

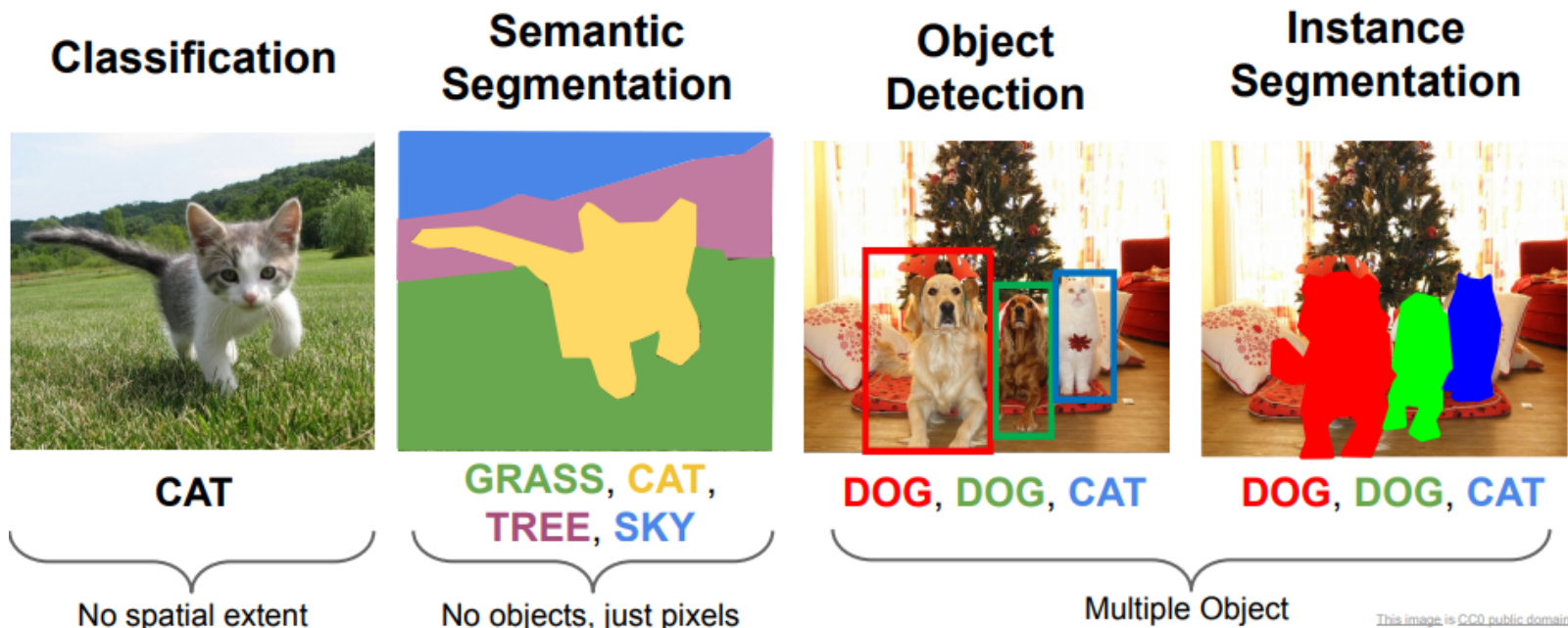
Computer vision

1. Giới thiệu chung về Computer vision

Computer vision là một lĩnh vực con của machine learning và deep learning giúp máy tính có khả năng hiểu và xử lý được thông tin dưới dạng hình ảnh.

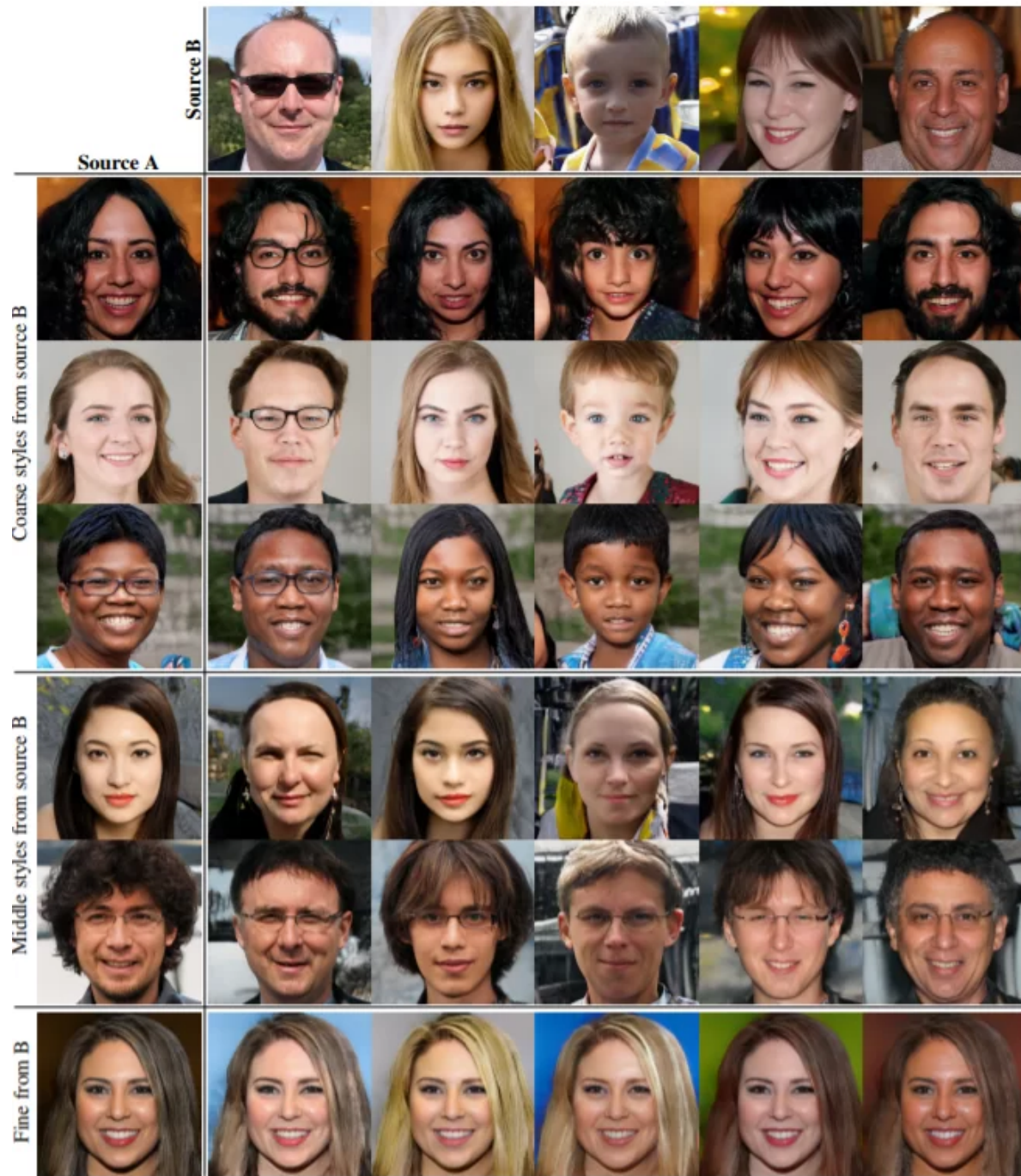
Các mô hình giải quyết bài toán Computer vision thường hoạt động trên dữ liệu dạng hình ảnh hoặc video. Bên cạnh hình ảnh và video được chụp bằng máy ảnh thông thường, các mô hình computer vision còn có thể sử dụng những dạng hình ảnh đặc biệt như ảnh chụp x quang, ảnh sóng âm, ảnh sóng điện từ ... Bên cạnh dữ liệu hình ảnh, các mô hình computer vision có thể còn sử dụng thêm các thông tin đến từ các dữ liệu loại khác để giúp gia tăng chất lượng của mô hình.

2. Các bài toán computer vision



Computer vision là một lĩnh vực gồm nhiều bài toán con, mỗi bài toán con lại có các đặc điểm khác nhau và yêu cầu label của dữ liệu train khác nhau:

- Classification: là bài toán phân lớp nhận đầu vào là một ảnh và trả đầu ra là lớp tương ứng với ảnh đó.
- Object detection: là bài toán kết hợp giữa hai bài toán object localization và object classification. Trong đó, mô hình giải quyết bài toán object detection đầu tiên sẽ định vị vị trí có thể chứa đối tượng trong ảnh (object localization), sau đó mô hình sẽ thực hiện phân lớp đối tượng để nhận diện đối tượng đó là đối tượng nào (object classification)
- Semantic segmentation: là một dạng khác của bài toán classification, trong đó, mô hình, thay vì phân lớp trên cả ảnh, sẽ phân lớp từng pixel trên ảnh thuộc lớp nào
- Instance segmentation: là một phiên bản cao hơn của semantic segmentation, bên cạnh việc phân lớp các pixel trên ảnh thuộc lớp nào, đối với những pixel thuộc cùng một lớp, mô hình cần phải phân lớp rõ pixel đó thuộc đối tượng nào



- Image generation: là bài toán yêu cầu mô hình sinh ra dữ liệu ảnh mới. Bên trong của bài toán Image generation gồm một số bài toán con như:
 - Image synthesis: sinh ảnh từ noise
 - Image-to-image translation: sinh ảnh mới từ ảnh ban đầu
 - Inpainting: vẽ thêm hình ảnh vào khoảng trống trong ảnh ban đầu
 - Outpainting: mở rộng vẽ thêm hình ảnh từ một hình ảnh ban đầu
- Image captioning: bài toán image captioning là bài toán xây dựng một câu hoặc một đoạn văn mô tả một bức ảnh ban đầu

<p>A young boy is playing basketball.</p> 	<p>Two dogs play in the grass.</p> 	<p>A dog swims in the water.</p> 	<p>A little girl in a pink shirt is swinging.</p> 
<p>A group of people walking down a street.</p> 	<p>A group of women dressed in formal attire.</p> 	<p>Two children play in the water.</p> 	<p>A dog jumps over a hurdle.</p> 

- Image aesthetics assessment: Đánh giá mức độ thẩm mỹ của ảnh, so sánh, sắp xếp và lựa chọn ảnh đẹp hơn trong danh sách các ảnh.



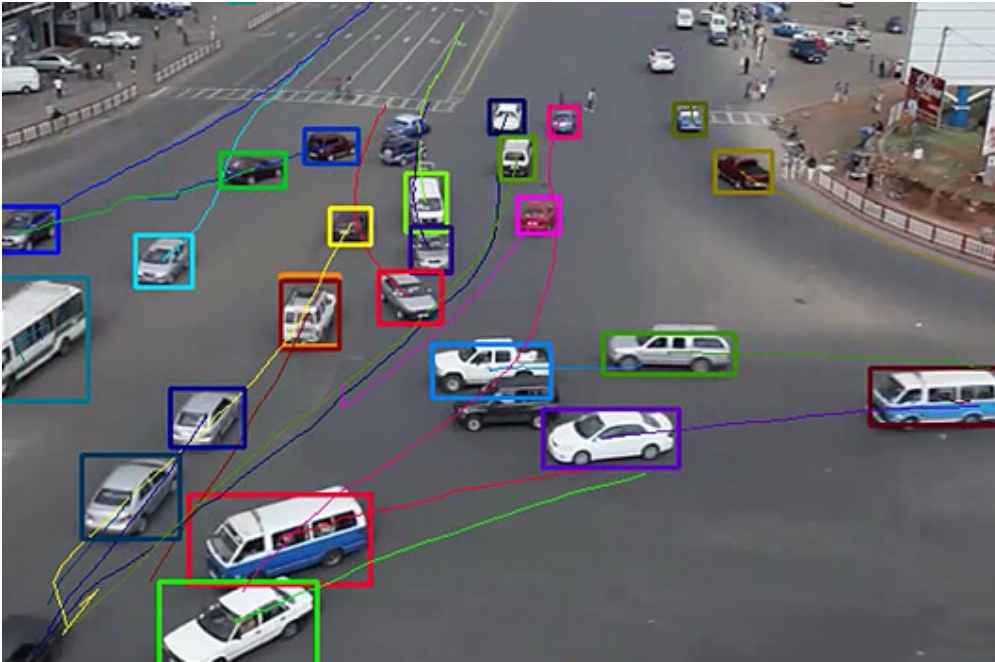
Đối với một số dạng dữ liệu phổ biến và quan trọng như dữ liệu khuôn mặt, ta cũng có một số bài toán như:

- Face detection: tương tự với object detection, tuy nhiên, bài toán face detection đòi hỏi mô hình đạt độ chính xác cao hơn và tốc độ chạy nhanh hơn, tiệm cận với real-time
- Face attribute classification: là bài toán phân tích các đặc tính trên khuôn mặt người như giới tính, cảm xúc, độ tuổi, chủng tộc ...
- Face recognition: là bài toán rất quan trọng và phổ biến hiện nay, face recognition là bài toán định danh khuôn mặt, đóng vai trò then chốt trong rất nhiều các hệ thống bảo mật hiện nay. Face recognition gồm hai bài toán con là:
 - Face identification: nhận đầu vào là một ảnh gương mặt người và trả đầu ra là định danh của gương mặt đó
 - Face verification: nhận đầu vào là hai ảnh gương mặt người và trả đầu ra là kết quả hai gương mặt có phải của cùng một người hay không.



Bên cạnh hình ảnh, video cũng là một loại dữ liệu quan trọng nhận được nhiều sự quan tâm trong computer vision. Về cơ bản, nếu coi video là một chuỗi các hình ảnh được sắp xếp với nhau theo trình tự thời gian, ta hoàn toàn có thể sử dụng các mô hình xử lý dữ liệu ảnh để áp dụng cho video. Tuy nhiên, rào cản và cũng là bài toán lớn nhất mà các mô hình xử lý video cần phải giải quyết là khối lượng tính toán lớn, dẫn đến thời gian xử lý lâu và chi phí vận hành hệ thống lớn. Điều này là động lực để các nhà nghiên cứu phát triển những mô hình xử lý video với tốc độ cao và chất lượng được đảm bảo.

Bài toán ứng phổ biến nhất đối với dữ liệu video là object tracking, theo dõi đối tượng trong suốt thời lượng của video.



3. Ảnh số - digital image

Ngày trước, hình ảnh sau khi được chụp sẽ được lưu giữ lại thông qua phim. Ngày nay, với sự phát triển của máy tính, hình ảnh đã được lưu giữ lại thông qua các thiết bị số như thẻ nhớ, ổ cứng.

Nhằm phục vụ cho các bài toán computer vision, các dữ liệu hình ảnh hay video thường được sử dụng dưới dạng ảnh raster. (Có một dạng ảnh số khác nhưng ít được sử dụng trong computer vision là ảnh vector). Các điểm ảnh được sắp xếp thành các ma trận. Trong đó, mỗi một điểm ảnh được đại diện bởi một hoặc nhiều giá trị số khác nhau. Với cùng một bức ảnh, các giá trị số tương ứng với mỗi điểm ảnh có thể khác nhau tùy thuộc vào không gian màu mà ta sử dụng để đọc ảnh.

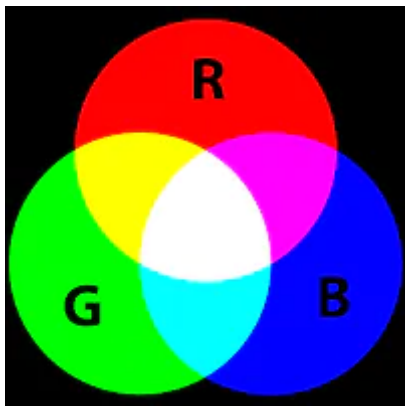
3.1. Không gian màu RGB

RGB là không gian màu phổ biến dùng trong máy tính, máy ảnh, điện thoại và nhiều thiết bị kỹ thuật số khác. Không gian màu này khá gần với cách mắt người tổng hợp màu sắc. Nguyên lý cơ bản là sử dụng 3 màu sắc cơ bản R (red - đỏ), G (green - xanh lục) và B (blue - xanh

lam) để biểu diễn tất cả các màu sắc. Mỗi một màu sắc cơ bản sẽ gồm các giá trị từ 0 đến 255, do đó, số lượng màu tối đa thường được sử dụng là

$$256 \times 256 \times 256 = 16,777,216$$

Với việc mỗi điểm ảnh được cấu thành từ ba giá trị màu R - G - B, một ảnh bất kỳ sẽ được biểu diễn bởi một ma trận có kích thước $H \times W \times 3$

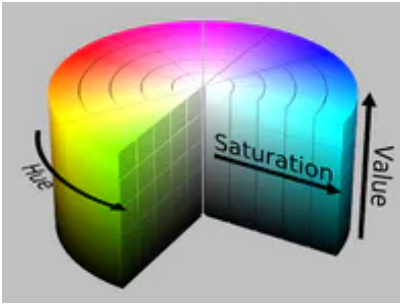


3.2. Không gian màu HSV hay HSB

Không gian màu HSV (còn gọi là HSB) là một cách tự nhiên hơn để mô tả màu sắc, dựa trên 3 số liệu:

- H: (Hue) Vùng màu
- S: (Saturation) Độ bão hòa màu
- B (hay V): (Bright hay Value) Độ sáng

HSV thường được sử dụng khi ta cần phân tích kỹ hơn về vùng sáng vùng tối của ảnh hay vùng có độ tương phản cao và vùng có độ tương phản thấp.



Một số không gian màu khác ít được sử dụng hơn như CIE LAB, CMYK ...

4. Computer vision workflow và vai trò của Image processing

Tương tự như workflow nói chung trong quá trình xây dựng mô hình deep learning, computer vision workflow cũng trải qua một số bước như chuẩn bị dữ liệu, tiền xử lý dữ liệu, xây dựng mô hình, huấn luyện và đánh giá mô hình, kiểm thử mô hình, triển khai mô hình.

Tuy nhiên, đối với computer vision, bước tiền xử lý dữ liệu đòi hỏi những thao tác đặc thù với dữ liệu hình ảnh. Những thao tác này được gọi chung là Image processing.

Trong thời gian trước, các kỹ thuật Image processing đã phát triển đến mức có thể giúp chúng ta giải quyết một số bài toán với dữ liệu đơn giản mà không cần xây dựng các mô hình machine learning. Ngày nay, với sự tiện dụng trong việc xây dựng các mô hình machine learning, việc trực tiếp sử dụng các kỹ thuật image processing trong việc giải các bài toán đã không còn quá phổ biến. Tuy nhiên, các kỹ thuật image processing vẫn đóng vai trò rất quan trọng trong việc làm sạch dữ liệu ảnh, gia tăng dữ liệu ảnh và từ đó, cải thiện độ chính xác của các mô hình machine learning hay deep learning trên dữ liệu ảnh.

