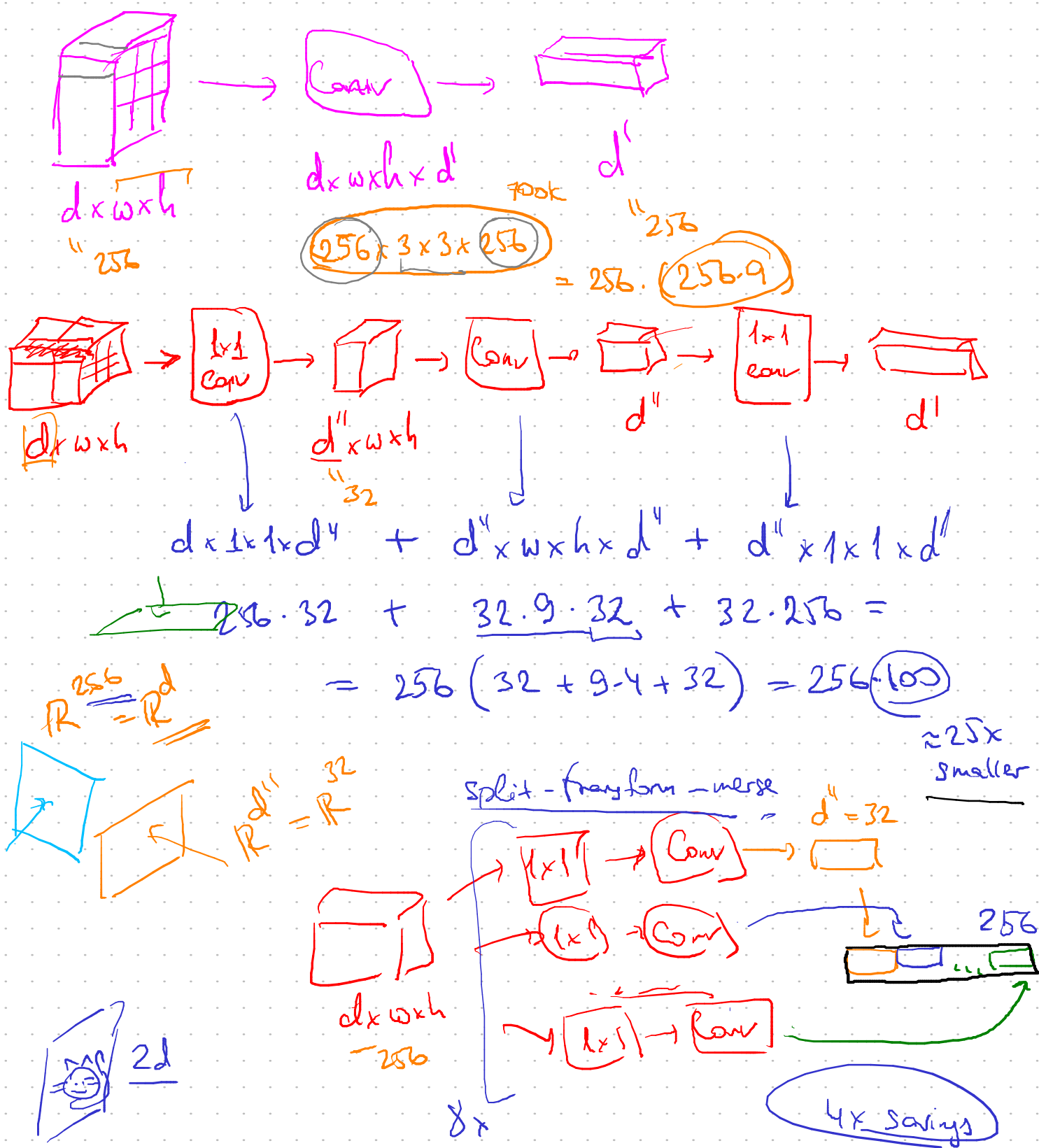


① Bottlenecks & split-transform-merge

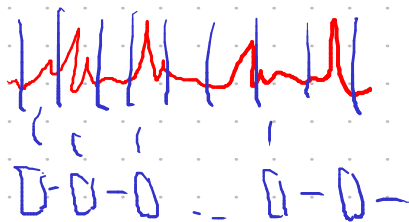


② RNNs

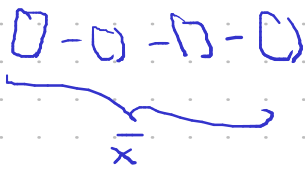
text

time series
sound

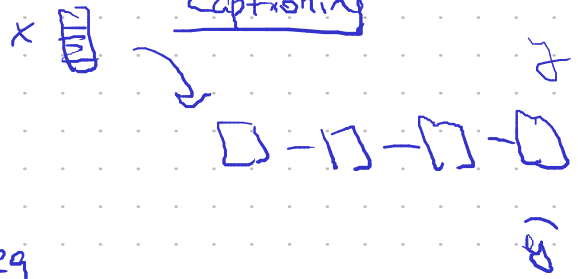
videos



classification



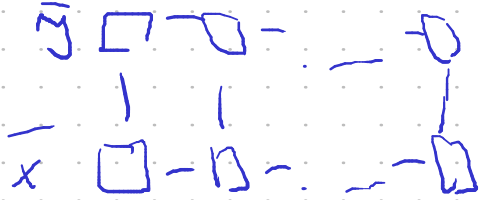
captioning



Seq2seq

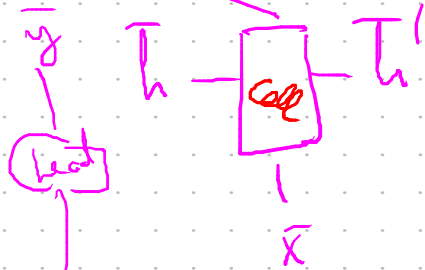
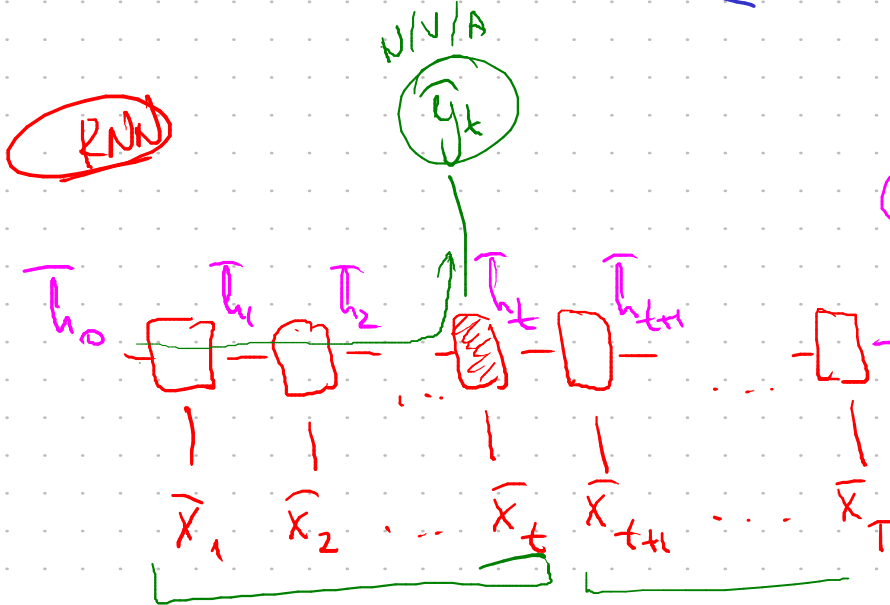


sequence labeling



pos tagging / subtitles / object tracking

$$f(x, h) = (h', \bar{y})$$

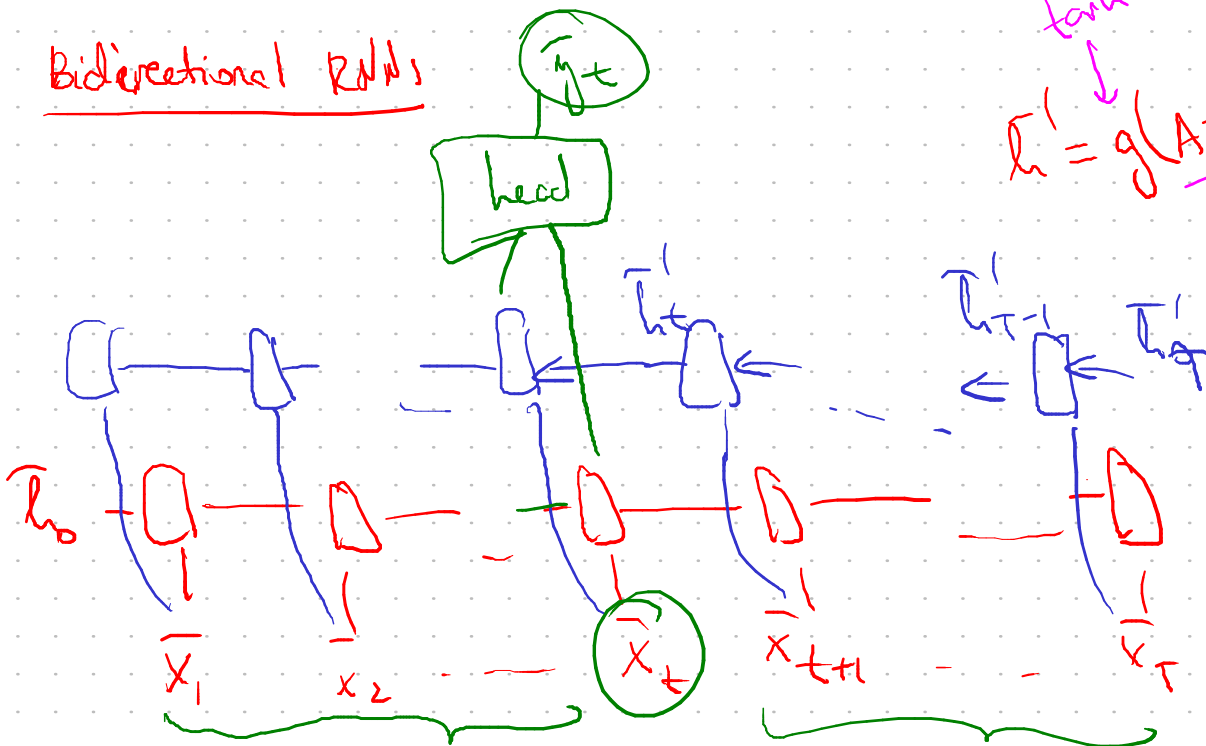


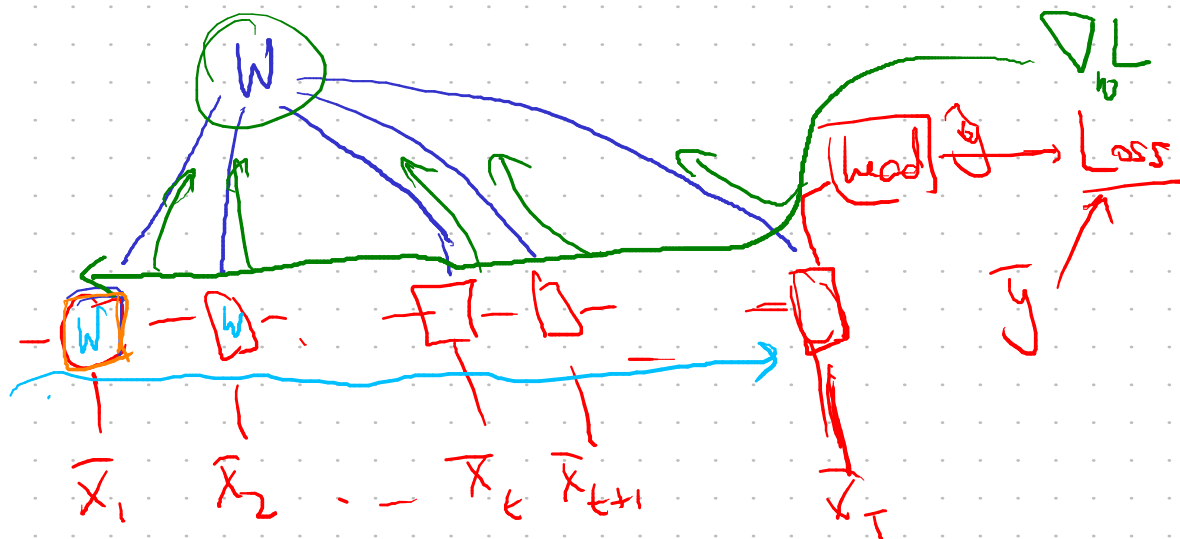
$$h' = g(A\bar{x} + B^T h)$$

ReLU tanh

weights

Bidirectional RNNs





$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial w_T} + \frac{\partial L}{\partial w_{T-1}} + \dots \rightarrow \frac{\partial L}{\partial w_1}$$

$$h_0 \rightarrow [W] \rightarrow h_1 \rightarrow [W] \rightarrow h_2 \rightarrow \dots$$

$$\bar{h}_1 = W \bar{h}_0, \bar{h}_2 = W^2 \bar{h}_0 \rightarrow \bar{h}_T = W^T \bar{h}_0$$

"spectral radius"

$$\|\bar{h}_t\| = \rho \cdot \|\bar{h}_{t-1}\|$$

$$\|\bar{h}_0\| = 1$$

$$\|\bar{h}_T\| = \rho^T \cdot \|\bar{h}_0\|$$

- exploding gradients

$$\rho > 1$$

/ clip the gradients

- vanishing gradients

$$\rho < 1$$