המאמר היומי של מייק: 28.04.25  
Around the World in 80 Timesteps: A Generative Approach to Global Visual Geolocation

היום נסקור מאמר לא רגיל וקצת מרענן האמת- הרי לא כל יום (ואפילו לא כל חודש ואולי בכלל) יוצא לי לסקור מאמר שמדבר על מודלי למידת מכונה ביישומים גיאוגרפיים. אכן שמעתם נכון - הרי ניתן למנף את הכלים העוצמתיים של למידת דיפ (deep learning) שפותחו בשנים האחרונות גם שם.

אוקיי אז המשימה שהמודל דן בה הוא זיהוי של מיקום על כדור הארץ שבו צולמה תמונה נתונה. כלומר עבור תמונה נתונה אני צריכים להגיד מה הקואורדינטות על כדור הארץ (שכמו שאתם בטח יודעים מהווה ספירה (sphere)). המאמר מאמן מודל דיפוזיה שהקלט בו היא תמונה והפלט הוא הקואורדינטות על כדור הארץ (אני מניח שניתן לתאר מיקום על ספירה באמצעות וקטור דו-ממדי).

אתם זוכרים שמודלי דיפוזיה מאומנים להסיר את רעש מדאטה בצורה הדרגתית כלומר כל הפעם המודל חוזה רעש שהתווסף לדאטה מאיטרציה הקודמת. כלומר בהינתן פיסת דאטה מורעשת ומספר איטרציה (בכל איטרציה מתווספת לדאטה כמות קטנה של רעש) המודל חוזה את הרעש שצריך להחסיר מהדאטה כדי להחזיר אותה (פיסת דאטה) ל״איטרציה הקודמת״ הפחות מורעשת. באינפרנס המודל מתחיל מרעש טהור והופך אותו על ידי הסרת רעש הדרגתית.

מודלי דיפוזיה האחרונים מבוססים על גישה שנקראת flow matching או FM בקצרה. FM מגדיר פונקצית מהירות שבאמצעותה ניתן לתאר את המסלול בין התפלגות הדאטה (המיוצגת על ידי דגימות בדאטהסט) לבין רעש טהור. מהירות זו יכולה להיות תלויה במספר האיטרציה t (כמו במאמר הזה) או קבועה כמו בלא מעט מאמרים אחרונים על מודלי דיפוזיה. המודל מאומן לשערך מהירות זאת בהינתן דגימה רועשת x\_t ומספר איטרציה t משערך את המהירות (v(t. אחרי שיש לנו אומדן זה ניתן לבצע אינפרנס על ידי פתרון משוואה דיפרנציאלית רגילה שהיא בעצם הגדרה של המהירות בתור נגזרת של x\_t לפי t.

אוקיי, כשיש לנו תמונה שהפיקסל שלה זה מספר כלשהו ביו -1 ל 1 די לבחור התפלגות הרעש בתור גאוסית. אבל אזרכיר כי אנו נמצאים על הספירה במשימה שלנו והרעש צריך להיות כזה שהוא עצמו יהיה על הספירה וגם הדאטה המורעש חייב להיות על הספירה גם כן. כלומר אנו נכנסים כאן לתחום של גיאומטריה רימנית (Riemann geometry) על הספירה. כלומר במקום להוסיף רעש לדאטה אנו מסובבים את הדאטה בזווית התלויה ב במספר איטרציה.

בעצם אנו מרעשים את הדאטה על ידי הנעתה בכיוון של משטח משיק עבור הספירה (שזה הכיוונים שאנו יכולים לנוע מבלי ליפול מהספירה). זה כמובן משנה את הגדרה המהירות (זה כבר לא נגזרת רגילה של x\_t לפי t) למשוואות קצת יותר מסובכות (החיים על ספירה לא פשוטים). ד״א לפי מה שאני הבנתי מהמאמר הרעש הטהור שמתחילים ממנו אינפרנס מפולג באופן אחיד על הספירה (למיטב ידיעתי זה לא לגמרי טריוויאלי להגדיר את זה מתמטית - ניתן לעשות זאת בכמה צורות).

אבל דבר אחד נותר ללא שינוי - המודל מאומן לאמוד את המהירות עבור האיטרציה t בהינתן קואורדינטה מורעשת על כדור הארץ של תמונה נתונה, שגם מהווה קלט למודל דיפוזיה. התמונה מוזנת לרשת אחרי העברת דרך אנקודר שלא מאומן (נותר מוקפא).

המאמר מאוד מעניין - מי שבקיא בגיאומטריה רימנית מוזמן לצלול ולהינות :)

https://arxiv.org/pdf/2412.06781