# Review 80: [Short] SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis

**Paper: https://arxiv.org/abs/2307.01952v1**

זהו השדרוג של מודל SD2.0 המגנרט תמונות באיכות מטריפה מטקסט. המאמר הציע כמה שפצורים לארכיטקטורה של SD המקורי, שכלל את האימון והשתמש ב-SDEdit לשיפור איכות התמונות

המודל מבוסס על אותם העקרונות כמו SD המקורי אך מוספים refiner באמצע המבוסס על SDEdit (דגימה ממודל דיפוזיה דרך פתרון נומרי של מישדיפ סטוכסטי :).

כלומר תהליך הדגימה נראה כך:

שיבוץ (embedding) הטקסט במרחב הלטנטי עם ה-encoder ודגימה מרעש גאוסי

יצירה של שיבוץ התמונה עם מודלי דיפוזיה לטנטי LDM (כמו ב-SD)

SDEdit לשיפור וקטור לטנטי מסעיף 2

יצירת תמונה מהשיבוץ של באמצעות ה-decoder

כאמור יש שינוים בארכיטקטורה (עדיין מבוססת על UNet ומנגנון attention). הם גם "העשירו״ את תהליך האימון:

הכניסו תמונות בגדלים שונים אך העבירו לרשת את הגודל שלהם

עשו crop לתמונות והזינו לרשת את גודל הקיצוץ.

https://arxiv.org/abs/2307.01952