Интерактивные примеры работы нейронных сетей

Найдите объекты в кадре, отделите их от фона и превратите голос в текст.

Нейросеть может обвести кота на фотографии, отделить его шерсть от дивана или описать всю сцену словами. Фотография одна — задачи разные. Начнём с опытов: будем менять входные данные, наблюдать ответы и разбираться, что именно удалось модели. Устройство нейронов подождёт до следующей главы.

Здесь работают готовые обученные модели. Начните с поиска объектов: у него самая небольшая загрузка. Скачивайте только тот пример, который хотите попробовать; фото, звук и ответы обрабатываются на вашем устройстве.

1. Найти объекты и проследить за ними

На фотографии можно спросить не только «есть ли кот?», но и «где каждый кот?». Это задача **обнаружения объектов**. Вход — изображение. Выход — список найденных объектов: для каждого нужны название категории, прямоугольная рамка и оценка уверенности модели.

Попробуйте. Скачайте модель и запустите её на фотографии. Затем откройте «Пример видео». Посмотрите, сохраняются ли номера котов при движении. Можно загрузить собственное фото или видео.

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

YOLO26n



Рамки появятся на изображении

Фото или видео Камера Пример видео

Скачать · 10 МБ

▼ О модели и ограничениях

Нейросеть распознаёт объекты на каждом кадре. Простой алгоритм сопоставляет рамки между кадрами; при перекрытии или резком движении номера могут меняться. Скорость зависит от устройства.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

YOLO26n · исходники и лицензия ↗

Рамки показывают расположение объектов; номера помогают проследить за ними на видео.

Интерактивный эксперимент — в электронной версии урока.

Рамка охватывает объект целиком и обычно захватывает немного фона. Модель ищет категории из своего заранее заданного списка: отсутствие рамки не доказывает, что на снимке ничего нет. Проверьте, все ли заметные объекты найдены и правильно ли подписаны.

В видео сеть заново находит объекты на каждом кадре. Дополнительный алгоритм сопоставляет соседние рамки и присваивает им номера — это **отслеживание объектов**. При перекрытии или резком движении номера могут перепутаться. Встроенный пример специально прост: движется фотография, а не сами коты. Собственное видео позволяет устроить более трудную проверку.

Заметка
Рамка и подпись — предсказание, а не доказательство. Иногда модель уверенно путает категории. Для проверки сравните подпись с самим изображением.

> Что нужно получить от модели, чтобы найти каждого кота на фотографии?
>
> ○ Только слово «животные» для всей фотографии
>
> ○ Точное расстояние до дивана в метрах
>
> ○ Имена котов и их характеры
>
> ○ Отдельную рамку и категорию для каждого найденного кота

2. Отделить объект от фона

Для наклейки с котом прямоугольника недостаточно: вместе с ним вырежется диван. Нужен ответ для каждой точки изображения — сохранить её или сделать прозрачной. Задача **выделения переднего плана** даёт гораздо более точный контур, чем рамка.

Попробуйте. Уберите фон и рассмотрите усы, шерсть и промежуток между котами. Затем откройте свою фотографию: легче ли выделяется предмет на однотонном фоне?

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

BiRefNet Lite



Своё фото

Результат без фона

Скачать · 124 МБ

▼ О модели и ограничениях

Сложные края, шерсть и прозрачные предметы могут выделяться неточно. Шахматная подложка обозначает прозрачность.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

BiRefNet Lite · исходники и лицензия ↗

Шахматная подложка обозначает прозрачность. Результат можно сохранить как PNG.

Интерактивный эксперимент — в электронной версии урока.

Вход — фотография, выход модели — маска: значения, определяющие, какую долю каждой точки сохранить. По этой маске программа делает фон прозрачным. На сложных краях возможны пропуски и остатки фона. Сеть выбирает заметный передний план; это не инструмент точного выбора любого предмета по вашему указанию.

Сравните этот результат с рамками из прошлого опыта. Обе сети работают с изображением, но отвечают на разные вопросы: «где объект?» и «какие точки оставить?». Именно нужный результат определяет выбор задачи.

> Нужно сделать наклейку в форме кота, без дивана вокруг. Какой результат полезнее всего?
>
> ○ Прямоугольная рамка вокруг кота
>
> ○ Подпись «на фотографии кот»
>
> ○ Маска, задающая прозрачность каждой точки
>
> ○ Расшифровка звуков из комнаты

3. Увидеть глубину на плоской фотографии

Фотография плоская, но мы замечаем, что кот ближе стены, а спинка дивана — дальше его морды. Нейросеть тоже может искать признаки взаимного расположения объектов: перспективу, перекрытия, знакомые формы. Задача называется **оценкой глубины**.

Попробуйте. Сначала сами назовите самый близкий и самый далёкий участок фотографии. Затем получите карту глубины. Совпало ли ваше предположение с ответом сети?

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

Depth Anything 3 Small



Карта глубины

Своё фото

Скачать · 106 МБ

▼ О модели и ограничениях

Светлое — ближе, тёмное — дальше. Это относительная глубина по одному снимку, без расстояний в метрах.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

Depth Anything 3 Small · исходники и лицензия ↗

На карте светлое означает ближе, тёмное — дальше. Сравните границы предметов с фотографией.

Интерактивный эксперимент — в электронной версии урока.

Вход — один снимок, выход — карта относительной глубины, где каждой точке сопоставлена оценка её удалённости. Такая карта помогает отделять ближний план от дальнего, например для эффекта размытого фона. Это другая задача, чем выделение кота: глубина описывает всю видимую сцену.

4. Превратить голос в текст

Теперь поменяем тип данных. Для **распознавания речи** на вход поступает звук, а на выходе нужна последовательность слов. Такая модель полезна для субтитров и черновика конспекта. Она должна

восстановить произнесённое, включая имена и числа, а не сочинить подходящую по смыслу фразу.

Попробуйте. Откройте и прослушайте русский пример, затем запустите распознавание. После этого запишите короткую фразу с необычным именем или числом. Сверьте расшифровку с тем, что действительно прозвучало.

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

Whisper Large V3 Turbo

Записать Аудиофайл Пример

Расшифровка появится здесь

Скачать · 760 МБ

▼ О модели и ограничениях

Обрабатываются первые 30 секунд. На тишине и шуме модель иногда выдумывает слова. Сравните расшифровку с записью.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

Whisper Large V3 Turbo · исходники и лицензия ↗

Можно использовать готовую запись, свой файл или микрофон. В опыте обрабатываются первые 30 секунд.

Интерактивный эксперимент — в электронной версии урока.

Плавный и грамматически правильный текст ещё не означает точную расшифровку. На шуме, тишине и непривычных словах модель может пропустить или добавить слова. Проверять результат нужно по записи, особенно если ошибка в имени или числе меняет смысл.

> В расшифровке появилась гладкая фраза, которой вы не помните. Как проверить её?
>
> ○ Попросить модель сделать фразу ещё красивее
> ○ Прослушать соответствующий участок записи и сравнить слова
> ○ Убедиться, что в ней нет опечаток
> ○ Посмотреть на размер скачанной модели

Заметка

По одному снимку нельзя однозначно восстановить размеры и расстояния: маленькая модель предмета вблизи может выглядеть как большой предмет издалека. Это демо не измеряет метры. Оно оценивает относительную глубину, опираясь на закономерности из обучения, и может ошибаться на зеркалах, стекле и необычных ракурсах.

5. Задать вопрос о фотографии или кадре с камеры

Детектор выдаёт рамки и названия категорий. Но иногда нужен другой ответ: «Что лежит на диване?» или «Чем отличаются два кота?». Здесь работают вместе изображение и язык. **Мультимодальная модель** получает фотографию и текст вопроса, затем составляет текстовый ответ.

Попробуйте. Попросите описать фотографию одним предложением. Затем включите камеру и спросите, какие предметы видны или что человек держит в руке. Измените сцену и дождитесь следующего ответа. Быстрый вариант отвечает по-английски; для русского есть

отдельный компактный вариант. Каждая модель видит отдельные кадры, поэтому ответы обновляются с задержкой, а не на каждом кадре видео.

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

FastVLM · быстрее · English



Ответ на вопрос появится здесь

Фото или видео Камера

Пример видео

Вопрос · English

What is in the image? Answer briefly.

Быстрый вариант отвечает лучше по-английски.

Скачать · 842 МБ

▼ О модели и ограничениях

Отвечает лучше по-английски. Видит отдельные кадры; частота обновления зависит от устройства. Ошибается в деталях и не определяет характер человека по внешности.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

FastVLM · 0.5B · исходники и лицензия ↗

Сопоставляйте каждое утверждение с изображением. Уверенный тон не подтверждает, что деталь существует.

Интерактивный эксперимент — в электронной версии урока.

По внешности нельзя установить, «хороший» человек или «плохой», каков его характер и можно ли ему доверять. Проверяемые вопросы касаются видимых предметов и действий.

Модель может связать изображение с вопросом и сформулировать удобный ответ. Но она также может неверно посчитать предметы или придумать правдоподобную деталь. Сравните её описание с результатом детектора: подпись словами и список рамок имеют разные возможности и разные ошибки.

6. Поручить текстовую задачу

Языковая модель получает текст и генерирует продолжение. После специального обучения на заданиях и ответах она может выполнять инструкции: сокращать, переводить, объяснять или предлагать названия. В отличие от распознавания речи здесь часто допустимо много разных хороших ответов.

Попробуйте. Придумайте названия для кружка, затем попросите сократить готовый абзац. В первом случае оцените выразительность, во втором — сохранился ли смысл и не добавились ли новые сведения.

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

Qwen3.5 · 0.8B

Задание

Придумай три коротких названия для школьного кружка робототехники. Используй только русские слова.

Ответ появится здесь

Придумать Объяснить

Перевести Сократить

Скачать · 648 МБ

▼ О модели и ограничениях

Это маленькая модель, а не большой облачный ассистент. Попробуйте короткие задания; проверьте факты и счёт. Каждый запуск — новый диалог.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

Qwen3.5 · 0.8B · исходники и лицензия ↗

Начните с маленькой модели Qwen. Кнопки под заданием предлагают несколько разных задач.

Интерактивный эксперимент — в электронной версии урока.

Для сравнения ниже есть более крупная модель. Дайте ей **точно то же задание** и заранее выберите критерий: соблюдение ограничения, точность пересказа или удачность названий. Размер модели сам по себе не определяет победителя в конкретном опыте. Скачивание второго варианта необязательно.

Для этого опыта нужен браузер с WebGPU, например Chrome на компьютере.

Gemma 4 E2B · QAT

Задание

Придумай три коротких названия для школьного кружка робототехники. Используй только русские слова.

Ответ появится здесь

Придумать Объяснить

Перевести Сократить

Скачать · 2,3 ГБ

▼ О модели и ограничениях

Загрузка около 2,3 ГБ, рабочей памяти нужно больше. Эта мобильная QAT-версия сжимает веса до двух бит. Нужна поддержка shader-f16; лучше запускать на компьютере.

Веса скачиваются с Hugging Face, вычислительная библиотека — с jsDelivr. Фото, звук и тексты обрабатываются на устройстве. Рабочей памяти требуется больше, чем размер загрузки. При загрузке другой модели эта выгружается; ответ остаётся на странице.

Gemma 4 E2B · QAT · исходники и лицензия ↗

Около 2,3 ГБ загрузки и дополнительная рабочая память. При запуске другого примера предыдущая модель выгружается, а её ответ остаётся на странице.

Интерактивный эксперимент — в электронной версии урока.

Во всех этих опытах мы **применяли готовые сети**: загрузка фотографии или новый вопрос меняют вход, но не обучают модель заново. Это позволяет использовать одну обученную сеть много раз, для разных примеров подходящей ей задачи.

От результата — к устройству

Как сеть превращает числа изображения в рамку, цвет или слово? В следующей главе «Искусственный нейрон: входы, веса и выход» разберём самый маленький вычисляющий блок и начнём собирать ответ на этот вопрос.

Задачу определяют вход, нужный выход и критерий качества: рамка, маска, глубина и текст требуют разных проверок.

При обучении меняются настройки сети. При применении готовая сеть получает новые входные данные.

Правдоподобный результат нужно сверять с изображением, записью или условиями задания.

Домашняя работа

1. Выберите доступный опыт и испытайте его на двух разных входах. Запишите, что модель получила, что выдала и как вы проверили результат. Найдите удачный ответ и ошибку или ограничение. Если модели недоступны на вашем устройстве, разберите два результата с компьютера учителя.

2. Вы загрузили другое фото в готовый детектор. Что произошло в этом опыте?

   Любая ошибка на новом фото стала невозможна

   Список известных модели категорий автоматически расширился

   Модель заново обучилась на всех фотографиях мира

   Модель применила прежние настройки к новому входу

Источники

- TensorFlow.js: задачи и браузерные примеры (https://www.tensorflow.org/js/demos)
- Google Teachable Machine: обучение собственным примерам (https://teachablemachine.withgoogle.com/)
- Карточка конкретной модели и её лицензия доступны под каждым опытом, в пункте «О модели и ограничениях».