

פרויקט יישומי למידה- זיהוי והסרת אובייקטים

רקע הפרויקט

המחקר שלנו עוסק בפיתוח שיטות מתקדמות לעריכת תמונות דיגיטליות, במיוחד באמצעות אוטומציה של תהליכי זיהוי והסרת אובייקטים מתוך תמונה. המוטיבציה מאחורי הפרויקט היא ללמוד בעצמנו ולשים יד על דאטא גדול ולאמן מודלים ברשתות עמוקות ולשלב עם עוד מודל שאין לנו ידע מוקדם עליו וגם מהצורך להעלות את רמת היעילות והדיוק בתחום תוך כדי שיפור של חווית משתמש שלא יצטרך לעשות הרבה פעולות עד שמשיג את מטרותיו, שכן תהליכים אלו משפרים את האפשרויות לשימוש בתמונות עבור מטרות שונות - משיפור אסתטי ועד שמירה על פרטיות וכלה בפיתוח מערכות למידת מכונה. התפיסה המסורתית, המבוססת על עבודה ידנית ותלויה במיומנות המבצע, אינה יעילה ודורשת משאבים רבים. הפרויקט שלנו מתמקד בהתמודדות עם אתגרים אלו תוך חקר היכולת לזהות ולהסיר אובייקטים ספציפיים, כגון כלי רכב, מתמונות דיגיטליות, כשלב ראשון לפני הרחבה לאובייקטים נוספים.

תיאור הפרויקט

הפרויקט שלנו מורכב משני חלקים ראשיים:

1. זיהוי אובייקטים בתמונה
2. הסרת אובייקטים מתמונה

זיהוי אובייקטים בתמונה

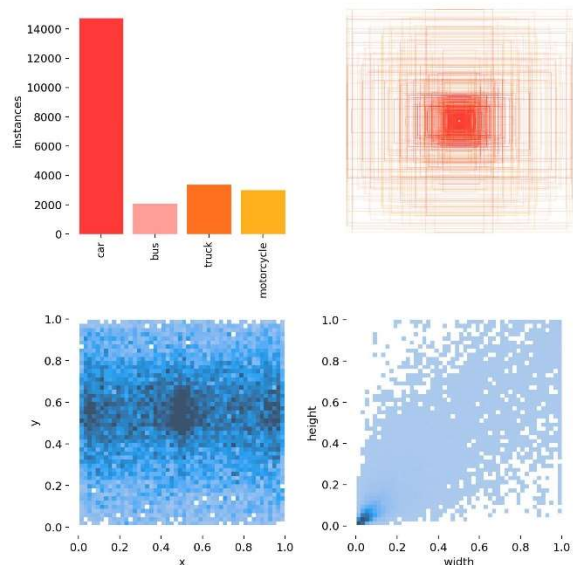
1. דאטא:

במסגרת הפרויקט בחרנו להשתמש בסט נתונים של תמונות מהדאטא-סט COCO-2017. הבחירה ב-COCO-2017 מתבססת על הגיוון של תמונות שמיכלות עצמים בקונטקסטים שונים, מה שהופך אותו לאידיאלי להכשרת מודלים שיכולים להבין ולזהות עצמים בתמונות. בנוסף השתמשנו בספריית FiftyOne לצורך הורדה ועיבוד מקדים של הדאטא שנפרט עליהם בהמשך.

הדאטא סט כלל 6000 תמונות אימון ו-2000 תמונות ולידציה. לנוכח אילוצים של כוח עיבוד, החלטנו לצמצם את סט הנתונים לתמונות עם כלי רכב בלבד (מכוניות, אוטובוסים, משאיות, אופנועים), והספרייה הגבילה את סט הולידציה ל-796 תמונה.

במהלך שלב הניסוי הראשוני, תהליך הורדת התמונות והעיבוד המקדים שלהן נעשה באופן ידני. עם זאת, זיהינו כי מספר רב מהתמונות אינן מתאימות לצרכי המחקר שלנו, משום שהן לא כללו את עצמי העניין – רכבים, אשר היו אמורים להוות את נושא הלמידה של המודל. התמודדות עם אתגר זה הובילה אותנו לחפש ולבחון שיטות נוספות, יעילות יותר, לביצוע המשימה. במסגרת חקרנו, התוודענו לספריית FiftyOne אשר הפכה לכלי משמעותי בפרויקט שלנו.

ספריית FiftyOne מספקת פתרון מתקדם להורדה ועיבוד ראשוני של תמונות, על ידי הפשטת התהליכים אלו והקלתן משמעותית. היא מאפשרת למשתמש להגדיר בצורה מדויקת אילו סוגים של אובייקטים יופיעו במערך הנתונים, ובכך לשלוט על הרלוונטיות והאיכות של הנתונים המתקבלים. כמו כן, הספרייה תומכת ביכולת לבצע ויזואליזציות של התמונות, מה שמאפשר לנו להעריך את חומר הגלם לפני תחילת תהליך האימון הממושך של המודל.



2. עיבוד מקדים:

שלב העיבוד המקדים של הנתונים בפרויקט זה כלל שני שלבים מרכזיים:

1. המרת פורמט הנתונים: המשימה הראשונה הייתה להמיר את מערך הנתונים מפורמט COCO, המבוסס על JSON, לפורמט YOLO, אשר משתמש בקובצי TXT. בשלב התחלתי, התהליך זה נעשה ידנית (מצורף בקובץ: converting-try1), כאשר התמונות נשמרו בפורמט המקורי וכל אחת מלווה בקובץ TXT מתאים, המפרט נתונים על אובייקטים ספציפיים שהוזוהו בתמונה. עם זאת, לאחר שהוצגה פונקציונליות מתקדמת בחבילת FiftyOne, המאפשרת המרה אוטומטית ומהירה לפורמט YOLOv5, נוטש השיטה הידנית לטובת השיטה האוטומטית, מה שחסך זמן ומאמץ רב.

שלבים אלה נמצאים במחברת האימון DI-proj-10k.ipynb

2. טיהור הנתונים: בהמשך האימון הראשוני, התגלו ביצועים לא מספקים, בחלקם כתוצאה מנוכחות אובייקטים אירלוונטיים בתמונות, שאינם נכללים בקטגוריות כלי הרכב שאותן בחרנו למקד. אובייקטים אלו גרמו למודל להתמקד באלמנטים שגויים, ובכך לפגוע בדיוק הזיהוי. כדי לפתור את הבעיה, הוחלט לעבור שוב על תהליך ההמרה, הפעם תוך כדי הגדרת כל אובייקט שאינו כלי רכב כ- class=other ואז להסיר אותו, מה שמנע את הכללתם של אלו בנתוני האימון. פעולה זו סייעה למקד את למידת המודל ולשפר את היכולת לזהות בצורה מדויקת את כלי הרכב הרלוונטיים.

3. אימון של מודל זיהוי האובייקטים:

במסגרת הפרויקט, בחרנו בYOLOv5, הגרסה החמישית מסדרת You Only Look Once, כמודל העיקרי לזיהוי אובייקטים בזמן אמת. זוהי אחת מהטכנולוגיות החדשות והמתקדמות ביותר בתחום, המציעה שיפורים ניכרים במהירות ובדיוק הזיהוי לעומת גרסאות קודמות. הפופולריות של YOLOv5 בקרב מפתחי מערכות ראייה ממוחשבת מוסברת ביתרונות הבאים:

1. מהירות ויעילות: עיבוד התמונות במהירות גבוהה תוך שמירה על דיוק מקסימלי, הופך את YOLOv5 למתאים ליישום במצבי זמן אמת וסביבות עם משאבי מחשוב מוגבלים.
2. דיוק גבוה: המודל מסוגל לזהות אובייקטים בצורה מדויקת, אפילו בתנאים מורכבים, כמו תמונות עם חפיפות של אובייקטים וברקעים מסובכים.
3. גמישות וגישות: פשטות השימוש והתאמה למגוון רחב של משימות זיהוי, תוך תמיכה באימון מותאם אישית על סטים נתונים ספציפיים.
4. תמיכה קהילתית רחבה: נהנה מקהילה גדולה ופעילה של מפתחים ומדעני נתונים, המספקת מקורות מידע רבים, כלים ותוספות.

התאמת המודל לזיהוי רכבים:

החלטנו לאמן את מודל YOLOv5 על מערך הנתונים הספציפי שלנו, במטרה לייעל אותו לזיהוי רכבים בתמונות. דרך האימון המותאם, המודל פיתח הבנה מעמיקה של המאפיינים הייחודיים לרכבים, מה שהגביר את דיוק הזיהוי והפחית את הסיכוי לשגיאות. המטרה הייתה לשפר את היכולת לזהות רכבים בצורה נכונה ומהירה, ללא תלות במגוון התנאים או בסוגי הרכבים השונים.

טיורינג המודל:

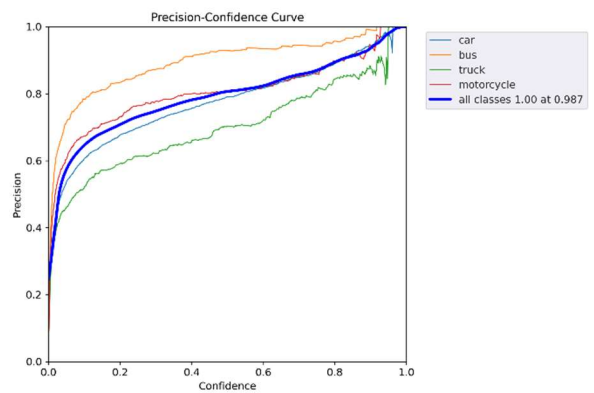
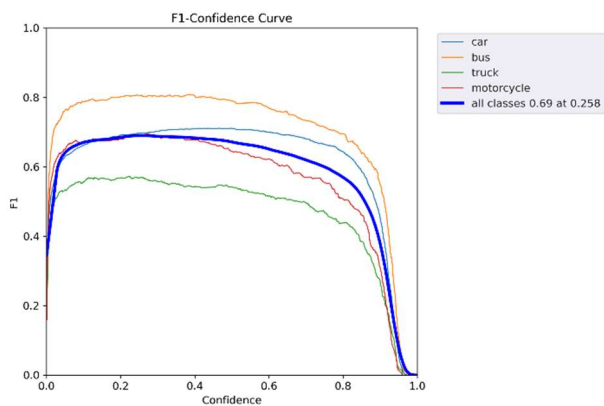
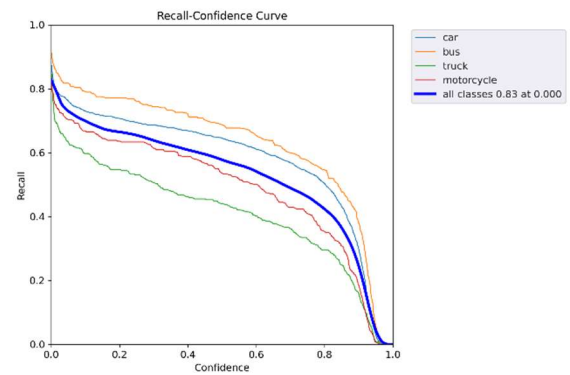
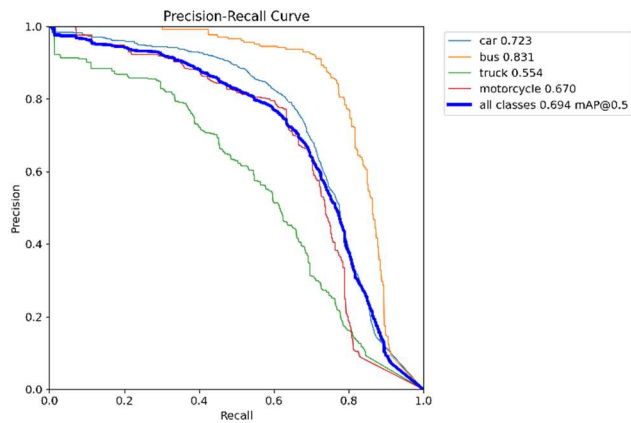
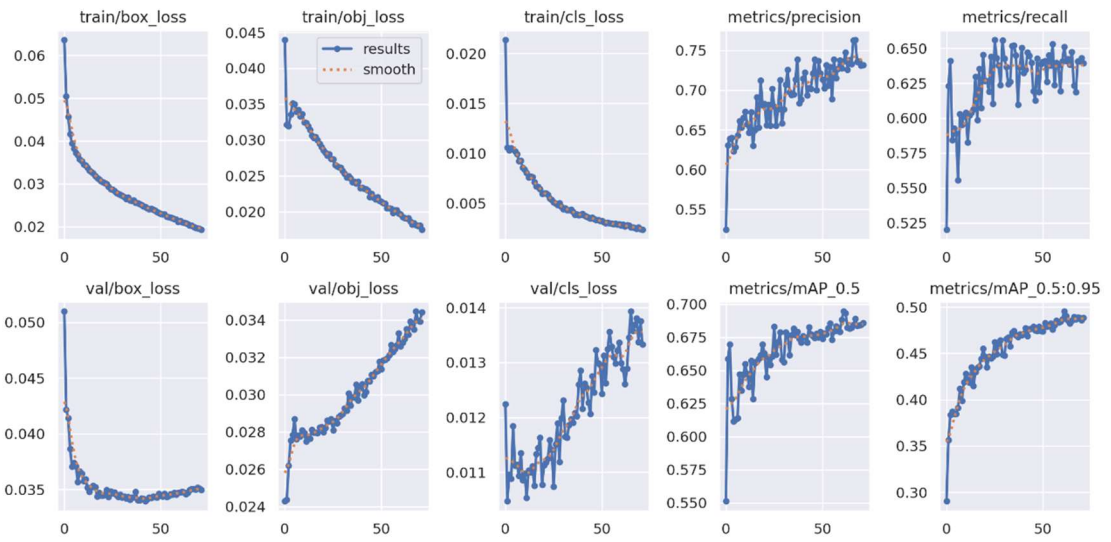
התחלנו את האימון של המודל עם גודל תמונה של 640, עם אצווה (batch) של 8 ו-100 אפוקים (epochs). קיבלנו דיוק (Precision) ושיחזור (Recall) בממוצע של 65%.

לכן, התחלנו לשפר את גודל התמונה ל-720 ולאחר מכן ל-1024, עם אצווה של 16, והפעלנו הפסקה מוקדמת (early stopping) כדי להקטין את זמן הריצה במקרה שלא היה שיפור, והגענו לממוצע של 75%.

נקודה חשובה ביותר היא שהנתונים שלנו לא היו מאוזנים בין הקלאסים: המכונות היוו 60% מהמקרים, ו-20% לאוטובוסים, 20% למשאיות, ו-20% לאופנועים, מה שגרם לדיוק נמוך יחסית בגלל התמקדות ראשית במכונות. אך בסך הכול, כאשר בדקנו את המודל על תמונות חדשות, הוא זיהה בצורה מושלמת כפי שנדרש.

בניתוח התוצאות שהתקבלו מהאימון של שני המודלים, נמצא כי המודל שעבר תהליך התאמה אישית של YOLOv5 הציג ביצועים עליונים במשימת זיהוי העצמים בתמונות. ההצלחה של המודל הזה בזיהוי עצמים בדיוק וביעילות רבה יותר הובילה אותנו לקבל את ההחלטה להמשיך ולהשתמש בו כבסיס לשלבים המתקדמים והחשובים הבאים במסגרת הפרויקט. (לא הצלחנו לשחזר את התוצאות מבעיה בקולאב ננסה בהצגה להראות)

ביצועי המודל:



1. אובדן (Loss) באימון ואימות (Training and Validation): האובדן בקופסא (Box), האובדן בעצם (Object), והאובדן במחלקה (Class) ירדו באופן עקבי במהלך האימון, מה שמרמז על שיפור והתכנסות המודל. האובדן בעת אימות הוא גבוה יותר בהשוואה לאימון, אך גם הוא ירד עם הזמן, סימן לכך שהמודל מכליל היטב.

2. דיוק (Precision) ושיחזור (Recall): הדיוק והשיחזור במהלך האימון היו יציבים עם מגמה קלה של שיפור, סביב הערכים של 0.65 לדיוק ו-0.60 לשיחזור. עבור אימות, ראינו עלייה בדיוק ובשיחזור, עם נטייה להתכנסות סביב הערכים של 0.625 ו-0.675 בהתאמה.

3. ממוצע דיוק (mAP): מדד ה-mAP באימון ובאימות שיפר באופן עקבי והגיע לשיעור של כ-0.50 באימון וכ-0.45 באימות עבור mAP ברזולוציית סף של 0.5 עד 0.95.

4. עקומת זיכרון-ביטחון (Recall-Confidence): העקומה מראה כי עבור רמות ביטחון גבוהות, הזיכרון (שיחזור) עבור כלל המחלקות היה גבוה ומתכנס לערך של 0.83 ברמת ביטחון של 0.000.

5. עקומת דיוק-זיכרון (Precision-Recall): המודל הראה דיוק גבוה עבור מחלקת האוטובוסים עם ערך של 0.831, בעוד מחלקת המשאיות הראתה את הדיוק הנמוך ביותר עם ערך של 0.554. הדיוק הממוצע עבור כל המחלקות עמד על 0.694 ברזולוציית סף של 0.5.

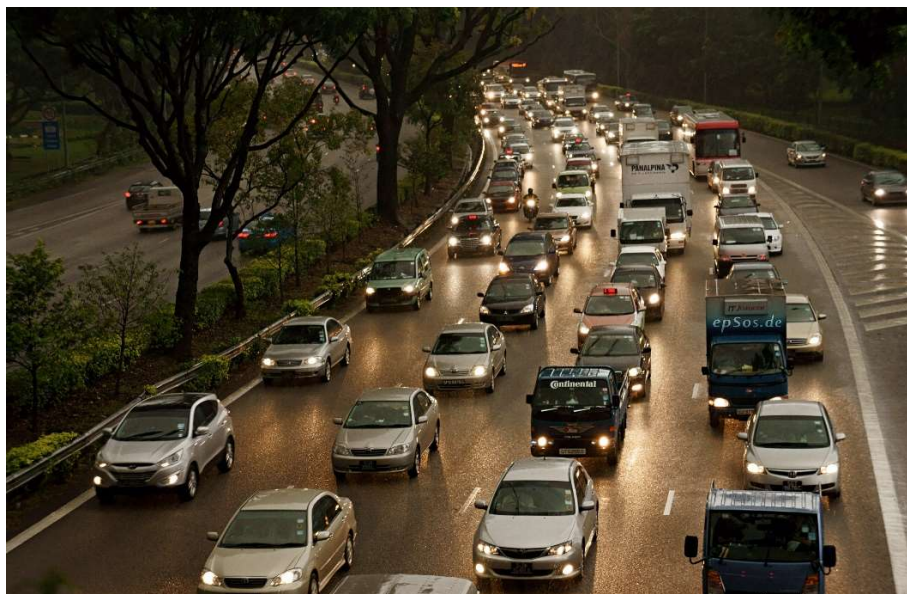
6. עקומת דיוק-ביטחון (Precision-Confidence): הדיוק שיפר ככל שהרמת הביטחון עלתה,

עד לערך מושלם של 1.00 ברמת ביטחון של 0.987 עבור כלל המחלקות.

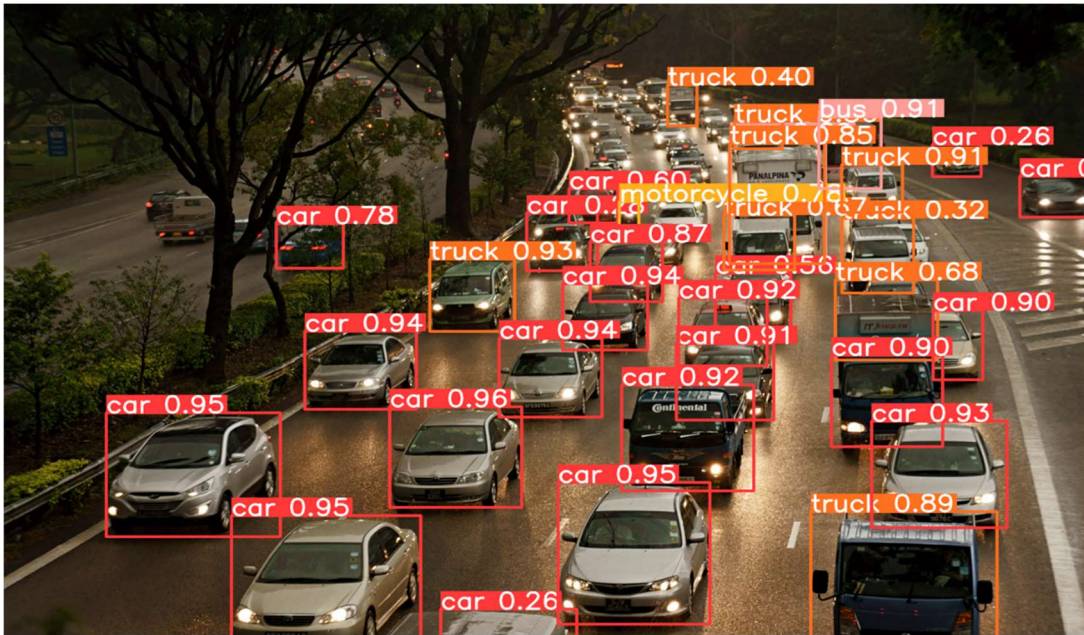
7. עקומת F1-Confidence: מציגה את הממוצע ההרמוני של הדיוק והרקול עבור סוגי רכבים שונים. הניקוד הגבוה ביותר לאוטובוסים הוא 0.831, מה שמראה על זיהוי עקבי ומדויק. עבור כל המחלקות, הניקוד המיטבי הוא 0.69 ברמת ביטחון של 0.258.

בסה"כ, המודל שלנו מציג ביצועים טובים עם יכולת זיהוי טובה במיוחד עבור מחלקות ספציפיות כמו רכבים ו אוטובוסים, ועם יכולת גמישה להתאמה לרמות ביטחון שונות, תוך שמירה על דיוק גבוה עם המגבלות של כסף וזמן שיהיו לנו שהתאמנו על 6000 מתוך 120,000 ועקפנו את המודילים המאומנים בביצועים אנחנו מסתפקים בזה.

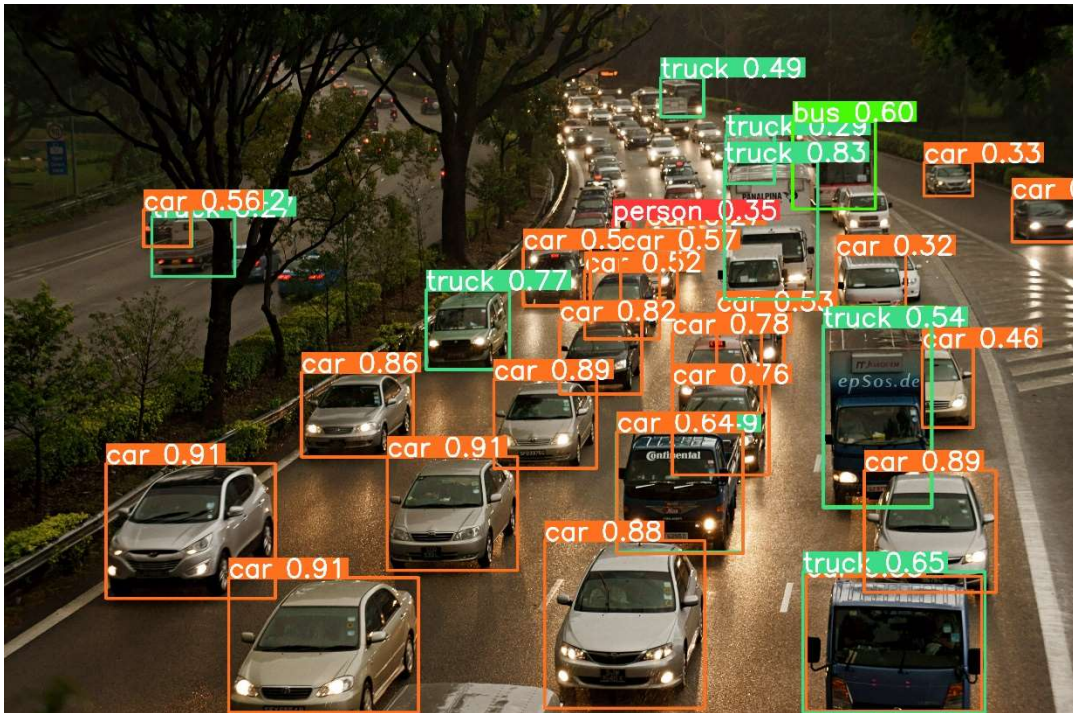
אצרף דוגמה להשוואה בין ביצועי המודל שלנו לעומת המודל הכי גדול ומוכשר של yolov5XL בתמונה הראשונה המודל שלנו מזהה את הרכבים עם הסתברויות יותר גדולות מאשר המודל השני ומצליח לתפוס אובייקטים מורכבים יותר.



YOLOV5-Custom



YOLOV5XL



סוג חומרה ועלויות לאימון :

1. CPU-Tesla V100
2. CPU-Nvidia A100
3. RAM 50-80 GB
4. Computing Units+Drive Storage= 80 Shekel

הסרת אובייקטים מתמונה

מודל מילוי אובייקטים:

לאחר שהתמקדנו במודל המותאם אישית של YOLOv5, המשכנו את התהליך עם שימוש ב-Mask R-CNN, שאפשר לנו לזהות באופן מדויק את האובייקטים המזהים בתמונה ולהפריד בינם לבין הרקע. כל אובייקט שזוהה על ידי המודל התווסף לו מסכת סגמנטציה המתארת את צורתו המדויקת. בשלב זה, נעזרנו בשתי פונקציות חשובות לניתוח התוצאות: הפונקציה 'iou' והפונקציה 'percent_within'.

הפונקציה (Intersection over Union) 'iou' מחשבת את ה IOU-בין כל זוג קופסאות, מהווה כלי מרכזי להערכת הדיוק בזיהוי אובייקטים על ידי חישוב החפיפה ביניהם. פונקציה זו איפשרה לנו למדוד באופן מדויק את איכות הזיהוי של המודל ולהשוות בין מסכות שונות שנוצרו.

הפונקציה 'percent_within' מחשבת את האחוז של נקודות שנמצאות בתוך מלבן מסוים, מה שאיפשר לנו להעריך כמה מהאובייקט או המסכה נמצא בתחום הרצוי. שימוש בפונקציה זו היה קריטי לניתוח המידע המתקבל מהמסכות ולהבנת היכולת של המודל למקד את האובייקטים באופן דיוק.

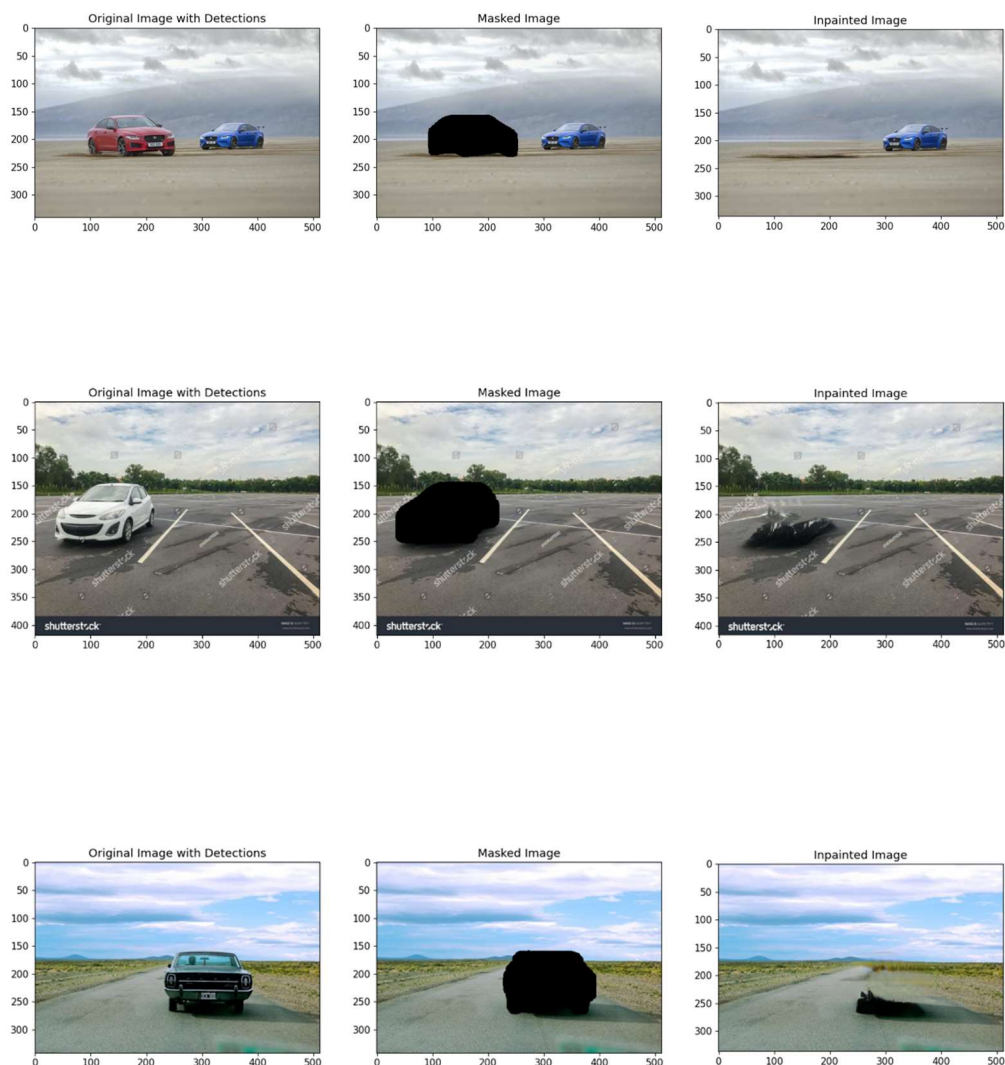
לאחר שהשגנו מסכות סגמנטציה מדויקות, התהליך המשיך עם השימוש בטכנולוגיית DeepFill, שנועדה למלא חללים שנוצרו לאחר הסרת אובייקטים מהתמונה. באמצעות רשתות נוירונים התחרותיות (GANs), DeepFill משחזרת את האזורים החסרים בצורה טבעית, שומרת על רצף הטקסטורה והאור, ומבטיחה שהתמונה הסופית תיראה כאילו העצם לא היה קיים בה מלכתחילה.

בכדי לשמור על יציבות התהליך ולהבטיח את יעילותו, נעשה שימוש בנורמליזציה ספקטרלית בסביבה של הדיסקרימינטור. הדיסקרימינטור, המשמש כ"שופט" במהלך האימון של רשת נוירונים התחרותיות, נדרש להיות מאוזן ויציב כדי להבטיח את הצלחת הלמידה. הנורמליזציה הספקטרלית מבצעת משימה זו על ידי הגבלת נורמת המשקלים של השכבות, מה שמונע "התפוצצות" בערכיהם ומאפשר אימון מאוזן ויעיל יותר. השימוש בטכניקה זו הוא חיוני לקיום אינטראקציה תקינה ומוצלחת בין הגנרטור לדיסקרימינטור, מה שמאפשר למודל ללמוד ולהשתפר באופן מתמיד.

בשלב הסופי של הפרויקט, פותח ממשק משתמש אשר מאפשר למשתמשים להעלות תמונות ולקבל את התמונה המעובדת ללא העצמים שבחרו להסיר. הפרויקט מדגים את הפוטנציאל של שימוש בטכנולוגיות למידה עמוקה לשיפור ואוטומציה של תהליכים ויזואליים, מה שמבטיח לשנות באופן מהותי את הדרך בה אנו עובדים עם תמונות דיגיטליות.

תוצאות:

הפלט של המודל הוא התמונה ללא האובייקט שברצוננו להסיר, להלן דוגמא:



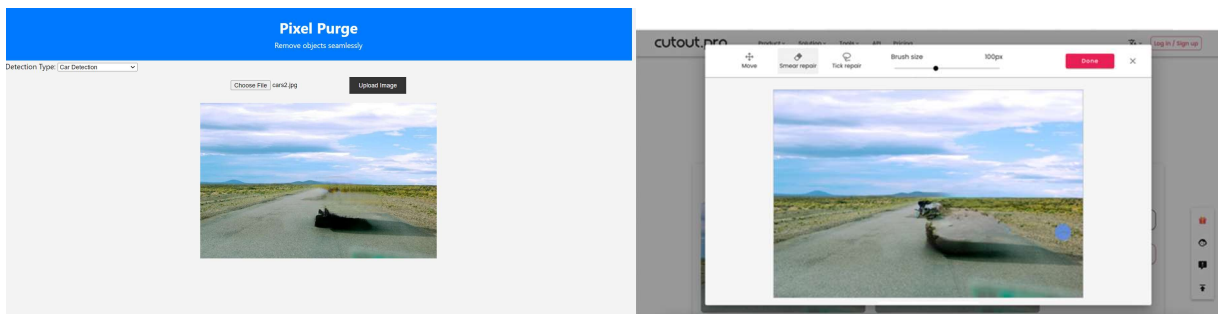
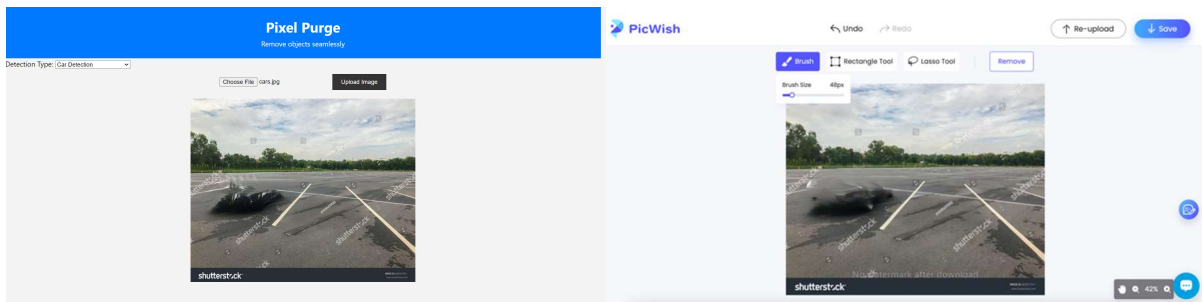
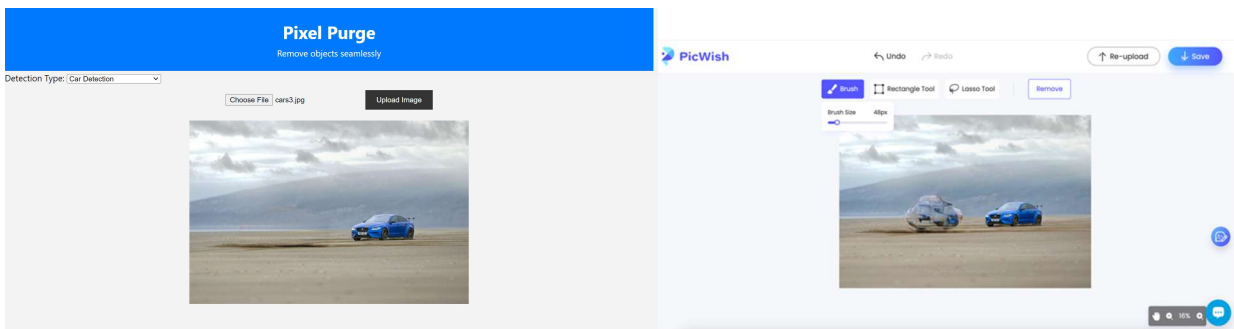
מדדי ביצוע:

במהלך פיתוח המודל שלנו, אחד היעדים העיקריים היה להשיג ביצועים גבוהים במיוחד בתחום הסרת אובייקטים מתמונות ומילוי החלל הנוצר באופן שישאיר את התמונה נראית טבעית וללא סימני עריכה. לשם כך, ביצענו השוואה בין התוצאות שהושגו על ידי המודל שלנו לבין תוצאות של מודלים מתחרים, המוצעות על ידי אתרים ושירותים אחרים באינטרנט.

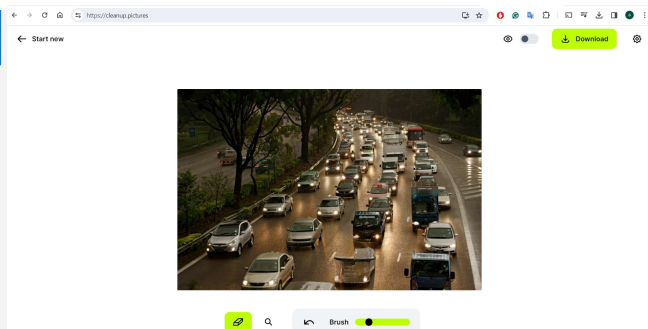
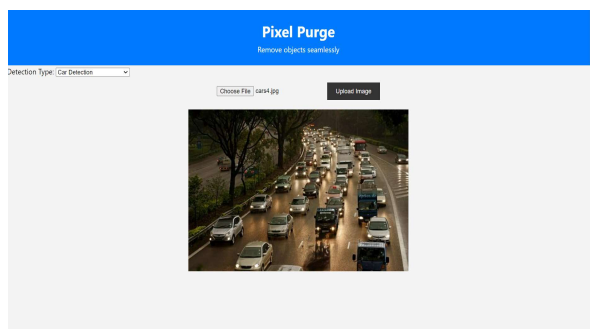
ההשוואה התמקדה במספר מדדים חשובים: דיוק הזיהוי והסרת האובייקט, רמת הטבעיות והאחידות של החלל שנוצר לאחר הסרה, והיכולת לשמור על רצף הטקסטורה והאור באופן שהתמונה המעודכנת תיראה כמו שלמה ולא עברה עריכה.

במבחני ההשוואה, תמונות שעברו עיבוד על ידי המודל שלנו הציגו מראה טבעי ואחיד באופן משמעותי יותר מאלו שהופקו על ידי מודלים אחרים. האובייקטים הוסרו בדיוק גבוה, כאשר החללים שנוצרו מולאו בצורה המשקפת את הסביבה המקורית באופן טבעי, ללא חולשות ברורות או תסכולים של עריכה עד כדי שגיאה קטנה.

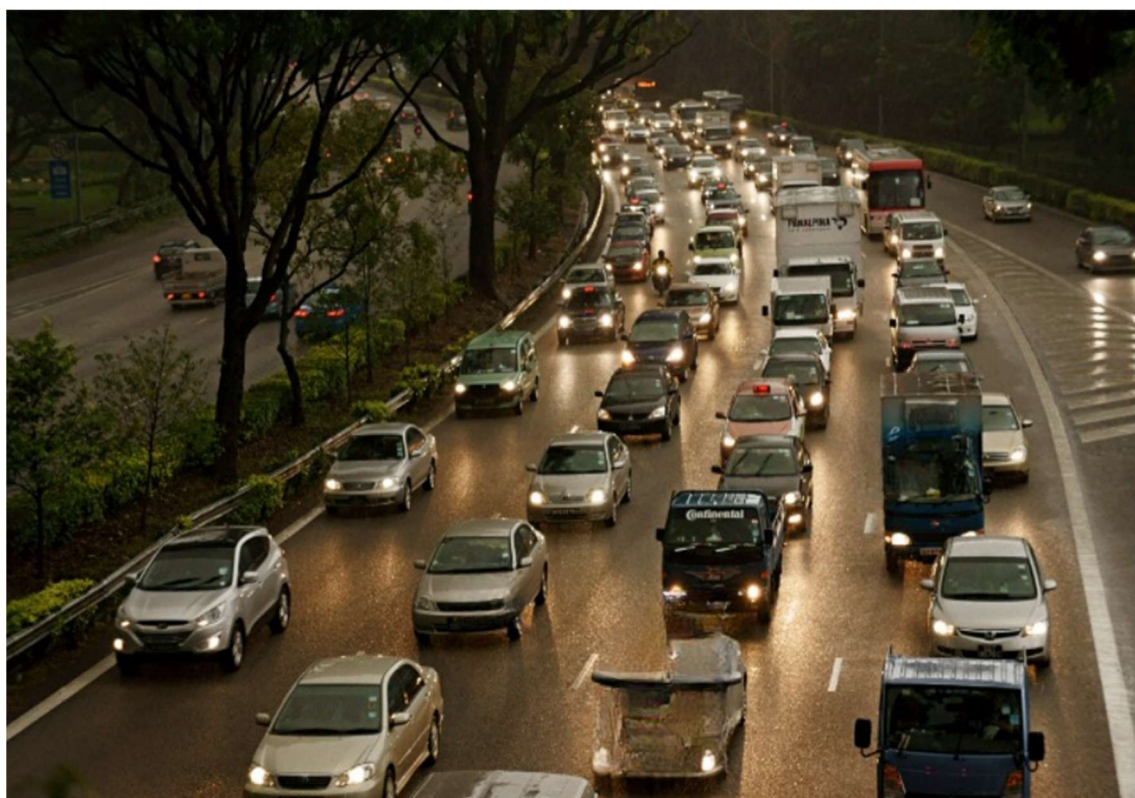
להלן דוגמאות מאתרים אחרים:



השוואה בין התוצאות שנוצרו על ידי אתר הנחשב לאחד המדויקים והמועדפים על ידי רוב המשתמשים (מצד ימין) לתוצאות שהושגו באמצעות המודל שפיתחנו (צד שמאל):



Cleanup image



Our website(Pixel Purge) model image



אחרי שסיימנו ראינו שאנחנו יכולים להוסיף עוד פיצ'ר קטן שהוא יזהה ויסיר כמות רחבה יותר של סוגי אובייקטים מאשר כלי רכב והצלחנו להשיג את זה!

דוגמה עם הסרת הבן אדם מספר 7 😊



לסיום נהגנו מאוד ולמדנו הרבה בפרויקט הזה בין אם זה מבחינת איך להתעסק בדאטא ענקי לבין אימון מודילים ברשתות עמוקות לבין שילוב אותם עם מודילים מאומנים אחרים, בהחלט זה יהיה דבר ראשון שנדבר עליו בראיון עבודה וזה בהחלט הדבר הכי מעניין ושימושי שיצא לנו לעשות כמובן בעזרה ותמיכה שלכם אז תודה רבה מכל הלב לאורן ועודד בכלל לא חשבנו שנגיע לרמה כזאת בהתחלת הקורס !