Digital Object Identifier 10.1109/ACCESS.2020.3035461

EDITORIAL

IEEE ACCESS SPECIAL SECTION EDITORIAL: AI-DRIVEN BIG DATA PROCESSING: THEORY, METHODOLOGY, AND APPLICATIONS

With the rapid development of network infrastructures and personal electronic products, big data generated from the Internet, sensing networks, and other equipment are rapidly growing and have received increasing attention in recent years. Recently, artificial intelligence (AI)-driven big data processing technologies based on pattern recognition, machine learning, and deep learning have been intensively applied to dealing with large-scale heterogeneous data. However, challenges still exist in the development of AI-driven big data processing. In order to meet the existing challenges, it is important to consider how to analyze and process big data in a way that is more effective and reduces costs, how to discover and understand knowledge from the data, and how to generalize and transfer these discoveries into other application fields.

Sixty-six high-quality articles have been accepted by this Special Section after an extensive and rigorous peer-review process.

In modern society, the demand for radio spectrum resources is increasing. As the information carriers of wireless transmission data, radio signals exhibit characteristics of big data in terms of volume, variety, value, and velocity. How to uniformly handle these radio signals and obtain value from them is a challenge that needs to be studied. In the article, ''Big data processing architecture for radio signals empowered by deep learning: Concept, experiment, applications, and challenges,'' by Zheng *et al.*, a big data processing architecture for radio signals is presented and a new approach of end-to-end signal processing based on deep learning is discussed in detail. The radio signal intelligent search engine is used as an example to verify the architecture, and system components and experimental results are introduced. In addition, the applications of the architecture in cognitive radio, spectrum monitoring, and cyberspace security are introduced. Finally, challenges are discussed, such as unified representation of radio signal features, distortionless compression of wideband sampled data, and deep neural networks for radio signals.

The selection of semantic concepts for modal construction and data collection remains an open research issue. It is highly demanding to choose good multimedia concepts with small semantic gaps to facilitate the work of cross-media system developers. However, very little work has been done in this area. The article by Rehman *et al.*, ''A benchmark data set and learning high-level semantic embeddings of multimedia for cross-media retrieval,'' contributes the a new, real-world web image data set for cross-media retrieval called FB5K. The proposed FB5K data set contains the following attributes: 1) 5130 images crawled from Facebook; 2) images that are categorized according to users' feelings; and 3) images independent of text and language rather than using feelings for search. Furthermore, they propose a novel approach through the use of Optical Character Recognition and explicit incorporation of high-level semantic information. They comprehensively compute the performance of four different subspace-learning methods and three modified versions of the Correspondence Auto Encoder, alongside numerous text features and similarity measurements comparing Wikipedia, Flickr30k, and FB5K. To check the characteristics of FB5K, they propose a semantic-based cross-media retrieval method. To accomplish cross-media retrieval, they introduced a new similarity measurement in the embedded space, which significantly improved system performance compared with the conventional Euclidean distance. Their experimental results demonstrated the efficiency of the proposed retrieval method on three different data sets to simplify and improve general image retrieval.

In the field of instant voice communication (IVC) steganalysis, the traditional detecting methods are mainly based on a supervised learning scheme that results in a large amount of complex manual pre-processing training data sets. The accuracy of supervised learning schemes can be easily destroyed by the difference between the distribution of training and testing data sets in the actual voice application. The disadvantages of this method are obvious in the big data environment. In this regard, the article, ''SSLSS: Semi-supervised learning-based steganalysis scheme for instant voice communication network,'' by Tu *et al.*, initially introduces a novel semi-supervised hybrid learning detection model for the IVC network. This provides the progress of manually annotating a training data set that has been removed to solve the problem of complex operations and poor applicability in

classifiers. Therefore, this model has a simpler structure and more extensive detection scope with a huge amount of data. Later, the authors designed a multi-criteria fusion module that can automatically generate the pseudo-label set from testing data set to train the classifier model. Thus, their scheme will not be affected by the distribution shift. In this module, they defined the confidence level and representative level to judge the feature vector for pseudolabeled. Through the experimental analysis, the low bit-rate speech coding steganalysis (G.723.1/G.729/iLBC speech codecs) is analyzed on quantization index modulation that are common codecs in the IVC network. The results show that their method has higher accuracy than the unsupervised method. The proposed approach is less affected and more accurate than the previous supervised methods through the distribution of different training and testing data sets. The experiments also proved that their method can be deployed in different kinds of the IVC codec by considering huge amounts of data sets.

Neural network models are popularly used in Chinese word segmentation tasks. The capsule architecture was proposed recently, which has solved some defects of the convolutional neural network. The article by Li *et al.*, ''Capsules based Chinese word segmentation for ancient Chinese medical books,'' first introduces capsule architecture to Chinese word segmentation. The authors utilize capsules as neural units. Before doing the routing algorithm, they make a sliding capsule window to select the features which are extracted from the primary capsule layer. The sliding capsule window is proposed to adapt to the capsule architecture to the sequence labeling task. The experiment results show that their proposed capsule-based Chinese word segmentation model achieves competitive performance with the previous state-of-the-art methods. Ancient Chinese medical books record a lot of valuable experiences from ancient medical workers. However, research about automatic text analysis on ancient Chinese medical documents is just the beginning. Due to the lack of annotated data for Chinese medicine, the authors develop word segmentation guidelines for ancient Chinese medical documents and select ten genres, 30 ancient Chinese medical books to set up the annotation data set. With the annotated data, the authors develop the segmenter for the ancient Chinese medical texts. Experiments show that the F_1 measures of their model on the two data sets are 94.9% and 81.4% on Chinese Treebank6.0 and Ancient Chinese Medical Books, respectively.

Three-dimensional (3-D) point clouds are important for many applications, including object tracking and 3-D scene reconstruction. Point clouds are usually obtained from laser scanners, but their high-cost impedes the widespread adoption of this technology. The article by Chu *et al.*, ''Generative adversarial network-based method for transforming single RGB image into 3-D point cloud,'' proposes a method to generate a 3-D point cloud corresponding to a single red–green–blue (RGB) image. The method retrieves high-quality 3-D data from two-dimensional (2-D) images captured by conventional cameras, which are generally less expensive. The proposed method comprises two stages. First, a generative adversarial network generates a depth image estimation from a single RGB image. Then, a 3-D point cloud is calculated from the depth image. The estimation relies on the parameters of the depth camera employed to generate the training data. The experimental results verify that the proposed method provides high-quality 3-D point clouds from single 2-D images. Moreover, the method does not require a PC with outstanding computational resources, further reducing implementation costs, as only a moderate-capacity graphics processing unit can efficiently handle the calculations.

Text in natural images contains rich semantics that is often highly relevant to objects or scenes. In the article ''Integrating scene text and visual appearance for fine-grained image classification,'' by Bai *et al.*, the authors focus on the problem of fully exploiting scene text for visual understanding. The main idea is combining word representations and deep visual features in a globally trainable deep convolutional neural network. First, recognized words are obtained by a scene text reading system. Next, they combine the word embedding of recognized words and the deep visual features into a single representation that is optimized by a convolutional neural network for fine-grained image classification. In their framework, the attention mechanism is adopted to compute the relevance between each recognized word and the given image, which further enhances the recognition performance. The authors performed experiments on two data sets: the context data set and the drink bottle data set, which are proposed for fine-grained classification of business places and drink bottles, respectively. The experimental results consistently demonstrate that the proposed method of combining textual and visual cues significantly outperforms classification with only visual representation. Moreover, they have shown that learned representation improves the retrieval performance on drink bottle images by a large margin, making it potentially powerful in product search.

Maturity level-based classification systems play an essential role in the design of tomato harvesting robots. Traditional knowledge-based systems are unable to meet the current production management requirements of precision picking, because they are time-consuming and have low accuracy. The article by Zhang *et al.*, ''Deep learning based improved classification system for designing tomato harvesting robot,'' proposes an improved deep learning-based classification method that improves the accuracy and scalability of tomato ripeness with a small amount of training data. This study is on the relationship between different data set augmentation methods and prediction results of the final classification task. They implement classification systems based on convolutional neural network (CNN), by training and validating the model on different augmented data sets and trying to choose an optimal augmentation method for data sets. The experimental results show an average accuracy of 91.9% with a less than 0.01-s prediction time. Compared to the existing methods, their solution achieves better prediction results both in terms of accuracy and time consumption. Moreover, this

is a versatile method and can be extended to other related fields.

Educational data-mining is an evolving discipline that focuses on the improvement of self-learning and adaptive methods. It is used for finding hidden patterns or intrinsic structures of educational data. In the arena of education, heterogeneous data is being involved and continuously growing in the paradigm of big-data. To extract meaningful information adaptively from big educational data, some specific data mining techniques are needed. The article by Kausar *et al.*, ''Integration of data mining clustering approach in the personalized e-learning system,'' presents a clustering approach to partition students into different groups or clusters based on their learning behavior. Furthermore, a personalized e-learning system architecture is presented, which detects and responds to teaching contents according to the students' learning capabilities. The primary objective includes the discovery of optimal settings, in which the learners can improve their learning capabilities. Moreover, the administration can find essential hidden patterns to bring effective reforms to the existing system. The clustering methods K-Means, K-Medoids, Density-Based Spatial Clustering of Applications with Noise, Agglomerative Hierarchical Cluster Tree and Clustering by Fast Search and Finding of Density Peaks via Heat Diffusion (CFSFDP-HD) are analyzed using educational data mining. It has been observed that more robust results can be achieved by the replacement of existing methods with CFSFDP-HD. The data mining techniques are equally effective in analyzing big data to make education systems vigorous.

In the article ''CarvingNet: Content-guided seam carving using deep convolution neural network,'' by Song *et al.*, the authors propose an improved content-aware image resizing method that uses deep learning. The proposed method is extended from seam carving, which is another image resizing method. Seam carving uses the energy map from an image. It also removes a seam where the energy is at the minimum. They propose a method for creating a deep energy map using an encoder–decoder convolution neural network. A deep energy map preserves important parts or boundaries in an image without distortion. Furthermore, it has the characteristic that the uniform intensity of edges is displayed for all images. Four well-known resizing methods and the proposed method are evaluated in terms of aspect ratio similarity. In such an objective evaluation, the proposed method demonstrates better results than the other four algorithms. The proposed method can reduce the size of an image without damaging the overall structure or losing important information.

Traditional surveillance systems for observing water levels are often complex, costly, and time-consuming. In the article ''Deep learning-based unmanned surveillance systems for observing water levels,'' by Pan *et al.*, the authors develop a low-cost unmanned surveillance system consisting of remote measuring stations and a monitoring center. The system uses a map-based Web service, as well as video cameras, water level analyzers, and wireless communication routers necessary to display real-time water level measurements of rivers and reservoirs on a Web platform. With the aid of a wireless communication router, the water level information is transmitted to a server connected to the Internet via a cellular network. By combining complex water level information from different river basins, the proposed system can be used to forecast and prevent flood disasters. In order to evaluate the proposed system, the authors conduct experiments using three feasible methods, including the difference method, dictionary learning, and deep learning. The experimental results show that the deep learning-based method performs best in terms of accuracy and stability.

Aiming at the shortages of basic flower pollination algorithm (FPA) with slow convergence speed, low search precision, and easy to fall into local optimum, a new adaptive FPA based on opposition-based learning and t-distribution (OTAFPA) is proposed in the article ''Improved flower pollination algorithm and its application in user identification across social networks,'' by Li *et al.* First, the opposition-based learning strategy is utilized to increase the diversity and quality of the initial population. Then, adaptive dynamic switching probability is introduced, which can effectively balance the global and local search according to the current number of iterations. Finally, the t-distribution variation is used to increase the population diversity and to help the algorithm jump out of the local optimum. The simulation experiments on eight classical test functions show that OTAFPA has better global optimization abilities, which improves the convergence speed and the solution accuracy of the algorithm. The OTAFPA also shows superior performance in practical applications of user identification across social networks.

Fog computing is an encouraging technology in the coming generation to pipeline the breach between cloud data centres and the Internet of Things (IoT) devices. Fog computing is not a counterfeit for cloud computing but a persuasive counterpart. It also accredits by utilizing the edge of the network while still rendering the possibility of interacting with the cloud. Nevertheless, the features of fog computing are encountering novel security challenges. The security of end users and/or fog nodes brings about major dilemmas in the implementation of real-life scenarios. Although there are several works investigated in the security challenges, physical layer security (PLS) in fog computing is not investigated in the above. The distinctive and evolving IoT applications necessitate new security regulations, models, and evaluations disseminated at the network edge. Notwithstanding, the achievement of the current cryptographic solutions in the customary way, many aspects, i.e., system imperfections, hacking skills, and augmented attack, has upheld the inexorableness of the detection techniques. In the article ''Security in fog computing: A novel technique to tackle an impersonation attack,'' by Tu *et al.*, they investigate PLS that exploits the properties of channel between end user and fog node to detect impersonation attacks in the fog computing

network. Moreover, it is also challenging to achieve accurate channel constraints between the end user and the fog node. Therefore, the authors propose Q-learning algorithm to attain the optimum value of the test threshold in the impersonation attack. The performance of the proposed scheme validates and guarantees to detect impersonation attacks accurately in fog computing networks.

Relation classification is a crucial ingredient in numerous information-extraction systems and has attracted a great deal of attention in recent years. Traditional approaches largely rely on feature engineering and suffer from the limitations of domain adaption and error propagation. To overcome the abovementioned problems, many deep neural network-based methods have been proposed; however, these methods cannot effectively locate and utilize the relation trigger features. To locate the relation trigger features and make full use of them, the article by Zhang *et al.*, ''Multi-gram CNN-based self-attention model for relation classification,'' proposes a novel multi-gram convolution neural network-based self-attention model with a recurrent neural network framework. The multi-gram conventional neural network attention model can learn the adaptive relational semantics of inputs based on the fact that a relationship can be totally defined by the shortest dependency path between its two entities. With the learned relational semantics, the proposed method can obtain the corresponding importance distribution over input sentences and locate the relation trigger features. For effective information propagation and integration, they utilize a bidirectional gated recurrent unit to encode high-level features during recurrent propagation. The experimental results on two benchmark data sets demonstrate that the proposed model outperforms most of the state-of-the-art models.

The idea of big data has gained extensive attention from governments and academia all over the world. It is especially relevant for the establishment of a smart city environment combining complex heterogeneous data with data analytics and artificial intelligence (AI) technology. Big data is generated from many facilities and sensor networks in smart cities and is often streamed and stored in cloud storage platforms. Ensuring the integrity and subsequent auditability of such big data is essential for the performance of AI-driven data analysis. Recent years have witnessed the emergence of many big data auditing schemes that are often characterized by third party auditors (TPAs). However, the TPA is a centralized entity, which is vulnerable to many security threats from both inside and outside the cloud. To avoid this centralized dependency, the article by Yu *et al.*, ''Decentralized big data auditing for smart city environments leveraging blockchain technology,'' proposes a decentralized big data auditing scheme for smart city environments featuring blockchain capabilities supporting improved reliability and stability without the need for a centralized TPA in auditing schemes. To support this, they have designed an optimized blockchain instantiation and conducted a comprehensive comparison between the existing schemes and the proposed scheme through both theoretical analysis and experimental evaluation. The comparison shows that lower communication and computation costs are incurred with their scheme than with existing schemes.

Sketch-based image retrieval (SBIR) finds natural images according to the features and rules defined by human beings. The retrieval results are generally similar in contour; however, their complete semantic information of the image is missing. From the user's point of view, the same hand-drawn image may represent many different things. Due to the semantic ''one-to-many'' category mapping relationship between the hand-drawn image and the natural image, which is the inherent ambiguity of hand-drawn images. In addition, the user's drawing has many different characteristics, so the retrieval results generally cannot fully match with his intent. For the abovementioned challenges, a personalized SBIR architecture is proposed in the article ''Personalized sketch-based image retrieval by convolutional neural network and deep transfer learning,'' by Qi *et al.* The proposed method includes a deep full convolutional neural network as a general model, and a personalized model using transfer learning to achieve fine-grained image semantic features. On the basis of the pretrained general model and the history of images selected by the user, they construct a personalized model training data set. Moreover, the user history feedback is combined with the current hand-drawn image as the input for the transfer learning model to fine-tune the distribution of features in vector space so that the neural network can learn personalized semantic information. The experiments show that the general model has strong generalization ability with the mean average precision as 0.64 on the Flickr15 K data set. The migration model can realize fine-grained image semantic vector space division, which perfectly satisfies the personalized retrieval requirements by a hand-drawn sketch-based image input.

Recent network research has demonstrated that the performance of convolutional neural networks can be improved by introducing a learning block that captures spatial correlations. In the article ''Multiple feature reweight DenseNet for image classification,'' by Zhang *et al.*, the authors propose a novel multiple feature reweight DenseNet (MFR-DenseNet) architecture. The MFR-DenseNet improves the representation power of the DenseNet by adaptively recalibrating the channel-wise feature responses and explicitly modeling the interdependencies between the features of different convolutional layers. First, in order to perform dynamic channel-wise feature recalibration, they construct the channel feature reweight DenseNet (CFR-DenseNet) by introducing the squeeze-and-excitation module (SEM) to DenseNet. Then, to model the interdependencies between the features of different convolutional layers, they propose the double squeeze-and-excitation module (DSEM) and construct the inter-layer feature reweight DenseNet (ILFR-DenseNet). In the last step, they designed the MFR-DenseNet by combining the CFR-DenseNet and the ILFR-DenseNet with an ensemble learning approach. Their experiments demonstrate the effectiveness of CFR-DenseNet, ILFR-DenseNet, and MFR-DenseNet. Their 100-layer MFR-DenseNet (with 7.1M

parameters) model achieves competitive results on CIFAR-10 and CIFAR-100 data sets, with test errors of 3.57% and 18.27%, respectively, achieving a 4.5% relative improvement on CIFAR-10 and a 5.09% relative improvement on CIFAR-100 over the best result of DenseNet (with 27.2M parameters).

Incentive-based demand response can fully mobilize a variety of demand-side resources to participate in the electricity market, but the uncertainty of user response behavior greatly limits the development of demand response services. The article by Liu *et al.*, "Analysis and accurate prediction of user's response behavior in incentive-based demand response," first constructs an implementation framework for incentive-based demand response and clarifies how a load-serving entity aggregates demand-side resources to participate in the power market business. Then, the characteristics of the user's response behavior are analyzed; it is found that the user's response behavior is variable, and it has a strong correlation on the timeline. Based on this, a prediction method of user response behavior based on long short-term memory (LSTM) is proposed after the analysis of the characteristics of the LSTM algorithm. The proposed prediction method is verified by simulation under the simulation environment setup by TensorFlow. The simulation results show that, compared with the traditional linear or nonlinear regression methods, the proposed method can significantly improve the accuracy of the prediction. At the same time, it is verified by further experiments that the proposed algorithm has good performance in various environments and has strong robustness.

Three-dimensional (3-D) data acquisition and real-time processing are critical issues in artificial vision systems. The developing time-of-flight (TOF) camera as a real-time vision sensor for obtaining depth images has now received wide attention, due to its great potential in many areas, such as 3-D perception, computer vision, robot navigation, human–machine interaction, augmented reality, and so on. The article by Yu and Chen, "Recent advances in 3-D data acquisition and processing by time-of-flight camera," surveys advances in TOF imaging technology mainly from the last decade. They focus only on the recent progress in overcoming limitations such as systematic errors, object boundary ambiguity, multipath errors, phase wrapping, and motion blur, and address the theoretical principles and future research trends as well.

In today's information age, the development of hot events is timely and rapid under the influence of the powerful Internet. Online social media, such as Weibo in China, has played an important role in the process of spreading public opinions and events. Sentiment analysis of social network texts can effectively reflect the development and changes of public opinions. At the same time, prediction and judgment of public opinion development can also play a key role in assisting decision-making and effective management. Therefore, sentiment analysis for hot events in online social media texts and judgment of public opinion development have become popular topics in recent years. At present, research on textual sentiment analysis is mainly aimed at a single text, and there is little-integrated analysis of multiuser and multidocument in unit time for time series. Moreover, most of the existing methods are focused on the information mined from the text itself, while the feature of identity differences and time sequence of different users and texts on social platforms are rarely studied. The article by Li *et al.*, "Time+user dual attention based sentiment prediction for multiple social network texts with time series," works on public opinion texts about some specific events on social network platforms and combines the textual information with sentiment time series to achieve multi-document sentiment prediction. Considering the related features of different social user identities and time series, they propose and implement an effective time+user dual attention mechanism model to analyze and predict the textual information of public opinion. The effectiveness of the proposed model is then verified through experiments on real data from a popular Chinese microblog platform called Sina Weibo.

Pedestrian detection is a key problem for automatic driving, and the results have been improved significantly via deep convolutional networks. However, there is still room to improve the performance of pedestrian detection by carefully dealing with some critical issues. To take advantages of more discriminative information for pedestrian detection, the article by Chen *et al.*, "Deep feature fusion by competitive attention for pedestrian detection," proposes a novel architecture to auto-choose semantic as well as specific information among the feature maps at different levels and integrate valuable information among the feature maps in multiscales. In particular, the proposed architecture consists of feature maps concatenating in different levels and feature maps integrating with multi-scales. Both the operations are equipped with a competitive attention block. The architecture has the ability to obtain more efficient and discriminating features for pedestrian detection. In comparison with the other prevailing models, the proposed architecture provides superior performance. The promising results achieved through experimentation with this architecture achieve a new state-of-the-art on Caltech data set.

The rapid developments in mobile Internet are reshaping our lives and activities. Understanding user behaviors and dynamics in such a large-scale network is essential for better system design, service provisioning, and network management. In the article "Coupled tensor decomposition for user clustering in mobile Internet traffic interaction pattern," by Yu *et al.*, the authors focus on the interaction pattern between mobile users and servers based on traffic flow data. Real traffic flow data is collected from the public network of ISPs by high-performance network traffic monitors. A traffic flow-based heterogeneous information network (TF-HIN) is introduced to represent the traffic interaction pattern, and node correlation characteristics are mined from TF-HIN. Based on the empirical analysis of traffic interaction patterns, the authors propose the coupled flow tensor to represent the relations among the user, server, and time, by incorporat-

ing correlations of user and server as auxiliary information. Two iterative algorithms, i.e., FTD and FTD-NFS, are proposed for coupled flow tensor decomposition and the latent factors are used for user clustering. They evaluate the proposed user clustering algorithms by using benchmark data sets and also analyze the user clustering results from real traffic flow data sets. The numerical experiments show that using coupled flow tensor with auxiliary information provides a novel and scalable user clustering method and improves the clustering accuracy.

Salient object detection is receiving more and more attention from researchers. An accurate saliency map will be useful for subsequent tasks. However, in most saliency maps predicted by existing models, objects regions are very blurred and the edges of objects are irregular. The reason is that hand-crafted features are the main basis for existing traditional methods to predict salient objects, which results in different pixels belonging to the same object often being predicted different saliency scores. In addition, the convolutional neural network (CNN)-based models predict saliency maps at patch scale, which causes the object edges of the output to be fuzzy. In the article ''Convolutional edge constraint-based U-Net for salient object detection,'' by Han *et al.*, the authors attempt to add an edge convolution constraint to a modified U-Net to predict the saliency map of an image. The network structure they adopt can fuse the features of different layers to reduce the loss of information. The proposed SalNet predicts the saliency map pixel-by-pixel, rather than at the patch scale as the CNN-based models do. Moreover, in order to better guide network mining of object edge information, they design a new loss function based on image convolution, which adds an L1 constraint to the edge information of saliency map and ground-truth. Finally, experimental results reveal that the proposed SalNet is effective in salient object detection tasks and is also competitive when compared with 11 state-of-the-art models.

When misalignment, deformation, and tracking failures occur, the appearance of the target tends to change significantly. Effectively learning the change of the target's appearance is an essential problem in visual tracking. Recently, most trackers based on convolutional neural networks update the tracker online to learn changes of the target's appearance. These methods collect tracking results as online training samples. Thus, the reliability of training samples is very important for online updates. The article by Ge *et al.*, ''Self-paced dense connectivity learning for visual tracking,'' proposes a self-paced selection model which integrates the self-paced learning model into the tracking framework with the goal of distinguishing reliable samples from the tracking results. It estimates the reliability of the tracking results by the self-paced function. They design a method that adaptively calculates the value of the pace, which determines the number of samples selected. This method is based on the number of tracking results. At the same time, the quality of the target's features plays a key role in the performance of the tracker. They employ dense connectivity learning to enhance the flow of information throughout the network, which makes the target's features represent better. The extensive experiments demonstrate that the self-paced dense connectivity learning tracker (SPDCT) performs favorably against the state-of-the-art trackers over four benchmark data sets.

When people work on reading comprehension, they often try to find words from passages which are similar to the question words first. Then, they deduce the answer based on the context around these similar words. Therefore, position information may be helpful in finding answer rapidly, and is useful for reading comprehension. However, previous attention-based machine reading comprehension models typically focus on the interaction between the question and the context representation without considering position information. In the article ''Enhancing machine reading comprehension with position information,'' by Xu *et al.*, the authors introduce position information to machine reading comprehension and investigate the performance of the position information. The position information is experimented in three different ways: 1) position encoder; 2) attention mechanism; and 3) position mapping embedding. By experimenting on the TriviaQA data set, the authors demonstrate the effectiveness of position information.

The article by Lu *et al.*, ''2-D-to-stereo panorama conversion using GAN and concentric mosaics,'' describes a learning-based technique to automatically convert a 2-D panorama to its stereoscopic version. In particular, the authors train a generative adversarial network using perspective stereo pairs as inputs. Given a 2-D panorama, they partition it into overlapping local perspective views. To satisfy the panoramic stereo condition, the authors generate a sequence of left and right stereo view pairs and stitch them to produce concentric mosaics. They also describe experiments on synthetic and real data sets, as well as comparisons with competing state-of-the-art techniques, which validate their technique.

GSA is badly suffering from a slow convergence rate and poor local search ability when solving complex optimization problems. To solve this problem, a new hybrid population-based algorithm is proposed with the combination of dynamic multi swarm particle swarm optimization and gravitational search algorithm (GSADMSPSO) in the article ''An improved hybrid method combining gravitational search algorithm with dynamic multi swarm particle swarm optimization,'' by Nagra *et al.* The proposed algorithm divides the main population of masses into smaller sub-swarms and also stabilizes them by presenting a new neighborhood strategy. Then, by adopting the global search ability of the proposed algorithm, each agent (particle) improves the position and velocity. The main idea is to integrate the ability of GSA with the DMSPSO to enhance the performance of exploration and exploitation of a proposed algorithm. In order to evaluate the competences of the proposed algorithm, benchmark functions are employed. The experimental results confirmed a better performance of GSADMSPSO as compared with

the other gravitational and PSO variants in terms of fitness rate.

Urban areas have focused recently on remote sensing applications since their function closely relates to the distribution of built-up areas, where reflectivity or scattering characteristics are the same or similar. Traditional pixel-based methods cannot discriminate the types of urban built-up areas well. The article by Li *et al.*, ''Deep learning-based classification methods for remote sensing images in urban built-up areas,'' investigates a deep learning-based classification method for remote sensing images, particularly for high spatial resolution remote sensing (HSRRS) images with various changes and multi-scene classes. Specifically, to help develop the corresponding classification methods in urban built-up areas, the authors consider four deep neural networks (DNNs): 1) convolutional neural network (CNN); 2) capsule networks (CapsNet); 3) same model with a different training rounding based on CNN (SMDTR-CNN); and 4) same model with different training rounding based on CapsNet (SMDTR-CapsNet). The performance of the proposed methods is evaluated in terms of overall accuracy, kappa coefficient, precision, and confusion matrix. The results reveal that SMDTR-CNN obtains the best overall accuracy (95.0%) and the kappa coefficient (0.944) while improving the precision of parking lot and resident samples by 1% and 4%, respectively.

The unstable nature of radio frequency signals and the need for external infrastructure inside buildings have limited the use of positioning techniques, such as Wi-Fi and Bluetooth fingerprinting. Compared to these techniques, the geomagnetic field exhibits stable signal strength in the time domain. However, existing magnetic positioning methods cannot perform well in a wide space because the magnetic signal is not always discernible. In the article ''Geomagnetic field based indoor landmark classification using deep learning,'' by Bhattarai *et al.*, the authors introduce deep recurrent neural networks (DRNNs) to build a model that is capable of capturing long-range dependencies in variable-length input sequences. The use of DRNNs is brought from the idea that the spatial/temporal sequence of magnetic field values around a given area will create a unique pattern over time, despite multiple locations having the same magnetic field value. Therefore, they can divide the indoor space into landmarks with magnetic field values and find the position of the user in a particular area inside the building. They present long short-term memory DRNNs for spatial/temporal sequence learning of magnetic patterns and evaluate the positioning performance on their testbed data sets. The experimental results show that their proposed models outperform other traditional positioning approaches with machine learning methods, such as support vector machine and k-nearest neighbors.

Softmax cross-entropy loss with L2 regularization is commonly adopted in the machine learning and neural network community. Considering that the traditional softmax cross-entropy loss simply focuses on fitting or classifying the training data accurately but does not explicitly encourage a large decision margin for classification, some loss functions are proposed to improve the generalization performance by solving the problem. However, these loss functions enhance the difficulty of model optimization. In addition, inspired by regularized logistic regression, where the regularized term is responsible for adjusting the width of decision margin, which can be seen as an approximation of support vector machine, the article by Li *et al.*, ''Large-margin regularized softmax cross-entropy loss,'' proposes a large-margin regularization method for softmax cross-entropy loss. The advantages of the proposed loss are twofold as follows: the first is the generalization performance improvement, and the second is easy optimization. The experimental results on three small-sample data sets show that their regularization method achieves good performance and outperforms the existing popular regularization methods of neural networks.

The sparsity of data is one of the main factors restricting the performance of recommender systems. In order to solve the sparsity problem, some recommender systems use auxiliary information, especially text information, as a supplement to increase the prediction accuracy of the ratings. However, the two mainstream approaches based on text analysis have some limitations. The bag-of-words-based model is one of them, being difficult to use the contextual information of the paragraph effectively so that only a shallow understanding of the paragraph can be parsed. Another model based on deep learning can extract the contextual information of the paragraph, and it also increases the complexity of the model. The article by Xie *et al.*, ''Unsupervised learning of paragraph embeddings for context-aware recommendation,'' proposes a novel context-aware recommendation model, named paragraph vector matrix factorization (P2VMF), which integrates the unsupervised learning of paragraph embeddings into probabilistic matrix factorization (PMF). Therefore, P2VMF can capture the semantic information of the paragraph and can improve the prediction accuracy of the ratings. The authors' extensive experiments on real-world data sets show that the performance of the P2VMF model is preferable as compared with those multiple recommendation models in the situation, where the ratings are quite sparse.

With the aim of interpolating the geomagnetic data from under-sampled or missing traces, this article presented an approach based on recurrent neural network (RNN) techniques to avoid the time and labor-intensive nature of the traditional manual and linear interpolation approaches. In the article ''Recurrent neural network-based approach for sparse geomagnetic data interpolation and reconstruction,'' by Liu *et al.*, a deep learning algorithm, long short-term memory (LSTM), is employed to build the precise model for sparse geomagnetic data interpolation. First, a continuous regression hyperplane is specified to recognize the probably intrinsic relationships between sparse and integral traces by inputting training data. Afterward, the trained model is tested with 20% of the trained geomagnetic data and other new untrained data for validation. Finally, extensive experiments are conducted for 2-D and 3-D field data. The results

demonstrate that their RNN-based approach is more superior than a classic linear method and a state-of-the-art method, support vector machine (SVM), as the interpolation precision, is approximately improved by 10%.

Effective and efficient fruit detection is considered crucial for designing an automated robot (AuRo) for yield estimation, disease control, harvesting, sorting, and grading. Several fruit detection schemes for designing AuRo have been developed during the last decades. However, conventional fruit detection methods are deficient in real-time response, accuracy, and extensibility. The article by Zhang *et al.*, ''Multi-task cascaded convolutional networks based intelligent fruit detection for designing automated robot,'' proposes an improved multi-task cascaded convolutional network-based intelligent fruit detection method. This method has the capability to make the AuRo work in real time with high accuracy. Moreover, based on the relationship between the diversity samples of the data set and the parameters of neural networks' evolution, this article presents an improved augmented method, a procedure that is based on image fusion to improve the detector performance. The experiment results demonstrated that the proposed detector performed immaculately both in terms of accuracy and time-cost. Furthermore, the extensive experiment also demonstrated that the proposed technique has the capacity and good portability to work with other akin objects conveniently.

City-scale traffic speed prediction provides a significant data foundation for intelligent transportation systems, which enrich commuters with up-to-date information about traffic conditions. However, predicting on-road vehicle speed accurately is challenging, as the speed of vehicles on the urban road can be affected by various factors. These factors can be categorized into three main aspects: temporal, spatial, and other latent information. In the article ''A novel spatio-temporal model for city-scale traffic speed prediction,'' by Niu *et al.*, the authors propose a novel spatio-temporal model named L-U-Net, based on U-Net, as well as long short-term memory architecture, and develop an effective speed prediction model which is capable of forecasting city-scale traffic conditions. It is worth noting that their model can avoid the high complexity and uncertainty of subjective feature extraction and can be easily extended to solve other spatio-temporal prediction problems such as flow prediction. The experimental results demonstrate that the proposed prediction model can forecast urban traffic speed effectively.

Recognizing prohibited items automatically is of great significance for intelligent X-ray baggage security screening. Convolutional neural networks (CNNs), with the support of big training data, have been verified as powerful models capable of reliably detecting expected objects in images. Therefore, building a specific CNN model working reliably on prohibited item detection also requires large amounts of labeled item image data. Unfortunately, the current X-ray baggage image database is not big enough in count and diversity for CNN model training. In the article, ''Data augmentation for X-ray prohibited item images using generative adversarial networks,'' by Yang *et al.*, the authors propose a novel method for X-ray prohibited item data augmentation using generative adversarial networks (GANs). The prohibited items are first extracted from X-ray baggage images using a K-nearest neighbor matting scheme. Then, the poses of the obtained item images are estimated using a space rectangular coordinate system and categorized into four or eight classes for constructing a training database. For generating realistic samples reliably, different GAN models are evaluated using Frechet Inception Distance scores, and some important tips for handling GAN training in X-ray prohibited item image generation are reported. Finally, to verify whether the generated images belong to its corresponding class or not, a cross-validation scheme based on a CNN model is implemented. The experimental results show that most of the generated images can be classified correctly by the CNN model. This implies that the generated prohibited item images can be used as the extended samples to augment an X-ray image database.

Sketch-based image retrieval (SBIR) is a challenging task due to the cross-domain gap between sketch queries and natural images. In the article ''Edge-guided cross-domain learning with shape regression for sketch-based image retrieval,'' by Song *et al.*, the authors propose a novel edge-guided cross-domain learning network to reduce the domain gap. In particular, edge maps extracted from natural images are introduced as the bridge between two domains, and the inputs from different domains are embedded into a common feature space. An edge guidance module is proposed to fuse natural images and the corresponding edge maps, which guides the network to generate more discriminative features in the domain alignment process. Meanwhile, a shape regression module is proposed to capture the inherent shape similarity between sketches and natural images. By training the proposed network in an end-to-end process, the sketch and natural image domains can be effectively associated, which potentially overcomes the challenge of the common feature learning for two heterogeneous domains. The experimental results on the SBIR data set demonstrate that the proposed method achieves superior performance compared with the state-of-the-art methods.

Multiple object tracking has been a challenging field, mainly due to noisy detection sets, an identity switch caused by occlusion, and similar appearance among nearby targets. Previous works rely on appearance models that are built on an individual or several selected frames for the comparison of features, but they cannot encode the long-term appearance changes caused by pose, viewing angle, and lighting conditions. In the article ''MOANA: An online learned adaptive appearance model for robust multiple object tracking in 3-D,'' by Tang and Hwang, the authors propose an adaptive model that learns online a relatively long-term appearance change of each target. The proposed model is compatible with any feature of fixed dimension or their combination, whose learning rates are dynamically controlled by the adaptive update and spatial weighting schemes. To handle occlusion

and nearby objects that are sharing a similar appearance, they also design cross-matching and re-identification schemes based on the application of the proposed adaptive appearance models. In addition, the 3-D geometry information is effectively incorporated in their formulation for data association. The proposed method outperforms all the state-of-the-art on the MOTChallenge 3-D benchmark and achieves real-time computation with only a standard desktop CPU. It has also shown superior performance over the state-of-the-art on the 2-D benchmark of MOTChallenge.

In terms of biometrics, a human finger possesses trimodal traits including fingerprint, finger-vein, and finger-knuckle-print, which provides convenience and practicality for finger trimodal fusion recognition. The scale inconsistency of finger trimodal images is an important reason affecting effective fusion. It is, therefore, very important to develop a theory to give a unified expression of finger trimodal features. In the article, ''Graph fusion for finger multimodal biometrics,'' by Zhang *et al.*, a graph-based feature extraction method for finger biometric images is proposed. The feature expression based on graph structure can effectively solve the problem of feature space mismatch for the finger three modalities. The authors provide two fusion frameworks to integrate the finger trimodal graph features together, the serial fusion and coding fusion. The research results can not only promote the advancement of finger multimodal biometrics technology but also provide a scientific solution framework for other multimodal feature fusion problems. The experimental results show that the proposed graph fusion recognition approach obtains a better and more effective recognition performance in finger biometrics.

The increasing use of invoicing has created an unnecessary burden on labor and material resources in the financial sector. The article by Sun *et al.*, ''Template matching-based method for intelligent invoice information identification,'' proposes a method to intelligently identify invoice information based on template matching, which retrieves the required information by image preprocessing, template matching, optical character recognition, and information exporting. The original invoice image is preprocessed first to remove the useless background information by secondary rotation and edge cutting. Then, the region of the required information in the obtained regular image is extracted by template matching, which is the core of the intelligent invoice information identification. The optical character recognizing is utilized to convert the image information into text so that the extracted information can be directly used. The text information is exported for backup and subsequent use in the last step. The experimental results indicate that the method using normalized correlation coefficient matching is the best choice, demonstrating a high accuracy of 95% and the average running time of 14 milliseconds.

Multi-task learning (MTL) is a machine learning method to share knowledge for multiple related machine learning tasks via learning those tasks jointly. It has been shown to be capable of effectively improving the generalization capability of each single task (learning just one task at a time). In the article ''A 3-D-CNN and LSTM based multi-task learning architecture for action recognition,'' by Ouyang *et al.*, the authors propose a novel MTL architecture that first combines 3-D convolutional neural networks (3-D CNN) plus the long short-term memory (LSTM) networks together with the MTL mechanism, tailored to information sharing of video inputs. They split each video into several clips and apply the hybrid deep model of 3-D CNN and LSTM to extract the sequential features of those video clips. Therefore, the proposed MTL model can share visual knowledge based on those video-clip features among different categories more efficiently. They evaluate the proposed method on three popular public action recognition video data sets. The experimental results show that the proposed MTL method can efficiently share detailed information in video clips among multiple action categories and outperforms other multi-task methods.

Promising results have been obtained using deep neural networks for traffic classification. However, most of the previous work was confined to one specific task of the classification, which restricts the classifier potential performance and application areas. The traffic flow can be labeled from a different perspective, which might help to improve the accuracy of classifiers by exploring more meaningful latent features. In addition, a deep neural network (DNN)-based model is hard to adapt to the changes in new classification demands because training such a new model costs not only many computing resources but also lots of labeled data. The article by Sun *et al.*, ''Common knowledge based and one-shot learning enabled multi-task traffic classification,'' proposes a multi-output DNN model to simultaneously learn multi-task traffic classifications. In this model, the common knowledge of traffic is exploited by the synergy among the tasks and improves the performance of each task separately. Also, it is showed that this structure shares the potential of meeting new demands in the future, and meanwhile, being able to achieve the classification with advanced speed and fair accuracy. One-shot learning, which refers to the learning process with scarce data, is also explored and their approach shows notable performance.

The article by Zhao *et al.*, ''Deep learning for risk detection and trajectory tracking at construction sites,'' investigates deep learning for risk detection and trajectory tracking at construction sites. Typically, safety officers are responsible for inspecting and verifying site safety due to many potential risks. Traditional target detection algorithms depend heavily on hand-crafted features. However, these features are difficult to design, and detection accuracy is poor. To solve these problems, this article proposes a deep-learning-based detection algorithm that uses pedestrian wearable devices (e.g., helmets and colored vests) to identify pedestrians. The authors trained a special data set by labeling helmets and colored vests to detect the two features among construction workers. Specifically, the Kalman filter and Hungarian matching algorithms are employed to track pedestrian trajectories. The testing experiment was run on an NVIDIA GeForce GTX 1080Ti with a detection speed of 18 frames/s. The mean average

precision was able to reach 0.89 when the intersection over union was set at 0.5.

Sign language recognition aims to recognize meaningful movements of hand gestures and is a significant solution in intelligent communication between the deaf community and hearing societies. However, until now, the current dynamic sign language recognition methods still have some drawbacks with difficulties of recognizing complex hand gestures, low recognition accuracy for most dynamic sign language recognition, and potential problems in larger video sequence data training. In order to solve these issues, the article by Liao *et al.*, ''Dynamic sign language recognition based on video sequence with BLSTM-3-D residual networks,'' presents a multimodal dynamic sign language recognition method based on a deep 3-D residual ConvNet and bi-directional LSTM networks, which is named BLSTM-3-D residual network (B3D ResNet). This method consists of three main parts. First, the hand object is localized in the video frames in order to reduce the time and space complexity of network calculation. Then, the B3D ResNet automatically extracts the spatiotemporal features from the video sequences and establishes an intermediate score corresponding to each action in the video sequence after feature analysis. Finally, by classifying the video sequences, the dynamic sign language is accurately identified. The experiment is conducted on test data sets, including the DEVISIGN_D data set and the SLR_Data set. The results show that the proposed method can obtain state-of-the-art recognition accuracy (89.8% on the DEVISIGN_D data set and 86.9% on the SLR_Data set). In addition, the B3D ResNet can effectively recognize complex hand gestures through larger video sequence data, and obtain high recognition accuracy for 500 vocabularies from Chinese hand sign language.

Finite mixture models based on the symmetric Gaussian distribution have been applied broadly in data analysis. However, not all the data in real-world applications can be safely supposed to have a symmetric Gaussian form. The article by Lai *et al.*, ''Positive data modeling using a mixture of mixtures of inverted beta distributions,'' presents a new mixture model that includes the inverted Beta mixture model (IBMM) as a special case to analyze the positive non-Gaussian data. The advantage of the proposed model is that the number of model parameters is variable and infinite. Consequently, the proposed model is adaptable to the size of the data. On the basis of the recently proposed extended variational inference (EVI) framework, the authors develop a closed-form solution to approximate the posterior distributions. The performance and effectiveness of the proposed model are demonstrated with the real data generated from two challenging applications, namely image classification and object detection.

Recently, deep learning has become a preferred choice for performing tasks in diverse application domains such as computer vision, natural language processing, sensor data analytics for healthcare, and collaborative filtering for personalized item recommendation. In addition, the Generative Adversarial Networks (GANs) have become one of the most popular frameworks for training machine learning models. Motivated by the huge success of GAN and deep learning on a wide range of fields, the article by Chae *et al.*, ''Collaborative adversarial autoencoders: An effective collaborative filtering model under the GAN framework,'' explores an effective way to exploit both techniques into the collaborative filtering task for accurate recommendations. The authors noticed that the IRGAN and GraphGAN are pioneering methods that successfully apply GAN to recommender systems. They point out an issue regarding the employment of standard matrix factorization (MF) as their basic model, which is linear and unable to capture the non-linear, subtle latent factors underlying user-item interactions. The proposed recommendation framework, named Collaborative Adversarial Autoencoders (CAAE), significantly extends the conventional IRGAN and GraphGAN as summarized: 1) they use Autoencoder, which is one of the most successful deep neural networks, as their generator, instead of using the MF model; 2) they employ Bayesian personalized ranking (BPR) as their discriminative model; and 3) they incorporate another generator model into their framework that focuses on generating negative items, which are items that a given user may not be interested in. They empirically test the proposed framework using three real-life data sets along with four evaluation metrics. Owing to those extensions, the proposed framework not only produces considerably higher recommendation accuracy than the conventional GAN-based recommenders (i.e., IRGAN and GraphGAN) but also outperforms the other state-of-the-art top-N recommenders (i.e., BPR, PureSVD, and FISM).

The article by Wang *et al.*, ''Knowledge base question answering with attentive pooling for question representation,'' presents a neural network model for a knowledge base (KB)-based single-relation question answering (SR-QA). This model is composed of two main modules, i.e., entity linking and relation detection. In each module, an embedding vector is computed from the input question sentence to calculate its similarity scores with entity candidates or relation candidates. This article focuses on attention-based question representation in SR-QA. In the entity linking module, two attentive pooling methods, inner-sentence attention and structure attention, are employed to derive question embeddings, and the performances are compared in experiments. In the relation detection module, a new attentive pooling structure, named multilevel target attention (MLTA), is proposed to utilize the multilevel descriptions of relations. In this structure, the attention weights for aggregating the hidden states of question sentences are calculated using relation candidates as queries at the relation level, word level, and character level. Then, the similarity scores for relation detection are computed by matching questions to relation candidates at all three levels. The experimental results show that the proposed model achieves a state-of-the-art accuracy of 82.29% on the simple questions data set.

Furthermore, the results of ablation tests demonstrate the effectiveness of the proposed MLTA method for question representation.

Internet application providers now have more incentive than ever to collect user data, which greatly increases the risk of user privacy violations due to the emergence of deep neural networks. In the article ''TensorClog: An imperceptible poisoning attack on deep neural network applications,'' by Shen *et al.*, the authors propose TensorClog, a poisoning attack technique that is designed for privacy protection against deep neural networks. TensorClog has three properties, with each of them serving a privacy protection purpose: 1) training on TensorClog poisoned data results in lower inference accuracy, reducing the incentive of abusive data collection; 2) training on TensorClog poisoned data converges to a larger loss, which prevents the neural network from learning the privacy; and 3) TensorClog regularizes the perturbation to remain a high structure similarity so that the poisoning does not affect the actual content in the data. Applying the proposed TensorClog poisoning technique to CIFAR-10 data set results in an increase in both converged training loss and test error by 300% and 272%, respectively. It manages to maintain data's human perception with a high SSIM index of 0.9905. More experiments, including different limited information attack scenarios and a real-world application transferred from pre-trained ImageNet models, are presented to further evaluate TensorClog's effectiveness in more complex situations.

With the explosive growth of image big data in the field of agriculture, image segmentation algorithms are confronted with unprecedented challenges. As one of the most important image segmentation technologies, the fuzzy c-means (FCMs) algorithm has been widely used in the field of agricultural image segmentation as it provides simple computation and high-quality segmentation. However, due to the large amount of computation, the sequential FCM algorithm is too slow to finish the segmentation task within an acceptable time. The article by Liu *et al.*, ''A spark-based parallel fuzzy c-means segmentation algorithm for agricultural image big data,'' proposes a parallel FCM segmentation algorithm based on the distributed memory computing platform, Apache Spark, for agricultural image big data. The input image is first converted from the RGB color space to the lab color space and generates point cloud data. Then, point cloud data are partitioned and stored in different computing nodes, in which the membership degrees of pixel points to different cluster centers are calculated, and the cluster centers are updated iteratively in a data-parallel form until the stopping condition is satisfied. Finally, point cloud data are restored after clustering for reconstructing the segmented image. On the Spark platform, the performance of the parallel FCMs algorithm is evaluated and reaches an average speedup of 12.54 on ten computing nodes. The experimental results show that the Spark-based parallel FCMs algorithm can obtain a significant increase in speedup, and the agricultural image testing set delivers a better performance improvement of 128% than the Hadoop-based approach. This article indicates that the Spark-based parallel FCM algorithm provides faster speed of segmentation for agricultural image big data and has better scale-up and size-up rates.

Quality-of-Service (QoS) value is usually unknown in service recommendation practices. There are some matrix factorization approaches for predicting the unknown value with a user-service model, which uses a single collaboration with the user's neighbor when looking for different services. However, the QoS value is highly related to the service provider and participants. The services are considered in various collaboration based on different users. By considering the context of services, the article by Guo *et al.*, ''Personalized QoS prediction for service recommendation with a service-oriented tensor model,'' proposes a QoS prediction model using tensor decomposition based on service collaboration called Service-Oriented Tensor (SOT). The prediction approach analyzes service collaboration from other similar services and relevant users by using a three-order tensor. Compared with the traditional model, the experiment results show that the proposed model achieves better prediction accuracy.

Despite the rapid progress over the past decade, person re-identification (reID) remains a challenging task due to the fact that discriminative features underlying different granularities are easily affected by illumination and camera-view variation. Most deep learning-based algorithms for reID extract global embedding as the representation of the pedestrian from the convolutional neural network. Considering that person attributes are robust and informative to identify pedestrians. The article by Zhang *et al.*, ''Part-based attribute-aware network for person re-identification,'' proposes a multi-branch model, namely part-based attribute-aware network (PAAN), to leverage both person reID and attribute performance, which not only utilizes ID labels visible to the whole image but also utilizes attribute information. In order to learn discriminative and robust global representation which is invariant to the above-mentioned fact, the authors resort to global and local person attributes to build global and local representation, respectively, utilizing the proposed layered partition strategy. Their goal is to exploit global or local semantic information to guide the optimization of global representation. In addition, in order to enhance the global representation, they design a semantic bridge replenishing mid-level semantic information for the final representation, which contains high-level semantic information. Extensive experiments are conducted to demonstrate the effectiveness of the proposed approach on two large-scale person re-identification data sets, including Market-1501 and DukeMTMC-reID, and this approach achieves rank-1 of 92.40% on Market-1501 and 82.59% on DukeMTMC-reID showing strong competitiveness among the start of the art.

With the development of deep learning, image super-resolution has made great breakthroughs. However, compared with a color image, the performance of depth map super-resolution is still poor. To address this problem, multi-

level recursive guidance and progressive supervised network (MRG-PS) is proposed in the article ''Depth map super-resolution via multilevel recursive guidance and progressive supervision,'' by Yang *et al.* First, a multilevel recursive guidance architecture is presented to extract features of a color stream and depth stream, in which the depth stream is guided by the color features at each level. Second, a progressive supervision module is developed to supervise the multilevel recursion to obtain depth-residual information on different levels. Finally, a residual fusion and construction strategy is designed to fuse all residual information and reconstruct the high-resolution depth map. The experimental results demonstrate that the proposed method outperforms the state-of-the-art methods.

The actual driving condition and fuel consumption rate gaps between lab and real-world are becoming larger. In the article, ''Multilayer perceptron method to estimate real-world fuel consumption rate of light duty vehicles,'' by Li *et al.*, the authors demonstrate an approach to determine the most important factors that may influence the prediction of the real-world fuel consumption rate of light-duty vehicles. A multilayer perceptron (MLP) method is developed for the prediction of fuel consumption since it provides accurate classification results despite the complicated properties of different types of inputs. The model considers the parameters of external environmental factors, the manipulation of vehicle companies, and the drivers' driving habits. Based on the BearOil database in China, 2,424,379 samples are used to optimize the proposed model. They indicate that differences exist between real-world fuel consumption and standard fuel consumption under simulation conditions. This study enables the government and policy-makers to use big data and intelligent systems for energy policy assessment and better governance.

Satellite image semantic segmentation, including extracting roads, detecting buildings, and identifying land cover types, is essential for sustainable development, agriculture, forestry, urban planning, and climate change research. Nevertheless, it is still unclear how to develop a refined semantic segmentation model in an efficient and elegant way. In the article ''Toward accurate high-resolution satellite image semantic segmentation,'' by Wu *et al.*, the authors propose an attention dilation-LinkNet (AD-LinkNet) neural network that adopts encoder–decoder structure, serial–parallel combination dilated convolution, channel-wise attention mechanism, and pre-trained encoder for semantic segmentation. Serial–parallel combination dilated convolution enlarges receptive field as well as assemble multi-scale features for multi-scale objects, such as long-span roads and small pools. The channel-wise attention mechanism is designed to take advantage of the context information in the satellite image. The experimental results on road extraction and surface classification data sets prove that the AD-LinkNet shows a significant effect on improving the segmentation accuracy. They defeated the D-Linknet algorithm that won the first place in the CVPR 2018 DeepGlobe road extraction competition.

Identifying shoe-print impressions at the scene of crime (SoC) from database images is a challenging problem in forensic science due to the complicated impressing surface, the partial absence of on-site impressions, and the huge domain gap between the query and the gallery images. The existing approaches pay much attention to feature extraction while ignoring its distinctive characteristics. In the article ''Shoe-print image retrieval with multi-part weighted CNN,'' by Ma *et al.*, the authors propose a novel multi-part weighted convolutional neural network (MP-CNN) for shoe-print image retrieval. Specifically, the proposed CNN model processes images in three steps: 1) dividing the input images vertically into two parts and extracting subfeatures by a parameter-shared network individually; 2) calculating the importance weight matrix of the sub-features based on the informative pixels they contained and concatenating them as the final feature; and 3) using the triplet loss function to measure the similarity between the query and the gallery images. In addition to the proposed network, they adopt an effective strategy to enhance the quality of the images and to reduce the domain gap using the U-Net structure. The experimental evaluations demonstrate that the proposed method significantly outperforms other fine-grained cross-domain methods on SPID data set and obtains comparative results with the state-of-the-art shoe-print retrieval methods on the FID300 data set.

Building an empathic conversation agent in open-domain is a key step toward affective computing and intelligent interactions. However, most current methods either focus on the consistency of content or the controllability of emotion and handling both factors are not yet properly solved. In the article ''Generating emotional controllable response based on multi-task and dual attention framework,'' by Xu *et al.*, the authors propose the multitask and dual attentions (MTDA) framework for generating an emotional response. The MTDA framework decomposes the input utterance into the content layer and emotional layer and then encodes and decodes them separately, which makes this end-to-end model more interpretable and controllable. A multi-task learning based encoder is employed in the MTDA framework, which can obtain the representation of the content and the emotion through unsupervised learning and supervised learning. A dual-attention mechanism is adopted for decoding, which ensures that specific emotional responses are coherent with the content and the emotion of the input. They also combine the MTDA framework with state-of-the-art generative models to train emotional generation systems. Extensive experiments show that the proposed model can not only adapt to different target emotion goals but also generate coherent and informative responses.

Shot boundary detection is essential to detect the position of frames where the shot changes. It has been actively studied in video analysis and management for convenience,

which becomes a key technique with the rapid proliferation of rich and diverse videos. With respect to the complex characteristics of different shots in varying length and content variation property, the article by Wu *et al.*, ''Two stage shot boundary detection via feature fusion and spatial-temporal convolutional neural networks,'' presents a two-stage method for shot boundary detection (TSSBD) which distinguishes abrupt shot by fusing color histogram and deep features, and locate gradual shot changes with C3D-based deep analysis. Abrupt shot changes are detected first as it occurs between two frames, which divides the complete video into segments containing gradual transitions. Over these video segments, gradual shot change detection is implemented using a 3-D convolutional neural network, which classifies clips into specific gradual shot change types. Finally, an effective merging strategy is proposed to locate positions of gradual shot transitions. The experimental analysis illustrates that the proposed progressive method is capable of detecting both abrupt shot transitions and gradual shot transitions accurately.

Chinese grammatical error correction (CGEC) is practically useful for learners of Chinese as a second language, but it is a rather challenging task due to the complex and flexible nature of Chinese language such that existing methods for English cannot be directly applied. In the article ''Chinese grammatical error correction based on convolutional sequence to sequence model,'' by Li *et al.*, the authors introduce a convolutional sequence to sequence model into the CGEC task for the first time, since many Chinese grammatical errors are concentrated between three and four words and a convolutional neural network can better capture the local context. A convolution-based model can obtain the representations of the context by fixed size kernel. By stacking convolution layers, long-term dependences can be obtained. The authors also propose two optimization methods, shared embedding and policy gradient, to optimize the convolutional sequence to sequence model through sharing parameters and reconstructing loss function. In addition, they collate the existing Chinese grammatical correction corpus in detail. The results show that the proposed optimization methods both achieve large improvement compared with the natural machine translation model based on a recurrent neural network.

Clustering is one of the most important topics in data mining and machine learning. The density peaks clustering (DPC) algorithm is a well-known density-based clustering method that can efficiently and effectively deal with non-spherical clusters. However, the computational methods of the local density and the distance measure are simple and easily ignore the correlation and similarity between samples, and the manual setting of parameters has a great influence on the clustering results; therefore, the clustering performance of DPC is poor on the high-dimensional data sets. To address these issues, the article by Sun *et al.*, ''An adaptive density peaks clustering method with fisher linear discriminant,'' presents an adaptive DPC algorithm with Fisher linear discriminant for the clustering of complex data sets, called ADPC-FLD. First, the kernel density estimation function is introduced to calculate the local density of the sample points. Pearson correlation coefficient between samples as weight is employed to construct a weighted Euclidean distance function to measure the distance between samples. This considers both the spatial structure and the correlation of the samples. Then, a novel density estimation entropy is proposed, and based on the minimization of density estimation entropy, the density estimation parameters are adaptively selected according to the distribution characteristics of the data, which can efficiently eliminate the influence of manual setting. Third, an adaptive strategy of cluster center selection is designed to avoid the error caused by the noise data as the cluster centers and the uncertainty of manually selecting the cluster centers. Finally, Fisher linear discriminant algorithm is used to eliminate the irrelevant information and reduce the dimensionality of high-dimensional data, following on which an adaptive DPC method is implemented on six synthetic data sets, thirteen UCI data sets, and seven gene expression data sets for comparing with other related algorithms. The experimental results on 26 data sets show that the proposed algorithm significantly outperforms several outstanding clustering approaches in terms of clustering accuracy and efficiency.

Graph-based semi-supervised learning (GSSL) has attracted great attention over the past decade. However, there are still several open problems: 1) how to construct a graph that effectively represents the underlying structure of data and 2) how to incorporate label information of the labeled samples into a procedure of label propagation. The solution in the article ''Semi-supervised classification with graph structure similarity and extended label paropagation,'' by Ma *et al.*, mainly focuses on two aspects: 1) the authors propose a new graph construction technique by fusing local and global structural similarity (FLGSS). Based on an initial graph structure such as K-nearest neighbors (KNN), they utilize different types of link prediction algorithms to extract local and global graph structure information. These two types of structure information are fused into a graph structure that enhances the ability to represent the data correlation and 2) by incorporating the label correlation with feature similarity of samples, they propose an extended label propagation algorithm (ELP). Through experiments on three different types of data sets, it is shown that the proposed method outperforms other widely used graph construction methods. The extended label inference algorithm achieves better classification results than some state-of-the-art methods. The proposed FLGSS method starts from KNN graph, and two link prediction algorithms are performed to construct the graph. With the time complexity analysis, the authors theoretically deduce that the time complexity of FLGSS is not beyond that of KNN. Meanwhile, the time complexity of ELP remains the same as that of the traditional LP algorithm.

Local Outlier Factor (LOF) outlier detecting algorithm has good accuracy in detecting global and local outliers. However, the algorithm needs to traverse the entire data set

when calculating the local outlier factor of each data point, which adds extra time overhead and makes the algorithm execution inefficient. In addition, if the K-distance neighborhood of an outlier point P contains some outliers that are incorrectly judged by the algorithm as normal points, then P may be misidentified as a normal point. To solve the above problems, the article by Yang *et al,*, ''An outlier detection approach based on improved self-organizing feature map clustering algorithm,'' proposes a Neighbor Entropy Local Outlier Factor (NELOF) outlier detecting algorithm. First, the authors improve the Self-Organizing Feature Map (SOFM) algorithm and use the optimized SOFM clustering algorithm to cluster the data set. Therefore, the calculation of each data point's local outlier factor only needs to be performed inside the small cluster. Second, this article replaces the K-distance neighborhood with relative K-distance neighborhood and utilizes the entropy of relative K neighborhood to redefine the local outlier factor, which improves the accuracy of outlier detection. Experimental results confirm that this optimized SOFM algorithm can avoid the random selection of neurons, and improve the clustering effect of traditional SOFM algorithm. In addition, the proposed NELOF algorithm outperforms the LOF algorithm in both accuracy and execution time of outlier detection.

Mobile application developers are getting more concerned due to the importance of quality requirements or non-functional requirements (NFR) in software quality. Developers around the globe are actively asking question(s) and sharing solutions to the problems related to software development on Stack Overflow (SO). The knowledge shared by developers on SO contains useful information related to software development such as feature requests (functional/non-functional), code snippets, reporting bugs, or sentiments. Extracting the NFRs shared by iOS developers on the programming Q&A website SO has become a challenge and a less researched area. To identify and understand the real problems, needs, trends, and the critical NFRs or quality requirements discussed on Stack Overflow related to iOS mobile application development, the article by Ahmad *et al*., ''Toward empirically investigating non-functional requirements of iOS developers on stack overflow,'' extracts and uses only the iOS posts data on SO. The authors apply the well-known statistical topical model Latent Dirichlet Allocation (LDA) to identify the main topics in iOS posts on SO. Then, they label the extracted topics with quality requirements or NFRs by using the wordlists to assess the trend, evolution, hot and unresolved NFRs in all iOS discussions. Their findings reveal that the highly frequent topics the iOS developers discussed are related to usability, reliability, and functionality, followed by efficiency. Interestingly, the most problematic areas unresolved are also usability, reliability, and functionality, though followed by portability. In addition, the evolution trend of each of the six different quality requirements or NFRs over time is depicted through comprehensive visualization. Their first empirical investigation on approximately 1.5 million iOS posts and comments on SO gives an insight on comprehending the NFRs in iOS application development through the lens of real-world practitioners.

Vehicle detection based on unmanned aerial vehicle (UAV) images is a challenging task due to the small size of objects, complex backgrounds, and the imbalance of various vehicle samples. The article by Yang *et al.*, ''Effective contexts for UAV vehicle detection,'' proposes a high-performance UAV vehicle detector. The authors use the single-shot refinement neural network (RefineDet) as a base network, which employs top-down architecture to offer contextual information, achieving accurate detection. However, for the small size of vehicles, the top-down architecture introduces too much context, which brings surrounding interference. The authors present a multi-scale adjacent connection module (ACM) to provide effective contextual information and reduce interference for vehicle detection. In addition, they adopt an alternate double loss training strategy (ADT) to solve the problem of imbalance between hard and easy examples during training and they design suitable default boxes according to the distribution of the UAV data set to improve the recall rate. The proposed method achieves 92.0% and 90.4% accuracy on the collected UAV data set and the publicly available Stanford drone data set, respectively. The proposed detector can run at 58 FPS on a single GPU.

It is a universal consensus that autonomous vehicles (AVs) and human-driven vehicles will share the roads in the coming decades. Trajectory planning of AVs has been extensively studied from the perspective of driving safety and riding comfort. However, human-like trajectory planning has rarely been studied. In the article ''Learning human-like trajectory planning on urban two-lane curved roads from experienced drivers,'' by Li *et al.*, the authors characterize and model human driving trajectories using real vehicle field test data collected on five two-way and two-lane urban curved roads with 20 experienced drivers and three experimental vehicles. A differential global positioning system (GPS) and an inertial navigation system (INS) are used to measure the vehicle positions and velocities in high precision. The authors study the trajectory characteristics of experienced drivers on curved two-lane roads, especially the relationships between the vehicle trajectories on bidirectional two lanes. Based on the long short-term memory neural network (LSTM NN), they develop a data-driven trajectory model to generate human-like driving trajectories. By comparing with three other modeling methods, the LSTM NN model is validated and tested in various cases with promising performance.

The Internet improves the speed of information dissemination, and the scale of unstructured text data is expanding and increasingly being used for mass communication. Although these large amounts of data meet the infinite demand, it is difficult to find public focus in a timely manner. Therefore, information extraction from big data has become an important research issue, and there are many published studies on big data processing at home and abroad. In the article ''An artificial intelligence driven multi-feature extraction scheme

for big data detection,'' by Wan *et al.*, the authors propose a multi-feature keyword extraction method. Based on this, an artificial intelligence-driven big data MFE scheme is designed, then an application example of the general scheme is expanded and detailed. Taking news as the carrier, this scheme is applied to the algorithm design of hot event detection. As a result, a multifeature fusion clustering algorithm is proposed based on user attention with two main stages. In the first stage, a multi-feature fusion model is developed to evaluate keywords. This model combines term frequency and part of speech features. The authors use it to extract keywords for representing news and events. In the second stage, they perform clustering and detect hot events in accordance with the procedure, and during the composition of news clusters, they analyze several variadic parameters in order to explore the optimal effectiveness. Then, experiments on the news corpus are conducted and the results show that the approach presented herein performs well.

WLAN-based indoor positioning algorithm has the characteristics of simple layout and low price, and it has gradually become a hotspot in both academia and industry. However, due to the poor stability of Wi-Fi signals, the received signal strength (RSS) fingerprints of some adjacent reference positions are difficult to evaluate similarity when utilizing traditional distance-based calculation methods. By clustering these RSS fingerprints into one region, the commonly utilized KNN algorithm in the past cannot achieve accurate positioning in the region. The article by Wang *et al.*, ''Learning to improve WLAN indoor positioning accuracy based on DBSCAN-KRF algorithm from RSS fingerprint data,'' introduces a concept of the insensitive region of the RSS fingerprint and a new algorithm named DBSCAN-KRF. This algorithm can delete noise sample and detect insensitive region. Then, different methods are selected to achieve indoor positioning by judging the region of the estimated fingerprint sample. The KNN algorithm is selected when the region is sensitive, and a random forest algorithm is selected when the region is insensitive. The experimental results show that the DBSCAN-KRF algorithm is superior when compared with other alternative indoor positioning algorithms.

Customer service is critical to all companies, as it may directly connect to the brand's reputation. Due to a great number of customers, e-commerce companies often employ multiple communication channels to answer customers' questions, for example, Chatbot and Hotline. On the one hand, each channel has limited capacity to respond to customers' requests; on the other hand, customers have different preferences over these channels. The current production systems are mainly built based on business rules that merely consider the tradeoffs between resources and customers' satisfaction. To achieve the optimal tradeoff between resources and customers' satisfaction, the article by Liu *et al.*, ''Which channel to ask my question?: Personalized customer service request stream routing using deep reinforcement learning,'' proposes a new framework based on deep reinforcement learning that directly takes both resources and user model into account. In addition to the framework, they also propose a new deep-reinforcement-learning-based routing method-double dueling deep Q-learning with prioritized experience replay (PER-DoDDQN). They evaluate the proposed framework and method using both synthetic and real customer service log data from a large financial technology company. They show that the proposed deep-reinforcement-learning-based framework is superior to the existing production system. Moreover, they also show that the proposed PER-DoDDQN is better than all other deep Q-learning variants in practice, which provides a more optimal routing plan. These observations suggest that the proposed method can seek the trade-off, where both channel resources and customers' satisfaction are optimal.

Change in a data stream can occur at the concept level and at the feature level. Change at the feature level can occur if new additional features appear in the stream, or if the importance and relevance of a feature changes as the stream progresses. This type of change has not received as much attention as concept-level change. Furthermore, a lot of the methods proposed for clustering streams (density-based, graph-based, and grid-based) rely on some form of distance as a similarity metric, and this is problematic in high-dimensional data where the curse of dimensionality renders distance measurements and any concept of ''density'' difficult. To address these two challenges, the article by Fahy and Yang, ''Dynamic feature selection for clustering high dimensional data streams,'' proposes combining them and framing the problem as a feature selection problem, specifically a dynamic feature selection problem. They propose a dynamic feature mask for clustering high dimensional data streams. Redundant features are masked, and clustering is performed along unmasked, relevant features. If a feature's perceived importance changes, the mask is updated accordingly; previously unimportant features are unmasked, and features which lose relevance become masked. The proposed method is algorithm-independent and can be used with any of the existing density-based clustering algorithms, which typically do not have a mechanism for dealing with feature drift and struggle with high-dimensional data. The authors evaluate the proposed method on four density-based clustering algorithms across four high-dimensional streams; two text streams and two image streams. In each case, the proposed dynamic feature mask improves clustering performance and reduces the processing time required by the underlying algorithm. Furthermore, change at the feature level can be observed and tracked.

In conclusion, we would like to thank all the authors who submitted their research articles to our Special Section. We highly appreciate the contributions of the reviewers for their constructive comments and suggestions. We also would like to acknowledge the guidance from the Editor-in-Chief and staff members.

ZHANYU MA, *Associate Editor*
Beijing University of Posts and Telecommunications
Beijing 100876, China

SUNWOO KIM, *Guest Editor*
Hanyang University
Seoul 04763, South Korea

YI-ZHE SONG, *Guest Editor*
University of Surrey
Guildford GU2 7XH, U.K.

PASCUAL MARTÍNEZ-GÓMEZ, *Guest Editor*
Amazon
Seattle, WA 98109, USA

HUIJI GAO, *Guest Editor*
LinkedIn
Sunnyvale, CA 94085, USA

JALIL TAGHIA, *Guest Editor*
KTH Royal Institute of Technology
114 28 Stockholm, Sweden

ZHANYU MA (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the KTH-Royal Institute of Technology, Sweden, in 2011. From 2012 to 2013, he was a Postdoctoral Research Fellow with the School of Electrical Engineering, KTH-Royal Institute of Technology. He was an Associate Professor with the Beijing University of Posts and Telecommunications, Beijing, China, from 2014 to 2019, where he has been a Full Professor since 2019. He has also been an Adjunct Associate Professor with Aalborg University, Aalborg, Denmark, since 2015. His research interests include pattern recognition and machine learning fundamentals with a focus on applications in computer vision, multimedia signal processing, and data mining.

SUNWOO KIM (Senior Member, IEEE) received the B.S. degree from Hanyang University, Seoul, South Korea, in 1999, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA, in 2005. Since 2005, he has been working with the Department of Electronic Engineering, Hanyang University, where he is currently a Professor. He was a Visiting Scholar with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, from 2018 to 2019. He is also the Director of the 5G/Unmanned Vehicle Research Center, funded by the Ministry of Science and ICT of Korea. His research interests include wireless communication/positioning/localization, signal processing, vehicular networks, and location-aware communications. He is an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

PASCUAL MARTÍNEZ-GÓMEZ received the Ph.D. degree with the University of Tokyo, Japan. He was a Research Scientist with the Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), and an Assistant Professor with Ochanomizu University. He is currently an Applied Scientist with Amazon, Seattle, WA, USA. His main research interests include natural language processing, multi-modal user interfaces, and machine learning. In the past, he researched on machine translation, speech and image recognition, and readability diagnosis.

JALIL TAGHIA received the Ph.D. degree in electrical engineering from the KTH Royal Institute of Technology in 2014. From June 2014 to August 2015, he was a Postdoctoral Researcher with the Neural Information Processing Group, Technical University of Berlin. He is currently a Postdoctoral Researcher with the Cognitive and Systems Neuroscience Laboratory, Stanford University.

YI-ZHE SONG (Senior Member, IEEE) received the Ph.D. degree in computer vision and machine learning from the University of Bath in 2008. He was a Senior Lecturer with the Queen Mary University of London, and a Research and Teaching Fellow with the University of Bath. He is currently a Reader of Computer Vision and Machine Learning with the Centre for Vision Speech and Signal Processing (CVSSP), U.K.'s largest academic research center for artificial intelligence with approximately 200 researchers. He received the Best Dissertation Award from the M.Sc. degree with the University of Cambridge in 2004, after getting the First Class Honours degree from the University of Bath, in 2003. He is a Fellow of the Higher Education Academy. He is a Full Member of the review college of the Engineering and Physical Sciences Research Council (EPSRC), the U.K.'s main agency for funding research in engineering and the physical sciences. He serves as an Expert Reviewer for the Czech National Science Foundation.

HUIJI GAO received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications and the Ph.D. degree in computer science from Arizona State University, with a research focus on data mining and machine learning. He is currently an Engineering Manager leading the AI Algorithms Foundation team, LinkedIn. He has broad interests in machine learning/AI and its applications, including large-scale recommender systems, computational advertising, search ranking, natural language processing, and conversational AI. He has filed more than ten U.S. patents and published 40 publications in top-tier journals and conferences, including KDD, AAAI, WWW, CIKM, ICDM, SDM, and DMKD with thousands of citations.