

Concepte și Aplicații în Vederea Artificială - Tema 2

Detectarea și recunoașterea facială a personajelor din serialul de desene animate Familia Simpson

Obiectiv

Scopul acestei teme este implementarea unui sistem automat de detectare și recunoaștere facială a personajelor din serialul de desene animate *Familia Simpson* (*The Simpsons*) folosind algoritmi de Vedere Artificială discutați la curs și implementați parțial la laborator.

Familia Simpson

Familia Simpson este un serial de desene animate lansat în anul 1989 și care rulează și în prezent cu noi sezoane. Serialul prezintă viața unei familii disfuncționale, având ca personaje principale pe Homer Simpson, un tată de familie lenș, dar inventiv, Marge Simpson, soția lui Homer și o mamă grijulie, și copiii Bart, Lisa și Maggie. De-a lungul sezoanelor apar mai mult de 150 de personaje diferite în episoadele serialului. Mai toate personajele (Figura 1) sunt desenate într-un stil inconfundabil cu față predominant galbenă, ochii rotunzi și având diverse elemente specifice (conturul feței, barba, păr).

Task 1 - detectarea facială

Prima problemă pe care o aveți de rezolvat constă în detectarea facială a tuturor fețelor personajelor care apar în imagini. Pentru fiecare imagine de intrare algoritmul vostru trebuie să returneze o mulțime de detecții asociate (ferastră dreptunghiulară și scor) ce localizează *toate* fețele dintr-o imagine. Figura 1 arată câteva exemple din mulțimea de antre-



Figura 1: Detectare facială a personajelor din *Familia Simpson*: fiecare față de interes este adnotată cu o fereastră dreptunghiulară de culoare roșie.



Figura 2: Recunoașterea facială a personajelor din Familia Simpson: fiecare față de interes este adnotată cu o fereastră dreptunghiulară de culoare specifică clasei personajului (albastru - Homer, galben - Marge, verde - Bart, violet - Lisa).

nare și adnotările corespunzătoare, ce constau în ferestre dreptunghiulare de culoare roșie ce încadrează perfect fiecare față.

Task 2 - recunoaștere facială

A doua problemă pe care o aveți de rezolvat constă în recunoașterea facială a numai anumitor personaje. Din cele peste 150 de personaje diferite ce apar de-a lungul episoadelor vom considera problema recunoașterii faciale numai pentru patru dintre ele și anume Homer, Marge, Bart și Lisa. Pentru fiecare imagine de intrare algoritmul vostru trebuie să returneze o mulțime de detecții asociate (numele personajului, fereastra dreptunghiulară și scor) ce localizează fețele celor patru personaje de interes (Homer, Marge, Bart, Lisa) din imagine. Figura 2 arată câteva exemple din mulțimea de antrenare și adnotările corespunzătoare, ce constau în ferestre dreptunghiulare ce încadrează perfect fețele de interes (pentru Homer, Marge, Bart și Lisa). Fiecare detecție are o culoare specifică clasei personajului (albastru - Homer, galben - Marge, verde - Bart, violet - Lisa).

Descrierea datelor

Arhiva cu materiale (disponibilă aici: <https://tinyurl.com/CAVA-2021-TEMA2>) conține patru directoare: *antrenare*, *validare*, *testare* și *evaluare*.

Directorul *antrenare* conține datele de antrenare. Pentru fiecare din cele patru personaje (Bart, Homer, Lisa, Marge) există un director corespunzător cu 1101 imagini de antrenare. Fiecare imagine conține adnotate toate fețele personajelor care apar în imagine. Fiecarei fețe adnotate îi corespunde o linie din fișierul text corespunzător (Figura 3) și are formatul:

```
nume_imagine xmin ymin xmax ymax nume_personaj
```

unde:

- `nume_imagine` reprezintă numele imaginii din directorul corespunzător personajului;
- `xmin ymin xmax ymax` reprezintă coordonatele ferestrei dreptunghiulare ce încadrează

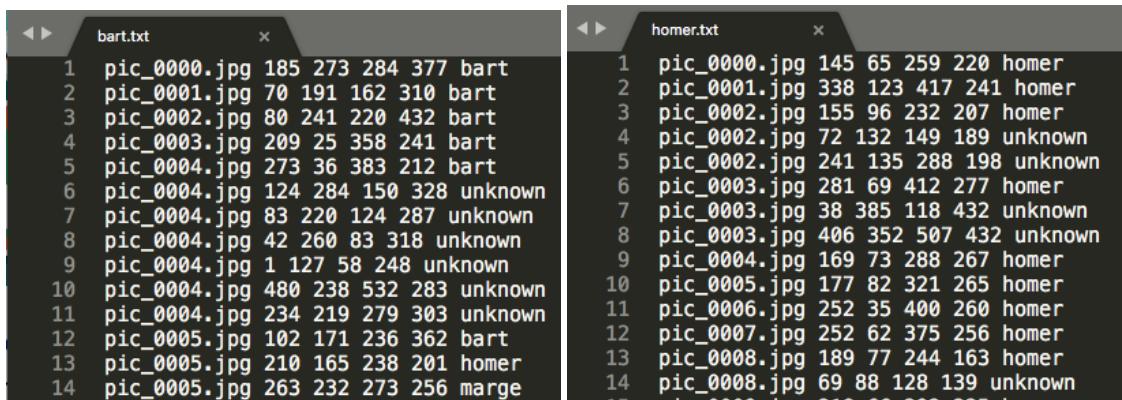


Figura 3: Fiecărei fețe adnotate din imaginile de antrenare îi corespunde o linie în fișierul text corespunzător cu adnotări ce urmează formatul nume_imagine xmin ymin xmax ymax nume_personaj.

față în imagine. Colțul din stânga sus are coordonatele (x_{\min}, y_{\min}) iar colțul din dreapta jos are coordonatele (x_{\max}, y_{\max}) .

- nume_personaj reprezintă numele personajului ce are față adnotată. Folosim pentru adnotarea fețelor cinci clase posibile: *bart*, *homer*, *lisa*, *marge* și *unknown* (pentru toate fețele personajelor diferite de cele patru de interes).

În total sunt 4404 imagini de antrenare ce conțin 5454 fețe adnotate.

Directorul *validare* conține datele de validare. Acestea constau din 200 de imagini adnotate în formatul descris mai sus. Puteți să vă folosiți de aceste date de validare pentru a estima performanțele diverselor voastre soluții.

Directorul *testare* are aceeași structură ca directorul *validare*. Vom face publice datele de test după **prima fază** (detalii mai târziu). Vom evalua performanța algoritmului vostru pe 200 de imagini de test, similare ca distribuție a fețelor cu imaginile de validare.

Directorul *evaluare* vă indică cum să vă scrieți codul astfel încât să respectați formatul fișierelor cu rezultate impus pentru faza de evaluare pe datele de test (**faza a doua**) ce va avea loc după trimiterea codului cu soluția fiecărui student. Conține următoarele sub-direcții:

- *fake_test* - acest director exemplifică cum vor arăta datele de testare, el păstrează aceeași structură ca cea descrisă pentru directorul *validare* descris anterior. Acest director va fi similar cu directorul *testare* în care vom pune imaginile de testare pentru faza a doua de evaluare.
 - *fisiere_solutie* - acest director exemplifică formatul fișierelor cu rezultatele pe care trebuie să le trimiteți în faza a doua. Veți trimite rezultatele voastre în acest format, încărcând o arhivă zip a unui director similar cu cel numit *Alexe_Bogdan_331*;

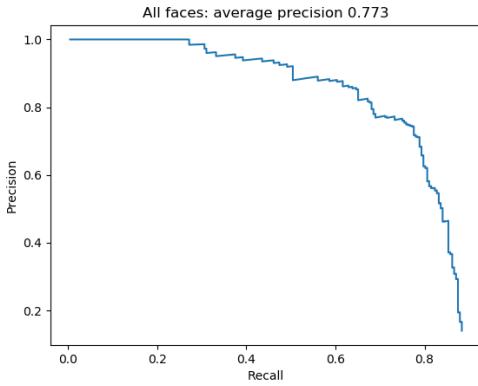


Figura 4: **Grafic precizie-recall pentru task-ul 1.** Performanța unei soluții realizate de noi pentru task-ul 1 pe datele de validare atinge un scor precizie-medie de 0.773.

- *cod_evaluare* - acest director conține codul care va fi folosit pentru evaluarea automată a rezultatelor voastre folosind adnotările soluțiilor corecte (ferestrele ground-truth). Asigurați-vă că acest cod rulează pe fișierele voastre. Puteți folosi în acest sens datele de validare. Adnotările soluțiilor corecte vor fi disponibile după faza a doua.

Protocolul de evaluare și măsura de performanță

Cuantificăm performanța algoritmilor voștri pentru task-ul 1 de detectare facială și task-ul 2 de recunoaștere facială în imagini test prin grafice de tip precizie-recall. Aceste grafice combină două valori:

- *precizia*: procentul de detecții returnate de algoritmul vostru ca fiind corecte (ele conțin o față de interes). În cazul ideal, algoritmul vostru are o precizie = 1 = 100%, adică fiecare detecție furnizată de algoritm reprezintă o față de interes. Pentru task-ul 1 toate fețele din imagine sunt fețe de interes, pentru task-ul 2 numai fețele unui personaj specific (Bart, Homer, Lisa sau Marge) sunt de interes.
- *recall* (= rată de detectare): procentul de fețe de interes din imaginile test localizate corect. În cazul ideal algoritmul vostru are un recall = 1 = 100%, adică localizează corect toate fețele de interes din imagine.

Fiecare punct de pe graficul precizie-recall reprezintă precizia și recall-ul algoritmului vostru obținute pentru toate detecțiile (ordonate descrescător după scor) care depășesc un anumit scor prag (threshold). De aceea, este foarte important ca atât pentru task-ul 1 cât și pentru task-ul 2 să aveți un scor asociat detecțiilor voastre. Sumarizăm întregul grafic prin *precizia medie* care reprezintă aria de sub grafic. Funcțiile *eval_detections* (pentru task-ul 1) și *eval_detections_character* (pentru task-ul 2) realizează aceste grafice și calculează precizia medie. Este important de reținut că fără un scor asociat pentru detecții nu putem calcula aria de sub graficul precizie-recall întrucât în acest caz graficul va fi reprezentat de un singur punct.

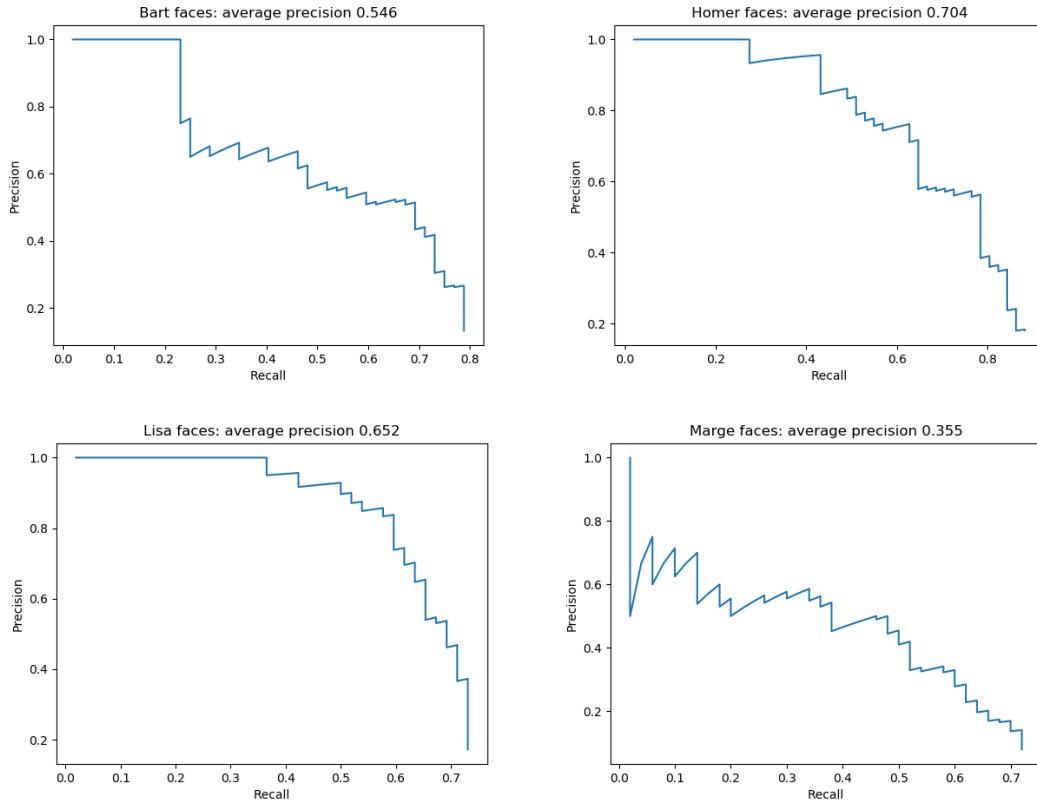


Figura 5: Grafic precizie-recall pentru task-ul 2. Performanța unei soluții realizate de noi pentru task-ul 2 pe datele de validare atinge un scor precizie-medie de 0.546 pentru Bart, 0.704 pentru Homer, 0.652 pentru Lisa și 0.355 pentru Marge.

Figura 4 ilustrează graficul precizie-recall și precizia medie = 0.773 pentru task-ul 1 pe setul de date de validare obținută de una din soluțiile noastre. Un detector facial perfect ar avea precizia medie = 1, cel mai din dreapta punct de pe graficul funcției aflându-se în acest caz în colțul din dreapta sus.

Figura 5 ilustrează graficele precizie-recall pentru fiecare personaj în parte pentru task-ul 2 pe setul de date de validare obținute de una din soluțiile noastre.

Restricții în implementarea soluției

Implementarea voastră trebuie să urmeze paradigma de glisare a unei ferestre (sliding-window), extragerea de caracteristici pentru fiecare fereastră și folosirea unui clasificator. Puteți folosi o paletă întreagă de caracteristici: histograme de gradienți orientați, caracteristici obținute folosind rețele convoluționale, caracteristici bazate pe culoare, etc.

Nu puteți folosi algoritmi de tipul Faster-RCNN sau YOLO sau rețele pre-antrenate pe fețe care v-ar ușura rezolvarea problemei. Dacă există neclarități ne puteți întreba dacă aveți sau nu voie să folosiți anumite lucruri.

Din punct de vedere al timpului de rulare, restricția este ca soluția voastră să poate fi rulată pe calculatorul nostru dotat cu un GPU GeForce RTX 3090 24 GB la un timp mediu de maxim 1 minut pe imagine.

Notare

Pentru această temă, vom folosi următoarele reguli de notare:

- **Task 1 - 4 puncte** - vom evalua performanța algoritmului vostru în problema detectării faciale pe o mulțime de 200 imagini de testare. Vom folosi ca măsură de performanță precizie-medie (Average Precision = AP) obținută pe baza graficului precizie-recall. Vom calcula punctajul corespunzător soluției voastre astfel:
 - o soluție care atinge un prag de $x\%$ AP va obține $\frac{4*x}{80}$ puncte. Practic acordăm 4 puncte pentru soluțiile care ating un prag de 80% AP.
 - soluția propusă de noi (Figura 4) ar obține $\frac{4*77.3}{80} = 3.87$ puncte.
 - o soluție care atinge un prag de 40% AP va obține 2 puncte.
 - o soluție care depășește pragul de 80% AP poate obține maxim 1 punct bonus în plus. Astfel, dacă ați avea o soluție perfectă cu 100% AP veți obține 5 puncte (1 punct bonus în plus).
- **Task 2 - 4 puncte** - vom evalua performanța algoritmului vostru în problema recunoașterii faciale pe aceeași mulțime de 200 imagini de testare. Vom folosi ca măsură de performanță media preciziilor-medii (mean Average Precision = mAP) obținută pe baza celor patru grafice precizie-recall. Vom calcula punctajul corespunzător soluției voastre astfel:
 - o soluție care atinge un prag de $x\%$ mAP va obține $\frac{4*x}{60}$ puncte. Practic acordăm 4 puncte pentru soluțiile care ating un prag de 60% mAP.
 - soluția propusă de noi (Figura 5) ar obține un $mAP = (0.546 + 0.704 + 0.652 + 0.355)/4 = 0.564$ care ar însemna $\frac{4*56.4}{60} = 3.76$ puncte.
 - o soluție care atinge un prag de 30% mAP va obține 2 puncte.
 - o soluție care depășește pragul de 60% mAP poate obține maxim 1 punct bonus în plus. Astfel, dacă ați avea o soluție cu mAP de cel puțin 75% mAP veți obține 5 puncte (1 punct bonus în plus).
- **documentație - 1 punct** - descrieți într-un fisier pdf de minim o pagină soluția voastră pentru rezolvarea celor două task-uri. Puteți ilustra aspecte cheie ale soluției voastre adăugând secvențe de cod și vizualizări ale imaginilor pentru soluția voastră. Acest fișier ar trebui să conțină suficientă informație astfel încât un student de nivel mediu de la cursul nostru să poată reimplementa soluția descrisă de voi.

- **oficiu - 1 punct** - primiți acest punct dacă formatul fișierelor voastre urmează formatul impus iar codul nostru de evaluare rulează pe datele primite de la voi fără a face modificări în fișierele voastre;
- **BONUS - 2 puncte** - puteți primi adițional 1 punct în plus pentru fiecare task, dacă depășiți pragurile de 80% AP și 60% mAP descrise mai înainte.

Termene limită

Prima fază - trimitera codului. Încărcați o arhiva zip cu codul soluției voastre și un fișier pdf ce descrie soluția voastră până vineri, 7 ianuarie, ora 23:59 la link-ul acesta <https://tinyurl.com/CAVA-2021-TEMA2-SOLUTII>. Codul vostru ar trebui să includă un fișier README (vedeți exemplul din materiale) cu următoarele informații: (i) librăriile folosite de voi și necesare pentru rularea soluției voastre; (ii) indicații despre cum ar trebui rulat codul pentru fiecare task. Studenții care nu încarcă un fișier pdf cu descrierea soluției lor vor avea nota scăzută cu 1 punct. Termenul limită este strict, nu vom accepta soluții primite mai târziu.

A doua fază - trimitera rezultatelor. Sâmbătă, 8 ianuarie, vom publica datele de test în directorul *testare* de la adresa <https://tinyurl.com/CAVA-2021-TEMA2>. Veți rula soluția voastră pe imaginile de test și veți încărca rezultatele în aceeași zi ca o arhiva zip folosind următorul link <https://tinyurl.com/CAVA-2021-TEMA2-REZULTATE> în formatul recomandat.