# CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

Trong chương này chúng tôi sẽ trình bày những kiến thức liên quan trong bài toán phân loại hình ảnh kết cấu bao gồm các nội dung:

* Trích xuất đặc trưng cục bộ hình ảnh: LBP, LDZP
* Trích xuất đặc trưng hình ảnh kết cấu dựa trên mạng nơ-ron học sâu Convolution Neural Network (CNN)
* Trích xuất đặc trưng cục bộ kết hợp với bộ phân loại sử dụng kĩ thuật học sâu.

Kết cấu được sử dụng để mô tả một khu vực trong đó các yếu tố kết cấu được đặc trưng hóa trong mối quan hệ không gian. Một hình ảnh có thể bao gồm một hoặc nhiều kết cấu. Nếu nhiều kết cấu tồn tại trong một hình ảnh, ranh giới giữa hai kết cấu có thể được phát hiện và phân biệt bằng cách sử dụng các độ đo kết cấu được định nghĩa trước. Các độ đo kết cấu có thể cung cấp thông tin quan trọng để phân đoạn hình ảnh, trích xuất đặc trưng và phân loại hình ảnh. Độ đo kết cấu rất hữu ích trong việc giải đoán ảnh chụp từ vệ tinh viễn thám, ảnh cộng hưởng từ y tế, khoa học vật liệu và ảnh địa hình hàng không. Ví dụ, việc nghiên cứu sự phát triển đất ở thành thị và nông thôn trong ảnh vệ tinh có thể được hưởng lợi từ việc sử dụng phân tích kết cấu ảnh.

Có nhiều biện pháp kết cấu để xác định đặc điểm của kết cấu. Quá trình Đặc điểm hóa kết cấu sẽ tạo ra một tập hợp các đặc trưng của kết cấu trong ảnh. Các đặc trưng khu vực như độ thô, tính đồng nhất, mật độ, độ mịn, độ thẳng, hướng, độ chi tiết và tần số thường được sử dụng làm đặc trưng kết cấu. Nhiều phương pháp tiếp cận bao gồm các hàm tự tương quan, ma trận đồng xuất hiện thang độ xám và mẫu nhị phân địa phương (LBP) được sử dụng để mô tả và trích xuất các đặc trưng kết cấu trong một hình ảnh. Tất cả các cách tiếp cận này thuộc bốn loại: phương pháp thống kê, phương pháp cấu trúc, phương pháp dựa trên mô hình và phương pháp dựa trên biến đổi [9], [10], [11], [12]. Trong hầu hết các trường hợp, các đặc trưng của kết cấu được biểu diễn bằng số bằng các vectơ đặc trưng, ​​bao gồm các thành phần đặc trưng bắt nguồn từ một vùng lân cận của lớp kết cấu tương ứng.

Mỗi cách tiếp cận này đều có những ưu điểm và nhược điểm riêng. Ví dụ, cách tiếp cận thống kê phù hợp với kết cấu vi mô (tức là kết cấu ngẫu nhiên) trong khi phương pháp tiếp cận theo cấu trúc phù hợp với kết cấu vĩ mô hoặc các mẫu kết cấu được xác định rõ như kết cấu tuần hoàn [13], [14]. Nhiều phương pháp để mô tả các đặc điểm kết cấu phụ thuộc vào các tham số được sử dụng như kích thước vùng lân cận cho vùng kết cấu, định lượng mức xám và định hướng để đo lường mối quan hệ giữa các pixel như khoảng cách và góc [11].

Một khi đặc trưng kết cấu được trích xuất từ một kết cấu, bước tiếp theo là thực hiện phân loại kết cấu của đối tượng trong ảnh. Phương pháp phân loại ảnh kết cấu sẽ bao gồm hai nhóm chính: Nhóm thứ nhất dựa trên đặc trưng với mức độ địa phương hóa không gian cao. Trong nhóm này, hầu hết các phương pháp phát hiện cạnh có thể sử dụng các đặc trưng kết cấu để địa phương hóa không gian. Vấn đề với nhóm phương pháp này là khó phân biệt ranh giới kết cấu và các cạnh rìa được tìm thấy trong cùng một kết cấu. Nhóm thứ hai dựa trên các hàm phân loại với đầu vào là các đặc trưng kết cấu. Độ chính xác phân loại của nhóm này phụ thuộc vào khả năng của hàm phân loại và mức độ phù hợp của đặc trưng được trích xuất với hàm phân loại. Do đó, bước quan trọng nhất trong nhóm phương pháp thứ hai là trích xuất đặc trưng, các đặc trưng được trích chọn này phải có khả năng phân biệt giữa các nhóm kết cấu khác nhau [15], [16]. Nhóm tác giả trong bài báo [9], [10], [11] định nghĩa một nhân tố (primitive) là một tập hợp các điểm ảnh có kết nối với nhau được đặc trưng bởi một danh sách các thuộc tính. Mỗi một primitive được gọi là một nhân tố kết cấu có thể được gọi là “texel” hoặc “texton”. Một mẫu kết cấu có thể được mô tả bởi một hoặc nhiều primitive có liên quan đến nhau về mặt không gian. Một primitive nhỏ nhất có thể là chính các điểm ảnh. Đặc trưng của một primitive được phân phối trong các vùng lân cận nó. Những khái niệm này được sử dụng rất nhiều và tỏ ra hiệu quả cao trong các bài toán phân loại kết cấu đối tượng trong ảnh [17], [18], [19], [20]. Một trong những phương pháp nổi tiếng nhất được sử dụng rộng rãi cho phân loại kết cấu là mẫu nhị phân cục bộ (Local Binary Pattern - LBP). Phương pháp LBP này chỉ quan tâm đến các mối quan hệ về không gian mà không quan tâm đến các mối quan hệ về màu sắc của các điểm ảnh.

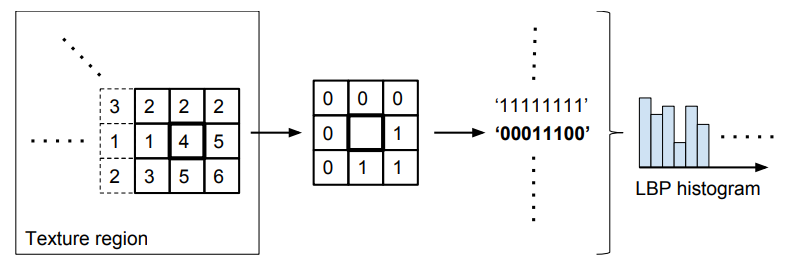
Ngày nay, ảnh màu trở thành một trong những nơi lưu trữ thông tin phổ biến trong các ngành nghề truyền thông, công nghệ thông tin. Đặc trưng màu cũng là một trong những đặc trưng mạnh, tiềm năng cho phân loại kết cấu của đối tượng trong ảnh. Hơn nữa, các đối tượng khác nhau sẽ có những đặc trưng màu sắc khác nhau, do đó, có nhiều phương pháp sử dụng đặc trưng màu sắc để phân loại ảnh kết cấu của đối tượng. Đặc trưng màu sắc được sử dụng kết hợp với đặc trưng không gian của kết cấu để phân loại kết cấu của đối tượng. Đặc trưng LBP và đặc trưng dựa trên phân bố màu sắc của điểm ảnh được kết hợp với nhau cho bài toán phân loại cấu trúc của ảnh màu. Mỗi nhóm kết cấu có một mẫu đặc trưng có thể phân biệt nhóm đó với những nhóm khác trong bài toán phân loại. Để xác định một điểm ảnh thuộc về một nhóm kết cấu nào đó, những điểm ảnh lân cận của điểm ảnh đó được lấy ra để đo mức độ tương tự của nó với một kết cấu và nhóm phổ màu.

* 1. **TRÍCH CHỌN ĐẶC TRƯNG CỤC BỘ HÌNH ẢNH KẾT CẤU**
     1. **Local binary patterns (LBP)**

Mặc dù các đặc trưng toàn cục của hình ảnh kết cấu đã đạt được những kết quả đáng mong đợi bằng cách khai thác các phân phối chung các giá trị pixel trong các vùng lân cận cực kì nhỏ (ví dụ 3 x 3). Phương pháp trích xuất đặc trưng hình ảnh cục bộ cung cấp một phân tích kết cấu mạnh mẽ giá trị cường độ pixel trong các vùng lân cận cục bộ. Trong đó tiêu biểu kể đến là Local Binary Pattern (LBP) và các biến thể của nó.

Bằng cách tính toán biểu đồ tần suất xuất hiện (histogram) các bộ mô tả cục bộ, các phương pháp tiếp cận giống như LBP kết hợp cấu trúc và các phương pháp thống kê góp phần lớn vào việc tăng hiệu suất trong vấn đề phân tích hình ảnh kết cấu. Biểu đồ histogram là một vector thưa, nó tóm tắt sự xuất hiện của các bộ mô tả cục bộ tại mọi vị trí pixel bất kể vị trí không gian của chúng do đó nó sẽ loại bỏ các hình dạng và bố cục hình ảnh toàn cục.

LBP được bắt nguồn ý tưởng từ sự kết hợp của việc phân tích các cấu trúc cục bộ trong phương pháp phân tích cấu trúc và phân tích sự xuất hiện của các phương pháp thống kê như ma trận đồng xuất hiện (GLCM- Gray Level Co-occurrence Matrix). LBP ban đầu được giới thiệu với ý tưởng cơ bản để tóm tắt một vùng kết cấu hoặc một hình ảnh bằng cách so sánh từng pixel với các pixel lân cận (ban đầu là 3 x 3). Đối với mỗi pixel mã nhị phân được tính bằng cách tạo ngưỡng của các pixel lân cận dựa trên giá trị pixel trung tâm như hình 2.1. Biểu đồ tần suất xuất hiện (histogram) sau đó được tính toán dẫn đến một vector đặc trưng 256 chiều ( = 256 mã có thể có). Bộ mô tả này rất phổ biến do sự đơn giản, dễ triển khai, chi phí tính toán thấp của nó và sự bất biến của nó đối với sự thay đổi ánh sáng đơn điệu.



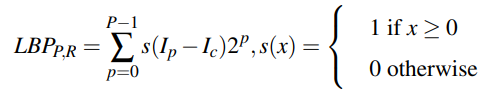
**Hình 2.1 Quá trình trích xuất biểu đồ histogram từ bộ mô tả LBP**

Tuy nhiên, LBP được mô tả có một số nhược điểm lớn: Đầu tiên, biểu đồ histogram

nhạy cảm với việc xoay hình ảnh. Thứ hai, nó có hỗ trợ không gian nhỏ (vùng lân cận 3 × 3) không trích xuất thông tin kết cấu quy mô lớn. Thứ ba, LBP mất thông tin kết cấu cục bộ (ví dụ: độ tương phản) bằng cách chỉ xem xét dấu hiệu của sự khác biệt của các pixel lân cận. Cuối cùng, nó rất nhạy cảm với nhiễu và làm mờ do các dao động nhỏ trên hoặc dưới giá trị trung tâm làm thay đổi mã nhị phân tương tự như một sự tương phản. Một số lượng lớn các biến thể của bản gốc LBP đã được đề xuất để cố gắng khắc phục những nhược điểm này.

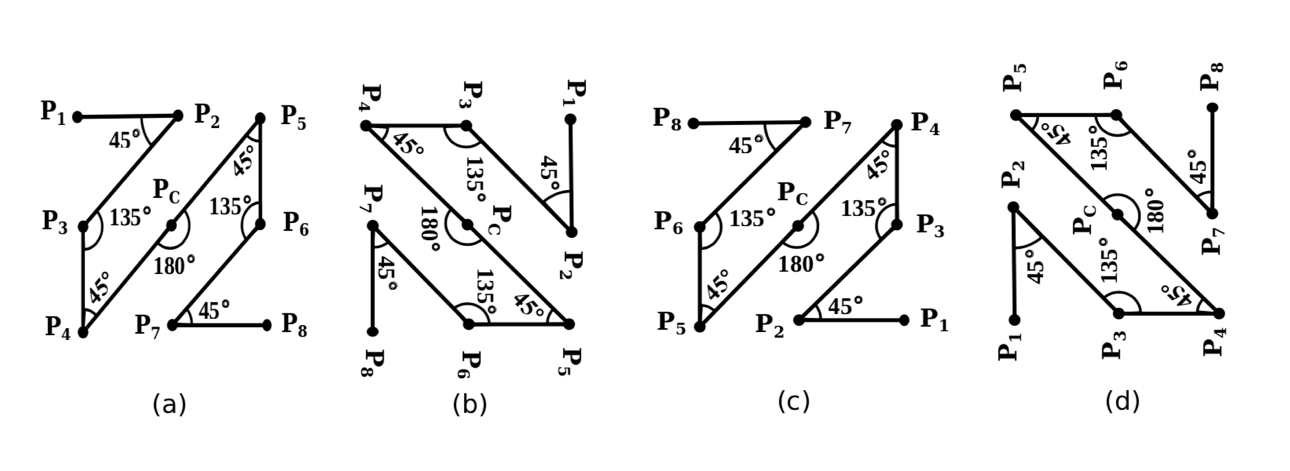
Một phiên bản bất biến xoay vòng có tên LBPROT được thực hiện bằng cách xoay vòng các mã nhị phân giống nhau. Điều này dẫn đến giảm kích thước của biểu đồ tần suất xuất hiện (histogram) biểu thị các lần xuất hiện của 36 vòng quay duy nhất bất biến các mẫu. Tuy nhiên, lượng tử hóa không gian góc với tám pixel trong vùng lân cận không đối xứng (bình phương) hình tròn không được áp dụng cho tính toán của bất biến quay. Hơn nữa, sự xuất hiện của 36 vòng quay duy nhất bất biến các mẫu nhị phân thay đổi phần lớn vì một số mẫu khó có thể xảy ra.

LBP bất biến xoay lần đầu tiên được cải thiện bằng cách sử dụng các giá trị cường độ nội suy song tuyến được lấy mẫu trên một vòng tròn có bán kính thay đổi xung quanh pixel trung tâm. Lượng tử của không gian góc có thể được sửa đổi bằng cách thay đổi số lượng giá trị nội suy và cấu trúc cục bộ có thể được mô tả ở nhiều quy mô bằng cách thay đổi bán kính. LBP tại một vị trí pixel (xc, yc) được tính như sau:



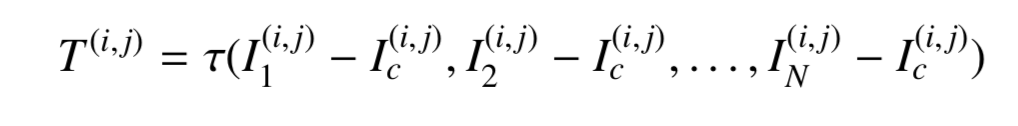
Trong đó: , P là giá trị nội suy được xem xét trong vùng lân cận và R là bán kính hình tròn mà các giá trị này nằm trên đó.

* + 1. **Local Directional ZigZag Pattern (LDZP)**

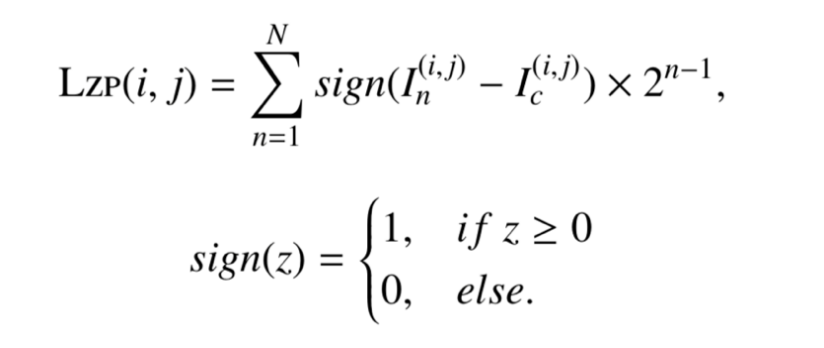
****

**Hình 2.2 Các cấu trúc ZigZag**

ZigZag LBP biểu diễn hiệu quả đặc trưng của mối quan hệ giữa điểm ảnh trung tâm và các điểm ảnh hàng xóm của nó theo cơ chế quét điểm theo hình Zigzag được gọi là mẫu Zig Zag. LZP là một bộ mô tả đặc trưng kết cấu mức xám cục bộ biểu diễn các cấu trúc Zig Zag không gian cục bộ của các ảnh kết cấu như được trình bày trong hình 2.2. Ảnh xám đầu vào được ký hiệu là I, Pc(i,j) là điểm ảnh trung tâm của cửa sổ 3x3 có giá trị mức xám Ic(i,j) và điểm ảnh hàng xóm thứ n của Pc(i,j) được ký hiệu là Pn(i,j) có giá trị mức xám là In(i,j), n là một số nguyên dương ở trong khoảng từ [1,N]. Giá trị N trong trường hợp này là 8 thể hiện có 8 điểm ảnh hàng xóm của điểm ảnh trung tâm. Kết cấu cục bộ của hình ảnh đơn sắc I được biểu thị bằng sự phân phối chung của chênh lệch giá trị xám giữa pixel trung tâm và N lân cận của nó (N> 0) được xác định như sau,



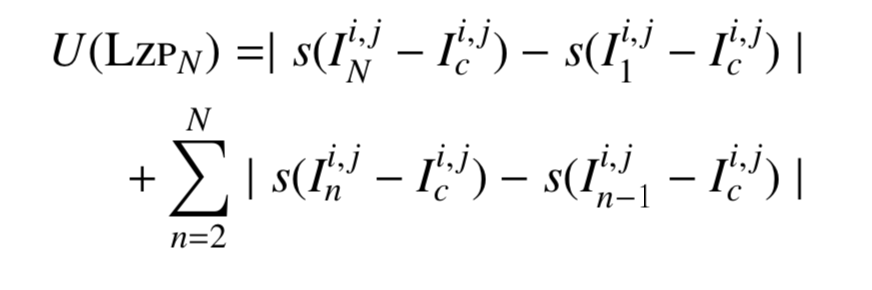
Nơi hàm τ () biểu diễn hàm phân phối chung. Để mã hóa thông tin cấu trúc sử dụng các mẫu Zig Zag chỉ cần quan tâm đến dấu của sai lệch giữa giá trị mức xám của điểm ảnh trung tâm và điểm ảnh hàng xóm sign(In(i,j) – Ic(i,j)), đây là điều làm cho LZP bất biến theo sự biến đổi của giá trị mức xám. Do đó, toán tử LZP là tối ưu cho với các biến đổi về cường độ sáng và được định nghĩa như sau:



Dấu của giá trị hiệu giữa giá trị mức xám của điểm ảnh trung tâm và các điểm ảnh hàng xóm được mô tả bằng một chuỗi số nhị phân có độ dài N bit, nơi 2n-1 biểu diễn giá trị trọng số của bit thứ n, kết quả là sẽ tạo ra 2N giá trị khác nhau cho mã LZP. Do cửa sổ thường sử dụng để tính toán giá trị LZP là 3x3 do đó mỗi điểm ảnh trung tâm sẽ có 8 điểm ảnh hàng xóm xung quanh và giá trị mã LZP cho mỗi điểm ảnh sẽ nằm trong khoảng từ 0-255. Hình 3.2 mô tả cấu trúc Zigzag của một cửa sổ 3x3, biểu diễn điểm ảnh dựa vào cấu trúc Zigzag, trọng số của mẫu Zig Zag cục bộ của cửa sổ, mẫu kết cấu ban đầu và mẫu LZP của mẫu kết cấu. Sau khi các mã LZP cho mỗi điểm ảnh được tính toán theo công thức bên trên của một ảnh có kích thước

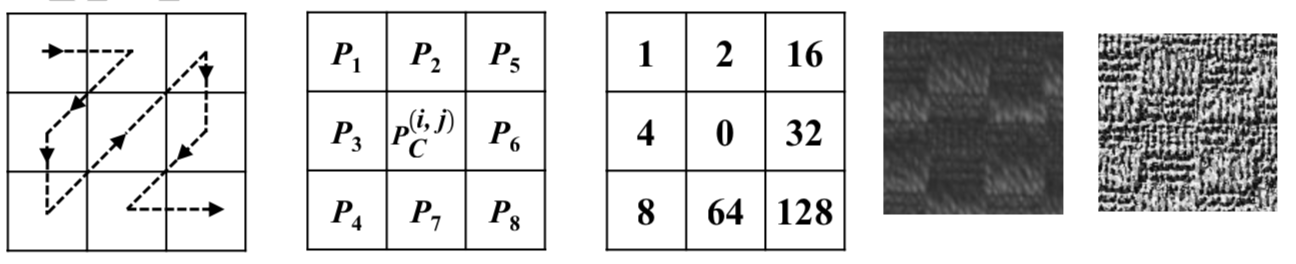
Mx x My, phân bố của các mẫu kết cấu mức xám được biểu diễn bằng việc xây dựng một phân bố rời rạc 256 cột của mã LZP được tính toán theo công thức dưới đây:

Với L là giá trị lớn nhất của mẫu LZP. Bộ mô tả đặc trưng LZP được mở rộng ra thành mẫu đồng nhất, trong đó thước đo tính đồng nhất U mã hóa số lần chuyển đổi từ 0 sang 1 và từ 1 sang 0 trong một mẫu N-bit, được định nghĩa như sau:



Ví dụ, giá trị U của chuỗi LZP 11111111 và 00000010 lần lượt là 0 và 2. Mẫu LZP đồng nhất đề cập đến sự xuất hiện đồng nhất của chuỗi đã hạn chế chuyển tiếp (U ≤ 2) trong chuỗi nhị phân tròn N-bit. Tất cả chuỗi nhị phân N-bit không đồng nhất (U> 2) được nhóm thành một danh mục "linh tinh". Ánh xạ từ LZPN đến LZPu2, N trong đó chỉ số trên “u2” biểu thị các mẫu đồng nhất có giá trị U nhiều nhất là 2 và có tổng N \* (N - 1) + 3 nhãn phân biệt. Việc tính toán LZPu2 được thực hiện bằng cách sử dụng một bảng tra cứu gồm 2N phần tử riêng biệt. Người ta quan sát thấy rằng biểu diễn thống nhất của LZP ổn định hơn (ít nhạy cảm hơn với nhiễu) và số lượng cột trở nên ít hơn đáng kể khi tính toán biểu đồ phân bố mã LZP, điều này làm cho việc so khớp hiệu quả về mặt tính toán.

Do đó trong thực tế, khi nhắc đến mã LZP người ta sẽ thường sử dụng mã đồng nhất. Nghiên cứu của đồ án này cũng sử dụng LZP để trích chọn đặc trưng như đã trình bày ở đầu chương dựa trên những cơ sở lý thuyết b và tính hiệu quả của mã LZP.



**Hình 2.3 Cấu trúc Zigzag của một cửa sổ 3x3 và mã hóa ảnh**

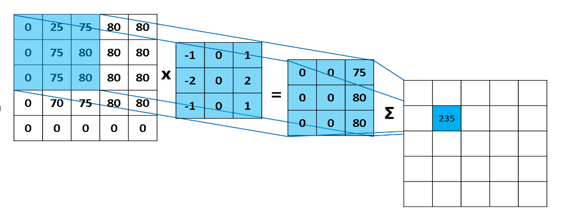
* 1. **TRÍCH CHỌN ĐẶC TRƯNG SỬ DỤNG MẠNG NƠRON HỌC SÂU (CNN)**

CNN là một cấu trúc rất phổ biến và quen thuộc trong học sâu. Ứng dụng đơn giản của nó là vào bài toán phân loại khuôn mặt, khi nhìn vào một bức ảnh sẽ biết đây là ai hoặc khi tìm kiếm từ khóa “Mèo” trong tab “Hình ảnh”, GOOGLE sẽ hiển thị ra rất nhiều ảnh có mèo trong đó. Mạng nơ ron của học sâu cũng được coi như một phương pháp trích chọn đặc trưng, điển hình là mạng nơ ron cơ bản CNN (Convolution Neural Network), thường được sử dụng với các bài toán liên quan xử lý ảnh và cho độ chính xác tốt hơn nhiều các phương pháp truyền thống.

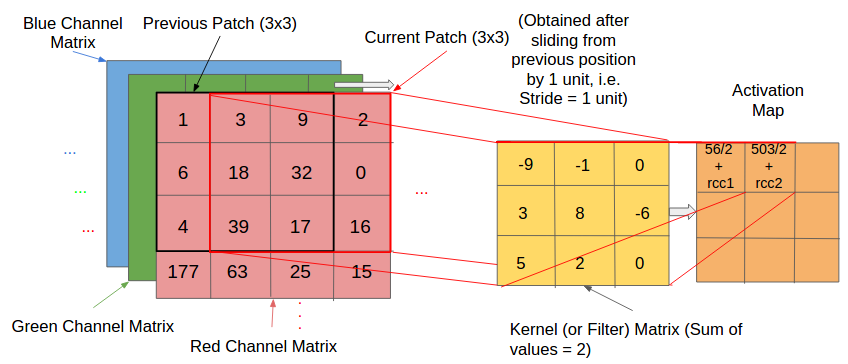
Các thành phần cơ bản của mạng CNN tiêu chuẩn:

**Convolution (tích chập):**

Tích chập được sử dụng đầu tiên trong xử lý tín hiệu số (Signal processing). Nhờ vào nguyên lý biến đổi thông tin, các nhà khoa học đã áp dụng kĩ thuật này vào xử lý ảnh và video số. Tích chập gồm 2 khái niệm khác là bộ lọc nhân chập và lớp tích chập. Trong mạng nơ ron thông thường, từ đầu vào, ta cho qua các lớp ẩn rồi ra được đầu ra. Với mô hình CNN, lớp tích chập cũng chính là lớp ẩn, khác ở chỗ, lớp tích chập là một tập các bản đồ đặc trưng và mỗi bản đồ đặc trưng này là một bản quét của đầu vào ban đầu, nhưng được trích xuất ra các đặc trưng cụ thể. Quét như thế nào thì lại dựa vào lớp tích chập hay kernel. Đây là một ma trận sẽ quét qua ma trận dữ liệu đầu vào, từ trái qua phải, trên xuống dưới, và nhân tương ứng từng giá trị của ma trận đầu vào mà ma trận kernel rồi cộng tổng lại, đưa qua các hàm kích hoạt, kết quả sẽ là một con số cụ thể, tập hợp các con số này lại là 1 ma trận nữa, chính là bản đồ đặc trưng. Hãy nhìn vào ví dụ sau. Tôi có 1 ma trận đầu vào và 1 kernel:



**Hình 2.4 Thao tác nhân tích chập**

****

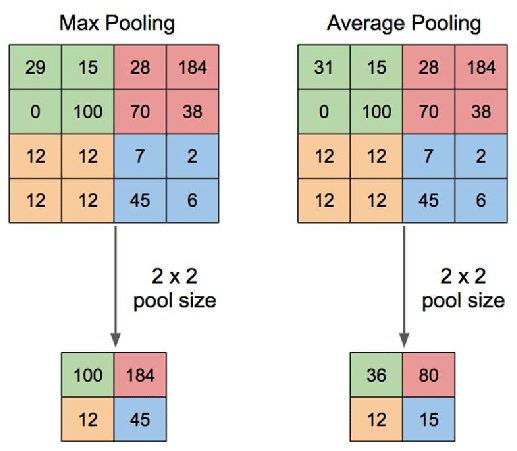
**Hình 2.5 Biểu đồ đặc trưng nhận được sau khi nhân tích chập.**

Nhân lần lượt dịch kernel từ trái quả phải từ trên xuống, ta sẽ thu được một ma trận đặc trưng đầy đủ. Và đó cũng chính là nhiệm vụ của CNN, nó sẽ tự động tìm các kernel, tự dò ra các đặc trưng. Ta có thể thay đổi Stride để thay đổi khoảng cách giữa 2 kernel khi quét, chọn stride và kích thước của kernel càng lớn thì kích thước của bản đồ đặc trưng càng nhỏ.

**Pooling:**

Lớp chứa hay lớp tổng hợp (Pooling layer): Ngoài các lớp tích chập vừa mô tả, mạng nơron tích chập cũng chứa các lớp pooling. Lớp pooling thường được sử dụng ngay sau lớp tích chập. Những gì các lớp pooling làm là đơn giản hóa các thông tin ở đầu ra từ các lớp tích chập. Điều này giúp giảm kích thước của biểu đồ đặc trưng.

Có 2 loại pooling layer phổ biến là: max pooling và average pooling:



**Hình 2.6 Minh họa nguyên lý hoạt động của lớp Pooling.**

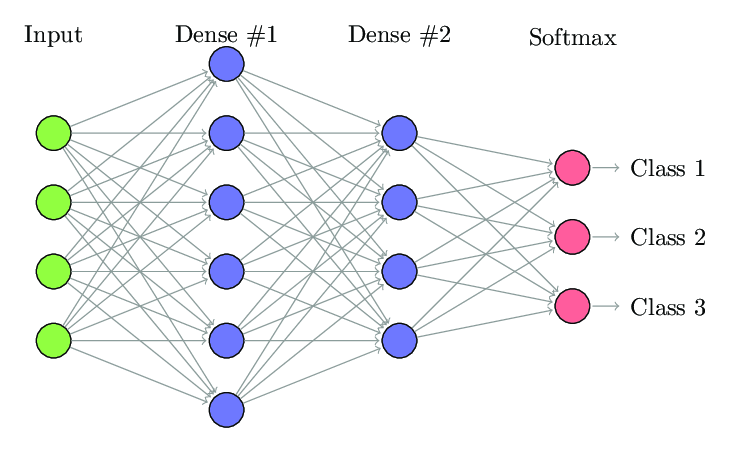
Mục đích của pooling rất đơn giản, nó làm giảm số siêu tham số mà ta cần phải tính toán, từ đó giảm thời gian tính toán, tránh overfitting. Loại pooling ta thường gặp nhất là max pooling, lấy giá trị lớn nhất trong một cửa sổ pooling. Pooling hoạt động gần giống với tích chập, nó cũng có 1 cửa sổ trượt gọi là pooling window, cửa sổ này trượt qua từng giá trị của ma trận dữ liệu đầu vào, chọn ra một giá trị từ các giá trị nằm trong cửa sổ trượt (với max pooling ta sẽ lấy giá trị lớn nhất). Hãy cùng nhìn vào ví dụ sau, Tôi chọn pooling window có kích thước là 2 \* 2, stride = 2 để đảm bảo không trùng nhau, và áp dụng max pooling:

**Fully Connected**

Giống trong mạng nơ ron đã nghiên cứu ở mục trước thì mỗi lớp ẩn được gọi là kết nối đầy đủ - fully connected. Thường sau lớp kết nối đầy đủ sẽ là 2 lớp kết nối đầy đủ, 1 lớp để tập hợp các lớp đặc trưng mà ta đã tìm ra, chuyển đổi dữ liệu từ 3D, hoặc 2D thành 1D, tức chỉ còn là 1 vector. Còn 1 lớp nữa là đầu ra, số nơron của lớp này phụ thuộc vào số đầu ra mà ta muốn tìm ra. Giả sử với tập dữ liệu MNIST chẳng hạn, ta có tập các số viết tay từ 0 -> 9. Vậy đầu ra sẽ có số nơ ron là 10.

Sau khi ảnh được truyền qua nhiều lớp tích chập và lớp pooling thì mô hình đã học được tương đối các đặc điểm của ảnh (ví dụ mắt, mũi, khuôn mặt) thì tensor của đầu ra của lớp cuối cùng, kích thước H\*W\*D, sẽ được chuyển về 1 vector kích thước (H\*W\*D)

￼

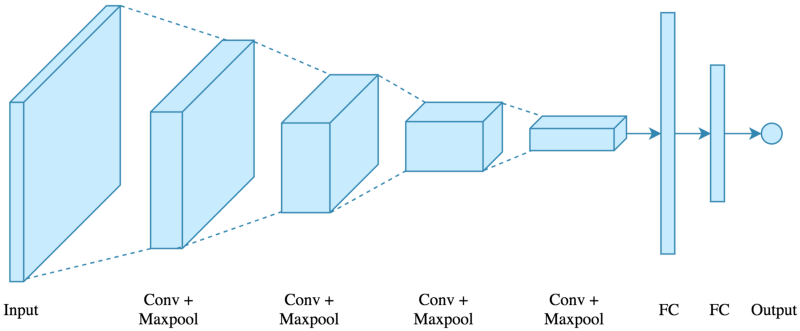


**Hình 2.7 Minh họa lớp kết nối đầy đủ.**

Sau đó ta dùng các lớp kết nối đầy đủ để kết hợp các đặc điểm của ảnh để ra được đầu ra của mô hình.

1. ￼

**Ví dụ cấu trúc một mạng CNN tiêu chuẩn:**



**Hình 2.8 Kiến trúc của một mạng CNN tiêu chuẩn.**

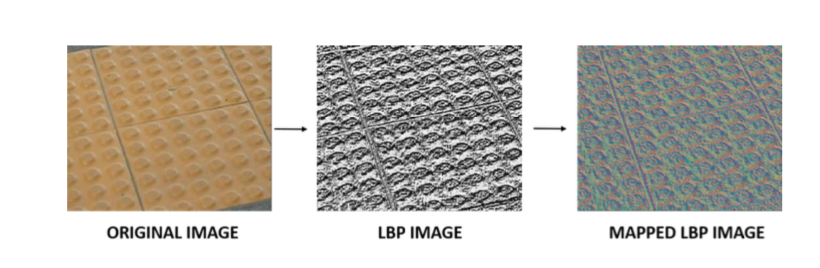
Những kết quả hứa hẹn của mô hình CNN trong bài toán phân loại hình ảnh, cũng như sử dụng như là một thuật toán trích chọn tổng hợp các đặc điểm hình ảnh thành vector đặc trưng đầu ra với số chiều xác định. Sau đó sử dụng các đặc trưng đó như là một tập vector mô tả đặc trưng hình ảnh để kết hợp các bộ phân loại khác nhau.

Bên cạnh đó tận dụng quá trình huấn luyện trên các tập dữ liệu kiểm tra, đánh giá với số lượng dữ liệu lớn ImageNet học chuyển đổi (transfer learning) được áp dụng tốt hơn giúp việc huấn luyện nhanh hơn và hiệu quả hơn, nâng cao hiệu quả phân loại.

* 1. **KẾT HỢP BỘ TRÍCH XUẤT ĐẶC TRƯNG CỤC BỘ VÀ MẠNG NƠRON HỌC SÂU (CNN)**

Phương pháp trích xuất đặc trưng cục bộ và trích xuất đặc trưng dựa trên mạng CNN học sâu đều có những ưu điểm riêng. Nếu như đặc trưng cục bộ điểm mạnh là đơn giản, chi phí tính toán thấp, hiệu quả trong phạm vi giới hạn dữ liệu, trong khi đó trích xuất đặc trưng CNN bị giới hạn về dữ liệu cần có một lượng lớn dữ liệu để huấn luyện một mô hình CNN có thể trích xuất đặc trưng hình ảnh một cách khái quát. Ngoài ra bộ phân loại thành phần trong CNN rất hiệu quả trong việc phân loại các đặc trưng nhằm đưa ra kết quả trong bài toán. Chính vì ￼vậy việc kết hợp bộ trích xuất đặc trưng cục bộ làm đầu vào cho một mạng CNN có đầy đủ bộ phân loại là cần thiết để cân bằng các vấn đề nêu trên.

Tuy nhiên, do cách tính LZP không phù hợp làm đầu vào trực tiếp của mạng CNN, đầu ra của LZP cần phải được biến đổi thông qua một phương pháp ánh xạ các mã LZP. Phương pháp này ánh xạ các mã LZP tới các điểm trong không gian số liệu 3D trong đó khoảng cách Euclidian xấp xỉ khoảng cách giữa các mã LZP. Sau khi chuyển đổi mã LZP, chúng có thể được tính trung bình với nhau bằng cách sử dụng các phép toán tích chập trong các mô hình CNN. Minh họa của ảnh trưng được ánh xạ được trình bày trong hình 3.3.



**Hình 2.9 Ảnh mẫu qua các bước xử lý**

* 1. **TỔNG KẾT CHƯƠNG**

Trong chương 2 chúng tôi đã trình bày những cơ sở lý thuyết về các kỹ thuật trích chọn đặc trưng cho hình ảnh kết cấu dựa trên những cơ sở lý thuyết về đặc trưng cục bộ, hay mạng học sâu và kết hợp cả hai phương pháp trên trong việc trích chọn đặc trưng cho hình ảnh kết cấu. Trong chương 3,4 chúng tôi sẽ triển khai các cơ sở lý thuyết và xây dựng ứng dụng phân loại hình ảnh kết cấu.

# CHƯƠNG 3. ĐỀ XUẤT MÔ HÌNH NHẬN DẠNG ẢNH KẾT CẤU

Trong chương này, tôi sẽ trình bày một đề xuất trong việc trích xuất đặc trưng cho hình ảnh nhằm phục vụ phân loại hình ảnh kết cấu. Các phương pháp này được tiếp cận theo hai hướng học máy và học sâu.

* 1. **HƯỚNG TIẾP CẬN DỰA TRÊN HỌC MÁY**

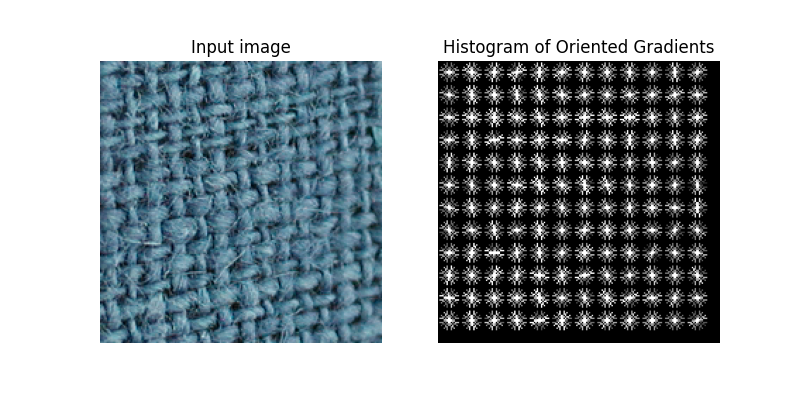
Trong hầu hết các bài toán phân loại dựa trên học máy, các kĩ thuật rút trích đặc trưng có vai trò hết sức quan trọng, ảnh hưởng lớn đến kết quả phân loại sau cùng. Chính vì điều này trong một thời gian dài, các nhà khoa học dữ liệu luôn tìm kiếm phương án trích xuất đặc trưng cho nhiều loại dữ liệu với nhiều loại đặc điểm khác nhau như: màu sắc sẽ được thể hiện qua biểu đồ histogram, đặc trưng hình dạng dựa theo sự biến đổi mức xám để xác định các vị trí biên như HOG, hay kỹ thuật trích xuất đặc trưng kết cấu GLCM.

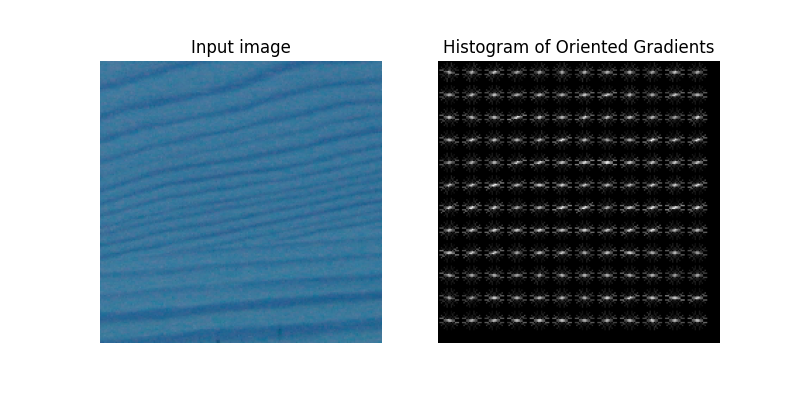
Trong khuôn khổ đề tài này, về phân loại hình ảnh kết cấu được cấu thành nên các loại vật liệu khác nhau, chúng tôi sẽ trình bày những đề xuất giúp thích hợp nhất trong việc trích rút đặc trưng kết cấu sau đó sử dụng một bộ phân loại học máy để phân loại các đặc trưng đó trong tập dữ liệu của chúng tôi.

* + 1. **Phương thức miêu tả đặc trưng kết cấu cục bộ**

Kết cấu có mặt ở khắp nơi trong các hình ảnh tự nhiên mạng cơ bản đặc trưng của sự xuất hiện của tất cả các bề mặt tự nhiên. Phân loại kết cấu là một trong những vấn đề tích cực và đầy thách thức đối với toàn bộ các nhà khoa học dữ liệu.

Điều đặc biệt của hình ảnh kết cấu so với các bài toán phân loại các hình ảnh khác ở điểm hình ảnh kết cấu thể hiện các chi tiết kết cấu nhỏ lẻ theo từng khu vực nhỏ lặp lại và đều nhau trên toàn bộ hình ảnh. Theo những phương pháp truyền thống được áp dụng cho bài toán phân loại bao gồm việc sử dụng các kĩ thuật rút trích đặc trưng: màu sắc, hình dạng … Những phương pháp này thường mô tả dữ liệu toàn cục sử dụng cái nhìn tổng quát về đối tượng để đưa ra bộ các thuộc tính của đối tượng. Chính vì vậy phương pháp này không hiệu quả trong việc phân loại hình ảnh kết cấu.



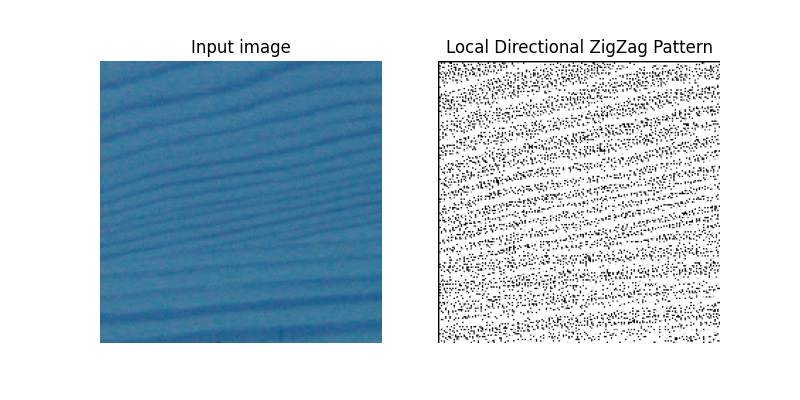


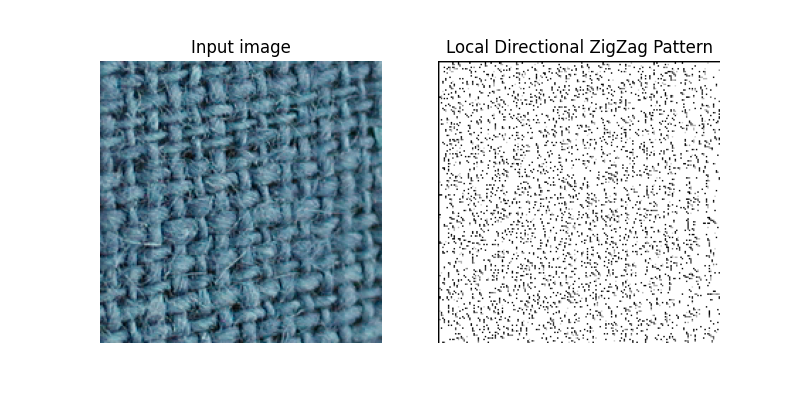


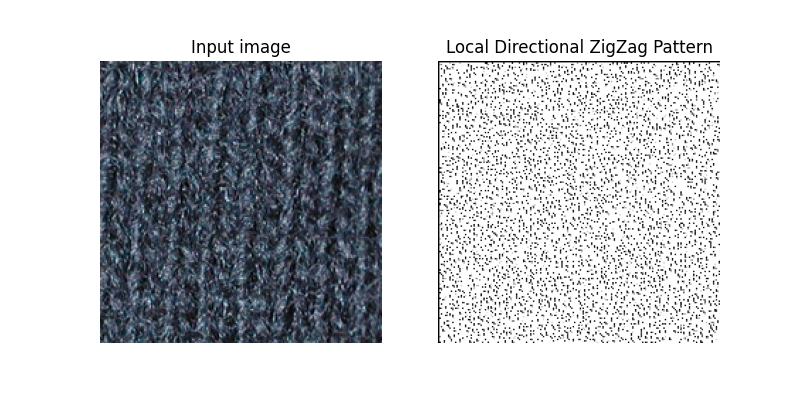
**Hình 3.1 Hình ảnh sau khi áp dụng phân tích HOG**

Trong các hình 3.1 đều được áp dụng kĩ thuật rút trích đặc trưng HOG cho hình ảnh kết cấu của 3 loại vật liệu khác nhau vải lanh, gỗ, vải thông thường. Kết quả đặc trưng HOG cho ra của 3 hình đều có sự tương đồng rõ rệt. Khi đưa chúng qua mô hình quân loại gần như không thấy được sự khác biệt giữa chúng khiến mô hình đạt hiệu xuất dự đoán kém ở mức 40%.

Biểu diễn kết cấu bằng các phương pháp cục bộ (local descriptor) là các tiếp cận rất phổ biển tiêu biểu như LBP và cho kết quả tương đối ổn định trên hai phương diện thời gian và độ chính xác. Tuy nhiên LBP gặp vấn đề đối với loại kết cấu phức tạp và biến đổi mạnh mẽ. Chính lẽ đó cần có một phương pháp để có thể rút trích các đặc trưng kết cấu của hình ảnh kết cấu biến đổi mạnh mẽ và liên lục. Trong tài liệu này chúng tôi sẽ trình bày phương pháp LDZP (Local Directional ZigZag Pattern). Kết quả thực nghiệm cho thấy phương pháp này cải thiện lên đến 50% độ chính xác so với các phương pháp trích xuất đặc trưng khác dựa trên màu sắc và hình dạng.







**Hình 3.2 Ảnh gốc và ảnh được mã hóa bằng LDZP**

Hình ảnh kết cấu sau khi được áp dụng kĩ thuật trên đã thể hiện được rõ hơn những đặc điểm về kết cấu cũng như mật độ kết nối của vật liệu, cũng như vân của vật liệu. Từ quan sát trên khi áp dụng bộ phân loại nó cho những kết quả khả quan hơn so với cách trích xuất đặc trưng theo hình dạng và màu sắc.

* + 1. **Xây dựng đặc trưng dựa trên biểu đồ Histogram**

Biểu đồ Histogram (trong xử lý ảnh gọi là biểu đồ mức xám) là một loại biểu đồ có ứng dụng rộng rãi trong thị giác máy tính và thống kê. Trục hoành mô tả phạm vi và quy mô của phạm vi mức xám (giá trị của các điểm ảnh) và trục tung hiển thị số lượng điểm ảnh với các cường độ sáng khác nhau, tức là tần suất xuất hiện của các giá trị điểm ảnh. Biểu đồ cho phép trực quan hóa dữ liệu số bằng cách chỉ ra số điểm ảnh nằm trong một phạm vi giá trị, được gọi là một lớp hoặc một bin. Tần suất của giá trị điểm ảnh rơi vào mỗi lớp được mô tả bằng cách sử dụng một thanh. Biểu đồ không thay đổi đối với các thao tác hình ảnh như xoay, dịch nhưng chúng cũng thay đổi một chút với sự thay đổi về tỷ lệ, góc xem hoặc với khuôn mặt trang điểm, đeo kính, đeo khẩu trang. Biểu đồ mức xám hoạt động kém trong các điều kiện hình ảnh hoặc ánh sáng khác nhau. Chúng cũng không hiệu quả trong việc phân biệt các hình ảnh khác nhau có sự phân bố màu sắc tương tự và chúng có số chiều cao.

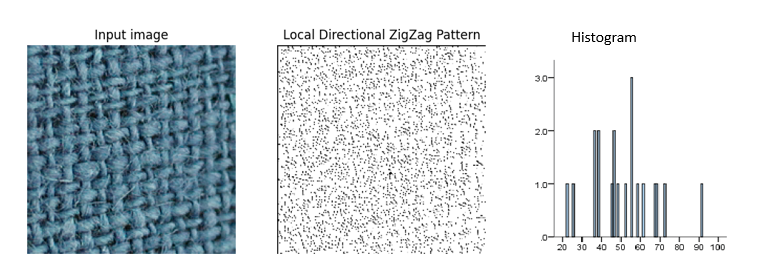
Cho ảnh kết cấu *IMxN*. Sau khi biểu diễn đặc trưng bằng LDZP ta thu được ảnh *ILDZP*. Ảnh thu được , *ILDZP*, sẽ được biểu diễn bằng một lược đồ mức xám theo công thức dưới đây.

, (3.1)

, (3.2)

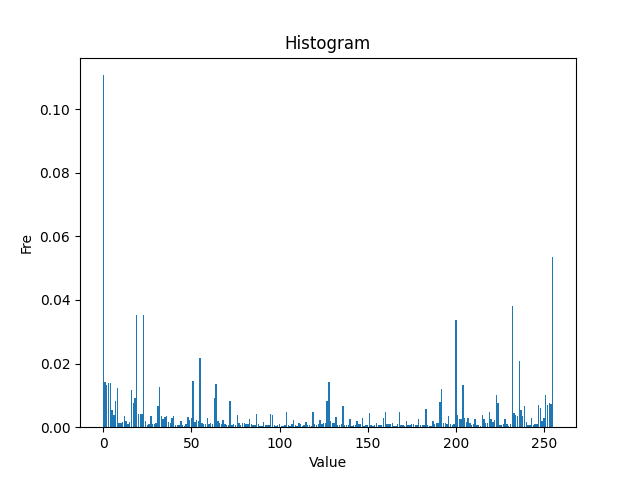
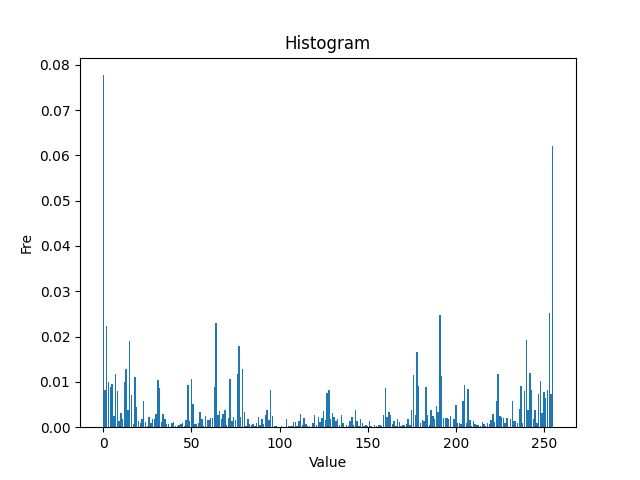
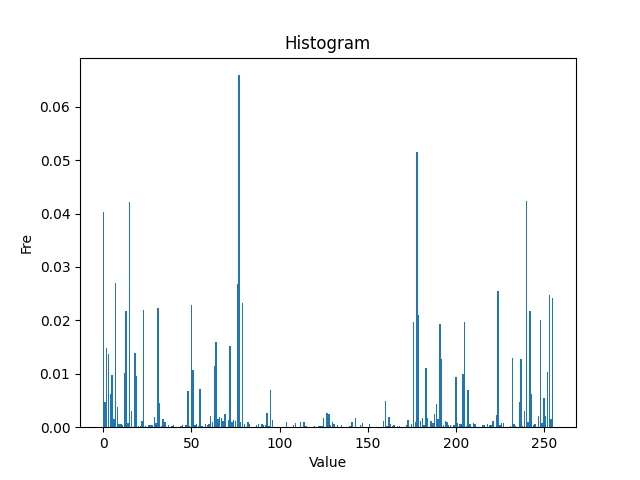
Việc tạo biểu đồ mức xám thể hiện sự phân bố các mức xám trong hình ảnh. Các giá trị này sẽ không ảnh hưởng khi xoay ảnh hay dịch chuyển hình ảnh. Từ đó không đánh mất đi đặc điểm của hình ảnh kết cấu.

Sơ đồ mô tả quá trình trích xuất đặc trưng của hình ảnh kết cấu:



**Hình 3.3 ảnh gốc, ảnh mã hóa LDZP và biểu đồ histogram của hình ảnh mã**

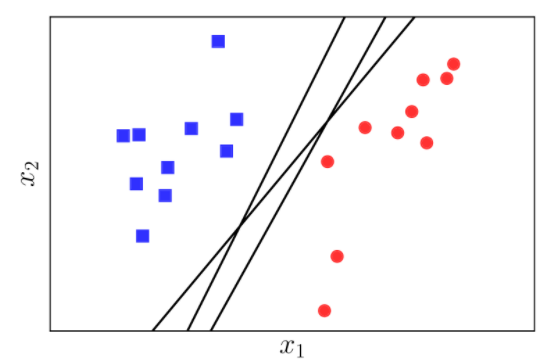
**hóa LDZP**



**Hình 3.4 Hình ảnh gốc và biểu đồ histogram của ảnh mã hóa zigzag từ ảnh gốc**

* + 1. **Mô hình phân loại kết cấu dựa trên đặc trưng Histogram**

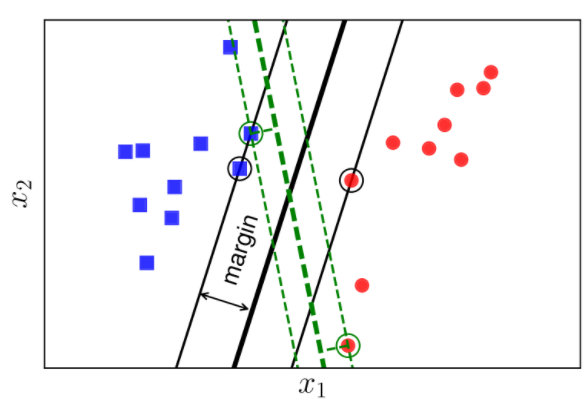
Thông thường sau khi xây dựng đặc trưng histogram của hình ảnh kết cấu việc đánh giá phân loại hình ảnh kết được dựa trên phương pháp tìm kiếm lân cận KNN. Các thước đo đánh giá sự tương đồng của hai biểu đồ histogram sử dụng khoảng cách L2 giữa hai vector. Phương pháp này ưu điểm đó là sự đơn giản và tương đối hiệu quả. Tuy nhiên nó gặp một số vấn đề về tốc độ nếu trong cơ sở dữ liệu có nhiều hình ảnh, việc tìm kiếm sẽ tiêu tốn thời gian. Bên cạnh đó mô hình phân loại trên chưa vận dụng được hết các đặc điểm để đưa ra sự phân loại cho các hình ảnh kết cấu. Chính lẽ đó chúng tôi sử dụng một bộ phân loại hiệu quả hơn đó là SVM (Support Vector Machine).



**Hình 3.5 Các mặt phân cách hai lớp sử dụng đường thẳng**

Tư tưởng chính:

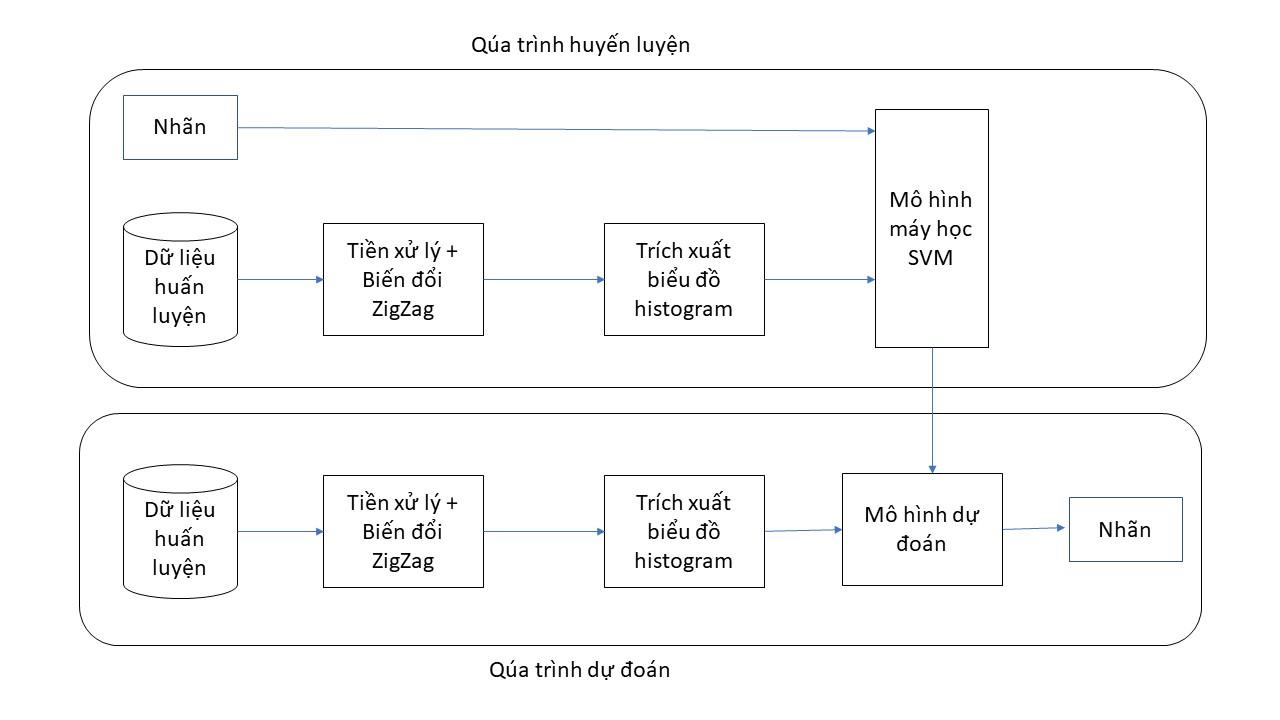
Luôn tồn tại nhiều đường thẳng giúp chia mặt phẳng như hình trên được thành hai phần. Điều này giúp cho phân biệt giá trị tròn màu đỏ và hình vuông màu xanh. Nhưng một câu hỏi đặt ra: trong vô số các mặt phân chia đó, đâu là mặt phân chia tốt nhất theo một tiêu chuẩn nào đó? Trong ba đường thẳng minh họa trong Hình 3.8 phía trên, có hai đường thẳng khá lệch về phía class hình tròn đỏ. Nếu mà đường boundary nghiêng về một vùng lãnh thổ thì gần nó chưa đủ tốt do rất dễ bị sai lệch khi dữ liệu thử nghiệm nó cách một chút xíu so với các đường phân cách. Vì vậy để công bằng, trong vô số các đường boundary thì ta sẽ chọn ra đường nào mà khoảng cách từ điểm gần nhất của mỗi class tới đường phân chia là như nhau. Khoảng cách như nhau này được gọi là margin (lề).



**Hình 3.6 Các đường boundary cho margin bằng nhau**

* + 1. **Sơ đồ tổng quát hệ thống phân loại kết cấu**

Phương pháp phân loại hình ảnh kết cấu được áp dụng kĩ thuật biến đổi hình ảnh zigzag sau đó mô tả hóa đặc trưng dữ liệu bằng biểu đồ histogram và sau đó được phân lớp sử dụng SVM (Support Vector Machine). Hoạt động của quá trình được mô tả như dưới đây:



**Hình 3.7 Sơ đồ xây dựng hệ thống phân loại hình ảnh kết cấu**

**Pha huấn luyện:**

Đầu vào: Tập dữ liệu huấn luyện kèm nhãn (m ảnh)

Đầu ra: một mô hình dự đoán (SVM model)

Các bước thực hiện:

Bước 1: Hình ảnh tiền xử lý bao gồm: resize kích thước hình ảnh 112 x 112, chuyển ảnh về ảnh mức xám. Sau đó hình ảnh xám được biến đổi dựa trên kỹ thuật Zigzag.

for *k* ← 1 to *m* do

*zigk* ← ZigZag(*I*k)

Bước 2: Tính toán biểu đồ histogram từ hình ảnh sau phép biến đổi dựa trên zigzag.

Bước 3: Huấn luyện mô hình SVM phân loại với các vector đặc trưng đã được xây dựng từ bước 2 kèm nhãn ban đầu.

**Pha đánh giá**

Bước 1: Hình ảnh tiền xử lý bao gồm: resize kích thước hình ảnh 112 x 112, chuyển ảnh về ảnh mức xám. Sau đó hình ảnh xám được biến đổi dựa trên kỹ thuật Zigzag.

for *k* ← 1 to *m* do

*zigk* ← ZigZag(*I*k)

Bước 2: Tính toán biểu đồ histogram từ hình ảnh sau phép biến đổi dựa trên Zigzag.

Bước 3: Hình ảnh được đưa vào bộ phân loại để đưa ra nhãn dự đoán.

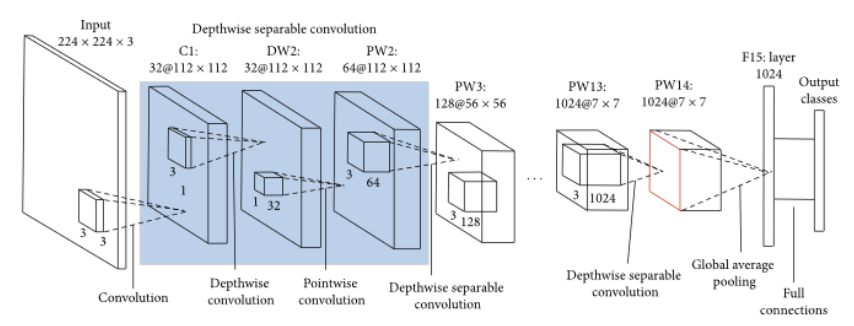
* 1. **PHƯƠNG PHÁP TIẾP CẬN DỰA TRÊN HỌC SÂU**

Trong vài năm trở lại đây học sâu đang là một trong những trào lưu được ưa chuộng trong vận dụng giải quyết vấn đề trong trí tuệ nhân tạo. Các giá trị của học sâu luôn có được những vận dụng giá trị trong cuộc sống. Sự vượt trội của học sâu so với học máy truyền thống trên dữ liệu hình ảnh lớn được chứng minh trong nhiều nghiên cứu. Trong khuôn khổ đề tài này chúng tôi vận dụng mô hình học sâu dựa trên sự mô phỏng neuron thần kinh trong não bổ để phân loại hình ảnh kết cấu.

* + 1. **Mạng nơ-ron tích chập CNN**

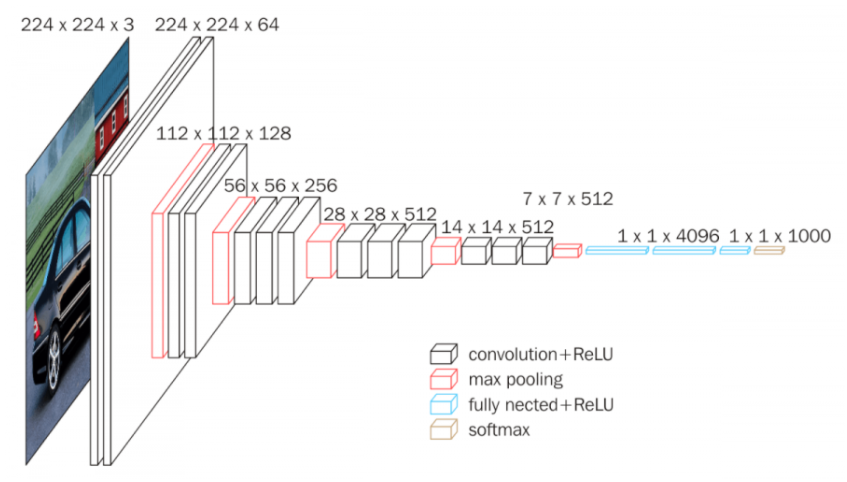
Mạng nơ-ron tích chập tìm hiểu đặc trưng của hình ảnh thông quá các lớp tích chập trên hình ảnh. Từ nhiều lớp tích chập và trên hình ảnh mô hình có thể nắm bắt các thông tin đặc trưng cơ bản nhất của hình ảnh. Những đặc trưng ấy được học hỏi và rèn luyện thông qua kỹ thuật tối ưu. Từ đó mạng CNN có thể nắm bắt nhiều thông tin hơn của hình ảnh và đưa ra một vector đặc trưng có chiều sâu bao hàm nhiều loại đặc trưng khác nhau trong hình ảnh như: màu sắc, hình dạng, kết cấu … Từ đó khi áp dụng một bộ phân loại như MLP hay SVM như phần trên kết quả được cải thiện rõ rệt so với các tiếp cận dựa trên học máy.

Trong khuôn khổ đề tài này chúng tôi đề xuất sử dụng ba mô hình học sâu trong việc xây dựng mô hình phân loại với các trường hợp sử dụng khác nhau: di động kiến trúc MobileNet, máy tính phổ thông với kiến trúc VGG16, máy tính chuyên dụng với kiến trúc ResNet.



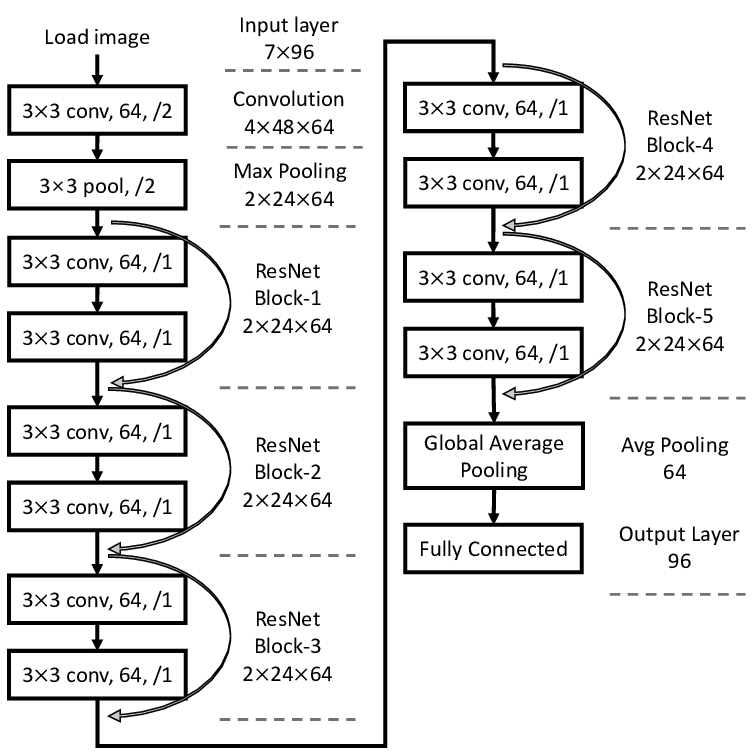
**Hình 3.8 Sơ đồ kiến trúc mô hình MobileNet**

Mô hình mobilenet được xây dựng nhằm phục vụ đưa các mô hình học sâu có thể hoạt động được với các thiết bị di động. Chính vì vậy nó có những cải tiến giúp mô hình nhẹ hơn đó là tách kích thước chiều sâu khỏi chiều ngang (chiều rộng \* chiều cao) cho chúng ta phép chập phân tách theo chiều sâu trong khi chúng ta thực hiện phép tích chập chiều sâu. Sau đó, chúng tôi sử dụng bộ lọc 1 \* 1 để bao phủ kích thước chiều sâu.



**Hình 3.9 Kiến trúc mô hình VGG16**

VGG16 là một mô hình mạng nơ-ron tích chập được đề xuất bởi K. Simonyan và A. Zisserman từ Đại học Oxford trong bài báo “Mạng lưới kết hợp rất sâu để nhận dạng hình ảnh quy mô lớn”. Mô hình đạt được độ chính xác trong bài kiểm tra top 5 là 92,7% trong ImageNet,là một tập dữ liệu của hơn 14 triệu hình ảnh thuộc 1000 lớp. Đây là một trong những mô hình nổi tiếng được nộp cho ILSVRC-2014 . Nó cải tiến hơn AlexNet bằng cách thay thế các bộ lọc có kích thước hạt nhân lớn (11 và 5 ở lớp chập đầu tiên và thứ hai, tương ứng) bằng nhiều bộ lọc kích thước hạt nhân 3 × 3 lần lượt. VGG16 đã được đào tạo trong nhiều tuần và đang sử dụng GPU NVIDIA Titan Black.



**Hình 3.10 Kiến trúc mô hình ResNet-12**

ResNet (Residual Network) được giới thiệu đến công chúng vào năm 2015 và thậm chí đã giành được vị trí thứ 1 trong cuộc thi ILSVRC 2015 với tỉ lệ lỗi top 5 chỉ 3.57%. Không những thế nó còn đứng vị trí đầu tiên trong cuộc thi ILSVRC and COCO 2015 với ImageNet Detection, ImageNet localization, Coco detection và Coco segmentation.Hiện tại thì có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152,...Với tên là ResNet theo sau là một số chỉ kiến trúc ResNet với số lớp nhất định. Điều đặc biệt của ResNet đã giải quyết đó là vấn đề học tập sau nhiều lớp của mô hình không còn hiệu quả. Cải tiến của nó giúp cho hình thành mô hình mạng nơron siêu sâu lên đến hàng trăm lớp. Từ đó giúp phát hiện được siêu chi tiết về đối tượng quan sát.

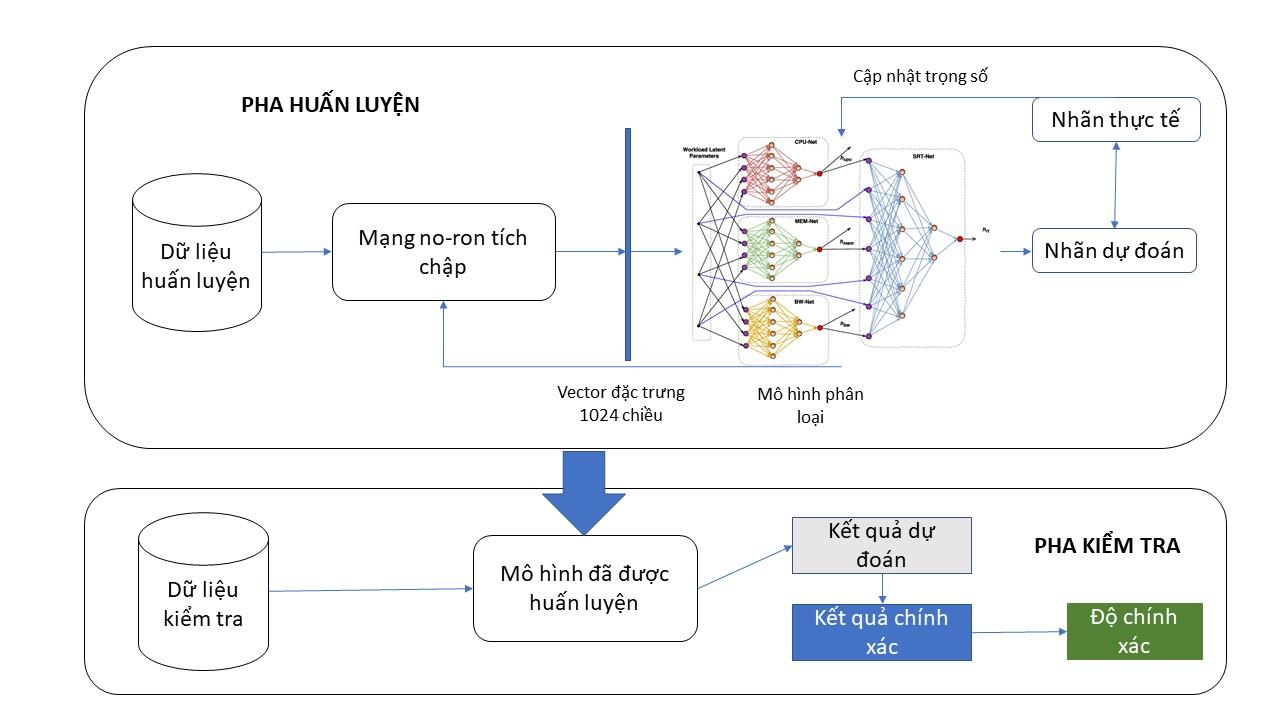
Bảng bên dưới hiển thị thông tin về các mô hình học sâu chúng tôi sử dụng trong việc xây dựng mô hình phân loại kết cấu. Chúng tôi huấn luyện các mô hình trên Google Colab sau đó được chạy dự đoán trên thiết bị máy tính phổ thông: Dell Vostro 3568 i5 7200U.

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Số lượng tham số | Dung lượng | Thời gian dự đoán 1 mẫu (i5 7200U) |
| MobileNet | 3.8 triệu | 14.8 MB | 0.46s |
| VGG16 | 15.1 triệu | 57.4 MB | 2.21s |
| ResNet50 | 24.7 triệu | 94.7 MB | 4.16s |

Cả ba mô hình đều có những kết quả tốt trên tập dữ liệu kiểm tra và đánh giá của chúng tôi. Trong quá trình thực hiện vận dụng mô hình trong bài toán phân loại chúng tôi điều chỉnh đầu vào mô hình hình ảnh kết cấu kích thước 112 x 112 điều này khiến chúng tôi tăng tốc độ huấn luyện trong khi không có sự ảnh hưởng hiệu năng với hình ảnh 224 x 224 do đặc điểm lặp lại cấu trúc của dữ liệu kết cấu.

Bên cạnh đó để có thể nâng cao kết quả huấn luyện và phân loại chúng tôi sử dụng các kiến trúc mạng trên đã huấn luyện sắn với tập dữ liệu ImageNet 1000, và đóng băng một số lớp của mô hình trong quá hình huấn luyện.

* + 1. **Sơ đồ tổng quát hệ thống phân loại kết cấu sử dụng học sâu**



**Hình 3.11 Sơ đồ xây dựng mô hình phân loại ảnh kết cấu sử dụng học sâu**

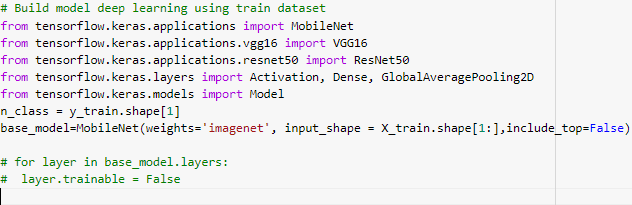
* Pha huấn luyện

Chúng tôi chia tập dữ liệu thành hai phần sử dụng để huấn luyện mô hình và để đánh giá độ chính xác của mô hình. Trong pha huấn luyện chúng tôi thực hiện các bước:

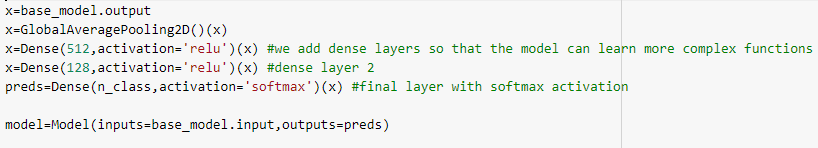
* Bước 1: Hình ảnh được gán nhãn các hình ảnh đưa về cùng một kích thước 112 x 112 tương ứng chiều rộng và chiều cao. Các nhãn của hình ảnh được mã hóa sử dụng phương pháp one hot vector để có dạng một vector đầu ra có dạng (1,n\_class) trong đó n\_class là số lớp cần phân loại.



* Bước 2: Mỗi hình ảnh chúng tôi cho đi qua mô hình mạng no-ron MobileNet, VGG16 hay ResNet50 để tạo ra vector 1024 chiều.



* Bước 3: Vector 1024 được sử dụng như là đặc trưng của hình ảnh sẽ được đưa ra qua một bộ phân loại tuyến tính.



* Bước 4: Kết quả dự đoán nhận được, được so sánh với kết quả sự thật và sử dụng hàm loss focal loss đánh giá. Từ đó cập nhật lại các trọng số mô hình bằng thuật toán lan truyền ngược.



* Pha kiểm tra

Trong pha này chúng tôi cho phần dữ liệu kiểm tra qua mô hình và đánh giá độ chính xác mô hình trên thang đo accuracy hoặc F1-score.

**3.3 PHƯƠNG PHÁP BIẾN ĐỔI HÌNH ẢNH CỤC BỘ VÀ MẠNG CNN**

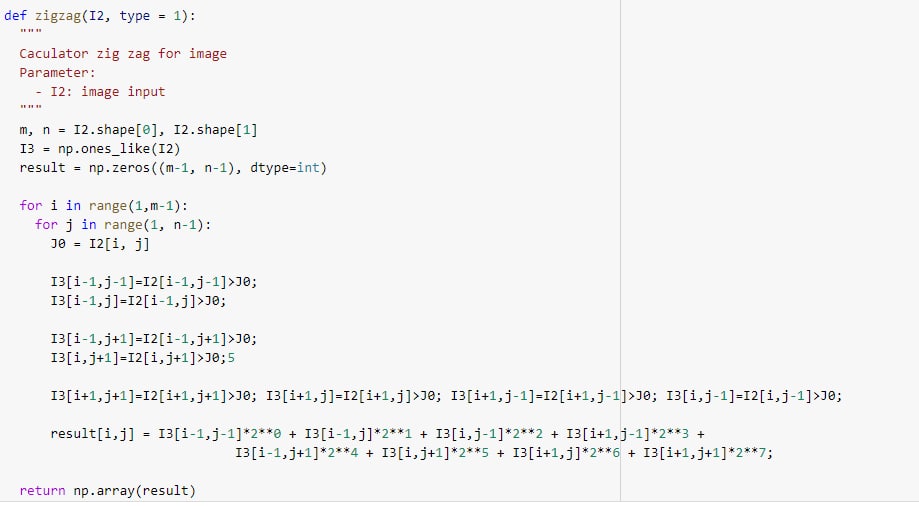
Trong phần trước chúng tôi sử dụng hình ảnh mặc đình là đầu vào cho mạng CNN.

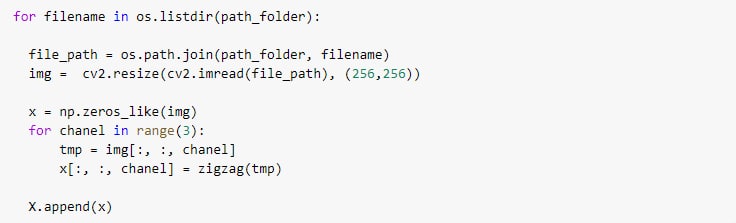
Chúng tôi nhận thấy một vài ưu điểm của việc sử dụng hình ảnh cục bộ đó là hình ảnh được biến đổi sáng một miền dữ liệu mới như ZigZag hay LBP. Miền dữ liệu đó là sự biễu diễn sự phụ thuộc các pixel vào các pixel xung quanh nó. Dẫn đến một hình ảnh đầu ra phụ thuộc thay vì riêng lẽ. Chúng tôi thực hiện nó như là một bước tiền xử lý cho đầu vào nhận dạng bằng mạng phân loại CNN. Nhằm giúp CNN có thể phân loại được hình ảnh kết cấu từ những đặc điểm mới.

Các bước thực hiện:

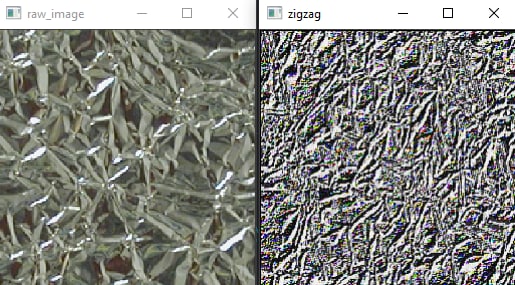
* **Giai đoạn 1: Biến đổi hình ảnh cục bộ sử dụng biến đổi ZigZag**

Mỗi hình ảnh được tách từng kênh riêng lẽ R, G, B sau đó mỗi kênh được biến đổi cục bộ ZigZag, sau đó cuối cùng được tổng hợp lại thành một hình ảnh cục bộ mới.





Một vài kết quả so sánh hình ảnh trước và sau khi biến đổi cục bộ của hình ảnh thông qua biến đổi cục bộ ZigZag.

 (a)

 (b)

Hình 3.12 (a) (b) Minh hoa biến đổi cục bộ sử dụng ZigZag

* **Giai đoạn 2: Sử dụng hình ảnh biến đổi cục bộ là hình ảnh đầu vào mạng CNN**

Hình ảnh sau khi bến đổi cục bộ được resize về kích thuớc 224 x 224 sau đó được đưa vào mạng CNN để phân loại tương tư như với hình ảnh mặc định trong phần 3.2

**3.4 KẾT CHƯƠNG**

Trong chương 3 tôi đã trình bày những phương án tiếp cận trong việc xây dựng mô hình phân loại hình ảnh kết cấu dựa trên những cách tiếp cận dựa trên học máy và học sâu. Trong chương tiếp theo chúng tôi vận dụng các phương án đề xuất để thực nghiệm các phương án từ đó có những nhận xét trong bài toán phân loại hình ảnh kết cấu.

# CHƯƠNG 4. XÂY DỰNG CHƯƠNG TRÌNH VÀ THỬ NGHIỆM

Chương này trình bày các kết quả thực nghiệm và các thảo luận đánh giá, so sánh giữa phương pháp đề xuất và các phương pháp cũ. Bên cạnh đó, một số giao diện chính của chương trình thực nghiệm cũng được giới thiệu tại phần cuối của chương. Chương này sẽ trình bày về các kết quả thực nghiệm và các thảo luận đánh giá, so sánh giữa các mô hình đã đề xuất.

* 1. **CƠ SỞ DỮ LIỆU DÙNG CHO THỰC NGHIỆM**

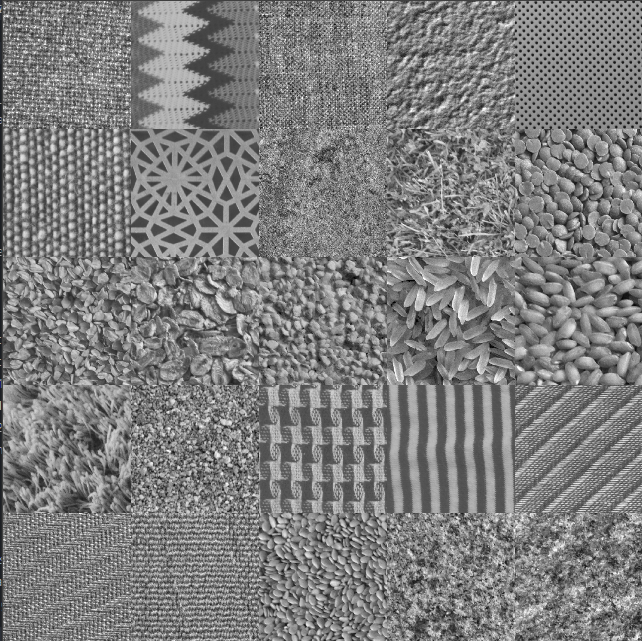
Trong mục này chúng tôi sử dụng 3 tập cơ sở dữ liệu về kết cấu để đánh giá hiệu quả của phương pháp so với cách tiếp cận phân loại hình ảnh kết cấu trực tiếp. Ba tập dữ liệu đó là KTH, Kyberge, và UIUC.

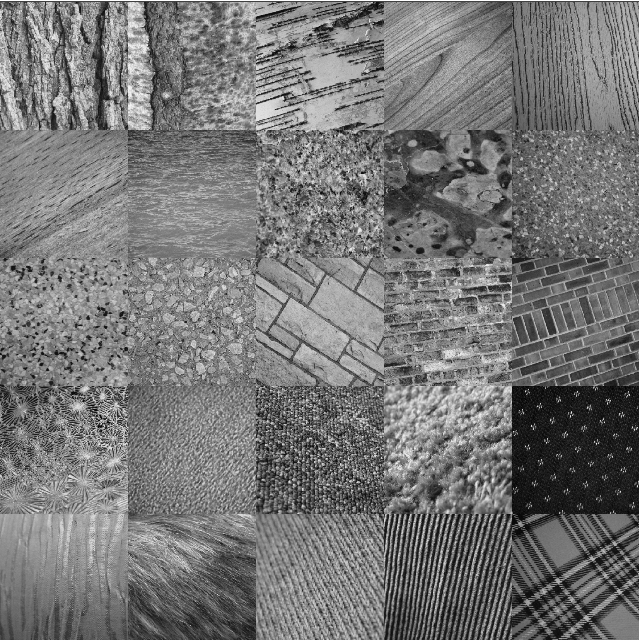
Các thông số chi tiết về các tập dữ liệu được mô tả trong bảng bên dưới:

Bảng 4.1 Các thông số chi tiết về các tập dữ liệu cho thực nghiệm

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| STT | Tên dataset | Số lượng mẫu | Số lượng huấn luyện | Số lượng kiểm tra | Số loại kết cấu | Kích thước ảnh |
| 1 | KTH | 3195 | 1595 | 1600 | 11 | 331 x 331 |
| 2 | Kyberge | 4480 | 2240 | 2240 | 28 | 331 x 331 |
| 3 | UIUC | 1000 | 500 | 500 | 25 | 331 x 331 |
| 4 | Tổng hợp | 8675 | 4335 | 4340 | 64 | 331 x 331 |

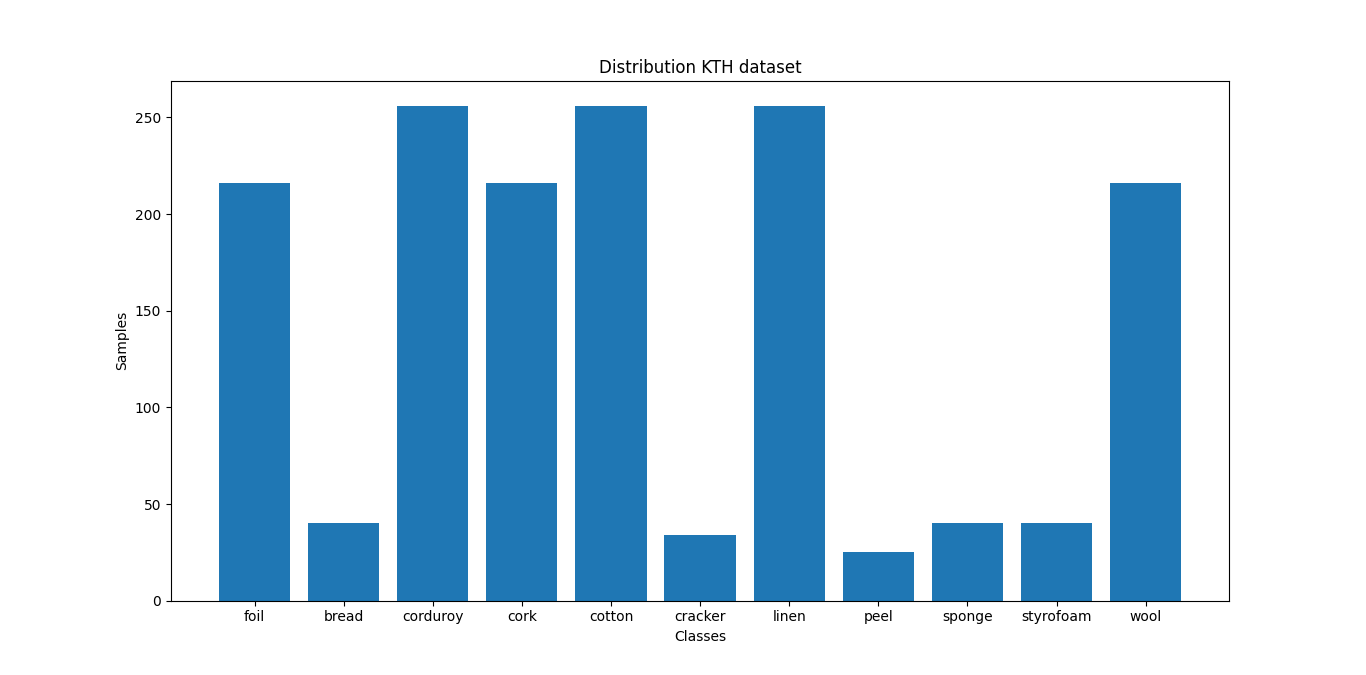
  
(a)

  
(b)

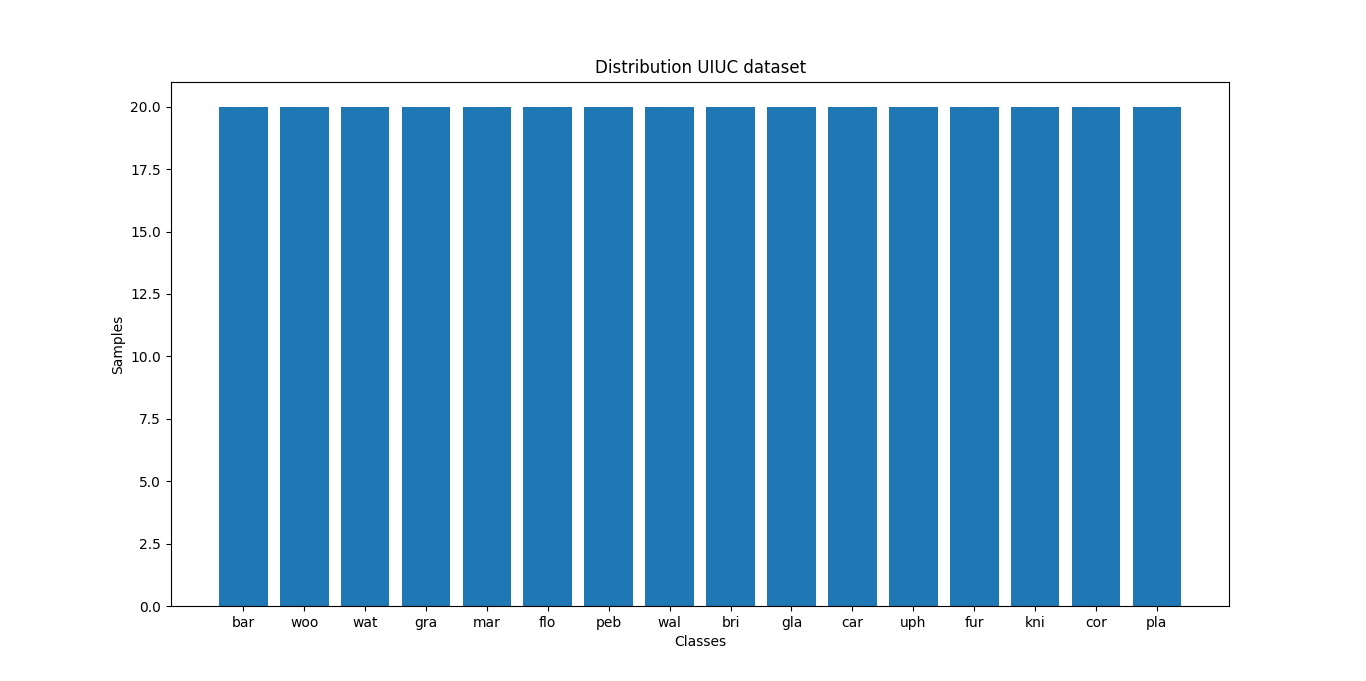
   
(c)

**Hình 4.1 Minh họa hình ảnh trong tập dữ liệu. (a) tập dữ liệu KTH, (b) tập dữ liệu Kyberge, (c) tập dữ liệu UIUC**

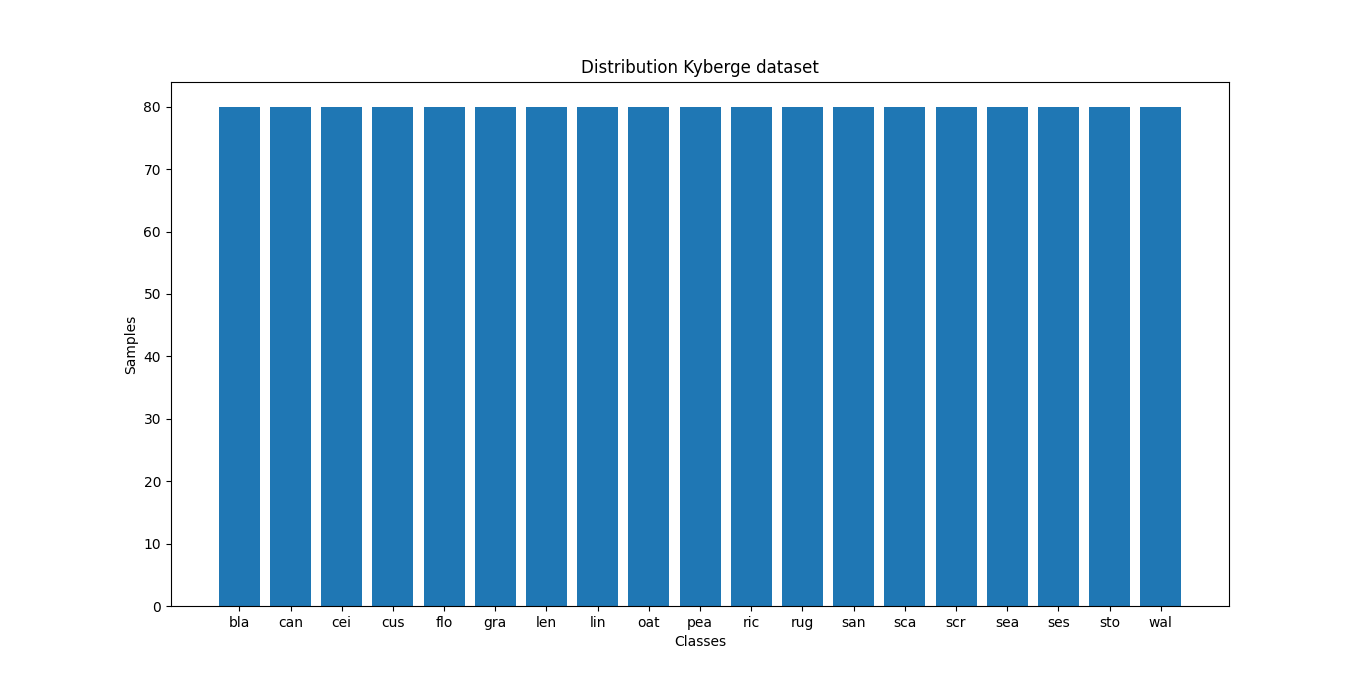
Trong các tập dữ liệu đều có thách thức riêng đối với nhà khoa học. Trong đó tiêu biểu lên là hai vấn đề: cân bằng dữ liệu, và số lượng dữ liệu. Trong đó tập KTH có sự mất cân bằng rõ rệt, có những lớp vài trăm hình ảnh nhưng có lớp chỉ có vài chục ảnh. Sự mất cân bằng dữ liệu là một trong vấn đề và thách thức đối với nhà nghiên cứu khoa học, khiến các mô hình dễ bị thiên vị đối với các lớp có nhiều mẫu và bỏ quên các lớp có số lượng mẫu. Bên cạnh đó tập dữ liệu UIUC lại đặt ra một thách thức khác đó là sự khan hiếm và thiếu hụt dữ liệu. Mỗi một loại kết cấu trong tập UIUC chỉ có 20 hình ảnh. Điều này khiến cho các mô hình học sâu chưa thể phát huy được khả năng cũng như dễ rơi vào trường hợp học quá mức (overfitting) và học chưa tốt (under fitting). Chưa dừng lại ở đó các mô hình học sâu có số lượng tham số lớn khiến cho việc sử dụng các tập dữ liệu với số lượng ít không đem lại hiệu quả và mô hình chưa thể khái quát và tìm hiểu các đặc điểm của dữ liệu.



**Hình 4.2 Sơ đồ phân bố số lượng mẫu/lớp trong tập dữ liệu KTH**

****

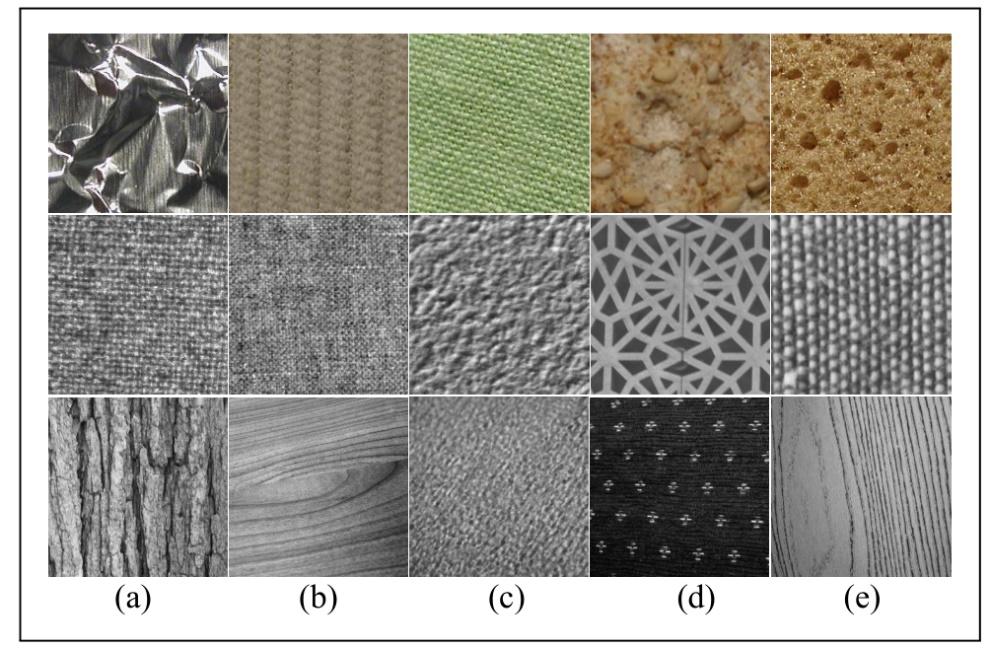
**Hình 4.3 Sơ đồ phân bố dữ liệu mẫu/lớp tập dữ liệu UIUC**

****

**Hình 4.4 Sơ đồ phân bố dữ liệu mẫu/lớp của tập dữ liệu Kyberge**

Chúng tôi tích hợp tập dữ liệu về kết cấu bề mặt (8675 hình ảnh), chứa 64 lớp từ 3 tập dữ liệu công khai. Các hình ảnh được thay đổi kích thước đồng nhất thành 331 x 331. Đáng chú ý, 11 lớp (3195 hình ảnh) từ dữ liệu bề mặt KTH (bao gồm KTH-TIPS và KTH-TIPS2), 28 lớp (4480 không có bản vá xoay) từ tập dữ liệu Kyberge và 25 lớp (1000 hình ảnh) từ tập dữ liệu UIUC. Chi tiết được hiển thị trong hình.

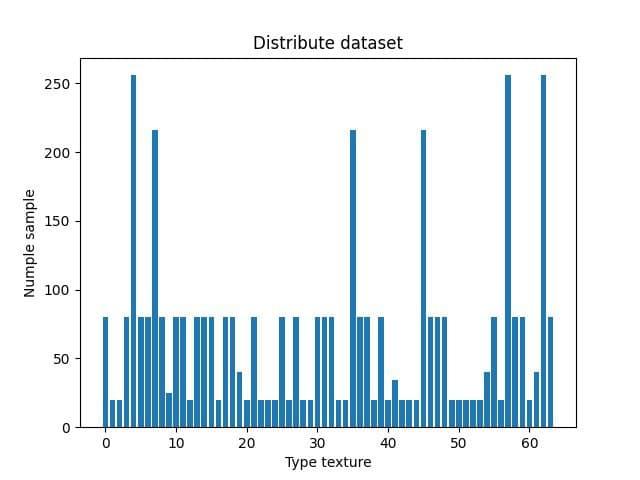
Những hình ảnh này là kết cấu bề mặt trong thế giới thực từ gỗ, chăn, vải, da, v.v., nó có thể được sử dụng để đánh giá năng lực của các mô hình hoặc được sử dụng làm tập dữ liệu sơ bộ để cải thiện hiệu suất của các mô hình phân loại.



**Hình 4.5 Minh họa ảnh kết từ tập Dataset kết hợp**

Bộ dữ liệu kết cấu. Mỗi hình ảnh là một mẫu của một lớp họa tiết, hàng đầu tiên từ [KTH](http://www.nada.kth.se/cvap/databases/kth-tips) , hàng thứ hai từ [Kyberge](http://www.cb.uu.se/~gustaf/texture/) và hàng thứ ba từ [UIUC](http://www-cvr.ai.uiuc.edu/ponce_grp/data/) .Tập dữ liệu được chia thành một nửa ngẫu nhiên, một cho tập huấn luyện và một tập khác để đánh giá.

Biểu đồ phân bố dữ liệu từ tập dữ liệu kết hợp:



**Hình 4.6 Phân bố dữ liệu tập dữ liệu kết hợp**

* 1. **CÁC THIẾT LẬP THÔNG SỐ THỰC NGHIỆM VÀ KẾT QUẢ THỰC NGHIỆM**

Trong phần này chúng tôi sẽ trình bày về những toàn bộ tham số và kỹ thuật để cấu hình cho mô hình trong quá trình triển khai mô hình.

* Phương pháp tiếp cận dựa trên học máy

Trong quá trình xây dựng mô hình phân loại với học máy chúng tôi sử dụng mô hình phân loại Linear SVM được triển khai trong sklearn.

Tham số cấu hình cho mô hình Linear SVM bao gồm:

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tham số | Ý nghĩa | Giá trị |
| 1 | penalty | Chỉ định tiêu chuẩn được sử dụng trong hình phạt | L2 |
| 2 | Loss | Hàm mất mát | squared\_hinge |
| 3 | C | Tham số điều chỉnh tương ứng tốc độ học tập | 1 |
| 4 | Class\_weight | Cân bằng trọng số cho các class khi dữ liệu mất cân bằng. | balance |
| 5 | Max\_iter | Số lần lặp tối đa chạy trên tất cả các lớp. | 10000 |

* Phương pháp tiếp cận dựa trên học sâu

Tham số trong việc huấn luyện mô hình học sâu:

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tham số | Ý nghĩa | Giá trị |
| 1 | epochs | Số lô lần lặp qua toàn bộ dữ liệu | 50 |
| 2 | Bacth\_size | Số lượng dữ liệu sử dụng cho một lần đào tạo | 16 |
| 3 | Learning\_rate | Tốc độ đào tạo | 0.01 |
| 4 | Loss | Hàm mất mát | Focal\_loss |
| 5 | Optimer | Trình tối ưu | Adam |

Độ chính xác của hệ thống nhận dạng được tính theo tỷ lệ phần trăm của các phân loại đúng, được tính như sau:

 (4.1)

Trong đó: Accuracy là độ chính xác tính theo phần trăm. #of correct classification là số ảnh kiểm tra được phân loại đúng. #of total testing images là tổng số ảnh được sử dụng cho kiểm tra.

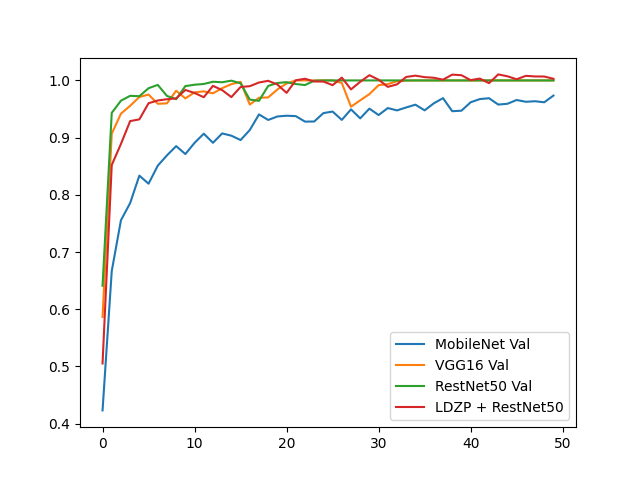
***Kết quả thực nghiệm:***

Kết quả đánh giá hiệu quả các mô hình dựa trên tiêu chí F1-score với từng tập dữ liệu được hiển thị bảng bên dưới:

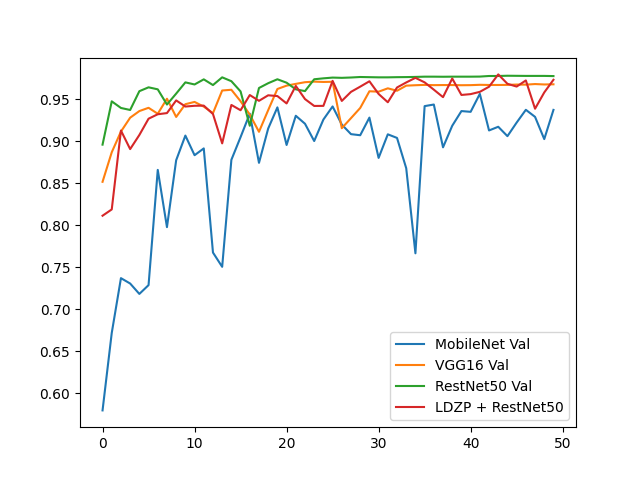
**Bảng 4.2 Kết quả thực nghiệm**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STT | Phương pháp | Dataset 01  KTH | Dataset 02  Kybers | Dataset 03  UIUC | Dataset 04  Tổng hợp |
| 1 | Histogram + SVM | 59.62% | 74.15% | 55.00% | 55.10% |
| 2 | LBPH + SVM | 77.5% | 98.44% | 80.23% | 84.75% |
| 3 | LDZP + SVM | 84.13% | 98.57% | 78.10% | 85.71% |
| 4 | MobileNetV1 | 98.14% | 99.02% | 88.62% | 95.67% |
| 5 | VGG16 | 98.89% | 99.37% | 93.60% | 97.13% |
| 6 | RestNet50 | 98.51% | 99.46% | **95.60%** | **97.83%** |
| 7 | LDZP + RestNet50 | **99.66%** | **99.91%** | 95.45% | 97.79% |

Ba mô hình học sâu hoạt động một cách hiệu quả và ổn định trong tập dữ liệu.



**Hình 4.8 Quá trình huấn luyện của các mô hình học sâu**



**Hình 4.9 Quá trình dự đoán của các mô hình học sâu**

**4.3 XÂY DỰNG CHƯƠNG TRÌNH THỬ NGHIỆM**

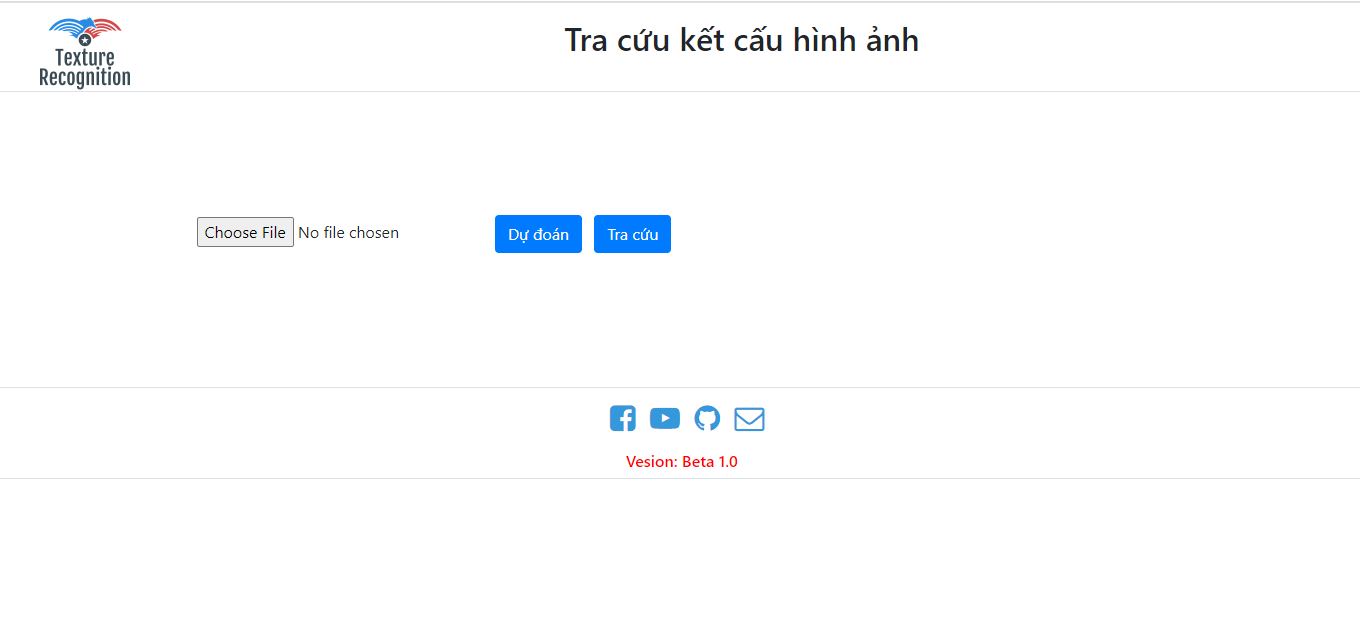
Những kết quả trên đã tạo một tiền đề trong việc vận dụng học sâu vào trong bài toán phân loại kết cấu trong thực tiễn. Từ những kết quả trên chúng tôi xây dựng một ứng dụng thực nghiệm để có thể tra cứu và phân loại hình ảnh kết cấu.

Chương trình có hai chức năng chính:

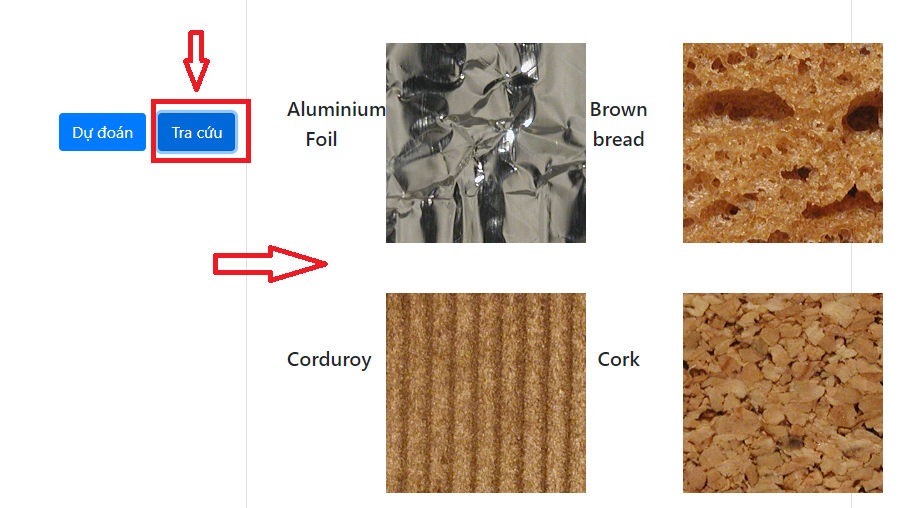
- Tra cứu bằng hình ảnh

- Tra cứu toàn bộ danh sách loại kết cấu

Dưới đây là giao diện chính bắt đầu của chương trình:

  
**Hình 4.10 Giao diện chính của ứng dụng**

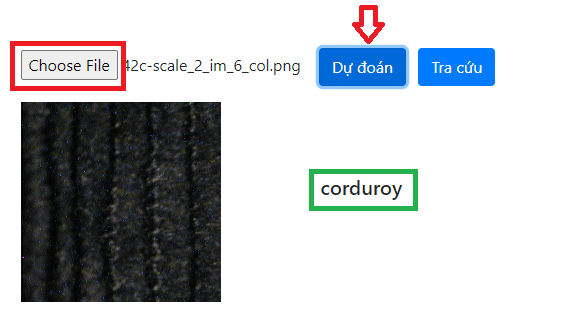
Để thực hiện tra cứu danh sách các loại kết cấu có sẵn thì người dùng click vào nút Tra cứu hình ảnh các loại kết cấu sẽ hiện bên tay phải của giao diện



**Hình 4.11 Hướng dẫn tra cứu các loại kết cấu**

Để sử dụng tính năng tra cứu loại kết cấu bằng hình ảnh người dùng cần làm theo các bước sau:

Bước 1: Click Choose File và chọn file ảnh cần tra cứu



**Hình 4.12 Minh họa chức năng tra cứu hình ảnh kết cấu bằng hình ảnh.**

Bước 2: Người dùng sau khi thấy được upload lên xong thì click nút Dự đoán và kết quả trả về bên dưới như hình minh họa được khoanh màu xanh.

**4.4 ĐÁNH GIÁ**

Phương pháp sử dụng biểu đồ histogram của hình ảnh không hiệu quả với dữ liệu kết cấu do nó chỉ xem xét được sự phân bố của mức xám nên gần như chỉ quan sát được đặc trưng màu sắc của hình ảnh nên đối tập dữ liệu có màu sắc tương đồng như UIUC thì mô hình đạt mức 55%, không có ý nghĩa mặt thực tiễn.

Sự hoạt động vượt trội của các mô hình học sâu đã thể hiện rõ nét bằng sự chênh lệch trong bảng tại mục (4.2). Các mô hình học sâu cải thiện đáng kể khả năng học và dự đoán của mô hình. Hiệu suất đánh giá trên các mô hình học sâu luôn đạt trên 90% đối với toàn bộ các dữ liệu với các mô hình các mức khác nhau từ dành cho thiết bị di động MobileNet, đến máy tính phổ thông VGG16, hay các thiết bị có hỗ trợ GPU ResNet50.

Trong đó kể đến mô hình ResNet50 cho khả năng dự đoán tốt hơn hết trong hầu hết các tập dữ liệu đánh giá cao hơn 2-5% so với mô hình còn lại. Điều này phản ánh chính xác khả năng của mô hình học sâu ResNet50 với bộ trọng số lớn và có thể trích xuất đặc trưng sâu hơn của kết cấu so với hai mô hình MobileNet và VGG16. Tuy vậy một mô hình cho di động với số lượng tham số chỉ bằng 1/8 so với RestNet và 1/5 so với VGG16, thì MobileNet đang cho thấy đây là một mô hình nhỏ gọn nhưng độ chính xác tương đối cao hiệu suất trung bình đạt 95% đối với tập dữ liệu tổng hợp với 64 class. Điều này mở ra những ứng dụng dễ dàng được tích hợp và xử lý ngay trên các thiết bị di động.

Các mô hình học sâu cho kết quả ổn định đối với các tập dữ liệu. Tuy vậy các phương pháp tiếp cận dựa trên học máy có ưu điểm: tốc độ huấn luyện nhanh, thời gian dự đoán nhanh và độ chính xác ổn trên 85% đối với mô hình LDZP. Điều này minh chứng các mô hình học máy có khả năng đạt được hiệu quả với các mô hình học sâu. Trong khi đó các mô hình này có khả năng huấn luyện nhanh, và không yêu cầu lượng dữ liệu lớn và ít bị ảnh hưởng với dữ liệu không cân bằng như trong các mô hình học sâu.

Bên cạnh đó, trong quá trình huấn luyện từ sơ đồ tốc độ huấn luyện chúng ta thấy hai mô hình ResNet50 và VGG16 cho tốc độ hội tụ nhanh hơn trong khoảng 5 epochs so với MobileNet và độ chính xác tốt hơn. Do cả hai mô hình kia được huấn luyện với tập ImageNet 1000 lớp trước đó. Các lớp cơ sở được đóng băng nên thời gian đào tạo và hiệu quả đề cao hơn MobileNet.

Trong các tập dữ liệu thì tập dữ liệu UIUC thường cho kết quả kém nhất cụ thể đối với cả mô hình RestNet và phương pháp kết hợp ảnh đặc trưng cục bộ và RestNet lí do là do bộ dữ liệu này gặp vấn đề hạn chế về dữ liệu nên các kết quả thử nghiệm tối đa đạt độ chính xác 95%.

Một trong những điểm sáng của phương pháp biến đổi ảnh cục bộ sau đó sử dụng bộ phân loại là nhận diện khá tốt trong tập dữ liệu KTH và Kyberge kể cả trường hợp mất cân bằng dữ liệu, nếu đảm bảo được một lượng dữ liệu đủ lớn mô hình hoạt động cực kì hiệu quả với độ chính xác lên đên ~99.9%. Trong tập dữ liệu KTH nó cao hơn gần 2% so với phương pháp sử dụng mạng CNN RestNet. Đạt được những kết quả như vậy mô hình có phép biến đổi giúp thể hiện rõ các đặc điểm mối tương quan giữa các pixel với nhau. Sau đó mô hình phân loại dễ dàng có thể tiếp cận và phân loại hình ảnh hơn. Nên kết quả khá ấn tượng. Còn đối với tập dữ liệu UIUC kết quả kém hơn một chút so với sử dụng duy nhất một mạng RestNet. Kết quả này phản ánh mô hình khi kết hợp yêu cầu một lượng dữ liệu lớn và và tập dữ liệu biến đổi này khi được ánh xạ sang bằng phép biến đổi cục bộ đã làm hao tổn đi một phần thông tin.

* 1. **KẾT CHƯƠNG**

Trong chương này tác giả trình bày giao diện chính của chương trình Demo, các giá trị tham số và kết quả thực nghiệm của mô hình đề xuất. Bên cạnh đó là các phân tích so sánh dựa trên các kết quả thu được.

Từ những phân tích trên có thể khẳng định các mô hình học sâu đang có những ảnh hưởng tích cực tới mô hình phân loại hình ảnh kết cấu. Bên cạnh đó độ chính xác của mô hình khi áp dụng học sâu đều cho kết quả rõ rệt. Từ những nhận định triển khai so sánh có thể khẳng định có thể vận dụng các mô hình học sâu vào các bài toán trong thực tế một cách hiệu quả và tiện ích.

Tuy vậy các mô hình dựa trên học máy có những kết quả tích cực với lượng dữ liệu nhỏ thì mô hình học máy có kết quả phân loại ngang ngửa các mô hình học sâu và tránh các trường hợp overfitting trong các mô hình học sâu.

# KẾT LUẬN

1. **Tổng kết các việc đã thực hiện**

Sau thời gian nghiên cứu lý thuyết về bài toán phân loại kết cấu, xây dựng các thuật toán, cài đặt và thử nghiệm đạt được một số kết quả sau đây:

* Đề xuất phương pháp biểu diễn đặc trưng hình ảnh kết cấu giúp nâng cao độ chính xác trong bài toán phân loại kết cấu.
* So sánh chứng minh đề xuất hiệu quả hơn đáng kể so với việc không sử dụng phép biến đổi zigzag cho hình ảnh kết cấu.
* Xây dựng ứng dụng web cho phép phân loại, tra cứu các loại kết cấu bao gồm 64 loại kết cấu từ 3 tập dữ liệu đã được xuất bản.

Như vậy, có thể kết luận rằng phương thức đề xuất có thể đưa vào áp dụng trong một vài điều kiện thực tế với quy mô vừa và nhỏ.

1. **Hạn chế của đề xuất**

Từ kết quả thực nghiệm có thể thấy đề xuất còn một số hạn chế như sau:

* Độ chính xác chưa cao trên tập dữ liệu UIUC do sự nhầm lẫn giữa các loại vật liệu có tính chất tương tự như: vải và vải lanh…
* Do thực hiện phương pháp trích xuất và biến đổi dựa theo zigzag nên sẽ chậm hơn so với cách tiếp cận phân loại trực tiếp từ ảnh xám đầu vào.
* Biểu đồ histogram còn nhiều vector thưa nên dẫn đến hiệu suất so sánh bị giảm trong một số trường hợp.

1. **Các hướng phát triển tiếp theo**

Dựa trên những kết quả đạt được, chúng tôi xin đưa ra một số hướng phát triển đề tài như sau:

* Tiếp tục đánh giá cách tiếp cận đề xuất với các loại phương thức biểu diễn đặc trưng khác;
* Tiếp tục nghiên cứu và đề xuất cách khắc phục trường hợp phân loại cùng một loại vật liệu với hiệu suất kém.
* Sử dụng mô hình học sâu với hiệu suất cao và máy tính với khả năng tính toán lớn để nâng cao hiệu suất dự đoán.
* Bổ sung tập dữ liệu kết cấu với 100 loại.

# TÀI LIỆU THAM KHẢO

[1] J.C. Russ, The Image Processing Handbook, 3rd edition, CRC Press, Florida, 1999.

[2] IEEE Standard 610.4-1990.

[3] R.M. Haralick, K. Shanmugam, I. Dinstein, IEEE Trans. Syst. Man Cybern. 3 (1973) 610–621.

[4] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison-Wes- ley, Reading, MA, 1992.

[5] A. Sarkar, K.M.S. Sharma, R.V. Sonak, IEEE Trans. Image Process. 6 (1997) 407–413.

[6] G. Cross, A. Jain, IEEE Trans. Pattern Anal. Mach. Intell. 5 (1983) 25–39.

[7] U. Indhal, T. Næs, J. Chemometr. 12 (1998) 261 – 278.

[8] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S. and Vedaldi, A., 2014. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3606-3613).

[9] Haralick RM (1979) Statistical and structural approaches to texture. In: Proceedings of IEEE, vol 67, Issue 5, pp 786–804.

[10] Haralick RM, Bosley R (1973) Texture features for image classification. In: Third ERTS symposium, NASA SP-351, pp 1219–1228.

[11] Haralick RM, Shapiro LG (1992) Computer and robot vision, vol 1. Addison Wesley, Boston.

[12] Tso B, Mather P (2009) Classification methods of remotely sensed data, 2nd edn. CRC Press, Boca Raton.

[13] Petrou M, Sevilla PG (2006) Image processing: dealing with texture. Wiley, New York.

[14] Pietikainen MK (2000) Texture analysis in machine vision. Series in machine perception and artificial intelligence, vol 40. World Scientific, Singapore.

[15] Dubes RC, Jain AK (1989) Random field models in image analysis. J Appl Stat 16:131–164.

[16] Frankot RT, Chellapa R (1987) Lognormal random-field models and their applications to radar image synthesis. IEEE Trans Geosci Remote Sens GE-25(2).

[17] He DC, Wang L (1989) Texture unit, texture spectrum, and texture analysis. In: Proceedings of IGARSS’ 89/12th Canadian symposium remote sensing, vol 5, pp 2769–2772.

[18] He D-C, Wang L (1990) Texture unit, texture spectrum, and texture analysis. IEEE Trans Geosci Remote Sens 28(4).

[19] Maenpaa T (2003) The local binary pattern approach to texture analysis – extensions and applications. Oulun Yliopisto, Oulu.

[20] Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Recognit Mach Intell 24 (7).

[21] T. Ojala, M. Pietikäinen, T. Mäenpää, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns”. IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971-987, 2002.

[22] C. M. Gurney, J. R. G. Townshend,"The Use of Contextual Information in the Classification of Remotely Sensed Data", Photogrammetric Engineering and Remote Sensing, vol. 49, No. 1, 55-64, 1983.

[23] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009.

[24] Roy, S.K., Chanda, B., Chaudhuri, B.B., Banerjee, S., Ghosh, D.K. and Dubey, S.R., 2018. Local directional ZigZag pattern: A rotation invariant descriptor for texture classification. *Pattern Recognition Letters*, *108*, pp.23-30.

[25] Guo, Zhenhua, Lei Zhang, and David Zhang. "A completed modeling of local binary pattern operator for texture classification." IEEE transactions on image processing 19, no. 6 (2010): 1657-1663.

[26] Guo, Zhenhua, Lei Zhang, and David Zhang. "Rotation invariant texture classification using LBP variance (LBPV) with global matching." Pattern recognition 43, no. 3 (2010): 706-719.

[27] Liu, Li, Yunli Long, Paul W. Fieguth, Songyang Lao, and Guoying Zhao. "BRINT: binary rotation invariant and noise tolerant texture classification." IEEE transactions on Image Processing 23, no. 7 (2014): 3071-3084.

[28] Xi, Meng, Liang Chen, Desanka Polajnar, and Weiyang Tong. "Local binary pattern network: A deep learning approach for face recognition." In 2016 IEEE international conference on Image processing (ICIP), pp. 3224-3228. IEEE, 2016.

[29] M. Cimpoi, S. Maji, I. Kokkinos, A. Vedaldi, Deep filter banks for texture recognition, description, and segmentation, IJCV 118 (1) (2016) 65–94.

[30] F. Hu, G.-S. Xia, J. Hu, L. Zhang, Transferring deep convolutional neural networks for the scene classification of highresolution remote sensing imagery, Remote Sensing 7 (11) (2015) 680–707.

[31] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, W. Gao, Wld: A robust local image descriptor, PAMI 32 (9) (2010) 1705–1720.