# Review 101: GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models

**Paper: https://arxiv.org/abs/2112.10741v3**

המלצת קריאה:

מומלץ במום לאוהבי מודלים גנרטיביים ליצירת דאטה ויזואלי מדאטה טקסטואלי. חובה לעוסקים במודלי דיפוזיה.

תחומי מאמר:

מודלים גנרטיביים ליצירת תמונה מתיאורה (כמו CLIP ו- DALL-E)

מודלי דיפוזיה לגנרוט דאטה ויזואלי

-------------------------------------------------------------------------------------------------------------------------------

כלים מתמטיים וידע מוקדם:

הבנה טובה ב- Diffusion Denoising Probabilistic Models(DDPM)

ידע בסיסי ב- Conditional Image Generation

תהליך קדמי ואחורי (backward and forward processes)

-------------------------------------------------------------------------------------------------------------------------------

פרטי המאמר:

מאמר: https://arxiv.org/abs/2112.10741

קוד: https://github.com/openai/glide-text2im

פורסם בתאריך: 08.03.2022

-------------------------------------------------------------------------------------------------------------------------------

מבוא:

בשנתיים האחרונות יצאו מספר עבודות המציעות גישות ליצירת פיסות דאטה ויזואלי (בעיקר תמונות אבל גם קטעי וידאו). המפורסמות ביניהן הן DALL-E ו-CLIP שהצליחו ליצור תמונות באיכות מרהיבה מהתיאור (למשל כיסא בצורה של אבוקדו). מרבית עבודות אלו התבססו על פרדיגמה הלמידה הניגודית (contrastive learning) כדי למדל את המיפוי בין מרחב השפה הטבעית לבין דומיין התמונות. באותה התקופה תחום המודלים הגנרטיביים לא קפא על שמריו, ובנוסף מודלי דיפוזיה הסתברותיים (DDPM) התחילו לכרסם מהשליטה הבלתי מעורערת של Generative Adversarial Networks (גאן) ו-Variational AutoEncoders (VAE) בתחום הזה. למיטב ידיעתנו העבודה הראשונה שבה פותח DDPM שהצליח להכות את ביצועי SOTA (שהושגו עם הגאן כמובן) בגנרוט תמונות היא מאמר זה שנסקר כאן. אחריו יצאו מספר מאמרים שעוד שכללו את ביצועיהם של מודלי דיפוזיה

המאמר הנסקר מציע לרתום את מודל הדיפוזיה, שקיבל את השם GLIDE, ליצירת תמונות מתיאורם. בנוסף ניתן לכייל את המודל שפותח במאמר בשביל לבצע ״השלמת״ תמונות (image inpainting) בהתבסס על התיאור. כלומר אם יש לנו תמונה שמתוארת באמצעות ״צפרדע רוכב על אופניים בשדה חמניות״ ובתמונה עצמה יש לנו צפרדע באמצע שדה חמניות ללא אופניים, אז GLIDE מכויל מסוגל להשלים את התמונה ולהוסיף אופניים שעליו.

שיתוף פעולה: הסקירה נכתבה יחד עם אברהם רביב