

SUNS - Zadanie 1 Dokumentácia

Funkcionalita celého zadanie bola implementovaná v knižnici MyLibs/data.py ktorej funkcie sa starajú o spracovanie celej databázy.

Po tom ako sa stiahnuté a rozbalené dáta uložia používateľom do priečinku ./data/Test alebo ./data/Traning, program načíta uloží do pamäte všetky obrázky. Túto funkcionality zastrešuje funkcia `cache_images("folder")` kde parameters folder je názov priečinku v ktorom sú obrázky v rámci priečinku data uložené (Test alebo Training).

Následne sa náhodne vyberú 4 obrázky z každého druhu ovocia a zobrazia sa na obrazovku ako test toho, že obrázky boli správne načítané

Po tom čo sa obrázky načítajú a skontrolujú tak sa pomocou $img = img / depth - 0,5$ kde `img` je obrázok načítaný ako matica a `depth` je hĺbka obrázku, v našom prípade 255. Následne sa normalizované obrázky uložia do .pickle súboru tak, aby s nimi v budúcnosti bolo možné pracovať. Pre rýchlu kontrolu sa tieto obrázky načítajú do pamäte a znova sa zobrazia ako pri predchádzajúcej vizuálnej kontrole.

Načítané znormalizované obrázky sa preusporiadajú v rámci ich kategórie a ich počet sa zníži na počet prvkov kategórie s najnižším počtom prvkov tak aby všetky kategórie mali rovnaký počet prvkov.

Obrázky z trénovacej množiny sa znova premiešajú, pridá sa k nim názov ich kategórie a sú pripravené na uloženie. Pred tým sa ale testovacie obrázky premiešajú pridelia sa k nim názvy kategórií a rozdelia sa na testovacie a validačné dáta v pomere 50:50. Na záver sa tieto dáta uložia do ďalšieho .pickle súboru aby boli rýchlo dostupné pre ďalšie použitie. Z každej kategórie sa vyberie jeden obrázok aj s názvom a zobrazí sa.