

# **위암 판독 모델 개발**

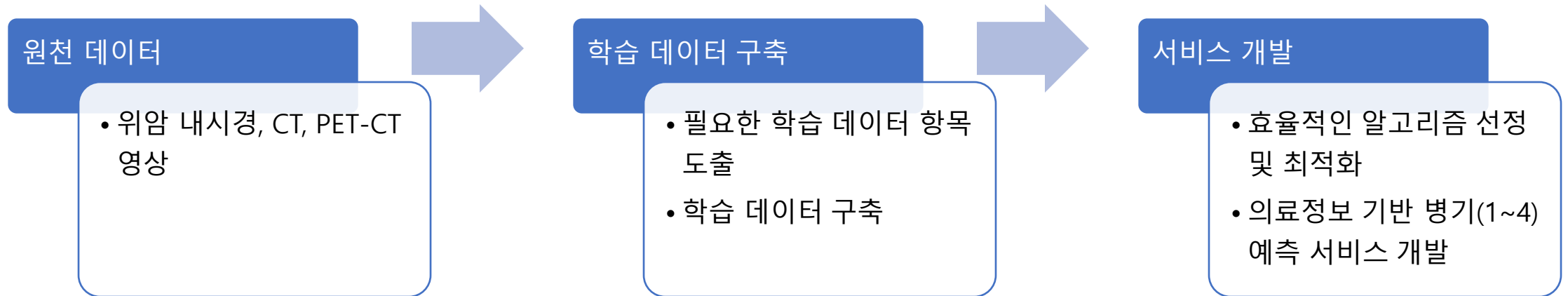
**Gastric Cancer Detection**

## 프로젝트 선정 동기

- 보건의료 분야: 머신러닝, 딥러닝의 가장 흥미로운 응용 분야 중 하나
- 위암은 전 세계적으로 다섯 번째 다빈도 악성종양, 암 관련 사망 원인 세 번째 우리나라 단위 인구당 위암 발생률 매우 높음(남자 2위, 여자 3위)
- 다양하고 실질적인 환자 상황 반영할 수 있는 요인들 찾아 적용, 분석할 필요성

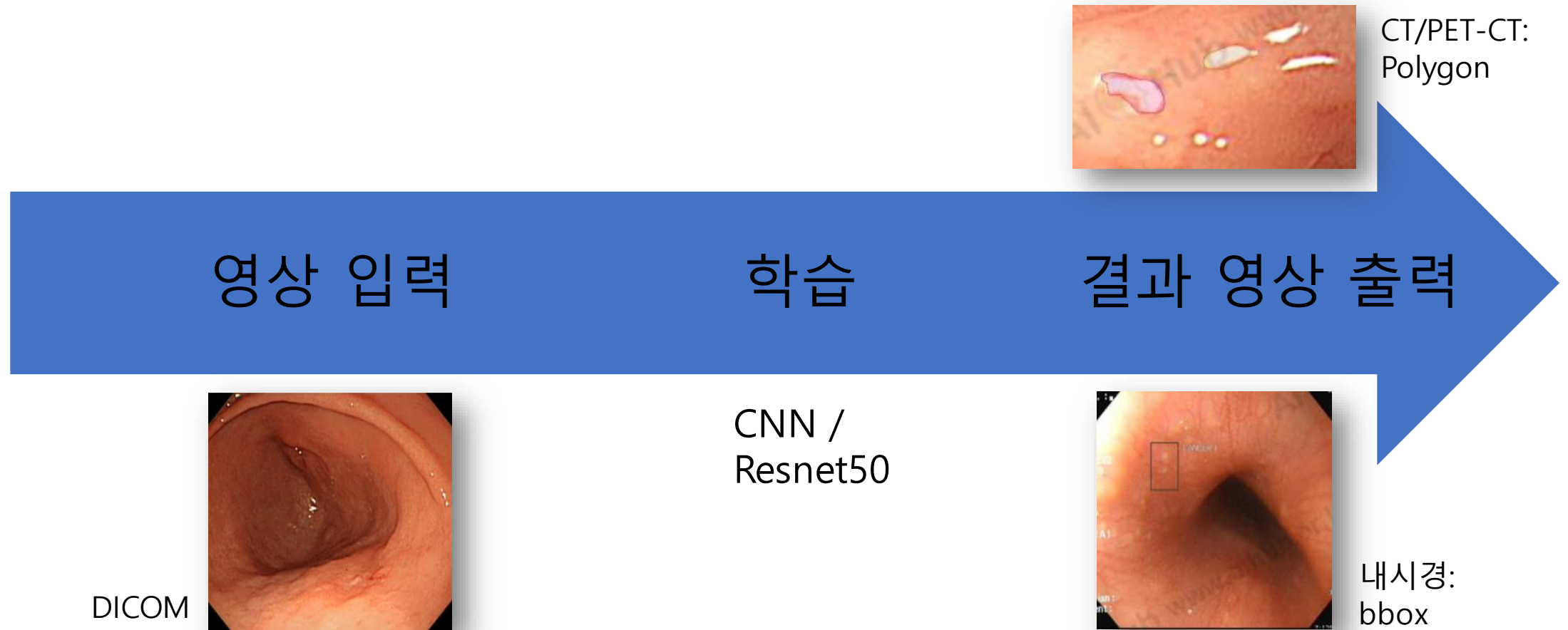
# 목표

- 위암 환자 영상에서 위암 관련 질환과 암 기수 자동 인식하는 인공지능 개발



# 목표

- 위암 환자 영상에서 위암 관련 질환과 암 기수 자동 인식하는 인공지능 개발



# 데이터셋

## • 위암 진단 의료 영상

- 가톨릭성모병원 EMR, PACS 시스템으로부터 추출, 정제된 내시경, CT, PET-CT 영상 데이터 구축
- AI 응용서비스 고려한 데이터 항목(나이, 성별, 병리소견 등) 포함
- 내시경, CT, PET-CT는 소화기계 암 진단에 중요한 정보 -> 위암 진단 및 병기 판단에 활용
- 실제 위암 환자 기수별 분포  
1기: 62.2%, 2기: 6.5%, 3기: 6.5%, 4기: 24.7%
- 데이터셋 성별 분포  
남:여 약 61% : 39%



실제 분포에 근접한 데이터

# 데이터셋

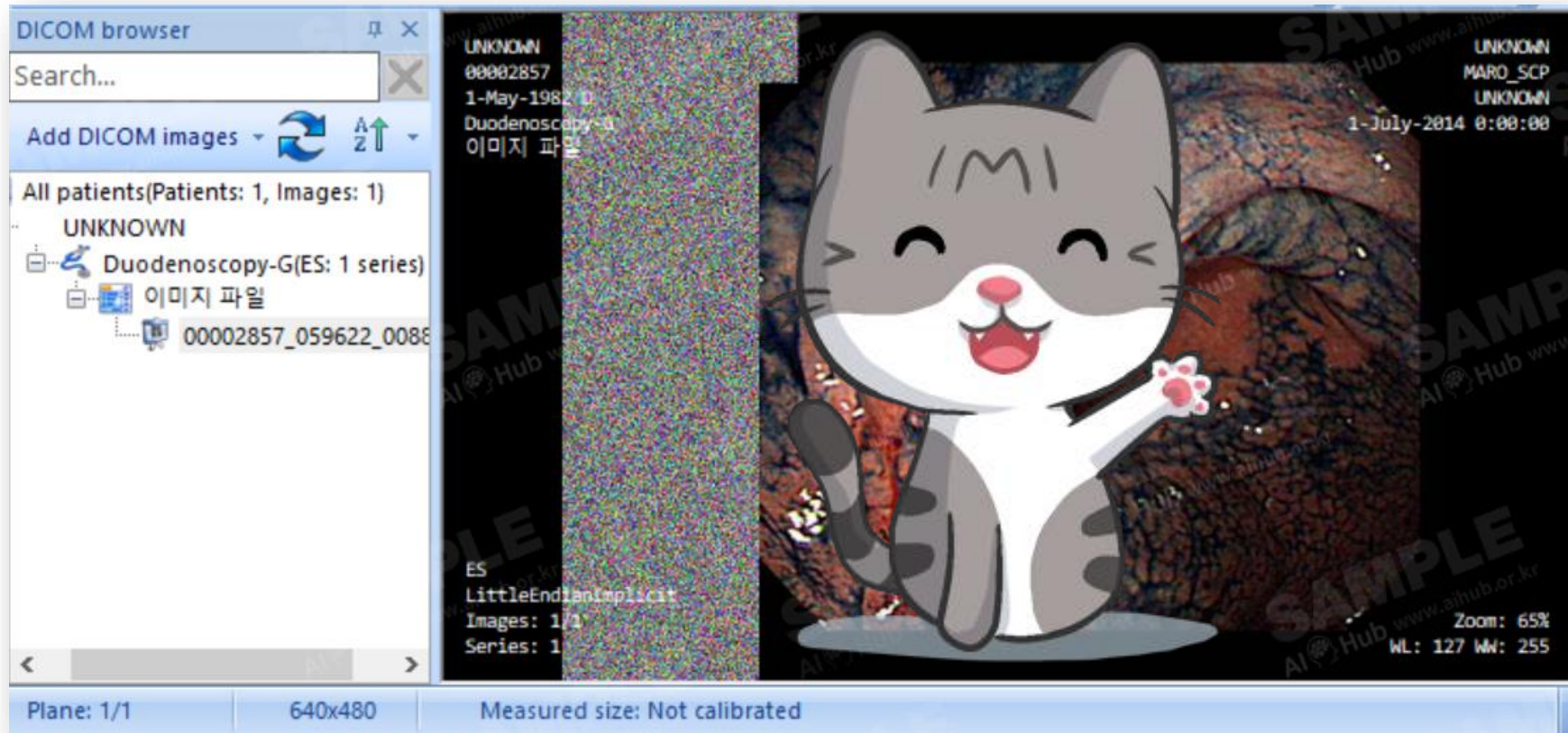
## • 데이터 구축 규모

- 원천데이터 11만 장(위암 내시경 2만 장, 위암 CT 2만 장, 위암 PET-CT 7만 장) 추출
- 내시경 사진 레이블링은 암 조직에 시행,  
이외 정상 부위/암이 아닌 양성종양 부위는 정상 부위로 판단

데이터 종류	데이터 형태	수량
위암 내시경	DICOM 영상	21,957장
위암 CT	DICOM 영상	22,978장
위암 PET-CT	DICOM 영상	71,384장
총계		116,319장

# 데이터셋

- 대표 도면



# 데이터셋

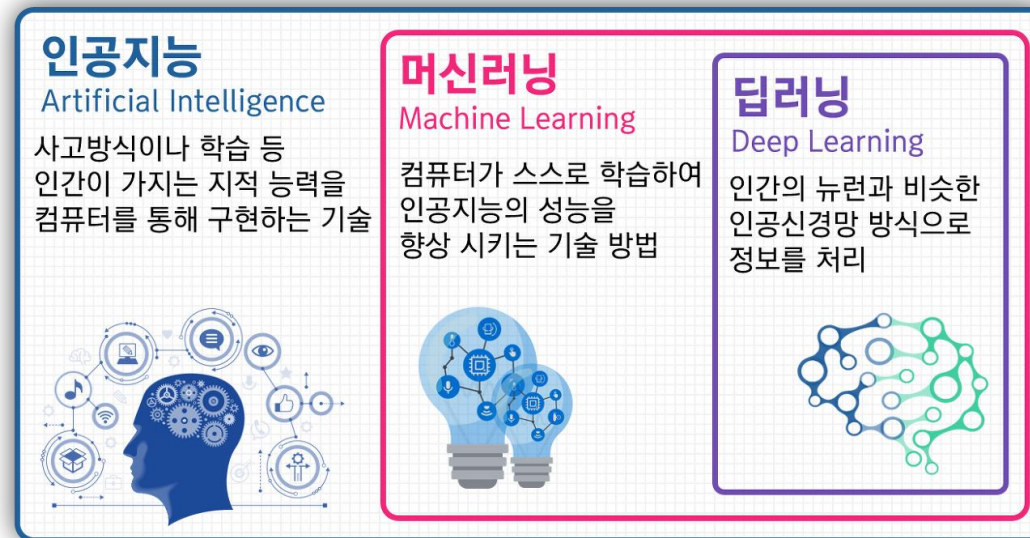
- 대표 도면

```
{
  "image": {
    "person_id": "00002857",
    "sex": "F",
    "image_name": "Gastric_Cancer/원천데이터/00002857/ENDO/00002857_059622_0088.dcm",
    "image_type": "DICOM",
    "image_seq": 88,
    "image_count": 109,
    "width": 640,
    "height": 480,
    "phase_id": 1,
    "annotations": [
      {
        "category": "cancer",
        "label_type": "polygon",
        "polygon": [
          [
            453.494140625,
            139.611328125
          ]
        ],
      }
    ],
  },
}
```



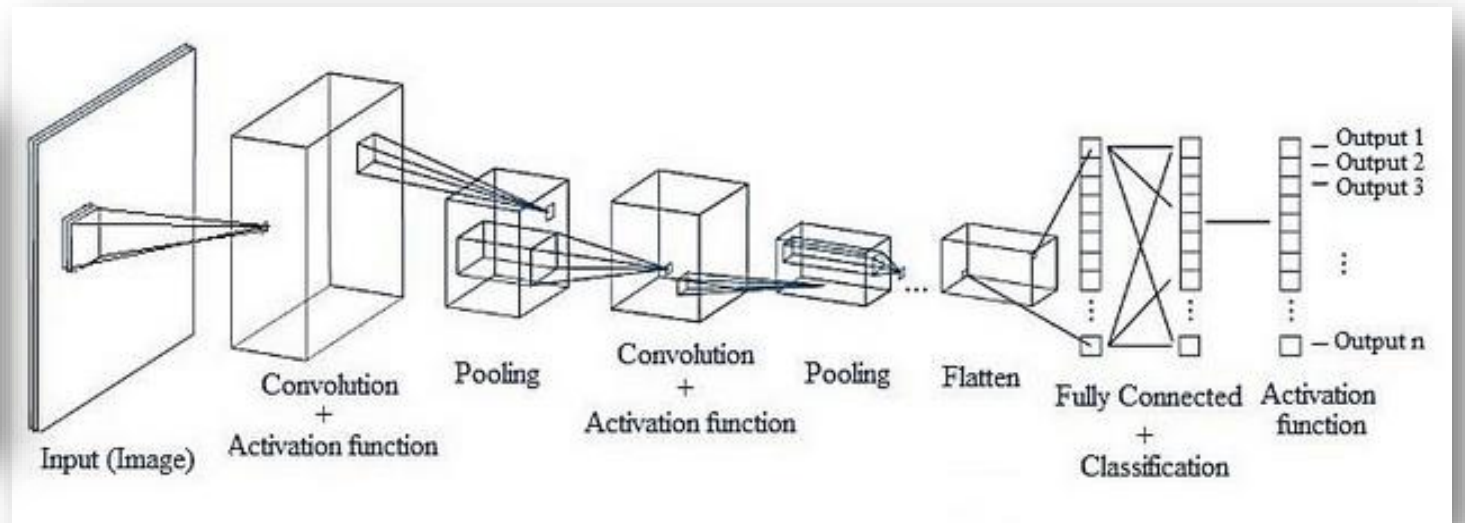
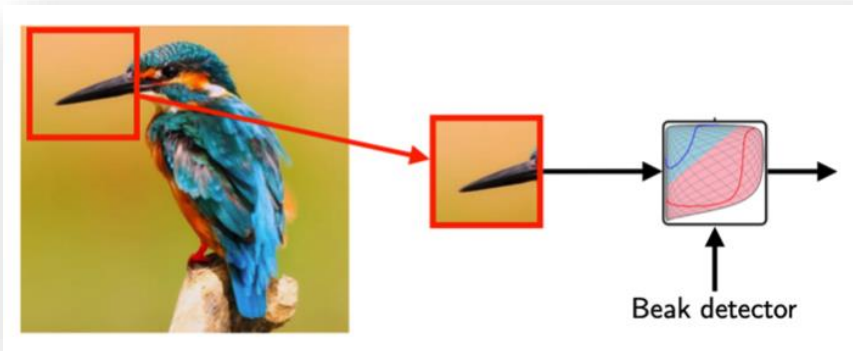
# 딥러닝

- 특징 추출부터 패턴까지 모든 과정을 사람 개입 없이 심층신경망 토대로 학습
  - 회귀: 연속 값 예측
  - 분류
    - 이진분류: 두 가지로 판단해야 할 때
    - 다중분류: 두 가지 이상으로 판단해야 할 때(여러 개 중 하나 구분)



# 모델

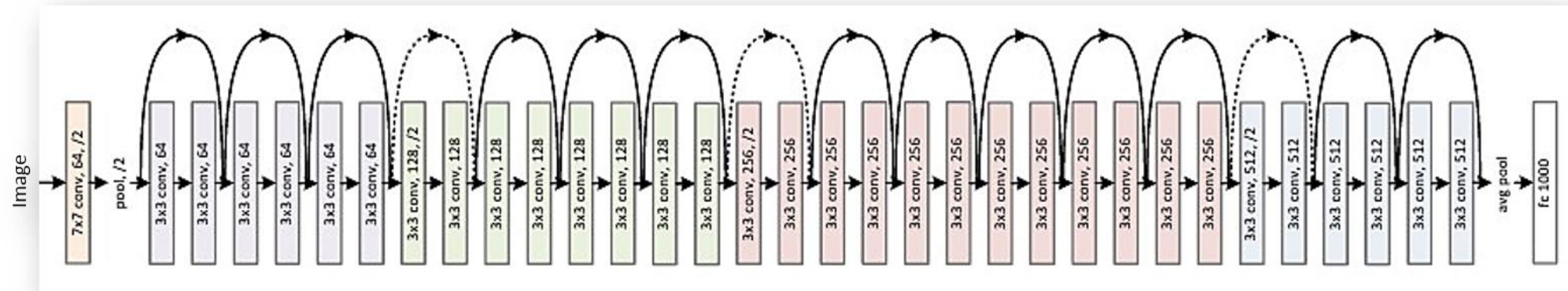
- **CNN**(Convolutional Neural Networks)
  - Convolution 전처리 작업 수행되는 Neural Network 모델
  - DNN(Deep Neural Network)의 문제점(이미지 공간/지역 정보 손실) 해결책
    - > 공간/지역 정보 유지한 채 feature 계층 빌드업
  - CNN 포인트: 영상 전체보단 부분, 영상의 한 픽셀과 주변 픽셀들의 연관성 살피기



# 모델

## • Resnet50

- ImageNet(ILSVRC) 2015 대회에서 우승한 모델
- 50개 레이어 있는 딥러닝 신경망 모델
- 이미지 분류, 이미지에서의 객체 감지 위한 백본 모델로 자주 사용
- 반복적인 곱셈은 그라디언트 극도로 작게 만들고 모델 학습 중지하게 함
  - > 입력이 활성화 계층으로 건너뛰기해 도달하는 연결로 그라디언트 소실 문제 해결



## 기대효과

- 내시경 영상만으로 조직 검사 전 암 기수 예측  
-> 검진 시간 단축
- 육안검사에서 누락 가능성 있는 경우 재확인 가능  
-> 병변 존재 누락되는 검진 감소시켜 진단 정확도 높임
- 위암 검진 효율성 향상 위한 보조도구로 활용

# 진행 계획

- 원천데이터 시각화
- 모델 학습 위한 데이터 전처리
- 알고리즘 선택, 모델 생성
- 훈련데이터로 모델 훈련  
-> 테스트 데이터에 적용
- 성능 평가

**감사합니다**

Contact:  
ods04193@gmail.com