

创建数据集

Dataset_Creator_200M

- 输入：数据库路径，数据集存储路径，存储训练集的 tfrecords 文件数量
(Default: database_path = 'db/lydhdb', output_path = '/tmpdata/', trainfile_num = 9)
- 输出：约 200 万样本训练集和约 7 万样本测试集

创建~200 万样本的训练样本集和~7 万样本的测试集，每个样本均为 512 个采样点的片段。训练集包含纯噪声片段，无噪声片段和规律噪声（方波，正弦波，三角波）片段。其中含 R 波样本~50%，无噪声无 R 波样本~35%，纯噪声样本~13.5%，规律噪声各类 747 个。测试集不包含规律噪声样本，其他样本类型比例与训练集一致。

每个记录人工标记的 R 波位置前后 20 个采样点都被视为 R 波。若样本正中间的采样点为 R 波，则被标为 1。其余均标为 0。

Dataset_Creator_YOLO

- 输入：数据库路径，数据集存储路径，存储训练集的 tfrecords 文件数量
(Default: database_path = 'db/lydhdb', output_path = '/tmpdata/', trainfile_num = 9)
- 输出：约 230 万样本训练集和约 8 万样本测试集

创建~230 万样本的训练样本集和~8 万样本的测试集，每个样本为 512 个采样点的片段。训练集包含纯噪声片段，含 R 波片段（N，A，V）和规律噪声（方波，正弦波，三角波）片段。其中含 R 波样本~200 万（N：47%，V：31%，A：22%），纯噪声样本~30 万，规律噪声各类 747 个。测试集不包含规律噪声样本，其他样本类型比例与训练集一致。

此样本集已经过数据增强，每个含 R 波的原始片段都与 10 种 em 和 10 种 ma 噪声进行人工叠加。N 类和 V 类原始片段的选取是通过对随机区域进行以 8 个采样点为步长的滑窗（512 点）进行选取，每个随机区域总共做滑窗 32 次。A 类原始片段是通过对随机区域进行以 8 个采样点为步长的滑窗（512 点）进行选取，每个随机区域总共做滑窗 16 次。

每个样本都被划分为 8 个等宽区域。样本标签为 8*2 的矩阵，8 表示 8 个区域，2 表示是否有 R 波和 R 波的位置对于区域宽度的比例。

Dataset_Creator_YOLO_400M（窄高噪声选取还没完成，其余已完成）

- 输入：数据库路径，数据集存储路径，存储训练集的 tfrecords 文件数量
(Default: database_path = 'db/lydhdb', output_path = '/tmpdata/', trainfile_num = 9)
- 输出：约 430 万样本训练集和约 8 万样本测试集

创建~430 万样本的训练样本集和~8 万样本的测试集，每个样本为 512 个采样点的片段。训练集包含纯噪声片段，含 R 波片段（N，A，V，超小波，传导阻滞）和规律噪声（方波，正弦波，三角波）片段。其中含 R 波样本~400 万（N：40%，A：20%，V：15%，超小波：15%，传导阻滞：10%），纯噪声样本~30 万，规律噪声各类 747 个。测试集不包含规律噪声样本，其他样本类型比例与训练集一致。

此样本集已经过数据增强，每个含 R 波（除超小波）的原始片段都与 10 种 em 和 10 种 ma 和 10 种窄高噪声进行人工叠加。N 类，V 类和传导阻滞类原始片段的选取是通过对随机区域进行以 8 个采样点为步长的滑窗（512 点）进行选取，每个随机区域总共做滑窗 32 次。A 类原始片段是通过对随机区域进行以 8 个采样点为步长的滑窗（512 点）进行选取，每个随机区域总共做滑窗 16 次。超小波片段为人工选取。

每个样本都被划分为 8 个等宽区域。样本标签为 8*2 的矩阵，8 表示 8 个区域，2 表示是否有 R 波和 R 波的位置对于区域宽度的比例。

搭建模型

Model_Builder

搭建论文中提到的卷积神经网络。该网络包含 attention layer。和论文中的网络不同的是，该网络中第四个和第五个 conv unit 后分别加入了 0.1 和 0.2 的 dropout，且网络最后新添加了一个 0.5 的 dropout 和一个 dense layer。

Model_Builder_YOLO

搭建 YOLO 的卷积神经网络。

模型训练

Train_and_Test_200M

搭建 data pipeline 直接读取 tfrecords 数据。设置模型 callback 并 compile 模型。训练模型并绘制每个 metric 的曲线图。还包括了绘制训练后 attention layer 权重和绘制所有预测错的样本的功能。

Train_and_Test_YOLO

搭建 data pipeline 直接读取 tfrecords 数据。设置模型 callback 并 compile 模型。训练模型并绘制每个 metric 的曲线图。

记录预测

Predict_200M

- 输入：ecg 记录路径，model 路径
- 输出：该记录的阳性预测值和敏感度。

根据论文中提到的方法，对用论文中提到的模型预测后的数据进行处理。记录中每个点在模型预测后都得到一个 0 到 1 之间的值，表示该点为 R 波的可能性。首先对数据进行 median filter 并根据设置的 threshold 把结果转化成 0 或 1，再把过于窄的 1 区间去掉。再次排查数据，若两段为 1

的区间距离过小，则去掉更窄的区间；若两段为 1 的区间距离过大，则用更小的 threshold 再次搜索可能遗漏的 1 区间。二次排查结束后，再次过滤掉过于窄的 1 区间。最后取每个 1 区间的中点为 R 波波峰，并计算阳性预测值和敏感度。

Predict_YOLO

- 输入： model 路径，数据库路径
- 输出： 打印出数据库所有数据的平均阳性预测值和敏感度。

把数据库的每条记录截取成多个为 512 点宽的片段。用训练好的 YOLO 模型对每条片段进行预测，并对模型输出数据进行 nonmax suppression。最后根据模型输出数据的相对位置，计算每个预测 R 波的绝对位置。把 R 波预测位置与真实位置进行比较，计算记录的阳性预测值和敏感度。

Trained Model

Model_v1: 论文中提到的模型，但已去掉 attention layer，并加上了一些 dropout。已经训练好，为正确率最高的模型。

Model_YOLO: 已经训练好，为正确率最高的 YOLO 模型。