Ship License Number Layout Normalization Based on Regional Texts Fine Localization

BaoLong Liu\textsuperscript{1,2}, Sanyuan Zhang\textsuperscript{1}, Zhenjie Hong\textsuperscript{2} and Xiuzi Ye\textsuperscript{2,*}

\textsuperscript{1}College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
\textsuperscript{2}College of Mathematics, Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, China
\textsuperscript{*}Yexiuzi@wzu.edu.cn

Abstract. Ship license numbers (SLNs) are frequently used to identify ships. An automatic SLNs recognition system makes it convenient and fast to recognize SLNs. A significant prerequisite of many characters recognition algorithms is that characters need to be recognized should be segmented as single ones or at least be listed in one horizontal line. However, SLNs from different ships may have different layouts and the characters of SLNs are often written in more than one lines. In order to recognize these multi-layout SLNs, it is momentous to normalize then into one line. In this paper, we present a SLN layout normalization framework based on regional texts fine localization. For an input SLN image, a series of pre-processing operations are first conducted to enhance the quality of the input SLN image. Then, a single shot text detector with regional attention is used to fine detect characters in that SLN. The normalization algorithms are finally proceeded to normalize the input SLN image based on the detected text bounding boxes and the layout priors of SLNs. At last, the proposed framework is tested on 3216 SLN images, the proposed framework is proved to be effective with a normalization rate of 87.38%.

1. Introduction

What are ship license numbers? Simply speaking, SLNs are ships names. Every ship has a name. Ship names are not randomly given, they are generally given by following the ship naming conventions issued by the local official authorities. Therefore, it is more reasonable to call them “license”.

Similar with a car license plates recognition system, an automatic ship license numbers recognition system can be designed to identify ships by recognizing the characters in SLNs. The recognition pipeline of SLNs are similar with car license plates’, it is mainly comprised of image acquisition, license location, tilt correction, character segmentation and character recognition.
Figure 1. Some sample normalization results of SLN images (note that some images may be resized for the convenience of display). (a) Left: input SLNs. Right: Normalization results. (b) Left: input SLNs. Right: Normalization results.

Several papers concerning SLNs recognition have been published [1-3]. [1] used a transferred deep CNN model in combination with prior features to localize multi-style ship license numbers in nature scenes. Liu et al. [2] presented an effective coarse-to-fine approach for locating various SLNs in the wild. [3] proposed a horizontal tilt correction method for SLNs.

A significant prerequisite of many license (or generally characters) recognition algorithms is that characters need to be recognized should be segmented as single characters. Even for the most advanced deep convolutional neural network-based characters recognition algorithms, e.g., CRNN [4], the characters need to be recognized also should be listed in one horizontal line. However, as shown in Figure 1, the styles of SLNs are multiple. Characters contained by these SLNs may be written in different lines with different angles. To boost the accuracy of SLNs recognition, in this paper, we present a SLNs layout normalization framework based on regional texts fine localization. The main objective is to normalize those SLNs that are of multiple styles into one nearly horizontal line.

SLNs from different countries or different ships may have different layouts. In this paper, the SLNs from Chinese cargo ships are used as experimental samples to demonstrate the pipeline of the proposed framework. SLNs of Chinese cargo ships from one same ship may have different layouts. It is thus reasonable to use SLNs of Chinese cargo ships as experimental samples.

The core difficulties of SLNs layout normalization mainly lie in two aspects: 1) Characters contained in the same SLN may be different in script, angle, size, aspect ratio and stroke width, and these characters, especially Pinyin, are of low resolution, at the same time. 2) Characters of SLNs may be written in different lines. The most difficult thing is that, some characters in an SLN are written in two line, while others are written in one line. Therefore, it is hard for image processing-based methods, e.g., MSER or Connected-Component analysis, to distinguish these cluttered characters out.

In this paper, for an input SLN image, a series of pre-processing operations are first conducted to enhance the quality of the input SLN image. Then, a single shot text detector with regional attention (SSTD) [5] is used to fine detect characters in that SLN. The normalization algorithms are finally
proceeded to normalize the input SLN image based on the detected text bounding boxes and the layout priors of SLNs. Figure 2 shows the procedures of the proposed framework.

The proposed framework is experimented on 3216 collected SLN images. The experimental results in Sec. 3 indicate that the proposed approach can effectively normalize SLNs with a normalization rate of 87.38%.

2. The proposed framework

For an input SLN image with cluttered characters contained, the objective of the proposed framework is to normalize the characters need to be recognized into an approximately horizontal line and discard the interference characters, e.g., Pinyin, the phonetic transcriptions of Chinese characters.

![Figure 2. The processing procedures of the proposed framework.](image)

SLNs from Chinese cargo ships can be classified into three categories according to the line number, namely, one-line SLNs, one-and-a-half-line SLNs and SLNs comprised of multiple (greater or equal to two) lines. One-and-a-half-line SLNs mean that some characters contained in one SLN are written in two line, while the others are written in one line.

Figure 2 shows the processing procedures of the proposed framework. It mainly consists of six steps. The technical details of each step are described as follows:

1) **Aspect ratio-based SLN classification.** SLNs from Chinese cargo ships generally contains at least 5 characters. In this situation, one-line and one-and-a-half-line SLNs are often have greater aspect ratios than multiple-line SLNs. In our experiments, if the aspect ratio of an input SLN image is greater than 2.0, this SLN will be classified as a multiple-line SLN, otherwise it will be classified as a one-line/one-and-a-half-line SLN. The reason of conducting aspect ratio-based SLN classification is that the horizontal tilt correction method [3] we used can only handle one-line or one-and-a-half-line SLNs.

2) **Horizontal tilt correction.** The horizontal tilt SLNs correction method presented in [3] is used to correct the input SLN because the SSTD-based text region detection method [5] can only detect axis-oriented horizontal text lines.

3) **Laplacian-based SLN image enhancement.** Ships, especially cargo ships, usually have large volume, while SLNs contained by them have very small sizes and low resolutions by comparison [2]. It is hard even for the state-of-the-art convolutional neural network (CNN)-based text detector, e.g., SSTD [5], to directly detect these small characters out. Therefore, a Laplacian-based enhancement algorithm is used to enhance the input SLN images.

4) **SLN image padding.** Padding is often used in deep convolutional neural networks to preserve the size and features of some convolutional layers. As is well known, the fine information captured by each conv layer of a CNN becomes coarser with its receptive field size increasing [1]. To boost the detection accuracy, the input SLN image is padded in advance. Assume that $w$ and $h$ represent the width and height of the input SLN image, respectively, the padding size of width and height are $w/2$ and $h/2$. The black rectangle in Figure 3(a) is the padding result of the SLN image in Figure 3(a) (red rectangle).

5) **SSTD-based regional texts fine detection.** SSTD [5] is a single-shot text detector that directly
outputs word-level bounding boxes in a natural image. It uses an attention mechanism which roughly identifies text regions via an automatically learned attentional map, which is helpful to substantially suppresses background interference and produce accurate inference of words, especially for the words at extremely small sizes. SLNs usually are of very small and low resolutions, it is thus reasonable to use SSTD to detect the characters, especially Pinyin, and then discarded them.

Figure 3. Layout prior-based SLN normalization.

6) Layout prior-based SLN normalization. The normalization procedures of one-line, one-and-a-half-line and multiple-line SLNs will be described in this part, respectively. (1) one-line and one-and-a-half-line SLNs. Denoting rectangle $S(S_x, S_y, S_w, S_h)$ as the input SLN image (red rectangle in Figure 3(a), for example), where $S_x$ and $S_y$ are the top-left coordinates of $S$, $S_w$ and $S_h$ are the width and height of $S$ respectively. $D(D_x, D_y, D_w, D_h)$ denotes the rectangle with the greatest width from the text rectangles detected by SSTD (green rectangles in Figure 3(a)). $T(T_x, T_y, T_w, T_h)$ represents the bottom-left rectangle of $S$. First of all, the bottom-left rectangle is defined as $T'(S_x, S_y + 0.5 \times S_h, 0.67 \times S_w, 0.5 \times S_h)$. based on the layout prior of one-and-a-half-line SLNs, the Pinyin characters would appear inside this area very likely. Thus, the Pinyin characters (green rectangles inside of $T'$ in Figure 3(a)) will be masked with black masks to reduce noises. Then, $D'$ is adjusted as $D'(S_x, D_y, S_w, D_h)$. $D'$ is treated as the final normalization result (Figure 3(b)). (2) Multiple-line SLNs. Multiple-line SLNs will be normalized into one line by chaining the detected bounding boxes one by one from top to bottom. Figure 4 shows the detail processing procedures of some SLN instances of the proposed framework.

Figure 4. SLNs layout normalization (note that some images may be resized for the convenience of display). (a) The input SLN images. (b) SLN images after tilt correction, enhancement and padding. (c) Texts fine detection results. (d) The normalization results.
3. Experimental results

3.1 Testing dataset
The effectiveness of the proposed framework was tested on 3216 SLN images. These SLN images were collected in the practical application environment of the proposed framework — Beijing Hangzhou the Grand Canal (Hangzhou). The cameras were set up along the river, and photographed SLNs from thousands of ships over day and night in different seasons. The background of them are complex, illuminations and characters are various (This SLNs recognition dataset will be released publicly in the near feature). Therefore, these testing SLN images are challenging enough. Figure 1 shows some testing SLN images.

3.2 Experiment results
The input SLNs are considered successfully corrected if the cluttering Pinyin characters of one-and-a-half-line SLNs are successfully cropped, and all the characters contained in those SLNs are observed to be in one fairly horizontal line. Otherwise, they will be treated as failures. The quantitative experiment results are listed in Table 1. Table 1 tells that the proposed framework achieves a normalization rate of 87.38%, suggesting the proposed approach has good SLNs normalization ability. One can also recognize the impressive SLNs normalization performances of the proposed framework from Figure 1.

Table 1. SLNs normalization results of the proposed framework

| Number of SLN images | Number of valid normalizations | Number of failures | Normalization rate |
|----------------------|-------------------------------|--------------------|-------------------|
| 3216                 | 2810                          | 406                | 87.38%            |

4. Conclusion
An effective layout normalization framework of SLNs have been proposed in this paper. The proposed framework was tested on 3216 collected SLN images, as the first work concerning the layout normalization of SLNs, the proposed framework is proved to be effective with a layout normalization rate of 87.38% and is very helpful to the characters recognition algorithms of SLNs.

Acknowledgments
This work was supported partly by National Natural Science Foundation of China (61272304 and 61772374), and Zhejiang Provincial Natural Science Foundation of China (LY16F020023 and LQ18F030010).

References
[1] B. Liu, X. Lyu, C. Li, S. Zhang, Z. Hong, and X. Ye, “Using transferred deep model in combination with prior features to localize multi-style ship license numbers in nature scenes,” in IEEE International Conference on TOOLS with Artificial Intelligence, 2018, pp. 506–510.
[2] B. Liu, J. Sheng, J. Dun, S. Zhang, Z. Hong, and X. Ye, “Locating various ship license numbers in the wild: An effective approach,” IEEE Intelligent Transportation Systems Magazine, vol. 9, no. 4, pp.102–117, 2017.
[3] B. Liu, S. Zhang, Z. Hong, and X. Ye, “A Horizontal Tilt Correction Method for Ship License Numbers Recognition,” in IOP Journal of Physics: Conf. Series, vol. 976, 012013, 2018.
[4] B. Shi, X. Bai and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp. 2298-2304, Nov. 1 2017.
[5] P. He, W. Huang, T. He, Q. Zhu, Y. Qiao, and X. Li, “Single Shot Text Detector with Regional Attention,” in IEEE International Conference on Computer Vision, 2017, pp. 3066-3074.