# Review 147: Generative Image Dynamics https://huggingface.co/papers/2309.07906

**Paper: https://arxiv.org/abs/2309.07906v3**

נתחיל מזה שזה מאמר מאוד יפה רק בגלל זה מופיע בו התמרת פוריה (עבדתי בתחום עיבוד אותות כמה שנים ויש לי זכרונות חמים ממנו). שנית, המאמר עוסק במודלי דיפוזיה וביצירת וידאו עם תמונה שאני מאוד אוהב. אוקיי.  
  
אז היום ב-#shorthebrewpapereviews סוקרים מאמר ש״מוסיף״ תנועה לאובייקט בתמונה, כלומר מגנרט וידאו קצר המראה אובייקט בדינמיקה( למשל פרח ברוח). יש כמה גישות למידול של תנועה בוידאו למשל שדה תנועה שמסמן תזוזה של כל פיקסל בתמונה. כלומר וידאו המורכב מ-T פריימים ניתן לייצוג על ידי T x H x W מספרים לתמונה בגודל HxW.   
  
כלומר יש לנו HxW מערכים באורך T מספרים המתארים את המסלול של כל פיקסל בתמונה. אז הנה באה הקטע של התמרת פוריה שאני כה אוהב. במקום לחזות את המערך הזה בואו נעשה זאת עם התמרת הפוריה שלו. כמובן שבמקרה שלנו ההתמרה דו מימדית כי כל פיקסל יכול לנוע בשני הכיוונים (למעשה 4 מספרים ממשים פר פיקסל כי התמרת פוריה היא מרוכבת).   
  
אוקיי, אז מאמנים מודל דיפוזיה שמטרתו לגנרט התמרת פוריה של תנועתו של כל פיקסל. הבעיה אבל שקצת קשה לאמן רשת שחוזה וקטור ארוך (של תדרים) לכל פיקסל בתמונה (לטענת המאמר). אז הם שמים לב שתנועה של כל פיקסל מתוארת בעיקר על ידי תדרים נמוכים (תנועה איטית) כאשר התדרים הגבוהים מקבלים ערכים נמוכים. אז המחברים מחליטים לקחת רק 16 התדרים הנמוכים ביותר לגנרוט (יש איזה נרמול של התדרים כדי למנוע ערכים גבוהים מדי של תדרים מסוימים).   
  
טוב, איך מאמנים מודל דיפוזיה. לוקחים תמונה בדינמיקה (סרטון קצר), מחשבים את שדה התמונה, מעבירים את זה דרך התמרת פוריה וזה מהווה הקלט למודל דיפוזיה שאותו מרעישים ואז מסירים את הרעש לאט לאט עם המודל. בנוסף מכניסים למודל את התמונה וגם תיאור טקסטואלי (אופציאונלי). ב-inference מעבירים את התמרת פוריה של שדה התנועה דרך התמרת פוריה הפוכה ומרנדרים את הוידאו. והדבר האחרון כמובן משתמשים במודל דיפוזיה לטנטי….