

דו"ח סיכום לתרגיל 3 בראייה ממוחשבת
נושאים: Segmentation
תאריך: 01.06.20

מגשים:
אופק חררי 305199879
שני ישראלוב 204679153

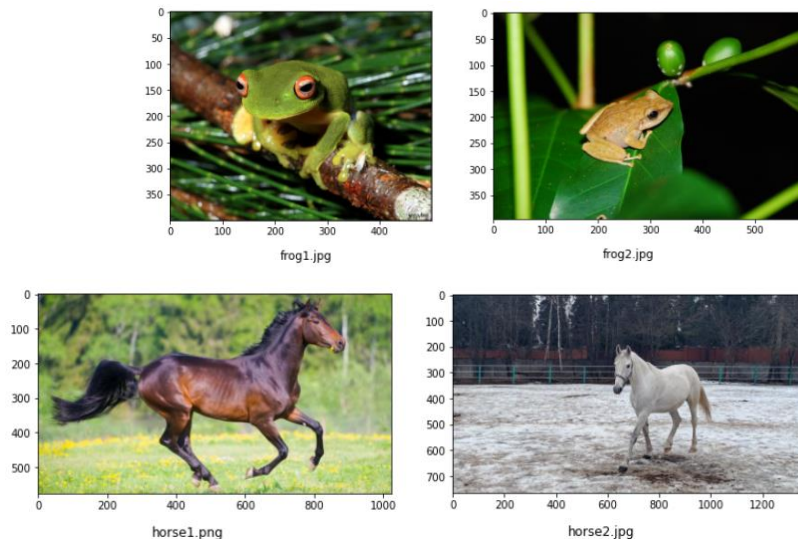
הבהרות על אופן העבודה שלנו: עבדנו ב- google Collaboratory notebooks, חוץ מתיקיית data שניתנה בתרגיל השתמשנו בתיקיות נוספות שמצורפות להגשה לתרגיל זה. בחלק 1: בתיקיית my_data יש שתי תיקיות: section3 עם הקבצים הרלוונטים לסעיפים 3,4,5 ותקייה שנייה section7 עם הקבצים הרלוונטים לסעיפים 7,8,9,10. מחלק 1 תמונת התוצאה של סעיף 9 מופיעה בתיקיית output. בחלק 2: בתיקיית my_data מופיע הסרטון שלנו שצילמנו ותמונת הרקע לסרטון שיצרנו. בחלק זה אנחנו יוצרים תיקיות חדשות לאחסון הפריימים של הסרטונים בתיקיית datasets/OTB/img/Custom, מחולק לשתי תיקיות עם הפריימים של הוידאו שלנו ששמופיע ב-my_data ושל האיש הרוקד מתיקיית data שניתנה בתרגיל. לאחר הסגמנטציה כל הפריימים הללו מופיעים בתיקיית output_path_after_segmentation, בתוך התיקייה הזאת כל תמונה שמופיעה עם הסימנת 01 היא mask. סרטון התוצאה מסעיף 4 מופיע בתיקיית output.

Part 1 - Classic Vs. Deep Learning-based Semantic Segmentation

בחלק הזה אנחנו משווים מתודות קלאסיות לסגמנטציה עם שיטות מבוססות למידה עמוקה. נראה את האפקט של הרקע על הסיווג.










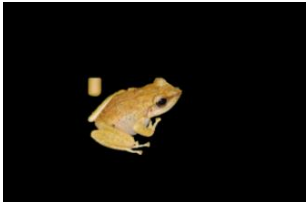


סעיף 1:

התמונות מתיקיות frogs ו-horses:



סעיף 2:

השיטה הקלאסית לסגמנטציה שבחרנו היא grabCut. בחרנו שיטה מבוססת למידה עמוקה לסגמנטציה והיא deeplabv3 כפי שהוצגה בתרגול. אלה כל התוצאות עם ההשוואה:

שיטה קלאסית	למידה עמוקה	תמונות מקוריות
		 Horse2
		 Horse1
		 Frog1
		 Frog2

*תמונות של ה-masks משני השיטות מופיעים בנספח בסוף הדו"ח.

סיכום של כל שיטה ומה היתרונות והחסרונות של כל שיטה.

השיטה הקלאסית – grabCut:

סיכום השיטה:

שיטה שמבוססת על חתך הגרף. בשיטה אנחנו מציינים bounding box מסביב לאובייקט שרוצים לעשות לו סגמנטציה. האלגוריתם מעריך האלגוריתם מעריך את התפלגות הצבעים של אובייקט היעד ושל הרקע באמצעות מודל גאומטרי ומפעיל אופטימיזציה מבוססת חתך גרף כדי להסיק את הערכים שלהם. הליך זה חוזר על עצמו עד להתכנסות.

יתרונות:

- לא תלויה בתגיות.
- ניתן לראות בתמונות שלנו שהחיתוך של horse1 בוצע בצורה יותר יפה בשיטה הקלאסית, השיטה הקלאסית לא הכניסה לשפות של הסוס את הרקע הירוק. זה קורה כי שיטת grabCut מתקנת את התוצאות לשימור הקצוות.

חסרונות:

- צריך לסמן באופן ידני את מיקום האובייקט שבו אנחנו מעוניינים בתמונה. במקרה שלנו עשינו מלבן שהכנסנו לו את הפיקסל השמאלי העליון ואז את גודל המלבן.

שיטת למידה עמוקה - deeplabv3:

סיכום השיטה:

שיטה זו מחלצת features ומשתמשת ב- atrous convolution בבלוקים האחרונים ברשת. לאחר מכן רשת ASPP מתווספת כדי לסווג כל פיקסל למחלקה המתאימה לו. הפלט של רשת ASPP מועבר לקונבולוציה 1×1 כדי לקבל את הגודל המקורי של התמונה ומקבלים מסכה עם הסגמנטציה של האובייקט בתמונה. השתמשנו כשלב בקוד שניתן בתרגול וביצענו עריכה לפי הצרכים שלנו בתרגיל כמו קבלת התגית כקלט.

יתרונות:

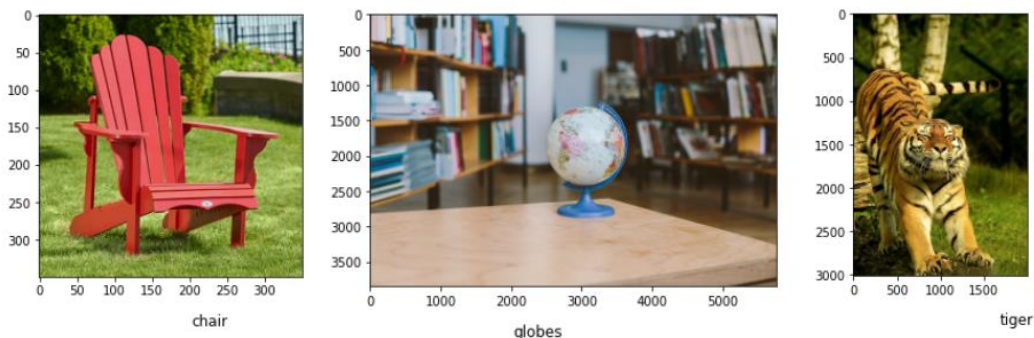
- אין אובייקטים או חלקים מיותרים שאינם קשורים לאובייקט. ניתן לראות את היתרון הזה בצורה יפה ב-horse2 ששם הסוס זוהה בצורה מאוד יפה בעוד שבשיטה הקלאסית האדמה הלבנה זוהתה כחלק מהאובייקט.
- אפשר ליישם את השיטה על הרבה תמונות, אין צורך בחיתוך ידני.

חסרונות:

- השימוש בתגיות מקשה לזהות אובייקטים שלא מופיעים ברשימת התגיות. במקרה שלנו קיבלנו את הצפרדעים למרות שאין תגית של "צפרדע" על ידי שימוש בתגית "ציפור". כמובן שזה שגוי על אף שקיבלנו תוצאות סבירות.
- ניתן לראות בתמונות frog2, horse1 שהשפות הן לא מאוד חלקות וכוללות גם חלקים מהרקע של האובייקטים.







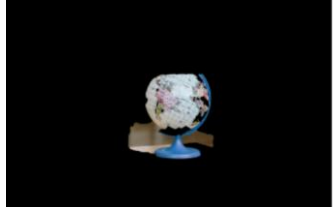
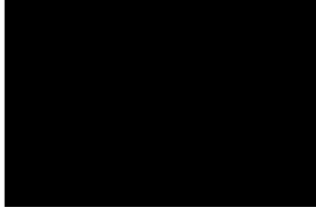

לסיכום, אנחנו מסיקים מהתוצאות של ארבעת תמונות אלו שהתוצאות של למידה עמוקה הן טובות יותר, ובאופן מפתיע גם עבור אובייקטים שאין להם תגית כמו הצפרדע.

סעיף 3:



סעיף 4:

התוצאות של השיטות מסעיף 2 על התמונות מסעיף 3:

שיטה קלאסית	למידה עמוקה	תמונה מקורית
		
		
		

*תמונות ה-masks מופיעות בנספח בסוף הדו"ח.

מי מהשיטות נתנה תוצאות טובות יותר על כל תמונה? מה דעתכם על למה שיטה יותר טובה מאחרת.

- בתמונה של הבעל חיים, טיגריס, השתמשנו ברשת עמוקה עם תגית "חתול", רואים שקיבלנו את האובייקט אבל בצורה מקוטעת, לעומת זאת בשיטה הקלאסית קיבלנו את הטיגריס בצורה יפה יותר, על אף שיש חלק מהרקע שהתווסף בטעות. במקרה זה היינו מעדיפים את השיטה הקלאסית, על אף החיסרון הבולט בה שזה הצורך הידני להגדיר את ה-bounding box.
 - בתמונה של הכסא, השתמשנו ברשת עמוקה עם תגית "כסא", קיבלנו תוצאה טובה. גם בשיטה הקלאסית קיבלנו תוצאה טובה. בשני המקרים קיבלנו חלקים מהרקע של הכסא. במקרה זה היינו מעדיפים את התוצאה של הרשת העמוקה שהכניסה פחות חלקים מהרקע.
 - בתמונה של הגלובוס, השתמשנו ברשת עמוקה, מאחר ואין תגית מתאימה, ניסינו כמה תגיות, לא הצלחנו לקבל סגמנטציה של הגלובוס. הצלחנו לקבל סגמנטציה של השולחן עליו הוא נמצא. זה חיסרון של השיטה מבוססת למידה לגבי אובייקטים שהם פחות נפוצים ואין הרבה דאטה. השיטה הקלאסית לא הניבה תוצאה מאוד יפה אבל היא כן זיהתה את הגלובוס ולכן היינו מעדיפים להשתמש בה.
- לסיכום, בסעיף זה ראינו טווח נוסף של תמונות מעבר למה שראינו בסעיף הקודם. גם כאן תוצאות של הלמידה עמוקה היו טובות יחסית עבור תגיות שמופיעות ברשת, אבל עבור הגלובוס לא הצלחנו למצוא תגית שזיהתה אותו וזה חיסרון גדול של השיטה הזאת וזה ממחיש את יתרונות השיטה הקלאסית על פני השיטה של למידה עמוקה.

סעיף 5:

כפי שראינו סגמנטציה יכולה להיות מחוספסת בשפות, כלומר ה-mask לא מושלמת ועלולה להיות רועשת בשפות. מה ניתן לעשות כדי לתקן או להפחית את הבעיה?
ניתן להשתמש בשיטה לזיהוי שפות, לדוגמא canny, ואז כתלות בשפות שהתקבלו להשתמש ב-dilation כדי ליצור שפות עבות יותר. לאחר מכן ניתן להסיר את השפות מהתמונה. אם השפות שהתקבלו אינן רציפות עושים edge-linking, ראינו בכיתה שיטות להשלמת "חורים" בשפות לדוגמא באמצעות hysteresis thresholding. נציין עבור החלק השני של התרגיל שעבור אובייקטים לא ירוקים שנמצאים על רקע ירוק אחיד נוכל גם להקטין את הבהירות של הערוץ הירוק על מנת להוריד את הרעש בשפות בפעולת הסגמנטציה.

סעיף 6:

המסווג pre-trained שבחרנו הוא VGG16.

סעיף 7:

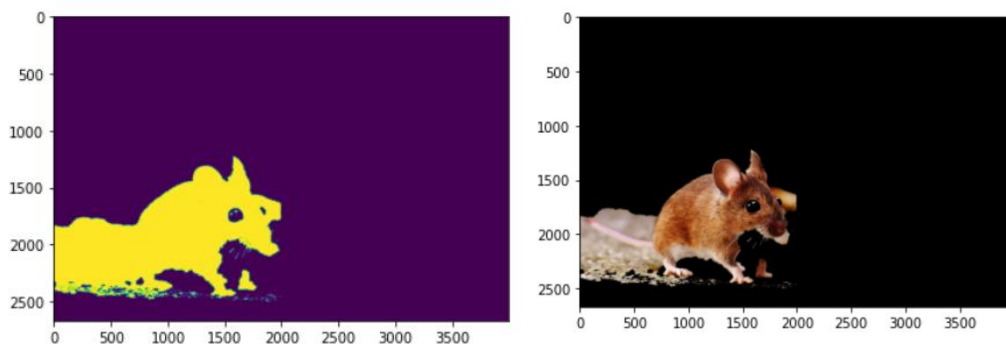
התמונה שבחרנו זה עכבר בשדה.



החיזוי של תמונה זו היא wood rabbit בהסתברות 55.41%. זה חיזוי סביר שכן מדובר במכרסמים בצבע חום.

סעיף 8:

השתמשנו בסגמנטציה בשיטה הקלאסית שהשתמשנו בסעיפים קודמים grabCut. המסיכה והתוצאה:



אנחנו רואים שיש לנו תוצאה טובה יחסית, אמנם קיבלנו גם חלק מהרקע אבל לפחות יש זיהוי ברור של העכבר, לא השתמשנו בשיטה של למידה עמוקה כי ברשת שהשתמשנו בה בסעיף הקודם אין תגית "עכבר" ושימוש בתגיות אחרות לא הניב תוצאות טובות.

סעיף 9:

בחרנו לשים את העכבר ברקע של מטבח כמחווה לסרט רטווי. לצורך זה התאמנו את הגודל של המטבח לתמונת העכבר, השתמשנו ב-mask של הסגמנטציה של העכבר, ובאמצעות פקודת paste הדבקנו את העכבר על רקע המטבח. להלן התוצאה:



תוצאה זו מצורפת בתיקיית output שהגשנו.

רפרנס:



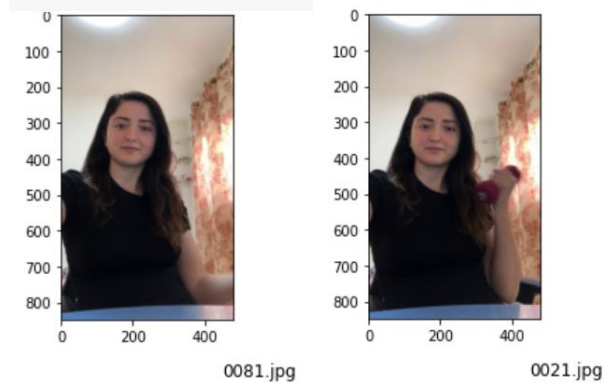
סעיף 10:

החיזוי שקיבלנו לתמונת התוצאה הוא espresso maker בהסתברות 54.65%. העכבר נמצא בסביבה לא טבעית שלו, זוהי סביבה של מטבח, יש כוסות קפה וקומקום ברקע. ייתכן והרשת מזהה את האובייקטים ברקע כאובייקטים יותר דומיננטיים ולכן מסווגת את זה כמכונת קפה. ציפינו לקבל תוצאה שונה כשכן רשת עמוקה מאומנת על דאטה קודם וסביר להניח שאין הרבה תמונות עם עכבר בסביבה של מטבח. המסקנה שלנו היא שהדבקת אובייקט על רקע לא צפוי והכנסה לרשת עמוקה תיתן תוצאות לא מדוייקות ואולי לא תזהה את האובייקט המרכזי אלא חלקים מהרקע.

Part 2 - Jurassic Fishbach

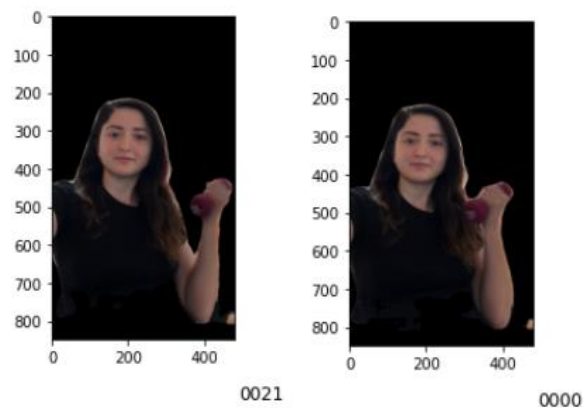
סעיף 1:

צילמנו וידיאו שלנו והשתמשנו בפונקציית `frame_video_convnet.py` שניתנה על ידי סגל הקורס כדי להמיר את הוידאו לפריימים. סה"כ יש לנו 93 פריימים להלן 2 מתוכם:



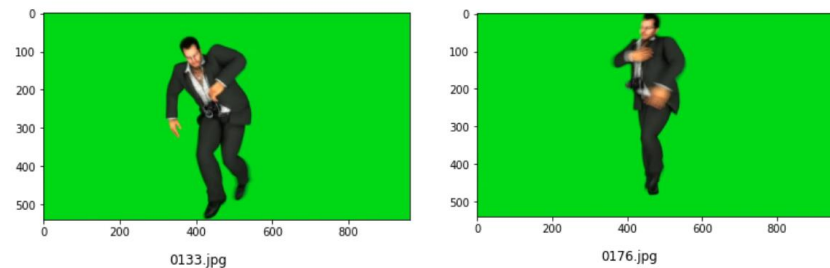
סעיף 2:

ביצענו סגמנטציה לפריימים של הוידאו. השתמשנו בלמידה עמוקה בשיטה שהשתמשנו בחלק הראשון ושיניתנה לנו בתרגול `deeplabv3`. השתמשנו ב-`label='person'`.

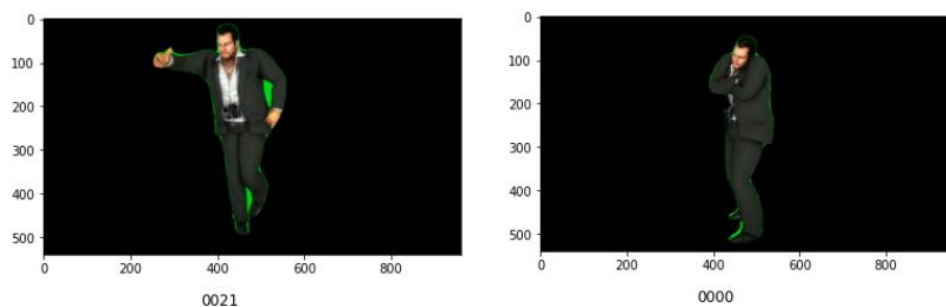


סעיף 3:

בחרנו בסרטון האדם הרוקד, באופן דומה השתמשנו בפונקציה `frame_video_convnet.py` ולאחר מכן במודל `deeplabv3` כדי לבצע סגמנטציה. יש 201 פריימים, להלן 2 מתוכם:



לאחר סגמנטציה:

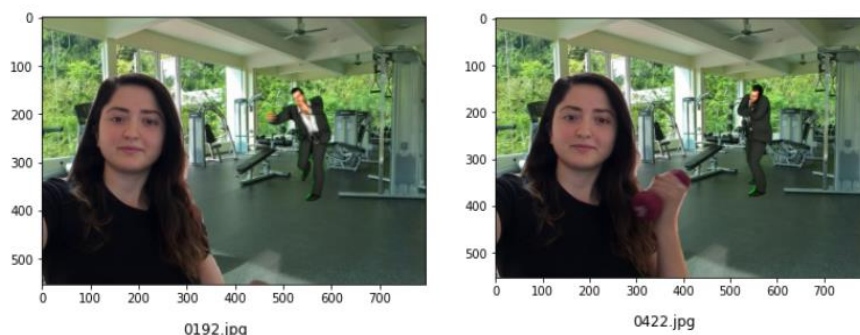


הסבר איך ביצעת את הסגמנטציה:

השתמשנו בשיטה של למידה עמוקה כי יש לנו שם את התגית של "person" שנתנה תוצאות טובות מאחר והאדם נמצא ברקע ירוק אין חלקי רקע נוספים שהתווספו בטעות.

סעיף 4:

בחרנו רקע סטטי של חדר כושר, שמנו את התמונה לאחר סגמנטציה שלנו וגם את התמונה של האיש הרוקד לאחר סגמנטציה. עשינו כך לאורך כל הפריימים. להלן 2 תמונות:




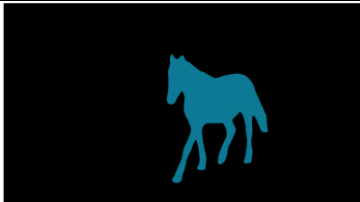


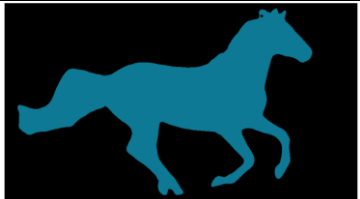







לאחר מכן המרנו לקובץ video באמצעות הפונקציה שניתנה לנו. ביצענו resize לתמונות של האיש הרוקד כדי שייראה נכון מבחינת פרספקטיבה עם הרקע. בעזרת פונקציית paste חיברנו את כל התמונות, השתמשנו ביכולת של פונקציה זו לבחור את הפיקסל שבו יודבק האובייקט כדי למקם את האובייקטים במקומות שרצינו. הכפלנו את התמונות כדי ליצור וידיאו איטי יותר.

קישור ליוטיוב: [shani in the gym](#)










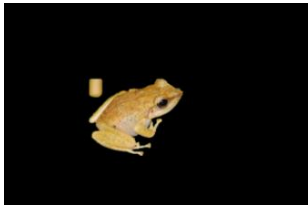


נספחים:

חלק 1 – סעיף 2 :

שיטת של ה-deep learning

למידה עמוקה	mask	תמונות מקוריות
		 Horse2
		 Horse1
		 Frog1
		 Frog2


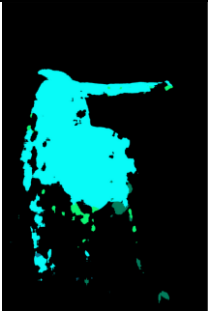







שיטת של ה-grabCut

קלאסית	mask	תמונות מקוריות
		 Horse2
		 Horse1
		 Frog1
		 Frog2

המשך נספחים:

חלק 1 – סעיף 4:

התוצאות של השיטות מסעיף 2 על התמונות מסעיף 3 בעזרת למידה עמוקה:

תמונה מקורית	mask	למידה עמוקה
		
		
		
*תיוג של שולחן		

התוצאות של השיטות מסעיף 2 על התמונות מסעיף 3 בעזרת grabCut:

קלאסית	mask	תמונה מקורית
