

# The Method of Feature Point Detection for Image Matching Based on Deep Neural Network

Miaomiao Zhu, Kehua Zhang\*

College of Engineering, Zhejiang Normal University, Jinhua Zhejiang  
Email: zhangkh207@outlook.com

Received: Mar. 9<sup>th</sup>, 2020; accepted: Mar. 20<sup>th</sup>, 2020; published: Mar. 27<sup>th</sup>, 2020

---

## Abstract

In order to achieve accurate image matching in robot recognition navigation, a new method for detecting matching feature points using a new deep neural network as a detector is proposed. Build a deep neural network model based on the data set to implement the mapping from the image to the feature point probability map. Then extract the position of the key feature points through non-maximum suppression of the obtained feature map. Finally, match the detected key feature points accurately to find the most matching image. Experiments show that the method of key feature points detected by deep neural network can make the matching rate high, and further experiments have found that it can also achieve accurate image correction. Compared with the traditional method, the method of deep neural network as a detector has significant advantages.

## Keywords

Deep Neural Network, Image Matching, Detection Feature Points, Key Points

---

# 基于深度神经网络的图像匹配特征点检测方法

朱苗苗, 张克华\*

浙江师范大学工学院, 浙江 金华  
Email: zhangkh207@outlook.com

收稿日期: 2020年3月9日; 录用日期: 2020年3月20日; 发布日期: 2020年3月27日

---

\*通讯作者。

## 摘要

为实现机器人识别导航中图像的准确匹配, 提出一种以新型深度神经网络模型作为检测器进行匹配特征点检测的方法。根据数据集搭建深度神经网络模型实现从图像到特征点概率图的映射; 将得到的特征图通过非极大抑制提取关键点的位置; 最后根据检测到的关键点进行准确匹配, 找到最为匹配的图像。实验表明, 通过该方法检测到的关键点匹配率高, 且通过匹配特征点可以实现准确的图像校正。相比传统的图像匹配方法, 深度神经网络模型作为检测器优势显著。

## 关键词

深度神经网络, 图像匹配, 特征点检测, 关键点

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在科技的快速发展中, 愈发智能化的移动机器人为人类的生产生活带来极大便利。移动机器人由计算机控制和操作, 通过多种传感器实现功能, 其中图像识别和处理是机器人越来越智能化发展的必要技术。图像匹配算法作为图像处理领域的关键技术之一, 在移动机器人的自主导航、图像拼接、单双目视觉等领域都得到广泛的应用。移动机器人的高性能需求对匹配算法的精度、鲁棒性、实时性等都提出了较高要求。

图像匹配是将模板图像与待匹配图像比较, 通过一定的方法进行配准, 从而确定待匹配图像类别并根据匹配点得到图像的相对位移。现有的图像匹配算法主要有基于区域的方法和基于特征的方法。基于区域的方法是依靠相似度度量方法, 在模板图像中通过一定步长的滑动窗口比较图像区域的相似度, 其中相似性度量包括互相关、互信息等。Barnea 等人就两个结构化数据之间的局部相似性提出序列相似性检测算法[1], 可以节省多个数量级的计算时间, 显著提高匹配速度, 但在匹配精度上有所欠缺。Lucas 和 Kanade 使用空间强度梯度信息指导搜索匹配的位置, 提出适用于立体视觉系统的技术[2], 该算法被称为 Lucas-Kanade (LK) 算法, 对于有噪声的图像匹配能力也比较强。而后由 LK 算法扩展和优化的算法在图像匹配领域得到广泛研究。然而, 基于区域的方法由于计算量大且图像受到光照或者重叠区域等影响鲁棒性差, 在应用中存在一定的挑战。基于特征的方法在强度、比例等方面更加稳定, 这是一种根据算子提取图像上的特征点检测关键点之间的距离的方法。传统的特征匹配算法主要以 Harris [3]、Sift [4]、Surf [5] 等算子的检测算法为主。Harris 角点检测算法用以提取像素级的角点, 对尺度变化较为敏感; Sift 是目前图像匹配中最为著名的检测算法, 使用局部梯度提取特征点, 对各种变化都具有强鲁棒性; Surf 是更快速的改进方法, 主要在速度和光照不变性上具有优势。在经典算法基础上发展了多种改进算法, 如 PCA-Sift [6]、SRP-Sift [7]、Harris-Sift [8]、Harris-Laplacian [9] 等, 改进算法的提出不断地在提高匹配精度和速度, 但效果仍有提高空间。

近几年, 随着深度学习的快速发展, 基于深度学习的新方法以其独特的特征学习和特征提取能力在计算机视觉中表现出优于传统方法的性能。Fischer 等人将通过卷积神经网络学习的特征在分类和识别任务上的表现与标准的 Sift 算法作比较[10], 证明了深度学习方法的特征提取能力优异。Melekhov 等人将

神经网络学习图像特征用于图像缺失区域的匹配[11],实验结果证实深度学习方法可以提高特征提取的性能。Ezeobiejesi 更是将深度学习的方法用于指纹的匹配[12],极大地提高了缺失指纹和完整指纹的匹配率。李源熠使用深度学习的常用模型对图像匹配算法进行深度研究[13],探究了不同模型的优缺点。万尧在经典神经网络模型的基础上,还加入了 yolo3 模型的研究[14],对图像进行去误差的精确匹配。综上,许多研究者已经证明了深度学习方法提高了对图像特征点的提取,使图像匹配更加精确。

本文提出一种以新型深度神经网络模型作为检测器进行特征点准确检测与匹配的方法,为图像匹配在移动机器人图像识别上的应用提供新的研究思路。首先根据数据集搭建深度神经网络模型实现从图像到特征点概率图的映射;然后将得到的特征图通过非极大抑制提取关键点的位置;最后根据检测到的关键点进行准确匹配,找到最为匹配的图像。此外,基于深度神经网络的图像匹配特征点检测方法在完成对图像的精确特征点提取后,不仅可以在图像匹配的同时确定图像类别,并且还可以根据匹配的特征点位置实现图像的准确校正。

## 2. 算法实现与模型建立

为实现图像的准确匹配,本文提出以新型深度神经网络模型作为检测器提取特征点的方法。通过训练好的模型,可以在图像上预测出关键的特征点,用以实现图像的准确匹配。

### 2.1. 算法实现

本文以深度神经网络作为检测器提取特征的图像匹配改进算法完成数据集的训练和应用,具体流程如图 1 所示。第一步,生成数据对,其中一张为原始图像  $g$ ,另一张为经过图像变换的生成图像  $g'$ ;第二步,通过深度神经网络从数据对中学习到对应两张特征点概率图( $s$  和  $s'$ );第三步,在特征点概率图中通过非极大抑制的方法提取关键特征点;第四步,对检测到的关键点使用蛮力匹配的方法进行匹配,找到准确的匹配点;第五步,将匹配失败的特征点进行误差反向传播,不断迭代更新网络模型,得到新的特征点概率图;第六步,完成匹配算法后根据对于不同类别图像的匹配率判断图像类别,并且根据图像匹配的特征点位置实现图像的校正。

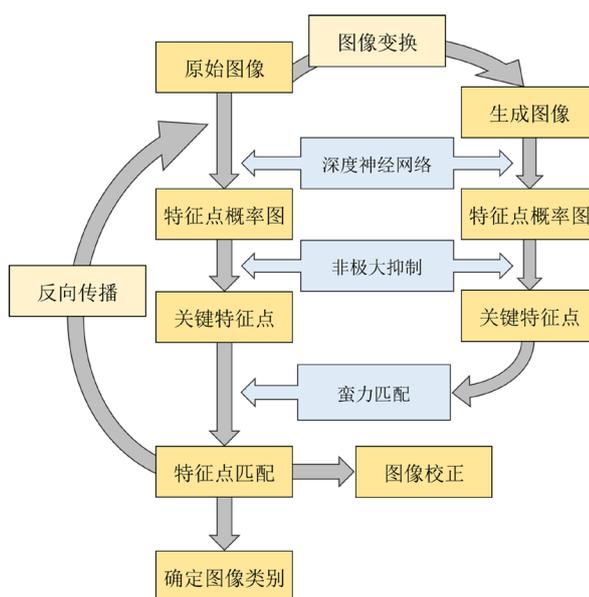


Figure 1. Algorithm implementation flowchart

图 1. 算法实现流程图

深度学习模型在图像匹配算法中主要完成如式(1)的图像对到特征点概率图的映射, 其中  $f$  表示网络的非线性计算。整个网络模型的算法实现是最小化成本函数的过程, 在模型的训练中, 通过不断迭代优化保存成本函数最优的模型参数值。

$$\begin{cases} s = f(g) \\ s' = f(g') \end{cases} \quad (1)$$

在深度学习算法计算得到特征点概率图后, 使用非极大抑制的方法在概率图  $s$  中提取关键特征点, 将其集合记为  $D$ , 再使用蛮力匹配方法进行特征点匹配, 匹配成功的样本点集合记为  $T$ , 最后得到与图像大小相同的关键特征点图像(式(2))与匹配的二分类图像(式(3))。

$$U_{x,y} = \begin{cases} 1 & (x,y) \in D \\ 0 & \text{其它} \end{cases} \quad (2)$$

$$V_{x,y} = \begin{cases} 1 & (x,y) \in T \\ 0 & \text{其它} \end{cases} \quad (3)$$

因此, 损失函数的计算公式可以表示为:

$$Loss = \sum (U_{x,y} - V_{x,y})^2 \quad (4)$$

与传统监督学习的深度学习模型不同, 这里并没有给定训练数据的目标值, 而是采用半监督学习的方式, 在训练过程中, 生成匹配的二分类图像, 以不断减少匹配失败的特征点为目标, 通过式(4)的损失函数进行网络模型的训练。

## 2.2. 数据预处理

图像在形成训练数据时需要经过变换形成数据对, 一张为原始图像, 另一张经过图像变换得到, 图像变换过程是从图像预处理方式中随机选择一种方式生成图像。其中, 图像预处理采用图像增强方法, 如几何变换、噪声变化、对比度变化等, 具体数据预处理参数如表 1 所示。

**Table 1.** Data preprocessing parameters

**表 1.** 数据预处理参数

参数名称	含义	参数取值
gaussian_noise	增加高斯噪声	(0, 1)
translation_horizontal	水平移动像素距离	(-20, 20)
translation_vertical	垂直移动像素距离	(-20, 20)
shearing_x	沿 x 方向裁剪	(-0.2, 0.2)
shearing_y	沿 y 方向裁剪	(-0.2, 0.2)
scaling_x	沿 x 方向放大倍数	(0.7, 1.3)
scaling_y	沿 y 方向放大倍数	(0.7, 1.3)
rotation	旋转角度	(-20, 20)
random_brightness	随机改变亮度	(0, 100)
random_contrast	随机改变对比度	(1, 3)

### 2.3. 模型的设计

模型的架构设计受 U-Net 网络[15]的启发, 采用如图 2 所示的先缩小后扩大的结构。网络模型的训练任务是图像匹配中的特征点提取, 通过网络计算生成特征点概率图。特征点概率图本质上是与原图像相同尺寸的分割网络, 其中特征点与非特征点形成二分类。因此, 在网络的设计中, 与一般的 U-Net 网络不同, 最后一层卷积层采用 Sigmoid 激活函数, 可以将数据压缩至 0~1, 生成像素级概率分布图。整个网络模型通过四次下采样提取图像的高维特征, 然后经过反卷积完成上采样, 恢复与原图像尺寸相同的特征点概率图。在整个过程中, 卷积层全部使用  $3 \times 3$  小卷积核, 并在每次卷积层后加入 Relu 激活函数[16]以及批量归一化操作[17]。为防止过拟合, 在网络中加入了 L2 权重衰减正则化。

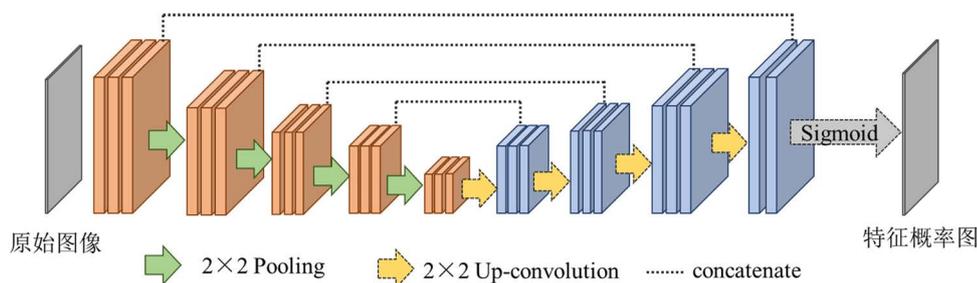


Figure 2. Structure of deep neural network  
图 2. 深度神经网络模型结构

## 3. 实验

### 3.1. 模型超参数

在模型训练过程中, 超参数的设定对模型的结果和性能都有一定的影响。在实验中, 深度神经网络模型的搭建基于 TensorFlow 框架, 采用正态分布初始化模型参数, 训练图像大小均为  $256 \times 256$ , 每批次 8 张图像数据, 总共循环 100 个周期。模型的保存采用每个周期一次验证, 通过最优模型的保存保证模型性能。在误差反向传播过程, 使用自适应矩估计梯度下降的优化算法, 保证快速找到最小优化点。

### 3.2. 实验数据

为验证提出的深度神经网络模型作为检测器提取匹配特征点方法的有效性, 采用寻宝机器人视力定位的图像做实验验证。寻宝机器人在工作过程中, 通过图像匹配判断图像类别即确定宝藏点位置号, 从而完成位置的定位。在此应用中, 需要寻宝机器人对六种不同图形的图像进行匹配识别, 六种不同类别的图形及其所表示的藏宝点位置号如图 3 所示。模型训练所用的数据集是将六种类别的图像做预处理变换。

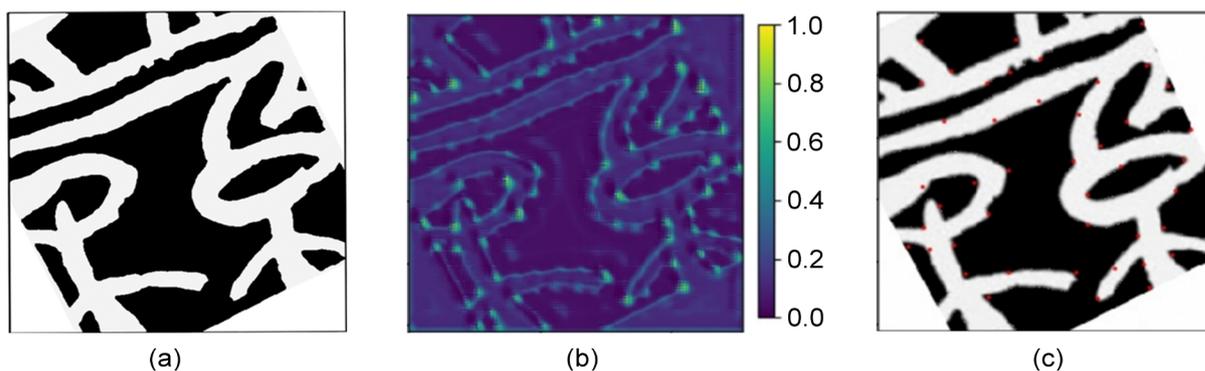
图像						
类别	3	4	5	6	7	8

Figure 3. Template images for six categories  
图 3. 六种类别的模板图像

而在训练过程中, 每个批次训练时数据又将被进一步处理, 随机裁剪选取图像中的部分区域进行提取特征的模型训练, 从而增加数据量, 避免数据重复。在完成模型的训练后, 采用完整的数据进行特征点的提取和与模板图像的匹配, 使用与训练数据不同的图像测试模型的结果, 以验证方法的有效性。在机器人的应用中, 在将图像经过深度神经网络计算得到特征点概率图之后, 还需经过匹配算法判断图像的种类, 从而实现准确的图像匹配功能。

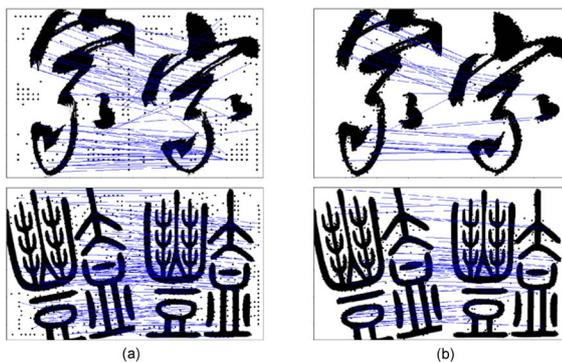
### 3.3. 实验结果分析

在实验中, 以深度神经网络模型作为检测器可以促进图像匹配特征点的检测, 经过训练有素的网络模型计算, 可以以端对端的形式将图像映射得到特征点分布概率图。完整的图像匹配过程还包括非极大抑制算法从概率图中提取关键特征点和蛮力匹配方法对关键特征点进行匹配。图 4 是采用深度神经网络的方法进行图像匹配的过程。图 4(b)是由图 4(a)的原始图像经过网络模型计算得到的特征点概率图, 从中可以看到特征点概率图实质上是由 0 到 1 的概率数值组成的像素级图像。通过神经网络模型的计算, 可以准确提取到图像中每个像素点是特征点的概率, 而概率图上的点经过非极大抑制的方法可以提取关键特征点, 也就是图 4(c)中的红色点表示的图像匹配关键点。将关键点与原图叠加, 可以明显看到提取的关键特征点大多是图像中像素变化较大的边缘点、角点等, 并且关键点的分布比较均匀, 这些特征可以很好的代表图像的特征。



**Figure 4.** Extraction of key feature points using deep neural network: (a) Original image input to the deep neural network model; (b) Feature point probability map calculated by the deep neural network model; (c) Key feature points extracted by non-maximum suppression

**图 4.** 深度神经网络模型提取关键特征点: (a) 输入到深度神经网络模型的原始图像; (b) 经过深度神经网络模型计算得到的特征点概率图; (c) 经过非极大抑制的方法提取的关键特征点



**Figure 5.** Image matching results: (a) Image matching using deep neural network as a detector; (b) Image matching using Sift method

**图 5.** 图像匹配结果: (a) 深度神经网络模型作为检测器的图像匹配; (b) Sift 方法进行图像匹配

为证明基于深度神经网络的特征点提取方法跟传统方法相比具有优势, 将神经网络模型作为检测器提取特征点后进行匹配和直接采用 Sift 特征提取进行匹配的算法做实验对比。图 5 给出了对比结果, 显然, 以神经网络模型作为检测器提取特征点进行匹配的成功匹配点多于直接使用 Sift 进行匹配的算法, 这表明基于神经网络模型先进行特征点的提取, 可以极大程度减少匹配点的丢失, 实现图像的最大化匹配。

准确的特征点提取可以提高图像的匹配率, 为进一步验证基于深度神经网络的特征点检测方法在图像匹配中的作用, 将提出的方法运用于寻宝机器人的视力定位中, 完成图像的匹配分类以及位置点的定位。对 100 张不同尺寸的六种类别图像进行结果验证, 在实现图像匹配的基础上, 通过神经网络模型提取待匹配图像和其余六张模板图像的特征点, 并将提取得到的特征点依次作图像匹配, 匹配点最多的模板图像则为最匹配图像, 以该模板的类别作为待匹配图的类别实现定位。同样将该方法与直接通过 Sift 匹配方法得到的类别作对比, 实验结果表明基于神经网络的方法准确率为 91%, Sift 图像匹配的方法准确率为 85%。

另外, 以神经网络模型作为检测器提取的特征点, 在完成图像匹配功能之后, 也可以实现图像的校正。图 6 展示了通过图像匹配后进行图像校正的两个例子, 可以看到对于图像的旋转、缩放等几何变化, 在确定匹配点位置后可以进行正确的校正。



Figure 6. Image correction results  
图 6. 图像校正结果

#### 4. 结束语

本文提出了一种图像匹配的优化算法, 以神经网络模型作为检测器实现特征点的提取来进行图像匹配算法的优化。模型自主学习从图像到特征点概率图的映射, 将得到的特征图通过非极大抑制提取关键点的位置, 最后通过关键点匹配确定目标图像类别。通过与传统方法的对比, 模型在实验数据上的检测和匹配结果证明了优化方法的优势。在今后的工作中, 将进一步扩展提出的方法, 将优化的图像匹配算法应用到更加复杂的图像匹配和检测中。

#### 参考文献

- [1] Barnea, D.I. and Silverman, H.F. (1972) A Class of Algorithms for Fast Digital Image Registration. *IEEE Transactions on Computers*, **100**, 179-186. <https://doi.org/10.1109/TC.1972.5008923>
- [2] Lucas, B.D. and Kanade, T. (1981) An Iterative Image Registration Technique with an Application to Stereo Vision.
- [3] Harris, C.G. and Stephens, M. (1988) A Combined Corner and Edge Detector. *Alvey Vision Conference*, New York, 10-5244. <https://doi.org/10.5244/C.2.23>
- [4] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [5] Bay, H., Ess, A., Tuytelaars, T., et al. (2008) Speeded-up Robust Features (SURF). *Computer Vision and Image Understanding*, **110**, 346-359. <https://doi.org/10.1016/j.cviu.2007.09.014>

- 
- [6] Ke, Y. and Sukthankar, R. (2004) PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, 27 June-2 July 2004, 2.
- [7] 杨飒, 杨春玲. 基于压缩感知与尺度不变特征变换的图像配准算法[J]. 光学学报, 2014, 34(11): 1110001.
- [8] 冯政寿, 王美清. 基于 Harris 与改进 SIFT 算法的图像匹配算法[J]. 福州大学学报(自然科学版), 2012, 40(2): 37-41.
- [9] Mikolajczyk, K. and Schmid, C. (2001) Indexing Based on Scale Invariant Interest Points. *Proceedings Eighth IEEE International Conference on Computer Vision*, Vancouver, 7-14 July 2001, 525-531. <https://doi.org/10.1109/ICCV.2001.937561>
- [10] Fischer, P., Dosovitskiy, A. and Brox, T. (2014) Descriptor Matching with Convolutional Neural Networks: A Comparison to Sift. arXiv Preprint arXiv:1405.5769.
- [11] Melekhov, I., Kannala, J. and Rahtu, E. (2016) Image Patch Matching Using Convolutional Descriptors with Euclidean Distance. *Asian Conference on Computer Vision*, Taipei, 20-24 November 2016, 638-653. [https://doi.org/10.1007/978-3-319-54526-4\\_46](https://doi.org/10.1007/978-3-319-54526-4_46)
- [12] Ezeobiejese, J. and Bhanu, B. (2018) Patch Based Latent Fingerprint Matching Using Deep Learning. 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 7-10 October 2018, 2017-2021. <https://doi.org/10.1109/ICIP.2018.8451567>
- [13] 李源熠. 基于深度学习的图片匹配算法实现[D]: [硕士学位论文]. 北京: 北京交通大学, 2018.
- [14] 万尧. 基于深度学习的图像匹配算法研究[D]: [硕士学位论文]. 南昌: 南昌大学, 2019.
- [15] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [16] Petersen, P. and Voigtlaender, F. (2018) Optimal Approximation of Piecewise Smooth Functions Using Deep ReLU Neural Networks. *Neural Networks*, **108**, 296-330. <https://doi.org/10.1016/j.neunet.2018.08.019>
- [17] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv Preprint arXiv:1502.03167.