|  |  |
| --- | --- |
| ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH  **TRƯỜNG ĐẠI HỌC**  **CÔNG NGHỆ THÔNG TIN** | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc Lập - Tự Do - Hạnh Phúc** |

# ĐỀ CƯƠNG CHI TIẾT

|  |  |
| --- | --- |
| **TÊN ĐỀ TÀI TIẾNG VIỆT: Hiện thực mạng nơ-ron tích chập Inception Resnet v2 bằng ngôn ngữ phần cứng Verilog** | |
| **TÊN ĐỀ TÀI TIẾNG ANH: Implementing Convolutional Neural Network Inception Resnet v2 with Verilog** | |
| **Giảng viên hướng dẫn:**  [Trương Văn Cương](https://courses.uit.edu.vn/user/view.php?id=7747&course=7594) | |
| **Thời gian thực hiện:** Từ ngày 04/04/2021 đến ngày 22/06/2021 | |
| **Sinh viên thực hiện:**  **Nguyễn Hoàng Nghĩa - 18521143**  **Trần Cao Khải - 18520877** | |
| **Nội dung đề tài:**  **Tổng quan đề tài:**  **Sơ lược về CNN**  Những năm gần đây, ta đã chứng kiến được nhiều thành tựu vượt bậc trong ngành Thị giác máy tính (Computer Vision). Các hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay drone giao hàng tự động. Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay.  CNNs chỉ đơn giản gồm nhiều layer của convolution 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 layer tiếp theo.  Kiến trúc truyền thống của một mạng CNN: Mạng neural tích chập (Convolutional neural networks), còn được biết đến với tên CNNs, là một dạng mạng neural được cấu thành bởi 2 thành phần sau:  - Công cụ tích chập dùng để phân tách và nhận dạng các đặc tính từ ảnh để từ đó phân tích ảnh trong quá trình được gọi là trích xuất đặc tính của ảnh.  - Lớp fully connected dùng để tối ưu hóa đầu ra và dự đoán được các class của ảnh dựa trên những đặc tính đã được trích xuất ra ở phần trên.  Tầng tích chập và tầng pooling có thể được hiệu chỉnh theo các siêu tham số (hyperparameters) được mô tả ở những phần tiếp theo.  **Inception Resnet V2**  Inception Resnet V2 là cấu trúc mạng nơ ron tích chập được xây dựng dựa trên cấu trúc của Inception và có bao gồm liên kết tránh (Residual Connections). Được thử nghiệm trên hệ thống cơ sở dữ liệu của Imagenet, với 164 layers và được ứng dụng để phân loại ảnh lên đến 1000 vật thể khác nhau. Có đầu vào là một bức ảnh có kích thước 299\*299\*3 và đầu ra sẽ là tên các vật thể có độ chính xác cao nhất được nhận dạng từ bức ảnh đầu vào.  Được xây dựng và phát triển bởi 4 kĩ sư phần mềm chính của Google:  - Christian Szegedy.  - Sergey Ioffe.  - Vincent Vanhoucke.  - Alex Alemi.  Cùng với đội ngũ hỗ trợ:   * Mô hình kiến trúc: - Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi * Hạ tầng hệ thống: - Jon Shlens, Benoit Steiner, Mark Sandler, và David Andersen * Thư viện phần mềm TensorFlow-Slim: - Sergio Guadarrama và Nathan Silberman * Hiện thực hóa mô hình: - Fernanda Viégas và James Wexler   Thư viện phần mềm TensorFlow-Slim hỗ trợ mạnh mẽ các phép toán học để tính toán trong machine learning và deep learning đã giúp việc tiếp cận các bài toán trở nên đơn giản, nhanh chóng và tiện lợi hơn.  **Các kiểu tầng của CNN:**  **Tầng tích chập (CONV)**  Tầng tích chập (CONV) sử dụng các bộ lọc để thực hiện phép tích chập khi đưa chúng đi qua đầu vào I theo các chiều của nó. Các siêu tham số của các bộ lọc này bao gồm kích thước bộ lọc F và độ trượt (stride) S.  Padding  Mỗi một lần áp dụng phép nhân chập, kích thước của ảnh bị giảm xuống, và vì thế chúng ta chỉ có thể thực hiện nó một vài lần trước khi ảnh trở nên quá nhỏ.    Để khắc phục hai nhược điểm trên, một đường viền phụ (padding) được thêm vào xung quanh ma trận đầu. Việc thêm đường viền phụ làm tăng kích thước của ma trận đầu vào, dẫn tới tăng kích thước ma trận đầu ra. Từ đó độ chênh lệch giữa ma trận đầu ra với ma trận đầu vào gốc giảm. Những ô nằm trên cạnh/ góc của ma trận đầu vào gốc cũng lùi sâu vào bên trong hơn, dẫn tới được sử dụng nhiều hơn trong việc tính toán ma trận đầu ra, tránh được việc mất mát thông tin.  Stride  Trong phép nhân chập ở trên, bộ lọc trượt trên ma trận đầu vào 1 hàng/ cột trong mỗi bước di chuyển. Tuy nhiên, giá trị này có thể bằng 2, 3 hoặc lớn hơn.  Với stride = 1:    Với stride = 2:    v.v  **Tầng Pooling:**  CNNs có tính bất biến và tính kết hợp cục bộ (Location Invariance and Compositionality). Với cùng một đối tượng, nếu đối tượng này được chiếu theo các góc độ khác nhau (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị ảnh hưởng đáng kể. Pooling layer sẽ cho bạn tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling).  Tầng pooling (POOL) là một phép downsampling, thường được sử dụng sau tầng tích chập, giúp tăng tính bất biến không gian. Cụ thể, max pooling và average pooling là những dạng pooling đặc biệt, mà tương ứng là trong đó giá trị lớn nhất và giá trị trung bình được lấy ra.  **Tầng Fully Connected:**  Tầng cuối cùng của mô hình CNN trong bài toán phân loại ảnh là tầng fully connected layer. Tầng này được dùng để đưa ra kết quả, sau khi Convolutional Layer và Pooling Layer đã nhận được các ảnh truyền qua nó, thì lúc đó ta đã thu được kết quả là model đã đọc được khá nhiều thông tin về ảnh. Vì vậy, để liên kết các đặc điểm đó lại và cho ra output chúng ta sẽ sử dụng Fully connected Layer.  **Tầng ReLU:** ReLU Layer chính là một hàm kích hoạt trong neural network. Hàm kích hoạt còn được gọi là activation function. Tác dụng chính của hàm kích hoạt này là việc mô phỏng các nơ ron có tỷ lệ truyền xung qua axon. Hiện nay, hàm ReLU đang được sử dụng khá phổ biến và thông dụng vì nó có nhiều ưu điểm khác nổi bật trong việc huấn luyện các mạng nơ ron và giúp cho việc tính toán trở nên nhanh hơn.  ReLU layer được sử dụng sau khi mỗi filter map được tính toán ra và áp dụng hàm ReLU lên tất cả các giá trị của filter map.  **Các sơ đồ khối của Inception Resnet v2**  Sơ đồ khối Stem (299x299x3 → 35x35x384):  Sơ đồ khối Inception-ResNet-A (35x35x384 → 35x35x384):  Sơ đồ khối Reduction-A (35x35x384 → 17x17x1152):  Sơ đồ khối Inception-ResNet-B (17x17x1152 → 17x17x1152):  Sơ đồ khối Reduction-B (17x17x1152 → 8x8x2144):  Sơ đồ khối Inception-ResNet-C (8x8x2144 → 8x8x2144):  **Mục tiêu của đề tài:**  Input: là một bức ảnh màu với độ phân giải 299\*299\*3.  Output: là tên các vật thể có độ chính xác cao nhất được nhận dạng từ bức ảnh đầu vào với 3 class: bóng chày, bóng rổ, bóng bầu dục.  **Đánh giá kết quả và kết luận:**  Chi phí tính toán của Inception Resnet V2 gần bằng chi phí tính toán của Inception-v4. Inception-ResNet-v2 thì tốc độ train nhanh hơn nhiều và đạt độ chính xác cuối cùng tốt hơn một chút so với Inception-v4. Tuy nhiên, nếu ReLU được sử dụng làm đơn vị kích hoạt trước, Inception Resnet V2 có thể phát triển sâu hơn nữa.  **Multi-model**  **Single-Crop Single-Model** | |
| **Xác nhận của CBHD**  (Ký tên và ghi rõ họ tên) | **TP. HCM, ngày….tháng …..năm…..**  **Sinh viên**  (Ký tên và ghi rõ họ tên) |