

שם הפרויקט : Fletcher

שמות הסטודנטים:

אמיר סילברמן 213108749

רז דיימונד 322281528

רועי סדן 324075381

שחר זנזורי 209295096

התמחות: למידה עמוקה Deep Learning

מאשר הפרויקט: ד"ר משה בוטמן

שמות הסטודנטים :

אמיר סילברמן amirsil2002@gmail.com

רז דיימונד raz.diamond@cs.colman.ac.il

רועי סדן roisadan@gmail.com

שחר זנזורי zanzurishahar@gmail.com

מטרת הפרויקט :

מטרת הפרויקט היא לפתח מודל למידה עמוקה המבוסס על רשת נוירונים, המסוגל לעבד קובץ אודיו של שיר שהוקלט בלייב – בין אם באמצעות הקלטה ישירה או צילום – בתנאים שאינם סטרייליים וללא בידוד רעשי רקע, כגון מחיאות כפיים, קריאות קהל ורעשים סביבתיים אחרים.

המודל יפיק גרסה נקייה של האודיו, המשמרת את הקלטת ההופעה עצמה, בדומה להקלטה שהייתה נשמעת אם רק הצלילים של הזמרים והנגנים היו מוקלטים בצורה נקייה ומדויקת, ללא הפרעות חיצוניות.

המטרה היא ניקוי ושחזור חוויית אולפן מתוך הקלטת לייב, תוך שמירה על האותנטיות של ההופעה.

מה קיים היום:

בשנים האחרונות חלה התקדמות משמעותית בפיתוח תוכנות ומודלים לעיבוד אודיו, במיוחד בכל הנוגע לבידוד רעשי רקע מהקלטות חיות, כמו הופעות.

להלן מספר כלים וטכנולוגיות בולטים בתחום:

מה הם הסוגים השונים של הטכנולוגיות היום?

1. מעבדי אודיו (Audio Processors):

מעבד אודיו הוא כלי תוכנה וחומרה שתוכנן במיוחד לעיבוד בזמן אמת של אודיו והקלטה, מעבדים אלה פועלים בשכבות נמוכות של נתוני הסאונד ומשתמשים באלגוריתמים מתקדמים כדי לנתח, לשנות או לשפר את הצליל.

מעבד לדוגמא : AudioMix

- תפקיד מרכזי: AudioMix הוא מעבד אודיו מתקדם שמשתמש בטכנולוגיות כמו למידת מכונה ורשתות נוירונים לעיבוד נתונים קוליים. ניתן להשתמש בו לצרכים כמו:
 - שיפור איכות הצליל.
 - הפרדת כלי נגינה ושירה (source separation).
 - סינתזה של אפקטים קוליים או מוזיקה.
 - הפחתת רעשים והסרת הפרעות מהקלטות.
- מאפיינים בולטים:
 - עיבוד בזמן אמת: מאפשר להופעות חיות להישמע ברמה אולפנית באמצעות שיפורי אודיו מידיים.
 - שימושים עיקריים:
 - בהופעות חיות: איזון סאונד דינמי והתאמת רמות צליל כדי לשפר את חוויית ההאזנה.
 - באולפנים: ניקוי רעשים ושיפור איכות הקלטות.
- למה זה טוב:
 - מבצעת סנכרון טוב בזמן אמת ויכולת של הקלטה איכותית בlive
 - זה התוצר הטוב ביותר שקיים כרגע בשוק להקלטה בלייב.
- פחות טוב:
 - זה כלי חדיש, זה אומר שקיימות הופעות ישנות יותר, שלא השתמשו במעבד הזה ולכן האיכות של הסאונד גם אם הוא מתוך ההופעה, הוא פחות איכותי, ולא מבצע בידוד כמו שצריך.
 - זה המצוי ולא הרצוי: יש בידוד אבל לא מלא, עדיין גם אם אני אכנס ליוטיוב ואשמע הופעה בלייב של להקה או זמר, אני אשמע טוב אבל אם הקהל מספיק רועש, לא נשמע טוב, ועדיין יהיה רעשי רקע של הקהל.
 - דורש המון זמן וכלים איכותיים: כדי שבאמת נוכל לקבל את האודיו נקי צריך הרבה כלים טכנולוגיים בזמן ההופעה, וגם אחריה (מיקרופונים מיוחדים, ורגישים, כיול ועבודה נוספת גם אחרי שההופעה הוקלטה וניסיון של בידוד רעשים נוסף)
- למה צריך אותנו:
 - הקלטות של הופעות שלא הוקלטו עם המעבד הן הקלטות פחות איכותיות, המעבד הוא חדש ומתחדש כל הזמן, קיימות הקלטות שהשתמשו במעבד פחות מתקדם ולכן הקובץ שמע או השיר שהיה בלייב הוא תוצר פחות איכותי.
 - עבודה בדיעבד
 - תלות בחומרה, כדי לבצע את ההקלטה של ההופעה בצורה טובה, נדרש לכך, ציוד היקפי וטכנולוגי בעלויות גבוהות מאוד, מיקרופונים מיוחדים עם חיישנים שהם בעלי רגישות

גבוהות יותר. המעבד עצמו, משתנה ומתחדש כל הזמן, הוא רכיב חומרה לכל דבר, ואנחנו מביאים אי תלות בחומרה.

2. תוכנות עריכה לעיבוד קובצי אודיו :

תוכנות עריכה לעיבוד קבצי אודיו הן כלי תוכנה המאפשרים למשתמשים לבצע שינויים, תיקונים, ושיפורים בקבצי אודיו. הן מיועדות לעבודה ידנית על נתונים קוליים ומספקות ממשקים גרפיים ליצירת תהליכי עבודה נוחים.

תוכנות לדוגמה : Adobe Audacity ו Adobe Audition

- יכולות עיקריות:
 - עריכת אודיו בסיסית: חיתוך, הדבקה, והוספת אפקטים.
 - הסרת רעשים: מתאים לניקוי הקלטות משירה או מהופעות חיות.
 - מאפשרת עבודה על מספר שכבות אודיו בו-זמנית, מושלם לערבוב (mixing) והפקה מוזיקלית.
 - תיקון ותזמון אודיו: ניתן לשנות את קצב השירה או לשפר חוסר התאמה בקובץ הקלטה.
 - שילוב עם וידאו: עריכה מסונכרנת של אודיו להפקות וידאו.
- למה זה טוב:
 - עוזרת לעיבוד אודיו, עבודה עם קבצי הקלטות, שינויים בתדרים ומשמשת כלי מדהים לאולפני הקלטות ומפיקים. וגם לניקוי ועבודה נוספת על אודיו מהופעות חיות.
 - תומכת באפקטים מתקדמים כמו EQ, reverb, ו-compression.
 - אינטגרציה עם שאר הכלים של Adobe (כמו Premiere Pro).
- פחות טוב:
 - תשלום, המערך הבסיסי של התוכנה שהוא החינמי, לא נותן הרבה יכולות עבודה ומאוד מצומצם.
 - נדרשת עבודה ידנית, מקצועיות במוזיקה, ובעריכה של קבצי אודיו.
 - דורשת עבודה רבה והתעסקות.
- למה צריך אותנו:
 - עבודה יותר פשוטה ונוחה, גם למי שלא יודע איך לעבוד עם הקלטות, לא מצריך ידע או עקומת למידה כלשהי.
 - חינמית
 - העבודה מתבצעת בפחות זמן.

שתי הטכנולוגיות שדיברנו עליהן משתמשות באלגוריתמים שעוזרים להם לעבד מידע ולהלן אלו המבוססים למידה עמוקה :

אלגוריתמים מבוססי למידה עמוקה

- Recurrent Neural Networks - RNNs :
 - מזהים דפוסים לאורך זמן באודיו ומנתחים כיצד רכיבי הקול משתנים.
 - מתאימים במיוחד להסרת רעשים מתמשכים כמו מחיאות כפיים או זמזום.
- Convolutional Neural Networks - CNNs :
 - מנתחים את הספקטרום של האודיו (Spectrogram) כדי לזהות ולהסיר רעשים.
 - משמשים להפרדת שירה מכלים מוזיקליים או לניקוי הקלטות רועשות.

- U-Net Architectures :
 - רשתות נוירונים המסוגלות לבצע הפרדת מקורות קול על ידי יצירת "מסיכה" שמסננת רעשים ומשאירה רק את הצלילים הרצויים.
- Generative Adversarial Networks - GANs ו Variational Autoencoders - VAEs :
 - מתמקדים בלמידה כיצד "לשחזר" את האודיו הרצוי תוך דחיית רעשי הרקע.
 - משמשים לעיתים קרובות בסביבות שבהן האודיו המקורי נפגע משמעותית.

תיאור הפרויקט:

הבעיה אליה אנחנו ניגשים היא הפרדת הרעשים מן האודיו המקורי ופלט שהוא הסאונד הרצוי מן ההקלטה, תוך שימור על מאפיינים ייחודיים שיכולים להיות שונים בין הקלטה להקלטה (למשל במצב שבו הזמר שר מילים קצת שונות או מוסיף הערות נרצה לשמר אותן ולא לסווגן כרעש)

אופן הפתרון:

- **פירוק** אודיו הקלט לגלים מבודדים על ידי שימוש ב**התמרת פורייה (FFT)**
- מעבר על כל גל לאחר הפירוק ו**סיווג** להסתברות שהוא רעש על ידי **רשת קונבולוציה** רציפה חד מימדית בעומק Pooling מסויים ולאחר מכן **רשת נוירונים FC** על מנת לנצל Joint learning של Classification ו Feature Extraction.
- **הרכבת** הגלים המפורקים לסאונד חדש על ידי נטרול הגלים שבהסתברות גבוהה הם רעש על ידי הוספת גל הפוך
- החזרת האודיו הנקי

אימון המודל:

כדי שנוכל לאמן את המודל ולמדוד עם פונקציית Loss את הטעות שלו נזדקק לצמדים של סאונד נקי וסאונד זזה עם רעשים מגוונים.

לשם כך את הדאטא אנחנו ניצור בעצמנו על ידי איסוף ויצירה של מגוון רעשים מבודדים, אותם נרכיב בצורה ציקלית ואוטומטית על שירים נקיים, כך שייוצרו לנו צמדים של אודיו נקי ואודיו זזה אך מלוכלך.

כך, בתהליך האימון נשווה את פלט המודל לאודיו הנקי שהיה ונוכל למדוד את הטעות עם פונקציית Loss, שלאחריה נבצע Backpropagation כדי לתקן לאחר את הפרמטרים של המודל ולהגדיל את הדיוק שלו בסיווג רעשים והסרתם.

טכנולוגיות וסביבת עבודה:

- Jupiter Notebook
- Python (Keras, SciPy, numpy, etc...)
- Google Colab environment for GPU runtime