Raport ştiinţific şi tehnic

Denumire proiect (EN)	Object recognition in images using curriculum learning
Denumire proiect (RO)	Recunoaşterea obiectelor din imagini folosind învăţarea automată bazată pe curiculă
Acronim	CURL
Cod proiect	PN-III-P1-1.1-PD-2016-0787
Număr contract	15/2018
Contractor	UNIVERSITATEA BUCUREŞTI
Tip proiect	Proiect de Cercetare Postdoctorală (PD)
Autoritatea contractantă	Unitatea Executivă pentru Finanţarea Învăţământului Superior, a Cercetării, Dezvoltării şi Inovării
Perioada de raportare	01.05.2018 - 28.12.2018
Etapa de executie	1/2018
Director de proiect	Radu Tudor Ionescu

În conformitate cu activităţile prevăzute în Etapa 1 de raportare din Anexa II a contractului 15/2018, am efectuat următoarele:

- **Activitatea 1.1. Cercetarea, dezvoltarea şi evaluarea unui estimator de dificultate al imaginilor.**

Această activitate corespunde cu obiectivul 1 din cadrul propunerii de proiect, anume "Build a better image difficulty predictor". În vederea îndeplinirii acestul obiectiv, am utilizat o reţea neuronală convoluţională bazată pe arhitectura VGG-f [1]. Această reţea este pre-antrenată pentru problema recunoaşterii obiectelor în imagini. Pentru a adapta reţeaua în vederea estimării dificultăţii unei imagini, ultimul strat de neuroni al reţelei, anume stratul de clasificare "softmax", a fost înlocuit cu un strat de regresie (regression layer). Mai apoi, reţeaua astfel modificată a fost antrenată folosind funcţia de pierdere dată de media pătratelor erorilor (MSE). Rezultatele obţinute cu această reţea pe setul de date de testare, sunt mai slabe decât cele obţinute cu estimatorul de dificultate prezentat în [2], conform tabelului de mai jos.

Metodă	**MSE**	**Coeficientul Kendall Tau**
[2]	0,231	0,472
VGG-f + regression layer	0,243	0,420

Tabel 1. Rezultatele estimatorului de dificultate a imaginilor propus în comparaţie cu rezultatele estimatorului state-of-the-art.

Probabil că motivul pentru care modelul nostru atinge performanţe mai slabe este faptul că setul de date de antrenare cu aproximativ 5.000 de imagini este prea mic, reţeaua reuşind să supra-înveţe cu uşurinţă etichetele exemplelor de antrenare. Pentru prevenirea supra-învăţării am încercat două tehnici uzuale, anume oprirea timpurie (early stopping) şi regularizarea. În fapt, rezultatele prezentate în Tabelul 1 sunt obţinute utilizând aceste două tehnici de oprire a supra-învăţării. Mai precis, antrenarea reţelei a fost oprită după 20 de epoci, iar regularizarea a fost obţinută prin adăugarea unei rate de dropout de 50%. Cu toate aceste, rezultatele nu sunt satisfăcătoare. De menţionat, ca această situaţie a fost deja prevăzută în propunerea de proiect. Conform celor menţionate în propunerea de proiect, pentru îndeplinirea următoarelor obiective am recurs la utilizarea estimatorului de dificultate a imaginilor propus în [2].

În plus față de cele prevăzute în obiectivul 1, am antrenat un sistem pentru gradarea automată a eseurilor în limba engleză, rezultatele fiind publicate în articolul [3]. Acest sistem poate fi folosit pe viitor ca sistem de estimare a dificultății de înțelegere a textelor din perspectiva corectitudinii limbii și modului de exprimare în scris a autorului. Asemenea estimatorului de dificultate a imaginilor, estimatorul de dificultate a înțelegerii textelor poate fi utilizat pentru a antrena modele neuronale pentru clasificarea textelor folosind paradigma învățării pe bază de curiculă. Trebuie menționat că cercetarea în direcția de analiză a textelor a fost prevăzută doar ca o dezvoltare ulterioară a proiectului, după ce direcția de analiză a imaginilor va fi explorată. Totuși, rezultatele din [3] sunt un pas important în această direcție.

- **Activitatea 1.1. Cercetarea, dezvoltarea și evaluarea unei rețele neuronale convoluționale.**

Această activitate corespunde cu obiectivul 2 din cadrul propunerii de proiect, anume "Train a CNN model on ILSVRC from scratch in order to replicate the state-of-the-art results". În vederea îndeplinirii acestui obiectiv, am antrenat nu una ci mai multe arhitecturi de rețele neuronale convoluționale pentru detectarea și recunoașterea obiectelor în imagini. Scopul antrenării mai multor modele este pe de o parte pentru a ne familiariza cu modele state-of-the-art, iar pe de altă parte pentru a vedea care sunt avantajele și dezavantajele acestor modele în raport cu timpul de calcul, dar și cu acuratețea în detectarea și recunoașterea obiectelor. Astfel, modele antrenate sunt:

(a) Un detector de obiecte de tipul "single-stage detector" bazat pe arhitectura MobileNet-SSD [4], ce recunoaște cele 20 de clase de obiecte din setul de date PASCAL VOC 2007 [5]. Acest detector este foarte rapid, timpul necesar pentru procesarea unei imagini fiind de 0,07 secunde pe un CPU. În schimb, rata de recunoaștere este scazută, anume 66,68%.

(b) Un detector de obiecte de tipul "single-stage detector" bazat pe arhitectura SSD300 [6], ce recunoaște cele 20 de clase de obiecte din setul de date PASCAL VOC 2007 [5]. Acest detector este destul de rapid, timpul necesar pentru procesarea unei imagini fiind de 0,56 secunde pe un CPU. Rata de recunoaștere este de asemenea mică, anume 69,00%.

(c) Un detector de obiecte de tipul "two-stage detector" bazat pe arhitectura Faster R-CNN [7], ce recunoaște cele 20 de clase de obiecte din setul de date PASCAL VOC 2007 [5]. Acest detector este mai lent, timpul necesar pentru procesarea unei imagini fiind de 7,74 secunde pe un CPU. Totuși, procentul de recunoaștere este mult mai ridicat, anume 78,37%.

În plus față de obiectivul 2 din cadrul propunerii de proiect, am încercat combinarea acestor modele în **etapa de testare**, pentru a obține un compromis optim între viteză și timp. În acest sens am propus mai multe metode de combinare a modelelor de detectare a obiectelor, prin împărțirea imaginilor de test mai ușoare către detectorii (a) sau (b) și imaginilor mai grele către detectorul (c), folosind diverse criterii, anume:

(1) Distribuirea imaginilor în mod aleator.

(2) Distribuirea imaginilor conform scorului de dificultate dat de estimatorul de dificultate a imaginilor din [2].

(3) Distribuirea imaginilor conform numărului de obiecte detectate de către un detector de obiecte rapid, anume cel descris la punctul (a). O imagine cu mai multe obiecte este considerată a fi mai dificilă.

(4) Distribuirea imaginilor conform mediei mărimii obiectelor detectate de către un detector de obiecte rapid, anume cel descris la punctul (a). O imagine cu obiecte mai mici este considerată a fi mai dificilă.

(5) Distribuirea imaginilor conform raportului dintre numărul de obiecte detectate și media mărimii obiectelor detectate de către un detector de obiecte rapid, anume cel descris la punctul (a). O imagine cu multe obiecte mici este considerată a fi mai dificilă.

Dacă primul criteriu reprezintă o abordare de bază (neinformată), celelalte abordări merg pe ipoteza învățării bazată pe curiculă, anume pe distribuirea exemplelor ușoare către modelul mai rapid (dar cu performanță scăzută) și exemplelor grele către modelul mai lent (dar cu performanță ridicată). Rezultatele publicate în lucrările [8] și [9] demonstrează că metodele bazate pe curiculă obțin performanță mult superioară în comparație cu abordarea de bază de la punctul (1).

Pe lângă faptul că am studiat problema generală de detectare și recunoaștere a obiectelor, am încercat să studiem performanța unor modele state-of-the-art pe o problemă mai specifică, anume detectarea fețelor în imagini. Modele antrenate în acest caz, sunt:

(d) Un detector de fețe de tipul "single-stage detector" bazat pe arhitectura MobileNet-SSD [4]. Acest detector este foarte rapid, timpul necesar pentru procesarea unei imagini fiind de 0,28 secunde pe un CPU. În schimb, rata de recunoaștere este mai scazută, anume 89,10%.

(e) Un detector de obiecte de tipul "single-stage detector" bazat pe arhitectura S^3FD [10]. Acest detector este mai lent, timpul necesar pentru procesarea unei imagini fiind de 1,89 secunde pe un CPU. Totuși, procentul de recunoaștere este mult mai ridicat, anume 99,67%.

Criteriile pentru distribuirea imaginiilor de test între detectorul de la punctul (d) și detectorul de la punctul (e) sunt asemănătoare cu cele prezentate anterior în cazul problemei mai generale de detectare și recunoaștere a obiectelor:

(1) Distribuirea imaginilor în mod aleator.

(2.i) Distribuirea imaginilor conform scorului de dificultate dat de estimatorul de dificultate a imaginilor din [2].

(2.ii) Distribuirea imaginilor conform scorului de dificultate dat un estimator de dificultate a imaginilor antrenat doar pe imagini ce conțin persoane. Acest model este adaptat problemei detectării fețelor în imagini.

(3) Distribuirea imaginilor conform numărului de fețe detectate de către un detector de fețe rapid, anume cel descris la punctul (d). O imagine cu mai multe fețe este considerată a fi mai dificilă.

(4) Distribuirea imaginilor conform mediei mărimii fețelor detectate de către un detector de fețe rapid, anume cel descris la punctul (d). O imagine cu fețe mai mici este considerată a fi mai dificilă.

(5) Distribuirea imaginilor conform raportului dintre numărul de fețe detectate și media mărimii fețelor detectate de către un detector de fețte rapid, anume cel descris la punctul (d). O imagine cu multe fețe mici este considerată a fi mai dificilă.

Rezultatele prezentate în lucrarea [11] demonstrează că metodele propuse la punctele (2.i) și (2.ii) aduc cele mai bun rezultate în practică. Totuși, și criteriile de distribuire de la punctele (3), (4) și (5) produc rezultate peste metoda de bază de la punctul (1).

Per total, rezultatele prezentate în detaliu în lucrările [8], [9] și [11] indică faptul că paradigma bazată pe curiculă este utilă în **faza de testare**. Conform următoarelor obiective din cadrul proiectului, în viitor dorim să demonstrăm utilitate paradigmei bazate pe curiculă în **faza de antrenare**.

- **Activitatatea 1.2. Diseminarea rezultatelor printr-un articol științific.**

În urma activităților de cercetare fundamentală efectuate, au rezultate 4 articole publicate în volume ale unor conferințe internaționale, dintre care 1 articol într-o conferință de categoria A* (ACL 2018) și 1 articol într-o conferință de categoria A (ICONIP 2018). Astfel, au fost îndeplinite cerințele minimale de diseminare a rezultatelor pe anul 2018, care prevedeau publicarea cel puțin a unui articol într-o conferință de categoria A sau A*. Articolele finanțate prin proiectului de cercetare sunt listate în continuare:

1. Mădălina Cozma, Andrei M. Butnaru, Radu Tudor Ionescu. Automated essay scoring with string kernels and word embeddings. In Proceedings of ACL, pp. 503–509, 2018. **(Rank A* Conference)**

2. Petru Soviany, Radu Tudor Ionescu. Optimizing the Trade-off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction. In Proceedings of SYNASC, 2018. **(Rank C Conference)**

3. Petru Soviany, Radu Tudor Ionescu. Frustratingly Easy Trade-off Optimization between Single-Stage and Two-Stage Deep Object Detectors. In Proceedings of CEFRL Workshop of ECCV, 2018. **(Rank B Workshop)**

4. Petru Soviany, Radu Tudor Ionescu. Continuous Trade-off Optimization between Fast and Accurate Deep Face Detectors. In Proceedings of ICONIP, 2018. **(Rank A Conference)**

Totodată, articolele finanțate prin proiectul PN-III-P1-1.1-PD-2016-0787 sunt listate și pe pagina oficială proiectului, situată la adresa: **http://curl-proj.herokuapp.com**

Referințe bibliografice:

[1] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In Proceedings of BMVC, 2014.

[2] R.T. Ionescu, B. Alexe, M. Leordeanu, M. Popescu, D. Papadopoulos, V. Ferrari. How hard can it be? Estimating the difficulty of visual search in an image. In Proceedings of CVPR, pp. 2157–2166, 2016.

[3] M. Cozma, A.M. Butnaru, R.T. Ionescu. Automated essay scoring with string kernels and word embeddings. In Proceedings of ACL, pp. 503–509, 2018.

[4] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

[5] M. Everingham, S.M. Eslami, L. van Gool, C.K. Williams, J. Winn, A. Zisserman. The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision, 111(1): 98–136, 2015.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg. SSD: Single Shot MultiBox Detector. In Proceedings of ECCV, pp. 21–37, 2016.

[7] S. Ren, K. He, R. Girshick, J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of NIPS, pp. 91–99, 2015.

[8] P. Soviany, R.T. Ionescu. Optimizing the Trade-off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction. In Proceedings of SYNASC, 2018.

[9] P. Soviany, R.T. Ionescu. Frustratingly Easy Trade-off Optimization between Single-Stage and Two-Stage Deep Object Detectors. In Proceedings of CEFRL Workshop of ECCV, 2018.

[10] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, S.Z. Li. S^3FD: Single shot scale-invariant face detector. In Proceedings of ICCV, pp. 192–201, 2017.

[11] P. Soviany, R.T. Ionescu. Continuous Trade-off Optimization between Fast and Accurate Deep Face Detectors. In Proceedings of ICONIP, 2018.

Data,
11.11.2018

Întocmit,
Radu Tudor Ionescu