Vintilescu Andreea-Alexandra

1. Definirea problemei

Scopul acestui proiect este de a identifica obiectele relevante din picturi, printr-o abordare de clasificare multi-label, în care fiecare imagine poate fi asociată cu mai multe etichete. Spre deosebire de modelele tradiționale de clasificare, care atribuie o singură etichetă fiecărei imagini, modelul propus în această lucrare are ca obiectiv recunoașterea tuturor obiectelor ce sunt prezente într-o pictură.

1. Specificul bazei de date

Setul de date utilizat, "The Paintings Dataset", cuprinde imagini cu picturi și etichete ce descriu obiectele prezente în fiecare lucrare. Acesta este împărțit în trei subseturi: *train*, *test* și *validation*, fiecare contribuind la antrenarea și evaluarea modelului. În cadrul acestui proiect, am descărcat și organizat imaginile în directoare corespunzătoare fiecărui subset, utilizând fișierele de etichete pentru a identifica obiectele din fiecare imagine. Un aspect mai puțin favorabil întâmpinat a fost etichetarea greșită a unor imagini, cum ar fi confuzia între "horse" și "house".

1. Arhitectura

Acest proiect a avut ca scop utilizarea arhitecturii ResNet-18, însa pentru a adapta acest model la cerințele temei curente de clasificare multi-label, am modificat ultimul strat al rețelei (stratul fully connected) pentru a genera un vector de scoruri pentru fiecare obiect în parte.

De asemenea, am folosit două funcții de pierdere: Focal Loss și BCEWithLogitsLoss. Focal Loss este un algoritm eficient în cazul datelor dezechilibrate, deoarece pune accent pe exemplele mai dificile de clasificat. În ceea ce privește optimizarea, am ales algoritmul Adam, ajustând parametrii acestuia, cum ar fi learning rate și dimensiunea batch-ului, pe parcursul experimentelor, pentru a maximiza performanța modelului.

1. Librăriile utilizate

În acest proiect, am folosit PyTorch ca librărie principală pentru construirea și antrenarea modelului. De asemenea, am utilizat torchvision pentru accesarea arhitecturii pre-antrenate ResNet-18 și pentru realizarea transformărilor necesare imaginilor (redimensionare, conversie în tensor, etc.).

1. Performanțe

Rețeaua neuronală a fost antrenată utilizând datele din setul de imagini din "**The Paintings Dataset**", mai exact 15k imagini, împărțite în subseturi pentru antrenare, testare și validare. Dintre acestea, ~8300 de imagini au fost utilizate pentru antrenare, iar restul pentru evaluarea performanței modelului.

În urma experimentelor realizate a fost intocmit tabelul de mai jos, unde se poate observa că funcția de pierdere FocalLoss() a fost mai eficientă în îmbunătățirea performanței pe termen lung, în special în ceea ce privește F1 score-ul și recall-ul. Comparativ cu BCEWithLogitsLoss(), care a obținut o precizie mai mare, dar cu un recall mai scăzut, FocalLoss() a demonstrat o performanță mai echilibrată în identificarea tuturor obiectelor din picturi.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Learning Rate | Batch Size | Criterion | Accuracy | F1 Score | Precision | Recall |
| 1e-4 | 32 | BCEWithLogitsLoss() | 0.265 | 0.273 | 0.847 | 0.202 |
| 1e-4 | 32 | FocalLoss() | 0.349 | 0.443 | 0.661 | 0.36 |
| 1e-4 | 64 | BCEWithLogitsLoss() | 0.281 | 0.29 | 0.852 | 0.241 |
| 1e-4 | 64 | FocalLoss() | 0.381 | 0.465 | 0.697 | 0.376 |
| 1e-3 | 32 | BCEWithLogitsLoss() | 0.426 | 0.542 | 0.735 | 0.456 |
| 1e-3 | 32 | FocalLoss() | 0.44 | 0.559 | 0.661 | 0.513 |
| 1e-3 | 64 | BCEWithLogitsLoss() | 0.419 | 0.541 | 0.72 | 0.0469 |
| 1e-3 | 64 | FocalLoss() | 0.433 | 0.562 | 0.691 | 0.489 |

Tabel 1. Rezultate obținute la rularea a 20 de epoci de antrenare

1. Comparația cu alte rezultate

O imagine care conține text, captură de ecran, Font, linie

Descriere generată automatAlt articol relevant in care se regăseste baza de date este "[*The State of the Art: Object Retrieval in Paintings using Discriminative Regions*"](https://www.robots.ox.ac.uk/~vgg/publications/2014/Crowley14/)*- E. J. Crowley, A. Zisserman .* Obiectivul principal al lujcrării este de a identifica obiectele din picturi atunci când se folosesc clasificatori învățați pe imagini naturale, cum sunt cele din setul de date PASCAL VOC vs. imagini din Google. Aceștia au constatat că performanța scade atunci când este aplicat pe imaginile din Google, comparativ cu setul de date PASCAL VOC.

Figura 1. Tabel privind rezultatele obținute în articolul mai sus menționat

După cum se poate observa în tabel, acuratețea obținută în lucrarea prezentă este semnificativ mai mică decât cea din articolul citat, în comparație cu precizia unde valoarea maximă atinge 82,5%.

1. Concluzii

Pe baza rezultatelor obținute, lucrării de față îi pot fi aduse posibile îmbunătățiri precum aplicarea unor transformări suplimentare pentru realizarea unui model mai robust și antrenarea pe un număr de epoci mai mare pentru îmbunătățirea treptată a performanței.

1. Bibliografie

[1]: *E. J. Crowley, A. Zisserman,* [**The State of the Art: Object Retrieval in Paintings using Discriminative Regions**](https://www.robots.ox.ac.uk/~vgg/publications/2014/Crowley14/)*, British Machine Vision Conference, 2014*