ראייה ממוחשבת בחדר ניתוח – 097222

דו"ח תרגיל בית 1

## מגישים:

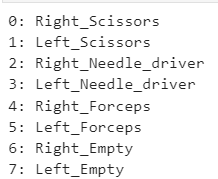
## עמית ניסן – 207108820

## גבריאל גוזל – 302770573

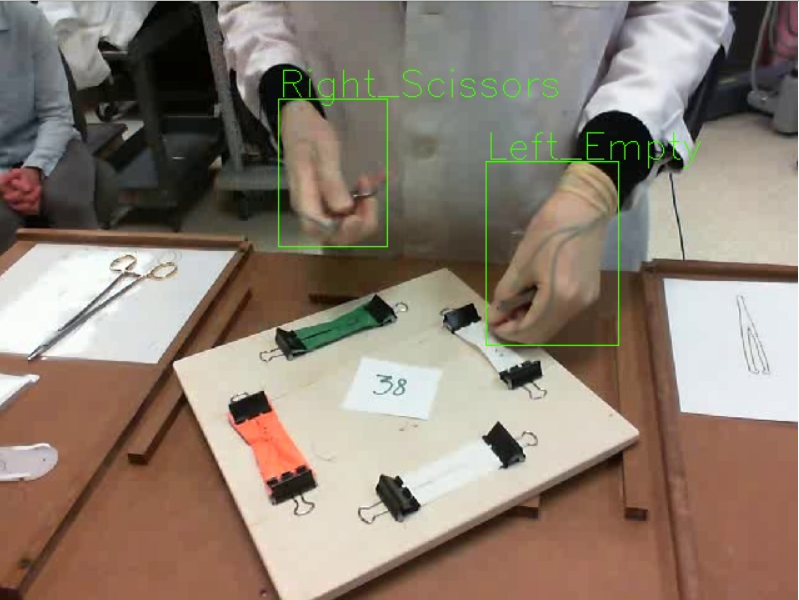
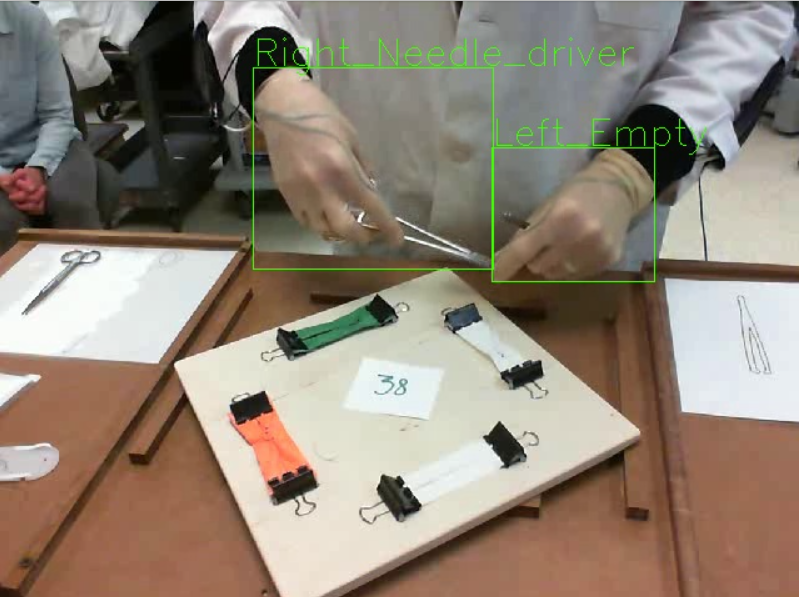
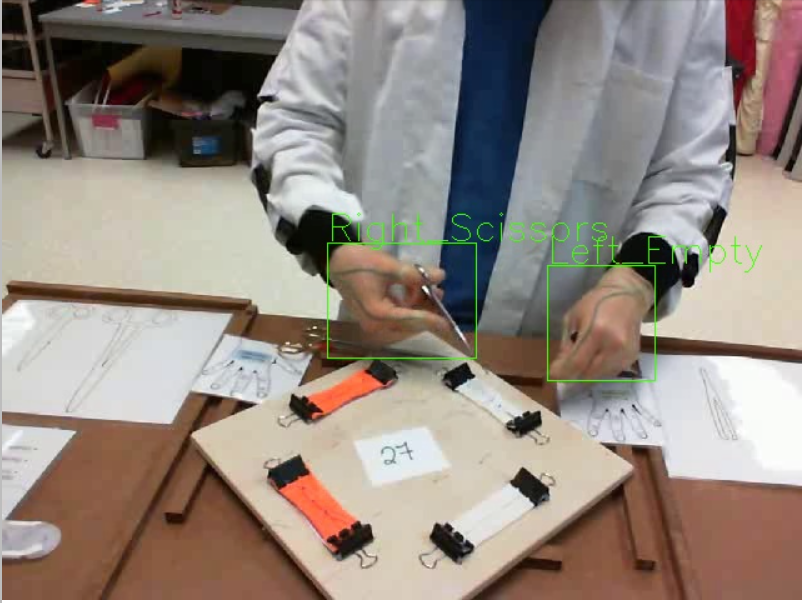
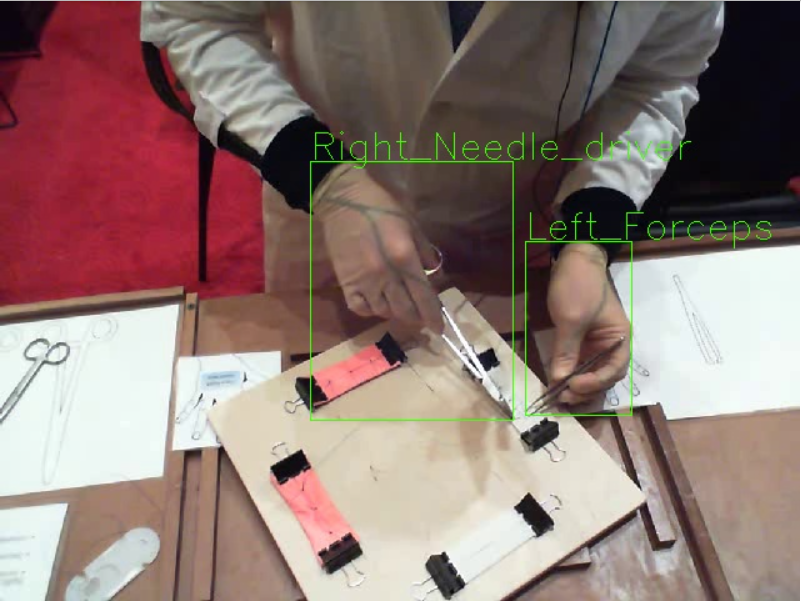
# Exploratory Data Analysis

# Visualization of some images

ה-Dataset עליו נעבוד בתרגיל מורכב מתמונות שנלקחו מסרטונים שצולמו בזמן שרופאים ביצעו תפירה של נייר טישו חתוך במרכז, וחתיכת בלון חתוך באמצע במטרה לדמות תפירה של רקמה עדינה יותר ופחות בהתאמה (איור 2 תמונות מתוך מאגר התמונות**)**. במהלך סרטונים אלו, הרופאים עשו שימוש במספר אביזרים לטובת התפירה: מספריים (scissors), מכשיר להחזקת המחט (needle driver) ומלקחיים (forceps). הקטלוג של ה-data מחלק אותו למעשה לשמונה קטגוריות שונות, ארבע לכל יד: ימין (שמאל) ריקה, ימין (שמאל) עם מספריים, ימין (שמאל) עם מחזיק מחט וימין (שמאל) עם מלקחיים. לכל פריים בסרטון, ישנו קובץ טקסט המכיל את הקטלוג לכל יד ומיקום המלבן המכיל את היד עם (בלי) האביזר במרחב הפריים. הקטגוריות השונות ממסופרות כפי שמוצג באיור 1 מספור הקטגוריות השונות במאגר כפי שנתון.



איור 1 מספור הקטגוריות השונות במאגר כפי שנתון

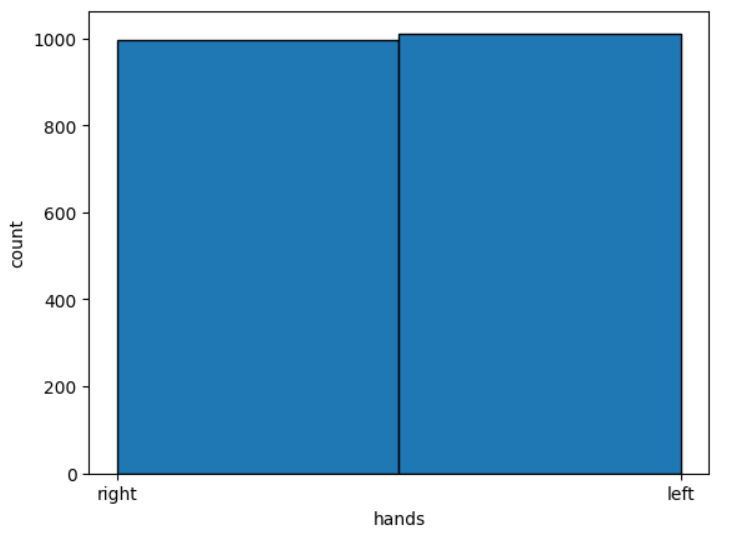


איור 2 תמונות מתוך מאגר התמונות עם הקטלוג האמיתי

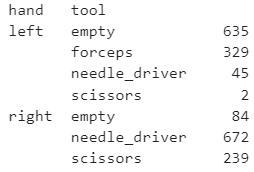
# Insights from simply “looking” at the data

תחילה בדקנו את כמות התמונות לכל יד (איור 3 מספר תמונות עבור כל יד). קיבלנו כי יש 1011 תמונות עבור יד שמאל, ו-995 עבור יד ימין. לכאורה חלוקה יחסית הוגנת בין שתי הידיים.

איור 3 מספר תמונות עבור כל יד



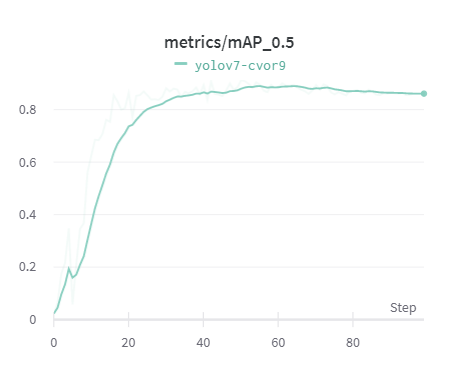
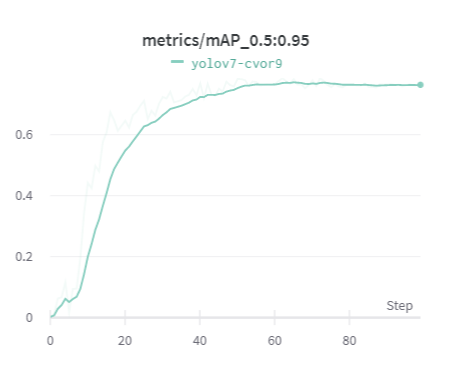
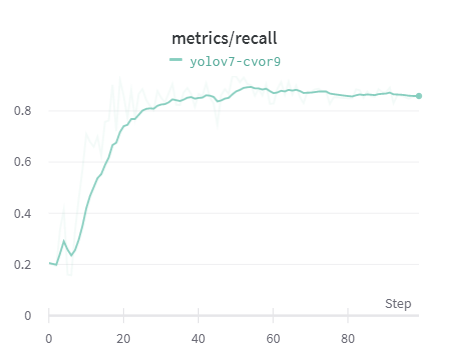
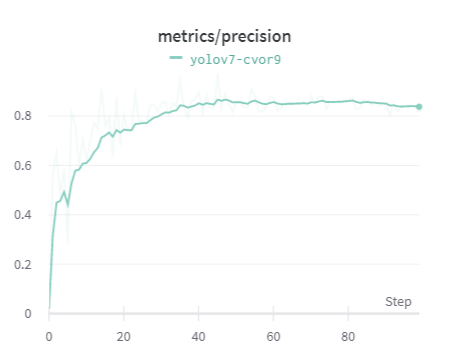
בשלב זה בדקנו את כמות התמונות שיש לנו לכל קטגוריה (איור 4 כמות התמונות לכל קטגוריה). תחילה ניתן לראות כי אין תמונות עבור קטגוריית right forceps וכן יש מעט תמונות של left scissors. ניתן להסיק מכך שרוב הסובייקטים שצולמו במאגר הינם דומיננטיים ביד ימין (אנו מניחים שאנשים דומיננטיים ביד ימין ישתמשו במספריים עם היד החזקה שלהם). מכך אנו גם מסיקים כי ה-needle driver הינו מכשיר עזר שמשתמשים בו במקביל למכשיר אחר ולכן יש בו תמונות רק עבור יד שמאל.



איור 4 כמות התמונות לכל קטגוריה

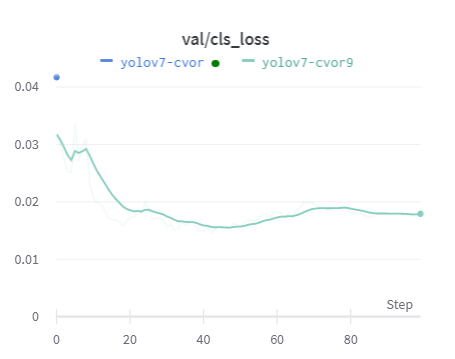
# Experiments

על מנת לעמוד במשימות התרגיל, אימנו מודל [YOLOv7](https://github.com/WongKinYiu/yolov7), אלגוריתם עדכני לזיהוי אובייקטים [1], בעזרת חלוקת מאגר התמונות ל-valid, train ו-test . הרצנו אותו עם batch size = 8, למשך epochs = 100 (סף עבור ה-IoU = 0.2, ו-SGD כ-optimizer) וקיבלנו את התוצאות המוצגות באיור 5 תוצאות האימון הראשוני על [YOLOv7](https://github.com/WongKinYiu/yolov7).



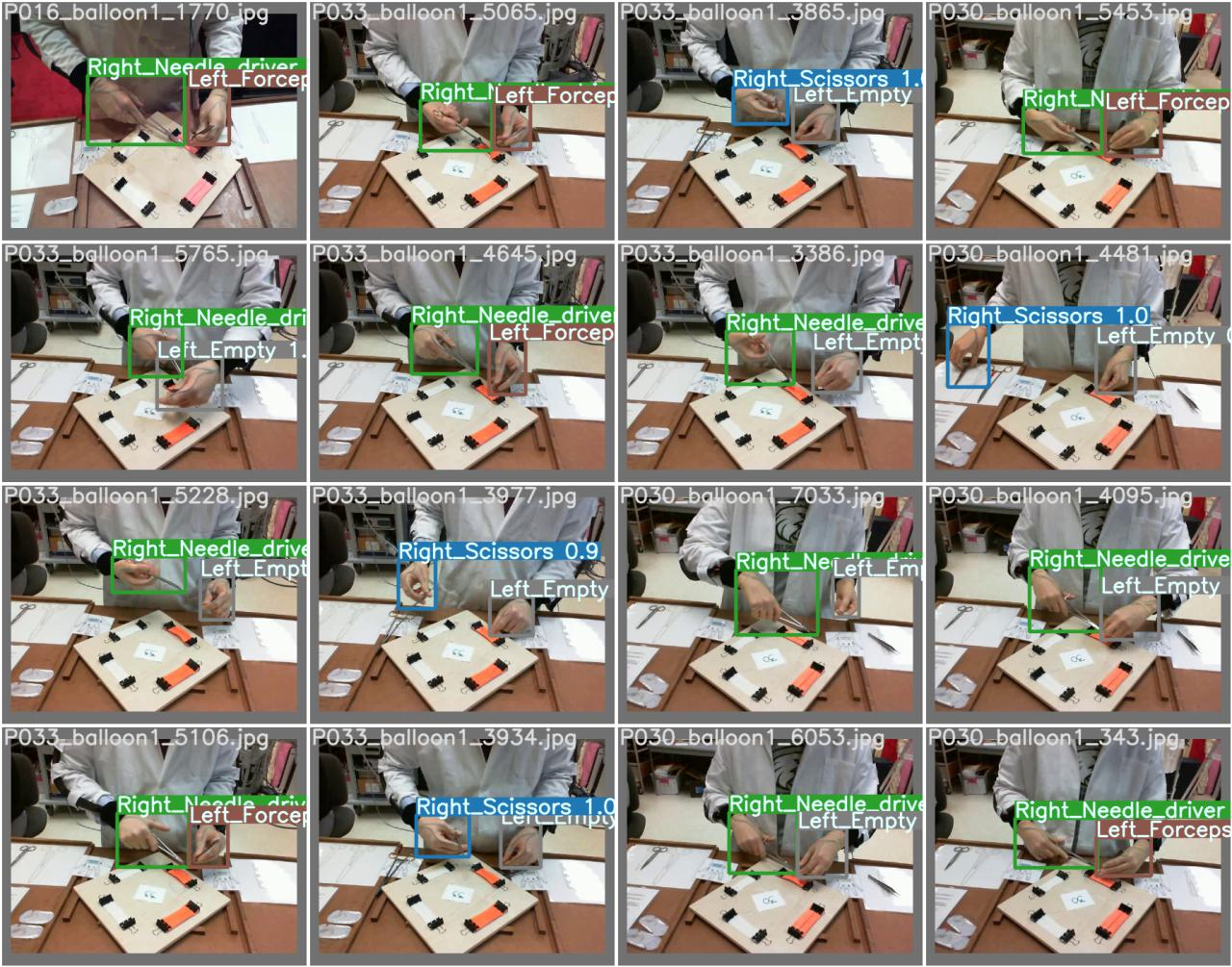
איור 5 תוצאות האימון הראשוני

מבחינת loss על סט הtest וה-valid, ניתן לראות כי הגדלת כמות ה-epochs כנראה לא תשנה באופן משמעותי את התוצאות שלנו כפי שמשתקף באיור 6 loss עבור הסטים של test ו-validation.



איור 7 loss עבור הסטים של test ו-validation

איור 6 תוצאות על סט הולדיציה



# Adam optimizer

# IoU Threshold change

# Tool usage evaluation on test set

כאשר בדקנו את הביצועים של המודל שלנו על ה-test, קיבלנו את התוצאות הבאות:

**Label precision recall f1-score support**

0 0.80 0.86 0.83 42

1 0.00 0.00 0.00 1

2 0.95 0.92 0.93 137

3 0.60 0.30 0.40 10

4 0.00 0.00 0.00 0

5 0.75 0.97 0.85 62

6 0.87 0.65 0.74 20

7 0.93 0.89 0.91 126

**micro avg** 0.88 0.88 0.88 398

**macro avg** 0.61 0.57 0.58 398

**weighted avg** 0.88 0.88 0.88 398

**f1 macro:** 0.66, **Accuracy:** 0.879

כאשר עמודת ה-support מייצגת כמה דוגמאות יש לכל קטגוריה במאגר מידע. ניתן לראות שעבור פחות מכ- 100 דוגמאות, התוצאות יורדות מתחת ל-89%.

----------------------------------------------------------------

**Label | AP@0.25 | AP@0.5 | AP@0.75 |Support**

----------------------------------------------------------------

Left\_Empty | 0.960157 | 0.960157 | 0.945975 | 126

Left\_Forceps | 0.967222 | 0.967222 | 0.967222 | 62

Left\_Needle\_driver | 0.500000 | 0.500000 | 0.400000 | 10

Left\_Scissors | 0.000000 | 0.000000 | 0.000000 | 1

Right\_Empty | 0.537762 | 0.537762 | 0.537762 | 20

Right\_Needle\_driver | 0.960336 | 0.951323 | 0.944032 | 137

Right\_Scissors | 0.891657 | 0.891657 | 0.891657 | 42

----------------------------------------------------------------

**mAP@k:** | 0.688161 | 0.686874 | 0.669521 | 398

----------------------------------------------------------------

כאשר חישבנו את ערכי ה-AP פר קטגוריה ואת הממוצע על פני כל הקטגוריות (mAP) קיבלנו כי עבור פחות מ-50 דוגמאות התוצאות יורדות מתחת ל-90%. ניתן לראות כי ה-mAP יורד באופן משמעותי בגלל שתי הקטגוריות בהן אין מספיק דוגמאות ובהן אנו מקבלים AP נמוך (משפיע באופן אחיד על הממוצע, שכן הממוצע מתעלם מחלקו היחסי של הclass ומבצע ממוצע על כמות הקלאסים).

# Discussion and Conclusion

בשלב בחינת המידע הנתון (התמונות וה-labels) שמנו לב כי יש הטיה חד צדדית של הדוגמאות כלפי כנראה אנשים עם יד ימין דומיננטית. כמו כן, ראינו שאין דוגמאות של שימוש ב-forceps ביד ימין ומעט מאוד דוגמאות של שימוש במספריים ביד שמאל. חשבנו שיש צורך לעשות ניסוי של אוגמנטציה של הסט של האימון, אך בחרנו לבחון קודם שינויים בפרמטרי המודל וקיבלנו\_\_\_\_\_\_\_ עבודה בהמשך עשויה לכלול שימוש בשיטות לאוגמנטציה של ה-data.

אימנו מודל SOTA לזיהוי אובייקטים (YOLOv7) ובתוצאות הראשונות קיבלנו mAP@50=68% ו-accuracy=87.9%. כאשר אנו מפרקים את הביטוי של ה-mAP ל-AP פר קטגוריה, אנו מבינים שכנראה לכמות הדוגמאות היתה השפעה משמעותית (יותר דוגמאות גורר תוצאות טובות יותר).

ברירת המחדל של YOLOv7 הוא שימוש בSGD כ-optimizer. כיוון וידוע כי adam הוא המאפטם המתקדם ביותר, ביצענו ניסוי כאשר שינינו את הoptimizer, וקיבלנו\_\_\_\_\_\_\_\_\_\_\_.

ניסוי נוסף שביצענו כלל שינוי של ערך הסף של ה-IoU, שברירת מחדל שלו הוא 0.2. קיבלנו\_\_\_\_\_

בזמן ביצוע inference על הסרטונים הנתונים, אנו משתמשים במנגנון החלקה, אשר לכל פריים מתייחס ל-8 פריימים שלפני ואחריו ולפי זה מקבל החלטה לגבי השערוכים השונים (bbox, class). המנגנון החלקה כולל בחינה של כמות הפרדיקציות שניתנו ע"י המודל, במידה והם שונים מ-2 (עבוד כל יד), אנו מבינים שהפרידקציה הנוכחית לא מדוייקת ויש לתקן גם אותה. כאשר בחנו את התוצאות בעין, נראה שזה הצליח להחליק את הפרדיקציות אך לא באופן מושלם.

לסיכום, אימנו מודל על כמות קטנה יחסית של תמונות (790 תמונות אימון ו-132 תמונות ולדיציה) וקיבלנו יחסית תוצאות מעודדות (מעל 90% AP ל-class כאשר היו מעל 50 דוגמאות). הגדלת הדוגמאות באמצעות הוספה של סובייקטים עם יד שמאל דומיננטית או באמצעות אוגמנטציות צפויה לשפר באופן דרמטי את התוצאות, בעיקר עבור המחלקות בהן יש פחות מ-50 דוגמאות.

# References

[1] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” *arXiv preprint arXiv:2207.02696*, 2022.