

# 2025 SSAFY AI 챌린지

모델 성능 향상을 위한 분석 및 개선 과정

A123\_쭈미정석

박준우(쭈누) 서미영 한정희 이주석

## CONTENTS

과제 개요 파악 및 모델 선정

---

데이터 중심 접근

---

모델 최적화 및 튜닝

---

결과 및 향후 과제

01

# 과제 개요 파악 및 모델 선정

## 01 과제 개요 파악 및 모델 선정

과제의 목적을 정확히 이해하고, 그에 맞는 특성을 가진 모델 선택하기

# [ VQA (Visual Question Answering) ]

### VQA

이미지 속 장면을 읽고 이해하여  
주어진 질문에 대해  
A/B/C/D 중 하나의 정답을 도출하는  
멀티모달 **Visual Question Answering**

### Qwen3-VL 계열 모델

시각적 인식 능력과 언어적 추론 능력이  
모두 뛰어난 Qwen3-VL 계열 모델

### Qwen3-VL-32B-Instruct

초기 모델 비교(4B vs 32B) 결과  
추론 성능이 우수한 **32B를 기반모델로 선정**

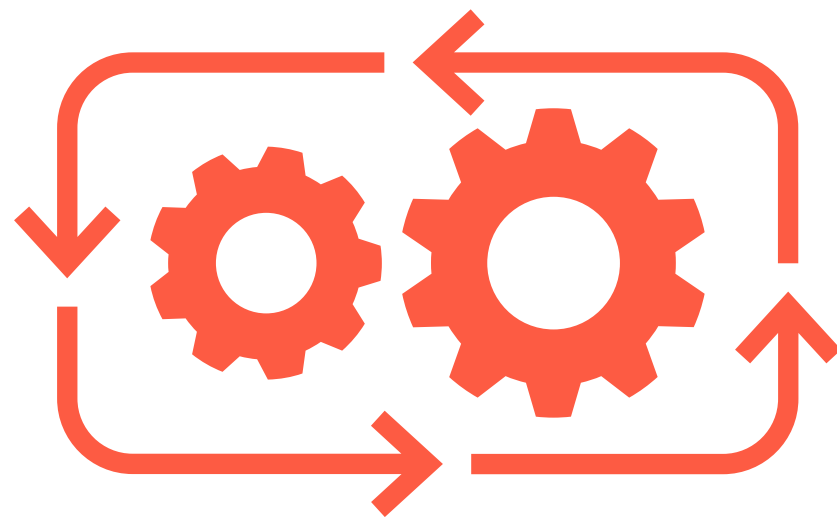
\*VL (Vision-Language): 이미지와 텍스트를 함께 처리

\*Instruct: '질문에 답하도록' 추가로 튜닝 (질문-답변 형태)

02

# 데이터셋 확인

## 02 데이터셋 확인



**Data Governance**

### 데이터 학습에 기반한 기술인 AI

단순히 모델을 개선하는 것보다,  
데이터를 얼마나 체계적이고 신뢰성 있게 관리하느냐가 더 중요

→ 모델 성능의 핵심인 **데이터 품질** 확보를 위해, **학습 전 데이터셋 분석** 선행

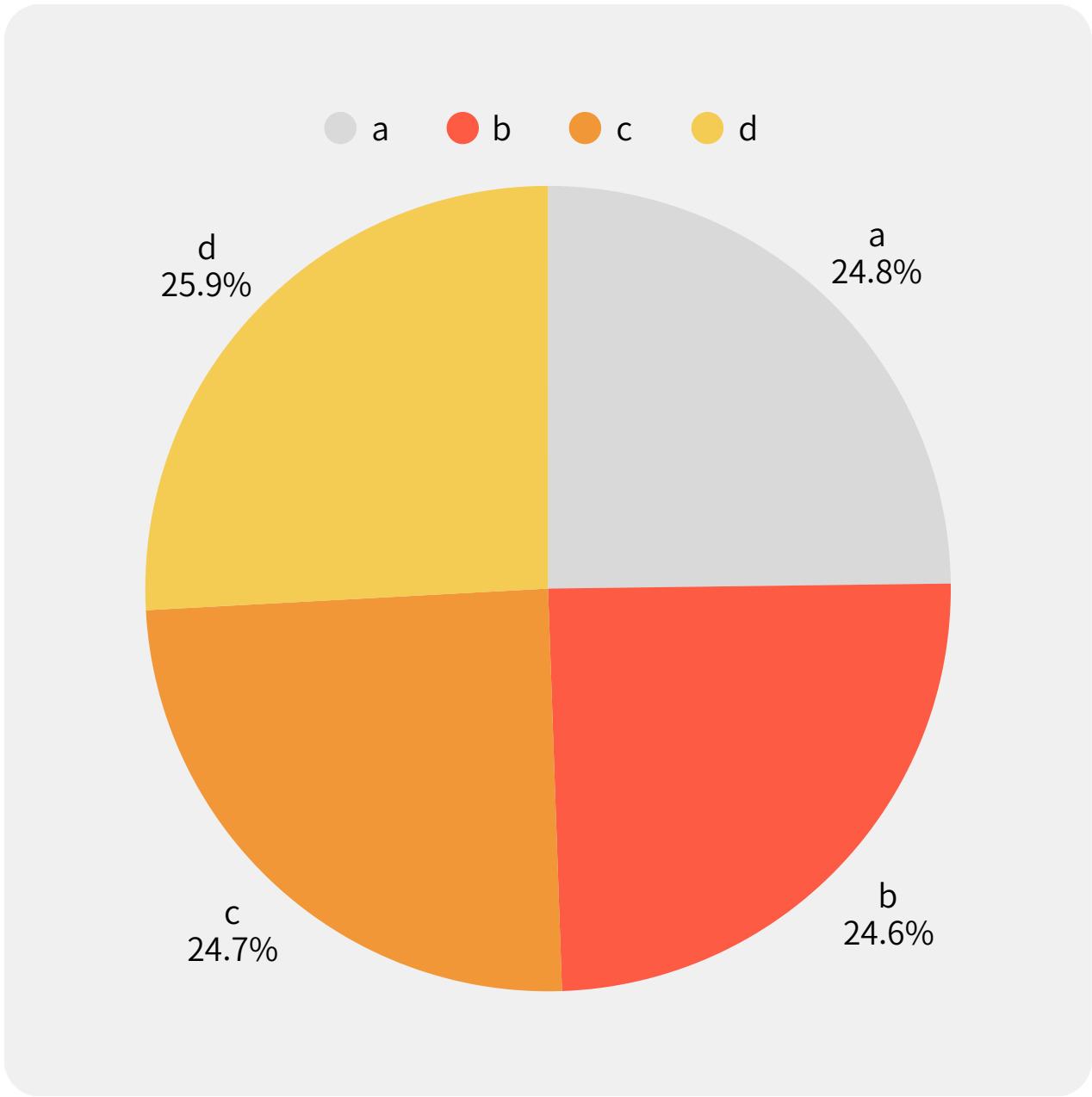
## O2-1 데이터 불균형 여부 확인 결과 “균등 분포”

데이터 불균형(오버샘플링/언더샘플링 필요) 여부를 판단하기 위해, 라벨 분포의 균형 상태 확인

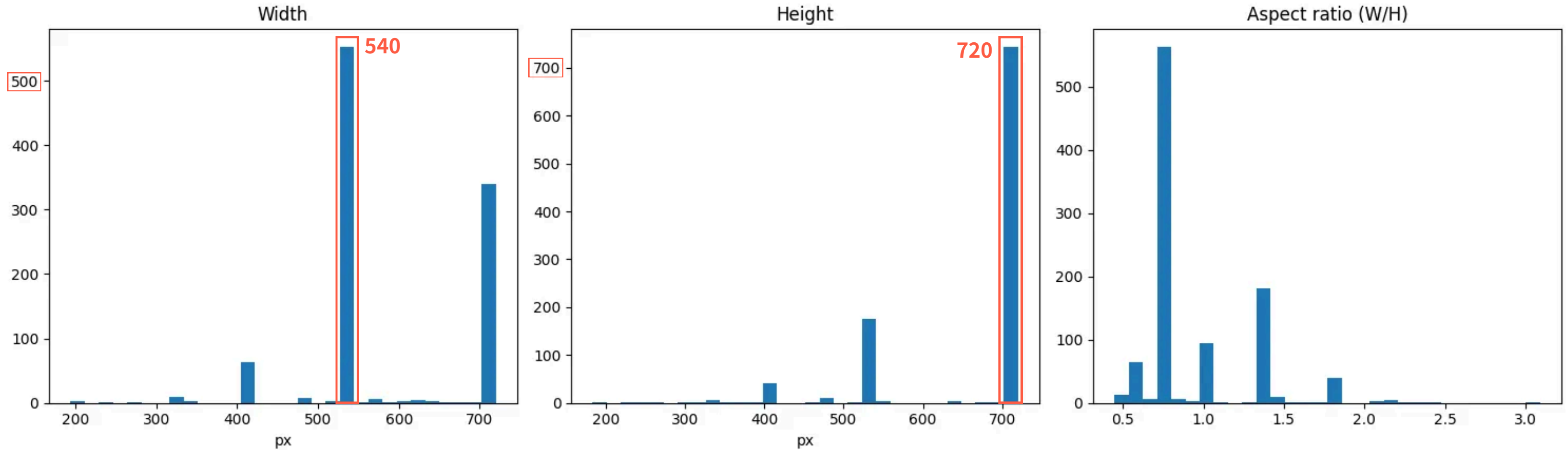
Answer Distribution (Train)



Answer Proportions (Train)



## 02-2 이미지 해상도(가로, 세로 사이즈) 확인



### 동적 해상도 (Dynamic Resolution)

모델의 AutoProcessor가 스스로 이미지의 크기를 판단해  
**원본 비율을 유지**한 채로 자동으로 리사이즈  
ex) 600 X 800 → 384 X 512

→ 모델이 사물의 형태나 위치 관계를 더 자연스럽게 이해

```
IMAGE_SIZE = 384
processor = AutoProcessor.from_pretrained (
    MODEL_ID,
    min_pixels=IMAGE_SIZE*IMAGE_SIZE,
    max_pixels=IMAGE_SIZE*IMAGE_SIZE,
    trust_remote_code = True,
)
```

초기 코드

→ 원본 이미지의 비율 무시, 시각적 정보 왜곡 발생 가능

```
IMAGE_SIZE = 384
processor = AutoProcessor.from_pretrained (
    MODEL_ID,
    trust_remote_code = True,
)
```

수정 후 코드



03

# 데이터셋 증강

03 데이터셋 증강

기존 이미지를 다양한 방식으로 변형해서, 모델이 더 폭넓은 상황을 학습할 수 있도록 도와주는 과정

1

VQA 과제의 특성

단순한 이미지 분류나 객체 검출과 달리,  
이미지 안의 위치 관계나 색상, 형태 단서가  
정답에 직접적으로 영향을 주는 과제

2

일반적인 증강의 위험성

VQA 특성을 고려한 설계  
과도한 회전, 색상 왜곡, 심한 잘라내기는  
정답의 핵심 단서를 훼손

3

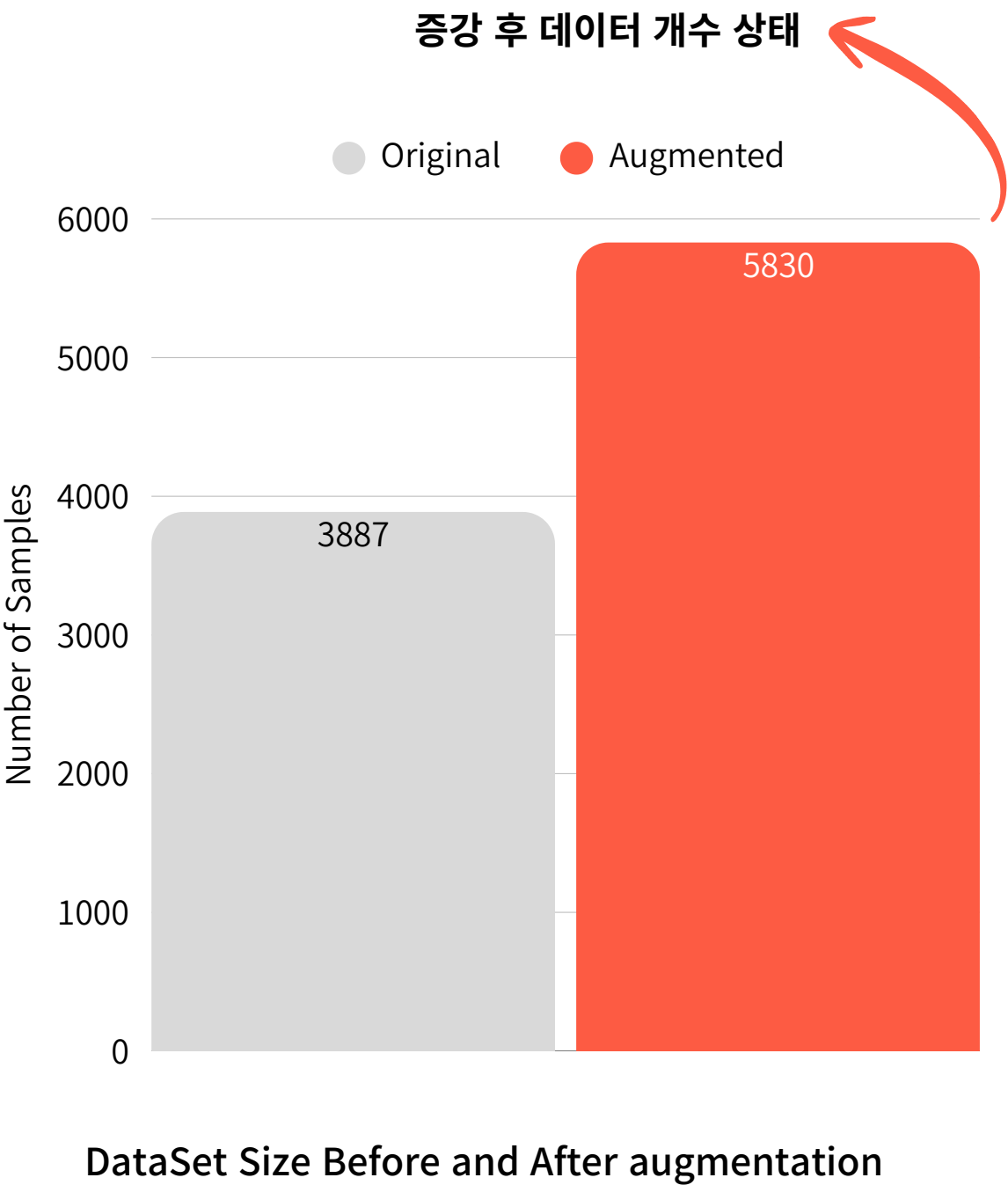
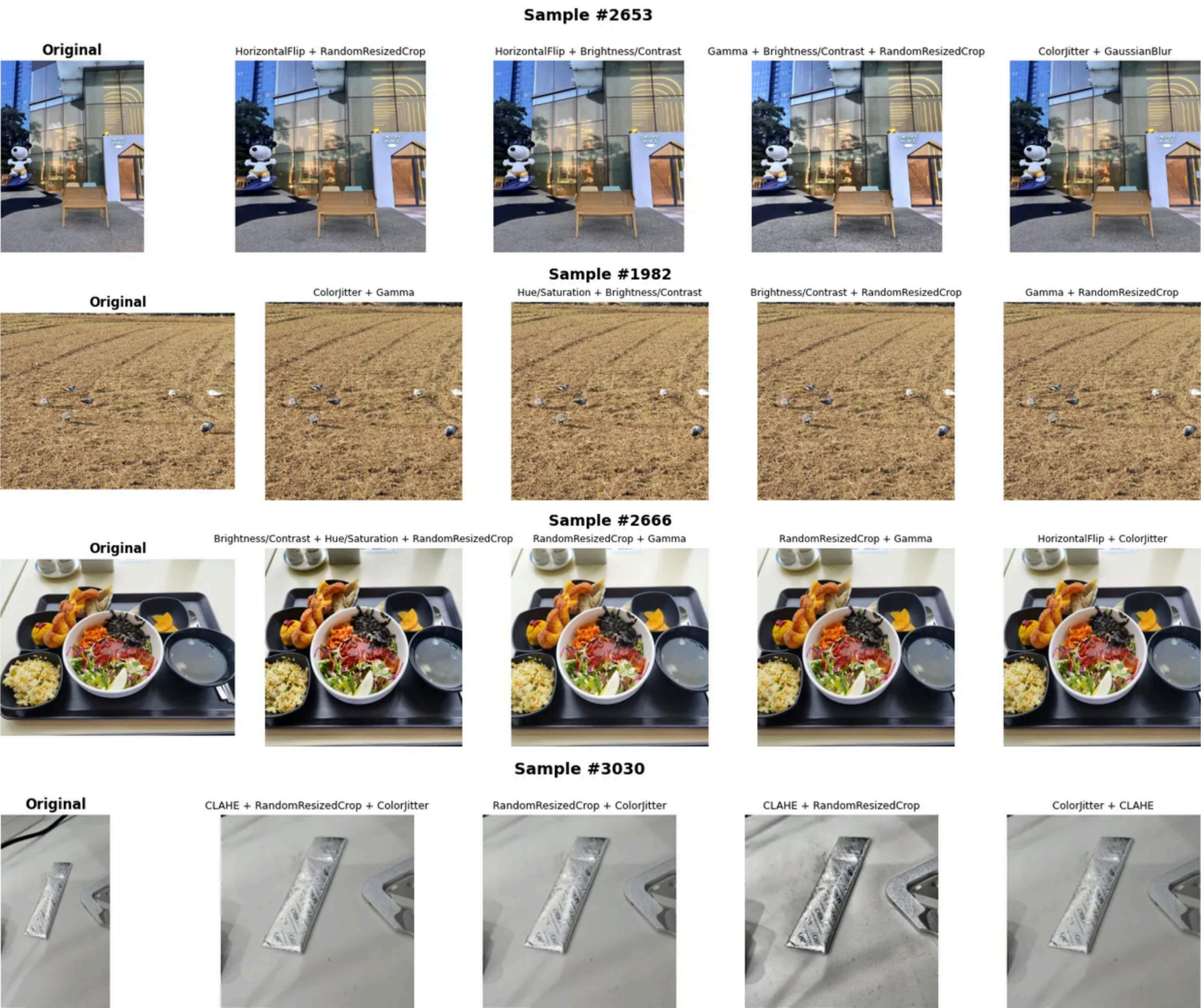
우리의 전략: 정보 보존

자연스러운 변형 위주로만 증강 적용,  
각 변환마다 서로 다른 확률 부여 후 랜덤 조합

적용 순서	변환 종류	핵심 파라미터	이미지 변화 (패턴)	적용 이유 (VQA 관점)
1	RandomResizedCrop	size=(384,384), scale=(0.9,1.0), ratio=(0.85,1.15), p=1.0	살짝 크롭 후 384×384로 리사이즈	구도·프레이밍 변화에도 강건하도록, 핵심 객체 보존을 위해 scale을 최소 0.9로 설정
2	ColorJitter (밝기/대비만)	brightness=0.08,contrast=0.08, saturation=0.0, hue=0.0, p=0.4	밝기/대비 ±8%만 변화 (색상/채도는 고정)	조명·노출 변화에는 강하게, 색상 단서는 유지
3	CLAHE	clip_limit=2.5, tile_grid_size=(8,8), p=0.3	어두운 영역 국소 대비 향상 → 작은 글자/패턴이 또렷	텍스트·패턴 등 세부 인식력 향상, 색 왜곡 없이 디테일 보존

→ 모델이 특정 구도나 조명 조건에 과적합되지 않고, 더 다양한 상황에 대응할 수 있도록 학습 데이터의 다양성을 높임

03 데이터셋 증강 - VQA 특성을 고려한 설계 결과



04

# 프롬프트 설정

## 04 프롬프트 설정: 기존 프롬프트 구성

# 모델 지시사항

SYSTEM\_INSTRUCT = (

"You are a helpful visual question answering assistant.\n"

"Return exactly one lowercase letter: a, b, c, or d.\n"

"Do not output any other text, punctuation, parentheses, or spaces."

)

# 프롬프트

def build\_mc\_prompt(question, a, b, c, d):

return ( f"Question: {question}\n"

f"(a) {a}\n(b) {b}\n(c) {c}\n(d) {d}\n"

"Choose the correct option and answer with one letter only: a b c d."

)

모델의 역할 설정 (VQA 전문 AI 역할)

출력 형식 강제


반드시 소문자 알파벳 하나만 출력

“답은 b입니다.”

같은 불필요한 텍스트 금지



## 04 프롬프트 추가 설정: 모델 지시사항



- 1. 역할 부여
- 2. 원하는 출력 형식 지시
- 3. 명확하고 구체적으로 묻기
- 4. 단계적 접근
- 5. 필요한 맥락 제공
- 6. Chain-of-Thought (분석 단계 명시)

SYSTEM\_INSTRUCT = (

# 1. 역할 부여 (가이드북 "역할 부여" 패턴 적용)

# - 명확한 전문가 역할 정의로 답변의 전문성과 일관성 확보

"You are an expert visual analysis AI specialized in multiple-choice questions.\n"

"Your capabilities include:\n"

"1. Precise image analysis\n"

"2. Contextual understanding\n"

"3. Logical reasoning\n\n"

# 2. 명확한 출력 형식 (가이드북 "원하는 출력 형식 지시" 원칙 적용)

# - 단일 문자만 출력하도록 강제하여 후처리 필요성 최소화

"OUTPUT REQUIREMENTS:\n"

"- Return exactly one lowercase letter (a, b, c, or d)\n"

"- No other text, punctuation, or explanations\n"

"- No spaces or line breaks\n\n"

# 3. 품질 기준 (가이드북 "명확하고 구체적으로 묻기" 원칙 적용)

# - 분석 품질 기준을 명시하여 더 신중한 답변 유도

"QUALITY STANDARDS:\n"


"1. Examine all visual details thoroughly\n"

"2. Consider the specific context of each question\n"

"3. Evaluate all options systematically\n"

"4. Choose the single most accurate answer based on visual evidence" )

## 04 프롬프트 추가 설정: 프롬프트 템플릿



1. 역할 부여
2. 원하는 출력 형식 지시
3. 명확하고 구체적으로 묻기
4. 단계적 접근
5. 필요한 맥락 제공
6. Chain-of-Thought (분석 단계 명시)

```
def build_mc_prompt(question, a, b, c, d):
```

```
# 1. 단계적 접근 (가이드북 "단계적 접근" 패턴 적용)  
# - 분석 과정을 명확한 단계로 구분하여 더 체계적인 답변 유도
```

```
return (
```

```
    "TASK: Visual Question Analysis\n\n"
```

```
# 2. 맥락 제공 (가이드북 "필요한 맥락 제공" 원칙 적용)
```

```
# - 질문과 선택지를 구조화하여 제시
```

```
f"QUESTION TO ANALYZE:\n{question}\n\n"
```

```
f"OPTIONS TO EVALUATE:\n"
```

```
f"a) {a}\n"
```

```
f"b) {b}\n"
```

```
f"c) {c}\n"
```

```
f"d) {d}\n\n"
```

```
# 3. 분석 단계 명시 (가이드북 "Chain-of-Thought" 패턴 적용)
```

```
# - 모델의 분석 과정을 단계별로 안내
```

```
"ANALYSIS STEPS:\n"
```

```
"1. Examine all visual elements in the image\n"
```

```
"2. Understand the specific requirements of the question\n"
```

```
"3. Consider each option against the visual evidence\n"
```

```
"4. Select the most accurate answer\n\n"
```

```
# 4. 최종 출력 형식 재강조
```

```
# - 마지막에 한 번 더 출력 형식을 강조하여 정확한 응답 유도
```

```
"RESPONSE FORMAT:\n"
```

```
"Provide exactly one lowercase letter (a, b, c, or d) representing  
the most accurate answer." )
```

05

# 하이퍼 파라미터 튜닝



## 05 하이퍼 파라미터 튜닝

### 하이퍼 파라미터

학습 중에 모델이 스스로 바꾸는 값이 아니라, 학습 이전에 사용자가 직접 설정해주는 값  
→ 하이퍼 파라미터는 머신러닝 모델의 성능에 중대한 영향을 미치며, 이를 조절하는 과정을 **하이퍼 파라미터 튜닝**이라고 합니다.

### 최적의 파라미터 조합 확인하기

구분	하이퍼 파라미터	역할 / 의미
LoRA 관련 하이퍼 파라미터	r	LoRA에서 사용하는 저차원 행렬의 랭크(rank). 작을수록 가벼워지고, 클수록 표현력 ↑
	lora_alpha	LoRA의 스케일링 계수( $\alpha$ ) — 저차원 업데이트의 세기를 조절
	lora_dropout	LoRA 모듈 내부에만 적용되는 드롭아웃 비율
학습 관련 하이퍼 파라미터	GRAD_ACCUM	Gradient Accumulation 스텝 수. 여러 미니배치의 기울기를 누적 후 한 번에 업데이트
	LR (learning rate)	파라미터 업데이트 속도를 결정하는 학습률
	Warm-up step	학습 초반 일정 step 동안 LR을 0→최대값까지 선형 증가

05 하이퍼 파라미터 튜닝

최적의 파라미터 조합 확인하기

파라미터	조합 1	조합 2	조합 3
r	8	12	16
lora_alpha	16	24	32
lora_dropout	0.05	0.07	0.1
GRAD_ACCUM	4	6	8
LR (learning rate)	5e-5	6e-5	8e-5
Warm-up step	0.03	0.05	0.1
제출 결과	0.95781	0.95987	0.95524

미세한 차이지만  
가장 높은 성능을 보인 파라미터 조합 2로 결정

## 05 하이퍼 파라미터 튜닝: 에폭(Epoch) 설정

에폭(Epoch)이란? 모델이 전체 데이터를 몇 번 반복 학습할지 사람이 직접 지정하는 값

### 일반화 성능 유지를 위해 Epoch 1로 설정

#### Epoch 1

안정적인 학습 (두 Loss가 비슷하게 유지)

#### Epoch 2

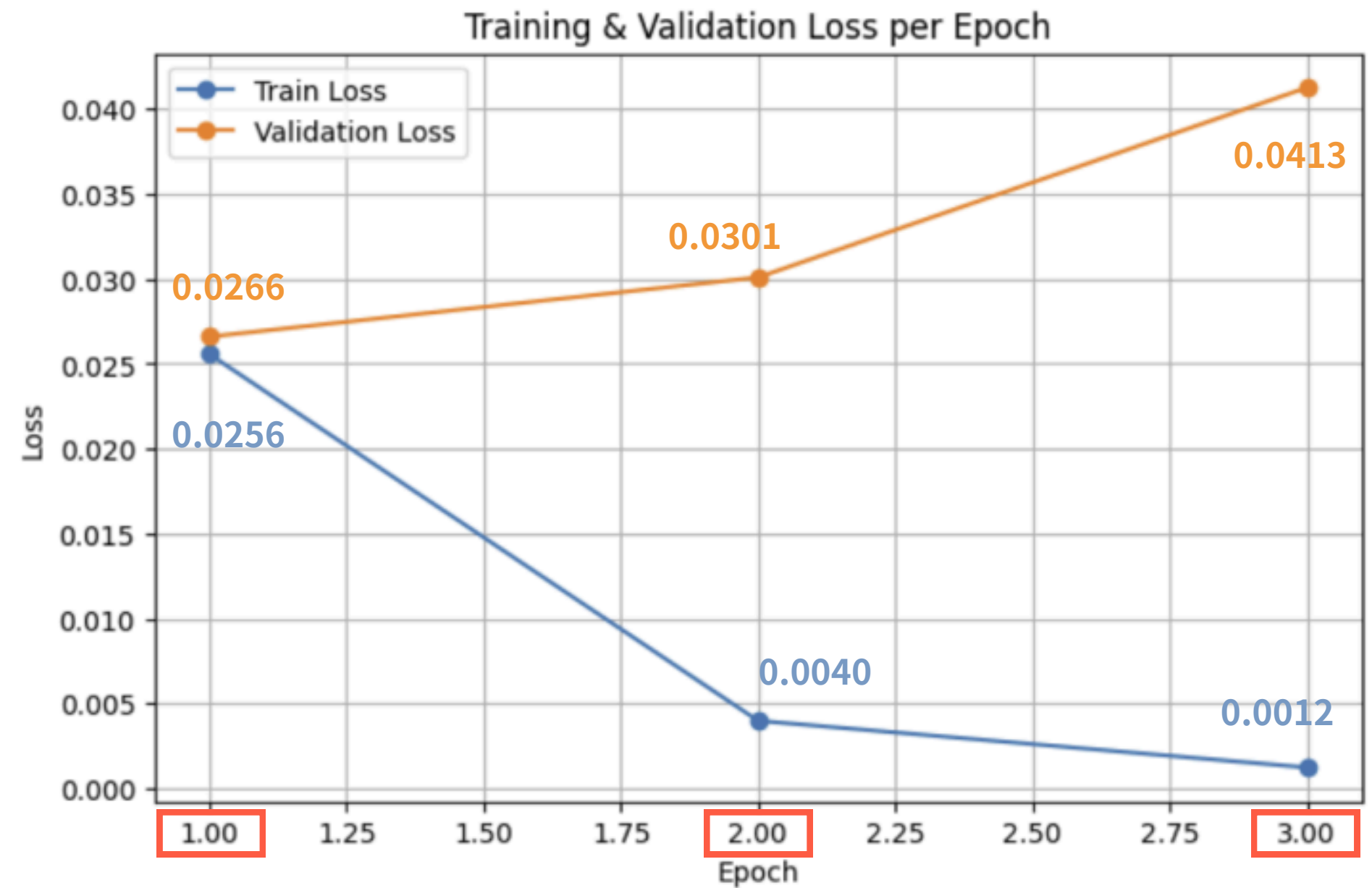
Train Loss ▼ Validation Loss ▲  
즉, 모델이 학습 데이터에 과도하게 맞춰짐

#### Epoch 3

Validation Loss ▲  
일반화 성능 저하, 과적합 발생

=== Epoch Loss Summary ===

Epoch	Train Loss	Validation Loss
1	0.0256	0.0266
2	0.0040	0.0301
3	0.0012	0.0413



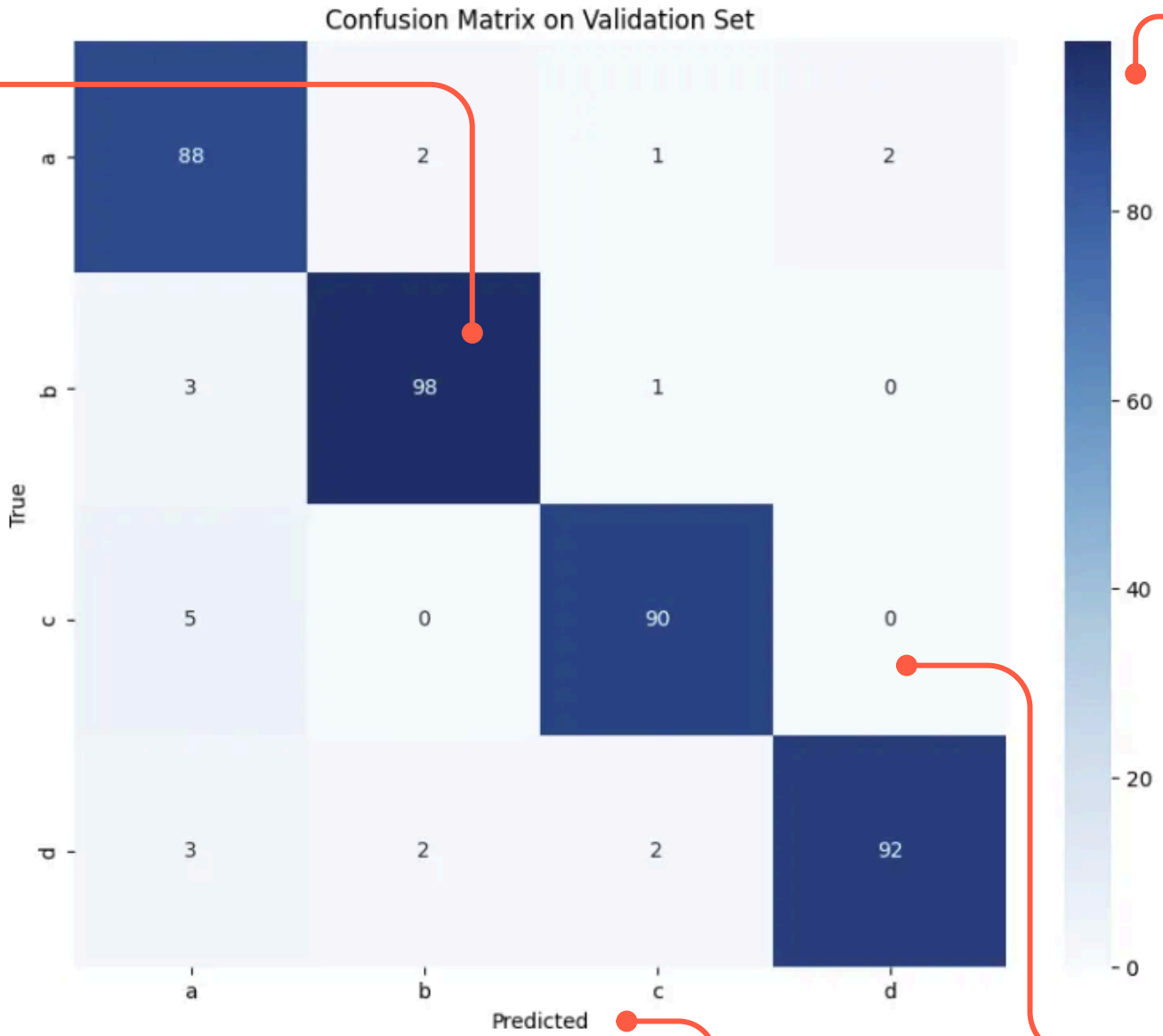
06

# 향후 과제 및 결과

06-1 추후 시도해보고 싶은 부분

검증 데이터셋에 대한  
모델의 예측 분포를 나타내는 **혼동행렬**

정답을 맞힌 개수  
값이 클수록 정확히 예측한 것



실제 정답  
세로축 (ground truth)

모델의 예측  
가로축 (prediction)

오답의 개수  
어떤 클래스가 어떤 클래스로  
잘못 예측됐는지 보여줌

대부분의 샘플이 대각선에 집중 → **정확한 예측 분포 확인**

06-2 추후 시도해보고 싶은 부분

분류 보고서(Classification Report)

Classification Report:

	precision	recall	f1-score	support
a	0.89	0.95	0.92	93
b	0.96	0.96	0.96	102
c	0.96	0.95	0.95	95
d	0.98	0.93	0.95	99
accuracy			0.95	389
macro avg	0.95	0.95	0.95	389
weighted avg	0.95	0.95	0.95	389

항목	관찰	해석
Accuracy = 0.95	95% 정확히 분류	전반적 성능 우수
Precision/Recall/F1 ≈ 0.95	균일한 성능	편향 없이 안정적 학습
Class a의 Precision = 0.89	일부 오분류 발생	시각적 유사-데이터 부족 가능
Macro ≈ Weighted avg	분포 균등	데이터 불균형 문제 없음


=== Error Pattern Analysis ===

- Q: 이 한식 세트 메뉴에 포함되지 않은 음식은 무엇인가요?  
True: c, Predicted: a
- Q: 이 음식 세트에 포함되지 않은 것은 무엇인가요?  
True: c, Predicted: a
- Q: 이 골목길에서 볼 수 없는 것은 무엇인가요?  
True: a, Predicted: c
- Q: 이 한식 상차림에서 보이는 주된 반찬은 무엇인가요?  
True: d, Predicted: b
- Q: 이 이미지는 어디에서 촬영된 것일 가능성이 가장 높을까요?  
True: a, Predicted: d

▶ 잘못 예측한 질문과 응답에 대한 출력 샘플

전반적으로 높은 성능을 보였으나, 라벨 a에서 약간의 오분류 발생

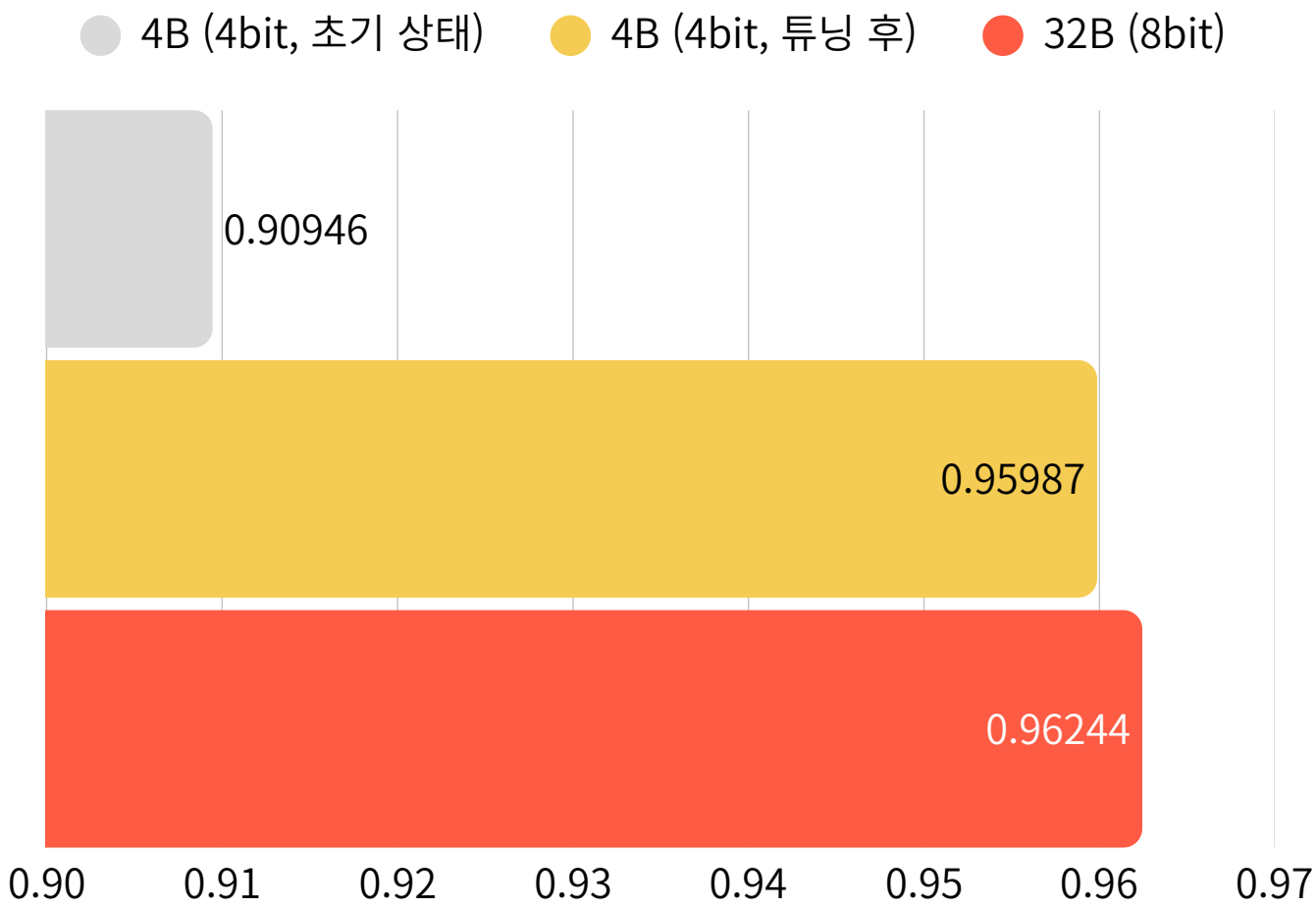
O6-3 2025 SSAFY 14기 AI 챌린지 [A123\_쭈미정석] 팀 결과

4	A171_4202122		0.96965	45	12m
5	A175_ACT		0.96862	24	2h
6	A172_할래말래		0.96759	11	14m
7	A205_지원이도집갈래		0.96759	32	1h
8	A176_미어캣트리오		0.96656	25	7m
9	A102_WeAreThere		0.96450	3	3d
10	A162_서현수대뇌피질연구소		0.96347	18	3m
11	C026_OBYG		0.96296	25	1h
12	A091_서울9반드림팀		0.96296	23	2h
13	A123_쭈미정석		0.96244	14	14h



Your Best Entry!  
Your most recent submission scored 0.96244, which is an improvement over your previous score of 0.95987. Great job!

[Tweet this](#)



마감 시간 기준 243팀 중 13위 달성

## A123\_쭈미정석

박준우(쭈누) 서미영 한정희 이주석