Интерактивные примеры работы нейронных сетей

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

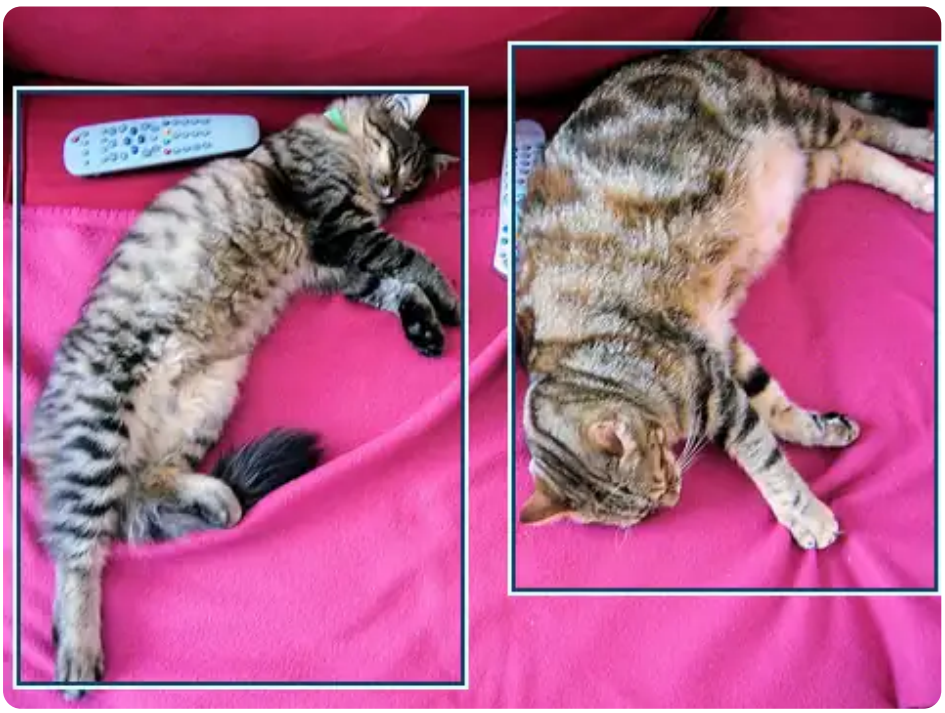
На розовом диване спят два кота. Хочется сделать из них наклейку для чата. Представьте, что вокруг каждого кота провели самую тесную прямоугольную рамку, в которую он помещается целиком. Получится ли наклейка без дивана, если вырезать оба прямоугольника?

Для первого занятия на 45 минут сравните рамки и маску на одном снимке, затем сверьте субтитры с записью. Результаты моделей уже в уроке. Если захотите проверить собственный снимок или видео, запустите модель на 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.

Посмотрите на левого кота. Что помимо него оказалось внутри рамки? Проследите лапу, вытянутую вниз. Целиком ли она войдёт в миниатюру?



Рамки YOLO26n для класса «кот».

Сверить наблюдение на снимке
Слева в рамке видны пульт и ткань между лапами.
А кончик вытянутой вниз лапы оказался чуть ниже нижней границы. При обрезке по рамке пульт сохранится, а часть лапы пропадёт.

В опыте ниже можно запустить детектор на другом фото. В «Примере видео» посмотрите, сохраняются ли номера котов при движении.

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.

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.

> Из фотографии нужны две прямоугольные миниатюры — по одной на каждого кота. Какого ответа детектора достаточно, чтобы редактор узнал, где провести границы обрезки?
>
> - ○ Координаты центра каждого кота
> - ○ Координаты отдельной рамки для каждого кота
> - ○ Координаты одной рамки вокруг обоих котов и число котов внутри
> - ○ Ширина и высота каждого кота без его положения на снимке

Заметка

Детектор ищет объекты из своего списка категорий. Если на фотографии виден предмет без рамки, модель могла его пропустить, а могла вообще не искать предметы этого вида.

2. Отделить объект от фона

Внутри рамки левого кота, между его лапами, видна розовая ткань. Для поиска кота такая рамка вполне удачна. А в наклейке лапы должны остаться, ткань между ними — исчезнуть. Теперь программе нужно решить, какую часть снимка оставить видимой. Эту задачу называют **выделением переднего плана**.

На результате второй сети найдите то место, где между лапами была розовая ткань. Исчезла ли ткань? Не пропала ли вместе с ней шерсть?



BiRefNet Lite обработала тот же снимок. В прозрачных местах виден фон страницы.

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

BiRefNet Lite



Результат без фона

Своё фото

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.

Сеть выдаёт **маску** — для каждого пикселя она оценивает, насколько тот должен остаться видимым. По этой маске программа делает фон прозрачным. Если на наклейке осталась полоска розовой ткани, сеть сохранила часть дивана. Если укоротился ус, она стёрла часть кота.

А если для наклейки нужен только левый кот? Даже когда сеть идеально убрала диван, общая маска может сохранить обоих. Она отделяет передний план от фона, но не указывает, где кончается один кот и начинается другой.

> Для наклейки нужно сохранить обоих котов и сделать прозрачной ткань между лапами. Какой результат позволит редактору вырезать такой контур?
>
> ○ Число пикселей, занятых котом внутри каждой рамки
> ○ Прямоугольная рамка вплотную к краям шерсти каждого кота
> ○ Маска с долей непрозрачности для каждого пикселя
> ○ Общее число пикселей с шерстью в каждой горизонтальной строке

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

Замена одного слова в субтитрах может изменить смысл фразы. При **распознавании речи** модель восстанавливает слова по звуку.

Какой из двух вариантов субтитров точно передаёт фразу? Сверьтесь с записью или с её текстом ниже.

Текст записи
«Сегодня мы проверяем, как нейронная сеть распознаёт русскую речь».

Оба варианта составлены автором для сравнения.

- **А.** Сегодня мы проверим, как нейронная сеть распознаёт русскую речь.
- **Б.** Сегодня мы проверяем, как нейронная сеть распознаёт русскую речь.

Сверить выбор
В записи слышно «проверяем» — говорящий описывает то, чем занят сейчас. Вариант Б передаёт его слова. В варианте А подмена на «проверим» превращает их в сообщение о планах.

А как справится сама нейросеть? В опыте ниже запустите распознавание той же записи или запишите фразу с необычным именем. Сверьте спорное слово с тем, что прозвучало. Для запуска потребуется скачать модель.

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.

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.

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 субтитрах разошлись две версии одного слова. Обе подходят по смыслу. Как узнать, какую оставить для точной цитаты?
>
> - ○ Выбрать слово, которое чаще встречается в субтитрах к таким видео
> - ○ Взять слово, которому программа распознавания приписала большую уверенность
> - ○ Сверить обе версии с кратким пересказом видео
> - ○ Прослушать соответствующий участок записи и сравнить слова

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 ком-

пактный 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 удач-

ность 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.

На чём учится сеть?

Для обучения детектора можно взять снимок с рамками вокруг котов. Если учить сеть убирать фон, к тому же снимку понадобится маска. Сравнивая ответ с образцом, программа настраивает веса — числа, от которых зависит результат сети. Чтобы проследить такое вычисление вручную, удобно взять один искусственный нейрон всего с двумя входами, как в задаче о координатах робота на карте.

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 модель не запускается, возьмите готовые рисунки с котами из первых двух опытов. Выберите деталь на фотографии и покажите на ней, какой из результатов лучше подходит для вашей наклейки.

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

   Список известных модели категорий автоматически расширился

   Любая ошибка на новом фото стала невозможна

   Модель применила прежние настройки к новому входу

   Модель заново обучилась на всех фотографиях мира

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».

Повторить эту тему