



Deep learning based Feature Selection Study

Sanghyuck Lee

2021. 08 03.

Outline

- Knockoffs Pruning for Segmentation Model
- 중견 연구 과제
- 갑상선 연구 과제
- 산학 협력 과제

Knockoffs Pruning for Segmentation Model

In Supplementary Material of SCOP

- Train GAN to construct knockoff data.
- The knockoff data are generated by the generator and then sent to the discriminator to **verify whether the knockoff condition** holds.
- The generator and discriminator are optimized alternately and the **loss function** provided by [1] is adopted.
- Considering that images are high dimension data, we take **multiple pixels** as a whole when **exchanging elements** between real data and knockoff data.
- To reduce the training cost on ImageNet, knockoff data with size 64×64 are generated first and then upsampled to 224×224 .
- For the network architectures, **multiple convolutional (deconvolutional) layers are stacked to construct the generator and discriminator on CIFAR-10**, and **ResNet-like architectures are adopted on ImageNet**. All the models are optimized with Adam [5] optimizer.
- On CIFAR-10, the learning rate, batchsize and number of iterations are set to 0.001, 128 and 20K, which are set to 0.001, 512 and 30K on ImageNet.

중견 연구 제안서 작성

- 국내 의료 AI 관련 정책 과제 조사

- 보건복지부가 주관하여 2025년까지 향후 5년간 총 945억원 사용해, 의료데이터셋과 의료 인공지능 기술 개발을 진행함. => 총 3가지 사업으로 요약됨.

1) 중환자 특화 빅데이터 (Korean Medical Information Mart for Intensive Care, K-MIMIC) 구축 및 AI 기반 임상 의사결정 지원 시스템 (Clinical Decision Support System, CDSS) 개발

- 데이터 정제를 통해 데이터 사용자 수요에 따라 제공 가능한 데이터셋 구성
- 확보된 중환자 데이터를 공유 분석하는 플랫폼 구축
- 중환자 CDSS를 위한 설명 가능한 AI 알고리즘 개발: 임상적 인과관계에 근거하여 중증도 평가, 재입원 가능성 등을 설명할 수 있는 AI 알고리즘 개발
- 임상 활용 검증 및 실증 연구

중견 연구 제안서 작성

2) 디지털 병리 기반의 암 전문 AI 분석 솔루션 개발

: 병리 진단이 필수적인 암에 대하여, AI 병리 분석 솔루션 개발

=> 병리 판독 업무 효율화, 환자 진단 및 치료 효과성 제고

- 병리 데이터, 주석 처리를 비롯한 표준화 처리, 의료 기관 간의 이미지 색상 차이, 스캐너 제조사에 따른 차이 등을 위한 표준화 기술 개발

- 설명 가능한 AI 알고리즘 기반 SW 개발, 임상 검증 및 활용 평가

3) 합성 의료 데이터 생성 및 임상 활용 실증 연구

: 민감 정보인 의료 데이터 정보 보호 및 활용성 제고를 위해, 공개 목적의 합성 의료 데이터 생성 및 임상적 유효성 및 활용성 검증 연구 지원

- 시나리오 기반 고품질 데이터셋 구축: 합성 데이터 생성 모델 개발

- 합성 데이터의 임상적 유효성 및 활용성 실증 연구

중견 연구 제안서 작성

요약 및 의견:

현재 국가적 차원에서 관심 있어 하는 내용은,

- 실생활에서 사용 가능, 설명 가능한 의료 AI 개발 => 진단 효율성 증대 및 의료 서비스 격차를 줄임.

가 하나의 최종 목표로 보이고, 이를 위한 기반 사업으로,

- 데이터셋 표준화 기술 및 공유 플랫폼, 합성 의료 데이터 생성 기술 개발 등을 진행하고자 함을 알 수 있었고,

이와 같은, 국가적 차원의 큰 그림에서, 당장 표준화된 데이터셋을 확보하기는 어려울 수 있으나, 전처리 과정과 표준화 방식, 학습 모델 제안 등을 포함한 우리만의 프레임워크가, 실생활에서 어느 정도 사용 가능한 세그멘테이션 성능을 보여줌으로써, 추후 표준화 기술에 대한 방향성 제시나, 국가 표준 세그멘테이션 모형의 초석을 다지는 데에 이바지함이 우리의 역할로 있을 수 있음.

중견 연구 제안서 작성

- 가제:

한글: 의료 분야내 주요 관심 구조물 세그멘테이션을 위한 사전훈련 최신 딥러닝 모델 개발

영문: Development of Pre-trained Deep Learning Model for Segmentation of Structures of Major Interest in the Medical Field

- 연구 목표:

본 과제에서는 몇몇 주요한 해부학적 구조에 대해서 세계 최고 수준의 성능을 갖는 사전학습된 세그멘테이션 모델을 배포함으로써, 다양한 의료분야 응용에 성공적으로 기여하는 것을 최종 목표로 하며, 이를 위해 다음과 같은 연차별 연구 목표를 설정하여 추진한다.

중견 연구 제안서 작성

1. 연구의 목표 (2페이지)

- 최종 연구 목표
- 연차별 목표 및 개요도
- 연차별 연구 개발 내용, 창의성 및 도전성, 추진 전략, 주요 결과물

2. 연구의 필요성 (2페이지)

- 4~5개로 나누어 총 2페이지

3. 연구자의 연구 수행역량 (1페이지)

- 연구실 스펙 및 노하우 등 자랑거리

중견 연구 제안서 작성

4. 연구의 추진 전략 및 방법 (연차별 1.5페이지)

- 연차별 상세 내용 및 주요 기술 설명 (그림 포함하여, 연차당 3가지 정도)

5. 연구결과의 중요성 (0.5페이지)

- 학문적/기술적 파급 효과
- 경제적/사회적 파급 효과

6. 연구기간 및 연구비 적정성 (0.5페이지)

- 추진일정 등
- 연구비 적정성

중견 연구 제안서 작성

- 진행 중인 내용:

- 다운 가능한 18개(/53개) 데이터셋(추가 확보 조사 긍정적), 단순 배열 파일(npz) 및 이미지 파일 (jpg) 형식으로 저장하여, 직접 확인 후, 중견 연구 계획에 포함할지 결정

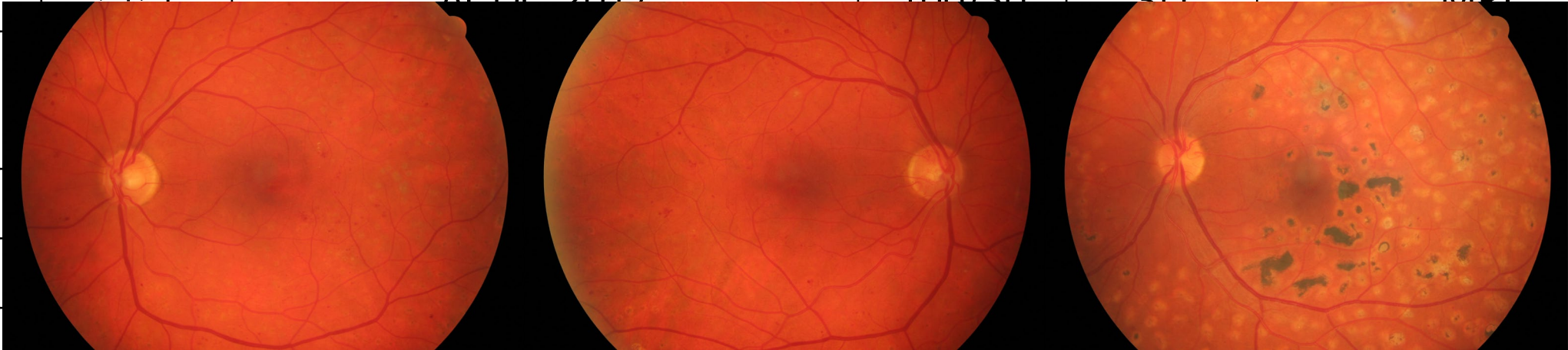
진행한 해외 의료 데이터셋 조사 + 국내 의료 AI 과제 조사

=> 현실적인 제안서 작성을 위한 응용

중견 연구 제안서 작성

연번	장기	데이터셋 이름	샘플 수	차원	촬영법
1	뇌	BraTS 2020	369	3D	MRI
2	뇌	IBSR	18	2D	MRI
3	망막	HRF	45	2D	사진 (Photograph)
4	머리, 목	Head&Neck	25/15	2D	CT
5	심장	ACDC	150	2D	MRI
6	심장	ACDC 2017	100/50	3D	MRI
7	유방	INbreast	116	2D	유방조영술 (Mammography)
8	폐	JSRT	247	2D	CT
9	폐	LIDC-IDRI	244527	2D	CT
10	폐	LUNA16	880	2D	CT

중견 연구 제안서 작성

연번	장기	데이터셋 이름	샘플 수	차원	촬영법
1	뇌	BraTS 2020	369	3D	MRI
2	뇌	IBSR	18	2D	MRI
3	망막	HRF	45	2D	사진 (Photograph)
4	머리, 목	Head&Neck	25/15	2D	CT
5	심장	ACDC	150	2D	MRI
6	신장	ACDC 2017	100/50	3D	MRI
7		LONATO	660	2D	CT
8					
9					
10					

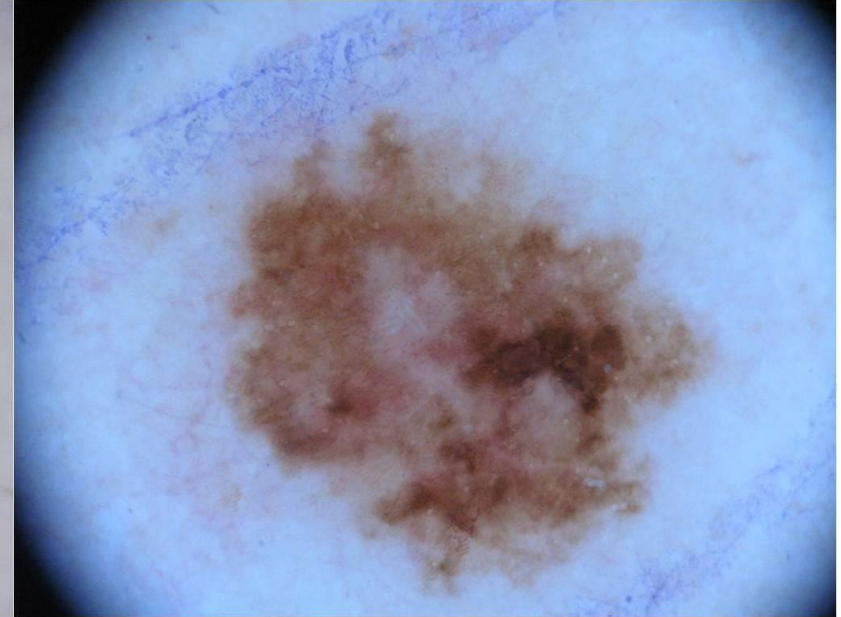
중견

연번					촬영법
1					MRI
2					MRI
3					(Photograph)
4					CT
5	심장	ACDC	150	2D	MRI
6	심장	ACDC 2017	100/50	3D	MRI
7	유방	INbreast	116	2D	유방조영술 (Mammography)
8	폐	JSRT	247	2D	CT
9	폐	LIDC-IDRI	244527	2D	CT
10	폐	LUNA16	880	2D	CT

중견 연구 제안서 작성

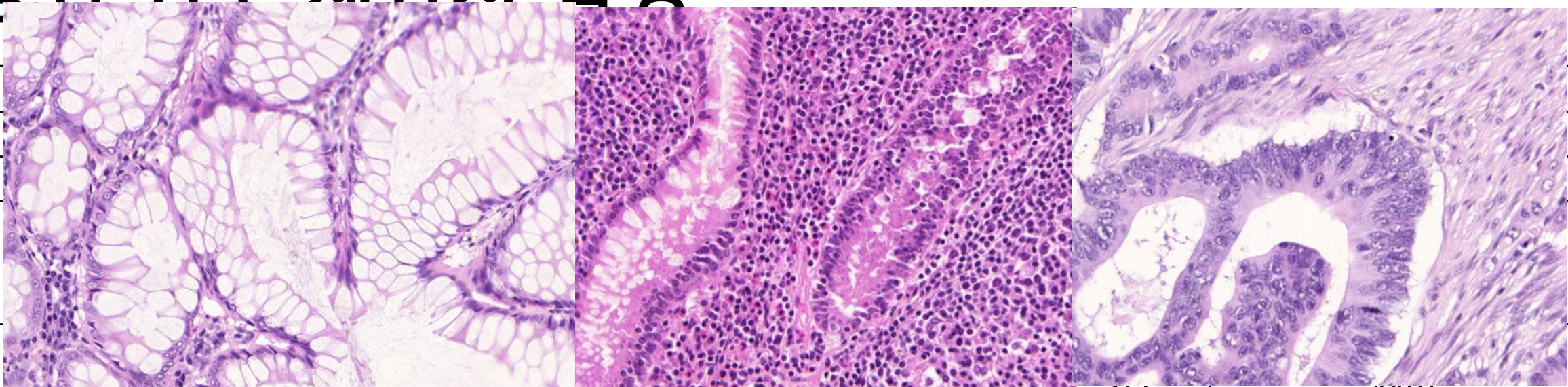
연번	장기	데이터셋 이름	샘플 수	차원	촬영법
11	간	SLiver07	30	3D	CT
12	간	3DIRCADb	20	3D	CT
13	복부 (Abdominal)	NIH-CT-82	82	3D	CT
14	복부 (Abdominal)	UFL-MRI-79	79	3D	MRI
15	피부	ISBI 2016	1250	2D	피부 확대경 (Dermoscopic image)
16	피부	ISIC 2017	2000/600	2D	피부 확대경 (Dermoscopic image)
17	피부	ISIC 2018	2594/1000	2D	피부 확대경 (Dermoscopic image)
18	영양관(분비선, Gland)	GlaS	85/80	2D	병리 이미지 (Histopathology image)

주요 연구 데이터셋



14	(Abdominal)	ISBI-MIRF-75	75	3D	MIRF
15	피부	ISBI 2016	1250	2D	피부 확대경 (Dermoscopic image)
16	피부	ISIC 2017	2000/600	2D	피부 확대경 (Dermoscopic image)
17	피부	ISIC 2018	2594/1000	2D	피부 확대경 (Dermoscopic image)
18	영양관(분비선, Gland)	GlaS	85/80	2D	병리 이미지 (Histopathology image)

중견 연구 제안서 작성



14	(Abdominal)	UPL MRI 7.5	7.5	3D	MRI
15	피부	ISBI 2016	1250	2D	피부 확대경 (Dermoscopic image)
16	피부	ISIC 2017	2000/600	2D	피부 확대경 (Dermoscopic image)
17	피부	ISIC 2018	2594/1000	2D	피부 확대경 (Dermoscopic image)
18	영양관(분비선, Gland)	GlaS	85/80	2D	병리 이미지 (Histopathology image)

중견 연구 제안서 작성

연번	장기	데이터셋 이름	샘플 수	차원	촬영법
1	뇌	BraTS 2020	369	3D	MRI
2	뇌	IBSR	18	2D	MRI
3	망막	HRF	45	2D	사진
4	머리, 목	Head&Neck	25/15	2D	CT
5	심장	ACDC	150	2D	MRI
6	심장	ACDC 2017	100/50	3D	MRI
7	유방	INbreast	116	2D	유방조영술
8	폐	JSRT	247	2D	CT
9	폐	LIDC-IDRI	244527	2D	CT
10	폐	LUNA16	880	2D	CT
11	간	SLiver07	30	3D	CT
12	간	3DIRCADb	20	3D	CT
13	복부	NIH-CT-82	82	3D	CT
14	복부	UFL-MRI-79	79	3D	MRI
15	피부	ISBI 2016	1250	2D	피부 확대경
16	피부	ISIC 2017	2000/600	2D	피부 확대경
17	피부	ISIC 2018	2594/1000	2D	피부 확대경
18	영양관	GlaS	85/80	2D	병리 이미지

갑상선 연구

- 새로운 세그멘테이션 진행 내용 – AX 1, 2, 4, 5

: 기존의 슬라이스 번호에 기반한 AX 슬라이스 선택 방법

-> DCM파일내 'Image Position' 정보에 기반한 선택 방법

기존 방식의 문제점:

근육 부피가 Activity 분류의 중요 정보로써 전달되길 의도함

-> But, 3mm와 7.5mm는 유의미하게 입력 간의 의도를 방해하는 변동이 될 것으로 판단함

해결 방안:

Dicom파일에 포함된 'Image Position' 정보에 기반해서

AX 1과 AX5: 1.875mm~7.5mm -> 6.73mm~8.16mm

AX 2와 AX4: 0.625mm~2.5mm -> 1.77mm~2.5mm 로 변동 줄임

갑상선 연구

- 세그멘테이션 진행 예정
 - 기존 1,237 환자 중, 732 환자의 AX 1, 2, 4, 5 재진행
 - AX1과 AX5: 732개 슬라이스에 대해서 눈, 내직근, 외직근, 시신경
($732 \times 4 \times 2 = 5,856$ 장)
 - AX2과 AX4: 703개 슬라이스에 대해서 눈, 내직근, 외직근, 시신경
($703 \times 4 \times 2 = 5,624$ 장)
- Total = $5,856 + 5,624 = 11,480$ 장

산학 협력 프로젝트

- 제출 실적

- ~~SW 제작 (산학 1), 국제학술대회 논문 1편 이상 게재 (산학 2), 국내 특허출원 1건 이상 (산학 2)~~

- 현재 진행 사항

- ~~—SW 등록: 완료~~

- 특허 출원: 특허 담당자 분 메일 수신 대기중

- ~~—국제 컨퍼런스 논문 작업: 아성이 PlatCon 컨퍼런스 논문 작업~~

Reference

[1] Jordon, James, Jinsung Yoon, and Mihaela van der Schaar. "KnockoffGAN: Generating knockoffs for feature selection using generative adversarial networks." *International Conference on Learning Representations*. 2018.