

Normalized Averaged Gradients of the Last Time-Step Loss:  $\nabla_{\mathbf{h}_t} \mathcal{L}_T$  for RNN

