Image Retrieval Based on Deep Learning

江俊广 School of Software, Tsinghua University
2015011584
13126830206@163.com

李沁恬 School of Software, Tsinghua University
2016013268
liqintian2016@163.com

刘译键 School of Software, Tsinghua University
2016013239
liuyijian16@163.com

ABSTRACT

本文采用了逐步求精的策略进行图像的检索，首先通过迁移学习训练图像的分类网络，根据分类网络的输出进行图像的粗选，然后采用注意力网络对粗选出的候选者进行精挑。

Keywords

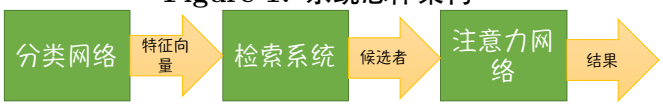
attention network, image classification, image retrieval

1. INTRODUCTION

基于深度学习的图像分类已经较为成熟，但是图像分类只关注不同类图像间的差异（[4]），直接将其迁移到图像检索时往往会忽视同一类图像间的差异。基于注意力网络的图像检索 [7] 通过图像级别的标签学习如何提取图像中的关键点，通过比对关键点的相对位置和特征来对图像进行比较，在大规模数据集上取得了较好的效果（在在 Oxford105k 的准确度为 88.5%），但是该方法存在的一大问题是速度较慢。因此我们将图像分类网络和注意力网络相结合（图1），利用分类网络效率高的特点进行图像的粗选，利用注意力网络准确度高的特点进行图像的精挑。

2. IMAGE CLASSIFICATION NETWORK

Figure 1: 系统总体架构



我们将现有的图像分类网络 (resnet18，参考了 [3]) 作为预训练网络，基于当前的数据集对该网络进行训练。训练时将给定数据集中的 80% 图片作为训练集，20% 图片作为验证集，采用随机梯度下降算法，最终在验证集上的分类准确度达到 98.76%。

然后使用训练好的分类网络对所有的图像进行全局特征提取。其中，特征提取有两种策略：

- 从分类网络的倒数第 2 层获取一个高维（512 维）向量，对该向量进行 PCA 降维（到 32 维）。
- 直接用最后一层的输出（10 维）向量。

图像分类网络的倒数几层包含了该图像大量的信息，我们关心的是，使用最后一层和最后第二层在进行图像检索时准确度的差异，以及 PCA 降维对图像检索的影响。具体的实验见5。

3. BALL TREE

Ball tree 是为了克服 KD 树高维失效而发明的，其构造过程是以质心 C 和半径 r 分割样本空间，每一个节点是一个超球体。通过这种方法构建的树要比 KD 树消耗更多的时间，但是这种数据结构对于高结构化的数据是非常有效的（[6]），鉴于图像中提取出来的特征向量维数较高，我们选取 ball tree 来搭建数据检索系统。至此，我们可以通过图像分类网络结

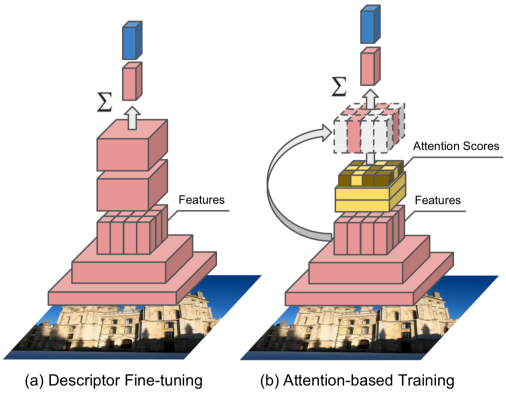
合 Ball tree 获得若干粗选后的候选者。

4. ATTENTION NETWORK

注意力网络（见图4）的训练过程分为两步:

1. 让网络学习用 N 个 d 维向量对图像进行表达。
2. 让网络学习得分函数，输出加权的求和特征。

Figure 2: Large-Scale Image Retrieval with Attentive Deep Local Features



我们通过修改 [5] 的注意力网络，得到图像的关键点及其特征。然后对每个候选者逐一验证，验证方法是用 RANSC 算法计算两张图片中有多少关键点的位置符合 2D 仿射变换 AffineTransform，符合点的数目越多，说明这两张图片越相像。由此得到每个候选者的验证得分，按照验证得分返回最近的 k 张图片。

5. EXPERIMENTS

5.1 特征向量、注意力机制对检索准确度的影响

我们采用了不同的特征向量进行粗选，并对比了有无注意力网络对检索准确度的影响。由于检索准确度不仅取决于图片的类别，还取决于两张图片的相似程度，因此我们从数据库中挑出 14 张较难检索的图片，采用 6 种方法进行检索，然后用人工的方式对返回的 10 张图片进行评价，以此衡量检索的准确度。所有方法返回的标签和查询图片的标签基本一致，准确度指的是和查询结果相似的图片的比例。从表1中可以看出，分类时的特征向量直接使用神经网络倒数第二层的效果最好，而有注意力网络在大多数图片上的检索效果好于没有注意力网络。

Table 1: 特征向量、注意力机制对检索准确度的影响

图片名	只用分类网络			分类 + 注意力网络		
特征向量维度	10	32	512	10	32	512
n01613177_1805	0.3	0.7	0.6	0.6	1	0.9
n01923025_3201	0.7	0.8	0.9	0.8	0.9	0.8
n02278980_5577	0.5	0.6	0.5	0.6	0.8	1
n03767203_3741	0.6	0.7	0.7	0.6	0.8	0.9
n03877845_5041	0.4	0.5	0.5	0.5	0.7	0.7
n03877845_5487	0.2	0.5	0.5	0.4	0.6	0.7
n04515003_16807	0.6	0.5	0.5	0.6	0.4	0.5
n04515003_37361	0.4	0.3	0.5	0.5	0.7	0.8
n04583620_4028	0.7	0.7	0.6	0.7	0.6	0.9
n07897438_1733	0.6	0.6	1	0.7	0.9	0.9
n07897438_4679	0.3	0.4	0.6	0.3	0.5	0.6
n10247358_14658	0.5	0.7	0.4	0.3	0.6	0.5
n11669921_12332	0.7	0.8	0.9	0.6	0.7	0.8
n11669921_43145	0.4	0.4	0.6	0.4	0.4	0.6
平均分数	0.49	0.58	0.63	0.54	0.69	0.76

5.2 候选者大小对检索准确度的影响

在分类网络 + 注意力网络模型中非常重要的一个参数是候选者的大小，候选者越多，可供注意力网络精挑的数目就越多，但与此同时检索的速度也越慢。为了找到一个合适的大小，我们进行了表2中的实验。通过注意力网络我们可以对两张图片的相似度打分，我们用返回的 10 张图片的平均相似度作为分类网络得到的候选者的准确度。从图3可以看出，随着候选者的增加，返回结果的平均相似度不断增加，为了兼顾检索效率，候选者的数目不妨取为 4k。

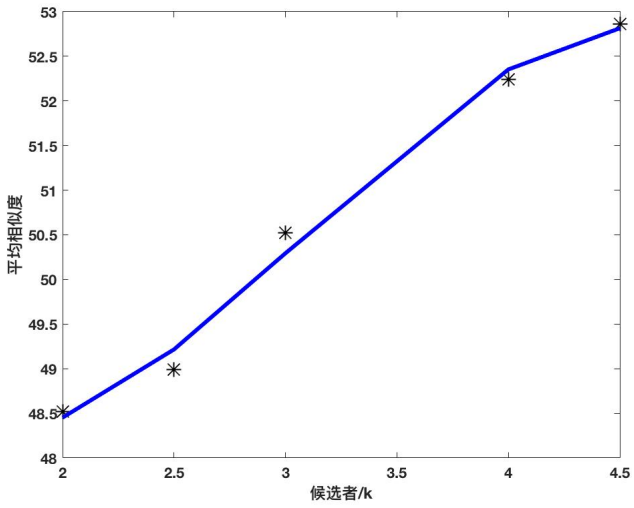
6. CONCLUSIONS

通过逐步求精的策略，我们将分类网络和注意力网络相结合，并通过实验证明注意力网络确实能提高检索准确度。与此同时，为了加快检索，用注意力网络进行验证的候选者数目不能太多。

Table 2: 候选者大小对平均相似度的影响

候选者/k	2	2.5	3	4	4.5
n11669921_12332	115.2	114	115.7	117.9	117
n11669921_43145	26	26.5	29.8	30.2	33
n10247358_14658	35.2	35.3	35.4	37	37.3
n07897438_4679	19.1	19.4	19.8	20.6	21.5
n07897438_1733	40.1	40.2	40.3	41.4	42.2
n04583620_4028	69	68.6	70.2	73.4	73.9
n04515003_37361	44.9	44.8	45.1	46	47.9
n04515003_16807	37.8	38.6	41.4	42.4	43
n03877845_5487	66.9	68.7	75.3	77.3	77.7
n03877845_5041	45.8	45.7	47.1	48.4	49
n03767203_3741	52.8	53.8	54.2	54.6	54.7
n02278980_5577	50.3	51.4	51.7	54.4	55
n01613177_1805	44.5	47.1	49.1	52.9	52.7
n01923025_3201	31.7	31.7	32.2	34.9	35.1
平均	48.52	48.99	50.52	52.24	52.86

Figure 3: 候选者大小对平均相似度的影响



不足与改进. 通过注意力网络进行验证也有其局限性，当数据库的规模较小或者数据库中与查询图片相似的图片较少时，通过验证得到的查询结果的准确度反而不如直接通过分类网络得到的结果准确度。为了克服这个弊端，返回结果首先考虑通过注意力网络得到的相似度较高的图片，如果数目不够，再加入分类网络获得的相似度较高的图片。经过实验，这种方法确实有效地提高了准确度。

7. REFERENCES

[1] Babenko, A., Slesarev, A., Chigorin, A., and Lempitsky, V. Neural codes for image retrieval. In *Computer Vision – ECCV 2014* (Cham, 2014), D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Springer International Publishing, pp. 584–599.

[2] Bhatia, N., and Vandana. Survey of nearest neighbor techniques. *CoRR abs/1007.0085* (2010).

[3] Chilamkurthy, S. Transfer learning tutorial. https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html.

[4] Gordo, A., Almazán, J., Revaud, J., and Larlus, D. Deep image retrieval: Learning global representations for image search. *CoRR abs/1604.01325* (2016).

[5] Hyeonwoo Noh, Andre Araujo, J. S. T. W. B. H. P. I. Large-scale image retrieval with attentive deep local features. https://github.com/JunguangJiang/models/tree/master/research/delf.

[6] Kumar, N., Zhang, L., and Nayar, S. What is a good nearest neighbors algorithm for finding similar patches in images? In *Computer Vision – ECCV 2008* (Berlin, Heidelberg, 2008), D. Forsyth, P. Torr, and A. Zisserman, Eds., Springer Berlin Heidelberg, pp. 364–378.

[7] Noh, H., Araujo, A., Sim, J., and Han, B. Image retrieval with deep local features and attention-based keypoints. *CoRR abs/1612.06321* (2016).