



HUST

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

ONE LOVE. ONE FUTURE.



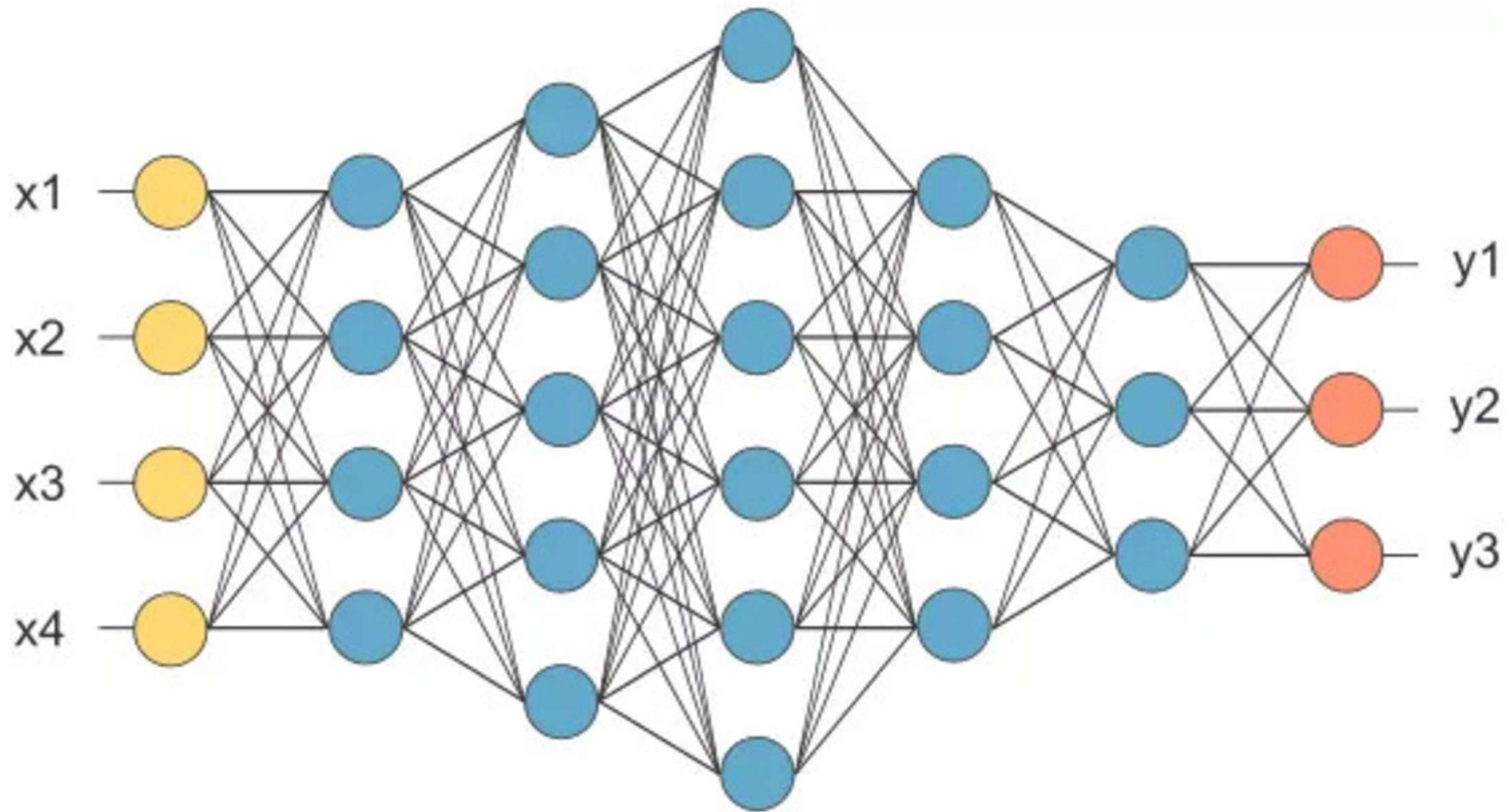
**ĐẠI HỌC
BÁCH KHOA HÀ NỘI**
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Convolutional Neural Networks

ONE LOVE. ONE FUTURE.

- Convolution
 - Translation Equivariance
 - Locality
 - Cross-Correlation
- Padding, Stride
- Pooling
- Ứng dụng bài toán thực tế

Convolution



- Convolutional layer (lớp tích chập) được thiết kế đặc biệt cho dữ liệu có cấu trúc dạng lưới như hình ảnh với hai tính chất quan trọng:
 - Translation Equivariance: Có thể nhận ra vùng ảnh dù ở các vị trí khác nhau
 - Locality: Tập trung vào các điểm ảnh ở gần hơn là điểm ảnh ở xa

$$\begin{aligned} H(i, j) &= b(i, j) + \sum_m \sum_n W(i, j, m, n) X(m, n) \\ &= b(i, j) + \sum_a \sum_b W^o(i, j, a, b) X(i + a, j + b) \end{aligned}$$

$$\begin{aligned} H(i, j) &= b(i, j) + \sum_a \sum_b W^o(i, j, a, b) X(i + a, j + b) \\ &= b^o + \sum_a \sum_b W^o(a, b) X(i + a, j + b) \end{aligned}$$

$$\begin{aligned} H(i, j) &= b^o + \sum_a \sum_b W^o(a, b) X(i + a, j + b) \\ &= b^o + \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} W^o(a, b) X(i + a, j + b) \end{aligned}$$

Cross-Correlation

Input Kernel Output

0	1	2
3	4	5
6	7	8

*

0	1
2	3

=

19	25
37	43

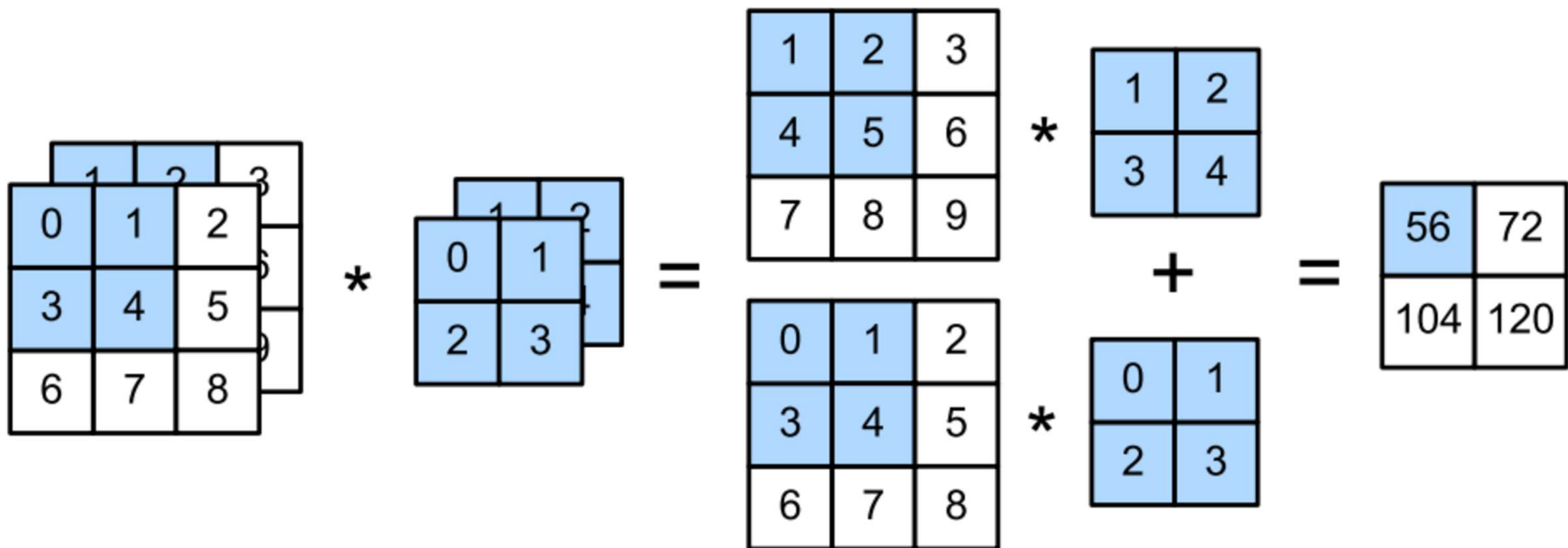
- Công thức của Cross-Correlation

$$H(i, j) = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} W^o(a, b) X(i + a, j + b)$$

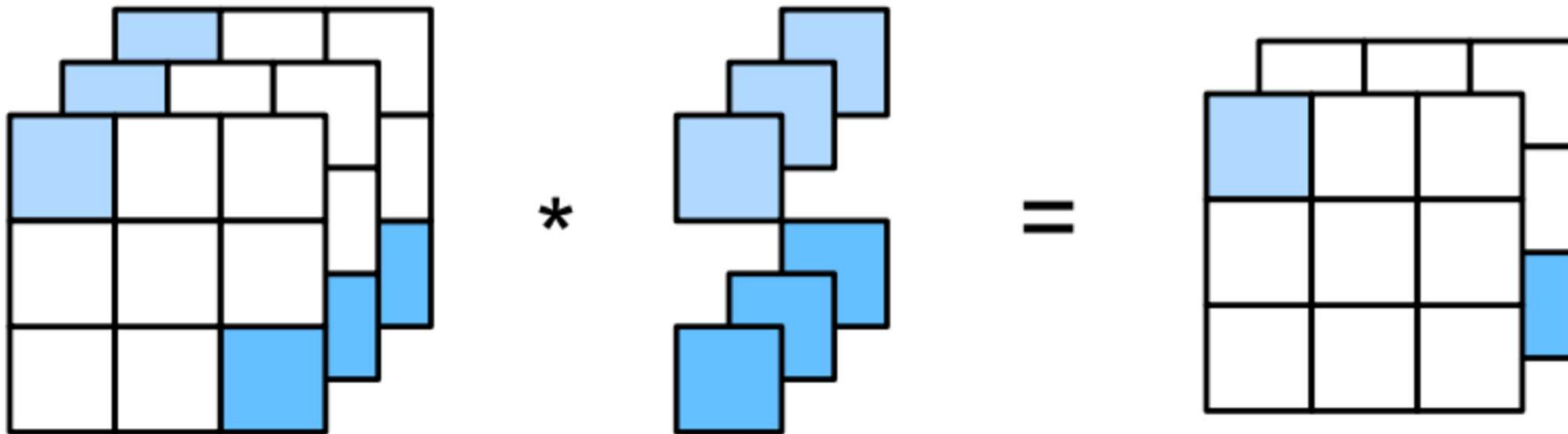
- Công thức của Convolution

$$H(i, j) = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} W^o(a, b) X(i - a, j - b)$$

Cross-Correlation



Cross-Correlation



Padding, Stride

0	1	2
3	4	5
6	7	8

$n_h \times n_w$

*

0	1
2	3

$k_h \times k_w$

=

19	25
37	43

$(n_h - k_h + 1) \times (n_w - k_w + 1)$

- Padding và Stride là hai phương pháp phổ biến để có thể chủ động điều chỉnh kích thước đầu ra

Padding

- Padding là phương pháp thêm các pixel bổ sung xung quanh biên của ảnh đầu vào, nhờ đó làm tăng kích thước ảnh đầu ra

$$\begin{array}{|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 2 & 0 \\ \hline 0 & 3 & 4 & 5 & 0 \\ \hline 0 & 6 & 7 & 8 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|} \hline 0 & 1 \\ \hline 2 & 3 \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|c|c|} \hline 0 & 3 & 8 & 4 \\ \hline 9 & 19 & 25 & 10 \\ \hline 21 & 37 & 43 & 16 \\ \hline 6 & 7 & 8 & 0 \\ \hline \end{array}$$

$(n_h + p_h) \times (n_w + p_w)$ $k_h \times k_w$ $(n_h - k_h + p_h + 1) \times (n_w - k_w + p_w + 1)$

- Các dạng padding phổ biến:
 - Zero padding: giả định đơn giản “ngoài biên là trống rỗng”
 - Reflection padding: tránh tạo “đứt gãy” ở biên, giữ tính liên tục về độ sáng và màu
 - Replication padding: coi như pixel biên kéo dài ra ngoài
 - Circular padding: xử lý cho dữ liệu chu kỳ, tuần hoàn

Stride

- Stride là phương pháp di chuyển kernel nhiều hơn một phần tử mỗi lần để tăng hiệu quả tính toán hoặc để kích thước ảnh đầu ra

0	0	0	0	0
0	0	1	2	0
0	3	4	5	0
0	6	7	8	0
0	0	0	0	0

*

0	1
2	3

$k_h \times k_w$

=

0	8
6	8

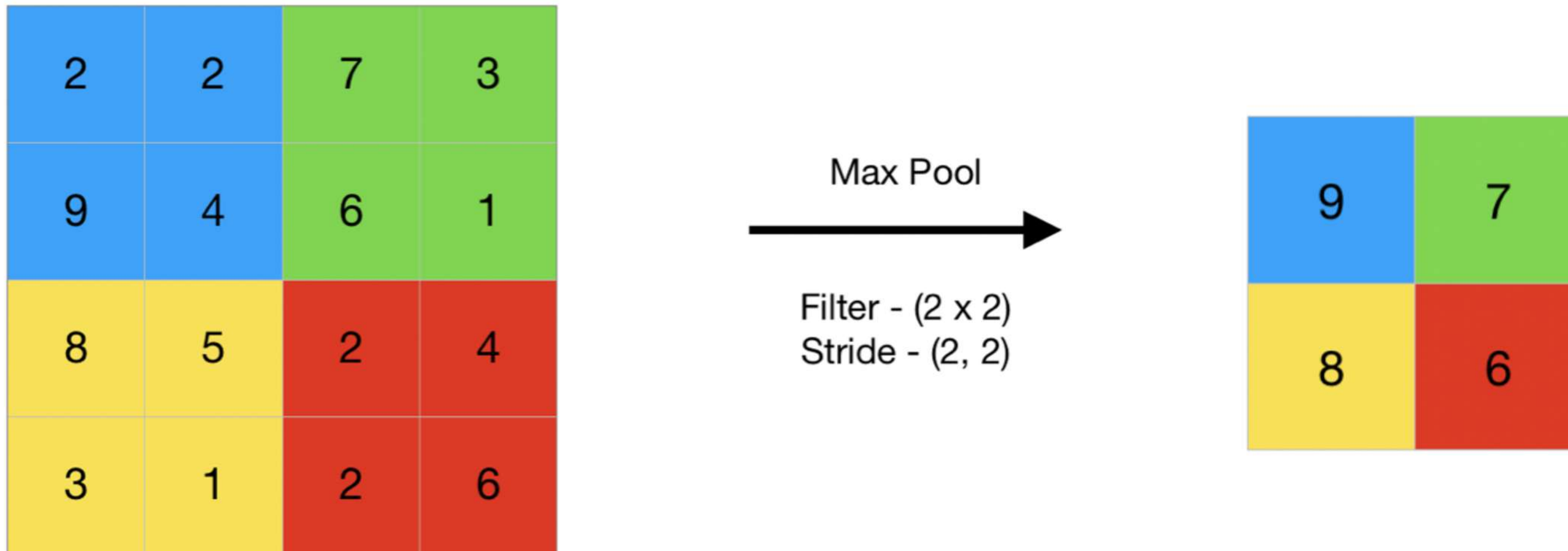
$$\left(\frac{n_h - k_h + p_h + s_h}{s_h} \right) \times \left(\frac{n_w - k_w + p_w + s_w}{s_w} \right)$$

$$(n_h + p_h) \times (n_w + p_w)$$

- Convolution layer mang tính chất Translation Equivariance:
 - Nếu input dịch đi 1 pixel thì output cũng dịch theo 1 pixel
 - Quan tâm đặc trưng ở đâu
- Pooling layer mang tính chất Translation Invariance:
 - Nếu input dịch đi 1 pixel thì output gần như giữ nguyên
 - Quan tâm có đặc trưng hay không

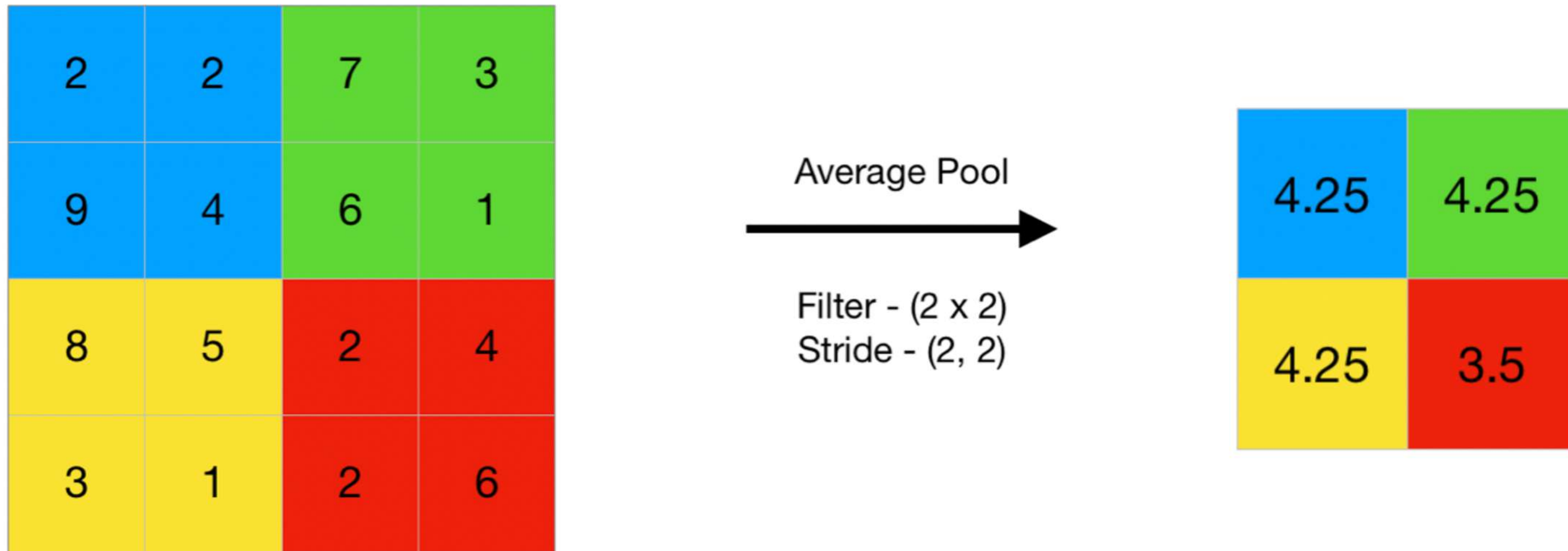
Max Pooling

- Max Pooling giữ lại các đặc trưng quan trọng nhất (giá trị lớn nhất đại diện cho "mức độ kích hoạt mạnh nhất" của đặc trưng) và đem lại kết quả tốt nhất trong nhiều trường hợp



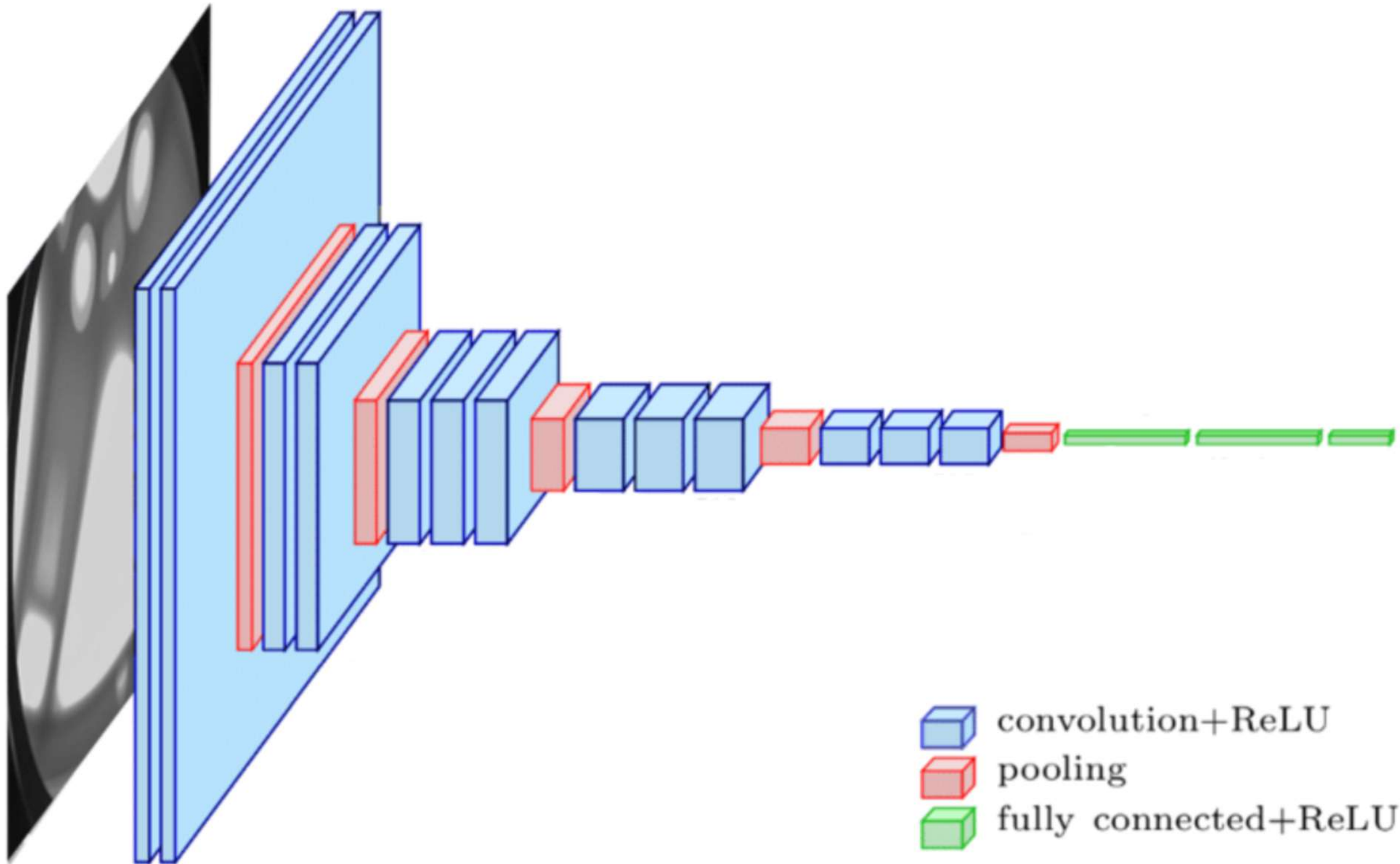
Average Pooling

- Average Pooling giữ giá trị trung bình của vùng, không quá thiên lệch vào đặc trưng mạnh, phù hợp hơn khi cần thông tin tổng quát



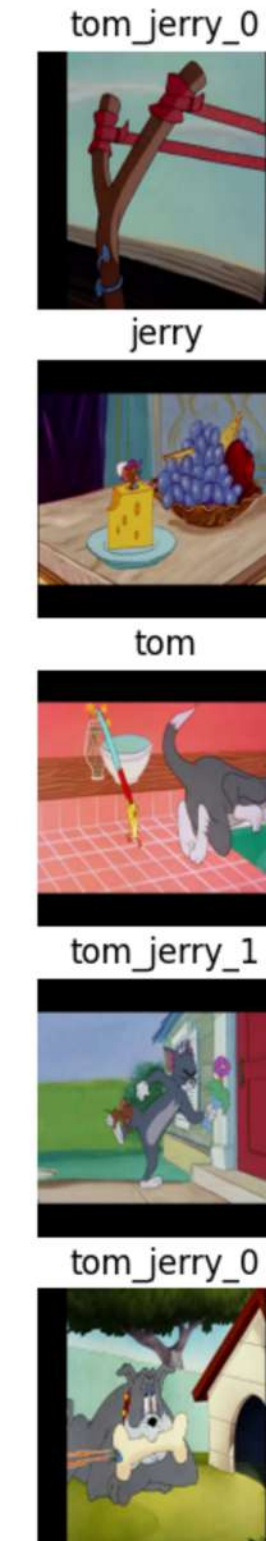
- Global Pooling giúp nén đặc trưng toàn cục, tóm tắt toàn bộ thông tin trong một kênh bằng một số duy nhất và tăng khả năng tổng quát của mô hình
- Global Pooling gồm có:
 - Global Average Pooling: tính giá trị “mức độ xuất hiện trung bình” của đặc trưng trong toàn ảnh
 - Global Max Pooling: tính giá trị “sự hiện diện mạnh nhất” của đặc trưng

CNN Architecture



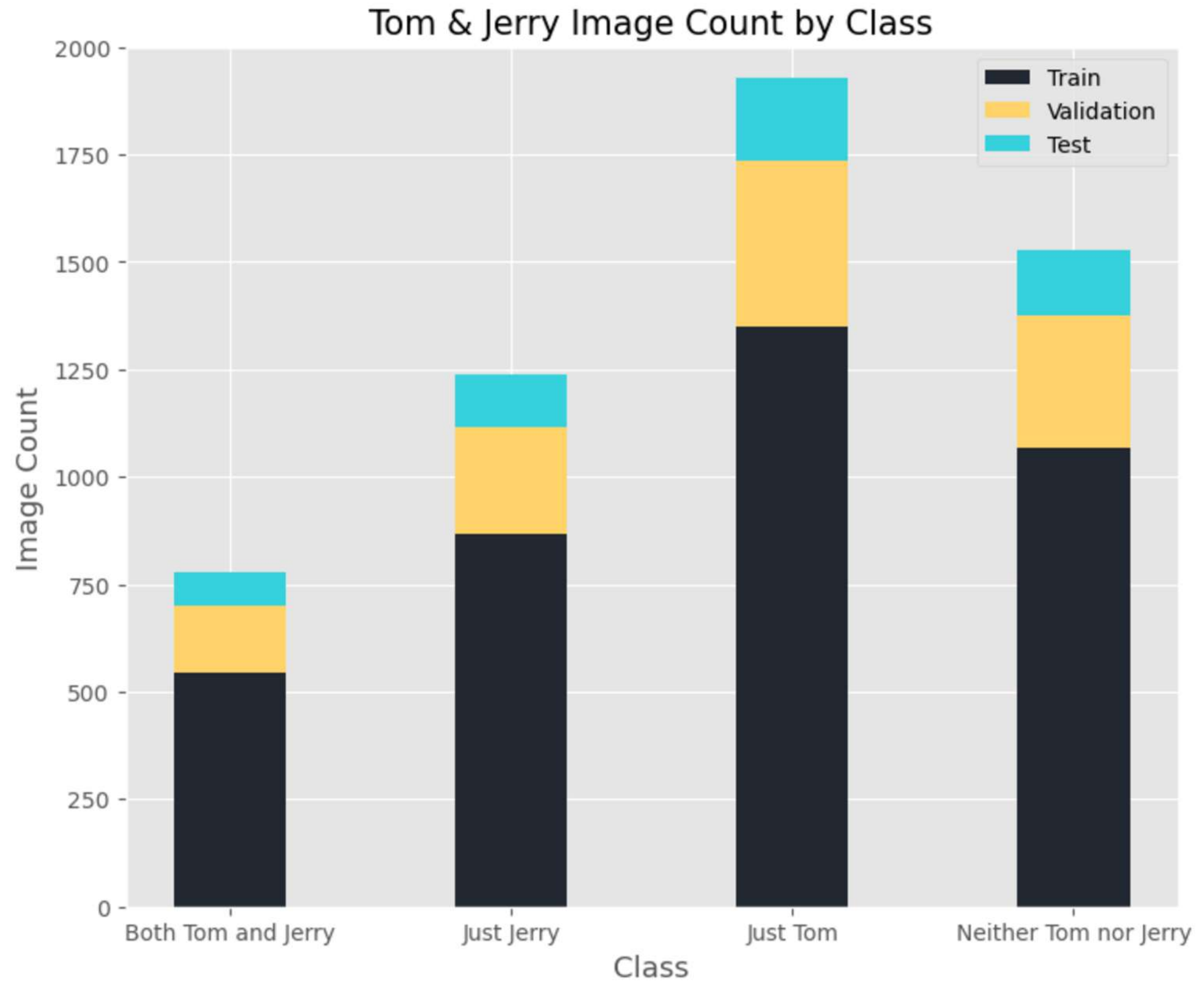
Ứng dụng trong bài toán thực tế

- Bộ dữ liệu: Tom and Jerry Image classification
- Bộ dữ liệu gồm 5478 ảnh lấy từ các tập phim Tom and Jerry được chia làm 4 nhãn: ảnh chỉ có Tom, ảnh chỉ có Jerry, ảnh có cả Tom và Jerry, ảnh không có cả hai



Ứng dụng trong bài toán thực tế

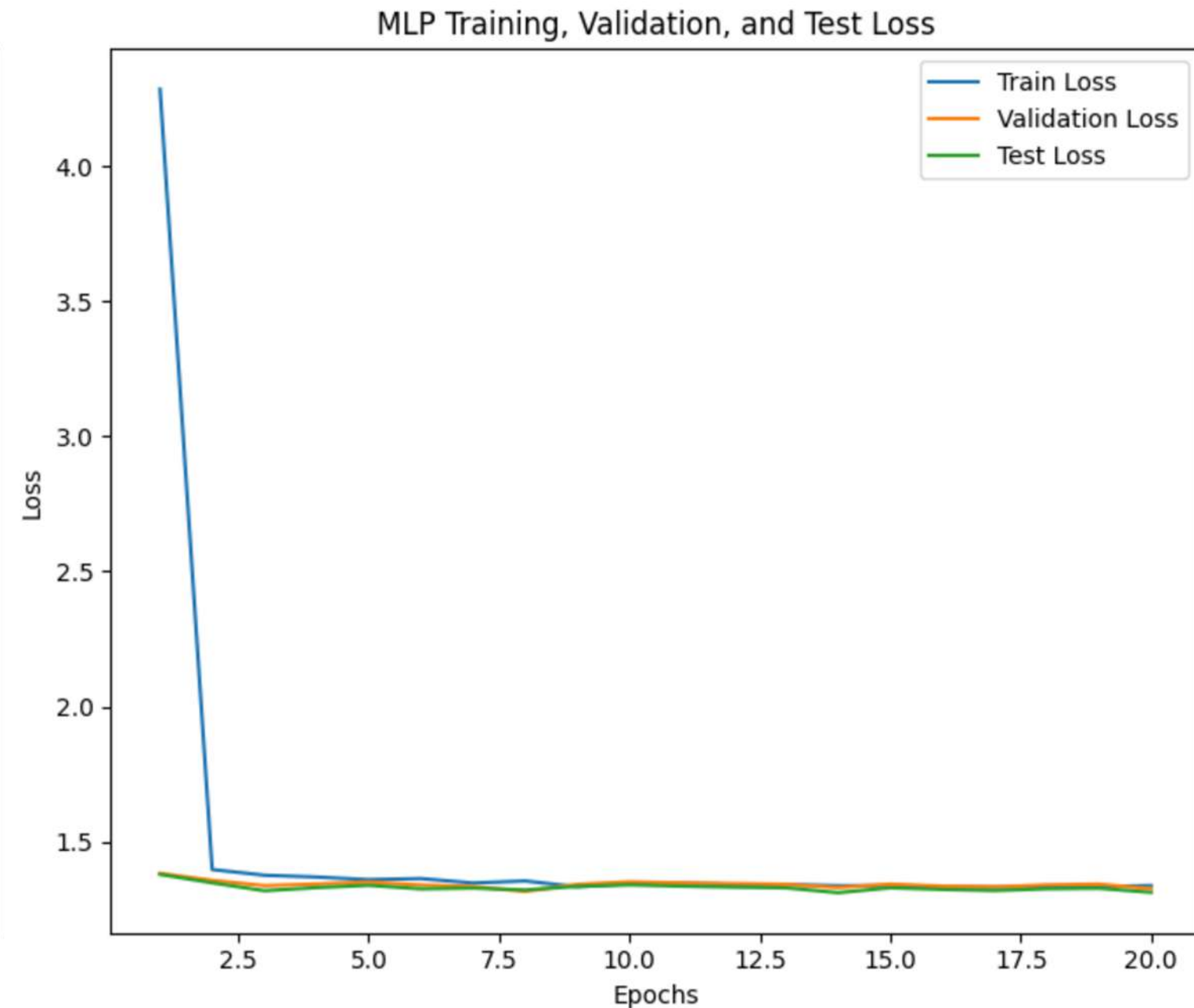
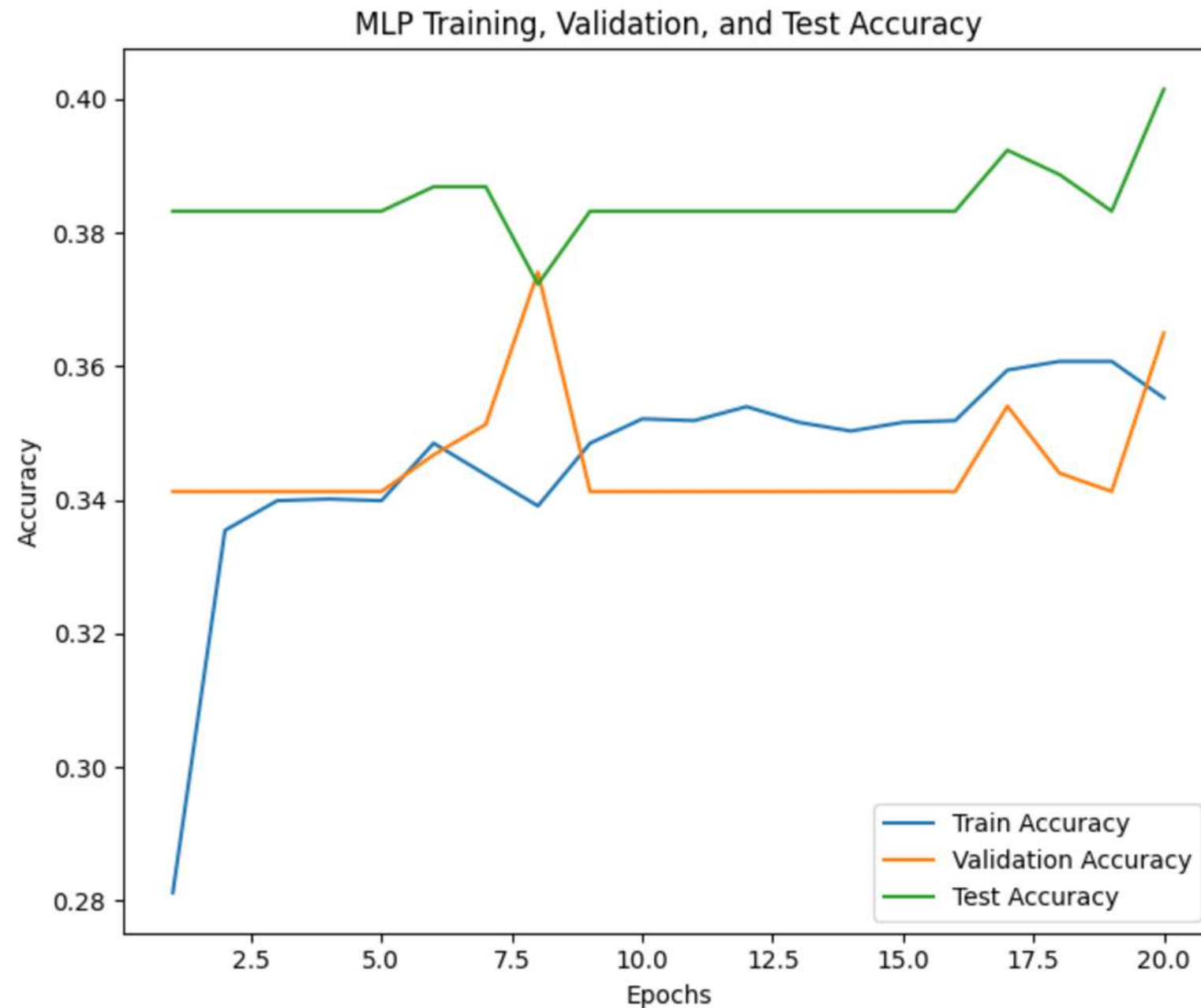
- Ba bộ train, validation và test với tỉ lệ 7:2:1



- MLP:
 - Flatten \rightarrow Linear(512) \rightarrow ReLU \rightarrow Dropout(0.5) \rightarrow Linear(128) \rightarrow ReLU \rightarrow Dropout(0.5) \rightarrow Linear(4)

Ứng dụng trong bài toán thực tế

- Validation set - Best Loss: 1.3149 - Best Accuracy: 0.3741
- Test set - Best Loss: 1.3102- Best Accuracy: 0.4015

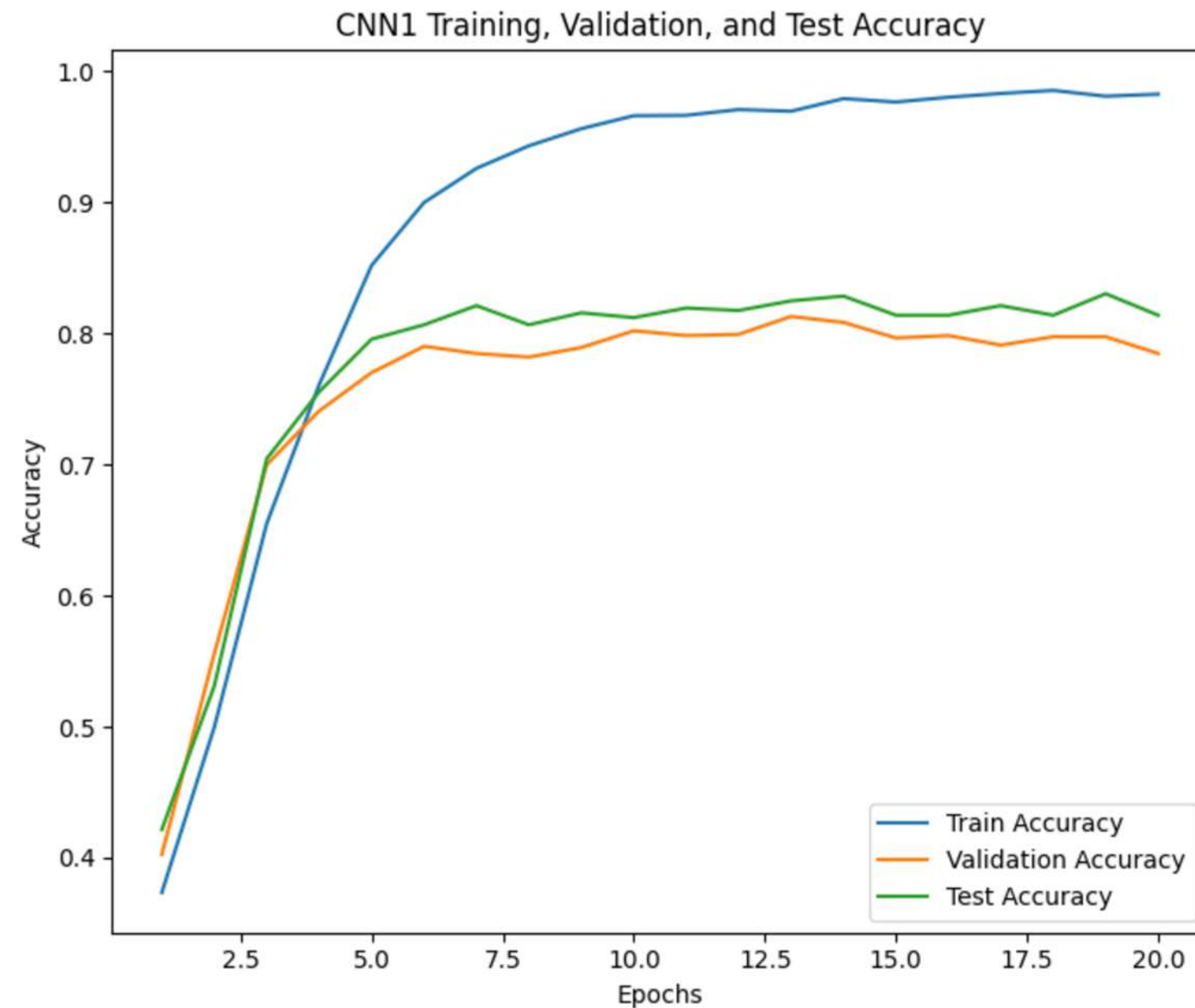


- CNN1:

- [Conv(3x3, 16) - ReLU - MaxPool(2x2)] →
[Conv(3x3, 32) - ReLU - MaxPool(2x2)] →
[Conv(3x3, 64) - ReLU - MaxPool(2x2)] →
Flatten → Linear(128) → ReLU → Linear(4)

Ứng dụng trong bài toán thực tế

- Validation set - Best Loss: 0.6130 - Best Accuracy: 0.8130
- Test set - Best Loss: 0.5647 - Best Accuracy: 0.8303

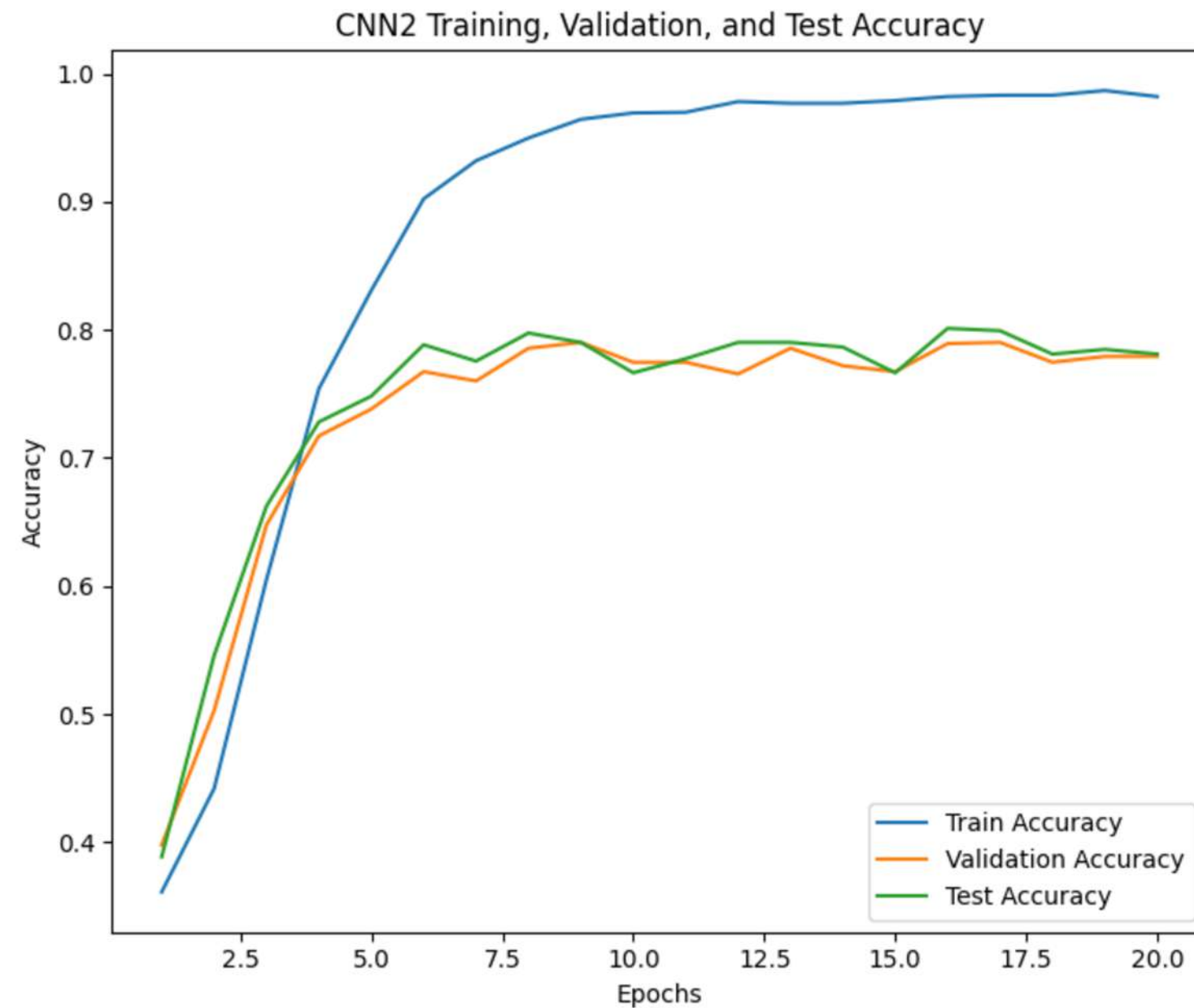


- CNN2:

- [Conv(3x3, 16) - ReLU - Conv(3x3, 16) - ReLU - MaxPool(2x2)] → [Conv(3x3, 32) - ReLU - Conv(3x3, 32) - ReLU - MaxPool(2x2)] → Flatten → Linear(128) → ReLU → Linear(4)

Ứng dụng trong bài toán thực tế

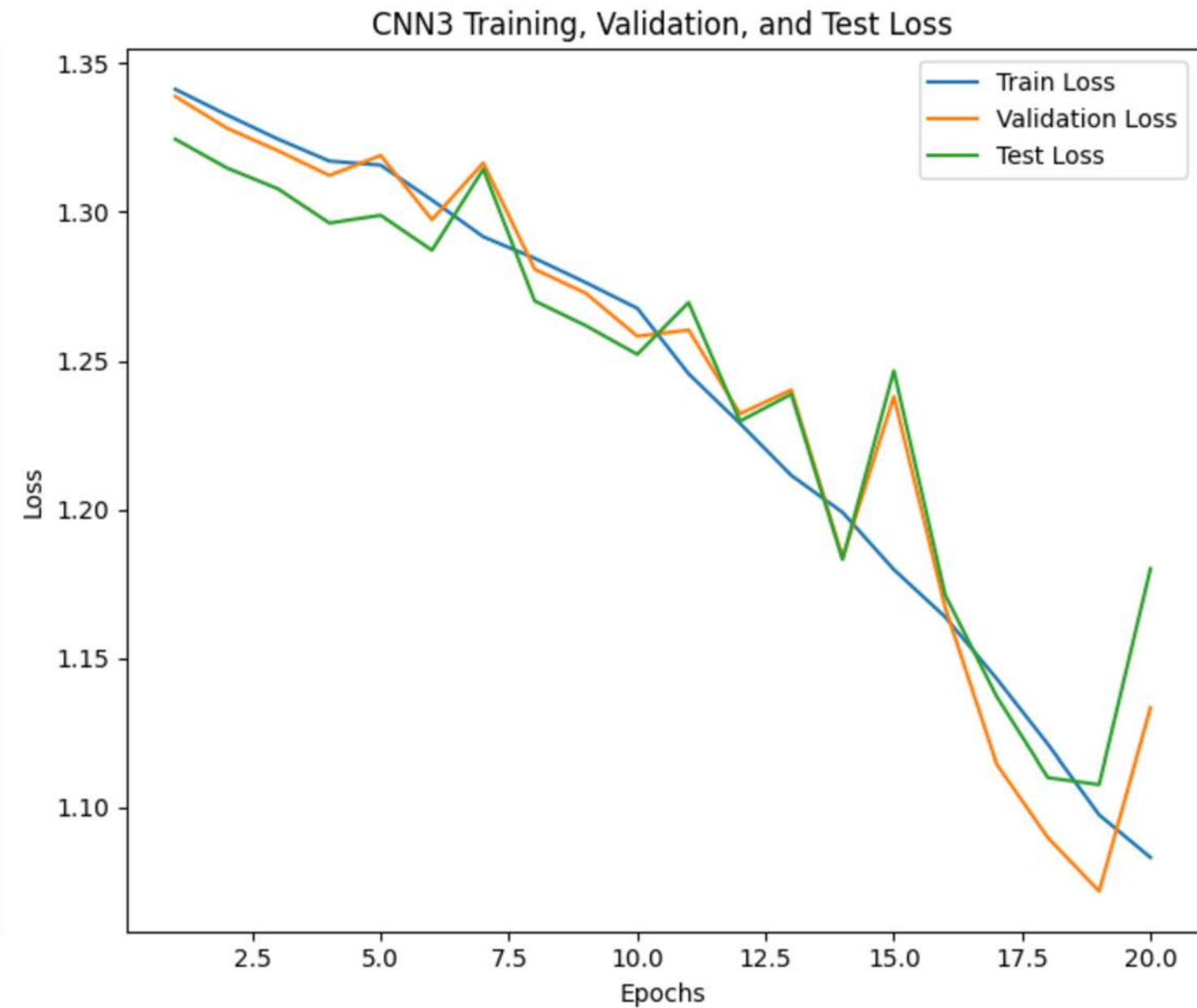
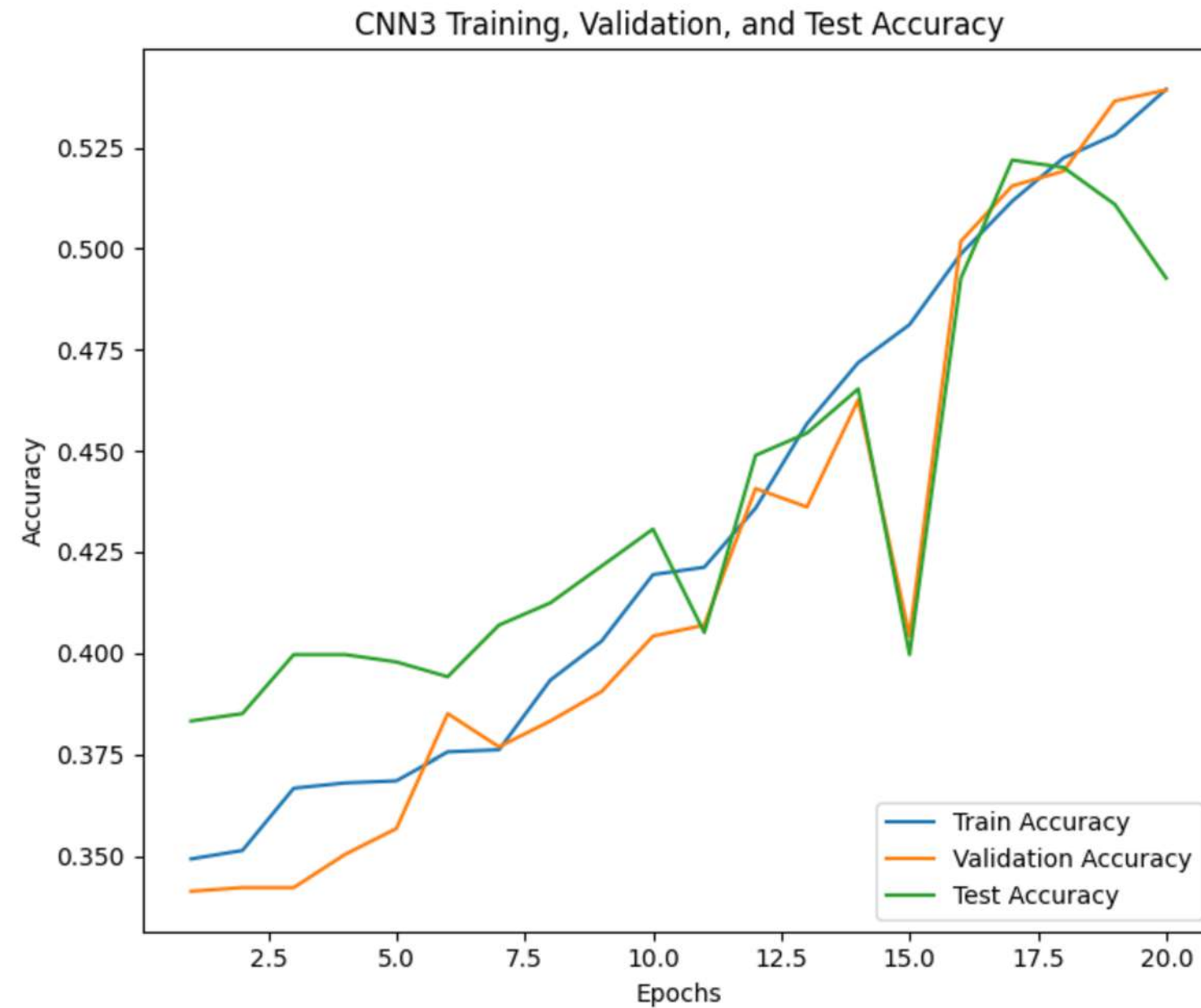
- Validation set - Best Loss: 0.7582 - Best Accuracy: 0.7801
- Test set - Best Loss: 0.7005 - Best Accuracy: 0.8084



- CNN3:
 - [Conv(3x3, 16) - ReLU - MaxPool(2x2)] →
 - [Conv(3x3, 32) - ReLU - MaxPool(2x2)] →
 - [Conv(3x3, 64) - ReLU - MaxPool(2x2)] →
 - GlobalAvgPool → Flatten → Linear(4)

Ứng dụng trong bài toán thực tế

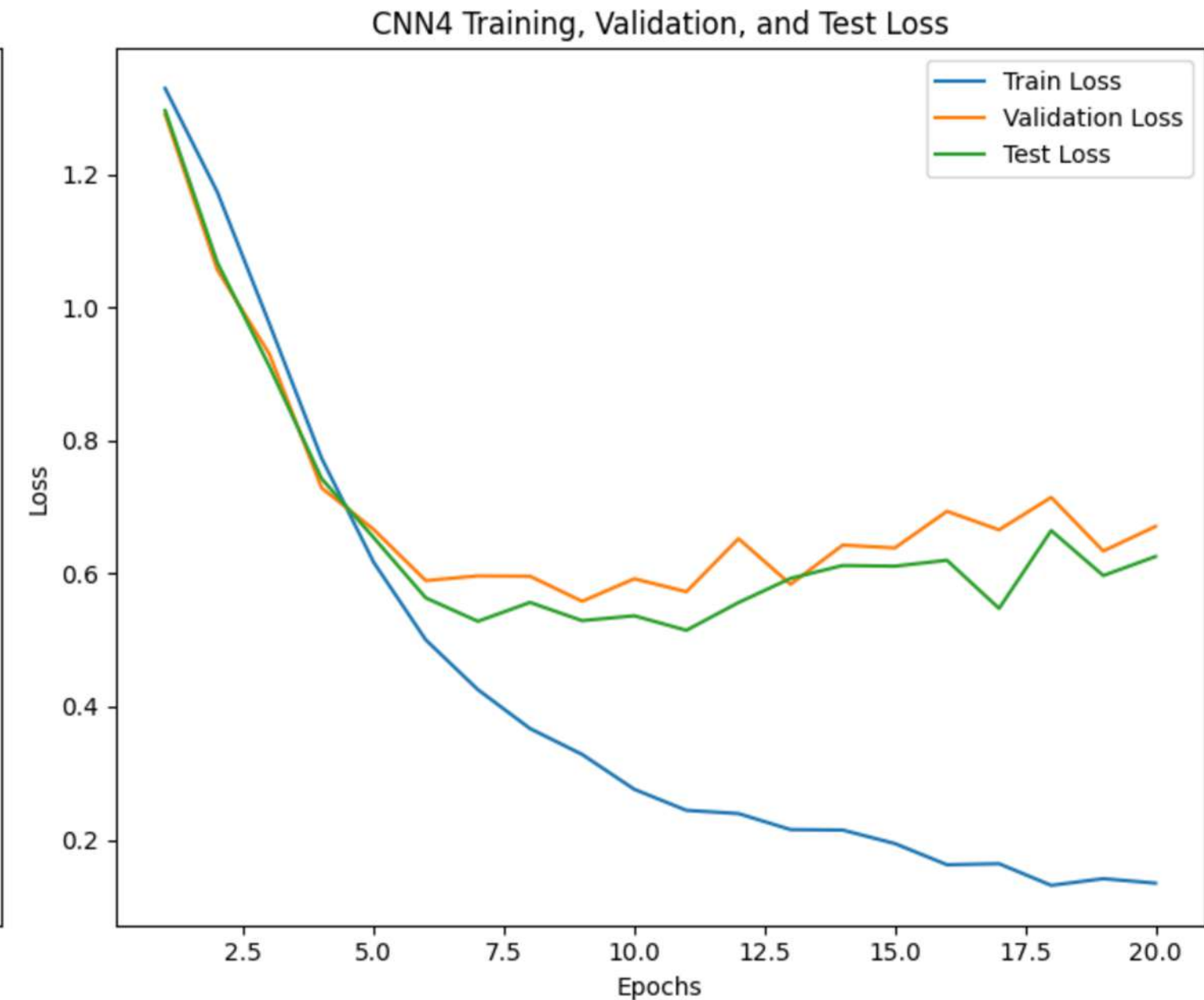
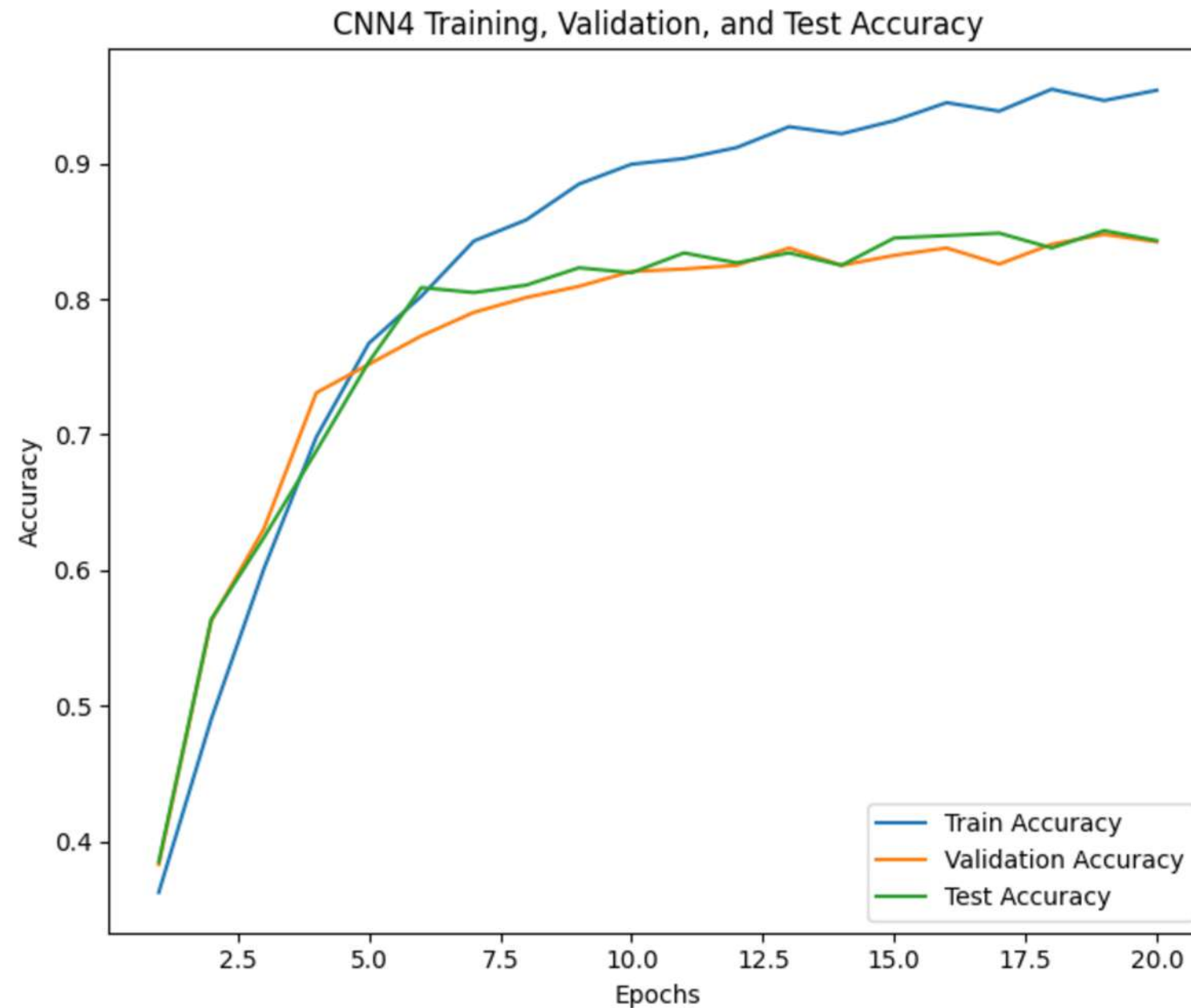
- Validation set - Best Loss: 1.0718 - Best Accuracy: 0.5392
- Test set - Best Loss: 1.1077 - Best Accuracy: 0.5219



- CNN4:
 - [Conv(3x3, 16) - ReLU - MaxPool(2x2)] →
[Conv(3x3, 32) - ReLU - MaxPool(2x2)] →
[Conv(3x3, 64) - ReLU - MaxPool(2x2)] →
Flatten → Dropout(0.5) → Linear(128) → ReLU →
Dropout(0.5) → Linear(4)

Ứng dụng trong bài toán thực tế

- Validation set - Best Loss: 0.5582 - Best Accuracy: 0.8476
- Test set - Best Loss: 0.5146 - Best Accuracy: 0.8504



Ứng dụng trong bài toán thực tế

Model	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy	Runtime
MLP	1.3149	0.3741	1.3102	0.4015	16m19s
CNN1	0.6130	0.8130	0.5647	0.8303	18m23s
CNN2	0.7582	0.7801	0.7005	0.8084	16m39s
CNN3	1.0718	0.5392	1.1077	0.5219	15m19s
CNN4	0.5582	0.8476	0.5146	0.8504	17m33s

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