

מציגים - יהב בר דוד וליאור אריאל רוז  
מנחה - ד"ר אירינה רבייב

### תקציר

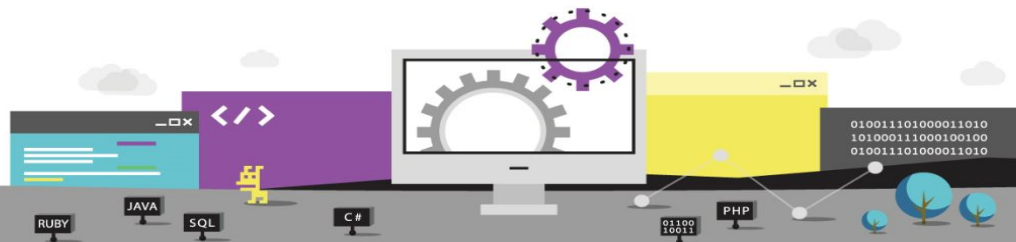
מסמכים היסטוריים יכולים לחשוף מידע רב (צורת כתיבה, ניסוח, תוכן שלא היה קיים ועוד). על מנת לבצע למידה חישובית (Machine Learning) יש צורך בכמות אדירה של נתונים מסווגים (Classified Data). תהליך יצירת הנתונים המסווגים (Annotations) הינו עבודה יקרה ומייגעת, ולכן בתחום של מסמכים היסטוריים, מאגרי המידע שקיימים לאימון מודלים קטנים בגודלם. גודל כזה אינו מאפשר לאמן מודלים עמוקים במיוחד כדי לקבל תוצאות גבוהות. על מנת ליצור דאטא סט גדול, בצורה קלה המצריכה פחות משאבים, יש צורך ביצירת דאטא סינטטי. בפרויקט "מחולל סגנון היסטורי" חקרנו שיטה ליצירת דאטא היסטורי סינטטי ופיתחנו מערכת (אתר) המאפשרת לכל משתמש לסנתז מסמכים בעצמו.

### טכנולוגיות



### שיטות

למידה חישובית - מודל Neural Style Transfer המבוסס על למידה עמוקה. ראייה ממוחשבת - טכניקות של בינאריזציה (Binarization), הרחבה (Dilation), עיבוד תמונה (Image Processing).



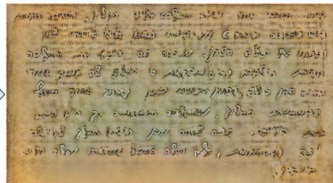
## דוגמאות



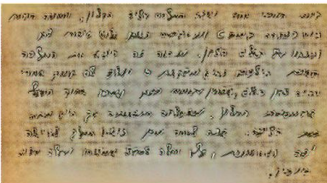
הנה תבין את זה שאתה שומע שיש חוקים, ואתה חושב  
שהם רק דברים שאתה צריך להקפיד עליהם, אבל  
אין לך שום חלק בהם. אתה חושב שאתה יכול  
לעבור עליהם בלי שום בעיה. אבל אתה צריך  
להבין שיש לך חלק בהם, ואתה צריך  
להקפיד עליהם. אתה צריך להבין שיש  
לך חלק בהם, ואתה צריך להקפיד עליהם.

[illegible]

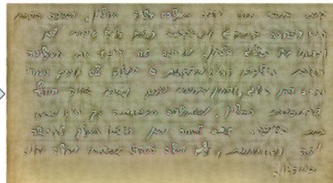
weight 4 - block3 conv2



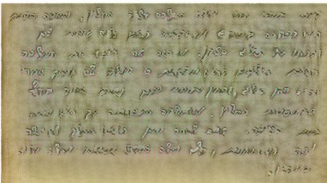
weight 5 - block3 conv2



weight 3 - block3\_conv2



weight 4 - block3\_conv2



## ניסויים

ביצענו ניסויים ע"י הרצת השיטה על מסמך היסטורי ומסמך מודרני, כך שבכל ניסוי שינו את האלמנטים הבאים:

- השכבה ברשת הנוירונית ממנה אנו מחלצים את ייצוג תמונת המסמך המודרני.
- המשקל למסמך היסטורי ולמסמך מודרני.
- מסמך מודרני (התוכן) ע"י החלת טכניקות של ראייה ממוחשבת.

## מסקנות

לאחר ביצוע מספר רב של ניסויים וניתוח התוצאות, התוצאות המיטביות ביותר מתקבלות ע"י החלת הרחבה (Dilation) לתוכן במסמך המודרני וחילוץ המסמך (התמונה) משכבות block3-conv2 ומשקל תוכן של 150-175.

