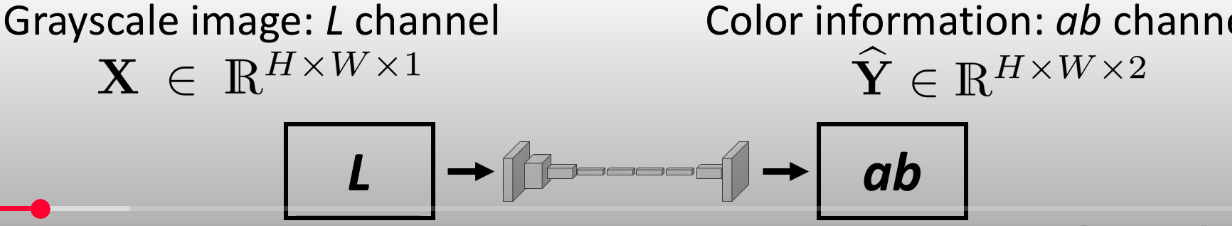
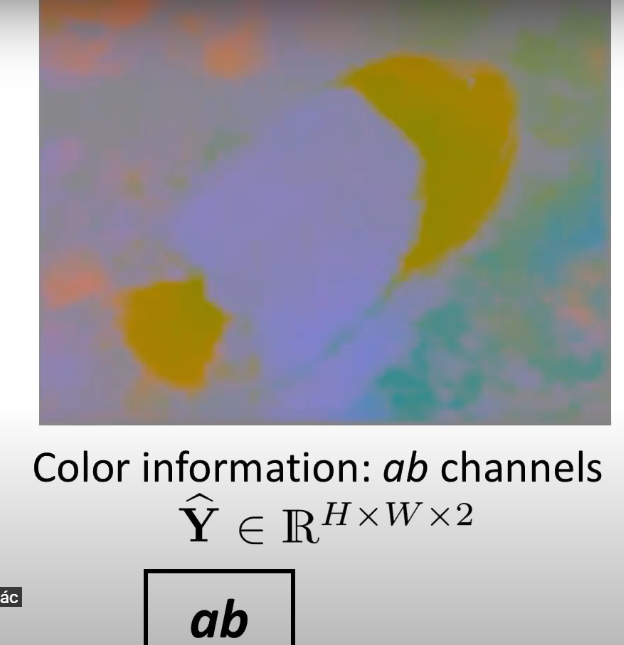
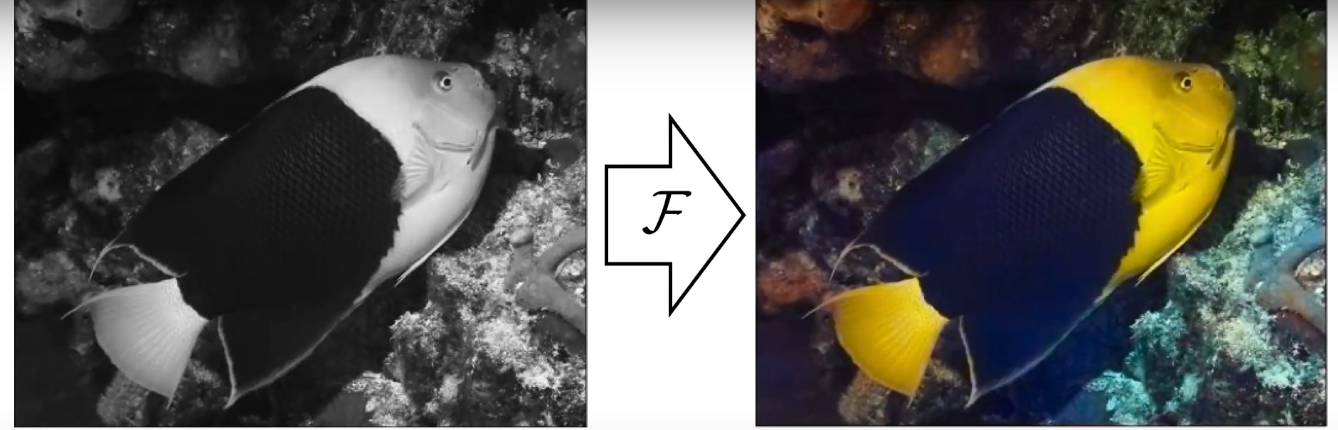
Chúng ta đang làm việc tỏng khong gian màu lab, 

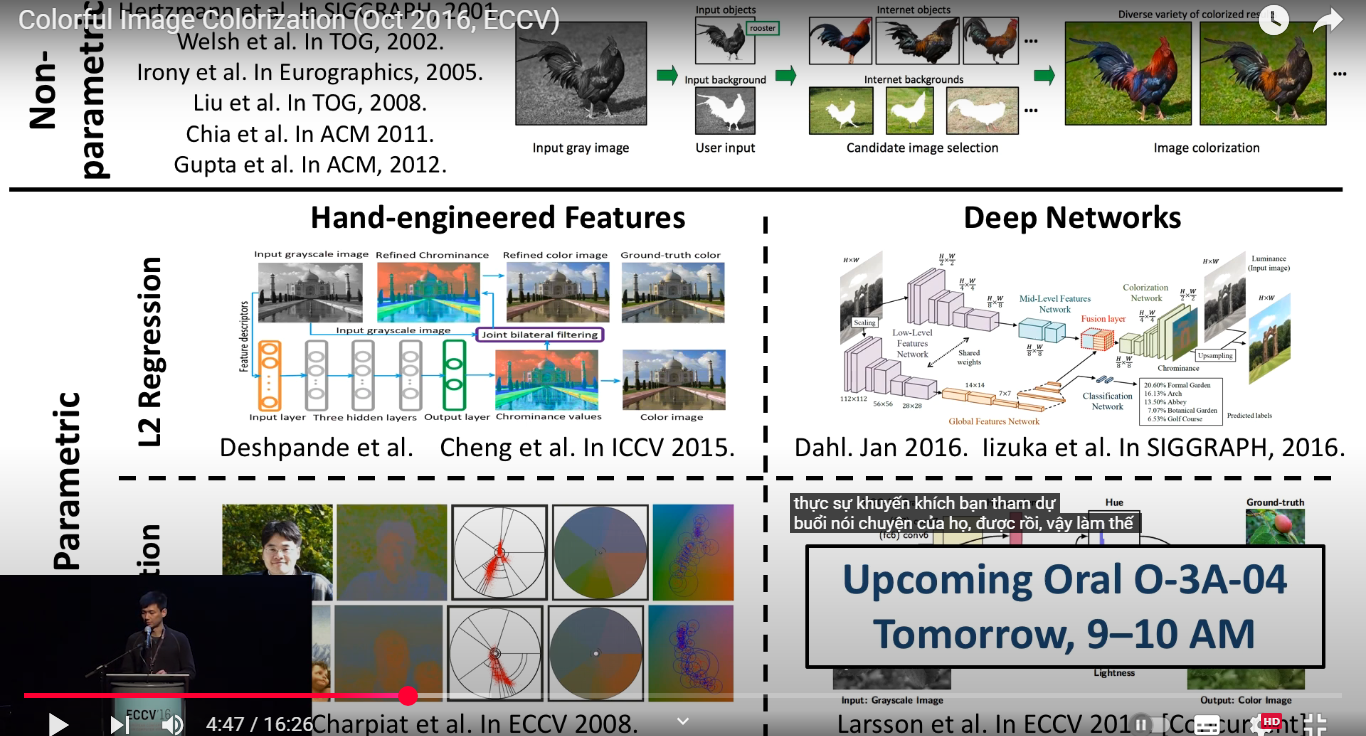
vì vậy kênh L có thông tin thang độ xám và đóng vai trò là đầu vào cho hệ thống của chúng ta

  
- Chúng ta đang tìm cách dự đoán các kênh A b hoặc thông tin màu và chúng ta học ánh xạ từ LAB bằng cách sử dụng CNN, sau đó, chúng ta có thể lấy đầu ra là các kênh a b đã dự đoán, nối

Chúng với đầu vào và hy vọng có được màu sắc hợp lý của ảnh thang độ xám đầu vào



- Các bài báo liên quan



- Làm thế nào để chúng ta ánh xạ từ hình ảnh độ sáng đầu vào thành màu đầu ra?

- Đầu tiên chúng ta lưu ý rằng chúng ta đã chuyển đổi vấn đề này thành vấn đề mà chúng ta dự đoán phân phối có thể xảy ra cho mọi pixel và như vậy, chúng ta có thể sử dụng nhiều hiểu biết sâu sắc và tiến bộ trong tài liệu phân đoạn ngữ nghĩa để giúp giải quyết vấn đề của mình, vì vậy, chúng ta bắt đầu với mạng VGG ( VGG Sử dụng nó để trích xuất đặc trưng nhằm đánh giá chất lượng màu sắc của ảnh tô màu, Khi đánh giá chất lượng ảnh tô màu, họ sử dụng VGG-16 đã được huấn luyện trên ImageNet để tính toán độ giống nhau giữa ảnh tô màu và ảnh gốc

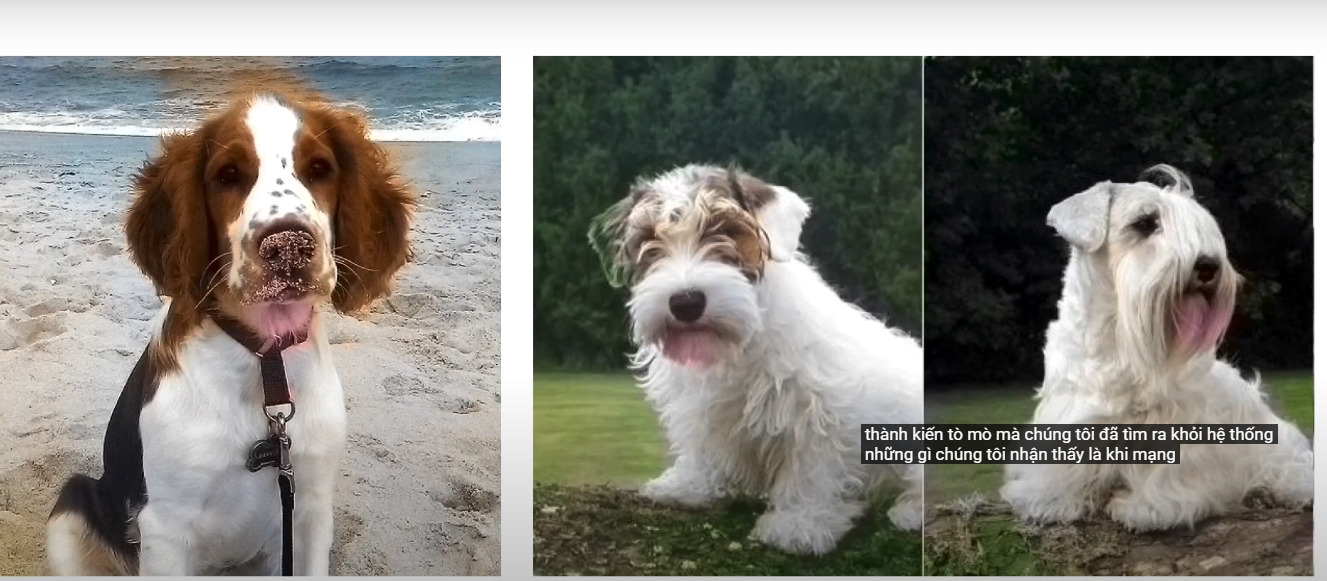
.|| Vậy VGG không phải là mô hình chính được dùng để tô màu, nhưng nó đóng vai trò quan trọng trong việc đánh giá chất lượng màu sắc bằng cách trích xuất đặc trưng cấp cao)

Đây là 1 chuyến tàu chính từ đầu, vì vậy, chỉ là kiến trúc, chúng ta loại bỏ các lớp FC trong VGG, chúng ta thêm 1 số độ phân giải không gian bổ sung và nút thắt cổ chai bằng cách sử dụng các phép tích chập thực hoặc giãn nở, chúng ta thêm một số lớp tích chập bổ sung lên trên cùng, sau đó ánh xạ các tính năng thành phân phối dự đoán cho mọi pixel. Bây giờ có một bước cuối cùng, là chuyển từ phân phối dự đoán thành ước tính điểm đơn để thực hiện việc này, chúng ta thức hiện nội suy giữa giá trị trị trung bình và chế độ và điều này cho phép chúng ta giữ được độ sống động của màu sắc đầu ra trong khi vẫn duy trì một số tính nhất quán về không gian.

- Được rồi, vậy làm thế nào để chúng làm tốt, ( Hồi quy L2 giúp mô hình **ổn định hơn**, tránh học quá mức (overfitting), và là một phần quan trọng trong việc huấn luyện và đánh giá chất lượng ảnh tô màu. )

- Trình bày ra ví dụ khi không có L2 và khi có L2 ( Nhưng ta hài lòng với kết quả là ví màu sắc hợp lý

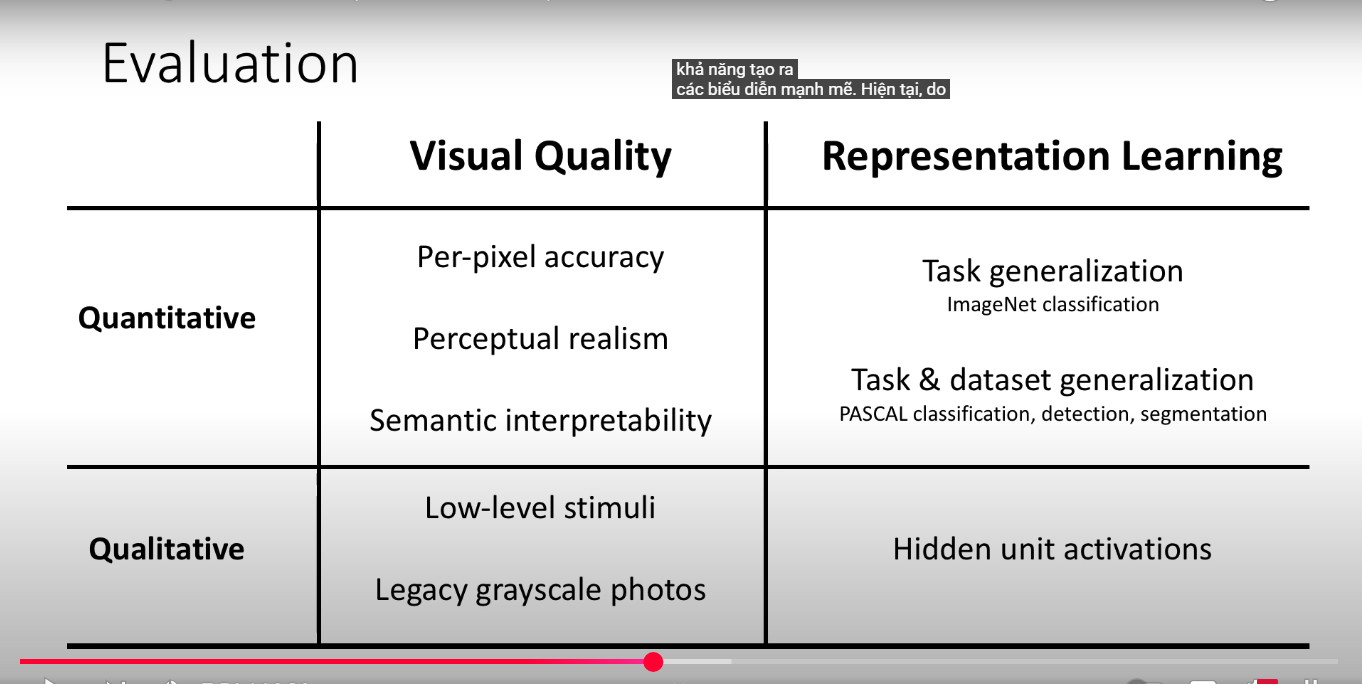
- Vẫn có những trường hợp lỗi mà chúng tôi đã quan sát thấy trong hệ thống của mình vi trong 1 trường hợp lỗi phổ biến là các vật thể do con người tạo ra thực sự có thể là một trong nhiều màu khác nhau trong mạng của chúng tôi đôi khi gặp khó khăn trong việc chọn một màu để sử dụng cuối cùng và điều này có thể dẫn đến đầu ra có một số loại mẫu nhuộm thắt nút

- Lỗi dự đoán sai ngữ cảnh liên quan đến con chó Nó đang dự đoán con chó thè lưởi ra nên tô miệng con chó màu hồng

- Vì vậy 1 trong những đóng góp của chúng tôi là thực sự suy nghĩ cẩn thận về cách đánh giá cần thận vấn đề tô màu, vì vậy các bài báo trước đây đã sử dụng các số liệu như độ chính xác trên mỗi pixel của chúng tôi chúng tôi cũng đánh giá nhứng số liệu này trong bài báo cáo của mình nhưng trên thực tế, trên mỗi pixel không nói lên sự tương tác chung giữa các pixel cũng như chất lượng nhấn thức

- **Định lượng** đề cập đến là: Độ chính xác màu ở từng điểm ảnh so với ảnh gốc, Đánh giá mức độ ảnh trông có thật hay không, có thể thông qua mạng VGG hoặc đánh giá người dùng. , Kiểm tra xem màu sắc có **hợp lý theo ngữ cảnh** không

- Định tính là: Quan sát cách mô hình xử lý các chi tiết nhỏ như đường viền, bóng đổ. , Kiểm tra cách mô hình tô màu các ảnh đen trắng cũ, đảm bảo hình ảnh trông **tự nhiên và có ý nghĩa**.



- Vậy chúng ta sẽ định lượng như thế nào về số liệu này, nếu chúng ta tạo ra màu sắc thực tế hoàn hảo x, chúng ta sẽ đạt được gần 50% theo định nghĩa nếu chúng ta thêm màu từ một hình ảnh ngẫu nhiên, chúng ta sẽ nhận được 13 % với hồi quy L2, chúng ta sẽ nhận được 21,2 % với phân loại, chúng ta sẽ nhận được 24 và toàn bộ hệ thống của chúng ta nhân được 32 hệ thống do Larson đề xuất ở tất cả mọi nơi, đó là công việc đồng thời sử dụng khuôn khổ phân loại đa thức với một kiến trúc khác nhận được 27 và điều này gợi ý với chúng ta rằng các quyết định thiết kế của chúng ta khi phân loại và sử dụng một thuật ngữ cân bằng lại lớp đã cho chúng ta phép tô màu x hợp lý

- **Trục tung**: % ảnh mà con người (trên Amazon Mechanical Turk - AMT) đánh giá là ảnh thật (real).

- **Trục hoành**: Các phương pháp tô màu ảnh khác nhau.

- **Tổng số ảnh kiểm tra**: 1600 ảnh cho mỗi thuật toán.

- Ground Truth (Ảnh gốc, màu xanh dương nhạt - 50%) : Ảnh thật nên có điểm tối đa 50% (do người dùng chọn ngẫu nhiên giữa ảnh thật và ảnh tô màu).

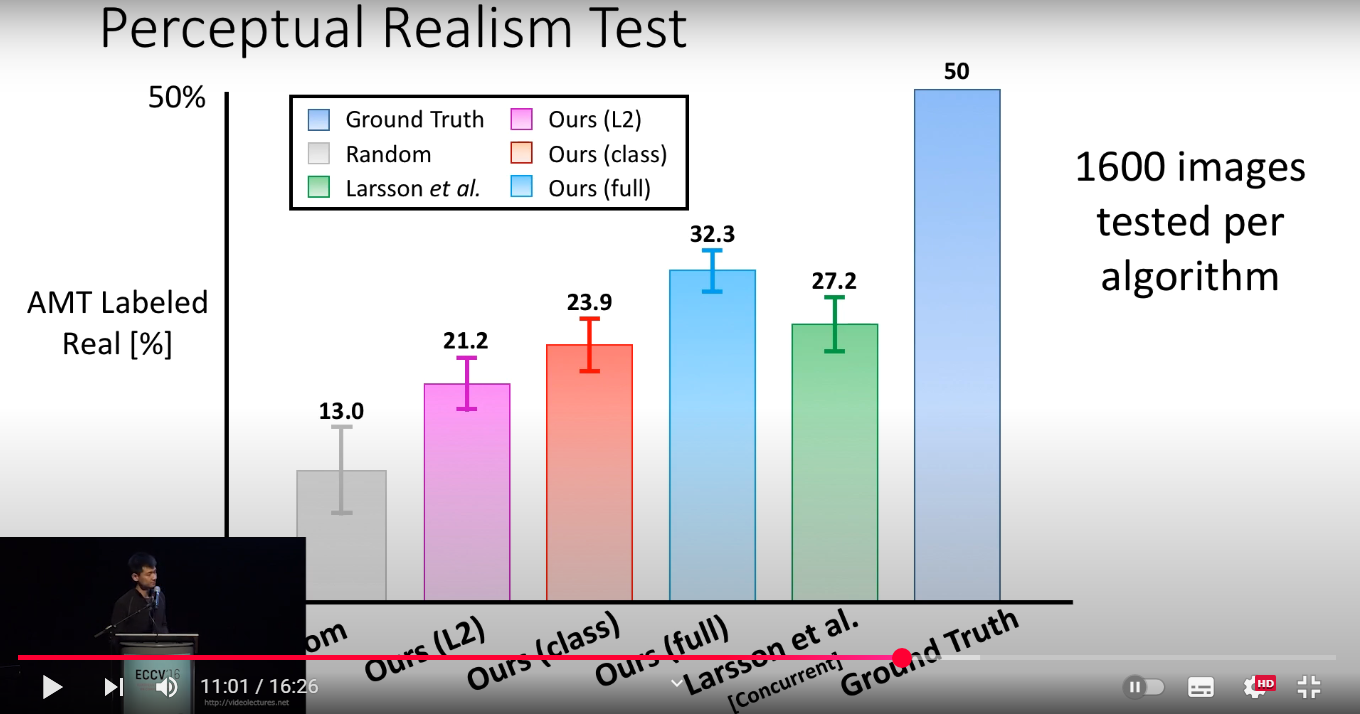
- Ours (full) - 32.3% Phương pháp đầy đủ của Zhang et al. cho kết quả cao nhất trong số các mô hình tự động

- Larsson et al. - 27.2% Một phương pháp tô màu khác, kém hơn "Ours (full)"

- Ours (class) - 23.9% Phiên bản mô hình sử dụng thông tin lớp đối tượng (class-based).

- Ours (L2) - 21.2% Phiên bản mô hình sử dụng hồi quy L2 (ít chân thực hơn so với class-based).

- Random - 13.0% Mô hình tô màu ngẫu nhiên, có kết quả thấp nhất.



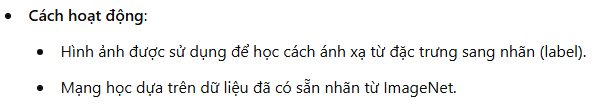
- Thông thường để có được biểu diễn nhận dạng các đối tượng, chúng ta có thể trực tiếp đào tạo cho các tác vụ bằng nhau sử dụng nhãn đối tượng, ví dụ như trong một hình ảnh , khuôn khổ phân loại đó và đây là trong một **thiết lập có giám sát**. Mặt khác, tô màu là một ví dụ không giám sát, đôi khi được gọi là học tự giám sát và trong mô hình này, chúng ta chia đầu vào thành hai phần, sau đó yêu cầu mạng dự đoán một phần được giữ lại từ phần còn lại

- Đây là sơ đồ minh họa 2 pp học máy: **học có giám sát (Supervised training)** và **học không giám sát/tự giám sát (Unsupervised/Self-supervised training)**, trong bối cảnh xử lý hình ảnh.

- Học có giám sát: Đầu vào là một hình ảnh từ tập dữ liệu ImageNet, chứa các phần x0 và x1.

**+ Mạng học đặc trưng (Learned feature hierarchy)**: Mô hình học cách trích xuất đặc trưng từ ảnh.

+ **Đầu ra**: Nhãn **y** của ảnh (ImageNet labels), tức là mô hình dự đoán loại đối tượng trong ảnh (ví dụ: chó, mèo, xe hơi).

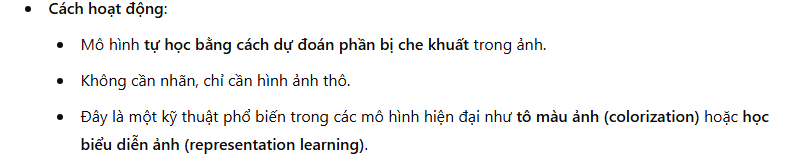
+ 

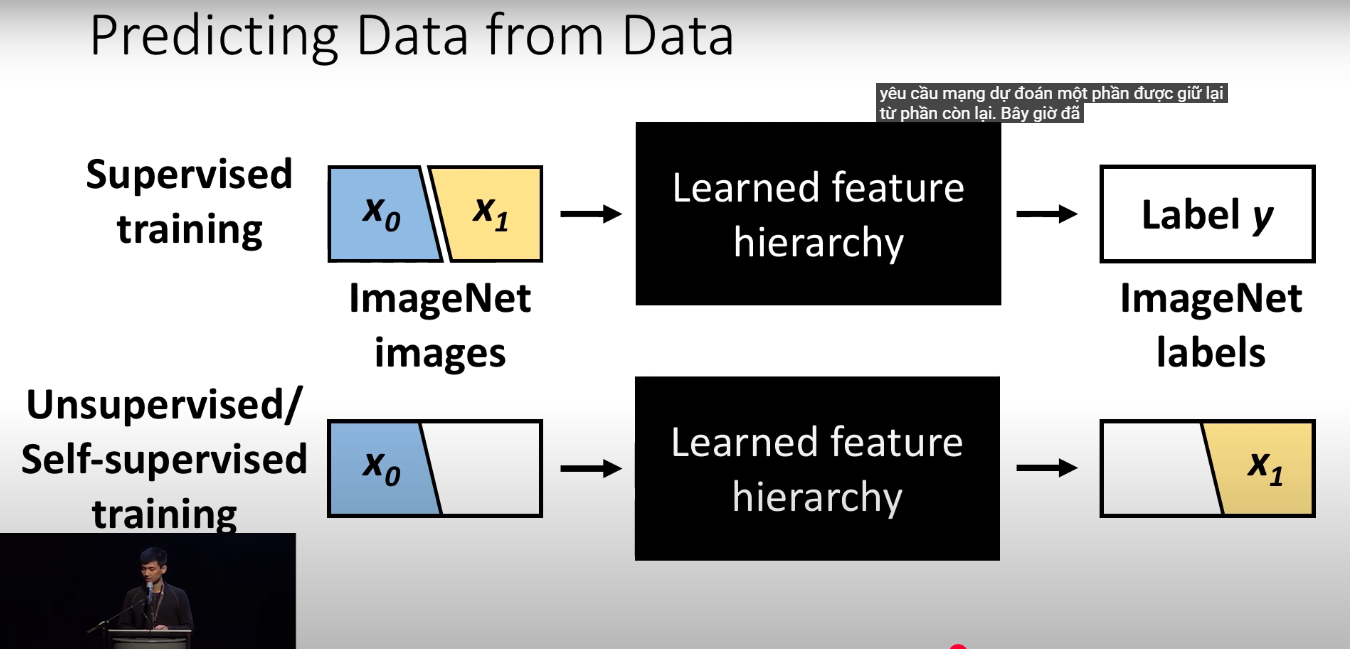
- Học không giám sát/tự giám sát

**+ Đầu vào**: Chỉ có một phần của ảnh **x0x\_0x0​**, không có nhãn.

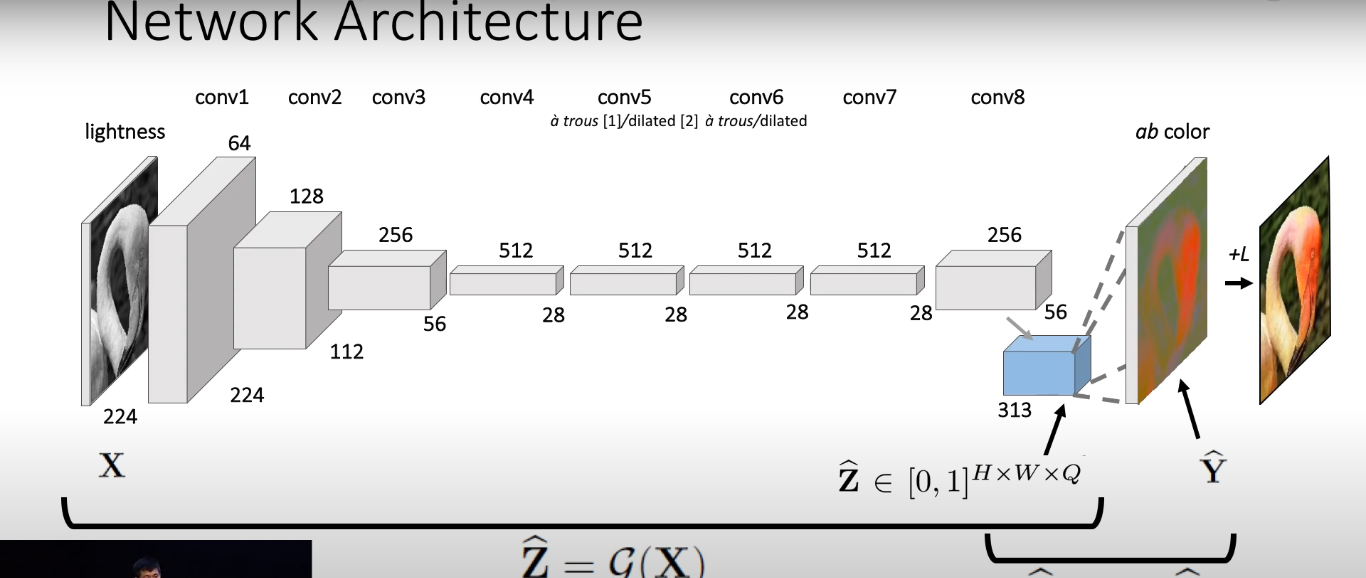
+ **Mạng học đặc trưng (Learned feature hierarchy)**: Mô hình vẫn học cách trích xuất đặc trưng nhưng không có nhãn hỗ trợ.

+ **Đầu ra**: Dự đoán phần còn lại của ảnh **x1​** (thay vì nhãn).

+ 



* Giải thích sơ đồ chuyển đổi



- Đầu vào là Ảnh đầu vào **X** là ảnh đen trắng (chỉ chứa **L - Lightness**) với kích thước **224×224**

- Mạng CNN trích xuất đặc trưng

+ **conv1**: 64 kênh đặc trưng → 224×224 => Giảm kích thước ảnh qua kỹ thật Maxpooling được dùng trong quá trình dowsampling

+ **conv2**: 128 kênh đặc trưng → 112×112 => Giảm kích thước ảnh qua kỹ thật Maxpooling

+ **conv3**: 256 kênh đặc trưng → 56×56=> Giảm kích thước ảnh qua kỹ thật Maxpooling

+ **conv4 - conv7**: Dùng kỹ thuật **dilated convolution (à trous)** giúp mở rộng trường quan sát mà không làm mất độ phân giải. => Giảm kích thước ảnh qua kỹ thật Maxpooling

+ **conv8**: Đưa số kênh về 256, giữ kích thước **56×56**

- Dự đoán màu sắc

+ Một khối **313 kênh** được tạo ra (biểu diễn trong khối màu xanh) để dự đoán giá trị màu sắc **ab**.

+ **313** là số lượng cụm màu (clusters) trong không gian **ab** (Lab color space), được lượng tử hóa.

- Kết hợp với kênh L để tạo ảnh màu

+ Ảnh đầu ra **Y** chính là ảnh màu, được tạo bằng cách **ghép kênh sáng L** với **màu ab** dự đoán.