Optimizing Long Distance Character Recognition Using Retro-Reflective License-Plating

Marshall Hageman
University of Maryland,
Baltimore County
marsh9@umbc.edu

David A. Opie
University of Maryland,
Baltimore County
daopie1@umbc.edu

I. Abstract

***Abstract*—There are hazardous wildlife effects caused by speeding watercraft. There is no current solution to remotely watch and enforce speed and area regulations without an officer on the scene. Speed and area controls require a system capable of functioning accurately during night and daytime from long distances while remaining inexpensive. However, there is no methodology to enforcing these protection zones other than with a police man or woman patrolling every zone. To help protect our shores we developed the Retro-Reflective Vessel Identification System (RRVIS). RRVIS is a license plate detection system capable of working day or night in order to incentives the protection of the wildlife. We simulated the license plate detection system under low light conditions and we extrapolated data from our simulation using machine learning OCR modules.**

II. Introduction

Retro-reflective vessel identification system
Marine and vegetative wildlife are under stress from several factors, some of which are anthropomorphic or natural. Natural stresses such as predation are not new, but are amplified by the anthropomorphic stresses. Anthropomorphic stresses include global warming, pollution, over-fishing, and shoreline erosion. Together, these stresses result in an alarming decline in the population and health of the littoral and riparian plants and animals (need reference here). In this document we describe a means for reducing one of these stresses: shoreline erosion caused by the wakes from vessels. Commercial and recreational boating contributes to shoreline erosion caused by the wakes of vessels. Vessels moving at speeds near or greater than the hull speed imparts high-energy wakes along the shorelines that would otherwise not experience the disturbance caused by these wakes (reference). The detrimental wakes from boating typically impacts shoreline areas that would otherwise only experience low-energy waves. Reducing boat speed in these areas will reduce this source of anthropomorphic stress. While there are posted boat speed limits in some sensitive areas, there is no enforcement to ensure compliance with the speed limits. The introduction of a camera-based detection system will influence peoples behavior and reduce shoreline erosion and wildlife hazards on waterways identified as at-risk.

Marine and Vegetative wildlife have been seen a huge decline in health due to global warming, pollution, over-fishing, and erosion largely caused by wakes from watercraft. Boat wakes have been shown to have erosive effects on shorelines, vegetative wildlife, and disruption of faunal communities. This is easily seen with Manatees. Every year on average 91 manatees die from boat related deaths. Slower boat speeds have been proven to reduce risk and injury to manatee[1]. This has led states like Florida to create manatee protection zones that restrict the speed and operation of vessels in designated areas. These zones are made to create safe protected areas that manatees can live without boat interference.

The effects of boat wakes have been shown to have erosive effects on shorelines, vegetative wildlife, and disruption of faunal communities. However, there is no methodology to enforcing these protection zones other than with a police man or woman patrolling every zone. To create a cost and time efficient solution we incentives safe boating practices with our retro-reflective vessel identification system (RRVIS) on the water.

The use of image capture with low-power cameras and character recognition systems can be used to read watercraft identification numbers. The system, the Retro-reflective Vessel Identification System (RRVIS), uses image capture and optical character recognition to identify vessels that pass the camera. Once the image is captured, the information of the watercraft location and time can be transmitted for analysis. Being installed on buoys and pilings necessitates that cameras are powered with batteries that are charged with solar panels and that the cameras operate within these low-power budgetary limits. Another requirement of the RRVIS is that the watercraft identification numbers use a retro-reflective number and number background. A further requirement of the RRVIS is that the vessel license plates or at least the vessel license plate numbers are a specific size. Vessels speed can be detected with radar techniques or using the two-camera and

University of Maryland Baltimore County
Undergraduate Studies, Computer Science
491 Wireless Cities

average speed algorithms. Radar detectors have an additional power drain on the RRVIS system. An alternative method of watercraft speed measurement is the calculate the average speed between two cameras that are separated by a known distance. As the distance between the two cameras is known, the average speed can be calculated by dividing this by the time taken to travel between two cameras. The calculation of speed can be done at the central computer that is receiving and processing the camera data. As opposed to a normal automotive speed camera, the RRVIS relies on retro-reflective vessel number license plates and infrared cameras to make a low powered system capable of increased accuracy and distance as compared to traditional license plate cameras. The system must be capable of operating under low light levels and with low power consumption all while functioning with only a widely available and affordable camera and lighting system. The use of a retro reflective material increases optical character recognition at range in low light and low visibility environments, thereby allowing the system to capture vessel license plate data at greater distances with the necessary accuracy. Challenges that exist with the implementation of the RRVIS include lighting conditions and distance between camera and vessel license plate. Lighting conditions include low-light conditions, bright-light conditions, reflection of light from the water surface and water vapor. For the camera in the RRVIS system, Infrared (IR) cameras provide an off-the-shelf solution. IR or Infrared wavelengths, measured in nanometers (nm), describes the wavelength of electromagnetic radiation near used by the camera. The IR part of the light spectrum is outside of the visible part of the light spectrum. Visible light ranges from about 390 nm to 700 nm. As such, these wavelengths recorded by IR cameras fall outside of the visible range. IR cameras that operate at wavelengths longer than 700 nm are common in closed circuit television (CCTV) and security applications. While cameras that operate at the IR wavelengths have similar performance, there are a few important differences. The challenge of low-light conditions can be resolved by using an illuminator (camera flash) to illuminate the vessel license plate. The wavelength of the illuminator needs to match the camera wavelength. The common IR wavelengths produced by diode (LED) IR illuminators are 850 nm and 940 nm. However, water vapor in the marine environment will interfere with the IR light reaching the target vessel license plate and the return of the reflected light imaged by the camera. The graph shown in Figure 1 demonstrates absorbance of light by water vapor as a function of wavelength. From this graph, it is clear that the 940 nm illuminators and cameras will not have the ability to image vessel license plates at a long distance from the camera.

Distance calculation The distance between the vessel license plate and the camera can be determined with the triangle similarity method. With this method, the pixel size of the object is related to the focal length of the camera. For testing, and before deploying the camera to the field, the vessel license plate with width W is placed a known distance D from the camera. We take a picture of our object using our camera and

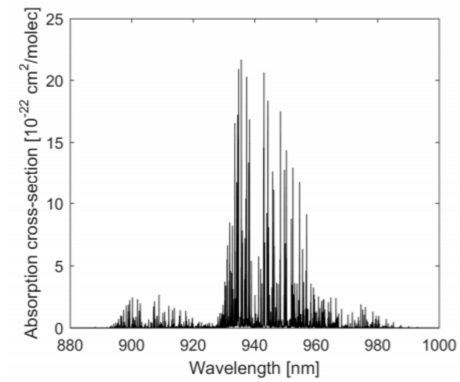

Fig. 1. Absorption of water in the IR band between 880 nm and 1000 nm. This graph shows that 940 nm IR cameras and illumination LEDs will be challenged by water vapor absorption, while 850 nm cameras and illuminators will not be challenged.

then measure the apparent width in pixels P. This allows us to derive the perceived focal length F of our camera: F = (P x D) / W For example, let's say I place a vessel license plate (horizontally; W = 2.5 feet) D = 50 feet in front of the camera and record the image. If the width W is 85 pixels, then the apparent focal length is F = (85 px x 50 ft.) / 2.5 = 1700 When the image of the vessel license plate found to have a image width of 25 pixels, then the distance to the vessel license plate is: D' = (W x F) / P = (2.5 ft. x 1700)/25 px = 170 ft. Therefore, the distance from the camera to the vessel license plate can be calculated.

III. Implementation

Our system is designed to be placed on a marker ("traffic signals" used for guiding vessel operators attached to posts or buoys). There is a motion detection IR camera and a radio to send the images to a server off-site. This low-powered system would be solar powered and only turn on when the motion detection camera is set off. Thus making it the perfect low maintenance sensor. On longer stretches of water it would be possible to use multiple sensors to capture more angles and distance to make sure that one of the cameras is in position to collect the license data the cameras currently set up to take 3 consecutive pictures when detecting movement with an offset of 3 seconds however this can be adjusted to capture video or single frames with greater or smaller offsets depending on the sensors position. On the software side, our current implementation is set up to receive an image and using python runs an EAST text detection CNN model to determine where the text is located (see figure 2). Once the detector has discovered an roi we then preprocess the region.

Preprocessing- The text becomes heavily distorted at range so it is necessary to preprocess these images effectively. Given more time, we had plans to next take a look at feature point algorithms or neural networks to straighten distorted text and further increase accuracy. The preprocessing method currently in use is a 3 step process (refer to figure 1). First the image is

Fig. 2. Preprocessing performed on our license plate before OCR

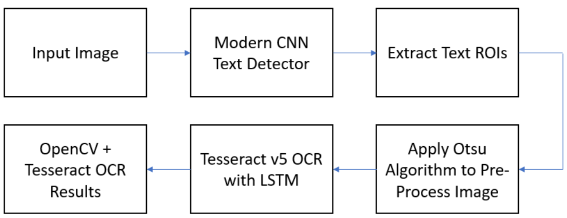

Fig. 3. General overview of our method for processing infrared photos and extracting region of interest to perform optical character recognition.

scaled to the correct size, then the image is inverted. Since our license plate reflects light only on the letters (The backing is designed as any infrared absorptive material coated black) we invert the colors so that the license resembles traditional text. From there we apply Otsu's method to do automatic image thresholding that seperates the pixels into two classes. Otsu's method is an algorithm that works by iteratively searching for the appropriate threshold to minimize intra-class variance.

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)$$

In Otsu's algorithm, the sigmas squared act as the variances for the weights 1 and 2 with a threshold of t. It works by computing a histogram with probabilities at every level of intensity then iterating through every threshold from 1 to the maximum intensity while updating the two weights and computing variance as it goes through. The desired weights have the maximum amount of variance.

After pre-processing the image we receive our extracted region of interest. We then need to denoise the image. To denoise the image we use Tesseract version 5 optical character recognition algorithm with the LSTM neural network. The Tesseract OCR engine is a robust optical character engine which includes a built-in deep learning machine. Tesseract processes the image and creates a list of characters it believes to have read.

A. Data Collection

As there is no data repository with retro reflective lettering publicly available for use it was necessary to capture imagery. When we started collecting data we initially used cameras with bright LED's behind visible light cameras. We were then able to test the lighting on three separate intensities to collect more data with different amounts of glare. We used a traditional low powered 12 MP camera in addition to a more powerful camera with a lens for focusing at range and 16 Mega Pixels. After conducting tests at measured intervals up to 60 ft. We then conducted tests on the water to ensure reflections would not have an effect on the back scatter of light affecting picture quality. After doing more research, we decided to purchase a cheap infrared camera for under 40 dollars capable of taking pictures using an infrared lens. The low budget 12 MP infrared camera features an infrared bulb array as well as a motion detector for capturing footage as it was originally intended for the purpose of monitoring wildlife (trail camera). We chose the camera not just for its budget friendly cost (which could be a major factor in its future deployment for restricted areas and speed detection). The camera features a built in full automatic IR filter for 1080P video during day (color) and night (black and white) so it would be usable when deployed during the daytime as well. Not only does Infrared allow more light to enter the camera lens making it ideal for night time but the bulbs are incredibly energy efficient and the light carries farther than normal visible light. In addition the illumination is able to carry farther allowing the license to be more visible. Since infrared wavelengths are not on the visible spectrum, the light is invisible to the human eye and would not distract or alert boaters. This makes it an ideal technology for sensor systems. Especially in high humidity areas such as alongside rivers and bodies of water the humidity is generally higher and thus a more penetrative light wave is necessary to produce optimal results. Unfortunately when testing this device we were unable to perform any testing along the water as our testing boat had been fully winterized and the temperature was below freezing. However we were able to collect many samples on a relatively straight stretch of land, taking infrared photos at 10 feet intervals for up to 70 feet away. We conducted testing both with and without the infrared bulb array due to challenges that arose with reflected infrared light being too harsh during the closer photos. Which led to the realization that an adaptive array scheme was necessary.

B. Challenges

Text detection and OCR Difficulty- The challenges faced during image processing were quite considerable as normal text processing and OCR are not equipped to handle the challenges of blurry low pixel images. While Tesseract is a powerful modern and open source optical character recognition engine, it is not trained on data models incorporating reflective text and the inconsistencies in shading and lighting that arise due to this. In addition we found EAST text detector to also struggle on picking up accurate text boxes. It would often detect the text but cut part of a letter out which would

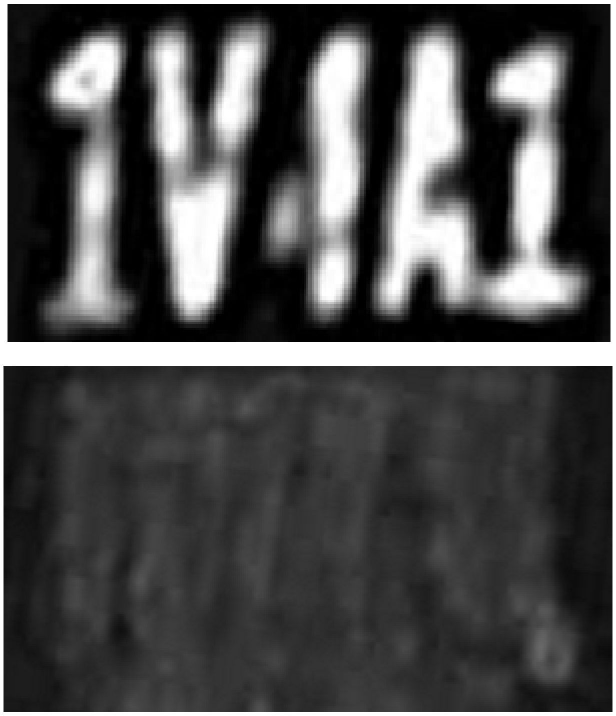

Fig. 4. Two unprocessed infrared photos taken in purely ambient light from 40 ft. The first photo is our retro reflective design, the second is taken with a standard black text on white background license. Both licenses are the same size lettering with identical characters (1V4A1).

(a) Excess Infrared Reflection (b) Closeup on License

Fig. 5. Infrared reflection problem caused by trail camera incapable of dimming infrared bulb array with respect to distance. Photo captured at a range of 20 feet.

make gauging the actual OCR much more difficult since it might have just been an inaccurate bounding box. One method we used was raising the tolerance on the bounding boxes surrounding the license plate, however this cannot be done too much due to the nature of tesseract when running on background images can decrease OCR precision significantly. Eventually to eliminate inconsistencies from EAST detection and test soley OCR I manually selected the roi for the license and then ran it through some code to preprocess the photo. Since the photos were blurry it was necessary to rescale, de-noise, and remove blur using my code before feeding the text back into Tesseract with parameters for better accuracy. Using this method I was able to produce data with reasonable accuracy to simulate if our text detection was using a better trained model for our data. In this way I could some of the benefits to be gained by incorporating retro-reflective into our license plate lettering.

Pixel Density- Our infrared camera and much of the data we collected during the initial data collection were shot with 12MP cameras however after reviewing the data it is clear that especially from the farther distances (mostly 40 and onwards) the text region of interest is going to remarkably small and often tilted or skewed in the case of a boat. For perspective, from a little more than 30 feet your license region of interest will appear with a width of 200 pixels and a height of 100 pixels. This will decrease continuously as the license is moved away and that is only if the license is directly head on. If tilted in any direction the region of interest becomes smaller. Although an affine transformation could be applied to increase readability of the license, the increased distortion would still negatively effect the OCR especially at more drastic angles. Infrared backs

Infrared Reflection- After collecting infrared data there was a noticeable problem with the data collected within a 20 foot range from the camera. The camera did not have built in software and sensors to decrease the brightness of the bulb array at closer distances. This meant that essentially everything near to the camera was becoming drastically over-exposed due to the intense reflection of infrared light. In figure 2 you can see the impact of too much infrared light at a range of 20 feet. To eliminate this effect we could implement software to detect range or subject size and dim the array accordingly. Although this is perceived as less important as it is unlikely (although possible) the boat would approach the camera that close depending on the setup of the camera which would most likely be positioned along the shoreline. In our testing we also tested the infrared camera in only ambient lighting without the benefit of the infrared LED's which led to promising results and eliminated the excess reflection.

IV. Evaluation

We noticed the performance did drop off considerably after around 40 feet. There are a number of factors that attribute to this decline but most is the drop in resolution. Much more data still must be collected in ambient settings for the preprocessing and settings of tesseract can be refined. The results were promising and presented some unique findings into how infrared retroreflection could increase photo clarity, however altogether our implementation lacked the refined touch it needed to really be able to process images from 60 feet with Precision as you can see from our OCR graph (figure 6). Much of the data we collected was botched due to the intensity of the infrared array bulb. Switching to ambient light and preprocessing the data significantly improved our results however the low resolution images were just not enough for our OCR engine to process which is why we now believe feature points and a trained learning model may be necessary to increase our accuracy dramatically.

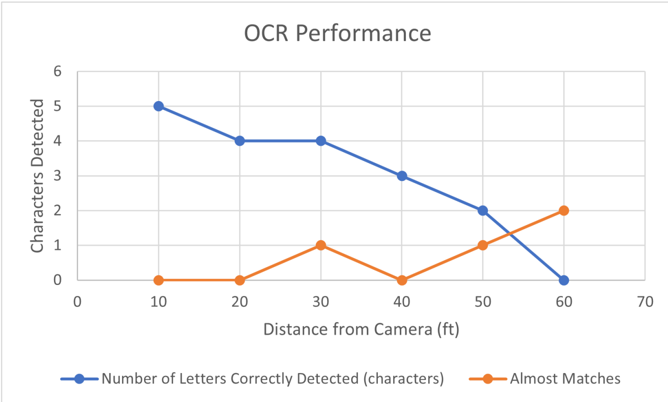

Fig. 6. Performance of Optical Character Recognition

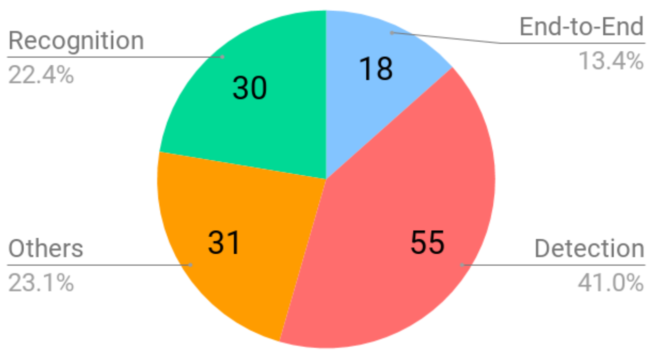

Fig. 7. Percentage of papers on Deep Learning Based Approach for OCR

V. Related Work

Text recognition using deep learning is a broad and growing field. Mainly, papers fall into three different categories: text recognition, text detection (optical character recognition), or end-to-end text detection which is a combination of the first two. In each of these three categories each group seems to continuously evolve and generate numerous new techniques each year largely due to the popularity of computer vision and document analysis [28,29,30] and their growing challenges that come with increased demand.

These challenges to text detection in natural scene images include "diverse text variabilities in color, fonts, orientations, languages and scales, extremely complex and text-like backgrounds, as well as some distortions and artifacts caused by image capturing like non-uniform illumination, low contrast, low resolution and occlusion."[25] Text detection methods can be roughly separated into two different methods: the bottom-up method, examples: [10,11,12,13,14,15], and the top-down method, examples: [16,17,18,19,20,21,22,23,24]. Recently, CNN (Convolutional Neural Network) "based top-down approaches have become the mainstream"[25] because of their superiority in terms of both accuracy and adaptability. These scene text detection (OCR) algorithms commonly use anchors to detect shapes like Faster R-CNN [26] or SSD [27], which are not flexible enough for proper text detection leading to poor performance[31]. Anchors are reference points spread out across the image and using each anchors Intersection-over-Union (IoU) that we overlay with a ground-truth table. In order to detect a character there needs to be a high value IoU overlay with the object. For intricate shapes, anchors are not flexible enough as a rectangular anchor will poorly overlay causing inefficiency.

Multiple papers have been published attempting to solve this inflexibility like the works by [21] and [22] who propose "quadrilateral anchors to hunt included text proposals which can better fit the multi-oriented text instances."[25] However, these works still use anchors and face inferior performance to works like [25] who propose anchor-free region text detection but they face their own unique challenges as well. Text Detection (OCR) is clearly evolving as more and more work is published.

Studying [2] we learned there are many factors affecting the accuracy of natural scene text recognition. Image and sensor noise are much higher than using things like scanners which operate at close range. On top of that, lower quality cameras often interpolate the pixels to produce real colors which just leads to more image noise. In a natural scenes such as a river or the ocean we can expect drastically different light condition where the sun will be shining brightly and may saturate our photo. Another concern is non planar or tilted viewing angles that may make the text hard or impossible to recognize. Non planar should not be as much of a concern as tilted due to the fact that the licenses can be designed to be rigid.

End-to-end text detection has similar challenges but "unlike character recognition for scanned documents, recognizing text in unconstrained images is complicated by a wide range of variations in backgrounds, textures, fonts, and lighting conditions."[32] After reviewing several works on end-to-end over the past 8 years [32,33,34] I found that each continuously builds off the past to achieve a greater precision value or a lower inaccuracy as seen in [33]; "[our detection method] produces a detection performance... averagely 2 percent higher than 'Ours DetOnly' [the detection method from 2011]."

VI. Conclusion

We took a broad investigation into the erosive effects of boats towards shorelines, vegetative wildlife, and marine wildlife like manatees. We performed several tests on the water and on land using normal view-able light and infrared light. We researched machine learning optical character recognition (OCR) modules for possible computer vision solutions. Our results show a proof of concept of how RRVIS incentives safe boating practices resulting in the protection of our natural wildlife.

Future Works- Currently our setup uses noise filtering but there is an incredible amount of pre-processing to still be explored. One step that is of interest would be adding a deep learning pre-processing step using a custom trained model for blurry images. One research into deep learning and computer vision techniques focuses on adapting the lighting the lighting and coloring of images to make them more visible in night time settings. Despite the number of papers working to decipher blurred and low resolution text detection there is frighteningly little knowledge on glare and reflective text, making a novel model or pre-possessing method necessary to further increase the accuracy beyond what we were able to produce with modern CNN engines. Custom training our own reflective sign detection to increase accuracy would be vital to further improving our results although the text visibility and detection at the extremely far ranges still holds promise and proof of concept.

Alternate Design- In addition, there is potential to discard text completely and rethink the license plate. We could build reflective bar codes or patterns for detection using infrared lighting or instead build more complicated patterns using multicolored reflective surfaces to utilize the visible light range. This would allow us to increase our accuracy by discarding traditional numerical codes. We could segment the license and determine shaded regions to produce a unique key with extreme accuracy. Reaping the benefits of increased range and low powered/cost detection in addition to increased precision as we are no longer relying on a finely tuned optical character recognition process.

Appendix A
References

[1] Calleson, CS, and RK Frohlich. 2007. Slower boat speeds reduce risks to manatees. Endangered Species Research. Vol. 3, pp. 295-304

[2] Thillou, Céline Gosselin, Bernard. 2007. Natural Scene Text Understanding. 10.5772/4966.

[3] Reference for Figure 1: Kern, C. The Difficulty of Measuring the Absorption of Scattered Sunlight by H2O and CO2 in Volcanic Plumes: A Comment on Pering et al. "A Novel and Inexpensive Method for Measuring Volcanic Plume Water Fluxes at High Temporal Resolution," Remote Sens. 2017, 9, 146. Remote Sens. 2017, 9, 534; doi:10.3390/rs9060534

[4] Reference for Figure 7: https://github.com/hwalsuklee/awesome-deep-text-detection-recognition/blob/master/overall$_{pi_c}hart.png$

[10] Lukas Nuemann and Jiri Matas. 2010. A method for text localization and recognition in real-world images. In ACCV, pp. 770-783.

[11] Boris Epshtein, Eyal Ofek and Yonatan Wexler. 2010. Detecting text in natural scenes with stroke width transform. In CVPR, pp. 2963-2970.

[12] Cong Yao, et al. 2012. Detecting texts of arbitrary orientations in natural images. In CVPR, pp. 1083-1090.

[13] Xucheng Yin, et al. 2014. Robust text detection in natural scene images. IEEE Trans. PAMI, 36(5):970-983.

[14] Xucheng Yin, et al. 2015. Multi-orientation scene text detection with adaptive clustering. IEEE Trans. PAMI, 37(9):1930-1937.

[15] Lei Sun, Qiang Huo, Wei Jia and Kai Chen. 2015. A robust approach for text detection from natural scene images. Pattern Recognition, 48(9):2906-2920.

[16] Max Jaderberg, Andrea Vedaldi and Andrew Zisserman. 2014. Deep features for text spotting. In ECCV, pp. 512-528.

[17] Zheng Zhang, et al. 2016. Multi-oriented text detection with fully convolutional networks. In CVPR, pp. 4159-4167.

[18] Cong Yao et al. 2016. Scene text detection via holistic, multi-channel prediction. arXiv preprint arXiv:1606.09002.

[19] Max Jaderberg, Karen Simonyan and Andrea Vedaldi and Andrew Zisserman. 2016. Reading text in the wild with convolutional neural networks. IJCV, 116(1):1-20.

[20] Ankush Gupta, Andrea Vedaldi and Andrew Zisserman. 2016. Sythetic data for text localization in natural images. In CVPR, pp. 2315-2324.

[21] Jianqi Ma, et al. 2017. Arbitrary-Oriented scene text detection via rotation proposals. arXiv preprint arXiv:1703.01086.

[22] Yuliang Liu and Lianwen Jin. 2017. Deep matching prior network toward tighter multi-oriented text detection. In CVPR, pp.1962-1969.

[23] Xinyu Zhou, et al. 2017. EAST: An efficient and accurate scene text detector. In CVPR, pp. 5551-5560.

[24] Kiaming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. 2016. Deep residual learning for image recognition. In CVPR, pp. 770-778.

[25] Zhuoyao Zhong, Lei Sun, Qiang Huo. 2018. An Anchor-Free Region Proposal Network for Faster R-CNN based Text Detection Approaches. arXiv preprint arXiv:1804.09003

[26] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, pp. 91-99.

[27] Wei Liu, et al. 2016. SSD: Single shot multiBox detector. In ECCV, pp. 21-37.

[28] Asif Shahab, Faisal Shafait and Andreas Dengel. 2011. ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. In ICDAR, pp. 1491-1496.

[29] Dimosthenis Karatzas, et al. 2013. ICDAR 2013 robust reading competition. In ICDAR, pp. 1484-1493.

[30] Dimosthenis Karatzas, et al. 2015. ICDAR 2015 robust reading competition. In ICDAR, pp. 1156-1160.

[31] Wenhao He, Xuyao Zhang, Fei Yin and Chenglin Liu. 2017. Deep direct regression for multi-oriented scene text detection. In ICCV, pp. 745-753.

[32] Tao Wang, David Wu, Adam Coates and Andrew Ng. 2011. End-to-end text recognition with convolutional neural networks. In ICPR.

[33] Liang Qiao, et al. 2020. Text perceptron: towards end-to-end arbitrary-shaped text spotting. arXiv preprint arXiv:2002.06820.

[34] Hui Li, Peng Wang, Chunhua Shen. 2017. Towards end-to-end text spotting with convolutional recurrent neural networks. arXiv preprint arXiv:1707.03985.