ANDREW AND ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING

SIPL
Signal and Image Processing Lab.

Technion Israel Institute of Technology

דו"ח סיכום פרויקט: ב'

ניתוח של משחק טניס

Tennis Game Analysis

מבצעים:

אלדד אוחיון **Eldad Ohayon**

אלכסנדר בלבנוב **Alexander Balabanov**

מנחה: **ערן שחר** **Eran Shachar**

סמסטר רישום: אביב תש"פ

תאריך הגשה: נובמבר, 2020

בשיתוף עם : **Baseline Vision**

תודות

אנו מודים מקרב לב לחברת Baseline שהציעה את הרעיון לפרויקט ולאיש הקשר בחברה עומרי כרמי .

אנו מודים גם ליאיר משה , נמרוד וצוות המעבדה על התמיכה , וכמובן למנחה שלנו – ערן שחר .

תקציר

הפרויקט נעשה בשיתוף עם חברת Baseline, מטרת הפרויקט הוא זיהוי של אובייקט קטן, במקרה זה כדור טניס, בצילום וידאו מצד המגרש. הפרויקט כלל סט נתונים של סרטוני וידאו מתויגים, הסרטונים צולמו מזוויות צילום שונות, בפרט צילום צד ימין ושמאל של המגרש, תנאי תאורה שונים וצבע מגרש שונה. סט האימון הניתן לנו קטן מידיי לאימון של ארכיטקטורה חדשה לגמרי, ומותאמת למשימה זאת ספציפית וביעלות גבוהה. מכאן התעסקנו במסגרת הפרויקט במנגנון למידה על רשת מאומנת מראש. ההנחה מאחורי זה היא שרשת מאומנת וספציפית רשת שלמדה לזהות כדורים במסגרת האימון על סט גדול מאוד, תוכל להיות מוסבת לאימון על סט נתונים קטן ולתת תוצאה טובה.

בחרנו לעבוד עם רשת בארכיטקטורת yolov5 שמאפשרת זיהוי ומיקום של אובייקטים בתמונה במהירות חישוב גובהה. הרשת לוקחת כל פריים בנפרד כאילו ללא ידע על הפריים הקודם לו ובכל זאת מגיע לביצועים בזמן אמת. לקחנו את הרשת כשהיא אומנה על דאטה סט של COCO מראש, שמכיל בתוכו קטגוריה של כדורים מעולם הספורט. תהליך אימון על הדאטה סט הנתון עם הפרדה של 6600 תמונות אימון (84.6% מכלל הסט) ו-1200 תמונתו מבחן (15.4%) אימנו רשת עם תוצר mAP @.5 = 0.725 על סט המבחן.

Abstract

Student project collaborated with SIPL and partnering with Baseline, set to achieve detection and classification of a small object, specifically a tennis ball in a sideview video footage. A dataset of tagged video was provided for training purposes, the videos were being short form different angles, specifically left and right points of view, different lighting conditions and differently colored courts. The provided dataset was too small for the task of training a deep neural network from scratch, which could have been task specific and efficient. Hence we engaged in transfer learning, using the presupposition that a pre-trained neural net, specifically one that was trained to recognize different types of balls, can be trained to achieve the set goal with a small dataset.

We picked a neural net with a YOLOv5 architecture, since it has low inference time and was trained on COCO dataset that includes sports balls as a class category. The network would take each frame independently and still could achieve real time inference. Training process included 6600 training images (84.6%) and 1200 test images (15.4%) that resulted in a network with mAP @.5 = 0.725 on the test set.

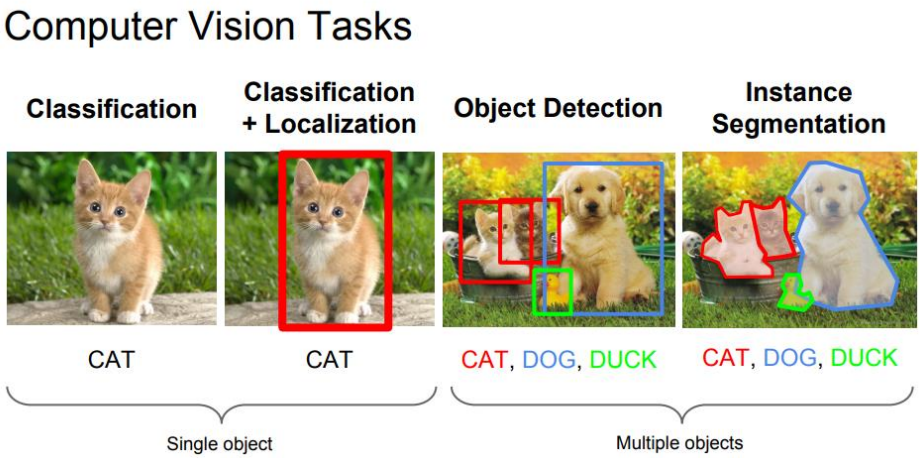
Contents

1. מבוא

נעבור על מושגים ועקרונות ששימשו אותנו לפתרון הבעיה הנתונה לנו.

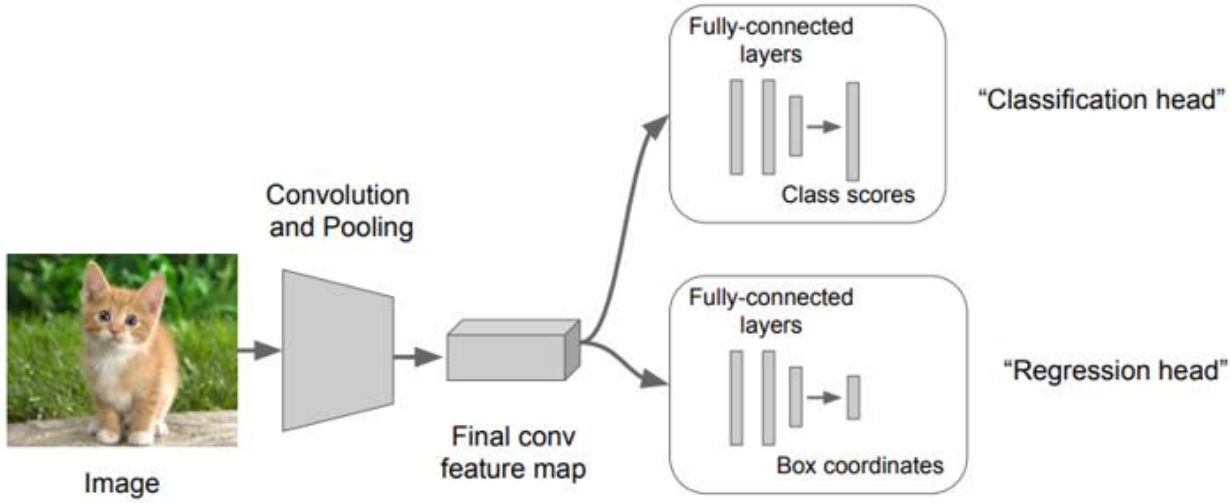
1.1. זיהוי אובייקטים

זיהוי אובייקטים זהו תחום בלמידה עמוקה שכולל בתוכו 2 פעולות נפרדות, קלסיפיקציה ואיתור האובייקט. בתחום למידה עמוקה פעמים רבות נקבל תמונה שמכילה עצמים רבים ושונים, לכן זיהוי איזה אובייקט בתמונה לא תמיד מספיק ונרצה להגיד שבאזור כלשהו קיים אובייקט ממחלקה מסויימת.



איור 1. הסבר על פעולות שונות בזיהוי אובייקטים *[1]*

האיור הנתון מסביר לנו בצורה ברורה מאוד את ההבדל בין משימות למידה שונות. מטרתנו בפרויקט להצליח לזהות אובייקטים רבים בתוך תמונה לתוך מחלקה נכונה והמיקום שלהם בצורה מדויקת ככל הניתן. נוכל סכמה כללית שמסבירה את דרך העבודה של רשתות מסוג זה



איור 2. סכמת פעילות של רשת זיהוי

נוכל לראות כאן בבירור כי התהליך יכול להיות מוגדר ל-3 פעולות כאשר 2 מהן נפרדות יחסית. חלק ראשון הוא שימוש באלגוריתמים שונים למציאת פיטצ'רים מתוך בתמונה, אותם פיטצ'רים יכולים להיות אחר כך מופנים לזיהוי מיקום האובייקט, וליזיהוי המחלקה של האובייקט בתוך הגבולות הניתנים. שימוש ברשתות לומדות פותרות בעיות רבות במשימת זיהוי אובייקטים, הסתרות חלקיות, הבדלים בצבעים ומיקומים שונים של האובייקט. רשת לומדת יודעת להתמודד עם הבדלי מיקום ע"י שימוש בקונבולוציה. רשתות גם יודעות להתמודד עם הסתרות חלקיות וצבעים שונים ע"י למידה של סט נתונים מדמה, לעומת אלגוריתמים קלאסיים שיודעים להתמודד עם מקרים מעטים.

1.2. מדדי טיב במערכות זיהוי

דיוק גבוה אינו ברור בצורה חד משמעית, נדרשת הגדרה של מדד טיב כלשהו שייתן לנו ציון על התוצאות שאנחנו מקבלים. קבוצת פיקסלים יכולים להיות חלק מאובייקט או לא, ואובייקטים יכולים להיות חלק ממחלקה או שלא, את 2 המקרים הנל אפשר להציג ע"י קלסיפיקציה לאחת מ-4 מקרים.

(TP) True Positive – זהו עצם שזוהה כפריט מהמחלקה שלנו, והוא אכן מהקבוצה אותה ניבאנו.

(FP) False Positive – זהו עצם שזוהה כפריט מהמחלקה, אך אינו חלק מהקבוצה.

(TN) True Negative – זהו עצם שלא זוהה כפריט, והוא אכן אינו שייך לקבוצה.

(FN) False Negative – זהו עצם שלא זוהה כפריט, אך הוא שייך לקבוצה.

1.2.1. דיוק precision

דיוק הוא מדד הנקבע ע"י מידת החפיפה בין מיקום "האמיתי" (לפי הסט המתויג, ייתכן ואינו מושלם) והמיקום אותו הרשת פלטה. קיים מדד בשם IoU – intersection over union אשר מוגבל בתחום 0-1 ופועל באופן הבא

איור 3. חפיפה בין האמת והניבוי

ניקח את הניבוי והחפיפה שלו מול האמת שניתנה לנו, נקבע את הקבוצה של הניבוי בתור A וקבוצת האמת בתור קבוצה B.

איור 4. איחוד בין האמת והניבוי

החישוב מתבצע ע"י יחס השטחים, מה שיוצא מדד אינווריאנטי לגודל האובייקט

$$IoU = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN}$$

חיסרון של מדד זה הוא שאם אין חפיפה כלל נקבל ערך IoU אפס, אך מצב זה אינו טוב לאימון, כי אין שום יכולת לדעת מהי גודל השגיאה, ייתכן והניבוי קרוב מאוד לחפוף ואין זה מקבל דירוג טוב יותר מניבוי שגוי לחלוטין. לכן בפרויקט נשתמש במדד [2] GIoU – Generalized Intersection over union.

בהבדל בין שני המדדים הוא ההתחשבות בשטח שמכיל את B,A אך אינו באף אחד מהם, נקרא לו קבוצה C. זהו בעצם "שטח מת" שיוצר מדד מרחק מהתאמה טובה. שימוש בתוספת זאת מאפשר ל-GIoU להיות שלילי ולהוות יד מנחה בתחילת האימון.

$$GIoU = IoU - \frac{C \ \setminus (A \cup B)}{C}$$

1.2.2. Recall

מדד זה מודד את כמות האובייקטים שהצלחנו לזהות נכונה לעומת כמות האובייקטים הקיימים.

אפשר להגדיר זאת בצורה מדוייקת יותר כ היחס הבא

$$Recall = \frac{TP}{TP + FN}$$

1.2.3 mAP

קיים יחס עדין בין הרצון למצוא את כל האובייקטים שקיימים לעומת אי מציאה של אובייקטים שאינם נכונים. אם נוריד את ערך ה-IoU המינימלי לזיהוי ייתכן שנזהה אובייקטים שלא קיימים (false positive) FP אם נגדל את המדד יותר מידיי נפספס נק' אמיתיות (false negative) FN. לכן נבצע חישוב ממוצא של הדיוק לרמות IoU שונות ונשתמש בזה כמדד גלובלי לביצועים של ניבוי [3].

1.3. העברת למידה

העברת למידה הידועה כ-Transfer learning היא שיטה אשר מאפשרת אימון של רשת עמוקה ומסובכת עם סט דאטה קטן יחסית. השיטה מתבסת על ההנחה כי אם נאמן מראש רשת על סט מידע גדול ושנגיש לכלל, משל COCO, הרשת ששמתקבלת היא גנרית מספיק ללמוד קשרים חדשים בקלות. בסט האימון שנשתמש בו קיימים מספיק אובייקטים שונים והלמידה היא לכל מיני פיצ'רים שונים של אובייקטים, כך שנוכל להסב את הרשת הקיימת ללמוד את האובייקט שלנו. לא צריך ללמוד פיצ'רים בתמונה אלא רק איך הפיצ'רים הקיימים מתקשרים לאובייקט אותו אנחנו מחפשים. שיטה פופולארית ל-TL היא "הקפאה" של הרשת במצב בו היא אומנה מראש ולאפשר שינוי משקלים רק בשכבה האחרונה או מספר שכבות קטן בסוף. שיטה זו מקצרת את זמן האימון כי אין צורך לחשב גרדיאנטים לאורך כל הרשת [4].

2. סקר ספרות

2.1. סט מידע קיים לרשת מאומנת

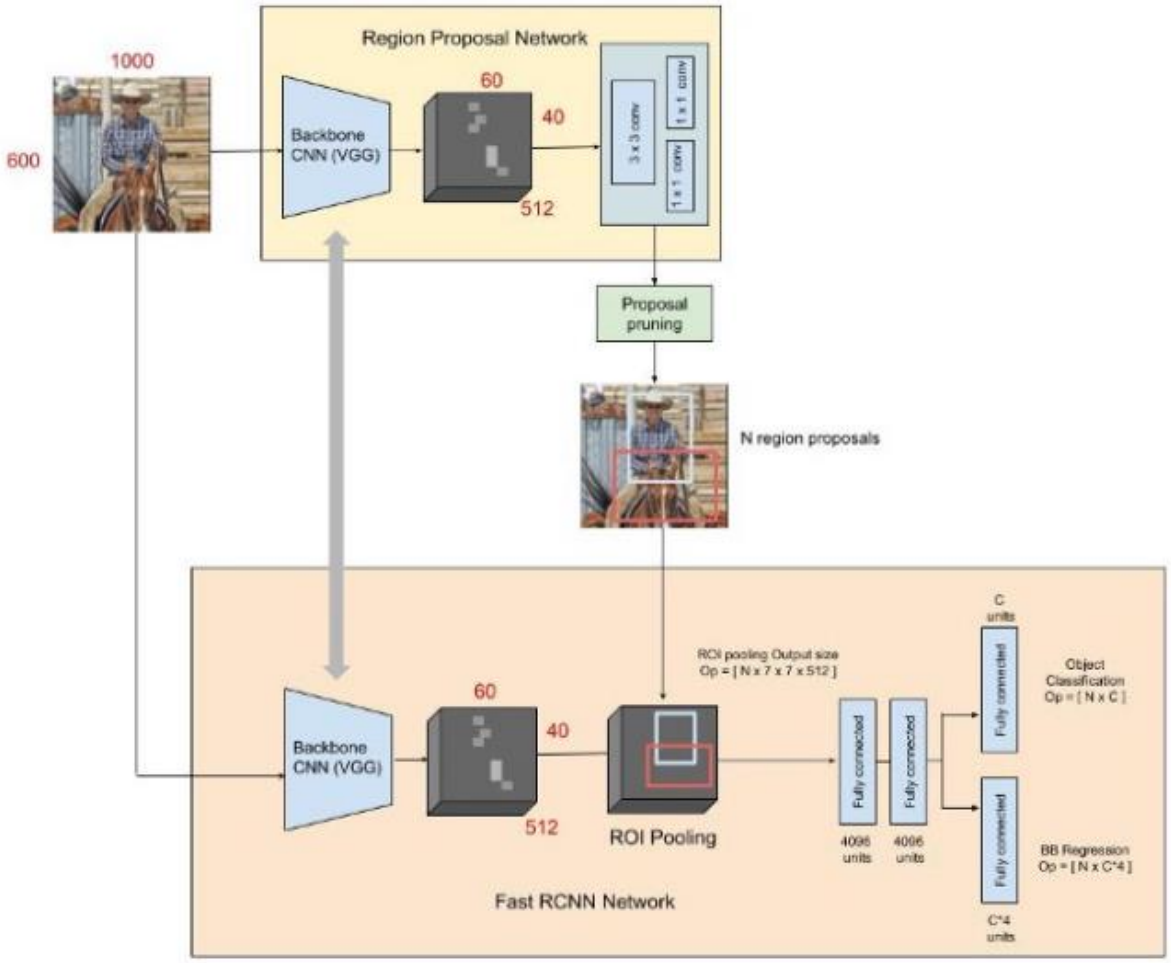
נראה כי קיימים סטי דאטה רבים באינטרנט שכל אחד מתעסק עם אובייקטים שונים. הצורך שלנו הוא זיהוי כדור טניס. קיימים סטי נתונים שונים ברשת , למשל COCO או VOC Pascal , אך למשימה שלנו של זיהוי אובייקט קטן ככדור טניס , נצטרך סט נתונים ייחודי התואם למשימה , כי ישנה שונות רבה בין גודל האובייקטים והמחלקות בין מה שדרוש לנו לבין אילו הקיימים שחיפשנו טרם ביצוע משימת זיהוי האובייקטים .

2.2. רשת מאומנת

קיימות רשתות מאומנות רבות באינטרנט, אך אנחנו צריכים רשתות שמצבעות זיהוי אובייקטים. רשתות כאלה קיימות בווריאציות שונות, אך ה-2 ארכיטקטורות הפופולאריות ביותר הן RCNN ו-YOLO.

2.2.1. RCNN

RCNN הוא ראשי תיבות של regional convolutional neural network, מה שרומז לנו על דרך הפעולה שלו. הרשת מחולקת ל-2 פעולות. זיהוי אזורי עניין, אזורים שאולי יהיה בתוכו אובייקט, וקלסיפיקציה. ארכיטקטורה זאת הייתה בין הראשונות שנתנה תוצאות טובות במשימת הזיהוי, ועברה איטרציות פיתוח רבות.



איור 5. מבנה *Faster-RCNN*

הרשת לוקחת את התמונה ומריצה אותה ב-2 ערוצים נפרדים. ערוץ אחד מזהה אזורי עיניין אפשריים בתוך התמונה, וערוץ שני אחראי לקלסיפיקציה של התמונה בתוך האזור עיניין שנבחר ע"י הערוץ הראשון. כך הרשת בעצם מריצה רשתות קלסיפיקציה פעמים רבות על אותה התמונה.

שיטה זו מוגבלת ע"י קצב מציאת אזורי העניין. תחילה הרשת הייתה מחפשת אזורי עניין ע"י אלגוריתמים קלאסיים, ולאט לאט פותחו גרסות מהירות יותר כמו fast RCNN ו – faster RCNN ששדרגו את האלגוריתם למציאת אזורי עניין והפיכה שלו לאלגוריתם לומד. רשת זו נותנת תוצאות ברמת דיוק גבוהה מאוד, אבל זמן העיבוד שלה ארוך מאוד עקב חזרה על אותה התמונה פעמים רבות.

2.2.2. YOLO

YOLO ראשי תיבות ל-You only look once מנסה לאחד את 2 הפעולות לפעולה אחת בו מתבצע הזיהוי וקלסיפיקציה ביחד. גישה זו מצמצמת את מספר הגישות אל התמונה וכך גם את זמן החישוב. הרשת משתמשת בנק' עוגן בתמונה וצופה את ה-BB (Bounding box) ע"י ניבוי של 4 ערכים. 2 הם מרכז ה-BB כהיסט מאותן נק' עיגון, 2 נוספות הם האורך והרוחב של ה-BB. הרשת מבצעת קלסיפיקציה לעומת כלל התמונה ולכל פיקסל נותנת הסתברות לכל מחלקה. הקלסיפיקציה הסבירה ביותר בתוך ה- BB נבחרת כמחלקה של אותו BB [5].

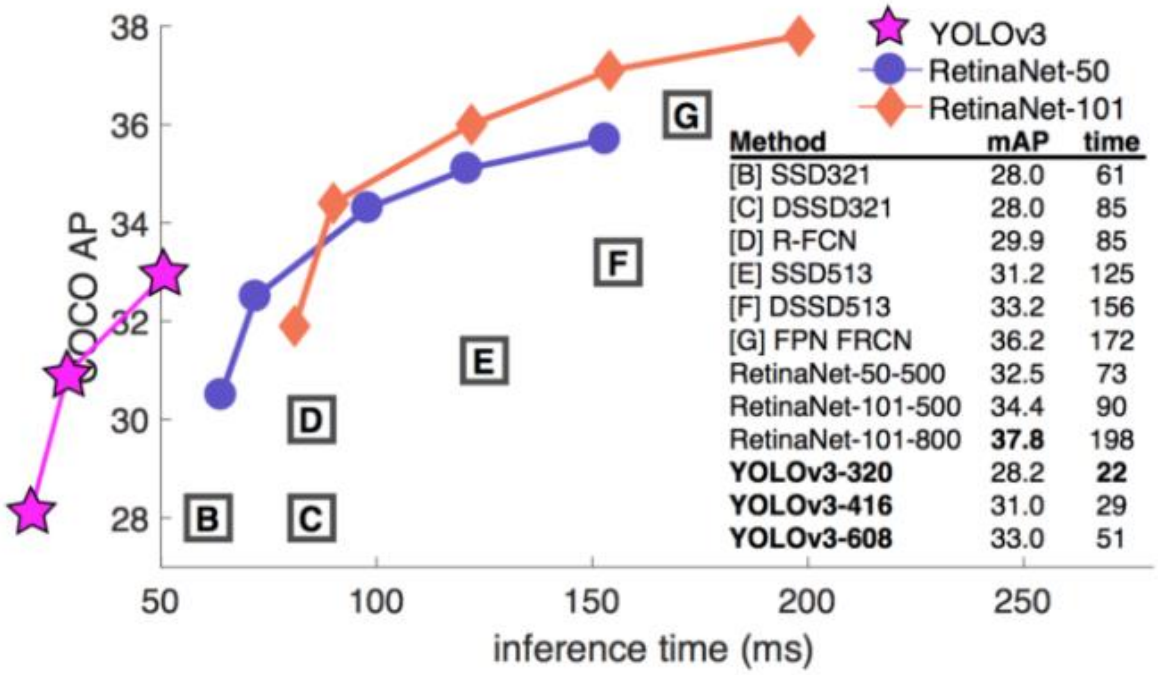
2.2.3. השוואה

נבדוק את 2 הרשתות אחת ביחס לשנייה.

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods							
Faster R-CNN+++ [3]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [6]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [4]	Inception-ResNet-v2 [19]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [18]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	**52.1**
One-stage methods							
YOLOv2 [13]	DarkNet-19 [13]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [9, 2]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [2]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [7]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [7]	ResNeXt-101-FPN	**40.8**	**61.1**	**44.1**	**24.1**	**44.2**	51.2
YOLOv3 608 × 608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

איור 6. השוואה של *AP* בין רשתות שונות *[7]*

נראה כי ל-YOLO יש תוצאות AP גרועות יותר על פני R-CNN, נבדוק איך זמני החישוב שונים בין הרשתות השונות.



איור 7. השוואה של מהירות בין רשתות שונות *[7]*

נראה כי למרות ההבדלים הקטנים במדדי mAP בין הרשתות השונות YOLO נותן תוצאות טובות ביחס לזמני חישוב, זה עקב המעבר היחיד על התמונה במקום מעברים רבים שרשתות אחרות מבצעות.

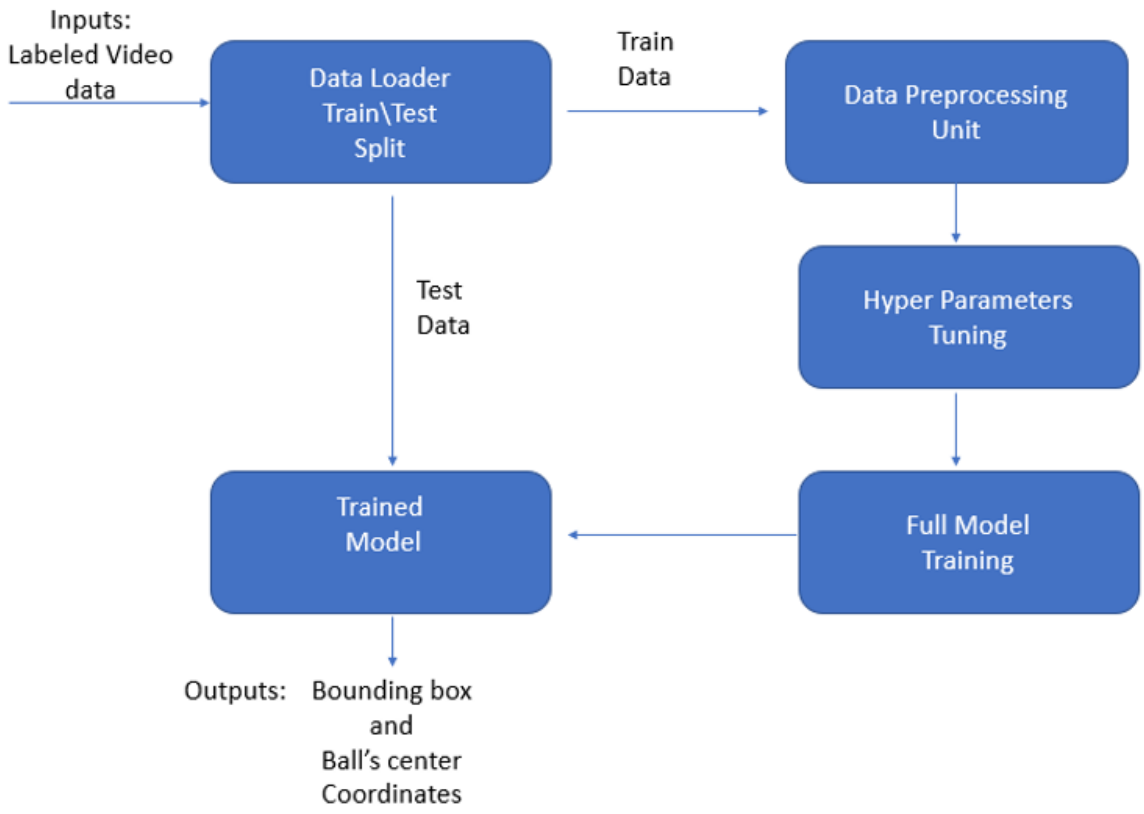
3. פתרון

3.1. סט אימון קיים

נתון לנו סט מידע של 7800 פריימים מסרטוני וידאו שונים. סט זה אינו מספיק בשביל לאמן רשת עמוקה מ"אפס" לכן השתמשנו בהעברת למידה. מצאנו סט הדאטה של COCO אשר מכיל מחלקה של כדורי ספורט ומכילה תמונות רבות ממשחקי טניס.

איור 8. דוגמא לתמונה מתוך סט המידע של *COCO*

אך תמונות אלו לא מספיקות לנו כי התמונות שם מצולמות ע"י מצלמה מקצועית מה שיוצר כדור גדול יותר ממנה שקיים אצלנו לרוב וגם חד יותר. אך זה סט טוב לאימון מקדים. הסט מידע שחברת Baseline סיפקה לנו היה מתויג ברובו בצורה די מדויקת. נשתמש בסט מידע שיש לנו לאמן רשת.



איור 9. סכמת בלוקים לאימון רשת

3.2. Yolov5

המודל yolov5 , יצא לאור בגרסתו הראשונית ב25 ביוני , בשנת 2020 , על ידי חברת Ultralytics .

כחלק מסקר הספרות בפרויקט , זו הרשת השנייה שבחנו והיא גם זו שנבחרה .

המודל , yolov5 , הוא מודל לביצוע object detection שהינו חלק מסדרת מודלים ממשפחה שנקראת yolo you only look once , שעקב ביצועיה רבים משתמשים בה לביצוע משימות הזיהוי.

כשאומרים ביצועים טובים הכוונה לדיוק וזמן inference.

כלומר ,נרצה מדד mAP כמה שיותר גבוה על ה dataset הנבחר , וזמן הסקה כמה שיותר קצר .

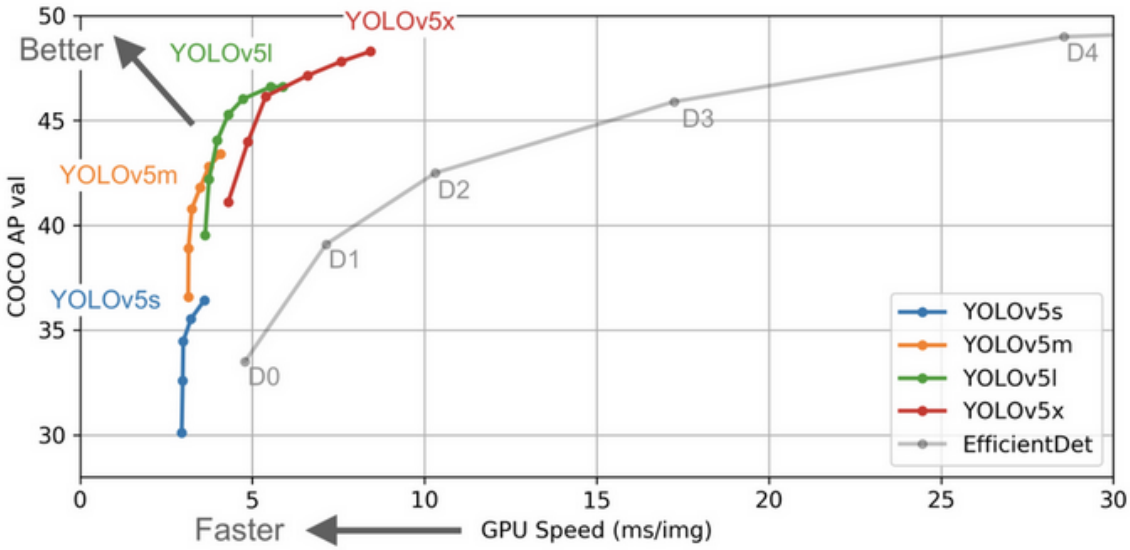
מודלים אחרים של object detection מבצעים תהליך שמבוסס על משימות מסובכות , כמו למשל שימוש רב בsliding window ובהמשך הזנה של הפלטים למסווגים (classifiers) שפועלים על מרווחי מקום שונים , בגדלים זהים , על כל התמונה , או על region proposals , במטרה לגנרט bounding boxes שבסבירות גבוהה יכילו אובייקטים ואז להזין אותם לרשתות קונבולוציה .

Yolo בניגוד אליהם מתייחס אל בעיית זיהוי האובייקטים כאל בעיית רגרסיה , שפתרונה נותן בו זמנית את הקור' של הbounding boxes ואת ערך הסתברות המחלקה עבור אותן תיבות .

נק' המפתח היא שמשפחת מודלי yolo מחלקת את התמונה למספר גרידים , שגורמים להאצה בזמן הריצה ובדיוק.

יש לציין , שישנו trade off בבחירה בשימוש בyolo לעומת מודלים אחרים , קיימים כיום מודלים מדויקים יותר מאילו של yolo , אך איטיים יותר . לכן על הלקוח להחליט כיצד לבצע חלוקה שתתאים לצרכיו .

דוג' לכך הוא מודל ממשפחת rcnn שבחנו קודם לכן , שהינו מדויק יותר .



איור 10. גרף ביצועים של רשתות מסוג *yolov5* שונות

מודל object detector תפקידו ליצור מאפיינים , features , מתמונות הקלט , ולאחר מכן להזין את ה features שיצר דרך מערכת פרדיקציה , שתפקידה לצייר bounding boxes סביב האובייקט , ולחזות את הסיווג שלו למחלקה . משפחת yolo , הייתה הראשונה לשלב בין הפרדיקציה של הbounding boxes עם הפרדיקציה של הסיווג למחלקה במערכת קצה לקצה , end to end.

המודלים במשפחה זו מורכבים משלושה חלקים עיקריים :

א. Backbone – רשת CNN שאוגרת מאפייני תמונות על ידי שימוש בגרעינים (מסננים) שונים. לרב משתמשים בארכיטקטורות state of the art שבקדמת המחקר .

למשל השימוש בvgg16 , שהוכח כיעיל במיוחד עקב residual blocks וskip connections .

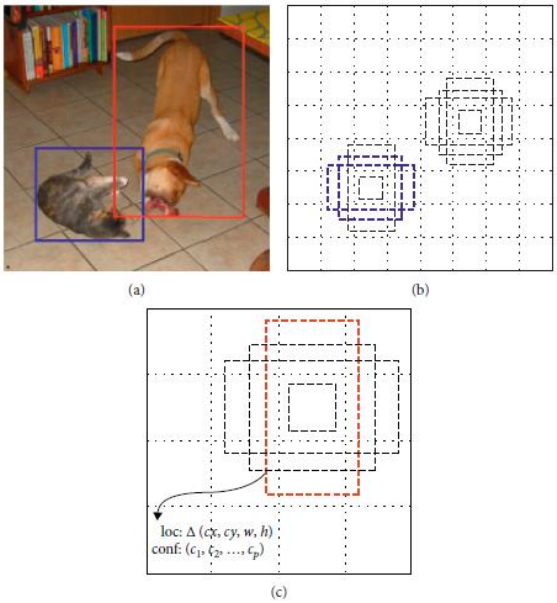
ב. Neck – סדרה של שכבות נוירונים שתפקידן לערבב ולשלב את מאפייני התמונות שנוצרו , features ולהעבירם לשלב הפרדיקציה .

ג. Head – לוקחת את המאפיינים שה Neck יצרה ומבצעת שלבים שבסופם מסיקה bounding box ו class.

התהליך שבו yolo פועלת הוא הינו כאמור חלוקת התמונה לgrid בגודל SxS , כאשר עבור כל גריד משוערכים מספר מסוים של anchor boxes , מספר ה anchor boxes שונה וגדל עם התפתחות משפחת yolo .

במהלך האימון מחושב הcost , ולבסוף המודל בוחר את הanchor box כbounding box הנבחרת בעלת הציון המשוקלל הטוב ביותר , מבחינת מדדי confidence , localization , objections .

בעמוד הבא מצורפת תמונה להמחשה .



איור 11. תמונה של *GT* וניבוי של *BB* שונים

4. שלב העיבוד המקדים – pre processing

בעבודות שמשתמשים בהן במאגרי נתונים , אותות שמע , מסמכי טקסט , תמונות וכו' , המידע שמגיע למשתמש מגיע לרב בתצורה לא מעובדת ומוכנה להרצה במערכת , לכן יש צורך בעיבוד מקדים .

דוג' לעיבוד מקדים הינן ניקוי ערכי outliers , המרה ליחידות / ביטויים רצויים , השלמת נתונים חסרים וכו'

במסגרת הפרויקט , עבדנו עם סוג נתונים שהינו תמונות דו ממדיות שצולמה במצלמה ברזולוציה גבוהה .

המודל שאיתו עבדנו ושמומש בPytorch כפה עלינו מספר דרישות שהיה עלינו לבצע על מאגר הנתונים שלנו.

ביצענו נורמליזציה כך שכל batch שהוזן למערכת היה בעל ממוצע 0 וסטיית תקן 1 , בין היתר הוכח שזה עשוי להאיץ את תהליך האימון ויכול למנוע את בעיית ה exploding gradients .

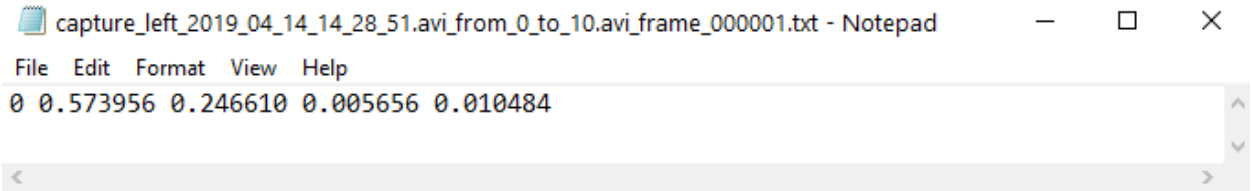
התאמנו את פורמט התמונות לתצורת PIL בחלק מהתמונות שלא היו בפורמט המתאים . ובנוסף המרנו אותם לטיפוס tensor שPytorch framework כופה על המודל הנוכחי .

לפני הרצה בדקנו כי כל התיוגים נכונים ולא מופיעים אובייקטים ממחלקות אחרות (היינו צריכים לנקות מקרים של תציוג של בן אדם), בתחילת הדרך התאמנו פורמטים של תיוגים, מפורמט pascal לפורמט yolo ויצירת קבצי CSV שמכילים את כל התיוגים הקיימים ביחד (פריימים רקים מתוגרמים לשורות ריקות). כל הפעולות האלו בוצעו ע"י סקריפט בפייתון.

5. אימון המודל

כמקובל בעולם הלמידה העמוקה חילקנו את סט הנתונים שלנו לסט אימון ולידציה ומבחן .

גודל סט האימון היה 84.6% מכלל הנתונים , סט הוולידציה שהיה זהה לסט האימון, מכיוון שהסט נתנון שלנו קטן מידי ורצינו להשאיר מספיק תמונות לסט מבחן. סט המבחן מהווה 15.4% הנותרים. סט המידע מורכב מ-13 סרטוני וידאו שפורקו לפריימים, לכן בשביל למנוע זליגה של סט האימון וסט המבחן לקחנו 11 סרטונים לסט האימון ועוד 2 סרטונים לסט המבחן. סט האימון גם כולל תנאי תאורה שונים, משחק ביום ומשחק מואר בלילה, משחקים מזוויות שונות וצבעי מגרש שונים. סט הנתונים שלנו הגיע בפורמט של Yolo שהוא פריים מהוידאו כקובץ *.png וקובץ טקסט בעל שם זהה וסיומת txt.



איור 12. דוגמא לקובץ תיוג

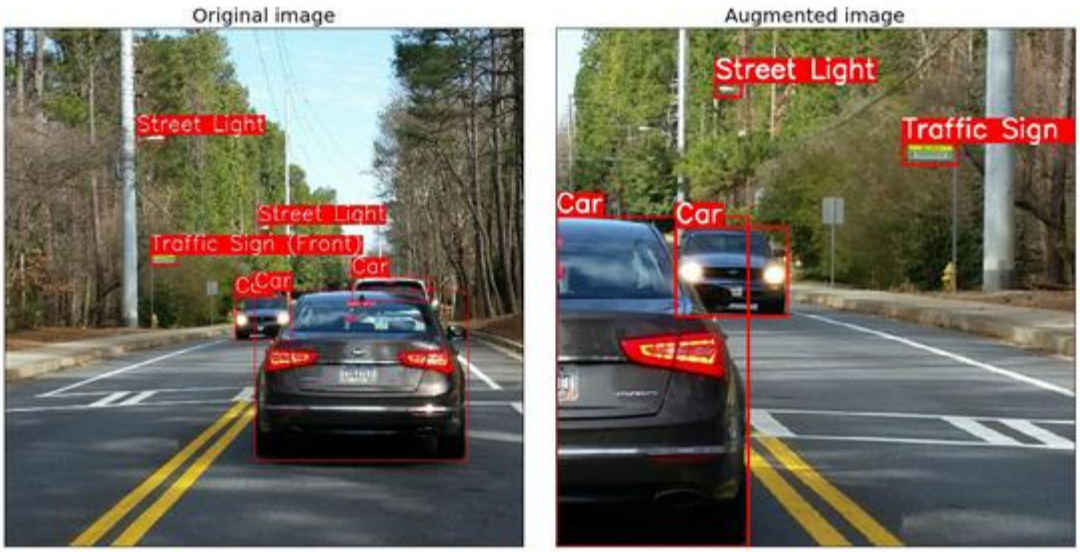
קובץ *.txt הוא קובץ תיוגים שמותאם לקובץ תמונה מאותו שם, הוא כולל את המידע הבא (בסדר משמאל לימין) [{label} , {bb x coordinate} , {bb y coordinate{} , {bb width} , {bb height}] כאשר התגית היא מספר מזוהה עם המחלקה שלנו בקובץ ההגדרות, והערכים של הגדלים הם יחסיים לגודל התמונה. בנוסף , על מנת לשפר את ביצועי המודל , בסט האימון והווליד‌ציה לאחר מעבר שלם על הסט , epoch , הוגרלו מחדש הmini batches , כאשר בכל אחד שכזה דגמנו בצורה אקראית , מהתפלגות אחידה , תמונה מסט הנתונים.

פעמים רבות סט האימון יכול לגרם לאימון יתר, במיוחד כאשר הסט אימון כל כך קטן, לכן מרבים להשתמש ב data augmentation. בפעולה זו שאנו מפעילים על תמונות הקלט שאנו מזינים למודל , אנחנו בעצם 'מעוותים' את תמונות הכניסה כדי שהמודל ילמד דוגמאות מגוונות ויתקשה להגיע למצב של התאמת יתר , overfitting, שתפגע בביצועים בזמן המבחן או ב real time.

השתמשנו בספרייה שנקראת albumentations שמספקת API מאוד נח להפעלת הטרנספורמציות .

בנוסף עקב כך שמדובר על Object detection מלבד הפעלת טרנספורמציות על תמונות הקלט , תמונת הקלט מכילה בנוסף גם bounding boxes ויש לזכור לבצע את הטרנספורמציה המתאימה גם עליהן – כלומר הזזה/scale וכו' במידה המתאימה גם עליהן . עבור שינוי צבע אין צורך לשנות .

איור 13. הדגמה של עיוותים על תמונה



איור 14. עיוות על תמונת קלט למערכת זיהוי אובייקטים

הרצנו את האימון עם הרשת המאומנת של yolov5, הרצנו את הפקודה הבאה בשביל אימון הרשת.

```
python train.py --img 736 --batch 16 --epochs 500 --data ball.yaml --cfg yolov5s.yaml --cache --workers 1
```

הפקודה רצה מתוך הספריה של yolov5, הגדרנו לאימון לרוץ על תמונות בגודל 736×736. כל באטצ' הינו בגודל של 16 תמונות בשביל לעדכן לעיתים יותר דחופות מהרגיל את המשקלים, כי הסט קטן וחילוק גס מידי יפגע בלימוד. הדאטה כולל את מיקום התמונות ותיוג ורשימת מחלקות. קונפיג כולל את הארכיטקטורה של הרשת, ספציפית yolov5s שהרשת הדיפולטית הקלה והמהירה ביותר בין הווריאציות, בחרנו לעבוד איתה והגדרנו את השכבה האחרונה כ-FC לניורון יחיד. הוספנו תגית של טעינה של המידע במטמון של המעבד וכך לזרז את תהליך האימון ע"י המנעות של משיכה של התמונות מהזיכרון. שימוש בתגית של עובדים בשביל להמנע מבעיות עבודה עם הדוקר. האימון כולו הורץ למשך 500 איפוקים , ועל כרטיסי gpu שבמחשבי המעבדה . להלן ההיפר-פרמטרים שנבחרו

```
hyp = {'lr0': 0.01,  # initial learning rate (SGD=1E-2, Adam=1E-3)
       'momentum': 0.937,  # SGD momentum
       'weight_decay': 5e-4,  # optimizer weight decay
       'giou': 0.05,  # giou loss gain
       'cls': 0.58,  # cls loss gain
       'cls_pw': 1.0,  # cls BCELoss positive_weight
       'obj': 1.0,  # obj loss gain (*=img_size/320 if img_size != 320)
       'obj_pw': 1.0,  # obj BCELoss positive_weight
       'iou_t': 0.20,  # iou training threshold
       'anchor_t': 4.0,  # anchor-multiple threshold
       'fl_gamma': 0.0,  # focal loss gamma (efficientDet default is gamma=1.5)
       'hsv_h': 0.014,  # image HSV-Hue augmentation (fraction)
       'hsv_s': 0.68,  # image HSV-Saturation augmentation (fraction)
       'hsv_v': 0.36,  # image HSV-Value augmentation (fraction)
       'degrees': 0.0,  # image rotation (+/- deg)
       'translate': 0.0,  # image translation (+/- fraction)
       'scale': 0.5,  # image scale (+/- gain)
       'shear': 0.0}  # image shear (+/- deg)
```

איור 15. רשימת היפרפרמטרים של המודל

הרשת רצה עם אופטימייזר של SGD בעל רכיב מומנטום 0.937 קצב למידה ראשוני של 0.01, עקב העובדה שהרשת אינה עמוקה כי רק השכבה האחרונה פתוחה לשינויים חישוב הגרדיאנטים מהיר מאוד.

6. תוצאות

נתבונן בביצועי הרשת לפני האימון על הסט הקטן שלנו.

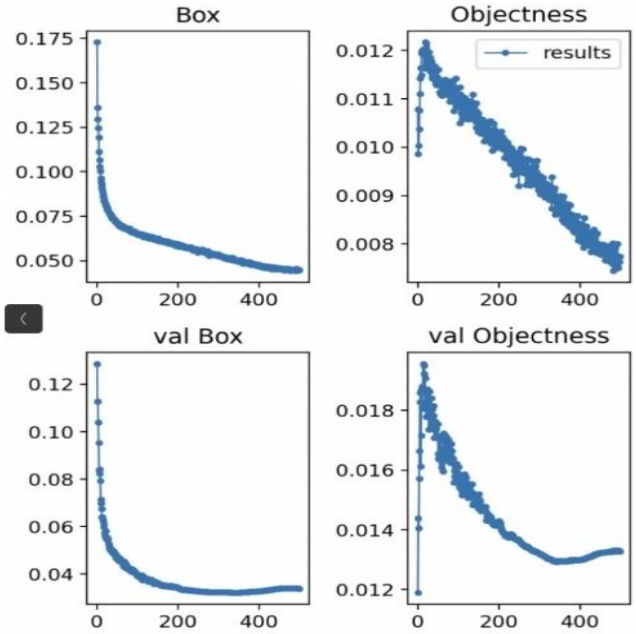
איור 16. תוצאות ניבוי מול *GT* ברשת מאומנת על *COCO* בלבד

נראה כי הרשת לא נותנת לנו תוצאות כלל, לעומת הרשת בסוף תהליך האימון שנציג את התוצאות שלה עכשיו.

נסתכל וננתח על תוצאות ההרצה הטובה ביותר מבחינת מדד mAP שקיבלנו וננתח אותה .

בכל הגרפים הציר המאוזן מציין את מספר הepoch.בשורה העליונה תוצאות על סט האימון ובתחתונה הווליידציה .

מבחינת מדדי פונקציית המחיר , אנו רואים שככל שאנו מאמנים את המודל , המודל מצליח להוריד את השגיאה הריבועית שמחושבת קור' הthe grount truth וקור' המשוערכות על ידי פלט ה bounding box שהמודל חזה .

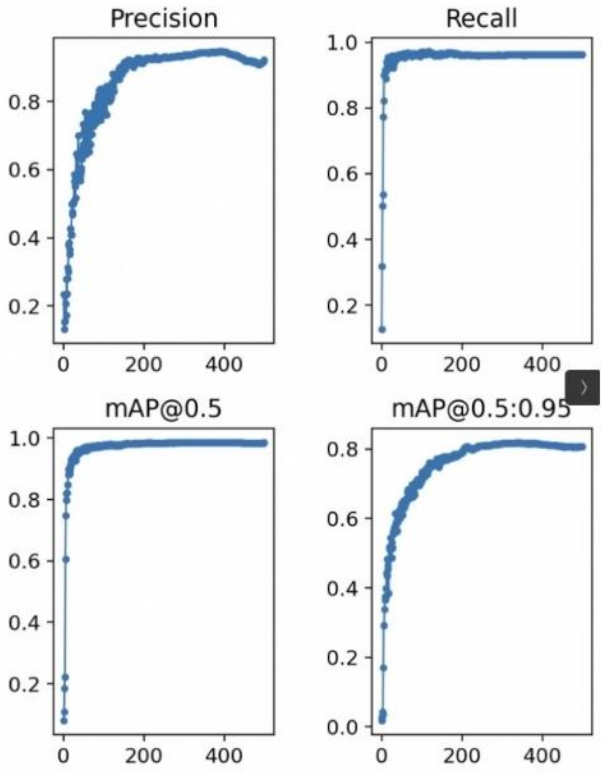


איור 17. פונקציות מחיר לאורך האימון

מבחינת תוצאות על מדדי הדיוק שנבחרו , ניתן לראות שבסט האימון הprecision עולה ככל שמתקדמים באימון , לכן המשקולות מתעדכנות בצורה נכונה וכאשר אנחנו מבצעים סיווג לאובייקט מסויים בסבירות גבוהה של כמעט תשעים אחוז הסיווג הינו נכון .

הגענו גם לתוצאות מרשימות במדד הrecall , אנו רואים ערך מאוד גבוהה לקראת סוף האימון , מה שאומר שכאשר ישנו אובייקט בתמונה אנחנו לרב נבצע חיזוי ונמנע מTN .

בנוסף על סט הולידציה ניתן לראות שעבור ערכי סף שונים של giou הערך המשקולל , mAP , עולה ככל שאנחנו מתקדמים בתהליך האימון ומגיע לתוצאות גבוהות .

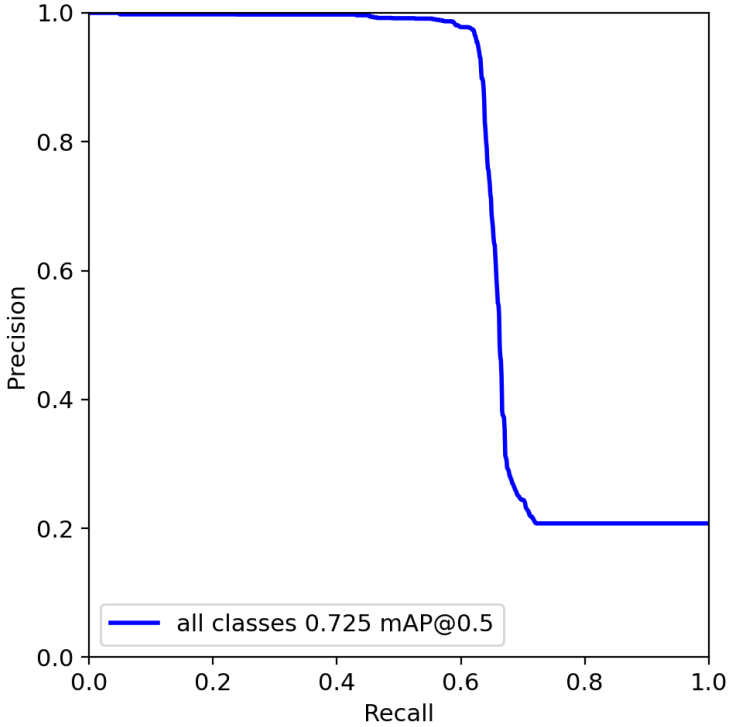


איור 18. מדדי דיוק לאורך האימון

נבדוק את התוצאות שקיבלנו במוצא של הרשת על הסט מבחן

איור 19. תוצאות ניבוי מול *GT* ברשת מאומנת *COCO* ודאטה יעודי

נראה כי התהליך מצליח לזהות כדורים הרבה יותר טוב מלפני האימון.



איור 20. גרף *precision/recall* של סט המבחן אחרי אימון

נראה כי ביצועי ה-recall אינם נהדים, אך בצפיה בסרטון של התוצאה אפשר להסיק כי קיים תיוג יתר, כדורים שבקושי ניתן להבחין בצידי המגרש מתוייגים ואין לנו את הציפייה או צורך לזהות אותם.

איור 21. תוצאת מבחן מרשת מאומנת

נראה תוצאה טובה גם בתנאי תאורה רעיים שרוב סט האימון היה בתאורת יום.

איור 22. תוצר רשת עם *FP*

אפשר לראות כי העקב של השחקן זוהה ככדור בפריימים בודדים, אפשר להסיק מתוך התנהגות הרשת כי כאשר קיים אובייקט שהוא בבירור כדור הוא גורם לסביבה שלו להיות גם יותר בטוחה בהמצאות כדור.

איור 23. תוצר רשת עם *FN*

שוב ניתן לראות כי הכדור הנמצא בפינה מזוהה רק כאשר הוא בקרבת הכדור הברור. והכדור שנמצא בפינה הימינית של התמונה אינו מזוהה ככדור והוא דוגמא ל-FN.

7. סיכום

במהלך הפרויקט , עקב סקר הספרות ואימון המודל למדנו על התהליך שבו מערכות לזיהוי אובייקטים פועלות , על שיטות שונות שמשפחות מודלים שונות מפעילות , ועל השיקולים ההנדסיים הכרוכים בנושא .

זיהוי אובייקטים קטנים –

במהלך הפרויקט אנו עסקנו בזיהוי כדור טניס , שמבחינת רזולוציה מהתמונות שניתנו לנו הינו אובייקט קטן , כלומר מכיל אחוז קטן מכלל גודל התמונה . ניתן להגיד שלמרות ההתקדמות בשנים האחרונות והשיפור ביכולת לבצע זיהוי אובייקטים , ישנו פער בין זיהוי אובייקטים בגודל קטן לעומת זיהוי אובייקטים בגודל בינוני/גדול .

בפרט, כשבחנו תוצאות על סט הנתונים של COCO התגלו פערים משמעותיים .

הפתרון לדעתנו יהיה שהקהילה החוקרת תשקיע בלייצר מספר סטי נתונים חדשים ,שיהיה בהם חלוקה מאוזנת יותר של אובייקטים לזיהוי בגדלים שונים מקטן ועד גדול .

דיוק מול מהירות –

קיימים כיום לא מעט מודלים לביצוע זיהוי אובייקטים ,

המודל שאנו בחנו , yolov5 הינו מודל end to end , שמגנרט עבור כל חלק מהתמונה anchor boxes במקום להציע אותן דרך הכנסת מאפייני מפתח לרשתות קונבולוציה . עקב כך מתקצר הזמן לביצוע זיהוי אובייקט לעומת משפחת מודלי rcnn למשל . החיסרון הינו שאנו משלמים מחיר בדיוק נמוך יותר .

המסקנה שלנו היא שיש trade off ועל הלקוח להחליט כיצד לבצע את החלוקה המתאימה .

אם ניקח מערכת לזיהוי כלי נשק בשדה תעופה , נרצה דיוק מירבי על מנת להציל חיי אדם , לעומת פרויקט כמו שלנו נוכל לבצע הנחה ביכולות הדיוק , על מנת שנעקוב בצורה מהירה יותר אחרי כדור הטניס .

סט הנתונים –

במהלך השלב הראשוני של הפרויקט סט הנתונים שלנו היה מצומצם , כתוצאה מכך המודל שאומן ביצע התאמת יתר לתמונות שראה ויכולת ההכללה שלו כשהוא נבחן על סרטונים שהוא לא ראה היו לא מספקות , המון פעמים הוא ביצע FN , כלומר שהיה כדור טניס בפריים אך הוא לא זיהה אותו .

בהמשך הסט נתונים שלנו גדל , כתוצאה מכך לאחר האימון , הבחנו בתוצאות טובות מאוד , אחוזי הconfidence עלו ומדד הmAP הגיע לרמה מאוד גבוהה של מעל 90% .

המסקנה היא החשיבות של בחירת סט הנתונים שתואם למשימה שנבחרה , ולדאוג שהוא יהיה עשיר בנתונים ממגוון מצבים וסיטואציות . סט נתונים שכזה יגרום להפחתת התאמת היתר ולשיפור ביכולת ההכללה .

8. משימות להמשך

בפרויקט זה ביצענו מימוש ראשוני לזיהוי כדור הטניס במשחק טניס , ישנם עוד אלמנטים רבים שניתן להוסיף על מנת לשפר את חווית האימון של שחקן טניס , וההתקדמות הענפה בעולם הלמידה העמוקה מצביעה על כך 'שהשמיים הם הגבול' .

- ביצוע זיהוי גבולות המגרש בעזרת התמרת Hough ולממש מנגנון שמזהה האם כדור שהוכה הינו חוקי או לא
- קיימים מודלים מאומנים של pose estimation , אפשר להשתמש בהם כדי למצוא את המנח של כל שחקן , ולפי הזוויות בין המחבט למפרקיו , לזהות איזו מכה הוא חובט ובהמשך להוציא סטטיסטיקה מתאימה לשחקן .
- שימוש ב reinforcment learning על מנת לאמן סוכן ללמוד את סגנון היריב שנגדו ישחק , והסקה האם לאחר אימון מוצלח של הסוכן , הוא יוכל לגבור על יריב ברמות משחק שונות .
- שימוש המיקומי הכדורים שזוהו והדאטה הרב הקיים וליצור משם רשת לומדת להשלמת מיקום הכדור בין פריימים שלא היה זיהוי, ובכך ליצור רציפות.

9. רשימת מקורות

[1] Available: .[מקוון] ”,Stanford cs231 lecture 8“ .http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

[2] ./Available: https://giou.stanford.edu .[מקוון] ”,GIoU“

[3] Available: https://jonathan-hui.medium.com/map-mean- .[מקוון] ”,mAP for Object detection“ ..average-precision-for-object-detection-4

[4] Available: .[מקוון] ”,Transfer learning for compute vision“ .https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html

[5] .”A. F. Joseph Redmon, “YoLov3: An incremental Improvement

[6] .Available: https://github.com/ultralytics/yolov5 .[מקוון] ”,ultralytics/yolov5“

[7] Qishuo Lu, Chonghua Liu, Zhuqing Jiang, Aidong Men, Bo Yang, "G-CNN: Object Detection via Grid Convolutional Neural Network", *Access IEEE*, vol. 5, pp. 24023-24031, 2017

[8] Wenming Cao, Jianhe Yuan, Zhihai He, Zhi Zhang, Zhiquan He, "Fast Deep Neural Networks With Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection", *Access IEEE*, vol. 6, pp. 8990-8999, 2018.

[9] Hung-Wei Hsu, Jian-Jiun Ding, "FasterMDNet: Learning model adaptation by RNN in tracking-by-detection based visual tracking", *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) 2017*, pp. 657-660, 2017.

[10] Lei Huang, Weijia Xu, Si Liu, Venktesh Pandey, Natalia Ruiz Juri, "Enabling versatile analysis of large scale traffic video data with deep learning and HiveQL", *Big Data (Big Data) 2017 IEEE International Conference on*, pp. 1153-1162, 2017.