


# 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)

17.11.2022 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

17.11.2022 Computer Vision 3, Ș.I. Mihai DOGARIU 3

3

## M3.1. Concept Supervised Deep Learning

17.11.2022 Computer Vision 3, Ș.I. Mihai DOGARIU 4

4

## Supervised Deep Learning

Învățarea mașinilor (machine learning) = spunem despre un sistem că „învată” 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      sarcina      dimensiunea

17.11.2022 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.

17.11.2022 Computer Vision 3, Ș.I. Mihai DOGARIU 6

6

Supervised Deep Learning

Sursă imagine: iStock

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

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

vs

$f(x_1) = y_1$

$f(x_1) = y_1$

17.11.2022 Computer Vision 3, p.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).

17.11.2022 Computer Vision 3, p.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.

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 9

9

Supervised Deep Learning

$x_i, \bar{y}_i$

model

$y_i$

$\mathcal{L}(y_i, \bar{y}_i)$

$J(w)$

$\nabla$

$\Sigma$

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 10

10

M3.2. Clasificarea imaginilor

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 11

11

Clasificarea imaginilor

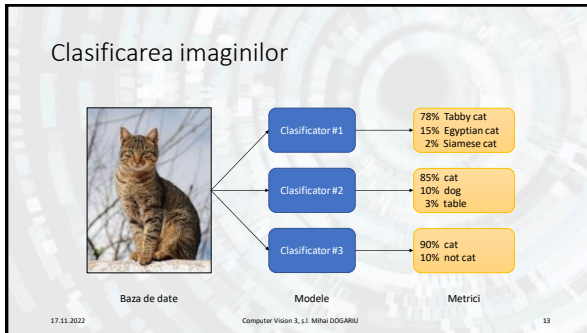
**Clasificarea imaginilor** = sarcina de a atribui o etichetă/clasă unei imagini.

Caracteristici:

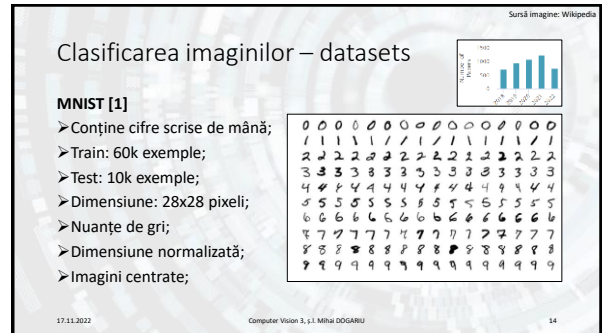
- Conținutul imaginii este tratat ca un întreg ⇔ eticheta descrie întreaga imagine, nu doar o porțiune din ea.
- Orice imagine aparține unei singure clase.
- Ieșirea unui astfel de clasificator este, de obicei, o probabilitate, nu o decizie categorică.
- De obicei, sunt clasificate doar imagini în care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

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 12

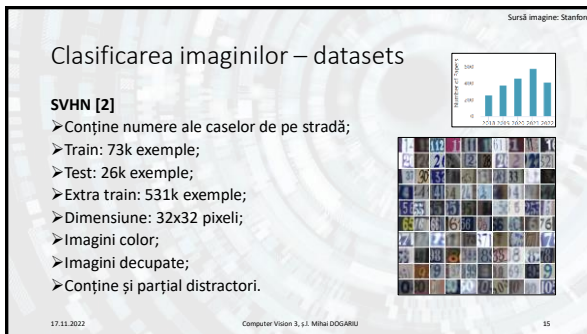
12



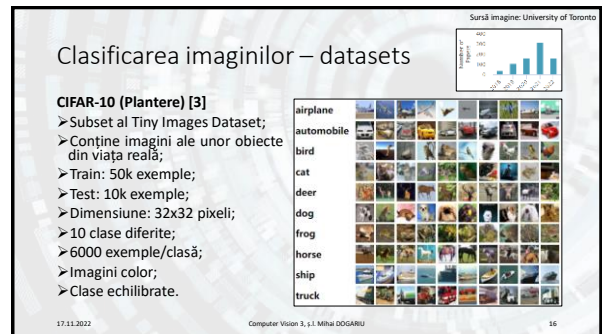
13



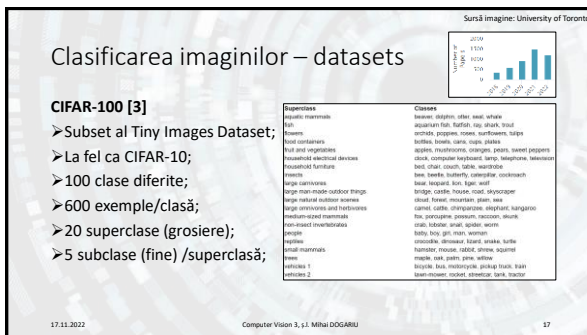
14



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.



17.11.2022 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.

17.11.2022 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!**

17.11.2022 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)

**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)** =  $\frac{PR}{P + R}$


**\*Doar pentru clasificare binară!**

17.11.2022 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

17.11.2022 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).

17.11.2022 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

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

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

25

## Clasificarea imaginilor – metrici

### Acuratețea pentru clasificarea multi-clasă

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

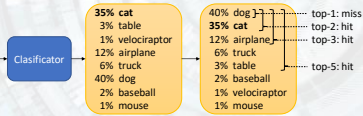
➤ E.g.: din 100 de exemple analizate, 83 au fost clasificate corect => 83% acuratețe.

### 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.

26

## Clasificarea imaginilor – metrici



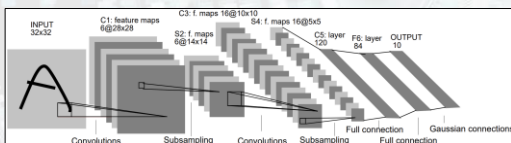
27

## 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

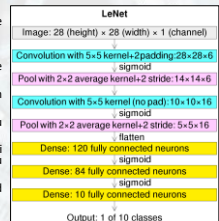
## Clasificarea imaginilor – LeNet [5]



29

## 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%.



30



Clasificarea imaginilor – AlexNet [6]

Sursă imagine: Neurorhive

17.11.2022 Computer Vision 3, s.l. Mihai DOGARU 31

31

Clasificarea imaginilor – AlexNet

Sursă imagine: Wikipedia

- 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 rețelei neuronale, funcția de activare ReLU, antrenarea distribuită pe GPU. A adus deep learning în atenția cercetătorilor.

AlexNet	
Image:	227 (height) × 227 (width) × 3 (channels)
Convolution:	with 11 × 11 kernel, stride 4, 54 × 54 × 64
Pool:	with 3 × 3 max kernel, stride 2, 26 × 26 × 64
Convolution:	with 5 × 5 kernel, stride 1, 12 × 12 × 128
Pool:	with 3 × 3 max kernel, stride 2, 12 × 12 × 128
Convolution:	with 3 × 3 kernel, stride 1, 12 × 12 × 128
Convolution:	with 3 × 3 kernel, stride 1, 12 × 12 × 128
Convolution:	with 3 × 3 kernel, stride 1, 12 × 12 × 128
Pool:	with 3 × 3 max kernel, stride 1, 5 × 5 × 128
Layer:	4096 fully connected neurons
Dropout:	4096 fully connected neurons
Dropout:	4096 fully connected neurons
Dropout:	1000 fully connected neurons
Output:	1 of 1000 classes

17.11.2022 Computer Vision 3, s.l. Mihai DOGARU 32

32

Clasificarea imaginilor – ZFNet [7]

17.11.2022 Computer Vision 3, s.l. Mihai DOGARU 33

33

Clasificarea imaginilor – ZFNet

- Câștigător al ILSVRC 2013;
- Arhitectură asemănătoare cu AlexNet – convoluții atât cu filtre, cât și pași de dimensiuni mai mici;
- Introduc 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.

17.11.2022 Computer Vision 3, s.l. Mihai DOGARU 34

34

Clasificarea imaginilor – ZFNet

17.11.2022 Computer Vision 3, s.l. Mihai DOGARU 35

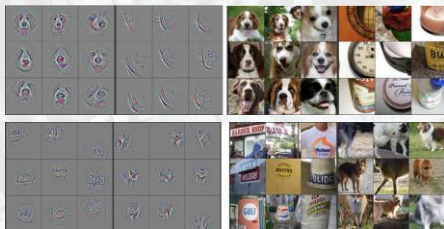
35

Clasificarea imaginilor – ZFNet

17.11.2022 Computer Vision 3, s.l. Mihai DOGARU 36

36

## Clasificarea imaginilor – ZFNet



17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

37

37

## Clasificarea imaginilor – NIN [8]

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



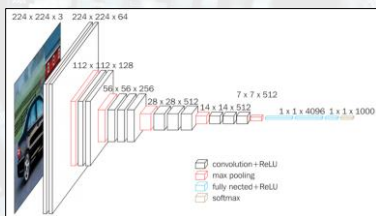
17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

38

38

## Clasificarea imaginilor – VGG [9]



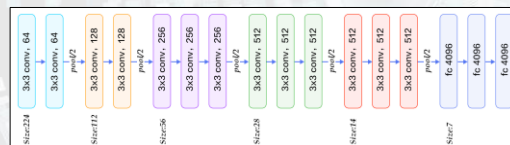
17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

39

39

## Clasificarea imaginilor – VGG



17.11.2022

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  $224 \times 224$  pixeli care sunt folosite împreună la antrenare, obținând un fel de regularizare.

17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

41

41

## Clasificarea imaginilor – GoogLeNet [10]



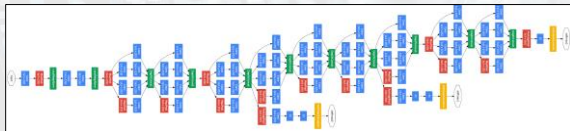
17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

42

42

## Clasificarea imaginilor – GoogLeNet [9]



17.11.2022

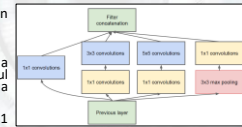
Computer Vision 3, p.1 Mihai DOGARU

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  $1 \times 1$  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.



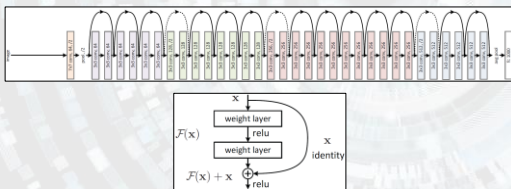
17.11.2022

Computer Vision 3, p.1 Mihai DOGARU

44

44

## Clasificarea imaginilor – ResNet [11]



17.11.2022

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.

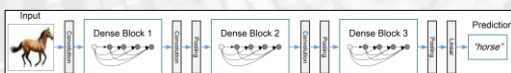
17.11.2022

Computer Vision 3, p.1 Mihai DOGARU

46

46

## Clasificarea imaginilor – DenseNet [12]



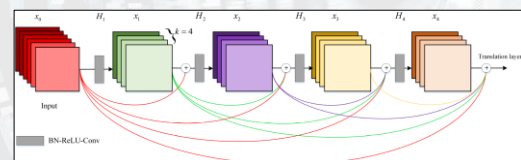
17.11.2022

Computer Vision 3, p.1 Mihai DOGARU

47

47

## Clasificarea imaginilor – DenseNet



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

17.11.2022

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.

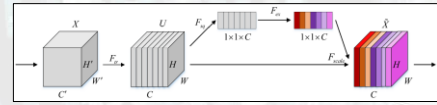
17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

49

49

## Clasificarea imaginilor – SENet [14]



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

$$s = F_{sq}(x, W) = \sigma(g(x, W))$$

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

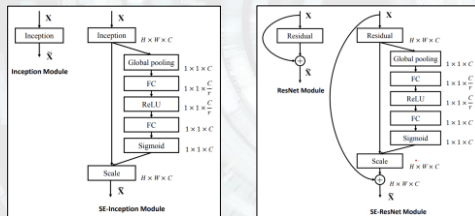
17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

50

50

## Clasificarea imaginilor – SENet



17.11.2022

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.

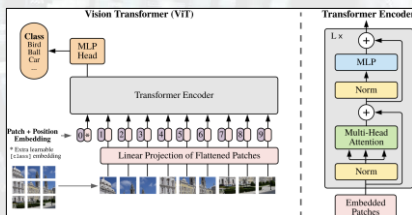
17.11.2022

Computer Vision 3, I.I. Mihai DOGARU

52

52

## Clasificarea imaginilor – ViT [15]



17.11.2022

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.



17.11.2022

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.

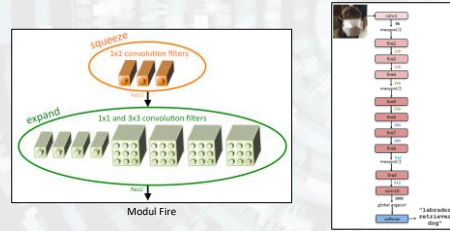
17.11.2022

Computer Vision 3, s.1. Mihai DOGARU

55

55

## Clasificarea imaginilor – SqueezeNet [18]



17.11.2022

Computer Vision 3, s.1. Mihai DOGARU

56

56

## Clasificarea imaginilor – SqueezeNet

- Datorită convoluțiilor  $1 \times 1$  în locul celor de  $3 \times 3$ , 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.

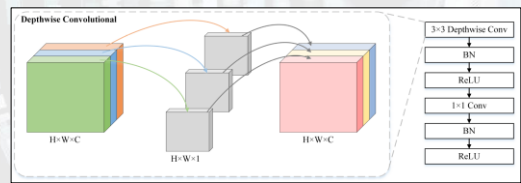
17.11.2022

Computer Vision 3, s.1. Mihai DOGARU

57

57

## Clasificarea imaginilor – MobileNet [20]



17.11.2022

Computer Vision 3, s.1. Mihai DOGARU

58

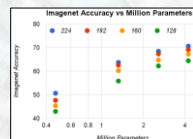
58

## Clasificarea imaginilor – MobileNet

- Folosește convoluții separabile pe adâncime formate din convoluții pe adâncime și convoluții punctuale ( $1 \times 1$ );

- 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



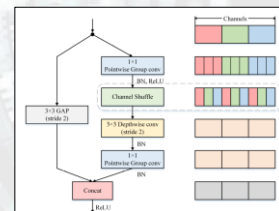
17.11.2022

Computer Vision 3, s.1. Mihai DOGARU

59

59

## Clasificarea imaginilor – ShuffleNet [21]



17.11.2022

Computer Vision 3, s.1. Mihai DOGARU

60

60

## Clasificarea imaginilor – ShuffleNet

- Folosește convoluțiile punctuale ( $1 \times 1$ ) 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.

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 61

61

## Clasificarea imaginilor – aplicații

### 1. Image captioning

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 62

62

## Clasificarea imaginilor – aplicații

### 2. Image retrieval [22]

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 63

63

## Clasificarea imaginilor – aplicații

### 3. Detecția automată a obiectelor

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 64

64

## Clasificarea imaginilor – aplicații

### 4. Segmentare semantică

17.11.2022 Computer Vision 3, p.1 Mihai DOGARU 65

65

## Clasificarea imaginilor - limitări

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

$x$	$+ .007 \times$	$\text{sign}(\nabla_x J(\theta, x, y))$	$=$	$x + \epsilon \text{sign}(\nabla_x J(\theta, x, y))$
"panda"		"nematode"		"gibbon"
57.7% confidence		8.2% confidence		99.3 % confidence

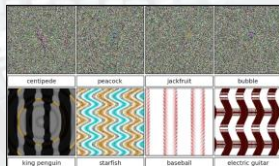
Perturbație liniară insesizabilă pentru oameni asupra unei imagini [16]

17.11.2022 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]

Computer Vision 3, 1.1 Mihai DOGARU

67

67

## Clasificarea imaginilor - limitări

Sursă imagine: OpenGenus IQ

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



Computer Vision 3, 1.1 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.

Computer Vision 3, 1.1 Mihai DOGARU

69

69

Sfârșit M3

Computer Vision 3, 1.1 Mihai DOGARU

70

70

## Bibliografie

- [1] <http://yann.lecun.com/exdb/mnist/>
- [2] <http://dld.stanford.edu/housenumber/>
- [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)*, 1-5.
- [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).

Computer Vision 3, 1.1 Mihai DOGARU

71

71

## 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, F., 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).
- [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*.

Computer Vision 3, 1.1 Mihai DOGARU

72

72

## 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](https://arxiv.org/abs/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.