

ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ СТВОРЕННЯ ТА ОБРОБКИ ВІДЕОКОНТЕНТУ

Ішу А.О.

Науковий керівник – Солодов В.Д.

Харківський національний університет радіоелектроніки,
каф. МІРЕС, м. Харків, Україна
e-mail: anastasiia.ishu@nure.ua

The article examines key aspects of the use of neural networks in the field of video processing. The work discusses neural networks and their main principles of operation. Particular attention is paid to image generators based on neural networks, diffusion models, highlighting their effectiveness in creating and optimizing video content. The features of stable diffusion are explored and modern services based on this technology are presented. This work is aimed at unlocking the potential of neural networks in video production and indicating ways for their optimal use.

У світі, який швидко розвивається, відео та зображення стали одними з основних джерел, з яких люди сприймають інформацію. І ось, із новим стрибком у розвитку нейронних мереж та штучного інтелекту, перед нами постають нові перспективи у відеопродакшені – створювати відео з використанням цих спеціалізованих інструментів. Досягнення в області генерування зображень на основі штучного інтелекту взяли світ штурмом. І це нові реалії, до яких нам поступово потрібно звикати.

Нині дана тема є дуже актуальною, адже використання нейромереж пришвидшують процес роботи, додають широкі можливості для створення різних концептів відео, що просуває розвиток медіаіндустрії на нові рівні.

Нейронні мережі (НМ) – це математичні моделі, які створені на основі функціональних можливостей нейронної мережі людського мозку. Штучні НМ складаються із базових одиниць – штучних нейронів (вузлів). Нейрон являє собою деяку нелінійну функцію, на яку впливають значення вхідних сигналів – ваг, та активатор – порогове значення, при якому активується вузол. Структурна одиниця отримує інформацію у вигляді електричних імпульсів, обробляє її та передає далі при перевищенні порогу, і таким чином взаємодіє з іншими. Вузли з'єднані між собою в шари. Вони поділяються на вхідні, що приймають данні; приховані, що опрацьовують інформацію; вихідні, що видають результат роботи. Якщо прихований шар один, мережа вважається неглибокою. І навпаки, якщо більше одного – глибокою [1].

Наразі налічується близько 30 різновидів НМ. Кожен із них має свої особливості та застосовується для вирішення конкретних завдань. Вибір архітектури НМ залежить від типу даних та необхідних результатів. Існує безліч архітектур нейромереж, спеціально призначених для обробки відео.

Деякі з них включають глибокі згорткові, рекурентні, генеративно-змагальні нейронні мережі та їх комбінації, призначені для вирішення завдань, таких як визначення об'єктів, відстеження об'єктів та генерація контенту.

При роботі з відео штучний інтелект (ШІ) може застосовуватися для виконання таких завдань: створення нових зображень з нуля (за текстом, картинкою, ескізом); підвищення якості (наприклад, до 4K); редагування (додавання субтитрів, покращення якості звуку) тощо. У сфері ШІ для створення зображень існує безліч унікальних інструментів, які мають індивідуальні особливості, можливості та сферу застосування. Генератори зображень (генеративні моделі) – це штучні НМ, що застосовуються для створення нових зображень, на основі вхідних параметрів та даних (текстових, візуальних). Вони продукують оригінальні та реалістичні картинки, поєднують стилі, концепції та атрибути у новому образі. Серед них можна виділити наступні найпопулярніші:

1. Нейронна передача стилю (англ. Neural Style Transfer, NST) – це оптимізаційна техніка, яка поєднує два зображення разом: вхідне зображення та еталонне зображення стилю. Вихідне зображення виглядає як вхідне, але згенероване у заданому стилі.

2. Генеративно-змагальні мережі (англ. Generative Adversarial Networks, GAN) – нейромережі, що складаються з двох нейронних мереж – генеративної та дискримінаційної, які між собою конкурують для створення реалістичних зображень. Коли дискримінатор отримує згенеровані генератором дані (підроблені), порівнює із реальними зображеннями на яких вчилася. Таким чином, якщо дискримінатор класифікує згенероване зображення як реальне, він оновлюється. І навпаки, якщо як підроблене, генератор оновлюється. Оновлення проводиться для збільшення ефективності та покращення якості генерації в наступних циклах.

3. Дифузійні моделі – це клас генеративних моделей, що генерують зображення, імітуючи поступову трансформацію шуму в структуровані візуальні образи за допомогою процесу, що нагадує дифузію частинок.

Основний принцип генеративних дифузійних моделей заснований на аналогічному дифузії процесі. Модель використовує ланцюги Маркова для покрокового додавання гаусового шуму до початкових даних. На кожному кроці дифузії до зображення додається певний рівень шуму.

Модель навчається пристосовуватися до цього шуму та відновлювати початковий стан, щоб отримати потрібні дані із шуму. Цей процес дозволяє створювати нові варіації даних на основі вхідних шаблонів. Стабільна дифузія (СД, англ. Stable Diffusion, SD) – це популярний підхід створення зображень, де модель використовує процес прихованої дифузії. Вона створена у 2022 році компанією Stability AI у співпраці з дослідниками, вченими та некомерційними організаціями.

Модель прихованої дифузії працює використовуючи багаторазове зменшення шуму в прихованому просторі. Цей простір слугує ефективною

альтернативою піксельному простору, що в свою чергу забезпечує кращі властивості масштабування та співвідношення просторової розмірності [2].

Стабільна дифузія включає в себе різні нейронні мережі: 1. Варіаційний автокодувальник (англ. variational autoencoder, VAE) – це модель, що кодує та декодує зображення в латентний простір, де інформація представляється не в повному розмірі, а в низькій розмірності для подальшого процесу прямої дифузії. 2. U-Net – модель, яка відновлює зображення із випадкового шуму за допомогою глибокої згорткової мережі. Дана модель враховує прихований шум та намагається спрогнозувати скільки шуму додано на кожному кроці. 3. Текстовий кодувальник – це модель, яка навчається витягувати семантичні ознаки з текстових описів. Він перетворює заданий текст в багатовимірні вектори з семантичними зв'язками, що зрозумілі НМ.

Загалом, СД відкриває широкі можливості для створення вражаючих та деталізованих зображень. За допомогою цієї технології можна не лише змінювати кольори, форму та інші атрибути зображення через текстові описи, але й генерувати різноманітні стилі: малюнок маслом, портрет, анімацію та інші.

Можливість комбінування та швидка ефективна генерація, разом з безкоштовним доступом та гнучкими налаштуваннями, роблять стабільну дифузію потужним інструментом для творчого використання.

Незважаючи на свої переваги, вона має свої мінуси. Для роботи потрібні значні обчислювальні ресурси. Також потрібен час, щоб навчитися правильно задавати параметри та формувати запити. Генерація зображень може виявитися неідеальною, особливо при неясних або складних вказівках, і супроводжуватися артефактами та деформованими деталями.

Найвідоміші сервіси для генерації зображень: 1. Веб-інтерфейси стабільної дифузії: Automatic1111, ComfyUI, Fooocus. 2. Веб-сервіси, що базуються на СД: Dall-E 2, Dall-E 3, Midjourney.

Завдяки дифузійним моделям, генерація зображень та відео продовжує завойовувати серця користувачів і набувати неабиякої популярності з кожним днем. Неперервне вдосконалення функціоналу, поява нових можливостей та моделей, підвищення якості вихідних зображень створюють стійку основу для широкого застосування цієї технології у сучасному світі.

Список використаних джерел: 1. Hlaiboroda M. Що таке нейронні мережі та де їх використовують? INCRYPTED. URL: <https://incrypted.com/ua/shcho-take-nejromerezhi/> (дата звернення: 28.02.2024). 2. Akruti A. An Introduction to Diffusion Models for Machine Learning. The Complete Data Engine for AI Model Development | Encord. URL: <https://encord.com/blog/diffusion-models/> (дата звернення: 28.02.2024). 3. Боюка М., Рижкова Є., Козловець С., Бобнів Р. Використання штучного інтелекту в обробці та покращенні якості зображень. Матеріали конференцій МНЛ, (17 березня 2023 р., м. Полтава), 156–158.