

Phân đoạn cấu trúc răng trong ảnh x-quang toàn cảnh (panoramic X-ray)

I. Tổng quan

Trong nha khoa, kiểm tra phóng xạ giúp các chuyên gia bằng cách hiển thị cấu trúc xương răng, qua đó có thể kiểm tra mức độ bất thường, gãy xương, khối u,... Tuy nhiên, chỉ dựa vào ý kiến chuyên gia có thể mang lại sự khác biệt trong chuẩn đoán dẫn đến cản trở quá trình điều trị. Mặc dù các công cụ chuẩn đoán hoàn toàn tự động vẫn chưa đạt được mong đợi nhưng việc nhận diện các mẫu hình được phát triển theo hướng hỗ trợ đưa ra quyết định, chủ yếu bắt đầu với việc phát hiện răng và các cấu trúc của chúng trong hình ảnh x-quang. Từ trước đến nay, phát hiện răng bất thường chủ yếu được xử lý bằng ngưỡng (threshold) và khu vực (region-based). Dưới đây là phương pháp học sâu, giúp phát hiện răng và phân đoạn từ một ảnh chụp x-quang toàn cảnh theo hướng hoàn toàn khác. Ảnh chụp x-quang toàn cảnh này còn chứa các hình ảnh của phần cằm, cột sống và hàm nên việc tách ảnh răng khá là khó. Hệ thống dựa trên hệ thần kinh xoắn ốc sử dụng mặt nạ (mask region-based convolutional neural network) để thực hiện phân đoạn cá thể. Hiệu suất đánh giá trên một tập có 1500 ảnh với độ biến thể cao và 10 loại hình ảnh chụp từ miệng bệnh nhân. Bằng cách chỉ đưa ra hệ thống đề xuất 193 hình ảnh miệng với trung bình 32 răng đạt được kết quả cao hơn so với nhiều hệ thống khác.

II. Giới thiệu

Từ những hình ảnh x-quang, nha sĩ có thể phân tích toàn bộ cấu trúc răng. Hình ảnh X-quang đã được các nha sĩ sử dụng để phân tích tình trạng răng, nướu, hàm,... để chuẩn đoán các bệnh về răng miệng. Thường có 2 loại chụp X-quang: ảnh chụp từ trong miệng bệnh nhân hoặc ảnh chụp từ bên ngoài miệng bệnh nhân và có 3 loại tia X thường được dùng trong nha khoa: ảnh chụp x-quang toàn cảnh (panoramic X-ray); bitewing X-ray; periapical intraoral radiography hoặc chỉ periapical X-ray.



(a)



(b)



(c)

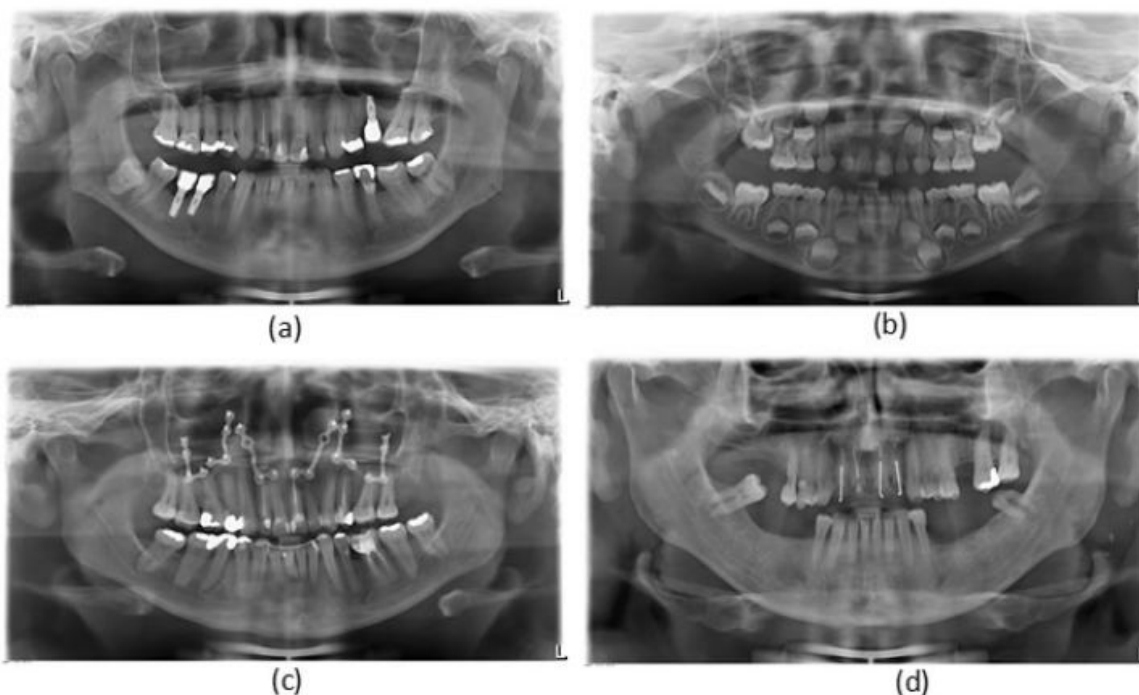
- (a) Bitewing X-ray;
- (b) Periapical X-ray;
- (c) Panoramic X-ray.

Hình 1.

Hình 1 là ví dụ để thể hiện các loại chụp ảnh x-quang. Và chụp X-quang toàn cảnh là phương pháp hữu hiệu để khám lâm sàng và chuẩn đoán bệnh răng miệng. Loại hình ảnh này cho phép trực quan hóa các vấn đề về răng miệng, gãy xương, u,... Thông thường, nha sĩ sẽ yêu cầu ảnh chụp để kiểm tra tiền phẫu.

III. Huấn luyện mask R-CNN

X quang chụp toàn cảnh không chỉ giới hạn ở phần răng bị cô lập, như trong hình ảnh chụp X quang, cũng cho thấy khớp giữa hàm và sọ, cằm, cột sống và các chi tiết khác bắt nguồn từ xương của vùng mũi và mặt. Các thông tin khác về chụp X quang toàn cảnh làm cho hình ảnh của họ khó phân tích, chẳng hạn như các biến thể của răng bệnh nhân, đồ tạo tác được sử dụng cho phục hình và giả, tính đồng nhất ở các vùng gần đối tượng quan tâm, không gian hiện có của răng bị thiếu và giới hạn của phương pháp chuyển đổi. Hình 2 mô tả một số ví dụ về các trường hợp này. Tóm lại, phân tích hình ảnh X-quang toàn cảnh phụ thuộc vào công việc cẩn thận của người chuyên nghiệp, những người không có công cụ tự động trong trợ giúp của mình. Bằng cách này, một phương pháp phân đoạn tự động để cô lập các bộ phận của hình ảnh X quang toàn cảnh có thể là một khởi đầu giúp các nha sĩ trong chẩn đoán của họ.



(a) Răng cấy ghép, răng phục hồi; (b) răng bị dư; (c) thiết bị cho chấn thương hàm dưới; (d) răng bị thiếu, hỏng.

Hình 2.

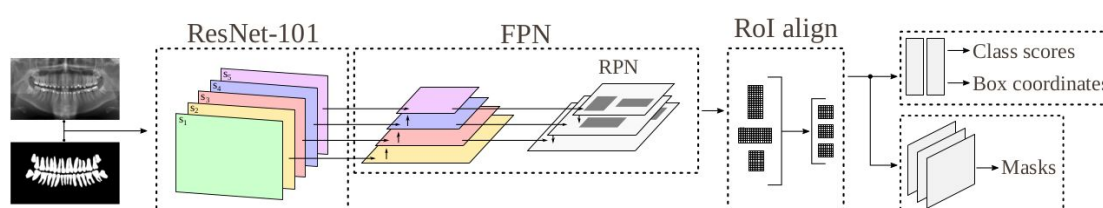
IV. Các công việc liên quan

Phân đoạn hình ảnh không giám sát trong ảnh chụp X-quang: phần lớn các nghiên cứu về hình ảnh nha khoa dựa trên phân đoạn pixel không được giám sát. Dựa trên hình ảnh bitewing images, periapical, or panoramic.

Phân đoạn răng trong hình ảnh x-quang toàn cảnh: khai thác phân đoạn cá thể dựa trên việc học sâu về hình ảnh X-quang toàn cảnh. Các phương pháp được so sánh chỉ sử dụng không hỗ trợ phân đoạn như : Cluster-based, threshold-based, region-based, boundary-based.

| STT | Danh mục | Số lượng ảnh | Số răng trung bình |
|-----|--|--------------|--------------------|
| 1 | Hình ảnh với tất cả răng chứa răng được phục hồi và các thiết bị nha khoa | 73 | 32 |
| 2 | Hình ảnh với tất cả răng chứa răng được phục hồi nhưng không thiết bị nha khoa | 220 | 32 |
| 3 | Hình ảnh với tất cả răng không chứa răng được phục hồi nhưng có các thiết bị nha khoa | 45 | 32 |
| 4 | Hình ảnh với tất cả răng không chứa răng được phục hồi và các thiết bị nha khoa | 140 | 32 |
| 5 | Hình ảnh răng có cấy ghép nha khoa | 120 | 18 |
| 6 | Hình ảnh có nhiều hơn 32 răng | 170 | 37 |
| 7 | Hình ảnh các răng bị thiếu, gồm có răng được phục hồi và các thiết bị nha khoa | 115 | 27 |
| 8 | Hình ảnh các răng bị thiếu, gồm có răng được phục hồi nhưng không có các thiết bị nha khoa | 457 | 29 |
| 9 | Hình ảnh các răng bị thiếu, không có răng được phục hồi nhưng gồm các thiết bị nha khoa | 45 | 28 |
| 10 | Hình ảnh các răng bị thiếu, không có răng được phục hồi và các thiết bị nha khoa | 115 | 28 |

Bảng 1. Danh mục dataset ảnh và số lượng răng trung bình ở mỗi tập danh mục.



Hình 3. Quá trình huấn luyện của một hệ thống phân đoạn. Từ trái qua phải: ảnh chụp X-quang và mặt nạ ảnh chú thích (annotation masks) được làm đầu vào, ResNet-101 trích xuất tính năng 5 tầng (từ S1 đến S5), nơi đầu ra của mỗi giai đoạn ResNet, nhưng S1, tạo thành một lớp trong mạng kim tự tháp tính năng (FPN); anchors được xác định trên FPN, và các vùng quan tâm (RoI) được tính toán (xác định mạng đề xuất vùng (RPN)) và cuối cùng,

được căn chỉnh (liên kết Rol). Đầu ra là điểm số của lớp và tọa độ hộp, được cung cấp bởi mạng được kết nối đầy đủ và mặt nạ, được cung cấp bởi fully convolutional network.

V. Đóng góp

Hiện nay, hầu hết các công trình đã giải quyết vấn đề phân khúc răng bằng cách sử dụng hình ảnh Bitewing và periapical images, do những khó khăn trong việc xử lý tầm nhìn toàn cảnh của răng miệng. Ảnh Bitewing và periapical images cho phép quan sát răng rõ ràng hơn và ít nhiễu hơn các cấu trúc xương khác. Các bộ dữ liệu được sử dụng trong các tác phẩm trình bày giới hạn để đánh giá các ứng dụng toàn diện của các phương pháp được đề xuất, mang lại một vài hình ảnh với các biến thể nhỏ, ngay cả trong một vài công trình đã khai thác toàn cảnh hình ảnh.

Dựa vào bộ dữ liệu như bảng 1, 1500 ảnh dữ liệu panoramic X-ray, với độ biến thiên cao, được phân loại thành 10 danh mục như trong bảng 1 để đào tạo dựa trên mask region-based convolutional neural network (Mask R-CNN). Kết quả cho thấy, bộ ảnh tập huấn là 193 ảnh răng miệng, hệ thống có khả năng đạt được 98% độ chính xác, 88% điểm F1, 94% độ chính xác, 84% thu hồi và, cuối cùng, 99% độ đặc hiệu trên 1224 hình ảnh của miệng. Kết quả đạt được bằng cách tăng răng ra khỏi các hình, và khai thác yếu tố nhiễu, nói cách khác là tăng đối tượng (răng thay vì miệng) đã hiệu suất so với các phương pháp được đánh giá ở dưới đây.

VI. Phương pháp phân đoạn hình ảnh x-quang

Nhiệm vụ là phát hiện đối tượng nhằm mục đích phân loại các đối tượng riêng lẻ. Mục tiêu phân đoạn ngữ nghĩa là phân loại từng pixel của đối tượng thành các mục đã biết mà không phân biệt các thể hiện đối tượng. Sử dụng phương pháp cổ điển là phát hiện và phân đoạn ngữ nghĩa. Phương pháp sử dụng ở đây là Mask R-CNN để phân loại hình ảnh x-quang.

Chi tiết kiến trúc của mạng deep network

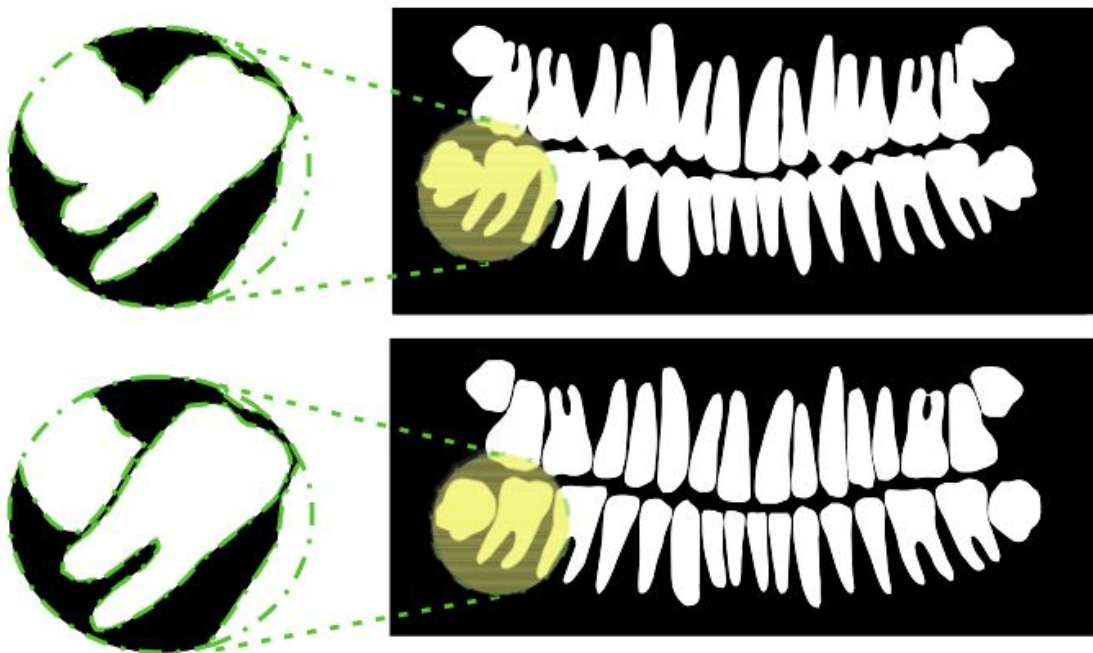
Kiến trúc của Mask R-CNN được thể hiện như hình 3, như là một mở rộng của **Faster R-CNN(*)** bao gồm một nhánh của mạng tích chập để thực hiện nhiệm vụ phân đoạn cá thể. Sau khi trích xuất các tính năng từ ResNet101, các tính năng này tạo ra một mạng kim tự tháp đặc trưng (FPN), nơi neo được xác định cuối cùng và các vùng quan tâm (Rols) được trích xuất. Hai giai đoạn này (FPN + neo) tạo thành mạng đề xuất vùng (RPN) được giới thiệu trong Faster R-CNN. Sau đó, các ROI được căn chỉnh để có cùng kích thước. Cuối cùng, mỗi fixed-size feature là: i) Phân loại răng và nền (class scores), ii) được local bằng cách hồi quy các tọa độ bounding box và iii) mỗi pixel được phân đoạn bằng bởi một mạng tích chập đầy đủ (full convolutional network (FCN)) trong mỗi bounding box răng phát hiện được (masks).

Trong tập dữ liệu chỉ có 193 hình ảnh được chú thích, không đủ để huấn luyện Mask R-CNN từ đầu do số lượng tham số free trong mạng học sâu. Vì vậy, trọng số được đào tạo trước trong tập dữ liệu **MSCOCO(**)**, có 80 đối tượng được chú thích cho ví dụ này. Chỉ sử dụng trọng số trong ResNet 101 của mạng Mask R-CNN. Các trọng số của top layer (RPN,...) được khởi tạo theo tập dữ liệu.

Các hyperparameters của việc huấn luyện Mask (ví dụ, tốc độ học tập, số lượng các kỷ nguyên) đã được xác định theo kinh nghiệm bằng cách quan sát việc đào tạo về một thí nghiệm với dữ liệu xác nhận hợp lệ. Trong bước này, chia 193 hình ảnh được chú thích thành hai bộ riêng biệt để đào tạo mạng phân đoạn, để xác thực các kết quả và điều chỉnh

các hyperparameters. Các hyperparameters được điều chỉnh đã được sử dụng trong bước đào tạo, kết quả của chúng được mô tả trong Phần VII.

Giai đoạn đào tạo dữ liệu được thực hiện các bước hai bước (two-fold steps). Trong bước đầu tiên, một **Adam optimizer(***)** được sử dụng với $\alpha = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. Các trọng số của các head layers được đào tạo 100 epoches. Trình tối ưu hóa Adam được sử dụng để có được kết quả nhanh chóng về việc điều chỉnh trọng số cho nhiệm vụ phân đoạn mới. Trong bước thứ hai, một trình tối ưu hóa dao động của đạo hàm (stochastic gradient-descent (SGD)), mà không có một định nghĩa nào về động lượng, với tốc độ học tập là 10^{-6} để thực hiện tinh chỉnh các trọng số. Trong bước cuối cùng này, trọng số của các giai đoạn 4 và 5 từ ResNet 101 (xem Hình 3) cũng như các lớp đầu được coi như là được huấn luyện. Mỗi giai đoạn của ResNet 101 tương ứng với một tập hợp các lớp liên kết tuần tự với cùng kích thước feature map. Việc đào tạo ới tỷ lệ lỗi là 10^6 . Trọng lượng mạng do giai đoạn đào tạo này được sử dụng trong việc đánh giá và so sánh kết quả so với các phương pháp khác.



Hình 4. Quá trình tách răng trong tập dữ liệu. Răng được thu phóng ở trên cùng, và ở phía dưới là vì nó đã được thay đổi ở đây.

| Category | Accuracy | Specificity | Precision | Recall | F1-score |
|----------------------|--------------------|-------------------|------------------|------------------|------------------|
| #1 (73) | 0.98 | 0.99 | 0.92 | 0.91 | 0.92 |
| #2 (60) | 0.98 | 0.99 | 0.94 | 0.91 | 0.92 |
| #3 (2) | 0.97 | 0.99 | 0.96 | 0.77 | 0.85 |
| #4 (67) | 0.98 | 0.99 | 0.96 | 0.89 | 0.93 |
| #5 (120) | 0.98 | 0.99 | 0.94 | 0.82 | 0.87 |
| #6 (170) | 0.97 | 0.99 | 0.94 | 0.83 | 0.88 |
| #7 (115) | 0.97 | 0.98 | 0.80 | 0.90 | 0.84 |
| #8 (457) | 0.97 | 0.99 | 0.97 | 0.80 | 0.87 |
| #9 (45) | 0.98 | 0.99 | 0.96 | 0.88 | 0.92 |
| #10 (115) | 0.99 | 0.99 | 0.93 | 0.89 | 0.91 |
| Weighted sum: | 1195.83 | 1215.73 | 1146.63 | 1032.57 | 1082.11 |
| Average±STD: | 0.98 ±0.008 | 0.99±0.006 | 0.94±0.06 | 0.84±0.07 | 0.88±0.05 |

Bảng 2. Tóm tắt các kết quả.

VII. Phân tích thực nghiệm

Chú thích của 276 ảnh trong bốn loại đầu tiên (1, 2, 3 và 4) của tập dữ liệu gốc [3] đã được sửa đổi bằng cách tách các răng theo Hình. 4. Các loại này được chọn vì chúng hiển thị hình ảnh với 32 răng, đặt nhiều hơn hoặc ít hơn ở các vị trí dự kiến; 193 hình ảnh đã được sử dụng cho đào tạo (6987 răng), trong khi 83 hình ảnh (3040 răng) đã được sử dụng như một bộ xác nhận để tinh chỉnh mạng sâu.

Quá trình tách răng trong mỗi bức ảnh được thực hiện để huấn luyện Mask R-CNN với nhiều mẫu hơn (bây giờ, các đối tượng là răng, chứ không phải toàn bộ vòm răng) so với tập dữ liệu thu thập ban đầu. Chiến lược này được chứng minh là có hiệu quả, thậm chí xem xét 1224 hình ảnh thử nghiệm đã được sử dụng làm bộ dữ liệu với toàn bộ chú thích vòm nha khoa. Sau khi huấn luyện Mask R-CNN với 6987 ảnh răng (từ 193 ảnh), và tinh chỉnh các tham số mạng với hình ảnh răng 3040 (từ 83 ảnh), 1224 hình ảnh vòm răng được sử dụng để đánh giá Mask R-CNN (sử dụng chú thích gốc của miệng).

***Phân tích định lượng

Để đánh giá hiệu suất, sử dụng các định lượng sau: $accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)}$, $specificity = \frac{TN}{(TN+FP)}$, $precision = \frac{TP}{(TP+FP)}$, $recall = \frac{TP}{(TP+FN)}$ và $F1-score = \frac{2*recall*precision}{(recall+precision)}$ (****), ở đây TP, TN, FN, FP lần lượt là true positive, true negative, false negative and false positive, respectively. Các đại lượng này được sử dụng trong kiểu pixel-wise.







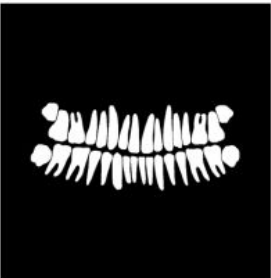
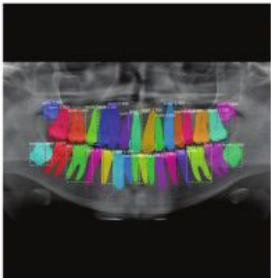
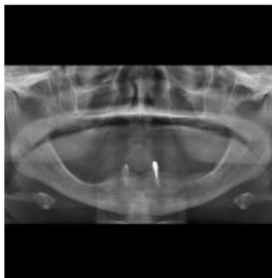
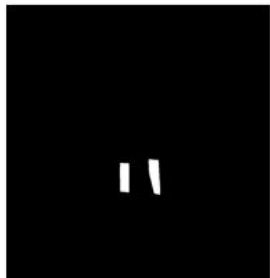
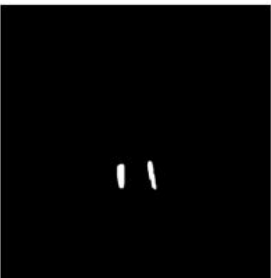
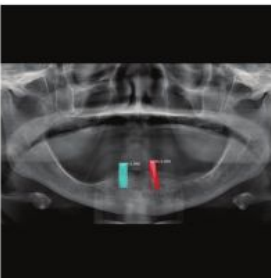
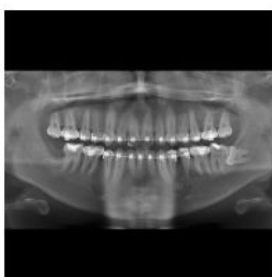


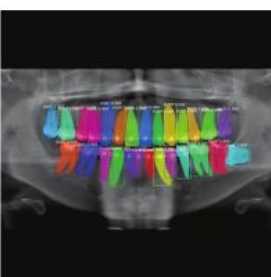



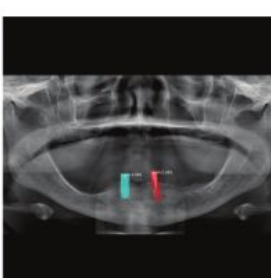
Theo phương pháp được đề xuất bởi "Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives," Expert Systems with Applications, vol. 107, pp. 15 – 31, 2018", một trung bình có trọng số được tính toán về số lượng hình ảnh cho mỗi danh mục (trọng số) và tổng kết số liệu kết quả cho từng hình ảnh, tất cả chia cho số lượng hình ảnh thử nghiệm. Bảng II tóm tắt các kết quả định lượng được hệ thống của chúng tôi tìm thấy. Tất cả các độ lệch chuẩn được tìm thấy rất thấp, chỉ ra rằng tất cả các kết quả riêng lẻ đều ở gần mức trung bình. Thực tế này chứng tỏ rằng, mặc dù các tập dữ liệu đã được thử thách, hệ thống đề xuất đạt được một sự tổng quát và thống nhất tốt trong các kết quả.

Tóm lại, kết quả cho thấy sự cân bằng tuyệt đối giữa true negative/false negative (specificity, accuracy) và true positive/false positive (accuracy, precision and recall, được chứng thực bằng điểm số F1, tính toán độ trung bình điều hòa giữa recall và precision.



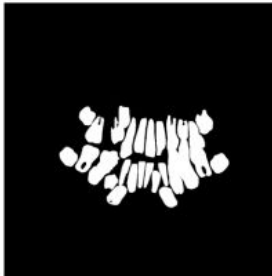
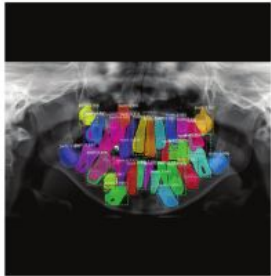
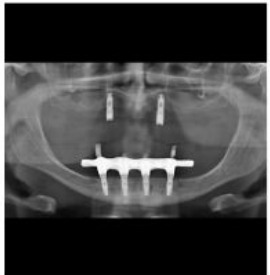


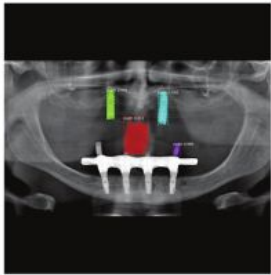

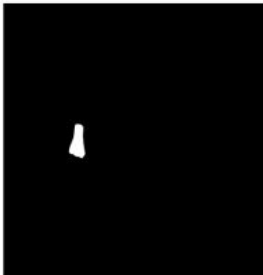

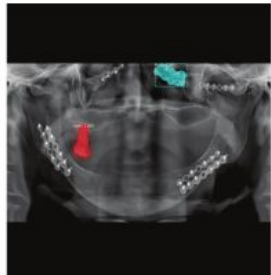



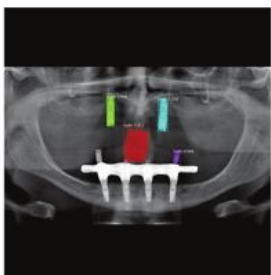
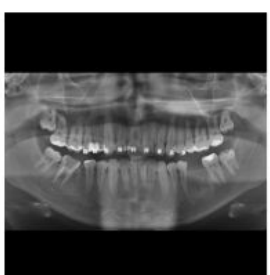


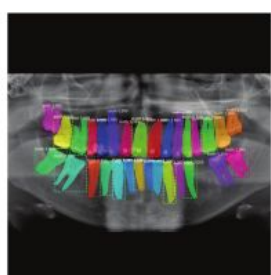
Mask R-CNN đã chứng minh kết quả cao hơn so với các phương pháp không giám sát được đánh giá trong sách được nêu trên. Bảng III trình bày kết quả của MASK R-CNN so với kết quả của các phương pháp phân đoạn không giám sát được đánh giá trong sách được nêu trên. Mặc dù số lượng hình ảnh thử nghiệm khác nhau giữa hai loại phương pháp, sự khác biệt không phải là quá cao để cản trở sự so sánh (1500 so với 1224). Trong Bảng III, kết quả tốt nhất của người không giám sát được đánh dấu cũng như kết quả của MASK R-CNN. Ngoại trừ tính đặc hiệu, kết quả của phương pháp học sâu đã gần với việc splitting/merging, xem xét các chỉ số khác, Mask R-CNN rõ ràng vượt trội. Ngoài ra, cần lưu ý rằng không có phương pháp không được giám sát nào khác trình bày tất cả các số liệu liên tục cao như cách tiếp cận học tập sâu.

| Method | Accuracy | Specificity | Precision | Recall | F1-score |
|------------------------------------|-------------|-------------|-------------|-------------|-------------|
| Region growing [12] | 0.68 | 0.69 | 0.35 | 0.63 | 0.44 |
| Splitting/merging [51] | 0.81 | 0.99 | 0.81 | 0.08 | 0.14 |
| Global thresholding [21] | 0.79 | 0.81 | 0.52 | 0.69 | 0.56 |
| Niblack method [34] | 0.81 | 0.81 | 0.51 | 0.82 | 0.61 |
| Fuzzy C-means [30] | 0.82 | 0.91 | 0.61 | 0.45 | 0.49 |
| Canny [35] | 0.79 | 0.96 | 0.45 | 0.11 | 0.17 |
| Sobel [35] | 0.80 | 0.99 | 0.66 | 0.03 | 0.06 |
| Active contours without edges [13] | 0.80 | 0.85 | 0.51 | 0.57 | 0.52 |
| Level set method [5] | 0.76 | 0.78 | 0.48 | 0.68 | 0.52 |
| Watershed [14] | 0.77 | 0.75 | 0.48 | 0.82 | 0.58 |
| Mask R-CNN | 0.98 | 0.99 | 0.94 | 0.84 | 0.88 |

Bảng 3. so sánh các kết quả học không giám sát được giới thiệu trong Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives," Expert Systems with Applications, vol. 107, pp. 15 – 31, 2018", và Mask R-CNN

| | Original | Groundtruth | Segmentation | Instance segmentation |
|----------------|---|---|--|---|
| Accuracy1.0 |  |  |  |  |
| F1-score0,94 |  |  |  |  |
| Precision1.0 |  |  |  |  |
| Recall0.96 |  |  |  |  |
| Specificity1.0 |  |  |  |  |

Hình 5. Ví dụ về một kết quả đơn lẻ tốt nhất.

| | Original | Groundtruth | Segmentation | Instance segmentation |
|-----------------|---|---|--|---|
| Accuracy0.91 |  |  |  |  |
| F1-score0.29 |  |  |  |  |
| Precision0.48 |  |  |  |  |
| Recall0.17 |  |  |  |  |
| Specificity0.96 |  |  |  |  |

Hình 6. Ví dụ về một kết quả đơn lẻ tồi nhất.

***Phân tích chất lượng

Hình 5 mô tả một kết quả tốt nhất, trong khi Hình 6 minh họa các kết quả tồi tệ nhất, trong mỗi chỉ số. Các hình ảnh ghép được bao gồm bởi các hình ảnh X-ray ban đầu, ground truth, phân đoạn và instance của nó của răng, từ cột trái sang phải, tương ứng. Các phân

đoạn tốt nhất (xem Hình 5) đạt được thông qua kết quả gần như hoàn hảo, ngoại trừ recall và điểm F1, mà cuối cùng đòi hỏi một kết quả rất cân bằng giữa recall và precision (trung bình điều hòa) cao. Trong các ví dụ phân đoạn xấu nhất (xem Hình 6), ngay cả accuracy và specificity đạt được giá trị cao; tuy nhiên, F1-score, precision và recall lại cho kết quả rất kém. Sau khi phân tích trực quan, kết quả này được mong đợi vì các hiện vật không học được trong quá trình đào tạo. Đây là trường hợp của hình ảnh trong các hàng hai, ba và bốn (xem từ trên xuống) trong Hình 6: Lưu ý rằng hình ảnh X-quang toàn cảnh hai và bốn là những hình ảnh giống nhau, trong đó hai số liệu, điểm F1 và nhớ lại, đạt được kết quả kém, trong khi ở hàng thứ ba, chúng ta có thể thấy một bộ phận giả được coi là răng, trong khi chỉ một chiếc răng được phát hiện bởi Mask R-CNN (đó là lý do chính xác 48%). Độ lệch chuẩn rất thấp (xem Bảng II) cho thấy phân khúc xấu hiếm khi được nhận thấy.

Chú thích:

(*)Faster R-CNN: <https://deeplml.com/rpn-explained.html>

(**)MSCOCO: trang web cung cấp tập dữ liệu. nguồn: <http://cocodataset.org/>

(***)Adam optimizer: Thuật toán tối ưu hóa Adam là một phần mở rộng cho stochastic gradient descent mà gần đây đã thấy sự chấp nhận rộng rãi hơn cho các ứng dụng học sâu trong tầm nhìn máy tính và xử lý ngôn ngữ tự nhiên. Adam là một thuật toán tối ưu hóa có thể được sử dụng thay cho thủ tục gốc stochastic gradient gốc để cập nhật trọng lượng mạng lặp lại dựa trên dữ liệu huấn luyện.

alpha. Còn được gọi là tỷ lệ học tập hoặc kích thước bước. Tỷ lệ mà trọng số được cập nhật (ví dụ: 0,001). Giá trị lớn hơn (ví dụ: 0,3) dẫn đến việc học ban đầu nhanh hơn trước khi tỷ lệ được cập nhật. Các giá trị nhỏ hơn (ví dụ: 1.0E-5) học tập chậm ngay trong khi đào tạo

beta1. Tỷ lệ phân rã theo cấp số nhân cho các ước tính thời điểm đầu tiên (ví dụ: 0,9).

beta2. Tỷ lệ phân rã theo cấp số mũ cho các ước tính thời điểm giây (ví dụ: 0,999). Giá trị này nên được đặt gần 1.0 về các vấn đề với độ dốc thừa thớt (ví dụ: NLP và các vấn đề về thị lực máy tính).

epsilon. Là một số rất nhỏ để ngăn chặn bất kỳ phân chia nào bằng 0 trong quá trình triển khai (ví dụ: 10E-8).

Nguồn:

<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>

(****) TP, TN, FP, FN: nguồn

http://www.academicos.ccadet.unam.mx/jorge.marquez/cursos/Instrumentacion/FalsePositive_TrueNegative_etc.pdf

Accuracy, Precision, Recall or F1: Nguồn:

<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

Nguồn bài báo cáo:

http://sibgrapi.sid.inpe.br/col/sid.inpe.br/sibgrapi/2018/08.29.19.07/doc/tooth_segmentation.pdf

Code đã được public trên link github: https://github.com/khangnlnbk/dental_images