

# *Extragerea și recunoașterea cifrelor într-un document scris de mână*

Gandacov Ivan

## **Rezumat**

Proiectul își propune implementarea și demonstrarea unui algoritm de segmentare și detecție a cifrelor într-un mediu structurat, cum ar fi o foaie de caiet. Rezultatul final reprezintă o aplicație *desktop* de tip *e-learning* prin intermediul căreia este prezentat fiecare pas de procesare. Aplicația oferă posibilitatea analizei performanței algoritmului prin utilizarea diverselor configurații de parametri.

Algoritmul este compus din două etape de bază: etapa de segmentare și etapa de recunoaștere. La baza etapei de segmentare stă procesarea imaginii utilizând proiecțiile histogramei pe axa X, respectiv axa Y. O proiecție a histogramei pe axa Y permite determinarea limitelor liniilor orizontale de caractere, pe când proiecția pe axa X permite determinarea limitelor fiecărui caracter din liniile determinate. Rezultatul acestei etape reprezintă setul de regiuni de încadrare (*bounding boxes* (eng.)) pentru toate caracterele segmentate. Fiecare regiune de încadrare reprezintă o subimagine a imaginii inițiale.

La baza etapei de recunoaștere stă o rețea neuronală complet conectată (*fully connected neural network* (eng.)) a cărei specializare este detecția unei cifre în imagine. Rețeaua este antrenată în prealabil utilizând setul de imagini MNIST<sup>1</sup>. O regiune de încadrare, rezultată din etapa de segmentare, este furnizată la intrarea rețelei neuronale. Rezultatul detecției cifrei de către rețeaua neuronală reprezintă rezultatul etapei de recunoaștere.

Aplicația este compusă din următoarele nouă module:

- Modulul cu imaginea inițială
  - Afișează imaginea inițială și informațiile aferente, precum: numele, locația, formatul, dimensiunea în pixeli, adâncimea și dimensiunea în *kbytes*.
- Modulul de transformare a imaginii color în imagine cu nuanțe gri (*grayscale* (eng.))
  - Oferă posibilitatea scalării în jos a imaginii inițiale (*downscaling* (eng.)) și convertirii imaginii astfel obținute în imagine cu nuanțe gri. Reducerea dimensiunii imaginii reprezintă o metodă de îmbunătățire a performanței algoritmului de procesare.
- Modulul de extragere a tuturor conturilor din imagine
  - Modulul utilizează operatorul *Sobel* pentru extragerea conturilor din imagine. Parametrii operatorului pot fi modificați.
- Modulul de extragere a conturilor ce aparțin cifrelor
  - Modulul utilizează clusterizarea *KMeans* a pixelilor după intensitatea acestora. Astfel sunt formate *k* clustere de pixeli. Primele *n* clustere au intensitatea cea mai mare, deci cel mai probabil aparțin caracterelor din imagine. Acestea sunt colorate cu alb. Ultimele (*k-n*) clustere au intensitate

---

<sup>1</sup> Bază de date cu imagini ale cifrelor scrise de mână. Aceasta conține 60000 imagini pentru antrenare și 10000 imagini de test. MNIST reprezintă o submulțime a unei colecții mai largi, numită NIST. Dimensiunile imaginilor originale alb-negru din NIST au fost normalizate pentru a putea fi cuprinse în regiuni de încadrare de dimensiunea 20x20 pixeli, în același timp păstrând proporția dintre înălțime și lățime. Imaginile au fost încadrate în imagini de dimensiunea 28x28 pixeli prin translarea centrului de greutate în centrul imaginii de 28x28 pixeli.

redusă, deci cel mai probabil aparțin fundalului. Acestea sunt colorate cu negru. Prin urmare, imaginea rezultată va conține doar contururile cifrelor.

- Modulul de dilatare a caracterelor pentru „umplerea” discontinuităților din contururile cifrelor
  - Faptul apariției discontinuităților în contururile cifrelor este foarte probabil. Pentru „umplerea” acestora este utilizată operația de dilatare a conturilor.
- Modulul de determinare a regiunilor de încadrare ale cifrelor din imagine
  - Acest modul utilizează proiecțiile histogramei Y și X pentru determinarea limitelor liniilor, respectiv caracterelor din linii.
- Modulul de combinare a regiunilor de încadrare care se suprapun
  - Faptul existenței mai multor regiuni de încadrare suprapuse pentru un singur caracter este foarte probabil. Modulul curent combină regiunile de încadrare care se suprapun.
- Modulul de afișare a rezultatului segmentării pe imaginea inițială
- Modulul de recunoaștere a caracterelor.
  - Modulul oferă posibilitate antrenării unei rețele neuronale și rulării detecției cifrelor din regiunile de încadrare, determinate în modulele anterioare.

Modulele comunică utilizând modelul *pipeline*. Astfel ieșirea / rezultatul modulului curent reprezintă intrarea modulului următor. Modularizarea utilizând modelul *pipeline* urmărește scopul facilitării viitoarelor posibile modificări ale aplicației.

Următoarea este corespondența dintre modulele aplicației și algoritmul de procesare: primele opt module corespund etapei de segmentare, iar ultimul, al nouălea, corespunde etapei de recunoaștere a caracterelor.

Aplicația de demonstrație este implementată utilizând *Qt C++*. Mediul de dezvoltare utilizat în scopul redactării codului *C++* este *Qt Creator*. Interfața grafică a aplicației este realizată utilizând *Qt Designer*, o parte componentă a *Qt Creator*. *Qt Designer* oferă posibilitatea adăugării elementelor ferestrei prin metoda *drag&drop*.

Modulul de recunoaștere conține o secțiune de antrenare a rețelei neuronale. Deși parametrii antrenării, precum: viteza de învățare, dimensiunea lotului, numărul de epoci și opțiunea de afișare a graficului de performanță (funcția de cost) în timp real, sunt specificați în cadrul modulului, însuși antrenarea rețelei are loc într-un program separat. Acest program de antrenare oferă o interfață grafică de tip linie de comandă și oferă posibilitatea controlului și monitorizării procesului de antrenare a rețelei neuronale. Programul a fost implementat utilizând limbajul *Python*, și anume distribuția *Python Anaconda*. Pentru redactarea codului sursă a fost utilizat *IDE-ul Spyder*.

Sintetizând toate cele descrise în paragrafele anterioare, rezultatul lucrării de față reprezintă o aplicație de tip *e-learning* care are rolul de a demonstra etapele de procesare implicate în algoritmul de extragere și detecție a cifrelor scrise de mână.

Performanța algoritmului implementat este variabilă și depinde de mai mulți factori, precum: nivelul de zgomot (*noise*) în imagine, contrastul<sup>2</sup> imaginii, dimensiunea cifrelor<sup>3</sup> în imagine. În cadrul etapei de recunoaștere a caracterelor redimensionarea unei regiuni de încadrare la dimensiunea  $28 \times 28$  pixeli poate duce la pierderea unor informații. Prin urmare poate exista o eroare în detecția cifrei din regiunea de încadrare respectivă. Algoritmul este destul de robust în ceea ce privește extragerea conturilor ce aparțin caracterelor. Datorită repetatelor operații de corecție determinarea regiunilor de încadrare este efectivă. Combinarea regiunilor demonstrează o încadrare exactă a caracterelor din imagine. Un avantaj este faptul că algoritmul este robust în cazul existenței liniilor de text ale căror orientare orizontală deviază de la linia orizontală dreaptă. Acest lucru a fost asigurat prin divizarea imaginii în porțiuni pe orizontală. Astfel algoritmul de segmentare este rulat pentru fiecare porțiune.

---

2 Un contrast ridicat oferă o mai bună diferențiere dintre intensitățile pixelilor de background și foreground.

3 Algoritmul a fost implementat pentru a funcționa în condițiile existenței în imagine a liniilor cu caractere de înălțime relativ constantă.