Raport științific și tehnic

Denumire proiect (EN)	Object recognition in images using curriculum learning
Denumire proiect (RO)	Recunoașterea obiectelor din imagini folosind învățarea automată bazată pe curiculă
Acronim	CURL
Cod proiect	PN-III-P1-1.1-PD-2016-0787
Număr contract	15/2018
Contractor	UNIVERSITATEA BUCUREȘTI
Tip proiect	Proiect de Cercetare Postdoctorală (PD)
Autoritatea contractantă	Unitatea Executivă pentru Finanțarea Învățământului Superior, a Cercetării, Dezvoltării și Inovării
Perioada de raportare	01.01.2019 - 30.12.2019
Etapa de executie	2/2019
Director de proiect	Radu Tudor Ionescu

În conformitate cu activitățile prevăzute în Etapa 2 de raportare din Anexa II a contractului 15/2018, am efectuat următoarele:

- **Activitatea 2.1. Cercetarea, devzoltarea și evaluarea unor rețele neuronale convoluționale state-of-the-art folosind diverse strategii de învățare bazată pe curiculă.**

Această activitate corespunde cu obiectivul 3 din cadrul propunerii de proiect, anume „Train better CNN models using curriculum learning". În vederea îndeplinirii acestul obiectiv, am studiat 4 probleme clasice din domeniul vederii artificiale și al procesării limbajului natural, anume: (1) generarea imaginilor, (2) detectarea obiectelor în imagini, (3) detectarea evenimentelor anormale în video și (4) dezambiguizarea sensului cuvintelor.

(1) Pentru problema de generare a imaginilor, am pornit de la două modele state-of-the-art pentru generare de imagini, bazate pe rețele neuronale generative adversariale (GAN) [1]. Primul model (SNGAN) [2] generează imagini pornind de la vectori cu zgomot aleator și se bazează pe normalizare spectrală. Acest model obține rezultate de top în generarea imaginilor pe setul de date CIFAR-10 [3], set ce a fost folosit și în studiile experimentale din cadrul proiectului. Al doilea model (Cycle-GAN) [4] generează imagini pornind de la alte imagini, realizând o transformare a stilului sau a obiectelor din imagini, asigurând consistența ciclică a rezultatelor obținute. Pentru îmbunătățirea rezultatelor generative ale acestor modele din literatura recentă, anume SNGAN și Cycle-GAN, am aplicat 3 metode de învățare bazată pe curiculă, conform planului specificat în propunerea de proiect. Cele 3 metode studiate sunt următoarele:
a) Adăugarea treptată a exemplelor dificile. Rețelele generative adversariale se antrenează în mod uzual pe un set de imagini neadnotate. În vederea antrenării bazate pe curiculă, imaginile au fost etichetate cu un scor de dificultate, folosind estimatorul prezentat în articolul [5]. Exemplele au fost adăugate în procesul de învățare treptat, în ordinea scorului de dificultate.
b) Ponderarea exemplelor cu scorul de dificultate. Am propus o nouă funcție de pierdere pentru rețele generative adversariale, care ține cont de dificultatea exemplelor. Scopul este acela de a forța modelul să acorde o atenție mai mare exemplelor ușoare la început, acest scop fiind realizat prin creșterea valorii funcției de pierdere pentru exemplele ușoare. Pe parcursul

procesului de învățare, valoarea funcției de pieredere se egalizează, astfel că exemplele ușoare și dificile devin la fel de importante.
c) Selectarea exemplelor conform unei distribuții de probabilitate care să favorizeze alegerea exemplelor ușoare la început. Pe parcursul procesului de învățare, distribuția de probabilitate se egalizează, astfel că exemplele ușoare și dificile au probabilitate egală de a fi selectate pentru procesul de antrenare.

Rezultatele obținute pe 3 seturi de date demonstrează că toate cele 3 metode de învățare bazată pe curiculă aduc îmbunătățiri, atât prin prisma calității imaginilor generate cât și prin eficientizarea timpului de antrenare. Rezultatele obținute în generarea imaginilor folosind metodele de învățare bazată pe curiculă sunt prezentate în lucrarea [6].

(2) Pentru problema de detectare a obiectelor, am utilizat o rețea neuronală convoluțională pentru detectare de obiecte, anume arhitectura Faster R-CNN [7]. Această rețea este pre-antrenată pentru problema detectării obiectelor în imagini pe un anumit set de date, denumit set sursă. În continuare, am studiat rezultatele obținute de Faster R-CNN pe un set de date dintr-o distribuție diferită, denumit set țintă. Fără a folosi imagini din setul țintă, rezultele modelului Faster R-CNN sunt slabe. Pentru creșterea performanței, modelul poate fi adaptat prin utilizarea unor imagini neadnotate din setul țintă. În vederea adaptării modelului, am studiat o variantă în care exemplele din setul țintă sunt alese aleator în comparație cu o varintă în care exemplele sunt luate în ordinea dificultății. Metoda de estimare aleasă pentru estimarea dificultății este de dată de numărul de obiecte detectate supra aria medie a acestor obiecte, această metodă fiind propusă în lucrarea [8] din Etapa 1 de implementare a proiectului. Rezultatele obținute pe 2 seturi de date demonstrează că metoda de adaptare bazată pe curiculă aduce îmbunătățiri considerabile asupra performanței, depășind tehnicile recente din literatură. Metodele și rezultatele obținute în detectarea obiectelor folosind metoda de adaptare bazată pe curiculă sunt prezentate în lucrarea [9].

(3) Pentru problema de detectare a evenimentelor anormale, am studiat două abordări prezentare în lucrările [10] și [11]. Algoritmul pentru detectarea și localizarea evenimentelor anormale din video prezentat în [10] se bazează atât pe trăsături ce reprezintă mișcarea din video cât și pe trăsături ce reprezintă înfățișarea sau postura obiectelor sau a persoanelor. Pentru reprezentarea mișcării, extragem gradienți de mișcare 3D pe care îi acumulăm în regiuni de dimensiune fixă. Aceste regiuni se numesc cuboizi spațio-temporali. Pentru procesarea ulterioară, păstrăm doar regiunile pentru care magnitudinea gradienților depășește un anumit prag. Pentru reprezentarea obiectelor și a posturii lor folosim o rețea neuronală convoluțională antrenată pe setul de date ImageNet pentru problema recunoașterii obiectelor din imagini. Trăsăturile extrase din rețeaua convoluțională sunt combinate cu gradienții de mișcare într-un vector de trăsături care reprezintă o sub-regiune spațio-temporală din video. În prima etapă de antrenare, folosim algoritmul de clusterizare k-means pentru a grupa vectorii de trăsături în funcție de similaritate. Grupurile de vectorii cu mai multe date vor fi mai ușor de învățat, în timp ce grupurile cu mai puține date vor fi mai dificile. În acest sens, utilizăm un prag prestabilit pentru a elimina grupurile cu mai puține elemente. Pentru fiecare grup rămas, antrenăm câte un clasificator SVM (Support Vector Machines) adaptat pentru o singură clasă. Fiecare clasificator este astfel antrenat pentru un nivel de dificultate diferit. Astfel, metoda rezultată realizează o învățare bazată pe curiculă în mod implicit, ea fiind propusă la punctul b) „Train specialized CNN models for easy, medium and difficult images.” al obiectivului 3 din propunerea de proiect. În faza de testare, clasificatorii se aplică pe cuboizi rezultați din fișierele video de test. Pentru fiecare exemplu considerăm scorul de anomalie ca fiind maximul dintre scorurile întoarse de clasificatorii SVM. În final, obținem scorurile la nivel de frame (cadru) considerând scorul maxim dintre cuboizii ce aparțin unui cadru din video. Scorurile astfel

obținute sunt netezite aplicând un filtru Gaussian pe dimensiunea temporală. Pentru detectarea și marcarea cadrelor ce conțin evenimente anormale, se aplică un prag prestabilit peste scorurile calculate la nivel de cadru. Algoritmul pentru detectarea și localizarea evenimentelor anormale din video prezentat în [11] este asemănător. Ca principale diferențe, enumerăm înlocuirea (i) cuboizilor spațio-temporali cu trăsături învățate în mod automat de modele de tip convolutional auto-encoders și (ii) clasificatorilor SVM pentru o clasă cu clasificatori SVM pentru discriminare binară. În cazul clasificatorilor SVM pentru discriminare binară, antrenarea pentru fiecare grup de exemple este realizată prin adaugarea de exemple negative din celelalte grupuri. Rezultatele experimentale arată că algoritmul prezentat în [11] produce rezultate mai bune decât cel prezentat în [10].

(4) Pentru problema de dezambiguizarea sensului cuvintelor, am pornit de la algoritmul ShotgunWSD [12], peste care am aplicat tehnica de eliminare a valorilor aberante descrisă în lucrarea [10]. Algoritmul ShotgunWSD are ca scop găsirea unei combinații de sensuri pentru cuvintele ambigue dintr-un document text. Algoritmul scufundă cuvintele din text și cuvintele din definițiile cuvintelor regăsite în baza de cunoștințe WordNet, într-un spațiu vectorial folosind abordarea word2vec [13]. Pentru fiecare sens al unui cuvânt se realizează un vocabular (rezultat din WordNet) și se calculează vectorul median al vocabularului. Pentru găsirea combinațiilor de sensuri potrivite, utilizăm similaritatea cosinus între acești vectori. Totuși, putem elimina sensurile cuvintelor depărtate de cuvintele din document prin clusterizarea vectorilor de cuvinte folosind algoritmul k-means, la fel ca în lucrarea [10]. Eliminarea acestor sensuri, dificil de potrivit, conduce la creșterea acurateții de dezambiguizare. Metoda propusă și rezultatele obținute sunt prezentate în lucrarea [14].

- **Activitatatea 2.2. Diseminarea rezultatelor printr-un articol științific.**

În urma activităților de cercetare fundamentală efectuate, au rezultate 2 articole publicate în volume ale unor conferințe internaționale, dintre care 1 articol într-o conferință de categoria A* (CVPR 2019) și 1 articol într-o conferință de categoria A (WACV 2019). Pe lângă aceste articole, a mai rezultat 1 articol publicat într-un jurnal de categorie A (top 25%, zona roșie). Totodată, putem menționa și 1 articol acceptat spre publicare la o conferință de categoria A (WACV 2020) și 1 articol trimis spre publicare. Astfel, au fost îndeplinite cerințele minimale de diseminare a rezultatelor pe anul 2019, care prevedeau publicarea cel puțin a unui articol într-un jurnal de categoria A. Articolele finanțate prin proiectului de cercetare sunt listate în continuare:

1. R.T. Ionescu, S. Smeureanu, M. Popescu, B. Alexe. Detecting abnormal events in video using Narrowed Normality Clusters. In Proceedings of WACV, pp. 1951–1960, 2019. **(Conferință rang A)**

2. R.T. Ionescu, F.S. Khan, M.I. Georgescu, L. Shao. Object-centric Auto-encoders and Dummy Anomalies for Abnormal Event Detection in Video. In Proceedings of CVPR, pp. 7842–7851, 2019. **(Conferință rang A*)**

3. A. Butnaru, R.T. Ionescu. ShotgunWSD 2.0: An improved algorithm for global word sense disambiguation. IEEE Access, 7(1):120961–120975, 2019. **(Jurnal top 25%, zona roșie)**

4. P. Soviany, C. Ardei, R.T. Ionescu, M. Leordeanu. Image Difficulty Curriculum for Generative Adversarial Networks (CuGAN). In Proceedings of WACV, 2020 (va apărea). **(Conferință rang A)**

5. P. Soviany, R.T. Ionescu, P. Rota, N. Sebe. Curriculum Self-Paced Learning for Cross-Domain Object Detection. Arxiv, 2019 (trimis).

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] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. Generative adversarial nets. In Proceedings of NIPS, pp. 2672–2680, 2014.

[2] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida. Spectral normalization for generative adversarial networks. In Proceedings of ICLR, 2018.

[3] A. Krizhevsky, G. Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[4] J.Y. Zhu, T. Park, P. Isola, A.A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of ICCV, pp. 2223–2232, 2017.

[5] 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.

[6] P. Soviany, C. Ardei, R.T. Ionescu, M. Leordeanu. Image Difficulty Curriculum for Generative Adversarial Networks (CuGAN). In Proceedings of WACV, 2020.

[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. Frustratingly Easy Trade-off Optimization between Single-Stage and Two-Stage Deep Object Detectors. In Proceedings of CEFRL Workshop of ECCV, 2018.

[9] P. Soviany, R.T. Ionescu, P. Rota, N. Sebe. Curriculum Self-Paced Learning for Cross-Domain Object Detection. Arxiv, 2019.

[10] R.T. Ionescu, S. Smeureanu, M. Popescu, B. Alexe. Detecting abnormal events in video using Narrowed Normality Clusters. In Proceedings of WACV, pp. 1951–1960, 2019.

[11] R.T. Ionescu, F.S. Khan, M.I. Georgescu, L. Shao. Object-centric Auto-encoders and Dummy Anomalies for Abnormal Event Detection in Video. In Proceedings of CVPR, pp. 7842–7851, 2019.

[12] A. Butnaru, R.T. Ionescu, F. Hristea. ShotgunWSD: An unsupervised algorithm for global word sense disambiguation inspired by DNA sequencing. In Proceedings of EACL, pp. 915–925, 2017.

[13] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS, pp. 3111–3119, 2013.

[14] A. Butnaru, R.T. Ionescu. ShotgunWSD 2.0: An improved algorithm for global word sense disambiguation. IEEE Access, 7(1):120961–120975, 2019.

Data,
29.11.2019

Întocmit,
Radu Tudor Ionescu