MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
http://www.comsoc.org/~mmc

MMTC Communications - Frontiers

Vol. 16, No. 1, January 2021

CONTENTS

Special Issue on Emerging Technologies for Smart Applications

Guest Editor: Ye Liu and Yirui Wu***
*Nanjing Agricultural University, China**
*Hohai University, China***
Email: yeliu@njau.edu.cn, wuyirui@hhu.edu.cn

This special issue of Frontiers focuses on recent progress of emerging technologies for smart applications. The research topics in this special issue include intelligent noise mapping for a smart environment, agricultural robots for smart agriculture, artificial intelligence for smart monitoring, and deep learning for medical image analysis.

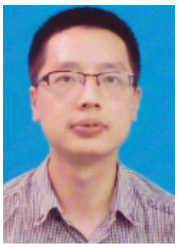
The first paper reviews intelligent noise mapping from the perspectives of computational models, software tools, and sound taxonomy. First, seven popular international models are introduced. Second, several commercial and open-source software tools are summarized regarding availability, features, and application scenarios. Finally, a classical urban sound taxonomy is presented.

The second paper summarizes the application status of machine vision in agricultural robots, summarizes the application of machine vision in agricultural production, and puts forward possible future research directions for the application of machine vision in agricultural robots in the future.

The third paper proposes a new spatial quality-control method based on the spatial regression test and random forest for identifying potential outliers in daily surface temperature observations. To evaluate the performance of the proposed method, a surface temperature dataset from different regions of China is applied. The results show that it improves the regression accuracy and reduces the runtime of the random forest through data refinement.

This fourth paper presents an end-to-end method for chest X-ray abnormality detection method using self-attention transformers, which doesn't rely on many hand-crafted components and doesn't have many pre and post processing steps. A dilated decoder layer was also designed to fuse context information from multiple scales. This method is extremely for the challenging abnormality detection task and its superior to some existing methods.

Ye Liu received M.S. and Ph.D. degrees in electronic science and engineering from Southeast University, Nanjing, Nanjing, China, in 2013 and 2018, respectively. He was a Visiting Scholar with Montana State University, Bozeman, MT, the USA, from October 2014 to October 2015. In addition, he was a visiting Ph.D. Student from February 2017 to January 2018 with the Networked Embedded Systems Group, RISE Swedish Institute of Computer Science. He is currently a Researcher with Nanjing Agricultural University, Nanjing, China. He has authored or co-authored papers in several prestigious journals and conferences, such as the IEEE Communications Magazine, IEEE Internet of Things Journal, ACM Transactions on Embedded Computing System, IEEE Network, INFOCOM, ICNP, EWSN, and ICC. His current research interests include wireless sensor networks, energy harvesting systems, and mobile crowdsensing. Dr. Liu was awarded 1st place in the EWSN Dependability Competition in 2019 and Macau Young Scholar in 2021.



Yirui Wu currently is an Associate Professor at Hohai University and working in the Hydrology Big Data Group. Before coming to Hohai, he obtained my Ph.D. degree from Nanjing University in 2016. He received my B.S. Degree from Nanjing University in 2011 as well. During my Ph. D study, he was with the IMAGE Lab under the supervision of Prof. Tong Lu and worked closely with Prof. Shivakumara Palaiahnakote. He visited Hong Kong University of Science and Technology twice in 2012 and 2014, supervised by Prof. Chiew-Lan Tai and Dr. Oscar Kin-Chung Au. He has published more than 40 journal/conference papers on Computer Vision, Pattern Recognition, Multimedia, and Intelligent Water Conservancy, including IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, Computer-Aided Geometry Design, Signal Processing: Image Communication, Expert Systems with Applications, ACM MM, ICME, BIBM, ICPR, ICDAR, ICIP, MMM, PCM and others. He has served as PC members for HPCC2021(CCF-C), CIKM2021(CCF-B), PIC2021, ACPR2021, DSAA2020, ML4CS2020, PIC2020, SECSOC2019 and others.

Intelligent Noise Mapping: Computational Models, Software Tools, and Sound Taxonomy

Ye Liu[1], Bangsong Du[1], Xing Yang[1], Lei Shu[1,2]
[1]Nanjing Agricultural University, China,[2]University of Lincoln, UK
yeliu@njau.edu.cn, bsongdu@gmail.com, harryyangx@gmail.com, lei.shu@ieee.org

1. Introduction

Noise pollution has risen to the second most significant environmental cause of public health problems next to air pollution. The official statistics [1], [2] from European Environment Agency show around 110 million people in Europe are exposed to harmful noise levels, leading to nearly 16,600 premature deaths each year. As many as 76 million of them suffer sleep disturbance at nighttime due to environmental noise. In addition, sufficient scientific studies [3], [4] demonstrate that noise pollution negatively impacts human health both psychologically and physiologically. For example, short-term low sleep quality causes irritability and nervousness. If one suffers sleep disturbance for a long time, it can result in depression, memory disorder, and neurological problems. Furthermore, animals and plants are also harmfully affected by noise pollution.

Many international organizations and governments have been making great efforts over the past years [5]. For example, the European Commission issues Directive 2002/49/EC (Environmental Noise Directive), which requires its members to produce noise maps and update them every five years. In October 2018, the World Health Organization (WHO) released new environmental noise guidelines [6] aimed at helping governments to evaluate the exposure of noise in our cities and make sustainable policy actions. The USA, China, Japan, and other countries also have been paying much attention to noise mapping.

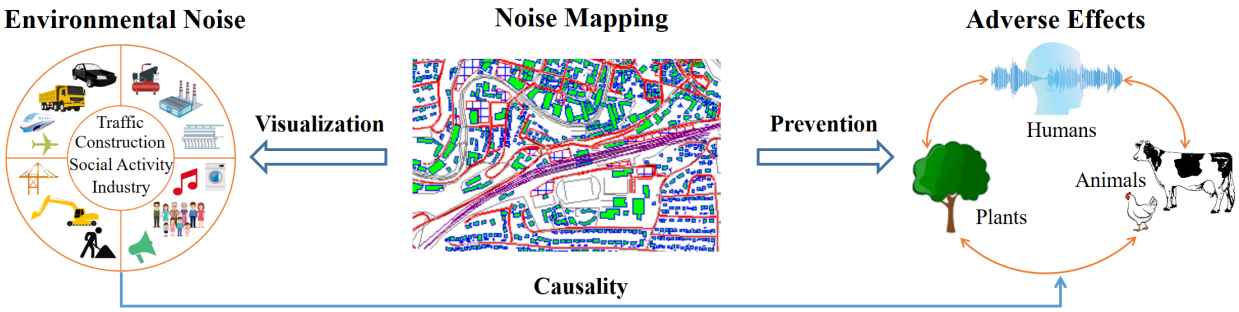


Figure 1. The relationship between environmental noise, adverse effects, and noise mapping: Environmental noise affects the whole ecosystems from humans, animals to plants in mental health, physiological health, and even physical health. Noise mapping can visualize the invisible disregarded sound noise and guide noise management action plans.

Noise mapping is usually required to assess the overall noise exposure in certain areas and the unique environmental noise emitted by road traffic, rail traffic, airports, and industrial activities. Moreover, it should present the number of people, dwellings, schools, and hospitals exposed to harmful noise levels. First, as shown in Figure 1, the invisible disregarded environmental noise can be visualized through a noise map, which is essential to gain a deep understanding of the sources, levels, and distribution of noise pollution. Secondly, the valuable map information can guide decision-makers to develop noise reduction plans precisely. Thirdly, public awareness can be raised through publishing urban noise maps to inform and consult the citizens about environmental noise exposure, harmful effect, and potential noise prevention measures. This is vital in the fight against noise pollution.

In recent years, a significant number of research and industrial works have contributed to this domain. The development of noise mapping technologies can be mainly divided into computational model-based noise calculation, noise measurement sensor network, and mobile crowdsensing approach. Each of the three methodologies has its advantages and limitations. In the future, they will complement each other and integrate into a cyber-physical system together with artificial intelligence, big data analytics, and cloud computing. Moreover, urban sound taxonomy is vital for intelligent noise mapping in the cyber-physical system. In the rest of this paper, we first review the proposed computational models and developed software tools for noise mapping. Then, sound taxonomy is discussed.

2. Computational Models

Model-based noise calculation is a kind of the main approach to generating noise maps. The standard modes include the ISO 9613-2, HARMONOISE/IMAGINE, Nord2000 (Nordic noise prediction method), NMPB-Routes-2008 (the standard method for road traffic noise calculation in France), ASJ RTN-Model (Acoustical Society of Japan Road Traffic Noise Model), CNOSSOS-EU (common noise assessment methods in EU), FHWA TNM (Federal Highway Administration Traffic Noise Model), and so on.

An example is given below. In the CNOSSOS-EU model, the frequency range for noise calculation is from 63 Hz to 8 kHz. The sound power level emitted by a category m vehicle at i th frequency band ($L_{W,i,m}$) is expressed as:

$$L_{W,i,m}(v_m) = 10 \times \lg(10^{\frac{L_{WR,i,m}(v_m)}{10}} + 10^{\frac{L_{WP,i,m}(v_m)}{10}}) \quad (1)$$

where v_m is the speed of the vehicle. $L_{WR,i,m}$ is the rolling noise caused by tyre/road interaction, and $L_{WP,i,m}$ is the propulsion noise produced from the engine, exhaust, and other driveline components. They are defined as:

$$L_{WR,i,m} = A_{R,i,m} + B_{R,i,m} \times \lg(\frac{v_m}{v_{ref}}) + \Delta L_{WR,i,m} \quad (2)$$

and

$$L_{WP,i,m} = A_{P,i,m} + B_{P,i,m} \times \frac{(v_m - v_{ref})}{v_{ref}} + \Delta L_{WP,i,m} \quad (3)$$

in which both $A_{R,i,m}$ and $B_{R,i,m}$ are reference coefficients. $\Delta L_{WR,i,m}$ is the sum of correction coefficients for specific road surface ($\Delta L_{WR,road,i,m}$), different vehicle conditions including studded tyres ($\Delta L_{WR,studdedtyres,i,m}$), speed variation ($\Delta L_{WR,acc,i,m}$) and air temperature ($\Delta L_{W,temp}$). Similarly, $A_{P,i,m}$ and $B_{P,i,m}$ correspond to reference coefficients of propulsion noise emission. $\Delta L_{WP,i,m}$ is the sum of correction coefficients in different regions ($\Delta L_{WP,road,i,m}$) and driving conditions ($\Delta L_{WP,grad,i,m}$ and $\Delta L_{WP,acc,i,m}$). Finally, the reference speed (v_{ref}) is equal to 70 km/h.

For the attenuation of noise during propagation, it needs to calculate the sound attenuation in favorable conditions (L_F), inhomogeneous conditions (L_H), and the long-term sound level for each path (L_{LT}). Then, the noise at the receptor ($L_{tot,LT}$) can be obtained by accumulating the long-term sound levels from all paths. The equations defined in the CNOSSOS-EU model are as follows:

$$L_{tot,LT} = 10 \times \lg(\sum_n 10^{\frac{L_{n,LT}}{10}}) \quad (4)$$

$$L_{LT} = 10 \times \lg(p \cdot 10^{\frac{L_F}{10}} + (1-p) \cdot 10^{\frac{L_H}{10}}) \quad (5)$$

$$L_F = L_{W,i,m} - (A_{div} + A_{atm} + A_{boundary,F}) \quad (6)$$

$$L_H = L_{W,i,m} - (A_{div} + A_{atm} + A_{boundary,H}) \quad (7)$$

where n is the total number of paths between sound emitter and receptor. p is the occurrence weight for favorable conditions and homogeneous conditions. A_{div}, A_{atm}, $A_{boundary,F}$, and $A_{boundary,H}$ is sound attenuation due to geometrical divergence, atmospheric absorption, and boundary of propagation medium in favorable conditions or homogeneous conditions, respectively. The detailed description of the CNOSSOSEU model for road, railway, and industry noise calculation can be found in Commission Directive (EU) 2015/996/EC [7].

3. Software Tools

The abovementioned models can well represent environmental noise. However, it is complicated and time-consuming to calculate them. To address the problem, many commercial and open-source software tools have been developed in these years. We summarize these software tools in TABLE I.

The Predictor-LimA software suite has now been updated to version 2019, which is released in October 2018. The developed LimA calculation cores have the ability of high capacity and breakneck computing speed so that large-scale noise maps can be generated quickly. The unique geometry analysis method has good performance for 3D noise mapping. With the integrated bookkeeping function, new input data can be imported at any time when the calculation

program is running to ensure the noise map keeps up-to-date. A Geographic Information System (GIS) module is integrated into the software for locally demographic analysis. The Predictor-LimA includes 30 noise calculation models, such as CNOSSOS-EU, HARMONOISE, ISO 9613, and NMPB-Routes-2008.

Another popular software is CadnaA, which provides noise calculation and result presentation. More than 30 calculation models are implemented to assess the noise of urban areas, traffics, and industrial plants. The Program Controlled Segmented Processing (PCSP)-based distributed calculation and multithreading support allows CadnaA to handle large-scale and fine-grained noise mapping. In addition, a dynamic noise map with predefined update intervals is available. CadanA is also able to connect with other GIS systems, where GIS provides the input data required in noise models as well as generating noise maps, and CadnaA is responsible for noise calculation.

SoundPLAN, noise3D, IMMI, and D-noise are some other commercial noise calculation software tools. Each of them has unique features. The modular design of SoundPLAN allows flexible packages combinations for customers. The multi-document application is beneficial for multi-sources input and results representation. The noise3D simulation tool provides an on-demand license service so that the cost is meager. The IMMI software can calculate shipping noise. The D-noise solution import the source data to a pre-structured database for data standardization and processing.
Apart from commercial software, researchers have been developing open-source tools for academia to study environmental noise [8]. NoiseModelling is an open-source software. It can calculate traffic noise with the NMPBRoute-2008 method. Industrial sound sources application is also available in the NoiseModelling tool. The opeNoise, a QGIS plugin, is another free tool for road traffic noise assessment. In summary, both commercial software and open-source ones are becoming increasingly powerful, that help researchers to conduct environmental noise studies more efficiently.

TABLE I. THE COMMERCIAL AND OPEN-SOURCE SOFTWARE TOOLS FOR NOISE MAPPING

Software Tool	Availability	Feature	Application Scenarios
Predictor-LimA	Commercial	LimA calculation Integrated GIS functionality Unique 3D geometry analysis Time-saving integrated bookkeeping 30 calculation models	Road, Railway, Aircraft Industry, Wind turbine
CadnaA	Commercial	Distributed calculation and multithreading Dynamic noise map with 3D visualization Connect to other GIS systems 30+ calculation models	Road, Railway, Aircraft Industry
SoundPLAN	Commercial	ArcView GIS interface Flexible packages customization Multi-document applications Distributed calculation and multithreading 40+ calculation models	Road, Railway, Aircraft Industry
Noise 3D	Commercial	Affordable and low cost 3D visualization	Road, Industry
IMMI	Commercial	Multiple input data formats Editing and visualizing objects Special functions for industrial noise 30+ calculation models	Road, Railway, Aircraft Industry, Shipping
D-noise	Commercial	Esri GIS-based system Pre-structured database and standardized input data Configurable reporting tool Support sonRAIL and sonROAD model	Road, Railway
NoiseModelling	Open-source	OrbisGIS plugin Support NMPB 2008 model	Road, Industry
OpeNoise	Open-source	QGIS plugin and 2D modelization Lateral diffraction and Flat terrain	Road

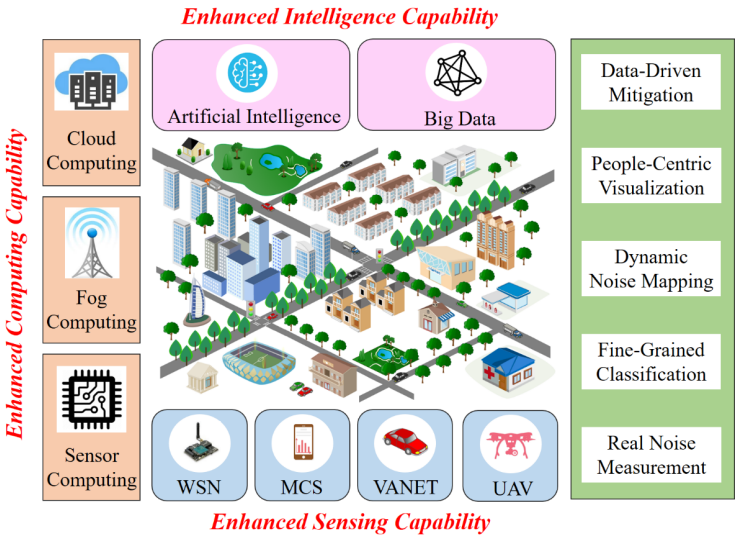


Figure 2. The paradigm of future intelligent noise mapping.

4. Urban Sound Taxonomy

Ubiquitous noise sensing and recognition can come true in the paradigm of future intelligent noise mapping [9]. As shown in Figure 2, all local acoustic sensor networks, mobile devices, edge nodes, and cloud platforms can perform this task. The deployment cost of an acoustic sensor network will reduce if part of data processing is assigned on the edge or cloud. In addition, sound data pre-processing on acoustic sensor networks could mitigate the burden of multimedia communications.

Sensor nodes usually compute the values of sound pressure level after capturing the sound noise. However, each noise value combines multiple sound sources such as road traffic noise, wind, and the human voice. To provide a precise noise map, it is essential to recognize sound sources and calculate their corresponding sound pressure levels. Moreover, these sound clips are valuable to both public and researchers. For example, the public could decide if they are willing to go to the hot spots. In addition, researchers can use these data for security prediction and social problem analysis.

The lack of a common sound taxonomy and real-world annotated sound datasets are the two barriers to sound pattern recognition. To address these issues, New York University proposes an urban sound taxonomy [10]. The taxonomy divides urban acoustic environments into four broad categories: human, nature, mechanical activities, and music. The four high-level semantic classes then go down to specific sound sources as detailed as possible. Figure 3 shows the first three layers of the urban sound taxonomy. Based on the taxonomy, two free large urban sound datasets (URBANSOUND and URBANSOUND8K) are then created, that contain 1302 and 8732 annotated recordings respectively. The datasets can be downloaded from the official website (https://urbansounddataset.weebly.com).

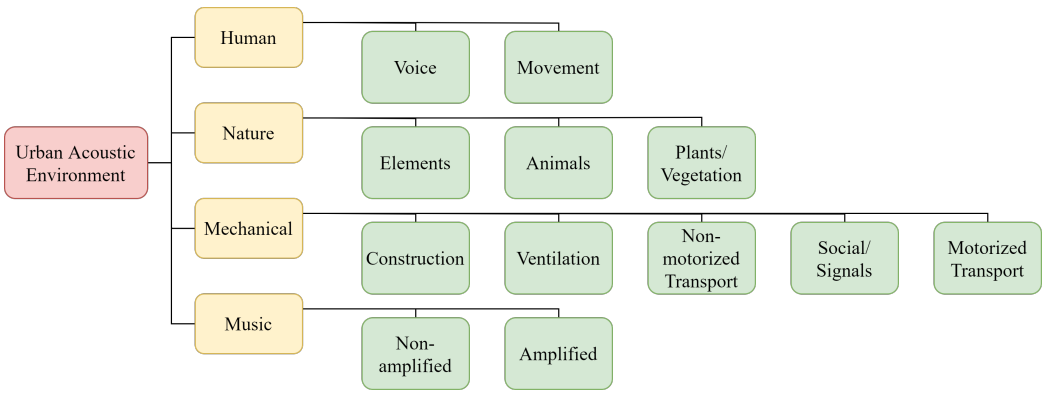


Figure 3. Urban sound taxonomy for intelligent noise mapping.

5. Conclusions

This paper presents a review of intelligent noise mapping from the perspectives of computational models, software tools, and sound taxonomy. Seven popular international models are introduced, along with a detailed illustration of the CNOSSOS-EU model. Next, some commercial and open-source software tools are summarized in terms of availability, features, and application scenarios. Finally, a classical urban sound taxonomy is discussed, as well as two helpful sound datasets, namely URBANSOUND and URBANSOUND8K.

References

[1] "2018 revision of world urbanization prospects," United Nations, May 2017, Accessed: Jan. 16, 2021. [Online]. Available: https://population.un.org/wup/

[2] "Population exposure to environmental noise," European Environment Agency, Nov. 2018, Accessed: Jan. 16, 2021. [Online]. Available: https://www.eea.europa.eu/data-and-maps/indicators/exposure-to-and-annoyance-by-2/assessment-4

[3] M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, and S. Stansfeld, "Auditory and non-auditory effects of noise on health." Lancet, vol. 383, no. 9925, pp. 1325–1332, 2014.

[4] S. A. Stansfeld and M. P. Matheson, "Noise pollution: non-auditory effects on health," British Medical Bulletin, vol. 68, no. 1, pp. 243–257, 2003.

[5] Y. Liu, L. Shu, Z. Huo, K. F. Tsang, and G. P. Hancke, "Collaborative Industrial Internet of Things for Noise Mapping: Prospects and Research Opportunities," in IEEE Industrial Electronics Magazine, vol. 15, no. 2, pp. 52-64, June 2021.

[6] "Environmental noise guidelines for the European region(2018)," World Health Organization Europe, Oct. 2018. [Online]. Available: http://www.euro.who.int/en/publications/abstracts/environmental-noise-guidelines-for-the-european-region-2018

[7] E. Directive, "Establishing common noise assessment methods according to directive 2002/49/EC of the European Parliament and the council," Official Journal of the European Union, 2015.

[8] J. Gulliver, D. Morley, D. Vienneau, F. Fabbri, M. Bell, P. Goodman, S. Beevers, D. Dajnak, F. J. Kelly, and D. Fecht, "Development of an open-source road traffic noise model for exposure assessment," Environmental Modelling & Software, vol. 74, pp. 183–193, 2015.

[9] Y. Liu et al., "Internet of Things for Noise Mapping in Smart Cities: State of the Art and Future Directions," in IEEE Network, vol. 34, no. 4, pp. 112-118, July/August 2020.

[10] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 1041–1044.

Ye Liu received M.S. and Ph.D. degrees in electronic science and engineering from Southeast University, Nanjing, Nanjing, China, in 2013 and 2018, respectively. He was a Visiting Scholar with Montana State University, Bozeman, MT, the USA from October 2014 to October 2015. He was a visiting Ph.D. Student from February 2017 to January 2018 with the Networked Embedded Systems Group, RISE Swedish Institute of Computer Science. He is currently a Researcher with Nanjing Agricultural University, Nanjing, China. He has authored or co-authored papers in several prestigious journals and conferences, such as the IEEE Communications Magazine, IEEE INTERNET OF THINGS JOURNAL, ACM Transactions on Embedded Computing System, IEEE NETWORK, INFOCOM, ICNP, EWSN, and ICC. His current research interests include wireless sensor networks, energy harvesting systems, and mobile crowdsensing. Dr. Liu was awarded 1st place in the EWSN Dependability Competition in 2019 and Macau Young Scholar in 2021.

Bangsong Du received a B.S. degree in engineering from Anhui Agricultural University, China, in 2018. He is currently working toward an M.S. degree in the college of engineering, Nanjing Agricultural University, China. His current research interests include continuous objects detection and tracking for wireless sensor networks and battery-free IoT.

Xing Yang received an M.S. degree in control engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2018. He is currently working toward a Ph.D. degree in the college of engineering, Nanjing Agricultural University. His current research interests include fault diagnosis in wireless sensor networks, agricultural Internet of Things reliability, and machine learning algorithm.

Lei Shu received the B.S. degree in computer science from South Central University for Nationalities, Wuhan, China, in 2002, the M.S. degree in computer engineering from Kyung Hee University, Seoul, South Korea, in 2005, and the Ph.D. degree in computer engineering from the Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland, in 2010. Until 2012, he was a Specially Assigned Researcher with the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. He is currently a Distinguished Professor with Nanjing Agricultural University, Nanjing, China, and a Lincoln Professor with the University of Lincoln, Lincoln, U.K. He is also the Director of the NAU-Lincoln Joint Research Center of Intelligent Engineering, Nanjing, China. Dr. Shu was on multiple Editorial Boards, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE Communications Magazine, IEEE Network Magazine, IEEE SYSTEMS JOURNAL, and IEEE/CCA JAS.

Recent Advance on Machine Vision for Agricultural Robots

Mingying Chen[1] and Lei Shu[1,2]
[1]*School of Artificial Intelligence, Nanjing Agricultural University, China*
[2]*University of Lincoln, UK*
ming@stu.njau.edu.cn, lei.shu@ieee.org

1. Introduction

Machine vision is a subject based on the engineering of computer vision theory. It is an interdisciplinary and comprehensive subject involving optical imaging, visual information processing, artificial intelligence, and mechatronics [1]. Machine vision can significantly improve the accuracy and real-time operation of the machine, the efficiency of work, the correctness of decision-making, which has brought further development of the machine, driving the machine automation, further intelligent innovation. The mature machine vision technology has become one of the important, indispensable technologies for social development. The agricultural robot is a flexible automatic or semi-automatic device that can be programmed repeatedly and takes agricultural products as the operation object and has the functions of human partial information perception and limb movement [2].

The agricultural robot is the most crucial part of modern agriculture. It combines mechanical, electronic, image processing, automatic control, artificial intelligence, computer technology, and agricultural machinery's main development direction [3]. Agricultural robots are mainly divided into walking series robots and manipulator series robots. In the part of information perception required by agricultural robots, the visual information provided by machine vision is an important part. The basic machine vision system, including image module, digital module and image processing, intelligent decision-making module [4]. The elemental composition of the machine vision system is shown in Fig.1.

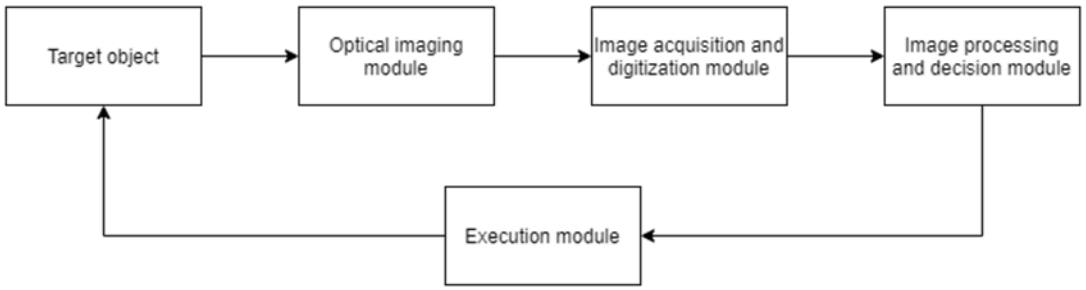


Fig. 1 Basic structure of machine vision system

2. Application of Machine Vision in Agricultural Robot

2.1 Application of Machine Vision in Agricultural Robot Navigation System

Walking agricultural robots often need automatic navigation functions in the field. Machine vision is one of the critical technologies in agricultural robot automated navigation systems. In the past, automated navigation technology, GPS navigation using absolute coordinates occupied an important position, which is reliable and can be used for large-scale path planning. However, the machine vision with relative coordinates is more flexible, real-time, and accurate in the complex agricultural environment. Through real-time monitoring and image processing, it provides necessary information support for the automatic navigation system. At the same time, it can also collect more accurate information about crops, pests, and diseases, to serve precision agriculture [5,6]. Therefore, in the complex environment, the construction of a stable and transparent visual perception platform can improve the comprehensive perception and analysis ability of agricultural robots for the harsh operating environment and target objects, realize the accurate identification of geographical markers, obstacles, crops, and other targets in the static and moving process of agricultural robots, and make up for the defects of the inertial navigation system. The structure of the machine vision navigation system in agricultural robots is shown in Fig. 2.

In 2013, Wang Baoliang [7] of Nanjing Agricultural University developed a field walking spraying robot based on an open system structure. In this study, the vertical projection method was used to locate the centerline of the crop row, and the crops and weeds were separated by image pre-processing, Canny operator edge extraction, morphological operation, and other operations. The actuator was controlled, and finally, the robot's autonomous

walking and drug spraying were realized. The average deviation of field walking was 38.5 mm. The robot must rely on manual real-time adjustment parameters to effectively identify the crop row centerline. Therefore, it isn't easy to ensure the continuity of walking and spraying operations, and there is still room for improvement in automation and intelligence. It can be seen that compared with industrial robots, the working environment of agricultural robots is more complex and different. Under such conditions, automatic navigation technology relying on machine vision still has particular challenges.

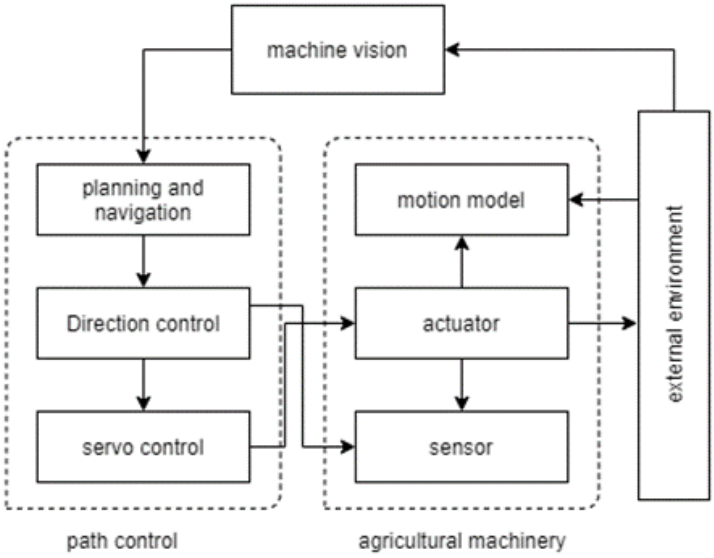


Fig. 2 Structure of agricultural machinery machine vision navigation system

2.2 Application of Machine Vision in Agricultural Robot Target Recognition

Manipulator agricultural robots usually need to operate after positioning the target, including some walking robots, such as fertilization robots, welding robots, and spraying robots. In the past, these links were allowed to be carried out in large areas without a difference, but under the requirements of precision agriculture, it is also necessary to identify and locate the target. Furthermore, most crops have distinct color and shape characteristics, which sharply contrast with the background, so the target and environment can be separated by the algorithm [8]. Therefore, the traditional machine vision technology is to identify the color, shape, and size differences of similar objects through image recognition to assist agricultural robots in positioning objects. In practical applications, recognition and positioning accuracy and processing speed are required [3].

Kondo-N [9,10] et al. developed a picking robot. The color camera was used to collect the image, and the fruit was segmented from the image background, and the number of fruit was identified by thresholding and filtering. Then, the 3D information of fruit was located by binocular stereo vision. Experiments show that the robot's fruit positioning success rate is high, but affected by the complex environment, obstacle avoidance picking completion is low, resulting in the overall picking success rate being affected. It can be seen that the automatic picking robot based on machine vision can reasonably identify the fruits and vegetables with a prominent 'color' signal. Still, it is challenging to locate the picking point due to the randomness of the growth position and direction of fruits and vegetables [4].

For a single feature, sometimes difficult to accurately identify the target, need to combine a variety of features to improve fruit recognition accuracy. For example, in texture analysis, Hayes J C [11] uses a gray-level co-occurrence matrix to detect plants. It is found that when the color characteristics of plants and background are not prominent, the segmentation effect of texture analysis is better than that of color analysis.

In addition, in recent years, in the image recognition of agricultural robots, target recognition based on deep learning has been introduced. Deep learning uses multiple feature layers for deep convolution, automatically learning image features from large amounts of data and extending them to actual scenes. For example, Fu et al. [12] used the backpropagation and stochastic gradient descent technology of the Zeiler and Fergus network (ZFNet) to train the Faster Convolutional Neural Network (R-CNN) end-to-end. The recognition rate of single kiwifruit

was 94.3%, and the recognition rates of occluded fruits and overlapping fruits were 82.5 % and 85.6 %, respectively. The model has a short processing time and good robustness to light variance and leaf occlusion.

3 Problems and research directions of machine vision in agricultural robot application

Compared with the industrial environment, the agricultural environment is more complex. There are considerable differences in different regions and even many differences in various locations in the same area. The characteristics of the target objects are different. At the same time, the operation of agricultural robots also needs to consider the influence of crop planting methods, crop morphology, and other aspects on robot operation. Therefore, there are still some problems in the application of machine vision in agricultural robots. In the future, the application research of machine vision of farming robots may focus on the following aspects:

(1) Accelerate dynamic positioning accuracy and timeliness. Objectively due to the complexity of the agricultural environment, so in the current machine vision in the navigation system of farming robots, there are still low navigation accuracy and slow processing speed, which leads to the overall work efficiency of agricultural robots being affected, even lower than the manual work efficiency, which is very unfavorable for the application and promotion of agricultural robots. Therefore, improving the accuracy and timeliness of visual navigation in agricultural scenes will be one of the research focuses.

(2) Three-dimensional vision technology. In agricultural scenes, it is sometimes not enough to obtain only two-dimensional images for the work of agricultural robots. For example, in the work of using the manipulator to extract the target object, the target object needs more accurate spatial positioning coordinates, and the performance of the two-dimensional visual system is easily affected by factors such as light. Three-dimensional vision technology can accurately locate the target object, accurately restore the target position, direction, and depth of information for agricultural work to provide more accurate location information.

(3) Target recognition algorithm based on multi-feature fusion. In the actual agricultural environment, only a single feature is used for recognition. External lighting conditions such as weather and light sources may often affect the recognition rate of the target. The multi-feature recognition algorithm can better utilize the advantages of different target features and improve the target recognition rate.

(4) Embedded vision system. The embedded vision system has the characteristics of compact structure, fast processing speed, and low cost, which will provide significant impetus for the development of machine vision systems in the agricultural background.

4. Conclusion

In this paper, the application of machine vision in agricultural robots is summarized from two aspects of automatic navigation and target recognition and positioning. Although it does not cover all the literature, the representative methods and applications are reviewed. According to these applications, some problems in the application are analyzed. Finally, the feasible and meaningful research direction of machine vision in agricultural robot application is discussed.

References

[1] SONF C H, PENG X Z. A summary of the research and development of machine vision[J]. Equipment Manufacturing Technology, 2019 (6): 213-216 (in Chinese).

[2] Wang Rujing, Sun Bingyu. Development status and expectation of agricultural robot[J]. Bulletin of Chinese Academy of Sciences, 2015, 30(6): 803 - 809. (in Chinese with English abstract)

[3] ZHAO N, ZHAO P, GAO Y. Study on Application of Machine Vision Technology to Modern Agriculture in China[J]. Journal of Tianjin Agricultural University, 2015: 02.

[4] Chen Ying. Study on development status and application trends of machine vision technology [J]. Wireless Internet Technology,2018,15(19):147-148. (in Chinese with English abstract)

[5] Diao Zhihua, Wang Huidan, Wei Wei. Summary of research on machine vision application in agricultural production[J]. Journal of Agricultural Mechanization Research, 2014, 36(3): 206 - 211. (in Chinese with English abstract)

[6] Yang Weimin, Li Tianshi, Jia Hengshe. Simulation and experiment of machine vision guidance of agriculture vehicles[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2004, 20(1): 160 - 165. (in Chinese with English abstract)

[7] Wang Baoliang, Development Of A multifunctional and Autonomous Agricultural Robot [D].Nanjing. Nanjing Agricultural University, 2013.

[8] Wu Rongda，Chen Tianci，Zhang Shiang, et al. Research Progression Key Technologies of Fruit and Vegetable Picking Robots [J]. Mechanical & Electrical Engineering Technology ，2021,50(09):128-132.

[9] Kondo N， Nishitsuji Y， Ling P， et at．Visual feedback guided robotic cherry tomato harvesting [J] ．Trans of the ASAE，1996(39): 2331 - 2338．

[10] Kondo N ，Monta M ，Fujiura T ． Fruit harvesting robots in Japan [J] ．Advances in Space Research， 1996， 18(1 - 2): 181 - 184．

[11] Hayes J C，Han Y J．Comparison of crop covers measuring system．ASAE Paper，St．Joseph，M ich．，1989

[12] Fu L, Feng Y, Majeed Y, et al. Kiwifruit detection in field images using Faster RNN with ZFNet[J]. IFAC-PapersOnLine,2018, 51(17):45-50.

Mingying Chen received a B.E. degree in the College of Engineering from Nanjing Agriculture University, Nanjing, China in 2020. He currently studies for a master s degree at the School of Artificial Intelligence, Nanjing Agricultural University. His research interests include deep learning and its applications.

Lei Shu received the B.S. degree in computer science from South Central University for Nationalities, Wuhan, China, in 2002, the M.S. degree in computer engineering from Kyung Hee University, Seoul, South Korea, in 2005, and the Ph.D. degree in computer engineering from the Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland, in 2010. Until 2012, he was a Specially Assigned Researcher with the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. He is currently a Distinguished Professor with Nanjing Agricultural University, Nanjing, China, and a Lincoln Professor with the University of Lincoln, Lincoln, U.K. He is also the Director of the NAU-Lincoln Joint Research Center of Intelligent Engineering, Nanjing, China. Dr. Shu was on multiple Editorial Boards, including the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE Communications Magazine, IEEE Network Magazine, IEEE SYSTEMS JOURNAL, and IEEE/CCA JAS.

A Quality Control Method for Daily Surface Temperature Observations via Spatial Regression and Random Forest Methods

Xuanchen Guo[1], Xing Yang[1,2], Lei Shu[1,2], Ye Liu[1]
[1]Nanjing Agricultural University, China,[2]University of Lincoln, UK
guoxuanchen1996@163.com, harryyangx@gmail.com, lei.shu@njau.edu.cn, yeliu@njau.edu.cn

1. Introduction

The continually increasing number of meteorological observation stations produces increasing amounts of meteorological data. These meteorological data are used to analyze the possible impacts of climate change on our environment, validate climate model simulations and provide initial conditions for numerical weather prediction [1]. However, surface observations are easily affected by the migration of stations and instrument failure; such factors influence the accuracy of surface observations and, ultimately, weather research. Therefore, quality control of surface observations is essential, especially the surface air temperature, because these have a more significant impact on people's daily lives [2]. Quality control is an essential part of data acquisition, transmission, and processing and is a precondition of international meteorological business communication.

Related work. Quality control methods for a single station mainly include extreme checks, internal consistency checks, and temporal outlier checks. These quality control methods for a single station strongly depend on the integrity of the observations, which is based on the time series used for quality control; quality control methods for a single station are not effective if observations are missing for an extended period. Recently, multiple stations in quality control have proven helpful for an increasing number of observation stations. Frequently used quality control methods for various stations mainly include the spatial regression test [3] (SRT) method, the inverse distance weighting (IDW) method, and polynomial interpolation (OI); SRT and IDW methods are the most widely used methods. The main idea of the SRT and IDW methods is to estimate the value of the target station based on the neighboring stations. The difference between the IDW method and the SRT method is the choice of weights between different adjacent stations. The SRT method is superior to the IDW method because the root-mean-square error is more effective for determining the weights of neighboring stations. In addition, the OI and SRT methods are consistent with the statistical information needed for quality control, as OI calculates the background field and uses a two-error covariance matrix to analyze the sensitivity of the target station observations based on statistical surface meteorological observations and the background field.

Contribution. Most of these traditional QC methods are based on mathematical or statistical methods. When mathematical or statistical methods fail to explain why the performance of a QC method for one station is affected by the terrain and climate, intelligent algorithms may be used instead. For example, the SRT method is still affected by the geographical environment of the target station and neighboring stations, but there is no specific formula to explain such impacts. SRT offers the advantage of selecting the neighboring stations with the most minor error and weighting neighboring stations according to historical data. Still, it is unable to use the regression method incorporating the observations of neighboring stations. Therefore, this paper proposes combining the random forest algorithm [4] (RF), which is more efficient for regression, with the SRT method, which is used to pre-process the dataset. Then, the dataset is trained by the RF method, and the estimated value of the target station is obtained using the SRF method.

2. Method

Spatial Regression Test Algorithm. The SRT assigns weight according to the root-mean-square error between the target station and neighboring stations. Then, for each adjacent station, a linear regression based on an estimate was used:

$$x_i = a_i + b_i y_i \tag{1}$$

where x_i is the estimate of the target station, the data of ith neighboring station (i=1, 2, …, n) is y_i, a_i and b_i are coefficients, and the weighted estimate x' is obtained by using the standard error of estimate s:

$$x' = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 s_i^{-2}}{\sum_{i=1}^{N} s_i^2}} \tag{2}$$

where N is the number of neighboring stations to be used in the estimate.

Then, the weighted standard error of estimate s' is calculated as follows:

$$s'^{-2} = N^{-1} \sum_{i=1}^{N} s_i^{-2} \tag{3}$$

The confidence intervals are formed as follows:

$$x' - fs' \le x \le x' + fs' \tag{4}$$

where f is the quality control parameter. If the relation in (4) holds, the observations pass the test.

Spatial Random Forest Algorithm. The RF algorithm is a combination of tree predictors, such that each tree depends on the values of a random vector sampled independently with the same distribution for all trees in the forest [4]. The RF algorithm grows an ensemble of trees. Each node split is selected randomly from among the best divisions; hence, it has a strong generalization ability and avoids over-fitting. In the surface meteorological observations, not all neighboring stations have a strong correlation with target stations, and the adjacent stations with weak correlation are equivalent to the invalid input of the datasets. Data types with weak inputs are complex for typical classifiers, such as neural nets and trees; thus, the RF algorithm is more suitable for surface meteorological observations. By using random feature selection in addition to bagging, the generalization error is estimated by out-of-bag (OOB) estimation, obtaining concrete results from otherwise theoretical values of strength and correlation.

The quality control model was constructed using the SRT and RF algorithm, and the SRF method can be divided into the following steps. First, dataset L is divided into the training sample L_{train} and testing sample L_{test}. The RMSE of ith neighboring stations s_i is calculated by using Formulas (1), according to the weighting coefficient, to calculate the new dataset L':

$$L'_i = \frac{s_i}{\sum_{i=1}^{N} s_i} L_i \tag{5}$$

where N is the number of neighboring stations to be used in the new dataset. Then, the RF regression model is used to train and regress the new dataset. Ultimately, the values predicted by the SRF method y_{est} are compared with observations of the target station y_{obs} that have inserted artificial errors. Finally, coefficient f is used to test whether the observed values fall within the confidence intervals:

$$\|y_{est} - y_{obs}\| \le f\sigma \tag{6}$$

If the observations of the target station fall within the confidence intervals, the observations pass the SRF test.

Figure 1. The spatial correlation of 14 target stations and the number of their neighboring stations.

3. Data Description

Daily mean temperature observations for 14 target stations from 2005 to 2014 were selected. The information for the 14 target stations is shown in Fig. 1, where a five-pointed star represents the stations and the number next to the stations' name indicate the number of neighboring stations, and the spatial correlation between the target station and neighboring stations are demonstrated as "high," "low" and "unknown." The 14 target stations are abbreviated as follows: Beihai (Bh), Chengdu (Cd), Guangzhou (Gz), Haikou (Hk), Hohhot (Hh), Jinghong (Jh), Lhasa (Ls), Lanzhou (Lz), Miyun (My), Mohe (Mh), Nanjing (Nj), Taiyuan (Ty), Urumqi (Um) and Changchun (Cc). Their neighboring stations were selected from those within 200 km.

To test the performance of the SRF method, artificial errors were randomly inserted into the observations of the target station. Approximately 3% of observations were selected for the insertion of random errors, and the formula is shown as:

$$K_\lambda = s_\lambda p_\lambda \quad (7)$$

where k is the value of the insertion error, s is the standard deviation of the observations for the target station. λ is the position for error insertion, and p is a random number with uniform distribution with a range of ±3.5.

4. Results

Spatial Correlation Analysis. For the study, daily mean temperature observations from 2005 to 2013 of 14 target stations and their neighboring stations were selected as a training sample. The 2014 observations were chosen as the testing sample. It was necessary to analyze the spatial correlation of the 14 target stations and neighboring stations because the spatial correlation of all stations in a region within 200 km may impact the performance of the quality control model. As shown in Table 1, the results of the spatial correlation were calculated with a semi-variogram and Moran's I, where the number of Mh and its neighboring stations is too small; thus, analyzing the semi-variogram in Mh was not possible. When R^2 and I are close to 1, a smaller *RSS* and larger *z-value* are associated with the higher spatial correlation between stations. It is clear that there are high spatial correlations of Bh, Cd, Gz, Hh, My, Ty, and Cc, while the spatial correlations of Jh, Ls, Lz, Mh, and Um are very low. An assessment of the different methods for different target stations showed that spatial correlation does impact the quality control of temperature observations.

Table 1. Spatial correlation indexes in the regions of 14 target stations.

	Co	Co+C	Ao	R^2	RSS	I	E(I)	mean	sd	z-value
Bh	0.101	0.715	1.17	0.848	0.095	0.649	-0.031	-0.032	0.097	7.031
Cd	0.153	2.398	1.22	0.849	1.2	0.721	-0.015	-0.012	0.071	10.39
Gz	0.188	1.163	2.07	0.922	0.076	0.581	-0.023	-0.023	0.093	6.47
Hk	0.157	0.781	1.53	0.804	0.085	0.478	-0.037	-0.037	0.118	4.348
Hh	0.004	2.85	0.61	0.843	2.72	0.545	-0.029	-0.026	0.112	5.096
Jh	0.001	2.385	0.3	0.326	13.3	0.162	-0.091	-0.083	0.158	1.556
Ls	0.01	6.4	0.55	0.385	79.5	-0.08	-0.1	-0.103	0.179	0.125
Lz	0.83	4.403	0.55	0.489	4.4	0.143	-0.24	-0.023	0.094	1.178
My	0.51	9.029	1.48	0.876	6.55	0.729	-0.013	-0.013	0.065	11.43
Mh	/	/	/	/	/	-0.411	-0.333	-0.349	0.29	-0.211
Nj	0.145	0.606	4.01	0.355	0.071	0.552	-0.12	-0.013	0.066	8.602
Ty	0.001	4.516	1.8	0.929	2.25	0.648	-0.012	-0.013	0.067	9.809
Um	0.01	11.58	1.54	0.371	375	0.446	-0.046	-0.046	0.123	4.014
Cc	0.269	1.069	1.38	0.736	0.287	0.615	-0.027	-0.027	0.104	6.174

Spatial Sensibility Analysis. *Several selected neighboring stations:* In general, ten adjacent stations were chosen as the reference stations for prediction by using the SRT method. However, it was unknown whether prediction would be improved with the 10 reference stations when the SRT method was combined with the RF method. Therefore, it was necessary to identify the appropriate number of selected neighboring stations to determine whether observations of reference stations should be weighted. Fig. 2(a-c) shows the performance of the RF, SRT, and SRF methods when 5, 10, 15, and 20 neighboring stations with the lowest standard error were selected as reference stations, where reference stations represented as SRF5, SRF10, SRF15, and SRF20.

The performances of the SRF and RF methods were found to be superior to the SRT method, as the SRF method required less time to run than the RF method. To achieve improved quality control, 15 reference stations were selected and weighted according to performance and runtime. Since the number of neighboring stations in Jh, Ls, and Mh was less than 15, the three target stations were tested separately, and the results are shown in Fig. 2(d-f), where the values of MAE in Jh, Ls, and Mh are 3.448, 2.747 and 1.276 and the importance of RMSE in Jh, Ls and Mh are 3.555, 2.927 and 1.736. The results show that the SRF and RF methods have better performance than the SRT method in regions with a low density of neighboring stations. Moreover, the MAE and RMSE of the SRF method were much lower than those of the SRT method, and this is also consistent with Hubbard's description of the SRT method, which does not apply to stations with few neighboring stations.

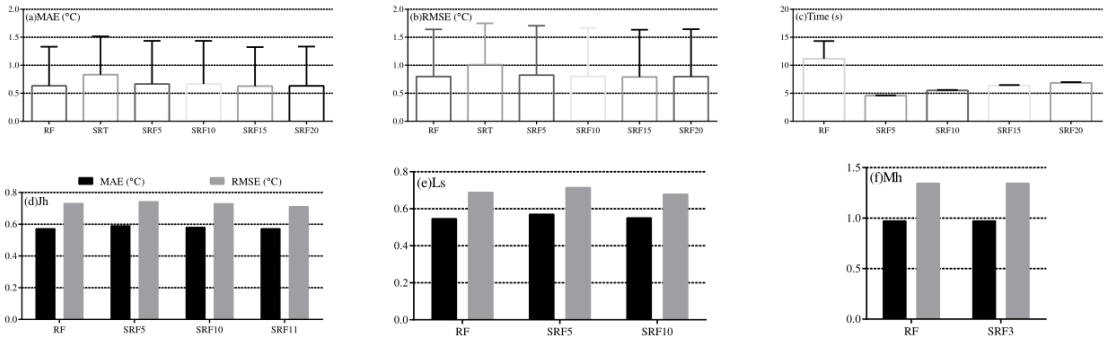


Figure 2. Performance of mean of the RF, SRT, and SRF methods: (a) MAE, (b) RMSE, (c) Time and performance of the RF and SRF methods in (d)Jh, (e)Ls, and (f)Mh.

The radius of neighboring stations: In the process of spatial quality control, the selection of the radius of neighboring stations also affects quality control; thus, different radii of neighboring stations were selected for testing. Fig. 3 depicts the performance difference of the RF, SRF, and SRT methods, where the MAE and RMSE of the RF and SRF method were lower than those of the SRT method, and the runtime of the SRF method was less than that of the RF method with the increase of radius. The performance of the SRT method fluctuates wildly when the radius to neighboring stations is less than 80 km, and it indicates that the SRT method relies on the radius of neighboring stations. In contrast, the SRF method is not affected by the radius of neighboring stations. In addition, the performance and runtime of the SRF method are relatively stable regardless of the change in radius.

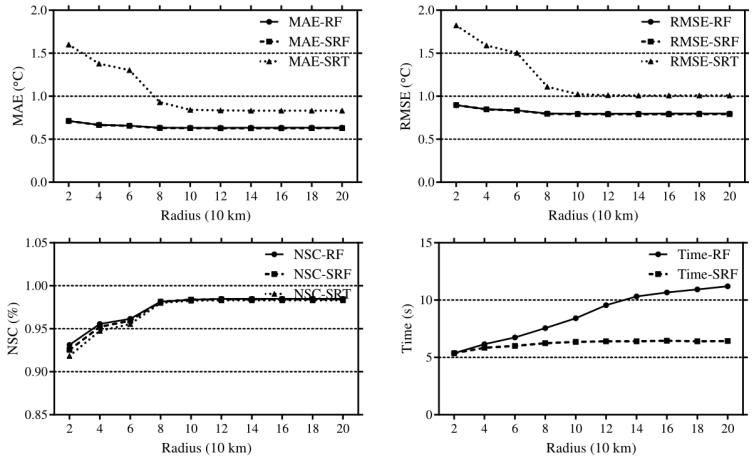


Figure 3. Performance of mean of the RF, SRF, and SRT methods under different radiuses.

Spatial Sensibility Analysis. The performances of the RF, SRF, and SRT methods in different target stations are shown in Fig. 4, which illustrates that the SRF method is superior to the RF and SRT methods. On the one hand, the runtime of the SRF method is shorter than that of the RF method, particularly for regions with a large number of neighboring stations. On the other hand, the MAE and RMSE of the SRF method are smaller than those of the SRT method, especially for regions with few neighboring stations, such as Jh, Ls, and Lz, which have a low spatial correlation. By comparing the performance of the three methods, the SRF method has an improved runtime over the RF method and improved accuracy in comparison to the SRT method.

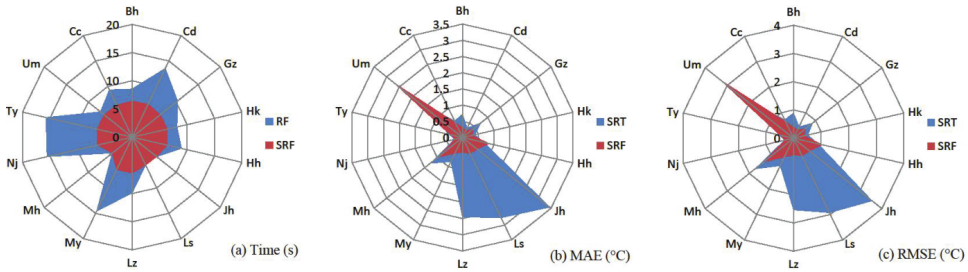


Figure 4. Performance of the RF, SRT, SRF methods for the 14 target stations.

The MAE and RMSE of the SRT method were lower than those of the SRF method in the Hh and Um target stations. To confirm whether this is a particular case or not, it was necessary to analyze the performance of the SRT and SRF methods at the Hh and Um target stations and their neighboring stations. The performances of the SRT and SRF methods are shown in Fig. 5, where the diamond indicates the station with a performance of the SRT method that is better than that of the SRF method and the dot indicates the opposite situation, and different colors indicate altitude. It is clear that the performance of the SRF method is better than that of the SRT method for most stations, but there are nine stations in these two regions which the performance of the SRT method is better than that of the SRF method. In future work, the selection of the quality control methods for these nine stations is worth considering.

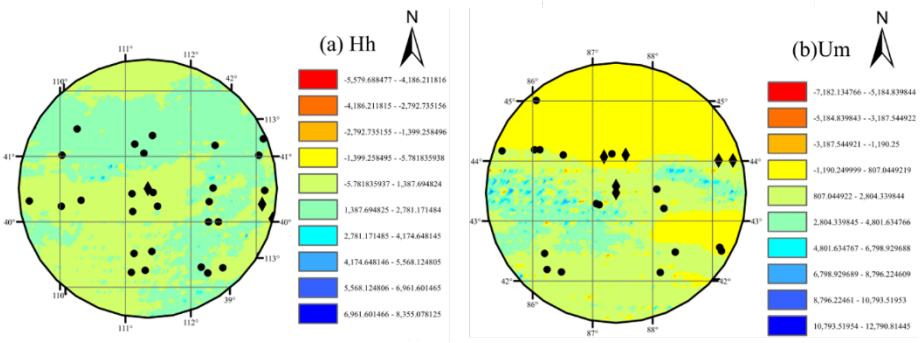


Figure 5. Performance of the SRT and SRF methods for (a) Hh, (b) Um target stations and their neighboring stations.

5. Conclusion and Discussion

A new quality control method is proposed to identify the outliers in daily surface temperature observations. The key to the SRF method is the data pre-processing of the SRT method and the regression of the RF method. The SRT method can effectively select the neighboring stations of high correlation, namely feature extraction; the random forest has the advantages of high precision and low generalization error. For small sample data, the random forest has better performance. A validation study by daily temperature observations in 14 target stations with seeded errors illustrates the excellent performance of the SRF method in spatial quality control, especially in regions with few neighboring stations.

The traditional spatial quality control methods have much to do with the geographical environment of stations; the dataset influences the runtime of the RF method and it increases with the increasing of neighboring stations. By considering the spatial correlation coefficients and performance of different ways, the spatial correlation may significantly affect the performance of the quality control model. For example, the performance of other methods in Bh, Cd, Gz, My, and Ty is much better than that of different approaches in Jh, Ls, Lz, Mh, and Um. However, in some unique stations such as Nj and Hk, it is difficult to determine its spatial correlation by using spatial correlation coefficients. Still, the performance of the SRF and SRT methods is good. The version of Mh, which has the lowest spatial correlation, has a lower MAE and RMSE than Jh and Ls. Therefore, the results show that spatial correlation may impact the performance of the quality control model, but there is no linear relationship between them.

It is important to note that the SRF method performs much better than the SRT method, indicating that the density of neighboring stations has a considerable impact on the performance of the SRT method. However, the thickness of neighboring stations has little effect on the performance of the SRF method. In addition, the SRF method has a lower MAE and RMSE than the SRT method in regions with a large number of neighboring stations. In general, the SRF method is more stable and accurate than the SRT method as the number of neighboring stations changes. The SRF method is more time-efficient than the RF method in regions with a large number of neighboring stations.

Compared with the SRT method, the SRF method can effectively eliminate the influence of the geographical environment and improve accuracy. Compared with the RF method, the SRF method can exploit the advantages of the SRT method to extract the most important information to construct a dataset with higher correlation, reducing the runtime of the quality control model while maintaining accuracy. The comprehensive comparison of the performance of the SRT, RF, and SRF methods demonstrates that the SRF method is superior to the SRT and RF methods with the same number of selected neighboring stations and selected radius. However, the SRF method has some limitations. (1) The SRF method does not effectively reduce the runtime and improve the model accuracy

compared with the RF method in the regions of a few neighboring stations such as Jh, Ls, and Mh. (2) The SRF method works in the target stations with fewer than 15 (or 10) neighboring stations. For example, there is no difference between the performance of the SRF method and the RF method in Mh because the station's filtering function of the SRT method does not make sense when the neighboring stations are fewer than 15 (or 10). (3) The SRF method evaluates the validity of the observations by thresholding, and simply divides the observations into correct or erroneous ones. The method of representing observations as credible probabilities is more reliable such as SRT-PS. However, because the two methods have different emphases and evaluation indexes, they cannot be compared between the SRF and SRT-PS methods.

It is recommended that the SRF method could be used for stations with a large number of neighboring stations. In future work, other regression models and interpolation methods will be selected to improve the model's accuracy, especially under extreme weather conditions such as typhoons. Furthermore, time-series effects have not been added to the model discussion. In future research, a model that combines temporal and spatial can be considered.

References

[1] Schneider, U. Gpcc's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor. Appl. Climatol.115, 15–40 (2014).

[2] Cheng, A. R., Lee, T. H., Ku, H. I. & Chen, Y. W. Quality control program for real-time hourly temperature observation in Taiwan. J. Atmospheric Ocean. Technol.33, 953–976 (2016).

[3] Hubbard, K. G. & You, J. Sensitivity analysis of quality assurance using the spatial regression approach—a case study of the maximum/minimum air temperature. J. Atmospheric Ocean. Technol.22, 1520–1530 (2005).

[4] Breiman, L. Random forests. Mach. Learn.45, 5–32 (2001).

Xuanchen Guo received a B.S. degree in electronic information science and technology from Nanjing Agricultural University, Nanjing, China, in 2019. She is currently working toward an M.S. degree in the college of engineering, Nanjing Agricultural University. Her current research interests include energy-efficient routing algorithms in wireless sensor networks and the agricultural Internet of Things.

Xing Yang received an M.S. degree in control engineering from Nanjing University of Information Science and Technology, Nanjing, China, in 2018. He is currently working toward a Ph.D. degree in the college of engineering, Nanjing Agricultural University. His current research interests include fault diagnosis in wireless sensor networks, agricultural Internet of Things reliability, and machine learning algorithm.

Lei Shu received the B.S. degree in computer science from South Central University for Nationalities, Wuhan, China, in 2002, the M.S. degree in computer engineering from Kyung Hee University, Seoul, South Korea, in 2005, and the Ph.D. degree in computer engineering from the Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland, in 2010. He is currently a Distinguished Professor with Nanjing Agricultural University, Nanjing, China, and a Lincoln Professor with the University of Lincoln, Lincoln, U.K. He is also the Director of the NAU-Lincoln Joint Research Center of Intelligent Engineering, Nanjing, China.

Ye Liu received M.S. and Ph.D. degrees in electronic science and engineering from Southeast University, Nanjing, Nanjing, China, in 2013 and 2018, respectively. He is currently a Researcher with Nanjing Agricultural University, Nanjing, China. He has authored papers in several prestigious journals, such as the IEEE Communications Magazine, IEEE Internet of Things Journal, ACM Transactions on Embedded Computing System, and IEEE Network. His current research interests include wireless sensor networks, energy harvesting systems, and mobile crowdsensing.

End-to-End Chest X-ray Abnormality Detection with Transformers

Qiran Kong[1], Yirui Wu[1]
[1]Hohai University, Nanjing, China
kds809545917@gmail.com, wuyirui@hhu.edu.cn,

1. Introduction

Since the first attempt to establish a computer-aided detection system in the late 1960s, several research groups have focused on developing CAD tools for information extraction from X-rays. Historically, most tools that are used to inference on chest X-rays are rule based and many commercial products have been developed for clinical usage. With the development of deep learning, researchers have made great progress in this field. However, superimposition and overlapping of different anatomical structures locate along the projection direction, leading to the diversity of chest abnormalities, which makes abnormality on chest X-rays difficult. Therefore, existing CADs haven't achieved a highly enough accuracy to work as a reliable and trustful diagnosis tool, leaving the automatic detection of chest abnormalities remains unresolved.

Facing these challenges, we propose our end-to-end transformer-based method for chest abnormality detection. Our proposed method consists of two different modules, a CNN-based feature extractor and a transformer-based detector. Our CNN-based extractor contains a novel dilated encoder, which not only enlarges the receptive fields, but also helps in fusing multiscale context information, improving detection performance further especially on small abnormalities, our transformer- based detector could map the feature maps directly into a set of predictions without handicraft components and multiple steps. We reframe chest the abnormality detection into direct set prediction. By adopting a bipartite matching of network predictions and ground truth label using Hungarian Algorithm to generate the specially designed loss function, our proposed method can be trained in an end-to-end manner. We hope this could make our network automatically find an effective pattern for chest abnormality detection and alleviate the problem of domain shift.

2. Method

Inspired by Detr[3] we design our proposed chest abnormality detection method. The overall structure is shown in Fig.1. Our proposed method uses CNN backbone and our specially designed dilated encoder to learn a 2D representation of an image. Then we flatten this 2D representation into a 1D sequence supplemented with positional encoding. Afterwards, we output a further 1D sequence representation of the image is output by the encoder and decoder layers. Finally, we feed this sequence into a fully connected network to get a set of predictions with bounding boxes and the corresponding classification label.

Feature Extractor. Feature Extractor is to obtain a 2D representation of an Input image which consists of two major parts, CNN backbone and dilated encoder that are shown in Fig.1. We adopt ResNet-50[1] as our CNN backbone and the dilated encoder will be introduced next. It takes a chest X-ray image as input $x \in \mathbb{R}^{3\times H_0\times W_0}$ and generates a lower revolution image but with much more channels output $h \in \mathbb{R}^C \times H \times W$, the input and output dimension of dilated remain consistent. Typical values we use are $\mathrm{C} = 2048$, and $H, W = \frac{H_0}{32}, \frac{W_0}{32}$. The input images were batched together if the batch size is not 1 and 0-padding is applied to ensure that all images in a mini-batch have the same dimensions $(\mathrm{H}_0, \mathrm{W}_0)$as the largest image dimension among each dimension. Then we use channel reduction which is typically a 1×1 convolution that reduces the channel to C/r. We add the channel reduction layer to reduce the tremendous memory usage of large numbers of channels.

Transformer Encoder and Decoder. We use the standard basic transformer which first occurs in [2]. The transformer expects a sequence as input, so we first flatten the channel reduced 2D image into an 1D sequence. That means we reshape the input from $\mathbb{R}^C/r \times H \times W$ to $\mathbb{R}^{\frac{C}{r}HW}$. Transformer can do computation in parallel and has the property of permutation-invariance, thus we use positional encoding to supplement the sequence with positional information. Each encoder layer has a standard architecture and consists of a multi-head self-attention module and a feed forward network (FFN). The important self-attention will be introduced later. We stack N transformer encoder layers on top of each other. Transformer decoder layer is similar, and we also stack Ndecoder layers together. It follows the standard architecture of self-attention transformer but can also decodes object decoders in parallel which significantly increases the speed and saves a large amount of computational resources. The transformer decoder layers take input as the N_{obj} object queries supplemented with positional encoding and

the output embedding of the last transformer embedding layer. Unlike many other transformer decoders, our transformer decoder also has the permutation invariance property, therefore positional encoding is also used in the decoder layer. The N_{obj} object queries are transformed into an output embedding by the decoder.

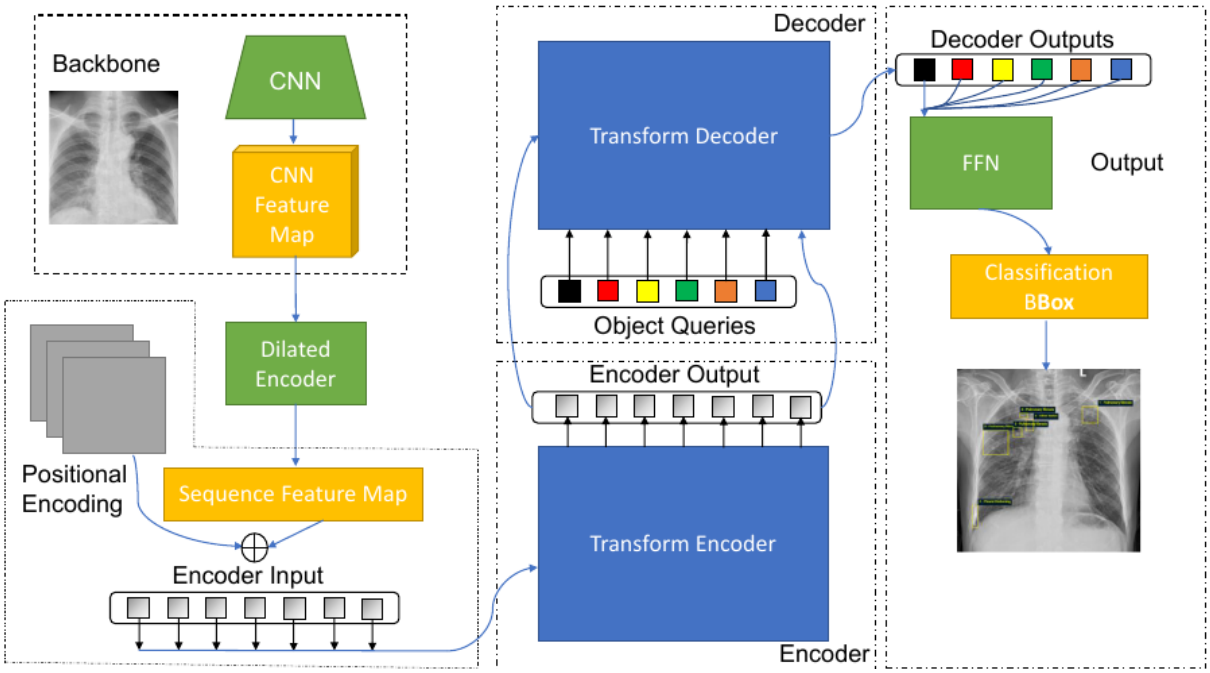


Figure 1. Network structure pipeline of our proposed method.

FFN Object Predictor. The goal of FFN object predictor is to output a set of bounding box predictions of the chest abnormalities of interest. The output embedding for the N_{obj}queries are then independently decoded into box coordinates (x, y, W, h) and the corresponding class labels by FFN and linear projector. We have a regression layer for box localization and classification layer for label generation (including background). We output N_{obj} bounding box predictions in parallel. Our FFN object predictor is composed of 3 perceptron layers with ReLu activation and a linear projection layer.

Context Aware Dilated Encoder. Our context aware dilated encoder is inspired by YoloF[7]. We show structure design of the dilated decoder subnetwork in Fig.2, which mainly consists of convolution, dilated convolution, and skip connections. It is common for Modern detectors use multiple levels of feature maps to achieve better performance. Different level of feature maps contains the contextual information in different granularity. Utilizing all the features maps could assure that out network could get all the contextual information need for a complicated task. These methods are mainly composed of convolutional layers while the key components of our network are transformer encoders and decoders. Due the limitation of computational resources, especially for graphic memory usage, it is impossible to process different features maps altogether, this made us to consider feature fusion instead of divide and conquer scheme like FPN[6].

We design our dilated decoder sub-network to extract context information. We first design a $1 \times 1$and a 3×3 standard convolution layer as a projector. This is used for feature refinement as for FPN. The main component is the residual block which consists of two 1×1 convolution layer with a 3×3 dilated convolution layer. Then we stack several residual blocks with residual connection to build a highway for gradient flow. Each residual block has a different dilated rate with different receptive field, covering all scales and extracting all contextual information. Dilated layers could enlarge receptive field and we stack up different dilated layers without weight sharing. Our goal is to extract information and add more scales out of the backbone.

3. Experiments

Datasets and measurements. Due to the access to medical data is difficult, we use only one open source dataset, Vinbig dataset for chest X-ray abnormality detection. We use a subset of the open source version with a total of

5000 training images and 1063 testing images with a lot of images with no abnormality filtered out. Furthermore, the images have been down-sized by a factor of 4. All images were labeled by a panel of experienced radiologists for the presence of 14 critical radiographic findings with annotations of bounding boxes and the corresponding class labels. The evaluation metric we use is the standard pascal voc mAP[8]. Any prediction that has an IoU that is larger than 0.5 will be considered true. The AP50 is calculated as follows:

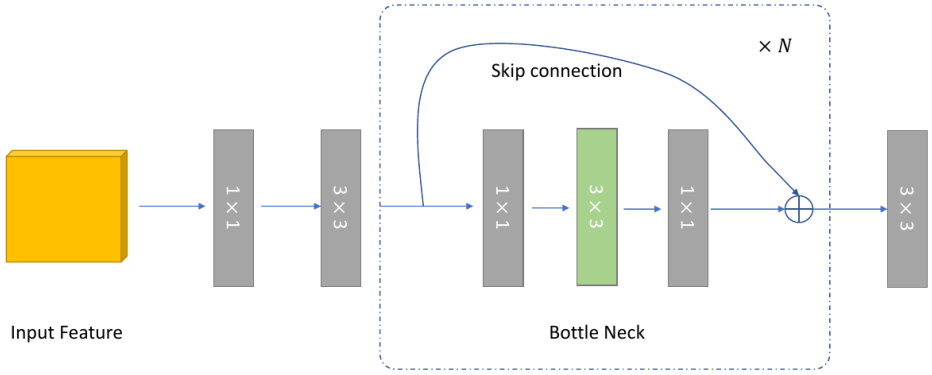


Figure 2. Structure Design of Context aware Dilated Encoder, where gray block represents standard convolution and green block represents dilated convolution.

$$AP_{50} = \frac{1}{num \times C} \sum_{i=1}^{C} \sum_{j=1}^{num} P(i, U_j)$$

where i refers to the index of the class and $P(i, U_j)$refers to the precision for the ith object class under the fixed IoU threshold of U_j. Furthermore, we define AR as average recall over detections on Q images

$$AR_q = \frac{1}{Q} \sum_{q=1}^{Q} R(q)$$

where Q is the number of total detected images. AR offers another view on robustness of detection results.

Comparison with existing methods. We report detection results achieved by the proposed method and existing methods in Table.1. Comparison methods include Faster R-CNN[4], Faster R-CNN FPN, Yolo v3[5]. Our method is a huge improvement over existing methods in this very challenging task. Especially when compared with Yolo V3, AP50 is increased from 26 to 34.3. Not surprisingly, the performance is increased a little bit from Faster R-CNN to Faster R-CNN FPN. This is because Feature pyramid could utilize multiple level of features, our dilated encoder acts similarly. We also believe that some of the improvement come from the end to end nature, as our method doesn't post process like nms. Some of the detection results might be highly overlapped and NMS may not only reduce the redundancy, but also remove some of the good results.

Some of the detection results are shown in Fig.3. We place the ground truth on top of our detection results and each color corresponds to a unique class. From the visualization of detection results, we can easily see that, our method performs well on some of simple classes like Cardiomegaly and Aortic enlargement. When facing much more complicated scenarios, the results become a little worse. Our method might misclassify the type of the abnormality, but our method could still detect the existence of the abnormality. This might be the benefit of the self-attention nature of our transformer-based method.

Experiments on number of transformer layers. As shown in Table.2, when we set the number of heads to 8 which is the default setting for most transformer models, we gained the maximum performance, both increasing and decreasing the number of heads will drop the AP and AR. When there are only 4 heads, the complexity of the network is not enough to deal with this task. In practice, we have also noted that the number of heads doesn't influence the memory usage as well, this confirms that with reduction of dimension for each head, we can save the computational cost.

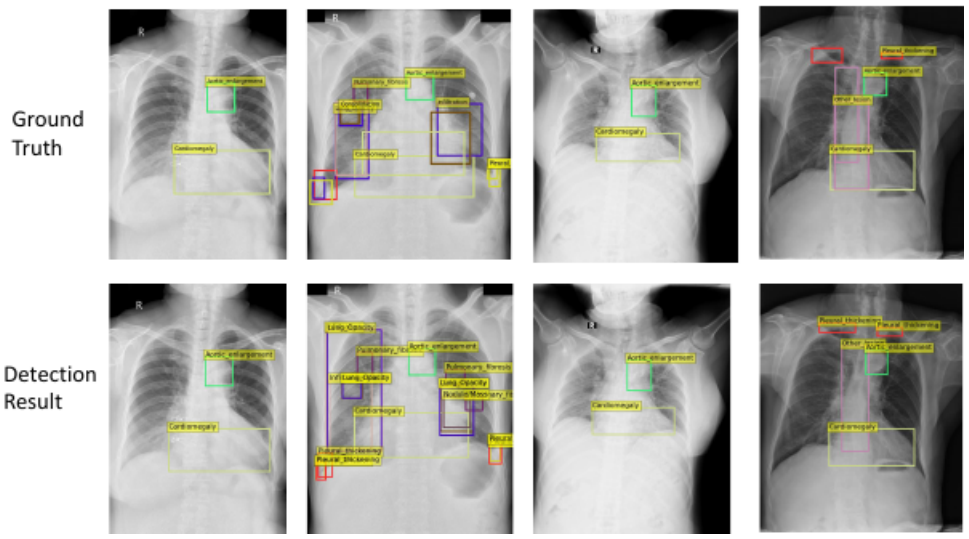


Figure 3. Comparison of detection results between our proposed method and ground truth

Table 1. Comparison with other methods

Method	AP50	AP75	AR	FPS
Faster R-CNN	28.5	13.4	28.2	12
Faster R-CNN FPN	29.1	13.6	29.1	8.6
Yolo V3	26.2	11.5	24	44
Detr	33.7	15.1	32.9	18.1
Proposed	34.3	15.8	33.4	16.5

Table 2. Experiments on number of heads

Num of heads	AP	AR
4	30.4	32.9
8	34.3	33.4
16	34.3	33.4

4. Conclusion

This paper presents an end-to-end method for chest X-ray abnormality detection method using self-attention transformers. Our method is designed be end-to-end and doesn't rely on many hand-crafted components we do not have many pre and post processing steps. We have also added a dilated decoder layer to fuse context information from multiple scales. Our network is extremely for this challenging task and we show its superiority to some existing methods. Our further work is to not only include information from the image, but also utilize multi-modal information to achieve better results.

References

[1] He, Kaiming, Xiangyu Zhang, and Shaoqing Ren. "Deep Residual Learning." Image Recognition (2015).

[2] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In Advances in neural information processing systems, pp. 5998-6008. 2017.

[3] Carion, Nicolas, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. "End-to-end object detection with transformers." In European Conference on Computer Vision, pp. 213-229. Springer, Cham, 2020.

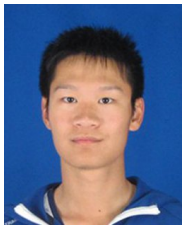
[4] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015): 91-99.

[5] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

[6] Kim, Seung-Wook, Hyong-Keun Kook, Jee-Young Sun, Mun-Cheon Kang, and Sung-Jea Ko. "Parallel feature pyramid network for object detection." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 234-250. 2018.

[7] Chen, Qiang, Yingming Wang, Tong Yang, Xiangyu Zhang, Jian Cheng, and Jian Sun. "You only look one-level feature." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13039-13048. 2021.

[8] Hoiem, Derek, Santosh K. Divvala, and James H. Hays. "Pascal VOC 2008 challenge." World Literature Today (2009).



Qiran Kong received a B.S. degree in computer science and technology from Hohai University, Nanjing, China, in 2019. He is currently working toward an M.S. degree in the college of computer and information, Hohai University. His current research interests include computer vision, scene text detection, biometrical image analysis, object detection and instance segmentation with transformers.

Yirui Wu currently is an Associate Professor at Hohai University and working in the Hydrology Big Data Group. Before coming to Hohai, he obtained my Ph.D. degree from Nanjing University in 2016. He received my B.S. Degree from Nanjing University in 2011 as well. During my Ph. D study, he was with the IMAGE Lab under the supervision of Prof. Tong Lu and worked closely with Prof. Shivakumara Palaiahnakote. He visited Hong Kong University of Science and Technology twice in 2012 and 2014, supervised by Prof. Chiew-Lan Tai and Dr. Oscar Kin-Chung Au. He has published more than 40 journal/conference papers on Computer Vision, Pattern Recognition, Multimedia, and Intelligent Water Conservancy, including IEEE Transactions on Image Processing, IEEE Transactions on Multimedia, Computer-Aided Geometry Design, Signal Processing: Image Communication, Expert Systems with Applications, ACM MM, ICME, BIBM, ICPR, ICDAR, ICIP, MMM, PCM and others. He has served as PC members for HPCC2021(CCF-C), CIKM2021(CCF-B), PIC2021, ACPR2021, DSAA2020， ML4CS2020， PIC2020, SECSOC2019 and others.

MMTC OFFICERS (Term 2020 — 2022)

CHAIR

Jun Wu
Fudan University
China

STEERING COMMITTEE CHAIR

Joel J. P. C. Rodrigues
Federal University of Piauí (UFPI)
Brazil

VICE CHAIRS

Shaoen Wu (North America)
Illinois State University
USA

Liang Zhou (Asia)
Nanjing University of Post and Telecommunications
China

Abderrahim Benslimane (Europe)
University of Avignon
France

Qing Yang (Letters & Member Communications)
University of North Texas
USA

SECRETARY

Han Hu
Beijing Institute of Technology
China

STANDARDS LIAISON

Weiyi Zhang
AT&T Research
USA

MMTC Communication-Frontier BOARD MEMBERS (Term 2020—2022)

Danda Rawat	Director	Howard University	USA
Sudip Misra	Co-Director	IIT Kharagpur	India
Guanyu Gao	Co-Director	Nanjing University of Science and Technology	China
Rui Wang	Co-Director	Tongji University	China
Lei Chen	Editor	Georgia Southern University	USA
Tasos Dagiuklas	Editor	London South Bank University	UK
ShuaiShuai Guo	Editor	King Abdullah University of Science and Technology	Saudi Arabia
Kejie Lu	Editor	University of Puerto Rico at Mayagüez	Puerto Rico
Nathalie Mitton	Editor	Inria Lille-Nord Europe	France
Zheng Chang	Editor	University of Jyväskylä	Finland
Dapeng Wu	Editor	Chongqing University of Posts & Telecommunications	China
Luca Foschini	Editor	University of Bologna	Italy
Mohamed Faten Zhani	Editor	University of Quebec	Canada
Armir Bujari	Editor	University of Padua	Italy
Kuan Zhang	Editor	University of Nebraska-Lincoln	USA
Bin Tan	Editor	Jinggangshan University	China