

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

Normalization và Các phương pháp chống Overfitting trong học sâu

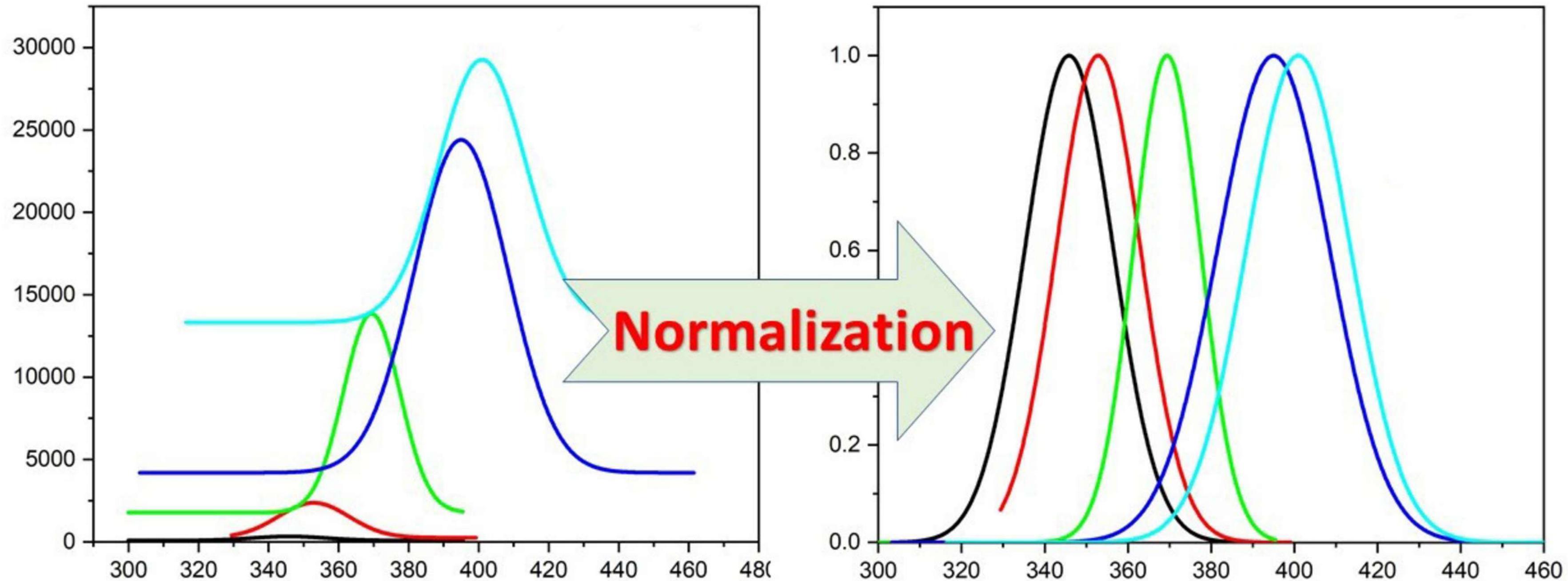
ONE LOVE. ONE FUTURE.

Nội dung chính

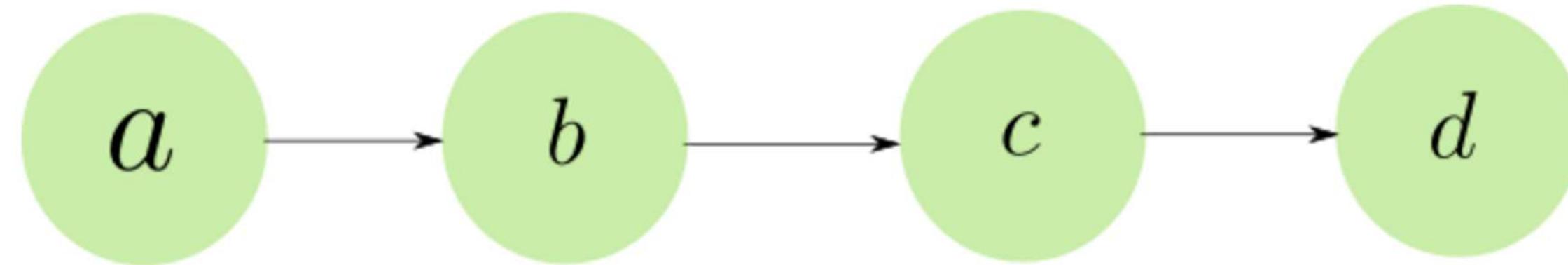
- Normalization
 - Normalization Methods
- Các phương pháp chống Overfitting
- Ứng dụng trong bài toán thực tế



Normalization

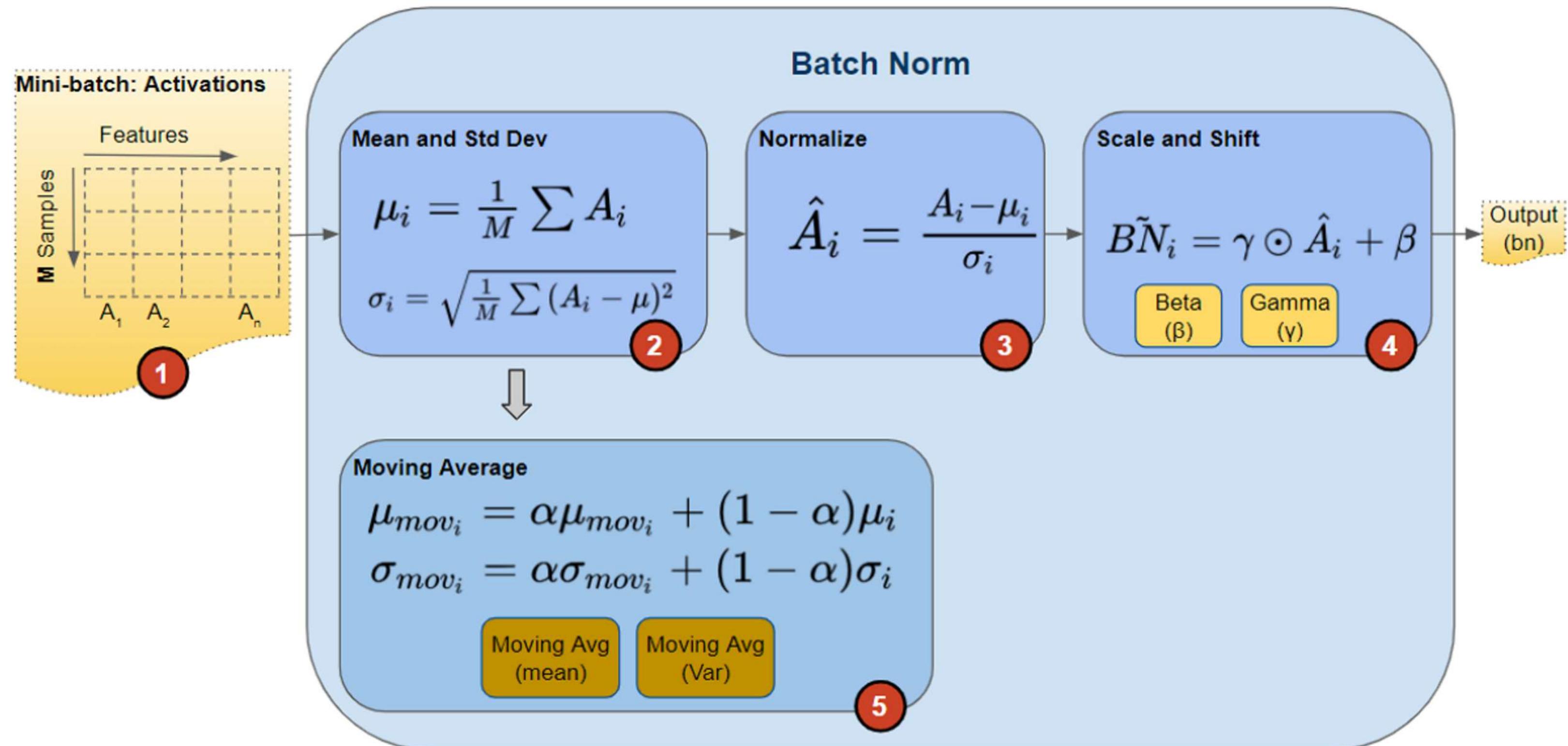


Batch Normalization



- Ở bước thứ i đầu ra ở lớp c là dữ liệu với phân phối p_c^i
- Lớp d cập nhật với giả định phân phối đầu vào là p_c^i
- Khi backprop, w_c cập nhật lại và cho ra phân phối p_c^{i+1} ở bước thứ $i + 1$
→ Internal Covariate Shift

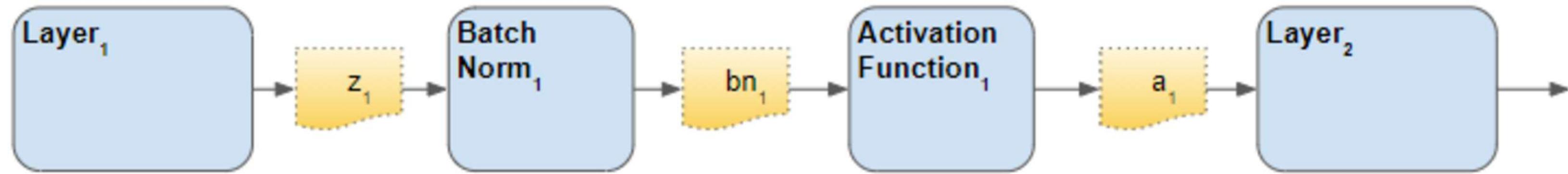
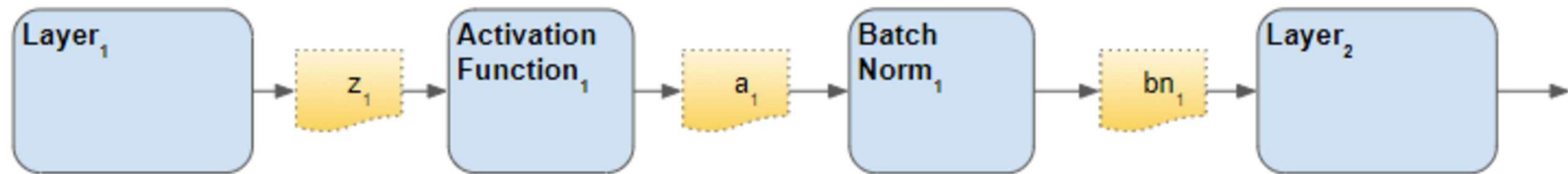
Batch Normalization



- Tại sao BatchNorm lại hoạt động tốt?
 - Internal Covariate Shift: giúp giảm “sự thay đổi phân phối đầu vào của từng lớp trong quá trình huấn luyện”
 - Gradient Stabilization: giúp gradient trở nên mượt hơn, ổn định hơn, tránh vanishing/exploding gradient
 - Smoothing the Loss: làm mượt không gian hàm loss, giúp gradient ít bị kẹt ở tối ưu cục bộ
 - Regularization Effect: có tác dụng tương tự dropout, vì khi tính batch mean/variance, BatchNorm bổ sung noise nhỏ vào mỗi batch

Batch Normalization

- Vị trí của BatchNorm trong mạng



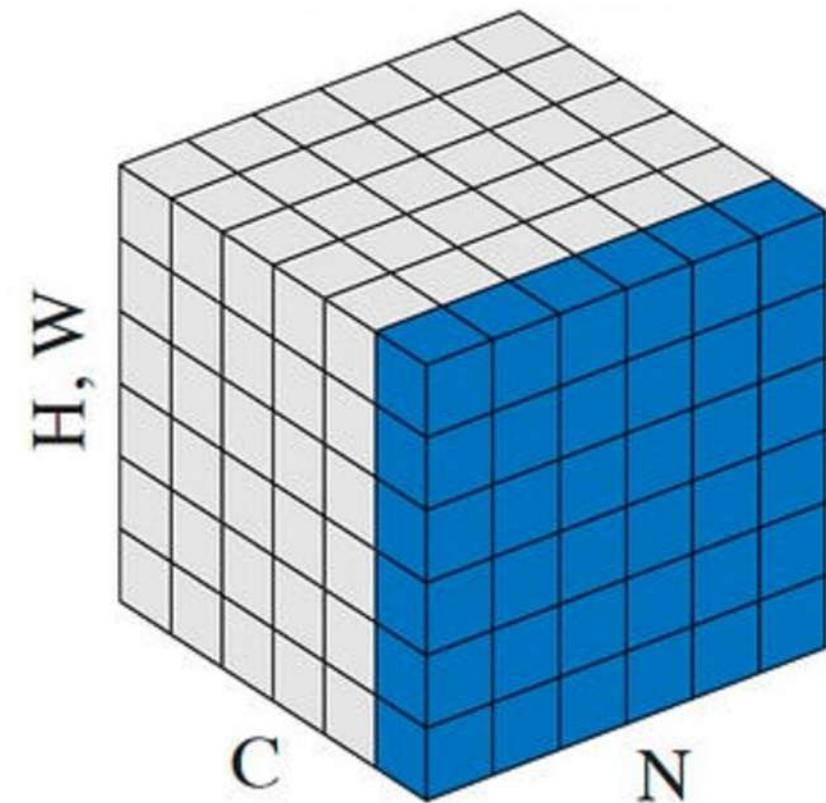
Batch Normalization

- Hạn chế của BatchNorm:
 - Phụ thuộc vào batch size
 - Không phù hợp với dữ liệu chuỗi thời gian
 - Không ổn định trong multi-domain training
- Để khắc phục hạn chế của BatchNorm, các biến thể khác được sử dụng:
 - LayerNorm
 - InstanceNorm
 - GroupNorm

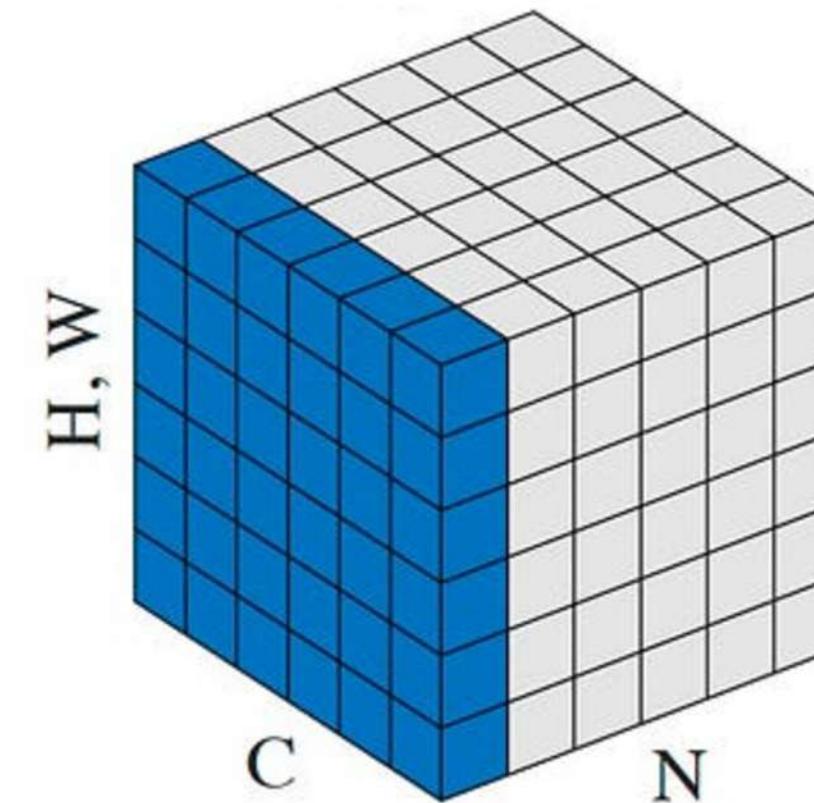


Normalization

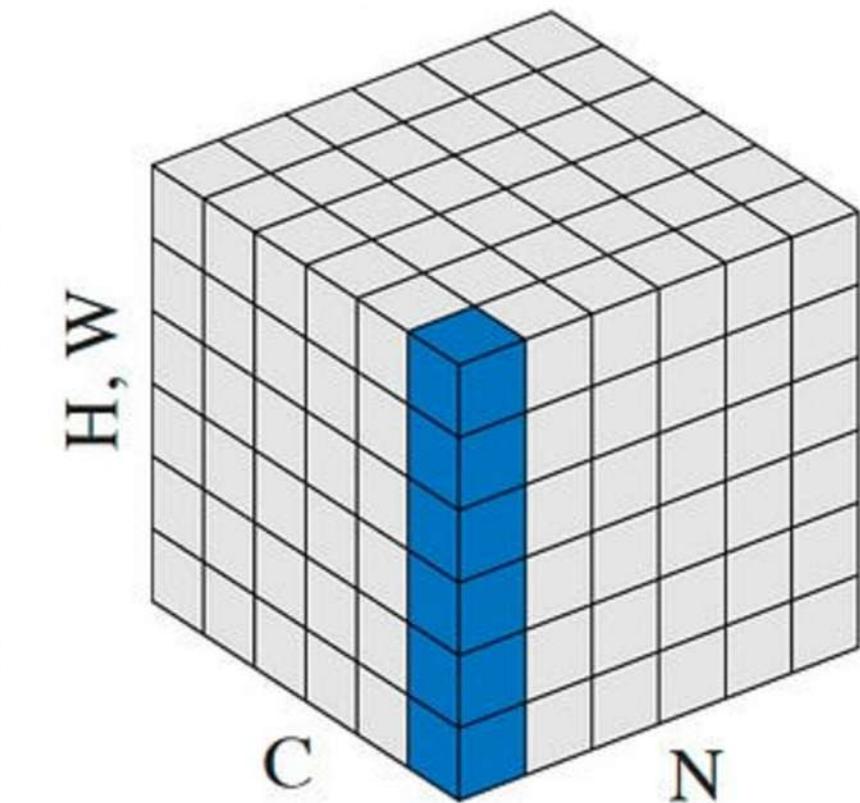
BatchNorm



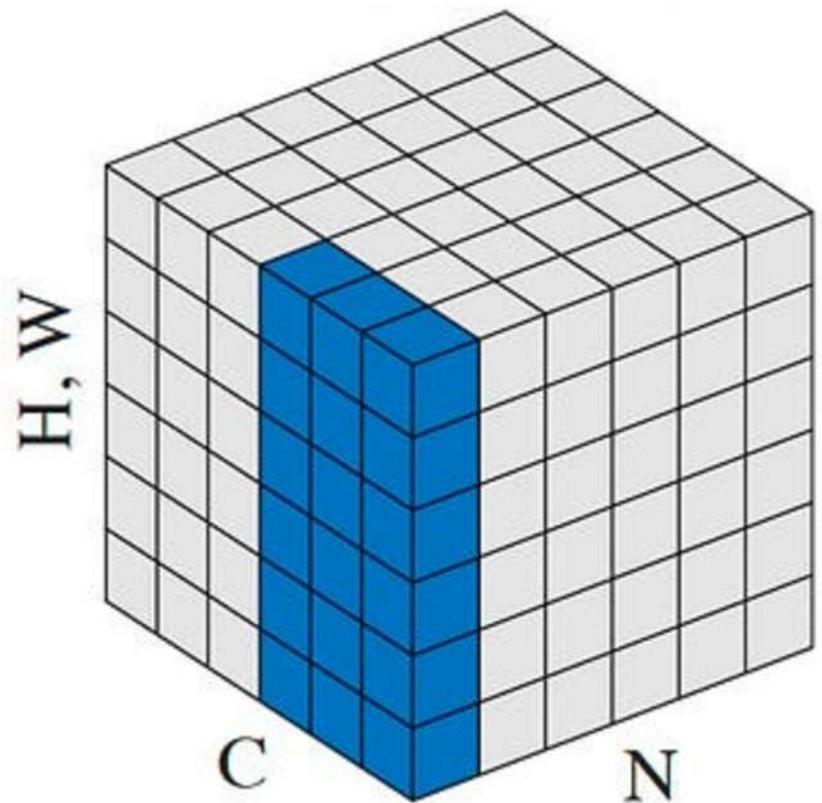
LayerNorm



InstanceNorm



GroupNorm



Weight Normalization

- WeightNorm tách riêng hướng và độ lớn của trọng số, giúp việc cập nhật hai yếu tố trở nên độc lập với nhau

$$y = wx + b \Rightarrow y = \frac{g}{\|v\|} vx + b$$

- WeightNorm giúp các bước cập nhật gradient đi đúng hướng hơn, đẩy nhanh tốc độ hội tụ. Tuy nhiên WeightNorm không đem lại hiệu quả tốt bằng BatchNorm mà thường phải kết hợp với Mean-Only BatchNorm

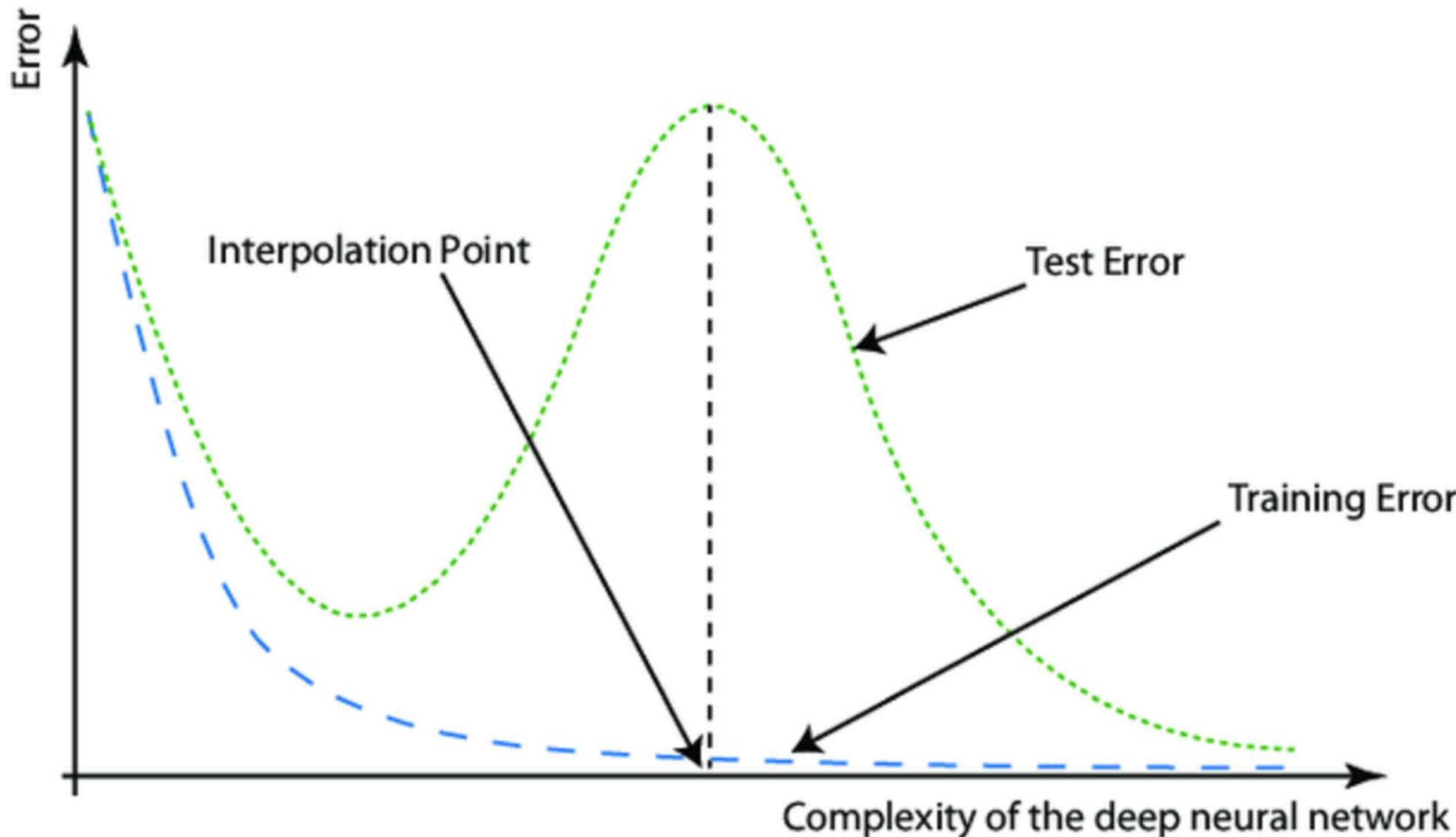


Các phương pháp chống Overfitting

- Các phương pháp chống Overfitting phổ biến là:
 - Validation/Cross-Validation
 - Regularization:
 - Early Stopping
 - L1/L2 Regularization
 - Dropout, DropConnect
 - Global Pooling
 - Normalization Methods
 - Data Augmentation

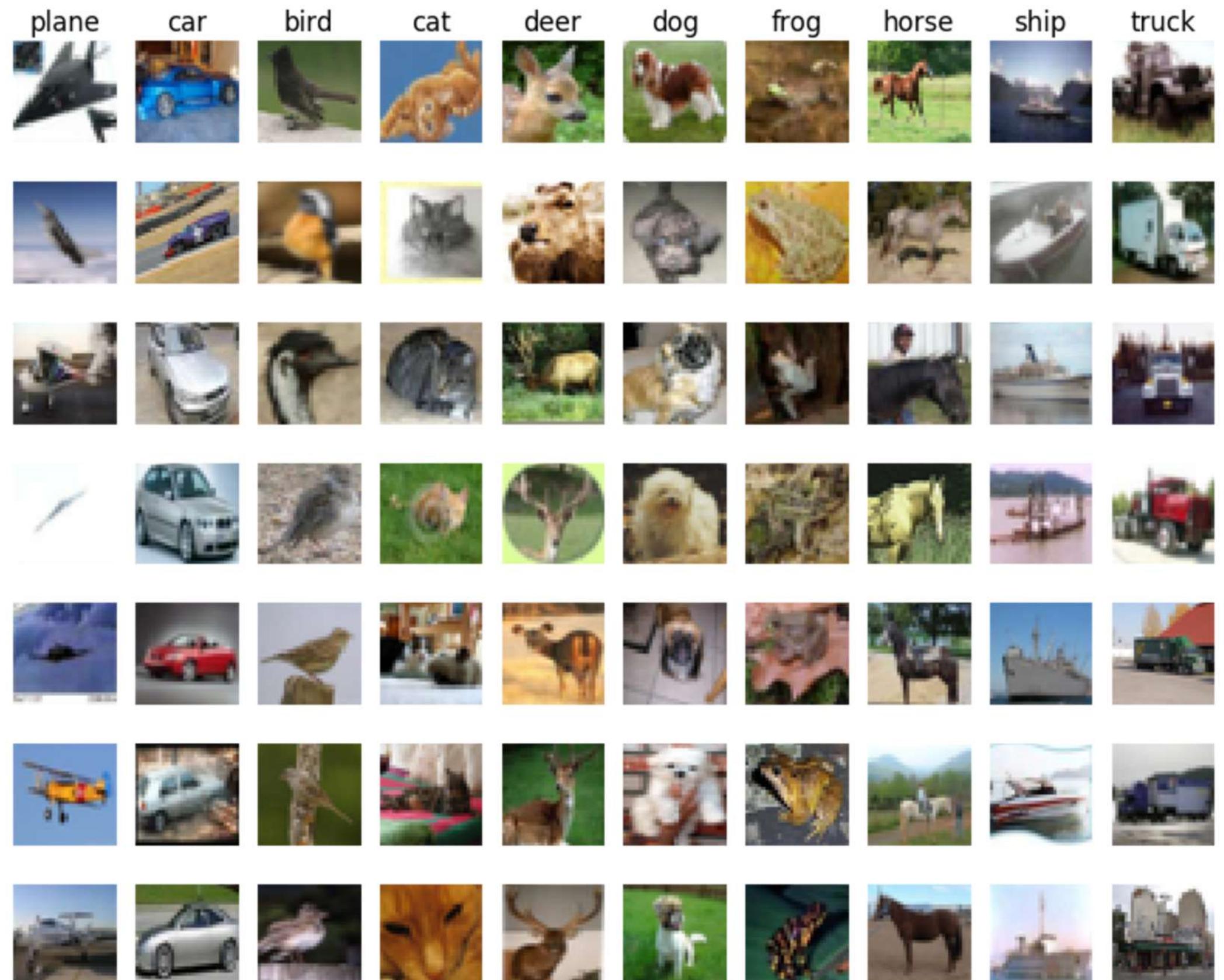


Double Descent



Ứng dụng trong bài toán thực tế

- Bộ dữ liệu CIFAR10
- Gồm 60000 ảnh, kích thước 32x32, chia làm 10 nhãn
- 50000 ảnh train và 10000 ảnh test



Ứng dụng trong bài toán thực tế

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



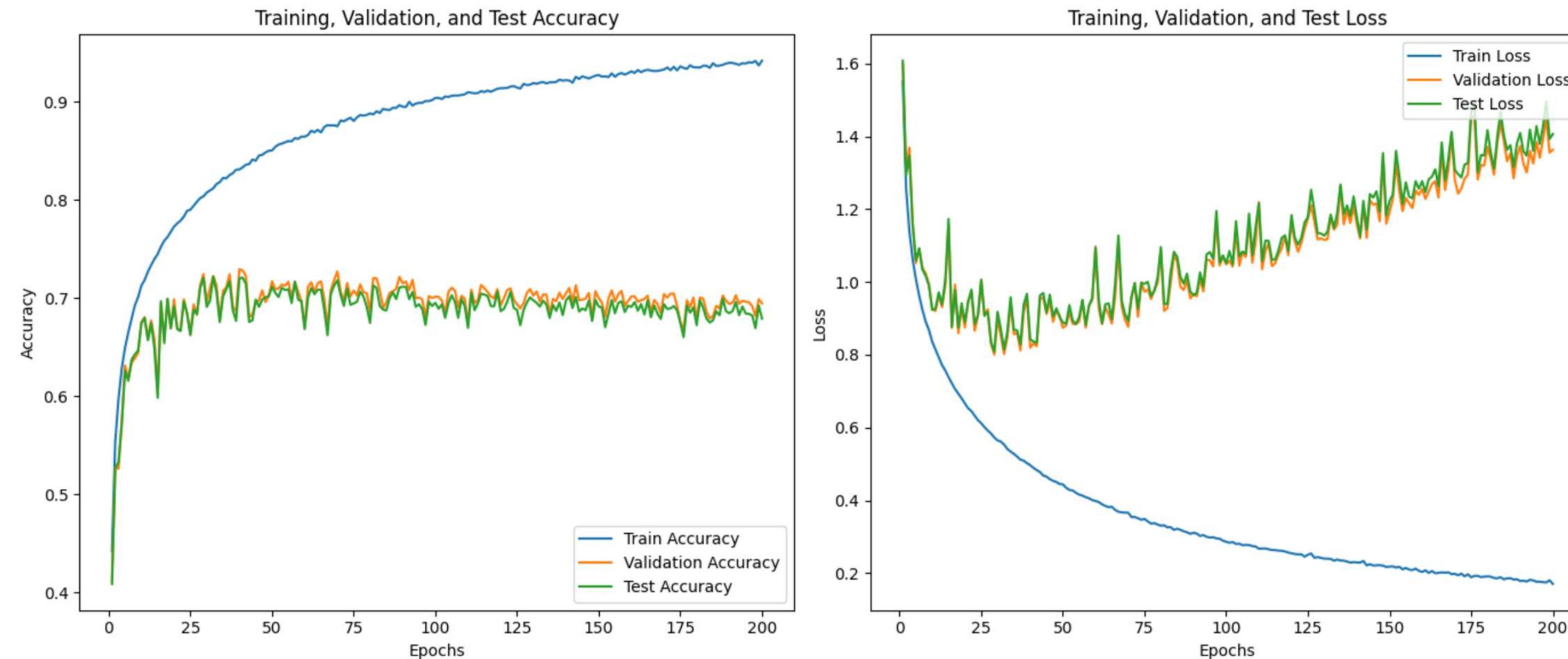
Ứng dụng trong bài toán thực tế

- CNN (No Normalization)
 - Validation set - Best Loss: 0.8768 - Best Accuracy: 0.7073
 - Test set - Best Loss: 0.8912 - Best Accuracy: 0.7071



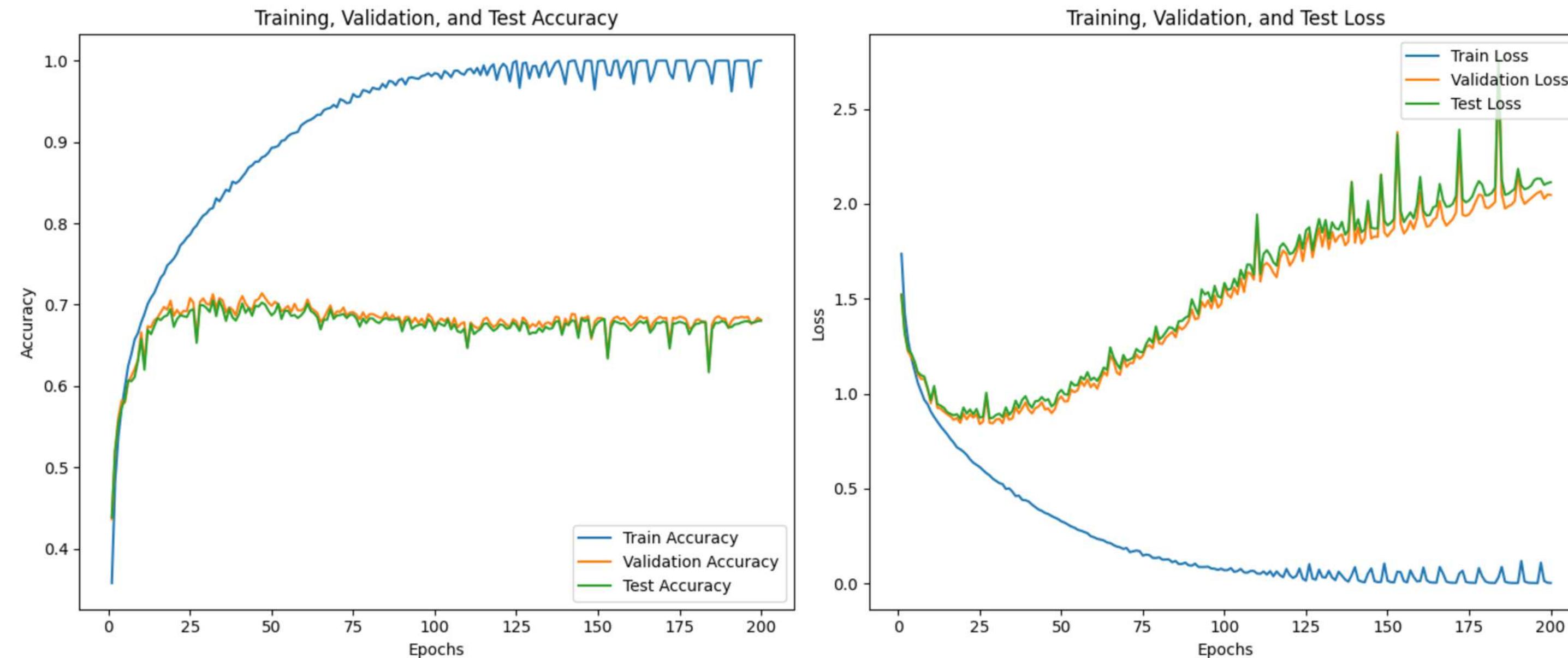
Ứng dụng trong bài toán thực tế

- CNN (BatchNorm)
 - Validation set - Best Loss: 0.8005 - Best Accuracy: 0.7293
 - Test set - Best Loss: 0.8080 - Best Accuracy: 0.7217



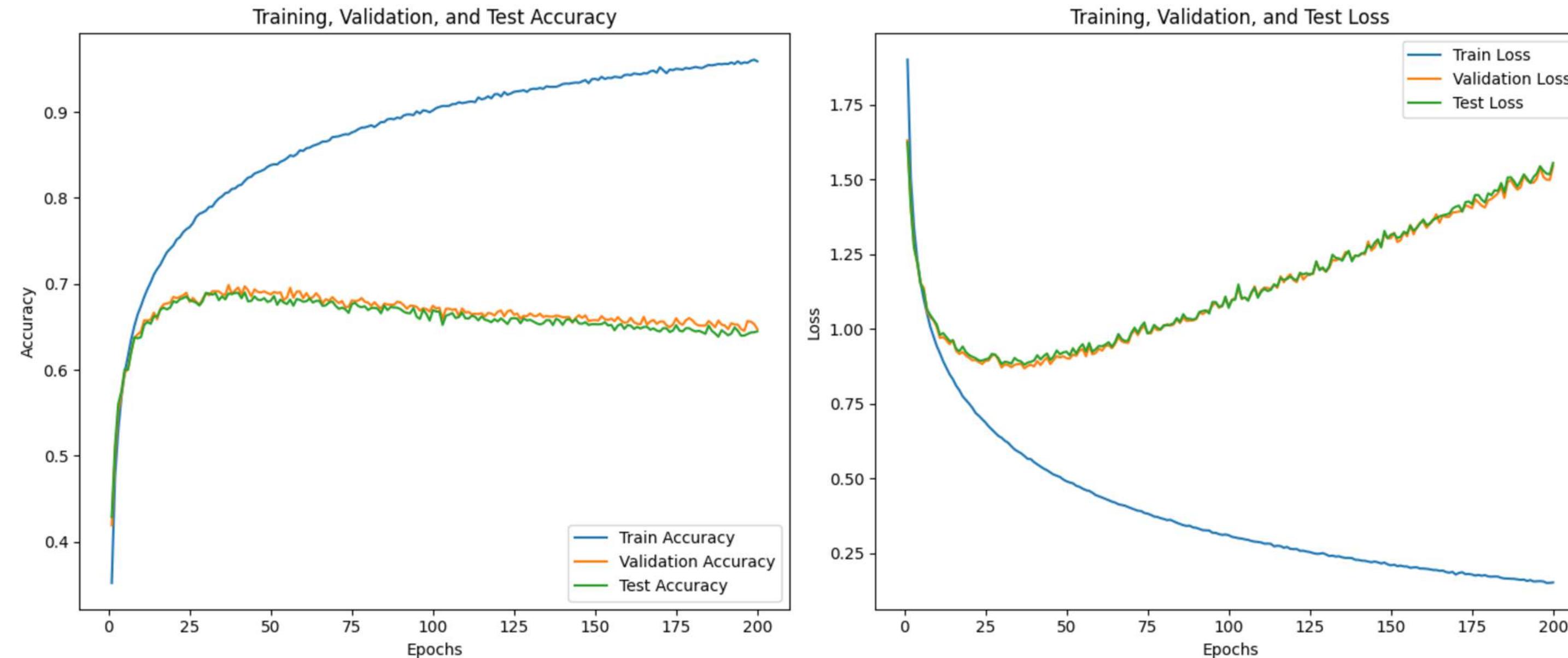
Ứng dụng trong bài toán thực tế

- CNN (LayerNorm)
 - Validation set - Best Loss: 0.8397 - Best Accuracy: 0.7140
 - Test set - Best Loss: 0.8681 - Best Accuracy: 0.7053



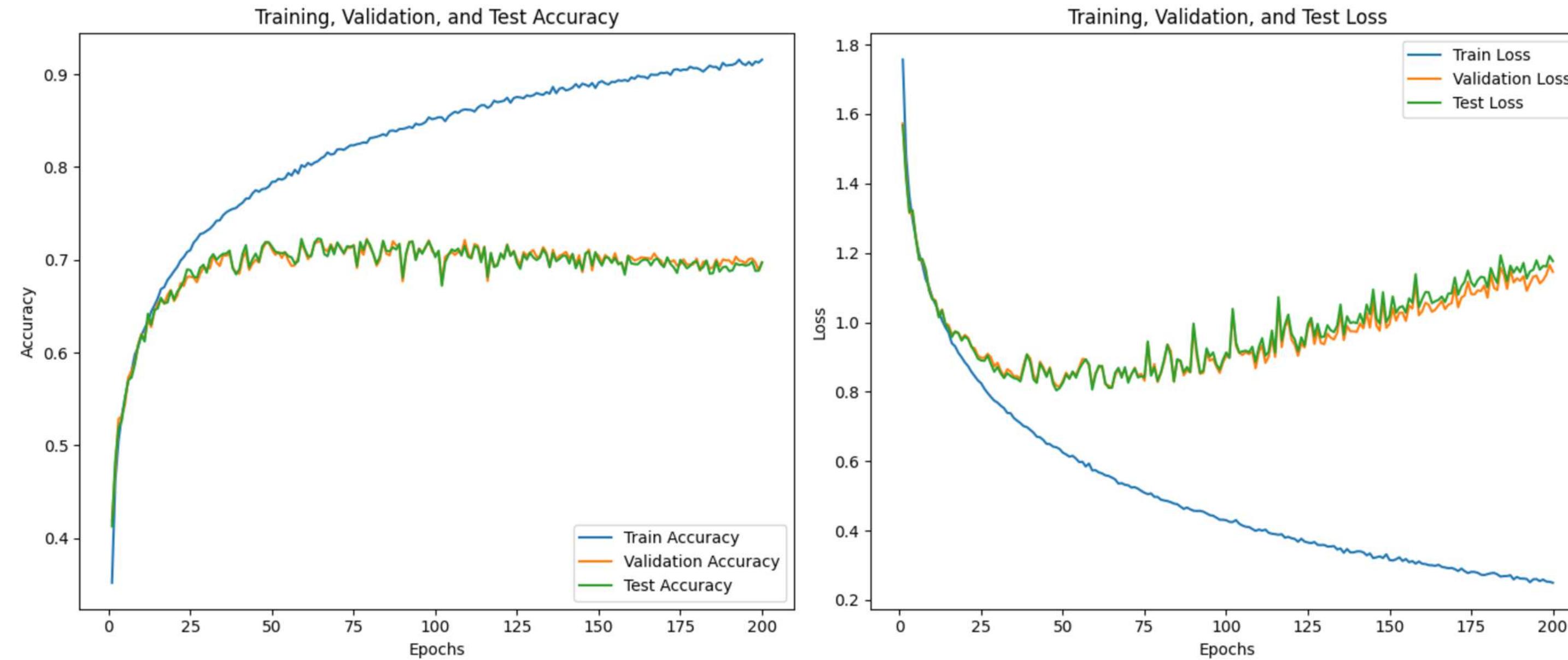
Ứng dụng trong bài toán thực tế

- CNN (InstanceNorm)
 - Validation set - Best Loss: 0.8677 - Best Accuracy: 0.6987
 - Test set - Best Loss: 0.8801 - Best Accuracy: 0.6901



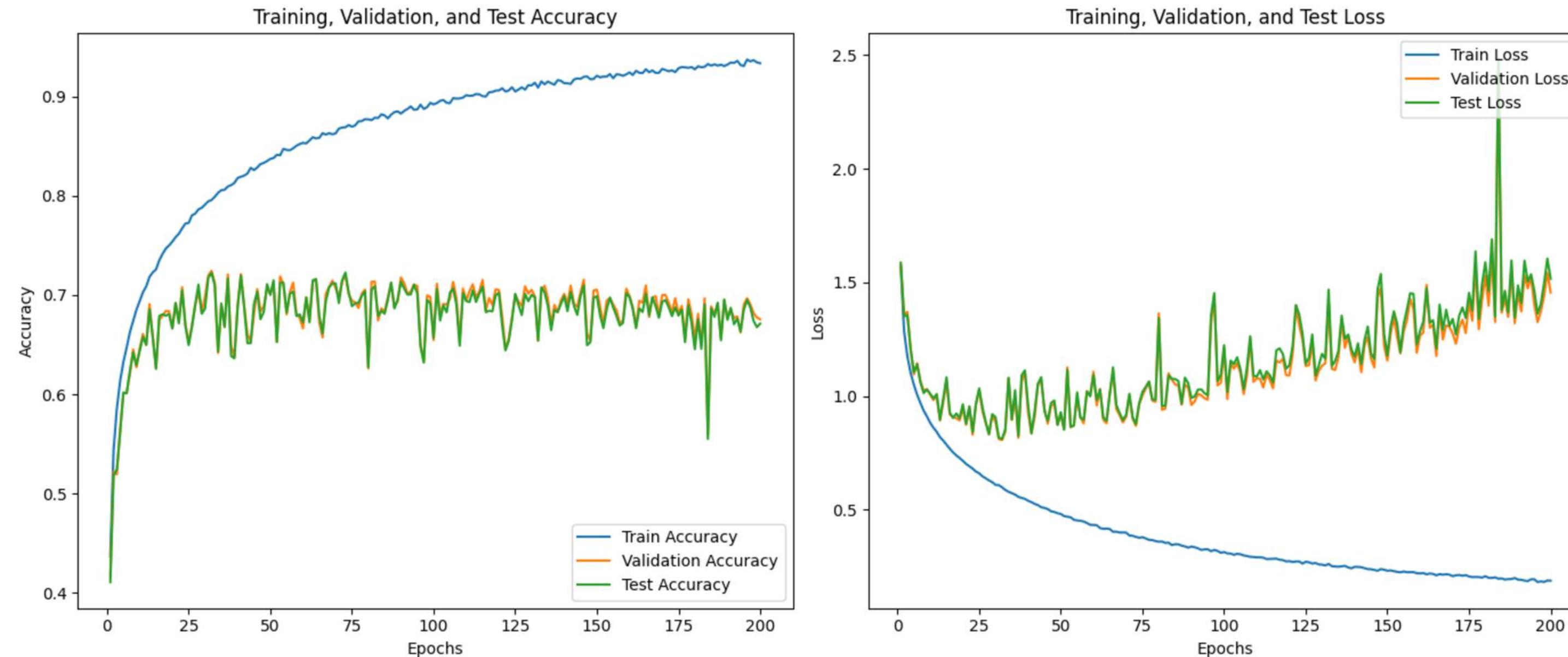
Ứng dụng trong bài toán thực tế

- CNN (GroupNorm)
 - Validation set - Best Loss: 0.8122 - Best Accuracy: 0.7227
 - Test set - Best Loss: 0.8041 - Best Accuracy: 0.7229

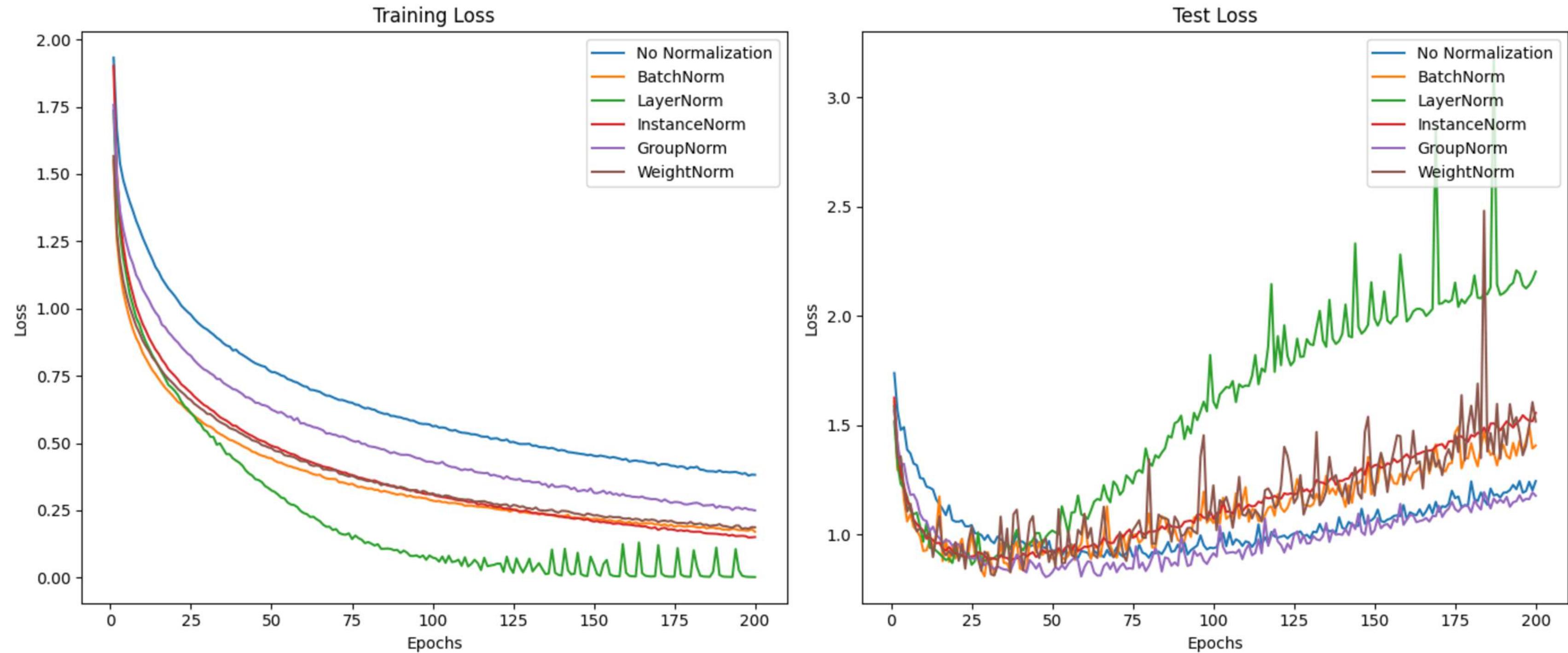


Ứng dụng trong bài toán thực tế

- CNN (WeightNorm)
 - Validation set - Best Loss: 0.8064 - Best Accuracy: 0.7243
 - Test set - Best Loss: 0.8111 - Best Accuracy: 0.7226



Ứng dụng trong bài toán thực tế



Ứng dụng trong bài toán thực tế

Model	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
No Normalization	0.8768	0.7073	0.8912	0.7071
BatchNorm	0.8005	0.7293	0.8080	0.7217
LayerNorm	0.8397	0.7140	0.8681	0.7053
InstanceNorm	0.8677	0.6987	0.8801	0.6901
GroupNorm	0.8122	0.7227	0.8041	0.7229
WeightNorm	0.8064	0.7243	0.8111	0.7226



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



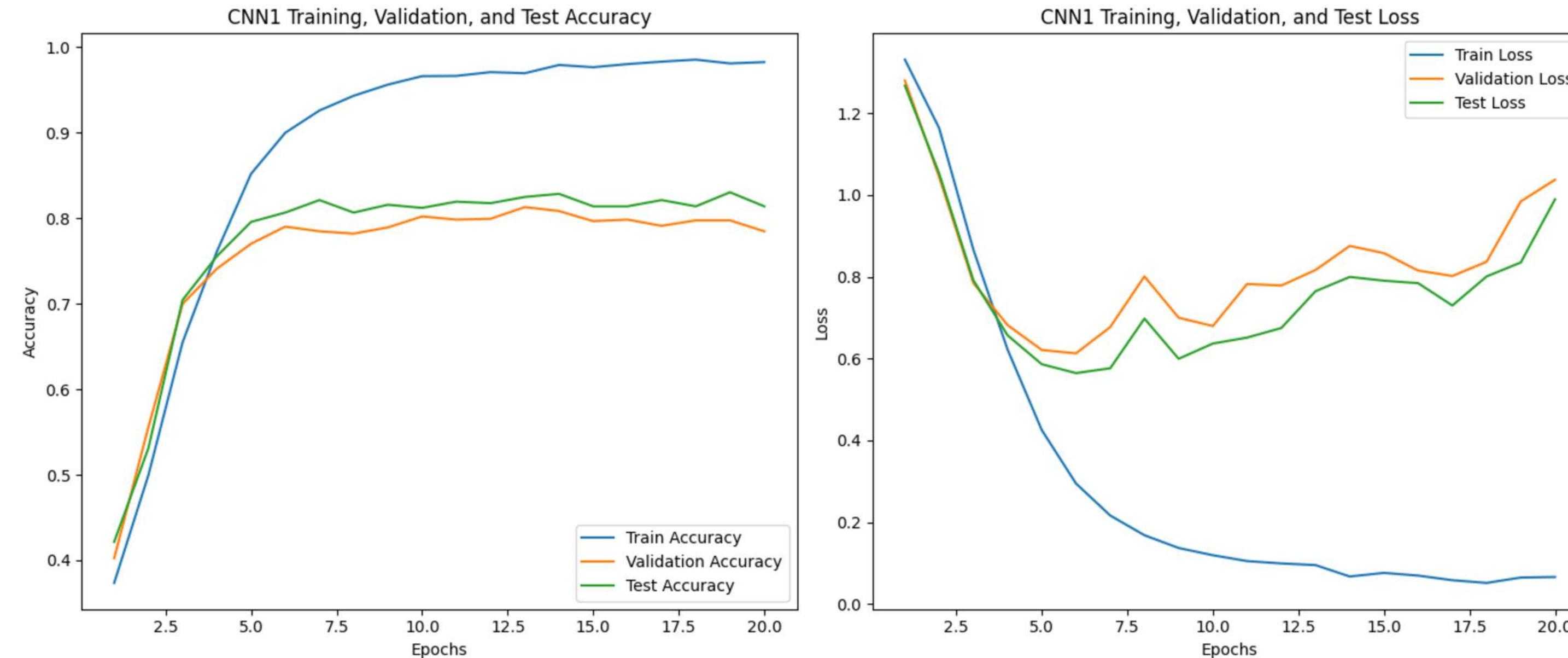
Ứng dụng trong bài toán thực tế

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



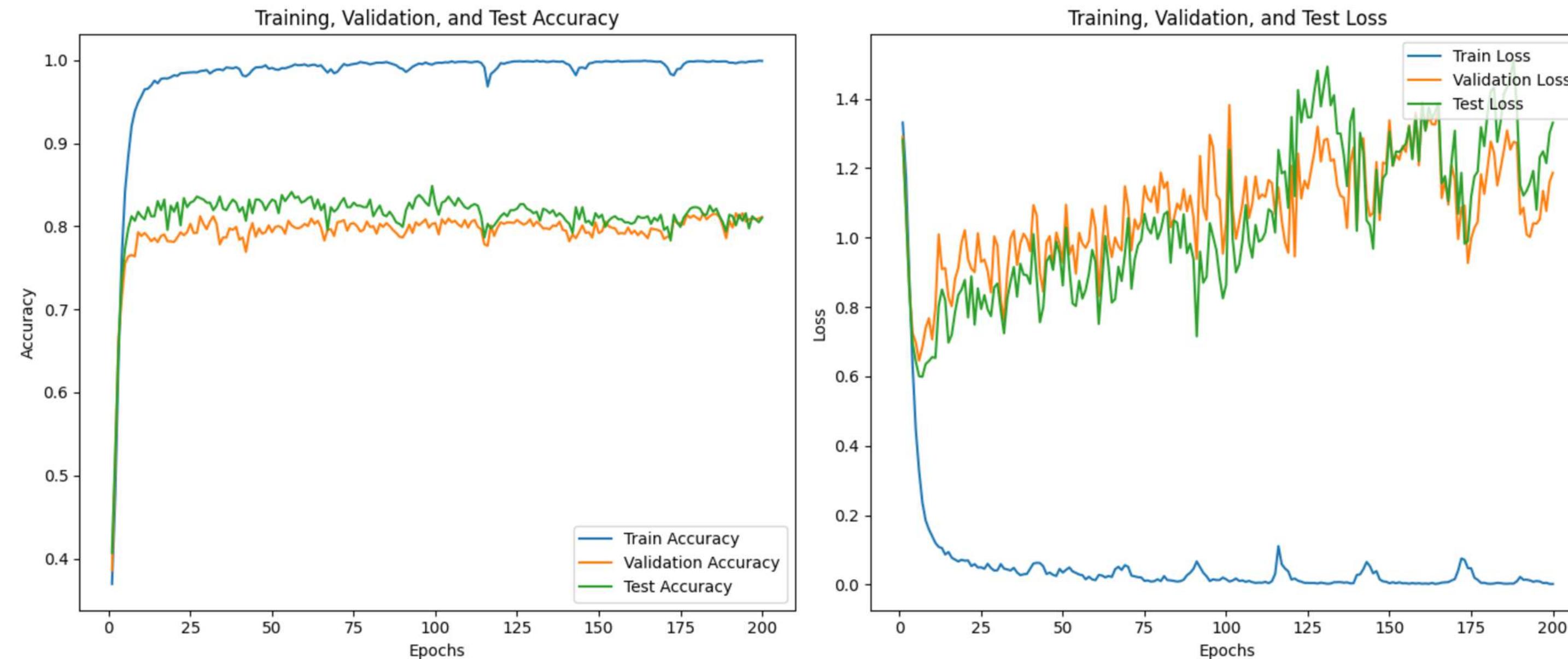
Ứng dụng trong bài toán thực tế

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



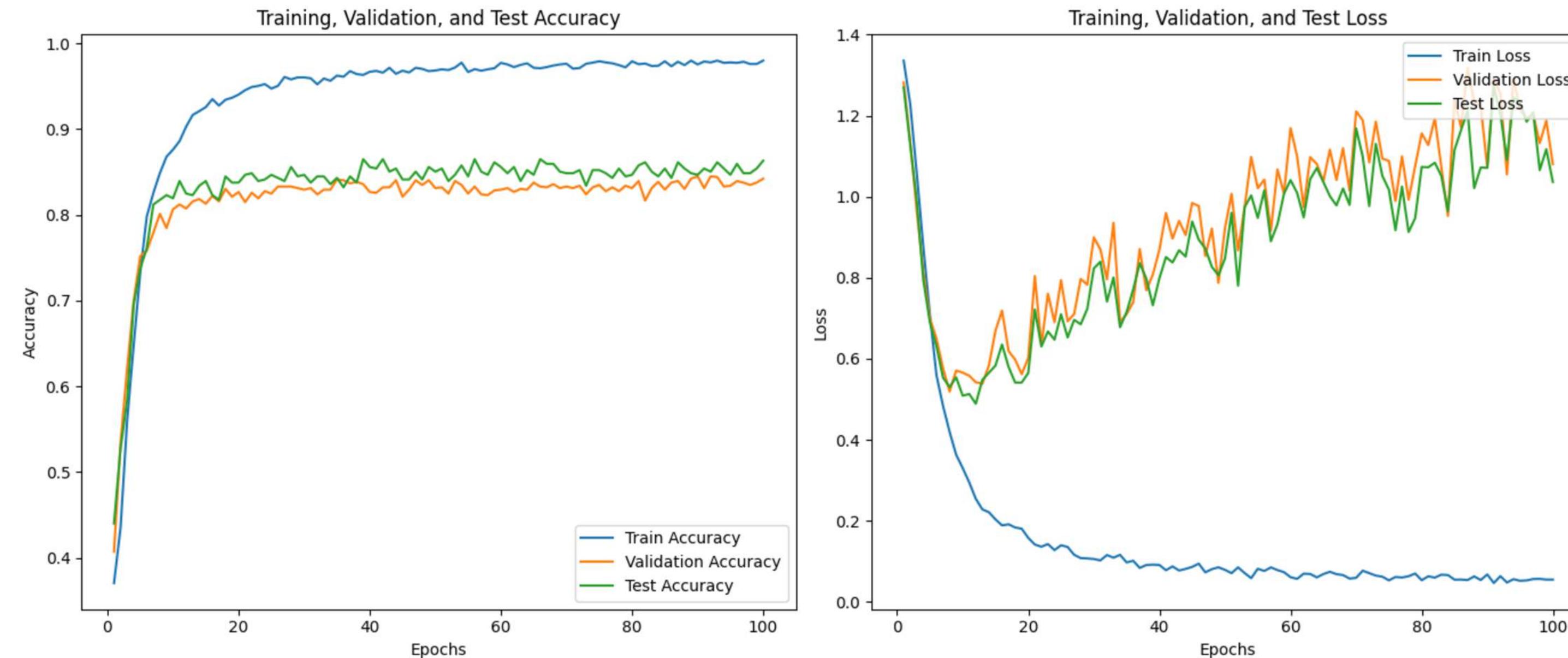
Ứng dụng trong bài toán thực tế

- CNN (L2 Regularization):
 - Validation set - Best Loss: 0.6452 - Best Accuracy: 0.8157
 - Test set - Best Loss: 0.5986 - Best Accuracy: 0.8485



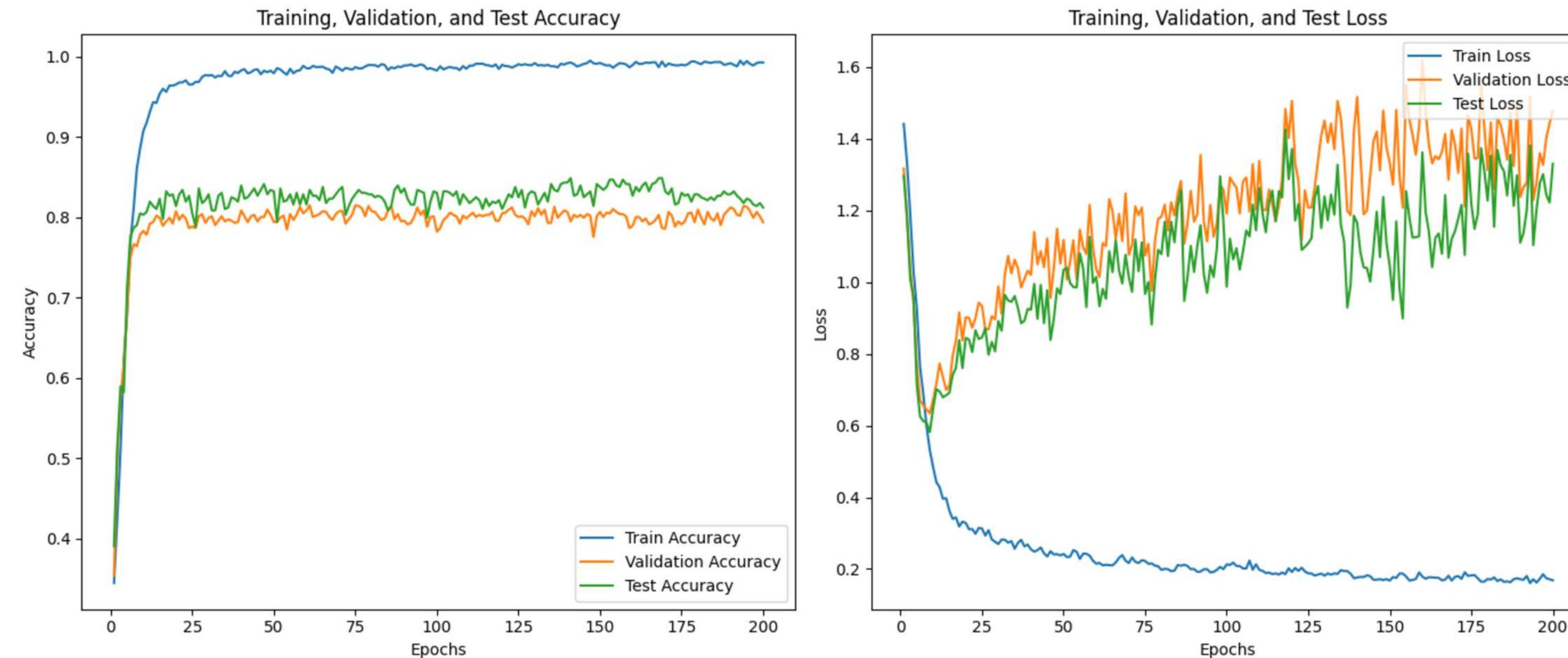
Ứng dụng trong bài toán thực tế

- CNN (Dropout):
 - Validation set - Best Loss: 0.5191 - Best Accuracy: 0.8449
 - Test set - Best Loss: 0.4890 - Best Accuracy: 0.8650



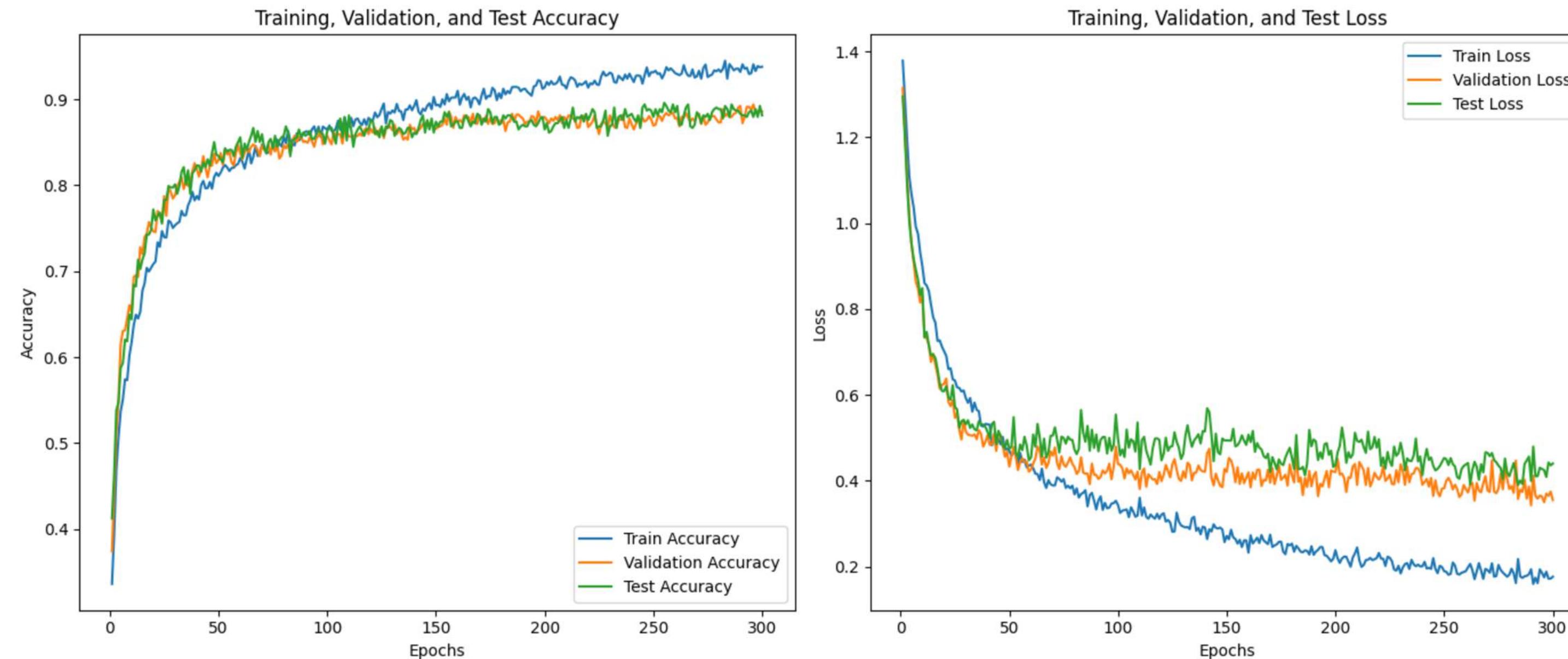
Ứng dụng trong bài toán thực tế

- CNN (DropConnect):
 - Validation set - Best Loss: 0.6345 - Best Accuracy: 0.8148
 - Test set - Best Loss: 0.5821 - Best Accuracy: 0.8485



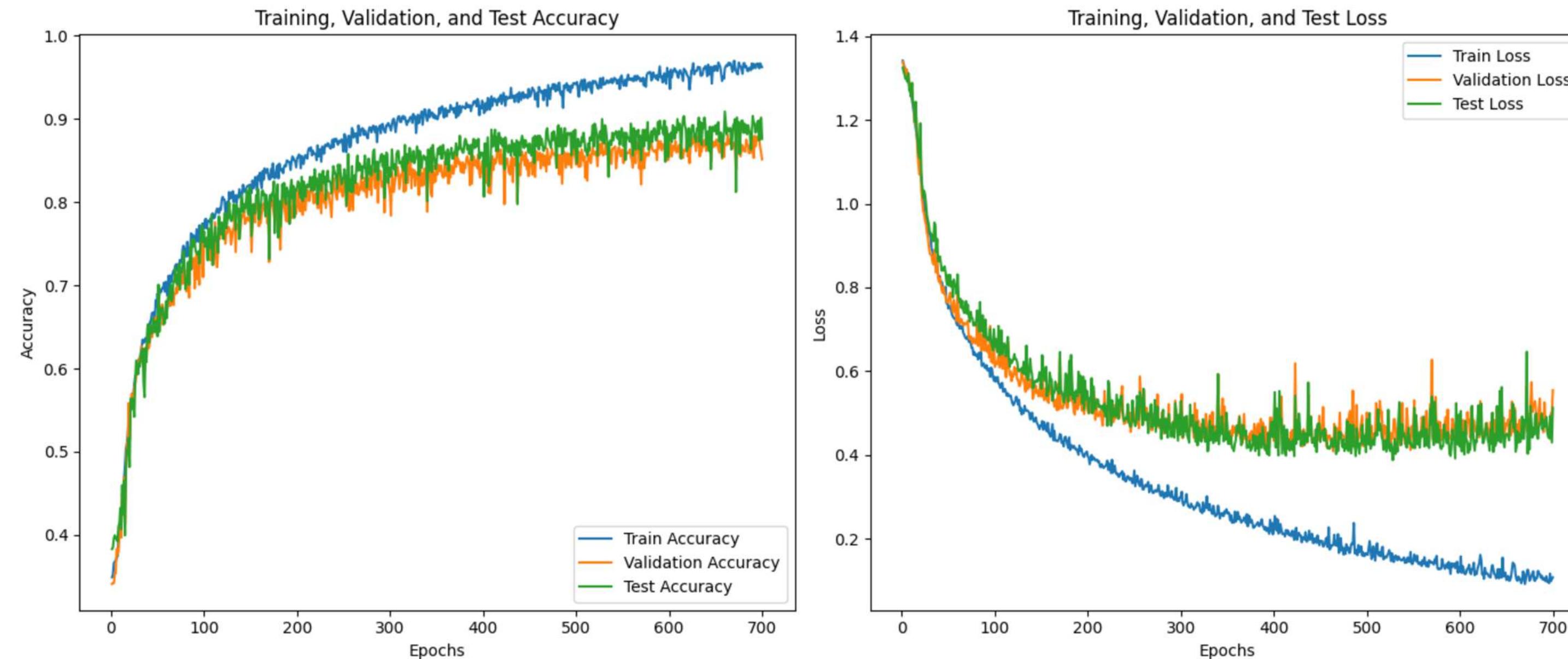
Ứng dụng trong bài toán thực tế

- CNN (Data Augmentation):
 - Validation set - Best Loss: 0.3431 - Best Accuracy: 0.8941
 - Test set - Best Loss: 0.3907 - Best Accuracy: 0.8960



Ứng dụng trong bài toán thực tế

- CNN (GlobalAvgPool):
 - Validation set - Best Loss: 0.4087 - Best Accuracy: 0.8813
 - Test set - Best Loss: 0.3875 - Best Accuracy: 0.9087



Ứng dụng trong bài toán thực tế

Model	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy
CNN	0.6130	0.8130	0.5647	0.8303
CNN + L2 Regularization	0.6452	0.8157	0.5986	0.8485
CNN + Dropout	0.5191	0.8449	0.4890	0.8650
CNN + DropConnect	0.6345	0.8148	0.5821	0.8485
CNN + GlobalPool	0.4087	0.8813	0.3875	0.9087
CNN + Data Augmentation	0.3431	0.8941	0.3907	0.8960

A large, faint watermark of the HUST logo is visible across the entire background of the slide.

HUST

THANK YOU !