

פרויקט 2 – עיבוד תמונה

קובץ הגדרות לפרויקט

הוספתי קובץ בשם `project_config.py` בו יש הגדרות של הפרמטרים של הפרויקט. ניתן להגדיר את תיקייה ה-`root`, וכמו כן בפרמטרים השונים של האלגוריתמים בפרויקט. למשל -

<code>sift_nfeatures</code>	בוחר את כמות הדיסקריפטורים שיחזרו עבור כל תמונה
<code>n_kmeans</code>	כמות המרכזים באלגוריתם K-means
<code>n_classes</code>	כמות הקטגוריות בדאטה
<code>n_KNN</code>	כמות השכנים שבעזרתם בוחרים את הקלאס באלגוריתם K nearest neighbors

כמו כן, ישנו `dictionary` של פייתון הממיר בין מספר הקלאס לשם הקלאס כמחזורת.

קובץ ראשון – PrepareData.py

הקובץ מכיל קריאה אחת לפונק' `prep_data()` הפונקציה מחשבת את הדיסקריפטורים עבור דאטה ה-`train` ולאחר מכן יוצרת את `voc` ואת `bow`

ניתן לשלוט בריצה של הפונק' בעזרת הדגלים:

```
flag_calc_train_des
flag_calc_test_des
flag_calc_voc
```

על מנת לא לחזור על חישובים מיותרים ומספר פעמים, כאשר רוצים לחשב רק חלק מהנדרש.

כמו כן ניתן להעביר פרמטר `use_subset_factor` כ-`None` או כמספר שלם על מנת לבצע דילול של הדאטה באלגוריתם `kmeans`.

הערה לגבי זמן הריצה:

כאשר מריצים על כל הדאטה במלואו החלק הזה של הכנת הדאטה יכול לקחת קצת זמן. זאת כיוון שכמות הדיסקריפטורים הכוללת היא 762342 וכל אחד באורך שך 128. וביצוע `kmeans` על דאטה גדול כזה לוקח קצת זמן. זו הסיבה שהוצע לנו לדלל את הדאטה עבור החלק הזה (אצלי בעזרת הפרמטר `use_subset_factor`)

קובץ שני – ClassifyData.py

הקובץ הזה מקבל כקלט ניתוב לתיקייה.

ניתן להריץ מה-`CMD` או לשנות בעורך הקוד את הניתוב בהתאם (תחת המשתנה `folder_path`) הקובץ קורא לפונק' `ClassifyImg` שמבצע את הסיווג, ובסופו של דבר מייצר את הקובץ `results.txt`

כמו כן – כתבתי פונק' נוספת שבודקת את הדיוק של המודל שלי עבור כל תיקייה ה-`test` שקיבלנו, ובעזרתה הערכתי את הדיוק של המודל שלי. (על מנת להשתמש בפונק' הזו יש להריץ את `prep_data` גם על תיקיית `(test)`)

תוצאות:

- הרצתי את המודל עם פרמטרים שונים :
- כמות מרכזים Kmeans – 50, 80, 100
 - כמו שכנים KNN – בטווחים של 5-15

רוב ההרצות הניבו תוצאות של סביב ה-40% דיוק בסיווג התמונות.
להלן טבלה מסכמת:

kmmeans	knn	test accuracy
50	5	38.26
50	9	39.70
50	15	39.66
80	3	36.55
80	5	37.62
80	7	38.56
80	9	38.36
80	11	39.56
80	13	39.03
80	15	38.83
80	17	39.33
100	5	37.02
100	7	37.62
100	9	37.72
100	15	38.16

לאור הבדיקות הללו, בחרתי להגיש את הפרויקט עם בחירת הפרמטרים הבאה :
כמות מרכזים Kmeans של 50
כמות שכנים KNN עבור 9