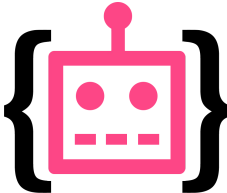


Review from  deepsystems.ai

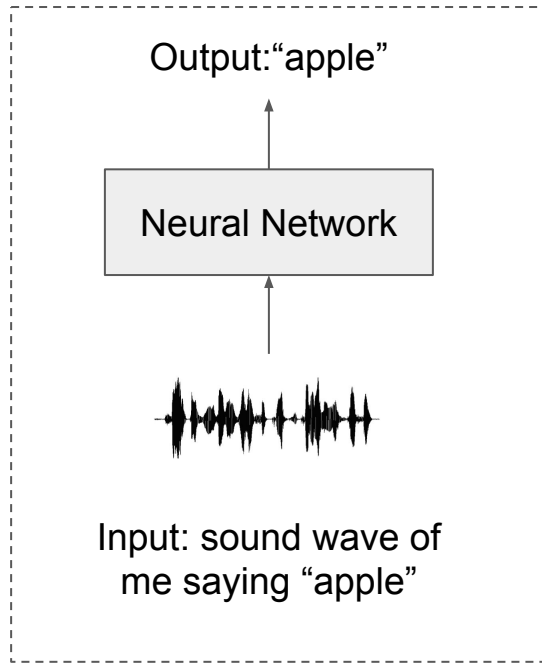
How to build end-to-end recognition system: best practices



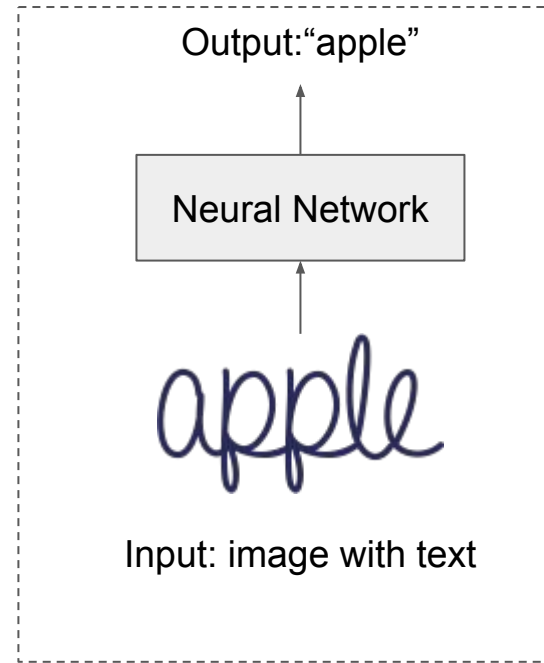
apple → "apple"

Task Description

Labelling unsegmented sequence data. i.e. training data is not pre-segmented.



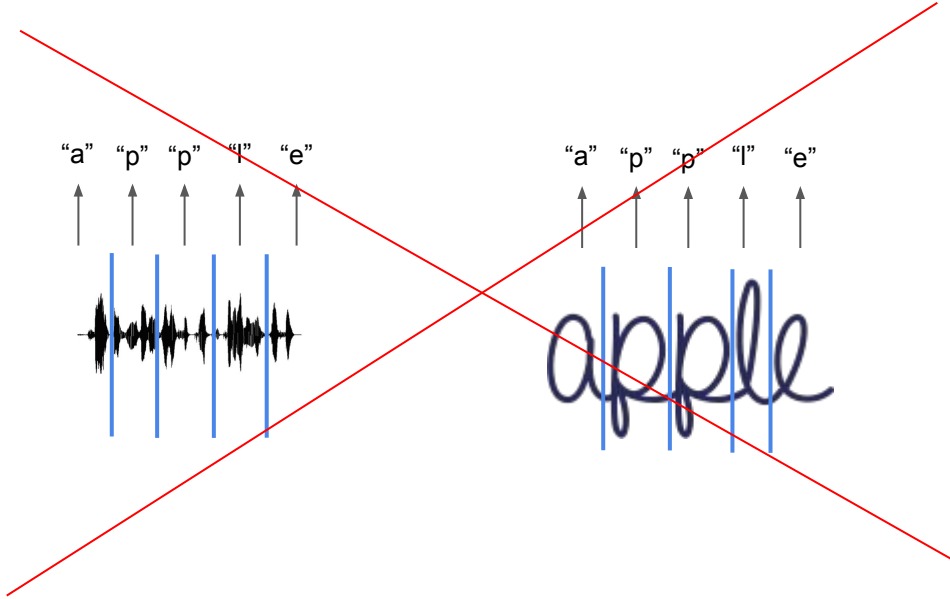
Example: speech recognition



Example: image ocr

Task Description

Labelling unsegmented sequence data. i.e. training data is not pre-segmented.



We can not pre-segment input data because:

- It is too time consuming
- It is too expensive
- It is impossible in most cases

Links

Alex Graves. CTC Loss: http://www.cs.toronto.edu/~graves/icml_2006.pdf

Keras example: [image_ocr.ipynb](#)

Big picture



Google: voice search



Baidu: Deep Speech



Dropbox: document scanner

Image OCR: model architecture

High-level overview

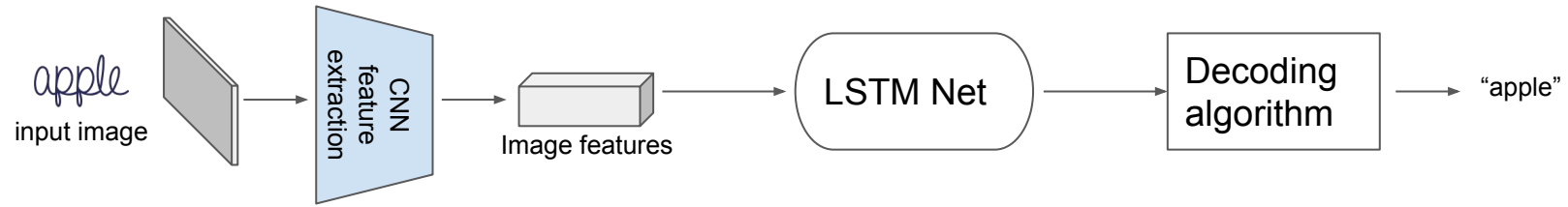
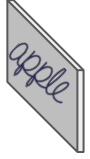


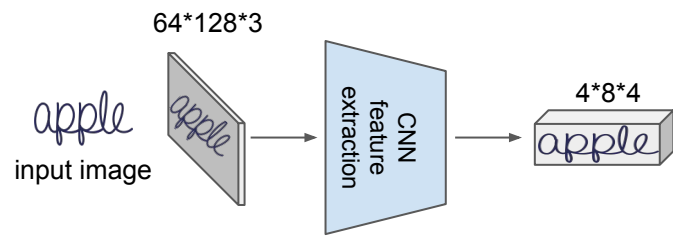
Image OCR: model architecture

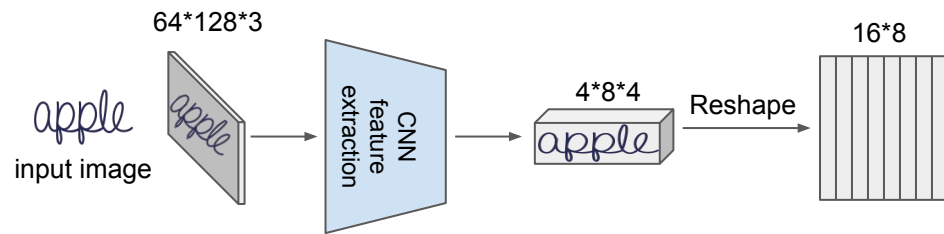
Detailed overview

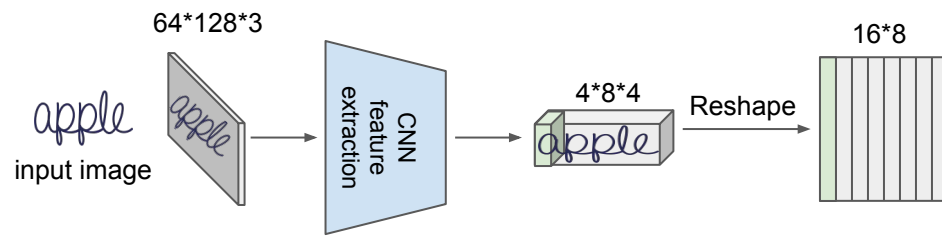
64*128*3

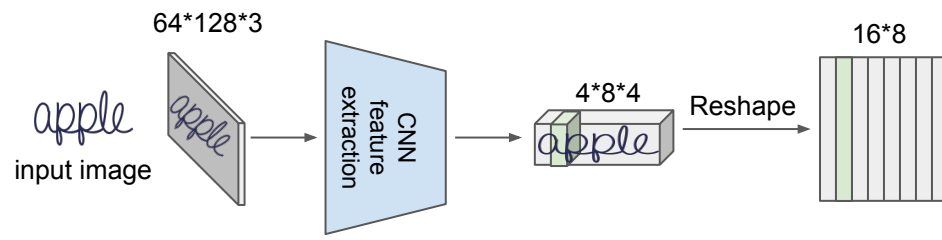
apple
input image

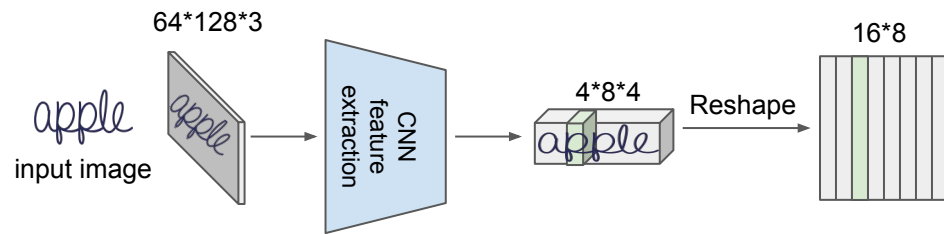


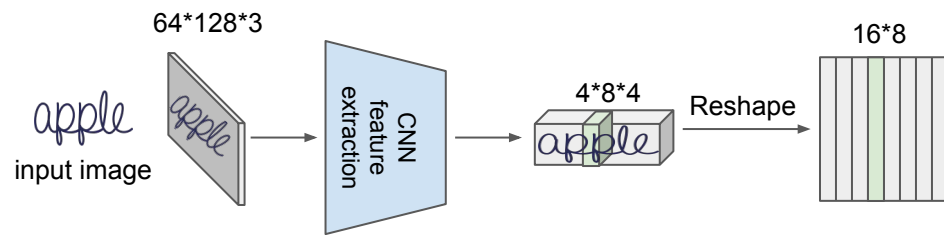


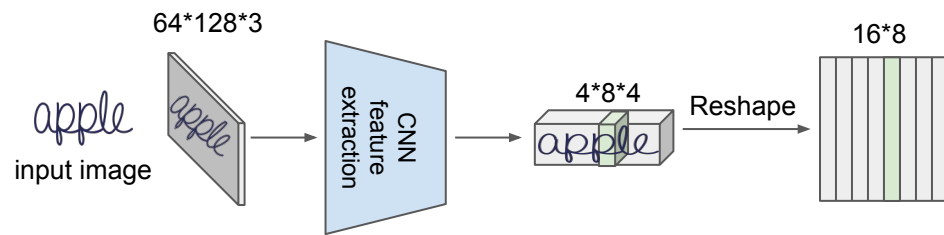


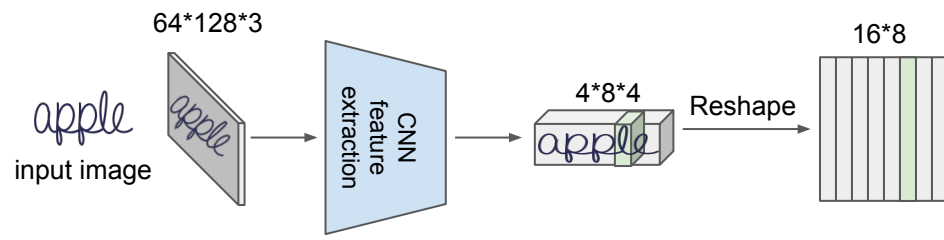


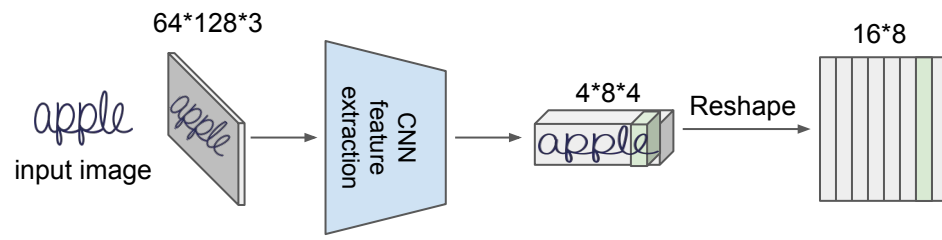


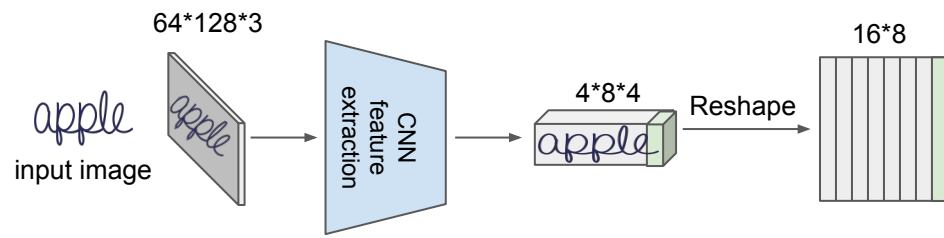


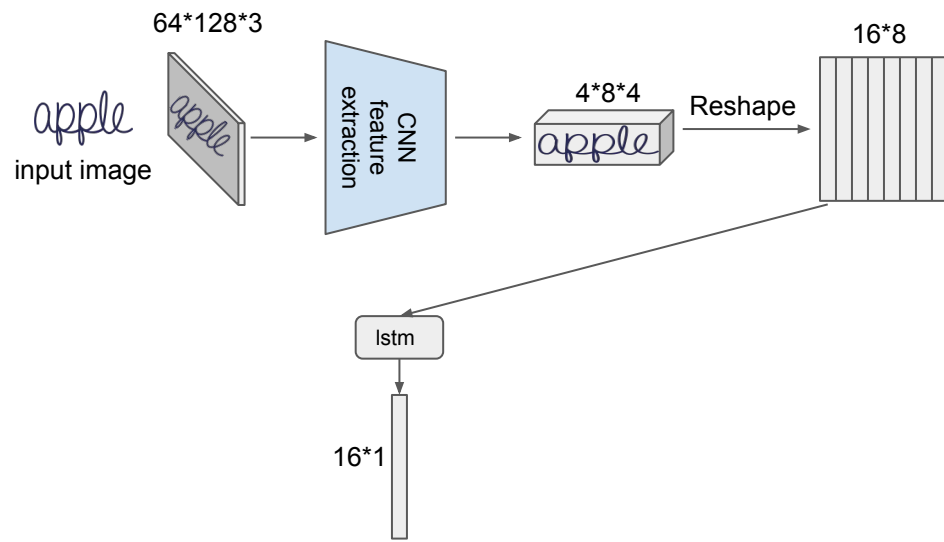


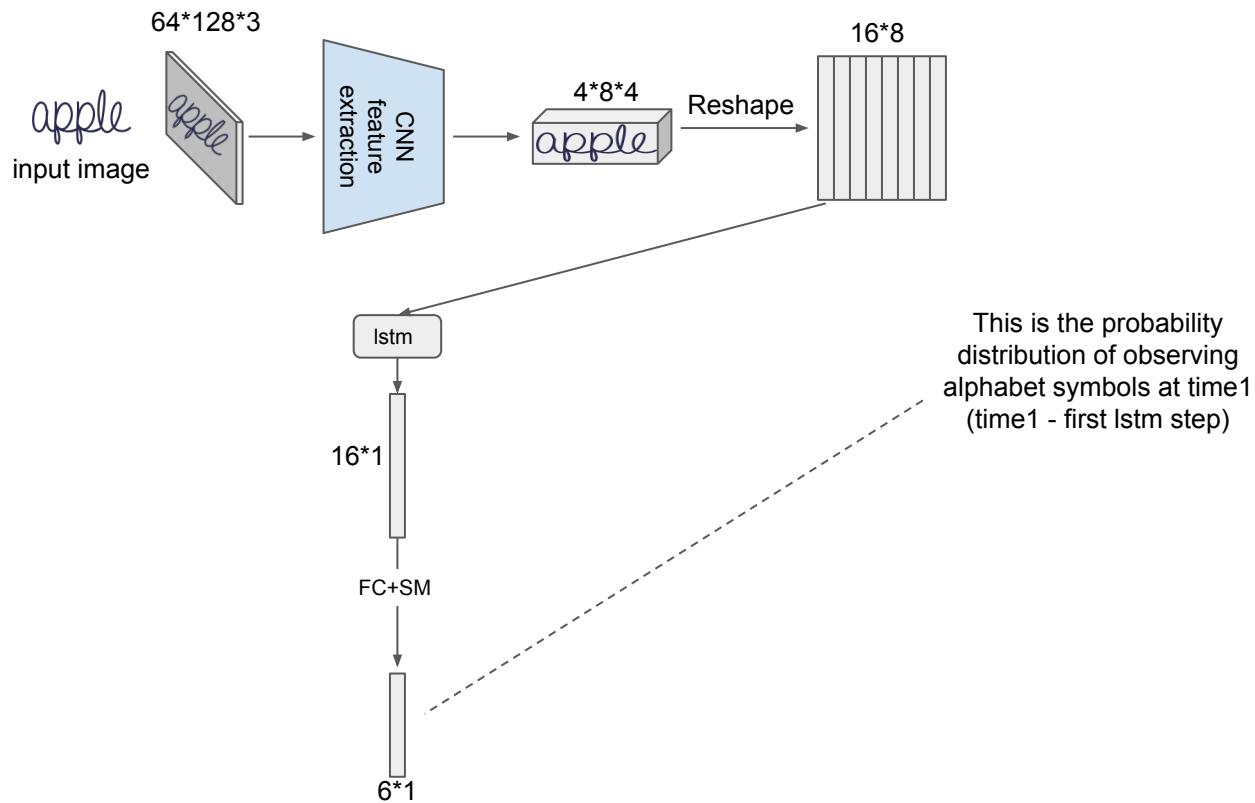


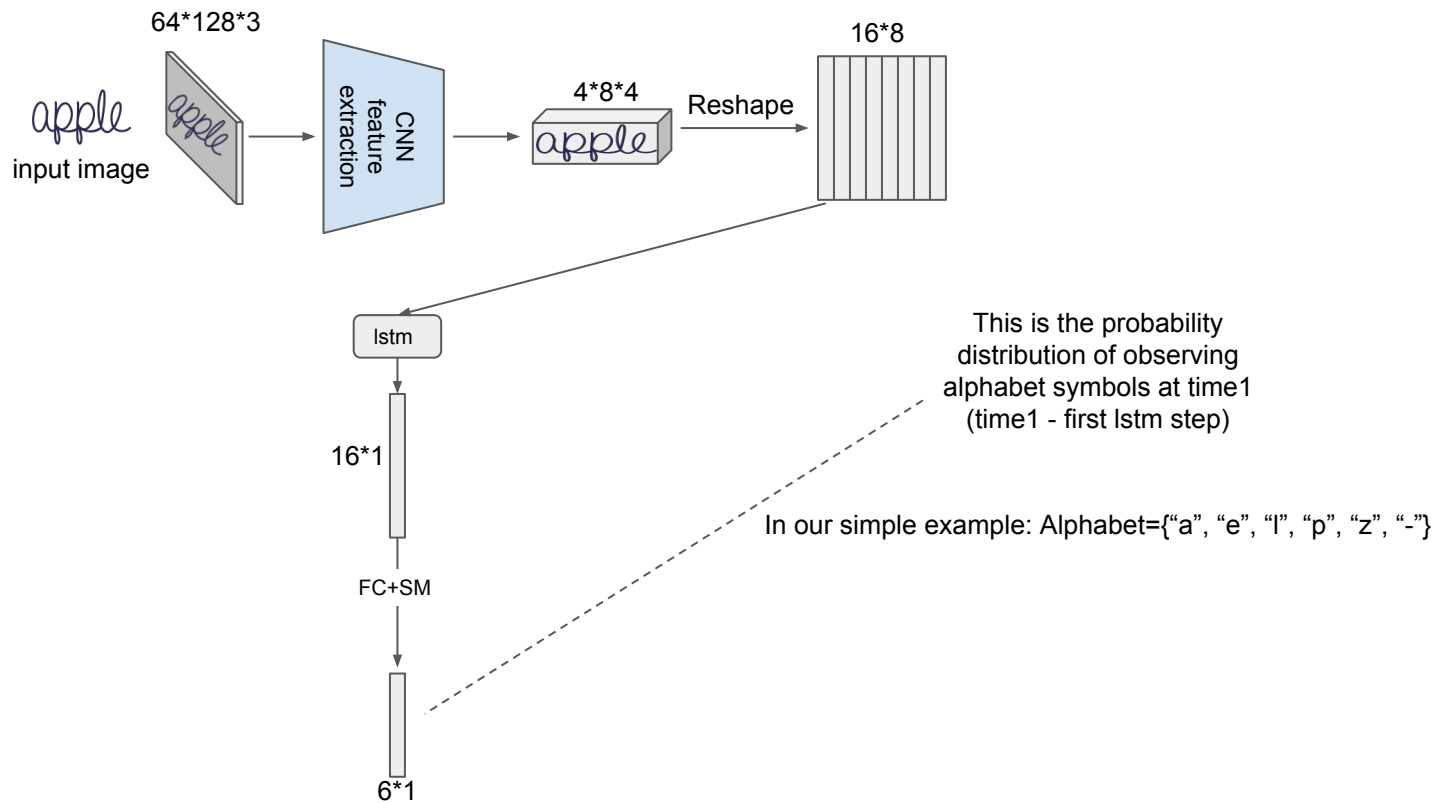


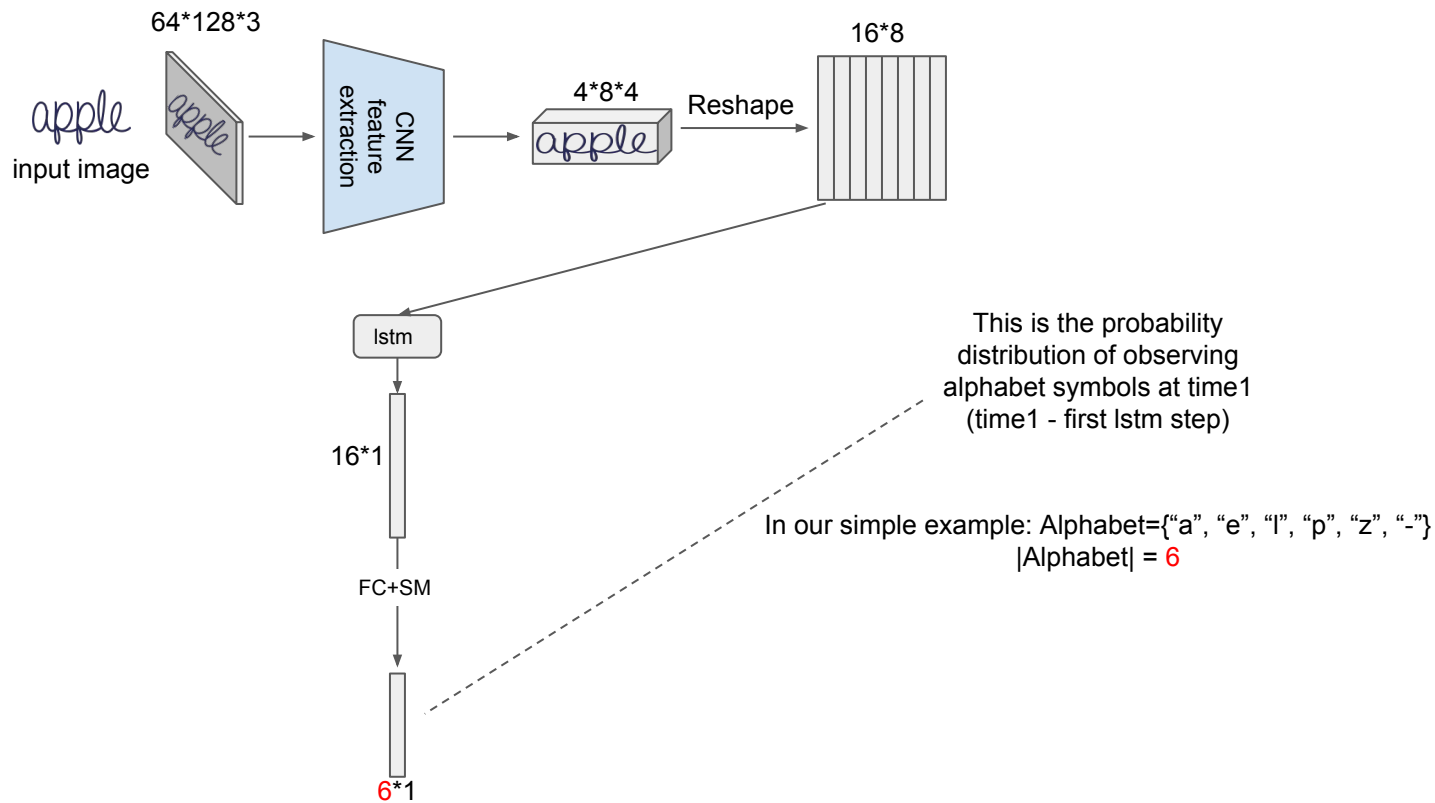


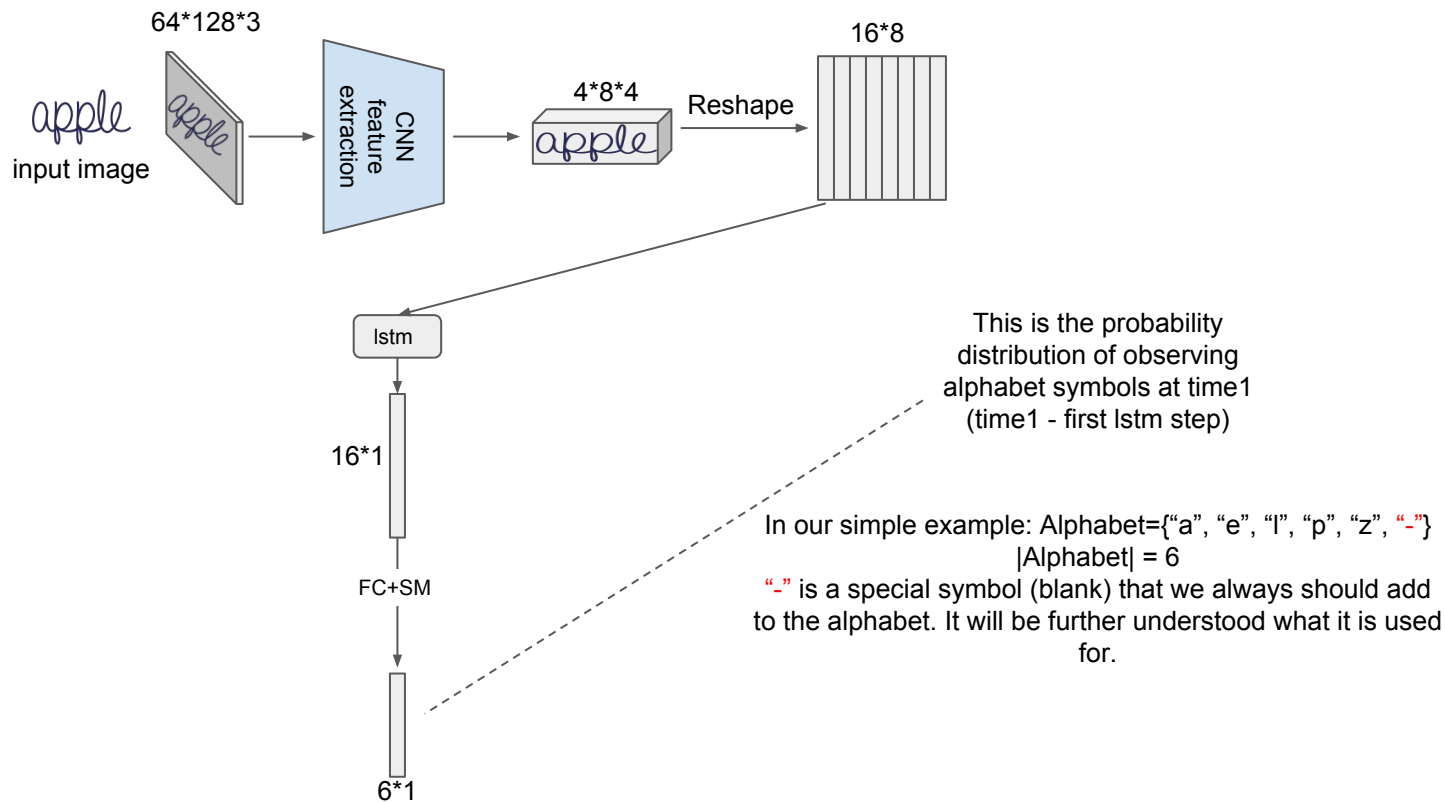


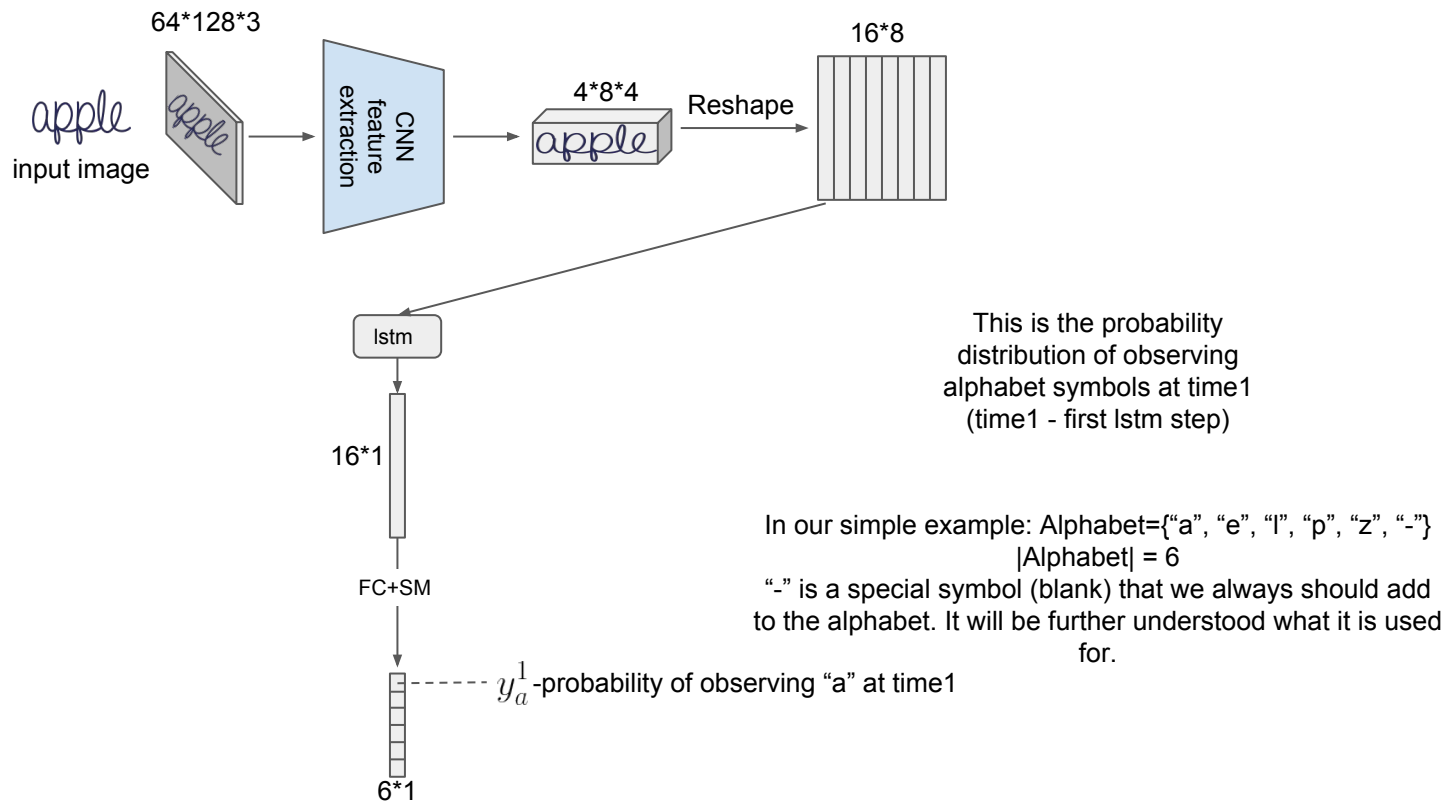


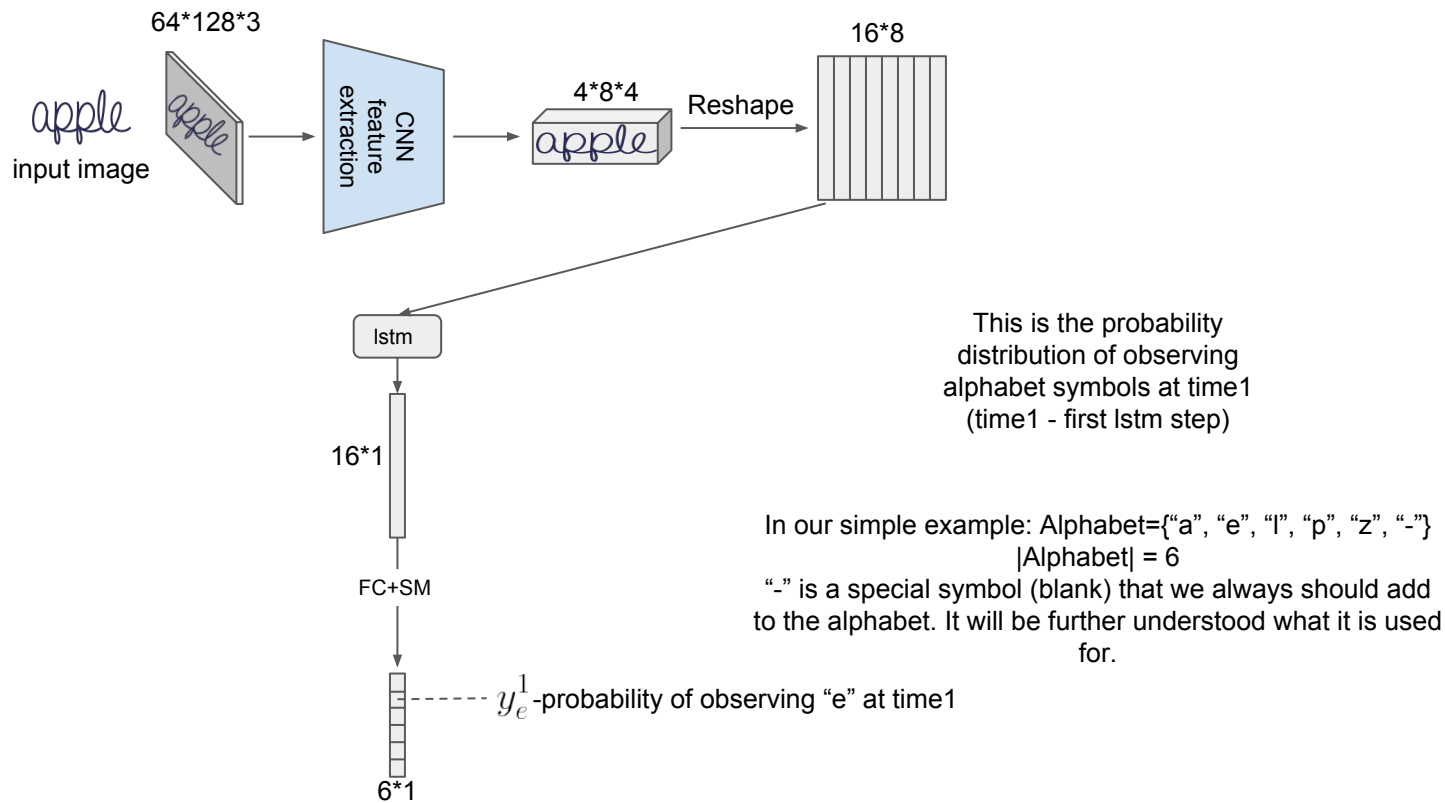


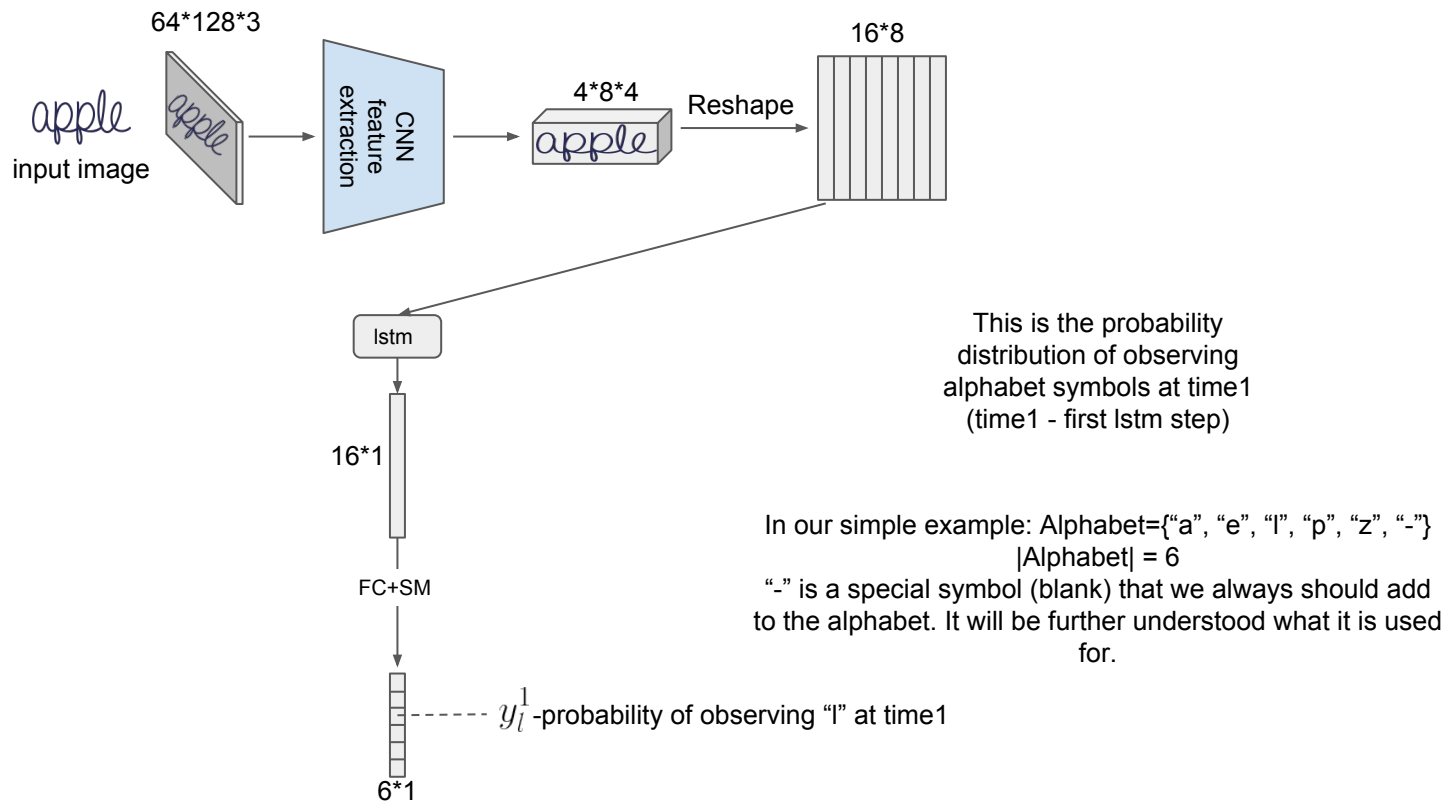


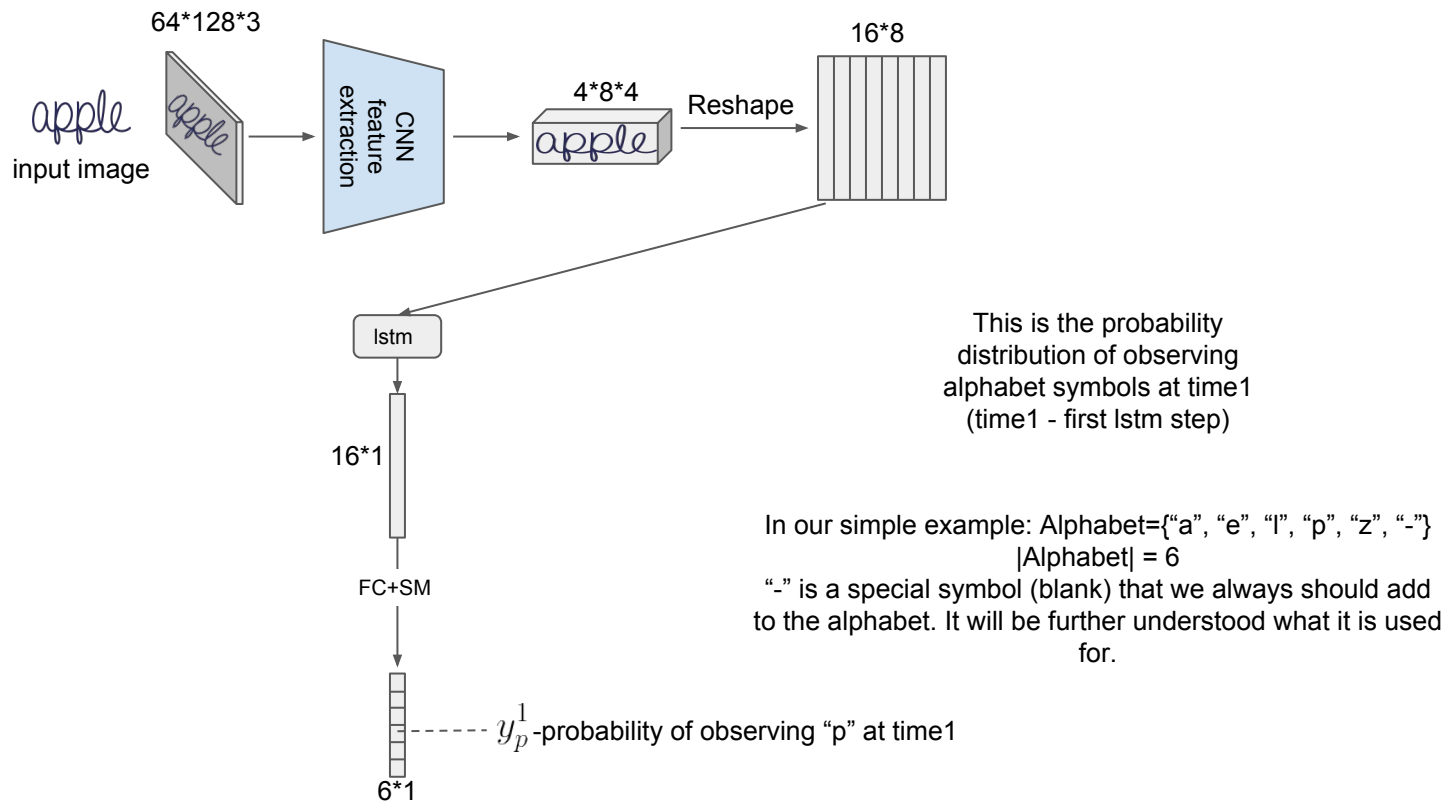


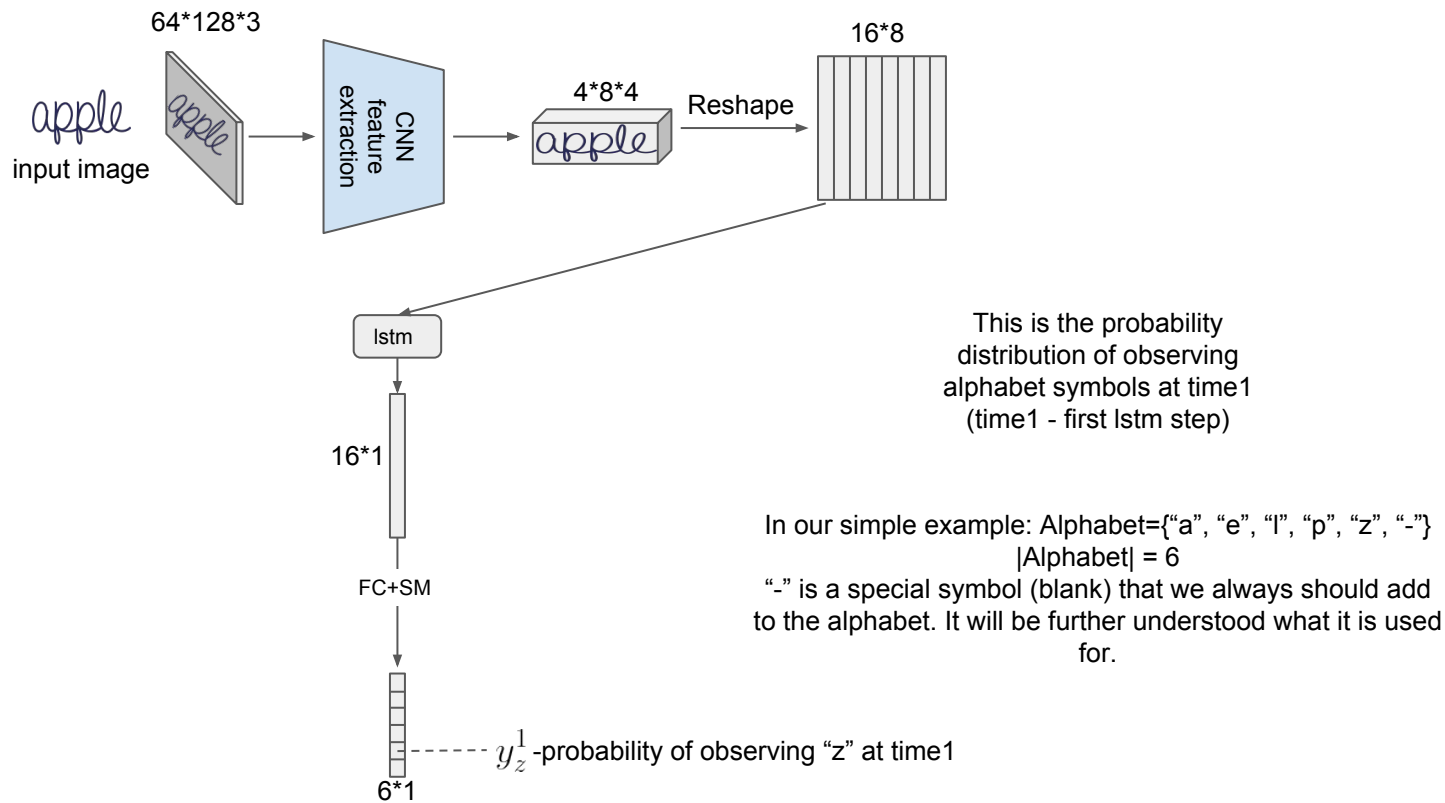


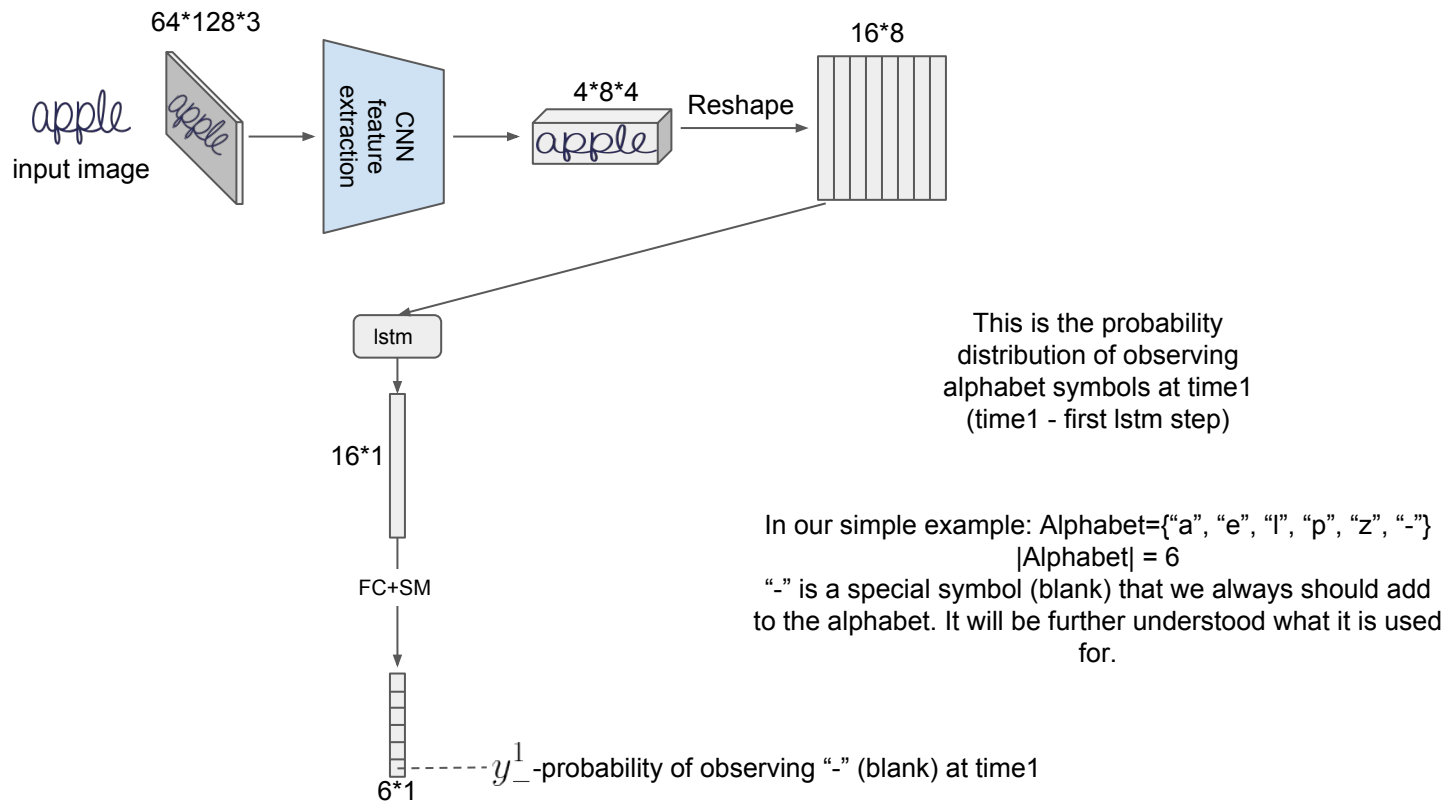


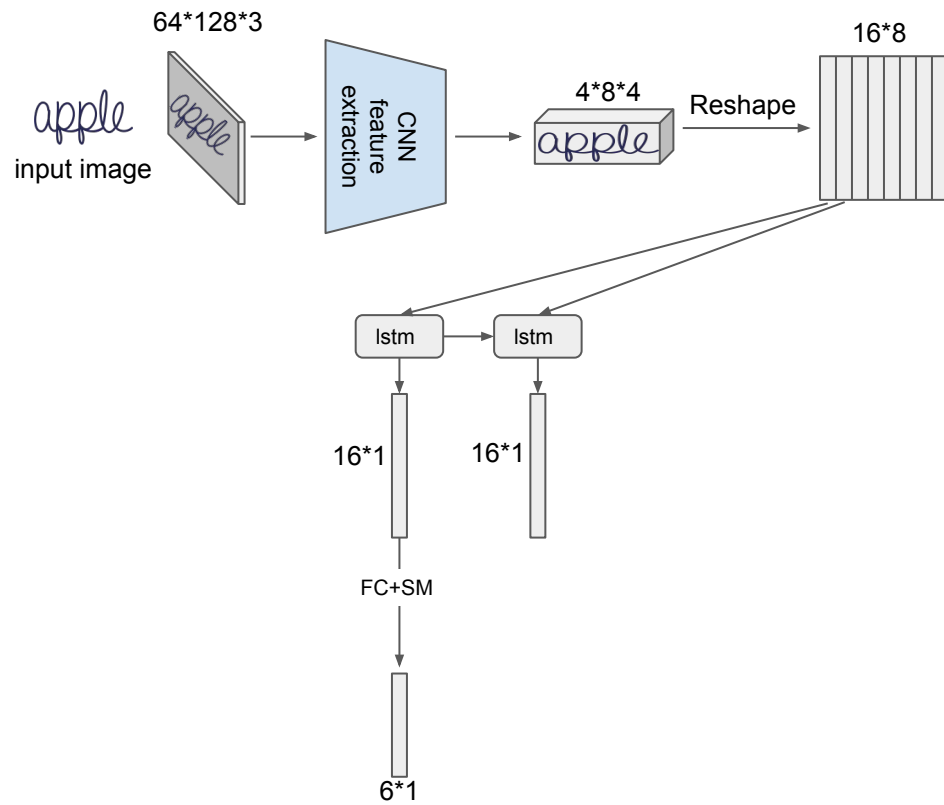


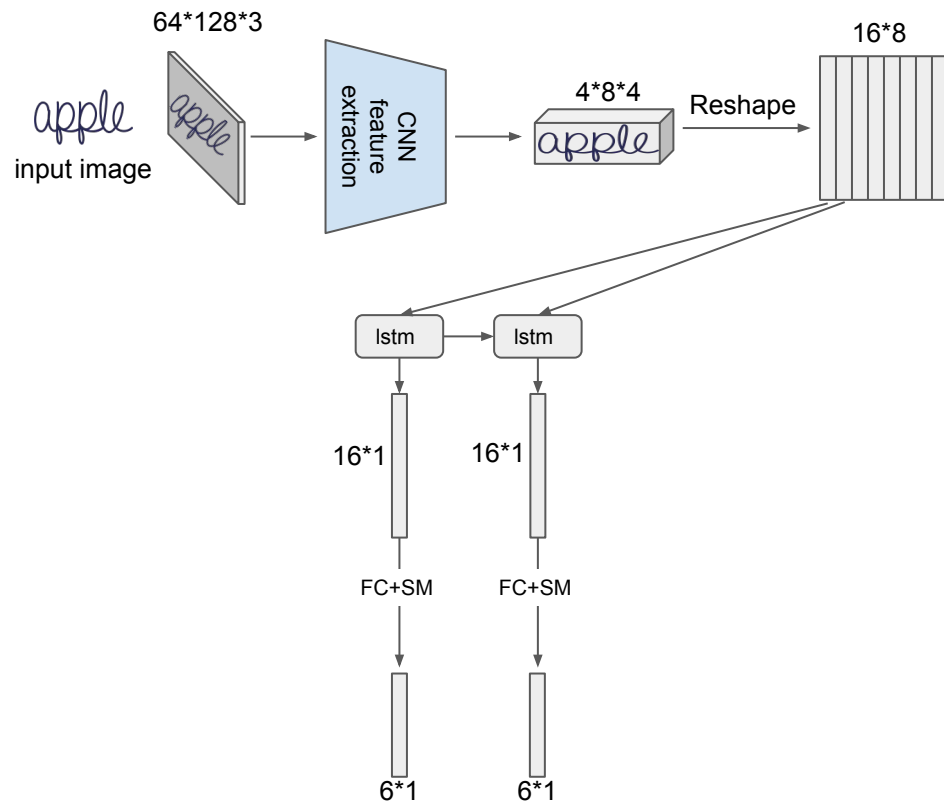


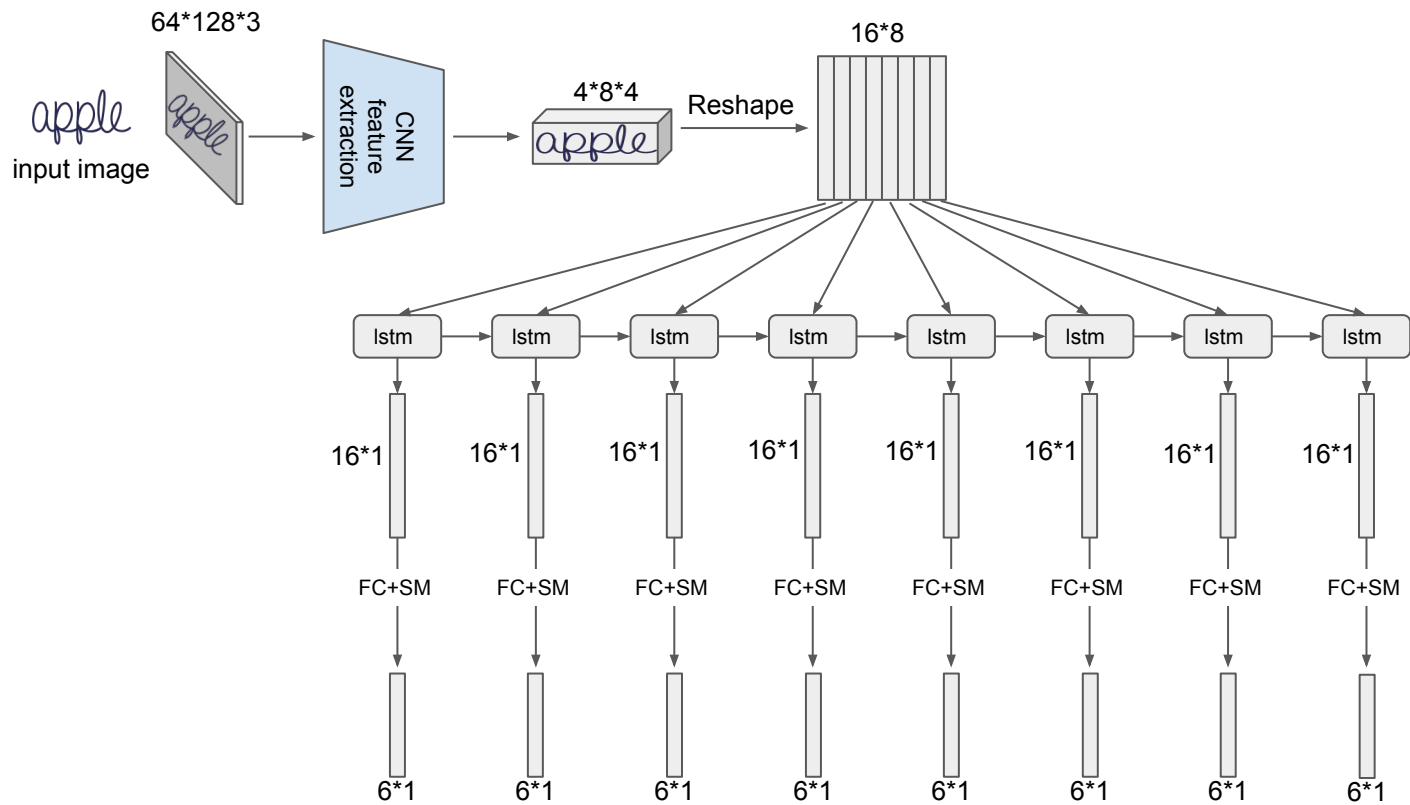


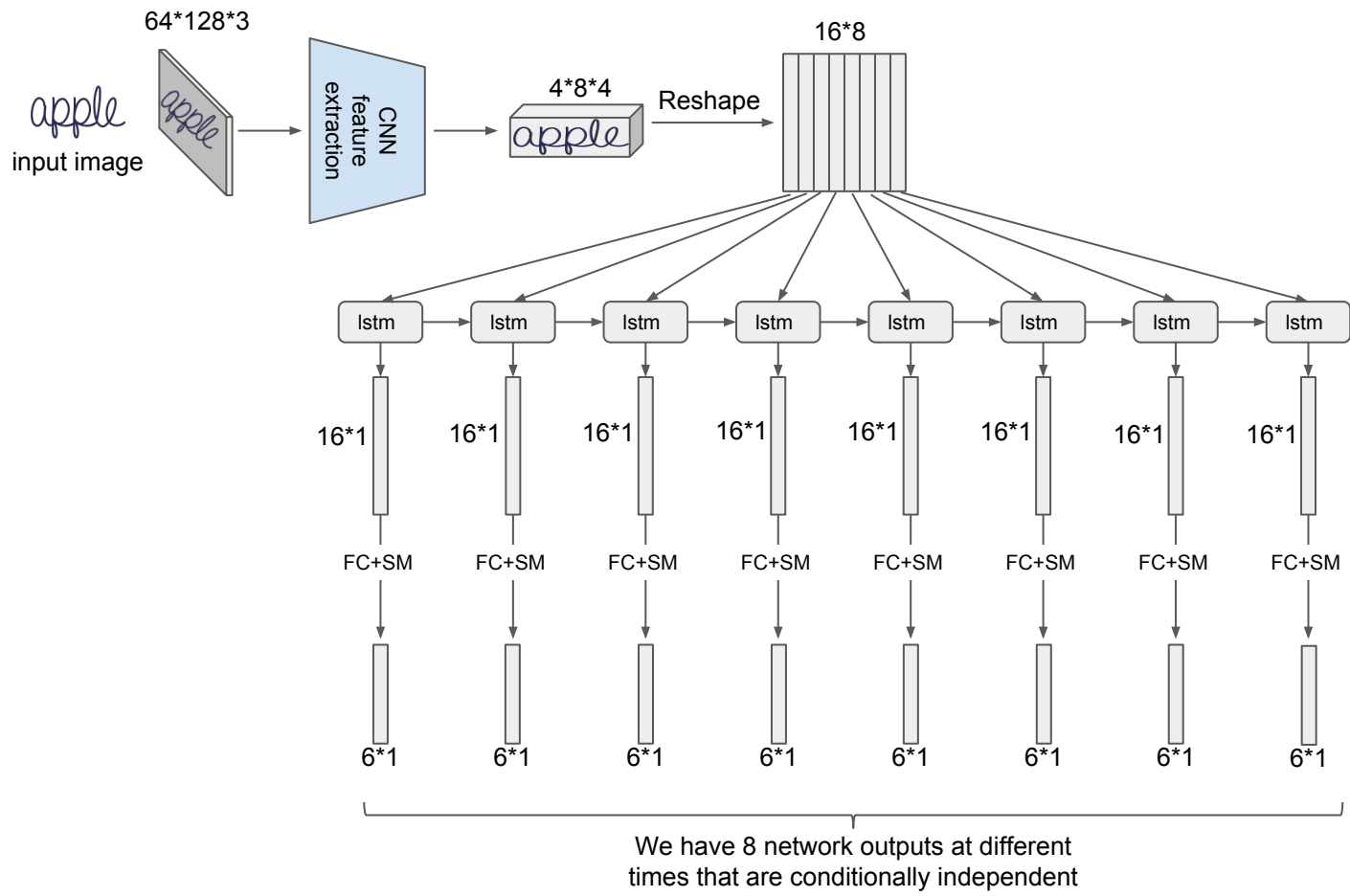


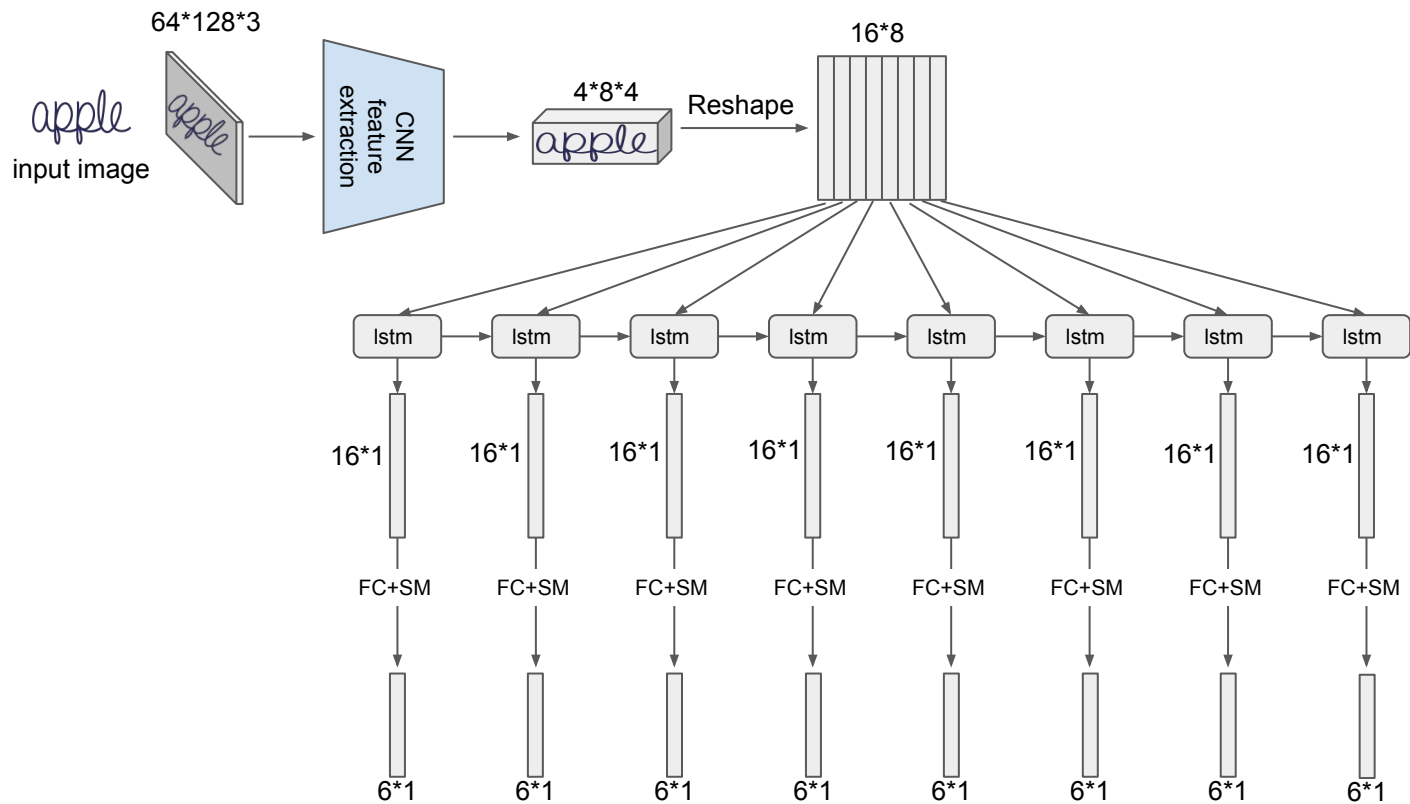






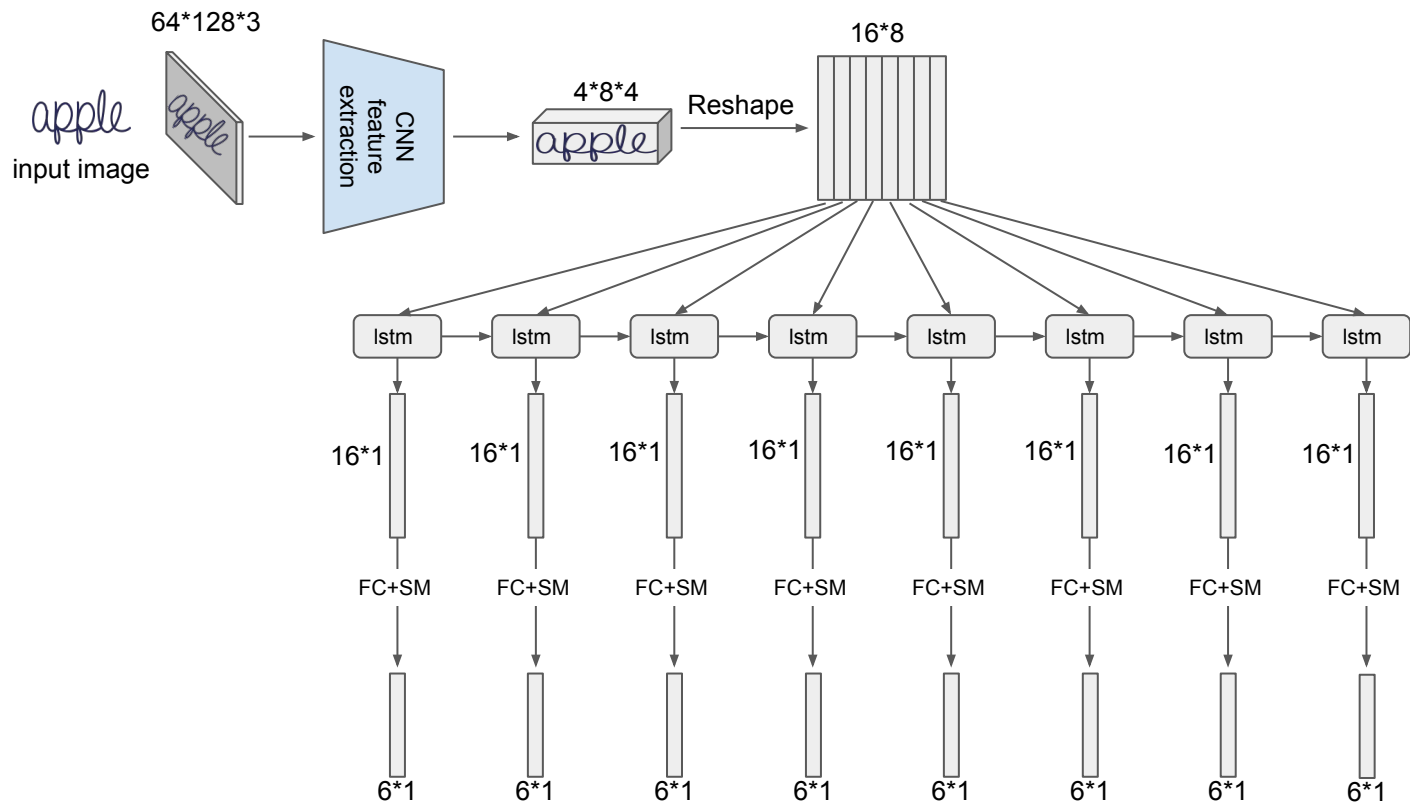






We have 8 network outputs at different times that are conditionally independent

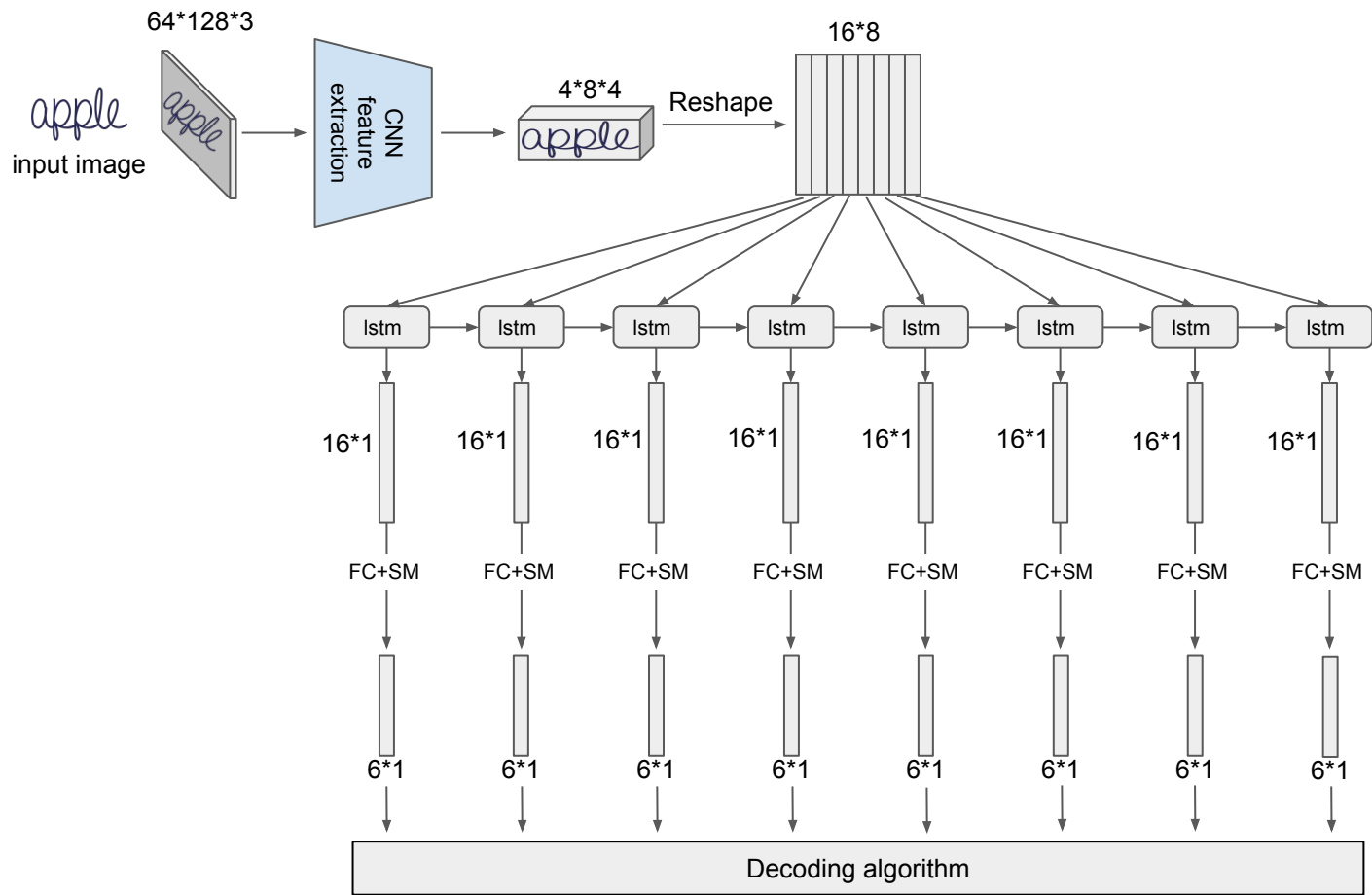
Note: We designed simplified neural network to have 8 outputs. It means that we can not recognize more than 8 characters per image.

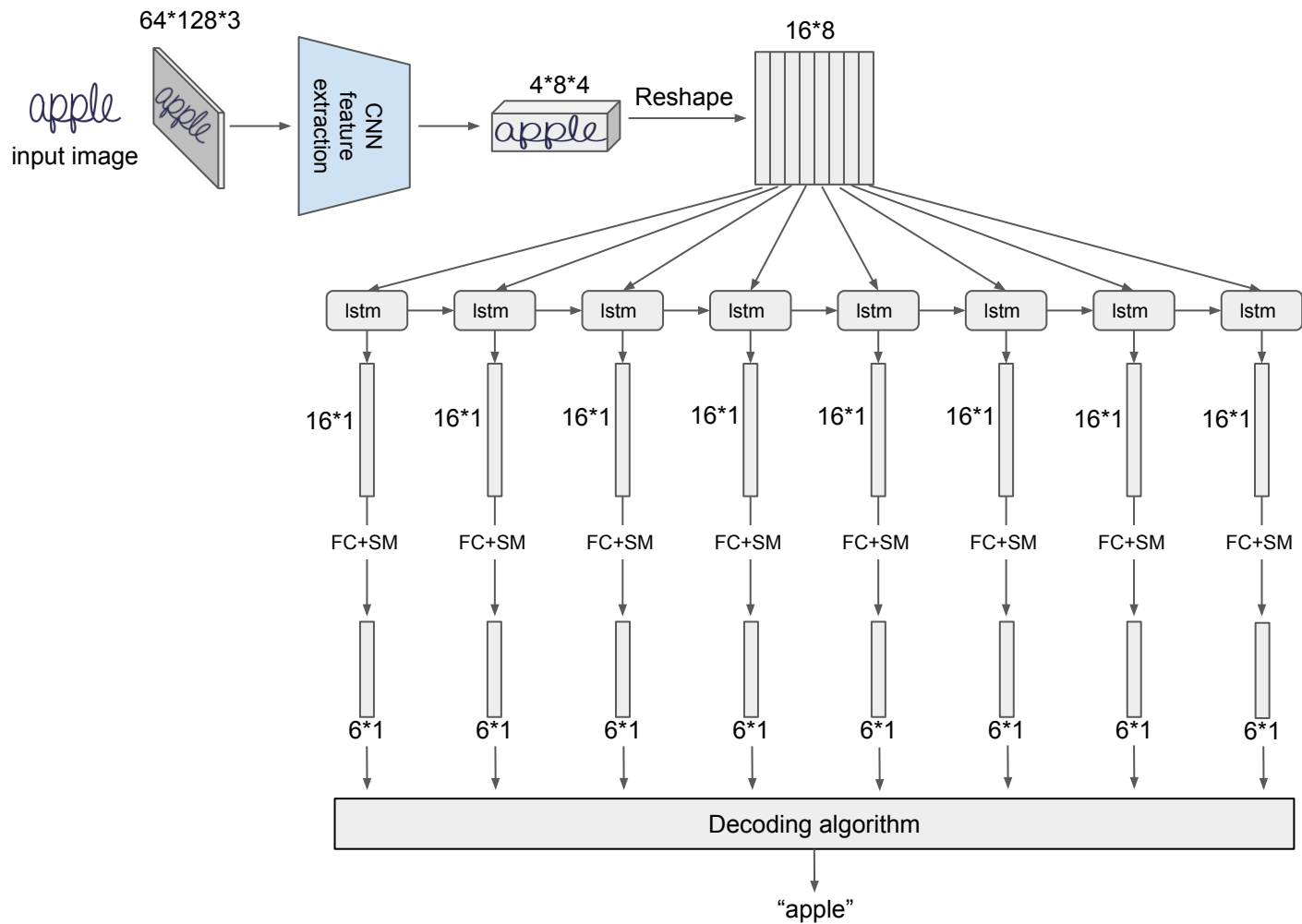


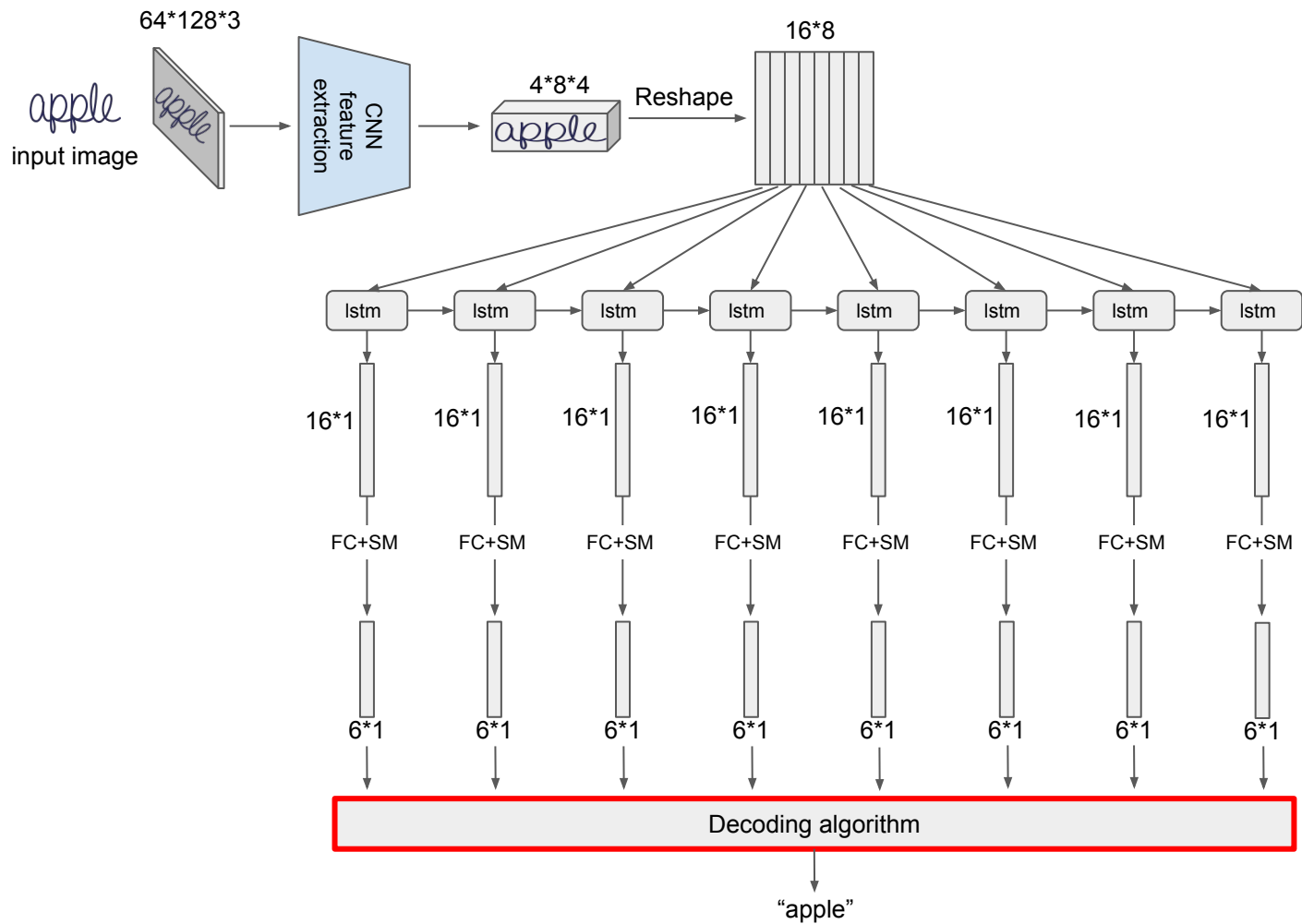
We have 8 network outputs at different times that are conditionally independent

Note: We designed simplified neural network to have 8 outputs. It means that we can not recognize more than 8 characters per image.

In practice, number of outputs can reach 32, 64 or more. The choice will depend on the specific task.







How does decoding
algorithm work?

Image OCR: model architecture

Decoding algorithm



6×1



6×1



6×1



6×1



6×1



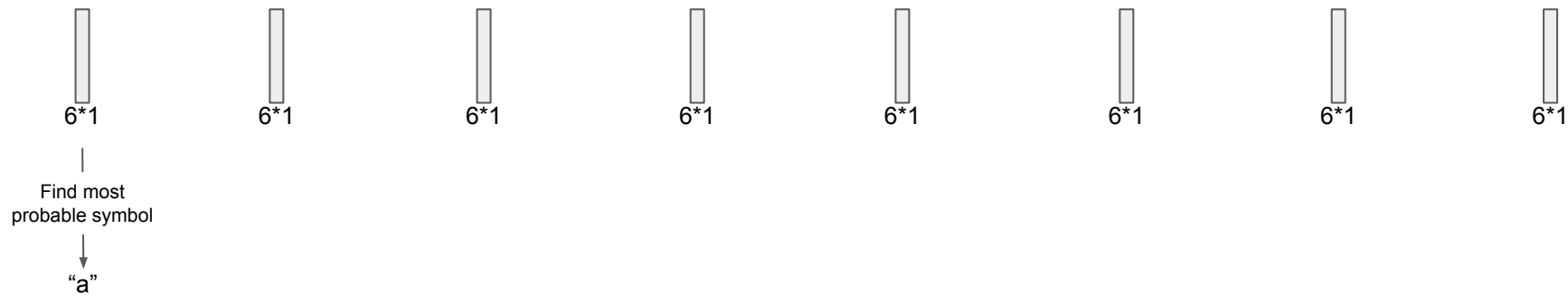
6×1

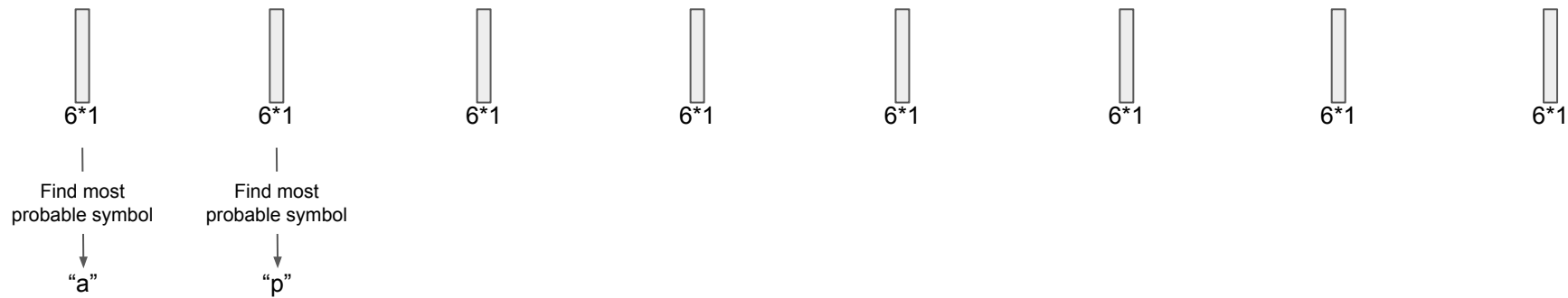


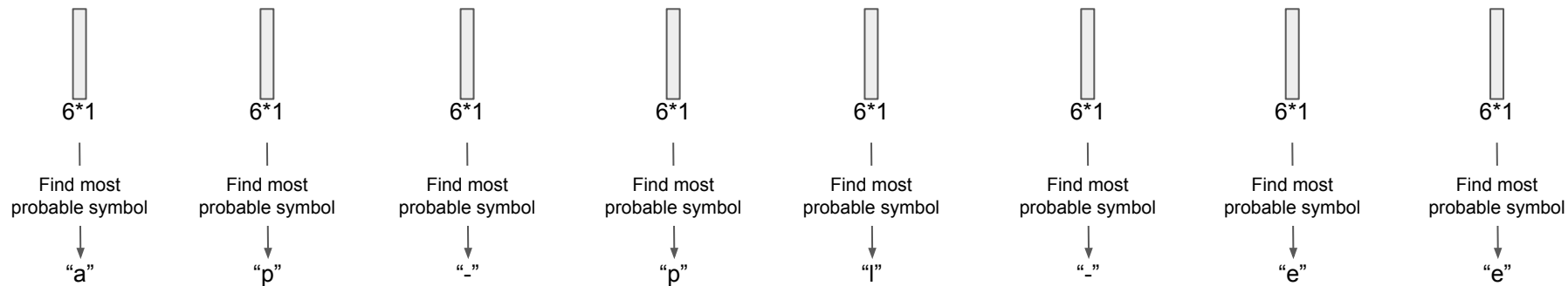
6×1

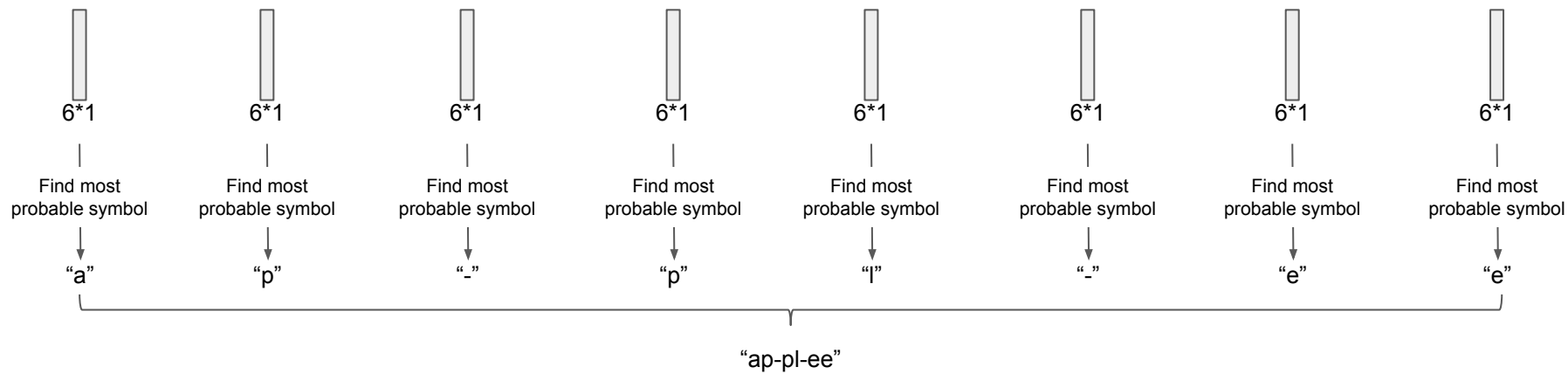


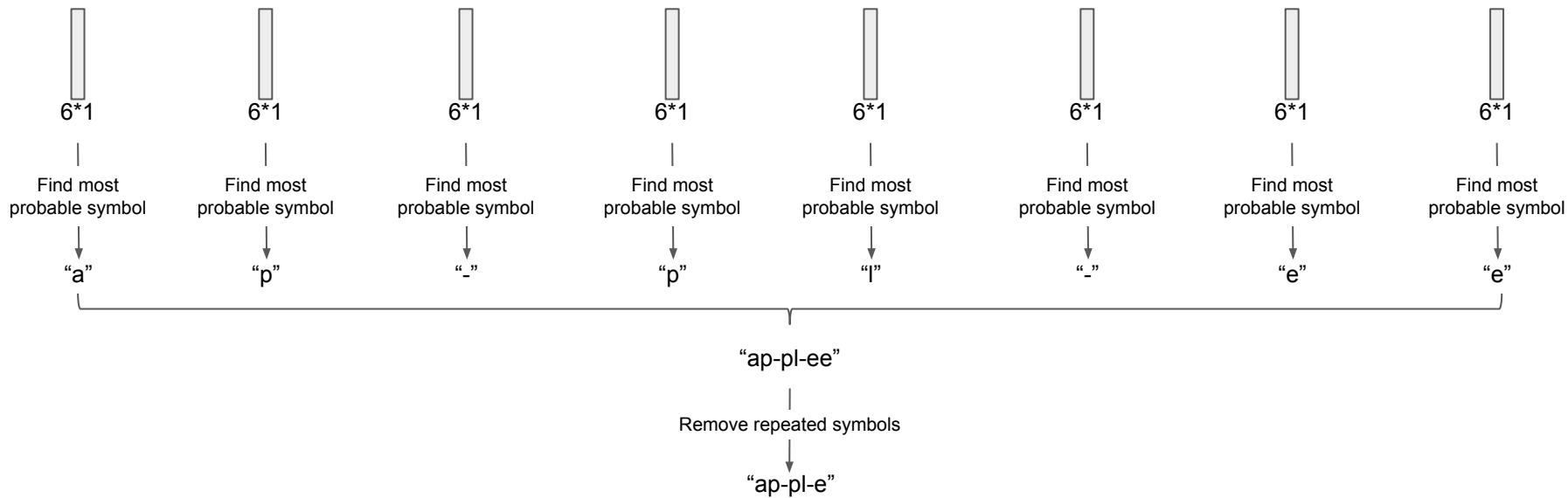
6×1

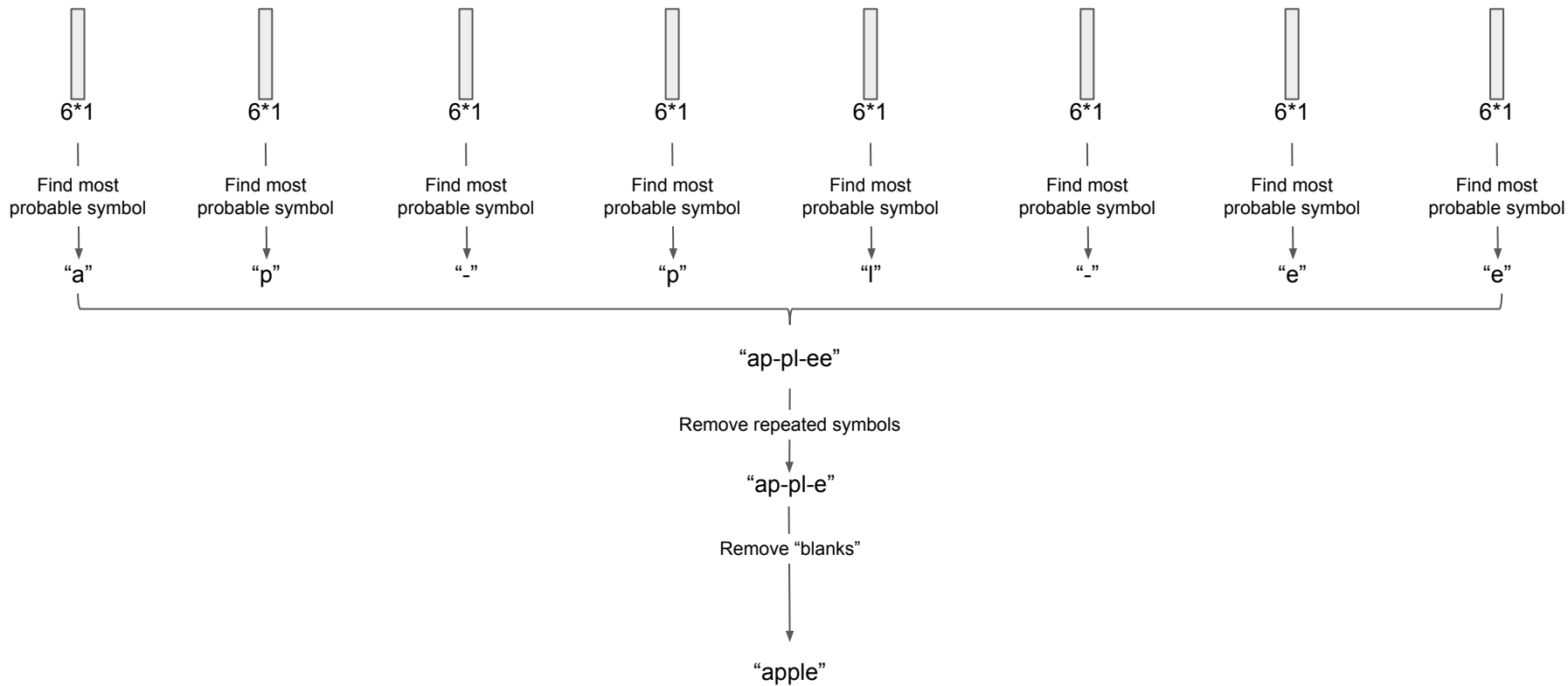


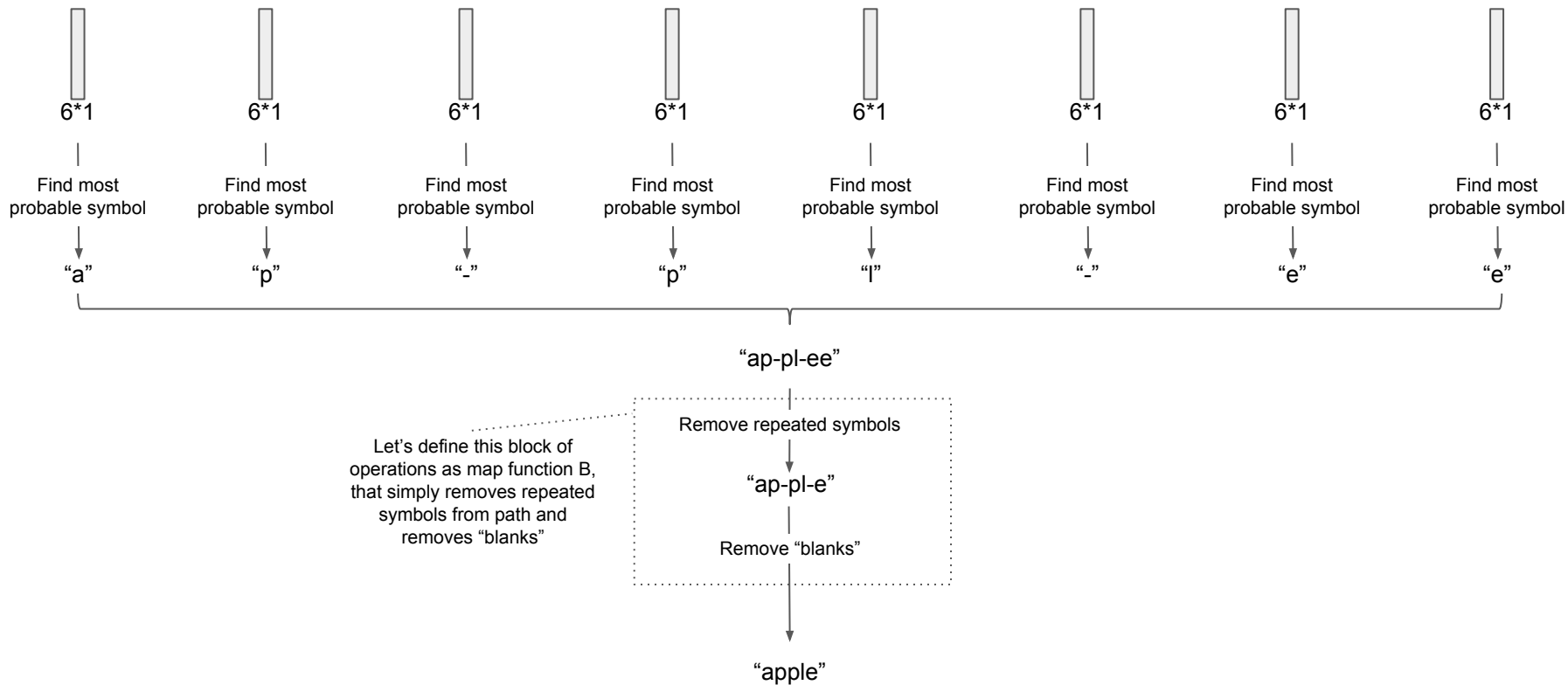


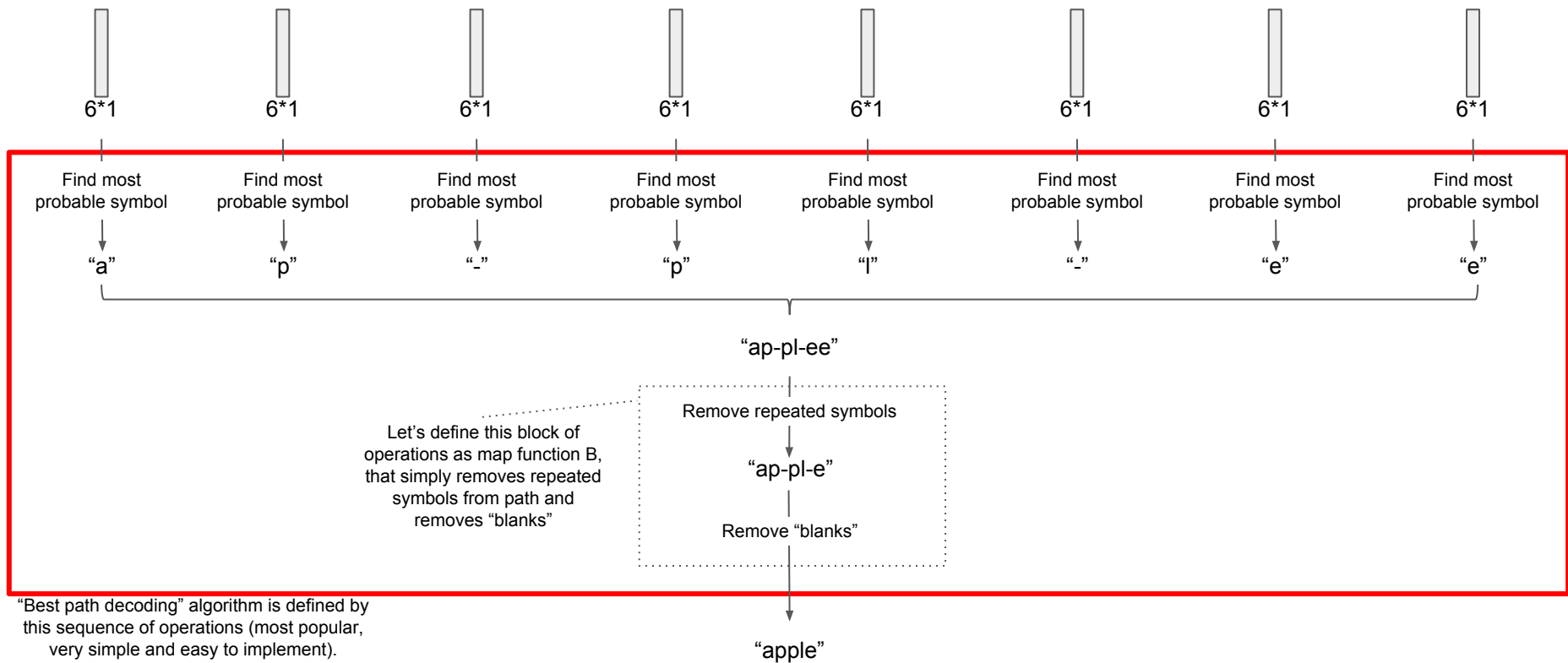


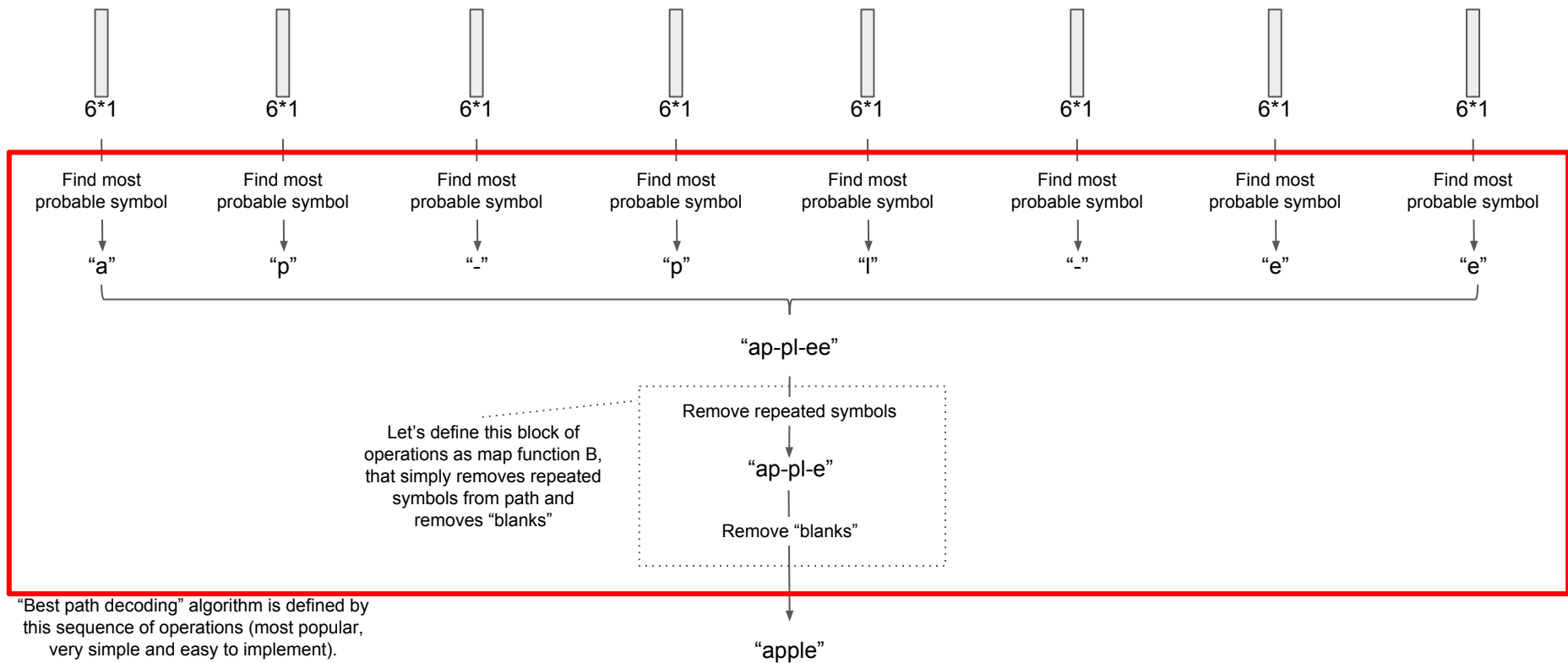












Note: there are a few other algorithms in literature.

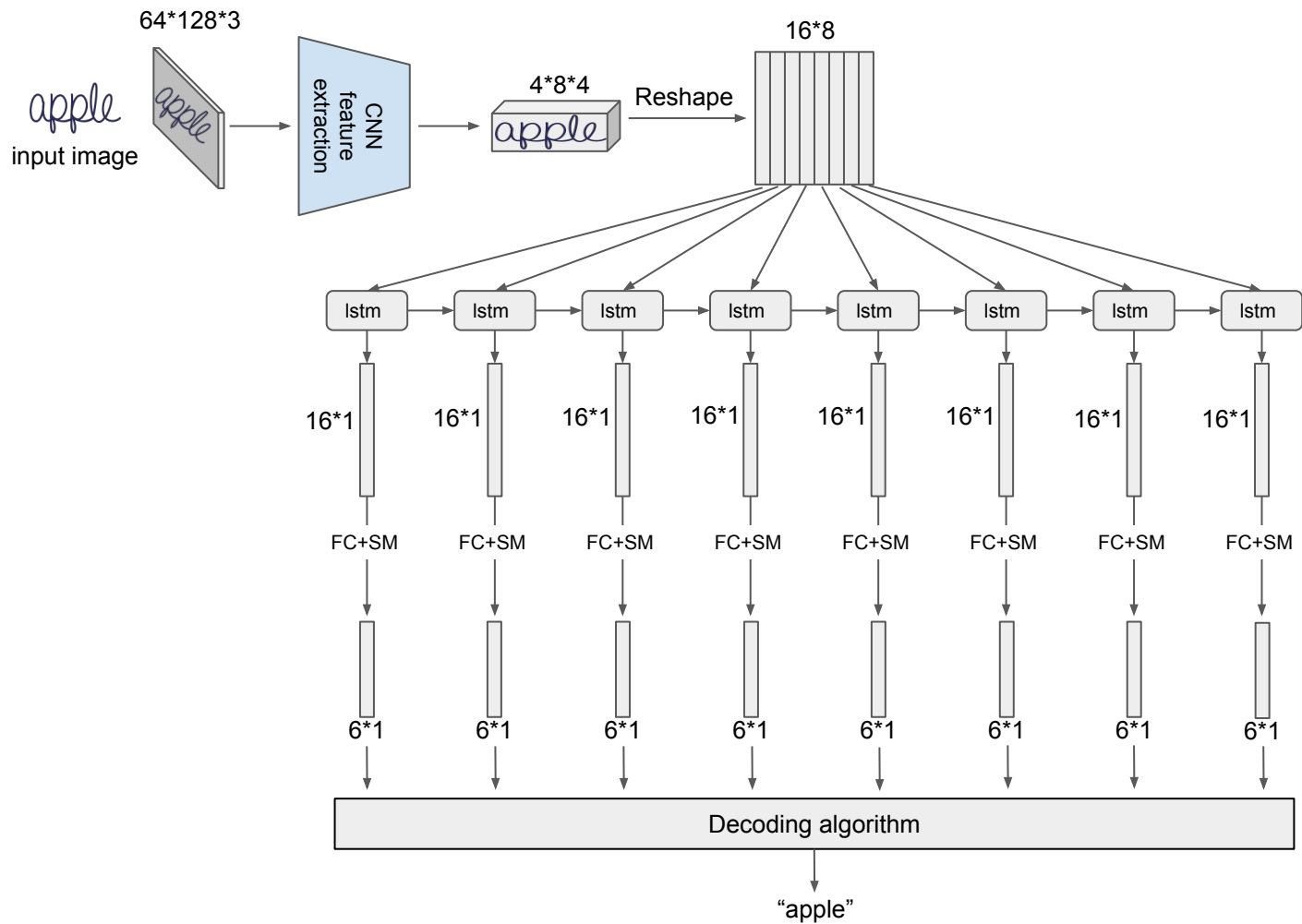
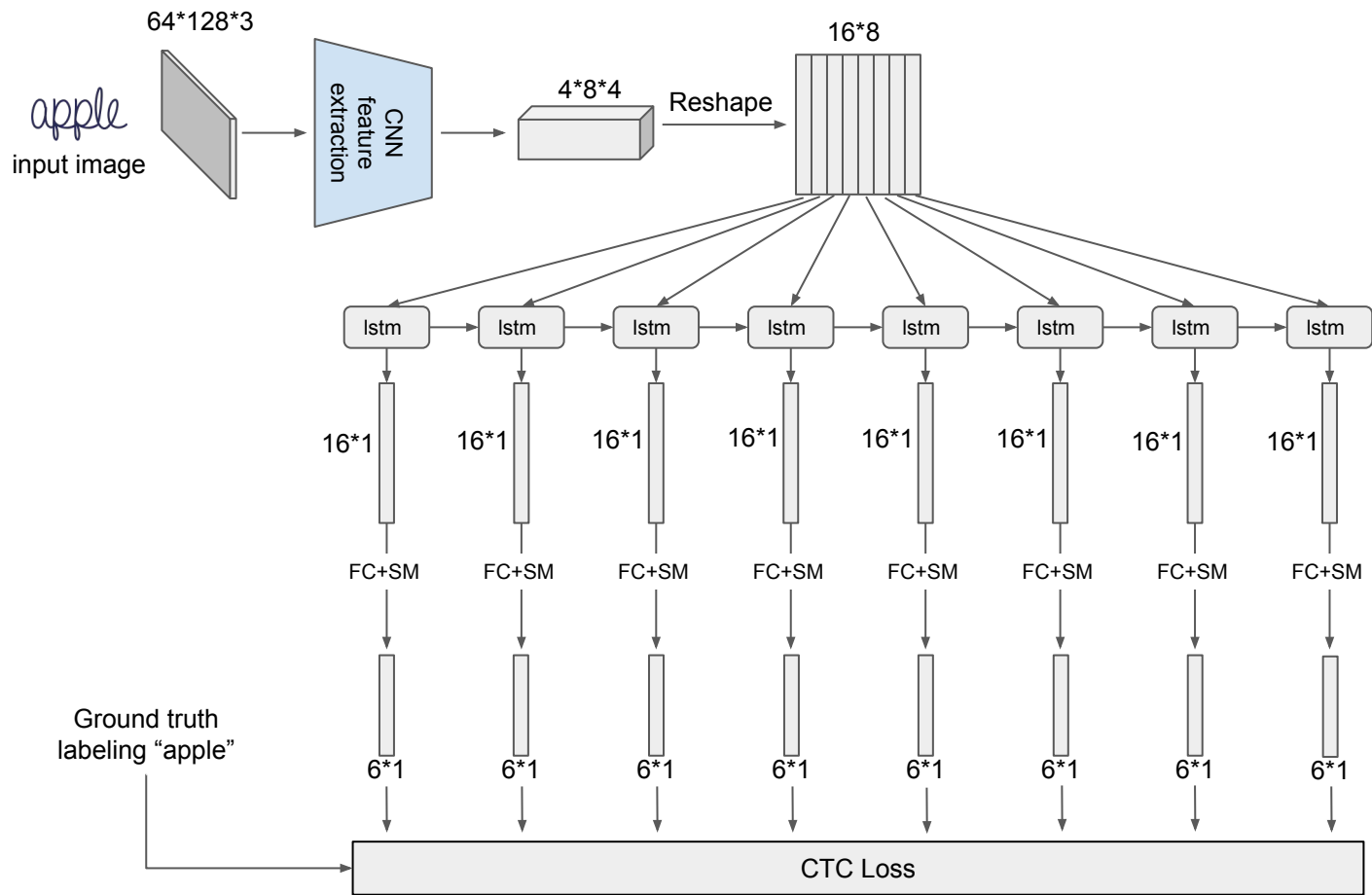


Image OCR: model architecture

Training: CTC Loss



Thank you

Our Website:

deepsystems.ai

Products:

supervise.ly - Dataset management, annotation and preparation service

movix.ai - Interactive, lstm-based movie recommender system

Outsource projects:

Our team is looking for business partners to make exciting deep learning solutions.