# CHƯƠNG 11: NHẬN DẠNG HÌNH ẢNH BẰNG SỬ DỤNG MÁY HỌC 3

## 11.1. Nhận dạng đối tượng

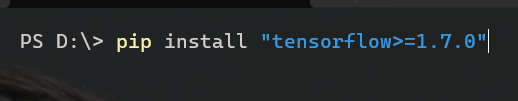
Thông qua việc dạy lại mô hình Googles Inception V3, nhận dạng hình ảnh của sự vật được thực hiện. Mô hình nhận dạng hình ảnh hiện đại có hàng triệu thông số. Để tìm một mô hình mới giúp nhận dạng hình ảnh mỗi lần, việc tìm một cấu trúc mạng Neural phù hợp và dạy lại nó đòi hỏi rất nhiều dữ liệu đào tạo và khả năng tính toán (thời gian). Do đó, cần có một kỹ thuật rút ngắn hầu hết công việc bằng cách lấy một mô hình đã được chứng minh đã hoạt động tốt cho một nhiệm vụ cụ thể và tái sử dụng nó trong mô hình mới. Trong chương này sẽ sử dụng lại các khả năng trích xuất tính năng của trình phân lớp hình ảnh mạnh mẽ đã học trong ImageNet và tìm hiểu các lớp phân lớp mới.

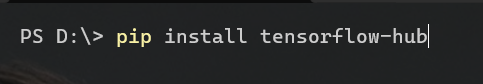
Phương pháp này hoạt động rất tốt trong nhiều ứng dụng, hoạt động với lượng dữ liệu đào tạo vừa phải (hàng trăm hình ảnh được gắn nhãn) và có thể chạy nhanh trên PC mà không cần GPU. Trong bài học này, sẽ sử dụng Trung tâm TensorFlow để thu thập các mô hình hoặc bộ phận mô hình đã xử lý trước. Sử dụng mô-đun trích xuất cấu trúc mô hình và tính năng Googles Inception V3 được học trong ImageNet để tiến hành đào tạo lại dữ liệu mới.

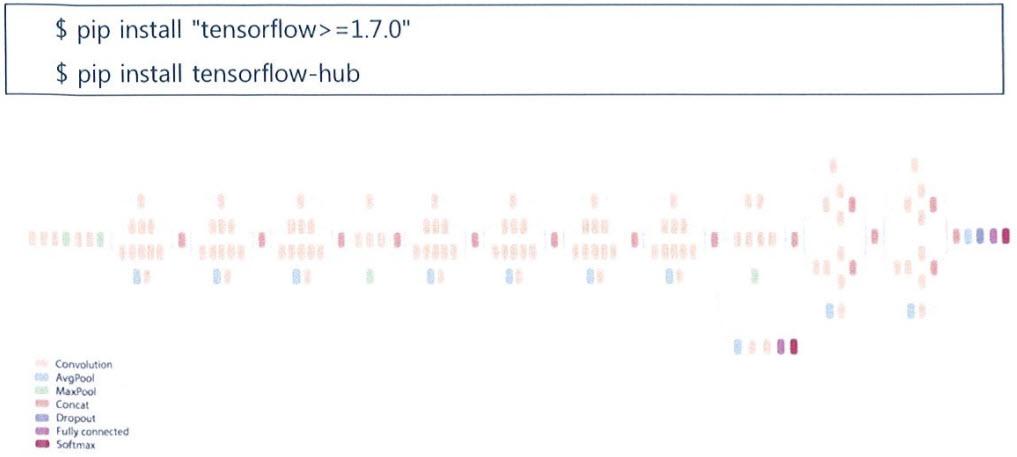
Bài học này được chạy trên hệ điều hành Windows.

## 11.2. Phương pháp nhận dạng đồ vật

Trước khi bắt đầu, cần cài đặt phiên bản mới nhất của TensorFlow và gói TensorFlow Hub. Do đó, để sử dụng TensorFlow Hub, cần cài đặt hoặc nâng cấp gói TensorFlow như sau: Mở cửa sổ lệnh Windows và chạy lệnh sau:







### 11.2.1. Học hình ảnh

Trước khi bắt đầu dạy học mô hình, cần hình ảnh của lớp mới mà muốn nhận ra. Có thể tìm hiểu bằng cách chuẩn bị hình ảnh mong muốn, nhưng trước tiên, hãy sử dụng hình ảnh chung đã chuẩn bị của một bông hoa. Để có được hình ảnh hoa, hãy tải xuống tệp nén của hình ảnh bông hoa bằng cách nhập nội dung sau vào thanh điều hướng trình duyệt web:

[*http://download.tensorflow.org/example\_images/flower\_photos.tgz*](http://download.tensorflow.org/example_images/flower_photos.tgz)

*hoặc* [*https://www.kaggle.com/datasets/alxmamaev/flowers-recognition*](https://www.kaggle.com/datasets/alxmamaev/flowers-recognition)

Nó cũng được bao gồm trong thư mục sau của CD được cung cấp:

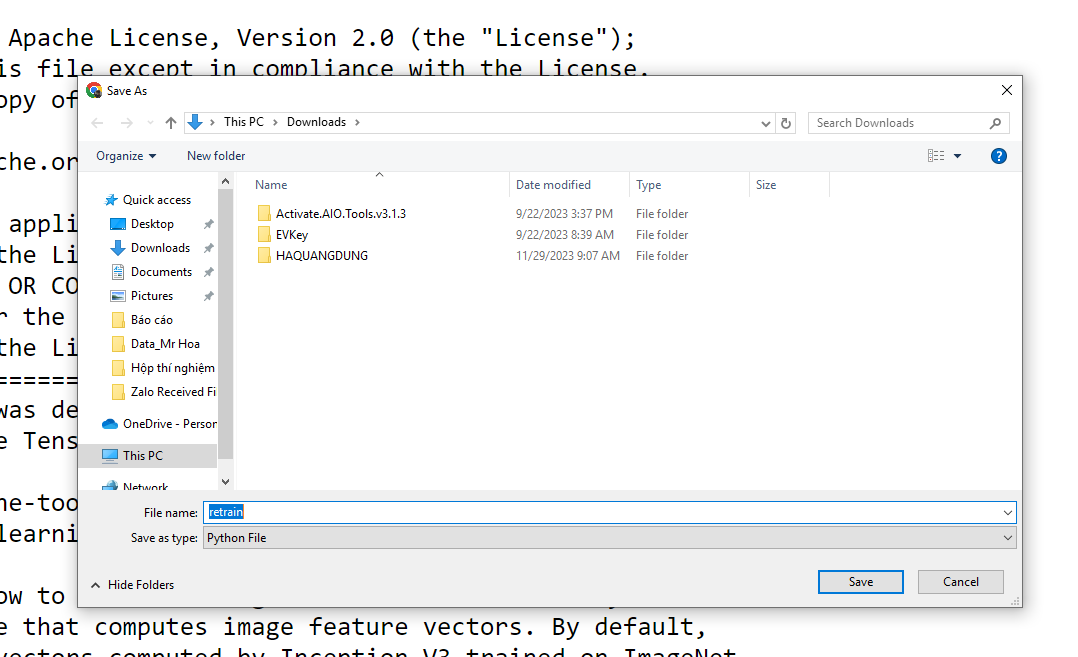
Giải nén các tập tin vào ổ đĩa cứng. Thư mục được tạo.

Trong thanh điều hướng trình duyệt web, nhập nội dung sau để gọi mã nguồn retrain.py, chọn toàn bộ mã nguồn và mở chương trình Notepad để sao chép mã.

<https://raw.githubusercontent.com/tensorflow/hub/r0.1/examples/image_retraining/retrain.py>

Trong Notepad, bấm vào File-Save As và lưu dưới dạng retrain.py ở vị trí tạo thư mục flower\_photos.

Thay vì phải sửa bằng notepad thì bạn cũng có thể nhấn tổ hợp phím Ctrl + S để lưu dưới dạng file .py

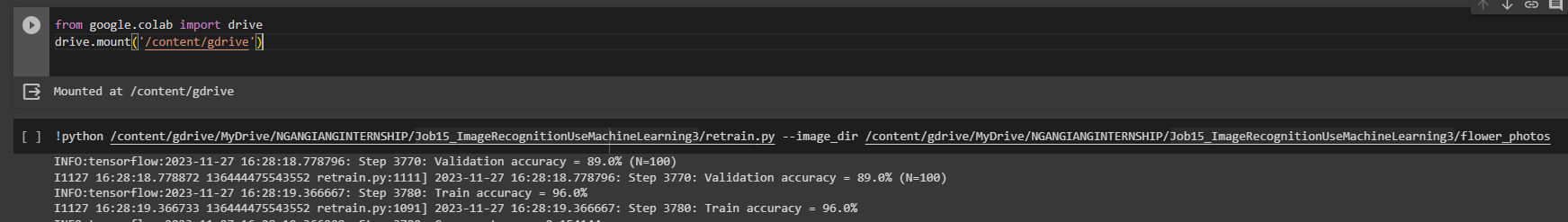


Sau đó chạy tệp bằng cách nhập lệnh sau: Chạy nó sẽ mất khoảng 30 phút tùy thuộc vào hiệu suất của PC.

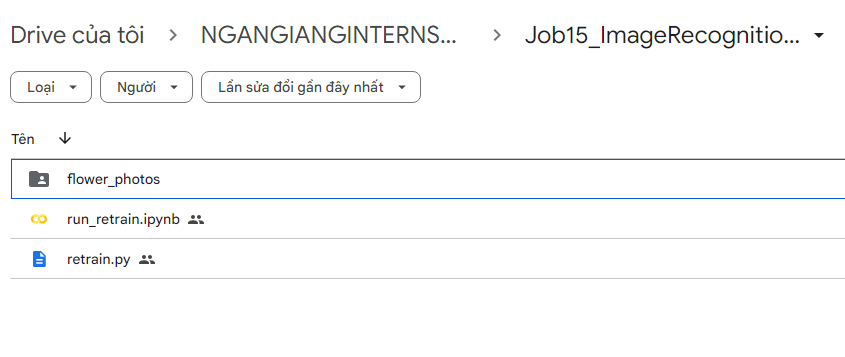


Còn nếu bạn chạy trên google colab thì hãy tạo ra 1 thư mục chứ tệp dataset flower\_photos và tệp retrain

| !python /content/drive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/retrain.py --image\_dir /content/drive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/flower\_photos |
| --- |



Ví dụ như sau:



Kịch bản này tải một mô-đun được dạy trước và dạy một trình phân lớp mới cho các bức ảnh hoa được tải xuống. Trong lớp ImageNet hiện có, nơi toàn bộ mạng được dạy, không có hình ảnh nào về các loại hoa được tải xuống ở trên. Chuyển đổi việc học có thể được sử dụng lại cho nhiều nhiệm vụ nhận thức mà không thay đổi các lớp con đã được dạy để phân biệt một số đối tượng.

### 11.2.2. Hiện tượng thắt cổ chai

Khi chạy chương trình trên, bước đầu tiên là phân tích tất cả các hình ảnh trên đĩa, tính toán nút cổ chai cho mỗi hình ảnh và lưu nó. Nút cổ chai là một thuật ngữ không chính thức đề cập đến lớp trước lớp cuối cùng của sự phân lớp. Lớp thứ hai của lớp cuối cùng được dạy để in ra các giá trị đầu ra để có thể đủ phân biệt giữa các lớp, có nghĩa là nó phải là một bản tóm tắt hình ảnh có ý nghĩa và ngắn gọn cho đến khi nó chứa đủ thông tin để bộ phân lớp đưa ra lựa chọn tốt trong một bộ sưu tập các giá trị rất nhỏ. Lớp cuối cùng của hoạt động học lại hoạt động tốt trong lớp mới vì rất hữu ích khi xác định loại đối tượng mới cần thiết để xác định thông tin cần thiết để phân biệt giữa 1000 lớp ImageNet.

Bởi vì tất cả các hình ảnh lặp đi lặp lại cần một lượng thời gian đáng kể để tính toán và dạy từng nút cổ chai, phải lưu trữ các giá trị nút cổ chai trên đĩa để chúng không được tính toán lại nhiều lần. Theo mặc định, các tập lệnh được lưu trữ trong thư mục / imp / bottleneck, vì vậy nếu chạy lại tập lệnh được sử dụng lại, không phải đợi nó nữa.

### 11.2.3. Dạy học cho máy

Khi nút cổ chai được hoàn thành, việc học thực tế của lớp trên của mạng được tiến hành. Có thể xem kết quả của các nhân viên khác nhau, mỗi người có “Độ chính xác đào tạo, “Độ chính xác xác thực” và “Entropy chéo” khác nhau. Nhìn vào từng giá trị,

* + Độ chính xác đào tạo: Tỷ lệ phần trăm dữ liệu được dự đoán và dán nhãn chính xác trong đợt đào tạo hiện tại
  + Độ chính xác xác thực: Độ chính xác của các nhóm hình ảnh được chọn ngẫu nhiên từ các bộ khác nhau.
  + Nếu Độ chính xác đào tạo được sử dụng như một chỉ số quan trọng, thậm chí tiếng ồn có thể được học và xảy ra quá khớp. Thước đo thực tế của hiệu suất mạng sẽ xác nhận hiệu suất của dữ liệu không có trong dữ liệu dạy học (= độ chính xác xác thực)
  + Độ chính xác cho dữ liệu học cao, trong khi độ chính xác của xác thực thấp, điều đó có nghĩa là nó bị quá khớp, là một kết quả không tốt.
  + Entropy chéo: Hàm mất mát cho thấy quá trình học tập tiến triển tốt như thế nào.
* Hàm mất mát cho thấy quá trình học tập tiến triển tốt như thế nào.

Theo mặc định, tập lệnh này thực hiện 4.000 bước đào tạo. Mỗi bước chọn 10 hình ảnh ngẫu nhiên từ tập dạy học, tìm từng nút cổ chai từ bộ đệm và đưa nó đến lớp cuối cùng để dự đoán. Bằng cách so sánh dự đoán này với nhãn thực tế, các trọng số của lớp cuối cùng được cập nhật thông qua quá trình lan truyền ngược.

Khi quá trình tiến triển, cần phải thấy độ chính xác tăng lên. Sau khi tất cả các bài học kết thúc, kiểm tra độ chính xác cuối cùng xuất hiện bằng cách dự đoán một bộ được đặt sang một bên để thử nghiệm. Đây là chỉ số tốt nhất về mô hình được đào tạo tốt như thế nào.

Con số chính xác từ 90 đến 95% phải được hiển thị, nhưng giá trị chính xác thay đổi tùy theo thực tế, vì mỗi nhân viên chọn ngẫu nhiên hình ảnh trong quá trình dạy học. Con số này dựa trên tỷ lệ phần trăm của hình ảnh trong bộ thử nghiệm với nhãn chính xác sau khi mô hình đã được dạy đầy đủ.

### 11.2.4. Sử dụng mô hình đào tạo lại

Kịch bản tạo một mô hình mới được đào tạo cho các danh mục trong **/tmp/output\_graph.pb** và một tệp văn bản có nhãn trong **/tmp/output\_labels.txt**. Mô hình mới được gắn với mô-đun TP-Hub và bao gồm một lớp phân lớp mới. Cả hai tệp có định dạng có thể được đọc bởi trình phân lớp hình ảnh C++ và Python để mô hình có sẵn ngay lập tức.

Nếu đã sử dụng label\_image vì đã thay thế lớp cấp cao nhất

Phải đặt cho script một tên mới với --output\_layer = final\_result.flag.

Sau đây là một ví dụ về cách chạy ví dụ label\_image với biểu đồ đào tạo lại. Theo quy ước, tất cả các mô-đun TensorFlow Hub chấp nhận đầu vào hình ảnh với các giá trị màu trong một phạm vi cố định [0,1]. Vì vậy, không cần thiết phải đặt --input\_mean hoặc --input\_std.flag.

Nhập nội dung sau vào Thanh điều hướng trình duyệt web để hiển thị mã nguồn label\_image.py, chọn tất cả. và mở chương trình Notepad.

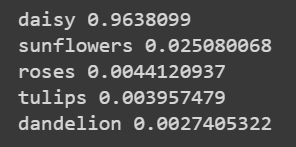
<https://raw.githubusercontent.com/tensorflow/tensorflow/master/tensorflow/examples/label_image/label_image.py>

Trong chương trình Notepad, nhấp vào File-Save As và lưu dưới dạng label\_image.py ở vị trí tạo thư mục flower\_photos.

Sau đó chạy tệp bằng cách nhập lệnh sau: Khi chạy nó. phải mất khoảng 30 phút tùy thuộc vào hiệu suất PC.

| python label\_image.py  --graph=/tmp/output\_graph.pb --labels=/tmp/output\_labels.txt  --input\_layer=Placeholder  --output\_layer=final\_result  --image=/content/gdrive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/flower\_photos/daisy/21652746\_cc379e0eea\_m.jpg |
| --- |

Có thể thấy kết quả như sau:



Hình ảnh hoa cúc được chia thành xác suất 96%. Trong hầu hết các trường hợp, hoa cúc nên xuất hiện trong danh sách nhãn hoa ở đầu. (Các mô hình được dạy lại có thể hơi khác nhau)

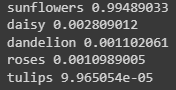
Cũng có thể thay thế tham số --image bằng hình ảnh của riêng mình. Nói cách khác, nó tải xuống và nhận ra hình ảnh mới của chính nó chứ không phải hình ảnh có trong thư mục hiện tại. Tại thời điểm này, hãy tải xuống hình ảnh của một trong những bông hoa chúng ta đã học trong thư mục flower\_photos và nhận ra như sau. Hình ảnh của hoa hướng dương đã được sử dụng ở đây.

| python label\_image.py --graph=/tmp/output\_graph.pb  --labels=/tmp/output\_labels.txt  --input\_layer=Placeholder  --output\_layer=final\_result  --image=C:Wflower\_photosWsun\_flower.jpg |
| --- |

Ví dụ như tôi sử dụng trên google colab

| !python /content/drive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/label\_image.py --graph=/content/drive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/tmp/output\_graph.pb --labels=/content/drive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/tmp/output\_labels.txt --input\_layer=Placeholder --output\_layer=final\_result --image=/content/drive/MyDrive/NGANGIANGINTERNSHIP/Job15\_ImageRecognitionUseMachineLearning3/flower\_photos/sunflowers/1022552002\_2b93faf9e7\_n.jpg |
| --- |

Sau đó có thể thấy kết quả như sau:



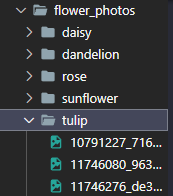
Hình ảnh hoa hướng dương được chia thành xác suất 99%. Nếu muốn sử dụng mô hình được dạy lại trong chương trình Python, tập lệnh nhãn ở trên là điểm khởi đầu tốt.

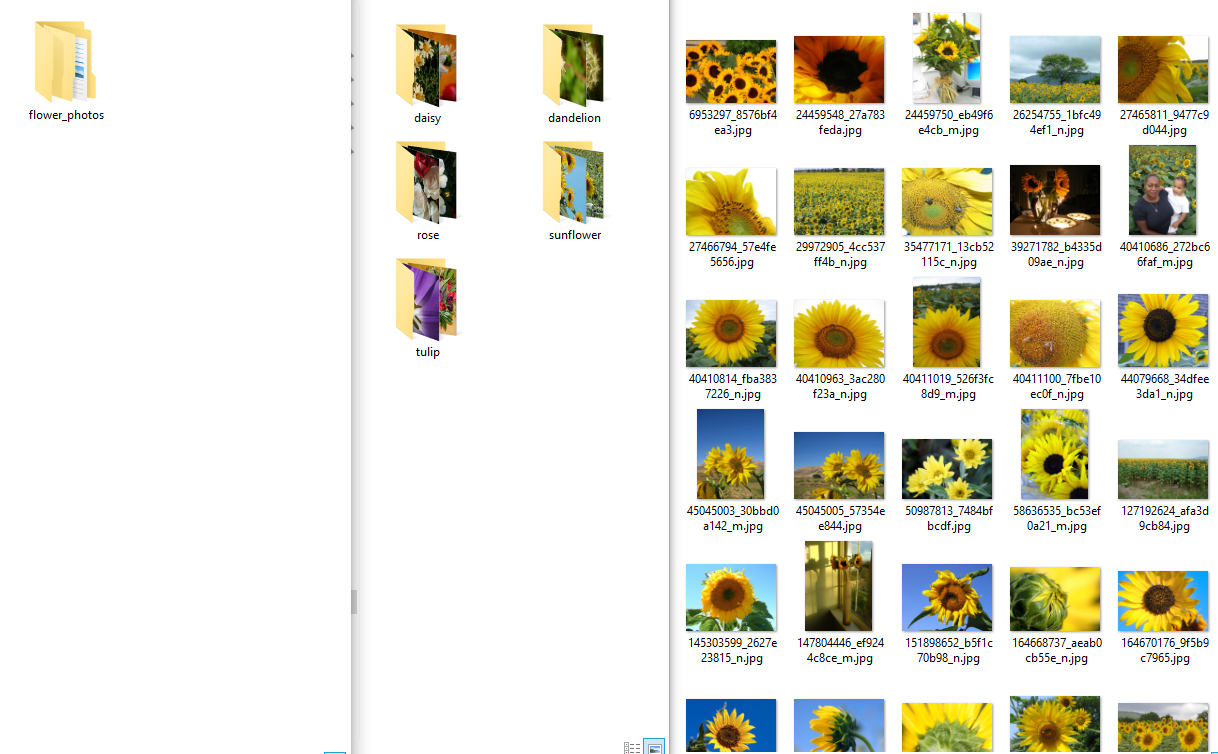
### 11.2.5. Tìm hiểu về thể loại riêng

Nếu có thể viết script từ hình ảnh ví dụ hoa, thì có thể bắt đầu học cách nhận ra các danh mục mà cá nhân mình quan tâm. Về lý thuyết, sẽ chỉ đến một bộ sưu tập thư mục con. Tập hợp các thư mục con đặt tên cho một trong mỗi loại và chỉ chứa hình ảnh của thể loại đó.

Nếu làm như vậy và chuyển thư mục gốc của thư mục con làm đối số cho -image\_dir, script sẽ làm tương tự như ví dụ hoa trước đó.

Dưới đây cho thấy cấu trúc thư mục của ví dụ hoa và một ví dụ về bố cục mà script đang tìm kiếm:





Trong thực tế, một số công việc có thể được yêu cầu để đạt được độ chính xác mong muốn. Dưới đây là một số vấn đề phổ biến có thể gặp phải trong khi tạo các danh mục của riêng mình:

**2.5.1. Tạo một loạt các hình ảnh đào tạo**

Điều đầu tiên để bắt đầu là nhìn vào những hình ảnh thu thập được. Vấn đề phổ biến nhất chúng ta thấy trong học tập đến từ dữ liệu được cung cấp.

Để được đào tạo tốt, phải thu thập ít nhất 100 bức ảnh của từng loại đối tượng muốn nhận ra. Càng thu thập được nhiều, mô hình được đào tạo càng chính xác. Ngoài ra, phải đảm bảo rằng các bức ảnh phản ánh cho những gì thực sự gặp phải trong ứng dụng. Ví dụ, nếu tất cả các hình ảnh phản ánh một đồ nội thất đang dựa vào một bức tường trống, thì việc cố gắng nhận ra các vật thể ngoài trời không mang lại kết quả tốt.

Để tránh một sai lầm khác, quá trình học tập nên chọn nhãn hình ảnh nào mà có điểm chung. Các tính năng không phổ biến không hữu ích cho việc học. Ví dụ: nếu chụp một vật trong phòng màu xanh lam và một vật khác trong phòng màu xanh lá cây, mô hình có thể hoàn thành dạy học bằng cách xem xét dự đoán màu nền mà không cần xem xét các đặc điểm của vật thể. Để tránh vấn đề này, hình ảnh nên được chụp ở những thời điểm khác nhau, với các thiết bị khác nhau và trong nhiều tình huống khác nhau.

Cũng nên suy nghĩ về các thể loại sử dụng. Tốt hơn là chia nhỏ thành các loại nhỏ hơn, khác biệt về mặt trực quan hơn các loại lớn hơn bao gồm các dạng vật lý khác nhau. Ví dụ: có thể nhập “ô tô”, “xe máy” và “xe tải” thay vì “phương tiện giao thông”. Để ứng dụng hình ảnh vào ứng dụng nhận dạng thực vật, thì cần xác định lại thực vật đó.

Cũng nên kiểm tra rằng tất cả các hình ảnh được dán nhãn đúng chưa. Thẻ do người dùng tạo thường không đáng tin cậy vì chúng phục vụ các mục đích khác nhau. Ví dụ: một bức ảnh được gắn thẻ “hoa cúc” có thể chứa người và chữ “hoa cúc” thay vì là ảnh bông hoa cúc. Nếu xem lại hình ảnh và nhặt ra bất kỳ lỗi nào, độ chính xác tổng thể có thể được cải thiện rất nhiều.

Một khi hài lòng với hình ảnh, có thể thay đổi các chi tiết của quá trình dạy học để cải thiện kết quả. Cách đơn giản nhất để thử là “--how\_many\_training\_steps”. Giá trị mặc định cho tùy chọn này là 4000, nhưng tăng nó lên 8000 sẽ tăng gấp đôi thời gian dạy học. Sự gia tăng độ chính xác chậm lại với các buổi đào tạo dài hơn (đôi khi nó dừng hoặc giảm do quá khớp), nhưng thử nghiệm sẽ cho thấy những gì hoạt động tốt nhất cho mô hình.

## 11.3. Biến dạng

Một cách phổ biến để cải thiện kết quả đào tạo hình ảnh là sửa đổi, cắt hoặc làm sáng đầu vào đào tạo theo bất kỳ cách nào. Điều này có lợi thế là mở rộng kích thước hiệu quả của dữ liệu dạy học với tất cả các biến thể có thể có của cùng một hình ảnh, giúp tìm hiểu cách mạng có thể đối phó với tất cả các biến dạng có thể phát sinh từ việc sử dụng bộ nhận dạng thực tế. Nhược điểm chính của việc tạo biến dạng này trong tập lệnh là thắt cổ chai không còn hữu ích vì hình ảnh đầu vào không được sử dụng lại một cách chính xác. Nói cách khác, quá trình đào tạo mất nhiều thời gian, vì vậy chỉ nên cố gắng sửa mô hình khi đã hài lòng với sự biến dạng hợp lý.

Script này sử dụng **--random\_crop**, **--random\_scale** và **--random\_brightness** để làm biến dạng. Các tùy chọn này là giá trị phần trăm và kiểm soát mức độ biến dạng được áp dụng cho mỗi hình ảnh. Giá trị bắt đầu hợp lý là 5 hoặc 10 và thử nghiệm xem chúng có giúp được gì không.

## 11.4. Siêu tham số

Có một vài thông số khác có thể được điều chỉnh để xem liệu chúng có giúp kết quả không. **--learning\_rate** kiểm soát kích thước cập nhật của lớp cuối cùng trong quá trình học. Theo trực giác, nếu nó nhỏ hơn thì việc học thì mất nhiều thời gian, nhưng nó có thể tăng độ chính xác tổng thể, nhưng không phải luôn luôn như vậy. Do đó, nên thử nghiệm cẩn thận để xem kết quả phù hợp với từng trường hợp.

**--train\_batch\_size** kiểm soát số lượng hình ảnh được kiểm tra trong mỗi bước đào tạo để đánh giá cập nhật của lớp cuối cùng.