

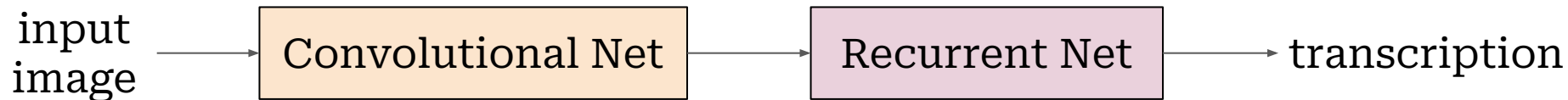
4.1 - HTR Models

Types of (DL) HTR Models

- Connectionist Temporal Classification (CTC) based
 - oldest but still state-of-the-art
- Sequence to Sequence
 - precursor to transformers
 - rarely used anymore
- Transformers
 - gaining popularity
 - sometimes issues with smaller datasets

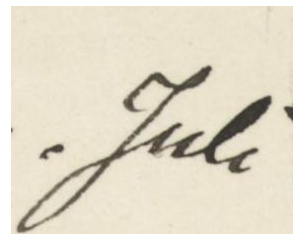
CTC-based Models

generally: Convolutional Recurrent Neural Networks (CRNN)



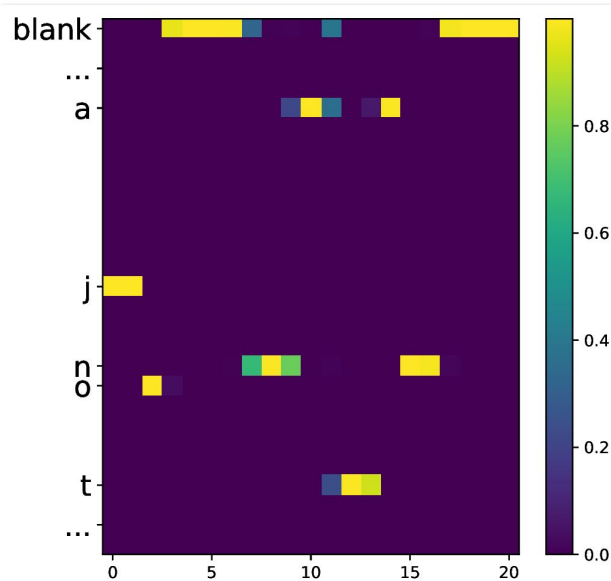
ConvNet

- main purpose: feature extraction
- output shape: batch size X filters X height X width
 - width = sequence length (timesteps)
- reshaped to: width X batch size X filters*height
- sequence length longer than actual transcription
 - each character spans more than one pixel column



RNN

- main purpose: text sequence “generation”
- output format: alphabet size X sequence length
 - one row per character, one column per timestep
 - probability of character X at timestep t



We'll see how to turn this representation into an actual transcription in a moment!

4.2 - The CTC Loss

CTC Loss

- prediction length \geq target length
 - no explicit alignment: prediction \leftrightarrow target
- scores all possible alignments between target and prediction
 - special blank character to allow for double-characters
- much more detailed explanation: <https://distill.pub/2017/ctc/>

CTC Loss - Alignment

target: hi, output length: 5

- hi###
- h#i##
- hhiii
- h#iii
- #hi##
- ##hi#
- ...

target: hello, output length: 7

- hel#lo#
- #hel#lo
- hhel#lo
- hel#loo
- heel#lo
- hell#lo
- ...

Hands-On Time!

HTR Models

Take a look at the HTR model in `htr/model.py`, to get an idea of what an implementation can look like.

Feel welcome to modify the model – in the interest of time, don't create a massive model.

If your laptop setup allows it, start training a model, e.g. using the provided Ibsen data, and observe how the training and validation loss behave. You can also take a look at `best_val_results.json` to see what the model recognises.

Alternatively, take a look at both files in the model folder that is included in the lesson material.