

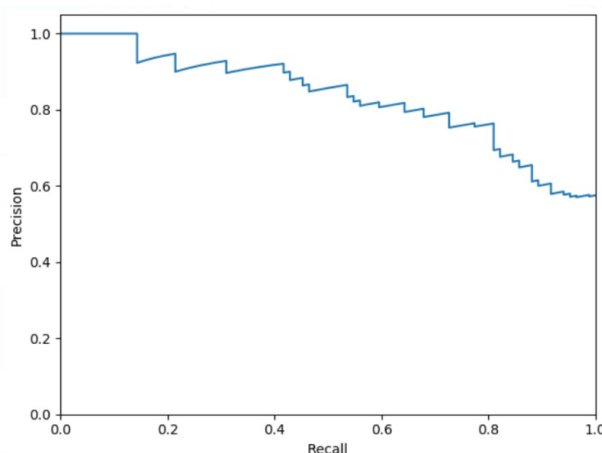
:Task Transfer

בפרויקט זה ביצענו Task Transfer לסיווג פרחים. הרשת שבעזרתה ביצענו את שינוי המשימה הינה ResNet50v2, שאומנה ובנתה לסיווג 1000 מחלקות, ובעזרתה נסווג באופן בינארי – פרח (1) או לא פרח (0). בכדי להתאים את המודל המאומן של ResNet50v2 לסיווג בינארי, הסרנו את השכבה האחרונה בעלת 1000 נירונים, ועל השכבה הלפני אחרונה (שכבת קונבולוציה $7*7*2048$) ביצענו average pooling, וקיבלנו וקטור בגודל 2048 (שכבת הavg_pool) שמחובר לנורון יציאה בודד. את השכבות שנלקחו ממודל ה ResNet50v2 קיבענו בכדי שלא יילמדו, ותהליך הלמידה בוצע רק ע"י שינוי משקולות השכבה החדשה שהוספנו. בסך הכל נלמדו ברשת הבסיסית של שינוי המשימה 2049 משקולות (2048 משקולות לavg_pool ו1 נוסף לחותך). (האימון בוצע לפי ה-SGD optimizer עם 10 epochs batch size 32).

בשלב ההתחלתי הרצנו את הרשת הבסיסית עם ההתאמה שתוארה. התקבלו התוצאות הבאות:

Test Error 0.407

:Recall precision curve



:Model Improvements

בכדי לשפר את אחוזי הדיוק על סט הבחינה, ביצענו מספר שינויים, חלקם כוללים שינויים במבנה הרשת, חלקם על הפרמטרים המאפיינים את הרשת וחלקם מאופיינים בהגדלת סט האימון.

שיפור מספר 1 Data augmentation

השינוי הראשון שבחרנו לבצע הוא הגדלת סט האימון. את ההגדלה מבצעים לא בעזרת לקיחת תמונות נוספות, שכן יש בידנו מספר סופי של תמונות איתן נצטרך לעבוד, אלא בעזרת יצירת גרסאות שונות שנגזרות מהתמונות המקוריות. לאחר חלוקה לסט אימון וסט בחינה (לפי החלוקה הנדרשת בהוראות), לקחנו את סט האימון ועבור כל תמונה בו הגרלנו שינוי מבין השינויים הבאים (הנהוגים והמקובלים לשימוש במשימות דומות):

1. הוספת רעש רנדומלי לתמונה. (לכל תמונה הגרלנו שונות רעש אחרת בין 0.001 ל0.01).
2. סיבוב התמונה (לכל תמונה הגרלנו זווית סיבוב אחרת בין 10 ל-90 מעלות).

בסט האימון החדש שמרנו את התמונות המקוריות ואת התמונות לאחר השינוי שהוגרל, כך הכפלנו את סט האימון פי 2 ($n=600$). לאחר ביצוע שינוי זה קיבלנו את שגיאת הבחינה הבאה:

Test Error 0.372

שיפור מספר 2 Max pooling & Data augmentation

בשינוי זה בדקנו שילוב של הגדלת סט האימון באותו אופן שתואר בשיפור הקודם, רק שכעת המודל שלנו מבצע max pooling במקום avg pooling בשלב שינוי המשימה. בחרנו לבדוק max pooling משום שבעזרתו יכולים לבוא לידי ביטוי באופן מובהק יותר פיצ'רים שונים (כמו קווי מתאר), שעלולים לדעוך כאשר עושים שימוש בaverage pooling. קיבלנו ששינוי זה עדיף על שינוי שכולל data augmentation בלבד. וכעת שגיאת הבחינה הינה:

Test Error 0.3488

שיפור מספר 3 הגדלת מספר שכבות נלמדות

במודל הראשוני, קיבענו את כל השכבות ולא למדנו את הפרמטרים שלהן (מלבד הפרמטרים בין שכבת הוקטור בגודל 2048 לנוירון היציאה). כעת בדקנו למידה של שכבות נוספות, כאשר בדקנו בצעדים את הוספת השכבות הנלמדות מהסוף. במודל המשופר שמוצע בחלק זה, בחרנו ללמוד עוד 3 שכבות (סך הכל את 4 השכבות האחרונות – מתוכן 2 שכבות שיש להן פרמטרים נלמדים), סה"כ פרמטרים נלמדים 6145 לעומת 2049 ברשת הבסיסית. טעות הבחינה במודל זה יצאה 40.1, לכן בדקנו אופציה להוסיף שכבות נוספות וכעת בחרנו לאמן 9 שכבות, סה"כ 3,417,089 משקולות והשגיאה על סט הבחינה המתקבלת כעת הינה 38.4%.

שיפור זה לא נראה מועיל יותר משאר השיפורים, שכן גם לאחר הוספת מספר רב של שכבות, הוא פחות עדיף משיפורים קודמים שנבחנו. ובנוסף, הפך את המודל להיות מורכב יותר ללמידה, וחשוף ל-overfit.

Test Error 0.384

שיפור מספר 4 – Tuning parameters

בכדי לבצע את שיפור זה, ביצענו תחילה כיוון פרמטרים עם סט האימון הרגיל, של 300 תמונות, אותו חילקנו לסט ולידציה וסט אימון (לפי חלוקה של 20%, 80%), ובחרנו את הפרמטרים שמזערו את הטעות על סט הולידציה. ביצענו כיוון בצורה איטרטיבית (כדי לאפשר לפרמטרים לשנות את ערכם מספר פעמים בהתאם לשינויים של פרמטרים אחרים), כך שבכל איטרציה עברנו על כלל הפרמטרים, ומצאנו מהו ערך הפרמטר האופטימאלי (מתוך טווח שהגדרנו מראש) הממזער את השגיאה על סט הולידציה, בהינתן סט פרמטרים מסוים (כל שאר הפרמטרים האחרים נשארים קבועים בכיוון פרמטר מסוים). הגבלנו ל-5 איטרציות ועצרנו במידה ובאיטרציה הנוכחית לא היה שינוי בסט הפרמטרים (כלומר זוהתה התכנסות). הפרמטרים אותם בחרנו לכיוון הינם:

- Threshold – הסף לפיו מסווגים 1 או 0. ראינו שבמודל ההתחלתי ההסתברויות בנוירון היציאה גבוהות ברוב המקרים, ולכן סף 0.5 לא בהכרח מבצע סיווג בצורה הטובה ביותר. לכן בחרנו לבדוק שינוי ערך זה כך שיוקדם את אחוזי הדיוק על סט הולידציה. במצב הנוכחי, עבור $t=0.5$ קיבלנו שאין טעות מסוג ראשון, כלומר אם המודל אמר "לא פרח" זה תמיד "לא פרח" ולכן למרות שלא מדובר בפרמטר סטנדרטי לכיוון, כן מצאנו לנכון ששינה מוטיבציה מספקת לנסות סף החלטה גבוה יותר.

שינוי הערכים הדיפולטיביים של ה-optimizer מסוג SGD:

- Learning Rate – פרמטר שמשפיע על גודל צעד הלמידה. ערך קטן מידי גורם לתהליך למידה ארוך מדי, וערך גדול מידי יגרום לתהליך למידה לא יציב.

- Decay – פרמטר שמונע מהמשקולות לגדול יותר על המידה, ומעניש על משקולות גדולות.
- Momentum – התחשבות בכיוון הקודם אליו הלכנו. היות והלימוד נעשה לפי batch אז נרצה לשמור מידע על הכיוון האחרון אליו הלכנו, ובכך לא להסתמך רק על כמות התצפיות בbatch הנוכחי.

לאחר התהליך האיטרטיבי מצאנו את הערכים האופטימאליים הבאים:

פרמטר אופטימאלי(הממצא שגיא על סט ולידיה)	טווח ערכים נבדק	פרמטר נבדק
0.8	(0.5,0.6,0.7,0.8,0.9)	Threshold
0.01	(0.0001,0.001,0.01,0.1,1)	Learning rate
0.1	(0,0.001,0.01,0.1,1)	Decay
0.5	(0,0.1,0.3,0.5,0.7,0.9)	Momentum

הפקנו גם גרף עבור כל אחד מפרמטרים אלו שמראה את שגיאת הולידציה עבור טווח הערכים הנבדק וכך גם היה ניתן לראות את הערך שמזער את השגיאה, אך הייתה טעות דקדוק בפונקציית ההדפסה, ולכן היא נעצרה. כלומר כן קיבלנו את הערכים המיטביים (כי הודפסו לפני), אך את הגרפים לא. לא ביצענו בשנית משום שזמן הריצה של הכיוון היה ארוך במיוחד (בסביבות ה-30 שעות) עקב בדיקת אפשרויות רבות של שילוב ערכי פרמטרים.

לאחר שמצאנו את הפרמטרים הטובים ביותר בשלב הכיוון, הרצנו את המודל וקיבלנו שגיאת בחינה:

Test Error 0.3258

שיפור מספר 5 – פרמטרים אופטימאליים + Data augmentation

לאחר שראינו שכיוון פרמטרים אכן מוריד את השגיאה על סט הבחינה, בחרנו לבדוק שילוב של הרצת מודל זה עם הפרמטרים הנבחרים יחד עם data augmentation שפורט בשיפור מספר 1.

עבור מודל זה קיבלנו:

Test Error 0.2558

סיכום שיפורים שנבחנו:

Test Error	תיאור	שיפור
0.407	Basic CNN	0
0.372	data augmentation	1
0.348	data augmentation & max pooling	2
0.384	more trained layers	3
0.325	tuning parameters	4
0.255	Best parameters & data augmentation	5

המודל הטוב ביותר הוא מודל שיפור מספר 5, שבו בחרנו להשתמש בפרמטרים שנמצאו אופטימאליים בשלב הכיוון (אלו שמזערו טעות על סט ולידיה), בשילוב data augmentation שמפורט בשיפור מספר 1.

תוצאות המודל הנבחר:

שינוי הערכים הדיפולטיביים של ה-optimizer מסוג SGD: Learning rate = 0.01, Decay = 0.1, Momentum = 0.5, שינוי ערך ה-threshold להיות 0.8. בשילוב עם data augmentation – (הגדלת סט האימון) על ידי אחד מהשינויים הבאים לכל אחת מהתמונות:

1. הוספת רעש רנדומלי לתמונה. (לכל תמונה הגרלנו שונות רעש אחרת בין 0.001 ל-0.01).
2. סיבוב התמונה (לכל תמונה הגרלנו זווית סיבוב אחרת בין 10 ל-90 מעלות).

עבור מודל זה קיבלנו טעות הבחינה הבאה:

Test Error 0.255

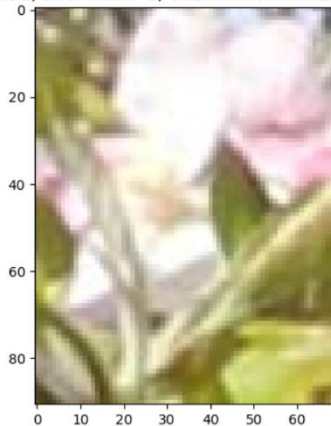
5 הטעויות הכי גדולות מסוג ראשון:

הערה: כל התמונות המוצגות הן התמונות כפי שהן בצורתן המקורית, לפני כל שינוי שבוצע בתהליך האימון.

טעות מסוג ראשון זו טעות בה סיווגנו לא פרח (0) ומדובר בפרח (1). לכן, מבין אלו, שגיאה גדולה ביותר הינה שגיאה שבה classification score הוא הנמוך ביותר.

```
Error Type 1, Error Index: 1, Image Name: 306, Classification Score: [0.3860206]
Error Type 1, Error Index: 2, Image Name: 376, Classification Score: [0.4458927]
Error Type 1, Error Index: 3, Image Name: 340, Classification Score: [0.56953764]
Error Type 1, Error Index: 4, Image Name: 335, Classification Score: [0.57142097]
Error Type 1, Error Index: 5, Image Name: 434, Classification Score: [0.66660833]
```

Error Type 1, Error Index: 1, Classification Score: [0.3860057]



Error Type 1, Error Index: 2, Classification Score: [0.44590455]



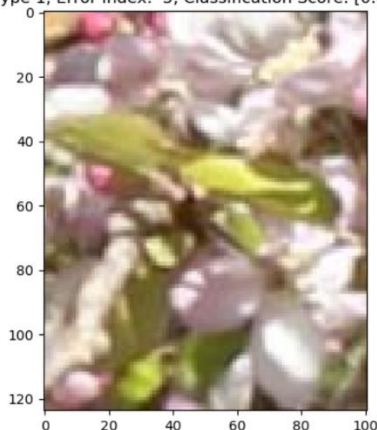
Error Type 1, Error Index: 3, Classification Score: [0.5695721]



Error Type 1, Error Index: 4, Classification Score: [0.57145447]



Error Type 1, Error Index: 5, Classification Score: [0.66655755]

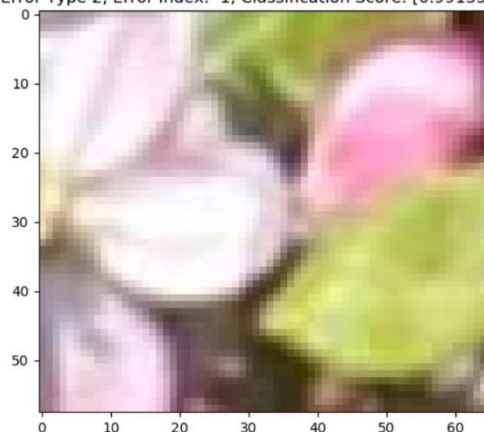


5 הטעויות הכי גדולות מסוג שני:

טעות מסוג שני זו טעות בה סיווגנו פרח (1) ולא מדובר בפרח (0). לכן, מבין אלו, שגיאה גדולה ביותר הינה שגיאה שבה classification score הוא הגבוה ביותר.

```
Error Type 2, Error Index: 1, Image Name: 352, Classification Score: [0.99133575]
Error Type 2, Error Index: 2, Image Name: 469, Classification Score: [0.9904426]
Error Type 2, Error Index: 3, Image Name: 455, Classification Score: [0.9870297]
Error Type 2, Error Index: 4, Image Name: 387, Classification Score: [0.98552394]
Error Type 2, Error Index: 5, Image Name: 407, Classification Score: [0.9772849]
```

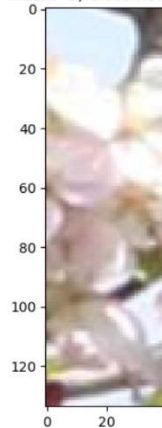
Error Type 2, Error Index: 1, Classification Score: [0.99133575]



Error Type 2, Error Index: 2, Classification Score: [0.9904426]



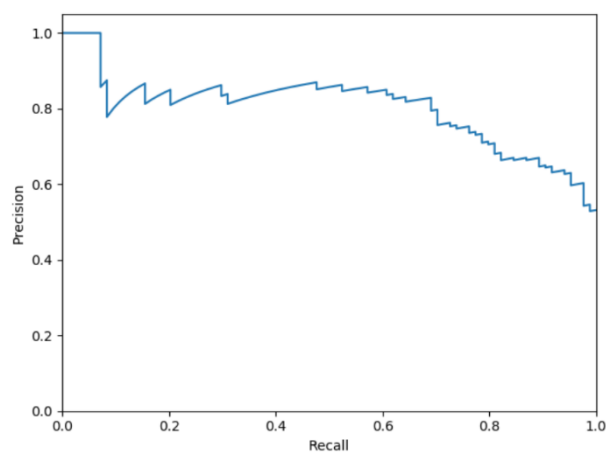
Error Type 2, Error Index: 3, Classification Score: [0.9870297] Error Type 2, Error Index: 4, Classification Score: [0.98552394]



Error Type 2, Error Index: 5, Classification Score: [0.9772849]



עקומת recall-precision curve:



עקומת זו מציגה את ה-recall וה-precision כתלות בthreshold שנע בין 0 ל1. עקומה זו יכולה לעזור לבחור את ה-trade off בחשיבות בין שני מדדים אלו שמציגים דיוק מודל בשני מישורים שונים, האחד (recall) מציג את אחוז האובייקטים שסווגו כפרחים, מתוך כל הפרחים האמיתיים, והשני (precision) מציג את אחוז הפרחים האמיתיים מתוך כל סיווגי הפרחים.