GoogLeNet là một mạng nơ-ron tích chập sâu được đề xuất bởi Szegedy và cộng sự [1]. Mạng này đã giành chiến thắng trong Thử thách nhận dạng hình ảnh quy mô lớn ImageNet 2014 (ILSVRC-2014) với hiệu suất phân loại đạt 92,3%. Đặc biệt, mô hình này được thiết kế theo kiến ​​trúc đặc biệt cho phép tăng độ sâu và chiều rộng của mạng nhưng vẫn giữ nguyên tài nguyên tính toán.

A graph of error rate

AI-generated content may be incorrect.

Mô hình VGG có tổng cộng 22 lớp và bao gồm 9 khối Inception. Mỗi khối Inception bao gồm bốn đường dẫn song song mà tại đó các lớp tích chập với các kích thước kernel khác nhau:

A diagram of convolutions

AI-generated content may be incorrect.

* Đường dẫn đầu tiên sử dụng một lớp tích chập với kích thước là 1 × 1.
* Trong đường dẫn thứ hai và thứ ba, một lớp tích chập có kích thước 1 × 1 được sử dụng trước khi đi qua lớp 3 × 3 và 5 × 5. Lớp tích chập 1 × 1 giúp giảm số kênh bộ lọc, do đó giảm độ phức tạp của mô hình.
* Đường dẫn thứ tư sử dụng một lớp max-pooling để giảm độ phân giải của đầu vào và tiếp theo là một lớp tích chập 1 × 1 để giảm chiều.

Bốn đường dẫn này sử dụng phần đệm thích hợp để đầu vào và đầu ra có cùng kích thước. Việc concat bốn đường dẫn này cho phép quét đầu vào ở các độ phân giải khác nhau. Đặc biệt, độ phức tạp của mô hình được giảm thiểu nhờ áp dụng lớp tích chập 1 × 1 trong mỗi đường dẫn.

A diagram of a flowchart

AI-generated content may be incorrect.

* Kích thước hình ảnh đầu vào là 224 × 224.
* Có chín khối Inception trong mạng này.
* Có bốn lớp max-pooling bên ngoài các khối Inception, trong đó có hai lớp nằm giữa các khối 3–4 và khối 7–8. Các lớp max-pooling này giúp giảm kích thước của dữ liệu đầu vào, do đó giảm độ phức tạp của mô hình cũng như chi phí tính toán.
* Mạng này kế thừa ý tưởng sử dụng lớp pooling trung bình từ NiN, giúp cải thiện hiệu suất mô hình và giảm hiện tượng overfitting.
* Một lớp dropout (với 40%) được sử dụng trước lớp linear. Đây cũng là một phương pháp regularization hiệu quả để giảm hiện tượng overfitting.
* Lớp đầu ra sử dụng hàm kích hoạt softmax để đưa ra 1000 đầu ra tương ứng với số lượng danh mục trong tập dữ liệu ImageNet.

Ngoài ra, một số mạng bổ sung được thêm vào bên cạnh, giúp khuyến khích sự phân biệt ở các giai đoạn thấp hơn trong bộ phân loại, tăng tín hiệu gradient nhận được backpropagation và cung cấp regularization bổ sung. Cấu trúc của các mạng này bao gồm:

* Một lớp pooling trung bình với kích thước pooling là 5 × 5 và stride là 3.
* Một lớp tích chập 1 × 1 với 128 bộ lọc để giảm chiều và kích hoạt RELU.
* Một lớp fully connected với 1024 units và kích hoạt RELU.
* Một dropout với tỷ lệ 70% đầu ra.
* Một lớp đầu ra sử dụng hàm kích hoạt softmax để phân loại đối tượng thành một trong 1000 classes.

A table of numbers and a number

AI-generated content may be incorrect.