# 摘 要

随着移动智能设备的普及，图像的获取越来越便捷，提取并理解智能设备拍摄的图像中的信息具有重要的应用价值。文字由于包含了大量丰富的语义信息，对于计算机理解图像具有重要意义。文本识别作为计算机视觉领域基础性研究工作，自然场景的文本识别具有深远的研究意义和实用价值。在图像检索、人机交互、车牌识别等领域有着广泛的应用。自然场景由于复杂的背景、多样化的字体、以及随意的分布等问题，使得传统的光学字符识别技术（OCR）难以被广泛应用。自然场景中的文字检测与文本识别仍然存在很多需要解决的技术难点。

得益于计算机硬件性能的提高，以及大数据获取越来越便捷，深度学习技术近年来取得了重大突破，在模式识别领域被广泛应用。本文借鉴图像检索、目标检测、自然语言处理等领域的进展，结合深度学习技术，提出了新的识别方法，本文主要工作如下：

（1）提出了基于CNN和RNN的文本识别方法。利用CNN提取原始图像的抽象特征，然后将特征送入RNN中建模图像全局信息。针对传统的文本识别需要较深的专业背景知识，文本检测与识别等过于复杂等问题，该方法采用端到端训练，无需过多的背景知识。将检测与识别融为一体。大大降低了训练难度。

（2）对自己生成的验证码图片数据集和Google的街景门牌号码SVHN数据集上进行了两部分实验。第一部分实验验证了单纯RNN对文本识别的有效性；第二部分实验探索CNN结合RNN对文本识别的提高以及不同的滑动窗对识别率的影响。

关键字：自然场景；文本识别；深度学习；RNN；CNN

# ABSTRACT

With the popularity of mobile smart devices, get images becomes more and more convenient. It has high application value that extract and understand the information in the images captured by mobile smart devices. It is significant important for computer to comprehend images due to text contains a lot of semantic information. As a basic research work in the field of computer vision, text recognition in natural scenes has great research and practical value. It has been widely used in the fields of image retrieval, human-computer interaction and license plate recognition. Due to the complexity of the background, the diversity of fonts, and the random distribution of natural scenes, making the traditional optical character recognition technology(OCR) is difficult to be widely used. Text detection and recognition in natural scenes still have many technical difficulties to be solved.

Thanks to the improvement of computer hardware performance and the convenience of large data acquisition, deep learning technology has made a great breakthrough in recent years, and is widely used in pattern recognition field. Based on the progress of image retrieval, object detection and natural language processing, this paper proposes a new recognition method combined with deep learning technology. As follows is the main work of this paper:

(1) This paper proposes a text recognition method based on CNN and RNN. The abstract features of original image are extracted by CNN, and the imported into RNN to modeled global information. Our method is training by end to end without too much background knowledge and integrated identification and detection, while text detection and recognition is too complex and requires a deeper professional background knowledge. This greatly reduce the difficulty of training.

(2) We conducted experiments on captcha image generated by ourselves and the famous Google Street View House Number dataset. In the first part, we demonstrate the effectiveness of RNN for text recognition. In the second part, we investigate the improvement of text recognition by RNN which combined with CNN, and the influence of different sliding windows on recognition rate.

**Key words:** text recognition; deep learning; RNN; CNN