



더 감쪽같은 이미지 번역을 위해 (파파고 이미지/AR 번역 개발기)



노영빈 / 김승재 NAVER Papago

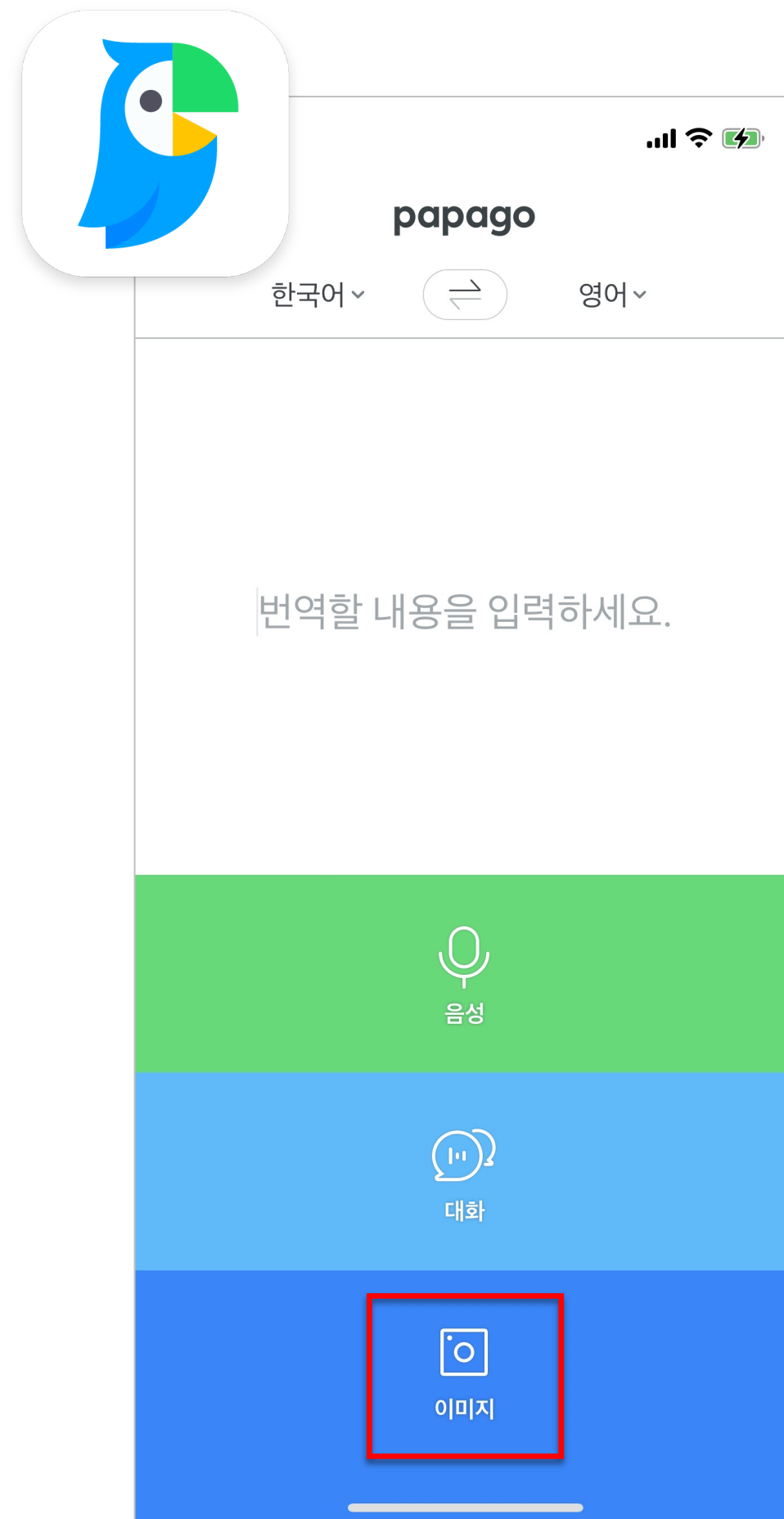
CONTENTS

1. 파파고 이미지 번역
2. 개선1: 리스트 형식의 텍스트 번역 고도화
3. 개선2: 텍스트 인페인팅 고도화
4. 더 나아가가기: 실시간 AR 이미지 번역

1. 파파고 이미지 번역

1.1 이미지 번역 소개

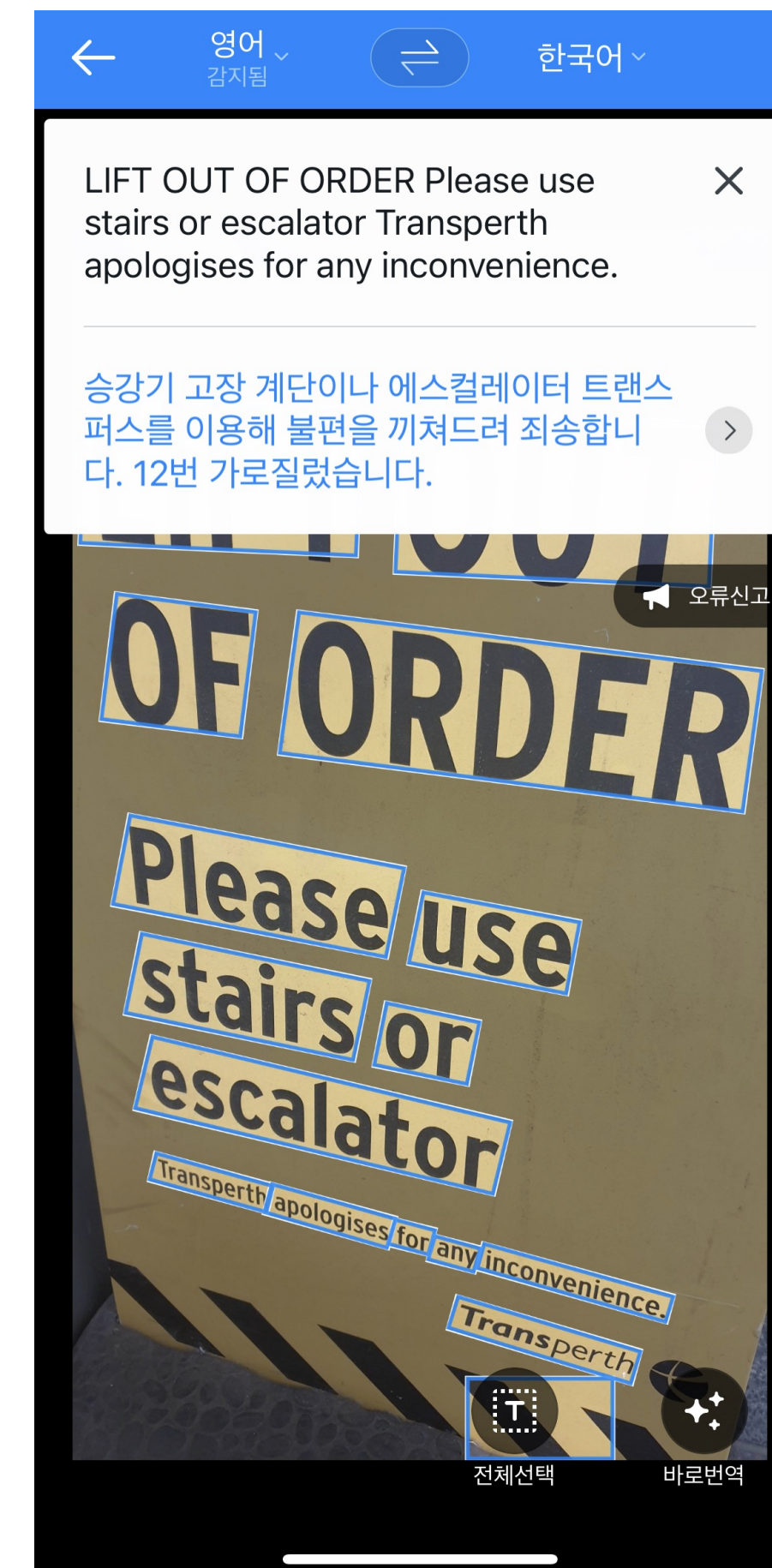
파파고 이미지 번역 분류



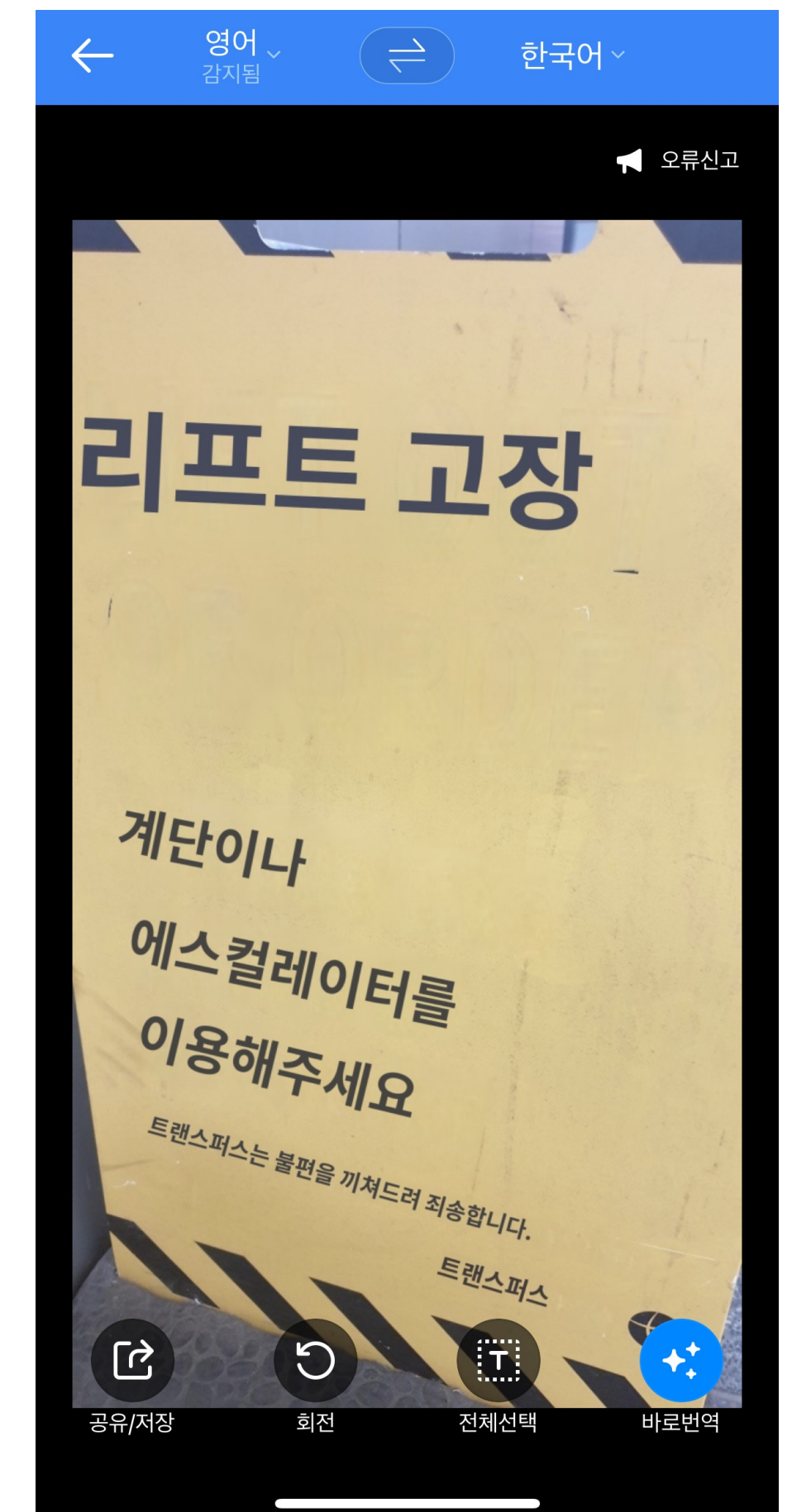
부분 번역



전체 번역

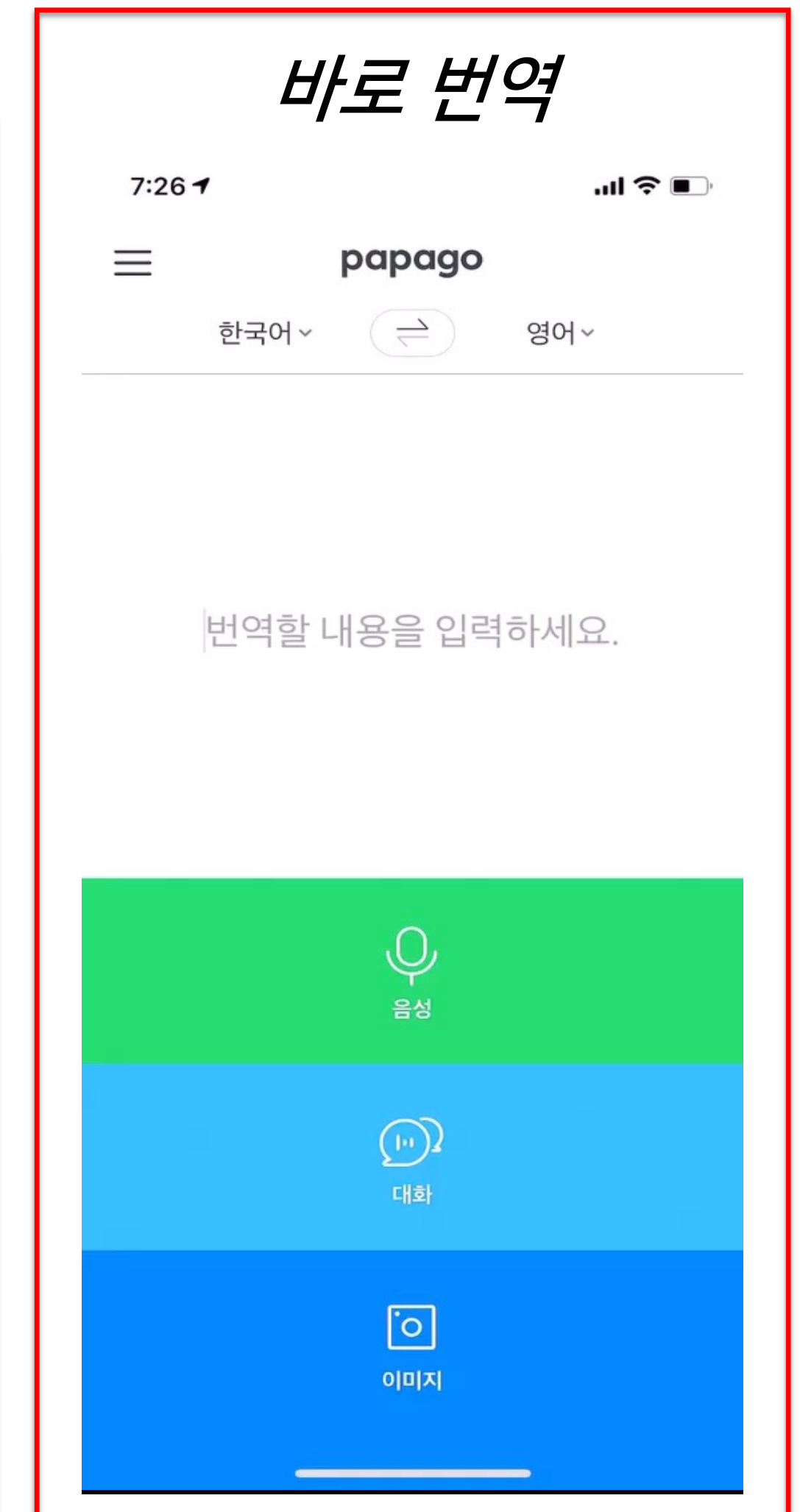
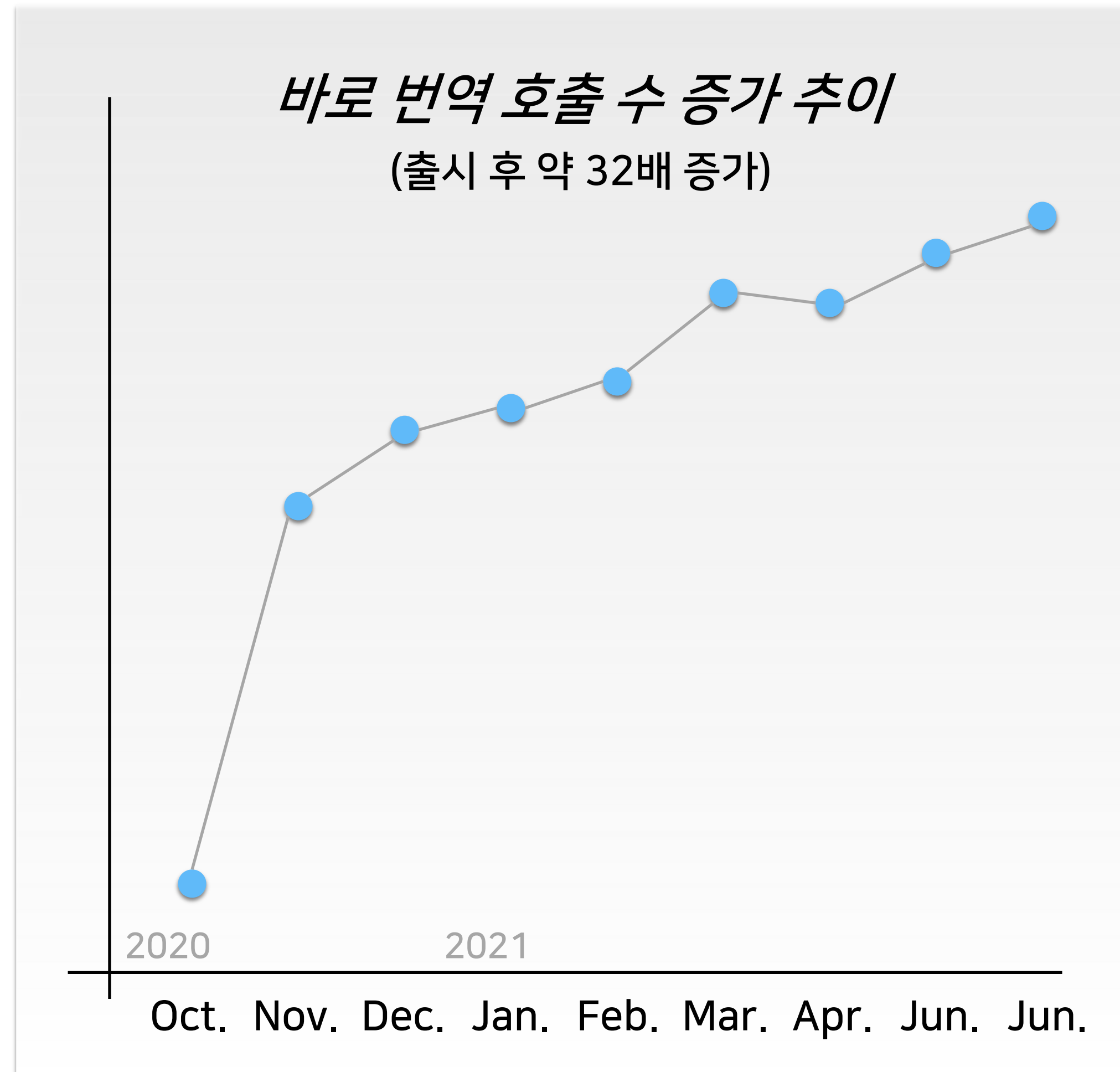
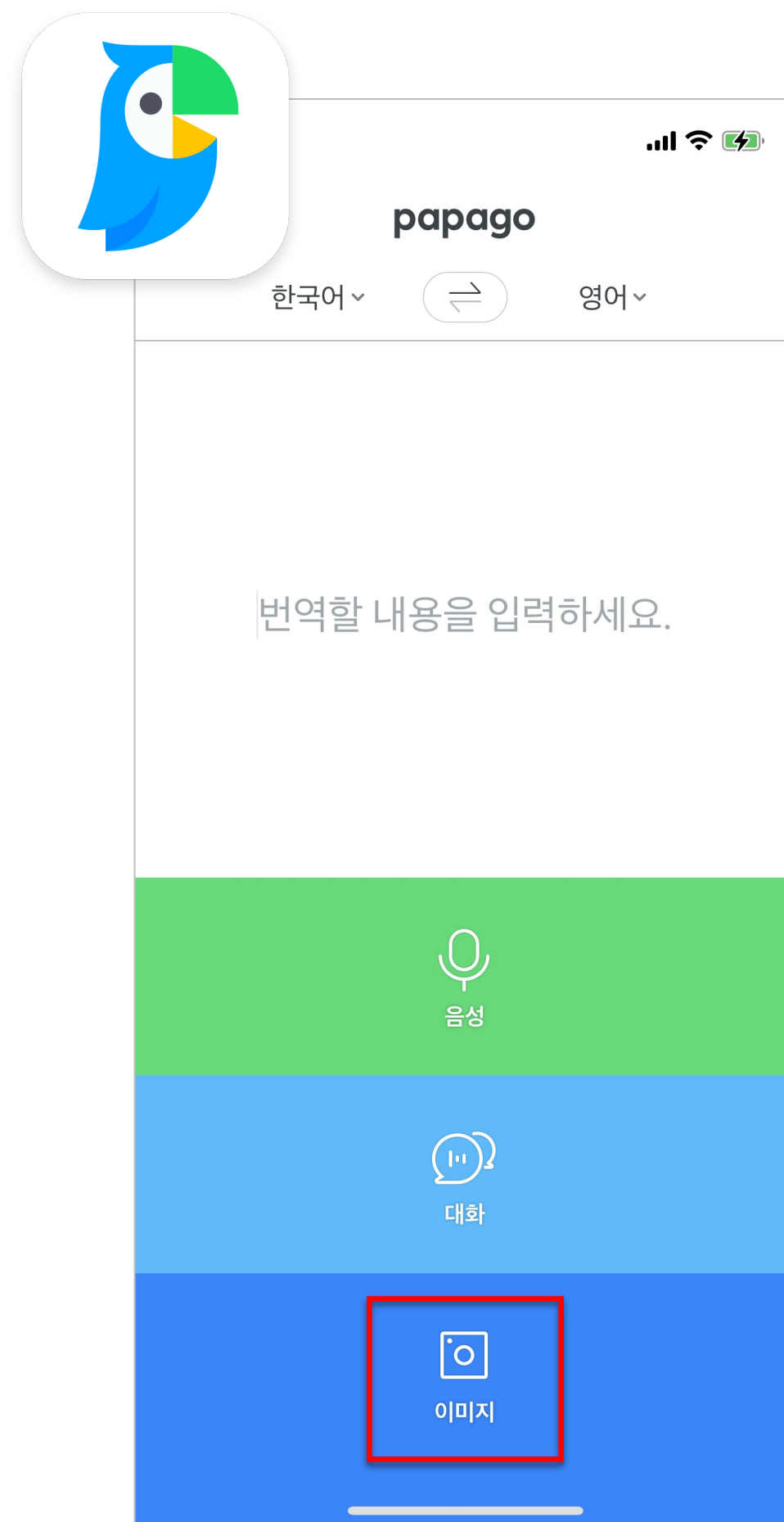


바로 번역



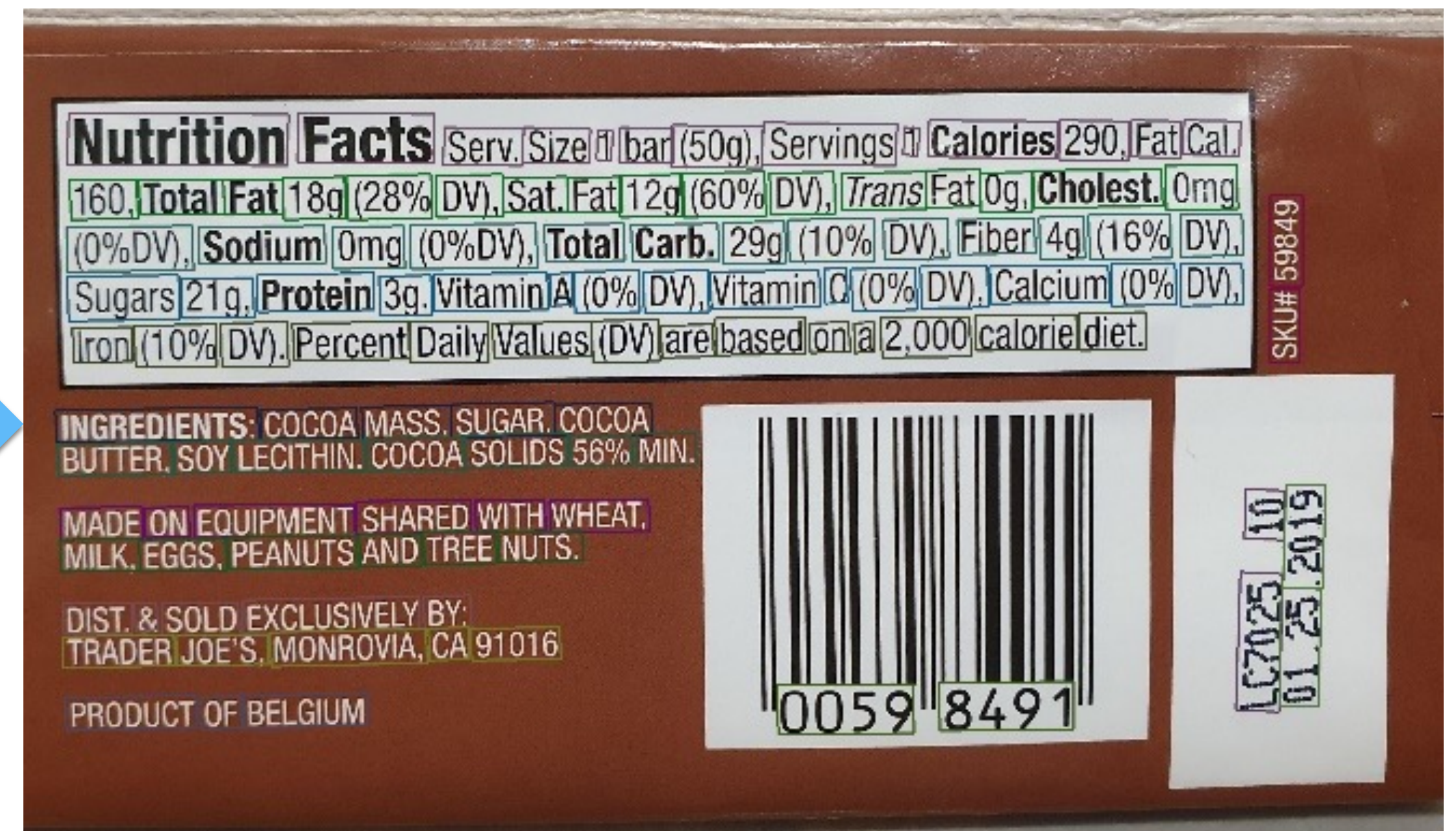
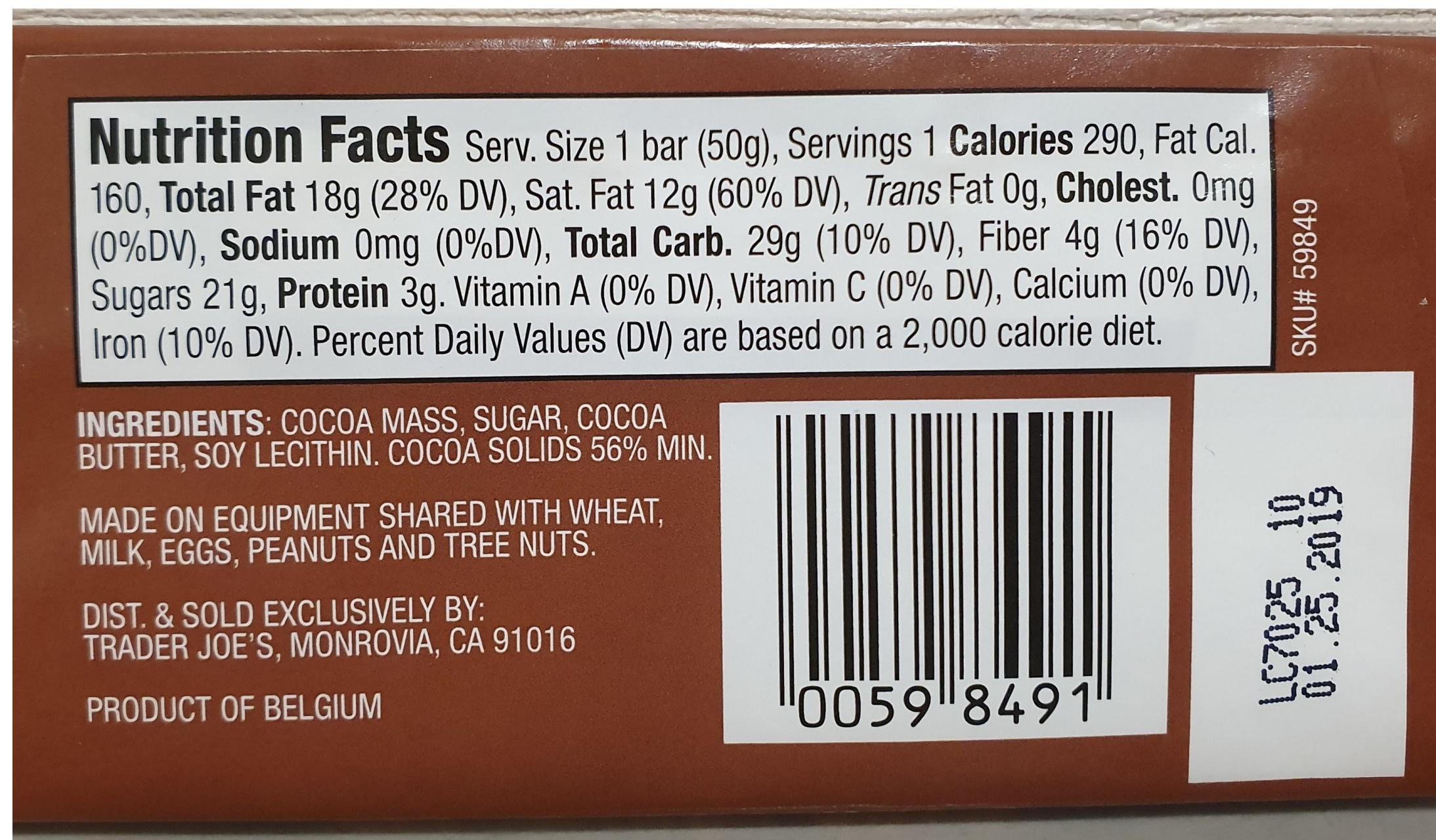
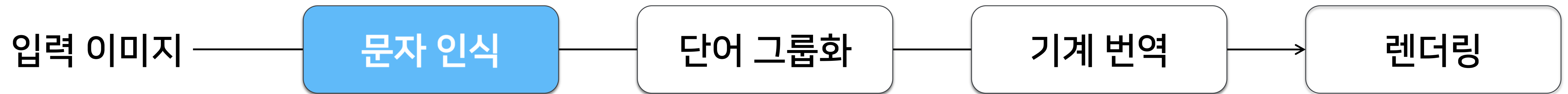
1.1 이미지 번역 소개

이미지 바로 번역을 중심으로 설명



1.2 바로 번역 과정 - 문자 인식

STEP1. 문자 인식(OCR)을 통해 텍스트를 검출합니다.



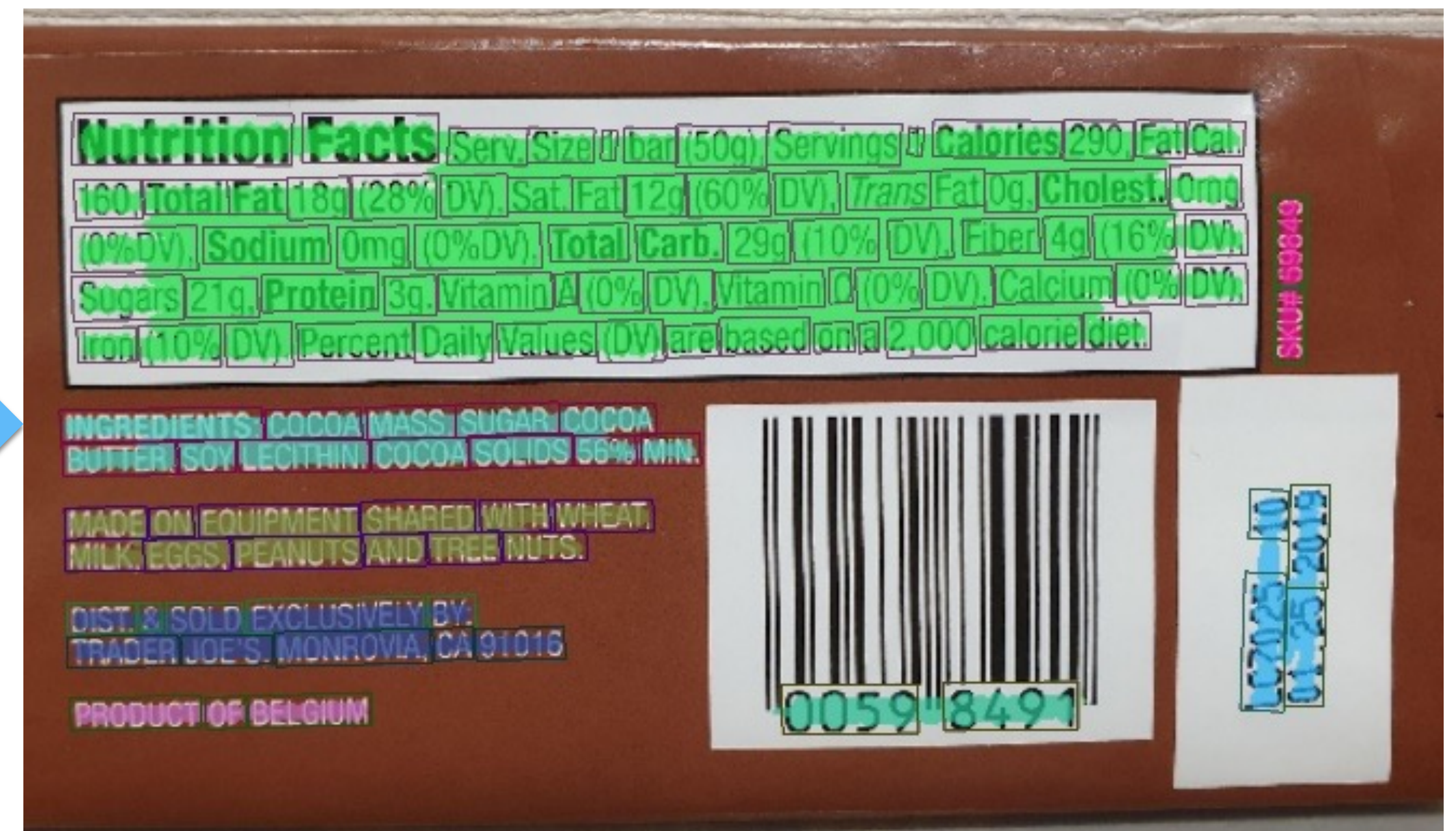
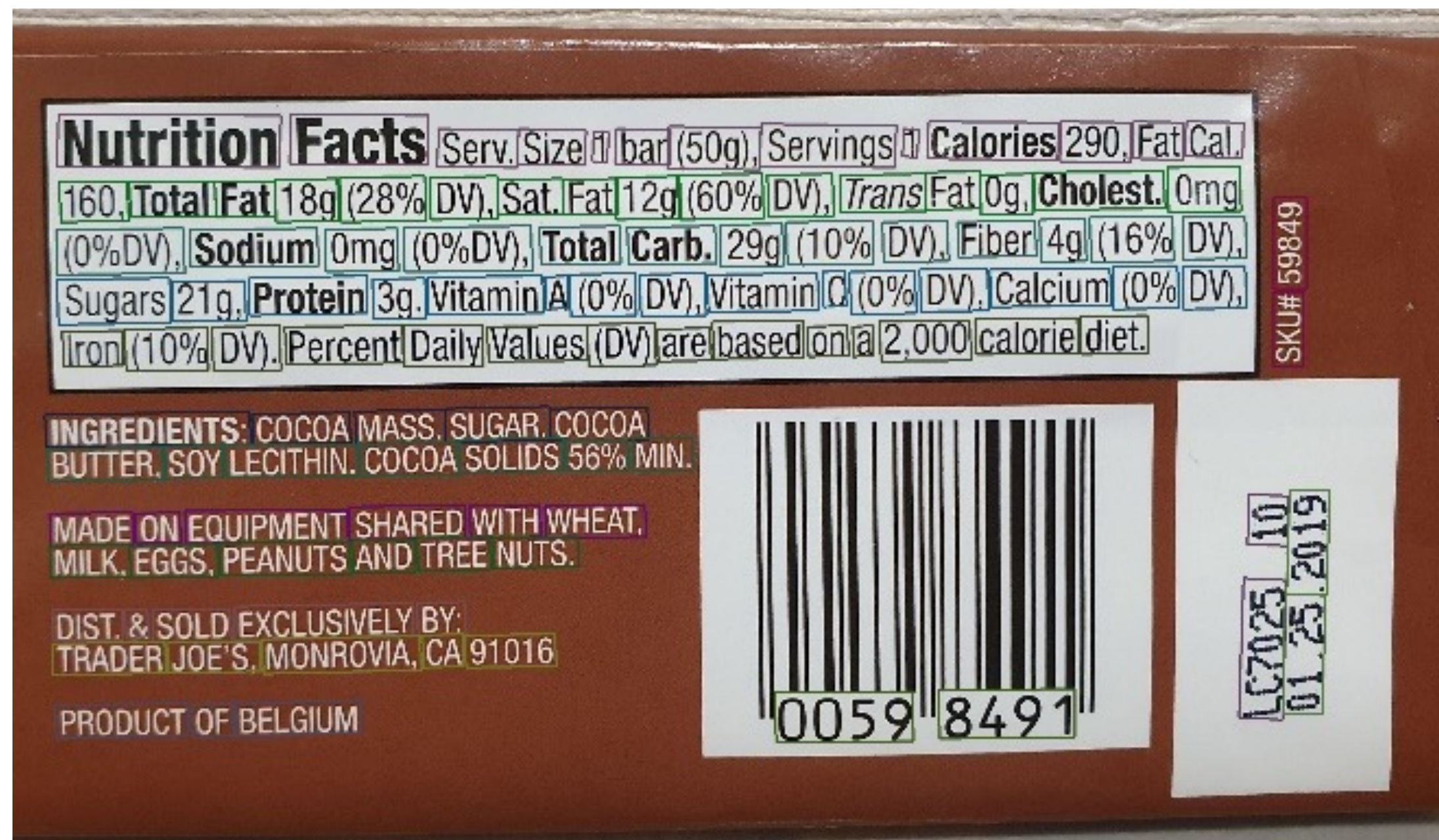
참조: [DEVIEW2019 <문자인식\(OCR\), 얼마나 정확하지?>](#)

1.2 바로 번역 과정 - 단어 그룹화

STEP2. HTS를 통해 같은 문장을 이루는 단어끼리 묶어줍니다.



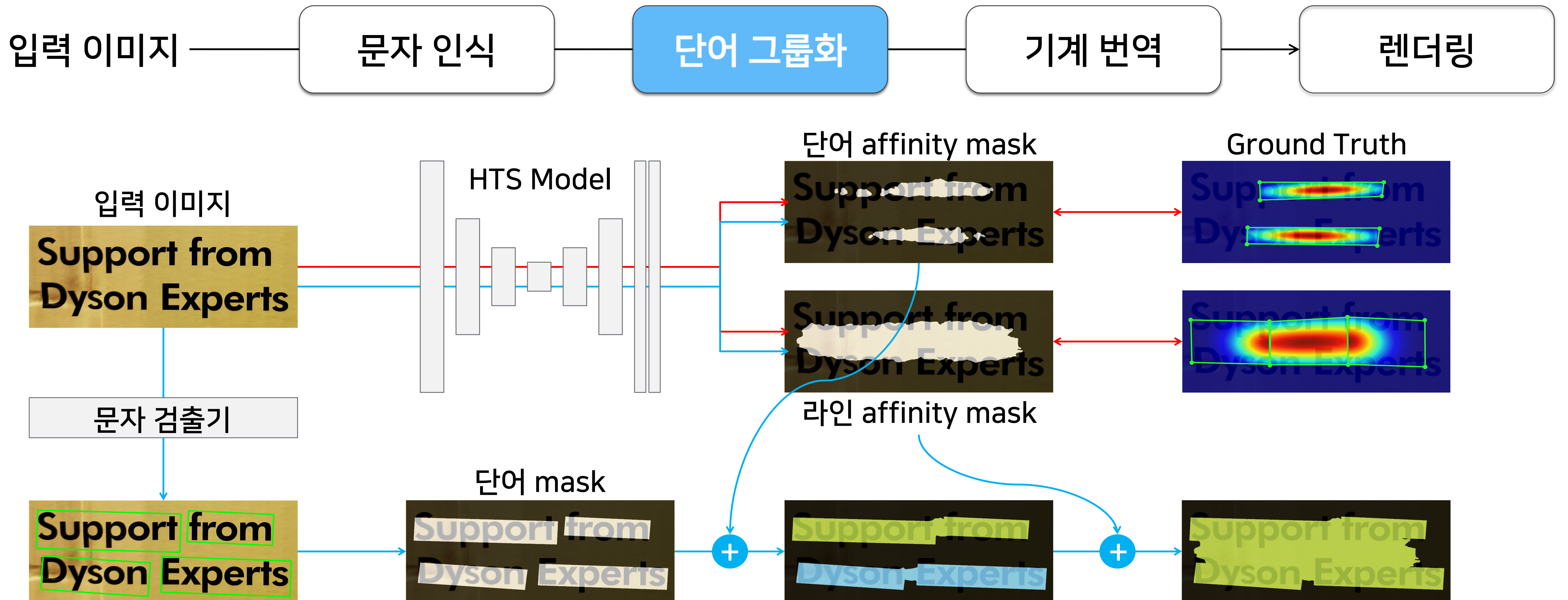
HTS: Hierarchical Text Structuring



참조: [DEVIEW2020 <외국어가 읽힌다 딱! \(파파고 이미지 번역\)>](#)

1.2 바로 번역 과정 - 단어 그룹화

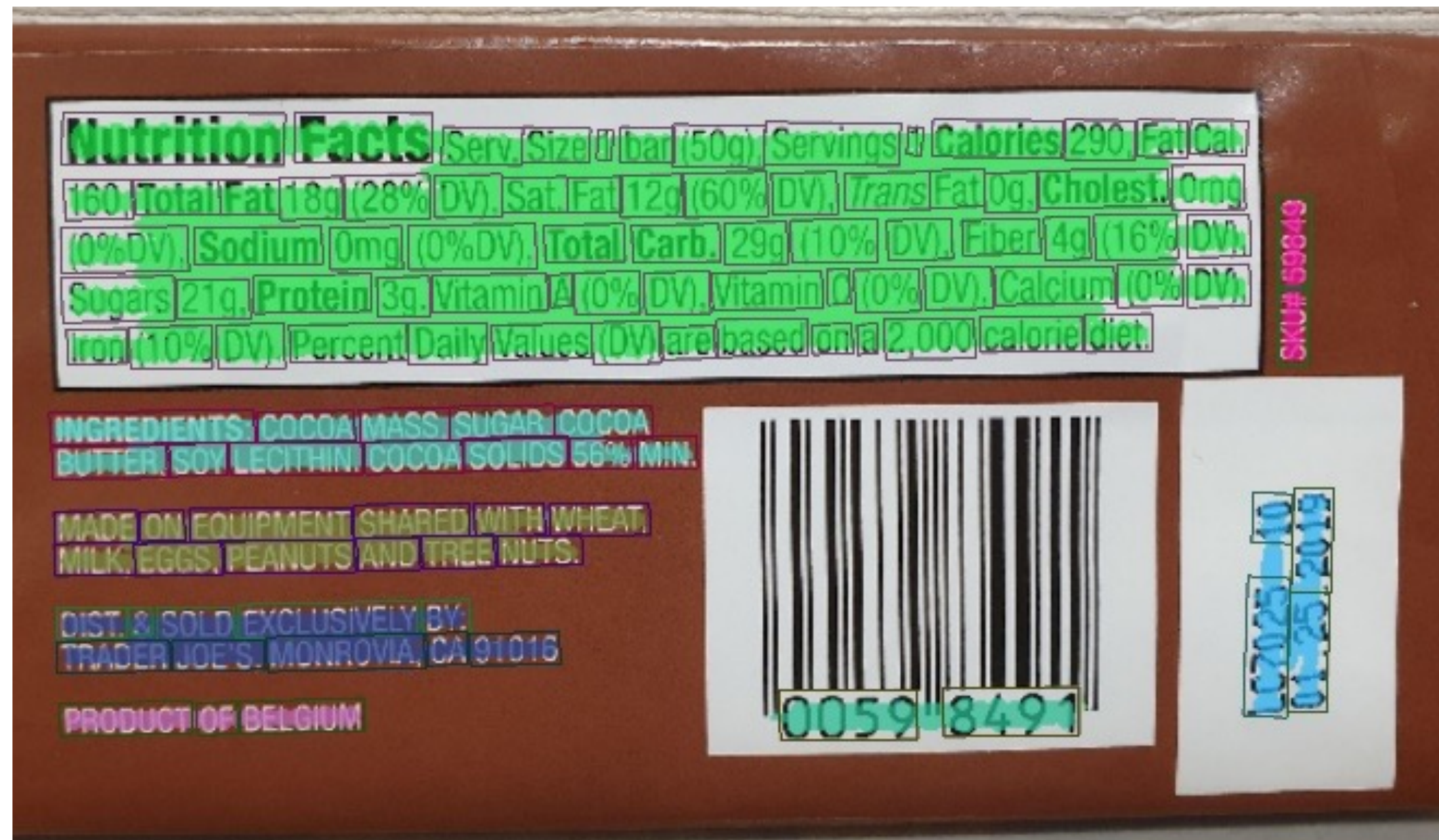
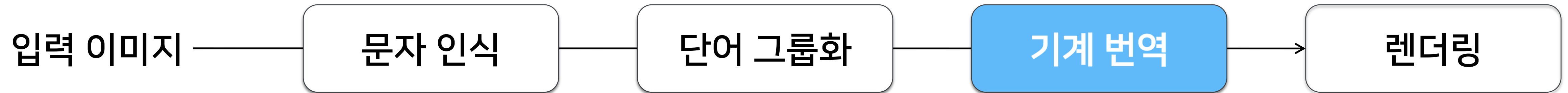
STEP2. HTS를 통해 같은 문장을 이루는 단어끼리 묶어줍니다.



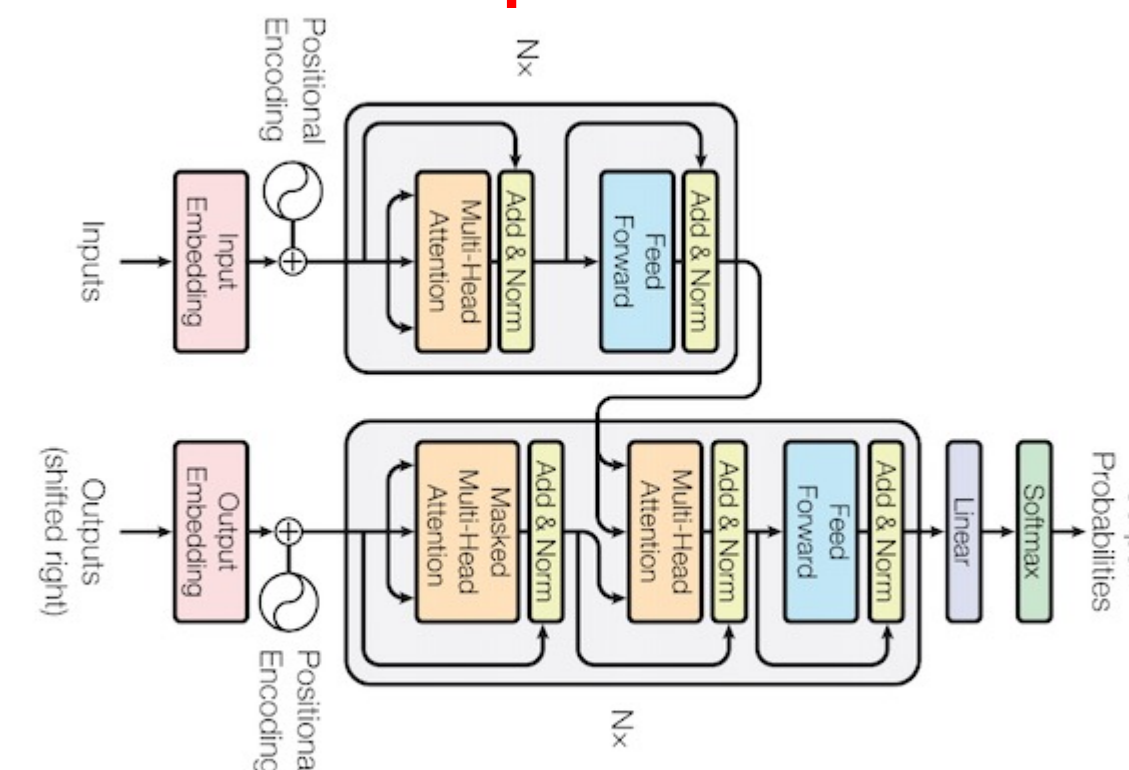
참조: [DEVIEW2020 <외국어가 읽힌다 딱! \(파파고 이미지 번역\)>](#)

1.2 바로 번역 과정 - 기계 번역

STEP3. 텍스트 블록별 기계 번역을 수행합니다.



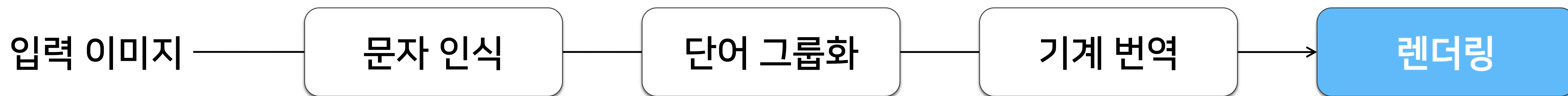
INGREDIENTS: COCOA MASS, SUGAR, COCOA BUTTER, SOY LECITHIN, COCOA SOLIDS 56% MIN



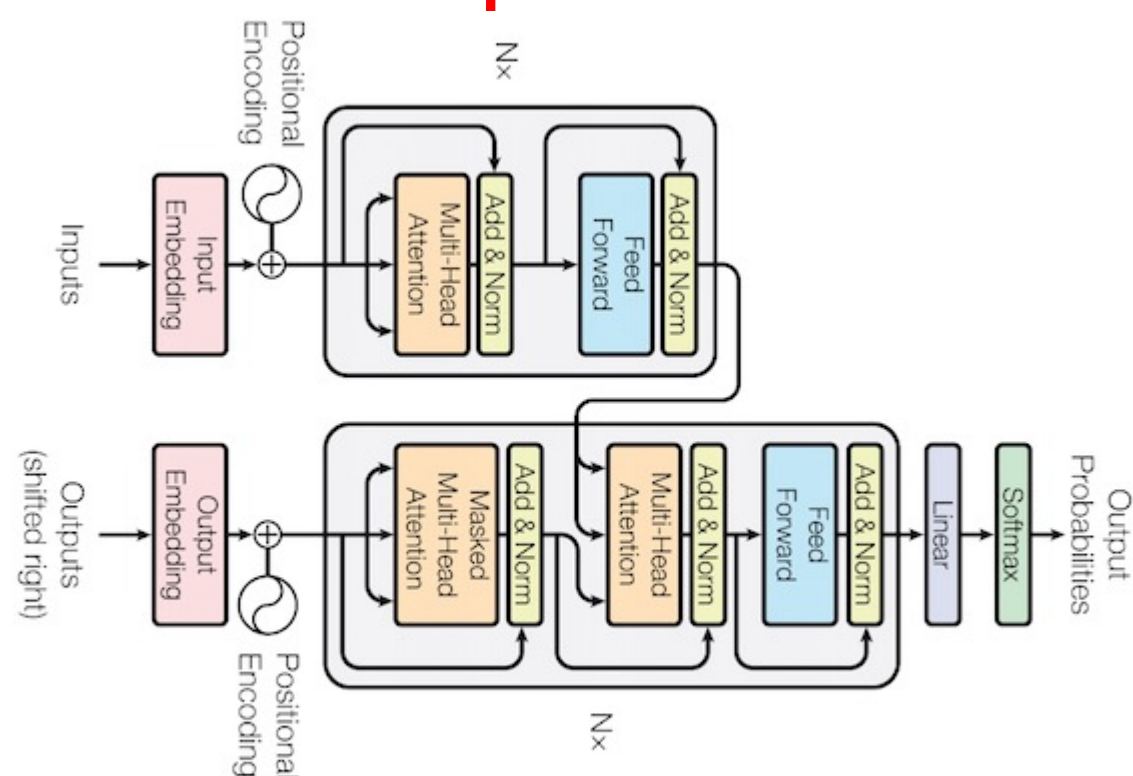
재료: 코코아 매스,
설탕, 코코아 버터, 콩
레시틴, 코코아
고형분 56%

1.2 바로 번역 과정 - 렌더링

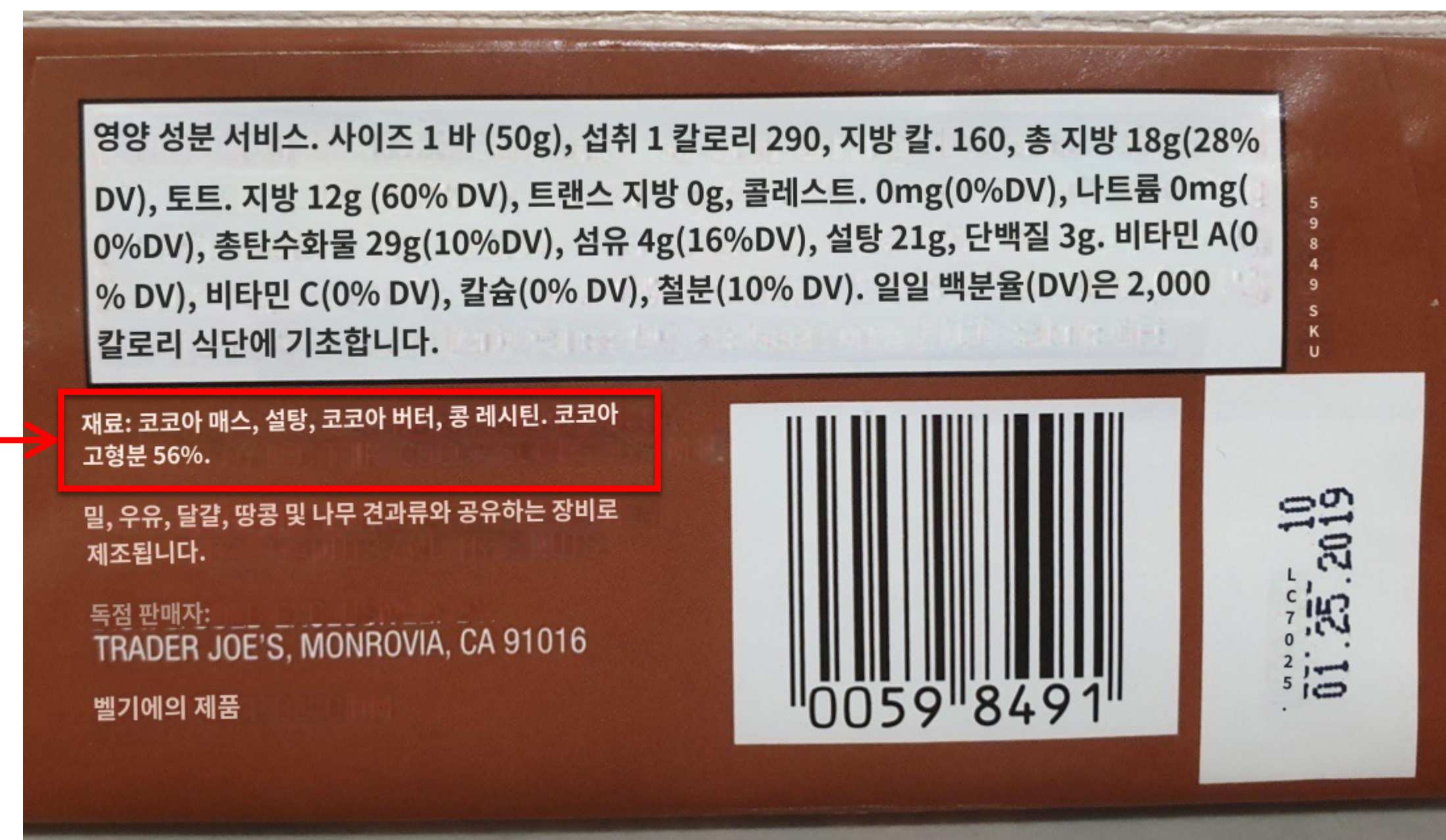
STEP4. 번역된 문장을 원본 이미지에 덧붙입니다.



INGREDIENTS: COCOA MASS, SUGAR, COCOA BUTTER, SOY LECITHIN, COCOA SOLIDS 56% MIN

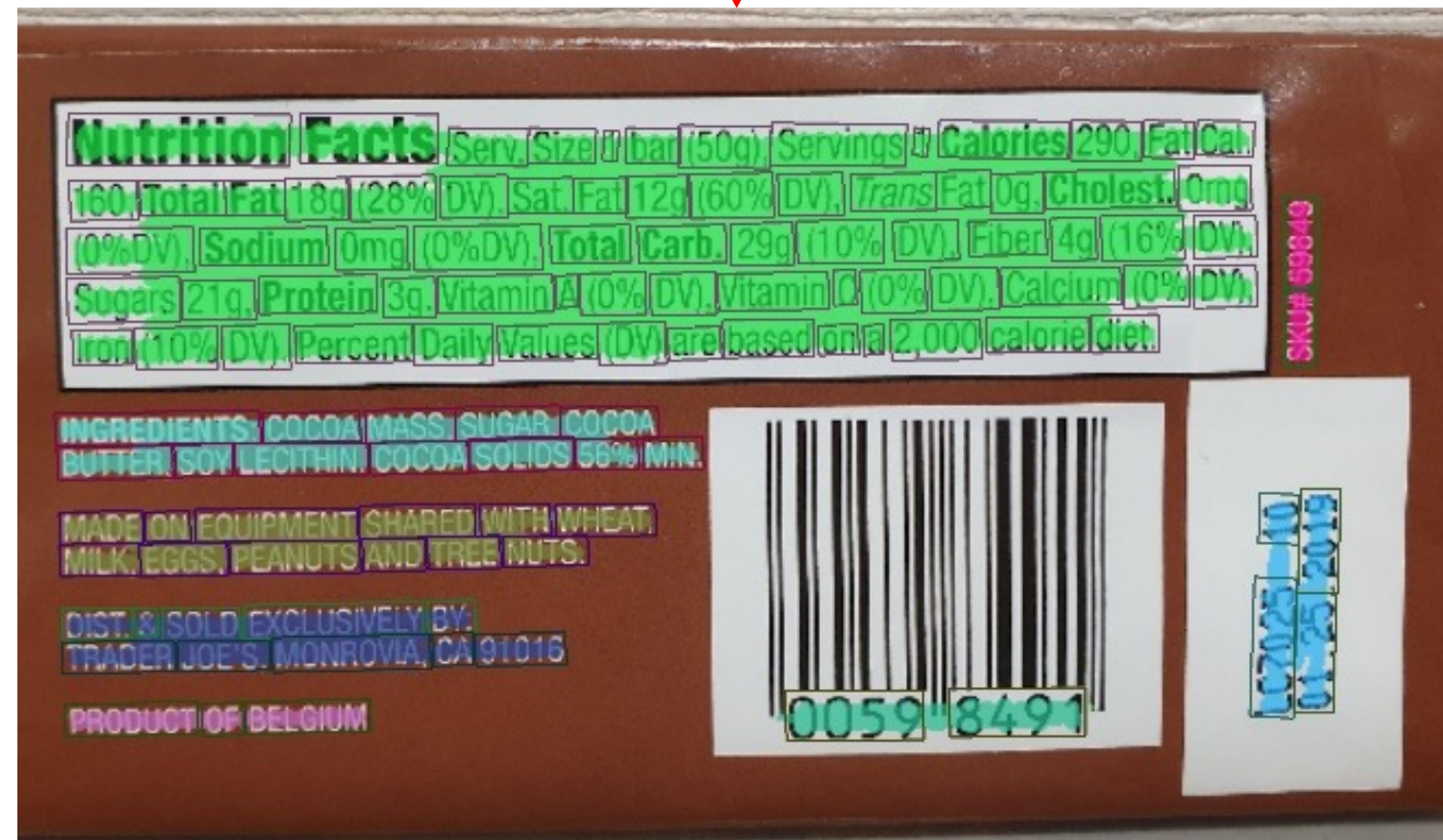
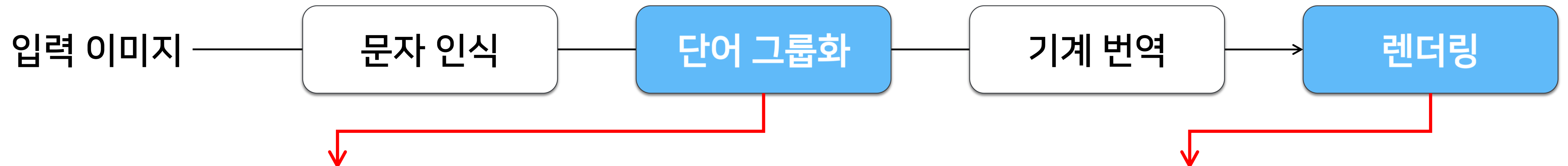


재료: 코코아 매스,
설탕, 코코아 버터, 콩
레시틴, 코코아
고형분 56%

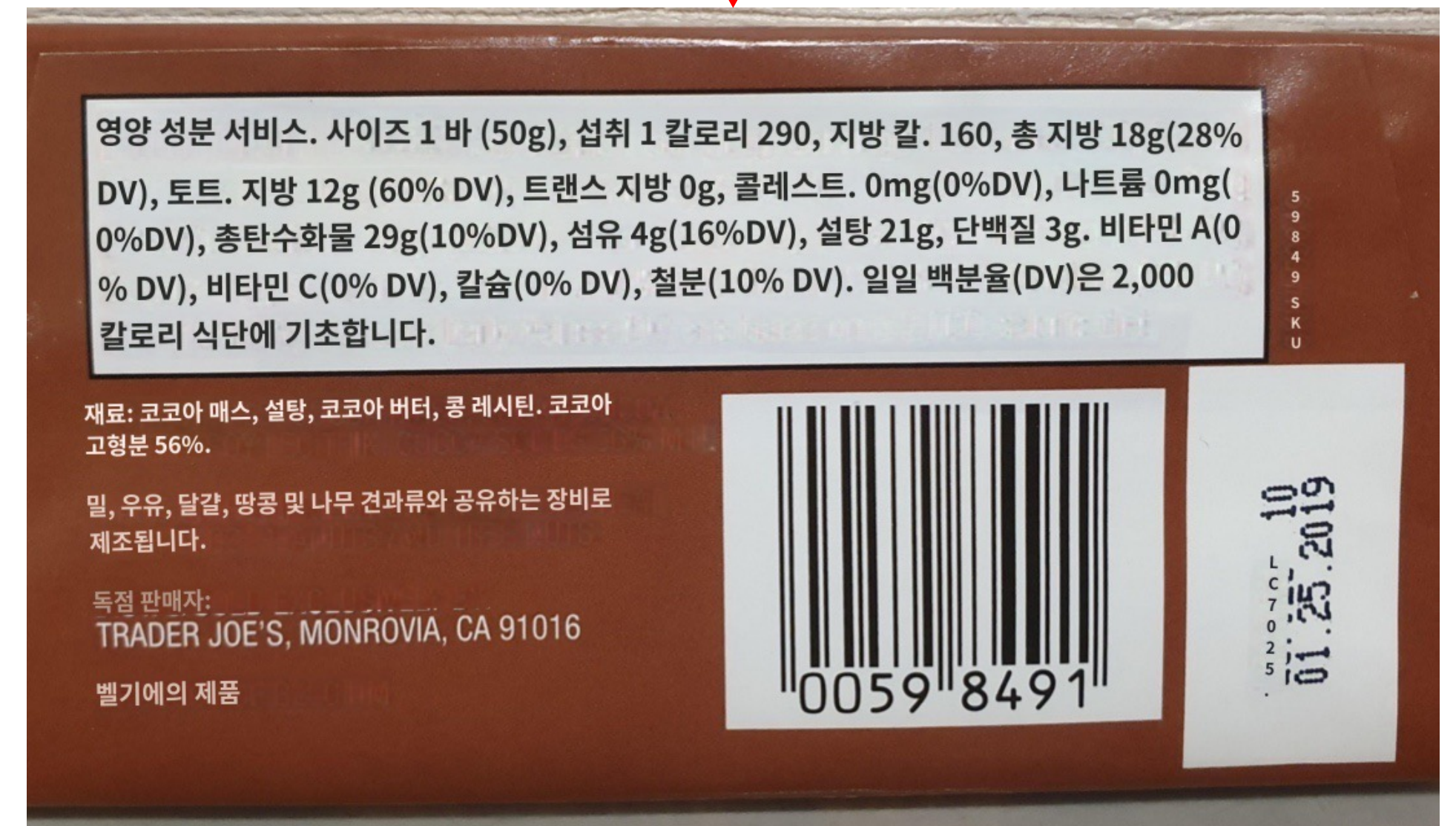


1.2 바로 번역 과정

바로 번역의 핵심은 단어 그룹화와 렌더링!



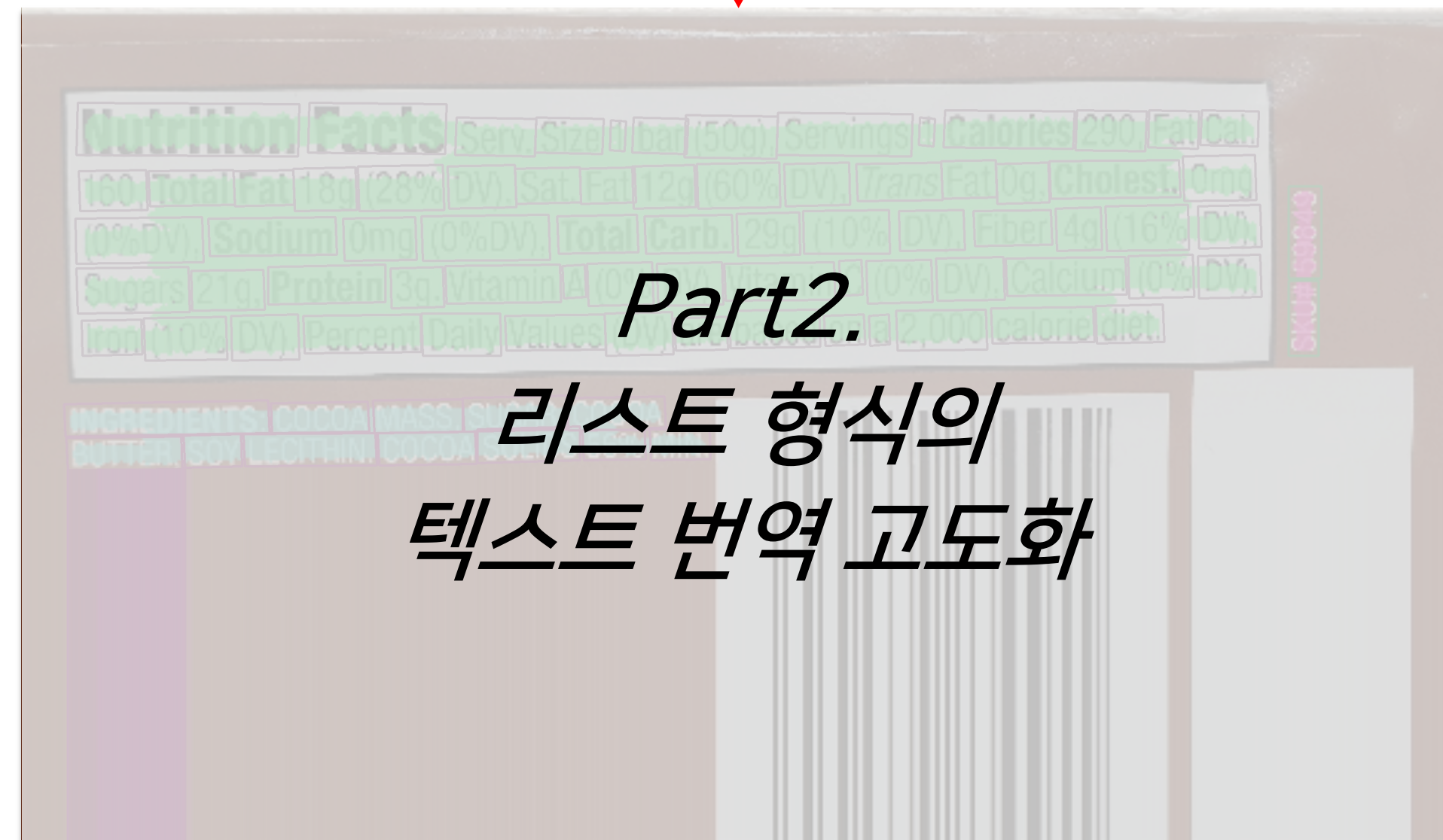
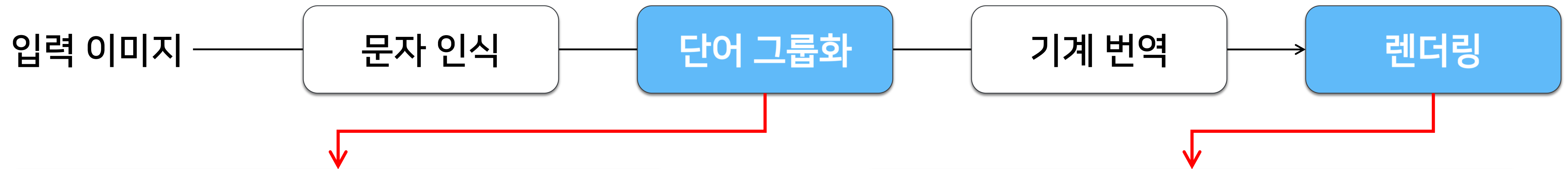
번역을 더 잘하기 위해서는
어떻게 단어를 묶어줘야 할까?



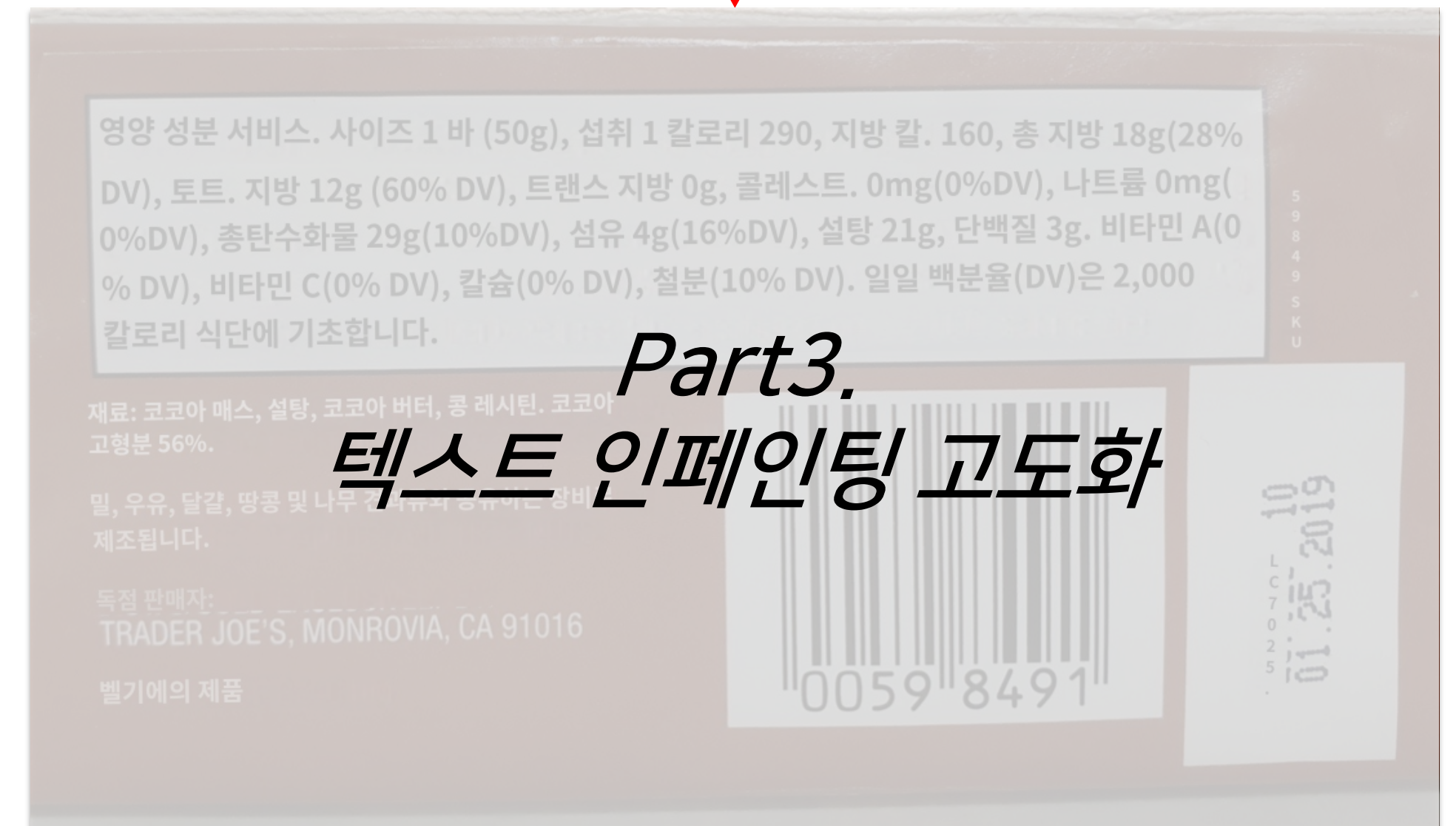
조금 더 예쁘고, 알아보기 쉽게
텍스트를 입힐 수 없을까?

1.2 바로 번역 과정

바로 번역의 핵심은 단어 그룹화와 렌더링!



번역을 더 잘하기 위해서는
어떻게 단어를 묶어줘야 할까?



조금 더 예쁘고, 알아보기 쉽게
텍스트를 입힐 수 있을까?

2.리스트 형식의 텍스트 번역 고도화

2.1 기존 방법 및 오류 예시

같은 스타일의 텍스트를 한 문단으로 묶어주는 Hierarchical Text Structuring (HTS), 개조식, 표 등 구조화된 문단에도 적합한 방식일까?

MOCKTAIL	60.000
Raspberry Soda, Apple Crisp, Virgin Mojito Kiwi, Passion Mojito, Milk Orange	
TASTY DRINK WITH ALOE VERA	60.000
Mix Fruit, Milk Tea, Mint Yogurt	
ICE BLENDER	75.000
Coffee, Matcha, Strawberry	
YOGHURT SMOOTHIES	75.000
Mango, Banana, Avocado & Coconut, Strawberry	
THICK SHAKE	75.000
Strawberry, Chocolate, Vanilla, Coffee, Snickers Special, Chocolate, Cookie Cream	
COFFEE	35.000
Vietnamese Black Coffee, Iced Coffee, Condensed Milk Coffee	
Espresso, Macchiato, Long Black, Regular Coffee	
Latte, Cappuccino, Flat White, Hot Chocolate	
Iced Cappuccino, Iced Chocolate, Iced Mochacino	
LIPTON TEA	25.000
Ice Tea By Glass	
Ice Tea Bottomless	
Ice Lemon Tea	
Pot Of Tea	
Pitcher Of Ice Tea	

메뉴판

Crown Auto Body
1717 WAVERLEY ST
WINNIPEG, MB, R3T 6A9
2044871860

TERM # 78038029
RECORD # 002845
HOST INVOICE # 0002391
HOST SEQ # 1001968
MERCH INVOICE 000010697

CARD ****4090
CREDIT/AMOUNT D
2018/11/22 16:51:10

PURCHASE

TOTAL \$200.00

AUTH#: 034636 B: 0217
HTS#: 20181122165115

TRANSACTION
APPROVED 000
THANK YOU

영수증

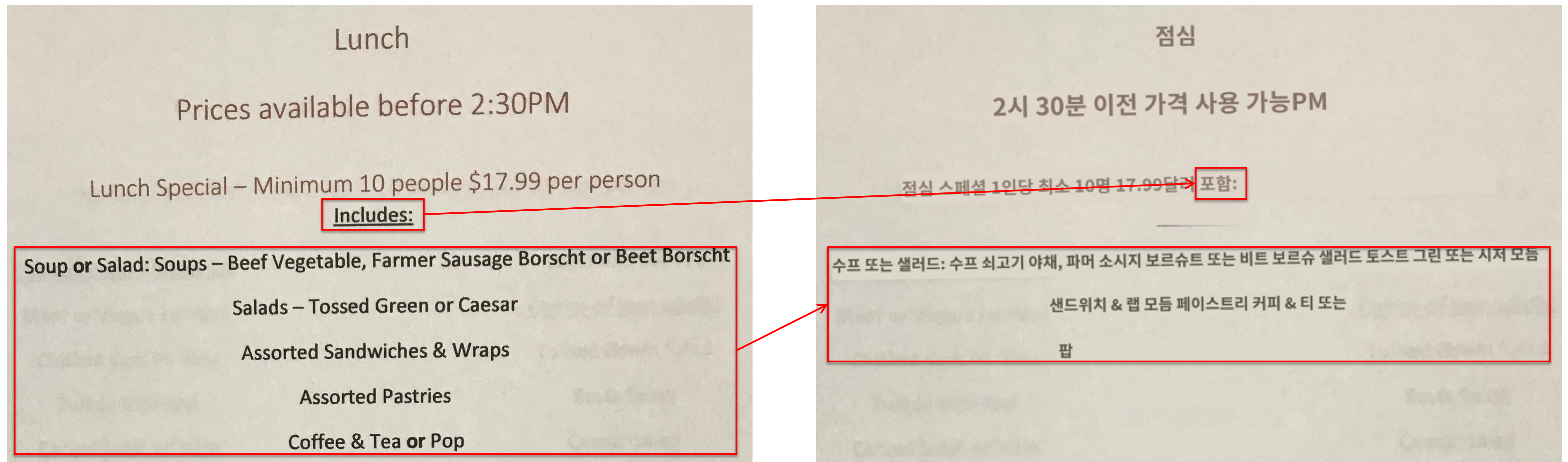


성분표

2.1 기존 방법 및 오류 예시

오류 예시1: 메뉴판

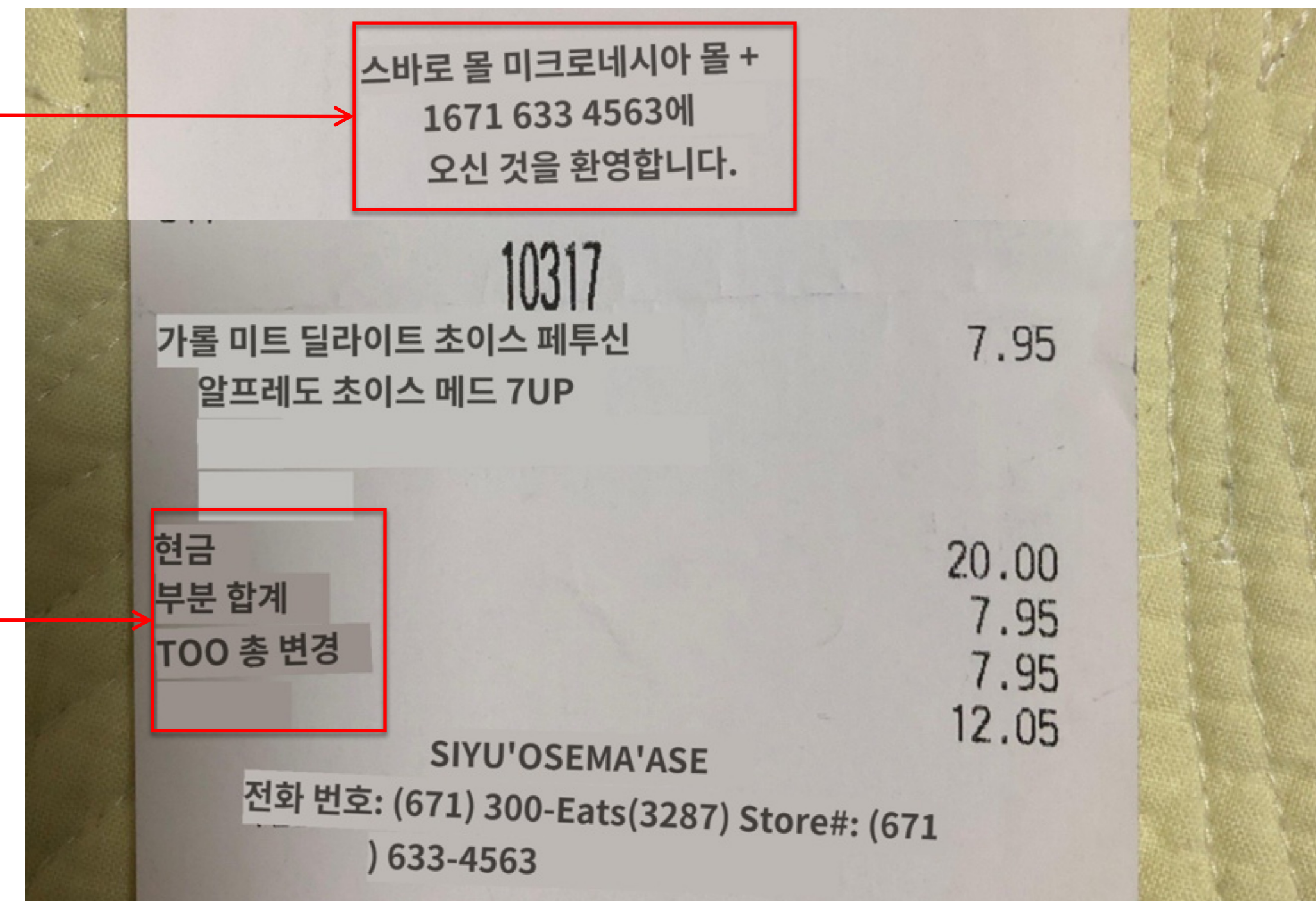
- 각기 렌더링 되어야 할 메뉴들이 함께 처리되어 식별이 어려움
- 서로 다른 색, 폰트의 텍스트가 함께 처리되어 잘못된 색과 폰트로 렌더링 됨



2.1 기존 방법 및 오류 예시

오류 예시2: 영수증

- 전화번호와 쇼핑몰 이름이 분리되지 못하여 어색한 번역 발생
- 영수증 내 각 항목이 적절한 위치에 렌더링되지 못하여 정확한 정보 파악 불가



2.2 문제 원인 파악

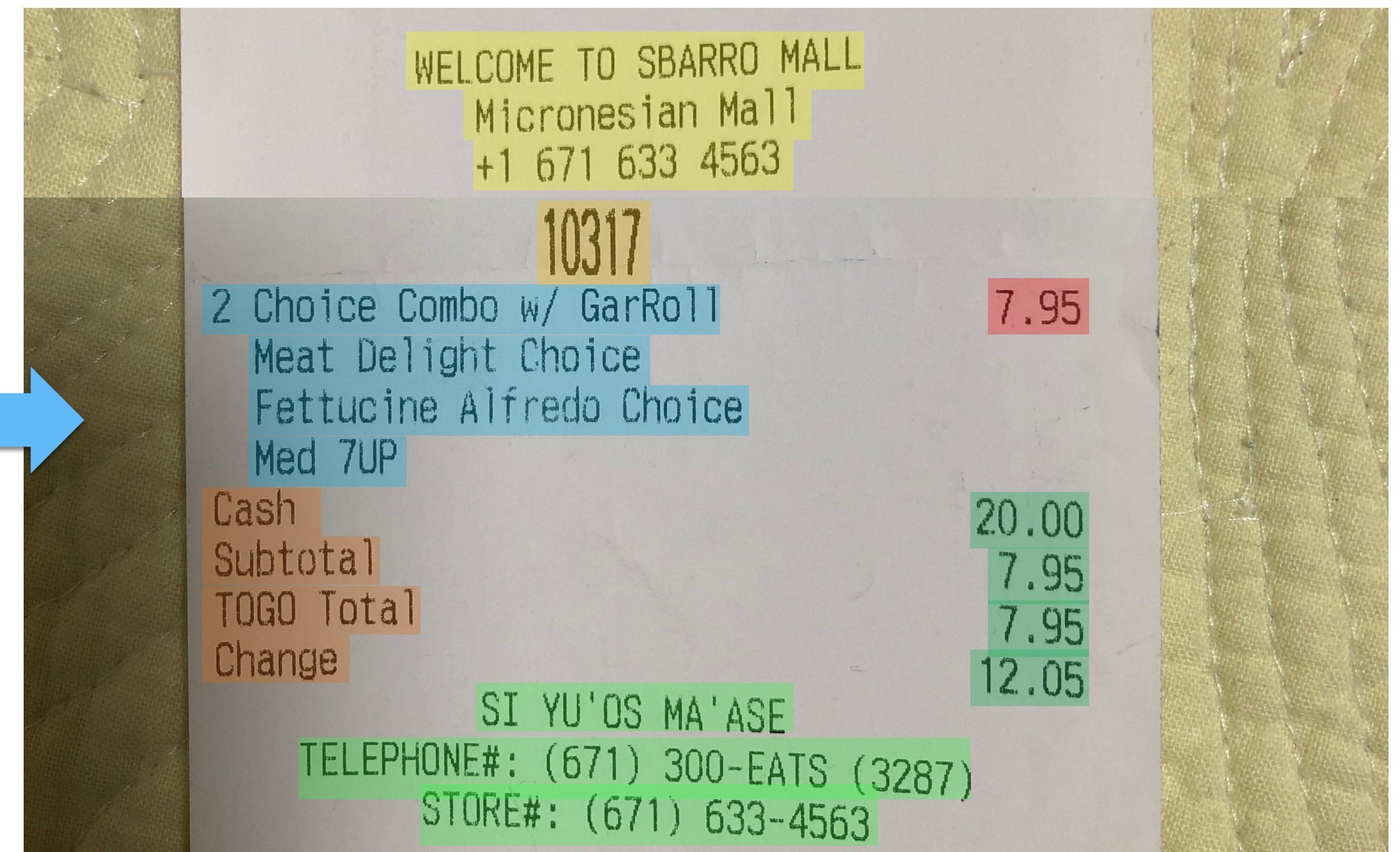
문제는 단어 그룹화 과정에서 발생!

- 입력 이미지가 들어오면 HTS 과정을 거쳐 텍스트 블록이 생성됨

입력 이미지



단어 그룹화 결과



2.2 문제 원인 파악

문제는 단어 그룹화 과정에서 발생!

- 모든 텍스트 블록을 마치 하나의 문장처럼 번역하므로 어색한 번역이 발생할 수 있음

블록 내 모든 텍스트를 병합 후 번역

Cash Subtotal TOGO Total Change

캐쉬 서브토크 토크고우 토크 토크인지


높임말 ☐

현금 부분 합계

TOO 총 변경

단어 그룹화 결과

WELCOME TO SBARRO MALL
Micronesian Mall
+1 671 633 4563

10317

2 Choice Combo w/ GarRoll 7.95
Meat Delight Choice
Fettucine Alfredo Choice
Med 7UP

Cash 20.00
Subtotal 7.95
TOGO Total 7.95
Change 12.05

SI YU'OS MA'ASE
TELEPHONE#: (671) 300-EATS (3287)
STORE#: (671) 633-4563

2.2 문제 원인 파악

문제는 단어 그룹화 과정에서 발생!

- 렌더링 시에도 하나의 문장처럼 처리되므로 각 텍스트의 고유한 위치를 잃어버림

블록 내 모든 텍스트를 병합 후 번역

Cash Subtotal TOGO Total Change

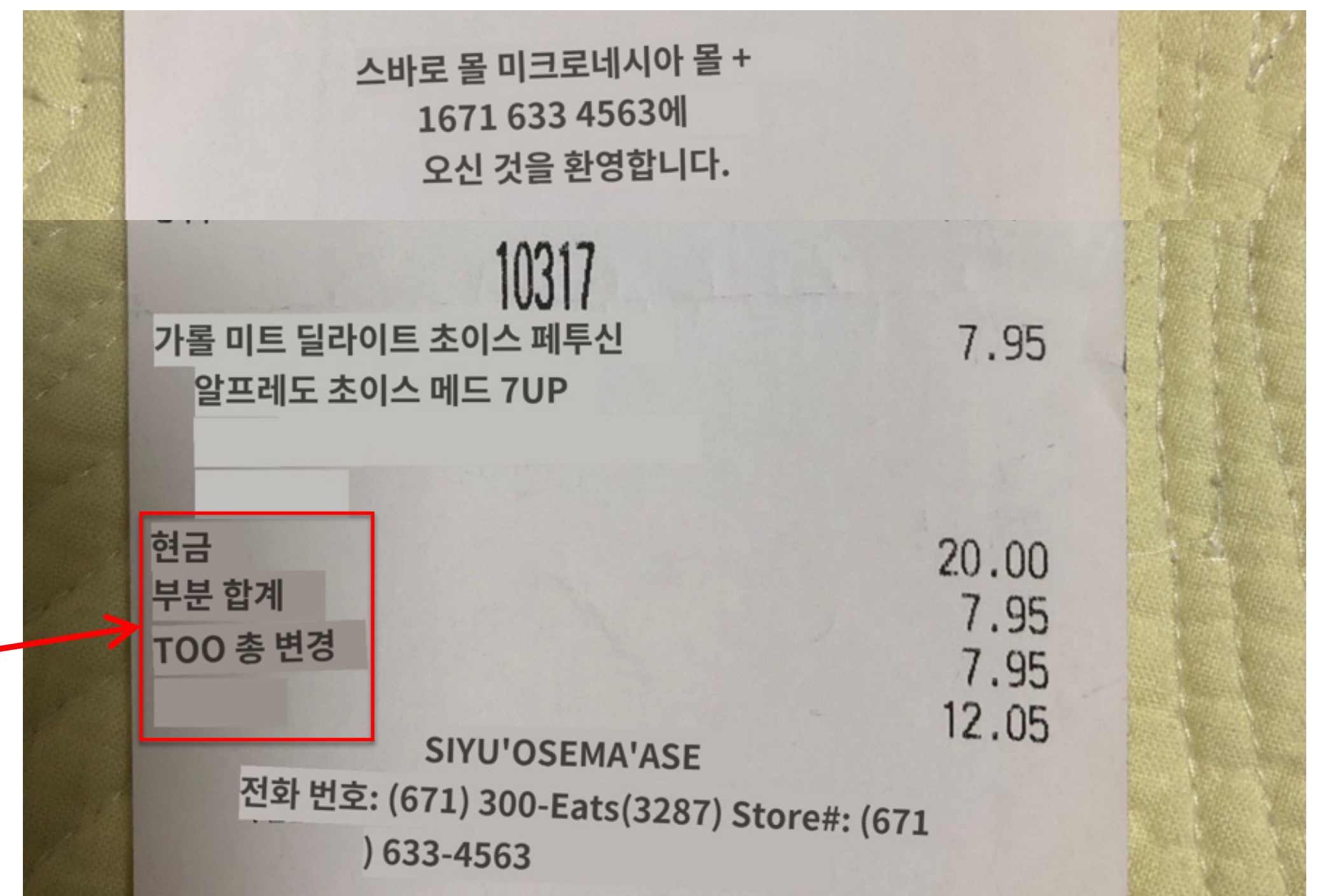
캐쉬 서브토크 토크고우 토크 토크인지


높임말 ☐

현금 부분 합계

TOO 총 변경

렌더링 결과

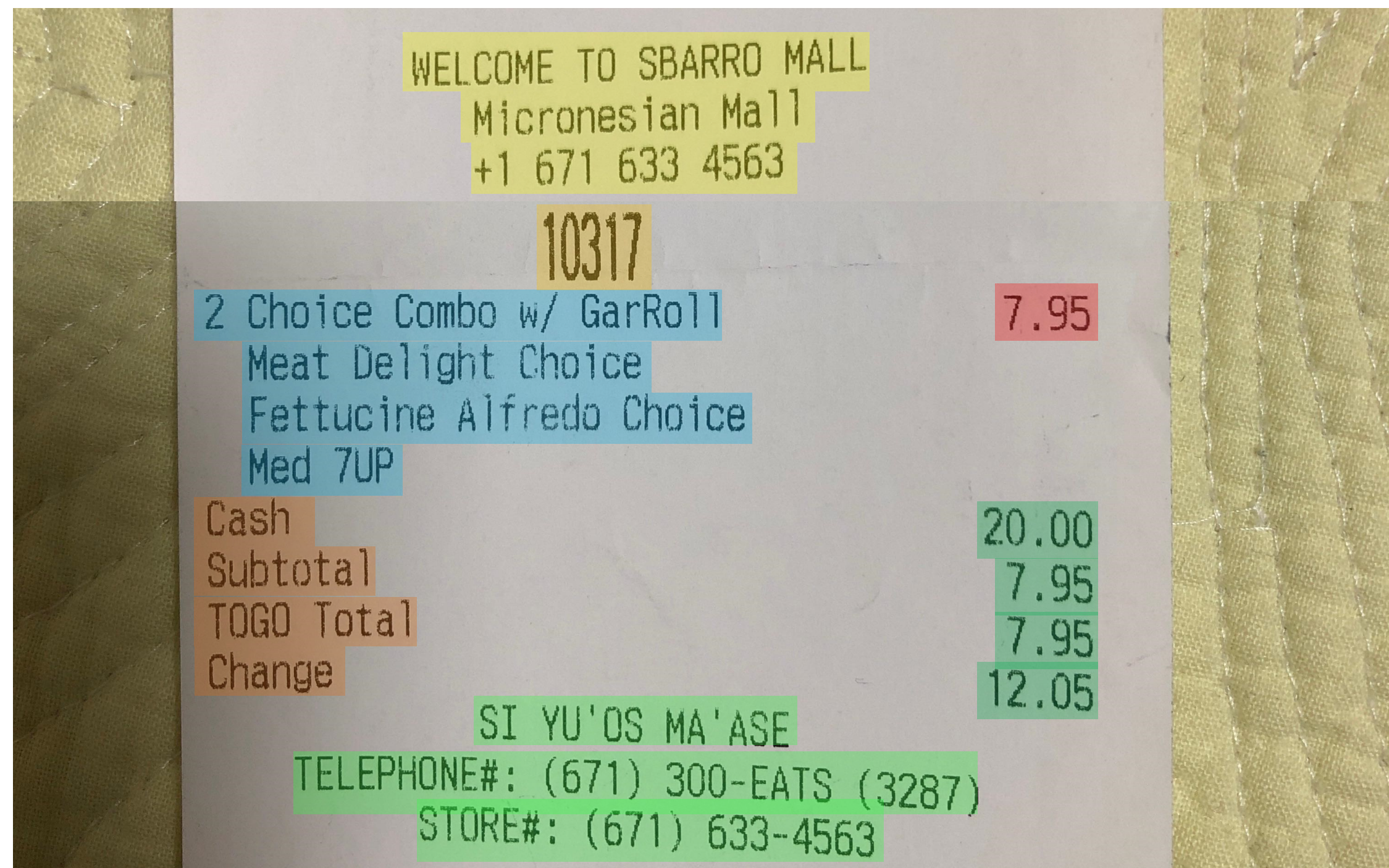


2.2 문제 원인 파악

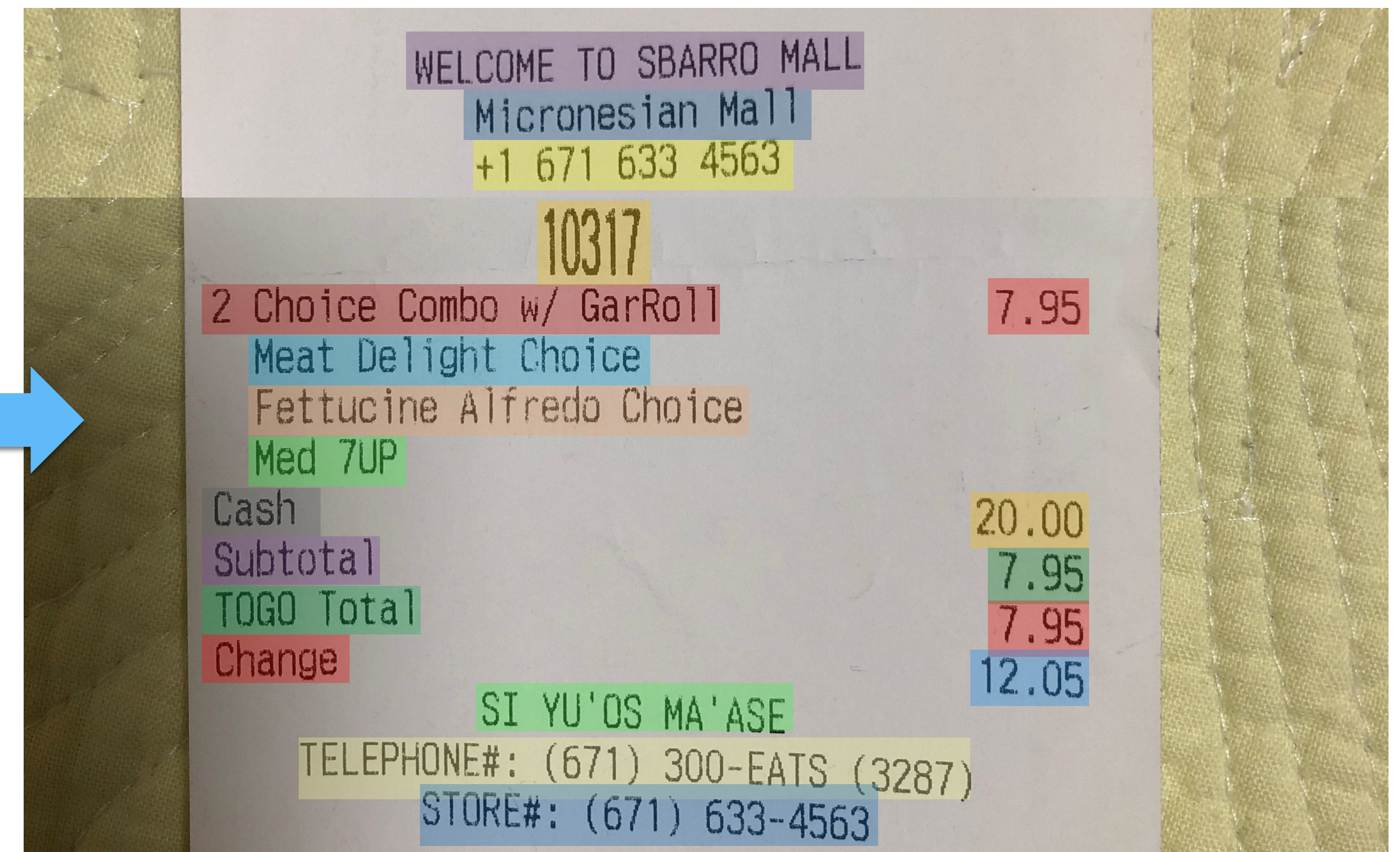
Solution: 합쳐진 블록을 분리하자!

- Visual feature를 기반으로 병합된 블록을 language feature를 기반으로 분리

As-is



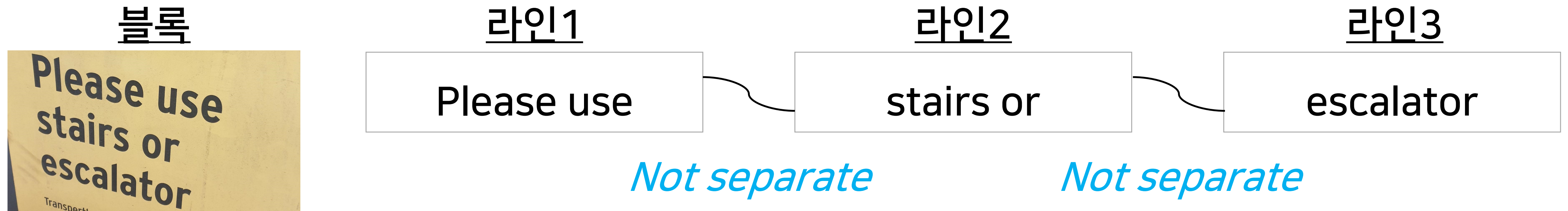
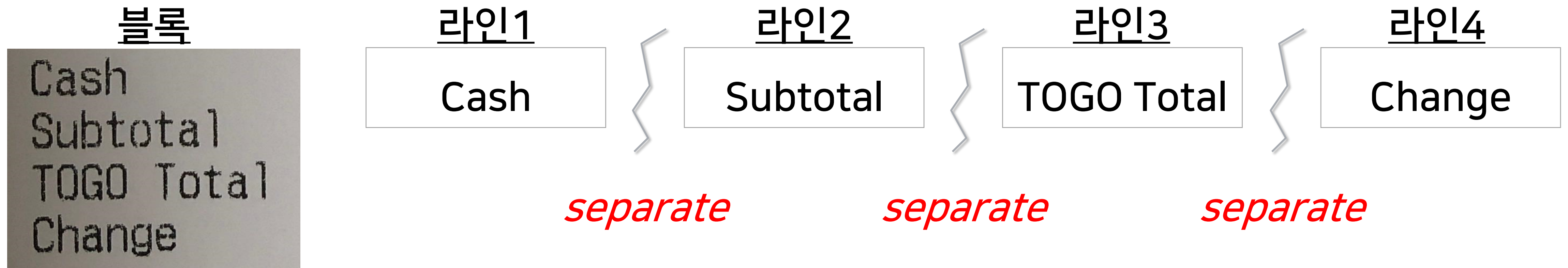
To-be



2.3 BERT를 활용한 텍스트 블록 분리

Task Formulation

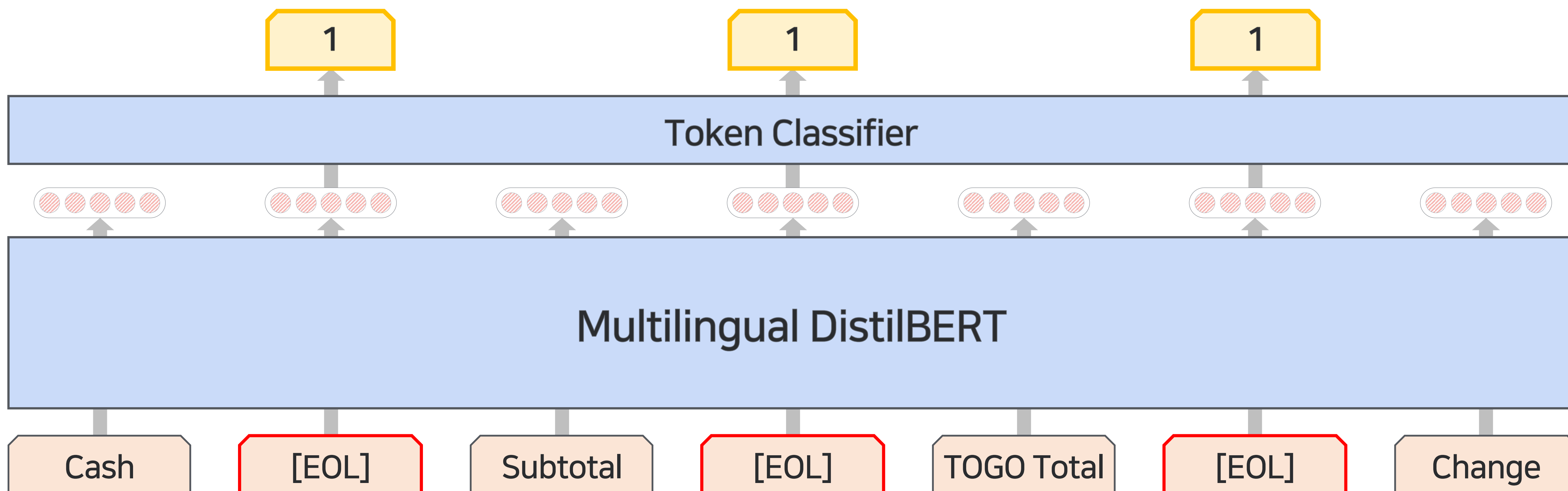
- 한 블록을 입력으로 받아 각 라인의 분리 여부를 판단하는 classification 문제로 정의



2.3 BERT를 활용한 텍스트 블록 분리

Methodology

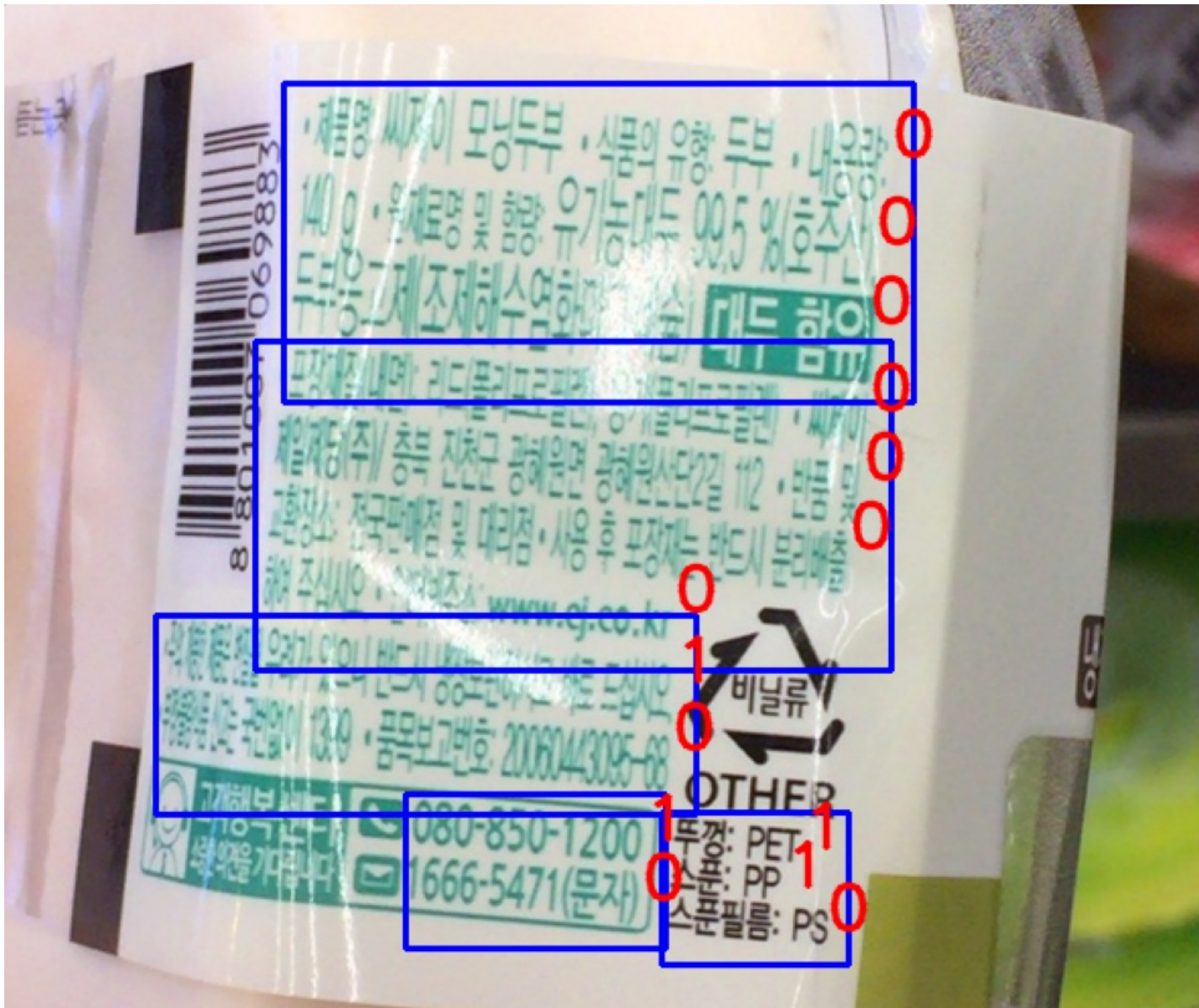
- Multilingual DistilBERT를 활용한 token classification 수행
- 라인 사이에 [End of Line] 토큰을 만들어 해당 토큰 위치에서 분리가 필요한지 판별



2.3 BERT를 활용한 텍스트 블록 분리

Data Annotation & Evaluation

- 각 텍스트 블록 내 라인별 binary 데이터 구축 (한, 중, 영, 일)
- 각 [End of Line] 토큰을 하나의 개별 데이터로 간주하여 F1-score 산출

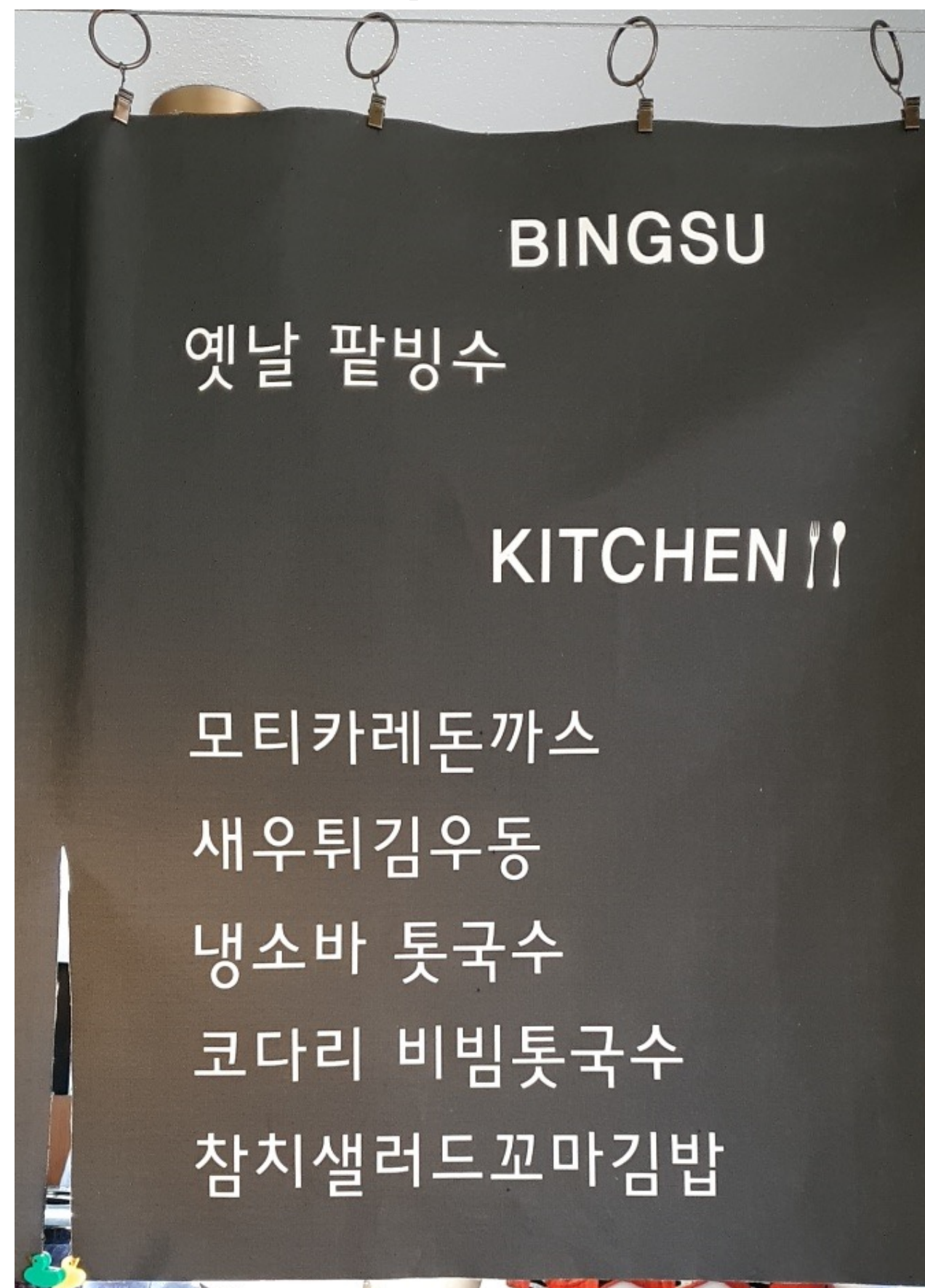


	Heuristic (PREV)	BERT (THIS)
Precision	89.60	91.71
Recall	41.25	96.46
F1-score	56.50	94.03

2.4 서비스 적용 결과

적용 예시1: 메뉴판

입력 이미지



PREV



THIS



2.4 서비스 적용 결과

적용 예시2: 상품 후면 설명

입력 이미지

USE PANAMAX FOR

Temporary relief of pain and discomfort associated with:

- Arthritis • Muscular and neuralgic conditions
- Headache

Reduces fever

DO NOT USE PANAMAX

- If using other products containing paracetamol
- For children under 7 years, except on medical advice
- For more than 48 hours for children and adolescents, aged 7-17 except on medical advice
- For more than a few days at a time in adults except on medical advice
- If foil seal over tablet is not intact

PREV

파나맥스 사용

다음과 관련된 통증과 불편의 일시적 완화:

- 관절염 • 근육 및 신경 상태
- 두통으로

발열 감소

Panamax 사용 안 함

- 다른 파라세타몰 함유 제품 사용 시 7세 미만 아동 및 청소년의 경우
- 의학적 조언 제외 48시간 이상, 의사의 조언 제외 7~17세
- 의학적 조언을 제외하고 성인의 경우 한 번에 며칠 이상, 알약 위에 포일 씰이 손상되지 않은 경우

THIS

파나맥스 사용

다음과 관련된 통증 및 불편함의 일시적 완화:

- 관절염 • 근육 및 신경통 상태
- 헤다체

열을 내린다.

파나맥스를 사용하지 마십시오.

- 파라세타몰 함유 다른 제품을 사용하는 경우
- 7세 미만 아동의 경우, 의사의 진찰을 받지 않는 경우
- 7~17세 소아청소년의 경우 의사의 진찰을 제외하고 48시간 이상 사용
- 성인의 경우 의사의 진찰을 제외하고 한 번에 며칠 이상 사용할 수 있습니다.
- 태블릿 위에 포일 실이 온전하지 않은 경우

3. 텍스트 인페인팅 고도화

3.1 기존 방법 및 문제점

규칙 기반의 Text Inpainting

- 배경색과 글자색을 추출한 뒤 텍스트 패치를 만들어 원본 이미지에 덧붙임



Text Block

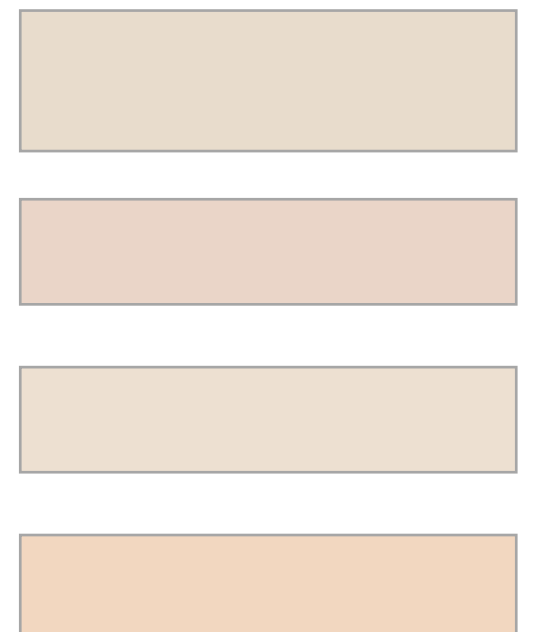
Welcome to the KIND community! Here at KIND, we think a
healthy *and* tasty, convenient *and* wholesome, economically
Kindly yours,

chewy with a crunch
Each Dark Chocolate Chunk
Healthy Grain Bar is packed with
snack you can feel good about.

배경색



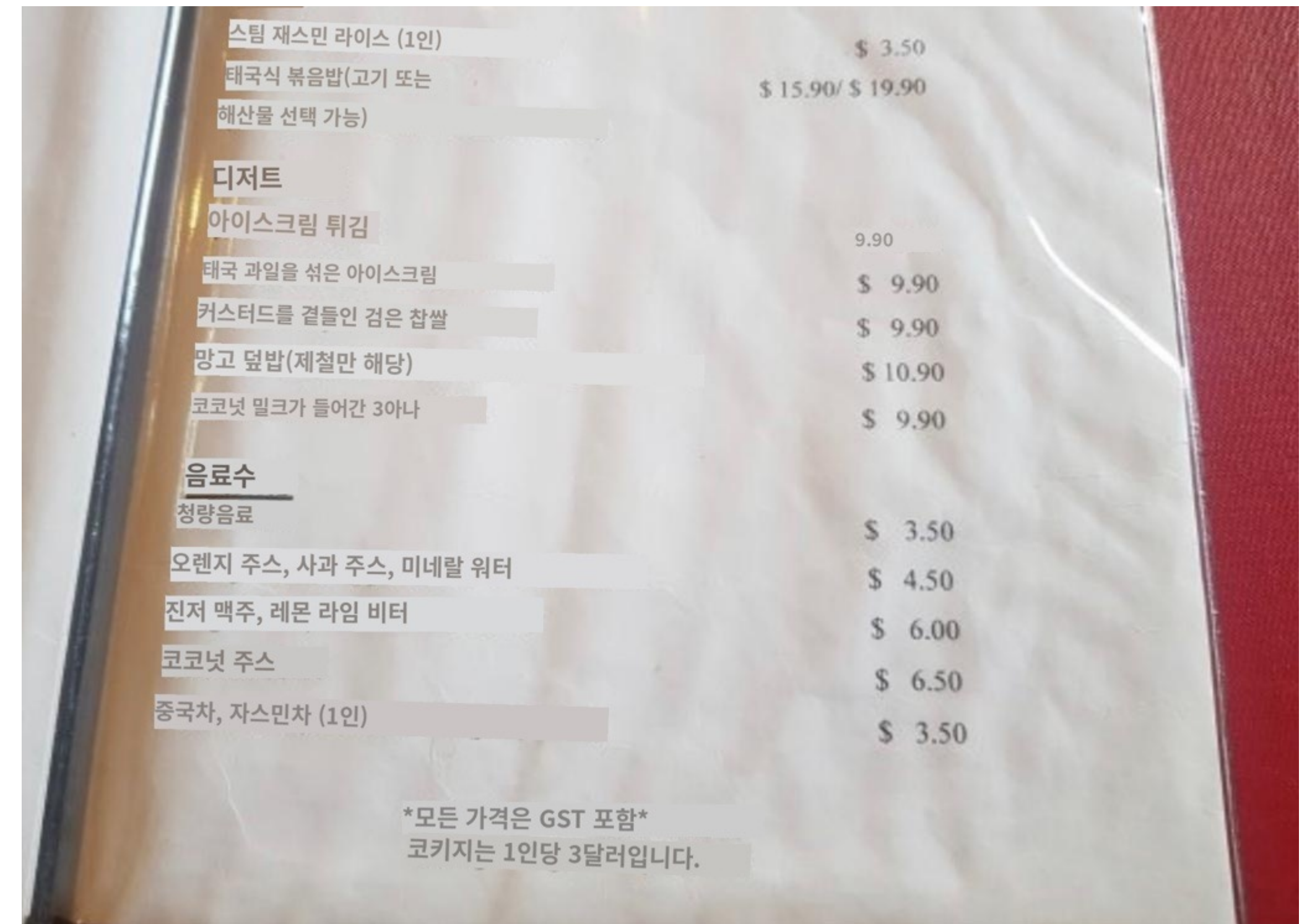
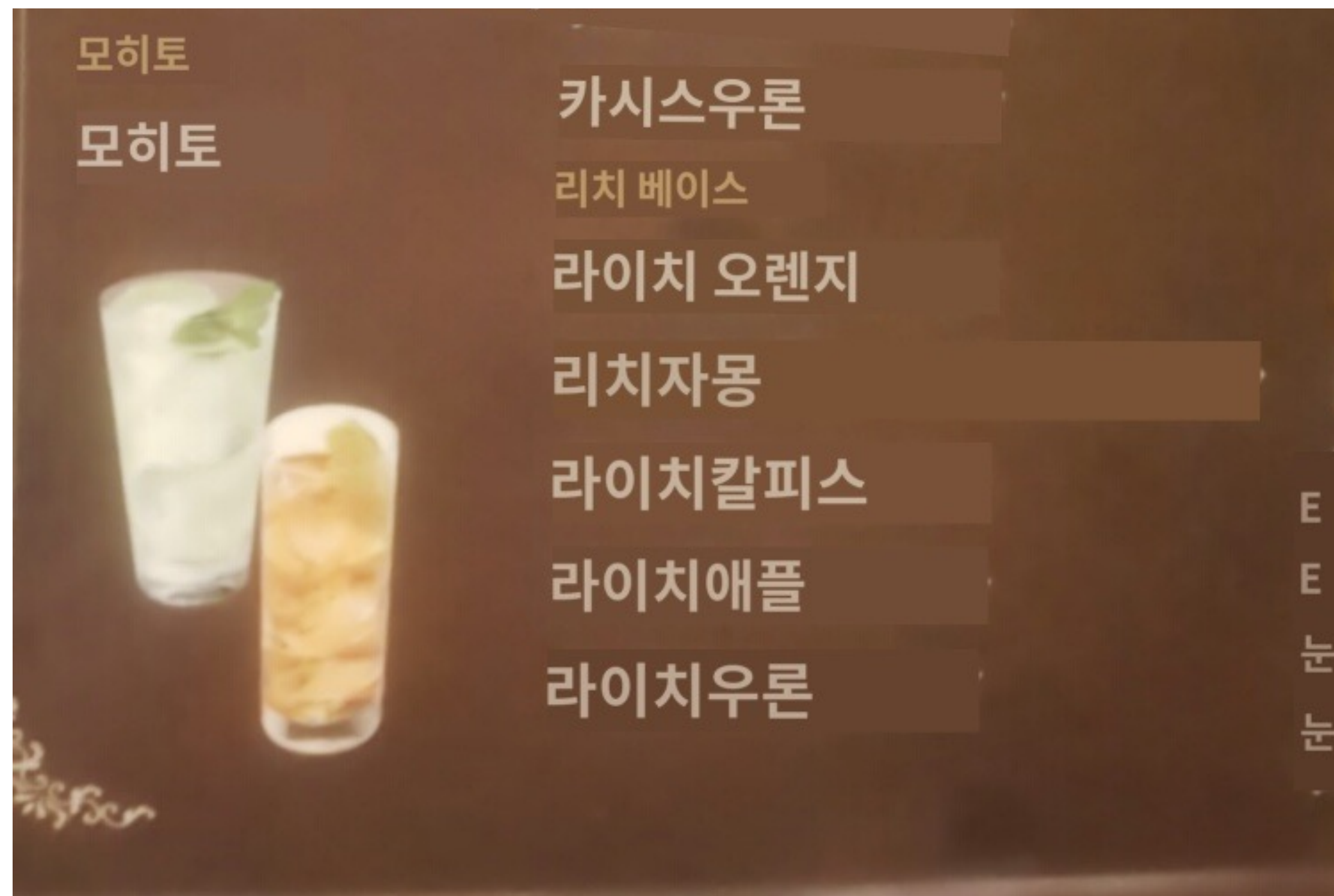
글자색



3.1 기존 방법 및 문제점

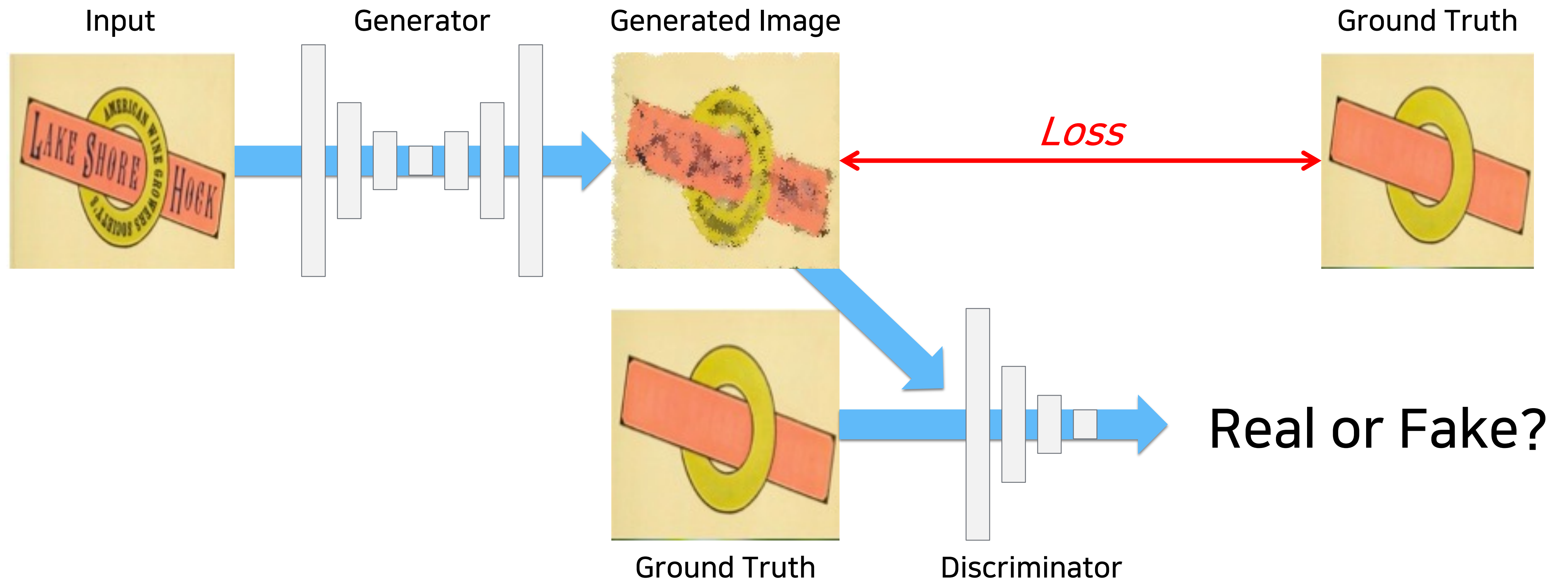
실제 배경색과 추출된 배경색의 차이로 인한 Artifact 발생

- 보다 자연스러운 텍스트 인페인팅 기술 필요



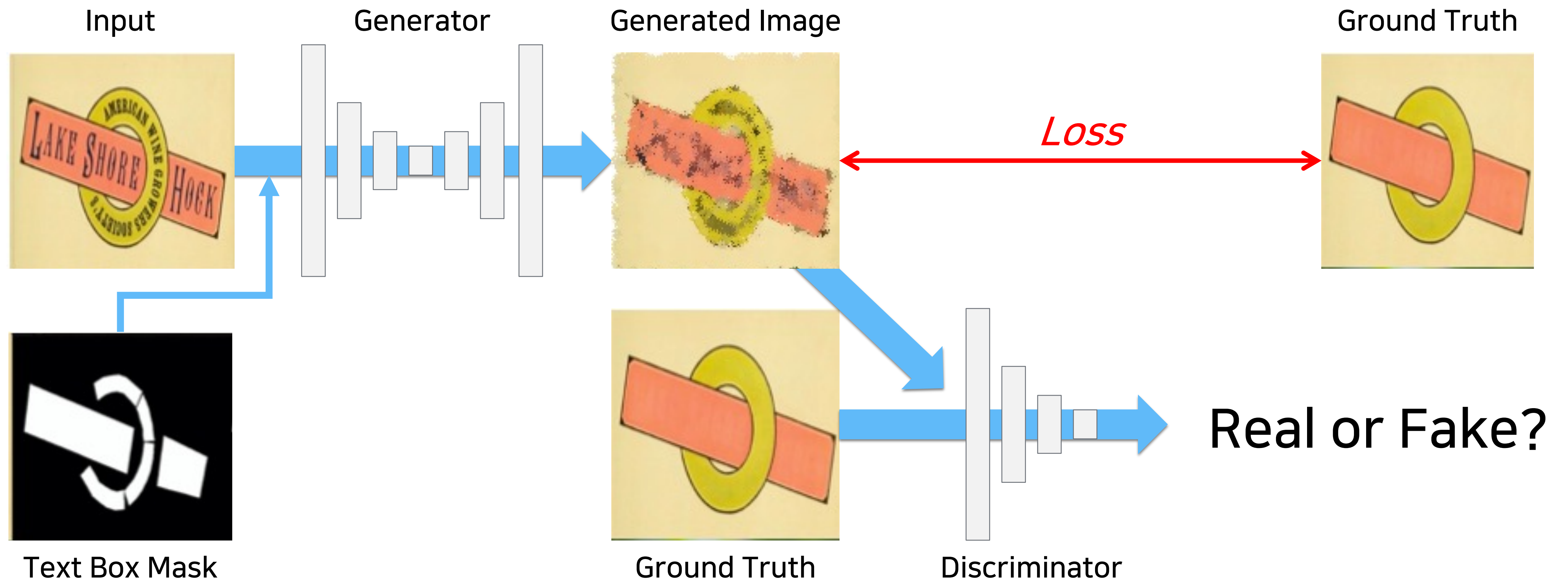
3.2 GAN을 활용한 텍스트 영역 제거

Baseline Method (ex. EnsNet)



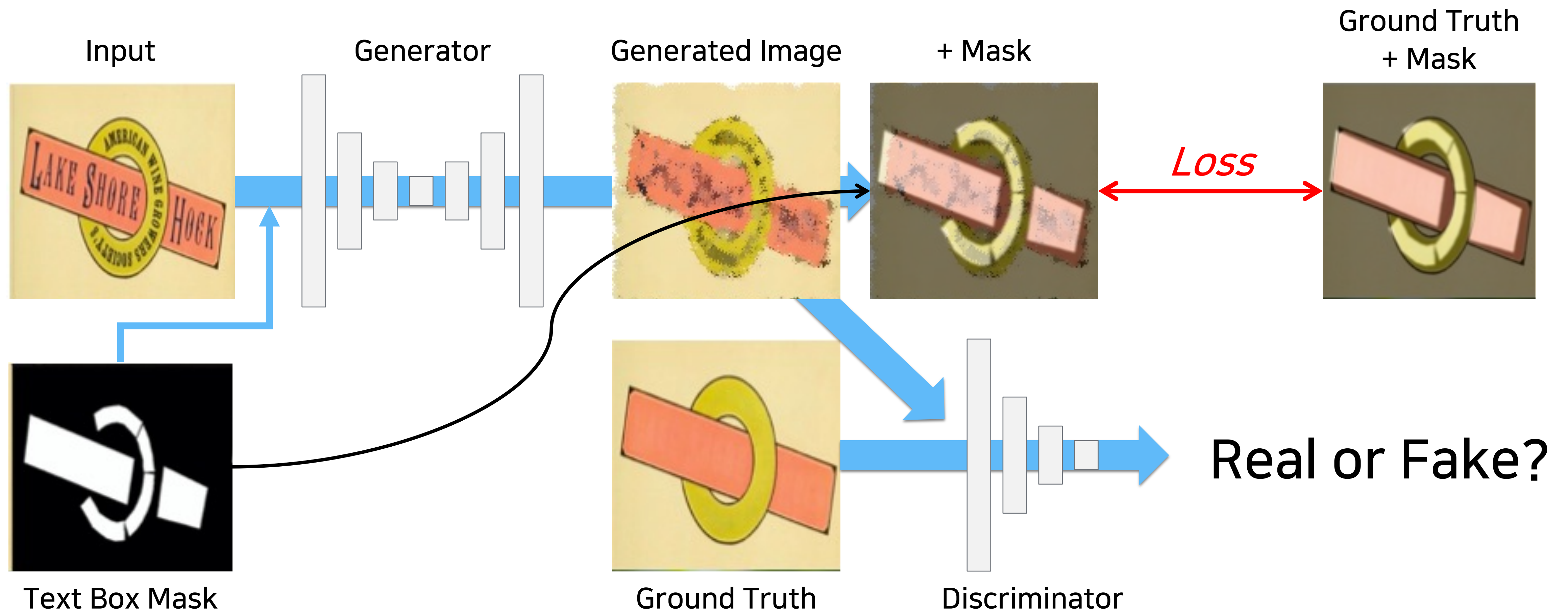
3.2 GAN을 활용한 텍스트 영역 제거

Proposed Method1: Text Box Mask 추가



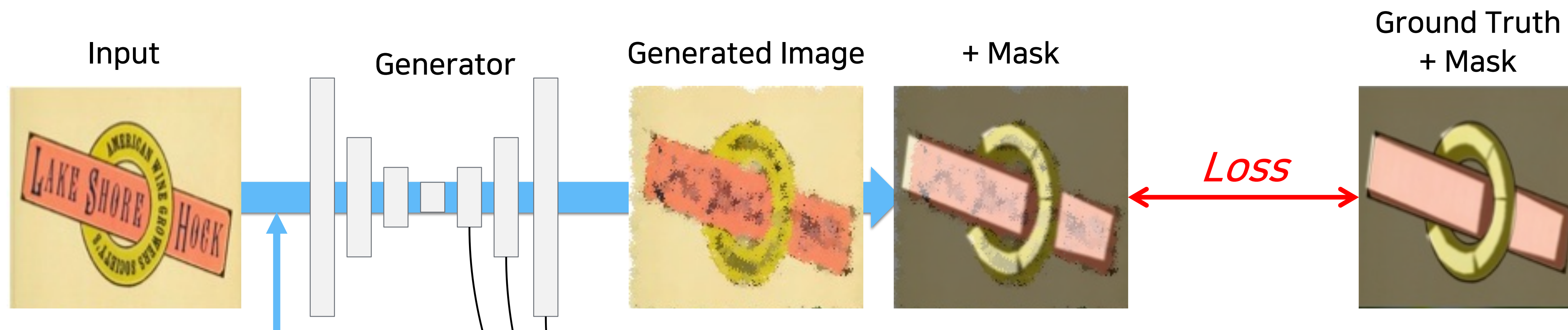
3.2 GAN을 활용한 텍스트 영역 제거

Proposed Method2: Region Regression Loss



3.2 GAN을 활용한 텍스트 영역 제거

Proposed Method2: Region Regression Loss



$$L_M(M, I_{out}, I_{gt}) = \sum_{i=1}^n \lambda_i \odot M_i \| I_{out(i)} - I_{gt(i)} \|_1$$

마지막 3개 deconvolution layer에 대해 RRL 적용

3.2 GAN을 활용한 텍스트 영역 제거

Dataset

- Real Dataset: 사람이 직접 포토샵을 이용하여 생성
- Synthetic Dataset: 글자가 없던 이미지에 글자를 합성하여 생성



Input

Ground Truth

Mask

Real Dataset



Input

Ground Truth

Mask

Synthetic Dataset

3.2 GAN을 활용한 텍스트 영역 제거

Quantitative Evaluation

- PSNR: 픽셀의 최대값 대비 잡음(MSE) 비율의 역수
- SSIM: 두 이미지간 구조적 유사도 측정
- MAE: 두 이미지간 평균 절대 오차
- Detection Eval: Text Detector를 적용하여 텍스트 제거 정도 측정

	PSNR	SSIM	MAE	Detection Eval
EnsNet (baseline)	32.99	95.16	0.0076	62.59
Ours + Mask	37.28	96.78	0.0039	43.58
Ours + Mask + RRL	40.82	98.19	0.0030	5.56

3.2 GAN을 활용한 텍스트 영역 제거

Qualitative Evaluation



Input

GT

EnsNet

MTRNet

MTRNet++

EraseNet

Ours

3.3 서비스 적용 결과

입력 이미지



Rosa Parks speaks to a reporter after her arrest. She was arrested because she refused to give her seat to a white person on a crowded bus.

That evening Martin and a number of other black leaders gathered together at Martin's church. They decided to call for a boycott on Monday, December 5. On that day they wanted all black riders to stay off the buses.

But how could they get this message to

PREV



로사 파스는 체포된 후 기자에게 말한다.
그녀는 만원 버스에서 백인 남자에게 자리를 양보하기를
거부해서 체포되었다.

그날 저녁 마틴과 많은 흑인 지도자들이 마틴의 교회에
모였다. 그들은 12월 5일 월요일에 보이콧을 요구하기로
결정했다. 그날 그들은 모든 흑인 승객들이 버스에서 떨어져
있기를 원했다.

하지만 어떻게 이 메시지를

THIS



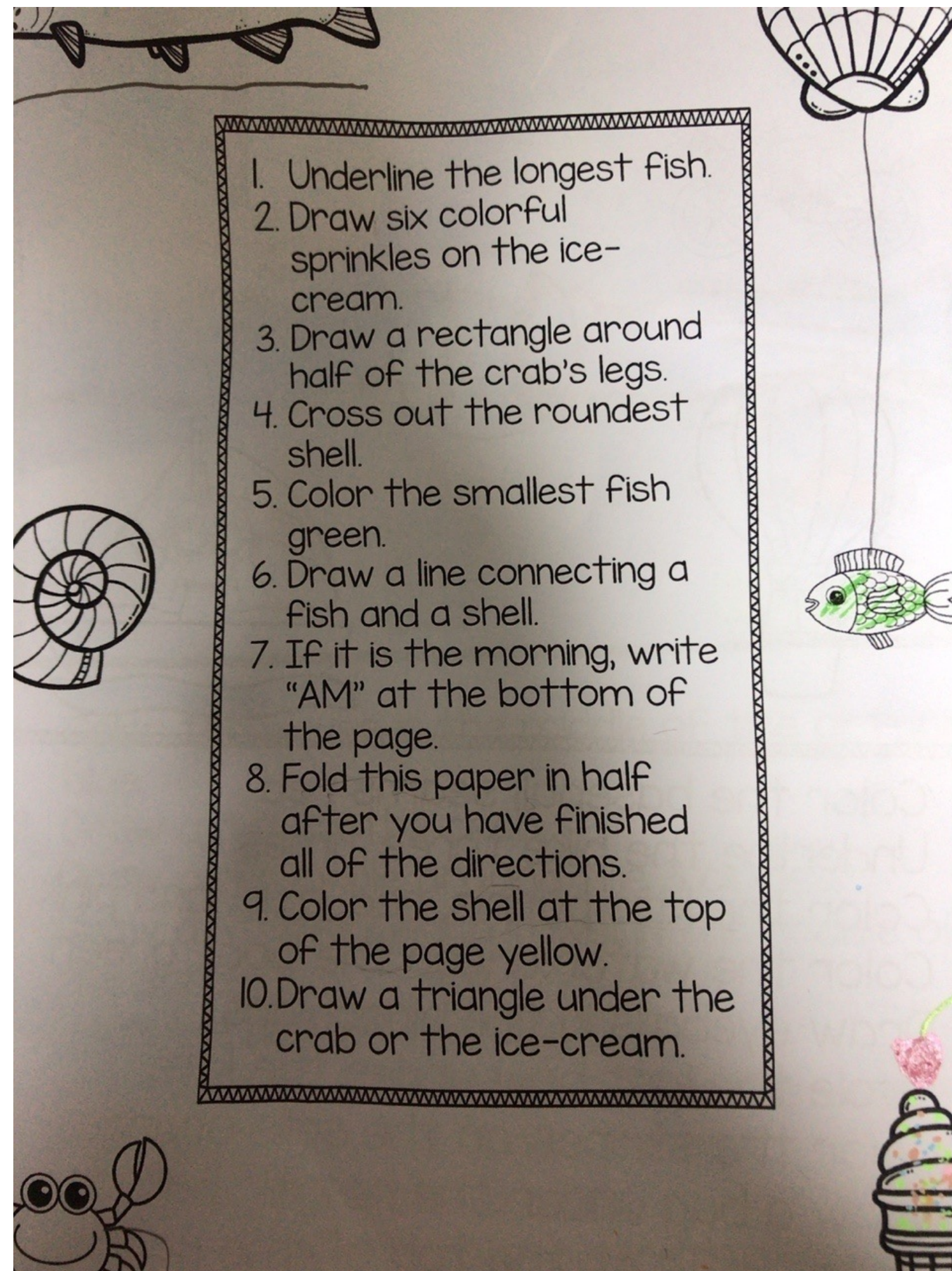
로사 파스는 체포된 후 기자에게 말한다.
그녀는 만원 버스에서 백인 남자에게 자리를 양보하기를
거부해서 체포되었다.

그날 저녁 마틴과 많은 흑인 지도자들이 마틴의 교회에
모였다. 그들은 12월 5일 월요일에 보이콧을 요구하기로
결정했다. 그날 그들은 모든 흑인 승객들이 버스에서 떨어져
있기를 원했다.

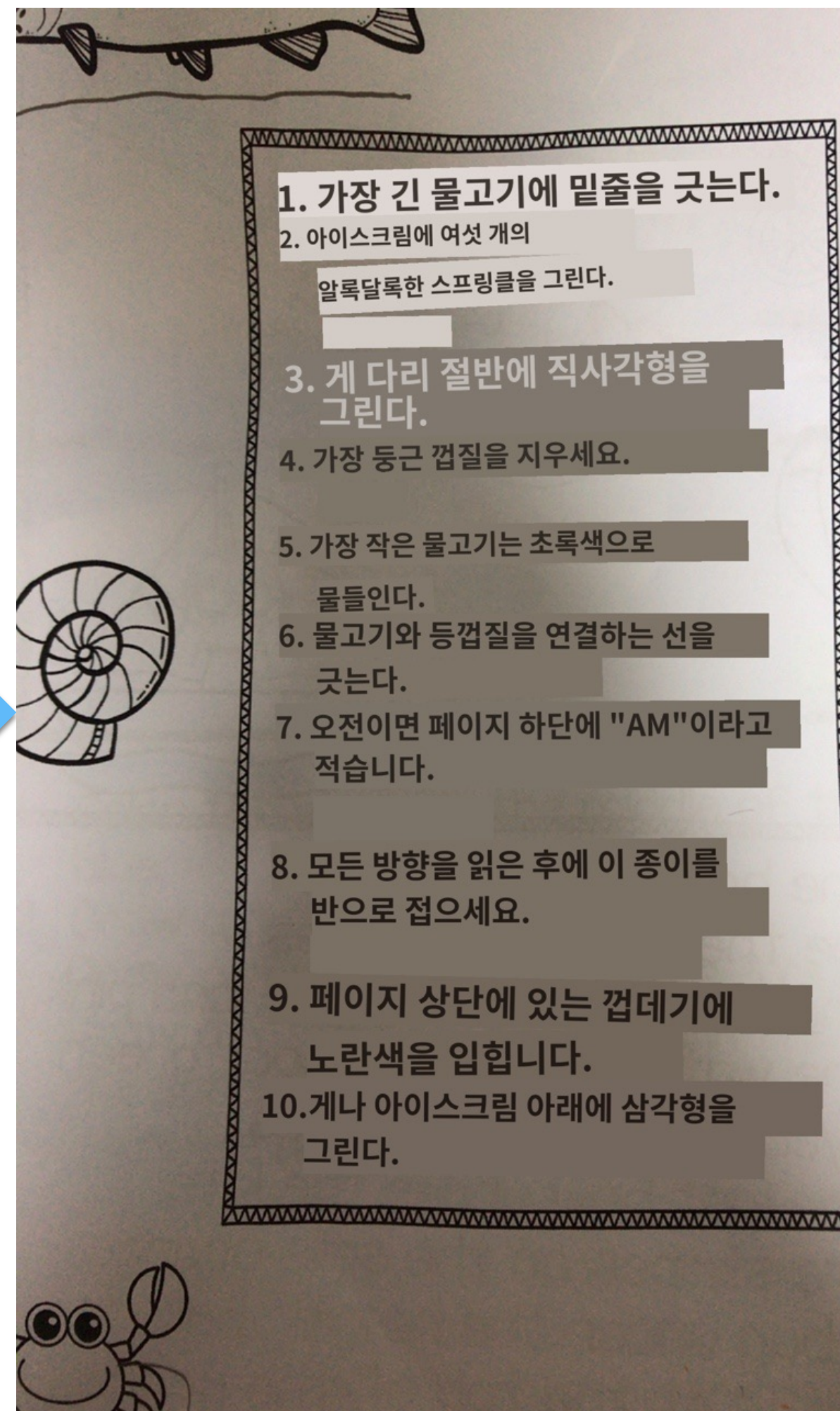
하지만 어떻게 이 메시지를

3.3 서비스 적용 결과

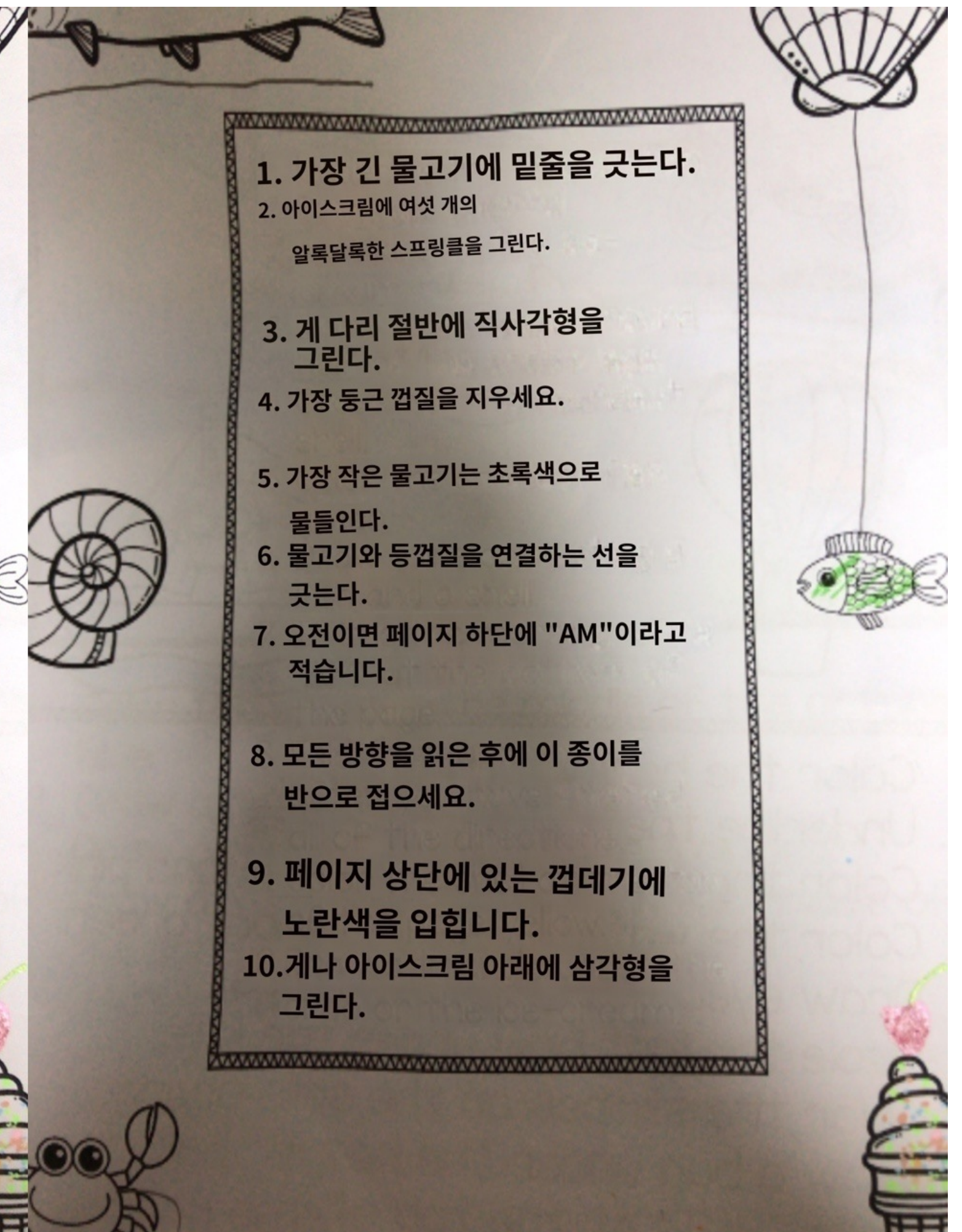
입력 이미지



PREV



THIS



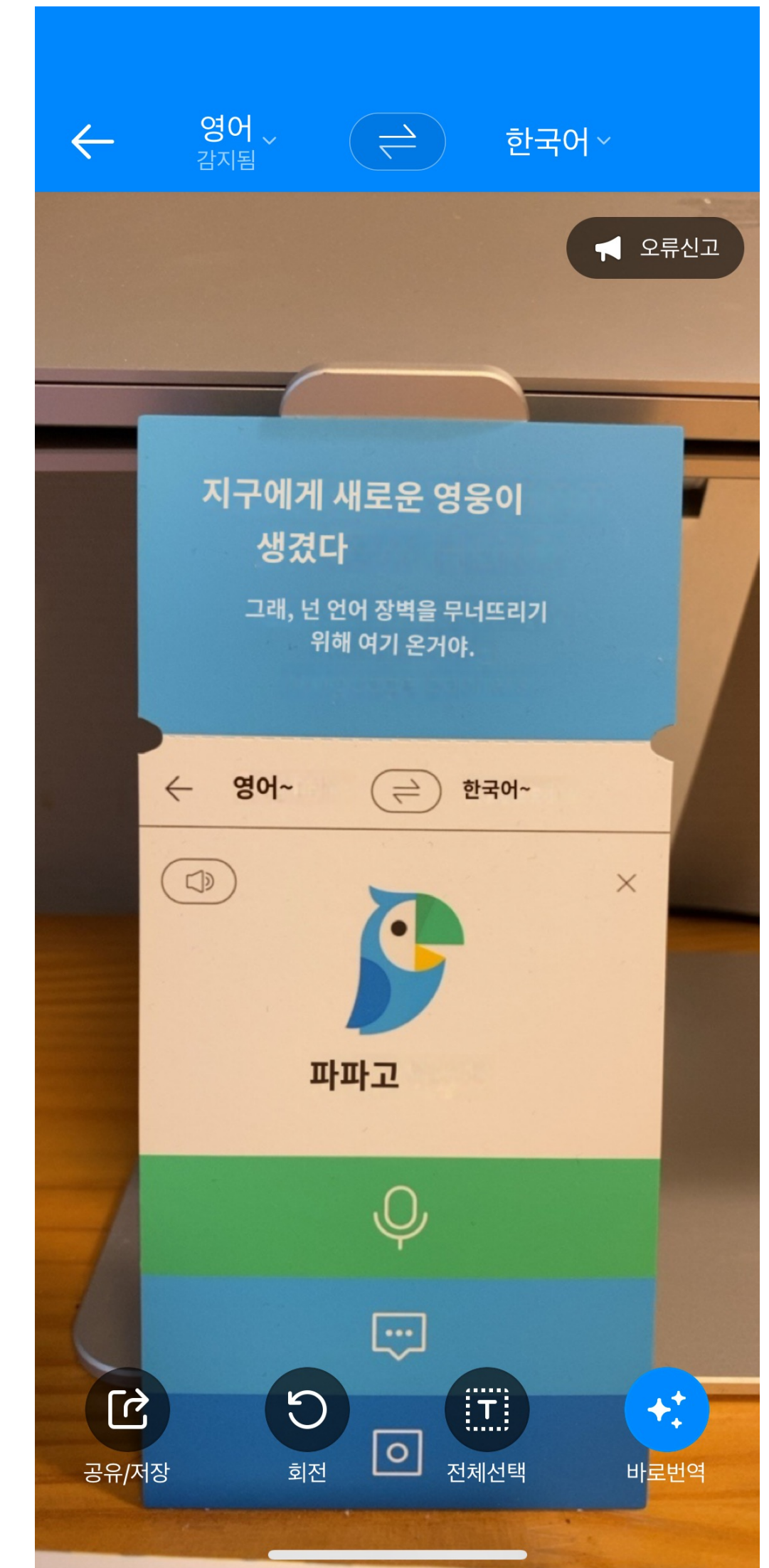
4.더 나아가기: AR 이미지 번역

4.1 AR 이미지 번역이란?

기존의 이미지 번역

- 촬영된 사진 한 장(스틸컷)에 대해 이미지 번역 결과를 보여주는 방식

매번 촬영 버튼을 누르는 불편함을 없애고
쉽게 여러 대상을 번역할 수 있을까?



4.1 AR 이미지 번역이란?

AR 이미지 번역

- 카메라에 비춰지는 영상에 대해
실시간으로 이미지 번역 결과를 보여주는 방식
- 파파고 이미지 번역 기술 + **Object tracking**(객체 추적)



4.2 AR 이미지 번역 엔진 설계

Object tracking(객체 추적)이란?

- 동영상의 연속된 프레임에서 시간에 따라 움직이는 객체의 위치를 추적하는 기술

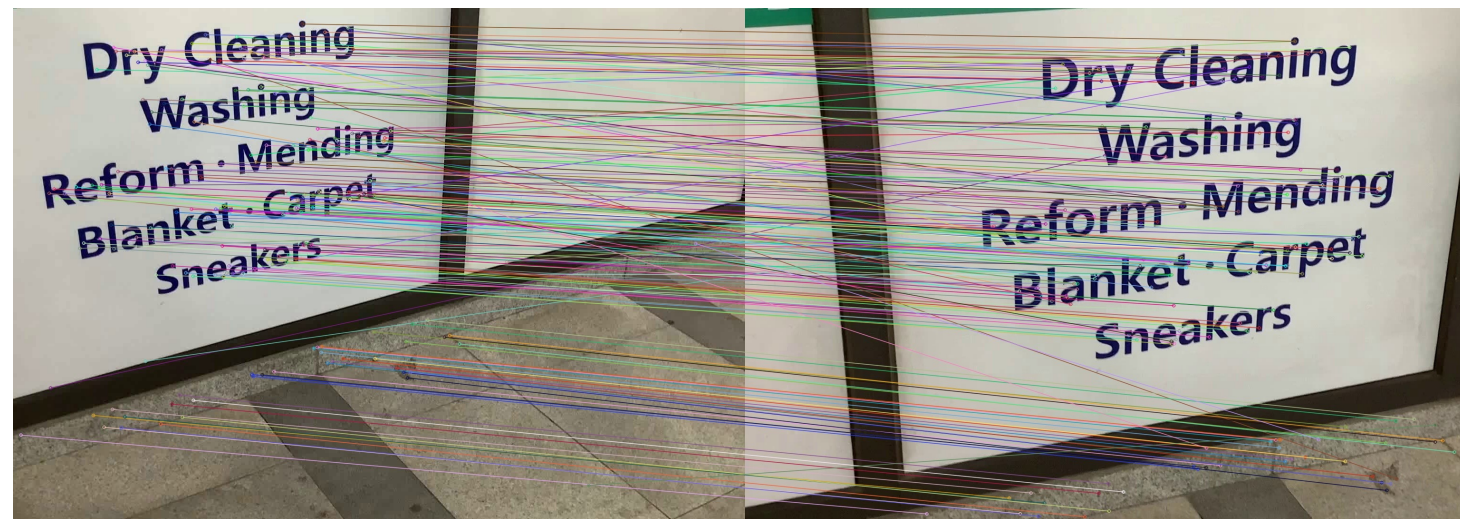


4.2 AR 이미지 번역 엔진 설계

Object tracking(객체 추적)이란?

- 동영상의 연속된 프레임에서 시간에 따라 움직이는 객체의 위치를 추적하는 기술

Image matching



Optical flow



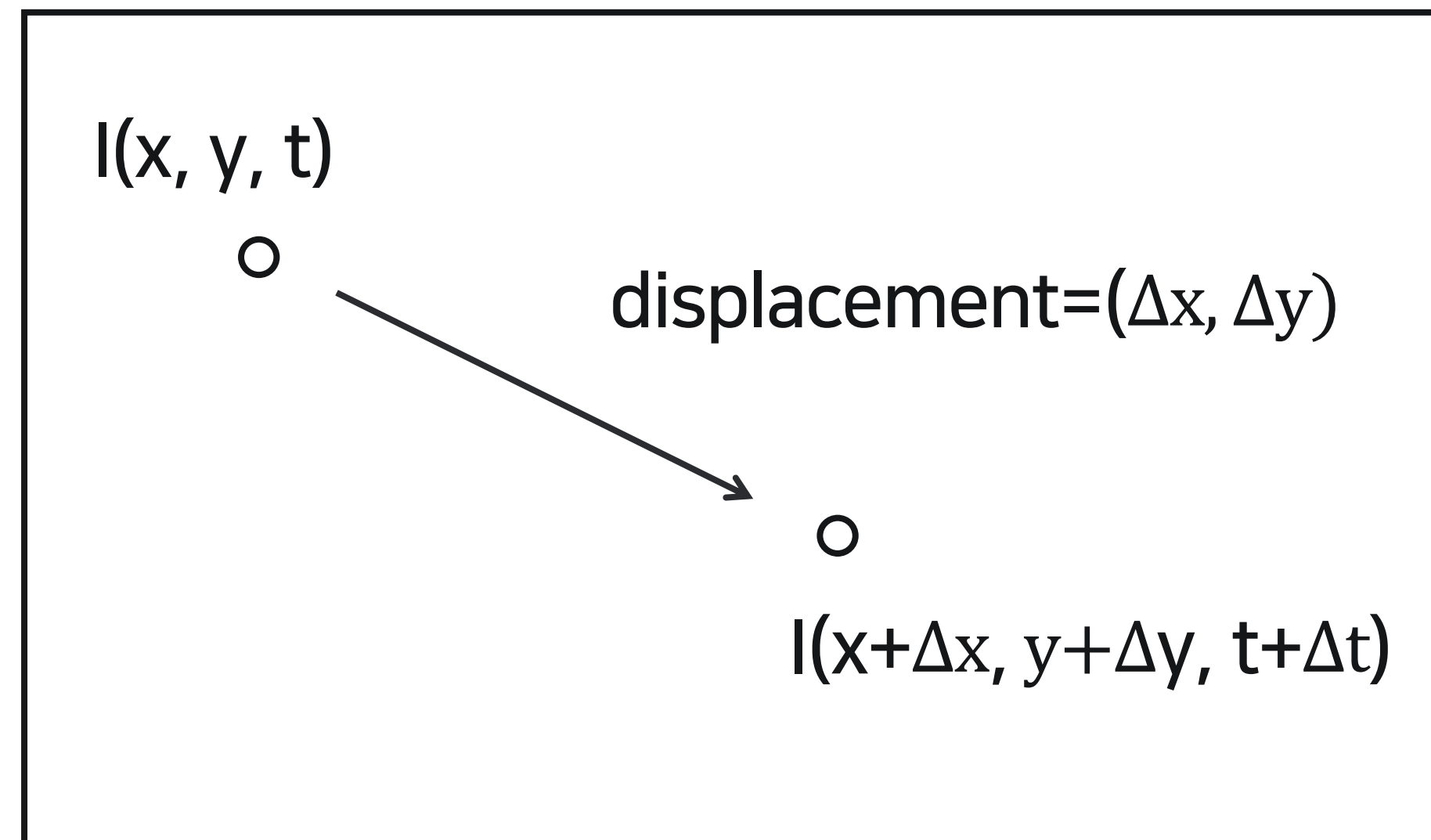
Filter 기반(ex.KCF)



4.2 AR 이미지 번역 엔진 설계

Optical flow

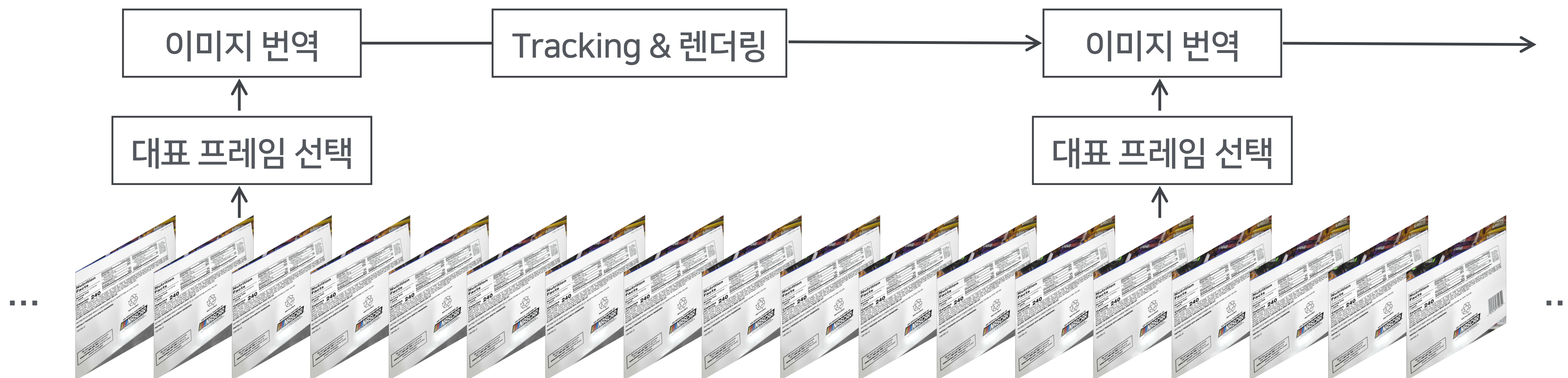
- 밝기 패턴의 가시적인 움직임(프레임 간 픽셀들의 움직임을 추정하는 데 사용)
- 모서리나 코너를 뚜렷하게 가지는 텍스트 이미지에 적용하기에 적합
- 적은 연산량



4.2 AR 이미지 번역 엔진 설계

엔진 동작 시나리오

- 1) 대표 프레임 선택
 - 2) 이미지 번역 수행
 - 3) Optical flow를 이용한 tracking & 렌더링 (반복)
- 1 Cycle

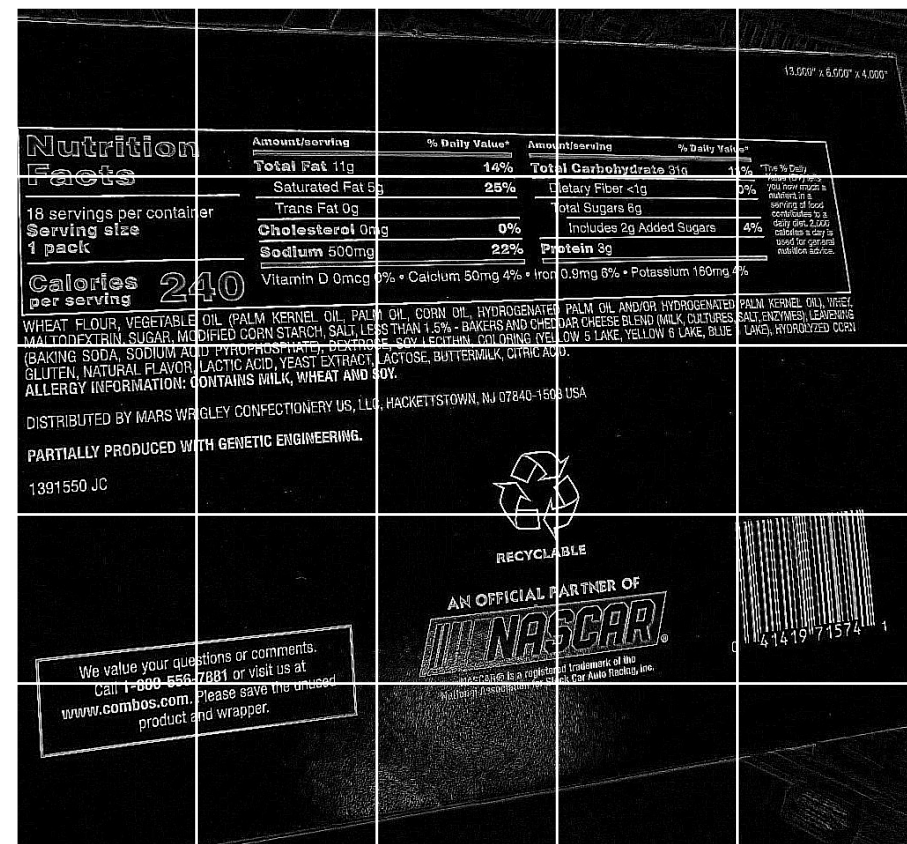


4.2 AR 이미지 번역 엔진 설계

엔진 모듈 구현

대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링

버퍼



17.3	58.2	29.5	7.3	10.8
22.4	59.6	51.6	33.4	18.6
24.0	48.1	36.2	40.3	20.0
43.2	20.4	37.6	45.3	26.7
28.1	42.5	35.9	31.5	15.7

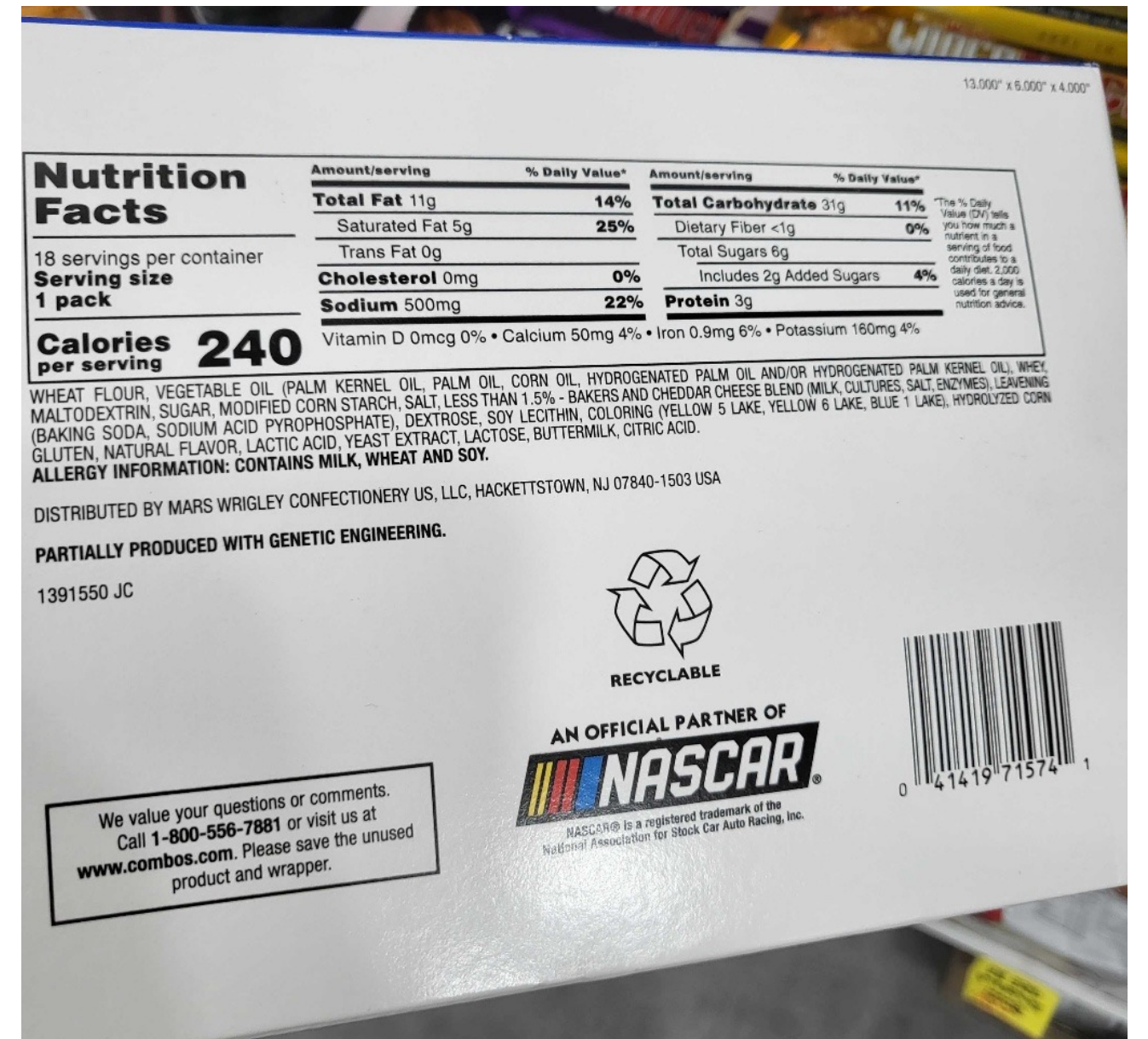
$$(H1+H2) - (L1+L2)$$

$$=(59.6 + 58.2)$$

$$- (7.3 + 10.8)$$

$$= 98.7$$

Laplacian filter → block 표준편차 계산 → Sharpness 계산 → Sharpness가 가장 높은 프레임 선택



4.2 AR 이미지 번역 엔진 설계

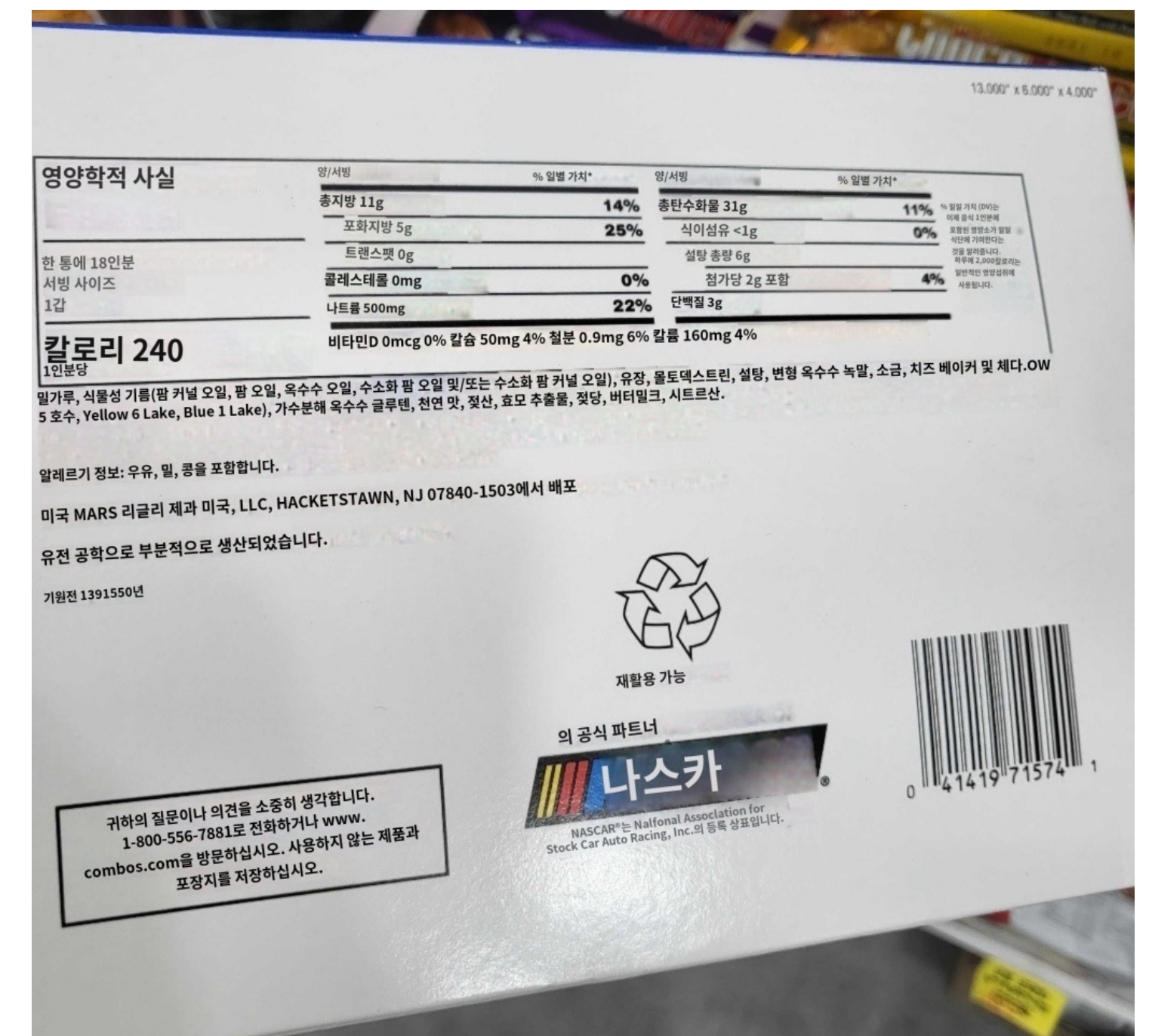
엔진 모듈 구현

대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링

대표 프레임에 대한
이미지 번역 결과 =



+



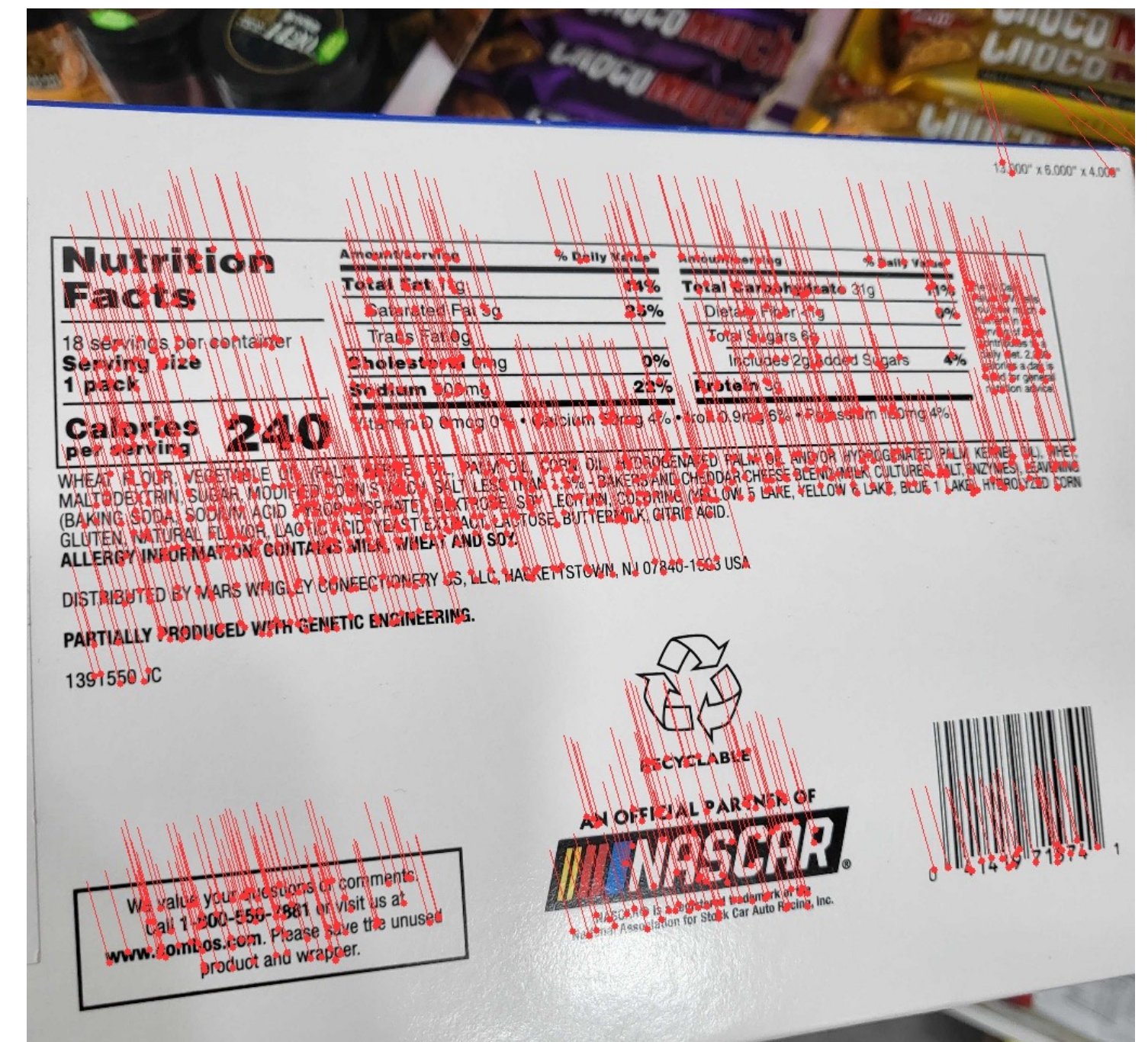
box 정보

렌더링 이미지

4.2 AR 이미지 번역 엔진 설계

엔진 모듈 구현

대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링



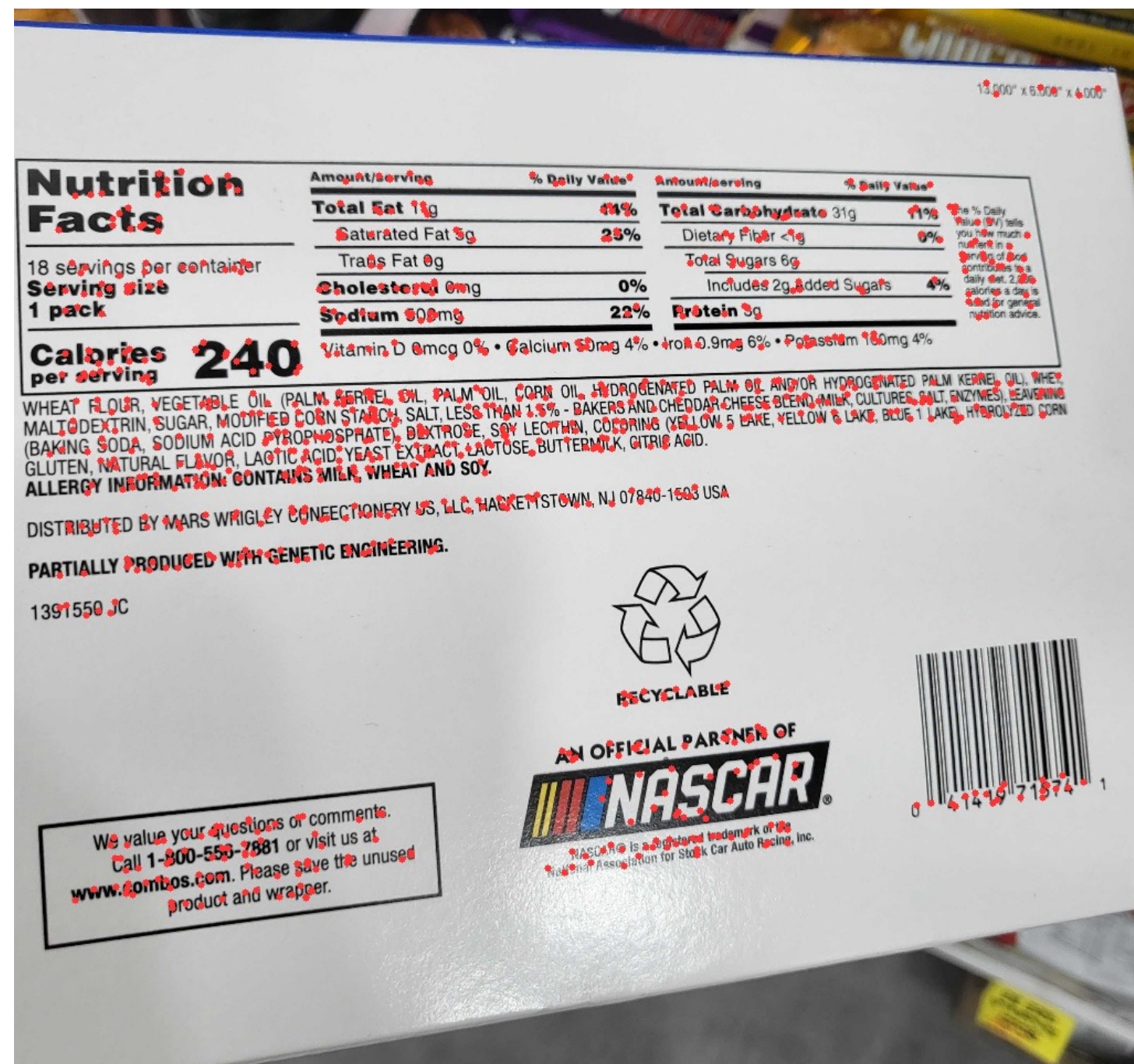
첫 프레임: box 내에서 corner points 추출

이후 프레임: optical flow 계산

4.2 AR 이미지 번역 엔진 설계

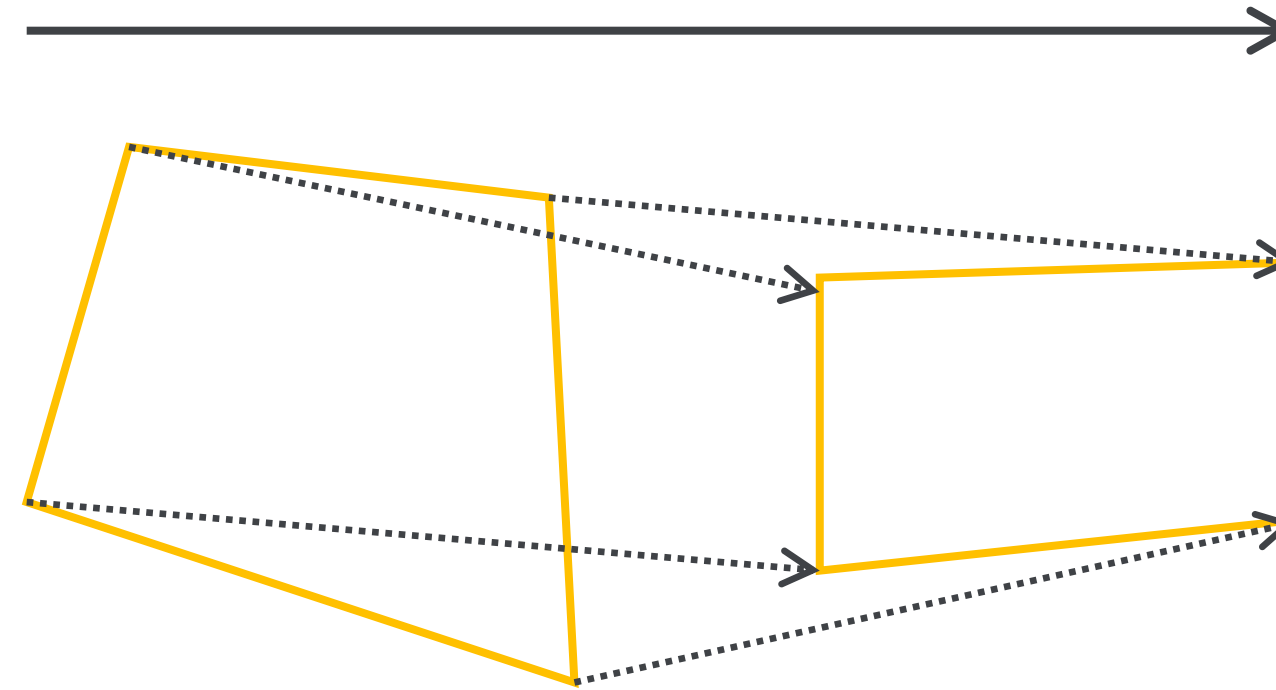
엔진 모듈 구현

대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링



첫 프레임

Homography 추정
(RANSAC)



*자유도 8의 변환 matrix

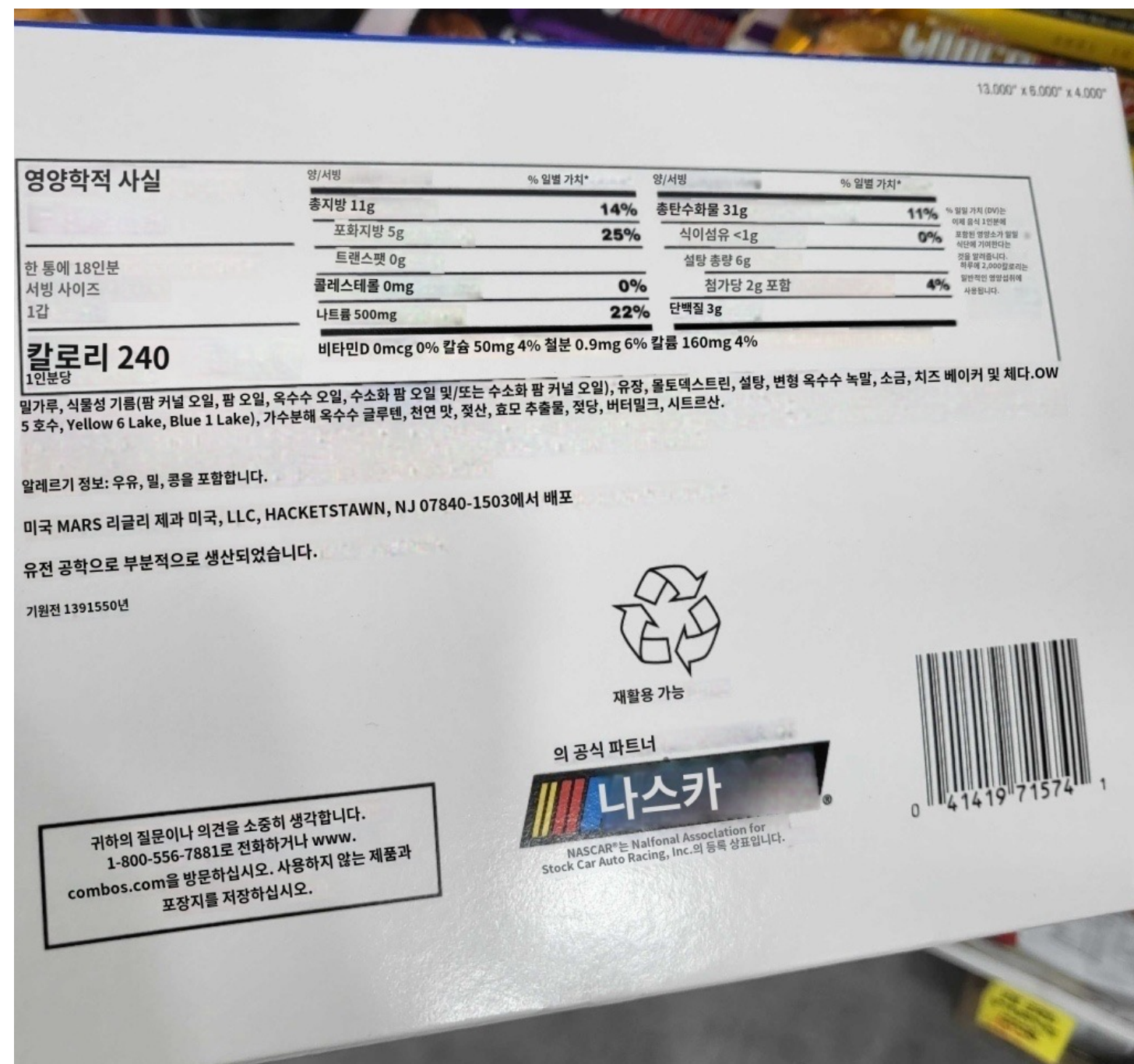


이후 프레임

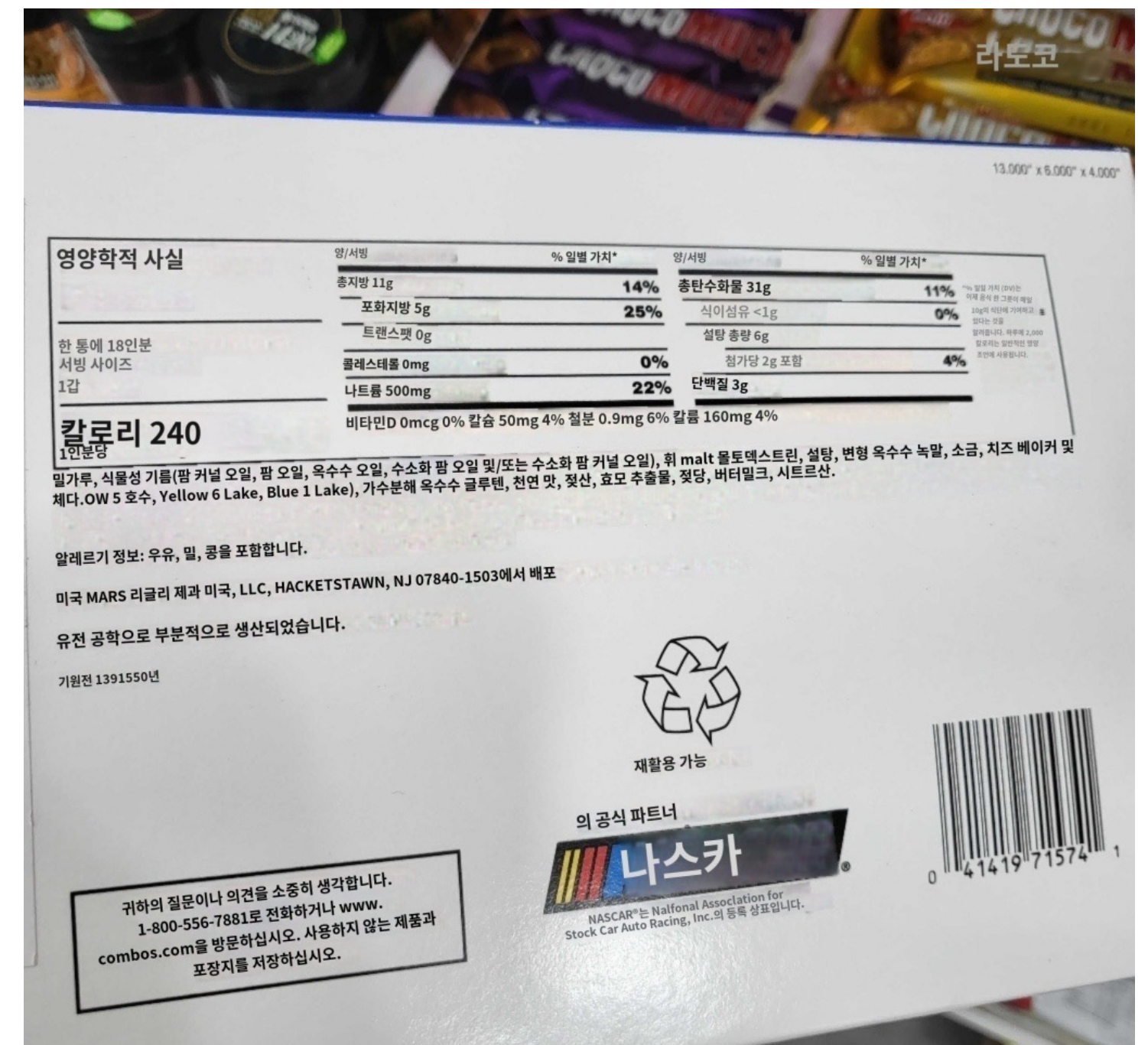
4.2 AR 이미지 번역 엔진 설계

엔진 모듈 구현

대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링



변환 및 렌더링



첫 프레임

이후 프레임

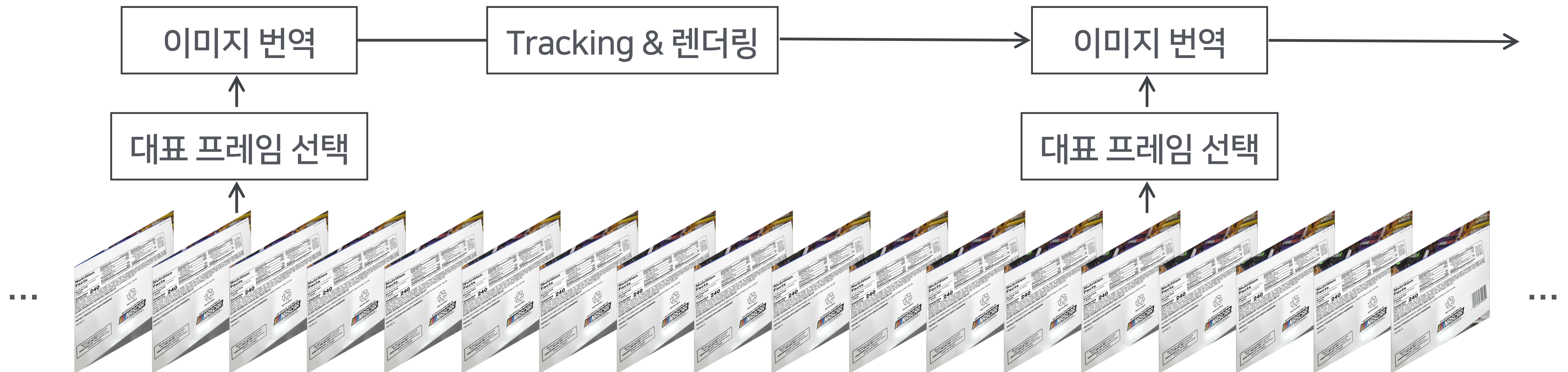
4.2 AR 이미지 번역 엔진 설계

엔진 모듈 구현

대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링



일정 시간 경과 or 품질 저하 감지 시 새로운 사이클 시작



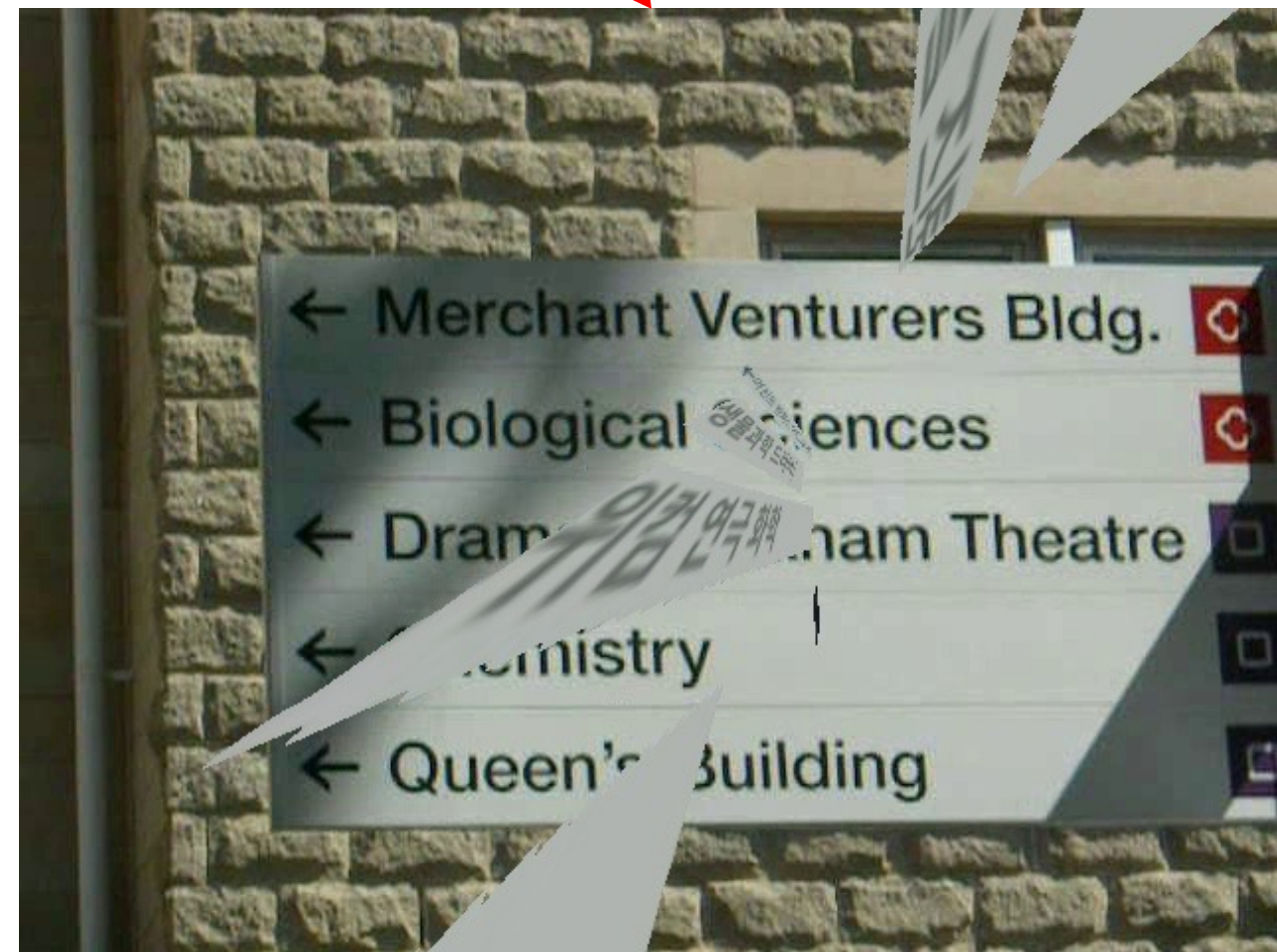
4.2 AR 이미지 번역 엔진 설계

엔진 모듈 구현

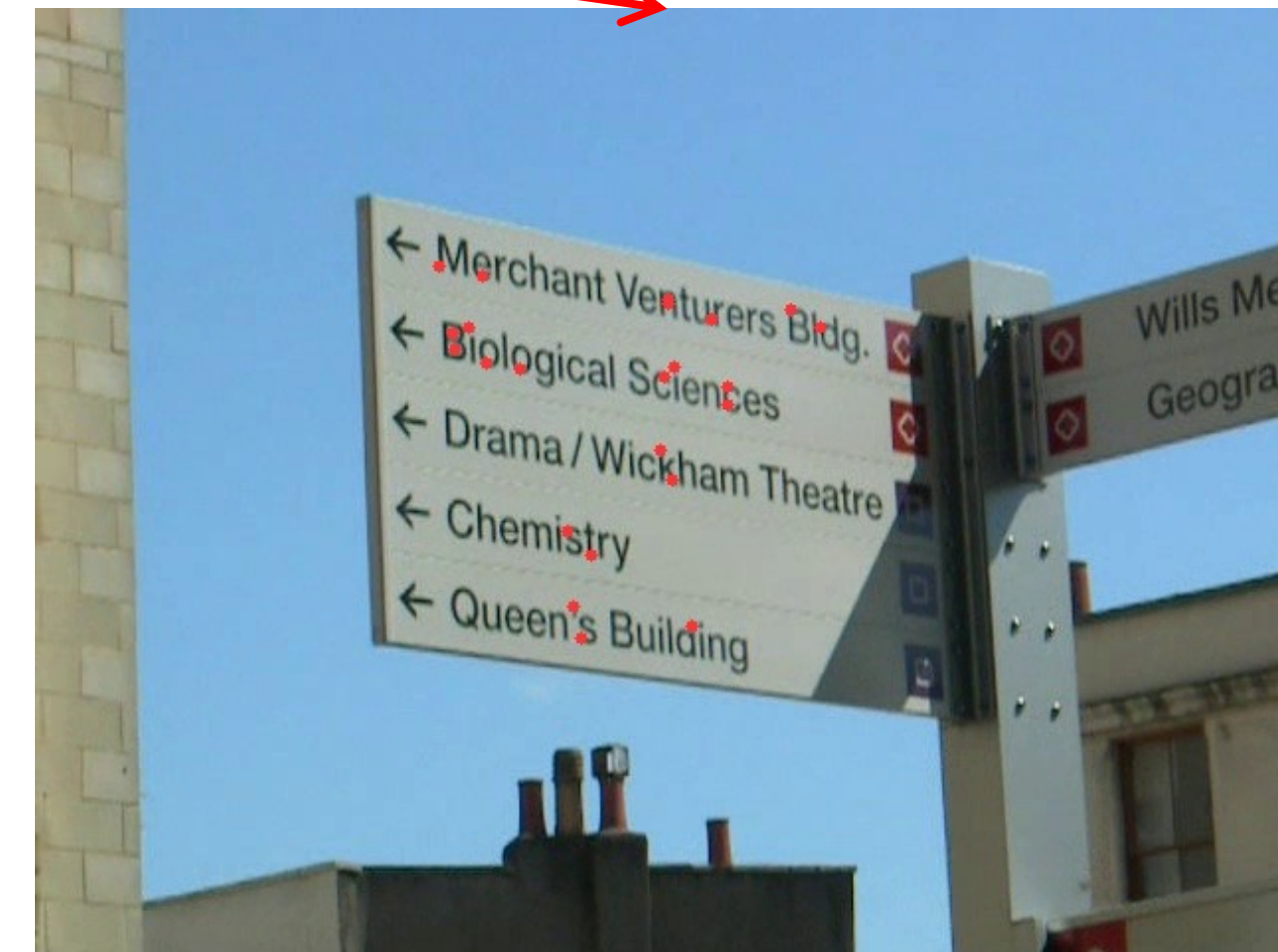
대표 프레임 선택 → 이미지 번역 → Tracking & 렌더링



일정 시간 경과 or 품질 저하 감지 시 새로운 사이클 시작



비정상 homography

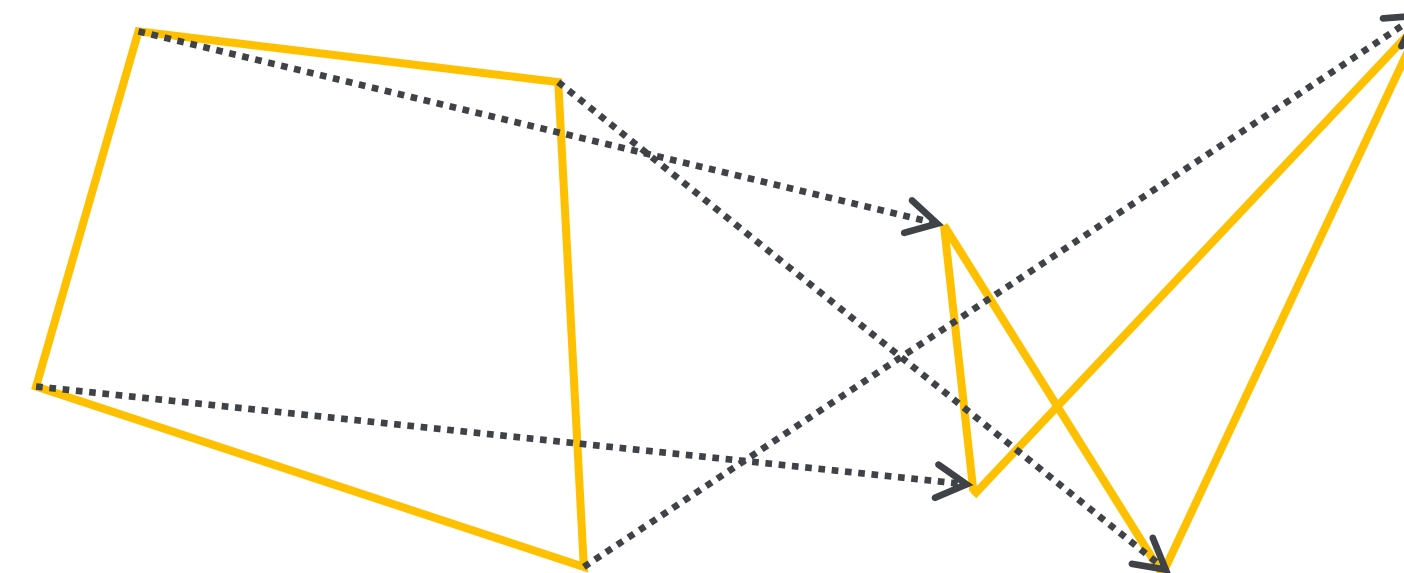
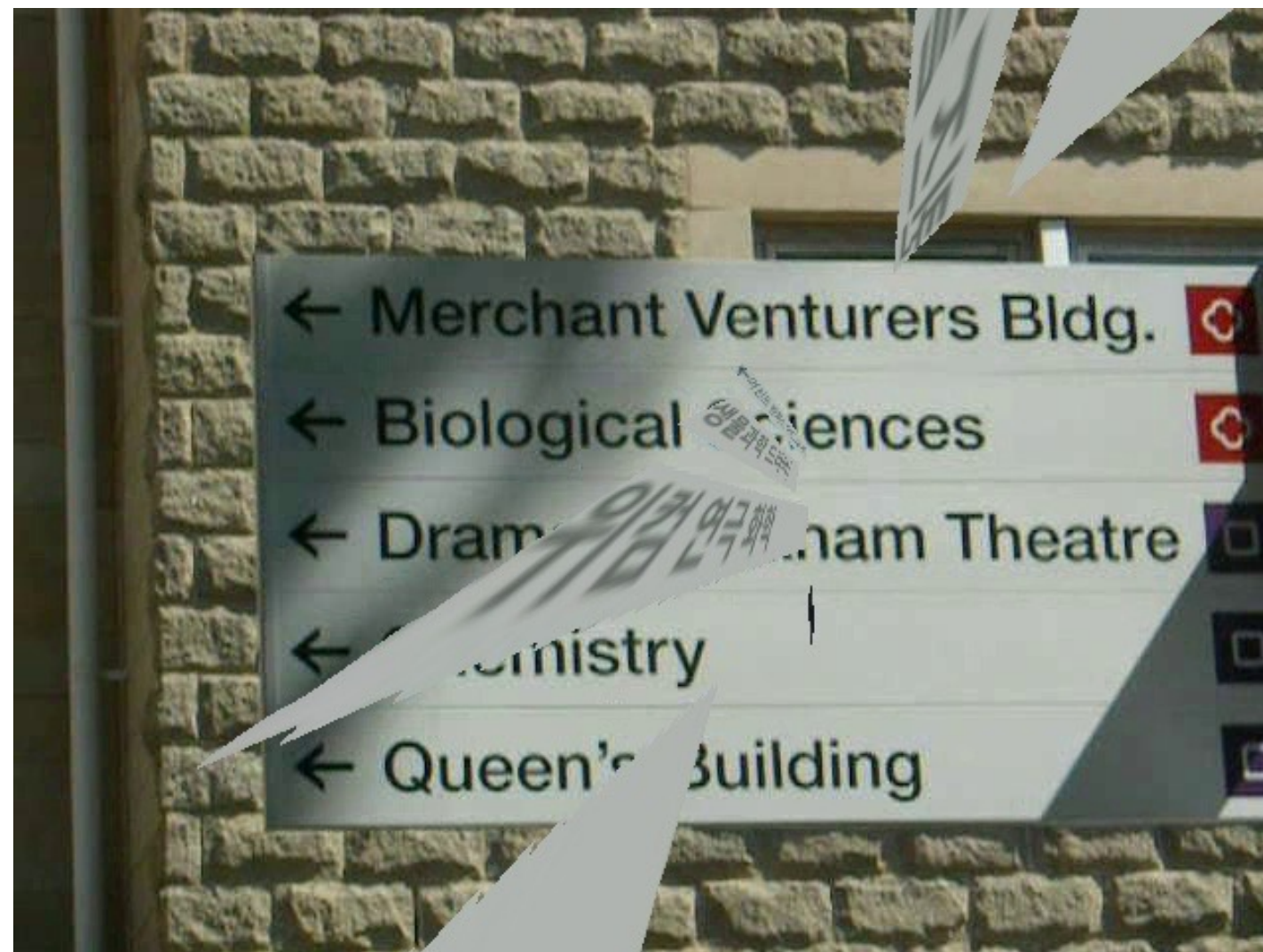


Tracking points 소멸

4.2 AR 이미지 번역 엔진 설계

문제 해결 1: 비정상 homography 검출

- 비정상적인 변환을 초래하는 homography들이 존재
(reflection, twist, excessive scale/perspective 등)
- Determinant 부호 검사, 원소 크기 검사 등을 이용해 검출 후 tracking 재시작

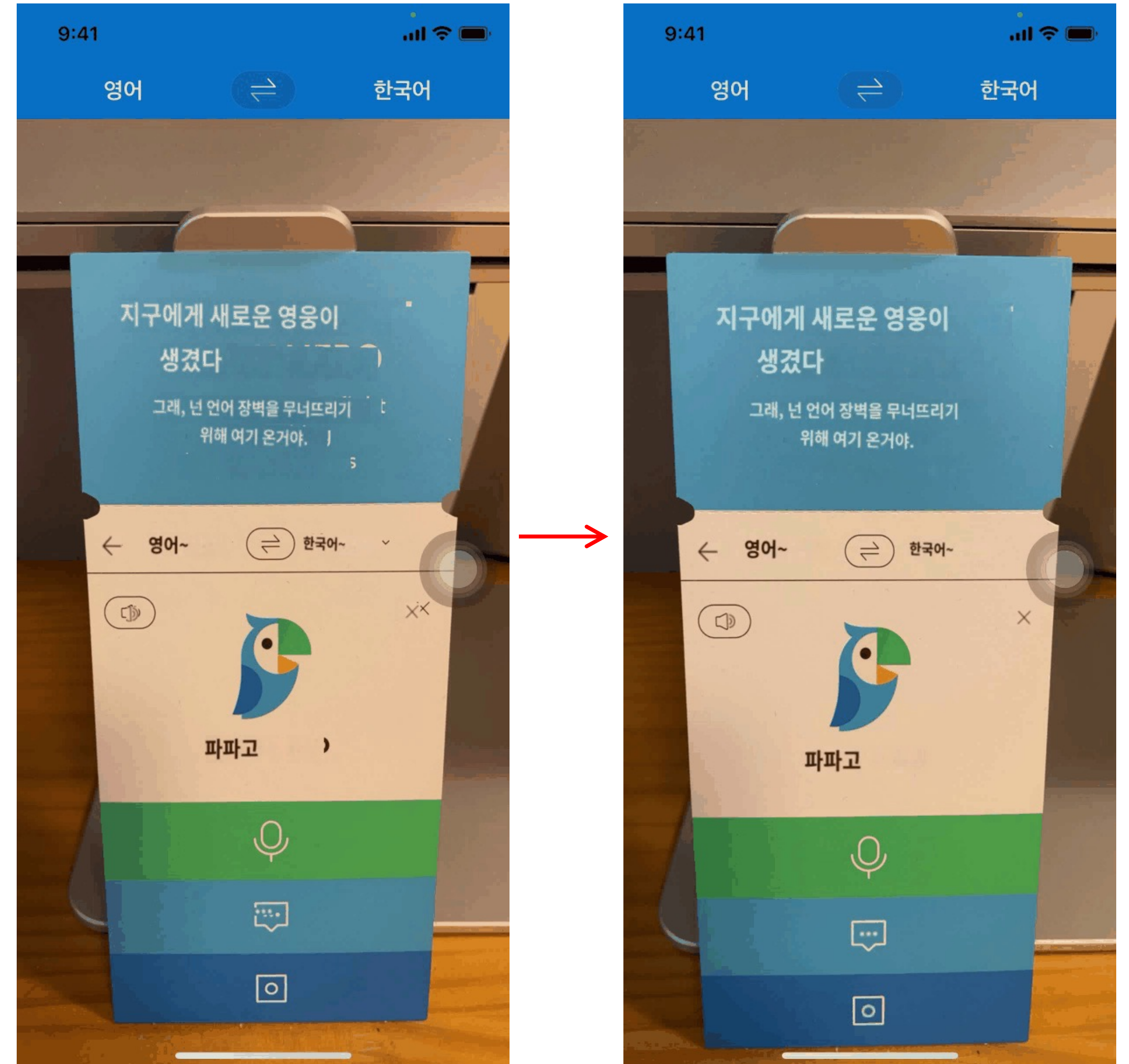


<Twist의 예>

4.2 AR 이미지 번역 엔진 설계

문제 해결 2: 떨림 현상(jitter) 완화

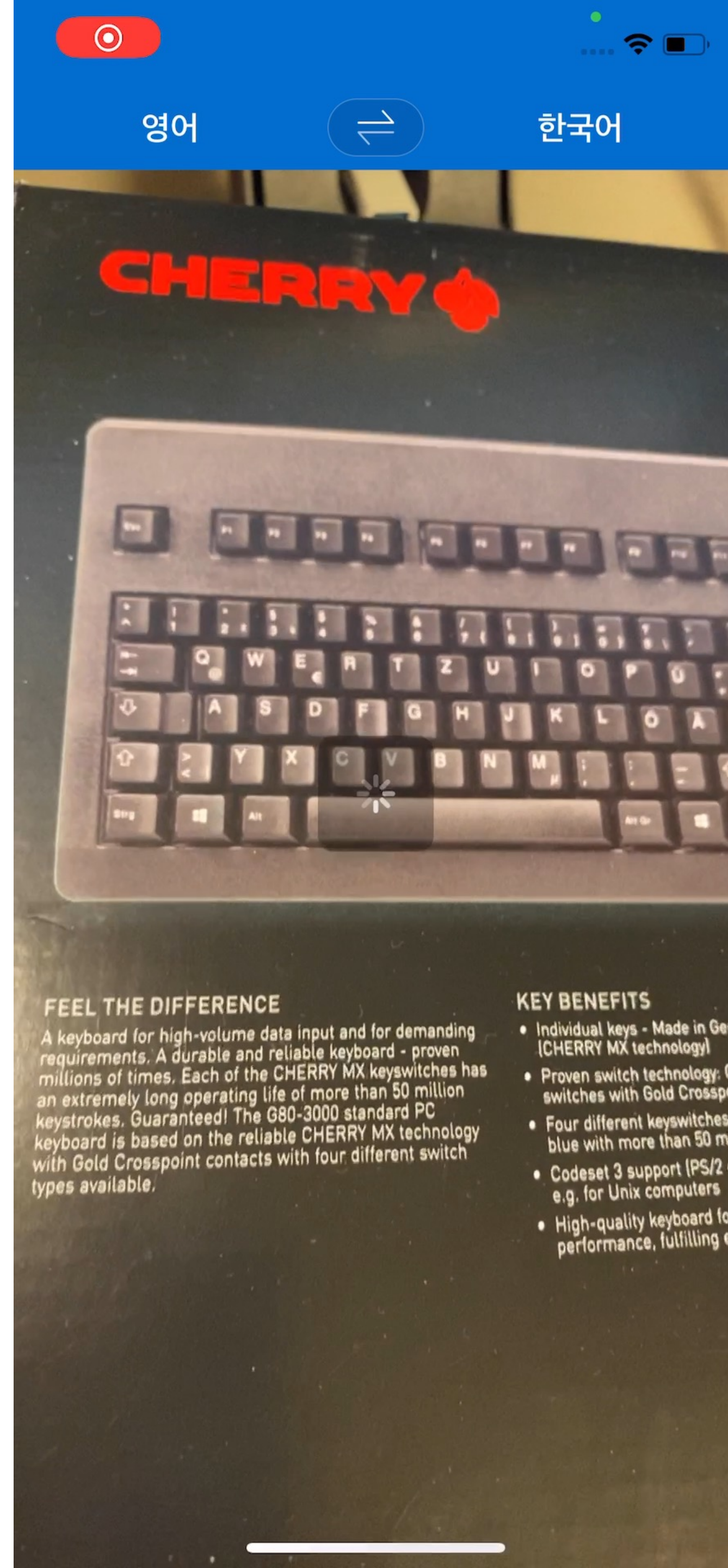
- Optical flow 결과의 부정확성, RANSAC 알고리즘의 랜덤 샘플링 등의 이유로 발생
- Optical flow와 RANSAC에서의 엄격한 outlier filtering을 통해 완화



4.3 데모 영상



4.3 데모 영상



발표 정리

- 파파고 이미지 바로 번역
- 리스트 형식의 텍스트 번역: BERT를 활용한 텍스트 블록 분리
- 텍스트 인페인팅: GAN을 활용한 텍스트 영역 제거
- AR 이미지 번역: 엔진 설계 및 데모



Thank You