PHÁT HIỆN DỤNG CỤ PHẪU THUẬT THỜI GIAN THỰC TRONG MỔ NỘI SOI DỰA TRÊN MẠNG NƠ-RÔN TÍCH CHẬP

REAL-TIME SURGICAL TOOL DETECTION IN MINIMALLY INVASIVE SURGERY USING CONVOLUTIONAL NEURAL NETWORK

Kim Đình Thái[1,*]

TÓM TẮT

Phát hiện dụng cụ phẫu thuật bao gồm việc xác định vị trí và loại dụng cụ phẫu thuật trong một bức ảnh hoặc một video. Đây là một bài toán quan trọng trong việc ứng dụng thị giác máy tính nhằm nâng cao hiệu quả của phẫu thuật nội soi. Bài báo này trình bày một thuật toán phát hiện dụng cụ phẫu thuật ở thời gian thực dựa trên mạng nơ-rôn tích chập (CNNs). Tập dữ liệu được sử dụng trong nghiên cứu này được tạo ra từ những video phẫu thuật cắt túi mật. Kết quả thực nghiệm cho thấy rằng thuật toán có thể hoạt động ở thời gian thực với tốc độ khung hình là 25,4 (fps) và độ chính xác trung bình của phát hiện dụng cụ (mAP) là 71,54%.

Từ khóa: *Phẫu thuật nội soi, CNN, phát hiện dụng cụ phẫu thuật, thị giác máy tính.*

ABSTRACT

The sugical tool detection is identifies the surgical tool category and locates the position using a bounding box for every known tool within an image or video. This is a significant issue in the use of computer vision to increase laparoscopic surgery efficacy. This paper presents a real-time surgical tool detection algorithm based on convolutional neural networks (CNNs). The dataset for this research was derived from cholecystectomy surgical videos. The experimental results show that the algorithm can operate in real-time at a frame rate of 25.4 (fps), with a mean average precision (mAP) of 71.54% over our dataset.

Keywords: *MIS, CNN, surgical tool detection, computer vision.*

[1]Trường Quốc tế, Đại học Quốc gia Hà Nội
[*]Email: thaidk@isvnu.vn
Ngày nhận bài: 15/8/2021
Ngày nhận bài sửa sau phản biện: 10/02/2022
Ngày chấp nhận đăng: 25/02/2022

1. GIỚI THIỆU

Ngày nay, phương pháp mổ nội soi đang dần thay thế phương pháp mổ hở truyền thống nhờ những ưu điểm vượt trội của nó, chẳng hạn như chẳng hạn như: ít đau sau mổ hơn, hồi phục nhanh hơn, thời gian nằm viện ngắn hơn, vết sẹo nhỏ hơn và nguy cơ nhiễm trùng thấp hơn so với mổ mở [1, 2]. Trong phẫu thuật nội soi, các bác sĩ sẽ tạo ra các vết rạch "đủ nhỏ" lên cơ thể bệnh nhân để cho phép các dụng cụ phẫu thuật và ống nội soi đi qua. Sau đó, nhà phẫu thuật sẽ thực hiện các thao tác cắt hoặc đốt bởi các dụng cụ cầm tay thông qua việc quan sát những hình ảnh trên một màn hình được cung cấp bởi camera gắn trên ống nội soi. Do không thể nhìn trực tiếp vào trong khoang bụng của bệnh nhân mà phải nhìn gián tiếp thông qua màn hình hiển thị để thực hiện các thao tác, cho nên kỹ thuật mổ nội soi thực sự khó hơn so với kỹ thuật mổ hở truyền thống rất nhiều, đặc biệt là với các bác sĩ ít kinh nghiệm [3]. Do vậy, thời gian cần thiết cho việc đào tạo một bác sĩ phẫu thuật nội soi thường khá dài. Hơn nữa, việc đánh giá kỹ năng sau quá trình đào tạo vẫn được thực hiện thủ công, dựa trên việc quan sát và đánh giá chủ quan của một chuyên gia.

Trong những năm gần đây, thị giác máy tính đã có những phát triển vượt bậc và do đó việc tích hợp kỹ thuật thị giác máy tính đã trở thành một phần quan trọng trong computer-assisted interventions (CAI) cho phẫu thuật nội soi [4]. Có thể lấy ví dụ như là việc áp dụng thị giác máy tính để phát hiện đầu của dụng cụ phẫu thuật (surgical tool's tip): Với những hình ảnh thu được từ camera nội soi có thể trích xuất được thông tin về loại dụng cụ và vị trí của đầu dụng cụ có trong bức ảnh đó. Từ đó, một công cụ đánh giá tự động về hiệu quả của một quá trình mổ (hoặc kỹ năng của một bác sĩ) được phát triển thông qua việc phân tích quỹ đạo chuyển động của đầu dụng cụ được sử dụng trong suốt quá trình phẫu thuật [5]. Bên cạnh đó, thông tin phản hồi về vị trí của đầu dụng cụ cũng có thể được sử dụng để điều khiển tự động camera nội soi tới vị trí mong muốn [6].

Trên thế giới đã có một số nghiên cứu trước đó đối với bài toán phát hiện dụng cụ nội soi. Có nhiều cách tiếp cận, có thể kể đến như Cai et al. [7] đã sử dụng những markers để đặt trên dụng cụ phẫu thuật cho việc phát hiện. Cách tiếp cận khác là sử dụng tần số radio cho việc phát hiện và theo dõi dụng cụ phẫu thuật ở thời gian thực [8]. Tuy nhiên, cả hai cách tiếp cận này đều yêu cầu một sự sửa đổi đối với dụng cụ được theo dõi [9].

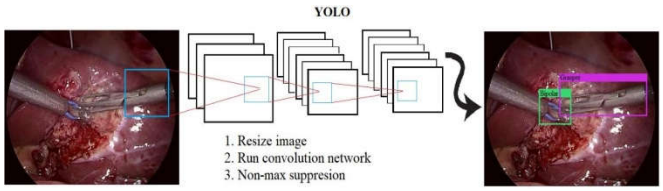
Do đó, một số nghiên cứu dựa trên thị giác máy tính đã được đề xuất. Cách tiếp cận này dựa trên những đặc trưng hình ảnh cho việc phát hiện, chẳng hạn như: dựa trên màu [6, 10], gradients [11] và texture [12]. Tuy nhiên, hầu hết các

nghiên cứu này không đủ mạnh mẽ để phát hiện các dụng cụ phẫu thuật với các điều kiện môi trường trong bụng bệnh nhân, nơi mà thường có sự xuất hiện của khói, máu, độ chói, độ bóng... Gần đây, đã có một số nghiên cứu dựa trên mạng nơ-rôn tích chập (CNN). Puta et al. [13] là người đầu tiên sử dụng CNN cho nhiều nhiệm vụ nhận dạng trong video nội soi. Một vài nghiên cứu [14-16] đã được đề xuất trong thách thức phát hiện sự xuất hiện dụng cụ trong M2CAI 2016 [17]. Jin et al. [5] sau đó đã phát triển công việc này bằng việc dựa vào Fast Region-based Convolutional Network (Faster R-CNN) [18] để nhận ra không chỉ sự xuất hiện mà còn định vị trí của đầu dụng cụ trong những video cắt túi mật. Tuy nhiên, ở Việt Nam, việc ứng dụng thị giác máy tính vào trong mổ nội soi nói chung, cũng như những nghiên cứu và ứng dụng về việc phát hiện dụng cụ phẫu thuật vẫn còn khá mới mẻ.

Trong nghiên cứu này, tác giả sử dụng một kiến trúc CNN rất nổi tiếng có tên là Only Look Once (YOLO) [19-21] cho việc phát hiện dụng cụ phẫu thuật. Kiến trúc này không chỉ phát hiện được sự xuất hiện của dụng cụ mà còn định được vị trí của dụng cụ đó trong một bức ảnh. Giai đoạn nghiên cứu hiện tại, tác giả chưa xây dựng được một tập dữ liệu đủ lớn cho nhiều loại dụng cụ phẫu thuật nội soi. Vì vậy, trong bài báo này, tác giả sử dụng tập dữ liệu m2cai16-tool-locations được cung cấp bởi [5] cho việc phát hiện bảy loại dụng cụ thường được sử dụng trong phẫu thuật nội soi cắt túi mật. Sau đó, tác giả thực hiện huấn luyện và đánh giá hiệu quả của mô hình đề xuất dựa trên tập dữ liệu này.

2. PHÁT HIỆN DỤNG CỤ PHẪU THUẬT

Phát hiện dụng cụ phẫu thuật gồm có hai nhiệm vụ. Một là phát hiện hoặc nhận dạng những dụng cụ nào xuất hiện và hai là xác định hộp bao quanh những dụng cụ đó, có trong một bức ảnh hoặc một video nội soi. Trong nghiên cứu này, chúng tôi sử dụng kiến trúc YOLOv3 (phiên bản 3) [21] cho việc phát hiện dụng cụ phẫu thuật, như được minh họa trong hình 1. YOLO là một kiến trúc CNN nổi tiếng được sử dụng cho những bài toán phát hiện đối tượng nói chung vì cân đối được cả yêu cầu về chất lượng cũng như tốc độ thực hiện.



Hình 1. Phát hiện dụng cụ phẫu thuật dựa trên YOLO

Như được biểu diễn trong hình 2, kiến trúc mạng YOLO bao gồm phần trích xuất đặc trưng (Feature Extractor) và phần phát hiện (Detector). Với đầu vào là một bức ảnh, sau khi qua khâu trích xuất đặc trưng, đầu ra sẽ là ba bản đồ đặc trưng (feature map) ở các tỉ lệ (scale) khác nhau. Sau đó, những bản đồ đặc trưng này sẽ được đưa đến khâu phát hiện để lấy được các thông tin về loại (class) và hộp bao quanh vật thể (bounding box).

Hình 2. Sơ đồ tổng quát kiến trúc mạng YOLO

Trong YOLOv3 [21], Darknet-53 được sử dụng làm feature Extractor để trích xuất các đặc trưng của một bức ảnh. Như được biểu diễn trong hình 3, Darknet-53 gồm có 23 khối dư (residual unit). Mỗi khối dư này gồm có một 3×3 và một 1×1 lớp tích chập (convolutional layer). Sau mỗi lớp tích chập là một batch normalization [22] và một hàm kích hoạt Leaky Relu [23]. Tại cuối mỗi khối dư, một phép cộng theo từng phần tử (element-wise) được thực hiện giữa vec-tơ đầu vào và vec-tơ đầu ra. Tiếp theo đó, sau mỗi khối dư là một lớp tích chập với bước nhảy là 2 để giảm kích thước bản đồ đặc trưng và do đó giảm số lượng tham số cho mô hình.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

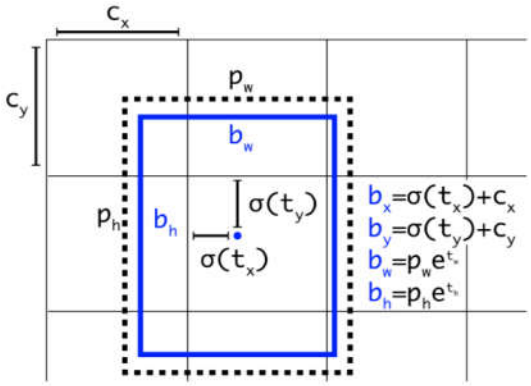
Hình 3. Kiến trúc Darknet-53 của YOLOv3

Khâu phát hiện của YOLOv3 sử dụng Feature Pyramid Network (FPN) [24] để phát hiện vật thể ở ba tỉ lệ khác nhau. Đối với ảnh đầu vào kích thước 416 × 416 thì đầu ra của YOLOv3 là ba bản đồ đầu ra (output map) có các kích thước 13 × 13, 26 × 26 và 52 × 52. Bản đồ đầu ra này có kích thước nhỏ được sử dụng để dự đoán những vật thể có kích thước lớn và những bản đồ đầu ra có kích thước lớn được sử dụng để dự đoán những vật thể có kích thước nhỏ. Mỗi ô lưới (grid cell) trên mỗi bản đồ đầu ra này sẽ dự đoán ba hộp bao quanh, như vậy số lượng hộp được dự đoán trên một bức ảnh sẽ là:

$$(13\times13+26\times26+52\times52)\times3=10647(\text{boxes}) \quad (1)$$

Để tìm được hộp bao quanh một vật thể trong một bức ảnh, YOLOv3 sử dụng các hộp mốc (anchor box) để làm cơ sở ước lượng. Những hộp mốc này sẽ được xác định trước và sẽ bao quanh vật thể một cách tương đối chính xác. Mỗi

một vật thể trong hình ảnh huấn luyện được phân bố về một hộp mốc. Trong trường hợp có từ hai hộp mốc trở lên cùng bao quanh vật thể thì hộp được lựa chọn là hộp có Intersection Over Union (IOU) với hộp sự thật (truth bounding box) là cao nhất.



Hình 4. Công thức ước lượng hộp bao quanh (màu xanh) từ hộp mốc (đường nét đứt) và ô lưới mà hộp đó thuộc về

Như được biểu diễn trong hình 4, một hộp mốc có kích thước (p_w, p_h) tại ô lưới nằm trên bản đồ đầu ra với góc trên cùng bên trái của nó là (c_x, c_y), YOLOv3 dự đoán bốn tham số (t_x, t_y, t_w, t_h), trong đó hai tham số đầu là độ lệch (offset) so với góc trên cùng bên trái của ô lưới và hai tham số sau là tỷ lệ so với hộp mốc. Các tham số này được sử dụng để xác định một hộp mốc với tọa độ tâm là (b_x, b_y) và kích thước là (b_w, b_h) theo công thức trong hình 4.

Đối với mỗi hộp dự đoán, YOLOv3 sẽ dự đoán xác suất mà hộp đó có chứa vật thể và xác suất lớp mà vật thể đó thuộc về. Do vậy, đầu ra của mô hình YOLOv3 là một véc-tơ sẽ bao gồm các thành phần sau:

$$y^T=\left[p_0,\langle t_x,t_y,t_w,t_h\rangle,\langle p_1,p_2,...p_c\rangle\right] \quad (2)$$

Trong đó, p_0 là là xác suất dự đoán vật thể xuất hiện trong hộp bao quanh.

$\langle t_x,t_y,t_w,t_h\rangle$ giúp xác định hộp bao quanh như được mô tả trong hình 4.

$\langle p_1,p_2,...p_c\rangle$ là véc tơ phân phối xác suất dự đoán của các lớp.

YOLOv3 có thể dự đoán ra rất nhiều hộp bao quanh có thể có trên một bức ảnh. Những ô lưới có vị trí gần nhau thì khả năng các hộp dự đoán bị chồng chéo là rất cao. Vì vậy, thuật toán non-max suppression (NMS) [25] được sử dụng để giảm bớt các hộp dự đoán này. NMS thực hiện theo hai bước như sau: Đầu tiên là loại bỏ các hộp có xác suất chứa vật thể nhỏ hơn 0,5. Sau đó lựa chọn những hộp có xác suất chứa vật thể là cao nhất và loại bỏ tất cả các hộp có IOU với hộp này lớn hơn một giá trị ngưỡng nào đó.

Quá trình huấn luyện của YOLOv3 là quá trình tối ưu hàm mất mát nhiều phần (multi-part loss function). Hàm mất mát này là tổng hàm mất mát của hộp dự đoán so với thực tế (L_{loc} - localization loss) và hàm mất mát của phân phối xác suất (L_{cls} - confidence loss):

$$L_{loc}=\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\begin{array}{l}\left(x_i-\hat{x}_i\right)^2+\left(y_i-\hat{y}_i\right)^2\\+\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\end{array}\right] \quad (3)$$

$$L_{cls}=\sum_{i=0}^{S^2}\sum_{j=0}^{B}\left[1_{ij}^{obj}+\lambda_{noobj}\left(1-1_{ij}^{noobj}\right)\right]\left(C_{ij}-\hat{C}_{ij}\right)^2+\sum_{i=0}^{1_{ij}^{obj}}\sum_{c\in C}1_i^{obj}\left(p_i(c)-\hat{p}_i(c)\right)^2 \quad (4)$$

Hàm mất mát: $L=L_{loc}+L_{cls}$ (5)

Trong đó:

$\left[(x,y),(w,h)\right]$: Kích thước ô mốc.

$\left[(\hat{x},\hat{y}),(\hat{w},\hat{h})\right]$: Kích thước ô dự đoán.

$1_i^{obj}=1$, nếu ô lưới thứ i có chứa vật thể.

$1_{ij}^{obj}=1$, nếu hộp thứ j của ô thứ i có chứa vật thể.

$1_{ij}^{noobj}=1$, nếu box thứ j của ô thứ i không chứa vật thể.

C_{ij}: Điểm tin cậy của ô thứ i.

$\hat{C}_{ij}$: Điểm tự tin dự đoán.

λ_{coord}, λ_{noobj}: Các hằng số điều chỉnh, có nhiệm vụ làm giảm giá trị của hàm mất mát.

$p_i(c)$: Xác suất có điều kiện: có hay không ô có chứa một đối tượng của lớp.

$\hat{p}_i(c)$: Xác suất có điều kiện dự đoán.

C: Tập hợp tất cả các lớp.

B: Số hộp dự đoán đối cho mỗi ô lưới.

S: Kích thước của feature map ở mỗi tỉ lệ.

L_{loc}: Hàm mất mát của hộp dự đoán so với thực tế.

L_{cls}: Hàm mất mát của phân phối xác suất.

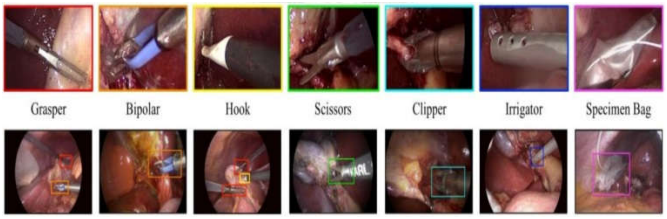
3. ĐÁNH GIÁ KẾT QUẢ

3.1. Tập dữ liệu

Để thực hiện nghiên cứu này, tác giả cần xây dựng một tập dữ liệu đủ lớn cho các loại dụng cụ với những chú thích (annotation) về tên và vị trí trong những hình ảnh nội soi. Giai đoạn nghiên cứu hiện tại, tập dữ liệu này vẫn chưa được hoàn thành. Vì vậy, trong bài báo này, tác giả sẽ sử dụng tập dữ liệu m2cai16-tool-locations được cung cấp trong [5] để huấn luyện và kiểm tra hiệu quả của mô hình đề xuất.

Tập dữ liệu m2cai16-tool-locations được xây dựng từ m2cai16-tool dataset [15] cho việc phát hiện bảy loại dụng cụ thường được sử dụng trong phẫu thuật nội soi cắt túi mật, như được biểu diễn trong hình 5. Tập dữ liệu m2cai16-tool-locations gồm có 2532 bức ảnh đã được gán nhãn về tên và tọa độ của những hộp bao quanh đầu của mỗi loại dụng cụ.

Bảng 1 mô tả số lượng ảnh cũng như tên và số lượng gán nhán của bảy loại dụng. Trong m2cai16-tool-locations, tác giả sử dụng dữ liệu được ghi từ video-1 tới video-7 cho việc huấn luyện (training), dữ liệu được ghi từ video-10 cho việc xác minh (validation) và dữ liệu được ghi từ video-8, video-9 cho việc kiểm tra (test).



Hình 5. Bảy loại dụng cụ trong phẫu thuật cắt túi mật (hàng trên) và chú thích về vị trí của đầu mỗi loại dụng cụ (hàng dưới)

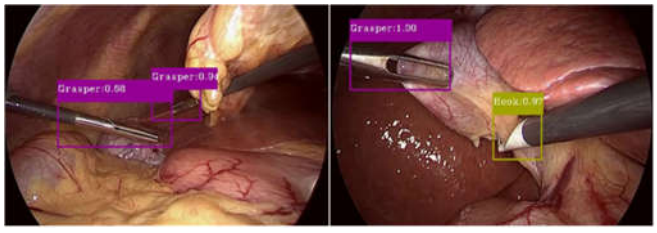
Bảng 1. Số lượng các hình ảnh đã được gán nhãn cho mỗi loại dụng cụ

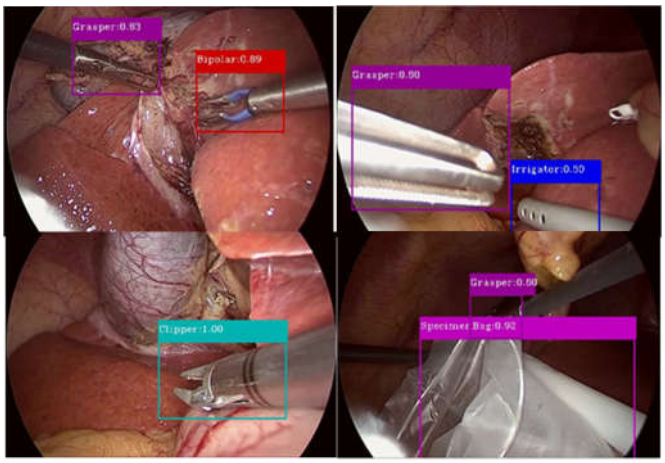
Dụng cụ	Số lượng chú thích
Grasper	923
Bipolar	350
Hook	308
Scissors	400
Clipper	400
Irrigator	485
Specimen Bag	275
Tổng chú thích	3141
Số lượng ảnh	2532

3.2. Kết quả đánh giá

Thí nghiệm được thực hiện trên hệ điều hành Ubuntu 16.04 với máy tính Intel i5-4590 CPU @ 3.40 GHz, RAM 16G và card màn hình GTX1060 Nvidia GPU.

Tác giả đã thực hiện chương trình dựa trên darknet framework [26], với các tham số được lựa chọn như sau: width = 416, height = 416 (kích thước ảnh đầu vào); classes = 7 (bảy loại dụng cụ) và filters = (classes+5)×3 = 36. Quá trình huấn luyện được thực hiện dựa trên tập training và validation như được mô tả trong phần (3.1). Sau đó, chúng tôi đã sử dụng dữ liệu được ghi từ hai video cắt túi mật (video-8, video-9) trong tập dữ liệu để xác nhận hiệu quả mô hình đề xuất. Hình 6 biểu diễn một số kết quả phát hiện dụng cụ phẫu thuật trên tập dữ liệu kiểm tra. Kết quả này cho thấy, mô hình đề xuất có thể nhận dạng và định vị trí đúng cho mỗi loại dụng cụ, mặc dù các dụng cụ này thường xuyên có sự thay đổi về hình dạng, hướng hoặc về góc nghiêng so với vị trí của camera quan sát. Mô hình cũng có thể phát hiện được dụng cụ đã bị che khuất một phần nào đó.



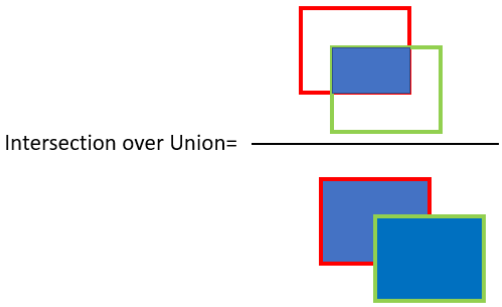


Hình 6. Một số kết quả phát hiện dụng cụ phẫu thuật. Mô hình có thể nhận dạng và định vị trí đúng cho mỗi loại dụng cụ mặc dù có sự thay đổi về hình dạng, hướng, góc và sự che khuất

Tiếp theo, tác giả thực hiện đánh giá định lượng hiệu quả của mô hình bằng các tham số recall, precision và mAP (mean Average Precision). Hình 7 mô tả khái niệm Intersection over Union (IoU), đó là tỉ lệ phần giao với phần hợp của hộp dự đoán và hộp sự thật. Công thức (6-7) mô tả định nghĩa các tham số recall và precision. Trong đó, True Positive (TP) là tổng số các phát hiện có IOU lớn hơn hoặc bằng 0,5; False Positive (FP) là tổng số các phát hiện có IOU nhỏ hơn 0,5; và False Negative (FN) là tổng số các không phát hiện được vật thể trong tập dữ liệu kiểm tra.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive+False positive}} \quad (6)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive+False Negative}} \quad (7)$$



Hình 7. Hộp sự thật (viền màu đỏ) và hộp dự đoán (viền màu xanh)

Bảng 2. Kết quả đánh giá hiệu quả của mô hình qua tham số recall và Precision

Testing Videos	Recall (%)	Precision (%)	FPS
video_08	78,5	85,2	25,5
video_09	80,6	90,5	25,3
Trung bình	79,55	87,85	25,4

Kết quả đánh giá cho tập dữ liệu kiểm tra thông qua tham số recall, precision và tốc độ khung hình (frame per second - FPS) được đưa ra trong bảng 2. Từ kết quả này có thể thấy rằng, khả năng phát hiện của mô hình (recall) là

khoảng 79,55% và tỉ lệ dự đoán chính xác của mô hình (precision) là khoảng 87,85%. Hơn nữa, mô hình có thể phát hiện được dụng cụ phẫu thuật ở thời gian thực với tốc độ khung hình là khoảng 25,4 (fps).

Bảng 3 mô tả kết quả phát hiện trung bình cho tất cả các loại dụng cụ phẫu thuật trong dữ liệu kiểm tra. Nhìn vào kết quả này có thể thấy rằng Bipolar và Irrigator có độ chính xác phát hiện là khá thấp. Điều này là do đặc điểm về hình dạng cũng như dữ liệu huấn luyện cho các hai loại dụng cụ này là chưa đủ. Với mô hình đề xuất, độ chính xác phát hiện trụng bình (mAP) cho các loại dụng cụ được xác định bằng 71,54%. Tỉ lệ này là khá cao khi được so sánh với các kết quả được công bố trong [15].

Bảng 3. Kết quả phát hiện dụng cụ trung bình (mAP) cho tất cả các dụng cụ phẫu thuật

Mô hình	Grasper	Bipolar	Hook	Scissor	Clipper	Irrigator	Specimanbag	mAP
YOLOv3	88,3	32,5	92,2	64,5	91,4	41,5	90,4	**71,54**

4. KẾT LUẬN

Trong bài báo này, tác giả đã giới thiệu về bài toán phát hiện dụng cụ phẫu thuật nội soi. Chúng tôi đã ứng dụng, kiểm tra và đánh giá hiệu quả của mô hình đề xuất dựa trên CNN (YOLOv3) đối với việc phát hiện bảy loại dụng cụ thường được sử dụng trong phẫu thuật nội soi cắt túi mật. Kết quả đánh giá cho thấy rằng precision là khoảng 87,85%, recall là khoảng 79,55%, mAP là khoảng 71,54% và tốc độ khung hình là khoảng 25,4 (fps).

Trong nghiên cứu tiếp theo, tác giả sẽ tăng cường tập dữ liệu hiện có thông qua các thuật toán xử lý ảnh (xoay, lật, kéo, dãn, làm mờ, làm bóng ảnh...). Hơn nữa, tác giả sẽ thu thập tập dữ liệu đủ lớn cho nhiều loại dụng cụ được sử dụng trong phẫu thuật nội soi nói chung, không chỉ riêng nội soi cắt túi mật. Trong bài báo này, tác giả mới chỉ ứng dụng mô hình YOLOv3 mà chưa có cải tiến nào. Vì vậy, trong nghiên cứu tiếp theo, tác giả sẽ cải thiện mô hình YOLOv3, đồng thời kết hợp thêm một số thuật toán xử lý ảnh, chẳng hạn như optical flow để nâng cao hiệu quả của sự phát hiện dụng cụ nội soi.

TÀI LIỆU THAM KHẢO

[1]. N. T. P. Dung (13/07/2018). *Loi ich cua viec mo noi soi*. Available: https://benh.vn/loi-ich-cua-viec-mo-noi-soi-4694/

[2]. M. Lan (13/2/2006). *Mo noi soi - lua chon so 1 cua bac si lan benh nhan*. Available: https://vnexpress.net/doi-song/mo-noi-soi-lua-chon-so-1-cua-bac-si-lan-benh-nhan-2261729.html

[3]. D. T. Kim, C. H. Cheng, D. G. Liu, K. C. J. Liu, W. S. W. Huang, 2019. *Designing a New Endoscope for Panoramic-View with Focus-Area 3D-Vision in Minimally Invasive Surgery*. Journal of Medical and Biological Engineering, pp. 1-16, 2019.

[4]. B. Münzer, K. Schoeffmann, L. Böszörmenyi, 2018. *Content-based processing and analysis of endoscopic images and videos: A survey*. Multimedia Tools and Applications, journal article vol. 77, no. 1, pp. 1323-1362.

[5]. A. Jin et al., 2018. *Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks.* 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 691-699.

[6]. L. Cheolwhan, W. Yuan-Fang, D. R. Uecker, W. Yulun, 1994. *Image analysis for automated tracking in robot-assisted endoscopic surgery.* in Proceedings of 12th International Conference on Pattern Recognition, vol. 1, pp. 88-92 vol.1.

[7]. K. Cai, R. Yang, Q. Lin, Z. Wang, 2016. *Tracking multiple surgical instruments in a near-infrared optical system.* Computer Assisted Surgery, vol. 21, pp. 46-55.

[8]. M. Kranzfelder et al., 2013. *Real-time instrument detection in minimally invasive surgery using radiofrequency identification technology*. The Journal of surgical research, vol. 185, 07/02 2013.

[9]. I. Laina et al., 2017. *Concurrent Segmentation and Localization for Tracking of Surgical Instruments.* International Conference on Medical Image Computing and Computer-Assisted Intervention.

[10]. A. Reiter, P. K. Allen, 2010. *An online learning approach to in-vivo tracking using synergistic features.* in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3441-3446.

[11]. D. Bouget, R. Benenson, M. Omran, L. Riffaud, B. Schiele, P. Jannin, 2015. *Detecting Surgical Tools by Modelling Local Appearance and Global Shape.* IEEE Transactions on Medical Imaging, vol. 34, pp. 1-1.

[12]. A. Reiter, P. K. Allen, T. Zhao, 2012. *Feature Classification for Tracking Articulated Surgical Tools.* Berlin, Heidelberg, pp. 592-600: Springer Berlin Heidelberg.

[13]. A. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, N. Padoy, 2016. *EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos*. IEEE Transactions on Medical Imaging, vol. 36..

[14]. M. Sahu, A. Mukhopadhyay, A. Szengel, S. Zachow, 2016. *Tool and Phase recognition using contextual CNN features.* arXiv:1610.08854 [cs.CV].

[15]. A. Raju, S. Wang, J. Huang, 2016. *M2CAI surgical tool detection challenge report*. University of Texas at Arlington, Tech. Rep.

[16]. A. P. Twinanda, D. Mutter, J. Marescaux, M. de Mathelin, N. Padoy, 2016. *Single-and multi-task architectures for tool presence detection challenge at M2CAI 2016.* arXiv preprint arXiv:1610.08851, 2016.

[17]. MCCAI, 2019. *Tool Presence Detection Challenge Result.*

[18]. R. B. Girshick, 2015. *Fast R-CNN.* 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015.

[19]. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, 2016. *You Only Look Once: Unified, Real-Time Object Detection.* in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788.

[20]. J. Redmon, A. Farhadi, 2017. *YOLO9000: Better, Faster, Stronger*. in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517-6525.

[21]. J. Redmon, A. Farhadi, 2018. *Yolov3: An incremental improvement.* arXiv preprint arXiv:1804.02767.

[22]. S. Ioffe, C. Szegedy, 2015. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.* ArXiv, vol. abs/1502.03167.

[23]. A. L. Maas, 2013. *Rectifier Nonlinearities Improve Neural Network Acoustic Models.*

[24]. T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, S. J. Belongie, 2017. *Feature Pyramid Networks for Object Detection.* 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936-944.

[25]. J. Hosang, R. Benenson, B. Schiele, 2017. Learning Non-maximum Suppression. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6469-6477

[26]. J. Redmon. Darknet, *Open Source Neural Networks in C. 2013–2016.* Available: https://pjreddie.com/darknet/Engieering, Hanoi University of Industry

AUTHOR INFORMATION

Kim Dinh Thai

International School, Vietnam National University, Hanoi