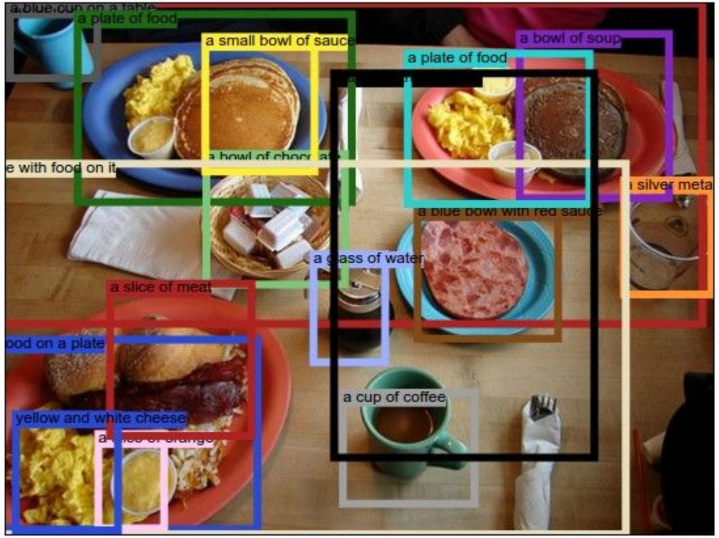
Цифровая обработка изображения

8. Применение рекурентных сетей в задачах анализа изображений

План занятия

- Рекурентные сети
- Автоматическая аннотация изображения
- Ответы на вопросы по изображению
- Распознавание текста

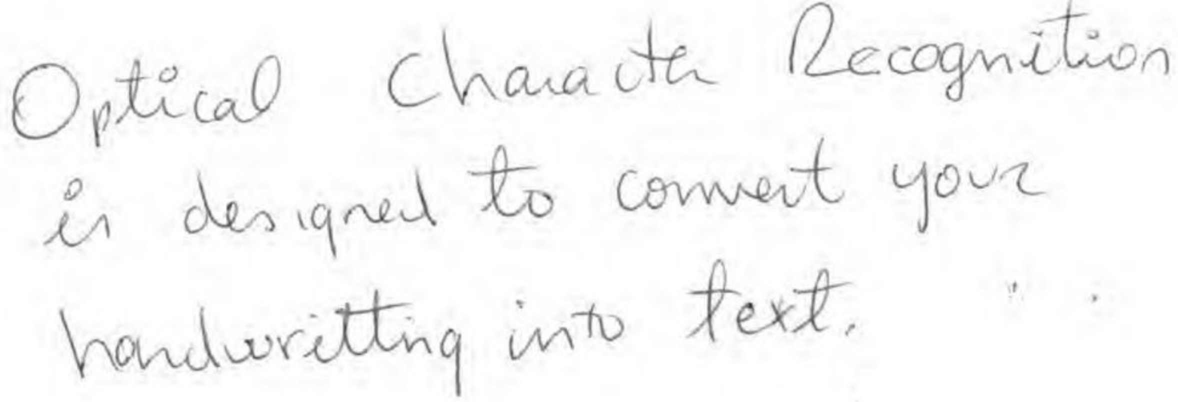
Описание изображения



a plate of food. food on a plate. a blue cup on a table. a plate of food. a blue bowl with red sauce. a bowl of soup. a cup of coffee. a bowl of chocolate. a glass of water. a plate of food. a silver metal container. a small bowl of sauce. table with food on it. a slice of orange. a table with food on it. a slice of meat. yellow and white cheese.

Распознавание текста на изображении

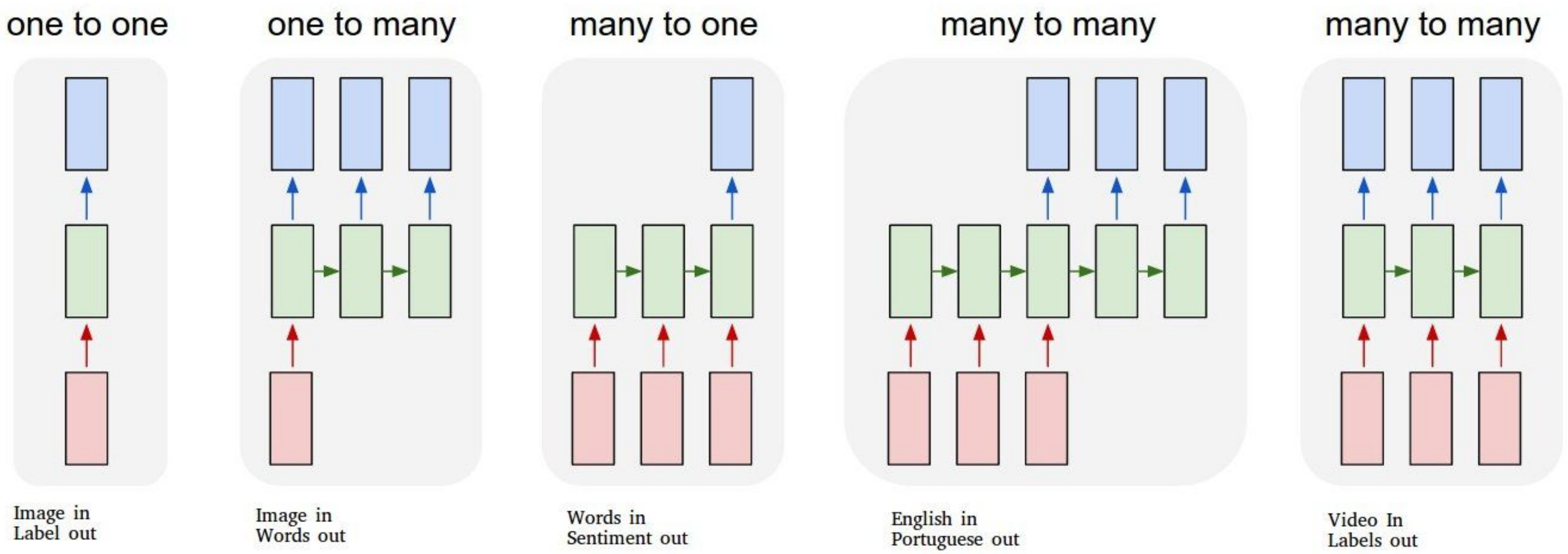
Распознавание рукописного текста



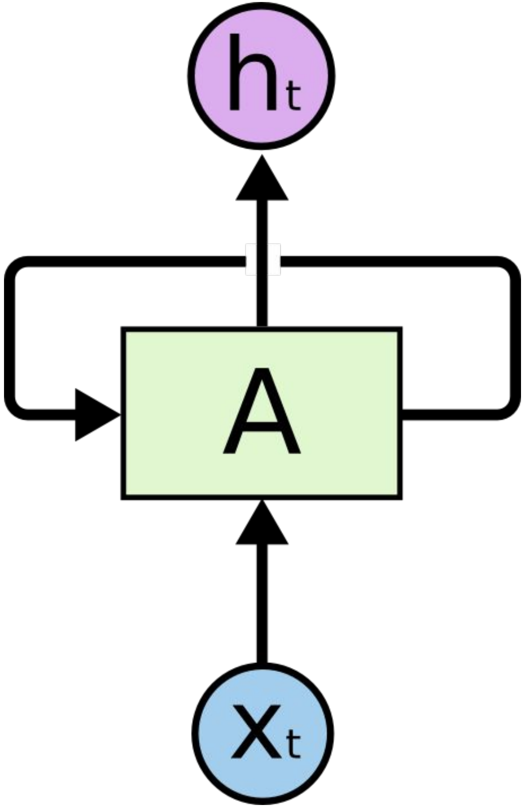
Optical Character Recognition
is designed to convert your
handwriting into text.

Рекурентные сети

Рекурентные сети



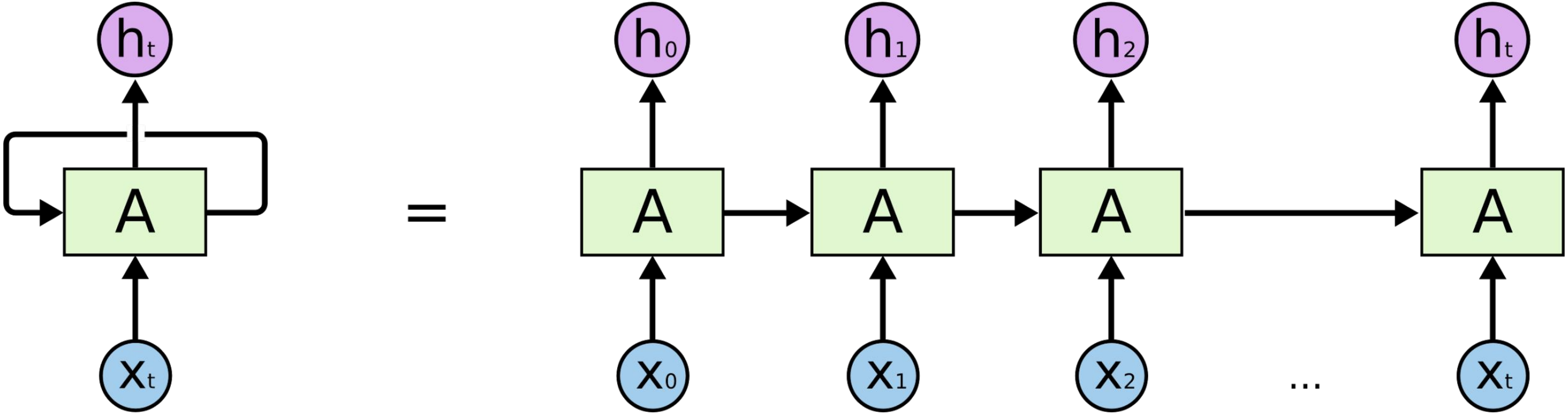
Рекурентные сети



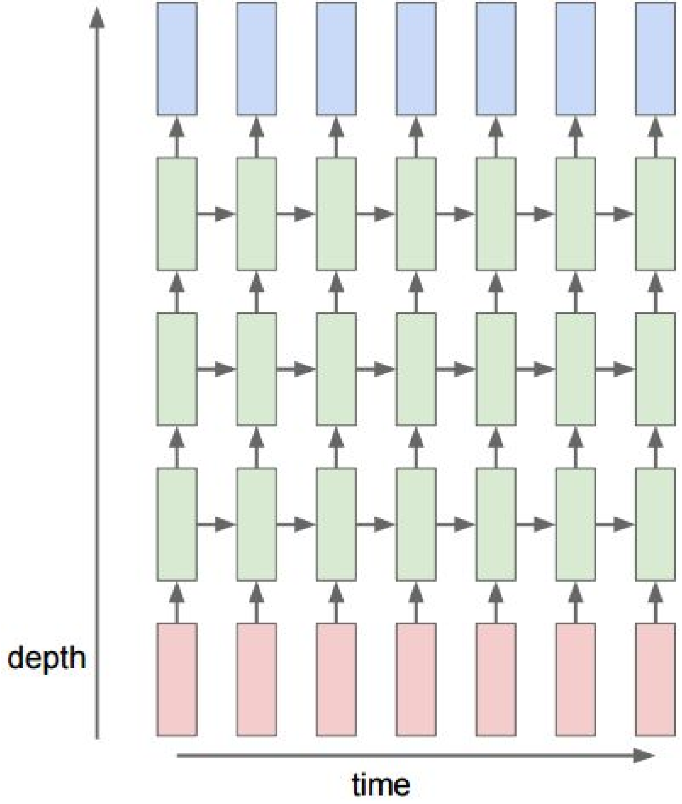
$$ht = f_{weights}(h_{t-1}, x_t) \therefore$$
$$h_t = tanh(W_{hh}.\,h_{t-1} + W_{xh}.\,x_t)$$

ht - состояние ячейки в момент времни t
xt - входной сигнал на шаге t
Whh - матрица преобразования состояния
Wxh - матрица преобразования входного сигнала

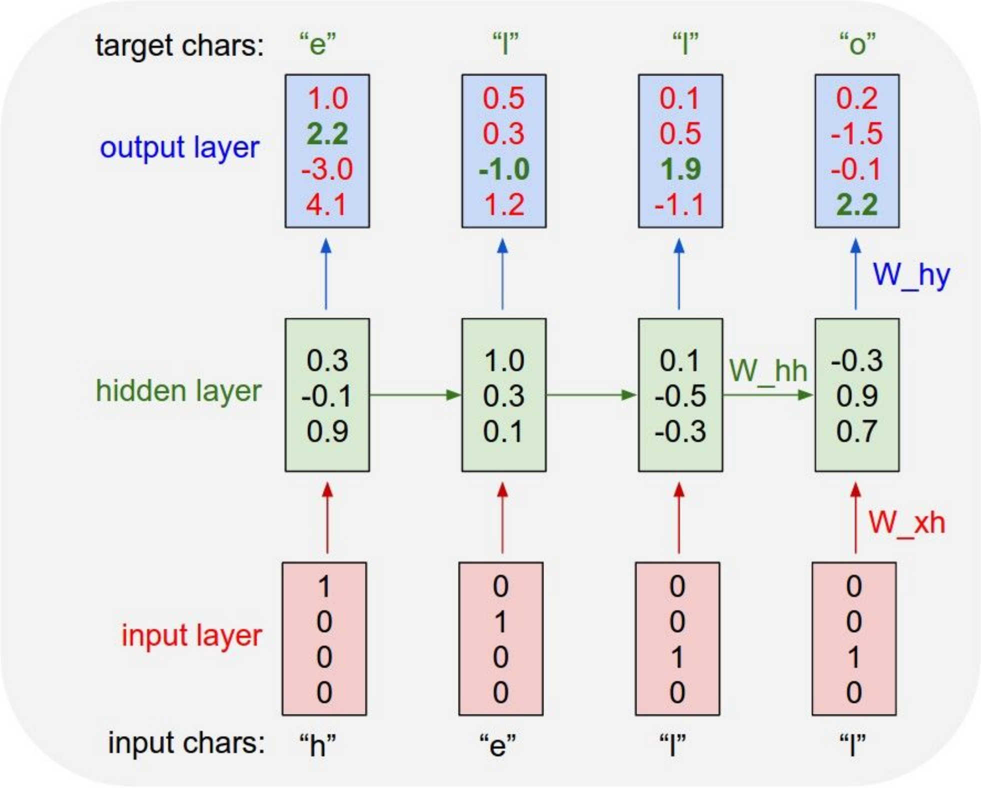
Рекурентные сети



Рекурентные сети



Генерация текстовой последовательности



Генерация текстовой последовательности

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma **??**. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

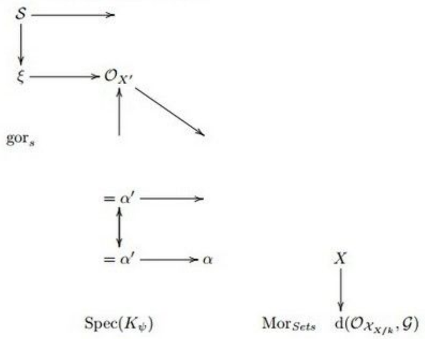
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

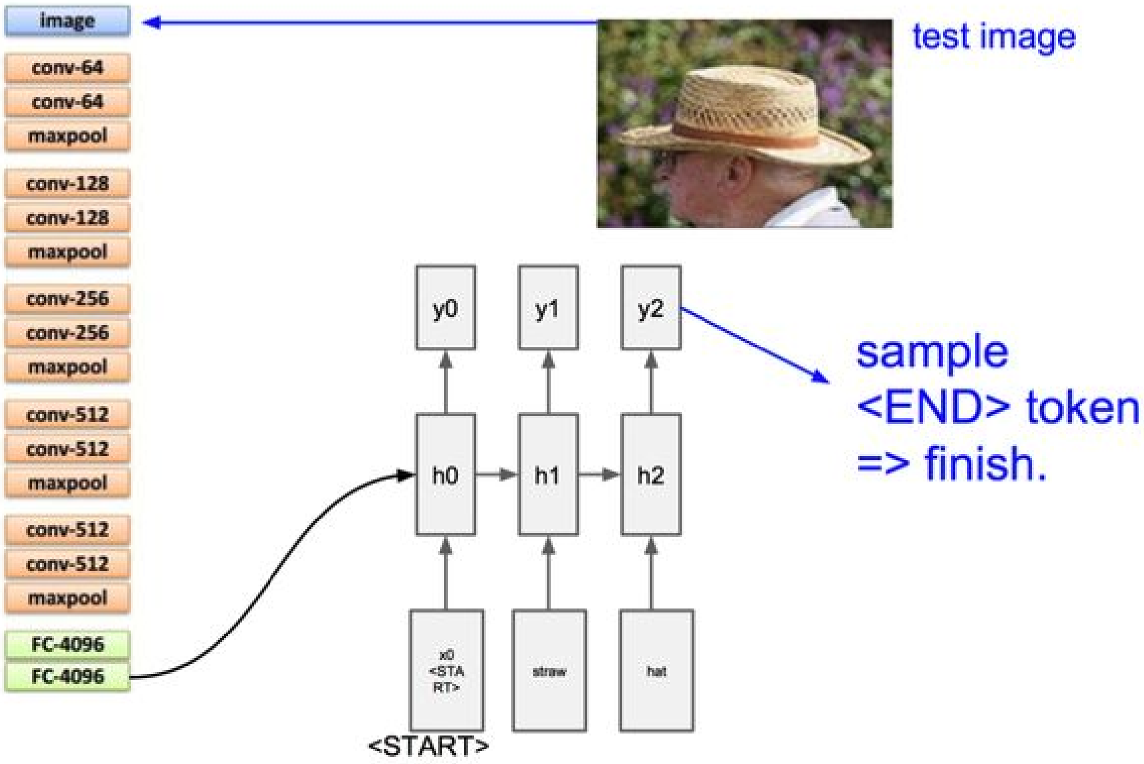
Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas **??**.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

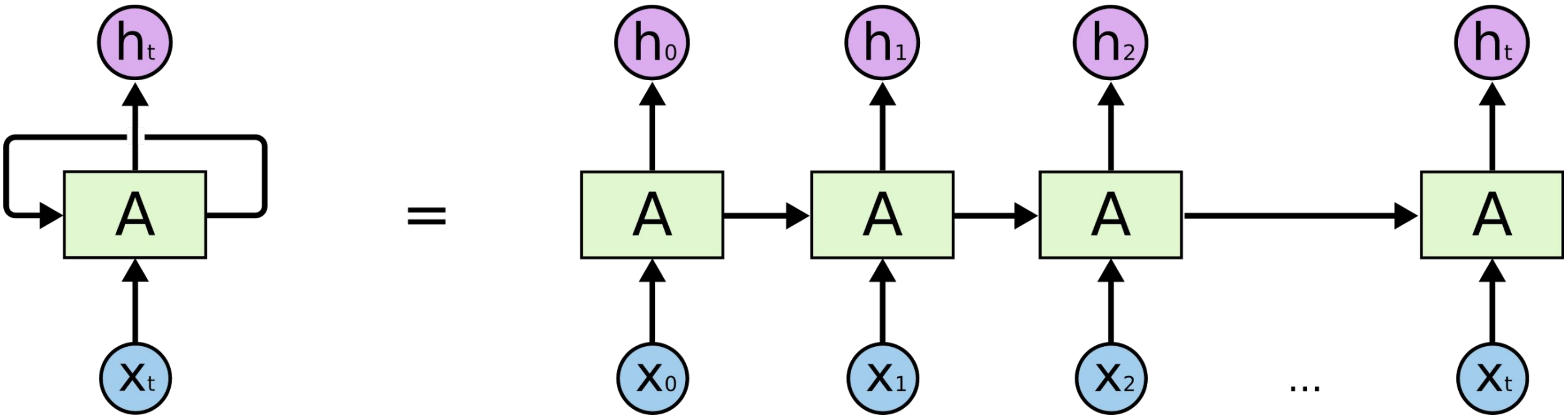
is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition **??** and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma **??**. This is a sequence of $\mathcal{F}$ is a similar morphism.

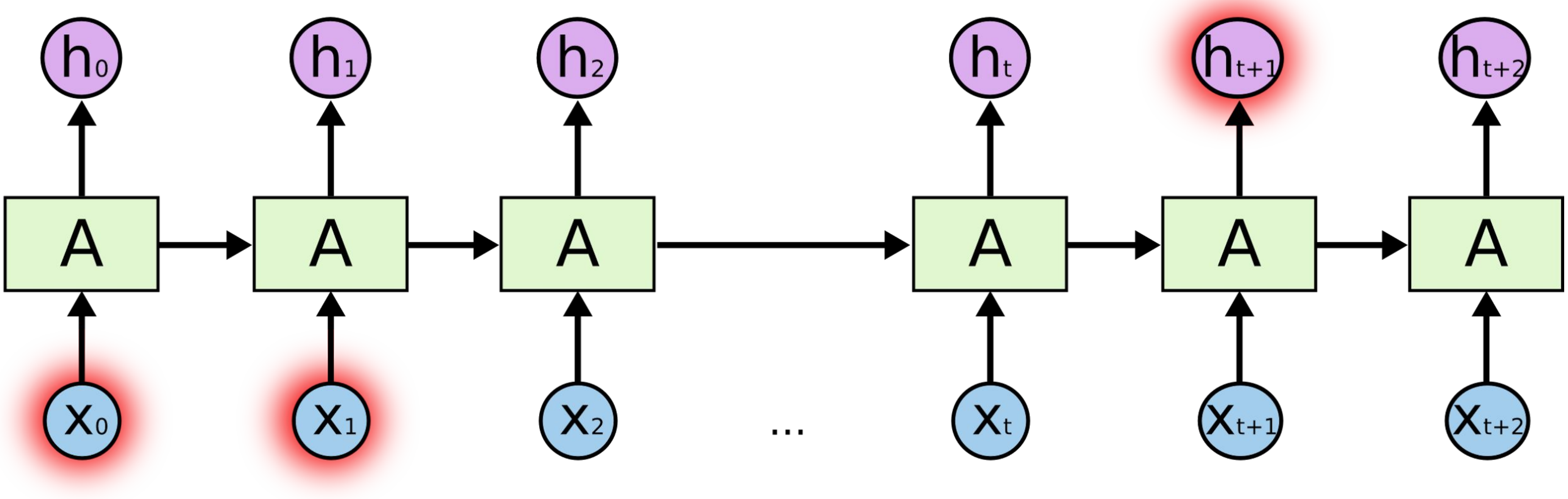
Генерация описания изображения



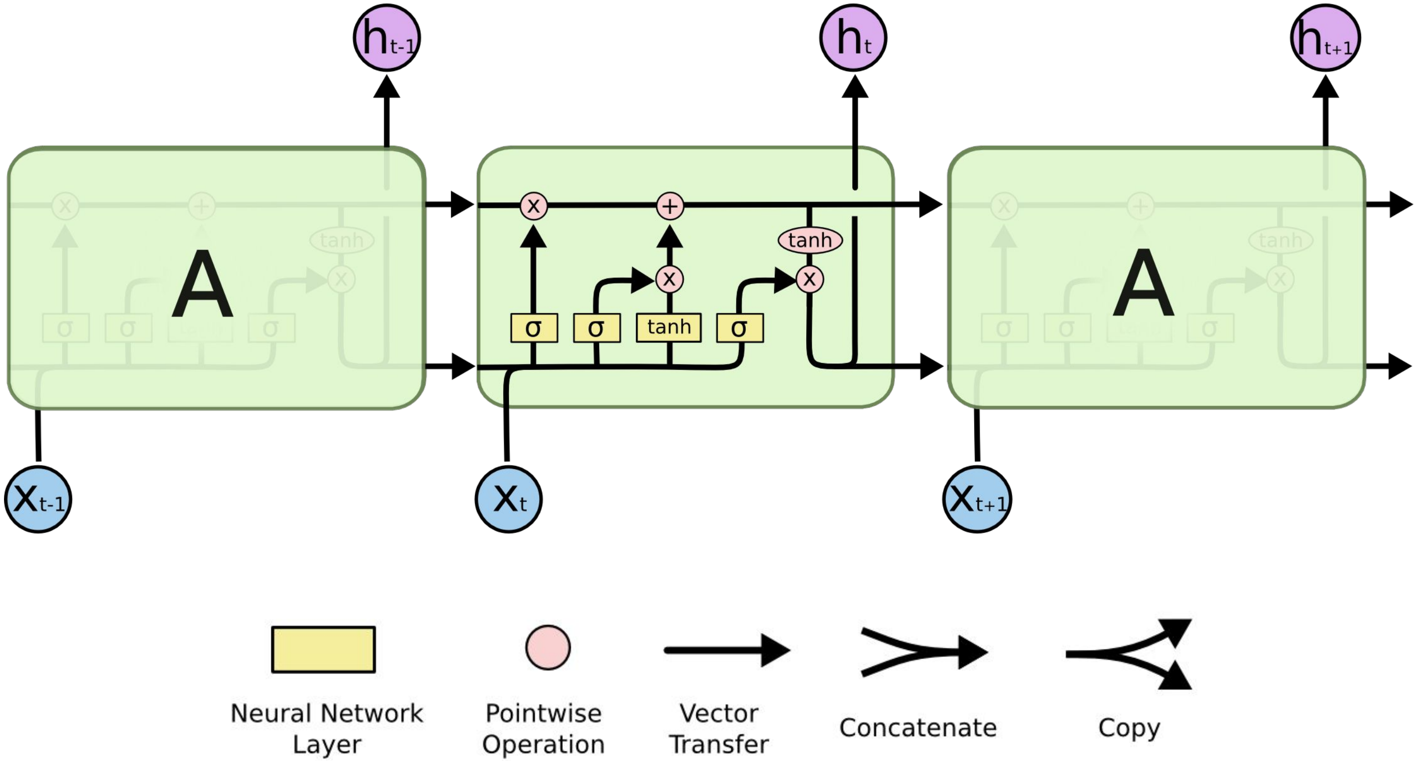
Рекурентные сети



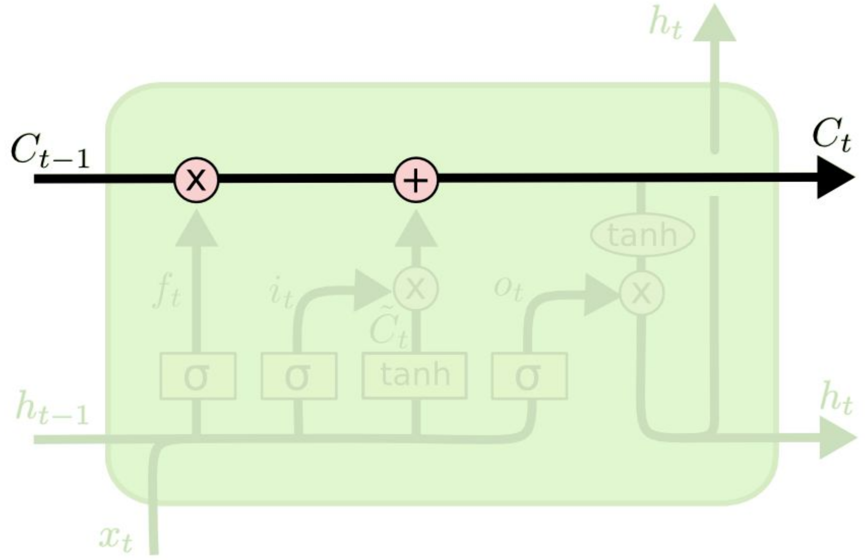
Проблема длительной памяти



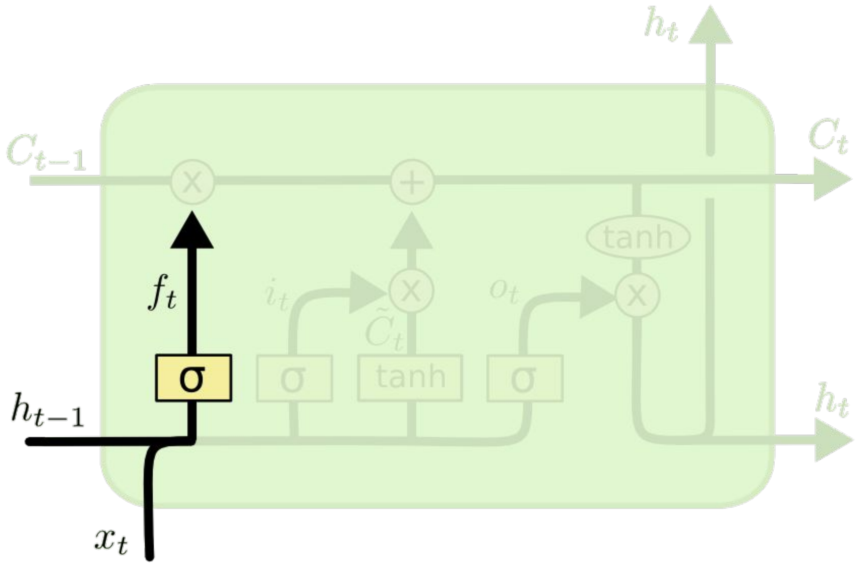
LSTM (Long Short Term Memory)



LSTM (Long Short Term Memory)

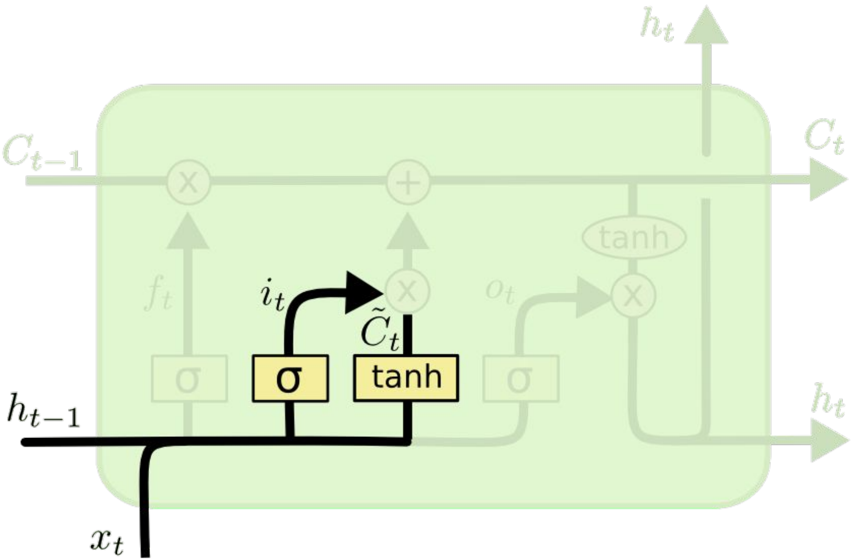


LSTM (Long Short Term Memory)



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

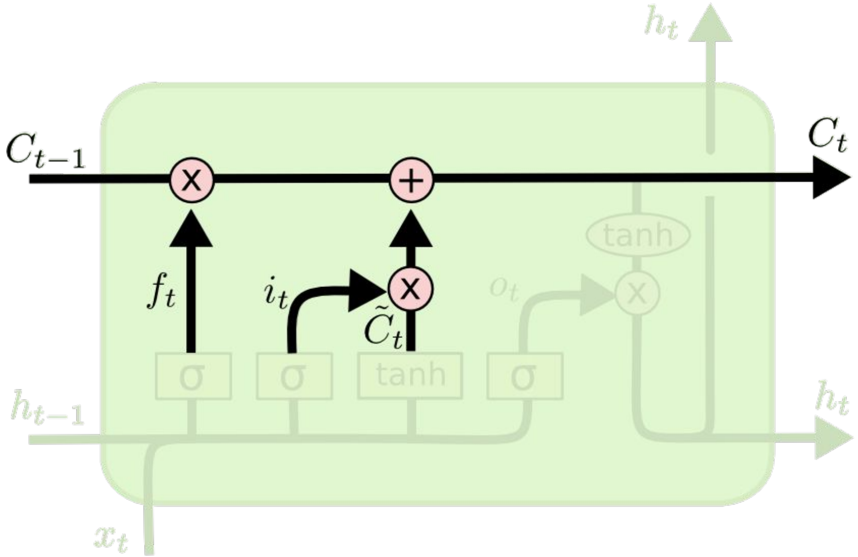
LSTM (Long Short Term Memory)



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

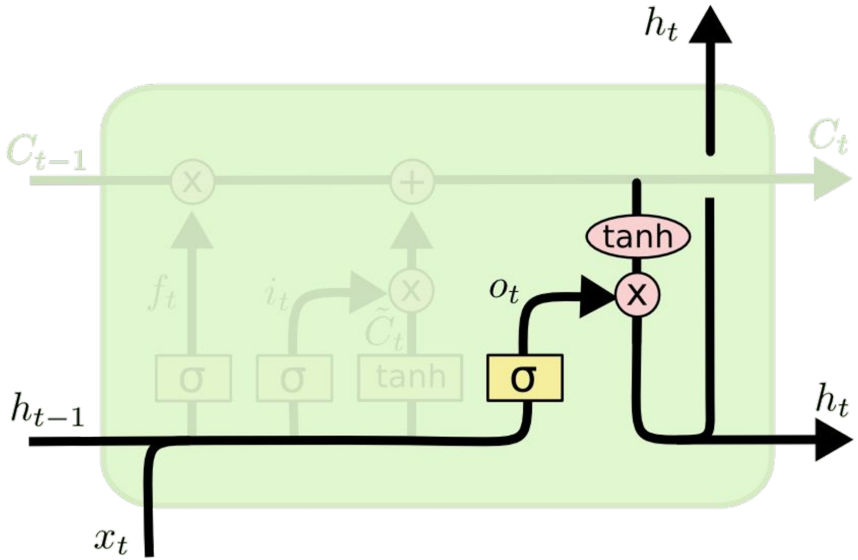
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

LSTM (Long Short Term Memory)



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

LSTM (Long Short Term Memory)



$$o_t = \sigma\left(W_o\ [\,h_{t-1}, x_t]\ +\ b_o\right)$$

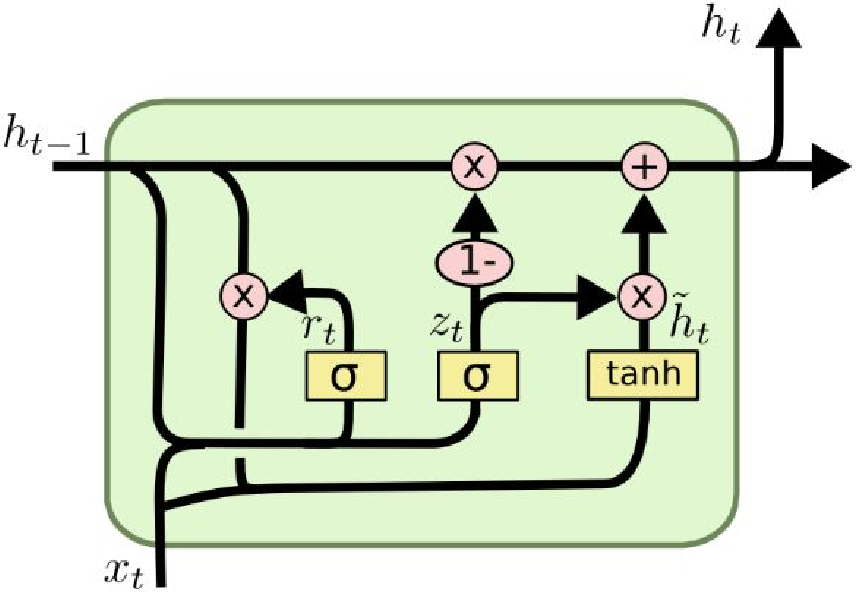
$$h_t = o_t * \tanh\left(C_t\right)$$

LSTM (Long Short Term Memory)

```
keras.layers.recurrent.LSTM(units, activation='tanh',
                            recurrent_activation='hard_sigmoid',
                            dropout=0.0,
                            recurrent_dropout=0.0)
```

```
units - размерность выхода
activation - активация выхода
recurrent_activation - активация внутреннего состояния
dropout - для компонент выходного вектора
recurrent_dropout - для компонент состояния
```

GRU Gated Recurrent Unit



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

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

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

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

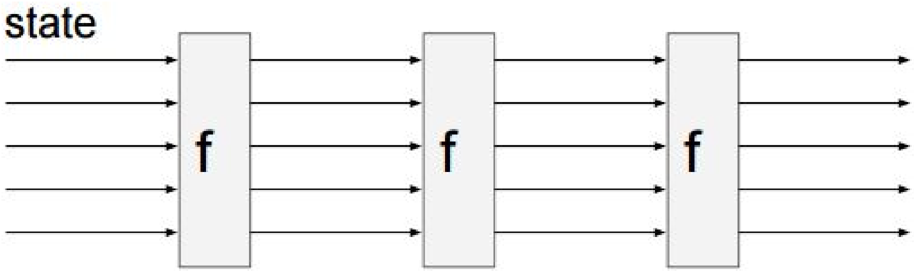
GRU Gated Recurrent Unit

```
keras.layers.recurrent.GRU(units, activation='tanh',
                           recurrent_activation='hard_sigmoid',
                           dropout=0.0, recurrent_dropout=0.0)
```

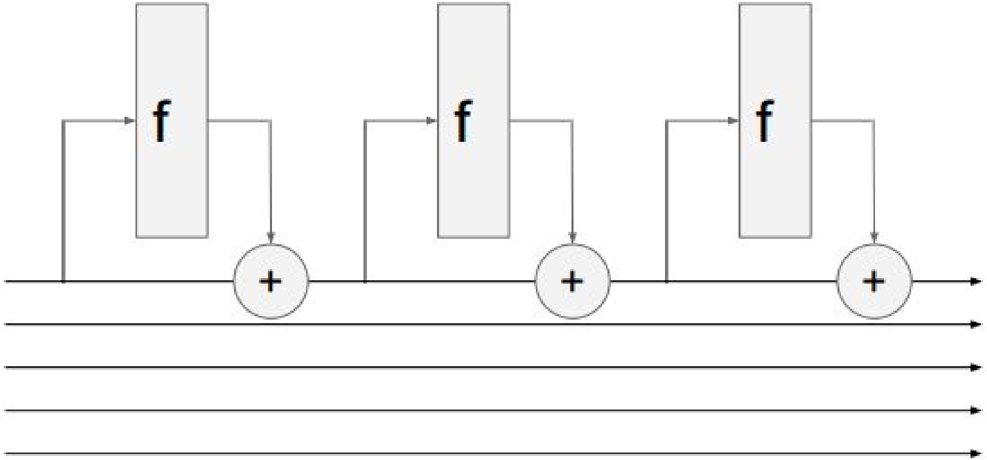
```
units - размерность выхода
activation - активация выхода
recurrent_activation - активация внутреннего состояния
dropout - для компонент выходного вектора
recurrent_dropout - для компонент состояния
```

RNN vs LSTM

RNN



LSTM
(ignoring
forget gates)



Автоматическая аннотация изображения

COCO 2015

The man at bat readies to swing at the pitch while the umpire looks on.

A large bus sitting next to a very tall building.

Метрики

Reference Sentences

R1: A bald eagle sits on a perch.

R2: An american bald eagle sitting on a branch in the zoo.

R3: Bald eagle perched on piece of lumber.

...

R50: A large bird standing on a tree branch.

Candidate Sentences

C1: An eagle is perched among trees.

C2: A picture of a bald eagle on a rope stem.

Triplet Annotation

Which of the sentences, B or C, is more similar to sentence A?

Sentence A : Anyone from R1 to R50

Sentence B : C1

Sentence C : C2

[CIDEr: Consensus-based Image Description Evaluation]

Метрики

$$\mathrm{CIDEr}_n(c_i, S_i) = \frac{1}{m} \sum_j \frac{\boldsymbol{g^n}(c_i) \cdot \boldsymbol{g^n}(s_{ij})}{\|\boldsymbol{g^n}(c_i)\| \|\boldsymbol{g^n}(s_{ij})\|},$$

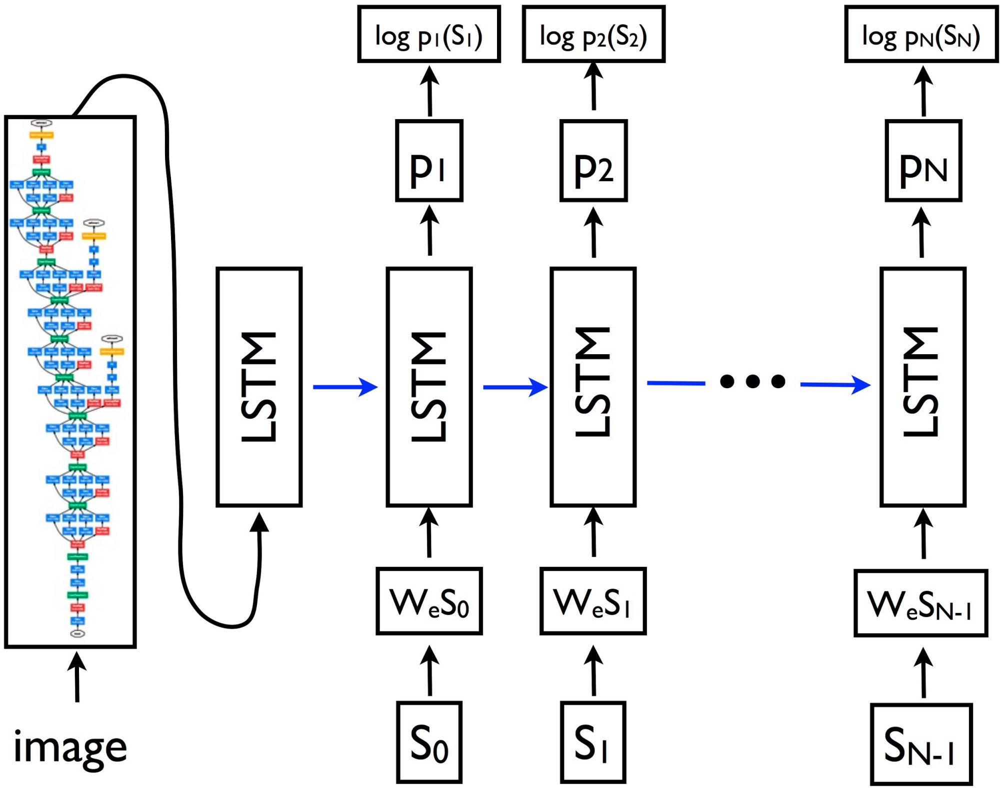
$$\mathrm{CIDEr}(c_i, S_i) = \sum_{n=1}^{N} w_n \mathrm{CIDEr}_n(c_i, S_i),$$

c - сгенерированное предложение
S - предложения из разметки
g - вектор n-грамм предложения

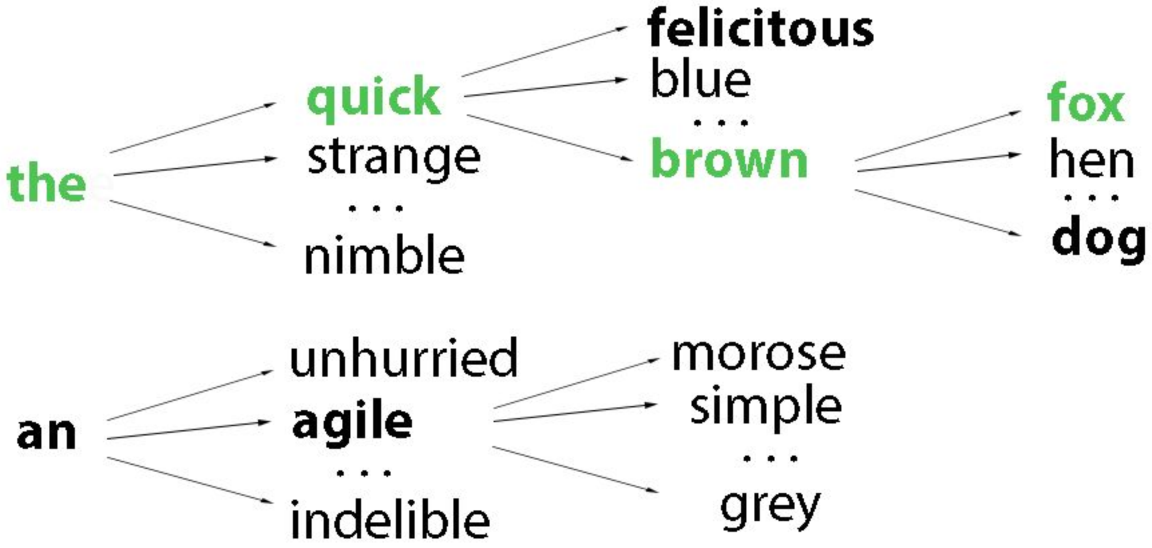
Im2Txt

https://github.com/tensorflow/models/tree/master/im2txt

Im2Txt

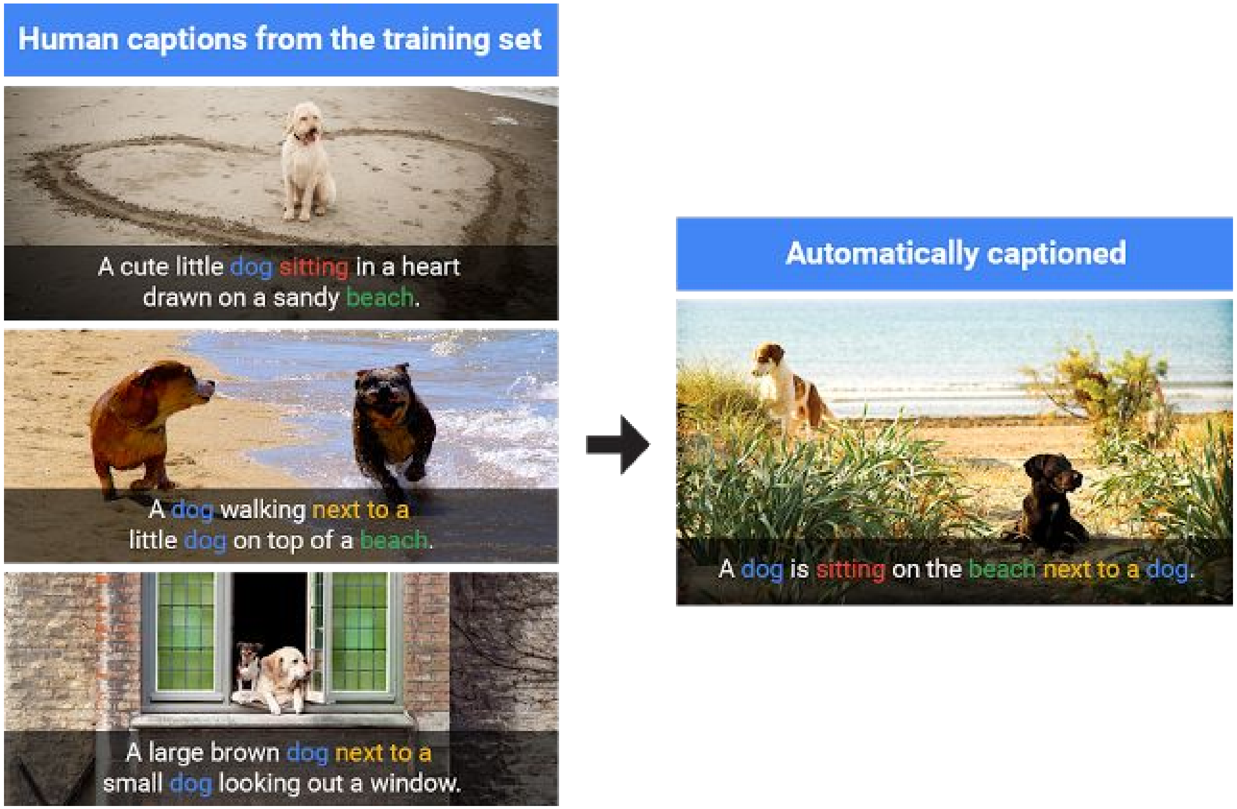

log p1(S1)
log p2(S2)
log pN(SN)
p1
p2
pN
LSTM
LSTM
LSTM
LSTM
WeS0
WeS1
WeSN-1
S0
S1
SN-1
image

BeamSearch



felicitous
blue
...
brown
quick
strange
...
nimble
the
fox
hen
...
dog
unhurried
agile
...
indelible
an
morose
simple
...
grey

Im2Txt


Human captions from the training set
A cute little dog sitting in a heart drawn on a sandy beach.
A dog walking next to a little dog on top of a beach.
A large brown dog next to a small dog looking out a window.
Automatically captioned
A dog is sitting on the beach next to a dog.

NeuralTalk

http://cs.stanford.edu/people/karpathy/densecap/

NeuralTalk

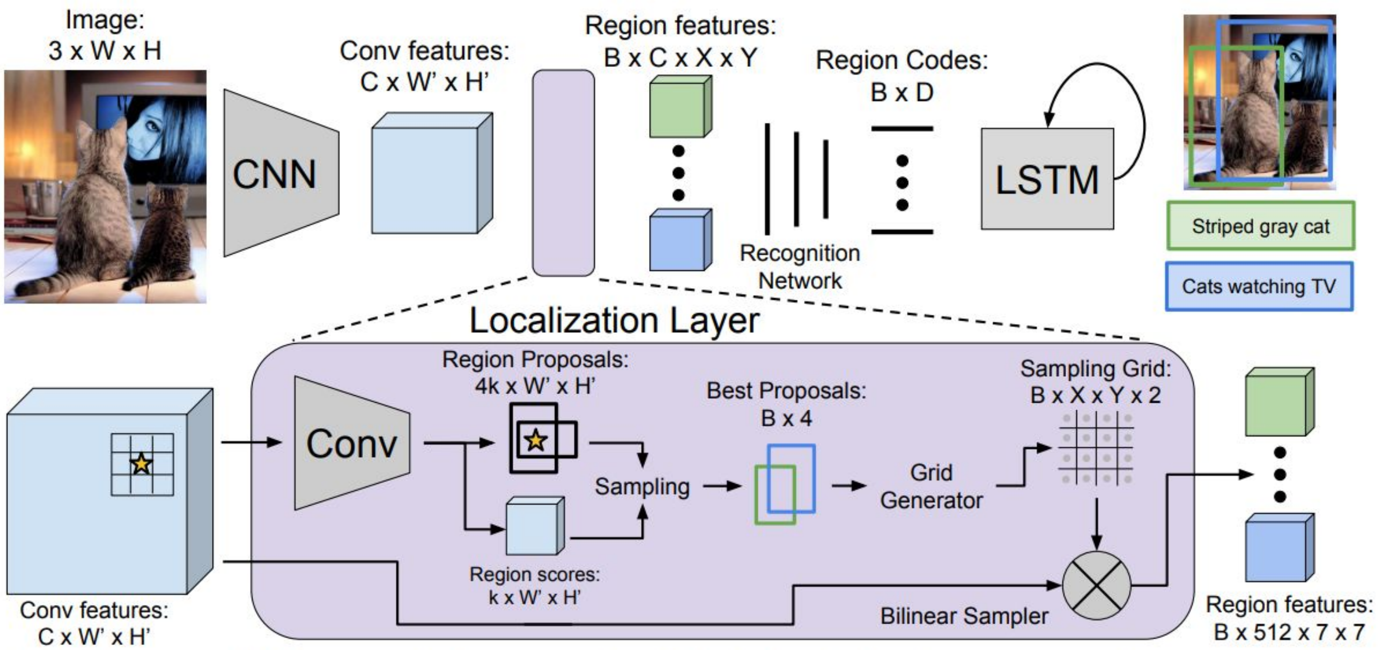


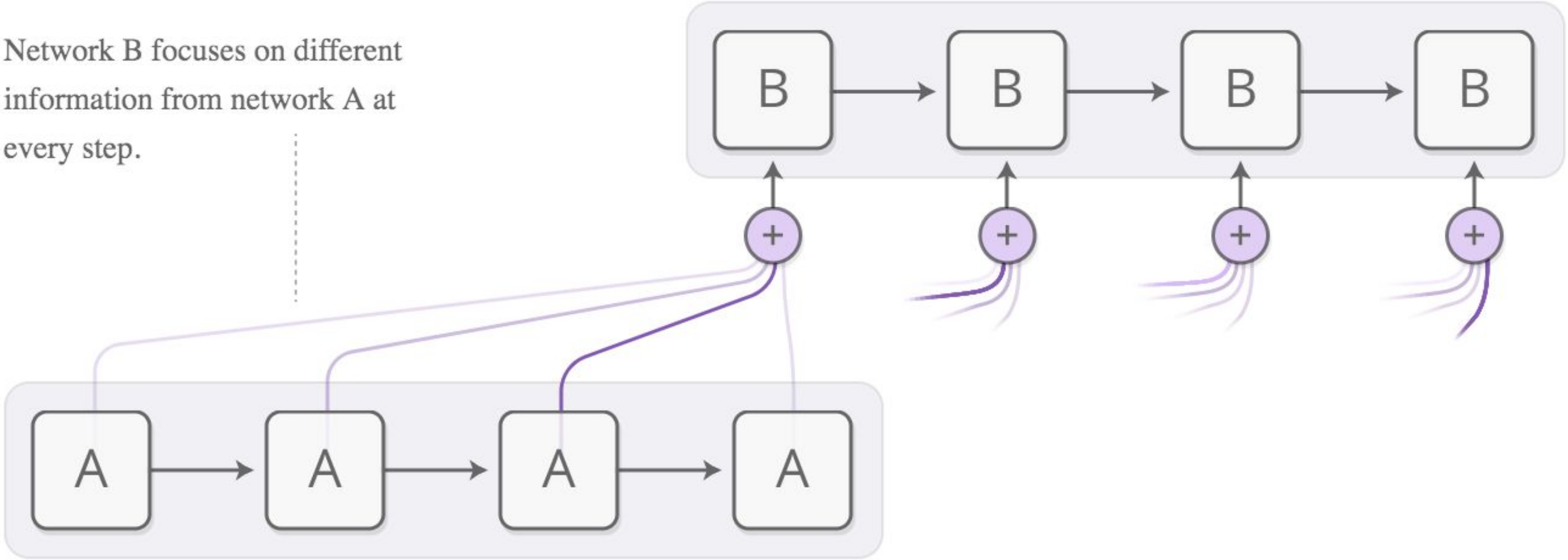
Figure 2. Model overview. An input image is first processed a CNN. The Localization Layer proposes regions and smoothly extracts a batch of corresponding activations using bilinear interpolation. These regions are processed with a fully-connected recognition network and described with an RNN language model. The model is trained end-to-end with gradient descent.

Attention

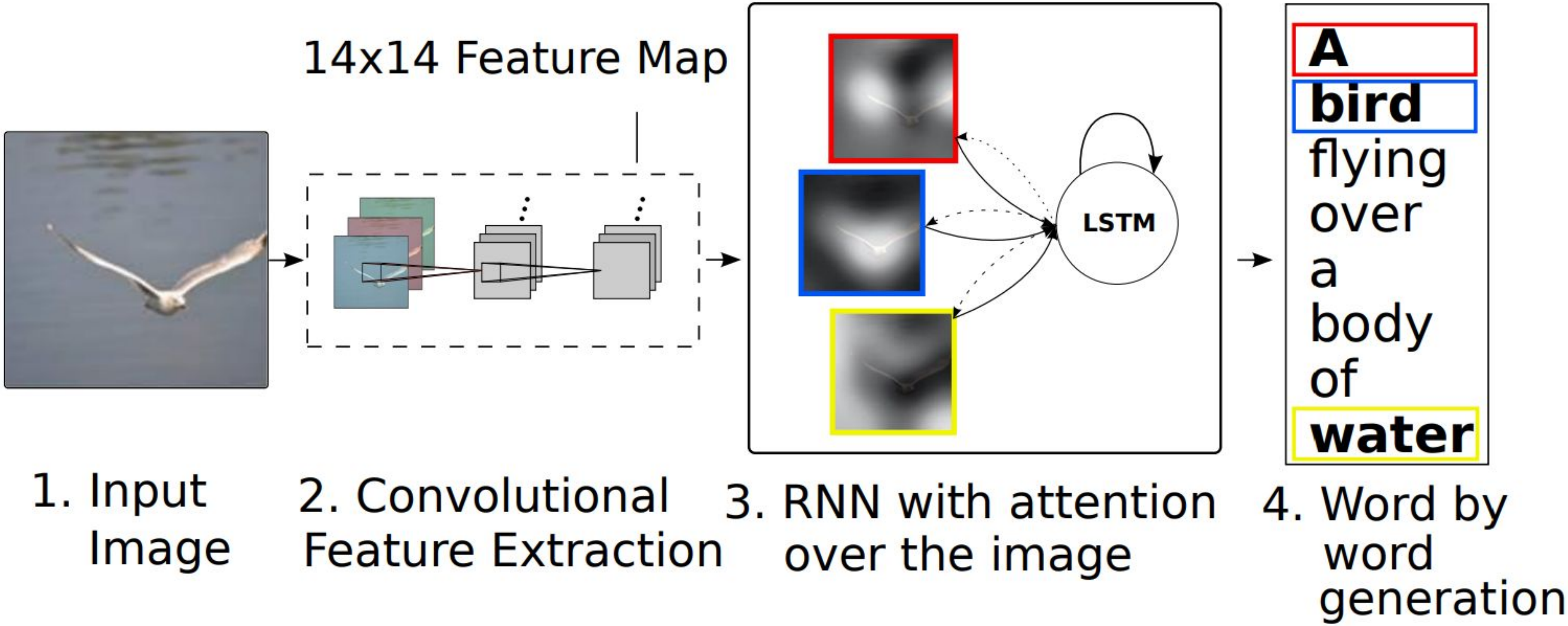
Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Attention

Network B focuses on different information from network A at every step.



Attention


14x14 Feature Map
LSTM
A
bird
flying
over
a
body
of
water
1. Input Image
2. Convolutional Feature Extraction
3. RNN with attention over the image
4. Word by word generation

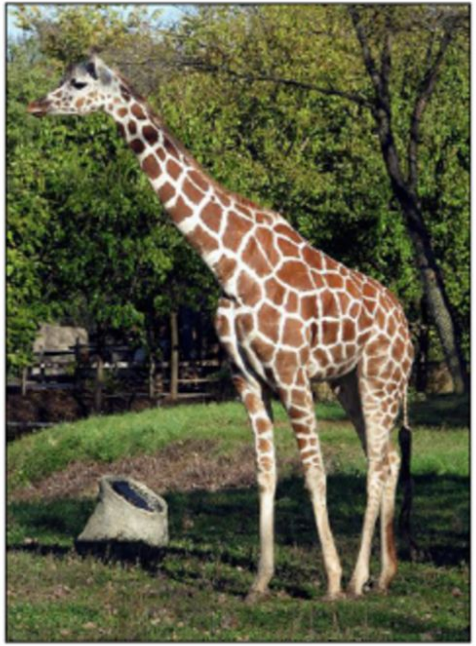
Attention

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

Автоматическая аннотация изображения

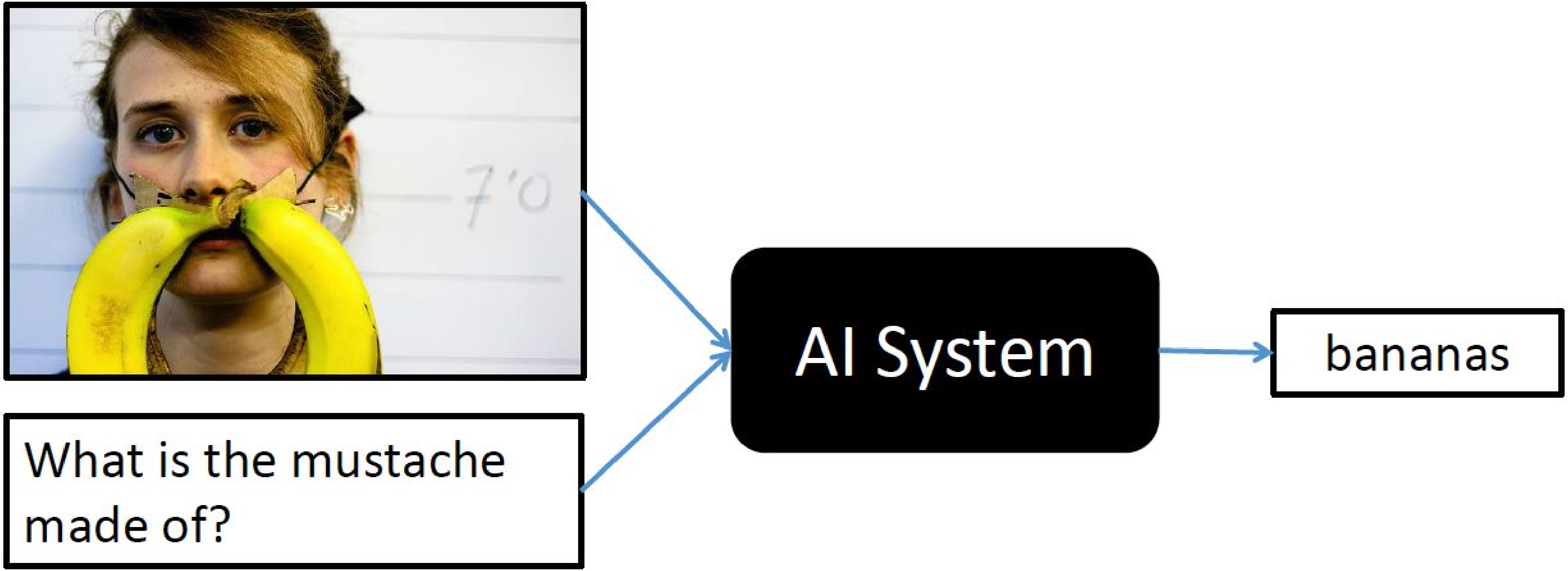


A giraffe standing next to a tree.

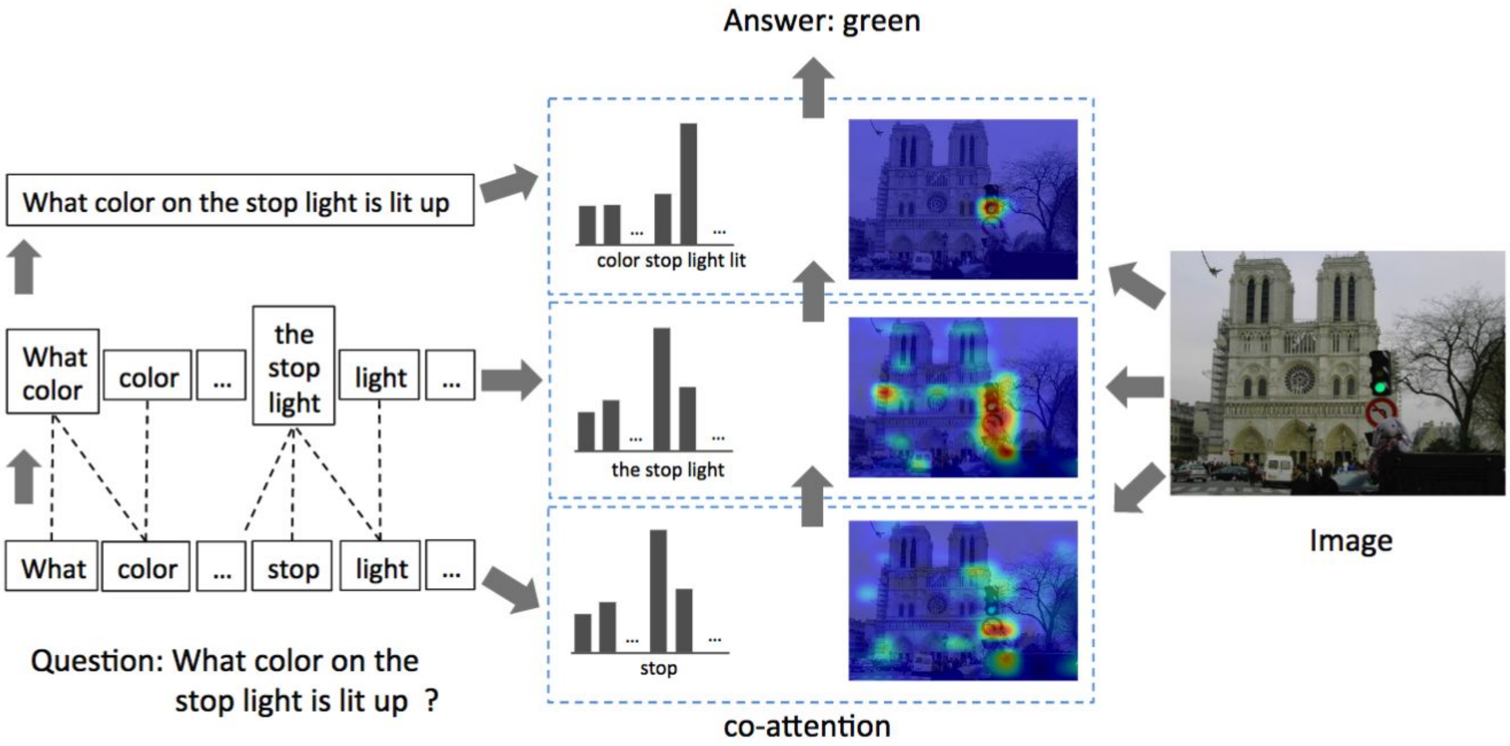
Ответы на вопросы по изображению
Visual Question Answering

https://arxiv.org/pdf/1705.03865.pdf

Ответы на вопросы по изображению



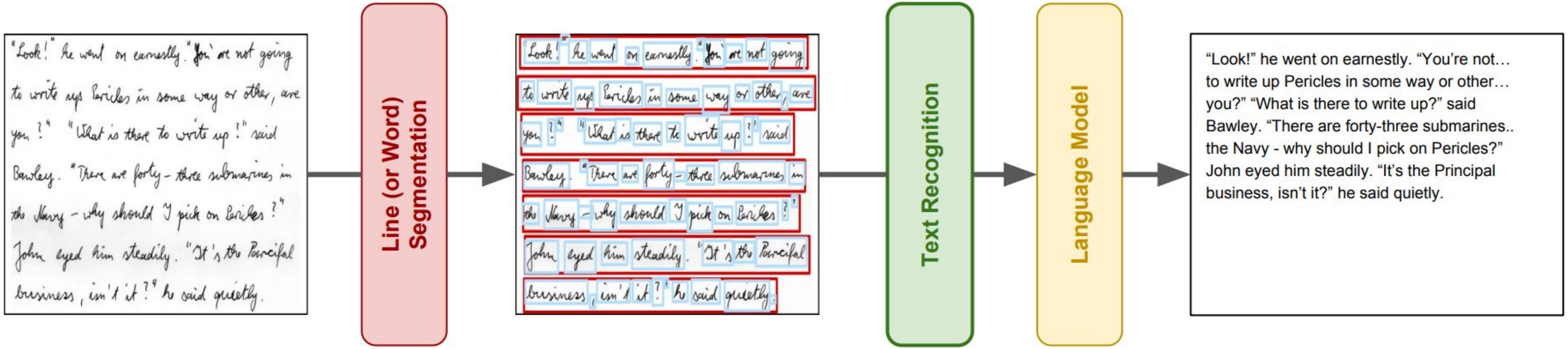
Ответы на вопросы по изображению



[Hierarchical Question-Image Co-Attention for Visual Question Answering]()

Распознавание текста

Распознавание текста

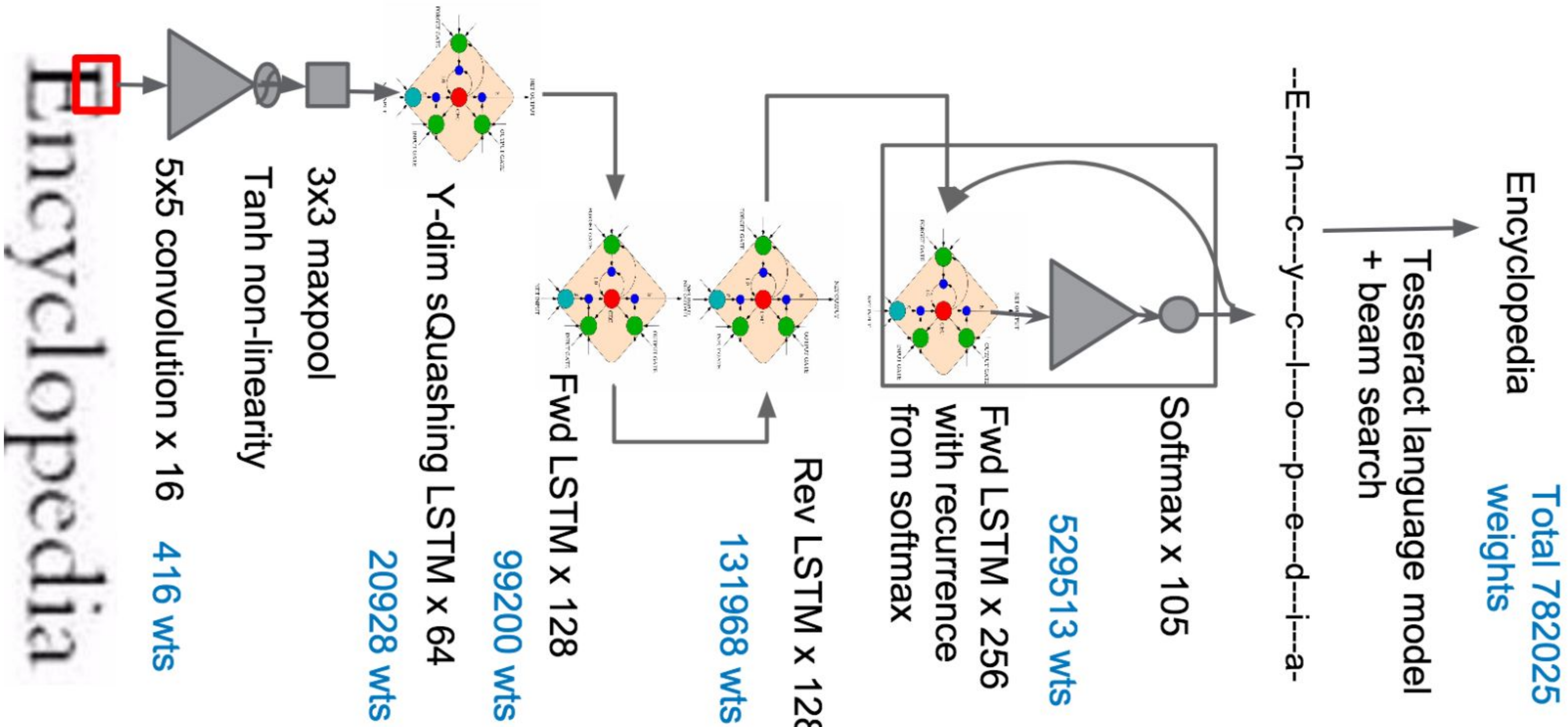


Tesseract

- открытая библиотека распознавания печатного текста
- поддерживает более 50 языков включая русский
- последняя версия модели построена на базе рекурентной сети

https://github.com/tesseract-ocr/tesseract

Tesseract

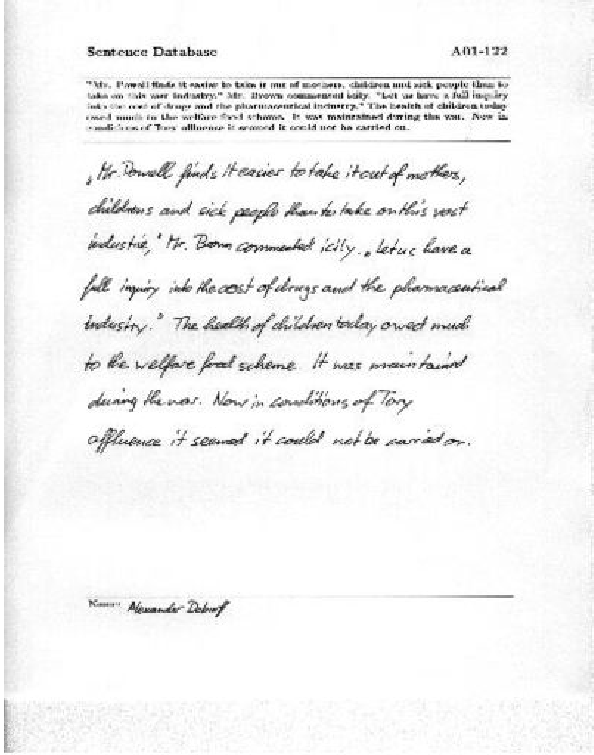


Encyclopedia
5x5 convolution x 16
416 wts
Tanh non-linearity
3x3 maxpool
20928 wts
Y-dim sQuashing LSTM x 64
99200 wts
Fwd LSTM x 128
131968 wts
Rev LSTM x 128
Fwd LSTM x 256
with recurrence
from softmax
529513 wts
Softmax x 105
-E--n--c--y--c--l--o--p--e--d--i--a-
Tesseract language model
+ beam search
Encyclopedia
Total 782025
weights

Распознавание рукописного ввода

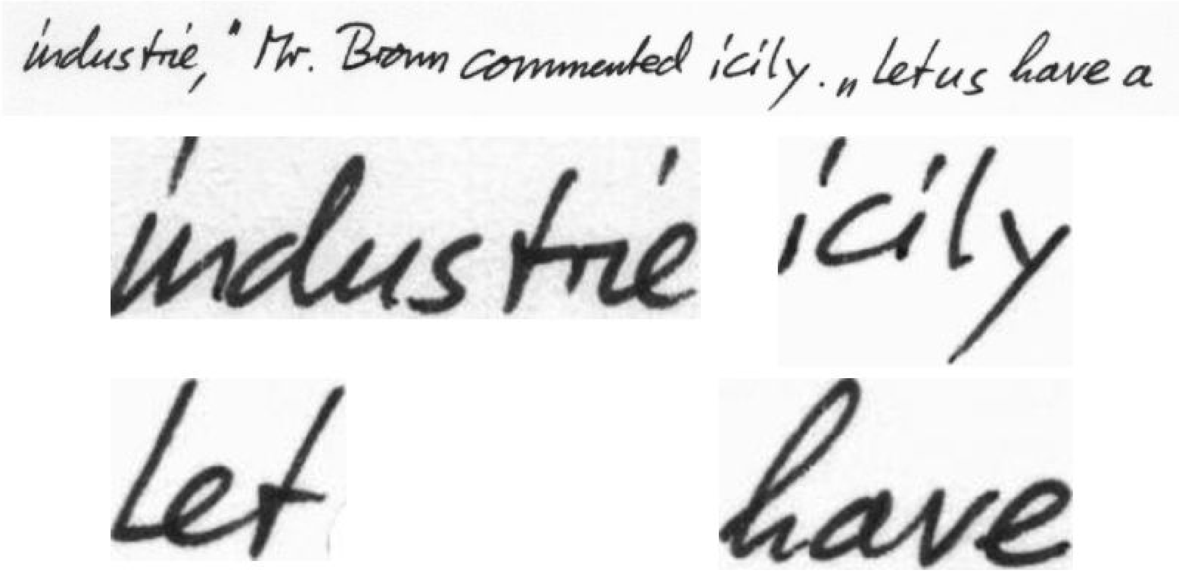
Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks (Graves)

IAM Handwriting Database


Sentence Database A01-122

„Mr. Powell finds it easier to take it out of mothers, childrens and sick people than to take on this vast industrie," Mr. Brown commented icily. „let us have a full inquiry into the cost of drugs and the pharmaceutical industry." The health of children today owed much to the welfare food scheme. It was maintained during the war. Now in conditions of Tory affluence it seemed it could not be carried on.

Name: Alexander Dobrof



[IAM Handwriting Database]

Распознавание рукописного ввода

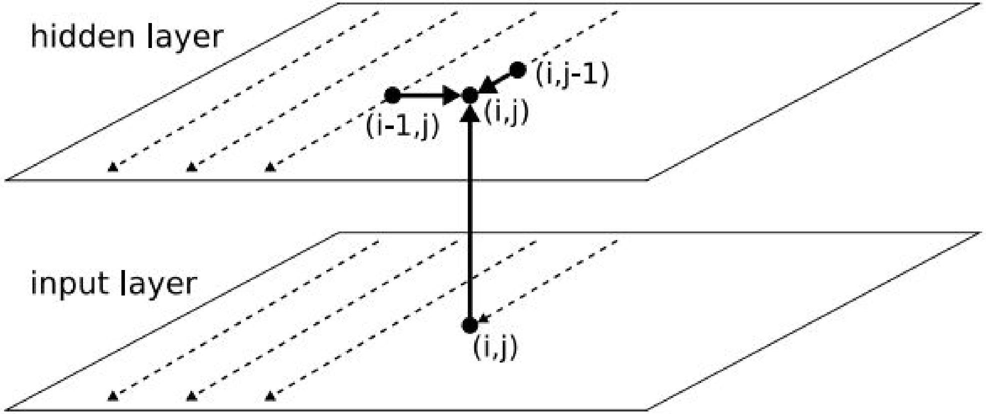
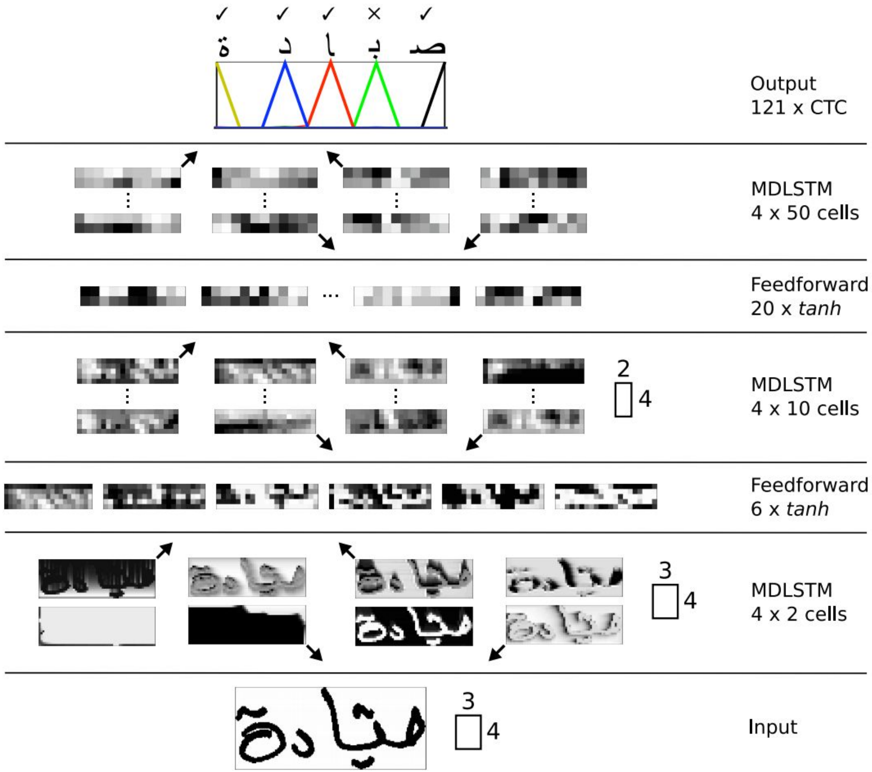


Figure 1: **Two dimensional MDRNN**. The thick lines show connections to the current point (i, j). The connections within the hidden layer plane are recurrent. The dashed lines show the scanning strips along which previous points were visited, starting at the top left corner.

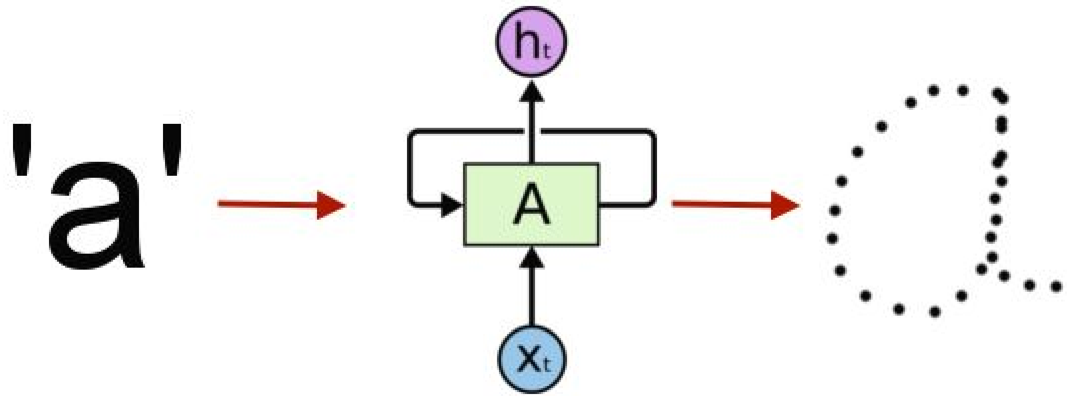
Распознавание рукописного ввода



Генерация рукописного текста

https://greydanus.github.io/2016/08/21/handwriting/

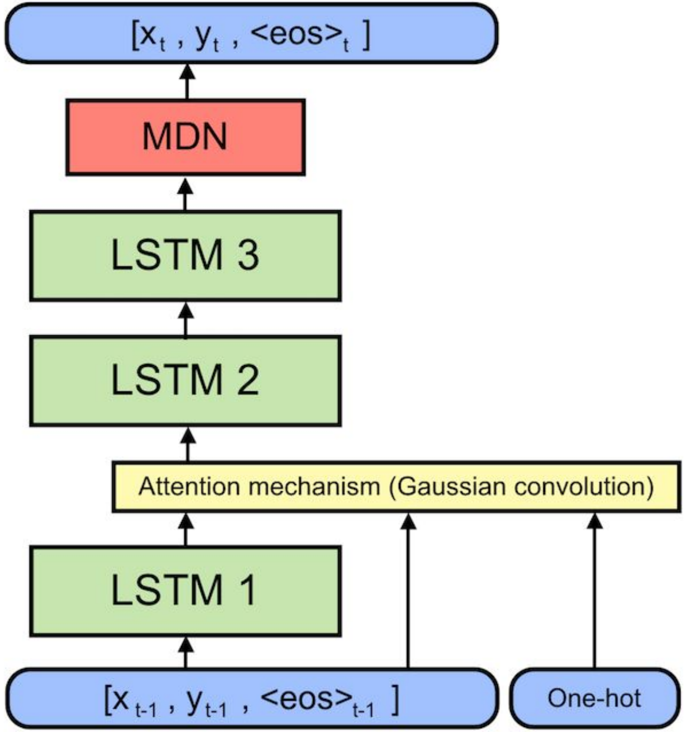
Генерация рукописного текста



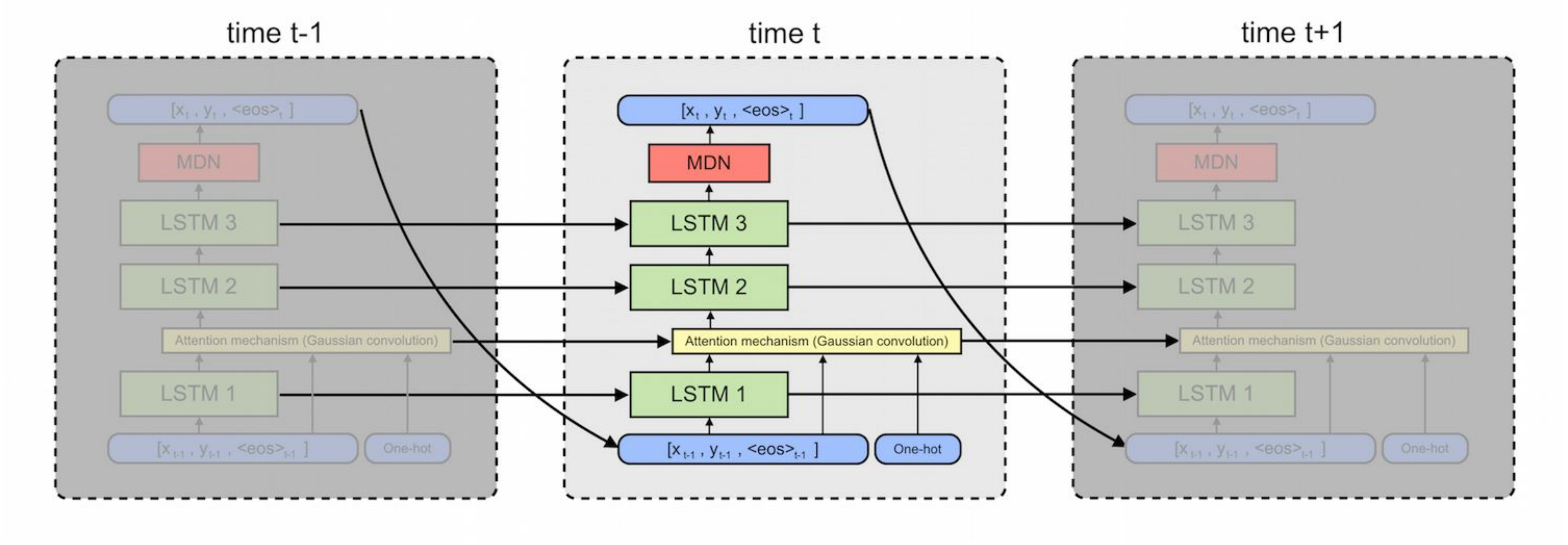

```
>>> print one_hot
[[ 1.  0.  0. ...,  0.  0.  0.]]
```

```
>>> print pen_points
[[ 0.03310043 -1.05923397  0.        ]
 [ 1.99788946 -0.55632969  0.        ]
 [-0.88192711 -1.66361628  0.        ]
              :
 [-0.78227638  1.64455155  1.        ]]
```

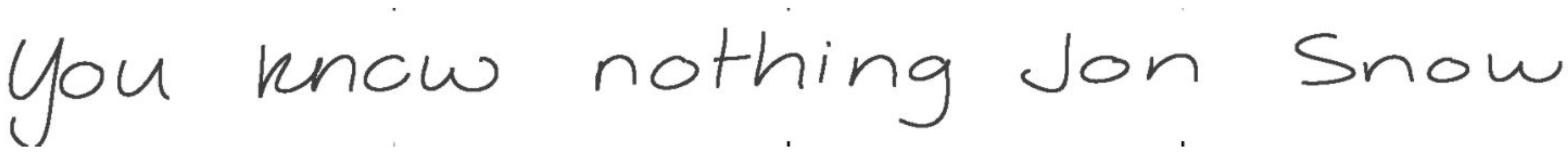
Генерация рукописного текста



Генерация рукописного текста



Генерация рукописного текста



You know nothing Jon Snow

cursive is still hard :(

Полезные материалы

- VISUALIZING AND UNDERSTANDING RECURRENT NETWORKS
- Attention and Augmented Recurrent Neural Networks
- Deep Neural Networks – Applications in Handwriting Recognition
- Awesome Recurrent Neural Networks