

VIETNAM NATIONAL UNIVERSITY - HO CHI MINH

UNIVERSITY OF SCIENCE

---



**BÁO CÁO ĐỒ ÁN CUỐI KÌ**

## **Real Estate Image Retrieval**

Môn học: Truy vấn thông tin thị giác

Lớp: 21TGMT

Giảng viên: Phạm Minh Hoàng

Nguyễn Trọng Việt

Võ Hoài Việt

Sinh viên: Lê Nguyễn Phương Uyên - 21127476

Đoàn Việt Hưng - 21127289

Email: lnpuyen21@clc.fitus.edu.vn

dvhung21@clc.fitus.edu.vn

**Ho Chi Minh City - 2024**

# MỤC LỤC

<b>1) GIỚI THIỆU</b>	<b>2</b>
1.1 Tóm lược: . . . . .	2
1.2 Bối cảnh: . . . . .	2
1.3 Động lực: . . . . .	3
1.4 Thách thức: . . . . .	3
1.5 Ý nghĩa khoa học . . . . .	5
1.6 Ý nghĩa thực tế: . . . . .	6
<b>2) NHỮNG NGHIÊN CỨU LIÊN QUAN</b>	<b>7</b>
<b>3) TẬP DỮ LIỆU REI</b>	<b>14</b>
<b>4) PHƯƠNG PHÁP</b>	<b>15</b>
4.1 Mô đun phân loại SVM: . . . . .	15
4.2 Mô đun rút trích đặc trưng EfficientNet: . . . . .	16
4.2.1 Compound scaling: . . . . .	16
4.2.2 Cấu trúc EfficientNet: . . . . .	18
<b>5) THỰC NGHIỆM</b>	<b>19</b>
<b>6) KẾT LUẬN</b>	<b>23</b>

# 1) GIỚI THIỆU

## 1.1 Tóm lược:

- Hệ thống truy xuất hình ảnh bất động sản nhằm nâng cao hiệu quả và độ chính xác trong việc tìm kiếm hình ảnh liên quan từ các cơ sở dữ liệu lớn, rất quan trọng đối với người mua và người bán.
- Báo cáo này nghiên cứu các kỹ thuật tiên tiến trong truy xuất hình ảnh, tập trung vào ứng dụng thuật toán học máy và công nghệ thị giác máy tính. Bằng cách sử dụng các phương pháp trích xuất đặc trưng và so khớp tương đồng, nghiên cứu khám phá các cách tiếp cận để phân loại và truy xuất hình ảnh.
- Hiệu suất của các phương pháp này được đánh giá qua tập dữ liệu hình ảnh bất động sản, xem xét độ chính xác, tốc độ và khả năng chống chịu trước các biến thể về chất lượng và nội dung hình ảnh. Kết quả cho thấy tích hợp kỹ thuật học máy tiên tiến cải thiện hiệu suất truy xuất so với các phương pháp truyền thống.
- Báo cáo cũng thảo luận về tác động thực tế của việc triển khai hệ thống này, bao gồm thách thức về quyền riêng tư dữ liệu và tài nguyên tính toán. Công việc tương lai sẽ tập trung vào tinh chỉnh độ chính xác của mô hình và khám phá các cách tiếp cận kết hợp dữ liệu văn bản và hình ảnh để truy xuất toàn diện hơn.

## 1.2 Bối cảnh:

- Sự phát triển nhanh chóng của công nghệ kỹ thuật số đã biến đổi ngành bất động sản, dẫn đến sự phụ thuộc ngày càng nhiều vào các nền tảng trực tuyến để tìm kiếm và tiếp thị. Trong bối cảnh này, khả năng truy xuất hiệu quả và chính xác các hình ảnh bất động sản liên quan đến các cơ sở dữ liệu rộng lớn là điều quan trọng.
- Các phương pháp truy xuất hình ảnh truyền thống, thường dựa vào gán thẻ thủ công và tìm kiếm theo từ khóa, đang trở nên không đủ do khối lượng và độ phức tạp của dữ liệu có sẵn. Báo cáo này khám phá tiềm năng của các thuật toán học máy tiên tiến và kỹ thuật thị giác máy tính để giải quyết những thách thức này.
- Bằng cách tự động hóa quá trình phân loại và truy xuất hình ảnh, các công nghệ này hứa hẹn cải thiện trải nghiệm người dùng, đơn giản hóa các giao dịch bất động sản và cung cấp kết quả tìm kiếm chính xác hơn. Nghiên cứu này đánh giá các cách tiếp cận hiện đại khác nhau đối với truy xuất hình ảnh, tập trung vào hiệu suất, tính thực tiễn và các tác động đối với sự phát triển trong tương lai của ngành bất động sản.

### 1.3 Động lực:

Bài báo "**Real Estate Image Retrieval**" giải quyết những bất cập trong việc chú thích hình ảnh bất động sản một cách thủ công. Dưới đây là những động lực chính:

- **Cải thiện hiệu quả thị trường bất động sản:** Việc phân loại chính xác và tự động hóa các hình ảnh bất động sản có thể làm tăng đáng kể quá trình liệt kê các tài sản trực tuyến. Điều này đảm bảo rằng người mua hoặc người thuê tiềm năng có thể nhanh chóng tìm thấy các bất động sản đáp ứng tiêu chí của họ, nâng cao trải nghiệm người dùng và hiệu quả của các nền tảng bất động sản.
- **Nâng cao chất lượng hình ảnh:** Hình ảnh bất động sản thường gặp phải các vấn đề về chất lượng như ánh sáng kém, bóng đổ, và mờ. Bằng cách áp dụng các kỹ thuật nâng cao hình ảnh như CLAHE (Contrast Limited Adaptive Histogram Equalization), bài báo nhằm cải thiện chất lượng hình ảnh, làm cho chúng trở nên hấp dẫn và thông tin hơn đối với người dùng.
- **Sử dụng Deep Learning để phân loại chính xác:** Bài báo giới thiệu một mô hình nhận dạng mới sử dụng mạng LSTM để học các mối tương quan theo cả chiều ngang và chiều dọc. Mô hình này được thiết kế để dự đoán chính xác các nhãn cho nhiều loại hình ảnh bất động sản khác nhau, một nhiệm vụ quan trọng do sự đa dạng lớn trong cùng một loại hình ảnh bất động sản.
- **Đóng góp dữ liệu mới:** Nghiên cứu giới thiệu một bộ dữ liệu mới gọi là Real Estate Image (REI), bao gồm một số lượng lớn hình ảnh được phân loại để đánh giá các thuật toán phân loại tiên tiến. Bộ dữ liệu này có giá trị đối với các nghiên cứu tiếp theo trong lĩnh vực phân loại hình ảnh bất động sản và có thể giúp phát triển các mô hình phân loại mạnh mẽ và chính xác hơn.
- **Ứng dụng thực tiễn:** Khả năng phân loại hình ảnh của các đặc điểm cụ thể như mặt bàn bếp và sàn nhà có thể hỗ trợ trong việc đánh giá tự động giá trị và chất lượng của bất động sản. Điều này có ý nghĩa thực tiễn đối với các đại lý bất động sản, người thẩm định và người mua, những người dựa vào thông tin hình ảnh chính xác để đưa ra quyết định thông minh.

Tóm lại, động lực của nghiên cứu này xuất phát từ nhu cầu nâng cao hiệu quả và độ chính xác của việc phân loại hình ảnh bất động sản, cải thiện chất lượng hình ảnh, tận dụng các kỹ thuật deep learning tiên tiến, đóng góp các bộ dữ liệu có giá trị, và giải quyết các ứng dụng thực tiễn trong thị trường bất động sản.

### 1.4 Thách thức:

#### a) Thu thập và chuẩn bị dữ liệu:

- **Đa dạng và chất lượng hình ảnh:** Hình ảnh bất động sản có thể rất đa dạng về chất lượng, góc chụp, ánh sáng, và nội dung. Việc thu thập và chuẩn bị dữ liệu đủ đa dạng và chất lượng cao là một thách thức lớn.

- **Gắn nhãn dữ liệu:** Việc gắn nhãn thủ công cho hình ảnh để tạo bộ dữ liệu huấn luyện cho các mô hình phân loại yêu cầu nhiều công sức và thời gian. Hơn nữa, sự chính xác trong việc gắn nhãn là rất quan trọng để đảm bảo hiệu quả của mô hình.

#### b) Xử lý và phân tích dữ liệu:

- **Phân loại chính xác:** Hình ảnh bất động sản có thể chứa nhiều đối tượng và chi tiết khác nhau (phòng khách, phòng ngủ, nhà bếp, nội thất, ngoại thất). Việc phân loại chính xác các đối tượng này đòi hỏi các mô hình AI phải rất tinh vi và được huấn luyện tốt.
- **Xử lý hình ảnh phức tạp:** Một số hình ảnh có thể bị nhiễu, không rõ ràng, hoặc có chất lượng kém, khiến việc phân loại trở nên khó khăn hơn.

#### c) Phát triển mô hình học máy:

- **Chọn mô hình phù hợp:** Có nhiều phương pháp học máy và mạng nơ-ron khác nhau (như CNN, RNN, Transformer). Việc chọn mô hình phù hợp và tối ưu hóa các tham số của nó là một thách thức kỹ thuật lớn.
- **Overfitting và Underfitting:** Đảm bảo mô hình không bị overfitting (quá khớp với dữ liệu huấn luyện) hoặc underfitting (không đủ khả năng mô hình hóa dữ liệu) đòi hỏi kỹ năng và kinh nghiệm trong việc điều chỉnh mô hình.

#### d) Tính toán và tài nguyên:

- **Tài nguyên tính toán:** Huấn luyện các mô hình deep learning yêu cầu tài nguyên tính toán lớn, bao gồm GPU mạnh mẽ và thời gian tính toán đáng kể. Điều này có thể đòi hỏi chi phí cao và cơ sở hạ tầng mạnh mẽ.
- **Xử lý dữ liệu lớn:** Lượng dữ liệu hình ảnh có thể rất lớn, đòi hỏi các hệ thống lưu trữ và xử lý dữ liệu hiệu quả.

#### e) Tính ứng dụng và triển khai:

- **Triển khai vào hệ thống thực tế:** Đưa mô hình phân loại hình ảnh vào các hệ thống thực tế đòi hỏi sự tích hợp chặt chẽ với các hệ thống hiện có và đảm bảo mô hình hoạt động ổn định, hiệu quả trong môi trường thực tế.
- **Bảo mật và quyền riêng tư:** Bảo vệ dữ liệu hình ảnh và đảm bảo quyền riêng tư của người dùng là một thách thức quan trọng, đặc biệt khi dữ liệu có thể chứa thông tin nhạy cảm.

#### f) Đánh giá và cải tiến liên tục:

- **Đánh giá hiệu quả mô hình:** Việc đánh giá hiệu quả của mô hình phân loại đòi hỏi các phương pháp đánh giá chính xác và đáng tin cậy. Ngoài ra, mô hình cần được cải tiến liên tục dựa trên phản hồi và dữ liệu mới.

- **Thích ứng với thay đổi:** Thị trường bất động sản và yêu cầu của người dùng có thể thay đổi theo thời gian. Mô hình phân loại cần phải linh hoạt và có khả năng thích ứng với những thay đổi này.

## 1.5 Ý nghĩa khoa học

### a) Công nghệ và trí tuệ nhân tạo:

- **Machine Learning và Deep Learning:** Việc phân loại hình ảnh bất động sản sử dụng các thuật toán học máy và mạng nơ-ron sâu, thúc đẩy sự phát triển của các mô hình AI tiên tiến và ứng dụng của chúng trong nhận dạng hình ảnh.
- **Computer Vision:** Chủ đề này giúp cải thiện các kỹ thuật nhận dạng và phân tích hình ảnh, như phân đoạn ảnh, phát hiện đối tượng, và trích xuất đặc trưng từ ảnh.

### b) Ứng dụng thực tiễn trong bất động sản:

- **Tối ưu hóa quá trình mua bán và cho thuê:** Việc phân loại hình ảnh tự động giúp các công ty bất động sản tổ chức và quản lý danh mục hình ảnh của họ một cách hiệu quả, cải thiện trải nghiệm người dùng khi tìm kiếm nhà đất.
- **Định giá bất động sản:** Hệ thống phân loại có thể hỗ trợ việc đánh giá giá trị bất động sản bằng cách phân tích các yếu tố trong hình ảnh như kích thước, tình trạng và kiểu dáng của tài sản.

### c) Dữ liệu lớn và phân tích:

- **Quản lý dữ liệu lớn:** Chủ đề này liên quan đến việc xử lý và phân tích lượng lớn hình ảnh bất động sản, cung cấp dữ liệu giá trị cho các phân tích thị trường, xu hướng bất động sản và ra quyết định chiến lược.
- **Phân tích thị trường:** Thông qua phân loại hình ảnh, các công ty có thể hiểu rõ hơn về xu hướng kiến trúc, sở thích của khách hàng, và các yếu tố ảnh hưởng đến giá trị bất động sản trên thị trường.

### d) Kinh tế và xã hội:

- **Tiết kiệm thời gian và chi phí:** Tự động hóa quá trình phân loại hình ảnh giúp giảm bớt nhu cầu về lao động thủ công, tiết kiệm thời gian và chi phí cho các doanh nghiệp bất động sản.
- **Phát triển đô thị và quy hoạch:** Việc phân loại hình ảnh bất động sản cung cấp dữ liệu hỗ trợ cho các dự án quy hoạch đô thị và phát triển cơ sở hạ tầng, góp phần vào việc xây dựng các cộng đồng bền vững và thông minh hơn.

## 1.6 Ý nghĩa thực tế:

### a) Nâng cao hiệu quả quản lý và tiếp thị:

- **Tự động hóa phân loại và tổ chức hình ảnh:** Giúp các công ty bất động sản quản lý và sắp xếp một lượng lớn hình ảnh bất động sản một cách tự động, giảm thiểu công việc thủ công và tăng cường độ chính xác.
- **Tối ưu hóa trải nghiệm người dùng:** Người mua và người thuê nhà có thể tìm kiếm và lọc các bất động sản một cách dễ dàng hơn dựa trên hình ảnh phân loại, như tìm nhà theo loại (căn hộ, biệt thự), theo phong cách thiết kế (hiện đại, cổ điển), hoặc theo tình trạng (mới, đã qua sử dụng).

### b) Cải thiện quy trình đánh giá và định giá bất động sản:

- **Phân tích hình ảnh để định giá:** Hệ thống phân loại hình ảnh có thể cung cấp thông tin chi tiết về tình trạng và đặc điểm của bất động sản, hỗ trợ quá trình định giá một cách chính xác và nhanh chóng hơn.
- **So sánh và đánh giá:** Dữ liệu hình ảnh phân loại giúp dễ dàng so sánh các tài sản tương tự, từ đó đưa ra quyết định mua bán hoặc đầu tư chính xác hơn.

### c) Tăng cường hiệu quả marketing và quảng cáo:

- **Tạo ra các quảng cáo chính xác và hấp dẫn hơn:** Việc phân loại hình ảnh giúp tạo ra các chiến dịch quảng cáo nhắm mục tiêu hiệu quả hơn, với hình ảnh phù hợp với sở thích và nhu cầu của khách hàng.
- **Cải thiện nội dung trực quan:** Hình ảnh được phân loại và sắp xếp giúp nâng cao chất lượng và sự hấp dẫn của nội dung trực quan trong các chiến dịch tiếp thị, thu hút khách hàng tiềm năng tốt hơn.

### d) Hỗ trợ quản lý tài sản và phát triển bất động sản:

- **Quản lý tài sản hiệu quả hơn:** Việc phân loại hình ảnh giúp các nhà quản lý tài sản theo dõi và quản lý tình trạng của các tài sản trong danh mục của họ một cách dễ dàng.
- **Phát triển bất động sản và quy hoạch đô thị:** Dữ liệu hình ảnh phân loại cung cấp thông tin quan trọng cho việc lập kế hoạch và phát triển các dự án bất động sản mới, đảm bảo chúng phù hợp với xu hướng thị trường và nhu cầu của người dân.

### e) Tăng cường độ chính xác và nhất quán trong kiểm tra và giám sát:

- **Kiểm tra tình trạng tài sản:** Hệ thống phân loại hình ảnh có thể được sử dụng để kiểm tra và giám sát tình trạng của bất động sản, phát hiện sớm các vấn đề cần sửa chữa hoặc bảo trì.

- **Nhất quán trong báo cáo:** Việc phân loại hình ảnh tự động đảm bảo rằng các báo cáo và dữ liệu liên quan đến bất động sản luôn nhất quán và chính xác, giảm thiểu sai sót do con người gây ra.

## 2) NHỮNG NGHIÊN CỨU LIÊN QUAN

Việc truy xuất hình ảnh bất động sản đã thu hút sự chú ý đáng kể trong những năm gần đây nhờ sự tiến bộ nhanh chóng của công nghệ thị giác máy tính và học máy. Bên dưới đây là một số nghiên cứu nổi bật đã đóng góp vào sự phát triển và hiểu biết về lĩnh vực này:

### a) Real Estate Image Analysis: A Literature Review[1]:

Tổng quan tài liệu của David Koch và những người đồng nghiên cứu (2020) cung cấp cái nhìn toàn diện về các kỹ thuật và phương pháp được sử dụng trong phân tích hình ảnh bất động sản. Bài báo phân loại các phương pháp hiện có thành các nhóm khác nhau dựa trên cách tiếp cận của chúng, chẳng hạn như học sâu, phân đoạn hình ảnh và trích xuất đặc trưng. Nó nhấn mạnh tầm quan trọng của các tập dữ liệu hình ảnh chất lượng cao và những thách thức liên quan đến sự biến đổi hình ảnh và che khuất trong nhiếp ảnh bất động sản. Các phương pháp chính được thảo luận bao gồm:

- Phân tích hình ảnh vệ tinh và hình ảnh từ trên cao:
- Phân loại và phân đoạn vùng đất: Nhận dạng các loại hình đất (đồng cỏ, rừng, đường, tòa nhà, v.v.).
- Nhận dạng và phân đoạn các tòa nhà: Xác định các loại tòa nhà (nhà ở, thương mại, nhà một gia đình, căn hộ, v.v.).
- Trích xuất thông số tòa nhà: Ví dụ, tính toán khả năng lắp đặt năng lượng mặt trời.

#### **Hình ảnh ngoài trời và từ Street View:**

- Truy xuất hình ảnh tòa nhà: Sử dụng các đặc điểm màu sắc và tính năng địa phương để nhận dạng và truy xuất hình ảnh tòa nhà.
- Nhận dạng yếu tố kiến trúc: Phân tích phong cách kiến trúc và các yếu tố ảnh hưởng đến chi phí xây dựng.

#### **Hình ảnh trong nhà và sơ đồ mặt bằng:**

- Nhận dạng và phân tích không gian nội thất: Sử dụng hình ảnh trong nhà để phân tích không gian nội thất và nhận dạng các yếu tố như đồ nội thất và vật liệu.

#### **Ưu điểm:**

- **Tăng cường khả năng tự động hóa:** Giúp tự động hóa các tác vụ như xác định nhu cầu năng



lượng sưởi ấm của một tòa nhà hoặc nhận dạng các yếu tố ảnh hưởng đến giá trị bất động sản, từ đó tiết kiệm thời gian và chi phí.

- **Trích xuất thông tin ẩn:** Giúp biến thông tin ẩn trong hình ảnh thành dữ liệu có thể sử dụng được, ví dụ như thông tin về đặc điểm tòa nhà hoặc khu vực lân cận.

- **Khả năng xử lý dữ liệu lớn:** Phương pháp thị giác máy tính và học sâu có khả năng xử lý một lượng lớn dữ liệu hình ảnh từ các nguồn khác nhau, bao gồm cả hình ảnh vệ tinh và hình ảnh từ Street View.

#### Nhược điểm:

- **Độ chính xác và tin cậy:** Một số phương pháp phân tích hình ảnh hiện tại có thể chưa đạt độ chính xác cao bằng các phương pháp truyền thống. Ví dụ, xác định nhu cầu năng lượng sưởi ấm của tòa nhà có thể chính xác hơn khi sử dụng các phương pháp truyền thống.

- **Thách thức về dữ liệu:** Các dữ liệu hình ảnh có thể không đầy đủ, không chính xác hoặc bị sai lệch, dẫn đến kết quả phân tích không đáng tin cậy.

- **Yêu cầu về chuyên môn và tài nguyên:** Việc tạo ra và duy trì các mô hình phân tích hình ảnh đòi hỏi chi phí và nguồn lực lớn, cũng như sự chuyên môn trong lĩnh vực này.

### **b) Real Estate Monitoring System Based on Remote Sensing and Image Recognition Technologies[2]:**

Sergejs Kodors và những người đồng nghiên cứu (2016) đã đề xuất một hệ thống giám sát bất động sản dựa trên công nghệ viễn thám và nhận diện hình ảnh. Hệ thống này tích hợp hình ảnh vệ tinh và ảnh chụp từ mặt đất để giám sát và phân tích các bất động sản, sử dụng các thuật toán nhận diện hình ảnh để phát hiện các thay đổi trong điều kiện bất động sản theo thời gian. Nghiên cứu này minh họa tiềm năng của việc kết hợp các nguồn dữ liệu khác nhau để phân tích bất động sản toàn diện. Bài báo mô tả một hệ thống giám sát bất động sản sử dụng công nghệ cảm biến từ xa và nhận diện hình ảnh, bao gồm các bước chính sau:

- **Thu Thập Dữ Liệu:** Hệ thống sử dụng cảm biến từ xa (như vệ tinh hoặc máy bay không người lái) để thu thập hình ảnh và dữ liệu về bất động sản. Những dữ liệu này bao gồm các yếu tố như hình dạng, kích thước và tình trạng của bất động sản.
- **Xử Lý Hình Ảnh:** Dữ liệu hình ảnh được xử lý bằng các thuật toán nhận diện hình ảnh. Các công nghệ như học máy và mạng nơ-ron tích chập (CNN) được sử dụng để phân tích và nhận diện các đặc điểm quan trọng của bất động sản, như tình trạng xây dựng, sự thay đổi trong kết cấu, và sự hiện diện của các yếu tố khác.
- **Tổng Hợp Dữ Liệu và Phân Tích:** Các dữ liệu thu được từ cảm biến và hình ảnh được tổng hợp và phân tích để cung cấp thông tin chi tiết về bất động sản. Điều này bao gồm việc phát hiện các thay đổi, đánh giá tình trạng và xác định giá trị của bất động sản.

- **Trình Bày Kết Quả:** Hệ thống cung cấp các báo cáo và hình ảnh được phân tích cho người dùng. Những thông tin này có thể được sử dụng để quản lý bất động sản, lập kế hoạch phát triển hoặc giám sát các vấn đề liên quan.

#### **Ưu điểm:**

- **Theo Dõi Liên Tục:** Công nghệ cảm biến từ xa cho phép theo dõi bất động sản liên tục và trong thời gian thực, giúp phát hiện sớm các thay đổi và vấn đề.
- **Tăng Độ Chính Xác:** Việc sử dụng nhận diện hình ảnh giúp cải thiện độ chính xác trong việc phân tích và đánh giá bất động sản so với các phương pháp truyền thống.
- **Giảm Thiểu Chi Phí:** So với các phương pháp giám sát truyền thống, việc sử dụng cảm biến từ xa và công nghệ nhận diện hình ảnh có thể giảm thiểu chi phí nhân công và thời gian thu thập dữ liệu.
- **Khả Năng Xử Lý Lớn:** Hệ thống có khả năng xử lý lượng dữ liệu lớn từ nhiều nguồn khác nhau, cung cấp cái nhìn tổng quan và chi tiết về bất động sản.

#### **Nhược điểm:**

- **Chi Phí Đầu Tư Cao:** Việc triển khai các công nghệ cảm biến từ xa và nhận diện hình ảnh có thể đòi hỏi đầu tư ban đầu lớn về thiết bị và phần mềm.
- **Độ Chính Xác Phụ Thuộc Vào Chất Lượng Hình Ảnh:** Độ chính xác của phân tích phụ thuộc vào chất lượng hình ảnh và dữ liệu thu thập được. Hình ảnh không rõ nét hoặc bị nhiễu có thể ảnh hưởng đến kết quả.
- **Yêu Cầu Kỹ Thuật Cao:** Việc triển khai và vận hành hệ thống yêu cầu các kỹ thuật và kiến thức chuyên môn cao, từ việc cài đặt thiết bị cảm biến đến xử lý và phân tích dữ liệu.
- **Khả Năng Đối Phó Với Thay Đổi Môi Trường:** Các yếu tố môi trường như thời tiết hoặc ánh sáng có thể ảnh hưởng đến chất lượng dữ liệu thu thập và độ chính xác của phân tích.

#### **c) Automatic Real-Estate Image Analysis for Retrieval and Classification[3]:**

Corneliu, Andrei, Laura và Bogdan (2022) đã tập trung vào truy xuất và phân loại tự động ảnh bất động sản bằng cách sử dụng mạng nơ-ron tích chập (CNN). Phương pháp của họ tự động phân loại hình ảnh dựa trên các kiểu kiến trúc và loại hình bất động sản, chứng minh hiệu quả của CNN trong việc xử lý các bộ dữ liệu hình ảnh quy mô lớn và đạt được độ chính xác cao trong các nhiệm vụ phân loại hình ảnh. Bài báo trình bày một hệ thống phân tích hình ảnh tự động dành cho bất động sản, với các bước chính bao gồm:

- **Trích Xuất Đặc Trưng:** Hệ thống sử dụng các kỹ thuật học máy và nhận diện hình ảnh để trích xuất các đặc trưng từ hình ảnh bất động sản. Các đặc trưng này có thể bao gồm màu sắc, hình dạng, kết cấu, và các yếu tố kiến trúc khác.

- **Phân Loại Hình Ảnh:** Các đặc trưng trích xuất được sử dụng để phân loại hình ảnh vào các nhóm hoặc danh mục khác nhau. Hệ thống có thể phân loại hình ảnh dựa trên loại bất động sản (như nhà ở, văn phòng, căn hộ) hoặc các thuộc tính khác (như tình trạng, kích thước).
- **Thu Hồi Hình Ảnh:** Hệ thống hỗ trợ tìm kiếm hình ảnh bằng cách so sánh các đặc trưng của hình ảnh truy vấn với các hình ảnh trong cơ sở dữ liệu. Kết quả là các hình ảnh tương tự hoặc phù hợp nhất với truy vấn sẽ được trả về.
- **Đánh Giá và Cải Tiến:** Hệ thống đánh giá hiệu suất của các mô hình phân loại và thu hồi, và sử dụng thông tin phản hồi để cải tiến các thuật toán và mô hình.

#### Ưu điểm:

- **Tự Động Hóa Quy Trình:** Việc tự động hóa quá trình phân tích và phân loại hình ảnh giúp giảm thiểu sự can thiệp của con người, tăng tính hiệu quả và độ chính xác trong việc xử lý dữ liệu.
- **Khả Năng Phân Tích Chi Tiết:** Hệ thống có khả năng phân tích và phân loại hình ảnh dựa trên nhiều đặc trưng khác nhau, cung cấp thông tin chi tiết về bất động sản.
- **Hỗ Trợ Tìm Kiếm Hiệu Quả:** Với khả năng thu hồi hình ảnh, hệ thống cho phép người dùng dễ dàng tìm kiếm và xác định các hình ảnh bất động sản phù hợp với yêu cầu của họ.
- **Cải Tiến Liên Tục:** Hệ thống có khả năng học và cải tiến liên tục dựa trên phản hồi và dữ liệu mới, nâng cao chất lượng phân tích và phân loại theo thời gian.

#### Nhược điểm:

- **Yêu Cầu Về Dữ Liệu Đầu Vào:** Độ chính xác của hệ thống phụ thuộc vào chất lượng và số lượng dữ liệu hình ảnh đầu vào. Hình ảnh không rõ nét hoặc thiếu thông tin có thể làm giảm hiệu quả của phân tích.
- **Chi Phí Đầu Tư Cao:** Việc triển khai hệ thống phân tích hình ảnh tự động có thể yêu cầu đầu tư lớn vào phần mềm và phần cứng, cũng như cần đội ngũ kỹ thuật để thiết lập và bảo trì hệ thống.
- **Phụ Thuộc Vào Các Thuật Toán:** Các thuật toán phân loại và thu hồi có thể gặp khó khăn trong việc xử lý các đặc trưng phức tạp hoặc không đồng nhất, ảnh hưởng đến độ chính xác của kết quả.
- **Khả Năng Xử Lý Hạn Chế:** Một số đặc trưng tinh vi hoặc khó phân loại có thể không được xử lý hiệu quả bởi các mô hình hiện có, dẫn đến hạn chế trong việc phân tích các loại hình bất động sản phức tạp.

#### **d) High-Level Feature Aggregation for Fine-Grained Architectural Floor Plan Retrieval[4]:**

Divya Sharma và Chiranjay Chattopadhyay (2018) đã nghiên cứu việc truy xuất bản vẽ sàn kiến trúc chi tiết thông qua việc tổng hợp đặc trưng cấp cao. Nghiên cứu của họ giải quyết thách thức của việc truy xuất các bản vẽ kiến trúc chi tiết từ các cơ sở dữ liệu lớn, nhấn mạnh tầm quan

trọng của việc nắm bắt các chi tiết thiết kế tinh vi. Họ đã đề xuất một phương pháp tổng hợp đặc trưng mới cải thiện đáng kể hiệu suất truy xuất bằng cách tập trung vào các đặc trưng chi tiết cụ thể của thiết kế kiến trúc. Bài báo đề xuất một phương pháp thu hồi kế hoạch mặt bằng chi tiết bằng cách sử dụng việc tổng hợp các đặc trưng cấp cao. Phương pháp này bao gồm các bước chính sau:

- **Trích Xuất Đặc Trưng:** Các đặc trưng của kế hoạch mặt bằng được trích xuất từ hình ảnh bằng cách sử dụng các mạng nơ-ron tích chập (CNN). Các đặc trưng này thường là các yếu tố cấp thấp như hình dạng, đường nét, và kết cấu.
- **Tổng Hợp Đặc Trưng Cấp Cao:** Các đặc trưng cấp thấp được tổng hợp thành các đặc trưng cấp cao hơn, cho phép hệ thống nhận diện và phân loại các đặc điểm kiến trúc phức tạp hơn.
- **So Sánh và Phân Tích:** Các đặc trưng cấp cao được so sánh với các đặc trưng của các kế hoạch mặt bằng trong cơ sở dữ liệu để tìm kiếm các kế hoạch phù hợp nhất với yêu cầu truy vấn.

#### **Ưu điểm:**

- **Tăng Độ Chính Xác:** Việc tổng hợp các đặc trưng cấp cao giúp cải thiện khả năng nhận diện và phân loại các kế hoạch mặt bằng chi tiết hơn, từ đó tăng độ chính xác của quá trình thu hồi.
- **Khả Năng Xử Lý Đa Dạng:** Phương pháp này có thể xử lý các kế hoạch mặt bằng có sự biến đổi lớn trong kiểu dáng và chi tiết, vì nó không chỉ dựa vào các đặc trưng đơn giản mà còn tổng hợp thông tin từ nhiều nguồn.
- **Khả Năng Tinh Chính:** Các đặc trưng cấp cao cho phép hệ thống phân tích các yếu tố tinh vi hơn, giúp nhận diện các đặc điểm kiến trúc phức tạp và chi tiết hơn.

#### **Nhược điểm:**

- **Tính Tính Toán Cao:** Phương pháp này yêu cầu một lượng tính toán lớn để trích xuất và tổng hợp các đặc trưng cấp cao. Điều này có thể dẫn đến thời gian xử lý lâu hơn và yêu cầu tài nguyên máy tính cao.
- **Phụ Thuộc vào Chất Lượng Dữ Liệu Đầu Vào:** Độ chính xác của phương pháp phụ thuộc vào chất lượng của các hình ảnh đầu vào và khả năng của mô hình mạng nơ-ron trong việc trích xuất các đặc trưng chính xác.
- **Khó Khăn Trong Việc Tổng Hợp Các Đặc Trưng:** Việc tổng hợp các đặc trưng cấp cao có thể gặp khó khăn trong việc duy trì thông tin chi tiết và phân biệt chính xác các đặc điểm kiến trúc tinh vi, đặc biệt là trong các trường hợp có sự khác biệt lớn giữa các kế hoạch mặt bằng.

#### e) **A system for Retrieving Images by content[5]:**

Venkat N. Gudivada và Vijay V. Raghavan đã giới thiệu một phương pháp sáng tạo cho phân tích ảnh bất động sản bằng cách kết hợp các kỹ thuật xử lý ảnh truyền thống và học máy. Phương pháp của họ bao gồm trích xuất cả các đặc trưng cấp thấp và cấp cao để nâng cao độ chính xác của các hệ thống truy xuất ảnh. Họ cũng thảo luận về việc tích hợp các đặc trưng này vào một khung làm việc thống nhất để phân loại và truy xuất ảnh hiệu quả hơn.

Những nghiên cứu này đã đóng góp quan trọng vào lĩnh vực truy xuất ảnh bất động sản bằng cách giải quyết các thách thức khác nhau như chất lượng bộ dữ liệu, trích xuất đặc trưng và độ chính xác phân loại. Chúng nhấn mạnh tiềm năng của việc tích hợp nhiều công nghệ và phương pháp tiếp cận để phát triển các hệ thống truy xuất ảnh bất động sản mạnh mẽ và hiệu quả. Bài báo đề xuất một hệ thống thu hồi hình ảnh dựa trên nội dung (Content-Based Image Retrieval - CBIR). Các thành phần chính của hệ thống bao gồm:

- **Trích Xuất Đặc Trưng:** Hình ảnh được phân tích để trích xuất các đặc trưng quan trọng như màu sắc, kết cấu, và hình dạng. Các đặc trưng này được mã hóa thành các vector hoặc biểu diễn số liệu.
- **Xây Dựng Cơ Sở Dữ Liệu:** Các đặc trưng trích xuất từ hình ảnh được lưu trữ trong một cơ sở dữ liệu hình ảnh. Mỗi hình ảnh trong cơ sở dữ liệu được gán một vector đặc trưng tương ứng.
- **Truy Vấn và So Sánh:** Khi người dùng thực hiện truy vấn bằng một hình ảnh hoặc mô tả nội dung, hệ thống so sánh các đặc trưng của hình ảnh truy vấn với các đặc trưng trong cơ sở dữ liệu để tìm ra những hình ảnh tương tự hoặc phù hợp nhất.
- **Hiển Thị Kết Quả:** Các hình ảnh có độ tương đồng cao nhất với truy vấn được hiển thị cho người dùng.

#### Ưu điểm:

- **Khả Năng Tìm Kiếm Chính Xác:** Hệ thống CBIR có khả năng tìm kiếm và phân loại hình ảnh dựa trên các đặc trưng nội dung thực tế, giúp nâng cao độ chính xác của các kết quả tìm kiếm.
- **Tính Tương Thích Cao:** Phương pháp có thể được áp dụng cho nhiều loại hình ảnh và lĩnh vực khác nhau, từ y tế đến giải trí và thương mại điện tử.
- **Độc Lập Với Metadata:** Phương pháp này không phụ thuộc vào các thông tin metadata của hình ảnh (như tiêu đề hoặc mô tả), mà chỉ dựa vào nội dung hình ảnh.

#### Nhược điểm:

- **Hiệu Suất Tính Toán Cao:** Việc trích xuất và so sánh các đặc trưng hình ảnh có thể đòi hỏi một lượng tính toán lớn, gây ra thời gian xử lý lâu và yêu cầu tài nguyên máy tính cao.
- **Độ Chính Xác Tùy Thuộc Vào Chất Lượng Đặc Trưng:** Độ chính xác của hệ thống phụ thuộc vào khả năng của thuật toán trích xuất đặc trưng. Nếu đặc trưng không đủ mạnh hoặc không chính xác, kết quả tìm kiếm có thể không đáp ứng được yêu cầu.

- **Khó Khăn Trong Việc Xử Lý Đặc Trưng Tinh Vi:** Một số đặc trưng phức tạp hoặc tinh vi (như cảm xúc hoặc ý nghĩa) có thể khó được trích xuất và phân tích, dẫn đến khả năng tìm kiếm hạn chế.

f) **Real Estate Image Classification For E-commerce Website[6]:**

Bài báo trình bày một phương pháp mới cho phân loại hình ảnh bất động sản trên các trang web thương mại điện tử. Hiện tại, hình ảnh là phần quan trọng nhất của các trang web thương mại điện tử và có ảnh hưởng lớn đến ngành bất động sản. Tác giả giới thiệu một dataset mới có tên REID, được thiết kế dựa trên yêu cầu thực tế từ trang web TopReal tại Việt Nam. Họ đề xuất hai mô hình mạng nơ-ron tích chập (ConvNets) cho bài toán phân loại hình ảnh: một mô hình dựa trên kiến trúc LeNet và một mô hình học chuyển giao sử dụng ResNet50 kết hợp với SVM.

**Đóng góp:**

- Đề xuất một dataset mới có tên REID, được mở rộng từ dataset REI, để phù hợp với các trang web thương mại điện tử ở Việt Nam.
- Xây dựng mô hình dựa trên LeNet để đánh giá hiệu suất trên 4 lớp tổng quát và 8 lớp cụ thể trong dataset REID.
- Đề xuất một framework kết hợp giữa ResNet50 và SVM cho bài toán phân loại hình ảnh, đạt được độ chính xác cao trên dataset REID.

**Phương pháp:**

- Mô hình LeNet: Được sử dụng như một mô hình cơ sở để phân tích và đánh giá dataset. Mô hình bao gồm bốn lớp tích chập và sử dụng max pooling để giảm kích thước biểu diễn và tăng tốc độ tính toán.
- Mô hình ResNet50 và SVM: Phương pháp học chuyển giao sử dụng ResNet50 như một mô-đun trích xuất đặc trưng và SVM như một bộ phân loại. ResNet50 được lựa chọn vì khả năng tránh vấn đề gradient biến mất khi số lượng lớp tăng lên.

**Ưu điểm:**

- Phương pháp đề xuất đạt độ chính xác cao, với kết quả là 97.6% và 87.3% cho các lớp 4 và 8 tương ứng.
- Dataset REID có thể được mở rộng và áp dụng cho nhiều bài toán khác trong ngành bất động sản.

**Nhược điểm:** Bài báo không đề cập chi tiết về các hạn chế tiềm tàng của phương pháp như yêu cầu về tài nguyên tính toán hoặc khả năng mở rộng của mô hình khi số lượng lớp hoặc kích thước dataset tăng lên.

### 3) TẬP DỮ LIỆU REI

REI Dataset là một tập dữ liệu mới được giới thiệu để phân loại hình ảnh bất động sản. Nó bao gồm hình ảnh thu thập từ các danh sách bất động sản thực tế và các nguồn trực tuyến, phục vụ cho ba nhiệm vụ phân loại chính:

- **Phân loại Cảnh (Place Classification):**

- Hình ảnh được thu thập từ cả danh sách bất động sản và các nguồn trực tuyến.
- Bao gồm sáu loại cảnh: phòng ngủ, phòng tắm, nhà bếp, phòng khách, sân trước, và sân sau.
- Mỗi loại chứa hơn 1.000 hình ảnh, tạo ra một thách thức lớn cho việc phân loại do sự đa dạng trong cùng một loại cảnh.

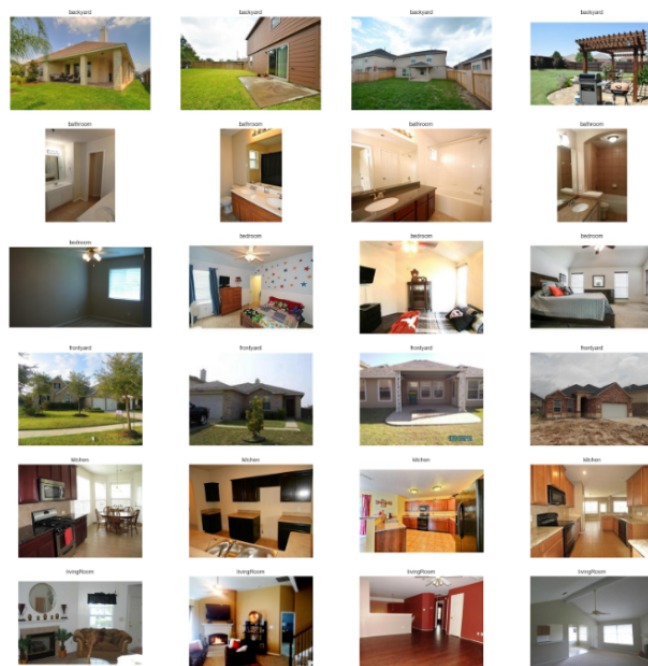
- **Phân loại Vật liệu cho Mặt Bàn (Material Classification for Counter-tops):**

- Sử dụng hình ảnh từ các nguồn trực tuyến.
- Tập trung vào việc nhận diện các loại vật liệu khác nhau như gạch, đá granite, Formica, v.v.

- **Phân loại Sàn (Flooring Classification):**

- Cũng sử dụng hình ảnh từ các nguồn trực tuyến.
- Nhấn đến việc phân loại các loại sàn khác nhau.
- Tập dữ liệu này có đủ số lượng mẫu cho mỗi loại để áp dụng các phương pháp học sâu, cho phép cải thiện khả năng phân loại và nhận dạng các đặc điểm quan trọng của bất động sản, hỗ trợ trong việc ước tính giá trị ngôi nhà.

Dưới đây là một số ảnh có trong tập dữ liệu:



## 4) PHƯƠNG PHÁP

Mục tiêu của đề án này là xây dựng một khung làm việc cho việc truy vấn hình ảnh bất động sản. Để đơn giản hóa khung làm việc, nhóm chia thành hai mô-đun: (a) mô-đun phân loại và (b) mô-đun truy vấn hình ảnh. Ở giai đoạn đầu nhóm tiến hành phân lớp tập dữ liệu bằng SVM và gán nhãn cho từng ảnh. Giai đoạn tiếp theo dùng EfficientNet-B7 để rút trích đặc trưng ảnh.

### 4.1 Mô đun phân loại SVM:

- Ta sẽ sử dụng lại phương pháp phân loại được đề cập đến trong [6], đó là phương pháp SVM (Support Vector Machine).

- SVM là một thuật toán học máy có giám sát được sử dụng phổ biến cho các bài toán phân loại và hồi quy. Mục tiêu chính của SVM là tìm một siêu phẳng (hyperplane) tối ưu để phân tách các mẫu dữ liệu thuộc các lớp khác nhau. SVM đặc biệt hiệu quả khi giải quyết các bài toán phân loại hai lớp (binary classification), nhưng có thể được mở rộng để giải quyết các bài toán phân loại đa lớp (multi-class classification) bằng cách sử dụng các phương pháp như "one-vs-one" hoặc "one-vs-all."

- Một trong những đặc điểm mạnh mẽ của SVM là khả năng áp dụng các hàm kernel để xử lý các bài toán phân loại phi tuyến tính. Khi dữ liệu không thể phân tách tuyến tính trong không gian ban đầu, các hàm kernel sẽ chuyển dữ liệu sang một không gian đặc trưng có chiều cao hơn, nơi dữ liệu có thể phân tách tuyến tính được.

- Các loại kernel phổ biến bao gồm:

- **Linear Kernel:** Sử dụng khi dữ liệu có thể phân tách tuyến tính.
- **Polynomial Kernel:** Sử dụng khi dữ liệu có mối quan hệ phi tuyến đơn giản.
- **Radial Basis Function (RBF) Kernel:** Sử dụng cho các trường hợp dữ liệu phức tạp và phi tuyến tính.

- Ta sẽ sử dụng SVM-RBF Kernel trong mô hình thay vì các loại kernel khác bởi vì độ phức tạp của tập dữ liệu REI.

#### SVM với RBF Kernel[7]:

- RBF Kernel, còn được gọi là Gaussian Kernel, là một trong những loại kernel được sử dụng phổ biến nhất khi dữ liệu không thể phân tách tuyến tính. RBF Kernel chuyển dữ liệu từ không gian ban đầu sang một không gian đặc trưng có chiều cao hơn, nơi siêu phẳng có thể phân tách các lớp khác nhau một cách hiệu quả.

- Cách hoạt động của SVM với RBF Kernel:

- **Chuyển đổi không gian:** RBF Kernel chuyển đổi các điểm dữ liệu từ không gian đầu vào sang một không gian có chiều cao hơn, nơi mà một siêu phẳng có thể được sử dụng để phân tách các điểm dữ liệu của các lớp khác nhau.



- **Tối ưu hóa siêu phẳng:** Sau khi chuyển đổi, SVM tìm kiếm siêu phẳng tối ưu trong không gian đặc trưng này. Quá trình này tối đa hóa khoảng cách giữa siêu phẳng và các support vectors (các điểm gần nhất từ mỗi lớp với siêu phẳng).

- **Gamma ( $\gamma$ ) trong RBF Kernel:** Tham số điều khiển độ ảnh hưởng của một điểm dữ liệu đơn lẻ. Nếu  $\gamma$  lớn, mỗi điểm dữ liệu sẽ có một ảnh hưởng lớn, dẫn đến khả năng phân tách rất chi tiết nhưng có thể dẫn đến overfitting. Nếu  $\gamma$  nhỏ, các điểm dữ liệu sẽ có ảnh hưởng ít hơn, dẫn đến một ranh giới phân chia mượt mà hơn nhưng có thể không nắm bắt được các chi tiết quan trọng trong dữ liệu.

- Ưu điểm của SVM với RBF Kernel:

- **Khả năng phân loại phi tuyến tính:** RBF Kernel cho phép SVM xử lý các bài toán phân loại phi tuyến tính một cách hiệu quả mà không cần phải tìm kiếm trực tiếp một hàm phi tuyến.
- **Hiệu suất tốt trên dữ liệu phức tạp:** SVM với RBF Kernel có thể hoạt động tốt trên các bộ dữ liệu có cấu trúc phức tạp, nơi mà các phương pháp phân loại tuyến tính có thể thất bại.
- **Tính mềm dẻo:** Tham số  $\gamma$  của RBF Kernel cho phép điều chỉnh mô hình để đạt được sự cân bằng tốt giữa độ chính xác và khả năng tổng quát hóa.

- Nhược điểm của SVM với RBF Kernel:

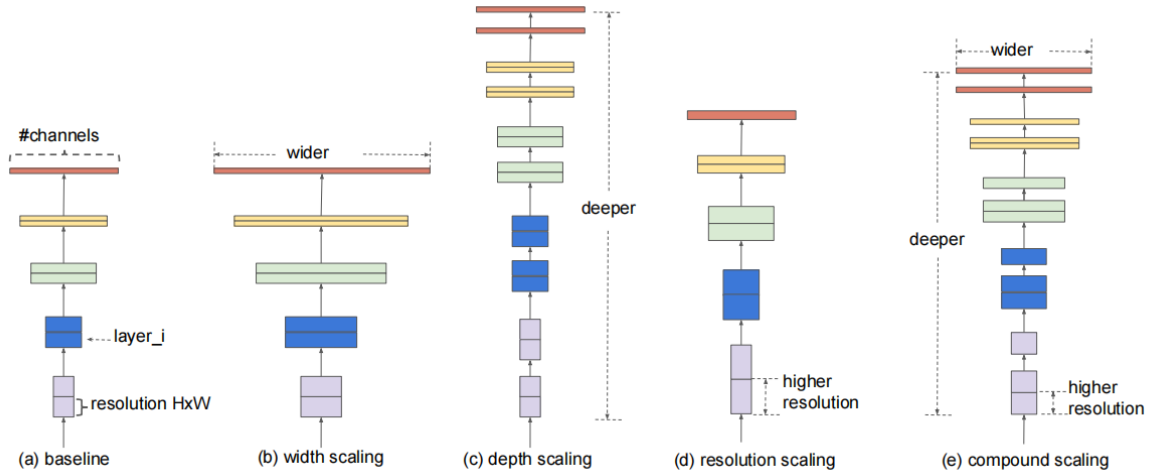
- **Cần tối ưu hóa tham số:** Việc lựa chọn tham số  $\gamma$  và  $C$  (tham số điều chỉnh độ mềm dẻo của ranh giới phân chia) cần phải được tối ưu hóa cẩn thận, thường thông qua quá trình tìm kiếm lưới (grid search) hoặc các phương pháp tối ưu khác.
- **Chi phí tính toán:** Khi số lượng điểm dữ liệu và số lượng đặc trưng tăng lên, chi phí tính toán cho việc huấn luyện SVM với RBF Kernel có thể tăng lên rất nhanh, dẫn đến thời gian huấn luyện dài.

## 4.2 Mô đun rút trích đặc trưng EfficientNet:

### 4.2.1 Compound scaling:

- Như ta đã biết, có ba kích thước tỷ lệ của CNN: depth, width, and resolution. Depth là độ sâu của mạng tương đương với số lớp trong đó, Width là độ rộng của mạng. Ví dụ: một thước đo chiều rộng là số kênh trong lớp Conv và Resolution là độ phân giải hình ảnh được chuyển đến CNN.

Hình bên dưới (từ bài báo [8]) sẽ cho chúng ta ý tưởng rõ ràng về việc thu phóng mô hình có nghĩa là gì trên các kích thước khác nhau. Chúng ta cũng sẽ thảo luận chi tiết về những điều này ở ngay sau đây.



- Các chiều mở rộng: Các phương pháp truyền thống thường mở rộng ConvNet theo một trong ba chiều:

- **Chiều sâu (depth):** Tăng số lớp để tăng khả năng nắm bắt các đặc điểm phức tạp, nhưng việc tăng quá nhiều sẽ gặp vấn đề về gradient biến mất.
- **Chiều rộng (width):** Tăng số kênh để cải thiện khả năng nhận diện chi tiết, nhưng nếu quá rộng mà nông thì khó nắm bắt đặc điểm cấp cao.
- **Độ phân giải (resolution):** Tăng độ phân giải để nắm bắt chi tiết tốt hơn, nhưng lợi ích giảm dần khi độ phân giải quá cao.

- Tuy nhiên khi ta tăng các giá trị depth, width, resolution đến một ngưỡng nào đó thì độ chính xác không thể tăng thêm được nữa, do đó việc chúng ta tìm ra bộ tham số  $d, w, r$  phù hợp để mô hình đạt được độ chính xác cao nhất gọi là Compound Scaling - Thu phóng mô hình kết hợp. Các chiều mở rộng không độc lập với nhau; cần cân bằng giữa chiều rộng, chiều sâu và độ phân giải để đạt hiệu quả tốt nhất.

- Định hình vấn đề: Một lớp ConvNet có thể được biểu diễn như một hàm:  $Y_i = F_i(X_i)$ , trong đó  $F_i$  là toán tử,  $Y_i$  là đầu ra và  $X_i$  là đầu vào với kích thước không gian  $\langle H_i, W_i, C_i \rangle$ , trong đó  $H_i$  và  $W_i$  là kích thước không gian còn  $C_i$  là số chiều của kênh. Một ConvNet  $N$  gồm nhiều lớp có thể được biểu diễn như một chuỗi các hàm:

$$\mathcal{N} = \bigodot_{i=1}^s \mathcal{F}_i^{L_i}(X(H_i, W_i, C_i))$$

- Mục tiêu là mở rộng mạng theo chiều dài  $L_i$ , chiều rộng  $C_i$ , và/hoặc độ phân giải  $(H_i, W_i)$  mà không thay đổi kiến trúc lớp cơ sở  $F_i$ . Bài toán tối ưu có thể được mô hình hóa như sau:

$$\max_{\alpha, \beta, \gamma} \text{Accuracy}(N(\alpha^\phi, \beta^\phi, \gamma^\phi)) \quad \text{với} \quad \begin{cases} \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (1), \\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad (2), \\ \text{Memory}(N) \leq \text{target memory} \quad (3), \\ \text{FLOPS}(N) \leq \text{target flops} \quad (4) \end{cases}$$

Trong đó:

$\alpha^\phi$  : chiều sâu

$\beta^\phi$  : chiều rộng

$\gamma^\phi$  : độ phân giải

$\phi$  : Hệ số compound

#### 4.2.2 Cấu trúc EfficientNet:

- Nhóm tác giả đã phát triển một mạng cơ sở mới bằng cách áp dụng tìm kiếm kiến trúc mạng nơron đa mục tiêu để tối ưu hóa cả độ chính xác và số lượng phép tính (FLOPS). Cụ thể, họ sử dụng cùng một không gian tìm kiếm như trong nghiên cứu MnasNet, nhưng thay vì tối ưu hóa độ trễ như nghiên cứu trước, nhóm này tập trung tối ưu hóa FLOPS. Công thức tối ưu hóa là:

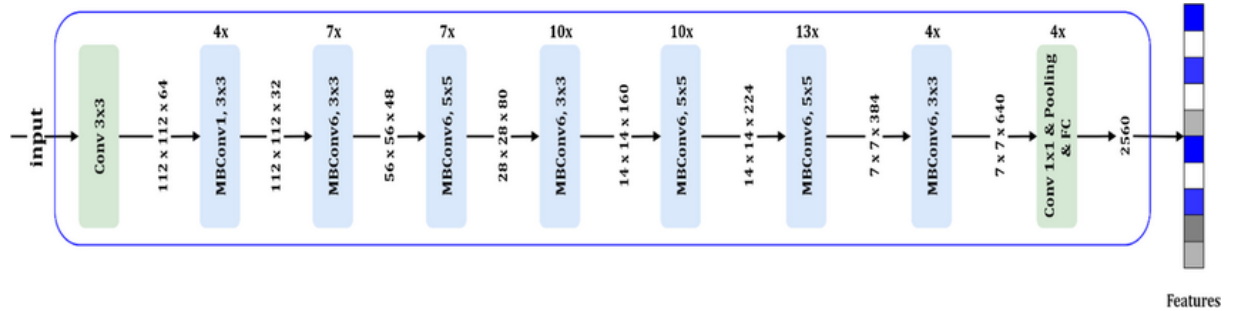
$$\text{ACC}(m) \times \left[ \frac{\text{FLOPS}(m)}{T} \right]^w$$

Trong đó,  $\text{ACC}(m)$  là độ chính xác,  $\text{FLOPS}(m)$  là số phép tính của mô hình,  $T$  là mức FLOPS mục tiêu, và  $w = -0.07$  là siêu tham số điều chỉnh sự cân bằng giữa độ chính xác và FLOPS. Kết quả là một mạng hiệu quả mới được gọi là EfficientNet-B0.

Stage $i$	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	$224 \times 224$	32	1
2	MBConv1, k3x3	$112 \times 112$	16	1
3	MBConv6, k3x3	$112 \times 112$	24	2
4	MBConv6, k5x5	$56 \times 56$	40	2
5	MBConv6, k3x3	$28 \times 28$	80	3
6	MBConv6, k5x5	$14 \times 14$	112	3
7	MBConv6, k5x5	$14 \times 14$	192	4
8	MBConv6, k3x3	$7 \times 7$	320	1
9	Conv1x1 & Pooling & FC	$7 \times 7$	1280	1

- Bắt đầu từ mô hình cơ sở EfficientNet-B0, nhóm tác giả áp dụng phương pháp thu phóng phức hợp của mình để thu phóng quy mô với hai bước bao gồm:

1. Đặt cố định giá trị của  $\phi$  bằng 1, nhóm tác giả thu được bộ giá trị tối ưu:  $\alpha = 1.2; \beta = 1.1; \gamma = 1.15$  theo ràng buộc (1).
2. Cố định  $\alpha, \beta, \gamma$  dưới dạng các hằng số và thu phóng mạng cơ sở với các  $\phi$  khác nhau từ đó thu được để thu được từ EfficientNet-B1 đến EfficientNet-B7.



- Sau khi có được mô hình EfficientNet ta sẽ pretrain lại trên tập data REI để mô hình học cách rút trích đặc trưng.

## 5) THỰC NGHIỆM

- Đầu vào của chương trình là một ảnh truy vấn:



- Đầu ra là tập 20 ảnh giống nhất:

```

Enter the path to an image (type 'exit' to quit):
/kaggle/input/rei-dataset/REI-Dataset_train/bathroom/bathroom (1).jpg
1/1 _____ 10s 10s/step
1/1 _____ 12s 12s/step

```



exit

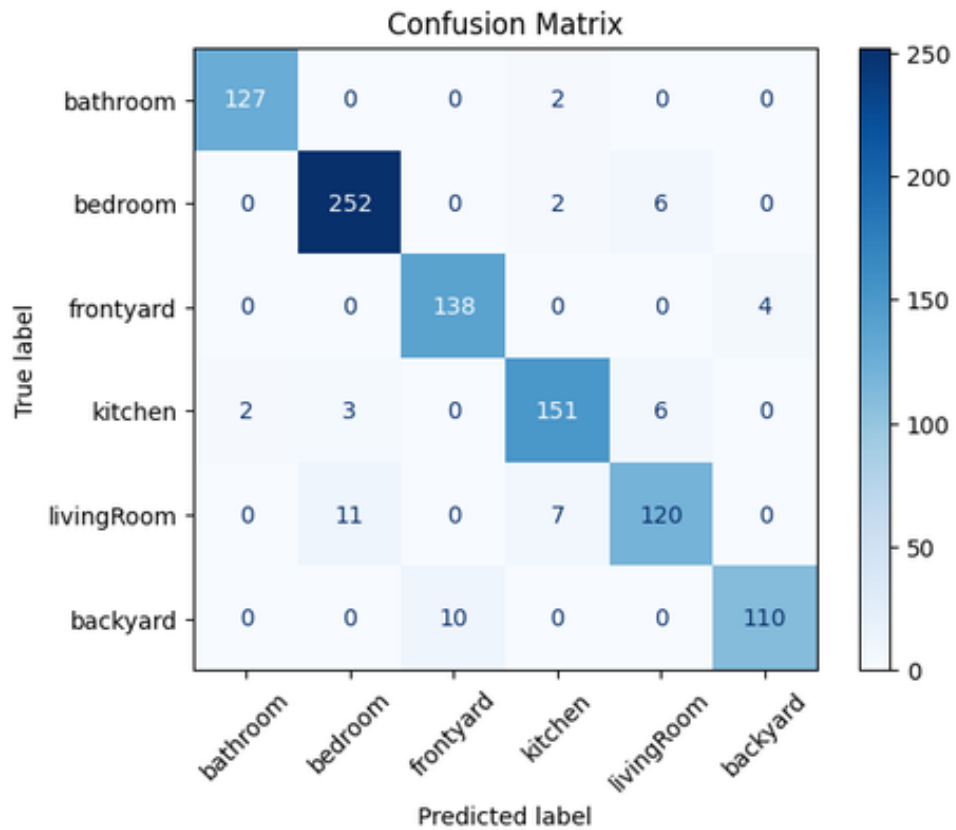
- Để so sánh độ chính xác của mô hình, nhóm sẽ chạy thực nghiệm trên cả ba mô hình: ResNet50, EfficientNet-B0 và EfficientNet-B7 trước khi được phân lớp và sau khi được phân lớp.

Mô hình	Chưa phân lớp	Đã phân lớp
RestNet50	86,3%	94,3%
EfficientNet-B0	85,6%	93,2%
EfficientNet-B7	92,4%	96%

#### Bổ sung kiến thức về kiến trúc ResNet50:

- ResNet-50 bao gồm 50 lớp, trong đó có 49 lớp tích chập (convolutional layers) và một lớp kết nối đầy đủ (*fully connected layer*).
- Điểm nổi bật của ResNet là các *residual blocks* (khối dư), cho phép các thông tin từ đầu vào của một lớp được truyền trực tiếp tới đầu ra của lớp đó thông qua một kết nối tắt (*skip connection*). Điều này giúp giảm thiểu mất mát thông tin và cho phép huấn luyện các mạng rất sâu mà không gặp phải vấn đề về gradient.
- Kiến trúc của ResNet-50 gồm nhiều *residual blocks*, mỗi block có ba lớp tích chập và một kết nối tắt.

- Dưới đây là kết quả thực nghiệm của ResNet50:

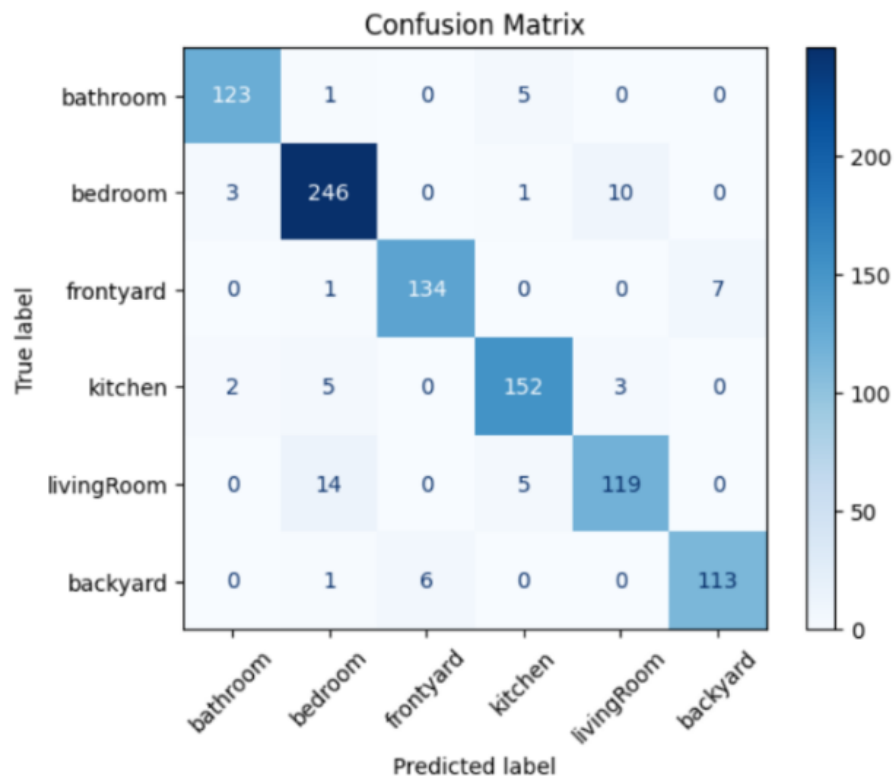


- Mô hình ResNet50 cho thấy độ chính xác khá cao ở các lớp "bedroom" và "frontyard," với số lượng dự đoán đúng lần lượt là 252 và 138.

- Tuy nhiên, có một số lỗi dự đoán ở lớp "livingRoom," nơi mô hình nhầm lẫn với lớp "bedroom" và "kitchen."

- Tổng thể, ResNet50 có hiệu quả cao trong việc phân loại các lớp chính, nhưng gặp khó khăn trong việc phân biệt giữa một số lớp như "livingRoom" và "bedroom."

### EfficientNet-B0:

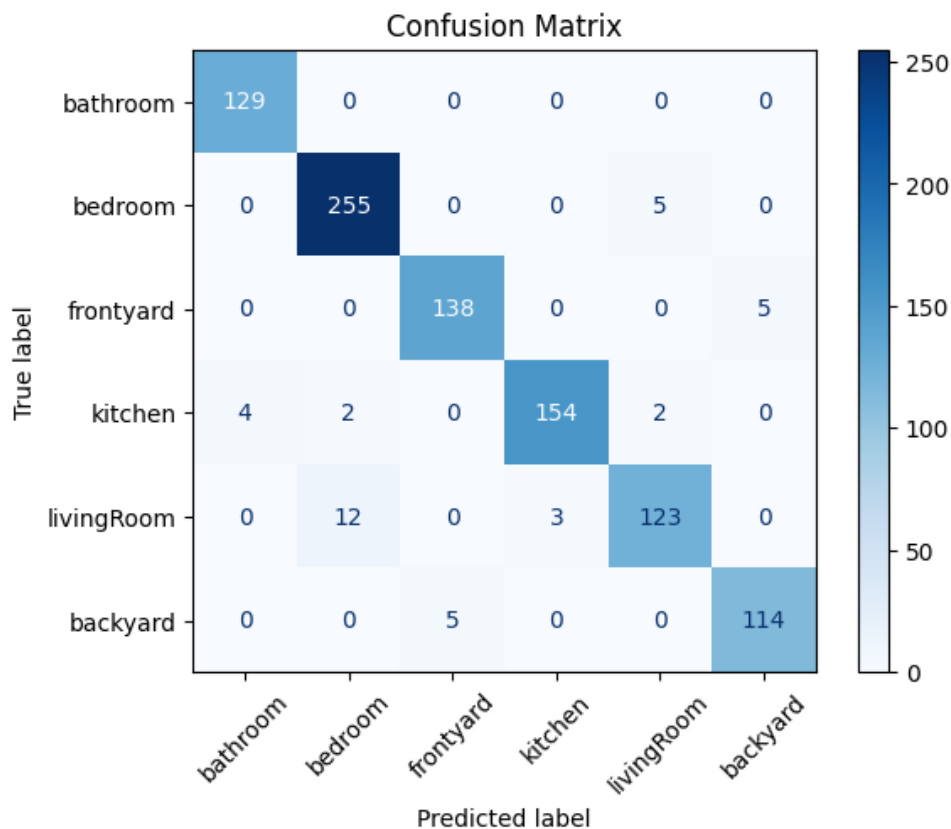


- EfficientNetB0 có sự cải thiện nhẹ trong độ chính xác ở lớp "kitchen" và "backyard," nhưng vẫn gặp lỗi dự đoán ở các lớp như "livingRoom."

- Lớp "bedroom" vẫn được dự đoán với độ chính xác cao (246), nhưng có một số lỗi nhỏ hơn so với ResNet50.

- Nhìn chung, EfficientNetB0 hoạt động tốt nhưng không có sự cải thiện đáng kể so với ResNet50.

### EfficientNet-B7:



- EfficientNetB7 cho thấy sự cải thiện rõ rệt so với hai mô hình trước đó. Lớp "bedroom" được phân loại với độ chính xác rất cao (255), và các lớp khác như "kitchen" và "backyard" cũng có độ chính xác tốt hơn.

- Số lượng lỗi dự đoán đã giảm, đặc biệt là ở lớp "livingRoom," nơi chỉ có rất ít lỗi nhầm lẫn so với hai mô hình trước.

- EfficientNetB7 có vẻ là mô hình hiệu quả nhất trong số ba mô hình, với độ chính xác cao hơn và ít lỗi dự đoán hơn.

- Còn nếu ta nói về chỉ số AP (average precision), kết quả thực nghiệm đối với 20 ảnh đầu ra cho thấy đối với kiến trúc ResNet50, chỉ số ap của nó chỉ dừng ở mức khoảng 0.85, đối với kiến trúc EfficientNet-B0 thì cũng chỉ khoảng 0.83. Cả hai con số đều khá khiêm tốn khi so sánh với kiến trúc EfficientNet-B7 với ap là 1. Có nghĩa là nếu chỉ truy suất 20 ảnh thì EfficientNet-B7 luôn cho ra kết quả đúng toàn diện nhất.

## 6) KẾT LUẬN

Với tập dữ liệu REI còn khá nhỏ nên chưa nhìn thấy rõ được sự vượt trội của EfficientNet so với ResNet. Tuy nhiên có thể thấy rằng EfficientNetB7 cho kết quả tốt nhất với độ chính xác cao hơn và ít lỗi



dự đoán hơn, chứng tỏ rằng mạng phức tạp hơn có thể học đặc trưng tốt hơn từ dữ liệu hình ảnh nhưng nó đòi hỏi tài nguyên tính toán nhiều hơn. Nếu cân cân bằng giữa hiệu suất và tài nguyên, ResNet50 là lựa chọn hợp lý. Bên cạnh đó việc phân lớp trước tập dữ liệu giúp cho quá trình truy vấn thực hiện chính xác hơn.

## References

- [1] David Koch et al. “Real Estate Image Analysis: A Literature Review”. In: Journal of Real Estate Literature 27.2 (2019), pp. 269–298. DOI: 10.22300/0927-7544.27.2.269. URL: [https://www.researchgate.net/publication/338955653\\_Real\\_Estate\\_Image\\_Analysis\\_A\\_Literature\\_Review](https://www.researchgate.net/publication/338955653_Real_Estate_Image_Analysis_A_Literature_Review).
- [2] Sergejs Kodors et al. “Real Estate Monitoring System Based on Remote Sensing and Image Recognition Technologies”. In: Procedia Computer Science. Vol. 104. ICTE 2016. Elsevier B.V., 2017, pp. 460–467. DOI: 10.1016/j.procs.2017.01.160. URL: <https://www.sciencedirect.com/science/article/pii/S1877050917302603>.
- [3] Corneliu Florea et al. “Automatic real-estate image analysis for retrieval and classification”. In: Bulletin of the Polytechnic Institute of Iași (2022). DOI: 10.2478/bipie-2022-0009. URL: <https://sciendo.com/article/10.2478/bipie-2022-0009>.
- [4] Divya Sharma<sup>1</sup> and Chiranjoy Chattopadhyay. “High-level feature aggregation for fine-grained architectural floor plan retrieval”. In: IET Computer Vision. IET. 2018. DOI: 10.1049/iet-cvi.2017.0581. URL: <https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.1049/iet-cvi.2017.0581>.
- [5] Venkat N. Gudivad and Vijay V. Raghavan. “A System for Retrieving Images by Content”. In: RIAO’94: Intelligent Multimedia Information Retrieval Systems and Management (1994). URL: <https://dl.acm.org/doi/pdf/10.5555/2856823.2856859>.
- [6] Vo Hoai Viet, Nguyen Nhat Khoa, and Pham Huynh Nhat. “Real Estate Image Classification For E-commerce Website”. In: Science Technology Development Journal 25.1 (2022), pp. 2224–2238. DOI: 10.32508/stdj.v25i1.3443.
- [7] Johar M. Ashfaq and Amer Iqbal. “Introduction to Support Vector Machines and Kernel Methods”. In: (Apr. 2019). URL: <https://www.researchgate.net/publication/332370436>.
- [8] Mingxing Tan and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. In: International Conference on Machine Learning (2019). DOI: 1905.11946.