

Concepte și Aplicații în Vederea Artificială - Tema 2

Detectarea și recunoașterea facială a personajelor din serialul de desene animate Laboratorul lui Dexter

Obiectiv

Scopul acestei teme este implementarea unui sistem automat de detectare și recunoaștere facială a personajelor din serialul de desene animate *Laboratorul lui Dexter* folosind algoritmi de Vedere Artificială discutați la curs și implementați parțial la laborator.

Laboratorul lui Dexter

Laboratorul lui Dexter este un serial animat creat de Genndy Tartakovsky și produs de Cartoon Network, care a debutat în 1996. Serialul urmărește aventurile unui băiat-geniu pe nume **Dexter**, care are un laborator secret, ultratehnologizat, ascuns în spatele bibliotecii din camera sa. Laboratorul este un loc spectaculos, plin de mașinării futuriste, roboți și invenții științifice incredibile, unde Dexter își petrece timpul lucrând la experimente complexe. Deși Dexter este un geniu, viața lui nu este deloc ușoară din cauza surorii sale mai mari, **DeeDee**. Ea este veselă, naivă și extrem de curioasă, intrând mereu în laboratorul secret al lui Dexter, chiar dacă el încearcă să o țină departe prin tot felul de măsuri de securitate. Dee Dee reușește întotdeauna să provoace haos, stricând invențiile fratelui ei și enervându-l la culme cu întrebările și dansurile ei. Pe lângă cei doi frați, părinții lor completează dinamica familiei și aduc un plus de comedie. **Mama (Mom)** lui Dexter este o gospodină devotată, obsedată de curățenie și mereu văzută purtând mănușile galbene de menaj. Ea este calmă și iubitoare, dar nu bănuiește nimic despre laboratorul secret al fiului său. Dexter face tot posibilul să își ascundă activitățile științifice, de teamă că mama lui ar putea descoperi din greșeală totul în timp ce face curățenie. **Tatăl (Dad)** lui Dexter, pe de

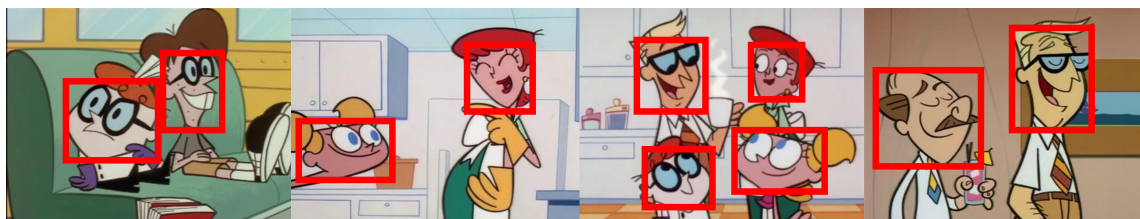


Figura 1: Detectare facială a personajelor din Laboratorul lui Dexter: fiecare față de interes este adnotată cu o fereastră dreptunghiulară de culoare roșie.

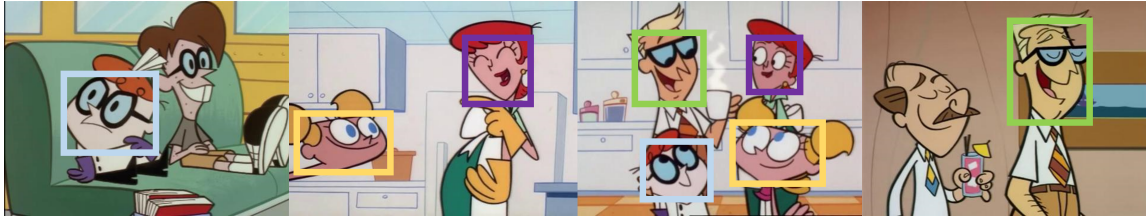


Figura 2: Recunoașterea facială a personajelor din Laboratorul lui Dexter: fiecare față de interes este adnotată cu o fereastră dreptunghiulară de culoare specifică clasei personajului (albastru - Dexter, galben - DeeDee, verde - Dad, violet - Mom).

altă parte, este relaxat și vesel. El este pasionat de activități simple precum grădinăritul, pescuitul sau privitul la televizor. Tatăl nu observă nimic neobișnuit în comportamentul lui Dexter și nu bănuiește că sub casa lor se află un laborator uriaș.

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 (fereastră dreptunghiulară și scor) ce localizează *toate* fețele dintr-o imagine. Figura 1 arată câteva exemple din mulțimea de antrenare 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. Alături de Dexter, personajul principal din serial, apar cu preponderență alte trei personaje: DeeDee - sora lui Dexter, Mom - mama lui Dexter, Dad - tatăl lui Dexter. Vom considera problema recunoașterii faciale numai pentru aceste patru personaje, Dexter, DeeDee, Mom și Dad. Pentru fiecare imagine de intrare algoritmul vostru trebuie să returneze o mulțime de detecții asociate (numele personajului, fereastră dreptunghiulară și scor) ce localizează fețele celor patru personaje de interes (Dexter, DeeDee, Mom, Dad) 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 Dexter, DeeDee, Mom și Dad). Fiecare detecție are o culoare specifică clasei personajului (albastru - Dexter, galben - DeeDee, verde - Dad, violet - Mom).

Descrierea datelor

Arhiva cu materiale (disponibilă aici: <https://tinyurl.com/CAVA-2024-TEMA2>) conține patru directoare: *antrenare*, *validare*, *testare* și *evaluare*.

Directorul *antrenare* conține datele de antrenare. Pentru fiecare din cele patru personaje (Dexter, DeeDee, Mom, Dad) există un director corespunzător cu 1000 imagini de antrenare. Fiecare imagine conține adnotate toate fețele personajelor care apar în imagine.

1	0001.jpg	191	196	233	224	dexter
2	0001.jpg	77	176	127	215	unknown
3	0001.jpg	293	153	323	209	unknown
4	0001.jpg	362	190	426	226	unknown
5	0002.jpg	192	159	302	251	dexter
6	0003.jpg	41	125	210	262	unknown
7	0003.jpg	341	190	436	259	dexter
8	0004.jpg	151	89	367	223	dexter
9	0005.jpg	185	179	324	248	dexter
10	0006.jpg	10	93	131	190	dexter
11	0007.jpg	231	149	287	183	dexter
12	0008.jpg	182	101	292	200	dexter
13	0009.jpg	187	129	420	302	dexter
14	0009.jpg	3	65	155	277	unknown
15	0010.jpg	82	55	401	278	dexter
16	0011.jpg	175	123	423	318	dexter
17	0012.jpg	70	239	194	336	dexter
18	0012.jpg	228	239	372	337	unknown
19	0013.jpg	99	124	254	252	dexter
20	0013.jpg	250	64	355	199	unknown
21	0014.jpg	264	62	432	214	dexter
22	0015.jpg	113	89	174	158	unknown

1	0001.jpg	275	89	408	177	deedee
2	0002.jpg	92	61	241	160	deedee
3	0002.jpg	335	29	444	105	dexter
4	0003.jpg	13	72	234	202	deedee
5	0003.jpg	307	190	468	321	dexter
6	0004.jpg	134	106	366	249	deedee
7	0005.jpg	229	81	440	233	deedee
8	0006.jpg	136	173	234	252	dexter
9	0006.jpg	325	166	430	246	deedee
10	0007.jpg	59	111	401	337	deedee
11	0008.jpg	63	34	353	256	deedee
12	0009.jpg	125	109	225	171	deedee
13	0010.jpg	46	45	322	248	deedee
14	0011.jpg	180	229	295	302	dexter
15	0011.jpg	276	65	423	168	deedee
16	0012.jpg	1	1	477	342	deedee
17	0013.jpg	64	89	216	216	deedee
18	0013.jpg	300	215	455	326	dexter
19	0014.jpg	72	140	234	245	deedee
20	0014.jpg	279	43	430	201	unknown
21	0015.jpg	82	100	202	200	deedee
22	0015.jpg	275	259	393	313	unknown

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`.

Fiecărei 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 fața în imagine. Colțul din stânga sus are coordonatele (`xmin`, `ymin`) iar colțul din dreapta jos are coordonatele (`xmax`, `ymax`).
- `nume_personaj` reprezintă numele personajului ce are fața adnotată. Folosim pentru adnotarea fețelor cinci clase posibile: *dexter*, *deedee*, *mom*, *dad* și *unknown* (pentru toate fețele personajelor diferite de cele patru de interes).

În total sunt 4000 imagini de antrenare ce conțin 5813 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-directoare:

- *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.
- *fișiere_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 *331_Alexe_Bogdan*;
- *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.

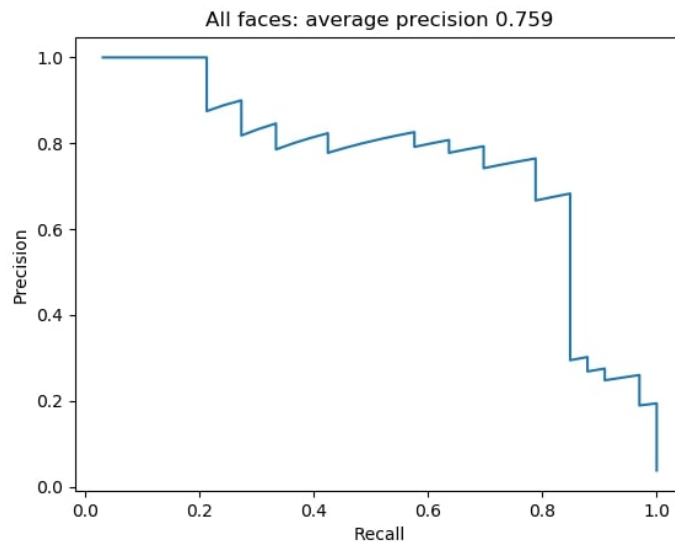


Figura 4: **Grafic precizie-recall pentru task-ul 1.** Performanța unei soluții realizate de noi pentru task-ul 1 pe datele de validare (primele 20 de imagini) atinge un scor precizie-medie de 0.759.

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 (Dexter, DeeDee, Mom sau Dad) 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 4 ilustrează graficul precizie-recall și precizia medie = 0.759 pentru task-ul 1 pe setul de date de validare (primele 20 de imagini) 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 (primele 20 de imagini) obținute de una din soluțiile noastre.

Restricții în implementarea soluției

Implementarea voastră pentru soluția de bază 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 gradienti orientați, caracteristici obținute folosind rețele convoluționale, caracteristici bazate pe culoare, etc.

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

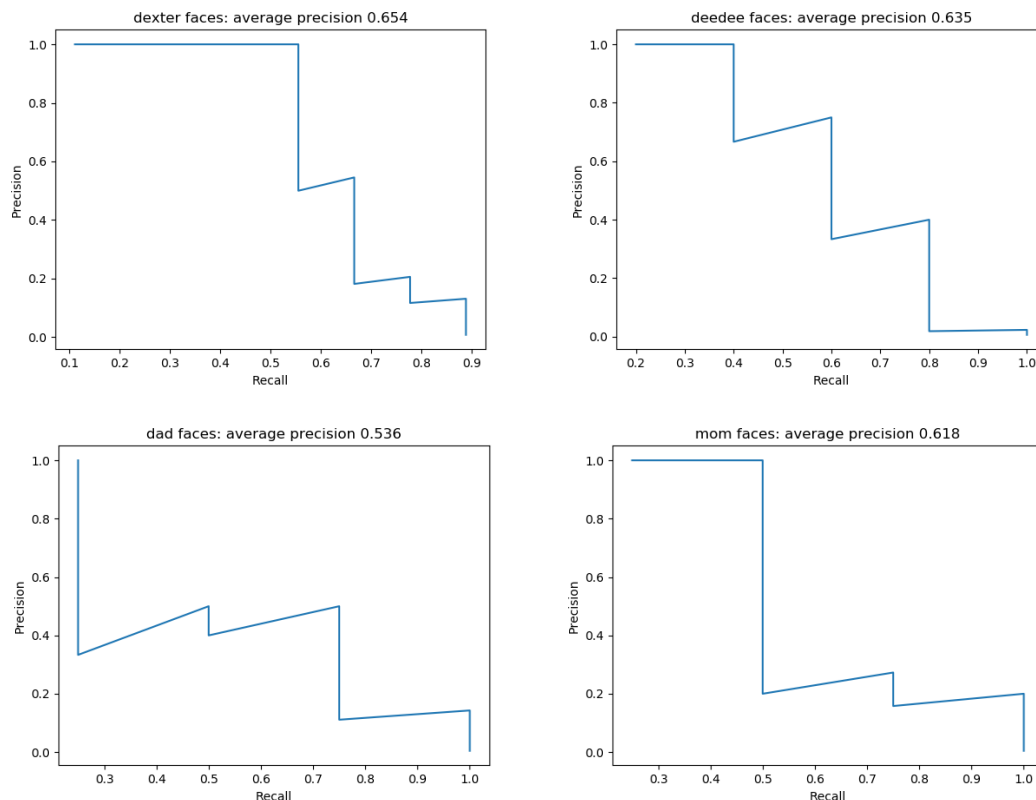


Figura 5: **Grafic precizie-recall pentru task-ul 2.** Performanța unei soluții realizate de noi pentru task-ul 2 pe datele de validare atinge un scor precizie-medie de 0.654 pentru Dexter, 0.635 pentru Deedee, 0.536 pentru Dad și 0.618 pentru Mom.

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 $\frac{4 \cdot x}{80}$ puncte, dar nu mai mult de 4 puncte. Practic acordăm 4 puncte pentru soluțiile care ating pragul de 80% AP.
 - soluția propusă de noi (Figura 4) ar obține $\frac{4 \cdot 75.9}{80} = 3.79$ puncte (am presupus că obținem o performanță similară pe întregul set de 200 de imagini).
 - o soluție care atinge un prag de 40.0% AP va obține 2 puncte.
 - o soluție care depășește pragul de 80% AP va obține 4 puncte, nu se iau puncte bonus aici.

- **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 $\frac{4 \cdot x}{60}$ puncte, dar nu mai mult de 4 puncte. Practic acordăm 4 puncte pentru soluțiile care ating pragul de 60% mAP.
 - soluția propusă de noi ar obține un $mAP = (0.654 + 0.635 + 0.536 + 0.618) / 4 = 0.610$ (Figura 5) care ar însemna 4 puncte (am presupus că obținem o performanță similară pe întregul set de 200 de imagini).
 - o soluție care atinge un prag de 30% mAP va obține 2 puncte.
 - o soluție care depășește pragul de 60% mAP va obține 4 puncte, nu se iau puncte bonus aici.
- **BONUS - 25% din punctajul pentru fiecare task** - vă puteți crește punctajul cu 25% pentru fiecare task, dacă folosiți o soluție secundară (un detector *state-of-the-art* de obiecte/fețe) de tip Faster-RCNN sau YOLO, pe care să o adaptați la problema voastră și a cărei performanță să depășească pragul impus pentru fiecare task (80%AP pentru task-ul 1 și 60% mAP pentru task-ul 2). Astfel, dacă obțineți cu soluția de bază 2 puncte la un task, puteți crește punctajul la 2,5 puncte condiționat de faptul că adaptați un detector pentru task-ul vostru și reușiți să depășiți pragul menționat.
- **documentație - 1 punct** - descrieți într-un fișier pdf de minim două pagini soluția voastră de bază 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. Dacă folosiți și o soluție secundară pentru bonus includeți detalii despre aceasta în documentație.
- **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;

Termene limită

Prima fază - trimiterea codului. Încărcați o arhivă zip cu codul soluției voastre și un fișier pdf ce descrie soluția voastră până duminică, 19 ianuarie 2025, ora 23:59 la link-ul acesta <https://tinyurl.com/CAVA-2024-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. Implementarea voastră trebuie să fie gata de rulat pe setul de testare, astfel includeți toate fișierele ce conțin modele, descriptori etc. în arhiva voastră pentru a putea fi încărcate direct. De asemenea, la fel ca și la Tema 1, nu acceptăm modificarea codului după trimiterea lui. Asigurați-vă ca arhiva trimisă conține un cod sursă care poate fi rulat într-un environment nou, că aveți toate fișierele incluse și că parametri aleși sunt robuști și generalizează pe datele de validare. Sub nicio formă nu aveți voie să vă modificați parametri după trimiterea soluției.

A doua fază - trimiterea rezultatelor. Luni, 20 ianuarie, vom publica datele de test în directorul *testare* de la adresa <https://tinyurl.com/CAVA-2024-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-2024-TEMA2-REZULTATE> în formatul recomandat.