https://arxiv.org/pdf/1801.04381

1. Hàm loss và thuật toán tối ưu

nn.CrossEntropyLoss():

tính xác xuất softmax : từng kết quả chưa chuẩn hóa chia cho tổng các kết quả chưa chuẩn hóa

tính Negative Log-Likelihood (NLL): nếu mô hình càng gần 1 thì cái này càng nhỏ ngược lại càng xa 1 nghĩa là không chắc thì sẽ lớn

2 Hàm tối ưu adam (dùng cho tất cả mô hình ) có ý tưởng là dùng momentum(nhớ hướng) + RMSProp(tối ưu lr theo từng trọng số theo 1 công thức quy ước)

3 Các đặc điểm phổ biến : residual structure ( phần dư đảo ngược ) nghĩa là tăng chiều → xử lý → giảm chiều (khác với ResNet). **Shortcut connections between the thin bottleneck layers**: Có nghĩa là kết nối tắt không đi qua toàn bộ block mà **nối trực tiếp đầu vào và đầu ra (hai lớp mỏng)** nếu chúng có cùng kích thước. Dẫn chứng mực Abstract đoạn 2