

# Сегментация объектов на видео и изображениях по текстовому описанию

Курсовая работа

Айрапетьянц Каринэ Арсеновна

*Научный руководитель:*

*Малоян Нарек Гагикович*

Московский государственный университет имени М. В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра информационной безопасности

5 апреля 2022 г.



# Содержание

Referring object segmentation

Актуальность

Цель

Language as Queries for RVOS

Архитектура

Vision-Language Transformer and Query Generation for Referring Segmentation

Архитектура

Примеры

Loss

Метрики

IoU

mAP



## Referring object segmentation

- ▶ **Referring video object segmentation (R-VOS)** -задача сегментации объекта на видео по его описанию на натуральном языке.
- ▶ **Referring image segmentation** - задача сегментации объекта на изображении по его описанию на натуральном языке.



# Актуальность

- ▶ видеонаблюдение
- ▶ приложения для редактирования изображений и видео
- ▶ обнаружение патологий на медицинских изображениях
- ▶ взаимодействие человека и робота посредством языка

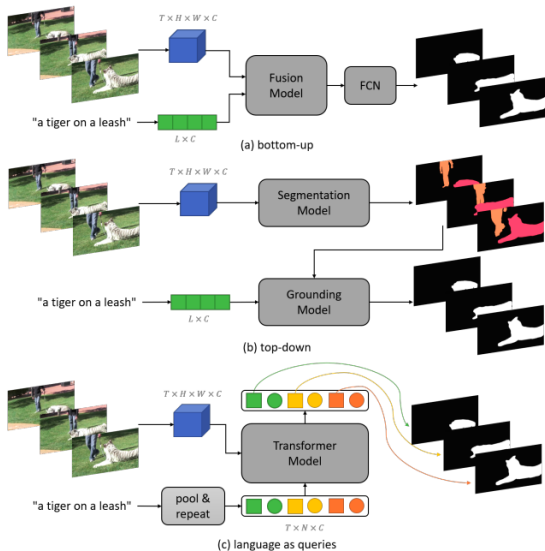


## Цель

- ▶ Дан видеоклип  $V = \{I_t\}_{t=1}^T$ , содержащий  $T$  фреймов и текстовое описание  $E = \{e_l\}_{l=1}^L$ , состоящее из  $L$  слов.  
Цель: создать маску сегментации описываемого объекта  $S = \{s_t\}_{t=1}^T, s_t \in \mathbb{R}^{H \times W}$ , для каждого фрейма.
- ▶ Дано изображение  $I$  и текстовое описание  $E$ .  
Цель: создать маску сегментации описываемого объекта.



# Language as Queries for RVOS



# Содержание

Referring object segmentation

Актуальность

Цель

Language as Queries for RVOS

Архитектура

Vision-Language Transformer and Query Generation for Referring Segmentation

Архитектура

Примеры

Loss

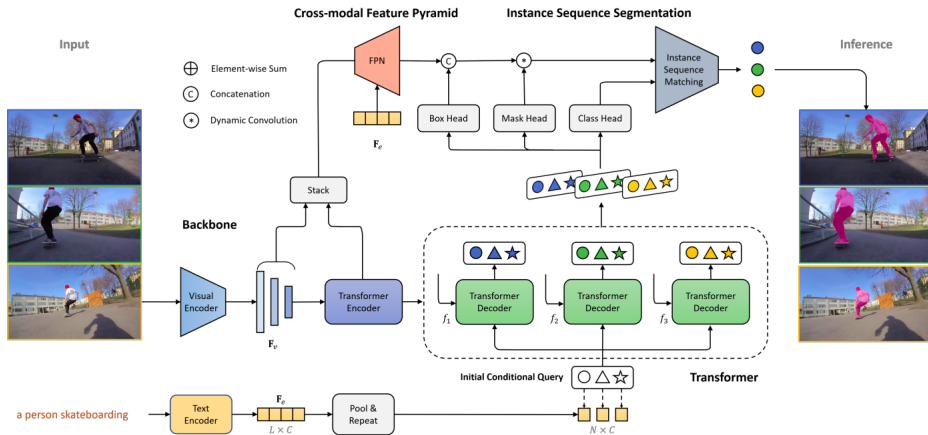
Метрики

IoU

mAP

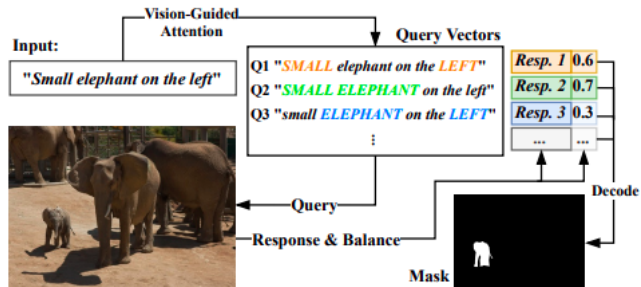


# Language as Queries for RVOS. Архитектура





# Queries



# Содержание

Referring object segmentation

Актуальность

Цель

Language as Queries for RVOS

Архитектура

Vision-Language Transformer and Query Generation for Referring Segmentation

Архитектура

Примеры

Loss

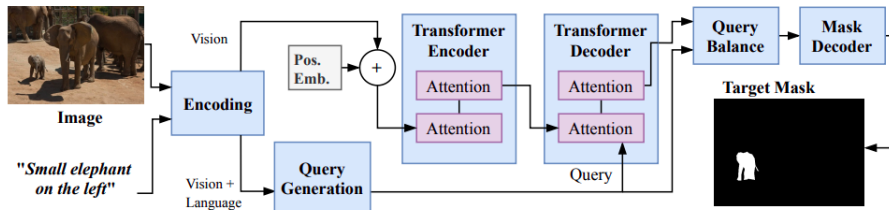
Метрики

IoU

mAP



# Vision-Language Transformer and Query Generation for Referring Segmentation. Архитектура



# Примеры

- ▶ Language as Queries for RVOS
- ▶ MTTR



Image (a)



"White bowl on corner"



"Bowl of carrots"



Image (b)



"Black cat"



"Lighter color cat"



Image (c)



"Guy with stripes"



"White shirt"



Image (d)



"Floral pattern"



"Green shirt"



Image (e)



"Curled tail"



"Elephant with rider"



Image (f)



"woman at 9 o'clock with white coat"



"Man kneeling in gray suit"



# Loss

- ▶  $\bar{y} = \{\bar{y}_i\}_{i=1}^N$  - множество предсказаний
- ▶ для каждого  $i$  предсказание имеет вид:  $\bar{y}_i = \{\bar{p}_i^t, \bar{b}_i^t, \bar{s}_i^t\}_{t=1}^T$
- ▶ правильный ответ имеет вид:  $y = \{c^t, b^t, s^t\}_{t=1}^T$ 
  - ▶  $\bar{y}_{pos} = \underset{\bar{y}_i \in \bar{y}}{\operatorname{argmin}} L_{match}(y, \bar{y}_i)$
  - ▶  $L_{match}(y, \bar{y}_i) = k_{cls} L_{cls}(y, \bar{y}_i) + k_{box} L_{box}(y, \bar{y}_i) + k_{mask} L_{mask}(y, \bar{y}_i)$



# Содержание

Referring object segmentation

Актуальность

Цель

Language as Queries for RVOS

Архитектура

Vision-Language Transformer and Query Generation for Referring Segmentation

Архитектура

Примеры

Loss


Метрики

IoU

mAP



► IoU:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


The diagram illustrates the components of the IoU formula. The top part shows two overlapping squares; the intersection of the two squares is shaded blue, representing the 'Area of Overlap'. The bottom part shows the union of the two squares, which is the total area covered by both squares, represented by a single blue shape.



# Содержание

Referring object segmentation

Актуальность

Цель

Language as Queries for RVOS

Архитектура

Vision-Language Transformer and Query Generation for Referring Segmentation

Архитектура

Примеры

Loss

Метрики

IoU

mAP



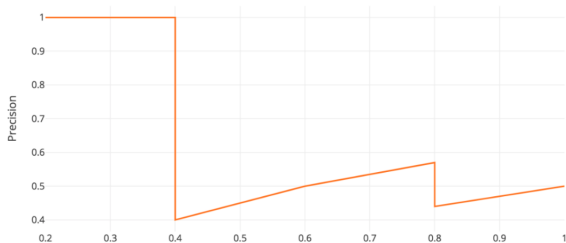


►  $mAP$ :

►  $p = \frac{TP}{TP+FP}, \quad r = \frac{TP}{TP+FN}$

►  $AP = \int_0^1 p(r) dr$

►  $mAP = \frac{1}{n} \sum_{i=1}^n AP_i$ ,  $n$ —количество классов,  $AP_i$  —  $AP$  для  $i$ -го класса



СПАСИБО ЗА ВНИМАНИЕ!

