

# 结合图像语义分割的增强现实型平视显示 系统设计与研究

安喆, 徐熙平, 杨进华, 乔杨, 刘洋

长春理工大学光电工程学院, 吉林 长春 130022

**摘要** 为了提高驾驶员在车辆行驶过程中的安全性,设计了一种结合图像语义分割的增强现实型平视显示(AR-HUD)系统。首先,提出一种改进的单发多框检测器网络对道路场景图像进行语义分割,网络前端采用 VGG-16 提取图像特征,网络后端对获取的特征图进行上采样,从而对特征图进行像素分割。通过对网络的训练,得到场景目标的像素级分类结果,即环境的语义内容信息。随后,通过分析真实场景、光学显示系统、驾驶员之间的关系,将计算机产生的虚拟信息叠加到真实场景,并将显示内容注册到驾驶员视野中,从而提高行车安全。实验结果表明,语义分割算法的准确率能达到 77.8%,虚实注册算法处理每帧图像的时间平均为 45 ms,约 22 frame·s<sup>-1</sup>。

**关键词** 图像处理; 增强现实; 图像语义分割; 虚实注册

中图分类号 TH741

文献标识码 A

doi: 10.3788/AOS201838.0710004

## Design of Augmented Reality Head-up Display System Based on Image Semantic Segmentation

An Zhe, Xu Xiping, Yang Jinhua, Qiao Yang, Liu Yang

School of Optoelectronic Engineering, Changchun University of Science and Technology,  
Changchun, Jilin 130022, China

**Abstract** In order to improve the security of drivers, an augmented reality head-up display (AR-HUD) system is designed based on image semantic segmentation. Firstly, we propose an improved single shot multibox detector network for semantic segmentation of road scene images. The front end of the network uses VGG-16 to extract the image features, and the back ends of the network are sampled on the feature maps. Thus, the feature map is segmented. Through the training of the network, the pixel level classification results of the scene objects are obtained, namely, the semantic content information of the environment. Then, with analysis of the relationship among real scene, optical display system, and drivers, the virtual information generated by computer is added to the real scene. In this way, the content is registered into the driver's view to improve the safety of driving. Experimental results show that the accuracy of the semantic segmentation algorithm can reach 77.8%, and image processing time of the algorithm for each frame is 45 ms, in other words, about 22 frame·s<sup>-1</sup>.

**Key words** image processing; augmented reality; image semantic segmentation; virtual-real registration

**OCIS codes** 100.3020; 170.0170; 170.3010

## 1 引 言

增强现实(AR)技术是一种以真实场景信息为基础,以虚拟对象作为补充信息的一种技术。中国是全球交通和路况最复杂的地区之一,约 80% 的交

通事故是由于驾驶员的注意力分散导致的。汽车平视显示器(HUD)的出现,使驾驶员不用低头就可以看到相关信息,可以将更多的精力放到观察路面情况上,与此同时,也可以减少驾驶员由于要观察远处道路情况与近距离查看导航、车辆信息是而须频繁

收稿日期: 2018-01-25; 收到修改稿日期: 2018-03-08

基金项目: 国家自然科学基金(61605016)

作者简介: 安喆(1992—),女,博士研究生,主要从事光电检测技术方面的研究。E-mail: 2016200046@mails.cust.edu.cn

导师简介: 徐熙平(1969—),男,博士,教授,博士生导师,主要从事光电检测与质量控制、图像处理等方面的研究。

E-mail: xxp@cust.edu.cn(通信联系人)

转换视线所引发的视觉疲劳<sup>[1]</sup>。一些现有的 HUD 系统仅在驾驶员视线前方显示虚拟信息<sup>[2]</sup>,没有与真实环境进行融合,这种虚拟内容与真实场景之间的不匹配,会在一定程度上导致驾驶员注意力分散。为此,本文提出一套增强现实型平视显示 (AR-HUD) 系统,以实现虚拟图像与环境的真正融合,提高行车安全性。

增强现实系统中,首先需要对周围环境进行理解,才能将虚拟物体叠加到场景中的正确位置,即解决虚实注册问题<sup>[3]</sup>。现有的虚实注册方法可分为有标识<sup>[4]</sup>和无标识虚实注册<sup>[5]</sup>两大类。有标识注册方法通常在场景中放置人工标识物,或通过识别自然标识进行虚实注册,此类方法完全依赖于标识信息,若视角不在标识区域,则无法完成注册。无标识的虚实注册通过对场景中的内容进行获取与理解,计算虚实注册矩阵,从而实现虚拟信息与真实环境的融合,克服了前述问题。获取场景的内容信息,即解决场景的语义分割问题。图像的语义分割是指利用计算机自动分割并识别出图像中的内容<sup>[6]</sup>。一般的 AR 系统只对场景进行识别,并未对目标之间的像素所属类别进行细致的划分,也有部分系列利用边缘检测、区域提取等方法得到了目标间的像素分类结果,但这些方法需要人为设计算法,事先提取特征,计算过程往往较复杂。随着深度学习<sup>[7-10]</sup>的发展,出现了一系列语义分割算法,这些基于深度卷积网络的语义分割算法不需要事先提取特征,而是采用学习的方法对网络进行训练。Long 等<sup>[11]</sup>提出了基于深度学习的全卷积网络 (FCN) 语义分割方法,该网络将最后的全连接层替换为卷积层,再经过上采样得到图像的像素分类结果,并且可以输入任意大小的图像。该方法虽然实现了对目标的分割与标记,但由于在上采样过程中丢失了图像部分像素,精度不是很高。Badrinarayanan 等<sup>[12]</sup>提出了逐像素语义分割 (SegNet) 网络,将最大池化指数转移至解码器中,改善了分割分辨率,并应用于道路和车辆的分割。Chen 等<sup>[13]</sup>提出前端为 FCN、后端采用条件随机场 (CRF) 优化前端输出的深度卷积网络对图像进行语义分割,采用空洞卷积,在不增加参数数量的情况下增大了感受野,改善了分割网络,并实现多尺度处理,后端通过 CRF 实现结构化预测,使分割性能得到进一步提升。

本文提出一种基于语义分割和 AR 相结合的车辆辅助系统。首先设计了改进的单发多框检测器 (SSD) 网络用于图像的语义分割,获取场景的语义

内容信息,并结合 AR-HUD 系统,分析系统间各部分之间的关系,通过计算虚实注册矩阵,将虚拟图像注册到真实场景中,实现对道路信息的增强,提高驾驶安全性。

## 2 AR-HUD 系统总体架构

系统总体架构如图 1 所示,包括图像语义获取模块、光学透射式虚像显示模块和汽车处理器模块。首先由跟踪相机获取道路场景图像,再经过改进的 SSD 语义分割网络对道路场景图像进行语义内容获取。经过汽车处理器模块对人眼、虚像与场景之间的对应关系进行计算分析,得出虚拟图像应投射的位置。然后将需要投射的信息通过数据线传递给光学系统的投影组件,投影组件将要显示的信息投射到一个特殊设计的自由曲面反射镜,以矫正挡风玻璃产生的图像畸变,并将图像反射到驾驶员的视野中,从而将虚拟图像注册到真实道路环境,并使驾驶员在视线前方看到虚实融合的增强场景。

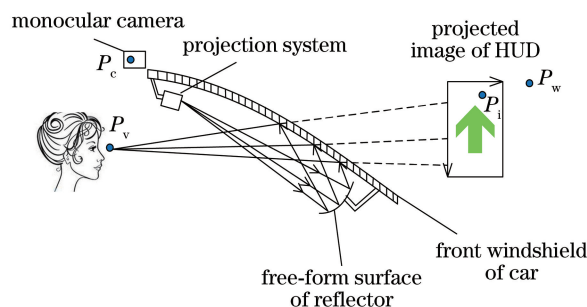


图 1 AR-HUD 系统

Fig. 1 AR-HUD system

## 3 结合语义分割的虚实注册方法

### 3.1 改进的 SSD 语义分割网络

为了获取场景的内容信息,从而将虚拟内容叠加到真实场景,提出一种基于改进的 SSD 深度学习网络的场景语义内容获取方法。

SSD 是一种基于回归的目标检测与识别方法<sup>[14]</sup>,使用单个回归器预测真实边框。这种深度学习模型将检测和识别过程在同一个网络中完成,训练时用牛津大学视觉几何课题组提出的 16 层深度卷积网络 (VGG-16) 提取特征,通过将 VGG-16 后两层的全连接层替换为新的卷积层,能够提取不同层次特征图上的特征,再经过特征回归预测,得到不同尺度目标的分类结果。由于得到了多尺度特征,可以对图像中的目标进行多层次判断,在提高精度的同时,保持了算法的速度。

对 SSD 网络进行改进,不仅在网络中输出目标所属类别结果,并且对目标的像素所属类别进行分类。改进的 SSD 网络的主要过程分为:1)通过 VGG-16 深度神经网络提取输入图像的深度特征;2)根据不同尺度的深度特征图设计不同大小的特征

抓取盒;3)用转置卷积对深度特征图进行上采样,逐层恢复特征图尺寸;4)通过提取特征抓取盒所对应深度图特征预测目标的所属类别;5)采用非最大值抑制(NMS)筛选预测结果,同时输出像素分割结果。整个网络模型如图 2 所示。

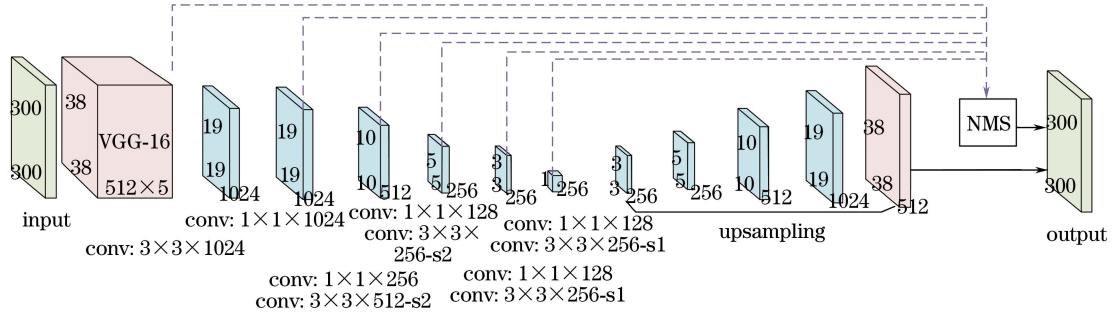


图 2 改进的 SSD 网络示意图

Fig. 2 Diagram of improved SSD network

网络的输入是整张图片,然后经过 VGG-16 网络对输入图像提取深度特征,在 VGG-16 网络后面添加不同尺度的卷积神经网络(CNN)层,用来提取图像的多尺度特征,卷积过程采用空洞卷积算法<sup>[15]</sup>。其中,conv:3x3x256-s2 表示卷积核,其卷积步长为 2,用来处理 CNN 层。经过几次卷积操作,特征图数量变为 128,得到用于检测识别的特征图,然后利用双线性插值的方法对特征图进行上采样操作,逐层恢复特征图大小,最终输出目标像素的分类结果。通过对改进 SSD 网络的训练,可以得到真实场景的目标分类结果,为后续虚拟图像的注册提供环境信息。

### 3.2 结合语义内容的虚实注册算法

经过深度卷积神经网络对道路图像进行处理,得到了目标的像素分类结果。在获得环境信息后,就可以通过对环境的理解,将虚拟图像注册到真实场景,使驾驶员观察到虚拟信息与道路场景融合的图像。此时,需要获取驾驶员、光学系统、道路场景之间的关系。如图 1 所示,设  $P_w = (x, y, z)$  为目标上任意一点, $P_c$  为在跟踪相机下与之相对应的点, $P_v$  为在人眼坐标系下的对应点,人眼通过光学系统观察到虚拟图像,两者构成组合成像系统,将其定义为虚拟相机坐标系, $P_i$  为在虚拟图像平面上所对应的像素点。它们之间的关系可以表示为

$$P_i = \mathbf{K}_v^{3 \times 4} \begin{pmatrix} \mathbf{r}_{v \leftarrow c}^{3 \times 3} & \mathbf{t}_{v \leftarrow c}^{3 \times 1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_{c \leftarrow w}^{3 \times 3} & \mathbf{t}_{c \leftarrow w}^{3 \times 1} \\ 0 & 1 \end{pmatrix} P_w, \quad (1)$$

式中  $\mathbf{K}_v$  为虚拟相机内参, $\mathbf{r}$ 、 $\mathbf{t}$  为相机的旋转平移矩阵。

在求解(1)式时,首先用单目深度估计<sup>[16]</sup>的方

法对图像中目标像素点的深度  $d$  进行估计。在获取像素点的深度后,将二维图像的像素点转换为空间点;若经过 3.1 节的改进 SSD 网络获取的道路场景目标上的某一像素点坐标为  $(u, v, d)$ ,则将这一点转换为空间点  $P_w$  的方程为

$$\begin{cases} x = \frac{z}{f_x}(u - c_x) \\ y = \frac{z}{f_y}(v - c_y) \\ z = \frac{d}{s} \end{cases}, \quad (2)$$

式中  $s$  为深度缩放因子, $f_x, f_y, c_x, c_y$  为相机内参。

将(2)式所得结果代入(1)式,经过矩阵求解即可得到空间坐标  $P_w$  相对于虚拟平面上的对应点  $P_i$  的变换矩阵,将其定义为初始变换矩阵  $\mathbf{A}$ 。

当有新帧到来时,相机的姿态改变导致虚实注册矩阵也相应改变。此时对前后两关键帧之间的输入图片进行特征点提取,再对两幅图片的特征点进行匹配,结合点的深度数据二维点转换为三维点,得到特征点集合,记为  $p = \{p_1, \dots, p_m\}$  和  $p' = \{p'_1, \dots, p'_m\}$ 。经过前面对输入图像中被检测目标像素的分类,可以对两帧之间的目标进行匹配。假设在场景中检测到有  $k$  类目标,则匹配后目标之间的空间像素点集合为  $q_k = \{q_1, \dots, q_n\}$  与  $q'_k = \{q'_1, \dots, q'_n\}$ 。由于目标像素与特征点之间必然有重合点,将二者取并集去掉重合点,得到: $l = p \cup q_k$  和  $l' = p' \cup q'_k$ 。匹配的空间点经  $\mathbf{R}$ 、 $\mathbf{T}$  变换,满足如下关系:

$$l_i = \mathbf{R}l'_i + \mathbf{T}. \quad (3)$$

此时的问题为相机姿态估计问题,即估计(3)式的  $\mathbf{R}$ 、 $\mathbf{T}$ 。此类问题一般用迭代最近点(ICP)<sup>[17]</sup>进行求解,设

$$e_i = l_i - (\mathbf{R}l'_i + \mathbf{T})。 \quad (4)$$

则相机姿态矩阵  $\mathbf{R}$ 、 $\mathbf{T}$  可用最小二乘问题进行求解:

$$\min J = \frac{1}{2} \sum_{i=1} \| e_i \|_2^2 = \frac{1}{2} \sum_{i=1} \| [l_i - (\mathbf{R}l'_i + \mathbf{T})] \|_2^2。 \quad (5)$$

这里的相机姿态估计不仅考虑了特征点之间的约束关系,而且增加了同一类别目标物体之间的约束关系。记当前相机相对初始位置的注册矩阵  $\mathbf{R}$ 、 $\mathbf{T}$  为  $\mathbf{B}$ 。则虚实注册矩阵  $\mathbf{T}_v$  可用初始注册矩阵  $\mathbf{A}$  与  $\mathbf{B}$  的合成矩阵表示:

$$\mathbf{T}_v = \mathbf{A} \cdot \mathbf{B}。 \quad (6)$$

求出虚实注册矩阵,即可将虚拟图像叠加到道路场景,使驾驶员能够看到虚实融合的图像。

## 4 实验过程与分析

### 4.1 语义分割算法实验及性能测试

实验的网络训练环境为 64 位 Windows 系统, GPU 为 GTX1060,实验平台为 Caffe。输入图像经

过 VGG-16 网络提取特征后,再经过 6 层特征提取层提取此时特征图的特征,由于每个特征层的尺度不同,这样就能够获得图像的多尺度特征。为了对图像的像素进行分类,用双线性插值的方法对特征图进行上采样,恢复原图大小。通过上述网络训练过程,即可在网络后端输出图像中不同目标的像素分类结果。模型在训练时采用 CamVid 数据集<sup>[18]</sup>,训练时初始学习率为 0.001,权重惩罚项为 0.0005,动量项为 0.9,一次处理图像批量为 32。在训练时间方面,第一次使用训练集中的全部样本训练,耗时 129.8 s,第二次为 98.7 s,此后使用训练集中的全部样本训练一次的时间均在 95 s 左右。图 3(a)、(b)分别为训练过程的迭代次数与准确率和损失函数之间的关系。随着迭代次数的增加,训练损失快速下降,与此同时准确率上升,基本满足了实际需求。

采用两种方法对实验结果进行定量评价。

1)假设输入图像中能够被检测到的物体类别数为  $k$ ,将未被检测到的物体记为一类,则总的类别数为  $k+1$ 类。记  $p_{ii}$  为被正确分类的像素数量,  $p_{ij}$  为被错误分类的像素数量。通过计算被准确分类的像素个数与所有像素个数的比值,并由总类数取平均值,对像素准确率平均值(MPA)作出评价。

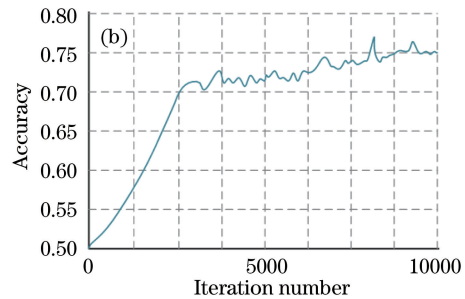
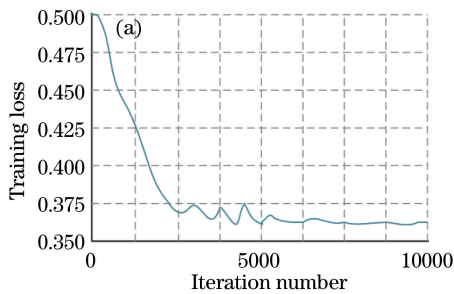


图 3 改进的 SSD 网络训练过程。(a)迭代次数与训练损失的关系;(b)迭代次数与准确率的关系

Fig. 3 Training process of improved SSD network. (a) Relationship between iteration number and training loss; (b) relationship between iteration number and accuracy

$$V_{\text{MPA}} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}。 \quad (7)$$

$$V_{\text{MIoU}} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}。 \quad (8)$$

2)像素的分类问题可以看成集合之间的交集与其并集之间的重合比例计算问题,其中交集为真实例的数量,并集为错误正例。将真实正例,错误正、负例的和定义为总数量,通过计算真实正例与总数量的比值,并由总类数取平均值,对像素分类结果的准确度进行平均交并比(MIoU)评价。

表 1 为本文方法与其他同类方法的评价结果及处理每帧图像的平均时间对比。可见,相比其他方法,本文算法对目标的分类准确度有所提高,且对图像处理的速度更快。

### 4.2 视觉虚实注册算法性能测试

通过改进的 SSD 网络对道路场景进行语义分

表 1 网络输出结果评价及平均处理时间对比  
Table 1 Evaluation of network output results and comparison of average processing time

Algorithm	MPA / %	MIoU / %	$t / (\text{ms} \cdot \text{frame}^{-1})$
DeepLab-CRF <sup>[13]</sup>	76.2	70.8	436
SegNet <sup>[12]</sup>	70.1	66.4	515
Proposed	77.8	72.1	410

割,然后结合深度信息将图像坐标转化为空间坐标点,获取目标的空间位置,再通过(6)式计算获得虚拟图像在虚拟平面上注册的位置。实验结果如图4所示。

图4(a)为原始图像,图4(b)为文献[12]的语义分割图像,图4(c)为采用本文算法的语义分割结果,图4(d)为虚实注册结果图像。其中第一行为晴朗天气下的结果,第二行为能见度200 m的雾天行

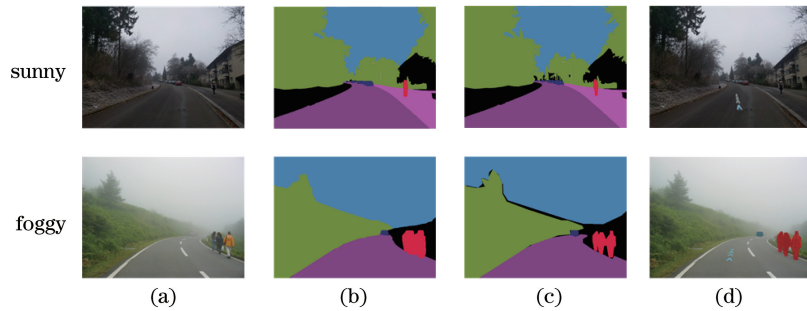


图 4 不同环境下语义分割及虚实注册结果。(a)原始图像;(b)文献[12]方法结果;(c)本文算法结果;(d)虚实注册结果

Fig. 4 Results of semantic segmentation and virtual-real registration in different environments.

(a) Original images; (b) method of Ref. [12]; (c) proposed method; (d) virtual-real registration images

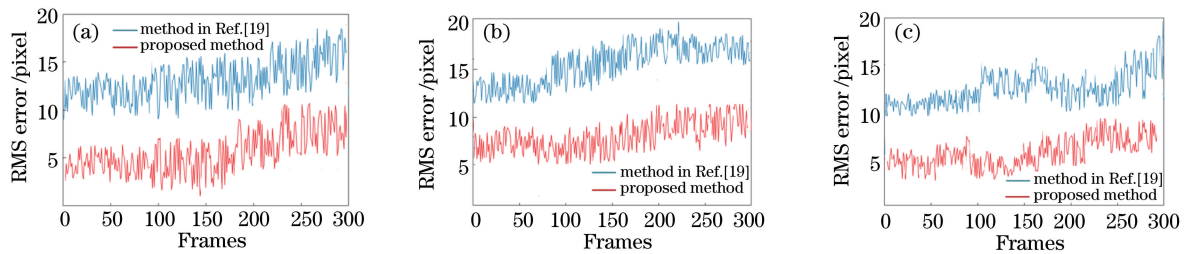


图 5 虚实注册在不同方向上的 RMS 误差及对比。(a)  $x$ ; (b)  $y$ ; (c)  $z$

Fig. 5 RMS error of virtual-real registration and comparison in different directions. (a)  $x$ ; (b)  $y$ ; (c)  $z$

图5表明,本文方法的注册误差基本小于10 pixel,与文献[19]方法相比,其误差更小。本文算法改善了由于虚拟图像与真实环境之间的不匹配导致的注意力分散问题,与其他方法相比,得到了更好的结果,最终实现了虚拟图像与现实场景的真正融合,有助于提高驾驶的安全性。

## 5 结 论

建立结合图像语义分割的 AR-HUD 系统。提

驶环境下的结果。AR-HUD 系统将行人、车辆、路面等目标进行分类和像素分割,在获取环境信息后,通过计算将虚拟图像叠加到现实场景中,为驾驶员提供更多有用信息,以提高行车安全。通过与其他方法进行对比,结果表明本文算法效果更好。相对于晴朗天气下,雾天环境虽然对语义分割及注册效果有一定影响,但基本满足了应用需求。虚实注册算法对每帧图像处理的时间约为 45 ms,约合  $22 \text{ frame} \cdot \text{s}^{-1}$ ,可满足实时性的要求。实验通过计算目标的真实坐标与虚实注册矩阵重投影坐标的均方根(RMS)误差,对本文算法的性能进行测试,并与基于单应性矩阵的注册方法<sup>[19]</sup>进行对比,图5(a)~(c)分别为  $x$ 、 $y$ 、 $z$  方向上的 RMS 误差,测试图像帧数为 300。

出改进的基于 SSD 网络的语义分割算法,通过分析 AR-HUD 系统各部分之间的关系,结合改进的 SSD 网络得到的场景语义信息,将虚拟内容叠加到真实场景中,实现了对场景信息的增强。实验结果表明,语义分割算法的准确率为 77.8%,虚实注册算法的速度约为  $22 \text{ frame} \cdot \text{s}^{-1}$ ,与现有的其他方法相比,其注册精度更高,且基本满足实时性要求,驾驶员无须过多分散注意力,即能获得道路信息的有效提示,从而提高行车安全。

研究过程中发现,光照、遮挡、抖动问题等因素会影响虚实注册的稳定性和准确率,这些将作为下一步研究的重点。

## 参 考 文 献

- [1] Park H S, Min W P, Won K H, *et al.* In-vehicle AR-HUD system to provide driving-safety information[J]. *ETRI Journal*, 2013, 35(6):1038-1047.
- [2] Yu Y H. Research status and development trend of augmented reality technology[J]. *Journal of Hunan Mass Media Vocational Technical College*, 2016, 16(1): 55-57.  
余艳红. 增强现实技术的研究现状及发展趋势[J]. *湖南大众传媒职业技术学院学报*, 2016, 16(1): 55-57.
- [3] Gui Z W, Liu Y, Chen J, *et al.* Online learning of tracking and registration based on natural scenes[J]. *Journal of Software*, 2016, 27(11): 2929-2945.  
桂振文, 刘越, 陈靖, 等. 基于自然场景在线学习的跟踪注册技术[J]. *软件学报*, 2016, 27(11): 2929-2945.
- [4] Fiala M. ARTag, a fiducial marker system using digital techniques [C] // *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, 2005, 2:590-596.
- [5] Zhang G, Chen H S, Ye Y D. A LoG operator based markerless augmented reality algorithm: LoG-PTAMM[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(9):1577-1586.  
张格, 陈昊升, 叶阳东. 一种基于 LoG 算子的无标识增强现实算法: LoG-PTAMM[J]. *计算机辅助设计与图形学学报*, 2016, 28(9): 1577-1586.
- [6] Gao K J, Sun S Y, Yao G S, *et al.* Semantic segmentation of night vision images for unmanned vehicles based on deep learning[J]. *Journal of Applied Optics*, 2017, 38(3): 421-428.  
高凯珺, 孙韶媛, 姚广顺, 等. 基于深度学习的无人车夜视图像语义分割[J]. *应用光学*, 2017, 38(3): 421-428.
- [7] Liu C H, Li Z, Xu C, *et al.* BRDF model for commonly used materials of space targets based on deep neural network[J]. *Acta Optica Sinica*, 2017, 37(11): 1129001.  
刘程浩, 李智, 徐灿, 等. 基于深度神经网络的空间目标常用材质 BRDF 模型[J]. *光学学报*, 2017, 37(11): 1129001.
- [8] Lu Y S, Li Y X, Liu B, *et al.* Hyperspectral data haze monitoring based on deep residual network[J]. *Acta Optica Sinica*, 2017, 37(11): 1128001.
- 陆永帅, 李元祥, 刘波, 等. 基于深度残差网络的高光谱遥感数据霾监测[J]. *光学学报*, 2017, 37(11): 1128001.
- [9] Huang H, He K, Zheng X L, *et al.* Spatial-spectral feature extraction of hyperspectral image based on deep learning[J]. *Laser & Optoelectronics Progress*, 2017, 54(10): 101001.  
黄鸿, 何凯, 郑新磊, 等. 基于深度学习的高光谱图像空-谱联合特征提取[J]. *激光与光电子学进展*, 2017, 54(10): 101001.
- [10] Wu S C, Zhao H T, Sun S Y. Depth estimation from monocular infrared video based on bi-recursive convolutional neural network[J]. *Acta Optica Sinica*, 2017, 37(12): 1215003.  
吴寿川, 赵海涛, 孙韶媛. 基于双向递归卷积神经网络的单目红外视频深度估计[J]. *光学学报*, 2017, 37(12): 1215003.
- [11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(4): 640-651.
- [12] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for scene segmentation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(12): 2481-2495.
- [13] Chen L C, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFs [J]. *Computer Science*, 2014, 11(4): 357-361.
- [14] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector [C] // *Proceedings of 2016 European Conference on Computer Vision*, 2016, 9905: 21-37.
- [15] Zhang K. Research on a rapid fusion method for remote sensing images based on an improved atrous wavelet decomposition [D]. Kaifeng: Henan University, 2016: 66-72.  
张凯. 基于改进 atrous 小波分解的遥感影像快速融合方法的研究[D]. 开封: 河南大学, 2016: 66-72.
- [16] Xu L, Zhao H T, Sun S Y. Monocular infrared image depth estimation based on deep convolutional neural networks [J]. *Acta Optica Sinica*, 2016, 36(7): 0715002.  
许路, 赵海涛, 孙韶媛. 基于深层卷积神经网络的单目红外图像深度估计[J]. *光学学报*, 2016, 36(7): 0715002.
- [17] Besl P J, Mckay N D. Method for registration of 3D shapes [C] // *Proceedings of 1992 Robotics-DL tentative International Society for Optics and Photonics*, 1992, 14: 239-256.

- [18] Brostow G J, Shotton J, Fauqueur J, *et al.* Segmentation and recognition using structure from motion point clouds [C] // Proceedings of 2008 European Conference on Computer Vision. 2008, 5302: 44-57.
- [19] Fei T, Liang X H, He Z Y, *et al.* A registration method based on nature feature with KLT tracking algorithm for wearable computers [C] // Proceedings of 2009 International Conference on Cyberworlds. 2009, 1: 416-421.