

文章编号: 1005-5630(2020)02-0039-06

DOI: 10.3969/j.issn.1005-5630.2020.02.007

基于迁移学习的单目菌落深度提取算法

邓相舟, 张荣福

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 高通量菌落挑选仪是生物制药行业中菌种筛选的重要设备, 但其只能识别二维位置信息。为了解决菌落的三维信息提取问题, 提出基于迁移学习的单目成像菌落深度提取算法。该算法以残差网络为基础, 结合多尺度的网络结构提取特征, 采用无监督的迁移学习训练方式, 使网络能够估计菌落深度信息。实验结果表明, 该算法的平均相对误差为 0.171, 均方根误差为 6.198, 对数均方根误差为 0.256, 在 1.25 阈值下的预测准确率提高到了 76.4%, 算法能够同时获取菌落深度信息及其表面特征, 为进一步提高筛选精度和有效挑选菌落提供了参考。

关键词: 菌落挑选; 单目视觉; 深度估计; 迁移学习

中图分类号: TP 391.9 **文献标志码:** A

Monocular colony depth extraction algorithms based on transfer learning

DENG Xiangzhou, ZHANG Rongfu

(School of Optical-Electrical and Computer Engineering, University of
Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: High-throughput colony sorter is an important equipment for bacteria screening in the biopharmaceutical industry. It uses colony image for intelligent identification and selection, but at present, the equipment only recognizes two-dimensional location information. In order to solve the problem of three-dimensional colony information extraction, this paper proposes a monocular image colony depth extraction algorithm based on transfer learning. The algorithm is based on residual network, combined with multi-scale network structure to extract features, and adopts unsupervised transfer learning training mode, so that the network can estimate the colony depth information. The experimental results show that the average relative error of the algorithm is 0.171, the root mean square error is 6.198, and the log root mean square error is 0.256. The accuracy of the results under the threshold value of 1.25 is increased to 76.4%. The algorithm can obtain the depth information and surface characteristics of the colony at the same time, which provides a reference for further improving the screening accuracy and effectively selecting the colony.

Keywords: colony selection; monocular vision; depth estimation; transfer learning

收稿日期: 2019-10-09

作者简介: 邓相舟(1995—), 男, 硕士研究生, 研究方向为计算机视觉。E-mail: ustbdxz2013@outlook.com

通信作者: 张荣福(1971—), 男, 教授, 研究方向为光电检测。E-mail: zhangrongfu@usst.edu.cn

引 言

在现代生物制药行业中，菌落经培养和诱变后将产生数以万计的单克隆菌落样本，其中仅有一小部分是符合要求的高性能菌落。为提高挑选通量^[1]，需借助菌落挑选设备。目前这类设备是通过光学系统成像和图像处理来获得菌落的位置坐标、面积、周长等二维特征，并进行分类、定位和挑选^[2]，但只能获得菌落二维位置信息。

由于培养皿中注入的培养基体积不一，导致菌落距离探针的距离存在差异：如果探针下降太深，针头可能直接触碰培养皿底部而造成培养皿破裂或针头弯曲；但探针下降过浅，会因为无法触及菌落表面而造成菌落漏选。因此对菌落进行深度估计能够有效地解决这些问题，另外，菌落的形态、纹理等信息同样可以通过深度估计来进行提取。

在图像处理中，通常采用双目立体匹配^[3]来进行深度估计，该方法需要经过相机标定、图像匹配、深度计算等过程，存在对设备要求较高、实时计算量大等问题。近年来，随着深度学习在计算机视觉领域的飞速发展，出现了基于卷积神经网络(convolutional neural networks, CNN)^[4]的单目深度估计方法。这类方法舍弃了一定的精度，具有计算量小、泛化性强等优点，同时应用场景也更加普遍。基于 CNN 进行单目图像深度估计的方法分为两种，即监督学习方法和非监督学习方法。在监督学习方法的应用方面：Eigen 等^[5]证明了使用图像和其对应的深度图构建神经网络的有效性；Shelhamer 等^[6]提出的 FCN 网络用卷积层替换了全连接层，在图像分割与深度估计上取得了显著的成果；此外，Laina 等^[7]基于残差网络^[8]的结构对网络进行改进，取得了精度更高的结果。但这类方法需要获取大量的图片所对应的深度真值作为约束，成本较高，且获取大量数据的工作实现起来比较困难。而对于无监督学习方法，只需要对双目图像进行训练，例如 Garg 等^[9]采用的利用双目视差关系对神经网络进行训练的方法。这一类方法对环境、设备条件要求较低，是最贴近实际情况、应用最灵活的方法。菌落挑选仪配备的单目相机用于挑选前对目标菌落进行信息的提取。为了在不改变现有挑选

仪设备硬件结构的情况下，提取菌落的深度信息，同时考虑到挑选所需的实时性要求，本文基于残差网络设计了一种改进的神经网络结构，提出基于迁移学习的单目菌落深度提取算法。

1 单目深度估计的基本原理

常用的立体匹配方法需要通过左、右双目图像匹配特征点来得到深度，根据双目视觉的几何原理，可以得到深度值 z ^[3] 的表达式

$$z = \frac{fB}{x_l - x_r} = \frac{fB}{d} \quad (1)$$

式中： B 为双目相机投影中心的间距； f 为相机焦距； x_l 和 x_r 分别为相同物体在左右两个不同相机中成像的坐标，两者之差即为视差 d 。CNN 可以通过卷积运算，模拟出图像与视差间的关系，从而在预测时使用单幅图像得到视差 d 。菌落挑选仪结构如图 1 所示，挑选仪使用机械臂控制探针的移动，通过单目相机采集菌落图像，图像经计算机处理后获得符合使用单目图像进行深度估计的场景。

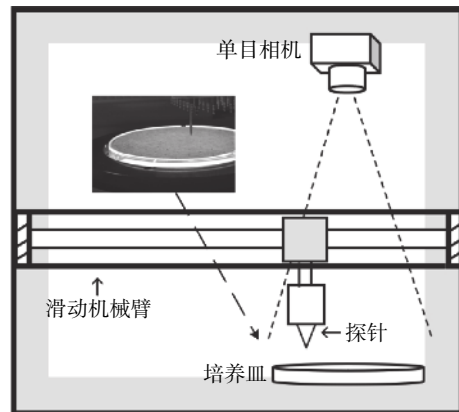


图 1 挑选仪设备结构图

Fig. 1 Structural diagram of selecting equipment

2 菌落深度信息及其位置获取算法

2.1 网络结构

为满足菌落挑选需求，需要搭建更深层的神经网络来获取精度更高的深度信息。因此，引入残差网络 ResNet 作为编码网络的基础网络。残差网络由多个提取特征的残差块(Resblock)组成，残差模块的计算形式为

$$H(x) = F(x) + x \quad (2)$$

式中: x 为模块输入; $H(x)$ 为模块输出; $F(x)$ 为残差函数。残差模块结构如图 2 所示, 网络训练转换为学习残差函数 $F(x)$ 。这样的结构让网络对微小的变化更灵敏, 解决了深层网络的梯度消失、爆炸以及退化问题。通过这种方式可以搭建更深层的神经网络, 加强网络的前向传播, 提取更加丰富的图像信息。

然而深层的网络结构也带来了一些问题, 随着网络层数的增加, 整体的深度预测精度有所提高, 但网络会逐渐丢失浅层的语义信息, 并且残差网络有初期降维较快的特点, 使得这一问题表现得更加严重。这对于提取菌落的表面特征有一些不利的影响, 如果仅仅基于残差网络的最后一层特征进行深度估计, 会导致输出的深度图由于缺少浅层的语义信息, 造成菌落形态特征丢失、边缘轮廓模糊化等问题。为了解决这个问题, 本文采用了多尺度的网络结构进行改进, 将编码网络的浅层特征通过横向连接的方式整合到解码网络中, 达到保留浅层语义信息的效果。解码网络的主体是上采样和卷积层相结合 UpConv 层, 即

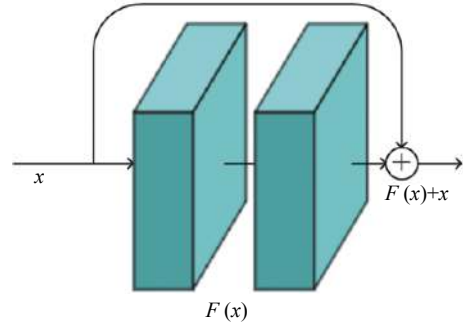


图 2 残差模块结构图解

Fig. 2 Structural diagram of residual module

先进行上采样然后进行卷积运算。上采样的主要目的是用插值的方式放大图像, 然后通过卷积层使图像特征更加的平滑。将残差块和 UpConv 层相同尺度的特征通过横向连接整合, 整体网络结构如图 3 所示, 图中 C 为 Conv 卷积层, R 代表残差网络层, Up 代表 UpConv 层。通过这种组网方式, 使解码网络融合编码网络中浅层的语义信息, 并在最后基于不同尺度进行了深度估计, 使得网络在提升性能的同时保证了菌落表面特征的相对完整性。

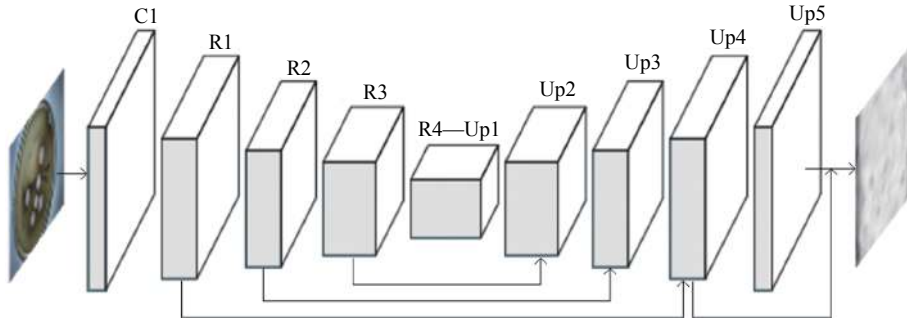


图 3 深度估计网络结构图

Fig. 3 Deep estimation network structure diagram

2.2 损失函数

网络的输入选定为双目图像中的左视图 I_l , 网络的输出为预测的视差 d 。训练时利用输出视差和右视图 I_r , 通过双线性采样的方式重构出新的左视图 \tilde{I}_l , 用重构的左视图 \tilde{I}_l 与原左视图 I_l 的损失作为约束, 将网络训练为输出视差 d 的可靠模型。本文使用 SSIM 损失函数计算左视图 \tilde{I}_l 与原左视图 I_l 的损失, SSIM 损失函数用于度量两幅图像之间的结构相似性, SSIM 损失越大代表着图像的视觉效果更好。SSIM 损失对亮度和色彩的变化可能会更迟钝, 但是可以较好地保持高

频信息。 $L1$ 范数损失函数, 也被称为最小绝对值偏差, 则可以较好地比对颜色亮度特征。因此, 损失函数采用 SSIM 损失^[10] 与 $L1$ 损失的结合, 即

$$E = \frac{1}{N} \sum_{i,j} \alpha (1 - \text{SSIM}(\tilde{I}_{i,j}, I_{i,j})) + (1 - \alpha) \|\tilde{I}_{i,j} - I_{i,j}\| \quad (3)$$

式中: N 为图像像素点个数; α 为比例系数; $I_{i,j}$ 为原左视图 I_l 中坐标点 (i, j) 处的值; $\tilde{I}_{i,j}$ 为重构出的新左视图 \tilde{I}_l 中坐标点 (i, j) 处的值。

2.3 基于迁移学习的网络训练方法

利用挑选仪实验平台来获取实验需要的菌落

数据, 由于真实的菌落样本过于单一, 这样训练出来的网络缺少泛化性, 因此本文选用了模拟菌落来做数据填充, 以提高数据样本的多样性。对相机进行标定以获得相机参数, 并对图像对进行

校正, 最终获得两百对图片作为菌落数据集, 如图 4 所示, 图中显示了三个菌落样本的左视图和右视图。

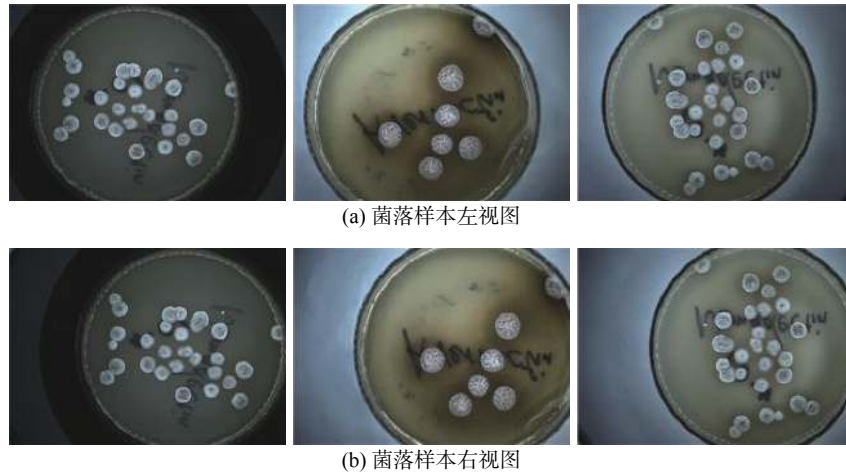


图 4 菌落数据集图像

Fig. 4 Colony dataset images

由于菌落图像数据量较为有限, 直接用这些数据从头训练神经网络效果较差, 还会导致过拟合、网络的泛化性能不足等问题, 所以本文选择使用迁移学习^[11]的方法来训练网络。迁移学习在解决一个问题时, 先将使用到的知识储存起来, 再将其应用到一个具有一定相关性的领域中, 可以用少量的有标签的数据建立一个可靠的模型, 然后对目标领域数据进行预测。这是一个迁移已有的知识来支撑起可靠模型的方法, 是解决目标任务中存在标签的训练数据不足的方法。

网络的训练分为两步来完成。首先用 KITTI 数据集^[12]对网络进行预训练, KITTI 数据集包含大量的经过校正的双目图像对, 实验中挑选出 5 000 对图像, 其中 90% 用作训练集, 10% 用作验证集, 通过训练可以有效地提取图像特征并建立深度估计的模型。下一步使用准备的菌落数据对预训练的网络进行 Fine-tune 训练, 使网络针对菌落数据进行调整, 以此来建立一个针对菌落的可靠的深度估计模型。

3 实验与结果分析

为了验证所用方法的有效性, 我们基于 Python 语言, 结合 Facebook 公司的深度学习框架 PyTorch

搭建了网络。在网络训练过程中, 以图像对的形式进行数据增强, 例如: 以随机概率对图像对进行翻转; 随机改变图像对的亮度、对比度和饱和度, 提高模型的泛化能力与鲁棒性。网络的输入尺寸为 512×256。用 KITTI 数据集训练时, 采用 Adam 损失函数优化器^[13], 初始学习率设置为 1×10^{-4} 。训练设定的 batch size 为 8, 最大迭代次数设定为 100 000 次。以经过 KITTI 数据集预训练的模型作为基础, 使用菌落数据对模型进行 Fine-tune 训练, 此时网络整体已经位于局部最优点附近, 迭代训练较少的次数即可收敛。设定初始学习速率与 batch size 不变, 训练集大小为 200, 训练最大迭代次数为 5 000 次。

为了对网络的性能进行量化评估, 将本文训练的网络与现有的单目深度估计网络在 KITTI 数据集上进行对比, 采用文献^[5, 14]等常用经典的衡量指标对实验结果进行评估, 这些指标的形式如下:

绝对值平均相对误差为

$$\delta_{\text{Rel}} = \frac{1}{T} \sum_i \frac{|d_i - \tilde{d}_i|}{\tilde{d}_i} \quad (4)$$

式中: d_i 为像素点 i 的估计深度值; \tilde{d}_i 为标签深度值; T 为所有测试图像的像素点数目总和。

均方根误差为

$$\delta_{\text{RMS}} = \sqrt{\frac{1}{T} \sum_i \|d_i - \tilde{d}_i\|^2} \quad (5)$$

对数均方根误差为

$$\delta_{\log\text{RMS}} = \sqrt{\frac{1}{T} \sum_i \|\ln(d_i) - \ln(\tilde{d}_i)\|^2} \quad (6)$$

某一阈值 t 下的预测准确率, 即预测结果中满足如下条件的像素占有所有像素的百分比为

$$\delta = \max\left(\frac{d_i}{\tilde{d}_i}, \frac{\tilde{d}_i}{d_i}\right) < t \quad (7)$$

式中: δ 为像素点 i 的估计深度值 d_i 与标签深度

值 \tilde{d}_i 的比值; t 为阈值, 小于该阈值为预测正确, 大于该阈值为预测错误。

各性能指标对比如表 1 所示, 前三列为损失值, 数值越低说明训练结果越好, 后三列为不同阈值 t 下的准确率, t_1 、 t_2 、 t_3 分别表示阈值 t 为 1.25、1.25²、1.25³ 时的结果, 数值越高说明预测结果越准确, 较优的结果加粗标出。从表中可以看出, 本文改进的网络在误差和准确率上均有一定优势, 证明了网络结构的有效性, 网络相对于传统网络模型有更好的结果。

视差图的质量直接反映了算法的性能, 图 5

表 1 基于 KITTI 数据集上的实验结果对比
Tab. 1 Comparison of experimental results on KITTI dataset

方法	δ_{Rel}	δ_{RMS}	$\delta_{\log\text{RMS}}$	t_1	t_2	t_3
Make3D ^[14]	0.282	8.734	0.361	0.601	0.820	0.926
Eigen Coarse	0.214	6.563	0.292	0.673	0.884	0.957
Eigen Fine	0.203	6.307	0.282	0.702	0.890	0.958
Laina	0.201	6.471	0.273	0.680	0.898	0.967
Ours	0.171	6.198	0.256	0.764	0.912	0.968

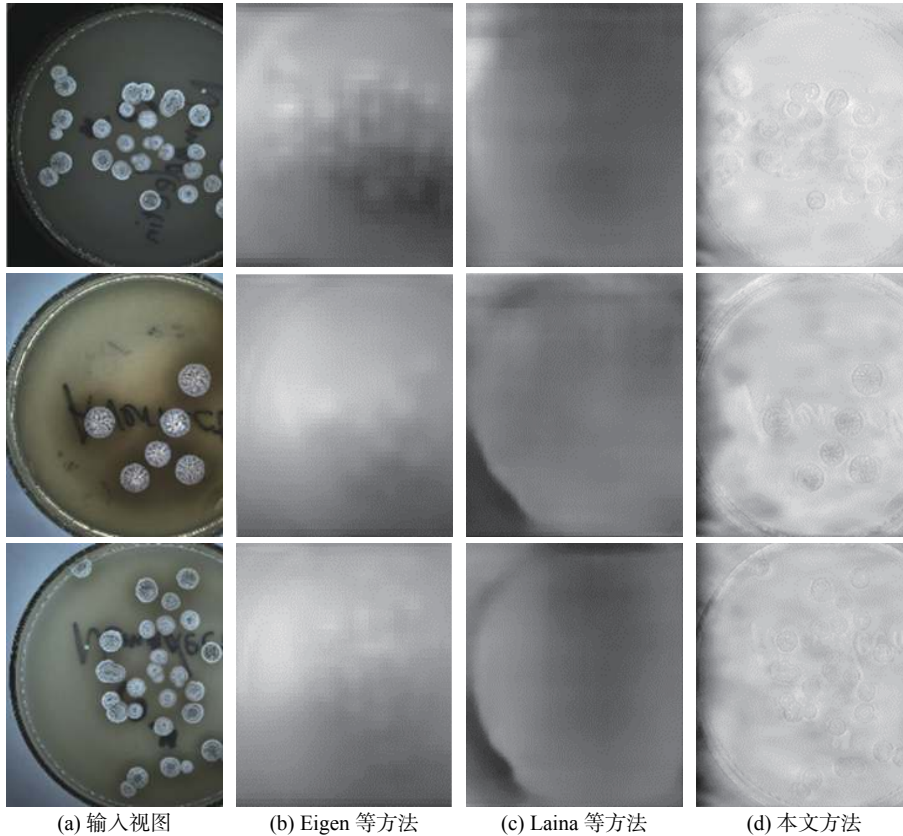


图 5 视差图效果对比

Fig. 5 Comparison of disparity map effects

列出了使用不同方法得到的 3 种菌落图像。图中第二列为 Eigen 等^[5]所用的方法估计的深度图，使用的基础网络为 VGG16，网络相对较浅，从图中可以辨识出菌落的大致位置，但是边缘轮廓比较模糊，算法精度也有待提高。第三列为 Laina 等^[7]所用的方法得到的结果，虽然他们也使用了残差网络获得了更加出色的结果，但是由于深层网络所带来的影响，导致输出的深度图并不理想，从图中无法辨识出菌落的位置、形态等。最后一列为本文算法得到的结果，预测的深度图保留了较为完整的细节特征，图中可以辨识出菌落所在位置，相对以前的单目估计方法在性能上也有所提升，实验结果证实了本文方法的可行性与有效性。

4 结论与展望

本文针对菌落挑选中的菌落深度提取问题，提出了一种基于迁移学习的单目深度恢复方法。该方法基于无监督的深度估计方法和迁移学习的思路，对传统的神经网络结构进行了改进与优化，将编码网络中不同尺度的特征进行整合，在减少误差、提高准确率的同时，保证了菌落表面特征的完整性。实验结果表明，本文的深度提取算法具有可行性并在性能上有所提升。算法没有使用成本较高的真实深度数据进行训练，而是使用比较容易获得的双目图像对，得到了能够预测视差的网络模型，提取了菌落的深度信息并保留了表面特征，对菌落挑选仪的自动化挑选工作有一定帮助。

尽管提出的方法对菌落挑选精度有一定的提高，但由于菌落挑选环境的复杂性和培养皿表面的不确定性，菌落挑选仪的自动化挑选工作还有很多方面可以做，后面会进一步关注多任务网络与菌落三维重建方向的研究。

参考文献：

- [1] 冯向东, 罗玮, 张荣福. 菌落挑选仪挑选通量提升算法研究[J]. 软件导刊, 2017, 16(7): 48 – 51.
- [2] 何睿, 陈戈. 基于细胞自动机的菌落识别的研究[J]. 标准科学, 2012(6): 50 – 53, 62.
- [3] 靳盼盼. 双目立体视觉测距技术研究[D]. 西安: 长安大学, 2014.
- [4] WANG H H, RAJ B. A survey: time travel in deep learning space: an introduction to deep learning models and how deep learning models evolved from the initial ideas[J]. Computer Science, 2015, 226(1–4): 23 – 34.
- [5] EIGEN D, PUHRSCHE C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network[C]//Proceedings of the 27th international conference on neural information processing systems. Montreal, Canada: MIT Press, 2014.
- [6] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640 – 651.
- [7] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks[C]//Proceedings of the 4th international conference on 3D vision (3DV). Stanford: IEEE, 2016.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016.
- [9] GARG R, VIJAY KUMAR B G, CARNEIRO G, et al. Unsupervised CNN for single view depth estimation: geometry to the rescue[C]//Proceedings of the 14th European conference. Cham: Springer, 2016: 740–756.
- [10] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600 – 612.
- [11] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345 – 1359.
- [12] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231 – 1237.
- [13] KINGMA D P, BA J. Adam: a method for stochastic optimization[J]. Computer Science, 2014.
- [14] SAXENA A, SUN M, Ng A Y. Make3D: learning 3D scene structure from a single still image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 824 – 840.

(编辑: 刘铁英)