BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

NIÊN LUẬN CƠ SỞ
NGÀNH CÔNG NGHỆ THÔNG TIN

Đề tài

TÌM HIỂU CNN(Convolutional Neural Network) & ỨNG DỤNG VÀO THỰC TIỄN

Sinh viên: Nguyễn Phước Thành
Mã số: B1610669
Khóa: 42

Cần Thơ, 09/2019

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ
KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG
BỘ MÔN CÔNG NGHỆ THÔNG TIN

NIÊN LUẬN CƠ SỞ
NGÀNH CÔNG NGHỆ THÔNG TIN

Đề tài

TÌM HIỂU CNN (Convolutional Neural Network) & ỨNG DỤNG VÀO THỰC TIỄN

Người hướng dẫn
TS Lâm Nhựt Khang

Sinh viên thực hiện
Nguyễn Phước Thành
Mã số: B1610669
Khóa: 42

Cần Thơ, 09/2019

1. Lời cảm ơn

Cảm ơn TS. Lâm Nhựt Khang, Bộ môn Công nghệ Thông tin, Khoa Công nghệ Thông tin và Truyền thông, trường Đại học Cần Thơ đã tích cực hướng dẫn, giúp đỡ nghiên cứu đề tài này.

2. Mục lục

3. Danh mục đồ thị, biểu bảng và màn hình

Tóm lượt

Ngày nay, cùng với sự phát triển của công nghệ thông tin trong thời đại công nghệ 4.0 như hiện nay, các ngành kĩ thuật đang từng bước vươn lên khẳng định bản thân cùng với sự thay đổi, chuyển mình đem lại nhiều giá trị trong mọi mặt của cuộc sống thông qua việc áp dụng những tiến bộ khoa học kĩ thuật vào việc cải tiến quy trình tạo ra sản phẩm tăng năng suất lao động. Trong lĩnh vực trí tuệ nhân tạo, xử lý ảnh là một thành phần không thể không nhắc đến, việc giúp máy tính nhận diện hình ảnh một cách phi trực quan thông qua các giải thuật và thuật toán. Có nhiều phương pháp được sử dụng trong việc sử lý hình ảnh, mạng neural thần kinh được sử dụng để kết hợp với các giải thuật hồi quy như CNN, RNN …

4. Phần giới thiệu

Ngày nay, con người tiếp cận thông tin thông qua nhiều kênh khác nhau, internet, phim ảnh, báo đài … Các thông tin này tác động lên các giác quan làm cho con người nhận biết được thế giới xung quanh, trực quan nhất có thể nói đến là dữ liệu về hình ảnh. Con người có thể dễ dàng nhận biết và mô tả nội dung hình ảnh một cách dễ dàng, cũng như phát hiện chính xác vị trí các vật thể trong ảnh. Tuy nhiên việc này (đọc và hiểu một bức ảnh) sẽ khó hơn nhiều với máy tính khi máy tính chỉ có thể nhận biết hay xử lý dữ liệu thông tin với hai số 0 và 1. Thông qua việc tìm hiểu về CNN sẽ là cầu nối giúp ta hiểu được máy tính nhận diện một bức ảnh và xử lý chúng như thế nào.

Mục tiêu của đề tài là tìm hiểu cách hoạt động của một phương pháp trong lĩnh vực giúp máy tính nhận diện và phân loại ảnh.

Bố cục của bản báo cáo gồm 3 phần: phần giới thiệu, phần nội dung và phần kết luận. Trong đó phần nội dung gồm có 4 chương:

Chương 1 - Đặc tả yêu cầu: CNN giúp máy tính trong việc xác định đối tượng trong ảnh .

Chương 2 - Thiết kế giải pháp Khái quát về CNN: Trình bày các kiến thức liên quan đến CNN, các kĩ thuật được dùng trong xử lý ảnh.

Chương 3 - Cài đặt giải pháp: Mô tả hoạt động của CNN thông qua ví dụ đơn giản

Chương 4 - Đánh giá kiểm thử: Đánh giá và hướng phát triển

5. Phần nội dung

5.1. Chương 1 - Đặc tả yêu cầu:

Mục tiêu chính của thị giác máy tính (Computer vision) - một nhánh của trí tuệ nhân tạo. Thị giác máy tính tập trung giải quết các vấn đề như:

- **Phân loại ảnh, miêu tả ảnh.**
- **Phát hiện vật thể trong ảnh:** Xe, con người, đèn giao thông, ...
- **Tạo ảnh với những phong cách khác nhau:** Hiển thị nội dung ngữ nghĩa của ảnh gốc theo những phong cách khác nhau.

Mạng nơ-ron truyền thống (Neural Network) hoạt động thực sự không hiệu quả với dữ liệu đầu vào là ảnh. Nếu thước ảnh đầu vào quá lớn (1000px x 1000px) thì số thuộc tính sẽ là 1000x1000x3 thuộc tính. Điều này đòi hỏi khối lượng tính toán lớn và thường dẫn đến overfitting do không đủ điều kiện. Mạng CNN (Convolutional Neural Network) được thiết kế dựa theo tầm nhìn của sinh vật sống, nguồn cảm hứng đến từ "**neocognitron**" được giới thiệu bởi **Kunihiko Fukushima** năm 1979. CNN sử dụng phương pháp các lớp, mỗi lớp nhận một khối 3D đầu vào và biến thành một khối 3D đầu ra có chức năng khác biệt. Lớp phía sau là kết quả của lớp trước.

5.2. Chương 2 - Thiết kế giải pháp:

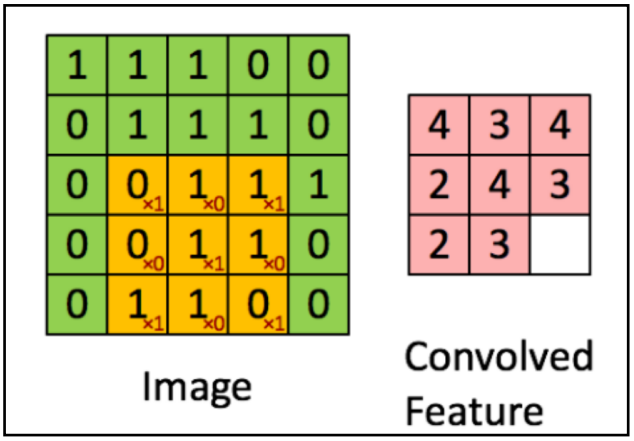
Mạng nơ-ron tích chập (CNN hay ConvNet) là mạng nơ-ron phổ biến nhất được dùng cho dữ liệu ảnh. Bên cạnh các lớp liên kết đầy đủ (FC layers), CNN còn đi cùng với các lớp ẩn đặc biệc giúp phát hiện và trích xuất những đặc trưng - chi tiết (patterns) xuất hiện trong ảnh gọi là Lớp Tích chập (Convolutional Layers). Chính những lớp tích chập này làm CNN trở nên khác biệt so với mạng nơ-ron truyền thống và hoạt động cực kỳ hiệu quả trong bài toán phân tích ảnh.

5.2.1. Lớp tích chập (Convolutional):

Tích chập được xử lý đầu tiên trong xử lý tín hiệu số (Signal Processing). Nhờ vào sự biến đổi thông tin các nhà khoa học đã áp dụng kỹ thuật này vào xử lý ảnh và video số.

Phép tích chập là một phép biến đổi bằng cách đưa một bộ lọc lên hình ảnh ban đầu và bộ lọc bắt đầu chạy quanh ảnh. Về bản chất bộ lọc sẽ biến đổi mà trận f và g (được gọi là ma trận đặc trưng và ma trận đầu vào) tạo thành một ma trận mới. Ta sẽ lần "trượt" ma trận g trên ma trận f cùng với việc nhân tương ứng từng thành phần của ma trận sau đó tính tổng các tích.

Để dễ hình dung, ta có thể xem



Hình 1: Tích chập

5.2.2. Đặc trưng của ảnh:

Đặc trưng của ảnh là các chi tiết xuất hiện trong ảnh, bao gồm các chi tiết đơn giản như góc, cạnh, đường và các chi tiết phức tạp như mắt, tai, mũi, con vật … Các bộ lọc sẽ giúp phát hiện và tìm ra các đặc trưng của ảnh.

Bộ lọc càng sâu thì các đặc trưng được phát hiện càng phức tạp.

$$f=\begin{bmatrix}1&1&1&0&0\\0&1&1&1&0\\0&0&1&1&1\\0&0&1&1&0\\0&1&1&0&0\end{bmatrix};g=\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}\Rightarrow output=\begin{bmatrix}4&3&4\\2&4&3\\2&3&4\end{bmatrix}$$

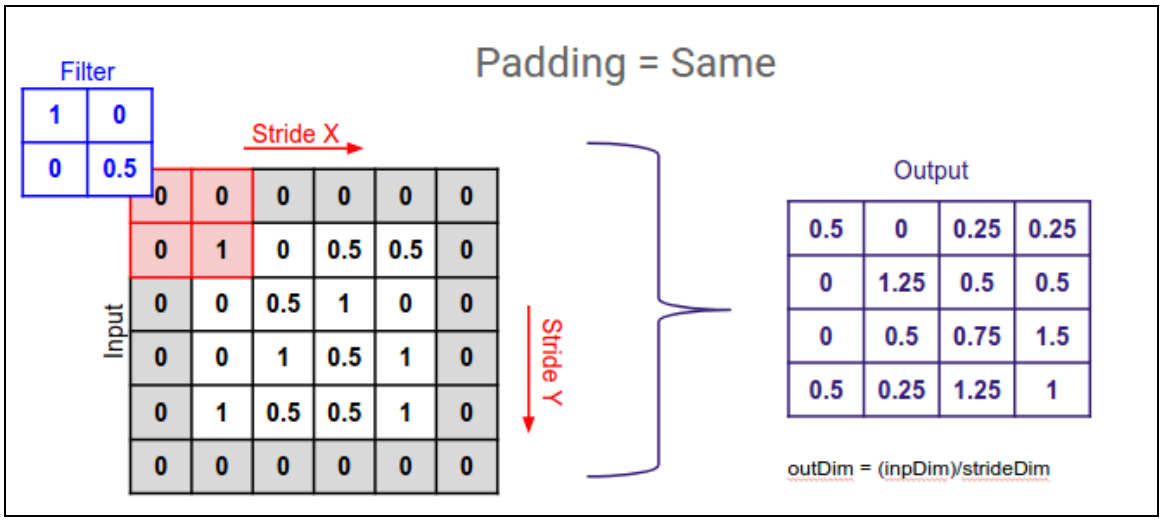
Thông thường một ma trân có kích thước NxN và bộ lọc có kích thước FxF thì ma trận kết quả có kích thước (N-F+1) x (N-F+1).

5.2.3. Padding:

Ảnh hưởng của phép tích chập:

- Làm giảm kích thước ma trận và để giảm nhiễu, chúng ta chỉ có thể tính tích chập vài lần trước khi ma trận trở nên quá nhỏ.

- Những điểm ảnh ở trung tâm được bao phủ bởi nhiều bộ lọc, nghĩa là được sử dụng để tính nhiều giá trị đầu ra, trong khi những điểm ảnh ở góc hoặc cạnh thì chỉ được sử dụng 1 – 2 lần. Vì thế chúng ta dễ bị mất thông tin tại các vùng gần của ảnh.

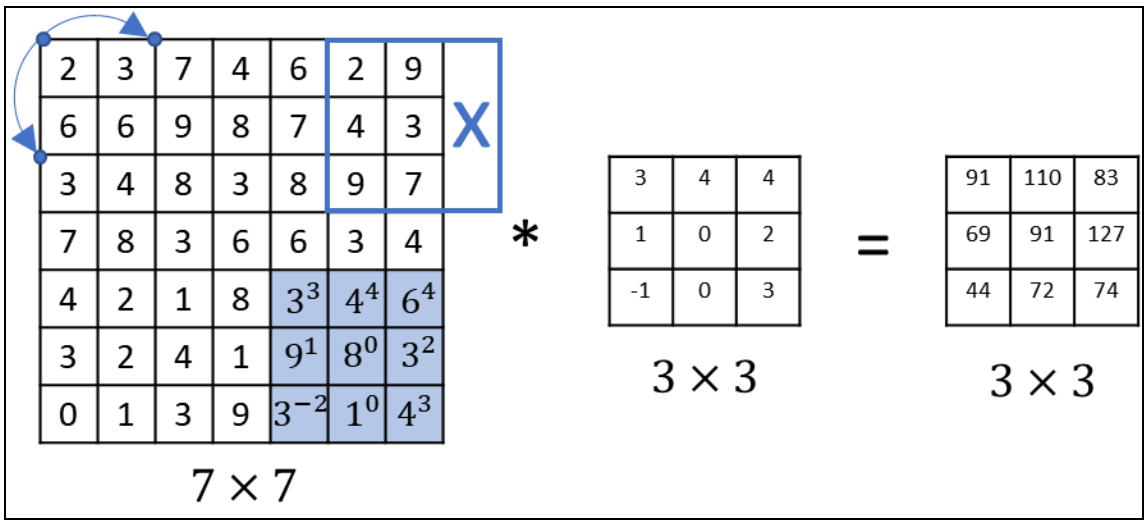


Hình 2: Đường viền phụ

Một đường viền phụ được thêm vào để tăng kích thước ma trận đầu vào làm cho các đường điểm trên đường viền lùi vào sâu một chút.

Theo qui ước đường viền phụ có giá trị bằng không, điều này giúp cho các đặc trưng của ảnh ở viền sẽ được giữ lại.

5.2.4. Nhân chập sải (Strided convolutions):



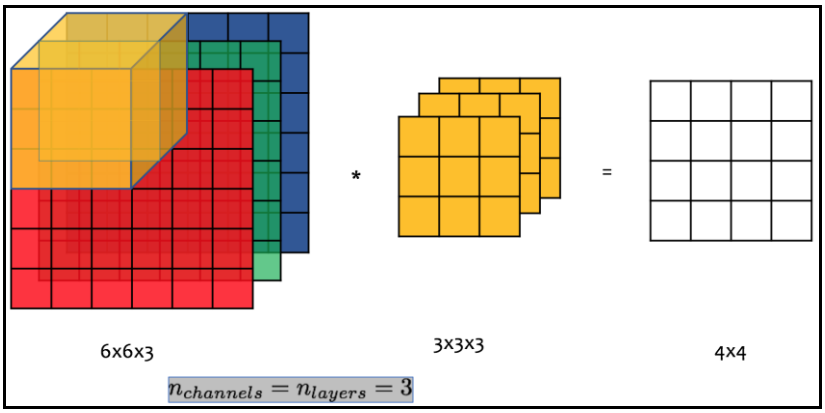
Hình 3: Convolution với stride bằng 2

"Trong lĩnh vực toán học thuần tuý, phép toán nhân chập được định nghĩa hơi khác so với phía trên. Trước khi thực hiện nhân chập (element-wise/ dot-product) và lấy tổng của các kết quả thu được, bộ lọc (filter) được lật lần lượt theo trục ngang và trục dọc (flipped filter). Ma trận đầu ra được tính dựa trên ma trận đầu vào và bộ lọc đã được lật này. Phép toán "nhân chập" được trình bày ở trên (thực hiện trực tiếp trên ma trận đầu vào và bộ lọc gốc) được gọi là tương quan chéo (cross-correlation). Tuy nhiên, theo quy ước trong ML và DL, phép tương quan chéo (cross-correlation) được gọi là phép nhân chập (convolution)"

5.2.5. Phép chập khối:

Phép chập khối với một bộ lọc:

Phép chập khối được dùng nhiều trong việc trích xuất đặc trưng của ảnh màu (3D images). Giải sử chúng ta có kích thước ảnh đầu vào 6x6 được biểu diễn theo hệ màu RGB. Ma trận đầu vào có kích thước 6x6x3 vì chúng ta có 3 kênh màu đỏ, xanh lục và xanh lam nên bộ lọc phải có kích thước tương ứng với 3 kênh màu.

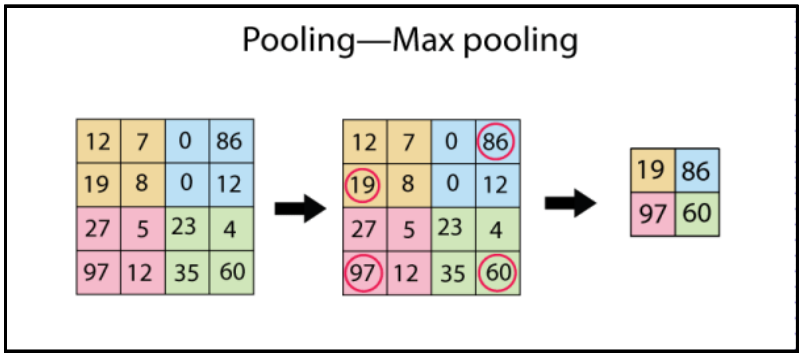


Hình 4: Sử dụng bộ lọc với ảnh màu RGB

Mỗi lớp của phép nhân chập với phần diện tích bao phủ bởi nó trên kênh tương ứng đầu vào. Tại một vị trí cụ thể của khối lọc, giá trị tại ô tương ứng của ma trận đầu ra (ma trận 2 chiều) là tổng của ba tích thu được.

5.2.6. Lớp pooling:

Lớp pooling được sử dụng trong CNN để giảm kích thước đầu vào, tăng tốc độ tính toán và hiệu năng trong việc phát hiện đặc trưng. Có nhiều hướng Pooling được sử dụng, trong đó phổ biến nhất là pooling theo giá trị cực đại (max pooling) và pooling theo giá trị trung bình (average pooling).



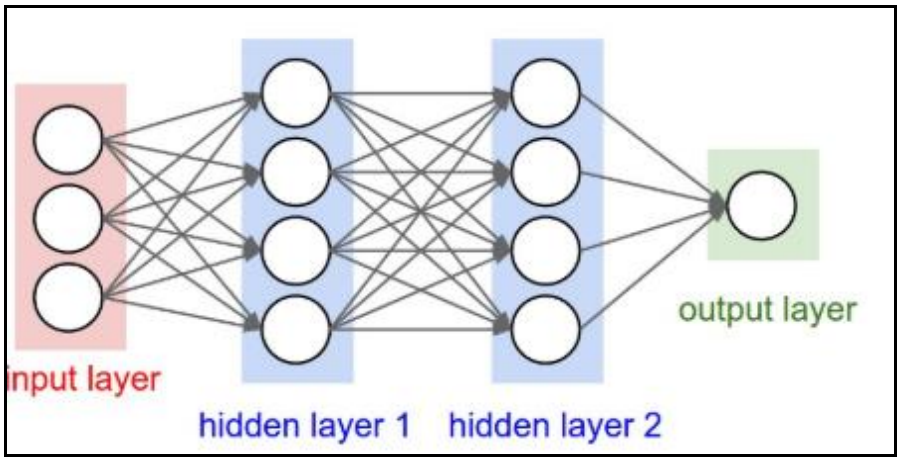
Hình 5: Lớp max pooling

Ta nhận thấy với cách sử dụng **max pooling** thì một đặc tính trong ảnh nếu được phủ bởi một bộ lọc sẽ được giữ lại sau khi sử dụng **max pooling**.

5.2.7. Lớp liên kết đầy đủ (Fully-connected layer):

Lớp liên kết đầy đủ giống như các lớp khác trong mạng thần kinh, mỗi nơ-ron trong lớp kết tiếp nhận các nơ-ron của lớp trước đó như giá trị đầu vào.

Cái tên fully-connected thể hiện cho việc tất cả các nơ-ron trong mang tiếp theo luôn kết nối với tất cả các nơ-ron của lớp trước



Hình 6: Các lớp ẩn trong lớp liên kết hoàn toàn

FC layer thường được sử dụng vào mỗi cuối của CNN. Vì vậy khi chúng ta đến giai đoạn này thì chúng ta có thể làm phẳng ma trân thành mảng 1 chiều, sau đó chúng ta có thể áp dụng các lớp ẩn (hiden layer) như bình thường.

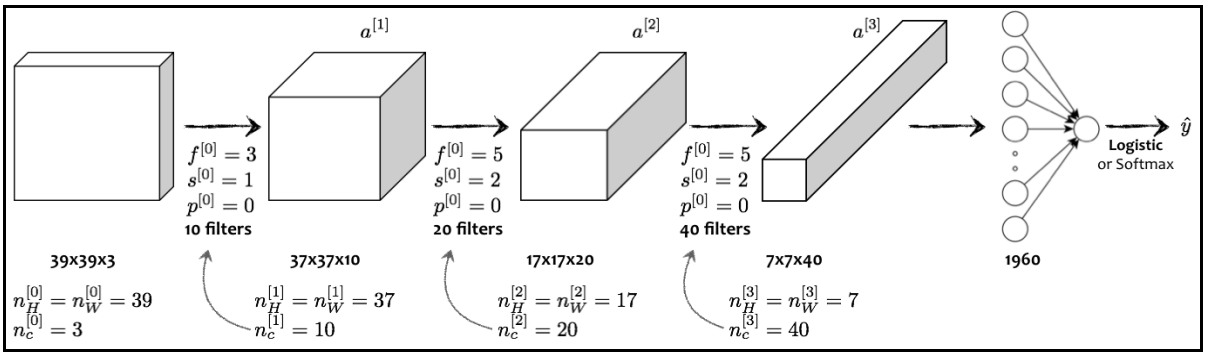
5.3. Chương 3 - Cài đặt giải pháp:

5.3.1. Mạng CNN một lớp:

Mạng nơ-ron tích chập (CNN) chỉ đơn giản gồm một vài lớp (layer) tích chập (convolutional) kết hợp với các hàm kích hoạt phi tuyến (nonlinear activation function) như ReLU hay tanh để tạo ra thông tin trừu tượng hơn (abstract/higher-level) cho các lớp (layer) tiếp theo.

Các lớp liên kết được với nhau thông qua cơ chế tích chập. Lớp tiếp theo là kết quả của lớp trước, nhờ vậy ta có đước các nối kết cục bộ. Nghĩa là mọi nơ-ron ở lớp (layer) tiếp theo sinh ra từ ma trận lọc (filter) áp đặt lên một vùng ảnh cục bộ của nơ-ron lớp (layer) trước đó.

Trong suốt quá trình huấn luyện, mạng nơ-ron tích chập (CNN) sẽ tự động học được các thông số cho các ma trận lọc. Ví dụ trong việc phân lớp ảnh, CNN sẽ cố gắng tìm ra thông số tối ưu cho các ma trận tương ứng theo thứ tự **điểm ảnh thô (raw pixel) > cạnh (edges) > hình dạng bề mặt (shapes) > các điểm trong bề mặt (facial) > các đặc trưng ở mức cao hơn (high-level features)**. Lớp cuối cùng là lớp kết nối đầy đủ được dùng để phân lớp ảnh.



Hình 7: Sơ đồ CNN đơn giản

Ta nhận thấy càng về cuối của CNN, kích thước của ảnh càng giảm trong khi số chiều tăng dần.

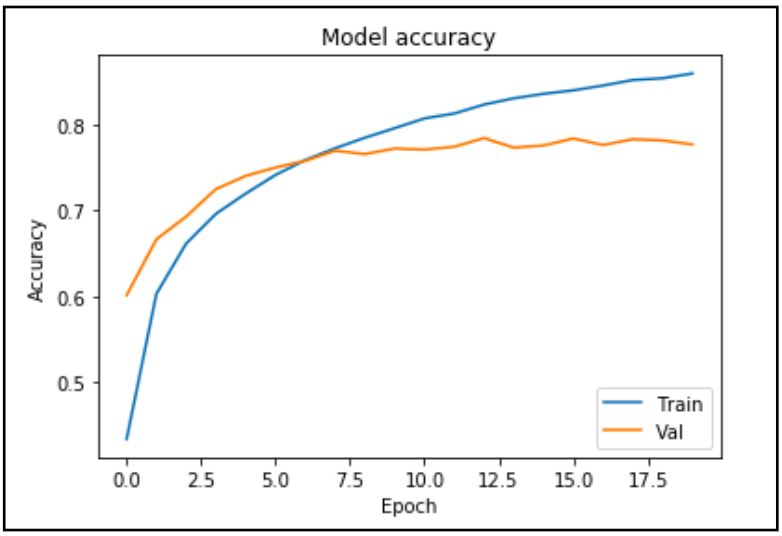
Ba lớp cơ bản được sử dụng trong một CNN:

- Lớp tích chập: Convolution (CONV)
- Lớp Pooling: Pooling layer (POOL)
- Lớp liên kết đầy đủ: Fully Connected (FC)

5.3.2. Ví dụ một CNN cụ thể:

Tập dữ liệu CIFAR-10 được sử dụng cho việc huấn luyện mô hình và kiểm thử. Tập dữ liệu là một bộ các ảnh thuộc 10 lớp khác nhau. Các lớp đại diện cho máy bay, xe hơi, chim, mèo, chó, nai, ếch, ngựa, tàu và xe tải.

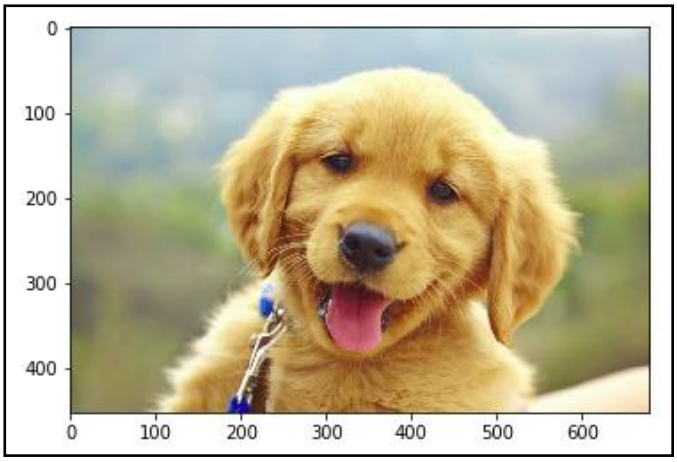
Thư viện Tensorflow cùng các thư viện khác như numpy, keras sẽ được sử dụng. Sau khi huấn luyện mô hình có kết quả khá khả quan.



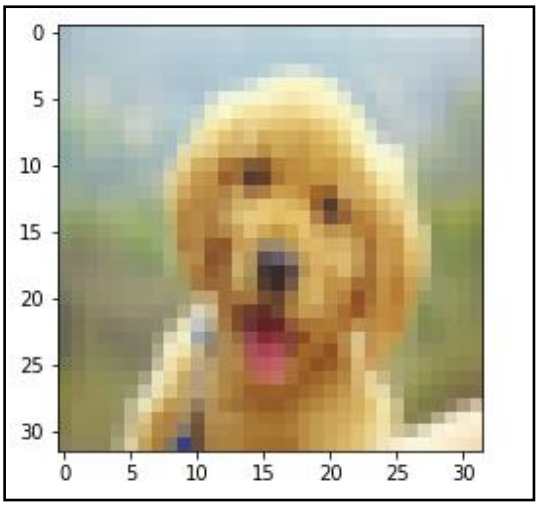
Biểu đồ sai số qua các lần lặp

** Kết quả đánh giá:*

Hình ảnh đầu vào



Hình ảnh được resize với kích thước 32 x 32



Kết quả dự đoán

```
Most likely class: dog -- Probability: 0.9338305
Second most likely class: bird -- Probability: 0.04670123
Third most likely class: cat -- Probability: 0.010684356
Fourth most likely class: horse -- Probability: 0.0066039423
Fifth most likely class: frog -- Probability: 0.0017065258
```

**Trong đó vẫn còn có các kết quả dự đoán chưa chính xác*

6. Phần kết luận:

Việc sử dụng mạng nơ-ron tích chập (CNN) có thể giúp ta trong việc phân loại ảnh, trích xuất đặc trưng của ảnh mang lại kết quả khả quan.

Do là mô hình máy học nên có thể cải tiến độ chính xác bằng nhiều phương pháp khác nhau như tiền xử lý ảnh, cải tiến số lượng các lớp trong CNN, tốc độ học, …. Tương lai có thể sử dụng các thuật toán xử lý ảnh khác để lấy đặc trưng ảnh tốt hơn, ….

7. Tài liệu tham khảo:

[1] Deshpande, A. (2018). *The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part 3).*

[2] H. Lee, R. R. (2009). *Convolutional deeb belief networks for scalable unsupervised learning of hierarchial representation.* Proceedings of the Twenty-Sixth International Conference on Machine Learing (ICML).

[3] (n.d.). *Keras Documentation.* keras.io.

[4] Krizhevsky, A. (2009). *Learning multiple layers of features from tiny images*. Master's Thesis, Department of Computer Science, University of Toronto.

[5] LeCun, Y., & Bengio, Y. (1995). *Convolutional networks for images, speech, and time series.* In Micheal A. (ed). The handbook of brain theory and neural networks (Second ed.). The MIT press, pp. 276-278.

[6] Patrice Y. Simard, D. S. (2003). *Best pratice for Convolutional Neural Networks Applied to Visual Document Analysis.* In International Conference on Document Analysis and Recogntion (ICDAR), IEEE Computer Society, Los Alamitos, page 958-962.

[7] Zhang, W. (1988). *Shift-invariant pattern recognition neural network and its optical architecture.*