OCR

Capcha recognition and other problems

Alexander Barmin

Области применения OCR

- Распознавание автомобильных номеров
- Распознавание документов
- Создание электронных версий бумажных документов
- Перевод текста по фото
- …

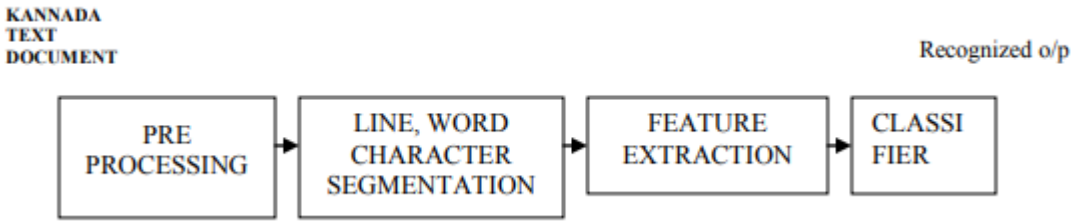
Самый первый в OCR

Пионер!!!!

Густав Таушек

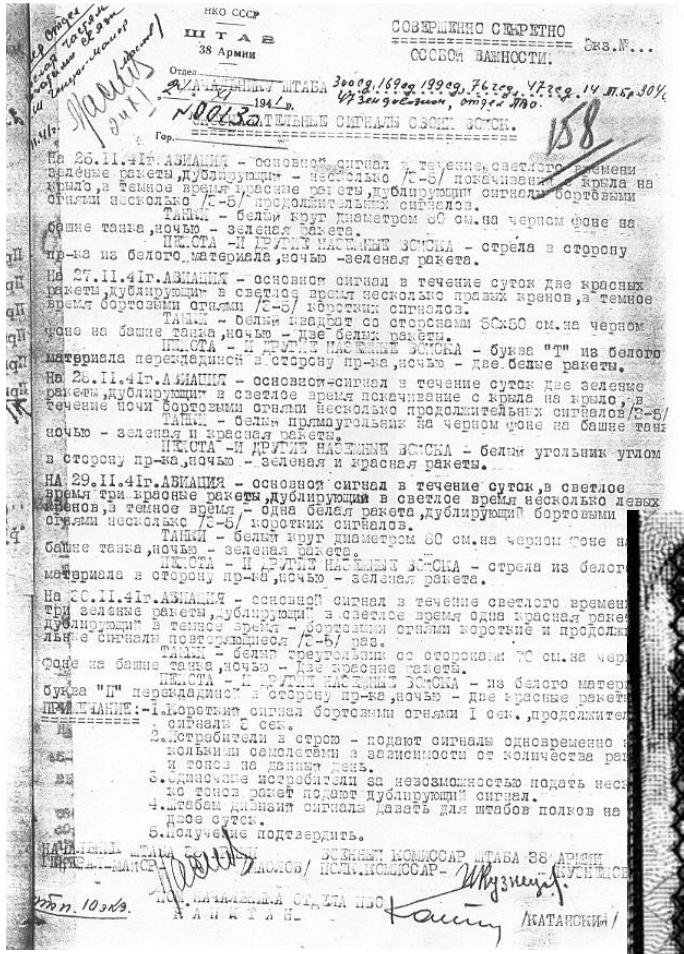
Эпоха до DL

- Сегментация изображения, выделение отдельных слов/символов
- Препроцессинг – бинаризация, морфологические операции, выделение признаков (любого рода)
- Классические методы МК (KNN, SVM, Bayesian)
- N-граммы вместо NLP



Kannada Character Recognition System: A Review

Качество сканирования



НКО СССР
ШТАБ
38 Армии
Отдел
194 г.
Гор.

СОВЕРШЕННО СЕКРЕТНО
Экз.№...
ОСОБОЙ ВАЖНОСТИ.

ДЛЯ ЗАМЕНЫ ШТАБА 300сд.169сд.199сд.76гсд.47гсд.14 т.бр.30ч.
47 Зен.дивизион, отдел Мо.

№0013 ...ТЕЛЬНЫЕ СИГНАЛЫ СВОИХ ВОЙСК.

158

На 26.II.41г.АБМАЦИЯ - основной сигнал в течение светлого времени зелёные ракеты,дублирующим - несколько /3-5/ покачиваний крыла на крыло,а темное время красные ракеты,дублирующий сигнал бортовыми огнями несколько /3-5/ продолжительных сигналов.
ТАНКИ - белый круг диаметром 80 см.на черном фоне на башне танка,ночью - зеленая ракета.
ПЕХОТА -И ДРУГИЕ НАЗЕМНЫЕ ВОЙСКА - стрела в сторону пр-ка из белого материала,ночью -зеленая ракета.

На 27.II.41г.АБМАЦИЯ - основной сигнал в течение суток две красных ракеты,дублирующим в светлое время несколько правых кренов,а темное время бортовыми огнями /3-5/ коротких сигналов.
ТАНКИ - белый квадрат со сторонами 80х80 см.на черном фоне на башне танка,ночью - две белых ракеты.
ПЕХОТА - И ДРУГИЕ НАЗЕМНЫЕ ВОЙСКА - буква "Г" из белого материала перекладиной в сторону пр-ка,ночью - две белые ракеты.

На 28.II.41г.АБМАЦИЯ - основной сигнал в течение суток две зеленые ракеты,дублирующим в светлое время покачивание с крыла на крыло,а в течение ночи бортовыми огнями несколько продолжительных сигналов /2-5/
ТАНКИ - белый прямоугольник на черном фоне на башне танка ночью - зеленая и красная ракеты.
ПЕХОТА -И ДРУГИЕ НАЗЕМНЫЕ ВОЙСКА - белым угольник углом в сторону пр-ка,ночью - зеленая и красная ракеты.

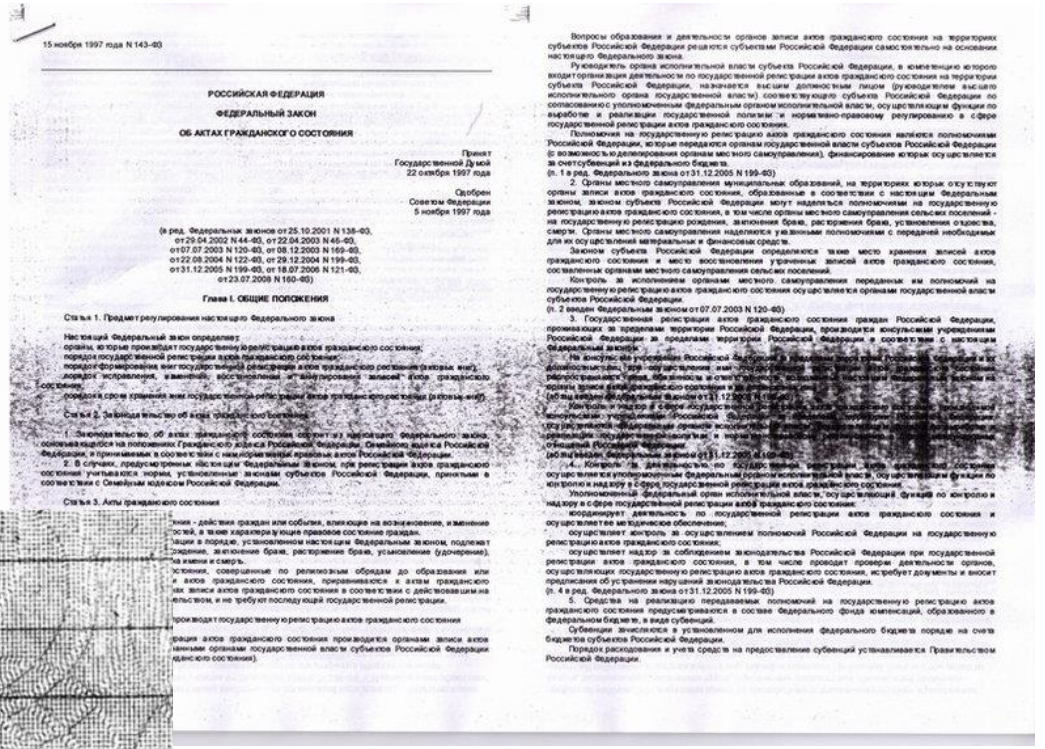
НА 29.II.41г.АБМАЦИЯ - основной сигнал в течение суток,в светлое время три красные ракеты,дублирующий в светлое время несколько левых кренов,а темное время - одна белая ракета,дублирующий бортовыми огнями несколько /3-5/ коротких сигналов.
ТАНКИ - белый круг диаметром 80 см.на черном фоне на башне танка,ночью - зеленая ракета.
ПЕХОТА - И ДРУГИЕ НАЗЕМНЫЕ ВОЙСКА - стрела из белого материала в сторону пр-ка,ночью - зеленая ракета.

На 30.II.41г.АБМАЦИЯ - основной сигнал в течение светлого времени три зеленые ракеты,дублирующий в светлое время одна красная ракета дублирующим в темное время - бортовыми огнями короткие и продолжительные сигналы повторяющиеся /3-5/ раз.
ТАНКИ - белый треугольник со сторонами 70 см.на черном фоне на башне танка,ночью - две красные ракеты.
ПЕХОТА - И ДРУГИЕ НАЗЕМНЫЕ ВОЙСКА - из белого материала буква "Г" перекладиной в сторону пр-ка,ночью - две красные ракеты.

ПРИМЕЧАНИЕ:-1.Короткий сигнал бортовыми огнями 1 сек.,продолжительный сигналы 5 сек.
2.Истребители в строю - подают сигналы одновременно несколькими самолетами в зависимости от количества ракет и тонов на данный день.
3.Одиночные истребители за невозможностью подать несколько тонов ракет подают дублирующий сигнал.
4.Штабам дивизий сигналы давать для штабов полков на двое суток.
5.Получение подтвердить.

НАЧАЛЬНИК ШТАБА 5...
ГЕНЕРАЛ-МАЙОР ...ЛОВ/
ВОЕННЫЙ КОМИССАР ШТАБА 38 АРМИИ
ИСПР.КОМИССАР- /КУЗНЕЦОВ/
НАЧАЛЬНИК ОТДЕЛА НКО
/КАТАНСКИЙ/

тп. 10 экз.



15 ноября 1997 года N 143-ФЗ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ФЕДЕРАЛЬНЫЙ ЗАКОН

ОБ АКТАХ ГРАЖДАНСКОГО СОСТОЯНИЯ

Принят
Государственной Думой
22 октября 1997 года

Одобрен
Советом Федерации
5 ноября 1997 года

(в ред. Федеральных законов от 25.10.2001 N 138-ФЗ,
от 29.04.2002 N 44-ФЗ, от 22.04.2003 N 45-ФЗ,
от 07.07.2003 N 120-ФЗ, от 08.12.2003 N 169-ФЗ,
от 22.08.2004 N 122-ФЗ, от 29.12.2004 N 199-ФЗ,
от 31.12.2005 N 199-ФЗ, от 18.07.2006 N 121-ФЗ,
от 23.07.2008 N 160-ФЗ)

Глава I. ОБЩИЕ ПОЛОЖЕНИЯ

Статья 1. Предмет регулирования настоящего Федерального закона

Настоящий Федеральный закон определяет:
органы, которые производят государственную регистрацию актов гражданского состояния;
порядок государственной регистрации актов гражданского состояния;
порядок формирования книг государственной регистрации актов гражданского состояния (актовых книг),
порядок исправления, изменения, восстановления и аннулирования записей актов гражданского состояния;
порядок и сроки хранения книг государственной регистрации актов гражданского состояния (актовых книг).

Статья 2. Законодательство об актах гражданского состояния

1. Законодательство об актах гражданского состояния состоит из настоящего Федерального закона, основанных на положениях Гражданского кодекса Российской Федерации, Семейного кодекса Российской Федерации, а принимаемых в соответствии с ними нормативных правовых актов Российской Федерации.

2. В случаях, предусмотренных настоящим Федеральным законом, при регистрации актов гражданского состояния учитываются нормы, установленные законами субъектов Российской Федерации, принятыми в соответствии с Семейным кодексом Российской Федерации.

Статья 3. Акты гражданского состояния

...ники - действия граждан или события, влекущие на возникновение, изменение ...остей, а также характеризующие правовое состояние граждан.
...ации в порядке, установленном настоящим Федеральным законом, подлежат ...ождение, заключение брака, расторжение брака, усыновление (удочерение), ...ие имени и смерть.
...остояния, совершенные по религиозным обрядам до образования или ...х актов гражданского состояния, приравниваются к актам гражданского ...ких записи актов гражданского состояния в соответствии с действовавшими на ...льством, и не требуют последующей государственной регистрации.

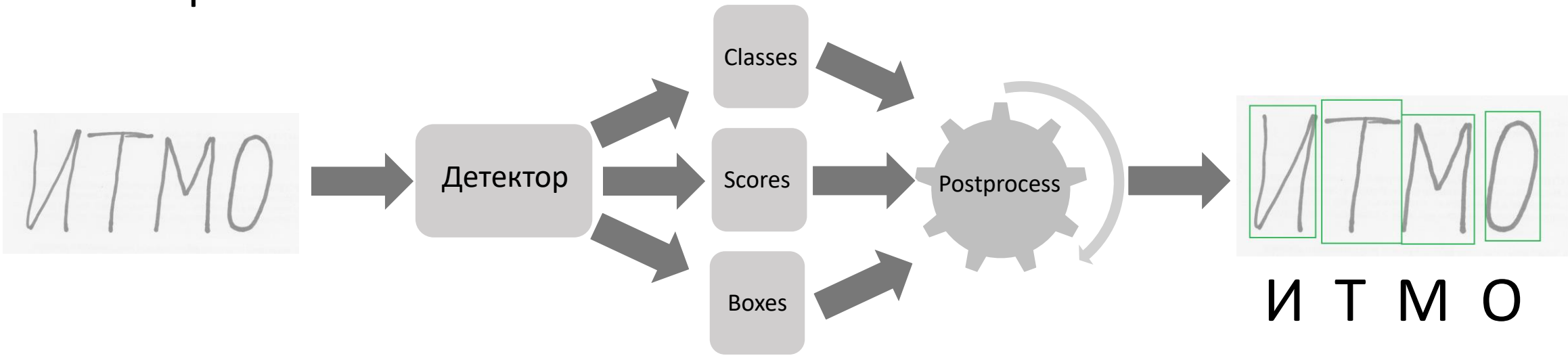
...роизводят государственную регистрацию актов гражданского состояния

...рация актов гражданского состояния производится органами записи актов ...ными органами государственной власти субъектов Российской Федерации ...данского состояния).

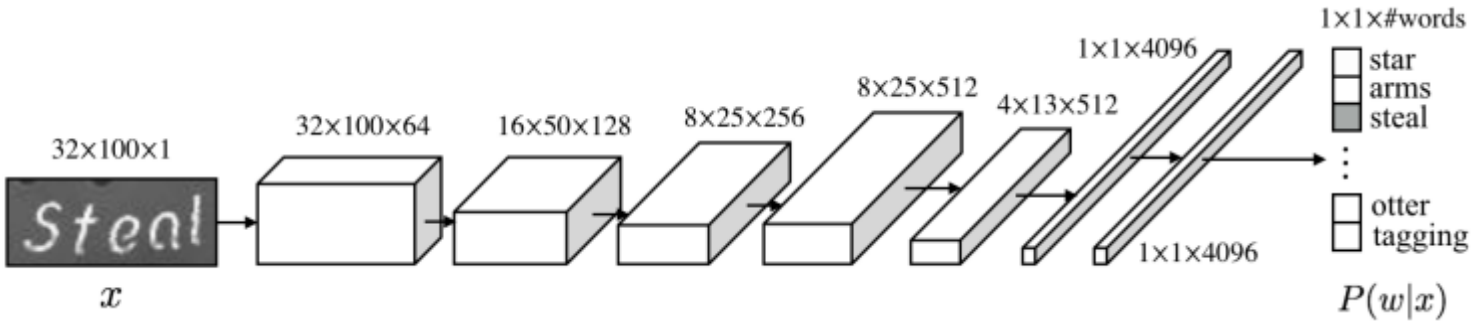
Вопросы образования и деятельности органов записи актов гражданского состояния на территориях субъектов Российской Федерации решаются субъектами Российской Федерации самостоятельно на основании настоящего Федерального закона.
Руководитель органа исполнительной власти субъекта Российской Федерации, в компетенцию которого входит организация деятельности по государственной регистрации актов гражданского состояния на территории субъекта Российской Федерации, назначается высшим должностным лицом (руководителем высшего исполнительного органа государственной власти) соответствующего субъекта Российской Федерации по согласованию с уполномоченным федеральным органом исполнительной власти, осуществляющим функции по выработке и реализации государственной политики и нормативно-правовому регулированию в сфере государственной регистрации актов гражданского состояния.
Полномочия на государственную регистрацию актов гражданского состояния являются полномочиями Российской Федерации, которые передаются органам государственной власти субъектов Российской Федерации (с возможностью их делегирования органам местного самоуправления), финансирование которых осуществляется за счет субвенций из федерального бюджета.
(п. 1 в ред. Федерального закона от 31.12.2005 N 199-ФЗ)
2. Органы местного самоуправления муниципальных образований, на территориях которых отсутствуют органы записи актов гражданского состояния, образованные в соответствии с настоящим Федеральным законом, законом субъекта Российской Федерации могут наделяться полномочиями на государственную регистрацию актов гражданского состояния, в том числе органы местного самоуправления сельских поселений - на государственную регистрацию рождения, заключения брака, расторжения брака, установления отцовства, смерти. Органы местного самоуправления наделяются указанными полномочиями с передачей необходимых для их осуществления материальных и финансовых средств.
Законом субъекта Российской Федерации определяются также место хранения записей актов гражданского состояния и место восстановления утраченных записей актов гражданского состояния, составленных органами местного самоуправления сельских поселений.
Контроль за исполнением органами местного самоуправления переданных им полномочий на государственную регистрацию актов гражданского состояния осуществляется органами государственной власти субъектов Российской Федерации.
(п. 2 введен Федеральным законом от 07.07.2003 N 120-ФЗ)
3. Государственная регистрация актов гражданского состояния граждан Российской Федерации, проживающих за пределами территории Российской Федерации, производится консульскими учреждениями Российской Федерации за пределами территории Российской Федерации в соответствии с настоящим Федеральным законом.
На консульские учреждения Российской Федерации [illegible]
...ельствах), при осуществлении ими [illegible]
...распространяется время, обозначенное в отчетных документах [illegible]
...ации записи актов гражданского состояния и ее [illegible]
(об [illegible] Федеральным законом от 31.12.2005 N 199-ФЗ)
...Контроль за деятельностью по государственной регистрации актов гражданского состояния, ...осуществляется уполномоченным федеральным органом исполнительной власти, осуществляющим функции по ...контролю и надзору в сфере государственной регистрации актов гражданского состояния.
Уполномоченный федеральный орган исполнительной власти, осуществляющий функции по контролю и надзору в сфере государственной регистрации актов гражданского состояния:
...координирует деятельность по государственной регистрации актов гражданского состояния и ...осуществляет методическое обеспечение;
...осуществляет контроль за осуществлением полномочий Российской Федерации на государственную регистрацию актов гражданского состояния;
осуществляет надзор за соблюдением законодательства Российской Федерации при государственной регистрации актов гражданского состояния, в том числе проводит проверки деятельности органов, осуществляющих государственную регистрацию актов гражданского состояния, истребует документы и вносит предписания об устранении нарушений законодательства Российской Федерации.
(п. 4 в ред. Федерального закона от 31.12.2005 N 199-ФЗ)
5. Средства на реализацию передаваемых полномочий на государственную регистрацию актов гражданского состояния предусматриваются в составе федерального фонда компенсаций, образованного в федеральном бюджете, в виде субвенций.
Субвенции зачисляются в установленном для исполнения федерального бюджета порядке на счета бюджетов субъектов Российской Федерации.
Порядок расходования и учета средств на предоставление субвенций устанавливается Правительством Российской Федерации.

Применяем DL “в лоб”

- Обучаем мультиклассовый детектор
- Сортируем боксы по оси x
- Собираем ответ



Применяем DL “в лоб”

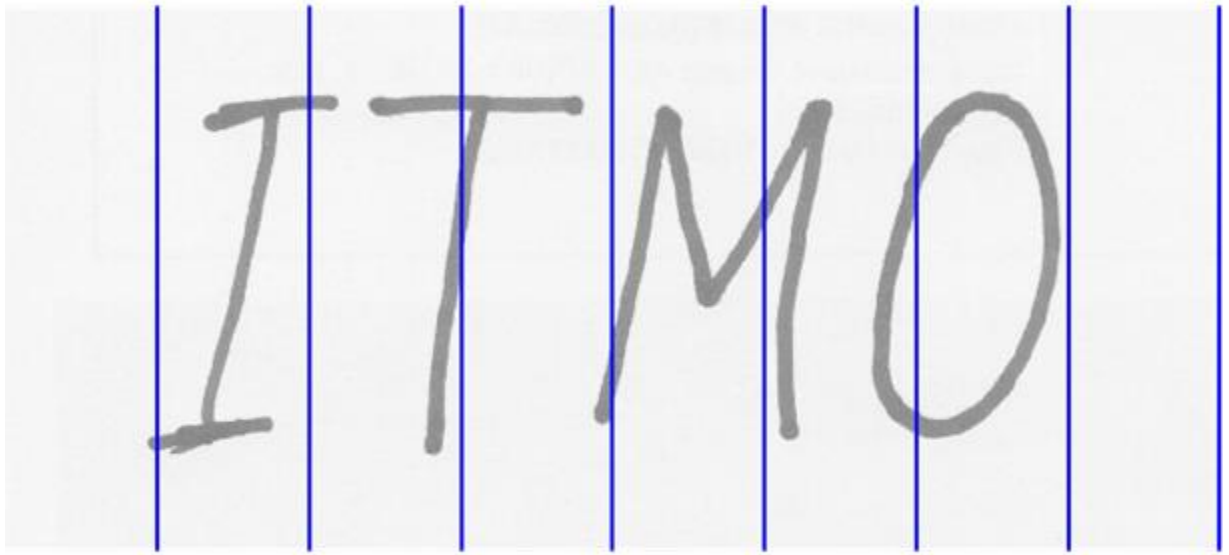


Распознавание 90K слов

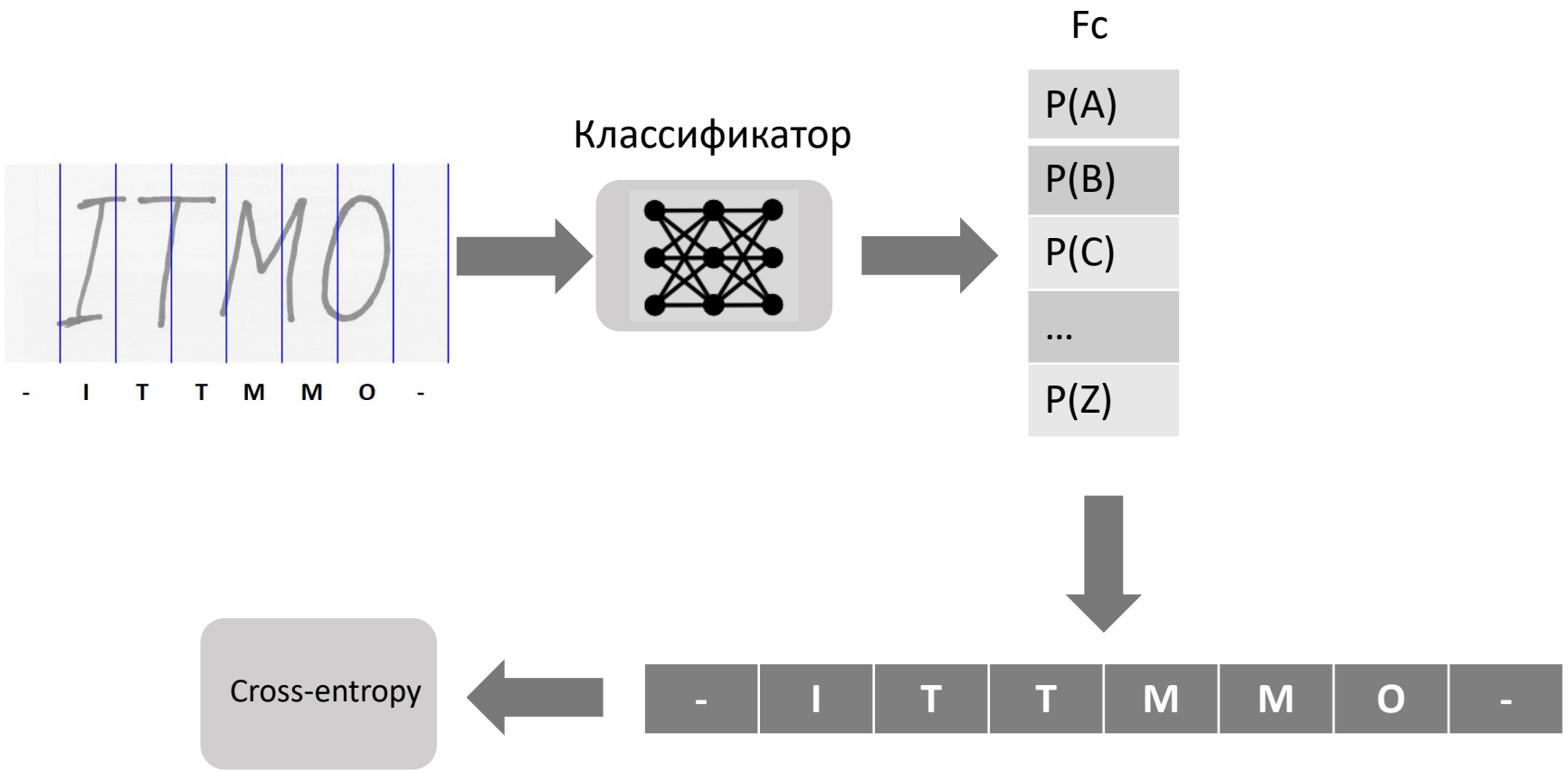
[Reading Text in the Wild with Convolutional Neural Networks]

Какие еще есть идеи?

Какие еще есть идеи?



Как это будет работать



CTC Loss
Connectionist Temporal Classification

- Решение seq2seq задач
- Не требуется выравнивание данных (Weakly aligned data)
- Задачи ASR / OCR

https://www.cs.toronto.edu/~graves/icml_2006.pdf

Что мы хотим получить?

$$X \rightarrow Y$$

- Speech recognition

“Hello world”

- Optical character recognition (OCR)

“Hello world”

Компоненты системы с CTC Loss

1. Visual feature extraction (CNN)
2. Sequential modeling based on visual feature sequence (RNN)
3. CTC layer to map input sequence (visual feature sequence) to output sequence (character sequence)

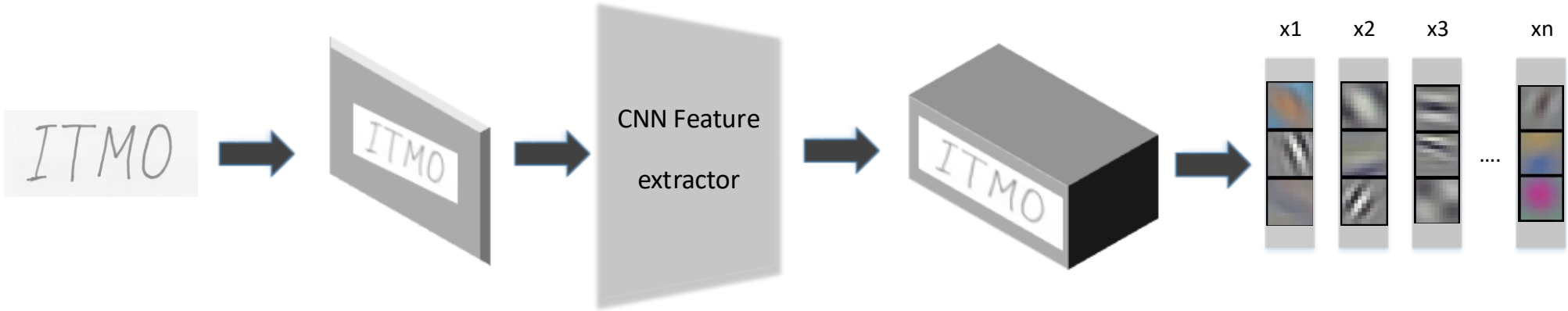
Формальности

- Конечный алфавит $\mathrm{A} = \{a, b, c \ldots z, -\}$
- Целевая последовательновть $Y = (a, b, c)$
- X - изображение
- Выход нейронной сети $f_\theta(X) = M$

$$M = \begin{pmatrix} P(a)_{t1} & \cdots & P(a)_{tn} \\ \vdots & \ddots & \vdots \\ P(-)_{t1} & \cdots & P(-)_{tn} \end{pmatrix}$$

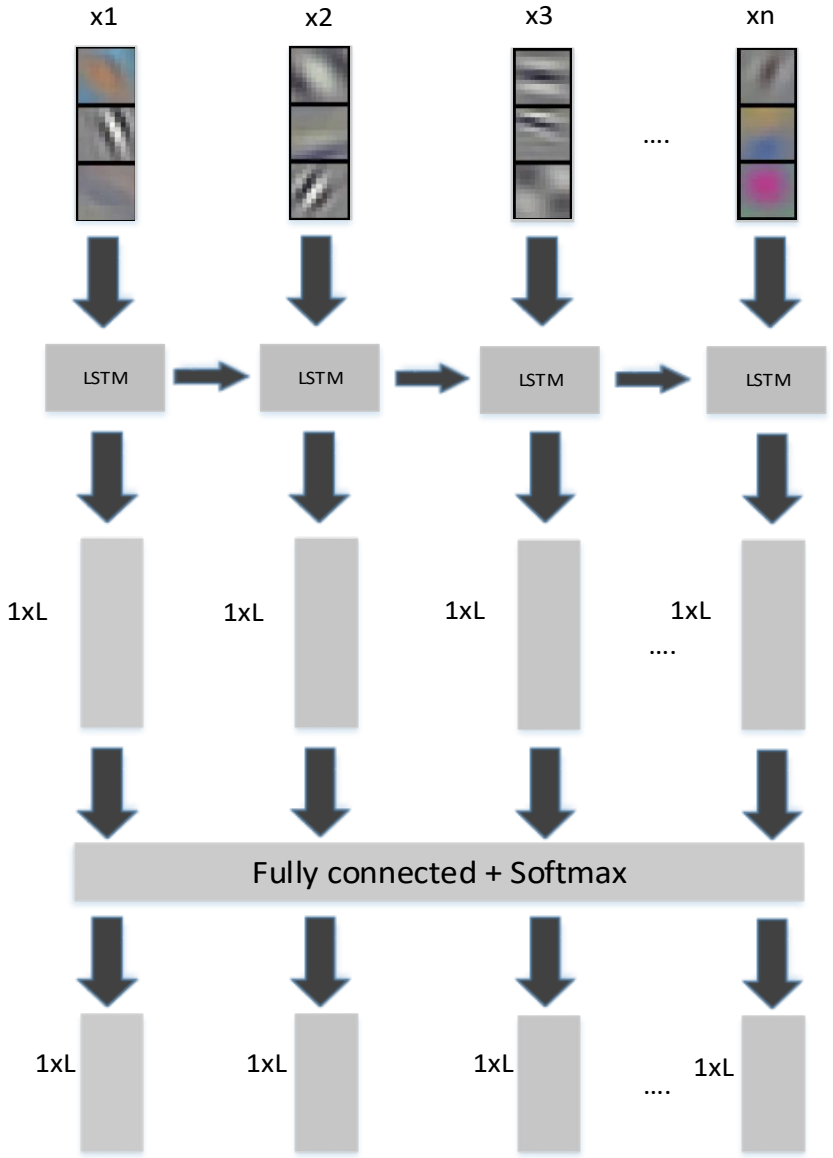
Типовая архитектура

Visual feature extraction



Типовая архитектура

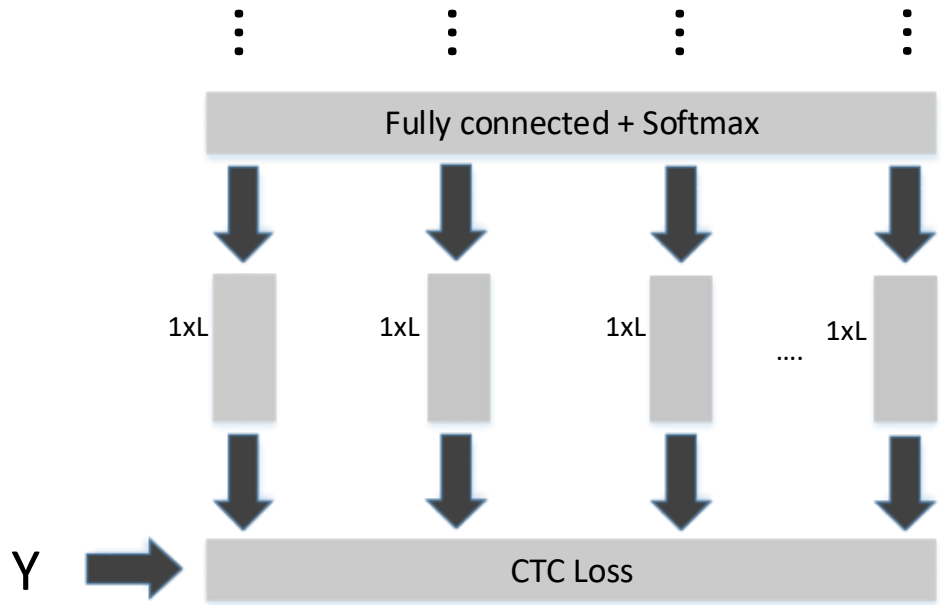
Sequential modeling based on visual feature sequence (RNN)



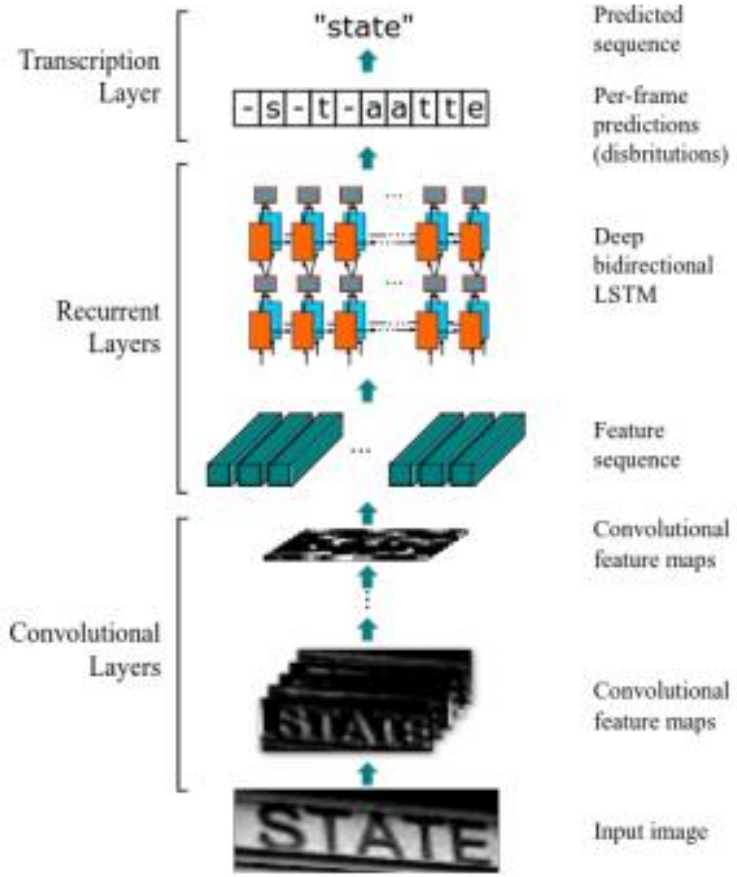
	x1	x2	x3	…	xn
P(A)	0.01	0.1	0.05	0.1	0.009
P(B)	0.2	0.05	0.07	0.005	0.01
…	0.8	0.006	0.005	0.001	0.3
P(-)	0.03	0.006	0.22	0.2	0.004

Типовая архитектура

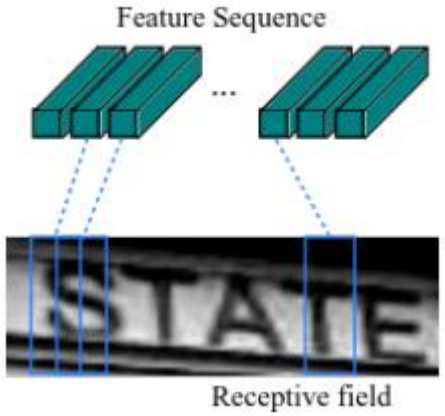
CTC layer to map input to output sequence



An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition



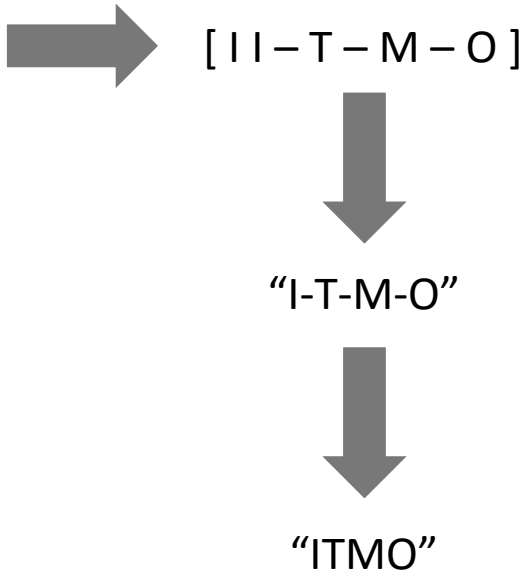
Архитектура



Признаки строго соответствуют участку изображения

Пример

	x1	x2	x3	x4	x5	x6	x7	x8
P(I)	0.8	0.7	0.1	0.2	0.0258	0.175	0.1	0.075
P(M)	0.05	0.075	0.1	0.13	0.0258	0.3	0.1	0.05
P(O)	0.05	0.075	0.1	0.13	0.0258	0.175	0.1	0.0416
P(T)	0.05	0.075	0.1	0.4	0.0258	0.175	0.1	0.0416
P(-)	0.05	0.075	0.6	0.13	0.5	0.175	0.6	0.0416



Blank

- Blank – спецсимвол, для разделения букв (не пробел!)

Intellectual Information Technologies → IIT → [I I T] → “IT”

CTC Loss
Собираем ответ

	x1	x2	x3	x4	x5	x6	x7	x8
P(I)	0.8	0.7	0.1	0.2	0.0258	0.175	0.1	0.13
P(M)	0.05	0.075	0.1	0.13	0.0258	0.3	0.1	0.1
P(O)	0.05	0.075	0.1	0.13	0.0258	0.175	0.1	0.5
P(T)	0.05	0.075	0.1	0.4	0.0258	0.175	0.1	0.13
P(-)	0.05	0.075	0.6	0.13	0.5	0.175	0.6	0.13

$$P(\text{"II}-T-M-O\text{"}) = p_I^1 p_I^2 p_-^3 p_T^4 p_-^5 p_M^6 p_-^7 p_O^8$$

"II–T–M–O" – *путь*

CTC Loss
Примеры путей

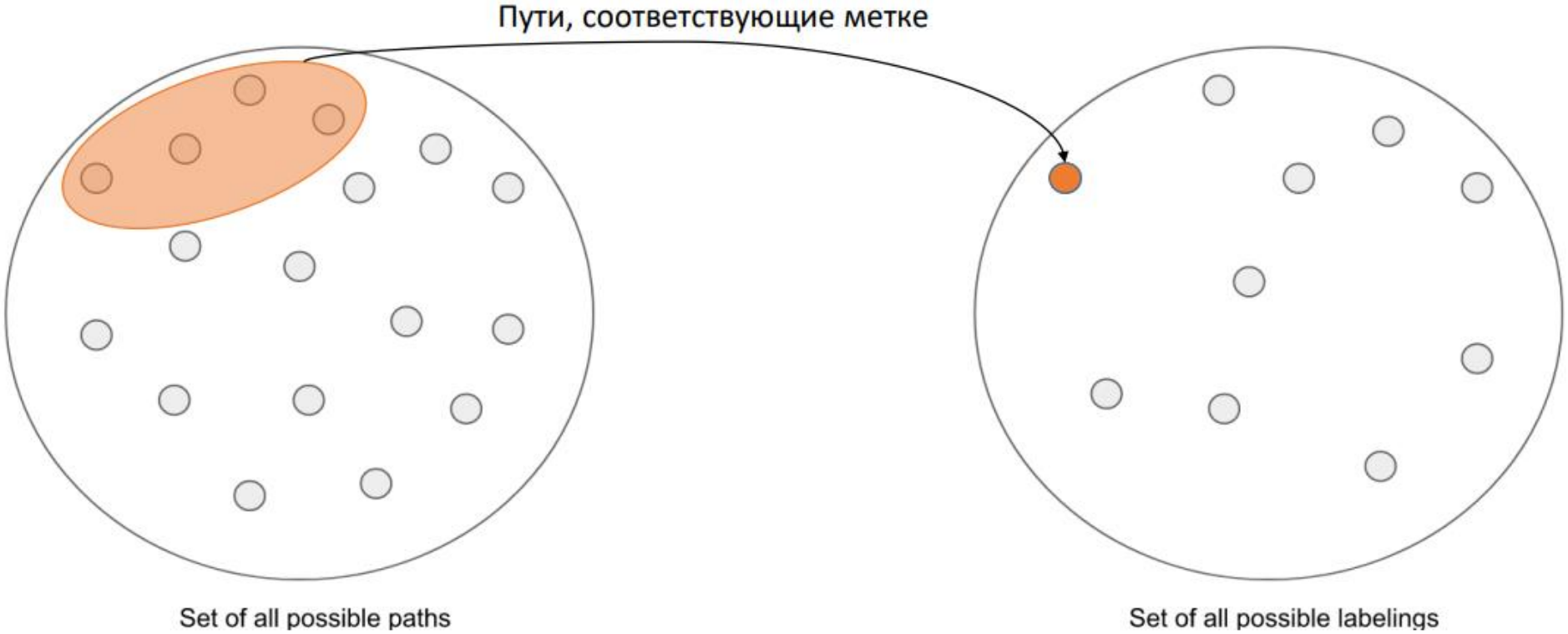
$$P(\text{"}II{-}T{-}M{-}O\text{"}) = p_I^1 p_I^2 p_-^3 p_T^4 p_-^5 p_M^6 p_-^7 p_O^8 \rightarrow \text{"ITMO"}$$

$$P(\text{"}IT{-}MM{-}{-}O\text{"}) = p_I^1 p_T^2 p_-^3 p_M^4 p_M^5 p_-^6 p_-^7 p_O^8 \rightarrow \text{"ITMO"}$$

CTC Loss
Примеры путей

$P(\text{"}II{-}T{-}M{-}O\text{"}) = p_I^1 p_I^2 p_-^3 p_T^4 p_-^5 p_M^6 p_-^7 p_O^8 \rightarrow$ "ITMO"

$P(\text{"}IT{-}MM{-}{-}O\text{"}) = p_I^1 p_T^2 p_-^3 p_M^4 p_M^5 p_-^6 p_-^7 p_O^8 \rightarrow$ "ITMO"



CTC Loss
Вероятность лейблига Y

$$p(y|x) = \sum_{x_\pi \in B^{-1}(y)} p(\pi|x)$$

$\pi -$ путь

$x -$ изображение

$y -$ лейбел (последовательность)

$B -$ функция сжатия

Вероятность лейблинга – сумма вероятностей всех путей, которые ведут в данный лейбел

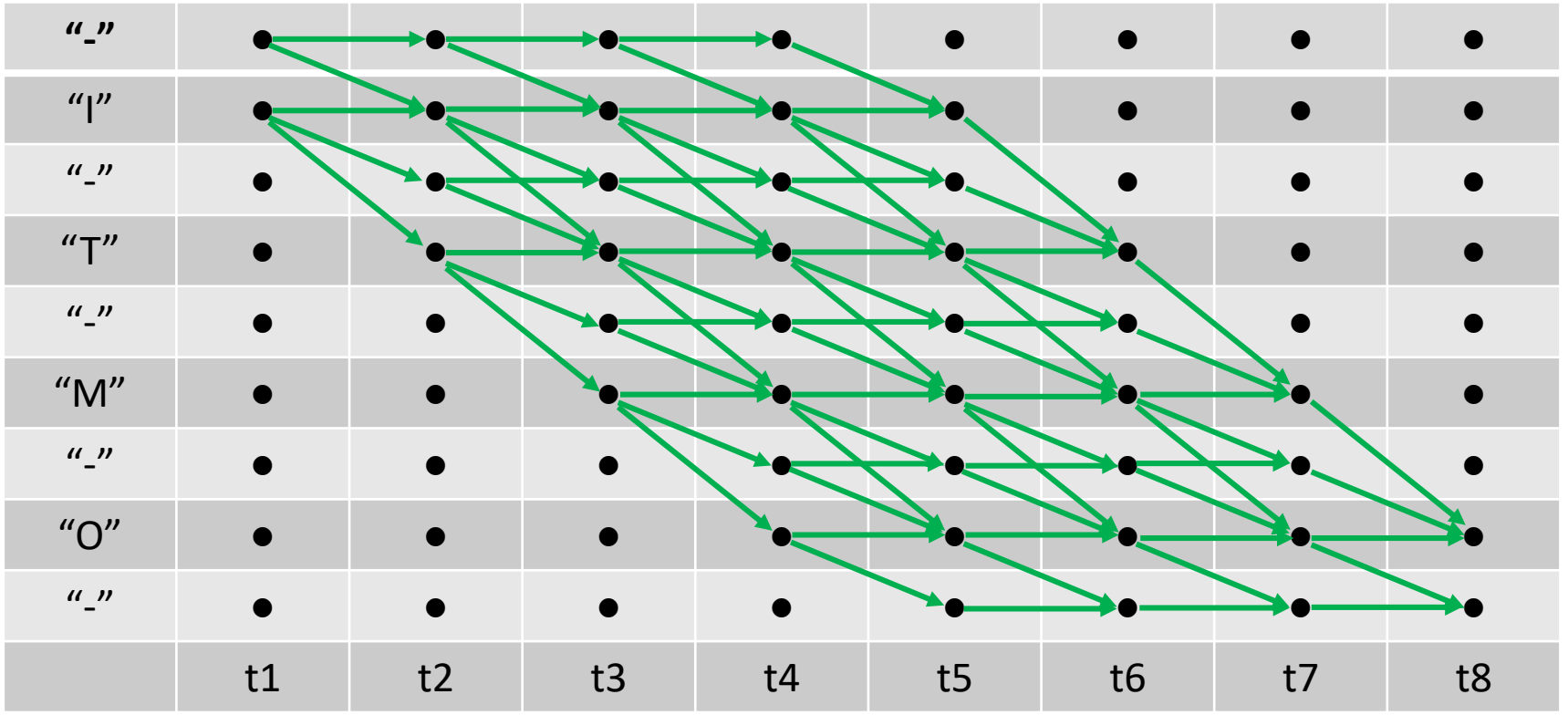
CTC Loss

$$CTC\ Loss = -\ln P(\text{"}ITMO\text{"})$$

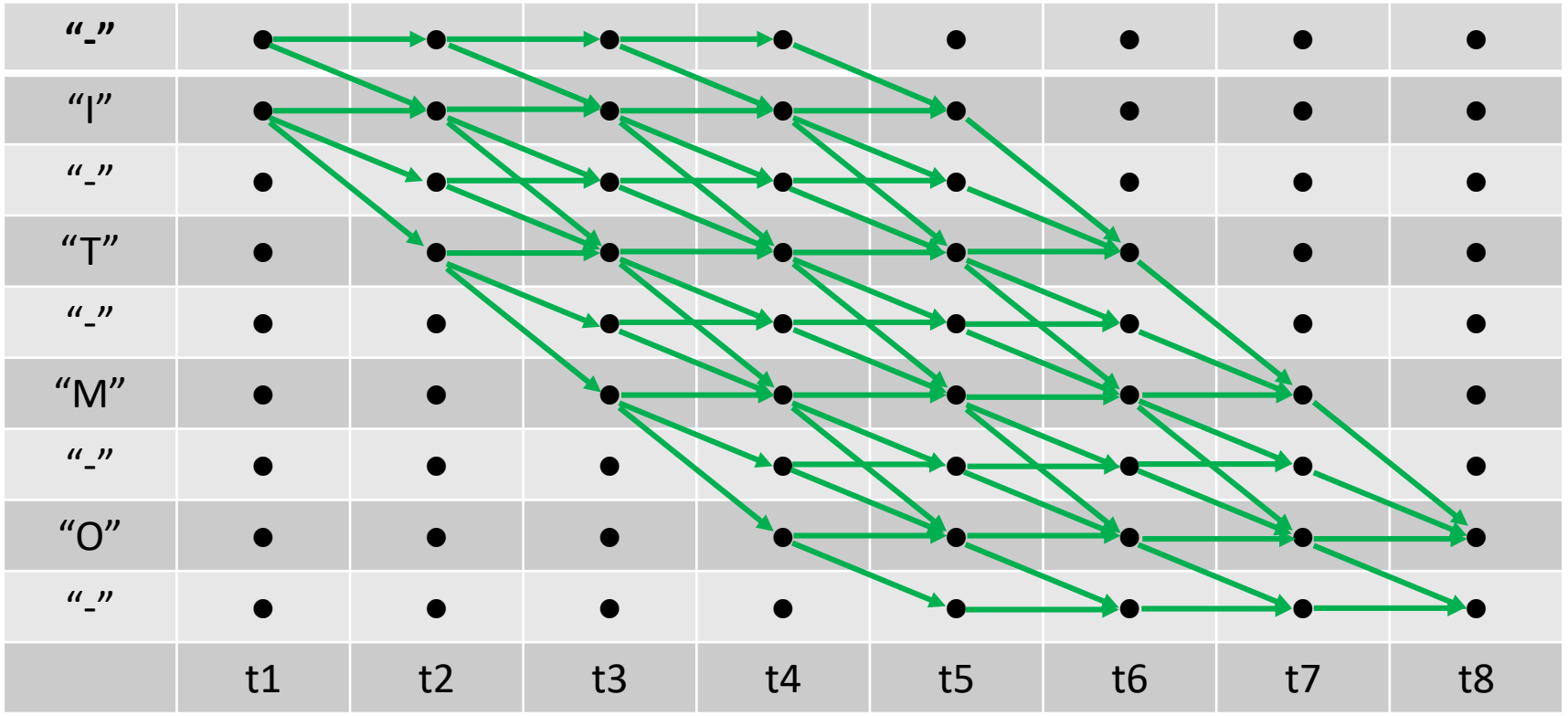
$$N_{paths} = |A|^k$$

$k -$ количество шагов RNN

CTC Loss
Как считать

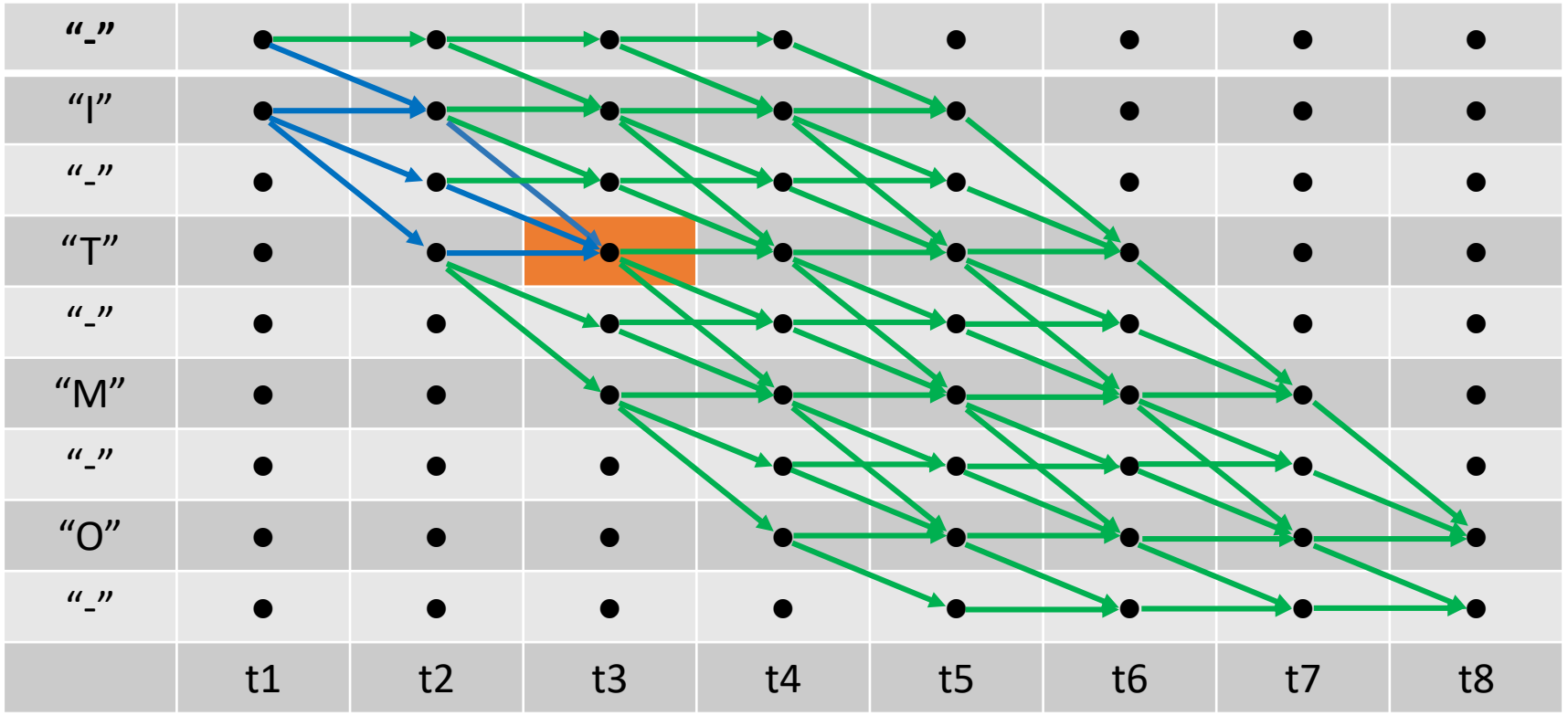


CTC Loss
Как считать



$\alpha_t(s)$ – суммарная вероятность всех подпутей для символа с индексом s в момент времени t

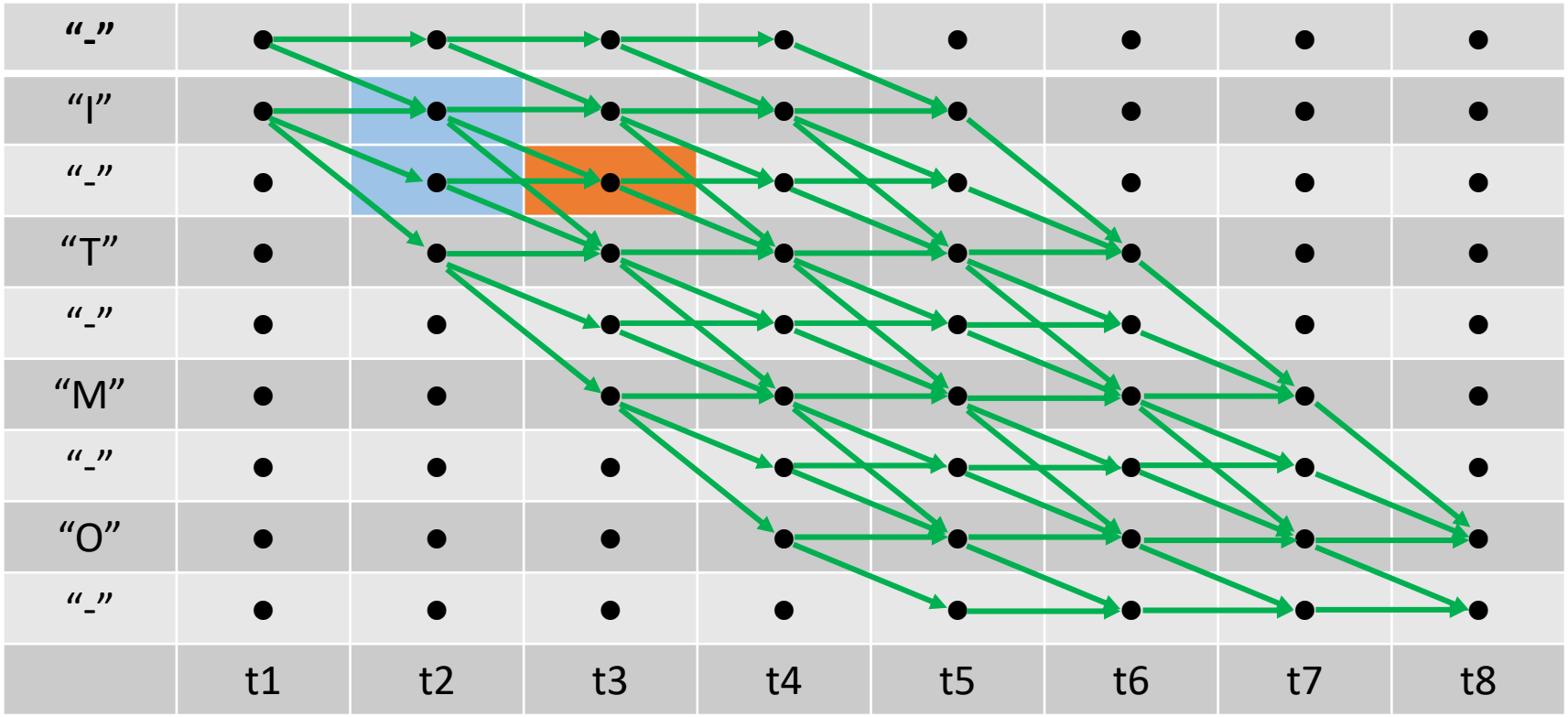
CTC Loss
Как считать



$$\alpha_3(4) = p(" - IT") + p("IIT") + p("I - T") + p("ITT")$$

Подсчет вероятностей для произвольной ячейки

1. В текущий момент времени прогнозируем blank

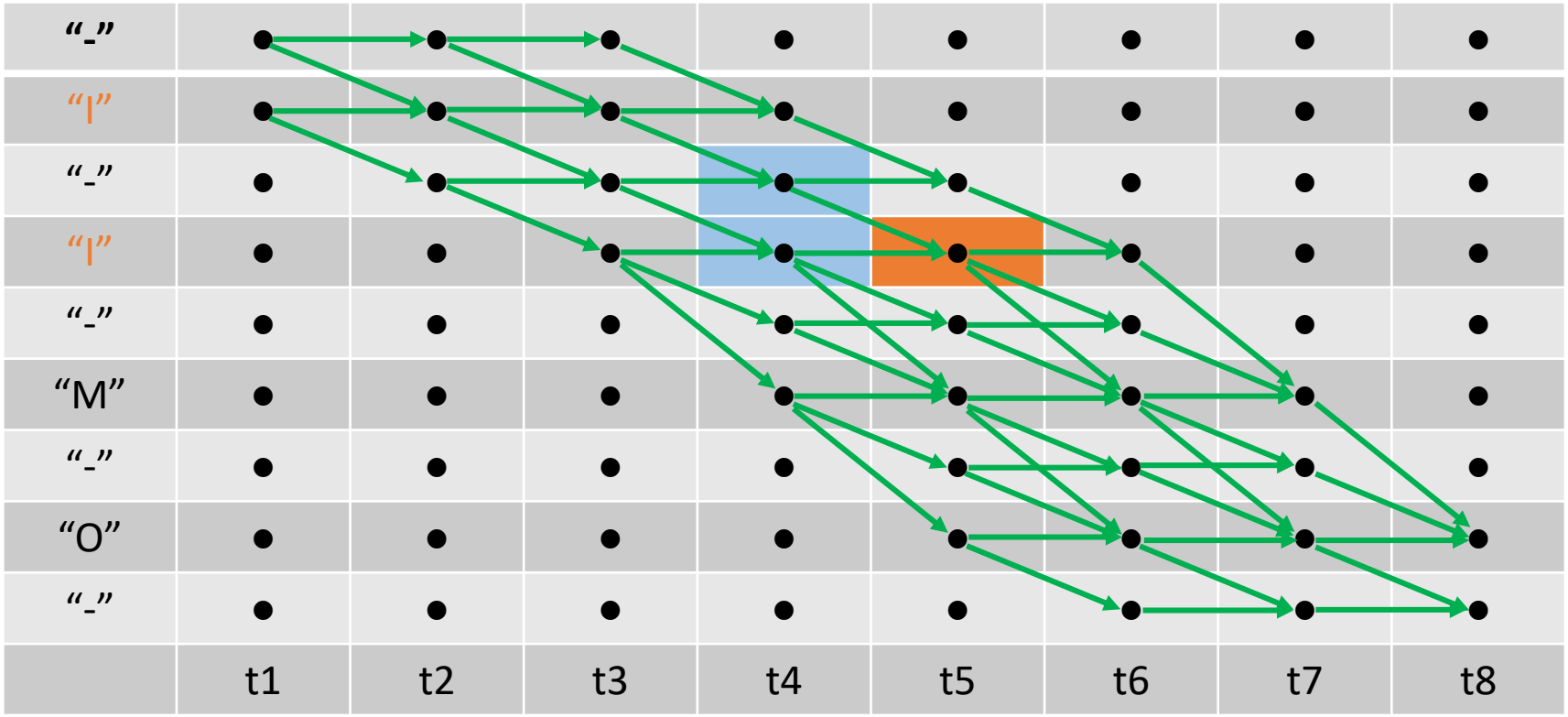


$$\alpha_t(s) = \left(\alpha_{t-1}(s) + \alpha_{t-1}(s-1)\right) \cdot p^t_{seq(s)}$$

$$\alpha_3(3) = \left(\alpha_2(3) + \alpha_2(2)\right)p^3_{\_}$$

Подсчет вероятностей для произвольной ячейки

2. Символ, аналогичный тому, который был два шага назад

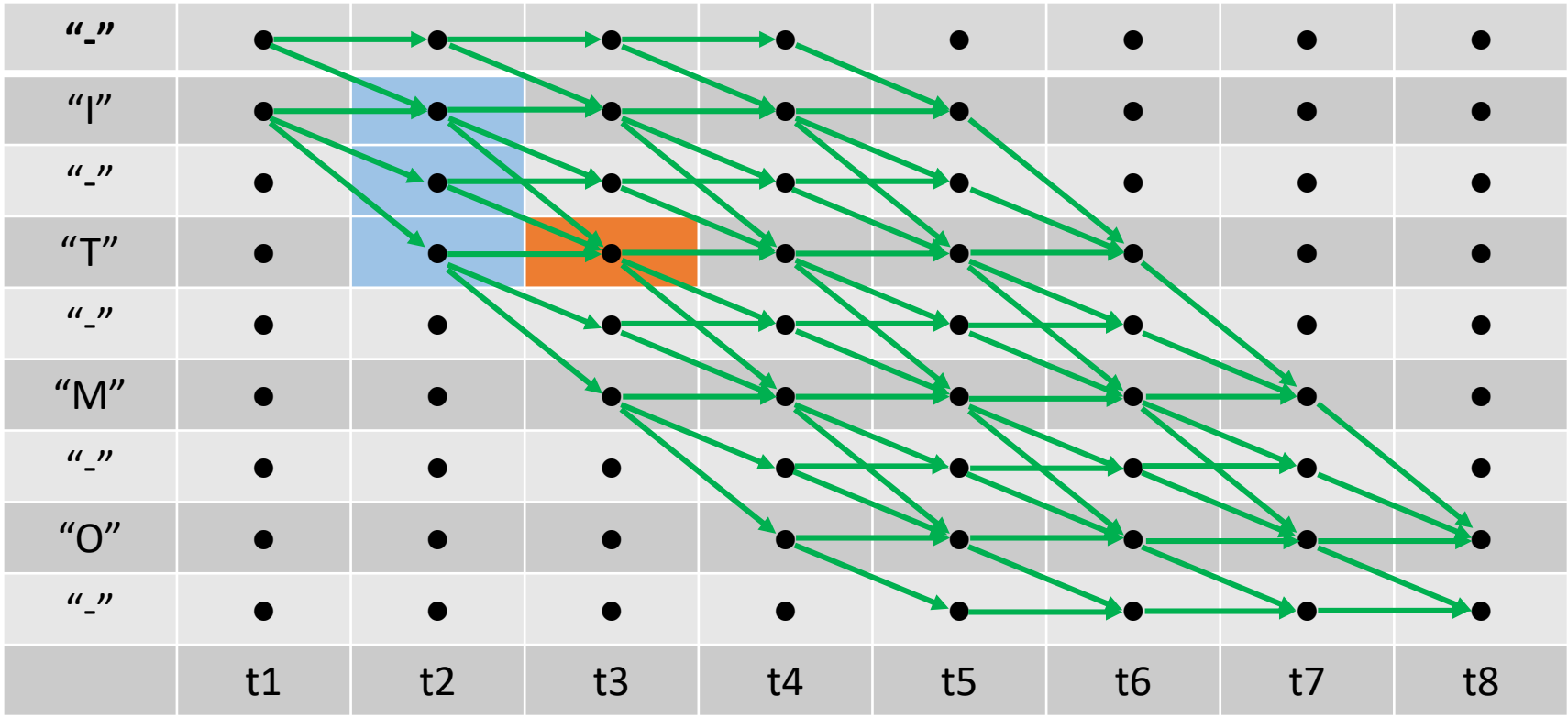


$$\alpha_t(s) = \left(\alpha_{t-1}(s) + \alpha_{t-1}(s-1)\right) \cdot p^t_{seq(s)}$$

$$\alpha_5(4) = \left(\alpha_4(4) + \alpha_4(3)\right) \cdot p^5_I$$

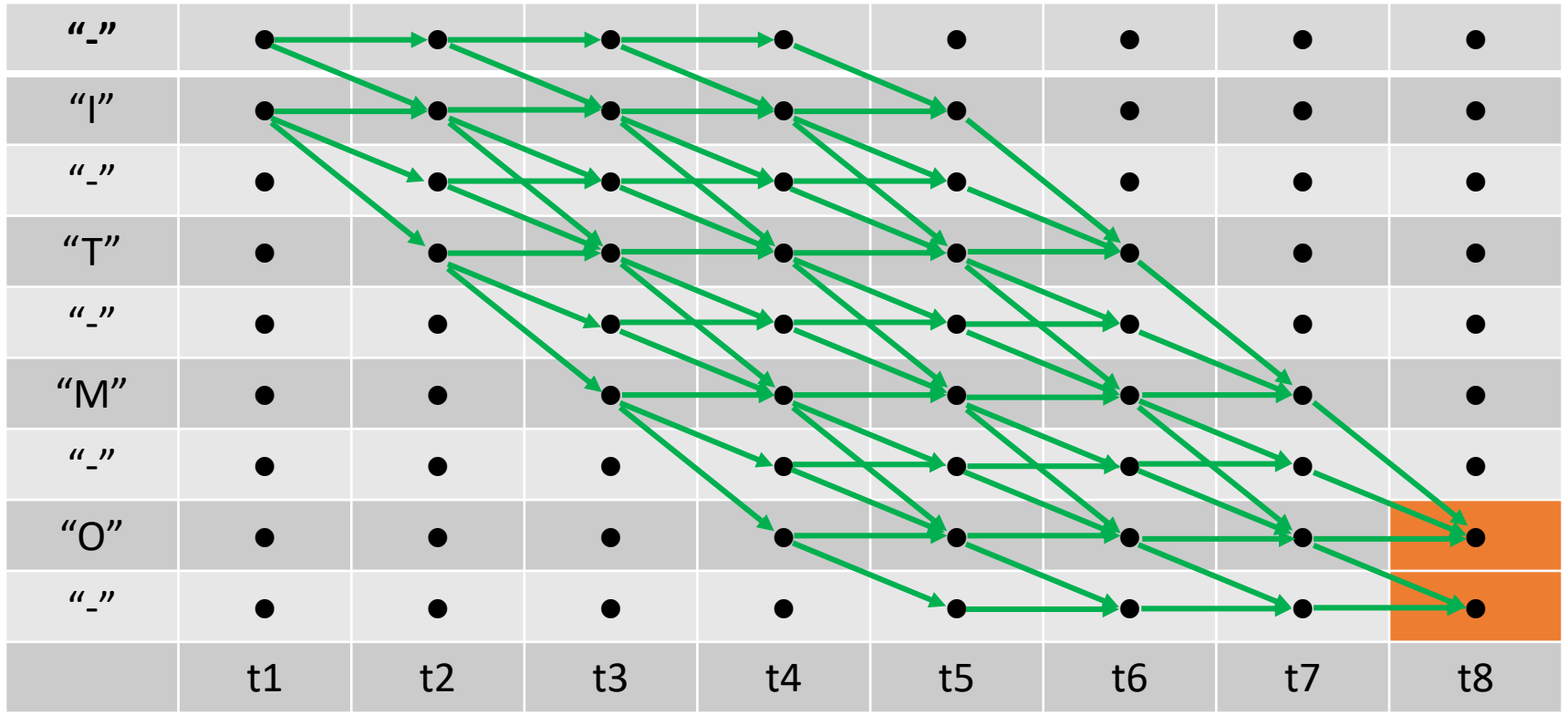
Подсчет вероятностей для произвольной ячейки

3. Все остальные случаи



$$\alpha_t(s) = \left(\alpha_{t-1}(s) + \alpha_{t-1}(s-1) + \alpha_{t-1}(s-2)\right) \cdot p^t_{seq(s)} \qquad \alpha_3(4) = \left(\alpha_2(4) + \alpha_2(3) + \alpha_2(2)\right) \cdot p^3_T$$

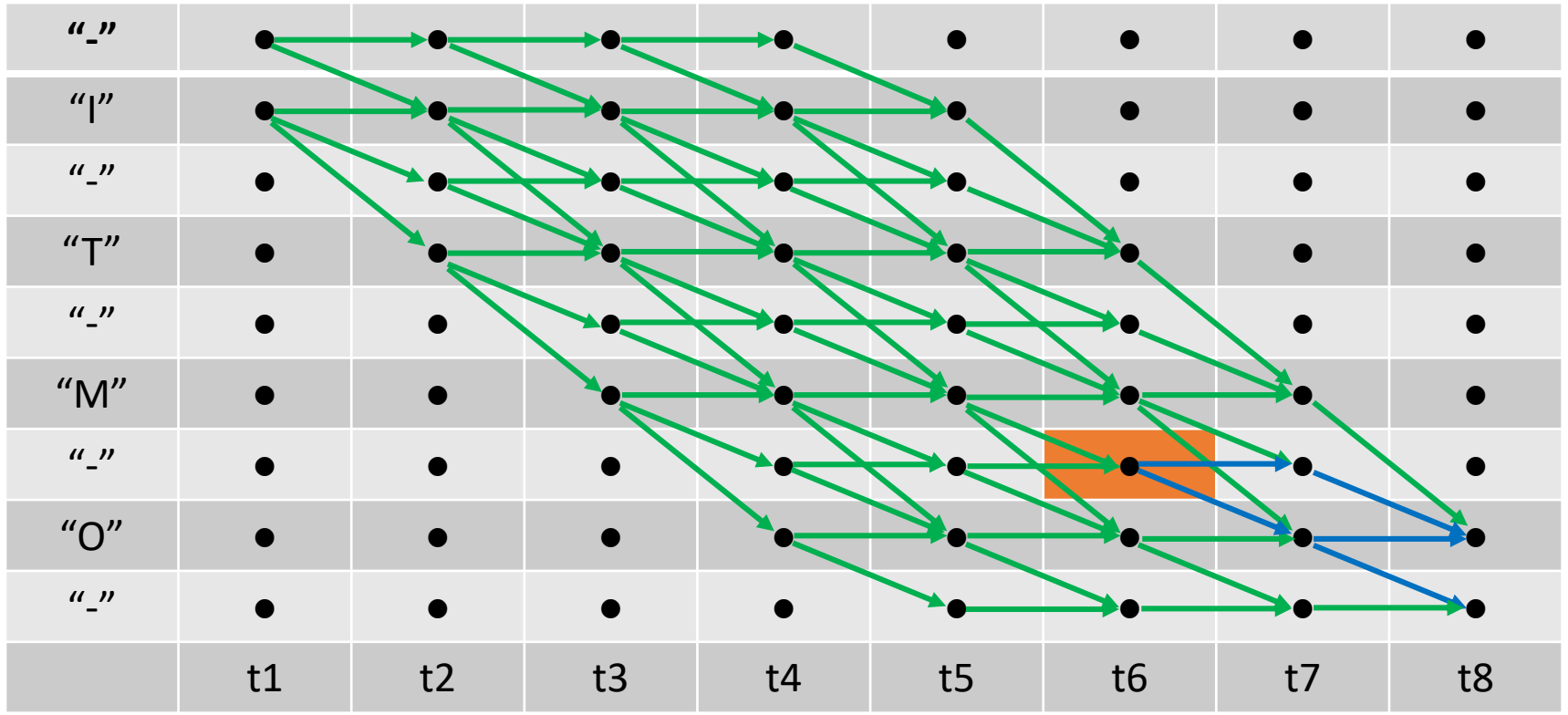
Итоговая вероятность



Forward pass done!

$$p("ITMO") = \alpha_8(8) + \alpha_8(9) \qquad CTC\ loss = -\ln p("ITMO") = -\ln\left(\alpha_8(8) + \alpha_8(9)\right)$$

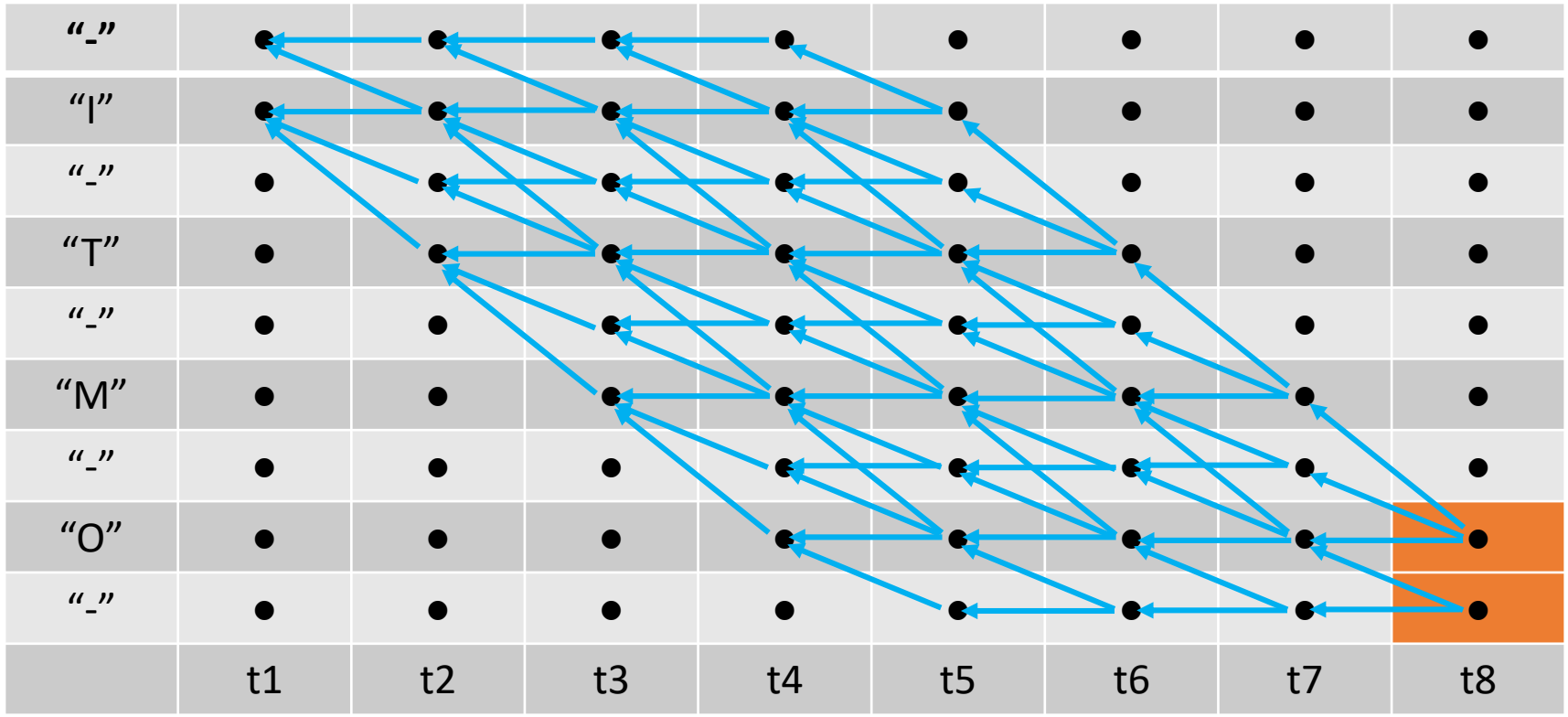
Backward pass



$\beta_t(s)$ – вероятность всех подпутей, которые приводят к $ground\ true$ лейбел

$\beta_6(7) = p("--O") + p("-\ O\ O") + p("-O-")$

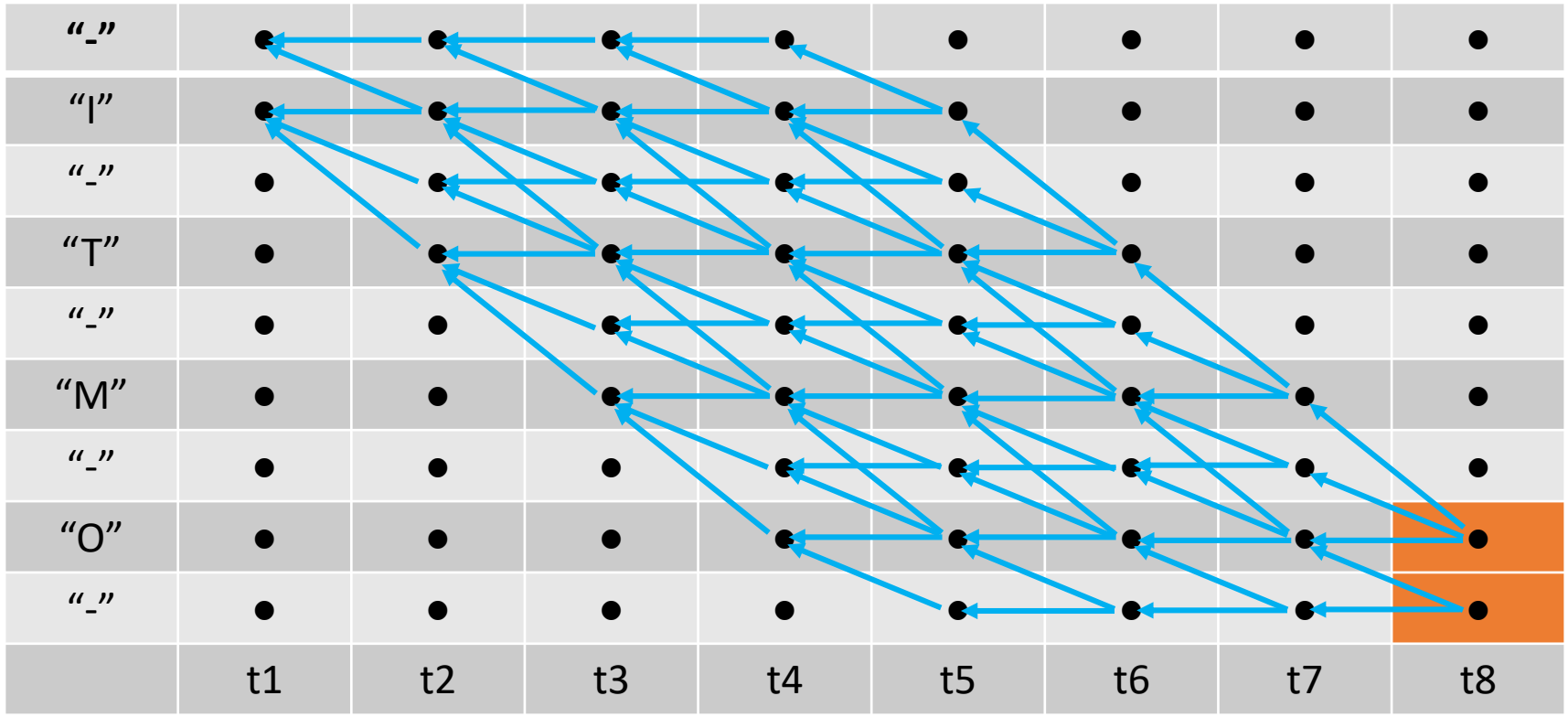
Backward pass



Для вычисления $\beta_t(s)$:

1. Изменяем связи на противоположные
2. Применяем ту же логику, что и для подсчета $\alpha_t(s)$

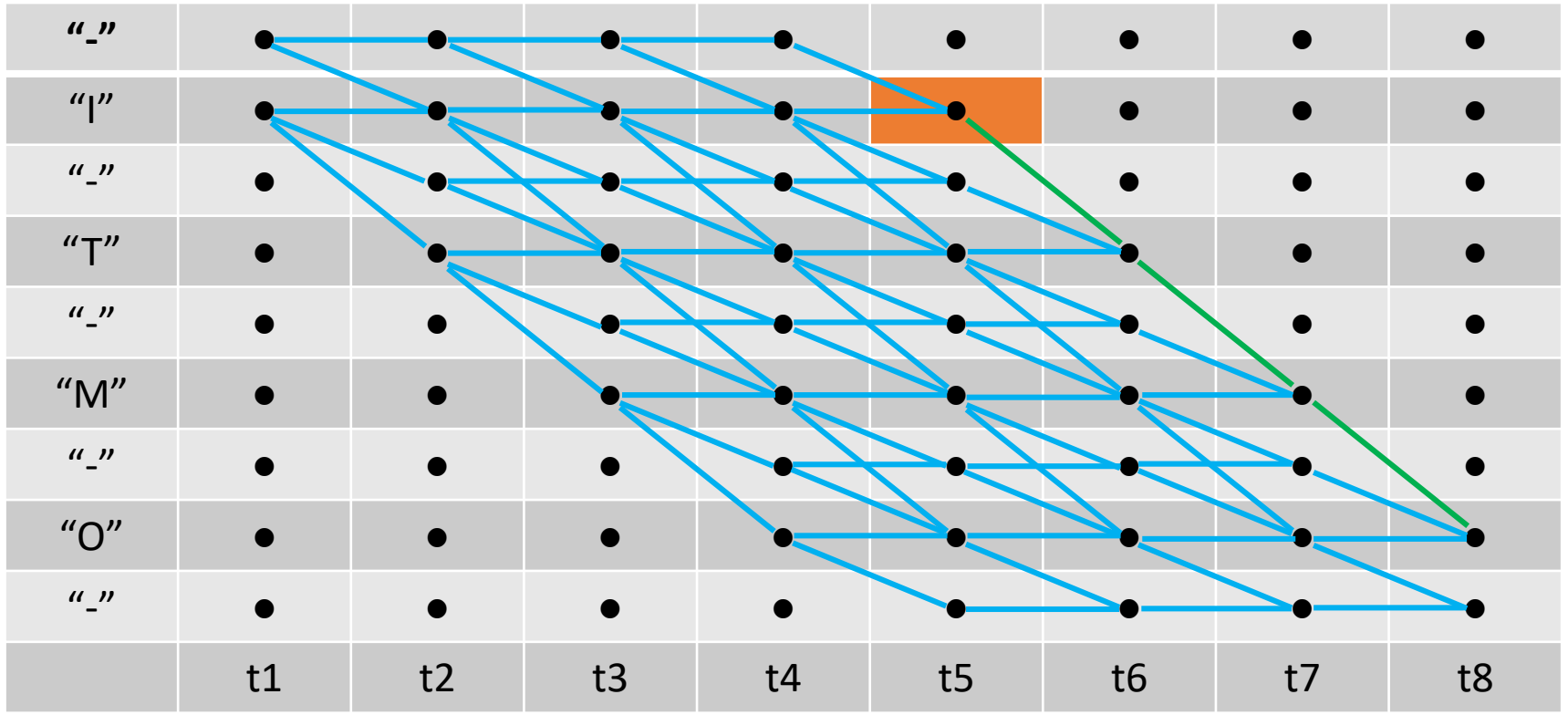
Backward pass



Для вычисления $\beta_t(s)$:

1. Изменяем связи на противоположные
2. Применяем ту же логику, что и для подсчета $\alpha_t(s)$

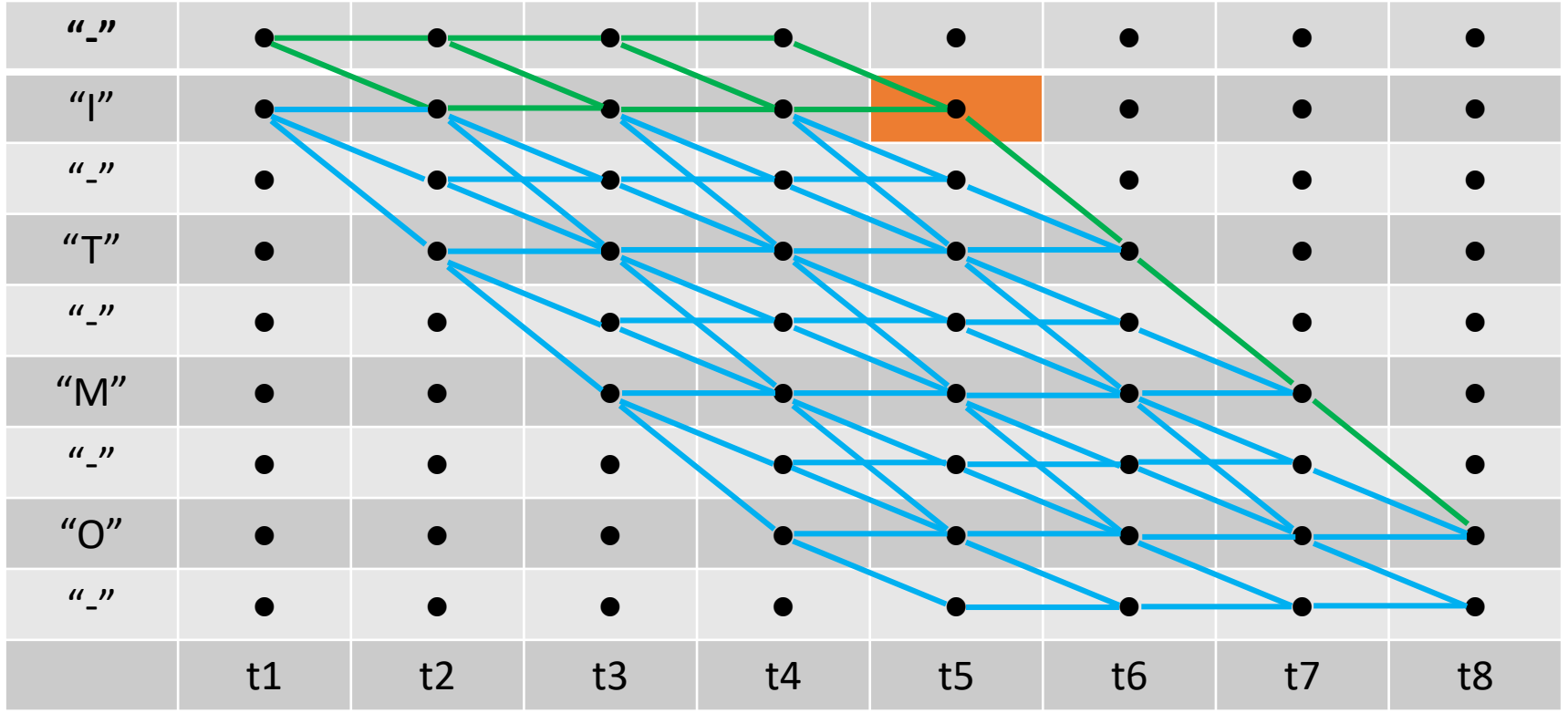
Вероятность пути через ячейку



$$\alpha_5(2) = p("---\ -I") + p("--\ -I\ I") + p("-\ -I\ I\ I") + p("-\ I\ I\ I\ I") + p("I\ I\ I\ I\ I")$$

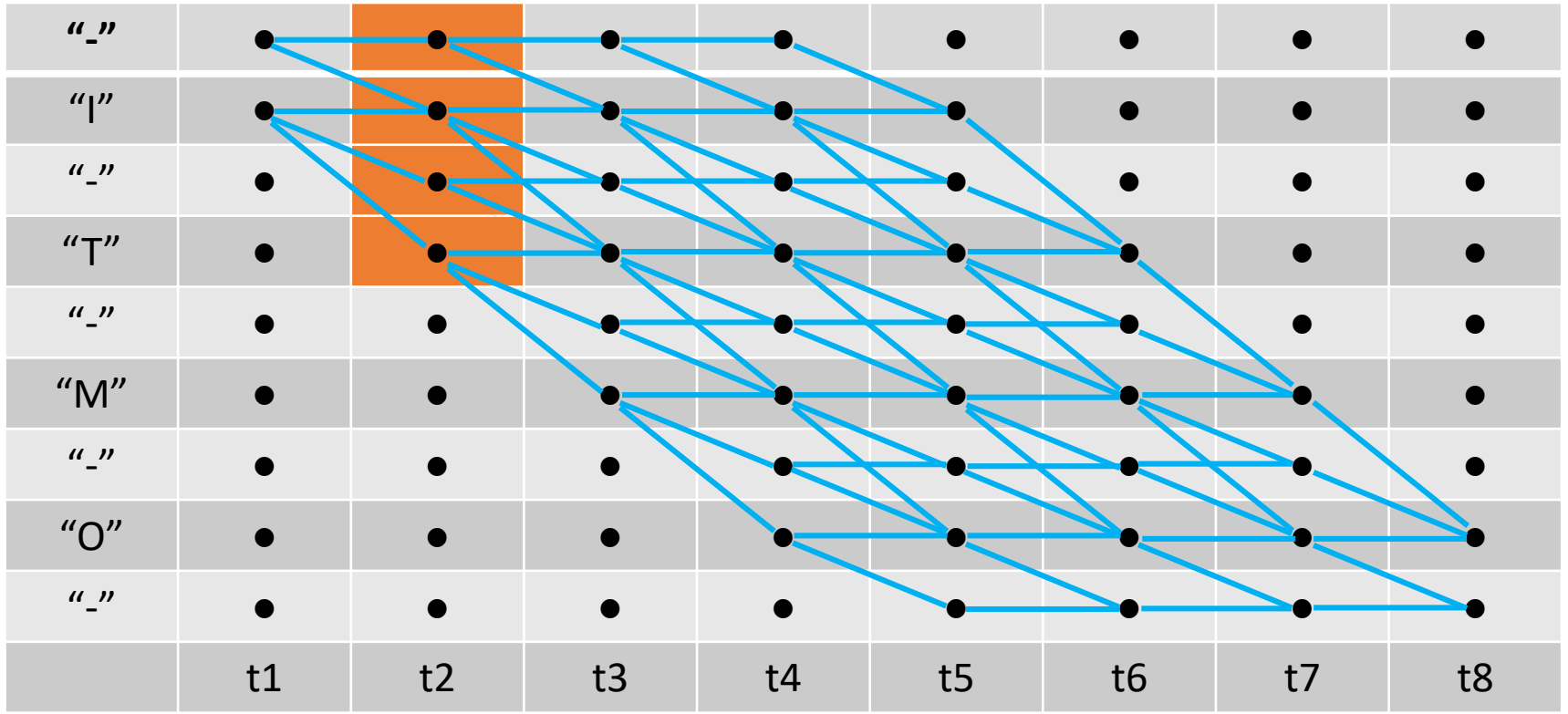
$$\beta_5(2) = p("ITMO") = p_I^5 \cdot p_T^6 \cdot p_M^7 \cdot p_O^8$$

Вероятность пути через ячейку



$\frac{\alpha_5(2) \cdot \beta_5(2)}{p_I^5}$ – сумма верояностей всех путей, прохожящих через ячейку

Вероятность лейблинга в момент времени



$$p("ITMO") = \sum_{s=1}^{4} \frac{\alpha_2(s) \cdot \beta_2(s)}{p^2_{seq(s)}}$$

Backprop

$$\frac{\partial\left(-\ln\left(p("ITMO")\right)\right)}{\partial y_k^t} = -\frac{1}{p("ITMO")} \cdot \frac{\partial p("ITMO")}{\partial y_k^t}$$

$$\frac{\partial\left(-\ln\left(p("ITMO")\right)\right)}{\partial y_k^t} = -\frac{1}{{y_k^t}^2} \cdot \sum_{s:seq(s)=k} \alpha_t(s) \cdot \beta_t(s)$$

Заключение

Вычисляем α для вычисления ctc loss

Вычисляем β – можно посчитать градиенты

Возможные улучшения

- Накрутить поверх распознавания Encoder-Decoder

Хороший пример по данной теме – соревнование 2020 года “**Digital Peter: recognition of Peter the Great's manuscripts**”

Baseline – модель с CTC loss, улучшение – приделать трансформер

Ссылка на гитхаб соревнования

Еще более новые подходы

AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

Encoder – Transformer-based vision model
Decoder – NLP model

Библиотека с реализацией
https://huggingface.co/docs/transformers/model_doc/vision-encoder-decoder

Полезные ссылки

- https://www.youtube.com/watch?v=SAfJ6nP2rrI (**How to build end-to-end recognition system (Part 1): best practices [RU]**)
- **https://www.youtube.com/watch?v=eYIL4TMAeRI** (**How to build end-to-end recognition system (Part 2): CTC Loss [RU]**
- https://www.youtube.com/watch?v=ZPNsYTs2Zx4 (**Как нейронная сеть распознает текст? Лекция 1 по Advanced Computer Vision)**
- **https://www.youtube.com/watch?v=F2ERzhLFeuo (Распознавание текста. Обратное распространение через CTC Loss. Лекция 2 по Advanced Computer Vision)**
- **https://cseweb.ucsd.edu/classes/wi19/cse291-g/student_presentations/CTC_OCR.pdf (**Connectionist Temporal Classification (CTC) with application to Optical Character Recognition)
- **https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/02-teaching/2021s_dla/2021_DLA-09_ZalkowMueller_CTC.pdf (**Connectionist Temporal Classification (CTC) Loss)
- **http://www.machinelearning.ru/wiki/images/c/c6/Digital_Signal_Processing%2C_lecture_7.pdf (**Распознавание речи. Современные подходы)
- https://ogunlao.github.io/blog/2020/07/17/breaking-down-ctc-loss.html (Breaking down the CTC Loss)

Спасибо за внимание!