现有问题

- 类内变化很大

  > high intra-class variation

- 现有分类方法如SIFT, HOG, and LBP是手工方法，在受控条件下捕获的图像数据集上表现相当不错，但在具有更多图像变化和部分面部的更具挑战性的数据集上表现不佳

  > Traditional approaches for this problem rely on hand-crafted features such as SIFT, HOG, and LBP, followed by a classifier trained on a database of images or videos. Most of these works perform reasonably well on datasets of images captured in a controlled condition but fail to perform as well on more challenging datasets with more image variation and partial faces.

现有工作不足

- 改进深度神经网络依赖于
  - 添加更多的层/神经元
  - 促进网络中的梯度流动（例如，通过添加跳过层）
  - 更好的正则化（例如，谱归一化）

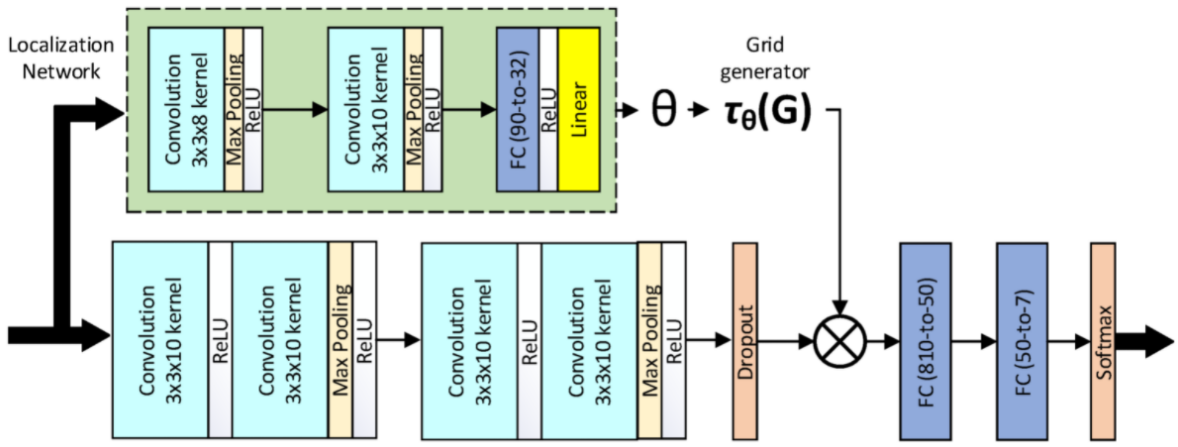
  这种方法对于具有大量类的分类有效，但是表情识别只有6类

改进方案

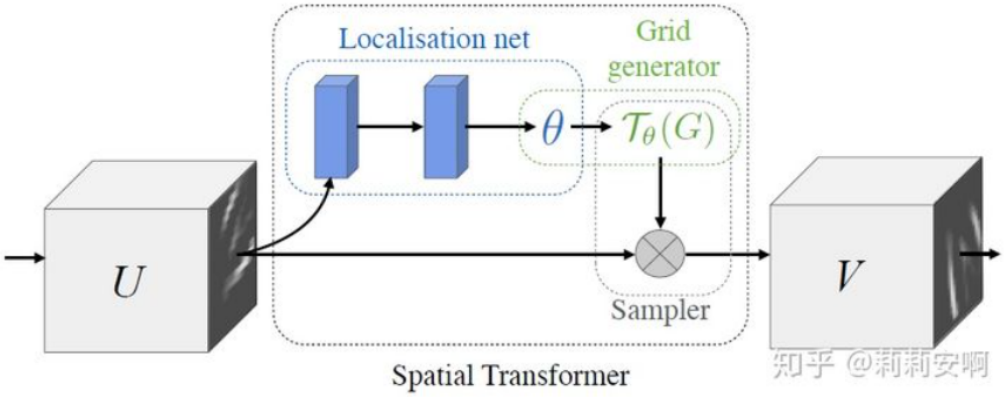
- 用少于十层的卷积网络
- 用spatial transformer network实现的attention

> using a convolutional network with less than 10 layers and attention (which is trained from scratch) is able to achieve promising results, presenting better results than state-of-the-art models for several databases.

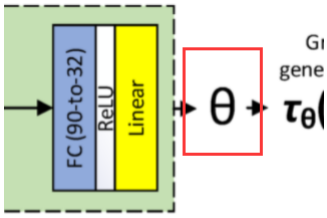
模型



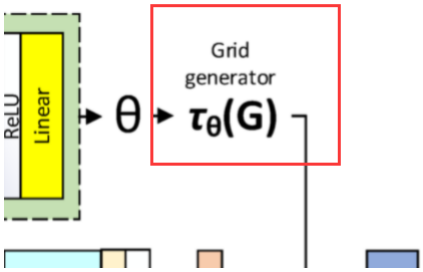
- **上面的是定位网络localization network**
  - 用处：定位到需要注意的部分
  - 结构：两层卷积网络，每个卷积网络有个最大池化和ReLU；还有两个全连接层
- **上面的定位网络部分其实是spatial transformer network（STN）**



  - STN包含三个部分：
    - Localisation Network-局部网络
    - Parameterised Sampling Grid-参数化网格采样
    - Differentiable Image Sampling-差分图像采样
  - 局部网络：生成一个θ，也就是上图所示θ



    - 它用于对原特征图U进行变换，是一个变换参数矩阵
    - 它有初始值，通过全局损失函数来调整到最好值
  - 参数化网格采样：局部网络生成后，原特征图U输入这个变换函数，就可以得到一个注意力在表情关键部位的输出特征图V了



  - 差分图像采样：计算对应点的值（好像是把小数处理为整数）
- **下面的是特征提取部分的网络**
  - 结构：由四个卷积层组成，每两个卷积层后跟一个最大池化层和一个ReLU)激活函数；然后是一个 dropout 层和两个全连接层
- **损失函数**
  - 随机梯度算法
  - 分类损失和正则化项的损失的两项总和

实验结果

- FER-2013——70.02%

Table 1. Classification accuracies on the FER 2013 dataset.

Method	Accuracy Rate
Unsupervised Domain Adaptation [47]	65.3%
Bag of Words [48]	67.4%
VGG+SVM [49]	66.31%
GoogleNet [50]	65.2%
FER on SoC [51]	66%
Mollahosseini et al. [9]	66.4%
The proposed algorithm	70.02%
Aff-Wild2 (VGG backbone) [52]	75%

- FERG_99.3%

Table 2. Classification accuracy on the FERG dataset.

Method	Accuracy Rate
DeepExpr [2]	89.02%
Ensemble Multi-feature [53]	97%
Adversarial NN [54]	98.2%
The proposed algorithm	99.3%

- 其他：

Table 3. Classification accuracy on the JAFFE dataset.

Method	Accuracy Rate
LBP+ORB features [55]	88.5%
Fisherface [56]	89.2%
Deep Features + HOG [57]	90.58%
Salient Facial Patch [58]	91.8%
CNN+SVM [59]	95.31%
The proposed algorithm	92.8%

Table 4. Classification accuracy on CK+.

Method	Accuracy Rate
MSR [60]	91.4%
3DCNN-DAP [61]	92.4%
LBP+ORB features [55]	93.2%
Inception [9]	93.2%
Deep Features + HOG [57]	94.17%
IB-CNN [37]	95.1%
IACNN [38]	95.37%
DTAGN [62]	97.2%
ST-RNN [63]	97.2%
PPDN [64]	97.3%
Dynamic cascaded classifier [65]	97.8%
The proposed algorithm	98.0%