

CÁC NHÀ NGHIÊN CỨU PHÁT TRIỂN PHẦN MỀM NHẬN DIỆN HÌNH ẢNH MỚI

Một nhóm các nhà nghiên cứu thuộc MIT đã phát hiện ra rằng chỉ cần một vài pixel thông tin là có thể nhận diện đối tượng trong một bức ảnh. Khám phá này có thể dẫn đến những tiến bộ to lớn trong viện nhận diện tự động các hình ảnh trực tuyến

Một nhóm các nhà nghiên cứu thuộc MIT đã phát hiện ra rằng chỉ cần một vài pixel thông tin là có thể nhận diện đối tượng trong một bức ảnh. Khám phá này có thể dẫn đến những tiến bộ to lớn trong viện nhận diện tự động các hình ảnh trực tuyến và nhất là cung cấp nền tảng cho máy tính nhìn được như con người.

Antonio Torralba, giảng viên tại Phòng thí nghiệm Trí thông minh nhân tạo và Khoa học máy tính của MIT, và cộng sự đã cố gắng tìm ra đâu là lượng thông tin nhỏ nhất – tức biểu hiện số ngắn nhất – có thể trích xuất được từ một hình ảnh mà đem lại được dấu hiệu hữu ích phần nội dung.

Thu được một phần biểu hiện ngắn như thế sẽ là một bước quan trọng trong việc hiện thực hóa khả năng phân loại hàng triệu hình ảnh trên Internet tự động. Hiện tại, cách duy nhất tìm ảnh được dựa trên lời chú thích ảnh bằng chữ mà mọi người nhập vào mỗi bức ảnh, mà nhiều bức ảnh lại thiếu thông tin này. Nhận diện tự động sẽ đem lại một phương pháp chú thích những bức ảnh mọi người tải từ máy ảnh kỹ thuật số vào máy tính mà không phải điền chú thích cho từng bức ảnh bằng tay. Và đặc biệt nó có thể dẫn đến hình ảnh máy chân thực, điều mà một ngày nào đó có thể cho phép hình dung được dữ liệu nhập vào từ máy ảnh và biết được chúng là gì.

Theo Torralba “Chúng tôi đang cố gắng tìm ra những đoạn mã rất ngắn cho hình ảnh để nếu có hai hình ảnh có chuỗi số tương tự, chúng có lẽ được chụp gần giống nhau hoặc gần như cùng một đối tượng, trong gần như cùng hoàn cảnh.” Nếu một ảnh được nhận diện với chú thích hoặc tiêu đề thì những tấm ảnh khác phù hợp với mã số của nó có lẽ sẽ về cùng một vật thể (ví dụ như ô tô, cây hoặc người) và vì vậy tên liên quan đến một bức ảnh có thể được chuyển đến cho những bức ảnh khác.

Theo Torralba “Với một lượng ảnh vô cùng lớn, thậm chí những phép toán khá đơn giản cũng có thể thực hiện khá tốt” trong việc nhận diện ảnh theo cách này. Torralba sẽ trình bày phát hiện của mình vào tháng 5 ở hội thảo về Nhận diện Mẫu và hình ảnh máy tính, Alaska. Công trình này được thực hiện cùng với Rob Fergus tại Viện Courant, ĐH New York và ĐH Do thái Yair Weiss, Jerusalem.

(Ảnh: Antonio Torralba)

Câu hỏi: Bạn nhìn thấy gì trong những vòng tròn màu đỏ? Một cái chai, chiếc điện thoại di động, một người và một chiếc giày? Trả lời: Chúng chỉ là một! Giảng viên Antonio Torralba đã tạo ra những bức ảnh có độ phân giải thấp này, sau đó chèn vào đối tượng trong vòng tròn đỏ, qua đó cho thấy ngoại cảnh tác động đến quá trình nhận biết vật thể của chúng ta ra sao. Ngay cả chiếc ô tô trong tấm ảnh bên dưới phía trái cũng là cùng một thứ.

Để tìm ra bao nhiêu lượng thông tin tối thiểu là đủ để mọi người nhận ra vật thể trong một bức ảnh, Torralba và đồng tác giả cố gắng giảm hình ảnh xuống độ phân giải càng lúc càng thấp, và xem mọi người có thể xác định được bao nhiêu ảnh ở mỗi cấp độ.

“Chúng tôi có thể nhận ra cái gì nằm trong bức ảnh, thậm chí độ phân giải là rất thấp vì chúng tôi biết rất nhiều về ảnh. Lượng thông tin mà bạn cần để xác định phần lớn ảnh là khoảng 32×32 .” Ngược lại, thậm chí ảnh thu nhỏ hiển hiện trên kết quả tìm kiếm của Google thường là 100×100 .

Thậm chí một máy ảnh kỹ thuật số phải chăng hiện nay cũng cho ra những bức ảnh chứa đơn vị megapixel dữ liệu – và mỗi pixel thường chứa 24 bit (0 hoặc 1) dữ liệu. Nhưng Torralba và cộng sự đã tạo nên một hệ thống toán học có thể giảm dữ liệu từ mỗi ảnh nhiều hơn và hóa ra nhiều bức ảnh đều có thể nhận diện được thậm chí khi được mã hóa thành biểu thị toán học chứa 256 đến 1024 bit dữ liệu.

Sử dụng lượng thông tin nhỏ như thế trên mỗi ảnh khiến ta có thể tìm kiếm những bức ảnh tương tự nhau trong hàng triệu bức ảnh trong cơ sở dữ liệu, sử dụng một máy tính cá nhân với thời gian ít hơn 1 giây. Và không như những phương pháp khác đòi hỏi phải phân tách từng bức ảnh thành những mục nhỏ chứa những vật thể khác nhau, phương pháp này sử dụng toàn bộ ảnh, khiến cho

việc áp dụng nó vào những bộ dữ liệu lớn dễ dàng hơn mà không cần sự can thiệp của con người.

Ví dụ, sử dụng hệ thống mã hóa mà họ phát triển, Torralba và cộng sự có thể hiển thị 12.9 triệu ảnh từ Internet chỉ với 600 megabyte dữ liệu – vừa đủ nhỏ để vừa với bộ nhớ RAM của phần lớn máy tính cá nhân hiện tại, và có thể lưu trữ trên một thẻ nhớ. Dữ liệu ảnh và phần mềm tìm kiếm dữ liệu hiện đang được công bố trên mạng.

Dĩ nhiên, một hệ thống giảm thiểu tối đa lượng thông tin không thể đạt đến nhận diện hoàn toàn. Hiện tại, phương pháp ghép nối hoạt động trên các loại hình ảnh phổ biến nhất. “Không phải tất cả các ảnh được tạo ra như nhau.” Hình ảnh càng phức tạp hoặc lạ lùng thì nó càng khó ghép nối. Nhưng đối với phần lớn các vật thể phổ biến nhất trong ảnh – người, hoa, ô tô, tòa nhà – kết quả thu được rất ấn tượng.

Công trình này là một phần của nghiên cứu được thực hiện bởi hàng trăm nhóm trên toàn thế giới, nhằm vào việc phân tích nội dung của thông tin hình ảnh. Torralba cũng cộng tác trong các công trình liên quan với những nhà nghiên cứu MIT khác như William Freeman, giảng viên tại Khoa Điện tử và Khoa học máy tính; Aude Oliva, giảng viên Khoa Khoa học não và nhận thức, nghiên cứu sinh tiến sĩ Bryan Russell và Ce Liu, thuộc CSAIL. Công trình của Torralba được hỗ trợ một phần thông qua Quỹ Tài trợ khoa học quốc gia.

Torralba nhấn mạnh rằng công trình này vẫn đang ở giai đoạn sơ khai và còn có nhiều vấn đề trong việc nhận dạng những vật thể bất thường hơn. Nó cũng giống như cách chúng ta nhận biết ngôn ngữ. “Có nhiều từ bạn nghe rất thường xuyên, nhưng cho dù bạn có sống ở đây bao lâu đi chăng nữa, cũng sẽ có từ bạn chưa hề biết đến. Bạn luôn luôn cần hiểu được một điều gì mới từ một ví dụ.”