

# 슬라이드 시스템

---

Incremental learning – 추천 모듈 테스트

2023.02.23. 발표자료  
- 작성자 : 석사과정 강현구

슬라이드 추천 시스템 |

진행사항(Summary)

1. 재진단 일정 협의 중

- 씨젠 내, Stomach 모델 업데이트에 대한 요구 식별
- 재진단 관련 세부사항 논의중

2. AL 시스템 세부 사항 수정

- Patch generator <-> Patch train 모듈 조율
- False N 자동 라벨링 조건 변경
- Colon Threshold 값 수정

3. WSI / Patch train 모듈 점검 완료

4. 추후 일정

5. 요청 사항

1. 재진단 일정 협의중

- 김무진 연구원이 고영신 부장과 1차 미팅을 진행했습니다
- 씨젠 측에서, Stomach 모델 업데이트를 희망한다는 점을 확인했으며,
- 재진단 대상 / 방법 등 세부 사항에 대해서 추가 협의가 필요합니다.
- 다음 주 중으로 마무리될 것으로 판단됩니다.

2. AL 시스템 세부 사항 수정

- 전반적인 AL 시스템 점검을 진행했습니다.

3. WSI / Patch train 모듈 점검 완료

- 몇 차례 하드웨어적 오류가 발생하였으나, 원인 식별 및 조치했습니다

4. 추후 일정

- 세부사항은 뒤에서 상세히 설명하겠습니다

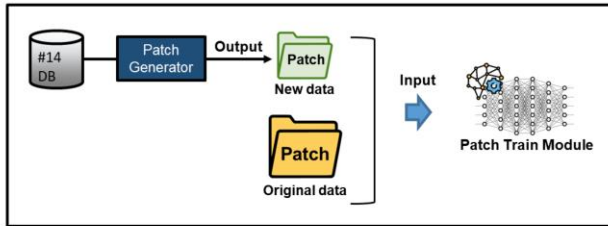
5. 요청 사항

- 세부사항은 뒤에서 상세히 설명하겠습니다

## 슬라이드 추천 시스템 |

### 2. AL 시스템 세부 사항 수정

#### 1) Patch generator <-> Patch train 모듈 조율



- AL 시스템의 각 모듈이 유기적으로 연결되기 위해, 각 단계별 입력/결과값 조율이 필요함
- 신규 데이터의 경우, 학습 기간에 따른 데이터 필터링 기능 필요

#### 결론

##### <Before>

- 양식: 날짜/ anatomy/ 데이터 타입/ 추천자/ Ground truth
- 예시 : 20221220/colon /patch/system/D



##### <After>

- 양식: 날짜/ anatomy/ **데이터 타입**/ 추천자/ Ground truth
- 예시 : 20221220/colon / **patch** /system/D

### 2. AL 시스템 세부 사항 수정

- 각 모듈 간 세부사항에 대해 조율이 필요합니다.
- 한 예로, Patch generator 모듈의 결과값을, Patch train 모듈이 입력값으로 받아 조율이 필요합니다.
- 또한 AL 시스템을 통해 확보한 신규 데이터의 경우,
- 추후 모델 업데이트 간 사용한 데이터를 구분하기 위해 학습 기간에 따른 데이터 필터링 기능이 필요합니다.
- 위의 상황을 모두 고려하여,
- 기존의 Patch generator의 결과 양식에서 "데이터 타입" 만 제거하며
- Patch Train의 입력값 양식을 위 양식에 맞추도록 결론을 내렸습니다.

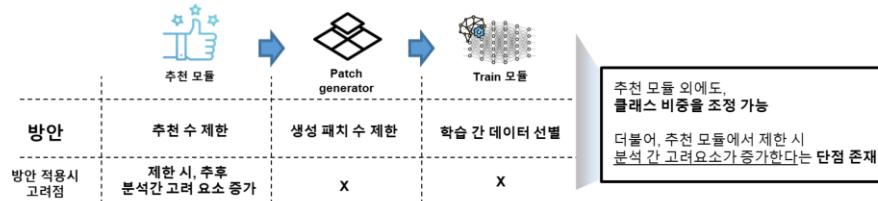
## 슬라이드 추천 시스템 |

### 2. AI 시스템 세부 사항 수정

고려점 : 모듈 학습 시, 클래스별 비중을 맞춰야(Balance) 학습 편향을 방지할 수 있음

기존 방안 : Patch 추천 단계에서 패치 수를 조절

- Ex)- Threshold 값 하향 조정, WSI 당 추천 Patch 수 조절



#### 수정 사항

- 2) [WSI GT N (Prediction D/M)] 자동 라벨링 대상 변경
  - 기존 : [Patch] 모델의 D/M 로 잘못 예측한 경우 중에서 Over-confident 상위 20개
  - 현재 : [Patch] 모델의 D/M 로 잘못 예측한 경우 전체
- 3) Colon Threshold 값 수정(0.55 -> 0.8)

## 2. AI 시스템 세부 사항 수정(계속)

- 또한, 2가지 사항에 대해 추가 수정을 진행했습니다.
- 모델을 학습할 때, 특정 클래스에 대한 학습 편향을 방지하기 위해 Balance train 데이터셋을 마련하는 것이 중요합니다.
- 이를 위해, 기존에는 Patch 추천 단계에서 패치 수를 조절하고자 했습니다.
- 하지만 추천 이후, Patch generator / Train 모듈 단계에서도 클래스 비중을 조절 가능하며,
- 오히려 추천 모듈에서 패치 수를 제한 시, 추후 결과 분석 간 고려 요소가 증가할 수 있는 단점이 있습니다.
- 따라서 추천 단계에서 패치 수를 제한하던 것을 대폭 완화하며,
- Train 모듈 단계에서 클래스간 비중을 조율하는 쪽으로 결론을 내렸습니다.

#### 슬라이드 추천 시스템 |

### 3. WSI / Patch train 모듈 점검 완료

#### 학습간 문제 식별 [조치 완료]

- #226 Server GPU 사용간, 메모리 오류 발생
- [원인] 사용가능한 GPU를 혼용함
- [조치] 사용가능한 GPU 확인 및 설정 완료. 이후 모듈 정상 작동 확인

#### 제한점

- WSI 모델 학습 간, Feature cube txt 파일 형성간 대부분의 시간 소요

#### 추후 방안

- 1) [메모리 부족 이슈] GPU 사용 간, 각자 할당된 GPU를 구체화하여 공유하기
- 2) [Feature cube 이슈] WSI 모델 재학습 전, Feature cube txt 파일 생성 과정 추가하기

### 3. WSI / Patch train 모듈 점검 완료

- WSI / Patch train 모듈 모두 정상 작동 확인했습니다.
- 단, 점검 간 #226 서버 GPU 메모리 이슈가 발생했습니다.
- 현재 원인 식별 및 조치했으며, 모듈 정상 작동 확인했습니다.
- 유의점으로, WSI 모델 학습 간, Feature cube 형태의 txt 파일을 생성하는데 많은 시간이 소요됩니다.
- 따라서 이번 점검 간 경험한 문제점들을 해소하고자
  - 1) 공용 서버의 GPU 사용간, 각자 할당된 GPU를 구체화 및 공유
  - 2) WSI 모델 재학습 전 Feature cube txt 파일을 생성하는 과정 추가
- 방안을 강구하겠습니다.

#### 슬라이드 추천 시스템 |

##### 4. 추후 일정

- 1). 재진단 일정 구체화 (협의중)
- 2). 모델 업데이트 기준 구체화 (설계중)
- 3). Patch generator / Patch train 모듈 업데이트
  - 2.1) AL 시스템 모듈 간 연동 설계 반영
- 4). AL 시스템 – 모델 관리 설계
  - [DB] model\_management 항목 고도화

##### 5. 논의 사항

- 1) [데이터 위치] WSI 테스트 데이터셋
- 2) [데이터 위치] Patch generator 결과물

#### 4. 추후 일정

- 앞으로의 일정은 전문의 재진단 이후, 즉각 모델 성능 점검을 준비할 수 있는 데 초점을 잡고 있습니다.
- 1) 재진단 일정은 김무진 연구원이 고영신 부장과 협의중이며, 다음 주 중으로 구체화 하겠습니다.
- 2) 또한 실제 현업에 적용할 모델의 선정 기준을 설계하고 있습니다. 다음 주 중으로 구체화하겠습니다.
- 3) 이번 주에 논의한 내용을 기반으로 각 모듈 업데이트 하겠습니다.
- 4) 업데이트 된 판독 모델을 관리하기 위해 추가 DB를 설계하겠습니다.

#### 5. 논의사항

- 1)WSI 테스트 데이터셋 / 2) Patch generator 결과물 저장 위치를 씨젠 측과 논의하여 구체화하겠습니다.

감사합니다.