

基于深度学习的模糊遮挡船号文本识别

余胜任

南京信息工程大学数学与统计学院, 江苏 南京

收稿日期: 2023年10月30日; 录用日期: 2023年11月27日; 发布日期: 2024年1月19日

摘要

随着我国船运业的蓬勃发展, 船号作为船舶身份的唯一标识, 确保准确识别船舶船号对于船舶管理至关重要。然而, 在实际场景中, 船号常常存在文字模糊或者被遮挡的情况, 这会极大地降低识别的精准度。由于传统的图像处理和文字识别方法不能很好地解决这一问题, 而且现有的大多数方式通常分二个步骤解决这个问题, 先进行图像恢复再对恢复图像进行文字识别, 但这种方式忽视了图像恢复和文字识别间的关联性, 因此本文提出了一种联合生成对抗网络(GAN)和卷积循环神经网络(CRNN)的针对模糊遮挡船号文本的双分支耦合文字识别框架, 称为SRC-GAN, 通过对抗性学习将文字识别和图像恢复集成起来。通过将识别模型和GAN模型联合训练, 学习更多图像的共性特征, 从而对低质量的图像有更好的识别性能。在船舶数据集和CTW数据集上的识别实验表明, 该方法相较于原始CRNN识别精度平均分布提升了11.98%和10.68%, 相较于二阶段识别模型也有一定的优势, SRC-GAN对于模糊遮挡文本图像有着更好的识别效果。

关键词

深度学习, 文字识别, GAN, CRNN

Deep Learning-Based Recognition of Blurred Ship Identification Text in Maritime Context

Shengren Yu

School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing Jiangsu

Received: Oct. 30th, 2023; accepted: Nov. 27th, 2023; published: Jan. 19th, 2024

Abstract

With the vigorous development of China's shipping industry, the ship number serves as the unique identifier of the ship's identity, ensuring accurate identification of the ship number is crucial for ship management. However, in practical scenarios, ship numbers often have blurred or obstructed text, which greatly reduces the accuracy of recognition. Due to the fact that traditional image

文章引用: 余胜任. 基于深度学习的模糊遮挡船号文本识别[J]. 建模与仿真, 2024, 13(1): 450-459.

DOI: 10.12677/mos.2024.131043

processing and text recognition methods cannot effectively solve this problem, and most existing methods usually solve this problem in two steps, first performing image restoration and then performing text recognition on the restored image, this approach ignores the correlation between image restoration and text recognition. Therefore, this article proposes a dual branch coupled text recognition framework called SRC-GAN, which combines Generative Adversarial Network (GAN) and Convolutional Recurrent Neural Network (CRNN) for fuzzy occluded ship number texts. It integrates text recognition and image restoration through adversarial learning. By jointly training the recognition model and GAN model, more common features of images can be learned, resulting in better recognition performance for low-quality images. The recognition experiments on the ship dataset and CTW dataset show that this method has improved the average recognition accuracy distribution by 11.98% and 10.68% compared to the original CRNN, and also has certain advantages over the two-stage recognition model. SRC-GAN has better recognition performance for fuzzy occluded text images.

Keywords

Deep Learning, Text Recognition, GAN, CRNN

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着我国航运行业的蓬勃发展,对航运船舶的管理也提出了更高的要求。每个船舶都拥有唯一的身份标识号,一种准确、高效的船舶识别方案对于智能交通系统(ITS)在水路运输场景中的应用具有重要意义[1]。目前,我国主要依赖船舶自动识别系统(AIS)来确定船只的身份,其中船号作为船舶的唯一身份标识,是船舶自动识别系统的关键输入信息之一。然而,对于一些违规货船或 AIS 系统发生故障的情况,只能依靠人工观测,这种方法对于流量较大的水域,容易出现漏检误判问题,且效率低下。因此,本文希望通过深度学习与图像处理技术与 AIS 系统配合,制定一种更高效的文字识别解决方案。目前,规则的文档文字识别研究已经相对完善,并具有较高的识别准确率和鲁棒性。然而,船号识别由于其场景的复杂性,容易受到光照不均、透视变形和残缺遮挡等因素的影响[2] [3],从而导致得到的文本图像常常存在模糊遮挡的问题,如图 1,这些问题给船号的精准识别带来了相当大的困难。以往对于模糊遮挡文本的识别通常分为二个阶段[4],先对图像进行恢复,再对恢复后的图像进行识别,这种方式可以提高文本图像的质量,也提高了文本的识别精度,但是忽略了图像恢复和文字识别间的关联性,不同图像在特征空间中的分布是不同的,将恢复模型和识别模型分开,生成的恢复后的图像不一定能够生成适合文字识别的特征分布。



Figure 1. Blurred and obscured ship identification numbers

图 1. 模糊和被遮挡的船号

随着深度学习方法在图像处理领域的快速发展,现代场景文字识别模型将文字识别过程分为文本区域检测和文本识别二个步骤[5],以船号识别为例,先通过船号区域检测预测几何边界框(BBox)来定位船

船舶号文本行,再基于检测模型定位的感兴趣区域(RoIs)进行船号文本识别。目前较为主流的文字区域检测算法有CTPN、EAST、PSENet、YOLO这几类。CTPN是由Zhi等[6]人提出的结合了CNN和LSTM深度网络,用于检测复杂场景下的横向分布的文字。该算法在对水平文字和小区域文字检测时,具有较高的准确度,但是对于倾斜或不规则文字识别准确率较低,且由于算法分两步完成,因此检测速度较慢。EAST是由Zhou等[7]人提出的基于像素点的场景文本检测算法。它省去了很多复杂的后处理操作直接对字符区域进行分割和对字符像素到字符边界框的距离进行回归,简单高效。但是受制于感受野的大小,在对每一个字符像素到边框四周的距离进行回归的时候,对于长边的回归会出现错误。PSENet是由Li等[8]人提出的一种基于分割的文字检测算法,它可以检测和分割具有不规则形状的区域,但对于噪声或复杂场景仍然不够鲁棒。YOLO是由Joseph[9]等人提出的单阶段目标检测模型,它将目标检测任务转化为单个神经网络的回归问题,能够同时预测图像中多个物体的位置和类别,相较于其他方法具有更快的推理速度和更高的性能,因此本文选用YOLOv5目标检测算法进行船号文本区域检测。

一旦获得了目标文字区域,接下来的任务是对目标区域中的文本进行识别。基于深度学习的文本识别方法大致可以分为两类。它们的相似之处在于使用CNN作为图像特征提取器,但不同之处在于文本内容解码。目前广泛使用的两种技术是CTC[10]和Seq2Seq[11]。由于文本线条通常具有倾斜、旋转或弯曲等非水平形状,文本识别也经历了从水平文本到任意形状文本的演变。CRNN[12]是最常用的文本识别模型之一。它基于全卷积结构的CNN提取图像特征序列,并利用双向长短期记忆(Bi-LSTM)[13]来预测每帧的标签分布。然后使用连接时间分类(CTC)将每帧的预测转化为最终的标签序列。CRNN能够识别任意长度的文本,并且只需要行级别而不是字符级别的注释。由于其快速和轻量级的特性,本文采用CRNN作为船舶船号的文本识别模型。然而,在实际应用中,由于场景的复杂性,船号文字区域常常存在遮挡和模糊等问题,从而极大地影响了识别精度。解决此问题,传统方法通常先对文字区域图片进行矫正和形态学处理再进行识别,但是文字矫正对于文字本身并没有进行有效的处理,通过形态学中的膨胀、腐蚀操作提高文字的分辨率,或者对二值图进行归一化等操作可以提高文字与背景的对比度,可提高识别准确度,但是这种方法的缺点在于鲁棒性过低,面对复杂的船舶场景文字,不能有效地处理所有目标文字。而且现有的大多数方式通常分二个步骤解决这个问题,先进行图像恢复再对恢复图像进行文字识别,但这种方式忽视了图像恢复和文字识别间的关联性。因此为了应对遮挡和模糊对文字识别的挑战,本文提出了一种联合GAN和CRNN的针对模糊遮挡船号文本的文字识别框架,称为SRC-GAN,旨在去除遮挡并提高图像质量,以实现图像的复原并提升文字识别精度。

2. 联合图像恢复的文字识别网络

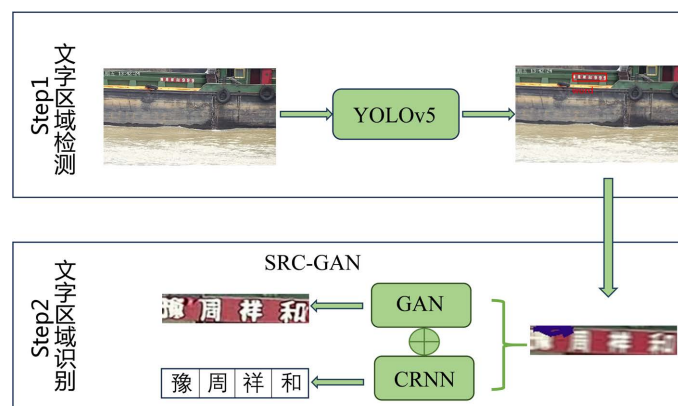


Figure 2. Overall flowchart
图 2. 整体流程图

在船舶场景下的文字识别任务中，首要步骤是文字区域的检测。本文采用 YOLOv5 网络来执行这一任务，该网络在文字区域检测上能够达到超过 95% 的准确率。通过 YOLOv5 模型预测到目标文字区域后再根据预测几何边界框进行裁剪和矫正后得到文字区域图像，再将图像作为输入进入 SRC-GAN 模型进行文字识别，最终生成高质量文字图像和得到精准文字识别结果。模型整体流程见图 2。

2.1. 基于 YOLOv5 的船号文字区域检测

本文选择 YOLOv5 目标检测算法进行船舶船号文字区域的检测，这一选择基于其卓越的实时性、高精度、适用于小目标的优化。YOLOv5 的轻便模型结构使其易于部署在资源受限的设备上，如嵌入式系统或移动设备。这一研究为船舶监控和导航系统提供了一种高效且可行的解决方案。利用 YOLOv5 的可调控模型宽度和深度的特性，以实现模型参数量和计算量的精准掌控。这样的调整导致了不同版本的 YOLOv5 网络结构，包括 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 等。其中，YOLOv5s 具有相对较少的参数量和计算量，使其成为适合嵌入式部署的理想选择，因此，本文将 YOLOv5s 作为船号文字区域检测的基准模型。总体来看 YOLOv5 由以下组成部分构成：输入层(input)、主干特征提取网络(backbone)、特征融合网络(neck)以及检测头(head)见图 3。

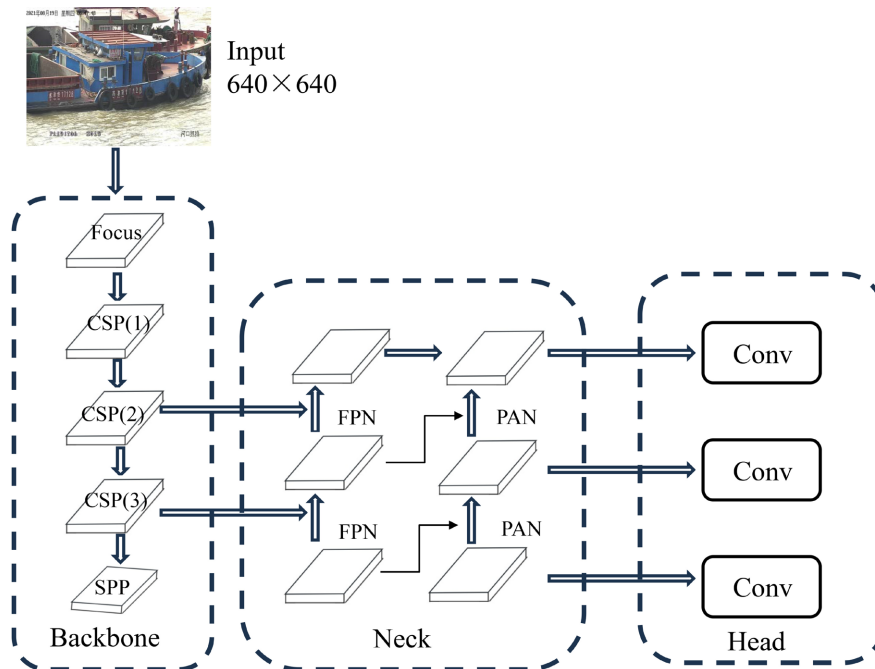


Figure 3. YOLOv5 network architecture diagram

图 3. YOLOv5 网络结构图

在具体结构方面，YOLOv5 的输入层经过了多方面创新，包括多尺度输入、Mosaic 数据增强算法、AutoML 技术的应用、CIoU 归一化策略以及通道顺序优化等，这些创新点共同提高了模型的灵活性、性能和鲁棒性。主干网络采用了 CSPDarknet53 结构，这一结构具备强大的特征提取能力。另外，Neck 模块结合了特征金字塔结构(FPN)和路径聚合网络结构(PAN) [14]，以增强网络对不同缩放尺度特征的融合能力，从而提高了模型的准确性和鲁棒性。Head 模块则负责输出目标检测结果，采用 GIOU 边界框损失函数和非极大值抑制算法，以进一步提升多目标和小目标的检测性能。这些模块使 YOLOv5 成为一个强大的工具，适用于船舶船号目标检测和其他多类别目标检测应用。

2.2. 联合 GAN 和 CRNN 的船号文字识别

2.2.1. CRNN

本文所选用的文字识别基础方法为 CRNN+CTC 的方法，见图 4。这种方法作为目前使用最广泛的文字识别算法之一，具有高精度和快速推理速度的优点。它支持变长输入，因此适合处理文本长度变化较大的生产环境。此算法对于各种不同场景下的文字都具有很好的适用性。

通过 CRNN 算法对船舶船号文字识别有以下优点：1) 无字符长度限制：CRNN 不受字符长度的限制，只需输入包含船号文字的图片和对应的标签序列。这意味着它可以有效地处理不同长度的船号文字，无需事先知道文本的字符数目。2) 无需字符分割：传统的文字识别算法通常需要对字符进行分割，以便单独识别每个字符。但 CRNN 通过结合 CNN [15]层和 LSTM [16]层，可以直接根据输入图片得到整个标签序列，无需字符分割。这样降低了复杂性和错误的风险，特别是在字符分割不明确或字符之间有重叠的情况下。3) 模型参数量小：CRNN 模型的参数相对较小，这使得它易于部署在移动终端或资源受限的环境中。小模型不仅占用更少的内存，还可以更快地进行推理，适用于实时或实时要求较高的应用场景。总之，CRNN 算法在船舶船号文字识别中具有灵活性、鲁棒性和高效性，无需字符分割，适用于处理不同长度和包含特殊字符的文本，使其成为一个有吸引力的选择。其核心部分包括卷积层、循环层和转录层，这些层结合在一起实现了端到端的文字识别。

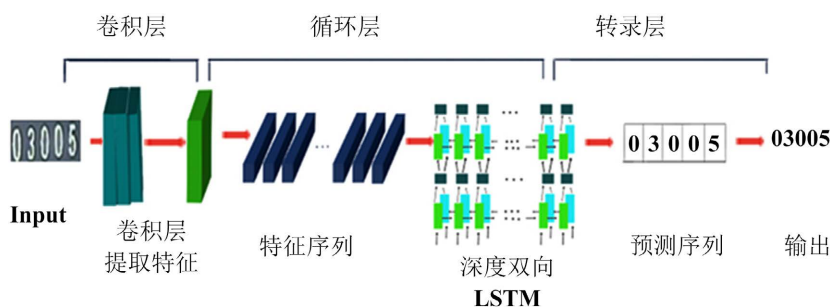


Figure 4. CRNN network architecture diagram

图 4. CRNN 网络结构图

2.2.2. SRCGAN

尽管 CRNN 算法进行船号识别存在一定的优势，但是实际应用中，船舶船号常常受到遮挡和文字模糊等问题的影响，从而导致了复杂场景下识别准确率普遍较低的情况。传统的图像处理方法和文字识别方法通常先对模糊遮挡的图像进行预处理然后对处理后的图像进行文字识别，这一方法虽然可以提高识别准确率，但其局限性在于，重构图像不一定生成适于文字识别的特征分布。针对此问题，本研究通过对 GAN 模型和文字识别器进行联合优化，提出了一种联合 GAN 和 CRNN 的针对模糊遮挡船号文字识别框架 SRC-GAN。模型框架见图 5：

SRC-GAN 模型框架由几个部分组成：一对生成器 G 、 F 以及它们各自对应的判别器 D_x 、 D_y ，还有一个 CRNN 识别器 R ， D_y 和 R 共享同一特征提取层，经由同 CNN 特征值提取层提取到的特征被分别送到鉴别器 D_y 和识别器 R 。模型的训练需要使用两个不同的数据集：一个是清晰无遮挡的船号文本图像，另一个是存在模糊遮挡的文本图像。这两种不同类型的图像数据被输入到模型中，通过训练，模型将能够生成清晰无遮挡的文本图像。生成器 G 、 D_1 以及生成器 F 、 D_2 具有相同的架构，不同之处在于它们的输入。生成器 G 从模糊遮挡的图像生成清晰无遮挡的图像，而生成器 F 则从清晰无遮挡的图像生成模糊遮挡的图像。这两个生成器需要进行双向翻译，以确保循环一致性。在每个方向上，都有一个专门的鉴别器，用来评估生成的图像的逼真程度，以确保生成更为真实的清晰无遮挡图像。

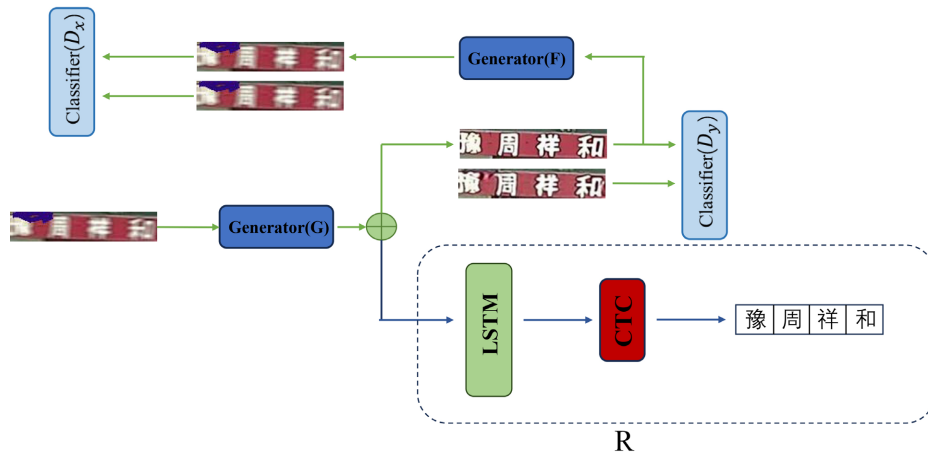


Figure 5. Text recognition framework diagram
图 5. 文字识别框架图

2.2.3. 损失函数

总体损失函数包括对抗损失、循环一致性损失、Identity 损失、识别损失，具体定义如下：

$$L_{total} = L_{GAN}(G, D_y, X, Y) + L_{GAN}(F, D_x, Y, X) + \lambda_1 L_{cyc}(G, F) + \lambda_2 L_{identity}(G, F) + \lambda_3 L_r \quad (1)$$

式中 λ_1 、 λ_2 和 λ_3 是超参数，在本实验中设为 1。

对抗损失包括二部分，分别是生成器 G 和判别器 D_y 以及生成器 F 和判别器 D_x 的对抗损失。定义如下：

$$L_{GAN}(G, D_y, X, Y) = E_{y \sim P_{data}(y)} [\log D_y(y)] + E_{x \sim P_{data}(x)} [\log(1 - D_y(G(x)))] \quad (2)$$

$$L_{GAN}(F, D_x, Y, X) = E_{x \sim P_{data}(x)} [\log D_x(x)] + E_{y \sim P_{data}(y)} [\log(1 - D_x(F(y)))] \quad (3)$$

式中： $P_{data}(x)$ 是源域 X 的分布， $P_{data}(y)$ 是目标域 Y 的分布。

循环一致性损失的目的是确保生成的图像在尽力保留源域 X 的原始内容的同时，能够接近目标域 Y 的风格，从而避免丢失了源域 X 中的重要信息。定义如下：

$$L_{cyc}(G, F) = E_{x \sim P_{data}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)} [\|G(F(y)) - y\|_1] \quad (4)$$

Identity 损失的主要作用是帮助保留输入图像的原始特征，以避免不必要的变化。定义如下：

$$L_{identity}(G, F) = E_{x \sim P_{data}(x)} [\|F(x) - x\|_1] + E_{y \sim P_{data}(y)} [\|G(y) - y\|_1] \quad (5)$$

在输入图像 I 上，表示识别损失为 L_r ，使用 CTC 损失进行文本识别。给定输出序列 l 和标签 y ，识别损失表示为：

$$L_r = -E_{I \sim P_{train}} [p(l|y; \theta_r)] \quad (6)$$

式中： θ_r 为识别器 R 的参数。

3. 实验

3.1. 数据集

实验采用了两个不同的数据集。首先，一个数据集来自长江卡口摄像头拍摄的非公开船舶数据。这

个数据集包含了 5600 张真实船只的详细图像，这些图像都是通过卡口摄像头拍摄获得，大小统一为 1920*1080 像素。每张图像都包含了有关船舶的信息，包括中文船号和数字船号。另一个数据集是 CTW 数据集，其中包含了 32,285 张高分辨率图像和 1,018,402 个中文字符实例，涵盖了 3850 种字符类别。这些图像来自腾讯街景摄像头，覆盖了中国的数十个不同城市。实验中，非公开船舶数据集被划分为训练集、验证集和测试集，比例为 8:1:1。而 CTW 数据集则按照 7:2:1 的比例划分为训练集、验证集和测试集。图 5 展示了部分样本图像，左侧为船舶数据集样本，右侧为 CTW 数据集样本。数据集样图见图 6。

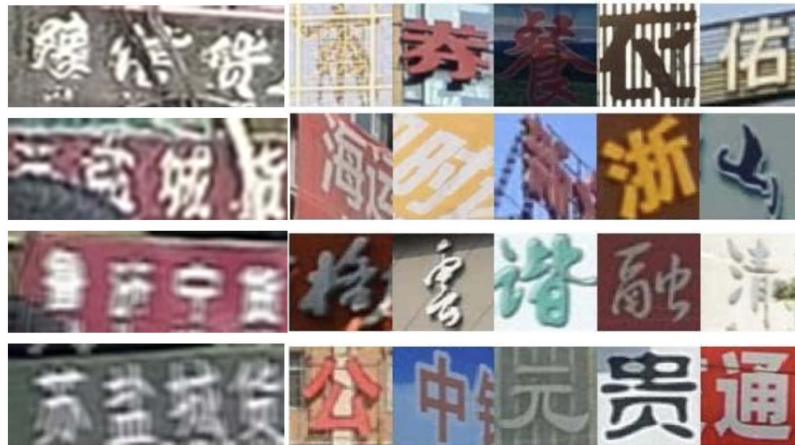


Figure 6. Sample images from ship dataset and CTW dataset
图 6. 船舶数据集和 CTW 数据集样图

3.2. 船号文字区域检测

对于 YOLOv5 模型，实验选用的模型是 YOLOv5s，其中模型训练参数设置如下：初始学习率为 0.001，权重衰减系数为 0.0005，batchsize 为 16。训练所有输入图像的尺寸均为 640 × 640。以 ADAM 优化器进行 300 次迭代训练。

为了评估所训练的模型的性能，以便进一步改进算法、提高准确率和泛化能力，本文采用了多种评估指标。这些指标包括精度(precision)、召回率(recall)、平均精度(AP)，以及图像处理时间。通过这些指标，能够在不同方面全面评估模型的性能，包括目标定位、目标检测和整体性能。精度用于评估算法的目标定位和检测能力，而召回率则用于评估模型检测目标的效能。平均精度主要用于衡量模型预测框和真实框之间的匹配程度。具体的计算表达式如下：

$$\text{precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 p(r) dr \quad (9)$$

在上式中，TP 代表正确分类的正样本数量，FP 代表错误分类的负样本数量，FN 代表错误分类的正样本数量，而 $p(r)$ 则表示 PR 曲线函数的表达式，用于评估不同召回率下的精度。PR 曲线的积分可以帮助综合评估算法在不同召回率情况下的性能表现，提供了更全面的性能指标。

如图 7 所示，利用 YOLOv5 可以准确在原图中预测出目标文字区域，得到目标文字区域后根据预测几何边界框进行裁剪和矫正后得到文字区域图像，然后进行下一步文字识别。

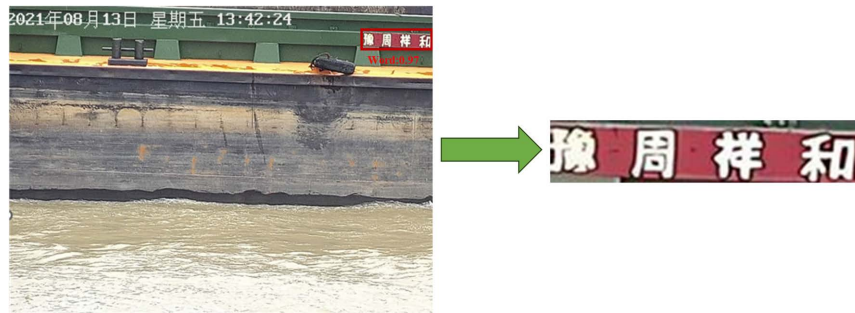


Figure 7. Example of ship number text region detection results
图 7. 船号文字区域检测结果示例

根据表 1 结果, 通过比较三种文字检测模型的文字区域检测效率(检测速度和检测精度), YOLOv5 模型在检测精度方面优于另外二种模型, 而且在检测速度上也是遥遥领先。这意味着 YOLOv5 在目标检测任务中实现了很高的准确性, 同时保持了较低的计算成本, 具有非常高的实时性能, 因此其更适合用于进行船舶船号文字区域检测。

Table 1. Performance comparison results between YOLOv5 and other models
表 1. YOLOv5 和其他模型性能对比结果

Model	AP/%	t/s
YOLOv5	95.54	0.014
SSD	91.54	0.255
CPTN	90.64	0.264

3.3. 船号文本识别

通过文字区域检测得到文字区域图像后, 就需要对文字区域进行识别。对于 SRC-GAN 模型, 训练时图片输入大小统一为 256*256, 模型选用 ADAM 优化器训练 200 轮, batchsize 设为 32, 初始学习率为 0.02, 前 100 轮学习率不变, 后 100 轮逐渐递减。

本文的评判文本识别效果主要文本识别准确率。文本识别准确率的定义如下:

$$F = \frac{P}{T} \times 100\% \quad (10)$$

式中, 文本识别率(F)表示被正确识别出来的样本数(P)与整体样本数(T)的比值, F 越大表示算法的识别效果越好。

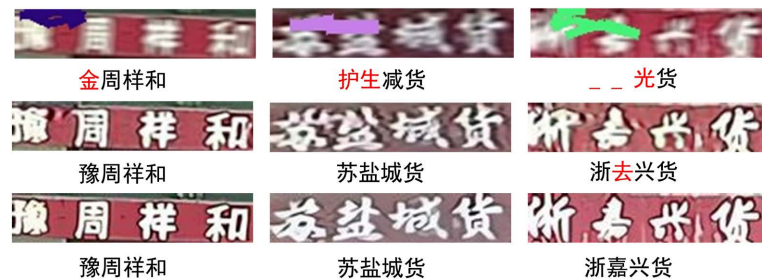


Figure 8. Comparison of recognition between the original image and SRC-GAN generated image

图 8. 原图像和 SRC-GAN 生成图像的识别对比

图 8 中第一行为存在模糊遮挡文本图像，第二行为通过 SRC-GAN 生成的文本图像，第三行为无遮挡清晰的文本图像，根据原图和恢复后的图像识别对比结果，通过 SRC-GAN 生成的图像的文字识别准确率相较于原始模糊遮挡图像得到了显著的提升。

表 2 进行了不同模糊遮挡程度下几种文字识别方法对于船舶数据集和 CTW 数据集的识别精度比较，结果表明在不同模糊遮挡程度下，SRC-GAN 对于二个数据集的识别精度相较于原始的 CRNN 得到了显著的提升，在船舶数据集上平均提升精度为 11.98%，在 CTW 数据集上平均提升精度为 10.68%。相较于另外二个识别模型的识别精度也有显著的提升，而且通过精度变化说明，对于模糊遮挡程度越高的图片，SRC-GAN 的识别效果优势越明显。同时通过比较将恢复模型和识别模型分开训练以及将二者联合训练的结果，SRC-GAN 的识别结果优于 TWO-Stage，也验证了联合训练的优越性。

Table 2. Performance comparison results between YOLOv5 and other models

表 2. YOLOv5 和其他模型性能对比结果

数据集	模糊遮挡程度	CRNN	CRNN+Seq2Seq	TWO-Stage	SRC-GAN
船舶	2	74.32%	77.45%	78.34%	80.12%
船舶	4	54.21%	55.63%	63.92%	66.52%
船舶	6	31.28%	30.78%	45.63%	49.13%
CTW	2	80.61%	81.36%	83.34%	85.90%
CTW	4	61.45%	60.49%	72.81%	76.45%
CTW	6	45.66%	44.41%	54.77%	57.41%

4. 结论

为了解决模糊遮挡问题给场景文字识别带来的挑战，同时也为了将图像恢复与文字识别联合起来，生成的图像可以有适合文字识别的特征分布。本文提出了一个联合生成对抗网络(GAN)和卷积神经网络(CRNN)的针对模糊遮挡船号文本的双分支耦合文字识别框架，称为 SRC-GAN，用于联合训练图像恢复模型和文字识别模型。通过对抗性学习将文字识别和图像恢复集成起来，学习更多图像的共性特征，从而对低质量的图像有更好的识别性能。与传统的文字识别方法相比，该方法有着显著的优势，与将图像恢复和文字识别分开进行的二阶段模型相比，该方法也有着一定的优越性。通过在船舶船号和 CTW 数据集上的识别实验表明，该方法相较于原始 CRNN 识别精度平均分布提升了 11.98%和 10.68%，验证了该方法在模糊遮挡文本图像上有着更好的识别效果。

参考文献

- [1] 陈丹涌, 陈观凤, 刘明明. 航运公司数字化安全管理体系系统设计[J]. 广州航海学院学报, 2020, 28(4): 23-27.
- [2] Liu, X.Y., Meng, G.F. and Pan, C.H. (2019) Scene Text Detection and Recognition with Advances in Deep Learning: A Survey. *International Journal on Document Analysis and Recognition (IJ DAR)*, **22**, 143-162. <https://doi.org/10.1007/s10032-019-00320-5>
- [3] Lin, H., Yang, P. and Zhang, F. (2020) Review of Scene Text Detection and Recognition. *Archives of Computational Methods in Engineering: State of the Art Reviews*, **27**, 433-454. <https://doi.org/10.1007/s11831-019-09315-1>
- [4] Aggarwal, A., Mittal, M. and Battineni, G. (2021) Generative Adversarial Network: An Overview of Theory and Applications. *International Journal of Information Management Data Insights*, **1**, Article 100004. <https://doi.org/10.1016/j.jjime.2020.100004>
- [5] Mahajan, S. and Rani, R. (2021) Text Detection and Localization in Scene Images: A Broad Review. *Artificial Intelligence Review*, **54**, 4317-4377. <https://doi.org/10.1007/s10462-021-10000-8>

-
- [6] Tian, Z., Huang, W., He, T., He, P. and Qiao, Y. (2016). Detecting Text in Natural Image with Connectionist Text Proposal Network. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, Springer, Cham. https://doi.org/10.1007/978-3-319-46484-8_4
- [7] Zhou, X.Y., *et al.* (2017) East: An Efficient and Accurate Scene Text Detector. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 5551-5560. <https://doi.org/10.48550/arXiv.1704.03155>
- [8] Wang, W., Xie, E., Li, X., *et al.* (2019) Shape Robust Text Detection with Progressive Scale Expansion Network. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Los Angeles, 16-20 June 2019, 9336-9345. <https://doi.org/10.48550/arXiv.1903.12473>
- [9] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 26 June-1 July 2016, 779-788. <https://doi.org/10.48550/arXiv.1506.02640>
- [10] Graves, A., Fernández, S., Gomez, F., *et al.* (2006) Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, 25-29 June 2006, 369-376. <https://dl.acm.org/doi/abs/10.1145/1143844.1143891>
- [11] Sutskever, I., Vinyals, O. and Le, Q.V. (2014) Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, **27**. <https://arxiv.org/abs/1409.3215>
- [12] Shi, B., Bai, X. and Yao, C. (2016) 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*, **39**, 2298-2304. <https://doi.org/10.1109/TPAMI.2016.2646371>
- [13] Graves, A., Mohamed, A. and Hinton, G. (2013) Speech Recognition with Deep Recurrent Neural Networks. 2013 *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, 26-31 May 2013, 6645-6649. <https://doi.org/10.1109/ICASSP.2013.6638947>
- [14] Cao, L., Li, H., Xie, R. and Zhu, J. (2020) A Text Detection Algorithm for Image of Student Exercises Based on CTPN and Enhanced YOLOv3, *IEEE Access*, **8**, 176924-176934. <https://doi.org/10.1109/ACCESS.2020.3025221>
- [15] Graves, A., Fernández, S. and Schmidhuber, J. (2005) Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. *International Conference on Artificial Neural Networks*, Berlin, 11-15 September 2005, 799-804. https://doi.org/10.1007/11550907_163
- [16] Yu, Y., Si, X., Hu, C., *et al.* (2019) A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, **31**, 1235-1270. https://doi.org/10.1162/neco_a_01199