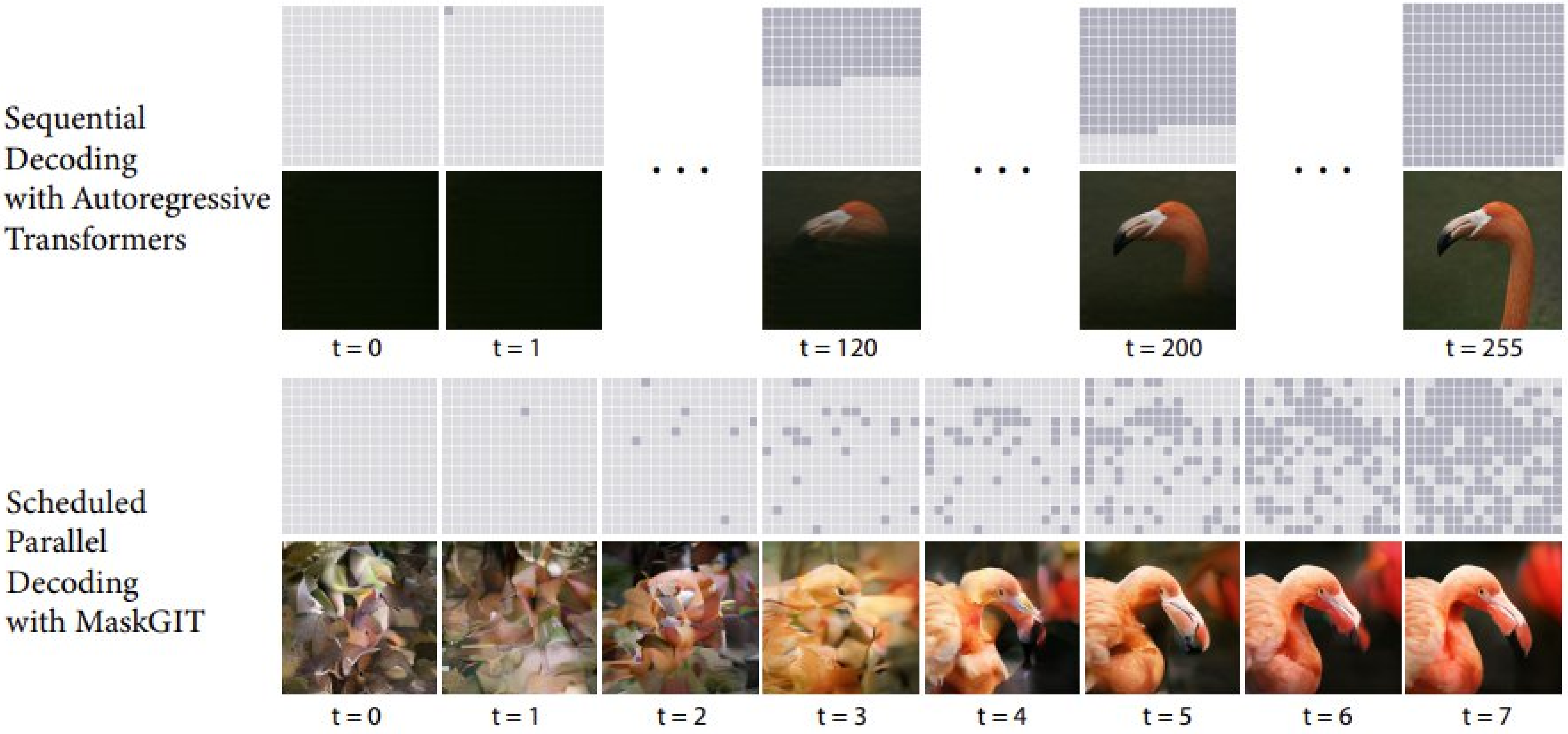
Non-autoregressive approach in different tasks.

Семаков Андрей

MaskGIT: Masked Generative Image Transformer

Задача



MaskGIT: Masked Generative Image Transformer

Идея

Пайплайн

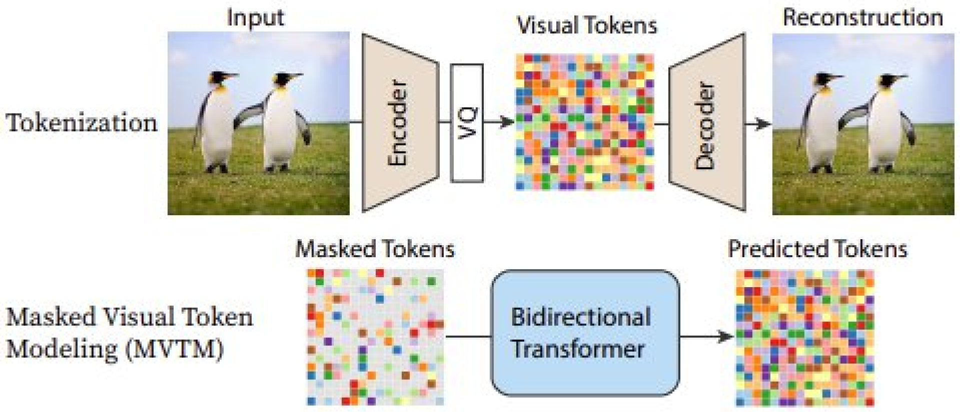


Figure 3. **Pipeline Overview.** MaskGIT follows a two-stage design, with 1) a tokenizer that tokenizes images into visual tokens, and 2) a bidirectional tranformer model that performs MVTM, i.e. learns to predict visual tokens masked at random.

Маскирование

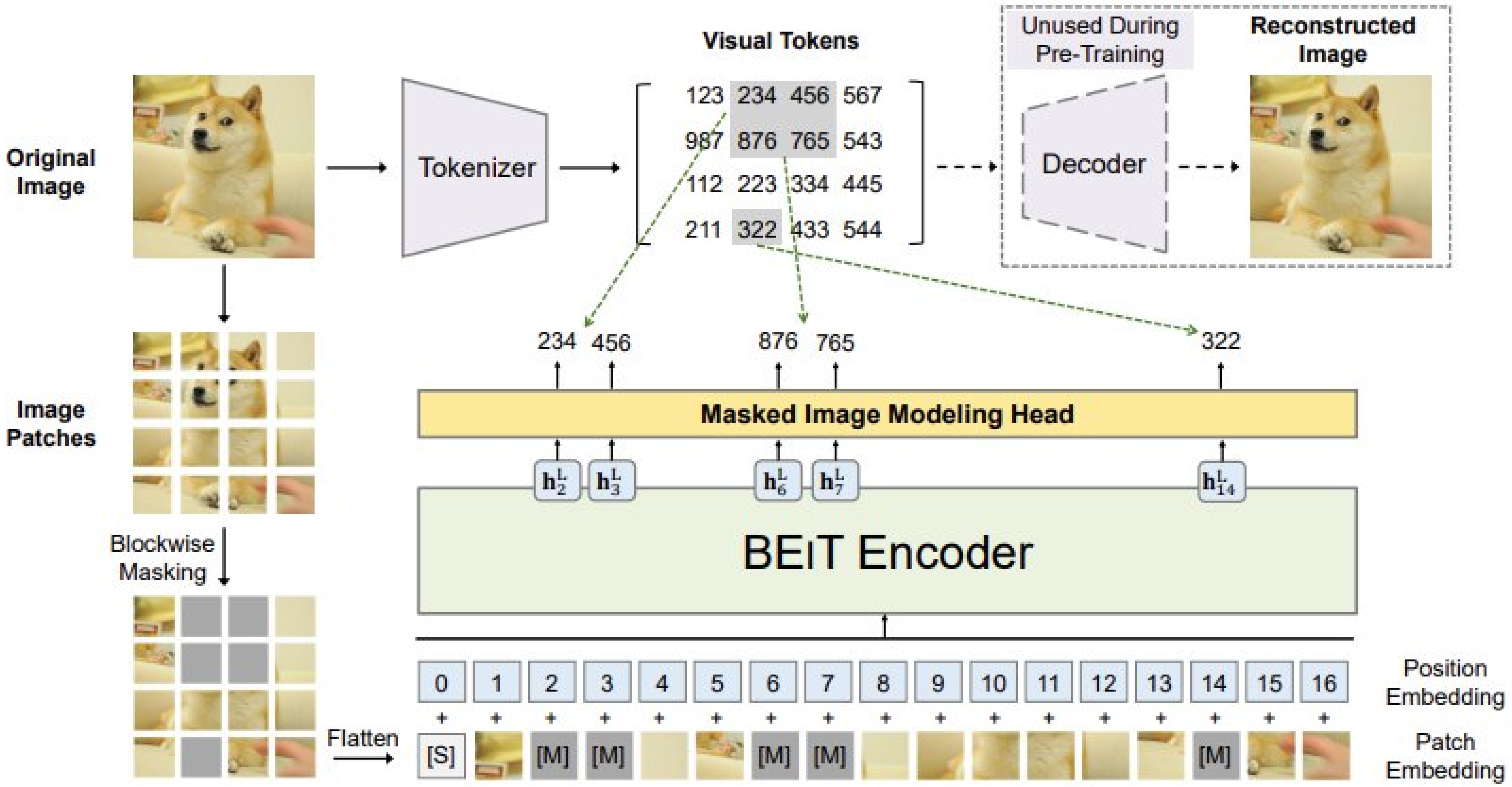
$$n = \left\lceil \gamma\left(\frac{t}{T}\right) N \right\rceil$$

$$\gamma(0) \rightarrow 1 \text{ and } \gamma(1) \rightarrow 0.$$

$$\mathcal{L}_{\text{mask}} = - \mathop{\mathbb{E}}_{\mathbf{Y} \in \mathcal{D}} \left[\sum_{\forall i \in [1,N], m_i = 1} \log p(y_i | Y_{\overline{\mathbf{M}}}) \right]$$

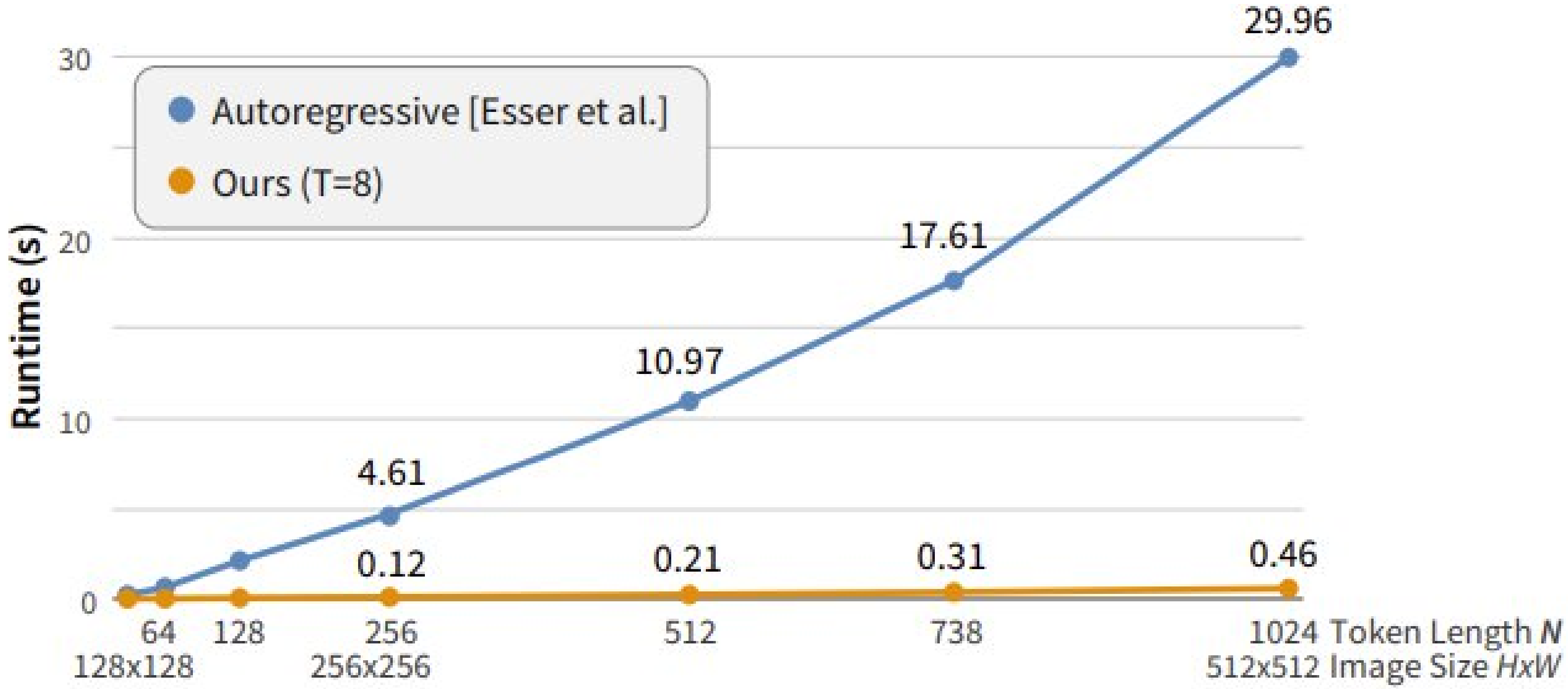
MaskGIT: Masked Generative Image Transformer

Идея



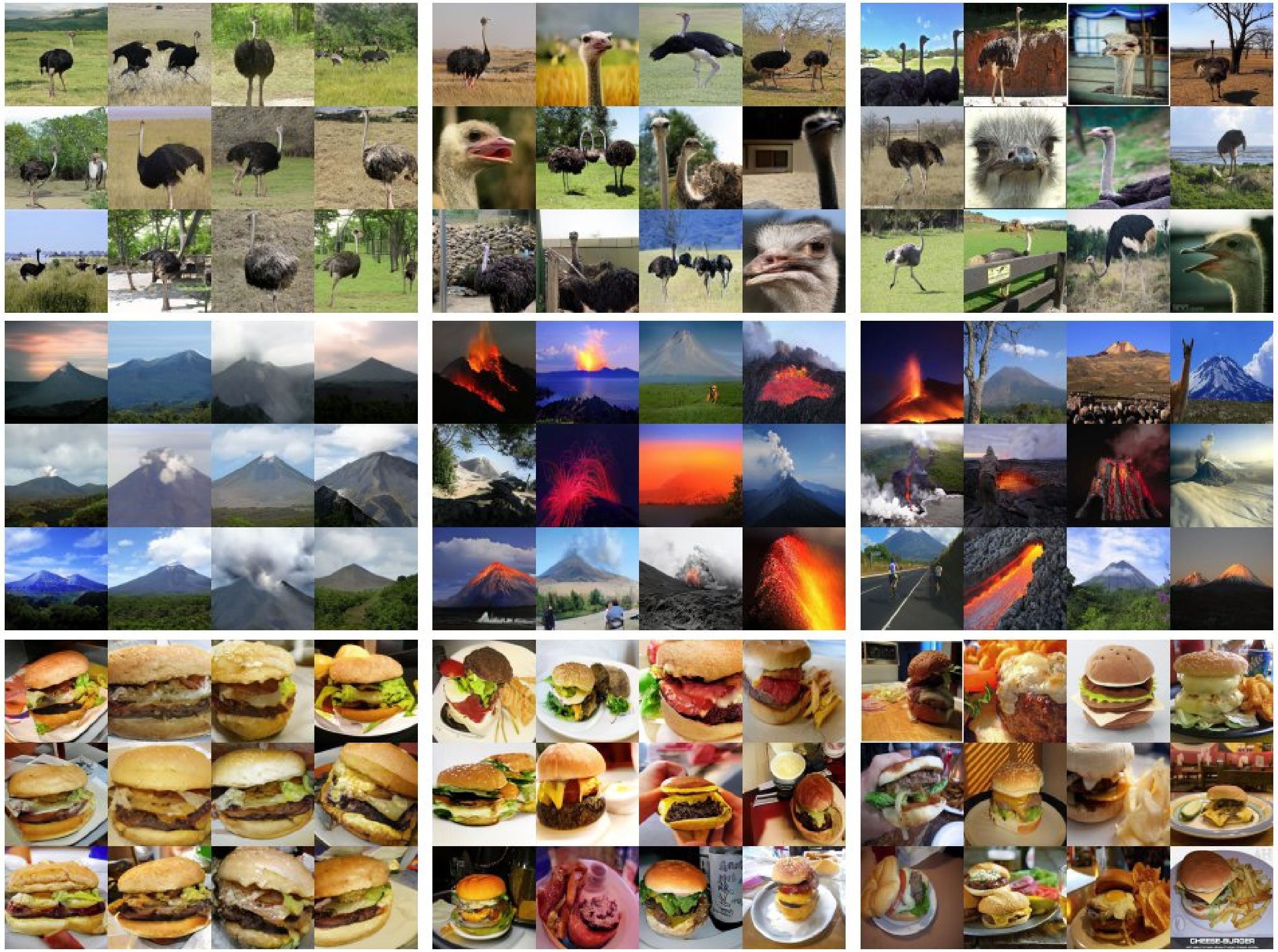
MaskGIT: Masked Generative Image Transformer

Результаты



MaskGIT: Masked Generative Image Transformer

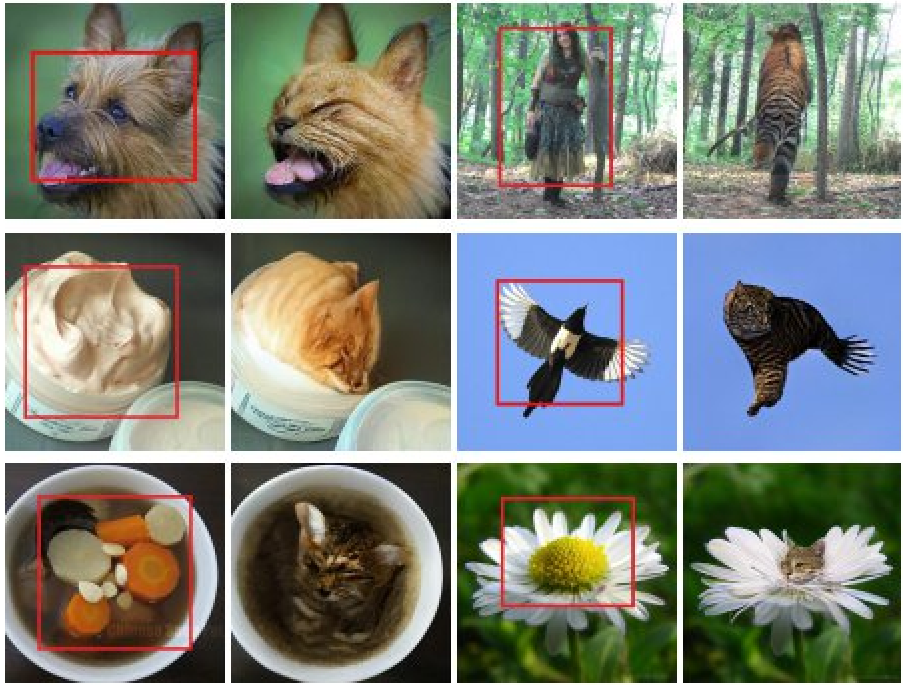
Результаты

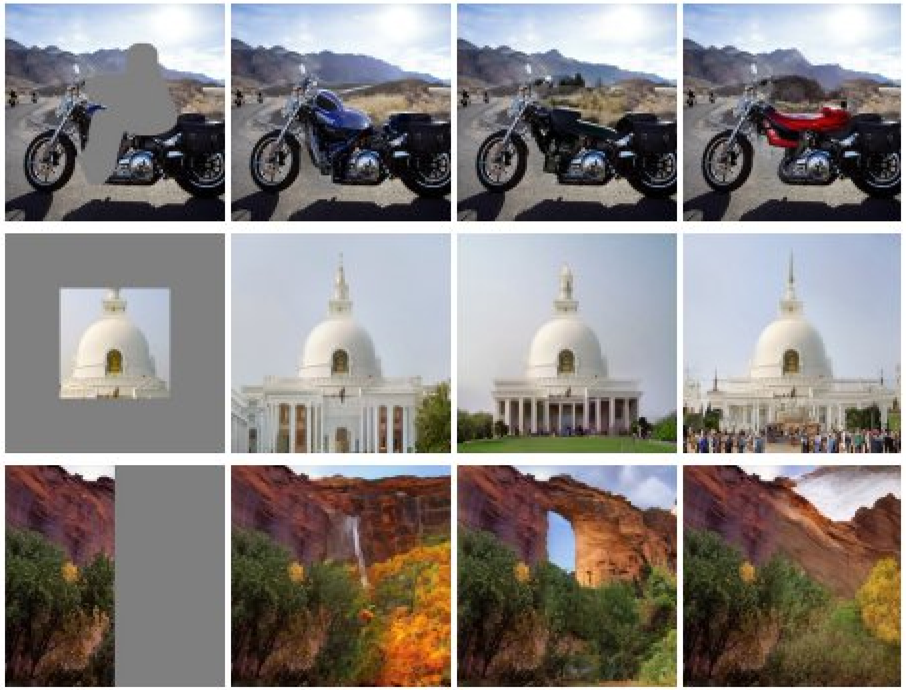


BigGAN-deep (FID=6.95) MaskGIT (FID=6.18) Training Set

MaskGIT: Masked Generative Image Transformer

Результаты





Input —— MaskGIT (Our Samples) ——

Non-Autoregressive Semantic Parsing for Compositional Task-Oriented Dialog

Задача

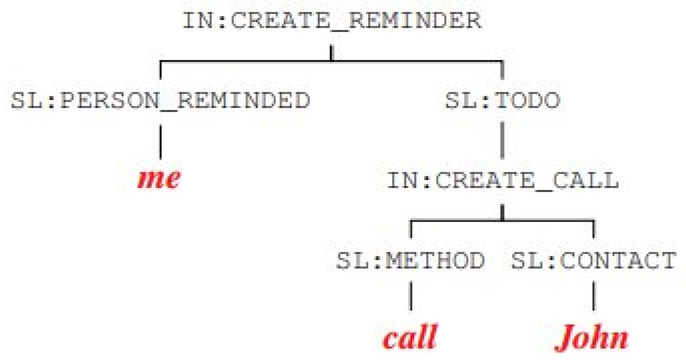
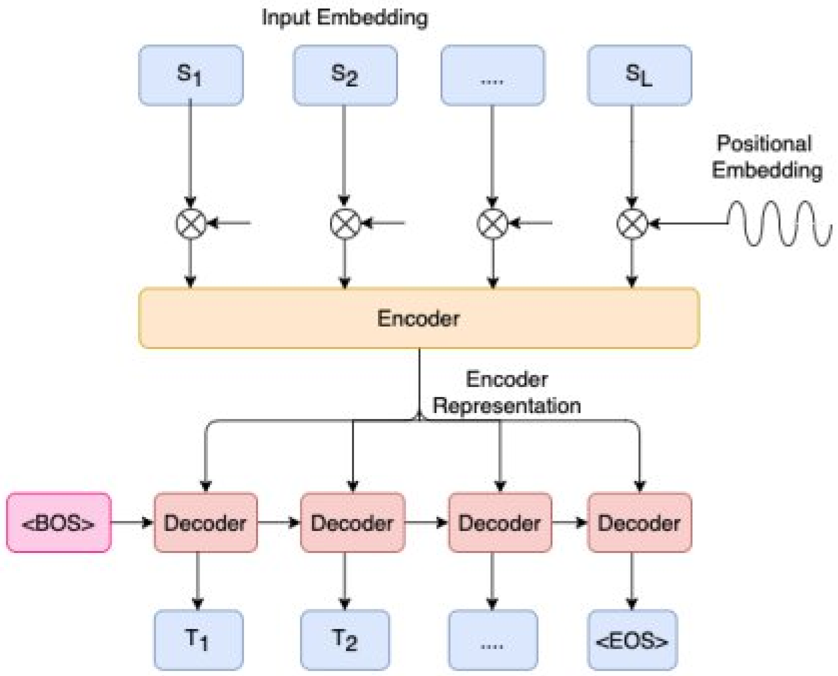


Figure 1: Decoupled semantic representation for the single utterance *"Please remind me to call John"*.

Non-Autoregressive Semantic Parsing for Compositional Task-Oriented Dialog

Идея

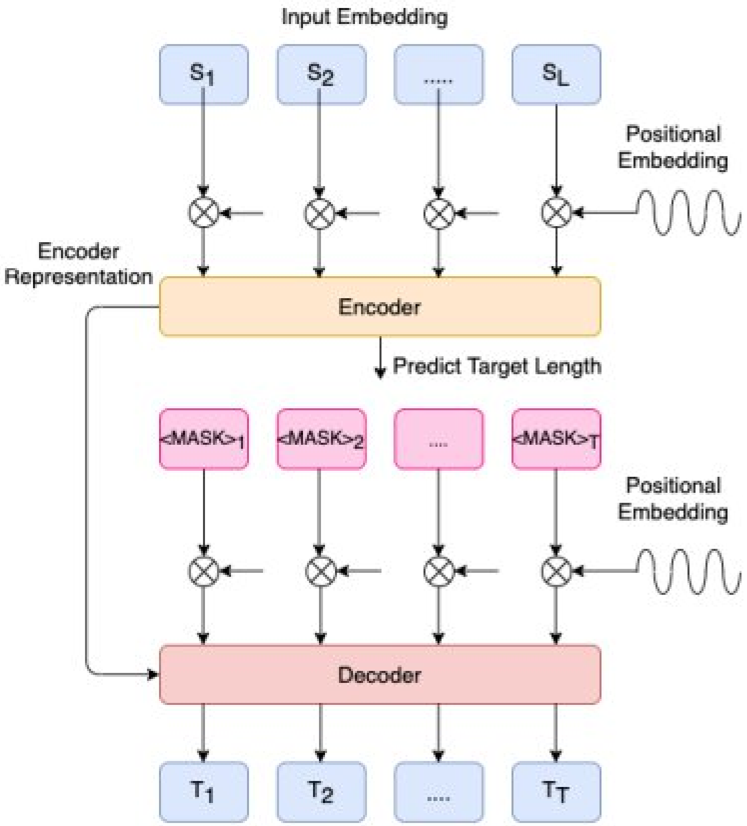
Традиционный подход

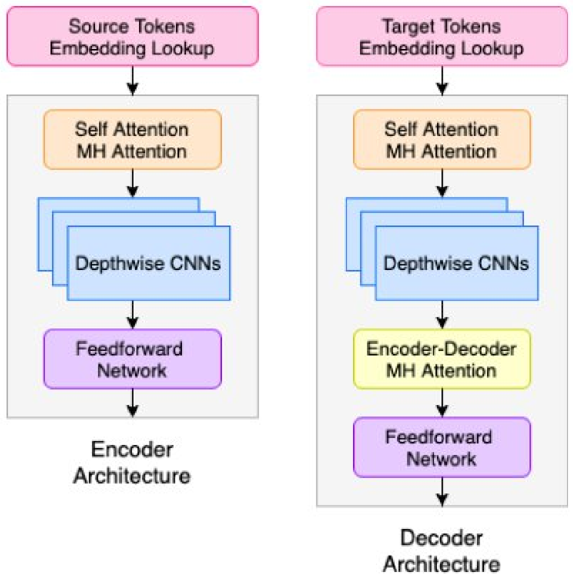


Non-Autoregressive Semantic Parsing for Compositional Task-Oriented Dialog

Идея

Предложение авторов





$$e_1, \dots, e_L = \text{Encoder}(s_1, \dots, s_L)$$

$$T = \text{PredictLength}(e_1, \dots, e_L)$$

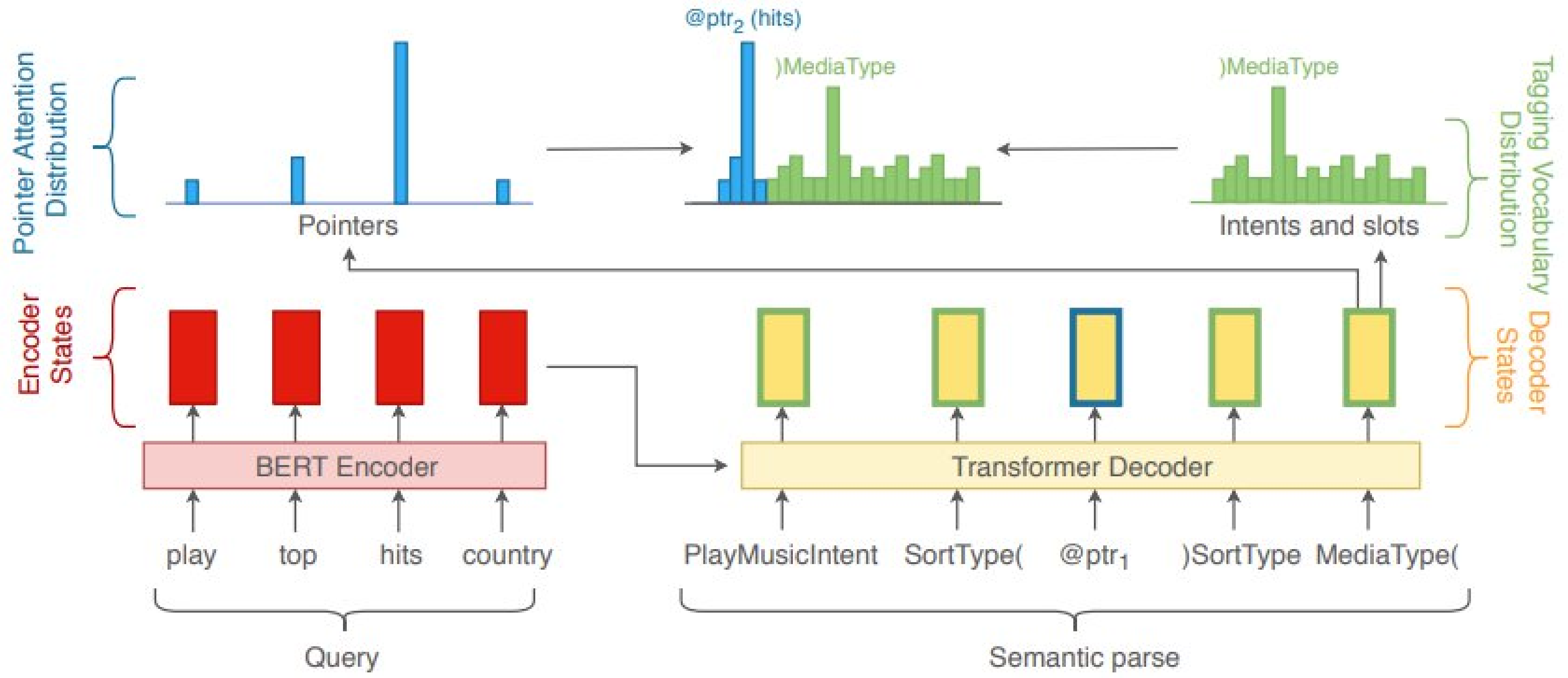
$$x_1, ..., x_T = \text{Dec}(\text{MASK}_1, \dots, \text{MASK}_T; e_1, ..., e_L)$$

$$y_1, ..., y_T = \text{PtrProj}(x_1, .., x_T; e_1, .., e_L)$$

```
Source: How far is the coffee shop
Target: [IN:GET_DISTANCE @ptr0 @ptr1 @ptr2 [SL:DESTINATION
    [IN:GET_RESTAURANT_LOCATION @ptr3 [SL:TYPE_FOOD
    @ptr4 SL:TYPE_FOOD] @ptr5 IN:GET_RESTAURANT_LOCATION]
    SL:DESTINATION] IN:GET_DISTANCE]
```

Non-Autoregressive Semantic Parsing for Compositional Task-Oriented Dialog

Идея

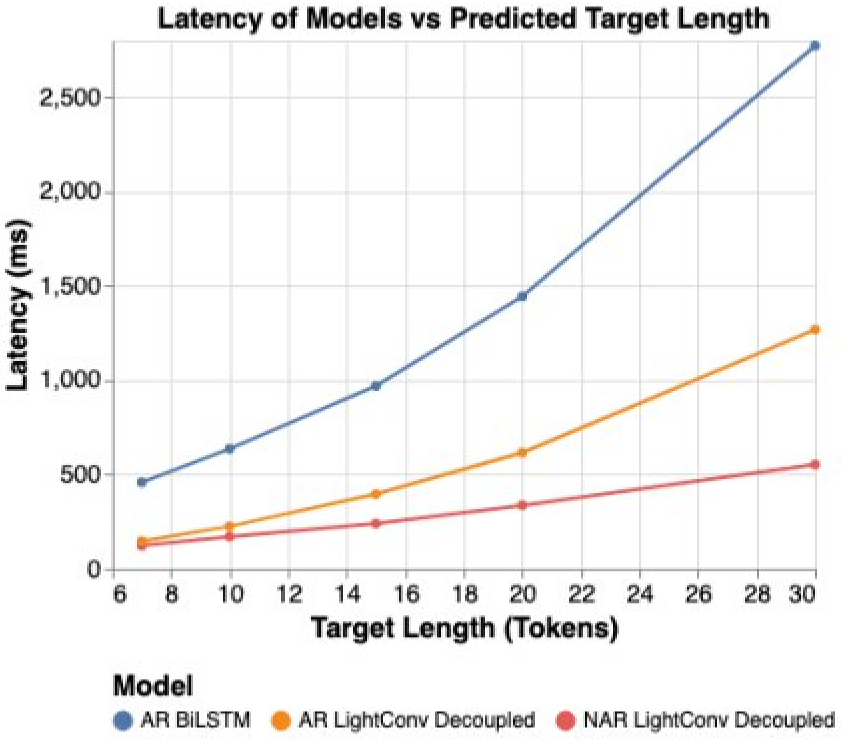


Non-Autoregressive Semantic Parsing for Compositional Task-Oriented Dialog

Результаты

	Exact Match Accuracy		
Model	**TOP**	**DSTC2**	**SNIPS**
RNNG (Einolghozati et al., 2018)	80.86	-	-
Ptr Transformer (Rongali et al., 2020)	79.25	-	85.43
Ptr BiLSTM (Aghajanyan et al., 2020)	79.51	88.33	-
GLAD (Zhong et al., 2018)	-	79.4	-
JointBiRNN (Hakkani-Tür et al., 2016)	-	-	73.20
Slot Gated (Goo et al., 2018)	-	-	75.50
Capsule NLU (Zhang et al., 2018)	-	-	80.90
	Ours		
NAR LightConv Pointer	80.20	88.16	80.86
AR LightConv Pointer	80.23	88.58	76.43

(a) Exact Match Accuracy on TOP, DSTC2, and SNIPS



(b) Median latency on TOP dataset

SpellMapper: A non-autoregressive neural spellchecker for ASR customization with candidate retrieval based on n-gram mappings

Задача

«Came from schleiddorf in wurteenberg»

User vocab: {“schlaitdorf”, “wurttemberg”}

«Came from schlaitdorf in wurttemberg»

(не очень впечатляющий слайд, я знаю)

SpellMapper: A non-autoregressive neural spellchecker for ASR customization with candidate retrieval based on n-gram mappings

Идея

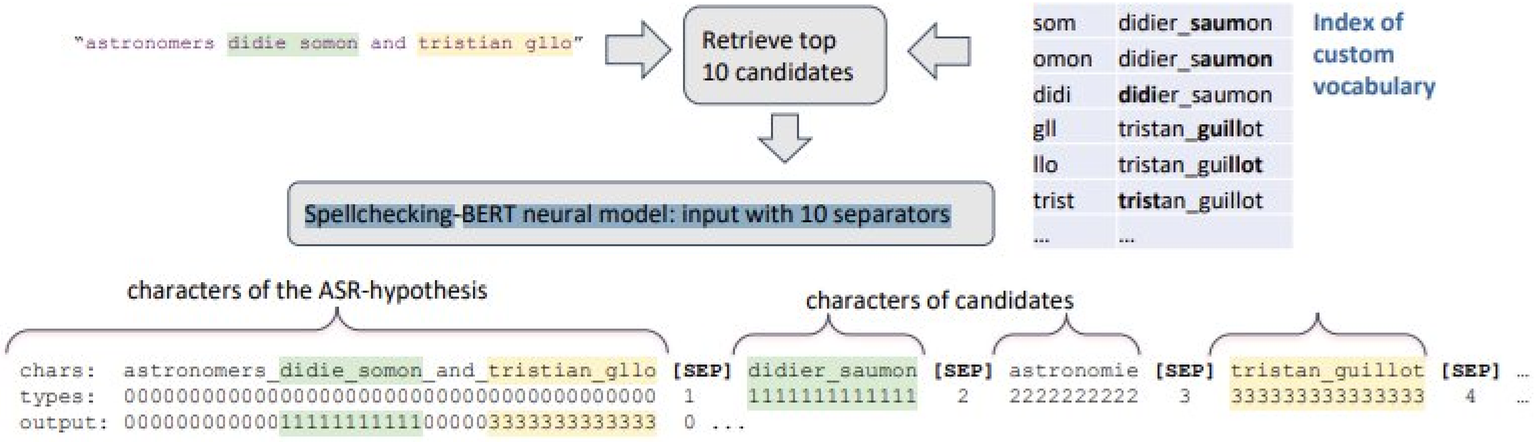


Figure 1: *Inference pipeline. Candidate retrieval ingests ASR output fragment of 10-15 words and selects top 10 user phrases with many matching n-grams in the user vocabulary index. ASR-hypothesis and candidates are combined to a single BERT input sequence with separators. The model predicts labels of correct candidates for each character of ASR-hypothesis, or 0.*

SpellMapper: A non-autoregressive neural spellchecker for ASR customization with candidate retrieval based on n-gram mappings

Результаты

Dataset	Users	Utte-rances	Max/Avg vocab size	ASR Model	Baseline WER %	Spellcheck WER %	"Ideal" WER %	Recall %	Precision %	Top 10 recall %
SWC	1341	61370	1341/172	CTC	6.69	5.26	4.46	67	87.4	90.4
				RNNT	6.07	4.93	4.12	62.9	85.6	88.1
SPGI	1114	39341	86/21	CTC	5.88	5.50	5.33	55	88.7	84.5
				RNNT	5.71	5.32	5.12	53.2	88.5	82.7
UserLibri	99	5559	192/43	CTC	3.35	2.89	2.52	61.7	82.3	83.9
				RNNT	2.77	2.44	2.09	54.5	77.6	74.1

SpellMapper: A non-autoregressive neural spellchecker for ASR customization with candidate retrieval based on n-gram mappings

Результаты

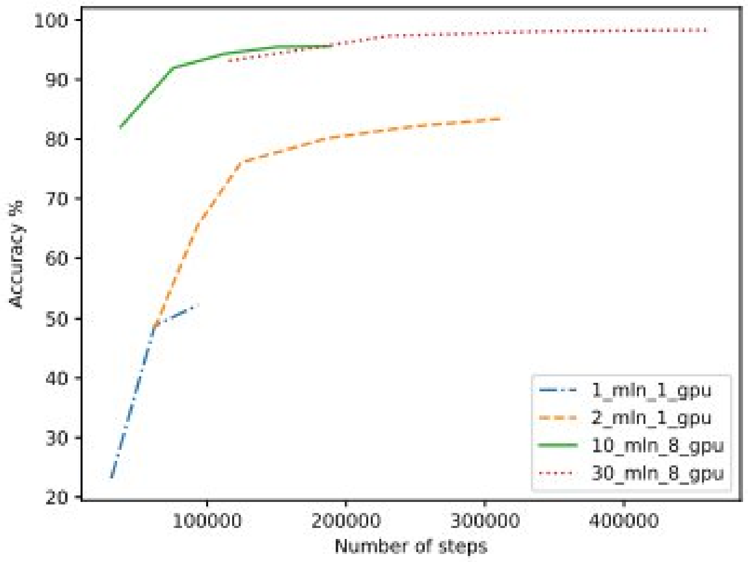


Figure 4: *SpellMapper accuracy on classes 1-10 depending on the number of training examples (1, 2, 10, 30M) evaluated on 20K validation examples. The whole phrase is considered correct if all symbols are tagged correctly. On class 0 (not presented on the graph) accuracy per symbol is 93%, 95.9%, 97.3%, 97.5% for corresponding corpus sizes*

Всем спасибо!