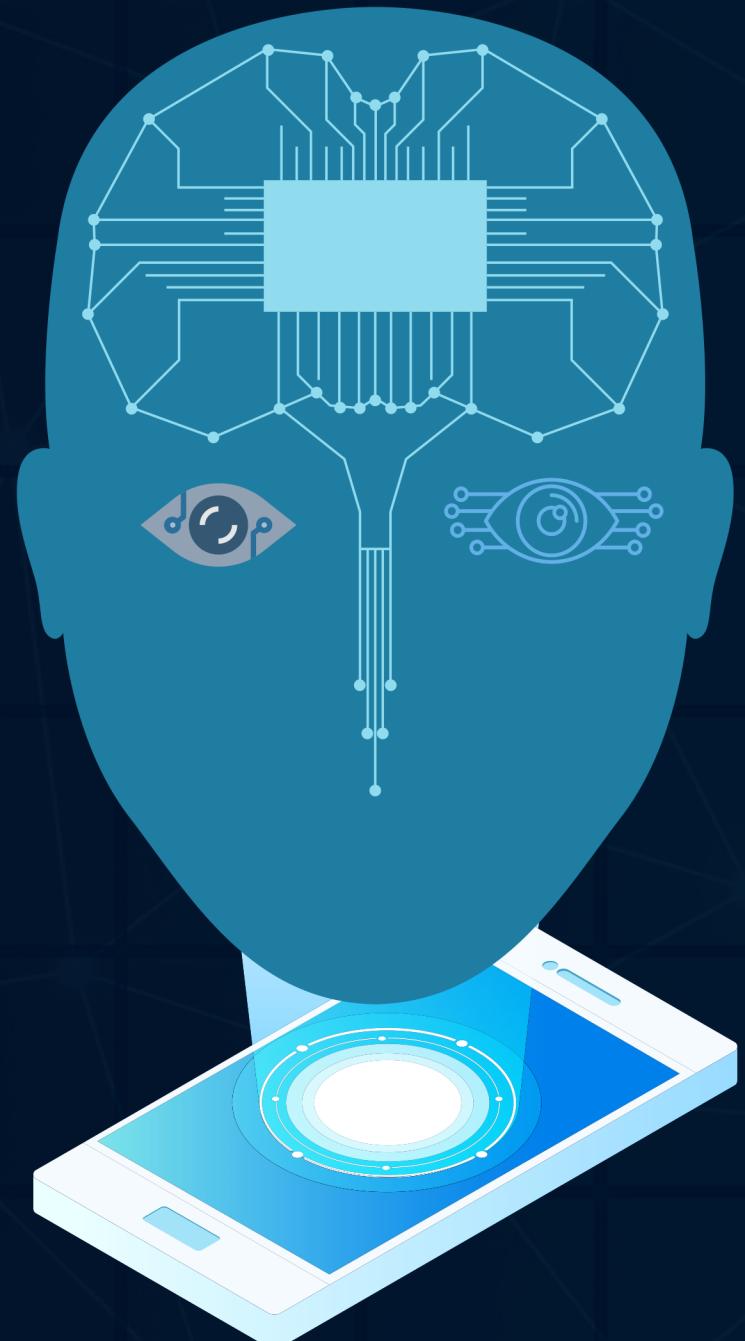




Thị giác máy tinh nâng cao CS331.N21



# Bài toán Captcha recognition

GV hướng dẫn: Ts Lê Minh Hưng

Thành viên:  
Ngô Mai Quốc Thắng  
20520757



# NỘI DUNG THỰC HIỆN

01 Giới thiệu Dataset

02 Convolutional  
recurrent neural  
network (CRNN)

03 Connectionist  
Temporal  
Classification

04 Kết quả

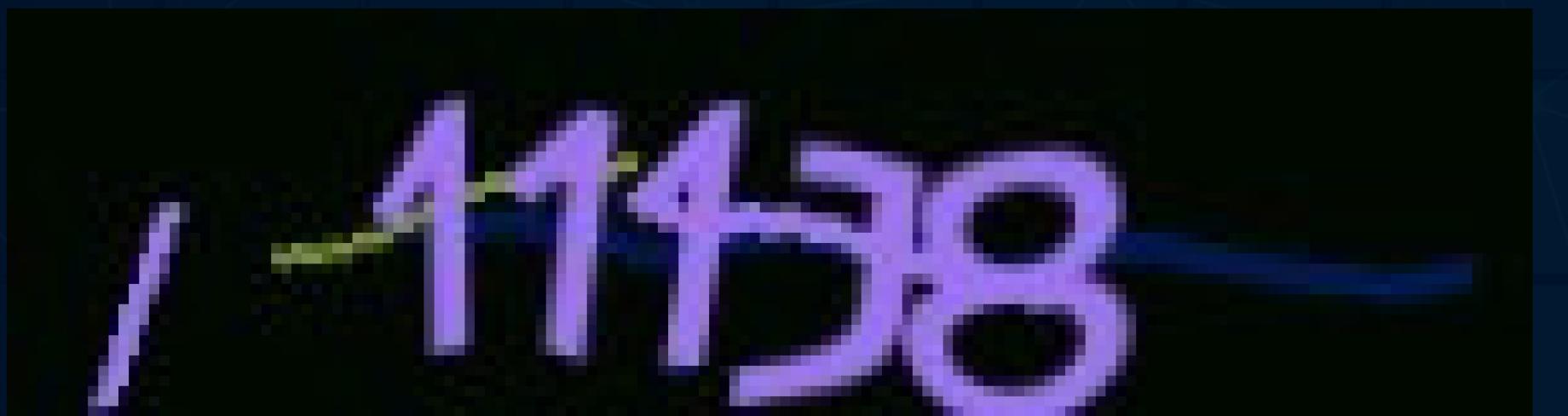
# INTRODUCTION

Convolutional Recurrent Neural Network (CRNN) là một kiến trúc được thiết kế chuyên biệt để giải quyết nhiệm vụ Text Recognition trong bài toán OCR. Xuất hiện trong bài báo An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition xuất bản năm 2015, cho đến nay, nó vẫn được coi là một trong những model hiệu quả nhất cho việc thực hiện Text Recognition.

# DATASET

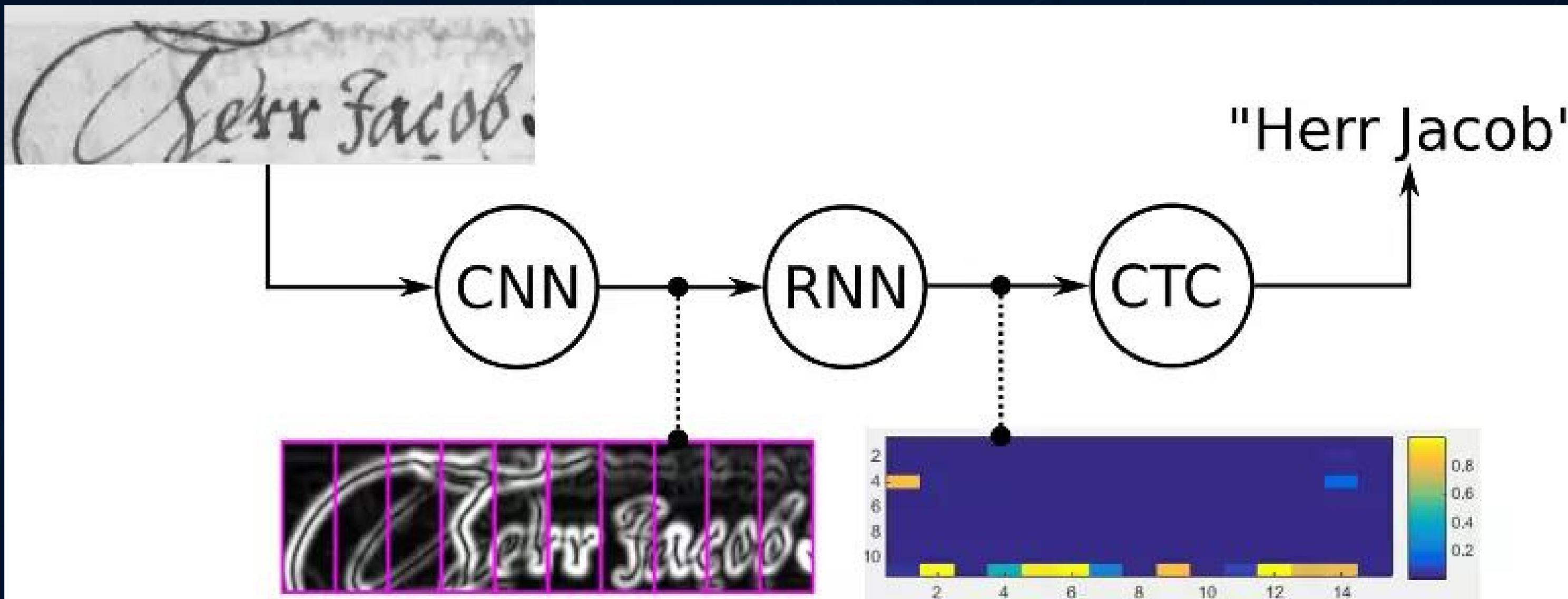
Sử dụng bộ dataset CAPTCHA Dataset có sẵn trên Kaggle.

Bộ dữ liệu có tổng cộng 113062 file ảnh có màu và định dạng .jpg





# CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN)

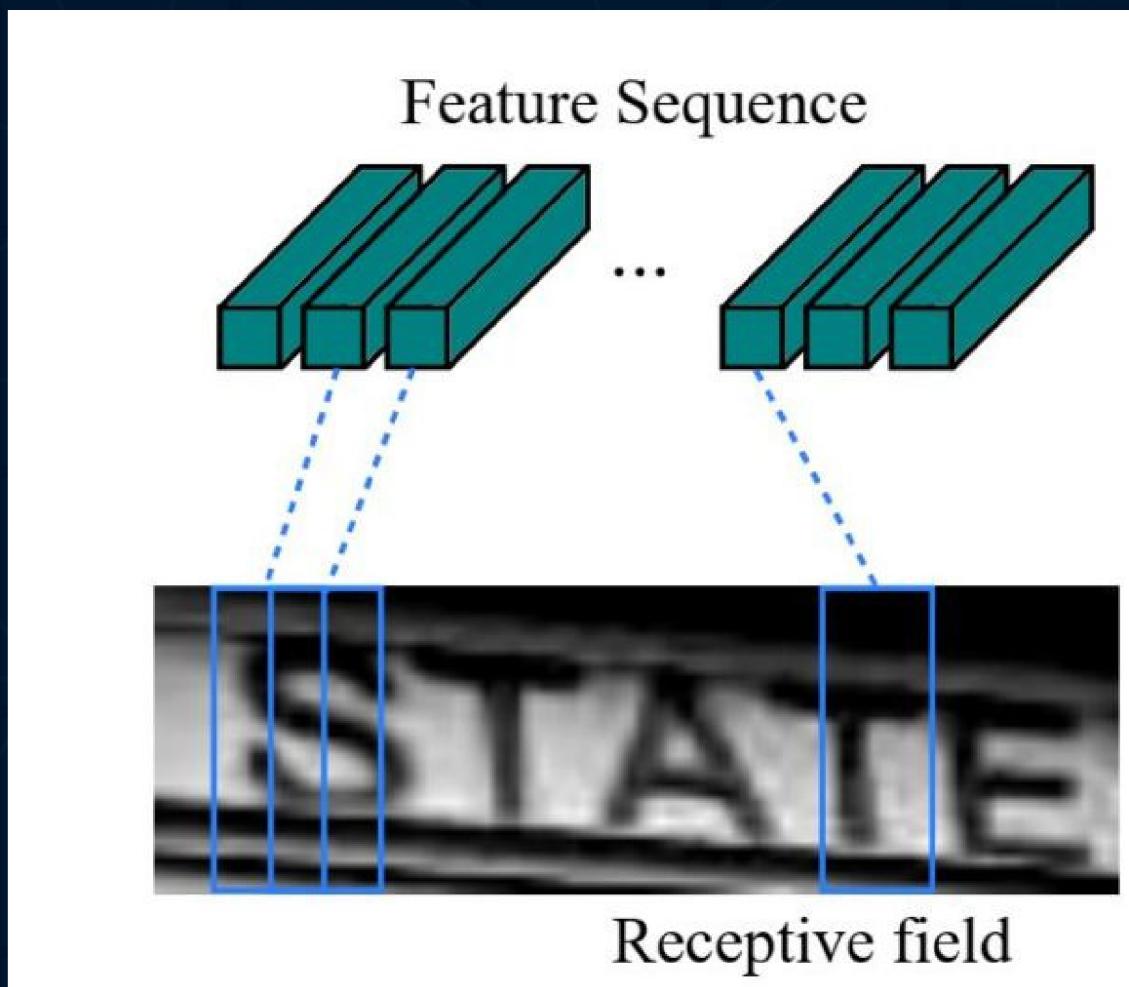


# CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN)

1

## Convolutional Layers

ẢNH ĐẦU VÀO ĐƯỢC CHO ĐI QUA CÁC LỚP CONV, SINH RA CÁC FEATURE MAPS. CÁC FEATURE MAPS SAU ĐÓ LẠI ĐƯỢC CHIA RA THÀNH MỘT CHUỖI CỦA CÁC FEATURE VECTORS (CÁC TIMESTEPS), GỌI LÀ FEATURE SEQUENCE.

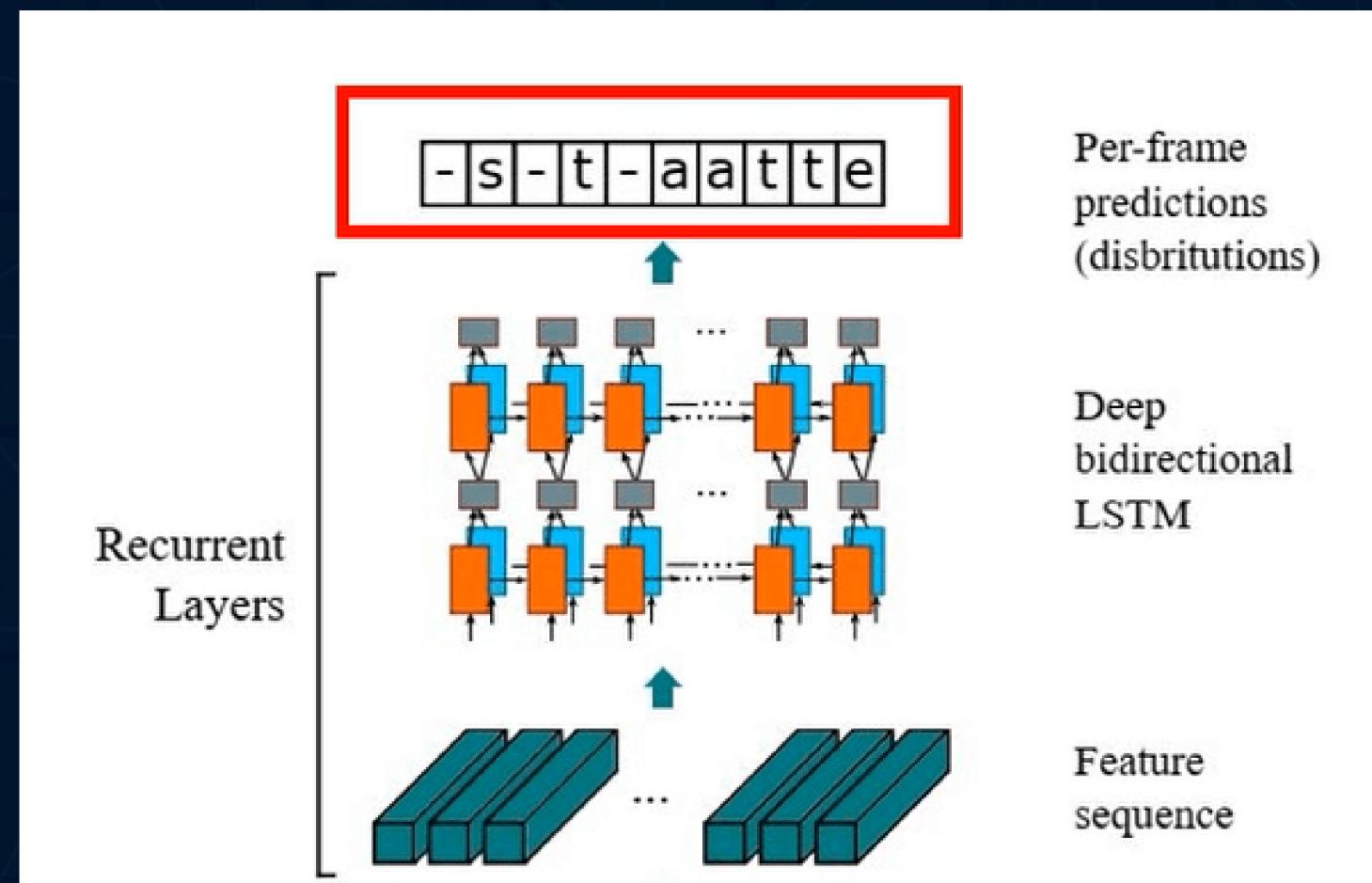


# CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN)

2

## Recurrent Layers

- Feature Sequence được đưa vào các lớp Bidirectional LSTM, sinh ra một chuỗi các ký tự - mỗi ký tự tương ứng với một TimeStep trong Feature Sequence.
- Tuy nhiên, chuỗi đầu ra của LSTM cũng rất lộn xộn: trùng lặp, không có ký tự, ...



# CONNECTIONIST TEMPORAL CLASSIFICATION

**CTC (Connectionist Temporal Classification)** là một hàm mất mát được sử dụng trong huấn luyện các mô hình Deep Learning.

- **Mục tiêu chính của CTC là tìm cách ánh xạ (alignment) giữa một đầu vào X và đầu ra Y.**
- **CTC không yêu cầu dữ liệu được gắn nhãn theo từng TimeStep cụ thể, mà nó có khả năng đưa ra xác suất cho mỗi khả năng ánh xạ từ X sang Y.**
- **CTC chỉ yêu cầu đầu vào là một hình ảnh (ma trận Feature của hình ảnh) và đoạn Text tương ứng với hình ảnh đó.**

# CONNECTIONIST TEMPORAL CLASSIFICATION

1

## Encoding the text

- Lý tưởng thì mỗi TimeStep sẽ tương ứng với một ký tự. Nếu một ký tự tồn tại trong cả 2 TimeSteps, CTC giải quyết vấn đề này bằng cách gộp tất cả các ký tự trùng nhau thành một.
- Tuy nhiên, nếu làm như vậy với từ mà bản thân nó có các ký tự trùng nhau thì kết quả bị sai lệch. Để tiếp tục xử lý vấn đề, CTC sử dụng một ký tự giả, gọi là blank và ký hiệu là “-”. Tất cả những ký tự lặp lại ký tự lặp sát nhau cần phải được thêm blank vào giữa chúng
- Ví dụ hello → hel-lo hoặc hel--lo.

# CONNECTIONIST TEMPORAL CLASSIFICATION



## Loss Calculate



Hàm mất mát CTC là negative log likelihood

$$O^{ML}(S, N_w) = - \sum_{(x,z) \in S} \ln(p(l|x))$$

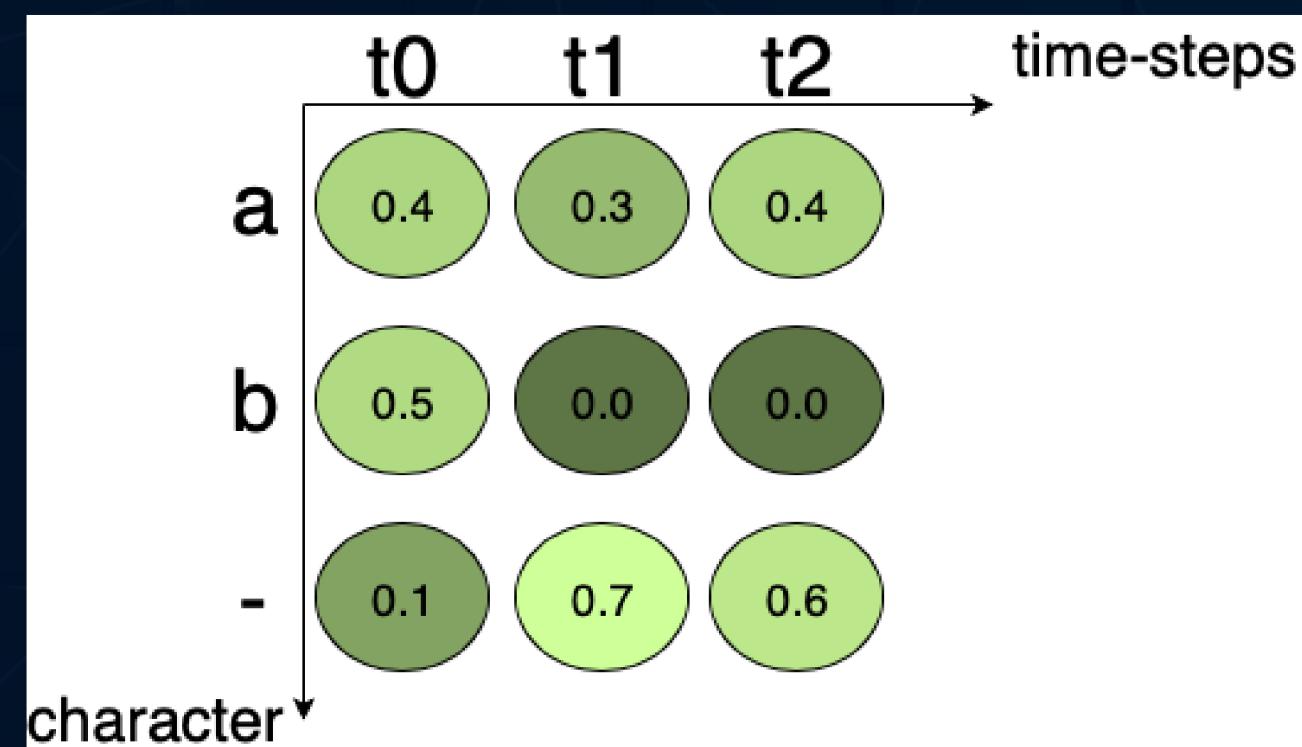
# CONNECTIONIST TEMPORAL CLASSIFICATION



## Loss Calculate



Giả sử chúng ta có một ma trận Score là Output của CRNN như sau:



- ☞ Các khả năng Alignment của ký tự *a* là: *aaa*, *a-*, *-a*, *aa-*, *-aa*, *--a* → Score của *\*a* =  $0.4 \times 0.3 \times 0.4 + 0.4 \times 0.7 \times 0.6 + 0.4 \times 0.7 + 0.4 \times 0.3 \times 0.6 + 0.1 \times 0.3 \times 0.4 + 0.1 \times 0.7 \times 0.4 = 0.608$ . \* → Loss =  $-\log_{10} 0.6084 = 0.216$
  - ☞ Các khả năng Alignment của ký tự *b* là: *bbb*, *b-*, *-b*, *bb-*, *-bb*, *--b* → Score của *b* =  $0.5 \times 0.0 \times 0.0 + 0.5 \times 0.7 \times 0.6 + 0.5 \times 0.7 + 0.5 \times 0.0 \times 0.6 + 0.1 \times 0.0 \times 0.0 + 0.1 \times 0.7 \times 0.0 = 0.56$  → \*Loss =  $\log_{10} 0.56 = 0.25$
  - ☞ Các khả năng Alignment của ký tự *blank* là: *-*, *--* → Score của *blank* =  $0.1 \times 0.7 + 0.1 \times 0.7 \times 0.6 = 0.112$  → Loss =  $-\log_{10} 0.112 = 0.95$
- Tổng Loss =  $0.216 + 0.25 + 0.95 = 1.416$ .

# CONNECTIONIST TEMPORAL CLASSIFICATION



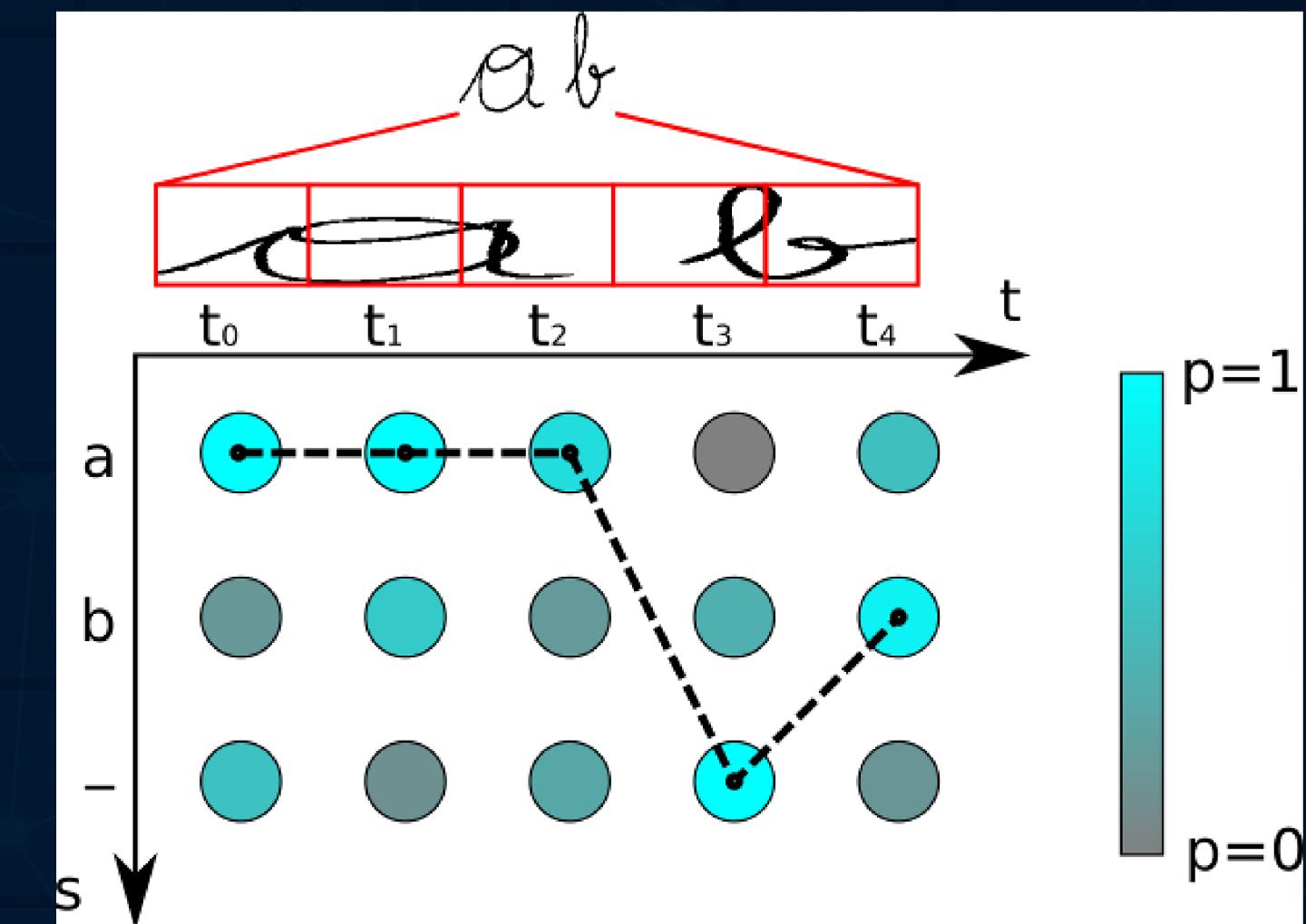
## Decoding Text

Quá trình Decoding một diễn ra như sau:

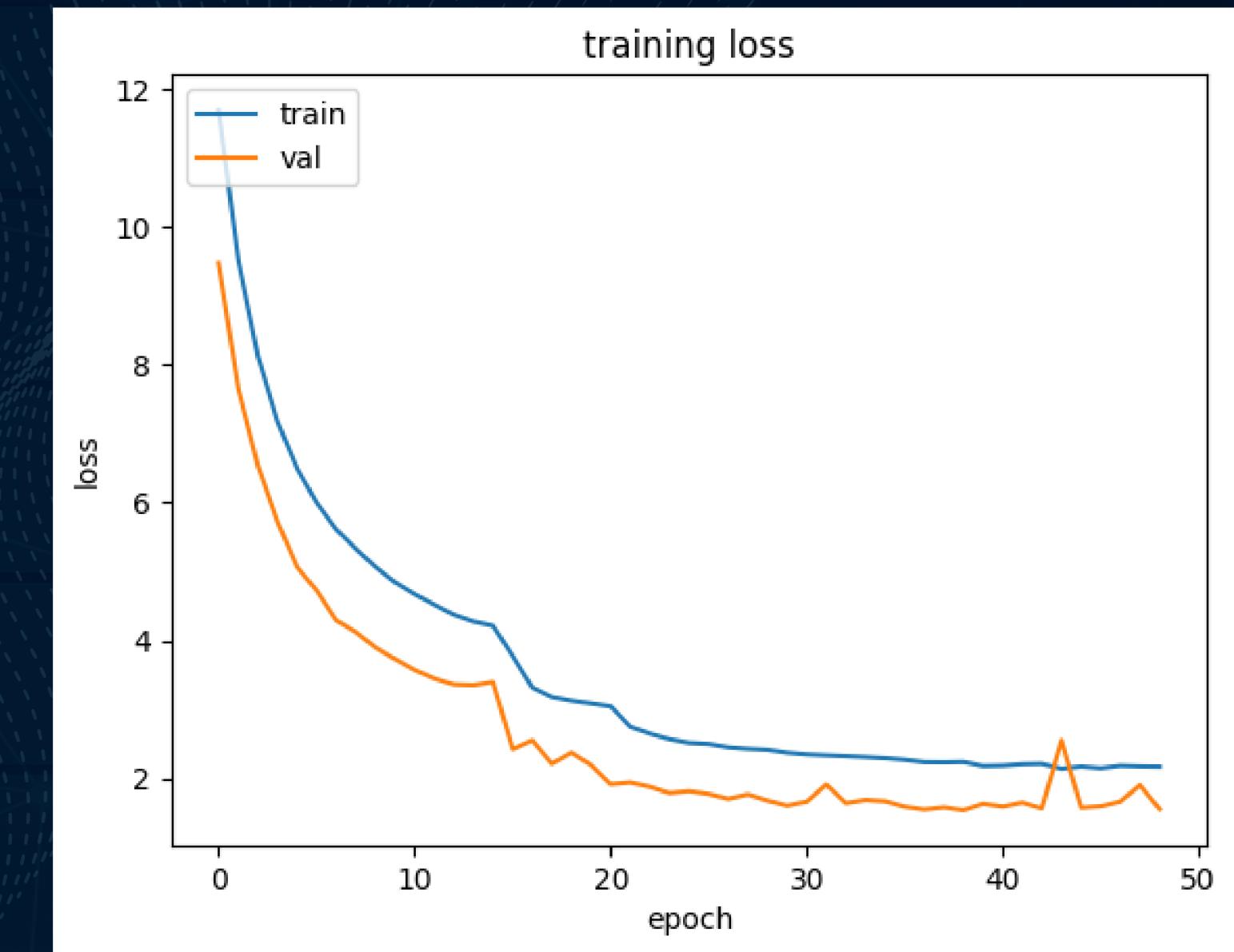
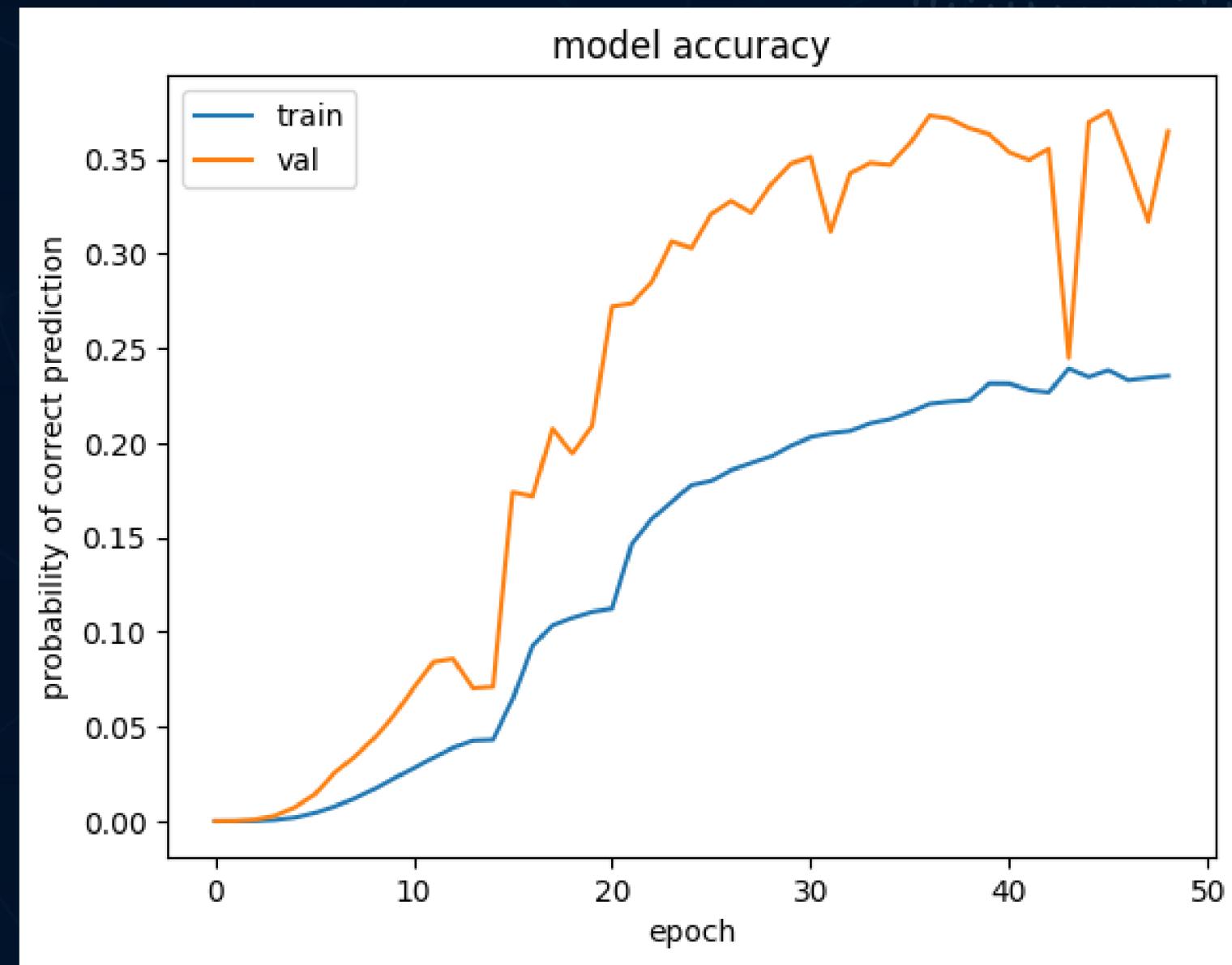
- Tìm đường đi tối ưu nhất từ Score Matrix bằng cách chọn các ký tự có Score cao nhất tại mỗi TimeStep.
- Xóa bỏ các ký tự trùng, ký tự trùng lặp.

Ví dụ: các ký tự là {a, b, -} và có 5 time-steps.

- tại time-step t<sub>0</sub>, ký tự hợp lý nhất là 'a', tương tự với t<sub>1</sub>, t<sub>2</sub>, cũng vậy.
- Ký tự blank có xác suất cao nhất tại time-step t<sub>3</sub>
- time-step t<sub>4</sub> là 'b'. Đường dẫn hợp lý sẽ là 'aaa-b'.
- Bây giờ ta bỏ các ký tự lặp và ký tự blank được kết quả cuối cùng là 'ab'.



# KẾT QUẢ THỰC NGHIỆM





# TÀI LIỆU THAM KHẢO

OCR MODEL FOR READING CAPTCHAS

[https://keras.io/examples/vision/captcha\\_ocr/](https://keras.io/examples/vision/captcha_ocr/)

NHẬN DIỆN TEXT TRONG HÌNH ẢNH VỚI CRNN+CTC

<https://viblo.asia/p/nhan-dien-text-trong-hinh-anh-voi-crnnctc-Eb85o9rBZ2G>

THANKS!

