



ОБРАБОТКА ВИДЕО

Дмитрий Коробченко

ВИДЕО



ВИДЕО

Последовательность кадров



Time

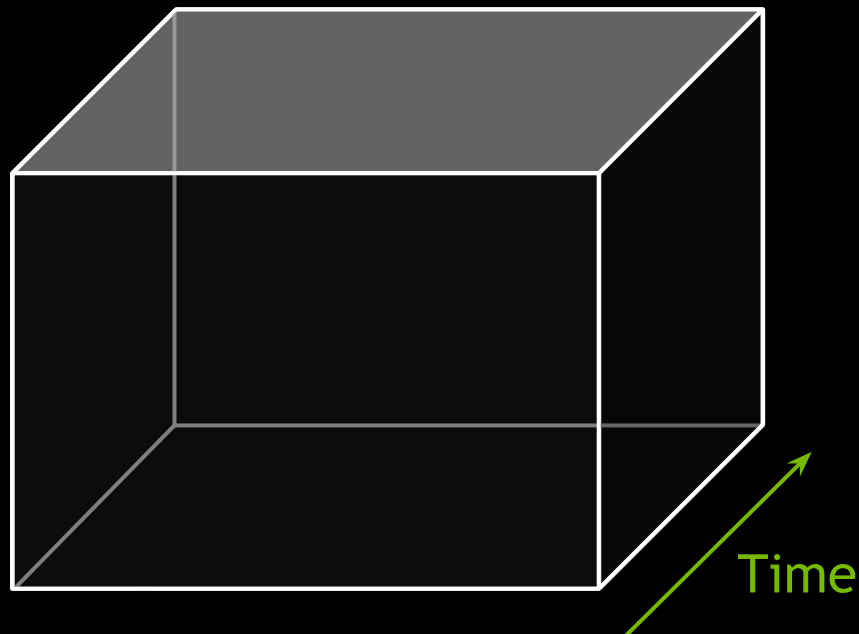
ВИДЕО

Последовательность кадров



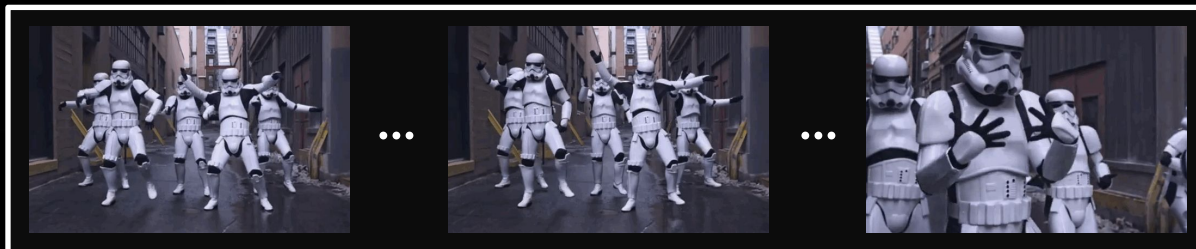
Time

Трёхмерный тензор



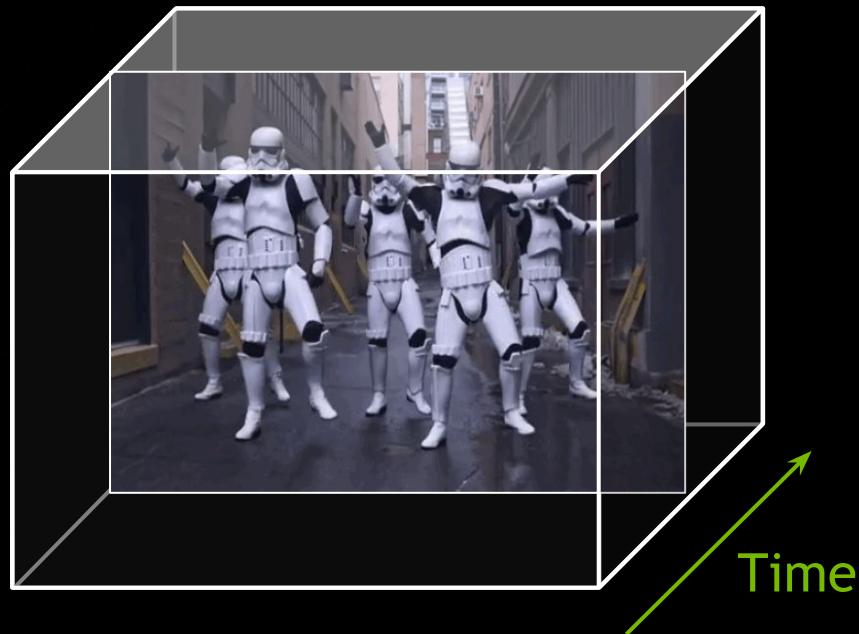
ВИДЕО

Последовательность кадров



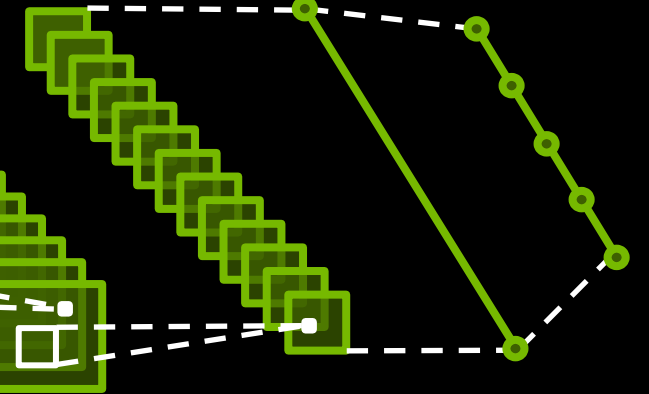
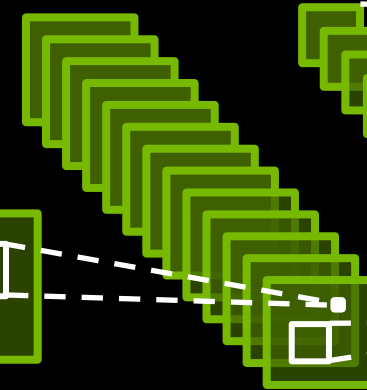
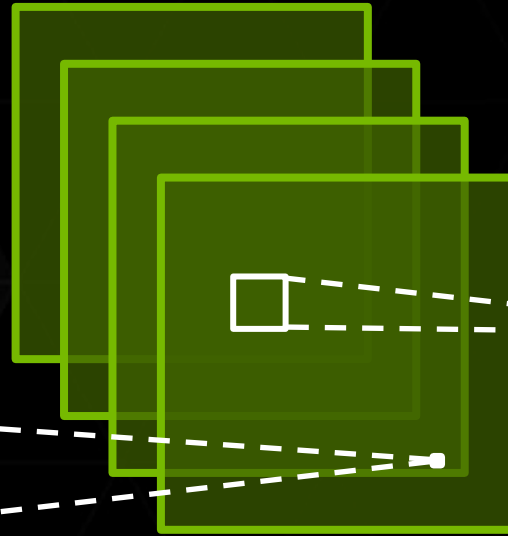
Time

Трёхмерный тензор



СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ (CNN)

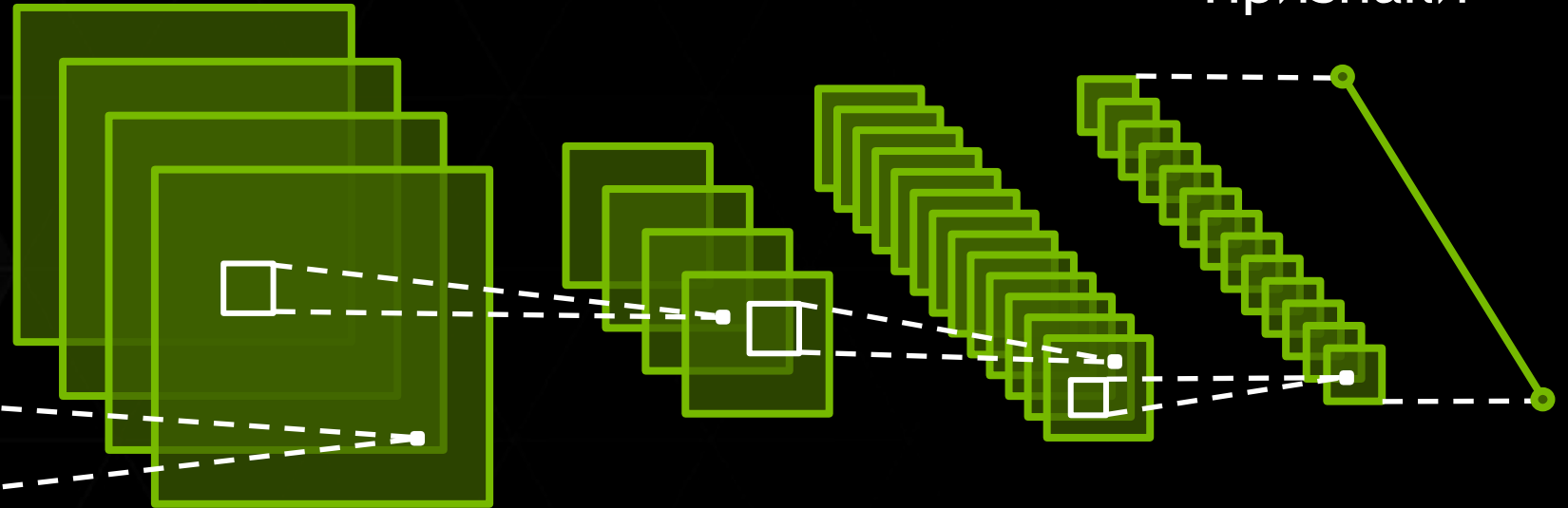
Изображение (кадр)



Предсказание

СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ (CNN)

Изображение (кадр)

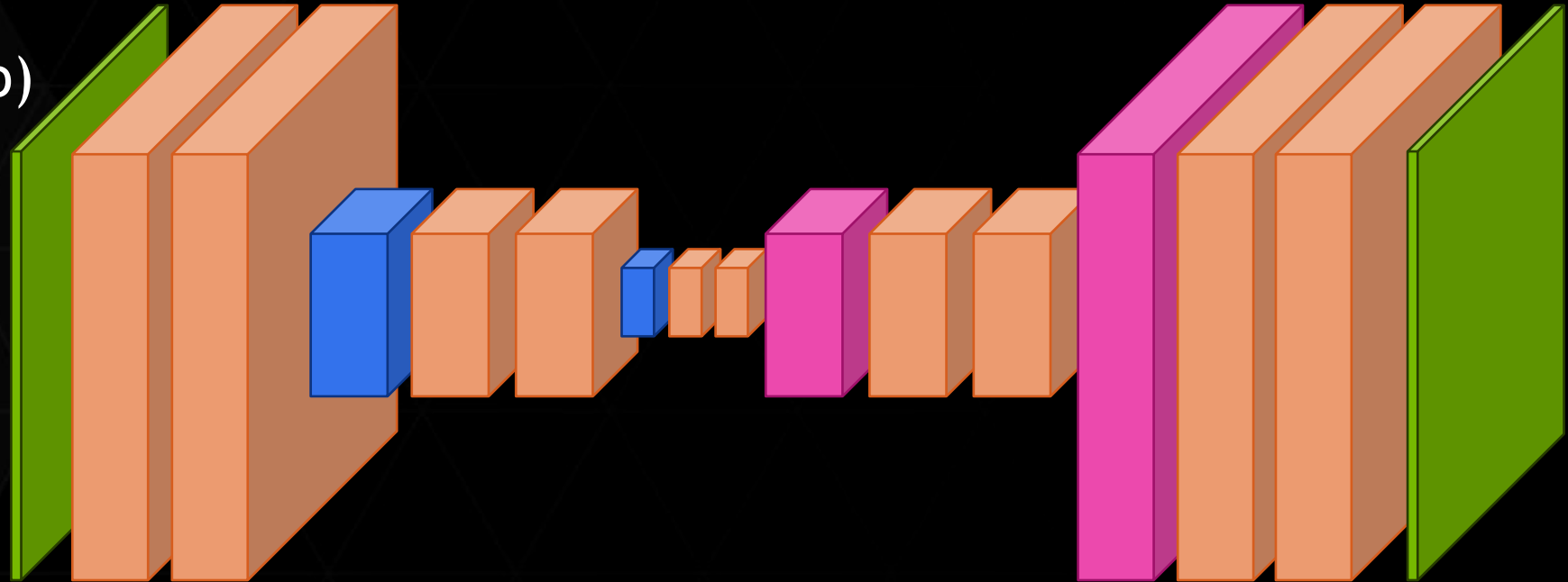


Признаки

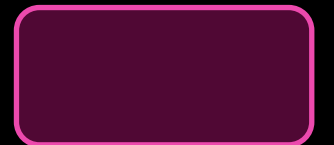
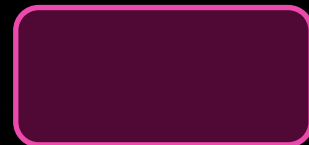
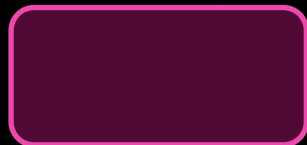
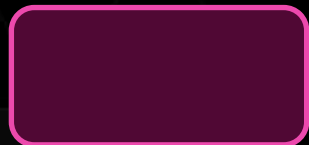
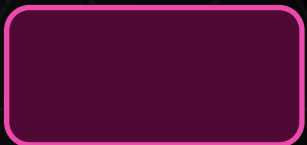
СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ (CNN)

Другое
изображение

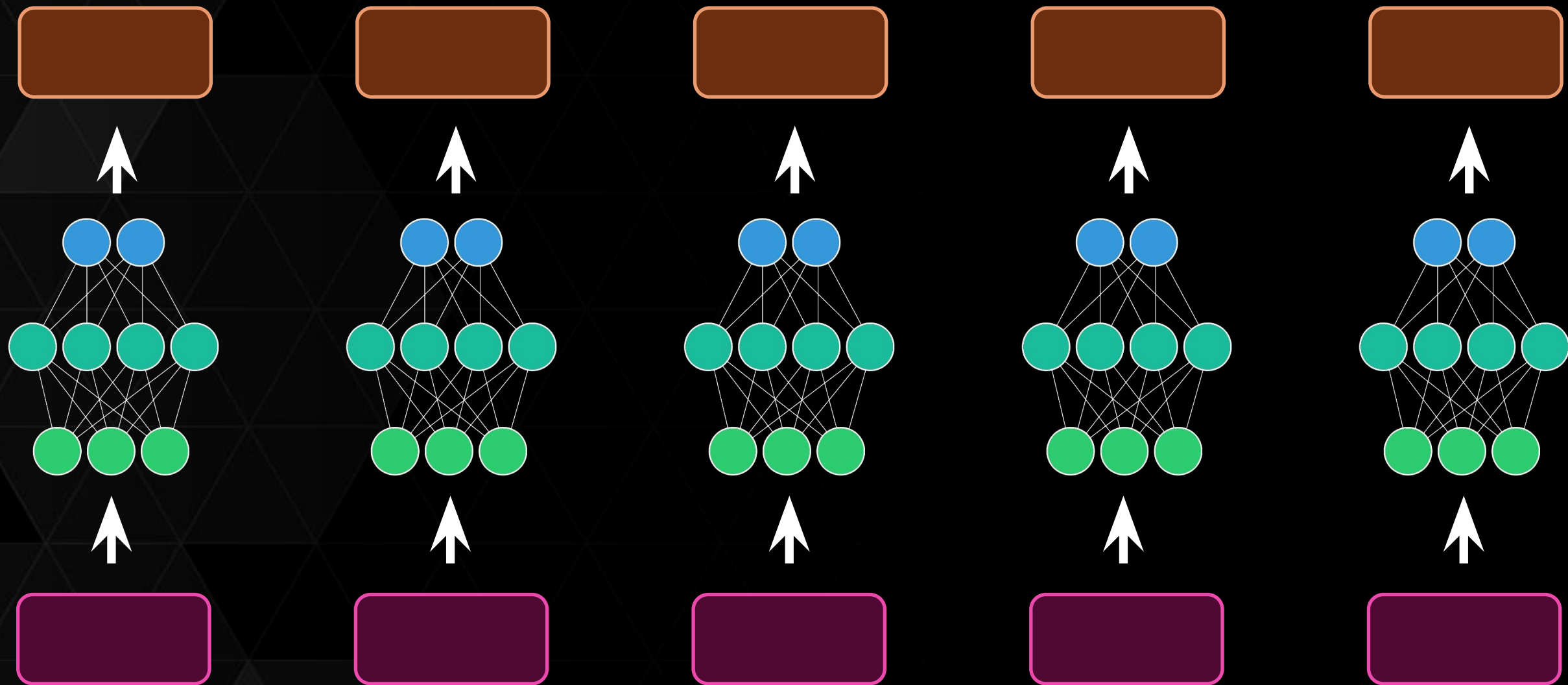
Изображение (кадр)



ОБРАБОТКА ПОСЛЕДОВАТЕЛЬНОСТЕЙ



ОБРАБОТКА ПОСЛЕДОВАТЕЛЬНОСТЕЙ



ПОКАДРОВАЯ ОБРАБОТКА



Кадр 1



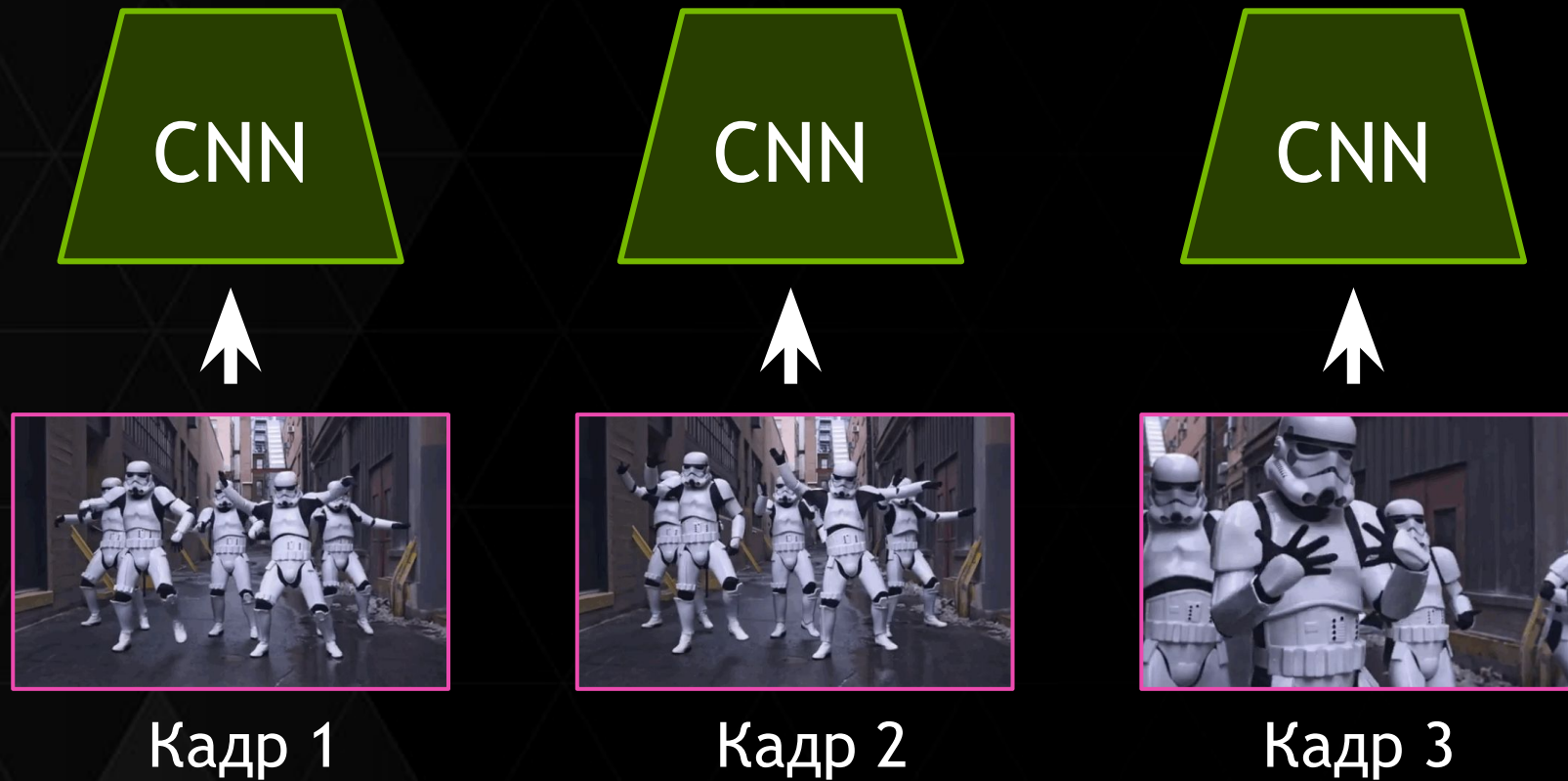
Кадр 2



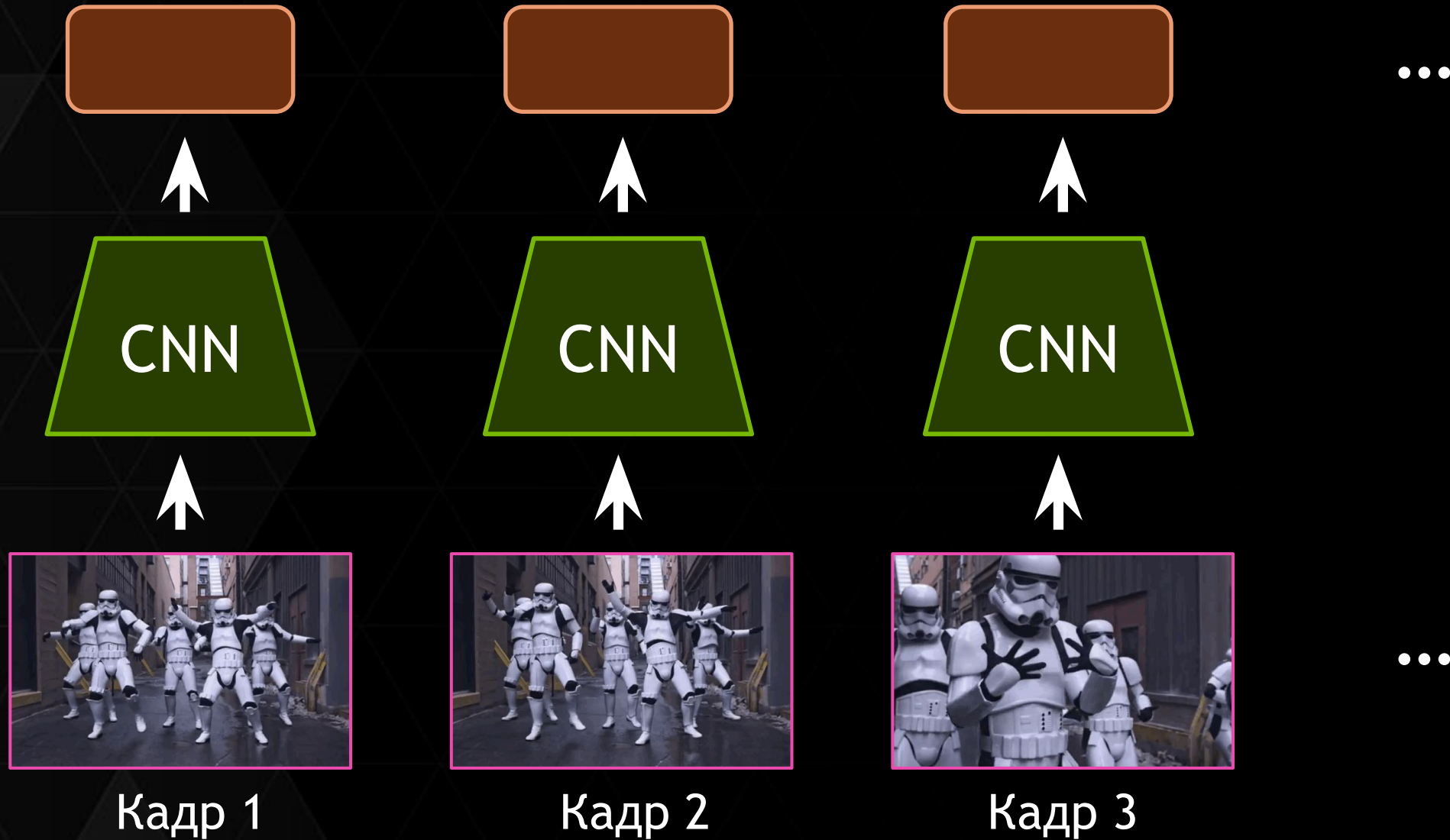
Кадр 3

...

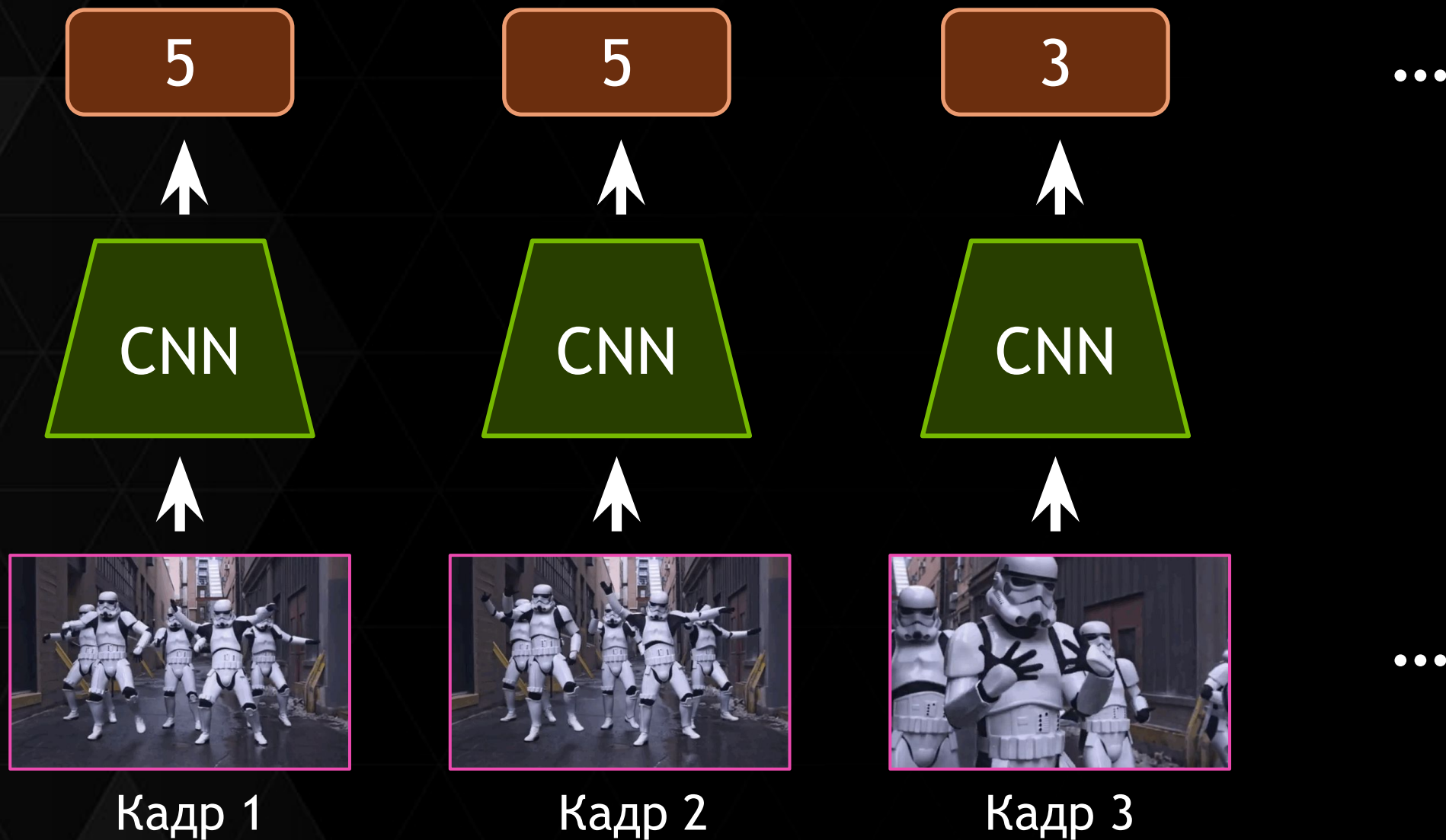
ПОКАДРОВАЯ ОБРАБОТКА



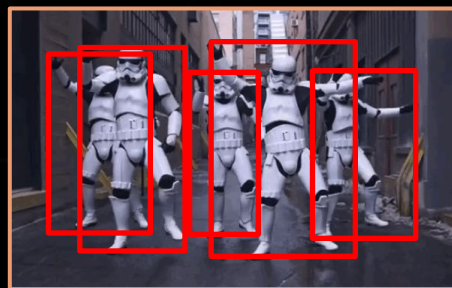
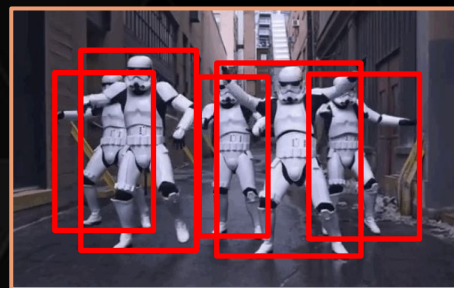
ПОКАДРОВАЯ ОБРАБОТКА



КЛАССИФИКАЦИЯ КАДРА



ДЕТЕКТИРОВАНИЕ ОБЪЕКТОВ



...



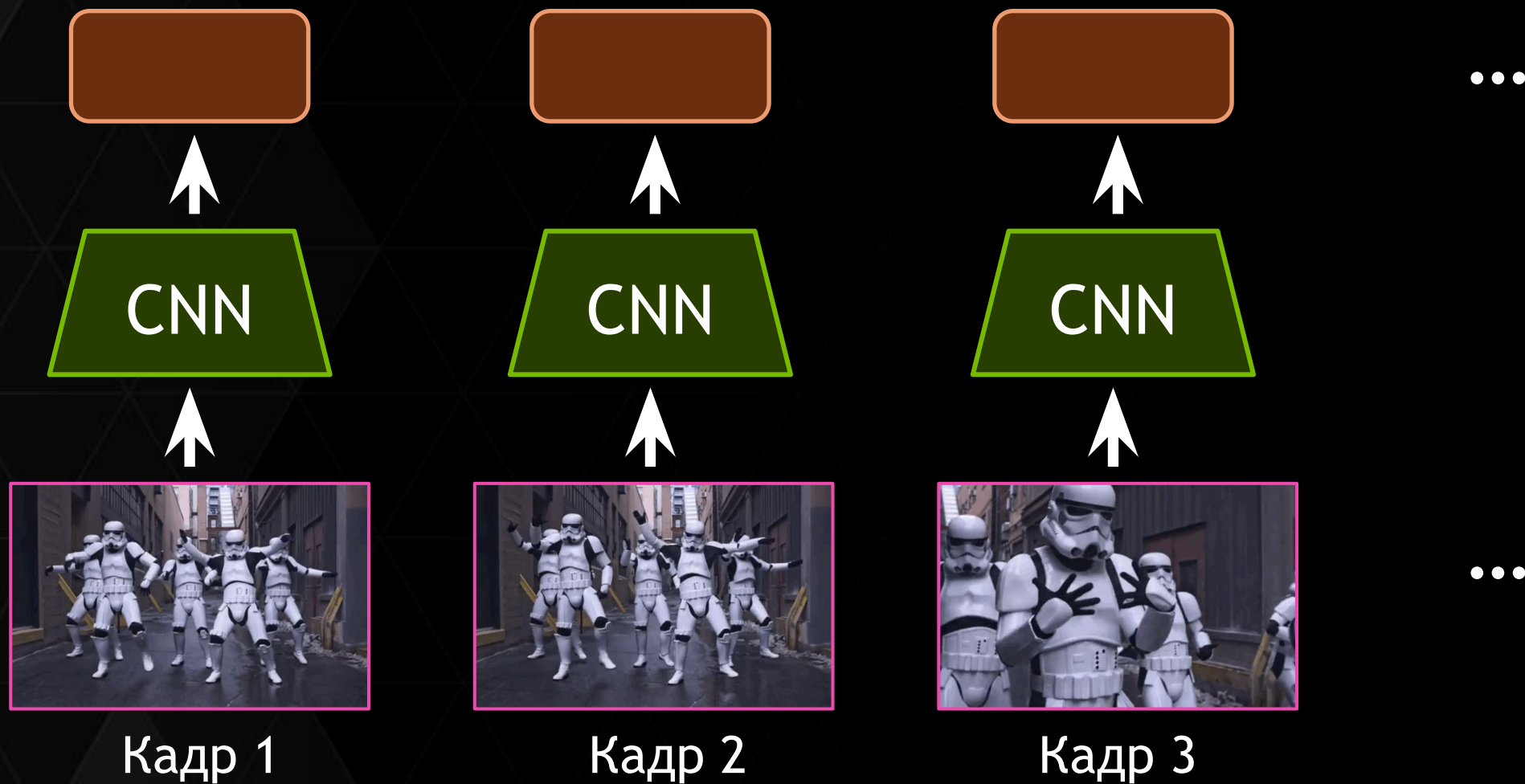
...

Кадр 1

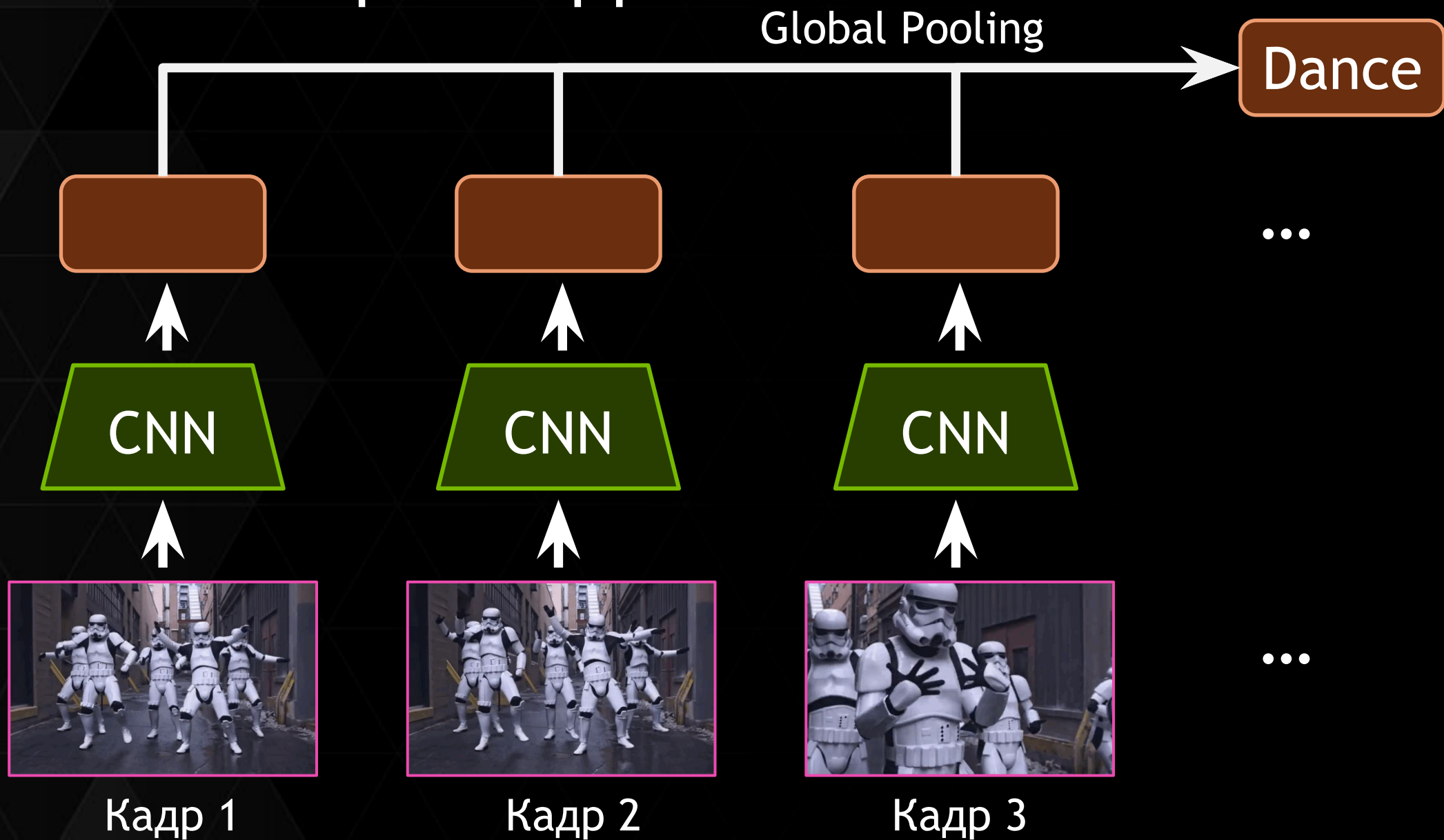
Кадр 2

Кадр 3

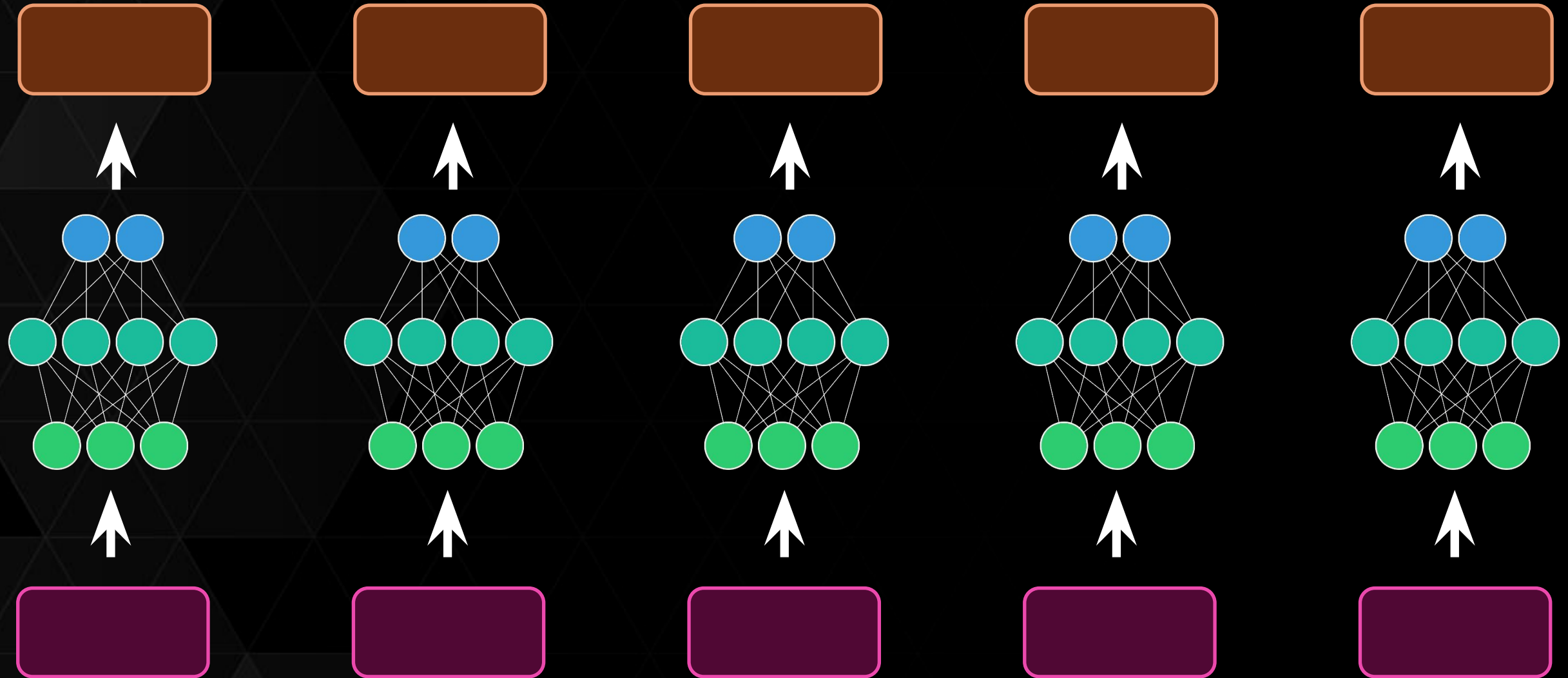
КЛАССИФИКАЦИЯ ВИДЕО



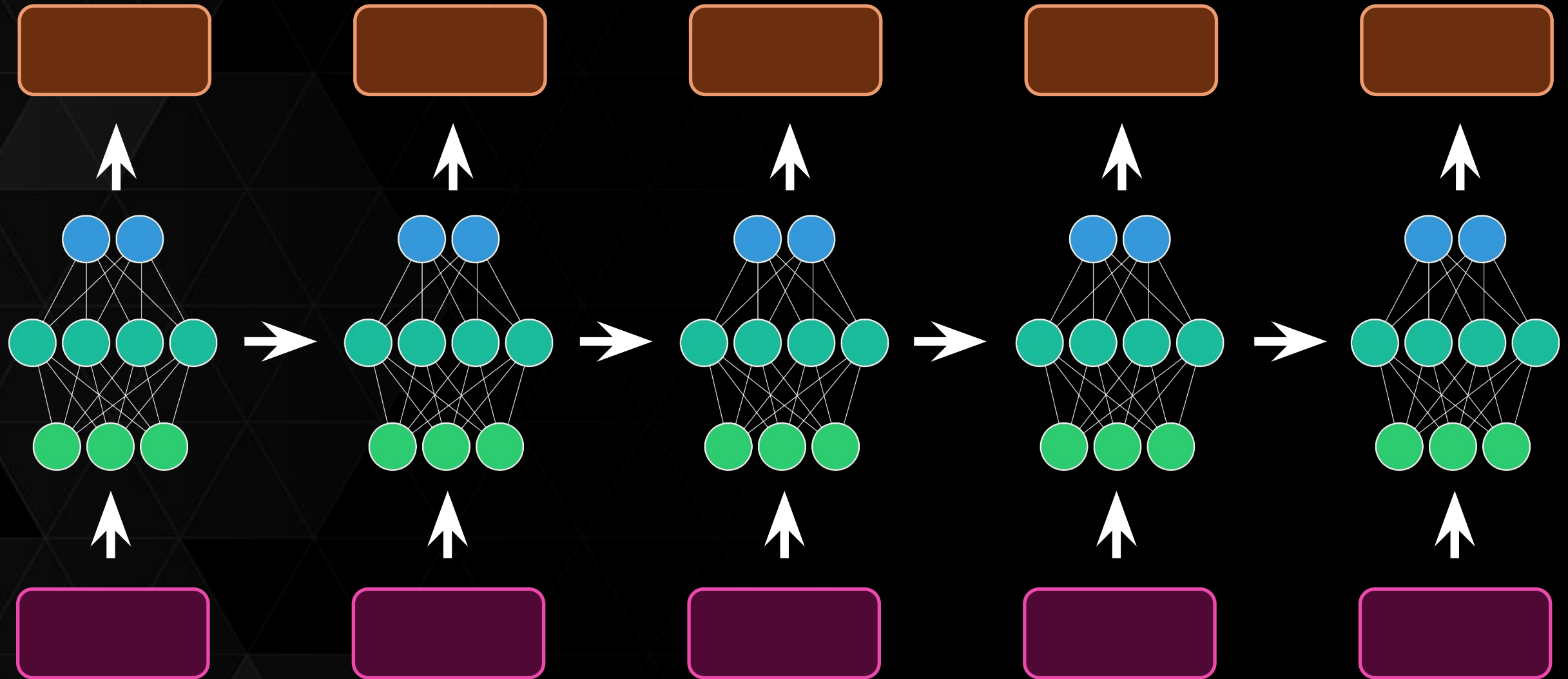
КЛАССИФИКАЦИЯ ВИДЕО



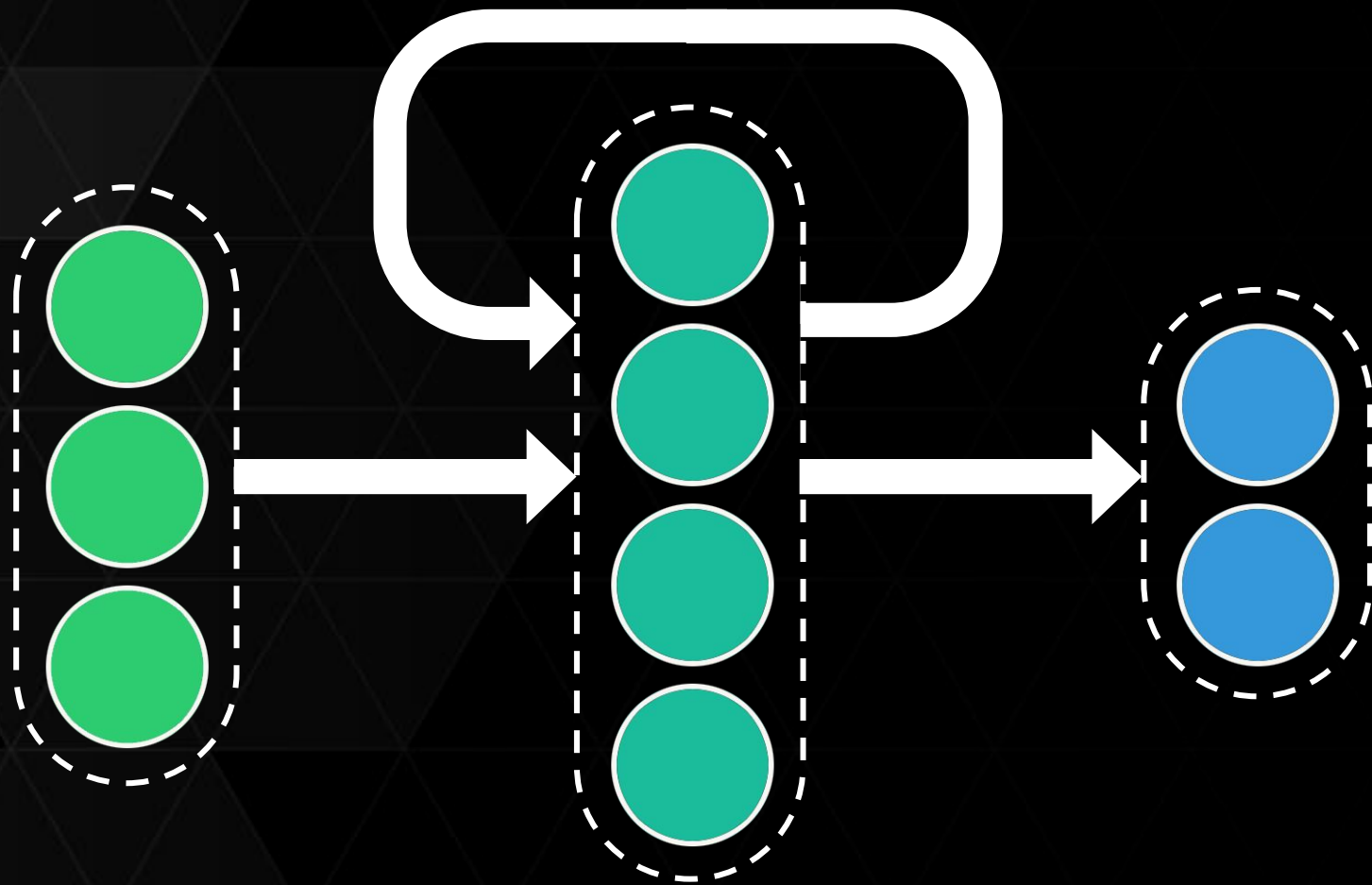
ОБРАБОТКА ПОСЛЕДОВАТЕЛЬНОСТЕЙ



РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ (RNN)

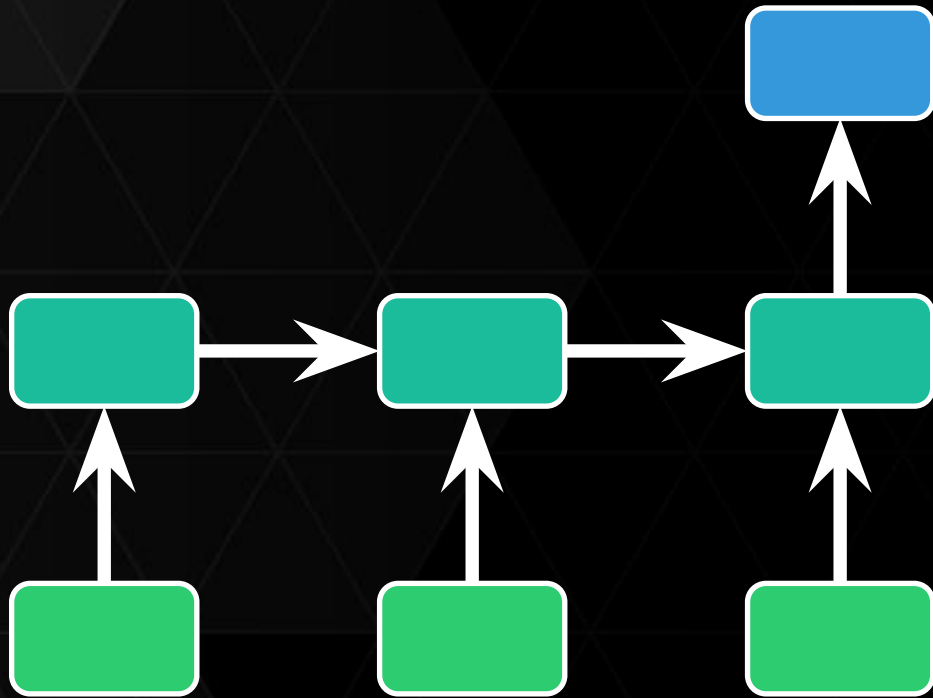


РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ

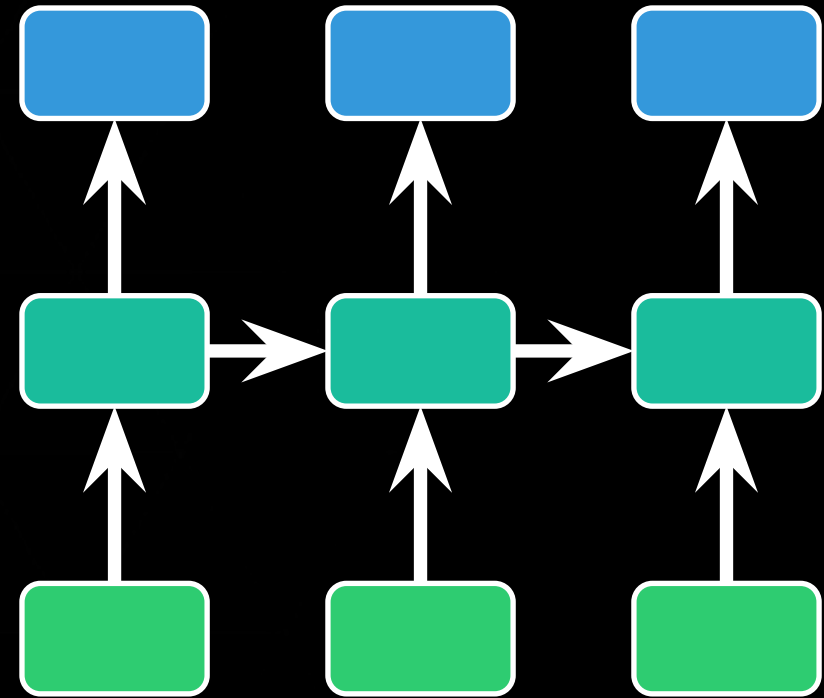


ТИПЫ ОБРАБОТКИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Many-to-One



Many-to-Many

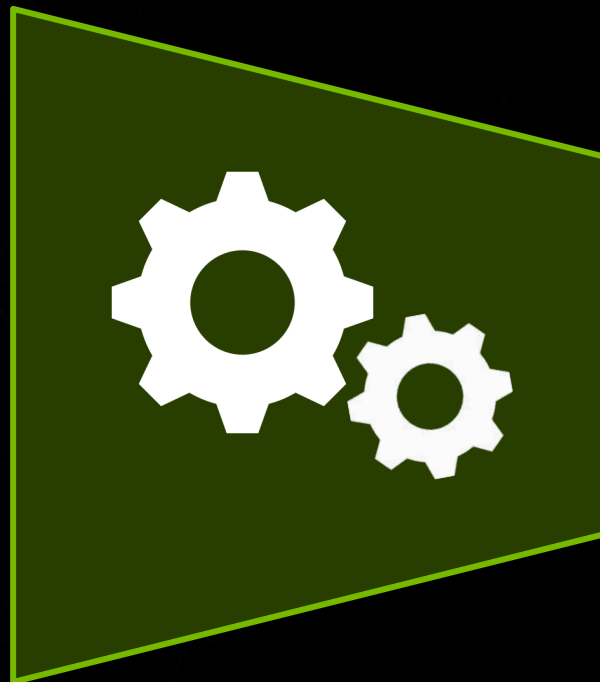


ПРИЗНАКИ ИЗОБРАЖЕНИЯ

Изображение



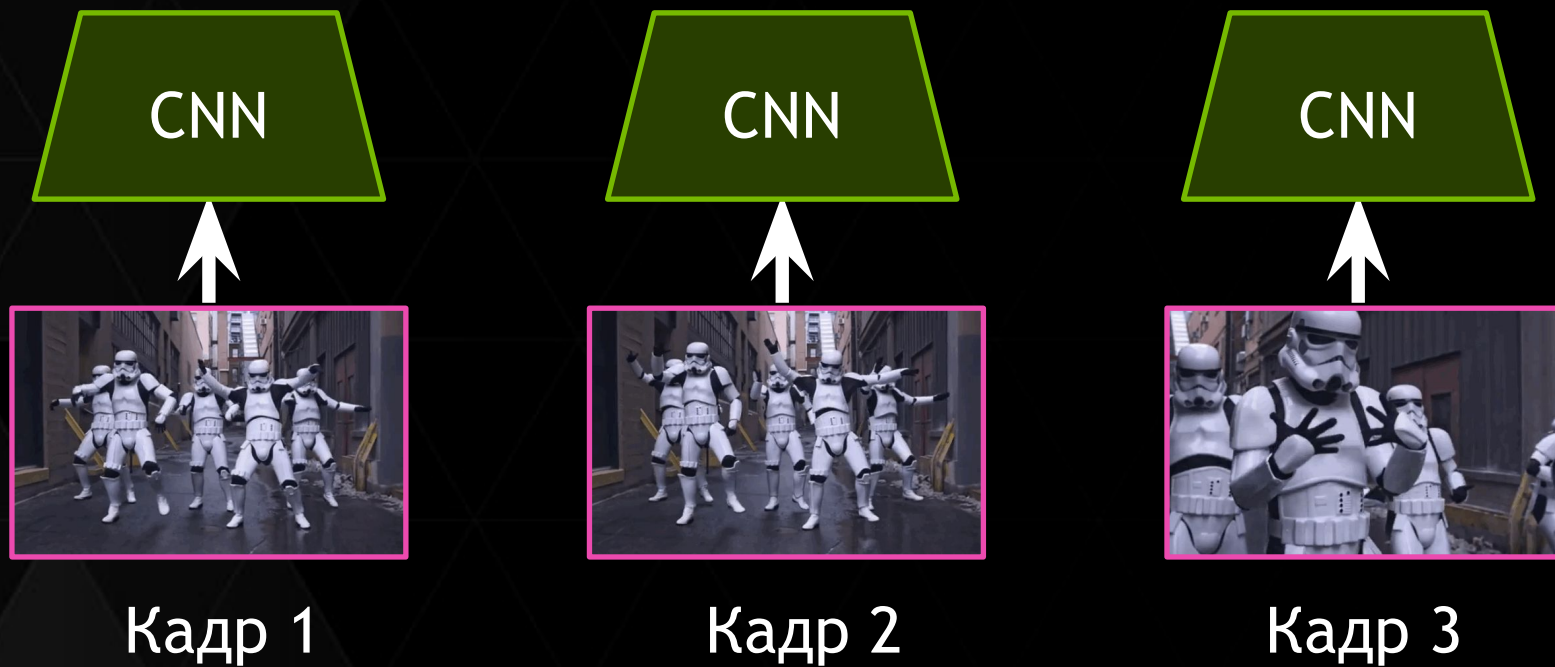
CNN



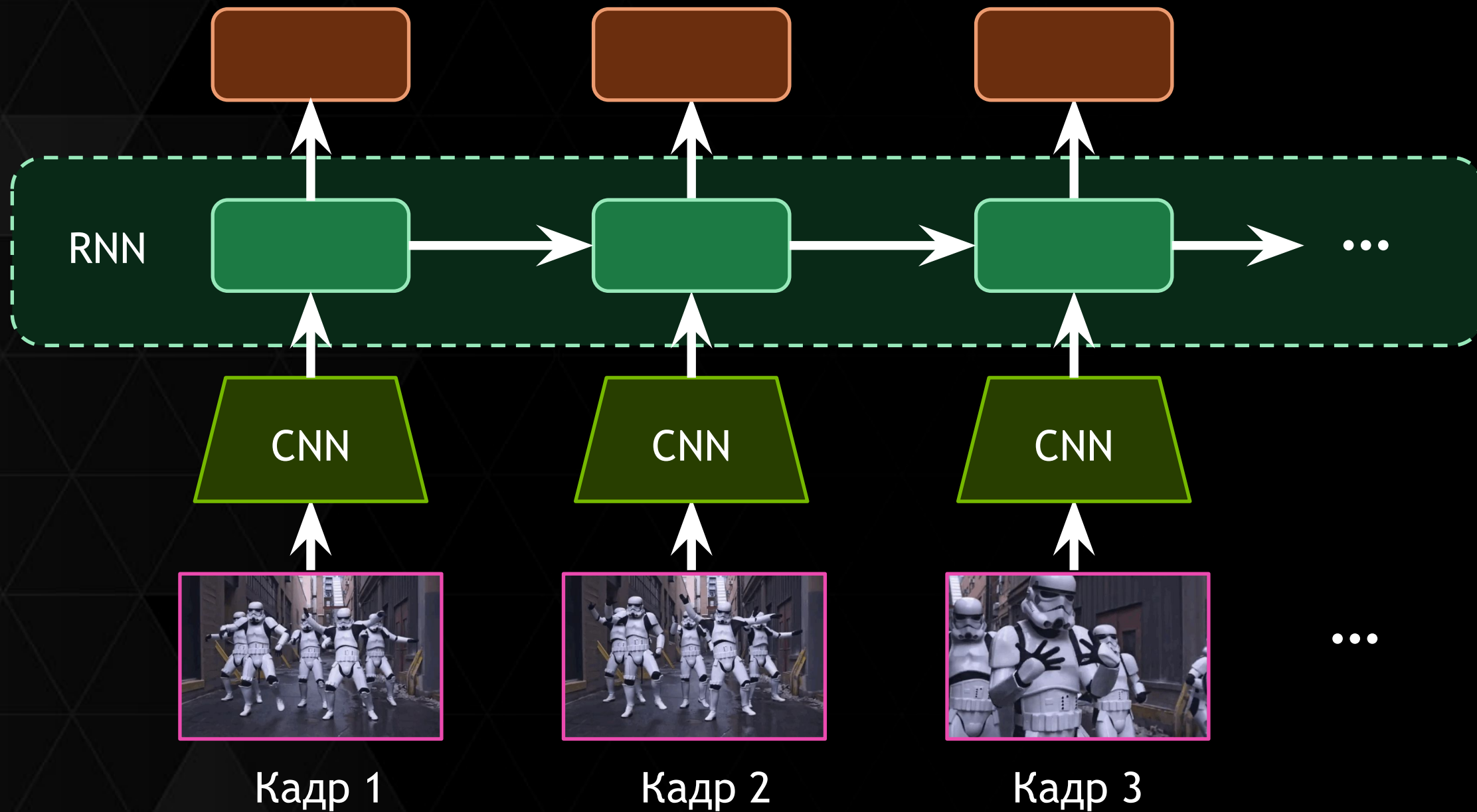
Эмбеддинг



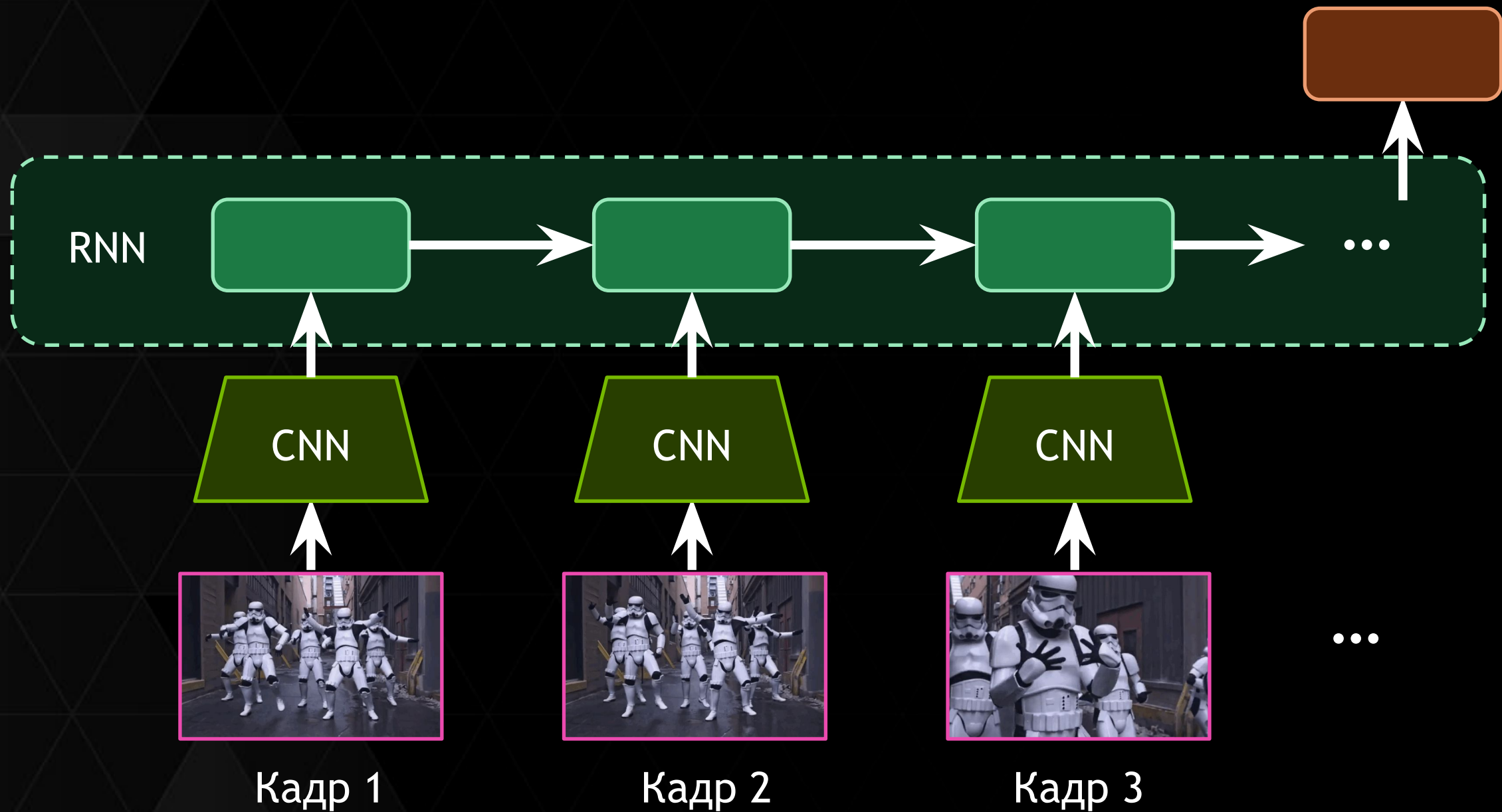
CNN + RNN



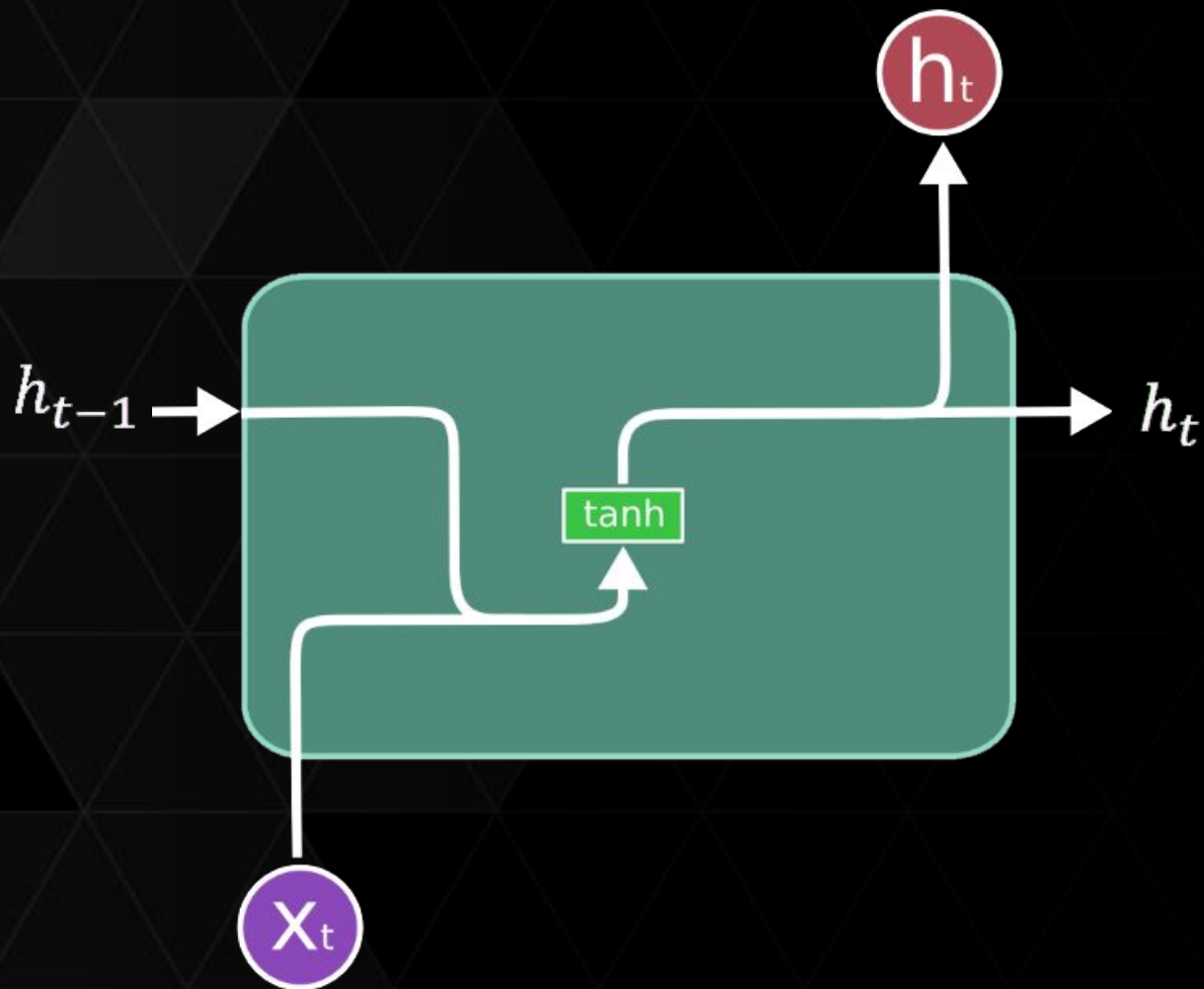
CNN + RNN



CNN + RNN

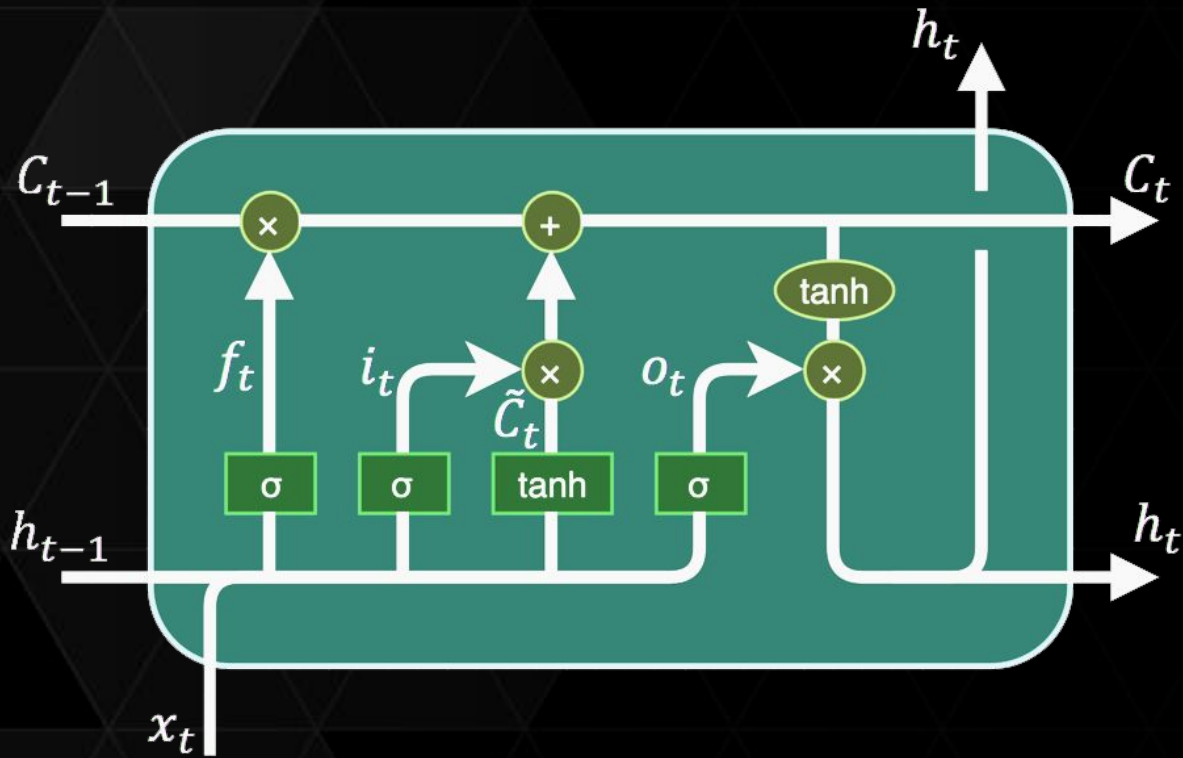


ПРОСТАЯ ЯЧЕЙКА RNN



$$h_t = \tanh(W_x x_t + W_h h_{t-1} + b)$$

LONG SHORT-TERM MEMORY (LSTM)



$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

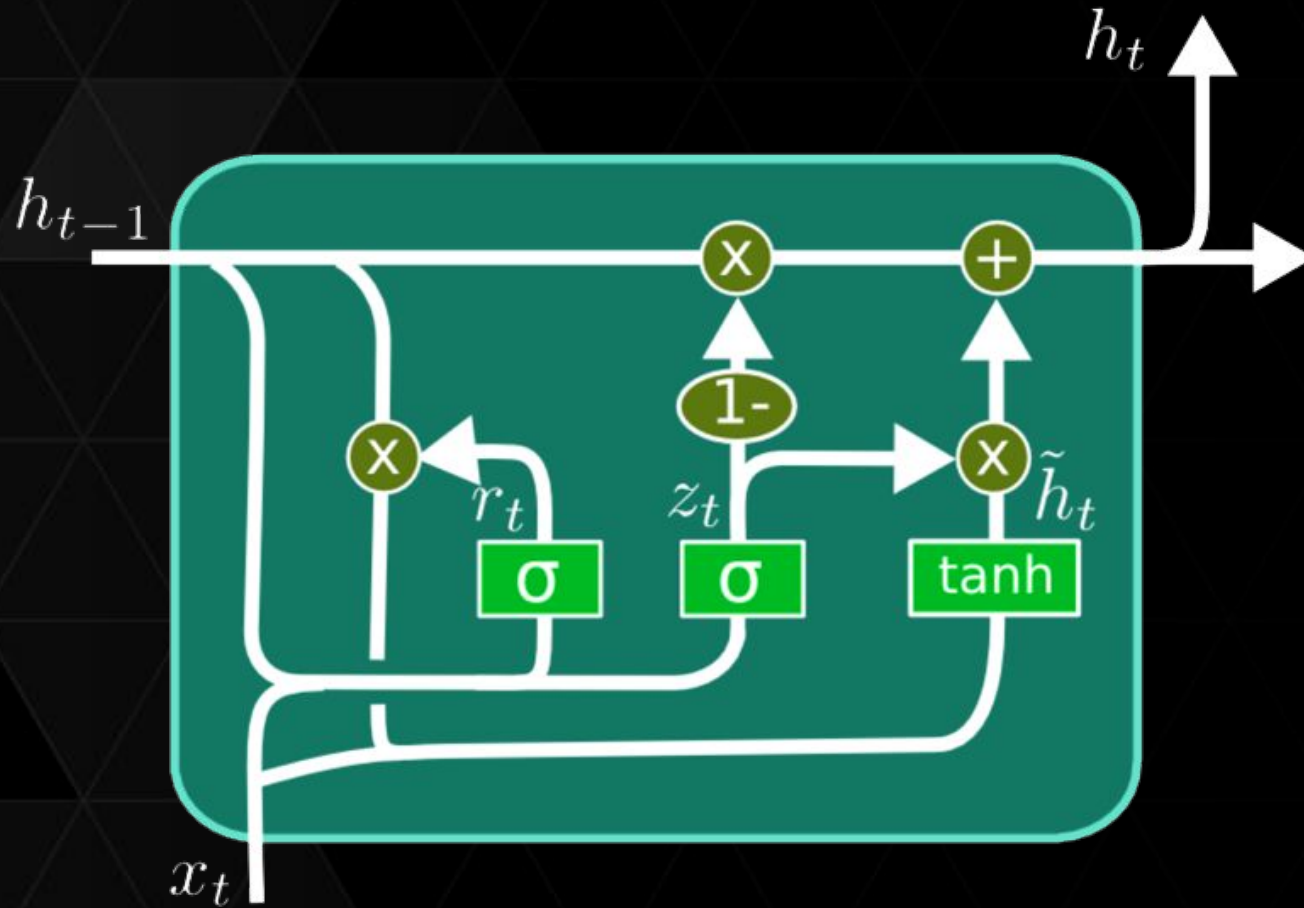
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$$

$$h_t = o_t \tanh(c_t)$$

GATED RECURRENT UNIT (GRU)



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

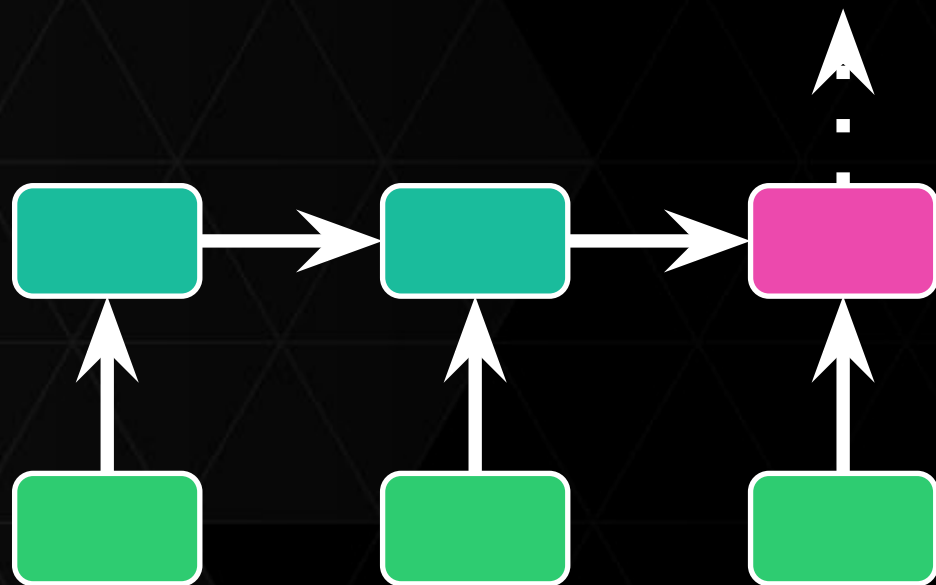
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

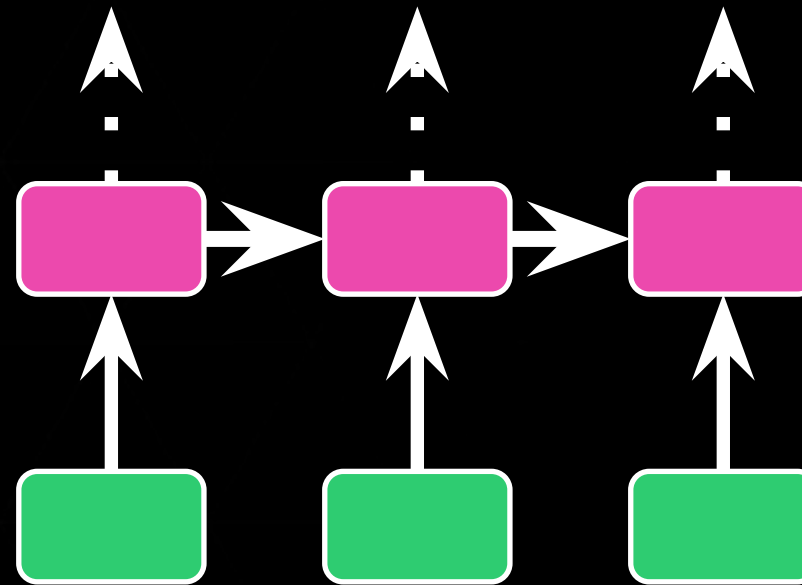
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

ОДНОНАПРАВЛЕННЫЕ RNN

Many-to-One

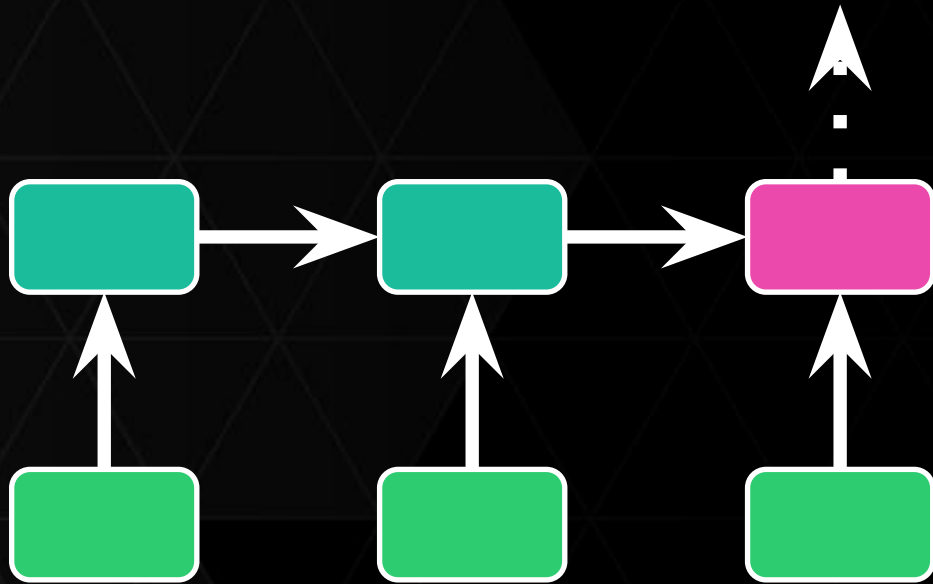


Many-to-Many



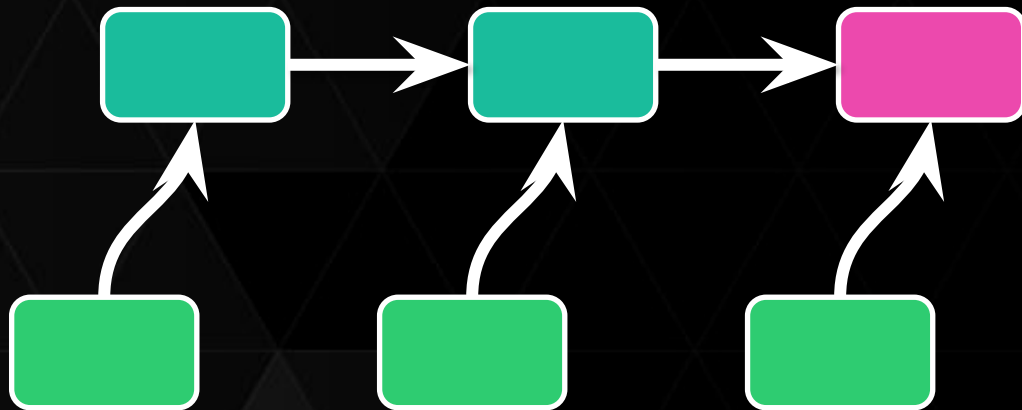
ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One



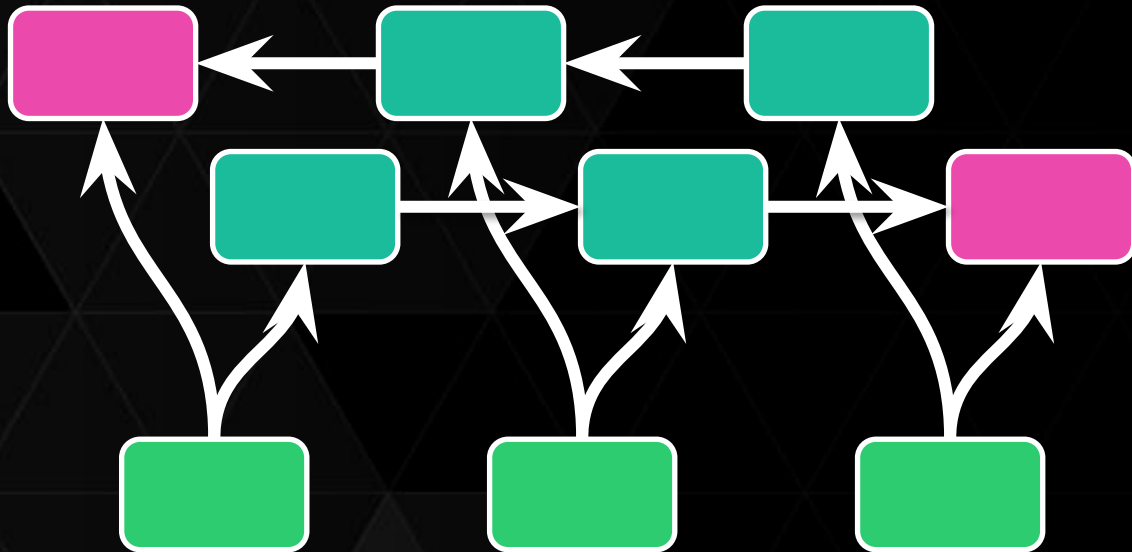
ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One



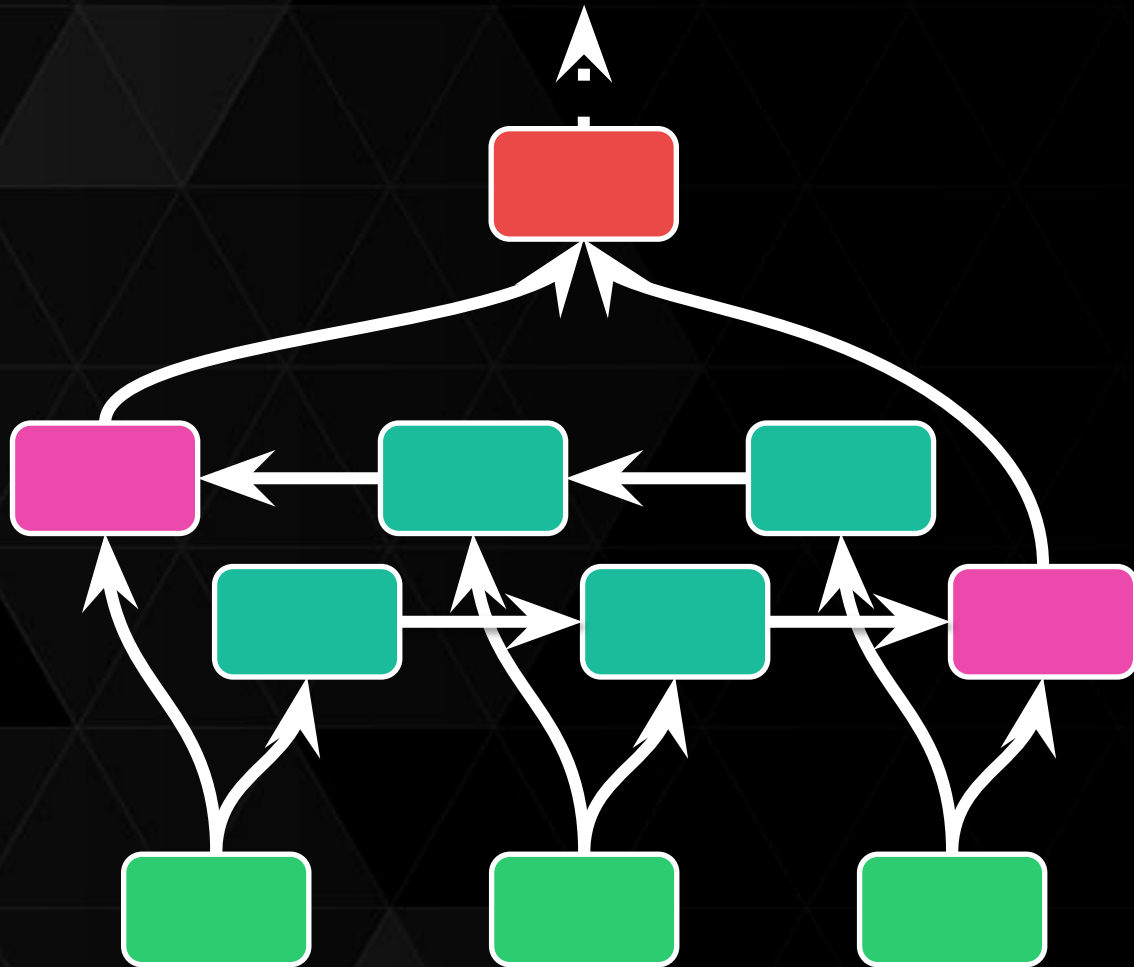
ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One



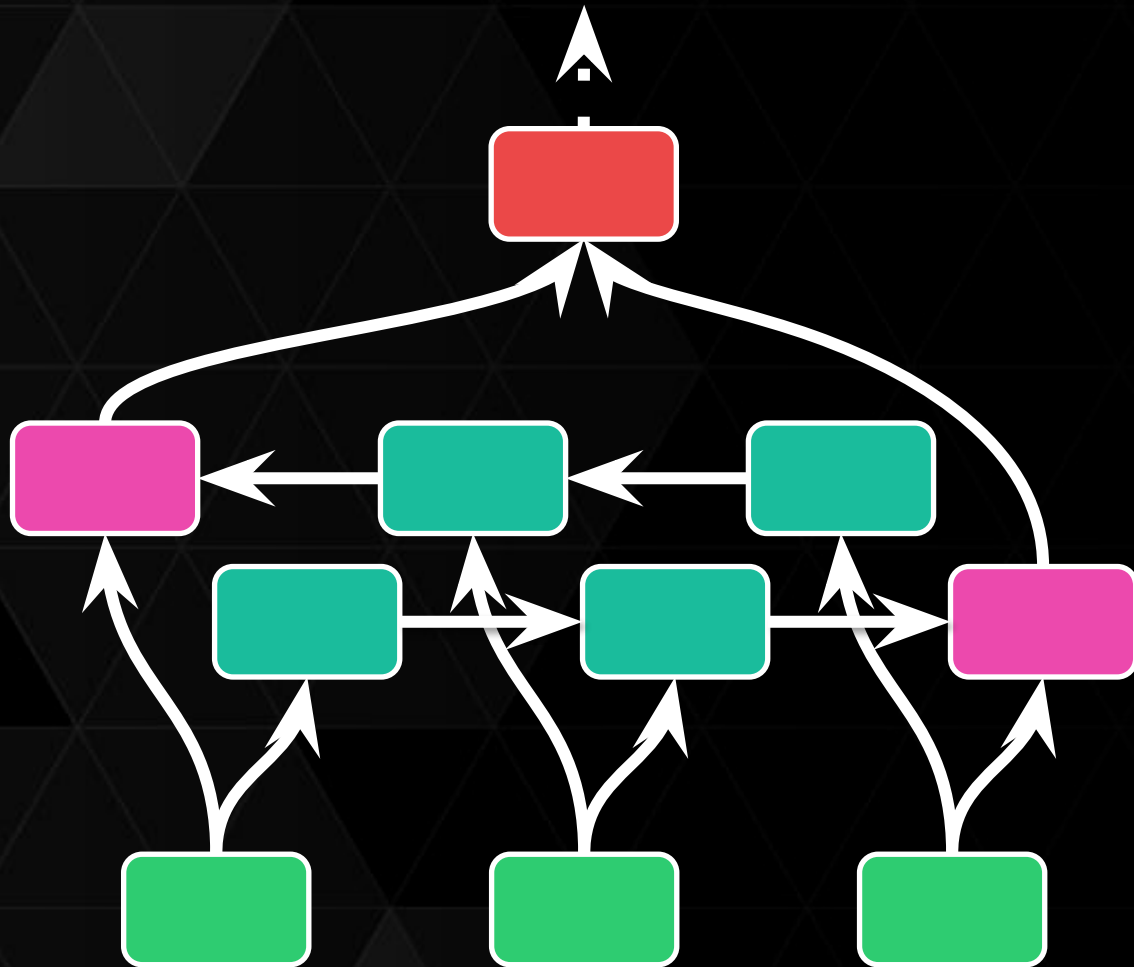
ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One

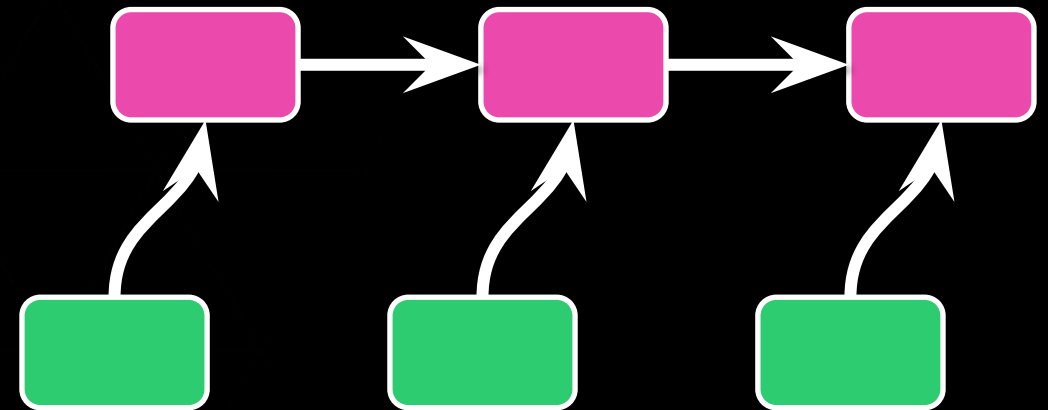


ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One

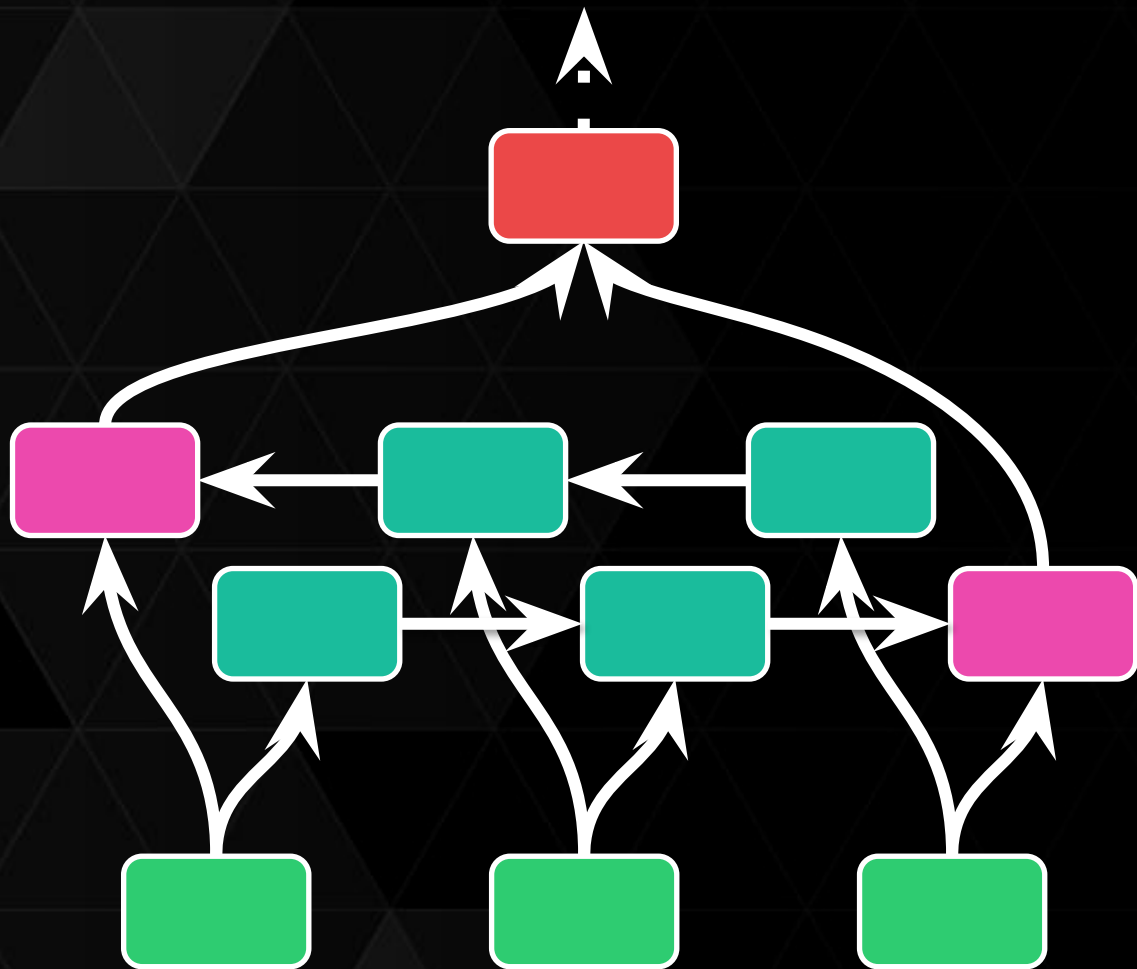


Many-to-Many

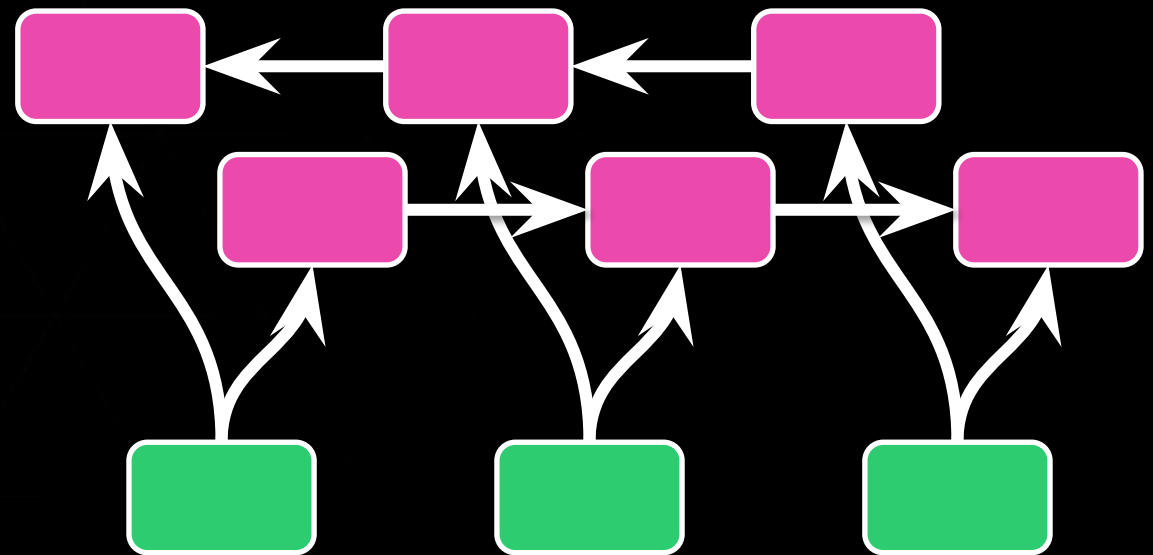


ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One

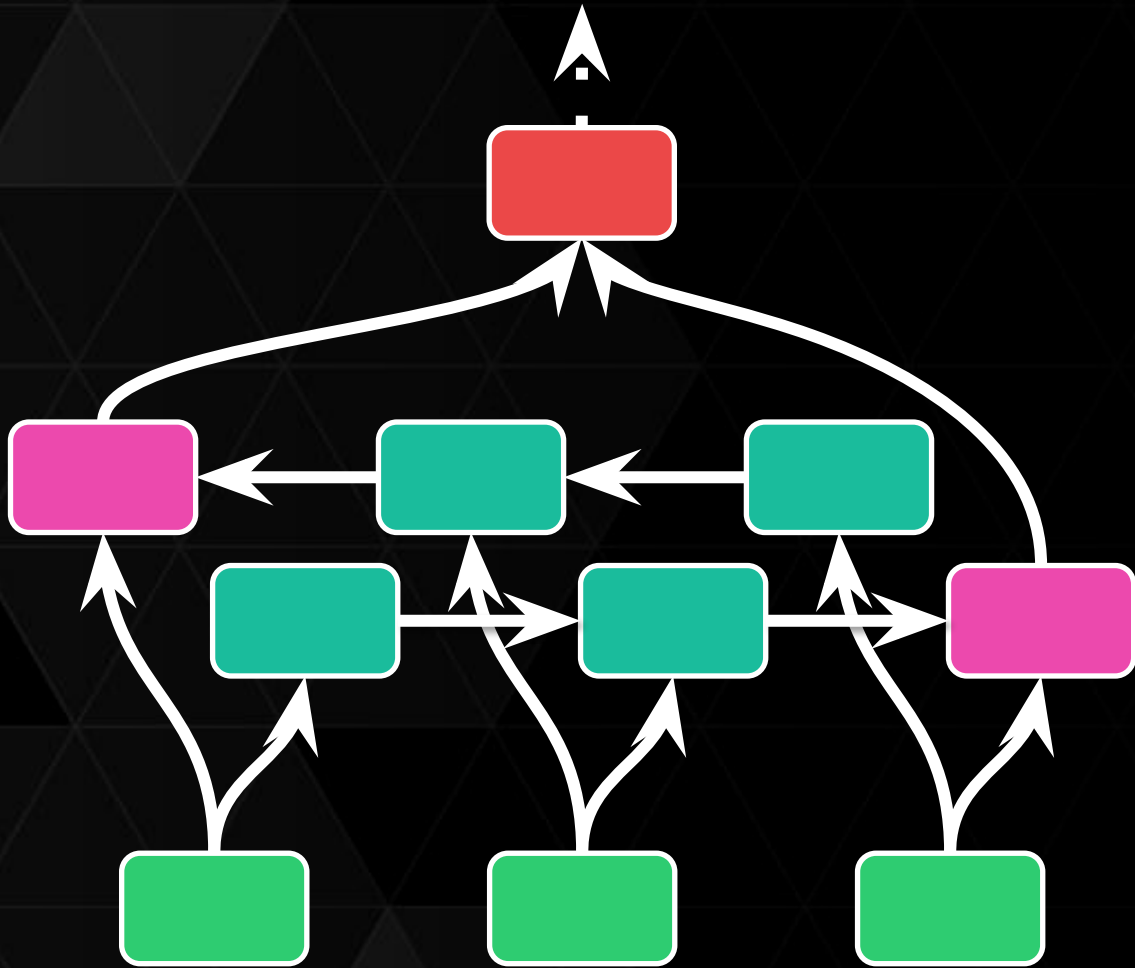


Many-to-Many

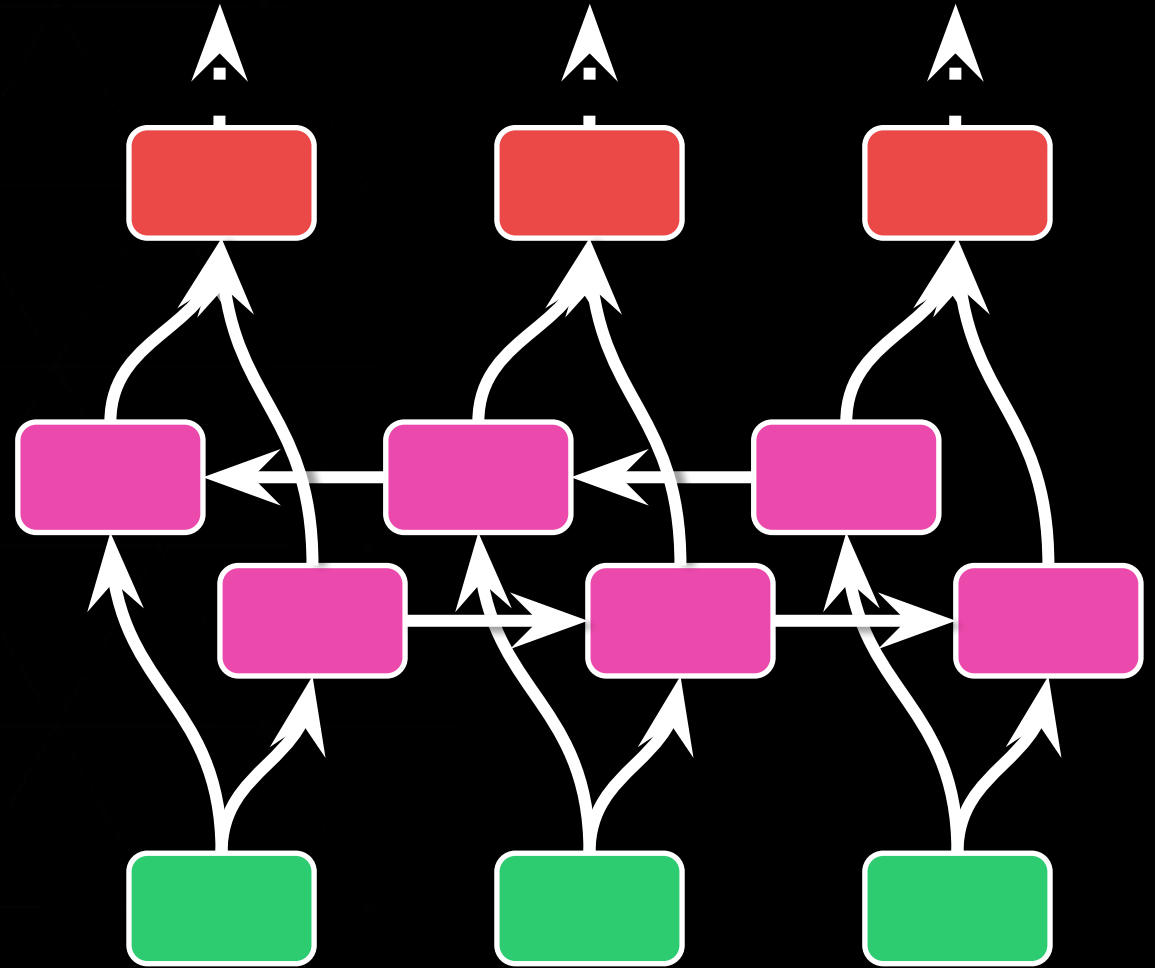


ДВУНАПРАВЛЕННЫЕ RNN (BIDIRECTIONAL)

Many-to-One



Many-to-Many



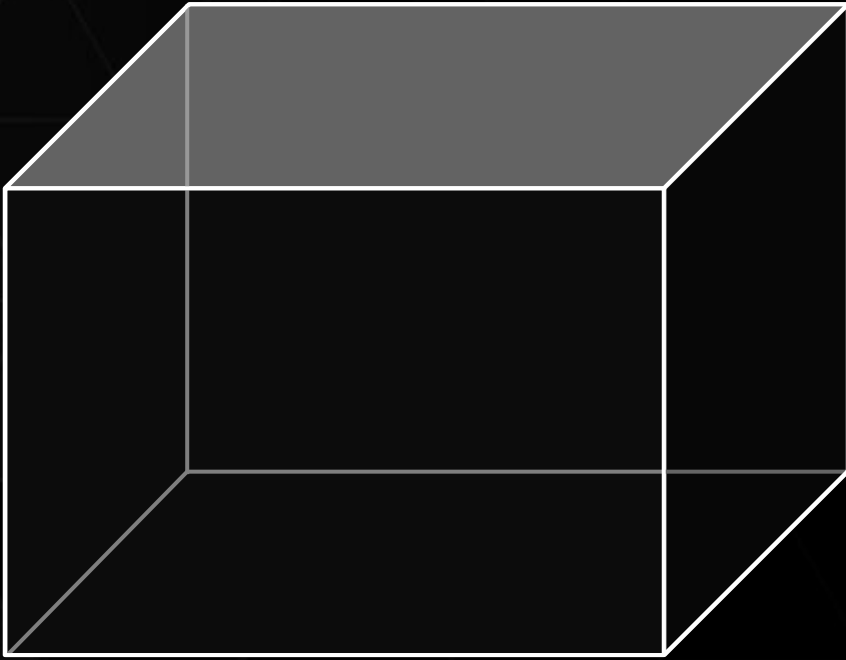
ТРЁХМЕРНЫЕ СВЁРТОЧНЫЕ СЕТИ

Видео как трёхмерный тензор



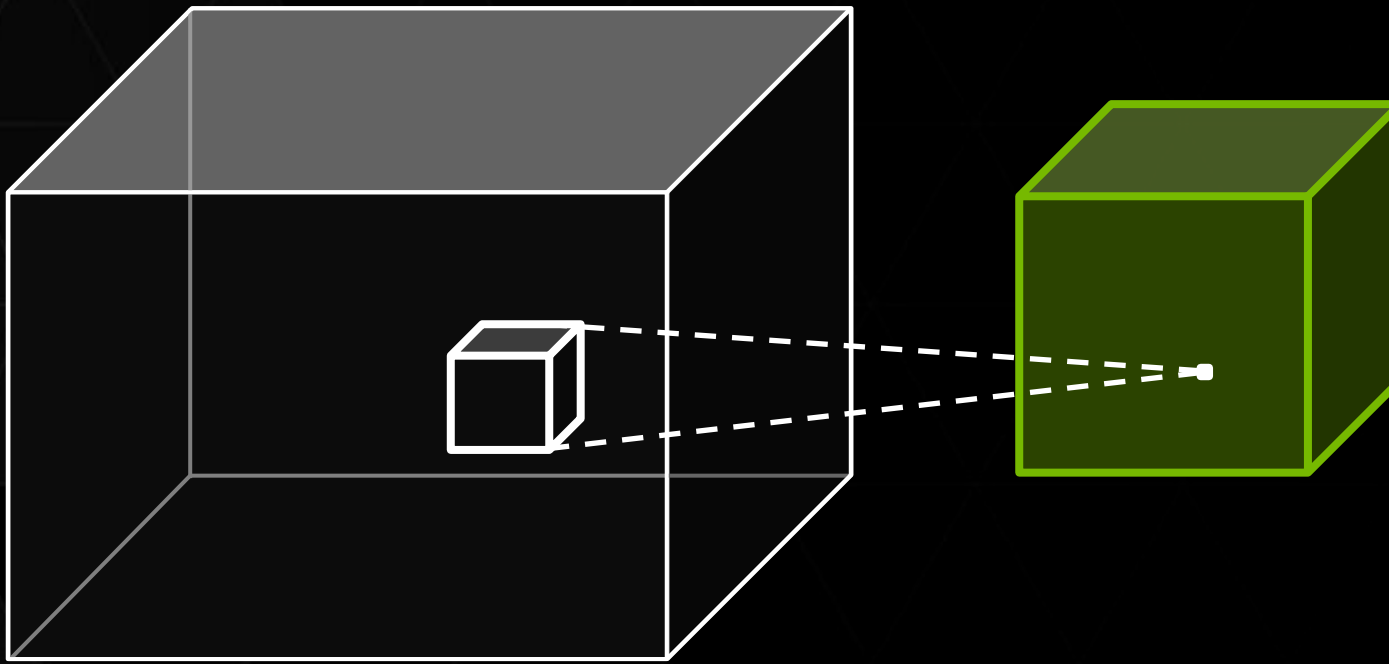
ТРЁХМЕРНЫЕ СВЁРТОЧНЫЕ СЕТИ

Видео как трёхмерный тензор



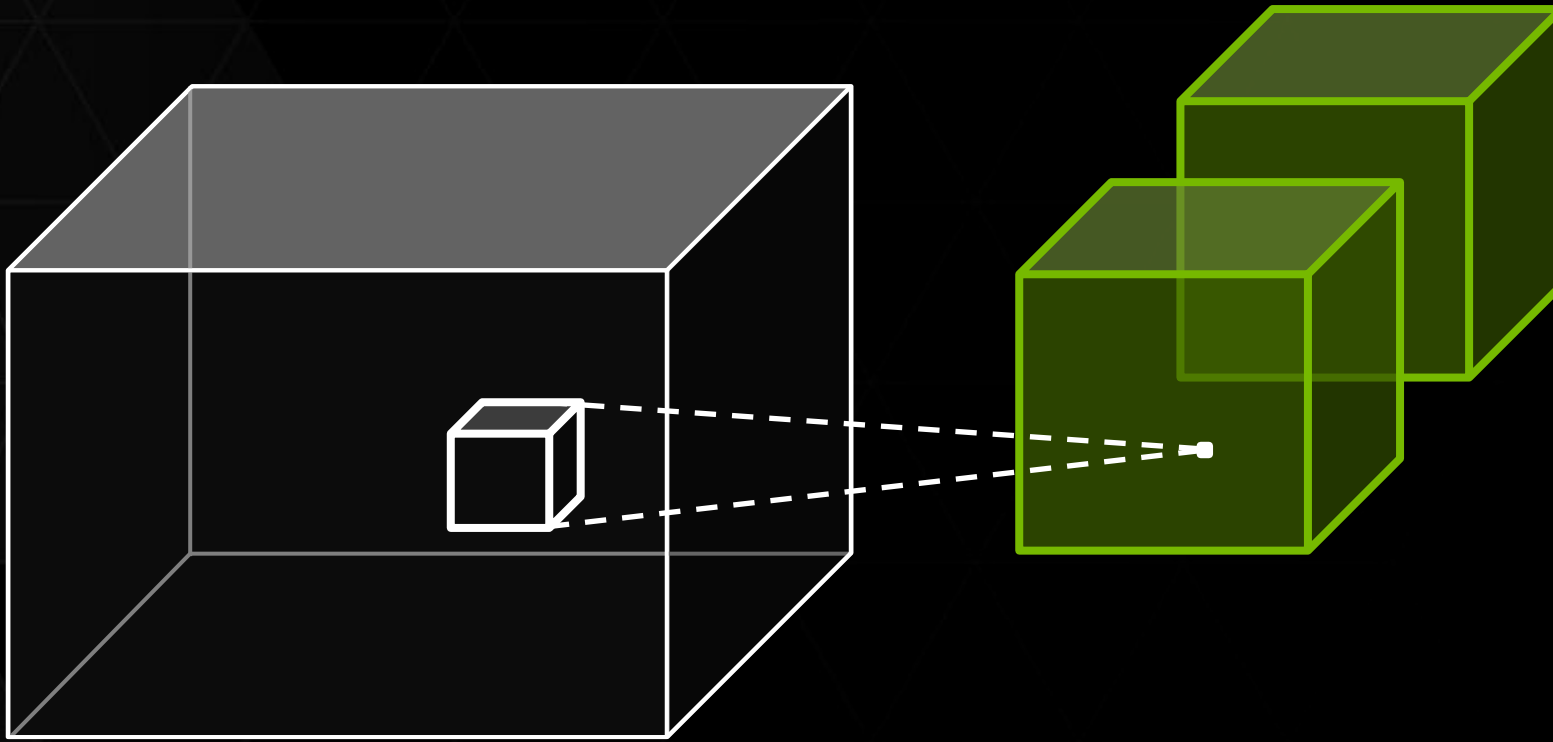
ТРЁХМЕРНЫЕ СВЁРТОЧНЫЕ СЕТИ

3D Convolution



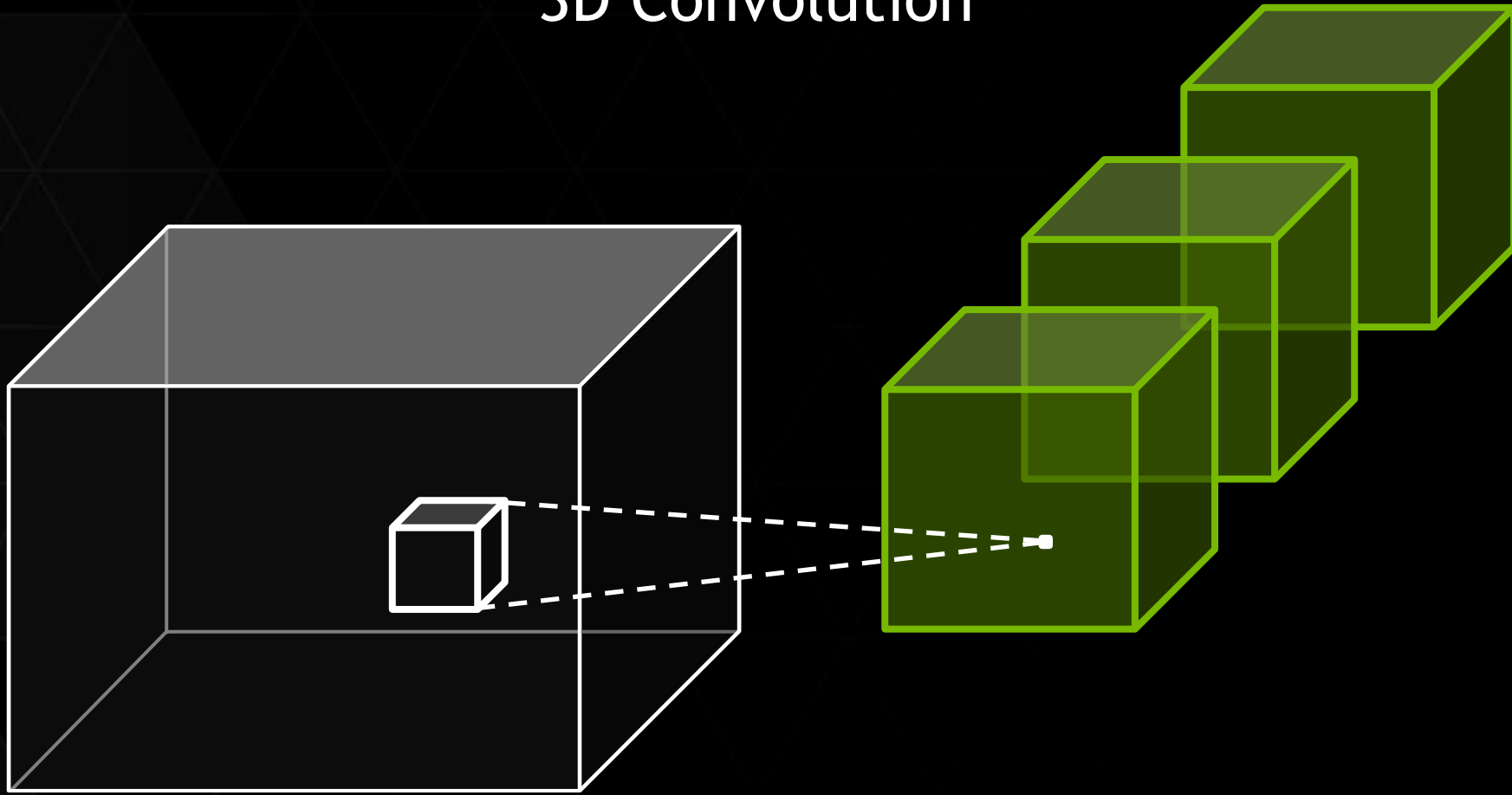
ТРЁХМЕРНЫЕ СВЁРТОЧНЫЕ СЕТИ

3D Convolution



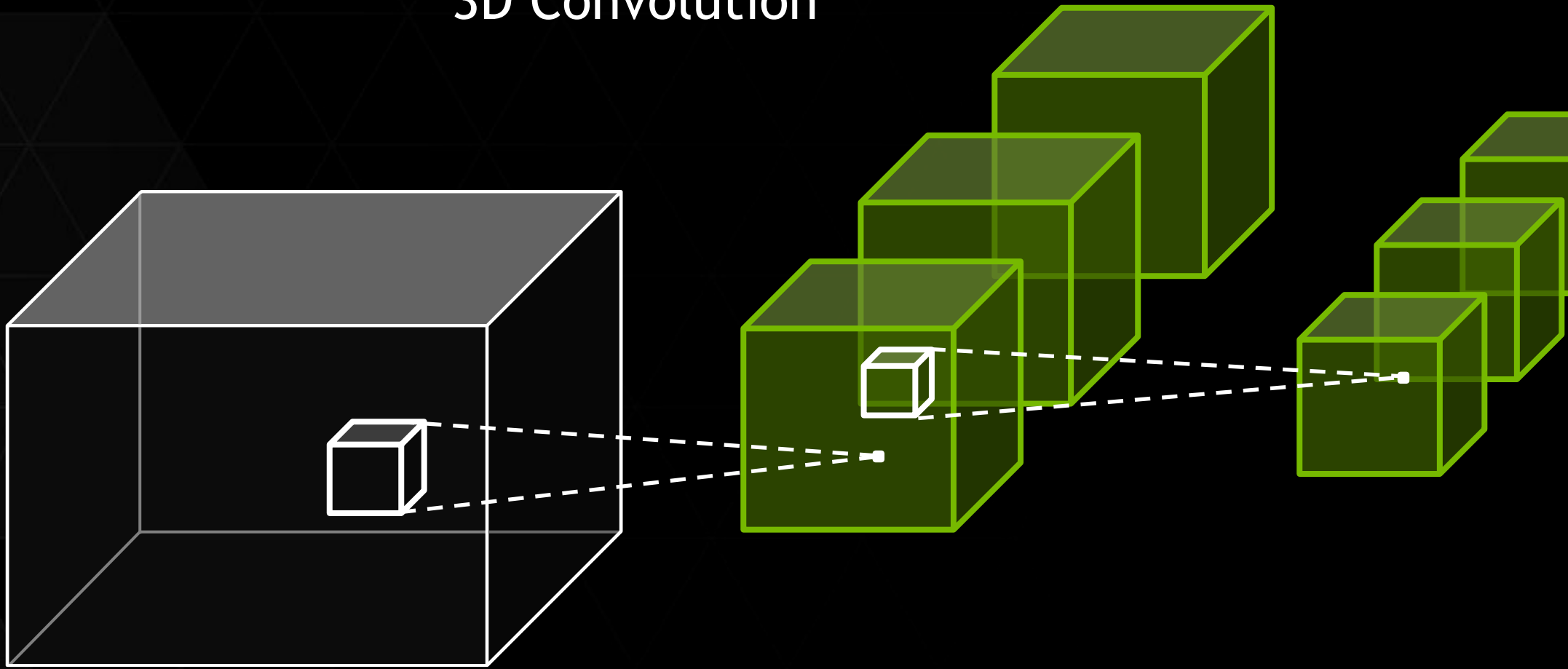
ТРЁХМЕРНЫЕ СВЁРТОЧНЫЕ СЕТИ

3D Convolution



ТРЁХМЕРНЫЕ СВЁРТОЧНЫЕ СЕТИ

3D Convolution



ВЕКТОРА ДВИЖЕНИЯ / ОПТИЧЕСКИЙ ПОТОК



ВЕКТОРА ДВИЖЕНИЯ / ОПТИЧЕСКИЙ ПОТОК

Sparse

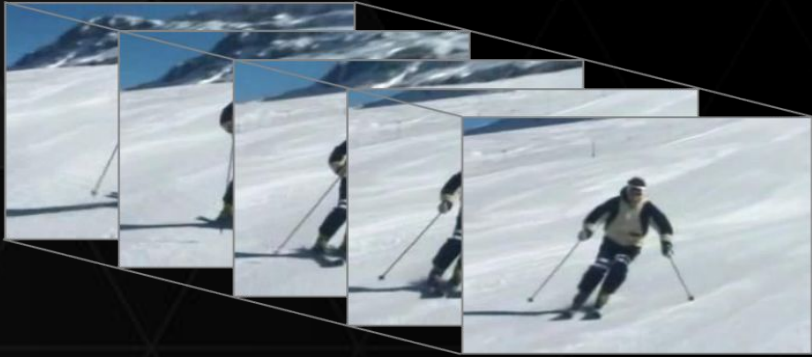


Dense



ДВУХКАНАЛЬНАЯ ОБРАБОТКА (TWO-STREAM)

RGB кадры

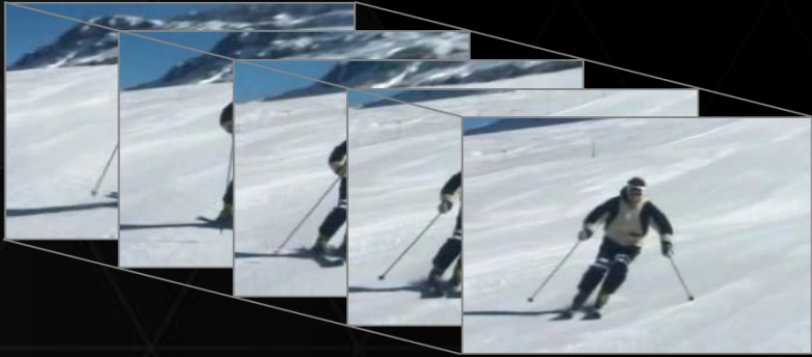


CNN / RNN



ДВУХКАНАЛЬНАЯ ОБРАБОТКА (TWO-STREAM)

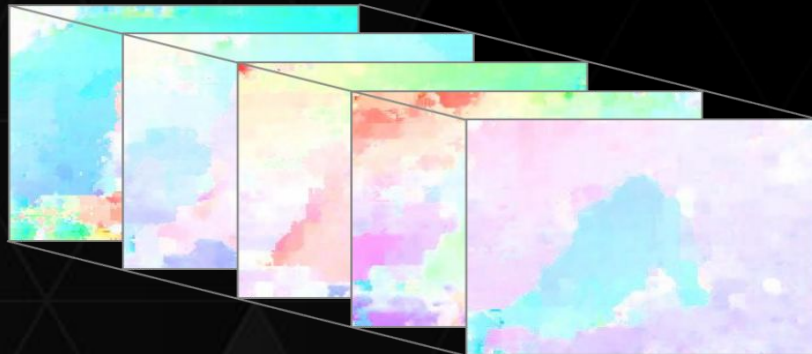
RGB кадры



CNN / RNN



Оптический поток

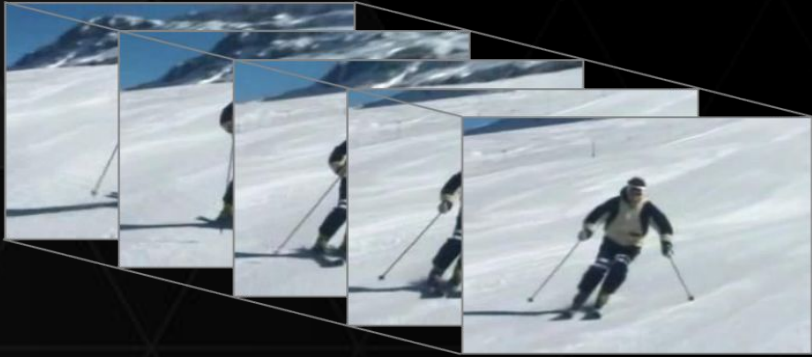


CNN / RNN



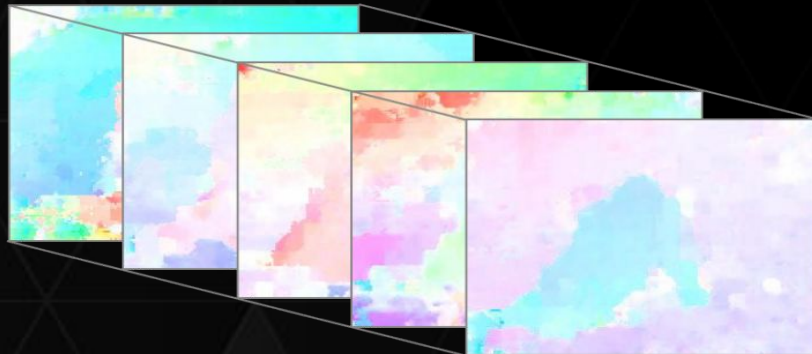
ДВУХКАНАЛЬНАЯ ОБРАБОТКА (TWO-STREAM)

RGB кадры



CNN / RNN

Оптический поток



CNN / RNN



+



The background is a dark blue field with a complex network of thin, light green lines. These lines connect various points, some of which are highlighted as bright green dots. The overall effect is a sense of a digital or neural network. The word "СПАСИБО" is centered in the middle of the image in a white, sans-serif font.

СПАСИБО