

CV-Final-开题报告的题目

刘颖凡

熊双宇

2020 年 5 月 9 日

1 选题背景及意义

1.1 问题

图像处理算法有很多实际运用, 如交通流量监控, 城市规划等, 但是由于现实中视角错位、相机拍摄角度和位置变化等问题, 而在实际使用中充满挑战. 如对不同医院里的新型冠状病毒肺炎患者的 CT 图片进行建模, 以诊断目标医院中 CT 图照片的持有人是否患病时, 不得不考虑由于设备差异 (相机拍摄角度和位置等) 对模型的预测结果可能造成的影响. 因此, 本文借鉴 one-shot 方法, 基于图像处理算法实际运用下的不同场景进行训练, 再利用目标场景中一张或少量带有标注的图片, 调整模型以减少设备差异导致的影响.

1.2 意义

1.2.1 使用 one-shot scene-specific 方法

给定一个目标场景, 这个场景下的所有图片, 由位置和角度保持不变的相机拍摄. 假设在这个场景下, 已有一张有标注的图片, 我们的目的是, 基于这一张标注的图片, 调整模型, 去适应该场景.

1.2.2 对比其他 one-shot 模型更符合现实运用

其他模型	现实条件	scene-specific 模型
一个模型通常被用于所有场景, 即在测试中, 同一个模型需要处理可能遇到的各种场景	测试中的图片通常是由固定角度和位置的相机拍摄 (即需要处理的图片可视为都属于单一场景)	能够调整模型使之适应单一的特定场景
假定一个已经学习完毕的模型, 无法被调整去适应某个特定场景	被安装在一个新环境下的相机, 本身就存在一个校准的过程	在校准过程中, 采取少量照片进行人工标注, 是可行的 . 因此, 可在校准的过程中, 利用 少量有标注的图片 , 调整模型, 去适应目标场景

1.2.3 举例

- 假设一个公司要利用图像处理算法为不同的客户提供服务, 我们可以将同一个客户提供的图片, 归属为一个场景
- 公司可以标注这些图片, 并且线下训练一些模型

- 假设公司要使用该服务给一个新的客户 (即新的场景), 可以合理地假设我们有新场景的一张或几张带标注的图片. 如果有 one-shot scene-specific 方法, 公司只需要基于这几张带标注的图片, 快速调整原线下训练好的模型即可

2 相关工作参考

1. One-Shot Scene-Specific Crowd Counting: 利用 **CSRNet** 作为框架搭建 **one-shot scene-specific model**

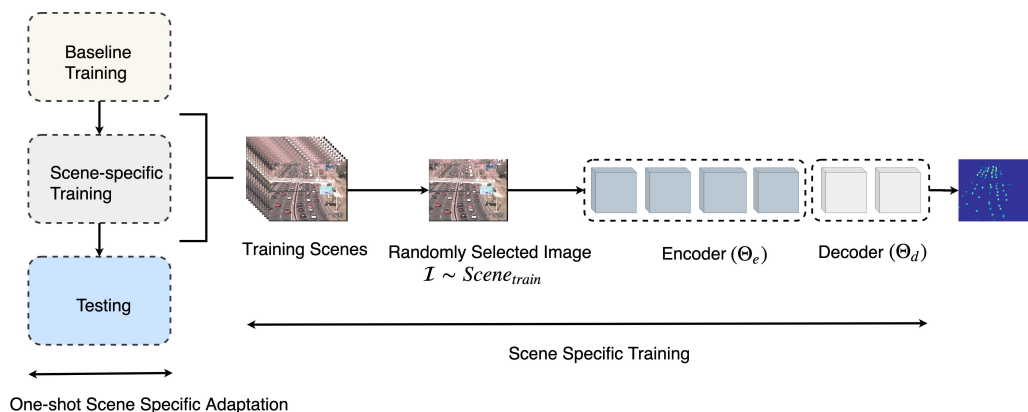


图 1: 网络框架

- (a) 将训练过程分为两部分:
 - i. baseline training: 联合训练 encoder 和 decoder, 得到 baseline model
 - ii. scene-specific training: 使用 baseline model 的权重初始化 encoder 和 decoder, 再次训练 decoder, 使模型能够有效地适应新的目标场景
 - (b) 在测试过程中, 使用目标场景的一张带标注图片, encoder 权重保持不变, 微调 decoder 的权重
2. COVID-CT-Dataset: A CT Scan Dataset about COVID-19: 创建用 COVID-CT 图片集, 建立二分类模型, 预测是否感染 COVID-19
 - (a) 训练: 由于数据集较小, 使用迁移学习 (transfer learning) 和数据增强 (data augmentation) 两种算法:
 - i. 迁移学习: 使用 ChestX-ray14 数据集, 预训练 DenseNet 后, 在 COVID-CT 数据集上微调模型
 - ii. 数据增强: 随机仿射变换 (包括 transformation 和 rotation), 随机裁剪和翻转
 - (b) 测试: 模型的 $F1$ score 达到 0.85
 3. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements and Prognosis of COVID-19 Pneumonia Using Computed Tomography
 - (a) 构造一个较大的 CT 数据集, 包括新型冠状病毒肺炎, 其他常见肺病和正常肺部
 - (b) 建立一个 AI 系统, 用于辅助专家进行准确诊断
 - i. 使用 Unet, SegNet 等神经网络用于肺部病变分割
 - ii. 基于肺部病变分割等模型, 建立诊断系统, 用于区分 NCP 和其他常见肺病和正常肺部
 - iii. 在中外人群肺部图片上测试 AI 系统

- (c) 为了帮助内科医生提供工具, 结合 CT 和临床参数, 为新冠肺炎患者提供预后指征,
 - i. 基于病变的定量测量, 根据病变大小和体积在用药前后的变化, 评价药物作用
 - ii. 观测病变特征和临床关键参数的相关性

3 课题研究内容

1. 基本功能: 基于新冠肺炎患者的 CT 影像数据进行建模, 训练模型, 尽可能提高识别出新冠肺炎患者的准确性
2. 附加功能:
 - (a) 对于同一的患者, 对比其从健康时到确诊过程中肺部 CT 影像的区别, 描述新冠病毒的扩散方式; (若可以搜集到同一个患者的数据)
 - (b) 可以对比无症状患者和有发热特征患者的 CT 影像数据, 找出其中是否有差异存在

4 实验设计思路

4.1 前期准备

1. 数据集准备:
 - (a) 2019 Novel Coronavirus Resource (2019nCoV)
 - (b) COVID-19 cases with chest X-ray or CT images
 - (c) CT Scan Dataset about COVID-19
2. 阅读相关论文

4.2 模型设计

4.2.1 模型概述

1. **训练:** 假设有一些带标注的数据, 这些数据来源于几个不同的场景, 运用这些数据训练模型
2. **测试:** 在一个与训练场景不同的特定的目标场景中, 利用该场景下一张带有标注的图片, 使用 crowd counting 算法调整模型后, 预测图片的 density map
3. **网络框架:** CSRNet[encoder, decoder]

4.2.2 使用到的技术

1. One-Shot Scene-Specific Adaptation

在测试过程中: 更新 θ_d , 虽然叫测试, 但是所有图片都有 *ground truth density map* 作为标注

- (a) 从属于目标场景的 M 张图片的数据集 $x_{i=1}^M$ 中给出一张图片 x_1 , 和对应的标注 density map y_1 (用于微调 decoder)
- (b) 记 encoder 的参数为 θ_e , decoder 的参数为 θ_d , 得到的 predicted density map 为 $F(x; \theta_e, \theta_d)$. 可见参数满足性质: θ_e 适用于所有场景, θ_d 适用于目标场景

(c) **微调 decoder**: 使用 y 和 F 之间的 L_2 distance 作为损失函数:

$$\begin{aligned}\theta_d^* &= \operatorname{argmin}_{\theta_d} L_d(\theta_d), \\ \text{where } L_d(\theta_d) &= \|\operatorname{vec}(F(x_1; \theta_e, \theta_d)) - \operatorname{vec}(y_1)\|_2\end{aligned}\quad (1)$$

- i. $\operatorname{vec}(\cdot)$ 将输入的矩阵转化为一个向量
- ii. $L_d(\theta_d)$ 是关于 θ_d 的函数 (因为被微调的只有 decoder network, θ_e 可视作常数), 并且定义在目标场景中的 (x_1, y_1) 上
- iii. 微调过程: 运用反向传播, 计算 L 对 θ_d 的梯度后, 用梯度下降更新 θ_d
- iv. 当使用目标场景中一张图片微调 decoder 后, 预测该场景中剩下的图片的 density map

2. Model Learning 在训练过程中: 学习 θ_e 和 θ_d

- (a) 使用 T 个场景中所有图片和对应标注, 假设每个场景中有数量为 N 的图片
- (b) 当得到了 θ_e , 可以通过 (1) 得到 θ_d , 则 $F(x_i^{(t)}; \theta_e, \theta_d^*)$ 应当尽可能接近 y_i , 但如果 x 不是属于 t 场景, 不要求 $F(x; \theta_e, \theta_d^*)$ 在 x 上表现很好
- (c) 对于固定的 θ_e , 针对 t -th 场景的 decoder 的参数被标记为 $\theta_d^{(t)}$, 并且 $\theta_d^{(t)}$ 由下式得到

$$\begin{aligned}\theta_d^{(t)} &= \operatorname{argmin}_{\theta_d} L_d^{(t)}(\theta_d), \\ \text{where } L_d^{(t)}(\theta_d) &= \|\operatorname{vec}(F(x_1^{(t)}; \theta_e, \theta_d)) - \operatorname{vec}(y_1^{(t)})\|_2\end{aligned}\quad (2)$$

- (d) 假设 (2) 中的解 $\theta_d^{(t)}$ 对于固定的 θ_e 是唯一的, 则 $\theta_d^{(t)}$ 可视为 θ_e 的函数, 标记为 $\theta_d^{(t)}(\theta_e)$
- (e) 优化过程

对于 t -th 场景中的训练数据 $\{x_i^{(t)}, y_i^{(t)}\}_{i=2}^N$, 损失函数:

$$L_e^{(t)}(\theta_e) = \sum_{i=2}^N \|\operatorname{vec}(F(x_i^{(t)}; \theta_e, \theta_d^{(t)}(\theta_e))) - \operatorname{vec}(y_i^{(t)})\|_2 \quad (3)$$

对于所有场景的总损失函数:

$$L_{final}(\theta_e) = \sum_{t=1}^T L_e^{(t)}(\theta_e) \quad (4)$$

迭代以下两步直到收敛:

- i. 固定 θ_e , 使用 (2) 算出 $\theta_d^{(t)}$, ($where\ t = 1, 2, \dots, T$)
- ii. 固定 $\theta_d^{(t)}$, ($where\ t = 1, 2, \dots, T$), 通过最小化 $L_{final}(\theta_e)$ 得到 θ_e

4.3 模型优化

- 1. 尝试 3D 神经网络结合 One-Shot Scene-Specific Adaptation 方法

4.4 模型评价

- 1. 评价指标: $F1\ score, MSE$

5 预期成果

5.1 一个具有科普性质的网站

- 1. 功能: 输入一张肺部 CT 影像图, 给出是新冠病毒的匹配率, 可以查看无症状患者和发热患者的图片对比图 (给出两张图的相似程度), 给用户提