

I-SUNS: Zadanie č.3

KONVOLUČNÉ NEURÓNOVÉ SIETE A PRENOS VEDOMOSTÍ

Vo vybranom programovacom jazyku implementujte program, ktorý bude vedieť klasifikovať druhy jedál. Na klasifikáciu aplikujte konvolučnú neurónovú sieť. Dataset je prístupný na adrese [na Google Drive](#). Čas odovzdania je určený časom vloženia do AIS. Deadline pre získanie 15 bodov je **2.12.2021 o 13:00/15:00**. Každý týžden omeškania je penalizovaný stratou dvoch bodov. **Idete o predposledné cvičenie. Do 9.12.2021 je potrebné mať odovzdané všetky zadania. Neodovzdanie akéhokoľvek zadania do 9. decembra sa považuje za nesplnenie podmienok zápočtu.**

Dáta

Dáta obsahujú obrazy rôznych druhov jedál na tanieri. K dispozícii máte 27.000 obrázkov v trénovacej a 3.000 obrázkov v testovacej množine. Triedy určujete podľa priečinka, v ktorom sa obrázok nachádza.

Úlohy

- Spoznajete, analyzujete a pripravte dáta na ďalšie spracovanie:
 - Analyzujete dataset. Vykreslite 1 obrázok z každej triedy. Pripravte si generátor dát, ktorý bude vedieť pripraviť batch obrázkov. Nezabudnite dáta normalizovať a vhodne zmeniť ich veľkosť. Vytvorte trénovaciu, validačnú a testovaciu množinu. Overte si správne načítanie obrázkov (spätným načítaním a zobrazením, výpismi z debuggera a pod). **2b**
- Natrénujte konvolučnú neurónovú sieť na riešenie tohoto problému.
 - Natrénujte konvolučnú neurónovú sieť. Zobrazte priebehy tréningu pre trénovacie aj validačné dáta. Vyhodnoťte na testovacej množine, získaný model opíšte pomocou konfúznej matice. Analyzujte výsledky. **2b**
 - Počas tréningu vytvárajte checkpointy, z ktorých bude možné tréning znova spustiť. **1b**
- Sledujte vplyv regularizátorov na neurónovú sieť:
 - Meňte parametre siete, kým nebudete pozorovať pretrénovanie (ak sa vám to nestane hneď na začiatku). Ukážte na grafoch priebehov a úspešnosti.

- Pokúste sa pretrénovanie vyriešiť aspoň dvoma metódami regularizácie. Pre každú metódu sledujte, aký vplyv majú parametre metódy na proces tréovania. Vyskúšajte aspoň 3 hodnoty a výsledky zobrazte do grafu/tabuľky (sledujete minimálne dĺžku tréovania, úspešnosť na tréovacích dáta a úspešnosť na testovacích dátach). **2x2b**
- Nájdite predtrénovanú sieť (napr. na datasete Imagenet) a prepoužite ju na riešenie problému:
 - Upravte sieť tak, aby miesto výsledku klasifikácia vracala príznaky zo siete. Použite takto pripravenú sieť na zakódovanie vášho celého datasetu. **1b**
 - Použite metódy na zmenšenie dimenzie a vizualizujte príznaky z datasetu v 2D alebo 3D priestore. Vzorky zafarbite podľa triedy, do ktorej patria. **2b**
 - Na pripravenom datasete natrénujte iný ľubovoľný nenáhodný klasifikátor na riešenie pôvodného klasifikačného problému. Snažte sa získať čo najlepšiu úspešnosť. Vyhodnoťte úspešnosť na tréovacej aj testovacej množine. Porovnanie s pôvodnou (najlepšou) neurónovou sieťou. **3b**

Nepovinné úlohy

- Napíšte program na čistenie datasetu - hľadajte rovnaké obrázky, chyby v označeniach a pod. **1-3b**
- Pripravený batch augmentujte pomocou minimálne troch rôznych augmentácií. Augmentovaný batch vizualizujte. Po augmentácií musia byť objekty stále jednoznačne identifikovateľné. **2b**
- Namiesto pôvodného klasifikačného problému natrénujte ľubovoľný nenáhodný klasifikátor na nájdenie dezertov a nedezertov. **1b**
- Na zakódovaný dataset aplikujte ľubovoľný zhlukovací algoritmus. Nájdene zhluky vizualizujte do 2D alebo 3D priestoru a analyzujte. **1-3b**