

大作业实现:

组员: 姜霁恒 5130379033

采用算法: 朴素贝叶斯分类器

首先读取对应照片, 分别存放于不同的容器中。接下来对于各照片进行高斯滤波, 这里我选取的高斯滤波窗口大小为 (3,5), 考虑到由于原图并不完全对称, 各点对于周围的影响是各向异性的, 因此对于略长的 y 轴上选取的窗口大小就要长于 x 轴上的窗口大小。在进行高斯滤波降噪以后, 对于每一个组对应点生成对应向量 (flair, t1, t1c, t2) → (truth), 当 truth 的值不等于 0 的时候即算作为有肿瘤的区域, 其他情况则算作无肿瘤。为了减小误差, 我还将背景点 (即 flair, t1, t1c, t2 均为 0 的点) 舍弃。另外, 虽然样本点的灰度值是离散的, 但由于变量数太多已经接近于连续的情况, 因此我将 256 个灰度级均匀分成了 32 组, 每组 8 个灰度级, 这样可以视为原变量是离散值, 直接利用出现频率除以总次数即可得到其概率。

在统计完所有向量后, 对得到的结果进行概率的计算, 只需要将某一组中的频数除以总数即可得到其出现的概率。在计算被测图像的时候, 我简化了这一计算公式:

$x = \{a_1, a_2, a_3, a_4\}$ ,  $C = \{y_0, y_1\}$ ;

其中 x 是变量特征, 四个特征分别对应于上面的向量的四个维度。而 C 代表类别集合, 我在这里做了简化, 只有两种情况, 肿瘤区域和非肿瘤区域。

我们知道贝叶斯公式可以写成如下形式:

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

因为分母  $P(x)$  对于所有类别皆为常数, 在比较大小时可以略去, 因此我们得到如下公式:

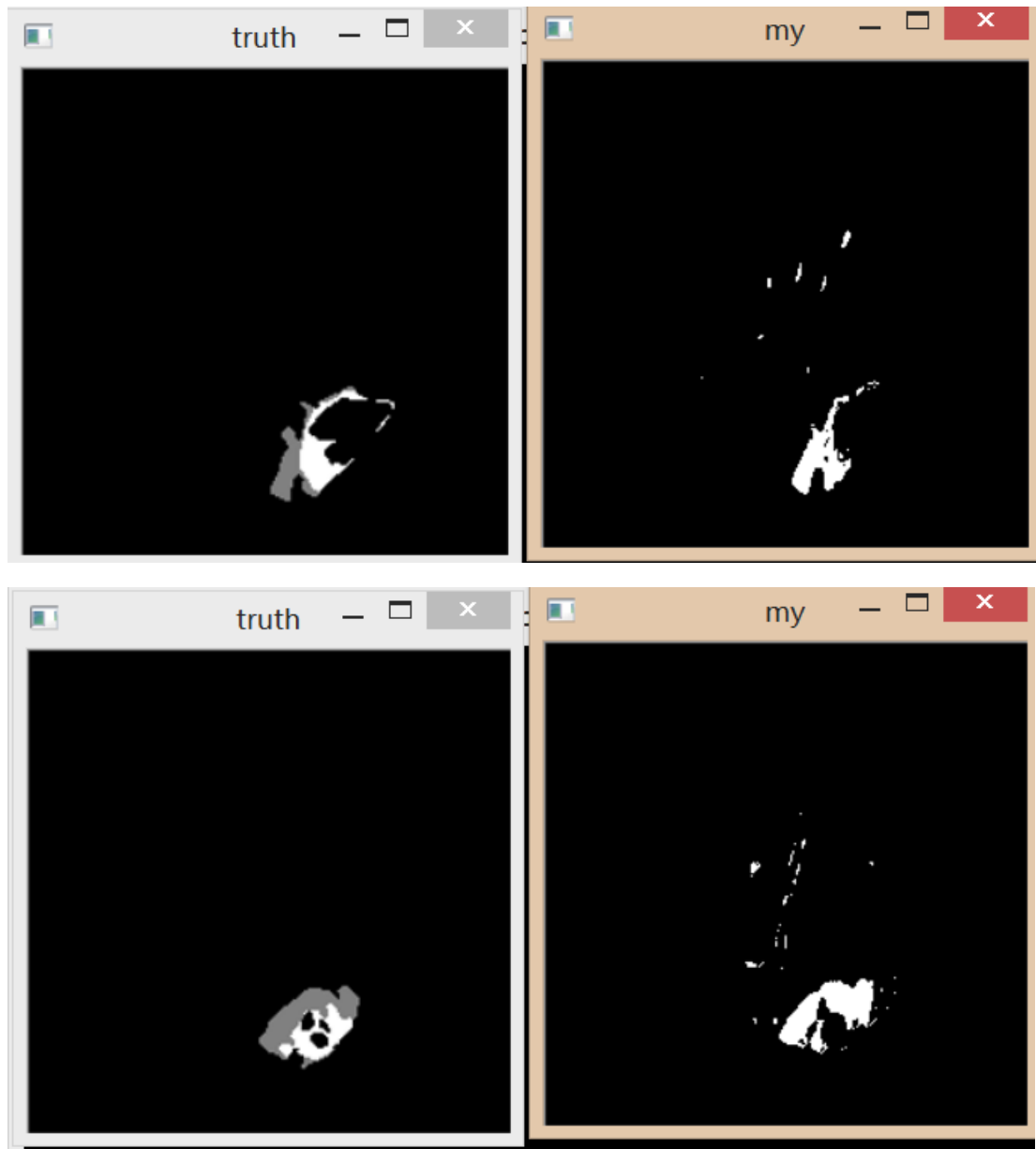
$$P(x|y_i)P(y_i) = P(a_1|y_i)P(a_2|y_i)...P(a_m|y_i)P(y_i) = P(y_i) \prod_{j=1}^m P(a_j|y_i)$$

如果  $P(y_k|x) = \max\{P(y_1|x), P(y_2|x), ..., P(y_n|x)\}$ , 则  $x \in y_k$

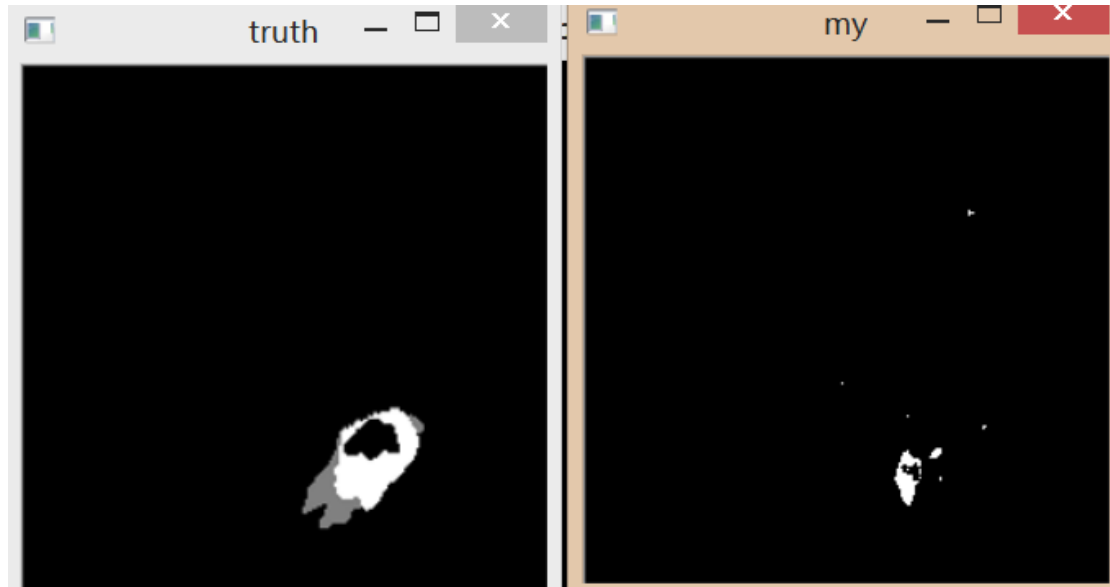
重复的对于本图进行判断, 最后绘制出图像。

算法执行效果:

当原图肿瘤区域与非肿瘤区域差别较为明显时, 效果较好 (图片 97 和 122):



而当原图肿瘤区域与非肿瘤区域相差不大时，算法会有稍大一些的误差（图 86）：



出现这种情况的可能原因是，由于朴素贝叶斯算法要求各变量相互独立，而在本例中显然各变量并非相互独立，尤其 `t1`, `t2`, `t1c` 三变量关系十分明显（未经数学证明，仅观察得），所以会对结果造成一定的影响。也有可能是由于我训练时使用了全部的数据所导致的。

程序使用：

输入 `det [number]` 来训练某一张图。

输入 `exit` 退出。

关于高斯滤波：

我在 `canny edge detector` 中自己实现了一个简易的各向同性的高斯滤波器，代码已经附在大作业的代码中了。但是由于运行下来感觉效果并非最理想，因此最后我还是选择了调用 OpenCV 的库函数。

环境：

由于在学期中间重新安装了 VS2015，导致我的 OpenCV 库也重新编译过。

OpenCV: 3.0.0 vc12/vc14 x86

IDE: VS2013 Ultimate/VS2015 Enterprise（当然，全都是破解版）