



## 2016./2017. mozgalo

Zadatak, na četvrtom po redu natjecanju Mozgalo, je **nenadzirana kategorizacija slika** (engl. unsupervised image classification) tj. **grupiranje** (engl. clustering) **slika prema njihovom sadržaju**. Dobit ćete neoznačeni skup podataka sa slikama koje treba svrstati u određeni broj kategorija. Zadatak je odrediti broj kategorija te svaku sliku svrstati u jednu od njih. Imajte na umu da Vaše rješenje ne smije imati nikakvu prethodnu informaciju o pripadnosti slika pojedinim kategorijama jer se tada ne bi radilo o problemu grupiranja nego (nadzirane) klasifikacije.

Dopušteno je raditi promjene na slikama kako biste ih prilagodili svom programu. Ako se odlučite prilagođavati slike to mora biti dio Vašeg programa. Jedino nije dopuštena promjena naziva slika.

Skup podataka sastoji se od gotovo 7000 slika u boji i njegova raspakirana veličina je 95MB.

Slike su različitih dimenzija, uz ograničenje da je najveća dimenzija (visina ili širina) svake pojedine slike 200 piksela.


Slike su pohranjene u različitim formatima (JPEG, PNG, GIF, TIFF).

Vaše rješenje, uz ugađanje određenih hiperparametara, mora raditi na proizvoljnom skupu

podataka, dok će se kvaliteta rezultata provjeravati na skupu podataka kojeg ćete dobiti. Rješenje problema treba dati konačni izlaz u obliku tekstualne datoteke koja sadrži dva stupca. U prvom stupcu nalazi se naziv slike, a u drugom broj kategorije kojoj slika pripada. Ovakav način ispisa nije jedini mogući, ali označava smjernicu kako bi mogao izgledati, u slučaju daljnje potrebe obrade podataka.

Prilikom pisanja dokumentacije potrebno je navesti svu literaturu koja se koristila. Detalje vezane uz dokumentaciju možete pronaći u dokumentu Bodovanje projektnog zadatka.

**SRETNNO!**



Neki postupci i pojmovi koji bi Vam mogli biti korisni prilikom rješavanja su (no nemojte se ograničiti samo na njih):

- redukcija dimenzionalnosti slike
- konvolucijske neuronske mreže, deep belief mreže, autoenkoderi, deep embedding
- K-means grupiranje, hijerarhijsko grupiranje
- postupci određivanja broja grupa (elbow method, AIC, BIC)
- image modeling
- joint segmentation
- spectral clustering
- metoda potpornih vektora (SVM, engl. Support Vector Machine)

Ako ne ostvarujete vrlo "dobre" rezultate na zadanom skupu podataka nemojte misliti da Vaše rješenje nije ispravno jer je dani skup podataka potencijalno "težak" za rješavanje ovog zadatka. Ako ipak želite pokazati da Vaše rješenje daje dobre rezultate na nekom drugom, "lakšem" skupu podataka slobodni ste to učiniti. Time ćete ujedno pokazati da Vaše rješenje dobro radi na različitim skupovima podataka. Ipak, imajte na umu da je kvaliteta rješenja na zadanom skupu podataka najvažnija.

**TIM MOZGALO**