

A SOLUTION FOR AUTOMATED WATER METER READING FROM IMAGES BY APPLYING DEEP LEARNING

Pham Xuan Tich*, Nguyen Dinh Duong

University of Transport and Communications

| ARTICLE INFO | | ABSTRACT |
|------------------------------|-------------------|--|
| Received: | 11/9/2023 | In this paper, we propose an automated meter reading (AMR) method applied to water meters by applying deep learning. We design a two-stage method using Rotational Region Convolutional Neural Networks (R2CNN). The first stage uses a R2CNN network to detect the digit region; the second stage applies another R2CNN network to recognize digits on an image that has only the alphanumeric region. The digits after identification are processed and sorted to obtain the counter meter. In ARM studies, most datasets are not available to the research community because the images belong to service companies. Therefore, in this study, we created a new dataset for the proposed method using it for training and testing. The result is a process with deep learning models that determine water meter readings from images of the watch face with high accuracy, and this process has been integrated into Citywork software to initially help developers. The manager audits the accuracy of the readings recorded by the employee manually to see if they match the readings in the images. |
| Revised: | 06/11/2023 | |
| Published: | 06/11/2023 | |
| KEYWORDS | | |
| Automated meter reading | | |
| Deep learning | | |
| Convolutional neural network | | |
| Recurrent neural network | | |
| Rotational Region | | |

MỘT GIẢI PHÁP TỰ ĐỘNG ĐỌC CHỈ SỐ ĐỒNG HỒ NƯỚC TỪ ẢNH ÁP DỤNG HỌC SÂU

Phạm Xuân Tích*, Nguyễn Đình Dương

Trường Đại học Giao thông vận tải

| THÔNG TIN BÀI BÁO | | TÓM TẮT |
|-------------------------|-------------------|--|
| Ngày nhận bài: | 11/9/2023 | Trong bài báo này, chúng tôi đề xuất một phương pháp tự động đọc chỉ số đồng hồ (AMR) áp dụng cho đồng hồ đo nước bằng cách áp dụng học sâu. Chúng tôi thiết kế một phương pháp hai giai đoạn sử dụng mạng nơ ron tích chập vùng quay (Rotational Region Convolutional Neural Networks - R2CNN). Giai đoạn thứ nhất sử dụng một mạng R2CNN để phát hiện vùng chứa chỉ số; giai đoạn thứ 2 áp dụng một mạng R2CNN khác để nhận dạng các chữ số trên ảnh chỉ có vùng chỉ số. Các chữ số sau khi nhận dạng được xử lý và sắp xếp để thu được chỉ số của đồng hồ. Trong các nghiên cứu về AMR, hầu hết các bộ dữ liệu đều không có sẵn cho cộng đồng nghiên cứu vì hình ảnh thuộc về các công ty dịch vụ. Do đó, trong nghiên cứu này, chúng tôi đã tạo ra một tập dữ liệu mới cho phương pháp đề xuất sử dụng nó để huấn luyện và thử nghiệm. Kết quả thu được một quy trình với các mô hình học sâu xác định chỉ số của đồng hồ đo nước từ ảnh chụp ảnh mặt đồng hồ với độ chính xác cao và quy trình này đã được tích hợp vào phần mềm Citywork bước đầu giúp các nhà quản lý kiểm toán tính chính xác của chỉ số do nhân viên ghi chỉ số đồng hồ theo phương pháp thủ công có đúng với chỉ số trong ảnh chụp của nó không. |
| Ngày hoàn thiện: | 06/11/2023 | |
| Ngày đăng: | 06/11/2023 | |
| TỪ KHÓA | | |
| Đọc đồng hồ tự động | | |
| Học sâu | | |
| Mạng nơ ron tích chập | | |
| Mạng nơ ron hồi quy | | |
| Vùng quay | | |

DOI: <https://doi.org/10.34238/tnu-jst.8741>

* Corresponding author. Email: tichpx@utc.edu.vn

1. Giới thiệu

Đọc đồng hồ tự động là việc tự động ghi lại mức tiêu thụ năng lượng điện, ga, nước,... để theo dõi và thanh toán với khách hàng sử dụng. Mặc dù đã có đồng hồ thông minh được áp dụng cho việc đọc chỉ số tự động, tuy nhiên chúng không phổ biến ở nhiều nước đang phát triển, nơi chủ yếu sử dụng đồng hồ đo cơ học, việc đọc chỉ số đồng hồ vẫn là công việc thủ công. Do đó thao tác này dễ xảy ra sai sót nên người ghi chỉ số cần chụp ảnh làm bằng chứng, trên cơ sở đó người quản lý kiểm tra tính chính xác của kết quả đọc hàng tháng. Việc kiểm tra này tốn kém về công sức và thời gian, hiệu quả thấp. Hơn nữa, do số lượng hình ảnh cần xử lý lớn nên việc kiểm tra thường được thực hiện bằng cách lấy mẫu và có thể không phát hiện ra sai sót, việc thực hiện kiểm tra đồng hồ tự động sẽ giảm thiểu sai sót do yếu tố con người và tiết kiệm nhân lực.



Hình 1. Ví dụ về hình ảnh đồng hồ nước

Đọc đồng hồ tự động là một ứng dụng nhận dạng văn bản trong khung cảnh. Ngày nay, nhận dạng văn bản trong thực tế đã nhận được nhiều sự quan tâm của nhiều nhà nghiên cứu [1], [2], nó có nhiều ứng dụng trong thực tế như nhận dạng biển số xe tự động, đọc biển báo giao thông trên xe tự hành. Có nhiều công bố của các nhóm nghiên cứu tập trung vào chủ đề này. Các phương pháp truyền thống tập trung vào phương pháp Nhận dạng ký tự quang học (Optical Character Recognition - OCR) [3], [4], phương pháp này trước tiên phân đoạn các ký tự riêng lẻ và sau đó nhận dạng từng ký tự riêng biệt. Sự đa dạng của các mẫu văn bản, ảnh mờ và nền làm tăng độ khó của việc phân đoạn ký tự. Vì vậy, hiệu suất bị giới hạn, do thiếu chính xác của giai đoạn phân đoạn ký tự. Hơn nữa, việc nhận dạng từng ký tự riêng lẻ sẽ bỏ qua mối quan hệ giữa các ký tự. Trong những năm gần đây, việc áp dụng học sâu vào nhận dạng đã đạt được nhiều thành công với nhiều ứng dụng thực tế [5], [6]. Các nghiên cứu gần đây liên quan đến nhận dạng văn bản trong khung cảnh được xem là bài toán nhận dạng chuỗi mà không cần phân đoạn. Shi và cộng sự [7] đã đề xuất mạng nơ ron tích chập hồi quy (CRNN) tích hợp mạng nơ ron tích chập (CNN) và mạng nơ ron hồi quy (RNN) để nhận dạng văn bản. Lee và cộng sự [8] đã thiết kế một mô hình tiếp cận dựa trên mạng nơ ron hồi quy chú ý cho ORC với trọng số mô hình chuỗi. Rayson Laroca và cộng sự [9] sử dụng một phiên bản thu gọn của mạng YOLO, được gọi là Fast-YOLO,

để phát hiện bộ đếm và mạng nơ ron tích chập hồi quy (CRNN) để phân đoạn chữ số. Trong [10], Concio và cộng sự phát triển một giải pháp đọc chỉ số đồng hồ nước dựa trên học sâu với 2 giai đoạn, giai đoạn thứ nhất xác định vùng chứa bằng mô hình U-Net với mạng xương sống Resnet34, giai đoạn thứ 2 xác định chỉ số bằng mạng Fast-RCNN với mạng xương sống là ResNet101. Liang và cộng sự [11] cũng phát triển một giải pháp đọc chỉ số đồng hồ nước với 2 giai đoạn, giai đoạn sử dụng Fast-RCNN để xác định vùng chứa chỉ số, giai đoạn 2 sử dụng SDD (Single Shot Multi-Box Detector) hoặc YOLOv3 để xác định chỉ số. Trong nghiên cứu này, chúng tôi đề xuất cách tiếp cận hai giai đoạn cho AMR cũng như trong [10], [11]. Giai đoạn thứ nhất thực hiện phát hiện vùng chữ số và giai đoạn thứ hai phát hiện chữ số và nhận dạng. Chúng tôi sử dụng mạng R2CNN cho cả 2 giai đoạn. R2CNN được đề xuất bởi Yingying Jiang và cộng sự trong [12]. Mạng nơ ron này được thiết kế để phát hiện các văn bản có hướng tùy ý trong khung cảnh. Chúng tôi sử dụng mạng nơ ron này để thiết kế mô hình tự động xác định chỉ số đồng hồ nước (Automated Water Meter Reading - AWMR) vì các hình ảnh đồng hồ nước thu thập được có các hướng khác nhau và rất phức tạp, xem Hình 1.

Một trong những yếu tố mang lại thành công cho các mô hình học máy là việc huấn luyện mạng với số lượng lớn dữ liệu nhằm đạt được các mô hình có tính khái quát hóa và mang lại độ chính xác cao cho những dữ liệu mới cần nhận dạng. Chúng tôi đã thu thập 3.250 hình ảnh từ một số công ty cấp nước trên địa bàn thành phố Hà Nội, chia ngẫu nhiên 80% tổng số hình ảnh cho giai đoạn huấn luyện và 20% còn lại cho giai đoạn kiểm tra mô hình. Kết quả thử nghiệm cho kết quả xác định chỉ số chính xác lên đến 92,6% trên tập kiểm tra. Giải pháp được áp dụng như một chức năng trong phần mềm Citywork (xem citywork.vn), giúp nhà quản lý tự động kiểm tra độ chính xác kết quả ghi chỉ số đồng hồ nước của nhân viên. Citywork là phần mềm do Công ty TNHH Phần mềm Citywork Việt Nam phát triển, giúp người dùng quản lý hệ thống mạng lưới cấp nước và khách hàng tiêu thụ nước.

2. Phương pháp nghiên cứu

2.1. Mạng nơ ron tích chập vùng quay (R2CNN)

Trong [12], Yingying Jiang và cộng sự đã giới thiệu một khung làm việc mới để phát hiện các văn bản có hướng tùy ý trong khung cảnh (xem Hình 2) dựa trên mạng Faster R-CNN [13]. Mạng đề xuất vùng (region proposal network - RPN) được sử dụng để đề xuất vùng văn bản và mô hình Fast R-CNN được cải tiến để thực hiện phân loại vùng văn bản, sàng lọc và dự đoán các vùng văn bản nghiêng. Bài toán phát hiện văn bản có hướng bất kỳ là bài toán đa nhiệm vụ. Nhiệm vụ mấu chốt của hướng tiếp cận này là dự đoán chấm điểm khả năng vùng đó là văn bản và độ nghiêng của hộp chứa văn bản của vùng ảnh được đề xuất bởi mạng RPN. Để tận dụng tối đa các đặc điểm dòng chữ, các tác giả đã thực hiện một số ma trận ROI Pooling với các kích thước khác nhau (7×7 , 11×3 , 3×11) cho mỗi đề xuất của RPN và sau đó kết hợp các đặc trưng khác để phát hiện thêm. Cải tiến mạng Faster R-CNN của họ bằng cách thêm một mỏ neo (anchor) nhỏ hơn để phát hiện các văn bản có kích thước nhỏ và sử dụng tính năng ức chế không tối đa (non-maximum suppression) có xu hướng để hậu xử lý các ứng cử viên nhằm đạt được kết quả cuối cùng.

Jiang sử dụng RPN để tạo ra hộp hình chữ nhật bao quanh vùng văn bản, các hình này được căn chỉnh theo trục bao quanh các văn bản có hướng tùy ý. Điều này là hợp lý vì văn bản trong hộp hình chữ nhật được căn chỉnh theo trục thuộc một trong các trường hợp sau: i) văn bản nằm theo hướng ngang; ii) văn bản theo hướng dọc; iii) văn bản theo hướng chéo của hộp.

Mạng R-CNN thực hiện nhanh hơn với ROI Pooling có bản đồ đặc trưng (feature map) kích thước 7×7 cho mỗi đề xuất từ RPN. Tuy nhiên, trên thực tế, chiều rộng của một số văn bản lớn hơn nhiều so với chiều cao của chúng, trong R2CNN các tác giả đã cố gắng sử dụng ba ROI Pooling có các kích thước khác nhau nhằm bắt được nhiều đặc trưng của văn bản hơn. Cụ thể, họ đã thêm hai bản đồ đặc trưng với kích thước: 11×3 và 3×11 . Kích thước 3×11 được cho là sẽ bắt được nhiều đặc trưng ngang hơn và giúp phát hiện các văn bản ngang có chiều rộng lớn

hơn nhiều so với chiều cao của nó. Kích thước 11×3 được cho là sẽ thu được nhiều đặc điểm đọc hơn và hữu ích cho việc phát hiện các văn bản đọc có chiều cao lớn hơn nhiều so với chiều rộng.



Hình 2. Ví dụ kết quả xác định các hình chữ nhật bao quanh vùng chứa văn bản của mạng R2CNN [12]

Theo cách tiếp cận của Jiang, sau RPN, họ phân loại đề xuất do RPN tạo ra thành hai loại văn bản hoặc phi văn bản được giới hạn trong các hình chữ nhật căn chỉnh theo trục chứa văn bản được định hướng tùy ý và dự đoán các hộp hình chữ nhật nghiêng. Mặc dù mục tiêu phát hiện của chúng là các hộp hình chữ nhật nghiêng, nhưng họ cho rằng việc thêm các ràng buộc bổ sung các hộp hình chữ nhật có cạnh song song với các trục tọa độ có thể cải thiện hiệu suất.

Kỹ thuật non-max suppression được áp dụng để loại bỏ những hình chữ nhật bị trùng, chồng lên nhau của cùng một đối tượng bằng cách chọn chiếc hộp có tính đặc trưng nhất. Sau khi loại bỏ tất cả các hộp có xác suất dự đoán nhỏ hơn 0,6, những bước tiếp theo được lặp lại khi vẫn còn tồn tại những hộp khác.

Hàm mất mát của quá trình huấn luyện trên RPN tương tự như Faster R-CNN [13]. Hàm mất mát được xác định trên mỗi đề xuất là tổng mất mát của việc phân loại văn bản/phi văn bản mất mát hồi quy trên các hộp. Mất mát hồi quy trên các hộp bao gồm hai phần: mất trên các hộp được căn chỉnh theo trục bao quanh các văn bản có hướng tùy ý và mất các hộp nghiêng có diện tích tối thiểu. Hàm mất trên mỗi đề xuất được định nghĩa là:

$$L(p, c, v, v^*, u, u^*) = L_{cls}(p, c) + \lambda_1 c \sum_{i \in \{x, y, w, h\}} L_{reg}(v_i, v_i^*) + \lambda_2 c \sum_{i \in \{x1, y1, x2, y2, h\}} L_{reg}(u_i, u_i^*) \quad (1)$$

λ_1 và λ_2 là các tham số kiểm soát cân bằng giữa ba thông số ở trên. Dự đoán các hộp chỉ tiến hành trên văn bản, c là chỉ báo của nhãn lớp. Văn bản được gán nhãn là 1 ($c = 1$) và nền được gán nhãn là 0 ($c = 0$). Tham số $p = (p_0, p_1)$ là xác suất trên các lớp văn bản và nền được tính toán bởi hàm softmax. $L_{cls}(p, t) = -\log p_c$ là log của giá trị mất mát cho phân lớp đúng c . $v = (v_x, v_y, v_w, v_h)$ là một bộ giá trị xác định giới hạn của hình chữ nhật đúng được đóng theo các trục tọa độ, bao gồm tọa độ của điểm trung tâm cũng như chiều rộng và chiều cao của điểm đó, và $v^* = (v_x^*, v_y^*, v_w^*, v_h^*)$ là bộ giá trị được dự đoán cho nhãn văn bản.

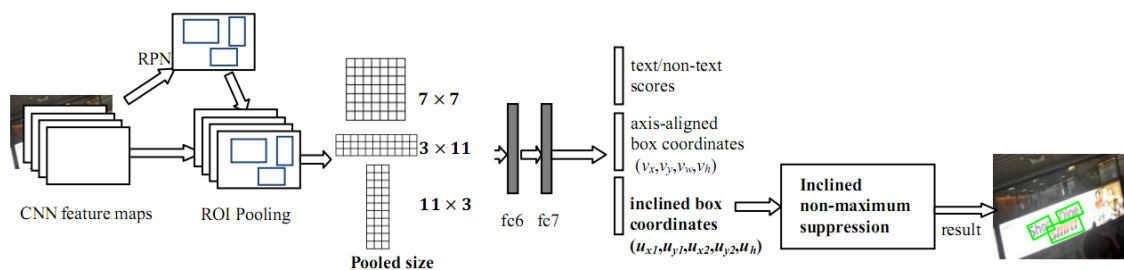
$\mathbf{u} = (\mathbf{u}_{x1}, \mathbf{u}_{y1}, \mathbf{u}_{x2}, \mathbf{u}_{y2}, \mathbf{u}_h)$ là bộ giá trị gồm các giá trị xác định hình giới hạn của hình chữ nhật nghiêng thực sự bao quanh vùng chữ gồm tọa độ của hai đỉnh đối diện của hộp nghiêng và chiều cao, $\mathbf{u}^* = (\mathbf{u}_{x1}^*, \mathbf{u}_{y1}^*, \mathbf{u}_{x2}^*, \mathbf{u}_{y2}^*, \mathbf{u}_h^*)$ là bộ giá trị được dự đoán cho nhãn văn bản. Tham số hóa cho v và v^* được đưa ra trong [14], trong đó v và v^* xác định tỉ lệ chuyển đổi bất biến và sự thay đổi chiều cao/chiều rộng của không gian log liên quan đến đề xuất đối tượng. Đối với các hộp nghiêng, tham số hóa của $(\mathbf{u}_{x1}, \mathbf{u}_{y1})$, $(\mathbf{u}_{x2}, \mathbf{u}_{y2})$, $(\mathbf{u}_{x1}^*, \mathbf{u}_{y1}^*)$ và $(\mathbf{u}_{x2}^*, \mathbf{u}_{y2}^*)$ giống với v_x , v_x . Và cách tham số hóa của \mathbf{u}_h và \mathbf{u}_h^* cũng giống với cách tham số hóa của v_h và v_h^* .

Đặt $(\mathbf{w}, \mathbf{w}^*)$ biểu thị (v_i, v_i^*) hoặc $(\mathbf{u}_i, \mathbf{u}_i^*)$, $L_{reg}(\mathbf{w}, \mathbf{w}^*)$ được định nghĩa là:

$$L_{reg}(\mathbf{w}, \mathbf{w}^*) = smooth_{L1}(\mathbf{w} - \mathbf{w}^*) \tag{2}$$

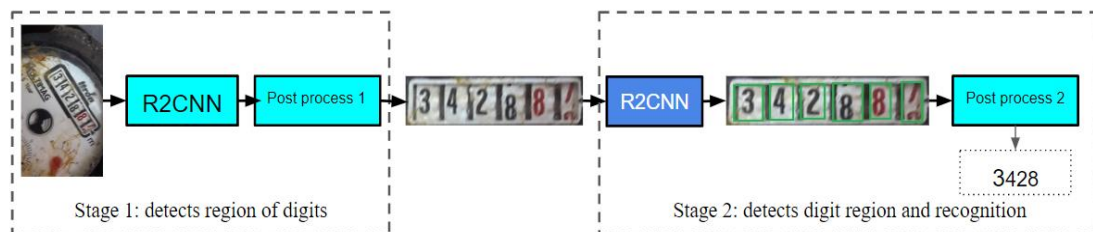
$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{Trường hợp ngược lại} \end{cases} \tag{3}$$

2.2. Giải pháp đề xuất



Hình 3. Kiến trúc mạng R2CNN [12]

Như hình 3, hình ảnh đồng hồ nước được đưa vào mô hình R2CNN đầu tiên, mô hình sẽ dò tìm vùng chữ số. Sau khi xác định được vùng chữ số chúng tôi thực hiện bước xử lý tiếp theo là cách xoay ngang để vùng chữ số song song với trục hoành nếu nó đang bị xiên, góc quay được tính toán dựa trên việc phân tích các chữ số ngược hoặc xuôi và góc nghiêng của hình chữ nhật bao quanh nó. Tiếp theo, hình ảnh vùng chữ số đã được cắt ra làm ảnh đầu vào của mạng R2CNN thứ hai thực hiện phát hiện vùng chứa từng chữ số và nhận dạng. Các chữ số được nhận dạng được sắp xếp từ trái qua phải theo trục ox để được chỉ số công tơ. Trong thực tế các chữ số trên các đồng hồ đo nước có màu đen và màu đỏ. Các chữ số màu đỏ đo phần lẻ vì vậy trong quá trình đo người ta không quan tâm đến số màu đỏ nên sau khi xác định được dãy số trên đồng hồ chúng tôi thực hiện một quá trình xử lý tiếp theo là loại bỏ các số màu đỏ bằng phân tích màu và chỉ giữ lại số màu đen cho kết quả cuối cùng. Hình 4 dưới đây mô tả quá trình thực hiện đọc chỉ số đồng hồ đo nước từ ảnh chụp.



Hình 4. Quy trình đọc đồng hồ nước tự động từ hình ảnh

3. Kết quả thử nghiệm và bàn luận

Hiện nay trong cộng đồng nghiên cứu chưa có bộ dữ liệu hình ảnh đồng hồ nước xuất bản công cộng, để xây dựng mô hình và đánh giá hiệu quả của giải pháp đề xuất, chúng tôi thu thập

hình ảnh đồng hồ nước bằng điện thoại di động của nhân viên ghi chỉ số nước. Chúng tôi đã thu thập được một bộ dữ liệu đa dạng và phức tạp về hình ảnh mặt đồng hồ đo nước bao gồm 3640 hình ảnh với 6 chữ số do một số công ty cung cấp nước sạch tại Hà Nội chụp lại. Những hình ảnh này rất đa dạng về góc xoay của vùng chỉ số bao gồm hình ảnh nằm ngang, dọc hoặc quay nghiêng, nhiều ảnh có chất lượng khác nhau về độ sáng của ảnh, lộn xộn, môi trường phức tạp, đồng hồ mới quay được nửa chữ số và bị mờ bởi bụi bẩn hoặc hơi nước,....

Việc xây dựng tập dữ liệu huấn luyện và kiểm tra bằng phần mềm labelme, chúng tôi tiến hành khoanh vùng chứa chỉ số và vùng của từng chữ số và gán nhãn bằng phần mềm labelme với 21 lớp gồm cả nửa chữ số (là một chữ số đang thay đổi, ví dụ 1=>2, 2=>3,...) và vùng các chữ số, chi tiết bảng 1.

Bảng 1. Gán nhãn cho dữ liệu

| Đối tượng chữ số/ vùng chỉ số | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Vùng chỉ số |
|----------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------------|
| Nhãn | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | CS |
| Đối tượng chữ số | 0->1 | 1->2 | 2->3 | 3->4 | 4->5 | 5->6 | 6->7 | 7->8 | 8->9 | 9->0 | |
| Nhãn | 01 | 12 | 23 | 34 | 45 | 56 | 67 | 78 | 89 | 90 | |

Để xây dựng và thử nghiệm mô hình, chúng tôi chia ngẫu nhiên 3640 hình ảnh đã được khoanh vùng và gán nhãn thành hai phần: 3000 hình ảnh làm tập huấn luyện (xấp xỉ 80%) và 640 hình ảnh làm tập kiểm tra. Các tham số của R2CNN được thiết lập tương tự trong [12], tuy nhiên chúng tôi thay đổi một số thông số để phù hợp với số liệu bài toán, thể hiện trong bảng 2. Kết quả thực nghiệm được trình bày ở bảng 3.

Bảng 2. Một số tham số của mạng R2CNN

| Parameter name | Value | |
|--------------------|---------|---------|
| | Stage 1 | Stage 2 |
| IMG_SHORT_SIDE_LEN | 1.000 | 100 |
| IMG_MAX_LENGTH | 1.500 | 400 |
| MAX_ITERATION | 500.000 | 500.000 |

Bảng 3. Kết quả thử nghiệm trên tập huấn luyện và tập kiểm tra

| Tập dữ liệu | Số ảnh | Số hình ảnh được đọc chính xác | Độ chính xác (%) |
|----------------|--------|--------------------------------|------------------|
| Tập huấn luyện | 3.000 | 2.852 | 95% |
| Tập kiểm tra | 640 | 593 | 92,6% |

Từ kết quả được thống kê sau quá trình thử nghiệm chúng tôi nhận thấy phần lớn lỗi nhận dạng xuất hiện ở trường hợp nửa chữ số, hình ảnh quá lộn xộn hoặc mờ. Từ kết quả thống kê trong bảng 2 cho thấy mô hình đề xuất cho kết quả khá chính xác với độ chính xác 92,6% trên tập kiểm tra. Do đề xuất trong bài báo này và trong các bài [10], [11] không thử nghiệm trên cùng một tập dữ liệu nên chúng tôi không thể so sánh trực tiếp kết quả thử nghiệm với các kết quả thử nghiệm của các giải pháp được đề xuất trong [10] và [11]. Tuy nhiên chúng ta có thể so sánh một cách tương đối giữa kết quả đạt được của các giải pháp. Giải pháp đề xuất trong [10] kết quả thử nghiệm đạt độ chính xác 91,5% với đồng hồ nhập ngoại và 75% với đồng hồ sản xuất tại Philipine, như vậy trung bình độ chính xác khoảng 83,25%. Giải pháp trong [11] kết quả thử nghiệm đạt độ chính xác 90,61% với giải pháp sử dụng mạng YOLOv3 cho pha 2 và 80,42% với giải pháp sử dụng SSD cho pha 2. Như vậy trong kết quả thử nghiệm của các mô hình thì giải pháp của chúng tôi cho kết quả thử nghiệm cao nhất.

Với kết quả này mô hình thu được chúng tôi đã tích hợp vào phần mềm Citywork bước đầu hỗ trợ khách hàng kiểm toán chất lượng chỉ số đồng hồ được ghi bởi các nhân viên của công ty.

4. Kết luận

Trong bài báo này, chúng tôi đề xuất phương pháp đọc chỉ số đồng hồ nước tự động, bao gồm hai quá trình phát hiện vùng chỉ số và nhận dạng chữ số. Chúng tôi sử dụng mạng R2CNN cho cả nhiệm vụ phát hiện vùng ảnh chứa chỉ số đồng hồ và nhận dạng các chữ số. Áp dụng mạng R2NN cho phép trích xuất vùng chứa chỉ số một cách chính xác và hiệu quả. Phương pháp đề xuất đã được huấn luyện và thử nghiệm trên tập dữ liệu do chúng tôi tự thu thập và xây dựng với 3000 hình ảnh được sử dụng để huấn luyện mô hình và 640 hình ảnh phục vụ kiểm tra mô hình. Kết quả thử nghiệm cho thấy mô hình đề xuất của chúng tôi cho kết quả với độ chính xác cao với độ chính xác 95% và 92,6% lần lượt trên tập huấn luyện và kiểm tra. Giải pháp đề xuất đã được áp dụng trong phần mềm Citywork như một chức năng AI hỗ trợ nhà quản lý kiểm tra tính chính xác của dữ liệu do nhân viên đi đọc và ghi nhận vào hệ thống. Mặc dù kết quả đọc chỉ số của mô hình đạt được có độ chính xác khá cao tuy nhiên vẫn chưa thể áp dụng vào việc đọc tự động trên các thiết bị smart phone do thực tế đòi hỏi mô hình phải có độ chính xác cao hơn.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Giao thông vận tải trong đề tài mã số T2023-CN-001.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] L. Neumann and J. Matas, "Real-time scene text localization and recognition," In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3538–3545.
- [2] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," *Computer Science*, June 2014, doi: 10.48550/arXiv.1406.2227.
- [3] Wang, T., Wu, D.J., Coates, A. and Ng, A.Y. "End-to-end text recognition with convolutional neural networks," *21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 3304–3308.
- [4] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "Photoocr: Reading text in uncontrolled conditions," *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 785–792.
- [5] T. D. Le, D. T. Nguyen, and Q. B. Truong, "Identification of some types of longan (through leaves) using image and deep learning technology," *TNU Journal of Science and Technology*, vol. 228, no. 02, pp. 128 – 135, 2023.
- [6] Q. T. Nguyen, Q. U. Nguyen, K. P. Phung, M. T. Nguyen, and M. S. Nguyen, "Detecting and measuring environmental disasters based on image segmentation deep learning technique," *TNU Journal of Science and Technology*, vol. 227, no. 16, pp. 140 – 148, 2022.
- [7] 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," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2298-2304, Nov. 2017.
- [8] C.-Y. Lee and S. Osindero "Recursive Recurrent Nets with Attention Modeling for OCR in the Wild," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2231–2239.
- [9] R. Laroca, V. Barroso, M. A. Diniz, G. R. Goncalves, W. R. Schwartz, and D. Menotti "Convolutional Neural Networks for Automatic Meter Reading," *Journal of Electronic Imaging*, vol. 28, no. 01, pp. 1-14, 2019, doi: 10.1117/1.JEI.28.1.013023.
- [10] M. L. W. Concio, F. S. Bernardo, J. M. Oplencia, G. L. Ortiz, and J. R. I. Pedrasa "Automated Water Meter Reading Through Image Recognition," *TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON)*, 01-04 November 2022, doi: 10.1109/TENCON55691.2022.9977678.
- [11] Y. Liang, Y. Liao, S. Li, W. Wu, T. Qiu, and W. Zhang "Research on water meter reading recognition based on deep learning," *Scientific Reports*, vol. 12, 2022, Art. no. 12861, doi: 10.1038/s41598-022-17255-3.
- [12] Y. Jiang, X. Zhu, X. Wang, S. Yang, W. Li, H. Wang, P. Fu, and Z. Luo, "R2CNN: Rotational Region CNN for Orientation Robust Scene Text Detection," *Computer Science*, June 2017, doi: 10.48550/arXiv.1706.09579.
- [13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015, doi: 10.48550/arXiv.1506.01497.
- [14] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, doi: 10.48550/arXiv.1311.2524.