⚡️🚀המאמר היומי של מייק -22.10.24: ⚡️🚀  
Sample what you can’t compress

לא היה לי הרבה זמן להקדיש לסקירה אז בחרתי במאמר הזה שניתן לסקור אותו די בלקוניות בלי לפגוע בחוויית הקוראים. המאמר מציע שיטה נחמדה לבניית ייצוג דאטה ויזואלי (קרי תמונות) באמצעות שכלול של אוטו-אנקודר. מכיוון שהייצוג הזה בד״כ במימד נמוך יותר מהדאטה עצמו אז ניתן להתייחס אליו בתור דחיסה של דאטה. ד״א ניתן לאמן ייצוגים שלאו דווקא ״מעבירים״ את הדאטה למרחב בעל מימד נמוך יותר ב-denoising AE ולפעמים ב-sparse AE.

אוטו-אנקודר זו דרך לבנות ייצוג מקומפרס של דאטה עם השילוב של האנקודר והדקודר כאשר האנקודר ממפה את הדאטה למרחב הייצוג והקודר משחזר את הדאטה המקורי מייצוגו הדחוס. מאמנים AE דרך מזעור של לוס השחזור (עד כמה טוב הצלחנו לשחזר את הדאטה מייצוגו הלטנטי) ולפעמים מוסיפים רגולריזציה במטרה לגרום לייצוג להיות בעל תכונות מסוימות (כגון דליל).

כמובן שלא תמיד מצליחים להגיע לייצוג חזק (ששומר את כל התכונות האינהרנטיות של פיסת דאטה) עם AE והמחברים מציעים לשכלל אותו על ידי הוספתו של מודל הדיפוזיה לסיפור. כזכור (או שלא ואז אני אזכיר) מודל דיפוזיה מאומן להסיר רעש מפיסת דאטה ואם מאמנים אותו טוב אז מקבלים מודל שיודע לגנרט דאטה מרעש טהור(על ידי הסרת רעש הדרגתית).

המחברים מציע לקחת את מודל הדיפוזיה (המחברים משתמשים במודל דיפוזיה המקורי שבונה את התמונה עצמה בתהליך דיפוזיה ולא ייצוגה הלטנטי). המודל הזה מורכב מסדרת של U-Nets (ולא טרנספורמרים כמו שאנו רואים היום במודלי דיפוזיה) שקודם מקטינים את מימד התמונה (כלומר ניתן לראות את זה כאוטו-אנקודר) ולאחר מכן בונים מהייצוג הזה את התמונה.

המחברים מזינים את התמונה המשוחזרת אחרי הדקודר של AE יחד עם התמונה המורעשת(המקורית) למודל דיפוזיה שמאומן כאמור להסיר רעש מהדאטה (יחד עם AE). הלוס מורכב מסכום משוקלל של הלוס הרגיל של מודל הדיפוזיה, הלוס הרגיל והלוס ה-perceptual ששניהם מופעלים לתמונה המשוחזרת אחרי השלב הראשון של ה-AE (לפני מודל הדיפוזיה). הלוס ה-perceptual בודק עד כמה התמונה המשוחזרת נראית ״טבעית למבט האנושי״ (משווים את האקטיבציות שלה ברשת מאומנת עם אלו של התמונות הטבעיות).

הייצוג הסופי של פיסת דאטה מתקבל אחרי ה״אנקודר״ של מודל דיפוזיה (ה-bottleneck). וכמובן יש טענות לדחיסה טובה יותר משיטות SOTA עם הגישה המוצעת…

https://arxiv.org/abs/2409.02529