Project

김 상현
(KIM SANGHYUN)

Contents

reCAPTCHA

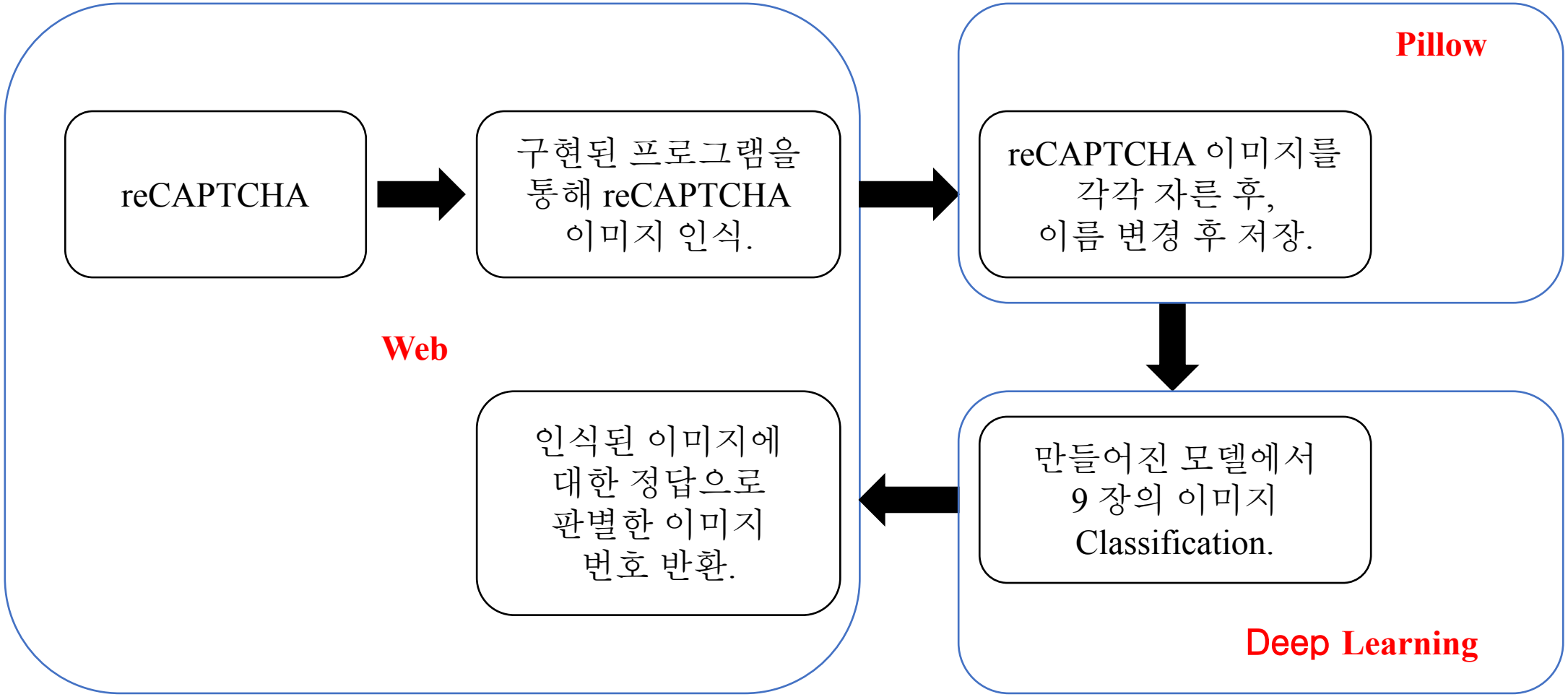
기간: 2019.10 ~ 2019.12

- **Goal**
  - **reCAPTCHA를 풀기위한 툴을 개발하고자 함.**
- **Main Contribution**
  - **pillow를 사용하여 reCAPTCHA 이미지 추출.**
  - **Deep Learning: DenseNet**

reCAPTCHA

- **Process**

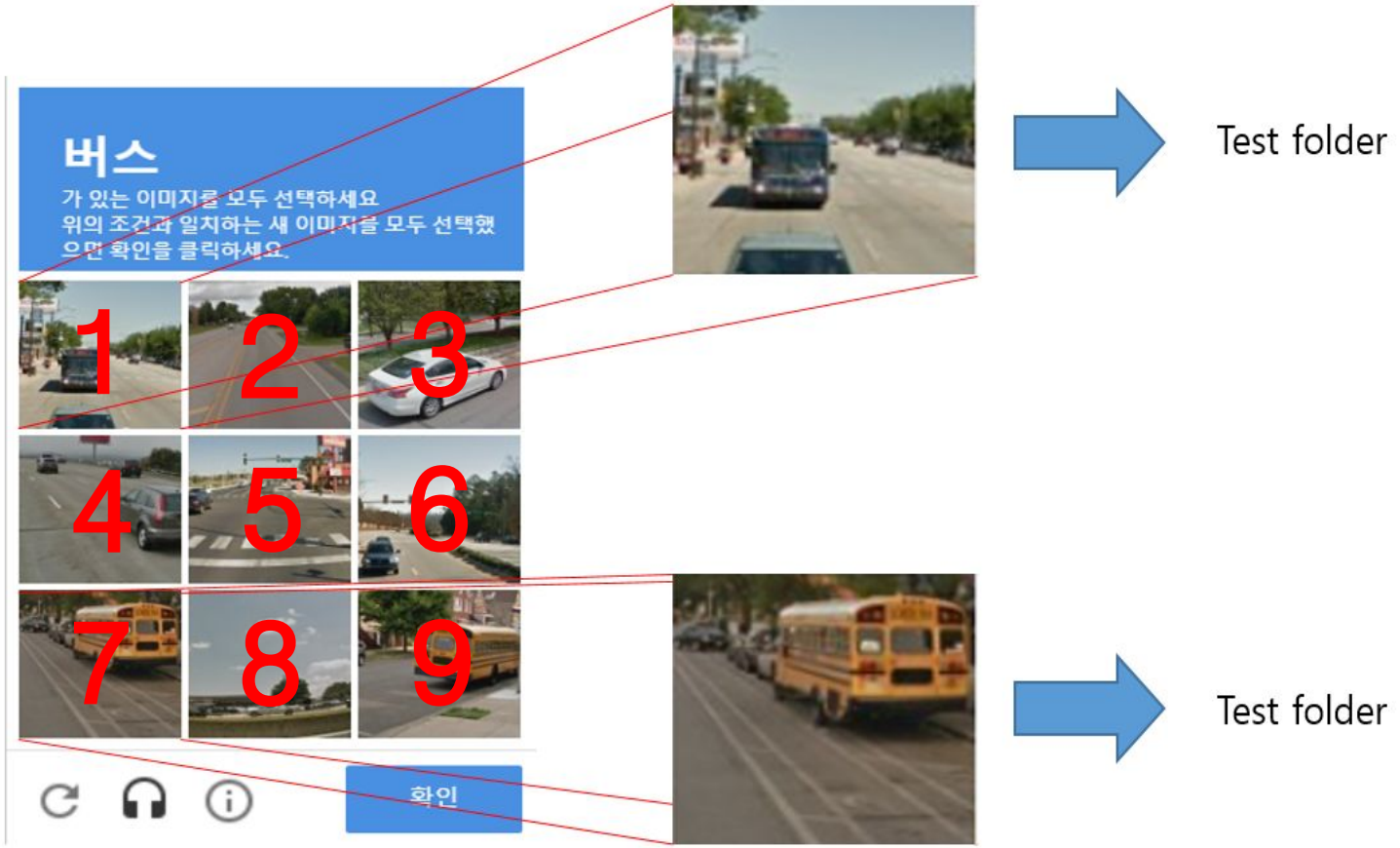
기간: 2019.10 ~ 2019.12



reCAPTCHA

기간: 2019.10 ~ 2019.12

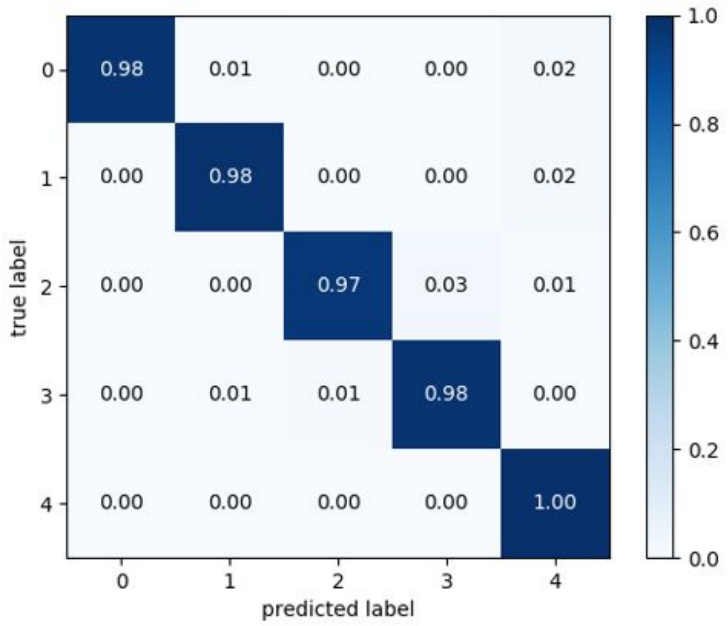
- **Pillow**
  - **Pillow를 사용하여 reCAPTCHA 이미지 추출**
  - **테스트 폴더로 이동 (이미지 이름 변경(숫자))**



reCAPTCHA

기간: 2019.10 ~ 2019.12

- **Deep Learning**
  - **DenseNet**
  - **Dataset (total 5 class)**
    - **train: 3715 (743 / 1 class)**
    - **test: 955 (191 / 1 class)**
  - **Acc: 99 %**



Bus : 0, car : 1, cat : 2, dog : 3, ship : 4

```
Prediction :  bus , Image_name :  9.jpg
Probability 0.84% => [bus]
Probability 0.07% => [ship]
Probability 0.06% => [car]
Probability 0.04% => [dog]
Probability 0.00% => [cat]
bus :  ['1.jpg', '6.jpg', '9.jpg']
car :  []
cat :  []
dog :  ['3.jpg', '4.jpg', '5.jpg']
ship :  ['2.jpg', '7.jpg', '8.jpg']
```

Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

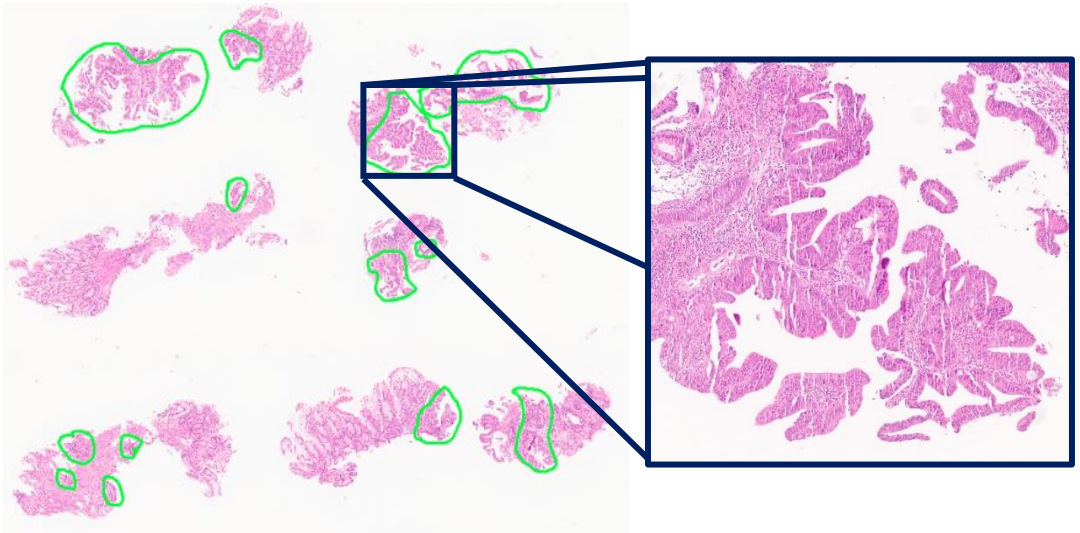
- **Goal**

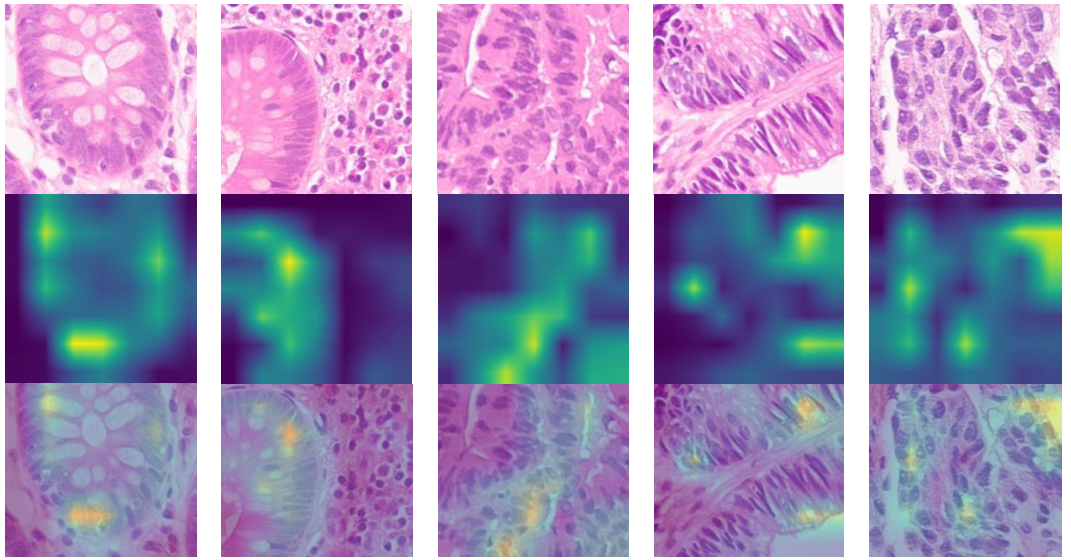
기간: 2019.06 ~ 2020.12

  - **AI를 활용해 대장암 판별 가능성 입증.**

- **Main Contribution**

  - **Normal, Adenoma, Adenocarcinoma(well, moderately, poorly)로 세분화하여 진단.**





Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

기간: 2019.06 ~ 2020.12

- **Problem**
  - 가공 되지 않은 데이터.
  - 오픈 되지 않은 데이터.

- **Solution**
  - **EDA 작업 진행.**

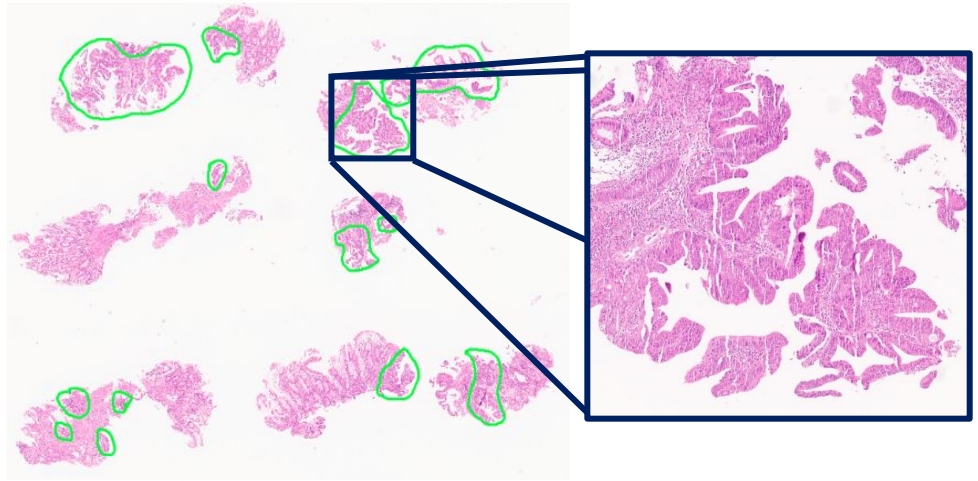
    - Data Histogram

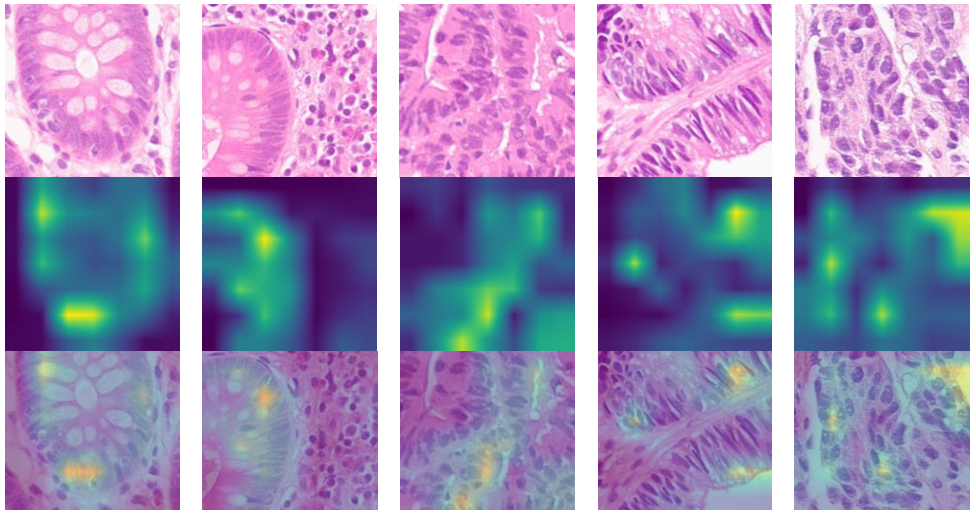
    - Class 분포.

  - **Data size.**

    - 400 배율, 400 x 400

  - **Class 균등화.**





Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

- **Dataset**

기간: 2019.06 ~ 2020.12

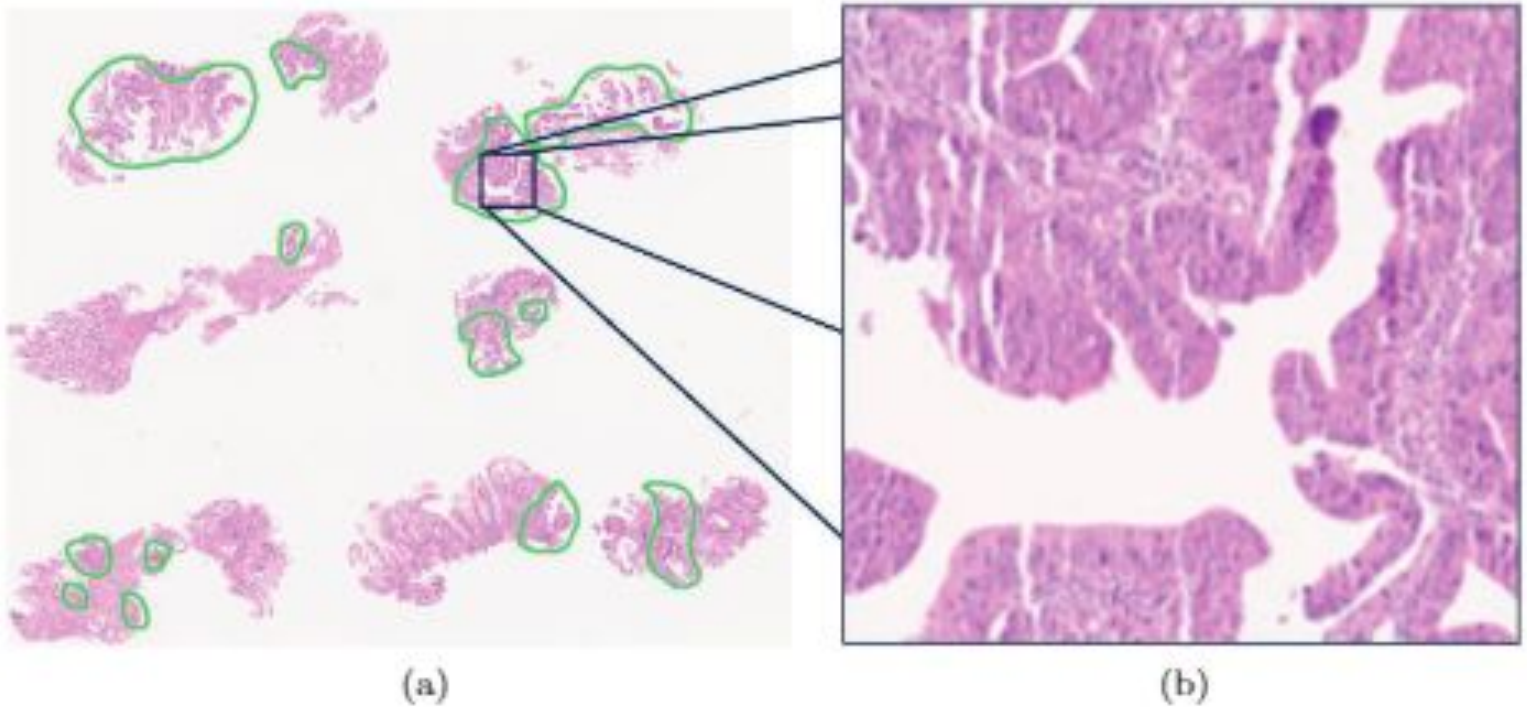


Fig. 2 (a) An example of an hematoxylin and eosin (H&E) images labeled by a pathologist and (b) a magnified image of data areas used to generate the training data.

Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

- **Dataset**

기간: 2019.06 ~ 2020.12

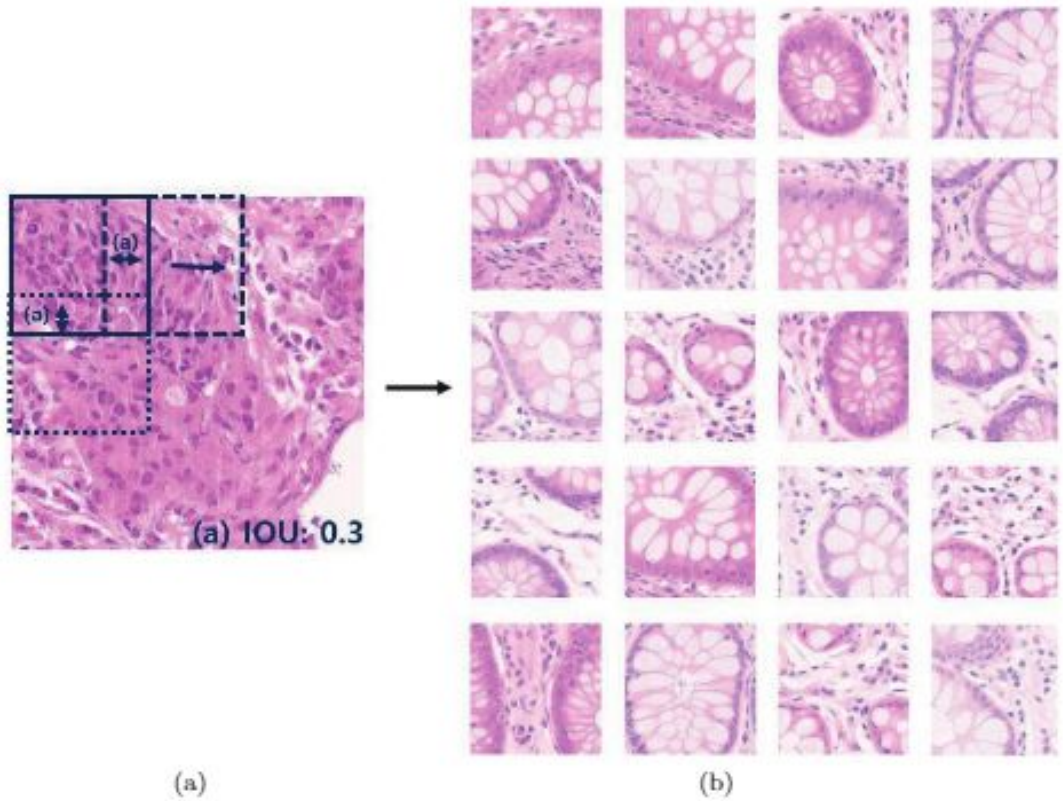


Fig. 3 (a) Window sliding-based training data patch generation process and (b) example of a training data patch extracted from the same data area.

Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

- **Dataset**

기간: 2019.06 ~ 2020.12

Table 1 Data Composition

	Normal	Adenoma	Adenocarcinoma			Total
			Well	Moderately	Poorly	
# of original slides	152	362	100	42	37	693
# of data areas	348	353	360	342	348	1751
# of training data patches	935	916	880	880	1000	4611
#of test data patches	121	121	111	109	109	571

Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

기간: 2019.06 ~ 2020.12

- **Hyper parameters**
  - **Optimizer: Adam**
  - **Learning Rate: 1e-3**
  - **Scheduler: StepLR**
  - **Loss: CrossEntropy**

Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

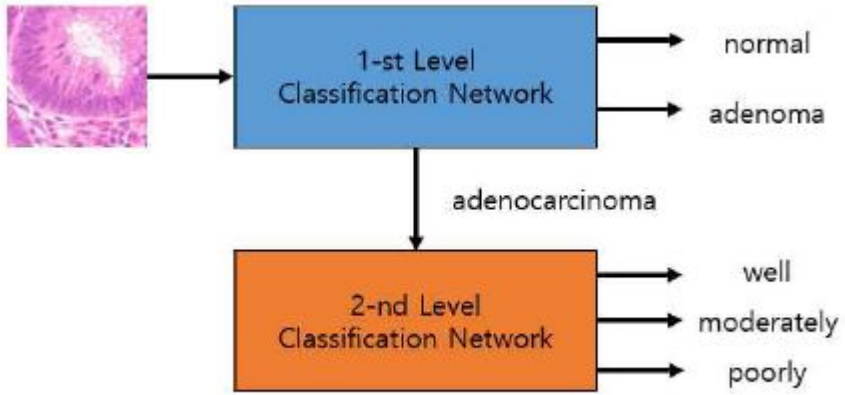
기간: 2019.06 ~ 2020.12

- **Another Experiment**
  - **Dataset**

(표 1) 데이터 세트

	정상	선종	선암		
			고분화형	중분화형	저분화형
원본	348	353	360	342	348
패치수	25,111	25,452	25,920	24,624	25,096

  - **2-stage classification**



(그림 2) 2단계 분류

"심층신경망을 이용한 조직검사 영상에서의 대장암 병변 분류," 2019 한국인터넷정보학회 추계학술발표대회, 2019.

Classification of Colorectal Cancer in Histological Images using Deep Neural Networks

- **Result**

기간: 2019.06 ~ 2020.12

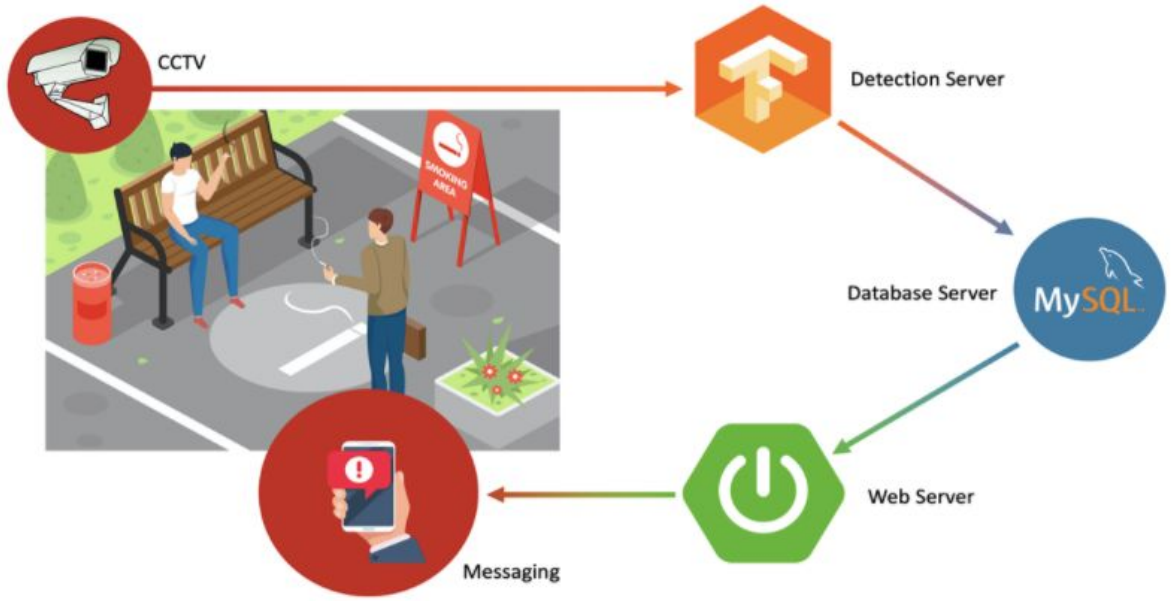
	mAP	Precision	Recall	Specificity	NVP
Res-net	**0.949**	**0.942**	**0.944**	**0.986**	**0.986**
Dense-net	0.936	0.927	0.925	0.981	0.981
Inception V3	0.922	0.914	0.914	0.978	0.978

"심층신경망을 이용한 조직검사 영상에서의 대장암 병변 분류," 2019 한국인터넷정보학회 추계학술발표대회, 2019.

Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

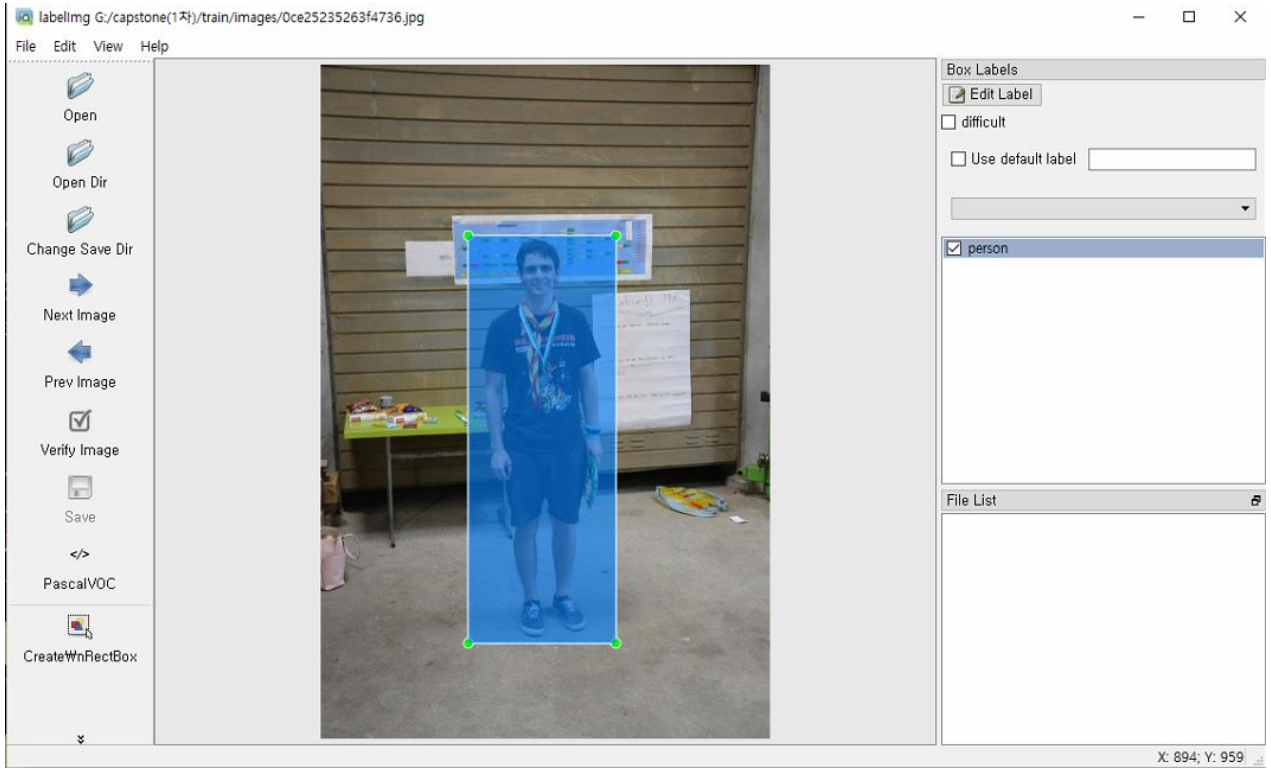
- **Goal**
  - 비 흡연 구역에서의 흡연행위를 적발하는 프로그램.
- **Problem**
  - 직접 촬영한 데이터
  - 담배(small object) 찾기
- **Solution**
  - **Multi scale training**
  - **Instaboost**



Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Dataset**
  - 사람
    - **COCO Dataset**
    - **Google AI Open Image**



	Train	Validation
Data 수	300	65

Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Dataset**
  - 담배
    - 촬영된 동영상에서 추출

	Train	Validation
Data 수	300	50

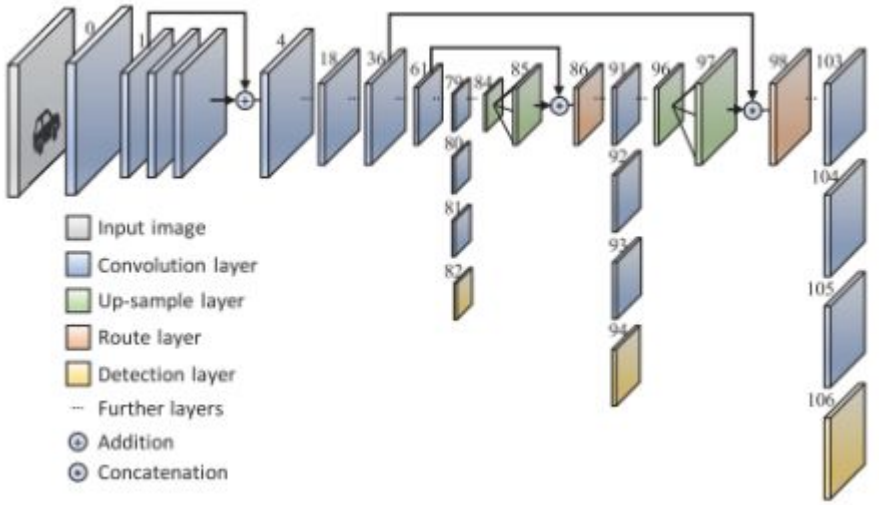
  - **Multi scale training**

  - **Instaboost**

Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Model**
  - **YOLO V3**
    - **속도가 빠르다고 알려져 있다.**
    - **다양한 scale로 학습을 하기 때문에 좀 더 일반화된 특징을 학습하여 다른 detection model들에 비해 높은 성능을 보여준다.**



Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Process**
  - **1. 동영상 촬영**

Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Process**
  - **2. 흡연구역 설정**

Detecting smoking outside the smoking area using object detection

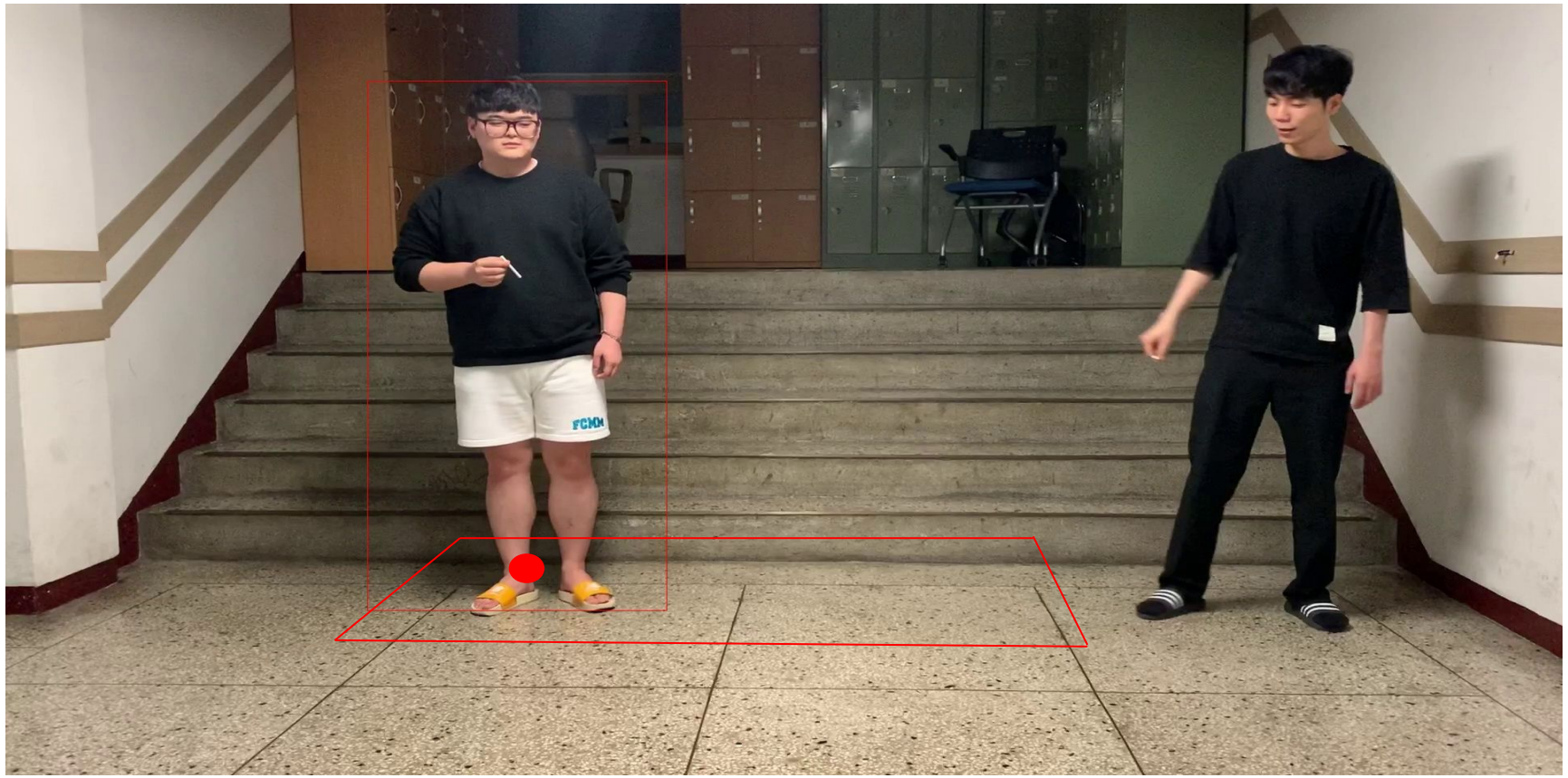
기간: 2020.03 ~ 2020.06

- **Process**
  - **3. 지정된 흡연 구역**

Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Process**
  - **4. 사람 탐지**
    - **기준: bounding 박스 하단의 중간 값**



Detecting smoking outside the smoking area using object detection

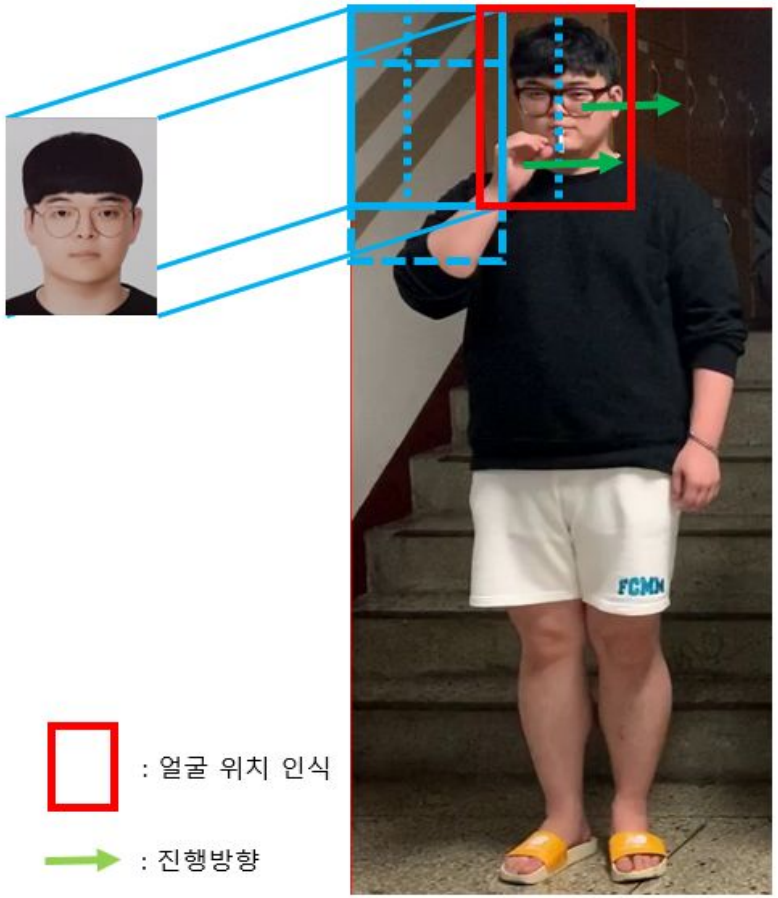
기간: 2020.03 ~ 2020.06

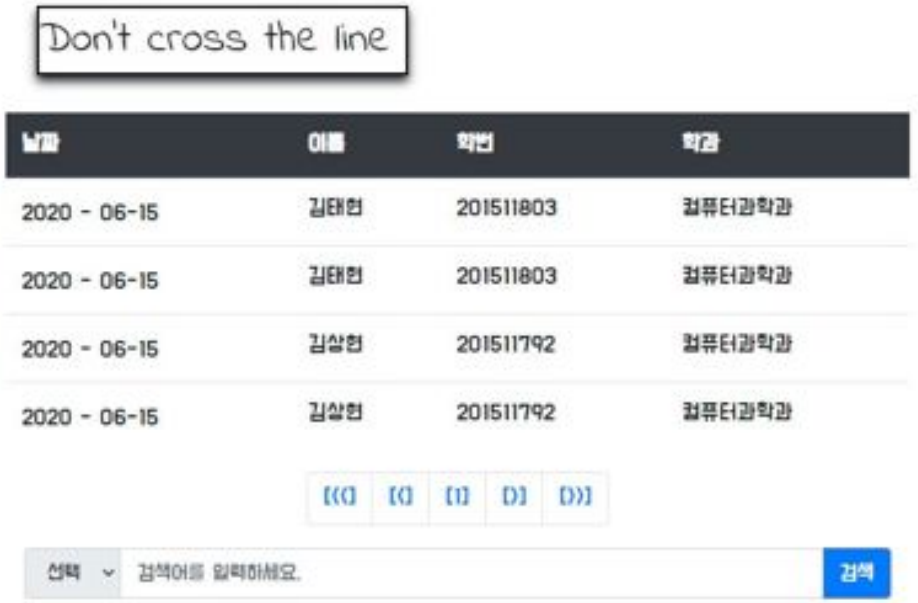
- **Process**
  - **5. 담배 탐지**

Detecting smoking outside the smoking area using object detection

기간: 2020.03 ~ 2020.06

- **Process**
  - **6. 흡연 구역 밖에서의 흡연자의 얼굴 인식. (sliding window 방식)**





- **Dataset**
  - **TACO Dataset**

	Battery	Clothing	Glass	Metal	Paper	Paperpack	Plastic	Plasticbag	Styrofoam
개수	566	1464	2533	3617	14939	3159	8218	15102	3380

Data Imbalance!!!

NAVER Boostcamp AI tech (Object Detection) (Team ranking 2)

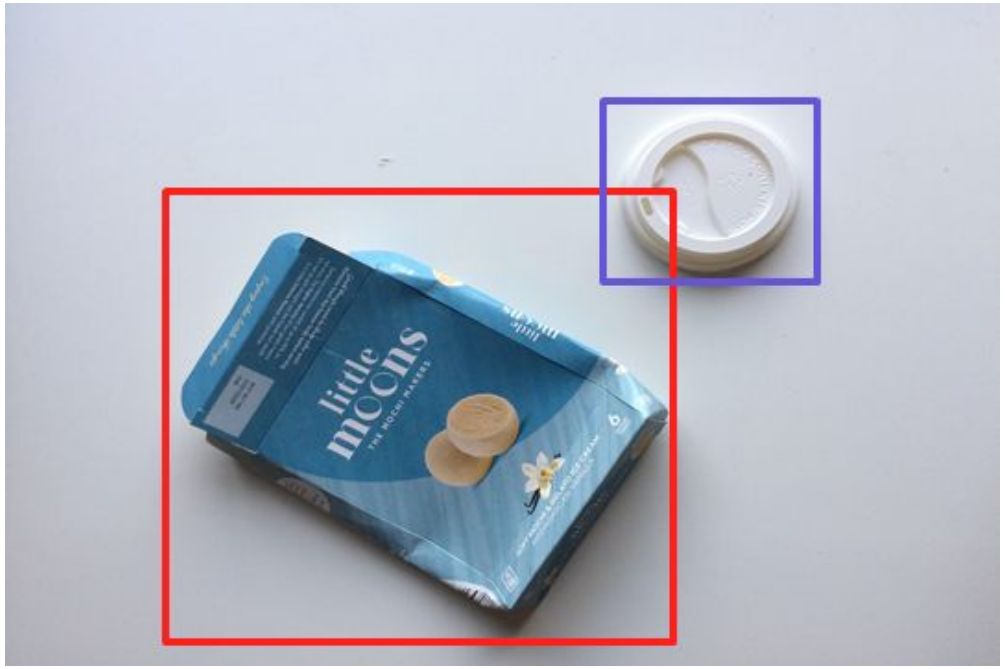
기간: 2021.05 ~ 2021.05 (2주)

- **Goal**
  - **쓰레기가 찍힌 사진에서 쓰레기를 Detection 하는 모델.**

- **Solution**
  - **DetectoRS + resnet50**
  - **InstaBoost**

- **Trial**
  - **Swin transformer**
  - **Mosaic**

- **Result**
  - **mAP: 0.6074**



NAVER Boostcamp AI tech (Object Detection) (Team ranking 2)

기간: 2021.05 ~ 2021.05 (2주)

- **Hyper parameters**
  - **Optimizer: AdamW**
  - **Learning Rate: 1e-4**
  - **Scheduler: StepLR**
  - **Loss: (0.4*F1) + (0.6*Focal)**
  - **Batch size: 8**
  - **Augmentation: Normalize, HorizontalFlip, Instaboost**

NAVER Boostcamp AI tech (Segmentation) (Team ranking 4)

기간: 2021.05 ~ 2021.05 (2주)

- **Goal**
  - **쓰레기가 찍힌 사진에서 쓰레기를 Segmentation 하는 모델.**
- **Solution**
  - **DeepLabV3+ + resnext101_32x16d**
  - **Copyblob**
  - **CutMix**
- **Result**
  - **mIoU: 0.6982**

NAVER Boostcamp AI tech (Segmentation) (Team ranking 4)

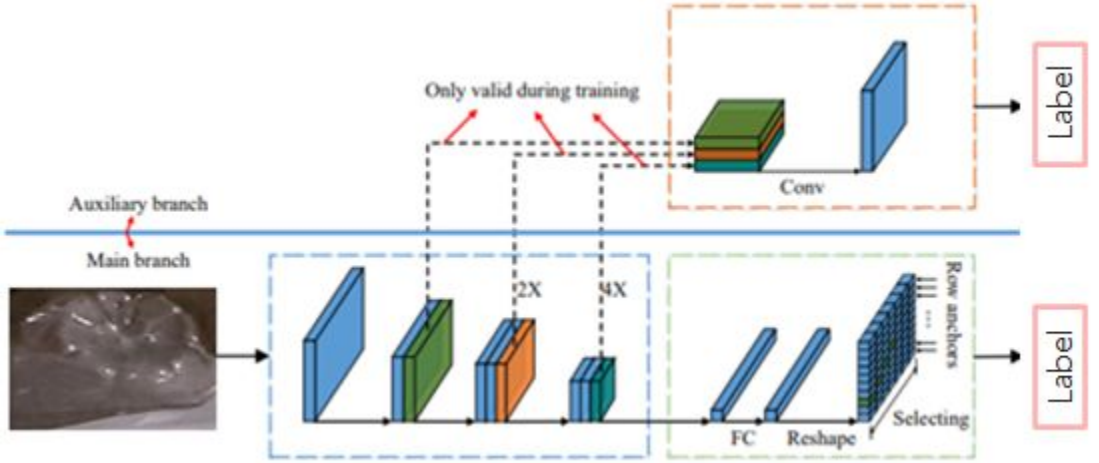
기간: 2021.05 ~ 2021.05 (2주)

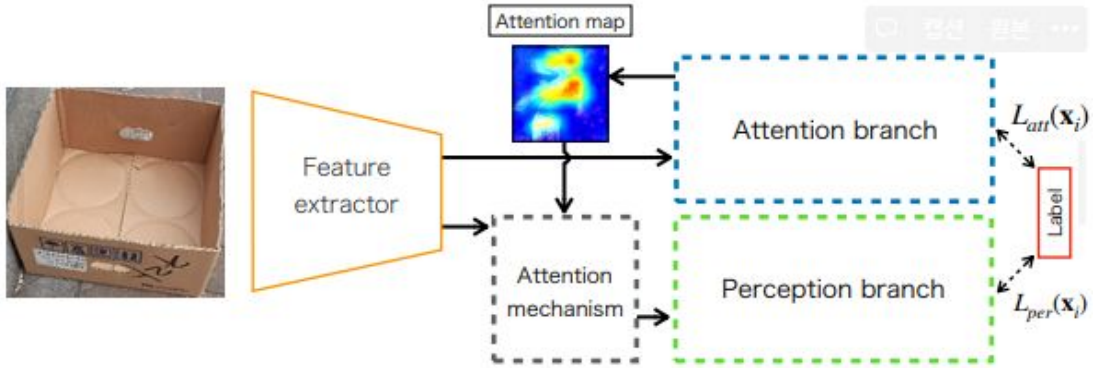
- **Hyper parameters**
  - **Optimizer: AdamW**
  - **Learning Rate: 1e-4**
  - **Scheduler: StepLR**
  - **Loss: (0.4*F1) + (0.6*Focal)**
  - **Batch size: 8**
  - **Augmentation: Normalize, HorizontalFlip, Rotate, RandomGridShuffle, OpticalDistortion, Cutout**

NAVER Boostcamp AI tech (Model compression) (Team ranking 2)

기간: 2021.06 ~ 2021.06

- **Goal**
  - **쓰레기 Classification model을 경량화**
- **Solution**
  - **ShuffleNet**
  - **Structured pruning**
- **Result**
  - **F1 score: 0.61**
  - **MACs: 약 100만**
- **Trial**
  - **Auxiliary training**
  - **Attention Branch Network**





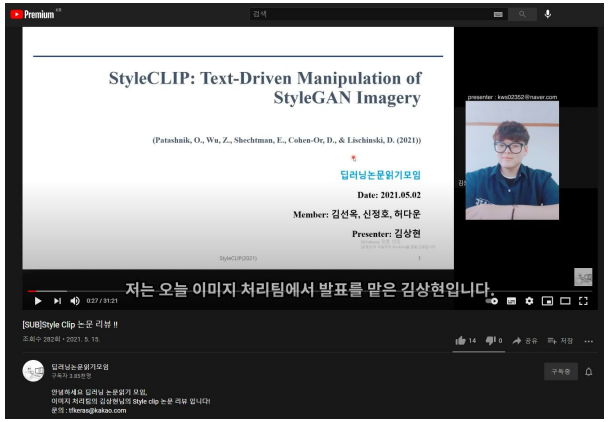
NAVER Boostcamp AI tech (Model compression) (Team ranking 2)

기간: 2021.06 ~ 2021.06

- **Hyper parameters**
  - **Optimizer: AdamP**
  - **Learning Rate: 1e-4**
  - **Scheduler: Cosineanneling**
  - **Loss: (0.4*F1) + (0.6*Focal)**
  - **Batch size: 64**
  - **Augmentation: RandAugmentation(“Identity”, “AutoContrast”, “Equalize”, “Rotate”, “Solarize”, “Color”, “Posterize”, “Contrast”, “Brightness”, “Sharpness”, “ShearX”, “ShearY”, “TranslateX”, “TranslateY”), Normalize, Resize, Cutout, Normalize**

Additional Activities

- **Youtube ‘딥러닝논문읽기모임’ 이미지 처리 팀 3기 멤버**
  - **Mar 2021 - Dec 2021 (10 months)**



- **NAVER Boostcamp AI Tech 1기 camper**
  - **Jan 2021 - Jun 2021 (6 months)**

- **서울아산병원 마취통증의학과 AI 파트 담당 연구원**
  - **Contract worker.**
  - **Jul 2021 - Jan 2022 (6 months)**

서울아산병원
Asan Medical Center

THANK YOU