

Depth Estimation

(a) Architecture of CNN+FC network



(b) Architecture of pure CNN network



(c) Architecture of CNN+Residual network



CNN + FC:

Mạng bao gồm hai phần, các lớp chập và các lớp FC, hình ảnh RGB đầu tiên đi qua các lớp tích

chập với các bộ lọc 11×11 , 5×5 , 3×3 . Chuẩn hóa hàng loạt, các lớp ReLU và các lớp gộp tối đa 2×2 tuân theo các lớp tích chập. Sau 6 lớp chập, dữ liệu đi qua 2 lớp được kết nối đầy đủ và đầu ra cuối cùng được thay đổi kích thước thành kích thước của bản đồ độ sâu chân lý mặt đất. Dropout được sử dụng sau lớp đầu tiên được kết nối đầy đủ để tránh trạng thái quá nhiều.

Dropout: kỹ thuật dropout là việc chúng ta sẽ bỏ qua một vài unit trong suốt quá trình train trong mô hình để chống over-fitting

BatchNorm : Trong gradient descent, mạng NN tính giá trị đạo hàm và giảm trọng số của nó dựa vào hướng đi của đạo hàm. Nhưng do các layer được xếp chồng lên nhau, phân phối của dữ liệu đầu vào sẽ bị thay đổi dần do việc cập nhật trọng số của các layer trước đó, làm cho phân phối của đầu vào của các layer phía sau sẽ khác xa so với phân phối của data input. BN giúp cố định phân phối của dữ liệu về

phân phối chuẩn, qua tất cả các lớp, dẫn tới tính chất phân phối của dữ liệu không thay đổi qua các lớp.

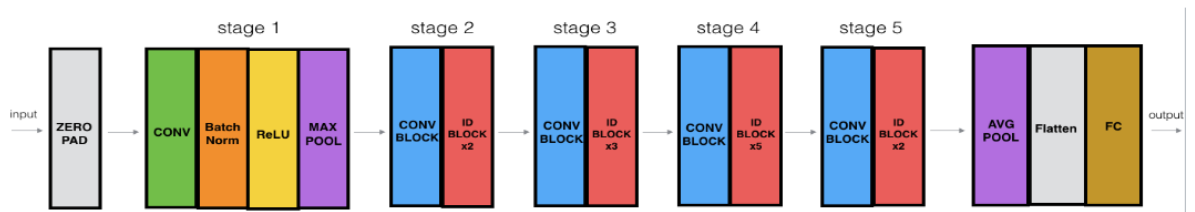
Pure CNN:

Để khắc phục sự cố overfitting, chúng tôi thay thế các lớp fully connected bằng các lớp chập

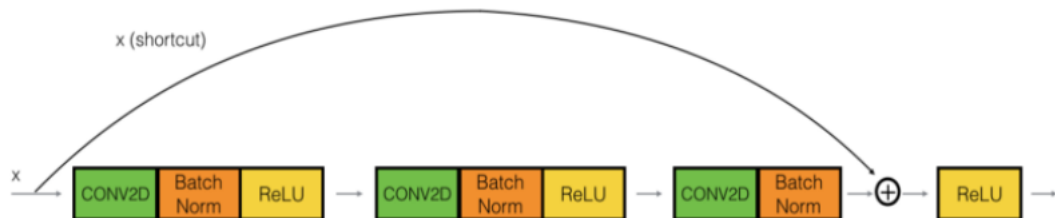
Mạng tích chập này bao gồm 8 lớp tích chập, và quá trình chuẩn hóa hàng loạt và kích hoạt ReLU sẽ hạ thấp mỗi lớp tích chập

CNN + Residual:

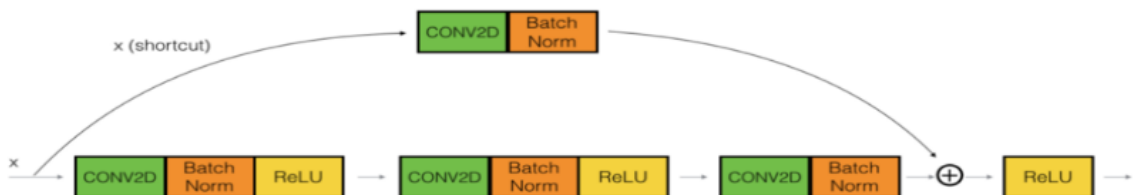
ResNet50:



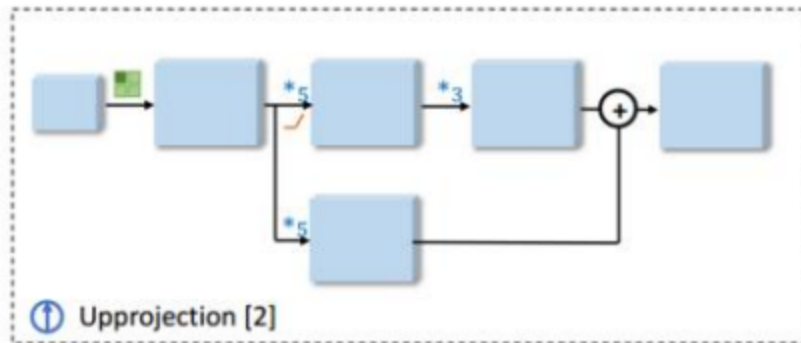
The identity block:



The convolutional block:



Upprojection:



Trong mạng ResNet-50