



AI 賦能 建築設計：以穩定擴散模型 提升 創意 與 生產力

陳上元^{*}／國立聯合大學建築學系 教授兼系主任、所長

吳建禾／吳建禾建築師事務所 建築師

鄭泰昇／國立成功大學建築系 教授

在大模型時代，應用人工智慧生成內容（Artificial Intelligence Generated Content, AIGC）賦能創作，即調用在雲端的大型資料庫裡預訓練（Pre-Training）的模型、進行推理（Inference），將成為設計產業必備的技術能力。本文旨在探索 AIGC 應用的合理配置，並以擴散模型（Diffusion model）的應用為核心，探索其對建築設計創作與決策所產生的影響。

大模型時代、AIGC 與邊緣運算：協同發展

隨著技術的演進，大模型時代的崛起正深刻地影響著人工智慧的發展，並與 AIGC 及邊緣運算形成緊密的協同關係。這三者並非獨立存在，而是共同構成了一個更高效、更普及的智慧應用生態系。

AIGC 技術的爆炸性成長，例如：ChatGPT 和 Stable Diffusion 等，正是大模型時代最顯著的成果^[1]。從文獻來看，早期的人工智慧模型受限於運算能力與資料量，其生成內容的精細度和多樣性有限。然而，近年來隨著算力大幅提升及大規模數據集的累積，研究者得以訓練出數十億乃至數千億參數的大型語言模型（LLMs）和生成式模型。這使得 AI 能夠更精準地模仿人類的創作與表達，從而實現了從簡單的規則生成到複雜、具原創性的 AIGC 飛躍。因此，大模型可以被視為推動 AIGC 從概念走向大規模應用的核心驅動力。

儘管大模型展現了驚人的能力，但其龐大的體積和運算需求也帶來了挑戰。傳統上，這類模型主要部署

在雲端，需要使用者透過網路連接進行運算。然而，這可能導致延遲、網路頻寬限制及資料隱私問題。這時，邊緣運算的價值便凸顯出來。

Janakiram^[2]等文獻指出，邊緣運算的核心在於將數據處理推向接近資料源頭的網路邊緣，與雲端運算形成互補。從機器學習的角度來看，這種協同模式（cloud-edge collaboration）完美地應用於大模型與 AIGC 的場景中。其運作模式可歸納為：

- 雲端（Cloud）：負責進行大模型的訓練與測試。由於模型訓練需要極高的運算資源和海量數據，因此在雲端進行集中處理是最有效率的方式。
- 邊緣（Edge）：負責部署與推理。一旦模型訓練完成，輕量化或優化後的模型便可部署到智慧手機、物聯網設備或終端裝置等邊緣設備上。這使得使用者能夠在本地球端即時調用模型進行 AIGC，例如：即時生成圖像或文字，而無需每次都與雲端進行大量數據傳輸。

這種雲、邊協同的模式，不僅大幅降低了延遲，提升了使用者體驗，也為資料隱私提供了更好的保障，同時還減輕了雲端伺服器的負擔（圖 1）。

^{*} 通訊作者，syuanchen@nuu.edu.tw

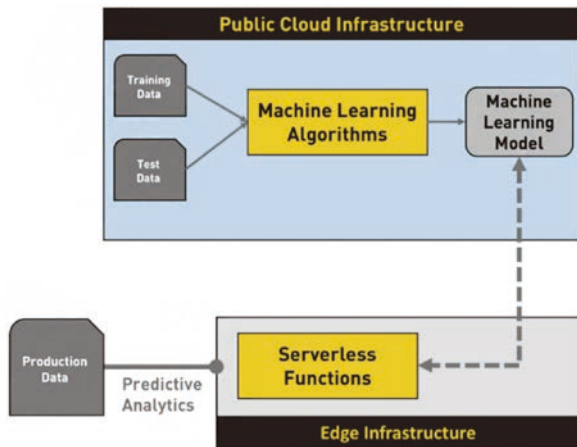


圖 1 在機器學習中，公有雲設施與邊緣運算設施的任務分派^[2]

邊緣推理：擴散模型（Diffusion Models）的核心原理、發展與應用

擴散模型是當前圖像生成領域最主流的模型之一，其獨特的運作機制使其有別於其他深度學習模型。本段將回顧擴散模型的發展歷程、核心原理以及其應用。

擴散模型最早由 Sohl-Dickstein 等人^[3]於 2015 年提出，靈感源於非平衡熱力學中的擴散過程。與生成對抗網絡（GANs）、變分自動編碼器（VAEs）和流模型等其他深度學習模型不同^[4]，擴散模型的訓練過程並非基於對抗學習。其核心原理是透過兩個步驟來學習如何生成圖像^[5]：

- 加噪（Forward Process）：逐步向原始訓練圖像中添加高斯雜訊，直到圖像完全變成隨機雜訊。
- 降噪（Reverse Process）：訓練一個神經網路來學習如何反向地逐步去除雜訊，最終從純雜訊中還原出清晰的圖像。

自此概念提出以來，擴散模型經歷了多次重大技術突破。2021 年，Nvidia^[6]提出基於潛在空間（latent space）的變分自編碼器，大幅提升了模型的訓練效率。同年，Google 引入了基於機率的優化模式，並強

化了模型的變分特性^[7]。隨後，KIM 等人^[8]於 2021 年將對比語言－圖像預訓練（CLIP）技術與擴散模型結合，使得模型能夠根據文字指令生成圖像，為文生圖（text-to-image）應用奠定了基礎。這一切的技術累積，最終在 2022 年迎來了里程碑式的發展「穩定擴散模型（Stable Diffusion）」的發表^[9]，該模型因其高效、高品質的生成能力而迅速普及。

穩定擴散 SD 模型於 2022 年問世，由 StabilityAI、CompVis 與 Runway 聯手開發。該模型基於潛在擴散模型，利用降維運算先將訓練圖像壓縮到低維潛在空間，再進行運算，最後還原成原始尺寸。此技術顯著降低計算資料量與硬體需求，令運算速度大幅提升，突破了標準擴散模型龐大計算量的限制。由此，圖像訓練與生成的實用性與普及度大為增強，為人工智慧在圖像處理領域奠定了堅實基礎^[6,9,10]（圖 2）。

主模型

穩定的擴散 SD 是一個條件生成模型，其主模型^[11]，包括三個主要構件：

對比語言－圖像預訓練文本編碼器（Contrastive Language-Image Pre-Training, CLIP, Text Encoder）：

用於解析提示詞的 CLIP 模型，是 SD 的條件生成器，包括文本編碼和圖像編碼 2 個部分，分別用於提取文本和圖像的特徵。通過搜集大量網路上的圖像和文字資訊再對 CLIP 模型進行訓練，可以實現文本和圖像的對應關係^[8,12]（圖 3）。

潛在擴散模型（Latent Diffusion Model）：

如前述，改良過的擴散模型，以翻轉 C 型的路徑，在潛在空間中進行，其路徑的前向過程加噪擴散（Diffusion），而返向過程則降噪（Denoising）復原。前向（由左到右）和返向過程（由右到左）如圖 4 所示，圖 4 右的條件（conditioning）元件，用來描述模型生成圖像的文本提示^[13]。除此之外，潛在擴散模型圖像生成

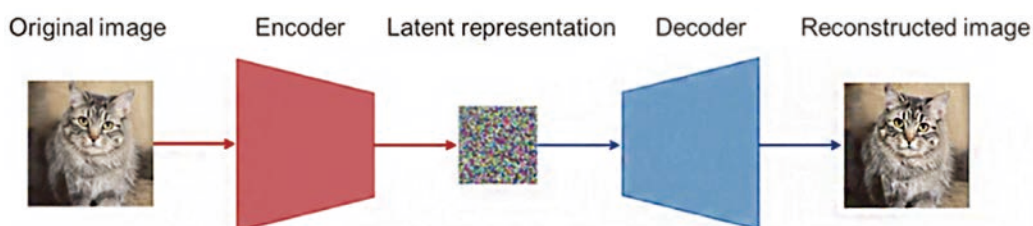


圖 2 潛在擴散模型^[14]

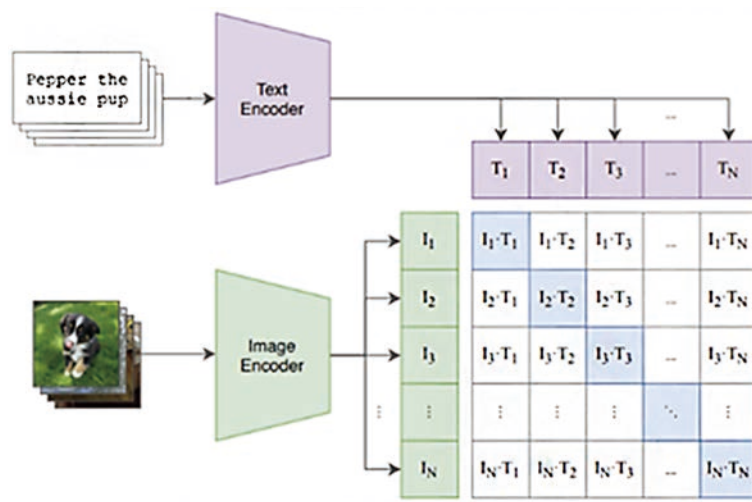


圖 3 CLIP 模型結構包含一個圖片編碼和對應的文字編碼^[15]

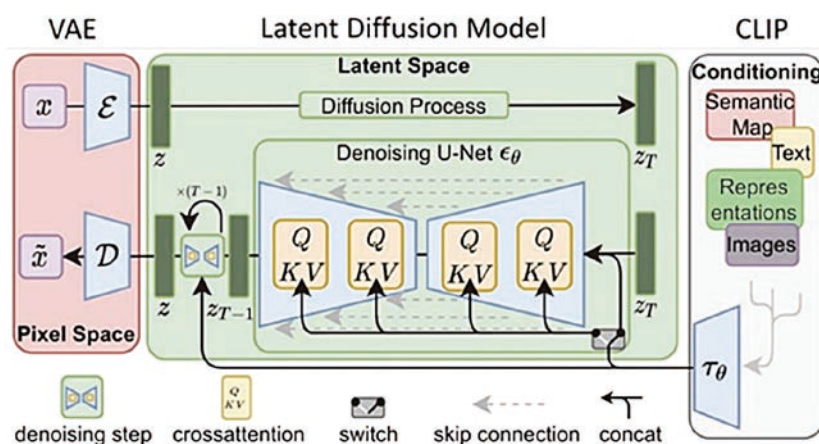


圖 4 穩定擴散模型^[13]

器裡還包含了 (a) U-Net 神經網路以及 (b) 排程與採樣法 (Scheduler and Sampler) 兩個部分：

(a) U-Net 神經網路：

U-Net 是一種卷積神經網路 (CNN)。它的結構特點是有一個「U」形的對稱結構，因此得名 U-Net。這種結構允許網路在多個尺度上學習圖像特徵，同時通過跳躍連接將不同層次的特徵圖結合起來，從而提高分割的準確性^[16]。在 SD 的返向降噪復原的過程中，U-Net 透過先收縮 (contraction)、再舒張 (Expansion) 兩階段過程，將雜訊圖像逐步轉化為清晰的目標圖像。它通過學習從雜訊到圖像的映射，逐步細化圖像的細節，直到生成高品質的圖像^[17] (圖 5)。

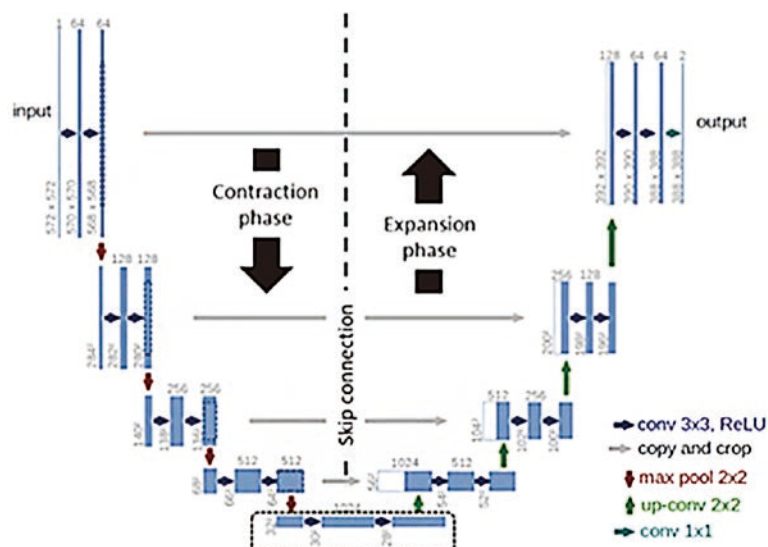


圖 5 U-Net architecture^[17]

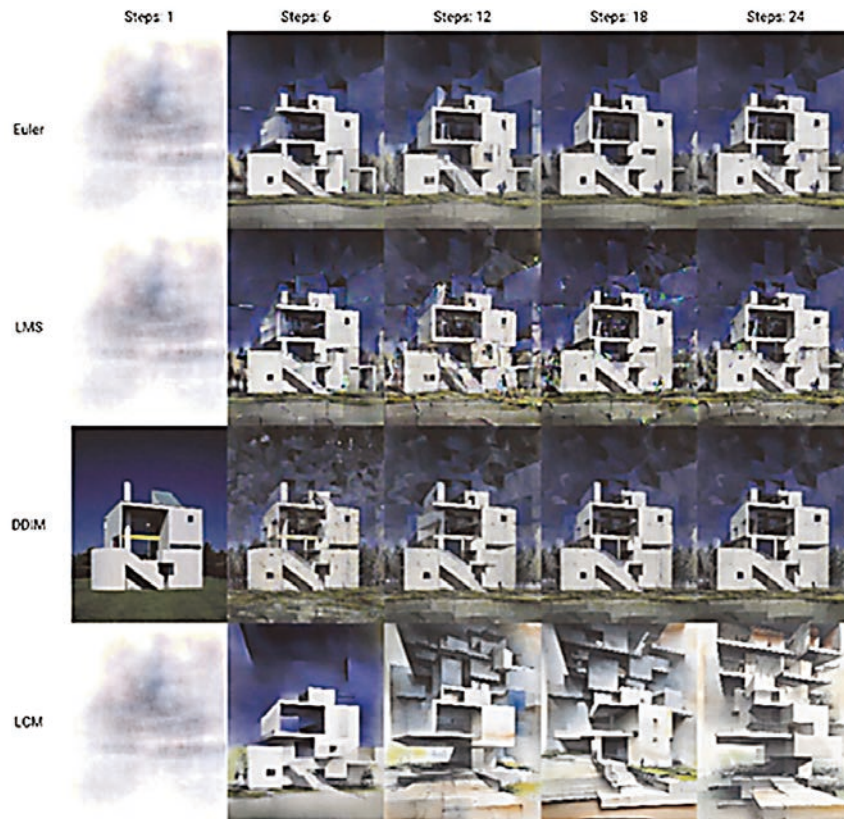


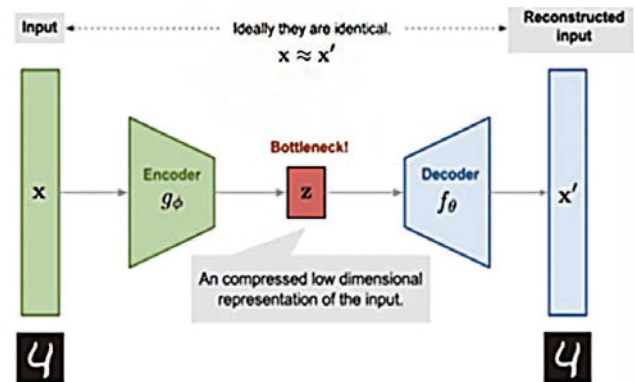
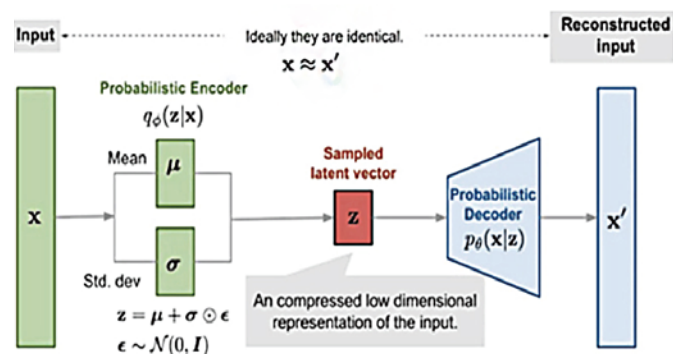
圖 6 不同的採樣法 (Y 軸) 比對不同排程 (X 軸) 的表現

(b) 排程與採樣法 (Scheduler and Sampler) :

在訓練擴散模型過程中，排程 (Scheduler) 是負責調控去噪的程度，決定每一步去除多少噪聲，以確保整個過程高效和精準。然而，採樣法 (sampler)，是負責控制去噪的具體操作的算法。使用者可以根據圖像類型和使用的模型來選擇不同的採樣法 (Sampler)，從而達到更佳の出圖效果。如圖 6 所示，在相同條件下，LCM 較其他採樣法有差異性更大的變化。

變分自編碼器 (Variational Auto Encoder, VAE) :

變分自編碼器 (VAE) 由自編碼器 (AE) 演變而來，結合了 AE 與機率建模生成模型。自編碼器是一種非監督式深度神經網絡，包含編碼器 (Encoder) 和解碼器 (Decoder)，編碼器將圖像資訊降維至潛空間，解碼器則將潛在資料升維還原為原始圖像 (圖 7)。VAE 在編碼器中加入數學運算，將輸入分為平均數和標準差 (經指數函數處理) 兩部分並相加，使其具機率性 (probabilistic) (圖 8)；相較之下，AE 為確定性 (deterministic)。通過潛在變數

圖 7 自編碼器^[19]圖 8 變分自編碼器^[19]

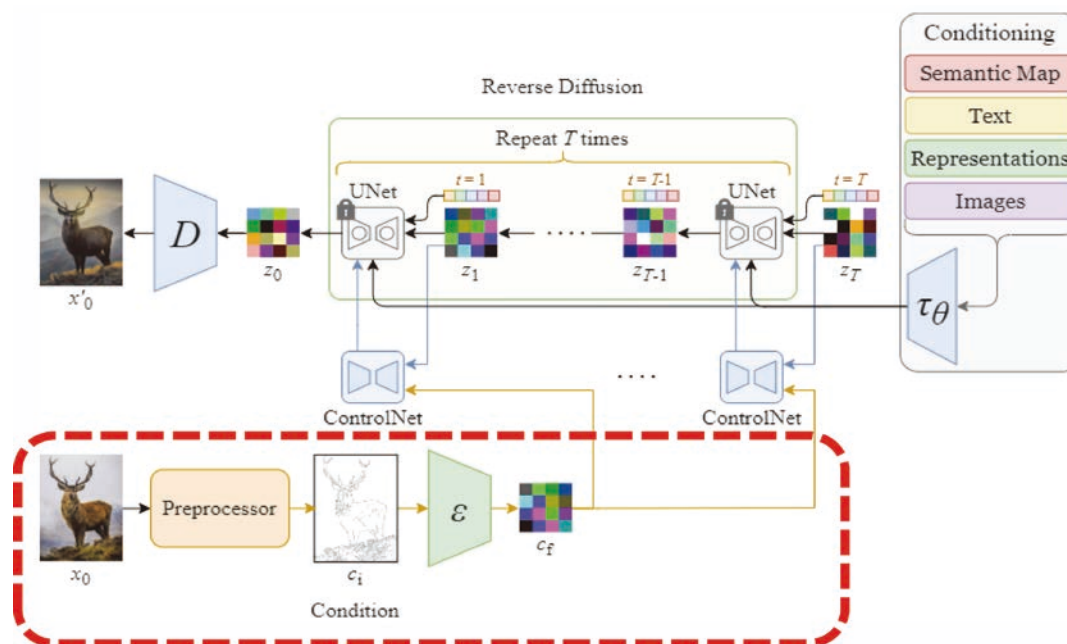


圖 9 控制網路 ControlNet [24]

與隨機性，VAE 能生成多樣化樣本 [18]。

穩定擴散 SD 模型由三個主要構件組成，稱為大模型，亦稱主模型或底模型。SD 必須有主模型才能進行繪圖，檔案格式為「.safetensors」，通常大小超過 2GB。訓練完整主模型成本高昂，非一般設備可負擔。因此，當僅需調整特定風格或主題時，無需重新訓練整個模型，而是透過更簡潔高效的微調模型（Fine-tuned models）來實現特定需求，提升運算效率。

微調模型 (Fine-tuned models)

在電腦視覺和自然語言處理領域，微調模型是指使用預訓練模型，並在特定任務上進行進一步的訓練，以使其適應特定的資料集或問題。通過微調，模型可以適應不同領域的特定資料分佈，從而提高其泛化能力和適應性。在 SD 裡常見的微調模型包括：(1) 文本反轉（Textual Inversion）模型、(2) 變分自動編碼（Variational Auto Encoder, VAE）模型；(3) 低秩調適（Low-Rank Adaptation of Large Language Models, LoRA）模型 [20] 等。這三種微調模型各有特點。文本反轉增強模型對新文本概念的理解，適用於快速適應新詞彙。VAE 強調圖像生成與重建，適合高質量圖像需求。LoRA 透過低秩調適，能在有限資源下高效微調大模型，適用於資源受限但需調整大模型的情境。根據需求選擇適當模型，可在不同應用場景中發揮最佳效果 [21]。

控制網路 (ControlNet)

在 SD 的基礎上，控制網路引入額外的控制模型，

使模型能夠根據特定條件（Condition）輸出圖像 [22]（如上圖 9 所示）。控制網路 ControlNet 模型通常由兩個部分組成：預處理器（Preprocessor）與模型。預處理器的功能是将送入的參考圖進行檢測，轉換成可供模型使用的控制圖像，在控制網路，針對邊緣檢測進行線條約束的類型分別有 Canny、Lineart、SoftEdge、Scribble、MLSD 以及 Segment 等控制網路模型 [23]。

使用者介面

為應用 SD，程式開發者 AUTOMATIC1111 開發了方便操控 SD 的圖形化介面（Graphical User Interface, GUI）命名為 WebUI，是目前流傳最廣、最為友善使用的（圖 10），其它還有強調以節點建構工作流的 ComfyUI（圖 11）等。

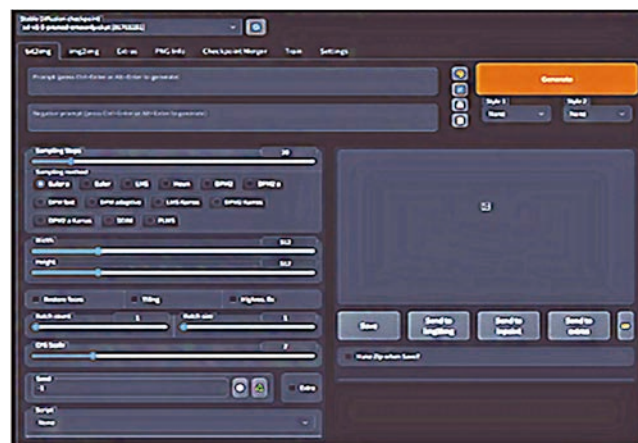


圖 10 WebUI

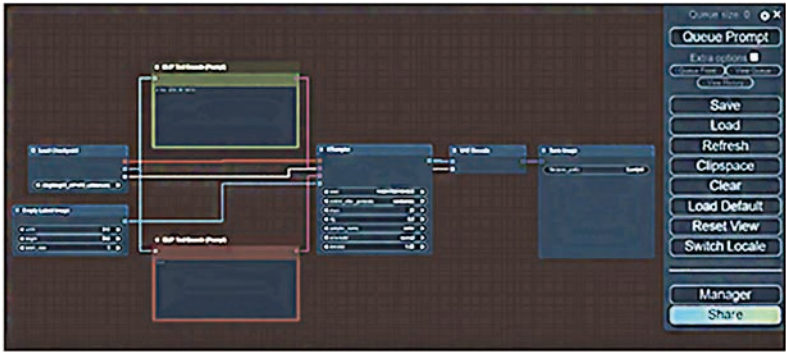


圖 11 ComfyUI

基於 SD：AI 推理的建築設計

SD 具備降維運算的能力為擴散模型的普及化，奠定了實用的基礎；其主模型核心的三大構件：(1) 對比語言 – 圖像預訓練文本編碼器 (CLIP)、(2) 潛在擴散模型 (Latent Diffusion Model)、(3) 變分自編碼器 (VAE)；輔以微調模型與控制網路；以及可選擇的使用者介面 GUI，使得 SD 賦能建築設計上的能力包括：

- 1. 文生圖：理解意念，推理意圖，並發展多樣性。
基於 CLIP 的對比語言 – 圖像能力，以及 VAE 變分自編碼器生成多樣性的設計樣本。
- 2. 圖生圖（圖生文、文生圖）：根據意圖，詮釋意念，展開對話，推衍意圖。

同理基於 CLIP 的對比語言 – 圖像能力，圖生文、文生圖。因此，根據意圖，詮釋意念，展開對話，推衍意圖。

- 3. 建構創意的工作流程。
應用 ComfyUI 建構基於 SD 模型的設計思考流程。有別於使用廣泛的 WebUI，ComfyUI 以基於節點 (node-based) 的設計理念，每個生成圖的步驟都是一個節點，使用者能夠觀察和精確控制 AI 生成圖的工作流程。
- 4. 微調與控制設計意圖與意念。
基於可擴充的微調模型與控制網路，設計得以更精準地控制，修改呈現的風格或者局部修正內容。

- 5. 擴充協作的能力。
SD 包括主模型、微調模型以及控制網路模型。根據實際需求，在使用端、個人終端設備可從雲端擴充下載差異性的預訓練模型，下載方法是通過 Civitai 站 (civitai.com) 和擁抱臉 (huggingface.co) 站來獲取。

- 6. 訓練客製化需求的模型。
SD 提供了可擴充性，使用者可以根據具體需求，選擇合適的微調模型類型、控制網路模型進行訓練。深度學習模型的訓練，在硬體需求與訓練時間有沉重的負荷，使用者可視實際狀況，決定部署於本機或者雲端。

實作演練

接下來本文觀察吳建禾建築師以 AI 推理「前鎮第一公有零售市場新建工程前期規劃案」的過程。以完成專案的設計文件交付 (Design Document) 為任務目標，其歷程包括了：前置設計 (Pre- Design, PD)、初步設計 (Schematic Design, SD)、設計發展 (Design Development, DD)、3D 模型建置 (3D Modeling, 3D)、施工文件 (Construction Documents, CD) 等階段。研究輔以 AI 賦能建築設計推理「設計文件」的檢核矩陣，梳理出穩定擴散在不同服務階段所提供的賦能項目，以及工具支援的不足之處，以提供在下一階段訓練、開發控制工具等客製化需求的參考依據 (表 1)。

表 1 檢核矩陣

	PD	SD	DD	3D	CD
理解意念，推理意圖，並發展多樣性					
根據意圖，詮釋意念，展開對話，推衍意圖					
建構設計的思考流程					
微調與控制設計意圖與意念					
擴充協作的能力					

T

Textual-Inversion

VAE

LoRA

前鎮第一公有零售市場新建工程前期規劃案的階段性發展：

前置設計與資料分析（Pre-Design, PD）

在前置設計與資料分析階段，團隊針對基地條件、功能需求與競圖要求進行了解，並運用 BIM-Revit 建模，建立基本量體模型與周邊環境的數據，進一步結合 AI 生成技術 進行方案推演。

1. BIM 建構初步市場建築量體，整合基地條件、法規與動線，並完成牆、樓板與結構系統的基本規劃與建模（圖 12）。
2. 文生圖快速產出多樣化市場的設計概念，透過導引詞：「Modern / traditional design facade and supermarket building」，導入到 ChatGPT 中，輔助生成具備豐富架構的提示詞 Prompt：

「A supermarket building blending modern and traditional design. The facade features sleek glass, steel, and concrete with warm wood, brick accents, and classic arches. Large windows provide natural light, while a mix of flat and sloped roofs creates balance. The setting includes pedestrian walkways, greenery, and a spacious parking area. Rendered in photorealistic detail with dynamic lighting.」，其產出成果，進而導入在「Rendair pro」中生成不同風格的模擬圖（圖 13）。

3. 圖生文解析 AI 生成圖像，提取風格與設計語意，以優化 BIM 設計方向，並且透過分析出的文字內容，與設計參與者進行討論及細節調整的方向（圖 14）。
4. Revit + AI 簡報設計版面擬定：透過 Revit 圖紙進行排版，分析需求及加入建築設計相關重點說明，針對市場功能需求進行區域大樣 3D 及空間配對需求調整。

初步設計及進階模型建置 （Schematic Design, SD）

初步設計階段主要是將概念轉化為具體的平面布局與立面方案，透過 ChatGPT & Claude & Notebook LM 相互輔助，將規劃設計內容多元輔助資料，並利用 Revit 持續建立進階市場建築模型，搭配運用 AI 生成的圖象，重新回饋調整視覺表現與創意發展，是 BIM + AI 的協同應用：

1. 市場內部空間規劃：基於 AI 輔助進行法規與設計規劃前置資料的統整與設計，並透過 AI 設計推理測試不同排列組合（圖 15）。

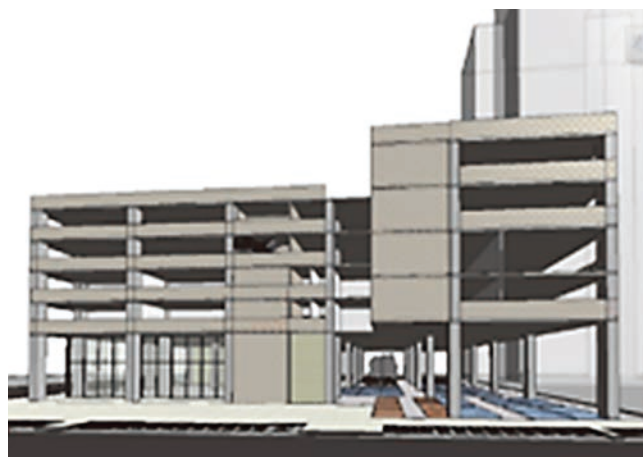


圖 12 BIM 初步建模



圖 13 多元風格模擬圖

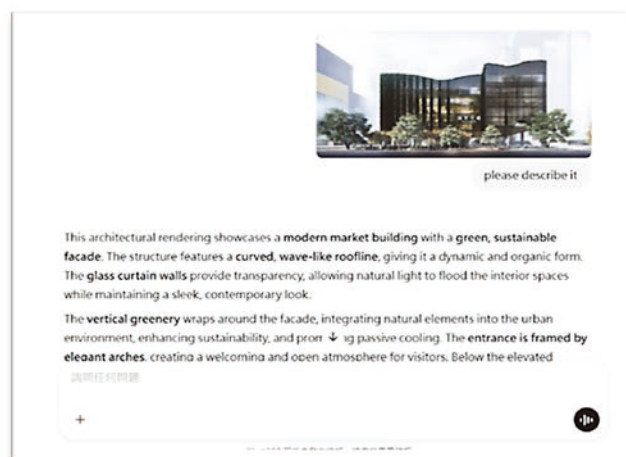


圖 14 圖生文解析，參與者進行溝通討論

2. AI 賦能的建築形式推理：透過 AI 風格轉換（Style Transfer）在經過法規分析與模型整合的進度下，持續生成不同的市場外觀風格。如圖 16 依提示詞 #01 ~ #08 產生的外觀樣貌。

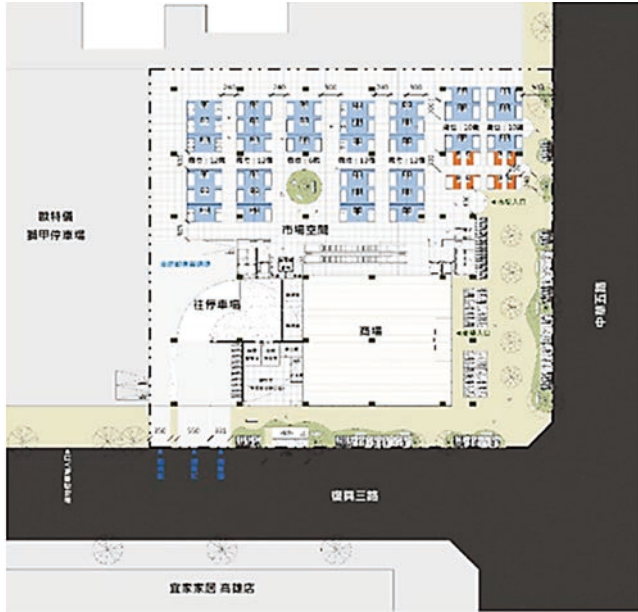


圖 15 BIM-Revit 規劃平面與分析圖說

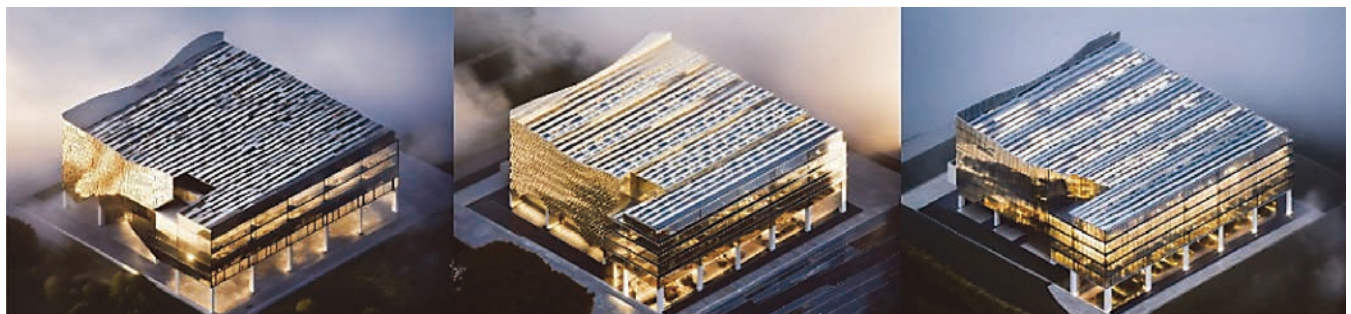
設計發展，多元方案的設計討論 (Design Development, DD)

設計發展進一步深化初步設計方案，並對應預算限制進行調整，確保最終提案兼顧美學與實用性。

1. 材質與光影：依據提示詞進行模型調整，並測試不同表面材質，如金屬構件、木構及混凝土等不同材質，評估其對建築外觀與環境光影的影響（圖 17）。
2. 設計的精緻化：透過 Enscape 進行初步材料表現處理，並交給 Veras AI 進行渲染。

3D 模型建置 (3D Modeling, 3D)

本案一開始就使用 BIM 建構 3D 量體。然而，在 AI 協作，完成多元方案的設計討論後，整合規劃平面，建 BIM 的 3D 模型，以完成後續文件的交付工作。



#01 流動形態－自然律動

#02 編織立面－科技美學

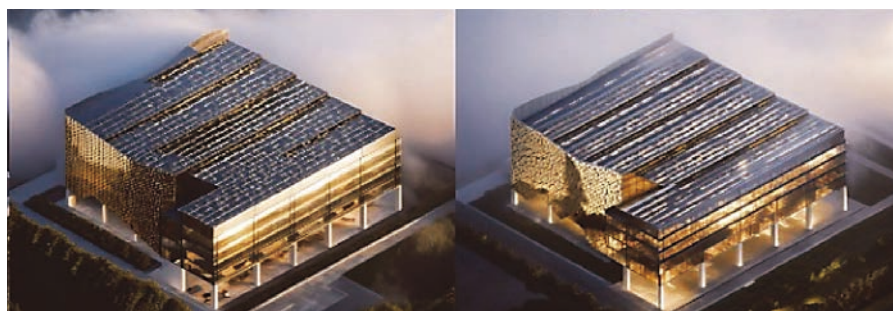
#03 幾何秩序－理性美學



#04 數字肌理－未來象徵

#05 光影層次－動態節奏

#06 生態融合－自然共生



#07 立體光雕－創新結構

#08 晶格紋理－參數設計

圖 16 AI 重新依提示詞產生的外觀樣貌

施工文件（Construction Documents, CD）階段

本階段為求構造與尺寸的準確度，全程以 BIM 建模、出圖。

經過實作觀察驗證，透過設計階段、對照 SD 賦能矩陣檢核，歸納分析如表 2。



圖 17 調整不同表面材質、評估外觀影響

表 2 透過檢核矩陣的檢討與改進

階段	賦能項目	優勢	不足	改進方向
PD	文生圖生成概念方案	快速提供多樣化靈感	方案細節模糊	訓練 Textual Inversion 模型：描述建築風格、類型和專業術語。例如：綠建築、智慧建築、低碳建築等。
SD	圖生圖（圖生文、文生圖）；控制網路等工具的交互應用	迭代優化設計	多元性、差異性不足；構造尺寸的精準度不足	訓練 LoRA 模型：差異性特徵的建築風格（依建築師、建築史、建築機能、地域性等）；搭配 3D 建模軟體，如 Revit、Sketchup 等
DD	採用微調模型，建議材質與採光方案	縮短專案溝通與決策時間	微調模型的適切性不足	訓練 VAE 模型：局部精緻化各式建築元件、細部大樣、建材與質感。
3D	SD 繪圖作為準確建模的參考	加速建模流程	精密度有限	開發客製化的 ControlNet
CD	結合 ControlNet（Segment 模式）生成結構分割圖；文生文功能，整理技術規格。	輔助文件整理	規範性不足	整合建築規範數據

結論

人工智慧的進展，正逐步改寫建築設計的工作方式。以「穩定擴散模型」(Stable Diffusion)為例，設計師僅需透過簡單的文字描述或影像提示，即能快速獲得各種設計草圖與風格模擬。這不僅大幅縮短了前期構想的時間，也讓更多可能性在短時間內被展現，為設計討論帶來前所未有的效率。

在實務操作中，AI 早已能協助不同階段的建築工作：從概念生成、方案比較，到材質與光影的模擬，甚至協助整理施工文件。它的角色並非取代專業，而是成為設計師的「智慧助理」，讓人力得以集中在更需要判斷與創造力的核心問題。

未來，隨著技術的不斷成熟，AI 將像 CAD 或 BIM 一樣，逐漸成為建築專業的日常工具。它的價值在於同時提升創意的廣度與實務的精度，開啟一個設計更靈活、更高效，並能快速回應需求的時代。對建築而言，AI 的出現不僅是一場工具革新，更是一個拓展想像力邊界的契機。

參考文獻

1. Cao, Y., et al. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. arXiv:2303.04226. Retrieved from <https://doi.org/10.48550/arXiv.2303.04226>
2. Janakiram, M.S.V. (2017, April 16). How edge computing and serverless deliver scalable machine learning services. Forbes. Retrieved from <https://www.forbes.com/sites/janakirammsv/2017/04/16/how-edge-computing-and-serverless-deliver-scalable-machine-learning-services/?sh=15625c4a2ebd>
3. Sohl-Dickstein, J., et al. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In Proceedings of the 32nd International Conference on Machine Learning. Retrieved from <https://proceedings.mlr.press/v37/sohl-dickstein15.html>
4. Czxttkl. (2023, November 14). Diffusion models. Retrieved from <https://czxttkl.com/2023/11/14/diffusion-models/>
5. Croitoru, F.-A., Hondru, V., Ionescu, R.T., and Shah, M. (2023). Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(9), 10850-10869. doi:10.1109/TPAMI.2023.3261988
6. Nvidia. (2021). Score-based generative modeling in latent space. arXiv:2106.05931. Retrieved from <https://doi.org/10.48550/arXiv.2106.05931>
7. Google. (2021). Variational diffusion models. arXiv:2107.00630. Retrieved from <https://arxiv.org/pdf/2107.00630>
8. Kim, G., et al. (2021). DiffusionCLIP: Text-guided diffusion models for robust image manipulation. arXiv:2110.02711. Retrieved from <https://doi.org/10.48550/arXiv.2110.02711>
9. Rombach, R., et al. (2022). High-resolution image synthesis with latent diffusion models. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 10674-10685). doi:10.1109/CVPR52688.2022.01042
10. Yang, L. (2024). Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys, 56(4), Article 105, 1-39. doi:10.48550/arXiv.2209.00796
11. Kochanowicz, J., et al. (2023, March 29). Diffusion models in practice. Part 1: A primers. Retrieved from <https://deepsense.ai/diffusion-models-in-practice-part-1-the-tools-of-the-trade/>
12. Gan, Y., et al. (2024). InstructCV: Instruction-tuned text-to-image diffusion models as vision generalists. In ICLR 2024. Retrieved from <https://doi.org/10.48550/arXiv.2310.00390>
13. Ulhaq, A., et al. (2024). Efficient diffusion models for vision: A survey. IEEE Transactions. Retrieved from <https://doi.org/10.48550/arXiv.2210.09292>
14. DZ. (2023, October 8). Variational Autoencoder. Medium. Retrieved from <https://python.plainenglish.io/variational-autoencoder-1eb543f5f055>
15. Openai. (2024, July 23). CLIP. Retrieved from <https://github.com/openai/CLIP>
16. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.
17. Pun, S.N. (2022). Modality specific U-Net variants for biomedical image segmentation: A survey. Artificial Intelligence Review. doi:10.1007/s10462-022-10152-1
18. Kingma, D.P. and Welling, M. (2013). Auto-encoding variational Bayes. CoRR, abs/1312.6114.
19. Allenu2019. (2021, May 7). Math AI – Variational Autoencoder (VAE). ALU – AI, Language, Universe. Retrieved from <https://alu2019.home.blog/2021/05/07/math-ai-variational-autoencoder-vaebian-fen-zi-b/>
20. Hu, E.J., et al. (2021). LoRA: Low-rank adaptation of large language models. arXiv:2106.09685. Retrieved from <https://doi.org/10.48550/arXiv.2106.09685>
21. 杰克艾米立 (2024) · Stable Diffusion · 台北市：旗標。
22. Zhang, L., Rao, A., and Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In IEEE/CVF International Conference on Computer Vision (ICCV). doi:10.1109/ICCV51070.2023.00355
23. Steins. (2023, June 6). Stable Diffusion — ControlNet clearly explained. Medium. Retrieved from <https://medium.com/@steinsfu/stable-diffusion-controlnet-clearly-explained-f86092b62c89>
24. Robot 9, 2024, <https://robot9.me/controlnet-diffusers/> 