

Архитектуры:

1. Диффузионные модели:

Stable Diffusion, Imagen, DreamFusion: В основе этих моделей лежит принцип скрытой диффузии. Модель сначала "портит" обучающие изображения, добавляя гауссовский шум, а затем учится восстанавливать исходное изображение, убирая шум. Во время генерации модель применяет этот же процесс "очистки" к случайному шуму, постепенно превращая его в осмысленное изображение, соответствующее текстовому описанию.

Плюсы: Высокое качество и разнообразие генерируемых изображений.

Минусы: Требуют значительных вычислительных ресурсов и времени на обучение и генерацию.

2. Трансформеры:

DALL-E 2, Muse: Эти модели основаны на архитектуре трансформеров, они учатся связывать между собой текстовое описание и части изображения.

Плюсы: Эффективное обучение на больших объемах данных, способность понимать сложные взаимосвязи между текстом и изображением.

Минусы: Могут быть менее эффективны, чем диффузионные модели, в задачах, требующих высокого уровня фотореализма.

3. Комбинированные архитектуры:

Midjourney: Использует комбинацию методов, включая диффузионные модели и генеративно-состязательные сети (GAN). Эта комбинация позволяет создавать изображения с высоким уровнем детализации и художественной выразительности.

DreamBooth: Сочетает диффузионные модели с техникой обучения с малым числом примеров (few-shot learning). Это позволяет модели генерировать новые изображения конкретного объекта, используя всего несколько его фотографий.

4. Другие архитектуры:

Imagen: Google не раскрывает точную архитектуру Imagen, но известно, что она отличается от DALL-E 2.

Предположительно, она также основана на диффузионных моделях, но с рядом уникальных модификаций.