Default architecture from Laia for handwritten text recognition
<https://github.com/jpuigcerver/Laia>.

image
Image of a single cropped line of text. Shape:
CHANNELS(fixed) × HEIGHT(fixed) × WIDTH(variable).

$3 \times 64 \times W$

Sequential
id: conv1
blocks: 4

Conv2d
id: 0
kernel_size: 3
padding: 1

$16 \times 64 \times W$

BatchNorm2d
id: 1

$16 \times 64 \times W$

LeakyReLU
id: 2
negative_slope: 0.01

$16 \times 64 \times W$

MaxPool2d
id: 3
kernel_size: 2
stride: 2

$16 \times 32 \times W/2$

Sequential
id: conv2
blocks: 4

Conv2d
id: 0
kernel_size: 3
padding: 1

$16 \times 32 \times W/2$

BatchNorm2d
id: 1

$16 \times 32 \times W/2$

LeakyReLU
id: 2
negative_slope: 0.01

$16 \times 32 \times W/2$

MaxPool2d
id: 3
kernel_size: 2
stride: 2

$16 \times 16 \times W/4$

Sequential
id: conv3
blocks: 3

Conv2d
id: 0
kernel_size: 3
padding: 1

$32 \times 16 \times W/4$

BatchNorm2d
id: 1

$32 \times 16 \times W/4$

LeakyReLU
id: 2
negative_slope: 0.01

$32 \times 16 \times W/4$

Sequential
id: conv4
blocks: 4

Conv2d
id: 0
kernel_size: 3
padding: 1

$32 \times 16 \times W/4$

BatchNorm2d
id: 1

$32 \times 16 \times W/4$

LeakyReLU
id: 2
negative_slope: 0.01

$32 \times 16 \times W/4$

MaxPool2d
id: 3
kernel_size: 2
stride: 2

$32 \times 8 \times W/8$

Reshape
id: to_1d
reshape_spec: [2, [0, 1]]

$W/8 \times 256$

LSTM
id: s3blstm
bidirectional: True
dropout: 0.5
num_layers: 3
hidden_size: 256

$W/8 \times 512$

Linear
id: fc

$W/8 \times S$

symbprob
Sequence of posterior probabilities for known
symbols. Shape: SEQ_LENGTH(variable) ×
NUM_SYMBOLS(fixed).