# Review 171: Matryoshka Diffusion Models

**Paper: https://arxiv.org/abs/2310.15111v2**

https://huggingface.co/papers/2310.15111

סקירה של היום היא על שילוב של מילה ברוסית (מטריושקה או בבושקה בעברית) ומודלי דיפוזיה. משתמע מכך נדבר על הרבה מודלי דיפוזיה אחת בתוך השנייה כמו שמקובל במטריושקה. אז היום ב-#shorthebrewpapereviews סוקרים מאמר שלקח רעיון של ProGAN והטיל אותו על מודלי דיפוזיה.

לצעירים בינינו ֿ ProGAN הוא גישה, מבוססת על GANs (שיטה גנרטיבית שולטת לפני מודלי הדיפוזיה) שמתחילה יצירת תמונה בעלת רזולוציה גבוהה מיצירת תמונה מרזולוציה נמוכה מאוד. לאחר מכן היא יוצרת ממנה תמונה ברזולוציה גבוהה יותר (נגיד פי 2) בכל שלב עד שמגיעים לתמונה ברזולוציה הנדרשת. אז איך בעצם מטילים את הרעיון הנחמד הזה על מודלי דיפוזיה?

כמו שאתם זוכרים מודלי דיפוזיה יוצרים תמונות מרעש טהור כאשר בכל שלב (איטרציה) מורידים קצת רעש מהתמונה עד שמגיעים לתמונה הנקייה. אז בשיטה המוצעת מציעים לבצע את התהליך הזה על תמונה מרזולוציות שונות בו זמנית. כלומר כאשר אנו מורידים רעש באיטרציה t (כדי לקבל תמונה מאיטרציה t-1) ברזולוציה מסוימת R אנחנו משתמשים לא רק בתמונה מאיטרציה t של התמונה מרזולוציה R אלא בתמונות מכל הרזולוציות האפשריות.

קודם כל זה משפר את איכות השערוך כי למודל יש מידע נוסף לגבי התמונה. בנוסף המאמר מציע לאמן את מודל בצורה פרוגרסיבית (בקטע טוב כאן). זאת אומרת מתחילים לאמן מודל דיפוזיה החל מרזולוציה נמוכה ואז ממשיכים לרזולוציות גבוהות יותר תוך כדי ניצול מודלי דיפוזיה מאומנים מרזולוציות נמוכות (לא לגמרי ברור האם המודלים מרזולוציות נמוכות מאומנות תוך כדי אימון של רזולוציות גבוהות).

התוצאות די מרשימות אבל לא ראיתי התייחסות לזמן יצירה גבוה יותר (או יותר משאבי חישוב) ממודל דיפוזיה סטנדרטי. הסיבה לכך נעוצה בעובדה כי בשביל לגנרט תמונה מרזולוציה גבוהה צריך כל פעם ליצור תמונות מרזולוציה נמוכה בכל איטרציה. אבל עדיין רעיון נחמד מאוד.