

УДК: 004.021

EDN: [OZBBQH](#)



## Коммуникация параметризованных нейронными сетями агентов посредством рисования и распознавания набросков

**А.А. Гараев\***

Казанский национальный исследовательский технический университет им.  
А.Н. Туполева – КАИ, г. Казань, Россия

\*E-mail: [30101998a@gmail.com](mailto:30101998a@gmail.com)

**Аннотация.** В данной статье рассматриваются принципы и особенности разработки нейронных сетей для задач рисования и распознавания изображений. Изучается возможность использования набросков изображений в качестве средства коммуникации агентов искусственного интеллекта, что позволяет создать канал связи между интеллектуальными агентами и глубже изучить возможности искусственного интеллекта к коммуникации. В качестве средств разработки использован язык программирования Python совместно с библиотекой искусственного интеллекта Pytorch. Итогом работы стала разработанная модель, состоящая из двух агентов - отправителя и получателя, где отправитель рисует полученное изображение в виде наброска, а цель получателя - распознать по наброску реальное изображение. Была доказана возможность создания визуального канала связи в виде набросков изображений между агентами, продемонстрированы результаты ее работы.

**Ключевые слова:** нейронные сети, распознавание изображений, Python, наброски, PyTorch

## Communication of parameterized by neural networks agents through drawing and recognition of image sketches

**A.A. Garaev\***

Kazan National Research Technical University named after A. N. Tupolev – KAI

\*E-mail: [30101998a@gmail.com](mailto:30101998a@gmail.com)

**Abstract.** This paper discusses the principles and features of the development of neural networks for the tasks of drawing and image recognition. The possibility of using image sketches as a means of communication for artificial intelligence agents is being studied, which allows a deeper study of the possibilities of artificial intelligence for communication. As development tools, the Python programming language was used in conjunction with the Pytorch artificial intelligence library. The result of the work was the developed model, consisting of two agents - the sender and the recipient, where the sender draws the received image in the form of a sketch, and the recipient's goal is to recognize the real image from the sketch. The possibility of creating a visual communication channel in the form of sketches of images between agents was proved, and the results of its work were demonstrated.

**Keywords:** neural networks, image recognition, Python, sketches, PyTorch

## 1. Введение

Научная задача, решаемая в данной работе, заключается в разработке методов искусственного интеллекта, позволяющих агентам использовать изображения в качестве способа коммуникации друг с другом и внешним миром. Объектом исследования являются методы и средства разработки искусственного интеллекта, а именно архитектуры нейронных сетей для решения целей и задач работы. Целью работы является создание визуального канала связи между агентами искусственного интеллекта, анализ результатов их коммуникации.

Для достижения поставленной цели решаются следующие задачи:

- 1) Обзор предметной области, а именно способов и методов искусственного интеллекта распознавания изображений.
- 2) Определение архитектуры модели для решения поставленной задачи.
- 3) Разработка модели с использованием выбранного языка программирования и необходимых библиотек.
- 4) Оценка результатов работы модели.

## 2. Обзор предметной области

Нейронные сети уже сегодня позволяют компьютерам принимать разумные решения при ограниченной помощи человека. Они способны сформировать и смоделировать отношения между сложными и нелинейными входными и выходными данными [2]. Например, они могут выполнять следующие задачи:

- Делать обобщения и выводы

Нейронные сети способны распознать, что два разных входных предложения имеют одинаковое значение. Например: «Не подскажете как произвести оплату? Как перевести деньги?» Несмотря на то, что данные предложения различны по своей структуре, нейронная сеть распознает, что оба предложения означают одно и то же и будет знать какой ответ предоставить.

- Создавать автономные самообучающиеся системы

Нейронные сети могут обучаться и улучшаться с течением времени в зависимости от поведения пользователя. Например, рассмотрим нейронную сеть, которая автоматически исправляет или предлагает слова, анализируя ваше поведение при наборе текста. Предположим, что модель обучена английскому языку и может проверять

орфографию английских слов. Однако, если вы часто вводите неанглийские слова, нейронная сеть также может автоматически запоминать и исправлять и эти слова [2].

- Изучать и моделировать сложные и изменчивые данные

Некоторые наборы данных, например суммы погашения кредита в банке, могут быть вариативными и иметь большие различие. Нейронные сети способны моделировать и такие данные. Например, они могут анализировать финансовые транзакции и помечать некоторые из них для обнаружения мошенничества. Они также могут обрабатывать сложные данные, которые являются ключом к решению сложных биологических проблем, таких как сворачивание белков, анализ ДНК и т. д. Нейросети для распознавания изображений также широко используются для распознавания различных патологий головного мозга, рака [5, 6].

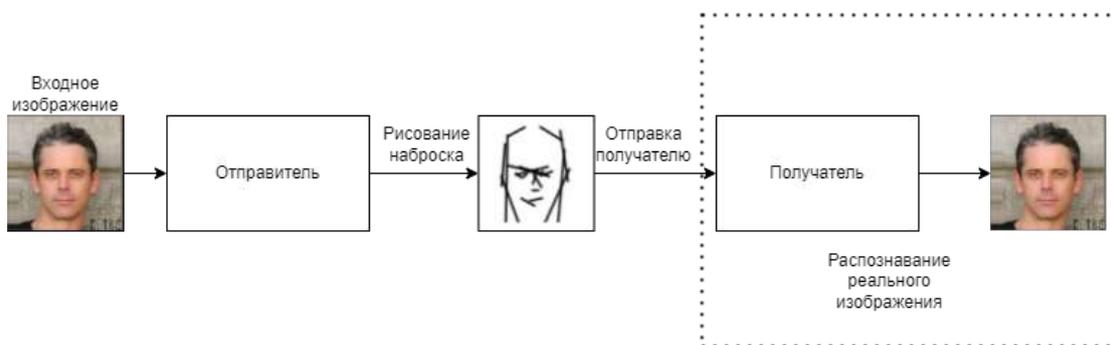
Как видно из всего вышеописанного, нейронные сети для распознавания изображений сегодня используются в самых различных сферах нашей жизни, совмещая самые разные подходы и способы применения. Искусственный интеллект уже сегодня позволяет упростить некоторые задачи человека и сделать его жизнь проще. Но как достичь того, чтобы искусственный интеллект действительно начал обретать подобие человеческого разума, стал способен к общению и коммуникации? Существуют неопровержимые доказательства того, что визуальная коммуникация предшествовала появлению письменного языка и послужила основой для его появления [1,3]. Древние люди использовали пещерную живопись и наскальные рисунки для общения друг с другом задолго до появления языков. Данная работа направлена на изучение того, как агенты могут научиться общаться, чтобы совместно решать задачи. Существующие работы в данной области были сосредоточены на передаче последовательностей дискретных токенов между агентами [1,4]. Данная работа посвящена исследованию того, как в качестве канала коммуникации можно использовать рисование простых набросков (или скетчей) изображений.

### 3. Разработка модели и результаты

Коммуникация возникает, когда два или более участников вовлечены в процесс общения, разделяют общую цель, решают определенную задачу или хотят достичь определенной цели. Этого можно достичь путем передачи информации. Различные исследования происхождения человеческих языков подтверждают тот факт, что первопричиной для появления языка стала кооперация людей для решения совместных

задач. «Сотрудничество» людей ускорило процесс эволюции общения, так как эволюция подразумевает, что несколько людей должны самоорганизоваться и приспособиться к одному и тому же языку.

Представим, что мы играем с кем-то в игру, в которой вы должны заставить своего друга угадать предмет в комнате, нарисовав его. Никакое другое общение, кроме нарисованного изображения, не допускается. Чтобы играть в эту игру, вам нужно научиться рисовать так, чтобы наш друг мог понять, что мы имеем в виду. В этой работе исследуется, как интеллектуальные агенты, параметризованные нейронными сетями, могут научиться играть в похожие игры с рисованием. Более конкретно, что один агент рисует эскиз данной фотографии, а второй агент угадывает на основе рисунка соответствующую фотографию из набора изображений (рисунок 1).



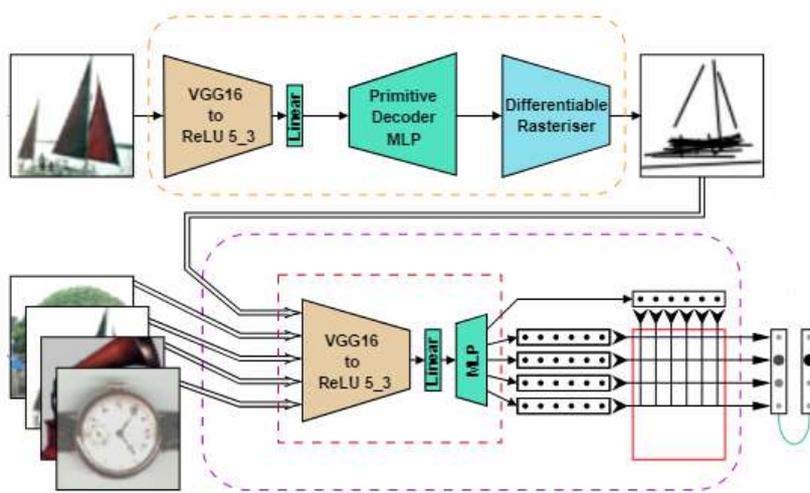
**Рисунок 1.** Общая структура работы разработанной модели.

Цель отправителя – создать набросок (скетч) из исходной фотографии. В экспериментах рисование скетчей ограничивается 20 черными прямыми линиями одинаковой ширины на белом фоне. Размер фона для рисования такой же, как у входных изображений. Получив входное изображение, ранняя зрительная система агента создает векторную кодировку, которая затем обрабатывается трехслойным многослойным перцептроном (MLP), который декодирует данные параметры. Данный перцептрон имеет функцию активации ReLU на первых двух слоях и активацию tanh на последнем слое. Первый слой состоит из 64 нейронов, второй из 256 нейронов. Выходной слой выдает четыре значения для каждой линии, которая будет нарисована; значения представляют собой начальные и конечные координаты каждой линии на холсте изображения с исходной точкой в центре. Для рисования изображения используется библиотека.

Эксперименты проводились с датасетом Caltech-101. Он состоит из изображений объектов, принадлежащих к 101 классу, плюс один класс фоновых помех. Каждое

изображение помечено одним объектом. Каждый класс содержит примерно от 40 до 800 изображений, всего около 9 тысяч изображений. Изображения имеют переменные размеры, с типичной длиной края 200-300 пикселей [7].

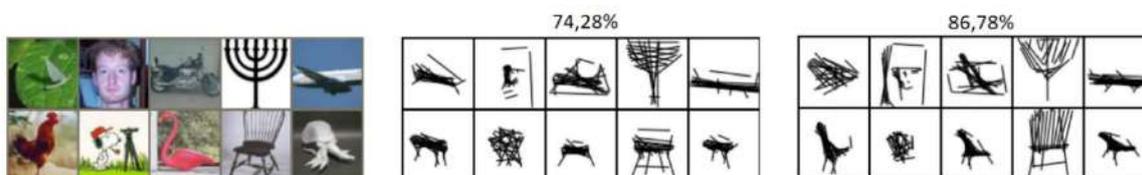
Разработанная модель состоит из отправителя и получателя. Цель отправителя нарисовать полученное изображение, а получатель учится правильно распознать правильное изображение по нарисованному. Агент «отправитель» представлен с изображением и зарисовывает его содержимое с помощью обучаемой процедуры рисования. Агент «получатель» предъявляется эскиз и коллекция фотографий, и он должен научиться правильно ассоциировать эскиз с соответствующей фотографией, предсказывая оценки, которые сравниваются с одним горячим целью. Оба агента параметризованы нейронными сетями, обученными сквозным методом с использованием градиентных методов. Архитектура модели представлена на рисунке 2.



**Рисунок 2.** Архитектура разработанной модели.

Агенту-получателю предоставляется набор фотографий и эскиз изображения, и он отвечает за предсказание того, какая фотография соответствует эскизу в соответствии с правилами конкретной игры. Зрительная система приемника связана с двухслойным MLP с нелинейностью ReLU на первом слое (последний слой не имеет функции активации). Если не указано иное, во всех экспериментах используется 64 нейрона в первом слое и 64 в последнем слое. Эскизное изображение и каждая фотография независимо проходят через визуальную систему и MLP для создания представления вектора признаков соответствующего ввода. Вектор оценки  $x$  создается для фотографий путем вычисления скалярного произведения элемента эскиза с элементом каждой соответствующей фотографии. Этот вектор оценок не нормализован, но его можно

рассматривать как распределение вероятностей, пропустив его через softmax. Фотография с наивысшим баллом является правильно предсказанной. Оптимизация параметров обоих агентов выполняется с помощью оптимизатора Adam с начальной скоростью обучения  $1 \times 10^{-4}$  для всех экспериментов. Также были проведены эксперименты с более широкой моделью. В ней агент отправитель кодирует целевую фотографию в 1024-мерный вектор (базовая модель кодирует в 64-мерный вектор), а емкость MLP приемника также увеличена с 64 до 1024. Результаты работы базовой и более широкой модели показана на Рисунке 3. Более широкая модель позволяет рисовать более понятные для человеческого глаза изображения, а также демонстрирует более высокую точность распознавания – 86.78% против 74.28% у базовой.



**Рисунок 3.** Результаты экспериментов.

#### 4. Заключение

В результате работы была разработана модель нейронной сети для решения задачи коммуникации между агентами искусственного интеллекта. Была доказана возможность создания визуального канала связи в виде набросков изображений между агентами, продемонстрирована рабочая модель и результаты ее работы.

#### Список литературы

1. Шолле, Франсуа. Глубокое обучение на Python / Ф. Шолле. – New York, 2018. – 400 с.
2. Hoffmann, Dirk L. U-th dating of carbonate crusts reveals neandertal origin of iberian cave art / Dirk L Hoffmann, Christopher D Standish, Marcos García-Diez, Paul B Pettitt, James A Milton, João Zilhão, Javier J Alcolea-González, Pedro Cantalejo-Duarte, Hipólito Collado, Rodrigo De Balbín, et al. // Science. – 2018. – № 359(6378). – P. 912-915.
3. Amores, Jaume. Multiple instance classification: Review, taxonomy and comparative study / Jaume Amores // Artificial intelligence. – 2013. – № 201. – P. 81-105.

4. Cao, Kris. Emergent communication through negotiation / Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z Leibo, Karl Tuyls, and Stephen Clark // In International Conference on Learning Representations. – 2018.
5. comidor.com: сайт. – 2019. – URL: <https://www.comidor.com/knowledge-base/machine-learning/image-recognition-use-cases/> (дата обращения: 01.07.2022).
6. insights.daffodilsw.com: сайт. – 2021. – URL: <https://insights.daffodilsw.com/blog/7-amazing-examples-of-image-recognition-apps> (дата обращения: 01.07.2022).
7. tensorflow.org: сайт. –2018. –URL: <https://www.tensorflow.org/datasets/catalog/caltech101> (дата обращения: 03.07.2022).