**1. Introduction**

基于内容的图像检索（CBIR）作为检索视觉数据最有效的方式之一，在各个领域都有广泛的应用。CBIR大致检索流程是通过提取图像的低级特征，如颜色，纹理，形状等或更具语义性的高级特征来描述图像，根据这些特征从图像数据库中搜索内容[1]。随着信息技术的高速发展以及大数据时代的到来，图像不仅数量多，而且所含也更加丰富，传统的内容图像检索由于缺少高层语义，已经无法满足用户的需求。

随着医院数字化影像数据的广泛应用，世界范围内的医学影像越来越多地被采集，传输和存储，从而导致医院存储的医疗图像数量大幅度增加。一个基于内容的医学图像检索（CBMIR）的系统可以有效的辅助各种诊断和治疗。医学图像不同于其他图像，具有灰度分辨率高、图像相似性大等特性，医学领域的CBIR系统，图像之间的细微差别往往是最重要的，因此，能够提取不变性特征（关于任何变换）的医学图像检索（CBMIR）系统是有价值的[2, 3]。在2012之前，基于内容的图像检索领域的研究重点在图像的手工视觉特征提取，其中的代表是局部二值图（LBP）描述符[4]和SIFT特征[5, 6]。基于LBP 特征也衍生出来许多变种，比如 LMP[7]和LBDP[8]等。Murala和Wu等人[7]提出了一种对一个给定的参考像素与其直接邻居像素之间的关系进行编码的局部网格图（LMP）描述符。Dubey等人[8]通过使用中心像素强度和局部位平面变换值构造了局部位平面解码模式（LBDP）算法的中心像素局部特征描述子。基于SIFT特征的方法大多依赖于BoW模型。BoW模型最初是为解决文档建模问题而提出的，因为文本本身就是由单词组成的。它通过累加单词响应到一个全局向量来给文档建立单词直方图。在图像领域，尺度不变（SIFT）特征[9]的引入使得BoW模型变得可行[10], 这在2003年后的近10年时间里，BoW模型给图像检索任务带来了各种提升。在2012年，Krizhevsky[11]等人使用的AlexNet神经网络模型在ILSRVC 2012上取得最高识别准确率，后面出现大量的深度神经网络模型，其中具有代表性的有VGGNet[12]，GoogleNet[13]以及ResNet[14]。机器学习，特别是深度学习成为了计算机视觉领域的研究重点。使用深度学习提取的特征能够获取图像像素中的高语义信息，在检索性能上已经超过了传统的手工特征[15]。

受到深度学习在图像检索上能够弥补语义鸿沟的启发，本文的主要贡献如下：

1. 对比几种传统特征和深度卷积特征在CBMIR任务上进行定量分析。
2. 微调医学图像特征提取网络，提取到的特征能在不平衡的医学数据集上有良好的表现。

文章后面部分安排如下，第2节是相关工作，第3节介绍了几种重要的传统特征和深度卷积特征，第4节显示实验结果，讨论和总结在第5节。

**2. Related work**

**2.1 Deep learning for CBMIR**

基于内容的医学图像检索是基于内容的图像检索与医学图像的有机结合，对于医学图像，由于临床上有益的信息在图像的小区域中高度局部化，基于全局特征提取的系统因此未能有效提供紧凑的特征表示[16]。

局部描述符使用一系列检测区域对局部图像内容进行编码。128维的SIFT向量在匹配准确率上从众多描述符中脱颖而出，成为了使用最多的描述符。在图像检索中通常结合Bag of Visual Words（BoVWs）模型使用。在[17]中，使用SIFT特征表达脑磁共振图像（MRI），再结合Bag of Visual Words（BoVWs）的方法，用于检索诊断阿尔茨海默病，Laguerre Circular Harmonic Functions coefficients（LG-CHF）作为检索的特征向量，优于基于SIFT和SURF特征的检索系统。

深度学习是机器学习的一个子领域，它使用一组算法，试图通过使用具有多个处理层的深层架构来模拟数据中存在的高级抽象，具有线性和非线性变换函数[18]。

一些研究人员成功将深度学习用于医学图像检索。Adden[19]等人通过训练深度神经网络对24类器官医学影像进行分类，最后结果表明深度神经网络能从图片中学到高级语义信息。Kim[20]等人结合原始特征和深度学习特征识别肺结叶CT，实验中最大准确率达到95.5%。Desai A D[21]等人使用在ImageNet2015上训练的ResNet-152网络提取CNN特征，使用SVM分类器实现胎龄估计。通常在处理小数据集时，通常会使用一些预训练的CNN网络迁移到目标数据集，对于医学图像，大部分数据是不平衡的且高度相似，而大部分预训练的CNN模型使用的是自然图像，例如ImageNet，使用这些预训练的模型也许效果并不好。受到CBIR和其他人的深入学习结果的启发，我们提出了有效的CBMIR深度学习特征提取方法。

**2.2 Retrieval Metric**

设 是一组图像的特征集合。 要查询图像与数据库每个图像的相似度，使用一个恰当的距离函数计算出与距离序列。根据它们距离大小，例如对数据库图像进行排序。常用的距离函数有欧氏距离(Euclidean Distance)，余弦距离，马氏距离(Mahalanobis Distance)等。

本文使用precision-recall，average precision 和平均精度（mean average precision, mAP）来评估CBIR的性能：

给定一张待查询图像和返回列表，可以根据P-R值绘制P-R曲线。

是每个查询检索到的相关项的精度分数平均值（ The average precision AP for a single query q is the mean over the precision scores after each retrieved relevant item）：

是检索到个相关图像的召回，是待查询图像总数。它相当于精准度-召回率曲线下的面积。通常，较大的AP值意味着更高的精准度-召回率曲线，亦即更好的检索性能。

平均精度是所有的平均值：

是查询图像集合。

**3. Feature for CBMIR**

**3.1 Gabor Filters**

Gabor Filters是最受欢迎的纹理描述符之一，在模式分析中也得到了广泛的应用。例如，它被用来研究脊柱内多孔海绵小梁骨内的方向性分布[22]。这里使用的Gabor特征是从图像的8个方向5个尺度提取的。

**3.2 SIFT + BoW**

尺度不变特征变换（SIFT），可在图像中检测出关键点，从尺度不变关键点获得特殊图像特征，提取关键点并计算其描述符，是一种局部特征描述子。SIFT特征是基于物体上的一些局部外观的兴趣点而与影像的大小和旋转无关。对于光线、噪声、微视角改变的容忍度也相当高。本文使用的是128维SIFT特征，通过聚类获得视觉字典，然后使用BoW生成固定长度的图像特征表达。

**3.3 灰度共生矩阵（GLCM）**

GLCM是描述具有某种空间位置关系两个像素灰度的联合分布。

**3.4 DCNN feature**

**3.4.1 Network Architecture**

Inceptionv3[23]是由GoogleNet[13]改进后的网络结构，能够很好地提取图像特征，所以使用Inceptionv3的特征层提取图像feature map（8 x 8），然后用Global Average Pooling生成一个2048维向量作为图像特征描述符。

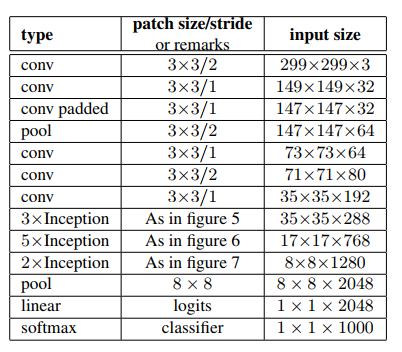


Figure 1. The Inceptionv3 network architecture

**3.4.2 Training details**

所有的图像resize到299x299大小输入到网络中，对标签使用one-hot方式编码。loss function使用交叉熵去拟合one hot的分布,交叉熵的公式是:

其中是预测的分布，而是真实的分布.考虑到训练数据不平衡，相似度十分大的影响，训练loss function使用了加权代价交叉熵函数：

其中是正样本系数，是代价系数。训练优化器使用的Adam算法[24]，整个模型初始学习率为0.01，每迭代50次在学习率上乘以0.1，总共迭代200次。

**4 实验结果**

**4.1 Dataset**

实验数据集IRMA来源于RWTH Aachen University Hospital，数据集由12000个116类的医学射线照片组成，其中有1000未公布的数据作为测试集，另外11000（10000训练+1000验证）作为训练数据。数据不均匀分布，每一类有至少10张图片提供训练[25]。

**4.2 Retrieval performance**

使用最常见的Precision and Recall评价检索性能。图1是各种特征的P-R图。

**5 分析和结论**

1. Yue, J., et al., *Content-based image retrieval using color and texture fused features.* 2011. **54**(3): p. 1121-1127.

2. Srinivas, M., et al., *Content based medical image retrieval using dictionary learning.* 2015. **168**(C): p. 880-895.

3. Yang, W., et al., *Content-Based Retrieval of Focal Liver Lesions Using Bag-of-Visual-Words Representations of Single- and Multiphase Contrast-Enhanced CT Images.* 2012. **25**(6): p. 708-719.

4. Ojala, T., et al., *Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns*. 2000: Springer Berlin Heidelberg. 404-420.

5. Zheng, L., et al., *SIFT Meets CNN: A Decade Survey of Instance Retrieval.* 2018. **40**(5): p. 1224.

6. Liu, L., et al., *From BoW to CNN: Two Decades of Texture Representation for Texture Classification.* 2018.

7. Murala, S. and Q.M.J.I.J.B.H.I. Wu, *Local mesh patterns versus local binary patterns: biomedical image indexing and retrieval.* 2014. **18**(3): p. 929-938.

8. Dubey, S.R., et al., *Local Bit-Plane Decoded Pattern: A Novel Feature Descriptor for Biomedical Image Retrieval.* 2016. **20**(4): p. 1139-1147.

9. Lowe, D.G.J.I.J.o.C.V., *Distinctive Image Features from Scale-Invariant Keypoints.* 2004. **60**(2): p. 91-110.

10. Philbin, J., et al. *Object retrieval with large vocabularies and fast spatial matching*. in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*. 2007.

11. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. in *International Conference on Neural Information Processing Systems*. 2012.

12. *<Very deep convolutional net- works for large-scale image recognition,” in ICLR, 2015..pdf>.*

13. Szegedy, C., et al., *Going Deeper with Convolutions.* 2014: p. 1-9.

14. He, K., et al., *Deep Residual Learning for Image Recognition.* 2015: p. 770-778.

15. Wan, J., et al. *Deep Learning for Content-Based Image Retrieval: A Comprehensive Study*. in *Acm International Conference on Multimedia*. 2014.

16. K., K.K. and T.V. Gopal. *Multilevel and multiple approaches for Feature Reweighting to reduce semantic gap using relevance feedback*. in *International Conference on Contemporary Computing and Informatics*. 2015.

17. Mizotin, M., et al. *Feature-based brain MRI retrieval for Alzheimer disease diagnosis*. in *IEEE International Conference on Image Processing*. 2013.

18. Deng, L. and D.J.N.P. Yu, *Deep Learning for Signal and Information Processing.* 2013.

19. Qayyum, A., et al., *Medical image retrieval using deep convolutional neural network.* 2017.

20. Kim, B.C., S.S. Yu, and H.I. Suk. *Deep feature learning for pulmonary nodule classification in a lung CT*. in *International Winter Conference on Brain-Computer Interface*. 2016.

21. Desai, A.D., et al., *Open-source, machine and deep learning-based automated algorithm for gestational age estimation through smartphone lens imaging.* Biomedical Optics Express, 2018. **9**(12): p. 6038-6052.

22. Gdyczynski, C.M., et al., *On estimating the directionality distribution in pedicle trabecular bone from micro-CT images.* 2014. **35**(12): p. 2415-28.

23. Szegedy, C., et al. *Rethinking the Inception Architecture for Computer Vision*. in *Computer Vision & Pattern Recognition*. 2016.

24. Kingma, D. and J.J.C.S. Ba, *Adam: A Method for Stochastic Optimization.* 2014.

25. Müller, H., et al. *Overview of the ImageCLEFmed 2007 Medical Retrieval and Medical Annotation Tasks*. in *Advances in Multilingual & Multimodal Information Retrieval*. 2008.