

Ghi chép bài giảng: Stanford CS231n

Lecture 1: Introduction to Computer Vision and Deep Learning

Ho Hong Phuc Nguyen

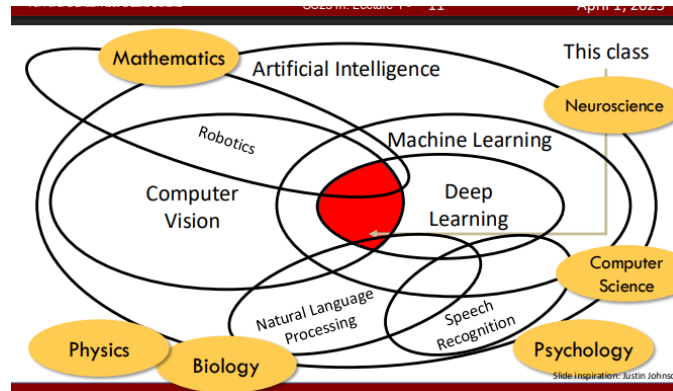
23 tháng 1, 2026

Mục lục

1	Giới thiệu chung về Computer Vision (CV)	3
2	Sơ lược lịch sử nghiên cứu Computer Vision	3
2.1	Thập niên 1950 - 1960: Nền tảng thần kinh học	3
2.2	Thập niên 1960 - 1990: Những bước đi đầu tiên	3
3	Sự trỗi dậy của Deep Learning và ImageNet	4
3.1	AI Winter và sự trở lại	4
3.2	Dự án ImageNet	4
3.3	AlexNet và Bước ngoặt thay đổi lịch sử AI	4
3.3.1	AlexNet là gì?	4
3.3.2	Tại sao AlexNet giúp AI mạnh trở lại?	4
4	Các bài toán chính của computer vision	5

1 Giới thiệu chung về Computer Vision (CV)

Giáo sư Fei-Fei Li nói rằng "Nếu AI là một bong bóng lớn, thì CV là một phần không thể tách rời, sử dụng các công cụ từ học máy (Machine Learning) và đặc biệt là học sâu (Deep Learning), thị giác máy tính không chỉ là một phần của trí tuệ nhân tạo, mà nó chính là nền tảng của trí tuệ." Computer Vision là một lĩnh vực đa ngành, giao thoa với toán học, thần kinh học, khoa học máy tính, tâm lý học và có ứng dụng rộng rãi từ y tế, luật pháp đến giáo dục.



2 Sơ lược lịch sử nghiên cứu Computer Vision

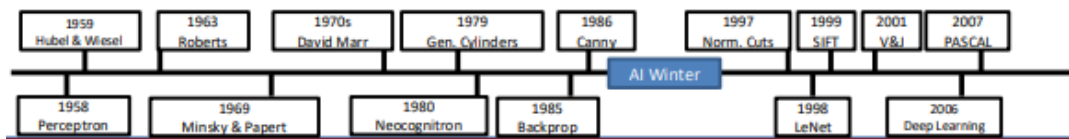
2.1 Thập niên 1950 - 1960: Nền tảng thần kinh học

Thí nghiệm của Hubel và Wiesel trên mèo đã phát hiện ra hai điều quan trọng:

- **Receptive Field (Trường thụ cảm):** Mỗi neuron chỉ phản ứng với một vùng không gian nhỏ và các mẫu đơn giản (như các cạnh định hướng).
- **Hierarchical path (Đường dẫn phân cấp):** Tín hiệu được truyền từ các tầng thấp (nhận diện cạnh) lên các tầng cao hơn để nhận diện các cấu trúc phức tạp.

2.2 Thập niên 1960 - 1990: Những bước đi đầu tiên

- **Larry Roberts (1963):** Luận án tiến sĩ đầu tiên về CV, nghiên cứu việc hiểu các hình khối 3D từ hình ảnh 2D.
- **Dự án mùa hè MIT (1966):** Một nỗ lực (quá lạc quan) nhằm giải quyết bài toán thị giác máy tính chỉ trong một mùa hè.
- **David Marr (1970s):** Đề xuất quy trình xử lý từ sơ đồ thô (edges), sơ đồ 2.5D (độ sâu) đến mô hình 3D hoàn chỉnh.
- **Fukushima (1980s):** thiết kế Neocognitron - mạng thần kinh tích chập sơ khai.
- **Rumelhart và Hinton. (1986s):** Thiết kế và chuẩn hóa thuật toán Backpropagation
- **Yann LeCun (1990s):** thiết kế CNN để nhận diện chữ số bưu điện (LeNet-5).



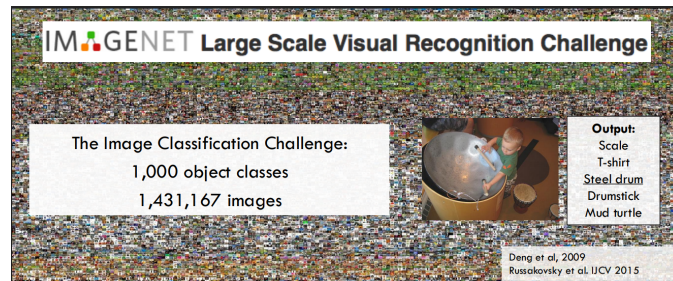
3 Sự trỗi dậy của Deep Learning và ImageNet

3.1 AI Winter và sự trở lại

Sau giai đoạn AI winter, giai đoạn mà sự quan tâm, kinh phí dành cho AI bị suy giảm nghiêm trọng do những kỳ vọng quá cao không được đáp ứng, lĩnh vực này đã bùng nổ trở lại nhờ sự kết hợp của ba yếu tố: thuật toán, tính toán và dữ liệu.

3.2 Dự án ImageNet

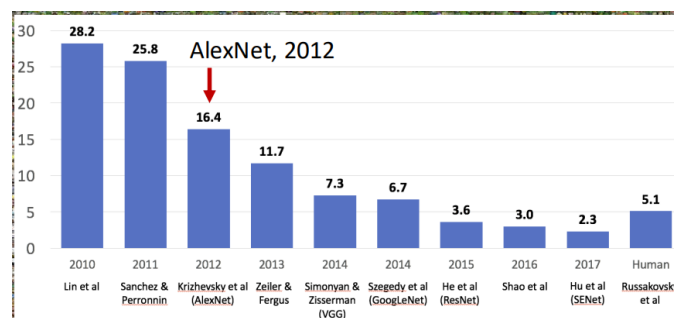
Giáo sư Fei-Fei Li và các cộng sự đã nhận ra tầm quan trọng bị đánh giá thấp của dữ liệu. Họ tạo ra ImageNet với hơn 14 triệu ảnh thuộc 22.000 danh mục. Cuộc thi ImageNet (ILSVRC) ra đời để thách thức các thuật toán nhận diện vật thể.



3.3 AlexNet và Bước ngoặt thay đổi lịch sử AI

3.3.1 AlexNet là gì?

- **Tác giả:** Được phát triển bởi Alex Krizhevsky, Ilya Sutskever và Geoffrey Hinton (Đại học Toronto) vào năm 2012.
- **Cấu trúc:** Là một mạng thần kinh tích chập (CNN) sâu 8 lớp. Mặc dù về lý thuyết không khác biệt quá lớn so với các kiến trúc cũ như Neocognitron, nhưng nó được thiết kế để chạy cực nhanh trên phần cứng hiện đại.
- **Thành tích:** Chiến thắng cuộc thi ImageNet (ILSVRC 2012) với tỷ lệ lỗi chỉ khoảng 16.4%, trong khi các phương pháp truyền thống trước đó đều có tỷ lệ lỗi trên 25%.



3.3.2 Tại sao AlexNet giúp AI mạnh trở lại?

Sự thành công của AlexNet đã đánh dấu sự kết thúc của "mùa đông AI" và khẳng định tầm quan trọng tối thượng của 3 yếu tố hội tụ (The Big Three):

- **Tầm quan trọng của Dữ liệu (Data):** Trước AlexNet, các nhà khoa học tin rằng thuật toán là tất cả. AlexNet chứng minh rằng **ImageNet** (Dữ liệu lớn) mới là "nhiên liệu" cần thiết để các mô hình lớn có thể học được những đặc trưng phức tạp mà con người không thể thiết kế thủ công.

- **Sức mạnh của Phần cứng (GPU):** AlexNet là mô hình đầu tiên tận dụng tối đa sức mạnh của GPU NVIDIA để huấn luyện mạng thần kinh. Điều này cho phép thực hiện hàng tỷ phép tính song song, biến việc huấn luyện từ hàng tháng xuống còn vài ngày.
- **Học đặc trưng tự động (End-to-End Learning):** Thay vì con người phải tự nghĩ ra các đặc trưng (Hand-crafted features) như SIFT hay HOG, AlexNet sử dụng Backpropagation để tự học xem đặc trưng nào là quan trọng nhất từ dữ liệu thô.

4 Các bài toán chính của computer vision

Giáo sư Ehsan Adeli trình bày các task quan trọng trong CV:

- **Image Classification:** Gán một nhãn cho toàn bộ ảnh.
- **Semantic Segmentation:** Gán nhãn cho từng pixel (không phân biệt cá thể).
- **Object Detection:** Xác định vị trí vật thể bằng khung hình (bounding box).
- **Instance Segmentation:** Kết hợp nhận diện và phân đoạn từng cá thể riêng biệt.

