

基于深层卷积神经网络的单目红外图像深度估计

许 路¹ 赵海涛¹ 孙韶媛²

¹华东理工大学信息科学与工程学院, 上海 200237

²东华大学信息科学与技术学院, 上海 201620

摘要 针对从单目红外图像中恢复深度信息的问题,提出了一种基于深层卷积神经网络(DCNN)的深度估计方法。用劳斯掩膜和梯度检测器分别提取不同尺度下红外图像的纹理能量与纹理梯度,并将这两种纹理信息作为红外图像的第一种特征;提取图像中像素及其邻域的灰度值,以及统计其灰度直方图作为另外两种特征;分别用三种特征和深度信息标签训练 DCNN,得到三种训练后的 DCNN 分别对单目红外图像进行深度估计。实验结果表明,相比较另外两种特征,用纹理信息训练的 DCNN 能够更有效地估计深度,并且优于现有的估计方法,尤其能较好地表现局部场景的深度变化。

关键词 机器视觉;卷积神经网络;深度估计;单目红外图像;纹理信息

中图分类号 TP391.9 **文献标识码** A

doi: 10.3788/AOS201636.0715002

Monocular Infrared Image Depth Estimation Based on Deep Convolutional Neural Networks

Xu Lu¹ Zhao Haitao¹ Sun Shaoyuan²

¹*Institute of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China*

²*Institute of Information Science and Technology, Donghua University, Shanghai 201620, China*

Abstract In order to recover depth information from monocular infrared image, a depth estimation algorithm based on novel deep convolutional neural networks (DCNN) is proposed. The texture energy and texture gradient of infrared images are extracted by using Laws' masks and the gradient detector at different scales. These two types of texture information are considered as the first kind of features. The selected gray values and their statistical histogram in specific areas are considered as another two kinds of features. The DCNN are trained on these three kinds of features with the corresponding depth labels respectively. The trained DCNN are then utilized to estimate the depths of testing monocular infrared images respectively. Experimental results show that compared with other methods, the DCNN trained by texture information can estimate the depth much better than those of the existing methods, especially in the depth changes of local scenes.

Key words machine vision; convolutional neural networks; depth estimation; monocular infrared image; texture information

OCIS codes 150.1135; 100.2960; 350.4600

1 引 言

对二维图像进行深度估计^[1-2]与深度重建^[3-4]是计算机视觉领域中的一个重要课题,也是目前应用最广泛,发展最快的研究方向之一。多目图像的深度估计和单目图像的深度估计是两类主要的研究问题。基于

收稿日期: 2016-01-25; **收到修改稿日期:** 2016-03-30

基金项目: 国家自然科学基金(61375007)、上海市科委基础研究项目(15JC1400600)

作者简介: 许 路(1992—),男,硕士研究生,主要从事深度神经网络与计算机视觉方面的研究。

E-mail: luxu92@163.com

导师简介: 赵海涛(1974—),男,博士,教授,主要从事模式识别、计算机视觉方面的研究。

E-mail: haitaozhao@ecust.edu.cn(通信联系人)

多目图像的深度估计是根据同一场景在不同图像中的特征差异来估计深度信息。而单目图像的深度估计只能依据有限的先验知识和图像本身的特征来提取深度线索进行深度估计。

早在 20 世纪 70 年代就有人提出根据阴影计算形状^[5]算法来恢复单幅图像的深度信息。阴影计算形状是指根据图像中像元点的亮度来计算其法向量,从而得到景物几何形状的算法。但该算法需要知道光源的位置和物体表面的反射特性等大量的先验知识,从而使得其应用的局限性增加。Saxena 等^[6-8]在 Make3D 项目的研究中,从马尔可夫随机场(MRF)提取了彩色图像在不同尺度下的局部和全局特征,并使用拉普拉斯模型等监督学习方法来训练模型进行深度估计。而丁伟利等^[9]则通过统计图像中不同封闭区域的边缘和颜色信息来推断城市道路图像中的深度信息。上述研究工作都是围绕着彩色图像展开的,本文主要研究无彩色信息的单目红外图像的深度估计。

红外图像是反映场景红外辐射分布和温度变化的一种灰度图像,近年来被应用于辅助夜间驾驶^[10]等领域。与彩色图像相比,红外图像的分辨率较低、清晰度较差,且缺乏深度信息和空间感^[11-12]。红外图像的深度信息有助于驾驶者直观地发现目标物体的距离远近,从而提高夜间驾驶的安全性。另外,由于深度信息能够有效地帮助观察者从周围景物中辨识出目标物体,所以对红外图像的深度估计研究也能够提高军事目标侦查的准确率。目前单目红外图像深度估计的方法有基于支持向量机(SVM)的红外深度估计方法^[13],通过独立成分分析(ICA)提取与深度信息线性相关性最大的特征,用来训练估计模型。但该特征筛选方法忽视了实际场景与深度特征之间的非线性关系。基于核主成分分析(KPCA)和主成分分析(PCA)相结合提取红外特征的方法^[14],通过反向传播(BP)神经网络建立深度估计模型。但其深度估计的准确度非常有限,尤其对天空和地面难以做到较好地区分。

由于红外图像和彩色图像的成像机制不同,彩色图像的深度特征提取方法不完全适用于红外图像。而现有的单目红外图像的深度估计方法稳健性较差,易受红外图像中景物变化的影响。本文提出了用深层卷积神经网络(DCNN)来进行单目红外图像的深度估计。深层卷积神经网络近几年来成功地应用于图像分类^[15]、视频分类^[16]、物体识别^[17]等领域,它是一种包含多层非线性运算的网络结构,能够很好地反映数据的复杂非线性映射关系。红外图像的深度估计问题的本质是研究红外特征与深度值之间的复杂非线性关系,因此 DCNN 适用于这类问题的研究。本文在 make3D^[6-8]特征提取方法的基础上,提取红外图像中三种不同的特征,分别用来训练基于 DCNN 的深度估计模型。这三种输入特征分别为:1) 通过劳斯掩膜和梯度检测器获得基于多尺度的纹理信息(包含纹理能量和纹理梯度);2) 红外图像中像元及其邻域的灰度值;3) 红外图像中景物的灰度统计直方图。本文着重研究不同的输入特征图和 DCNN 结构的变化对深度估计准确率的影响程度,实验结果表明,相比较另外两种特征,用多尺度纹理信息训练的网络模型能够更有效地估计单目红外图像的深度。

2 红外深度线索提取

2.1 纹理信息

Lee 等^[18]认为纹理信息(包含纹理能量和纹理梯度)与图像深度存在密切关系,Liu 等^[19]揭示了在三维(3D)重建过程中纹理信息对于 3D 图像质量的重要影响。Saxena 等^[6-8]在 make3D 项目中采取基于多尺度的方法对彩色图像进行深度估计。

上述工作主要是针对彩色图像进行的 3D 重建,理论分析和实验结果表明纹理特征具有较好的效果。针对红外图像,选择纹理信息和清晰度等辅助信息作为特征,特征的提取是在红外图像的不同尺度下完成的,具体过程如图 1 所示。选取 9 种劳斯掩膜 $M_n, n=1, \dots, 9$, 分别与原图像 I 进行卷积,得到卷积后的图像 $P_n = I * M_n$, 其中 $*$ 表示卷积运算,再选择以 20° 为间隔的 9 种边缘检测器 $F_n, n=10, \dots, 18$, 提取图像的梯度图 $P_n = I * F_n$, 共获得 18 张图片。

图 1(b)为图 1(a)中的图像块提取部分。在图 1(b)中,尺度 2 中的虚线部分为尺度 1 中的 X_0 , 尺度 3 中的虚线部分为尺度 2 中的 Y_0 。纹理信息中各种特征所属的尺度和图像如表 1 所示。

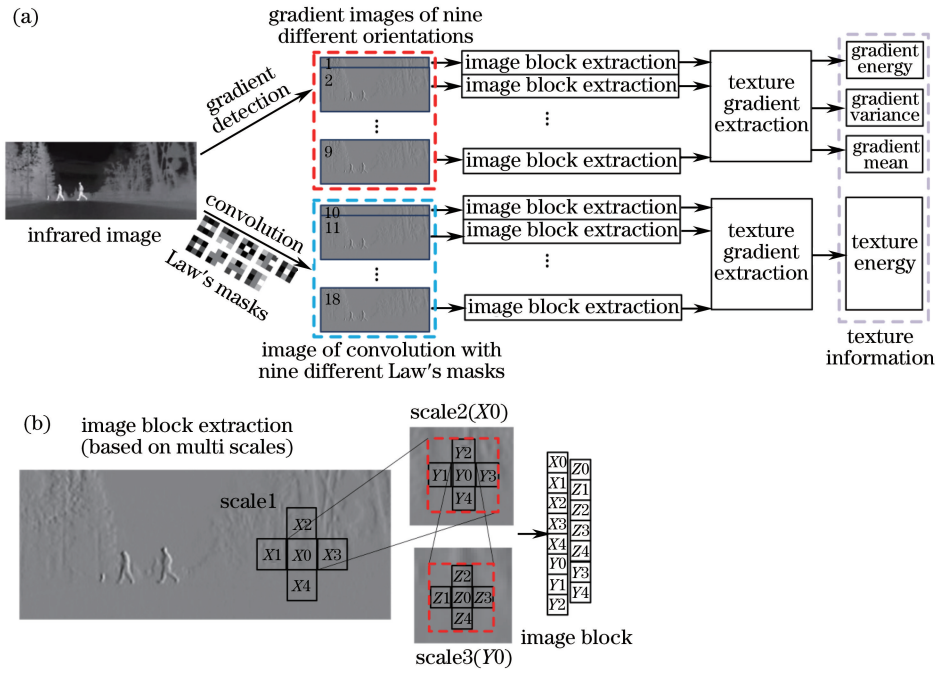


图 1 纹理信息的提取

Fig. 1 Extraction of texture information

表 1 纹理信息的组成

Table 1 Composition of texture information

Feature	From scale	From images
Texture energy	Scale 1,2,3	Images of convolution with Laws' masks
Texture square energy	Scale 2,3	Images of convolution with Laws' masks
Gradient energy	Scale 1,2,3	Gradient images
Gradient square energy	Scale 2,3	Gradient images
Gradient mean	Scale 2	Gradient images
Gradient variance	Scale 2	Gradient images

具体提取方式如下:1) 纹理能量和梯度能量的提取。根据图 1(b)的方法,计算图 1(a)的 18 幅图像(9 幅梯度图像和 9 幅劳斯掩膜卷积后的图像)中每幅图像的像元点 $P_{n_\lambda}(i, j)$ 在三个尺度上各个图像块的能量总和在尺度 2,3 上各个图像块的平方能量总和。其中前 9 幅的结果为梯度能量 $E_{m_1}^{n_1}$,后 9 幅的结果为纹理能量 $E_{m_2}^{n_2}$,其表示为

$$E_{m_k}^{n_\lambda} = \sum_{i,j \in m_k} |P_{n_\lambda}(i, j)|^k, \quad (1)$$

式中 m_k 为图 1(b)中不同尺度下的图像块的集合,当 $k=1$ 时, m_1 包含 $X_0, \dots, X_4, Y_0, \dots, Y_4, Z_0, \dots, Z_4$, 这 15 个图像块;当 $k=2$ 时, m_2 包含 $Y_0, \dots, Y_4, Z_0, \dots, Z_4$ 这 10 个图像块。 n_λ 为图像序号,当 $\lambda=1$ 时, n_1 为图 1(a)中 9 幅梯度图像的序号;当 $\lambda=2$ 时, n_2 为图 1(a)中劳斯掩膜卷积后所得 9 幅图像的序号。2) 梯度均值和梯度方差的提取。分别计算 P_{n_λ} 在尺度 2 上各个图像块的梯度均值 $M_{\eta}^{n_1}$ 和梯度方差 $S_{\eta}^{n_1}$, 两者分别表示为

$$M_{\eta}^{n_1} = \frac{1}{\text{size}(\eta)} \sum_{i,j \in \eta} P_{n_1}(i, j), \quad (2)$$

$$S_{\eta}^{n_1} = \frac{1}{\text{size}(\eta)} \sum_{i,j \in \eta} [P_{n_1}(i, j) - M_{\eta}^{n_1}]^2, \quad (3)$$

式中 η 为包含图 1(b) 中的 Y_0, \dots, Y_4 这 5 个图像块的集合, $\text{size}(\cdot)$ 用来计算图像块所包含的像元数。

最后计算在原红外图像中两个较大尺度上的清晰度为

$$C_v = \frac{1}{\text{size}(v)} \sum_{i,j \in v} [I(i,j) - \overline{I(i,j)}]^2, \quad (4)$$

式中 v 为包含图 1(b) 中的 $X_0, \dots, X_4, Y_0, \dots, Y_4$ 这 10 个图像块的集合。加上清晰度特征, 共计得到 550 个特征, 为了保证深度神经网络的深度估计效果, 去掉其中较大的 21 个值, 将剩余的 529 个特征值组成分辨率为 $23 \text{ pixel} \times 23 \text{ pixel}$ 的特征图, 用来训练卷积神经网络。

2.2 场景灰度值和灰度统计直方图

红外图像的特殊成像机制导致其分辨率较低, 局部信息缺乏, 所以简单的使用局部特征不足以实现对深度的估计。为了与纹理信息特征数目保持一致, 选取待估计深度的像元周围的 529 个灰度值组成特征图并统计灰度值的直方图, 作为两类特征输入分别训练卷积神经网络。

灰度直方图 G 是表示各种灰度级别的像元出现次数的统计图, 能够有效地反映不同灰度出现的频率。由于灰度分布过多地集中在狭窄区间会影响图像中细节的深度估计效果, 从而需要进行灰度直方图均衡化。其表达式为

$$G(a) = \frac{q_a}{q}, a = 0, \dots, 255, \quad (5)$$

$$T(a) = \sum_{j=0}^a G(j), \quad (6)$$

式中 q 为灰度图像中像元的总和, n_a 为灰度值 a 出现的次数, T 为均衡后的灰度统计直方图。

3 DCNN

3.1 网络结构

DCNN 是一种热门的人工神经网络, 与传统的神经网络相比, 它的权值共享特性有效地减少了权值的数量。其中卷积是一种线性运算, 能够有效地降低噪音并增强特征, 而采样层则是清除冗余数据的过程。DCNN 中的特征图在隐层间传递的过程中, 分辨率不断减小, 而数量却不断递增, 产生了更丰富的特征信息, 使得 DCNN 能够更好地表现红外图像特征与红外图像深度的关系^[20-21]。因此, 选择 DCNN 进行红外图像的深度估计。

图 2 是 DCNN 结构图。经过多次实验, 确定最佳的隐层层数为 6, 其中包含 3 个卷积层, 2 个采样层和 1 个全连接层(第一层卷积层有 6 张特征图, 第二层和第三层分别有 16 张和 120 张), 卷积层与采样层交替相连。仿照 LeNet-5 神经网络的结构, 将卷积层的特征图数量分别设定为 6, 16, 20。由于纹理信息和灰度值的特征数都为 529, 所以需要将其转化成 23×23 的正方形才能作为卷积网络的输入图。而灰度统计直方图的特征维度为 256, 需要将其转化成 16×16 的正方形才能作为卷积网络的输入图。由于不同特征的输入图大小不同, 根据以下方法分别确定卷积神经网络各个特征图的尺寸:

1) 卷积层特征图的边长 L_{map} 大小与输入特征图的边长 L_{input} 符合以下关系,

$$L_{\text{map}} = L_{\text{input}} - L_{\text{con}} + 1, \quad (7)$$

式中 L_{con} 为卷积核的边长;

2) 由于采样层的作用是将 4 个特征融合为一个特征, 所以采样层特征图的边长为卷积层的一半;

3) 为了保证卷积核能够感受到更多的特征, 将前两个卷积核的边长设为最大;

4) 由于输出层是全连接层, 所以采用边长为 2 的卷积核将采样层的输出转化为一个向量。

在网络中, 特征图的像元值(神经元)可计算为

$$\mathbf{X}^{l+1} = f(\mathbf{W}^l * \mathbf{X}^l + \mathbf{b}), \quad (8)$$

式中 \mathbf{W} 为卷积核, \mathbf{b} 为偏置, $l = 1, \dots, 5$ 为当前所在层, $f(\cdot)$ 为激活函数。同一张特征图上的不同神经元共享同一个卷积核, 卷积层所对应的卷积核是可训练的, 而采样层的卷积核则是一个值为 $1/4$ 的常数矩阵, 及平均采样。除第一层外, 卷积层中每一个特征图的卷积核与前一采样层的所有特征图分别相连。

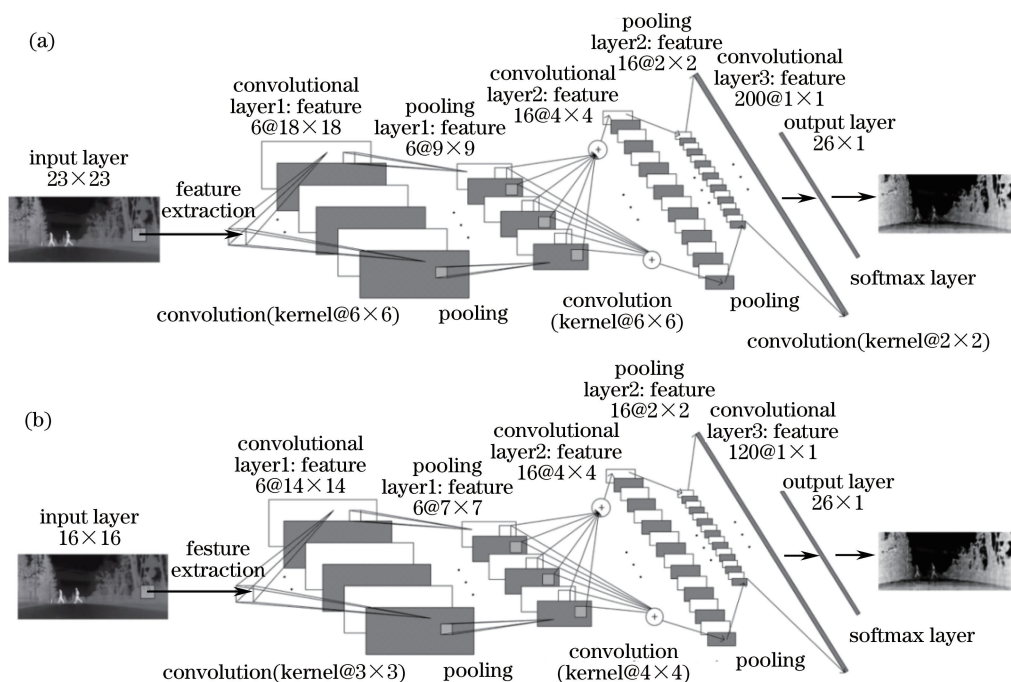


图 2 卷积神经网络结构。(a)纹理信息和灰度值所对应的卷积神经网络结构；(b)灰度直方图所对应的卷积神经网络结构
 Fig. 2 Convolution neural network structure. (a) Convolution neural network structure for texture information and gray value; (b) convolution neural network structure for gray histogram

经过训练,图 3(a)为使用纹理信息训练的卷积神经网络的第一个卷积层的 6 个输出,图 3(b)为第二个卷积层的 16 个输出。可以看出,相比图 3(a),图 3(b)中特征图的边缘清晰,所包含的图像信息更为丰富,且已经有了明显的深度变化,说明了深层的卷积神经网络可以有效地提取深度信息。

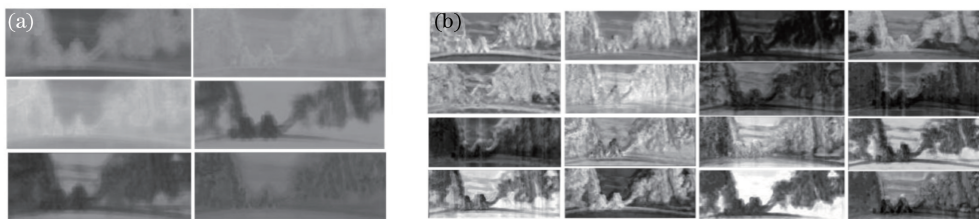


图 3 卷积神经网络隐层的输出。(a)第一个卷积层的输出；(b)第二个卷积层的输出
 Fig. 3 Output of the hidden layer of convolutional neural network. (a) Output of first convolutional layer; (b) output of second convolutional layer

卷积运算不但能够有效地增强信号(图 3),也能够平滑和降低图像中的噪音。图 4 为卷积层 1 的 6 个卷积核和卷积层 2 的部分卷积核。很明显,图 4(b)的卷积核比图 4(a)平滑很多,表明 DCNN 能够通过训练有效地消除图片中的噪声。

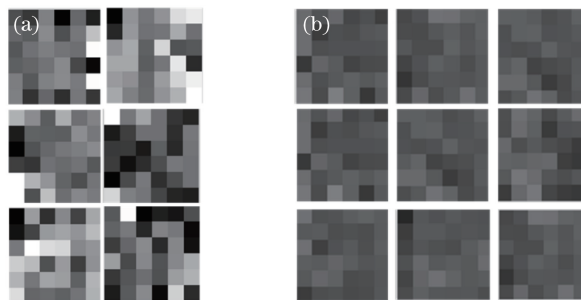


图 4 卷积核。(a)第一个卷积层的卷积核；(b)第二个卷积层的卷积核
 Fig. 4 Convolution kernel. (a) Kernel of first convolutional layer; (b) kernel of second convolutional layer

3.2 参数选择

数据的过拟合是 DCNN 在训练过程中较容易产生一个问题,为了避免这种情况发生,将学习率 η 设定为 0.5,并确定深度标签为

$$\mathbf{d} \left\{ \text{floor} \left[\frac{D_{\text{epth}}(i, j)}{10} \right] + 1, 1 \right\} = 1, \quad (9)$$

式中 \mathbf{d} 为 26 维的列向量,其中第 $\text{floor} \left[\frac{D_{\text{epth}}(i, j)}{10} \right] + 1$ 个值为 1,其余都为 0, $D_{\text{epth}}(i, j)$ 为输入特征所对应的深度图像的深度值, $\text{floor}[\cdot]$ 为向下取整运算。DCNN 的输出采用“1 of N ”编码($N=26$)。卷积网络中所采用的激活函数 sigmoid 函数为

$$f(x) = \frac{1}{[1 + \exp(x)]}, \quad (10)$$

该函数的输出范围为 0~1,为了消除量纲不统一对深度估计的影响,需要对输入输出的数据进行归一化。将输入特征与对应的深度标签组成样本用来训练卷积神经网络。

3.3 网络训练与深度估计

网络是通过深度标签来训练的。在训练的过程中,输出层的误差不断地向前传递,实现对网络参数的更新。误差是以灵敏度的形式传递的,其中灵敏度 δ 是平方误差相对于偏置 b 的变化率。 l 层的灵敏度 δ^l 可以求得,

$$\delta^l = f'(u^l) * \delta^{l+1}, \quad (11)$$

式中 $l=1, \dots, 5$ 为当前网络所在层的序号, $u^l = \mathbf{W}^{l-1} * \mathbf{X}^{l-1} + \mathbf{b}$ 为 $(l-1)$ 层激活函数的输入。与隐层不同,输出层神经元的灵敏度为

$$\delta^5 = f'(u^5) * (y - d), \quad (12)$$

式中 y 为是网络的输出。然后,通过灵敏度来计算卷积核值 \mathbf{W} 的梯度 $\frac{\partial \epsilon}{\partial \mathbf{W}^l}$ 和偏置 b 的梯度 $\frac{\partial \epsilon}{\partial b}$ 为

$$\frac{\partial \epsilon}{\partial \mathbf{W}^l} = \mathbf{X}^{l-1} * (\delta^{l-1})^T, \quad (13)$$

$$\frac{\partial \epsilon}{\partial b} = \frac{\partial \epsilon}{\partial u} \frac{\partial u}{\partial b} = \delta, \quad (14)$$

式中 $\epsilon = \frac{1}{2} (y - d)^2$ 为平方误差代价函数。最后,将得到的梯度与学习率相乘,就可以用来更新每一层的参数。迭代大约 2500 次,得到深度估计模型。

将待测红外图像作为深层神经网络的输入,图像中的每个像元值都会得到一个深度标签 d ,选取深度标签中最大值所对应的下标 h ,并将它转换成深度值,

$$D_{\text{depth}} = (h - 1) \times 10, \quad (15)$$

按上述公式将每个深度标签转化成深度值,并依次排序,就可以得到对应的深度估计图。

4 实验与结果分析

本文算法步骤如下:1) 以一个像元单位为步长,将红外图像分割成若干个大小为 27 pixel \times 27 pixel 的图像块;2) 提取每个图像块中的纹理信息,场景灰度值并统计灰度直方图分别作为红外图像的三种特征;3) 分别用这三种特征训练卷积神经网络得到三个深度估计模型,对比它们的深度估计效果。

实验采用的红外图像为 FLIR 公司提供的车载红外图像,为了对比深度估计的效果,将 Saxena 等^[7]基于马尔科夫随机场对彩色图像进行深度估计的方法用于红外图像的深度估计,并与基于 SVM 的方法^[13]和基于 KPCA 与 BP 神经网络的方法^[14]一起与本文方法进行对比。其中基于马尔科夫随机场的方法^[7]是将提取的基于多尺度的纹理特征,用来训练拉普拉斯模型进行深度估计(称之为拉普拉斯方法)。基于 SVM 的方法^[13]是通过 ICA 对大量的深度特征进行筛选(称之为 SVM 方法),而基于 KPCA 和 BP 神经网络的方法^[14]则是通过 KPCA 与 PCA 相结合进行特征的进一步提取,并训练 BP 网络进行深度估计(称之为 KPCA

与 BP 神经网络方法)。在深度估计的结果对比中,除了给出深度估计图以外,还给出了期望深度与估计深度的样本点的拟合结果。

4.1 不同特征训练的网络的深度估计结果对比

图 5 给出使用灰度特征,灰度统计直方图和纹理信息训练卷积神经网络得到不同的深度估计模型后,对红外图像进行深度估计的对比实验结果。图 5(a)和(b)分别为用来作为训练数据的原始红外图像及其对应的深度图。图 5(c)为采用灰度特征进行训练得到的模型对红外图像进行深度估计的结果,图 5(d)为采用灰度统计直方图特征进行训练得到的模型对红外图像进行深度估计的结果,图 5(e)为采用纹理信息特征进行训练得到的模型对红外图像进行深度估计的结果。从图 5 可以看出:1)使用灰度特征图训练的卷积网络在进行深度估计时,产生了较多的估计错误(两旁的树木出现了较多的白斑块);2)使用灰度统计直方图训练的卷积网络的估计错误较少,但对树枝和人等细节的深度估计效果不佳;3)使用纹理信息训练的卷积网络对细节估计效果很好且估计错误较少。

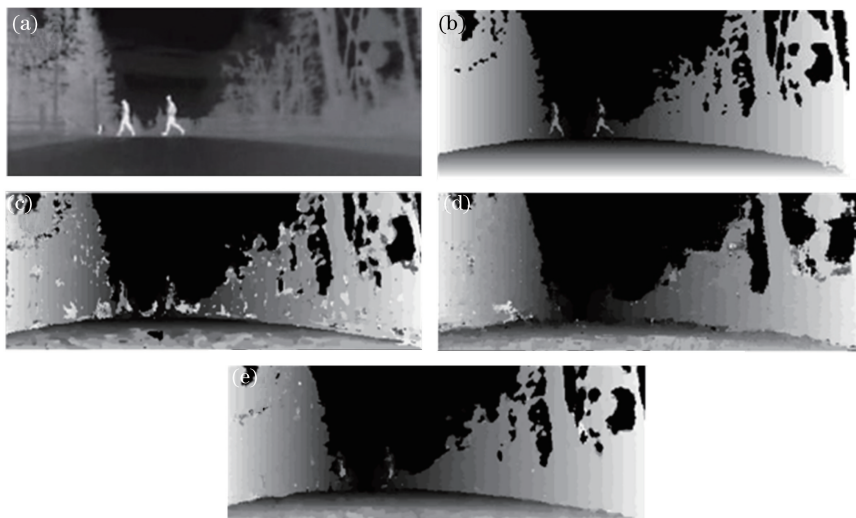


图 5 卷积神经网络模型的深度估计结果。(a)原红外图像;(b)对应的深度图像;(c)使用灰度特征训练网络的估计结果;(d)使用灰度统计直方图训练网络的估计结果;(e)使用纹理信息训练网络的估计结果

Fig. 5 Depth estimation results for the convolutional neural network model. (a) Original infrared image; (b) corresponding depth image; (c) estimation result for network trained by gray feature; (d) estimation result for network trained by gray histogram; (e) estimation result for network trained by texture information

图 6 为从待估计红外图中随机抽取的 100 个样本点的深度估计结果对比。其中,黑色实线为标准深度值,红色点划线为用灰度值训练的卷积神经网络的深度估计值,绿色虚线为用灰度统计直方图训练的网络的深度估计值,蓝色圈划线为用纹理信息训练的网络的深度估计值。从图中可以发现蓝色圈划线与黑色实线拟合的最好,表明用纹理信息训练的卷积神经网络的深度估计值能够与标准深度更加相似。

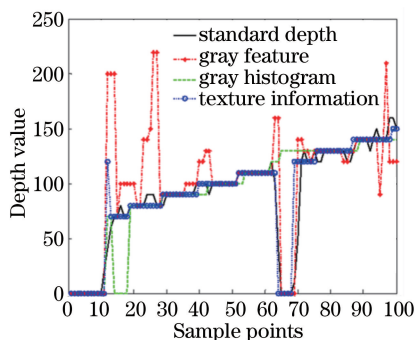


图 6 不同特征训练模型的深度估计结果对比

Fig. 6 Comparison of depth estimation results of model trained by different features

为了更好地反映不同特征的深度估计效果,统计了不同特征训练的深度估计模型的准确率,如表 2 所示。

表 2 不同特征训练模型的深度估计准确率对比

Table 2 Comparison of depth estimation accuracy of model trained by different features

Feature	Accuracy		
	$\zeta < 1.25$	$\zeta < 1.25^2$	$\zeta < 1.25^3$
Gray value	0.523	0.748	0.841
Gray histogram	0.544	0.737	0.869
Texture information	0.624	0.815	0.912

表中准确率的计算方法在彩色图像进行深度估计的相关工作中已经得到应用^[20],其中参数 ζ 为

$$\zeta = \max\left(\frac{\alpha}{\beta}, \frac{\beta}{\alpha}\right), \quad (16)$$

式中 α 为原深度值, β 为估计的深度值。由表 2 可知,用纹理信息训练的深度估计模型能够实现更高的准确率。

4.2 与其他方法的深度估计结果对比

图 7 给出 4 种方法对三幅不同红外图像的深度估计结果比较。其中第一行为三幅不同的街道场景的原始红外图像,第二行到第五行分别给出 SVM 方法^[13]、KPCA 与 BP 神经网络方法^[14]、拉普拉斯方法^[7]和本文方法对三幅红外图像进行深度估计的对比实验结果。可以看出,第二行的 SVM 方法能较好地表现树木、城墙和汽车等景物的轮廓,但并不能有效地区分景物之间的深度,对天空进行深度估计的效果受云的影响较大。第三行的 KPCA 和 BP 神经网络方法对天空的深度估计效果有了明显的改善,且能够与地上景物较好的区分开,但深度估计的准确率仍然较低。第四行的拉普拉斯方法能够有效地分割不同景物并区分它们之间的深度值,但却不能反映树木等景物内部的深度变化,且对于天空的深度估计也不准确。第五行本文方法的深度估计结果不仅能区分不同景物的深度,而且能够反映树木和道路的深度值由近及远的变化过程,如图 7(a)和图 7(b)所示。在图 7(c)中,相比较前三种方法,本文方法能明显表现出近处树木和远处景物的深度差别。

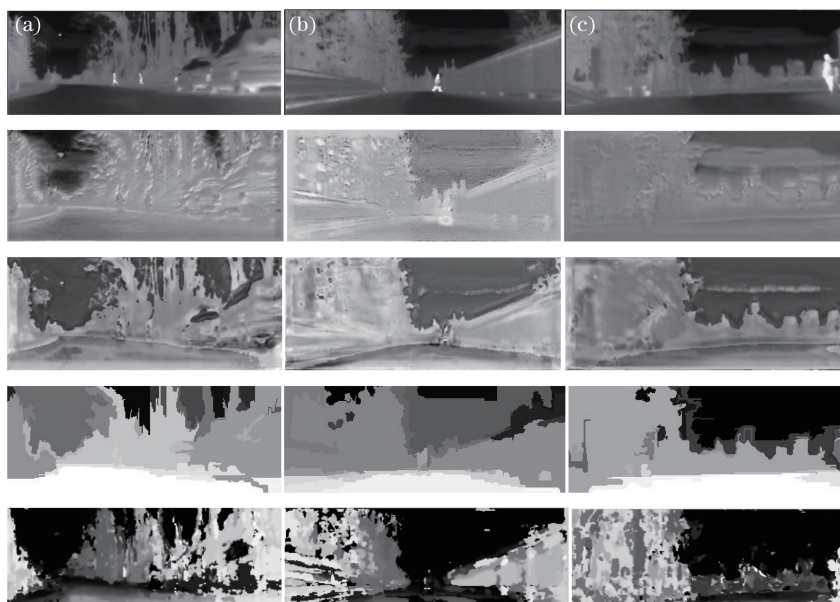


图 7 三种场景的深度估计结果

Fig. 7 Depth estimation results for three scenarios

图 8 为随机抽取的 100 个样本点的深度估计结果对比,为了便于观察,将 SVM 方法^[13]、BP 神经网络方法^[14]和拉普拉斯方法^[7]分别画图与本文方法对比。可以看出,相比较其他三种方法,本文方法能够与原深度值更好地拟合。

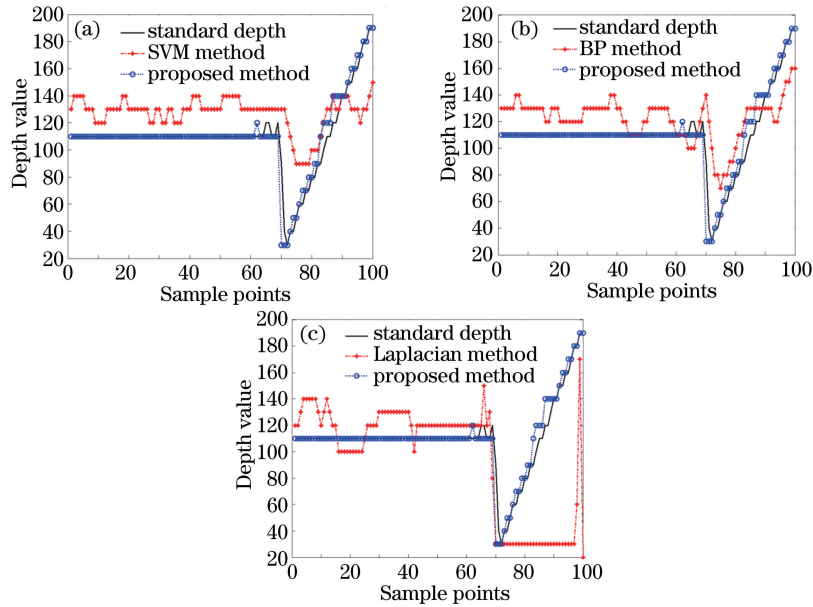


图 8 不同方法的深度估计结果对比。(a)SVM方法与本文方法对比；(b)BP神经网络方法与本文方法对比；
(c)拉普拉斯方法与本文方法对比

Fig. 8 Comparison of depth estimation results with different methods. (a) Comparison of SVM method and the proposed method; (b) comparison of BP method and the proposed method; (c) comparison of Laplacian method and the proposed method

拉普拉斯方法^[7]是针对彩色图像深度估计设计的方法,因此对红外图像的深度估计效果较差。表 3 为对 FLIR 公司提供的车载图像集采用定量的方法,统计了不同方法对红外图像进行深度估计的准确率。

表 3 不同方法的深度估计准确率对比

Table 3 Comparison of depth estimation accuracy with different methods

Method	Accuracy		
	$\zeta < 1.25$	$\zeta < 1.25^2$	$\zeta < 1.25^3$
SVM	0.447	0.658	0.783
BP	0.456	0.671	0.807
Laplacian	0.386	0.578	0.681
Proposed	0.624	0.815	0.912

由表 3 可知,本文方法能够更好地实现对红外图像的深度估计。

5 结 论

提出一种基于深层卷积神经网络的单目红外图像的深度估计模型。采用红外图像中提取的灰度特征、纹理特征、灰度统计直方图特征分别训练深层卷积神经网络,并进行红外测试图像的深度估计。实验结果表明,相比较灰度特征和灰度统计直方图特征,纹理信息更适合用来训练深层卷积神经网络模型。运用纹理信息训练的深层卷积神经网络能够有效地对单目图像进行深度估计,尤其对场景内部的深度变化等细节有更强的分辨能力。

根据单目红外图像深度估计问题的特点,通过劳斯掩膜与梯度检测器提取不同尺度下的纹理信息并构造与之相对应的深层卷积神经网络模型,最后将训练好的深度卷积神经网络模型用于单目红外图像的深度估计。在今后的研究中,将采集更大规模,场景更为复杂的单目红外图像库并进行红外图像的深度估计研究。目前用到的深层神经网络在今后的研究中也进一步深化,并通过分析不同特征层的特点,采用不同的参数模型和训练方法;在网络前向传播的过程中引入反馈,并采用不同的步长设置,通过全连接层代替现有的采样层来进一步提高深度预测的精度。

参 考 文 献

- 1 Saxena A, Chung S H, Ng A Y. 3-D depth reconstruction from a single still image[J]. International Journal of Computer Vision, 2008, 76(1): 53-69.
- 2 Tian Yanbing, Bai Jian, Huang Zhi. Depth estimation with a panoramic stereo imaging system[J]. Acta Optica Sinica, 2013, 33(6): 0611002.
田延冰, 白 剑, 黄 治. 基于全景环带立体成像系统的深度信息估计[J]. 光学学报, 2013, 33(6): 0611002.
- 3 Ding Weili, Ma Pengcheng, Lu Ming, *et al.*. High resolution light field depth reconstruction algorithm based on priori likelihood[J]. Acta Optica Sinica, 2015, 35(7): 0715002.
丁伟利, 马鹏程, 陆 鸣, 等. 基于先验似然的高分辨率光场图像深度重建算法研究[J]. 光学学报, 2015, 35(7): 0715002.
- 4 Jia Songmin, Wang Ke, Li Xiuzhi, *et al.*. Monocular camera three dimension reconstruction based on variation model[J]. Acta Optica Sinica, 2014, 34(4): 0415002.
贾松敏, 王 可, 李秀智, 等. 基于变分模型的单目视觉三维重建方法[J]. 光学学报, 2014, 34(4): 0415002.
- 5 Horn B K P. Obtaining shape from shading information[M]. New York: MIT Press, 1989: 123-171.
- 6 Saxena A, Chung S H, Ng A Y. Learning depth from single monocular images[C]. Advances in Neural Information Processing Systems, 2005: 1161-1168.
- 7 Saxena A, Sun M, Ng A Y. Make 3D: Learning 3D scene structure from a single still image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 824-840.
- 8 Saxena A, Schulte J, Ng A Y. Depth estimation using monocular and stereo cues[C]. International Joint Conference on Artificial Intelligence, 2007: 2197-2203.
- 9 Ding Weili, Li Yong, Wang Wenfeng, *et al.*. Depth estimation of urban road image based on contour understanding[J]. Acta Optica Sinica, 2014, 34(7): 0715001.
丁伟利, 李 勇, 王文锋, 等. 基于轮廓特征理解的城市道路图像深度估计[J]. 光学学报, 2014, 34(7): 0715001.
- 10 Han Bo, Li Yibin. Development of drive assistant in military[J]. Laser & Infrared, 2010, 40(3): 238-240.
韩 波, 李宜斌. 军用车辆辅助驾驶仪的发展[J]. 激光与红外, 2010, 40(3): 238-240.
- 11 Leng Hanbing, Yi Bo, Xie Qingsheng, *et al.*. Adaptive nonuniformity correction for infrared images based on temporal moment matching[J]. Acta Optica Sinica, 2015, 35(4): 0410003.
冷寒冰, 易 波, 谢庆胜, 等. 基于时域矩匹配的红外图像自适应非均匀性校正[J]. 光学学报, 2015, 35(4): 0410003.
- 12 Wu Zhensen, Dou Yuhong. Visible light scattering and infrared radiation of spatial object[J]. Acta Optica Sinica, 2003, 23(10): 1250-1254.
吴振森, 窦玉红. 空间目标的可见光散射与红外辐射[J]. 光学学报, 2003, 23(10): 1250-1254.
- 13 Xi Lin, Sun Shaoyuan, Li Linna, *et al.*. Depth estimation from monocular infrared images based on SVM model[J]. Laser & Infrared, 2012, 42(11): 1311-1315.
席 林, 孙韶媛, 李琳娜, 等. 基于 SVM 模型的单目红外图像深度估计[J]. 激光与红外, 2012, 42(11): 1311-1315.
- 14 Sun Shaoyuan, Li Linna, Zhao Haitao. Depth estimation from monocular vehicle infrared images based on KPCA and BP neural network[J]. Infrared and Laser Engineering, 2013, 42(9): 2348-2352.
孙韶媛, 李琳娜, 赵海涛. 采用 KPCA 和 BP 神经网络的单目车载红外图像深度估计[J]. 红外与激光工程, 2013, 42(9): 2348-2352.
- 15 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems, 2012: 1106-1114.
- 16 Karpathy A, Toderici G, Shetty S, *et al.*. Large-scale video classification with convolutional neural networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1725-1732.
- 17 Liang M, Hu X. Recurrent convolutional neural network for object recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3367-3375.
- 18 Lee S C, Nevatia R. Extraction and integration of window in a 3D building model from ground view images[C]. IEEE Computer Conference on Computer Vision and Pattern Recognition, 2004: 113-120.
- 19 Liu L, Yu G, Zokai S, *et al.*. Multiview geometry for texture mapping 2D images onto 3D range data[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2006, 2: 2293-2300.
- 20 Liu F, Shen C, Lin C. Deep convolutional neural fields for depth estimation from a single image[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5162-5170.
- 21 Li B, Shen C, Dai Y, *et al.*. Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1119-1127.