# Review 87: [Short] TokenFlow: Consistent Diffusion Features for Consistent Video Editing

**Paper: https://arxiv.org/abs/2307.10373v3**

סוקרים המאמר כחול-לבן היום קצרות ב #shorthebrewpapereviews

חוסר רציפות בין הפריימים: מה הבעיה הגדולה ביותר בעריכה של וידאו עם באמצעות מודלי דיפוזיה גנרטיביים? מודלי דיפוזיה מסתדרים די יפה עם עריכה של תמונות לפי תאור טקסטואלי אבל עם הוידאו הסיפור הוא יותר מסובך כי נדרשת רציפות בין הפריימים.

הדרך הנאיבית לבצע עריכת וידאו בהתאם לתיאור טקסטואלי היא לערוך כל פריים (תמונה). אבל איך נשמור על קוהרנטיות בין הפריימים הערוכים? המחברים לוקחים פיצרים של הפריימים הסמוכים ולהשתמש בהם ולהחליק את הוידאו ערוך בעזרת אינטרפולציה של הפיצ'רים של הפריימים הקרובים לו.

אבל מה הם הפיצ'רים של הפריימים שכדאי לקחת? קודם כל המחברים לוקחים את את השאילתות, מפתחות וערכים (queries, keys, values) ממנגנוני ה-attention מכמה פריימים ערוכים סמוכים של הוידאו המקורי. לאחר מכן עבור פריים i מפעילים מנגונן -attention על השאילתה שלו ועל המפתחות והערכים של הפריימים.

ככה למעשה מחושב ״צוג הרציפות״ של וידאו (המורכב מייצוג של כל פריים ביחס לפריימים האחרים). לכל פריים מחפשים את הפריים הבא לפניו וזה שבא אחריו עם ה-attention הקרוב ביותר לפי מרחק הקוסינוס מבחינת ייצוג הרציפות. ואז עבור כל פריים שאנחנו עורכים אנו משפרים את רציפות בין הפריימים תוך שמירה על אותו ״ייצוג רציפות״ כמו בוידאו המקורי על ידי אינטרפולציה שלו על ידי שני ייצוגי הרציפות של הפריימים שמצאנו.

https://arxiv.org/abs/2307.10373