# Review 106: Interpolating between Images with Diffusion Models, 25.07.23 https://arxiv.org/abs/2307.12560.pdf

**Paper: https://arxiv.org/abs/2307.12560v1**

האם אתם יודעים שניתן להפוך תמונה אחת לתמונה אחרת הצורה רציפה וחלקה באמצעות מודלי דיפוזיה. כלומר לוקחים תמונה של רוק ועל ידי שינויים קטנים והדרגתיים הופכים אותו לתמונה של שרק. היום ב- #shorthebrewpapereviews מדברים על איך עושים זאת באמצעות מודלי דיפוזיה.

קודם כל לוקחים את שתי התמונות ומעבירים אותן למרחב הלטנטי. הגישה הנאיבית (שלא מצריכה מודלי דיפוזיה) היתה לבצע אינטרפולציה לינארית הדרגתית מהייצוג הלטנטי של התמונה הראשונה בכיוון של הייצוג הלטנטי של התמונה השנייה. ואז מעבירים את הייצוג הלטנטי המשוערך לתמונת ביניים עם הדקודר. הגישה הנאיבית הזו לא עובדת כל כך טוב. במקום זאת מוספים רעש לייצוגים, משתמשים באותה אינטרפולציה עבורם ואז משתמשים במודלי דיפוזיה כדי לנקות את הרעש (בסוף מעבירים את התוצאה דרך הדקודר כדי ליצור תמונה).

כלומר כדי ליצור תמונה N/2 מתוך N תמונות ב״שרשרת השינוי״ ממצעים את הייצוגים הלטנטיים של שתי תמונות המקור, כדי ליצור תמונה בשלב 3N/4 ממצעים ייצוגים מורעשים בשלב N ו- N/2. כמה רעש מוסיפים? ככל שהתמונות בשרשרת השינוי רחוקים יותר מוסיפים יותר רעש כי זה מאפשר לבחון ״יותר אפשרויות לאינטרפולציה״ במרחב הלטנטי. האם זה מספיק?

לא תמיד. כאשר יש לנו תיאור טקסטואלי לתמונות האלו ניתן למנף אותו לשיפור איכות ״שרשרת השינוי״. כדי להשיג זאת המאמר גם מכייל את האנקודר עבור כל תמונה במטרה להתאים אותו יותר לתמונה (תוך מזעור של שגיאת השחזור הרעש של מודל דיפוזיה מאומן SD). המאמר גם משתמש ב-pose של התמונות כדי לשפר את ביצועי המודל שלהם. כדי לקבל pose של תמונה הביניים עושים אינטרפולציה של ה-poses של התמונות המקוריות (מוסיפים אותו כהתניה נוספת למודל דיפוזיה המנקה את הרעש).