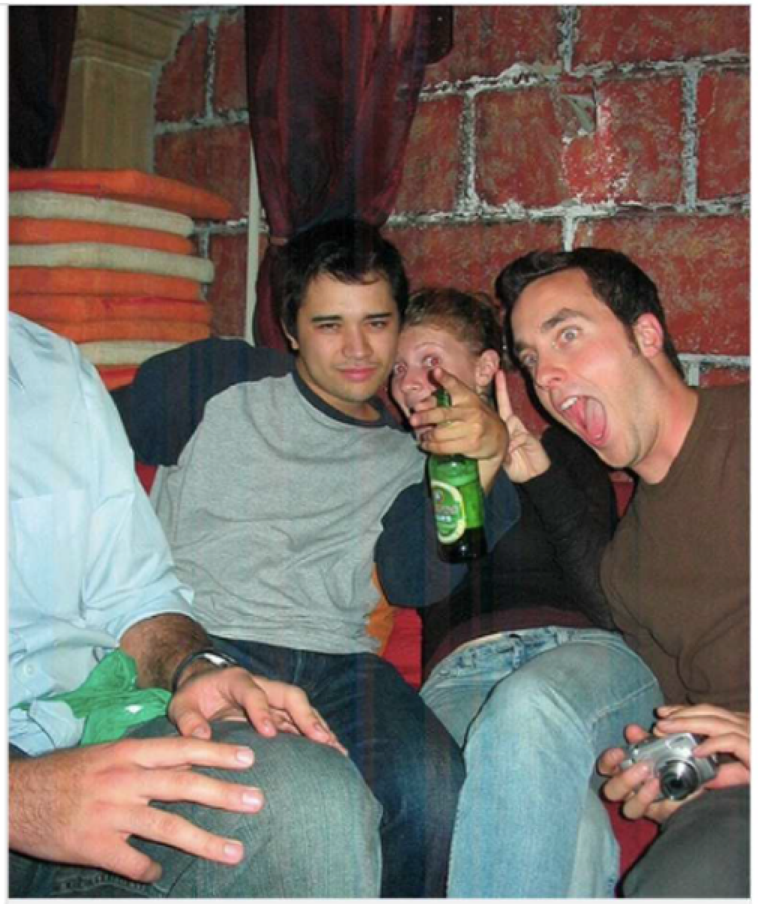
DACON x UOS

2023 AICOSS HACKATHON

- Satellite Image Multilabel Classification

University of Seoul
Jun Park, Minsu Kwon, Wonseok Choi

Multilabel Classification



2 positive label: Person, Camera
78 negative labels
Missing label: Bottle

사진에 등장하는 모든 Category를 분류하는 Task

Challenges

- 전체 Class 경우의 수가 수천 개 이상 존재함
- 학습 데이터에 Missing label이 존재함

Solution

수천 개의 Class를 볼 수 있는 **큰 모델**과 Missing label을 잘 걸러내는 **최적화** 전략 필요!

Satellite Image

Positive label : Building, Cars, ...

Negative label : Airplane, Airport, ...

Positive label : Bare soil, grass, ...

Negative label : Airplane, Airport, ...

Train dataset : 65496개

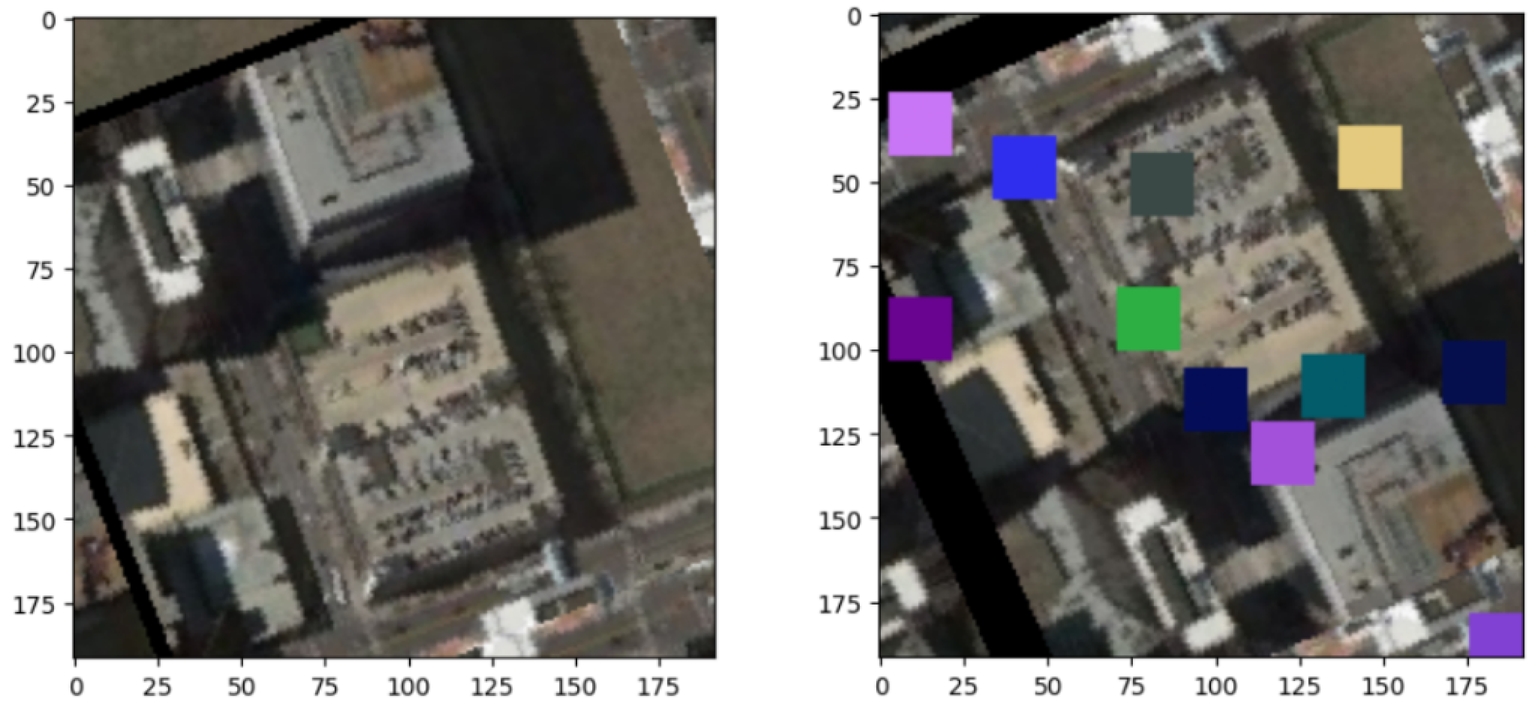
Public test dataset : 43665개

총 60개 클래스 존재

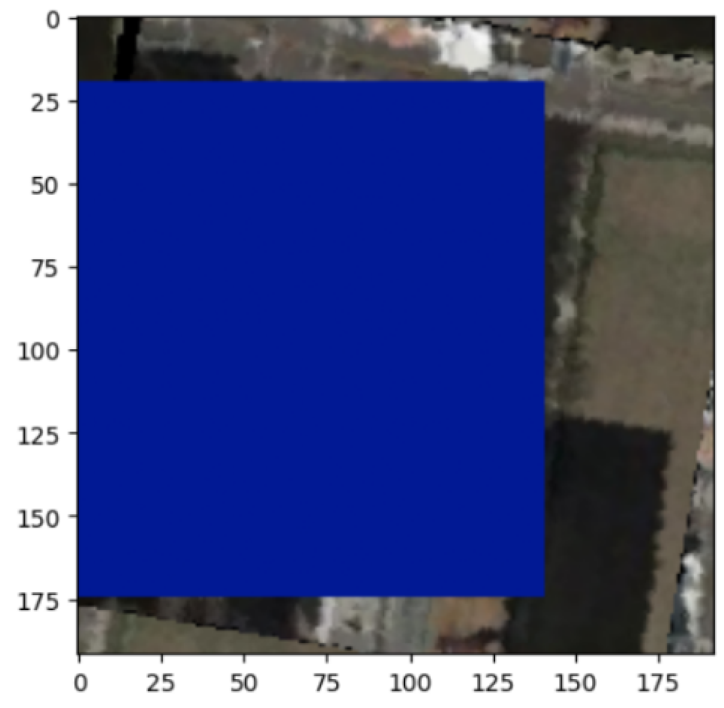
Data Augmentation Strategy

Insight

항공 이미지 다중 분류 문제에서, 과한 Augmentation 기법은 정보손실을 일으킴!



[Fair Augmentation]



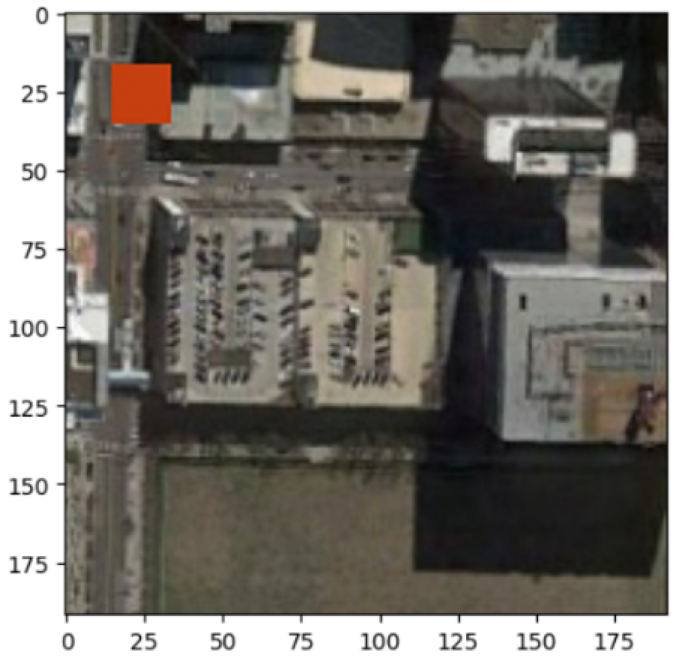
[Wrong Augmentation]

무결성을 깨트리지 않는 최적의 Augmentation 탐색

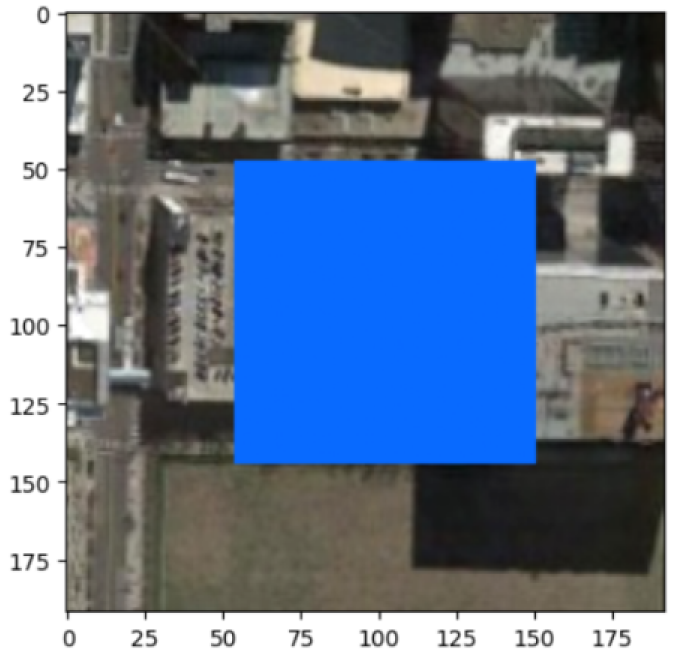
Data Augmentation Strategy

Ablation Study - Cutout Factor

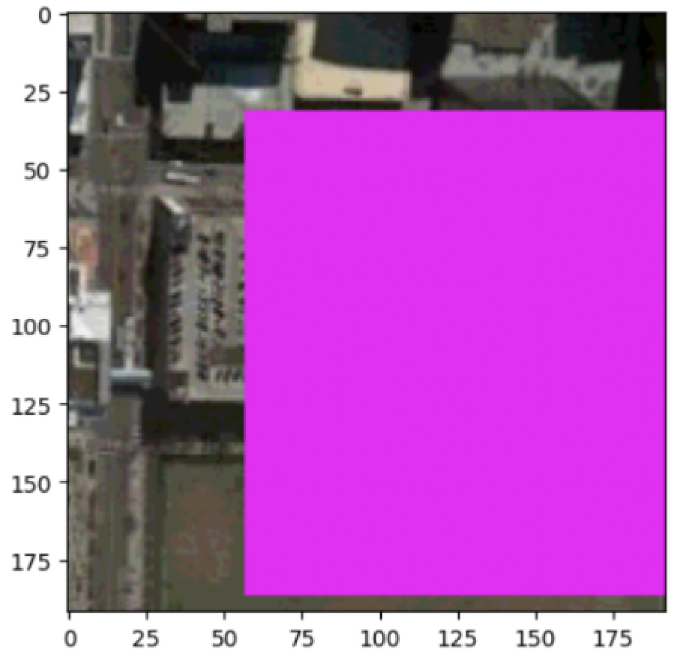
Cutout factor를 조절해가며 최적의 Cutout factor 탐색



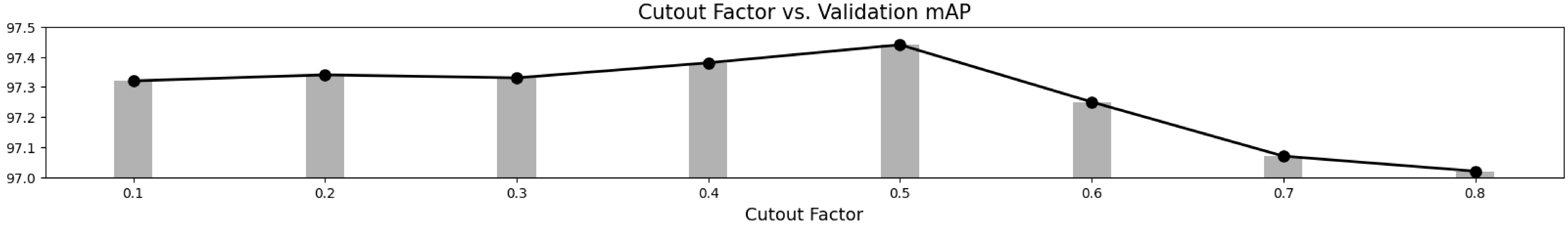
0.1



0.5



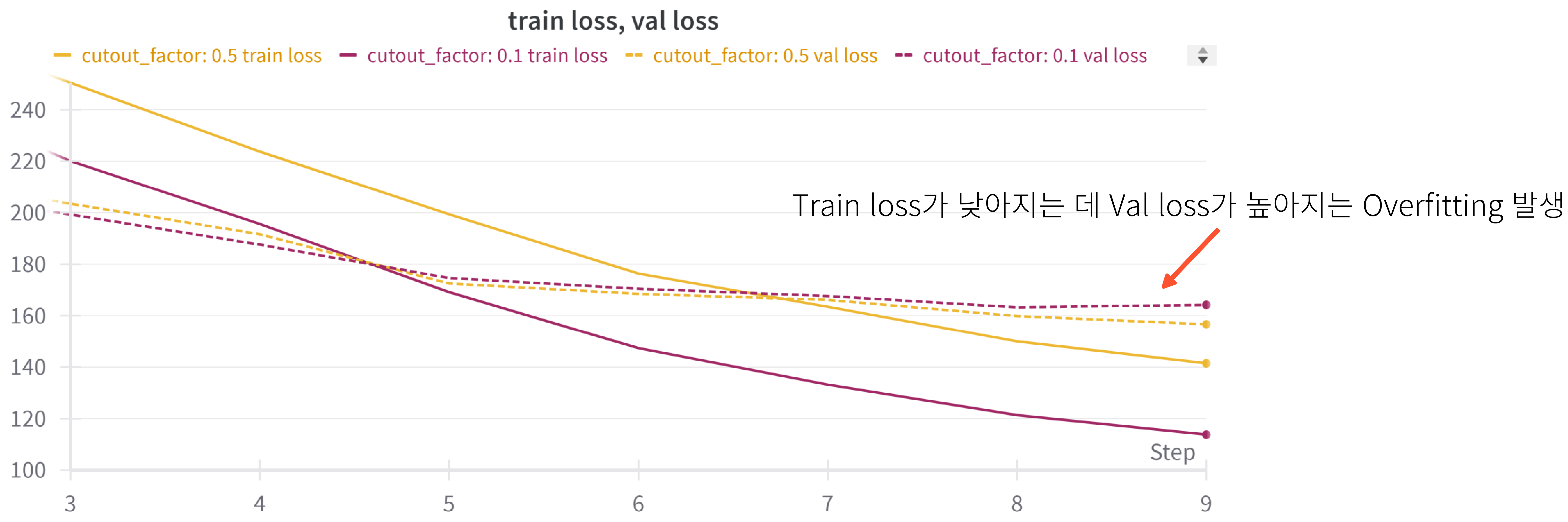
0.8



Data Augmentation Strategy

Ablation Study - Cutout Factor

Cutout factor가 너무 낮은 경우 Overfitting 문제가 발생 (일반화 성능 저하)



Data Augmentation Strategy

Ablation Study

Greedy search 방식으로 모델 마다 최적화된 Augmentation 방식을 찾음

Model Name	Loss	Prior	LR	MIN_LR	Augment	Batch Size	DDP
TResNet_XL_ML-Decoder	P-ASL	X	3e-4	1e-6	Weak	128	X
TResNet_XL_ML-Decoder	Two-Way	-	3e-4	1e-6	Weak	128	X
TResNet_v2_L_ML-Decoder	ZLPR	-	3e-4	1e-6	Weak	128	X
CvT_21_Q2L	P-ASL	X	3e-4	1e-6	Weak	400	O
TResNet_XL_Learnable_ML-Decoder	P-ASL	O	3e-4	1e-6	Strong	128	X
TResNet_XL_Learnable_ML-Decoder	P-ASL	O	3e-4	1e-5	Strong	128	X
Swin_v2_L_ML-Decoder	ASL+P-ASL	O	1.2e-4~1e-4	0	Weak	456	O

[Weak] Cutout(factor=0.5) + RandAugment(magnitude=9)

[Strong] RandomHorizontalFlip + RandomVerticalFlip + Cutout(factor=0.5) + AutoAugment

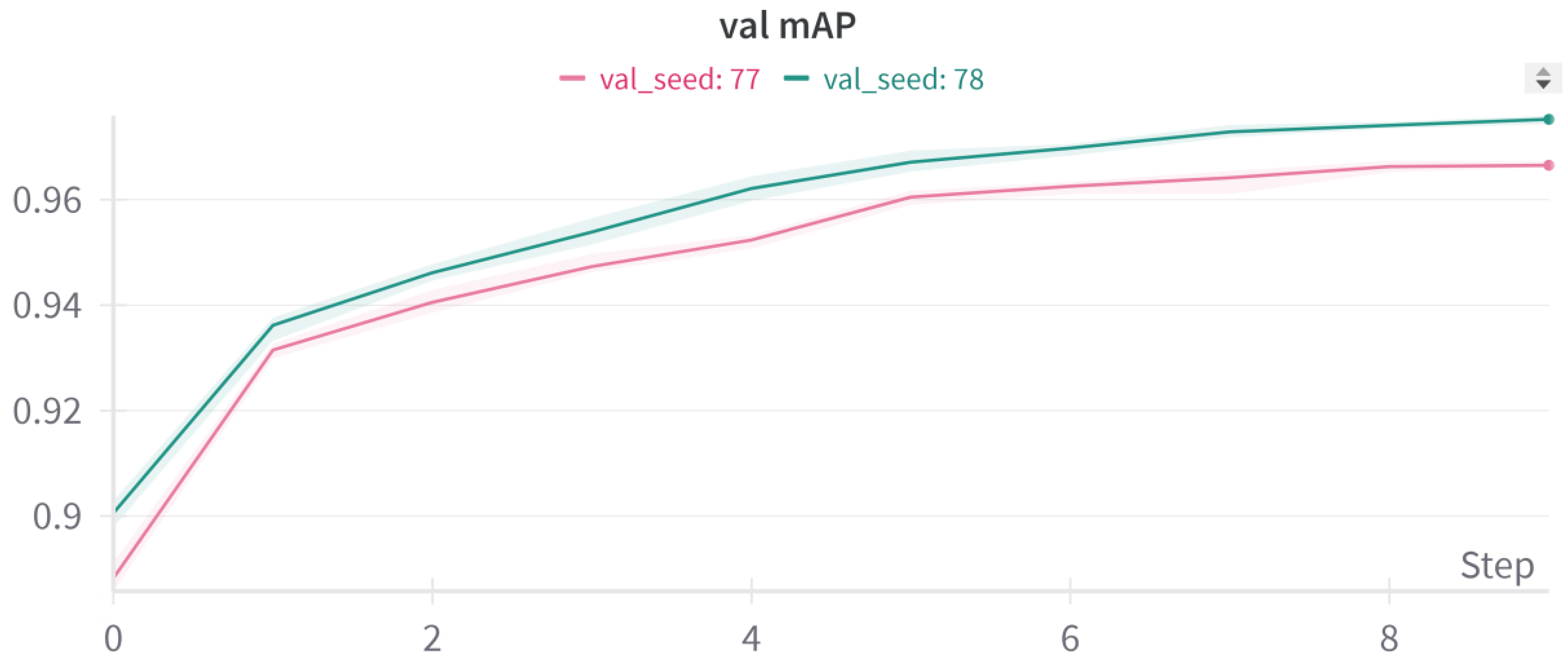
Train/Val Dataset Split Strategy

Insight

무작위하게 100번 Split 후 Train과 Validation의 클래스 별 비율 차이의 표준편차가 가장 작은 Split을 선택

Ablation Study - Best/Worst split

Best split과 Worst split으로 각각 학습한 후 val mAP 비교



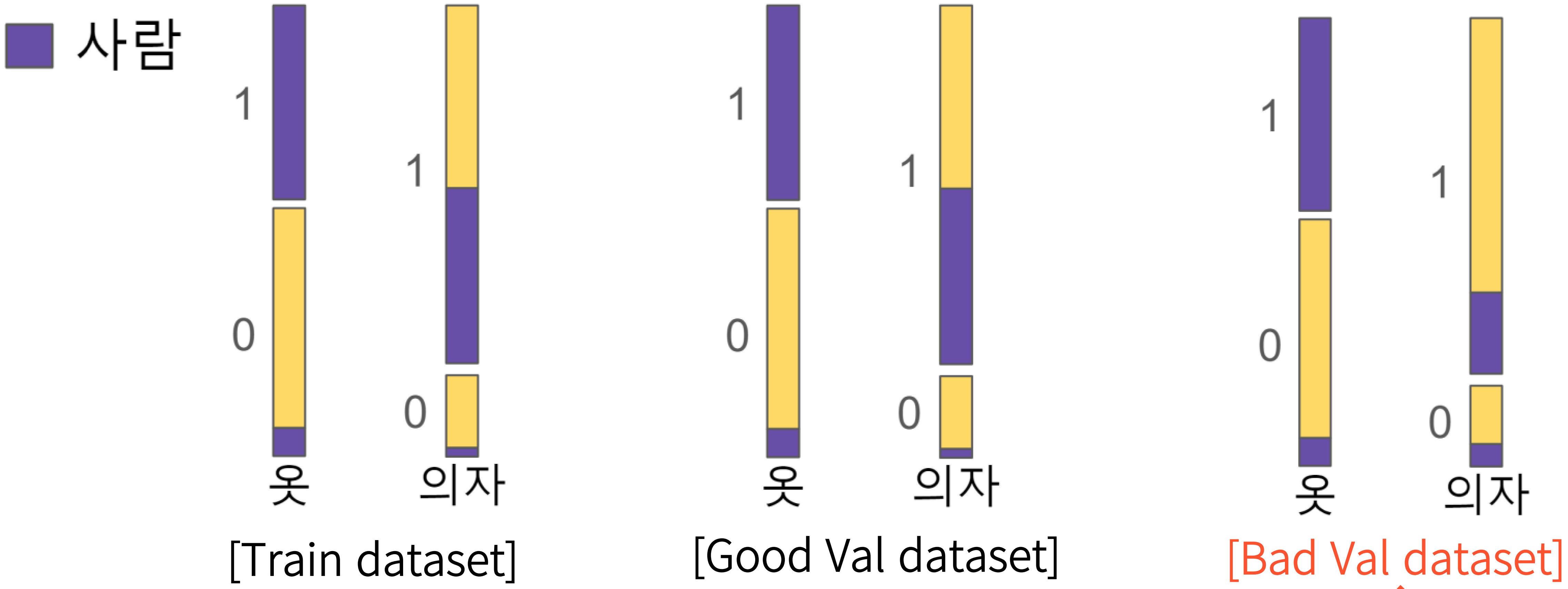
Random Seed	78 (Best)	77 (Worst)
$\sigma(\lvert \mathbf{P}_{train}(\mathbf{c}) - \mathbf{P}_{val}(\mathbf{c}) \rvert)$	0.22	0.41
val mAP (%)	97.52	96.65

$$where\ \mathbf{c} = [1, 2, \dots 60]$$

Train/Val Dataset Split Strategy

Challenges

Multilabel dataset은 경우의 수가 많아 정합된(Stratified) Split을 얻기 힘듬



'의자'에 앉아있는 '사람'을 Train dataset보다 적게 담았음

Loss Strategy

💡 *Insight*

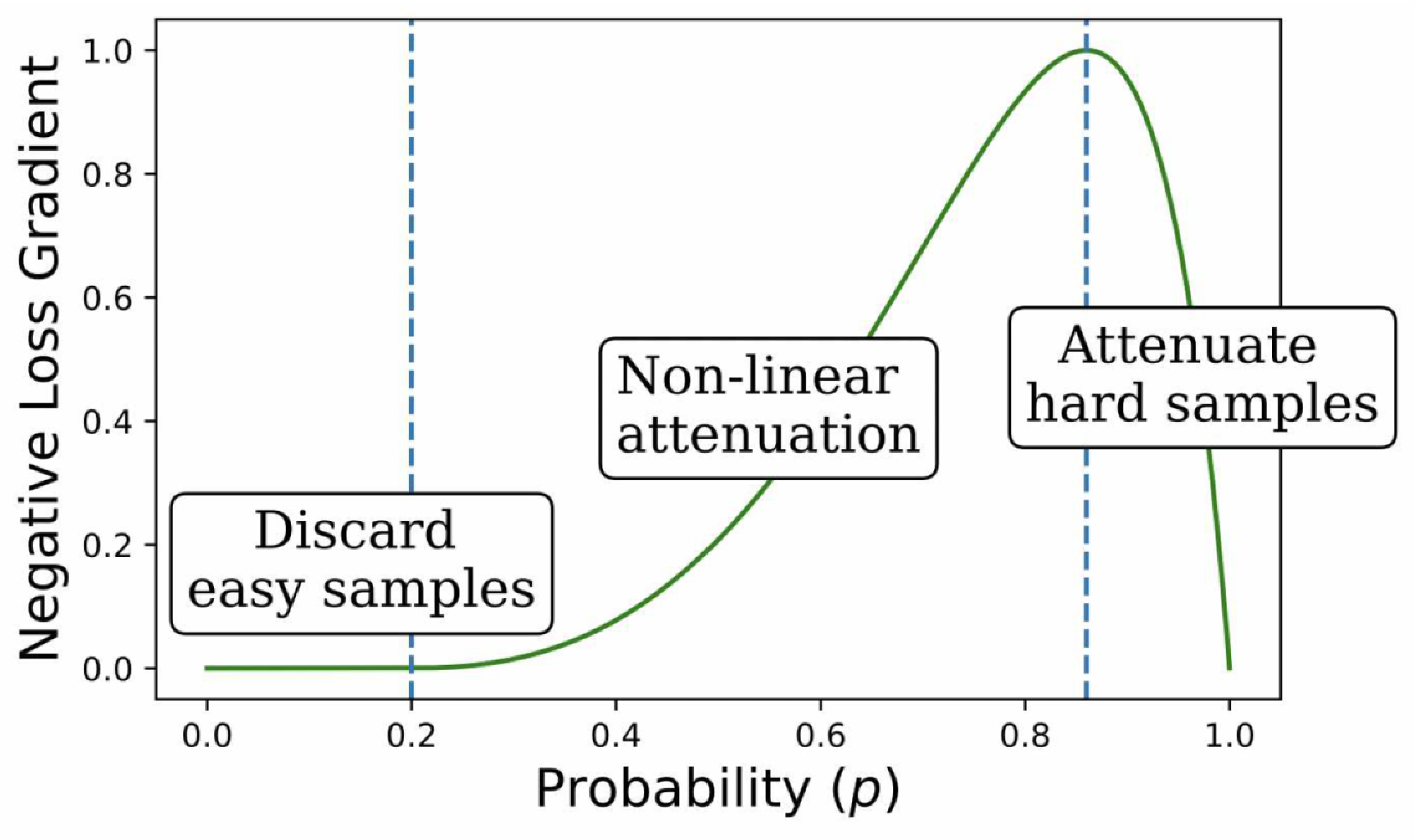
Multilabel Classification에 주로 사용되는 Binary Cross Entropy (BCE) Loss는 한계가 존재함

📌 *Solution*

	Class imbalance를 고려했는가?	Missing label을 고려했는가?	Unannotated data를 다룰 수 있는가?	Classification margin을 고려했는가?
BCE [Baseline]	❌	❌	❌	❌
Focal Loss [ICCV 2017]	✅	❌	❌	❌
ZLPR [ARXIV 2022]	✅	❌	❌	❌
ASL [ICCV 2021]	✅	✅	❌	❌
PASL [CVPR 2022]	✅	✅	✅	❌
TWL [CVPR 2023]	✅	✅	❌	✅

Loss Strategy

💡 *Asymmetric Loss (ASL) 분석*



Prior works

- Easy sample에 대해 Exponential decay를 사용해 Loss를 줄임 (Focal loss)

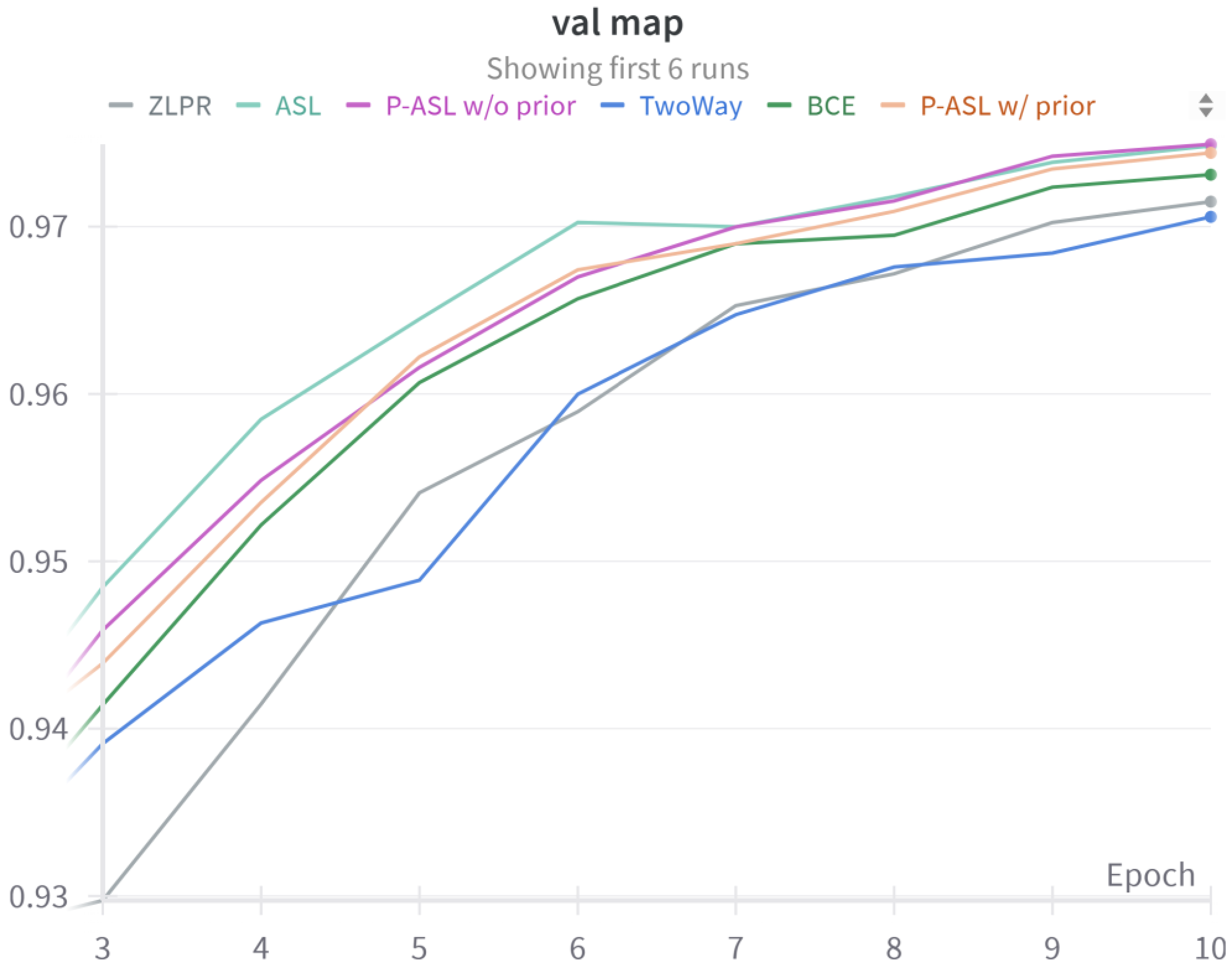
ASL

- Postive sample과 Negative sample에 대해 다른 Decay factor를 사용함 (γ^+, γ^- 가 Hyperparameter로 추가됨)
- Hard negative sample(Missing label)에 대해 Loss gradient를 감소시킴

* Loss Gradient : 학습시에 모델에 영향을 주는 정도

Loss Strategy

Ablation Study - Loss



	val mAP (%)
PASL (w/o prior)	**97.49 (+0.18)**
ASL	97.48 (+0.17)
PASL (w/ prior)	97.44 (+0.13)
BCE *(Baseline)*	97.31
ZLPR	97.15 (-0.16)
TWL	97.06 (-0.25)

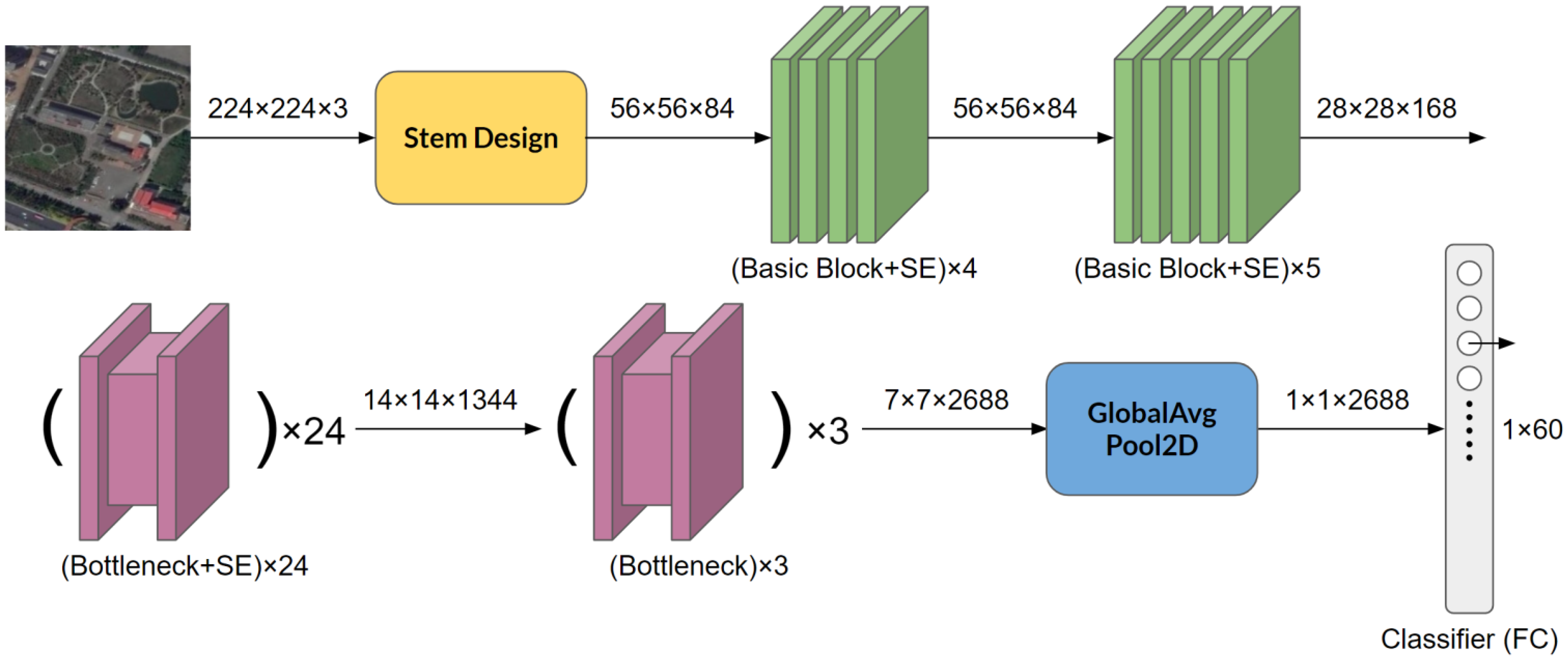
Model Selection

💡 *Insight*

ResNet[CVPR 2016]은 Image classifcation에서 널리 사용되는 논문이나, 2024년에 사용하기에 너무 오래되었음

📌 *Solution*

ResNet 보다 향상된 Accuracy와 GPU Throughput을 제안한 TResNet-XL[WACV 2021]을 Backbone으로 사용



Models	Top Training Speed (img/sec)	Top Inference Speed (img/sec)	Max Train Batch Size	Top-1 Acc. [%]
ResNet50	**805**	2830	288	79.0
TResNet-M	730	**2930**	**512**	80.8
TResNet-L	345	1390	316	81.5
TResNet-XL	250	1060	240	**82.0**

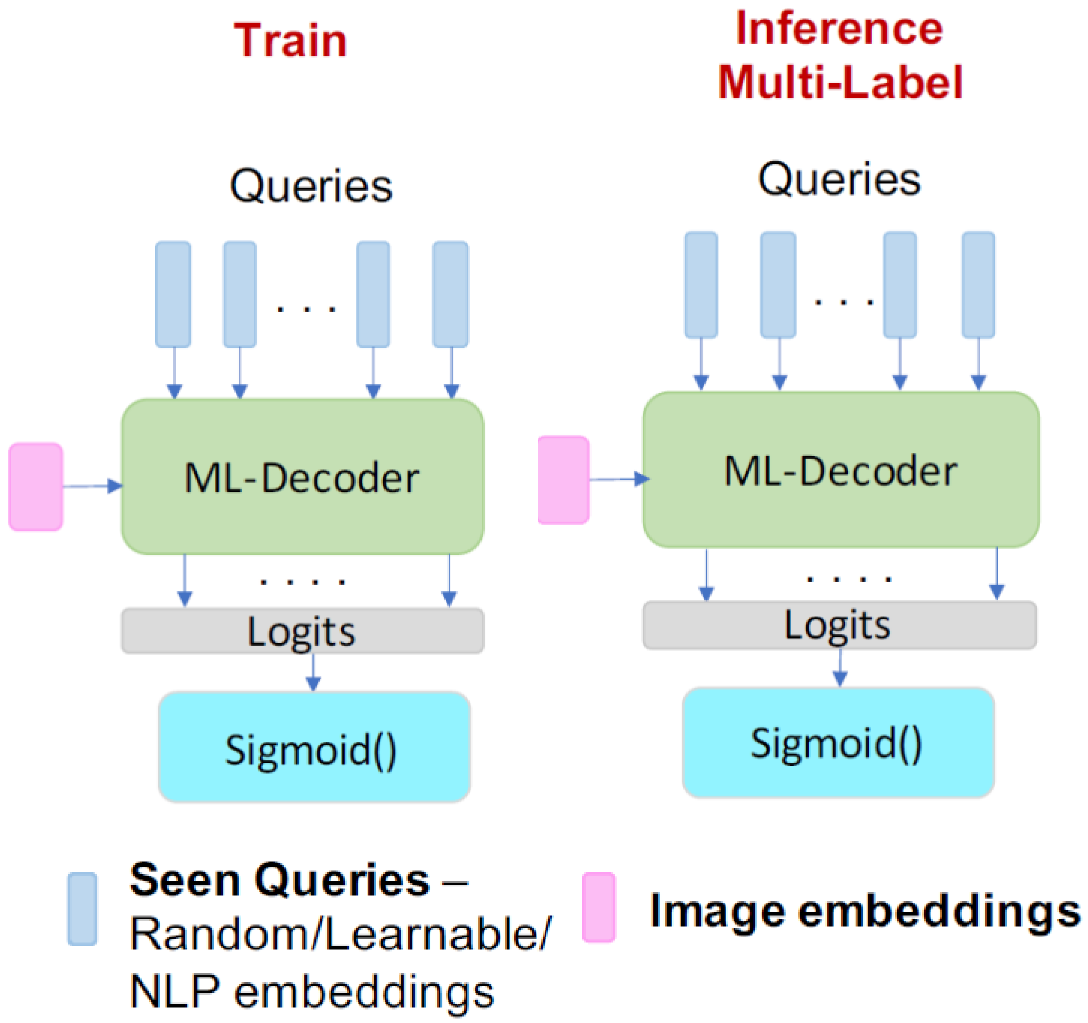
Multi-head Attention-based Decoder

💡 *Insight*

Multi-head attention based decoder를 Multi label classification의 Classifier로 사용하면 성능이 향상됨 (Q2L, Arxiv 2021)

📌 *Solution*

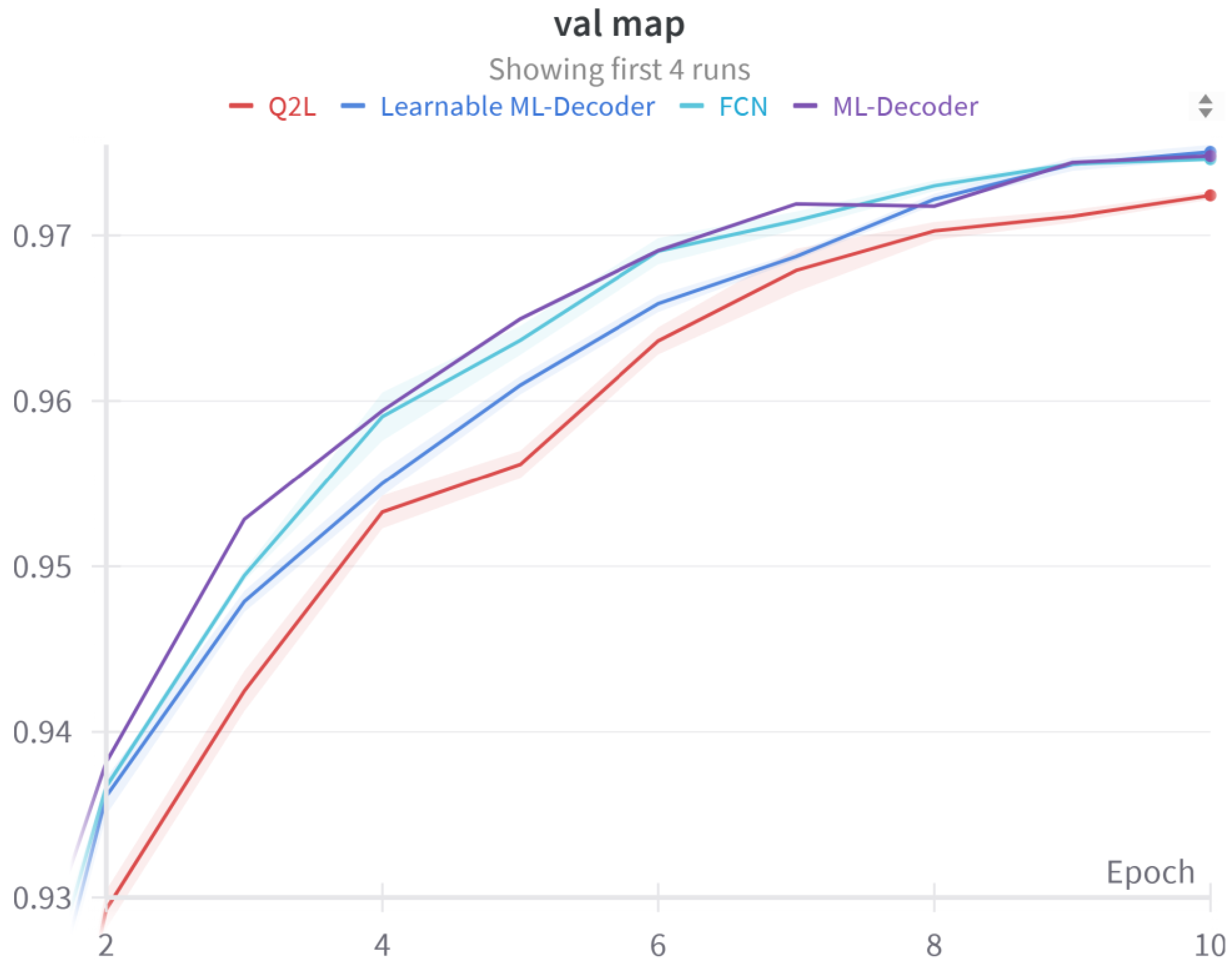
Q2L[Arxiv 2021]과 ML Decoder[WACV 2023]을 Classifier로 사용



Classification Head	Num of Classes	Num of Queries	Flops [G]	mAP [%]
GAP	80	—	23.0	87.0
Transformer-Decoder	80	80	24.1	**88.1**
ML-Decoder	80	20	23.6	88.0
ML-Decoder	80	80	23.9	**88.1**

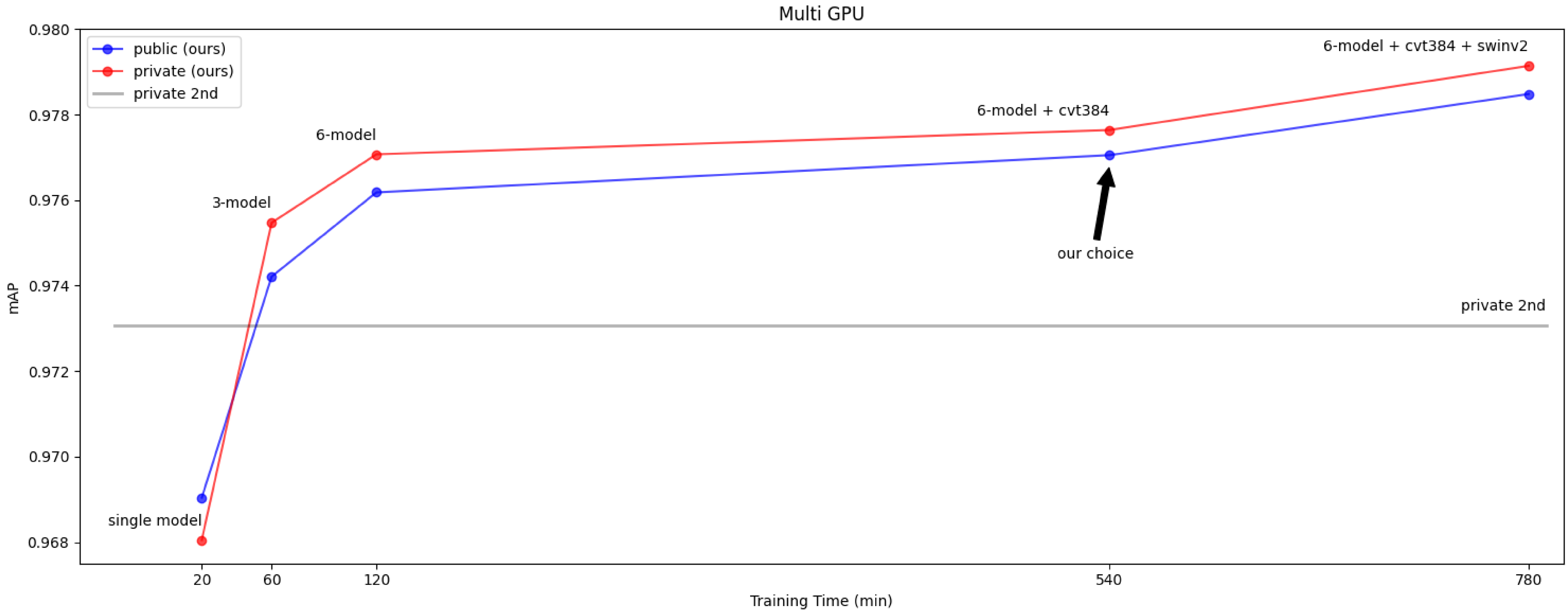
Multi-head Attention-based Decoder

Ablation Study - ML Decoder



	Val mAP (%)
Learnable ML Decoder	**97.50 (+0.05)**
ML Decoder	97.49 (+0.04)
FCN (*Baseline*)	97.45
Q2L	97.20 (-0.25)

Ensemble


Multi GPU
public (ours)
private (ours)
private 2nd
6-model + cvt384 + swinv2
6-model + cvt384
6-model
3-model
our choice
private 2nd
single model
0.980
0.978
0.976
0.974
0.972
0.970
0.968
mAP
20
60
120
540
780
Training Time (min)

Model Output Analysis (6-ensemble)

class	AP	class	AP	class	AP	class	AP	class	AP	class	AP
0	0.98626	10	0.99951	20	0.99883	30	0.99603	40	0.98061	50	0.96596
1	0.97931	11	0.99986	21	0.99986	31	0.99216	41	0.98198	51	0.98176
2	0.94066	12	0.96859	22	0.97454	32	0.98679	42	0.9955	52	0.97913
3	0.9753	13	0.9959	23	0.99806	33	0.94237	43	0.99213	53	0.97839
4	0.99166	14	0.99934	24	0.98751	34	0.97778	44	0.94854	54	0.94804
5	0.99991	15	0.99961	25	0.99724	35	0.99704	45	0.98082	55	0.98392
6	0.97906	16	0.95873	26	0.99804	36	0.97692	46	0.97995	56	0.99126
7	0.98946	17	0.98809	27	0.99798	37	0.97354	47	0.98132	57	0.98424
8	0.97953	18	0.83334	28	0.99515	38	0.98412	48	0.96624	58	0.97153
9	0.99653	19	0.98382	29	0.99612	39	0.98357	49	0.9616	59	1

Model Output Analysis (6-ensemble)

💡 *Insight*

mAP가 98% 이상으로 올라가지 않는 이유는 Val Dataset에 Missing label이 존재하기 때문

🔍 *Analysis*

Val Dataset에 18번째 Class (Football Field)가 있는데 없다고 표시된 것을 직접 확인

[Missing Label]

Inference Time Reduction Strategy

📌 *Enhancing efficiency in model inference*

Half precision (FP16)을 사용해 성능 저하 없이 Inference speed를 70% 늘릴 수 있음

	FP32	FP16	diff.
Inference time [sec]	493	289	-41.38%
Inference speed [img/sec]	101.85	173.75	+70.59%
Public mAP	0.97618	0.97614	-0.00410%
Private mAP	0.97707	0.97702	-0.00512%

Thank you!

University of Seoul
Jun Park, Minsu Kwon, Wonseok Choi

코드 & Weight는 Github에
공개되어 있습니다!