

영양 성분 인식을 위한 FOTS 기반 OCR 구현

이나현⁰, 신재영*, 이수민*, 정유철*

*금오공과대학교 컴퓨터공학과,

⁰금오공과대학교 컴퓨터공학과

e-mail: {nahyun286⁰, njs05036*, suminlee2323*}@gmail.com, jyc@kumoh.ac.kr*

FOTS based OCR Implementation for Nutritional Component Recognition

Na-hyeon Lee⁰, Jae-young Shin*, Su-min Lee*, Yu-chul Jung*

*Dept. of Computer Engineering, Kumoh National Institute of Technology,

⁰Dept. of Computer Engineering, Kumoh National Institute of Technology

요약

사람들이 체중을 조절하고 건강을 관리하기 위한 방법 중 하루 영양소 섭취량을 조절이 있다. 현대 사회에선 가공식품의 섭취량이 증가함에 따라 자연스럽게 가공식품들의 영양소를 파악하고 기록하는 것도 중요한 문제가 되었다. 본 논문에서는 실제 가공 식품의 포장지에 인쇄되어있는 영양성분 표 이미지를 인식할 수 있는 OCR을 FOTS 기반으로 구현 및 실험을 진행하였다. 실제로 시중에서 파는 영양성분 표는 한글과 영어가 섞여 있기 때문에 한글을 인식하는 모델과 영어와 숫자를 인식하는 모델을 따로 학습한 뒤 생성하여 각 언어에 대한 인식률을 향상시켰다.

키워드: OCR(Optical Character Recognition), FOTS, 영양성분 (Nutritional Component)

I. Introduction

사람들이 영양 섭취량에 따른 맞춤형 건강 및 식단을 관리하는 서비스를 많이 이용하는 추세에 맞추어 제품에 기재된 영양정보를 읽는 OCR을 구현한다.

본 논문에선 기존 알고리즘을 학습 유형을 달리하여 학습 시, 손실 정도가 낮은 방법을 선택한다. 실험 결과 언어 별로 따로 학습한 뒤 각각 인식하는 것이 더 좋은 결과를 보였다.

영양정보	중간영양성분표	160 kcal	1243, 81, 1321, 76, 1325, 146, 1247, 151, ml
총 칼로리	100%	160 kcal	1135, 1611, 1251, 1613, 1250, 1725, 1134, 1724, 26
총 지방	100%	27g	833, 78, 898, 74, 895, 144, 829, 140, g
나트륨	110 mg	6%	1210, 302, 1331, 303, 1331, 390, 1210, 389, the
탄수화물	29g	9%	634, 314, 785, 304, 790, 385, 640, 395, 19u
당류	27g	27%	615, 596, 639, 593, 632, 700, 618, 785, 299
지방	2.4 g	4%	35, 314, 258, 298, 264, 384, 41, 400, F1e
트랜스지방	0 g	0%	256, 1475, 494, 1478, 492, 1604, 254, 1600, 159
포화지방	2g	13%	147, 103, 487, 210, 484, 249, 144, 242, 알정보
콜레스테롤	5 mg	2%	118, 1088, 498, 1086, 499, 1132, 119, 1135, 트랜스지방
단백질	5g	9%	53, 1448, 395, 1457, 391, 1581, 49, 1572, 단백질
칼슘	180 mg	26%	34, 419, 302, 414, 305, 540, 37, 545, 나트륨

Fig. 1. 인식 결과

II. Preliminaries

1. Outline

본 연구에서는 영양성분 표 인식을 위해 FOTS 알고리즘[1]을 사용한다. 이 알고리즘은 End-to-End 방식으로 이미지에서 문자 영역을 찾는 Detection과 해당 문자 영역에서 텍스트로 문자를 읽어오는 Recognition인 두 단계를 하나로 통합하여 약 2배의 시간 절감과 더 많은 특징 맵 학습이 가능하다. 또한 ROI Rotate를 사용하여 비스듬한 글자에 대해서도 인식을 할 수 있다.

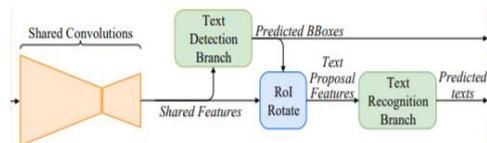


Fig. 2. FOTS End-to-End 구조

2. Dataset

한글 데이터 셋은 AI Hub[2]에서 제공하는 한국어 글자체 이미지 데이터 약 89,000장을 사용한다.

영어, 숫자 이미지는 ICDAR 2017MTL[3], 2015[4]와 SynthText 데이터[5] 약 438,200장을 사용한다.

추가 데이터 셋은 영양성분 표 내의 문자를 포함하여 약 600장의 데이터를 제작하여 보충하였다.

III. The Proposed Learning Method

한글, 영어, 숫자 데이터 셋을 한꺼번에 학습하는 경우와 한글, 영어숫자 데이터 셋으로 나누어 각각 학습시키는 경우로 나누어 실험을 진행하였다

1. Detection

최저 loss는 한글 데이터만으로 학습 시 0.005, 한글과 영어로 학습 시 0.0053이 나왔다. 모든 언어로 학습시킨 체크포인트에 한글로 튜닝 시 0.0038로 가장 낮은 loss를 보였다.

2. Recognition

최저 loss는 한글 데이터만으로 학습 시 3.323, 한글과 영어로 학습 시 3.868이 나왔다. 모든 언어로 학습시킨 체크포인트에 한글로 튜닝 시 0.204로 가장 낮은 loss를 보였다.

모든 언어를 한꺼번에 학습할 때보다 언어별로 나누어 학습할 때 더 낮은 loss를 보였다. 모든 데이터로 학습시킨 체크포인트를 이용해 한글 인식 전용 체크 포인트와 영어와 숫자 인식 전용 체크 포인트를 생성한다. 실제 이미지 인식 시, 두 체크 포인트를 이용하여 한글과 영어, 숫자를 각각 인식한다.

IV. Experiment

1. 한글 학습

Table 1. 한글 학습 시 최저 loss

	최저 loss
Detection	0.00372
Recognition	0.18625

한글 인식에서 최저 loss가 위와 같이 측정되었다. Recognition의 경우 최저 loss가 낮아졌다.

2. 영어, 숫자 학습

Table 2. 영어, 숫자 학습 시 최저 loss

	최저 loss
Detection	0.00234
Recognition	0.12773

영어와 숫자 인식에서 최저 loss가 위 표와 같이 측정되었다. Recognition의 경우도 한글과 유사하게 최저 loss가 줄어든 것을 알 수 있다.

3. 실제 이미지 인식

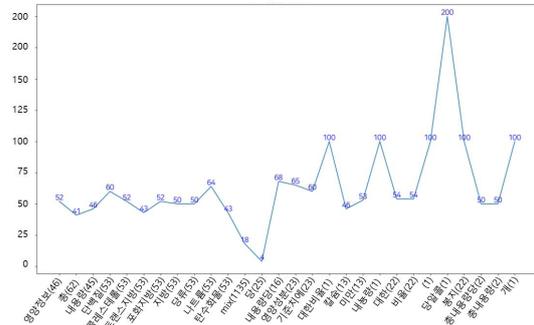


Fig. 3. 전체 테스트 데이터 셋에 대한 인식 비율

테스트 이미지 53장에 대해 인식한 단어와 정답 단어를 비교하여 정확도는 아래와 같이 계산하였다.

$$\text{정확도} = \frac{\text{단어를 올바르게 인식한 횟수}}{\text{전체 이미지속 단어의 출현 빈도}} * 100 \dots \text{식 (1)}$$

Fig. 3의 x축 레이블은 ‘단어(총 출현 빈도수)’로 표시하였다. 레이블 중 “mix”는 영어와 한글이 같이 인식된 단어이다. ‘당’과 같은 경우는 4%의 아주 낮은 인식률을 보였고, “당알콜”의 경우 실제 이미지에선 1장에 1단어 밖에 없었지만, 2번 인식되어 200%라는 수치가 나오게 되었다.

V. Conclusion

테스트 데이터 셋에 대해 약 59%의 인식률을 보였다. 영양성분 표의 핵심 단어들은 잘 읽어왔다. 현재, 한글, 영어보다 숫자 데이터가 많이 부족해 실제 인식에서도 숫자를 정확하게 읽어오지 못하는 경우가 있어서, 향후 숫자 데이터를 확장하여 추가로 성능개선을 계획하고 있다.

REFERENCES

[1] Xuebo Liu, Ding Liang, Shi Yan, Dagui Chen, Yu Qiao, and Junjie Yan, FOTS: Fast Oriented Text Spotting with a Unified Network, 2018.
 [2] AI Hub, <https://www.aihub.or.kr/aidata/133>
 [3] ICDAR2017 Competition on Multi-lingual scene text detection and script identification, <https://rrc.cvc.uab.es/?ch=8>
 [4] Incidental Scene Text, <https://rrc.cvc.uab.es/?ch=4>
 [5] SynthText in the Wild Dataset, <https://www.robots.ox.ac.uk/~vgg/data/scenetext/>