|  |
| --- |
|  |
| ResNet50 – Transfer learning report |
|  |
|  |

Transfer Learning from ResNet50

**פירוט המשימה:**

**לבנות רשת נוירונים שיודעת להבדיל בין100 קלאסים שונים של אתרים בעולם.**

**המשימה התחחלקה לשניים:**

**1. ליצור Deep Neural Network שתדע לסווג את התמונות לקלאסים השונים.**

**2. ליצור מסווג KNN שלפי הפיצ’רים של הרשת העמוקה שייצרנו תמצא את K התמונות הקרובות לה ביותר ותציג אותן.**

**פירוט הניסוי:**

**בעיה זו של הוצאת פיצ’רים מתמונה וסיווג התמונה לקלאס המתאים, היא בעיה דומה מאוד לאתגר ש imageNet מציג שמסווגת בעיקר בעלי חיים צמחים ועצמים דוממים (לא נופים כמו באתגר שלנו).**

**מהסיבה הזו החלטנו שדרך טובה לפתור את האתגר שלנו תהיה ביצוע Transfer Learning מרשת שכבר אומנה באתגר של imageNet. הרשת שבחרנו לצאת ממנה היא רשת ResNet50**

**שלבי ההרצה של ניסוי בודד:**

**- טעינת התמונות של ה TrainingData וטעינת התמונות של ה ValidationData**

**- טעינת המודל של ResNet50**

**- הוספת שכבות ה dense האחרונות בכדי ליצור את הפלט הרצוי (מערך של 100 קלאסים עם ציונים עבור כל אחד מהם לתמונה שהתקבלה)**

**- הגדרת פונקציית ה loss ה optimizer וה Hyper Parameters הרלוונטים לאותו optimizer.**

**(learning rate, momentum וכו’)**

**- יצירת ה callbacks עבור שמירת ההיסטוריה של תוצאות ה loss וה accuracy בסיום כל epoch + שמירת המשקלים.**

**- אימון המודל.**

**- יצירת הגרפים ומדידת תוצאות האימון.**

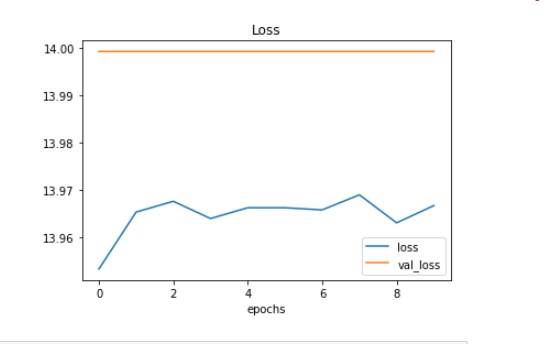
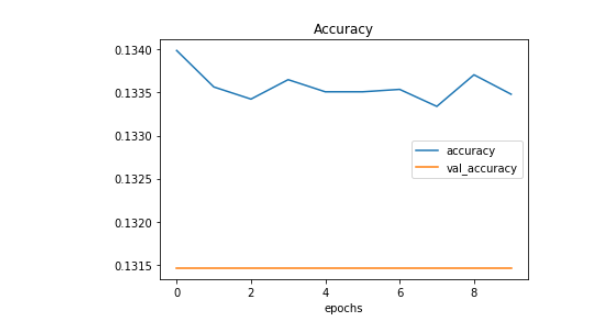
**חלק ראשון קיבוע ה Learning Rate:**

**בשלב זה ננסה למצוא Learning rate מתאים עבור ה optimizer שבחרנו (Stochastic Gradient Descent). על פי מה שנלמד בשיעור Learning rate יהיה בד"כ בטווח הערכים בין 0.00001 ל 1.**

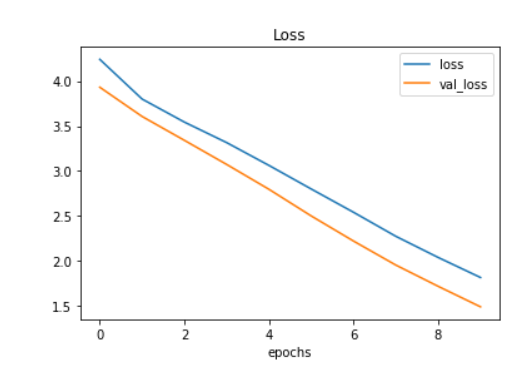
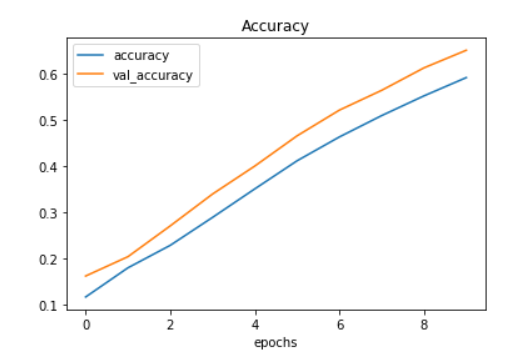
**ניקח את הקצוות תחילה ולאחר מכן ערך באמצע ונראה את תוצאות האימון עבור כל אחד מהערכים.**

**בנוסף ניקח כמות קטנה יחסית של epochs מכיוון שבשלב זה נרצה בעיקר לראות מגמות של איך אנחנו מתפקדים עם learning rates שונים.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.9 | 10 | 1 | SGD - Momentum | categorical\_crossentropy |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.9 | 10 | 0.00001 | SGD - Momentum | categorical\_crossentropy |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.9 | 10 | 0.001 | SGD - Momentum | categorical\_crossentropy |

****

**סיכום תוצאות:**

**- ניתן לראות שכאשר הstep היה גדול מידי (Learning Rate = 1) כל batch שהסתיים הרשת עשתה שינוי גדול מידי על המשקולות. השינוי הזה גרם לכך שנפספס את נקודות המינימום שאנו מחפשים על ה loss. הlearning Rate הגבוה גרם לכך שהצעד שאנחנו עושים בכיוון המנוגד ל Gradient יהיה גדול מידי, מה שגרם לשינוי לא נכון של המשקולות על הרשת.**

**- ניתן לראות שכאשר ה step היה קטן מידי (Learning Rate = 0.00001) האימון על הרשת גם לא הצליח. אך הגרף נראה שונה לגמרי הוא כן מתקדם לכיוון הנכון פשוט בצורה איטית מאוד. ולכן 10 epochs בלבד לא מספיקים לו.**

**- כאשר ערך ה Learning Rate היה 0.01 ניתן לראות שתוצאות האימון לאחר 10 epochs היה לא רע בכלל. הגענו לכמעט90% Accuracy על קבוצת הולידציה שלנו.**

**חלק שני – קיבוע ה batch Size וה Momentum:**

**כעת נרצה לקבע את שני ה Hyper Parameters הנוספים ה Batch Size וה Momentum.**

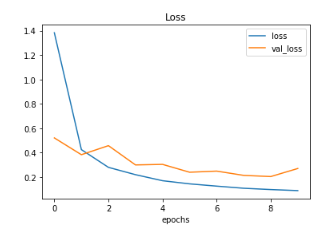
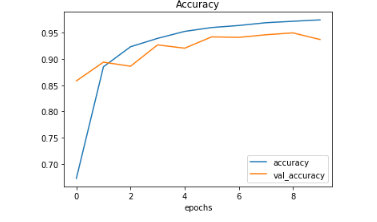
**- לאחר מחקר באינטרנטמצאנו שאת המומנטום יש להגדיר בין 0 ל 1 (על פי הנוסחא המטמטית). בנוסף ראינו שלרוב מגדירים את המומנטום בערך0.9. בניסויים שלנו ננסה את ערך זה ובנוסף ערך קטן יחסית של 0.5**

**- ה Batch Size הוא משתנה יותר עדין ותלוי גם בגודל ה Training Data שלנו. לרוב הוא גם יהיה תחום בין הערך 1 וכמה מאות בודדים. כאשר ערכים נמוכים ייגרמו לכך שאימון הרשת ידרוש פחות Epochs מכיוון שבכל epoch נעשה כמות גדולה יותר של שינויים על המשקולות (שינוי נעשה בסיום כל batch), אך Batch קטן עלול לעשות יותר תנועות בכיוונים לא נכונים. זאת מכיוון שהשינוי שייעשה יהיה על מעט מידי דוגמאות ולא יהיה מייצג. לעומת זאתbatch גדול מידי יאריך משמעותית את זמן האימוןמכיוון שנקבל מעט מאוד שינויים על המשקולות בכל epoch דבר שיידרוש כמות גדולה יותר של epochs מה שיגרור זמן אימון ארוך הרבה יותר. נבדוק את האימון על Batch Size של 10, 16, 32. (ממגבלות של הזיכרון שיש לנו על ה GPU לא נוכל להריץ Batch גדול מ 32)**

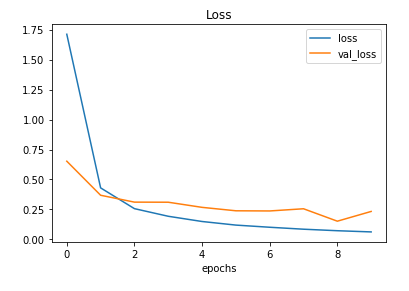
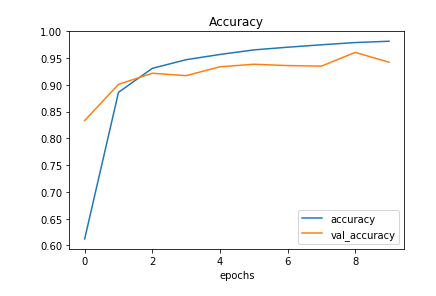
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.9 | 10 | 0.001 | SGD - Momentum | categorical\_crossentropy |



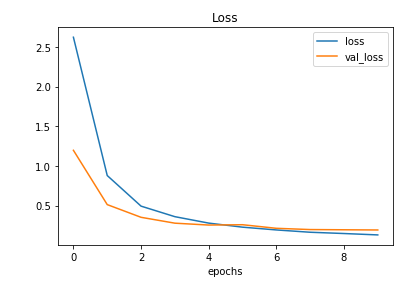
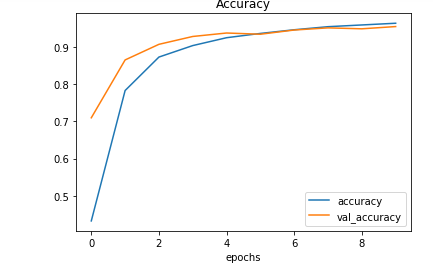
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.9 | 16 | 0.001 | SGD - Momentum | categorical\_crossentropy |



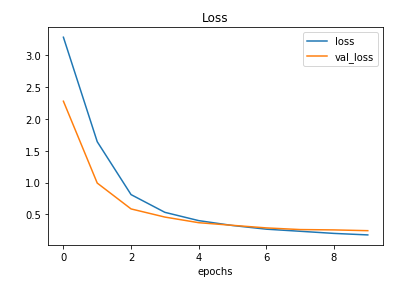
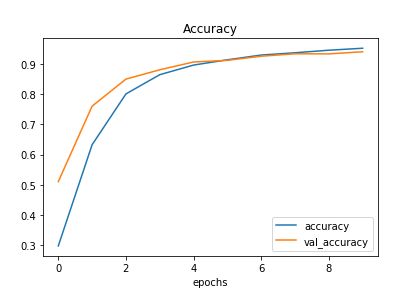
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.9 | 32 | 0.001 | SGD - Momentum | categorical\_crossentropy |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.5 | 16 | 0.001 | SGD - Momentum | categorical\_crossentropy |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 10 | 0.5 | 32 | 0.001 | SGD - Momentum | categorical\_crossentropy |

****

**סיכום תוצאות:**

**- אפשר לראות מגמה מאוד מעניינת בהקשר של המומנטום, שבניסויים בהם המומנטום היה 0.5 המרווח בין ה trainingData ל Validation data היה קטן מאוד כמעט אפסי. לא הצלחתי למצוא צידוקים לזה מהחומרים שקראתי אבל זה לא נראה כמו צירוף מקרים. להערכתי יכול להיות שזה קשור לעובדה שעבור מומנטום קטן אנחנו עושים "צעדים" קטנים יותר בכיו ה Gradient**

**- בנוסף ניתן לראות שעבור המומנטום הגבוה הגרף נע בצרוה יותר חדה ומגיע כבר ב Epochs הראשונים לאחוזי דיוק יחסית גבוהים ולאחר מכן קצב השיפור קטן בצורה חדה.**

**מסקנה: יכול להיות שעבור המומנטום כמו ה Learning Rate נרצה להתחיל את האימון בערך גבוהה יחסית של המומנטום ולאחר מכן כשנרצה לגרד את אחוזיהדיוק הגבוהים נקטין את המומנטום.**

**- עבור גודל ה Batch אפשר לראות מגמה ברורה ולא מפתיעה, שכאשר גודל ה Batch היה מאוד גבוה אימון הרשת היה יותר איטי אבל באופן מובהק אם פחות טעויות, אפשר לראות את זה בצורה החלקה יחסיתשיש לגרפים כאשר גודל ה Batch הוא 32. כאשר גודל ה batch הוא 10 ו 16 קל לראות שהיו epochs שבסוף ה accuracy אפילו ירד.**

**- כמו כן ניתן לראות שעבור Batch בגודל 10 קיבלנו ביצועים פחות טובים בצורה משמעותית. ולכן נשתמש בעיקר בגדלי ה batch של 16 ו -32 בהמשך האימון**

**מסקנה:**

**כצפוי ככל שגודל ה batch היה גדול יותר לקח יותר זמן לאמן את המודל, מכיוון שבכל Epoch עשינו פחות שינויים על המשקולות. מצד שני השינויים היו יותר מדויקיים מכיוון שעשינוו ממוצע של השינוי הנדרש על יותר דגימות.**

**חלק שלישי – אימון הרשת:**

**- כעת נאמן את הרשת עם ערכי ה Hyper Parameters שמצאנו כמתאימים ביותר לאימון הרשת:**

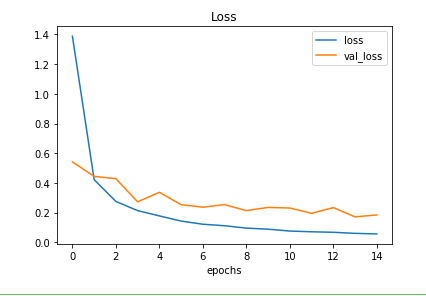
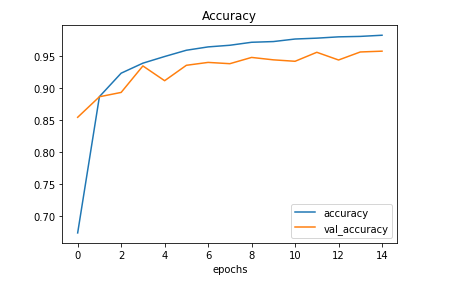
**- Learning Rate = 0.001**

**- Momentum = 0.9**

**- Batch Size = 16**

**את האימון הזה של הרשת נבצע עם כמות גדולה יותר של Epochs**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 15 | 0.9 | 16 | 0.001 | SGD - Momentum | categorical\_crossentropy |



**סיכום תוצאות:**

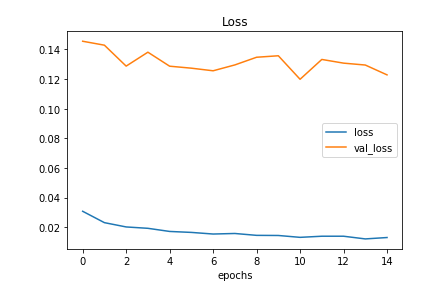
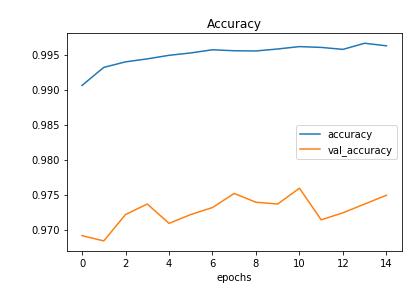
**- ניתן לראות שהגענו לאימון לא רע בכלל של כמעט 95% דיוק על קבוצת ה validation**

**בשלב הבא ננקטין את ה Learning Rate וה Momentum ונגדיל את ה BatchSize מכיוון שבאחוזים הללו יש סיכוי יותר גבוהה להתקדם בכיון הלא נכון**

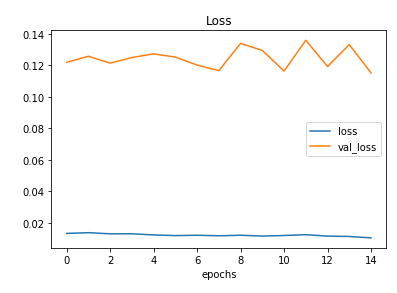
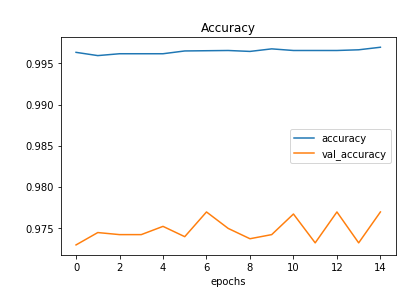
**חלק רביעי – הקטנת ה Learning Rate והמומנטום והגדלת ה Batch Size:**

**כעת הרשת שלנו נמצאת עם למעלה מ 90% accuracy על ה validation שלנו. מה שייגרום לכך שהסיכוי שנעשה צעד בכיוון הלא נכון לגבוה יותר. בשלב זה כדי לאמן את הרשת בצורה טובה יותר נבצע עוד 3 הרצות בהן נקטין בצורה משמעותית את ה Learning Rate שלנו ונגדיל את ה Batch Size ונוריד את המומנטום.**

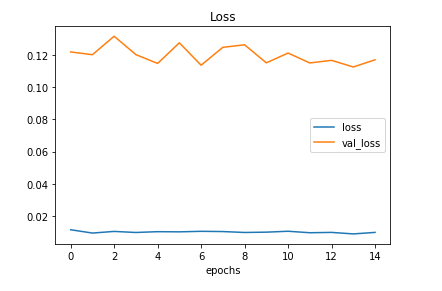
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 15 | 0.5 | 24 | 0.001 | SGD - Momentum | categorical\_crossentropy |

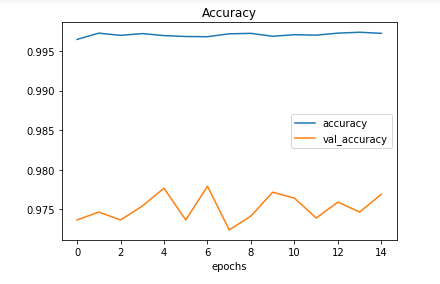


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 15 | 0.5 | 24 | 0.0006 | SGD - Momentum | categorical\_crossentropy |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epochs | momentum | Batch Size | Learning Rate | Optimizer | Loss Function |
| 15 | 0.5 | 32 | 0.0001 | SGD - Momentum | categorical\_crossentropy |

****

****

**סיכום תוצאות:**

**- הגענו לאחר הניסויים לאחוז דיוק מצוין של 97.5% דיוק על ה Validation Data.**