

# Nhận dạng thực vật trong thế giới mở (LifeCLEF 2016)

Hervé Goëau<sup>1</sup>, Pierre Bonnet<sup>4</sup>, và Alexis Joly<sup>2,3</sup>

<sup>1</sup> IRD, UMR AMAP, Pháp, herve.goeau@cirad.fr

<sup>2</sup> Inria ZENITH, Pháp, alexis.joly@inria.fr

<sup>3</sup> LIRMM, Montpellier, Pháp

<sup>4</sup> CIRAD, UMR AMAP, Pháp, pierre.bonnet@cirad.fr

Trừu tượng. Thử thách nhận dạng thực vật LifeCLEF nhằm mục đích đánh giá các phương pháp và hệ thống nhận dạng thực vật ở quy mô rất lớn, gần với các điều kiện của kịch bản giám sát đa dạng sinh học trong thế giới thực. Ấn bản năm 2016 thực sự được thực hiện trên một bộ hơn 110 nghìn hình ảnh minh họa 1000 loài thực vật sống ở Tây Âu, được xây dựng thông qua nền tảng cảm biến có sự tham gia quy mô lớn được khởi xướng vào năm 2011 và hiện có sự tham gia của hàng chục nghìn người đóng góp. Điểm mới lạ chính trong những năm trước đó là nhiệm vụ nhận dạng được đánh giá là một *bộ mớ vấn đề nhận dạng*. Một vấn đề trong đó hệ thống nhận dạng phải mạnh mẽ đối với các danh mục chưa biết và chưa từng thấy. Ngoài việc phân loại bạo lực trên các lớp đã biết của tập huấn luyện, thách thức lớn là tự động từ chối các lần truy cập phân loại dương tính giả do các lớp chưa biết gây ra. Phần tổng quan này trình bày chính xác hơn các nguồn lực và đánh giá thách thức, tóm tắt các phương pháp tiếp cận và hệ thống được các nhóm nghiên cứu tham gia sử dụng và đưa ra phân tích về các kết quả chính.

Từ khóa: LifeCLEF, thực vật, lá, lá, hoa, quả, vỏ cây, thân, cành, loài, thu hồi, hình ảnh, bộ sưu tập, nhận dạng loài, khoa học công dân, phân loại chi tiết, đánh giá, điểm chuẩn

## 1. Giới thiệu

Nhận dạng thực vật dựa trên hình ảnh là giải pháp hứa hẹn nhất hướng tới thu hẹp khoảng cách phân loại thực vật, như được minh họa bằng sự phát triển của công việc nghiên cứu về chủ đề [7], [4], [13], [10], [1] cũng như sự nổi lên của các ứng dụng di động chuyên dụng như LeafSnap [14] hay Pl@ntNet [12]. Mặc dù những ứng dụng này đầy hứa hẹn nhưng hiệu suất của chúng vẫn còn kém xa so với yêu cầu của một kịch bản giám sát sinh thái hoàn toàn tự động. Việc cho phép đồng đảo người dân thực hiện các quan sát thực vật chính xác đòi hỏi phải trang bị cho họ những công cụ nhận dạng hiệu quả hơn nhiều. Như một minh họa, vào năm 2015, 2.328.502 truy vấn đã được người dùng ứng dụng di động Pl@ntNet gửi nhưng cuối cùng chỉ có dưới 3% trong số đó được chia sẻ và xác thực hợp tác. Việc cho phép khai thác các quan sát chưa được xác thực có thể mở rộng quy mô thu thập hồ sơ thực vật trên toàn thế giới lên nhiều bậc độ lớn. Đo lường và tăng cường

do đó hiệu suất của các công cụ nhận dạng tự động là rất quan trọng. Là bước đầu tiên hướng tới việc đánh giá tính khả thi của mô hình giám sát đa dạng sinh học tự động như vậy, chúng tôi đã tạo và chia sẻ một nền tảng thử nghiệm mới hoàn toàn bao gồm nhật ký tìm kiếm hình ảnh của ứng dụng di động Pl@ntNet (trái ngược với các phiên bản trước của tiêu chuẩn PlantCLEF dựa trên các quan sát thực vật được chia sẻ và xác nhận rõ ràng).

Như một kịch bản cụ thể, chúng tôi tập trung vào việc giám sát các loài thực vật ngoại lai xâm lấn. Ngày nay, những loài này gây ra tổn thất kinh tế lớn cho xã hội chúng ta (ước tính gần 12 tỷ euro mỗi năm ở châu Âu) và là một trong những mối đe dọa chính đối với việc bảo tồn đa dạng sinh học [22]. Chi phí này thậm chí có thể còn quan trọng hơn ở cấp quốc gia, chẳng hạn như ở Trung Quốc, nơi được ước tính là khoảng 15 tỷ đô la Mỹ hàng năm [23], và hơn 34 tỷ đô la Mỹ ở Mỹ [17]. Việc phát hiện sớm sự xuất hiện của các loài này cũng như theo dõi những thay đổi trong phân bố và hiện tượng học của chúng là những yếu tố then chốt để quản lý chúng và giảm chi phí quản lý.

## 2 Bộ dữ liệu

### 2.1 Tập dữ liệu huấn luyện

Đối với tập huấn luyện, chúng tôi đã cung cấp bộ dữ liệu PlantCLEF 2015 được bổ sung thêm các chú thích thực tế cơ bản của các hình ảnh thử nghiệm (được giữ bí mật trong chiến dịch năm 2015). Chính xác hơn, bộ dữ liệu PlantCLEF 2015 bao gồm 113.205 bức ảnh thuộc 41.794 quan sát của 1000 loài cây, thảo mộc và đường xỉ sống ở khu vực Tây Âu. Dữ liệu này được thu thập bởi 8.960 người đóng góp khác nhau. Mỗi bức ảnh thuộc về một và chỉ một trong 7 loại chế độ xem được bảo cáo trong siêu dữ liệu (toàn bộ cây, quả, lá, hoa, thân, cành, quét lá) và được liên kết với một mã định danh quan sát thực vật duy nhất cho phép liên kết nó với các hình ảnh khác của cùng một loại cây (được cùng một người quan sát trong cùng ngày).

Điểm độc đáo của bộ dữ liệu PlantCLEF là bản chất xã hội của nó khiến nó gần với các điều kiện của kịch bản nhận dạng trong thế giới thực: (i) hình ảnh của cùng một loài đến từ các loài thực vật khác nhau sống ở các khu vực riêng biệt, (ii) hình ảnh được chụp bởi những người dùng khác nhau có thể không sử dụng cùng một giao thức thu nhận hình ảnh, (iii) các bức ảnh được chụp vào các thời điểm khác nhau trong năm. Mỗi hình ảnh của tập dữ liệu được liên kết với siêu dữ liệu theo ngữ cảnh (tác giả, ngày tháng, tên địa phương, mã nhà máy) và dữ liệu xã hội (xếp hạng của người dùng về chất lượng hình ảnh, tên đơn vị phân loại được xác thực cộng tác, tên bản địa) được cung cấp trong tệp xml có cấu trúc. Cài đặt thiết bị và định vị địa lý gps chỉ khả dụng cho một số hình ảnh. Chính xác hơn, mỗi hình ảnh được liên kết với siêu dữ liệu sau: (i) hình ảnh của cùng một loài đến từ các loài thực vật khác nhau sống ở các khu vực riêng biệt, (ii) hình ảnh được chụp bởi những người dùng khác nhau có thể không sử dụng cùng một giao thức thu thập hình ảnh, (iii) hình ảnh được chụp vào các thời điểm khác nhau trong năm. Mỗi hình ảnh của tập dữ liệu được liên kết với siêu dữ liệu theo ngữ cảnh (tác giả, ngày tháng, tên địa phương, mã nhà máy) và dữ liệu xã hội (xếp hạng của người dùng về chất lượng hình ảnh, tên đơn vị phân loại được xác thực cộng tác, tên bản địa) được cung cấp trong tệp xml có cấu trúc. Cài đặt thiết bị và định vị địa lý gps chỉ khả dụng cho một số hình ảnh. Chính xác hơn, mỗi hình ảnh được liên kết với siêu dữ liệu sau: (i) hình ảnh của cùng một loài đến từ các loài thực vật khác nhau sống ở các khu vực riêng biệt, (ii) hình ảnh được chụp bởi những người dùng khác nhau có thể không sử dụng cùng một giao thức thu thập hình ảnh, (iii) hình ảnh được chụp vào các thời điểm khác nhau trong năm. Mỗi hình ảnh của tập dữ liệu được liên kết với siêu dữ liệu theo ngữ cảnh (tác giả, ngày tháng, tên địa phương, mã nhà máy) và dữ liệu xã hội (xếp hạng của người dùng về chất lượng hình ảnh, tên đơn vị phân loại được xác thực cộng tác, tên bản địa) được cung cấp trong tệp xml có cấu trúc. Cài đặt thiết bị và định vị địa lý gps chỉ khả dụng cho một số hình ảnh. Chính xác hơn, mỗi hình ảnh được liên kết với siêu dữ liệu sau: (iii) ảnh được chụp vào các thời điểm khác nhau trong năm. Mỗi hình ảnh của tập dữ liệu được liên kết với siêu dữ liệu theo ngữ cảnh (tác giả, ngày tháng, tên địa phương, mã nhà máy) và dữ liệu xã hội (xếp hạng của người dùng về chất lượng hình ảnh, tên đơn vị phân loại được xác thực cộng tác, tên bản địa) được cung cấp trong tệp xml có cấu trúc. Cài đặt thiết bị và định vị địa lý gps chỉ khả dụng cho một số hình ảnh. Chính xác hơn, mỗi hình ảnh được liên kết với siêu dữ liệu sau: (iii) ảnh được chụp vào các thời điểm khác nhau trong năm. Mỗi hình ảnh của tập dữ liệu được liên kết với siêu dữ liệu theo ngữ cảnh (tác giả, ngày tháng, tên địa phương, mã nhà máy) và dữ liệu xã hội (xếp hạng của người dùng về chất lượng hình ảnh, tên đơn vị phân loại được xác thực cộng tác, tên bản địa) được cung cấp trong tệp xml có cấu trúc. Cài đặt thiết bị và định vị địa lý gps chỉ khả dụng cho một số hình ảnh. Chính xác hơn, mỗi hình ảnh được liên kết với siêu dữ liệu sau:

(iii) ảnh được chụp vào các thời điểm khác nhau trong năm. Mỗi hình ảnh của tập dữ liệu được liên kết với siêu dữ liệu theo ngữ cảnh (tác giả, ngày tháng, tên địa phương, mã nhà máy) và dữ liệu xã hội (xếp hạng của người dùng về chất lượng hình ảnh, tên đơn vị phân loại được xác thực cộng tác, tên bản địa) được cung cấp trong tệp xml có cấu trúc. Cài đặt thiết bị và định vị địa lý gps chỉ khả dụng cho một số hình ảnh. Chính xác hơn, mỗi hình ảnh được liên kết với siêu dữ liệu sau:

– Mã quan sát: ID quan sát thực vật mà từ đó có thể liên kết một số hình ảnh

- Tên tệp
- ID phương tiện:id của hình ảnh
- XemNội dung:Toàn bộ hoặc Cảnh hoặc Hoa hoặc Quả hoặc Lá hoặc Lá Quét hoặc Thân
- Mã lớp:ID số lớp phải được sử dụng làm thông tin cơ bản. Đó là một số phân loại được sử dụng bởi Tela Botanica
- Giống loài:tên loài (gồm 3 phần: tên chi, tên loài, tác giả đã phát hiện hoặc sửa đổi tên loài)
- Giống:Tên của Chi, cao hơn một bậc so với Loài trong hệ thống phân loại được Tela Botanica sử dụng
- Gia đình:tên của Họ, hai cấp trên Loài trong hệ thống phân cấp được Tela Botanica sử dụng
- Ngày: (nếu có) ngày quan sát cây
- Bỏ phiếu:mức trung bình (làm tròn) xếp hạng của người dùng về chất lượng hình ảnh
- Vị trí: (nếu có) tên địa phương, thị trấn trong hầu hết thời gian
- Kinh độ vĩ độ: (nếu có) tọa độ GPS của quan sát trong siêu dữ liệu EXIF hoặc nếu không tìm thấy thông tin GPS trong EXIF thì tọa độ GPS của địa phương nơi nhà máy được quan sát (chỉ dành cho các thị trấn thuộc vùng đô thị của Pháp)
- Tác giả:tên tác giả bức tranh
- NămTrongCLEF:ImageCLEF2011, ImageCLEF2012, ImageCLEF2013, Plant-CLEF2014, PlantCLEF2015 chỉ định thời điểm hình ảnh được tích hợp trong thử thách
- PersonalPlantId2014:ID quan sát thực vật được sử dụng năm ngoái trong nhiệm vụ thực vật LifeCLEF2014
- ImageID2014:hình ảnh id.jpg được sử dụng vào năm 2014.

## 2.2 Tập dữ liệu thử nghiệm

Đối với tập thử nghiệm, chúng tôi đã tạo một tập dữ liệu có chú thích mới dựa trên các truy vấn hình ảnh được gửi bởi những người dùng đã được xác thực của ứng dụng di động Pl@ntNet vào năm 2015 (các truy vấn không được xác thực phải bị xóa vì vấn đề bản quyền). Một phần trong số các truy vấn đó đã được liên kết với tên loài hợp lệ vì chúng đã được các tác giả chia sẻ rõ ràng và được cộng tác sửa đổi. Chúng tôi đã đưa vào tập thử nghiệm 4633 loài có liên quan đến một loài thuộc 1000 loài của tập huấn luyện (điền vào các lớp đã biết). Những bức ảnh còn lại được phân phát cho một nhóm các nhà thực vật học chịu trách nhiệm chú thích thủ công chúng bằng tên loài hợp lệ hoặc bằng các thẻ mới được tạo theo lựa chọn của họ (và được chia sẻ giữa họ). Trong khoảng thời gian dành cho quá trình này, họ có thể chú thích thủ công 1821 bức ảnh có trong bộ thử nghiệm. Do đó, 144 thẻ mới đã được tạo ra để xác định các lớp chưa biết, chẳng hạn như *vật thể không phải thực vật, chân hoặc bàn tay, UVQ* (Đối tượng thực vật không xác định), *cây nhân tạo, họ xương rồng, nấm, động vật, đồ ăn, rau* hoặc tên chính xác hơn của các loại cây trồng trong vườn như hoa hồng, hoa phong lữ, cây ficus, v.v. Vì lý do riêng tư, chúng tôi đã phải xóa tất cả các hình ảnh được gắn thẻ là *mọi người* (khoảng 1.1% số truy vấn được gắn thẻ). Cuối cùng, để hoàn thành số ảnh thử nghiệm thuộc về chưa rõ

các lớp, chúng tôi đã chọn ngẫu nhiên một tập hợp 1546 truy vấn hình ảnh có liên quan đến tên loài hợp lệ không thuộc hệ thực vật Tây Âu (và do đó, không thuộc về 1000 loài của tập huấn luyện hoặc các loài có khả năng rất giống nhau). Cuối cùng, tập kiểm tra bao gồm 8.000 ảnh, 4633 ảnh được gắn nhãn là một trong 1000 lớp đã biết của tập huấn luyện và 3367 được gắn nhãn là các lớp mới chưa biết. Trong số 4633 hình ảnh về các loài đã biết, có 366 hình ảnh được gắn thẻ là *xâm lấn* theo danh sách chọn lọc gồm 26 loài có khả năng xâm lấn. Danh sách này được xác định bằng cách tổng hợp một số nguồn (chẳng hạn như Viện bảo tồn thực vật quốc gia và Chương trình các loài xâm lấn toàn cầu) và bằng cách tính toán điểm giao với 1000 loài của tập huấn luyện.

### 3 Mô tả nhiệm vụ

Dựa trên nền tảng thử nghiệm được mô tả trước đó, chúng tôi đã tiến hành đánh giá theo định hướng hệ thống với sự tham gia của các nhóm nghiên cứu khác nhau đã tải xuống dữ liệu và chạy hệ thống của họ. Để tránh những người tham gia điều chỉnh thuật toán của họ theo kịch bản các loài xâm lấn và giữ cho đánh giá của chúng tôi có thể khái quát hóa cho những người khác, chúng tôi đã không cung cấp danh sách các loài sẽ được phát hiện. Những người tham gia chỉ biết rằng các loài mục tiêu được bao gồm trong một tập hợp lớn hơn gồm 1000 loài mà chúng tôi đã cung cấp bộ đào tạo. Những người tham gia cũng biết rằng (i) hầu hết dữ liệu thử nghiệm không thuộc danh sách các loài mục tiêu (ii) một phần lớn trong số chúng không thuộc tập huấn luyện của 1000 loài và (iii) một phần trong số đó thậm chí có thể không phải là thực vật. Về bản chất, nhiệm vụ cần giải quyết có liên quan đến cái mà đôi khi được gọi là *bộ mở hoặc mở rộng thế giới* các vấn đề về nhận dạng [3,18], tức là các vấn đề trong đó hệ thống nhận dạng phải mạnh mẽ đối với các danh mục chưa biết và chưa từng thấy. Ngoài việc phân loại bạo lực trên các lớp đã biết của tập huấn luyện, một thách thức lớn là tự động từ chối các lần truy cập phân loại dương tính giả do các lớp chưa biết gây ra (*I E* bởi những kẻ gây phân tâm). Để đo lường khả năng này của các hệ thống được đánh giá, mỗi dự đoán phải được liên kết với điểm tin cậy trong  $P \in [0,1]$  định lượng xác suất dự đoán này là đúng (độc lập với các dự đoán khác).

Mỗi nhóm tham gia được phép gửi tối đa 4 lượt chạy được xây dựng từ các phương pháp khác nhau. Các phương pháp tiếp cận bán giám sát, tương tác hoặc sử dụng nguồn lực cộng đồng được cho phép nhưng được so sánh độc lập với các phương pháp hoàn toàn tự động. Do đó, bất kỳ sự trợ giúp nào của con người trong việc xử lý các truy vấn kiểm tra đều phải được báo hiệu trong các lần chạy được gửi.

Những người tham gia thử thách được phép sử dụng dữ liệu đào tạo bên ngoài với điều kiện thử nghiệm hoàn toàn có thể tái sản xuất được, tức là tài nguyên bên ngoài đã sử dụng được tham chiếu rõ ràng và có thể truy cập được đối với bất kỳ nhóm nghiên cứu nào khác trên thế giới và tài nguyên bổ sung không chứa bất kỳ quan sát thử nghiệm nào. Đặc biệt nghiêm cấm thu thập dữ liệu đào tạo từ các tên miền sau:

<http://ds.plantnet-project.org/>

<http://www.tela-botanica.org> <http://identify.plantnet-project.org> <http://publish.plantnet-project.org/> <http://www.gbif.org/>

## 4 số liệu

Số liệu được sử dụng để đánh giá hiệu suất của hệ thống là độ chính xác trung bình trung bình phân loại, sau đây được gọi là "mAP-open", xem xét từng lớp  $CToi$  của tập huấn luyện dưới dạng một truy vấn. Cụ thể hơn, đối với mỗi lớp  $CToi$ , chúng tôi trích xuất từ tập chạy tất cả các dự đoán với  $Id$  lớp dự đoán  $= CToi$ , xếp hạng chúng theo xác suất giảm dần  $P \in [0,1]$  và tính Độ chính xác trung bình cho lớp đó. Giá trị trung bình sau đó được tính toán trên tất cả các lớp. Các yếu tố gây phân tâm liên quan đến các giá trị xác suất cao (tức là cảnh báo sai) có khả năng làm suy giảm mAP ở mức độ cao, do đó, điều quan trọng là phải thử loại bỏ chúng. Để đánh giá cụ thể hơn kịch bản sử dụng mục tiêu ( $I$  các loài xâm lấn), một mAP thứ cấp ("mAP-xâm lấn mở") đã được tính toán bằng cách chỉ xem xét một tập hợp con của các loài thuộc danh sách đen các loài xâm lấn dưới dạng truy vấn.

## 5 Đối tượng và phương pháp

94 nhóm nghiên cứu đã đăng ký tham gia cuộc thi LifeCLEF plant challenge 2016 và tải xuống bộ dữ liệu. Trong số lượng lớn khán giả thô này, 8 nhóm nghiên cứu đã thành công trong việc gửi chạy,  $I$  các tập tin chứa dự đoán của (các) hệ thống mà chúng đã chạy. Chi tiết về các phương pháp và hệ thống được sử dụng trong các cuộc chạy được phát triển thêm trong ghi chú làm việc riêng của những người tham gia (Bluefield [9], Sabanci [5], CMP [20], LIIR, Floristic [6], UM [15], QUT [16], BME [2]). Bảng 1 cung cấp kết quả đạt được sau mỗi lần chạy cũng như tổng hợp ngắn gọn về các phương pháp được sử dụng trong mỗi lần chạy. Bổ sung, các đoạn văn sau đây cung cấp thêm một số chi tiết về các phương pháp và chiến lược tổng thể được mỗi người tham gia sử dụng.

Hệ thống Bluefield, Nhật Bản, 4 lần chạy, [9]: Một hệ thống dựa trên VGGNet [19] có bổ sung Spatial Pyramid Pooling, Parametric ReLU và loại bỏ lớp không xác định dựa trên điểm dự đoán tối thiểu của dữ liệu huấn luyện (Run 1). Lần chạy 2 giống như lần chạy 1 nhưng có cách từ chối hơi khác bằng cách sử dụng bộ xác thực. Lần chạy 3 và 4 tương ứng giống với Lần chạy 1 và 2 nhưng điểm của các hình ảnh thuộc cùng một quan sát đã được tính tổng và chuẩn hóa.

Hệ thống BME TMIT, Hungary, 4 lần chạy, [2]: Nhóm này đã cố gắng kết hợp ba phương pháp phân loại: (i) một phương pháp dựa trên các đặc điểm SIFT dày đặc, vector đánh giá và SVM (Run 2), (ii) phương pháp thứ hai dựa trên AlexNet CNN (Run 1) và (iii), phương pháp cuối cùng dựa trên SVM được huấn luyện trên siêu dữ liệu. Lần chạy 3 tương ứng với sự kết hợp của ba phân loại (sử dụng trung bình có trọng số) và Lần chạy 4 đã thêm hai cơ chế loại bỏ vào Run3 (loại bỏ dựa trên khoảng cách đối với

Bảng 1: Kết quả của nhiệm vụ xác định thực vật LifeCLEF 2016. Cột “Từ khóa” & “Từ chối” cố gắng đưa ra ý chính về phương pháp được sử dụng.

Chạy	Từ khóa	Sự từ chối	bản đồ- mở	bản đồ- mở- xâm lấn	bản đồ- đóng cửa
Bluefield Run4	VGGNet, kết hợp đầu ra từ cùng một quan sát	ngưỡng theo lớp (đào tạo + xác nhận)	0,742	0,717	0,827
SabancıU GebzeTU Run1	2x(VGGNet, GoogleNet) điều chỉnh với sự tôn trọng. 70k, 115k hình ảnh đào tạo	GoogleNet 70k/70k Thực vật/Mạng hình ảnh	0,738	0,704	0,806
SabancıU...Run3	SabancıUGebzeTU Run1	Đã xóa thủ công 90 hình ảnh thử nghiệm	0,737	0,703	0,807
Bluefield Run3	Bluefield Chạy 4	ngưỡng theo lớp	0,736	0,718	0,82
SabancıU...Run2	SabancıUGebzeTU Run1	-	0,736	0,683	0,807
SabancıU...Run4	SabancıUGebzeTU Run1	-	0,735	0,695	0,802
CMP Chạy1	Đóng gói 3xResNet-152	-	0,71	0,653	0,79
LIIR KUL Run3	CaffeNet, VGGNet16, 3xGoogleNet, thêm 12k hình ảnh thực vật bên ngoài	ngưỡng	0,703	0,674	0,761
LIIR KUL Run2	LIIR KUL Chạy 3	ngưỡng	0,692	0,667	0,744
LIIR KUL Run1	LIIR KUL Chạy 3	ngưỡng	0,669	0,652	0,708
UM Run4	VGGNet16	-	0,669	0,598	0,742
CMP Chạy2	ResNet-152	-	0,644	0,564	0,729
CMP Chạy3	ResNet-152 (đào tạo 2015)	-	0,639	0,59	0,723
QUT Run3	1 "chung" GoogleNet, 6 "cơ quan" GoogleNets, kết hợp quan sát	-	0,629	0,61	0,696
Hoa Run3	GoogleNet, siêu dữ liệu	-	0,627	0,533	0,693
UM Run1	VGGNet16	-	0,627	0,537	0,7
Hoa Run1	GoogleNet	-	0,619	0,541	0,694
Bluefield Run1	VGGNet	ngưỡng theo lớp	0,611	0,6	0,692
Bluefield Run2	VGGNet	ngưỡng theo lớp	0,611	0,6	0,693
Hoa Run2	GoogleNet	ngưỡng theo lớp	0,611	0,538	0,681
QUT Chạy1	GoogleNet	-	0,601	0,563	0,672
UM Run3	VGGNet16 chuyên dụng và cơ quan kết hợp & lớp loài	-	0,589	0,509	0,652
QUT Run2	6 "cơ quan" GoogleNets, kết hợp quan sát	-	0,564	0,562	0,641
UM Run2	VGGNet16 từ đầu (không có ImageNet2012)	-	0,481	0,446	0,552
QUT Run4	QUT Run3	ngưỡng	0,367	0,359	0,378
BMETMITRun4	AlexNet & BVW & metadata	-	0,174	0,144	0,213
BMETMITRun3	AlexNet & BVW & metadata	ngưỡng bằng sự phân loại	0,17	0,125	0,197
BMETMITRun1	AlexNet	-	0,169	0,125	0,196
BMETMITRun2	BVW (vectơ ngư dân)	-	0,066	0,128	0,101

vectơ ngư dân và điểm dự đoán tối thiểu của dữ liệu huấn luyện cho CNN).

Hệ CMP, Cộng hòa Séc, 3 lần chạy:Nhóm này đã xây dựng hệ thống của họ bằng cách tiếp cận CNN dư rất sâu ResNet với 152 lớp [11] đã đạt được kết quả tốt nhất trong cả hai thử thách ILSVRC 2015 và COCO 2015 (Đối tượng chung trong bối cảnh) vào năm ngoái. Họ đã thêm một lớp được kết nối đầy đủ với 512 nơ-ron trên đầu mạng, ngay trước bộ phân loại softmax có chức năng kích hoạt tối đa [8]. Do đó, họ đã có được lần chạy đầu tiên (lần 2) bằng cách sử dụng tất cả tập dữ liệu huấn luyện năm 2016 trong khi họ chỉ sử dụng tập dữ liệu huấn luyện năm 2015 trong lần chạy 3. Lần chạy 1 đạt được hiệu suất tốt nhất bằng cách sử dụng phương pháp đóng bao của 3 ResNet-152: tập dữ liệu huấn luyện là được chia thành ba phần và mỗi CNN sử dụng một phần khác nhau để xác thực và hai phần còn lại để tinh chỉnh.

Hệ thống hoa, Pháp, 3 chạy, [6]:Người tham gia này đã sử dụng kiến trúc GoogleNet đã sửa đổi bằng cách thêm chức năng chuẩn hóa hàng loạt và kích hoạt ReLU thay vì chức năng PRELU (chạy 1). Trong Lần chạy 2, các ngưỡng thích ứng (một cho mỗi lớp) dựa trên dự đoán của các hình ảnh huấn luyện trong CNN được tinh chỉnh đã được ước tính để loại bỏ dự đoán quá thấp trên các hình ảnh thử nghiệm. Lần chạy 3 đã sử dụng tìm kiếm tương tự về mặt hình ảnh để giảm tỷ lệ dự đoán CNN ban đầu khi hình ảnh thử nghiệm đưa ra các knn không đồng nhất theo siêu dữ liệu (thẻ cơ quan, GPS, cấp độ chi và họ).

Hệ thống LIIR KUL, Bỉ, 3 lần chạy:Nhóm này đã sử dụng một bộ phân loại tổng hợp gồm 5 mô hình được tinh chỉnh: một CaffeNet, một VGGNet16 và 3 GoogLeNet. Họ đã thêm 12k dữ liệu đào tạo bên ngoài từ bộ hoa Oxford, LeafSnap và trunk12, đồng thời cố gắng khai thác thông tin trong siêu dữ liệu, chủ yếu là các bản đồ phạm vi từ tọa độ GPS so sánh dự đoán với thẻ nội dung. Để làm tiêu chí loại bỏ, họ đã sử dụng ngưỡng tin cậy của dự đoán tốt nhất, một ngưỡng khác nhau cho mỗi lần chạy (lần chạy 1: 0,25, lần chạy 2: 0,2, lần chạy 3: 0,15).

Hệ thống QUT, Úc, 4 lần chạy, [16]:Người tham gia này đã so sánh cách tiếp cận tinh chỉnh CNN tiêu chuẩn dựa trên GoogleNet (chạy 1) với cách tiếp cận đóng bao "mixDCNN" (chạy 2) được xây dựng trên 6 GoogleNet tinh chỉnh trên 6 tập hợp con đào tạo tương ứng với 6 cơ quan riêng biệt ("lá Các hình ảnh đào tạo " và "leafscan" thực sự được hợp nhất thành một tập hợp con). Các đầu ra được tính trọng số bởi "xác suất nghề nghiệp" mang lại cho mỗi CNN sự tin cậy về dự đoán của họ. Lần chạy 3 đã hợp nhất hai cách tiếp cận, lần chạy 4 cũng vậy nhưng có một ngưỡng cố gắng loại bỏ các kết quả dương tính giả.

Hệ thống Sabanci, Thổ Nhĩ Kỳ, 4 lượt chạy, [5]:Nhóm này đã sử dụng hệ thống dựa trên CNN với 2 cấu hình chính. Lần 1: tập hợp GoogLeNet [21] và VGGNet [19] được tinh chỉnh trên dữ liệu LifeCLEF 2015 (để nhận dạng các loài mục tiêu), cũng như GoogleNet thứ hai được tinh chỉnh về vấn đề loại bỏ nhị phân (sử dụng 70 nghìn hình ảnh của PlantCLEF2016 tập huấn luyện cho nhãn lớp đã biết và 70K hình ảnh bên ngoài từ bộ dữ liệu ILSVRC cho nhãn lớp chưa xác định).

Lần chạy 2 giống với lần chạy 1 nhưng không có sự từ chối. Lần 3 giống với Lần 1 nhưng có sự loại bỏ thủ công 90 hình ảnh rõ ràng không phải thực vật.

Hệ thống UM, Malaysia & Anh, 4 giải:Nhóm này đã sử dụng hệ thống CNN dựa trên 16 lớp VGGNet. Họ đã sửa đổi mức độ chập cao hơn để học cùng lúc sự kết hợp giữa các loài và cơ quan. VGGNet16 với các lớp cơ quan và loài kết hợp và chuyên dụng với bảy nhãn cơ quan: cành, toàn bộ, hoa, quả, lá, quét lá và thân.

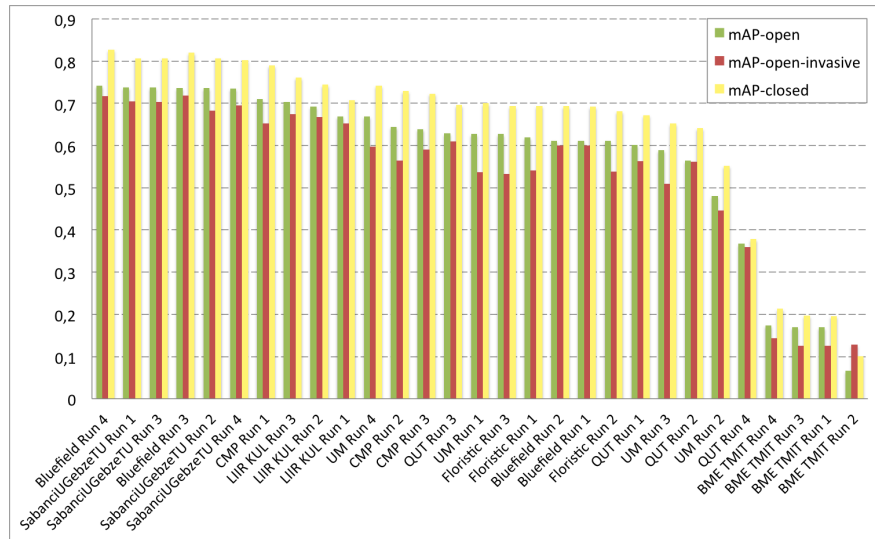
## 6 kết quả chính thức

Chúng tôi báo cáo trong Hình 1 số điểm đạt được sau 29 lần chạy được thu thập cho hai chỉ số đánh giá chính thức (mAP-mở và mAP-mở-xâm lấn). Để đánh giá tốt hơn tác động của những yếu tố gây phân tâm (*I* Ecác hình ảnh trong bộ thử nghiệm thuộc các lớp không xác định), chúng tôi cũng báo cáo mAP thu được khi xóa chúng (và được ký hiệu là mAP đã đóng). Là một nhận xét đáng chú ý đầu tiên, 26 hoạt động hàng đầu hoạt động tốt nhất đều dựa trên Mạng thần kinh chuyển đổi (CNN). Điều này chắc chắn khẳng định tính ưu việt của các phương pháp học sâu so với các phương pháp trước đây, đặc biệt là phương pháp dựa trên các tính năng thủ công (chẳng hạn như BME TMIT Run 2). Các hệ thống dựa trên CNN khác nhau chủ yếu khác nhau ở (i) kiến trúc của CNN đã sử dụng, (ii) cách quản lý việc loại bỏ các lớp chưa biết và (iii), các cải tiến thiết kế hệ thống khác nhau như tập hợp phân loại, đóng gói hoặc tổng hợp mức độ quan sát. Một mAP ấn tượng bằng 0.718 (đối với kịch bản giám sát các loài xâm lấn có chủ đích) đã đạt được nhờ cấu hình hệ thống tốt nhất của Bluefield (chạy 3). Tuy nhiên, mức tăng đạt được trong lần chạy này liên quan nhiều hơn đến việc sử dụng tổng hợp ở cấp độ quan sát (xem Bluefield chạy 1 để so sánh) hơn là loại bỏ tốt các yếu tố gây phân tâm. So sánh số liệu mAP-mở với mAP-đóng, hình trên thực tế cho thấy rằng sự hiện diện của các lớp chưa biết làm giảm hiệu suất của tất cả các hệ thống theo cách gần giống nhau. Khó khăn trong việc loại bỏ các lớp chưa biết này được xác nhận bởi sự khác biệt rất thấp giữa các lần chạy của những người tham gia đã thử nghiệm hệ thống của họ có hoặc không có sự từ chối (*ví dụ* Sabanci Run 1 so với Run 2 hoặc FlorisTic Run 1 so với Run 2). Mặt khác, người ta có thể nhận xét rằng tất cả các hệ thống đều khá mạnh mẽ trước sự hiện diện của các lớp chưa biết do hiệu suất giảm không quá cao. Trên thực tế, vì tất cả các CNN được sử dụng đều đã được đào tạo trước về một tập dữ liệu tổng quát lớn (ImageNet), nên có khả năng là họ đã học được một tập hợp các mẫu hình ảnh đủ đa dạng để tránh bị trang bị thiếu.

## 7 Phân tích bổ sung: Tác động của mức độ mới lạ

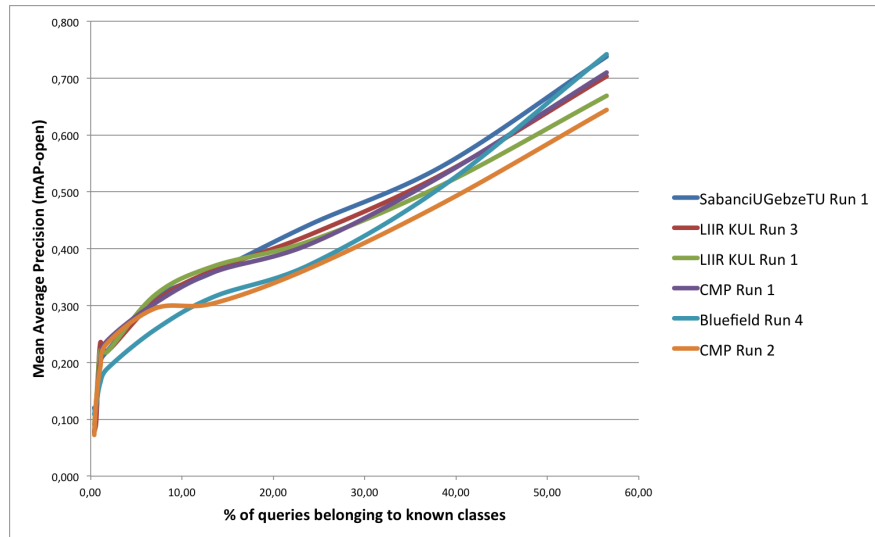
Trong quá trình đánh giá được tiến hành, tỷ lệ các lớp chưa biết trong bộ kiểm tra vẫn ở mức hợp lý (thực tế chỉ là 42%) do quy trình được sử dụng để





Hình 1. Điểm đạt được của tất cả các hệ thống được đánh giá trong nhiệm vụ nhận dạng thực vật của LifeCLEF 2016, mAP-mở: có nghĩa là Độ chính xác trung bình trên 1000 loài của tập huấn luyện và các yếu tố phân tâm trong tập kiểm tra, mAP-mở-xâm lấn: có nghĩa là Độ chính xác trung bình có yếu tố phân tâm nhưng chỉ giới hạn ở 26 loài xâm lấn, mAP đã đóng: có nghĩa là Độ chính xác trung bình trên 1000 loài nhưng không có yếu tố gây phân tâm trong bộ thử nghiệm

tạo ra nó. Trong luồng dữ liệu tìm kiếm trên thiết bị di động thực, tỷ lệ hình ảnh thuộc các lớp chưa xác định thực sự có thể cao hơn nhiều. Để mô phỏng mức độ mới lạ cao hơn như vậy, chúng tôi đã lấy mẫu dần dần các hình ảnh thử nghiệm thuộc các lớp đã biết và tính toán lại chỉ số đánh giá mAP-open. Kết quả của thí nghiệm này được cung cấp trong hình. Để rõ ràng, chúng tôi chỉ báo cáo các đường cong của các hệ thống tốt nhất (đối với nhiều mức độ mới lạ khác nhau). Kết luận đầu tiên, biểu đồ cho thấy rõ ràng rằng mức độ mới trong bộ thử nghiệm có ảnh hưởng mạnh mẽ đến hiệu suất của tất cả các hệ thống. Ngay cả khi 25% truy vấn vẫn thuộc về một lớp đã biết, không có hệ thống nào được đánh giá đạt được độ chính xác trung bình trung bình lớn hơn 0.45 (để so sánh với 0.83 trong một thế giới đóng). Tuy nhiên, một số hệ thống có khả năng chống lại tính mới tốt hơn những hệ thống khác. Chẳng hạn, hiệu suất của lần chạy Bluefield tốt nhất trên bộ thử nghiệm chính thức sẽ nhanh chóng suy giảm với tỷ lệ mới lạ cao hơn (mặc dù đã sử dụng chiến lược từ chối). Nhìn vào hoạt động tốt nhất của Sabanci, người ta có thể thấy rằng việc sử dụng lớp từ chối có giám sát là chiến lược có lợi nhất cho tỷ lệ mới vừa phải nhưng sau đó hiệu suất cũng giảm ở tỷ lệ cao. Điều thú vị là, việc so sánh LIIR KUL Run1 và LIIR KUL Run3 cho thấy rằng chỉ cần sử dụng ngưỡng từ chối cao hơn áp dụng cho xác suất CNN sẽ có lợi hơn trong bối cảnh tỷ lệ lớp chưa xác định cao. Như vậy,



Hình 2. Tác động của mức độ mới lạ: Độ chính xác trung bình trung bình so với tỷ lệ hình ảnh thử nghiệm thuộc các lớp đã biết

## 8 Kết luận

Bài viết này trình bày tổng quan và kết quả của thử thách nhận dạng thực vật LifeCLEF 2016 sau 5 thử thách trước đó được thực hiện trong diễn đàn đánh giá CLEF. Điểm mới lạ chính so với năm trước là nhiệm vụ nhận dạng được đánh giá là một *bộ mới* vấn đề nhận dạng, *I Em* một vấn đề trong đó hệ thống nhận dạng phải mạnh mẽ đối với các danh mục chưa biết và chưa từng thấy. Kết luận chính là CNN có vẻ khá mạnh mẽ một cách tự nhiên trước sự hiện diện của các lớp chưa biết trong tập thử nghiệm nhưng không có phương pháp loại bỏ nào được những người tham gia sử dụng bổ sung đã cải thiện được độ mạnh đó. Ngoài ra, tỷ lệ tính mới trong bài kiểm tra vẫn ở mức vừa phải. Do đó, chúng tôi đã tiến hành các thử nghiệm bổ sung cho thấy rằng hoạt động của CNN bị ảnh hưởng mạnh mẽ bởi tỷ lệ hình ảnh thuộc các lớp chưa xác định cao hơn và vấn đề rõ ràng vẫn còn bỏ ngỏ. Cuối cùng, nghiên cứu của chúng tôi cho thấy rằng vẫn còn một số điểm cần cải thiện trước khi có thể chia sẻ các quan sát thực vật được xác định tự động trong các nền tảng đa dạng sinh học quốc tế.

Người giới thiệu

1. Aptoula, E., Yanikoglu, B.: Đặc điểm hình thái để nhận dạng thực vật dựa vào lá. Trong: Proc. IEEE quốc tế Conf. Quá trình hình ảnh., Melbourne, Australia. P. 7 (2013)
2. Bálint Pál Tóth, Márton Tóth, DP, Szúcs, G.: Học sâu và phân loại svm để nhận dạng thực vật trong truy xuất hình ảnh quy mô lớn dựa trên nội dung. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)

3. Bendale, A., Boulton, T.E.: Hướng tới sự công nhận của thế giới mở. CoRR (2014), <http://arxiv.org/abs/1412.5687>
4. Cerutti, G., Tougné, L., Vacavant, A., Coquin, D.: Một đa giác hoạt động tham số để phân đoạn lá và ước tính hình dạng. Trong: Hội nghị chuyên đề quốc tế về máy tính trực quan. trang 202–213 (2011)
5. Ghazi, M.M., Yanikoglu, B., Aptoula, E.: Nhận dạng thực vật trong bối cảnh mở bằng cách sử dụng một tập hợp các mạng lưới thần kinh tích chập sâu. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)
6. Goëau, H., Champ, J., Joly, A.: Sự tham gia trồng hoa tại nhiệm vụ xác định thực vật lifeclef 2016. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)
7. Goëau, H., Joly, A., Selmi, S., Bonnet, P., Mouysset, E., Joyeux, L., Molino, J.F., Birnbaum, P., Bathelemy, D., Boujemaa, N.: Nhận dạng các loài thực vật dựa trên hình ảnh từ dữ liệu có nguồn gốc từ cộng đồng. Trong: Hội nghị ACM về Đa phương tiện. trang 813–814 (2011)
8. Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y.: Maxout Networks. Bản in điện tử ArXiv (Tháng 2 năm 2013)
9. Hang, S.T., Tatsuma, A., Aono, M.: Bluefield (kde tut) tại nhiệm vụ xác định thực vật lifeclef 2016. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)
10. Hazra, A., Deb, K., Kundu, S., Hazra, P., và cộng sự: Lựa chọn đặc điểm định hướng hình dạng để nhận dạng cây cà chua. Tạp chí Quốc tế về Công nghệ và Nghiên cứu Ứng dụng Máy tính 2(4), 449–meta (2013)
11. He, K., Zhang, X., Ren, S., Sun, J.: Học dư sâu để nhận dạng hình ảnh. bản in trước arXiv arXiv:1512.03385 (2015)
12. Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J.F., và cộng sự: Nhận dạng thực vật tương tác dựa trên dữ liệu hình ảnh xã hội. Tin học sinh thái 23, 22–34 (2014)
13. Kebapci, H., Yanikoglu, B., Unal, G.: Truy xuất hình ảnh thực vật bằng cách sử dụng các đặc điểm màu sắc, hình dạng và kết cấu. Tạp chí Máy tính 54(9), 1475–1490 (2011)
14. Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.B.: Leafsnap: Một hệ thống thị giác máy tính để nhận dạng loài thực vật tự động. Trong: Hội nghị Châu Âu về Thị giác Máy tính. trang 502–516 (2012)
15. Lee, S.H., Chang, Y.L., Chan, C.S., Remagnino, P.: Hệ thống nhận dạng thực vật dựa trên mạng lưới thần kinh tích chập cho nhiệm vụ phân loại thực vật lifeclef 2016. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)
16. McCool, C., Ge, Z., Corke, P.: Học tính năng thông qua hỗn hợp dcnn để phân loại thực vật hạt mịn. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)
17. Pimentel, D., Zuniga, R., Morrison, D.: Cập nhật về chi phí kinh tế và môi trường liên quan đến các loài ngoại lai xâm lấn ở Hoa Kỳ. Kinh tế sinh thái 52(3), 273–288 (2005)
18. Scheirer, W.J., Jain, L.P., Boulton, T.E.: Mô hình xác suất nhận dạng tập mở. Giao dịch của IEEE về Phân tích Mẫu và Trí thông minh Máy (T-PAMI) 36 (2014)
19. Simonyan, K., Zisserman, A.: Mạng tích chập rất sâu để nhận dạng hình ảnh quy mô lớn. CoRR abs/1409.1556 (2014)
20. Šulc, M., Mishkin, D., Matas, J.: Mạng lưới dư lượng rất sâu với mức tối đa để nhận dạng thực vật trong tự nhiên. Trong: Ghi chú làm việc của hội nghị CLEF 2016 (2016)
21. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Đi sâu hơn với sự cuộn xoáy. Trong: Kỷ yếu của Hội nghị IEEE về Tầm nhìn Máy tính và Nhận dạng Mẫu. trang 1–9 (2015)

22. Weber, E., Gut, D.: Đánh giá nguy cơ các loài thực vật có khả năng xâm lấn ở Trung Âu. Tạp chí Bảo tồn Thiên nhiên 12(3), 171–179 (2004)
23. Weber, E., Sun, SG, Li, B.: Thực vật ngoại lai xâm lấn ở Trung Quốc: sự đa dạng và hiểu biết sâu sắc về sinh thái. Xâm lược sinh học 10(8), 1411–1429 (2008)