

在智能医疗中的深度估计算法的实现与优化

邓远桥 邹逸轩

自我介绍

老师同学们大家好

我们项目的主题是智能医疗中的深度估计，由于我们组均为大一新生，所以此项目以了解，理解为主，并在此基础上提出一些自己的看法。首先我们需要了解的就是【智能医疗】和【深度估计】的基础概念。何为【智能医疗】？“智能医疗是通过打造健康档案区域医疗信息平台，利用最先进的[物联网技术](#)，实现患者与医务人员、医疗机构、医疗设备之间的互动，逐步达到信息化。”而通过对一些文献的阅读，我们自己的理解是：所谓智能医疗，就是利用当下不断进步的技术（物联网，云计算等技术）与医疗结合，来实现更高效，强力，精准的医疗服务（疾病治疗，就医手续等）。通过继续深入了解发现，现在智能医疗可以分为三个阶段：1.0-医疗信息化——通过计算机、宽带网络等技术实现医院信息共享和区域医疗信息共享；2.0-联网医疗——借助可穿戴设备、4G 网络、云计算、大数据等技术，实现以在线导流、问诊为主要模式的互联网医疗以及医院内部融合医保的全流程移动；3.0 智能医疗——人工智能技术全面融入医疗健康全环节，借助医疗机器人、虚拟现实、增强现实、5G 网络、人工智能等技术，实现人工智能辅助诊断、远程手术等业务模式，实现医疗健康全流程智能化。而当下，正是从联网医疗 2.0 到智能医疗 3.0 的过渡阶段，智能医疗的实现是时代的大趋势。而【深度估计】又是什么呢，简单来说，就是利用一张或唯一、多个视角下的 RGB 图像，估计图像中每个像素相对拍摄源的距离。为什么我们想研究深度估计呢，因为它是一个很实用，能给人们带来便利的东西。例如，在科技发展能通过 CT 等手段拍摄出人体内部图像的今天，一个能做到对图像的准确深度估计很有可能让医生做出更加合理的诊疗方案，甚至对病症部位的准确估计能极大的提高一场手术的成功率与成功度。例如肿瘤切除手术中通过深度估计技术的运用来对肿瘤位置进行准确判断便能让医生做出更合理，有针对性的诊疗方案，减少失败风险从而提高整个手术的成功率。于是乎，我们项目的主要内容就是对深度估计的学习与理解：单目深度估计的研究现状，基于线索或机器学习的传统方法，以及基于无监督的深度学习方法，计算机视觉基础，相关网络模型，常用的损失函数，深度估计的评价方法以及指标等等

算法

1. 监督算法：以输入图像，输出深度图真实值作为训练。主要有回归和分类俩类。

（1）线性回归算法

线性回归算法通常用来构建一个预测模型。例如，根据房子年限、房屋面积这两个特性，来预测房屋价格，就可以构建一个线性回归算法。利用已有的数据训练模型，再用训练好的模型预测新的房价。

如果有一个输入 x ，对应一个输出 y ，就可以构建一个一维线性回归模型，通常表现为一条直线或取现，当给定一个 x 值时，对应可以求出 y 值。当有多个特征共同影响 y 值时，就可以构建一个多元的线性回归模型。

（2）逻辑回归算法

逻辑回归算法用来进行对数据的分类。它和线性回归算法类似，不同之处在于，输出值 y 是给定的几类。最常见的是分成两类，如好和坏。使用逻辑回归算法，可以根据输入数据的特征，判断该条数据的输出是哪一类的。

2. 无监督算法：仅使用俩个摄像机采集的双目图像数据进行联合训练。其中双目数据可彼此预测对方，从而获得相应的视差数据，在根据视差与深度的关系进行演化亦或是将双目图像中各个像素点的对应问题看作是立体匹配问题进行训练。因深度数据的获取难度高，所以大量算法都基于无监督模型。

（1）Kmeans 聚类

聚类是一种非监督学习，它和分类的不同之处在于，分类是有标签的，而聚类是无标签的。分类的结果是知道哪个好哪个坏，而聚类是根据特性，将相似的事物聚集到一起，不考虑它们的好坏。

Kmeans 算法是聚类算法中的一种，可以根据输入的特性，将一些数据聚集为任意多个类别。Kmeans 算法使

用距离的远近来聚集一类数据。

（2）降维

有的时候，一条数据的输入特性可能有很多。比如，在预测房价的问题上，可能会输入房间面积、年限、地理位置相关信息等近百个属性。使用这么多属性进行分析，会为分析过程带来麻烦。降维，就是将这些 D 个维度的输入属性，缩小成 d 个维度的输入属性。它将一些输入特性合并或进行某些操作，来减少变量的数量。

（3）异常检测

异常检测用来判断某一个数据，其输入特性是否异常。例如，有多个属性来描述一个飞机的发动机，在大量的数据输入中，通过异常检测，来发现哪条数据记录存在异常。

3. 基于视频的深度估计：既包含了单帧视频的单目深度估计，也包含了多帧间视频帧的像素的立体匹配，从而近似获取多视角图像，对相机位姿进行估计。

深度学习

深度算法的研究离不开深度学习，特别是深度学习领域的最新进展，对医学图像的识别、分类和量化的发展起到了促进作用。

深度学习正迅速成为一种最先进的学习方式，在各种医学应用场景中表现突出。

在过去几十年中，如 CT, MRI, PET 等医学成像技术已被广泛用于疾病的早期发现, 诊断和治疗。这些解释大多是由医生进行的，因为病理上的差异以及医生的潜在疲劳，所以效率不高。但随着机器学习技术的引进，这种情况被不断改善。

深度学习取得成功离不开以下因素：

1. 高性能的 CPU 和 GPU
2. 大数据的可行性
3. 算法的发展

在大量的样本数量下，深度学习的方法是非常有效的但在医学应用中图像却成了问题，为克服图像过少的问题。研究人员设计了以下几种策略：

1. 将二维 (2D) 或三维 (3D) 图像块 (image patches) 而不是全尺寸图像作为输入，以降低输入维数并减少模型构建所需参数；
2. 通过仿射变换人工生成样本 (即数据增强) 来扩展数据集，然后用增强的数据集从头开始训练它们的网络；
3. 使用在计算机视觉中通过对大量自然图像训练生成的深度模型作为“现成”特征提取器，用目标任务样本训练最终分类器或输出层；
4. 使用来自非医学或自然图像的预训练模型的模型参数初始化模型参数，然后用与任务有关的样本微调网络参数；
5. 通过将全连接层中的权重转换为卷积核，对任意大小的输入使用用小尺寸图像作为输入训练出的模型

我们研究的基于视频深度估计则是研究显微镜的内外参数以及基于显微镜单目图像进行深度估计（模型，

猪眼实验，病人手术图像及视频)

由于双目图像需要利用立体匹配进行像素点对应和视差计算，所以计算复杂度也较高，尤其是对于低纹理场景的匹配效果不好。相较于双目，单目深度估计成本更低也跟容易普及。

单目深度估计即利用一张或唯一视角下的 RGB 图像，估计图像中每个像素相对拍摄源的距离。通过人眼需从图像中提取出大量深度信息的例子得出，单目深度估计不仅需要从二维图像中学会客观的深度信息，还要提取一些经验信息。

完善

我们认为单目深度估计的发展仍将集中在提高精度，可传输性和实时性上

精度：新的函数或网路框架。更有效的几何约束来改进无监督算法

可传输性：同一网络在不同摄像机，场景和数据集的性能。但因为领域或摄像机的不同可能会导致性能的下降

实时性：更快的推理速度，保证实时性的条件下提高精度

思考

‘深度估计’，顾名思义，它是一种估计，只要是估计就一定会存在误差，而我们要做的就是通过手段不断地去缩小这个误差。那么为了实现这个目的就需要找到 RGB 图和深度图像之间存在的映射关系，并以此设计合理的算法。介于刚读大一知识面有限，在三种常见算法方向（基于几何，基于传感器，基于深度学习）中我们选择了在课堂上有所了解的深度学习者一方向来继续深入学习研究。那么是不是可以采用 CNN 来进行各种医学图像的训练，包括放射线医学，病理学，皮肤病学和眼科医学等数据，并通过 BP 反馈来不断调整优化算法函数呢？

深度估计的应用是否能运用到其他领域呢？例如无人驾驶汽车，海底探测等等

参考文献：

1. Deep Learning in Medical Image Analysis--《Annual Review of Biomedical Engineering》
2. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
3. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
4. Laina I, Rupprecht C, Belagiannis V, et al. Deeper depth prediction with fully convolutional residual networks[C]//2016 Fourth international conference on 3D vision (3DV). IEEE, 2016: 239-248.
5. Fu H, Gong M, Wang C, et al. Deep ordinal regression network for monocular depth estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2002-2011.
6. Godard C, Mac Aodha O, Brostow G J. Unsupervised monocular depth estimation with left-right consistency[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 270-279.

7. Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: Learning optical flow with convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2758-2766.
8. Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2462-2470.
9. Mayer N, Ilg E, Hausser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4040-4048.
10. Xie J, Girshick R, Farhadi A. Deep3d: Fully automatic 2d-to-3d video conversion with deep convolutional neural networks[C]//European Conference on Computer Vision. Springer, Cham, 2016: 842-857.
11. Luo Y, Ren J, Lin M, et al. Single View Stereo Matching[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
12. Zhou T, Brown M, Snavely N, et al. Unsupervised learning of depth and ego-motion from video[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1851-1858.
13. Yin Z, Shi J. Geonet: Unsupervised learning of dense depth, optical flow and camera pose[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1983-1992.
14. Zhan H, Garg R, Saroj Weerasekera C, et al. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 340-349.
15. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
16. Radford A , Metz L , Chintala S . Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer Science, 2015.