Computer Vision 3

Ș.l. dr. ing. Mihai DOGARIU
www.mdogariu.aimultimedialab.ro

1

Structura cursului

M1. Introducere
M2. Fundamentele Învățării Adânci (Deep Learning Fundamentals)
M3. Învățare Adâncă Supervizată (Supervised Deep Learning)
M4. Învățare Adâncă Nesupervizată (Unsupervised Deep Learning)
M5. Învățare Consolidată (Reinforcement Learning)

2

M3. Învățare Adâncă Supervizată (Supervised Deep Learning)

3.1. Concept Supervised Deep Learning
3.2. Clasificarea imaginilor

3

M3.1. Concept Supervised Deep Learning

4

Supervised Deep Learning

Învățarea mașinilor (machine learning) = spunem despre un sistem că „învață" din experiența E cu privire la o clasă de sarcini de lucru T și o măsură de performanță P, dacă performanța sa în rezolvarea sarcinilor T, măsurată prin P, crește cu experiența E.

Bază de date = o grupare de elemente cu proprietăți comune. Reprezintă „experiența" pe care o întâlnește un algoritm de învățare conform definiției de mai sus.

$$D = \{((x_i, y_i)|T), 1 \le i \le M\}$$

input output sarcina dimensiunea

5

Supervised Deep Learning

$$D = \{((x_i, y_i)|T), 1 \le i \le M\} = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_M, y_M)\}$$

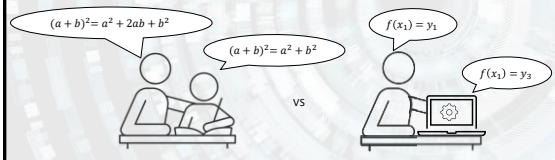
$$\left.\begin{array}{c} f(x_1) = y_1 \\ f(x_2) = y_2 \\ f(x_3) = y_3 \\ \dots \\ f(x_M) = y_M \end{array}\right\} f = ?$$

- Fiecare pereche (x_M, y_M) se mai numește și exemplu de antrenare;
- x_M = vector de intrare;
- y_M = ieșirea reală/eticheta.

6

Sursă imagine: iStock

Supervised Deep Learning



7

Supervised Deep Learning

Învățarea supervizată = paradigmă de învățare a mașinilor în care datele de antrenare sunt etichetate. Fiecare exemplu de antrenare este format dintr-un descriptor de trăsături și o etichetă. Scopul învățării supervizate este de a învăța funcția de asociere dintre trăsăturile de intrare și eticheta corespondentă.

Învățarea supervizată adâncă = paradigma învățării supervizate aplicată pe rețele neuronale adânci (adânci = mai multe (>3, >7, >30, >50) straturi).

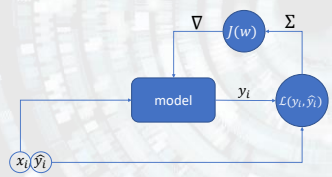
8

Supervised Deep Learning

- Toate datele de antrenament conțin o etichetă proprie;
- 2 subcategorii principale:
  1. Clasificare – ne dorim să prezicem o valoare discretă reprezentând cărei clase îi aparține un eșantion de intrare.
  2. Regresie – ne dorim să prezicem valori continue adaptate modelului care descrie baza de date.
- Coliziuni ale definițiilor:
  - Un clasificator poate prezice o valoare continuă sub forma unei distribuții de probabilitate.
  - Un regresor poate prezice o valoare discretă sub forma unei cantități întregi.

9

Supervised Deep Learning



10

M3.2. Clasificarea imaginilor

11

Clasificarea imaginilor

Clasificarea imaginilor = sarcina de a atribui o etichetă/clasă unei imagini.

Caracteristici:

- Conținutul imaginii este tratat ca un întreg ⇔ eticheta descrie întreaga imagine, nu doar o porțiune din ea.
- Orice imagine aparține unei singure clase.
- Ieșirea unui astfel de clasificator este, de obicei, o probabilitate, nu o decizie categorică.
- De obicei, sunt clasificate doar imagini în care obiectul/conceptul de interes ocupă o pondere semnificativă din imagine sau în care se găsește doar obiectul/conceptul de interes.
- Depinde foarte mult de aspectele calitative și cantitative ale bazei de date de antrenare.
- Clasificatorii multi-clasă au, de obicei, ultimul strat complet conectat și activare de tipul softmax.

12

Clasificarea imaginilor

Clasificator #1 → 78% Tabby cat, 15% Egyptian cat, 2% Siamese cat

Clasificator #2 → 85% cat, 10% dog, 3% table

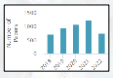
Clasificator #3 → 90% cat, 10% not cat

Baza de date | Modele | Metrici

13

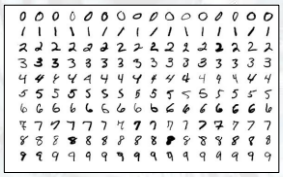
Sursă imagine: Wikipedia

Clasificarea imaginilor – datasets



MNIST [1]

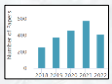
- Conține cifre scrise de mână;
- Train: 60k exemple;
- Test: 10k exemple;
- Dimensiune: 28x28 pixeli;
- Nuanțe de gri;
- Dimensiune normalizată;
- Imagini centrate;



14

Sursă imagine: Stanford

Clasificarea imaginilor – datasets



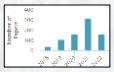
SVHN [2]

- Conține numere ale caselor de pe stradă;
- Train: 73k exemple;
- Test: 26k exemple;
- Extra train: 531k exemple;
- Dimensiune: 32x32 pixeli;
- Imagini color;
- Imagini decupate;
- Conține și parțial distractori.

15

Sursă imagine: University of Toronto

Clasificarea imaginilor – datasets



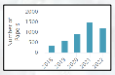
CIFAR-10 (Plantere) [3]

- Subset al Tiny Images Dataset;
- Conține imagini ale unor obiecte din viața reală;
- Train: 50k exemple;
- Test: 10k exemple;
- Dimensiune: 32x32 pixeli;
- 10 clase diferite;
- 6000 exemple/clasă;
- Imagini color;
- Clase echilibrate.

16

Sursă imagine: University of Toronto

Clasificarea imaginilor – datasets



CIFAR-100 [3]

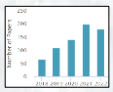
- Subset al Tiny Images Dataset;
- La fel ca CIFAR-10;
- 100 clase diferite;
- 600 exemple/clasă;
- 20 superclase (grosiere);
- 5 subclase (fine) /superclasă;

Superclass	Classes
aquatic mammals	beaver, dolphin, otter, seal, whale
fish	aquarium fish, flatfish, ray, shark, trout
flowers	orchids, poppies, roses, sunflowers, tulips
food containers	bottles, bowls, cans, cups, plates
fruit and vegetables	apples, mushrooms, oranges, pears, sweet peppers
household electrical devices	clock, computer keyboard, lamp, telephone, television
household furniture	bed, chair, couch, table, wardrobe
insects	bee, beetle, butterfly, caterpillar, cockroach
large carnivores	bear, leopard, lion, tiger, wolf
large man-made outdoor things	bridge, castle, house, road, skyscraper
large natural outdoor scenes	cloud, forest, mountain, plain, sea
large omnivores and herbivores	camel, cattle, chimpanzee, elephant, kangaroo
medium-sized mammals	fox, porcupine, possum, raccoon, skunk
non-insect invertebrates	crab, lobster, snail, spider, worm
people	baby, boy, girl, man, woman
reptiles	crocodile, dinosaur, lizard, snake, turtle
small mammals	hamster, mouse, rabbit, shrew, squirrel
trees	maple, oak, palm, pine, willow
vehicles 1	bicycle, bus, motorcycle, pickup truck, train
vehicles 2	lawn-mower, rocket, streetcar, tank, tractor

17

Sursă imagine: MIT

Clasificarea imaginilor – datasets



Places-365 [4]

- Utilizată pentru recunoașterea scenelor;
- >10M imagini;
- 434 clase;
- 5-30k imagini/clasă;
- Imagini color;
- Dimensiuni variabile;
- Generată automat => există zgomot.

18

Sursă imagine: Stanford

Clasificarea imaginilor – datasets

ImageNet [5]

- Conține imagini corespunzătoare synsets ale ierarhiei WordNet;
- 14M imagini color;
- 21841 clase;
- ~650 imagini/clasă;
- Dimensiuni variabile;
- Utilizată pentru competiția ILSVRC (ImageNet Large Scale Visual Recognition Challenge): 1.4M imagini, 1000 clase;
- Generată automat => zgomot;
- Cea mai populară în procesarea imaginilor.

19

Clasificarea imaginilor – metrici

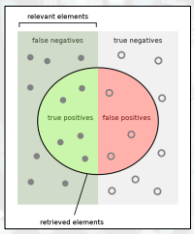
Metrică = metodă cantitativă de a măsura performanța unui sistem. Oferă valoari numerice ordonabile pentru cuantificarea progresului făcut de un model antrenabil.

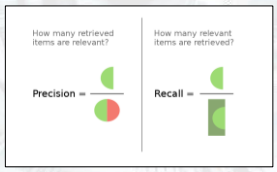
- Se calculează la finalul unei epoci, pe întreaga bază de date (train, val, test).
- Depinde de sarcina de lucru – diferite metrici pentru diferite tipuri de sisteme.
- De obicei, nu este diferențiabilă, deci nu poate fi utilizată pentru a conduce procesul de învățare.
- În unele cazuri, poate fi sinonimă cu funcția de cost (e.g. MAE, MSE).
- Trebuie interpretată în context. De obicei, sunt menționate valorile/intervalele de valori ce reprezintă situația dorită.

20

Sursă imagine: Wikipedia

Clasificarea imaginilor – metrici





*Doar pentru clasificare binară!

21

Clasificarea imaginilor – metrici

Matrice de confuzie:

Real \ Prezis	Pozitiv	Negativ
Pozitiv	True positive (tp)	False negative (fn)
Negativ	False positive (fp)	True negative (tn)

$$precizie(P) = \frac{tp}{tp + fp}$$

$$amintire(R) = \frac{tp}{tp + fn} = rata\ TP\ (TPR)$$

$$rata\ TN\ (TNR) = \frac{tn}{tn + fp}$$

$$rata\ FP\ (FPR) = \frac{fp}{tn + fp}$$

$$acuratețea\ (ACC) = \frac{tp + tn}{tp + tn + fp + fn}$$

$$F - score\ (F) = 2\frac{PR}{P + R}$$

*Doar pentru clasificare binară!

22

Sursă imagine: Wikipedia

Clasificarea imaginilor – metrici

ROC = Receiver Operating Characteristic
AUC = Area Under the (ROC) Curve
AUC=1 => clasificator perfect
AUC=0.5 => clasificator complet aleator
AUC=0 => clasificator anti-perfect

23

Clasificarea imaginilor – metrici

Acuratețea pentru clase dezechilibrate (imbalanced dataset)

- Pentru un clasificator binar, acuratețea este definită astfel:

$$ACC = \frac{tp + tn}{tp + tn + fp + fn}$$

- Scenariu:
  - Considerăm o bază de date cu 95 exemple negative și 5 exemple pozitive.
  - Un clasificator care prezice clasa negativă în 100% din cazuri obține o acurateț de 95%, ceea ce este înșelător.
  - Soluție: folosim acuratețea echilibrată (balanced accuracy).

24

Clasificarea imaginilor – metrici

Acuratețea pentru clase dezechilibrate (imbalanced dataset)

- Acuratețea echilibrată este definită astfel:

$$ACC = \frac{\text{TPR} + \text{TNR}}{2} = \frac{\frac{tp}{tp + fn} + \frac{tn}{tn + fp}}{2}$$

- Scenariu:
  - Considerăm o bază de date cu 95 exemple negative și 5 exemple pozitive.
  - Un clasificator care prezice clasa negativă în 100% din cazuri

		Prezis: Pozitiv	Prezis: Negativ
Real	Pozitiv	tp = 0	fn = 5
	Negativ	fp = 0	tn = 95

$$ACC = \frac{\text{TPR} + \text{TNR}}{2} = \frac{\frac{0}{0+5} + \frac{95}{95+0}}{2} = 0.5$$

25

Clasificarea imaginilor – metrici

Acuratețea pentru clasificarea multi-clasă

$$ACC = \frac{clasificări\ corecte}{clasificări\ totale}$$

- E.g.: din 100 de exemple analizate, 83 au fost clasificate corect => 83% acuratețe.

Acuratețea top-k

- Pentru un exemplu din baza de date se calculează probabilitatea relativă de apartenență la fiecare clasă.
- Se ordonează probabilitățile de ieșire.
- Dacă n între primele k cel mai bine cotate clase se numără și clasa reală => clasificare corectă.

26

Clasificarea imaginilor – metrici

Clasificator

35% cat	40% dog	top-1: miss
3% table	**35% cat**	top-2: hit
1% velociraptor	12% airplane	top-3: hit
12% airplane	6% truck	top-5: hit
6% truck	3% table	
40% dog	2% baseball	
2% baseball	1% velociraptor	
1% mouse	1% mouse	

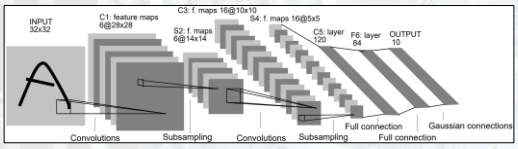
27

Clasificarea imaginilor – modele

- Modelele de rețele neuronale reprezintă partea centrală a sistemelor de clasificare a imaginilor.
- Tradițional, s-au concentrat pe rețele convoluționale (complet convoluționale sau conv + fully-connected).
- Au reprezentat punctul de atracție al domeniului de deep learning.
- Au fost preluate și în alte domenii (audio, text, meta).
- Au o gamă largă de aplicații, nu doar clasificare.

28

Clasificarea imaginilor – LeNet [5]

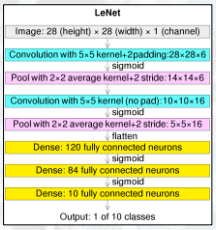


29

Sursă imagine: Wikipedia

Clasificarea imaginilor – LeNet

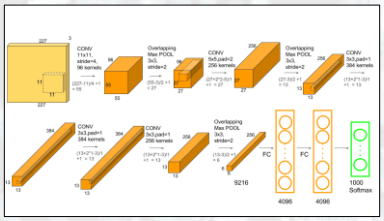
- Utilizată pentru a detecta automat codurile poștale scrise de mână de pe plicuri poștale;
- Prima rețea în care s-a folosit propagarea înapoi;
- A fost introdusă în același timp cu baza de date MNIST (Digits);
- Fiecare strat convoluțional este format din convoluție, activare și pooling;
- Este utilizat mean/average pooling pentru subeșantionare;
- C5 este un strat convoluțional cu nucleul de aceeași dimensiune cu trăsăturile de intrare, echivalent cu un strat fully connected.
- Pe ultimul nivel este folosit un strat fully connected pentru clasificare.
- Varianta LeNet-5 a obținut o acuratețe de 99.2%.



30

Sursă imagine: Neurohive

Clasificarea imaginilor – AlexNet [6]



31

Sursă imagine: Wikipedia

Clasificarea imaginilor – AlexNet

- Utilizată pentru clasificarea imaginilor naturale în cadrul competiției ILSVRC 2012 (câștigător);
- O variantă îmbunătățită a LeNet-5;
- Este utilizat max pooling pentru subeșantionare;
- Utilizează funcția de activare ReLU, în defavoarea sigmoid și tanh;
- A obținut o eroare top-5 de 15.3% (locul 2 – 26.1%);
- Utilizează dropout ca mod de regularizare;
- A demonstrat superioritatea mai multor elemente cheie: adâncimea rețelelor neuronale, funcția de activare ReLU, antrenarea distribuită pe GPU. A adus deep learning în atenția cercetătorilor.

32

Sfârșit M3

33

Bibliografie

[1] http://yann.lecun.com/exdb/mnist/

[2] http://ufldl.stanford.edu/housenumbers/

[3] https://www.cs.toronto.edu/~kriz/cifar.html

[4] http://places2.csail.mit.edu/

[5] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[7] Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

[8] Lin, M., Chen, Q., & Yan, S. (2014). Network in network. 2nd International Conference on Learning Representations (ICLR 2014).

[9] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR 2015), 1–14.

[10] Szegedy, C. et al. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

34

Bibliografie

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[12] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[13] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. Remote Sensing, 13(22), 4712.

[14] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

[15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[16] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015) Explaining and harnessing adversarial examples. International Conference on Learning Representations (poster).

[17] Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 427-436).

35