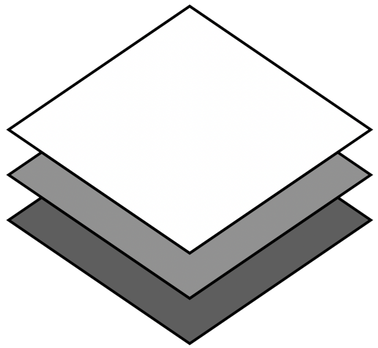
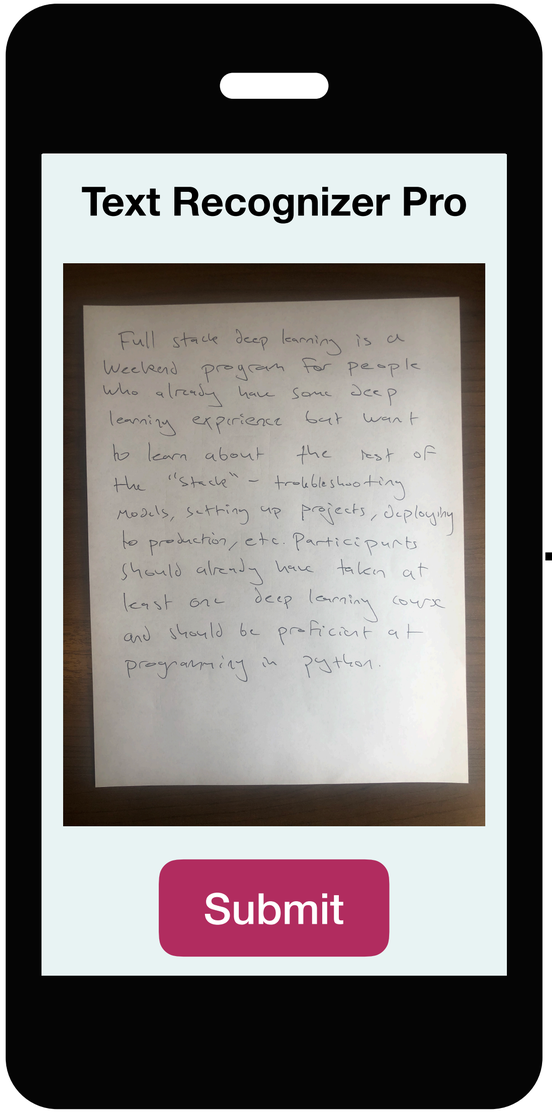


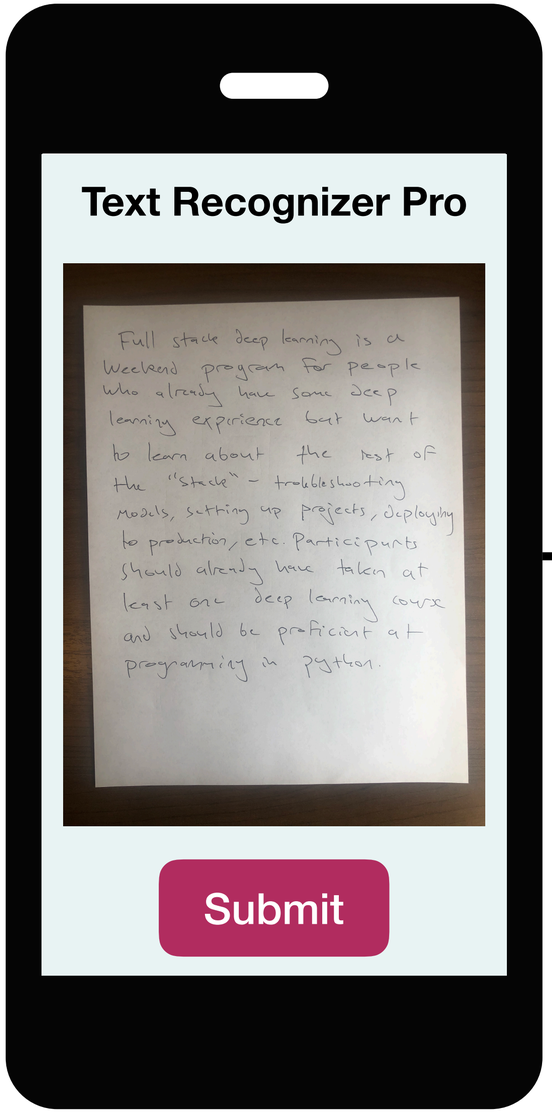
Full Stack Deep Learning

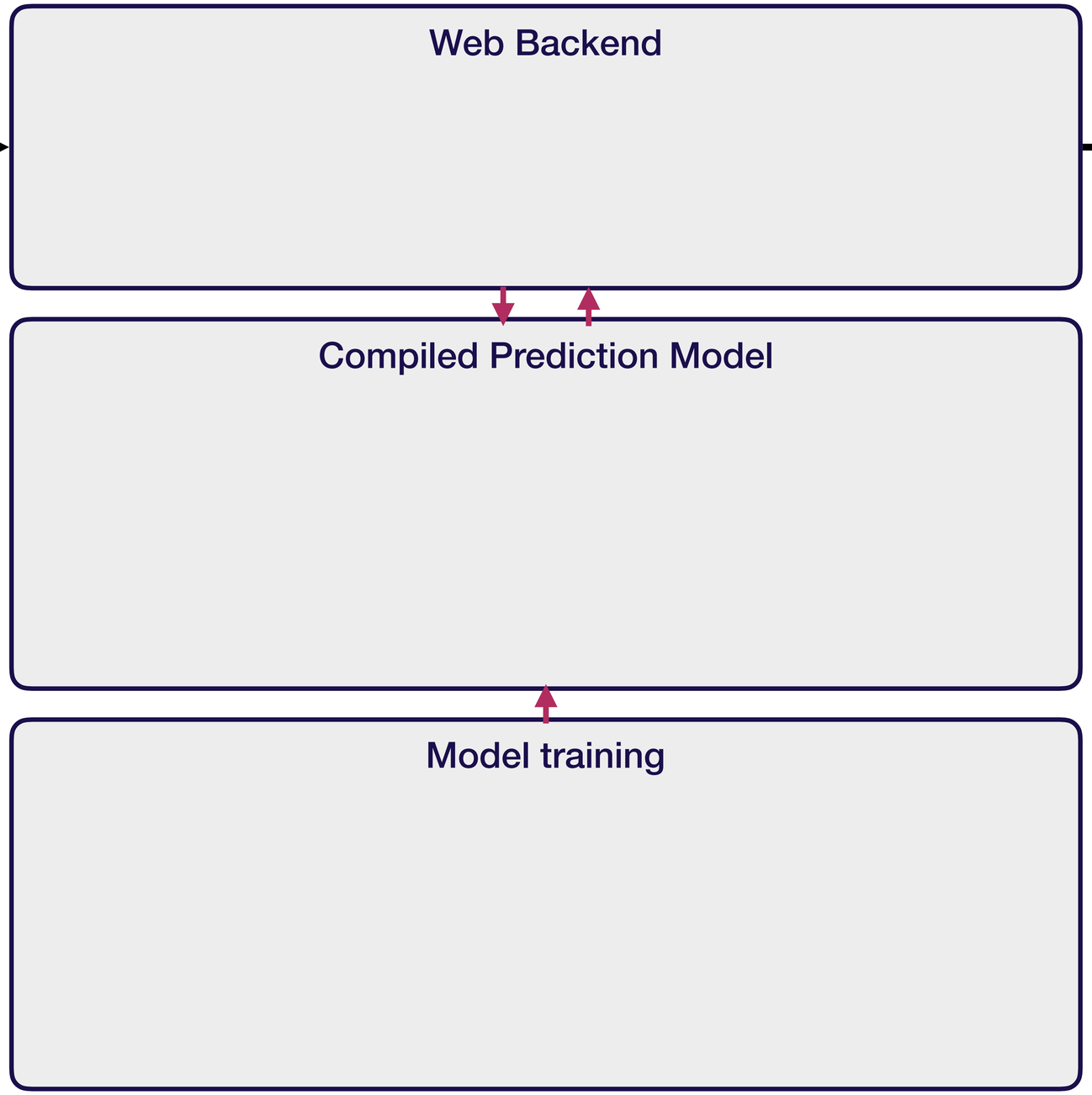
Introduction to the Text Recognizer Lab

Josh Tobin, Sergey Karayev, Pieter Abbeel

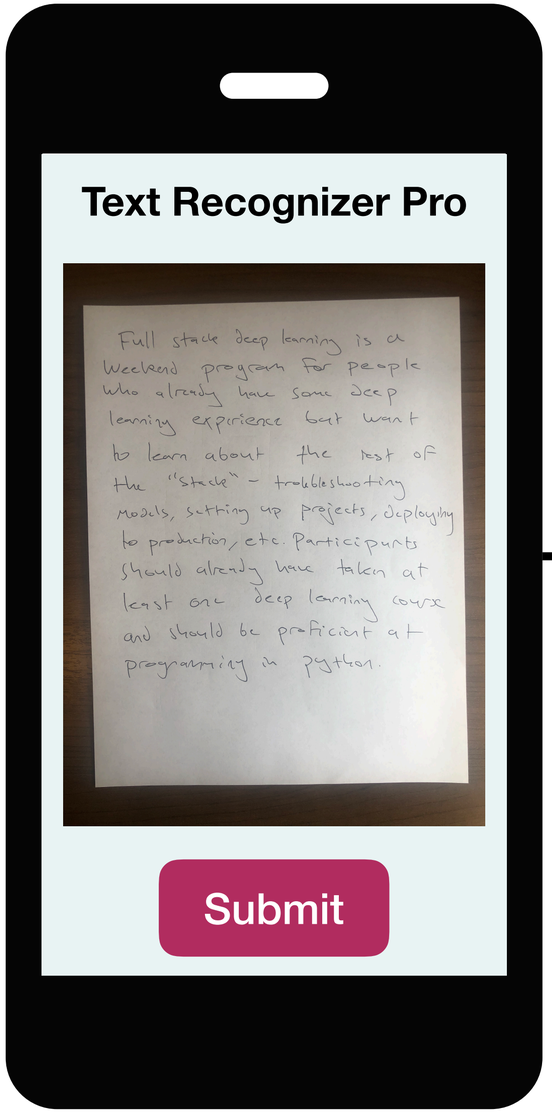


Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" ...





Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …



Web Backend

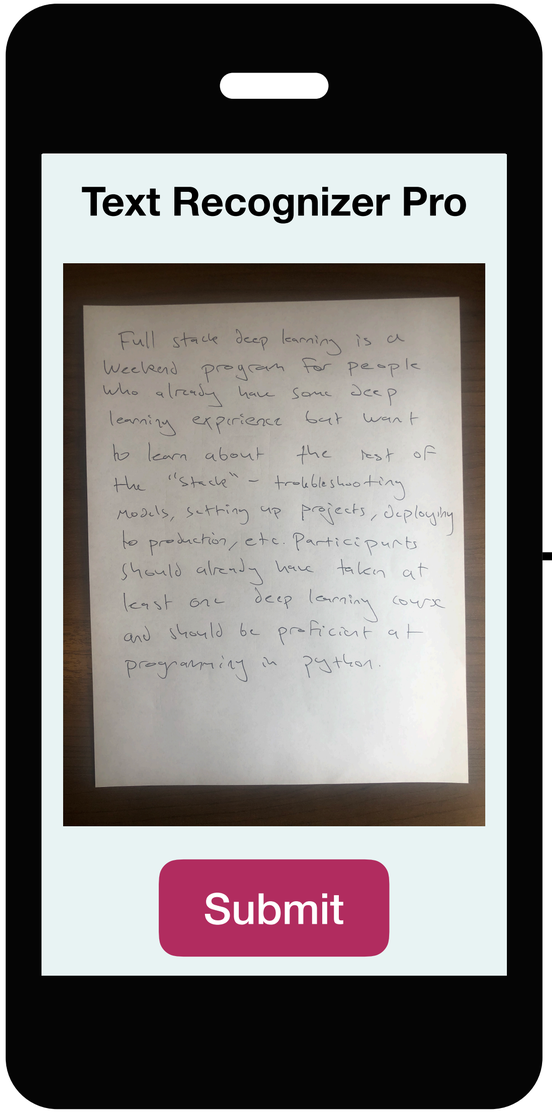
'POST' request

{'image': b'%\xe3L\xa6r<\x1b\xdc\x0c\x11:\xfb\xd9\xeb …}

Compiled Prediction Model

Model training

Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …

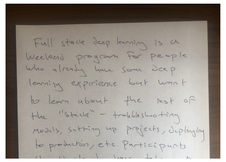


Web Backend

'POST' request

{'image': b'%\xe3L\xa6r<\x1b\xdc\x0c\x11:\xfb\xd9\xeb …}

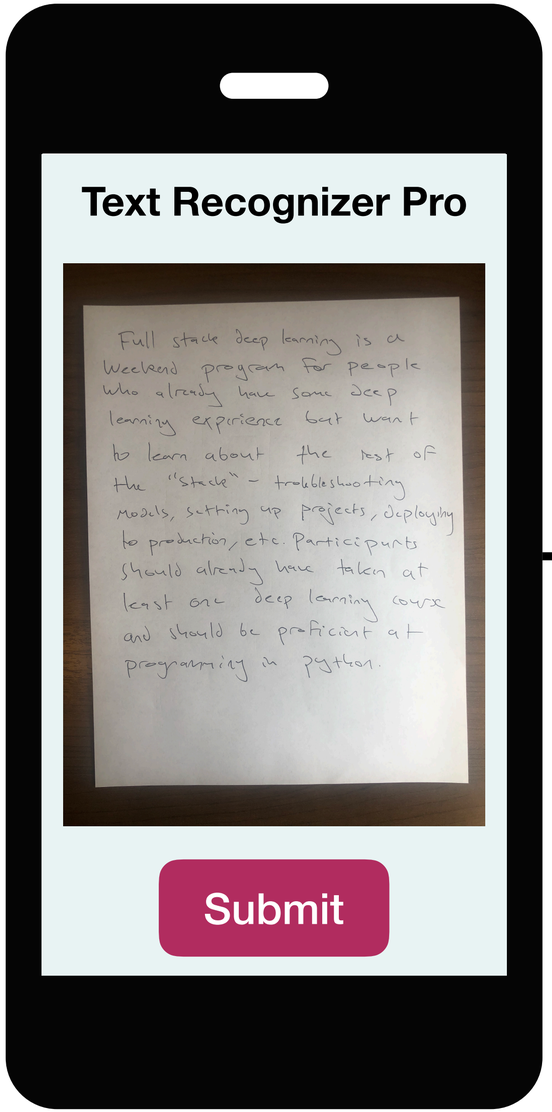
Decode image

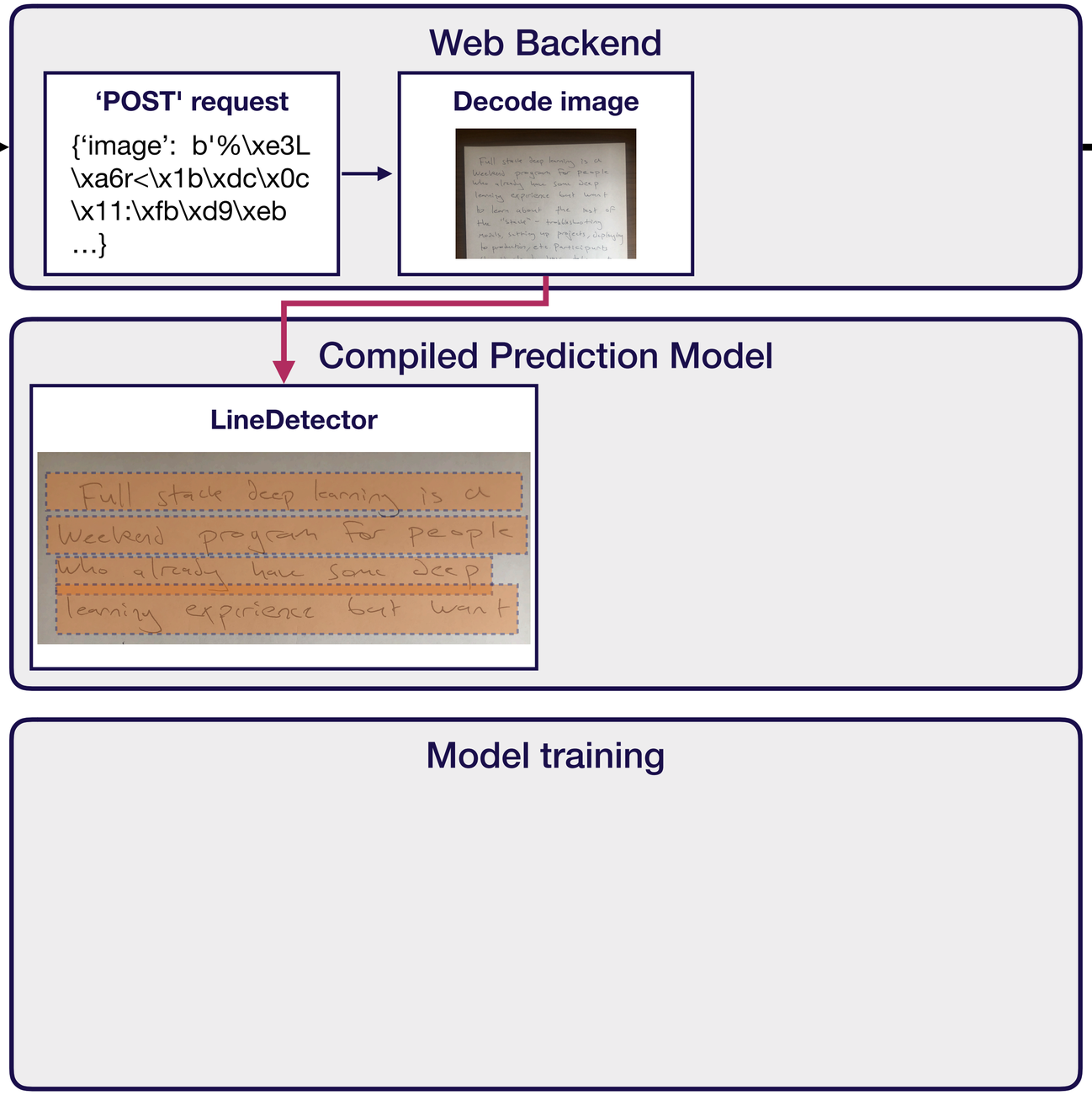


Compiled Prediction Model

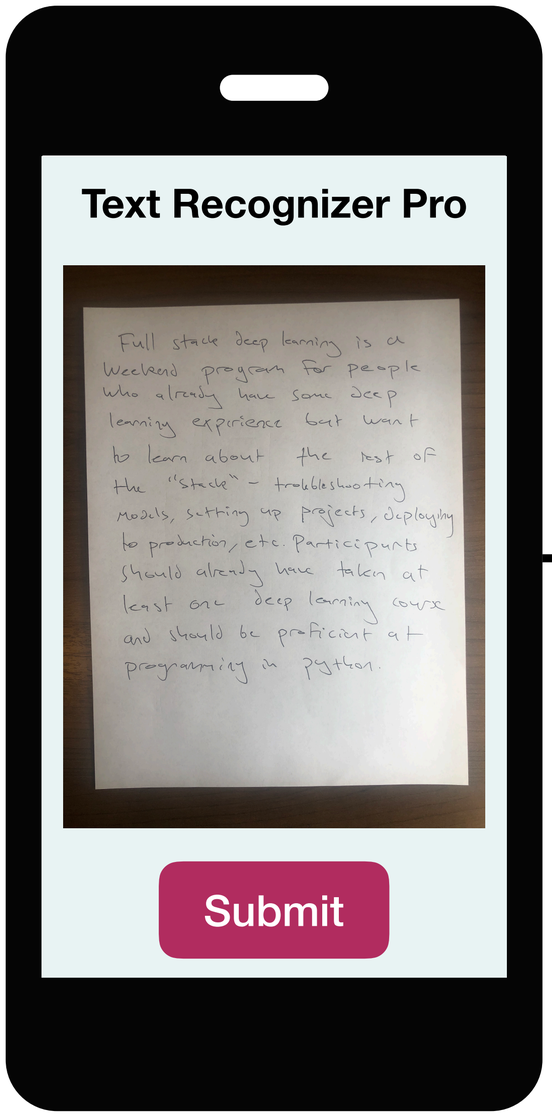
Model training

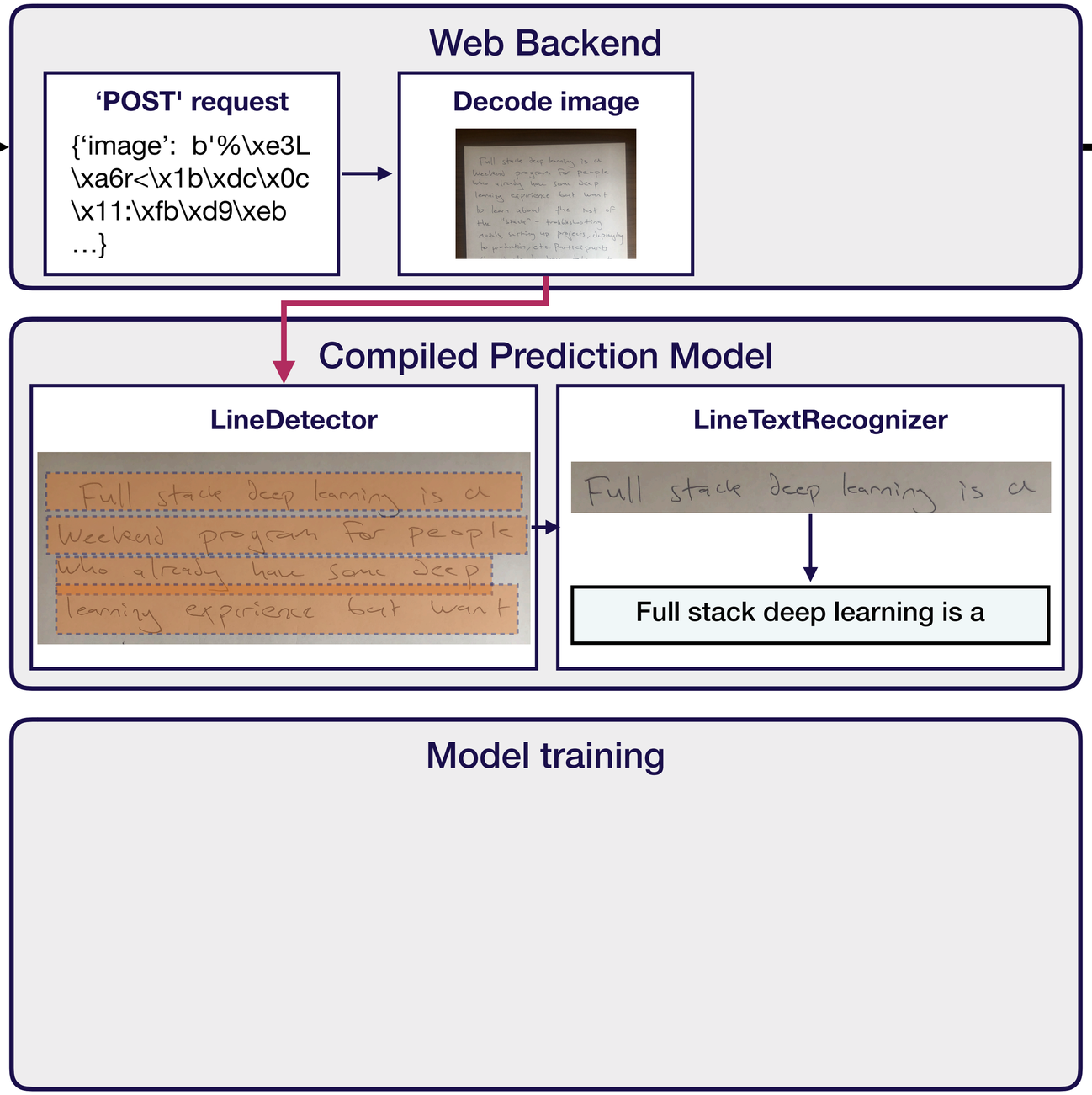
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …



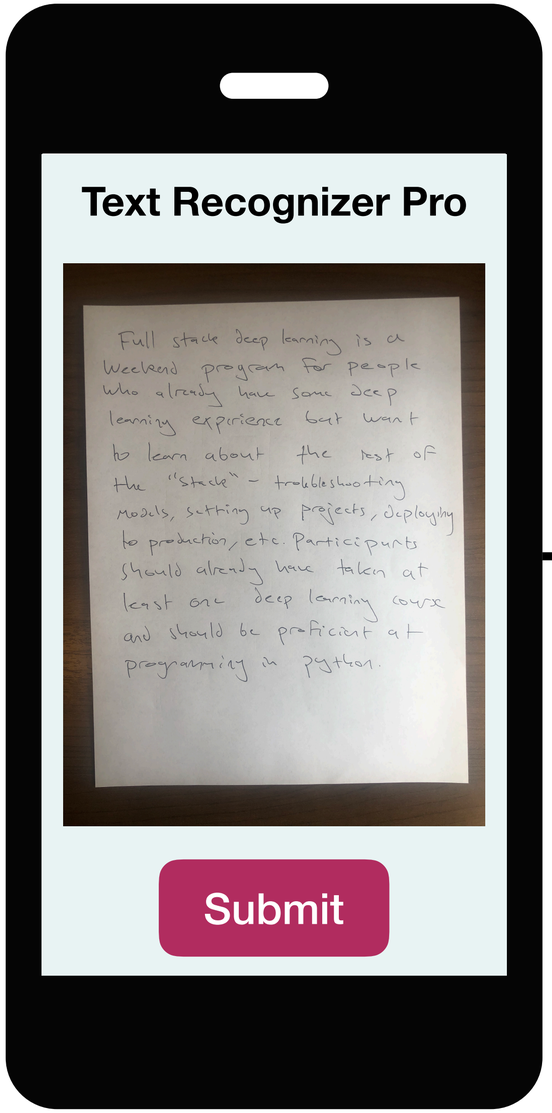


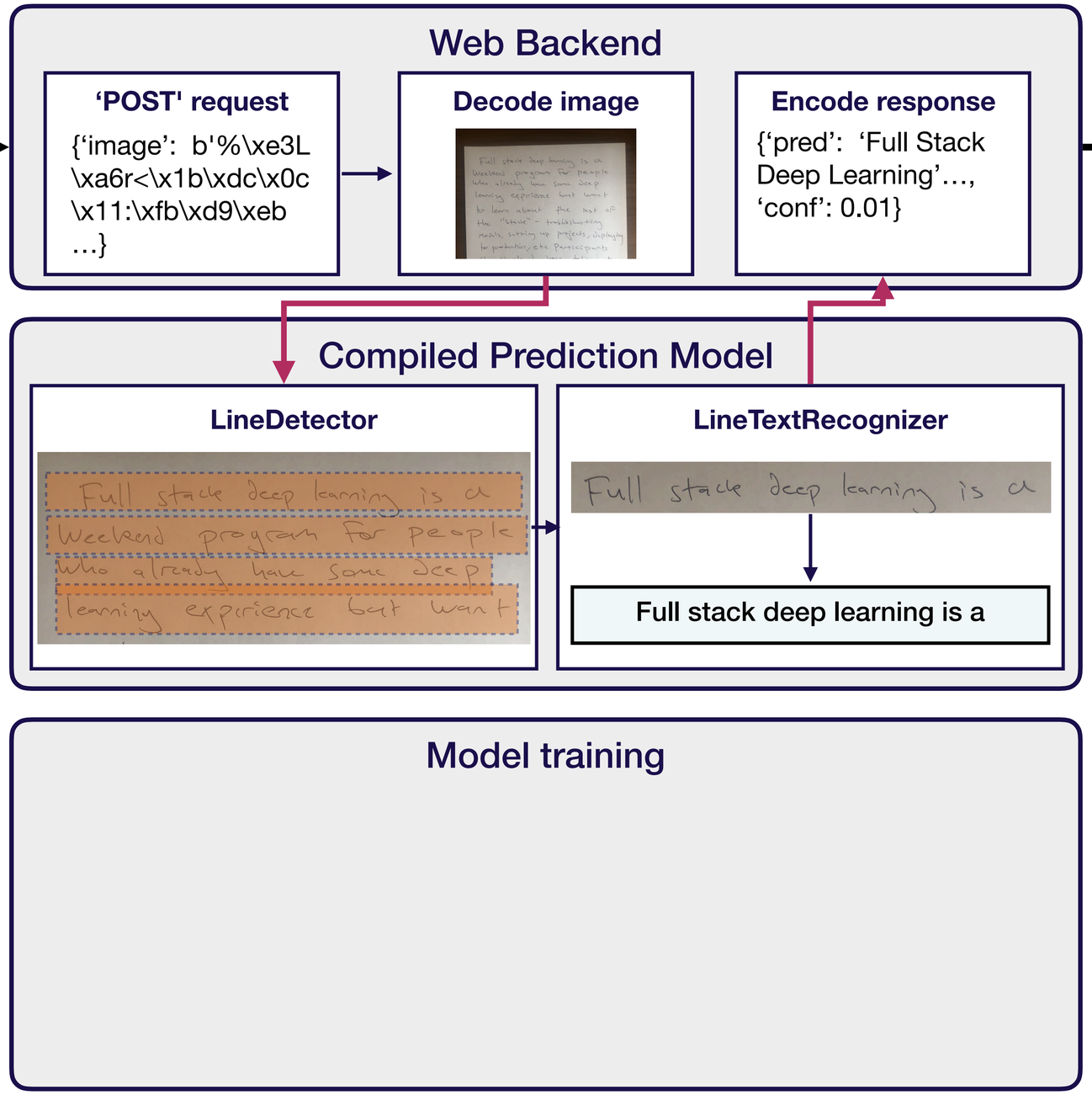
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" ...



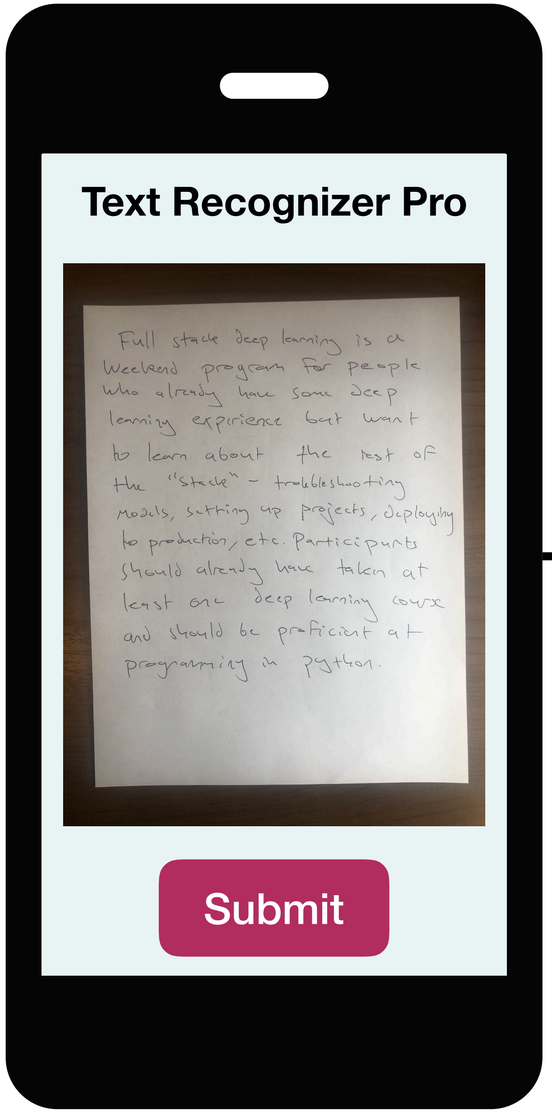


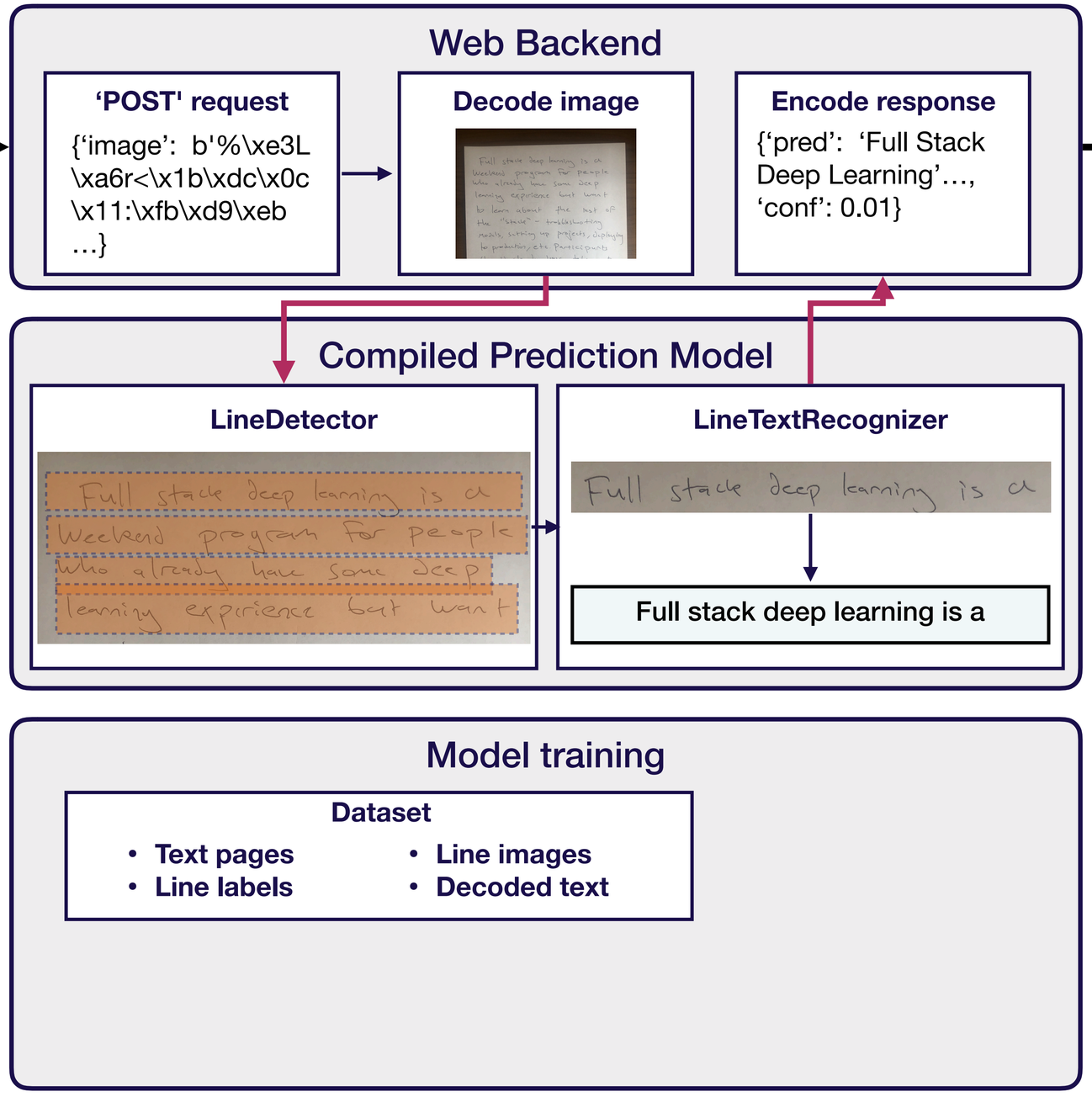
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" ...



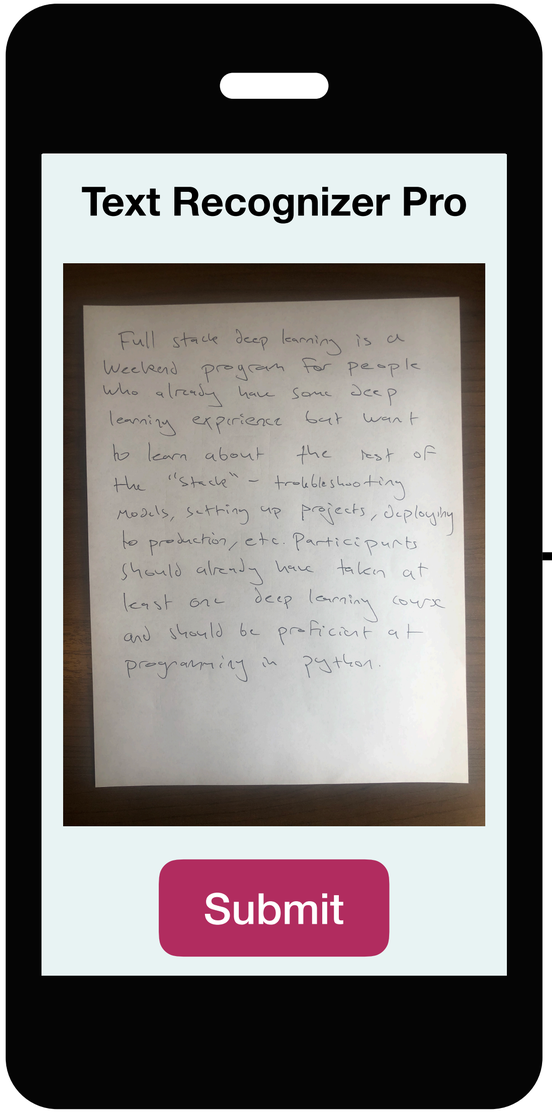


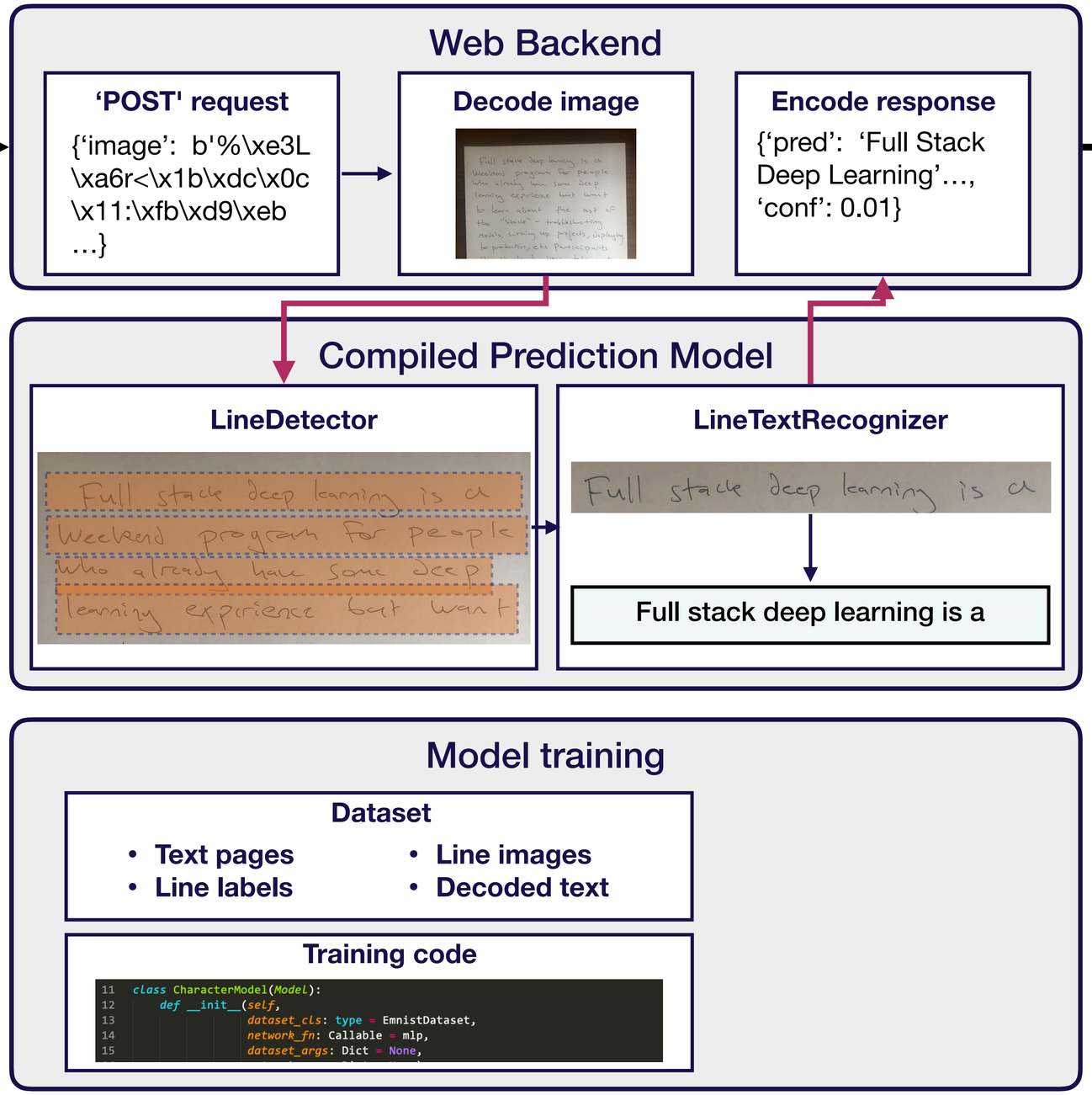
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" ...



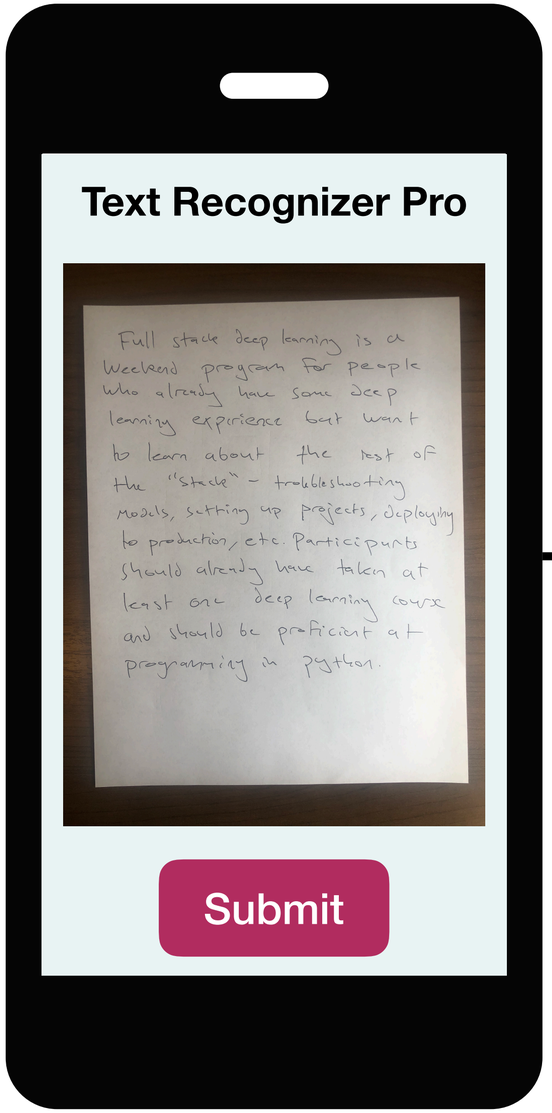


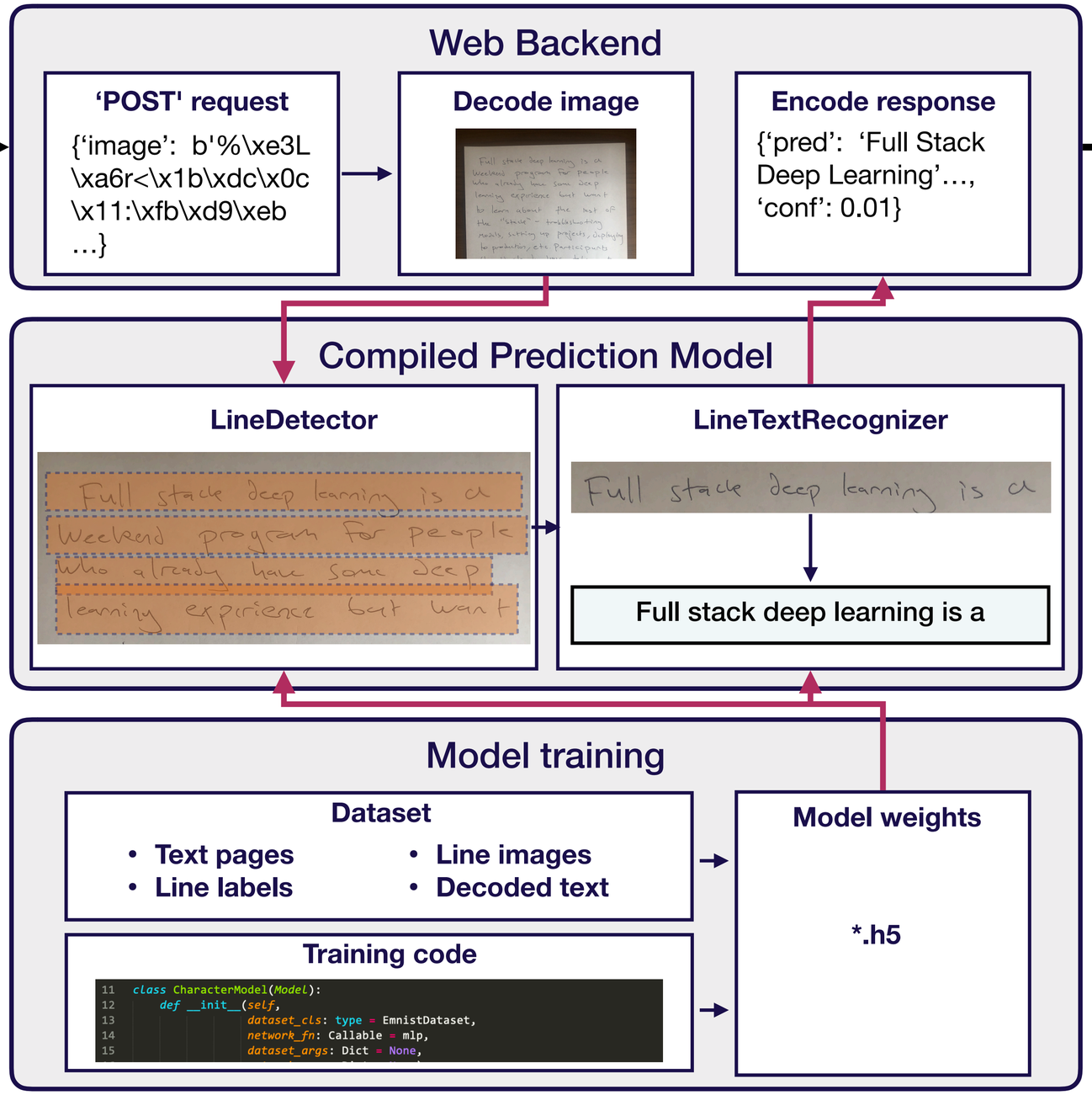
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …



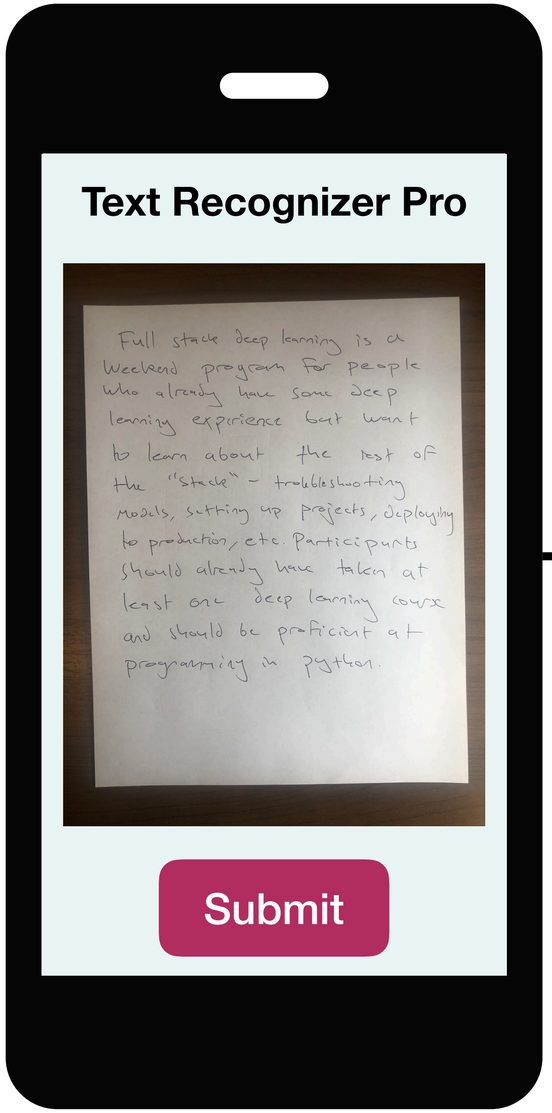


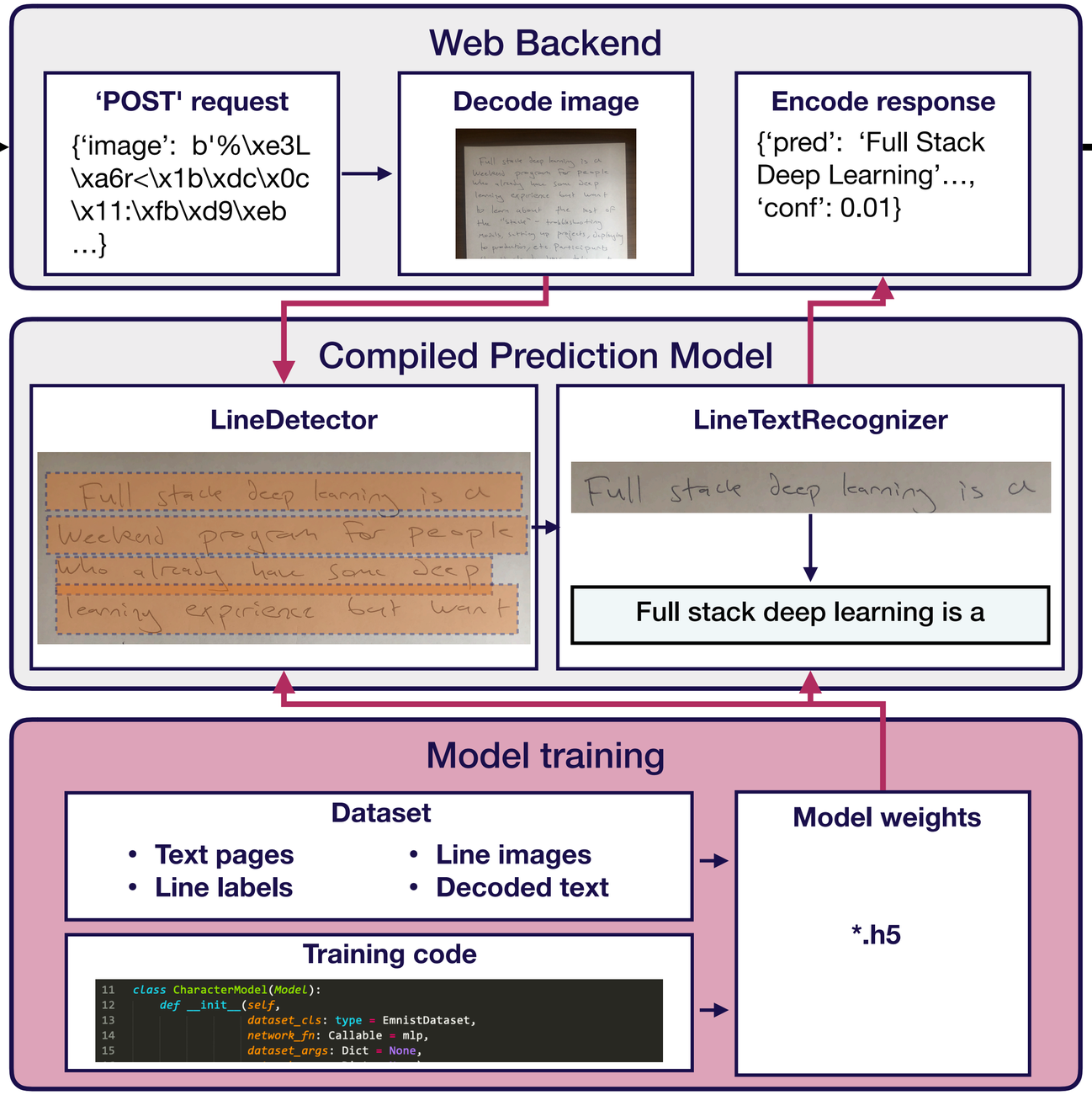
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …



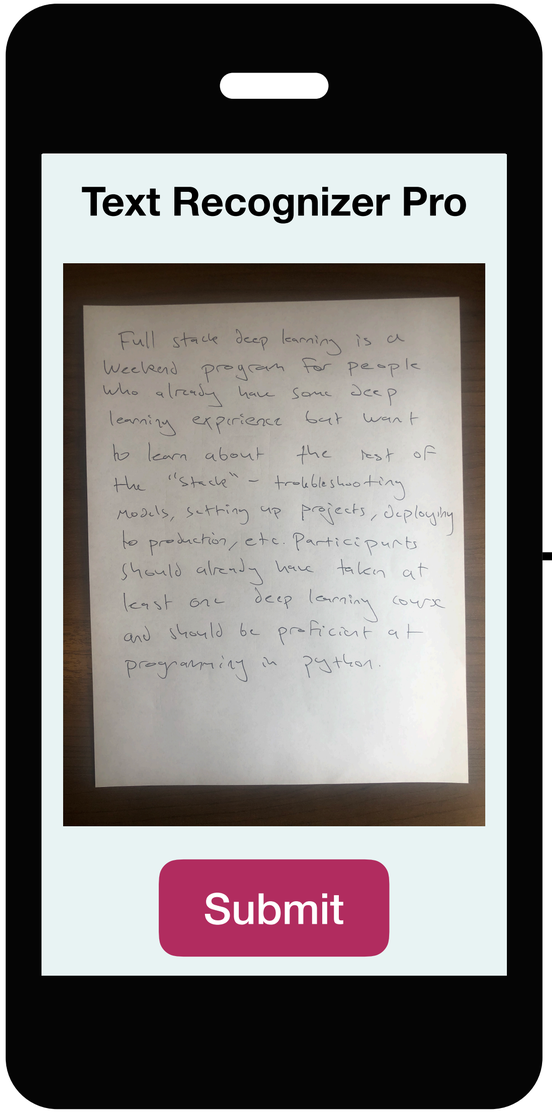


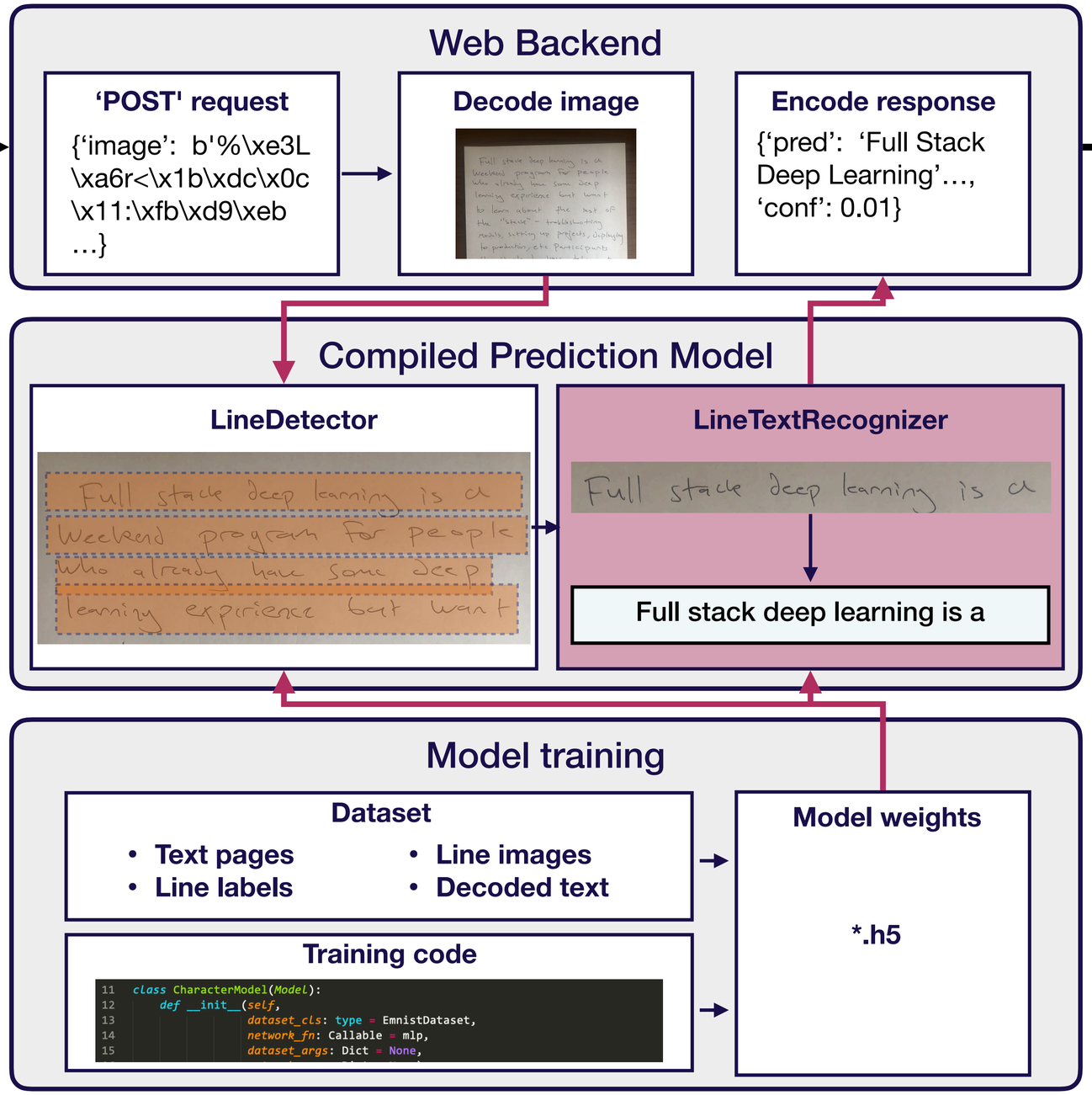
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …



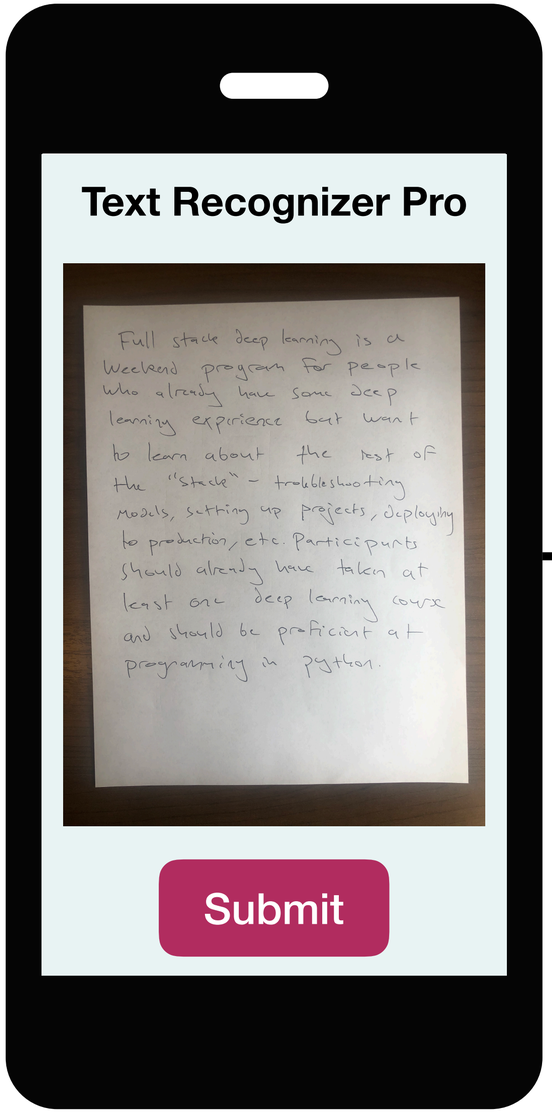


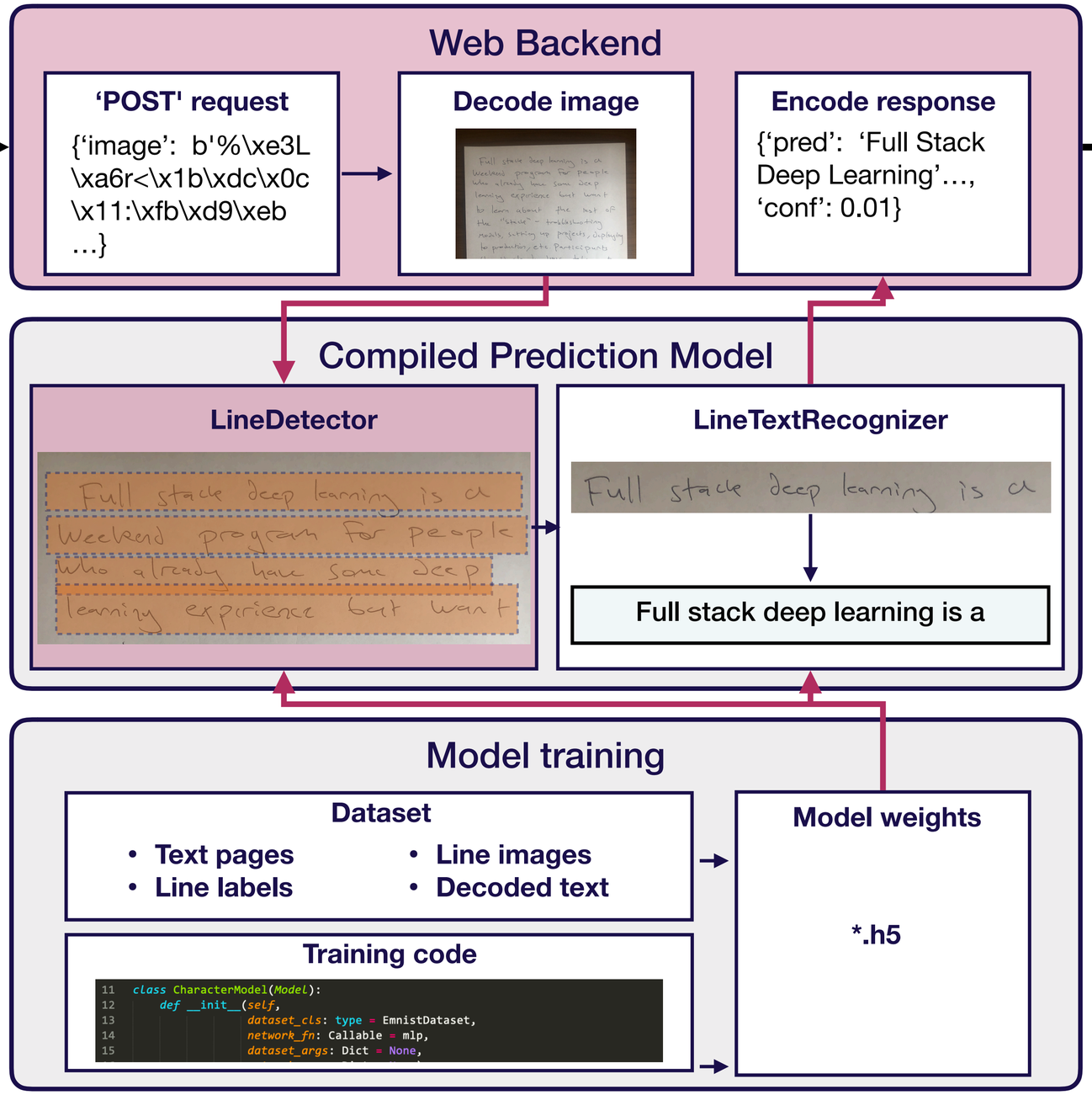
Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" …





Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" ...





Full stack deep learning is a weekend program for people who already have some deep learning experience but want to learn about the rest of the "stack" ...

Outline of the labs

- **Lab 1: Codebase walkthrough.**
- **Lab 2: Single-character prediction.** Before predicting full lines, try a simpler problem.
- **Lab 3: LineTextRecognizer.** Build a synthetic dataset. LSTM + CTC loss model.
- **Lab 4: Tools for experimentation.** Weights & Biases, parallel experiment running.
- **Lab 5. Experimentation.** Try some things and run some things overnight.
- **Lab 6. LineDetector.** Train the line detection model.
- **Lab 7. Data.** Managing data, including label and versioning.
- **Lab 8. Continuous integration.** Testing your model.
- **Lab 9. Deployment.** Put the model into production.

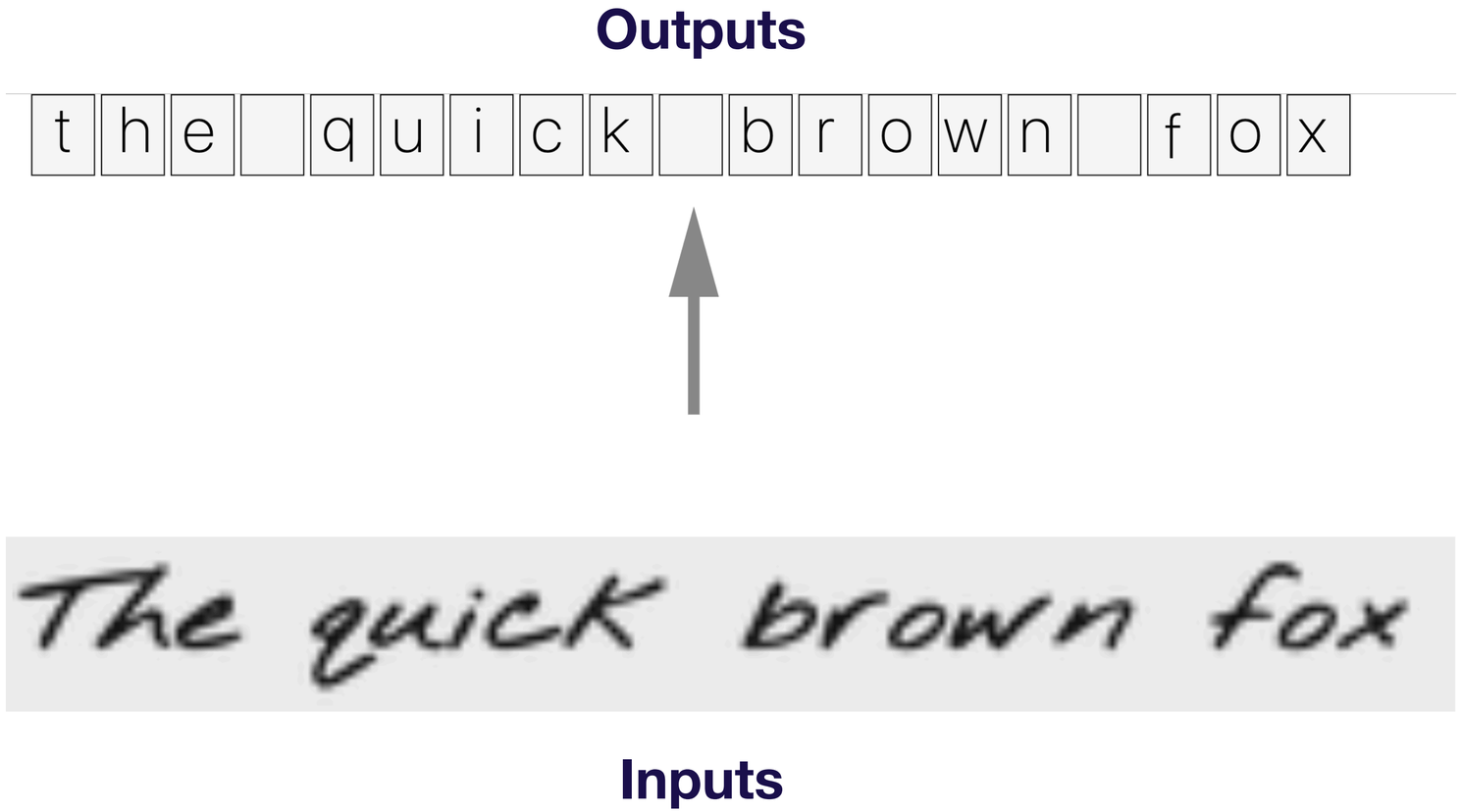
Go here

https://github.com/full-stack-deep-learning/fsdl-text-recognizer-project

W&B Access Code

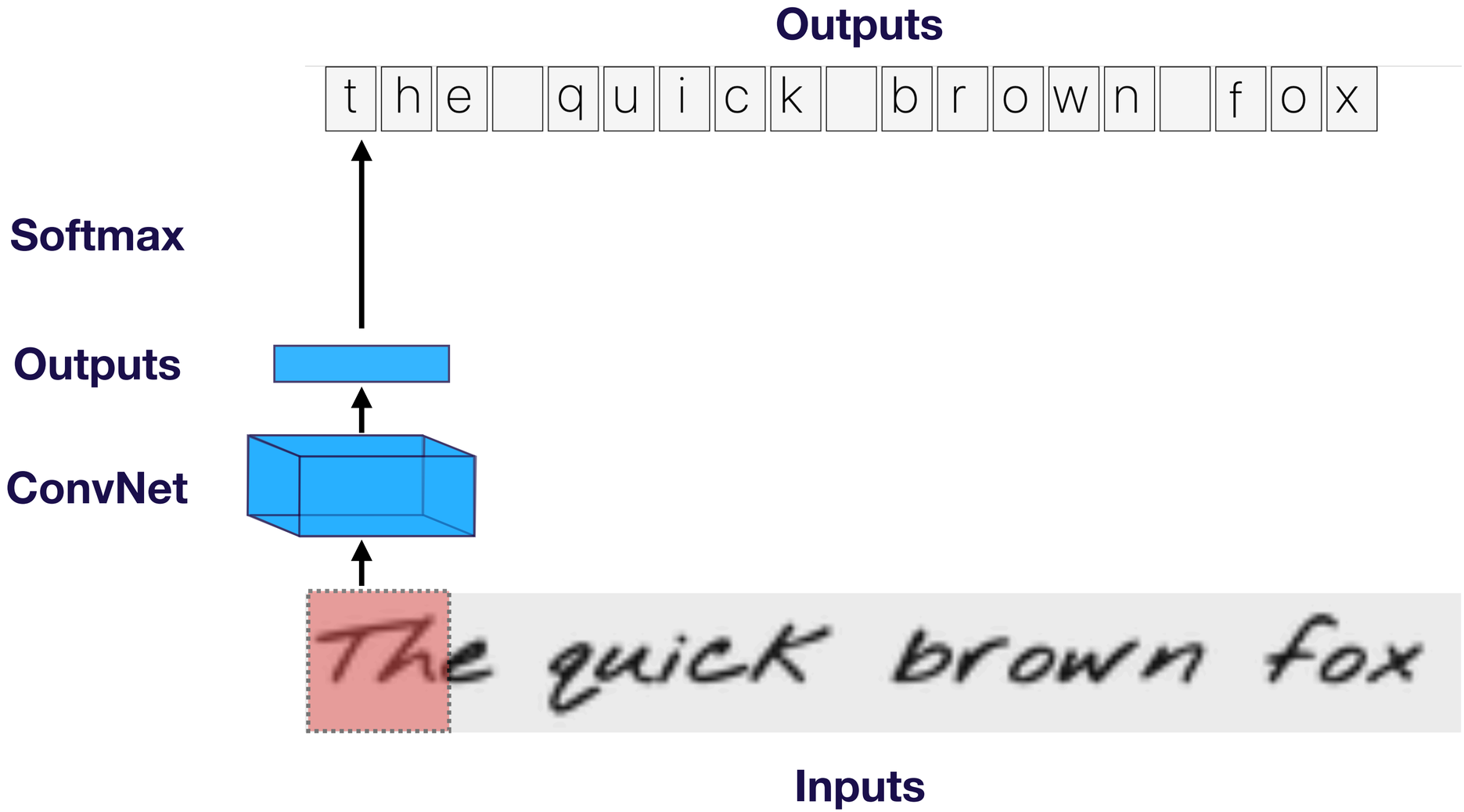
fsdl-deep-cajon

LineTextRecognizer Model architecture

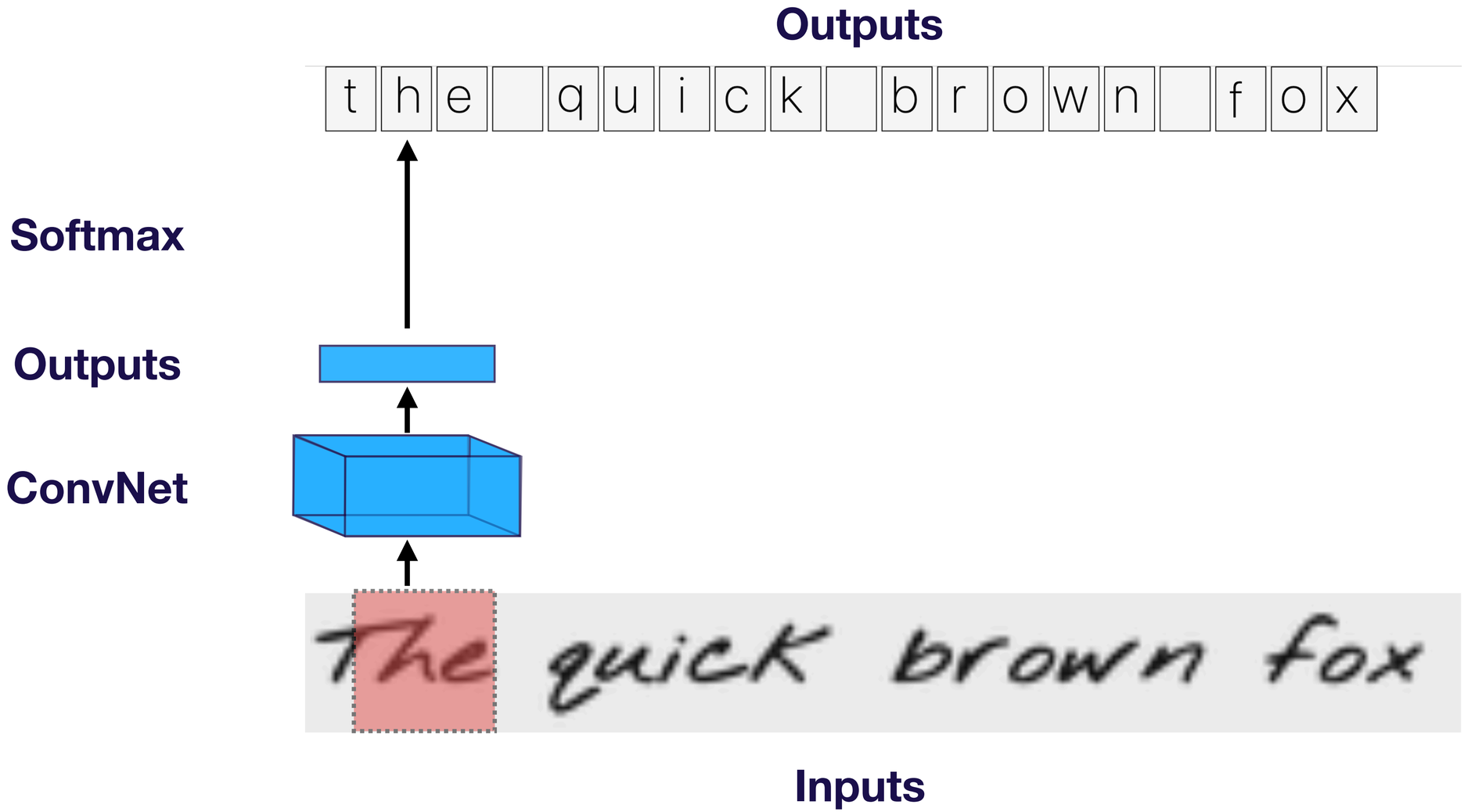


Sequence Modeling with CTC, Awni Hannun, Distill 10.23915/distill.00008

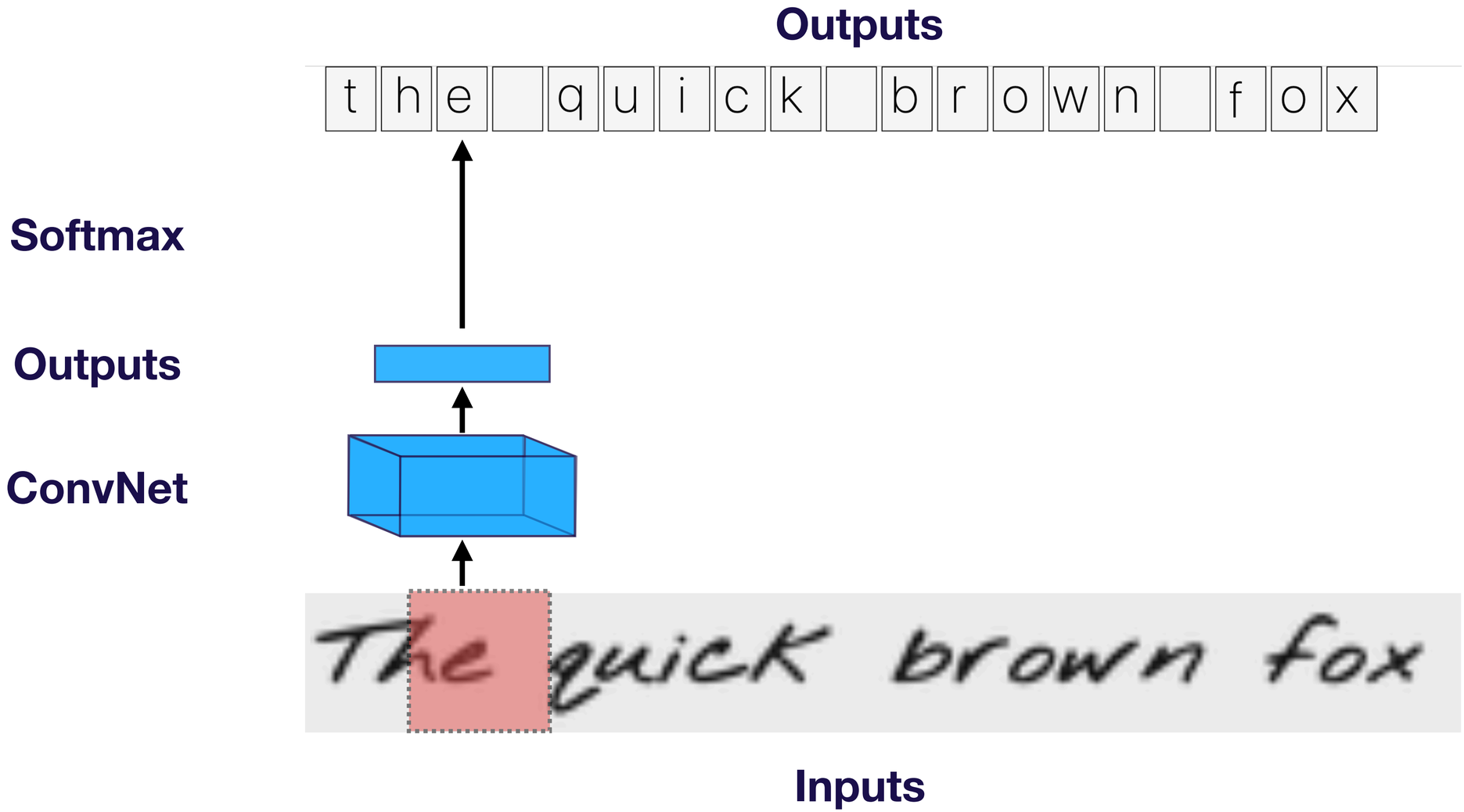
A baseline model architecture



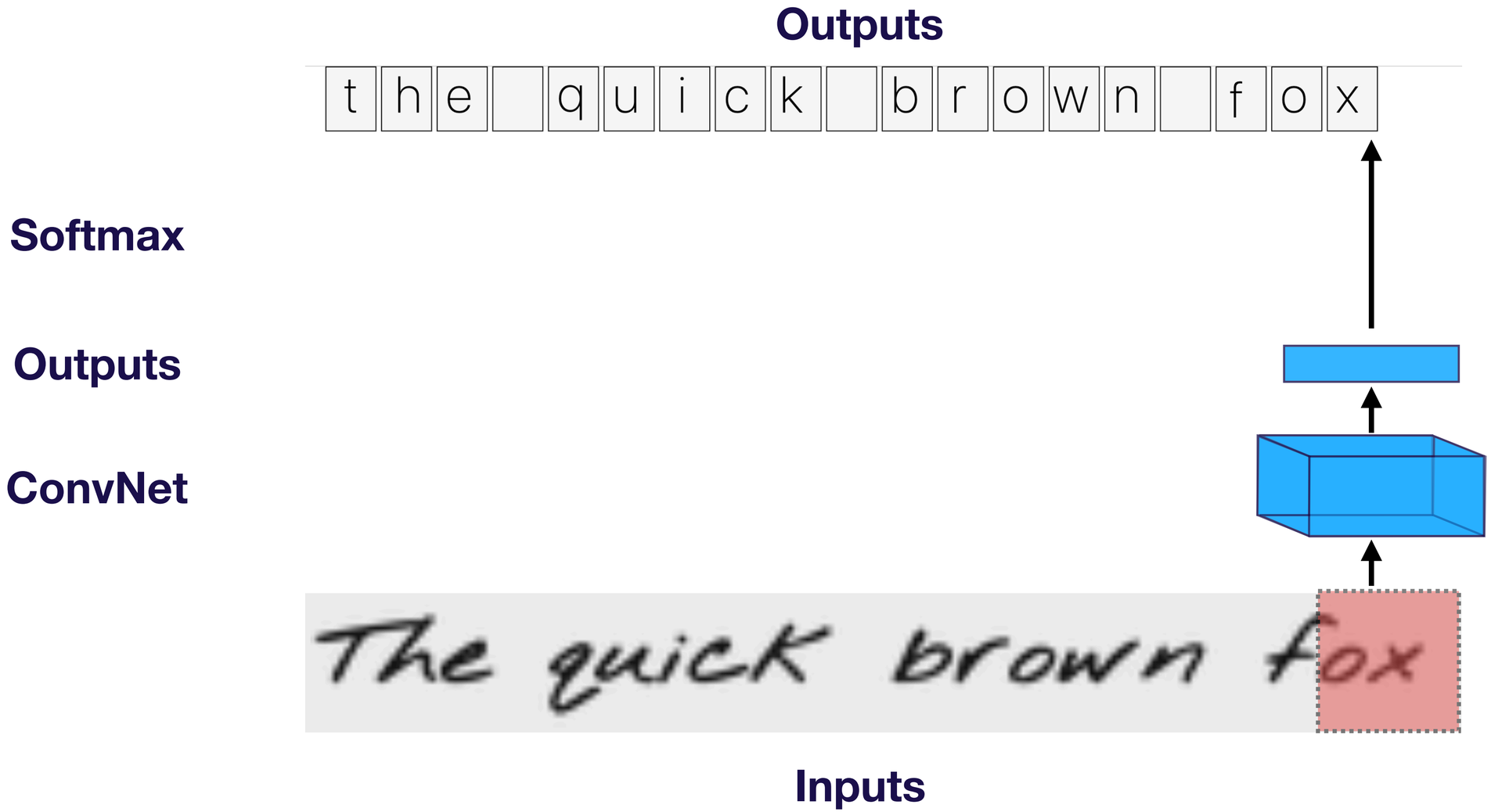
A baseline model architecture



A baseline model architecture



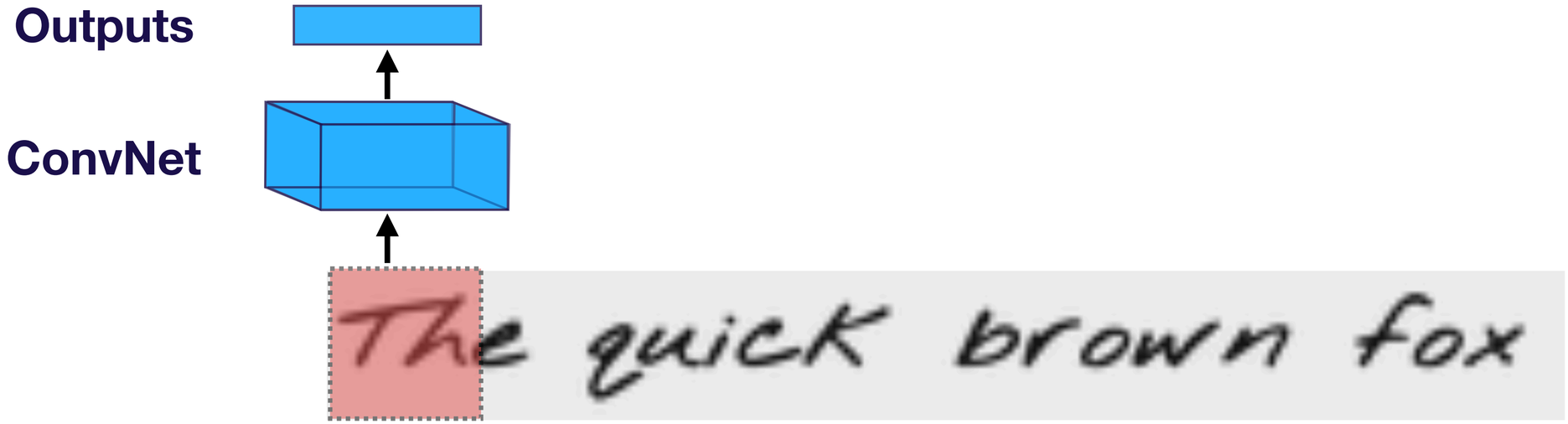
A baseline model architecture



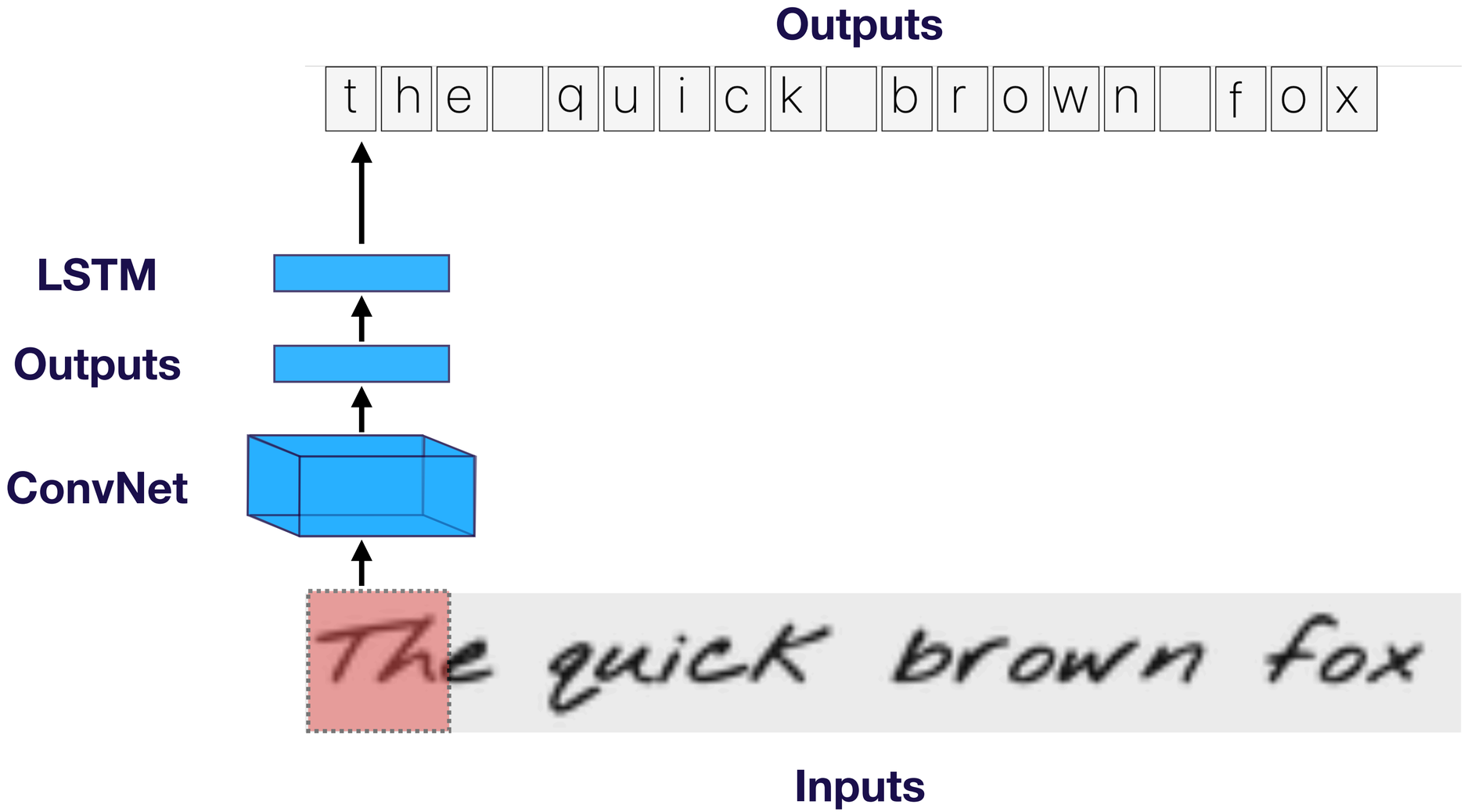
Let's give the model some context

Outputs

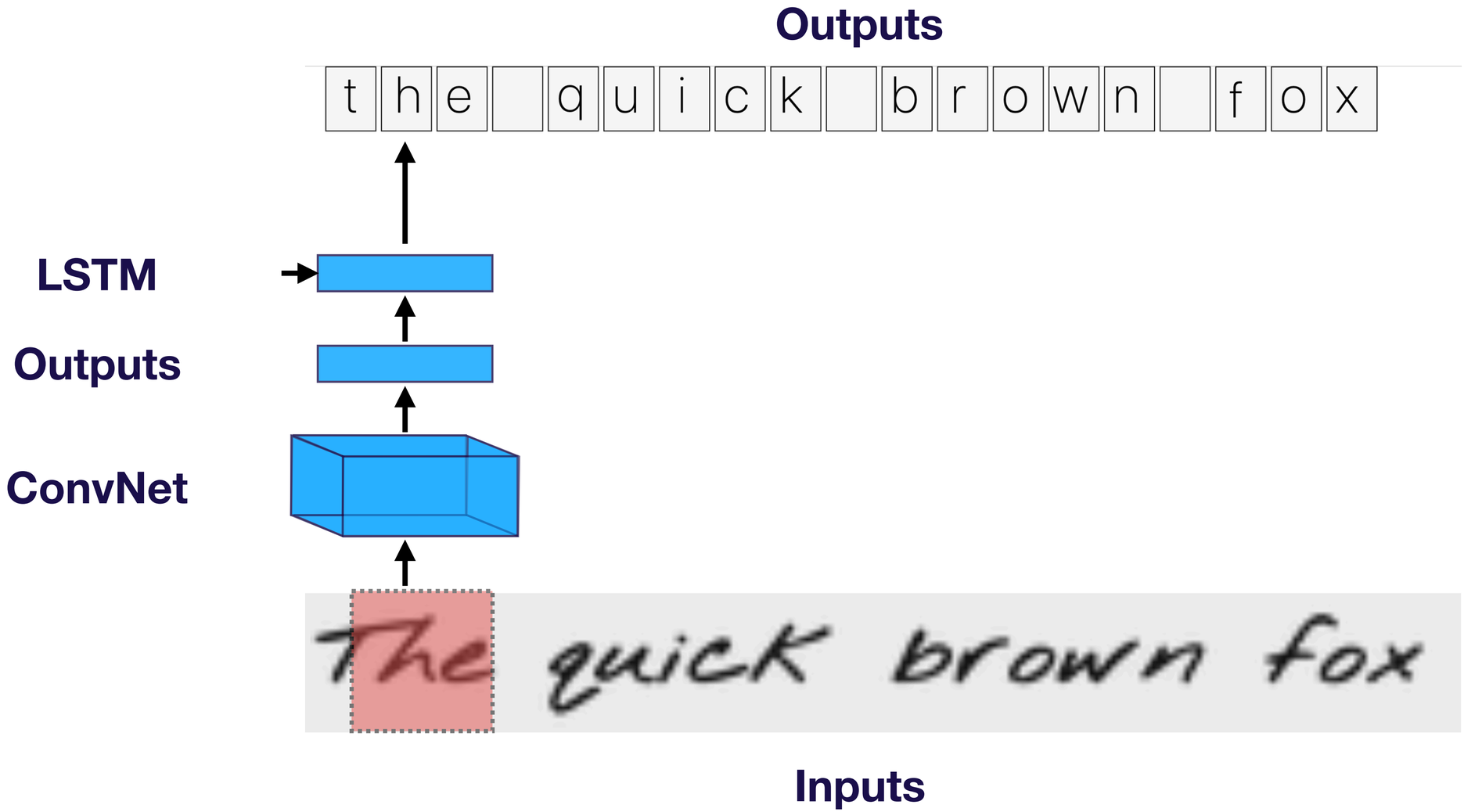
t h e _ q u i c k _ b r o w n _ f o x



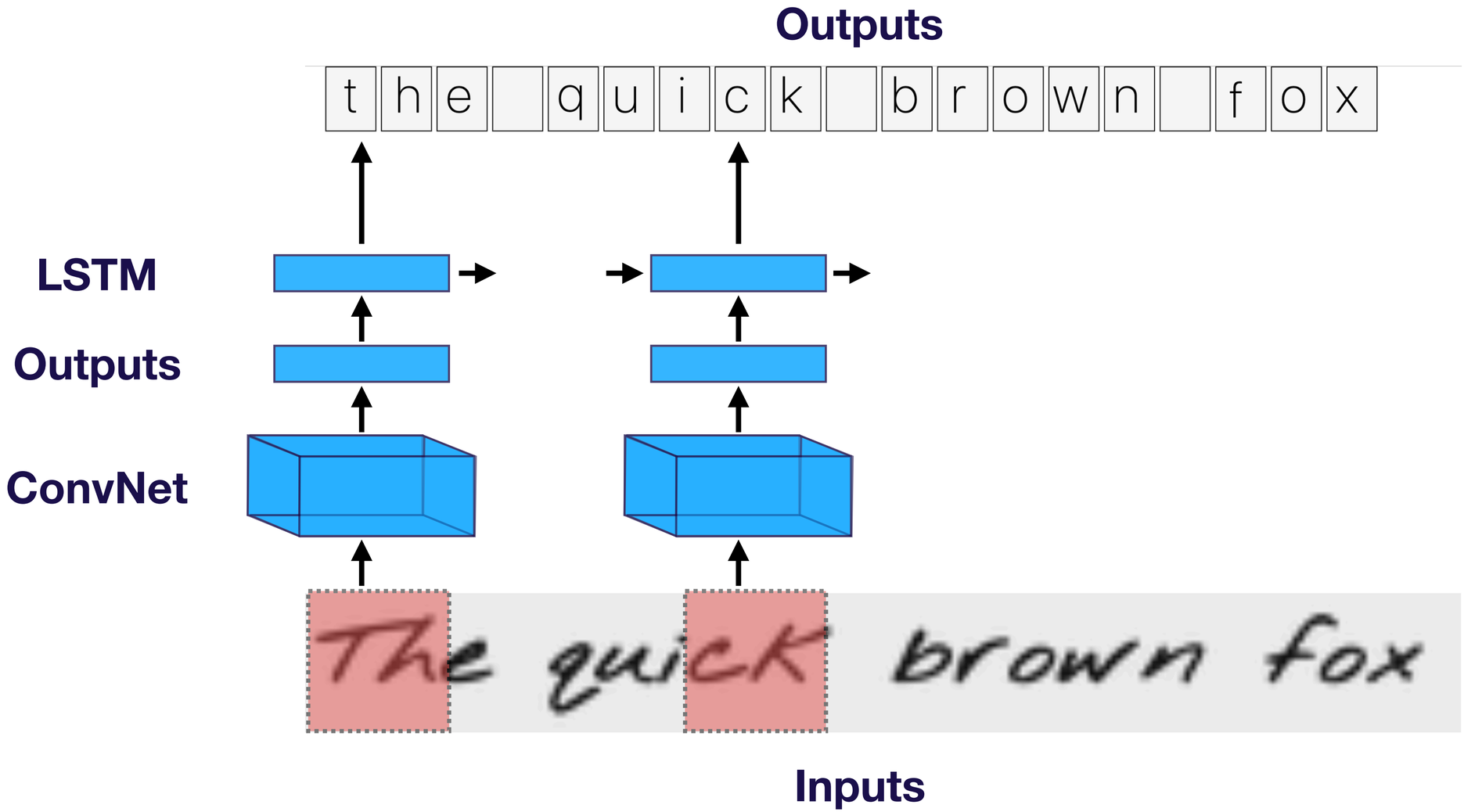
Let's give the model some context



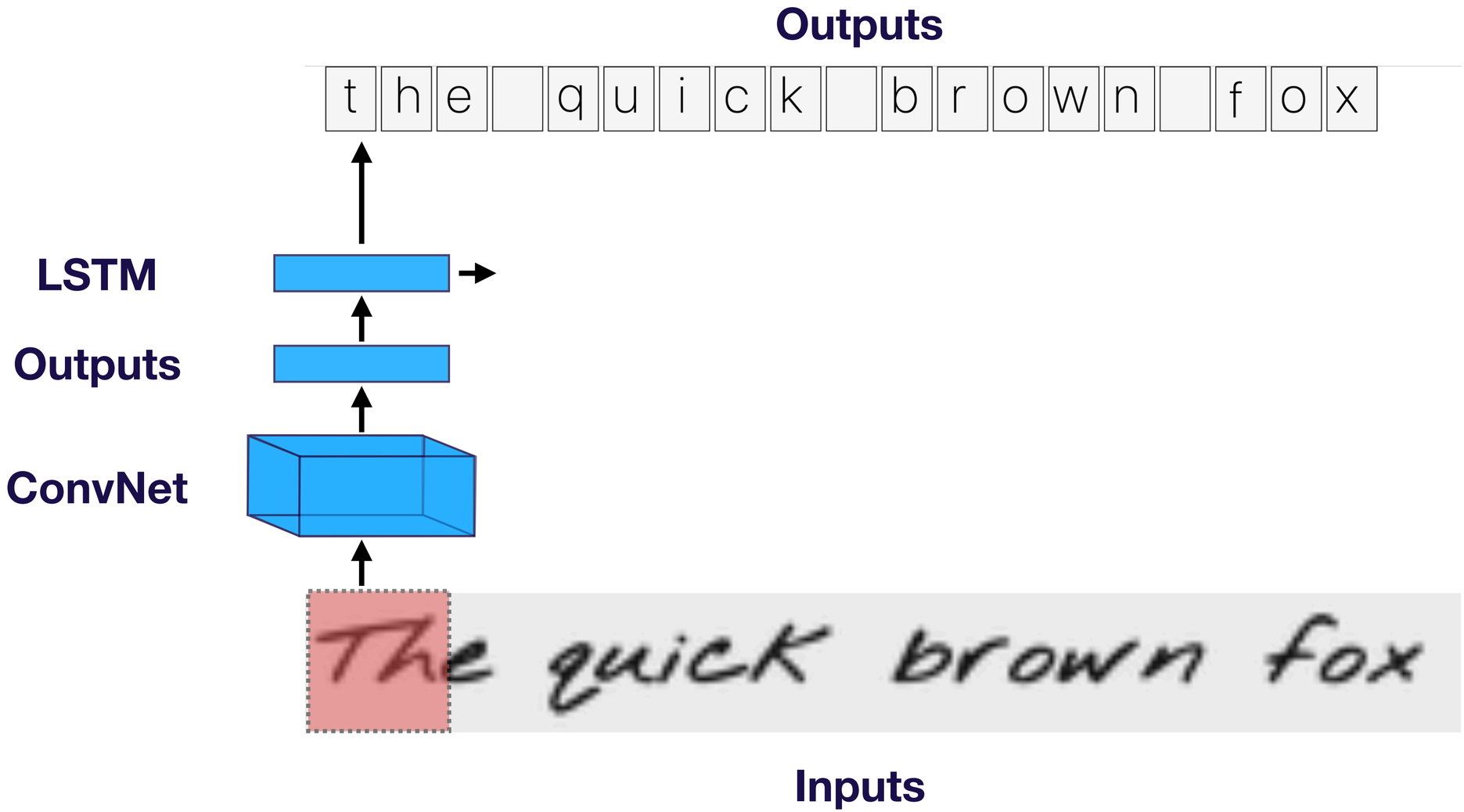
Let's give the model some context



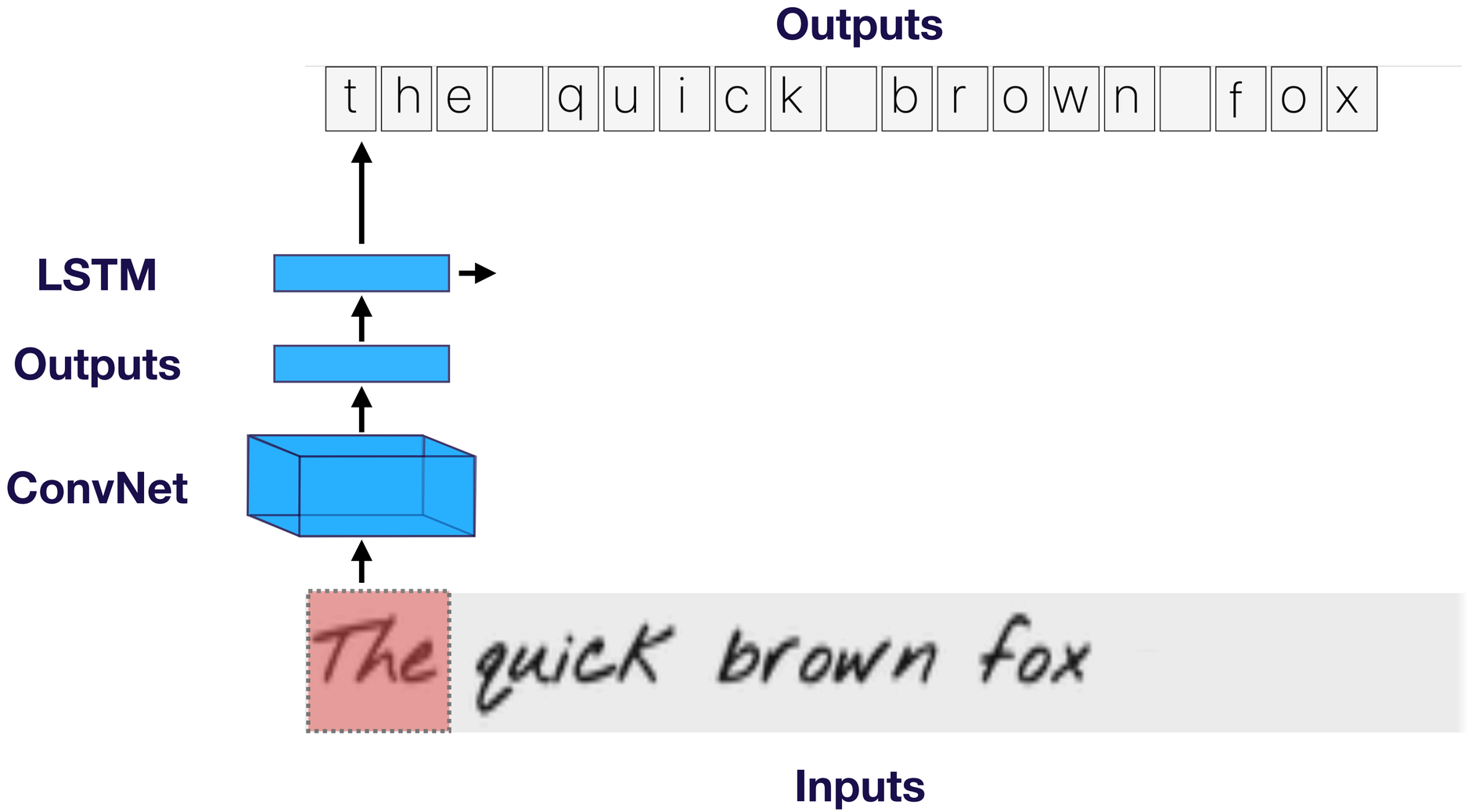
Let's give the model some context



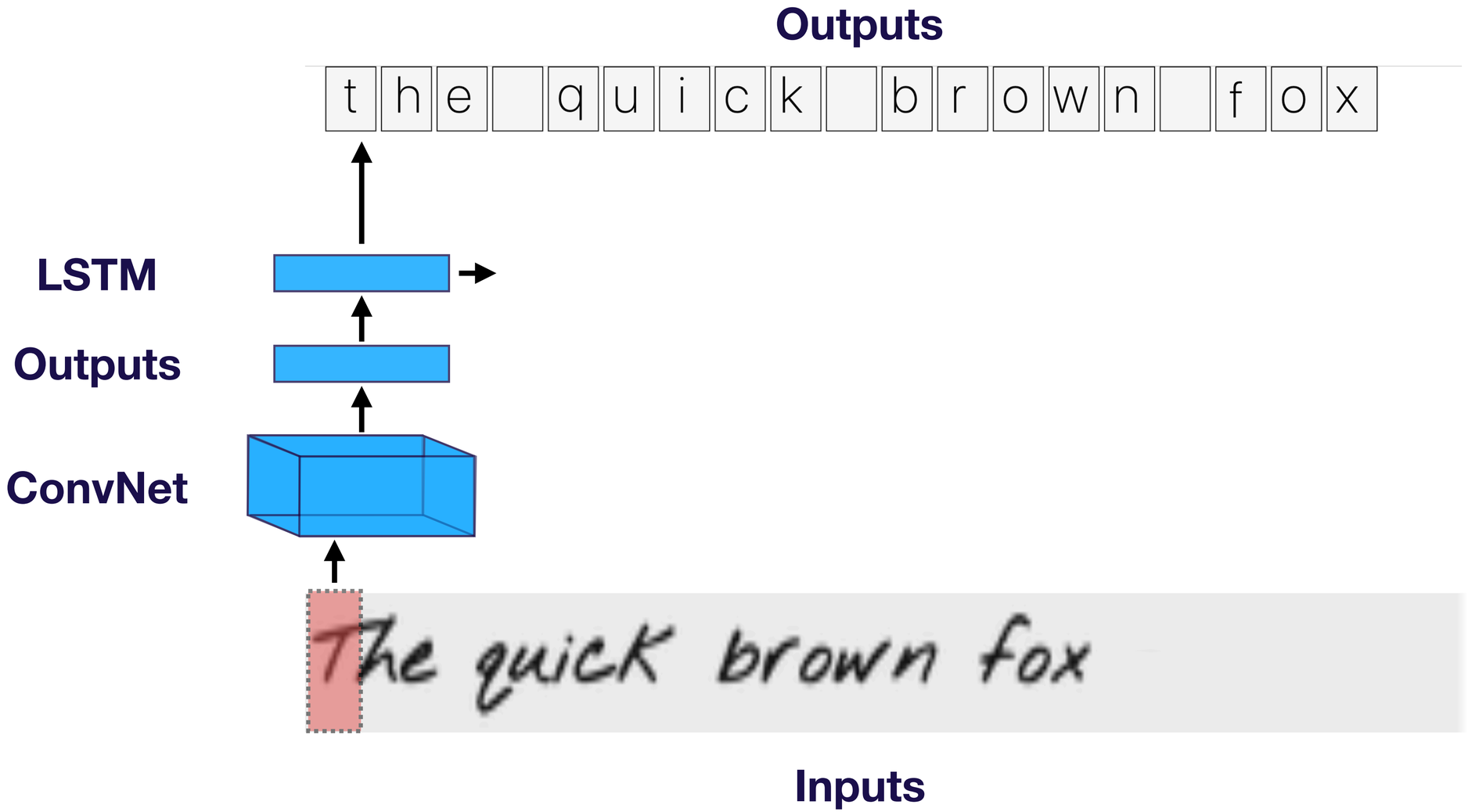
Challenge: different line spacing



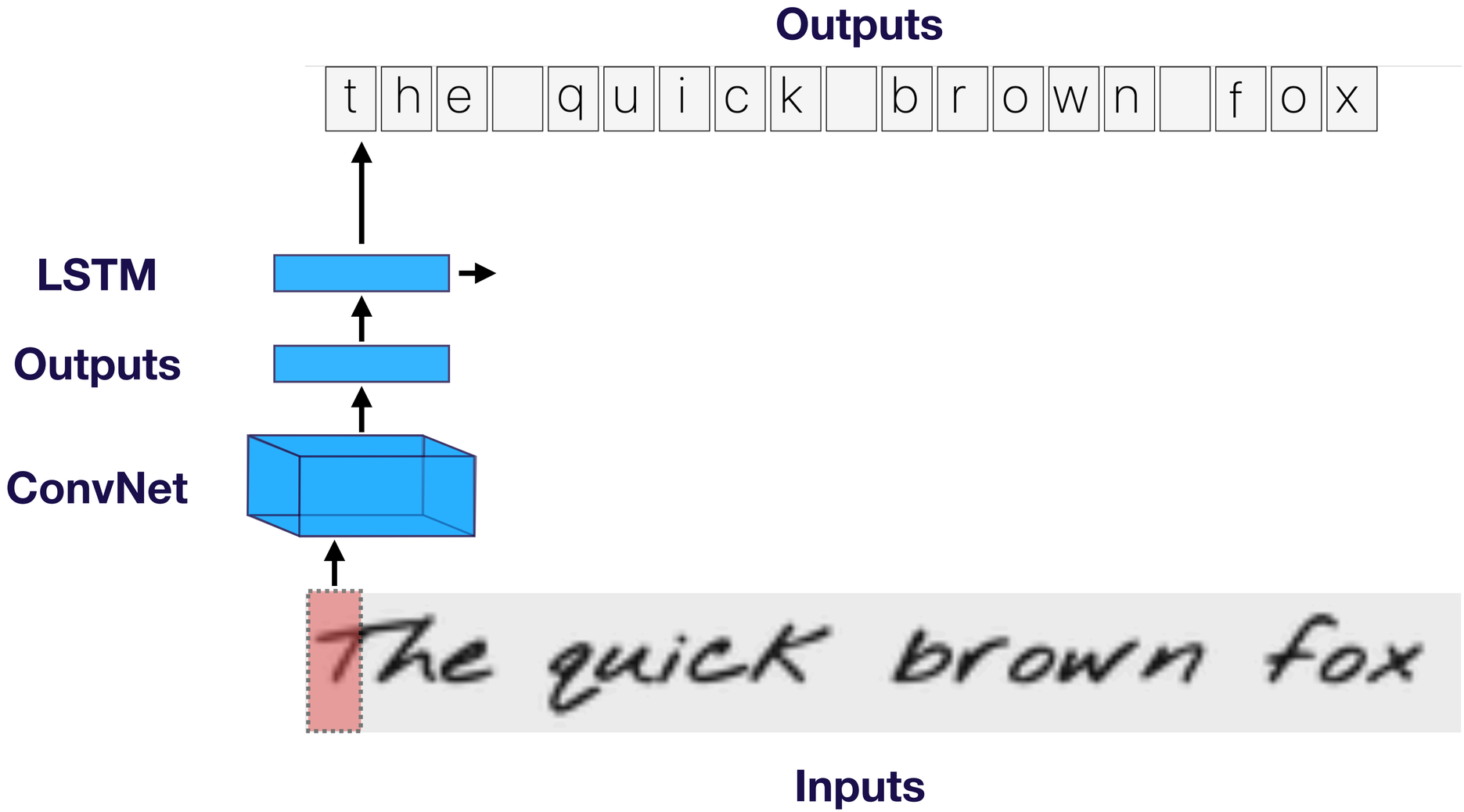
Challenge: different line spacing



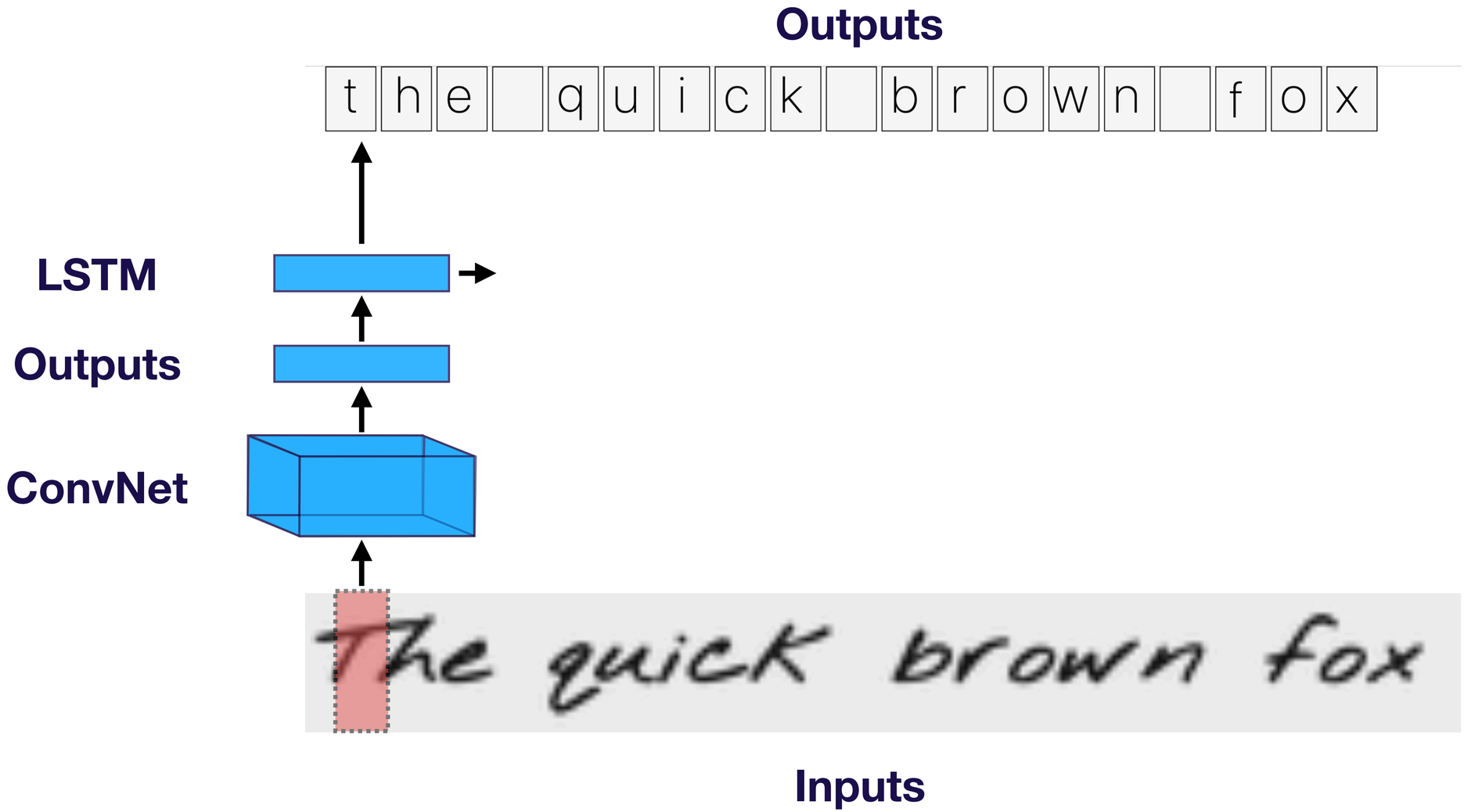
Challenge: different line spacing



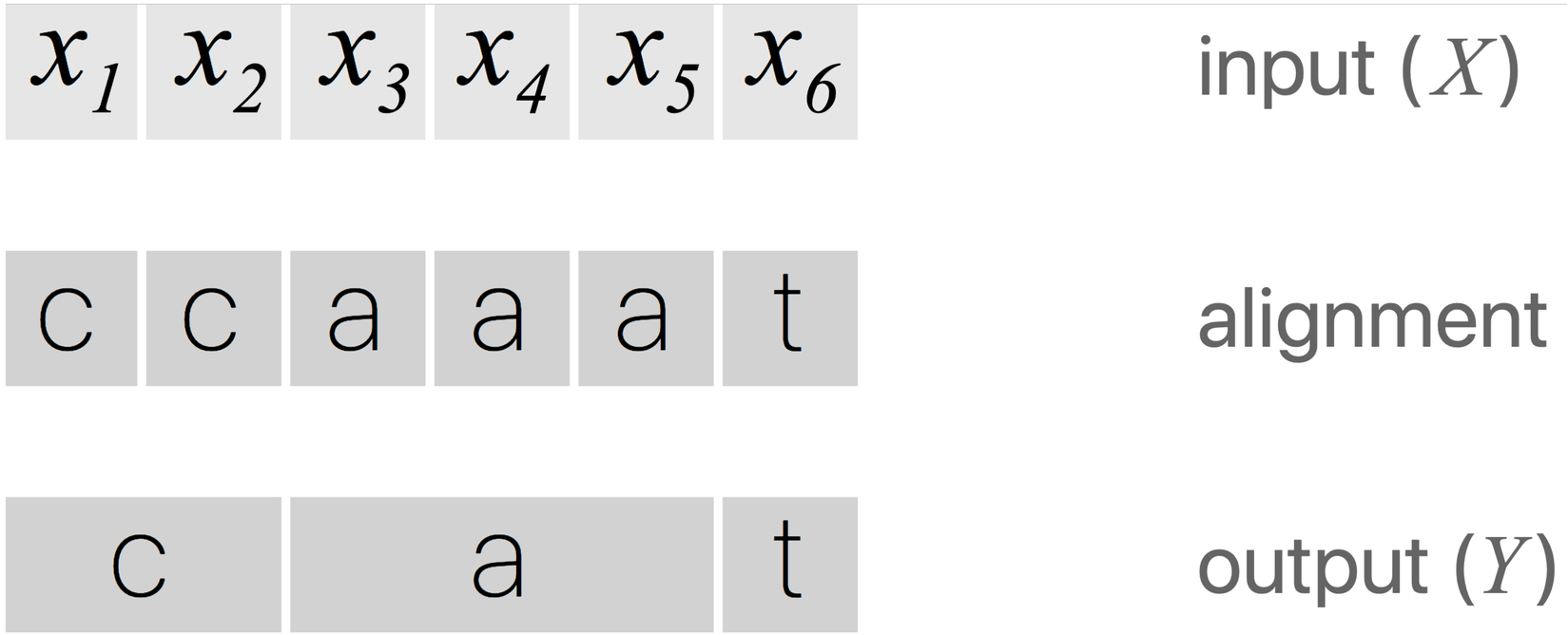
Challenge: different line spacing



Challenge: different line spacing

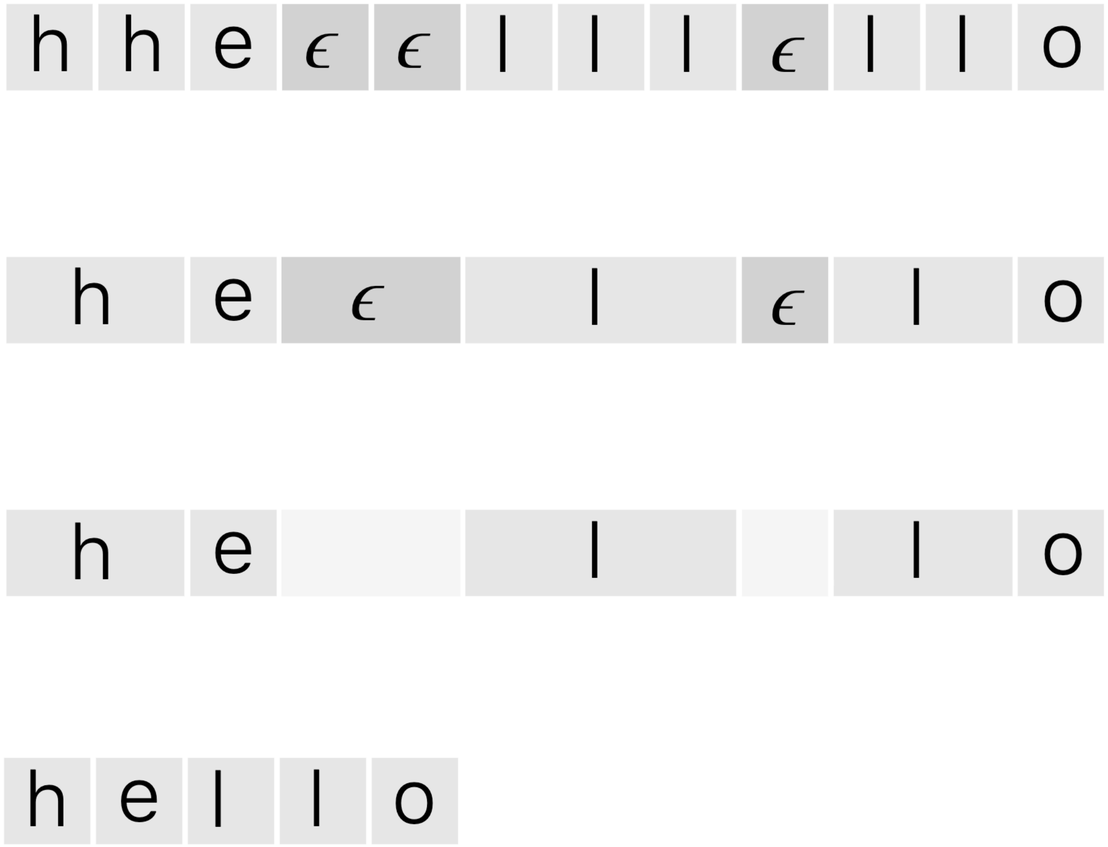


Intuition for the solution



Sequence Modeling with CTC, Awni Hannun, Distill 10.23915/distill.00008

CTC Loss



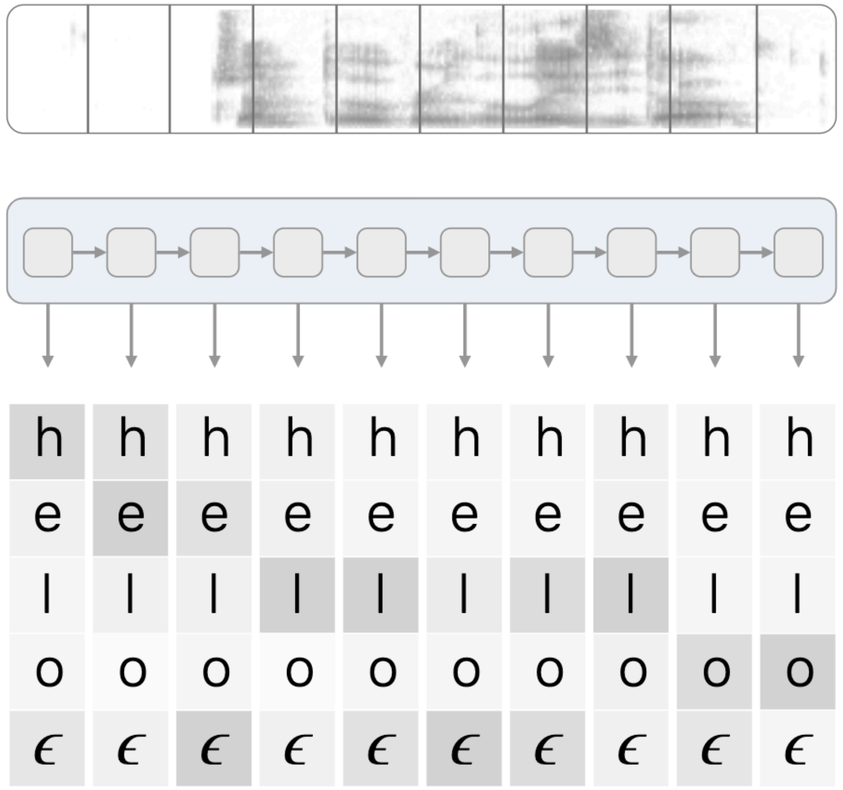
First, merge repeat characters.

Then, remove any ϵ tokens.

The remaining characters are the output.

Sequence Modeling with CTC, Awni Hannun, Distill 10.23915/distill.00008

Training with the CTC loss



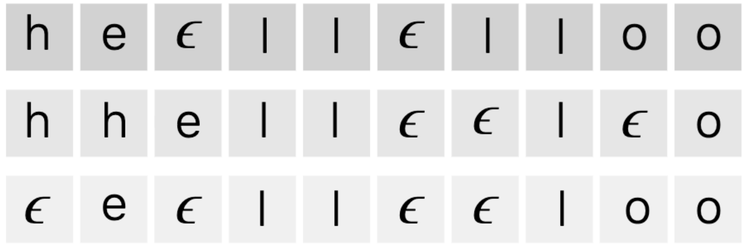
We start with an input sequence, like a spectrogram of audio.

The input is fed into an RNN, for example.

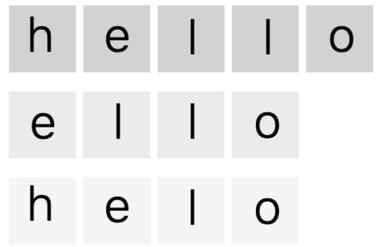
The network gives $p_t(a \mid X)$, a distribution over the outputs {h, e, l, o, ϵ} for each input step.

Sequence Modeling with CTC, Awni Hannun, Distill 10.23915/distill.00008

Training with the CTC loss



With the per time-step output distribution, we compute the probability of different sequences



By marginalizing over alignments, we get a distribution over outputs

Sequence Modeling with CTC, Awni Hannun, Distill 10.23915/distill.00008

For more of the technical details

Sequence Modeling with CTC, Awni Hannun, Distill 10.23915/distill.00008