

STN-YOLO: 공간 변환 네트워크를 이용한 견고한 뇌종양 탐지

김정은^o 이효정 김현진 이기정 이지은 박상현[†]

연세대학교 컴퓨터과학과

{wjddms2216, hyojoy, hjkim99, rlwj4177, jieun199624, sanghyun}@yonsei.ac.kr

STN-YOLO: Robust Brain Tumor Detection with Spatial Transformer Network

Jeongeun Kim^o Hyojeong Lee Hyunjin Kim Kijung Lee Jieun Lee Sanghyun Park[†]

Dept. of Computer Science, Yonsei University

요약

딥러닝 기반의 뇌종양 탐지 분야에서는 뇌종양 탐지를 위해 you only look once (YOLO)와 같은 객체 탐지 모델이 널리 사용되고 있다. 그러나, 뇌종양 탐지를 위한 의료 영상 촬영 시, 환자의 움직임 등으로 인해 변형이 일어날 수 있는데 반해, 기존의 객체 탐지 모델들은 이미지의 변형이나 환경 변화에 견고하지 않다. 일반적으로 이런 문제를 해결하기 위해 데이터 증강 기법이 사용되지만, 이는 모델이 변형에 대한 이해 없이 단순히 변형에 대응하도록 학습되기 때문에 새로운 변형에 대한 견고성을 확보하기 어렵다. 따라서 본 연구에서는 YOLO 앞 단계에 Spatial Transformer Network(STN)를 도입함으로써, 모델이 입력 이미지의 공간적 변형을 동적으로 조정하고 적용할 수 있는 STN-YOLO를 제안한다. STN-YOLO는 BraTS2020 데이터 세트에 다양한 변형(Translate, Rotation, Shear-X)을 적용한 이미지에 대해 기존 데이터 증강 기법을 적용한 YOLO 대비 mPC_{0.5:0.95}에서 7.9% 향상된 성능을 보였다. 이는 STN-YOLO가 데이터 증강 기법보다 모델의 견고함을 향상시킨다는 것을 보여준다.

1. 서론

뇌종양 탐지(Brain Tumor Detection)는 딥러닝 모델을 활용하여 뇌 내부의 비정상적인 세포를 식별하고 분류하여 뇌종양을 탐지하는 분야이다. 뇌종양 탐지 기술은 magnetic resonance image(MRI), computed tomography(CT)와 같은 의료 영상에서 신속한 뇌종양 탐지를 가능하게 한다.

뇌종양 탐지를 위해 대표적으로 you only look once(YOLO) [1]와 같은 객체 탐지(object detection) 모델들이 사용된다 [2]. 객체 탐지 모델은 이미지에 존재하는 객체의 종류를 분류하고 객체의 위치를 찾아 bounding box 형태로 나타내는 역할을 한다. 그러나 이러한 객체 탐지 모델들은 다양한 환경 변화나 이미지 내 변경에 대해 충분히 견고(robust)하지 못하다는 문제점이 있다 [3]. 뇌종양 탐지를 위한 의료 영상 촬영 과정에서 크기, 위치, 회전 등의 단순한 변형 이외에도 환자의 자세가 변화에 따른 다양한 변형이 존재하기 때문에 이러한 객체 탐지 모델의 문제점은 잘못된 뇌종양 탐지 결과를 불러올 수 있다. 따라서 뇌종양 탐지를 위한 딥러닝 모델이 이러한 변형들을 정확하게 인식하고 일관되게 처리할 수 있도록 학습하는 것이 중요하다.

일반적으로 딥러닝 모델이 이러한 변형을 학습할 수 있도록 데이터 증강(augmentation) 기법을 사용한다. 데이터 증강 기법은 확률적으로 회전, 좌우반전, 크기 변형과 같은 방식을 입

력 데이터에 적용하여 인위적으로 데이터의 수를 늘리는 기법이다. 데이터 증강 기법은 데이터의 수를 늘려 모델이 다양한 입력을 학습할 수 있지만, 모델이 실제 변형을 이해하기보다 변형에 대응할 수 있도록 학습하기 때문에 새로운 유형의 변형 혹은 복잡한 변형에 대해서는 견고성을 확보하기 어렵다.

이러한 문제점을 해결하기 위해, 본 연구에서는 입력 영상의 다양한 변형에 대한 객체 탐지 모델의 견고성을 향상시키기 위해 객체 탐지 모델인 YOLO의 백본(backbone) 네트워크에 Spatial Transformer Network(STN) [4]를 추가하는 STN-YOLO를 제안한다. STN은 입력 이미지에 대해 스스로 최적의 아핀 변환(affine transform)을 학습한다. 따라서 변형된 이미지가 입력되더라도 이를 동적으로 조정하여 견고한 객체 탐지 결과를 보여준다. 모델의 견고함을 확인하기 위해 BraTS2020 데이터 세트에 다양한 변형을 적용한 결과, STN-YOLO가 데이터 증강 기법만을 적용한 YOLOv5보다 mean performance under corruption(mPC)_{0.5}에 대해 2.6%, mPC_{0.5:0.95}에 대해 7.9% 향상된 결과를 보여준다.

2. 본론

2.1 STN

STN은 딥러닝 모델 내에 삽입 가능한 모듈로서, 입력 데이터

[†] 교신 저자: sanghyun@yonsei.ac.kr

* 이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00229822).

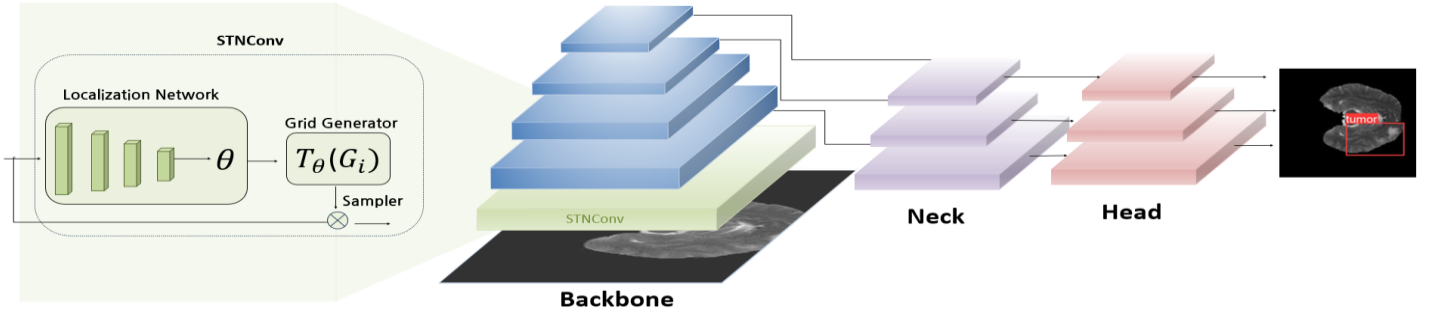


그림 1 STN-YOLO 모델 구조

에 적용되는 공간적 변형을 동적으로 학습하여 모델의 성능을 향상시키는 모듈이다. STN은 Localization Network f_{loc} , Grid Generator G , Sampler f_{samp} 로 구성되어 있다. f_{loc} 는 Convolutional Neural Network(CNN)과 Fully Convolution Layer로 구성되며, 입력 이미지 I 로부터 아핀 변환에 필요한 파라미터 θ 를 학습한다. G 는 f_{loc} 에서 제공한 θ 를 바탕으로, I 에서 샘플링 할 픽셀의 위치 i 를 정해주는 샘플링 그리드 $T_{\theta}(G_i)$ 를 생성하며 수식 (1)과 같다.

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = T_{\theta}(G_i) = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (1)$$

(x_i^s, y_i^s) 는 아핀 변환된 좌표, (x_i^t, y_i^t) 는 입력 이미지 그리드 위의 좌표를 나타낸다. f_{samp} 는 $T_{\theta}(G_i)$ 를 사용하여 I 로부터 새롭게 변형된 출력 이미지 I' 를 생성한다. 이때, 새로운 위치의 픽셀 값은 쌍선형 보간법(bilinear interpolation)을 통해 결정된다.

2.2 STN-YOLO

본 논문에서는 데이터 증강기법을 사용한 YOLO가 다양한 환경 변화나 이미지 내 변형에 대해 충분한 견고성을 보이지 못하는 문제를 개선하기 위해 그림 1과 같이 STN-YOLO를 제안한다. 제안 모델은 YOLO의 백본 네트워크 앞 단에 STN을 추가하여 모델이 I 의 공간적 변형을 동적으로 학습하여 조정하고 적용할 수 있다.

STN은 2.1절에서 설명한 것과 같이 f_{loc} 가 I 를 통해 θ 를 추정하고, G 는 θ 를 사용하여 $T_{\theta}(G_i)$ 를 생성한다. 이후 f_{samp} 는 $T_{\theta}(G_i)$ 와 I 를 사용하여 I' 를 재구성한다. 이는 수식 (2), (3)과 같다.

$$\theta = f_{loc}(I) \quad (2)$$

$$I' = f_{samp}(T_{\theta}(G_i), I) \quad (3)$$

STN을 통해 생성된 I' 는 YOLO v5 f_{yolo} 에 입력되어 객체의 위치와 클래스를 탐지한다. 객체 탐지 결과 O 를 얻는 과정은 수식 (4)와 같다.

$$O = f_{yolo}(I') \quad (4)$$

이를 통해 STN-YOLO는 입력 이미지의 공간적 변형을 동적으로 조정하도록 학습하고 적용함으로써, 다양한 환경 변화나

이미지 내 변형에도 견고한 객체 탐지 결과를 제공한다.

3. 실험

3.1 실험 환경

본 연구에서는 뇌종양 탐지 성능을 평가하기 위해 뇌 MRI 데이터 세트인 BraTS 2020 [5] 학습 데이터 세트 중 T2를 사용했다. 총 369명의 환자 영상으로 구성되어 있으며, 모델의 학습 및 검증을 위해 7:2:1의 비율로 나누어 학습, 검증 및 테스트 집합으로 사용했다.

2D 뇌종양 탐지 모델 학습을 위해, 3D 형태인 T2 영상을 Z축을 따라 잘라내어 만든 2D 슬라이스를 입력으로 사용했다. 또한 정답 영상에서 뇌종양의 위치를 box 형태로 표현하기 위해, box의 왼쪽 위 꼭지점의 x, y 좌표 및 box의 넓이(width)와 높이(height)를 정답으로 사용했다.

3.1 실험 결과

뇌종양 탐지 성능 평가에는 mean average precision(mAP)와 mPC 지표가 사용되었다. mAP는 예측된 bounding box와 실제 간의 IoU를 기반으로 계산되며, 이때 정답으로 판단하는 임계값을 0.5로 고정하면 mAP_0.5, 0.5에서 0.95로 0.05씩 변화하며 모든 값의 평균을 사용하면 mAP_0.5:0.95이다. mPC는 모델의 견고성을 평가하기 위해 변형된 데이터에 대한 mAP의 평균이다.

우리는 입력 변형에 대한 STN의 우수성을 입증하기 위해, STN-YOLO를 YOLOv5와 비교한다. 표 1의 YOLOv5는 데이터 증강 기법을 적용하지 않은 모델이고, Yolov5(Aug)와 STN-YOLO는 학습 시에 좌우 반전(flipLr), 혼합(mixup), 모자이크(mosaic) 등을 사용하여 학습했다. 테스트 시에는 각 모델의 입력 변형에 대한 견고성을 확인하기 위해, 학습 시에 사용하지 않았던 이동(translate), 축 변형(shear), 회전(rotation)을 각각 적용하여 성능을 비교했다.

표 1은 각 모델의 견고성에 대한 실험 결과이다. STN-YOLO는 데이터 증강을 적용한 YOLOv5와 비교하여 mPC_0.5에서 2.6%, mPC_0.5:0.95에서는 7.9% 향상된 성능을 보였다. 임계값이 0.5로 고정된 mAP_0.5에 대해서는 원본 이미지에 대한 뇌종양 탐지 성능은 유지하면서 모든 변형에 대해서 데이터 증

표 1 STN-YOLO 성능 비교

	mAP_0.5				mPC	mAP_0.5:0.95				mPC
	Original	Translate	Shear	Rotate		Original	Translate	Shear	Rotate	
YOLOv5	0.686	0.663	0.666	0.597	0.642	0.414	0.351	0.364	0.307	0.359
YOLOv5 (Aug)	0.803	0.777	0.783	0.731	0.764	0.503	0.396	0.437	0.349	0.394
STN-YOLOv5	0.802	0.789	0.792	0.751	0.777	0.474	0.44	0.431	0.413	0.428

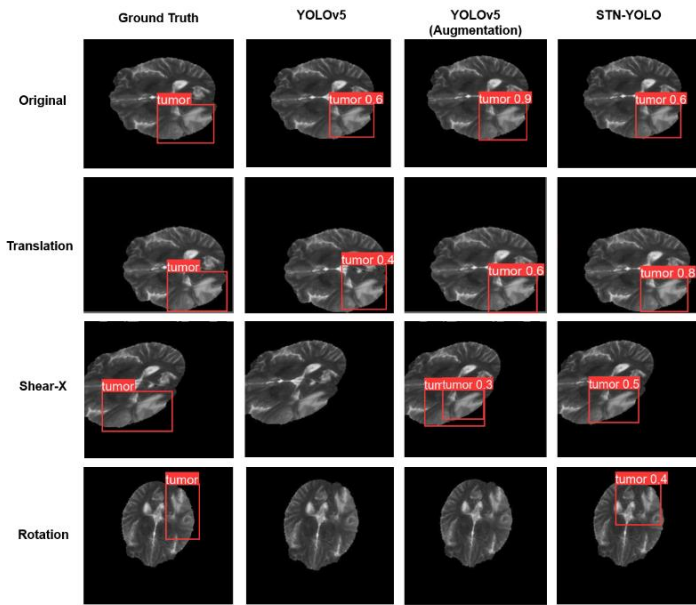


그림 2 변형된 데이터에 대한 뇌종양 예측 결과 시각화

강 기법을 사용한 YOLOv5 보다 향상된 성능을 보였다. 임계값을 0.5에서 0.95까지 변화시키는 mAP_0.5:0.95에 대해서는 이동과 회전 변형에 대해서 다른 모델들보다 향상된 성능을 보였으며 특히 회전 변형은 데이터 증강 기법을 사용한 YOLOv5 보다 15.49% 향상된 것을 확인했다.

그림 2는 각 모델의 뇌종양 위치 예측 결과를 시각화한 것이다. 모든 모델이 원본과 이동에 있어서 잘 예측하는 것을 확인할 수 있다. 하지만 원본 데이터만 학습한 YOLOv5는 축 변형, 회전에 대해서는 뇌종양 탐지에 실패한 것을 볼 수 있다. 데이터 증강 기법을 사용한 YOLOv5는 축 변형은 뇌종양을 두 개로 탐지하고 회전에 대해서는 탐지에 실패한 것을 볼 수 있다. 반면 STN-YOLO는 모든 변형에 대해 탐지를 성공한 것을 확인할 수 있다. 이러한 결과는 이동과 같은 단순한 변형에는 딥러닝 모델의 특징에 따라 뇌종양 탐지가 가능하지만 축 이동이나 회전과 같은 복잡한 변형은 STN과 같은 변형에 특화된 모듈이 필수적인 것을 의미한다.

4. 결 론

뇌종양 탐지 분야에서는 실제 의료 환경을 위해 다양한 조건과 상

황에서 뇌종양을 정확하게 탐지할 수 있는 모델의 견고성이 매우 중요하다. 일반적으로 견고성을 높이기 위해 데이터 증강 기법을 활용하지만 새로운 변형에 대한 견고성을 확보하기 어렵다는 한계점이 존재한다. 본 논문에서는 이를 해소하기 위해 STN-YOLO를 제안했다. STN을 통해 모델의 구조적 변형에 효과적으로 적응할 수 있게 함으로서 데이터 증강기법을 사용한 YOLOv5 보다 변형 데이터에 대해서 견고한 결과를 보여주었다. 그러나 변형에 특화된 학습이 원본 데이터에 대한 성능을 상대적으로 낮게 만든 것을 확인했다. 따라서 향후 연구에서는 원본데이터에 대한 성능을 개선할 수 있는 방법을 찾아, 모델의 일반화 성능을 높이고 의료 영상 분석 분야에서의 적용 가능성을 더욱 확장할 수 있는 연구를 진행하고자 한다. 특히, 신약 개발에 있어서 신약 후보물질이 뇌에 미치는 영향을 정밀하게 파악하기 위해 뇌 영상에서의 변화를 정확하게 탐지함으로써, 신약의 효능 및 안전성 평가에 있어 보다 신뢰할 수 있는 결과를 제공할 수 있을 것으로 기대한다.

참 고 문 헌

- [1] Wang, Hai, et al. "YOLOv5-Fog: A multiobjective visual detection algorithm for fog driving scenes based on improved YOLOv5." IEEE Transactions on Instrumentation and Measurement 71 (2022): 1-12.
- [2] Dipu, Nadim Mahmud, Sifatul Alam Shohan, and K. M. A. Salam. "Deep learning based brain tumor detection and classification." 2021 International conference on intelligent technologies (CONIT). IEEE, 2021.
- [3] Wang, Hai, et al. "YOLOv5-Fog: A multiobjective visual detection algorithm for fog driving scenes based on improved YOLOv5." IEEE Transactions on Instrumentation and Measurement 71 (2022): 1-12.
- [4] Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in neural information processing systems 28 (2015).
- [5] H. Jia, W. Cai, H. Huang, Y. Xia, H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task, 2020. arXiv:2012.15318