AntennaNet: Antenna Parameters Measuring Network for Mobile Communication Base Station Using UAV

Yikui Zhai, *Member, IEEE*, Qirui Ke, Xinglin Liu, Wenlve Zhou, Ying Xu, Zilu Ying, Junying Gan, Junying Zeng, Ruggero Donida Labati, *Member, IEEE*, Vincenzo Piuri, *Fellow, IEEE*, and Fabio Scotti, *Member, IEEE*

***Abstract*—In the field of measuring parameters of mobile communication base station antenna, most of its methods share some deficiencies to a different extent. The traditional ones suffer from inferior efficiency and low safety coefficient while the state-of-the-art ones reveal problems of unsatisfying operational speed and accuracy. To address these problems, a high-speed and high-precision measuring system has been proposed with AntennaNet to speed up instance segmentation. Firstly, the dual attention mechanism module is designed and adopted in AntennaNet, which can promote the accuracy and efficiency of the detection and segmentation of mobile base station antenna, while the PointRend module with down-tilt fitting is also employed to smooth the edges of masks segmented from antenna images and measure the parameters data. Secondly, transfer learning is introduced to ameliorate overfitting and training difficulty caused by small samples. Thirdly, threshold, pixel coordinates and least square are applied to ascertain the number of antennas and measure the relevant parameters. Finally, this high speed and high precision measuring system is complete with the combination of android App and UAV. Experimental results fully attest to the preponderance of the proposed method in its significant improvement of segmentation speed, fitting accuracy speed and fitting accuracy of the antenna, whose relevant parameters of the fitting also conform to industry standards. Our method not only ensures staff safety but also improves social efficiency, which revolutionized the way of measuring antenna parameters of mobile communication base stations.**

Index Terms*—AntennaNet; Dual attention mechanism; PointRend with down-tilt fitting; Transfer learning; Parameters measurement.

This work is supported by National Natural Science Foundation (No.61771347, 61672335); The Open Project of Guizhou Province Public Big Data Key Laboratory (No.2019BDKFJJ015); Basic Research and Applied Basic Research Key Project in General Colleges and Universities of Guangdong Province (No. 2018KZDXM073, 2019KZDZX1017, 2020ZDZX3 031); Guangdong Basic and Applied Basic Research Foundation (No. 2019A15 15010716).

Yikui Zhai, Qirui Ke, Xinglin Liu, Wenlve Zhou, Ying Xu, Zilu Ying, Junying Gan and Junying Zeng are with the Department of Intelligent Manufacturing, WuyiUniversity, Jiangmen, China, 529020.

Ruggero Donida Labati, Vincenzo Piuri, and Fabio Scotti are with the Departimento di Information, Universita, Degli Studi di Milano, via Celoria 18, 20133 Milano(MI), Italy.

Corresponding Author: Xinglin Liu. Email: jmxlliu@163.com

I. Introduction

BASE station antenna, as the final transmission part of mobile communication wireless network and the last link of base station, weighs heavily in the overall quality, of mobile communication network. The down-tilt angle of mobile communication base station antenna is adjusted as per the network coverage requirements, distribution of traffic, anti-interference requirements, and network service quality, which can optimize antenna down-tilt [1] to bring about the maximum coverage probability of each base station antenna density. For example, in practice, if a large-scale event needs to be held, the mobile operators need to test in advance whether the network coverage dependent on the antennas down-tilt can meet the requirement of the event. If the conventional method is adopted, it involves climbing of the towers and manual measurement of all the base stations one by one so as to test whether the network coverage in this area can meet the needs of holding the large-scale event. When the results show negative, the antennas down-tilt needs to be adjusted manually to meet the demand of ensuring user experience.

With the continuous advancement of the construction of 4G networks and the popularization of 5G networks, the requirements for capacity and coverage of mobile communication networks and the number of mobile communication base station are synchronously increasing. To accommodate these changes, the structure of the base station is converted from a single-layer antenna structure [2-3] to a multi-layer antenna structure [4]. Therefore, it greatly raises the difficulty of measuring the antenna down-tilt manually with workers ascending the tower. Concerning this issue, many researchers have contributed their efforts to propose various scientific methods. Xu et al [5] proposed a non-contact measurement system based on kinematic structure, in which the accurate posture can be estimated with only a few images of the antenna taken and some easy interaction performed on the smartphone. However, this method falls short of accuracy with an error of 2 degrees, far from satisfying the industry standards. Moreover, Pedras et al [6] introduced a novel quality of experience model. By evaluating several radio frequency channel indicators, the average opinion score level is used to estimate the quality perceived by the user to obtain the optimal antenna down-tilt

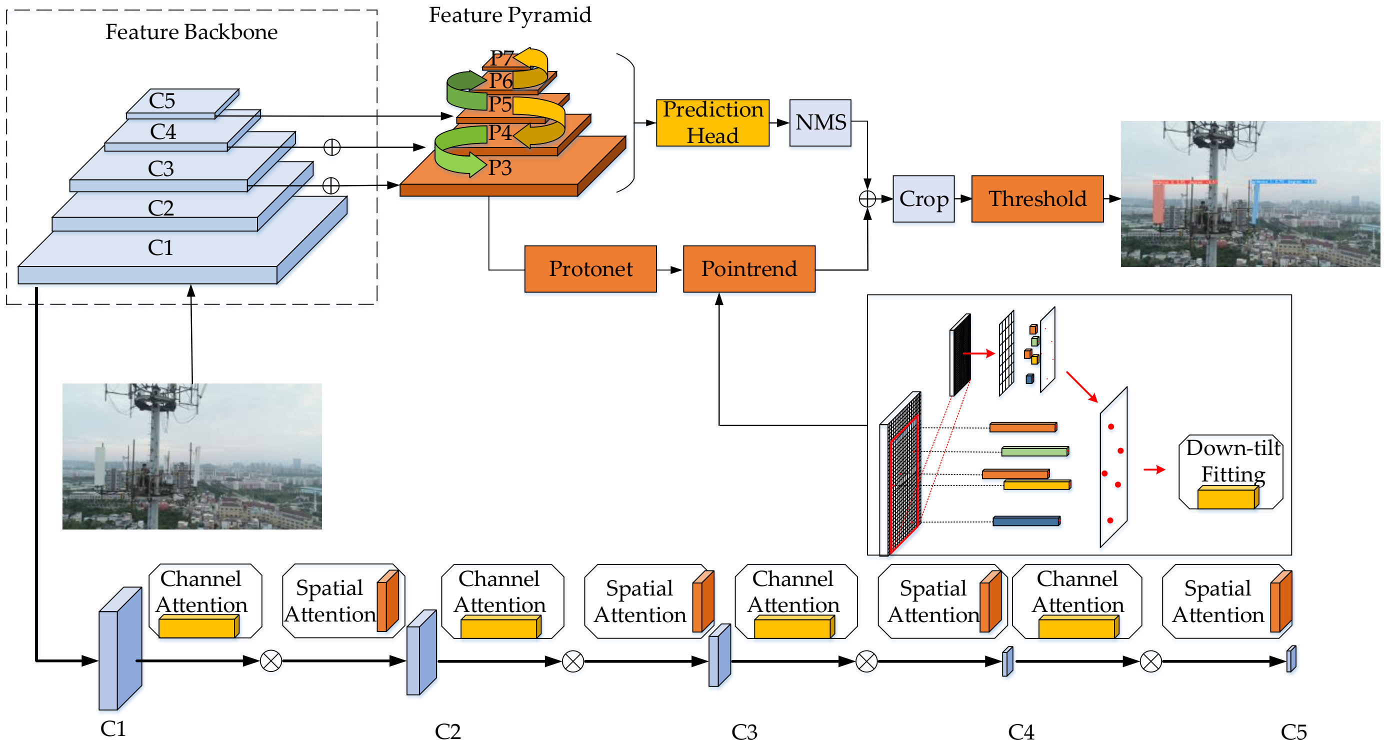


Fig. 1. The framework of AntennaNet. Among them, the RetinaNet based on Feature Pyramid Network (FPN) and resnet101 are used to build a basic network framework. And the dual attention module: channel and spatial attention at every convolutional block of deep networks is to refine the intermediate feature map, and PointRend module is added after Protonet to generate more refined masks.

followed by manual adjustment of the antenna down-tilt angle. Takashi et al [7] proposed a two-way direction detection unit to test the antenna in a state where the electromagnetic wave radiation plane points to a reference direction directly opposite to the measurement plane; then scanned and detected the two-way direction so as to test the relevant parameters of antenna. Nonetheless, in order to ensure the measurement accuracy, the measurement range must be controlled at the minimum, which greatly increases the difficulty of measurement and demands extremely high requirements on antennas. Undoubtedly the contributions of the methods mentioned above cannot be easily overlooked, but their problems such as large measurement errors, expensive labor costs, and harsh measurement environments similarly should also be clearly acknowledged. And our previous work [8] can solved these challenges just right by introducing a fully automatic and intelligent antenna measurement unmanned aerial vehicle (UAV) system for mobile communication base station. This system commences with the utilization of UAV to obtain antenna data to construct a self-made dataset and then completes antenna parameters measurement by combining instance segmentation and mathematical analysis, which can effectively strengthen the safety factor during measurement, efficiency, and productivity. Experiments certify its conformity to industry standards and its practicality in real world. In the past year, many successful cases of our system can be witnessed in lots of cities. With our system, the possibility of repetitive climbing tower by workers to measure antenna parameters has been completely ruled out. Instead, all they need is to utilize our system to measure the parameters of each antenna, mark the ones in need of adjustment, and then climb the tower to adjust. With the above comparison, our system greatly improves work efficiency and reduce risk factor. And we continued to work on the field of antenna parameters measurement for mobile communication base stations, and realized a revolutionary improvement based on the original method. AntennaNet, consisting of a one-stage instance segmentation network based on transfer learning [9] and dual attention mechanism is incorporated for much faster instance segmentation to automatically generate instance mask for each input according to their own characteristics and a PointRend module is added to for effectively improving the measurement accuracy.

The structure of AntennaNet is shown in Figure1. Among them, the RetinaNet [10] based on Feature Pyramid Network (FPN) [11] and resnet101 are used to build a basic network framework. And the dual attention mechanism: channel and spatial at every convolutional block of deep networks is to refine the intermediate feature map, and PointRend is utilized to generate a more meticulous mask for each instance.

Several empirical studies are implemented to certify the effectiveness of the proposed method. Meanwhile, we conducted numerous comparative experiments between the proposed method and the current antenna parameters measurement methods. The results prove that our method not only gets rid of the shackles of hardware, but also becomes the state-of-the-art method. Thus, the contributions of this work can be summarized as below.

1) A novel instance segmentation network named Antenna-Net is presented to measure the parameters of mobile communi-

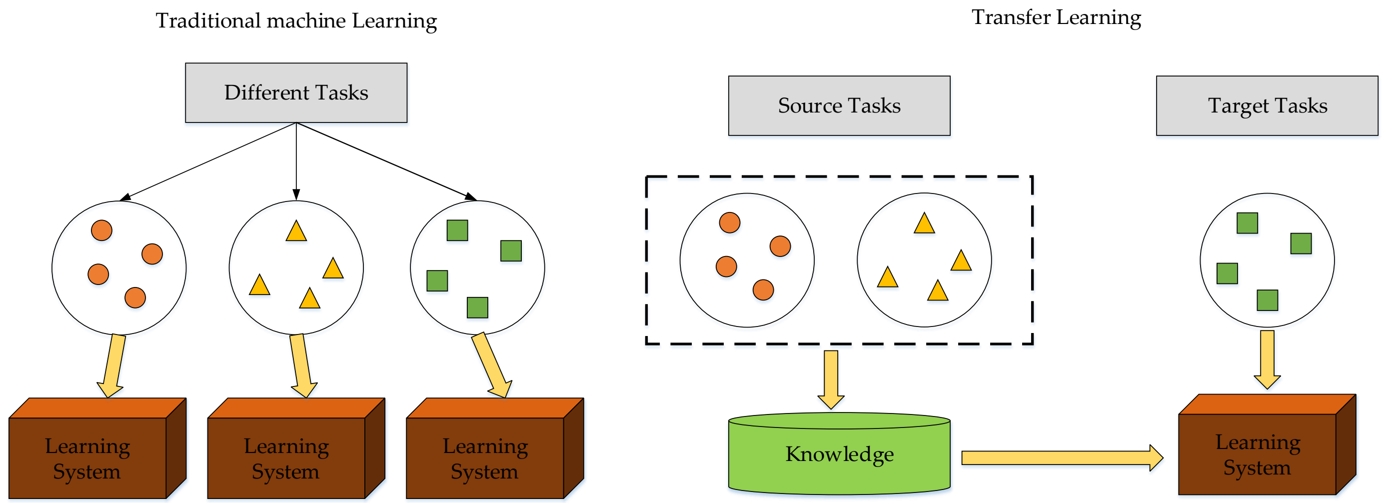


Fig. 2. Differences between traditional machine learning and transfer learning.

cation base station antenna, which combines the dual attention mechanism that poses channel attention in front of spatial attention based on backbone, the prediction module that generates category of region proposals and the segmentation module that produces high-quality mask of each antenna.

2) PointRend module with down-tilt fitting is deployed to make further predictions for fuzzy segmentation edge points of the mask for polishing a more meticulous mask of instance. The addition of PointRend with down-tilt fitting can effectively improve the accuracy of measured antenna parameters.

3) Transfer learning strategy confirmed to be effective for small samples tasks has also been incorporated to mobile communication base station antenna parameter measurement with limited labeled data to mitigate the influence of training difficulty and overfitting.

4) A generalized antenna measuring system is introduced to inspect communication base station, which includes rich appearance and posture information of base station with a novel instance segmentation network. It can automatically ascertain the quantity of antennas in base station and the industrial parameters of each antenna.

5) Extensive comparative experiments have been conducted with other advanced method focusing on speed as well as precision. Notably, our proposed network and system can outperform all the other existing methods on these two aspects.

This paragraph introduces the layout of this paper as follows. Section 2 provides a summary of related works tackling the relative problems of antenna parameters measurement while Section 3 introduces AntennaNet which operates with much faster and high-precision base station antenna parameters measurement method. The elaboration of the design and workflow of this system is presented in Section 4 while experimental results are explained in Section 5. At length, an overview of the existing work and our expectations in this field are to be delivered in Section 6.

II. Related Works

With the continuous outstanding performance of deep learning [12] methods in various fields [13-16], and the astonishing progress of the computing power in artificial intelligence (AI), more and more companies have undoubtedly adopted deep learning methods as the research and application tool for key technologies. For instance, Mathews et al [17] initiated a method based on deep learning methods for classifying single-lead electrocardiogram (ECG) signals, achieving an excellent performance. Moreover, Li et al [18] applied deep learning to autonomous driving, where multi-task learning was first adopted to take internal visual images as input and predict trajectory features and then control decisions were made based on the reinforcement learning control module. In addition, Etxeberria et al [19] suggested a method based on monocular vision and deep learning architecture, making use of visual odometer, attitude estimation and other technologies to provide a new technical solution for train parking in railway positioning. Inspired by these methods above and the previous works, we intend to use deep learning methods to measure antenna parameters, so as to increase the suitability of our method for the actual industrial requirement.

The flying capabilities of Unmanned Aerial Vehicle (UAV) can well address the difficulty of data acquisition arisen from the height of antenna, facilitating the acquirement of high-resolution data. In recent years, examples [20-21] of using UAV for deep learning have been common occurrences. Safadinho et al [20], based on Convolutional Neural Networks (CNN), designed a computer vision algorithm to supplement GPS, which could improve the safety of drones used for mission delivery while Wang et al [21] suggested an online distributed algorithm that used drones for tracking and searching, which took the problem of power into account and well settled the challenge of the best path. It can be predicted that UAV will continue to exert its unique advantages such as maneuverability, extendibility and stability in many fields. Varied from the mentioned studies, our method only makes use of the flying capabilities of UAV to obtain high-resolution image data, and simultaneously combines transfer learning, the attention mechanism [22,23] and the PointRend [24,25] module to complete the measurement of antenna parameters.

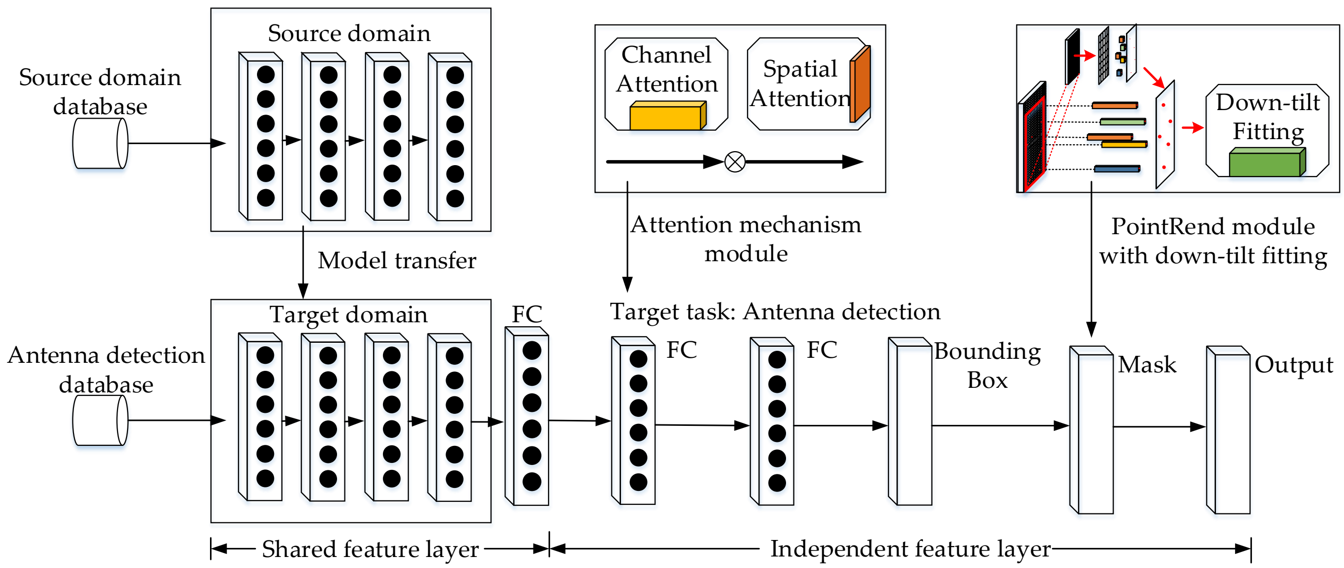


Fig. 3. The process of the proposed AntennaNet with transfer learning. Among them, FC is the abbreviation of fully connected layer.

As the name suggested, transfer learning serves to learn a general law (generalized knowledge) on the source domain and task, and then transfers this law to the target domain and task. Compared with traditional machine learning methods, transfer learning is dedicated to transferring data from the previous task to the target task which lacks high-quality training data, while traditional machine learning techniques devotes itself to capturing information from each task. The differences between them can be observed in figure 2. With the advent of transfer learning [26-27], more and more researchers utilize it to overcome various historical challenges, including detection, classification, and segmentation. For example, Talo et al [28] projected a method with deep transfer learning in the field of classification which could automatically distinguish normal and abnormal brain MR images, and obtained d 5-fold classification accuracy of 100% on 613 MR images.

In the field of computer vision, the attention mechanism operates through selectively allocating visual processing resources to visually significant areas to understand and imitate the behavior of the human visual system. Simultaneously, with the substantial enhancement of computing power and the establishment of saliency detection data sets, deep learning technology has gradually become a means of attention mechanism calculation and modeling. Therefore, the attention mechanism has received copious attention from the academic community as soon as it came out, and has revealed great potential for wide-range application. Li et al [29] introduced a convolutional neural network (CNN) with the attention mechanism, which could perceive the occlusion area of the face and focus on the unique distinguishing non-occlusion area. Experiments confirm the truth that this method augments non-occlusion recognition accuracy of areas and occlusion areas. Liu et al [30] proposed an attention-based bidirectional long short-term memory (LSTM) storage framework with convolutional layer, which can overcome uneven classification due to the high dimension and sparseness of text data. Woo et al [31] raised a convolutional block attention module that could be conducted with the attention map in turn along channel and spatial for feedforward convolutional neural network. It can be concluded from the above methods that no matter in classification tasks, segmentation tasks or detection tasks, the attention mechanism can effectively improve the accuracy. Therefore, based on this conclusion, we adopt transfer learning method and combine it with UAV, the attention mechanism, the PointRend module (detailed in Section3) and numerical analysis to complete the measurement task of mobile communication base station antenna parameters, and realized the original intention of a much faster and more accurate model.

III. Our Methodology

Developing a deep learning approach for base station antenna segmentation requires copious labeled training data, but the labeling and collection process happens to be the bottleneck. To overcome this difficulty, we aim to leverage transfer learning and to obtain parameters from a model of large databases trained by a deep convolutional neural network for instance segmentation of base station antennas. Afterwards, the dual attention mechanism is adopted to improve segmentation accuracy, while the PointRend module demonstrates its effectiveness in smoothing the edge of segmented mask support, followed by recuperating antenna parameters measurement accuracy. Ultimately, we utilize numerical analysis to measure antenna parameters. Figure 3 can be referred to the elaboration of the framework of our proposed method.

A. YOLACT

YOLACT [32] proposed by Bolya et al is one of the constructions for the proposed algorithm and measuring system in this paper. Therefore, we would like to fully introduce and explain the YOLACT method before introducing other parts of the algorithm. Three losses functions of YOLACT include category confidence loss $Lcls$, box regression loss $Lbox$, and mask loss $Lmask$, which are actually Softmax Loss, smooth-L1 Loss and the loss used to calculate the binary cross-entropy loss

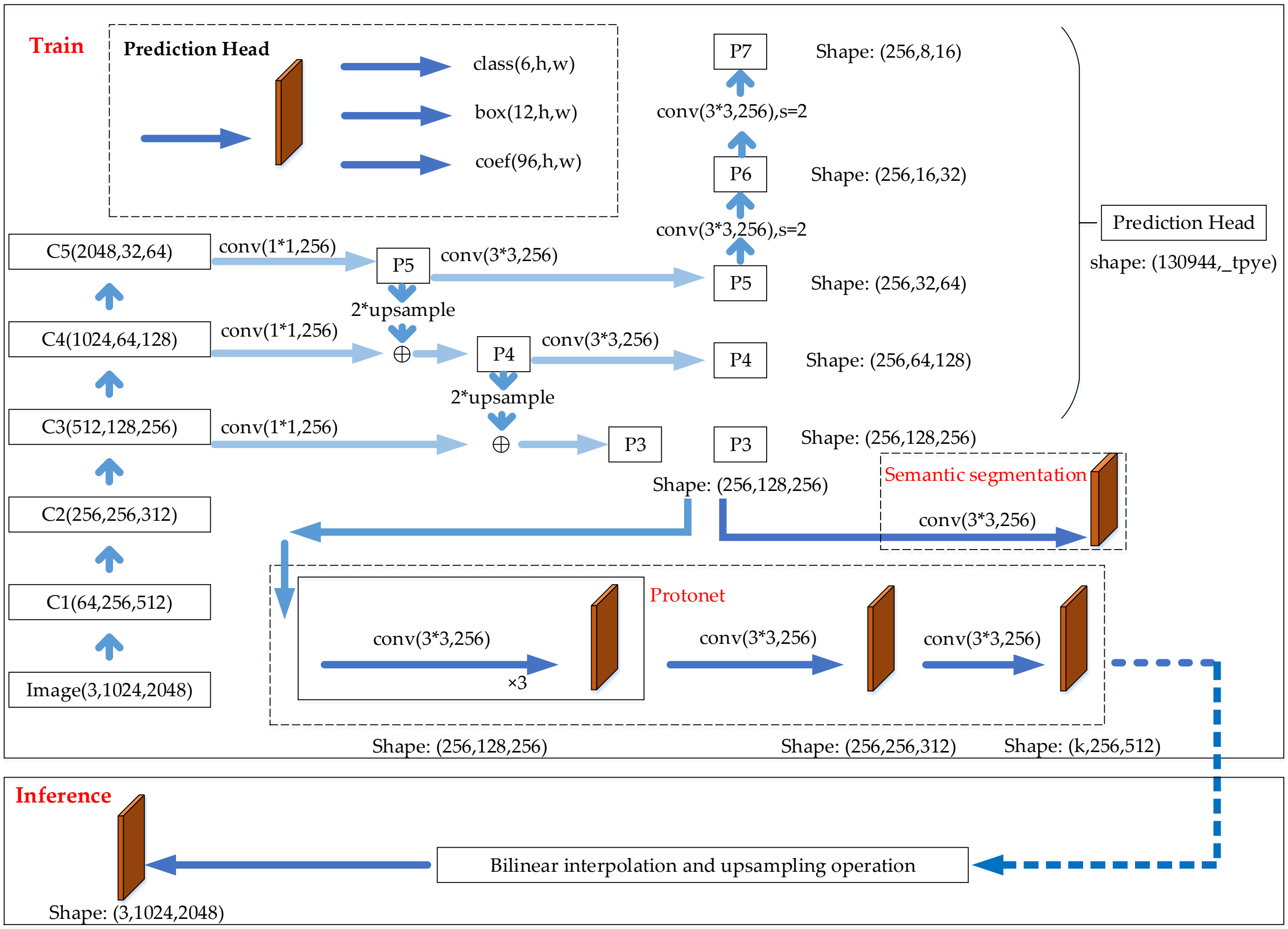


Fig. 4. Flow diagrams during proposed method.

between the integrated masks and ground truth masks respectively. The mask loss $Lmask$ can be drawn as:

$$Lmask = BCE(M, M_{gt}) \quad (1)$$

The weight coefficients of these three losses are 1, 1.5, 6.125. And during the training process, the online hard example mining algorithm ensures the ratio of positive and negative samples as 3:1. It is worth noting that since the size of the mask is 138×138, the mask data of the original image needs to be reduced accordingly by bilinear interpolation to the same size when calculating the mask loss. There is a Head architecture in [32] to process the feature map of the FPN network to generate three types of output: confidence of all categories with a dimension as $W \times H \times 3 \times 81$ because 81 categories in COCO; all position offset in a dimension of $W \times H \times 3 \times 4$; and all mask coefficient with a dimension as $W \times H \times 3 \times 32$, in all of which, 3 represents the number of anchors generated by each pixel.

In addition, a Protonet structure is designed which consists of multiple 3×3 convolutional layers, up-sampling layers and a 1×1 convolutional layer to output masks with a dimension of $138 \times 138 \times 32$. The difference between this structure and Fully Convolutional Networks (FCN) lies in its no specific loss for prototypes. On the contrary, all the supervision information of these prototypes comes from the mask loss after integration. generating semantic vectors which can also be utilized to engender a set of prototype masks of the same size as original

B. Basic Framework

Two-stage models [33-34] which are driving serial logic through re-pooling operations such as ROI (Region of Interest) aligned in [33] to map features to the bounding box. This explains the hardness to speed up the process for two-stage model. Albeit FCIS [35] has made some efforts, such as parallelizing the re-pooling operation, it still lacks a favorable speed due to too many post-processing steps. It was not until Bolya et al [32] provide a simple full convolution model for real-time instance segmentation that both speed and accuracy in instance segmentation could be considered at the same time. It is the first real-time one-stage instance segmentation model and its main idea is to directly add the mask branch to the one-stage object detection algorithm without adding any ROI pooling operation, Inspired by YOLACT [32] and Mask RCNN [33], the proposed AntennaNet combines one-stage object detection structures with their details listed in [36-38] and a mask branch to segment the targets that can increase segmentation speed. The training and inference process of proposed method are shown in Figure 4. Unlike traditional instance segmentation tasks, we construct masks for antennas through two simple parallel tasks seen in Figure 4. Since fully connected layers in FCN specializes in image, this becomes one of the parallel tasks. The other parallel task is to add an additional prediction head to one-stage object detection branch to forecast the vector of mask coefficients. Each anchor represented by instance in prototype space can be encoded with it, fully utilizing the ability of convolutional layers' edge in generating spatial coherent masks. Finally, it is assembled by matrix multiplication, which not only retains the spatial correlation, but also maintains the one-stage model structure, engaged with extremely fast operational speed.

C. Dual Attention Mechanism

This paper adopts a simple but effective dual attention mechanism module, with which, the attention weight can be inferred along channel and spatial in turn by an intermediate feature map, and then be multiplied by the original feature map to adjust feature adaptively. Because of its extremely lightweight assembly process, the dual attention mechanism module can be seamlessly incorporated into any CNN architecture with trifling additional overhead, and can be trained end-to-end with any basic CNN, whose details are displayed in Figure 5. Dual attention module has two sequential sub-modules: channel and spatial, in which both max-pooling and average-pooling outputs are fully employed with a shared network. Spatial attention module merges two similar outputs along the channel axis and maps them to convolutional layer while the channel attention module focuses on meaning feature, each channel of which represents a detector. In order to gather spatial features, global average pooling and max pooling are deployed to obtain different feature information which can be described below:

$$\begin{aligned} Mc(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma((W_1(W_0(F^C_{avg})) + (W_1(W_0(F^C_{\max}))) \end{aligned} \tag{2}$$

The input is feature F of $H\times W\times C$. First, a spatial global average pooling and a maximum pooling are performed respectively to obtain two $1\times 1\times C$ channel descriptions. Then, they are respectively sent to a two-layer neural network, where C/r denotes the number of neurons on the first layer, C represents the number of neurons on the second layer and ReLU refers to the activation function. This two-layer neural network is assigned to common use. Then, the two obtained features are added and passed through a sigmoid activation function to obtain the weight coefficient Mc. Finally, multiplication is executed on the weight coefficient and the original feature F to get the new feature after scaling. As for the spatial attention module, it aims to focus on meaningful features whose detailed information can be referred to the following formula.

$$\begin{aligned} M_s(F) &= \sigma(f^{7\times 7}([AvgPool(F), MaxPool(F)])) \\ &= \sigma(f^{7\times 7}([F^S_{avg}; F^S_{\max}])) \end{aligned} \tag{3}$$

Similar to channel attention, spatial attention module also possesses feature F of $H\times W\times C$ as the input. To begin with, an average pooling and a max pooling of a channel dimension are introduced to obtain two $H\times W\times 1$ channel description, and to splice these two descriptions together according to channels.

Then, after the handle of a 7×7 convolutional layer and sigmoid activation function, the weight coefficient Ms can be obtained. Lastly, the weight coefficient and the feature F are of channel attention and spatial attention can be combined in a multiplied to get the new feature after scaling. The two modules

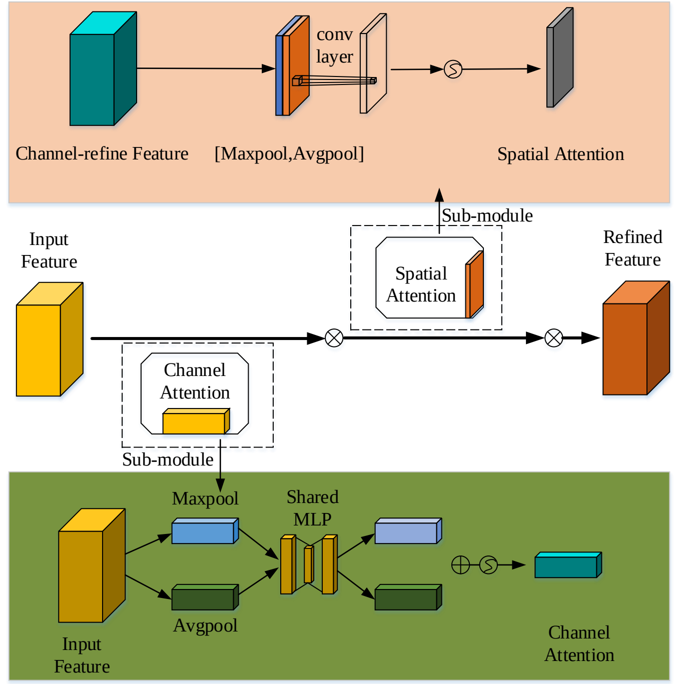


Fig. 5. The overview of dual attention mechanism module and each attention sub-module.

parallel or sequential manner with any basic CNN. Through experiments, it is validated that sequential combination with the channel attention mechanism in the front provides better results. These two independent attention modules not only can save parameters and computing power but also can be integrated into the existing network architecture as a plug-and-play module.

D. PointRend with Down-tilt Fitting

Because of the extreme sampling problems in the pixel labeling task, the mask branch cannot add a very meticulous and accurate mask for each instance, which will greatly undermine the accuracy of antenna parameters measurement in our work. Consequently, this paper resorts to an efficient and high-quality pointrend module to smooth the mask edges of segmented instances and to minimize measuring error of antenna parameters. Pointrend is mainly composed of three parts: point selection strategy, point-wise feature representation, and point head. The point selection strategy mainly relies on flexible and adaptive selection of suitable sampling points. It means that more points are sampled in border areas, and fewer points are sampled in non-border areas. After the point selection strategy, the most uncertain points are selected to re-predict the segmentation results of these points. Fine-grained features and coarse prediction features expounded in Figure 6 are used to calculate the feature of each point as the input of the point head to conduct re-prediction. After point-wise feature representation calculates the feature of each special point (the most uncertain point), the feature of each point is processed through point head to obtain a new prediction result. The steps above can be summarized as the special processing of points at the boundary position, aiming to further refine the result. Hence, the mask edge of antennas in mobile communication base station can be delicately smoothed by pointrend to enhance

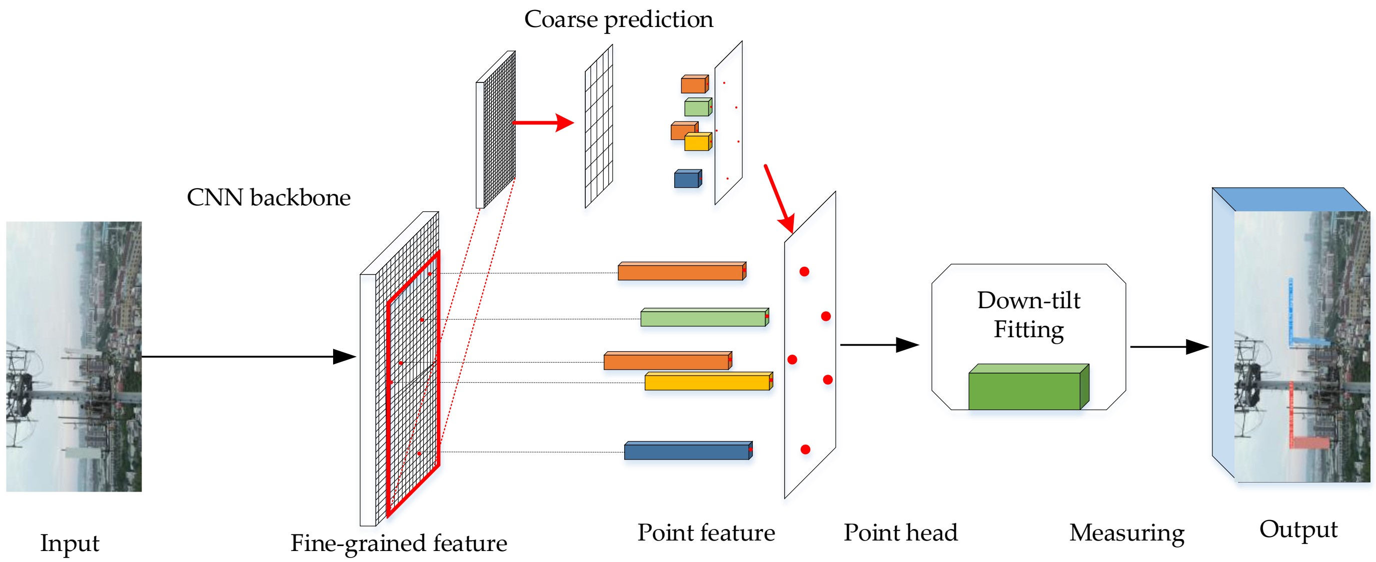


Fig. 6. The structure of PointRend with down-tilt fitting.

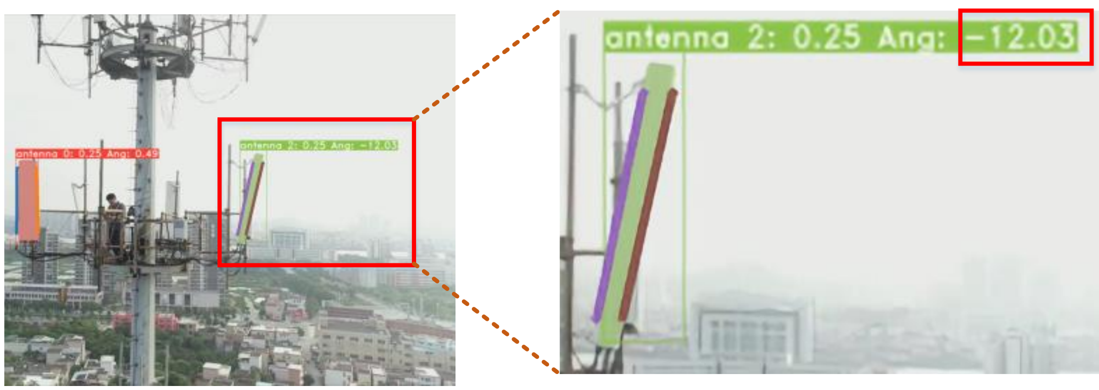


Fig. 7. Fitting pixel coordinates of the mask edge to calculate the antennas down-tilt angle.

fitting accuracy and parameters measurement accuracy. Then, Down-tilt fitting in pointrend module is presented here. Least squares method [39], the most basic and important data fitting method, is used to learn the relationship between the masked pixels to quantify the antenna parameters. With its wide range of applications in modeling, least squares method plays a critical role in the research of a large amount of data analysis. So as to abate the measurement error and calculation cost, the 3/5 pixel value coordinates (the top and bottom 1/5 of the mask are removed) of the edges on both sides of the mask are selected to fit the down-tilt angle of antennas in mobile communication base station. The process is demonstrated in figure 7.

Above all, assume the parameter estimation equation as (4), in which x and y represent the abscissa point and the ordinate point of the pixel value coordinate point selected in the mask respectively, β denotes the correlation coefficient, and μ refers to the bias.

$$y = \beta_0 + \beta_1 x + \beta_2 x + \cdots + \beta_k x + \mu \tag{4}$$

If m sets of pixel value coordinates are taken into (4), results can be written as (5).

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \mu_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \mu_2 \\ \vdots \\ y_m = \beta_0 + \beta_1 x_{m1} + \beta_2 x_{m2} + \cdots + \beta_k x_{mk} + \mu_m \end{cases} \tag{5}$$

The matrix expression form of (5) is:

$$y = BX + \mu \tag{6}$$

Thereby, the sum of squares of the difference between the observed value and the estimated value of the explained variable is:

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = e'e = (Y - X\hat{B})'(Y - X\hat{B}) \tag{7}$$

According to the principle of least squares, parameter estimate is the solution of following equations.

$$\frac{\partial}{\partial \hat{B}} (Y - X\hat{B})'(Y - X\hat{B}) = 0 \tag{8}$$

So the least squares estimate of parameter is:

$$\hat{B} = (X'X)^{-1} X'Y \tag{9}$$

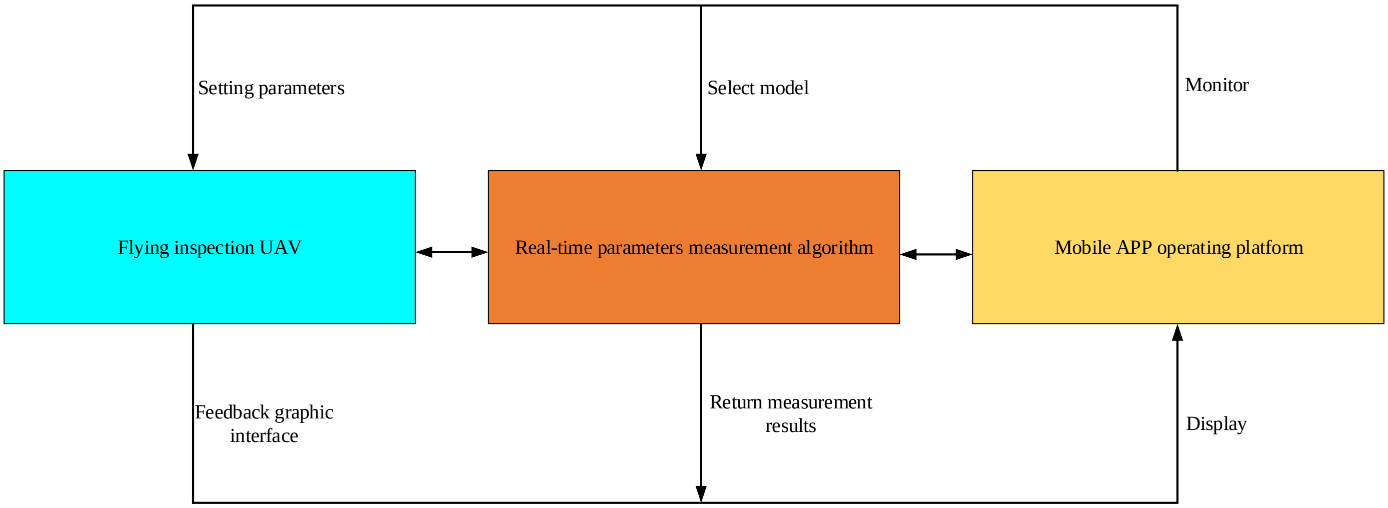
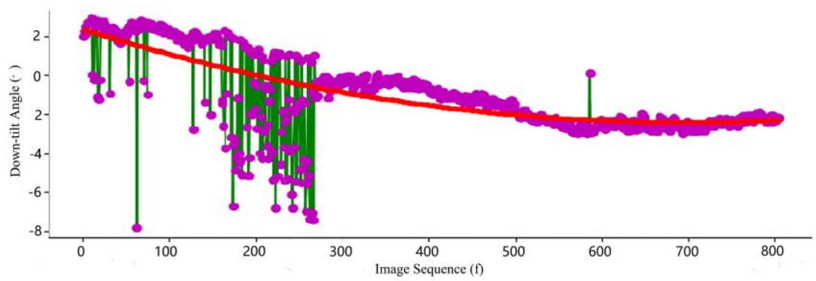
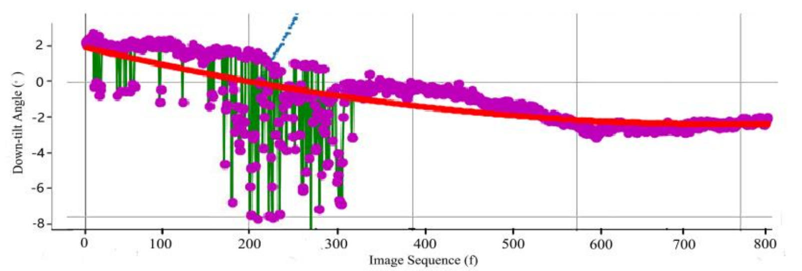


Fig. 8. The basic framework of UAV automatic measurement system.



(a)



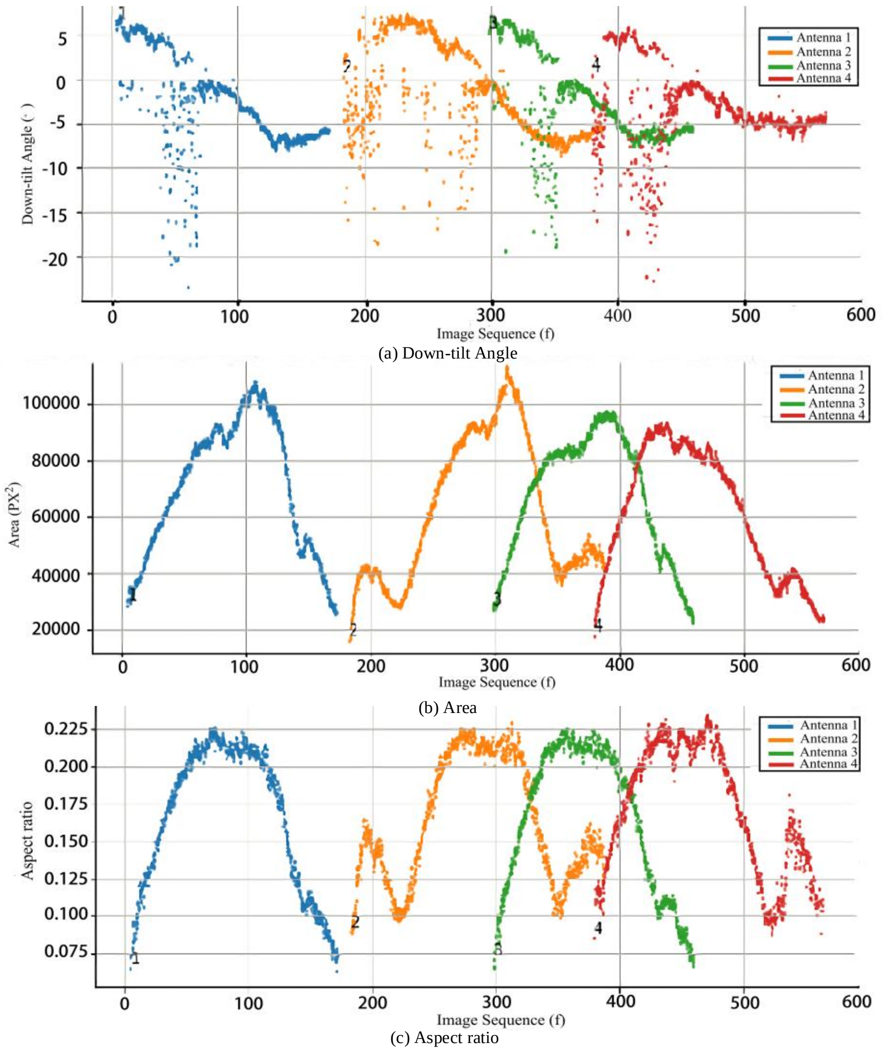
(b)

Fig. 9. Antennas number separation. (a) and (b) are the down-tilt angle measurement results of different antennas respectively. The unit of Down-tilt angle is degree expressed as (°), while the unit of Image Sequence is frame abbreviated as (f).

E. Transfer Learning

Despite of the extensive applications [40] of deep convolute-ion neural networks thanks to its ability to learn rich image representations, there still requires a large amount of data to help learn these features. And unfortunately, the dataset used in this paper precisely falls short of data which will cause overfitting and training difficulty, but transfer learning can dispose of the problem appropriately. Transfer learning refers to an approach that transfers knowledge, both low-level and mid-level features, from a related domain to a new problem and thus greatly decreases the amount of data from the new domain. Therefore, if transfer learning is applied to transfer an image recognition model trained on massive images to mobile communication base station antennas segmentation task, the obsession caused by small samples can be effectively alleviated and the same favorable performance can be obtained as massive data tasks do. In this work, the adoption of transfer learning can successfully offset the impact of small communication base station antenna samples on the performance of AntennaNet detailed in Figure 2 and Figure 3. The experimental results strongly prove the effectiveness of transfer learning.

The theory of transfer learning can be comprehended as follows. Transfer learning, given a source domain D_s and a target domain D_T, which correspond to learning tasks T_s and T_T, intends to improve the learning ability of target prediction function $f(T(\cdot))$ in D_T utilizing the knowledge of D_s and T_s, in which $D_s \neq D_T$ or $T_s \neq T_T$. In this study, small mobile communication base station antenna samples can be regarded as D_T, and massive images of multiple categories on MS COCO [41] or ImageNet [42] are considered as D_s. Transfer



(a) Down-tilt Angle

(b) Area

(c) Aspect ratio

Fig. 10. Process of measuring the number of antennas by different parameters. The unit of Area is pixel^2 donated to (PX^2).

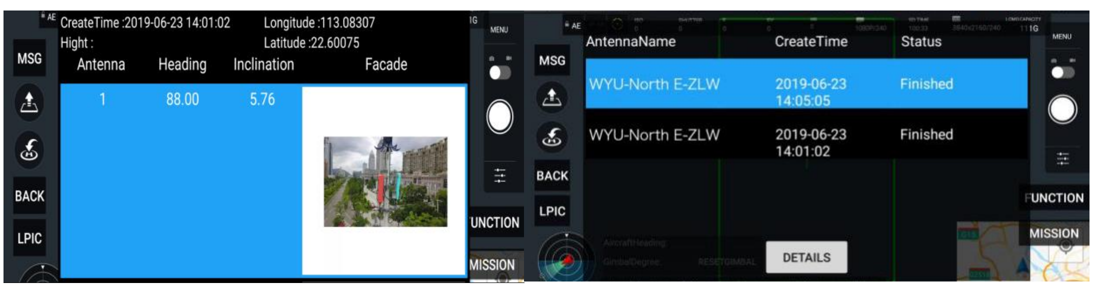


Fig. 11. Mobile terminals android APP interface.

learning strategy is implemented to share some parameters to make full use of the similarities between the models for better performance.

IV. Automatic Parameters Measurement System

With the proposal of AntennaNet, antenna can be effortlessly segmented in pixel-level to enable antenna parameters measurement. Fully playing to the strength of UAV in obtaining video data of mobile communication base station, we develop an Android app to adapt to AntennaNet and UAV so that we can output and display the communication base station video from UAV's perspective and the quantity of antennas and measurement results for each antenna on the mobile terminal.

A. Framework and Flowchart of Measurement System

Figure 8 explains the framework of the new faster measurement system that we designed and developed. The whole process starts with turning on the UAV, setting the UAV's orbiting height and radius on APP interface and selecting the parameters measurement model to be called during UAV's orbiting process. After UAV flies around the mobile communication base station, the algorithm proposed above processes the video taken by UAV, determines the

Algorithm 1:
Flowchart of automatic measurement of antenna parameters

Input:
Video data X of mobile communication base station with antenna captured by UAV.
Output:
Number n of antennas, down-tilt d, area a and aspect ratio r of each antenna are all displayed on the APP interface.
Step1: Given a video data X of mobile communication base station with antenna captured by UAV;
Step2: Frame video X and use them as the input to proposed algorithm;
Step3: Proposed algorithm processes the framed images in turn, and generates prototype masks p and mask coefficients m;

Step4: The most uncertain points in prototype masks are selected to re-predict the segmentation results to generate a new prototype masks p';
Step5: Multiply m and p', and get the mask M of each antenna after threshold filtering;
Step6: Perform linear fitting on the edge coordinates of obtained mask M to acquire the antenna parameters such as down-tilt d, area a and aspect ratio r;
Step7: Determine the number n of antennas by pixel coordinates and threshold (the empirical value of multiple verification is 50);
Step8: Combine all masked images into video in sequence and output the number of antennas n, down-tilt d, area a and aspect ratio r.

number of antennas through threshold and pixel coordinates, and measures the parameters (aspect ratio, down-tilt angle, area, etc.) of each antenna in a much faster speed. Finally, results of algorithm measurement and the antenna attitude calculated based on the UAV video are fed back to mobile APP interface.

Flowchart of automatic measurement of antenna parameters can be reviewed in algorithm 1 for mobile communication base station. It includes a flying inspection UAV, a much faster parameters measurement algorithm, and a mobile terminal Android APP operating platform.

B. Visualization of APP and Parameters Measuring results

Through the video captured by UAV during its flights around mobile communication base station, the number of antennas in the video and the relevant parameters of each antenna can be obtained with higher speed with the help of the algorithm proposed in this paper. Diagrams in Figure 9 list different parameters of the same antenna obtained by processing the video, which apparently certifies that wrong segmentation or missing segmentation barely affects the determination of the same antenna. And the horizontal axis of each sub-figure represents the number of frames, and the vertical axis represents the down-tilt angle. However, a different story is told in a multi-antenna video because we found that different antenna parameters can be used to ascertain the number of antennas, which is determined by threshold value and the discrepancy between pixels coordinate. Figure 10 reveals the process of determining the quantity of antennas with various parameters (down-tilt angle, area and aspect ratio). Among them, the horizontal axis of each sub-figure represents the number of frames, and the vertical axis represents the down-tilt angle, area, and aspect ratio. Finally, the video taken by UAV and the results measured via the algorithm are displayed in mobile terminal Android APP Interface which can be referred to Figure11.

V. EXPERIMENTAL RESULTS AND ANALYSIS

The computer employed in the experiments is configured as Core i9cpu with 64G running memory, and 1080Ti graphics cards with 11G memory. The experiments are conducted with ubuntu18.04 operating system, and pytorch1.1.0. We constructed a database named Antennascapes, which is a low-altitude natural scene database with 2categories of 1500 train images and 219 test images, some of which can be observed in Figure 12. All images are higher resolution (1024×2048 pixels) compared to ImageNet and MS COCO and have better pixel accuracy for ground truth instance segmentation, as they are captured by the flight of UAV around the mobile communication base station. 1500 train images are artificially accurately labeled by VGG Image Annotator (VIA), while 219 testing images are adjusted to a specific angle (1°,4°,5°,6°,7°,8°,9°,12°,15°) by professional workers to verify the accuracy of measurement results. Specific angle distribution of 219 testing images can be examined in Table 1.

TABLE 1
SPECIFIC ANGLE DISTRIBUTION OF TEST IMAGES

Testing set with angel information									
Angel	1°	4°	5°	6°	7°	8°	9°	12°	15°
Number	13	25	22	30	25	23	22	39	20

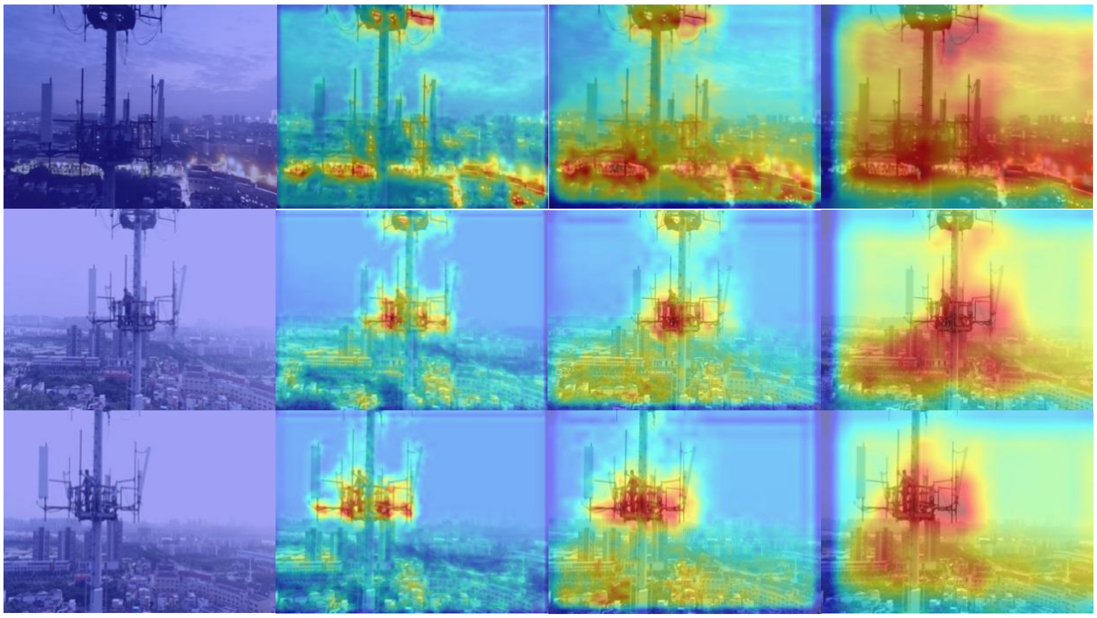
A. Experiment description

All models are trained with the batch size 16 on GPU using MS COCO [41] pre-trained weighs. We neither freeze the pre-trained batch specification, nor add an additional BN layer because the setting batch size can sufficiently satisfy the use of batch norm. We train with (Stochastic Gradient Descent) SGD for 100000 iterations at an initial learning rate of 10^{-4} and divide by 10 at iterations 20000, 70000, 80000, and 85000, with a weight decay of 5×10^{-5}, a momentum of 0.9, score threshold set to 0.247 and all data augmentations in SSD [38]. It takes 10 to12 days to handle our 1500 training images.

1) Analysis of Dual Attention Mechanism

As the distribution of section 3.2 and shown in Figure 1, the dual attention mechanism with the channel attention placed in front of spatial attention is inserted between the convolutional layers C1 and C2, C2 and C3, C3 and C4, C4 and C5. We combine different feature extraction networks (resnet101, resnet50, darknet53) with FPN network to verify the effectiveness of attention mechanism module in this paper under the pre-training model of MS COCO, with comparative results detailed in Table 2, in which segmentation accuracy indicates the ratio of segmented non-back antenna to all non-back antennas. Furthermore, since it is considered that less than 1-degree error exists between the measured value of the antenna down-tilt angle and the actual value, which can be ignored in the actual scene application, the fitting accuracy is defined as the ratio of the quantity of front-side antennas which are segmented and their down-tilt angles are measured within 1 degree with comparison to the real down-tilt angles of the number of detected front-side antennas. The difference between

Fig. 12. Samples of Antennascapes, the first four images are selected from training set; the last four from testing set. They are the antenna images of mobile communication base stations at different times, weathers, and locations to increase the generalization of model trained by these training images.



(a) (b) (c) (d)

Fig. 13. Attention map output by each dual attention mechanism module on resnet50. The first, the second and the third row are attention maps with antenna down-tilt angles of 5, 12 and 15. (a) is the attention map between Convolutional layer C1 and C2; (b) is the attention map between Convolutional layer C2 and C3; (c) is the attention map between Convolutional layer C3 and C4; (d) is the attention map between Convolutional layer C4 and C5.

TABLE 2
COMPARISON TABLE OF RESULTS OF DIFFERENT FEATURE EXTRACTION NETWORKS WITH OR WITHOUT ATTENTION MECHANISM

	Resnet101(%)		Resnet50(%)		Darknet53(%)	
	W	WO	W	WO	W	WO
Segmentation accuracy	93.54	93.22	**98.75**	87.73	98.71	97.07
Fitting accuracy	71.63	58.83	**94.50**	70.83	67.54	55.36

TABLE 3
SEGMENTATION ACCURACY COMPARISON

Method	Resnet101(%)	Resnet50(%)	Darknet53(%)
Basic	93.22	87.73	97.07
Basic+pointrend	93.51	87.69	96.89
Basic+attention	93.54	**98.75**	98.71
Basic+attention+ pointrend	93.62	**98.70**	98.74

TABLE 4
FITTING ACCURACY COMPARISON

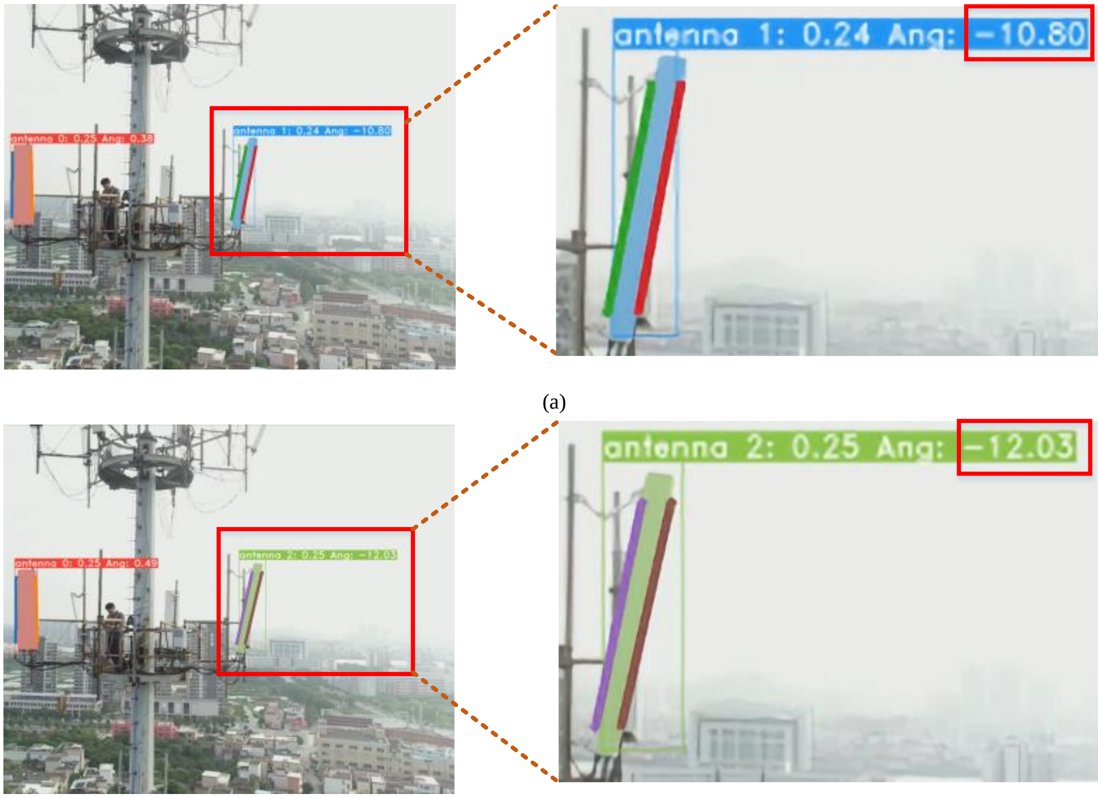
Method	Resnet101(%)	Resnet50(%)	Darknet53(%)
Basic	58.83	70.83	55.36
Basic+pointrend	62.38	76.14	69.17
Basic+attention	71.63	94.50	67.54
Basic+attention+ pointrend	78.85(+7.22)	**96.86(+2.36)**	71.42(+3.88)

TABLE 5
COMPARISON RESULTS OF TRANSFER DIFFERENT MODELS

	MS COCO[41]		ImageNet[42]	
	W	WO	W	WO
Segmentation accuracy	**87.73%**	72.90%	86.91%	72.90%
Fitting accuracy	**70.83%**	59.68%	70.09%	59.68%

“W” and “WO” lies in the presence and absence of a dual attention mechanism module in the network structure respectively.

Data in Table 2 illustrates that the highest segmentation accuracy reaching 98.75% and fitting accuracy surging to 94.50% occur in the feature extraction network called Resnet50 with dual attention module. When compared with the figures in the same network but without attention mechanism in its next column, they witness a significant improvement in both segmentation accuracy and fitting accuracy. Figure 13 involves the attention maps output by each dual attention mechanism module in the feature extraction network resnet50 with FPN. They express the idea that the redder the images seem, the greater contribution of these areas can be made to the final prediction result. The opposite theory can be applied to the blue area in these images.



(a)

(b)

Fig. 14. Measurement result of the same image whose antenna actual down-tilt angle is 12 degrees. (a) without pointrend module, the measurement result is 10.80; (b) with pointrend module, the measurement result is 12.03.

TABLE 6
PRECISION COMPARISON WITH RELATED METHODS ON THE SAME DATASET

Method	Backbone	Detection accuracy	Segmentation accuracy	Fitting accuracy
Mask R-CNN[33]	Resnet101	-	**99.44%**	58.2%
Faster R-CNN[34]	Resnet101	89.85%	-	-
YOLOv3[36]	Darknet53	92.72%	-	-
Basic	Darknet53	97.07%	97.07%	55.36%
Basic+pointrend	Darknet53	96.89%	96.89%	69.17%
Basic+attention	Darknet53	98.71%	98.71%	67.54%
AntennaNet(ours)	Darknet53	98.74%	98.74%	71.42%
Basic	Resnet101	93.22%	93.22%	58.83%
Basic+pointrend	Resnet101	93.51%	93.51%	62.38%
Basic+attention	Resnet101	93.54%	93.54%	71.63%
AntennaNet(ours)	Resnet101	93.62%	93.62%	78.85%
Basic	Resnet50	87.73%	87.73%	70.83%
Basic+pointrend	Resnet50	87.69%	87.69%	76.14%
Basic+attention	Resnet50	**98.75%**	**98.75%**	94.50%
AntennaNet(ours)	Resnet50	**98.70%**	**98.70%**	**96.86%**

2) Analysis of Pointrend with Down-Tilt Fitting

As the distribution of section 3.3 and shown in Figure 1, the PointRend module is added after protonet to form more refined masks. For verifying the effectiveness of pointrend module, this work respectively adds pointrend to basic structure and the basic structure with attention mechanism under different feature extraction networks to test their segmentation and fitting accuracy. In Table 3 and Table 4 we find segmentation accuracy comparison and fitting accuracy comparison, through which it can be concluded that the pointrend module contributes more to fitting accuracy than segmentation accuracy. It is obvious that fitting accuracy has been greatly improved after adding pointrend module to the proposed method regardless of the feature extraction network. On the contrary, the addition of the pointrend module does not come in significantly for the segmentation accuracy.

In the same feature extraction network resnet50 with attention mechanism, using the pointrend module as the control variable, we separately trained the network with and without the pointernd module for the same parameters. Then we tested the same image whose down-tile angle is 12 degree in testing set, and the results are shown in Figure 14. Without pointrend added to AntennaNet, we measure the down-tilt angle of front side antenna at 10.80 degree with a discrepancy of 1.2 degrees,

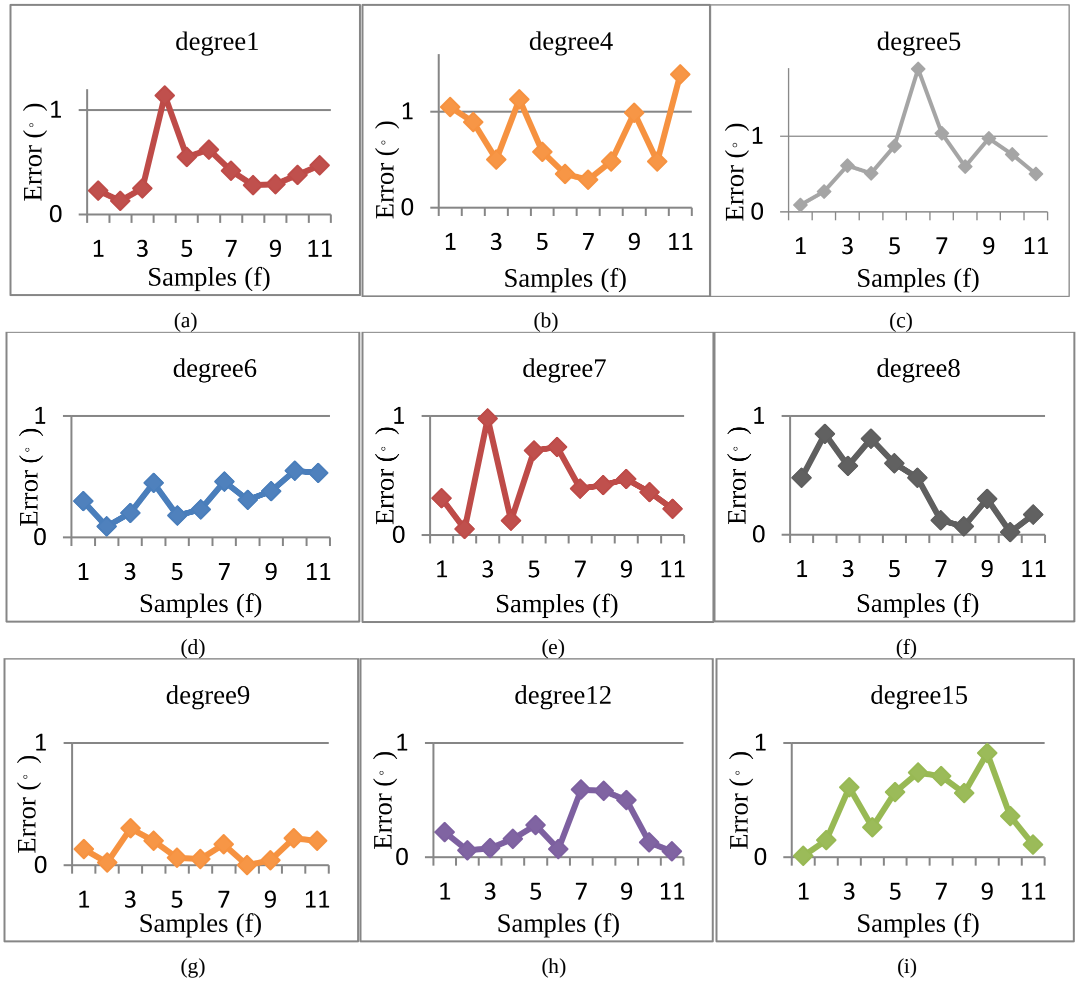


Fig. 15. Error analyses of measured value and the actual value of each antenna down-tilt angle in testing set. (a)-(i) represents the fluctuation of the measured value when the actual value from 1 degree to 15 degree respectively. Among them, the horizontal axis represents the number of testing samples, and the vertical axis donates to the absolute error between the measured value of the antenna parameters and the true value.

which deviates from the industry standards, while with pointrend added to AntennaNet, the down-tilt angle of front side antenna at 12.03 degree barely shares any difference from the actual value. The results prove the astonishing effectiveness of the PointRend module in improving fitting accuracy in antenna parameters measurement tasks.

3) Analysis of Transfer Learning

Since the amount of training samples positively correlates to the accuracy of the trained model, transfer learning method is drawn into to solve the problem of insufficient samples in this paper. We employ the weights in the model trained by MS COCO [41] and ImageNet [42] as our initialization weights, to avoid the need of random initialization for the weights. Table 5 delivers the data of the segmentation and fitting accuracy obtained by using only resnet50 and FPN modules to perform weight transfer on these two models. The "W" column indicates the introduction of transfer learning to AntennaNet while "WO" column does the opposite.

This comparison conveys the message that by initializing the weight of MS COCO dataset model, a higher segmentation and fitting accuracy can be acquired in MS COCO dataset model than those in ImageNet dataset model for the antenna segmentation of our mobile communication base station and antenna down-tilt measurement tasks. It is worth noting that assisted with transfer learning, both of the accuracies have witnessed a favorable growth. Apparently, transfer learning succeeds in tremendously raising the segmentation accuracy of small samples tasks.

TABLE 7
SPEED COMPARISON WITH RELATED METHODS ON THE SAME DATASET

Method	Backbone	Time (s)	
		Inference Time	Testing Time
Mask R-CNN[33]	Resnet101	4.74	1.21
Faster R-CNN[34]	Resnet101	2.21	0.81
YOLOv3[36]	Darknet53	1.46	0.01
Basic	Darknet53	0.68	0.12
Basic+pointrend	Darknet53	0.94	0.26
Basic+attention	Darknet53	0.71	0.13
AntennaNet(ours)	Darknet53	0.99	0.28
Basic	Resnet101	0.71	0.10
Basic+pointrend	Resnet101	1.02	0.31
Basic+attention	Resnet101	0.77	0.15
AntennaNet(ours)	Resnet101	1.03	0.33
Basic	Resnet50	0.61	0.10
Basic+pointrend	Resnet50	0.96	0.29
Basic+attention	Resnet50	0.65	0.14
AntennaNet(ours)	Resnet50	**0.98**	**0.31**

TABLE 8
COMPARISON OF ACTUAL AND MEASURED ANTENNA DOWN-TILT ANGLES BETWEEN PROPOSED METHOD AND PREVIOUS METHODS

Actual angle	1°	4°	5°	6°	7°	8°	9°	12°	15°
Mean error [8]	1.34	0.50	0.16	0.54	0.99	0.61	0.23	1.40	1.09
Mean error of proposed method	0.42	0.34	0.75	0.26	0.09	0.19	0.09	0.17	0.35
Variance[8]	0.25	1.60	0.96	1,40	9.19	1.14	2.61	4.85	0.54
Variance of proposed method	0.09	0.46	0.17	0.15	0.43	0.30	0.06	0.14	0.12

B. Precision Comparisons with Other SOTA Methods

Comparisons are conducted among detection, segmentation and fitting accuracy of our method with current one-stage, two-stage detection and instance segmentation algorithms based on deep learning methods as well as on different modules of our proposed method. Table 6 unravels that when the feature extraction network remains to be Resnet50, the method proposed in this paper performs best on all aspects, greatly surpassing the previous work [8] and some current one-stage and two-stage detection and instance segmentation algorithms based on deep learning methods in terms of fitting accuracy. Even though the detection accuracy of AntennaNet as a one-stage instance segmentation framework, is slightly lower than that of Mask R-CNN as a two-stage instance segmentation algorithm, it still surpasses that of Faster R-CNN YOLOv3 and hardly affects the proposed method in this paper to determine the number of antennas and output the parameters of each antenna.

C. Speed Comparisons with Other SOTA Methods

Afterwards, we defined that Time in Table 7 is expressed as the sum of the inference time and testing time, where the inference time includes the time to load the model and infer the testing data obtained based on the trained content, while the testing time includes the time for instance segmentation or detection and parameters measurement. Various time costs of our method with current one-stage, two-stage detection and instance segmentation algorithms based on deep learning methods are studied. Likewise, the same comparison is conducted on our proposed method with different modules. Table 7 suggests that the proposed method is less time-consuming than two-stage detection and instance segmentation algorithms, and can realize fast measurement, the measurement time is 0.31s in the feature extraction network named Resnet50. Also, it exceeds the detection speed of Faster R-CNN without the mask step. When taking not only the accuracy but also the speed into consideration, our proposed method bears the comprehensively best outcome among all the methods discussed in the experiments.

D. Antenna Measuring Error

It is inspiring that our method serves well to enhance the performance of the current state-of-the-art methods on mobile communication base station antenna parameter measurement task, which attests to and underlines the effectiveness and importance of dual attention mechanism and Pointrend by transfer learning. We intend to work out which model performs best through comparing the measured angle and the real angle with experimental data filled in Table 8. It is indicated that the error value of each group of the same angle image measurement result falls within 1 degree, which fully meets the industry standard. Meanwhile, the mean errors of the proposed method and the previous method [8] have been compared with each other and a variance analysis of the proposed method has been conducted. From table 8, it can be reached that the fitting accuracy summarized from mean error comparison and system stability generalized from variance comparison of the proposed method AntennaNet are much better than those of the previous methods. The error analysis using proposed network and system of measured value and actual value of each antenna down-tilt angle in testing set is shown in figure 15. Among them, horizontal axis represents the quantity of measured images at the same angle, and vertical axis denotes the error between measured value and actual value. To our surprise, no obvious fluctuation can be spotted for the data at the same angle, which bears out the extremely favorable reliability and stability of the system proposed in this paper.

VI. CONCLUSION

A fully automatic high-speed and high-precision measurement system for antenna parameters is unprecedentedly proposed to replace traditional mobile communication base station antenna parameters measurement methods which mainly involve manual measurement. We design a one-stage instance segmentation framework with dual attention mechanism module as well as the pointrend module. Combining transfer learning and linear fitting module can realize a much faster measurement of parameters for mobile communication base stations antenna. With comprehensive experiments, it has been confirmed that appropriate training strategies employed can engender prominent performance in automatic antenna parameters high-speed measurement. Compared with current one-stage, two-stage detection and instance segmentation algorithms based on deep learning methods including Mask R-CNN, Faster R-CNN and YOLOv3, our method significantly excels in shorter recognition time and higher fitting accuracy while compared with the traditional methods, our method outperforms in less artificial participation, greater work efficiency and higher security. In addition, it is estimated that our high-speed and high-precision antenna parameters measurement method will act an essential role in the 5G era characterized with high bandwidth, low latency and high concurrency.

Our proposed method, apart from its standalone utility, is trusted with providing an effective and practical solution for mobile communication base station antenna parameters measurement. It stands an extremely promising future as it is an unparalleled system that can simultaneously realize high-speed measurement and ensure high segmentation accuracy. Its strengths facilitate its usage in the field of low-altitude measurement in more complex scenarios and effectively reduce the dangers arisen from manual measurement.

REFERENCES

[1] Yang. J *et al.*, “Optimal base station antenna down-tilt in downlink cellular networks,” *IEEE Transactions on Wireless Communications.*, vol. 18, no. 3, pp. 1779-1791, March. 2019.

[2] Shen. F et al., “Low-Cost Dual-Band Multi-polarization Aperture-Shared Antenna with Single-Layer Substrate,” *IEEE Antennas and Wireless Propagation Letters.*, vol. 18, no. 7, pp. 1337–1341, July. 2019.

[3] Mohammed. A, Hussein. A, Abderraouf. M, and Sheikh. S, “Directive Wideband Cavity Antenna with Single-Layer Meta-Superstrate,” *IEEE Antennas and Wireless Propagation Letters.*, vol. 18, no. 9, pp.1771-1774, Sep. 2019.

[4] Bawri. S, Islam. S, Sahaq. K, *et al.*, “Multilayer Base Station Antenna at 3.5 GHz for Future 5G Indoor Systems,” *in Proceedings of 2019 First International Conference of Intelligent Computing and Engineering*. Mukalla, Yemen, 2019, pp. 978-981.

[5] Kun. X, Guo. T, Zhou. Y, *et al.*, “Antenna Mechanical Pose Measurem -ent Based on Structure from Motion,” *ZTE Communications*, vol. 4, no. 16, pp.38-45, 2020.

[6] Pedras. V, Sousa. M, Vieira. P, *et al.*, “Antenna tilt optimization using a novel QoE model based on 3G radio measurements”, *in Proceedings of 20th International Symposium on Wireless Personal Multimedia Communications* (WPMC), Indonesia, Yogyakarta, 2017, pp. 124-130.

[7] Massive-Mimo Antenna Measurement Device and Method of Measuring Directivity Thereof by Takashi. Kawamura. (2017, January 26). *Patent 20170222735*.[Online].Available:http://www.freepatentsonline.com/y2017/0222735.html.

[8] Zhai. Y, Ke. Q, Xu. Y, *et al.*, “Mobile Communication Base Station Antenna Measurement Using Unmanned Aerial Vehicle,” *IEEE Access*, vol. 7, pp. 119892-119903, Aug. 2019.

[9] Pan. S, and Yang. Q. “Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp.1345 – 1359, Oct. 2010.

[10] Lin. T, Goyal. P, Girshick. R, *et al.*, “Focal Loss for Dense Object Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[11] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 936-944.

[12] LeCun. Y, Bengio. Y, and Hinton. G, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May, 2015.

[13] L. Yang, L. Jing, J. Yu, and M, K, “Learning Transferred Weights From Co-Occurrence Data for Heterogeneous Transfer Learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2187-2200, Nov. 2016.

[14] S. Yildirim, D. Jothimani, C. Kavaklioglu, *et al.*, “Building Domain-Specific Lexicons: An Application to Financial News,” *in Proceedings of 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, Istanbul, Turkey,2019, pp. 23-26.

[15] W. Yang, W. Luo, W. Kang, Z. Huang and Q. Wu, "FVRAS-Net: An Embedded Finger-Vein Recognition and AntiSpoofing System Using a Unified CNN," *in IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 11, pp. 8690-8701, Nov. 2020.

[16] Mahdavifar. S, Ali. A. “Application of Deep Learning to Cybersecurity: A Survey,” *Neurocomputing*, vol. 347, pp. 149–176, June. 2019.

[17] Mathews. M, Chandra. M, and Barner. K, “A Novel Application of Deep Learning for Single-Lead ECG Classification,” *Computers in Biology and Medicine*, vol. 99, pp. 53–62, Aug. 2018.

[18] Li. D, Zhao. D, Zhang. Q, *et al.*, “Reinforcement Learning and Deep Learning Based Lateral Control for Autonomous Driving,” *IEEE Computational Intelligence Magazine*, vol. 14, no. 2, pp. 83–98, May. 2019.

[19] Etxeberria. M, Labayen. M, Zamalloa. M, *et al.*, “Application of Computer Vision and Deep Learning in the railway domain for autonomous train stop operation,” *in Proceedings of the 2020 IEEE International Symposium on System Integration (SII)*, Hawaii, USA, 2020, pp. 943-948.

[20] Safadinho. D, Ribeiro. R, *et al.*, “UAV Landing Using Computer Vision Techniques for Human Detection,” *Sensors*, vol. 20, no. 3, pp. 613-648, Jan. 2020.

[21] Wang. T, Qin. R, Chen. Y, *et al.*, “A reinforcement learning approach for UAV target searching and tracking,” *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4347-4364, Feb. 2019.

[22] J. Fu, et al., "Dual Attention Network for Scene Segmentation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019, pp. 3141-3149.

[23] Vaswani. A, Shazeer. N, Parmar. N, *et al.*, “Attention Is All You Need,” *in Proceedings of the International Conference on Neural Information Processing Systems*, Vancouver, Canada, 2017, pp. 5998–6008.

[24] Shan. B, Yong. F, “A Cross Entropy Based Deep Neural Network Model for Road Extraction from Satellite Images,” *Entropy*, vol. 22, no. 5, pp. 535-550, Apr. 2020.

[25] Kirillov. A, Wu. Y, He K, *et al.*, “PointRend: Image Segmentation As Rendering,” *in Proceedings of Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 2020, pp. 9799–9808.

[26] T. Phan, S. Sultana, T. Nguyen, and T. Bauschert, “Q- TRANSFER: A Novel Framework for Efficient Deep Transfer Learning in Networking,” *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Dalian, China, 2020, pp. 146-151.

[27] Sun. Q, Liu. Y, Chua. T, *et al.*, “Meta-Transfer Learning for Few-Shot Learning,”*in Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, California, USA,2019, pp. 403–412.

[28] Talo. M, Baloglu. U, Yildirim. O, *et al.*, “Application of Deep Transfer Learning for Automated Brain Abnormality Classification Using MR Images,” *Cognitive Systems Research*, vol. 54, pp. 176–188, May. 2019.

[29] Li. Y, Zeng. J, Shan. S, Chen. X. “Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism,” *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, May. 2019.

[30] Liu. G, Guo. J, “Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification,” *Neurocomputing*, vol. 337, pp. 325–338, Apr. 2019.

[31] Woo. S, Park. J, Lee. J, Kweon. I, “CBAM: Convolutional Block Attention Module,” *in Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 3–19.

[32] Bolya. Daniel, *et al.*, “YOLACT: Real-Time Instance Segmentation.” *in Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Korea, Seoul, 2019, pp. 9157–9166.

[33] He. K, Georgia. G, Piotr. D, “Mask R-CNN,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, Feb. 2020.

[34] Ren. S, He. K, Girshick. R, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, June. 2017.

[35] Li. Y, Qi. H, Dai. J, *et al.*, “Fully Convolutional Instance-aware Semantic Segmentation,” *in Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, University of Maryland at College Park, USA, 2017, pp. 4438-4446.

[36] Redmon. J, Ali. F, “YOLOv3: An Incremental Improvement,” *ArXiv* Preprint ArXiv:1804.02767, Apr. 2018.

[37] Redmon. J, Divvala. S, Girshick. R, “You Only Look Once: Unified, Real-Time Object Detection,” *in Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, LAS VEGAS, USA, 2016, pp. 779-788.

[38] Liu. W, Anguelov. D, Erhan. D, *et al.*, “SSD: Single Shot Multi-Box Detector,” in Proceedings of European Conference on Computer Vision (ECCV), Amsterdam, Netherlands, 2016, pp. 21–37.

[39] Tibshirani. R, Tibshirani. R, Taylor. J, “Least Angle Regression,” *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[40] Ying, Z, Xuan, C, Zhai, Y, *et al.*, “TAI-SARNET: Deep Transferred Atrous-Inception CNN for Small Samples SAR ATR,” *Sensors*, vol. 20, no. 6, pp. 1724-1744, Mar. 2020.

[41] Lin. T, Maire. M, Belongie. S, *et al.*, “Microsoft COCO: Common Objects in Context,” *in Proceedings of European Conference on Computer Vision(ECCV)*, Zurich, Switzerland, 2014, pp. 740–755.

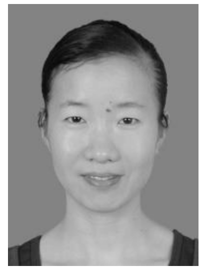
[42] Jia. D, Wei. D, Socher. R, *et al.*, “ImageNet: A Large-Scale Hierarchical Image Database,” *in Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Florida, USA, 2009, pp. 248–255.

Yikui Zhai is an Associate Professor at Wuyi University, Guangdong, China. He was a visiting scholar at the University of Milan, Department of Computer Science, since 2016. He received his Ph. D degree in Signal and Information Processing at Beihang University in June 2013. He received the Bachelor and Master degree in Optical Electronics Information and Communication Engineering, and Signal and Information Processing from Shantou University, Guangdong, China in 2004 and 2007 respectively. Since October 2007, he has been working at Department of Intelligence Manufacturing, Wuyi University, Guangdong, China, and his research interests include: image processing, deep learning and pattern recognition.

Qirui Ke is currently pursuing his Master in Department of Intelligence Manufacturing, Wuyi University. He received his B.S. degree at Hubei University of Arts and Science in 2018. His research interests include: pattern recognition and automatic measuring.

Xinglin Liu is an Associate Professor at Wuyi University, Guangdong, China. He received the B.S. at Wuyi University in 1999, received his M.S. degrees at ChongQing University in 2005, and also received his Ph. D degree in Computer Application Technology at South China University of Technology in 2012. His research interests include: text knowledge acquisition, intelligent calculation, intelligent recommendation, intelligent signal processing, and pattern recognition.

Wenlve Zhou is currently pursuing his Master in Department of Intelligence Manufacturing, Wuyi University. He received his B.S. degree in Beijing Institute of Technology, Zhuhai in 2019. His research interests include: image analysis and pattern recognition



Ying Xu received the B.S. and M.S. degrees in Automation, and Control Engineering from Wuhan University of Science and Technology in 2004 and 2008, respectively. She also received her Ph.D. in South China University of Technology in 2013. She joined Wuyi University in 2008, and her research interests include: intelligent signal processing and pattern recognition.

Zilu Ying is a Full Professor at Wuyi University, Guangdong, China. He received his B.S., M.S. and Ph.D. degrees in Electrical Information Engineering from Beihang University. He is an executive director of the Guangdong Society of Image and Graphics and a member of the Signal Processing Branch of the Chinese Institute of Electronics. His research interests include: signal processing and pattern recognition.

Junying Gan is a Full Professor at Wuyi University, Guangdong, China. She received her B.S., M.S. and Ph.D. degrees in Electrical Information Engineering from Beihang University in 1987, 1992 and 2003, respectively. She joined Wuyi University in 1992. She is the executive director of Guangdong image graphics association, and an IEEE member. She has published more than 50 journal papers, and has received several provincial technology awards. Her research interests include: biometric extraction and pattern recognition.

Junying Zeng was born in Ganzhou, Jiangxi province, China, in 1977. He got his Ph.D. degree in physical electronics from Beijing University of Posts and Tele-communications, Beijing, China, in 2008. His research interests include: intelligent signal processing and pattern recognition.

Ruggero Donida Labati received the Ph.D. degree in computer science from the Universitá degli Studi di Milano, Crema, Italy, in 2013. He has been an Assistant Professor in Computer Science with the Universitá degli Studi di Milano since 2015. He has been a Visiting Researcher with Michigan State University, East Lansing, MI, USA. Original results have been published in more than 50 papers in international journals, proceedings of international conferences, books, and book chapters. His current research interests include intelligent systems, signal and image processing, machine learning, pattern analysis and recognition, theory and industrial applications of neural networks, biometrics, and industrial applications. He is an Associate Editor with the Journal of Ambient Intelligence and Humanized Computing (Springer).

Vincenzo Piuri is IEEE Fellow, ACM Fellow, and Full Professor at the University of Milan, Italy (since 2000), where he was also Department Chair (2007-2012). He was Associate Professor at Politecnico di Milano, Italy (1992-2000), visiting professor at the University of Texas at Austin, USA (summers 1996-1999), and visiting researcher at George Mason University, USA (summers 2012-2016). He founded a start-up company, Sensure srl, in the area of intelligent systems for industrial applications (leading it from 2007 until 2010) and was active in industrial research projects with several companies. He received his M.S. and Ph.D degree in Computer Engineering from Politecnico di Milano, Italy. His main research and industrial application interests are: intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, internet-of-things, cloud computing, fault tolerance, application-specific digital processing architectures, and arithmetic architectures.

Fabio Scotti received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003. He has been an Associate Professor in Computer Science with the Universitá degli Studi di Milano, Crema, Italy, since 2015. Original results have been published in more than 100 papers in international journals, proceedings of international conferences, books, book chapters, and patents. His current research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, three-dimensional reconstruction, industrial applications, intelligent measurement systems, and high-level system design. He is an Associate Editor with the IEEE Transactions on Human-Machine Systems and Soft Computing (Springer).