

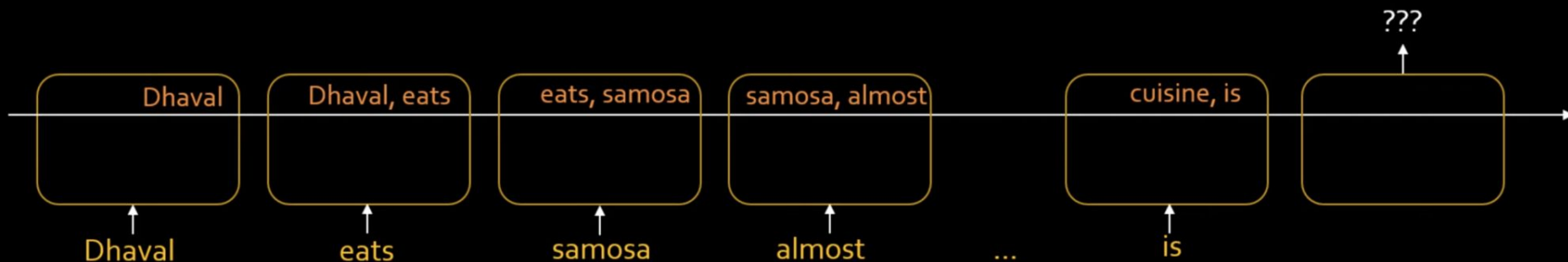


Basic RNN suffer from short term memory problem

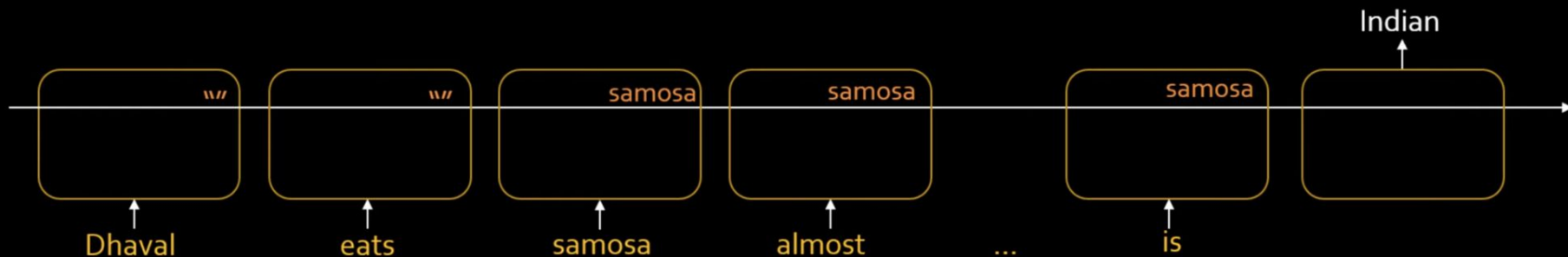
Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian



Traditional RNN



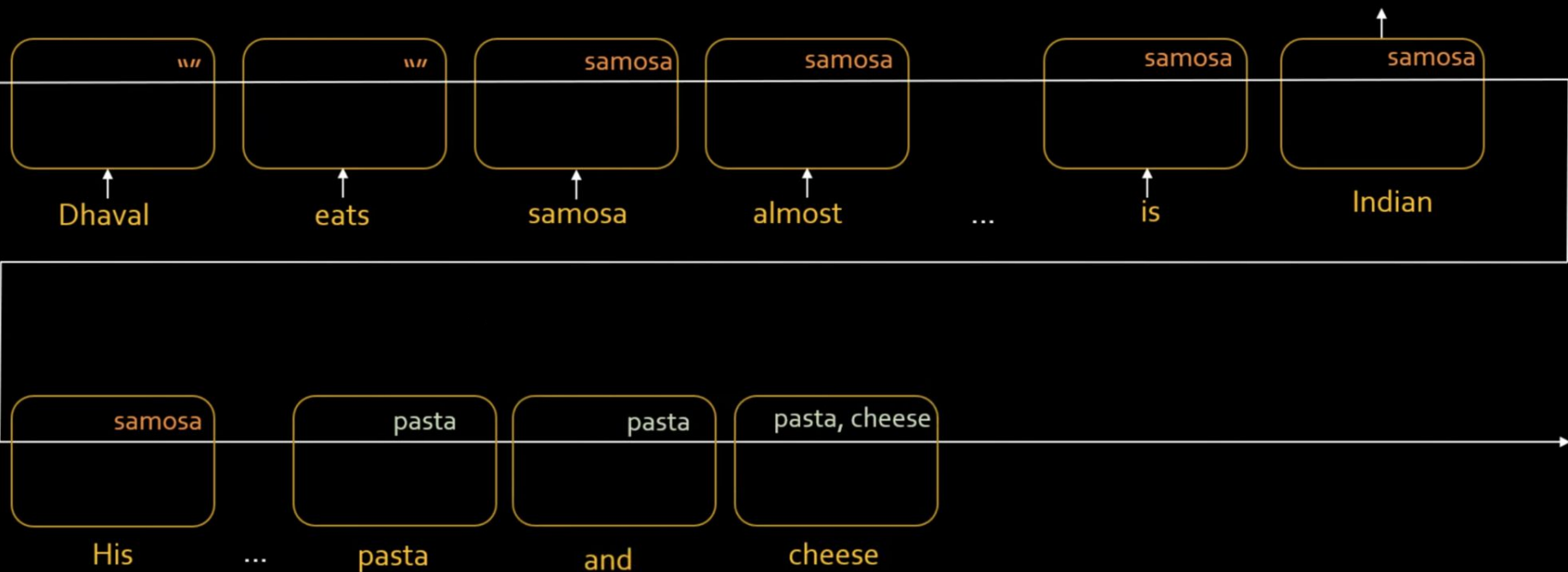
Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian



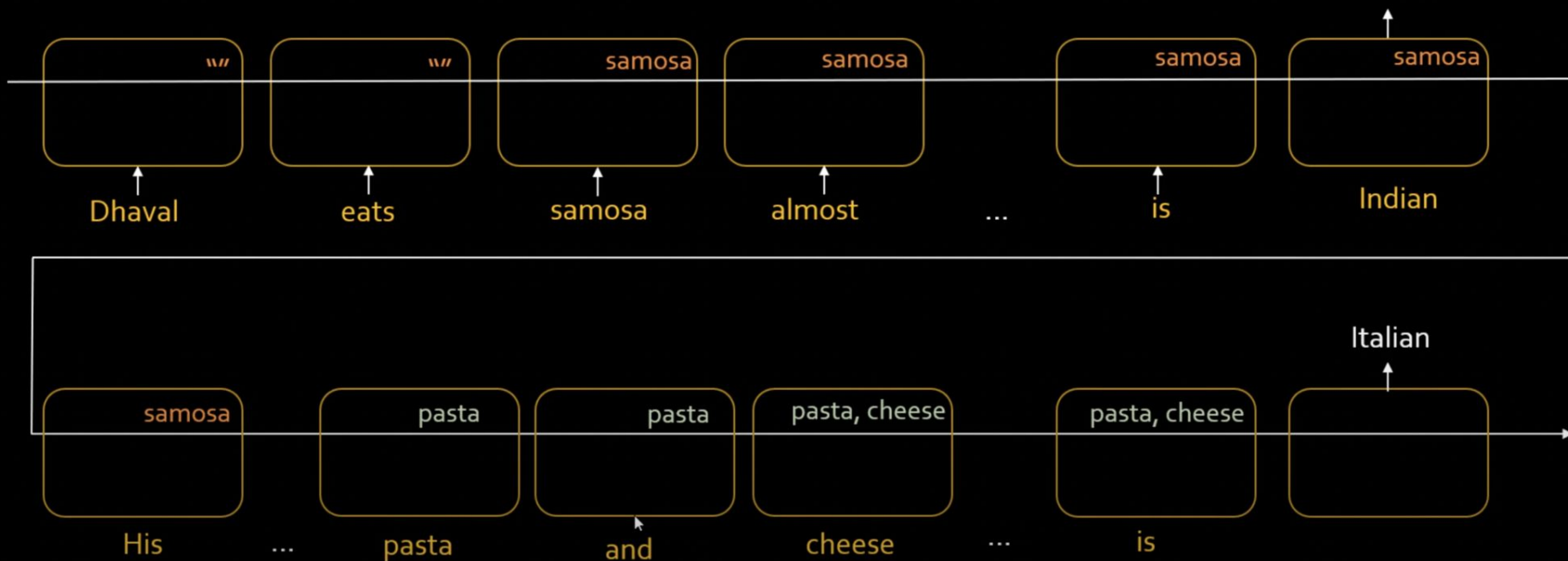
Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pastas and cheese that means Bhavin's favorite cuisine is ...

Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pastas and cheese that means Bhavin's favorite cuisine is Italian

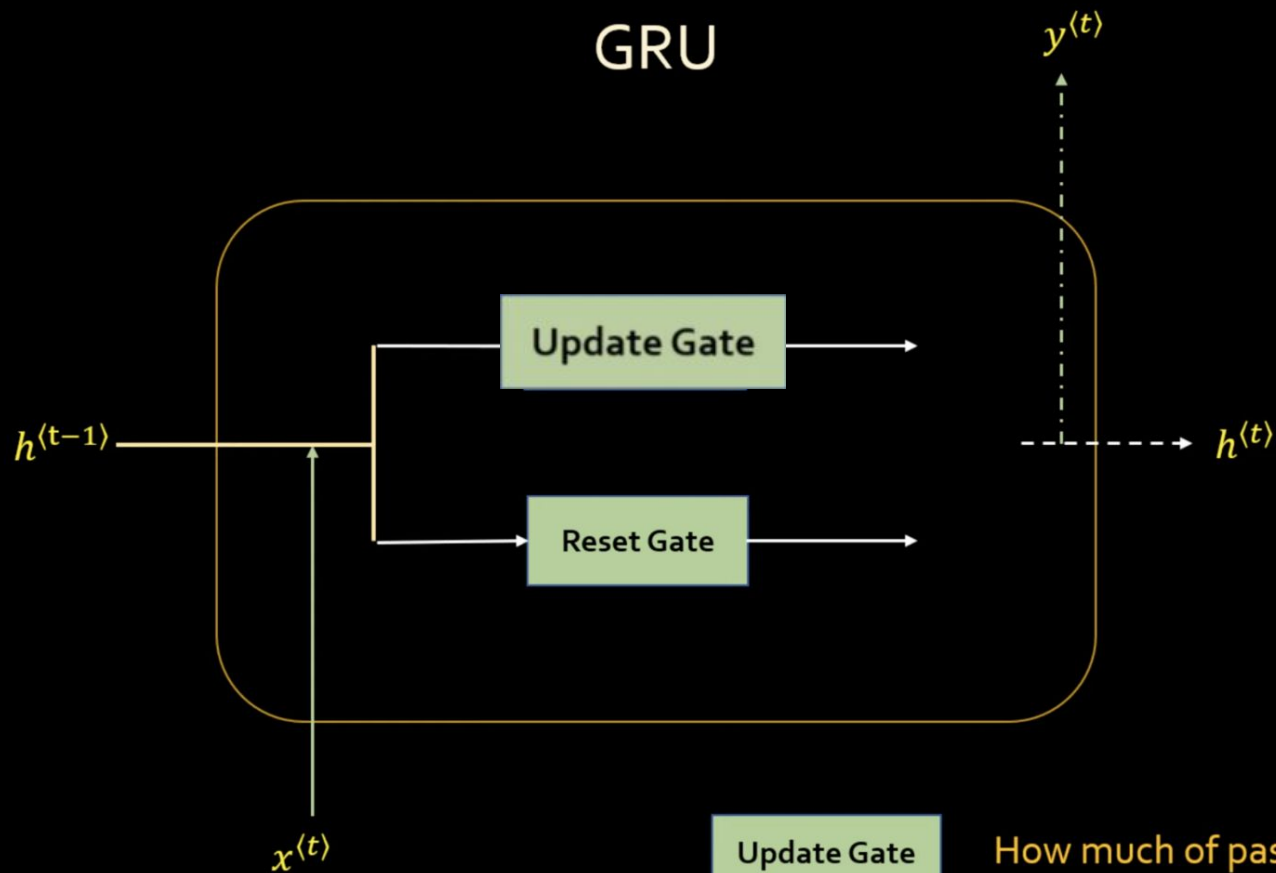
Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pasta and cheese that means Bhavin's favorite cuisine is Italian



Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pasta and cheese that means Bhavin's favorite cuisine is Italian



GRU



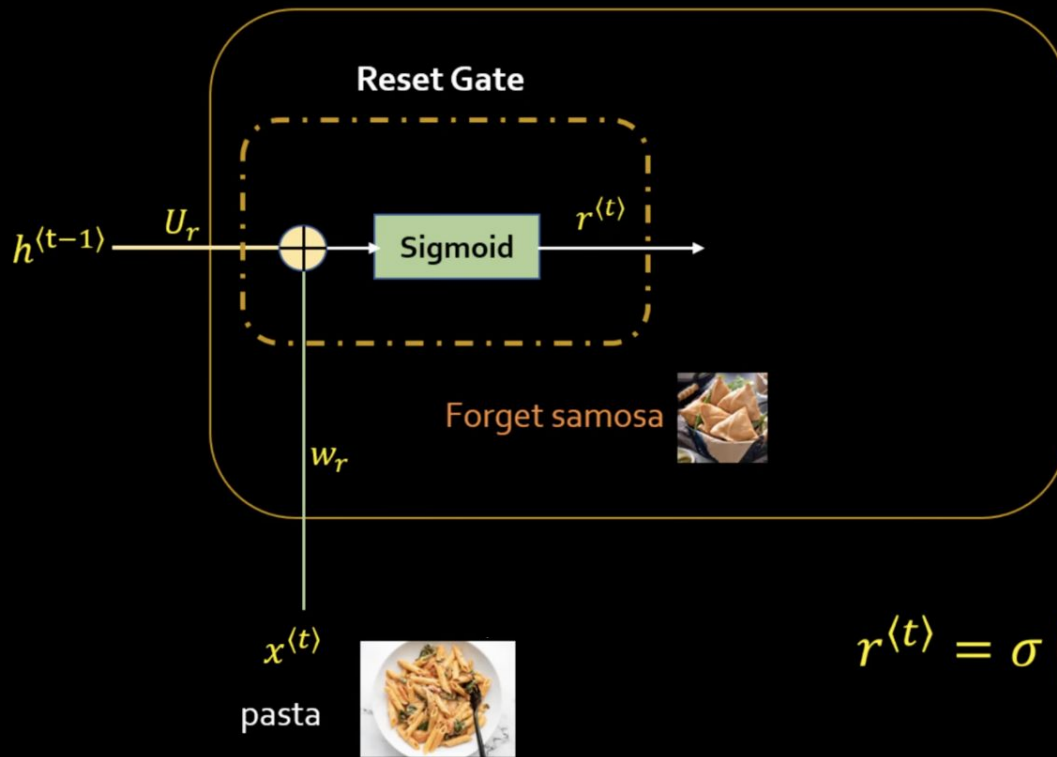
Update Gate

How much of past memory to retain

Reset Gate

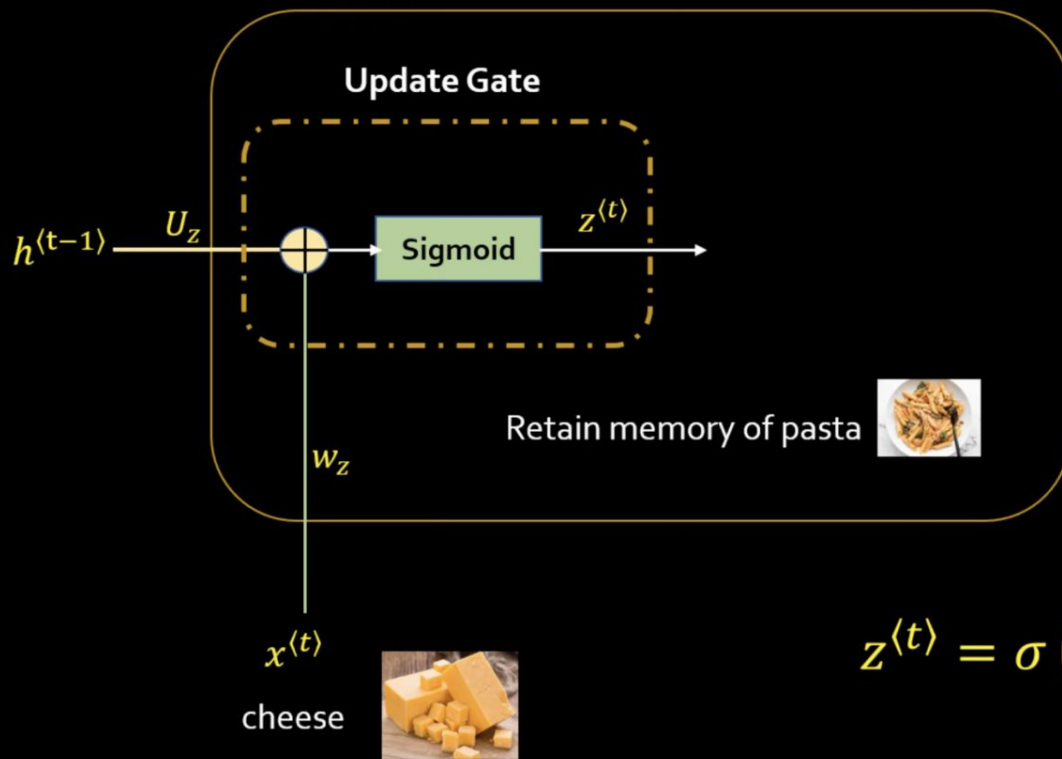
How much of past memory to forget

Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pasta and cheese that means Bhavin's favorite cuisine is Italian

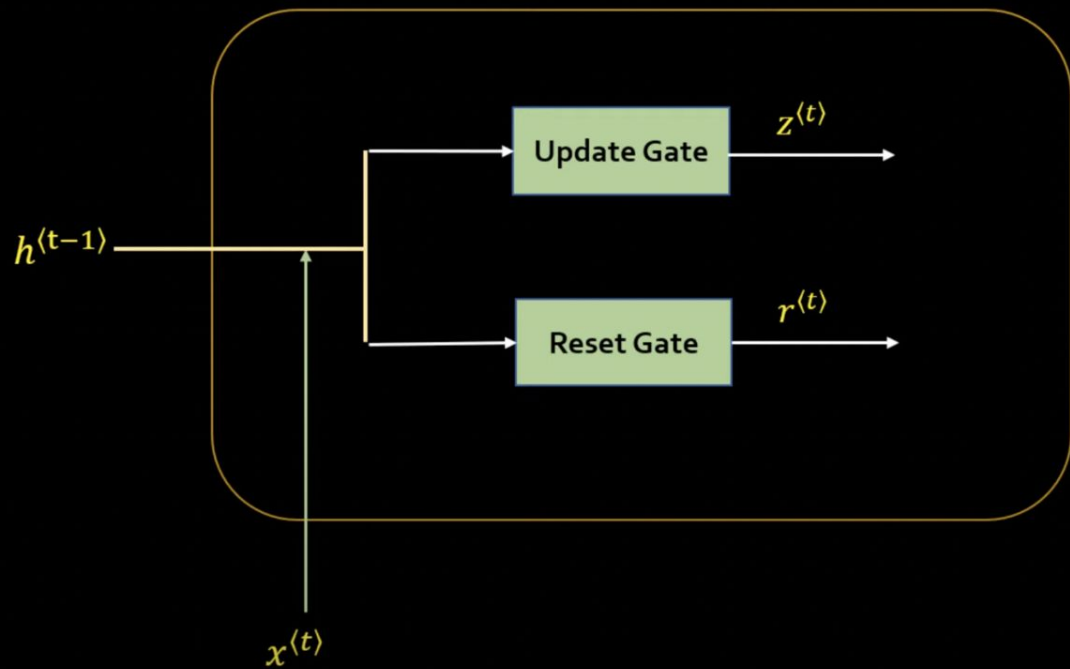


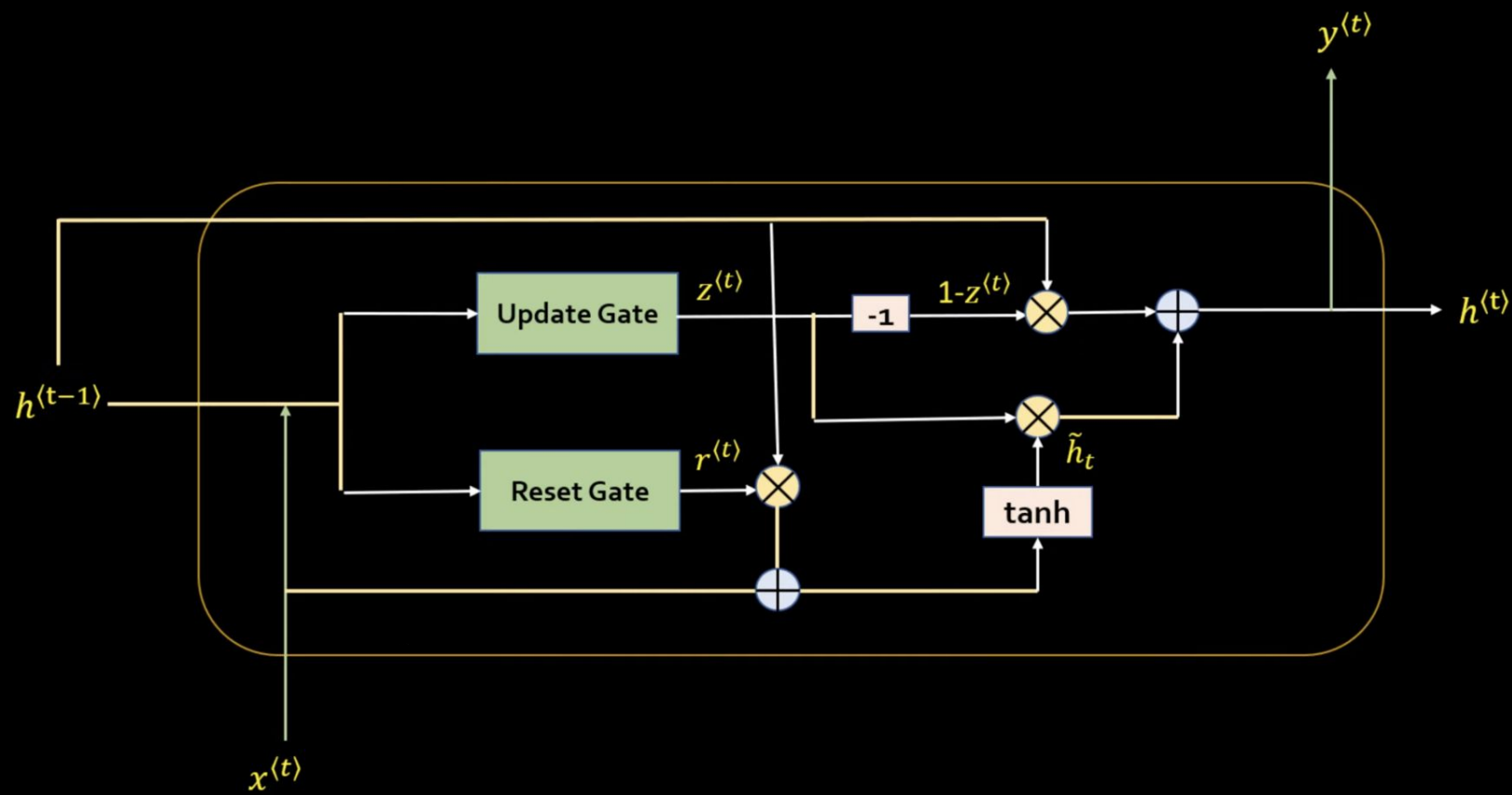
$$r^{(t)} = \sigma \left(w_r x^{(t)} + U_r h^{(t-1)} \right)$$

Dhaval eats samosa almost everyday, it shouldn't be hard to guess that his favorite cuisine is Indian. His brother Bhavin however is a lover of pasta and cheese that means Bhavin's favorite cuisine is Italian



$$z^{(t)} = \sigma \left(w_z x^{(t)} + U_z h^{(t-1)} \right)$$





LSTM	GRU
3 Gates: Input, output, forget	2 Gates: reset, update
More accurate on longer sequence, less efficient	More efficient computation wise. Getting more popular
Invented: 1995 - 1997	Invented: 2014