# Review 102: [Short] TokenFlow: Consistent Diffusion Features for Consistent Video Editing, 21.07.23 https://arxiv.org/abs/2307.10373

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

היום סוקרים מאמר כחול-לבן ב #shorthebrewpapereviews חוסר רציפות בין הפריימים: מה הבעיה הגדולה ביותר בעריכה של וידאו עם באמצעות מודלי דיפוזיה גנרטיביים? מודלי דיפוזיה מסתדרים יפה עם עריכת תמונות לפי תיאור טקסטואלי אבל עם הוידאו הסיפור יותר מסובך כי נדרשת רציפות בין פריימים. הדרך הנאיבית לבצע עריכת וידאו בהתאם לתיאור טקסטואלי היא לערוך כל פריים (תמונה).   
  
אבל איך נשמור על קוהרנטיות בין הפריימים הערוכים? המחברים לוקחים פיצ'רים של הפריימים הסמוכים ומשתמשים בהם כדי להחליק את וידאו ערוך בעזרת אינטרפולציה של הפיצ'רים שלו עם הפריימים הקרובים לו.   
  
אבל אלו פיצ'רים של הפריימים כדאי לקחת? קודם כל המחברים לוקחים את את השאילתות, מפתחות וערכים (queries, keys, values) ממנגנוני ה-attention מכמה פריימים סמוכים של הוידאו המקורי. לאחר מכן עבור פריים i מפעילים מנגנון-attention על השאילתה שלו ועל המפתחות והערכים של הפריימים האחרים. ככה למעשה מחושב ״צוג הרציפות״ של הוידאו המקורי (המורכב מייצוג של כל פריים ביחס לפריימים האחרים). לכל פריים מחפשים את הפריים שבא לפניו ואחד שבא אחריו עם ה-attention (בין q, k, v שהסברנו קודם) הקרוב ביותר לפי מרחק הקוסינוס מבחינת ייצוג הרציפות.   
  
ואז עבור כל פריים שאנחנו עורכים אנו משפרים את הרציפות שלו עם הפריימים אחרים תוך שמירה על אותו ״ייצוג רציפות״ כמו בוידאו המקורי על ידי אינטרפולציה שלו באמצעות שני ייצוגי הרציפות של הפריימים שמצאנו.