# Review 136: [Short] Any-Size-Diffusion: Toward Efficient Text-Driven Synthesis for Any-Size HD Images, 03.09.2023 https://arxiv.org/abs/2308.16582.pdf

**Paper: https://arxiv.org/abs/2308.16582v2**

אוקיי, אני חייב להתוודות בפניכם. למרות שעסקתי לא מעט במודלי דיפוזיה גנרטיביים לא ידעתי שהם מתקשים ליצור תמונות בגדלים שונים. אז היום גיליתי את זה דרך המאמר שאני הולך לסקור קצרות ב-#shorthebrewpapereviews.   
  
המאמר מציע גישה די פשוטה לאימון מודלים לגדלים שונים. קודם כל לוקחים דאטהסט עם תמונות (בעלות כותרת או תיאור) מכל מיני גדלים, ועבור כל תמונה מחשבים יחסי של אורך לרוחב (H/W). מגדירים את היחסים בין רוחב לגודל שהמודל מאומן עליהם ומחפשים את היחס הקרוב ביותר לזה של התמונה.   
  
אז עושים לתמונה סקיילינג לפי היחס הזה ומכניסים אותה למודל דיפוזיה מאומן (Stable Diffusion -SD). לאחר מכן מכיילים את המודל הזה (כמו שמאמנים מודל דיפוזיה רגיל) אבל בשיטת LoRA כאשר האנקודר והדקודר לא מתעדכנים (frozen). כלומר מאמנים מטריצת תוספות למטריצות משקלים של המודל המאומן (שנשאר מוקפא) כאשר מטריצת תוספת זו מאומנת בתור מכפלה של שתי מטריצות low-rank (לזירוז הכיול).  
  
איך מגנרטים תמונות לאחר מכן? יש להם שיטה מעניינת (הופתעתי קצת שהיא עובדת) – מחלקים תמונה לפאצ'ים זרים ואז בכל איטרציה של מודל דיפוזיה בוחרים כמה מהם, מזיזים אותם לרוחב ולאורך עם הזזה בגודל אקראי, משחזרים את הרעש הנוסף בפאצ'ים המוזזים ואת השאר מעתיקים מהאיטרציה הקודים. כאמור מפתיע שזה עובד כי זה קצת לא מתיישר עם אופן האימון של מודל דיפוזיה – כנראה מספר האיטרציות הוא די גבוה כאן.