

影像分割中Dice系数和Hausdorff距离的比较

何奕松, 蒋家良, 余行, 傅玉川
四川大学华西医院放疗科, 四川 成都 610041

【摘要】目的:比较影像分割技术中Dice系数和Hausdorff距离的评价效果,分析这两种相似性系数之间的联系,并设计相应的类型加以进一步论述。**方法:**采用Dice系数和Hausdorff距离对两种轮廓相似度进行评估。设计18个(9对)从临床靶区中抽象出的轮廓,计算并比较这两类相似性系数。**结果:**根据比较结果,总结出3种不同类型情况:(1)图像符合度好;(2)图像整体符合度好,但存在一小部分符合度差;(3)图像轮廓符合度差。**结论:**仅靠Dice系数作为唯一评判标准是不完善的,应在此基础上增加Hausdorff距离的筛选,这样才能最大限度地说明轮廓之间的符合程度。

【关键词】精确放疗;靶区勾画;轮廓相似性;Dice系数;Hausdorff距离

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2019)11-1307-05

Comparison of Dice similarity coefficient and Hausdorff distance in image segmentation

HE Yisong, JIANG Jialiang, YU Hang, FU Yuchuan

Department of Radiotherapy, West China Hospital, Sichuan University, Chengdu 610041, China

Abstract: Objective To compare the evaluation effects of Dice similarity coefficient and Hausdorff distance in image segmentation; analyze the relationships between two similarity coefficients; and design the corresponding contours for further discussion. **Methods** Dice similarity coefficient and Hausdorff distance were used to evaluate contour similarity. Moreover, 18 (9 pairs) virtual contours abstracted from clinical target areas were designed for calculating and comparing the two similarity coefficients, namely Dice similarity coefficient and Hausdorff distance. **Results** According to the comparison results, 3 different types of cases were summarized. (1) The image conformity was preferable; (2) the overall conformity of the images was good, but there was a small proportion of poor conformity; (3) the conformity of image contour was poor. **Conclusion** Using Dice similarity coefficient as the only criterion is imperfect, and Dice similarity coefficient should be combined with Hausdorff distance to reflect contour conformity to the greatest extent.

Keywords: precision radiotherapy; segmentation of target area; contour similarity; Dice similarity coefficient; Hausdorff distance

前言

在放疗计划的制定过程中,靶区勾画需要花费医生大量的时间和精力。随着人工智能技术的发展,通过计算机深度学习(卷积神经网络^[1-2]、全卷积神经网络^[3])来完成对靶区和危及器官的自动勾画已成为可能^[4],但在此过程中存在勾画的准确性评判问题。通常使用Dice系数(Dice Similarity Coefficient, DSC),即重叠比较法,评估两个轮廓之间相似程度,

一般认为DSC>0.7表示自动分割区域和手动分割区域两者重复度高,分割效果好^[5];另一种评估方法是Hausdorff距离(Hausdorff Distance, HD),它是一种对位置信息差异很敏感的参数,如果仅有一小部分偏离较大都会导致HD陡然增大,其值越小意味着重复度越高,但也和目标区域大小有关,越大的区域其HD值往往越高,比如,文献[6]中股骨头与膀胱的平均HD值之间的差异。图1~2分别为鼻咽癌CTV2^[7]自动勾画和手动勾画的图像实例,下方有其相应的DSC值和HD值,从其中很难看出这两种方法之间的联系和异同,因此有必要进行基于抽象轮廓的系统比较和分析。本研究设计了18个(9对)从临床靶区中抽象出的轮廓,系统计算和比较这两种相似性系数,从直观上说明DSC和HD这两种表达系数的相互联系和互补性。

【收稿日期】2019-06-10

【作者简介】何奕松,在读硕士,研究方向:医学物理,E-mail:627412679-@qq.com

【通信作者】傅玉川,博士,主任技师,研究方向:医学物理,E-mail:ychfu@hotmail.com

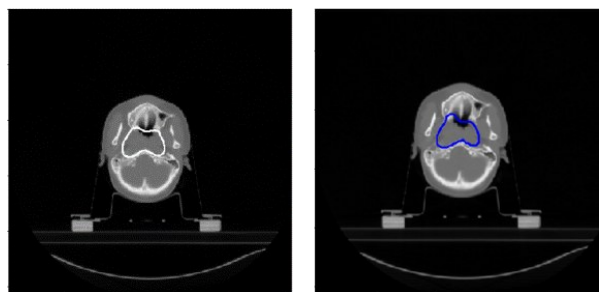


图1 较大偏差图像示例

Fig.1 Example of images with large deviations

左图为自动勾画图像,右图为手动勾画图像,两者的DSC为0.888 9, HD为8.000

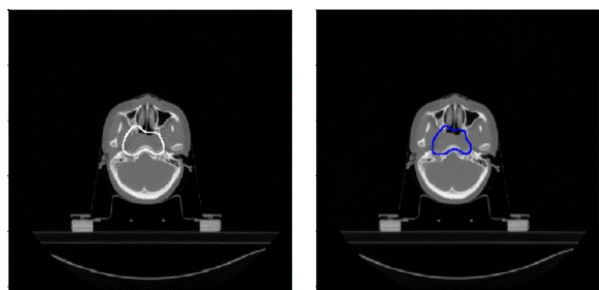


图2 小偏差图像示例

Fig.2 Example of images with small deviations

左图为自动勾画图像,右图为手动勾画图像,两者的DSC为0.928 8, HD为5.013

1 两种轮廓相似度评估方法

1.1 DSC

DSC通过估计两个轮廓的相互重叠部分体积(面积)与其总体积(面积)的百分比来描述它们之间的相似程度。计算公式为:

$$DSC(A, B) = \frac{2(A \cap B)}{A + B} \quad (1)$$

其中, A 和 B 分别表示两个目标区域; \cap 表示取二者共同的部分; $+$ 表示取二者之和。这种计算方法是运用得最多的比较方法^[8]。实际上,DSC是Kappa系数中的一个特例^[9],在极限情况下,Kappa系数的计算公式可以被简成现在的DSC计算公式。早在1945年,Dice^[10]就在研究生态学上不同物种之间数量时提到这个参数。

通过式(1)计算得到的DSC是一个取值范围在0到1之间的参数。数值为1的时候表示两个目标区域完全重合,数值为0则表示两个目标区域完全分离。为了拓宽DSC的值域(使其值不局限在0到1之间),可采用它的一种变形式将它的取值范围拓宽到整个实数域,这就是Logit转换^[11]:

$$\text{Logit}(DSC) = \ln\left(\frac{DSC}{1 - DSC}\right) \quad (2)$$

这是一种单调的转换,转换后的Logit(DSC)的取值范围是整个实数域,并且具有单调性。当Logit(DSC)=0的时候,相应的DSC为0.5,这样就能在大量相似度相近的数据中更加明显地看出其不同之处。

在描述空间重叠度时,Zijdenbos等^[9]认为一个好的空间重叠,对应的DSC应该大于0.700 0,相应的Logit(DSC)为0.847 3,这样我们观察的范围就可以从0.700 0到1.000 0变成0.847 3到正无穷了。

DSC对位置和大小差异性都很敏感,例如,大小相等的两个图形,把它们的一半区域重叠,彼此只有一半面积,导致DSC=0.5;或者一个区域完全覆盖在另一个比它小一半上的区域上,DSC就只有三分之二。在DSC测量图像相似性的过程中,位置的差异相比大小的差异更能反映出不同,重合相同体积(面积)的图像,其所在位置不同会导致最后的DSC不同。这样能够反映出直观的感觉,两个区域(其中一个完全包含另一个区域)比两个部分重叠区域更加相似。结合实际轮廓图像处理来看,虽然DSC在计算过程中具有简洁明了的优点,但它却不能描述所有的情况。

1.2 HD

距离比较法HD是受到几何建模、计算几何和计算机图形学研究领域学者关注的相似性度量参数。临床医学上CT图像的轮廓相似性比较实际上是比较CT图像上一系列坐标点围成的轮廓之间的相似性。当需要比较两个集合之间的相似度的时候,HD是对分割不好区域非常敏感的参数。HD作为一种相似性度量^[12],在图像匹配领域用来衡量二值图像之间的相似度,与以往大多数图像匹配度量准则不同的是这种度量不强调图像中的匹配点一一对应,点与点之间的关系是模糊的,同时距离也满足一些性质,如同一致性、对称性和三角不等性,其优点是这种度量本身具有很强的抗干扰能力和容错能力^[13]。使用HD时,由于不需要考虑被测点集之间点和点的对应关系,因此还可以有效地解决图像当中存在噪声和目标被部分遮挡时的识别问题^[14]。

与图像处理领域中距离的计算和应用研究形成鲜明对比的是针对连续几何对象间的距离的计算研究相对较少,截止目前,只有少数学者对自由曲线或曲面间的精确距离进行研究^[15-18]。CT图像轮廓的相似性比较实际上是两组坐标点连成的图像之间的比较,即两个不同的点集之间的比较。在二维平面中,HD是指一个集合到另一个集合之间最近点的所有距离的最大值。给定两个有限点集 $A = \{a_1, \dots, a_p\}$ 和 $B = \{b_1, \dots, b_q\}$,

它们之间的HD被定义为:

$$H(A, B) = \max \{ h(A, B), h(B, A) \} \tag{3}$$

其中, $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$; $h(B, A) = \max_{b \in B} \min_{a \in A} \|b - a\|$ 。这里的 $\| \ \|$ 是一种定义在点集A和点集B上的距离范数^[19], 含义是两个点之间的欧几里得距离, 简称欧式距离。 $h(A, B)$ 被称为A、B之间的间接HD, 这表示一个属于A的点a比任何其他属于B中的点的距离都要大于它到b的距离, 得到一系列这样的距离, 它们的最大值就是之前提到过的 $h(A, B)$ 。直观上来说, 假如 $h(A, B)=d$, 那么A中的每一个点到B中的每一个点的距离都不大于d。 $H(A, B)$ 是 $h(A, B)$ 和 $h(B, A)$ 中较大的一个。HD是一种非线性算子^[20], 能够测量两个集合之间的不匹配程度。换句话说, 这样的距离决定了模型和物体之间不匹配的程度。因此, 它可以通过比较两个重叠的图像来进行对象识别^[21]。

对HD来说, 不像DSC一样有着某个确定的值象征着重合度好或者坏的区分界限, 但正常情况下, 如果图形重合的很好, 其HD会很小, 倘若存在某一部分偏差较大, 那么HD会立刻飙升至几十甚至上百, 这足以用于评价其重合程度。只要HD不是异常大, 都可以认为重合度是很好的, 此处将20作为评判的标准。

2 材料与方法

2.1 材料

为了让结果看起来更加简单明了, 这里没有选用常规的CT勾画图像, 而是设计了一些由这些CT勾画图像抽象出来、线条单一、简洁、通过画图软件画出来的轮廓图像。图像分别有3组, 每组包含6幅图像, 总计18个轮廓(9对)。图3为其中的一对轮廓。

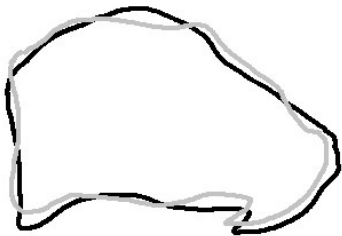


图3 轮廓比较示意图
Fig.3 Comparison of segmented contours

浅色轮廓和深色轮廓分别模拟自动勾画和手动勾画的轮廓, 手动勾画的轮廓往往表面更紧凑、更光滑, 自动勾画的轮廓则没有那么好的效果

2.2 方法

通过MATLAB提取出图像轮廓的坐标点, 将其

导入预先编辑好用来进行计算DSC和HD的Python脚本, 最后输出该幅图像中对应轮廓之间的DSC和HD。所有9对轮廓处理完毕后, 将其结果与图像进行统一比较。判断参数好坏的标准目前并没有一个准确的值, 由于HD本身的敏感性大, 再加上与目标区域的体积有关^[6], 所以很难得到一个绝对的标准。为了能更加直观地反映出本文所述的情况, 在模拟图像时, 放大了整个图像, 相应地HD的值也会提高。对本次研究设计模拟的3种类型的图像, 选定DSC与HD的边界值分别为0.700 0和20。

3 结果

实验中每幅图像都对应了两个输出参数, 分别为DSC和HD, 图4~6分别显示了3种典型的轮廓类别比较, 所有18个轮廓的输出结果如表1所示(表内数据保留4位有效数字)。由图表可知图像大致可以被分成3种不同的类别: DSC大于0.700 0且HD不超过20, DSC大于0.700 0但HD超过20, DSC小于0.700 0且HD超过20。

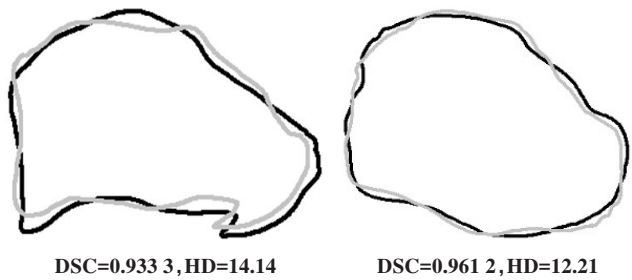


图4 DSC好且HD也好的图像示例(类别1)
Fig.4 Example of contours with good DSC and good HD (category 1)

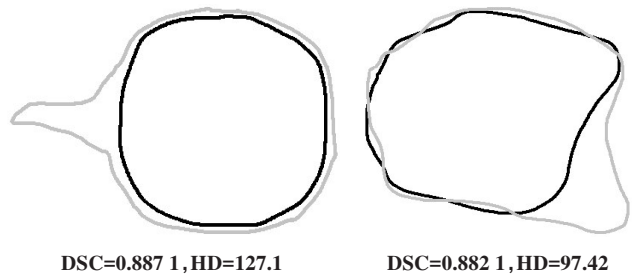


图5 DSC好但HD过高的图像示例(类别2)
Fig.5 Example of contours with good DSC and greater HD (category 2)

4 讨论

如图4所示的类别1中, 两个轮廓之间的DSC都达到并超过了0.900 0, 接近于1, 远超过之前设定的评判标准0.700 0, 说明轮廓之间有着非常好的形状相似性。同时, 它们的HD分别为14.14和12.21, 均小于评判标准20, 说明轮廓之间也有着非常好的距

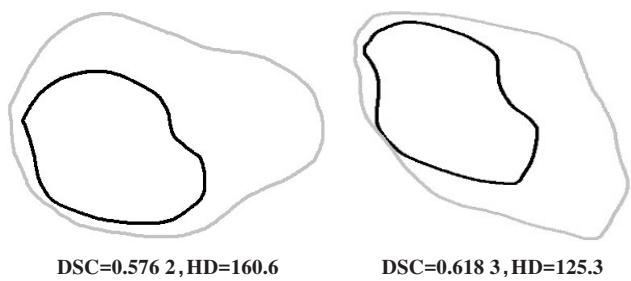


图 6 DSC 低且 HD 过高的图像示例(类别 3)
Fig.6 Example of contours with lower DSC and greater HD (category 3)

表 1 3 种典型类别轮廓相似性参数的比较
Tab.1 Comparison of contour similarity parameters of 3 typical categories

编号	类别 1		类别 2		类别 3	
	DSC	HD	DSC	HD	DSC	HD
1	0.933 3	14.14	0.887 1	127.10	0.576 2	160.60
2	0.961 2	12.21	0.882 1	97.42	0.618 3	125.30
3	0.962 6	11.05	0.917 8	106.60	0.538 4	110.80
4	0.942 9	17.72	0.877 7	61.00	0.595 8	99.32
5	0.962 5	10.05	0.932 2	63.29	0.529 8	157.80
6	0.956 5	11.70	0.885 7	72.20	0.581 0	110.70
平均值	0.953 1	12.81	0.897 1	87.94	0.573 3	127.40

离相似性。假设在临床的靶区勾画中,自动勾画结果和手动勾画结果满足这个条件,则可以认为两者具有很好的相似性。

图 5 的类别 2 在临床实践中不常见,但也是不可忽略的一类情况。两个轮廓之间的 DSC 都超过评判标准 0.700 0,并达到 0.887 1、0.882 1,说明轮廓之间也有着非常好的形状相似性,但反观其 HD 却不尽如人意,二者的 HD 值为 127.1 和 97.42,远远大于预设值 20,说明它们之间的距离相似性很差。综合两个参数来看,虽然 DSC 展示出很好的相似性,但 HD 却截然相反,说明存在这样一种情况,即轮廓在整体上的符合度很好,但存在某一小部偏差很大,导致 HD 变得很大,在临床实践中是存在这种情况的,比如机器在识别靶区或器官上存在一些误差,或者是医生有意增减受到照射的区域。此类情况说明不能把 DSC 作为判断轮廓相似性的唯一衡量参量。

一般会被排除的情况如图 6 所示,两个轮廓之间的 DSC 和 HD 分别为 0.576 2、160.6 和 0.618 3、125.3。无论从 DSC 还是 HD 的角度上看,都是偏离评价标准很大的值。除此之外,不存在 DSC 不好但 HD 很好的情况,一旦 DSC 很小,说明有很大一部分区域偏离

大,此时对应 HD 也应该很大。
本研究结果表明,在影像分割技术中,仅靠 DSC 系数作为唯一的轮廓相似度评判标准是不完善的,应在此基础上增加对距离敏感的参数 HD 的筛选,这样才能更加全面准确地评估影像分割轮廓的相似性。

【参考文献】

[1] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016: 770-778.

[2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Comput Sci, 2014, arXiv: 1409.1556v6.

[3] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, 2015: 3431-3440.

[4] KOSMIN M, LEDSAM J, ROMERA-PAREDES B, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer [J]. Radiother Oncol, 2019, DOI: 10.1016/j.radonc.2019.03.004.

[5] ANDREWS S, HAMARNEH G. Multi-region probabilistic Dice similarity coefficient using the Aitchison distance and bipartite graph matching [J/OL]. Comput Sci, 2015. <https://arxiv.org/pdf/1509.07244v3.pdf>.

[6] 王金媛,徐寿平,刘博,等.定量评价Atlas模板库病例数目对宫颈癌危及器官自动勾画的影响[J].中国医学物理学杂志,2019,36(7): 760-764.

WANG J Y, XU S P, LIU B, et al. Quantitative evaluation of Atlas-based auto-segmentation of organs-at-risk in patients with cervical cancer using different Atlas database sizes [J]. Chinese Journal of Medical Physics, 2019, 36(7): 760-764.

[7] 吴昕,刘磊,肖江洪,等.鼻咽癌靶区的自动勾画评价[J].四川医学,2015,36(6): 762-766.

WU X, LIU L, XIAO H H, et al. Evaluation of auto-segmentation of targets areas in nasopharyngeal carcinoma [J]. Sichuan Medical Journal, 2015, 36(6): 762-766.

[8] SHARP G, FRITSCHER K D, PEKAR V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy [J]. Med Phys, 2014, 41(5): 4871620.

[9] ZIJDENBOS A P, DAWANT B M, MARGOLIN R A, et al. Morphometric analysis of white matter lesions in MR images: method and validation [J]. IEEE Trans Med Imaging, 1994, 13(4): 716-724.

[10] DICE L R. Measures of the amount of ecologic association between species [J]. Ecology, 1945, 26: 297-302.

[11] ZOU K H, WARFIELD S K, BHARATHA A, et al. Statistical validation of image segmentation quality based on a spatial overlap index [J]. Acad Radiol, 2004, 11(2): 178-189.

[12] HUTTENLOCHER D P, KLANDERMAN G A, RUCKLIDGE W A. Comparing images using the Hausdorff distance [J]. IEEE Trans Pattern Anal, 1993, 15(9): 850-863.

[13] 曹京京. Hausdorff 距离的计算原理及其在二维匹配中的应用 [D]. 大连: 大连理工大学, 2013.

CAO J J. The calculation therapy of Hausdorff distance and its application to the matching of 2D geometrical objects [D]. Dalian: Dalian University of Technology, 2013.

- [14] 彭晓明, 丁明跃, 周成平, 等. 一种利用 Hausdorff 距离的高效目标搜索算法[J]. 中国图象图形学报, 2004, 9(1): 25-30.
PENG X M, DING M Y, ZHOU C P, et al. A highly efficient approach to locating objects using the Hausdorff distance[J]. Journal of Image and Graphics, 2004, 9(1): 25-30.
- [15] ALT H, SCHARF L. Computing the Hausdorff distance between curved objects[J]. Int J Comput Geom Appl, 2008, 18(4): 304-320.
- [16] ELBER G, GRANDINE T A. Hausdorff distance and minimal distances between parametric freeforms in R² and R³ [C]. International Conference on Geometric Modeling and Processing, 2008: 191-204.
- [17] CHEN X D, MA W Y, XU G, et al. Computing the Hausdorff distance between two B-spline curves[J]. Comput Aided Des, 2010, 42(12): 1197-1206.
- [18] KIM Y J, OH Y T, YOON S H, et al. Precise Hausdorff distance computation for planar freeform curves using biarcs and depth buffer [J]. Visual Comput, 2010, 26(6): 1007-1016.
- [19] 邱志敏, 李军, 葛军, 等. 基于 Hausdorff 距离的自动目标识别算法的研究[J]. 红外技术, 2006, 28(4): 199-202.
QIU Z M, LI J, GE J, et al. An automatic target recognition algorithm based on Hausdorff distance[J]. Infrared Technology, 2006, 28(4): 199-202.
- [20] HUTTENLOCHER D P, KLANDERMAN G A, RUCKLIDGE W J. Comparing images using the Hausdorff distance[J]. IEEE Trans Pattern Anal Machine Intell, 1993, 15(9): 850-863.
- [21] YOU J, PISSALOUX E, HELLEC J L, et al. A guided image matching approach using Hausdorff distance with interesting points detection [C]. Proceedings of 1st International Conference on Image Processing, 1994: 968-972.

(编辑:谭斯允)