**המחלקה להנדסת תוכנה**

**פרויקט גמר – תשפ"ב**

**אותנטיקציה בין שני כתבי יד**

**Authentication between two Handwriting**

**מאת**

**מועתז תמימי מוסטפא אבו גנאם**

**319019378 318680253**

**מנחה אקדמי: דר' יהודה חסין אישור: תאריך:**

**רכז הפרויקטים:דר' שפנייר אסף אישור: תאריך:**

מערכות ניהול הפרויקט:

|  |  |  |
| --- | --- | --- |
| # | מערכת | מיקום |
| 1 | מאגר קוד | https://github.com/MustafaAbuGhannam/Authentication\_between\_two\_hand\_writing |
|  |  |  |

מידע נוסף (מחק את המיותר)

|  |  |
| --- | --- |
| סוג הפרויקט | מחקרי ממרצה במכללה |
| פרויקט מח"ר | לא |
| פרויקט ממשיך | זה פרויקט חדש (מנסה לשפר תוצאות של פרויקט קודם) |
| פרויקט זוגי: | כן |

מבוא

לקביעת המחבר של טקסט כתוב יש מעשי משמעות בתחום הזיהוי הפלילי, אימות חתימה, והיסטוריה ספרותית. למשל, במבחן הפסיכומטרי אם רוצים לדעת אם אותו בן אדם עשה את שני המבחנים כדי למנוע הונאות .

לכן אם נוכל לזהות מי עשה את המבחן על ידי זיהוי אותו דרך כתב היד שלו בפרק של החיבור במבחן

נוכל למנוע הרבה הונאות .

ניתוחים מסוג זה מבוססים כולם על השוואה בין דוגמאות כתיבה שונות. הטכניקות בתחום נותרו סובייקטיביות במידה רבה, מה שמקשה על המעבר לכלים אוטומטיים.

המטרה העיקרית של פרויקט זה היא לפתח מערכת אוטומטית בעלת דיוק גבוה, מערכת שיכולה לקבוע אם שתי דוגמאות כתיבה נכתבו על ידי אותו אדם. בנוסף, המערכת שלנו צריכה להיות מסוגלת להתמודד עם מחברים שמעולם לא נתקלה בהם בעבר.

דרישות ואפיון הבעיה

## דרישות ואפיון הבעיה

המשתמש מכניס 2 חיבורים וירצה לדעת אם שני החיבורים נכתבו על ידי אותו בן אדם או לא

## הבעיה מבחינת הנדסת תוכנה

בהינתן שני חיבורים רוצים לדעת האם שניהם נכתבו על ידי אותו בן אדם בשימוש של עיבוד תמונה ו . Deep learning

בשביל זה צריך לחשוב על:

* דרך לאסוף Data מספיק שהרשת תלמד.
* לבחון ולבחור ארכיטקטורה לרשת שתהיה מתאימה לבעיה שלנו,
* לחשוב על דרך להכין את ה data כדי שיהיה מאוזן ונוכל להעביר אותו לרשת
* הדיוק של הרשת צריך להיות מספיק גבוהה כדי שנוכל להשתמש בה
* אחרי אימון הרשת זמן ההכרעה של הרשת על דוגמה חדשה צריך להיות נמוך

תיאור הפתרון

הפתרון מחולק ל 3 חלקים מרכזים:

1-הכנת Data : כדי שהרשת שאנחנו הולכים לעשות תגיע לאחוזים גבוהים של דיוק אנחנו צריכים לספק לה מספיק data לשלבי האימון והבחנה והמרכז הארצי לבחינות והערכה הסכים לספק לנו 39 זוגות של חיבורים בשפה הערבית ממבחנו הפסיכומטרי.

כל חיבור היה בעל 2 דפים וכל חיבור בערך היה בעל 35 שורות

כדי לבנות את ה pairwise dataset המשמש לאימון המודל שלנו, אנו מפרידים בין כותבים המשמשים בהכשרה, באימות ובמבחן שלנו. אנו עושים זאת על מנת לוודא שהרשת מאומנת להיות אגנוסטית לסגנונות כתב יד ספציפיים של מחבר. ל התאמה חיובית, אנו דוגמים באקראי מחבר ואז דגימה אקראית של שתי שורות שונות שנכתבו על ידי אותו מחבר. לאי התאמה שלילית, אנו דוגמים שני כותבים שונים ולאחר מכן דוגמים באופן אקראי שורה מכל אחד. אנחנו בונים dataset שלנו להיות מאוזן ( בעל מספר שווה של דוגמאות חיובים ושלילים ).

2- עיבוד הData : לקחנו כל חיבור של סטודנט וחילצנו את השורות ומכל השורות שמצאנו לקחנו שורות שמכילות לפחות 5 מילים ואז לכל שורה עשינו Cropping and Down-Sampling כך שכל השורות יהיו באותו אורך ורוחב כך שיהיה יותר קל לאמן את הרשת.

3- :Architecture בפרויקט זה אנחנו הולכים להשתמש ב Siamese Architecture ראו איור 1 ,

Graphical user interface

Description automatically generatedבארכיטקטורה זו שני תמונות מועברות לשתי רשתות CNN שוות ( אותם משקלים לשתי הרשתות ) ואז התוצאות מועברות לפונקציית מחיר וכך מאמנים את הרשת

איור 1

**סקירת עבודות דומות \ בספרות והשוואה**

1. לפני שנתיים נעשה פרויקט מחקרי דומה במכללה שמטרתו הייתה לעשות אותנטיקציה בין שתי כתבי יד בשפה העברית הם עשו את זה על ידי חילוץ אותיות מסוימות ואז אמנו רשת שמסוגלת למצוא הבדלים בין שתי כתבי היד על ידי איך כל אחד כותב את אותם אותיות שהם חילצו והם הגיעו לאחוז הצלחה טוב בשפה העברית, אך זה לא עבד טוב כשעברו לשפה הערבית בגלל שהשפה הערבית מחוברת וקשה לחלץ ממנה אותיות.
2. מצאנו מאמר [[1]](#footnote-1) דומה למה שאנחנו הולכים לעשות והם מנסים למצוא הבדלים בין שתי כתבי יד על ידי השוואת שורות והם הגיעו לאחוזים למאוד גבוהים על Dataset יחסי קטן לחיבורים באנגלית והם השתמשו בארכיטקטורה שאנחנו הולכים להשתמש בה .

1. http://cs231n.stanford.edu/reports/2017/pdfs/801.pdf [↑](#footnote-ref-1)