

MDETR - Modulated Detection for End-to-End Multi-Modal Understanding

Практик-Исследователь: Денисенко Наталья Андреевна

Задача, поставленная перед авторами, заключается в получении регионов изображения на основе его текстового описания. Основное достоинство метода перед своими предшественниками заключается в усовершенствовании пайплайна, для того, чтобы отказаться от предположения детектора как black box. Полученных улучшений было достаточно, чтобы MDETR попал в топ 3% по поданным на ICCV 2021 статьям, а также, оказался на [обложке](#). Несмотря на то, что финальная версия была выложена 12 октября 2021, у неё уже 22 цитирования. В основном, методы, описанные в статьях, упоминающих MDETR, довольно близки к нему по тематике:

- [End-to-End Referring Video Object Segmentation with Multimodal Transformers](#) - выделение объектов на видео по текстовому описанию
- [QVHIGHLIGHTS](#) - выделение хайлайтов видео по текстовому описанию
- [SORNet](#) - генерация описания расположения объектов на сцене по отношению других объектов

Никаких официальных рецензий на данную статью пока не выложено, однако авторы сами упоминают найденные проблемы в их методе. И хотя большая часть проблем относится к недочетам в датасете, можно выделить ухудшение качества предсказаний, в случае, когда описание не содержит все объекты представленные на изображении. Мне кажется, что это может быть направлением по развитию статьи.

Что касается применений данного метода, то на [сайте](#) приведено 5 возможных направлений:

Object Localization	1	20.00%
Phrase Grounding	1	20.00%
Question Answering	1	20.00%

Referring Expression Segmentation	1	20.00%
Visual Question Answering	1	20.00%

Наибольшее влияние на MDETR оказал метод представленный [тут](#) над ним работали двое из авторов (Nicolas Carion и Gabriel Synnaeve).

Основная информация по авторам:

- авторы либо как-то относятся к NYU, либо к Facebook AI research team, либо и к первому, и ко второму
 - **Aishwarya Kamath** - PhD из NYU, работает в лаборатории Prof. Yann LeCun and Prof. Kyunghyun Cho по направлению CV и NLP
 - **Mannat Singh** - *Facebook AI Research*; CalTech
 - **Yann LeCun** - *Silver Professor of Data Science, Computer Science, Neural Science, and Electrical Engineering* из NYU, а также директор *AI Research Team at Facebook*
 - **Gabriel Synnaeve** - *Facebook AI Research*; PhD in Bayesian modeling applied to real-time strategy games Grenoble (франция)
 - **Ishan Misra** - *Facebook AI Research*; PhD Carnegie Mellon University; 6 его статей были приняты на ICCV 2021 (3 из них для oral)
 - **Nicolas Carion** - *PostDoc at NYU Courant*; есть твиттер, где он пишет только про ML, зато почти 1000 подписчиков