

## 선명화 기법을 이용한 TextFuseNet 성능 향상

정지연\*, 천지은<sup>o</sup>, 정유철\*

\*금오공과대학교 컴퓨터공학과,

<sup>o</sup>금오공과대학교 컴퓨터공학과

e-mail: {jiyeon2781\*, wldms01125<sup>o</sup>, jyc\*}@kumoh.ac.kr

## Performance Improvement of TextFuseNet using Image Sharpening

Ji-Yeon Jeong\*, Ji-Eun Cheon<sup>o</sup>, Yuchul Jung\*

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

<sup>o</sup>Dept. of Computer Engineering, Kumoh National Institute of Technology

### ● 요약 ●

본 논문에서는 Scene Text Detection의 새로운 프레임워크인 TextFuseNet에 영상처리 관련 기술인 선명화 기법을 제안한다. Scene Text Detection은 야외 간판이나 표지판 등 불특정 배경에서 글자를 인식하는 기술이며, 그중 하나의 프레임워크가 TextFuseNet이다. TextFuseNet은 문자, 단어, 전역 기준으로 텍스트를 감지하는데, 여기서는 영상처리의 기술인 선명화 기법을 적용하여 TextFuseNet의 성능을 향상시키는 것이 목적이다. 선명화 기법은 기존 Sharpening Filter 방법과 Unsharp Masking 방법을 사용하였고 이 중 Sharpening Filter 방법을 적용하였을 때 AP가 0.9% 향상되었음을 확인하였다.

**키워드:** 이미지 선명화(Image Sharpening), 장면 텍스트 검출(Scene Text Detection)

## I. Introduction

최근 컴퓨터 비전의 기술적 발전으로 OCR 기능이 크게 향상됐다. 이에 따라 일상 환경에서 많이 접할 수 있는 Scene Text도 주목을 받고 있다. 문서 이미지 속 글자들은 정형화되어 있지만 보통 이미지 글자들은 비정형화되어 있어 분별이 어렵다. 이러한 비정형화된 글자를 검출하기 위해 Scene Text Detection 기술을 사용한다.

Scene Text Detection이란 야외 간판이나 표지판 등 불특정 배경에서 글자를 인식하는 기술이다. 이 중 하나의 프레임워크인 TextFuseNet에 선명화 기법 두 가지를 적용하여 확인해보았다.

먼저 Sharpening Filter를 사용하였을 때, 단어의 전반적인 성능은 기존 TextFuseNet에 비해 0.5% 정도 향상되었음을 알 수 있다. 알파벳 및 숫자에서도 전체적으로 기존 TextFuseNet에 비해 0.1%에서 10%까지 성능이 향상되었음을 확인할 수 있지만, 약 14개의 알파벳, 숫자에서 0.1%에서 5%까지 성능이 하락하였음을 확인할 수 있다.

다음 Unsharp Masking을 사용하였을 때, 단어의 전반적인 성능은 기존 TextFuseNet에 비해 0.5% 정도 하락하였다. 알파벳 및 숫자에서도 3분의 2 정도는 0.1%에서 5%까지 성능이 하락하였고 나머지는 0.1%에서 10%까지 성능이 향상되었다.

## II. Preliminaries

### 1. Related works

Scene Text Detection 관련 논문은 현재에도 계속 나오고 있으며, 문자를 추출하여 사용하는 시스템에서도 활용되고 있다. Benchmark를 확인해보면 여러 논문이 존재하는데, 그중 3가지를 선택해 비교해보았다.

Table 1. System Environment

	MaskText Spotter	TextFuse Net	SPCNet
Framework	PyTorch	PyTorch	Tensorflow
Power	Mask R-CNN	Detectron2	X
Independence	GRU를 통한 문자 예측	multi-path fusion 생성	단순 Layers

Benchmark의 순위를 확인해보면 TextFuseNet이 1위, SPCNet이 2위, MaskTextSpotter가 3위를 기록하였다. 이는 ICDAR2013 Dataset을 기준으로 Benchmark를 확인하였는데, 이 Dataset 뿐만 아니라 다른 Dataset에서도 논문이 계속해서 업데이트되고 있다. 또한, 본 연구의 성능을 개선시키기 위해 많은 연구가 진행되고 있다.

### III. The Proposed Scheme

본 논문에서는 TextFuseNet 성능 향상을 위해 이미지 전처리 단계에서 선명화 기법을 제안한다.

#### 1. Image Sharpening

##### 1.1 Sharpening Filter

Fig. 1에서는 OpenCV의 5x5 Sharpening Filter를 적용하여 cv2.filter2D 함수를 통해 적용시키는 방법이다. Fig. 2는 원본 이미지인 Input Image에 Sharpening Filter를 적용하였을 때 Output Image가 나온다.

$$M = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & 2 & 2 & 2 & -1 \\ -1 & 2 & 9 & 2 & -1 \\ -1 & 2 & 2 & 2 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

Fig. 1. Sharpening Filter



Fig. 2. Sharpening Filter Result

##### 1.2 Unsharp Masking

Unsharp Masking은 이미지 선명화의 기술 중 하나이며 해당 기술을 거치면 채도와 명도가 전체적으로 골고루 퍼지게 된다. Unsharp Masking의 과정은 Fig. 3에서 확인할 수 있다. 간단하게 설명하면 원본 이미지를 복제한 후 입력 이미지에 대한 Gaussian Filtering을 수행한다. 그리고 원본 이미지와 Gaussian Filtering을 수행한 이미지를 적절한 비율로 더하는 방법이다.

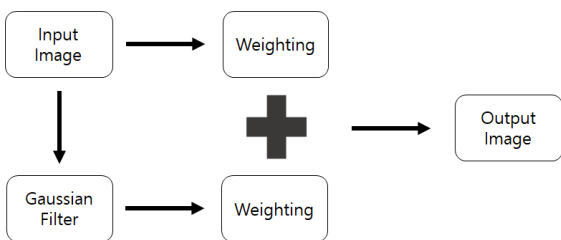


Fig. 3. Unsharp Masking Process

위의 과정을 거치면 이미지가 선명해지고 채도의 대비가 높아진다. Fig. 4는 원본 이미지인 Input Image에 Unsharp Masking을 적용하

였을 때 Output Image가 나온다.



Fig. 4. Unsharp Masking Process

### IV. Experiments

TextFuseNet 프레임워크에 SynthText Dataset으로 사전훈련된 모델을 바탕으로 ICDAR2013 Dataset[5]을 통해 Batch Size는 1로 설정하여 20K Iteration으로 미세조정(fine-tuning)을 진행했다.

ICDAR2013 Dataset은 Robust Reading Competition에서 제공된 데이터이다. Train 데이터는 229개이며 Test 데이터는 233개이다. 각 실험 결과는 이미지 전처리 부분에서 선명화 기법을 적용하지 않은 기존 TextFuseNet 프레임워크를 사용하였을 때와 각각의 선명화 기법을 적용하고 TextFuseNet 프레임워크를 사용하였을 때로 나뉜다.

#### 1. Unusing Image Sharpening

선명화 기법을 적용하지 않고 사전 훈련된 모델을 TextFuseNet 프레임워크를 이용하여 ICDAR2013 데이터를 대상으로 fine-tuning을 하였을 때의 결과이다.

Table 2. Unusing Image Sharpening - Bbox

AP	AP50	AP75	APs	APm	API
81.4	90.2	89.9	70.6	86.6	93.2

Table 3. Unusing Image Sharpening - Segmentation

AP	AP50	AP75	APs	APm	API
84.1	90.3	90.1	75.2	86.2	94.3

#### 2. Sharpening Filter

5x5 Sharpening Filter를 적용해서 이미지 전처리를 진행하고 사전 훈련된 모델을 TextFuseNet 프레임워크를 이용하여 ICDAR2013 데이터를 대상으로 fine-tuning을 하였을 때의 결과이다.

Table 4. Sharpening Filter - Bbox

AP	AP50	AP75	APs	APm	API
82.3	90.0	89.8	71.4	87.9	93.8

Table 5. Sharpening Filter – Segmentation

AP	AP50	AP75	APs	APm	API
84.2	90.3	89.8	74.8	88.9	94.3

### 3. Unsharp Masking

Unsharp Masking 기술을 적용해서 이미지 전처리를 진행하고 사전 훈련된 모델을 TextFuseNet 프레임워크를 이용하여 ICDAR2013 데이터를 대상으로 fine-tuning을 하였을 때의 결과이다.

Table 6. Unsharp Masking – Bbox

AP	AP50	AP75	APs	APm	API
80.1	90.1	89.3	69.1	85.5	92.0

Table 7. Unsharp Masking – Segmentation

AP	AP50	AP75	APs	APm	API
81.7	90.2	89.5	71.8	86.1	92.2

### 4. Detection Result

Fig. 5는 (a), (b), (c), (d) 각각 원본 이미지, 기존 TextFuseNet 프레임워크에서 제공된 모델, Sharpening Filter를 적용하여 120K Iteration으로 학습시킨 모델, Unsharp Masking을 적용하여 120K Iteration으로 학습시킨 모델의 Detection Result를 나타낸다.

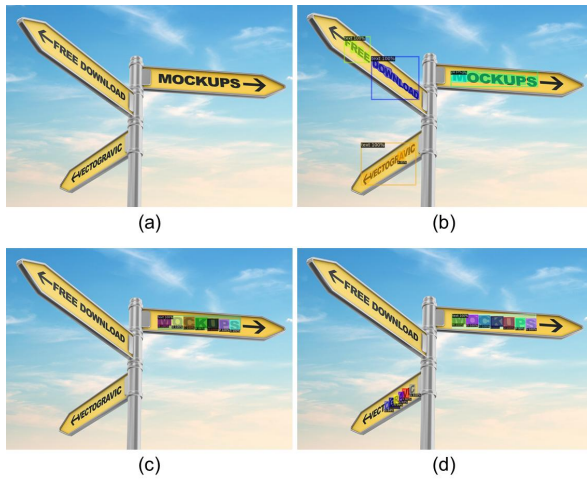


Fig. 5. Detection Result

(b)는 전체적인 단어는 잘 추출하지만, 문자는 잘 추출하지 못하였다. (c)는 (b)에 비해 문자는 잘 추출하지만, 굴곡진 문자는 잘 추출하지 못하였다. (d)는 (c)보다 굴곡진 문자를 조금 더 잘 추출하였다. 수치상으로 비교하자면 Bbox 기준 (c)는 (b)에 비해 AP가 0.9% 상승하였지만 (d)는 (b)에 비해 AP가 1.3% 하락하였다.

## V. Conclusions

Scene Text Detection의 프레임워크인 TextFuseNet에 Sharpening Filter, Unsharp Masking 등 두 가지의 선명화 기법을 적용하여 성능을 확인해보았다. 수치상으로 확인하였을 때 선명화 기법을 적용한 TextFuseNet 프레임워크는 성능이 그렇게 높지 않아 잘 추출하지 못할 것으로 생각했지만, 이미지를 입력해 Detection을 진행해보니 전반적으로 문자는 잘 추출되었다. 하지만 굴곡진 단어는 잘 추출하지 못하였다. 선명화 기법을 적용하면 문자와 배경 사이 채도의 대비가 더 극대화되어 Detection에서는 글자를 더 잘 찾을 수 있음을 확인할 수 있다. 하지만 앞서 말했듯이 굴곡진 단어는 전혀 추출하고 있지 못하므로 이를 보완하기 위해 추후 굴곡진 Dataset을 추가해 학습을 진행하여 성능 향상을 계획하고 있다.

## REFERENCES

- [1] Minghui Liao, Baoguang Shi, Xiang Bai, Xinggang Wang, Wenyu Liu, "TextBoxes: A Fast Text Detector with a Single Deep Neural Network", AAAI 2017, 21 Nov 2016.
- [2] Jian Ye, Zhe Chen, Juhua Liu, Bo Du, "TextFuseNet: Scene Text Detection with Richer Fused Features" IJCAI-20, 2020.
- [3] Minghui Liao, Pengyuan Lyu, Minghang He, Cong Yao, Wenhao Wu, Xiang Bai, "Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes" IEEE, August 2019.
- [4] Enze Xie, Yuhang Zang, Shuai Shao, Gang Yu, Cong Yao, Guangyao Li, "Scene Text Detection with Supervised Pyramid Context Network", AAAI 2019, 21 Nov 2018.
- [5] Robust Reading Competition ICDAR2013 Dataset, <https://rrc.cvc.uab.es/?ch=2&com=downloads>