چکیده مقالات

Code: mvip1005

Spectral-Spatial Anomaly Detection in Hyperspectral Imagery Based on Dual-Domain Autoencoders

Mohamad Ebrahim Aghili Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran
m.aghili@modares.ac.ir

Hassan Ghassemian* Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran
ghassemi@modares.ac.ir

mage processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran
maryam.imani@modares.ac.ir

Abstract—Hyperspectral anomaly detection is crucial for applications like aerial surveillance in remote sensing images. However, robust identification of anomalous pixels remains challenging. A novel spectral-spatial anomaly detection technique called Dual-Domain Autoencoders (DDA) is proposed to address these challenges. First, Nonnegative Matrix Factorization (NMF) is applied to decompose the hyperspectral data into anomaly and background components. Refinement of the designation is then done using intersection masking. Next, a spectral autoencoder is trained on identified background signature pixels and used to reconstruct the image. The reconstruction error highlights spectral anomalies. Furthermore, a spatial autoencoder is trained on principal component patches from likely background areas. Fused reconstruction error from the spectral and spatial autoencoders is finally used to give enhanced anomaly detection. Experiments demonstrate higher AUC for DDA over individual autoencoders and benchmark methods. The integration of matrix factorization and dual-domain, fused autoencoders thus provides superior anomaly identification. Spatial modeling further constrains the background, enabling accurate flagging of unusual local hyperspectral patterns. This study provides the effectiveness of employing autoencoders trained on intelligently sampled hyperspectral pixel signatures and spatial features for improved spectral-spatial anomaly detection.

Keywords—hyperspectral image, anomaly detection, autoencoder, matrix factorization, spectral-spatial features

Code:mvip1008

Tomato Leaf Disease Detection Using Transfer Learning: A Comparative Study

Mahsa Bahrami Digitalization SBU Management Mapna Electric & Control Engineering & Manufacturing (MECO) Alborz, Iran
Bahrami.Mahsa@mapnaec.com

Aliakbar Pourhatami Digitalization SBU Management Mapna Electric & Control Engineering & Manufacturing (MECO) Alborz, Iran
Pourhatami.Aliakbar@mapnaec.com

Mohsen Maboodi Digitalization SBU Management Mapna Electric & Control Engineering & Manufacturing (MECO) Alborz, Iran
Maboodi.Mohsen@mapnaec.com

Abstract— Tomato, an important food crop enriched with essential minerals and vitamins, is widely used in various cuisines. Farmers should accurately surveillance tomato farms to manufacture quality foods free from any type of pest. Tomato leaf disease is a common problem that impacts the quantity and quality of production. Therefore, accurate and robust algorithms are needed for the detection of tomato leaf disease. Among various methods, deep learning algorithms are taken an added importance in terms of automatic, accurate, and robust leaf disease detection. In this paper, we provide a comparative study of transfer learning, reusing a pre-trained model for a new task, networks including VGG19, ResNet-101, and MobileNetv2 on two benchmark datasets including PlantVillage and CCMT datasets. The best performance was achieved by VGG19 where the best accuracy, precision, recall, and F1-score on the test set of PlantVillage and CCMT datasets were 99.48%, 99.27%, 99.28%, 99.27%, and 92.76%, 92.74%, 95.09%, 90.86%, respectively. The results show that VGG19 algorithms have the potential for accurate and robust tomato leaf disease detection.

Keywords— detection, tomato, transfer learning, leaf disease, deep learning.

Code:mvip1010

Improved Stereo Depth Estimation Using Smoothness and Geometrical Attention

Seyedeh-Parisa Zarei School of Electrical and Computer Engineering, College of Eng. University of Tehran Tehran, Iran
sp.zarei@ut.ac.ir

Hamid Soltanian-Zadeh CIPCE, School of Electrical and Computer Engineering, College of Eng. University of Tehran Tehran, Iran
hszadeh@ut.ac.ir

Ahmah Kalhor School of Electrical and Computer Engineering, College of Eng. University of Tehran Tehran, Iran
akalhor@ut.ac.ir

Abstract— Depth estimation is a critical step for many computer vision tasks such as scene understanding, registration, and localization. The view synthesis-based method estimates depth in a self-supervised framework without any expensive ground truth. However, this method suffers from the so-called ill-posed problem. A general solution to ill-posed problems is to incorporate relevant constraints and regularizations. To this end, we propose a new attention module and a loss term enforcing causation between a 2-D image and the corresponding depth map. The results show that the proposed method has made overall improvements in terms of accuracy and time required for training. In particular, while converging 6 epochs faster than the base model, the model outperforms the base model, MonoDepth2, on standard metrics, e.g., by 6% on RMSlog.

Keywords— self-supervised depth estimation, geometrical attention, causality, deep learning

Code:mvip1015

Hybrid Approach for Automated Discrimination between COVID-19 and Similar Respiratory Diseases

Houman bahrami Rad
Department of Information Technology
Tarbiat Modares Tehran, Iran
b.houman@modares.ac.ir

Mehrdad Kargari
Department of Information Technology
Tarbiat Modares Tehran, Iran
M_kargari@modares.ac.ir

Meysam Alavi Department of Information Technology
Tarbiat Modares Tehran, Iran
meysamalavi@modares.ac.ir

Amir Behnam Kharazmy
Department of Internal Medicine Shahid Beheshti University of Medical Sciences
Tehran, Iran
Drkharazmy@sbmu.ac.ir

Abstract— One of the main challenges in accurately diagnosing COVID-19 is its clinical manifestations, which are similar to some respiratory diseases such as viral and bacterial pneumonia, and even influenza during cold seasons. These similarities may lead to misdiagnosis and periodically threaten public health. In order to achieve an accurate and rapid diagnosis of COVID-19 and differentiate it from other respiratory diseases with similar features, this research aims to provide a comprehensive system. This system utilizes artificial intelligence and machine learning methods to analyze lung CT scan images to distinguish COVID-19 from other lung diseases. To train this system, data from lung CT scan images of patients from Imam Hussein Hospital in Tehran were used. The data includes three categories of patients: COVID-19 patients, patients with lung pneumonia, and patients with other respiratory conditions. In this study, by using artificial intelligence and machine learning methods, two main phases of patient classification were conducted. The results show that deep learning models and hybrid models achieved acceptable performance with accuracies of 98.98% and 99.79%, respectively.

Keywords—COVID-19, Deep Learning, Hybrid model, CT Scan Images

Code:mvip1016

Weighted Features Based Classification of Polarimetric SAR Images

Fatemeh Saneipour Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran
fatemeh.saneipour@modares.ac.ir

Maryam Imani Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran
maryam.imani@modares.ac.ir

Hassan Ghassemian Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran
ghassemi@modares.ac.ir

Abstract— Today, classification of polarimetric images is an important topic where various statistical pattern recognition methods have been used to achieve the high accurate classification maps. In this work, weighting the polarimetric features according to their statistical behavior (the mean vector and variance values as the first and second statistics) is suggested to improve the PolSAR image classification. A weighted feature matrix is composed and applied to the popular classifiers such as maximum likelihood, K-nearest neighbor and support vector machine. The weighted feature matrix can be also implemented on other arbitrary classifiers to improve their discrimination ability. The experiments on the L-band AIRSAR dataset show appropriate classification results.

Keywords—polarimetric synthetic aperture radar (PolSAR), weighted features, classification.

Code:mvip1017

Design of a Vision-Based Dynamic Positioning Control System for a Work-class ROV Using Hardware-in-the-Loop Simulation

Ali Nourmohammadi Department of Mechanical Engineering Isfahan University of Technology Isfahan, Iran, 84156-83111 ali.nourmohamadi@alumni.iut.ac.ir

Mehdi Loueipour Research Institute for Subsea Science and Technology Isfahan University of Technology Isfahan, Iran, 84156-83111 loueipour@iut.ac.ir

Alireza Hosseinnajad Research Institute for Subsea Science and Technology Isfahan University of Technology Isfahan, Iran, 84156-83111 a.hosseinnajad@alumni.iut.ac.ir

Abstract—This paper introduces a vision-based dynamic positioning (DP) control system and develops a hardware-in-theloop (HIL) platform to validate the performance of the controller applied to a work-class remotely operated vehicle (ROV). The proposed platform consists of three main parts: hardware, image processing part and controller. The hardware included a calibrated camera that was connected to a dedicated computer via USB 2.0. In the image processing part after pre-processing a circular Hough transform was used to detect and determine the position of the target in the image plane. Furthermore, this paper proposed a feedforward proportional-integral-derivative (PID) controller. To evaluate the performance of the proposed controller, two scenarios were implemented. In the first scenario, the target was considered stationary and a disturbance was applied to the ROV in the simulation environment. In the second scenario, the target object has moved along a rectangular path, and the objective was to stabilize the ROV at the desired points. In both scenarios, the reference signal was acquired by a calibrated camera from the target and sent to the controller. The results showed the desirable performance of the proposed controller.

Keywords—Remotely operated vehicle, Hardware-in-the-loop simulation, Vision-based control, Dynamic positioning.

Code:mvip1020

An Improved U-Net Image Segmentation Network for Crankshaft Surface Defect Detection

Ashkan Moosavian
Department of Agricultural Engineering
Technical and Vocational University (TVU)
Tehran, Iran
a_moosavian@tvu.ac.ir

Elmira Bagheri School of Automotive Engineering
Iran University of Science and Technology (IUST)
Tehran, Iran
elmira_bagheri@alumni.iust.ac.ir

Alireza Yazdanijoo*
Irankhodro powertrain company (IPCo) Tehran, Iran
alireza_gorgvand@auto.iust.ac.ir

Amir Hossein Barshooi
School of Automotive Engineering Iran University of Science and Technology (IUST)
Tehran, Iran
a_barshooi@alumni.iust.ac.ir

Abstract— Crankshaft is one of the mechanical components of the vehicle engine, and quality control of it holds significant importance in the production line. In this paper, a vision-based system was developed to detect apparent structural defects on the crankshaft surface. By examining the different approaches in computer vision tasks, the semantic segmentation technique was chosen to solve this problem. In the first stage, a dataset consisting of 400 crankshaft experimental images with structural defects such as scratch, pitting, and grinding were collected. Then, the Convolutional Neural Network (CNN) with MobileNet architecture was trained to detect apparent defects, and an Intersection over Union (IoU) evaluation criteria of 64.7% was obtained. In the third stage, some image processing techniques were used to increase the performance. By applying the DexiNed edge detection filter on the train-set images, the IoU was increased by 8.4%. Considering the importance of this issue in the automotive industry, it has been tried again to boost the performance by augmenting the dataset images. On the other hand, this can also prevent overfitting of the model. By training the model under the same conditions as the previous stages, the IoU in this stage increased by 13.2% and reached 86.3%.

Keywords—defect detection, crankshaft, deep learning, segmentation, pitting, scratch

Code:mvip1021

Efficient Bitrate Ladder Construction using Transfer Learning and Spatio-Temporal Features

Ali Falahati[1] , Mohammad Karim Safavi[1] , Ardavan Elahi[1] , Farhad Pakdaman[2] , and Moncef Gabbouj[2]

[1] Avanco, Tehran, Iran

[2] Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland

Abstract— Providing high-quality video with efficient bitrate is a main challenge in video industry. The traditional one-size-fitsall scheme for bitrate ladders is inefficient and reaching the best content-aware decision computationally impractical due to extensive encodings required. To mitigate this, we propose a bitrate and complexity efficient bitrate ladder prediction method using transfer learning and spatio-temporal features. We propose: (1) using feature maps from well-known pre-trained DNNs to predict rate-quality behavior with limited training data; and (2) improving highest quality rung efficiency by predicting minimum bitrate for top quality and using it for the top rung. The method tested on 102 video scenes demonstrates 94.1% reduction in complexity versus brute-force at 1.71% BD-Rate expense. Additionally, transfer learning was thoroughly studied through four networks and ablation studies.

Keywords— HTTP Adaptive Streaming (HAS), Bitrate ladder, CRF, Transfer Learning, Pareto Front

Code:mvip1024

Highresolution Remote Sensing Image Classification With Limited Training Data

Mehdi Ariaei Image Processing and Information Analysis Lab Tarbiat Modares University Tehran, Iran Email: m.ariaei@modares.ac.ir

Hassan Ghassemian Image Processing and Information Analysis Lab Tarbiat Modares University Tehran, Iran Email: ghassemi@modares.ac.ir

Maryam Imani Image Processing and Information Analysis Lab Tarbiat Modares University Tehran, Iran Email: maryam.imani@modares.ac.ir

Abstract— Accurate classification of land cover from aerial images is one of the research topics in remote sensing and is also in high demand in industry. However, obtaining labeled data for training different classifiers that heavily depend on supervision is still a challenging and resource-intensive task. Unsupervised methods have emerged as a powerful alternative to overcome the limitations associated with labeled data. Such methods have a high ability to discover hidden patterns and structures in multi-spectral images and have the possibility of classifying various types of land cover without relying on labeled samples. Our research primarily involved the analysis of World-View3 satellite imagery. Our strategy was creating an advanced pipeline that extracted features using autoencoders. Through this approach, the hyperspectral images' key characteristics are efficiently extracted. Subsequently, we implement transfer learning to re-train the model with a limited number of labeled data. By applying transfer learning, our pipeline significantly enhances the capability of multispectral image processing, enabling a more comprehensive and accurate interpretation of satellite imagery data. Finally, we evaluate our results not only by providing a confusion matrix but also through a visual comparison between the class map and the RGB composition of the MSI image.

Keywords—World-View3, Autoencoders, Multi-Spectral images, Deep Learning, Transfer Learning, Limited Training Data

Code:mvip1025

Spatial-Spectral Neural Network for High Resolution Multispectral Image Classification

Roozbeh Tanha Image processing and Information Analysis Lab Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran roozbeh.tanha@modares.ac.ir

Hassan Ghassemian* Image processing and Information Analysis Lab Faculty of Electrical and Computer Engineering Tarbiat Modares University Tehran, Iran ghassemi@modares.ac.ir

Abstract— Classification of multispectral images in remotesensing area having the capability to analyze and categorize diversified land cover. In this issue, extracting suitable spatial, spectral and even temporal features is one of the main challenges. Also, the existence of sufficient data required for the classification training process is another challenge, because in many cases it may not be available and we may not even have a reliable classification map. The use of neural networks for simultaneous feature extraction and classification is very popular and significant progress has been made in this field, but these networks usually have a high computational cost and require significant training data in the training process. In this work we propose a neural network for multispectral image classification purpose which requires few training samples and less calculation without using filterbanks for spatial feature extraction and it can improve classification accuracy by fusion of spatial and spectral features. The simulations indicate that the proposed method shows an acceptable performance.

Keywords—Remote sensing; multispectral image classification; feature extraction; neural network;

Code:mvip-1026

بهسازی سامانه بازشناسی چهره با استفاده از نمایش تنک و پیش پردازش مبتنی بر آنالیز مولفه اساسی

حمیدرضا کاوه یزدی و مسعود رضا آقا بزرگی
کارشناسی ارشد – دانشگاه یزد – دانشکده مهندسی برق، hamidreza.k2015@gmail.com
استاد– دانشگاه یزد – دانشکده مهندسی برق ،aghabozorgi@yazd.ac.ir

چکیده ـ چهره نقشی اساسی در شناسایی افراد و نمایش احساسات آنها دارد. لذا بازشناسی چهره به یک موضوع مهم در سیستم -های امنیتی و شناسایی مجرمان تبدیل شده است. از مهمترین مسائل در راستای فرایند بازشناسی چهره، تعیین یک ویژگی مناسب و دیگری انتخاب یک معیار سنجش شباهت مناسب است. در این مقاله ابتدا روش نمایش تنک را بررسی کرده و سپس استفاده از آن در فرایند بازشناسی چهره به منظور افزایش دقت فرایند را بررسی می کنیم. در این میان فرضیاتی که در نظر گرفته میشود، همان فرضیات روش نمایش تنک است. با این تفاوت که سعی شده با استفاده از پیش پردازش مبتنی بر روش آنالیز مولفه اساسی، دادههایی که ناهمبستگی بیشتری با داده اصلی دارند کنار گذاشته شوند و عملیات محاسبه فاصله میان تصاویر با استفاده از نمایش تنک فقط روی تعداد کمی از دادههای دارای شباهت زیاد که از خروجی الگوریتم آنالیز مولفه اساسی بدست آمده است انجام گیرد. پس از آزمایش هم زمان معیار فاصله بر مبنای نمایش تنک و الگوریتم پیشنهادی بر روی یک پایگاه داده چهره به نام face T&AT شاهد افزایش دقت سامانه بازشناسی چهره به اندازه 6 درصد بوده ایم. همچنین زمان اجرای الگوریتم پیشنهادی نیز به میزان قابل مالحظهای کاهش یافته است.

کلید واژه– بازشناسی چهره، نمایش تنک، فاصله فشردگی، آنالیز مولفه اساسی

Code: mvip-1033

A Comparison Between CCTV and Industrial Cameras for Vehicle Attribute Recognition

Mohammadreza Asadi *AI Research Center HoopadVision Company* Isfahan, Iran
mohammadreza.asadi1992@gmail.com

Mohammad Yasin Fakhar *AI Research Center HoopadVision Company* Isfahan, Iran
yasin.fakhar34@gmail.com

Seyedeh Sogand Hashemi *AI Research Center HoopadVision Company* Isfahan, Iran
hashemisogand94@gmail.com

Safiyeh Rezaei *AI Research Center HoopadVision Company* Isfahan, Iran
safiyeh.rezaei73@gmail.com

Mohamad Kiani Abari *AI Research Center HoopadVision Company* Isfahan, Iran
kiani@hoopadvision.com

Seyed Alireza Abbaspour *AI Research Center HoopadVision Company* Isfahan, Iran
seyedalirezaabbaspour@gmail.com

Abstract— In machine/computer vision, cameras serve a major role in image acquisition. Surveillance scenarios typically rely on Closed-Circuit Television (CCTV) cameras. This study aims to evaluate industrial cameras within a surveillance application, contrasting their performance with that of CCTV cameras. We explore the comparative analysis of CCTV and industrial cameras for vehicle attribute recognition, specifically concentrating on the recognition of vehicle color and model using deep learning techniques. To train and evaluate the models, we have created datasets from images captured by both a CCTV and an industrial camera. Our findings indicate that the industrial camera outperforms the CCTV. However, employing advanced processing algorithms has the potential to minimize the performance gap between these two cameras. Our research represents one of the initial comparative analyses between these camera types, offering valuable guidance in selecting the most suitable camera for specific applications.

Keywords— CCTV, color recognition, deep learning, industrial camera, surveillance systems

Code:mvip-1034

Automatic Detection of Brain Tumor on MRI Images Using a YOLO-Based Algorithm

Masume Rahimi
Faculty of Electrical and Computer Engineering
University of Zanjan
Zanjan, Iran
masume.rahimi@znu.ac.ir

Mohammad Mostafavi
Faculty of Electrical and Computer Engineering
University of Zanjan
Zanjan, Iran
mmostafavi@znu.ac.ir

Abazar Arabameri
Faculty of Electrical and Computer Engineering
University of Zanjan
Zanjan, Iran
arabameri@znu.ac.ir

Abstract— This paper presents a study on the application of transfer learning and fine-tuning techniques to a deep learning model for the purpose of detecting three specific types of brain tumors from MRI images. The proposed approach utilizes the YOLO algorithm for automatic diagnosis. Specifically, the YOLOv4-tiny model, which is a smaller version of the YOLOv4 algorithm, was trained and evaluated due to its improved performance. The dataset utilized in this research is obtained from the figshare data repository, which comprises of labeled MRI images. The division of the dataset resulted in 80% for training, 10% for validation, and 10% for testing purposes. Additionally, a pre-processing technique was devised to enhance the features in the MRI images. The outcomes of the implementation demonstrate that the YOLOv4-tiny model obtained a mean average precision (mAP) of 0.8074 for the raw data and 0.8324 for the processed data.

Keywords—Object detection algorithm, YOLO algorithm, bounding box, transfer learning.

Code:Mvip 1035

Developing a novel deep learning approach to diagnosis COVID-19 disease using lung CT-scan images

Fatemeh Savei
Operation Management and Information Technology
Kharazmi University
Tehran, Iran
fatemee.savei@khu.ac.ir

*Omid Mahdi Ebadati E.
Operation Management and Information Technology
Kharazmi University
Tehran, Iran
ebadati@khu.ac.ir
*Corresponding Author

Seyed Hossein Siadat
Operation Management and Information Technology
Kharazmi University
Tehran, Iran
siadat.hossein@khu.ac.ir

Milad Masroor
Institute of Petroleum Engineering, School of Chemical engineering
University of Tehran
Tehran, Iran
milad.masroor7275@ut.ac.ir

Abstract— The aim of this research is to present a suitable method for feature extraction from chest CT scan images to enhance the accuracy and speed of detecting the coronavirus. In today's world, the role of engineering in medicine has significantly expanded, and with the advancement of imaging technology and image processing, disease diagnosis has become faster, easier, and more precise. The viral infection of COVID-19, which originated in Wuhan, China and spread worldwide, has resulted in the death of over 4.4 million people, despite the initiation of vaccination efforts. Due to the high demand for PCR kits and their severe shortage, radiographic techniques such as X-rays and CT scans can be utilized for diagnostic purposes. Rapid detection of the coronavirus in the early stages can significantly prevent mortality from this devastating disease. Given the difficulty of diagnosing this disease in the early stages, providing a method that facilitates the early diagnosis of COVID-19 is highly valuable. In the proposed method, a modified two-dimensional convolutional neural network called Residual 2D-CNN is employed for more accurate and faster detection of COVID-19 from CT scan images. The final results of this research demonstrate an approximate 97% accuracy in detecting this virus.

Keywords— CT scan, Deep Learning, Residual 2D-CNN, COVID-19.

Code:mvip-1038

Innovative Diagnosis of Dental Diseases Using YOLO V8 Deep Learning Model

1st Marzieh Razaghi Department of Electrical And Computer Engineering Kashan University Kashan, Iran m.razaghi@grad.kashanu.ac.ir

2nd Dr.Hossein Ebrahimpour Komleh Department of Electrical And Computer Engineering Kashan University Kashan, Iran ebrahimpour@kashanu.ac.ir *Corresponding Author

3rd Dr.Fereshteh Dehghani Department of Electrical And Computer Engineering Kashan University Kashan, Iran fdehghani@kashanu.ac.ir

4th Dr.Zahra Shahidi Department of Restorative Dentistry Tehran University of Medical Science Tehran, Iran zshahidi@sina.tums.ac.ir

Abstract—The diagnosis and identification of dental problems pose significant challenges. Traditionally, dental disease diagnosis was a manual and time-consuming process, requiring dentists to meticulously examine and evaluate the condition. The integration of artificial intelligence (AI) represents a transformative approach to aid in medical imaging diagnostics. Specifically, leveraging AI for diagnosing dental issues entails the automatic localization of lesions. In this study, the Yolo V8 deep learning model is employed to develop an innovative method for the detection and categorization of common dental problems. The primary objective of this approach is to establish a comprehensive database comprising two distinct categories of dental X-ray images: BiteWing X-ray Images and Orthopantomography X-ray (OPG). These categories aim to facilitate the diagnosis and classification of various dental diseases. The results of the experiments showed that the best performance in training YOLOv8m was achieved with mAP of 71.6%, recall of 90%, and precision of 90%.

Keywords— Dental Diseases, YOLO V8, Deep learning model, BiteWing, OPG.

Code:mvip-1039

Fuzzy-Based Algorithm for Efficient Vehicle License Plate Recognition in Iran's Transportation System

Masoud Karimi
Faculty of Electrical Engineering
Shahid Beheshti University,
Tehran, Iran
mas.karimi@mail.sbu.ac.ir

Mahdi Pourgholi
Faculty of Electrical Engineering
Shahid Beheshti University,
Tehran, Iran
m_pourgholi@sbu.ac.ir

Abstract—This paper presents a novel approach to license plate recognition in Iran's transportation system by employing a fuzzy-based algorithm. The proposed method involves a four-stage process, including the conversion of RGB images to grayscale, fuzzy thresholding using a Fuzzy Inference System (FIS), character segmentation, and application of fuzzy clustering for license plate number detection. Simulation results demonstrate the effectiveness of the proposed method in achieving rapid and accurate license plate number detection.

Keywords-Fuzzy logic; Image processing; Fuzzy edge detection; Fuzzy character recognition; Fuzzy classification.

Code: mvip1040

Adaptive Blind Watermarking in DWT-DCT Domain Based on Edge and Brightness Information

Arezoo PariZangeneh
Department of Electrical and Computer Engineering
Isfahan University of Technology
Isfahan, 84156-83111 Iran
a.zanganeh@ec.iut.ac.ir

Ghazale Ghorbanzade
Department of Electrical and Computer Engineering
Isfahan University of Technology
Isfahan, 84156-83111 Iran
gh_ghorbanzade@alumni.iut.ac.ir

Zahra Nabizadeh
Department of Electrical and Computer Engineering
Isfahan University of Technology
Isfahan, 84156-83111 Iran
z.nabizadeh@ec.iut.ac.ir

Nader Karimi
Department of Electrical and Computer Engineering
Isfahan University of Technology
Isfahan, 84156-83111 Iran
nader.karimi@cc.iut.ac.ir

Shadrokh Samavi
Department of Electrical and Computer Engineering
McMaster University
L8S 4L8, Canada
samavi@mcmaster.ca

Abstract— With the growth of editing and sharing of images through the internet, the importance of the protection of the images' authority has increased. Robust watermarking is a known approach to maintaining copyright protection. Robustness and imperceptibility are two factors that are tried to be maximized through the watermarking process, although increasing the robustness of the algorithms lessens the imperceptibility of the watermarking. In this paper, an adaptive method is proposed that determines the amount of watermark embedding in different parts of the cover image regarding its texture and brightness. Adaptive embedding increases the robustness while preserving the quality of the watermarked image. Experimental results also show that the proposed method can effectively reconstruct the embedded payload in different kinds of common watermarking attacks. Our proposed method has shown good performance in comparison with a method that have utilized the same watermark length and image sets.

Keywords—robust watermarking algorithms, digital cosine transform (DCT), digital wavelet transform (DWT)

Code: mvip-1041

ایجاد ساختار شبکه عصبی عمیق توزیع شده بر روی دستگاه‌های لبه شبکه و مه با بهره‌گیری از محاسبات شناختی

امید جمشیدی،
دانشجوی دکتری هوش مصنوعی
دانشگاه بوعلی سینا،
o.jamshidi@eng.basu.ac.ir

مهدی عباسی ،
عضو هیات علمی گروه
کامپیوتر دانشگاه بوعلی سینا،
abbasi@basu.ac.ir

عباس رمضانی،
عضو هیات علمی گروه برق
دانشگاه بوعلی سینا،
a.ramazani@basu.ac.ir

محمدرضا خسروی،
محقق دانشگاه خلیخ فارس ایران،
m.khosravi@mehr.pgu.ac.ir

چکیده —شبکه‌های عصبی کانولوشنی، به عنوان ابزاری کارآمد، در زمینه‌های گوناگون از جمله علوم مهندسی و پزشکی معرفی شده‌اند. اما یکی از چالش‌های اصلی استفاده از این شبکه‌های عصبی، تعداد زیاد پارامترها و نیاز به منابع رایانشی فراوان است. استفاده از منابع رایانشی در مه و دستگاه‌های لبه شبکه، روش مناسبی برای تامین منابع رایانشی مورد نیاز و توزیع شبکه‌های عصبی کانولشنی در این ساختار می‌باشد. اما توزیع شبکههای عصبی در این ساختار باعث ایجاد ازدحام در شبکه انتقال داده در لبه شبکه و مه می‌گردد. در این مقاله، یک ساختار توزیع شده از شبکه عصبی کانولوشنی در مه و دستگاه‌های پایانی لبه شبکه معرفی شده است و از محاسبات شناختی برای انتخاب دستگاههای مورد نیاز در لبه شبکه جهت مشارکت در ساخت مدل استفاده می‌شود. روش پیشنهادی میزان ازدحام ایجاد شده را قابل تنظیم و همچنین دقت مدل را حفظ می‌نماید. نتایج نشان می‌دهد که با کاهش حدود 60 درصدی تعداد دستگاه‌های لبه شبکه، در دقت مدل کاهش قابل توجهی ایجاد نمی‌شود.

کلید واژه— شبکه عصبی توزیع شده، محاسبات شناختی، یادگیری عمیق

Code: mvip1044

Presenting a two-dimensional Restricted Boltzmann Machine Network for Action Recognition in Video

Majid Joudaki
Electrical and Computer Engineering
University of Kashan
Kashan, Iran
m.joudaki@gmail.com

Hossein Ebrahimpour-Komleh
Electrical and Computer Engineering
University of Kashan
Kashan, Iran
ebrahimpour.kashanu@gmail.com

Abstract— In the realm of video analysis, particularly in identifying human activities, there has been a notable shift towards utilizing advanced neural network architectures. Among these, deep belief networks stand out for their efficient learning and layered configuration. This paper introduces a novel recurrent network structure that integrates deep belief networks, tailored to address three key challenges. Firstly, it processes twodimensional video frames directly. Secondly, it employs restricted Boltzmann machines to grasp short-term temporal patterns. Lastly, it incorporates recursive processing for long-term temporal learning. Tested on the KTH and UCFSports datasets, the model achieved significant accuracies of 94.17% and 91.86%, respectively, with an in-depth analysis of the results provided.

Keywords— Deep learning, deep belief networks, restricted Boltzmann machine, action recognition, recurrent networks.

Code: mvip1051

Ovarian Tumor Ultrasound Image Segmentation with Deep Neural Networks

Mahnaz Siahpoosh
Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
m.siahpoosh@modares.ac.ir

Maryam Imani
Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
maryam.imani@modares.ac.ir

Hassan Ghassemian
Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
ghassemi@modares.ac.ir

Abstract—The precise and automated segmentation of ovarian tumors in medical images plays a pivotal role in the treatment of ovarian cancer in women. U-Net has demonstrated remarkable success in the field of medical image segmentation. However, due to its small receptive field, U-Net faces challenges in extracting global context information. Moreover, due to the significant variation in scale and size among tumors, it is essential to employ a network capable of effectively extracting information at Multiple scales. In this study, we present a U-Net-based network named PCU-Net for the segmentation of ovarian tumors, incorporating ConvMixer and Pyramid Dilated Convolution (PDC) modules. The ConvMixer module captures global context information by utilizing large-size kernels. The PDC module integrates local and global contextual patterns through utilization of parallel dilated convolution with different dilation rate. Furthermore, our model is computationally economical, due to its small number of network parameters. We assess the proposed method's performance using the Multi-Modality Ovarian Tumor Ultrasound (MMOTU) dataset. The results indicate that in comparison to U-Net, our proposed PCU-Net exhibits an improvement of 4.23% in terms of Intersection over Union (IoU) and 2.99% in terms of Dice Similarity Coefficient (DSC).

Index Terms—Segmentation, Ultrasound Images, U-Net, ConvMixer, Pyramid Dilated Convolution

Code: mvip1052

Leveraging Swin Transformer for Local-to-Global Weakly Supervised Semantic Segmentation

Rozhan Ahmadi
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
roz.ahmadi@sharif.edu

Shohreh Kasaei
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
kasaei@sharif.edu

Abstract—In recent years, weakly supervised semantic segmentation using image-level labels as supervision has received significant attention in the field of computer vision. Most existing methods have addressed the challenges arising from the lack of spatial information in these labels by focusing on facilitating supervised learning through the generation of pseudolabels from class activation maps (CAMs). Due to the localized pattern detection of Convolutional Neural Networks (CNNs), CAMs often emphasize only the most discriminative parts of an object, making it challenging to accurately distinguish foreground objects from each other and the background. Recent studies have shown that Vision Transformer (ViT) features, due to their global view, are more effective in capturing the scene layout than CNNs. However, the use of hierarchical ViTs has not been extensively explored in this field. This work explores the use of Swin Transformer by proposing "SWTformer" to enhance the accuracy of the initial seed CAMs by bringing local and global views together. SWTformer-V1 generates class probabilities and CAMs using only the patch tokens as features. SWTformer-V2 incorporates a multi-scale feature fusion mechanism to extract additional information and utilizes a background-aware mechanism to generate more accurate localization maps with improved cross-object discrimination. Based on experiments on the PascalVOC 2012 dataset, SWTformer-V1 achieves a 1.27% mAP higher localization accuracy, outperforming state-of-the-art models. It also yields comparable performance by 0.86% mIoU on average higher than other methods in generating initial localization maps, depending only on the classification network. SWTformer-V2 further improves the accuracy of the generated seed CAMs by 5.32% mIoU, further proving the effectiveness of the local-to-global view provided by the Swin transformer. Code available at: https://github.com/RozhanAhmadi/SWTformer

Index Terms—Weakly Supervised Semantic Segmentation, Class Activation Map, Hierarchical Vision Transformer, Imagelevel label

Code: mvip1089

Ball Detection Algorithms Enhancement in Sport Robots

Mohammad Esfandiarpour
Department of Engineering Science
University of Tehran
Tehran, Iran
m.esfandiarpour@ut.ac.ir

Seyed Mani Mirshabani
Department of Engineering Science
University of Tehran
Tehran, Iran
s.manimirshabani@ut.ac.ir

Ehsan Maani Miandoab
Department of Engineering Science
University of Tehran
Tehran, Iran
e.maani@ut.ac.ir

Abstract— This paper presents a comparative analysis of four different ball detection algorithms that can be utilized in a wide range of sports robots. The study compares the performance of these algorithms in terms of accuracy, robustness, noise sensitivity, and computational time consumption. Studied algorithms are Color detection, Hough Circle, Frame differencing, and YOLOv8. Additionally, to overcome the algorithms' drawbacks, two different combinations of algorithms are proposed. The first approach is a combination of color detection and Hough circle, and the other one is a combination of color detection and YOLO. Then, these two hybrid methods are compared to previous algorithms and it is revealed that proposed algorithms are more accurate and efficient than well-known object detection algorithms. Employing these algorithms will lead to more accuracy in sports robots.

Keywords—Ball detection, Yolo, circle detection, Hough circle, Color detection, frame differencing

Code: mvip1090

Cultural-Aware AI Model for Emotion Recognition

Mehrdad Baradaran
Department of Computer and Data Sciences,
Shahid Beheshti University,
Tehran, Iran
meh.baradaran@mail.sbu.ac.ir

Payam Zohari
Department of Computer Engineering,
Khaje Nasir University Of Technology,
Tehran, Iran
m.zohari@email.kntu.ac.ir

Abtin Mahyar
High Performance Computing Laboratory,
School of Computer Science,
Institute for Research in Fundamental Sciences,
Tehran, Iran
abtinmahyar@gmail.com

Hossein Motamednia
High Performance Computing Laboratory,
School of Computer Science,
Institute for Research in Fundamental Sciences,
Tehran, Iran
h.motamednia@ipm.ir

Dara Rahmati
Faculty of Computer Science and Engineering,
Shahid Beheshti University,
Tehran, Iran
d rahmati@sbu.ac.ir

Saeid Gorgin
Department of Computer Engineering,
Chosun University, Gwangju, 61452,
South Korea
gorgin@irost.ir

Abstract—Emotion AI is a research domain that aims to understand human emotions from visual or textual data. However, existing methods often ignore the influence of cultural diversity on emotional interpretation. In this paper, we propose a multimodal deep learning model that integrates cultural awareness into emotion recognition. Our model uses images as the primary data source and comments from individuals across different regions as the secondary data source. Our results show that our model achieves robust performance across various scenarios. Our contribution is to introduce a novel fusion approach that bridges cultural gaps and fosters a more nuanced understanding of emotions. Due to the best of our knowledge, few works are using this approach, for Emotion AI, combining different types of data sources and models. We evaluate our model on the ArtELingo dataset, which contains image-comment pairs with Chinese, Arabic, and English annotations. The experimental results in the evaluation phase demonstrate an impressive 80% recognition accuracy for the model that merges image-text features.

Index Terms—Emotion Recognition, DeepFeature Extraction, Image Representation, Text Representation

Code: mvip1091

The Effect of Uncertainty on No-Reference Image Quality Assessment

Mohammadreza Raei
Department of Electrical and Computer Engineering
Faculty of Engineering
Kharazmi University
Tehran, Iran
Email: mraei@alumni.khu.ac.ir

Azadeh Mansouri
Department of Electrical and Computer Engineering
Faculty of Engineering
Kharazmi University
Tehran, Iran
Email: a_mansouri@khu.ac.ir

Abstract— No-reference Image Quality Assessment (NR-IQA) attained acceptable results through deep learning models. However, the overfitting, caused by complex deep models and insufficient labeled datasets, has become a primary challenge for the research community. Addressing this issue, various strategies such as data augmentation, transfer learning, and weakly supervised learning have been investigated. This paper introduces an approach, suggesting the use of a probability distribution instead of a rigid target to mitigate overconfidence issues. The proposed label uncertainty can provide acceptable results specially in terms of cross dataset validation.

Index Terms— image quality assessment, deep learning, convolutional neural network, no-reference, weakly supervised learning

Code: mvip1093

Human Identification via Dental Panoramic Images Using Simple Features

Mahdi Hariri
Department of Electrical and Computer Engineering
University of Zanjan
Zanjan, Iran
Ma.hariri@znu.ac.ir

Behnaz Tutunchi
Department of Computer Engineering
Islamic Azad University
Zanjan Branch
Zanjan, Iran
Behnaz.Tutunchi@gmail.com

In mass disasters, quick identification of the victims is important even in rescuing the survivors. Teeth are one of the hard parts of the human body, which are more resistant to decay and destruction than other biometric modalities, and due to the variety of numbers, types, and shapes, they are considered a suitable option for identification. Due to the lack of Panoramic dental images from people before their death, the number of database images for training today's modern identification systems is very limited which reduces the accuracy of recognition. In this article, three relatively time-robust features, the number of teeth, the number and position of restored teeth, and the interdental distances are used for identification. For the independence of identification to the variety of processes and the angle of photographs, we use spatial histograms of empty tooth areas as feature vectors. By comparing the Euclidean distance, the images of the data set can be recovered with a similarity of more than 85% to the input image, which is an acceptable accuracy for the simple and easy access of the features. Considering the expert's role in the final decision, this system helps facilitate and accelerate identification.

Keywords—Identification, Panoramic, Dental Images, Histogram, Simple Features.

Code: mvip1094

Brain Tumor Segmentation in MRI Images using Deformable and Dilated Convolutions

Nasim Amini
School of Computer Engineering
Iran University of Science and Technology
Tehran, Iran
nasim.7394@hmail.com

Mohsen Soryani
School of Computer Engineering
Iran University of Science and Technology
Tehran, Iran
soryani@iust.ac.ir

Mohammad Reza Mohammadi
School of Computer Engineering
Iran University of Science and Technology
Tehran, Iran
mrmohammad@iust.ac.ir

Abstract— Segmentation of brain tumor images is an important issue in medical image processing and can help surgeons to accurately assess the tumor area. Since tumors vary in shape, size, and location from patient to patient, segmenting brain tumors is a challenge. In addition, small tumors are more difficult to segment than larger ones. In this paper, we present a method based on deep convolutional networks to improve the segmentation accuracy of brain tumors, especially small tumors in MRI images. In this method we have increased the accuracy of tumor segmentation by adding a module to UNet model. The proposed module uses deformable and dilated convolutions, which provide more spatial information to the network and thus increase the accuracy of tumor segmentation. The results show that our method is able to achieve a Dice of 0.8877 in the whole tumor section. For the core and enhancing tumor sections, we were able to achieve Dice values of 0.8683 and 0.8176, respectively.

Keywords—Segmentation, Deep Learning, Machine Learning, tumor, MRI, Dilated Convolution, Deformable Convolution, UNet

mvip1028

Rethinking RAFT for Efficient Optical Flow

Navid Eslami, Farnoosh Arefi, Amir M. Mansourian, Shohreh Kasaei Department of Computer Engineering
Sharif University of Technology Tehran, Iran Email:

{navid.eslami, far.arefi, amir.mansurian, kasaei@sharif.edu}

Abstract—Despite significant progress in deep learning-based optical flow methods, accurately estimating large displacements and repetitive patterns remains a challenge. The limitations of local features and similarity search patterns used in these algorithms contribute to this issue. Additionally, some existing methods suffer from slow runtime and excessive graphic memory consumption. To address these problems, this paper proposes a novel approach based on the RAFT framework. The proposed Attention-based Feature Localization (AFL) approach incorporates the attention mechanism to handle global feature extraction and address repetitive patterns. It introduces an operator for matching pixels with corresponding counterparts in the second frame and assigning accurate flow values. Furthermore, an Amorphous Lookup Operator (ALO) is proposed to enhance convergence speed and improve RAFT's ability to handle large displacements by reducing data redundancy in its search operator and expanding the search space for similarity extraction. The proposed method, Efficient RAFT (Ef-RAFT), achieves significant improvements of 10% on the Sintel dataset and 5% on the KITTI dataset over RAFT. Remarkably, these enhancements are attained with a modest 33% reduction in speed and a mere 13% increase in memory usage. The code is available at: https://github.com/n3slami/Ef-RAFT

Index Terms—Optical Flow, Large Displacement, Repetitive Patterns, Attention Mechanism, Deep Neural Networks

Code: mvip1029

A Novel Loss Function Based on Clustering Quality Criteria in Spatio-Temporal Clustering

Farnoosh Arefi, Vida Ramezanian, Shohreh Kasaei Department of Computer Engineering Sharif University of Technology Tehran, Iran Email: {far.arefi, vidaramezanian, kasaei@sharif.edu }

Abstract—Video instance segmentation has diverse applications in the autonomous vehicle industry, image surveillance systems, production lines, and medical video analysis. There are two approaches, top-down and bottom-up, to address the task of instance segmentation in video. Top-down methods heavily rely on image-level segmentation and have separate processes for detection and tracking. They show a strong dependency on image-level segmentation. Bottom-up approaches leverage both spatial and temporal information simultaneously, aiming for a gradual transition from pixel-level features to spatiotemporal instances. This paper introduces a novel method to improve the performance of video instance segmentation based on the bottom-up approach. In this method, by utilizing the silhouette metric to assess clustering quality and introducing the central distance metric in loss functions, the values of embedding vectors are improved, leading to the generation of more distinct clusters in space and time. Experimental results demonstrate that this method achieved an approximately 2% improvement compared to the baseline method.

Index Terms—Video Instance Segmentation, Object Tracking, Spatio-temproal clustering

Code:mvip1030

Enhancing Weakly Supervised Semantic Segmentation through Patch-Based Refinement

Narges Javid Tajrishi Department of Computer Engineering Sharif University of Technology Tehran, Iran narges.javid1377@sharif.edu

Sepehr Amini Afshar Department of Computer Engineering Sharif University of Technology Tehran, Iran sepehr.amini@sharif.edu

Shohreh Kasaei, IEEE, Senior Member Department of Computer Engineering Sharif University of Technology Tehran, Iran kasaei@sharif.edu

Abstract—Weakly-Supervised Semantic Segmentation (WSSS) with image-level labels, commonly uses Class Activation Maps (CAM) to generate pseudo-labels. However, Convolutional Neural Networks (CNNs), with their limited local receptive field, often struggle to identify entire object regions. Recently, the Vision Transformer (ViT) architecture has been employed instead of CNNs to capture long-range feature dependencies, by using the self-attention mechanism. Despite its advantages, ViT tends to overlook local feature details, leading to attention maps with low quality and unclear object details. This paper introduces a novel method to enhance the local details in attention maps by leveraging local patches. These local patches are selected from regions that are more likely to contain the desired objects. By effectively utilizing these local patches during the training and generation stages, the model yields more detailed attention maps. Extensive experiments were conducted on the PASCAL VOC 2012 benchmark dataset to demonstrate the efficacy of the proposed approach. The results show significant improvements (+2.6% mIoU) with minimal computational overhead, underscoring the potential of the proposed method in the field of WeaklySupervised Semantic Segmentation.

Index Terms—Weakly-supervised Semantic Segmentation, Image Classification, Deep Learning, Vision Transformer.

Code:mvip1031

Unsupervised Classification of Remotely Sensed High resolution Images using RP-CNN

Maedeh Fallahreyhani Image processing and InformationAnalysis Lab, Faculty of Electrical, Tarbiat Modares University, Tehran, Iran, m.fallahreyhani@modares.ac.ir

Hassan Ghassemian Image processing and InformationAnalysis Lab, Faculty of Electrical, Tarbiat Modares University, Tehran, Iran, ghassemi@modares.ac.ir

Maryam Imani Image processing and InformationAnalysis Lab, Faculty of Electrical, Tarbiat Modares University, Tehran, Iran, maryam.imani@modares.ac.ir

Abstract—In recent years, various deep learning frameworks have been developed for the classification of remotely sensed images. However, the network models proposed in these frameworks exhibit high complexity and do not yield high classification accuracy when applied to unlabeled scenarios. This paper introduces a Multi spectral image (MSI) classification approach that combines the random patches network with self-supervised branch (RPSS) to extract informative deep features. The proposed method involves convolving image bands with random patches to obtain multi-level deep features. Subsequently, we use panchromatic image (PAN) to extract spatial features. The MS spectral features, the derived RPSS features and spatial features then merged to classify the MSI using a support vector machine (SVM) classifier. The experimental results on real remotely sensed images have been presented.

Keywords— Deep learning, Multi spectral data, PCA, Unsupervised classification.

Code : mvip-1057

تشخیص عارضه برگ درختان زرشک با استفاده از شبکه های عصبی کانولوشنی

جلال الدین زارعی، محمدحسین خسروی

دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران

zarei.jalaluddin@birjand.ac.ir

mohokhosravi@birjand.ac.ir

چکیده ــ کارشناسان کشاورزی سعی میکنند تا دفع آفات و بیماری ها را در کمترین زمان ممکن انجام دهند؛ اما محدودیتهایی نظیر کمبود نیروی انسانی، ضعف بینایی، عدم دانش کافی و همینطورمحدودیت های قرنطینه ای در انتقال آفات و بیماری ها به آزمایشگاه میتوانند دلیل خوبی برای استفاده از فناوری دیجیتال در تشخیص آفات و بیماری ها و در نهایت دفع آنها باشد. یکی از راه حل های موجود در این زمینه استفاده از شبکه های عصبی کانولوشنی است. در این مقاله یک مجموعه داده جدید از تصاویر برگ های سالم درخت زرشک، زنگ زرشک و شبپره پارانشیم خوار برگ زرشک را با نام مجموعه داده زرشک معرفی و سپس عملکرد تعدادی از شبکه های عصبی کانولوشنی معروف با استفاده از الگوریتم بهینه ساز گرادیان کاهشی Adam را روی این مجموعه داده بررسی میکنیم.

کلیدواژه‌ــ بهینه سازAdam ، زنگ زرشک، شبپره پارانشیم خوار، شبکه های عصبی کانولوشنی.

Code : mvip-1058

MaskChanger: A Transformer-Based Model Tailoring Change Detection with Mask Classification

Mohammad Ebrahimzadeh
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
mebrahimzadeh@ce.sharif.edu

Mohammad Taghi Manzuri
Department of Computer Engineering
Sharif University of Technology
Tehran, Iran
manzuri@sharif.edu

Abstract — Change detection in multi-temporal remote sensing data enables crucial urban analysis and environmental monitoring applications. However, complex factors like illumination variance and occlusion make robust automated change interpretation challenging. We propose MaskChanger - a novel deep learning paradigm tailored for satellite image change detection. Our method adapts the segmentation-specialized Mask2Former architecture by incorporating Siamese networks to extract features separately from bi-temporal images, while retaining the original mask transformer decoder. To our knowledge, this is the first study in which change detection is converted from the existing per-pixel classification approach into a mask classification approach. Evaluated on the LEVIR-CD benchmark of over 600 very high-resolution image pairs exhibiting real-world rural and urban changes, MaskChanger achieves F1-Score of 91.96%, outperforming prior transformer-based change detection approaches.

Keywords— Change Detection, Remote Sensing Image, Transformer

Code : mvip1059

Convergence of Deep Learning and Edge Computing using Model Optimization

Peyman Babaei
Computer Department
West Tehran Branch, Islamic Azad University
Tehran, IRAN
Peyman.Babaei@IAU.ac.ir

Abstract— Edge systems are undergoing a groundbreaking computing evolution to support artificial intelligence, deep learning, and complex computational algorithms. Using cloud servers to perform deep learning model inference poses challenges such as response delays, increased communication costs, and data privacy concerns. Therefore, significant efforts have been made to push the processing of deep learning models to edge systems, which has led to the creation of edge intelligence as the intersection of learning and edge computing. Learning models, especially deep convolutional neural networks, have made significant achievements in machine vision, which provide high accuracy and predictability by spending computing power and memory. If these models are optimized and deployed on edge systems, there will be a revolution in the applications of edge systems in real time. In this paper, by using optimization techniques such as quantization, weight pruning, and weight clustering, the possibility of deploying a typical convolutional neural network model on edge systems that have limited computing resources and memory is investigated. The results show that by using a collaborative algorithm, despite the slight decrease in the accuracy of the model, it is possible to achieve a small-sized model that can even be deployed on microcontrollers.

Keywords— Model optimization, Edge computing, Deep learning.

Code : mvip-1061

ارائه الگوریتم حریصانه در به دست آوردن فضای ویژگی مناسب در ایجاد گزارش برای تصاویر پزشکی

دکتر محمد احسان بصیری ،
دانشکده فنی و مهندسی، دانشگاه شهرکرد ،
basiri@sku.ac.ir

دکتر رضا روحانی ،
دانشکده فنی و مهندسی، دانشگاه شهرکرد ،
rrohani@sku.ac.ir

راضیه طاهری ،
دانشکده فنی و مهندسی، دانشگاه شهرکرد ،
Raziehtaheri47@gmail.com

چکیده ــ با پیشرفت روزافزون علم، محققان موفق شدند. با ترکیب دو زمینه ی بینایی کامپیوتر و پردازش زبان طبیعی به تولید گزارش های پزشکی دست یابند. تفسیر تصاویر پزشکی به زبان طبیعی یکی از زمینه های نوظهور هوش مصنوعی است که باوجود تحقیقاتی که در این زمینه صورت گرفته همچنان یکی از زمینه های پرچالش و فعال پژوهشی محسوب می گردد. گزارشهای اولیه هرچند، حاوی نقص های همچون نامرتبط بودن گزارش های تولیدشده با تصاویر بودند اما زیربنای خوبی برای کمک به پزشکان در امر خطیر درمان شدند به حدی که ما را بر این داشت گامی در راستای رفع محدودیت های گزارش نویسی تصاویر پزشکی برداریم. در این پژوهش سعی بر این شد که بتوانیم فضایی ویژگی مناسب تری در روند گزارش نویسی برای تصاویر پزشکی قفسه سینه Ray - X Chest داشته باشیم و گزارش های را تولید کنیم که ازلحاظ محتوا کامل تر و طولانی یتر باشند. در فرایند به دست آوردن فضایی ویژگی مناسب برای تولید این گزارش ها از روشهای یادگیری عمیق همچون سازوکار توجه، چارچوب رمزگذارــ رمزگشا و سپس اعمال الگوریتم جستجوی حریصانه در تولید گزارشها بهره برده ایم. آنگاه از معیارهای اندازهگیری زبان طبیعی مثل ROUGE,BLEUو METEOR استفاده کرده ایم و دریافتهایم که مدل پیشنهادی ما نسبت به کارهای قبلی قویتر عمل کرده است و توانسته گزارش های مفصل تری تولید کند.

کلیدواژهــ جستجوی حریصانه ، چارچوب رمزگذارــ رمزگشا، سازوکار توجه، شبکه عصبی پیچشی.

Code: mvip-1064

بازسازي منحنی مجموعه نقاط نمونه مرزي در صفحه با استفاده از مثلث بندي دلونی و ترکیب محدودیت طول اضلاع و درجه رئوس

فرناز شیخی ،
دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی
،
hanaie.sama@email.kntu.ac.ir

سما حنایی ،
دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی،
Sheikh.f@kntu.ac.ir

چکیده – بازسازي منحنی یکی از مسائل چالش برانگیز در هندسه محاسباتی می باشد که داراي کاربردهاي فراوان در بینایی ماشین و گرافیک کامپیوتري است. به ازاي مجموعه نقاط ورودي که از مرز یک شکل در صفحه نمونه برداري شده اند، مسئله بازسازي منحنی تلاش می نماید تا شکل اولیه را مشابه درك بصري انسان بازسازي نماید.ما در این مقاله بر بازسازي منحنی هاي باز و بسته، داراي حفره، چند مؤلفه اي و داراي گوشه هاي تیز تمرکز داشته و الگوریتمی ارائه می دهیم که مشکل بازسازي منحنی هاي باز را در الگوریتم هاي پیشین برطرف نموده و در بازسازي حفره ها و گوشه هاي تیز، دقیق تر از الگوریتم هاي پیشین عمل نماید. الگوریتم ارائه شده بر پایه ي مثلث بندي دلونی نقاط بنا شده و از محدودیت طول اضلاع و درجه رئوس براي اصلاح گراف دلونی بهره می برد. به ازاي مجموعه نمونه مرزي n نقطه در صفحه، الگوریتم پیشنهادي داراي o(nlogn) مرتبه زمانی می باشد.

کلید واژه– بازسازي شکل، بازسازي منحنی، حفره، مثلثبندي دلونی، مرز بیرونی، مرز داخلی، منحنی باز، منحنی بسته، هندسه محاسباتی

Code : mvip-1070

NEM: Nested Ensemble Model for scene recognition

Amir Hossein Saleknia, Elmira Bagheri, Amir Hossein Barshooi, Ahmad Ayatollahi School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran amir.h.saleknia@gmail.com, elmira_bagheri@alumni.iust.ac.ir, a_barshooi@alumni.iust.ac.ir, ayatollahi@iust.ac.ir

Abstract— Scene recognition predicts the category of the scene in which an image is taken. With ensemble learning as its basis, we propose a novel scene recognition method called the Nested Ensemble Model (NEM). As a first step, we employ ensemble learning to construct HybridNets that use object and scene information simultaneously. To take advantage of different neural network architecture paradigms, we developed HybridNet-Transformer, HybridNet-ConvNeXt, and HybridNetCNN. Our NEM is then constructed by fusing these models together based on ensemble learning. By doing this, NEM not only benefits from object and scene knowledge together, but also from CNNs' locality and global modeling of Transformers and ConvNeXts. On MIT-67 and Scene-15, two of the most popular scene recognition datasets, NEM achieved to the accuracy of 89.30% and 96.80%, respectively. These results indicate that this approach is superior to existing methods.

Keywords— Scene Recognition; Ensemble Learning; Transformers; ConvNeXts; CNNs

Code : mvip-1072

Residual Convolutional Neural Network With Autoencoder Based Attention For PoSAR Image Classification

Maryam Imani

Faculty of Electrical and Computer Engineering

Tarbiat Modares University

Tehran, Iran

Email: maryam.imani@modares.ac.ir

Abstract— The polarimetric synthetic aperture radar (PolSAR) images contain fine characteristics and abstract spatial features, which attention to them can improve the classification accuracy. In this work, the residual convolutional neural network with autoencoder based attention (RCNN-AA) is proposed for PolSAR image classification. The scaled difference of the convolutional autoencoder with the original input patch is used as the weight, which contains information about the fine spatial features. Multiplication of this normalized difference in the input patch provides the attention feature maps that can be concatenated with the original input and used as input of the RCNN. An ablation study is done, and also, the proposed RCNN-AA model is compared to some deep learning based models. The results show preference of the RCNN-AA with respect to the competitors.

Keywords—autoencoder, convolutional neutral network (CNN), residual learning, PolSAR classification.

Code: mvip-1073

The role of involution in lightweight super resolution

Akram Khatami-Rizi Cyberspace Research Institute Shahid Beheshti University Tehran Iran
a.khatamirizi@mail.sbu.ac.ir

Ahmad Mahmoudi-Aznaveh Cyberspace Research Institute Shahid Beheshti University Tehran Iran
a_mahmoudi@sbu.ac.ir

Abstract— In recent years, single-image super resolution (SISR) methods using deep convolutional neural networks (CNN) have achieved impressive results. Deep networks learn a complex nonlinear mapping between low-resolution (LR) image patches and high-resolution (HR) versions in a network. However, the enhanced performance of many CNN-based models often comes at the expense of increased memory consumption, making implementation challenging on resource-constrained devices. To minimize the number of parameters, an effective and efficient operator called Involution has been employed, which provides better performance at a lower cost compared to its convolutionbased counterparts. While Involution has previously found success in Image Classification, Object Detection, and Image Segmentation, its application in single-image super resolution is less explored. Information multi-distillation network (IMDN) is identified as a lightweight model providing remarkable performance in SISR. In this work, we propose a strategy that combines involution and convolution in the feature extraction blocks of IMDN. The proposed approach improves network performance with fewer parameters. More specifically, we proposed to use involution more often than convolution as the network gets deeper. Unlike many networks, the blocks of ours do not have the same structure. Our proposed lightweight model has the same performance as the IMDN network with 490k parameters, while has undergone fewer training epochs. The code is available at https://github.com/akramkhatami/IMDB_DeepInvo.

Keywords— Super resolution, Lightweight, Information distillation, Involution

Code : mvip-1074

Comparison of Different Pre-trained Models for Automatic Medical Report Generation

Saeed Niksaz Computer Engineering Department, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran saeedniksaz@eng.uk.ac.ir

Fahimeh Ghasemian Computer Engineering Department, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran ghasemianfahime@uk.ac.ir

Abstract— Automatic generation of text reports explores inter-modal connections between language and visual features. The CNN-RNN architecture is the most common approach for this task, but due to the complexity of medical images, its performance is not sufficiently reliable, and generating detailed reports for medical images remains a challenging problem. In this paper, we attempt to enhance a CNN-RNN architecture and examine the effects of different techniques for feature extraction. Our results on the Indiana University Chest X-ray Collection show that the combination of ResNeXt and BioBERT achieves the best results for this task.

Keywords—Image processing; Medical Report Generation; Deep Learning; Medical Image

Code : mvip-1075

Improving the Efficiency of Complex Convolution Networks

Maryam Sabokpa School of Computer Engineering Iran University of Science and Technology Tehran, Iran
maryam_sabokpa@comp.iust.ac.ir

Dr. Naser Mozayani School of Computer Engineering Iran University of Science and Technology Tehran, Iran
mozayani@iust.ac.ir

Dr. Morteza Garshasbi School of Mathematics & Computer Science Iran University of Science and Technology Tehran, Iran
m_garshasbi@iust.ac.ir

Abstract— Complex data is commonly encountered in several domains of signal and image processing, where two important components, namely the magnitude and phase of the complex data, convey valuable information. This paper introduces novel complex convolution and activation functions aimed at enhancing the efficiency of the complex convolution network. The suggested complex convolution function incorporates complex weights and exhibits equivariance to complex multiplication. Furthermore, a novel activation function is introduced that shows equivariance to complex multiplication, hence enhancing the efficiency of the complex convolution network.

Keywords— complex neural networks, equivariance, invariance, manifold

Code : mvip-1077

Enhancing Multi-view Mammography Image Classification: By using Breast Region Extraction Method and Statistical Features

Neda Shirany Bidabadi a,* Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran.
n.shirani@alumni.iut.ac.ir

Elham Mahmoudzadeh b Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran.
mamoudzadeh@iut.ac.ir.

Abstract— Breast cancer is one of the most dangerous diseases among women. Different methods are used to diagnose this cancer that among these, imaging and computer-aided systems are more common. In these systems, one of the most important step is preprocessing and removing unnecessary areas of the images, as well as extracting the chest area. In this paper, we present a method that consists of preprocessing, feature extraction, and using a machine learning classifier. In the preprocessing step, we propose a method to extract the region of interest in both angles of mammography images. The proposed novel method includes applying gamma correction thresholding to the images and obtaining two binary images based on the proposed threshold using the Otsu method. Results show the proposed method successfully removes the chest muscle with 98% accuracy. In the next, for feature extraction phase, we utilize three different methods for extracting features. Finally, by employing an Extra tree model classifier, we classify mammography images into normal and abnormal. By incorporating the block-based feature extraction method, we achieve 98% accuracy in classification. Overall, our approach demonstrates the effectiveness of preprocessing and feature extraction for diagnosing breast cancer using mammography images.

Keywords—Mammography Image, Region of Interest Extraction, Breast Cancer, Statistical Feature Extraction

Code:mvip-1078

دسته بندی گندله های سنگ آهن براساس اندازه با استفاده از یادگیری عمیق

اسما شمسی گوشکی ،
استادیار،
دانشگاه صنعتی سیرجان ،
a.shamsi@Sirjantech.ac.ir

مرضیه صالحی ،
دانشجو کارشناسی ارشد، دانشگاه صنعتی سیرجند ،
salehi.m@stu.Sirjantech.ac.ir

چکیده— در صنعت تولید فولاد ذرات ریز سنگ آهن در یک دیسک گردان بهم می چسبند و گلوله هایی را تشکیل می دهند که به آن ها گندله سبز سنگ آهن می گویند. گندله ها مواد اولیه تشکیل دهنده ی بسیاری از محصولات ، از جمله فولاد هستند و کیفیت آنها در سرعت تولید و عیار محصول نهایی تاثیر مستقیم دارد. یکی از معیار های تعیین کیفیت گندله ها ، اندازه ی آنها است که باید در محدوده خاصی باشند. یکی از روش های موجود برای این منظور، نظارت خودکار با دوربین است که نیازمند استفاده از روش های پردازش تصویر برای تشخیص و دسته بندی گندله های روی دیسک در حال چرخش ، با دقت و سرعت بالا است ، تا بتوان با توجه به اندازه گندله ها ، سرعت چرخش دیسک را تنظیم کرد. در این زمینه از روش های یادگیری عمیق مختلف در مقالات استفاده شده است. در این مقاله از شبکه عصبی کانولوشنی YOLOv8 برای این منظور استفاده شده است که به دقت 95 الی 100 درصد برای کلاس های مختلف رسیده است. روش مذکور با مدل های VGG16، Resnet50 ، Mobilenet و ترکیبی از آن ها که برای این مسئله در مراجع پیشنهاد شده است، مقایسه شده و عملکرد بهتری را نشان داده است.

کلیدواژه– پردازش تصویر، دسته بندی، گندله سنگ آهن، یادگیری عمیق.

Code:mvip-1079

Temporal Relations of Informative Frames in Action Recognition

Alireza Rahnama Department of Engineering Faculty of Electrical and Computer Engineering Kharazmi University Tehran, Iran Email: Alireza.rahnama@khu.ac.ir

Azadeh Mansouri Department of Engineering Faculty of Electrical and Computer Engineering Kharazmi University Tehran, Iran Email: A mansouri@khu.ac.ir

Abstract—This paper presents a simple approach leveraging temporal learning on informative frames for action recognition. We propose a training-free simple adaptive frame selection scenario employing just the similarity technique in a temporal window. The proposed frame selection method provides an appropriate strategy to capture informative frames and provide meaningful features. Moreover, we use transfer learning for spatial feature extraction and employ LSTM and GRU for temporal modeling. Our method is evaluated on two popular datasets, UCF11 and KTH, and it demonstrates acceptable results.

Keywords—Action recognition, Frame selection, deep temporal modeling, transfer learning, Spatial-Temporal features

Code:mvip-1082

Estimation of the Biological Age of the Human Brain Using Multitask Self-Supervised Learning

Zahrasadat Sajjadi School of Computer Engineering Iran University of Science and Technology Tehran, Iran
sajjadi_z@comp.iust.ac.ir

Soheil Hamzebeigi School of Computer Engineering Iran University of Science and Technology Tehran, Iran
soheil_hamzebeigi@comp.iust.ac.ir

Mohsen Soryani School of Computer Engineering Iran University of Science and Technology Tehran, Iran
soryani@iust.ac.ir

Abstract—As the global population ages, reliable methods for assessing brain health and its changes have become crucial. Predicting the biological age of the brain is a promising biomarker for evaluating overall health, and machine learning tools have made this possible. However, efficient training strategies are required to achieve high performance in predicting human brain age. We propose a self-supervised multitask model for predicting rotation class and brain age from the ADNI dataset's 2D magnetic resonance imaging (MRI). This model outperforms single-task models and improves brain age prediction by providing more general features through the self-supervised task of rotation class prediction. We have also examined how self-supervised pretraining implemented by DINO framework (knowledge distillation with no labels) affect our network and discover that this pretraining weights has a considerable impact. We indicated that our proposed model, performed equally well as state-of-the-art models, with a Mean Absolute Error (MAE) of 3.27 years.

Keywords—biological brain age, deep learning, DINO framework, multitask model, pretraining, self-supervised learning, transfer learning

Code:mvip-1083

Analyzing different loss functions for Single Image Super Resolution

Shakiba Shojaei
Shahid Beheshti University
Cyberspace Research Institute
Tehran, Iran
sh.shojaei@sbu.ac.ir

Ahmad Mahmoudi-Aznaveh
Shahid Beheshti University
Cyberspace Research Institute
Tehran, Iran
a_mahmoudi@sbu.ac.ir

Abstract— Single Image Super Resolution (SISR), which aims to recover a high-resolution (HR) image from a lowresolution (LR) image, is an ill-posed problem. CNNs have been used in low-level vision tasks such as SR, and inspired by impressive results in high-level tasks. By choosing the proper structure, the methods can be improved significantly. In this case, selecting an appropriate loss function is essential for any deep learning task, especially in SISR. The exploited loss function impacts the quality of the images produced by the SISR algorithms. Some loss functions can make the output image look blurred or unnatural, which goes against the purpose of SR. To ensure that the output image retains the content of the original photo while also improving the structure and texture, it is essential to choose a loss that is well suited for the task. In this paper, various loss functions for SISR are reviewed. Then, we present an overall analysis of loss functions for SISR based on our exploration.

Keywords— super-resolution, loss function, image quality assessment, feature extraction

Code:mvip-1084

Self-Supervised Contrastive Learning In Spiking Neural Networks

Yeganeh Bahari and Saeed Reza Kheradpisheh*

Departement of Computer and Data Sciences, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

Emails: bahary.yeganeh@gmail.com, s kheradpisheh@sbu.ac.ir * Corresponding author

Abstract—Spiking neural networks (SNNs), inspired by the biological neural processing of the brain, are vastly growing due to their higher potential to handle spatiotemporal patterns with lower energy consumption, especially, if implemented on neuromorphic devices. In this study, we propose self-supervised contrastive learning (SSL) for SNNs to learn informative latent representations from a large set of unlabeled data. The proposed SSL pre-trained SNN is then fine-tuned on a small set of labeled samples of a downstream supervised task. To evaluate the proposed method, we trained convolutional SNNs using SSL on MNIST and CIFAR10 datasets with 80% of images as unlabeled samples, then fine-tuned the networks on the remaining 20% images. The proposed SSL-based SNNs could reach 94.23% and 62.24% recognition accuracies on testing sets of MNIST and CIFAR10, respectively.

Index Terms—Spiking Neural Networks, Self-Supervised Learning, SimCLR, Data Augmentation.

Code:mvip-1085

Marginal Contrastive Loss: A Step Forward for Forward-Forward

Hossein Aghagolzadeh Faculty of Electrical and Computer Engineering Babol Noshirvani University of Technology Babol, Iran ho.golzadeh@gmail.com

Mehdi Ezoji Faculty of Electrical and Computer Engineering Babol Noshirvani University of Technology Babol, Iran m.ezoji@nit.ac.ir

Abstract— The Forward-Forward algorithm was developed to increase the resemblance of artificial neural network training processes to those occurring in the brain, in contrast to the backpropagation algorithm, which has been shown to have less similarity to brain processes. While Forward-Forward is a fascinating and novel idea, it significantly differs in performance from backpropagation. Forward-Forward strives to achieve this similarity by updating each layer independently of the others, through the introduction of a loss function that facilitates the separability of data from different classes. This inherent nature of creating discrimination between data of different classes inspired us to take advantages of contrastive learning to improve Forward-Forward performance. We modified the contrastive loss to be used in Forward-Forward, and our experimental results show that the proposed method improves the model accuracy and increases the convergence speed by more than 20 times.

Keywords—Forward-Forward, Backpropagation, Contrastive Learning, Brain-inspired Computing, Neural Network Training

Code:mvip-1086

An End-to-End Model for Chicken Detection in a Cluttered Environment

Komeil Sadeghi Chemme University of Science and Technology of Behshahr, Mazandaran, Iran
komilsadeghicemme@Gmail.com

Reza Javanmard Alitappeh University of Science and Technology of Behshahr, Mazandaran, Iran
Rezajavanmard64@Gmail.com

Abstract—The livestock industry requires live and online monitoring of animals for their growth and development. With the increasing demand for meat and chicken worldwide, the evolution of the livestock and poultry industry, and the increase in their area, it has become challenging for humans to manage the task due to human error and the increasing volume of work. However, with the advancements in technology, including the use of artificial intelligence, learning models, and closedcircuit cameras, most of the work can be entrusted to machines, thus avoiding human error. By installing and connecting this system globally, it can count the number of all animals at any given moment and provide us with the number of each animal along with its location. In this paper, we developed a model comprises a segmentation module preparing input data for a Yolov8 based detector model. This integrated end-to-end proposal is used to count and track chickens for analysing their behavior. Experimental results clarify the performance of the new proposal in the sense of accuracy in comparison to previous approaches with 8% improvement. This article aims to educate a learning model based on yolov8, which can be used to develop an integrated and automatic system for counting and evaluating the welfare of animals in the future.

Index Terms—Chicken detection, Object detection, Deep learning, Segmentation

Code:mvip-1087

Enhancing the Generalization of Synthetic Image Detection Models through the Exploration of Features in Deep Detection Models

Alireza Hajabdollah Javaheri
Cyberspace research Institute
Shahid Beheshti University
Tehran, Iran
alireza.javaheri1376@gmail.com

Hossein Motamednia High Performance Computing Laboratory, School of Computer Science, Institute for Research in Fundamental Sciences, Tehran, Iran h.motamednia@ipm.ir

Ahmad Mahmoudi-Azanveh
Cyberspace research Institute
Shahid Beheshti University
Tehran, Iran a
mahmoudi@sbu.ac.ir

Abstract—One of the major challenges of AI is the misuse of images generated by generative models. Advances in this field have reached a point where distinguishing between real and fake images can be impossible for humans and challenging even for machines. Although significant work has been done on detecting fake images, there is an ongoing competition between content generation and detection methods. However, a significant challenge for detection methods is their limitation to content generated by specific models. This study aims to enhance the generalization of fake image detection methods. Experimental results indicate that modifications made to the base model have contributed to improving its generalizability.

Index Terms—Synthetic image detection, Image forensics, GANs, Diffusion Models, CNN

Code:mvip-1088

Finger-hand Rehabilitation using DNN-based Gesture Recognition of Low-cost Webcam Images

Shayan Mesdaghi Dept of Electrical Engineering University of Guilan Rasht, Iran
mesdaghis@phd.guilan.ac.ir

Reza PR Hasanzadeh Dept of Electrical Engineering University of Guilan Rasht, Iran
hasanzadehpak@guilan.ac.ir

Farrokh Janabi-Sharifi Dept of Mechanical ,Industrial and Mechatronic Engineering, Tornoto Metropolitan University Toronto, Canada
fsharifi@torontomu.ca

Abstract— In this paper, an image-based gesture recognition system has been presented for finger hand rehabilitation using a low-cost camera. Since the goal is to set up a lowcost home rehabilitation system, the analysis of each finger should be easily possible for the user, and the user should be able to follow the information of the improvement process of one's treatment. Hence, first, the models governing the movement angles of the fingers were established, and then some criteria have been developed to evaluate the improvement of the performance of the fingers. Finally, several deep-learning models were initially implemented to extract the hand gesture and model parameters and based on the experimental results, the MediaPipe framework was found suitable due to its precision and robustness to determine the finger angles in low quality images during each exercise.

Keywords— rehabilitation, MediaPipe, hand gesture, pose estimation, deep learning