

0619 ინფორმაციისა და კომუნიკაციის ტექნოლოგიები
INFORMATION AND COMMUNICATION TECHNOLOGIES (ICTS)

ხელოვნური ინტელექტი და მაღალეფექტური კომპიუტერული ხედვა

გიორგი კაკაშვილი

დავით აღმაშენებლის სახელობის საქართველოს ეროვნული თავდაცვის აკადემია

David Aghmashenebeli National Defence Academy of Georgian

E-mail: giorgikakashvili1@gmail.com

რეზიუმე

სტატიაში განხილულია ხელოვნური ინტელექტის (AI) და მაღალეფექტური კომპიუტერული ხედვის კოლაბორაცია. კომპიუტერული ხედვის უპირველესი მიზანია ვიზუალური მონაცემების ინტერპრეტაციით და ანალიზით შეასრულონ დავალებები, რომლებიც საჭიროებენ ადამიანის მსგავს ვიზუალურ შესაძლებლობებს. სტატიაში ნათლადაა ნაჩვენები, რომ კომპიუტერული ხედვის ალგორითმები შეიძლება დაიყოს ტრადიციულ და ღრმა სწავლებაზე დაფუძნებულ მიდგომებად. კომპიუტერული ხედვის ტრადიციული ალგორითმები ეყრდნობა ხელით შექმნილ წესებს. მეორეს მხრივ, ღრმა სწავლებამ მოახდინა რევოლუცია კომპიუტერულ ხედვაში ხელოვნური ნეირონული ქსელების, განსაკუთრებით კონვოლუციური ნეირონული ქსელების (CNN) გამოყენებით, რომლის საშუალებითაც სწავლობენ უშუალოდ მონაცემებიდან. რაც მათ აძლევს შესაძლებლობას ამოიცნონ ობიექტები, დაახარისხონ სურათები და შეასრულონ რთული ამოცანები მაღალი სიზუსტით. კომპიუტერული ხედვის ალგორითმებში მიღწევებმა, განსაკუთრებით ღრმა სწავლებამ CNN-ების საშუალებით, შესამჩნევი პროგრესი გვაჩვენა კომპიუტერული ხედვის სხვადასხვა ამოცანებსა და აპლიკაციებში, გაამდიდრა AI სისტემების შესაძლებლობები და გახსნა ახალი ფანჯარა ინდუსტრიებისთვის.

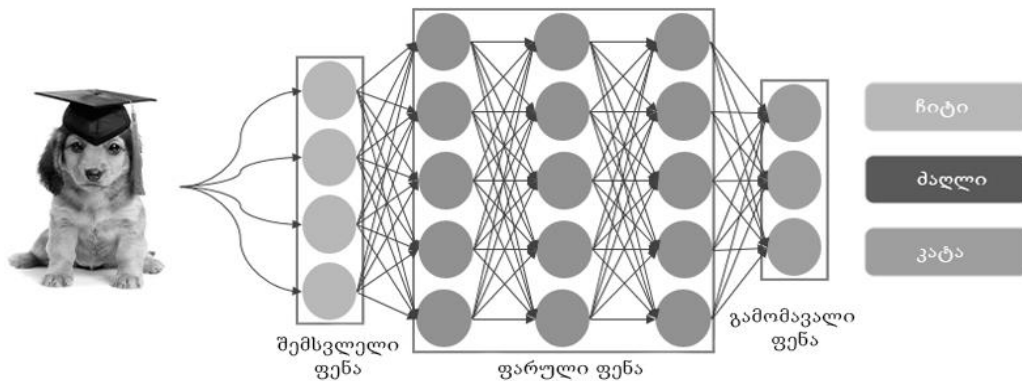
საკვანძო სიტყვები: ხელოვნური ინტელექტი, კომპიუტერული ხედვა, კონვოლუციური ნეირონული ქსელი, ღრმა სწავლება, თანამედროვე ტექნოლოგიები.

ხელოვნურმა ინტელექტმა (AI) და კომპიუტერულმა ხედვამ განიცადა მნიშვნელოვანი წინსვლა ბოლო წლებში, რამაც კომპიუტერული ხედვის სფერო ახალ საფეხურზე გადაიყვანა. ხელოვნური ინტელექტის ალგორითმების სიმძლავრის მაღალი ხარისხის გამოთვლებთან (HPC) გაერთიანების საშუალებით, მკვლევარებმა და ინჟინრებმა გამოავლინეს ინოვაციური შესაძლებლობები ვიზუალურ აღქმასა და ანალიზში. ნაშრომის მიზანია წარმოაჩინოს ხელოვნური ინტელექტისა და მაღალეფექტური ხედვის შერწყმის უპირატესობები, მისი სარგებელი და სამომავლო პერსპექტივები. კომპიუტერული ხედვა არის მულტიდისციპლინარული სფერო, რომელიც ფოკუსირებულია იმაზე, რომ კომპიუტერებმა შეძლონ ვიზუალური მონაცემების მაღალი დონის გაგება, ადამიანის ვიზუალური აღქმის სიმულაცია. ის მოიცავს ისეთ ამოცანებს, როგორცაა გამოსახულების და ვიდეოს ამოცნობა, ობიექტების ამოცნობა და თვალყურის დევნება, გამოსახულების სეგმენტაცია და სცენის რეკონსტრუქცია. ტრადიციულად, კომპიუტერული ხედვის ალგორითმები ეფუძნებოდა ხელნაკეთ წესებსა და ევრისტიკას, რაც ზღუდავდა მათ ეფექტურობას და მასშტაბურობას. თუმცა, ხელოვნური ინტელექტისა და ღრმა სწავლის ტექნიკის გაჩენამ რევოლუცია მოახდინა კომპიუტერულ ხედვაში, რაც საშუალებას აძლევს მანქანებს ისწავლონ უსაზღვრო რაოდენობის მონაცემებიდან.

კომპიუტერული ხედვის უპირველესი მიზანია მანქანებს საშუალება მისცეს ვიზუალური მონაცემების ინტერპრეტაცია და ანალიზი, რაც მათ შესაძლებლობას აძლევს შეასრულონ დავალებები, რომლებიც, როგორც წესი, მოითხოვს ადამიანის ვიზუალურ შესაძლებლობებს. ეს მოიცავს ამოცანებს, როგორცაა გამოსახულების ამოცნობა, ობიექტების ამოცნობა, გამოსახულების სეგმენტაცია, სახის ამოცნობა, სცენის გაგება, მოძრაობის ანალიზი და 3D რეკონსტრუქცია. კომპიუტერული ხედვის ალგორითმები შექმნილია ვიზუალური მონაცემების დასამუშავებლად და შესაბამისი მახასიათებლებისა და შაბლონების ამოღებისთვის. ეს ალგორითმები შეიძლება დაიყოს ორ ძირითად კატეგორიად: ტრადიციული კომპიუტერული ხედვა და ღრმა სწავლაზე დაფუძნებული მიდგომები [1].

ტრადიციული კომპიუტერული ხედვის ალგორითმები ეყრდნობა ხელით შექმნილ წესებს და ევრისტიკას სურათების მახასიათებლების ამოსაღებად და დასკვნების გასაკეთებლად. ისინი მოიცავს ისეთ ტექნიკას, როგორც კიდეების ამოცნობა, გამოსახულების გაფილტვრა, მახასიათებლების ამოღება და გეომეტრიული გარდაქმნები. მიუხედავად იმისა, რომ ეს მიდგომები ფართოდ იქნა გამოყენებული და ჯერ კიდევ აქტუალურია გარკვეულ აპლიკაციებში, ისინი ხშირად საჭიროებენ ექსპერტულ ცოდნას და შეიძლება ჰქონდეთ შეზღუდვები რთული ვიზუალური სცენარების მართვაში[1].

ღრმა სწავლებამ, მეორეს მხრივ, მოახდინა რევოლუცია კომპიუტერულ ხედვაში, ხელოვნური ნეირონული ქსელების გამოყენებით, რათა ისწავლონ უშუალოდ მონაცემებიდან. კონვოლუციური ნეირონული ქსელები (CNN) განსაკუთრებით პოპულარულია კომპიუტერული ხედვის ამოცანებში. CNN-ებს შეუძლიათ ავტომატურად ისწავლონ ვიზუალური მონაცემების იერარქიული წარმოდგენები, რაც მათ საშუალებას აძლევს ამოიცნონ ობიექტები, დაახარისხონ სურათები და შეასრულონ რთული ამოცანები განსაკუთრებული სიზუსტით. ღრმა სწავლის მოდელებმა მიაღწიეს შესანიშნავ შედეგებს კომპიუტერული ხედვის სხვადასხვა გამოწვევებში და გახდა მრავალი თანამედროვე აპლიკაციის ქვაკუთხედი (სურ.1)[2].

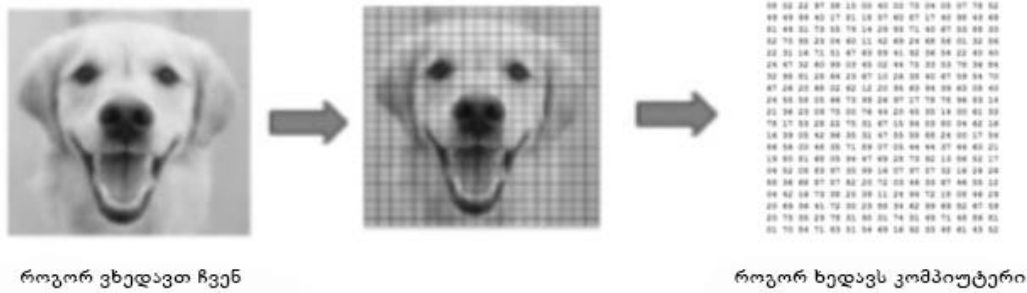


სურ. 1. კონვოლუციური ნეირონული ქსელი

გასული რამდენიმე ათწლეულის განმავლობაში, ღრმა სწავლება აღმოჩნდა ძალიან მძლავრი ინსტრუმენტი დიდი რაოდენობით მონაცემების დამუშავების უნარის გამო. ფარული ფენების გამოყენების ინტერესი აჭარბებს ტრადიციულ ტექნიკას, განსაკუთრებით შაბლონების ამოცნობაში. 1950-იანი წლებიდან, ხელოვნური ინტელექტის განვითარების ადრეული დღეებიდან, მკვლევარები იბრძოდნენ შეექმნათ სისტემა, რომელსაც შეუძლია ვიზუალური მონაცემების გაგება. მომდევნო წლებში ეს სფერო ცნობილი გახდა როგორც კომპიუტერული ხედვა. 2012 წელს კომპიუტერულმა ხედვამ კვანტური ნახტომი გააკეთა, როდესაც ტორონტოს უნივერსიტეტის მკვლევართა ჯგუფმა შეიმუშავა ხელოვნური ინტელექტის მოდელი, რომელმაც დიდი სხვაობით გადააჭარბა გამოსახულების ამოცნობის საუკეთესო ალგორითმებს.

AI სისტემა, რომელიც ცნობილი გახდა, როგორც AlexNet (მისი მთავარი შემქმნელის, ალექს კრიჟევსკის სახელი), გაიმარჯვა 2012 წლის ImageNet კომპიუტერული ხედვის კონკურსში საოცარი 85 პროცენტის სიზუსტით. AlexNet-ის ცენტრში იყო Convolutional Neural Networks სპეციალური ტიპის ნეირონული ქსელი, რომელიც უხეშად ასახავს ადამიანის ხედვას. წლების განმავლობაში CNN გახდა ძალიან მნიშვნელოვანი ნაწილი კომპიუტერული ხედვის მრავალი აპლიკაციისთვის [3],[4],[5].

როდესაც კონვოლუციური ნეირონული ქსელზე ვსაუბრობთ პირველი მნიშვნელოვანი რამ, რაც უნდა გვესმოდეს, არის ის, თუ როგორ ხედავს კომპიუტერი გამოსახულებას. ის არ ხედავს რეალურ ვიზუალურ შინაარსს, არამედ ის აღიქვამს როგორც პიქსელის რიცხვითი მნიშვნელობების მასივს (სურ. 2.).

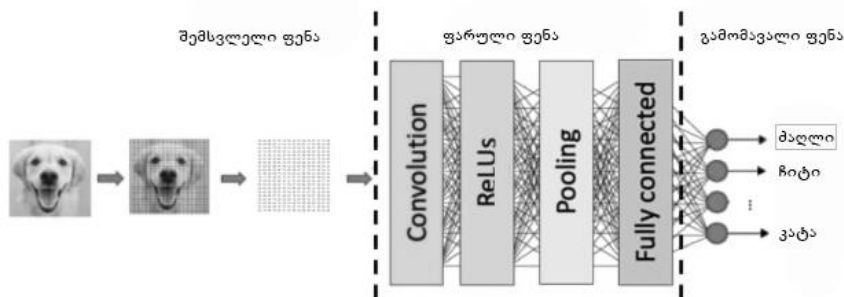


სურ. 2. ილუსტრაცია იმისა, თუ როგორ ხედავს კომპიუტერი სურათს

რით განსხვავდება კომპიუტერის გამოსახულების წარმოდგენა იმისგან, რასაც ჩვენ ვხედავთ?

თითოეული პიქსელი წარმოადგენს ერთ ფერს. თითოეული ფერი, თავის მხრივ, შედგება სამი ფერისგან წითელი, მწვანე და ლურჯისგან. ამ სამი ფერიდან თითოეულს კვლავ აქვს ათობითი ფერის მნიშვნელობა. ეს მნიშვნელობა შეიძლება მერყეობდეს 0-დან 255-მდე, 0 არის მთლიანად შავი და 255 მთლიანად თეთრი. აქედან გამომდინარე, თითოეული ფერის პიქსელი შეიძლება აღწერილი იყოს სამი ნომრით, ეგრეთ წოდებული RGB ათობითი მნიშვნელობებით. პიქსელი ძალის ღია ყავისფერი ბეწვიდან, მაგალითად, ითარგმნება როგორც RGB (255,248,220). ეს კოდი არის კომპიუტერისთვის ხელმისაწვდომი ერთადერთი პარამეტრი[6].

როდესაც ჩვენ ადამიანები ვხედავთ და ვახარისხებთ რაღაცას, ჩვენი ტვინი აცნობიერებს მას კონკრეტული შაბლონების ეტიკეტირების, პროგნოზირებისა და ამოცნობის გზით. ანალოგიურად, გამოსახულების კლასიფიკატორები ეყრდნობიან ტვინის მსგავს სტრუქტურებს: ეგრეთ წოდებულ კონვოლუციურ ნეირონულ ქსელებს (CNN). გამოსახულების კლასიფიკატორი იღებს გამოსახულების პიქსელის რიცხვით მნიშვნელობებს, გადასცემს მას CNN-ში და იღებს საბოლოო შედეგს. ეს გამომავალი შედეგი შეიძლება იყოს ერთი კლასი ან კლასების ალბათობა, რომელიც საუკეთესოდ აღწერს სურათს. პროცესი ასე გამოიყურება (სურ. 3).



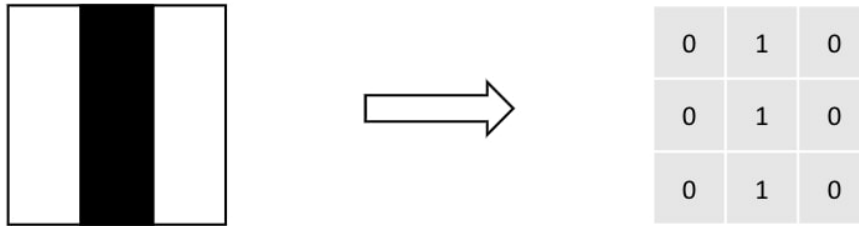
სურ. 3. როგორ ახარისხებს CNN სურათს

შემდეგ ფარული ფენები ასრულებენ თავის საქმეს და აკეთებენ პროგნოზს გამომავალ ფენაში. CNN-ის ფარული ფენების შიგნით არის "ჯადოსნური" ადგილი. სურათის შეტანიდან კლასის გამოტანამდე მოგზაურობისას CNN-ის თითოეული ფენის ტიპი ასრულებს კონკრეტულ დავალებას [2].

კონვოლუციის ფენა

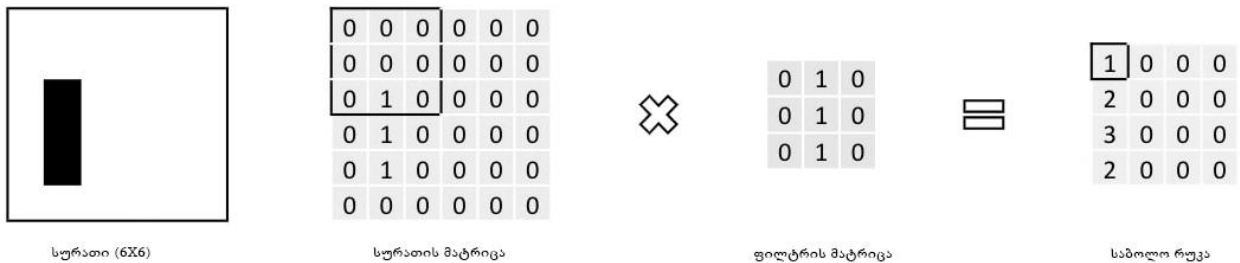
კონვოლუციის ფენები არის ძირითადი სამშენებლო ბლოკები გამოსახულების კლასიფიკატორებში. მათემატიკური ტერმინი კონვოლუცია ეხება ორი ფუნქციის ერთობლიობას (f და g), რომლებიც წარმოქმნიან მესამე ფუნქციას (z). კონვოლუციური ფენა იღებს შემსვლელ ობიექტს, იყენებს ფილტრს და გამოსცემს ფუნქციების რუკას. მახასიათებლის რუკა (z) არის შეყვანისა და ფილტრის (f და g) კომბინაცია, აქედან გამომდინარეობს სახელწოდება კონვოლუციის ფენა. კონვოლუციის მიზანია გამოსახულების მახასიათებლების ამოღება. მახასიათებელი არის ორიგინალური გამოსახულების სპეციფიკური მახასიათებელი, როგორიცაა წერტილები, კიდეები ან ძალის ცხვირის ფორმა. ციფრული სახით

დამუშავებული გამოსახულების მსგავსად, ფუნქცია ითარგმნება პიქსელის რიცხვითი მნიშვნელობების ყუთში. ეს მატრიცა ემსახურება როგორც მახასიათებლების დეტექტორს. წარმოიდგინეთ, რომ გსურთ ვერტიკალური ხაზის აღმოჩენა. ფუნქციების დეტექტორი ასე გამოიყურება (სურ.4)[2],[6].



სურ.4. ვერტიკალური ხაზის გამოვლენა

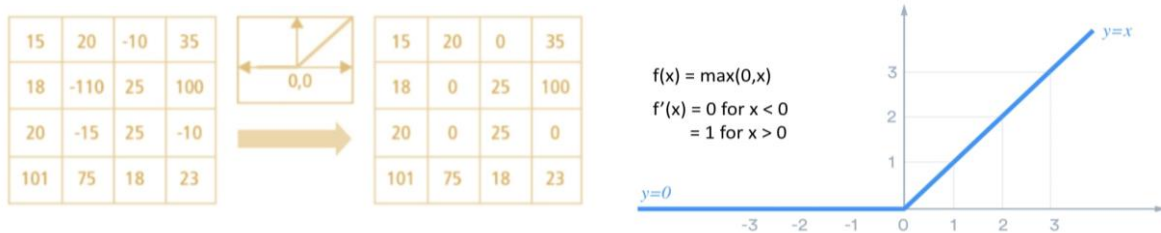
სურათის სკანირებისთვის ფილტრის მატრიცა მოძრაობს სურათზე, პიქსელის ბლოკი პიქსელის ბლოკით. თითოეული ქვერეგიონისთვის, ღირებულება გამოითვლება იმის საფუძველზე თუ რამდენად კარგი შესაბამისობაა ფილტრსა და სურათს შორის. საბოლოო რუკას ვიღებთ ორი მატრიცის მარტივად გადამრავლებით (სურ. 5).



სურ. 5. სურათის გადათარგმნა ფუნქციურ რუკაში

ReLU ფენა

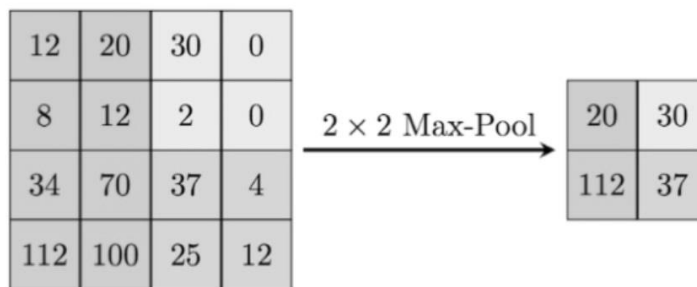
ReLU არის მოკლე გამოსწორებული ხაზოვანი ერთეულები და ე.წ. აქტივაციის ფუნქცია. აქტივაციის ფუნქციის გამოყენების მთავარი მიზანია გამოთვლაში არაწრფივობის დამატება. ReLU ფენა იღებს ფუნქციის რუკას (ანუ კონვოლუციის ფენის გამოსავალს) და ასწორებს ნებისმიერ უარყოფით მნიშვნელობას ნულამდე. დადებითი რიცხვები უცვლელი რჩება. არაწრფივობა ნიშნავს, რომ დახრილობა არ არის მუდმივი. ReLU ფენა არაწრფივია, რადგან ფუნქციის დახრილობა არის 0 ან 1(სურ. 6.) [6].



სურ. 6 ReLU ფენა

გაერთიანების ფენა

გაერთიანების პროცესში ფილტრი გადის შეყვანის მატრიცაზე და ანიჭებს ერთ მნიშვნელობას თითო ქვერეგიონზე ახალ გამომავალი მატრიცისთვის. გაერთიანების მთავარი მიზანია სურათის შემცირება. ეს გაზრდის გამოსახულების კლასიფიკატორის გამოთვლით სიჩქარეს. გაერთიანების ყველაზე გავრცელებული მიდგომა არის მაქსიმალური გაერთიანება (სურ. 7).



სურ. 7. max pooling-ში ფილტრი გამოსცემს კონკრეტული ქვერეგიონის მაქსიმალურ მნიშვნელობას.

ვიზუალიზაციაში, 2x2 ფილტრი გადის 4x4 შეყვანის მატრიცაზე. ფილტრი იღებს კონკრეტული პოზიციის მაქსიმალურ მნიშვნელობას და ანიჭებს მას, როგორც 1x1 მნიშვნელობას გამომავალ მატრიცას. მაგალითად, ზედა მარცხენა ოთხ ველში უმაღლესი მნიშვნელობა არის 20. შემდეგ ეს რიცხვი ენიჭება გამომავალი მატრიცის ერთ ველს. ოთხი ველიდან ერთამდე - ეს არის მონაცემთა შემცირება 75%-ით!

სრულად დაკავშირებული ფენა

სრულად დაკავშირებული ფენა მთავრდება CNN-ით. ბოლო გაერთიანების ფენის გამომავალი შეყვანის სახით, ის აერთიანებს ყველა ინფორმაციას და წარმოქმნის საბოლოო კლასიფიკაციას. ამ ეტაპზე, შეყვანის ვექტორის თითოეული მნიშვნელობა წარმოადგენს კონკრეტულ კლასს კუთვნილი მახასიათებლის ალბათობას. მაგალითად, ერთი მნიშვნელობა შეიძლება მიუთითებდეს, რომ თათი ძაღლს ეკუთვნის 90% ალბათობით. სხვა მნიშვნელობა შეიძლება შეესაბამებოდეს ცხვირს. სრულად დაკავშირებული ფენა იღებს ყველა ინფორმაციას, იყენებს მას წონაზე და გამოსცემს საბოლოო კლასიფიკაციას [6].

განვიხილოთ მაგალითი რომელიც გულისხმობს CIFAR-10 მონაცემთა ნაკრების გამოსახულების კლასიფიკაციას CNN-ით, Python პროგრამული ენის გამოყენებით. საწყის ეტაპზე ჩავტვირთოთ საჭირო CIFAR-10 მონაცემთა ნაკრები TensorFlow-დან.

```
import numpy as np
import matplotlib.pyplot as plt
import tensorflow as tf
cifar10 = tf.keras.datasets.cifar10
X_train, y_train, (X_test, y_test) = cifar10.load_data()
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

მოვახდინოთ მონაცემთა ნაკრების ვიზუალური გამოსახულების მიღება.

```
labels = ['Airplane', 'Automobile', 'Bird', 'Cat', 'Deer', 'Dog', 'Frog', 'Horse', 'Ship', 'Truck']
fig, axes = plt.subplots(ncols = 5, nrows = 4, figsize = (12, 12))
index = 0
for i in range(4):
    for j in range(5):
        axes[i,j].set_title(labels[y_train[index][0]])
        axes[i,j].imshow(X_train[index])
        axes[i,j].get_xaxis().set_visible(False)
        axes[i,j].get_yaxis().set_visible(False)
        index += 1
```

ნორმალიზაცია `X_train` და `X_test` მასივებში. თითოეული პიქსელის მნიშვნელობის 255.0-ზე გაყოფით, კოდი უზრუნველყოფს გამოსახულების მონაცემების ნორმალიზებას და მზადყოფნას გამოსაყენებლად ნეირონული ქსელის მოდელის სწავლებისთვის.

X_train = X_train/255.0

X_test = X_test/255.0

გამოვიყენოთ `to_categorical` ფუნქცია TensorFlow's Keras უტილიტებიდან ლეიბლის მასივების `y_train` და `y_test` ერთ ცხელ დაშიფრულ ვექტორებად გადასაყვანად.

```

from tensorflow.keras.utils import to_categorical
y_train = tf.keras.utils.to_categorical(y_train, 10)
y_test = tf.keras.utils.to_categorical(y_test, 10)
y_train

კონვოლუციური ნეირონული ქსელის (CNN) მოდელი Keras ბიბლიოთეკის გამოყენებით.
from keras.models import Sequential
from keras.layers import Conv2D, MaxPool2D, Flatten, Dense, Dropout, BatchNormalization
model = Sequential()
# Convolutional Layer
model.add(Conv2D(filters = 32,
                 kernel_size = (3,3),
                 input_shape = (32, 32, 3 ),
                 activation = 'relu',
                 padding='same'))
model.add(BatchNormalization())
model.add(Conv2D(filters=32, kernel_size=(3, 3), activation='relu', padding='same'))
model.add(BatchNormalization())
# Pooling layer
model.add(MaxPool2D( pool_size = (2,2)))
# Dropout Layer
model.add(Dropout(0.25))
model.add(Conv2D(filters=64, kernel_size=(3, 3), activation='relu', padding='same'))
model.add(BatchNormalization())
model.add(Conv2D(filters=64, kernel_size=(3, 3), activation='relu', padding='same'))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Conv2D(filters=128, kernel_size=(3, 3), activation='relu', padding='same'))
model.add(BatchNormalization())
model.add(Conv2D(filters=128, kernel_size=(3, 3), activation='relu', padding='same'))
model.add(BatchNormalization())
model.add(MaxPool2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
# Flattening
model.add(Flatten())
# Fully connected layers
model.add(Dense(128, activation = 'relu'))
model.add(Dropout(0.25))
model.add(Dense(10, activation = 'softmax'))

განსაზღვრული მოდელის კომპილაცია და ვარჯიში.
model.compile(loss = 'categorical_crossentropy', optimizer = 'adam',
metrics = 'accuracy')
model.fit(X_train, y_train, epochs=12)

```

მოდელის შეფასება

```
prediction = model.evaluate(X_test, y_test)  
print(f'Test Accuracy : {prediction[1] * 100:.2f}%')
```

გამომავალი შედეგი

```
313/313 [=====] - 3s 9ms/step - loss: 0.5228 - accuracy: 0.8300 Test Accuracy :  
83.00%
```

ხელოვნური ინტელექტისა და მაღალი ხარისხის კომპიუტერული ხედვის შერწყმამ მოახდინა რევოლუცია ჩვენი ვიზუალური მონაცემების აღქმისა და ანალიზის გზაზე. AI ალგორითმებმა ღრმა სწავლის ტექნიკით აჩვენეს შესანიშნავი სიზუსტე და მრავალფეროვნება კომპიუტერული ხედვის სხვადასხვა ამოცანებში. მაღალი ხარისხის გამოთვლითი სისტემები უზრუნველყოფენ აუცილებელ გამოთვლით რესურსებს ამ ალგორითმების ეფექტურად მომზადებისა და გამოყენებისთვის. უწყვეტი მიღწევებით, AI-ზე ორიენტირებული კომპიუტერული ხედვა მზად არის შეცვალოს მრავალი ინდუსტრია, გააძლიეროს ავტომატიზაცია, ეფექტურობა და გადაწყვეტილების მიღების პროცესები მომდევნო წლებში.

ლიტერატურა

1. Krishna, R., Computer vision: Foundations and Applications, published by Stanford university, December 2017
2. Alzubaidi, L., Zhang, J., Humaidi, A.J. *et al.* Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* **8**, 53 (2021). <https://doi.org/10.1186/s40537-021-00444-8>
3. Redmon, J., et al. "YOLO: Real-Time Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
4. Andrej Karpathy and Li Fei-Fei. "Deep Visual-Semantic Alignments for Generating Image Descriptions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
5. Liu, J., Shao, H., Jiang, Y. *et al.* CNN-Based Hidden-Layer Topological Structure Design and Optimization Methods for Image Classification. *Neural Process Lett* **54**, 2831–2842 (2022). <https://doi.org/10.1007/s11063-022-10742-8>
6. <https://poloclub.github.io/cnn-explainer/> (გადამოწმებულია 01.12.2022)

Abstract

The article discusses the collaboration of artificial intelligence (AI) and high-performance computer vision. The primary goal of computer vision is to interpret and analyze visual data to perform tasks that require human-like visual capabilities. The article highlights that computer vision algorithms can be divided into traditional computer vision and deep learning-based approaches. Traditional computer vision algorithms rely on hand-crafted rules. Deep learning, on the other hand, has revolutionized computer vision using artificial neural networks, particularly convolutional neural networks (CNNs), which learn directly from data. This enables them to recognize objects, classify images and perform complex tasks with high accuracy. Advances in computer vision algorithms, especially deep learning through CNNs, have shown remarkable progress in various computer vision tasks and applications, enhanced the capabilities of AI systems, and opened a new window for industries.

Key words: Artificial intelligence, computer vision, convolutional neural network, deep learning, modern technologies.