Урок 15. Автоэнкодер, GAN, авторегрессия и диффузия

9 класс · Урок 15 из 34 · 45 минут

Вся программа 9 класса

Четыре модели генерации изображений могут дать похожие картинки. Разберём, чем различаются их способы работы.

Содержание урока

- 01. Цель урока
- 02. Ход урока
- 03. Этап 1. Введение: Четыре мастерские · 5 мин
- 04. Этап 2. Код, состязание, последовательность, шум · 14 мин
- 05. Этап 3. Подбор модели по задаче · 15 мин
- 06. Этап 4. Что менялось в исследованиях · 6 мин
- 07. Этап 5. Итоги урока · 5 мин

01. Цель урока

Сравнить четыре подхода к генерации изображений. Объяснить, чему учится модель, как она строит изображение и по каким признакам сравнивать результаты.

02. Ход урока

Этап	Мин	Учитель	Ученики
Этап 1. Введение: Четыре мастерские	5	Показывает, почему одной оценки красоты мало.	Предлагают свойства, важные для разных проектов.
Этап 2. Код, состязание, последовательность, шум	14	Объясняет два шага каждого метода.	Сверяют рассказ с картой и приводят различие.
Этап 3. Подбор модели по задаче	15	Раздаёт четыре заявки на эскизы.	Выбирают и защищают подход, отмечают цену выбора.
Этап 4. Что менялось в исследованиях	6	Сводит результаты групп и показывает вехи.	Объясняют, какая идея пригодилась бы в другой задаче.

Этап	Мин	Учитель	Ученики
Этап 5. Итоги урока	5	Подводит итог и сверяет ответы.	Письменно отвечают и сверяют решение.
Итого	45		

03. Этап 1. Введение: Четыре мастерские 5 мин

ГОВОРИМ УЧЕНИКАМ

Школьная мастерская выбирает программу для эскизов эмблем. Одна программа даёт десятки похожих вариантов, другая позволяет плавно менять форму, третья дольше строит изображение. По какому свойству выбирать? Сначала разберём, какими путями модели генерации изображений получают новый рисунок.

Для учителя

Попросите назвать, где полезны скорость, разнообразие или управляемость. Подчеркните, что свойства конкретной программы надо измерять. Название семейства подсказывает принцип работы.

04. Этап 2. Код, состязание, последовательность, шум 14 мин

ГОВОРИМ УЧЕНИКАМ

Представьте, что у нас много эскизов простых светильников. Вариационный автоэнкодер учится переводить изображение в короткое числовое описание. Для каждого примера он задаёт область возможных кодов, а не единственную точку. Вторая часть сети, декодер, восстанавливает изображение по коду. При создании новых вариантов можно выбрать код из области, в которой модель училась работать. Соседние коды часто дают похожие формы, поэтому удобно искать плавные переходы между эскизами.

В генеративно-состязательной сети, GAN, при обучении работают две сети. Генератор создаёт образец, а дискриминатор учится отличать сгенерированное изображение от обучающего. Его оценка помогает генератору исправлять ошибки. После обучения новые изображения создаёт генератор. Иногда генератор находит несколько удачных решений и начинает повторять их. Тогда качество отдельных изображений и разнообразие всей подборки надо проверять отдельно.

Авторегрессионный подход превращает изображение в последовательность элементов. Это могут быть пиксели, как в Image GPT, или более крупные кодированные фрагменты в других моделях. Модель предсказывает очередной элемент с учётом предыдущих. Так можно продолжить заданное начало. При длинной последовательности шагов много, а ранний выбор влияет на то, что достроится дальше.

Диффузионная модель учится убирать шум. На обучении исходные изображения искажают шумом разной силы и показывают модели, как убрать часть искажения. При

генерации модель обычно начинает со случайного шума и постепенно получает более ясную сцену. В управляемых моделях на шаги влияет текст или изображение. Несколько последовательных шагов дают возможность уточнять детали, но требуют времени.

Подход	Чему учится	Как строит новый вариант
VAE	Кодировать и восстанавливать примеры	Декодирует выбранный код.
GAN	Генератор учится создавать новые изображения по оценке другой сети	Создаёт изображение генератором.
Авторегрессия	Предсказывать очередной элемент последовательности	Добавляет элементы один за другим.
Диффузия	Убирать шум с разных уровней искажения	Постепенно превращает шум в изображение.

05. Этап 3. Подбор модели по задаче 15 мин

ГОВОРИМ УЧЕНИКАМ

Теперь можно вернуться к мастерской. Если нужен плавный переход между формами, вам пригодится работа с соседними кодами. Если важно сохранить заданное начало, полезно сравнить подходы, которые достраивают последовательность или умеют удерживать условие. Для одной задачи могут подойти разные методы. Объясните, какой выберете первым и как проверите результат.

Консультация для мастерской

Выберите одну заявку. Объясните, какой подход испытали бы первым, как он создаёт изображение и какой результат заставил бы попробовать другой.

Плафоны. Нужна серия форм между круглой и вытянутой для обсуждения с мастером.

Узор. Есть первые строки пиксельного орнамента. Нужны продолжения, учитывающие начало.

Значок. На стенд требуется много разных вариантов одного знака за короткое время.

Иллюстрация. Нужны несколько вариантов по подробному тексту, где важны отношения предметов.

Обменяйтесь предложением с другой группой. Пусть она назовёт один риск вашего способа. Добавьте к предложению измеримый критерий.

Для учителя

Возможные первые пробы: VAE для переходов, авторегрессия для орнамента, GAN для быстрых вариантов, условная диффузия для иллюстрации. Принимаются другие обоснованные решения. Критерии: плавность изменений, сохранность

начала и ритма, число действительно разных вариантов за время, доля соблюдённых отношений.

06. Этап 4. Что менялось в исследованиях 6 мин

ГОВОРИМ УЧЕНИКАМ

Работы о VAE и GAN в 2013–2014 годах показали разные способы учить систему создавать новые примеры. В 2020 году авторы Image GPT перенесли идею предсказания следующего элемента с текста на пиксели. В том же году работа о диффузии показала, что модель может получать изображение через цепочку шагов очистки от шума. Позднейшие системы иногда сочетают идеи: например, диффузия может идти в сжатом представлении. Управляемая диффузия также позволяет дорисовывать выделенный участок по подсказке. Такой подход полезен в редакторе, когда меняется только часть изображения. Детали вдоль границы правки всё равно нужно проверить. Поэтому полезнее спрашивать, что именно делает каждый компонент системы, чем выбирать победителя по дате.

Для учителя

Попросите две группы назвать, что можно сочетать. Для задачи с длинной генерацией подходящий ответ — работа со сжатым представлением, если она уменьшает объём обрабатываемых данных. Не обещайте универсальной скорости или качества.

07. Этап 5. Итоги урока 5 мин

ГОВОРИМ УЧЕНИКАМ

Оценивайте модель по тому, что нужно в вашей задаче. Одна удачная картинка ещё ничего не говорит о скорости или разнообразии всей серии.

Главное за урок

- VAE строит варианты через код, а GAN использует состязание сетей при обучении.
- Авторегрессия дополняет последовательность, диффузия уточняет изображение через удаление шума.
- Результаты сравнивают по условиям задачи и нескольким пробам.

Итоговые вопросы

1. Что проверите у генератора, если по десяти попыткам он выдаёт почти один значок?
2. Чем шаг авторегрессии отличается от шага диффузии?

Для учителя

Проверяем разнообразие подборки и одинаковые условия попыток. Авторегрессия выбирает очередной элемент, а диффузия уменьшает шум и может уточнять всю сцену на каждом шаге.

Источники

1. Kingma, Welling. Auto-Encoding Variational Bayes · 20.12.2013 · Материал старше 2024 года. «a reparameterization of the variational lower bound» Источник
2. Goodfellow и соавторы. Generative Adversarial Networks · 10.06.2014 · Материал старше 2024 года. «simultaneously training two models» Источник
3. OpenAI. Image GPT · 17.06.2020 · Материал старше 2024 года. «trained on pixel sequences» Источник
4. Ho, Jain, Abbeel. Denoising Diffusion Probabilistic Models · 19.06.2020 · Материал старше 2024 года. «high quality image synthesis results using diffusion probabilistic models» Источник
5. Rombach и соавторы. High-Resolution Image Synthesis with Latent Diffusion Models · 20.12.2021 · Материал старше 2024 года. «inpainting, and semantic synthesis» Источник