

用於文字生成圖像模型的數值控制

本院覽號

32A-1130906

公告日期

2025-01-15

智財權狀態

美國臨時案已申請

摘要

本發明提出了一種基於潛在擴散模型（LDMs）的創新技術，來實現文字生成圖像的精確數值控制。該技術解決了現有文本到圖像生成模型無法有效處理數值參數（如焦距、光圈、ISO感光度、曝光時間等相機設定）的問題，但其應用並不僅限於相機設定。本方法將數值參數嵌入到通過潛在擴散模型（LDMs）的文本特徵空間中，使這些參數成為生成過程中的控制，用戶可以根據具體的數值輸入精確調整圖像生成過程。這種數值化的控制框架不僅提升了生成圖像的準確性與一致性，還可以應用於其他需要數值控制的領域。無論是專業攝影、藝術創作還是其他需要數值精確調整的場景，本技術都提供了一個靈活且強大的解決方案，大幅提升了用戶的控制能力和使用體驗。

技術優勢

- 精確數值控制：本技術可達成精確的數值控制，超越了傳統文字生成圖像模型的能力，使生成結果更為準確和一致。
- 高度靈活性：本技術具有高度靈活性，可應用於各種需要精確數值控制的領域，並可與其他來自社群的LoRA模型結合，擴展其應用範圍。
- 提升使用體驗：相比現有技術，該方法在生成圖像過程中提供了前所未有的數值控制力，允許用戶根據具體需求進行精確數值調整，提升了使用體驗。

應用範圍

- 圖像生成控制：本技術可用於開發精確數值控制圖像生成過程的應用，適合專業攝影師和藝術創作者，提供更多元的創作工具。
- 影像處理與編輯：本技術可以進一步應用於影像後期處理與編輯軟體，允許用戶根據精確的數值參數進行影像調整。
- 數值控制：本技術可擴展至其他領域，如3D建模、視覺特效等，通過精確數值參數來控制生成效果，提升應用的靈活性。

創作人

陳駿丞



中央研究院
ACADEMIA SINICA