# Review 139: Unsupervised Compositional Concepts Discovery with Text-to-Image Generative Models, 06.09.2023 https://arxiv.org/abs/2306.05357.pdf

**Paper: https://arxiv.org/abs/2306.05357v2**

מודלי דיפוזיה גנרטיביים מסוגלים ליצור תמונות (ולא רק) באיכות מרהיבה בהתאם לתיאור טקסטואלי. עכשיו השאלה האם ניתן לגרום למודלי דיפוזיה לעשות את הפעולה ההפוכה (סוג של). כלומר להפיק קונספטים ויזואליים (כלב, מכונית וכאלו) מסט תמונות נתון.   
  
היום ב-#shorthebrewpapereviews אנחנו סוקרים מאמר שמציע שיטה אלגנטית אינטואיטיבית לכך. קודם כל המאמר מגדיר מושג ״מודל דיפוזיה מרוכב״ (composite diffusion model או CDM). אתם זוכרים שבמהלך הגנרוט של תמונה אם מודלי דיפוזיה רגילים אנו מתחילים מרעש טהור ואז מסירים מנות קטנות של רעש, המחושב עם מודל מאומן, באיטרציות עד שמגיעים לתמונה נקיה.   
  
המודל שמשערך את הרעש מקבל גם את התיאור של התמונה (ייצוגו הוקטורי). עם CDM אנו יוצרים תמונה עם יותר מקונספט ויזואלי אחד כאשר הרעש המנוקה הוא סכום של שערוכי רעש בהינתן קונספט מינוס הרעש המשוערך ללא התניה בתיאור (unconditioned). זה די דומה (מקרה פרטי) של classifier-free guidance שיטה פופולרית לגנרוט תמונות עם מודלי דיפוזיה.   
  
אבל בואו נחזור לשאלה המקורית: איך ניתן להפיק קונספטים ויזואליים (או ייצוגם) מסט תמונות נתון? נניח שאנו רוצים ״ללמוד״ K קונספטים ויזואליים מסט של תמונות. אז כל תמונה ניתנת לייצוג בתור CDM כאשר במקום סכום של שערוכי הרעש בהינתן כל קונספט בונים את סכומם הקמור (סכום המקדמים שווה ל 1 וכולם אי שליליים) עם המקדמים נלמדים עבור כל תמונה (גם כאן מחסירים מהסכום את שערוך הרעש ה-unconditioned).   
  
אז בהינתן סט תמונות מכיילים מודל דיפוזיה כאשר המטרה ללמוד את ייצוגי הקונספטים הויזואליים וגם המקדמים של כל קונספט בכל תמונה. הלוס מכמת עד כמה טוב הצלחנו לשחזר כל תמונה. לאחר מכן ניתן לגנרט תמונה עבור כל קונספט בנפרד או כל מיני שילובים שלהם.