

CUT-GAN 과 Detectron2 기반 라벨지 자동 인식 및 추출 방법

End-to-End Table Content Recognition and Extraction Method: using CUTGAN and Detectron2

남석현
asht05166@naver.com

ggccillee222@naver.com
박도훈
ggp04186@naver.com

jhk@knu.ac.kr
정상훈
hoon3722@gmail.com

이세종

*김정홍

경북대학교 컴퓨터학부
대구광역시 북구 대학로 80

요약

산업에서 널리 쓰이는 재고관리 시스템은 형태의 라벨지를 사용함으로, 이를 컴퓨터가 자동으로 인식하는데 어려움이 있다. 본 연구에서는 이를 개선하기 위하여 재고 관리를 위한 라벨지 자동 인식을 수행하는 머신러닝 기반 방법을 제안하였다. 제안할 방법은 CUT-GAN 을 통하여 다양한 환경에서 라벨지 사진을 생성하고, Detectron2 를 통하여 라벨지 내 표의 자동 인식률을 높였다. 라벨지 내의 표를 인식하여 OCR 을 수행하는 이 방안은 기존의 OCR 방식보다 더욱 정확하게 디지털 인식을 수행함으로써 업계에서의 업무 디지털화를 앞당기는 효과가 있을 것으로 기대한다.

키워드: CUT-GAN, Detectron2, OCR, 라벨지, 자동 인식

1. 서론

최근 산업에서는 재고관리를 위해 스마트 기기를 통한 자동 입력 방식이 증가하고 있다. 그러나, 각 회사마다 통일되지 않은 형태의 라벨지를 가지고 있으며, 영세한 업체에서는 라벨지를 출력한 후, 수기로 내용을 기입하기도 한다. 이에 따라, 필기체와 인쇄체가 혼용된 라벨지를 자동으로 인식할 수 있도록 하는 오픈소스 기반의 OCR 모듈을 개발하고자 본 연구를 진행하였다.

라벨지의 대부분은 표 형태로 구성되는데, 이를 탐지하기 위해서 대표적으로 사용하는 딥러닝 모델로는 Detectron2[1]가 있다. 이 모델을 이용한

표의 인식을 연구한 사례로 TableBank[2]가 존재하는데, 해당 연구에서는 디지털 문서 내부의 Table Detection 을 목표로 Detectron2 에 학습시켰다. 이로 인하여 실생활에서 사용되는 다양한 표의 사진을 탐지하는 것에 있어서는 충분히 좋은 성능을 보이지 못한다는 한계가 존재한다. 이를 개선하기 위해서는 실생활에서 촬영된 표의 이미지를 인식할 수 있도록 추가적인 학습이 필요한데, 본 연구에서는 이것을 CUT GAN Model[2]을 통하여 해결하고자 하였다.

CUT 을 통하여 다양한 환경에서 촬영된 표의 이미지를 생성하고 이를 Detectron2 의 학습용 데이터로 활용하였다.

2. 관련연구

이번 연구에서는 기존 TableBank 의 표 이미지만을 사용했을 때보다 더 정확하게 테이블 구조 인식을 할 수 있게 하기 위해 GAN(Generative Adversarial Network)을 사용한다. 적대적 생성모델 GAN 은 역할에 따라 Generator(G), Discriminator(D) 두 모델로 나뉜다. G 는 실제로 촬영된 표 이미지를 따라한 가짜 표 이미지를 생성한다. D 는 G 가 생성한 가짜 표 이미지와 실제 표 이미지를 구별하는 역할을 한다. G 와 D 두 모델은 각각의 목적을 달성하기 위해 서로 경쟁 하며 학습한다. 이 두 모델의 목적을 하나로 합친 수식은 다음과 같다.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))].$$

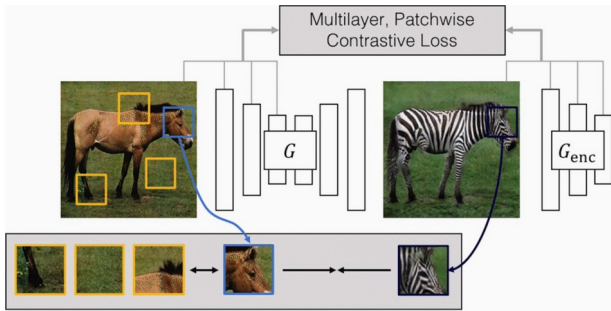


그림 1. CUT 모델 구성도

그림 1 은 CUT[3] 모델의 구성을 나타낸다. CUT 는 CycleGAN 의 문제로 알려진 Cycle Consistency 를 보완할 수 있는 Contrastive Learning 기법을 사용한다.

표 1. CUT 과 타 GAN Model 의 비교

Method	Cityscapes				Cat→Dog		Horse→Zebra	
	mAP↑	pixAcc↑	classAcc↑	FID↓	FID↓	FID↓	sec/iter↓	Mem(GB)↓
CycleGAN [89]	20.4	55.9	25.4	76.3	85.9	77.2	0.40	4.81
MUNIT [44]	16.9	56.5	22.5	91.4	104.4	133.8	0.39	3.84
DRIT [41]	17.0	58.7	22.2	155.3	123.4	140.0	0.70	4.85
Distance [4]	8.4	42.2	12.6	81.8	155.3	72.0	0.15	2.72
SelfDistance [4]	15.3	56.9	20.6	78.8	144.4	80.8	0.16	2.72
GCGAN [18]	21.2	63.2	26.6	105.2	96.6	86.7	0.26	2.67
CUT	24.7	68.8	30.7	56.4	76.2	45.5	0.24	3.33
FastCUT	19.1	59.9	24.3	68.8	94.0	73.4	0.15	2.25

표 1 은 CUT 모델의 논문[3]에서 발췌한 CycleGAN, GCGAN 등의 다양한 GAN Model 과 CUT 의 sec/iter, Mem, FID 등을 측정하여 비교한 결과물이다. CUT 은 CycleGAN 대비 학습에

걸리는 시간이 적고 GPU 메모리 자원을 적게 사용하면서 낮은 FID 값을 보인다. 또한, 다양성 높은 이미지 생성을 위한 학습에 적절하다고 판단되어 본 연구에서 사용할 GAN 모델로 채택되었다. CUT 으로 생성된 이미지를 Detectron2 에 훈련시킬 시, 본 연구에서 제안하는 모듈을 통하여 비정형화된 텍스트 이미지를 인식할 수 있을 것으로 기대한다.

3. 실험 및 결과

CUT 을 통해 생성된 이미지를 이용하여 Labeling 을 실시한 후, 이를 기반으로 Detectron2 모델을 학습시켜 기존 TableBank 의 데이터셋으로 학습된 Detectron2 대비 향상된 정확도를 가지는지를 실험한다. CUT 과 Detectron2 모델은 Google Colab Notebook 으로 작성된 기본 코드를 활용하였으며, Labeling 은 Roboflow 내의 COCO Format Annotation 을 사용하였다.

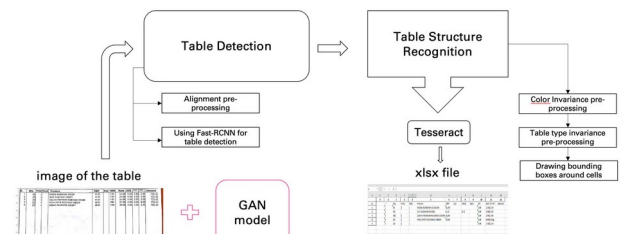


그림 2. 제안한 라벨인식 모듈 구성도

그림 2 는 라벨인식 모듈의 전체 구성도이다. 촬영된 표를 Detectron2 모델로 Table Structure Recognition 을 시행한 뒤, 생성된 Annotation 을 기반으로 해당 이미지를 자동으로 Crop 하고 Color Invariance Preprocessing, Table type invariance Preprocessing, Greyscale Conversion 등의 전처리를 거친 후 Tesseract 를 통하여 OCR 을 진행하였다.

3.1. 데이터셋

실생활에서 표를 촬영한 결과물은 별도의 보정 절차를 거치지 않으면 컴퓨터로 생성된 표와 큰

차이를 보인다. 이미지의 구도나 거리 등에 따라, 표의 형태가 평행사변형의 모양을 가지기도 하고, 표의 배경이 되는 종이의 범위 또한 달라진다. 또한, 화질의 열화도 이루어진다. 이러한 요소들을 Detectron2 에 학습시키기 위해, CUT 의 학습에는 배경의 상태와 구도 등에 변화가 있는 표의 이미지를 선정하였다.

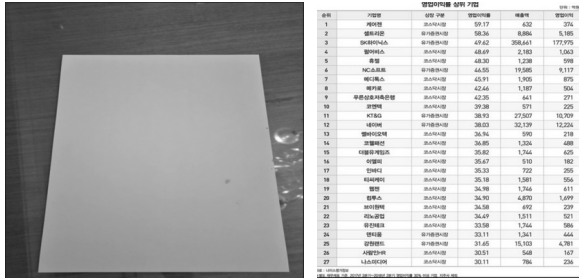


그림 3. (A)배경 이미지 (B)표 이미지

그림 3은 CUT 의 학습을 위한 배경 이미지와 표 이미지이다. 각 이미지를 imageA, imageB 로 구분하여 A 에서 B 로 점층적 변화를 일으키도록 학습하였다.



그림 4. CUT 으로 생성된 이미지

그림 4 는 배경 이미지인 A4 용지 내에 표 이미지가 삽입된 결과물이다.

3.2. 실험환경 및 결과

본 연구에서 별도로 확보한 800 개의 이미지를 기반으로 Detectron2 의 표 인식 정확도를 평가함으로써 TableBank 데이터셋에 추가로 CUT 을 이용해 생성된 600 개의 이미지를 학습하였을 때의 mAP(mean Average Precision)를

비교하였다. 데이터셋에 사용된 이미지 중 400 개는 컴퓨터로 생성한 표의 이미지를, 400 개는 실제로 촬영된 표의 이미지로 구성하여 각각 TableBank 데이터셋을 통한 학습만으로도 높은 인식률을 보장하는 이미지와 낮은 인식률을 보인 이미지들로 구분하여, 실생활에서 촬영하였을 때의 상황과 유사하게 배분하여 평가하였다.

표 2. mAP 비교

학습 데이터	TableBank	TableBank + CUT
데이터 개수	417234	417834
mAP	0.823	0.856

표 2 는 TableBank 로 학습된 모델과 CUT 을 통해 생성된 이미지로 추가 학습된 Detectron2 모델의 mAP 비교를 나타낸다. TableBank 데이터셋만을 학습하였을 때에 비하여 0.033 의 mAP 향상이 있었음을 확인하였다.

4. 결론

본 연구는 학습 데이터 생성에 CUT 을 사용하였고, 이를 통해 실제 표 이미지에 대한 인식률을 향상시키는 것을 목표로 하였다. 그 결과, 다양한 환경에서 촬영된 실제 표 이미지에 대한 인식 정확도는 기존의 데이터셋 대비 적은 양의 데이터로도 상당한 개선을 보였다. 그러나 이러한 방법에도 불구하고 몇 가지 문제점이 파악되었다. CUT 을 활용한 경우, 생성된 표 이미지의 해상도가 상대적으로 낮아지며, 항상 정확한 표 형태의 이미지를 만들어내지 못하는 문제가 있었다. 이러한 한계점을 극복하기 위한 연구 방향으로 표 이미지를 생성하는 알고리즘에 보다 적절한 GAN 모델을 적용하는 것을 제안한다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음(2021-0-01082).

참 고 문 헌

- [1] Yuxin Wu, Alexander Kirillov, Francisco Massa, WanYen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2/>, 2019.
- [2] Li, Minghao, et al. "Tablebank: A benchmark dataset for table detection and recognition." *arXiv preprint arXiv:1903.01949* (2019).
- [3] Park, Taesung, et al. "Contrastive learning for unpaired image-to-image translation." *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part IX 16*. Springer International Publishing, 2020.
- [4] R. Smith, "An Overview of the Tesseract OCR Engine," *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Curitiba, Brazil, 2007, pp. 629-633, doi: 10.1109/ICDAR.2007.4376991.
- [5] 손윤식, 이혜린, 박소희, 신지원.(2019).정보 활용성 확장을 위한 Tesseract OCR 기술 이용 방안. 한국정보과학회 학술발표논문집(),1278-1280.