

Computer Vision 3

Ș.I. 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)

03.12.2022

Computer Vision 3, Ș.I. Mihai DOGARIU

2

2

M3. Învățare Adâncă Supervizată (Supervised Deep Learning)

- 3.1. Concept Supervised Deep Learning
- 3.2. Clasificarea imaginilor
- 3.2. Detecția obiectelor

03.12.2022

Computer Vision 3, Ș.I. Mihai DOGARIU

3

3

M3.1. Concept Supervised Deep Learning

03.12.2022

Computer Vision 3, Ș.I. Mihai DOGARIU

4

4

Supervised Deep Learning

Învățarea mașinilor (machine learning) = spunem despre un sistem că „învăță” 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.

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

input output sarcină dimensiunea

03.12.2022

Computer Vision 3, Ș.I. Mihai DOGARIU

5

5

Supervised Deep Learning

$$D = \{((x_i, y_i) | T), 1 \leq i \leq M\} = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_M, y_M)\}$$
$$\left. \begin{array}{l} f(x_1) = y_1 \\ f(x_2) = y_2 \\ f(x_3) = y_3 \\ \dots \\ f(x_M) = y_M \end{array} \right\} f = ?$$

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

03.12.2022

Computer Vision 3, Ș.I. Mihai DOGARIU

6

6

Sursă imagine: iStock

Supervised Deep Learning

$(a+b)^2 = a^2 + 2ab + b^2$

VS

$f(x_1) = y_1$

$(a+b)^2 = a^2 + b^2$

$f(x_1) = y_1$

03.12.2022 Computer Vision 3, s.1. Mihai DOGARU 7

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 corespundătoare.

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

03.12.2022 Computer Vision 3, s.1. Mihai DOGARU 8

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.

03.12.2022 Computer Vision 3, s.1. Mihai DOGARU 9

9

Supervised Deep Learning

03.12.2022 Computer Vision 3, s.1. Mihai DOGARU 10

10

M3.2. Clasificarea imaginilor

03.12.2022 Computer Vision 3, s.1. Mihai DOGARU 11

11

Clasificarea imaginilor

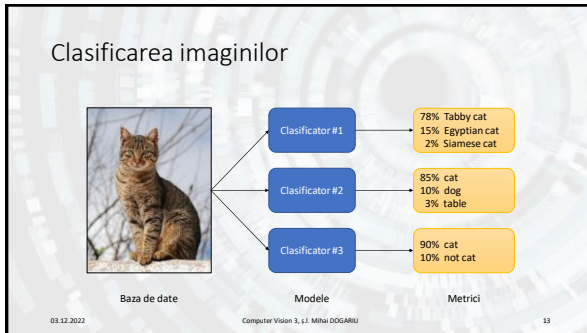
Clasificarea imaginilor = sarcina de a atribui o etichetă/clasă unei imagini.

Caracteristici:

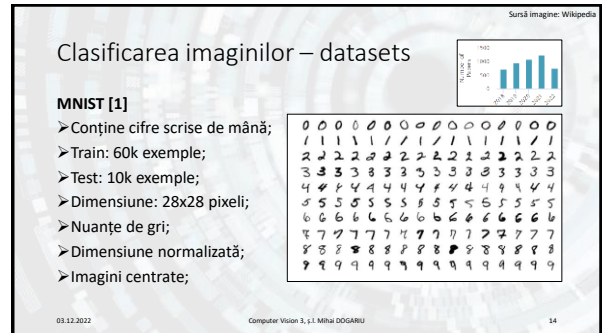
- Conținutul imaginii este tratat ca un întreg ⇔ eticheta descrie întreaga imagine, nu doar o porțiune din ea.
- Orice imagine aparține unei singure clase.
- Ieșirea unui astfel de clasificator este, de obicei, o probabilitate, nu o decizie categorică.
- De obicei, sunt clasificate doar imagini în care 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.

03.12.2022 Computer Vision 3, s.1. Mihai DOGARU 12

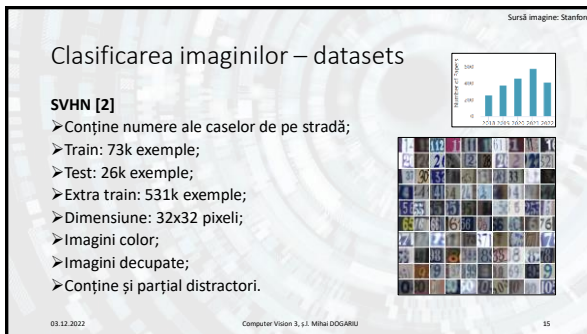
12



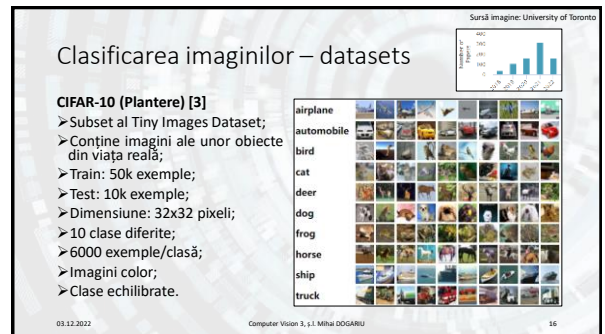
13



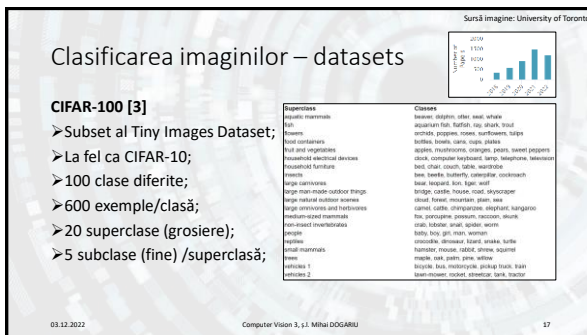
14



15



16



17




18

Clasificarea imaginilor – datasets

Sursă imagine: Stanford

ImageNet [5]

- Contine 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.



03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 19

19

Clasificarea imaginilor – metrici

Metrică – metodă cantitativă de a măsura performanța unui sistem. Oferă valori 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.

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 20

20

Clasificarea imaginilor – metrici

Sursă imagine: Wikipedia



Precision = $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$

Recall = $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$

**Doar pentru clasificare binară!*

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 21

21

Clasificarea imaginilor – metrici

Matrice de confuzie:

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

Formule:

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

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

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

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

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

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


**Doar pentru clasificare binară!*

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 22

22

Clasificarea imaginilor – metrici

Sursă imagine: Wikipedia



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

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 23

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 acuratețe de 95%, ceea ce este înșelător.
 - Soluție: folosim acuratețea echilibrată (balanced accuracy).

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 24

24

Clasificarea imaginilor – metrici

Acuratețea pentru clase dezechilibrate (imbalanced dataset)

➤ Acuratețea echilibrată este definită astfel:

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

➤ 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	Negativ
Real	Pozitiv	tp = 0	fn = 5
	Negativ	fp = 0	tn = 95

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

03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

25

25

Clasificarea imaginilor – metrici

Acuratețea pentru clasificarea multi-clasă

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

➤ 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.

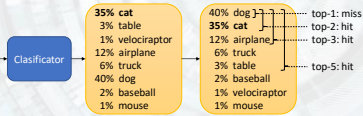
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

26

26

Clasificarea imaginilor – metrici



03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

27

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.

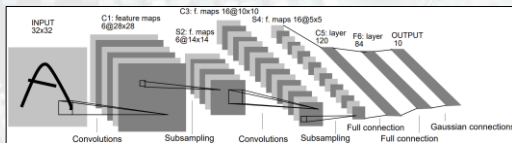
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

28

28

Clasificarea imaginilor – LeNet [5]



03.12.2022

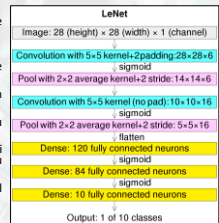
Computer Vision 3, I.I. Mihai DOGARU

29

29

Clasificarea imaginilor – LeNet

- Utilizată pentru a detecta automat codurile poștale scrise de mână de pe plăcuțe 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%.



03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

30

30

Clasificarea imaginilor – AlexNet [6]

03.12.2022 Computer Vision 3, şİ. Mihai DOGARU 31

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: adăncimea rețelelor neuronale, funcția de activare ReLU, antrenarea distribuită pe GPU. A dus la deep learning în atenția cercetătorilor.

The diagram illustrates the AlexNet architecture. It starts with an input image of size 224x224x3. This is processed by five convolutional layers (Conv1 to Conv5) and three fully connected layers (FC1 to FC3). The final output is a classification result showing the top-5 classes and their probabilities.

AlexNet	
Image:	224x224x3 → 227x227x3 (3 channels)
Conv1:	Convolution with 11x11 kernel, stride 4, 54x54x6
Conv2:	Convolution with 3x3 max kernel, stride 2, 28x28x6
Conv3:	Convolution with 5x5 kernel, pad 2, 28x28x6
Conv4:	Convolution with 3x3 max kernel, stride 2, 12x12x6
Conv5:	Convolution with 3x3 kernel, pad 1, 12x12x6
FC1:	Convolution with 3x3 kernel, pad 1, 12x12x6
FC2:	Convolution with 3x3 kernel, pad 1, 12x12x6
FC3:	Convolution with 3x3 kernel, pad 1, 12x12x6
Pool:	Pool with 3x3 max kernel, stride 4, 5x5x6
Results:	Dog (0.496) Rat (0.353) Sheep (0.006) Horse (0.003) Lion (0.001) Elephant (0.001)
Output:	1 out of 1000 classes

6/12/2022 Computer Vision 3, p.1 Mihai Dăscălușanu 32

03.12.2022 Computer Vision 3, ș.I. Mihai DOGARU 32

Clasificarea imaginilor – ZFNet [7]

03.12.2022 Computer Vision 3, ş.I. Mihai DOGARIU 33

Clasificarea imaginilor – ZFNet

- Câștigător al ILSVRC 2013;
- Arhitectură asemănătoare cu AlexNet – convoluții atât cu filtre, cât și pași de dimensiuni mai mici;
- Introduce vizualizarea componentelor învățate de hărțile de trăsături intermediare cu ajutorul operației de deconvoluție – rețelei de feed-forward i se atașează o rețea complementară, ce execută inversul operațiilor din rețea, în paralel. Reușesc să obțină o aproximare a intrării care a produs o activare anume.

6/3/2022 Computer Vision 3, s.1 Mihai OGDARU 34

03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 34

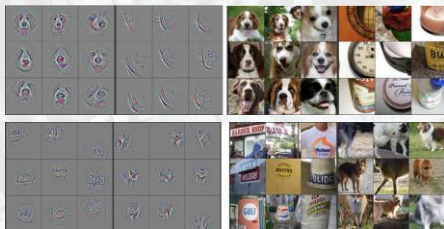
Clasificarea imaginilor – ZFNet

03.12.2022 Computer Vision 3, ş1. Mihai DOGARU 35

Clasificarea imaginilor – ZFNet

03.12.2022 Computer Vision 3, s.l. Mihai DOGARU 36

Clasificarea imaginilor – ZFNet



03.12.2022

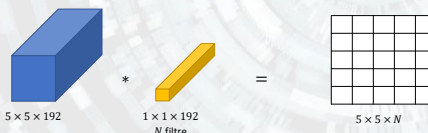
Computer Vision 3, I.I. Mihai DOGARU

37

37

Clasificarea imaginilor – NIN [8]

- Network in Network (NIN) introduce ideea de a folosi convoluții 1×1 pentru a micșora dimensionalitatea datelor.
- Convoluția 1×1 este echivalentă cu un strat fully connected.



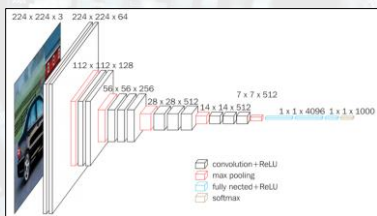
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

38

38

Clasificarea imaginilor – VGG [9]



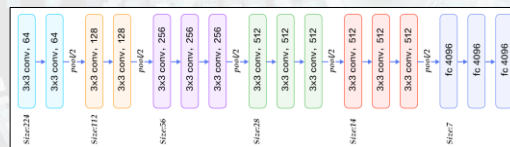
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

39

39

Clasificarea imaginilor – VGG



03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

40

40

Clasificarea imaginilor – VGG

- Model similar cu AlexNet;
- Folosește convoluții cu nucleu de dimensiune mică, asemănător ZFNet;
- Adaugă mai multă adâncime rețelei, demonstrând că rețelele adânci reușesc să capteze mai multă informație;
- Introduc o modularizare a rețelei prin repetarea unei configurații de convoluții de mai multe ori, într-un singur modul.
- 2 variante populare: VGG-16, VGG-19.
- Procesarea intrărilor multi-scală: imaginea este inițial scalată la o valoare între 256 și 512, după care se decupează ferestre de 224×224 pixeli care sunt folosite împreună la antrenare, obținând un fel de regularizare.

03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

41

41

Clasificarea imaginilor – GoogLeNet [10]



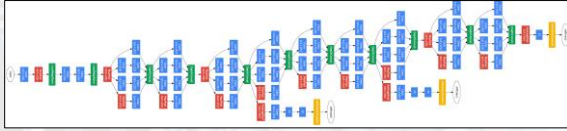
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

42

42

Clasificarea imaginilor – GoogLeNet [9]



03.12.2022

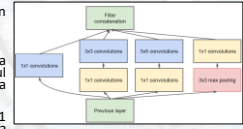
Computer Vision 3, p.1. Mihai DOGARU

43

43

Clasificarea imaginilor – GoogLeNet

- Câștigător al ILSVRC 2014;
- Cunoscută și sub denumirea de Inception (v1, v2, v3, v4);
- Se bazează pe modulele Inception;
- Clasificatorii auxiliari sunt utilizați pentru a combate vanishing gradient în timpul antrenării. În timpul testării se renunță la ei.
- Utilizează extensiv convoluțiile 1×1 pentru a reduce dimensionalitatea și a economisi resurse de calcul;
- Reușește să depășească AlexNet în performanțe, iar numărul de parametri este redus de 12 ori.



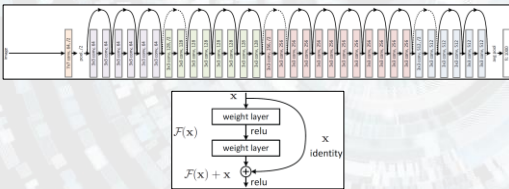
03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

44

44

Clasificarea imaginilor – ResNet [11]



03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

45

45

Clasificarea imaginilor – ResNet [11]

- Câștigător al ILSVRC 2015;
- Cel mai popular model la ora actuală (140k citări);
- Utilizează scurtături (skip connections) pentru a sări peste unele straturi;
- Folosește module, asemănător VGG și Inception;
- Rețelele prea adânci nu mai reușesc să propage cu succes gradientii înapoi și ajung să se „degradeze” – soluția: utilizarea skip connections. Acestea ajută și în cazul problemei „vanishing gradients”.
- Are multe variante asociate: ResNe[x]t-34/50/101/152.

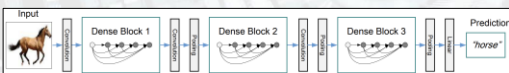
03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

46

46

Clasificarea imaginilor – DenseNet [12]



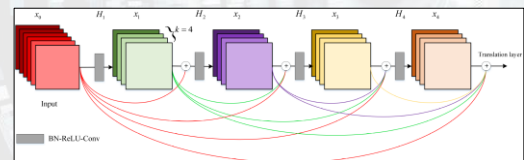
03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

47

47

Clasificarea imaginilor – DenseNet



Structura unui bloc conectat cu un factor de creștere $k=4$ [13]

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

48

48

Clasificarea imaginilor – DenseNet

- Continuă ideea blocurilor modulare, asemănător ResNet;
- Introduc conexiunile dense: fiecare strat dintr-un bloc este conectat la toate straturile care îi urmează din blocul respectiv;
- Blocurile dense sunt utilizate pentru a rezolva problema „vanishing gradient”: fiecare strat are acces direct la gradientii din straturile care îi urmează => nu mai există pericolul ca aceștia să se diminueze excesiv până când sunt propagați către straturile incipiente;
- Adaptarea numărului de canale se realizează cu convoluții 1 x 1;
- Micșorarea hărților de trăsături se realizează cu straturi de pooling.

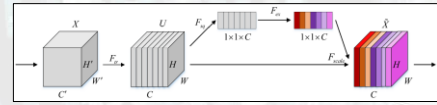
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

49

49

Clasificarea imaginilor – SENet [14]



$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j)$$

$$s = F_{ex}(z, W) = \sigma(g(z, W))$$

$$F_{scale} = (u_c, s_c) = s_c \cdot u_c$$

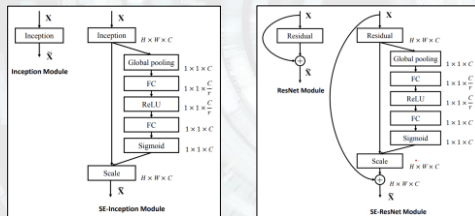
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

50

50

Clasificarea imaginilor – SENet



03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

51

51

Clasificarea imaginilor – SENet

- Câștigător al ILSVRC 2017 (ultima ediție);
- Introduc un bloc ce îmbunătățește interdependențele între canale fără aproape niciun cost adăugat;
- Modulele Squeeze-and-Excitation pot fi adăugate la orice arhitectură existentă;
- Scalează în mod diferit ponderile canalelor cu ajutorul mecanismului de „atenție”;
- Printre primele implementări de succes ale „atenției” în domeniul de computer vision.

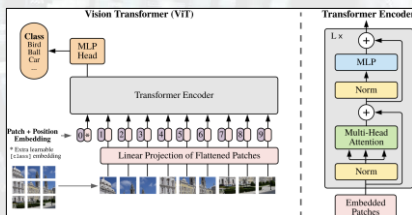
03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

52

52

Clasificarea imaginilor – ViT [15]



03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

53

53

Clasificarea imaginilor – ViT

- Vision Transformer reprezintă adaptarea arhitecturii recurente de tip Transformer la sarcini de computer vision;
- ViT funcționează mai bine decât clasicele rețele convoluționale doar pentru baze de date foarte mari (>100M exemple);
- ViT este mai rapid decât ResNet;
- Majoritatea modelelor de tip Transformer sunt antrenate pe JFT-300M – bază de date internă (closed source) a Google. Acest lucru reprezintă o mare constrângere pentru cercetare.
- Resursele de calcul necesare pentru reproducerea rezultatelor nu sunt disponibile publicului larg.



03.12.2022

Computer Vision 3, I.I. Mihai DOGARU

54

54

Clasificarea imaginilor – modele embedded

➤Motivație:

- Număr mare de dispozitive conectate (IoT);
- Industria bazată pe subansamble cât mai compacte: drone, telefoane mobile, roboți, mașini autonome etc. => platforme mobile cu putere redusă de calcul, baterie limitată;

➤Necesitate:

- Număr mic de parametri;
- Dimensiune redusă a modelului (în MB);
- Timp de inferență cât mai redus;
- Performanțe comparabile cu sisteme full-scale.

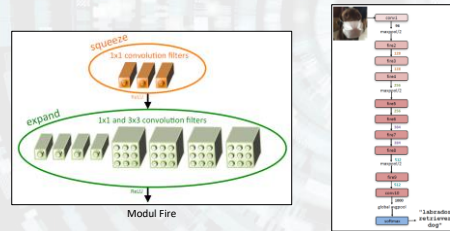
03.12.2022

Computer Vision 3, s.1. Mihai DOGARU

55

55

Clasificarea imaginilor – SqueezeNet [18]



03.12.2022

Computer Vision 3, s.1. Mihai DOGARU

56

56

Clasificarea imaginilor – SqueezeNet

➤Datorită convoluțiilor 1×1 în locul celor de 3×3 , se reduc parametri rețelei;

➤Este introdus un strat de global average pooling în partea de final a rețelei, astfel încât straturile convoluționale să aibă harta de activări cât mai mare (transformă trăsăturile de dimensiuni $N \times N \times c$ în trăsături de dimensiuni $1 \times 1 \times c$);

➤Utilizează DeepCompression pentru a reduce volumul modelului AlexNet cu un factor de 510x (prin cuantizare pe 6 biți în loc de 32 biți), de la 240MB la 0.47MB;

➤A redus numărul de parametri cu un factor de aproape 50;

➤Păstrează performanțele modelului original.

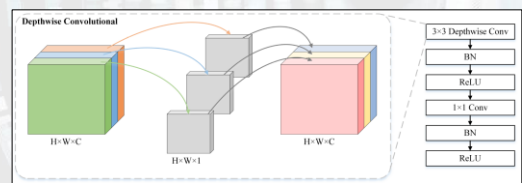
03.12.2022

Computer Vision 3, s.1. Mihai DOGARU

57

57

Clasificarea imaginilor – MobileNet [20]



03.12.2022

Computer Vision 3, s.1. Mihai DOGARU

58

58

Clasificarea imaginilor – MobileNet

➤Folosește convoluții separabile pe adâncime formate din convoluții pe adâncime și convoluții punctuale (1×1);

➤Folosește doi hiperparametri, multiplicator de lățime, respectiv multiplicator de rezoluție pentru a controla compromisul dintre timpul de procesare și acuratețe.

Model	ImageNet Accuracy	Million Multi-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138



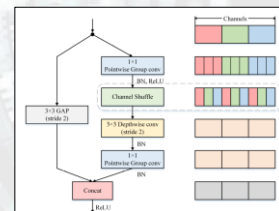
03.12.2022

Computer Vision 3, s.1. Mihai DOGARU

59

59

Clasificarea imaginilor – ShuffleNet [21]



03.12.2022

Computer Vision 3, s.1. Mihai DOGARU

60

60

Clasificarea imaginilor – ShuffleNet

- Folosește convoluțiile punctuale (1×1) de grup pentru a combate problema convoluțiilor punctuale normale, care reduc excesiv de mult numărul de canale => limitează complexitatea reprezentării datelor => rezultate slabe.
- Folosește amestecarea canalelor pentru a combate problema convoluțiilor de grup, care blochează schimbul de informații între canalele dintre grupuri diferite și limitează puterea de reprezentare.

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 61

61

Clasificarea imaginilor – aplicații

1. Image captioning

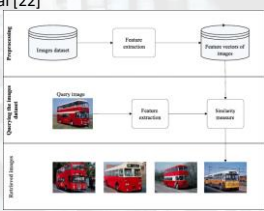


03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 62

62

Clasificarea imaginilor – aplicații

2. Image retrieval [22]



03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 63

63

Clasificarea imaginilor – aplicații

3. Detecția automată a obiectelor



03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 64

64

Clasificarea imaginilor – aplicații

4. Segmentare semantică

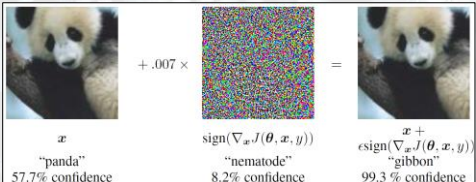


03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 65

65

Clasificarea imaginilor - limitări

➤ Rețelele neuronale sunt sensibile la atacuri adversariale (GoogLeNet):



x	$\text{sign}(\nabla_x J(\theta, x, y))$	$x + \epsilon \text{sign}(\nabla_x J(\theta, x, y))$
"panda"	"nematode"	"gibbon"
57.7% confidence	8.2% confidence	99.3 % confidence

Perturbație liniară insesizabilă pentru oameni asupra unei imagini [16]

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 66

66

Clasificarea imaginilor - limitări

➤ Rețelele neuronale pot fi păcălite cu imagini absurde generate de algoritmi evolutivi:

Imagini recunoscute greșit, cu scoruri >99.6% de către rețele state-of-the-art [17]

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 67

67

Clasificarea imaginilor - limitări

➤ Rețelele neuronale sunt sensibile la translații globale, rotații și scalare.

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 68

68

Clasificarea imaginilor - concluzii

➤ Sistemul de clasificare a imaginilor presupune 2 componente majore:

- Extragerea unor descriptori de trăsături din date;
- Clasificarea descriptorilor de trăsături.

➤ Sistemele de clasificare a imaginilor ocupă un rol central în dezvoltarea rețelelor neuronale cu aplicații în computer vision;

➤ Versatilitate mare – baza de date determină tipul aplicației;

➤ Susceptibile la „atacuri”;

➤ Potențial comercial ridicat.

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 69

69

M3.3. Detecția obiectelor

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 70

70

Detecția obiectelor

Definiție: **Detecția obiectelor** = sarcina de a determina unde anume se găsesc obiecte în imagine și de a determina cărei categorii aparțin.

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 71

71

Detecția obiectelor

➤ Procesul de detecție a obiectelor poate fi împărțit în 3 pași:

1. Găsirea regiunii informative (unde se află obiectul?);
2. Extragerea descriptorului asociat obiectului (cum este descris obiectul?);
3. Clasificarea obiectului (ce fel de obiect este?)

03.12.2022 Computer Vision 3, p.1 Mihai DOGARU 72

72

Detecția obiectelor

Detecția generică a obiectelor – M3.3.

Detecția obiectelor proeminente – M3.4.

03.12.2022 Computer Vision 3, I.I. Mihai DOGARU 73

73

Detecția obiectelor

➤ Detecția generică a obiectelor – caracteristici:

- Trebuie recunoscute toate obiectele (cu care a fost antrenat modelul) dintr-o imagine => baza de date este un punct critic;
- Fiecare obiect detectat este descris de:
 - Casetă de încadrare – 4 coordonate (2 colțuri opuse sau coordonatele unui colț/punct central + lățime și înălțime);
 - Clasa obiectului;
 - Scor $\in [0, 1]$;
- Contează atât scorul, cât și poziționarea casetei.

03.12.2022 Computer Vision 3, I.I. Mihai DOGARU 74

74

Detecția obiectelor

➤ Detecția generică a obiectelor

- Casetă de încadrare:

- 1) x_1, y_1, x_2, y_2 ;
- 2) x, y, w, h ;

- Clasa: „dog”;

- Scorul: 1.0;

- Casetă de încadrare:

- 1) x_1, y_1, x_2, y_2 ;
- 2) x, y, w, h ;

- Clasa: „cat”;

- Scorul: 1.0;

03.12.2022 Computer Vision 3, I.I. Mihai DOGARU 75

75

Detecția obiectelor

$bbox_1 = (x_1, y_1, x_2, y_2)$

$bbox_2 = (x_1, y_1, w, h)$

$bbox_3 = (x_1, y_1, w, h)$

03.12.2022 Computer Vision 3, I.I. Mihai DOGARU 76

76

Detecția obiectelor – datasets

PASCAL VOC 07/12 [23]

- Conține imagini naturale;
- 20 clase adnotate;
- Train: 5.7k imagini;
- Val: 5.8k imagini;
- Test: 10.9k imagini;
- 2.37 obiecte/img;
- Introduce mAP;

03.12.2022 Computer Vision 3, I.I. Mihai DOGARU 77

77

Detecția obiectelor – datasets

Pascal VOC


03.12.2022 Computer Vision 3, I.I. Mihai DOGARU 78

78

Detecția obiectelor – datasets

MS-COCO [24]

- Conține imagini naturale;
- 80 clase;
- 160k imagini adnotate;
- 940k obiecte adnotate;
- 7.27 obiecte/img;
- Cea mai populară;

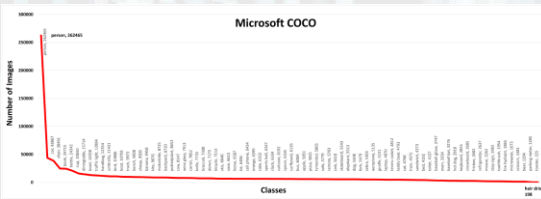


03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 79

79

Detecția obiectelor – datasets

Microsoft COCO



03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 80

80

Detecția obiectelor – datasets

Open Images [25]

- Conține imagini naturale;
- 600 clase;
- 1.9M imagini adnotate;
- 15M obiecte adnotate;
- 8.38 obiecte/img;
- Cea mai mare.

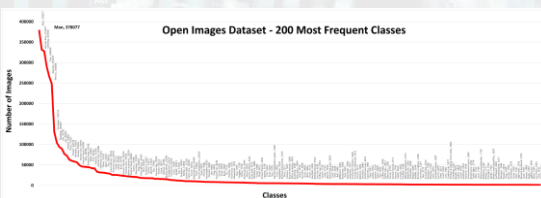


03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 81

81

Detecția obiectelor – datasets

Open Images Dataset - 200 Most Frequent Classes

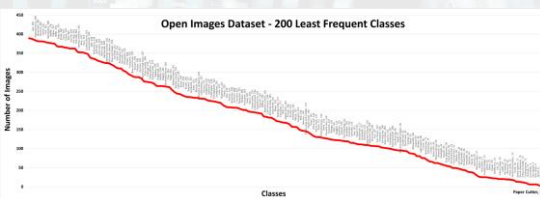


03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 82

82

Detecția obiectelor – datasets

Open Images Dataset - 200 Least Frequent Classes

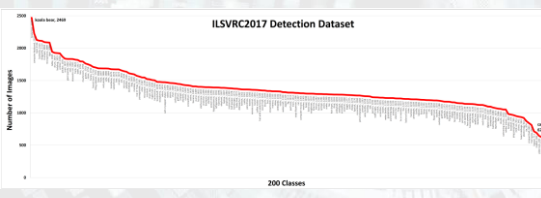


03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 83

83

Detecția obiectelor – datasets

ILSVRC2017 Detection Dataset



03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 84

84

Detecția obiectelor – datasets

Dataset	Classes	Train			Validation			Test
		Images	Objects	Objects/Image	Images	Objects	Objects/Image	
PASCAL VOC 12	20	5,117	11,669	2.38	5,833	11,841	2.37	10,981
MS-COCO	80	118,287	860,001	7.27	5,000	36,781	7.35	40,670
ILSVRC	200	496,567	478,807	1.05	20,121	55,501	2.76	40,152
OpenImage	600	1,743,042	14,610,229	8.38	41,620	204,621	4.92	125,436

03.12.2022

Computer Vision 3, 1.1. Mihai DOGARU

85

85

Detecția obiectelor – mAP

- Metrica folosită pentru a evalua sistemele de detecție a obiectelor se numește mean Average Precision (mAP).
- În componența ei, se iau în calcul mai multe metrici:
 - Intersecția supra reuniunea;
 - Precizia;
 - Reamintirea;
 - Precizia medie.

03.12.2022

Computer Vision 3, 1.1. Mihai DOGARU

86

86

Detecția obiectelor – mAP

- Intersecția supra reuniunea (Intersection over Union – IoU):

$$IoU = \frac{\text{Aria intersecției}}{\text{Aria reuniunii}} = \frac{\text{Aria intersecției}}{\text{Aria reuniunii}}$$


03.12.2022

Computer Vision 3, 1.1. Mihai DOGARU

87

87

Detecția obiectelor – mAP

- Se stabilește un prag (e.g. 0.5) în funcție de care se determină precizia și reamintirea pentru o clasă anume.

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

$$\text{reamintire}(R) = \frac{tp}{tp + fn}$$

- Cum definim tp , fp , fn ?

03.12.2022

Computer Vision 3, 1.1. Mihai DOGARU

88

88

Detecția obiectelor – mAP

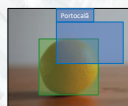
Groundtruth



Detecție clasă corectă
 $IoU=0.9 > \text{prag} \Rightarrow \text{true positive}$



Lipsă detecție /
Detecție clasă greșită
 $IoU = 0 \Rightarrow \text{false negative}$



Detecție clasă corectă
 $IoU=0.3 < \text{prag} \Rightarrow \text{false positive}$

03.12.2022

Computer Vision 3, 1.1. Mihai DOGARU

89

89

Detecția obiectelor – mAP

1. Se selectează o imagine;



03.12.2022

Computer Vision 3, 1.1. Mihai DOGARU

90

90

Sursă imagine: learpencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;




03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 91

91

Sursă imagine: learpencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;



03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 92

92

Sursă imagine: learpencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);

Detections							
Conf.	0.63	0.77	0.92	0.86	0.88	0.58	0.91
Matches GT by IOU?	TP	TP	TP	FP	TP	TP	FP

03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 93

93

Sursă imagine: learpencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);
5. Se sortează tabelul în ordinea scorului (confidence);

Detections							
Conf.	0.92	0.91	0.88	0.86	0.77	0.63	0.58
Matches GT by IOU?	TP	FP	TP	FP	TP	TP	TP

03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 94

94

Sursă imagine: learpencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);
5. Se sortează tabelul în ordinea descrescătoare a scorului (confidence);
6. Se calculează precizia și reamintirea pentru un prag IOU;

Prag	Conf.	Matches	Calcularea TP	Calcularea FP	Precizia	Recall
0.9	TP	1	0	$\frac{1}{1+0} = 1$	$\frac{1}{1} = 0.98$	
0.8	FP	1	1	$\frac{1}{1+1} = 0.5$	$\frac{1}{1} = 0.98$	
0.8	TP	2	1	$\frac{2}{2+1} = 0.66$	$\frac{2}{2} = 0.98$	
0.8	FP	2	2	0.5	0.16	
0.77	TP	3	2	0.6	0.25	
0.63	TP	4	2	0.66	0.33	
0.58	TP	5	2	0.71	0.41	

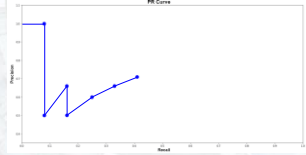
03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 95

95

Sursă imagine: learpencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);
5. Se sortează tabelul în ordinea descrescătoare a scorului (confidence);
6. Se calculează precizia și reamintirea pentru un prag IOU;
7. Se trasează graficul PR;



03.12.2022 Computer Vision 3, p.1. Mihai DOGARU 96

96

Sursă imagine: learnopencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);
5. Se sortează tabelul în ordinea descrescătoare a scorului (confidence);
6. Se calculează precizia și reamintirea pentru un prag IoU;
7. Se trasează graficul PR;
8. Se interpolează graficul PR în 11 (101) puncte;

Se selectează precizia maximă corespunzătoare valorilor de reamintire mai mari decât reamintirea curentă

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 97

97

Sursă imagine: learnopencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);
5. Se sortează tabelul în ordinea descrescătoare a scorului (confidence);
6. Se calculează precizia și reamintirea pentru un prag IoU;
7. Se trasează graficul PR;
8. Se interpolează graficul PR în 11 (101) puncte;
9. Se calculează precizia medie;

$$AP_{dog} = \frac{1}{11} (P(0) + P(0.1) + P(0.2) + \dots + P(1.0))$$

$$= \frac{1}{11} (1 + 4 \times 0.71 + 6 \times 0)$$

$$= 34.9\%$$

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 98

98

Sursă imagine: learnopencv

Detecția obiectelor – mAP

1. Se selectează o imagine;
2. Se rețin etichetele imaginii;
3. Se rulează detecția obiectelor pe imagine;
4. Se extrag detecțiile pentru clasa de interes (dog);
5. Se sortează tabelul în ordinea descrescătoare a scorului (confidence);
6. Se calculează precizia și reamintirea pentru un prag IoU;
7. Se trasează graficul PR;
8. Se interpolează graficul PR în 11 (101) puncte;
9. Se calculează precizia medie;
10. Se calculează media precizilor medii pe toate clasele.

CLASS	dog	person	sheep	truck	teddy
AP	0.349	0.545	0.00	1.00	0.50

$$mAP = \frac{1}{5} (AP_{dog} + AP_{person} + AP_{sheep} + AP_{truck} + AP_{teddy})$$

$$= \frac{1}{5} (0.349 + 0.545 + 0.00 + 1.00 + 0.50)$$

$$= 87.89\%$$

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 99

99

Detecția obiectelor – modele

➤ Fiecare model de detecție a obiectelor are nevoie de o arhitectură specializată în extragerea unui set de trăsături cât mai distinctiv – elementul în jurul căruia se construiește modelul (backbone). Arhitecturi “backbone” populare:

- AlexNet [6];
- VGG [9];
- GoLeNet/Inception [10];
- ResNet [11];
- ResNeXt [26];
- CSPNet [27];
- EfficientNet [28];

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 100

100

Detecția obiectelor – modele

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 101

101

Detecția obiectelor – modele în 2 etape

➤ Modelele în 2 etape presupun parcurgerea următoarelor...2 etape:

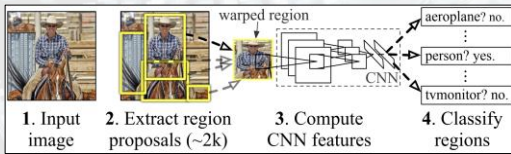
1. Propunerea și extragerea unor regiuni de interes (potențiale obiecte);
2. Clasificarea regiunilor propuse + regresia casetelor de încadrare.

03.12.2022 Computer Vision 3, 1.1 Mihai DOGARU 102

102

Detecția obiectelor – modele în 2 etape

Regions with CNN features (R-CNN) [29]



03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

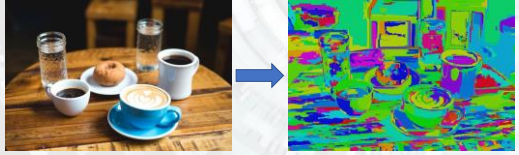
103

103

Detecția obiectelor – modele în 2 etape

R-CNN – extragerea propunerilor de regiuni:

1. Se realizează o suprasegmentare a fiecărei imagini de intrare (algoritmul Felzenszwalb & Huttenlocher [30])



03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

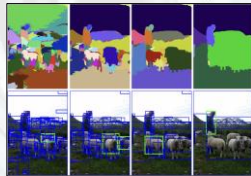
104

104

Detecția obiectelor – modele în 2 etape

R-CNN – extragerea propunerilor de regiuni:

2. Regiunile segmentate se grupează aglomerativ, ierarhic, în funcție de similitudine (algoritmul Selective Search [31]), până la obținerea numărului dorit de regiuni propuse (2k).



03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

105

105

Detecția obiectelor – modele în 2 etape

R-CNN – extragerea trăsăturilor din regiunile propuse:

- Fiecare regiune este redimensionată (warped) la 227 x 227 pixeli;
- Regiunile sunt propagate prin rețeaua de extragere de trăsături – backbone (AlexNet);
- Se obține un vector de trăsături de dimensiune 4096 pentru fiecare regiune.



03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

106

106

Detecția obiectelor – modele în 2 etape

R-CNN – clasificarea regiunilor propuse:

- Fiecare vector de trăsături este procesat de către un clasificator SVM antrenat pentru fiecare clasă, în parte => N clasificatori (N=nr. clase);
- Fiecare clasificator dă o decizie asupra aceleiași regiuni => N scoruri diferite. Se alege clasa corespunzătoare celui mai mare scor. Dacă acest scor este mai mare decât un prag predefinit (e.g. 0.7), atunci se consideră că a fost detectat un obiect din acea clasă. Altfel, se consideră că fiind „background”.

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

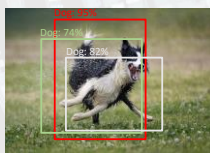
107

107

Detecția obiectelor – modele în 2 etape

R-CNN – clasificarea regiunilor propuse:

- Există situații când un obiect are asociate mai multe casete de încadrare asemănătoare:



03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

108

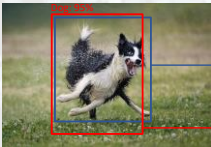
108

- Se utilizează mecanismul Non-Maximum Supression (NMS):
1. Se selectează caseta de încadrare cu cel mai mare scor (95%);
 2. Se calculează IoU dintre caseta de la pct. 1 și toate celelalte casete asociate aceleiași clase;
 3. Dacă IoU dintre caseta de scor maxim și o casetă țintă depășește un prag stabilit (e.g. 0.3), se elimină caseta țintă. Altfel, se trece mai departe.
 4. Se repetă algoritmul până la epuizarea tuturor casetelor de încadrare.

Detecția obiectelor – modele în 2 etape

R-CNN – clasificarea regiunilor propuse:

➤ După stabilirea detecțiilor, se rulează și o ramură de regresie a casetei de încadrare pentru rafinarea încadrării.



Casetă de încadrare obținută după regresie

Casetă prezisă de către algoritmul Selective Search

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

109

109

Detecția obiectelor – modele în 2 etape

R-CNN – antrenare:

1. Pre-train: rețeaua backbone (AlexNet) este pre-antrenată pe ImageNet, folosind adnotări la nivel de imagine (nu obiect);
2. Fine-tune: se înlocuiește ultimul strat al clasificatorului, de dimensiune 1000, cu un clasificator de dimensiune N+1 (N clase + background) și se antrenează rețeaua backbone pentru a o adapta la noua sarcină (obiecte decupate din imagini și redimensionate);
3. Se înlocuiește clasificatorul final cu N clasificatori SVM, câte unul pentru fiecare clasă și se optimizează.

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

110

110

Detecția obiectelor – modele în 2 etape

R-CNN – rezultate:

VOC 2010 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
DPM v5	49.2	53.8	13.1	15.3	35.5	53.4	49.7	27.0	17.2	28.8	14.7	17.8	46.4	51.2	47.7	10.8	34.2	20.7	43.8	38.3	33.4
LVA	56.2	42.4	15.3	12.6	21.8	49.3	36.8	46.1	12.9	32.9	30.0	36.5	43.5	52.9	32.9	15.5	41.1	31.8	47.0	44.8	35.1
Regionlets	65.0	48.9	25.9	24.6	24.5	56.1	54.5	51.2	17.0	28.9	30.2	35.8	40.2	55.7	43.5	14.3	43.9	32.6	54.0	45.9	39.7
SegDPM	61.4	53.4	25.6	25.2	35.5	51.7	50.6	50.8	19.3	33.8	26.8	40.4	48.3	54.4	47.1	14.8	38.7	35.0	52.8	43.1	40.4
R-CNN	67.1	64.1	46.7	32.0	30.5	56.4	57.2	65.9	27.0	47.3	40.9	66.6	57.8	65.9	53.6	26.7	56.5	38.1	52.8	50.2	50.2
R-CNN BB	71.8	65.8	53.8	36.8	35.9	59.7	60.9	69.9	27.9	50.6	41.4	70.8	62.6	69.8	58.1	28.5	59.4	39.3	61.2	52.4	53.7

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

111

111

Detecția obiectelor – modele în 2 etape

R-CNN – concluzii:

- A depășit toate modelele existente de detecție a obiectelor la acea vreme cu o diferență semnificativă;
- Algoritmul este încet (47s/imagine);
- Antrenarea este complexă – proces în 3 pași și durează mult (84 ore);
- Propunerea de regiuni se face conform unui algoritm fix, fără învățare.

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

112

112

Sfârșit M3

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

117

117

Bibliografie

- [1] <http://yann.lecun.com/exdb/mnist/>
- [2] <http://ufldl.stanford.edu/housenumbers/>
- [3] <http://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).

03.12.2022

Computer Vision 3, p.1. Mihai DOGARU

118

118

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., ... & Houtis, 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).
- [18] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.

03.12.2022

Computer Vision 3, c1. Mihai DOGARU

119

119

Bibliografie

- [19] Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *ICLR* 2016.
- [20] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [21] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6848-6856).
- [22] Singh, S., & Batra, S. (2020). An efficient bi-layer content based image retrieval system. *Multimedia Tools and Applications*, 79(25), 17731-17759.
- [23] <http://host.robots.ox.ac.uk/pascal/VOC/>
- [24] <https://cocodataset.org/#home>
- [25] <https://storage.googleapis.com/openimages/web/index.html>
- [26] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).

03.12.2022

Computer Vision 3, c1. Mihai DOGARU

120

120

Bibliografie

- [27] Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 390-393).
- [28] Tan, M., & Le, Q. (2019, May). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [29] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
- [30] Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International journal of computer vision*, 59(2), 167-181.

03.12.2022

Computer Vision 3, c1. Mihai DOGARU

121

121