**שאלה 2**

סעיף 1

בחרנו להמשיך עם VGG16 , מאומן על ImageNet .

סעיף 2

בחרנו את תמונת הכבשה בdata שסיפקתם לנו, ביצענו עיבוד מקדים לקראת כניסה לרשת בדומה לתרגיל בית 1 , והתקבלה הפרדיקציה הבאה: תמונה שמכילה טקסט, דשא, כבשה, חוץ

התיאור נוצר באופן אוטומטי

באופן מפתיע, אין לimageNet תיוג של כבשה. אף על פי-כן , הרשת הצליחה למצוא את החיה הכי דומה לה מתוך רשימת 1000 התיוגים והוא : האיל.

סעיף 3

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, ירוק, יונק, מכרסם

התיאור נוצר באופן אוטומטיתמונה שמכילה דשא, חוץ, שדה, יונק

התיאור נוצר באופן אוטומטיבחרנו להפעיל: סיבוב עם כיוון השעון, הפיכת גווני התמונה לגווני אפור והפעלת פילטר לפלסיאן.  
התקבלו התוצאות הנ"ל:

סעיף 4

לאחר עיבוד מקדים של כל אחת מהטרנספורמציות, הזנו את התמונות לרשת והתקבלו החיזויים הבאים: תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

לגבי שתי הטרסנפורמציות הראשונות, קיבלנו חיזוי דומה לסעיף 2. הדבר מעיד על יכולת הרשת להיות אדישה להזזות ולשינויי צבע של התמונות המוזנות לה. זוהי תכונת הweight\_sharing המאפיינת שכבות קונבולוציה המאפשרת לה לזהות קשרים מרחביים בתמונה ללא קשר להזזות/שינויי צבע.

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיהרשת **טועה** עבור התמונה המסוננת לאחר מעבר בפילטר לפלסיאן. התמונה חזרתה כי מדובר בjellyfish. נריץ חיפוש תמונות שחור-לבן של jellyfish ונקבל: תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

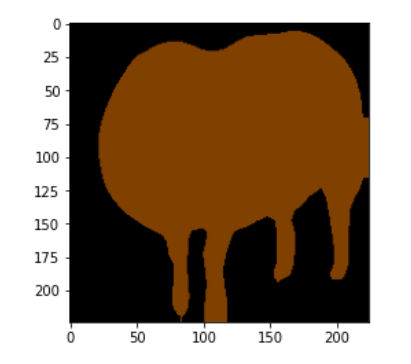
ניתן לראות כי הjellyfish בראש עגלגל שמתנקזות אליו זרועות דקיקות היוצרות צורות קשתיות של אוזניים/קרניים. במובן זה, ישנה התאמה קטנה לכבשה אך היעדר במבנה הצמרירי של הכבשה וכן שינויים בטקסטורת מבנה הפנים גורמות לרשת להכשל בחלק זה של התמונה. היעדרות זו נובעת כמובן מחוסר היכולת של הלפלסאין לזהותם שכן אינם כוללים מעברים חדים של שינויי צבע.

בנוסף, כפי שלמדנו בענ"ת , הלפסלסיאן רגיש לרעשים בתמונה. במקרה שלנו רואים הרבה נקודות באיזור הרגליים של הכבשה. כנראה כתוצאה מהדשא עליו היא עומדת ומשינויי צבע חלקיים בפרוותה התחתונה.  
דבר זה מכשיל את הרשת וגורם לה לחשוב שהכבש היא בעלת מספר רב של רגליים/זרועות כמו הjellyfish.

סעיף 5

השתמשנו בשיטת deep segmentation method של FCN עם רשת pretrained של resnet50. כפי שראינו בתרגול 6, בשיטה זו אנו נפרדים משכבות הFC וע"י הוספת שכבות המבצעות up-convolution - מקבלים תמונה בגודל זהה לתמונת הכניסה תוך שמירה על מפת המאפיינים העמוקה והמורכבת עם receptive field רחב שייצרה רשת הresnet. זה יאפשר סגמנטציה יותר מדוייקת של התמונה.

ביציאה מתקבלים מספר ערוצים כמספר מחלקות הסגמנטציה (21 אצלינו). למעשה , לכל פיקסל ישנו וקטור הסתברויות(המכיל אותו פיקסל מכל אחד מן הערוצים השונים) הנותן אינפורמציה על ההיתכנות לקבלת label מסויים.  
לכן, ביצענו argmax להשארת הlabel עם ההסתברות הכי גבוהה ולקבלת תמונת הפלט הסופית(ערוץ בודד).

לאחר המרה ויזואלית של הסגמנטציות לRGB , קיבלנו את התוצאה הבאה: 

סעיף 6

בסעיף זה ערבבנו רקע שונה עם מסיכת הכבשה שבנינו בסעיף 5.   
(\*)אנו יצרנו תמונה חדשה עבורה, לכל פיקסל, בדקנו האם הוא שייך לרקע השחור או למסיכת הכבשה החומה. בהתאמה לכל אחד מן המקרים, הצבנו לפיקסל את ערך הפיקסל המקביל לו בתמונת הרקע שטענו או את ערך הפיקסל בתמונת הכבשה(כמובן, נירמלנו את כל התמונות לאותו הגודל)

לאחר-מכן, שמנו לב כי מימדי הכבשה אינם פרופורציוניים למימדי הרקע, ולכן ביצענו scaling לאובייקט הכבשה.  
ביצענו זאת ע"י איתור bounding box המקיף את אובייקט מסיכת הכבשה מסעיף 5. האיתור נעשה ע"י מציאת הערכים המקסימליים והמינימליים של המסיכה, ולקיחת המלבן המתאים.

למלבן זה ביצענו הקטנה בפאקטור ,הזזנו אותו לתחתית התמונה וחזרנו על תהליך (\*) עם המסיכה המוקטנת.

התקבלה התוצאה הבאה:

תמונה שמכילה שמים, חוץ, הר, כביש

התיאור נוצר באופן אוטומטי

סעיף 7

לאחר עיבוד, הזנו את התמונה החדשה לרשת והתקבל החיזוי הבא: תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

הרשת חזתה חציר. נראה כי הרשת נתנה עדיפות דווקא לרקע בתמונה במקום לכבשה שבמרכזה. אמנם בצידי הכביש ישנם שדות הדומים לשדות קש (אם כי קשה לראות).

תמונה שמכילה כבשה, דשא, יונק, עמידה

התיאור נוצר באופן אוטומטינראה שהרצון שלנו להתאים את התמונה לגודל המתאים לרקע, הטעה את הרשת בחיזוי. ביצענו הכנסה של התמונה ללא scaling :



והתקבל חיזוי דומה לסעיף 2:

דבר המחזק את הטענה כי הרשת נתנה עדיפות לניתוח הרקע בתמונה המקורית שהכנסנו.

**שאלה 3**

סעיף 1

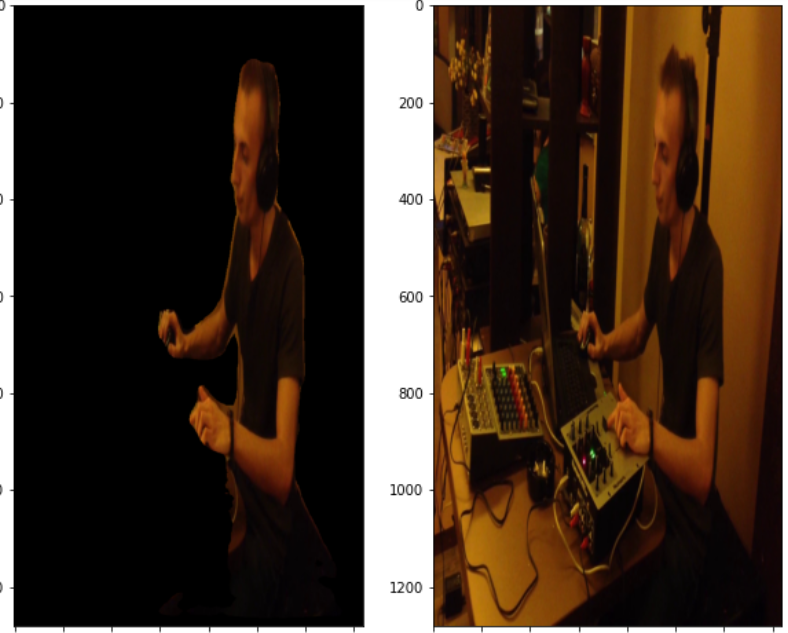
הסרטון נמצא בתוך תיקיית הoutput.חילצנו את הframes ע"י פונקציית העזר וביצענו עיבוד נדרש.

שתי frames לדוגמה: תמונה שמכילה טקסט, מקורה, מדף, שונה

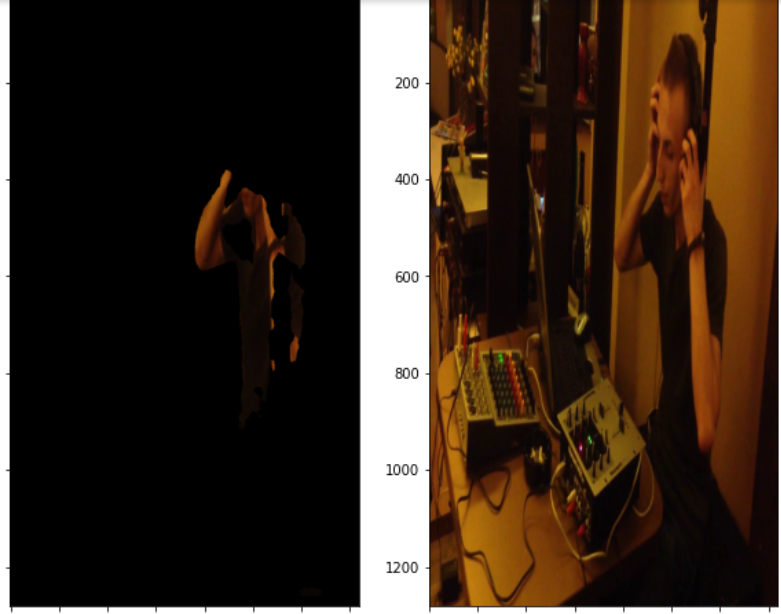
התיאור נוצר באופן אוטומטי

סעיף 2

בחרנו להשתמש במודל deep-learning של deep lab v3 . הפונקציה הממשת זהה לפונקציה משאלה 2 סעיף 6 , עד כדי שינוי מFCN ל deep lab v3 ושינוי הרשת pre-trained לresnet101.

ביצענו סגמנטציה לכל אחד מהframes. ראינו כי יש frames שהמודל מצליח לזהות בצורה מיטבית כמו:

עד כדי מספר פיקסלי רקע קטנים הצמודים לגופו של הבחור, ישנה הפרדה טובה מהרקע. הסגמנטציה אף מזהה את האוזניות כחלק אינטגרלי מהאובייקט.

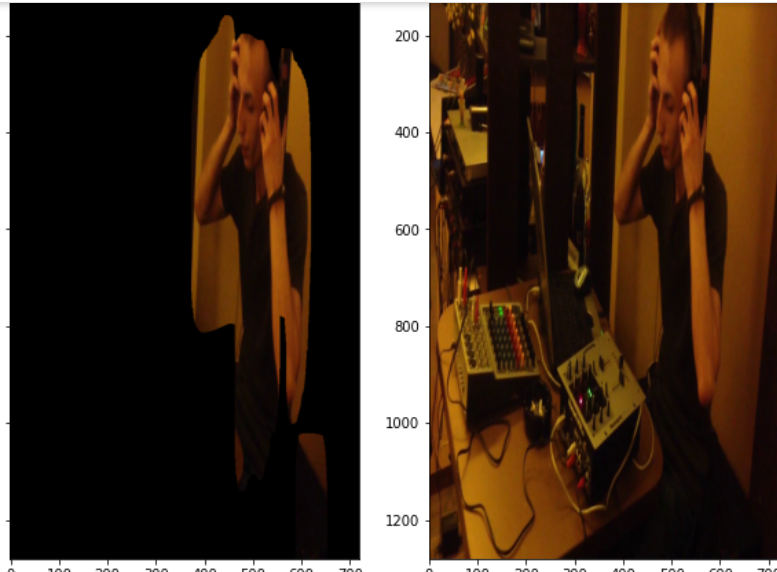
לעומת-זאת, כאשר הבחור שם את האוזניות: 

ניתן לראות כי יש זיהוי חלקי בלבד של האובייקט. כנראה תנוחת הידיים וצורתם על פני הראש גורמים להטעיית הרשת בנוגע לזיהוי חלק הפנים של האדם.  
דווקא בסוף הסרטון כאשר הבחור מוריד את האוזניות בידו הימינית בלבד, הסגמנטציה אינה נכשלת והמודל מצליח לזהות את פני הבחור וגופו כאחד.

**פתרון לבעיה:** זיהינו את התמונות הבעייתיות כתמונות אשר התמונה הסגמנטיבית שלהן היא בעלת כמות פיקסלים המהווה יותר מ90% מפיקסלי התמונה.  
לכן, עבור תמונות אלו ביצענו פעולה מורפולגית של הרחבה , כפי שלמדנו בענ"ת.   
הגרעין שלקחנו היה אליפסה פחוסה כגון:

הבחירה נבעה מצורתו הממוצעת של הבחור  
על פני הframes השונים אשר בקירוב   
כלוא בתוך אליפסה הדומה לציור.

כמובן שלאחר פעולה זו יתווספו עוד פיקסלי רקע מיותרים אבל לפחות לא חתכנו את האובייקט(כמו בתמונה מלעיל). בנוסף מדובר בכמות frames קטנה הסובלת מבעיה זו – לכן העין האנושית כמעט ולא תבחין בבעיה בנגן MP4 המריץ 24 frames בשניה.

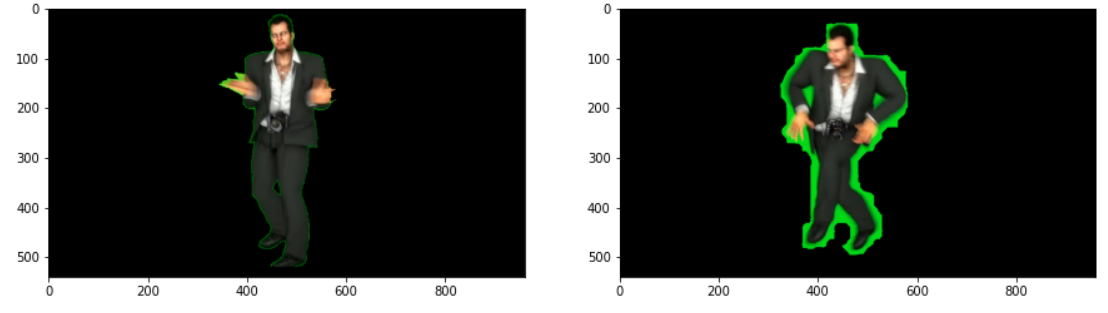
עבור הדוגמה מלעיל, לאחר הרחבה ,נקבל: 

סעיף 3

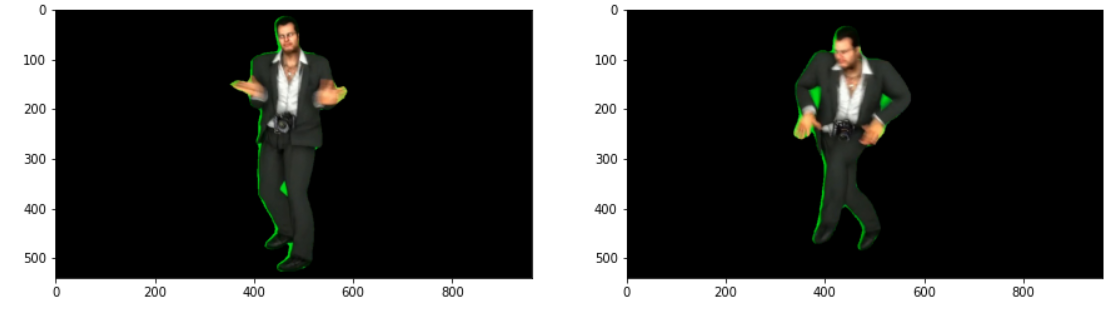
בחרנו בסרטון הרקדן.

בתחילה המחשבה היתה להשתמש בסגמנטציה קלאסית של MRF כפי שלמדנו בהרצאה. המוטיבציה נבעה מהעובדה כי הרקע הוא ירוק בלבד ולכן ישנה הפרדה ברורה בין הרקע לרקדן. האלגוריתם מנסה למזער פונקציית ארגיה המבטאת סטנדרטיזציה בין הפיקסלים בתמונה ויחס בין פילוג גאוסי של הרקע לפילוג הגאוסי של האובייקט. כפי שראינו בלומדות, מודל gaussian mixture שכזה אינו מבטיח הפרדה מדוייקת בין אובייקט לרקע, בין היתר – עקב מחוייבות לשמירה על פילוג חוקי.

ואמנם קיבלנו תוצאות עם רקע ירוק ברור:

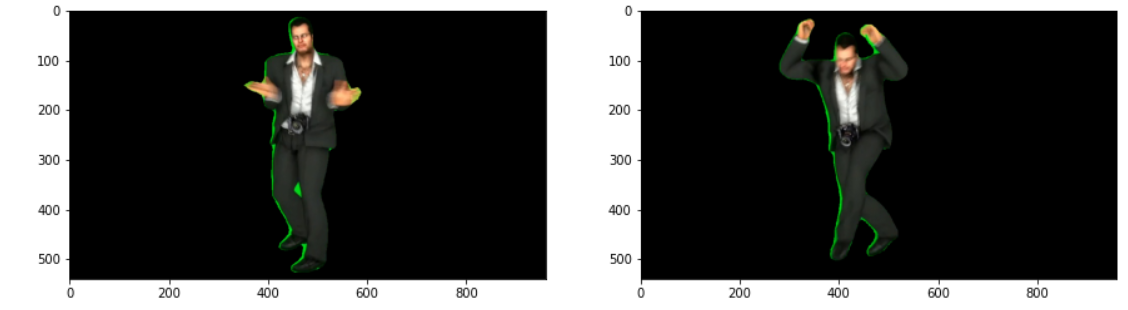


לכן, חזרנו לאופציית הסגמנטציה "העמוקה". הפעלנו את אותו המודל מסעיף 2. גם מודל זה לא הוציא לגמרי את הרקדן ללא רקע ירוק, אך עם תוצאות טובות יותר מאשר השיטה הקלאסית:



בניגוד לסעיף 2, עבור רקע ירוק מונוטוני זה, לרשת לא היתה בעיה עם קליטה מלאה של כלל האובייקט. כלומר היא לא חתכה חלק ממנו בניגוד לDJ מסעיף 2.

לכן, במקרה זה, כדי לנסות לאפטם את תוצאת הסגמנטציה – הפעלנו פעולה שחיקה עם גרעין 5X5. כאמור קיבלנו את כל האובייקט בתוספת רקע בניגוד לסעיף 2 ששם חלקים עיקריים מהאובייקט לא הופיעו, לפיכך הדבר הטבעי היה לשחוק מעט את הדמות שזוהתה. הקפדנו על גרעין קטן כדי לא להרוס את מראה הרקדן.

לפני שחיקה: 

אמנם קיבלנו תוצאות טובות יותר בהשוואה למקרה ללא שחיקה:

תמונה שמכילה טקסט, צג, צילום מסך, מחשב

התיאור נוצר באופן אוטומטי

(\*) קוד הMRF שהרצנו(ואותו כאמור- לא בחרנו): תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

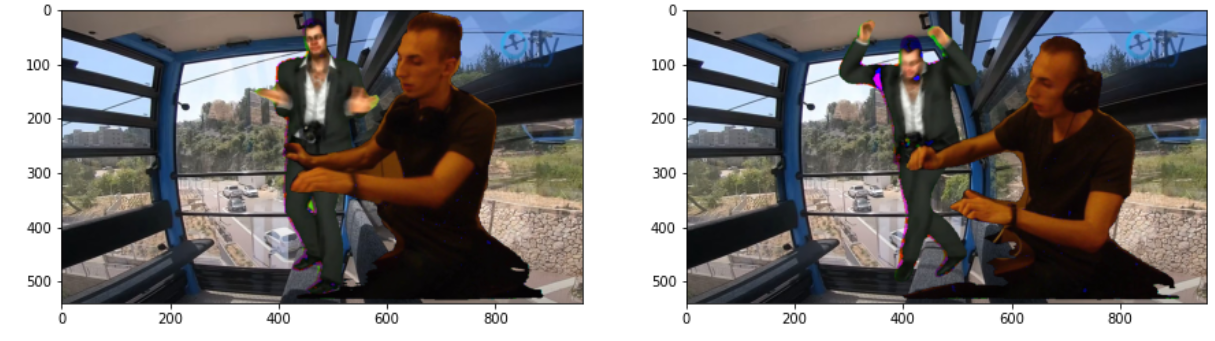
תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

סעיף 4

ראשית ביצענו התאמה למספר הפריימים. ראינו כי מספר הframes של הDJ הוא יותר גדול ולכן העדפנו להוריד מהDJ מספר frames לא מוצלחים של סגמנטציית רשת (בנוסף לשחיקה שביצענו).

בנוסף, ביצענו התאמת מימדים בדומה לסעיף 2.6 כך שהרקדן יוכל לעמוד על מושב הרכבל, והDJ אשר רגליו לא נראות גם בסרטון המקורי – ממקום קרוב לחזית עדשת המצלמה (כדי ליצור את התחושה שהמצלמה מפספסת את רגליו ולא כי רגליו פשוט אינן נמצאות).

לבסוף יצרנו את הסרטון בעזרת הפונקציה המשורטטת ו**המצאת תיקיה חדשה בmy\_data (mixImageQ3 ).**