

## Almog Ben Simon — ML2

### ארכיטקטורת המודל:

שכבה ראשונה – שכבת קונבולוציה בגודל של 3,16 עם kernel size של 5 ו padding של 2, לאחר הקונבולוציה ביצעתי maxpooling בגודל 2, ולאחר מכן אקטיבציית relu על הוקטור ולבסוף נרמול.

שכבה שנייה – שכבת קונבולוציה בגודל של 16,32 עם kernel size של 5 ו padding של 2, לאחר הקונבולוציה אקטיבציית relu על הוקטור, ואז maxpooling בגודל ולבסוף נרמול. (כמו הראשונה רק בסדר שונה)

שכבה שלישית – שכבת קונבולוציה בגודל של 32,16 עם kernel size של 5 ו padding של 2, לאחר הקונבולוציה אקטיבציית relu על הוקטור.

שכבה רביעית – שכבת קונבולוציה בגודל של 16,32 עם kernel size של 5 ו padding של 2, לאחר הקונבולוציה maxpooling בגודל 2, ואז אקטיבציית relu על הוקטור, ולבסוף נרמול, רק שלאחר הנרמול ביצעתי AdaptiveAvgPool בגודל 2.

שכבת fully connected – לאחר שכבות הקונבולוציה, צריך שכבה לסיווג התמונה בגודל 128 על 10 (כגודל המחלקות).

שכבת optimizer – Adam עם learning rate of 0.01.

Cross Entropy Loss – loss

batch size – 100, epochs – 50 אך כדי לחסוך זמן ריצה ביצעתי תנאי שכאשר עובר את ה80 אחוז תעצור אז בפועל יצא 21 אפוקים.

מספר הפרמטרים של המודל הינו 41146.

### אימון המודל:

תהליך האימון שביצעתי הינו תחילה רשת קונבולוציה בסיסית עם פחות מ50,000 פרמטרים. ראיתי כי אני מצליח לתוצאות גבוהות במיוחד. חקרתי את הדאטה עליו אני עובד והגעתי למסקנה כי התמונות מורכבות ואין מספיק דאטה כדי שהרשת תתפוס

ותסווג טוב את התמונות. לכן תקפתי את הבעיה עם הפתרון של data augmentation, שהינה דרך לייצר דאטה מדאטה קיים בעזרת מניפולציות וטרנספורמציות כדי לשכפל אותו. אם זה לסובב את התמונה, למתוח אותה, להוסיף בהירות וכל מיני משחקים שיוצרות תמונות עם תכונות טיפה שונות מאותן תמונות ובכך אני מגדיל את הדאטה. ביצתי 5 אוגמנטציות, כך שמ50,000 תמונות הכפלתי את הדאטה ל250,000 תמונות. כמובן שכל אוגמנטציה שונה.

בנוסף לאחר קריאה מעמיקה באינטרנט שמתי לב כי בשכבות קונבולוציה כאשר מנרמלים את הוקטור לפני כניסתו לשכבה, מוסיף המון, לעומת ישר אחרי השכבה. כלומר אחרי maxpooling ופעולות שעושים אחרי שכבת הקונבולוציה, נרמול לבסוף משפר משמעותית את התוצאות.

בנוסף חשבתי תחילה כי ביצוע dropout יעזור לתוצאות המודל, אך הגעתי לתוצאות טיפה יותר נמוכות איתו ולכן וויתרתי על החלק הזה (לאחר הfully connected). בנוסף ניסיתי להחליף אופטימיזר sgd אך adam השיג תוצאות טובות יותר.

## דיוק סופי הינו: 80.04%

להלן הגרפים של השגיאות והloss על סט האימון והמבחן: ניתן לראות כי שתיהן מתכנסות וכבר באפוקים הראשונים יש ירידה מאוד חדה בשניהם, וניתן לראות כי שתי הפונקציות מאוד דומות כלומר פונקציות הלזו מתאימה למודל.

