# Review 117b: [Short] Predicting masked tokens in stochastic locations improves masked image modeling, 07.08.23 https://arxiv.org/abs/2308.00566.pdf

**Paper: https://arxiv.org/abs/2303.00289v1**

היום ב-#shorthebrewpapereviews סוקרים מאמר של כמה חוקרים ישראלים עם Yann LeCun האגדי!! שיטות למידה self-supervised (או SSL) הפכו להיות מאוד פופולריות לבניית ייצוג עוצמתי עבור דאטה ויזואלי (תמונות) שניתן להשתמש בו למשימות מגוונות. שיטות אלו לא דורשות דאטה מתויג ולכן ניתן לאמן אותם על דאטהסטים ענקיים של תמונות מהאינטרנט.   
  
בדרך כלל שיטת SSL מהנדסת משימה שלא דורשת תמונות מתויגות. למשל אחד המאמר האחרונים של יאן לקון (I-JEPA) המשימה הייתה חיזוי הייצוג (embedding) של פאץ בתמונה נתונה בהינתן ייצוגים של פאצ'ים אחרים של התמונה. ככה ייצוג שנבנה לומד להפיק את המאפיינים הסמנטיים של הפאצ'ים מייצוג הויזואלי של הפאצ'ים באותה תמונה. במאמר I-JEPA המודל מקבל את הייצוגים של כמה פאצ'ים (ההקשר) יחד עם הקידוד המיקום שלו בתמונה (positional encoding) של המיקומים של הפאץ' שחיזויו היה צריך לחזות היה מיוצג עם וקטור המיסוך (הקבוע עבור כל הפאצ'ים) וגם קידוד המיקום שלו בתמונה.   
  
במאמר הנסקר המחברים מבקשים להכליל את הגישה של I-JEPA ובמקום קידוד מיקום מדויק להעביר למודל קידוק מקומי מורעש (גם עבור פאצ'י הקשר וגם עבור פאצ'י שייצוגם נחזים). איך זה נעשה? פשוט מוסיפים וקטור גאוסי עם מטריצת קווריאנס S נלמדת לוקטור קידוד מיקום. ככה אנו הופכים את משימת SSL מורכבת יותר וכתוצאה מכך הייצוגים המופקים באמצעותה משתפרים. מאחר וצריך ללמוד פרמטרים של ההתפלגות שממנה צריך לדגום את הוקטור המורעש המייצג מיקום אז בדומה ל-VAE משתמשים ב-reparameterization trick.