

# MÔ HÌNH HỌC SÂU VỚI SPARSE RCNN PHÁT HIỆN VÀ PHÂN LOẠI TẾ BÀO MÁU

Trần Đình Toàn<sup>1\*</sup>, Nguyễn Đức Toàn<sup>2</sup>

<sup>1</sup>Trường Đại học Công Thương Thành phố Hồ Chí Minh

<sup>2</sup>Công ty Cổ phần Y Chính Xác

\*Email: [toantd@huit.edu.vn](mailto:toantd@huit.edu.vn)

Ngày nhận bài: 16/10/2023; Ngày chấp nhận đăng: 08/12/2023

## TÓM TẮT

Nghiên cứu này đề xuất mô hình học sâu để phát hiện và phân loại tế bào máu như hồng cầu, bạch cầu và tiểu cầu dựa trên ảnh tế bào máu, sử dụng kiến trúc mạng SPARSE RCNN kết hợp các hàm mất mát GIOU, Focal và L1, nhằm hỗ trợ bác sĩ phát hiện sự phát triển bất thường của các loại tế bào máu, qua đó chẩn đoán sớm được các bệnh lý khác nhau như nhiễm trùng trong cơ thể, viêm họng, viêm phổi, viêm màng não và các bệnh lý khác. Ngoài ra, phát hiện sớm sự bất thường của các tế bào hồng cầu, bạch cầu và tiểu cầu có thể giúp bác sĩ đánh giá tổng thể về sức khỏe của bệnh nhân, chẳng hạn như mức độ thiếu máu hoặc bệnh lý liên quan đến hệ miễn dịch của cơ thể. Thử nghiệm được tiến hành dựa trên bộ dữ liệu Blood Cell Count and Detection - BCCD (COCO) lần lượt với các kiến trúc Single shot detector (SSD), SPARSE RCNN và kết hợp các hàm mất mát để giám sát, kết quả mô hình được xây dựng từ SPARSE RCNN đạt mAP là 0,591 tốt hơn mô hình SSD đạt mAP là 0,326 và các kết quả khác đã công bố.

*Từ khóa:* SSD, BLOOD CELL, SPARSE RCNN, BCCD.

## 1. MỞ ĐẦU

Các bệnh về máu có thể gây ra nhiều vấn đề về sức khỏe như suy giảm chức năng tế bào máu, tăng nguy cơ nhiễm trùng và các vấn đề về huyết áp [1, 2]. Bệnh lý về máu là nhóm bệnh liên quan đến hệ tuần hoàn và sản sinh máu trong cơ thể, một số bệnh thường gặp như: Thiếu máu là một căn bệnh khá phổ biến và xảy ra khi lượng tế bào hồng cầu (máu đỏ) trong cơ thể thiếu hụt, gây ra triệu chứng như mệt mỏi, hoa mắt, vàng da, suy nhược cơ thể, đặc biệt thai phụ khi bị thiếu máu sẽ tăng nguy cơ bị sinh non [3]. Bệnh bạch cầu là bệnh liên quan đến tế bào bạch cầu (máu trắng) phát triển bất thường thành ác tính, gồm viêm tủy xương và ung thư tủy xương, có hai loại bệnh bạch cầu là cấp tính và mãn tính [2, 4, 5]. Các bệnh về máu khác như thiếu máu bẩm sinh, bệnh Thalassemia, bệnh von Willebrand, bệnh bạch cầu loại trừ miễn dịch và bệnh Hodgkin [6]. Thông thường bệnh lý về máu rất khó phát hiện và dễ bị nhầm lẫn với các bệnh lý khác nên tùy thuộc vào từng loại bệnh lý với các triệu chứng mà cách điều trị có thể khác nhau. Việc giúp bác sĩ sớm đưa ra chẩn đoán bệnh chính xác và có phác đồ điều trị kịp thời là rất quan trọng nhằm giảm thiểu tác động của các bệnh lý này đến sức khỏe của bệnh nhân. Đó cũng là lý do tại sao việc phát hiện các tế bào máu phát triển bất thường thông qua các phương pháp xét nghiệm máu, soi đếm tế bào máu là một phương pháp chẩn đoán quan trọng và cần thiết [5, 7].

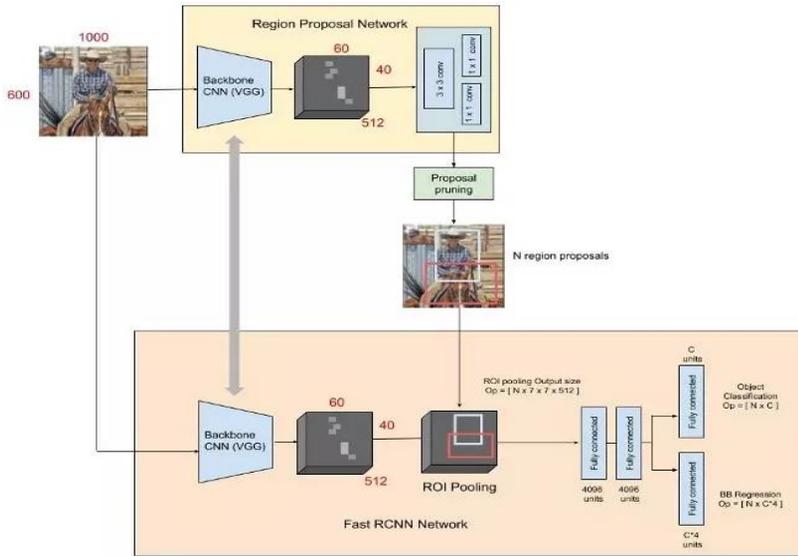
## 2. CÁC CÔNG TRÌNH LIÊN QUAN

Thời gian gần đây, có rất nhiều nghiên cứu ứng dụng khoa học máy tính, đặc biệt là học sâu đã trở thành một công cụ quan trọng được dùng vào phát hiện đối tượng trong ảnh y khoa. Các nghiên cứu chứng minh việc sử dụng các thuật toán học sâu để phát hiện đối tượng trong ảnh y khoa có thể đem lại kết quả rất tốt và đáng tin cậy. Chẳng hạn như nghiên cứu của Wang cùng cộng sự đã sử dụng mô hình Faster RCNN để phát hiện u xơ tử cung [8]. Liu cùng nhóm nghiên cứu phát triển một mô hình học sâu để phát hiện bệnh nhân ung thư gan dựa trên ảnh CT [9]. Y Sharifi và nhóm nghiên cứu đã sử dụng kỹ thuật học sâu để phát hiện các khối u trong ảnh siêu âm tuyến giáp [10]. L Zhang cùng cộng sự phát triển mô hình học sâu để phát hiện và định vị các khối u trên ảnh siêu âm buồng trứng [11]. Nghiên cứu của S Kulhare và cộng sự đã sử dụng kỹ thuật học sâu để phát hiện các khối u trên màng phổi [12]. Nhóm nghiên cứu của Bruno đã đề xuất mô hình mạng neuron tích chập đa tầng để phân loại các khối u trực tràng trên ảnh toàn bộ lát cắt mô [13]. Sara Hosseinzadeh Kassani cùng nhóm nghiên cứu đã đề xuất mô hình học sâu phát hiện và phân loại sự phát triển bất thường của bạch cầu dựa trên ảnh soi tế bào máu qua kính hiển vi [14]. Hagar Mohamed cùng nhóm nghiên cứu đã đề xuất hai mô hình với kiến trúc VGG16 và denseNet121 để phân loại các bệnh ung thư máu [15]. M Samir Abou El-Seoud cùng nhóm nghiên cứu đã sử dụng mạng CNN (Convolutional neural networks) để phát hiện và phân loại các loại tế bào máu trắng [16]. Chao Chen cùng nhóm nghiên cứu sử dụng mạng học sâu để phân tích tế bào máu [17]. David Opeoluwa Oyewola cùng cộng sự đã đề xuất mạng tăng cường dữ liệu DACNN (Data Augmentation Convolutional Neural Network) phân tích tế bào máu để xác định bệnh sốt rét [18]. Maria Delgado-Ortet cùng nhóm nghiên cứu đã sử dụng mạng CNN phân đoạn tế bào máu xác định có hay không nhiễm bệnh sốt rét [19]. Ashish Girdhar cùng cộng sự sử dụng mạng CNN để phân loại tế bào bạch cầu [20]. Hüseyin Kutlu cùng cộng sự đã sử dụng các kiến trúc mạng học sâu như RCNN (Region based Convolutional Neural Networks), AlexNet, VGG16, GoogLeNet, ResNet50 kết hợp học chuyển tiếp để phát hiện và phân loại bạch cầu [21].

Bên cạnh đó cũng có nhiều công trình trong học sâu được đề xuất để phát triển bài toán phát hiện đối tượng, cụ thể Ross Girshick cùng nhóm nghiên cứu đã đề xuất mạng RCNN để giải quyết bài toán phát hiện đối tượng [22]. Tiếp theo, Fast RCNN là kiến trúc mạng neuron đã được Ross Girshick cải tiến về mặt tốc độ so với RCNN [23]. Shaoqing Ren cùng cộng sự đề xuất kiến trúc mạng Faster RCNN [24]. Kiến trúc “You Only Look Once: Unified, Real-Time Object Detection” viết tắt là YOLO của Joseph Redmon cùng nhóm nghiên cứu đề xuất là một phương pháp tiếp cận thống nhất cho việc phát hiện đối tượng thời gian thực. Phương pháp này sử dụng một mô hình mạng neuron đơn giản để xác định các vùng quan tâm (ROI) trên ảnh, dự đoán lớp và hộp giới hạn cho các đối tượng trong vùng đó [25]. Kiến trúc RetinaNet là một mô hình phát hiện đối tượng tiên tiến được giới thiệu bởi Tsung-Yi Lin cùng nhóm nghiên cứu [26]. NAS-FPN (Neural Architecture Search Feature Pyramid Network) là một kiến trúc mạng neuron tiên tiến được giới thiệu bởi Golnaz Ghiasi cùng nhóm nghiên cứu. Mô hình này sử dụng kỹ thuật Neural Architecture Search (NAS) để tìm kiếm kiến trúc mạng neuron tối ưu cho việc trích xuất đặc trưng trong phát hiện đối tượng. NAS-FPN kết hợp hai thành phần chính: một mô-đun phát hiện đối tượng trên các đặc trưng đa tầng được tạo ra bởi FPN và một mô-đun NAS được sử dụng để tìm kiếm kiến trúc phù hợp nhất cho mô-đun phát hiện đối tượng. Mô-đun NAS được áp dụng trên tập dữ liệu huấn luyện để tìm kiếm kiến trúc mạng neuron tối ưu cho việc phát hiện đối tượng [27]. YOLOv3 là một trong những mô hình phát hiện đối tượng tiên tiến được giới thiệu bởi Joseph Redmon cùng nhóm nghiên cứu vào năm 2018 [28].

Các nghiên cứu trên cho thấy tiềm năng của học sâu dựa trên dữ liệu y khoa, đặc biệt là dữ liệu ảnh trong việc hỗ trợ bác sĩ phát hiện và chẩn đoán các bệnh lý, đồng thời khẳng định

rằng phương pháp này có thể giúp nâng cao độ chính xác và tốc độ của quá trình chẩn đoán bệnh đặc biệt là các bệnh liên quan đến phân tích tế bào máu.



Hình 1. Kiến trúc Fast RCNN Network

Để giải quyết vấn đề đặt ra, bài báo này tập trung giải quyết 2 công việc cụ thể như sau:

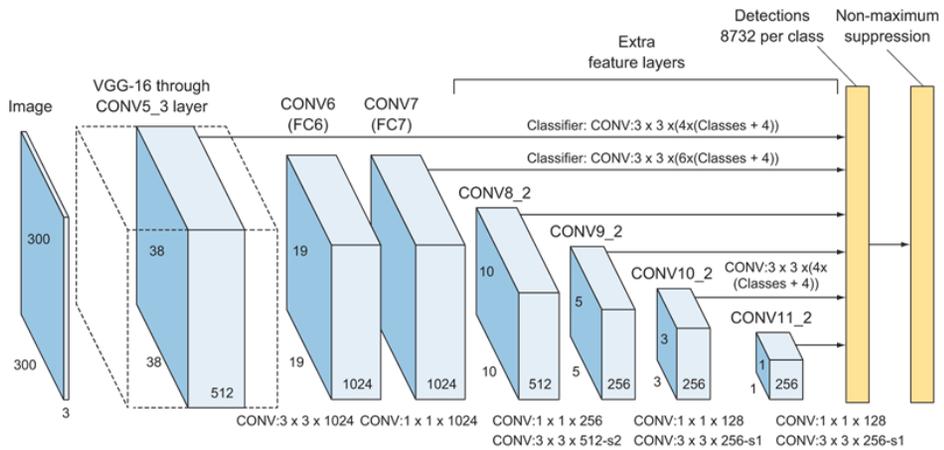
- Xây dựng mô hình học sâu dựa trên kiến trúc mạng RCNN kết hợp kiến trúc ma trận “thưa” SPARSE, để phát hiện và phân loại đối tượng dựa trên ảnh tế bào máu.
- Thực nghiệm: 1. Tiến hành với mô hình xây dựng từ kiến trúc SPARSE RCNN; 2. Tiến hành với mô hình được xây dựng từ kiến trúc SSD, cả 2 mô hình này đều kết hợp với các hàm mất mát: GIOU, Focal và L1 để giám sát tất cả các quá trình phát hiện và phân loại đối tượng; 3. Tiến hành với các mô hình được xây dựng trên kiến trúc YOLOF (You Only Live Once - level Feature) và YOLOv3 (phiên bản thứ ba của thuật toán YOLO) của các công trình khác đã công bố, đồng thời cũng tiến hành trên các phiên bản YOLOv7 và YOLOv8. Thực nghiệm trên cùng bộ dữ liệu ảnh tế bào máu BCCD (COCO).

Các kết quả này được trình bày trong phần 3, phần 4 minh họa những kết quả thực nghiệm cũng như các đánh giá. Phần cuối cùng là một vài kết luận và những hướng nghiên cứu còn bỏ ngỏ khi đề cập đến bài toán hỗ trợ phát hiện và phân loại tế bào máu.

### 3. PHƯƠNG PHÁP NGHIÊN CỨU

#### 3.1. SSD (Single Shot Multibox Detector)

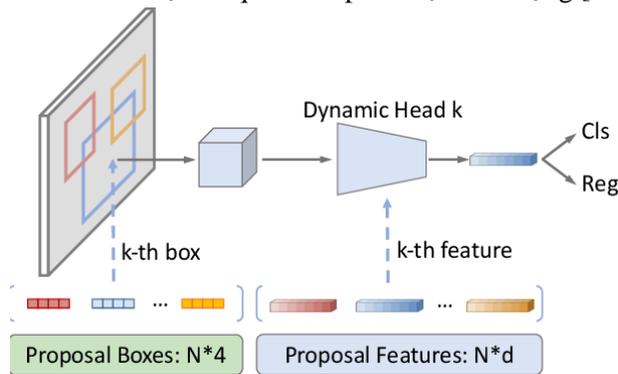
SSD là một thuật toán phát hiện đối tượng trong ảnh với tốc độ nhanh và độ chính xác cao. Thuật toán này sử dụng một mạng neuron tích chập (CNN) để trích xuất các đặc trưng của ảnh, sau đó sử dụng một bộ phân loại để xác định đối tượng có mặt trong ảnh và vị trí của nó. Điểm mạnh của SSD là tốc độ nhanh và khả năng phát hiện đa đối tượng trên cùng một ảnh. SSD sử dụng kỹ thuật anchor boxes để giảm số lượng vùng quan tâm cần phân loại. Kỹ thuật này giúp thuật toán tập trung vào các khu vực quan trọng trong ảnh, giảm thiểu thời gian xử lý và tăng độ chính xác. Tuy nhiên, điểm yếu của SSD là khả năng phát hiện đối tượng nhỏ hơn không tốt và thường cho kết quả không chính xác trên các đối tượng có hình dạng phức tạp. Do đó, SSD thường được sử dụng trong các ứng dụng yêu cầu tốc độ phát hiện nhanh [29, 30].



Hình 2. Kiến trúc SSD

### 3.2. SPARSE RCNN

Sparse RCNN là một mô hình mạng neuron sử dụng trong bài toán phát hiện đối tượng trong ảnh, được phát triển dựa trên RCNN. Sparse RCNN có khả năng phát hiện các đối tượng trong ảnh một cách nhanh chóng và hiệu quả hơn so với RCNN truyền thống. Khác với RCNN, Sparse RCNN sử dụng một cấu trúc mạng học sâu đơn giản hơn và tận dụng tính hiệu quả của việc thực hiện phép tích chập một cách "thưa" (sparse) trên ảnh đầu vào. Điều này giúp giảm đáng kể số lượng vùng quan tâm (region of interest - ROI) cần được xử lý và giúp tăng tốc quá trình huấn luyện. Đối với RCNN truyền thống, quá trình phát hiện đối tượng được chia thành 2 giai đoạn: Tìm các khu vực tiềm năng chứa đối tượng trong ảnh (region proposals). Phân loại đối tượng trong các khu vực tiềm năng đó bằng cách sử dụng một mạng CNN. Tuy nhiên, việc tìm kiếm các khu vực tiềm năng gặp nhiều vấn đề vì số lượng khu vực tiềm năng có thể lên đến hàng trăm nghìn trong một bức ảnh lớn. Để giải quyết vấn đề này, Sparse RCNN sử dụng một ma trận thưa, chỉ lưu trữ thông tin của một số khu vực tiềm năng quan trọng nhất (thông qua giá trị đánh dấu) để giảm độ phức tạp tính toán. Đồng thời, Sparse RCNN sử dụng một mô hình thuộc loại one-stage detector, tức là chỉ phân loại đối tượng một lần trực tiếp trên toàn bức ảnh, thay vì phân loại trên từng khu vực tiềm năng như RCNN truyền thống. Ngoài ra, Sparse RCNN cũng sử dụng một thuật toán mới để chọn và đánh giá các vùng quan tâm cần xử lý, giúp tăng độ chính xác và hiệu quả cho quá trình phát hiện đối tượng. Sparse RCNN là một mô hình phát hiện đối tượng trong ảnh cải tiến dựa trên RCNN, với tính năng "thưa" để giảm số lượng vùng quan tâm cần xử lý và thuật toán tối ưu hóa quá trình đánh giá vùng quan tâm, giúp tăng độ chính xác và tốc độ của quá trình phát hiện đối tượng [31].



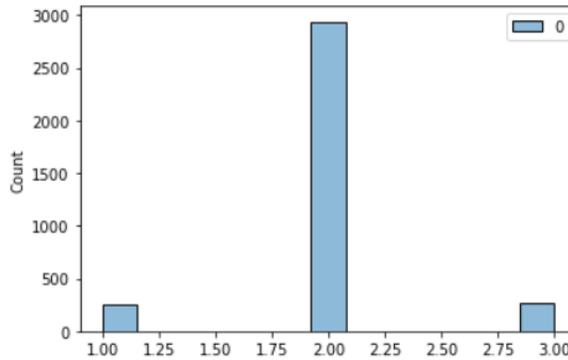
Hình 3. Kiến trúc Sparse RCNN

## 4. THỰC NGHIỆM VÀ KẾT QUẢ

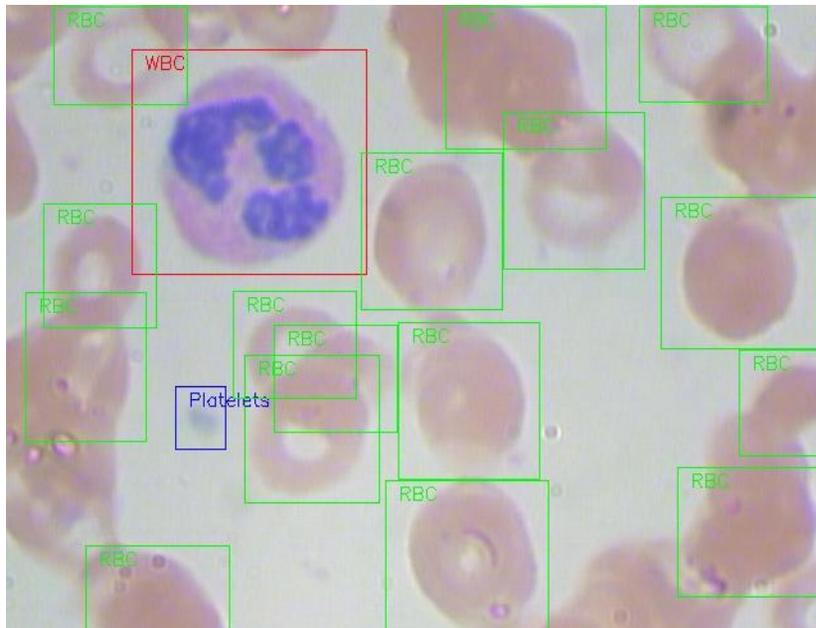
### 4.1. Dữ liệu

Bộ dữ liệu BCCD (COCO) chứa ảnh các tế bào máu, gồm 364 ảnh về hồng cầu, bạch cầu, tiểu cầu. Mỗi ảnh có kích thước  $640 \times 480$  pixel và có độ phân giải là 2 megapixel [32].

Bộ dữ liệu này gồm: 249 nhãn về tiểu cầu, 2938 nhãn hồng cầu, 263 nhãn là bạch cầu.



Hình 4. Trực quan hóa các lớp ảnh trong bộ dữ liệu BCCD (COCO)



Hình 5. Một ảnh đa đối tượng trong bộ dữ liệu BCCD (COCO)

### 4.2. Thực nghiệm

Bộ dữ liệu BCCD được chia thành ba phần gồm train, validation, test. Trong đó, 256 ảnh cho tập train, 74 ảnh dùng validation và 34 ảnh dùng để test.

Tiền xử lý dữ liệu sử dụng các kỹ thuật random crop, reszie, random flip, brightness transform, gaussian transform, saturation transform, hue transform.

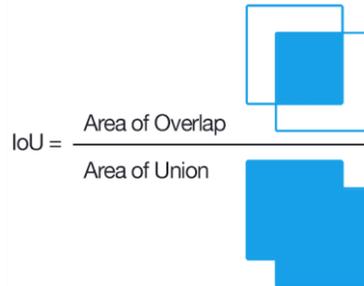
Với kiến trúc SSD và SPARSE RCNN, lần lượt thực hiện các chiến lược huấn luyện để phù hợp với từng mô hình, gồm các thông số như sau:

Kiến trúc SSD: Sử dụng optimizer SGD với learning rate là 0.02, momentum = 0.9, weight decay = 0.0005.

Kiến trúc SPARSE RCNN: Sử dụng optimizer ADAM với learning rate là 0.00005, weight decay = 0.001.

Trong nghiên cứu này, sử dụng độ đo AP (Average Precision) là độ đo dùng để kiểm tra độ chính xác của các phương pháp phát hiện đối tượng. IoU là độ đo dùng để kiểm tra độ khớp nhau giữa bounding box được dự đoán của đối tượng và bounding box thật của đối tượng đó.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$



Hình 6. Minh họa cho IoU

Với mỗi ngưỡng IoU xác định được một cặp precision và recall. Precision là tỉ lệ số lượng bounding box được dự đoán đúng trên tổng số bounding box được dự đoán. Recall là tỉ lệ số lượng bounding box được dự đoán đúng trên tổng số bounding box ground truth. Từ cặp precision và recall này, tính được AP và AP [0,1].

Bài viết này đánh giá kết quả mỗi phương pháp bằng cách tính  $AP^{IoU=0.5:0.95}$  và mAP là trung bình AP trên toàn bộ class.

Khi quan sát bộ dữ liệu BCCD cho thấy dữ liệu bị mất cân bằng khá lớn ở các lớp, mất cân bằng lớn giữa nhãn hồng cầu so với hai nhãn còn lại, và vị trí của hồng cầu bị trùng lặp khá nhiều dẫn đến mô hình có thể học lệch qua nhãn của hồng cầu, và số lượng bounding box dự đoán sẽ bị thiếu vì các trường hợp trùng lặp có thể bị bỏ qua.

Do đó, nghiên cứu này kết hợp ba hàm mất mát nhằm giải quyết các vấn đề trên: GIOU loss, Focal loss và L1 loss với chức năng giám sát từng nhiệm vụ của các kiến trúc mạng này. Cụ thể như sau:

Hàm mất mát GIOU: giúp mô hình có giám sát tốt hơn về độ tương quan của các bounding box và độ trùng lặp của các bounding box làm giảm thiểu sự thiếu sót khi dự đoán.

Gọi A và B lần lượt là hai bounding box Ground truth và dự đoán của mô hình

$$GIOU\ Loss = \frac{1}{N} \sum_{i=1}^N 1 - \left( IoU - 1 + \frac{|A \cup B|}{|C|} \right)$$

Trong đó, C là diện tích nhỏ nhất để một hình chữ nhật có thể bao quanh cả 2 bounding box A và B.

Hàm mất mát Focal: giúp cho mô hình không bị học lệch quá nhiều với nhãn hồng cầu.

$$Focal\ loss = \frac{1}{N} \sum_{i=1}^N -\alpha(1 - p_t)^\gamma \log(p_t)$$

Trong đó,  $\alpha$  và  $\gamma$  là hai siêu tham số,  $p_t$  là xác suất phân lớp.

Hàm mất mát L1:

$$L1 \text{ loss} = \frac{1}{N} \sum_{i=1}^N |y - \hat{y}|$$

Trong đó,  $y$  và  $\hat{y}$  lần lượt là nhãn và dự đoán của khối Regression tọa độ bounding box của kiến trúc này.

Ngoài thực nghiệm trên 2 kiến trúc SSD và SPARSE RCNN, trong nghiên cứu này cũng tiến hành thực nghiệm với kiến trúc YOLOv8, YOLOv7, YOLOF và YOLOv3 của các công trình đã công bố trước đó để có cơ sở đối sánh với kết quả nghiên cứu.

### 4.3. Kết quả

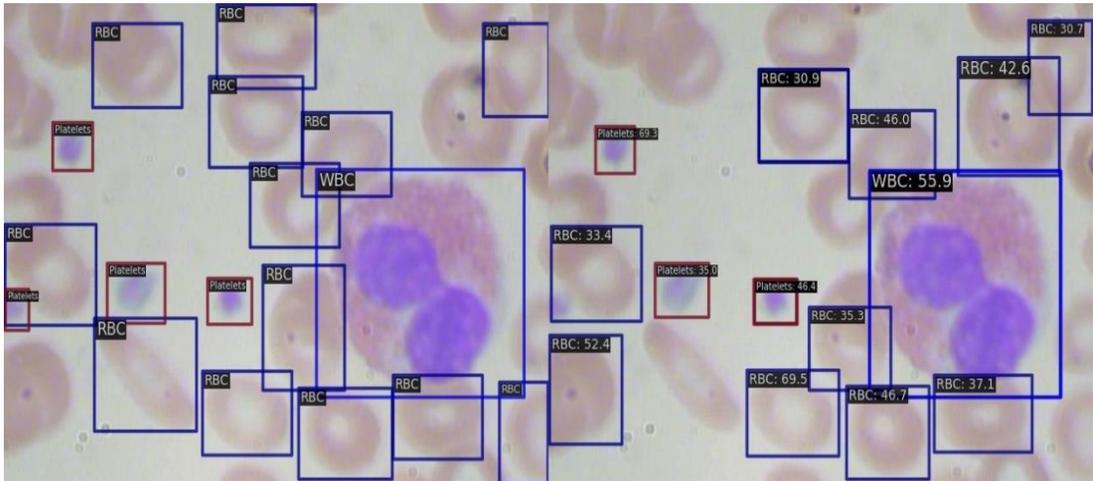
Từ kết quả các thực nghiệm (Bảng 1) cho thấy mô hình SPARSE RCNN nhờ tận dụng ma trận SPARSE để giảm được độ phức tạp của tính toán, tập trung vào các vùng cần quan tâm (ROI) và loại bớt các thông tin không quan trọng, kết hợp với Branch-out Network được sử dụng để dự đoán vị trí của các đối tượng trên những khu vực không được quan tâm đến trong bước lấy mẫu. Các dự đoán này được tính bằng cách sử dụng thông tin từ các khu vực quan tâm khác trên ảnh thông qua một mạng neuron riêng biệt và được sử dụng để cập nhật các trọng số của mô hình chính. Điều này giúp cho mô hình SPARSE RCNN dự đoán chính xác và tốt hơn so với mô hình được xây dựng từ kiến trúc SSD thiên về tốc độ xử lý hơn hiệu quả tính toán.

Bảng 1. Kết quả thực nghiệm và đối sánh

	<b>mAP 50</b>	<b>mAP 75</b>	<b>mAP</b>	<b>time</b>
<b>SSD</b>	0,638	0,278	0,326	0,137
<b>SPARSE RCNN</b>	0,831	0,715	<b>0,591</b>	0,343
YOLOv8	0,917	0,700	0,629	
YOLOv7	0,880	0,677	0,583	
YOLOF	0,890	0,657	0,575	
YOLOv3	0,867	0,453	0,459	

Ngoài ra, thực nghiệm cũng cho thấy kiến trúc được sử dụng trong nghiên cứu này cho kết quả tốt hơn so với các kiến trúc YOLOF và YOLOv3 được thực nghiệm trên cùng bộ dữ liệu đã công bố trước đó [33]. Và không thể so sánh với các kết quả đã công bố khác do sử dụng khác bộ dữ liệu.

Nghiên cứu này cũng thực nghiệm với YOLOv7 và YOLOv8 đây là hai kiến trúc mới được công bố năm 2022 và 2023. Kết quả độ chính xác tổng thể mAP của YOLOv7 và YOLOv8 lần lượt đã đạt được và vượt qua mức mAP của SPARSE RCNN kết hợp ba hàm mất mát. Tuy nhiên, nếu xét kỹ về mAP75 cho thấy SPARSE RCNN vẫn tốt hơn. Với các version mô hình YOLO thường được dùng cho các bài toán real-time với kiểu kiến trúc one-stage nhằm đạt được mAP50 cao. Tuy nhiên, với bài toán y khoa thông thường vị trí hay độ trùng lặp của đối tượng cần được khớp và chính xác hơn so với những bài toán khác nên người ta thường quan tâm đến độ đo mAP75 và SPARSE RCNN vẫn đạt được hiệu quả cao với kiến trúc Two-stage.



Hình 7. Bên trái là nhân, bên phải là ảnh được dự đoán từ mô hình SPARESE RCNN

## 5. KẾT LUẬN

Nghiên cứu này, đã đề xuất mô hình phát hiện và phân loại tế bào máu được xây dựng lần lượt dựa trên kiến trúc SPARSE RCNN và SSD kết hợp với 3 hàm loss GIOU, Facal, L1 với mục đích giám sát từng nhiệm vụ thực hiện của 2 kiến trúc này. Các thực nghiệm được tiến hành trên bộ dữ liệu ảnh tế bào máu BCCD (COCO). Trong đó thực nghiệm 2 với SPARSE RCNN cho kết quả tốt hơn về độ chính xác trên tất cả các độ đo. Ngoại trừ về tốc độ, kiến trúc SPARSE RCNN cho thấy chậm hơn so với kiến trúc SSD, tuy nhiên kết quả ở Bảng 1 có thể chấp nhận được và có thể sử dụng để xây dựng các ứng dụng yêu cầu về thời gian thực. Ngoài ra, kết quả thực nghiệm cũng cho thấy mô hình đề xuất trong nghiên cứu này cho kết quả tốt hơn so với các công trình khác đã công bố.

Hướng tiếp theo, sẽ tiếp tục nghiên cứu và cải tiến phương pháp phát hiện đối tượng để có độ chính xác cao hơn và thời gian nhanh hơn có thể đáp ứng được nhu cầu nghiên cứu và ứng dụng thực tế cho các bài toán liên quan dữ liệu y khoa, cũng như các bài toán áp dụng kỹ thuật phát hiện đối tượng khác.

## TÀI LIỆU THAM KHẢO

1. Yi Hyeon Gyu. - Nhận biết bất thường về bệnh lý máu. <https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/suc-khoe-tong-quat/nhan-biet-bat-thuong-ve-benh-ly-mau/>
2. - Kể tên các bệnh về máu và triệu chứng nhận biết. <https://medlatec.vn/tin-tuc/ke-ten-cac-benh-ve-mau-va-trieu-chung-nhan-biet-s195-n22770>
3. Ahmar Urooj Zaidi, Steven Buck, Manisha Gadgeel, Miguel Herrera-Martinez, Araathi Mohan, Kenya Johnson, Shruti Bagla, Robert M. Johnson, Yaddanapudi Ravindranath. - Clinical diagnosis of red cell membrane disorders: comparison of osmotic gradient ektacytometry and eosin maleimide (ema) fluorescence test for red cell band 3 (ae1, slc4a1) content for clinical diagnosis. *Front Physiol* **11** (2020). <https://doi.org/10.3389/fphys.2020.00636>
4. Rối loạn tế bào máu: Triệu chứng, phân loại và nguyên nhân. <https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/suc-khoe-tong-quat/roi-loan-te-bao-mau-trieu-chung-phan-loai-va-nguyen-nhan/>

5. Australia H. - Blood and blood vessels. (2021). <https://www.healthdirect.gov.au/blood-and-blood-vessels>
6. Thalassemia, von Willebrand, Hodgkin. <https://www.msdmanuals.com/vi-vn/chuy%C3%AAn-gia/SearchResults?query=thalassemia>
7. Ứng dụng trí tuệ nhân tạo trong phân tích hình thái tế bào máu - Bước tiến mới trong chuyên khoa huyết học. <https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/suc-khoe-tong-quat/ung-dung-tri-tue-nhan-tao-trong-phan-tich-hinh-thai-te-bao-mau-buoc-tien-moi-trong-chuyen-khoa-huyet-hoc/>
8. Aihua Zhao, Jian Zhang, Shixuan Wang, Yan Wang, Xin Zhu, Wenfeng Shen, Wenwen Wang. - Detection of hysteroscopic hysteromyoma in real-time based on deep learning. *Journal of Physics: Conference Series* **1861** (2021) 012079. <https://iopscience.iop.org/article/10.1088/1742-6596/1861/1/012079/pdf>
9. Hameedur Rahman, Tanvir Fatima Naik Bukht, Azhar Imran, Junaid Tariq, Shanshan Tu, Abdulkareem Alzahrani. - A deep learning approach for liver and tumor segmentation in CT images using resunet. *Bioengineering* **9** (8) (2022). <https://doi.org/10.3390/bioengineering9080368>
10. Sharifi Y, Bakhshali MA, Dehghani T, DanaiAshgzari M, Sargolzaei M, Eslami S. - Deep learning on ultrasound images of thyroid nodules. *Biocybernetics and Biomedical Engineering* **41** (1) (2021) 636-655. <https://doi.org/10.1016/j.bbe.2021.02.008>
11. Zhang L, Huang J, Liu L. - Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color ultrasound detecting system. *Journal of Medical Systems* **43** (8) (2019) 1-9. <https://doi.org/10.1007/s10916-019-1356-8>
12. Sourabh Kulhare, Xinliang Zheng, Courosh Mehanian, Cynthia Gregory, Meihua Zhu, Kenton Gregory, Hua Xie, James McAndrew Jones, Benjamin Wilson. - Ultrasound-based detection of lung abnormalities using single shot detection convolutional neural networks. *Lecture Notes in Computer Science*, 11042 LNCS, (2018) 65-73. [https://doi.org/10.1007/978-3-030-01045-4\\_8](https://doi.org/10.1007/978-3-030-01045-4_8)
13. Bruno Korbar, Andrea M. Olofson, Allen P. Miraflor, Catherine M. Nicka, Matthew A. Suriawinata, Lorenzo Torresani, Arief A. Suriawinata, Saeed Hassanpour - Deep learning for classification of colorectal polyps on whole-slide images. *Journal of pathology informatics* **8** (1) (2017). [https://doi.org/10.4103/jpi.jpi\\_34\\_17](https://doi.org/10.4103/jpi.jpi_34_17)
14. Kassani SH, kassani PH, Wesolowski MJ, Schneider KA, Deters R. - A hybrid deep learning architecture for leukemic b-lymphoblast classification. *Electrical Engineering and Systems Science* (2019). <https://doi.org/10.48550/arXiv.1909.11866>
15. Mohamed H, Kamal Elsheref F, Kamal SR. - A new model for blood cancer classification based on deep learning techniques. *International Journal of Advanced Computer Science and Applications* **14** (6) (2023). <http://dx.doi.org/10.14569/IJACSA.2023.0140645>
16. M Samir Abou El-Seoud. - Detection and classification of white blood cells through deep learning techniques. *iJOE*. **16** (15) (2020). <https://doi.org/10.3991/ijoe.v16i15.15481>
17. Chao Chen, Yuanjie Gu, Zhibo Xiao, Hailun Wang, Xiaoliang He, Zhilong Jiang, Yan Kong, Cheng Liu, Liang Xue, Javier Vargas, Shouyu Wang. - Automatic whole blood cell analysis from blood smear using label-free multi-modal imaging with deep neural networks. *Anal Chim Acta* **1229** (2022). <https://doi.org/10.1016/j.aca.2022.340401>

18. Oyewola DO, Dada EG, Misra S, Damaševičius R. - A novel data augmentation convolutional neural network for detecting malaria parasite in blood smear images. *Applied Artificial Intelligence*. **36** (1) (2022). <https://doi.org/10.1080/08839514.2022.2033473>
19. Delgado-Ortet M, Molina A, Alférez S, Rodellar J, Merino A. - A deep learning approach for segmentation of red blood cell images and malaria detection. *Entropy*. **22** (6) (2020) 1-16. <https://doi.org/10.3390/e22060657>
20. Girdhar A, Kapur H, Kumar V. - Classification of white blood cell using convolution neural network. *Biomed Signal Process Control* **71** (2022). <https://doi.org/10.1016/j.bspc.2021.103156>
21. Kutlu H, Avci E, Özyurt F. - White blood cells detection and classification based on regional convolutional neural networks. *Med Hypotheses* **135** (2020). <https://doi.org/10.1016/j.mehy.2019.109472>
22. Girshick R, Donahue J, Darrell T, Malik J. - Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2013) 580-587. <https://doi.org/10.48550/arXiv.1311.2524>
23. Girshick R. - Fast R-CNN. (2015). <https://doi.org/10.48550/arXiv.1504.08083>
24. Ren S, He K, Girshick R, Sun J. - Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39** (6) (2015) 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
25. Redmon J, Divvala S, Girshick R, Farhadi A. - You only look once: unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2016) 779-788. <https://doi.org/10.48550/arXiv.1506.02640>
26. Lin TY, Goyal P, Girshick R, He K, Dollar P. - Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42** (2) (2017) 318-327. <https://doi.org/10.48550/arXiv.1708.02002>
27. Ghiasi G, Lin TY, Le Q V. - NAS-FPN: learning scalable feature pyramid architecture for object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2019) 7029-7038. <https://doi.org/10.1109/CVPR.2019.00720>
28. Redmon J, Farhadi A. - YOLOv3: an incremental improvement. (2018). <https://doi.org/10.48550/arXiv.1804.02767>
29. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg. - SSD: single shot multibox detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9905 LNCS (2015) 21-37. <https://doi.org/10.48550/arXiv.1512.02325>
30. Li J, Hou Q, Xing J, Ju J. - SSD object detection model based on multi-frequency feature theory. *IEEE Access*, **8** (2020) 82294-82305. <https://doi.org/10.1109/ACCESS.2020.2990477>
31. Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka, Lei Li, Zehuan Yuan, Changhu Wang, Ping Luo. - Sparse R-CNN: End-to-end object detection with learnable proposals. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2020) 14449-14458. <https://doi.org/10.48550/arXiv.2011.12450>

32. BCCD (COCO). <https://www.kaggle.com/datasets/ammarnassanalhajali/bccd-coco>
33. Xu F, Li X, Yang H, Wang Y, Xiang W. - TE-YOLOF: Tiny and efficient YOLOF for blood cell detection. arXiv (2021) 1-10. <https://doi.org/10.48550/arXiv.2108.12313>

## ABSTRACT

### DEEP LEARNING MODEL WITH SPARSE RCNN DETECTS AND CLASSIFIES BLOOD CELLS

Tran Dinh Toan<sup>1\*</sup>, Nguyen Duc Toan<sup>2</sup>

<sup>1</sup>*Ho Chi Minh City University of Industry and Trade*

<sup>2</sup>*Precision Medical Joint Stock Company*

\*Email: [toantd@huit.edu.vn](mailto:toantd@huit.edu.vn)

This study proposes a deep learning model to detect and classify blood cells such as red blood, white blood, and platelets based on blood cell images, using the SPARSE RCNN network architecture and combining functions loss of GIOU, Focal, and L1, to assist doctors in detecting abnormal development of blood cell types, thereby early diagnosing various diseases such as infections in the body, sore throat, pneumonia, meningitis and other pathologies. In addition, early detection of abnormalities in red blood, white blood, and platelets can help doctors evaluate the patient's overall health, such as the degree of anemia or disease related to the body's immune system. Experiments were conducted based on the Blood Cell Count and Detection - BCCD (COCO) dataset with SSD, SPARSE RCNN architectures, and combined loss functions for monitoring, model results are built from SPARSE RCNN achieving an mAP of 0.591, better than the SSD model achieving an mAP of 0.326 and other published results.

*Keywords:* SSD, BLOOD CELL, SPARSE RCNN, BCCD.