1. **Tóm tắt:**

* Bài báo đưa ra phương pháp hợp lý để đưa ảnh trắng đen về ảnh có màu *hợp lý*.
* Các phương pháp trước cần tương tác với người nhiều hoặc màu không hợp lý.
* Nhiệm vụ phân loại và sử dụng cân bằng lại lớp trong quá trình huấn luyện.
* Sử dụng khái niệm truyền thẳng trong mạng CNN.
* Độ đo: bài kiểm tra Turing về tô màu (chọn giữa ảnh màu đã phục chế so với ảnh thực tế)
* **Key word:** Tô màu, Thị giác máy tính cho Đồ họa, CNN, Học tự giám sát.
* **Câu hỏi thắc mắc khi đọc:** Mạng CNN là gì?

1. **Giới thiệu:**

A collage of photos of different animals

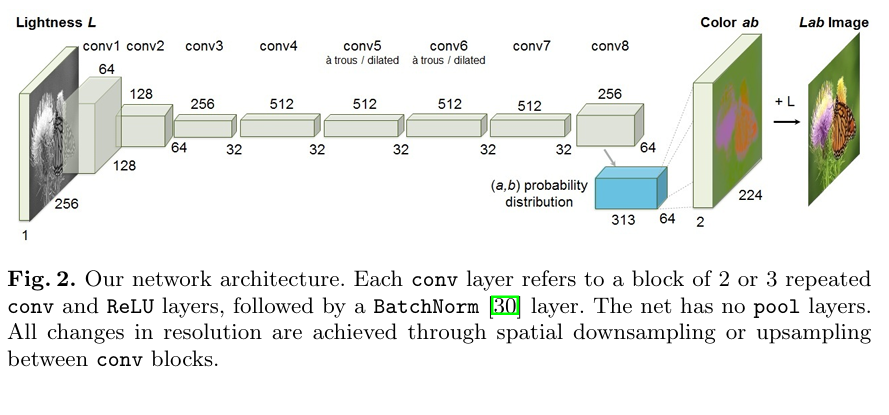
AI-generated content may be incorrect.

* Hình 1: bọ cánh cam thường có màu cam, cây cỏ thường màu xanh,… Những quả bóng có thể màu đỏ, tím hoặc những màu khác (không chắc chắn). Nhưng mục tiêu là: phụ thuộc thống kê ngữ nghĩa của ảnh màu đen trắng với phiên bản màu của chúng sao cho trực quan nhất chứ không phải phục chế thành màu ảnh gốc của ảnh.
* Trước đây người ta lấy dữ liệu trong CIE Lab colorspace [1][2] để khảo sát => màu nhạt nhòa vì kế thừa từ các bài toán hồi quy tiêu chuẩn, nơi mục tiêu là tối thiểu hóa sai số Euclid giữa ước lượng và giá trị thực tế.
* Rich.Zhang theo [3] sử dụng thuật toán riêng dự đoán phân phối màu sắc cho các điểm ảnh sao cho hợp lý  bằng cách lấy *trung bình ủ* của phân phối.
* Đánh giá theo độ đo: bài kiểm tra Turing về tô màu thì lừa được 32% người, cao hơn trước [2].
* Phương pháp có tên: *cross-channel encoder.*
* Đóng góp:
  + Tiến bộ trong vấn đề đồ họa của tô màu ảnh tự động bằng cách (a) thiết kế một hàm mục tiêu phù hợp xử lý sự không chắc chắn đa thức của bài toán tô màu và nắm bắt sự đa dạng rộng lớn của màu sắc, (b) giới thiệu một khung đánh giá mới cho các thuật toán tô màu, có thể áp dụng cho các nhiệm vụ tổng hợp ảnh khác, và (c) thiết lập một tiêu chuẩn mới cho nhiệm vụ bằng cách huấn luyện trên một triệu ảnh màu.
  + Giới thiệu nhiệm vụ tô màu như một phương pháp cạnh tranh và đơn giản để học biểu diễn tự giám sát, đạt kết quả hàng đầu trên một số tiêu chuẩn.

1. **Reference:**

* Các phương pháp không tham số, với một ảnh đen trắng đầu vào, đầu tiên xác định một hoặc nhiều ảnh màu tham chiếu (được cung cấp bởi người dùng hoặc tự động truy xuất) để sử dụng làm dữ liệu nguồn. Sau đó, theo khung Image Analogies [17], màu sắc được chuyển từ các vùng tương tự của ảnh tham chiếu sang ảnh đầu vào [18, 19, 20, 21].
* Bài toán như hồi quy vào không gian màu liên tục [22, 1, 2] hoặc phân loại các giá trị màu được lượng tử hóa [3]
* Bài toán tương tự: [23], [24], [25],  [5].

1. **Cách tiếp cận:**



* Ánh xạ từ ảnh đen trắng => phân phối giá trị màu đầu ra được lượng tử hóa bằng kiến trúc được hiển thị trong Hình 2 => Ảnh màu.

1. Hàm mục tiêu.
   1. A white background with black text

      AI-generated content may be incorrect.
   2. A white paper with black text and numbers

      AI-generated content may be incorrect.
2. Hàm cân bằng.

A white paper with black text and symbols

AI-generated content may be incorrect.

1. Xác xuất ước lượng điểm.

A text on a piece of paper

AI-generated content may be incorrect.

1. **Thí nghiệm.**
2. **Đánh giá chất lượng tô màu.**

A close-up of a collage of images

AI-generated content may be incorrect.

Hình 6 cho thấy ảnh thực tế và ảnh phục chế màu, 2 bên là số phần tram người chọn=> hơn 32% bị lừa (*p*<0.05 trong mỗi trường hợp).

**Khả năng diễn giải ngữ nghĩa (Phân loại VGG):** đưa các hình ảnh màu giả vào mạng VGG [5] được huấn luyện để dự đoán các lớp ImageNet từ ảnh màu thực.

**Độ chính xác thô(AuC):** Như một bài kiểm tra mức độ thấp, chúng tôi tính tỷ lệ phần trăm các màu điểm ảnh dự đoán nằm trong khoảng cách *L*2​ ngưỡng của giá trị thực tế trong không gian màu *ab*. Sau đó, chúng tôi quét qua các ngưỡng từ 0 đến 150 để tạo ra một hàm khối lượng tích lũy, như được giới thiệu trong [22], tích hợp diện tích dưới đường cong (AuC) và chuẩn hóa. Lưu ý rằng chỉ số AuC này đo lường độ chính xác dự đoán thô, trong khi phương pháp của chúng tôi nhắm đến tính hợp lý.

1. **Mã hóa đa kênh như học đặc trưng tự giám sát**

A diagram of a number of data

AI-generated content may be incorrect.

Đầu tiên là huấn luyện các bộ phân loại tuyến tính như hình 7, sau đó tinh chỉnh như bảng 2. Sử dụng mô hình AlexNet [38] để so sánh với các thí nghiệm trước thì thấy hiệu quả cáo hơn trong phân đoạn và phân loại.

1. Phân loại ImageNet

Từ dataset ImageNet (không có nhãn ngữ nghĩa). Các nhà khoa học cố định trọng số của mạng, thêm nhãn ngữ nghĩa và huấn luyện bộ phân loại tuyến tính trên từng lớp tích chập => AlexNet cho kết quả hiệu suất cao nhất.

1. PASCAL
   * **Phân loại PASCAL VOC 2007**: Mô hình đạt độ chính xác cao nhất khi cố định biểu diễn và tinh chỉnh phần còn lại, ngay cả khi chỉ dùng ảnh đen trắng.
   * **Phát hiện PASCAL VOC 2007**: Sử dụng Fast R-CNN, mô hình đạt 46.9% (đen trắng) và 47.9% (màu), cao hơn cơ sở k-means (45.6%) nhưng vẫn kém hơn mô hình huấn luyện trước với giám sát ImageNet (56.8%).
   * **Phân đoạn ngữ nghĩa PASCAL VOC 2012**: Với kiến trúc FCN, mô hình đen trắng đạt 35.0%, tương đương Donahue et al. [16], và tăng lên 35.6% khi thêm thông tin màu, vượt trội hơn các phương pháp khác. Nhiệm vụ tô màu của mô hình có sự tương đồng với phân đoạn ngữ nghĩa do cả hai đều liên quan đến phân loại theo điểm ảnh.
2. **Kết luận.**

Tô màu ảnh bằng CNN sâu không chỉ tạo ra kết quả gần giống ảnh màu thực mà còn học được biểu diễn hữu ích, hỗ trợ hiệu quả các nhiệm vụ phân loại, phát hiện và phân đoạn, cạnh tranh với các phương pháp tự giám sát khác.

1. **Referrence (có trong bài báo rồi)**
2. **Mạng có khai thác manh mối ở mức thấp không?**

A table with numbers and text

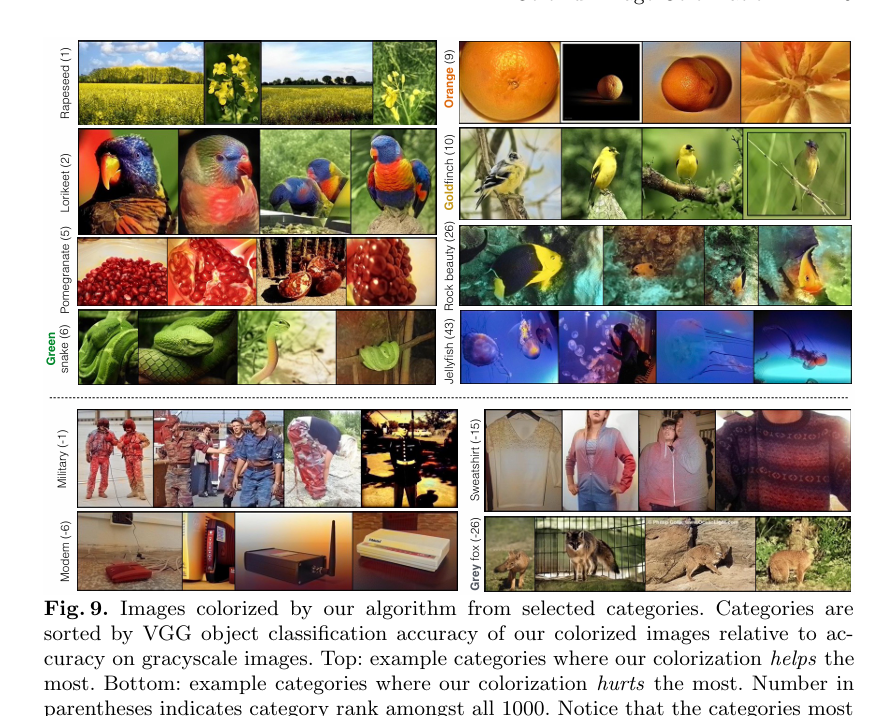
AI-generated content may be incorrect.

* Phương pháp này không chú đến mức thấp hay cao mà chú ý đến mức điểm ảnh và ngữ nghĩa.
* Nghiên cứu dựa trên độ sáng tối của điểm ảnh (ở ảnh trắng đen)
* Kết quả cho thấy mạng khó khôi phục màu sắc từ biểu đồ đen trắng chỉ dựa vào độ sáng, nhưng lại hiệu quả hơn với các đối tượng nhận diện được (như rau). Dự đoán ổn định với thay đổi độ sáng/thương phản, nhưng bị ảnh hưởng khi hình ảnh bị làm mờ do mất kết cấu.

1. Các mô hình khác như Noroozi et al. [43], Doersch et al. [14], và Donahue et al. [16] đã điều chỉnh kiến trúc (giảm stride, đổi thứ tự LRN/pooling, thay ReLU bằng LeakyReLU, bỏ nhóm tích chập) để cải thiện hiệu suất, nhưng mỗi mô hình có điểm mạnh ở các lớp khác nhau (conv1, conv2, conv3, conv4).
2. Mô hình của nhóm nghiên cứu (sử dụng đầu vào grayscale đơn kênh) cho thấy hiệu suất ổn định và nổi bật nhất ở lớp conv5.

Kết quả được tóm tắt trong Bảng 3, nhấn mạnh sự khác biệt trong thiết kế và hiệu quả của các phương pháp.

1. **Khả năng diễn giải ngữ nghĩa để tô màu.**

****

1. **Hiệu suất theo danh mục**

* Top lớp cải thiện: Phần lớn là đối tượng có tên chứa màu sắc (vd: *green snake*, *orange*, *goldfinch*).
* Top lớp giảm hiệu suất:
  + Đồ vật nhân tạo (vd: *modem*, *quần áo*) → màu không rõ ràng.
  + Một số động vật (vd: *gray fox*) → hệ thống dự đoán sai màu đặc trưng.

1. **Nhầm lẫn thường gặp.**

****

* *Minibus* → bị phân loại thành *school bus* (do tô màu vàng).
* *Standard schnauzer* → nhầm thành *Irish terrier* (màu lông thay đổi).

1. **Mạng có Khai thác Các đặc điểm Cấp thấp?**

A close-up of a color test

AI-generated content may be incorrect.

**Hình 12** kiểm tra giả thuyết: *"Mạng chỉ dựa vào mối quan hệ đơn giản giữa độ sáng (L) và màu sắc (ab)?"*

* **Thí nghiệm**:
  + Đầu vào là **bảng màu Macbeth** (các ô màu có độ sáng khác nhau) → mạng **không** tái tạo được màu chính xác.
  + Đầu vào là **rau củ có cùng độ sáng** (vd: zucchini, cà tím) → mạng dự đoán đúng màu nhờ nhận diện đặc trưng cao cấp.
* **Kết luận**: Mạng **không** chỉ phụ thuộc vào tín hiệu cấp thấp.

1. **Mạng có học được phân phối đa thức của màu sắc không?**

A screenshot of a computer game

AI-generated content may be incorrect.

**Hình 13** minh họa phân phối xác suất màu đầu ra:

* **Ví dụ**:
  + Chim (foreground) → phân phối đỉnh ở **đỏ/xanh lam**.
  + Nền (background) → phân phối ở **xanh lá/vàng/nâu**.
* **Annealed-mean** (T = 0.38) giúp cân bằng giữa **độ sống động** (*vibrancy*) và **tính nhất quán không gian** (*spatial coherence*).

1. **Kiến trúc mạng.**

A screenshot of a computer

AI-generated content may be incorrect.

**Bảng 4** chi tiết các lớp:

* **conv1 → conv5**: Tăng **effective dilation** (tích lũy stride × dilation) từ 1 → 16.
* **conv6 → conv8**: Giảm dilation về 8 → 4.
* **BatchNorm** được áp dụng sau mỗi khối convolution.

1. **So sánh các phương pháp khác.**
2. **So với LEARCH[22].**

**A graph on a white background

AI-generated content may be incorrect.**

* Dataset: 240 ảnh SUN (scenes).
* Kết quả (Bảng 5):
  + AuC: 90.1% (ours) > 88.8% (LEARCH).
  + AMT "real vs fake": 17.2% (ours) > 9.8% (LEARCH).

1. **So với Deep Colorization [1]**

* Ưu điểm của chúng tôi:
  + Pipeline đơn giản: Chỉ cần CNN, không cần DAISY features hay bilateral filter.
  + Đa dạng dữ liệu: Huấn luyện trên ImageNet (đối tượng + scenes) thay vì chỉ scenes.
  + Tốc độ: 0.022s/ảnh (GPU) so với 4.9s/ảnh (CPU) của [1].