

# Image Retrieval

taing@vinbrain.net

Ngày 12 tháng 4 năm 2021

## 1 Siamese Neural Network

Siamese Neural Networks (SNNs) là một loại neural networks với kiến trúc share weight. Kiến trúc này hiệu quả khi training với lượng dữ liệu hạn chế. SNN có thể dùng để predict xem 2 hình ảnh có phải tương tự nhau hay không. Điều này cho phép phân loại các lớp dữ liệu mới mà không cần huấn luyện lại mạng.

Để xây dựng một mạng Siamese đầu tiên cần phải tạo ra các cặp data ngẫu nhiên, nếu chúng cùng class thì gán bằng 1, ngược lại gán bằng 0.

## 2 Shuffle Neural Network

### 2.1 Pointwise Convolution

Pointwise convolution là phép tích chập của output feature map của layer phía trước với kernel có kích thước  $W \times H$  là  $1 \times 1$ . Đây là phương pháp dùng để giảm kích thước của feature map nhưng vẫn giữ được các đặc trưng.

### 2.2 Grouped Convolution

Ý tưởng của phương pháp này là thay vì tính tích chập cho toàn bộ các kênh như phép tích chập truyền thống là tích chập khối. Phương pháp này sẽ chia feature map ra thành nhiều nhóm nhỏ với số lượng kênh cố định. Thực hiện tích chập cho từng group nhỏ, sau cùng ghép các output feature map lại với nhau nhờ vậy số kênh ở output không khác gì với nhân tích chập truyền thống.

### 2.3 Channel Shuffle

Nhược điểm của Grouped convolution là một phép tính tính chập chỉ nhận một đầu vào một số channel cố định thay vì toàn bộ channel như tích chập khối do đó chỉ học được một phần đặc trưng của dữ liệu dẫn đến giảm hiệu suất. Channel Shuffle được sinh ra để giải quyết vấn đề này, nó giúp xáo

trộn ngẫu nhiên các output feature map của layer phía trước trước khi đưa vào layer tiếp theo.

## 2.4 Optimization

Kiến trúc của mạng Shuffle tối ưu hơn các mạng CNN thông thường nhờ vào việc giảm số lượng phép toán cũng như tham số mô hình.

Để tối ưu mô hình Shuffle sử dụng kiến trúc Residual tương tự như mô hình mạng ResNet bằng cách kết hợp thông tin giữa ngõ vào đã đi qua các hidden layer và ngõ vào gốc để tăng độ tin cậy của thông tin.

Tuy nhiên có nhiều sự cải tiến về ngõ vào gốc thay vì đi trực tiếp đến ngõ ra sẽ được đi qua một layer AVG Pool để giảm kích thước feature map và thay lớp Add bằng lớp Concat để có thể chồng hai đầu ra của hai nhánh với nhau giúp giữ nguyên kích thước channel so với khi không dùng lớp Pool.

## 3 Triplet Loss

Một cách khác để huấn luyện mạng Siamese là dùng hàm triplet loss. Hàm triplet loss được định nghĩa là khoảng cách dựa trên hàm loss với 3 đầu vào gồm anchor (điểm dữ liệu cần huấn luyện), positive (điểm dữ liệu cùng class với anchor) và negative (điểm dữ liệu khác class với anchor).

Công thức toán học:  $L = \max(d(a,p) - d(a,n) + \text{margin}, 0)$

Để tối ưu hàm loss về minimize cần dồn  $d(a,p)$  tiến về 0 và  $d(a,n)$  tiến ra vô cùng. Điều này có nghĩa sau khi training xong thì các mẫu positive sẽ gần với anchor hơn trong khi các mẫu negative sẽ ra xa anchor.