**פרויקט גמר 2022 – אוניברסיטת אריאל**

**מוגש ע"י מתן-בן נגר, רוני הראל, עשהאל כהן, אוריה שפיגל**

**מרצה: גיל בן ארצי**

**Introduction**

שפת סימנים הוא כינוי לכל אחת מהשפות הטבעיות שמשתמשות בתנועות ידיים, מחוות גוף והבעות פנים כאמצעים העיקריים להעברת מסרים וליצירת תקשורת בין הדוברים. לכן, השימוש בשפת הסימנים מאפשר שמירה על השקט. זאת בניגוד לשפה מדוברת, שבה האמצעי העיקרי הוא הקול המופק בחללי הפה, האף והגרון.

האוכלוסייה העיקרית שמשתמשת בשפת סימנים היא אוכלוסיית החירשים וכבדי-השמיעה, בעיקר אלה שנולדו כבדי-שמיעה או שלקו בשמיעתם בגיל צעיר. במקומות שונים בעולם משמשות שפות סימנים שונות, ואין בהכרח יכולת לדובר שפת סימנים אחת להבין דובר בשפת סימנים אחרת.

לפי אמונה שהייתה רווחת בעבר, שפת סימנים היא שפה דלה שאינה אלא פנטומימה והצגה חזותית ופשטנית של הדברים הנאמרים, ולא ניתן להביע באמצעותה רעיונות מופשטים ומורכבים. אולם בעשורים האחרונים של המאה ה-20 בלשנים התחילו לחקור ביתר רצינות את שפת הסימנים והגיעו למסקנה כי מדובר בשפה לכל דבר, בעלת אוצר מילים עשיר ודקדוק מורכב - ממש כמו כל שפה טבעית אחרת.

**Necessity**

למה אנחנו צריכים טכנולוגיות לזיהוי שפת הסימנים?

אנשים חירשים ואילמים משתמשים בשפת הסימנים של תנועות ידיים כדי לתקשר, ומכאן שאנשים רגילים נתקלים בבעיות בזיהוי שפתם על ידי סימנים שנעשו. מכאן שיש צורך במערכות המזהות את הסימנים השונים ומעבירות את המידע לאנשים רגילים.

**The Problem**

שפת סימנים אינה תלויה בשפה המדוברת ואינה שואלת ממנה בהכרח את אוצר המילים ואת כללי הדקדוק והתחביר, כלומר - אינה תרגום מילה-במילה של השפה המדוברת (בניגוד ל"שפות המסומנות"). דוגמה בולטת לכך היא שפת הסימנים הבריטית ושפת הסימנים האמריקאית, שהן שפות שונות מאוד והדובר בשפה אחת לא יבין את הדובר בשפה השנייה - זאת למרות העובדה שגם בבריטניה וגם בארצות הברית מדברים באותה שפה, אנגלית.

עובדה זאת גורמת למחסור במאגרי מידע גדולים ואיכותיים של מידע מתוייג בכל הקשור לשפת הסימנים בשפות שונות. כיום לא ניתן למצוא מגוון רב של מודלים איכותיים לזיהוי שפת הסימנים בשפות שאינן בשפה האנגלית.

**Suggestion Proposal – slGAN – Sign Language Generative Adverisal Networks**

האם ניתן להגדיל את מאגרי המידע שלנו באופן מלאכותי?

אנחנו מעוניינים לחקור את מודל ה-slGAN בסביבה מצומצמת של אלפב"ת בשפת הסימנים כדי לראות האם הוא מצליח להניב תוצאות טובות יותר ממודלים רגילים.

מהו מודל ה-slGAN? (תיאורטית).

המודל הוא למעשה מודל Semi-Supervised המכיל בתוכו את הרכיבים הבאים:

1. Genertator – אחראי על ייצור תמונות של שפת הסימנים
2. Discrimminator – האם התמונה נראית אמיתית

\* 2 רשתות נוירונים אלו יהיו אחראיות על ייצור המידע הסינטטי הלא מתוייג.

3. Classifier – רשת נוירונים שתאומן על 10% (אולי אפילו פחות) מתוך מידע מתוייג .

<https://www.kaggle.com/datasets/datamunge/sign-language-mnist>

כאן למעשה נכנס החלק של ה-Semi-Supervised, אנחנו מדמים סיטואציה בה כמות המידע שלנו הוא מצומצם.

בשאר המידע אנחנו נשתמש כדי להשוות מודל רגיל שאומן על 80% מהדאטסט לעומת המודל שלנו שאומן על 10% בלבד.

1. את התמונה שקיבלנו מה-Generator ועברה את מבחן ה-Discrimminator נעביר אל ה-Classifier לצורך תיוג ולבסוף נוסיף למאגר מידע שלנו.
2. נאמן שוב את המודל שלנו על הdataset החדש שיוצר.
3. נבצע השוואה בין מודלים שונים שאומנו על כל ה-80% מהדאטה סט, לעומת המודל שלנו שאימן את עצמו , נבחן את הביצועים של הGAN . תוצאות המודל יוצגו כך:

|  |  |  |  |
| --- | --- | --- | --- |
| Loss | Accuracy | Amount of training data |  |
|  |  |  | snGAN |
|  |  |  | CNN |
|  |  |  | Model1 |
|  |  |  | Model 2 |
|  |  |  | Model3 |