lstm output (w/ packed_sequence input):

one sequence

seq length

sum of sequence lengths of the batch

output of the last layer of every frame of every sequence

hidden_size   hidden_size

forward pass   backward pass

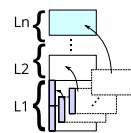includes batch_size indirectly
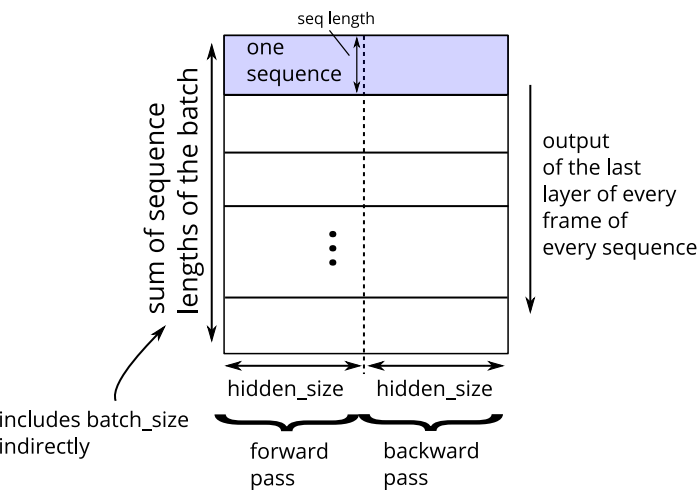
output = [sum of seq. lengths, D*hidden_size]

h_T output:

output for one sequence of the batch (last frame)

forward pass

backward pass

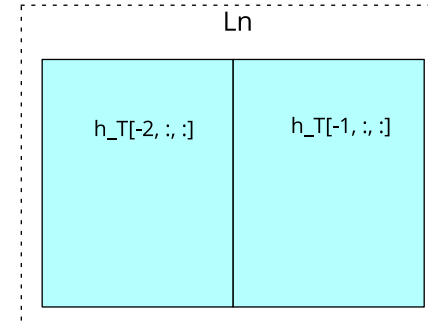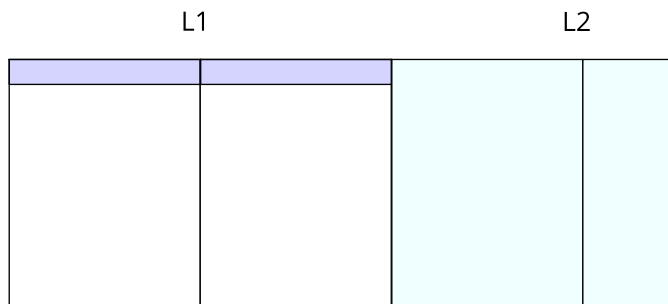contains the last output h_T of each sequence in the batch

hidden_size

batch_size

L1

L2

Ln

D*L
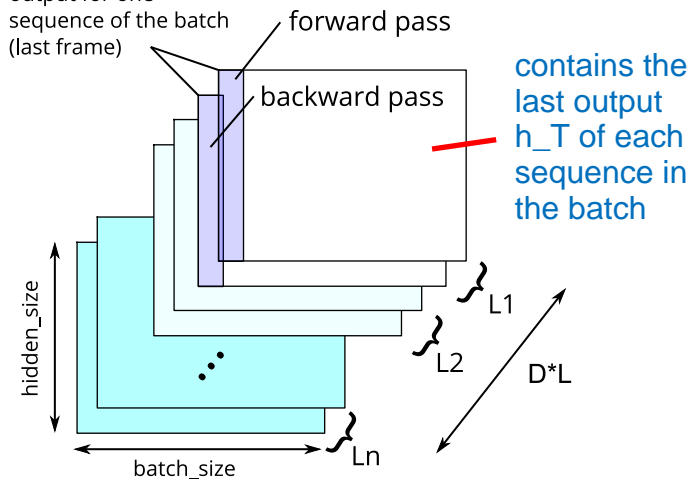
h_T = [D*L, batch_size, hidden_size]

Ln

L2

L1

Reshaping for follow-up linear layer

reshaped to [batch_size, D*L*hidden_size]:

L1   L2   Ln

...

$h_T[-2, :, :]$ $h_T[-1, :, :]$

relevant output
h_T = cat((h[-1,:,:], h[-2,:,:]), dim=-1)

(reshaping not really needed, simply use the last (for lstms) or last two (bilstms) layer's output as shown in blue

Notation:

[..., ..., dim 3]

[dim 1, ..., ...]

[..., ..., ..., dim 4]

[..., dim 2, ...]