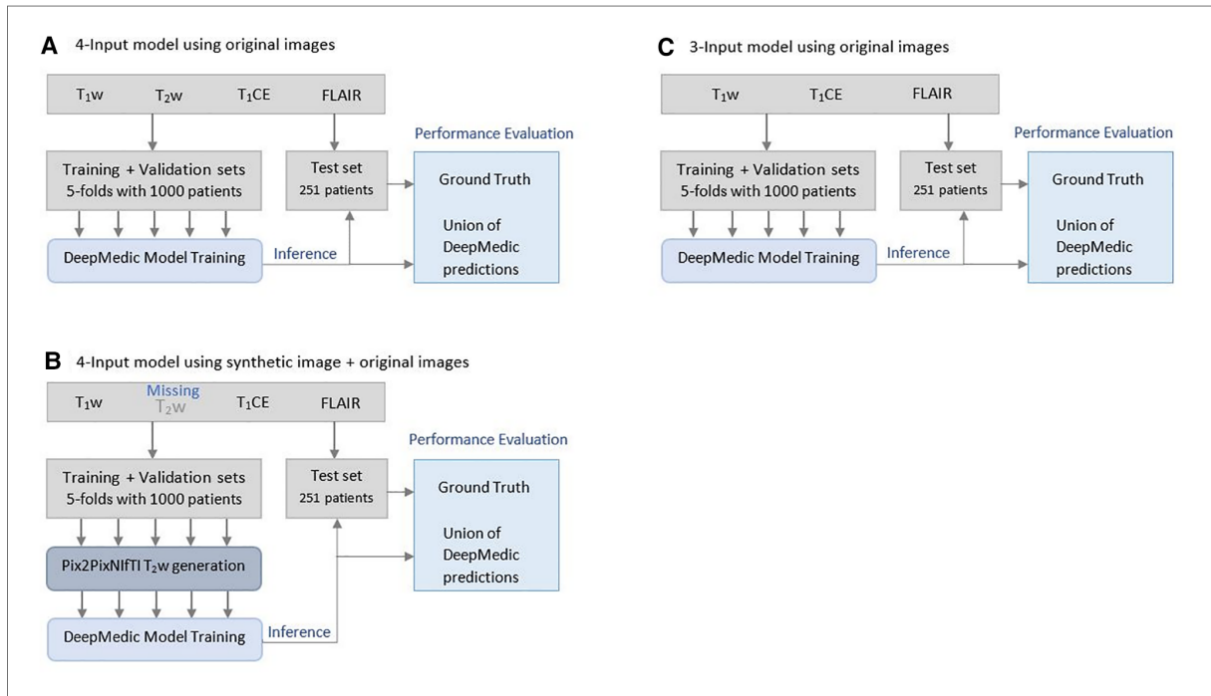


FIGURE 2
Adapted Pix2PixNIfTI architecture of conditional generative adversarial network (cGAN) for the creation of target image (FLAIR in the figure) using reference image (T_2w in the figure) (20).



2.3 Data processing

데이터 처리 섹션에서는 뇌 종양 분할을 위한 DeepMedic 분할 모델과 Pix2PixNifTI 모델의 훈련 방법론을 설명합니다. DeepMedic 모델은 네 가지 입력 이미지(T₁w, T₂w, T₁CE, FLAIR)를 사용하여 배치 크기 10으로 35 에포크 동안 훈련되었으며, 이를 네 입력 모델이라고 합니다. 또한, 세 가지 입력 이미지를 다양한 조합으로 훈련하여 세 입력 모델로도 알려져 있습니다.

누락된 입력 이미지를 해결하기 위해 Pix2PixNifTI 모델을 사용하여 합성 이미지를 생성했습니다. 이 모델은 동일한 데이터셋을 사용하여 5배 교차 검증 방식으로 훈련 및 테스트되었으며, 배치 크기는 1, 총 100 에포크로 설정되었습니다. 생성된 합성 이미지는 DeepMedic 모델에 통합되어 원본 이미지 중 하나를 대체하는 방식으로 사용되었습니다.

제공된 표는 DeepMedic 모델 훈련에 사용된 입력 이미지 조합과 Pix2PixNifTI 모델을 통해 합성 이미지를 생성하는 데 사용된 참조 이미지를 나타냅니다. 생성된 이미지는 누락된 원본 이미지를 보완하여 다중 클래스 자동 분할 성능을 향상시키기 위한 목적을 가지고 있습니다 [T4], [T6].

2.4 Statistical analysis

통계 분석은 세 가지 단계로 진행되었으며, 주로 다중 레이블 예측의 정확도, 정밀도 및 민감도를 평가하여 가능한 한 GT(정답)와 가까운 결과를 도출하는 데 중점을 두었습니다. GT와 5배 교차 검증 모델의 자동 분할 간의 유사성과 재현성은 전체 종양(WT) 및 개별 종양 레이블(NC, ED, CET)에 대해 다이스 계수를 사용하여 평가되었습니다.

다이스 계수는 모든 클래스에 대해 개별 환자를 기준으로 GT 분할을 고려하여 평가되었으며, 5개의 교차 검증 폴드에서 생성된 예측 결과를 통합하여 나중에 모든 환자에 대해 평균화되었습니다. 다이스 계수 측정 후, 종양 하위 영역 간의 차이는 표준 쌍체 t-검정을 사용하여 통계적 유의성을 테스트했습니다. 비교에는 모든 원본 이미지를 포함한 모델과 원본 이미지 하나를 합성 이미지로 대체한 모델의 자동 생성 분할 결과가 포함되었습니다. 통계적 유의성 기준

은 단위 분할에 대해 $p < 0.05$ 로 설정되었으며, 이후 다중 비교에 대해 보정(Bonferroni correction)이 적용되었습니다.

마지막으로, 평균 제곱 오차(MSE)는 이미지 품질의 정량적 척도로 사용되어 모든 합성 이미지에 대해 평가되었으며, 이는 생성된 이미지의 뇌 볼륨에 대한 원본 이미지와 예측 이미지 간의 차이를 반영하는 평균 오차를 측정하는 데 사용되었습니다.

연구의 논의 섹션에서는 자동 세그멘테이션을 위한 입력 채널 부족 문제를 해결하기 위해 채택한 두 가지 주요 접근 방식에 대해 설명하고 있습니다. 주요 내용은 다음과 같습니다:

1. **주요 접근 방식:** 첫 번째 방법은 GAN(Generative Adversarial Networks)의 변형을 사용하여 부족한 입력 채널을 생성하는 것이었습니다. 특히 3D Pix2PixNifTI 알고리즘을 사용하여, 세그멘테이션 모델에서 특정 입력이 누락된 경우 대체할 수 있는 합성 이미지를 생성하는 것을 목표로 하였습니다.
2. **합성 이미지 평가:** 연구에서는 이러한 합성 이미지의 효과를 시각적 및 정량적 평가를 통해 평가하였습니다. 목표는 특정 입력이 없을 때 세그멘테이션 모델에 가장 적합한 대체 이미지를 식별하는 것이었습니다.
3. **기존 모델과의 비교:** 두 번째 접근 방식은 세 개의 입력 채널로 훈련된 기존의 다중 클래스 세그멘테이션 모델을 사용하는 것이었습니다. 이를 통해 합성 이미지를 사용한 모델의 성능과 원본 이미지 전체 세트를 사용한 모델의 성능을 직접 비교할 수 있었습니다.
4. **결과 및 발견:** 결과적으로 합성 이미지를 사용할 경우 성능이 다소 감소했지만(다이스 점수에서 반영됨), 모델은 여전히 전체 종양 세그멘테이션에 대해 원본 입력 모델과 유사한 성능을 보였습니다. 연구는 DeepMedic 모델이 입력이 부족한 상황에서도 만족스러운 세그멘테이션 결과를 달성할 수 있음을 강조하며, 그 강건성과 유연성을 보여주었습니다.
5. **시각적 평가:** 생성된 합성 이미지에 대한 정성적 평가는 시각적으로 수용 가능하다는 것을 보여주었으며, 이는 특정 이미징 모달리티가 누락된 임상 응용에서 GAN으로 생성된 이미지를 사용할 수 있는 가능성을 뒷받침합니다.

이 섹션은 불완전한 데이터가 있는 상황에서 세그멘테이션 정확도를 개선하기 위한 혁신적인 접근 방식을 강조하며, 제안된 방법의 효과를 전통적인 모델과 비교하여 보여줍니다.

TABLE 6 Mean dice scores with SD evaluated for individual tumor sub-regions and whole tumor for predictions evaluated using a three-input DeepMedic model with multiple contrast combination of three original images and four-input DeepMedic prediction derived using original images for comparison purposes.

Three-input contrast combination for DeepMedic prediction	Dice \pm SD			
	Necrotic tumor core	Edema	Contrast-enhanced tumor	Whole tumor
T ₁ w-T ₁ CE-FLAIR	0.61 \pm 0.24	0.70 \pm 0.12	0.70 \pm 0.19	0.77 \pm 0.07
T ₁ w-T ₁ CE-T ₂ w	0.76 \pm 0.29*	0.81 \pm 0.14*	0.85 \pm 0.21*	0.90 \pm 0.07*
T ₁ CE-T ₂ w-FLAIR	0.77 \pm 0.28*	0.86 \pm 0.12*	0.85 \pm 0.20*	0.93 \pm 0.06*
T ₁ w-T ₂ w-FLAIR	0.57 \pm 0.29	0.79 \pm 0.15	0.62 \pm 0.23	0.92 \pm 0.07
Four-input original image DeepMedic prediction	0.76 \pm 0.29	0.86 \pm 0.13	0.86 \pm 0.20	0.93 \pm 0.06

The dice scores statistically not significant for tumor sub-regions are marked with a "*" sign ($p > 0.0025$).

TABLE 6은 세 가지 원본 이미지를 사용한 세 입력 DeepMedic 모델의 예측 결과에 대한 평균 다이스 점수와 표준 편차(SD)를 보여줍니다. 이 표는 각 종양 하위 영역과 전체 종양에 대한 예측을 평가하며, 원본 이미지를 사용한 네 입력 DeepMedic 예측과 비교합니다.

• 세 입력 대비 조합:

- **T1w-T1CE-FLAIR:** 괴사 종양 중심(Necrotic tumor core) 0.61 \pm 0.24, 부종(Edema) 0.70 \pm 0.12, 조영증강 종양(Contrast-enhanced tumor) 0.70 \pm 0.19, 전체 종양(Whole tumor) 0.77 \pm 0.07
- **T1w-T1CE-T2w:** 괴사 종양 중심 0.76 \pm 0.29*, 부종 0.81 \pm 0.14*, 조영증강 종양 0.85 \pm 0.21*, 전체 종양 0.90 \pm 0.07*
- **T1CE-T2w-FLAIR:** 괴사 종양 중심 0.77 \pm 0.28*, 부종 0.86 \pm 0.12*, 조영증강 종양 0.85 \pm 0.20*, 전체 종양 0.93 \pm 0.06*
- **T1w-T2w-FLAIR:** 괴사 종양 중심 0.57 \pm 0.29, 부종 0.79 \pm 0.15, 조영증강 종양 0.62 \pm 0.23, 전체 종양 0.92 \pm 0.07

- **네 입력 원본 이미지 DeepMedic 예측:** 괴사 종양 중심 0.76 \pm 0.29, 부종 0.86 \pm 0.13, 조영증강 종양 0.86 \pm 0.20, 전체 종양 0.93 \pm 0.06

표에서 통계적으로 유의미하지 않은 종양 하위 영역의 다이스 점수는 "*" 기호로 표시되어 있으며, p 값이 0.0025 보다 큼니다.

이 표는 세 입력 조합이 네 입력 조합에 비해 각 종양 하위 영역의 예측 성능이 어떻게 다른지를 보여줍니다.

TABLE 7 Mean MSE of individual synthetic image calculated for multiple reference images along with evaluated corresponding Pearson and Spearman correlation coefficient of synthetic image MSE and dice score of respective four-input model prediction for multi-class and whole tumor sub-region.

Target image	Reference image	Synthetic image MSE	Pearson coefficient		Spearman coefficient	
			Multi-class	Whole tumor	Multi-class	Whole tumor
FLAIR	T ₁ CE	0.517307	-0.09747	-0.07988*	-0.17191	-0.22446
	T ₁ w	0.494333	-0.10610	-0.08877*	-0.15752	-0.15511
	T ₂ w	0.342234	-0.07842	-0.00643*	-0.15566	-0.20277
T ₁ CE	FLAIR	0.896027	-0.01530*	0.09458*	-0.03816*	0.04452*
	T ₁ w	1.052703	0.01256*	0.13154	0.00432*	0.11043*
	T ₂ w	0.972855	-0.00596*	0.11747*	-0.02017*	0.12646
T ₁ w	FLAIR	0.764378	-0.05202*	-0.08146*	-0.11485	0.00669*
	T ₁ CE	0.628514	-0.02677*	0.02138*	-0.02819*	0.09077*
	T ₂ w	0.845746	-0.05245*	-0.07035*	-0.13704	-0.05480*
T ₂ w	FLAIR	0.454605	-0.08700	-0.02092*	-0.11868	-0.16521
	T ₁ CE	0.383218	-0.11272	-0.04269*	-0.17125	-0.20306
	T ₁ w	0.439892	-0.12217	-0.05132*	-0.15581*	-0.18053

The correlation coefficient statistically not significant for tumor sub-regions is marked with a "*" sign ($p > 0.05$).

TABLE 7은 여러 참조 이미지를 기반으로 계산된 개별 합성 이미지의 평균 평균 제곱 오차(MSE)와 해당 합성 이미지 MSE와 네 입력 모델 예측의 다이스 점수 간의 피어슨 및 스피어만 상관 계수를 보여줍니다. 이 표는 다중 클래스 및 전체 종양 하위 영역에 대한 정보를 포함하고 있습니다.

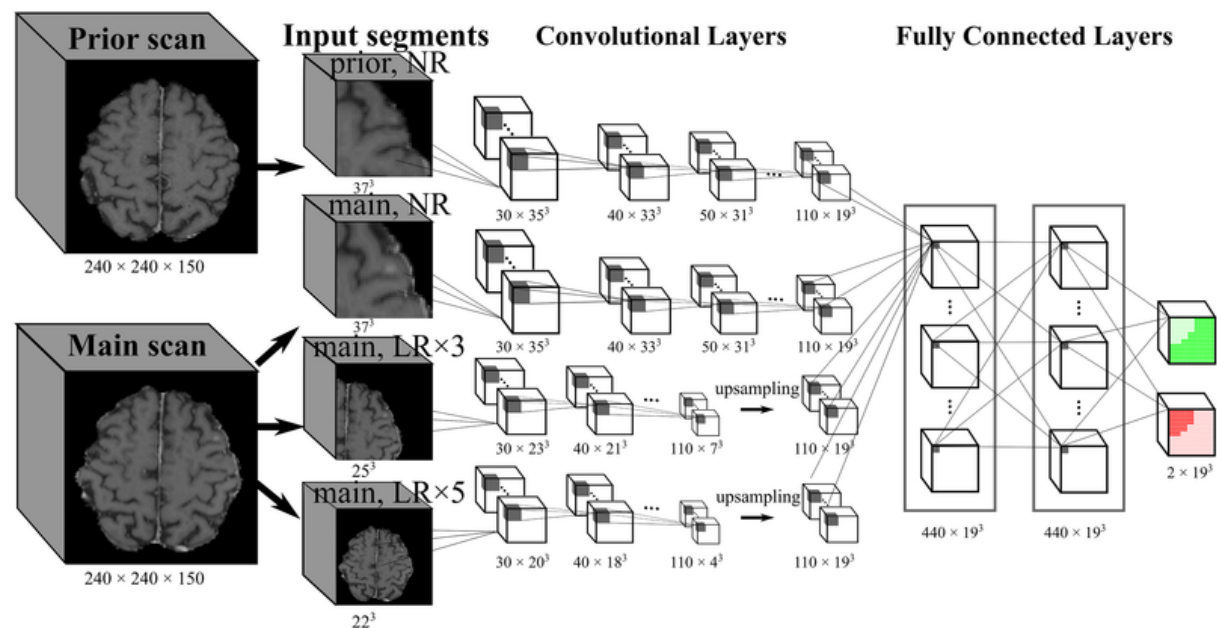
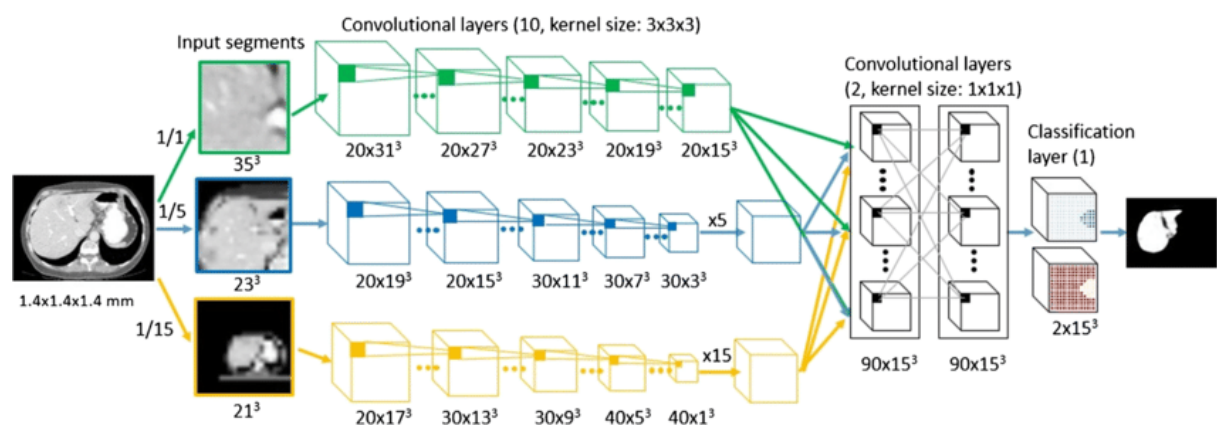
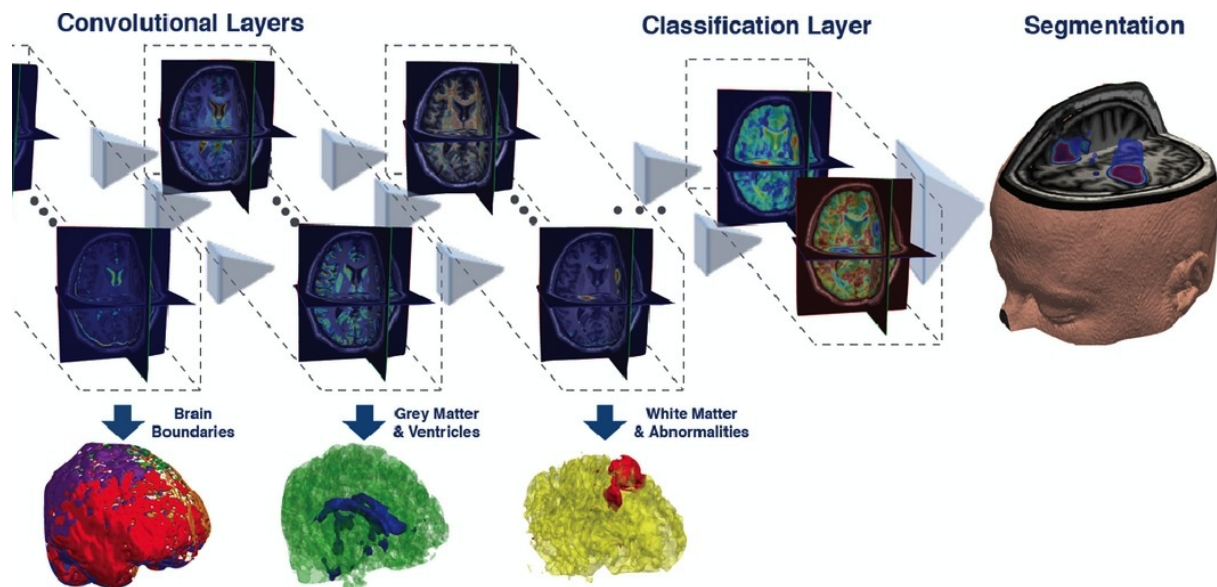
- MSE와 다이스 점수의 관계:** MSE가 낮을수록 다이스 점수가 높아지는 경향이 있습니다. 이는 합성 이미지의 품질이 예측 성능에 긍정적인 영향을 미친다는 것을 의미합니다. 즉, 합성 이미지가 실제와 더 유사할수록 모델의 예측 정확도가 높아진다는 것입니다.
- T2w 이미지의 성능:** T2w 이미지는 상대적으로 낮은 MSE(0.342234)를 보여주며, 이는 이 이미지가 다른 이미지들에 비해 더 나은 품질을 가지고 있음을 시사합니다. 따라서 T2w 이미지를 사용하는 것이 예측 성능을 향상시키는 데 유리할 수 있습니다.
- T1CE FLAIR 조합:** T1CE FLAIR 조합의 MSE는 0.896027로 상대적으로 높지만, 스피어만 상관 계수는 0.04452로 양의 상관관계를 보입니다. 이는 이 조합이 특정 상황에서 유용할 수 있음을 나타냅니다.
- 전반적인 경향:** 전반적으로, MSE가 낮고 다이스 점수가 높은 조합이 더 나은 예측 성능을 나타내며, T2w 이미지가 가장 유리한 것으로 보입니다. 반면, T1w와 T1CE 조합은 상대적으로 높은 MSE를 보여 예측 성능이 떨어질 수 있습니다.

결론적으로, T2w 이미지를 포함한 조합이 예측 성능을 향상시키는 데 가장 효과적일 것으로 판단됩니다.

학습 모델은 뭔데?

DeepMedic

<https://deepmedic.org/>



자세히는 공부하지 않았으니 3D CNN 을 기반으로 하여 전역적인 뇌구조를 이해하는데 집중한 아키텍처로 예상됨

그렇다면 다음은 3D 유닛을 공부해야 하는건가..

