

תרגיל בית 2

מגישים :

רם יזדי 305246308

גיא לורברבום 302773338

ספי לנגמן 305183527

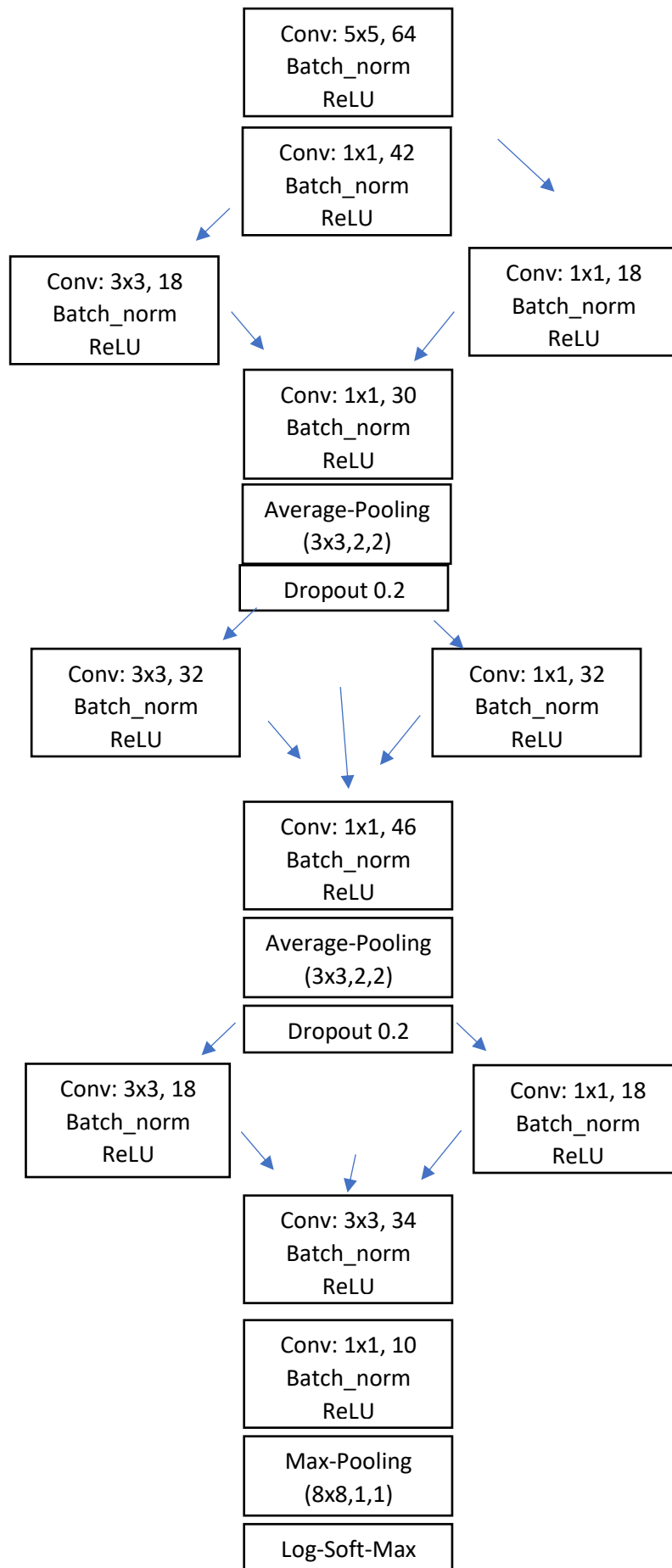
קישור לקובץ t7 :

https://drive.google.com/open?id=0B4kqDb_VN_TMTViSkxRaXY3aUE

קישור לgithub :

<https://github.com/sefilang/machine-learning>

ארכיטקטורת המודל:



תיאור המודל:

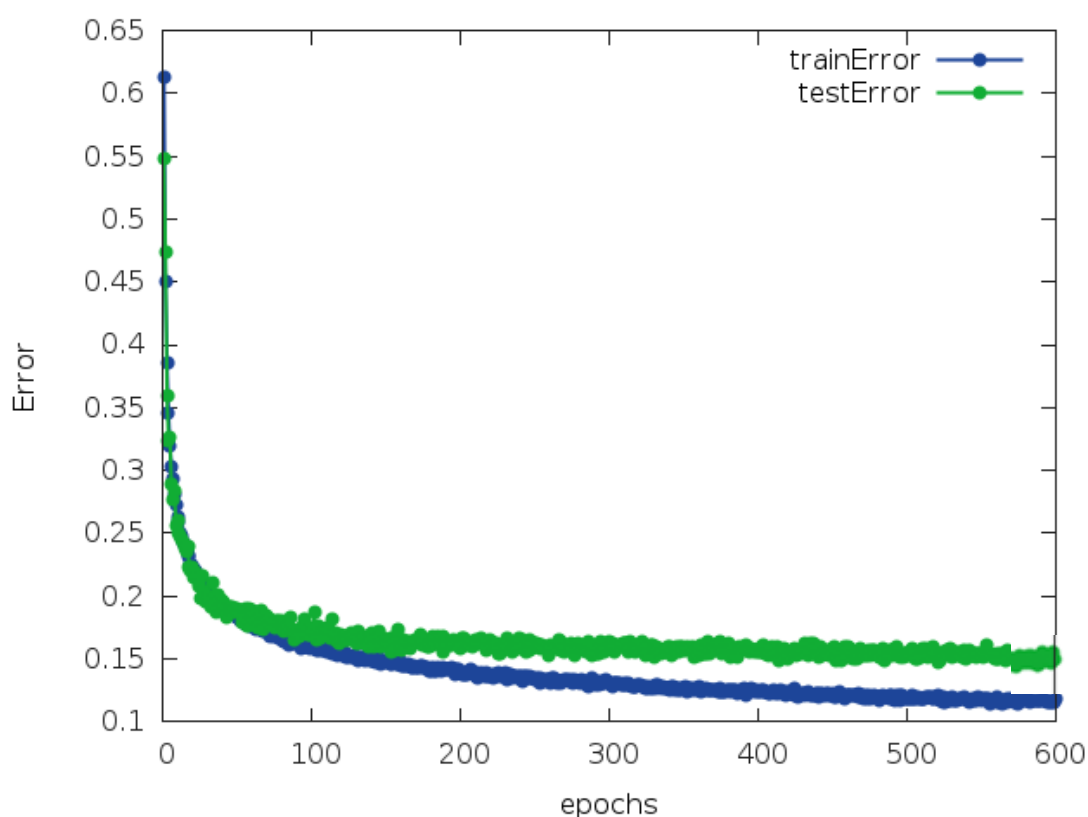
הרשת הנ"ל מכילה 9 שכבות קונבולוציה כאשר 3 שכבות מתוכן מורכבות משרשור של שתי קונבולוציות בעלות פילטרים בגדלים שונים. הורדנו את מימדי התמונות במהלך הרשת באמצעות שכבות Max/Average pooling בלבד כאשר בשכבות הקונבולוציה ריפדנו כך שמימדי הקלט של השכבה יהיו זהים למימדי הפלט. לצורך הימנעות מהתאמת-יתר השתמשנו במהלך האימון בשכבות Dropout עם הסתברות של 0.2 לכבות ניורון.

כמו כן, לצורך אוגמנטציה השתמשנו בהיפוך התמונה בצורה אופקית כך שלכל תמונה במהלך האימון הייתה הסתברות $1/3$ להתהפך. לצורך אוגמנטציה השתמשנו בהיפוך התמונה בצורה אופקית (hflip) כך שלכל תמונה במהלך האימון הייתה הסתברות $1/3$ להתהפך. בנוסף, בטרם תחילת האימון אתחלנו את הקשתות לפי התפלגות נורמלית עם תוחלת 0 ושונות 0.25 ואת ה bias ל-0.

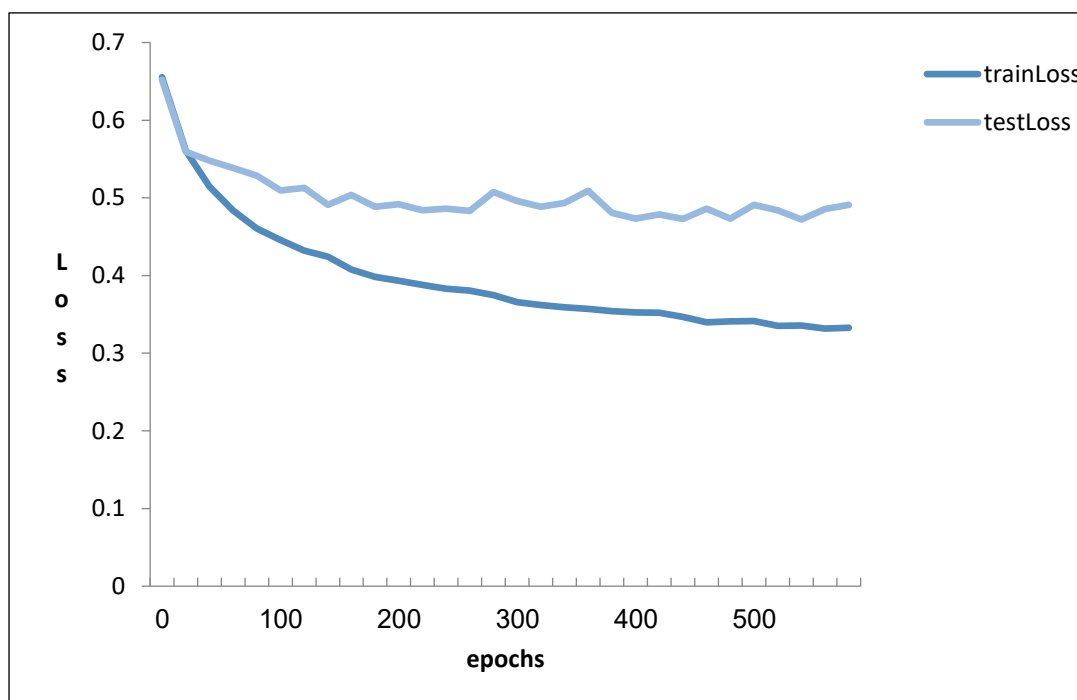
כמו כן, נרמלנו את מדגם האימון לפי כל ערוץ בנפרד. פונקציית המטרה (קריטריון) היא Log-loss (ClassNLLCriterion) והאופטימיזר הוא adam. מספר epochs שבחרנו עבור המודל הוא 600, וכן בחרנו שהמודל יתאמן על המדגם ב-batch בגודל 128. מספר הפרמטרים שהתקבלו הינו 49334.

ניתוח המודל:

להלן גרף המציג את טעות האימון והמבחן:



ניתן לראות כי השגיאה על מדגם האימון יורדת ככל שכמות ה-epoch-ים עולה.
 בנוסף, מ-epoch מספר 400 השגיאה על מדגם המבחן מתנהגת בצורה מונוטונית.
 להלן גרף המציג את ה-loss על מדגם האימון ומדגם המבחן : (גרף זה נכתב ידנית מהפלט שקיבלנו לקובץ out):



מסקנות וניסיונות:

בתחילת התהליך, בדקנו כיצד לאתחל את קשתות המודל בצורה היעילה ביותר- ניסינו לאתחל עם ערך אפס, וכן עם ערכים מהתפלגות נורמלית. גילינו כי האתחול השני תורם לקצב התכנסות המודל.

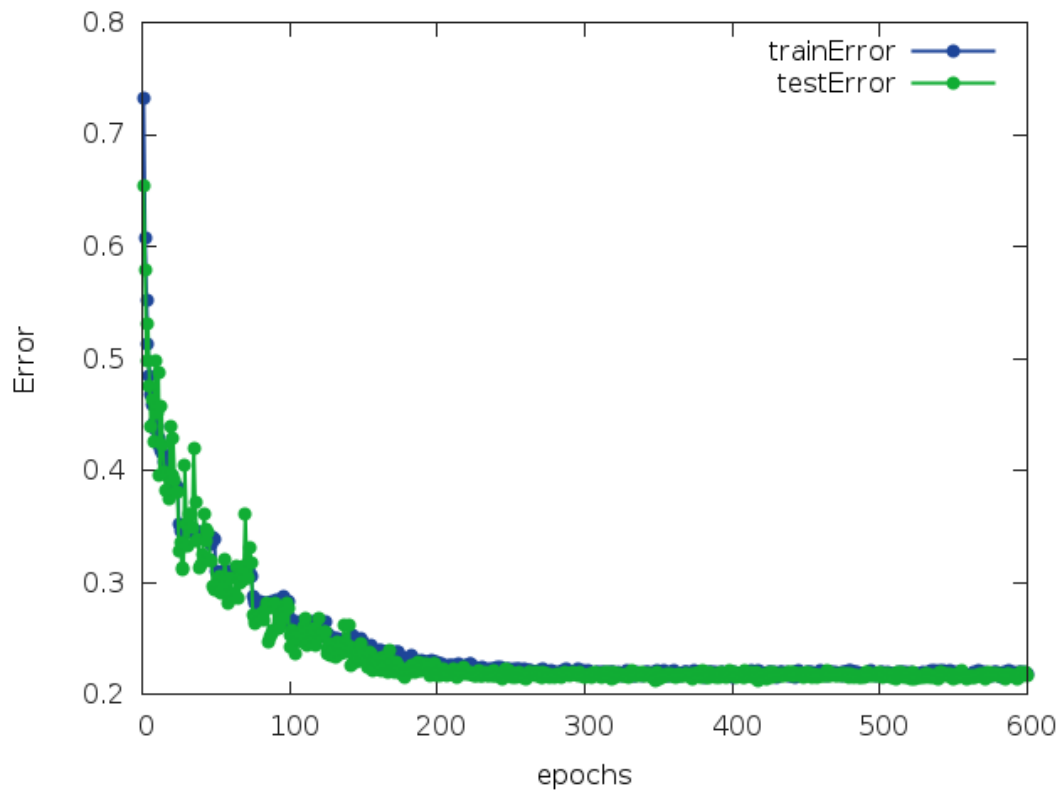
בהמשך, ניסינו לממש ארכיטקטורות שונות ביניהן: שינוי במספר הקונבולוציות כטרייד אוף בהוספת שכבה לינארית בסוף הרשת (הוספת שכבה לינארית העלתה את מספר הפרמטרים באופן ניכר).

מסקנתנו הייתה שהורדת השכבות הלינאריות מהמודל ועל ידי כך אפשרות להוסיף קונבולוציות נוספות (תוך הורדת ממדי התמונות ל-1X1) הגדילה את דיוק המודל מסדר גודל של 77% ל-84.5%.

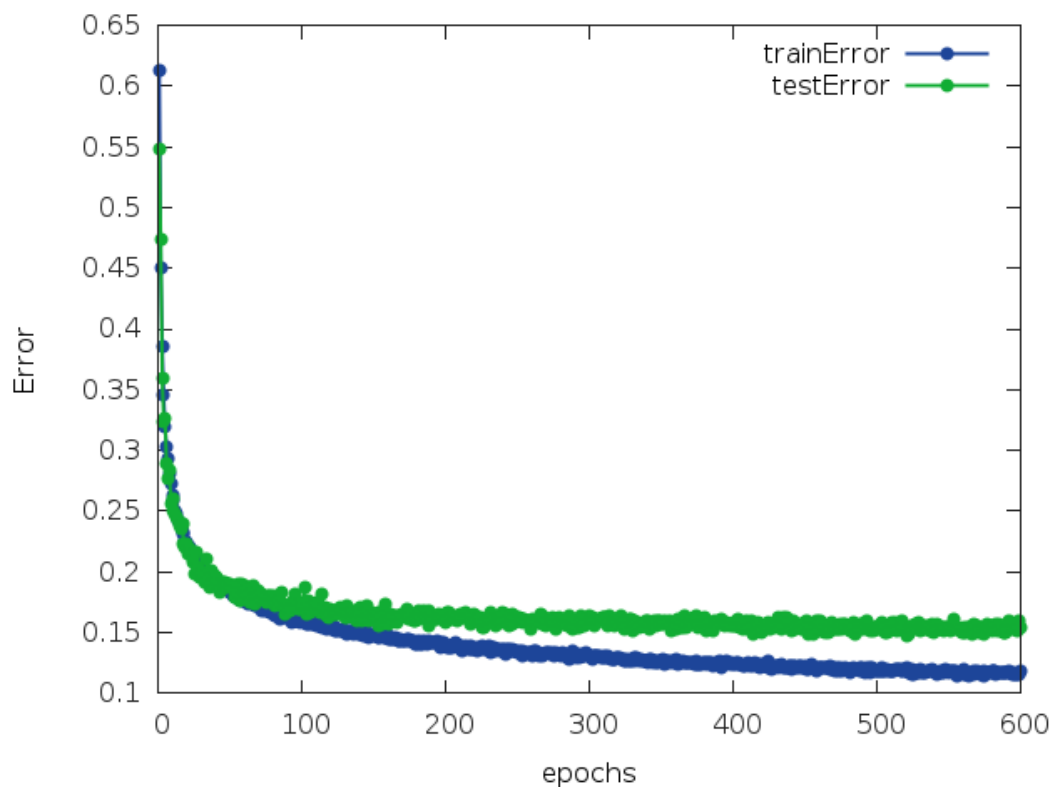
בנוסף, בארכיטקטורה דומה לזו שהגשנו אך ללא שימוש בשכבה מקבילית התקבלו תוצאות פחות טובות (83%).

במהלך ניסיונותינו, הבחנו כי המודל מרבה לטעות על תמונות של חיות לעומת שאר הלייבלים (ככל הנראה בגלל שחיות יכולות להופיע בצורות מגוונות יותר). בעקבות כך, ניסינו בשלב האוגמנטציה לבצע הטייה (rotate) לתמונות בעלות התיוג הנ"ל. ניסיון זה גרע מהדיוק ולכן מסקנותינו שלא כל אוגמנטציה יכול לתרום לתהליך האימון.

- במהלך ניסיונותינו השתמשנו בשני פונקציות אופטימיזציה שונות - adam ו sgd :
1. עבור sgd הדיוק הטוב ביותר שהתקבל הוא 78.5 לאחר אימון של epoch 200 – ים.
 2. עבור adam הדיוק הטוב ביותר שהתקבל הוא 84.5 לאחר אימון של epoch 500 – ים.
 3. להלן גרף המציג את טעות מדגם האימון והמבחן בעבור פונקציית sgd :



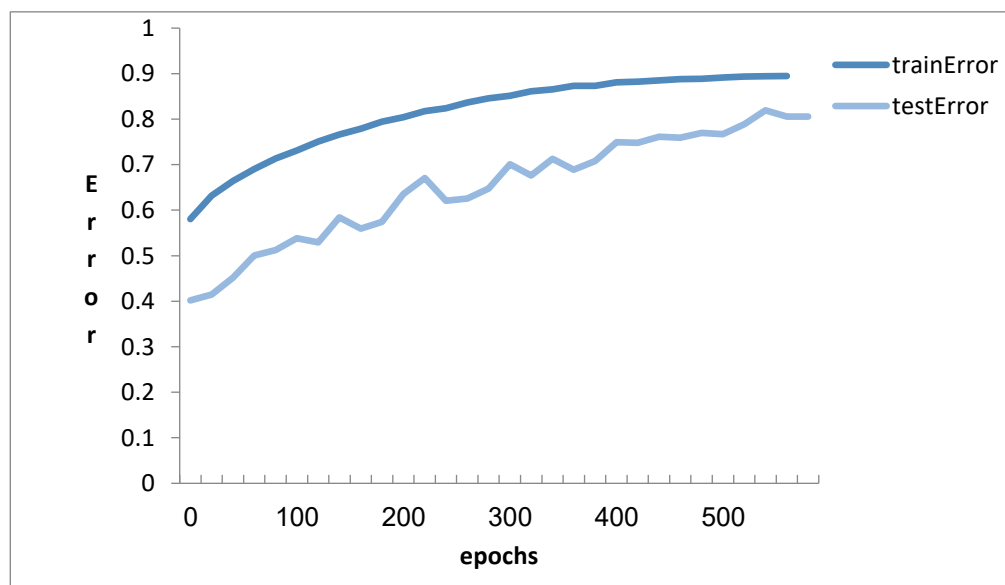
להלן גרף המציג את טעות מדגם האימון והמבחן בעבור פונקציית adam :



למרות שבאמצעות sgd מתקבל מודל יציב יותר (שגיאות המבחן והאימון דומות) התוצאות של פונקציית adam הינן טובות יותר וכן ניתן לראות כי למרות ההבדל בין שגיאות האימון לשגיאות המבחן אין הדרדרות בתוצאות שגיאת המבחן ולכן ניתן להסיק שאין overfitting.

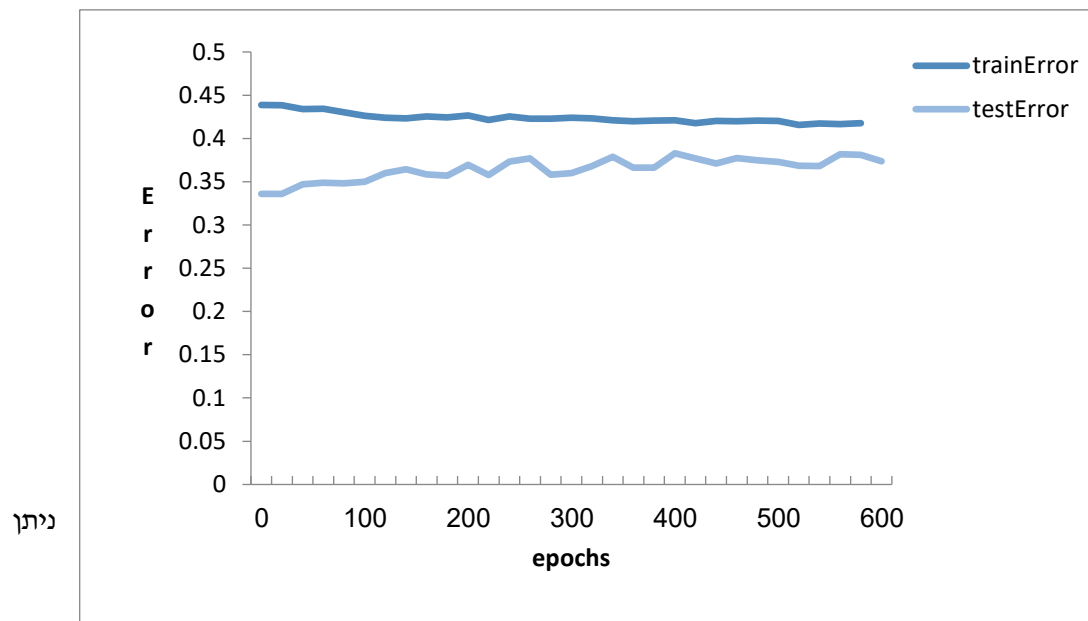
שיטות אוגמנטציה:

1. אוגמנטציה המכילה קומבינציה של hflip, crop עם ריפוד באפסים, crop עם ריפוד בהשקפות, rotate (רק על מחלקות החיות) בהסתברויות שונות. הדיוק ב- epoch החמישי היה מסדר גודל של 60, והדיוק הלך וירד עד שלאחר epoch 300-ים הגיע לתוצאה של 19.441.



ניתן לראות כי שגיאת האימון גבוהה יותר משגיאת המבחן לאורך כל ההרצה. דבר זה עומד בניגוד לעיקרון הלמידה וככל הנראה נגרם עקב ריבוי אוגמנטציה (האוגמנטציה נעשית על מדגם האימון בלבד).

2. אוגמנטציה המכילה קומבינציה של `hflip`, `rotate` (רק על מחלקות החיות) בהסתברויות שונות. הדיוק הטוב ביותר שהתקבל הינו 62.64% לאחר כ-epoch 580-ים.



להסיק כי אוגמנטציית הcrop אכן פגמה בתצורת המודל.

3. אוגמנטציית המודל שבחרנו- מכילה אך ורק `hflip` בהסתברות 1/3 עבור כל תמונה. הדיוק הטוב ביותר שהתקבל הינו 84.5% לאחר epoch 600-ים. (הגרף הרלוונטי מוצג למעלה). מסקנתנו היא שיש להשתמש באוגמנטציה במידה נכונה תוך כדי הכרות עם datan אתו אנו עובדים.