

Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection

Xiaolong Hui , Jiang Bian , Xiaoguang Zhao and Min Tan

Abstract

This article presents an autonomous navigation approach based on a transmission tower for unmanned aerial vehicle (UAV) power line inspection. For this complex vision task, a perspective navigation model, which plays an important role in the description and analysis of the flight strategy, is introduced. Based on the proposed navigation model, valuable cues are excavated from a perspective image, which enhances the capability of the perception of three-dimensional direction and simultaneously improves the safety of intelligent inspection. Specifically, for robust and continuous localization of the transmission tower, a developed detecting-tracking visual strategy—comprised tower detection based on a faster region-based convolutional neural network and tower tracking by kernelized correlation filters—is presented. Further, segmentation by fully convolutional networks is applied to the extraction of transmission lines, from which the vanishing point (VP), an important basis for determining the flight heading, can be obtained. For more robust navigation, the designed scheme addresses the scenario of a nonexistent VP. Finally, the proposed navigation approach and constructed UAV platform were evaluated in a practical environment and achieved satisfactory results. To the best of our knowledge, this article marks the first time that a navigation approach based on a transmission tower is proposed and implemented.

Keywords

Unmanned aerial vehicle, intelligent inspection, three-dimensional (3-D) perception, visual navigation

Date received: 30 October 2017; accepted: 6 November 2017

Topic: Special Issue – 3D Vision for Robot Perception

Topic Editor: Antonio Fernandez-Caballero

Associate Editor: Shengyu Chen

Introduction

High-voltage transmission lines are vital to economic development; however, they lack regular maintenance. The inefficiency of existing inspection and maintenance methods leads to a considerable amount of wasted resources. With the rapidly expanded scale of overhead transmission lines, intelligent inspection has received an increasing amount of attention. In recent years, many intelligent inspection studies have been conducted on two major platforms, specifically the unmanned aerial vehicle (UAV) and crawling robot.^{1–3} Compared with crawling robots, UAVs provide greater flexibility, a broader operation range, and higher speed.

Autonomous navigation, an important component of intelligent inspection, has been relying on accurate global positioning system (GPS) or reliable recognition of transmission lines. For GPS-based autonomous navigation systems, ground stations and real-time and accurate GPS data

Institute of Automation Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China

Corresponding author:

Xiaolong Hui, Institute of Automation Chinese Academy of Sciences, University of Chinese Academy of Sciences, No. 95 ZhongGuanCun East Road, HaiDian District, Beijing 100190, People's Republic of China. Email: huixiaolong2015@ia.ac.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

of transmission towers and UAVs are indispensable. Lu et al.⁴ achieved autonomous transmission-line inspection simply using GPS data of transmission towers and UAVs. To maintain protection of UAVs from obstacles, Hrabar et al.⁵ further applied both a stereo camera and light detection and ranging to the navigation system. However, the problem of a collision between a UAV and transmission lines has not been addressed.

Despite the excellent performance of GPS, transmission-line-based navigation studies have also been conducted in many perspectives. Cao et al.⁶ proposed a boundary search radon transform approach for real-time transmission-line detection. Chen et al.⁷ developed the cluster radon transform, an improved radon transform, to extract lines from a satellite image. Cerón et al.⁸ designed a method called circle-based search, which showed better performance on long lines of detection than some traditional methods, such as the Hough transform,⁹ a linear-time line segment detector (LSD),¹⁰ and a real-time line segment detection by edge drawing lines.¹¹ Meanwhile, Araar and Aouf¹² applied two different kinds of control strategies to the inspecting UAV and compared their performances. Zhou et al.¹³ proposed a model that solves the problem of automatic selection of the threshold during line tracking. Zhang et al.¹⁴ used a Kalman filter to track the power lines in the Hough space based on the continuity of a video sequence. Compared with GPS-based navigation, navigation based on transmission lines has many limitations and challenges. For example, extraction of narrow transmission lines from a complex background remains a challenge. Some artificial backgrounds cause adverse effects in recognition work. Nevertheless, in terms of existing methods, UAVs must inspect above the transmission lines, whereby overhanging lines are visually parallel, which simultaneously improves the robustness of recognition and navigation. Currently, autonomous inspection still faces severe security risks from the probability of a UAV crash. Thus, existing inspection approaches have not been applied to inspection of high-voltage transmission lines.

The transmission tower is another important visual cue in the electric corridor and has attracted considerable research attention. Steiger et al.¹⁵ adopted the implicit shape model¹⁶ approach for representation of transmission towers. In the detection phase, local maxima in the voting space indicate the likely locations of transmission towers. Sampedro et al.¹⁷ employed histogram of oriented gradient¹⁸ features and multilayer perception neural networks to distinguish transmission towers from backgrounds. By employing a sliding window algorithm, localization of towers is achieved. Moreover, Cerón et al.¹⁹ developed a new descriptor based on a grid of two-dimensional feature descriptors, and they used the support vector machine as a classifier. Their detection required less than 50 ms, which was faster than previous methods. However, the accuracy of existing transmission-tower detection continues to have

difficulty meeting the requirements of autonomous navigation. Since 2012, deep learning technology has shown impressive performance in various vision tasks. Some typical approaches, such as a faster region-based convolutional neural network (Faster-RCNN)²⁰ and a single shot multibox detector (SSD),²¹ outperformed other traditional methods and achieved excellent detection performance. Even though deep learning achieves breakthroughs in many fields, their performance in the transmission-tower detection still deserves an expectation, and to the best of our knowledge, an autonomous navigation approach based on the transmission tower has not yet been proposed.

To achieve transmission-tower-based autonomous navigation, UAVs must be capable of directional perception. According to the projection theory, spatial parallel lines intersect at infinity, which is imaged as a vanishing point (VP). By employing the VP, the direction of a spatially straight line can be uniquely determined. VP detection is widely used in road detection^{22–24} and robotics navigation.^{25–27} It can be generally categorized in two groups: edge-based and texture-based. Taking road detection as an example, edge-based VP detection relies on high-contrast edges, which are mainly from clear road boundaries and lane marks. In contrast, texture-based methods with relative high computational complexity show superiority in unstructured off-roads environments. In addition, some road segmentation methods based on deep learning have made outstanding progress.²⁸ Many new approaches have been applied to the practical automatic driving system with the aforementioned methods as a reference. Nonetheless, practical VP-based navigation methods, which are suitable for UAV autonomous inspection, need to be developed.

In this article, we focus on the problem of autonomous navigation based on the transmission tower, which enables power infrastructure inspection. To improve the navigation scheme description, a navigation model that leverages the perspective relation is introduced first. The scheme is fundamentally comprised of two components: reliable transmission-tower localization and robust transmission-line extraction. To achieve reliable real-time localization of the transmission tower, a detection-tracking visual strategy is employed. It combines faster R-CNN detection and the kernelized correlation filter (KCF) tracking algorithm. For the robust extraction of transmission lines in complex backgrounds, segmentation that applies deep learning is employed, followed by a proposed annotation method. The detected transmission lines are further processed by edge-based VP detection, which consists of LSD, random sample consensus (RANSAC),²⁹ and linear and nonlinear optimization, to obtain a precise flight heading defined by VP. Moreover, an experiential flight heading adjusted by the proposed navigation model is also adopted when a VP does not exist, which improves the flexibility of the scheme. To verify the effectiveness of the proposed navigation

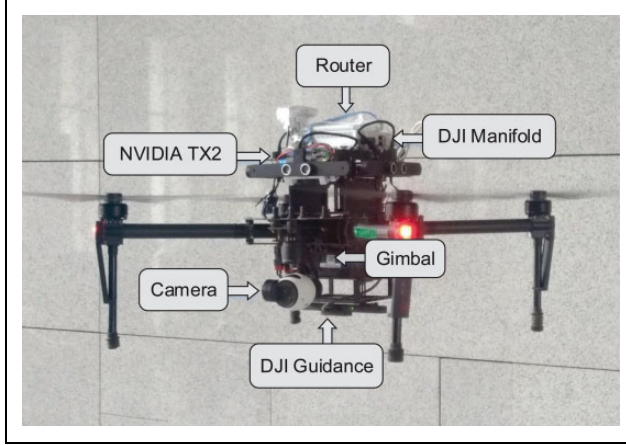


Figure 1. DJI Matrice 100 UAV inspection platform: TX2, DJI guidance, DJI Manifold, DJI Zenmuse X3 gimbal and monocular camera. UAV: unmanned aerial vehicle.

scheme, we built a UAV inspection platform and obtained satisfactory results in an evaluation. Compared with existing inspection systems and navigation approaches, the proposed transmission-tower-based navigation approach has key attributes and advantages. For one, the UAV is not required to navigate above the transmission lines, which reduces the risk of an inspection accident. Moreover, the UAV has a beyond visual range flight capability and requires no GPS or ground station. Furthermore, the presented approach is effective in relatively complex backgrounds. To the best of our knowledge, the presented approach marks the first time that UAV achieves transmission-tower-based autonomous navigation.

The remainder of the article is structured as follows. The whole system design of the inspection UAV is described in “hardware platform and system architecture” section. Details of our autonomous navigation model, detection-tracking visual strategy, VP detection that applies fully convolutional networks (FCNs), and the concrete navigation process are described in “proposed navigation scheme” section. The experimental results and analysis are described in “experiment and analysis” section 4. Finally, the conclusions are summarized in “conclusion” section.

Hardware platform and system architecture

The presented system employs a refitted DJI Matrice 100 (M100) quadrotor as the inspection platform, as shown in Figure 1. It is equipped with an advanced DJI Lightbridge high-definition image transmission system; a DJI guidance visual positioning system, which operates in GPS-deficient environments; and a DJI Zenmuse X3 (X3 pan-tilt-zoom (PTZ)) gimbal and monocular camera for tracking and detection.

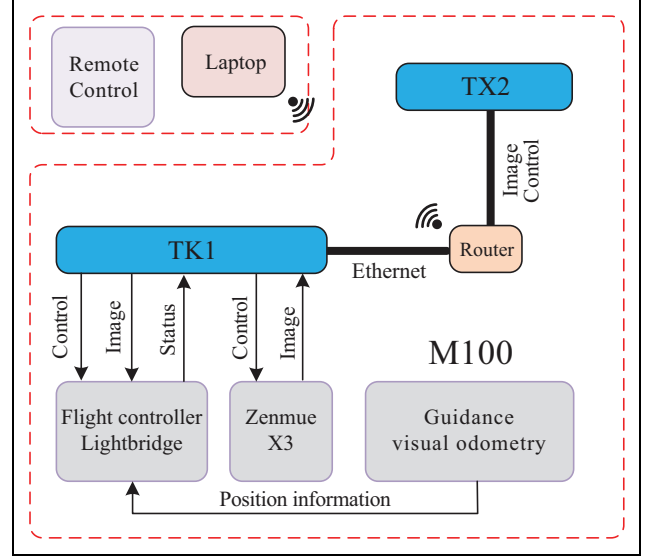


Figure 2. System architecture of M100.

Table 1. Performance of the inspection UAV.

Items	Characteristics
Maximum takeoff weight	3600 g
Hovering accuracy	Vertical: 0.5 m, Horizontal: 2.5 m
Hovering Ttime	1 kg payload: 13 min
Transmission distance	CE: 3.5 km
Maximum angular velocity	Yaw: 150°/s
Maximum speed	17 m/s (no payload, no wind)
Maximum speed of ascent	4 m/s (no payload, no wind)
Maximum speed of descent	5 m/s (no payload, no wind)
Maximum wind resistance	10 m/s

UAV: unmanned aerial vehicle.

In addition, two embedded processors, DJI Manifold (NVIDIA TK1) and NVIDIA TX2, are situated on the M100. TX2 has a high-performance graphics processing unit and is thus the main processor responsible for detection-tracking visual strategy, VP detection, and control strategies. DJI Manifold is simply used to capture sensor data. The robot operating system (ROS) is run on both processors, and an ROS network is constructed by a router to facilitate information exchange, as shown in Figure 2. Specifically, we run the DJI onboard SDK on Manifold, which builds a bridge between M100 and the upper level TX2. It communicates with flight control N1 and returns important data to TX2 via ROS topics, while simultaneously transmitting control commands from TX2 to N1 and X3 gimbal. Our algorithms, which are detailed in the next section, are implemented in TX2. In particular, we compress raw images with the resolution of 1280×720 for stable and real-time image transport from TK1 to TX2. In addition, we employ a laptop for the ROS system to initialize all nodes. Additional parameters and related performances of the UAV platform are listed in Table 1.

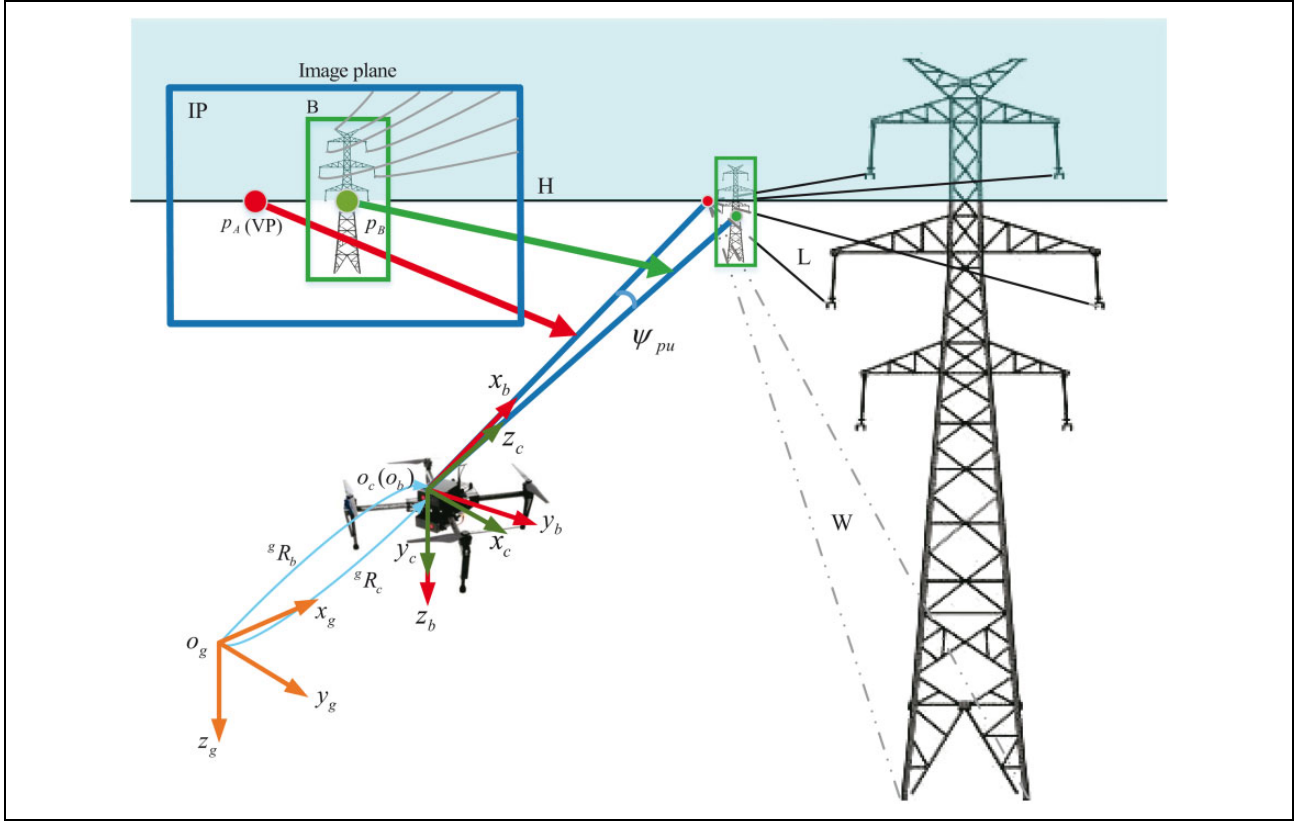


Figure 3. Perspective navigation model. $C_g = o_g x_g y_g z_g$ (orange), $C_b = o_b x_b y_b z_b$ (red) and $C_c = o_c x_c y_c z_c$ (green) correspond to inertial frame, body-fixed frame and camera frame, respectively. p_A and p_B denote the projection of flight heading and center of transmission towers, respectively. ψ_{pu} is the angle between flight heading x_b and the optical axis z_c . Other symbols that contribute to a better understanding of perspective relationship include IP = image plane, B = bounding box, H = the horizon line, L = overhead transmission lines, W = roadway (black dotted line). In the diagram, we assume that the transmission lines are parallel to the ground in order to more clearly express our navigation strategy, and p_A and VP coincide.

Proposed navigation scheme

To have a better description of proposed scheme for transmission-tower-based continuous flight, we first to introduce a perspective navigation model. Thereafter, two key issues comprising real-time reliable transmission-tower localization and robust VP computation are addressed. Finally, we detail the whole inspection process between adjacent towers.

Perspective navigation model

According to previous autonomous navigation models, the UAV has to fly above the transmission lines so that the parallel property of transmission lines can be preserved via the downward facing camera. Navigation depends only on the detected transmission lines. Nevertheless, the performance of these line-based methods is degraded by unreliable line extraction.

By contrast, our model, illustrated in Figure 3, is constructed in three-dimensional (3-D) space and takes the perspective relationship into account. The algorithm employs perspective images as input, and outputs the UAV

velocity control amount in 3-D motion space. To better elucidate this concept, we first to introduce three coordinate frames: global inertial frame $C_g = o_g x_g y_g z_g$, body-fixed frame $C_b = o_b x_b y_b z_b$ and camera frame $C_c = o_c x_c y_c z_c$, all of which follow the right-hand rule. To simplify the problem, we assume that coordinate origins o_b and o_c coincide. Next, let gR_c denotes the rotation matrix of the camera frame with respect to the inertial frame. The body frame with respect to the inertial frame is denoted by gR_b . The attitude rotation matrix of the body coordinate frame with respect to the camera coordinate frame is calculated as follows

$${}^cR_b = ({}^gR_c)^{-1} {}^gR_b \quad (1)$$

With respect to the camera coordinate frame C_c , we define two unit 3-vectors β_1, β_2 separately corresponding to the VP and UAV heading, which are deduced from the following equations

$$\beta_1 = K^{-1} \tilde{v} / \|K^{-1} \tilde{v}\| \quad (2)$$

$$\beta_2 = {}^cR_b x_b \quad (3)$$

where \tilde{v} is the homogeneous coordinate of VP v , x_b is a unit 3-vector parallel to the x -axis in the body frame, and K is the intrinsic matrix.

For a further description of the relationship between β_1 and β_2 , the rotation axis $l = (l_x, l_y, l_z)$ and the rotation angle γ are introduced. They are derived from $l = \beta_1 \times \beta_2$ and $\gamma = \arcsin \|l\|$. The rotation matrix ${}^{\beta_1}R_{\beta_2}$ between both vectors is constructed according to Rodriguez formula

$${}^{\beta_1}R_{\beta_2} = \cos\gamma E + (1 - \cos\gamma)ll^T + \sin\gamma l^\Lambda \quad (4)$$

where E is an identity matrix, l^Λ is a (3×3) skew-symmetric matrix associated with l . By the decomposition of the rotation matrix ${}^{\beta_1}R_{\beta_2}$, UAV can get the amount of rotation to the direction defined by VP.

Actually, on the one hand, the relative change between spatial 3-vectors can be reflected on the image plane. Such as, on the image plane, we calculate the projection of the UAV flight heading by

$$\lambda \tilde{p}_A = K^c R_b x_b \quad (5)$$

where λ is a scalar, \tilde{p}_A is the homogeneous coordinate of the projection point p_A . If p_A coincides with VP, UAV will fly along the spatial direction defined by VP. On the other hand, the spatial relative position between UAV and a certain object can be predicted by the plane position relationship between p_A and the detected object boundary. Such as, the relationship that p_A fall within the detected tower bounding box pointed by character B indicates that the UAV will fly towards the tower and collide with the tower in the end. To avoid this situation, the UAV should adjust heading to make its projection p_A fall outside the detected bounding box. By further analysis, it is understood that more information can be excavated from the perspective navigation model. Ideally, when the projection p_A is consistent with the VP of the transmission lines, the UAV will fly along the transmission lines. Furthermore, when PTZ camera is parallel to the ground and is at the same height as the tower center, the horizon line will be observed in the middle of the image plane, and the projection of tower center p_B is exactly located in the horizon line. Since the transmission lines are located above the center of the tower, they can be clearly observed in the upper right area of image, having a background of blue sky.

Based on the above analysis, we draw some important conclusions, which are guidelines for our design of navigation scheme and listed as follows: (1) Just by the reliable localization of transmission tower, UAV is capable of autonomous flight along one side of transmission lines between adjacent towers. (2) When UAV flies along the direction defined by VP, the flight route is parallel to transmission lines. (3) Benefiting from this distinct flight route, transmission-line extraction becomes much more robust and observation is much clearer. The detailed navigation scheme will be given in the last subsection after resolving

the real-time reliable localization of transmission tower and precise estimate for VP.

Detection-tracking visual strategy

In our navigation scheme, transmission tower, the only landmark in power corridor, must be timely and reliably localized. We select an optimal scheme to resolve this problem, comprehensively considering the limited computational capability of onboard embedded processor, the speed and accuracy of algorithms. Specifically, we picked out the most suitable deep-learning-based object detection algorithm, namely, faster R-CNN.²⁰ It is the only algorithm that achieves no false detection in our transmission-tower test data set among all tested algorithms. Faster R-CNN presents a region proposal network (RPN) that shares convolutional features with the detection network, thus enabling nearly cost-free region proposals. Shareable features within sliding windows are mapped to a lower-dimensional vector for further object classification and bounding box regression. Nevertheless, for faster R-CNN framework, the currently fastest ZF network³⁰ is just less than 2 fps on our TX2 platform.

To make up the shortcomings of low speed, we combine real-time tracking with the robust detection. For the tracking, we select the excellent KCFs.³¹ KCF obtains training samples by the matrix cycling of images and takes advantage of the Fourier transform to avoid complex matrix inversion operations. These properties of KCF lead to high tracking accuracy and low computational costs. Actually, real-time tracking is more robust in relatively short period due to considering successive property of motion and simultaneously contributes to smooth observation.

When result of tracking exists a significant drift away from target, the tracker is re-initialized from the recent detection result. Further, to address the problem of matching between detection and tracking of different running frequencies, tracking results are recorded in a queue with a first-in first-out property, as shown in Figure 4. The time stamp is attached to the current input frame. When the next frame arrives, the current frame with the distinguishing time stamp and corresponding tracking result is pushed into the queue. The detection result of the current frame will be obtained after several frames due to a delay of computation. According to the time stamp, detection can find the matched tracking in the queue.

In addition, in the section of experiment, we will make a detailed comparison and analysis of all tested detection algorithms, and give parameters setting of training faster R-CNN.

VP detection for precise flight heading

The VP of transmission lines is an important visual cue that provides a precise flight heading during autonomous navigation. Although the traditional approach for VP detection

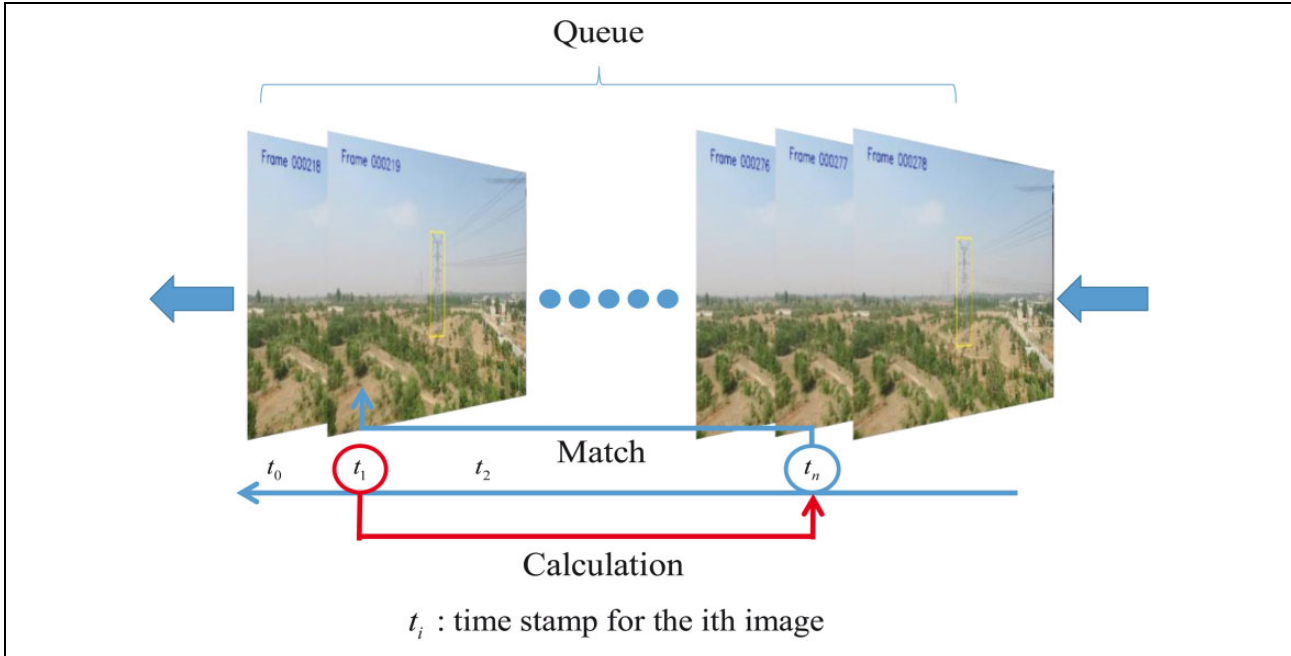


Figure 4. Detection-tracking visual strategy.

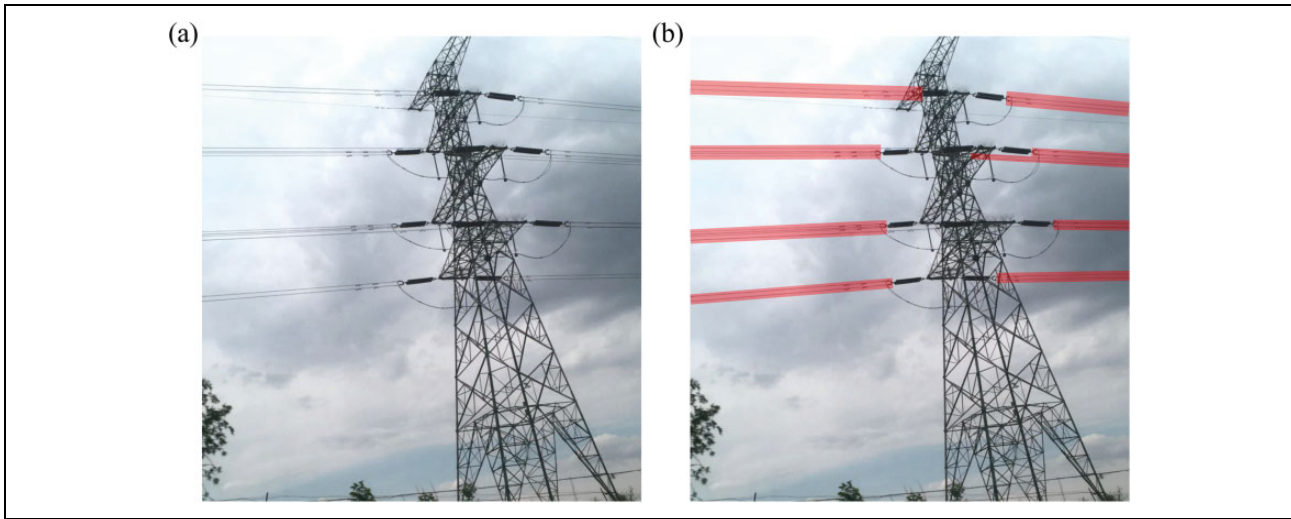


Figure 5. (a) Original picture. (b) Annotated picture (Red polygons correspond to the ground truth).

has achieved a satisfactory navigation effect, excessive usage of prior knowledge reduces the intelligence and application prospects of the system.

Therefore, inspired by CNN-based methods for semantic segmentation, including crowd segmentation,³² road segmentation,²⁸ and retinal vessel segmentation,³³ we adopt the excellent FCNs, which apply a deep learning pipeline and can be trained from end to end. Specifically, we apply this technology before the traditional VP detection method. This simultaneously avoids wasting visual resources caused by a prior-knowledge-based filter and relieves some difficulties involving narrow line extraction from complex backgrounds.

Currently, various FCNs in combination with residual networks³⁴ have reached state-of-the-art level. Considering both the heavy computational costs and immense manual labeling work—prompted by precise pixel-level segmentation—we decided to apply segmentation on the region level.

In view of the special distribution of transmission lines, we employ a simple but effective means of manual annotation. As shown in Figure 5, transmission lines are divided into several parallel line groups, each of which corresponds to a phase voltage. Transmission lines in the same group are covered with red polygons, which are regarded as ground truth labels for training and evaluating FCNs. With respect to the loss

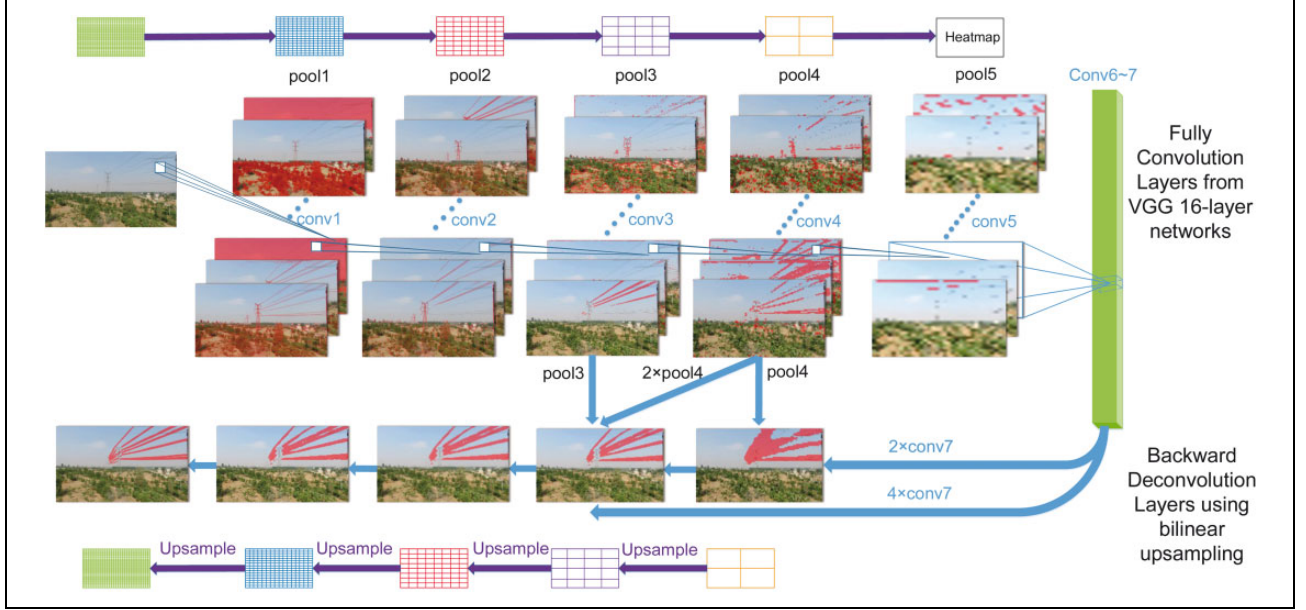


Figure 6. Structure of FCNs and feature maps for transmission-line segmentation. FCNs: fully convolutional networks.

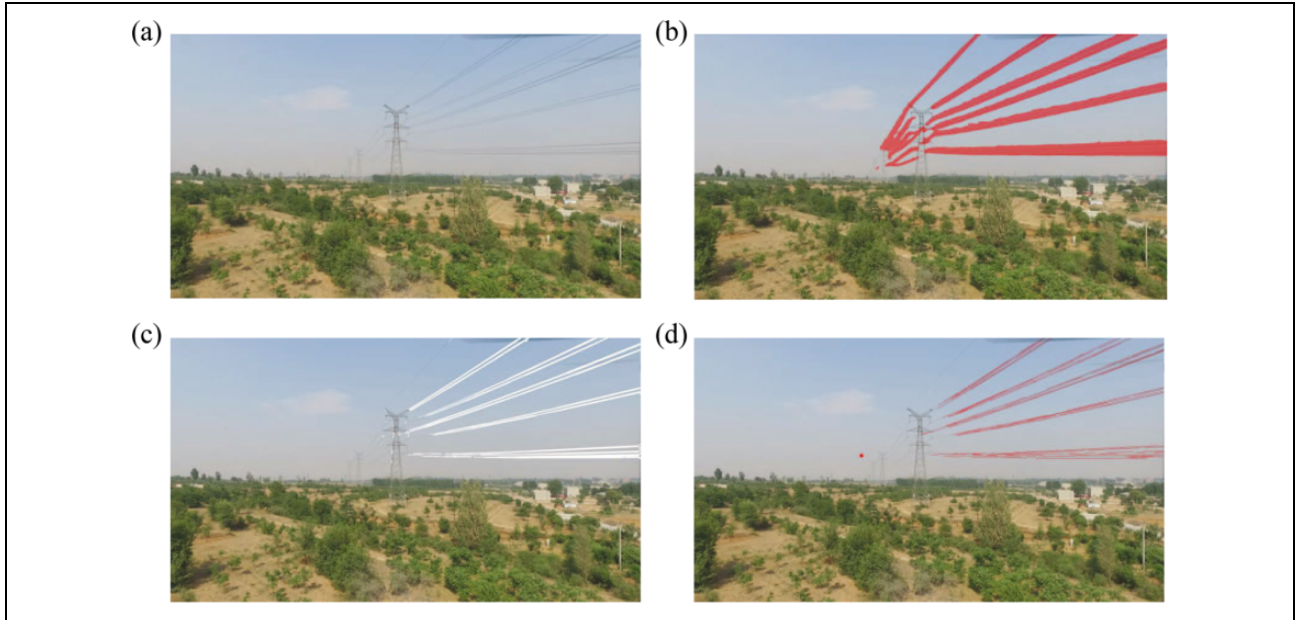


Figure 7. Process of proposed VP detection. VP: vanishing point.

function, we employ an effective cross-entropy approach, which is defined as equation (6) for a binary classification

$$\text{loss}(a, b) = \frac{-1}{N_o} \sum_{k=1}^{N_o} \left(a_k \log b_k + (1 - a_k) \log(1 - b_k) \right) \quad (6)$$

where N_o is the number of output neurons, a_k is the ground truth, and b_k represents prediction.

As shown in Figure 6, FCNs are composed of convolution and deconvolution networks.³⁵ Regarding the feedforward convolution part, VGG 16-layer networks³⁶ are

employed for multidimensional feature representation. Their fully connected layers are converted to convolutions for the production of a heat map. From that point, to recover the original input size, the heat map, whose size has been reduced to $\frac{1}{32^2}$ of input image, is bilinearly upsampled by the backward deconvolution layers. Meanwhile, the up-sampled coarse intermediate results are fused with summing scores of pooling layers for rectification to realize precise segmentation.

To combine with the traditional method for VP detection, as shown in Figure 7(b) [(a) is the raw image], within

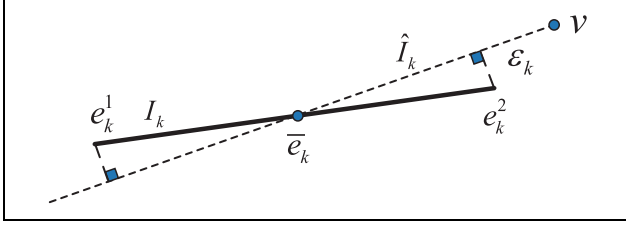


Figure 8. Illustration of the consistency measure.

the red transmission-line regions generated by FCNs, LSD¹⁰ is carried out, whose results correspond to the white regions displayed in Figure 7(c). The linear-time LSD, with no parameters tuning required, is regarded as an effective preprocessor, which algorithm is capable of adaption to a certain degree of environmental change. As a consequence, we apply LSD before the Hough transformation, which arrangement significantly improves quality of Hough line-detection. Finally, as displayed in Figure 7(d), the red straight line segments, available for computation of VP, are exactly the result of Hough line-detection.

Before estimation of VP, the robust algorithm RANSAC²⁹ is deployed. RANSAC returns the consensus set, which has the largest number of inliers. Here, an inlier is a straight line segment whose distance to the VP is lower than a specific threshold. For further optimization of VP, a typical consistency measure method is first to be introduced. As shown in Figure 8, e_k^1 and e_k^2 are the two endpoints of the k -th detected line segment I_k , whose midpoint is denoted by \bar{e}_k . The notation \hat{I}_k is considered an ideal straight line which corresponds to I_k and passes through the estimated VP denoted by v and the centroid \bar{e}_k . The difference between the detected I_k and the ideal \hat{I}_k is measured by the ε_k , which is the perpendicular distance from e_k^1 to \hat{I}_k denoted by

$$\varepsilon_k = \text{dis}(e_k^1, \hat{I}_k) = \text{dis}(e_k^1, \bar{e}_k \times v) \quad (7)$$

where $\text{dis}()$ is a distance function from a point to a line and \times denotes the cross product of two vectors.

Following the RANSAC result, the VP position is further refined by an iterative nonlinear optimization method. It is worth mentioning that there exist two common refinement models involving least mean square (LMS) model and maximum likelihood estimation (MLE) model.³⁷ Their objective functions are displayed as follows

$$v^* = \arg \min_v \sum_{k=1}^N \varepsilon_k^2(v) \quad (8)$$

$$v^* = \arg \max_v \sum_{k=1}^M \log P(I_k|v) \quad (9)$$

where N is the number of inliers and M denotes the number of all the detected lines. Specifically, $P(I_k|v)$ is a prior model that should be learned in advance for different databases and different consistency measure methods.

We finally select the LMS method based on the following consideration. LMS, using only the inlier lines, is faster to compute and does not need the prior probability model from the training set, even though MLE, using both inlier and outlier lines, achieves a slightly higher accuracy. For the initial estimation of the LMS objective function, linear equations $\Delta v = \eta$ deduced from the results of RANSAC are introduced. Each row of it corresponds to a straight line equation through VP, which is built by the detected straight segment. According to the Hough result, this $n \times 2$ set of equations has the following form

$$\begin{pmatrix} \alpha_1^0 & \alpha_1^1 \\ \vdots & \vdots \\ \alpha_n^0 & \alpha_n^1 \end{pmatrix}_{n \times 2} \begin{pmatrix} v_x \\ v_y \end{pmatrix}_{2 \times 1} = - \begin{pmatrix} \alpha_1^2 \\ \vdots \\ \alpha_n^2 \end{pmatrix}_{n \times 1} \quad (10)$$

where $v = (v_x, v_y)^T$ is the solution to the set of linear equations, $(\alpha_i^0, \alpha_i^1, \alpha_i^2)$ is a triplet of parameters corresponding to the i -th straight line. The linear least-squares solution to the above set of inhomogeneous equations is $\Delta^+ \eta$, where Δ^+ is the pseudo-inverse of Δ . Finally, by Levenberg–Marquardt³⁸ iteration, the precise VP can be obtained.

The navigation process

Based on the perspective navigation model, the detailed navigation scheme is illustrated in Figure 9. We decompose a long-distance inspection mission into several short sub-tasks between adjacent transmission towers. The short sub-task is further divided into two stages: initialization and inspection.

At initialization stage, by detection, the UAV determines first the position of the next tower. Thereafter, the PTZ adjusts the pitch and yaw angle to make the camera optical axis point to the center of the detected transmission tower. At the stage of UAV heading adjustment, p_A is constantly calculated by equation (5) until it reaches the same horizontal position as VP. Since PTZ actively tracks tower, VP position has almost no change during the adjustment of flight heading. In a practical application, we just calculate VP once and apply FCNs to the system successfully. In particular, we suggest an experiential flight heading to cope with the case that VP is nonexistent. The horizontal distance from the projection of experiential heading to the center of tower p_B is 1.2 times the width of the bounding box. Following the experiential direction, UAV may fly along a z-type trajectory in a power corridor.

At the inspection stage, the UAV flies along the expected heading. With the detection-tracking visual strategy enabled, the PTZ constantly adjusts the yaw and pitch to track the next tower until the relative yaw angle ψ_{pu} between camera optical axis and the expected heading reaches 90° . During the flight, once the tracking fails, UAV will immediately brake. After that, the tracker will be initialized again using the recent detection result.

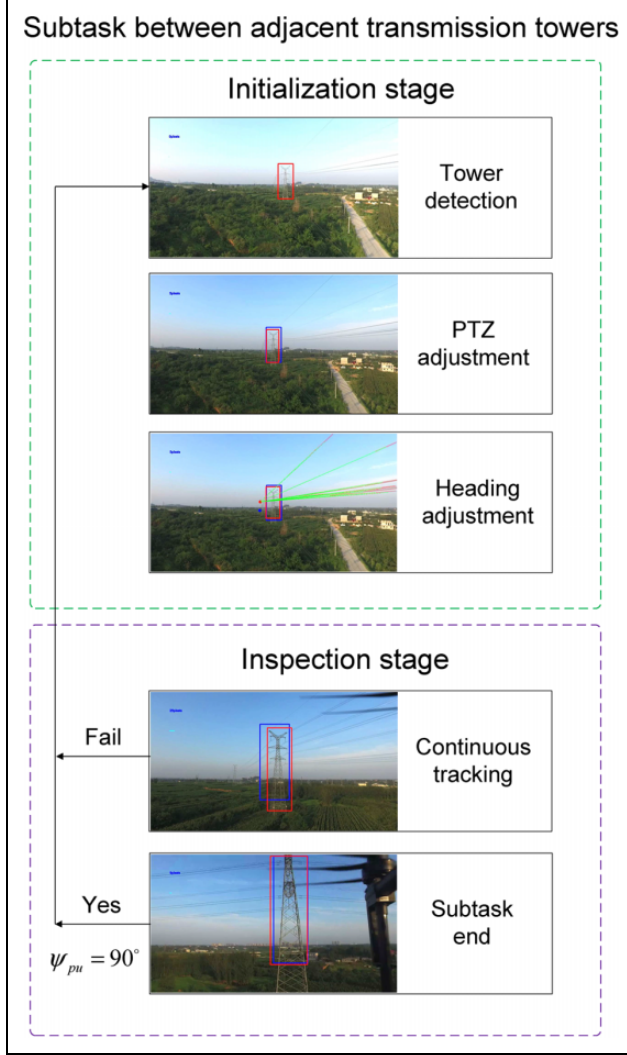


Figure 9. Flowchart of autonomous navigation between adjacent towers. Initialization stage and inspection stage are distinguished with green and purple dashed bounding boxes. The red bounding box represents the result of detection. The blue bounding box denotes the result of tracking. At the stage of heading adjustment, the blue circle denotes the projection of current heading, the red circle denotes the estimated VP, the green lines represent the estimated straight segments passing through VP.

Experiment and analysis

To verify the effectiveness of the proposed approach, three experiments were conducted. First, by comparison among state-of-the-art object detection algorithms, faster R-CNN (ZF) was demonstrated to be most appropriate for reliable localization of transmission tower. The remarkable improvement of both robustness and precision of VP detection by the proposed method was demonstrated by the second experiment. Finally, the whole system was tested in a practical inspection environment followed by a detailed discussion.

Table 2. Speed and average precision.^a

Method	AP, %	FPS
Faster R-CNN (ZF)	89.5	2
Faster R-CNN (VGG16)	89.6	0.8
SSD300	88.9	7
SSD512	89.2	2
YOLOv2	86.8	5.6

AP: average precision; FPS: frame per second; faster R-CNN: faster region-based convolutional neural network.

^aFor faster R-CNN, (ZF) and (VGG16) represent two different networks of feature extraction. SSD300 and SSD512 separately correspond to 300×300 and 512×512 input resolutions. YOLOv2 corresponds to the 544×544 input resolution. AP and FPS are is measured on NVIDIA TX2.

Tower detection experiment

For this experiment, we collected 1280 sheets of transmission-tower pictures from aerial photos, 1000 of which are used as training, and the rest of which are used as test. To find an appropriate detection method for reliable localization of transmission tower, we make a comparison among three state-of-the-art object detection frameworks comprising faster R-CNN,²⁰ SSD,²¹ and YOLOv2.³⁹ Comparison was conducted from the following three aspects: speed, average precision (AP), and the ratio of false detection (precision-recall curve). Following the detection evaluation scheme of the PASCAL visual object classes challenge,⁴⁰ precision and recall are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

where true positive (TP), false positive (FP), and false negative (FN) separately correspond to true positive (tower is correctly detected), false positive (background is mistaken as tower), and false negative (tower is not detected). AP is approximate as the area under the precision-recall curve.

As shown in Table 2, at the aspect of speed, SSD300 outperforms other methods and simultaneously achieves a relatively high AP. For the AP, faster R-CNN (VGG16) has a slight advantage over other methods. With respect to the ratio of false detection, as illustrated in Figure 10, overall, faster R-CNN (ZF), denoted by the green solid line, maintains a 100% precision, namely, no false detection, over a fairly wide range of recall, which clearly surpasses SSD and YOLO. At this point, both SSD and YOLO encounter different degrees of false detection, even at a low level of recall. In addition, all the curves break off when the recall is close to one. The reason is that, for all the tested algorithms, there exist some bounding boxes that have not been detected. In fact, detection-tracking visual strategy depends heavily on the detection accuracy. During the autonomous navigation, we can accept a certain degree of missed detection (FN), whereas cannot tolerate the false detection (FP),

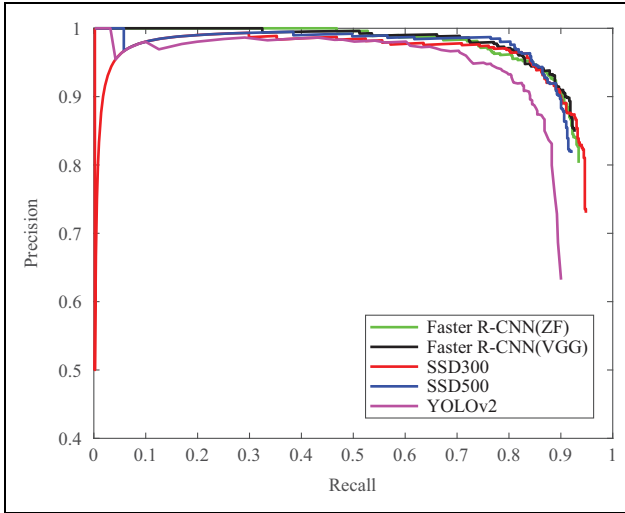


Figure 10. Precision-recall curve for detection.

which brings a significant threat to navigation safety. Thus, we decide to apply faster R-CNN (ZF) to this visual strategy. For the best test result, following the four-step training algorithm proposed by faster R-CNN, we selected 80,000, 40,000, 80,000, 40,000 iterations, respectively, corresponding to each step. ZF network is pretrained by ImageNet.⁴¹ Other parameters take the default settings.

For a more detailed explanation of the detection-tracking strategy, we selected a short typical excerpt from all the recorded videos, which involves the particular case of the tracking failure and initialization by detection again. In Figure 11, with the changes of visual angle shown in Figure 11(a)–(c), the projection of sun moves toward left relative to the target tower, in which case the tracking fails and the blue tracking bounding box drifts gradually away from the target tower. As shown in Figure 11(d), when the pixel disparity between both bounding box centers of tracking and detection are beyond 120 pixels, the tracking is

considered a failure. From that point, UAV begins to slow down until the velocity is zero. At the moment corresponding to the frame 617, as shown in Figure 11(h) by the red detection bounding box, the tracking is initialized again.

VP detection experiment

In this section, we show first the necessity of applying FCNs in transmission-line VP detection. A comparative experiment was performed to verify that the proposed method improved the detection precision. Unlike in tower detection, we did not draw a precision-recall curve because drawing this curve requires a pixel-level label instead of a region-level label. For VP detection of transmission lines, segmentation is simply a subtask; thus, the pixel-level label is believed to be difficult to use and unnecessary.

To train FCNs, 600 pictures were collected from aerial videos, 400 sheets of which were used for training and validation. The remainder was used as test items. The pictures of the training set were obtained from many different scenes covering rural areas, mountain areas, and suburbs, whose backgrounds included farmland, highways, blue skies, and high mountains. Additionally, changes of illumination, viewpoint, and distance were also taken into consideration. The test results are shown in Figure 12. The first four images were obtained from the PTZ camera with the same resolution of 1920×1080 . The remaining four sheets were selected from the processed results at the stage of VP detection with a resolution of 1280×720 and having JPEG compression. Figure 12(a) shows that transmission lines are partly occluded by the UAV landing gear. In the case of the bright light shown in Figure 12(b), the decreased contrast makes the transmission lines difficult to distinguish. However, the FCNs are still effective and achieves excellent performance. In Figure 12(c), the transmission lines are nearly submerged in farmland. However, the distant highway, which has distinct edges, and the nearby farmland,

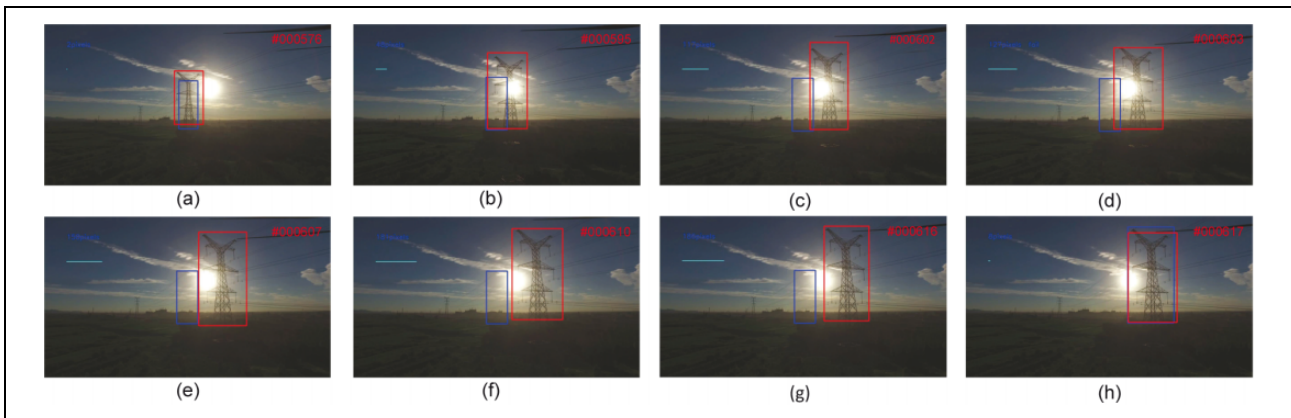


Figure 11. Example of detection-tracking visual strategy. The red bounding box represents the result of detection. The blue bounding box denotes the result of tracking. The red character string beginning with character # and displayed in right upper region indicates the frame number. The cyan straight line segment as shown in the upper-left region denotes the horizontal pixel errors between the centers of tracking and detection.

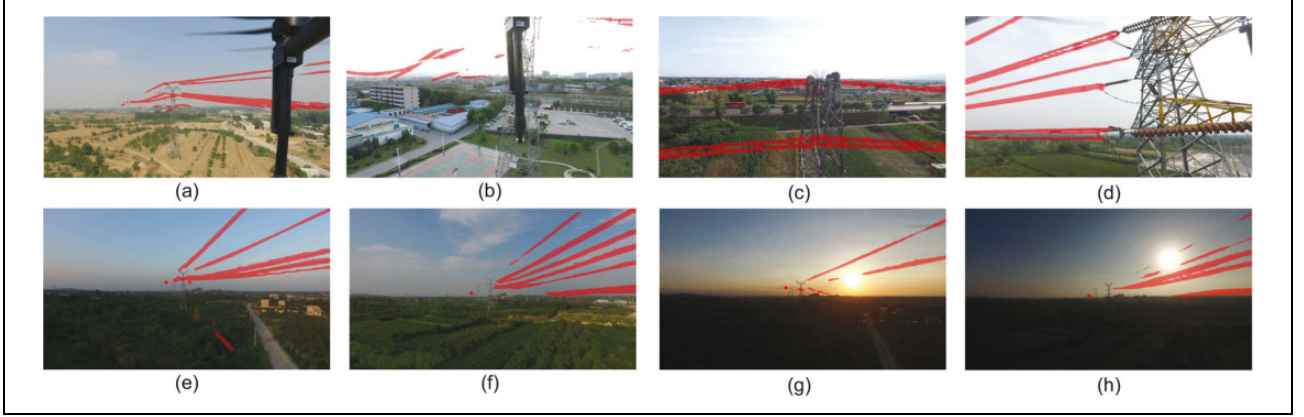


Figure 12. Extraction of transmission lines applying FCNs. The red circles in (e)–(h) are the detected VP. FCNs: fully convolutional networks; VP: vanishing point.

which has clear textural features, do not cause interference to the correct extraction work. Figure 12(d) shows that, despite the change of the observation distance and angle, the correct extraction can still be guaranteed. Constrained by the router bandwidth, the real processing images must be compressed with JPEG format to ensure a reliable transmission, which means poor image quality. Nevertheless, the outstanding FCNs have not been impacted by this negative factor. We tested the whole system at dusk and early morning, and selected four typical images processed at the stage of VP detection, which separately corresponded to Figure 12(e)–(h). In the course of practical inspection, the detected transmission lines are mostly with the background of blue sky. Nonetheless, the traditional VP detection may often fail on account of adverse weather and poor illumination. As shown in Figure 12(e) and (f), which suffer from cloudy weather, the VP detection combined with the FCNs properly functions. In the condition of dazzling illumination at dawn, as shown in Figure 12(g) and (h), the FCNs successfully extract a few useful straight line segments, which are sufficient for VP detection.

Many VP detection methods serving autonomous driving are effective because pavements, which own clear edges and more pixels, are easier to distinguish from background. In contrast, transmission lines have fewer pixels and are difficult to identify; thus, extraction is more readily impeded. We think that the precision of VP detection highly depends on the number of transmission lines detected. Different methods achieve similar accuracy on the same detection result.

As a consequence, we simply compared the edge-based method with its improvement that is joined with FCNs. As a reasonable contrast, we used the upper-right region of the overall image as input of the conventional method. This approach notably incurred the loss of some valuable transmission lines. Regarding the proposed method, the processing was carried out on the overall image. For the quantitative evaluation methodology, we selected the accumulative pixel error between the real detection and ground truth. This kind

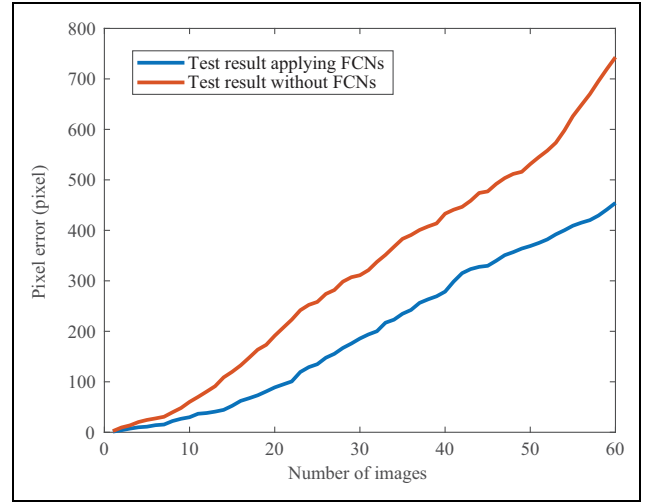


Figure 13. Accumulative pixel error between the estimated VP and the ground truth. VP: vanishing point.

of assessment was believed to be meaningful and sufficient to verify the effectiveness of the proposed method for the goal of transmission-line VP detection.

With respect to the test data set, a sequence of images with the resolution of 1280×720 was selected. These images were obtained during an autonomous flight between adjacent transmission towers. The curve of accumulative pixel errors is shown as Figure 13. It is shown that the accumulation errors relating to the proposed method are 450 pixels. This result is superior to that of the traditional method with an error of 750 pixels. The average of seven pixel errors is negligible for the 1280×720 resolution image, which results from the uncertainty of the annotation due to the intrinsic pendulous traits of transmission lines. In addition, a steep ascent occurs at the end of the curve for the traditional method owing to the change of illumination. In contrast, the curve for the proposed method is relatively smooth, which indirectly proves the capability of the proposed method to resist environmental interference.

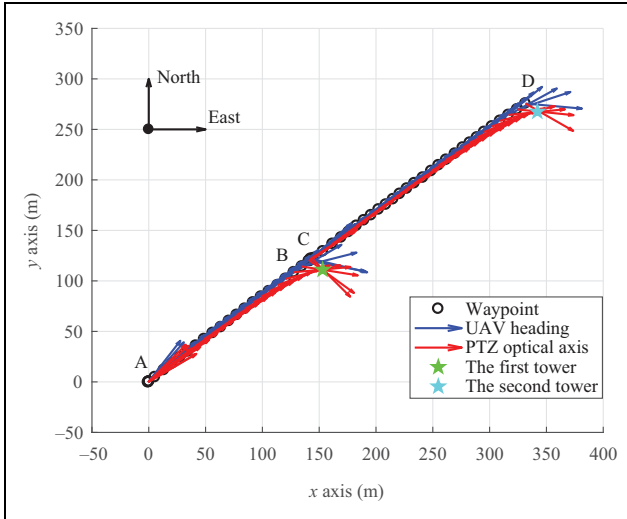


Figure 14. Planar view of a trajectory with direction annotation.

Comprehensive experiment

In addition to the performance analysis of each system part, we performed comprehensive experiments to demonstrate the feasibility of the entire system. The test distance was approximately 400 m, which covered two adjacent intervals. Although the method could achieve more distant autonomous navigation, we relinquished the more distant test with consideration of the safety of power infrastructures and UAV. Nonetheless, the test with a limited distance was adequate to demonstrate the effectiveness of our system.

UAV waypoint and PTZ orientation were synchronously recorded with respect to inertial coordinate frame from the initialization stage. The waypoint, which is the fusion result of visual positioning system (DJI guidance) and GPS, was measured relative to the takeoff position. Further, the waypoints were sampled with an interval of fifty to clearly describe the trajectory, as shown in Figure 14. The average flight duration between adjacent waypoints is approximately 4 s. We select four typical positions shown in the figure for a detailed explanation. At the position A, PTZ, whose camera optical axis is denoted by the red arrow, adjusts the yaw and pitch angle to situate the detected transmission tower in the image center. Then, UAV rotates to the expected heading defined by VP. In order to avoid the occlusion to the camera caused by landing gear, from waypoint B, UAV continues to fly along the expected direction and simultaneously begins to rotate. At the position C, the relative yaw angle ψ_{pu} reaches 90° , the inspection between adjacent towers ends. After that, the next inspection starts with waypoint C and ends with waypoint D.

Furthermore, at different initial positions, we conducted multiple tests. As shown in Figure 15, a total of nine trajectories were painted, each of which was similar to the trajectory, as shown in Figure 14. Figure 15 is a planar

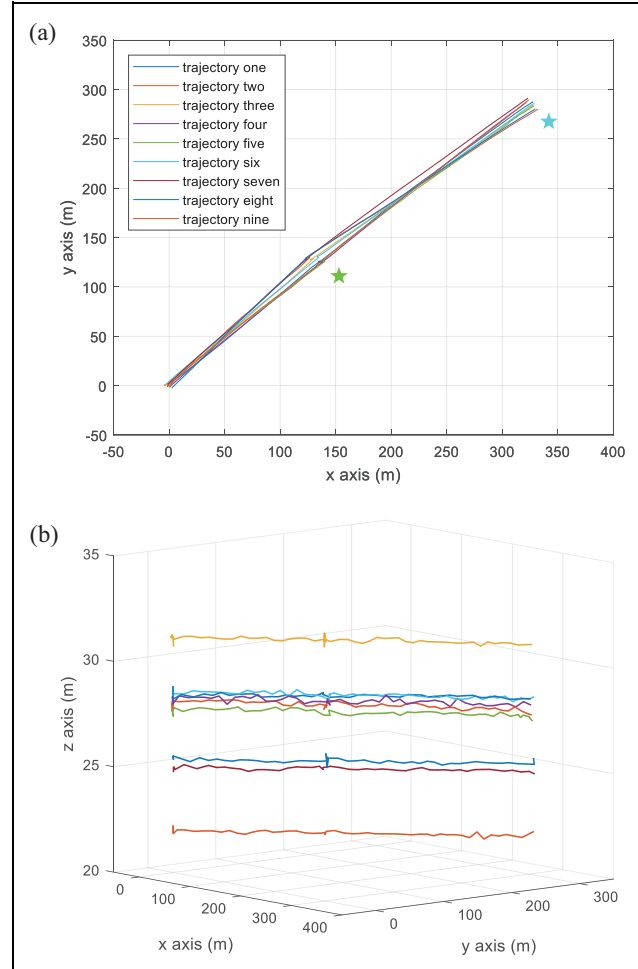


Figure 15. Planar and altitude views of multiple trajectories. (a) Planar view of multiple trajectories. (b) Altitude view of multiple trajectories.

view of flight trajectories. All practical trajectories are in agreement with the expected. Figure 15(b) displays the change in trajectories in terms of height. In particular, there exist skews between the directions of transmission lines and the trajectories 1 and 2. The occurrence of some poor trajectories is attributed to the following factors. Owing to the intrinsic pendulous traits of transmission lines, the difference between the computing orientation defined by VP and the ideal flight heading is inevitable. In addition, the number of detected transmission lines, which relates to many factors, such as the observation distance and angle, illumination, weather, and so on also has a significant impact on the computation of VP. Moreover, during the flight, in view of the general positioning precision, the flight course can likewise be influenced by wind.

Discussion

Owing to faster R-CNN, the false detection never came about during tests. Although the target is occasionally lost during inspection—likely on account of poor

illumination—it is not a severe issue because the currently effective tracking result is adequate to aid in navigation. With respect to tracking, a failure occasionally arises at the initialization stage, which is attributed to both the rapid rotation of the camera and the blurry image caused by the poor condition of illumination. In addition, as mentioned in Figure 11, when the background feature is more distinct than the foreground, the tracking may gradually drift with the background. Serious drift eventually causes the tracking to be initialized again by detection. Moreover, the strategy of flying with simultaneous rotation decreases the tracking drift to a great extent. The probability of a simultaneous failure of both tracking and detection is extremely low; thus, localization of the transmission tower is also very robust.

With respect to VP computation, it is actually not always the case that a greater number of detected transmission lines results in a more approximate computation heading to the ideal flight heading. This is because the inevitable pendulous trait of transmission lines becomes much more serious at the start and end of the transmission line. Thus, a determination should be implemented about whether the computation VP is located within the region of the detected bounding box. If the VP quality is poor, the UAV should fly along the direction of the convergence of transmission lines, or the experimental direction. Based on the aforementioned analysis, the extraction of transmission lines by applying segmentation via deep learning plays an important role in the determination of the flight heading.

Despite the excellent autonomous navigation performance of our system based on transmission towers and transmission lines, there exist some shortcomings. First, there is a lack of consideration of the distance to transmission lines; thus, the more distant navigation remains a challenge. Nowadays, at a secure distance of 20 m, it is difficult to directly measure the distance to transmission lines depending only on the existing sensors. Actually, the distance can be estimated by the position of the tracked transmission tower. The position of the transmission tower can be further estimated from several images obtained at multiple different positions. Second, our approach is just effective when meeting the following two hypotheses: the next tower to be positioned at the initialization stage (1) has a maximum detected bounding box relative to other detected towers and (2) must be within view. Navigating in more complex environments may require the assistance of GPS.

Conclusion

In this article, we proposed a safe and robust transmission-tower-based autonomous navigation approach for UAV transmission-line inspection. For the implementation of the proposed method, a UAV inspection platform carrying advanced embedded processors was developed. To the best of our knowledge, it is the first time that transmission-tower-based navigation method is put forward and is put

into practice. The transmission tower is regarded as an important landmark, by which the continuous inspection in the power corridor can be realized. Furthermore, the unique means of flying along one side of the transmission lines is adopted, which reduces the risk of inspection accidents. To achieve the aforementioned concept, an important navigation model was herein introduced, which addresses the problem that UAV will be capable of perception of a 3-D world simply by leveraging an image with perspective distortion.

Specifically, to achieve the real-time and robust localization of a tower, a detection-tracking visual strategy that integrates the detection and tracking is used. We finally apply faster R-CNN to the detection-tracking framework for a trade-off between speed, precision rate, and recall. At the stage of flight heading calculation, FCNs are first employed to realize the transmission-line extraction of great importance, which is effective, even in more complex environments. With respect to the precise flight defined by VP, the algorithm based on the RANSAC strategy, together with linear and nonlinear optimization, is employed. Additionally, an experimental flight heading is adopted to cope with the case in which a VP is non-existent. Finally, with the newly designed experimental platform, continuous flight experiments without GPS demonstrated the effectiveness of the navigation model and proposed methods.

Our ongoing and future work will focus on the maintenance of the safety distance between UAV and transmission lines to further improve the inspection safety. In addition, based on this novel navigation scheme, a transmission-line tracking algorithm is being developed for a high-quality observation of transmission line. Based on the earlier works, an online transmission-line fault diagnose system will come true soon.

Acknowledgements

We would like to thank Editage (www.editage.com) for English-language editing and publication support.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: supported by the National Natural Science Foundation of China under grant nos. 61673378 and 61421004.

ORCID iD

Xiaolong Hui  <http://orcid.org/0000-0003-0737-952X>
Jiang Bian  <http://orcid.org/0000-0002-5125-4882>

References

- Pagnano A, Höpf M and Teti R. A roadmap for automated power line inspection. Maintenance and repair. *Procedia CIRP* 2013; 12: 234–239.
- Tavares L and Sequeira JS. Riol-robotic inspection over power lines. *IFAC Proc Vol* 2007; 40(15): 108–113.
- Pouliot N, Richard PL and Montambault S. Linescout technology opens the way to robotic inspection and maintenance of high-voltage power lines. *IEEE Power Energy Technol Syst J* 2015; 2(1): 1–11.
- Lu M, Sheng G, Liu Y, et al. Research on auto-tracking algorithm for power line inspection based on unmanned aerial vehicle. In: *Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific*, Shanghai, China, March 2012, pp. 1–5. IEEE.
- Hrabar S, Merz T and Frousheger D. Development of an autonomous helicopter for aerial powerline inspections. In: *2010 1st International Conference on Applied Robotics for the Power Industry (CARPI)*, Montreal, Canada, October 2010, pp. 1–6. IEEE.
- Cao W, Zhu L, Han J, et al. High voltage transmission line detection for uav based routing inspection. In: *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, Wollongong, NSW, Australia, July 2013, pp. 554–558. IEEE.
- Chen Y, Li Y, Zhang H, et al. Automatic power line extraction from high resolution remote sensing imagery based on an improved radon transform. *Patt Recogn* 2016; 49: 174–186.
- Ceron A, Mondragon B, Prieto F, et al. Power line detection using a circle based search with uav images. In: *2014 International Conference on Unmanned Aircraft Systems (ICUAS)*, Orlando, FL, USA, May 2014, pp. 632–639. IEEE.
- Illingworth J and Kittler J. A survey of the hough transform. *Comput Vis Graph Image Process* 1988; 44(1): 87–116.
- Von Gioi RG, Jakubowicz J, Morel JM, et al. Lsd: a fast line segment detector with a false detection control. *IEEE Trans Patt Anal Mach Intell* 2010; 32(4): 722–732.
- Akinlar C and Topal C. Edlines: a real-time line segment detector with a false detection control. *Patt Recogn Lett* 2011; 32(13): 1633–1642.
- Araar O and Aouf N. Visual servoing of a quadrotor uav for autonomous power lines inspection. In: *2014 22nd Mediterranean Conference of Control and Automation (MED)*, Palermo, Italy, June 2014, pp. 1418–1424. IEEE.
- Zhou G, Yuan J, Yen IL, et al. Robust real-time uav based power line detection and tracking. In: *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, September 2016, pp. 744–748. IEEE.
- Zhang J, Liu L, Wang B, et al. High speed automatic power line detection and tracking for a uav-based inspection. In: *2012 International Conference on Industrial Control and Electronics Engineering (ICICEE)*, Xi'an, China, August 2012, pp. 266–269. IEEE.
- Steiger O, Lucas E and Maret Y. Automatic detection of transmission towers. In: *SENSORS, 2014 IEEE*, Valencia, Spain, Novembr 2014, pp. 1034–1037. IEEE.
- Leibe B, Leonardis A and Schiele B. Robust object detection with interleaved categorization and segmentation. *Int J Comput Vis* 2008; 77(1–3): 259–289.
- Sampedro C, Martinez C, Chauhan A, et al. A supervised approach to electric tower detection and classification for power line inspection. In: *2014 International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, July 2014, pp. 1970–1977. IEEE.
- Dalal N and Triggs B. Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, San Diego, CA, USA, June 2005, Vol. 1, pp. 886–893. IEEE.
- Cerón A, Mondragón I and Prieto F. Real-time transmission tower detection from video based on a feature descriptor. *IET Comput Vis* 2016; 11: pp. 33–42.
- Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, Montreal, Quebec, Canada, October 2015, pp. 91–99.
- Liu W, Anguelov D, Erhan D, et al. Ssd: single shot multibox detector. In: *European Conference on Computer Vision*, Amsterdam, Netherlands, October 2016, pp. 21–37. Springer.
- Fan X and Shin H. Road vanishing point detection using weber adaptive local filter and salient-block-wise weighted soft voting. *IET Comput Vis* 2016; 10(6): 503–512.
- Kong H, Sarma SE and Tang F. Generalizing laplacian of gaussian filters for vanishing-point detection. *IEEE Trans Intell Transp Syst* 2013; 14(1): 408–418.
- Shi J, Wang J and Fu F. Fast and robust vanishing point detection for unstructured road following. *IEEE Trans Intell Transp Syst* 2016; 17(4): 970–979.
- Elloumi W, Guissous K, Chetouani A, et al. Indoor navigation assistance with a smartphone camera based on vanishing points. In: *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Montbéliard-Belfort, France, October 2013, pp. 1–9. IEEE.
- Lu Y and Song D. Visual navigation using heterogeneous landmarks and unsupervised geometric constraints. *IEEE Trans Robot* 2015; 31(3): 736–749.
- Ji Y, Yamashita A and Asama H. Rgb-d slam using vanishing point and door plate information in corridor environment. *Intell Ser Robot* 2015; 8(2): 105–114.
- Teichmann M, Weber M, Zoellner M, et al. Multinet: Real-time joint semantic reasoning for autonomous driving. *arXiv preprint arXiv:161207695* 2016; .
- Hartley R and Zisserman A. *Multiple View Geometry in Computer Vision*. Cambridge, United Kingdom: Cambridge university press, 2003.
- Zeiler MD and Fergus R. Visualizing and understanding convolutional networks. In: *European conference on computer vision*, Zurich, Switzerland, September 2014, pp. 818–833. Springer.
- Henriques JF, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. *IEEE Trans Patt Anal Mach Intell* 2015; 37(3): 583–596.

32. Kang K and Wang X. Fully convolutional neural networks for crowd segmentation. *arXiv preprint arXiv:1411.4464* 2014;.
33. Fu H, Xu Y, Wong DWK, et al. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Prague, Czech Republic, April 2016, pp. 698–701. IEEE.
34. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, June 2016, pp. 770–778.
35. Zeiler MD, Krishnan D, Taylor GW, et al. Deconvolutional networks. In: *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA, June 2010, pp. 2528–2535. IEEE.
36. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* 2014; .
37. Zhang L, Lu H, Hu X, et al. Vanishing point estimation and line classification in a manhattan world with a unifying camera model. *Int J Comput Vis* 2016; 117(2): 111–130.
38. Nocedal J and Wright SJ. *Numerical Optimization, 2nd*. New York: Springer, 2006.
39. Redmon J and Farhadi A. Yolo9000: better, faster, stronger. *arXiv preprint arXiv:1612.08242* 2016.
40. Everingham M, Van Gool L, Williams CK, et al. The pascal visual object classes (voc) challenge. *Int J Comput Vis* 2010; 88(2): 303–338.
41. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015; 115(3): 211–252.