

UNIVERSITATEA "ALEXANDRU-IOAN CUZA" DIN IAȘI

FACULTATEA DE INFORMATICĂ



LUCRARE DE LICENȚĂ

**Identificarea similarității vizuale a imaginilor
folosind rețele neurale**

propusă de

Petru-Marian Gafițescu

Sesiunea: iulie, 2019

Coordonator științific

Conf. Dr. Sabin-Corneliu Buraga

UNIVERSITATEA "ALEXANDRU-IOAN CUZA" DIN IAȘI

FACULTATEA DE INFORMATICĂ

**Identificarea similarității vizuale a
imaginilor folosind rețele neurale**

Petru-Marian Gafițescu

Sesiunea: iulie, 2019

Coordonator științific

Conf. Dr. Sabin-Corneliu Buraga

Avizat,
Îndrumător lucrare de licență,
Conf. Dr. Sabin-Corneliu Buraga.

Data: Semnătura:

Declarație privind originalitatea conținutului lucrării de licență

Subsemnatul **Gafițescu Petru-Marian** domiciliat în **România, jud. Suceava, Sat. Arbore(Com. Arbore), nr. 1012**, născut la data de **30 iunie 1998**, identificat prin **CNP 1980630330527**, absolvent al Facultății de informatică, **Facultatea de informatică** specializarea **informatică**, promoția 2019, declar pe propria răspundere cunoscând consecințele falsului în declarații în sensul art. 326 din Noul Cod Penal și dispozițiile Legii Educației Naționale nr. 1/2011 art. 143 al. 4 și 5 referitoare la plagiat, că lucrarea de licență cu titlul **Identificarea similarității vizuale a imaginilor folosind rețele neuronale** elaborată sub îndrumarea domnului **Conf. Dr. Sabin-Corneliu Buraga**, pe care urmează să o susțin în fața comisiei este originală, îmi aparține și îmi asum conținutul său în întregime.

De asemenea, declar că sunt de acord ca lucrarea mea de licență să fie verificată prin orice modalitate legală pentru confirmarea originalității, consimțind inclusiv la introducerea conținutului ei într-o bază de date în acest scop.

Am luat la cunoștință despre faptul că este interzisă comercializarea de lucrări științifice în vederea facilitării falsificării de către cumpărător a calității de autor al unei lucrări de licență, de diplomă sau de disertație și în acest sens, declar pe proprie răspundere că lucrarea de față nu a fost copiată ci reprezintă rodul cercetării pe care am întreprins-o.

Data:

Semnătura:

Declarație de consimțământ

Prin prezenta declar că sunt de acord ca lucrarea de licență cu titlul **Identificarea similarității vizuale a imaginilor folosind rețele neurale**, codul sursă al programelor și celelalte conținuturi (grafice, multimedia, date de test, etc.) care însoțesc această lucrare să fie utilizate în cadrul Facultății de informatică.

De asemenea, sunt de acord ca Facultatea de informatică de la Universitatea "Alexandru-Ioan Cuza" din Iași, să utilizeze, modifice, reproducă și să distribuie în scopuri necomerciale programele-calculator, format executabil și sursă, realizate de mine în cadrul prezentei lucrări de licență.

Absolvent **Petru-Marian Gafițescu**

Data:

Semnătura:

Cuprins

| | |
|---|-----------|
| Rezumat | 2 |
| Introducere | 3 |
| 1 Fundamente | 5 |
| 1.1 Rețea neurală | 5 |
| 1.2 Rețele neurale convoluționale | 9 |
| 1.2.1 Convoluția | 10 |
| 1.2.2 <i>Pooling</i> | 10 |
| 1.2.3 <i>Clasificare</i> | 11 |
| 1.3 Keras și Tensorflow | 11 |
| 1.3.1 Tensorflow | 11 |
| 1.3.2 Keras | 12 |
| 1.4 OAuth | 12 |
| 2 Scopuri și cerințe | 14 |
| 2.1 Necesitatea aplicației | 14 |
| 2.2 Modul în care aplicația rezolvă problema vizată | 14 |
| 2.3 Comparația cu soluții existente | 15 |
| 2.3.1 SimilarityPivot | 15 |
| 2.3.2 Reverse image search | 15 |
| 2.3.3 Servicii pentru analiza imaginilor | 16 |
| 3 Analiză și proiectare | 17 |
| 3.1 Scenarii de utilizare | 17 |
| 3.1.1 Autentificare | 17 |
| 3.1.2 Alege sursa imaginilor | 17 |
| 3.1.3 Încarcă imagine | 17 |

| | | |
|----------|--|-----------|
| 3.2 | Arhitectura aplicației | 18 |
| 3.3 | Comunicarea între componentele aplicației | 22 |
| 3.4 | Modulul <i>back-end</i> | 23 |
| 3.4.1 | API-ul expus | 23 |
| 3.4.2 | Rețeaua neurală folosită | 23 |
| 3.4.3 | Baza de date | 26 |
| 4 | Detalii de implementare | 27 |
| 4.1 | Implementarea aplicației | 27 |
| 4.2 | Detaliile modulului front-end | 28 |
| 4.2.1 | HomeComponent | 28 |
| 4.2.2 | LoginComponent | 29 |
| 4.2.3 | ImageUploadComponent | 29 |
| 4.3 | Detaliile modulului back-end | 30 |
| 4.3.1 | Autentificarea cu OAuth | 30 |
| 4.3.2 | Informații despre profil | 30 |
| 4.3.3 | Analiza unei imagini | 31 |
| 4.3.4 | Preluarea imaginilor de pe Flickr | 33 |
| 4.3.5 | Optimizarea procesului de returnare a imaginilor | 34 |
| 5 | Manual de utilizare | 36 |
| | Concluzii | 37 |
| | Bibliografie | 39 |
| | Anexe | 41 |

Rezumat

Această lucrare descrie scopul, implementarea și rezultatele aplicației web "FlickrCloneFetch", o aplicație web ce permite utilizatorului să se conecteze cu ajutorului contului de *Flickr*, să încarce în aplicație o imagine, iar apoi să vizualizeze imaginile similare din punct de vedere vizual. În funcție de preferințele utilizatorului, imaginile vor fi preluate de pe profilul său, imaginile publice ale contactelor sale sau imaginile publice relevante de pe platforma *Flickr*.

Modulul de *back-end* al aplicației este scris în Python, folosind Flask pentru a expune un API. Analiza imaginilor este făcută cu ajutorul unui model de rețea neurală convoluțională, folosind *framework*-urile Tensorflow și Keras. *Front-end*-ul aplicației este realizat în Typescript, utilizând platforma Angular. Sursa imaginilor este platforma *Flickr*, imaginile fiind accesate folosind *Flickr API*.

Introducere

În zilele noastre, datorită faptului că majoritatea telefoanelor mobile au și funcția de cameră foto, imaginile au devenit principala formă de rememorare a evenimentelor trăite. Platforme întregi sunt dedicate fotografiilor, printre cele mai populare numărându-se Instagram, Flickr sau Pinterest. Multe din aceste platforme oferă posibilitatea de sortare, respectiv căutare a imaginilor în funcție de locație, mărime, data fotografiei sau descrierea adăugată de utilizator. Uneori aceste opțiuni nu sunt de ajuns, imaginile având mai multă relevanță din prisma conținutului decât după orice alte metadate. Totuși, opțiunea de a căuta imagini după similaritatea vizuală față de altă imagine, cu alte cuvinte, chiar după componența imaginii este o caracteristică rar întâlnită.

Aplicația "FlickrCloneFetch" își propune implementarea acestei caracteristici pentru platforma *Flickr*, oferind utilizatorului posibilitatea de a vedea imagini similare vizual cu o imagine încărcată atât de pe propriul profil, cât și de pe profilele contactelor sau chiar din imaginile publice de pe platformă.

Scopul "FlickrCloneFetch" este de a fi o unealtă utilă atât fotografilor care folosesc platforma (pentru a vedea lucrările din propriul profil care au aceeași tematică), cât și utilizatorului uzual (pentru a găsi fotografii legate de un anumit eveniment din viața sa, sau fotografii cu unele tematici alese).

Pentru realizarea funcționalității descrise au fost folosite diverse tehnologii web, precum Flak, Angular, împreună cu API-ul platformei *Flickr* și un model de rețea neurală specializată în clasificarea imaginilor, model disponibil prin *framework*-ul TensorFlow.

Capitolul *Fundamente* va oferi o descriere a principalelor concepte, tehnologii și biblioteci folosite în implementarea aplicației.

Scopul urmărit, o privire de ansamblu asupra aplicației și comparația cu unele soluții deja existente se vor regăsi în capitolul *Scopuri și cerințe*.

Următorul capitol, *Analiză și proiectare* conține descrierea arhitecturii aplicației, prezentarea componentelor necesare și a modului în care interacționează, precum și scenariile de utilizare ale aplicației.

În capitolul *Detalii de implementare* se prezintă aspecte privitoare la modul de implementare, motivarea alegerii modelului de rețea neurală, algoritmi și structurile de date folosite pentru a optimiza timpul de răspuns al aplicației precum și API-urile externe folosite.

Capitolul *Manual de utilizare* prezintă modul de utilizare a aplicației , fiind urmat de *Concluzii*, o serie de concluzii legate de modul în care aplicația răspunde cerințelor și posibile direcții de dezvoltare ulterioară.

Capitolul 1

Fundamente

1.1 Rețea neurală

O rețea neurală este un ansamblu interconectat de elemente de procesare simple, *unități* sau *noduri*, a căror funcționalitate este ușor bazată pe neuronul biologic. Abilitatea de procesare a rețelei este stocată în conexiunile inter-unități, sau *ponderi* sinaptice, obținute dintr-un proces de adaptare la învățare dintr-un set de date de antrenare.[8]

În biologie, fiecare neuron primește informație prin dendrite și transmite informație prin axon, doar atunci când a primit suficientă informație de la neuronii conectați la dendritele sale. Astfel, un neuron va transmite informația mai departe atunci când cantitatea de informație primită depășește un anumit *prag*.

Neuronii artificiali sunt construiți pe același principiu: dendritele vor fi echivalate de conexiuni către unitate, iar axonul de către o conexiune pornind de la acea unitate. Fiecare sinapsă(conexiune inter-neuronală) va fi caracterizată de o anumită *pondere* astfel încât datele de intrare primite prin acea conexiune vor fi multiplicare cu acea pondere înainte de a fi transmise la neuronul artificial. Aici, semnalele ponderate sunt însumate pentru a forma o *activare*.

În cadrul nodului, dacă *activarea* va depăși un anumit *prag*, unitatea va transmite mai departe o valoare(de obicei 1), altfel nu transmite nimic. Compararea cu pragul de activare se face folosind o funcție de activare. Matematic, pentru datele de intrare reprezentate de $x_1...x_n$ și ponderile reprezentate de $w_1...w_n$, neuronul artificial funcționează astfel:

$$output = \begin{cases} 0, & \text{dacă } \sum_i x_i * w_i \leq prag \\ 1, & \text{dacă } \sum_i x_i * w_i > prag \end{cases}$$

Procedeul descris (însumare, compararea cu pragul, apoi emiterea de 1 când pragul este mai mic și 0 altfel) reprezintă cel mai simplu model de neuron artificial, perceptronul (fig. 1.1).

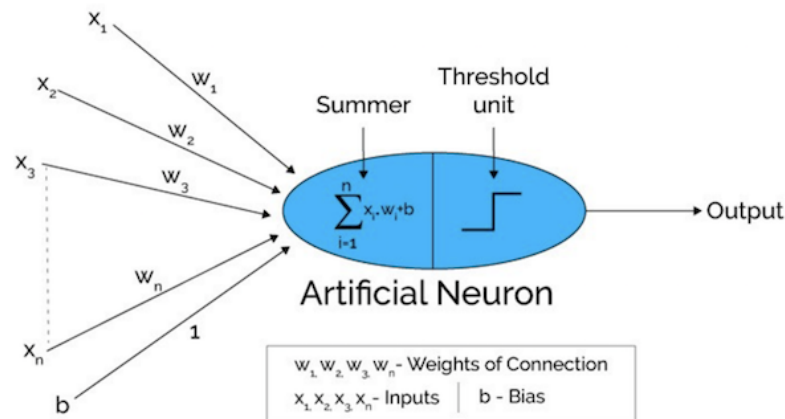


Figura 1.1: Un model de perceptron[6]

Modul de învățare a unui perceptron este foarte simplu: Dacă greșește în a clasifica un element (obține *output* 1 când valoarea elementului e 0, sau invers) creștem ponderile dacă activarea a fost prea mică comparativ cu pragul de activare, respectiv descreștem ponderile dacă pragul de activare a fost mai mic decât activarea. Astfel perceptronul își adaptează ponderile dinamic, în funcție de datele de antrenament, trecând printr-un proces de învățare.

De-a lungul istoriei, au fost folosite mai multe tipuri de neuroni, fiecare având avantajele și dezavantajele sale. Cum corpul neuronului este în principiu un sumator ponderat, ce se va schimba între tipurile de neuroni este funcția de activare.

Deoarece de unul singur un neuron nu are putere de decizie prea mare, o rețea neurală va folosi mai mulți, ieșirile unor neuroni constituind intrări pentru alți neuroni. Neuroni vor comunica între ei prin activarea/lipsa activării, astfel imitând transmiterea de semnale electrice din cadrul creierului uman.

O rețea neurală cu mai multe straturi de neuroni mai este numită și *MultiLayer-Perceptron*, deși de obicei sunt folosiți neuroni de tip sigmoid, nu perceptron. O rețea de tip este organizată tipic pe mai multe straturi, având următoarea structură:

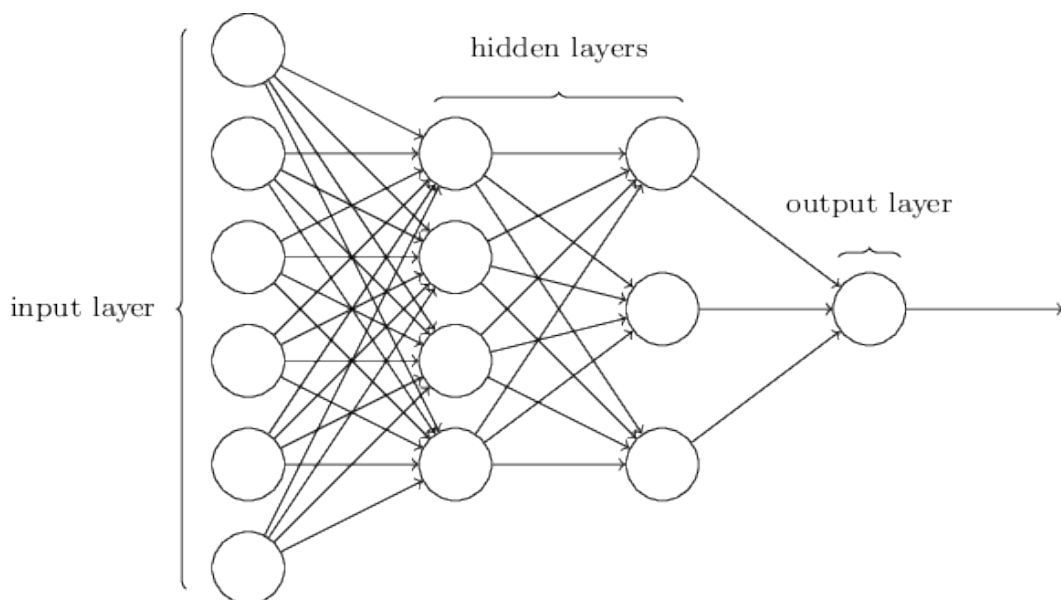


Figura 1.2: Structura unei rețele[10]

Straturile rețelei:

1. Stratul de intrare (*input*)

Stratul de *input* este format din neuroni de intrare, care preiau datele de intrare și le transmit mai departe rețelei. De cele mai multe ori, neuronii de intrare nu au ponderi și nici funcții de activare, având un rol pasiv în cadrul rețelei.

2. Stratul/Straturile de procesare (*hidden*)

Straturile *hidden* se ocupă de procesarea datelor de intrare. Există diferite moduri de a organiza straturile ascunse ale rețelei, în funcție de tipul rețelei: în cadrul unei rețele *feedforward* doar vor transmite datele prelucrate, iar în cadrul unei rețele convoluționale ce se ocupă de procesarea imaginilor vor exista straturi convoluționale și de *pooling* pentru a descoperi anumite trăsături în imagine, iar la rețelele de tip recurent, în straturile ascunse vor apărea conceptele de memorie și stări.

În toate cazurile, ideea de bază este ca straturile *hidden* vor prelua datele de intrare neprelucrate de la stratul de input, apoi vor prelucra acele date conform ponderilor asiguate și calculate în diverse moduri, pasând apoi datele procesate ultimului strat, cel de ieșire. Numărul de straturi *hidden*, organizarea lor, precum și tipurile de straturi și neuroni folosite depind atât de structura rețelei cât și de scopul în care este folosită aceasta.

3. Stratul de ieșire (*output*)

Neuronii din stratul de ieșire pot fi proiectați diferit față de straturile anterioare, reprezentând actorii finali ai rețelei și având rolul de a produce outputul programului. Astfel, stratul de ieșire se coalizează și produce în mod concret rezultatul final, cu rol în eficientizarea și îmbunătățirea acestuia. Referitor la cele menționate anterior, pentru a înțelege mai bine rețeaua neurală, cele trei nivele de straturi(stratul de intrare, straturile ascunse și stratul de ieșire)vor fi privite împreună ca întreg.

Pe lângă straturile descrise, o rețea neuronală mai are o funcție de cost folosită pentru a aproxima diferența dintre valoarea prezisă de rețea și valoarea reală. Mecanismul de antrenare utilizat la perceptron nu funcționează și aici, deoarece eroarea nu se propagă decât pe ultimul strat ascuns. Antrenarea unei rețele neurale are ca scop principal reducerea valorii funcției de cost, existând mai multe tehnici pentru a îmbunătăți acest procedeu. Un rol foarte important în antrenarea unei rețele îl are algoritmul de *Backpropagation*.

Backpropagation Când antrenăm un simplu perceptron, îi putem modifica instant ponderile în funcție de eroarea obținută (valoarea funcției de cost). În cazul unei rețele eroarea trebuie corectată la toate nivelele, nu doar la ultimul. Aici intervine algoritmul de *Backpropagation*. Algoritmul funcționează astfel:[5]

Pentru fiecare strat l , începând cu ultimul:

Pentru fiecare neuron i de pe stratul l :

1. Calculăm eroarea neuronului i de pe stratul l
2. Folosind eroarea calculată calculăm derivatele parțiale ale funcției de cost în funcție de pragul neuronului i și ponderile sale (derivatele parțiale ne arată cum variază funcția de cost în funcție de parametrii aleși)
3. Modificăm ponderile și/sau pragul pentru a corecta eroarea pentru acel neuron
4. Propagăm eroarea către neuronii de pe stratul $l-1$ și repetăm pașii de mai sus

Atât în timpul antrenării, cât și în timpul folosirii unei rețele (trecerea datelor de intrare prin rețea pentru a obține date de ieșire), se folosește foarte mult înmulțirea de matrici (pentru a înmulți ponderile neuronilor cu datele de intrare). Pentru anumite cazuri de input, putem optimiza numărul de parametri și procesul de funcționare. În continuare vom prezenta un tip de rețea optimizată în mod special pentru imagini.

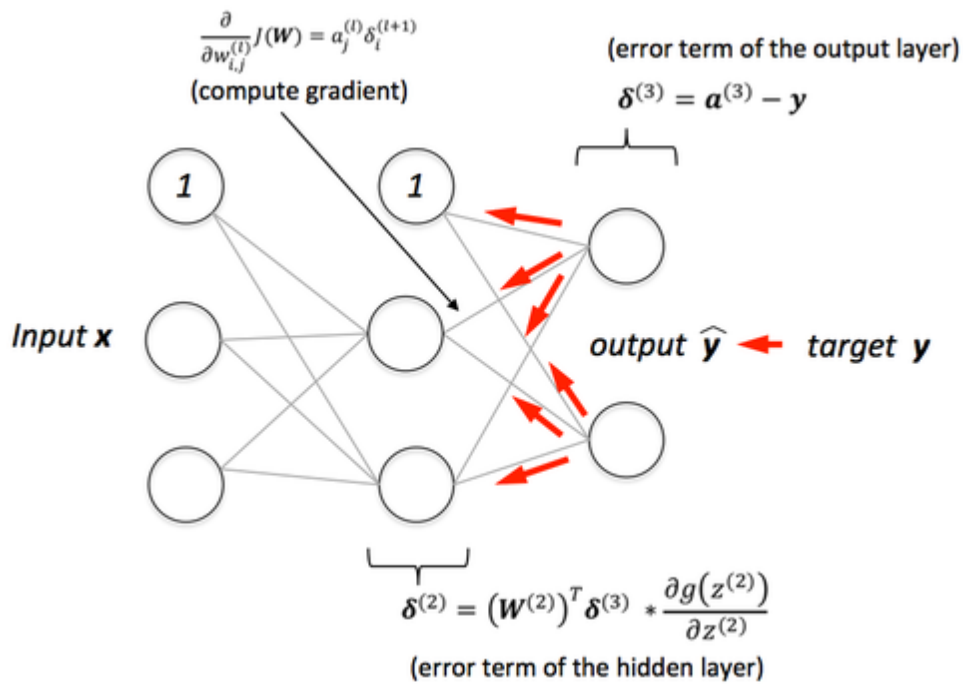


Figura 1.3: Algoritmul de Backpropagation[11]

1.2 Rețele neurale convoluționale

O rețea neurală convoluțională este un tip specializat de rețea neurală pentru procesarea datelor organizate sub o formă de tablou (un tablou unidimensional pentru datele temporale, un tablou bidimensional de pixeli pentru imagini) care folosește o operație matematică specializată numită convoluție în locul înmulțirii generale de matrici.[7]

O rețea convoluțională are anumite tipuri de straturi care nu apar într-o rețea clasică: straturi convoluționale și straturi de *pooling*, dar la final apar și straturi conectate (*fully-connected*) similare cu cele clasice. În cadrul straturilor convoluționale și celor de *pooling* neuronii din fiecare strat vor fi conectați doar la o mică regiune din stratul precedent, în loc de conexiuni între toți neuronii cum aveam la o arhitectură clasică.

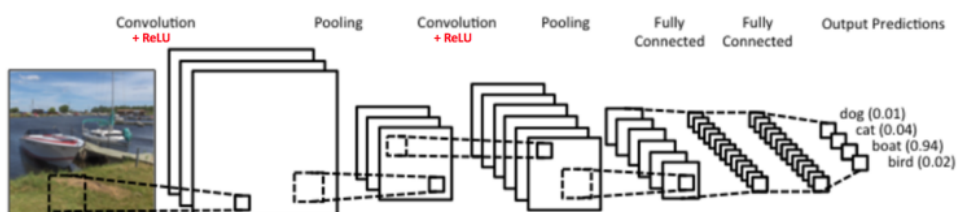


Figura 1.4: Arhitectura unei rețele convoluționale[2]

1.2.1 Convoluția

În cadrul unui strat de convoluție se folosește un filtru(sau *kernel*) pentru a extrage trăsături din cadrul imaginii. Filtrul se aplică unei mici regiuni din stratul precedent, realizând o mapare a imaginii în funcție de acel filtru. Cum imaginea este un tablou bidimensional de pixeli, un filtru va fi reprezentat tot de către un tablou bidimensional, mai mic decât cel de intrare. Aplicarea filtrului presupune calcularea produsului scalar între o parte din imagine de mărimea filtrului și filtru, efectuând această operație pentru toate părțile de aceeași mărime din imagine, luate la rând.

Un exemplu de aplicare a unui filtru pe o imagine:

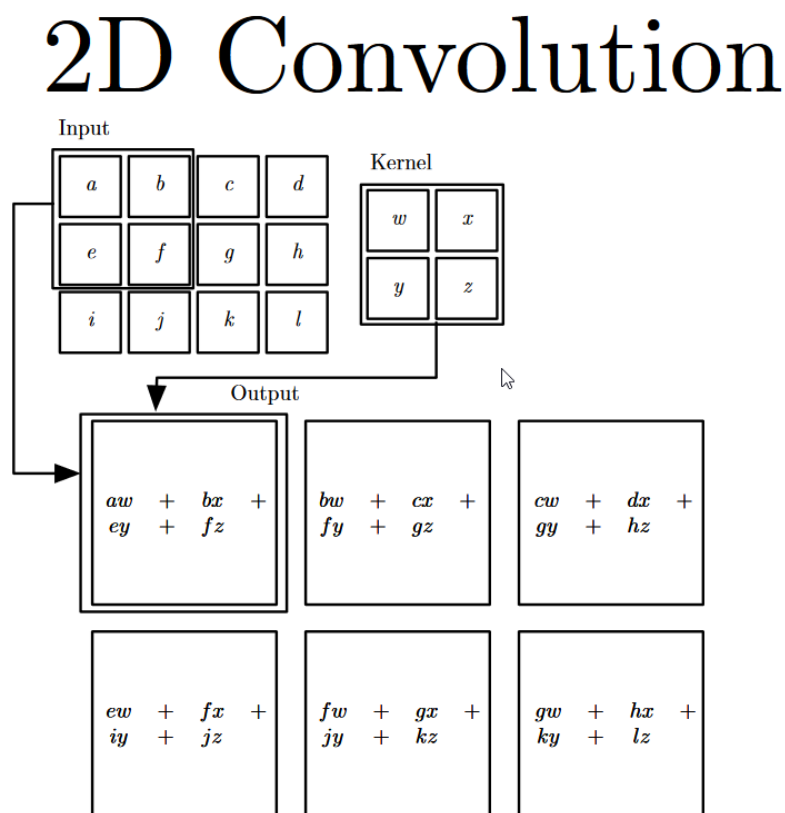


Figura 1.5: Operația de convoluție [7]

1.2.2 Pooling

Straturile de *pooling* (denumite și straturi de *subsampling* sau *downsampling*) vor prelua câte o mapare a imaginii dată de un filtru convoluțional și vor reduce dimensiunea fiecărei mapări, extrăgând astfel trăsăturile dominante. Pot exista mai multe tipuri de straturi de *pooling*: *MaxPooling*, *AveragePooling*, etc. De exemplu, în cazul operației

de *MaxPooling* ne alegem o zonă din mapare, din care extragem doar elementul maxim, astfel reducând dimensiunea mapării.

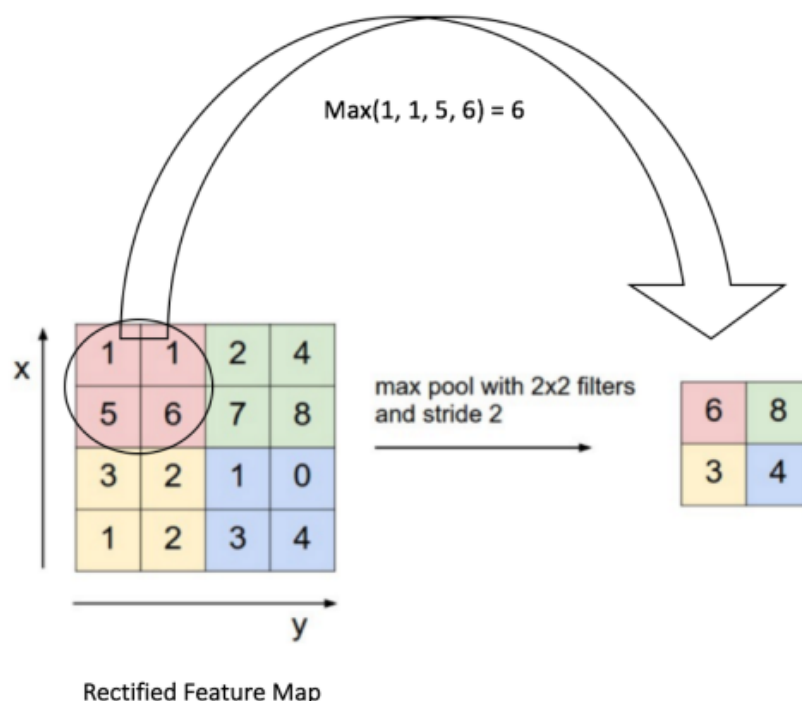


Figura 1.6: Exemplu de *MaxPooling*[2]

1.2.3 Clasificare

Adăugarea unui strat *fully-connected* ajută la clasificarea imaginii. Stratul de clasificare este de fapt un *MLP* tradițional, cu toți neuronii din el având conexiuni cu toți neuronii din stratul precedent. Stratul de clasificare va prelua trăsăturile de nivel înalt din imagine furnizate de straturile convoluționale și reduce ca dimensiune de straturile de pooling și va realiza o clasificare a imaginii bazată pe trăsăturile întâlnite.

Există multe tipuri de arhitecturi de rețele convoluționale, însă mai toate își au bazele pe operațiile descrise: convoluție, scalarea mapării (*pooling*), apoi clasificare.

1.3 Keras și Tensorflow

1.3.1 Tensorflow

Tensorflow este un *framework* cu sursă deschisă dezvoltat de Google și folosit pentru calcule numerice. Tensorflow oferă suport pentru învățare automată și *deep*

learning. Folosind diverse instrumente puse la dispoziție, utilizatorii pot construi modele de învățare automată la nivele diferite de abstractizare, de la modele construite pe baza unor serii de operații matematice la API-uri de nivel înalt pentru construirea de arhitecturi întregi disponibile, cum ar fi rețele neurale sau regresori liniari.

1.3.2 Keras

Keras reprezintă un API de nivel foarte înalt dedicat rețelelor neurale, scris în Python și care poate rula pe mai multe platforme de învățare automată cum ar fi TensorFlow, CNTK sau Theano. A fost dezvoltat pentru a abstractiza la maximum procesul de creare a unei rețele neurale, având ca scop atât experimentarea rapidă, cât și construcția într-un mod facil a unei arhitecturi solide. Keras poate fi văzut ca o interfață simplă, consistentă, folosită pentru cazurile uzuale, ce se pliază peste un *back-end* optimizat de *deep learning* și învățare automată. Este extrem de modular, permițând reutilizarea de blocuri configurabile, dar și extensibil, utilizatorii având posibilitatea de a crea noi straturi, funcții de cost sau arhitecturi pentru idei de cercetare.

1.4 OAuth

OAuth este un protocol deschis ce permite o metodă pentru o aplicație client să acceseze diverse resurse protejate de pe un server fără ca utilizatorul să își dezvăluie credențialele de pe server aplicației client.[9]

Procesul de autentificare OAuth folosește două tipuri de *token*[4]:

1. *Request token*

Folosit de aplicația client pentru a cere utilizatorului să autorizeze accesul la resursele protejate. *Token*-ul autorizat de utilizator este folosit o singură dată, este recomandat să expire după o limită de timp și va fi schimbat cu un *token* de acces.

2. *Access token*

Folosit de aplicația client să acceseze resursele protejate în numele utilizatorului.

Pașii folosiți în protocolul OAuth 1.0:

1. Aplicația trimite o cerere semnată pentru a primi un *Request token*
2. Serviciul furnizează aplicației *token*-ul solicitat

3. Aplicația redirectează utilizatorul spre o pagină de autorizare, unde utilizatorul își dă acordul, apoi se revine în contextul aplicației
4. Printr-o nouă cerere semnată, aplicația schimbă *Request Token*-ul cu un *token* de acces și un secret al acestui *token*
5. Aplicația folosește *token*-ul de acces și secretul acestuia pentru a accesa resursele protejate

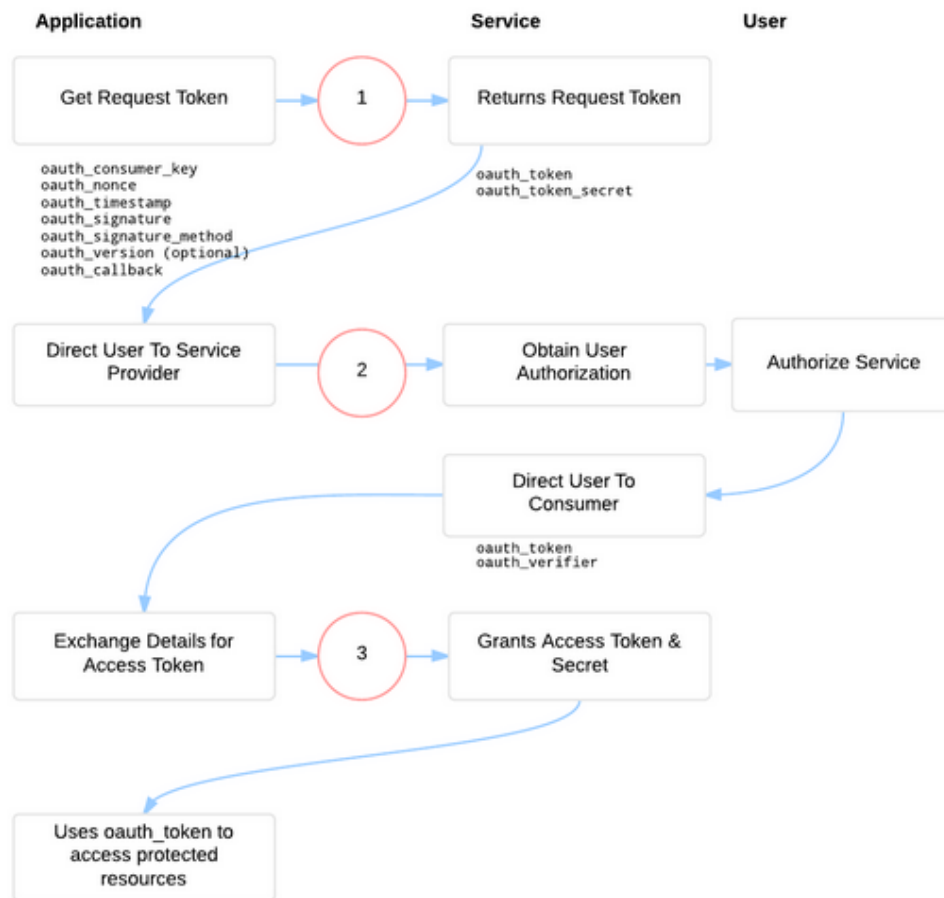


Figura 1.7: Autentificarea prin protocolul OAuth 1.0[3]

Capitolul 2

Scopuri și cerințe

2.1 Necesitatea aplicației

Aplicația "FlickrCloneFetch" permite utilizatorului să vizualizeze imagini similare vizual cu o imagine încărcată, având posibilitatea selecției sursei imaginilor: propriul profil de Flickr, conturile contactelor sale, sau imagini relevante publice de pe Flickr. Astfel, utilizatorul obține o selecție de imagini similare prin prisma conținutului, ci nu al altor meta-date(locația imaginii, data la care a fost realizată, etc).

Un exemplu în care "FlickrCloneFetch" se dovedește utilă este următorul: Utilizatorul dorește să revadă imaginile cu castelele vizitate postate pe profilul său de Flickr. Deoarece vizitele au fost la intervale diferite de timp, este clar că o căutare după o perioadă de timp nu îi va răspunde cerințelor. De asemenea, cum castelele se află în locuri diferite nici o căutare după locație nu va corespunde obiectivului. O căutare după titlu va eșua din nou, deoarece utilizatorul și-a numit fotografiile folosind doar numele castelelor vizitate, evident diferite. Folosind aplicația "FlickrCloneFetch", utilizatorul va încărca o imagine cu un castel și va selecta faptul că dorește să vadă doar imagini de pe propriul profil. Aplicația va analiza imaginile de pe profilul său, afișând doar cele în care apar castele și astfel răspunzând cerinței utilizatorului.

2.2 Modul în care aplicația rezolvă problema vizată

Utilizatorul se va autentifica în aplicație cu contul său de Flickr, astfel oferind aplicației acces la imaginile personale și la lista de contacte. După autentificare, utilizatorul va încărca o imagine în aplicație și va selecta sursa imaginilor ce vor fi analizate

(în mod general, sursa este profilul personal de Flickr).

Imaginea încărcată este analizată cu ajutorul unei rețele neurale, analiză în urma căreia se obțin anumite etichete ce descriu conținutul imaginii. Imaginile din sursa selectată de utilizator vor fi la rândul lor analizate cu aceeași rețea, etichetele obținute fiind comparate cu lista inițială de etichete, similaritatea conținutului fiind detectată prin existența etichetelor identice în imagini diferite.

Imaginile care au etichete ce apar și în lista de etichete a imaginii încărcate vor fi afișate în aplicație.

2.3 Comparația cu soluții existente

O soluție ce la prima vedere pare că ar rezolva problema expusă ar fi chiar căutarea oferită de *Flickr*. Însă, căutarea după anumite cuvinte cheie se realizează verificând existența cuvintelor în titlul, descrierea și lista de etichete ale imaginii. Astfel soluția de căutare nu oferă garanția faptului că ceea ce căutăm se află chiar în conținutul imaginii, titlul, descrierea și etichetele fiind alese de către utilizator după preferințele sale.

2.3.1 SimilarityPivot

SimilarityPivot este o unealtă oferită de Flickr care folosește rețele neurale pentru a analiza imaginile și a găsi imagini care au o anumită mărime, orientare sau paletă de culori. SimilarityPivot se poate combina cu căutarea după cuvinte cheie pentru a returna imagini ce conțin în titlu, descriere sau etichete acele cuvinte și în același timp respectă și cerințele vizuale descrise mai sus. Diferența între SimilarityPivot și "FlickrCloneFetch" constă în faptul că SimilarityPivot analizează în mod general imaginea, căutând doar anumite criterii vizuale (precum paleta de culori), în timp ce "FlickrCloneFetch" analizează imaginea pentru a descoperi etichete care să îi reprezinte conținutul. Astfel, cele două aplicații prezintă scenarii de utilizare diferite.

2.3.2 Reverse image search

Există mai multe sit-uri ce oferă posibilitatea de a efectua o căutare inversă după imagini, cel mai cunoscut pentru această funcționalitate fiind chiar Google. Funcționalitatea oferită de Google oferă utilizatorului posibilitatea să încarce o imagine, analizează imaginea și returnează o etichetă ce descrie conținutul imaginii. De asemenea, prezintă sit-

uri care corespund etichetei returnate și afișează imagini similare din punct de vedere vizual. Totuși, spre deosebire de o căutare uzuală, pentru acest tip de căutare nu se poate restrânge domeniul căutării la un singur sit, deci imaginile similare vor fi aduse de pe întreg web-ul, nu de pe o platformă specificată. De asemenea, căutarea poate găsi doar imagini publice, nu și imagini vizibile doar unui anumit utilizator (de pe profilul său, sau imagini ale contactelor cu vizibilitate doar pentru lista de prieteni). Astfel, diferența între soluțiile de căutare inversă după imagini și "FlickrCloneFetch" constă în domeniul căutării, precum și posibilitatea "FlickrCloneFetch" de a accesa fotografii vizibile doar unui utilizator autentificat.

2.3.3 Servicii pentru analiza imaginilor

Platformele de *cloud computing* oferă și servicii de analiză vizuală a imaginilor, precum CloudVision de la Google Cloud Platform, ComputerVision de la AmazonWeb-Services, și ComputerVision de la Microsoft Azure Cloud. Toate aceste servicii pot fi folosite pentru a realiza aplicații similare "FlickrCloneFetch", însă nu reprezintă în sine o soluție la problema expusă.

Capitolul 3

Analiză și proiectare

3.1 Scenarii de utilizare

Utilizatorul poate efectua următoarele acțiuni:

3.1.1 Autentificare

Utilizatorul se autentifică în aplicație, care îl va redirecta către platforma Flickr, pentru a efectua autentificarea prin intermediul acesteia.

3.1.2 Alege sursa imaginilor

Utilizatorul poate alege sursa imaginilor ce vor fi analizate: propriul profil, imaginile contactelor sale, sau imagini publice de pe platforma Flickr.

3.1.3 Încarcă imagine

Pentru a efectua această acțiune, utilizatorul trebuie să fi trecut prin pasul de autentificare și eventual să fi ales sursa imaginilor. Dacă acesta nu a ales sursa imaginilor, în mod implicit aceasta este propriul profil. Scenariul acesta de utilizare presupune ca utilizatorul să aleagă o imagine de pe propriul dispozitiv pe care să o trimită către aplicație.

Pentru a returna imaginile similare din punct de vedere vizual, aplicația va efectua următoarele acțiuni:

1. Analizează imaginea încărcată folosind o rețea neurală și returnează o serie de etichete descriptive

2. Aduce imaginile din sursa selectată de utilizator
3. Analizează imaginile aduse, obținând pentru fiecare o serie de etichete
4. Descoperă similaritatea între imagini comparând etichetele și obține o listă de imagini corespunzătoare
5. Afișează imaginile din lista obținută

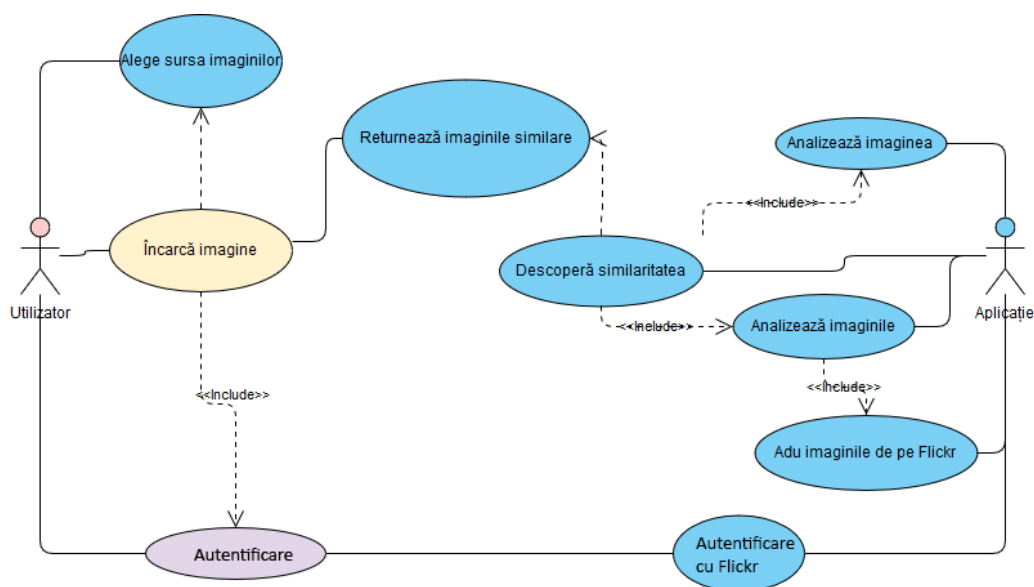


Figura 3.1: Diagrama de scenarii de utilizare a aplicației

3.2 Arhitectura aplicației

Aplicația este împărțită în două module, dar folosește și anumite servicii externe. Utilizatorul interacționează direct cu modulul *front-end*, o aplicație web realizată cu ajutorul *framework*-ului Angular în limbajul Typescript. La rândul său, modulul este împărțit în mai multe componente, acestea apelând unele servicii. Modulul de back-end este realizat cu ajutorul *framework*-ului Flask în limbajul Python. Se folosesc mai multe biblioteci pentru a efectua anumite operații(ex. *requests* pentru a apela API-uri, *pony.orm* pentru conectarea la baza de date) și *framework*-uri specializate pentru rețele neuronale(Tensorflow și Keras). Serviciile externe sunt reprezentate de API-urile platformelor Flickr(pentru autentificare, informații despre profil și imagini) și Data-Muse(pentru a obține sinonime). De asemenea se mai folosește o bază de date SQLite pentru reținerea *token*-ului de acces și identificatorului unic.

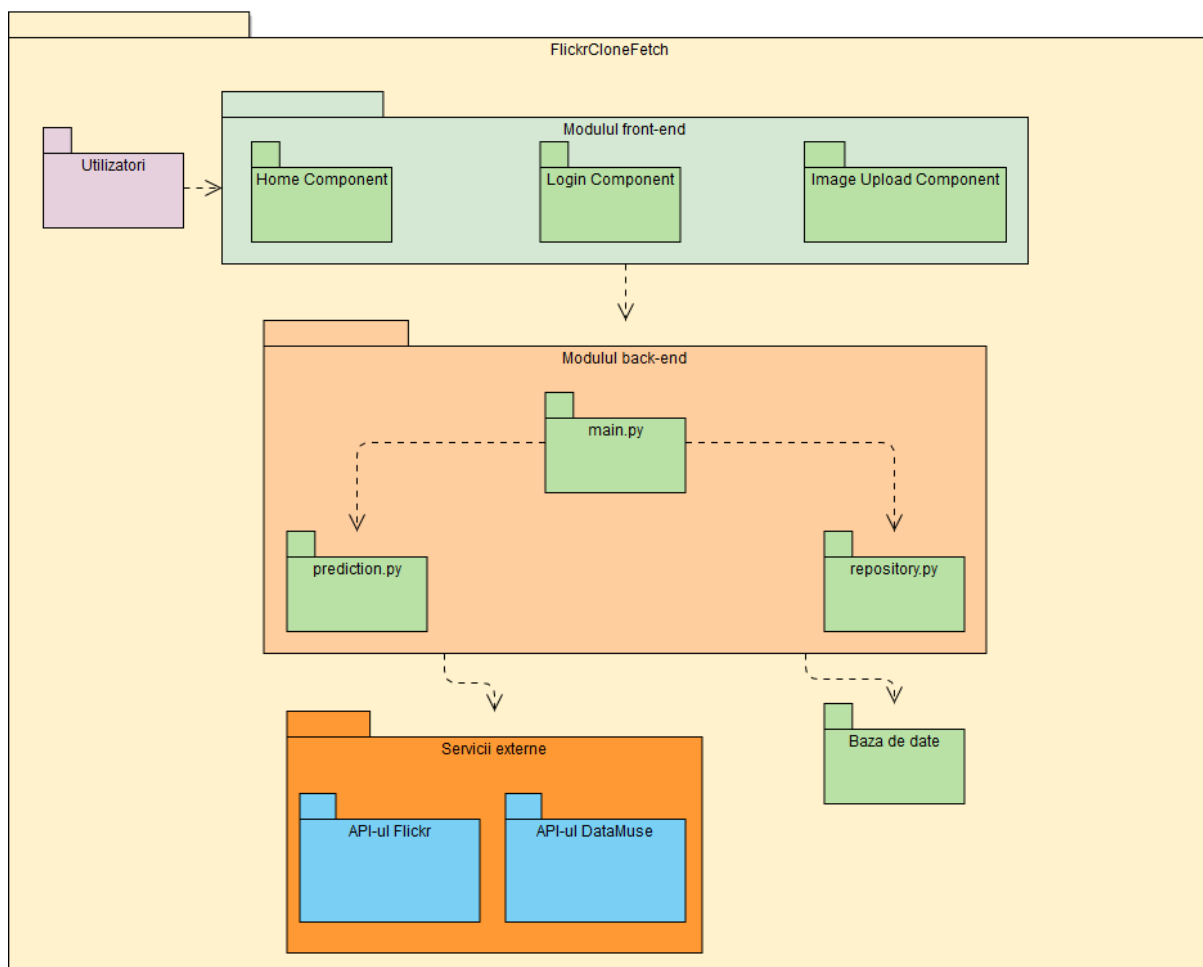


Figura 3.2: Arhitectura aplicației

Modulul *front-end* este împărțit în 3 componente: HomeComponent, LoginComponent și ImageUploadComponent, iar în cadrul modulului se folosesc și 3 servicii: AuthService, AuthGuardService și ImageService. HomeComponent are rolul de a oferi informații despre aplicație și a facilita procesul de autentificare. LoginComponent se ocupă de partea de autentificare în aplicație, folosind AuthService pentru a stoca anumite informații despre utilizator și a pregăti o experiență personalizată pentru acesta. ImageUploadComponent facilitează interacțiunea cu utilizatorul în cadrul scenariului de utilizare de încărcare a unei imagini. Folosind ImageUploadService, procesează cererile utilizatorului și răspunsul modulului de *back-end*, la final afișând informațiile cerute de către utilizator. AuthGuardService este folosit în procesul de securizare a aplicației, împiedicând accesarea ImageUploadComponent de către un utilizator neautentificat.

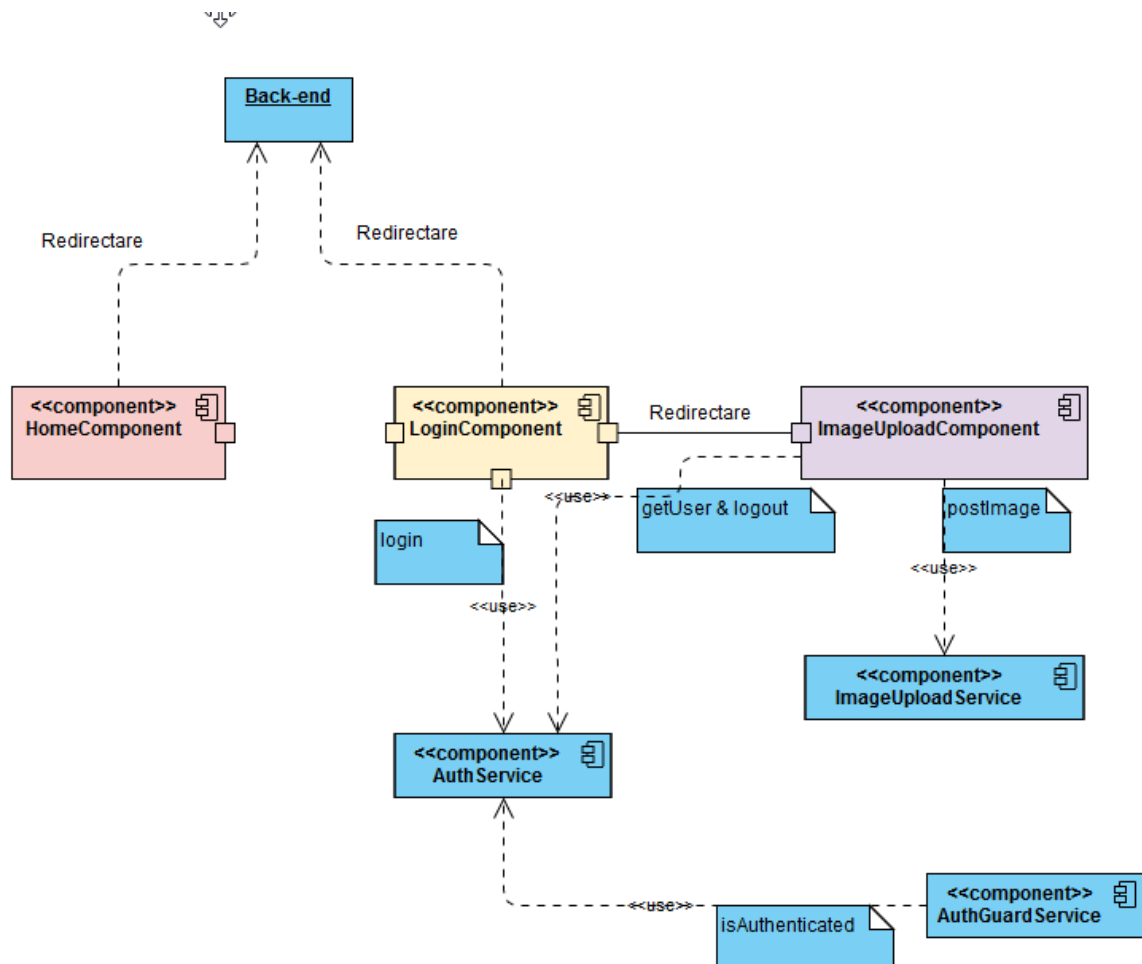


Figura 3.3: Arhitectura modulului front-end

Modulul *back-end* este împărțit în 3 componente principale: *main.py*, *prediction.py* și *repository.py*. Submodulul *main* are mai multe sarcini:

1. Autentificarea cu OAuth
2. Apelarea FlickrAPI
3. Apelarea DataMuse API
4. Comunicarea cu modulul *front-end*
5. Compararea listelor de etichete și detectarea similarității

Submodulul *repository* este folosit pentru comunicarea cu baza de date, folosindu-se biblioteca de cod *pony.orm* în acest scop. Submodulul *prediction* se ocupă de analiza imaginilor folosind o rețea neuronală și returnează lista de etichete ce descriu fiecare imagine în parte.

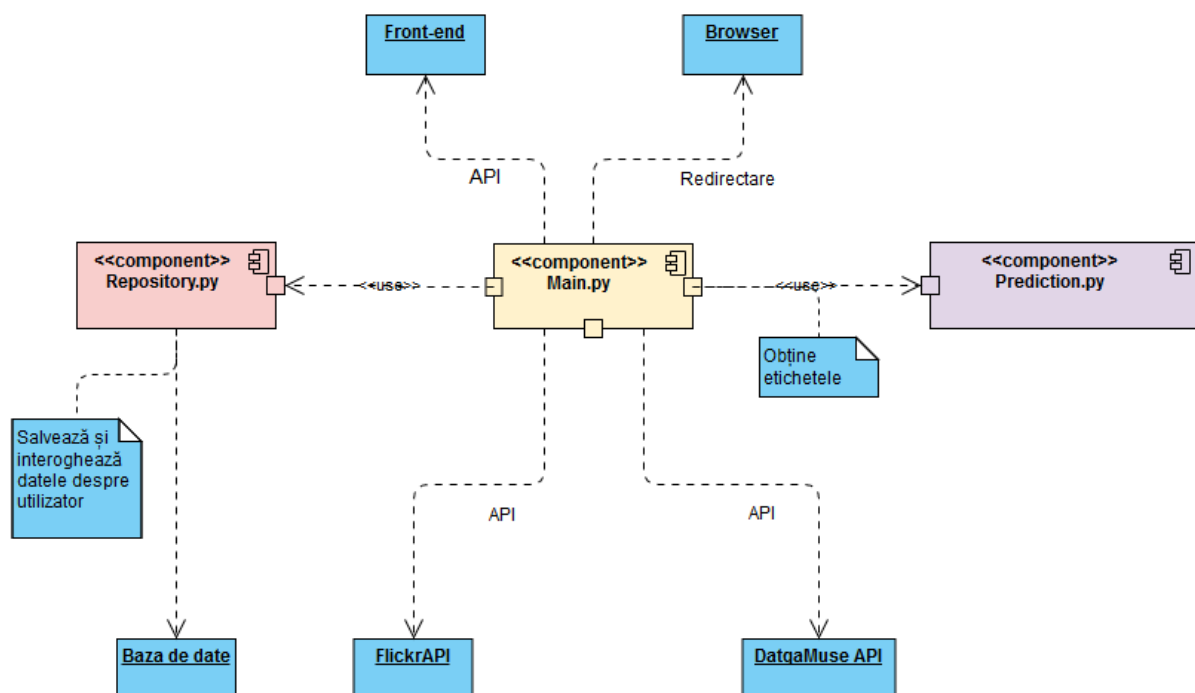


Figura 3.4: Arhitectura modulului back-end

3.3 Comunicarea între componentele aplicației

Utilizatorul va comunica cu modulul de front-end pentru a trimite cereri și a primi informațiile corespunzătoare cererii sale. De asemenea utilizatorul va interacționa cu pagini afișate de Flickr în browser pentru a autoriza aplicația "FlickrCloneFetch". Modulul front-end comunică cu utilizatorul și cu modulul back-end. Comunicare cu modulul back-end se realizează prin trimiterea de cereri HTTP căreiu un API pus la dispoziție de către back-end. Pe lângă comunicarea cu modulul front-end deja menționată, modulul back-end va mai apela doua API-uri: API-ul Flickr pentru preluarea de imagini și informații despre utilizator precum și autentificarea prin protocolul OAuth 1.0 și DataMuse API pentru preluarea de cuvinte similare ca înțeles unui cuvânt dat.

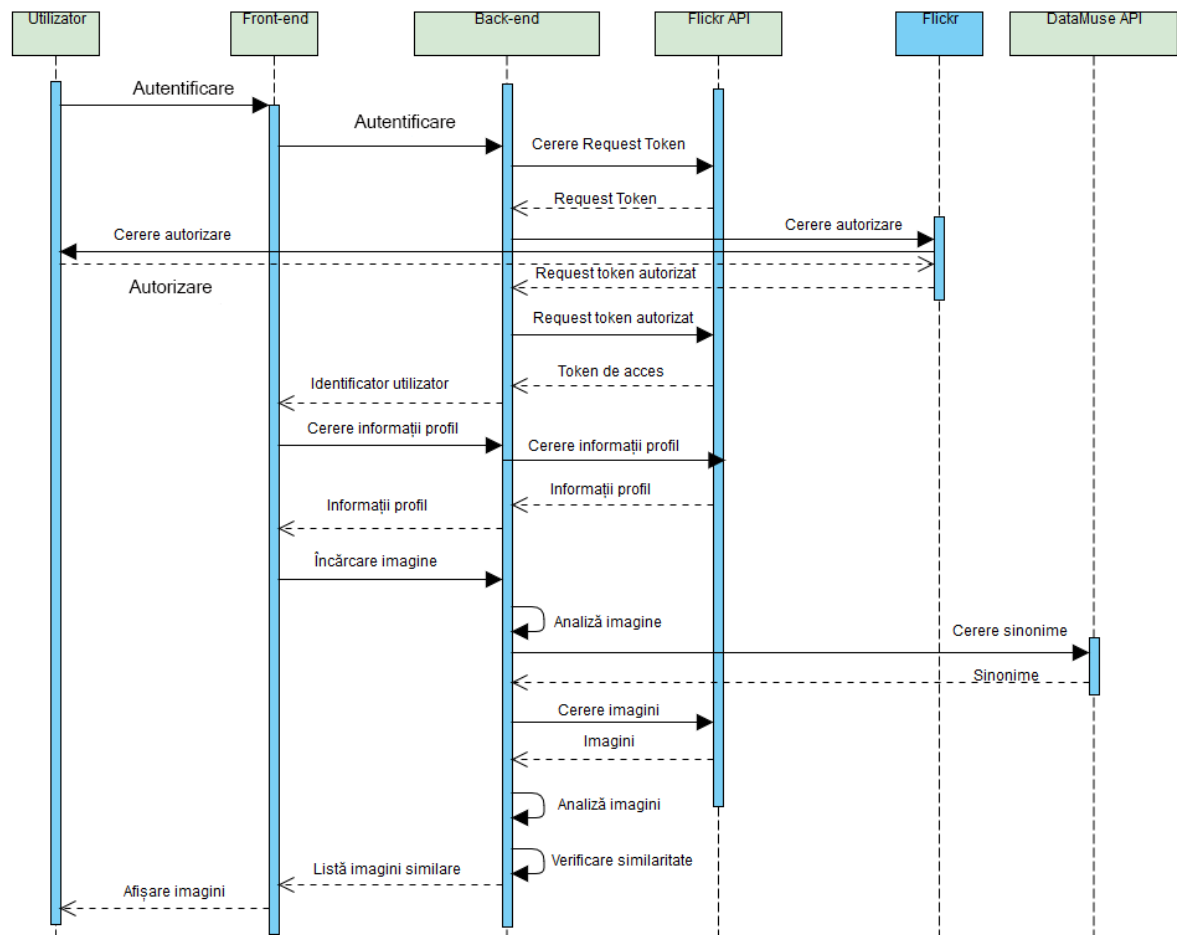


Figura 3.5: Diagrama de secvență a aplicației

3.4 Modulul *back-end*

3.4.1 API-ul expus

Specificația OPENApi a API-ului se găsește la Anexa 1

/login

Endpoint pentru autentificare, utilizatorul este redirectat aici de către modulul *front-end*. Aplicația va redirecta la rândul ei către */authorize*.

/authorize

Endpoint pentru autorizare, se ajunge aici din redirectarea efectuată de către aplicație. Are ca parametri *oauth-token* și *oauth-verifier* și va redirecționa către *front-end*.

/profile

Returnează informațiile despre profilul de Flickr al utilizatorului: numele utilizatorului și imaginea sa de profil. Are ca parametru identificatorul utilizatorului.

/addImage

Primește identificatorul utilizatorului, imaginea în base64 și opțiunea privitoare la sursa imaginilor ce vor fi analizate. Returnează lista de etichete a imaginii și lista de imagini similare din punct de vedere vizual.

3.4.2 Rețeaua neurală folosită

Rețeaua neurală folosită pentru analiza imaginilor este un model numit *NASNet Mobile*, obținut de Google prin *framework*-ul NAS(Neural Architecture Search) și discutat în lucrarea *Learning Transferable Architectures for Scalable Image Recognition*. [13]

Proiectul *AutoML*, prin *framework*-ul NAS, propune următorul lucru: o rețea neurală controler propune o arhitectură de rețea "copil" care va fi antrenată și evaluată pe o anumită sarcină. Evaluarea este folosită pentru a calibra rețeaua controler, care va propune idei îmbunătățite de arhitectură la iterația următoare. Procesul se repetă de mii de ori, generând noi arhitecturi, testându-le și evaluându-le până când se obțin arhitecturi de rețele neurale noi utile.[12]

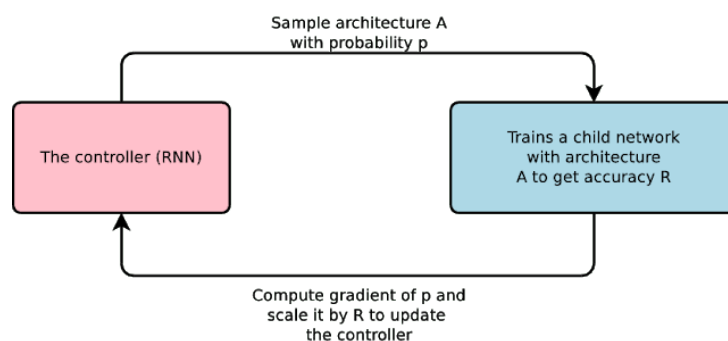


Figura 3.6: Procesul de obținere a noi rețele neurale folosind NAS [12]

Arhitectura *NASNet* este obținută prin procesul descris mai sus, cu unele mici modificări. Scopul rețelei e să clasifice imagini, deci a fost evaluată performanța pe două din cele mai cunoscute seturi de date în mediul academic: setul de detecție de obiecte *COCO* și setul de clasificare de imagini *ImageNet*. Deoarece seturile acestea de date sunt foarte mari, AutoML a fost reprogramat să caute straturi care ar fi atunci când sunt conectate în număr mare. Căutarea straturilor a avut loc pe un set de date mai mic, *CIFAR10*, apoi cea mai bună arhitectură a fost scalată și transferată pentru antrenamentul pe *COCO* și *ImageNet*. [13]

Arhitectura *NASNet* a fost predefinită, având două tipuri de straturi:

1. *Normal Cell* - un strat convoluțional care returnează o mapare de aceeași dimensiune a imaginii
2. *Reduction Cell* - un strat convoluțional care reduce dimensiunea mapării cu un factor de doi

Rețeaua neurală controler a trebuit doar să descopere cea mai bună arhitectură pentru cele două tipuri de straturi.

După aproximativ 4 zile de căutări au fost obținute mai multe arhitecturi de straturi candidat, care combinate produc arhitecturi cu rezultate comparabile cu alte arhitecturi din domeniu. Cele mai bune 3 arhitecturi obținute se numesc *NASNet-A*, *NASNet-B* și *NASNet-C*.

Modelul folosit în aplicația "FlickrCloneFetch", numit *NASNet Mobile*, este o versiune scalată a arhitecturii *NASNet*, mai mică, cu 5 326 716 parametri optimizată pentru un cost computațional redus și o viteză mai mare. Modelul folosit este pre-antrenat pe 1000 de categorii din setul de date *ImageNet*, având o acuratețe de 91.9 procente pe același set de date.

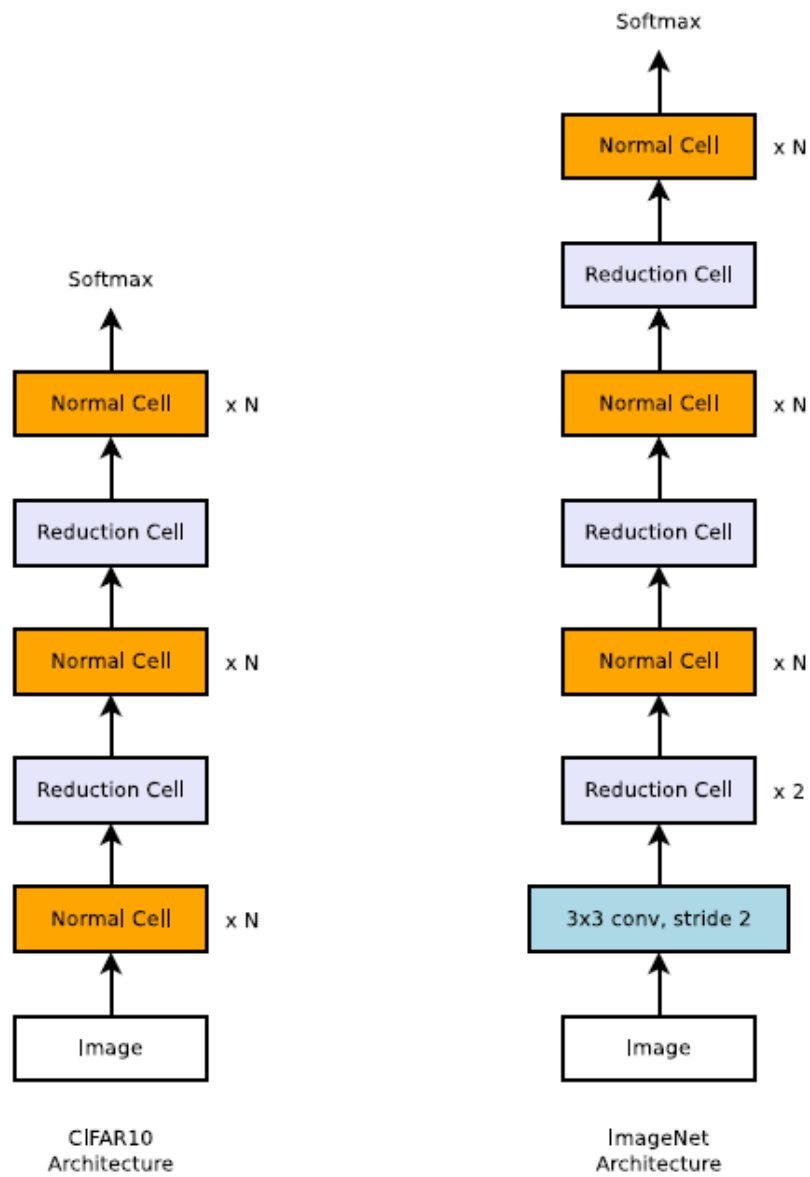


Figura 3.7: Arhitectura NASNet [13]

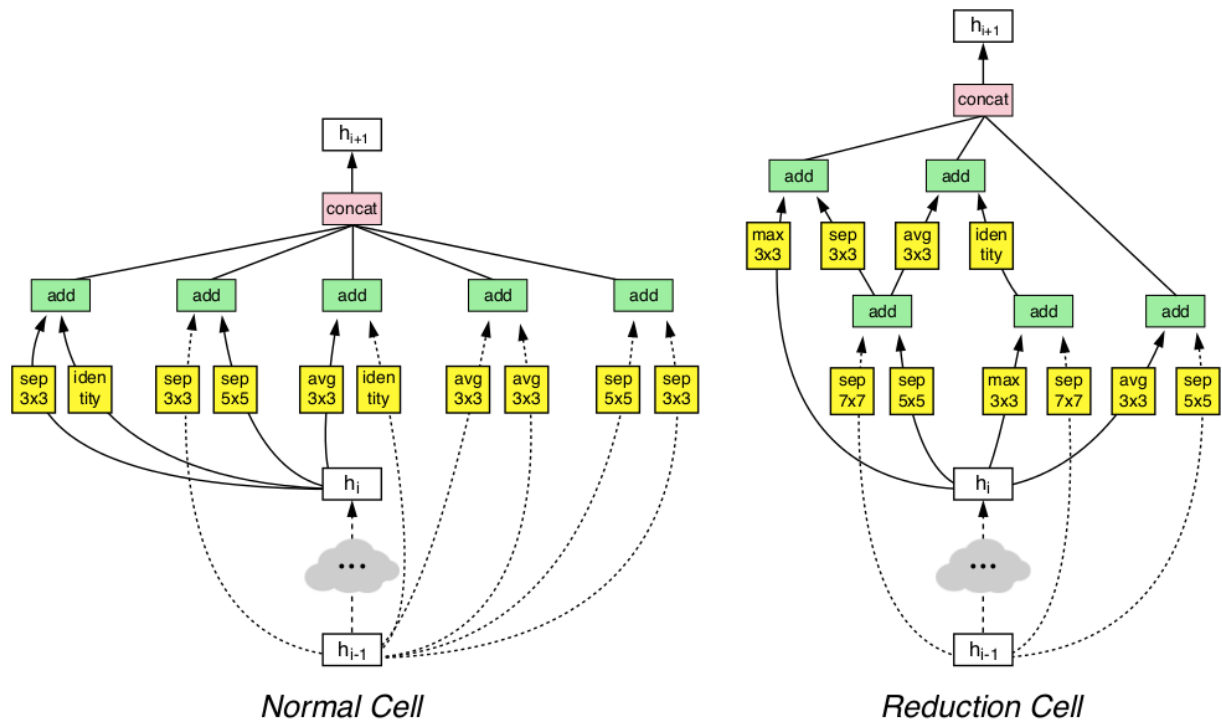


Figura 3.8: Celulele arhitecturii NASNet-A [13]

3.4.3 Baza de date

Baza de date folosită, de tip SQLite, are un singur tabel folosit pentru a stoca *token*-ul de acces și identificatorul unic al utilizatorului generat în relație cu acesta.

| users | |
|--------------------|---------|
| user_nsid | int |
| user_id | int |
| username | varchar |
| fullname | varchar |
| oauth_token | varchar |
| oauth_token_secret | varchar |

Figura 3.9: Schema bazei de date folosite

Capitolul 4

Detalii de implementare

4.1 Implementarea aplicației

Implementarea aplicației a fost realizată într-un mod care permite următoarele acțiuni în cadrul unui scenariu de utilizare uzual:

1. Modulul *front-end* afișează pagina de autentificare și redirecționează utilizatorul la procesul de autentificare de pe *back-end* (Home Component)
2. Modulul *back-end* trimite cererea de autentificare prin OAuth la API-ul Flickr
3. Se salvează *request token*-ului și redirecționează la autorizarea pagina de autentificare în Flickr și autorizare prin API-ul Flickr
4. Se trimite *request token*-ul autorizat de către utilizator pentru a obține un *token* de acces
5. Se generează un identificator unic pentru utilizator, se salvează în baza de date împreună cu *token*-ul de acces primit și apoi se face redirecționează către *front-end* trimițând și identificatorul unic
6. Modulul *front-end* reține identificatorul utilizatorului și cere datele despre profil
7. Modulul *back-end* furnizează datele despre profil (nume utilizator, imagine de profil)
8. *Front-end*-ul redirecționează utilizatorul la pagina de încărcare imagini (Login Component)

9. Aici i se oferă utilizatorului posibilitatea de a încărca o imagine și de a alege sursa imaginilor, apoi trimite la modulul *back-end* aceste informații (Image Upload Component)
10. Modulul *back-end* primește imaginea și opțiunile privitoare la sursă și analizează imaginea cu o rețea neurală, obținând o listă de etichete
11. Se apelează API-ul DataMuse pentru a obține cuvinte similare etichetelor obținute
12. Se apelează API-ul Flickr pentru a obține imagini și acestea se analizează cu rețeaua neurală
13. Se compară etichetele imaginilor cu etichetele imaginii încărcate și se creează o listă de imagini similare
14. Se returnează lista către *front-end*
15. *Front-end*-ul afișează imaginile returnate de *back-end* (Image Upload Component)

4.2 Detaliile modului front-end

Modulul *front-end* este realizat cu ajutorul *framework*-ului Angular. Angular este un *framework* cu sursă deschisă dezvoltat de către Google pentru crearea de aplicații client SPA (*single-page application*) folosind HTML, CSS și TypeScript. O aplicație Angular este compusă din module, componente și servicii.

Este împărțit în 3 componente: *HomeComponent*, *LoginComponent* și *ImageUploadComponent* și folosește 3 servicii: *AuthService*, *Auth-GuardService* și *ImageService*.

4.2.1 HomeComponent

Se află pe ruta `/home` și este prima componentă cu care interacționează orice utilizatorul. Rolul ei este de a oferi informații despre aplicație și a facilita procesul de autentificare. Are un singur buton, cel de "Login", care atunci când este apăsat va redirecta utilizatorul spre modulul de *back-end* pentru continuarea procesului de autentificare.

4.2.2 LoginComponent

Această componentă are ruta `/login/:id`, unde parametrul `id` va reprezenta identificatorul utilizatorului trimis de către serviciul de *back-end* prin redirectare. În cadrul componentei se va prelua valoarea parametrului din rută și se va apela funcția `login` din cadrul serviciului *AuthService*. Această funcție se ocupă de salvarea identificatorului. Inițial acesta era salvat în `LocalStorage`, dar din considerente de securitate, s-a ales ca soluție `SessionStorage`, care se golește la părăsirea sesiunii curente. Componenta apelează din nou *AuthService* și pentru a cere modulului de *back-end* informații despre profilul de Flickr ale utilizatorului, informații salvate în cadrul serviciului. După aceste acțiuni, se folosește routerul din Angular pentru a redirecta utilizatorul către următoarea componentă, *ImageUploadComponent*.

4.2.3 ImageUploadComponent

Componenta aferentă rutei `/upload`, care este o rută protejată prin o gardă de rută. Gărzile de rută în Angular sunt interfețe care îi spun routerului dacă poate sau nu naviga la o anumită rută. Routerul decide în funcție de o valoare booleană returnată de o clasă care implementează interfața gării. În cadrul aplicației s-a folosit garda `CanActivate`, a cărei interfață a fost implementată de serviciul `Auth-GuardService`. Acesta conține o metodă care returnează valoarea adevărat pentru când utilizatorul este autentificat (se determină cu ajutorul metodei `isAuthenticated` din cadrul *AuthService*, care caută dacă este salvat vreun identificator în `SessionStorage`) și fals pentru cazul în care nu este autentificat. Astfel *ImageUploadComponent* este accesibilă doar unui utilizator autentificat.

Componenta conține un buton "Choose Files" care odată apăsă deschide o fereastră de unde utilizatorul va alege o imagine, butonul având în spate un *input* HTML de tip `image`. După ce utilizatorul selectează o imagine, se va apela funcția *preview*, care va citi conținutul imaginii și o va afișa pe ecran, pentru ca utilizatorul să își vizualizeze selecția. În cadrul paginii există și un grup de 3 *radio-buttons*, care permit utilizatorului să selecteze de unde vrea să provină imaginile care vor fi analizate pentru a determina similaritatea. Opțiunile sunt profilul de Flickr utilizatorului, imaginile utilizatorilor din lista sa de contacte sau imaginile publice de pe Flickr, în mod implicit fiind selectată opțiunea corespunzătoare propriului profil.

Apăsarea butonului de "Upload" va apela funcția cu același nume. Funcția va

afișa un element de tip *spinner* pe ecran pentru a indica utilizatorului procesarea cererii. De asemenea va prelua identificatorul cu ajutorul funcției `getUser` din `AuthService`, funcție ce preia identificatorul din `SessionStorage`. Componenta va apela funcția `postImage` din serviciul `ImageService`, folosind imaginea în format `base64`, identificatorul utilizatorului și opțiunea acestuia ca parametri.

Folosind clientul de `http` din `Angular`, `ImageService` transmite o cerere la *back-end* și returnează răspunsul acestuia. În funcție de răspuns, componenta va afișa lista de etichete aferente imaginii și imaginile similare vizual, sau un mesaj explicativ în cazul în care nu au fost găsite etichete/imagini.

4.3 Detaliile modului back-end

Modulul de *back-end* este realizat cu ajutorul *framework*-ului `Flask` în limbajul `Python`. `Flask` este un *microframework* `WSGI`(*Web Server Gateway Interface*) pentru realizarea de aplicații web, scris în `Python`. Este cunoscut pentru simplitatea sa, dar și pentru posibilitatea de a se scala pentru aplicații oricât de complexe.

4.3.1 Autentificarea cu OAuth

Pentru a realiza autentificarea pe platforma `Flickr` prin protocolul `OAuth`, aplicația folosește o bibliotecă de cod numită *AuthLib* ce furnizează un client `OAuth` pentru `Flask`. Clientul este înregistrat în aplicație folosind anumite configurări (id-ul clientului, url-ul pentru autorizare, etc.) dintr-un fișier de configurări. Atunci când este accesată ruta `/login`, aplicația primește prin intermediul clientului `OAuth` un *request token* și generează un identificator unic pentru utilizatorul ce a accesat ruta, salvând *token*-ul într-un sistem de cache, apoi redirectează către ruta `/authorize`.

În cadrul funcției ce gestionează ruta `/authorize`, clientul `OAuth` schimbă *request token*-ul, cu un *token* `OAuth` de acces, pe care aplicația îl va salva împreună cu identificatorul generat într-o bază de date, apoi utilizatorul va fi redirectat la modulul de *front-end*.

4.3.2 Informații despre profil

Aplicația va face o cerere la una din metodele disponibile prin API-ul `Flickr`, va împacheta informațiile cerute (nume utilizator și imagine de profil) într-un `JSON` și le

va returna.

4.3.3 Analiza unei imagini

Pentru analiza unei imagini se folosește un model de rețea neurală disponibil prin biblioteca Keras. Numit *NASNetMobile* acest model este alcătuit din mai multe straturi convoluționale, fiind o versiune scalată a arhitecturii *NASNet*, versiune optimizată pentru viteză crescută. Modelul este preantrenat pe setul de date ImageNet, putând clasifica imagini în 1000 categorii. Biblioteca Keras oferă mai multe modele de rețele neurale preantrenate pe același set de date, figura 4.1 conținând specificațiile lor.

| Model | Size | Top-1 Accuracy | Top-5 Accuracy | Parameters | Depth |
|-------------------|--------|----------------|----------------|-------------|-------|
| Xception | 88 MB | 0.790 | 0.945 | 22,910,480 | 126 |
| VGG16 | 528 MB | 0.713 | 0.901 | 138,357,544 | 23 |
| VGG19 | 549 MB | 0.713 | 0.900 | 143,667,240 | 26 |
| ResNet50 | 99 MB | 0.749 | 0.921 | 25,636,712 | 168 |
| InceptionV3 | 92 MB | 0.779 | 0.937 | 23,851,784 | 159 |
| InceptionResNetV2 | 215 MB | 0.803 | 0.953 | 55,873,736 | 572 |
| MobileNet | 16 MB | 0.704 | 0.895 | 4,253,864 | 88 |
| MobileNetV2 | 14 MB | 0.713 | 0.901 | 3,538,984 | 88 |
| DenseNet121 | 33 MB | 0.750 | 0.923 | 8,062,504 | 121 |
| DenseNet169 | 57 MB | 0.762 | 0.932 | 14,307,880 | 169 |
| DenseNet201 | 80 MB | 0.773 | 0.936 | 20,242,984 | 201 |
| NASNetMobile | 23 MB | 0.744 | 0.919 | 5,326,716 | - |
| NASNetLarge | 343 MB | 0.825 | 0.960 | 88,949,818 | - |

Figura 4.1: Modelele de rețele disponibile prin Keras

Pentru a alege un anumit model, s-au testat mai multe din cele disponibile, apoi s-a încercat a se face un compromis între acuratețe și viteza de analiză a unei imagini (cunoscută ca și viteza de *forward-pass* în rețea). Ambii parametri depind de mărimea modelului, numărul de parametri, tipul de straturi folosite, precum și arhitectura modelului în general.

De asemenea viteza depinde și de mărimea imaginii analizate(aceeași indiferent de model) și mărimea la care trebuie dimensionată imaginea pentru a fi acceptată de model(diferită). În tabela 4.1 se găsesc comparații ale modelelor încercate ținând cont de viteza medie de analiza a unei imagini (prima imagine după instanțierea modelului

durează mai mult și nu a fost luată în considerare pentru medie), acuratețea modelului și mărimea datelor de intrare acceptate(în format (lățime, înălțime)).

| Model | Viteză medie | Acuratețe top-1 | Acuratețe top-5 | Format |
|-------------------|--------------|-----------------|-----------------|-----------|
| NASNetMobile | 0.25 secunde | 0.7444 | 0.919 | (224,224) |
| ResNet50 | 0.6 secunde | 0.749 | 0.921 | (224,224) |
| MobileNetV2 | 0.25 secunde | 0.713 | 0.901 | (224,224) |
| VGG16 | 0.5 secunde | 0.713 | 0.901 | (224,224) |
| InceptionResNetV2 | 1.23 secunde | 788 | 6344 | (299,299) |

Tabela 4.1: Comparație între modelele încercate

Fiecare model acceptă propriul tip de date de intrare, din acest motiv trebuie folosite funcțiile de preprocesare corespunzătoare, disponibile împreună cu modelul în biblioteca Keras. Modelul returnează o listă de predicții ce trebuie interpretată folosind din nou o funcție specifică modelului. După aceasta se obține o listă de tuple, primul element din tuplă fiind o eticheta ce descrie imaginea, iar al doilea fiind probabilitatea etichetei de a descrie corect imaginea.

Instantierea modelului, preprocesarea imaginii și obținerea predicțiilor se realizează în modul următor:

```

1 # obtinerea modelului preantrenat
2 model = NASNetMobile(weights="imagenet")
3 # decodarea imaginii din base64
4 image_response = base64.b64decode(image_response)
5 # citirea continutului imaginii
6 img = image.load_img(BytesIO(image_response), target_size=(224, 224))
7 # procesarea imaginii in numpy
8 img = image.img_to_array(img)
9 img = np.expand_dims(img, axis=0)
10 # preprocesarea conform specificatiilor modelului
11 img = preprocess_input(img)
12 # obtinerea predictiilor
13 predictions = model.predict(img)
14 #decodificarea predictiilor
15 predictions = decode_predictions(predictions)[0]
```

În funcție de cerințe, aplicația folosește și un parametru care precizează probabilitatea minimă acceptată pentru o etichetă ca aceasta să se regăsească în lista de etichete

finală, acest parametru având momentan valoarea de 0.25.

4.3.4 Preluarea imaginilor de pe Flickr

În funcție de opțiunea aleasă pentru sursa imaginilor ce vor fi analizate, API-ul de la Flickr va fi folosit în mod diferit.

Totuși orice metodă este folosită, va fi preluată o listă de fotografii cu un format specific, un exemplu fiind:

```
1 <photos page="2" pages="89" perpage="10" total="881">
2   <photo id="2636" owner="47058503995@N01"
3     secret="a123456" server="2" title="test_04"
4     ispublic="1" isfriend="0" isfamily="0" />
5   <photo id="2635" owner="47058503995@N01"
6     secret="b123456" server="2" title="test_03"
7     ispublic="0" isfriend="1" isfamily="1" />
8   <photo id="2633" owner="47058503995@N01"
9     secret="c123456" server="2" title="test_01"
10    ispublic="1" isfriend="0" isfamily="0" />
11  <photo id="2610" owner="12037949754@N01"
12    secret="d123456" server="2" title="00_tall"
13    ispublic="1" isfriend="0" isfamily="0" />
14 </photos>
```

[1]

Folosirea metodei getPhotos

Endpoint-ul flickr.people.getPhotos returnează imaginile de pe profilul unui anumit utilizator. Pentru a utiliza metoda, cererea trebuie să fie autentificată folosind *token*-ul de acces. Metoda returnează imaginile vizibile utilizatorului autentificat. Aplicația folosește această metodă pentru a prelua imaginile de pe profilul utilizatorului, utilizând valoarea "user_id=me" ca și parametru în cerere.

Folosirea metodei search

În cazul opțiunii de a prelua imaginile de pe profilele contactelor, găsirea listei de contacte a utilizatorului și folosirea metodei getPhotos descrise anterior ar afecta negativ performanța aplicației, o cerere putând să aducă mii de fotografii și crescând

în mod neacceptat viteza de răspuns. Pentru a optimiza acest proces și pentru a prelua și imagini publice relevante de pe platformă, aplicația va folosi endpoint-ul flickr.photos.search, care returnează o listă de imagini ce corespund unor anumite criterii.

Pentru aducerea fotografiilor relevante din lista de contacte, se va folosi "contacts=1" ca și parametru în cerere. Tot ca și parametru se va folosi "text", adăugându-i ca valoare eticheta după care vrem să căutăm imagini. Căutarea se va face în descriere, titlu și etichetele imaginilor, așa cum s-a menționat anterior. Pentru a primi cele mai relevante imagini, cererea va include și parametrul "sort=relevance".

Transformarea unei imagini din formatul returnat de cele două metode într-un url se face folosind o funcție din aplicație ce construiește url-ul folosind informațiile primite, url construit după un format prestabilit.

```
1 data["farm-id"] = photo.get("farm", "")
2 data["server-id"] = photo.get("server")
3 data["id"] = photo.get("id")
4 data["secret"] = photo.get("secret")
5 photo_url = "https://farm{farm-id}.staticflickr.com/{server-id}/{id}-{secret}.jpg".format(**data)
```

4.3.5 Optimizarea procesului de returnare a imaginilor

Pentru a optimiza procesul de returnare a imaginilor, atât din punct de vedere al similarității vizuale, cât și a timpului de răspuns, folosim două procedee:

API-ul DataMuse

API-ul DataMuse permite căutarea unor sinonime, cuvinte înrudite sau cuvinte ce respectă anumite reguli în raport cu un cuvânt dat. Deoarece, de exemplu, e de dorit ca și imaginile etichetate "castel" și cele etichetate "palat" să fie returnate, deoarece au similaritate vizuală mare, se folosește API-ul pentru a căuta cuvinte cu înțeles similar cu etichetele. Pentru fiecare etichetă, se vor lua primele 3 cuvinte similare(dacă există) cu eticheta și se vor introduce în lista de etichete a imaginii. Astfel se optimizează performanța aplicației, detectându-se mai multe imagini similare din punct de vedere vizual.

Sistem de cache

Deoarece analiza unei singure imagini durează în medie 0.25 secunde, întreg procesul, ce include aducerea imaginilor, deschiderea fiecărui url și preluarea imaginii, analizarea ei, folosirea API-ului WordMuse și detectarea similarității ajunge să dureze în medie 60-70 de secunde pentru o cerere în care se analizează 100 de imagini, timpul crescând în mod evident o dată cu creșterea numărului de imagini. Soluția pentru a optimiza acest proces este folosirea unui sistem de cache, astfel încât cererile consecutive cu imagini etichetate identic de către rețea se vor rezolva mult mai repede, fiind necesar doar timpul de analiză al imaginii inițiale și accesarea sistemului de cache.

Capitolul 5

Manual de utilizare

Pentru utilizarea aplicației "FlickrCloneFetch", utilizatorul va urma pașii:

1. Va apăsa butonul de "Login" de pe pagina principală
2. Se va autentifica în contul de Flickr, pe pagina de Flickr pe care a fost redirectat
3. Va autoriza accesul aplicației la contul său de Flickr, apăsând "Authorize" din pagina deschisă de Flickr
4. Va apăsa butonul de "Choose file", apoi va alege o imagine de pe propriul dispozitiv
5. Va alege una din opțiunile sursei imaginilor ce vor fi analizate, apăsând unul din cele trei *radio-buttons* disponibile
6. Va apăsa butonul "Upload"
7. Aplicația îi va procesa cererea, apoi va afișa etichetele corespunzătoare imaginii și imaginile similare vizual, sau mesaje explicative dacă nu au fost găsite etichete/imagini corespunzătoare

Procesul de utilizare este simplificat de faptul că nu există nevoia de crearea de credențiale separate pentru aplicație, ci se folosește direct rutina de autentificare a platformei Flickr.

Concluzii

În lucrarea de față au fost prezentate scopul și cerințele, arhitectura și implementarea aplicației web "FlickrCloneFetch". Aceasta își propune să ofere utilizatorilor posibilitatea de a vizualiza imagini similare vizual cu o imagine încărcată, imaginile provenind de pe platforma Flickr, fie de pe profilul utilizatorului, profilele contactelor sale, sau din imaginile publice postate pe platformă.

Aplicația este împărțită în două module, dar folosește și servicii externe. Autentificarea utilizatorului în aplicație se realizează prin intermediul platformei Flickr, folosind protocolul OAuth. Modulul *front-end* este realizat în Typescript cu ajutorul Angular și se ocupă de medierea interacțiunii între utilizator și modulul de *back-end*. Acesta este scris în Python, iar cu ajutorul Flask expune un API cu care comunică modulul de *front-end*.

Analiza imaginilor este realizată cu ajutorul *framework*-ului cu sursă deschisă Tensorflow și a API-ului de nivel înalt Keras. Se folosește un model de rețea neuronală disponibil prin Keras și preantrenat pe setul de date ImageNet, putând clasifica imaginile în 1000 de categorii. Modelul este o versiune scalată a arhitecturii *NASNet*, propusă de cei de la Google și obținută prin proiectul AutoML și *framework*-ului NAS (Neural Architecture Search) și este formată din diverse tipuri de straturi convoluționale.

Imaginile sunt căutate și accesate folosind API-ul oferit de platforma Flickr, API folosit și în protocolul de autentificare.

Pentru optimizarea funcționalității, se folosește un sistem de cache pentru rezolvarea rapidă a cererilor identice consecutive și API-ul oferit de DataMuse pentru a căuta cuvinte cu înțeles similar, astfel mărind plaja de căutare a imaginilor similare vizual.

Există posibilități de dezvoltare ulterioară a aplicației în mai multe direcții:

Îmbunătățirea analizei imaginilor

Pentru a îmbunătăți analiza imaginilor, se poate reantrena modelul de rețea neurală folosit pe un seturi de date mai mari, astfel introducând noi categorii de clasificare. De asemenea se poate alege un alt model de rețea neurală, care să îl înlocuiască pe cel existent. O altă posibilitate este folosirea unui serviciu de *Computer Vision* oferit de platformele de *cloud computing*, precum Cloud Vision de la Google Cloud Platform.

Integrarea cu alte platforme

Aplicația poate fi integrată cu alte platforme sau rețele sociale cu accent pe imagini, precum Instagram, Pinterest sau Pixabay. Astfel s-ar adapta funcționalitatea aplicației la mai multe surse de date, dovedindu-se utilă pentru o gamă mai largă de utilizatori.

Bibliografie

- [1] ***. Flickr api documentation. <https://www.flickr.com/services/api/>.
- [2] ***. An intuitive explanation of convolutional neural networks. <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>.
- [3] ***. Oauth bible. <http://oauthbible.com/>.
- [4] ***. Oauth core 1.0 revision a. <https://oauth.net/core/1.0a/>.
- [5] Răzvan Benchea. Neural networks. <https://sites.google.com/view/rbenchea/neural-networks?authuser=0>.
- [6] Adi Chris. From perceptron to deep neural nets. <https://becominghuman.ai/from-perceptron-to-deep-neural-nets-504b8ff616e>.
- [7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [8] Kevin Gurney. *An Introduction to Neural Networks*. Taylor & Francis, Inc., 1997.
- [9] Eran Hammer-Lahav. The oauth 1.0 protocol. RFC 5849, RFC Editor, April 2010. <http://www.rfc-editor.org/rfc/rfc5849.txt>.
- [10] Michael A. Nielsen. *Neural Networks and Deep Learning*. De3termination Press, 2015.
- [11] Sebastian Raschka. A visual explanation of the back propagation algorithm for neural networks. <https://www.kdnuggets.com/2016/06/visual-explanation-backpropagation-algorithm-neural-networks.html>.
- [12] Barret Zoph and Quoc V. Le. Neural Architecture Search with Reinforcement Learning. *arXiv e-prints*, page arXiv:1611.01578, Nov 2016.

- [13] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. *CoRR*, abs/1707.07012, 2017.

Anexa 1

```
1 {
2   "openapi": "3.0.2",
3   "info": {
4     "title": "Back-end",
5     "version": "1.0.0"
6   },
7   "paths": {
8     "/login": {
9       "summary": "Endpoint autentificare",
10      "description": "Enpoint-ul pentru autentificare,
11      ↪ utilizatorul este redirectat aici din browser.",
12      "get": {
13        "responses": {
14          "302": {
15            "description": "Redirecteaza catre /
16            ↪ authorize, incluzand oauth_token si oauth_verifier."
17          }
18        },
19        "summary": "GET pentru autentificare",
20        "description": "GET din browser, utilizatorul
21        ↪ fiind redirectat aici din front-end."
22      }
23    },
24    "/authorize": {
25      "summary": "Endpoint pentru autorizare",
```

```

23         "description": "Endpoint la care se ajunge din
↪ redirectarea efectuata de aplicatie.",
24         "get": {
25             "responses": {
26                 "302": {
27                     "description": "Redirecteaza catre
↪ front-end."
28                 }
29             },
30             "summary": "GET autorizare"
31         },
32         "parameters": [
33             {
34                 "name": "oauth_token",
35                 "description": "Tokenul OAuth",
36                 "schema": {
37                     "type": "string"
38                 },
39                 "in": "query",
40                 "required": true
41             },
42             {
43                 "name": "oauth_verifier",
44                 "description": "",
45                 "schema": {
46                     "type": "string"
47                 },
48                 "in": "query",
49                 "required": true
50             }
51         ]
52     },
53     "/addImage": {

```

```

54     "summary": "Adaugare imagine",
55     "description": "",
56     "post": {
57         "requestBody": {
58             "description": "Identificatorul
↪ utilizatorului, imaginea in base64 si sursa.",
59             "content": {
60                 "application/json": {
61                     "examples": {
62                         "post-image": {
63                             "value": {
64                                 "user": "id-ul
↪ utilizatorului",
65                                 "file": "imaginea in
↪ base64",
66                                 "option": "sursa
↪ imaginilor"
67                             }
68                         }
69                     }
70                 }
71             },
72             "required": true
73         },
74         "responses": {
75             "200": {
76                 "content": {
77                     "application/json": {
78                         "examples": {
79                             "full-example": {
80                                 "value": {
81                                     "labels": [
82                                         "label1",

```



```

83         "label2"
84     ],
85     "images": [
86         "url1",
87         "url2"
88     ]
89 }
90 },
91 "empty-example": {
92     "value": {
93         "labels": [
94         ],
95         "images": [
96         ]
97     }
98 }
99 }
100 }
101 },
102     "description": "Returneaza descrierea
    ↳ si imaginile."
103 },
104     "401": {
105         "description": "Daca utilizatorul nu
    ↳ este logat/ nu exista."
106     },
107     "400": {
108         "description": "JSON malformat/
    ↳ incomplet."
109     }
110 },
111     "summary": "POST imagine"
112 }

```

```

113     },
114     "/profile": {
115         "description": "Returneaza informatiile despre
↪ profilul de Flickr al utilizatorului.",
116         "get": {
117             "responses": {
118                 "200": {
119                     "content": {
120                         "application/json": {
121                             "examples": {
122                                 "profil": {
123                                     "value": {
124                                         "name": "nume-user
↪ ",
125                                         "buddyicon": "url
↪ imagine de profil"
126                                     }
127                                 }
128                             }
129                         }
130                     },
131                     "description": "Cazul in care
↪ utilizatorul exista si este logat."
132                 },
133                 "401": {
134                     "description": "Cazul in care
↪ utilizatorul nu este logat/ nu exista."
135                 }
136             },
137             "description": "Returneaza informatiile."
138         },
139         "parameters": [
140             {

```

```
141         "name": "user",
142         "description": "Identificatorul
↪ utilizatorului",
143         "schema": {
144             "type": "string"
145         },
146         "in": "query",
147         "required": true
148     }
149 ]
150 }
151 }
152 }
```