

MedMNIST 3D 딥러닝 모델 개발 결과 보고서

김윤우 (kyw2448@gmail.com)

1. 개발 목표

- ① PyTorch를 이용해 medical 3D image data를 deep learning 하는 모델을 직접 제작해보자.
- ② 보편적으로 사용하는 deep learning skill을 연습해보자.
- ③ AUC score와 accuracy 양쪽에서 높게 평가되는 network를 찾아보자.

2. 전략

- ① complexity를 다르게 하여 세 가지 버전의 network를 사용한다.
- ② 세 network를 사용하여 각각 여러 차례 학습을 진행한다. 이때 각종 hyper parameter의 값은 대부분의 test에서 좋은 결과를 얻어낸 값으로 통일한다.
- ③ 각각의 학습에서 epoch time에 따라 끝까지 진행한 last model과 valid test에서 최고의 AUC score를 기록한 best model을 저장한다.
- ④ valid test의 결과를 비교하여 모델 각각에 weight를 부여한 ensemble을 통해 최종 결과를 얻는다.

3. 깃허브 파일 요약 - <https://github.com/JakeYunwooKim/medmnist> & 별지 참조

- ① mnist3dnet.py : MedMNIST의 3D data를 deep learning 하기 위해 사용되는 기본 네트워크(Net1, Net2, Net3) 코드를 담은 파일
- ② main.py : deep learning과 관련된 주요 hyper parameter의 값과 불러올 data의 종류를 지정 관리하고 transformation과 data loader를 설정하는 코드를 담은 파일
- ③ train.py : main.py의 설정값을 불러와 deep learning을 반복 진행하고, 각각의 count에서 얻어낸 중간단계 모델의 parameter 값을 저장하는 코드를 담은 파일
- ④ ensemble.py : deep learning을 통해 얻은 중간단계 모델들의 test data에 대한 output 값들을 ensemble 하여 최종 결과를 얻어내는 코드를 담은 파일

4. 결과 - MedMNIST benchmarking score와 비교

Methods	OrganMNIST3D		NoduleMNIST3D		FractureMNIST3D		AdrenalMNIST3D		VesselMNIST3D		SynapseMNIST3D	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
ResNet-18 + 2.5D	0.977	0.788	0.838	0.835	0.587	0.451	0.718	0.772	0.748	0.846	0.634	0.696
ResNet-18 + 3D	0.996	0.907	0.863	0.844	0.712	0.508	0.827	0.721	0.874	0.877	0.820	0.745
ResNet-18 + ACS	0.994	0.900	0.873	0.847	0.714	0.497	0.839	0.754	0.930	0.928	0.705	0.722
ResNet-50 + 2.5D	0.974	0.769	0.835	0.848	0.552	0.397	0.732	0.763	0.751	0.877	0.669	0.735
ResNet-50 + 3D	0.994	0.883	0.875	0.847	0.725	0.494	0.828	0.745	0.907	0.918	0.851	0.795
ResNet-50 + ACS	0.994	0.889	0.886	0.841	0.750	0.517	0.828	0.758	0.912	0.858	0.719	0.709
auto-sklearn	0.977	0.814	0.914	0.874	0.628	0.453	0.828	0.802	0.910	0.915	0.631	0.730
AutoKeras	0.979	0.804	0.844	0.834	0.642	0.458	0.804	0.705	0.773	0.894	0.538	0.724
MyNet	0.997	0.933	0.914	0.868	0.757	0.604	0.854	0.815	0.973	0.969	0.925	0.878

(표 : MedMNIST benchmarking score와 개별 제작한 model의 score 비교)

- ① ensemble 전 best model의 결과는 기존의 model과 근사한 값을 보임.
- ② ensemble 후 대부분의 score가 MedMNIST의 benchmarking score보다 높게 나옴.
- ③ NoduleMNIST3D의 accuracy는 auto-sklearn에 조금 못 미침.
 - 원인 : nodule 한정으로 Net1에서의 score가 낮게 측정된 것이 결과에 반영됨
 - auto-sklearn 이외의 모델보다는 높음.