

# **Nghiên cứu ứng dụng thuật toán nhận dạng cấu trúc bảng dựa trên phát hiện đối tượng**

**Ngô Đại Dương<sup>1</sup>, Hải Hồng Phan<sup>1\*</sup>, Phạm Lê Phương<sup>2</sup>**

<sup>1</sup> Học Viện Kỹ thuật Quân Sự; hongpth@lqdtu.edu.vn; daiduong28789@hotmail.com

<sup>2</sup> Trung tâm Thông tin và Dữ liệu KTTV; phuongpl80@gmail.com

\*Tác giả liên hệ: hongpth@lqdtu.edu.vn; Tel: +84-372576968

Ban Biên tập nhận bài: 05/3/2021; Ngày phản biện xong: 10/5/2021; Ngày đăng bài: 25/7/2021

**Tóm tắt:** Nhận dạng cấu trúc bảng là vấn đề quan trọng trong số hóa tài liệu. Với sự phát triển của các kỹ thuật học sâu hiện nay việc phát hiện bảng đã có nhiều bước tiến lớn, trong khi đó nhận dạng cấu trúc bảng vẫn gặp rất nhiều khó khăn do cấu trúc bảng phức tạp, đặc biệt với dữ liệu thực tế. Bài báo này đề xuất một phương pháp ứng dụng mô hình Cascade mask R-CNN x101FPN deconv để nhận dạng hàng và cột. Bài báo cũng đề xuất sử dụng mô hình Faster R-CNN để nhận dạng các ô trong bảng, từ đó đưa ra cấu trúc bảng. Thuật toán đề xuất được đánh giá trên tập dữ liệu phổ biến như TabStructDB và các tài liệu thu thập được từ các trạm Khí tượng thủy văn. Kết quả thực nghiệm đạt 90% độ chính xác trên các tập dữ liệu này. Thuật toán có khả năng áp dụng hiệu quả vào việc nhận dạng cấu trúc bảng của các tài liệu thông thường; đặc biệt thuật toán có khả năng xử lý với các tài liệu lịch sử và các chữ viết tay, phù hợp với đặc điểm tài liệu của ngành Khí tượng thủy văn. Từ đó góp phần vào việc số hóa tài liệu, lưu trữ và truy xuất thông tin dữ liệu của ngành Khí tượng thủy văn.

**Từ khóa:** Nhận dạng cấu trúc bảng; Nhận dạng ô; Khí tượng thủy văn.

## **1. Mở đầu**

Hiện nay, chuyển đổi số là một trong những mục tiêu phát triển của quốc gia vì lợi ích mà nó mang lại cho sự phát triển kinh tế, xã hội. Một trong các nhiệm vụ quan trọng của chuyển đổi số là số hóa tài liệu. Các tài liệu sổ sách được scan hoặc chụp ảnh rồi lưu vào máy tính dưới dạng định dạng ảnh. Tuy nhiên, để khai thác dữ liệu hiệu quả hơn cần số hóa các file này thành các dạng văn bản. Tài liệu được chia thành nhiều vùng như: vùng đoạn văn bản, vùng tiêu đề, vùng ảnh, vùng bảng... Trong đó, vùng bảng thường chứa nhiều thông tin, nhất là với các tài liệu sổ sách, kỹ thuật. Do đó, trích rút thông tin bảng trong tài liệu là một khâu quan trọng, quyết định trong số hóa tài liệu.

Cấu trúc bảng trong tài liệu rất đa dạng với nhiều kích thước khác nhau và nhiều loại khác nhau (như bảng có viền, bảng ít viền, bảng không viền). Ngoài ra bảng có thể từ các tài liệu mới, cũ khác nhau và có nhiều loại cấu trúc rất phức tạp. Nhận dạng cấu trúc bảng là xác định các ô (cell), hàng, cột và mối quan hệ phân cấp giữa các ô. Đây là một bài toán rất phức tạp và hiện nay vẫn chưa có một giải pháp tổng thể nào giải quyết được toàn bộ các dữ liệu bảng trong thực tế và kết quả nhận dạng chưa cao.

Một số nghiên cứu tiêu biểu về lĩnh vực này có thể kể tới như: Năm 1997, nghiên cứu [1] lần đầu đề xuất một phương pháp trích rút bảng dựa trên cấu trúc dữ liệu được gọi là đồ thị căn chỉnh ký tự (CAG-Character Alignment Graph), một đồ thị căn chỉnh ký tự được hình thành bằng cách kiểm tra sự liên kết khoảng trắng trong các khối văn bản liền kề. Một số

tham số khác như lỗ (số khoảng trống giữa các cột) và khoảng trống (số khoảng trống liền kề như vậy trên mỗi dòng) được sử dụng để kiểm tra cấu trúc khoảng trống trong một khối dữ liệu. Cùng với sự phát triển của các kỹ thuật học sâu, năm 2017, nghiên cứu [2] đã đề xuất một hệ thống từ đầu đến cuối để hiểu bảng trong tài liệu được gọi là DeepDeSRT. Đầu tiên, nghiên cứu sử dụng mô hình mạng Faster-RCNN [3] cho nhiệm vụ phát hiện bảng, tiếp theo sử dụng mô hình phân đoạn ngữ nghĩa FCN [4] thực hiện nhiệm vụ nhận dạng hàng và cột ở trong bảng. Nghiên cứu thực hiện huấn luyện trên tập dữ liệu Marmot [5], và đánh giá trên tập dữ liệu ICDAR 2013 [6]. Kết quả đạt được  $F_1$  là 96,67% cho nhiệm vụ phát hiện bảng và 91,44% cho nhiệm vụ nhận dạng cấu trúc bảng. Năm 2019, nghiên cứu [7] đã đề xuất một mô hình học sâu tại hội nghị quốc tế về phân tích và nhận dạng tài liệu (ICDAR) dựa vào các phép tích chập biến dạng để nhận dạng hàng và cột của bảng, từ đó có thể xác định các ô và cấu trúc trong bảng.

Nghiên cứu [8] đã đề xuất một mô hình học sâu CascadeTabNet sử dụng để phát hiện và nhận dạng cấu trúc bảng từ đầu đến cuối. Các thử nghiệm của nghiên cứu được thực hiện trên các tập dữ liệu ICDAR 2013, ICDAR 2019 [9], Tablebank [10].

Các phương pháp kể trên phần lớn giải quyết trên ảnh tài liệu xây dựng cho các cuộc thi, chưa xử lý và chứng minh tính hiệu quả với dữ liệu thực tế. Hai là, các dữ liệu thường có một loại lịch sử hoặc hiện đại. Trong khi dữ liệu cần số hóa thực tế có thể gồm cả hai. Dữ liệu của các trạm Khí tượng thủy văn là một trong các dữ liệu như vậy, nó bao gồm cả các tài liệu mới và cũ. Bài báo này đề xuất một phương pháp dựa trên các kỹ thuật phát hiện đối tượng theo hướng học sâu để giải quyết bài toán thực tế. Phương pháp đề xuất được đánh giá trên tập dữ liệu phổ biến TabStructDB mở rộng [1] cho nhiệm vụ nhận dạng hàng và cột, đánh giá nhận dạng các ô trong cột trên tập dữ liệu DetectCell tự xây dựng và tập dữ liệu Khí tượng thủy văn.

Mục đích của nghiên cứu nhằm: (1) đề xuất thuật toán nhận dạng cấu trúc bảng dựa trên mô hình Cascade mask R-CNN x101FPN deconv cho nhiệm vụ nhận dạng hàng và cột và Faster R-CNN cho nhiệm vụ nhận dạng các ô từ các cột ở trong bảng; (2) phương pháp đề xuất được chứng minh hiệu quả trên các tập dữ liệu, đặc biệt dữ liệu thực tế là các tài liệu Khí tượng thủy văn.

## **2. Phương pháp nghiên cứu**

### *2.1. Dữ liệu nghiên cứu*

#### **2.1.1. Tập dữ liệu TabStructDB mở rộng và DetectCell**

Với nhiệm vụ nhận dạng cấu trúc bảng, bài báo sử dụng dữ liệu TabStructDB mở rộng để tăng cường dữ liệu. Dữ liệu này được gán nhãn cho từng hàng và từng cột của bảng. Tổng cộng bao gồm có 2079 ảnh (ban đầu có 1081 ảnh) trong đó có tăng cường thêm 868 ảnh từ tập dữ liệu Tablebank [10] và 130 ảnh của tập dữ liệu quan trắc KTTV. Dữ liệu này được sử dụng để nhận dạng hàng và cột cho bảng. Với nhiệm vụ nhận dạng ô (cell), bài báo đề xuất tập dữ liệu DetectCell với 1.172 ảnh. Mỗi ảnh gán nhãn cho từng ô, dữ liệu gồm cả tập dữ liệu mở rộng và tập dữ liệu quan trắc KTTV.

#### **2.1.2. Tập dữ liệu quan trắc KTTV**

Tập dữ liệu Khí tượng thủy văn thu thập bao gồm các sổ sách, bảng biểu từ nhiều năm trước, nó bao gồm cả dữ liệu hiện đại và lịch sử. Có 13 loại sổ sách kỹ thuật khác nhau: sổ quan trắc khí tượng cơ bản, sổ quan trắc mực nước, sổ dùng khi đo độ sâu, tổng bức xạ giờ, sổ quan trắc bốc hơi GGI-3000..., mỗi loại sổ sách thường có các bảng thông thường và các bảng đặc trưng cho từng loại sổ.

n	Meson	L	S	J	$J^{PC}$	our calculated $\sqrt{\langle r^2 \rangle}$	Theor. $\sqrt{\langle r^2 \rangle}$ with potential model [?]	$ R(0) ^2$	Người đo
						fm	fm	GeV <sup>3</sup>	
2 D	$\eta_{c2}(2^1D_2)$	2	0	2	2 <sup>+-</sup>	1.2595	—	≈ 0	Thủy
	$\psi(2^3D_1)$	2	1	1	1 <sup>+-</sup>	1.2112	—	≈ 0	Thủy
	$\psi_2(2^3D_2)$	2	1	2	2 <sup>+-</sup>	1.2556	—	≈ 0	Thủy
	$\psi_3(2^3D_3)$	2	1	3	3 <sup>+-</sup>	1.2793	—	≈ 0	Thủy
3 D	$\eta_{c2}(3^1D_2)$	2	0	2	2 <sup>+-</sup>	1.57397	—	≈ 0	Thủy
	$\psi(3^3D_1)$	2	1	1	1 <sup>+-</sup>	1.52397	—	≈ 0	Thủy
	$\psi_2(3^3D_2)$	2	1	2	2 <sup>+-</sup>	1.5706	—	≈ 0	Thủy
	$\psi_3(3^3D_3)$	2	1	3	3 <sup>+-</sup>	1.59398	—	≈ 0	Thủy
4 D	$\eta_{c2}(4^1D_2)$	2	0	2	2 <sup>+-</sup>	1.8596	—	≈ 0	Thủy
	$\psi(4^3D_1)$	2	1	1	1 <sup>+-</sup>	1.80785	—	≈ 0	Thủy
	$\psi_2(4^3D_2)$	2	1	2	2 <sup>+-</sup>	1.8565	—	≈ 0	Thủy
	$\psi_3(4^3D_3)$	2	1	3	3 <sup>+-</sup>	1.8798	—	≈ 0	Thủy
5 D	$\eta_{c2}(5^1D_2)$	2	0	2	2 <sup>+-</sup>	2.1248	—	≈ 0	Thủy
	$\psi(5^3D_1)$	2	1	1	1 <sup>+-</sup>	2.0714	—	≈ 0	Thủy
	$\psi_2(5^3D_2)$	2	1	2	2 <sup>+-</sup>	2.1218	—	≈ 0	Thủy
	$\psi_3(5^3D_3)$	2	1	3	3 <sup>+-</sup>	2.1452	—	≈ 0	Thủy
1 F	$h_{c3}(1^1F_3)$	3	0	3	3 <sup>+-</sup>	1.0878	—	≈ 0	Thủy
	$\chi_2(1^3F_2)$	3	1	2	2 <sup>+-</sup>	1.06896	—	≈ 0	Thủy
	$\chi_3(1^3F_3)$	3	1	3	3 <sup>+-</sup>	1.08628	—	≈ 0	Thủy
	$\chi_4(1^3F_4)$	3	1	4	4 <sup>+-</sup>	1.09818	—	≈ 0	Thủy
2 F	$h_{c3}(2^1F_3)$	3	0	3	3 <sup>+-</sup>	1.4253	—	≈ 0	Thủy
	$\chi_2(2^3F_2)$	3	1	2	2 <sup>+-</sup>	1.4062	—	≈ 0	Thủy
	$\chi_3(2^3F_3)$	3	1	3	3 <sup>+-</sup>	1.42387	—	≈ 0	Thủy
	$\chi_4(2^3F_4)$	3	1	4	4 <sup>+-</sup>	1.43578	—	≈ 0	Thủy
3 F	$h_{c3}(3^1F_3)$	3	0	3	3 <sup>+-</sup>	1.725	—	≈ 0	Thủy
	$\chi_2(3^3F_2)$	3	1	2	2 <sup>+-</sup>	1.7056	—	≈ 0	Thủy
	$\chi_3(3^3F_3)$	3	1	3	3 <sup>+-</sup>	1.72369	—	≈ 0	Thủy
	$\chi_4(3^3F_4)$	3	1	4	4 <sup>+-</sup>	1.73579	—	≈ 0	Thủy
4 F	$h_{c3}(4^1F_3)$	3	0	3	3 <sup>+-</sup>	2.0	—	≈ 0	Thủy
	$\chi_2(4^3F_2)$	3	1	2	2 <sup>+-</sup>	1.9802	—	≈ 0	Thủy
	$\chi_3(4^3F_3)$	3	1	3	3 <sup>+-</sup>	1.9988	—	≈ 0	Thủy
	$\chi_4(4^3F_4)$	3	1	4	4 <sup>+-</sup>	2.0111	—	≈ 0	Thủy
1 G	$\eta_{c4}(1^1G_4)$	4	0	4	4 <sup>+-</sup>	1.2589	—	≈ 0	Thủy
	$\psi_3(1^3G_3)$	4	1	3	3 <sup>+-</sup>	1.25026	—	≈ 0	Thủy
	$\psi_4(1^3G_4)$	4	1	4	4 <sup>+-</sup>	1.25878	(a)	≈ 0	Thủy
	$\psi_5(1^3G_5)$	4	1	5	5 <sup>+-</sup>	1.26409	—	≈ 0	Thủy

**Hình 1.** Ảnh của tập dữ liệu TabStrucDB và tập dữ liệu DetectCell (a) bảng có cấu trúc các hàng sát nhau của tập dữ liệu TabStrucDB (b) Một cột của bảng đo lưu lượng bằng máy ADCP tập dữ liệu DetectCell.

**Bảng 1.** Danh sách các số của dữ liệu KTTV.

STT	Tên số	Đặc điểm	Số mẫu
1	SKT-1	Số quan trắc khí tượng cơ bản	642
2	SKT-2	Số quan trắc khí tượng	7
3	SKT-3	Số quan trắc nhiệt độ đất	7
4	SKT-13a	Số quan trắc bốc hơi GGI-3000	7
5	BKT11	Tổng lượng bức xạ giờ	8
6	BKT11a	Cường độ bức xạ giờ	8
7	ADCP	Số ghi chép máy ADCP	6
8	Tài liệu chỉnh biên	Tài liệu chỉnh biên	41
9		Số dùng khi đo độ sâu	16
10		Số ghi quan trắc mực nước	19
11		Số ghi đo lưu lượng nước số 1	11
12		Biểu XL mẫu nước chất lơ lửng	7
13		Biểu ghi tốc độ và tính chất lưu lượng triều	5
<b>Tổng</b>			<b>784</b>

Với dữ liệu ban đầu, có thể phân tập dữ liệu KTTV là những bảng cũ của sổ quan trắc khí tượng có từ năm 1969 đến năm 1985, cũng có những bảng của sổ quan trắc khí tượng cơ bản từ năm 2000 trở về đây. Hình 2 mô tả hai loại bảng trong nhiều loại bảng của tập dữ liệu KTTV. Nghiên cứu cho thấy dữ liệu trong bảng, ngoài là các mực in, còn có các chữ viết tay, dành cho hai loại bảng là lịch sử và hiện đại. Ngoài ra, còn một số bảng kẻ tay cũng được xử lý trong bài báo này. Bước đầu nghiên cứu này sử dụng 130 ảnh đưa vào tập TabStructDB để tiến hành đánh giá thuật toán.

Độ sâu (cm)	5	10	15	20	Tên QTV
	Số đọc	Số đọc	Số đọc	Số đọc	
Giờ					
01	254	215	275	281	Hiện
07	254	259	266	270	Hiện
13	325	310	300	295	Giảng
19	280	290	300	305	Giảng
Tổng số	1410	1174	1441	1511	
TB	118	281	285	288	
Độ sâu					
13 <sup>h</sup>		(a)			

Cấp mực nước	Số ngày xuất hiện trong tháng												Số ngày xuất hiện trong năm	Tổng đến
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII		
Hmax, 5-12cm														
531 - 520										1			1	1
519 - 500														
799 - 780										1			1	2
779 - 760														
759 - 740														
739 - 720														
719 - 700														
699 - 680								1					1	3
679 - 660								1					1	4
659 - 640								1					1	5
639 - 620										1			1	6
619 - 600														7
599 - 580								1		1			2	9
579 - 560								2		1			3	12
559 - 540								1	2	1			4	16
539 - 520								1		1			2	18
519 - 500								2	2				4	22
499 - 480								1		1			2	24
479 - 460								1	2				3	27
459 - 440					1	1		2	3		1		6	35
439 - 420					1	2		4	3				10	45
419 - 400								2	7		1		10	55
399 - 380								2			1		10	65
379 - 360								2	3				5	73
359 - 340					1	1	5	3	2	1	3		16	89
339 - 320					2	2	1	2	2				12	101
319 - 300						2	1	2		10	3		23	124
299 - 280							3	3		5	3		14	138
279 - 260						1	3	12		4	7	4	25	163
259 - 240							3	10	3				13	176
239 - 220							4	4	3				11	187
219 - 200													4	191
199 - 180	5		7	1	5								21	212

Ký hiệu

(b)

+ Trị số hiệu chỉnh

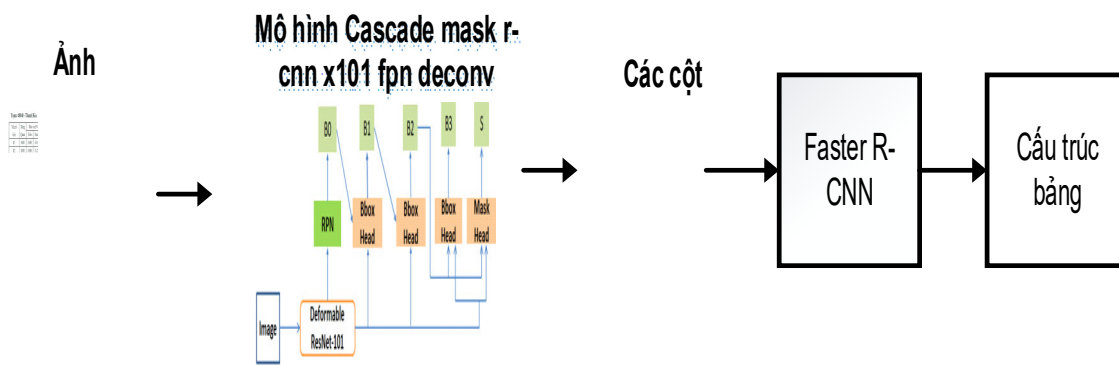
☒ Trị số khả nghi

☑ Trị số bỏ sung

**Hình 2.** Một số loại bảng thuộc tập dữ liệu Khí tượng thủy văn (a) bảng hiện đại, có các số viết tay (b) bảng lịch sử, có các số viết tay.

## 2.2. Phương pháp nghiên cứu

Nghiên cứu sử dụng mô hình Cascade mask r-cnn\_x101\_fpn\_deconv để nhận dạng hàng và cột. Do kết quả thực nghiệm ban đầu cho thấy kết quả nhận dạng hàng khá tệ, trong khi kết quả nhận dạng cột tốt. Điều này có thể là do các hàng chứa nhiều phần nền, ít đặc trưng cần thiết để nhận dạng và các hàng có khoảng cách gần nhau. Nghiên cứu này đề xuất phương pháp tiếp tục sử dụng một mạng Faster R-CNN để nhận dạng các cell được tách ra từ các cột vừa được nhận dạng. Ý tưởng này xuất phát từ việc nhận thấy rằng các cell trong một hàng có khả năng dễ dàng được nhận dạng hơn là trong toàn bảng, và các cell nhận dạng trong cột tương đương với một hàng trong bảng. Việc lựa chọn Faster R-CNN mục đích muốn lựa chọn một mô hình nhẹ và có hiệu quả cao.



**Hình 3.** Mô hình nhận dạng cấu trúc bảng.



Sau đây, sẽ mô tả về một số kỹ thuật được sử dụng trong nhận dạng cấu trúc bảng cụ thể là mô hình Cascade R-CNN kết hợp với tích chập biến dạng (*Cascade mask r-cnn\_r101\_fpn\_dconv\_c3-c5*). Mô hình này xuất phát từ một số nghiên cứu trước đây như về R-CNN [11], Fast R-CNN [12], Faster R-CNN [4], Mask R-CNN [13], Cascade R-CNN [14], tích chập biến dạng [15], ResNet [16], FPN [17]. Mô hình sẽ được huấn luyện trước với tập dữ liệu COCO [18] trước khi được tinh chỉnh lại với dữ liệu của nghiên cứu này. Điểm khác biệt của mô hình này so với các mô hình khác [3, 8] là sử dụng tích chập biến đổi. Ta có, phép toán tích chập thông thường như sau:

$$(F * I)(i, j) = \sum_{m=-K}^K \sum_{n=-K}^K F(m, n) \times I(i - m, j - n) \quad \forall i = 1, \dots, H, \forall j = 1, \dots, H \quad (1)$$

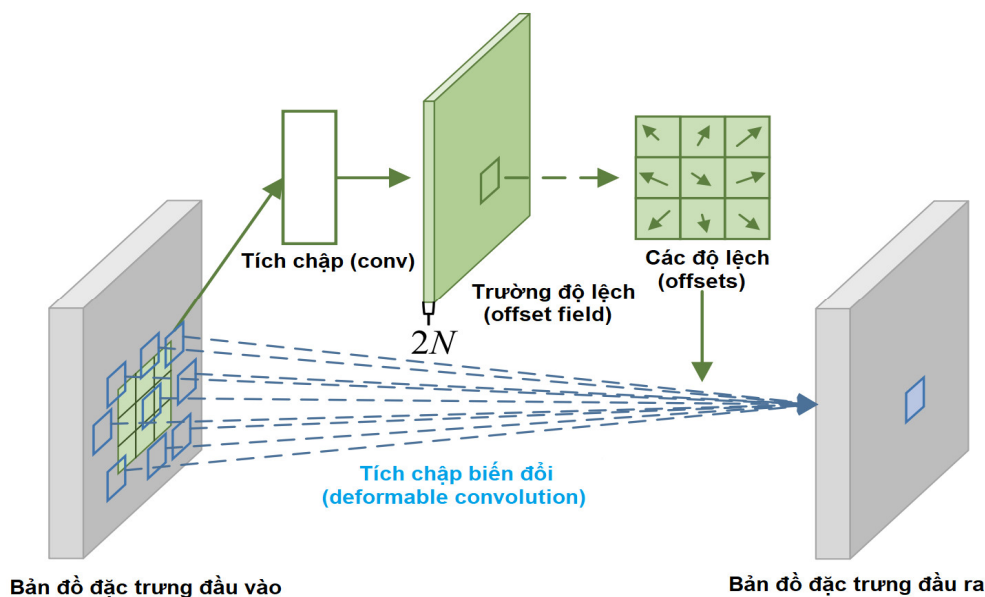
Dấu \* thể hiện phép toán tích chập, F là bộ lọc mà được học, I là ảnh, K là giá trị được tính như sau:  $\lfloor |F|/2 \rfloor$  trong đó  $|F|$  là kích thước của bộ lọc. H là chiều cao của ảnh, W là chiều rộng của ảnh, i, j đại diện cho vị trí đang thực hiện phép tích chập đã đề xuất sử dụng phép toán tích chập có thể biến dạng vào trong nhận dạng hàng và cột trong bảng. Phép tích chập biến dạng sử dụng các hệ số mở rộng thay vì sử dụng lưới cố định cho phép lớp tự điều chỉnh. Các hệ số này được tính toán dựa trên tập hợp các tầng tích chập khác, do đó chúng có thể học được. Công thức của phép tích chập biến dạng 2-D có thể được mô tả bằng công thức như sau:

$$(F \circ I)(i, j) = \sum_{m=-K}^K \sum_{n=-K}^K F(i, j) \times I(i - m + \delta_{i,j,m,n}^{vertical}, j - n + \delta_{i,j,m,n}^{horizontal}) \quad \forall i = 1 \dots H, \forall j = 1 \dots W \quad (2)$$

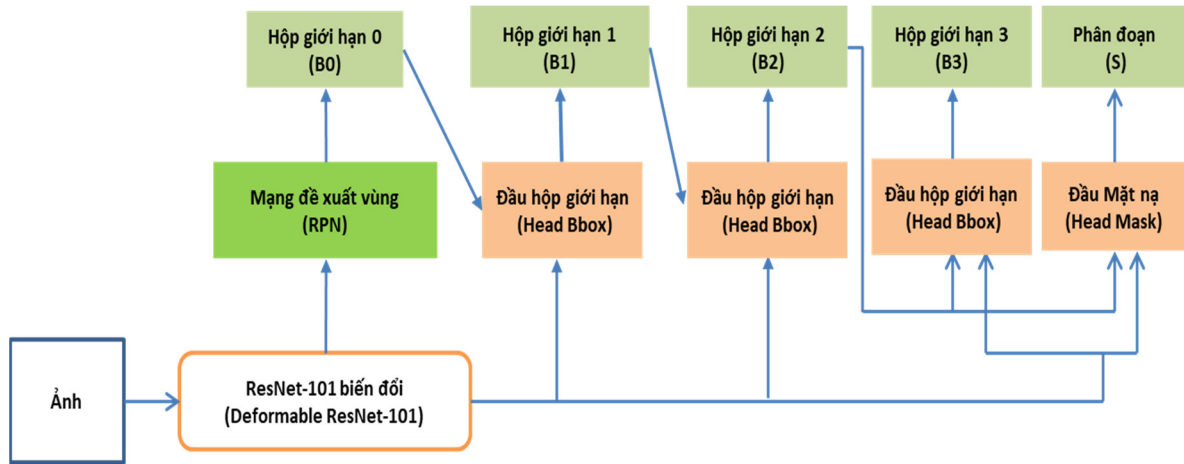
Kí hiệu  $\circ$  thể hiện cho phép tích chập biến dạng, các tham số khác hoàn toàn giống với công thức tích chập thông thường ở trên.  $\delta_{i,j,m,n}^{vertical}$ ,  $\delta_{i,j,m,n}^{horizontal}$  kí hiệu cho các tham số mở rộng. Hình 4 thể hiện phép tích chập biến đổi.

Hình 5 các độ lệch (offset) có thể đạt được bằng cách áp dụng một lớp tích chập trên bản đồ đặc trưng đầu vào. Hạt nhân của tích chập sử dụng độ phân giải, độ giãn nở không gian như của lớp tích chập hiện tại. Trường độ lệch (offset) đầu ra có cùng độ phân giải như bản đồ đặc trưng đầu vào và có 2N kênh trong đó 2N tương ứng với N hiệu số 2D.

Kết hợp tích chập biến đổi với mạng ResNet-101 [16] và với Cascade Mask R-CNN [14], và FPN [17] ta có mô hình mạng như trong Hình 5. Trong đó “I” là ảnh đầu vào, “conv” các tích chập khung xương, “pool” trích rút đặc trưng từng vùng, “H” đầu mạng, “B” hộp giới hạn và “C” là bộ phân loại. “B<sub>0</sub>” là các đề xuất trong toàn bộ kiến trúc.



Hình 4. Biểu diễn trực quan của phép tích chập biến đổi.



Hình 5. Mô hình mạng Cascade mask r-cnn\_r101 dconv.

### 3. Kết quả và thảo luận

#### 3.1. Tập dữ liệu đánh giá và độ đo

Nghiên cứu đánh giá trên hai tập dữ liệu TabStructDB mở rộng với 130 bảng thuộc tập dữ liệu KTTV, DetectCell và hiển thị một số kết quả nhận dạng cấu trúc bảng trong tập dữ liệu KTTV. Nghiên cứu này chia các tập dữ liệu theo tỷ lệ 60% cho tập huấn luyện, 10% cho tập kiểm định, 30 % cho tập kiểm tra. Kết quả thực nghiệm được thực hiện trên tập kiểm tra.

Độ đo được sử dụng dựa trên tham số IoU (*Intersection over Union*) [19] trên các phép đo Precision, Recall, F1. IoU bằng diện tích vùng chồng lấn giữa hộp dự đoán và hộp đúng (ground truth) chia cho diện tích của vùng được tạo bởi hợp của hộp dự đoán và hộp đúng.

$$IoU = \frac{IA}{(DA + GA - IA)} \quad (3)$$



Hình 6. Mô tả về IoU.

Công thức về độ đo P, R, F1 trung bình:

$$P = \frac{1}{N} \sum_{i=1}^N (IA_i / DA_i) \quad (4)$$

$$R = \frac{1}{N} \sum_{i=1}^N (IA_i / GA_i) \quad (5)$$

$$F_1 = 2 * ((P * R) / (P + R)) \quad (6)$$

Trong đó IA là diện tích vùng giao nhau giữa hộp dự đoán và hộp đúng, DA là diện tích hộp dự đoán, GA là diện tích hộp đúng, N là số ảnh trong tập kiểm tra.

#### 3.2. Kết quả đánh giá

Nghiên cứu này sử dụng ngôn ngữ lập trình Python, thư viện PyTorch, MMDetection [20] xây dựng chương trình và các mô hình thử nghiệm. Các thực nghiệm của nghiên cứu này được thực hiện trên nền tảng Google Colaboratory Pro với Tesla V100-SXM2 có bộ nhớ GPU 16GB, Intel(R) Xeon(R) CPU @2.30GHz và 24 GB Ram.

### 3.2.1. Đánh giá kết quả nhận dạng hàng và cột

Thực nghiệm việc nhận dạng cấu trúc bảng. Ta thấy theo như bảng 2, mô hình cho phép nhận dạng cột khá tốt, tỷ lệ nhận dạng có thể lên tới 97,20 % và độ đo F1 theo IOU lên tới 94,04 % (độ tin cậy 0,7). Tuy nhiên, việc nhận dạng hàng còn chưa tốt chỉ có 49,77% (độ tin cậy 0,4) hàng có thể nhận ra, những hàng có khả năng nhận ra thì có ngưỡng IoU trùng khớp cao  $F1 = 89,19\%$  (độ tin cậy 0,4). Các bảng có các hàng cách xa nhau (lớn hơn độ rộng của ký tự) có xu thế nhận dạng rất tốt, các hàng gần nhau thường không thể phân biệt và nhận dạng được.  $d = (\text{số đối tượng phát hiện ra có ngưỡng IoU} > 0,5) / (\text{số đối tượng thực có})$ .

**Bảng 2.** Kết quả nhận dạng hàng và cột.

IoU	Row				Column				All			
	P	R	F1	d	P	R	F1	d	P	R	F1	d
<b>0,4</b>	90,72	87,71	<b>89,19</b>	<b>49,77</b>	94,19	93,23	93,71	97,78	86,82	85,16	85,98	62,03
<b>0,5</b>	91,17	88,59	89,86	47,94	94,32	93,35	93,83	97,78	87,03	85,61	86,31	60,87
<b>0,6</b>	91,54	89,20	90,36	46,12	94,42	93,51	93,96	97,53	87,16	85,95	86,55	59,64
<b>0,7</b>	92,03	89,92	90,97	44,43	94,47	93,61	94,04	<b>97,20</b>	87,34	86,27	86,80	58,53

Nghiên cứu sử dụng một thủ thuật nhỏ là thay đổi kích thước của ảnh theo chiều cao. Nghiên cứu kéo dẫn ảnh 3 lần theo chiều cao  $h=h*3$ , kết quả thực nghiệm cho thấy việc kết hợp kéo dẫn ảnh dự đoán hàng và giữ nguyên ảnh gốc cho dự đoán cột cho kết quả tốt nhất như trong bảng 3. Ta cũng thấy mô hình cũng tốt hơn so với mô hình Cascade mask hrnetv2pw40.

**Bảng 3.** Kết quả thực nghiệm nhận dạng hàng cột sử dụng kéo dẫn ảnh.

Phương pháp	Mô hình	IoU				Wavg.
		0,6	0,7	0,8	0,9	
Kéo dẫn ảnh	Cascade mask r-cnn x101 dconv	83,87	84,40	85,06	85,67	<b>84,75</b>
Ảnh gốc (cột) và kéo dẫn (hàng)	Cascade mask r-cnn x101 dconv	84,75	85,34	85,58	<b>86,65</b>	<b>85,58</b>
Ảnh gốc (cột) và kéo dẫn (hàng)	Cascade mask hrnetv2pw40	82,53	83,55	84,52	86,08	84,17

Hình 7 minh họa kết quả nhận dạng hàng và cột tốt. Điều này là do các hàng ở đây có khoảng cách đủ để nhận dạng. Trong hình 8, mô hình nhận dạng thiếu hàng rất nhiều. Áp dụng kỹ thuật kéo dẫn ảnh được kết quả thực nghiệm nhận dạng tốt ở hình 9. Sau đây là các kết quả nhận dạng cụ thể:



**Hình 7.** Kết quả nhận dạng hàng cột tốt.

Ảnh kết quả

SkM*	$\beta_2$	$\gamma$	$r_n$	$r_p$	$\Delta_n$	$\Delta_p$	$-\lambda_n$	$-\lambda_p$
<sup>100</sup> Sn	0.00	—	4.33	4.42	0.00	0.00	17.00	3.15
<sup>102</sup> Sn	0.00	—	4.38	4.43	1.35	0.00	11.86	3.86
<sup>104</sup> Sn	0.00	—	4.42	4.45	1.70	0.00	11.24	4.59
<sup>106</sup> Sn	0.00	—	4.46	4.46	1.80	0.00	10.65	5.33
<sup>108</sup> Sn	0.00	—	4.50	4.48	1.84	0.00	10.13	6.07
<sup>110</sup> Sn	0.00	—	4.54	4.49	1.97	0.00	9.75	6.84
<sup>112</sup> Sn	0.00	—	4.58	4.51	2.05	0.00	9.43	7.60
<sup>114</sup> Sn	0.00	—	4.62	4.53	2.10	0.00	9.13	8.35
<sup>116</sup> Sn	0.00	—	4.66	4.55	2.06	0.00	8.86	9.09
<sup>118</sup> Sn	0.00	—	4.69	4.56	2.00	0.00	8.60	9.82
<sup>120</sup> Sn	0.00	—	4.72	4.58	1.94	0.00	8.36	10.54
<sup>122</sup> Sn	0.00	—	4.75	4.59	1.85	0.00	8.12	11.25
<sup>124</sup> Sn	0.00	—	4.78	4.61	1.75	0.00	7.87	11.95
<sup>126</sup> Sn	0.00	—	4.81	4.62	1.62	0.00	7.62	12.63
<sup>128</sup> Sn	0.00	—	4.84	4.63	1.42	0.00	7.34	13.31
<sup>130</sup> Sn	0.00	—	4.87	4.64	1.10	0.00	7.01	13.98
<sup>132</sup> Sn	0.00	—	4.89	4.66	0.00	0.00	8.48	14.64
<sup>134</sup> Sn	0.00	—	4.93	4.67	0.68	0.00	3.62	15.07
<sup>136</sup> Sn	0.00	—	4.97	4.68	0.78	0.00	3.36	15.51
<sup>138</sup> Sn	0.00	—	5.02	4.69	1.04	0.00	3.11	15.90
<sup>140</sup> Sn	0.00	—	5.07	4.71	1.74	0.00	3.06	16.43

Hình 8. Bảng nhận thiếu hàng.

SkI3	$\beta_2$	$\gamma$	$r_n$	$r_p$	$\Delta_n$	$\Delta_p$	$-\lambda_n$	$-\lambda_p$
<sup>100</sup> Sn	0.00	—	4.33	4.40	0.00	0.00	17.18	3.50
<sup>102</sup> Sn	0.00	—	4.38	4.42	1.54	0.00	11.00	4.33
<sup>104</sup> Sn	0.00	—	4.43	4.43	1.65	0.00	10.40	5.16
<sup>106</sup> Sn	0.00	—	4.48	4.45	2.20	0.00	10.34	6.00
<sup>108</sup> Sn	0.00	—	4.52	4.47	1.83	0.00	9.93	6.81
<sup>110</sup> Sn	0.00	—	4.56	4.49	2.37	0.00	9.76	7.61
<sup>112</sup> Sn	0.00	—	4.60	4.50	2.43	0.00	9.41	8.38
<sup>114</sup> Sn	0.00	—	4.64	4.52	2.31	0.00	9.09	9.13
<sup>116</sup> Sn	0.00	—	4.67	4.53	2.02	0.00	8.69	9.85
<sup>118</sup> Sn	0.00	—	4.71	4.55	2.04	0.00	8.28	10.55
<sup>120</sup> Sn	0.00	—	4.75	4.56	0.00	0.00	8.50	10.94
<sup>122</sup> Sn	0.00	—	4.78	4.57	1.70	0.00	7.61	11.94
<sup>124</sup> Sn	0.00	—	4.81	4.59	1.68	0.00	7.34	12.67
<sup>126</sup> Sn	0.00	—	4.84	4.60	1.61	0.00	7.10	13.40
<sup>128</sup> Sn	0.00	—	4.87	4.61	1.46	0.00	6.85	14.13
<sup>130</sup> Sn	0.00	—	4.90	4.62	1.19	0.00	6.54	14.86
<sup>132</sup> Sn	0.00	—	4.93	4.64	0.00	0.00	8.00	15.56
<sup>134</sup> Sn	0.00	—	4.99	4.65	0.78	0.00	2.00	15.93
<sup>136</sup> Sn	0.00	—	5.04	4.67	1.30	0.00	2.10	16.39
<sup>138</sup> Sn	0.00	—	5.10	4.69	1.61	0.00	2.09	16.63
<sup>140</sup> Sn	0.00	—	5.14	4.71	1.66	0.00	1.94	17.22

Hình 9. Bảng nhận đủ hàng sau khi thực hiện phép kéo dãn.

### 3.2.2. Kết quả nhận dạng ô trong bảng

Sau đó, nghiên cứu này tiến hành thực nghiệm việc nhận dạng ô với từng cột và kết hợp chúng lại với nhau. Bảng 4, thể hiện kết quả nhận dạng ô với tập dữ liệu là các cột.

Bảng 4. Kết quả nhận dạng cell.

Mô hình	IoU				WAvg.
	0,6	0,7	0,8	0,9	
Faster R-CNN	88,52	88,60	88,60	<b>88,82</b>	88,63



Dưới đây là một số kết quả thực nghiệm nhận dạng cấu trúc bảng từ đầu đến cuối. Qua trình xử lý gồm: phát hiện hàng, cột và phát hiện cell trong bảng. Qua kết quả thực nghiệm, thuật toán có khả năng nhận dạng cấu trúc bảng tương đối tốt. Theo bảng 10, các ô được nhận dạng chính xác bằng hiện đại, đặc biệt với cả các chữ viết tay. Bảng 11, bảng hiện đại với các chữ được đánh máy, kết quả nhận dạng các ô chính xác. Bảng 12, bảng vẽ tay có tính chất lịch sử, chữ viết tay, các ô được nhận dạng chính xác. Bảng 13, bảng lịch sử, giấy có vết gấp, và chữ là cả viết tay và đánh máy, thuật toán cho phép ta xác định được chính xác các ô.

TT	Ngày	Giờ	Phút	H	Q	Qhc	F	Vmc	Vdb	Tên File số liệu thô	Tên File Config	Người đo	Hướng chảy	Hướng đo	Ghi Chú
1	27	17	00	35	232,1	232	1880	0,18	0,17	TR1004.R.000	TR1RA.CFG	Thuyết	chảy 0	T-P	Ngược
2	27	17	10							TR1004.R.000	TR1RA.CFG		chảy 0	T-P	
3	27	17	15	102	362,4	362	1600	0,23	0,25	TR1002.R.000	TR1RA.CFG		chảy 0	T-P	
4	27	18	00	120	519,6	519	1690	0,31	0,32	TR1003.R.000	TR1RA.CFG		chảy 0	T-P	
5	27	19	00	140	597,9	598	1690	0,32	0,32	TR1004.R.000	TR1RA.CFG	Thuyết	chảy 0	T-P	
6	27	20	00	455	424,3	424	1730	0,25	0,27	TR1005.R.000	TR1RA.CFG	Thuyết	chảy 0	T-P	
7	27	20	15	159	426,5	426	1740	0,24	0,26	TR1006.R.000	TR1RA.CFG		chảy 0	T-P	(12)
8	27	20	45	166	323,5	324	1760	0,18	0,16	TR1007.R.000	TR1RA.CFG		chảy 0	T-P	
9	27	21	00	169	333,6	334	1770	0,19	0,19	TR1008.R.000	TR1RA.CFG		chảy 0	T-P	
10	27	21	59	173	7,6	7,6	1780	0,00	0,00	TR1009.R.000	TR1RA.CFG		chảy 0	T-P	Yếu
11	27	22	55	166	458,7	459	1750	0,26	0,22	TR1010.R.000	TR1RA.CFG		chảy 0	T-P	
12	27	23	53	152	420,4	420	1720	0,53	0,60	TR1011.R.000	TR1RA.CFG		chảy 0	T-P	
13	28	0	53	134	1059	1060	1680	0,63	0,73	TR1012.R.000	TR1RA.CFG		chảy 0	T-P	
14	28	1	52	119	1159	1160	1690	0,70	0,81	TR1013.R.000	TR1RA.CFG		chảy 0	T-P	
15	28	2	59	102	1200	1200	1680	0,75	0,87	TR1014.R.000	TR1RA.CFG		chảy 0	T-P	
16	28	3	59	87	1216	1220	1660	0,78	0,89	TR1015.R.000	TR1RA.CFG		chảy 0	T-P	
17	28	7	55	74	1172,5	1180	1530	0,77	0,80	TR1016.R.000	TR1RA.CFG		chảy 0	T-P	
18	28	5	58	63	1109,4	1110	1580	0,74	0,82	TR1017.R.000	TR1RA.CFG		chảy 0	T-P	
19	28	6	56		1132,1	1130				TR1019.R.000	TR1RA.CFG		chảy 0	T-P	
20	28	7	00	53	1133,5	1130	1480	0,76	0,83	TR1020.R.000	TR1RA.CFG	Thuyết	chảy 0	T-P	(4)
21	28	7	59	47	1087,8	1090	1460	0,75	0,85	TR1021.R.000	TR1RA.CFG		chảy 0	T-P	
22	28	9	00	43	1094,0	1090	1440	0,75	0,80	TR1022.R.000	TR1RA.CFG	Nguyen	chảy 0	T-P	
23	28	9			1063,5	1060				TR1023.R.000	TR1RA.CFG		chảy 0	T-P	(3)
24	28	10	00	39	1058,0	1060	1440	0,74	0,81	TR1024.R.000	TR1RA.CFG		chảy 0	T-P	

**Bảng 10.** Kết quả nhận dạng cấu trúc bảng kết quả đo lưu lượng bằng máy ADCP.

Yếu tố	Tổng		Bức xạ SN		Bức xạ SD		Trục xạ			Tổng xạ	Nhiệt độ	Bức xạ thực		Cân cân		Thời gian	Ghi chú
	Giờ	Quan	Trên	Dưới	Trên	Dưới	Tán xạ	Trục xạ	ngang			SN	SD	bức xạ	Ambedo		
01	0.00	0.00	0.15	148.49	148.68	0.00	0.00	0.00	148.49	18.90	-0.15	-0.19	-0.33	0.00			
02	0.00	0.00	0.13	148.70	148.86	0.00	0.00	0.00	148.70	19.00	-0.13	-0.16	-0.29	0.00			
03	0.00	0.00	0.05	148.55	148.68	0.00	0.00	0.00	148.55	18.93	-0.05	-0.13	-0.18	0.00			
04	0.00	0.00	0.01	148.50	148.65	0.00	0.00	0.00	148.50	18.90	-0.01	-0.15	-0.16	0.00			
05	0.00	0.00	0.00	148.33	148.44	0.00	0.00	0.00	148.33	18.82	0.00	-0.11	-0.11	0.00			
06	0.00	0.00	0.00	148.32	148.45	0.00	0.00	0.00	148.32	18.81	0.00	-0.13	-0.14	0.00			
07	0.21	0.31	0.23	148.10	148.43	0.26	0.00	0.00	148.42	18.71	0.08	-0.33	-0.25	0.75			
08	7.93	8.55	1.48	148.27	149.15	5.55	7.64	2.38	156.82	18.79	7.07	-0.87	6.20	0.17			
09	42.48	44.09	7.22	150.72	152.48	28.64	28.12	13.83	194.81	19.99	36.87	-1.76	35.11	0.16			
10	53.15	54.65	7.91	154.66	155.20	36.08	27.07	17.07	209.31	21.89	46.74	-0.54	46.20	0.14			
11	58.30	60.16	8.86	154.82	155.72	39.75	25.83	18.54	214.98	21.96	51.30	-0.90	50.40	0.15			
12	50.47	52.08	7.26	156.00	156.33	34.61	21.21	15.86	208.08	22.52	44.82	-0.33	44.49	0.14			
13	34.76	36.08	4.93	154.19	154.62	23.64	15.49	11.12	190.27	21.66	31.15	-0.43	30.72	0.14			
14	41.86	42.93	6.53	154.09	155.00	28.53	21.14	13.33	197.02	21.61	36.40	-0.91	35.49	0.15			
15	28.43	29.65	4.36	153.50	154.15	19.28	18.59	9.15	183.15	21.33	25.29	-0.65	24.65	0.15			
16	29.94	31.03	4.84	153.45	154.12	20.49	30.39	9.45	184.48	21.31	26.18	-0.67	25.51	0.16			
17	19.06	19.99	2.89	153.05	153.36	13.03	59.98	6.03	173.04	21.11	17.10	-0.31	16.79	0.14			
18	1.63	1.94	0.19	151.75	151.80	1.26	-2.96	0.37	153.70	20.49	1.76	-0.04	1.71	0.10			
19	0.00	0.00	0.07	150.67	150.99	0.00	0.00	0.00	150.67	19.96	-0.07	-0.32	-0.40	0.00			
20	0.00	0.00	0.14	150.35	150.72	0.00	0.00	0.00	150.35	19.81	-0.14	-0.37	-0.50	0.00			
21	0.00	0.00	0.13	150.06	150.44	0.00	0.00	0.00	150.06	19.67	-0.13	-0.38	-0.51	0.00			
22	0.00	0.00	0.09	149.74	150.22	0.00	0.00	0.00	149.74	19.51	-0.09	-0.48	-0.57	0.00			
23	0.00	0.00	0.12	149.39	149.87	0.00	0.00	0.00	149.39	19.34	-0.12	-0.48	-0.60	0.00			
24	0.00	0.00	0.14	149.26	149.76	0.00	0.00	0.00	149.26	19.28	-0.14	-0.50	-0.63	0.00			

**Bảng 11.** Kết quả nhận dạng cấu trúc bảng đo bức xạ giờ.



$\Delta Q$ T-1	H cm	$Q_c$ $m^3/s$	$F_{tra}$ $m^2$	$V_{tra}$ $m/s$	$Q = V \times F$ $m^3/s$	$\Delta Q = Q - Q_c$	$\frac{\Delta Q}{Q_c}$
	260	525	615	0,85	523	-2	0,38
	280	602	684	0,88	602	0	0
	300	687	752	0,91	684	-3	0,44
	320	777	820	0,95	779	2	0,26
	340	875	889	0,98	871	-4	0,46
	360	975	956	1,02	975	0	0
	380	1085	1025	1,06	1086	1	0,09
	400	1200	1095	1,10	1204	4	0,33
	420	1317	1162	1,13	1313	-4	0,30
	440	1455	1230	1,18	1451	-4	0,27
	260	505	615	0,82	504	-1	0,20
	280	555	684	0,81	554	-1	0,20
	300	615	752	0,82	617	2	0,32
	320	685	820	0,84	689	4	0,58
	340	760	889	0,85	756	-4	0,53
	360	847	956	0,88	841	-6	0,71
	380	940	1025	0,92	943	3	0,32
	400	1040	1095	0,95	1040	0	0
	420	1150	1162	0,99	1150	0	0
	440	1275	1230	1,04	1272	-3	0,31

Bảng 12. Kết quả nhận dạng bảng kiểm tra Q lịch sử viết tay.

Số hiệu thủy trực	Khoảng cách từ mốc khởi điểm (m)	ĐỘ SÂU (m)			Góc chệch của dây cáp (độ)	Trị số hiệu chỉnh	Độ sâu đã hiệu chỉnh (m)	Hiệu chỉnh để quy về mực nước dùng để tính	Độ sâu đã quy về mực nước dùng để tính	Độ cao đáy sông (m)	Đất ở đáy	GHI CHÚ
		I	II	Trung bình								
1	2	3	4	5	6	7	8	9	10	11	12	13
17	150	9,0	9,0	9,0			9,0			8,4		
20	140	9,2	9,2	9,2			9,2			8,6		
21	130	9,1	9,1	9,1			9,1			8,5		
22	120	6,5	6,8	6,6			6,8			6,7		
23	110	3,22	3,22	3,22			3,22			2,69		
1117	100,6			0,0			0,0			0,57		
C <sub>1</sub>	99,6									0,78		
C <sub>2</sub>	97,8									1,75		
C <sub>3</sub>	96,0									1,51		
C <sub>4</sub>	94,3									1,84		
C <sub>5</sub>	92,5									2,12		
C <sub>6</sub>	83,8									23,9		
KP	0,0									2,00		

Bảng 13. Kết quả nhận dạng bảng trắc đồ ngang.

#### 4. Kết luận

Nghiên cứu đã đề xuất thuật toán nhận dạng cấu trúc bảng dựa trên mô hình Cascade mask R-CNN x101FPN deconv để nhận dạng hàng và cột, sau đó sử dụng mô hình Faster R-CNN để nhận dạng các ô trong bảng từ đó đưa ra cấu trúc của bảng. Thuật toán cho thấy hiệu quả cao trong nhận dạng cấu trúc bảng với gần 90% độ chính xác. Trong thời gian tới, bài báo sẽ tiếp tục nghiên cứu phát triển thuật toán trên thông tin hàng để nâng cao kết quả nhận dạng. Đồng thời tiếp tục thử nghiệm các dữ liệu khác của KTTV để ứng dụng tốt hơn trong thực tế.

**Đóng góp của tác giả:** Xây dựng ý tưởng nghiên cứu: H.H.P., N.D.D., P.L.P.; Lựa chọn phương pháp nghiên cứu: H.H.P., N.D.D.; Xử lý số liệu: N.D.D.; Phân tích mẫu: N.D.D.; Lấy mẫu: H.H.P., P.L.P.; Viết bản thảo bài báo: H.H.P., N.D.D.; Chỉnh sửa bài báo: H.H.P., N.D.D., P.L.P.

**Lời cam đoan:** Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

#### Tài liệu tham khảo

1. Pyreddy, P.; Croft, W.B. Tinti: A system for retrieval in text tables title2. *Technical report, USA, 1997*.
2. Schreiber, S.; Agne, S.; Wolf, I.; Dengel, A.; Ahmed, S. DeepDeSRT: deep learning for detection and structure recognition of tables in document images. *Proceeding of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017*.
3. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149.
4. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object detection via region-based fully convolutional networks. *Proceeding of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 2016*, 379–387.
5. Marmot Dataset. Institute of Computer Science and Technology of Peking University and Institute of Digital Publishing of Founder R&D Center, China, [http://www.icst.pku.edu.cn/cpdp/data/marmot\\_data.htm](http://www.icst.pku.edu.cn/cpdp/data/marmot_data.htm), 2010.
6. Göbel, M.; Hassan, T.; Oro, E.; Orsi, G. ICDAR 2013 Table Competition. *Proceeding of the 12th International Conference on Document Analysis and Recognition, 2013*, 1449–1453.
7. Siddiqui, S.A.; Fateh, I.A.; Rizvi, S.T.R.; Dengel, A.; Ahmed, S. DeepTabStR: Deep Learning based Table Structure Recognition. *International Conference on Document Analysis and Recognition (ICDAR), 2019*, 1403–1409.
8. Prasad, D.; Gadpal, A.; Kapadni, K.; Visave, M.; Sultanpure, K. CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020*.
9. Gao, G.; Huang, Y.; Dejean, H.; Meunier, J.; Yan, Q.; Fang, Y.; Kleber, F.; Lang, E. Icdar 2019 competition on table detection and recognition (ctdar). *International Conference on Document Analysis and Recognition (ICDAR), 2019*, 1510–1515.
10. Li, M.; Cui, L.; Huang, S.; Wei, F.; Zhou, M.; Li, Z. Tablebank: Table benchmark for image-based table detection and recognition. *Proceedings of the 12th Conference on Language Resources and Evaluation, 2020*, 1918–1925.
11. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition, 2014*. ISBN:978-1-4799-5118-5.

12. Girshick, R. Fast R-CNN. Proceedings of the IEEE international conference on computer vision, 2015.
13. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask r-cnn. *ICCV. IEEE* **2017**, 2980–2988.
14. Cai, Z.; Vasconcelos, N. Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, 43, 1483–1498.
15. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. IEEE International Conference on Computer Vision (ICCV) 2017, 1, pp. 1-3.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, **2016**, 770–778.
17. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 936–944, doi: 10.1109/CVPR.2017.106.
18. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. 2014, pp. 740–755. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48).
19. Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 658–666.
20. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; Zhang, Z.; Cheng, D.; Zhu, C.; Cheng, T.; Zhao, Q.; Li, B.; Lu, X.; Zhu, R.; Wu, Y.; Dai, J.; Wang, J.; Shi, J.; Ouyang, W.; Loy, C.C.; Lin, D. Mmdetection: Open MMLab detection toolbox and benchmark. *CoRR, abs/1906.07155*, 2019.

## Research and apply table structure recognition algorithm based on object detection

Ngo Dai Duong<sup>1</sup>, Hai-Hong Phan<sup>1\*</sup>, Pham Le Phuong<sup>2</sup>

<sup>1</sup> Military Technical Academy, Viet Nam; hongpth@lqdtu.edu.vn;  
daiduong28789@hotmail.com

<sup>2</sup> Hydro-Meteorological Information and Data Center; phuongpl80@gmail.com

**Abstract:** Table structure identification is an important issue in document digitization. With the development of current deep learning techniques, the detection of tables has made great strides, while table structure identification still faces many difficulties due to complex table structure, especially with real data practice. This article proposes a method to apply the Cascade mask model R-CNN x101FPN deconv to identify rows and columns. The paper also proposes to use the Faster R-CNN model to identify the cells in the table, thereby giving out the table structure. The proposed algorithm is evaluated on popular datasets such as TabStructDB and documents collected from Hydrometeorology stations. The experimental results reached 90% accuracy on these datasets. The algorithm is capable of effectively applying to the identification of the table structure of common documents; especially, the algorithm is capable of dealing with historical documents and handwriting, in accordance with the document characteristics of the hydrometeorology industry. Since then, it contributes to the digitization of documents, storage and data retrieval of hydrometeorology industry.

**Keywords:** Structure table recognition; Cell recognition; Meteorological and hydrological.