Секція будівництва і архітектури

МЕТОД ШВИДКОЇ АРХІТЕКТУРНОЇ ВІЗУАЛІЗАЦІЇ ЗА ДОПОМОГОЮ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

Гетун Г. В.[1], *Іванченко Г. М.*[2], *Соломін А. В.*[3], *Ботвіновська С. І.*[4], *Гетун С. Ю.*[5]
[1,2,4,5]*Київський національний університет будівництва і архітектури*
[3]*НТУ України «Київський політехнічний інститут» ім. І.Сікорського*
E-mail: [1]*galinagetun@ukr.net,* [2]*ivgm61@gmail.com,* [3]*a.solomin@kpi.ua*
[4]*botvinovska.si@knuba.edu.ua,* [5]*sgetun@gmail.com*

1. Постановка проблеми. Архітектурна візуалізація як окремий напрям в роботі архітектора потребує спеціальних компетенцій. За класичного підходу архітектору потрібен 3d-художник, а час, витрачений на отримання перших фотореалістичних візуалізацій є суттєвим. Тому викликом для отримання фотореалістичних зображень архітектурних об'єктів є створення простого, швидкого і автономного методу, доступного ще на початкових етапах проєктування.

2. Можливості. Запропонований підхід використовує низку методів створення архітектурної візуалізації за текстовим описом і опціонально допоміжними зображеннями. Наприклад, на рис. 1 наведено результати одного з таких методів, що генерує зображення за ескізом (див. рис. 1, *а*) і текстовим описом. Інші методи, дозволяють змінювати стиль вихідного зображення або його частини, імплементувати (вмальовувати) в частину зображення бажані елементи, зберігаючи консистентність зображення, збільшувати розмір зображення, додаючи деталізацію, та інші результати, що є альтернативою ресурсозатратним з точки зору класичних методів роботи із зображеннями і 3d-графікою процесам. Слід зазначити, що типовий час генерації для цих методів вимірюється секундами.

3. Реалізація. В основі методу лежать моделі глибокого навчання для перетворення тексту в зображення, а саме моделі прихованої дифузії. У машинному навчанні дифузійні моделі, також відомі як ймовірнісні моделі дифузії, є класом моделей прихованих змінних. Це ланцюги Маркова, навчені за допомогою варіаційного висновку.

Перші моделі були навчені на основі загальнодоступного набору з п'яти мільярдів пар зображення та тексту, в подальшому використовувались й інші набори розмічених даних.

Рис. 1. Метод створення архітектурної візуалізації за текстовим описом і нарисом (*а*). Текстовий опис до генерації (*б*): «modern building, wooden panels, oaks, solar panels, neon design, daylight atmosphere»

Наразі для роботи з дифузійними моделями існує декілька інтерфейсів, найбільш розповсюдженими з яких є *Stable Diffusion WebUI*, *ComfyUI*, *Fooocus*, *VoltaML*, *SwarmUI*, *InvokeAI*. Надалі розглянуто реалізацію за допомогою середовища *ComfyUI*, перевагу якому автори надали за його потужність і гнучкість.

4. Приклади використання

4.1. Розглянемо базовий варіант, коли генерація зображень відбувається за допомогою лише текстового опису. Модель намагається

згенерувати зображення, що разом з текстовим описом сформують пару, яка органічно могла б доповнити датасет, за яким модель навчена. Формат текстового опису є суттєвим. Він має використовувати знайомі навченій моделі патерни (токени). Здебільшого токени – це звичайні слова і терміни англійською мовою. Прикладом текстового опису може бути: «*Modern townhouses in a residential area, new apartment buildings with green outdoor facilities in the city, detailed, high quality*». Цей базовий метод очікувано дає широкий спектр композиційних і стилістичних варіацій. Як з одного текстового опису отримуються різні варіанти (див. примітку 1).

Примітка 1. *Моделі прихованої дифузії генерують зображення в процесі ітераційного наближення в тензорному просторі всіх можливих зображень до локального мінімуму оціночної функції, тобто прийнятнішого зображення. За початкову точку в цьому багатовимірному просторі обирається зазвичай точка, що відповідає деякому довільному зображенню шуму. Якщо згенерувати це довільне зображення шуму за цілочисельною сигнатурою, ми отримаємо відповідність множини цілих чисел результатам процесу генерації, тобто множині зображень, що всі формально відповідають за логікою навчання моделі наведеному текстовому опису.*

4.2. Існує набір інструментів для більш точного контролю композиції. Наприклад, генерація зображень може повторювати композицію наданого креслення, схематичної комп'ютерної візуалізації чи навіть чернетки. Для цього використовуються моделі, навчені на спеціальних датасетах рисунків з ліній, карт глибини та інші. Надані зображення, що контролюють генерацію, перетворюються на референсне зображення типу, на якому навчалась контролююча модель, щоб сконфігурована система генерувала зображення з таким самим референсним зображенням. Приклади генерацій з текстовим описом і наданим контролюючим зображенням, а також відповідні референсні зображення, що автоматично створюються в процесі такої генерації, наведено на рис. 1, *б–і*.

4.3. Для керування стилем зображення, освітленням, оточенням, порою року тощо може використовуватись відповідний розширений текстовий опис, але для цього існують більш продуктивні методи. Один з таких методів використовує моделі адаптації низького рангу, навчені на відносно невеликих наборах стилістично споріднених зображень, наприклад, лише будівель деякого конкретного архітектурного стилю чи зображень лише будівель взимку, тощо.

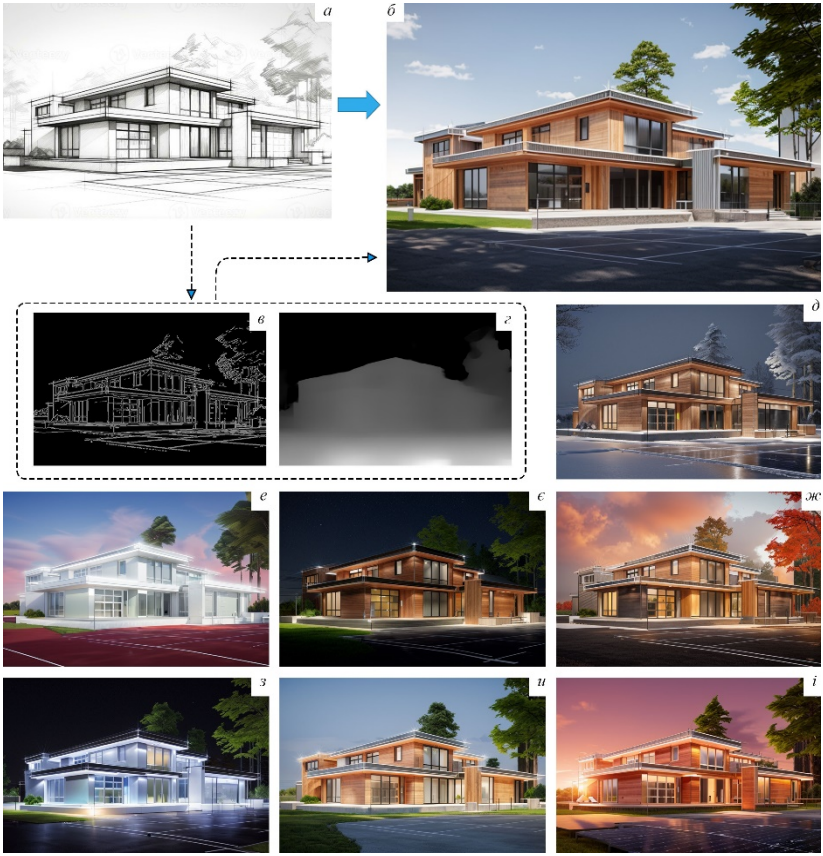

Рис. 2. Метод створення архітектурної візуалізації за текстовим описом, 3d-ескізом (*а*) і прикладом стилю (*б*). Текстовий опис до генерації (*в*): «Beautiful photo of half-timbered townhouse in a residential area, new apartment buildings with green outdoor facilities in the city, detailed, high quality»

Примітка 2. *Сигнатура малюнку лініями, що використовується для контролю кінцевої генерації* – (*г*), *сигнатура карти глибини (світліші точки відповідають ближчому положенню до точки зору, темніші – віддаленішому)* – (*д*). *Ці сигнатури генеруються автоматично за допомогою спеціально навчених моделей. Інші приклади генерації з тим самим текстовим описом, 3d-ескізом і прикладом стилю наведено на* (*е*) – (*і*). *Помітно, що за допомогою моделі зображення-текстового адаптера*

вдається ввести в генерацію елементи стилю фахверк, загальну кольорову гаму, рослинність та інші притаманні референсному зображенню деталі, тоді як компоновка будівлі відповідає 3d-ескізу.

Сигнатура малюнку лініями, що використовується для контролю кінцевої генерації – (*в*), сигнатура карти глибини (світліші точки відповідають ближчому положенню до точки зору, темніші – віддаленішому) – (*г*). Ці сигнатури генеруються автоматично за допомогою спеціально навчених моделей. Генерації з тим же контрольним нарисом та іншими текстовими описами наведено на (*д*)–(*і*). Наприклад, опис до (*д*): «modern building, wooden panels, oaks, solar panels, neon design, winter atmosphere»

Застосування в методі такої моделі адаптації низького рангу дозволяє отримувати генерації в стилі, що є навченим узагальненням відповідного датасету.

Іншим ефективним способом контролю стилю зображення, що генерується, є використання так званих зображення-текстових адаптерів. Це моделі, які навчені для виокремлення із зображення властивостей, що формулюються як допоміжний текстовий опис, і використовуються під час генерації разом з основним текстовим описом. Такий опис може включати розпізнані моделлю токени, що описують стиль, композицію, або, наприклад, кольорову палітру. Застосування зображення-текстових адаптерів є мультимодальними, тобто можуть бути використані одночасно декілька конкретних потрібних властивостей наданих зображень. На рис. 2, *а–і* надані приклади генерацій із застосування одночасно текстового опису, ескізу і контролю стиля.

5. Додаткові можливості. Як видно з вищенаведеного, різні інструменти використовуються окремо і в поєднанні. На рис. 3 наведено приклад конфігурації метода для отримання результатів, наведених на рис. 2. Самі згенеровані зображення також можуть бути використані як допоміжні для наступних генерацій. Це дозволяє конструювати ланцюги генерацій для уточнення та доопрацювання.

Серед інших корисних в практичній роботі інструментів слід відмітити наступні:

5.1. Заміна частини зображення із збереженням локальної або глобальної по усьому зображенню консистентності.

5.2. Збільшення розміру з додаванням необхідної деталізації зображення із збереженням консистентності.

5.5. Уніфікація серії зображень за стилем.

5.3. Генерація відео.

6. Обмеження.

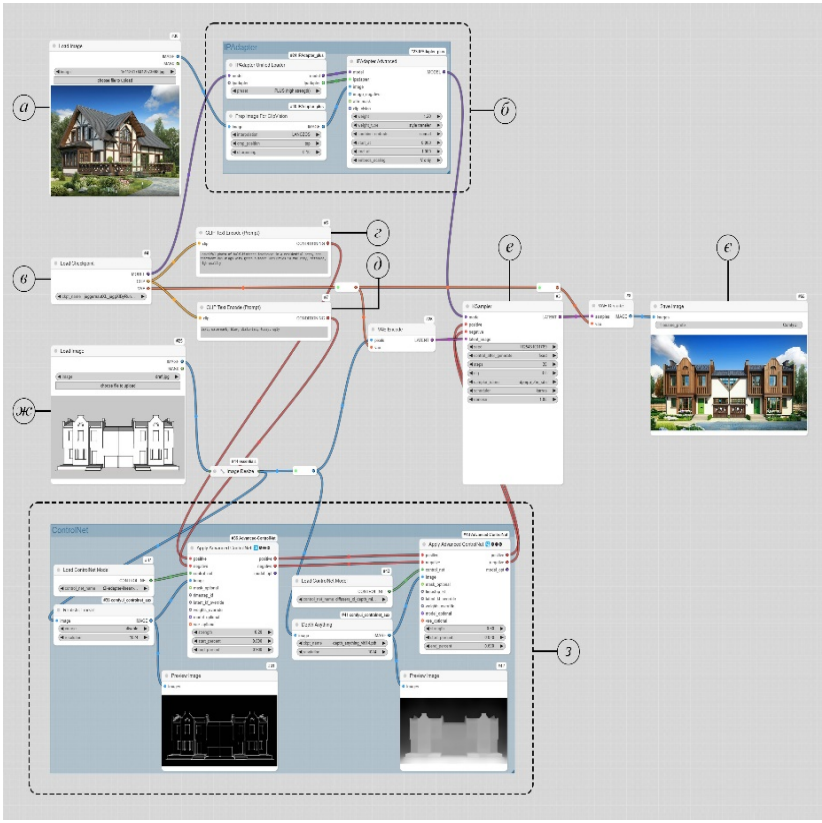

Рис. 3. Конфігурація методу для отримання результатів, наведених на рис. 2

Примітка 3. *В середовищі ComfyUI метод генерації конфігурується у вигляді графа з вузлів, що відповідають функціональним об'єктам, і поєднань між ними. Тут а – завантаження зображення із зразком стилю*; *б – блок, що відповідає зображення-текстовому адаптеру*; *в – завантаження основної моделі*; *г – текстовий опис*; *д – негативний текстовий опис* (*те, чого ми не хочемо бачити на результаті генерації*); *е – вузол, де задаються математично-процедурні параметри процесу генерації*; *ж – завантаження зображення, що контролюватиме компоновку генерації*; *з – блок, що відповідає за контроль компоновки, в даному випадку він складається з контролю за допомогою відповідності малюнку з ліній і карти глибини*; *є – результат генерації.*

Наразі на відміну від класичного 3d-підходу, за допомогою розглянутого сімейства методів важко добитись стабільної генерації серій зображень одного об'єкту (наприклад серії зображень будівлі з різних ракурсів так, щоб всі деталі на різних зображеннях співпадали, тобто щоб серія була консистентною). Для багатьох конкретних випадків таку задачу вдається вирішити, проте авторам не відомий відповідний метод, що можна було б вважати універсальним.

Висновок. Методи, засновані на машинному навчанні, хоча і не замінюють весь спектр інструментів архітектурної візуалізації, є вагомим додатком, а подекуди повноцінною альтернативою окремим його інструментам. Перевагами таких методів є доступність, автономність, швидкість, простота і продуктивність. До недоліків наразі слід віднести обмежену точність керування результатом і недостатню консистентність серій візуалізацій. Їх використання може бути рекомендованим архітекторам і всім спеціалістам, залученим до створення будівель та споруд, на стадії концепт-розробки, пошуку форм і просторових рішень та в інших процесах, що потребують швидкої та ресурсонезатратної візуалізації. Наведені методи можуть бути корисними як додаток до інших методів архітектурної візуалізації.

Література

1. *Song, Y., Sohl-Dickstein, J. N., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B.* Score-based generative modeling through stochastic differential equations. ArXiv, abs/2011.13456, 2020. URL https://api./semant icscholar.org/CorpusID:227209335.

2. *Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B.* High-resolution image synthesis with latent diffusion models. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022. DOI: 10.1109/cvpr52688.2022.01042. URL http://dx./doi.org/10.1109/CVPR52688.2 022.01042.

3. *Xiangyuan Xue, Zeyu Lu, Di Huang, Wanli Ouyang, Lei Bai.* (2024). GenAgent: Build Collaborative AI Systems with Automated Workflow Generation--Case Studies on ComfyUI. ArXiv:2409.01392

4. *Getun G.V., Kolhan A.V.,* On the importance of implementing Revit Autodesk in the educational process for construction students. Стаття н. т. збірник «Future in the results of modern scientific research '2024». Conference proceedings No 34 on August 20, 2024 p. 34–37. DOI: 10.30890/2709-1783.2024-34-00 3

5. https://github.com/comfyanonymous/ComfyUI