

Дерунец Роман

# Внимание!

## Лекция 10

# Роман Дерунец



Сибирские нейросети



сибирские  
нейросети

ЛабПЦТ НГУ

LABADT

ИИР НГУ

Институт  
Интеллектуальной  
Робототехники





# Обо мне



- Спикер Data Fest 2023, 2024, Data Fest Siberia 4, 5, Технопром 2024, IT-Город 2024, AINL 2025
- Стипендиат Mediascope (стипендия имени В.В. Гродского)
- Финалист Phystech GigaChat Challenge, Urbancode, Cup IT, Лига приключений, AI Journey 2023




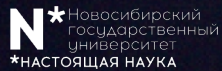

# Обо мне: science

- 1. Pisets: A Robust Speech Recognition System for Lectures and Interviews, NAACL 2025 
- 2. Интеллектуальная вопросно-ответная система на основе интеграции генеративной нейросетевой модели языка и неструктурированной базы знаний: Свидетельство о государственной регистрации программы для ЭВМ № 2025611584 
- 3. Knowledge as Recollection: Advancing Multimodal Retrieval-Augmented Generation, AINL 2025, принято к публикации; 
- 4. TabQA at SemEval-2025 Task 8: Column Augmented Generation for Question Answering over Tabular Data, ACL 2025, принято к публикации 
- 5. Соавтор MERA Code 



# Обо мне: educational



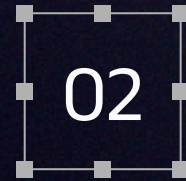
- Региональный центр "Альтаир": курсы по Python и машинному обучению, интенсивы в рамках кубка губернатора Новосибирской области  Альтаир  
Региональный центр
- Цифровая кафедра НГУ: машинное обучение и нейронные сети  N\* Новосибирский  
государственный  
университет  
\*НАСТОЯЩАЯ НАУКА
- Синьцзянский университет: зимняя школа искусственного интеллекта 2023 и 2024 



# Как мы уже поняли, word2vec



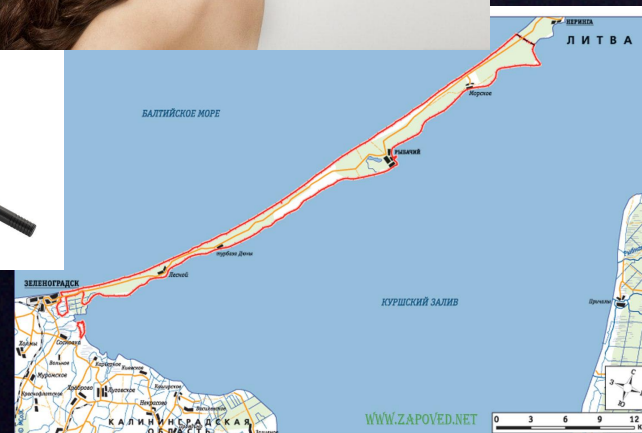
Не учитывают  
глубокую семантику



Не решают  
проблемы  
ОМОНИМИИ

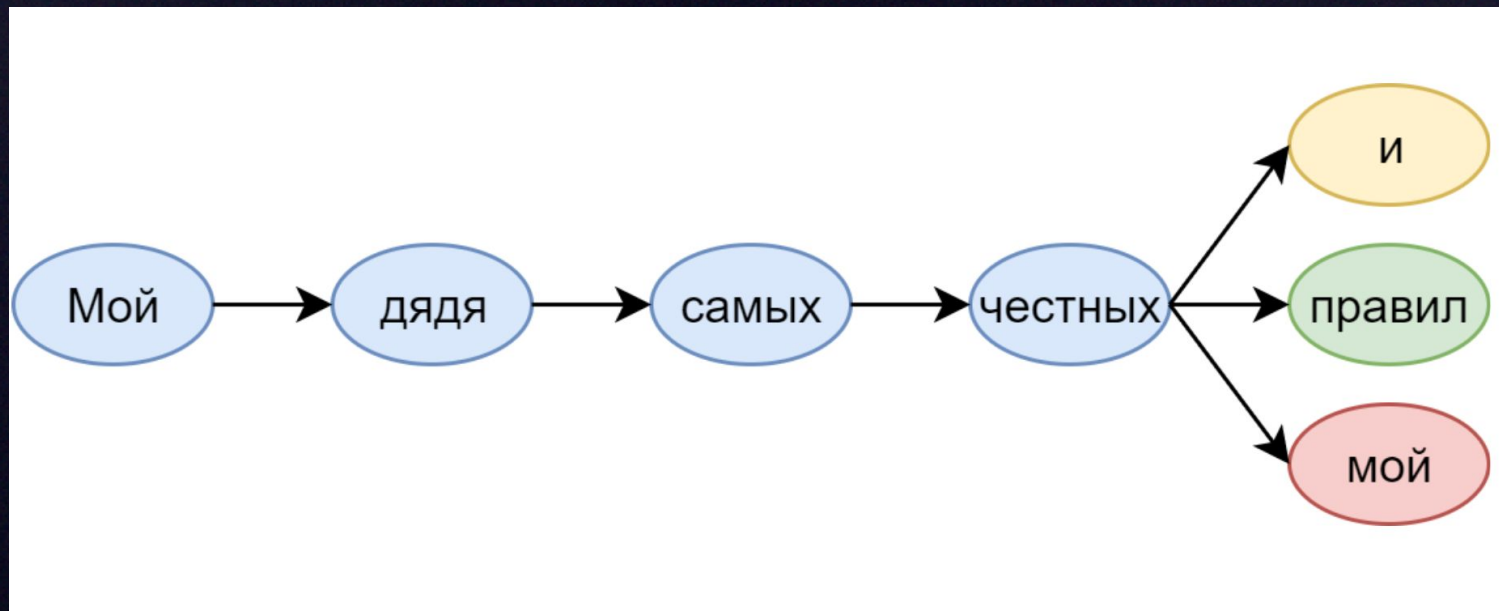
# Недостатки word2vec

## Коса



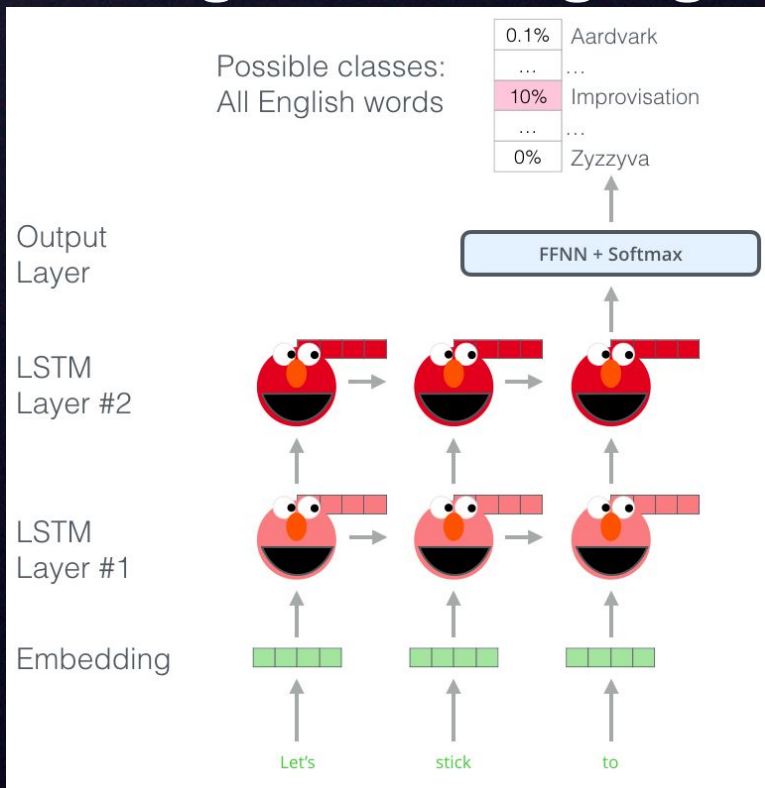


# Языковое моделирование





# Embeddings from Language Models



# LSTM и GRU – это хорошо! Но



- На длинных последовательностях контекст утекает....



# Модели с вниманием



*Белогубов. Мне, Аким Акимыч, только бы обратили внимание.*

*Юсов (строго). Что ты шутишь этим, что ли?*

*Белогубов. Как можно-с!..*

*Юсов. Обратили внимание... Легко сказать!*

*Чего еще нужно чиновнику? Чего он еще желать может?*

*Белогубов. Да-с!*

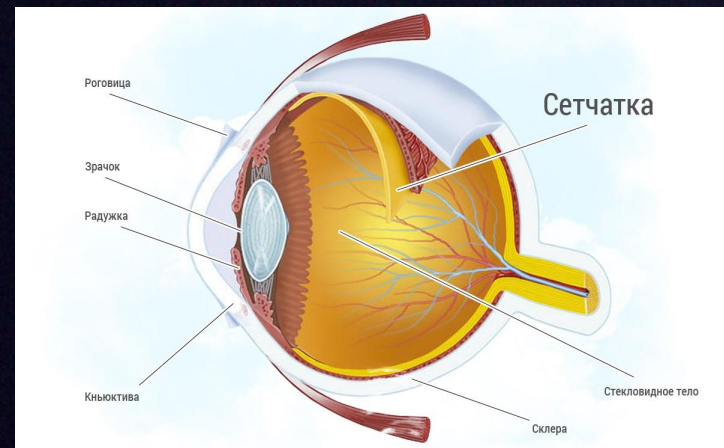
*А. Н. Островский. Доходное место*



# Модели с вниманием: шаг назад



С ранних шагов нейробиологии было понятно, что внимание — это сложно

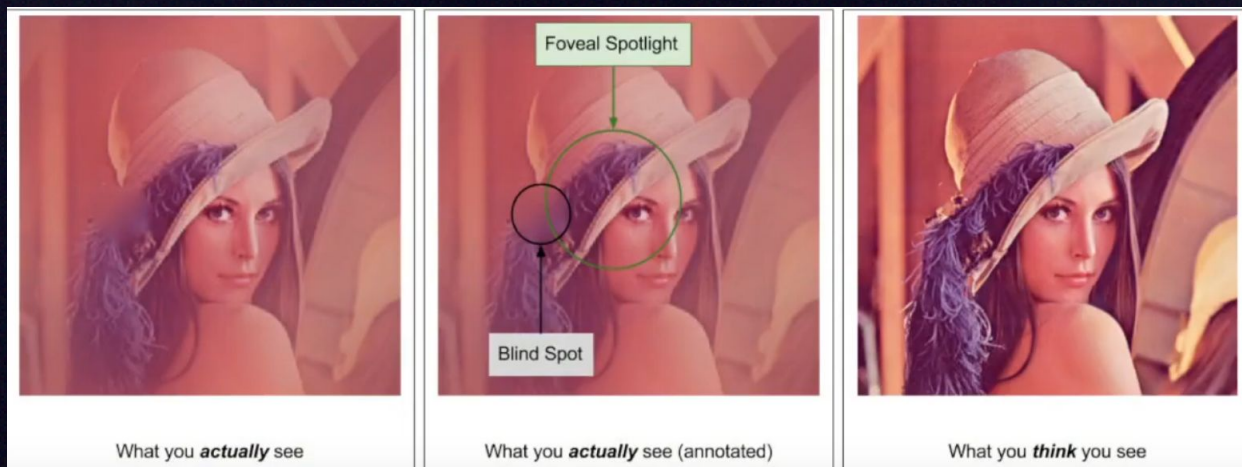




# Модели с вниманием: шаг назад

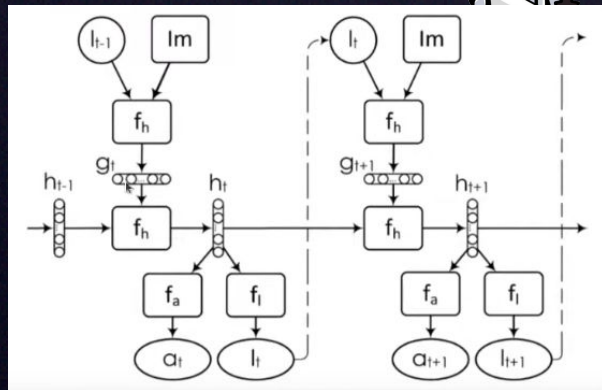


А вы внимательно смотрите на презентацию? Внимательно ли меня слушаете?  
С ранних шагов нейробиологии было понятно, что внимание — это сложно





# Recurrent models of visual attention 2014, Mnih et al.

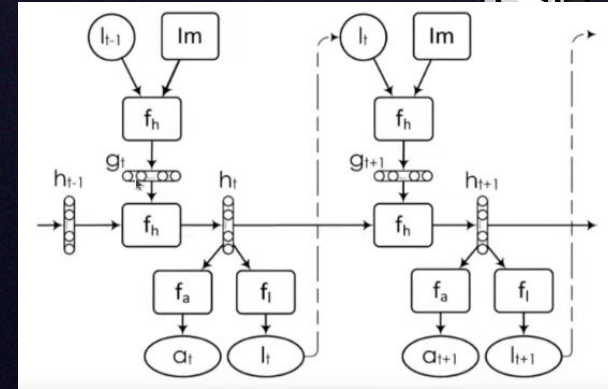


- каждый момент времени  $t$  на вход сети поступает предыдущее состояние  $h(t-1)$  и произведенное из него положение  $l$  для нового «взгляда»;
- этот новый «взгляд» с помощью функции  $f_h$  преобразуется в вектор признаков  $g_t$  (от слова *glimpse*), который служит входом на шаге  $t$ ;
- с помощью  $f$  получаем следующее скрытое состояние  $h$ , из него уже текущее «действие»  $a$  («действием» может быть, например, выдача ответа о том, какой объект удалось распознать) и положение следующего «взгляда»  $l+1$



# Recurrent visual attention

- Сеть должна произвести следующее действие, но при этом функция ошибки получается не сразу, а только после того, как все «взгляды» закончатся, и модель выдаст собственно ответ в виде очередного  $a$
- Нам нужно обучать последовательность взглядов, как нам это делать?



# Recurrent visual attention



- (Williams, 1992), алгоритм REINFORCE
- Погружаться в RL сегодня не будем

(a) 60x60 Cluttered Translated MNIST

Model	Error
FC, 2 layers (64 hidden each)	28.58%
FC, 2 layers (256 hidden each)	11.96%
Convolutional, 2 layers	8.09%
RAM, 4 glimpses, $12 \times 12$ , 3 scales	4.96%
RAM, 6 glimpses, $12 \times 12$ , 3 scales	4.08%
RAM, 8 glimpses, $12 \times 12$ , 3 scales	4.04%
RAM, 8 random glimpses	14.4%





# Машинный перевод

*Царь: Вызывает антирес*

*Ваш технический прогресс:*

*Как у вас там сеют брюкву —*

*С кожурою али без?..*

*Посол: Йес!*

Леонид Филатов. Сказка про Федота-  
стрельца, удалого молодца

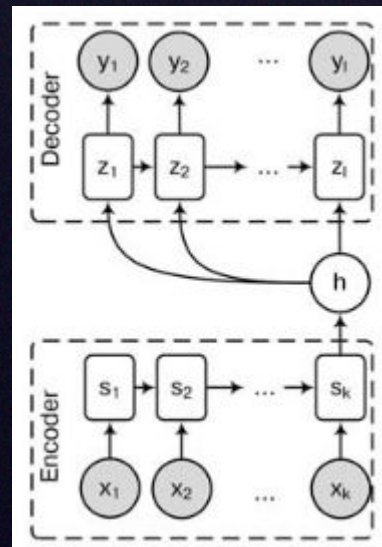




# Encoder-decoder (Sutskever et al., 2014; Cho et al. 2014)



- скрытое состояние используется как сжатое представление всей предшествующей истории
- следующее слово мы предсказываем из этого сжатого представления
- кодировщик сжимает некий вход в распределенное представление  $z$ , и оно подается каждый раз на вход рекуррентной сети, которая работает декодировщиком
- **Какой важный недостаток?**

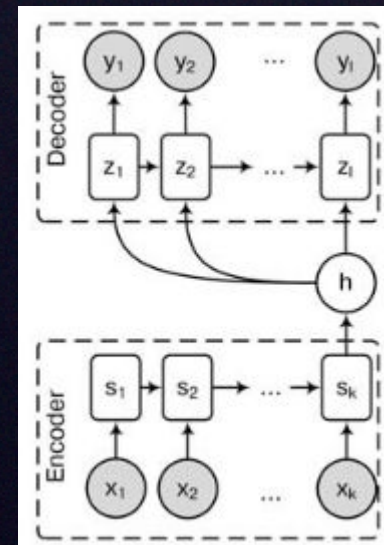




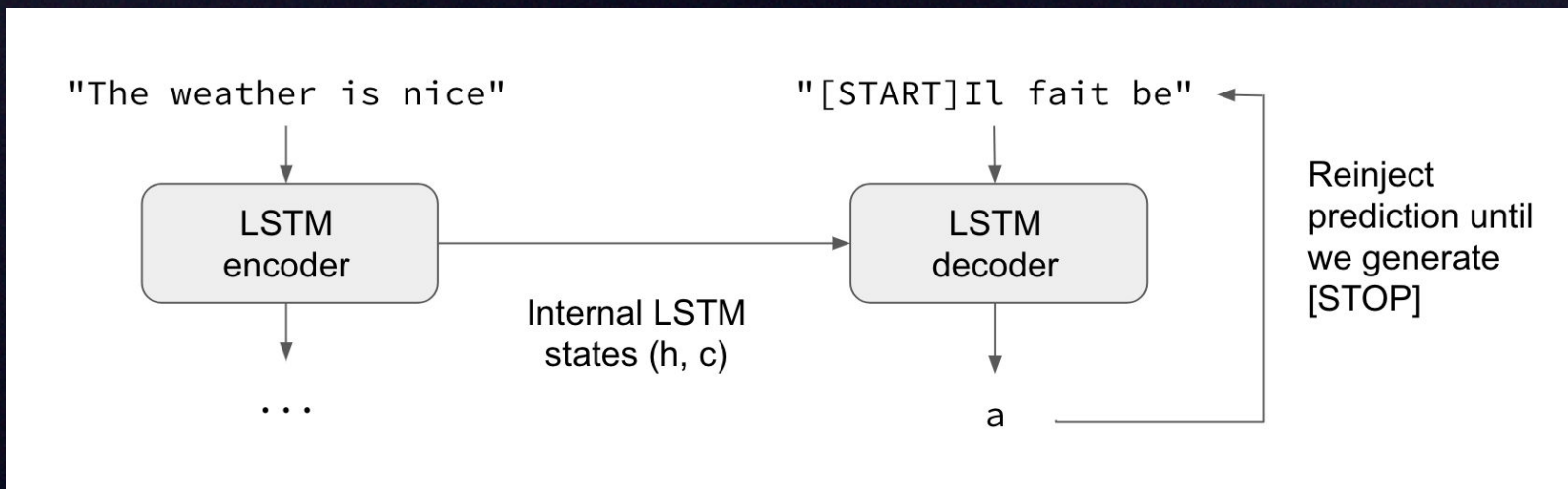
# Encoder-decoder



- С увеличением длины входа качество падает радикально из-за попытки уместить всё в один вектор
- Скрытое пространство тоже ограничено, тяжело увеличивать его размерность



# Encoder-decoder

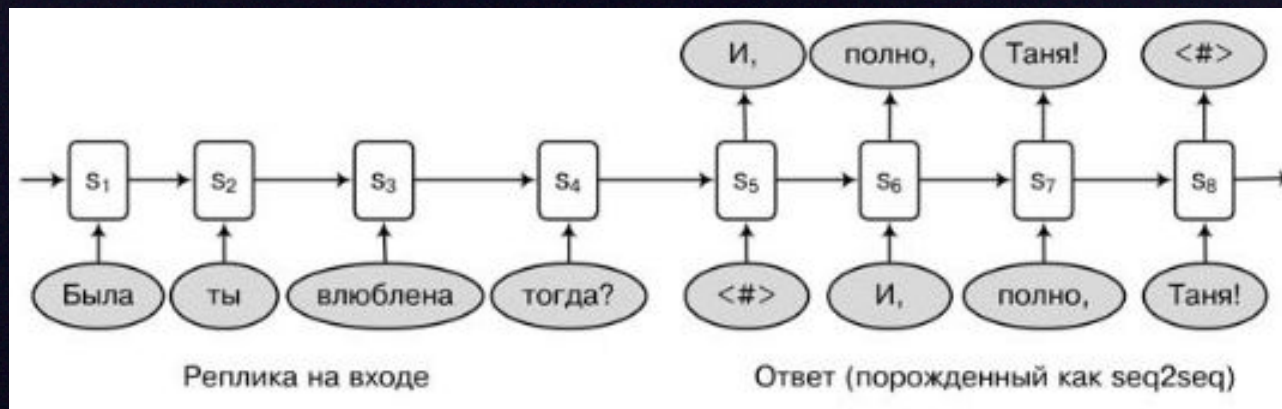




# Порождающие системы encoder-decoder



- Нет контекста между диалогами

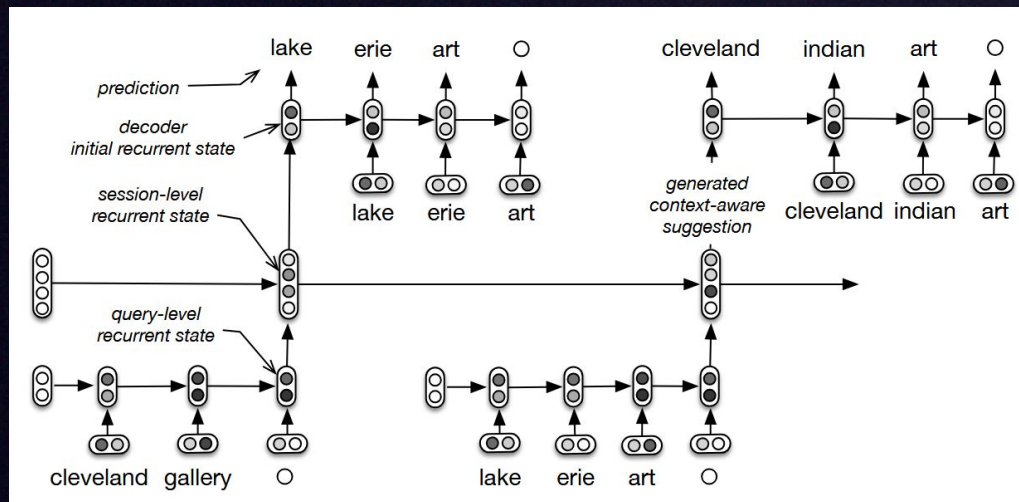


# HRED (Sordoni et al., 2015)



Диалог как двухуровневая система: последовательность высказываний  $\leftarrow$  последовательность слов

- Encoder RNN для сворачивания каждого высказывания
- Context RNN для вектора контекста
- Decoder RNN последовательно предсказывает слова системы



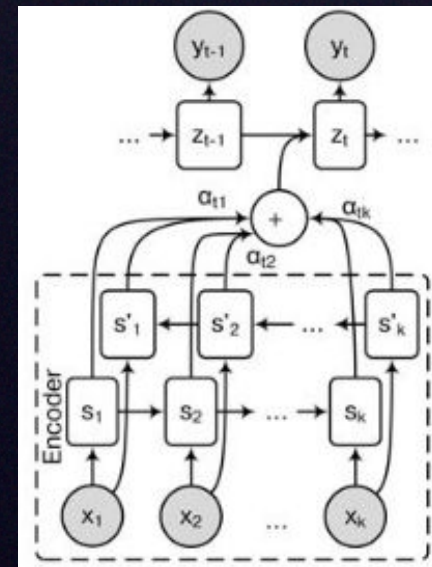


# Внимание (Bahdanau et al. 2014)



Решение проблем с длинным контекстом:

- Soft alignment model выдает веса  $a$ , которые показывают, насколько та или иная часть входа важна для текущего выхода
- $h$  для каждого слова получается с помощью двунаправленной LSTM
- всю модель можно обучать одновременно





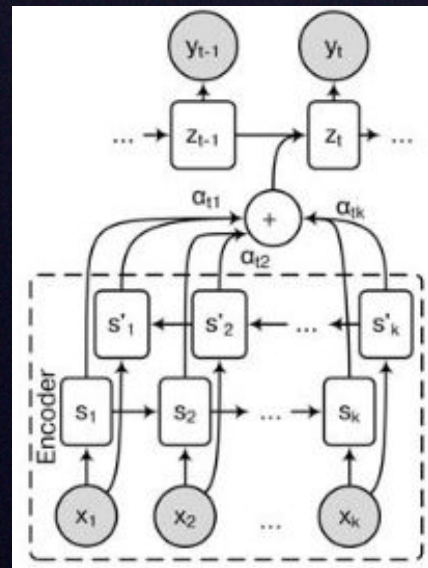
# Внимание! Это внимание



Решение:

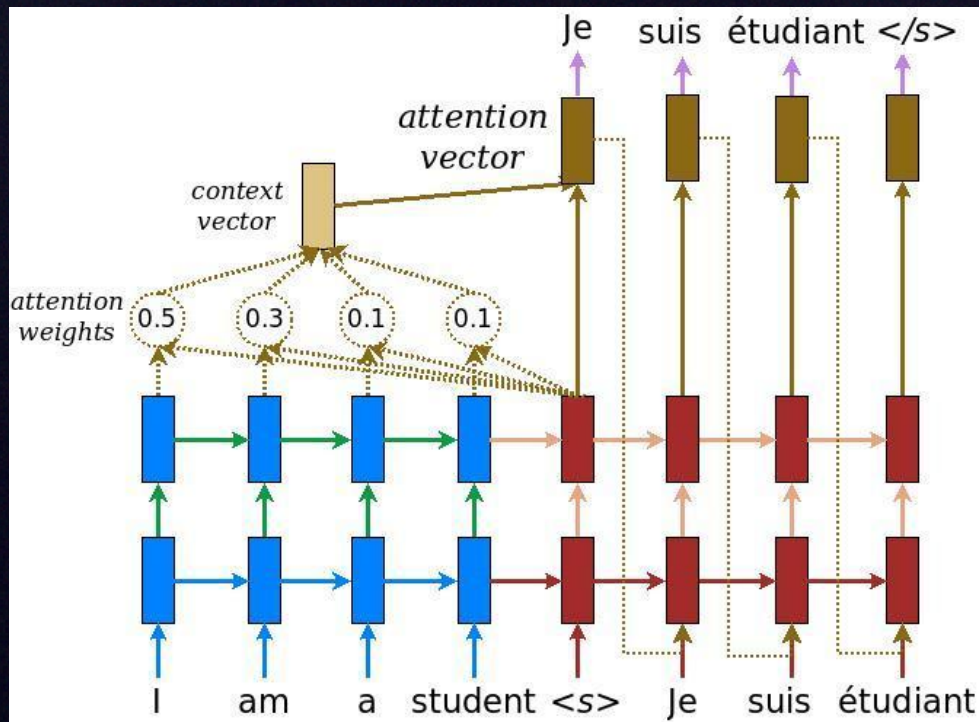
- Soft alignment model выдает веса  $a$ , которые показывают, насколько та или иная часть входа важна для текущего выхода
- $h$  для каждого слова получается с помощью двунаправленной LSTM
- всю модель можно обучать одновременно

Декодер получает на вход линейную комбинацию представлений каждого  $x$ , веса которые меняются со временем





# Внимание! Это внимание



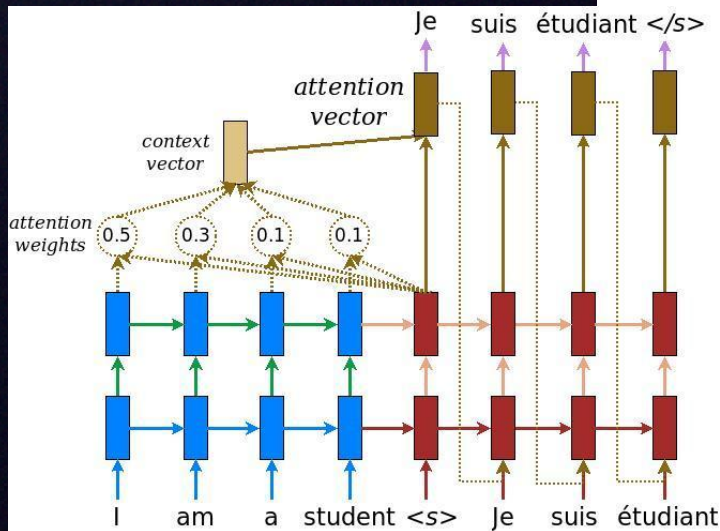
# Внимание! Это внимание



$$\alpha_{ts} = \frac{\exp(\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_s))}{\sum_{s'=1}^S \exp(\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_{s'}))} \quad \text{[Attention weights]} \quad (1)$$

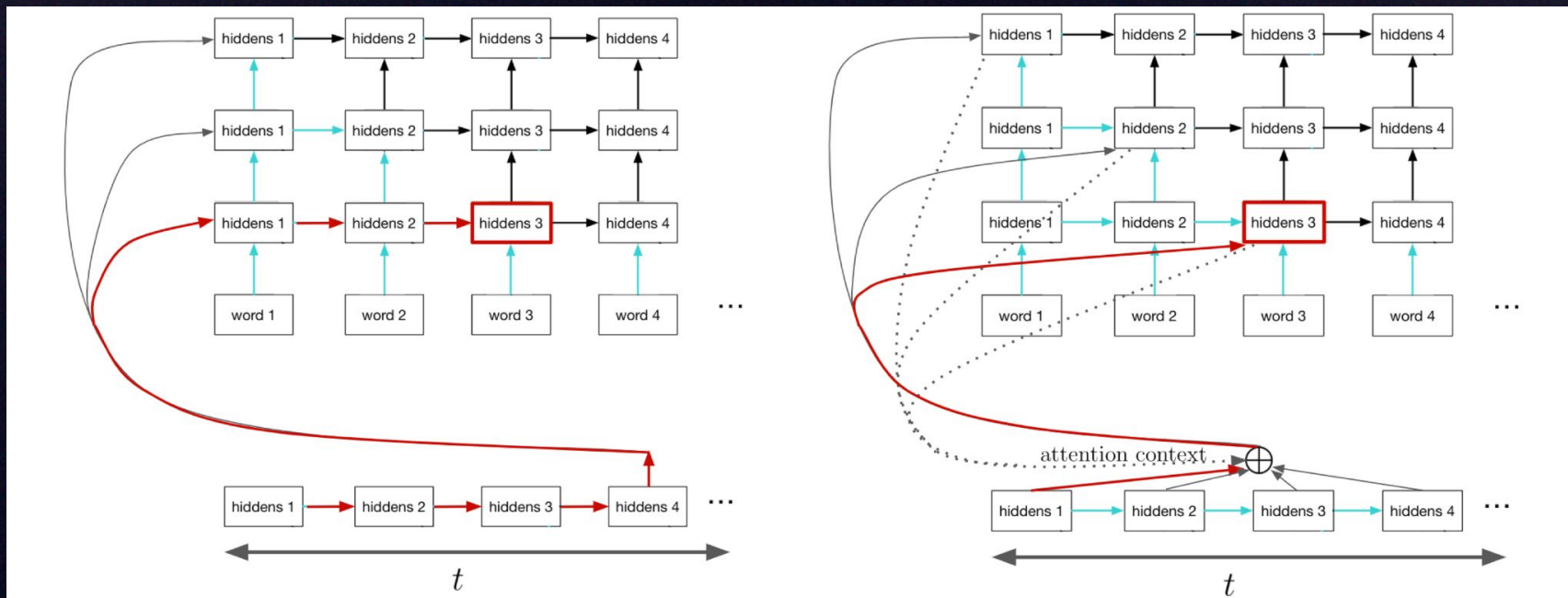
$$\mathbf{c}_t = \sum_s \alpha_{ts} \bar{\mathbf{h}}_s \quad \text{[Context vector]} \quad (2)$$

$$\mathbf{a}_t = f(\mathbf{c}_t, \mathbf{h}_t) = \tanh(\mathbf{W}_c[\mathbf{c}_t; \mathbf{h}_t]) \quad \text{[Attention vector]} \quad (3)$$





# Внимание! Это внимание



# Внимание! Это внимание



- $t$ : sequence length,  $d$ : # layers and  $k$ : # neurons at each layer.

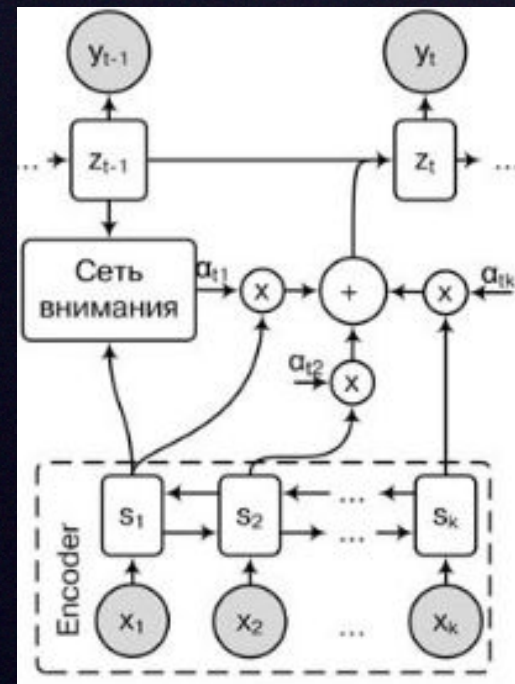
Model	training complexity	training memory	test complexity	test memory
RNN	$t \times k^2 \times d$	$t \times k \times d$	$t \times k^2 \times d$	$k \times d$
RNN+attn.	$t^2 \times k^2 \times d$	$t^2 \times k \times d$	$t^2 \times k^2 \times d$	$t \times k \times d$



# Мягкое внимание (Luong et al. 2015a, Jean et al. 2015)



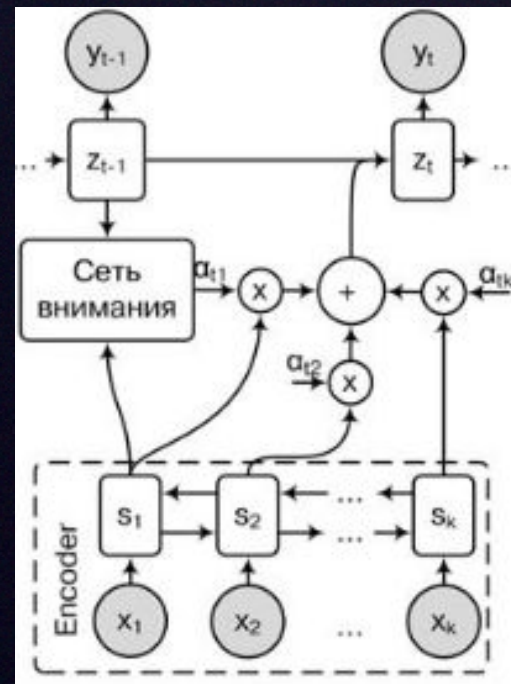
- Encoder – двунаправленная RNN, есть оба контекста
- Сеть внимания дает оценку релевантности



# Мягкое внимание

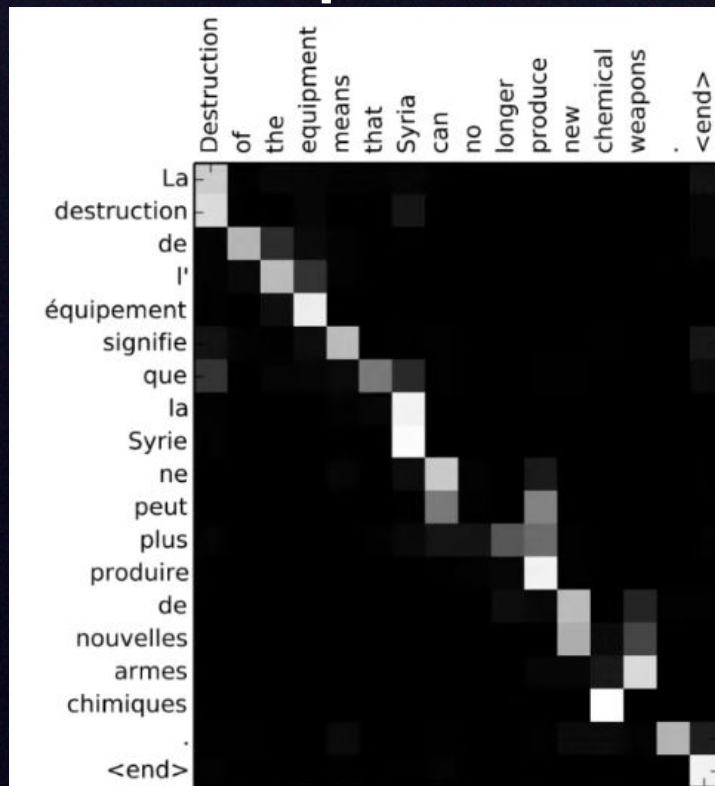


- Снова end-to-end обучение, лишь небольшой трюк
- Тем не менее, сеть внимания работает так, как мы хотим





# Становится понятнее, что происходит



# Лучше с порядком слов



Economic growth has slowed down in recent years .

Das Wirtschaftswachstum hat sich in den letzten Jahren verlangsamt .

Economic growth has slowed down in recent years .

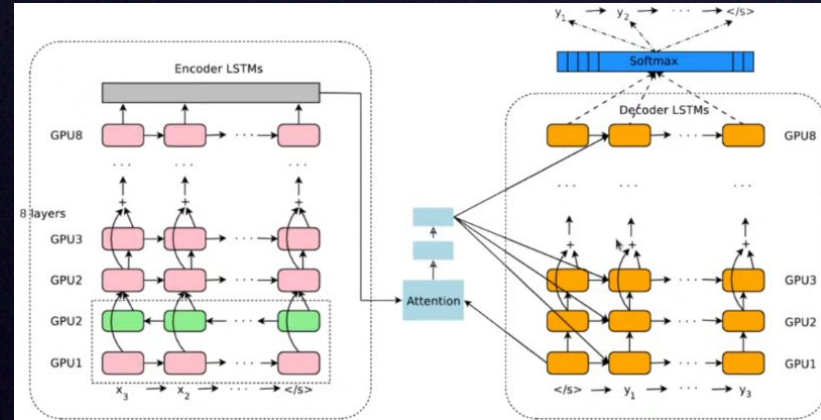
La croissance économique s' est ralentie ces dernières années .



# Google Translate 2016



- в GNMT используются по восемь уровней LSTM в кодировщике и декодировщике
- но помогает ли это?

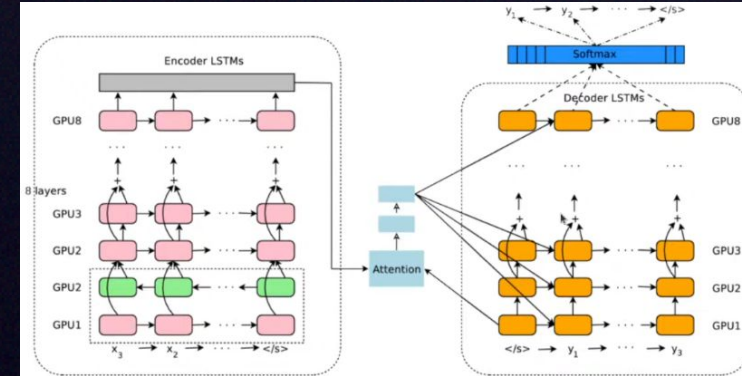




# Google Translate 2016

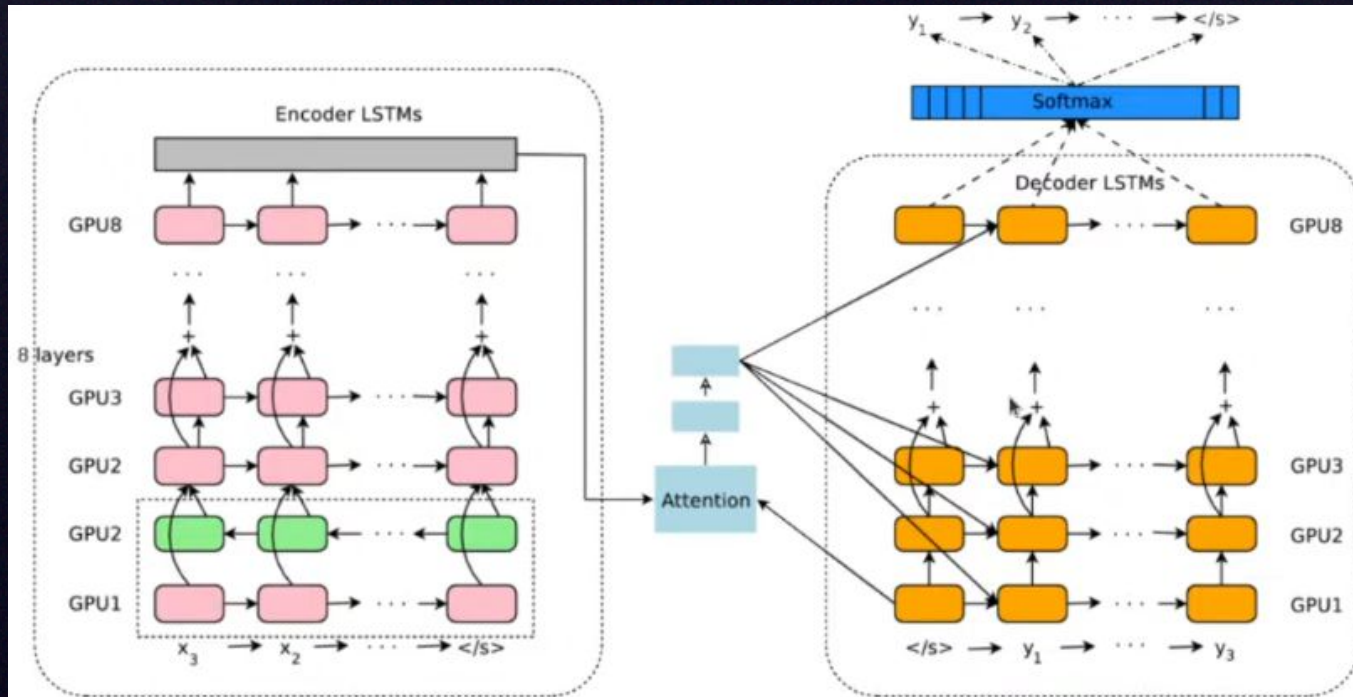


- в GNMT используются по восемь уровней LSTM в кодировщике и декодировщике
- добавляются остаточные связи между уровнями, как в ResNet
- нижний слой двунаправленный, чтобы был контекст слева и справа
- + сегментация слов





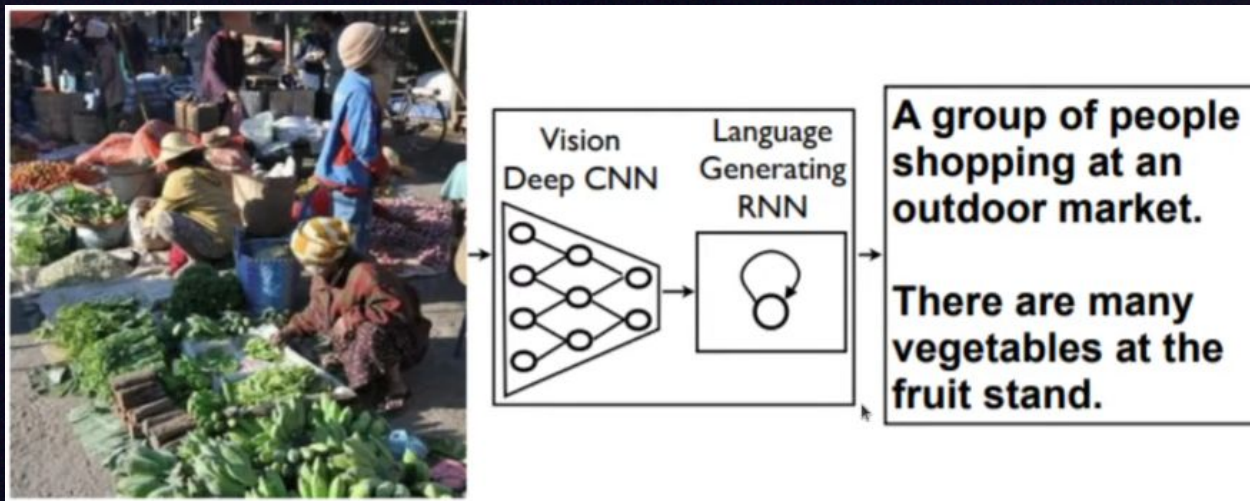
# Google Translate 2016



# Show and Tell (Vinyals et al. 2015)

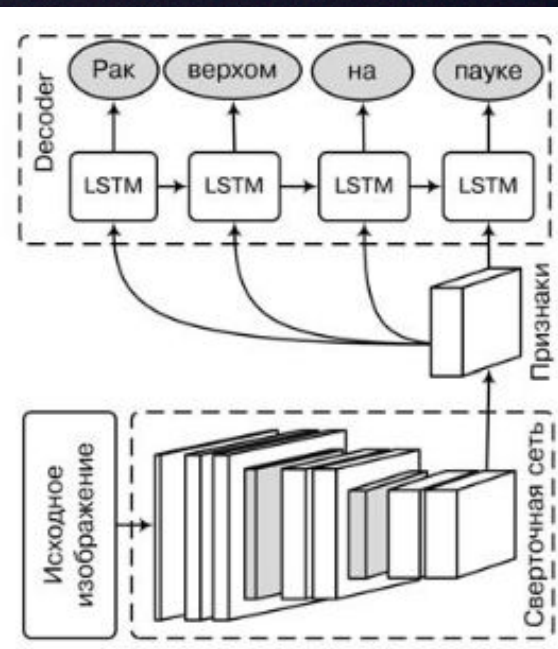


- Давайте делать подписи к картинкам





# Show and Tell

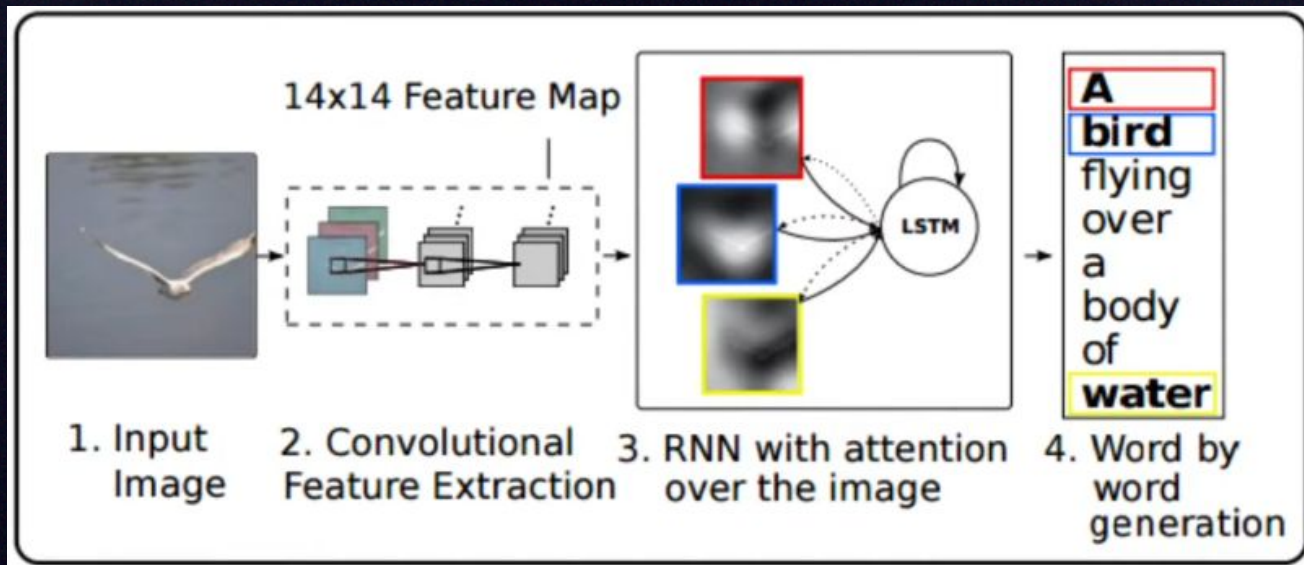


<p>A person riding a motorcycle on a dirt road.</p>	<p>Two dogs play in the grass.</p>	<p>A skateboarder does a trick on a ramp.</p>	<p>A dog is jumping to catch a frisbee.</p>
<p>A group of young people playing a game of frisbee.</p>	<p>Two hockey players are fighting over the puck.</p>	<p>A little girl in a pink hat is blowing bubbles.</p>	<p>A refrigerator filled with lots of food and drinks.</p>
<p>A herd of elephants walking across a dry grass field.</p>	<p>A close up of a cat laying on a couch.</p>	<p>A red motorcycle parked on the side of the road.</p>	<p>A yellow school bus parked in a parking lot.</p>
Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image

# Show, Attend, and Tell (Xu et al. 2015)



- Внимание может быть на что угодно!



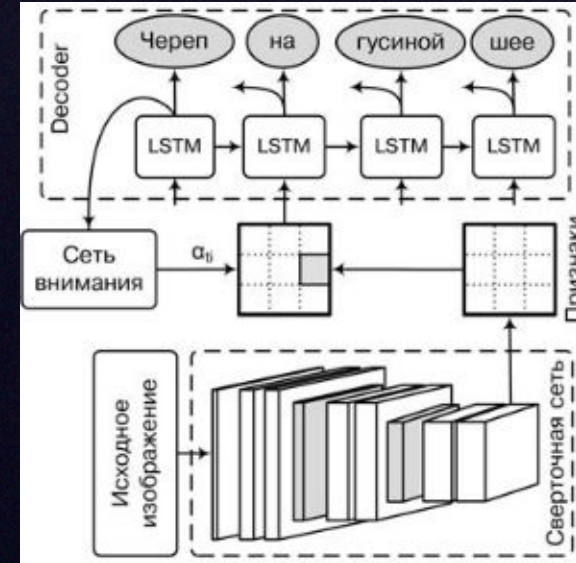


# Soft x Hard attention

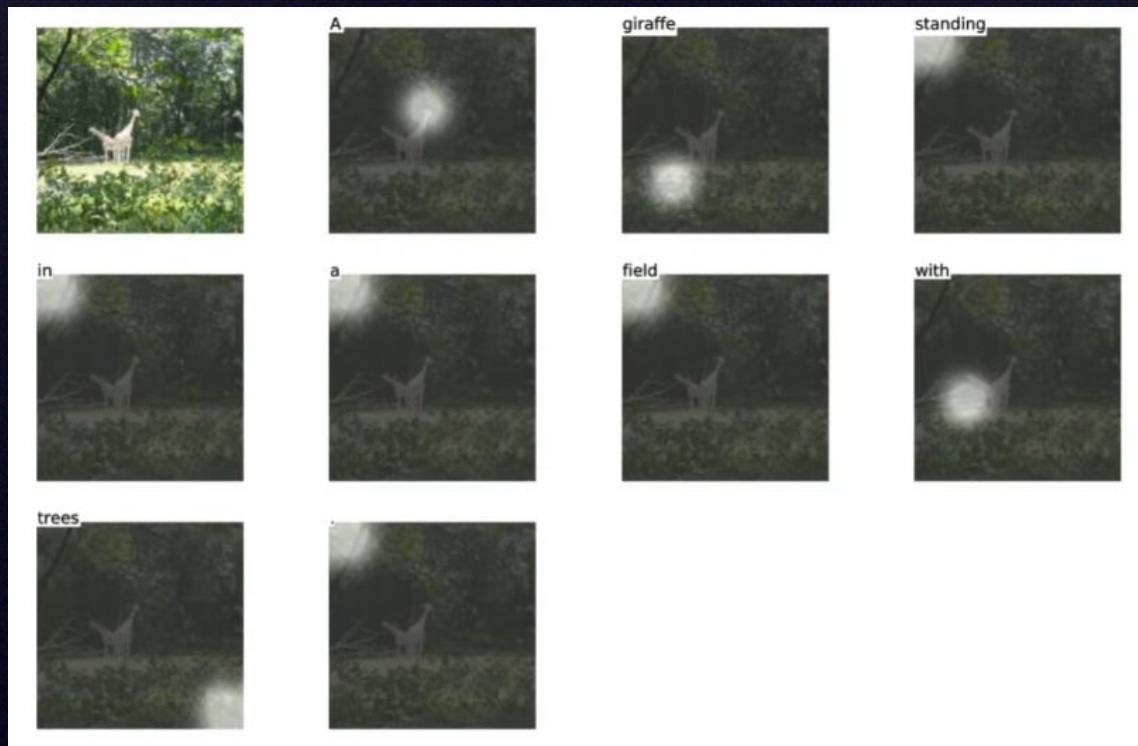


$$L_s = \sum_s p(s | \mathbf{a}) \log p(\mathbf{y} | s, \mathbf{a}) \leq \log \sum_s p(s | \mathbf{a}) p(\mathbf{y} | s, \mathbf{a}) = \log p(\mathbf{y} | \mathbf{a})$$

- Hard attention обучается максимизацией вариационной нижней оценки
- веса  $\alpha$  интерпретируются как вероятности событий  $s$  того, что модель «посмотрит» в момент времени  $t$  на часть изображения  $i$ , и на вход рекуррентной сети для порождения текста подается представление той части изображения  $\mathbf{a}^*$ , которая выпадет на кубике с вероятностями  $\mathbf{a}$



# Soft x Hard attention



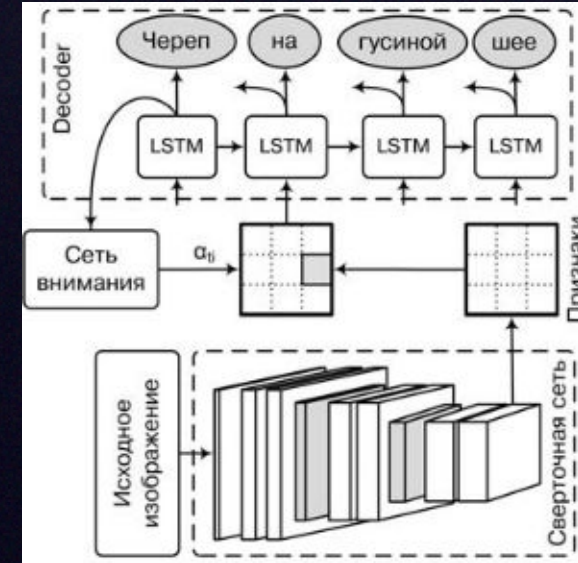


# Soft x Hard attention



$$\sum_{i=1}^L \alpha_{t,i} \mathbf{a}_i$$

- В мягком на вход RNN подается *ожидаие* вектора из  $s$
- Модель на каждом шаге “смотрит” на все части изображения, но некоторые из них более важны

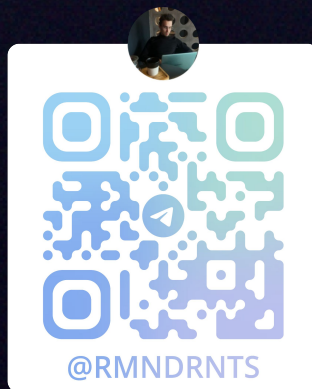


# Soft x Hard attention





# Внимание! Спасибо за внимание!

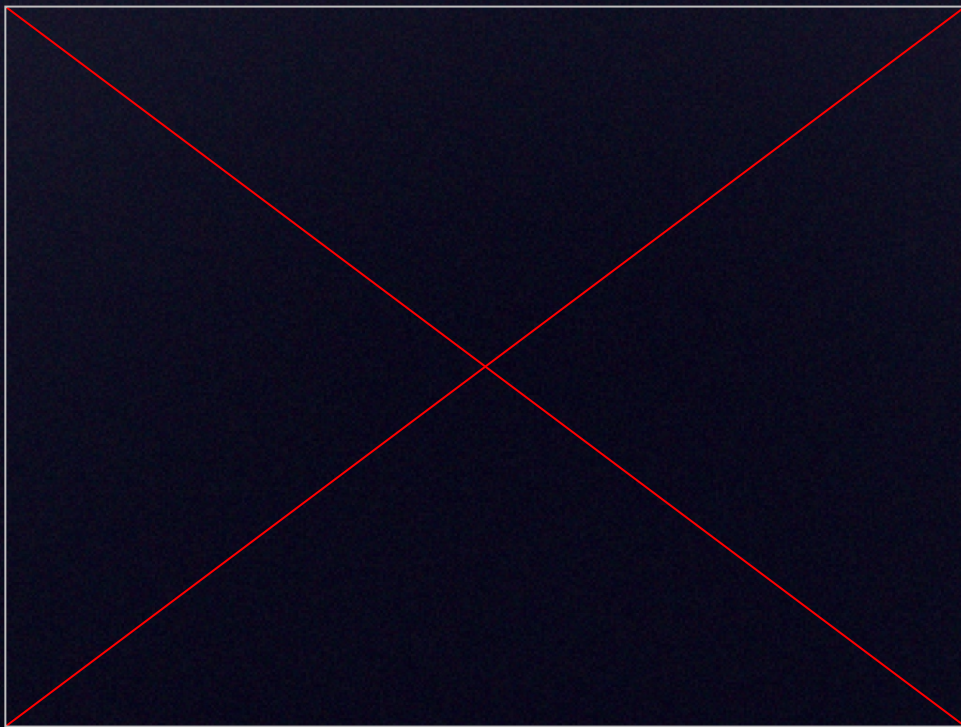


📍 **Дерунец Роман**

**N\*** Новосибирский  
государственный  
университет  
**\*НАСТОЯЩАЯ НАУКА**



# Внимание! Спасибо за внимание!





# Внимание! Спасибо за внимание!



## FOCUS EXPERIMENT

Based on the findings of J M Tangen, S C Murphy, M B Thompson in  
'Flashed face distortion effect', Perception, 40 (2011) p628-630



[www.ToThePointAtWork.com](http://www.ToThePointAtWork.com)

**N\***Новосибирский  
государственный  
университет  
**\*НАСТОЯЩАЯ НАУКА**