Universitatea din Bucureşti
Facultatea de Matematică şi Informatică
Departamentul de Informatică

Alexandra Diaconu
Bogdan Alexe
Radu Ionescu
Anii 3 şi 4, opţional

Concepte şi Aplicaţii în Vederea Artificială

Tema de laborator 11 - Clasificarea imaginilor folosind diverse caracteristici

Obiectiv:

Scopul acestei teme de laborator este multiplu: (i) extragerea de caracteristici diferite din imagini, caracteristici ce vor fi folosite apoi pentru clasificarea lor; (ii) lucrul cu rețelele convoluționale, modele computaționale state-of-the-art ce au revoluționat domeniul Vederii Artificiale; (iii) folosirea diverselor caracteristici extrase pentru realizarea unor experimente de clasificare binară pe două seturi de date diferite.

Date

În acest laborator vom lucra cu două seturi de date (Figura 1). Primul set de date conține imagini color din două clase, clasele *soccer-ball* (49 de imagini de antrenare și 15 imagini de testare) și *yin-yang* (45 de imagini de antrenare și 15 imagini de testare). Imaginile din cele două clase au fost extrase din setul de date inițial CALTECH 101 [1]. Cele două clase conțin obiecte similare în formă și culoare însă pot fi ușor discriminate folosind caracteristici atent alese. Al doilea set de date conține imagini cu fețe (1000 de imagini de antrenare și 100 de imagini de testare) și non-fețe (1000 de imagini de antrenare și 100 de imagini de testare) și este obținutt folosind datele din laboratoarele 9 și 10.

[1] https://data.caltech.edu/records/mzrjq-6wc02

Figura 1: **Exemple de imagini din cele două baze de date folosite**. **Stânga**: exemple din clasa yin-yang (rândul de sus) și clasa soccer-ball (rândul de jos). **Dreapta**: exemple din clasa fețe (rândul de sus) și clasa non-fețe (rândul de jos).

Caracteristici folosite

În acest laborator vom folosi mai multe tipuri de caracteristici:

- caracteristici la nivel de pixel = caracteristici ce derivă direct din valorile de intensitate sau de culoare ale fiecărui pixel din imagine, fără a lua în considerare structuri mai complexe sau relații între părți ale imaginii;
- caracteristici de nivel mediu = caracteristici ce cuprind informații care depășesc nivelul pixelilor și încep să captureze structuri locale din imagine, fără a ajunge încă la un nivel semantic înalt;
- caracteristici la nivel semantic = caracteristici obținute prin metode mai complexe (spre exemplu rețele neuronale convoluționale), care interpretează conținutul imaginii în termeni de obiecte, scene sau concepte. Astfel, ele integrează informații din întreaga imagine, captând relații spațiale complexe între părți diferite ale obiectelor.

Pentru fiecare imagine de antrenare extragem diverse caracteristici sub forma unor vectori, aceștia practic reprezintă descriptorul vizual al imaginii. Folosim apoi un SVM liniar pentru a învăța un clasificator pe datele de antrenare extrase.

Prezentăm în detaliu aceste caracteristici în ordinea complexității lor, corelat de altfel cu gradul de semnificație semantică.

Caracteristici la nivel de pixel

Caracteristicile extrase din imagini la nivel de pixel, precum histogramele de intensitate (în imaginile grayscale), histograme de culoare (în imaginile RGB) sau vectorii cu valorile de intensitate ale pixelilor (în engleză denumirea este *raw pixels*), reprezintă descrieri brute ale conținutului imaginii, fără a include informații semantice directe. Aceste caracteristici se concentrează pe proprietăți de bază ale imaginii, cum ar fi distribuția valorilor de luminanță sau cromatică, structura geometrică locală și alte aspecte cuantificabile numeric.

Histogramele de intensitate descriu distribuția nivelurilor de intensitate ale pixelilor într-o imagine fără a lua în considerare poziția sau contextul acestor valori. Sunt utile pentru identificarea tiparelor globale, cum ar fi contrastul sau luminozitatea generală.

Vectorii de intensitate sunt colecții directe ale valorilor de intensitate ale pixelilor dispuși într-un format unidimensional. Deși reprezintă imaginea completă, acești vectori formează o reprezentare sensibilă la transformări precum rotațiile, translațiile sau modificările de mărime ale obiectelor într-o imagine.

Aceste tipuri de caracteristici sunt esențiale în aplicațiile în care prelucrarea imaginii se bazează pe analize matematice sau statistice, însă, de obicei, sunt insuficiente pentru înțelegerea semnificației de nivel înalt (cum ar fi identificarea obiectelor sau interpretarea semantică). Ele servesc adesea ca bază pentru metode mai avansate de procesare, precum analiza principală a componentelor (PCA) sau pentru pregătirea datelor în modele computaționale mai complexe, cum ar fi rețelele neuronale convoluționale.

Caracteristici de nivel mediu

Acestea sunt caracteristici care combină informațiile brute de la nivelul pixelilor pentru a identifica tipare (pattern-uri) locale sau structuri geometrice relevante în imagine. Printre acestea se numără histogramele de gradienți orientați (HOG) sau caracteristici bazate pe textură precum LBP (Local Binary Patterns).

În acest laborator ne vom concentra pe histogramele de gradienți orientați, cu care am lucrat și în precedentele laboratoare. Histogramele de gradați orientați cuantifică informația despre direcția gradientului (schimbarea intensității) în regiuni locale ale imaginii. HOG este sensibil la contururi și muchii și este utilizat frecvent în probleme precum detectarea obiectelor (de exemplu, detectarea pietonilor). Spre deosebire de caracteristicile la nivel de pixel, HOG captează relații spațiale între pixeli dintr-o regiune.

Caracteristici de nivel semantic

Caracteristicile la nivel semantic reprezintă descrieri abstracte și bogate ale unei imagini, care codifică informații despre obiectele, scenele sau conceptele prezente în imagine, depășind interpretările pur geometrice sau locale. Acestea sunt esențiale pentru înțelegerea de nivel înalt a conținutului unei imagini și sunt utilizate în probleme precum recunoașterea de obiecte, clasificarea imaginilor sau generarea de descrieri textuale pentru imagini.

Aceste caracteristici se bazează pe relații complexe între părțile componente ale unei imagini. Nu se referă la părți individuale ale imaginii (cum ar fi pixeli sau contururi), ci la structuri compuse, cum ar fi obiectele sau contextul global.

Caracteristicile semantice pot fi reutilizate pentru diverse sarcini, cum ar fi clasificare, segmentare sau detectare de obiecte. Ele prezintă o robustețe mai mare la variații în iluminarea sau zgomotul imaginii, diverse posturi ale obiectului într-o imagine, comparativ cu caracteristicile la nivel de pixel sau de nivel mediu.

Aceste caracteristici sunt extrase în principal din rețele neuronale convoluționale (CNN), mai ales din straturile finale. Straturile inițiale ale CNN-urilor extrag caracteristici locale (precum muchii, texturi), în timp ce straturile intermediare combină aceste caracteristici pentru a construi descrieri mai complexe. Ultimele straturi (fully connected layers) codifică întreaga imagine într-un vector dens, care descrie conceptele din imagine. În acest laborator vom folosi vectorii obținuți din straturile complet conectate (fully-connected) ale rețelelor precum VGG-16, ResNet sau AlexNet pre-antrenate pe setul de date ImageNet [2].

Rezultate în clasificarea imaginilor de pe cele două seturi de date

Figura 2 prezintă rezultatele obținute folosind diversele caracteristici descrise pentru primul set de date (cu clasele *soccer-ball* și *yin-yang*). Caracteristicile la nivel de pixel bazate fie pe histograme de intensități grayscale sau histograme de culori obțin rezultate cu puțin

[2] `https://www.image-net.org/`

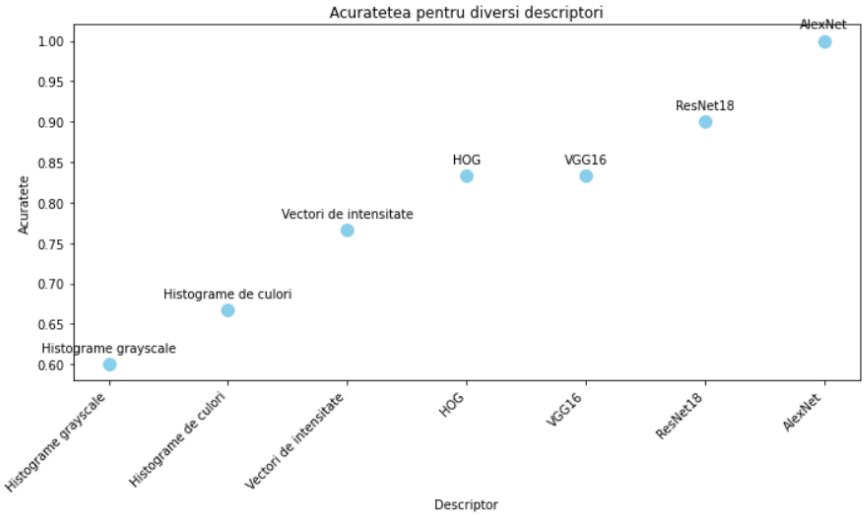

Figura 2: **Rezultate pentru primul set de date**.

peste scorul unui clasificator aleator, dovedind că nu sunt cele mai bune caracteristici de folosit în clasificarea imaginilor. Surprinzător, caracteristicile bazate pe vectori de intensități ale pixelilor obțin o performanță mult mai bună, apropiată de cea obținută folosind histograme de gradienți orientați. Performanța poate fi explicată prin alinierea destul de mare a obiectelor din cele două clase, fiecare imagine având obiectul dominant centrat în imagine, fără să mai existe alte obiecte în imagine. Caracteristicile de nivel semantic extrase din rețelele convoluționale AlexNet, VGG16 sau ResNet18 obțin performanțe superioare.

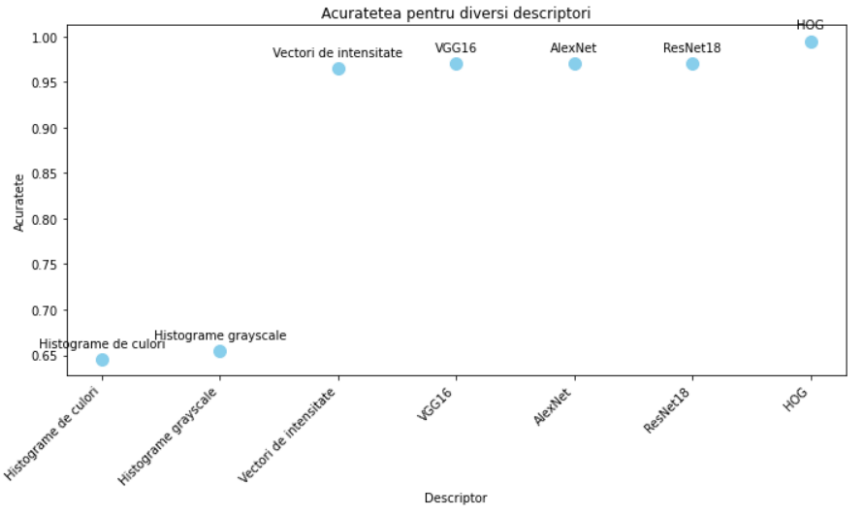

Figura 3: **Rezultate pentru al doilea set de date**.

Figura 3 prezintă rezultatele obținute folosind diversele caracteristici descrise pe al doilea set de date (*fețe* vs. *non-fețe*). Și de această dată caracteristicile bazate pe vectori de intensități ale pixelilor obțin o performanță mult mai bună decât cele bazate pe histograme grayscale sau de culori, profitând de alinierea unor structuri locale ale fețelor în interiorul unei ferestre care le încadrează perfect. Histogramele de gradienți orientați performează cel mai bine, mai bine chiar și decât caracteristicile de nivel semantic explicația fiind pe de o parte numărul mic de exemple de antrenare și testare și pe de altă parte faptul că rețelele convoluționale folosite conțin caracteristici generice, pre-antrenate pe ImageNet și nu specifice problemei de discriminare între fețe și non-fețe.