

Prepoznavanje slikara pomoću konvolucijskih neuronskih mreža

Ivana Pleše, Ivana Senkić, Vedran Vinković

1 Uvodni opis problema

Problem kojim ćemo se baviti bit će prepoznavanje autora umjetničkih slika, uz izostanak bilo kakve druge informacije o slikama osim njih samih. Donedavno je to bio zadatak rezerviran isključivo za povjesničare umjetnosti kojima je jedan od glavnih izazova bio raspoznati radove plagijatora od originala, no pojavom strojnog učenja, popraćenog velikim tehnološkim napretcima, otvorena su vrata bržem prepoznavanju i katalogiziranju svjetskih remek-djela.

Primjenom različitih metoda strojnog učenja kao što su stroj potpornih vektora (*SVM*), algoritam k -najbližih susjeda (k -*NN*) i hijerarhijsko grupiranje (*HCA*), taj posao dobiva priliku biti automatiziran, čime se uvelike smanjuje vrijeme utrošeno na identificiranje slika. Kao posebna novost u pristupu ovom problemu, javlja se korištenje konvolucijskih neuronskih mreža (*CNN*), za koje se pokazalo da u nekim slučajevima postiže bolje rezultate od stručnjaka, zbog čega će naš rad po uzoru na [1] biti usmjeren upravo na taj pristup.

Skup podataka na kojem ćemo vršiti takvu obradu je *Kaggle*^[2], koji se temelji na *WikiArt*^[3], a sadrži približno 100,000 slika 2,300 autora iz različitih povijesnih razdoblja i likovnih stilova.

2 Cilj i hipoteze istraživanja problema

Glavni cilj istraživanja je na temelju treninga različitih konvolucijskih neuronskih mreža, prepoznati onu koja daje najbolje rezultate, te detaljnije analizirati njezine elemente. Od posebnog interesa je razumjeti značajke prepoznavanja autora najuspješnije od mreža, odnosno analizirati na koji način ona vrši klasifikaciju po autorima i stilovima.

Hipoteza istraživanja je da složena i jedinstvena kombinacija različitih likovnih sposobnosti i osobina ličnosti svakog od umjetnika rezultira njegovim osobnim stilom. Promatranjem osobnog stila moguće je unaprijediti metode spomenutog raspoznavanja.

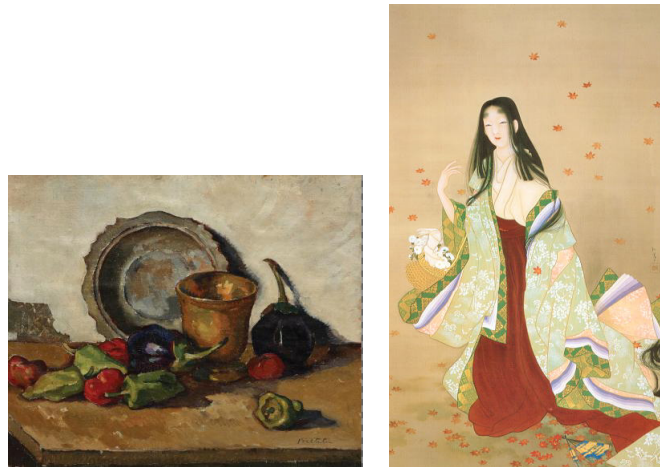
3 Pregled dosadašnjih istraživanja

Zajednički element svakog od dosadašnjih istraživanja ovog problema u području strojnog učenja bila je upotreba već unaprijed zadanog skupa značajki pa su korištenjem primjerice SIFT (engl.

Scale-invariant feature transform) i HOG (engl. *Histogram of oriented gradients*) metoda u postupku preuzete one njima svojstvene i uobičajene, što je dovelo do relativno loših rezultata. Također, većina opisanih istraživanja je, uz rijetke iznimke, kao nedostatak imala i korištenje manjeg skupa podataka, a uz to su se uglavnom bavila prepoznavanjem likovnih stilova i zajedničkih značajki različitih autora, dok je prepoznavanje samih autora slika ostalo zanemareno. Jedini za sada poznat i značajan rad koji se bavi isključivo prepoznavanjem autora slika je [1] koji se temelji na konvolucijskim neuronskim mrežama i s kojim ćemo se temeljitije upoznati kroz ovaj projekt.

4 Materijali, metodologija i plan istraživanja

Osim što se spomenuti skup podataka s platforme *Kaggle*^[2] sastoji od slika koje variraju u veličini i obliku, tako je i njihov broj po pojedinim umjetnicima različit pa je podatke potrebno modificirati prije početka samog treninga. U suprotnom, bez usklađivanja njihovih dimenzija ili odabira dovoljno velikog broja uzoraka, postoji rizik da odabrani skup neće biti dovoljno reprezentativan za trening, a na koncu i kvalitetno funkcioniranje istrenirane mreže.



Slika 1: Primjeri slika iz skupa podataka: Adam Baltatu (lijevo) i Uemura Shoen (desno)

Kako bismo izbjegli te probleme, uzimat ćemo samo one autore čiji je broj slika barem 300. Na taj način, naš skup sadržavat će slike od 57 autora (od njih sveukupno 2319) iz različitih likovnih razdoblja. Tih 300 slika od svakog autora ćemo podijeliti u 3 podskupa tako da prvom pripada 80%, a drugom i trećem 10% slika. Oni će predstavljati trening skup, validacijski skup te testni skup, što znači da će sadržavati 240 i po 30 slika respektivno. Također, slike ćemo centralizirati (engl. *zero-center*) i normalizirati, te uzeti kvadratni isječak dimenzija 224x224 te ćemo ih takve koristiti u trening i validacijskom skupu. Slikama ne mijenjamo veličinu kako bi održali njihovu detaljnost na razini piksela, koja nam je od velike važnosti, a koju bi takvom promjenom potencijalno izgubili.

Dodatno, za vrijeme treninga mreže, svaku sliku nasumično zrcalimo horizontalno s vjerojatnošću od 50% te uzimamo slučajan isječak, čime doprinosimo raznolikosti trening skupa i smanjujemo rizik od overfita. Taj postupak je utemeljen na hipotezi da je svaka slika ravnomjerno prožeta umjetničkim stilom promatranog autora, pa nasumičnim odabirom pojedinog dijela ne gubimo informacije potrebne za trening mreže. Potrebno je još napomenuti da je svaka slika označena svojim autorom u zasebnoj .csv datoteci.

Po završetku pripreme podataka, započinjemo s treningom konvolucijskih neuronskih mreža. Mreže koje ćemo trenirati bit će tri različita tipa: jednostavna konvolucijska neuronska mreža (engl. *Baseline CNN*), ResNet-18 trenirana otpočetak (engl. *ResNet-18 Trained from Scratch*) te ResNet-18 s prijenosom znanja (engl. *ResNet-18 with Transfer Learning*). Svaka od njih kao ulazne podatke uzima sliku 224x224x3 (u RGB modelu boja) te vraća rezultate za pojedinog umjetnika, koje u svakom prethodnom koraku treninga pokušava maksimizirati u odnosu na preostale.

Specijalnost ResNet-18 mreže istrenirane otpočetak jest da ona sama uči značajke isključivo u svrhu našeg zadatka, tj. prepoznavanja autora slika, dok je za ResNet-18 s prijenosom znanja svojstveno da je već od prije trenirana na velikim skupovima podataka (*ImageNet*) pa učenje započinje s *predznanjem* (tj. počinje s unaprijed određenim težinama). Arhitektura jednostavne konvolucijske mreže opisana je u Tablici 1.

Veličina ulaza	Sloj
3x224x224	3x3 CONV, <i>stride</i> 2, <i>padding</i> 1
32x112x112	2x2 <i>Maxpool</i>
32x56x56	3x3 CONV, <i>stride</i> 2, <i>padding</i> 1
32x28x28	2x2 <i>Maxpool</i>
1x6272	<i>Fully-connected</i>
1x228	<i>Fully-connected</i>

Tablica 1: Izgled jednostavne CNN

Analizom rezultata odredit ćemo najuspješniju mrežu te daljnje istraživanje vršiti na njoj i to primjenom različitih vrsta vizualizacije, a sve u svrhu boljeg razumijevanja njezine reprezentacije autora, odnosno autorova stila. Neke od takvih vizualizacija uključuju filtere mreža i korištenje *saliency map* (mape ispupčenosti).

Uz to, mrežu ćemo testirati korištenjem slika iz [4] koje su unaprijed digitalno modificirane, kako bi izgledale kao slike autora s kojima je naša mreža već upoznata.

Dodatno, najbolju mrežu ćemo ponovo trenirati, ali na drugačijem skupu podataka. Priprema podataka će ostati ista, ali broj slika po autoru ćemo smanjiti na 200 (tako ćemo dobiti 108 autora), a zatim pokušati isto i sa cijelim skupom dostupnim na platformi *Kaggle*^[2].

5 Očekivani rezultati predloženog projekta

Sudeći po rezultatima dobivenim u istraživanju opisanom u [1], očekujemo da će najbolja mreža i u našem slučaju biti ResNet-18 s prijenosom znanja. Predviđamo da će ona na manjem skupu podataka (300 slika po autoru) davati jako dobre rezultate.

Korištenjem cijelog skupa podataka, zbog neuravnoteženog broja slika po autoru, njeno prepoznavanje autora s vrlo malim brojem slika (< 50) neće biti jednako dobro, ali će i dalje s visokom preciznošću moći prepoznati one autore iz, prethodno spomenutog, manjeg skupa podataka.

6 Literatura

- [1] N. Viswanathan, *Artist Identification with Convolutional Neural Networks*, Stanford University, 2016.
- [2] Kaggle, <https://www.kaggle.com/c/painter-by-numbers>, (20.04.2018.)
- [3] Wikiart, <https://www.wikiart.org/>, (20.04.2018)
- [4] J. Johnson, *neural-style*, <https://github.com/jcjohnson/neural-style>, (20.04.2018.)