



HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



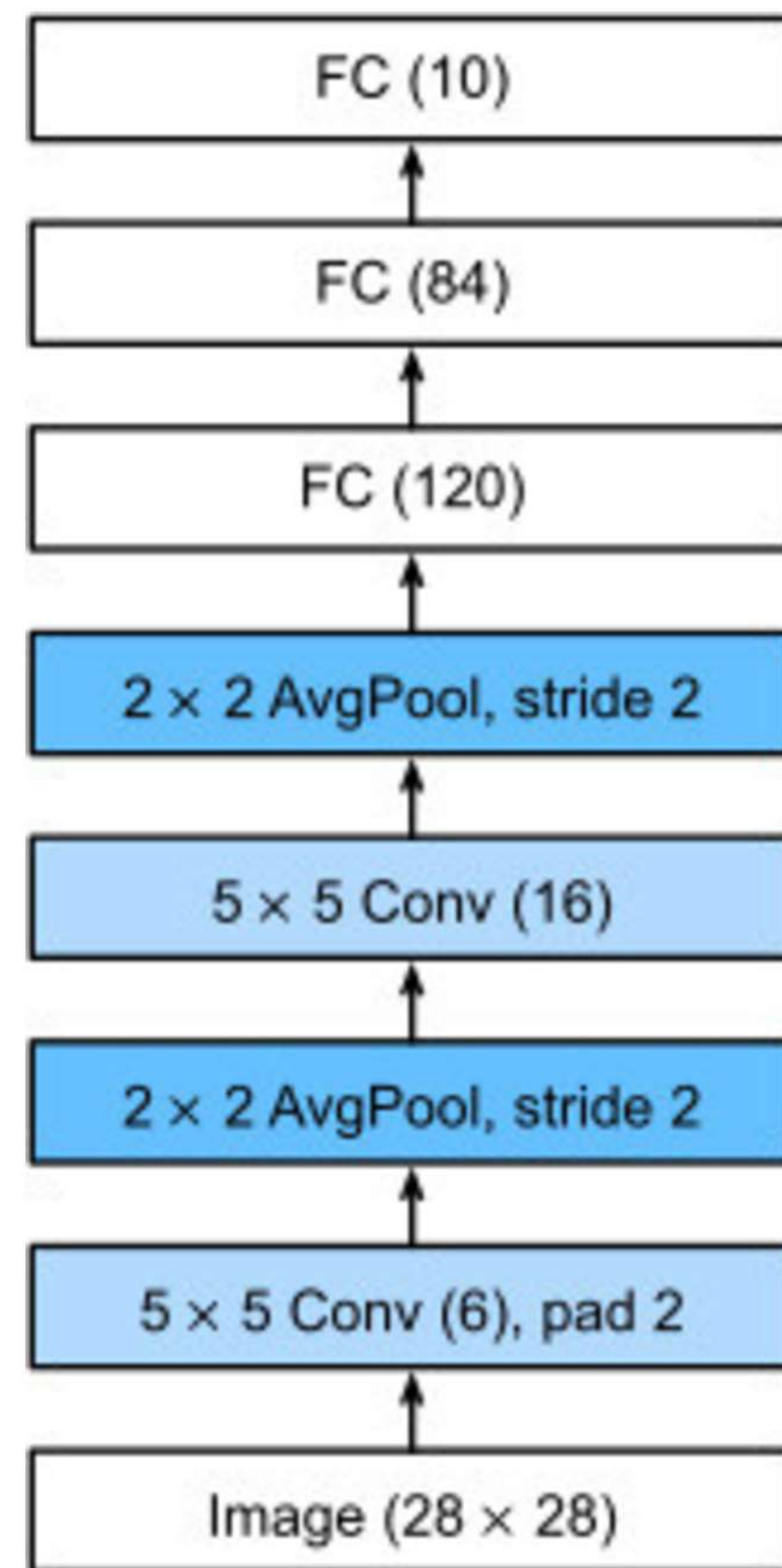
**ĐẠI HỌC
BÁCH KHOA HÀ NỘI**
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Các kiến trúc CNN cơ bản

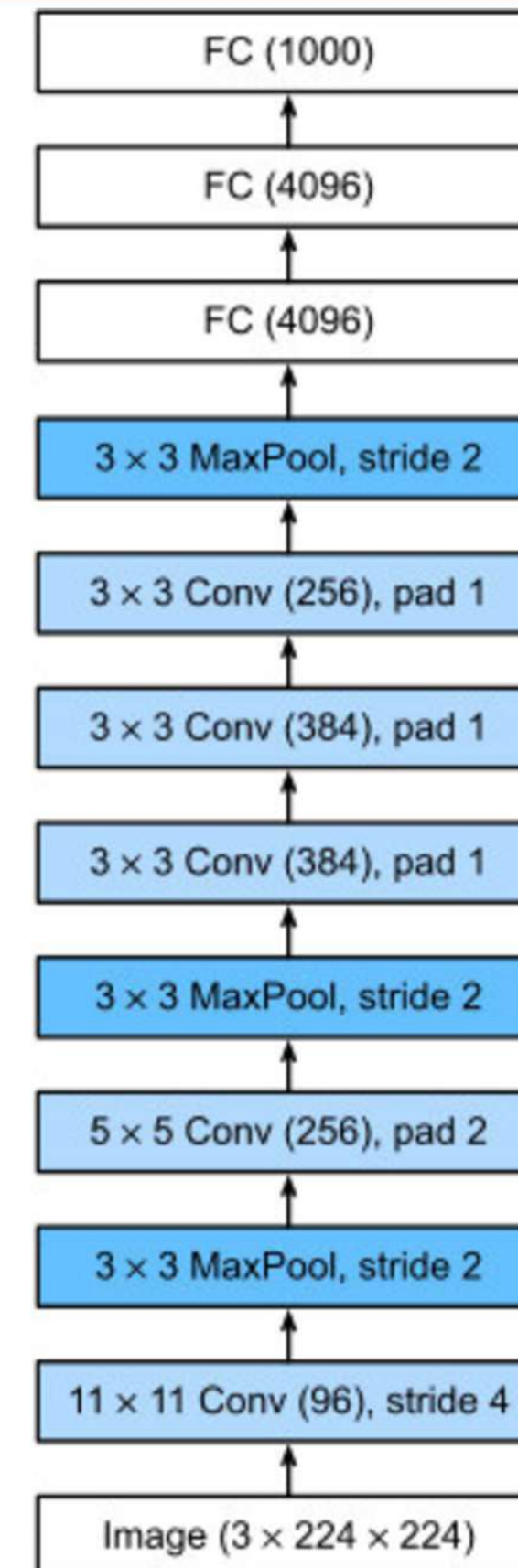
ONE LOVE. ONE FUTURE.

- Các kiến trúc CNN cơ bản:
 - LeNet
 - AlexNet
 - NiN
 - VGG
- Ứng dụng trong bài toán thực tế

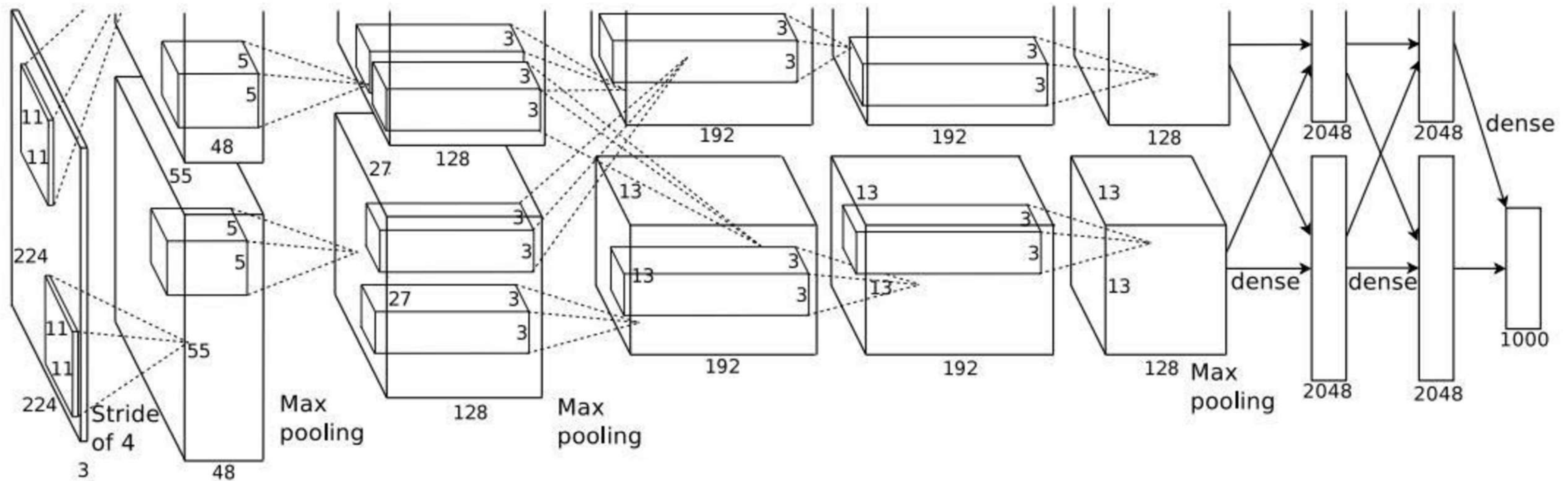
- LeNet ra đời năm 1998 giải quyết bài toán nhận dạng chữ số viết tay
- LeNet là mạng CNN đầu tiên hoạt động thực tế thương mại khi được ngân hàng Mỹ sử dụng trong hệ thống đọc séc



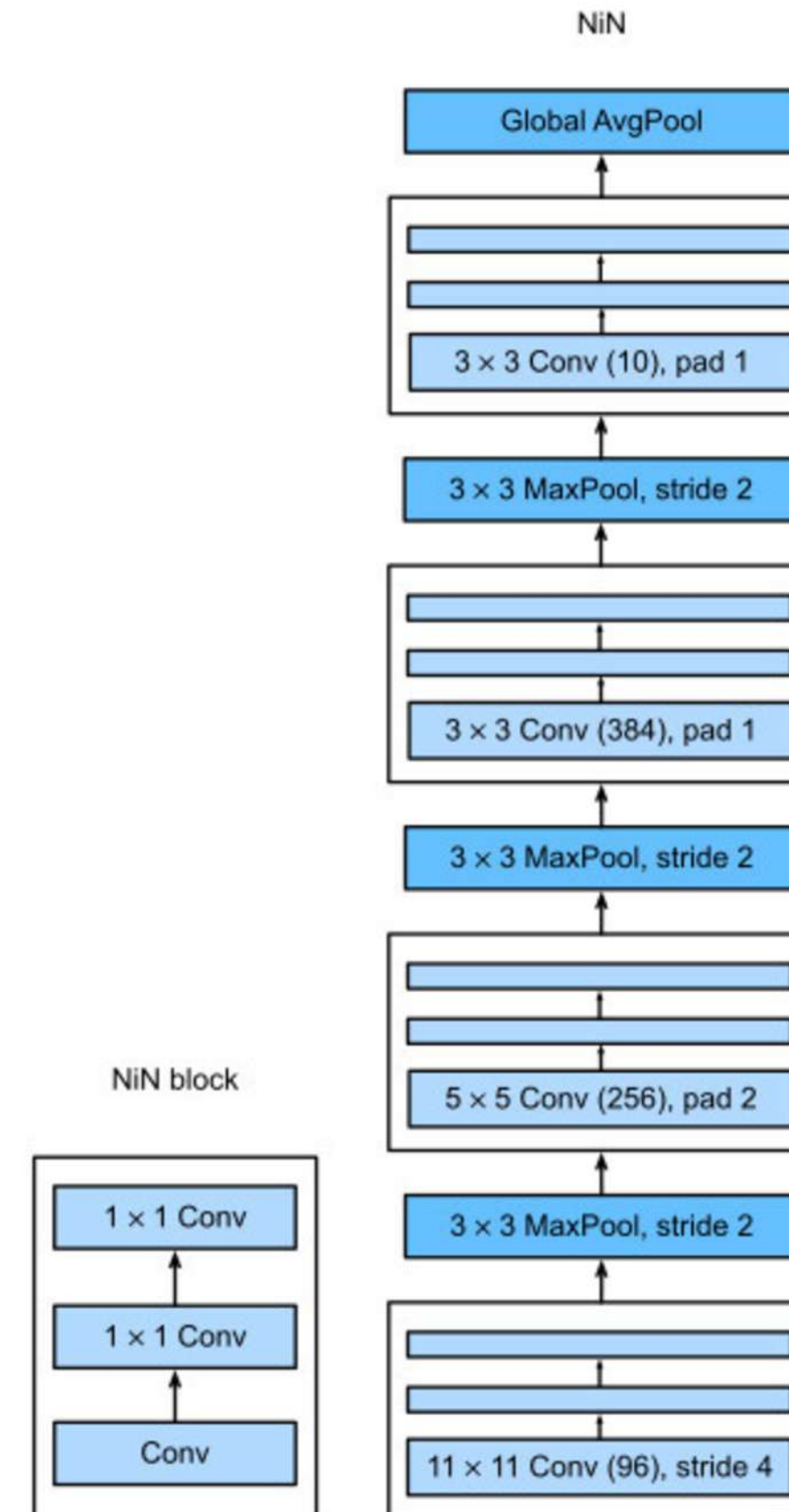
- AlexNet ra đời năm 2012 là mạng CNN đầu tiên được huấn luyện hiệu quả trên tập dữ liệu khổng lồ ImageNet (1.2 triệu ảnh, 1000 lớp)
- AlexNet chứng minh rằng deep learning trên GPU có thể vượt trội hoàn toàn so với các phương pháp truyền thống trong nhận dạng ảnh quy mô lớn, mở ra kỷ nguyên deep learning hiện đại



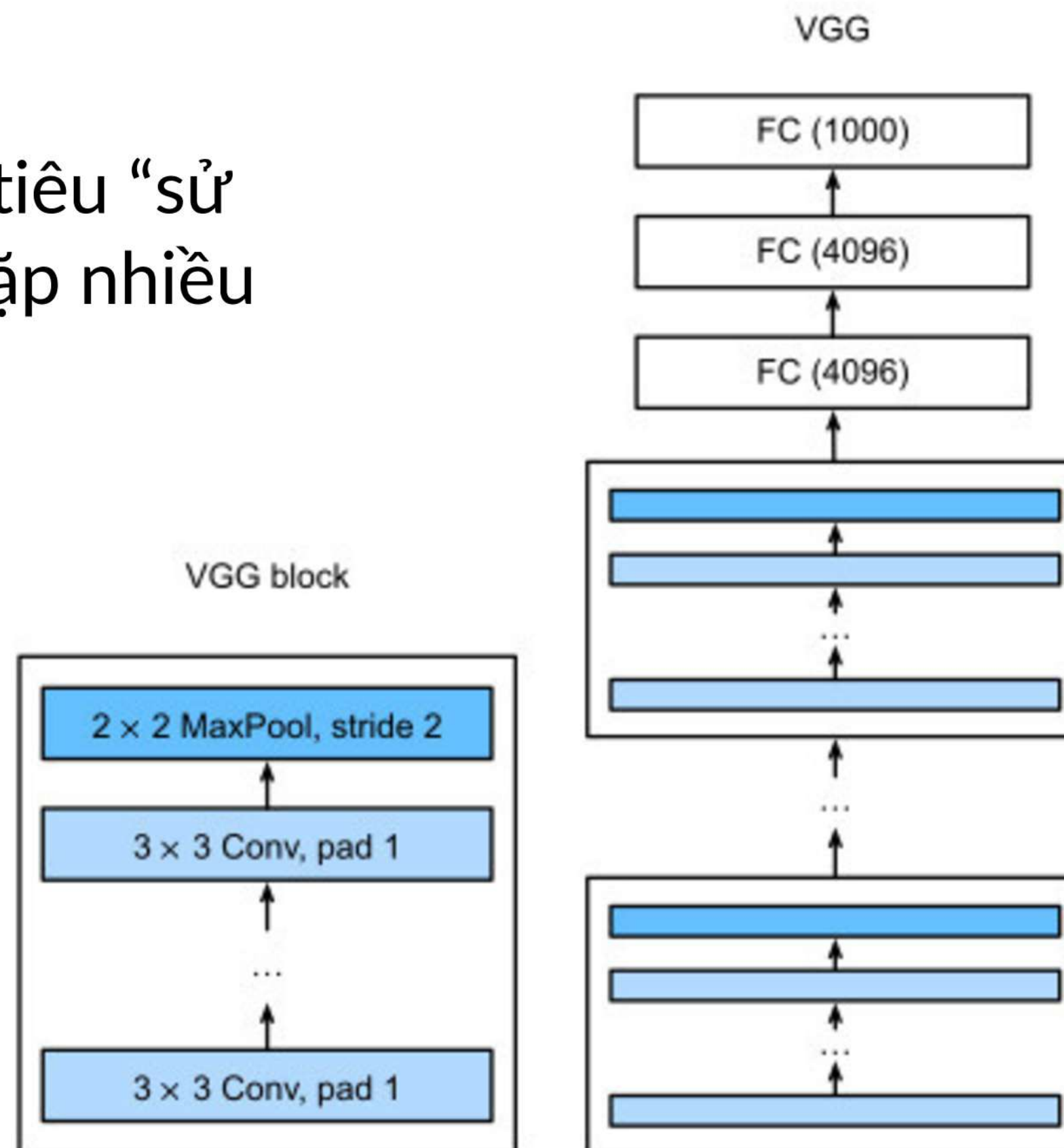
- Cấu trúc mạng AlexNet gốc năm 2012



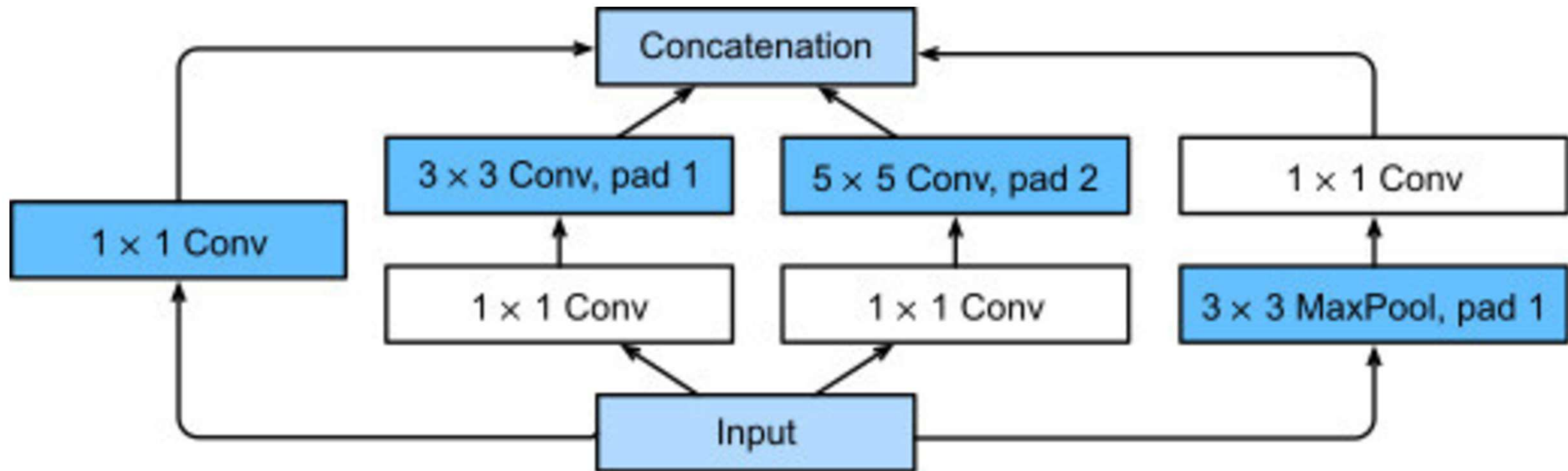
- NiN hay Network in Network ra đời năm 2013 bổ sung thêm một mạng con phi tuyến MLP sau các lớp convolution giúp tăng khả năng biểu diễn phi tuyến
- NiN sử dụng thêm cả GlobalAvgPool để giảm tham số mô hình

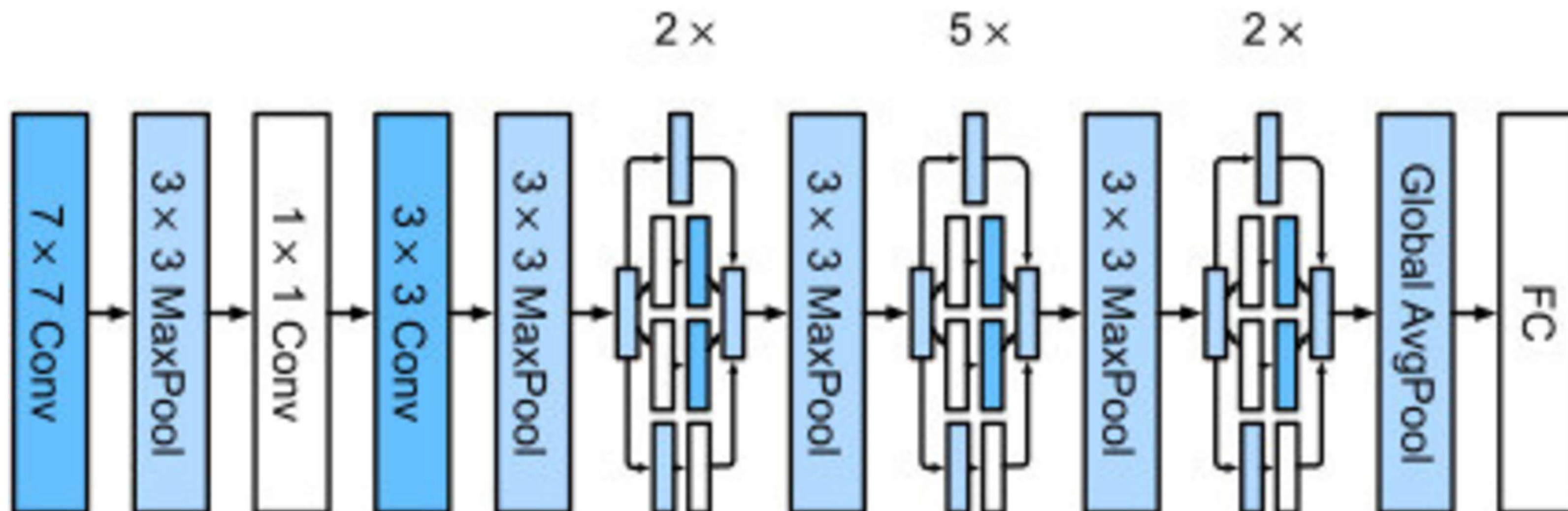


- VGG ra đời năm 2014 với mục tiêu “sử dụng cùng một khối CNN nhỏ lặp nhiều lần tạo mạng sâu và đồng nhất”

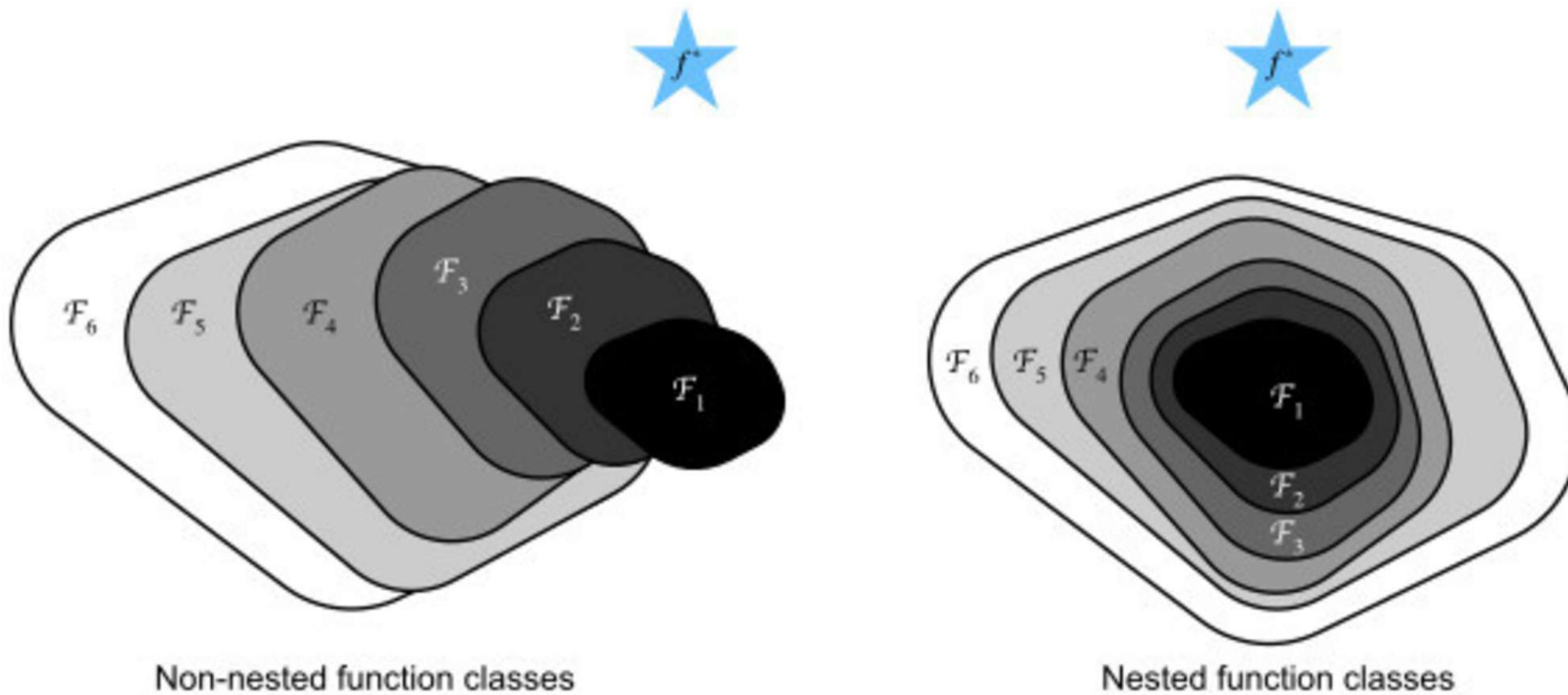


- GoogLeNet ra đời năm 2014, đề xuất một loại block là Inception block

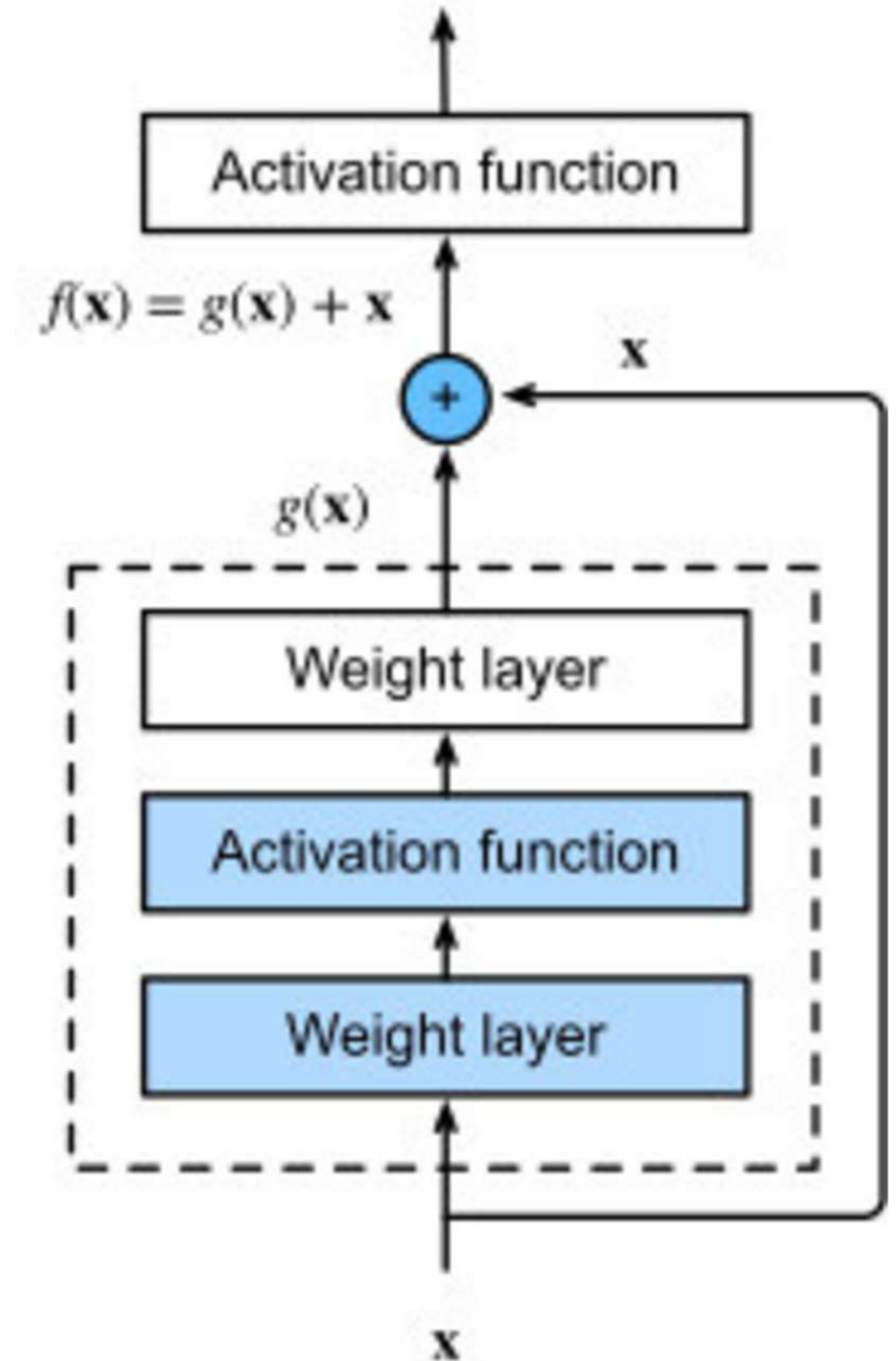
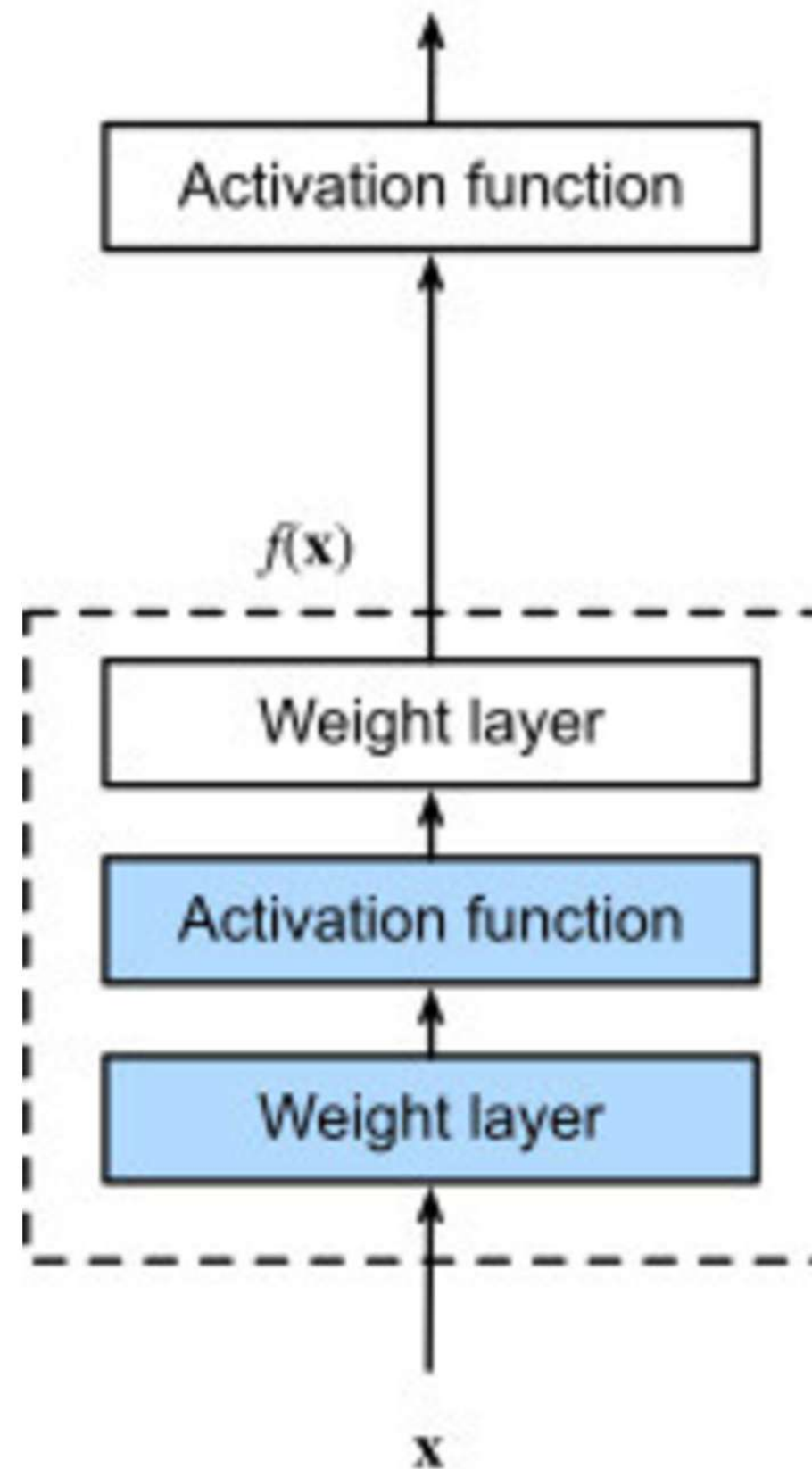




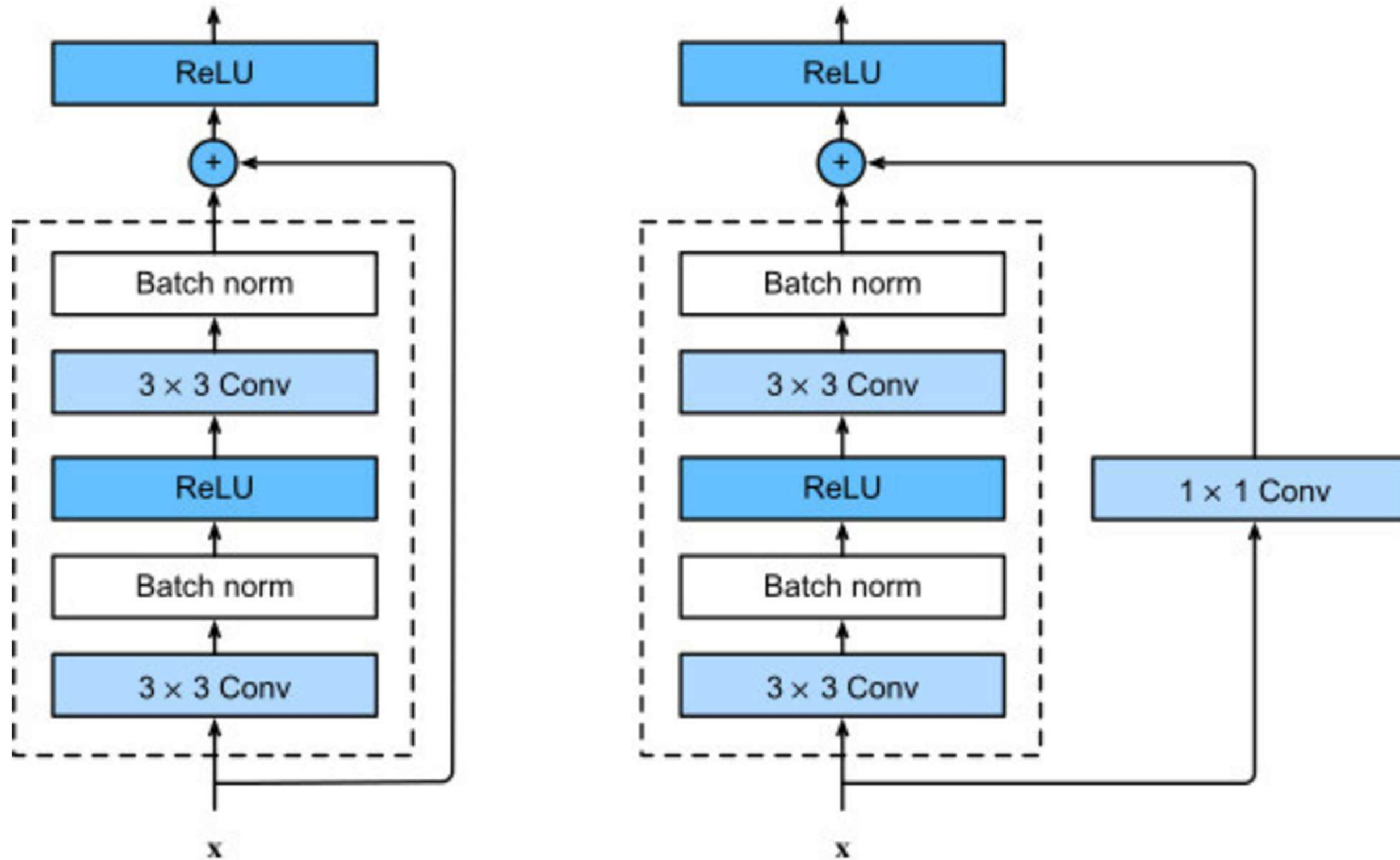
$$f^o = \underset{f \in F}{\operatorname{argmin}} L(f(x_i), y_i)$$



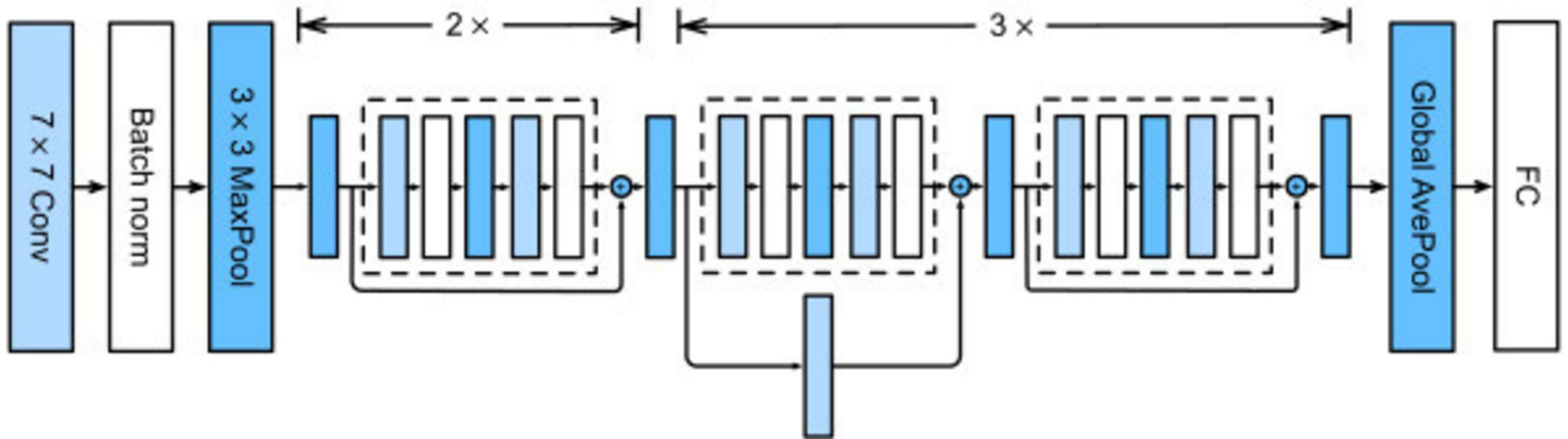
- Residual block đảm bảo mạng vẫn có thể hoạt động y như cũ nếu tầng mới không học được gì hữu ích



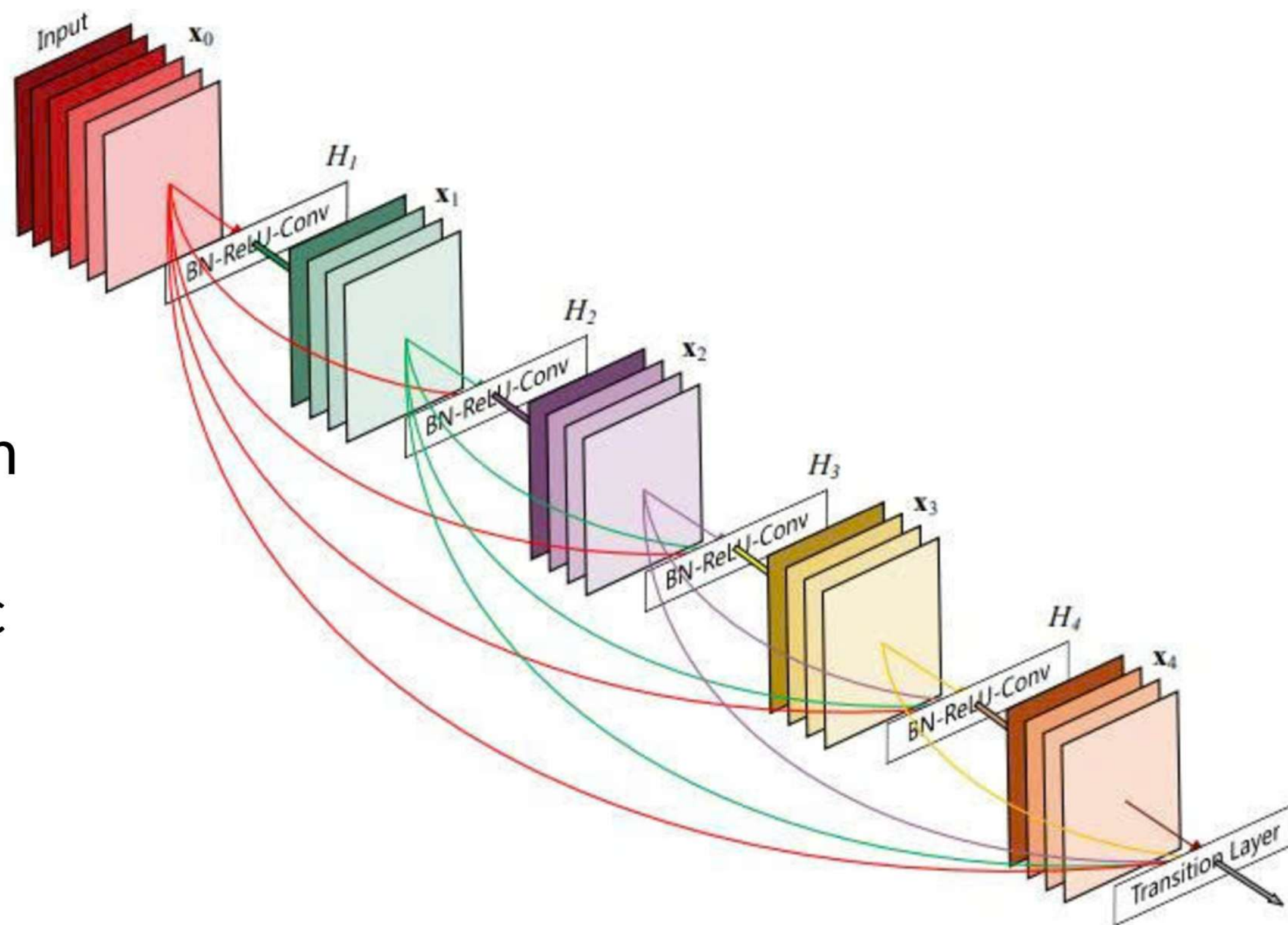
ResNet



ResNet



- DenseNet ra đời năm 2017 với ý tưởng: mỗi tầng kết nối trực tiếp tới tất cả các tầng trước nó



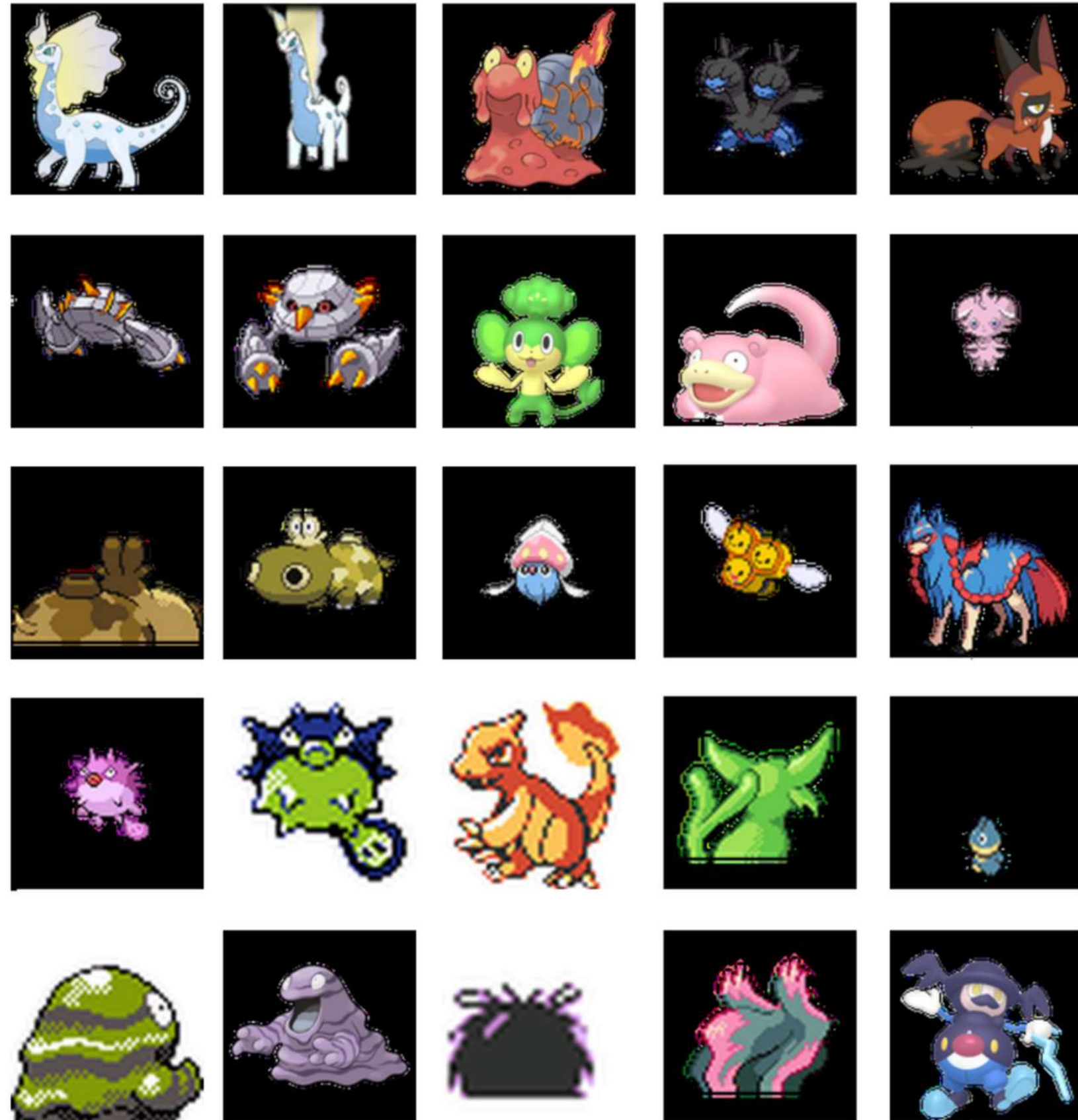
DenseNet

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

- Từ 2017 trở đi, các mô hình CNN chủ yếu cải tiến hiệu năng, tối ưu hóa và tự động hóa thiết kế hơn là thay đổi bản chất. Không có “bước nhảy lý thuyết” lớn như ResNet hay DenseNet
- Năm 2020, Vision Transformer ra đời, mở ra một kỷ nguyên mới, cạnh tranh và dần vượt qua CNN trong nhiều tác vụ thị giác
- Các mô hình hiện đại dần chuyển sang hướng hybrid CNN + Transformer

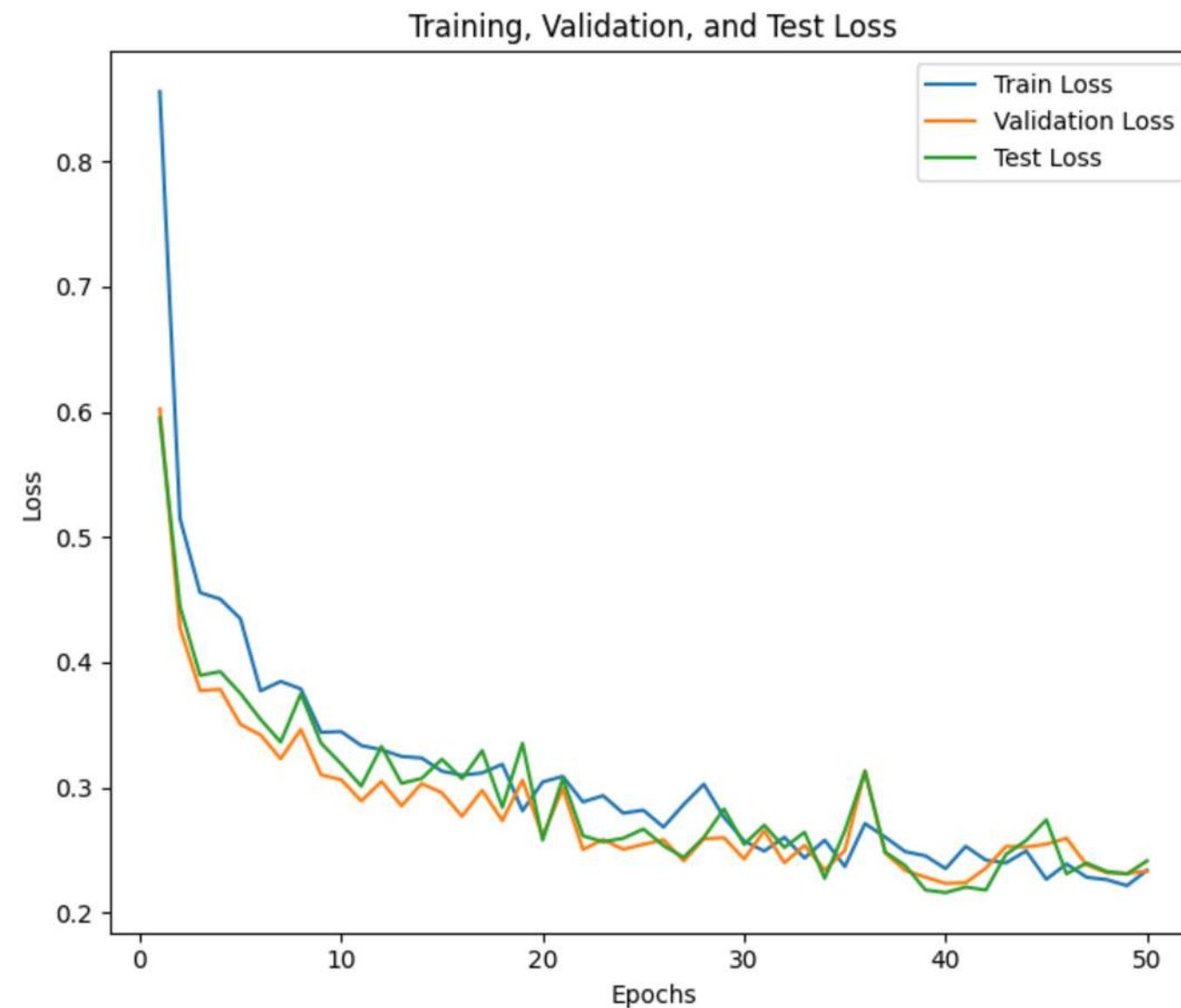
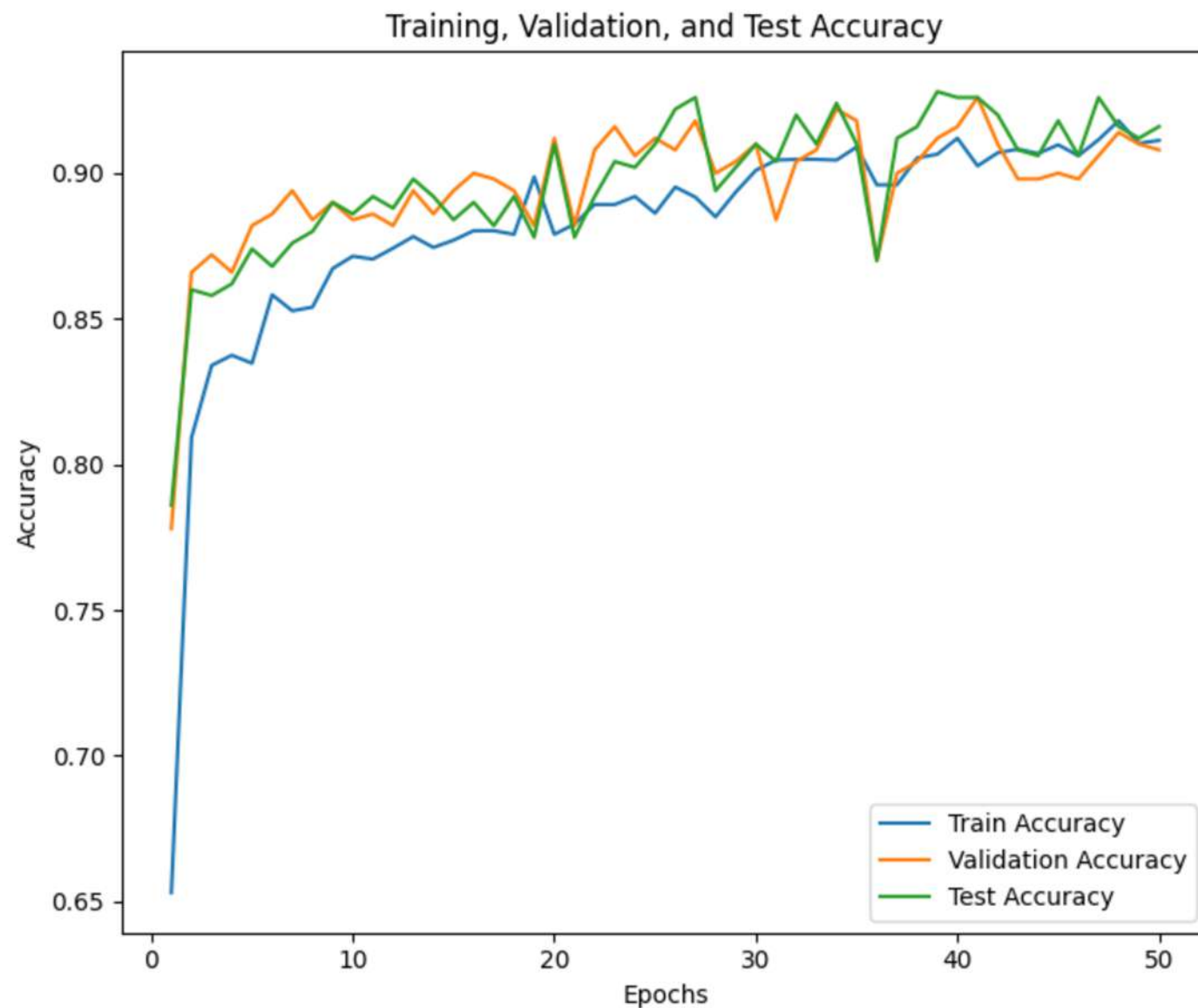
Ứng dụng trong bài toán thực tế

- Bộ dữ liệu: Pokemon Images Dataset
- Bộ dữ liệu gồm 400000 bức ảnh của 1000 loại pokemon khác nhau
- Ảnh RGB kích thước 224x224



Ứng dụng trong bài toán thực tế

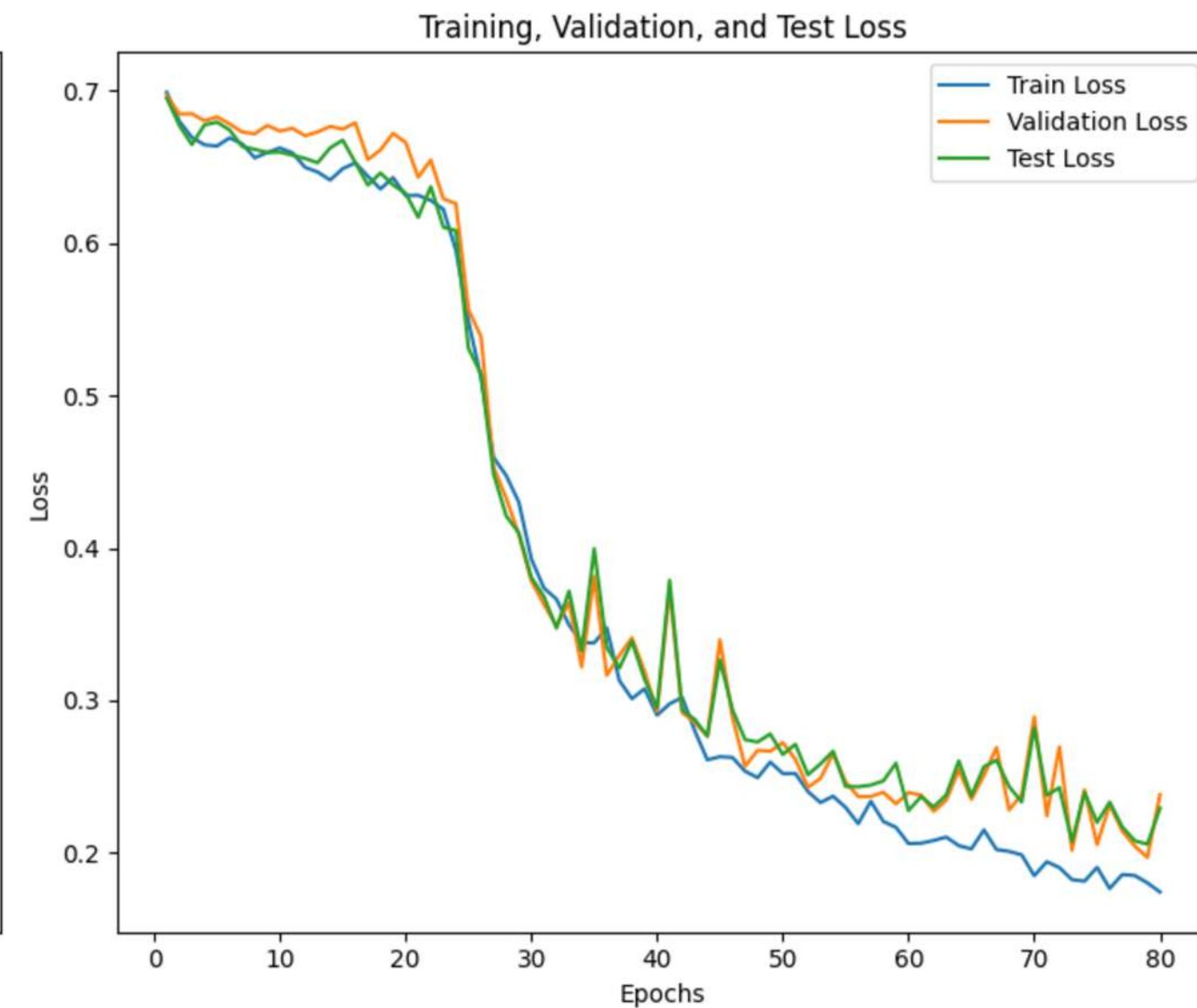
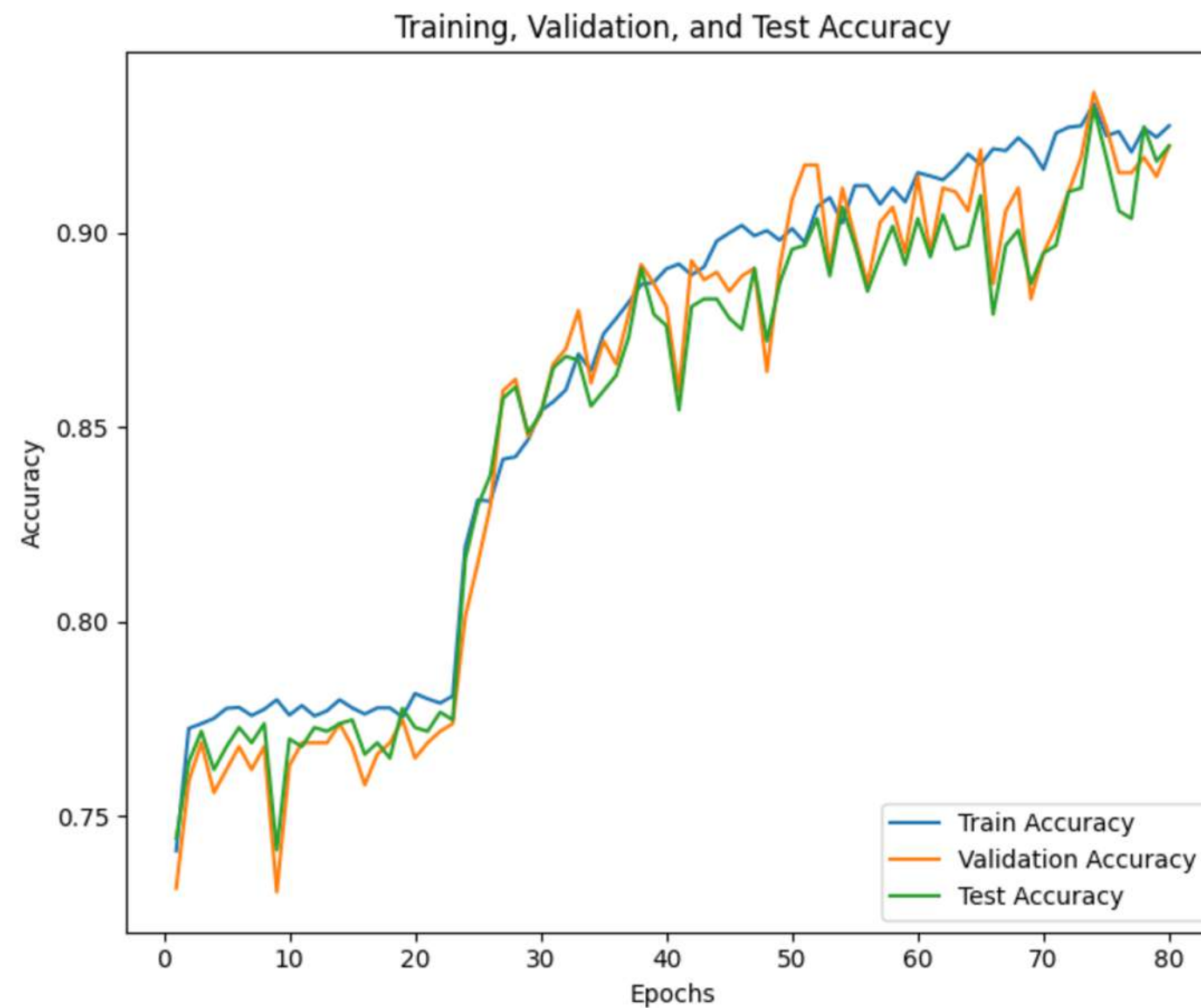
- AlexNet
 - Validation set - Best Loss: 0.2232 - Best Accuracy: 0.9260
 - Test set - Best Loss: 0.2159 - Best Accuracy: 0.9280



Ứng dụng trong bài toán thực tế

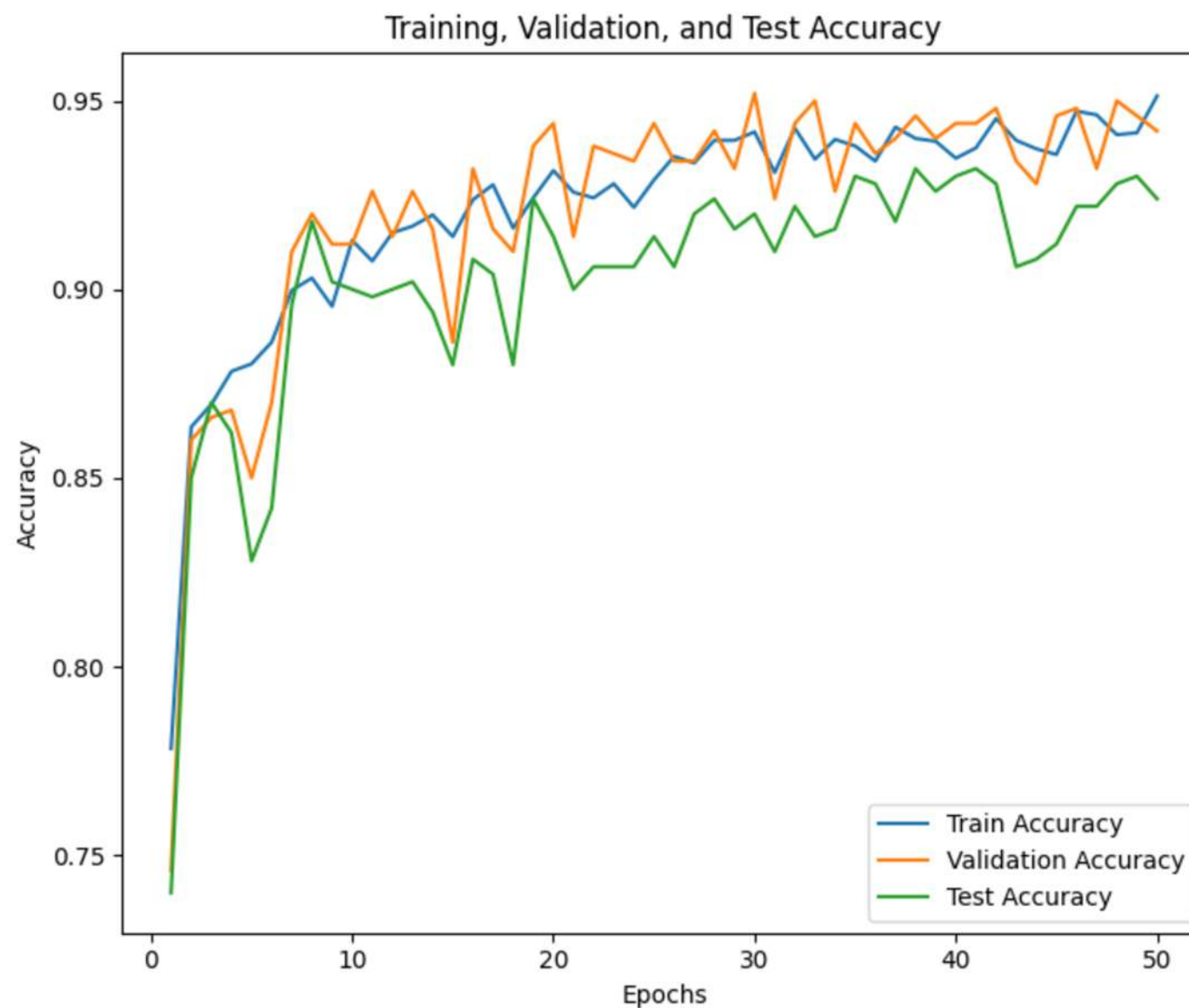
- NiN

- Validation set - Best Loss: 0.2008 - Best Accuracy: 0.9280
- Test set - Best Loss: 0.2030 - Best Accuracy: 0.9340



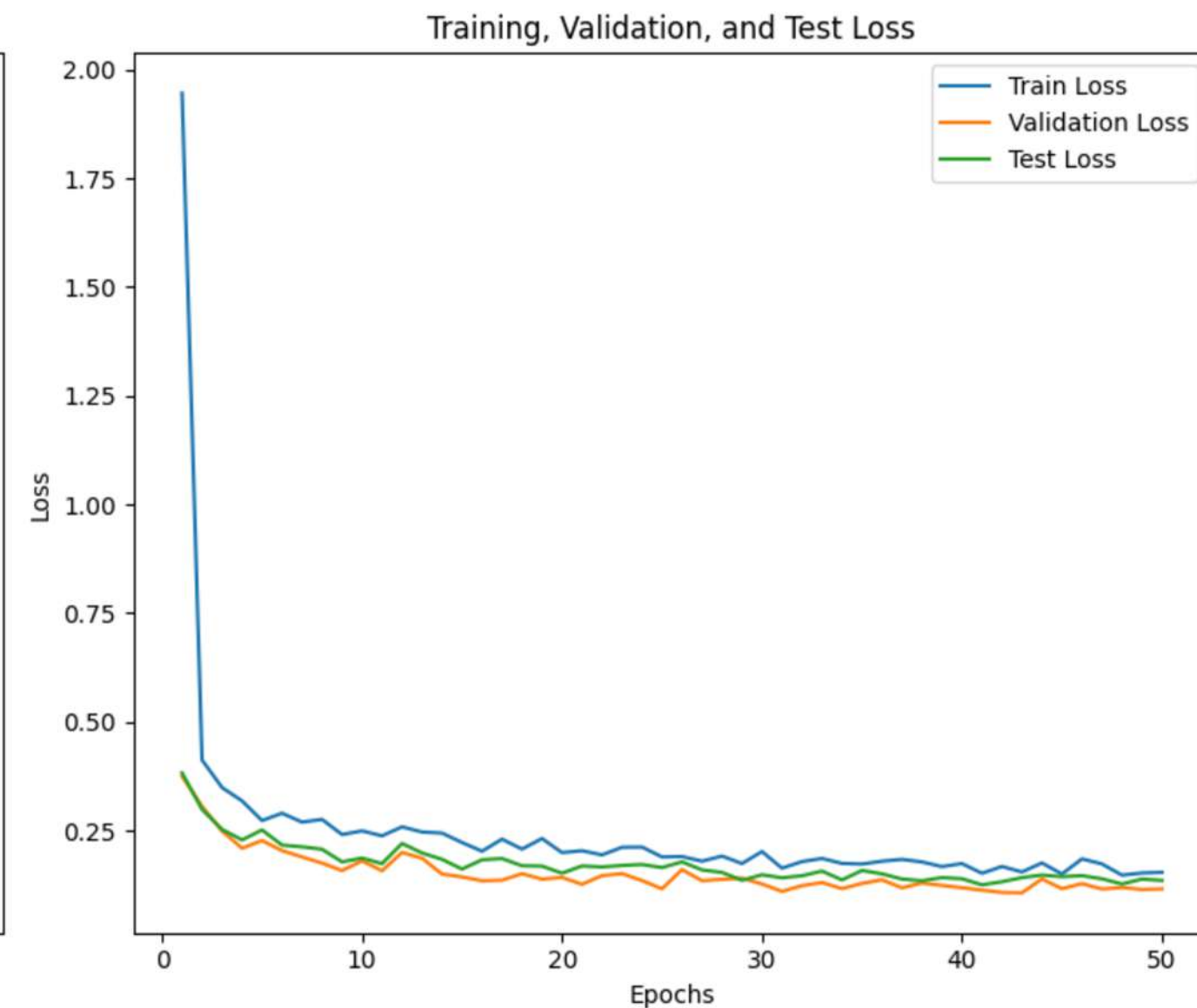
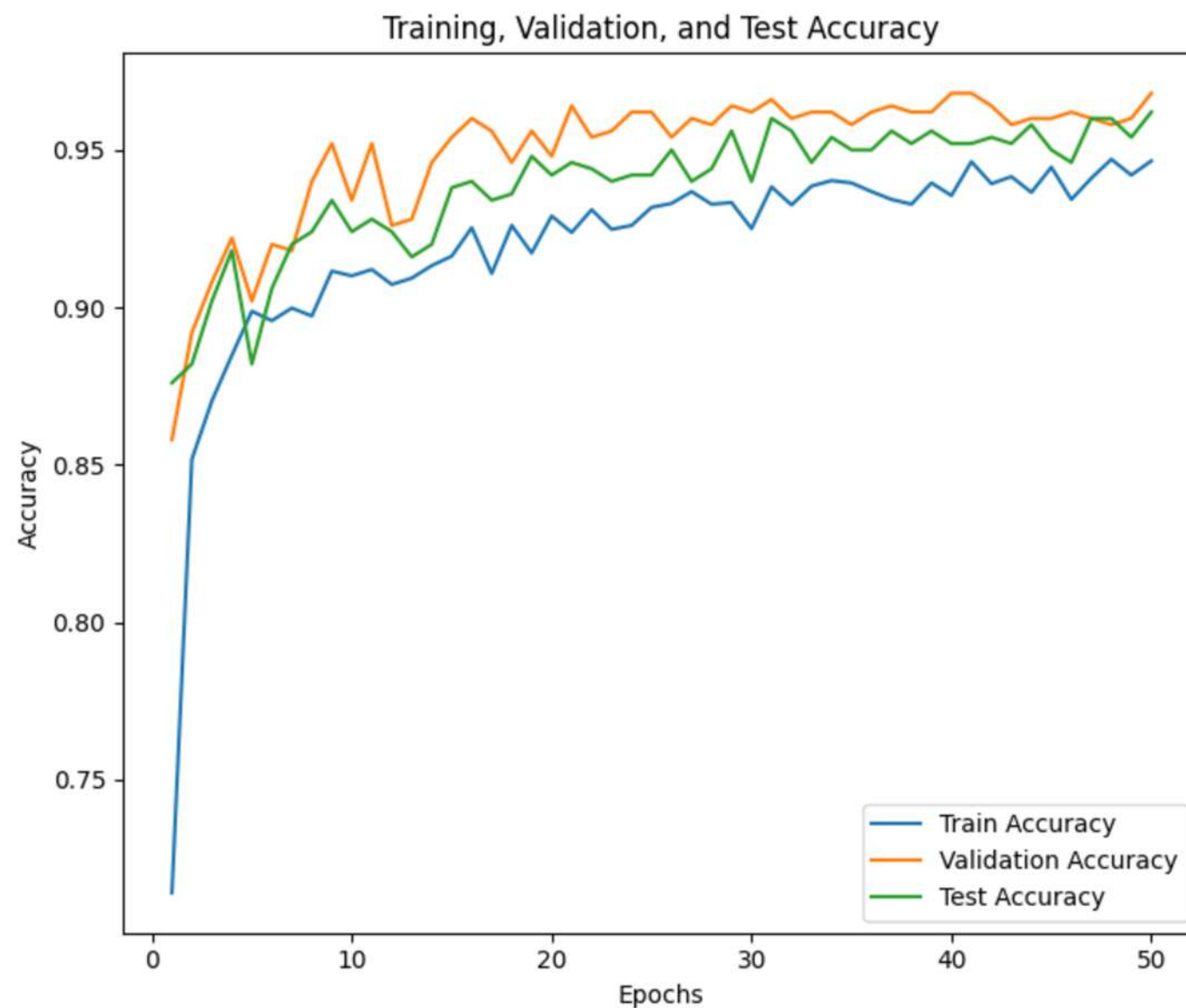
- VGG16

- Validation set - Best Loss: 0.1345 - Best Accuracy: 0.9520
- Test set - Best Loss: 0.1890 - Best Accuracy: 0.9370



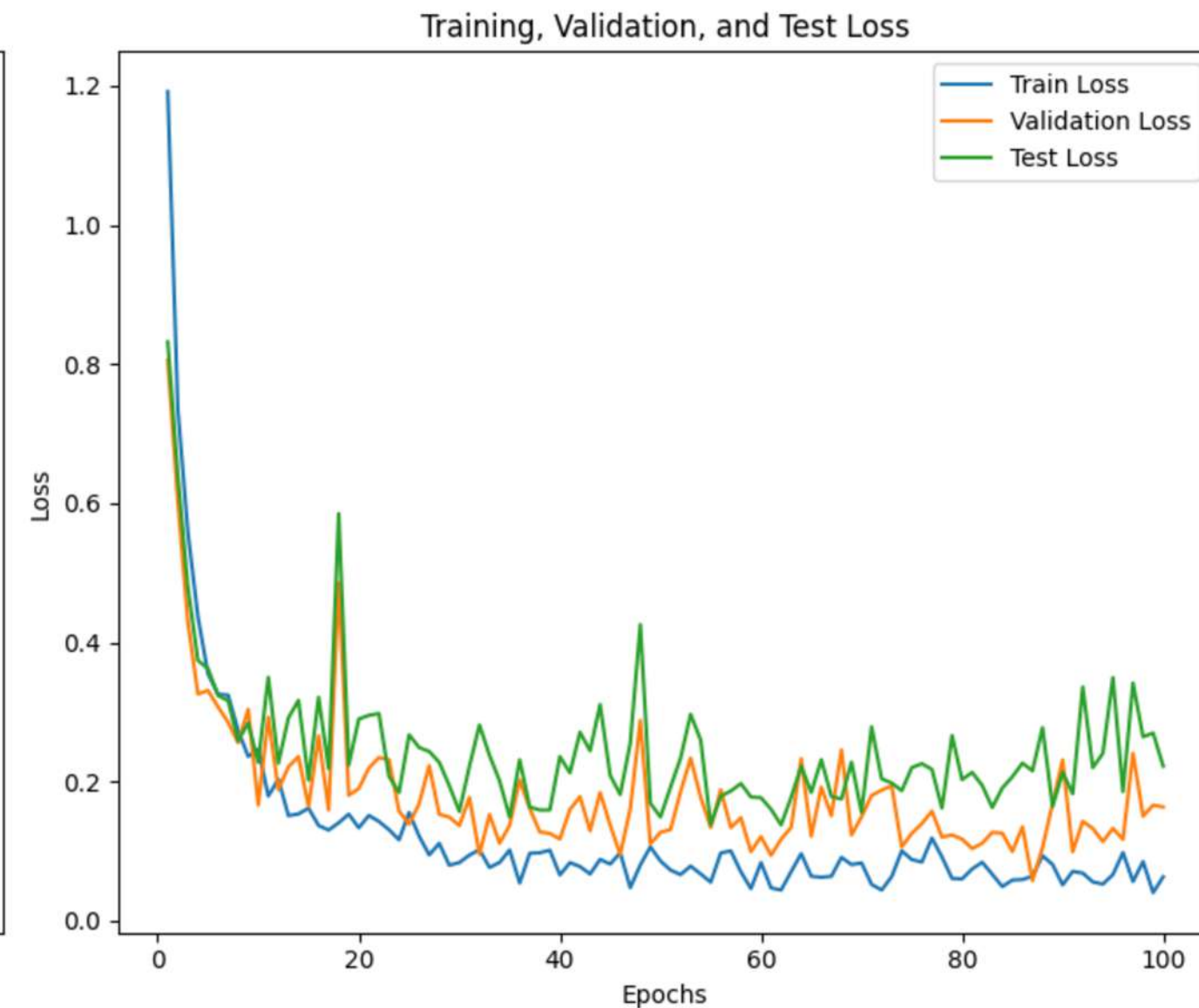
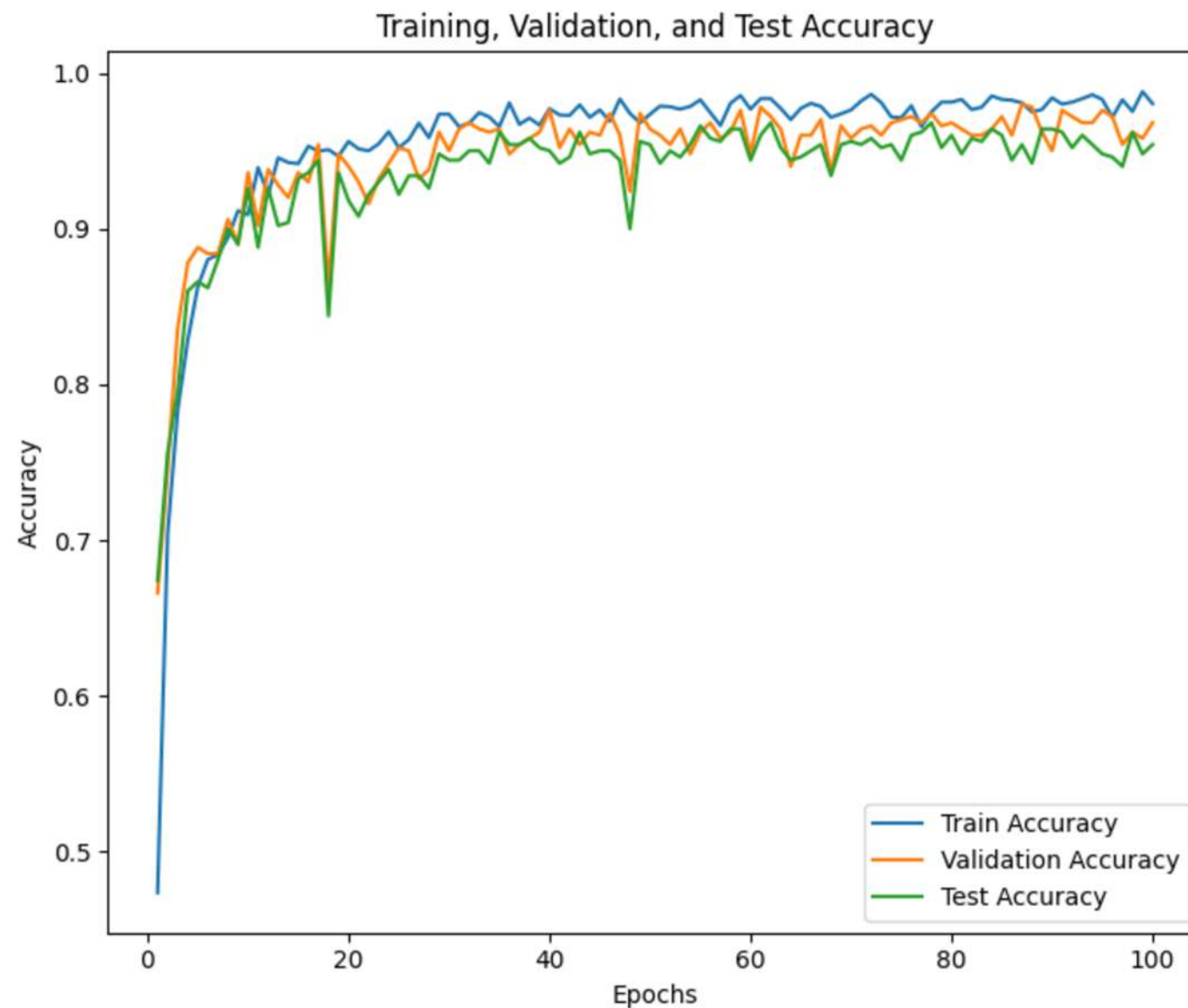
Ứng dụng trong bài toán thực tế

- GoogLeNet
 - Validation set - Best Loss: 0.1069 - Best Accuracy: 0.9680
 - Test set - Best Loss: 0.1377 - Best Accuracy: 0.9620



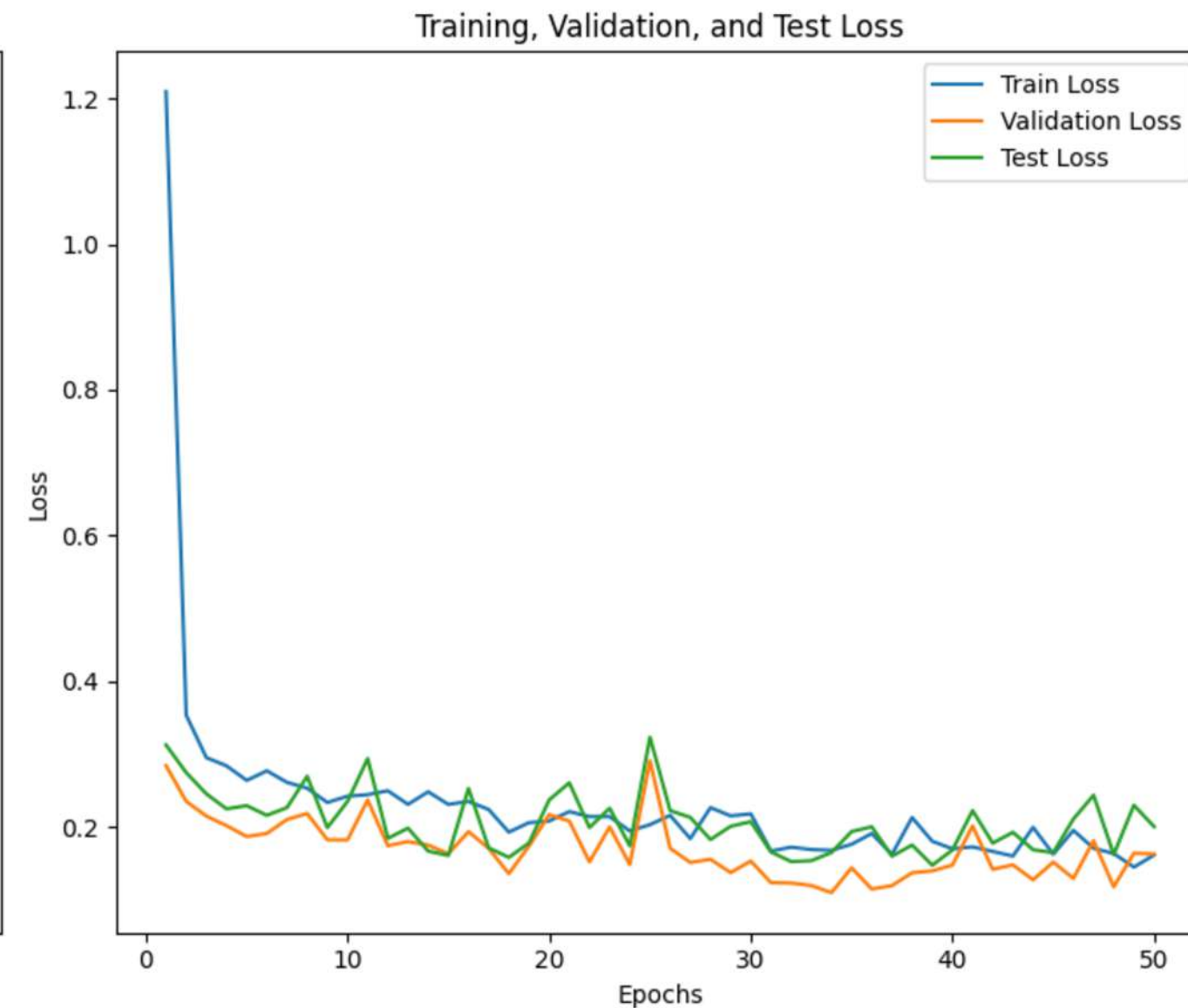
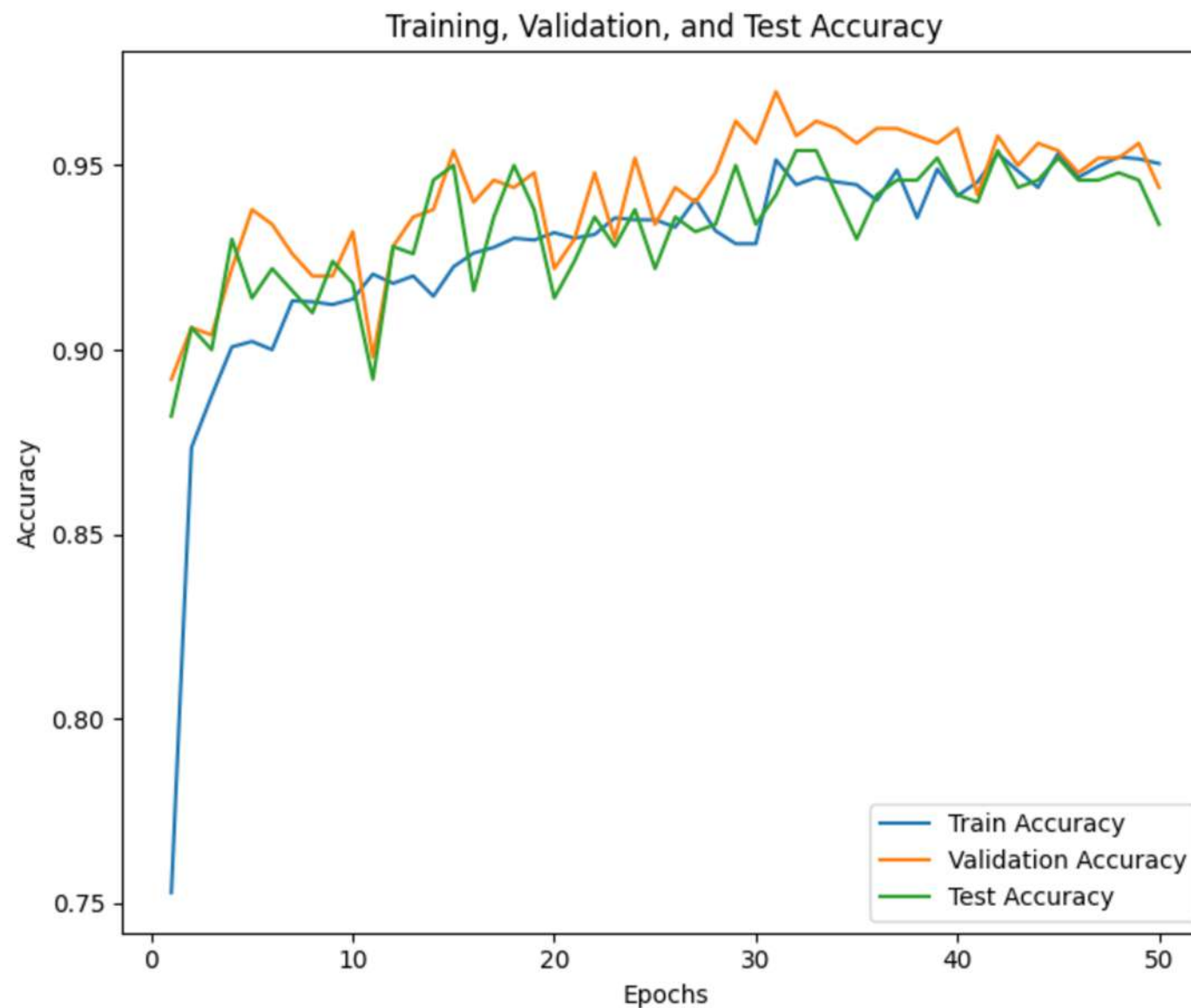
- ResNet52

- Validation set - Best Loss: 0.0576 - Best Accuracy: 0.9800
- Test set - Best Loss: 0.1252 - Best Accuracy: 0.9680



- DenseNet121

- Validation set - Best Loss: 0.1097 - Best Accuracy: 0.9700
- Test set - Best Loss: 0.1434 - Best Accuracy: 0.9540



Ứng dụng trong bài toán thực tế

Model	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
AlexNet	0.2232	0.9260	0.2159	0.9280
NiN	0.2008	0.9280	0.2030	0.9340
VGG16	0.1345	0.9520	0.1890	0.9370
GoogLeNet	0.1069	0.9680	0.1377	0.9620
ResNet52	0.0576	0.9800	0.1252	0.9680
DenseNet121	0.1097	0.9700	0.1434	0.9540

A large graphic on the left side of the slide. It features a dark blue background with a pattern of red dots of varying sizes arranged in concentric, slightly irregular circles, creating a sense of depth and movement. The word "HUST" is centered within this graphic in a white, bold, sans-serif font.

HUST

THANK YOU !



hust.edu.vn



fb.com/dhbkhn