# 第四期 人脸/头部姿态估计+图像去噪

# 一，[img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation](https://arxiv.org/pdf/2012.07791.pdf)



fig1

# 1.论文目的和创新点：

## 1）目的：

估计人脸的姿态（3个欧拉角：pitch,yaw,roll）

## 2）创新点：

a）直接估计图像中人脸的3D姿态（6DoF），无需人脸检测阶段。

b）可以通过估计出的3D姿态计算出图像中人脸的包围盒。

# 2.网络结构：image.png

fig2

## 1）网络主干(backbone)：

resnet18

## 2）方法框架(approach)：

Faster-RCNN +FPN

# 3.方法流程(pipline)：

输入：二维平面图像

输出：6DoF：h\_img=(rx,ry,rz,x,y,z)，即3个欧拉角和3个平移量

流程：1）输入一张包含人脸的二维图像，

2）通过主干网络提取共享特征图

3）Region Proposal Net(RPN)选取出候选区域

4）对选取的候选框进行RoI Pooling，并通过两个分支(FC layer)分别预测人脸的得分和6DoF，同时

作为预测的副产物，可以输出人脸的包围盒。

# 4.损失函数：

损失函数包含两大部分：

## 1）RPN网络的proprosal Loss：

和faster rcnn 中定义一样包含classification 和 box regression loss

## 2）RoI head 的 RoI Loss

人脸得分loss + 6DoF loss + 2维投影点的校正loss

人脸得分loss：二元交叉熵

6DoF loss：MSE

2维投影点校正loss：L1

# 5.概念：

## 1）二维投影点：

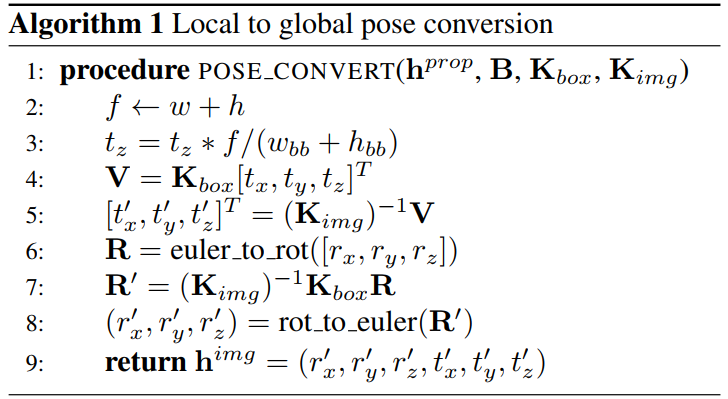
由于我们预测的是2维图像中人脸的姿态，因此计算是在2维空间。然而我们的真实标签信息给出的是3维空间的信息，因此需要做一个3D to 2D的投影变换： ------ eq1

公式中P: 3D人脸上的非共面点，R,t分别表示旋转矩阵和平移向量，Q：2D空间的投影点，K：论文提出的相机内参矩阵：，f=图像宽和高的和，cx,cy表示图像中心点坐标。

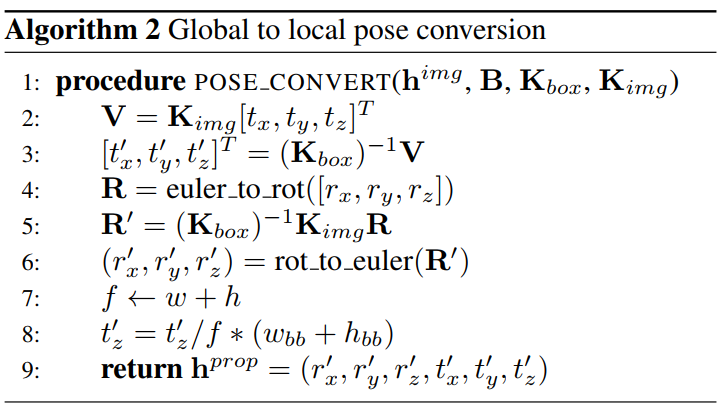
由于图像中人脸的大小是不一致的，通过这个公式就可以维持人脸和图像之间的一致性，即投影出的Q点保持在一个一致的尺度。

## 2）局部——全局的互转换：

由于我们的真实标签提供的是相对于全图（global）的信息，而我们的RPN和RoI head 的loss计算都是在proposal 层面进行运算（local）因此在计算RPN网络时需要进行全局到局部的转换：



相反在进行测试时候，我们需要输出的时相对于全图（global）的信息，因此需要进行局部到全局的转换：



## 3）副产物——人脸包围盒：

论文提出利用6DoF和eq1公式计算出bounding box，供proposal loss 计算和测试输出人脸box。这里的方法是使用68个3Dlandmark 和eq1进行投影，得到2D的landmark，然后计算其最小外接矩阵，即人脸box。这里的68个3D点是预先保存的一个模板。

## 4）真实标签的标注：

论文提出其方法是一种弱监督的方法，其弱表现在训练时候的label 并不是人工的标注，而是通过利用公开模型计算得到的。因为其真实标签需要6DoF信息，但是已有数据集并不满足要求，论文使用retinaFace在wide face人脸检测数据集上预测出landmark和box,通过这两个信息解算出6DoF标签。因此这是一个弱监督训练。这种方法能更好的根据任务控制人脸box和形状：

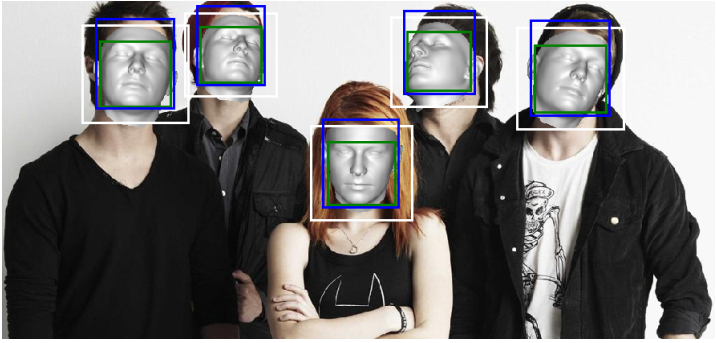


fig3

# 6.Demo:

[img2pose.avi](file:///C:/Program%20Files/Yuque/yuque-desktop/resources/app.asar/build/renderer/index.html#P1PxU)

# 二，[Neighbor2Neighbor: Self-Supervised Denoising from Single Noisy Images](https://arxiv.org/pdf/2101.02824.pdf)

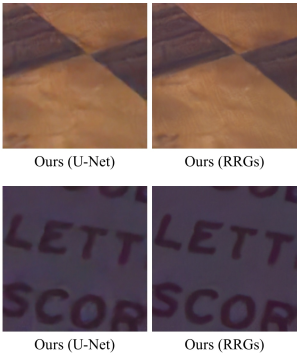
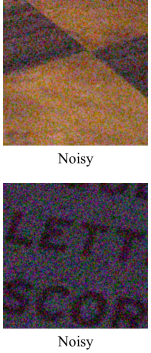


fig1

# 一，目的和创新：

## 1）目的：

去除图像中的噪声，提高图像质量

## 2）创新：

论文提出了一种自监督的方法进行训练，不需要提供干净的图像作为ground truth

# 二，网络结构：

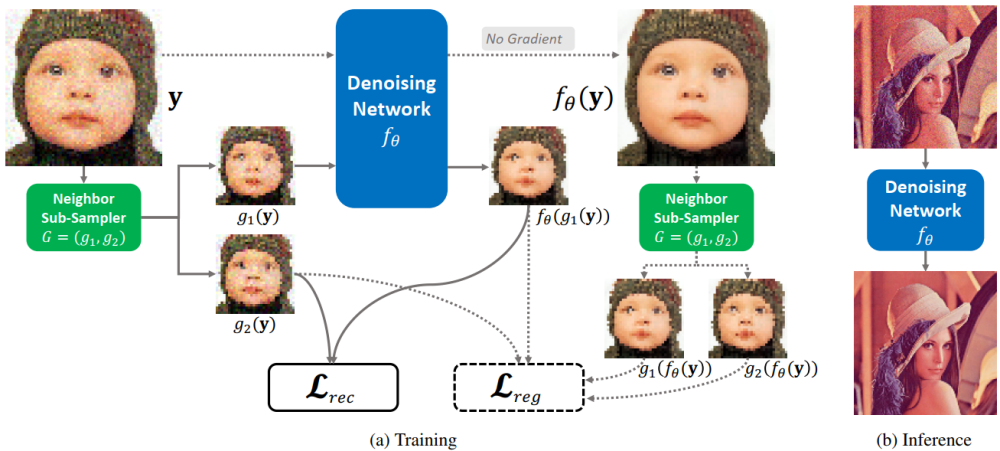
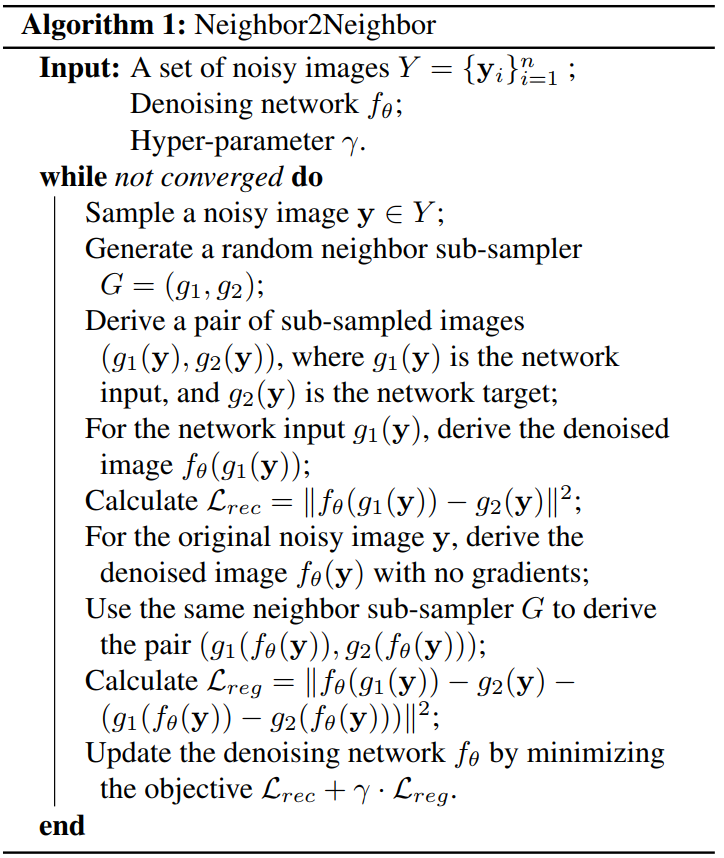
image.png

fig2

## 1) backbone:

Unet/RRG([cycleISP](https://openaccess.thecvf.com/content_CVPR_2020/papers/Zamir_CycleISP_Real_Image_Restoration_via_Improved_Data_Synthesis_CVPR_2020_paper.pdf))

## 2).training pipline：



# 三，训练损失函数：







# 四，随机下采样对的生成：

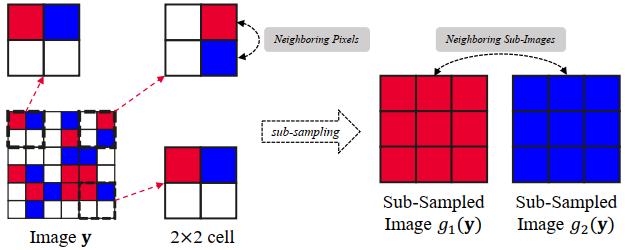


fig3

1）首先将图像划分成[W/k,H/k]的网格，k=2

2）在每个kxk的网格中随机选取两个不同的元素，分别组成下采样图

# 五，Demo:

| Noise | Clean |
| 0021_noisy.jpg | 0021_clean.jpg |