

Învățare prin consolidare pentru locomoția agenților 3D

May 27, 2020

Contents

1	Introducere	5
1.1	Contextul problemei	5
1.2	Soluția propusa	5
2	Fundamente teoretice	7
2.1	Inteligența Artificială și subdomeniile ei	7
2.2	Tipuri de învățare automată	7
2.3	Rețele neuronale	7
2.4	Învățare prin consolidare	9
2.4.1	Introducere	9

Chapter 1

Introducere

1.1 Contextul problemei

Reconstituirea digitala al unui obiect reprezinta o problema propusa de cateva decenii si este activ tratata in domeniul Viziunii Artificiale si al Graficii Computerizate, avand ca scop final obtinerea unei forme cat mai fidele al obiectelor reale. Procedeele existente se folosesc de varii tipuri de date pentru obtinerea unei reconstructii, acestea fiind obtinute prin uzul de tehnologii precum: Camere clasice, Camere RGB-D, LiDaR, Raze X, Ultrasunete, RMN, CT etc. Considerand costul aferent fiecarei alternative prin hardware-ul dedicat necesar, se poate deduce ca cea mai ieftina solutie ar necesita o simpla camera RGB, cu viziune monotipica, pentru a creea imagini 2D. Acest fapt concretizeaza dorinta

Recuperarea dimensiunii a 3-a din poze 2D a fost telul multor studii in ultimii ani. Prima generatie de metode au tratat perspectiva geometrica din punct de vedere matematic, observand proiectia 2D al obiectelor 3D, si incercand sa creeze un proces reversibil. Solutiile cele mai bune ale acestei abordari necesita multiple cadre ce capturau diferite unghiuri ale obiectelor, acestea avand nevoie de o calibrare meticuloasa. A doua generatie de metode a folosit o memorie ce continea cunostinte anterioare despre obiecte. Se poate trage o paralela la capacitatea umana prin care este posibila deducerea formei si geometriei unui obiect folosind si doar un singur ochi, cu ajutorul cunostintelor precedente despre obiecte asemanatoare celui in cauza. Din acest punct de vedere, problema de reconstructie al obiectelor 3D devine o problema de recunoastere. Considerand in prezent eficienta solutiilor de Invatare Profunda pentru problemele de recunoastere, cat si cresterea cuantumului de noi date ce pot fi folosite drept date de antrenare, se justifica tendinta comunitatii stiintifice de a folosi ramuri ale Invatarii Profunde precum Retelele Neuronale Convolutionale sau Retele Neuronale Recurente pentru obtinerea geometriei si structurii 3D al obiectelor.

1.2 Soluția propusa

Aceasta lucrare abordeaza reconstructia 3D a obiectelor print intermediul retelelor neuronale, folosind doar una sau mai multe imagini 2D. Deoarece natura acestei probleme face parte din domeniul Viziunii Artificiale, sunt folosite preponderent Retele Neuronale Convolutionale. De asemenea, exista 3 mari tipuri de volume ce pot fi reconstruite: ansamblu de fasii, ansamblu de puncte si ansamblu de voxelii. Solutia actuala se foloseste de ultimul tip de volum prezentat.

Arhitectura propusa este compusa din 3 module:

- Auto-codificator: format dintr-un codificator ce extrage diferite trasaturi ale imaginii

primitive si un decodicator ce interpreteaza trasaturile extrase in volume voxelizate

- Unificator: realizeaza contopirea multiplelor volume deduse intr-un volum mai robust
- Rafinor: realizeaza cizelarea volumului unificat

Spre deosebire de solutii din aceeasi familie precum Pix2Vox , ce alocă un Auto-Codicator complet convolutional clasic, 3D-R2N2 , ce introduce aspecte recurente pentru tratarea cazurilor cu multiple poze, prin blocuri LSTM 3D Convolutionale si GRU 3D Convolutionale, solutia prezentata introduce, din punct de vedere arhitectural, nivele aditionale de convolutie al har-tilor de caracteristici la ultimele 3 trepte ale Codicatorului, rezultand in 3 volume voxelizate ce captureaza diferite reconstructii ale obiectului real. Dupa trecerea prin celelalte module ale arhitecturii, rezultatul final este un singur volum. Adicional, adaugarea si integrarea procedeelor precum extractor de caracteristici MobileNet V2, functii de activare Mish, regularizare DropBlock si optimizator Ranger aduc rezultatele lui YOYO peste SotA-ul actual pe datasetul Data3D-R2N2.

Cu ajutorul bibliotecii kaolin, volumele pot fi vizualizate intr-un mediu 3D interactiv, in care se poate analiza din orice unghi volumul voxelizat reconstruit. De asemenea, este posibila reprezentarea acestuia sub forma unui ansamblu de plase, pentru o reprezentare mai neteda.

Chapter 2

Fundamente teoretice

2.1 Inteligența Artificială și subdomeniile ei

Odată cu creșterea popularității soluțiilor de Inteligență Artificială s-a creat un nivel ridicat de confuzie despre cum se definește și diferențiază Inteligența Artificială de alte precum Invatare Automată, Invatare Profundă, Viziune Artificială și

2.2 Tipuri de Invatare Automată

2.3 Rețele neuronale

Rețelele neuronale sunt o formă de învățare automată. Unitățile fundamentale ale unei rețele se numesc noduri, care sunt adesea asemănate cu neuronii creierului uman. O rețea neuronală învață prin antrenamente care duc la formarea conexiunilor între noduri.

Din punct de vedere structural, rețelele neuronale sunt formate din mai multe straturi care procesează date de intrare. Fiecare strat este reprezentat de o transformare non-liniară care conduce la învățarea unor caracteristici care aparțin de un anumit nivel de abstractizare.

Cel mai simplu exemplu de strat al unei rețele neuronale este stratul dens. Acesta are ca scop transformarea datelor de intrare printr-o funcție non-liniară parametrizată, dată de următoarea formulă:

$$h = A \cdot (W \cdot x + b)$$

În cazul de mai sus, W este o pondere de mărime $n_h \times n_x$ și b este un bias de mărime n_h , unde x este dimensiunea datelor de intrare, iar h este dimensiunea datelor de ieșire. Funcția A se numește funcție de activare, și este elementul care oferă caracteristica de non-liniaritate stratului. Această funcție este aleasă în funcție de sarcina stratului. În funcție de problema pe care o rezolvă, o rețea poate avea un număr arbitrar de straturi înălțate precum în Figura 2.2.

Pentru ca rețeaua să poată învăța, aceasta trebuie să poată face diferența dintre rezultatele corecte și cele greșite. În acest scop, se definește funcția de cost. Valorile obținute pe baza acestei funcții denotă diferența dintre estimările făcute de către rețea și rezultatul așteptat. Un exemplu simplu și des întâlnit de funcție de cost este abaterea pătratică medie, care este definită prin următoarea formulă [?]:

$$F = \frac{1}{n} \sum (x_i - x'_i)^2$$

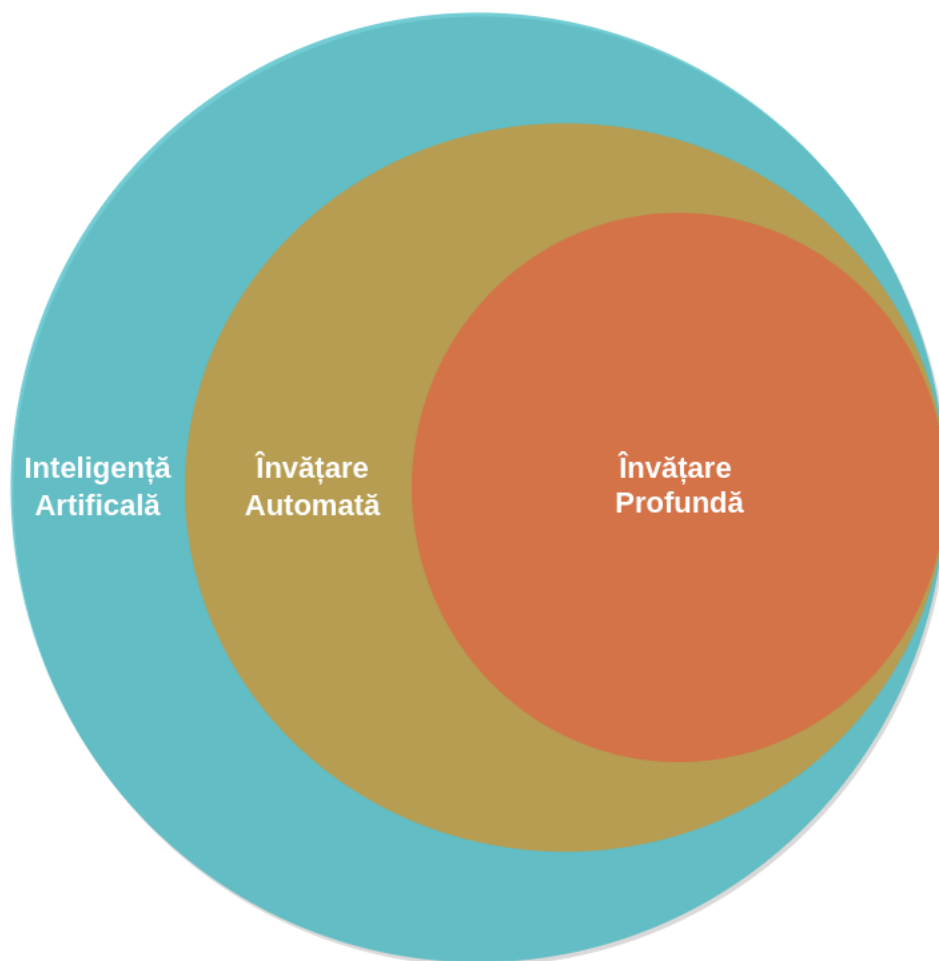


Figure 2.1: Subdomeniile Inteligentei Artificiale

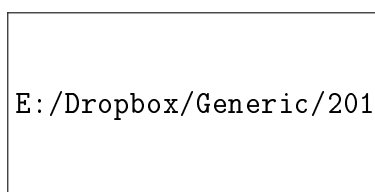


Figure 2.2: Exemplu de rețea neuronală simplă cu trei straturi

Odată ce funcția de cost a fost aleasă, rețeaua va învăța încercând să își optimizeze parametrii în mod iterativ astfel încât să minimizeze erorile rezultate. Cea mai populară metodă de optimizare este algoritmul de backpropagation [?], bazat pe direcția gradientilor funcției de cost.

2.4 Învățare prin consolidare

2.4.1 Introducere