Journal of Korea Robotics Society (2021) 16(3):276-282
https://doi.org/10.7746/jkros.2021.16.3.276 ISSN: 1975-6291 / eISSN: 2287-3961

밝기 변화에 강인한 적대적 음영 생성 및 훈련 글자 인식 알고리즘

Adversarial Shade Generation and Training Text Recognition Algorithm that is Robust to Text in Brightness

서 민 석[1] · 김 대 한[1] · 최 동 걸†
Minseok Seo[1], Daehan Kim[1], Dong-Geol Choi†

Abstract: The system for recognizing text in natural scenes has been applied in various industries. However, due to the change in brightness that occurs in nature such as light reflection and shadow, the text recognition performance significantly decreases. To solve this problem, we propose an adversarial shadow generation and training algorithm that is robust to shadow changes. The adversarial shadow generation and training algorithm divides the entire image into a total of 9 grids, and adjusts the brightness with 4 trainable parameters for each grid. Finally, training is conducted in a adversarial relationship between the text recognition model and the shaded image generator. As the training progresses, more and more difficult shaded grid combinations occur. When training with this curriculum-learning attitude, we not only showed a performance improvement of more than 3% in the ICDAR2015 public benchmark dataset, but also confirmed that the performance improved when applied to our's android application text recognition dataset.

Keywords: Text Recognition, Deep Learning, Smart Phone Application

1. 서 론

자연 장면에서 글자를 인식하는 시스템은 다양한 산업에서 응용되고있는 중요한 기술 중 하나이며 학계에 큰 주목을 받고있다[1-7]. 이러한 학계의 주목으로, 다양한 벤치마크 데이터셋들이[8-10] 공개되고 있는 동시에 딥러닝 기반의 글자 인식 성능도 크게 개선되었다[1-4][8,9][11-17]. 하지만 이러한 글자 인식 기술의 발전과 응용에도 불구하고 스마트폰을 활용한 근접촬영에서 발생하는 그림자, 빛 반사로 생기는 밝기 변화가 심한 환경에서는 글자 인식 성능이 유의미하게 저하된다[8][11-13]. 예를 들면 명함 인식[6], 카드 인식[18], 신분증 인식 기술은 정확한 관심 구역에 그림자 없이 글자를 넣어야 인식률이 높고, 자동차 번호판 [19] 인식 시스템은 빛 반사가 발생하면 인식률이 낮아진다[5,7].

딥러닝 기반의 글자 인식 시스템이 현실에 발생할 수 있는 밝기 변화에 약한 근본적인 이유는 딥러닝 모델은 훈련 데이터셋과 테스트 데이터셋 사이의 분포가 다르면 성능이 저하되기 때문이다. 이러한 문제를 해결하기 위하여 ImageNet-C[11], ImageNet-P[11]등 많은 벤치마크 데이터셋과 방법들이 제안되었지만, 글자 인식 기술에서는 제안되지 않았다.

우리는 글자 인식 분야에서 이러한 문제를 직접적으로 해결하기 위하여 적대적 음영 생성 및 훈련 글자 인식 알고리즘을 제안하고, 우리의 알고리즘을 검증하기 위하여 스마트폰 근접 촬영으로 그림자와 빛 반사를 포함하고 있는 우리의 안드로이드 어플리케이션 글자인식 데이터셋을 검증용 데이터셋으로 제안한다.

적대적 음영 생성 및 훈련 글자 인식 알고리즘은 입력 이미지를 9개의 그리드로 나누고 하나의 그리드당 학습가능한 4개의 파라메터로 밝기, 대비, 채도, 색조를 조절한다. 글자 인식기의 훈련과정이 진행될수록 적대적 음영 생성기는 9개의 그리드의 독립적인 밝기, 대비, 채도, 색조의 조절으로 글자 인식

Received : May. 17. 2021; Revised : Jun. 10. 2021; Accepted : Jul. 6. 2021
※ This research was supported by Korea Electric Power Corporation. (Grant number : 202100240001)
1. Master's Student, Department of Information and Communication Engineering, Hanbat National University, Daejeon 34158, Korea (minseok.seo, daehan.kim@edu.hanbat.ac.kr)
† Associate Professor, Corresponding author: Department of Information and Communication Engineering, Hanbat National University, Daejeon 34158, Korea (dgchoi@habat.ac.kr)

기가 인식하기 어려운 이미지를 생성한다. 따라서 처음에는 상대적으로 학습하기 쉬운 이미지를 훈련하다가 점점 어려운 이미지를 훈련하는 커리큘럼 학습을 진행한다. 적대적 음영 생성 및 훈련 글자 인식 알고리즘을 ICDAR2015[8] 공개 벤치마크 데이터셋에서 검증한 결과 3% 이상의 성능 향상을 관찰 하였다.

마지막으로 우리의 알고리즘을 현실 음영이 있는 데이터셋에서 검증하기 위하여 3,000장 규모의 안드로이드 어플리케이션 글자인식 데이터셋을 수집하고, 주석하여 우리의 알고리즘을 검증하였다.

2. 관련 연구

관련 연구 부분에서는 기존 딥러닝 기반의 인식 시스템에서 현실 부패와 왜곡을 극복하는 학습 방법에 대해서 조사하고, 다음으로는 기존 글자 인식 시스템에서 현실 부패와 왜곡를 어떻게 극복하였는지 조사한다.

2.1 현실 부패와 왜곡에 강인한 학습 방법

CNN 기반의 인식 시스템은 사람의 시각 인식 시스템과는 다르게 이미지의 내부에 작은 훼손, 노이즈가 존재하면 그 성능이 저하된다. 이러한 문제를 분석하고 해결하기 위하여 Hendrycks et al.은 ImageNet에 현실에서 발생할수 있는 다양한 부패와 왜곡을 추가한 가상 데이터셋 인 ImageNet-C, ImageNet-P[11]를 제안하였다. 이러한 데이터셋은 단순 작은 부패와 왜곡이 추가 되었을 뿐인데 성능이 평균 50% 이상 감소하는 것을 보였다.

Rusak et al.[12]은 적대적으로 다양한 종류의 노이즈를 인식 모델과 적대적으로 삽입하여 ImageNet-C, ImageNet-P에서 state-of-the-art를 달성하였다. 이러한 실험 결과는 네트워크의 변경 없이, 데이터셋 전처리 기법으로도 현실 부패와 왜곡을 극복할 수 있다는 것을 보여줬다. 우리의 적대적 음영 생성 및 훈련 방법도 Rusak et al.와 같은 데이터 전처리 방법이다. 하지만 Rusak et al.와는 다르게 노이즈를 이미지에 입력하지는 것이 아니라 그리드를 9개로 나누고 4개의 훈련 가능한 파라메터로 이미지를 조합한다. 또한 우리의 방법은 글자 인식 방법에 직접적으로 적용하고 검증 하였기 때문에 적대적 데이터 전처리 방법이 글자 인식 방법에서도 유의미하게 응용될 수 있음을 보였다.

2.2 딥러닝 기반의 글자인식 방법

딥러닝 기술이 발전하면서 글자인식도 크게 진보하였다. 하지만 딥러닝 기반의 글자인식 방법들은 기존의 딥러닝 기반의 인식 기술들과는 다르게 입력 이미지 내부에 포함되어 있는 각 글자의 순서와 의미가 결합되어 하나의 단어, 문장이 되기 때문에 각 글자 사이의 관계를 추론 할 수 있어야한다. 따라서 CNN과 RNN을 결합한 구조가 주류를 이루고 있고 대표적으로는 CRNN[5], RARE[14], R2AM[20], GRCNN, Rosetta[4]와 같은 글자인식 네트워크가 있다.

우리의 작업에서는 기존의 제안되어온 네트워크와 상호보완적인 관계로 같이 사용할 수 있는 적대적 음영 생성 및 훈련 알고리즘을 제안한다.

2.3 현실 왜곡과 부패에 강인한 글자 인식 방법

조명 변화로 발생하는 밝기 변화, 사람의 카메라 움직임으로 생기는 모션블러, 다양한 글자의 크기, 글자의 색상, 글자의 불규칙한 모양 등 현실에서 발생하는 다양한 글자 부패 및 왜곡은 글자 인식을 크게 하락 시킨다[1-4][8][11-14]. Shi et al.은 불규칙한 글자 모양 때문에 글자 인식 성능이 하락하는 문제를 해결하기 위하여 Special Transformer network (STN)[14]를 글자 인식 네트워크에 삽입하여 왜곡된 글자가 입력으로 들어와도 Thin- Plate-Spline 변형으로 자동적으로 글자 인식 네트워크가 인식하기 쉬운 모양으로 변경되게 하여 불규칙한 글자 모양 문제를 해결하였다. 하지만 Special Transformer network는 오직 불규칙한 글자 모양에서의 글자 인식 성능을 목적으로 하였기 때문에 다른 왜곡과 부패에서는 여전히 글자 인식 성능을 개선하지 못했다.

최근 Mout et al.[17]은 이미지 super resolution (SR)을 과 글자 인식 사이의 다중 작업 학습을 통하여 글자 인식 네트워크가 글자 사이즈가 너무 작거나 해상도가 안좋을 때에도 SR도 동시에 학습하였기 때문에 인식 성능을 높일 수 있었다. 하지만 Mout et al.가 제안한 방법은 조명 변화와 같은 부패와 왜곡은 고려하지 못했다.

우리의 적대적 음영 생성 및 훈련 알고리즘은 데이터 전처리 방법 중 하나이기 때문에 Mout et al. 방법과는 다르게 네트워크의 변경이 필요 없고, 그렇기 다른 딥러닝 전처리 방법들과 같이[20] 연산량 및 파라메터 증가가 없다. 또 우리의 방법은 다양한 음영 변화에 강인한 글자 인식기를 목표로 한다.

2.4 글자 인식 데이터셋

글자 인식 네트워크를 훈련시키기 위하여 직접 데이터셋을 구축하고 주석처리 하는 것 은 매우 노동 집약적이다. 예를 들어 일반적인 객체 탐지 데이터셋은 한 객체당 4개의 좌표와, 객체 정보 하나를 입력하면 충분하지만, 글자 인식 데이터셋은 한 단어당 그 단어를 구성하고 있는 글자, 좌표의 정보가 필

(a) Original image (b) Grid-separated image (c) Image with shade change applied

[Fig. 1] An image sample generated by a trained hostile shading generator. (a) When the original image is input (b) After dividing into 9 grids, (c) Applying individual shade changes to each grid to create a sample

요하다. 이러한 이유 때문에 ImageNet 규모의 글자 인식 데이터셋을 구성하는 것은 어렵다.

사람의 노동력을 최소한으로 하고 큰 규모의 데이터셋을 구성하는 가장 대중적인 방법 중 하나는 가상데이터를 생성하고 활용 하는 것이다. 글자 인식에서 대중적으로 가장 많이 사용되는 MJSynth[10]와 SynthText[9] 데이터셋 또한 가상으로 생성된 데이터셋이다. MJSynth은 9만장의 이미지에서 900만 개의 단어가 포함되어 있는 가상 데이터로, 실세계 이미지 배경에 다양한 폰트의 단어를 합성하여 그 글자의 위치와 글자 정보를 포함하고 있는 글자 탐지 및 인식 데이터 셋이다. SynthText 데이터셋은 550만개의 잘라진 단어로 구성되어 있고 오직 글자 인식만을 목적으로 이루어진 데이터셋이다.

따라서 보통의 글자 인식기 훈련에서 가상데이터로 훈련하기 때문에 아무리 인위적으로 부패와 왜곡을 추가 한다고 하여도 실세계 에서 발생하는 다양한 부패와 왜곡을 모두 포함하기 어렵다. 따라서 우리는 사람이 정해서 부패와 왜곡을 추가하는 것이 아닌 적대적 음영 생성기를 통하여 자동으로 글자 인식 네트워크가 어려워하는 샘플을 생성하여 학습한다.

마지막으로 우리의 학습 방법으로 훈련된 글자 인식기가 실세계 밝기 변화에 강인한 것을 확인하기 위하여 다양한 밝기 변화를 포함하고 있는 우리의 새로운 데이터셋인 안드로이드 어플리케이션 데이터셋에서 우리의 훈련 방법의 성능을 평가한다.

3. 연구 방법

연구 방법 부분에서는 적대적 음영 생성 및 훈련 글자 인식 알고리즘에 대하여 설명하고, 글자 인식기 네트워크 구조에 대하여 자세하게 설명한다. 또 우리의 안드로이드 글자 인식 데이터셋의 수집 방법에 대하여 설명하고, 마지막으로는 우리 네트워크의 전체적인 구조에 대하여 자세하게 설명한다.

3.1 적대적 음영 생성 및 훈련 방법

전체의 이미지에 밝기, 대비, 채도, 색조를 조절하여 글자 인식 네트워크를 훈련시키는 것은 글자 인식 네트워크가 다양

(a) Images without min, max constraints (b) Images without average constraints

[Fig. 2] Shading adjustment sample image

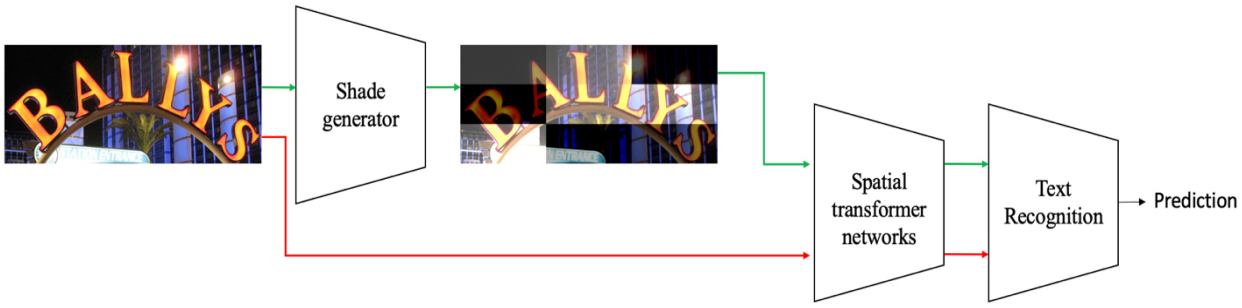
한 음영 변화에 강인해 질 수 있는 가장 간단한 방법 중 하나이다. 하지만 [Fig. 1(a)]에 보이는 것과 같이 대부분의 입력 이미지 내부에서도 음영 변화가 있기 때문에 모든 부분에 일괄적으로 음영을 조절하는 것은 최적이 아니다. 따라서 [Fig. 1(b)]와 같이 입력 이미지가 들어오면 9개의 구역을 나누고 각 구역마다 밝기, 대비, 채도, 색조 개의 요소를 조절한다. 이러한 과정을 통하여 이미지를 생성하면 [Fig. 1(c)]와 같은 이미지를 생성할 수 있다.

하지만 적대적으로 음영을 조절하게 되면, 음영 생성기는 글자 인식 네트워크의 성능을 저하시키게 하는 목적을 가지고 학습하기 때문에 [Fig. 2(a)]와 같이 배경을 흰색 또는 검정색으로 변경하려는 특성을 가지게 된다. 하지만 글자의 형태가 사라지는 것은 우리의 목적이 아니기 때문에 음영을 비율을 최저 0.2, 최대 0.8로 제약을 두어 적대적 음영 생성기를 훈련하였다.

또 음영 비율에만 제약을 두면 각 그리드가 독립적으로 적용되기 [Fig. 2(b)]와 같이 어두운 음영 밝은 음영 한쪽에만 적용 될 수 있다. 우리는 이러한 문제를 해결하기 위하여 각 그리드의 평균 값이 0.6이 되도록 제약을 주었다.

3.2 글자 인식 네트워크

우리의 글자 인식기는 훈련 과정과 테스트 과정이 서로 다르다. 훈련 단계에서는 적대적 음영 생성기가 부착되어 이미지를 점점 더 어려운 샘플로 만드는 과정이 포함되어 있고, 훈련이 완료된 테스트 단계에서는 적대적 음영 생성기가 없다. 따라서 우리의 글자 인식기는 테스트시 연산량 및 파라메터 증가 없이 성능을 향상시킬 수 있다.



[Fig. 3] Shading generation and training text recognition network pipeline. The red line represents the test phase, and the green line represents the training phase

3.3 안드로이드 어플리케이션 글자인식 데이터셋

우리는 안드로이드 어플리케이션 글자인식 데이터셋을 구성하기 위하여 총 9곳의 물류 창고에서 존재하는 물류 상품명, 태그 등을 스마트폰 카메라를 통하여 촬영하였다. 또 우리는 3,000장의 이미지에서 약 6,000개의 단어를 잘라내어 저장하고, 파일명에 해당 단어를 구성하고 있는 글자를 주석하여 적었다.

안드로이드 어플리케이션에서의 글자 인식 응용 상황에서는 사람의 움직임으로 생기는 모션 블러, 사람의 신체적 차이점으로 생기는 뷰 포인트 디스토션, 또 근접샷으로 생기는 사람의 그림자 등이 존재한다. 이러한 안드로이드 상황에서의 특성 때문에 우리가 수집한 데이터셋도 모션 블러, 뷰 포인트 디스토션 마지막으로 그림자가 포함되어 있다.

3.4 음영 생성 및 훈련 글자 인식 네트워크 구조

우리의 적대적 음영 생성 및 훈련 글자 인식 네트워크의 전체적인 구조는 [Fig. 3]과 같다. 그림에 초록색 선에 보이는 것과 같이 훈련 시 입력 이미지가 들어오면 적대적 음영 생성기에서 9개의 그리드에 음영을 조절하고 음영이 조절된 이미지는 다시 STN에 삽입되어 공간적 왜곡을 보정한 후 글자 인식 네트워크에 삽입되고 최종 예측 한다. 테스트 시에는 [Fig. 3]의 빨간선을 따라 음영 생성기를 거치지 않고 바로 STN에 삽입되어 공간적 왜곡을 보정한 후 글자 인식 네트워크에서 최종으로 예측한다.

4. 실험 결과

실험결과 부분에서는 우리의 실험의 자세한 환경 구성에 대하여 설명하고, ICDAR2015 데이터셋에서 우리의 알고리즘을 검증 후 , 안드로이드 어플리케이션 데이터셋에서 추가적으로 성능을 검증한다.

4.1 자세한 실험 환경 구성

우리는 글자 인식기를 학습시키기 위하여 MJSynth와 SynthText 데이터셋에서 훈련 하였다. 또 Adam 옵티마이저를 사용하였고 초기 러닝 레이트를 1.0으로 설정하였다. 배치사이즈는 192로 설정하였으며, 300,000 이터레이션을 학습하였다. 또 모든 입력 이미지를 높이 33, 너비 101로 재조정하였다.

4.2 ICDAR2015 데이터셋에서의 검증

우리는 우리의 음영 생성 및 훈련 글자 인식 방법의 효과를 검증하기 위해서 가장 대표적인 실제 현실 데이터인 ICDAR2015 데이터셋에서 검증하였다. ICDAR 2015 데이터셋은 ICDAR 2015 Robust Reading 대회를 위해 제작되었으며 교육용 이미지 4,468개와 평가 용 이미지 2,077개가 포함되어 있다. 이미지는 구글 안경을 통하여 착용자의 자연스러운 움직임 아래에서 추출되었다. 따라서 대부분은 모션 블러가 적용되어 있고, 일부는 해상도가 매우 낮다.

또 이런 데이터셋을 연구자들은 평가를 위해 두개의 데이터셋으로 나누어 사용하였는데 1,811개의 버전과 2,077개의 버전 이다. 1,811개의 이미지 버전은 매우 저 해상도 이미지나, 왜곡이 심한 이미지를 제거한 버전이고, 2,077은 저해상도와 왜곡이 심한 이미지를 포함한 버전이다.

첫 번째로 우리의 알고리즘을 글자 인식에서 가장 대표적으로 자주 사용되는 STAR-Net[16] 구조에서 검증 하였다. [Table 1]은 기본 STAR-Net과 우리의 알고리즘을 적용한 STAR-Net의 실험 결과이다.

표에 보이는 것과 같이 1,811, 2,077 각 0.2%, 3.7% 성능이 개선된 것을 확인할 수 있다. 1,811 데이터셋에서는 성능 향상폭이 적은 이유는 이미 데이터셋에 다수의 왜곡과 부패 이미지들이 이미 제거가 되어있기 때문이다. 하지만 2,077 데이터셋에서는 3.7%의 성능향상이 있는 것을 보아 우리의 알고리즘이 부패와 왜곡에 효과가 있다는 것을 확인할 수 있었다.

[Table 1] Results of our algorithm verification on STAR-Net

Method	Dataset	Accuracy
STAR-Net	ICDAR2015/1,811	76.1%
STAR-Net+Our's	ICDAR2015/1,811	**76.3%**
STAR-Net	ICDAR2015/2,077	70.3%
STAR-Net+Our's	ICDAR2015/2,077	**74.0%**

[Table 2] Results of verifying our algorithm of various text recognition networks

Method	Dataset	Accuracy
CRNN	ICDAR 2015/1,811	69.4%
RARE	ICDAR 2015/1,811	74.5%
R2AM	ICDAR 2015/1,811	68.9%
GRCNN	ICDAR 2015/1,811	71.4%
Rosetta	ICDAR 2015/1,811	71.2%
CRNN+Our's	ICDAR 2015/1,811	69.1%
RARE+Our's	ICDAR 2015/1,811	**74.9%**
R2AM+Our's	ICDAR 2015/1,811	68.2%
GRCNN+Our's	ICDAR 2015/1,811	70.3%
Rosetta+Our's	ICDAR 2015/1,811	70.9%
CRNN	ICDAR 2015/2,077	64.2%
RARE	ICDAR 2015/2,077	68.9%
R2AM	ICDAR 2015/2,077	63.6%
CRCNN	ICDAR 2015/2,077	65.8%
Rosetta	ICDAR 2015/2,077	66.0%
CRNN+Our's	ICDAR 2015/2,077	**67.1%**
RARE+Our's	ICDAR 2015/2,077	**70.9%**
R2AM+Our's	ICDAR 2015/2,077	**66.1%**
GRCNN+Our's	ICDAR 2015/2,077	**67.2%**
Rosetta+Our's	ICDAR 2015/2,077	**69.0%**

두 번째 실험으로는 다양한 글자 인식 네트워크에서 우리의 알고리즘이 꾸준한 성능이 있는지 확인하기 위하여 실험을 설계하고 성능을 측정하였다.

[Table 2]에 보이는 것과 같이 우리의 방법은 ICDAR 2015/2,077 데이터셋에서 꾸준한 성능 향상을 보였다. 하지만 ICDAR 2015/1,811에서는 RARE[14]구조 제외하고는 성능이 조금씩 하락하였다. 그 이유는 ICDAR2015/1811에서는 왜곡과 부패가 있는 데이터가 삭제되어 있기 때문이라고 분석하였다.

마지막으로 우리의 적대적 음영 생성 및 훈련 알고리즘의 밝기, 대비, 채도, 색조의 제한 등과 같은 방법들이 성능에 미치는 영향에 대해서 분석하였다. [Table 3]에 보이는 것과 같이 밝기, 대비, 채도, 색조의 최대값, 최소값 제한을 두지 않으면 성능이 2.1%, 1.9%로 학습이 되지 않는 것을 확인할 수 있었고

[Table 3] Results of verifying our algorithm of various text recognition networks

Method	Dataset	Accuracy
STAR-Net+Aug	ICDAR2015/1,811	75.9%
STAR-Net+Our's (w/o)	ICDAR2015/1,811	2.1%
STAR-Net+Our's (w/)	ICDAR2015/1,811	**76.3%**
STAR-Net+Aug	ICDAR2015/2,077	72.3%
STAR-Net+Our's (w/o)	ICDAR2015/2,077	1.9%
STAR-Net+Our's (w/)	ICDAR2015/2,077	**74.0%**

[Table 4] Results of verifying our algorithm of various text recognition networks

Method	Dataset	Accuracy
STAR-Net	Our's	88.1%
STAR-Net+Our's	Our's	**93.3%**

단순 밝기, 대비, 채도, 색조의 랜덤 적용으로는(augmentation) 2%의 성능 향상이 있는 것을 확인할 수 있었다. 이러한 실험 결과 우리는 단순하게 데이터에 brightness, contrast, saturation, hue의 랜덤 적용보다 적대적 음영 생성 및 훈련 방법 성능이 1.7% 높은 것으로 보아 적대적 음영 생성 및 훈련 방법이 효과가 있다고 말할 수 있다.

4.3 안드로이드 어플리케이션 데이터셋에서의 검증

현실 음영 변화를 다수 포함하고 있는 우리의 안드로이드 어플리케이션 데이터셋에서 우리의 적대적 음영 생성 및 훈련 방법을 검증하였다. [Table 4]에 보이는 것과 같이 우리의 알고리즘을 활용 하였을 때 베이스라인 성능보다 5.2% 향상된 성능을 볼 수 있었다. 이 결과 우리의 직관처럼 우리의 방법은 특히 음영 변화가 다수 포함된 안드로이드 어플리케이션 데이터셋에서 특히 효과가 있음을 보았다.

5. 결 론

우리는 글자 인식 시스템이 현실에서 발생할 수 있는 밝기 변화에 취약하다는 문제점을 발견하고, 이를 해결하기 위하여, 밝기 변화에 강인한 적대적 음영 변화 및 훈련 방법을 제안하였다. 또 우리의 알고리즘을 검증하기 위하여 ICDAR2015 데이터셋에서 실험하였으며, 다양한 글자 인식 네트워크 구조에서 우리의 알고리즘이 효과가 있음을 보였다. 마지막으로 우리의 알고리즘을 추가적으로 검증하기 위해서 안드로이드 어플리케이션을 통하여 수집한 안드로이드 어플리케이션 글자 인식 데이터셋에서 우리의 알고리즘을 검증하고 우리의 알

고리즘이 실제 데이터셋에서도 유효함을 실험적으로 보였다. 우리는 우리의 알고리즘이 자동차 번호판 인식기, 카드 번호 인식기와 같은 실제 응용에서 활용되어 기존 인식기에서 인식이 불가능 했던 부분을 해결하길 바란다.

References

[1] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee, "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 2019, DOI: 10.1109/iccv.2019.00481.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998, DOI: 10.1109/5.726791.

[3] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: An Attentional Scene Text Recognizer with Flexible Rectification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2035-2048, Sep., 2019, DOI: 10.1109/tpami.2018.2848939.

[4] F. F. Borisyuk, A. Gordo, and V. Sivakumar, "Rosetta: Large Scale System for Text Detection and Recognition in Images," *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, DOI: 10.1145/3219819.3219861.

[5] SHI, Baoguang; BAI, Xiang; YAO, Cong. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence, 2016, 39.11: 2298-2304, DOI: 10.1109/TPAMI.2016.2646371.

[6] J.-H. Kim and J. Lim, "License Plate Detection and Recognition Algorithm using Deep Learning," *Journal of IKEEE*, vol. 23, no. 2, pp. 642-651, Jun., 2019, DOI: 10.7471/IKEEE.2019.23.2.642.

[7] M. Seo, S. Lee, and D.-G. Choi, "Spatial-temporal Ensemble Method for Action Recognition," *Journal of Korea Robotics Society*, vol. 15, no. 4, pp. 385-391, Dec., 2020, DOI: 10.7746/jkros.2020.15.4.385.

[8] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, F. Shafait, S. Uchida, and E. Valveny, "ICDAR 2015 competition on Robust Reading," *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Tunis, Tunisia, 2015, DOI: 10.1109/icdar.2015.7333942.

[9] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic Data for Text Localisation in Natural Images," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, DOI: 10.1109/cvpr.2016.254.

[10] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," *NIPS DLW*, 2014, [Online], https://arxiv.org/pdf/1406.2227.pdf.

[11] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *International Conference on Learning Representations (ICLR)*, 2019, [Online], https://arxiv.org/pdf/1903.12261.pdf.

[12] E. Rusak, L. Schott, R. S. Zimmermann, J. Bitterwolf, O. Bringmann, M. Bethge, and W. Brendel, "A Simple Way to Make Neural Networks Robust Against Diverse Image Corruptions," *Lecture Notes in Computer Science*, pp. 53-69, 2020, DOI: 10.1007/978-3-030-58580-8_4.

[13] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," *2011 International Conference on Computer Vision*, Barcelona, Spain, 2011, DOI: 10.1109/iccv.2011.6126402.

[14] B. Shi, X. Wang, P. Lyu, C. Yao, and X. Bai, "Robust Scene Text Recognition with Automatic Rectification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, DOI: 10.1109/cvpr.2016.452.

[15] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *NIPS*, 2015, [Online], https://proceedings.neurips.cc/paper/2015/file/33ceb07bf4eeb3da587e268d663aba1a-Paper.pdf.

[16] W. Liu, C. Chen, K. Wong, Z. Su, and J. Han, "STAR-Net: A SpaTial Attention Residue Network for Scene Text Recognition," *British Machine Vision Conference 2016*, 2016, DOI: 10.5244/c.30.43.

[17] Y. Mou, L. Tan, H. Yang, J. Chen, L. Liu, P. Yan, and Y. Huang, "PlugNet: Degradation Aware Scene Text Recognition Supervised by a Pluggable Super-Resolution Unit," *European Conference on Computer Vision*, pp. 158-174, 2020, DOI: 10.1007/978-3-030-58555-6_10.

[18] J.-H. Kim, "Automatic Recognition of Bank Security Card Using Smart Phone," *The Journal of the Korea Contents Association*, vol. 16, no. 12, pp. 19-26, Dec. 2016, DOI: 10.5392/JKCA.2016.16.12.019.

[19] S. Lee and G. Park, "Proposal for License Plate Recognition Using Synthetic Data and Vehicle Type Recognition System," *Journal of Broadcast Engineering*, vol. 25, no. 5, pp. 776-788, Sep., 2020, DOI: 10.5909/JBE.2020.25.5.776.

[20] C.-Y. Lee and S. Osindero, "Recursive recurrent nets with attention modeling for ocr in the wild," *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2231-2239, DOI: 10.1109/CVPR.2016.245.

서 민 석

2019 한밭대학교 정보통신공학과(학사)

2020~현재 한밭대학교 정보통신공학과 (석사)

관심분야: Robotics, Vision Programing, Deep Learning

김 대 한

2020 한밭대학교 정보통신공학과(학사)

2021~현재 한밭대학교 정보통신공학과 (석사)

관심분야: Robotics, Vision Programing, Deep Learning

최 동 걸

2005 한양대학교 전자컴퓨터공학부(학사)

2007 한양대학교 전자전기제어계측공학과 (석사)

2016 KAIST 로봇공학학제전공(박사)

2018 KAIST 정보전자연구소 박사 후 연구원

2018~현재 한밭대학교 정보통신공학과 조교수

관심분야: Robot Vision, Sensor Fusion, Autonomous Robot System, Artificial Intelligence