⚡️🚀המאמר היומי של מייק 06.08.24: ⚡️🚀

TurboEdit: Text-Based Image Editing Using Few-Step Diffusion Models

חוזרים לסקור מאמרים על מודלי דיפוזיה עם מאמר כחול לבן של קבוצת חוקרים מאוניברסיטת תל אביב. הם מציעים שיטה מעניינת לעריכה מהירה של תמונה. כלומר בהינתן תמונה עם פרומפט נתון c אנו רוצים ליצור תמונה עם פרומפט אחר c1.

כמו שאתם זוכרים מודלי דיפוזיה מגנרטים תמונה על ידי הסרה רעש הדרגתית (denoising). בכל שלב המודל חוזה כמה רעש צריך להסיר מהתמונה והרעש המשוערך הזה מחוסר מהתמונה המורעשת באיטרציה הקודמת. השיטה הפשוטה לעשות עריכה של תמונה היא:

להחסיר מהתמונה(המקורית) באיטרציה t את הרעש הזה המשוערך עם פרומפט c (כמו שעושים כאשר אין עריכה)

להוסיף אל התוצאה את התוחלת המשוערכת של התמונה המורעשת(הערוכה) עם הפרומפט c1 החדש (עם התמונה המורעשת הערוכה.

כלומר בכל איטרציה מתקנים את הסרת הרעש בכיוון הפרומפט החדש.

דרך אגב ניתן שערוך הרעש הנוסף באיטרציה t ושערוך תוחלת התמונה אחרי הסרת הרעש אלו שתי בעיות שקולות, כלומר אחת מהן היא פשוט רפרמטריזציה של השנייה מבחינת השערוך.

הבעיה בשיטה הפשוטה לעריכת תמונות שהיא לא עובדת טוב ויוצרת ארטיפקטים בתמונה הערוכה. המחברים מנצלים מחקר קודם שמצא שהסקייל של הרעש (כלומר ההפרש בין התמונה המורעשת לתוחלתה) לא מתנהג לפי הסקייל של התהליך הקדמי של הדיפוזיה של התמונה המקורית (שבו מוסיפים רעש עם שונות עולה לתמונה עד שזו הופכת לרעש טהור). הרעש שנוצר במהלך עריכה כזו הוא בעל שונות משמעות גדולה יותר מאשר זה של התמונה המקורית.

אז המחברים מציעים להחסיר מהתמונה המורעשת המקורית באיטרציה t את שערוך התוחלת של התמונה המורעשת עבור האיטרציה t+d עבור d חיובי שהם מצאו. כלומר לוקחים תמונה x\_t ומזינים אותה למודל שערוך התוחלת עם מספר איטרציה t+d. בסוף מכוונים את התמונה עם שערוך תוחלת המשוערכת של התמונה הערוכה עם איטרציה t+d.

בנוסף המאמר מציע דרך מעניינת לווסת את ״עוצמת העריכה״ בצורה דומה ל classifier guidance כדי לכוון את התוצאה של מודל דיפוזיה גנרטיבי ללא פרומפט עבור פרומפט נתון. הפעם על ידי ניתוח של נוסחת העריכה המחברים משקול של מרחק cross-prompt (הפרש שערוך התוחלת עבור התמונה הערוכה המורעשת עבור פרומפטים c ו- c1) לבין מרחק cross-trajectory שמודד הפרש בין חיזוי התוחלת בין התמונה הרגילה לתמונה המשוערכת). משקול כזה מאפשר לבצע את העריכה בפחות איטרציות denoising.

מאמר כתוב יפה ובהחלט מומלץ

https://arxiv.org/abs/2408.00735