

Abstract

研究现状：收集多个注释，投票选出一个

问题：忽略了不同注释里医生共同的同意或不同意的丰富信息

提出：model multi-rater (dis)agreement, MRNet.

1. 专家推理模块 EIM

以嵌入专业水平的个人评分及作为先验知识

形成高级语义特征

2. 从粗略的预测中重建多评估梯度，进一步提高分割性能

创新：第一次在不同专业水平下为医学图像分割生成校准预测

实验：5个不同成像的医疗分割任务。

结果：优？

Introduction

动机 每个专家水平不一样，需要合适的分割策略来保证一致性

现有投票，忽略了不一致信息。

意义

贡献 1) 引入专家的概念，EIM.

2) 多评估者重建模块(MRM)，对先验和软预测进行重构。

3) 为了更好的利用多评估者的信息，在模型里加入了 Multi-rater Perception Module (MPM). 经验证的导致性能的提升。

优于世界上其它的多人评估策略，并且 29 FPS.

2. Related work

医疗图片分割：以上方法，没有充分利用多名评估员之间的
可变性

多评估者策略：抽样，多分支结构缺乏原则性的调查研究。

在利用丰富不一致信息，用于训练和预测上。

3. Methodology

3.1 Motivation.

专家水平不一。

定量测量这种差异，在基准数据集 RIGA 评估
训练多个网络，再投票

Table 1. 通过 Dice coefficient 检查一致性及专业水平。

3.2 Overall Framework

Fig 2. 分两阶段。

一. Unet \rightarrow Encoder

EIM 提取高级特征值到 Decoder

$F \rightarrow |X| \text{ Con} + \text{Sigmoid} \rightarrow P$

二. 细化。

MAM

MRM :	重建不确定图.
MPM :	具有注意力的, 利用不确定图预测 进行细化

3.3. Expertise-aware Inferring Module.

给专家加权重

$$GT^{\text{soft}} = \sum_{i=1}^N S_i V_i \rightarrow \psi(x, v)$$

↓ ↓ ↓ ↓
 mask 参数 input 权重向量

训练期间, 3种策略交替设置 V .
作为数据增强.

推理期间. 使用多数投票模式.

ConvLSTM: 因为语义各个信息到语义特征表示中.

作用 捕捉不同专家与特征的相关性.

方式

$$\text{公式 } h_t = \bigcup_{t=1}^T \text{ConvLSTM}(f^5, h_{t-1})$$

$t = 1, 2, \dots, T$.

empirically $\rightarrow 2$.

h 及 f 的增强

3.5. Multi-rater Reconstruction Module

固有的、关联性、有价值的一致

\downarrow
专家商量和预测 \downarrow
不同专家

该部分使用 VGG16

Encoder \downarrow 保留输入图像的拓扑和感知特征

$$\text{单个评分者} \quad \text{多个评分者} \quad \text{Loss}_{\text{rec}} = \frac{1}{N} \sum_{i=1}^N L_{\text{BCE}}(S_i, \bar{S}_i)$$

二值交叉熵损失

$$\text{两者融合 Loss}_{\text{con}} = \frac{1}{k} \sum_{i=1}^k \frac{1}{2} \|D_i - \bar{D}_i\|^2$$

Conv blocks num $\rightarrow k=5$

VGG 16

重建单个评分者的评分之后。

$$\text{通过 } U_{\text{map}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{S}_i - \frac{1}{N} \sum_{i=1}^N \bar{S}_i)^2}$$

像素标准差评估不一致且不确定图。

3.5. Multi-rater Perception Module.

采用 spatial attention strategy.

强调不为空区域。

$$\text{soft}(U_{\text{map}}) = \text{softmax}(F_{\text{Gauss}}(U_{\text{map}}, k), U_{\text{map}})$$

\downarrow
kernel

init : 32, 4

\downarrow
核 \downarrow
标准差

软注意力机制。向左分支，跳连接

$$\tilde{F}^j = F' + \text{Soft}(A_j) \otimes F'$$

$$M = \text{Conv}_{1 \times 1}(\text{Concat}(\tilde{F}', \tilde{F}^2, \tilde{F}^3, \tilde{F}^4))$$

$$\begin{aligned} L &= L_{BCE}(P', GT^{\text{soft}}) + L_{BCE}(M, GT^{\text{soft}}) \\ &+ \alpha \text{loss}_{\text{con}} + (1-\alpha) \text{loss}_{\text{rec}}. \end{aligned}$$

4. Experiments.

4.1 Datasets.

RIIGA: 公开的视网膜杯盘分割数据集,

三千来源, 共 750 张	} 195 BinRushed 460 Messidor 95 Magrabia	> train
二位专家		

U-Net → 预训练于 ImageNet.

QUBIQ: inter-rater variability

同类.

4.2. Experimental Setup.

4.2.1 Implementation Details.

主框架 Unet with ResNet34.

MRM DeepLab-v3 + VGG 16

4.2.2 Metric

D^s, IoU^s 越高, 表现越好

4.3. Experimental Results.

4.3.1. Performance of the Multi-rater Strategy.

不太理解. STAPLE strategy.

TABLE 2.

i证明了该框架的有效性.

动态专家推理和多评估者一致性建模.

5. Conclusion.

将多个专家评估的不一致纳入为线索.

生成更好地反映多个专家之间潜在一致性的校准模型预测.

{ 感知
重建