



# Upstage AI Lab

Computer Vision Seminar | 2025. 07. 11(금)

## 목차

- 01. 팀 소개
- 02. 경진대회 수행 절차 및 방법
- 03. 분석 인사이트 및 결과
- 04. 회고

01

# 팀 소개

---

팀장/팀원 소개  
협업 방식



[5조 - 티끌 모아 Tech]    각자의 작은 기술이 모여 힘을 발휘한다!



팀장

송규현

RAG / 경영정보  
EDA, 베이스라인 작성,  
2-stage 모델 실험



팀원

이상현

AI응용분야찾기/재료공학  
데이터 특성추출  
및 대책, EDA



팀원

이영준

MLOps / 컴퓨터공학  
자동화 구축 후 다양한 모델,  
옵션 조합 실험



팀원

조은별

고분자공학과  
마스킹 탐지 및 통계 기반 분석,  
파라미터 최적화, 스케줄러 실험



팀원

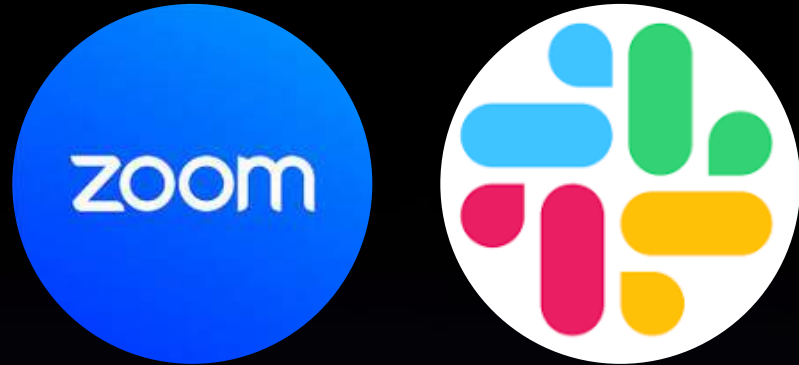
편아현

소프트웨어학과  
mixup 증강 코드 구현,  
오프라인 증강, Grad-CAM 시각화

# 경진대회 협업 방식

: Computer Vision [대회] Classification

회의 및 아이디어 공유



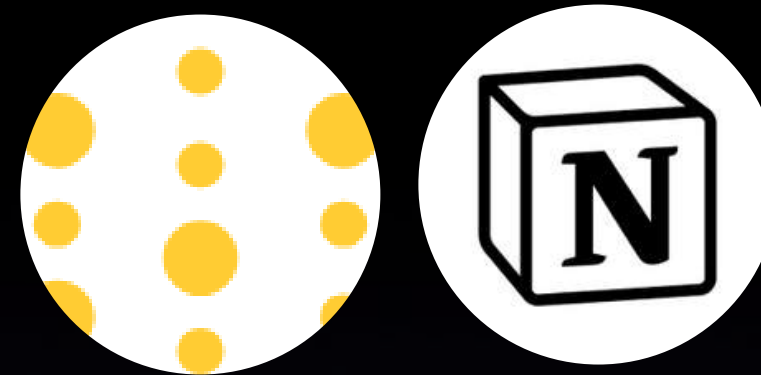
Zoom & Slack

TO-DO & 문서화



Github Issue & docs/

실험 정리



WandB & Notion

코드 작성



Claude Code



Gemini CLI



ChatGPT



02

## 경진대회 수행 절차 및 방법

---

목표 수립  
수행 내용 / 수행 결과

# 경진대회 목표 수립

: Computer Vision [대회] Classification

주제

Document Type Classification | 문서 타입 분류

문서는 금융, 보험, 물류, 의료 등 도메인을 가리지 않고 많이 취급됩니다. 이 대회는 다양한 종류의 문서 이미지의 클래스를 예측합니다.

목표

목표

- 데이터증강 - 모델 파인튜닝 파이프라인을 경험하기 [o]
- 다양한 이미지 증강 기법 실험하기 [o]
- 신뢰 가능한 Validation 전략 찾아내기 [o]
- 베이스라인 코드를 바탕으로 체계적인 파라미터 실험 [△]
- 목표 점수: F1-score 0.95 이상 [x]

개요

소개 및 배경 설명

17 종류의 각종 문서, 주민등록증, 여권, 운전면허증, 자동차 번호판, 계  
기판을 분류하는 과제

기간

2025. 06. 30 ~ 2025. 07.10

# 경진대회 수행 내용

## : Computer Vision [대회] Classification

1

### 개발 환경 구축

- UV로 빠르게 환경 구축
- python == 3.10
- torch == 2.5.0 (+cu12.1)
- timm
- albumentations
- wandb

2

### 데이터 분석

- EDA : train, test 이미지 분석
    - 밝기, 대비, RGB
    - Blur
    - Noise
    - Aspect
    - Rotate
    - 기타...
- > docs/Dataset EDA.md

3

### 데이터 증강

- 'eda' : EDA 기반 증강
- 'dilation', 'erosion'
- 'easiest', 'stilleasy' : 간단한 rotate, flip 위주의 증강
- 'basic' : rotate, flip, noise, blur, RGB
- 'middle' : Perspective, GridDistortion 등 강도를 높인 증강
- 'aggressive' : CoarseDropout 추가

4

### 모델 선택 학습 및 평가

- Transfer Learning : timm 사전학습 모델을 사용하여 full-finetuning
- Custom Classifier Head : timm 사전학습 모델을 backbone으로 사용하고 head는 따로 구성하여 예측
- 2-Stage Classifier : 어려운 클래스의 문서를 별도의 모델로 예측하는 구조

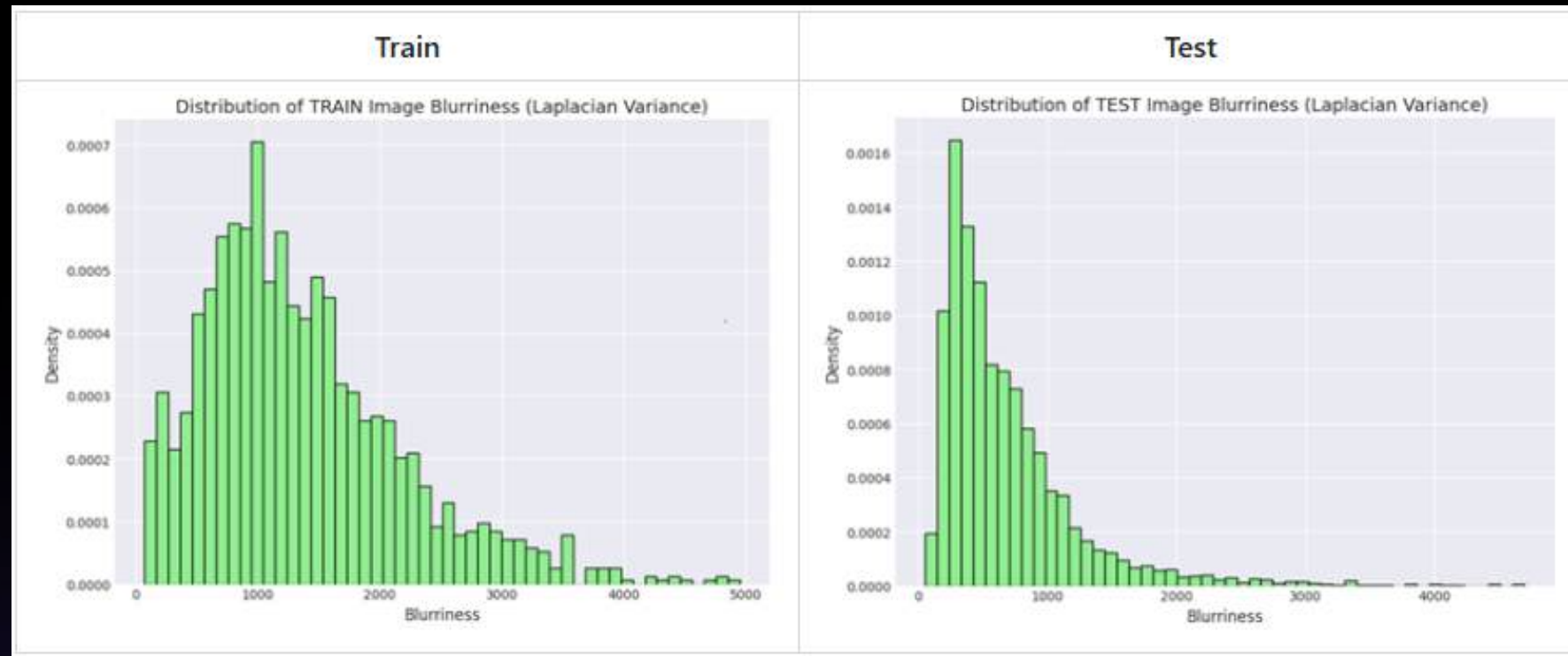


# 경진대회 수행 결과

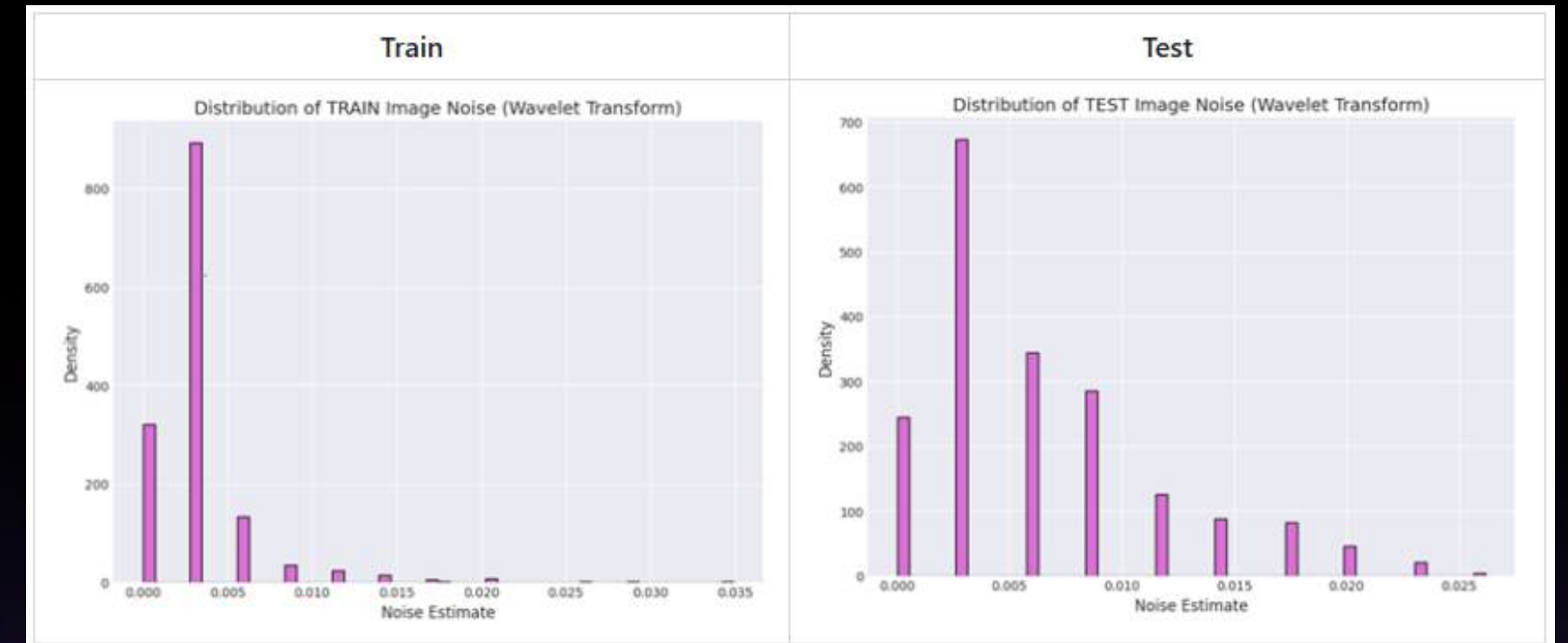
: Computer Vision [대회] Classification

## [데이터 분석] - Train & Test 이미지 EDA

- 밝기, 대비, RGB 분포에는 큰 차이가 없었으나 Blurriness, Noise, Rotation에는 큰 차이가 있었다.
- docs/Dataset EDA.md 에 구체적인 EDA 문서 작성.



<Blurriness>



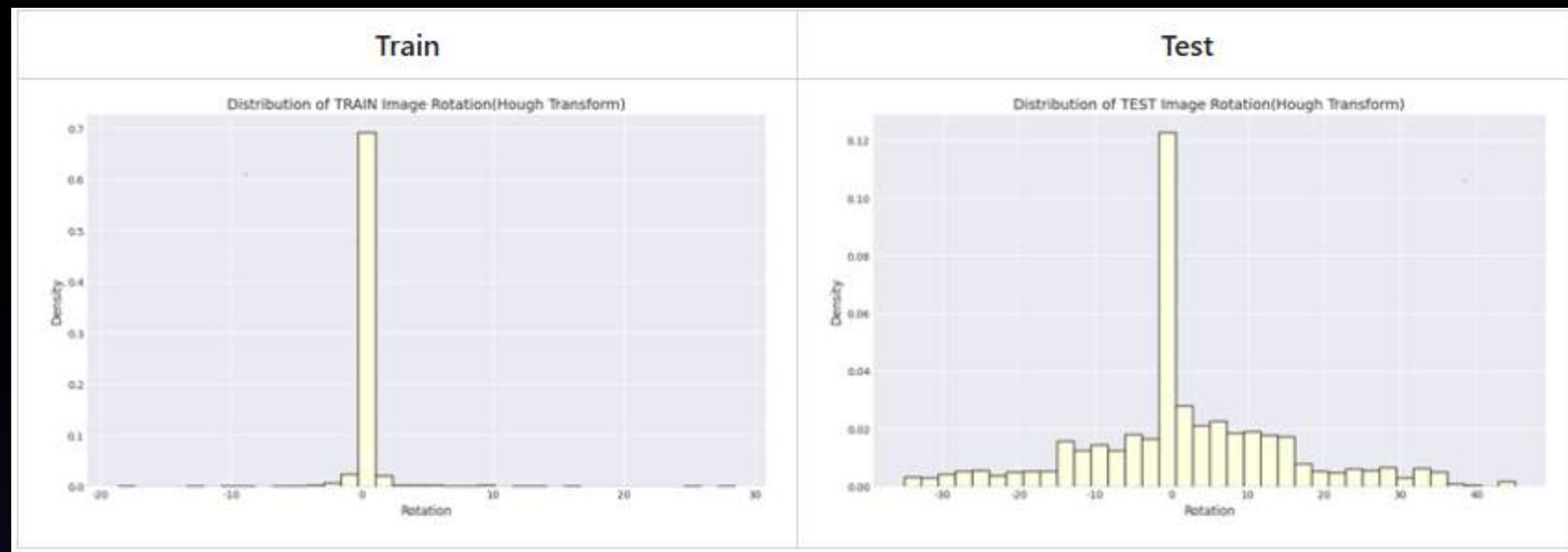
<Noise Estimation>

# 경진대회 수행 결과

: Computer Vision [대회] Classification

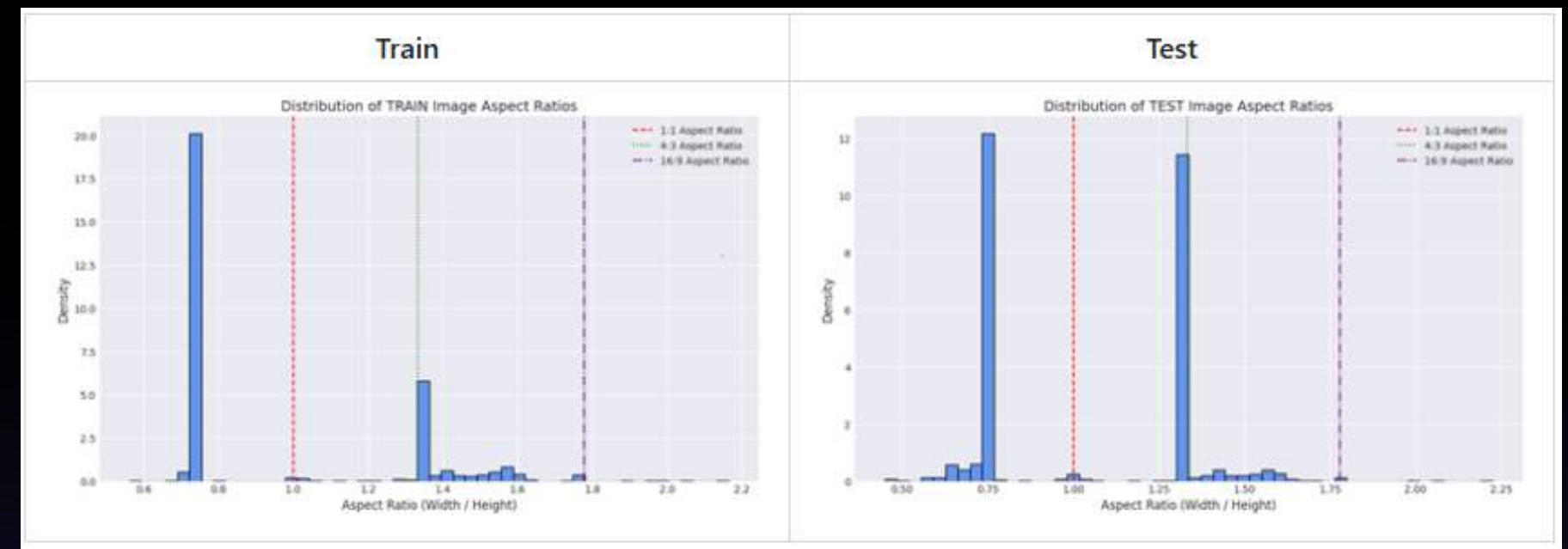
## [데이터 분석] - Train & Test 이미지 EDA

- 밝기, 대비, RGB 분포에는 큰 차이가 없었으나 Blurriness, Noise, Rotation에는 큰 차이가 있었다.
- docs/Dataset EDA.md 에 구체적인 EDA 문서 작성.



### <Rotation Estimate>

- Flip된 이미지가 다수이기 때문에 Test Rotation Estimate의 범위가 넓지 않다.



### <Aspect>

- Test 데이터셋에 적용된 증강의 대부분이 fill=(255,255,255)임을 예측 가능
- Rotate, Flip의 영향으로 위와 같은 aspect 분포 차이를 보이는 것 같고, 원본 이미지들의 aspect ratio는 train과 test가 유사한 것으로 보인다.

# 경진대회 수행 결과

## : Computer Vision [대회] Classification

### [데이터 분석] - 직접 눈으로 확인

#### [Train]

- 운전면허증, 주민등록증, 여권 : 증명사진 마스킹이 눈에 띄고, 유사한 종횡비를 갖는다.
- 의료 관련 문서 중 [3, 4, 7, 14]의 형태와 내용이 유사하여, 클래스의 분류가 어려워 보임.
  - 같은 클래스 내에서도, 문서의 제목이 상이한 데이터가 다수  
e.g. 3 : 입원 외래 진료확인서, 입원 진료확인서, 퇴원요약지, 입원요약지, 입원사실증명서, 입원퇴원증명서, ...  
7 : 진료확인서, 통원확인서, 통원 진료확인서, 통원증명서, ...
  - 서로 다른 클래스에 있어야 하는 문서가 존재 → **5개 데이터의 클래스를 직접 수정하여 이후 프로젝트 진행 (Public은 점수가 낮아졌지만, Final에서 수정한 버전의 점수가 더 높았다.)**
    - 7(진료확인서)에 '입(퇴)원 확인서' 존재
    - 4(진단서)에 '진료소견서'가 존재

#### [Test]

- 약 70% 정도의 이미지가 Flip 되어 글자가 반전됨.
- 약 50% 이상의 이미지에 Noise 및 Blur가 적용됨.



# 경진대회 수행 결과

: Computer Vision [대회] Classification

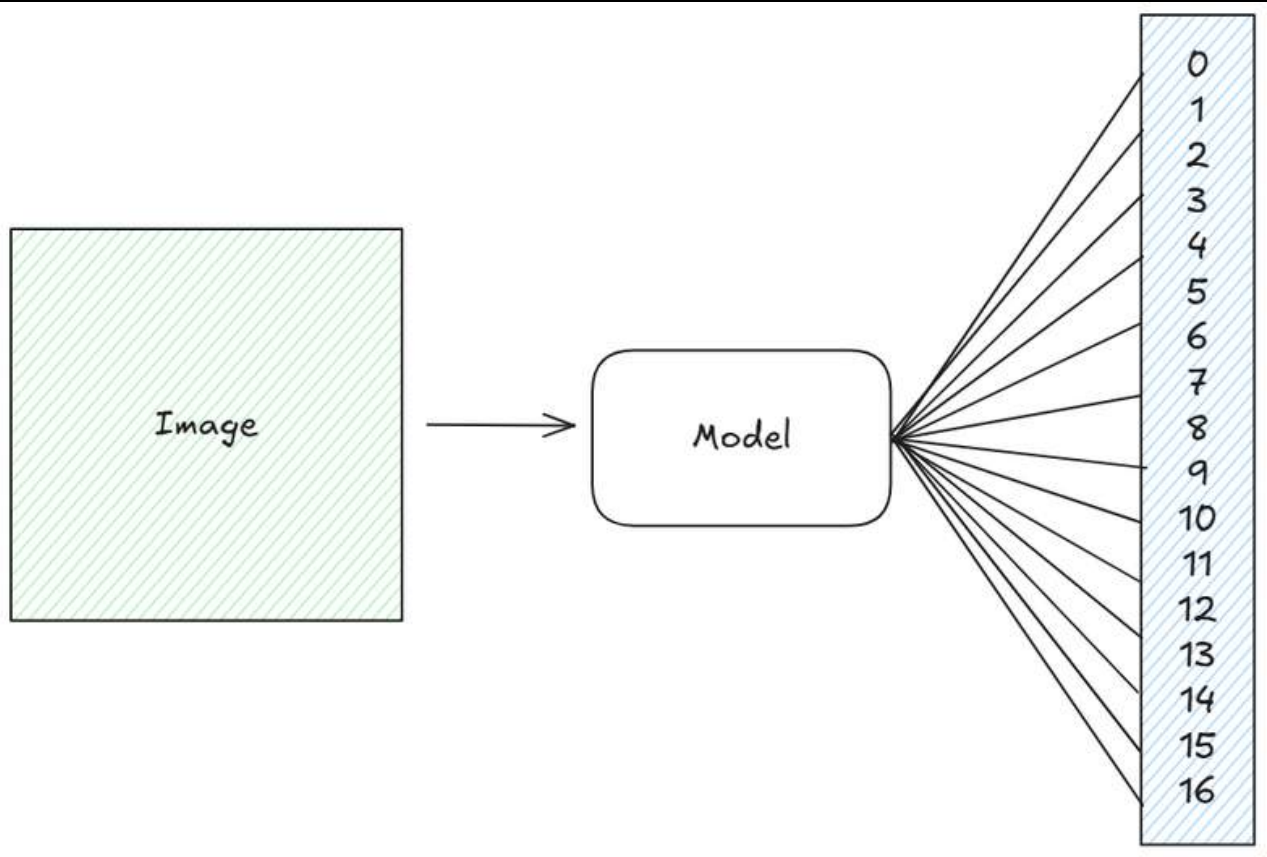
## [데이터 증강]

```
1  'eda': A.Compose([
2      # Brightness, Contrast, ColorJitter
3      A.ColorJitter(brightness=0.1, contrast=0.07, saturation=0.07, hue=0.07, p=1.0),
4      # 공간 변형에 대한 증강
5      A.Affine(
6          translate_percent=(-0.05,0.05),
7          rotate=(-20,30),
8          fill=(255,255,255),
9          p=1.0
10     ),
11     # x,y 좌표 반전
12     A.OneOf([
13         A.Compose([
14             A.HorizontalFlip(p=1.0),
15             A.VerticalFlip(p=1.0),
16         ]),
17         A.Transpose(p=1),
18     ], p=0.8),
19     # Blur & Noise
20     A.OneOf([
21         A.GaussianBlur(sigma_limit=(0.5, 2.5), p=1.0),
22         A.Blur(blur_limit=(3, 9), p=1.0),
23     ], p=1.0),
24     A.GaussNoise(std_range=(0.0025, 0.2), p=1.0),
25 ])
```

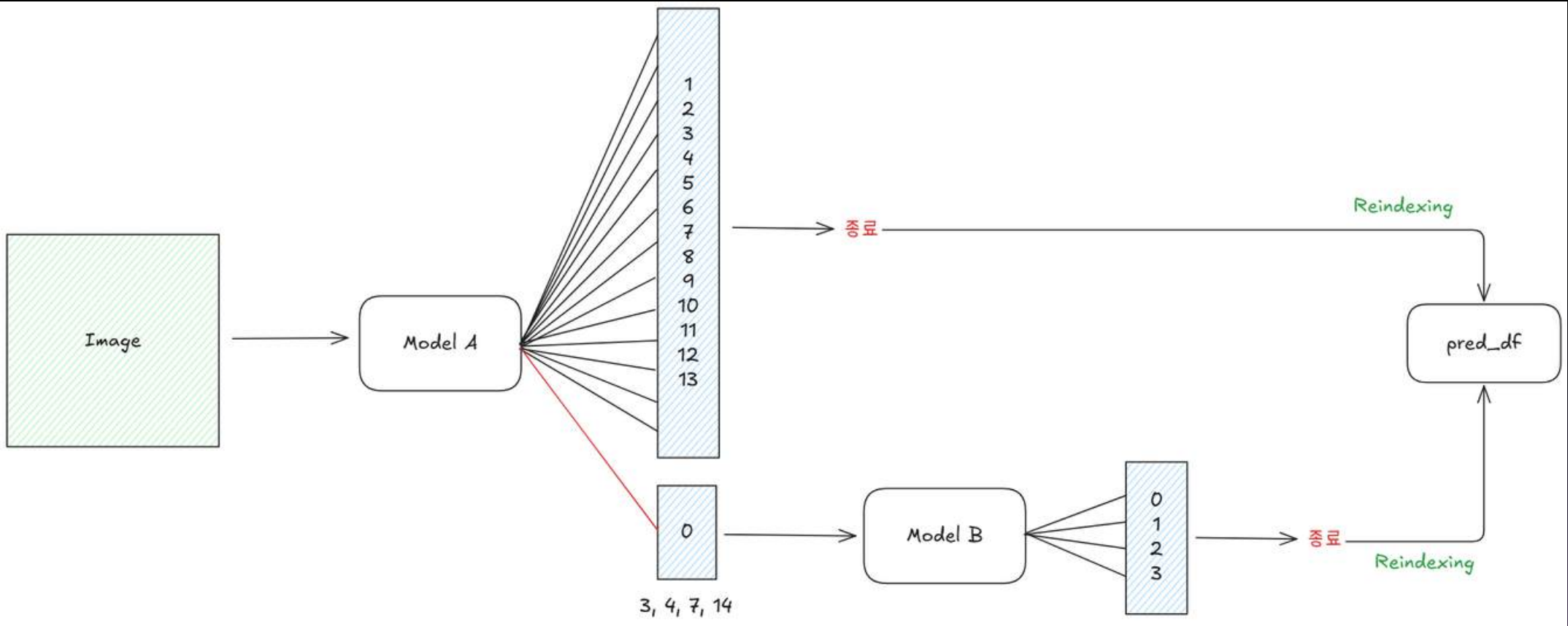
- **eda** : EDA를 기반으로 test 데이터셋과 비슷한 이미지를 만드는 증강
- dilation, erosion : 경계를 흐릿하게, 강조하는 증강
- **easiest** : Rotate, Flip만 적용한 증강
- **stilleasy** : Shift, Rotate, Flip 적용
- **basic** : 밝기, 대비, RGBshift, Shift, Rotate, Flip
- **middle** : Noise, Perspective / GridDistortion, Shift, Rotate, Flip
- **aggressive** : CoarseDropout, Noise, Perspective / GridDistortion, Shift, Rotate, Flip, 밝기, 대비, RGBshift

[모델링]

단일 모델 전략, Custom Classifier Head 전략



2-Stage Classifier 전략





03

## 분석 인사이트 및 결과

---

문제 및 인사이트 도출  
해결 방법 및 결과



# 경진대회 인사이트 공유

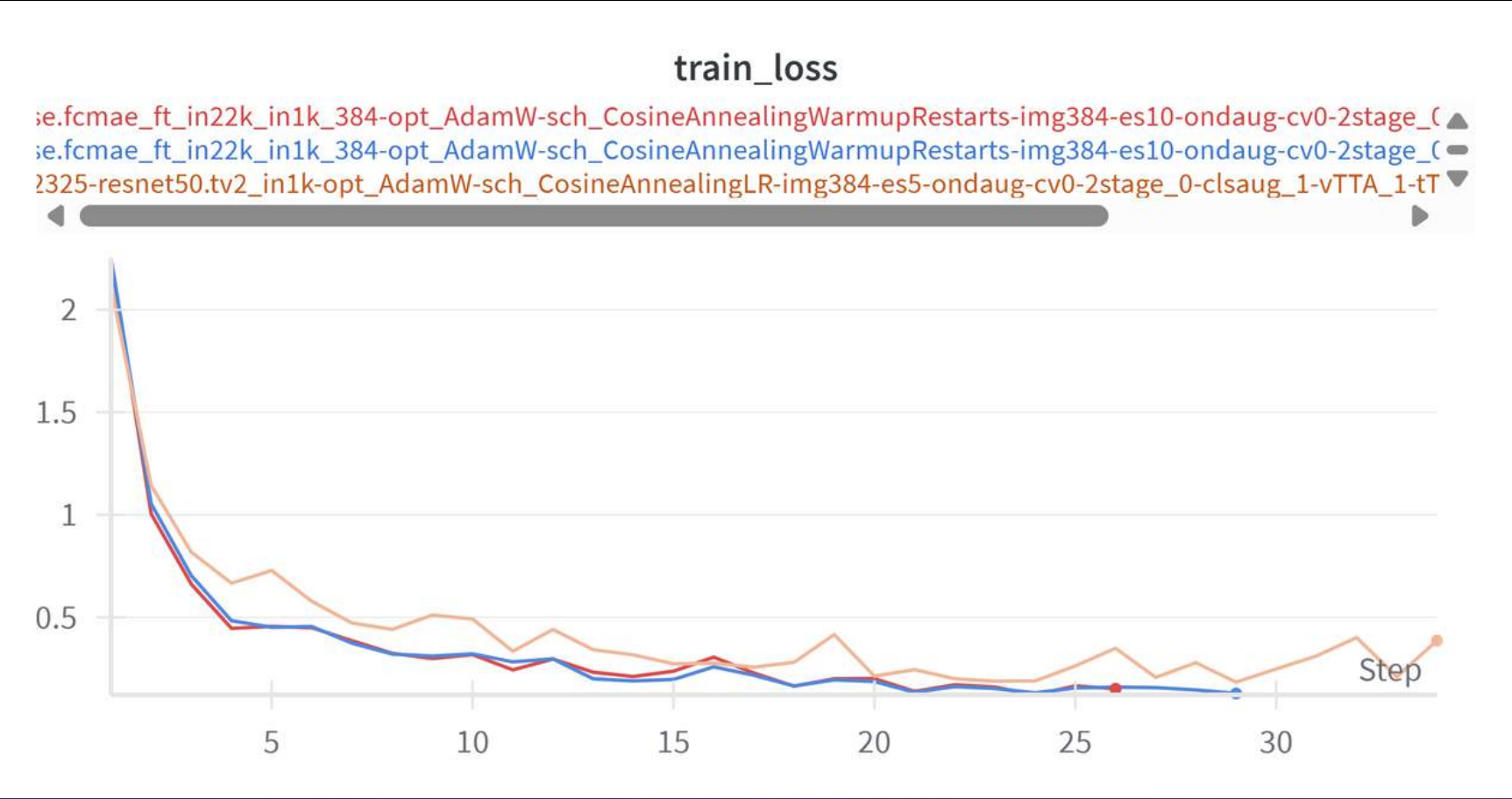
: Computer Vision [대회] Classification

## [데이터 증강]

- 문제 1 : train 데이터셋의 개수가 작다.
  - 문제 2 : [1, 13, 14] 클래스의 데이터 개수가 작다.
- 
- 해결 : 데이터 증강

### Online VS Offline 증강의 장단점

	Online	Offline
다양성	↑	↓
재현성	↓	↑
작은 데이터셋	↓	↑



# 경진대회 인사이트 공유

## : Computer Vision [대회] Classification

### [데이터 증강]

- 다중 Online 증강 : 여러 증강들을 A.OneOf 로 묶어 매 epoch마다 서로 다른 증강 기법을 사용해 Online 증강을 수행한다.
- Dynamic Online 증강 : epoch에 따라 서로 다른 증강 기법을 사용해 Online 증강을 수행한다.
  - 처음에는 약한 강도의 증강을 적용하다가 점차 강한 강도의 증강으로 변화시킨다.
- Offline 증강 : train 데이터셋을 미리 설정한 증강 기법들로 각 클래스별로 1000개까지 증강하여 활용한다.
  - 클래스 불균형 문제가 자연스럽게 해소된다.
- **Offline 증강 + Mixup Online 증강**
  - Offline 증강을 통해 클래스 불균형을 해고
  - Mixup 증강을 통해 결정 경계 강화하여 어려운 클래스에 대한 예측 성능을 높임.



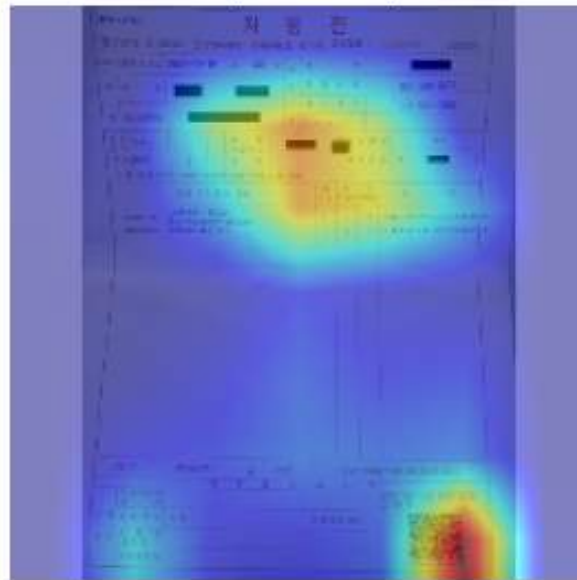
# 경진대회 인사이트 공유

: Computer Vision [대회] Classification

## [Grad-CAM]

- 정보가 마스킹된 부분에 모델이 과도하게 의존할까 걱정했는데, 훈련된 모델이 입력 이미지의 형태 자체를 기반으로 분류를 수행한다는 깨달았다.

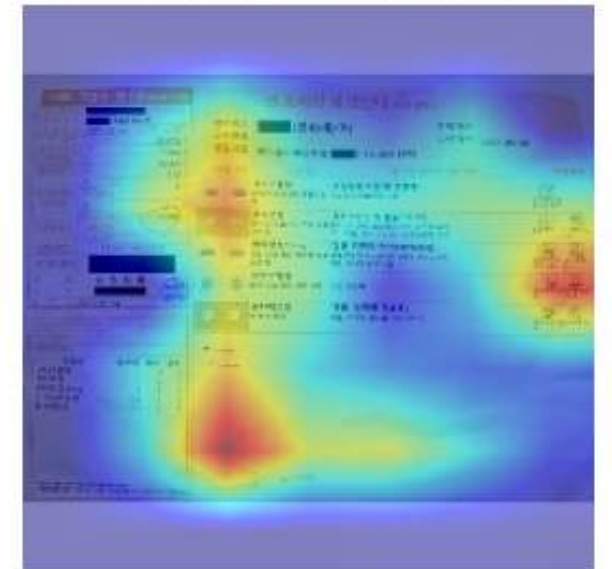
Grad-CAM | GT: 12 | Pred: 12



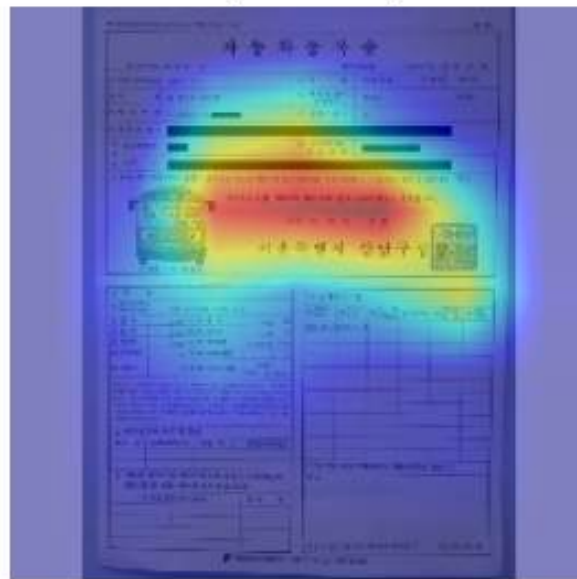
Grad-CAM | GT: 7 | Pred: 3



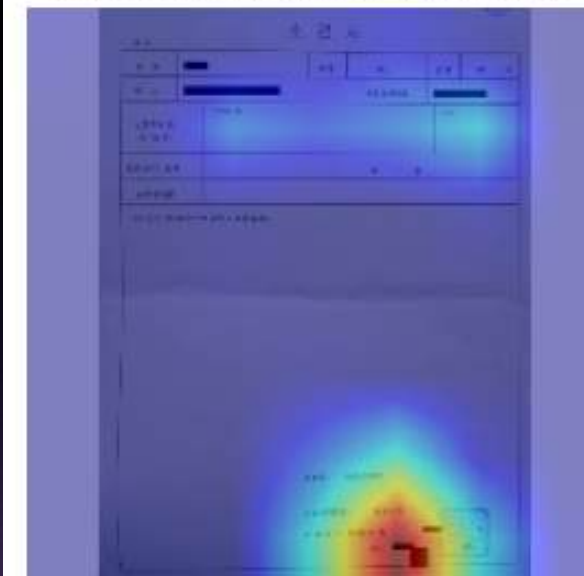
Grad-CAM | GT: 11 | Pred: 11



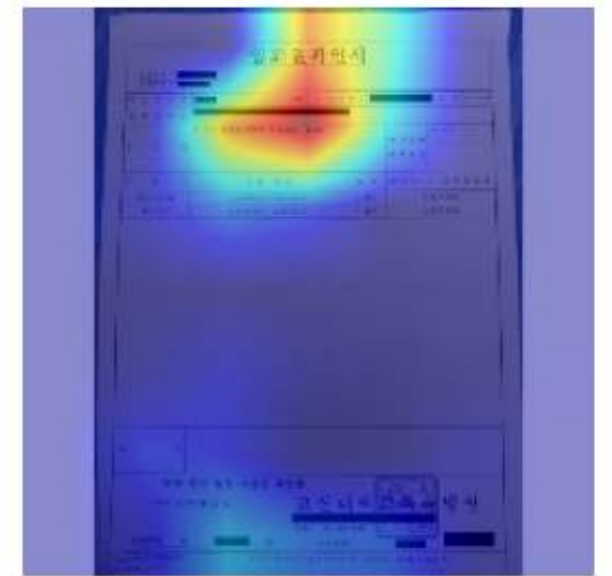
Grad-CAM | GT: 15 | Pred: 15



Grad-CAM | GT: 14 | Pred: 14



Grad-CAM | GT: 3 | Pred: 3





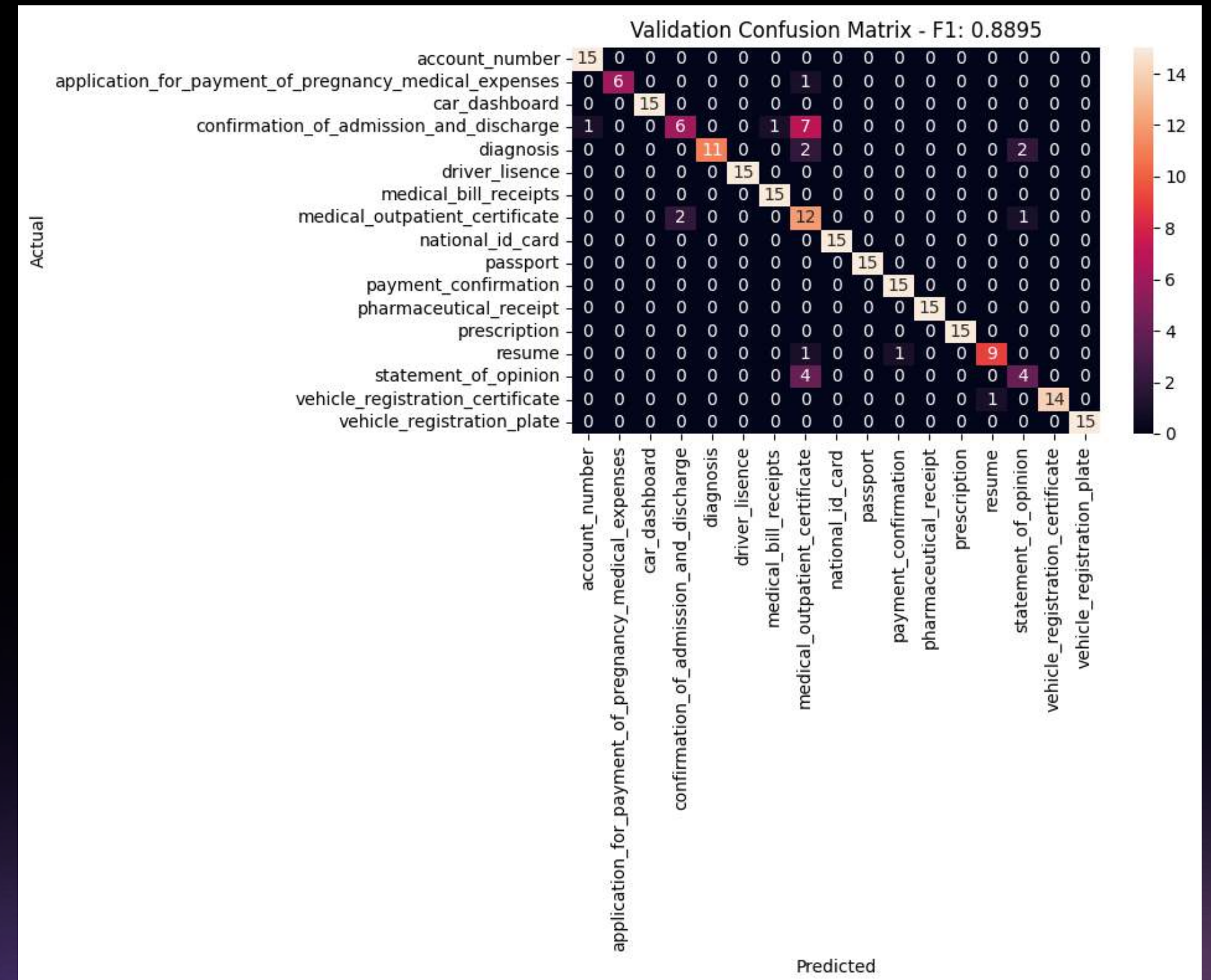
# 경진대회 인사이트 공유

## : Computer Vision [대회] Classification

### [Validation]

- 문제점 : validation data와 test data 간 괴리가 크기 때문에 validation f1-score와 실제 submission score에 차이가 컸다.
- 해결 : validation 시 TTA를 통해 validation 이미지를 test 데이터 셋과 유사하게 만들어 평가했다.
  - 해당 버전의 val-score = 0.8895,  
해당 버전의 제출 점수 = 0.9021(Public), 0.8737(Final)

+ ) Confusion Matrix 인사이트 : [3, 4, 7, 14] 클래스 간 혼동이 크다.  
→ 이후 모델링에 반영한다.

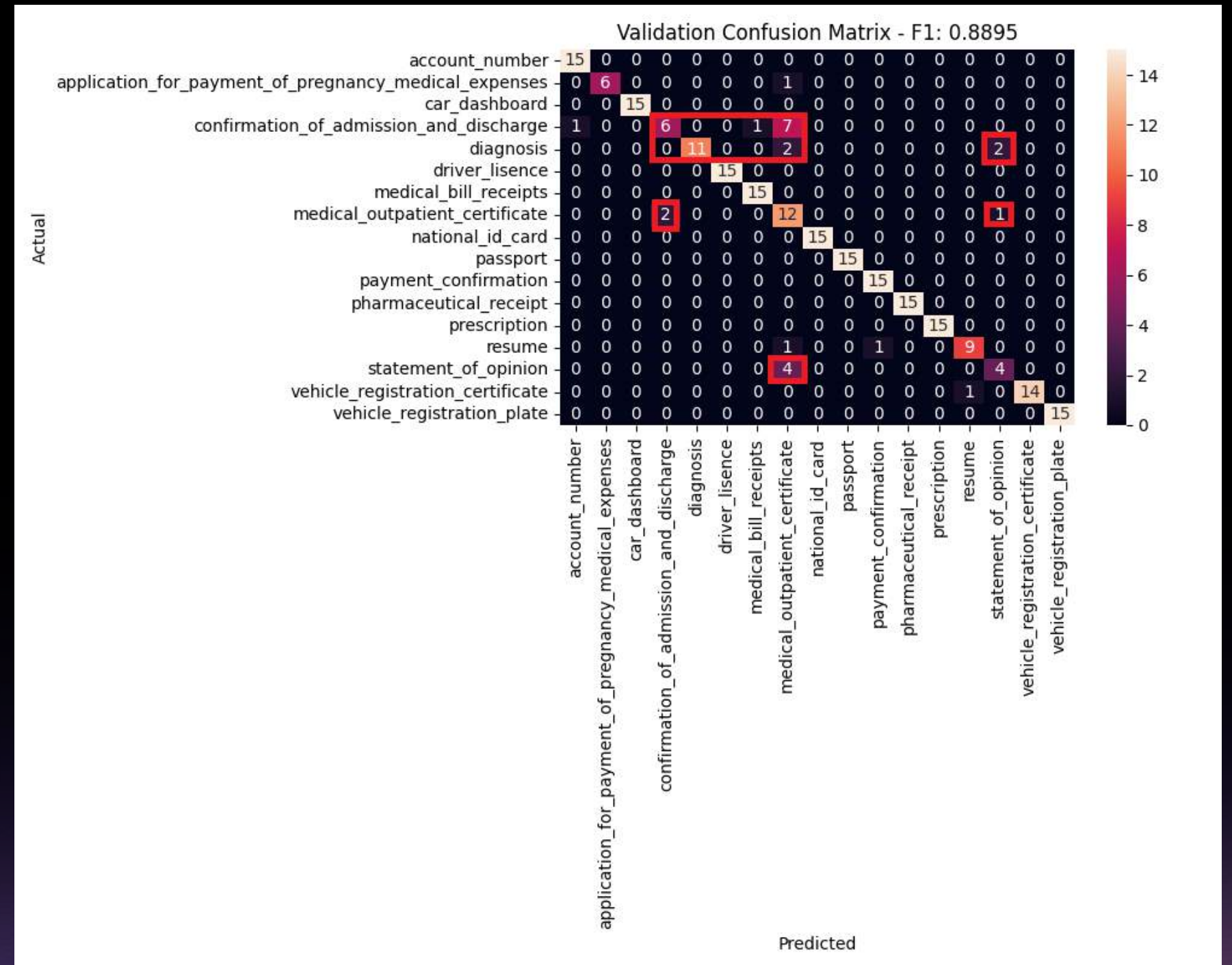


# 경진대회 인사이트 공유

## : Computer Vision [대회] Classification

### [모델링]

- 문제점 : Confusion Matrix를 확인했을 때 [3, 4, 7, 14] 에 해당하는 의료 관련 문서들을 서로 혼동하는 경우가 많았다.
- 시도 1) 2-Stage Classifier
  - Model A : easy class를 예측하는 모델
  - Model B : hard class [3, 4, 7, 14]를 예측하는 모델
- 시도 2) Custom Classifier Head
  - Claude의 도움을 받아 Classifier Head에 문서의 특징을 더 잘 추출할 수 있도록 모듈 구성
  - Standard-10-view 모듈
  - 3, 7번 클래스 특화 모듈



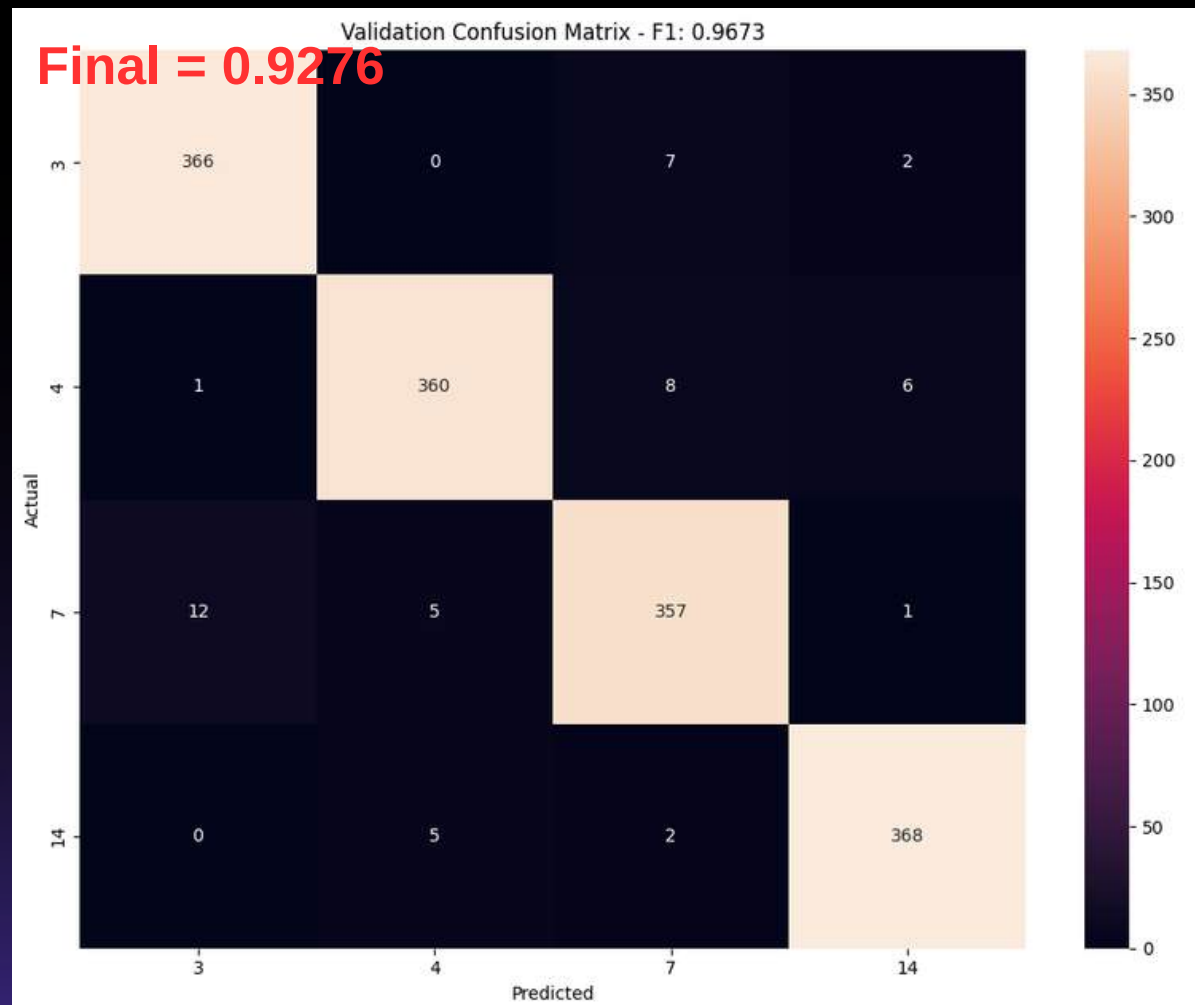
# 경진대회 인사이트 공유

## : Computer Vision [대회] Classification

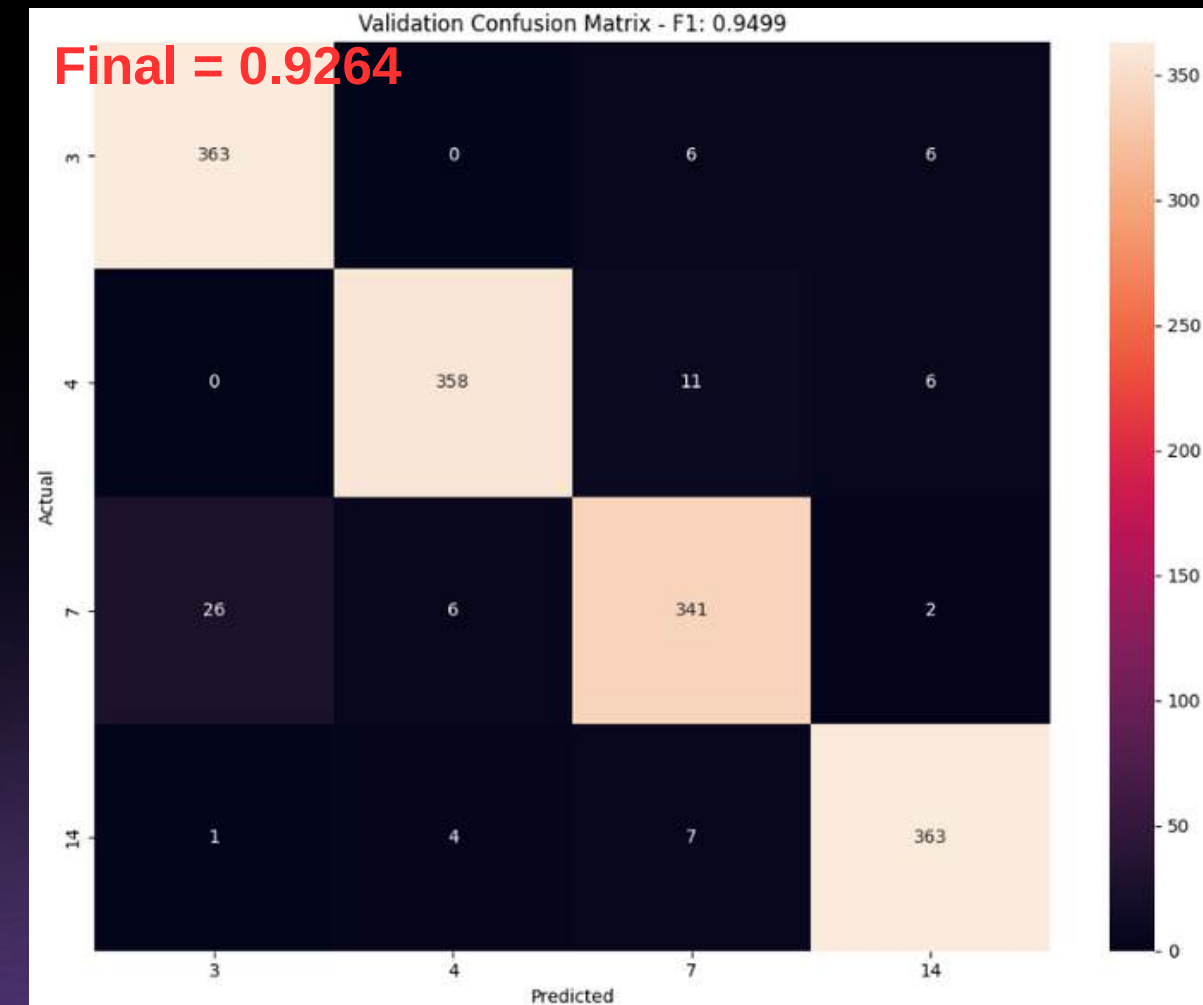
### [모델링]

- 시도 1) 2-Stage Classifier
  - Model A : easy class를 예측하는 모델
  - Model B : hard class [3, 4, 7, 14]를 예측하는 모델

- test 1)
  - Model A : convnextv2-base-384
  - Model B : convnextv2-nano-384
  - dataset : offline augment 500개



- test 2)
  - Model A : convnextv2-base-384
  - Model B : maxvit-base-384
  - dataset : offline augment 500개





# 경진대회 인사이트 공유

## : Computer Vision [대회] Classification

### [모델링] • 시도 2) Custom Classifier Head

1. hard class [3, 4, 7, 14] 를 구분하기 위해서는 문서의 제목에 집중할 필요가 있다.
  - 제목 인식을 위해 feature map의 상단 20%, 30%, 40% 부분을 crop하여 추가적으로 CNN 을 학습시키는 Head
2. test 데이터셋이 다양한 회전을 갖고 있으므로 상단 뿐만 아니라 다양한 view의 feature map에서 제목을 인식하도록 지시.
  - Standard-10-view Head로 변경 : 좌상단, 우상단, 좌하단, 우하단, 정중앙을 crop하고 각각 flip도 적용해서 10개의 view를 추출한 후, 분류 수행
3. 3(입퇴원확인서), 7(진료확인서) 간에 혼동이 심한 것을 인지하고, 이를 집중적으로 분류하는 Head를 설계하도록 지시.
  - 1번 Head 구조 사용
  - 입퇴원확인서의 “퇴”와 진료확인서의 “통”을 구분하기 위해 수직/수평 선을 더 잘 구분하도록 5x1 filter, 1x5 filter 로 filter 크기를 바꾼 모듈 추가
  - 입퇴원확인서에 “입 . 퇴”의 점(.)을 파악하기 위해 1x1 filter를 갖는 모듈 추가
4. 3번에서 제목 인식 모듈만 Standard-10-view로 변경

# 경진대회 인사이트 공유

## : Computer Vision [대회] Classification

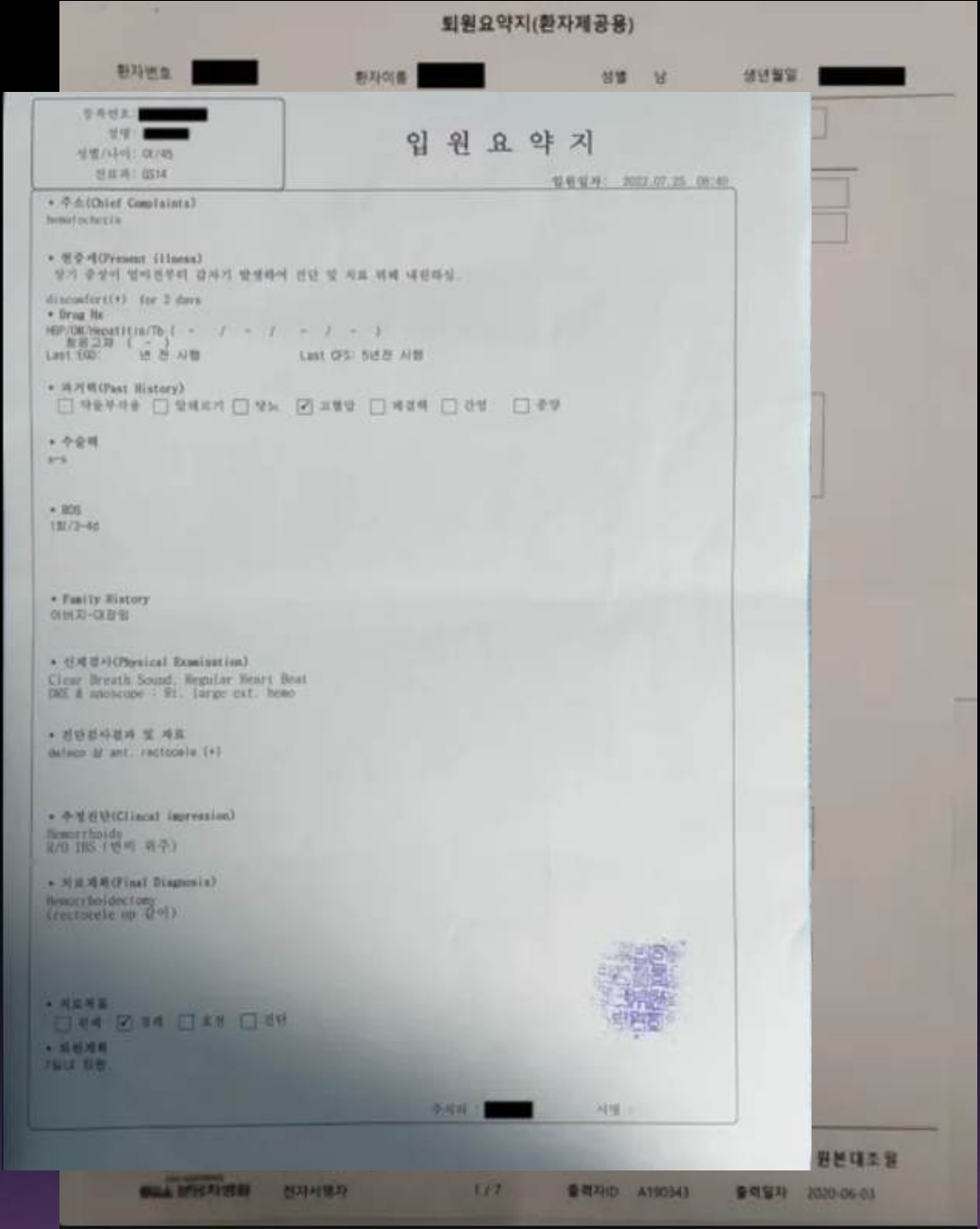
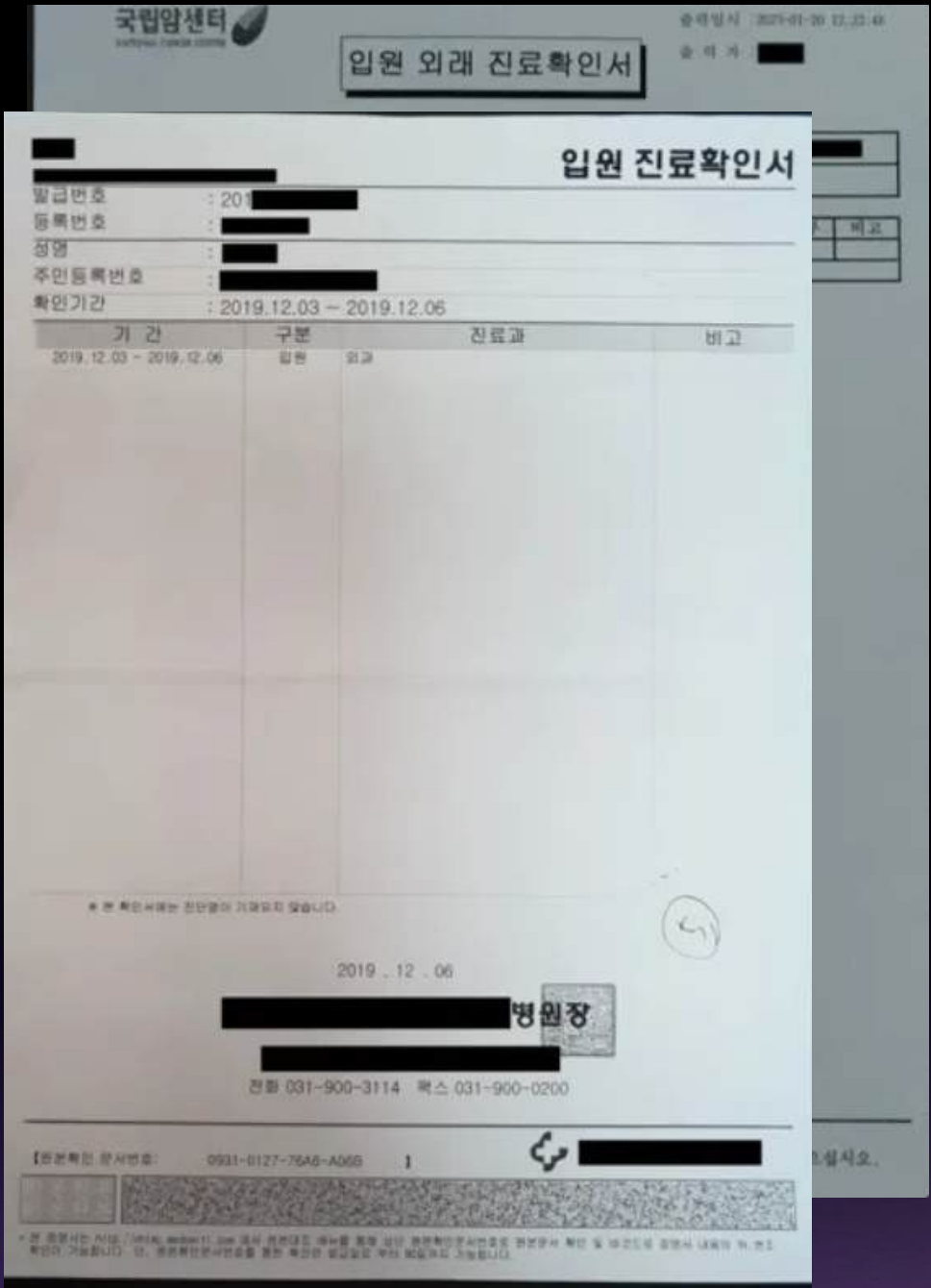
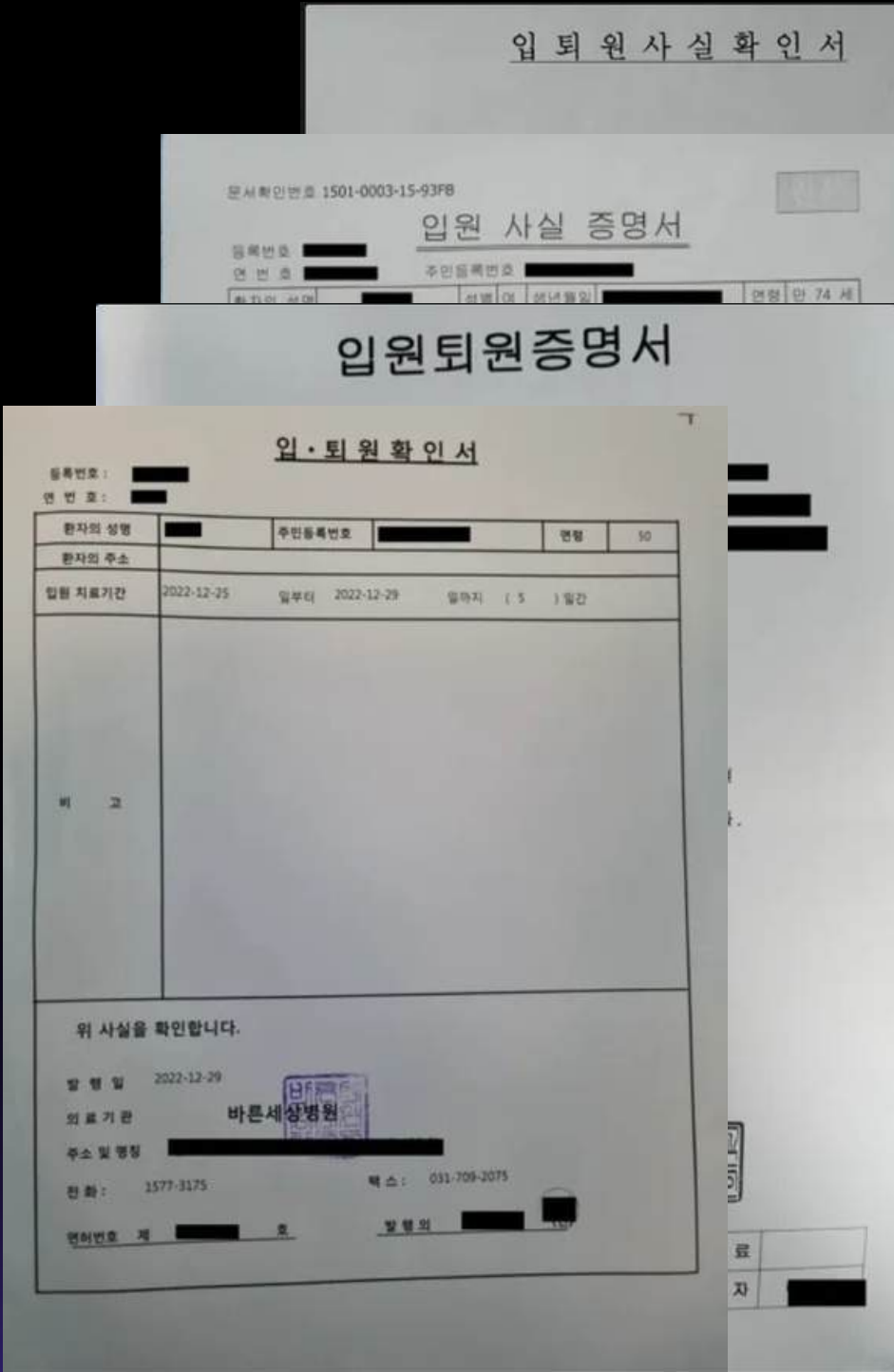
### [모델링] • 시도 2) Custom Classifier Head

1. hard class [3, 4, 7, 14] 를 구분하기 위해서는 문서의 제목에 집중할 필요가 있다. **Final = 0.9457**
  - 제목 인식을 위해 feature map의 상단 20%, 30%, 40% 부분을 crop하여 추가적으로 CNN 을 학습시키는 Head
2. test 데이터셋이 다양한 회전을 갖고 있으므로 상단 뿐만 아니라 다양한 view의 feature map에서 제목을 인식하도록 지시. **Final = 0.9107**
  - Standard-10-view Head로 변경 : 좌상단, 우상단, 좌하단, 우하단, 정중앙을 crop하고 각각 flip도 적용해서 10개의 view를 추출한 후, 분류 수행
3. 3(입퇴원확인서), 7(진료확인서) 간에 혼동이 심한 것을 인지하고, 이를 집중적으로 분류하는 Head를 설계하도록 지시. **실험 저장 못함**
  - 1번 Head 구조 사용
  - 입퇴원확인서의 “퇴”와 진료확인서의 “통”을 구분하기 위해 수직/수평 선을 더 잘 구분하도록 5x1 filter, 1x5 filter 로 filter 크기를 바꾼 모듈 추가
  - 입퇴원확인서에 “입 . 퇴”의 점(.)을 파악하기 위해 1x1 filter를 갖는 모듈 추가
4. 3번에서 제목 인식 모듈만 Standard-10-view로 변경 **학습이 불안정하여 수렴하지 못함**

[ 표준 입퇴원확인서 ]

[ 진료 확인서와 혼동 가능한 입퇴원확인서 ]

[ 입퇴원확인서 ?? ]





[ 표준 진료확인서 ]

진료 확인서

차트번호: [redacted]  
연번: [redacted]

성명	[redacted]	성별	남	연령	만 35 세
주민등록번호	[redacted]				
주소	[redacted]				
병명	신경부리병증을 동반한 요추 및 기타 추간판장애(G55.1*) M511				
입원	년 월 일 부터 년 월 일 까지 ( 간)				
통원	2023년 06월15일 부터 2023년 06월15일 까지 (1 일간)				
실통원일자	15일				

상기와 같이 진료 받았음을 확인합니다.

발행일 2023년 06월27일  
의사성명 [redacted]  
연번 [redacted]  
주소 및 영칭 [redacted] 활기전마취통증의학과의원  
전화번호 055-313-2500 (FAX) 055-313-2501

의료기관의 장 [redacted]

[ 다양한 진료 확인서 ]: 진료확인서 종류도 많네요~~

외래 진료 사실 확인서

문서확인번호: [redacted]

성명	[redacted]	성별	남	연령	49 세
주민등록번호	[redacted]				
주소	[redacted]				
병명	관절성 기원의 기타 및 상세불명의 위장관 및 장관의 두통 비급여 항목: 치료목적 수액				
입원	년 월 일 부터 년 월 일 까지 ( 일간)				
통원	2023 년 03 월 02 일 부터 년 월 일 까지 ( 일간)				

상기 환자는 본원에 내원하여 위와 같은 병명으로  
[redacted] 치료(하고) [redacted] 응급 확인 합니다.  
발행일: 2023-03-02

요양기관명: 오뚜기정의원  
주소: [redacted]  
전화번호: 031-275-0005  
연번: [redacted] 의사명: [redacted]

치 료 확 인 서

문서확인번호: [redacted]

(외래진료)진료사실증명서

연번: [redacted]  
문서번호: [redacted]

성명	[redacted]	성별	남
주민등록번호	[redacted]		
주소	[redacted]		
병명	[redacted]		
진료이력	2023-06-13 2023-05-30 2023-05-23 2023-05-11		
내원이력	[redacted]		
비고	[redacted]		

상기 환자는 위와 같이 본 병원에서 진료받은 사실을 증명합니다.  
발행일: 2023년 6월 22일  
의료기관: 한림대학교동탄성심병원  
주소 및 영칭: [redacted]  
대표전화: 1522-2500  
연번: [redacted] 의사명: [redacted]

## 경진대회 인사이트 공유

: Computer Vision [대회] Classification

[입퇴원확인서와 헛갈리기 쉬운 진료 확인서]: 문제의 “통”원 씨리즈~~~

통원치료사실확인서

입 퇴원 확인서

통원증명서

등록번호: [redacted]  
연번: [redacted]

1. 환자의 주소 [redacted]

2. 환자의 성명 [redacted] 성별 [redacted] 나이 [redacted] 생년월일 [redacted]

3. 병명 알레르기질환

4. 치료기간 2023.02.11 ~ 2023.02.17

위와 같이 진료하였음을 확인함.

발행일: 2023.02.17

포항세명기독병원

전화: 054)275-0005  
팩스: 054)275-0003

통원 진료 확인서

문서확인번호 1501-0005-01-45C0

등록번호 [redacted] 연번 [redacted] 주민등록번호 [redacted]

환자의 성명 [redacted] 성별 [redacted] 생년월일 [redacted] 연령 [redacted] 만 0 세

환자의 주소 [redacted]

진료과 어린이병원 소아심장과

진단명 심장병 심장폐렴 심장충격증

COMMENT

위 환자는  
2015/11/06[경사], 2015/11/06, 2015/10/08[경사], 2015/10/08  
(4) 차례

본원에서 위 상병으로 통원진료를 받은 사실이 있음을 확인함.

작성일 2015년 11월 06일

의사명 [redacted]

연번번호 [redacted]

서울아산병원장

\* 본서에 본문의 직인이 없으면 무효입니다.

본 문서는 최첨단 정보 기술이 적용된 문서이며, 정가와 같이 2차원 바코드와 일치하지 않은 문서는 복제본으로 사용할 수 없습니다. 또한, 내용과 전파 여부는 발급일로부터 90일간 확인 가능하며 서울아산병원 홈페이지(www.amsc.seoul.kr)에서 확인할 수 있습니다.

심각한 문제 직면: 예측된 분류를  
검산해보니, 입퇴원확인서에  
30장이 넘는 진료확인서 문서들이  
발견

입퇴원확인서와 진료확인서를 어떻게  
구분할 것인가?

전략 1) OCR 로 할까?

전략 2) 오프라인 데이터 증강을 할까?

전략 3) 딥러닝 고도화 기법에 도전할  
까?



04

## 회고

---

우리 팀의 목표 달성도  
느낀점 및 향후 계획



# 경진대회 회고

## : Computer Vision [대회] Classification

### Point 1

#### 우리 팀의 처음 목표에서 어디까지 도달했는가

목표했던 바를 어느 정도 이뤄냈기에 70%정도 도달했다고 생각한다.  
데이터 증강, 모델 실험 등 최대한 많은 시도를 했다.

### Point 2

#### 우리 팀이 잘했던 점

다음 회의 때까지의 각자의 역할을 잘 분배해서 진행한 것  
서로 의견 충돌을 하며 더 좋은 아이디어를 추출해낼 수 있었다. 이를 바탕으로 멘토링 시간에 양질의 정보를 얻었고 많은 시도를 할 수 있었다.

### Point 3

#### 협업하면서 아쉬웠던 점

마지막 날에 시간이 더 많았다면 점수를 더 올릴 수 있었지 않을까 싶다.

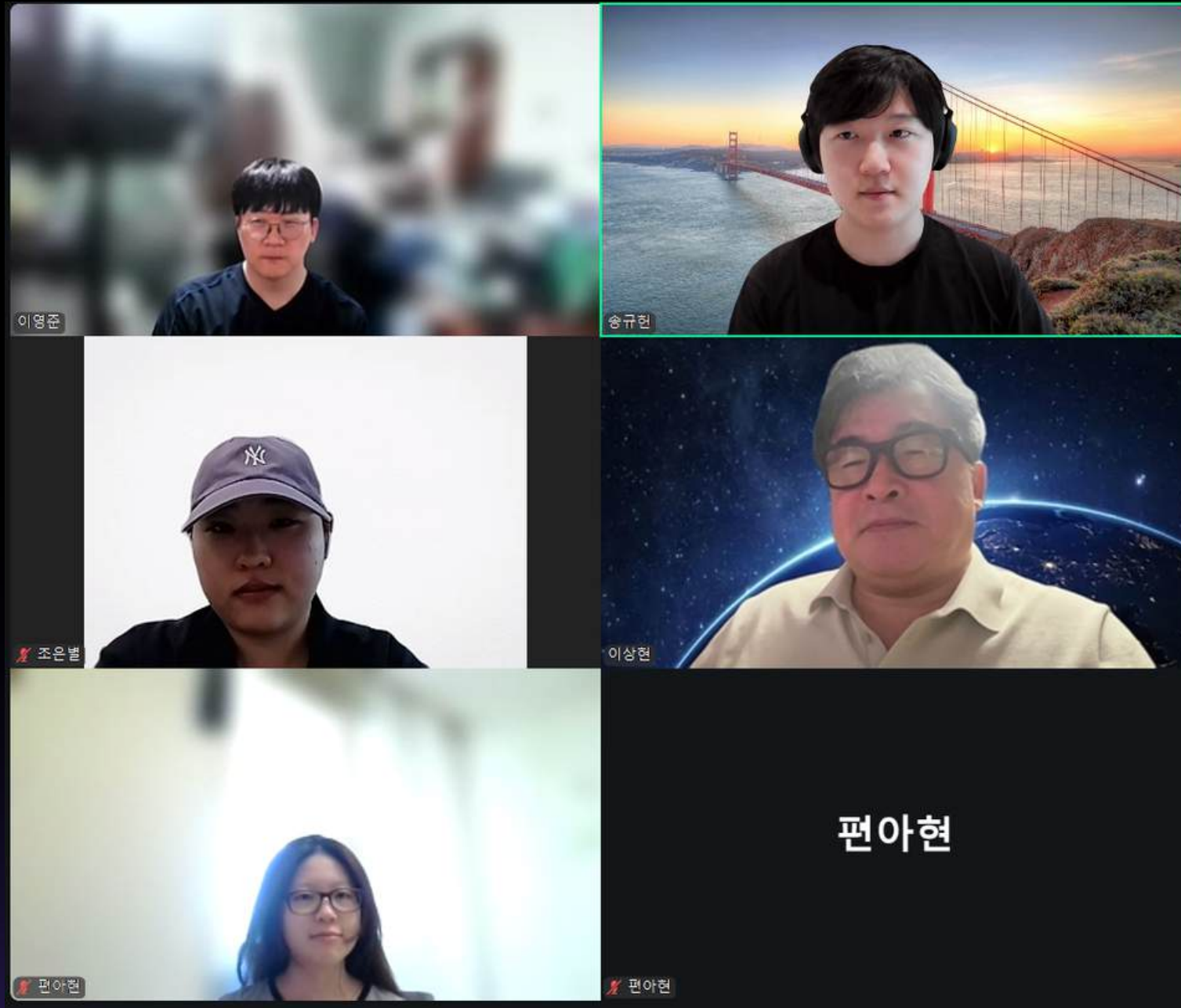
- 향후 계획 : 다음에도 이런 대회를 참여한다면 이번 인사이트를 기반으로 더 높은 점수를 노려보고 싶다.
- epoch를 1로 하여 테스트를 먼저 수행한 후 학습을 했어야 했는데, 너무 급하게 코드를 작성하느라 제대로된 테스트 없이 학습을 수행했고, 오류가 많이 발생해 시간을 많이 잡아 먹었다.
- 코드 준비가 늦어지니 다양한 모델, 옵티마이저, 스케줄러, 손실함수 실험을 하지 못했다.

이미지 수집과 라벨링의 중요성을 실감할 수 있는 좋은 기회였고, 이런 경험을 할 수 있도록 해주신 주최 측에  
“격하게” 감사를 표하고 싶습니다.

## 경진대회 진행 소감

: Computer Vision [대회] Classification

느낀점



송규현

짧은 시간 내에 많은 실험을 시도한 경험은 만족스럽지만, 체계적으로 실험을 통해 점수를 개선해 나가는 프로세스를 경험하지 못해 아쉽습니다. 코드 오류를 빨리 해결하지 못한 것이 아쉽고, 다음부터는 LLM을 더 잘 활용해 코드보단 데이터에 중점을 두고 경진대회를 진행하려 합니다.

이상현

이번 이미지 경진대회를 통해 이미지 데이터의 세계에 흠뻑 빠져본 느낌이다. CNN이나 여러 딥러닝 기법들이 도깨비 방망이가 아니라, 우리의 노력으로 이미지를 모아야하고, 그 이미지들을 일일이 라벨링을 한 수고가 뒤에 있어야 함을 실감했다. 수강생들이 좋은 경험을 하게 만드느라 어렵게 데이터셋을 만들어 주신 주최측에 격하게 감사를 표한다.

이영준

자동화를 구축하여 다양한 실험을 하려고 하였습니다. 마지막에는 Claude가 이상하게 코딩한 부분으로 마감 전 오래 진행된 실험 결과가 한 클래스만으로 이상하게 나왔고, 마감을 얼마 남지 않은 시점에 서버가 용량 초과로 죽어서 열심히 노력한 부분이 빛을 보지 못해 아쉬움이 많은데 다음에는 철저히 준비해서 팀에 더 도움이 되도록 하겠습니다.

조은별

마지막에 실험하던 MaxVit 모델을 더 다양한 조건에서 실험을 마무리하지 못해서 아쉬운 마음이 듭니다. 그럼에도 이번 문서 이미지 분류 경진대회를 통해 다양한 데이터 증강 기법, 스케줄러 전략, 그리고 Transformer 기반 모델들에 대해 깊이 있게 학습할 수 있어 매우 유익한 경험이었습니다.

편아현

조금이라도 더 높은 점수를 낼 수 있었을 것 같아 더 아쉬움이 남지만, 개인적으로 팀에 적절히 기여한 것 같아 만족스럽습니다. 팀원들 모두 각자의 위치에서 열심히 해서 6등이라는 등수까지 올랐던 것 같고, 이 점수까지도 올리기 힘들었기에 의미있는 등수라고 생각합니다.

Life-Changing Education

# 감사합니다.

---