다양한 조명 환경에 강인한 seven-segment OCR 방법

김진성[1], 노가은[1], 남현길, 박종일†
한양대학교
{towetous, shqmffl486, skagusrlf, jipark}@hanyang.ac.kr

Robust seven-segment OCR method for various illumination environments

Jinsung Kim Gaeun Noh Hyeongil Nam Jong-Il Park
Hanyang University

요 약

본 논문은 인식이 어려운 조명 환경에도 강인한 seven-segment 문자 인식을 위해서, 영상 내에 다양한 조명 연출이 가능하도록 합성 데이터 셋을 생성하고 학습할 수 있는 OCR 방법을 제안한다. 기존 연구에서는 deblurring 과 같이 영상 이미지의 해상도를 높여 문자 인식의 정확도를 향상시키는 것에 초점을 두었으나, 여러 조명 환경에 대비할 수 있는 OCR 관련 연구들은 부족하다. 이를 해결하기 위해 본 논문에서는 문자가 포함된 자연스러운 배경 영상에, seven-segment 문자를 합성시킨 후 relighting 을 적용함으로써 실제 환경과 유사한 장면을 연출해 새로운 합성 데이터 셋을 생성한다. 그리고 생성된 데이터 셋을 딥러닝 기반 학습시켜 다양한 조명에도 강인한 문자 인식기를 만들고자 한다. 합성 데이터 셋의 사용여부와 일반적인 데이터 augmentation 기법의 사용 여부를 비교하여, 본 논문에서 제안한 방법의 효과를 확인할 수 있었다. 이를 통해서 seven-segment 문자 인식 뿐만 아니라, 다양한 문자에 대해서도 적용될 수 있는 초석이 될 것으로 기대된다.

1. 서론

본 논문에서 언급할 OCR(Optical Character Recognition, 광학 문자 인식)은 이미지(사진) 속 문자 위치를 찾고 어떤 문자인지 자동으로 알아내는 기술을 뜻한다. 주로 Naver Clovaai OCR, Google drive OCR 등 여러 회사에서 주요 비즈니스 활용에 최적화된 OCR 인식 모델을 개발하였다. 최근에는 영수증, 신용카드, 사업자 등록증, 명함 그리고 신분증과 같은 문서의 주요 특징을 추출하는 document OCR 모델이 나오면서 점차 확장되고 있다. 하지만 정규화된 글자 이외로 사람마다 다양한 글씨체와 크기로 인해 좋은 성능을 가지고 있는 OCR 프로그램이어도 정확히 파악하고 인식하기 어렵다. 예를 들어 높고 낮은 해상도, 일정하지 않은 문서의 크기, 필기체의 경우에는 OCR 정확도가 현저히 떨어진다.

현재 seven-segment 영상에서 숫자와 배경을 명확히 하기 위해서 영상 이진화를 수행하거나 크기를 정규화 하는 연구들이 있다. 하지만 선행 연구에서는 OCR 정확도를 높이기 위한 연구들이 많고 활용된 딥 러닝 네트워크의 경우 CNN 과 같은 간소화된 네트워크 구조를 활용하는 것에 초점을 맞추었다[8]. 그리고 알고리즘의 문자 인식 범위는 대부분 깔끔한 영상에서 글자를 검출했기 때문에 성능을 개선하는 부분에서 한계가 있다. 또한 대부분 deblurring 과 같이 해상도를 높여 문자 인식 정확도를 높이는 것에 초점을 두지만, 불충분한 외부 광원으로 인한 조명 대비가 어두운 seven-segment 를 검출하는 연구들이 부족하다. 주로, 오픈소스인 Tesseract OCR 엔진을 사용하지만, Tesseract OCR 이 자동으로 전처리를 하지 못하기 때문에 인식률이 떨어지는 단점이 있다.

이를 해결하기 위해서 본 논문에서는 많은 OCR 문제들 중에 seven-segment 문제에 초점을 두고 방안을 제시한다. 문자가

[1] 공동저자, †교신저자

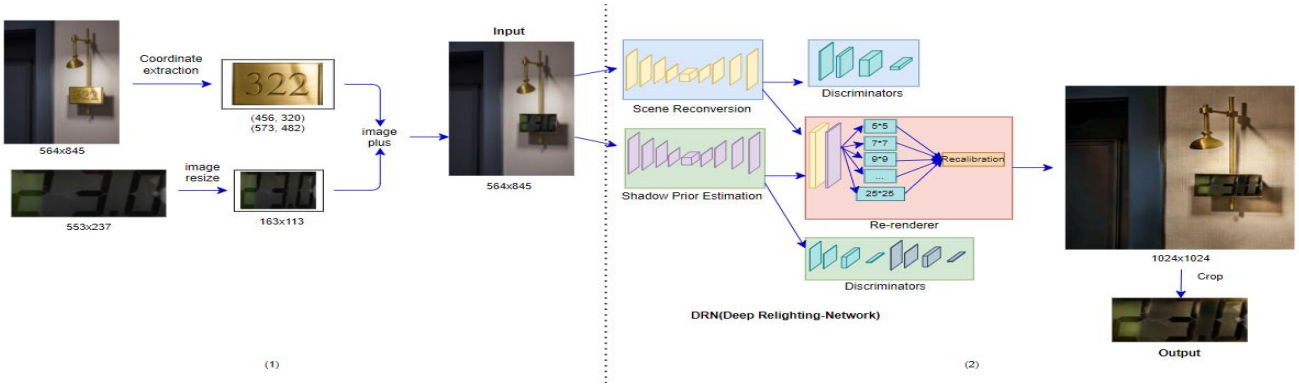

그림 1. Wild scene 에서 relighting 합성 데이터 셋 생성 과정

포함된 2D 영상에서 DRN(Deep Relighting Network)을 응용해 이미지의 광원을 OCR 에서 인식하기 어려운 실제 환경 기반의 seven-segment 합성 데이터 셋을 생성하고 조명 효과에 강인한 OCR 방법을 제안한다.

2. 합성 데이터 셋 생성

본 논문에서는 조명 효과에 강인한 seven-segment OCR 을 위한 합성 데이터 셋 생성 과정을 그림 1 과 같이 나타내었다. 전체 과정을 개괄적으로 살펴보면, 기존의 seven-segment 데이터 셋에 조명 효과를 주기 위해 DRN 을 통해 합성 데이터 셋을 만드는 구체적인 과정은 다음과 같다.

2.1 Relight

OCR 문자 인식기의 정확도는 다양한 조명 환경 및 조명의 강도에 크게 영향을 받는다. Seven-segment OCR 관련한 데이터 셋은 이미 많이 존재하지만 wild scene 에서의 다양한 조명환경 및 조명의 강도에 관련한 데이터 셋은 존재하지 않고 수집을 통해 데이터 셋을 만드는 것에는 시간적 비용적 한계가 존재한다. 본 논문에서는 이를 해결하기 위해 다양한 조명 효과를 줄 수 있는 relighting 효과를 기존의 데이터 셋에 적용시켜 실제환경과 유사한 새로운 합성 데이터 셋을 생성하는 방법을 제시한다. Relighting method 에는 U-Net[2], Retinex-Net[3], Pix2Pix[4], DRN(Deep Relighting Network)[5] 등이 존재한다. 본 논문에서는 DRN 이 이중 혹은 어두운 광원을 포함한 상황에서 relighting 이 효과적이므로 해당 방법을 활용한다(그림 1). 문자가 포함된 이미지를 선택하고 선택된 배경 이미지에서 관심 영역(배경 이미지의 문자가 포함되어 있는 위치)의 좌표 값을 구해 crop 하면 관심 영역의 사이즈를 구할 수 있다. Crop 된 seven-segment 데이터를 관심 영역의 사이즈에 맞게 조정하고 배경 이미지의 관심 영역 위치에 결합하면, DRN 네트워크의 input 이미지 형태로 전처리 할 수 있다. (그림 1. (1))

그 다음으로 전처리 된 input 이미지를 scene reconversion network 와 shadow prior estimation network 에 각각 통과시켜 그림자에 영향을 주는 구조물들에 대한 특징과 다른 방향의 광원으로부터 생성될 그림자에 대한 특징을 얻게 된다. 이 과정에서 discriminators 는 scene reconversion network 와 shadow prior estimation network 각각의 예측 값과 ground-truth(g.t)의 차이를 줄이는 방향으로 학습하는 것을 돕는다. 그리고 두 네트워크로부터 얻은 특징을 다양한 크기의 필터에 통과시켜 공간적 패턴들을 추출하고 recalibration 과정을 통해 패턴들 중 가중치가 큰 특징들을 추출할 수 있게 된다. 추출된 특징들을 이용하여 relighting 된 output 이미지를 얻을 수 있고 output 이미지로부터 관심 영역을 crop 해 새로운 합성 데이터 셋을 생성할 수 있다(그림 1. (2)). 또한 합성 데이터 셋 생성 시에 전처리 된 Input 이미지의 사이즈에 따라, recalibration 되는 과정에서 공간적 패턴들의 가중치가 다르게

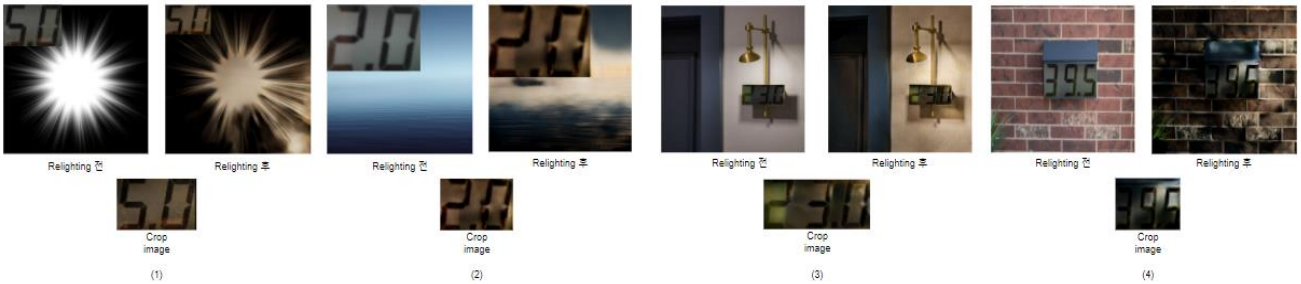

그림 2. 조명 환경에 따른 relighting 적용 전/후 영상
((1), (2) 문자가 없이 임의 배치로 합성된 데이터, (3), (4) 문자가 포함된 영상을 이용하여 합성된 데이터)

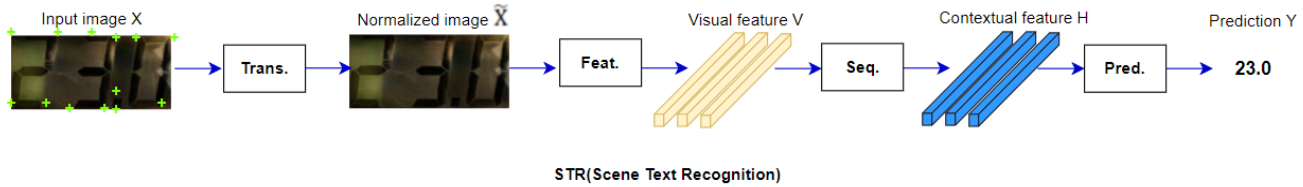

그림 3. Scene Text Recognition 네트워크 구조 및 예측 과정

적용되어 relighting 효과가 각기 다르게 나타난다. 이를 통해 동일한 배경 이미지에 대해서 입력 사이즈에 따라 relighting 효과가 다르게 적용되기 때문에 seven-segment 문자 데이터를 생성해 낼 수 있다.

실제 합성 데이터 셋 생성시, 핀 조명 혹은 자연광이 있는 영상 내에서 이미 문자가 있는 관심 영역에 seven-segment 데이터를 결합하여 relighting 을 적용하여 DRN 의 scene reconversion network 를 통해 구조물에 의해 relighting 되는 효과를 그림 2 와 같이 확인할 수 있다. (그림 2. (1))과 (그림 2. (2))는 문자가 포함되지 않은 배경 이미지에 seven-segment 문자를 임의의 위치에 합성시킨 후 relighting 을 적용한 것이고, (그림 2. (3))과 (그림 2. (4))는 문자가 포함되어 있는 자연스러운 배경 이미지에 seven-segment 문자를 관심 영역에 합성 시킨 후 relighting 을 적용한 것이다.

3. Scene Text Recognition(STR)

본 논문에서 사용한 Scene Text Recognition(STR) 프레임워크는 Naver Clovaai 에서 착안하였다[6]. STR 은 Transformation(Trans.), Feature extraction(Feat.), Sequence modeling(Seq.), Prediction(Pred.) 총 4 단계로 구성되며, 앞에서 언급한 합성 데이터 셋이 이를 거치게 된다(그림 3). 첫 번째 Transformation stage 은 다양한 STR 모델 중에서 thin-plate spline(TPS)을 사용하였다. 그림 3 에서의 Input 이미지로 relighting 을 적용시킨 영상이 들어가면 초록색으로 표시된 '+' 점인 기준점(fiducial points)을 찾은 다음에 문자 영역을 predefined rectangle 으로 바꾸기 때문에 휘어진 글자의 경우를 곧게 펼 수 있다. 두 번째 Feature extraction stage 에서는 RCNN, VGG 등 여러 모델 중 ResNet 모델을 사용하였다. 각각의 특징맵에서의 column 은 이미지 자체에 대응되는 receptive field 가 있기 때문에 column 으로 나누어서 receptive field 에서 글자를 예측할 수 있다. 세 번째 Sequence modeling stage 에서는 특징맵에서 각 column 은 sequence 구조로 사용된다. 하지만 이 sequence 에 문맥 정보가 없을 수 있기 때문에 Feat. 단계 이후에 BiLSTM (Bidirectional LSTM)을 사용해서 더 좋은 특징들을 뽑는다. 마지막으로 네 번째 Prediction 단계에서는 sequence 를 예측하는 부분이다. Prediction 을 위해 Attention-based sequence(Attn) 모델을 사용했다. 이 단계에서는 Attn 을 통해 input sequence 내의 정보 흐름들을 캡처해 output sequence 를 예측한다.

4. 비교 실험

본 논문의 실험에서는 문자가 포함된 2D 영상에서 DRN(Deep Relighting Network)을 응용해 이미지의 광원을 OCR 에서 인식하기 어려운 실제 환경 기반의 seven-segment 합성 데이터 셋을 생성했다. 기존의 인식기보다 더 나은 성능을 증명하기 위해 앞에서 언급한 relighting 적용 유무를 비교한다. 또한 일정량의 데이터를 relighting 과 같이 다양한 효과들을 augmentation 시켜 인식 정확도를 높이기 위해 일반적으로 사용되는 augmentation 유무를 비교하여 총 4 가지 경우로 실험한다. 본 실험에서 적용한 augmentation 기법으로 색의 주요한 세 속성인 색상, 채도, 명도를 임의로 변경하는 Color Jitter, 노이즈 필터를 이용한 Random Adjust Sharpness, Random Equalize, Random Auto Contrast, 커널 필터를 이용한 Gaussian Blur 등을 적용했다.

비교 실험을 위한 데이터 셋은 training set 2,152 개 validation set 1,818 개 testing set 50 개로 분류하여 총 4,020 개의 데이터를 만들었다. 또한 합성 데이터 셋 생성 시에 전처리 된 input 이미지의 사이즈에 따라 적용된 relighting 효과가 다르게 적용되기 때문에 이미지 사이즈를 146 x 70 (512 x 512, relighting 에서)와 296 x 135 (1024 x 1024, relighting 에서) 사이즈로 규격화 하였다. 총 4 가지 model 의 인식기에서 공통적으로 설정한 데이터 셋의 class 는 11 개(0, 1, 2 …, 9, 공백)로 세팅하고 반복 횟수는 300,000 번으로 지정했다. 실험결과는 표 1, 표 2 와 같다. 먼저 512 사이즈의 결과 경우 g.t 모델의 test case 의 정확도는 54%, augmentation 모델은 58%, relighting 모델은 70% 그리고 relighting 에 augmentation 을 결합한 모델은 58%로 relighting 을 적용한 모델들이 적용하지 않은 모델들에 비해 높은 정확도를 확인할 수 있었다. 다음은 1024 사이즈의 결과 경우 g.t 모델의 test case 는 58%, augmentation 모델은 60%, relighting 모델은 70% 그리고 relighting 에 augmentation 을 결합한 모델은 66%로 이 또한 relighting 을 적용한 모델들이 적용하지 않은 모델들에 비해

높은 정확도를 확인할 수 있다. 실험을 통해서 주목할 부분은 relighting 에 augmentation 을 결합한 모델의 결과가 relighting 만 적용시킨 모델보다 다소 성능이 낮았다는 것이다. 이는 relighting 자체에 왜곡이 가해진 상태에서 과도한 augmentation 을 하게 되면 오히려 인식기에서 제대로 학습이 어려운 것으로 판단된다. 또한, 실험에서 대부분의 숫자는 잘 인식하는 결과를 볼 수 있었지만, 소수점이 있는 영상에서 소수점의 위치를 제대로 인식하지 못하는 결과를 확인할 수 있었다.

표 1. 512x512 사이즈의 합성 데이터 셋으로부터 얻은 /crop 된 test case 비교(%)

	Relighting (Proposed method)	Non-Relighting
Augmentation	58	58
Non-Augmentation	70	54(g.t)

표 2. 1024x1024 사이즈의 합성 데이터 셋으로부터 얻은 /crop 된 test case 비교 (%)

	Relighting (Proposed method)	Non-Relighting
Augmentation	66	60
Non-Augmentation	70	58(g.t)

5. 결론

본 논문에서는 relighting 을 적용하여 만들어진 합성 데이터 셋을 학습시켜, 다양한 조명 환경에 강인한 seven-segment 문자 인식 방법을 제안하였다. 이를 위해, seven-segment 문자 데이터 셋에 직접 relighting 을 적용시키지 않고, 자연스러운 배경 이미지에 합성시킨 후 relighting 을 반영하여 실제 조명 환경과 유사한 상황의 seven-segment 문자 데이터 셋 생성하였다. 생성한 데이터 셋으로는 relighting 과 augmentation 각각의 유무에 따라 총 4 가지의 경우 만들어진 모델을 조명 효과들로 문자 인식이 어려운 test case 를 실험한 결과 relighting 효과를 적용시킨 모델들이 적용하지 않은 모델들에 비해 높은 정확도를 확인할 수 있었다. 이후의 연구에서는 실 내외 및 다양한 조명 환경에서 seven-segment 의 인식이 가능케하여 앞으로의 OCR 분야에서 활발히 적용될 수 있을 것이다.

감사의 글

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 ICT R&D 지원을 받아 수행된 연구임.
(No.2021-0-00917, 식물 성장 영상 정보를 이용한 식물공장 피노믹스 시스템)

참고문헌

[1] Zohair Al-Ameen, Faster Deblurring for Digital Images using an Ameliorated Richardson-Lucy Algorithm, IEIE Transactions on Smart Processing and Computing, vol. 7, no. 4, August 2018

[2] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation,

[3] Qi Wu; Maoling Qin; Jingqi Song; Li Liu, An Improved Method of Low Light Image Enhancement Based on Retinex, 2021 6th International Conference on Image, Vision and Computing (ICIVC)

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Nets, CVPR 2017

[5] Li-Wen Wang, Wan-Chi Siu, Zhi-Song Liu, Chu-Tak Li, Daniel P.K. Lun, Deep Relighting Networks for Image Light Source Manipulation, The 2020 European Conference on Computer Vision.

[6] Jeonghun Baek, Geewook Kim, Junyeop Lee, Sungrae Park, Dongyoon Han, Sangdoo Yun, Seong Joon Oh, Hwalsuk Lee. "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis." In The IEEE International Conference on Computer Vision (ICCV), 2019.

[7] Peng, X., & Wang, C. (2020, July). Building super-resolution image generator for OCR accuracy improvement. In International Workshop on Document Analysis Systems (pp. 145-160). Springer, Cham.