**מיני פרוייקט בנושא Video Jenre**

מגישים: רועי קצב ודקל שירן

**קישור ל- git repository:**

<https://github.com/roye52371/Video-Genre/tree/master>

**הקדמה:**

אני ושותפי יצרנו Video Jenre Classifier, כלומר באמצעות מודל  *, יצרנו קוד שלוקח סרטונים קיימים ומאמן את המודל באמצעותם, כך שהמודל לומד לסווג את הסרטונים לז'אנר המתאים.*

*ה- המקורי מכיל 320 סרטונים(20 סרטונים לכל אחד מ16 הז'אנרים השונים).*

*כל סרטון חולק בצורה אחידה ל-2040 פריימים אשר אותם חילקנו לקבוצות של 120 פריימים וכל קבוצה כיווצנו לסרטון שנמצא בתת-תיקייה שונה, המתאימה לסרטון המקורי ממנו כווץ.*

*את ההרצה תחילה עשינו על 16 ז'אנרים שונים ובהמשך צמצמנו את מספר הז'אנרים ל9 ז'אנרים ולאחר מכן ל-5 ז'אנרים.*

* *-CNN רשת קונבולוציה שמשמשת כ-Feature Extractor .*
* *LSTM - רשת RNN שמקבלת Sequence , ה-LSTM מחזיר משוב על התוצאות ולומד באמצעותו בנוסף.*

**אופן הרצת המודל:**

על מנת שתוכלו להוריד את הפרוייקט ולהריצו תחילה תצטרכו להתקין את הספריות:

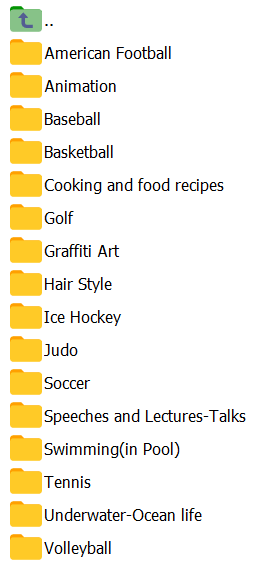
pytorch, PIL, os, glob, json, numpy, random, cv2, matplotlib, torchvision, tqdm, vit\_pytorch, sklearn, colorama, LSTMmodel, Video\_Dataset, pathlib, \_\_future\_\_, splitfolders.

שם הקובץ להרצה- smallDataset\_TestShow.py .

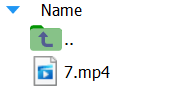
הכנו תיקיית לדוגמא למקרה מוקטן שניתן להריץ גם על ה אך מומלץ להריץ על הGPU.

ניתן להוסיף סרטונים שונים לכל תיקייה לפי הז'אנר המתאים.

התיקייה מכילה תתי תיקיות לפי סוגי הז'אנרים השונים:



ולאחר מכן כל תת תיקייה של ז'אנר מכילה סרטון יחיד:



לאחר הרצת המודל המאומן על תיקיית קיבלנו את התוצאות הבאות:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

**דו"ח תוצאות**:

מצורפים תוצאות confusion matrix, precision, Top k score, accuracy בעבור שלושה Datasets שונים עבור מודל CNN+LSTM.

אימנו מודל על שלושה *שונים.*

*כאשר ב- הראשון ישנם 16 ז'אנרים-*

Hair Style, Swimming (in Pool), Judo, Volleyball, BaseBall, Tennis, Animation, Underwater-Ocean life, Basketball, Ice Hockey, Soccer, Graffiti Art,

Cooking and food recipes, American Football, Golf, Speeches and Lectures-Talks.

*כאשר ב- השני ישנם 9 ז'אנרים-*

American Football, Baseball, Basketball, Cooking and food recipes, Golf, Graffiti Art, Hair Style, Speeches and Lectures-Talks, Underwater-Ocean life.

*כאשר ב- השלישי ישנם 5 ז'אנרים-*

Basketball, Cooking and food recipes, Speeches and Lectures-Talks, Underwater-Ocean life.

*כלל ה- מחולקים ל- 70% ול-30% .*

תוצאות ההרצה על 16 ז'אנרים:

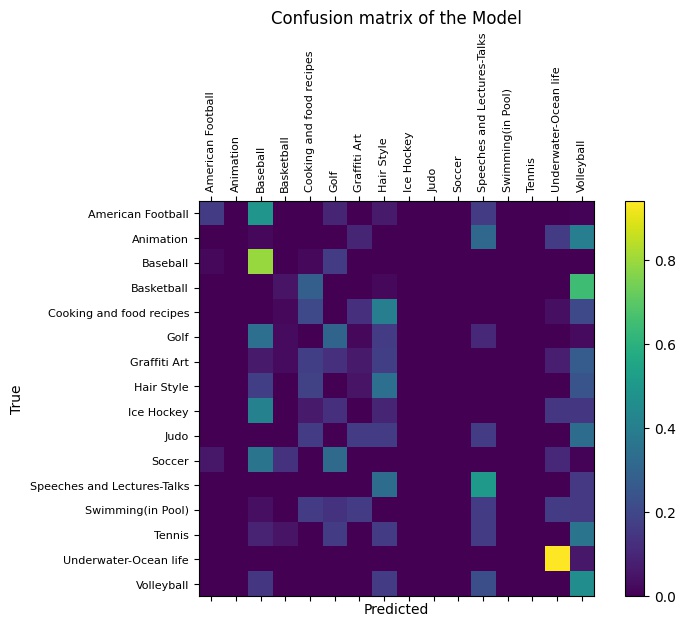
Top1 accuracy score: 0.24019607843137256

Top3 accuracy score: 0.3946078431372549

Top5 accuracy score: 0.49387254901960786

|  |  |
| --- | --- |
|  | precision recall f1-score support |
|  |  |
|  | American Football 0.68 0.17 0.27 102 |
|  | Animation 0.00 0.00 0.00 102 |
|  | Baseball 0.27 0.79 0.40 102 |
|  | Basketball 0.16 0.05 0.07 102 |
|  | Cooking and food recipes 0.16 0.21 0.18 102 |
|  | Golf 0.21 0.30 0.25 102 |
|  | Graffiti Art 0.10 0.07 0.08 102 |
|  | Hair Style 0.16 0.34 0.22 102 |
|  | Ice Hockey 0.00 0.00 0.00 102 |
|  | Judo 0.00 0.00 0.00 102 |
|  | Soccer 0.00 0.00 0.00 102 |
|  | Speeches and Lectures-Talks 0.28 0.51 0.36 102 |
|  | Swimming(in Pool) 0.00 0.00 0.00 102 |
|  | Tennis 0.00 0.00 0.00 102 |
|  | Underwater-Ocean life 0.57 0.94 0.71 102 |
|  | Volleyball 0.13 0.46 0.20 102 |
|  |  |
|  | accuracy 0.24 1632 |
|  | macro avg 0.17 0.24 0.17 1632 |
|  | weighted avg 0.17 0.24 0.17 1632 |

CNN+LSTM\_Dataset70\_30\_isBi\_\_True\_accuracy=24.019607843137255\_martix:



תוצאות ההרצה על 9 ז'אנרים:

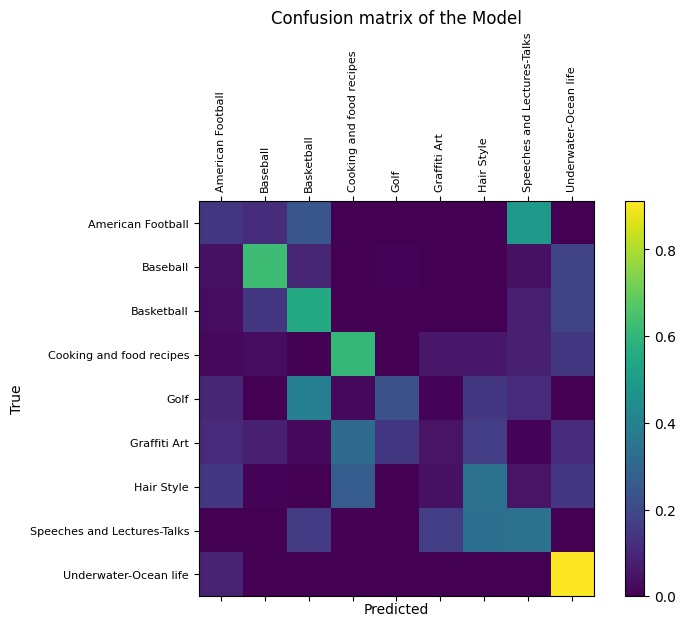
Top1 accuracy score: 0.42374727668845313

Top3 accuracy score: 0.6688453159041394

Top5 accuracy score: 0.8017429193899782

|  |  |
| --- | --- |
|  | precision recall f1-score support |
|  |  |
|  | American Football 0.22 0.15 0.18 102 |
|  | Baseball 0.62 0.63 0.62 102 |
|  | Basketball 0.38 0.56 0.45 102 |
|  | Cooking and food recipes 0.50 0.61 0.55 102 |
|  | Golf 0.59 0.23 0.33 102 |
|  | Graffiti Art 0.15 0.05 0.07 102 |
|  | Hair Style 0.33 0.34 0.33 102 |
|  | Speeches and Lectures-Talks 0.29 0.34 0.31 102 |
|  | Underwater-Ocean life 0.54 0.91 0.68 102 |
|  |  |
|  | accuracy 0.42 918 |
|  | macro avg 0.40 0.42 0.39 918 |
|  | weighted avg 0.40 0.42 0.39 918 |

CNN+LSTM\_Dataset70\_30\_isBi\_\_True\_accuracy=42.37472766884531\_martix:



תוצאות ההרצה על 5 ז'אנרים:

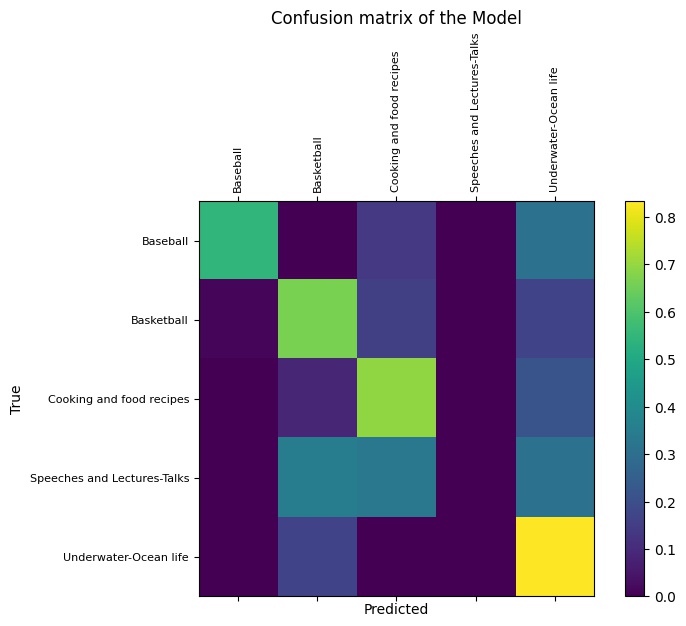
Top1 accuracy score: 0.5490196078431373

Top3 accuracy score: 0.7588235294117647

Top5 accuracy score: 1.0

|  |  |
| --- | --- |
|  | precision recall f1-score support |
|  |  |
|  | Baseball 0.98 0.55 0.70 102 |
|  | Basketball 0.52 0.67 0.59 102 |
|  | Cooking and food recipes 0.53 0.70 0.60 102 |
|  | Speeches and Lectures-Talks 0.00 0.00 0.00 102 |
|  | Underwater-Ocean life 0.45 0.83 0.59 102 |
|  |  |
|  | accuracy 0.55 510 |
|  | macro avg 0.50 0.55 0.50 510 |
|  | weighted avg 0.50 0.55 0.50 510 |

NumOfJenre\_5\_model\_CNN+LSTM\_Dataset70\_30\_isBi\_True\_accuracy=54.90196078431373\_martix:



**סיכום מפגשים ואבני דרך:**

סיכום פגישות והדרך שלנו במהלך המיני מהמפגש האחרון ועד ליום ההגשה.

17.10- מפגש ראשון עם פרופסור ג'יהאד, והסבר על מיני הפרויקט.

31.10- מפגש שני עם פרופסור ג'יהאד, שבו הוצגו הנושאים השונים לביצוע מיני פרויקט.

8.11- בחירת נושא פרויקט, Video Genre, נשלח מייל לפרופסור ג'יהאד כדי לעדכן על בחירה והמשך  עבודה.

14.11- שיחה עם פרופסור ג'יהאד בנוגע לפרויקט בפעם הראשונה, נאמר שנצטרך להכין דאטה סט של סרטונים עם מספר ז'אנרים שונים, כל סרטון באורך 2 עד 5 דקות, בנוסף נצטרך להשתמש

 ב-Deep Learning Model מסוג CNN+LSTM, ושנקבע פגישה להמשך.

21.11-שיחה עם פרופסור ג'יהאד, שהציע להפנות אותנו לדוקטורנט שלו, בוראק, כדי שניעזר בו בנושא Deep Learning Model using CNN+LSTM.

28.11- יצרנו דאטה סט ראשוני עם מספר סרטונים ל5 ז'אנרים , ותיאמנו פגישה עם בוראק.

1.12- נפגשנו עם בוראק, במהלך הפגישה בוראק תיאר בפנינו את המודל שבו נצטרך להיעזר, וסיכם איתנו על בניית דאטה סט, ושבהמשך נאמן את המודל ונבחן אותו באמצעות המחשוב באוניברסיטה.

כשדיברנו עם בוראק על הדאטה סט, עלתה התלבטות בנוגע לאורך הסרטונים, שעלולים להיות ארוכים מידיי(התבקשנו לסרטונים באורך 2 עד 5 דקות).

לאחר התייעצות עם פרופסור ג'יהאד , הוחלט להשאיר את אורך הסרטונים כמו שהוא, ובהמשך בהתאם לתוצאות, לקצר או לקחת חלק מהפריימים.

10.12 - עדכנו את הדאטה סט להיות, דאטה סט של 10 ז'אנרים שונים של סרטונים, החלטנו להתמקד בז'אנרים בנושאי ספורט.

בכל ז'אנר ישנם 10 סרטונים, כל אחד באורך 2-5 דקות. כל ז'אנר מיוצג בתיקייה עם שם הז'אנר(בתוך תיקיית דאטה סט) וכל סרטון בתיקייה של ז'אנר ספציפי ממוספר מ-0 עד 9.

הז'אנרים:

 Soccer, Basketball,American Football, Baseball, Tennis, Golf, Ice Hockey, Swimming, Volleyball, Judo.

12.12- עדכנו את פרופסור ג'יהאד לגבי הדאטה סט שיצרנו.

סוכם שנוסיף עוד 6 ז'אנרים שאינם מתחום הספורט, ובנוסף, נוסיף בין 5-10 סרטונים נוספים לכל ז'אנר, שכן 10 סרטונים לכל ז'אנר עלולים לא להספיק, סה"כ נצטך 15-20 סרטונים כעת ל-16 ז'אנרים.

18.12- עדכנו את הדאטה סט שיצרנו, והעלנו אותו לתיקיית דרייב.

**Dataset:**

הדאטה סט כעת מכיל את 16 הז'אנרים הבאים:

1. Hair Style
2. Swimming (in Pool)
3. Judo
4. Volleyball
5. BaseBall
6. Tennis
7. Animation
8. Underwater-Ocean life
9. Basketball
10. Ice Hockey
11. Soccer
12. Graffiti Art
13. Cooking and food recipes
14. American Football
15. Golf
16. Speeches and Lectures-Talks

**הערות חשובות לגבי הדאטה סט:**

* כל ז'אנר מיוצג על ידי תיקייה (בתוך תיקיית Dataset), כל ז'אנר מכיל 20 סרטונים, באורך 2-5 דקות, שממוספרים בתיקייה עם השמות 0-19. סה"כ 320 סרטונים שונים בדאטה סט.

* ז'אנר Speeches and Lectures-Talks, אינו מתעסק בהרצאות מסוג לימוד

כמו הרצאות מרצה בתיכון ובאוניברסיטה עבור קורס מסוים, אלא נאומים שונים, וסרטוני הרצאות בסגנון הרצאות העשרה כגון TED Talks וכו'.

* ז'אנר Cooking and food recipes, מתעסק בסרטוני בישול שונים של אוכל וקינוחים וכו', ובסרטונים של אנשים מבשלים מתכון ספציפי(הכוונה היא שרואים שמכינים משהו ספציפי,ושבנאדם מציג ונותן הוראות מה לבשל, לא שהסרטון מציג תמונה של מתכון מילולית, עם כמויות ומדידות), סרטונים במטבח הביתי או במטבח מסעדה ובאופן כללי סרטונים שמבשלים בהם.

* ז'אנר Swimming מודגש כשחייה בבריכה, ללא שחייה בים ובנהרות ומקומות שאינן בריכה.

27.12- נפגשנו בשנית עם בוראק, במשך כשעה דנו על הדאטא סט שהכנו, על  תהליך האימון שנצטרך לבצע בהמשך ועל אופייה. סיכמנו אימון בעזרת 3 מודלים שונים: CNN, LSTM, TRANSFORM.

חלוקת הדאטא תהא ב-2 אופנים שונים:

1. 70% train, 30% test.

2. 80% train, 20% test.

את המידע המתקבל אנו נכניס ל- confusion matrix ונבצע אנליזה על המידע המתקבל לצורך הדו"ח הסופי וניתוח התוצאות(תוצאה טובה = לפחות 70% הצלחה פר ז'אנר). לביצוע אנליזה וניתוח תוצאות נשתמש בספרייה SKLearn (נחשב בנוסף Precision, Recall לכל ז'אנר).

את ההרצה אנו נבצע דרך vpn על שרתי האוניברסיטה.

לאחר הפגישה עם בוראק שלחנו מייל לג'יהאד לגבי המשך העבודה והפגישה שהייתה.

9.1- קיימנו פגישה עם בוראק, שעסקה בהבנה של התאמת הארכיטקטורה של LSTM לבעיה שאנחנו פותרים ולהמרת הדאטה סט שלנו ל-  dataloader מתאים.

בנוסף שאלנו לגבי אופן ההרצה של הקוד בgpu האוניברסיטאי.

קבענו שנפגש בשבוע הבא להמשך תיקונים וניסיונות להרצה ואימון ראשון על מודל LSTM.

17.01- פגישה עם בוראק- דיברנו על איך לתקן את הdata loader ואת הארכיטקטורה של המודל LSTM+CNN .

25.01- דיברנו עם בוראק על בעיות הקצאות זיכרון על השרת והוא עזר וטיפל בהם.

01.02- פגישה עם בוראק- לאחר הרצה ארוכה מאוד של כ25 שעות קיבלנו תוצאה של דיוק 6%. לאחר קבלת התוצאות דנו עם בוראק על דרך לייעל את זמני הריצה ולשפר את התוצאות, הוחלט על ביצוע offline processing: עבור כל סרטון דגמנו 30 פריימים שונים בצורה אחידה ושמרנו אותם בתיקייה עם שם הסירטון, זה על מנת שבעת הרצה לא נצטרך לדגום מחדש את אותם הפריימים ונחסוך בזמן ריצה.

06.02- מייל עדכון לפרופ' ג'יהאד בנוגע לאחוזים נמוכים מאוד בתוצאות, לפני שהגדלנו את מספר הפריימים ל-30 פר סירטון. לאחר מכן נשאלנו חזר על מספר הז'אנרים זמני הרצה וכו' וכתבנו תשובה מיידית על כל הנתונים הרלוונטים.

08.02- פגישה עם בוראק- לאחר הרצה עם כלל השינויים הרלוונטיים בהם דנו עם בוראק בפגישה הקודמת, קיבלנו את אותם התוצאות של 6%. ולכן הוחלט להגדיל את מספר הפריימים לדגימה פר סירטון(כ-120). שינוי זה עזר לנו לעלות לרמת דיוק של 10%.

10.02- פגישה עם בוראק- לאחר דיון עם בוראק הוחלט על הגדלת הdataset באופן מלאכותי באמצעות augmentation על הdataset הקיים, באופן הבא: מבצעים offline processing בדרך אחרת. הוחלט לדגום מכל סרטון כ3600 פריימים ולחלק אותם לתתי סרטונים של 120 פריימים לסרטון, כל אחד מהם נשמר בתיקייה עם הפריימים שלו.

14.02-

1. לאחר שבוראק ניסה להריץ סקריפט שיוצר את הdataset באופן שתואר לעיל(10.02) על השרתים של האוניברסיטה, הוא עדכן כי אין מספיק זיכרון פנוי על מנת לאחסן את הdataset בצורתו החדשה. ולכן הורדנו את כמות הפריימים המלאה מתוך כל סרטון ל2040 פריימים.

2. מייל שנשלח לפרופ' ג'יהאד בנוגע לעזרה ועל רעיונות שיעזרו לשפר את אחוזי הדיוק שקיבלנו, במייל מתוארים כל הנתונים הרלוונטיים לאותו זמן בנוגע לדרך הרצת האלגוריתם, מספר הז'אנרים, סרטונים ופריימים.

20.02- ביקשנו מפרופ' ג'יהאד להוריד את מספר הז'אנרים ונענינו בחיוב. ובנוגע לזמני הריצה הארוכים התבקשנו להוריד את מספר הפריימים פר סירטון הנלקחים. בנוסף, פרופ' ג'יהאד הציע לנו לכווץ כלל הפריימים לתתי סרטונים כדי לחסוך בזיכרון.

28.02- עדכון לפרופ' ג'יהאד בנוגע למצב האלגוריתם שלנו מבחינת מספר הז'אנרים, סרטונים ופריימים ובקשה להגדלת הGPU שהוקצה לנו בשרת.

01.03- עדכון נוסף במייל על המודל והזמני ריצה הנוכחיים.

02.03- הרצנו את המודל עם batch size train=8 וגם batch size train=1 , ההרצה נמשכה קרוב ליום שלם וקבלנו דיוק של 24%.(הרצה עם 16 ז'אנרים).

03.03- חישבנו confusion matrix, precision וערכים נוספים למודל שקיבלנו. לאור ניתוח התוצאות שהתקבלו החלטנו לצמצם את מספר הז'אנרים לאחר התייעצות. הקטנו את מספר הז'אנרים מ16 ל9, בעקבות ז'אנרים רבים שלא זוהו כלל וז'אנר אחד שהיה לו overfitting והיווה מסיח לז'אנרים אחרים רבים.

04.03- לאחר הרצה של המודל החדש שהתארכה כ13 שעות קיבלנו רמת דיוק של כ42%.

06.03- פגישה עם בוראק, החלטה לבצע הרצה למודל של Top1(42%),Top3(67%),Top5(80%). מה שעזר לנו לקבל אחוזי דיוק גבוהים יותר.

10.03- פגישת סיכום עבודה עם בוראק.