Neural Network on Face Images

周伯威

2016213588  
The Institute of Computer Graphics and Computer-Aided Design

[hi@bowei.me](mailto:hi@bowei.me)

王安琪

2016213654  
The Institute of Computer Graphics and Computer-Aided Design

[waqthss2012@163.com](mailto:waqthss2012@163.com)

桑留芳

2016213652  
Institute of Information System and Engineering

[slf12thuss@163.com](mailto:slf12thuss@163.com)

**摘要**

我们完成了本次人工神经网络作业的必做内容及部分选做内容，两部分分别位于报告第**1**节与第**2**节。除此之外，在第**3**节中我们分析了表情识别任务正确率低下的可能原因。第**4**节中介绍了我们对于提高识别正确率做的尝试，包括使用一些简单的图像特征用作神经网络的输入。从结果上看，我们选取的图像特征可有效提升面部朝向识别、人脸识别的准确率，并在其他参数相同的前提下降低收敛所需轮数。

# 必做内容

## 代码实现

为了完成表情识别功能，我们修改了backprop\_face函数中神经网络各层的单元数、load\_target函数中神经网络输出层单元的输出以适应四种不同表情；修改了evaluate\_performance函数中的评估方法以正确地输出准确率。

需要注意的是，本文所介绍的功能均是在同一份代码中实现的，**附录A**中具体介绍了运行相关的选项。

## 问答题

我们使用了四个输出层单元，分别对应于*angry*、*happy*、*neutral*、*sad*四种表情。例如，结果为*sad*时，对应于四号输出层单元输出TARGET\_HIGH，而其他三个单元输出TARGET\_LOW。

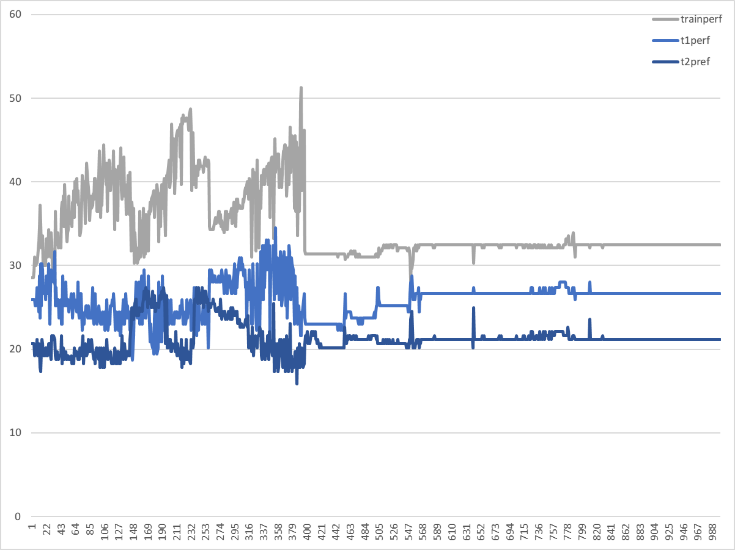


Figure 1 表情识别正确率变化。使用了的网络，，

对于表情识别任务，我们进行了轮的训练并得到如**Figure 1**、**Figure 2**所示的结果。在轮时，训练数据准确率达到最高的。而*test1*、*test2*数据集则分别在与轮时达到最高的与。从轮开始，三个数据集的正确率都趋于稳定。

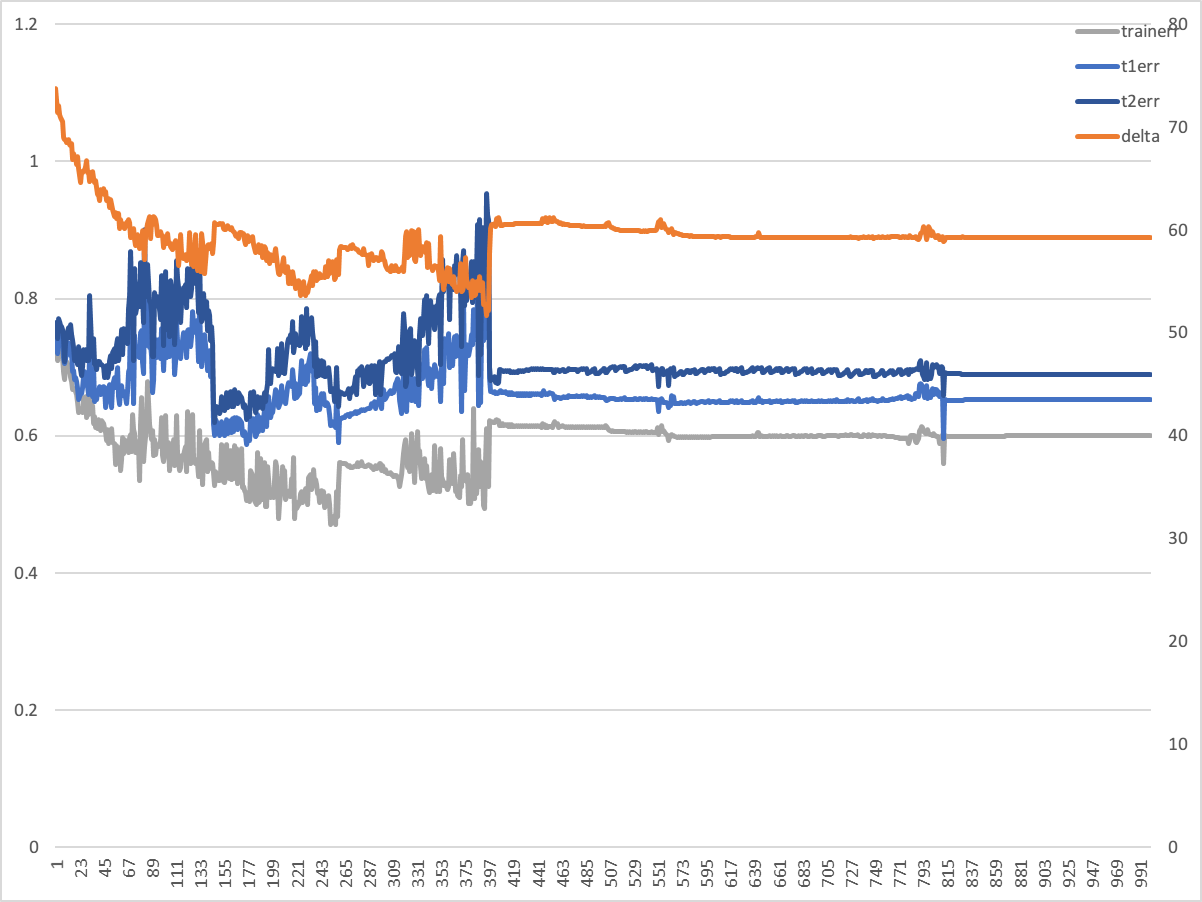


Figure 2 表情识别误差变化

我们尝试修改了隐藏单元个数、学习速率及动量，但识别准确率仍然较低，我们在第**3**节中分析了其中可能的原因。有趣的是，一次实验中，我们不小心将图像左上角的patch作为输入并得到了的正确率(高于上述实验)，进一步说明该方法存在一定问题。

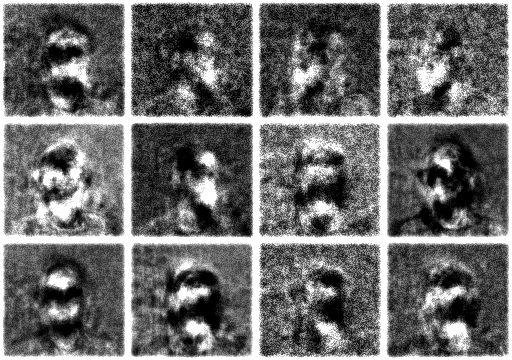


Figure 3 脸部朝向识别网络中，12个隐藏层单元接收的权重

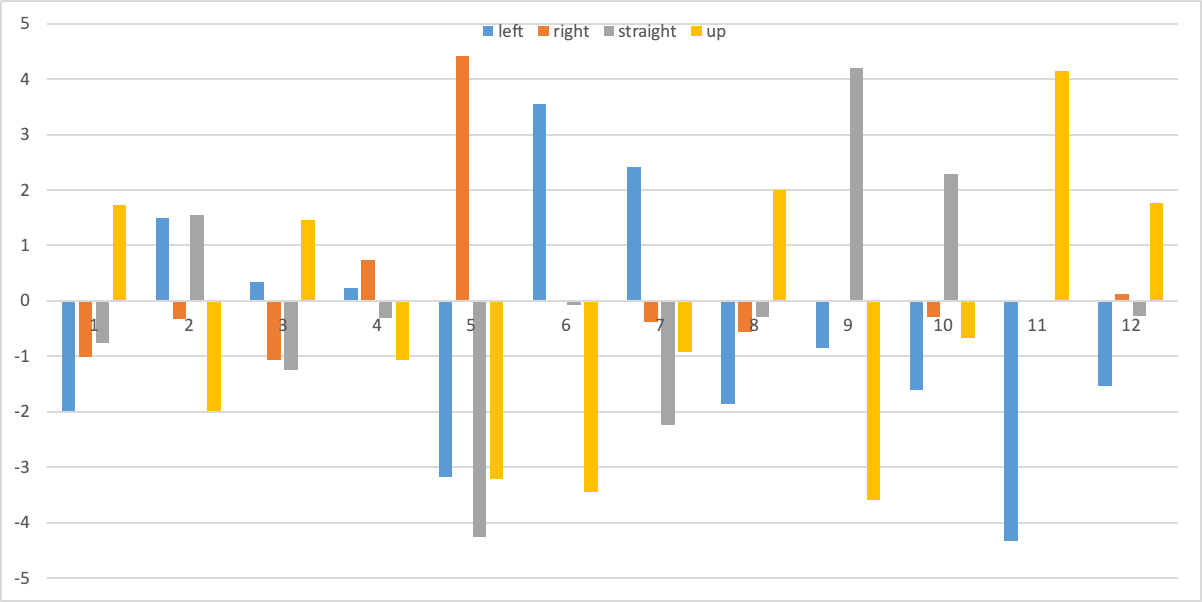


Figure 4脸部朝向识别网络中，12个隐藏层单元对四个输出层单元的贡献

因表情识别的不稳定性过高，我们对反向传播的观察是针对脸部朝向识别进行的。仅需要对表情识别程序稍加改动便可适用于脸部朝向识别。

**Figure 3**、**Figure 4**展示了脸部朝向识别网络的相关权重(图像进行了增强处理)。在**Figure 3**中，我们依稀可以认出第6幅图、第9幅图分别为向左、向前的朝向，在**Figure 4**中相应单元的高权重也印证了这点。然而，**Figure 3**中同样可辨认的图1(前)、图2(左)、图4(右)却没有在**Figure 4**中得到较高的权重。另一方面，**Figure 4**中5、11号单元具有较高权重，却未在**Figure 3**中得到相应体现。我们认为这体现了神经网络的不确定性，即神经网络的中间结果是难以分析与解释的。

# 选做内容

本节主要介绍我们实现的人脸识别功能，我们亦做了一些其他有趣的实验及分析，将在第**3**节、第**4**节介绍。

我们在隐藏层使用了个单元，学习速率与动量初始均设为。输出层使用了个单元，分别表示二十个人。

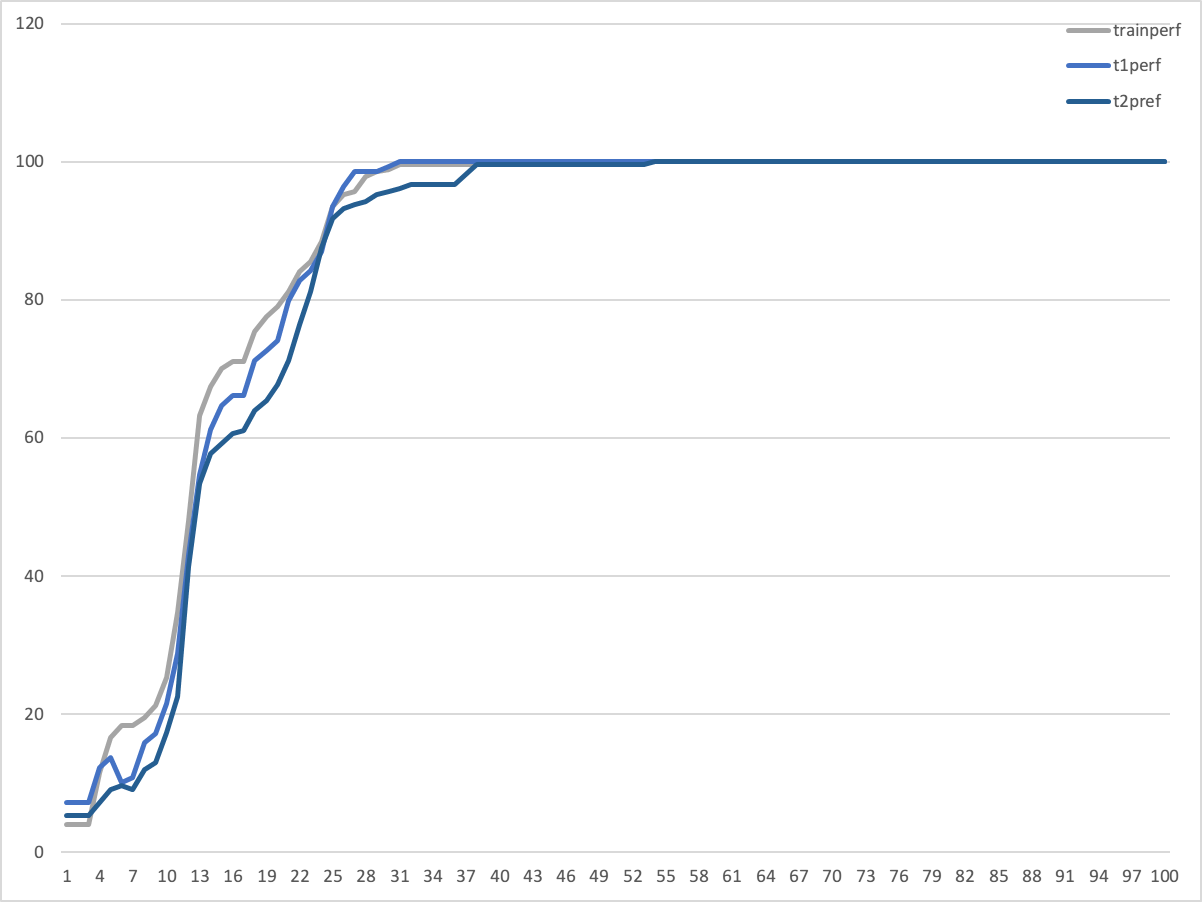


Figure 5 人脸识别(输入图像特征)正确率变化

以图像特征为输入(特征提取方法见第**4**节)，训练了的神经网络，其结果如**Figure5**、**Figure 6**所示。可见，从轮训练开始，三个数据集上的准确率均已达到，且误差变化没有上升趋势，可认为该网络能够很好地处理该数据集的人脸识别任务。从数据集上看，我们注意到不同人脸的区别主要是图像采集的背景，而非人脸，我们认为这是准确率如此之高的主要原因。

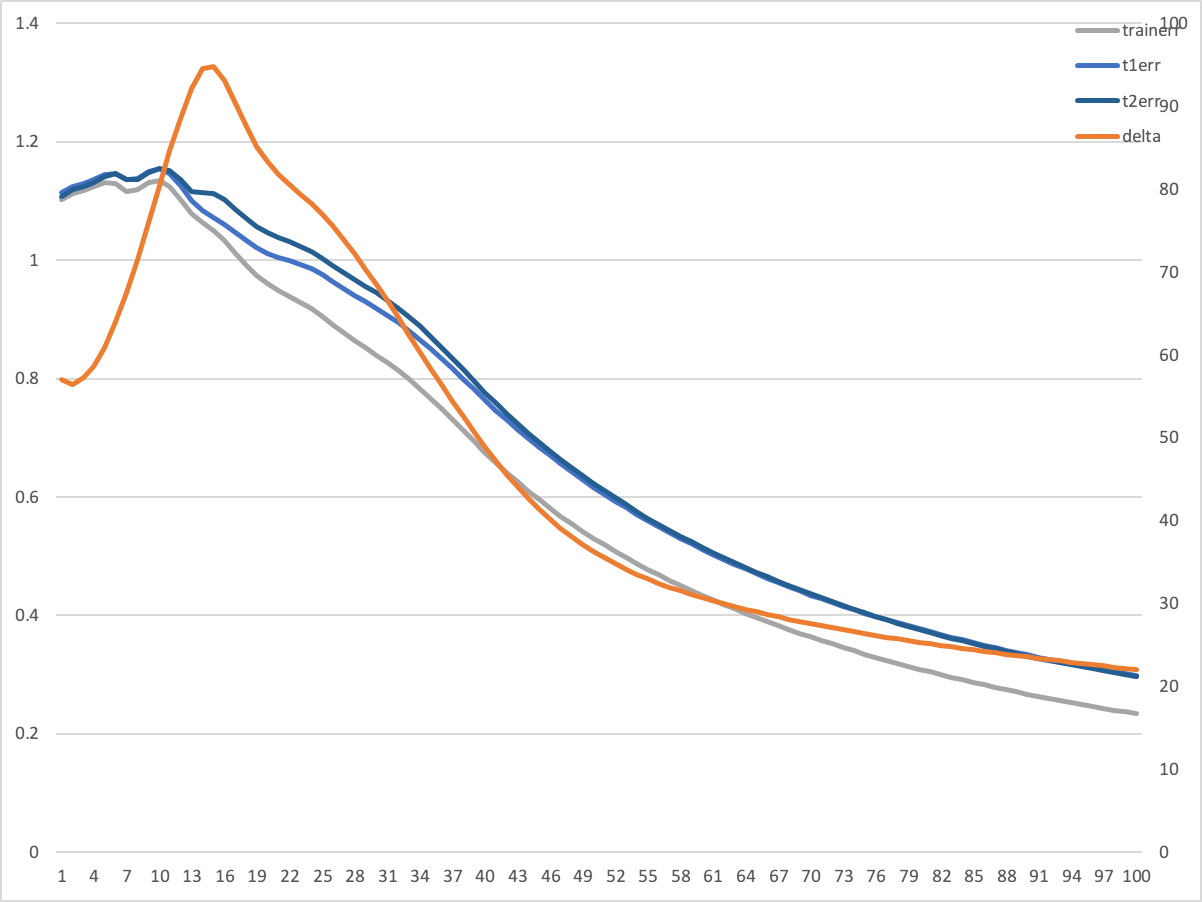


Figure 6 人脸识别(输入图像特征)误差变化

# 表情识别问题分析

相比于人脸识别、人脸朝向识别、墨镜识别，表情识别的训练过程收敛缓慢甚至不收敛，效果亦不如随机选取。我们认为其中包含多方面的原因。

**训练数据过少**。训练数据集仅包含277张图片，这对于机器学习方法来说是过少的，特别是对于不同图片变化巨大、且目标特征不明显的情况。

**图像中的人脸未对齐**。本次作业的神经网络是直接以图像作为输入的，这就意味着，输入是对位移十分敏感的。人脸的五官是表情识别的关键，不对齐也就无从进行进一步的识别。

**数据集包含四种脸部朝向**。数据集的图片数本身已经很少，却还包含四种脸部朝向。不同朝向的图片之间是无法“共享”表情信息的，这相当于图片数量减少到了原有的。

**表情难以以简单特征表达**。人脸、人脸朝向、墨镜均可用简单特征表达。例如，“戴墨镜的人的图像上部某些位置会很暗”、“大胡子danieln的图像下方显得很暗”等都是可用单个像素表达的。而表情则是一种高级属性，只能用高级特征描述。

**数据质量差**。我们注意到，数据采集员没有认真标注。比如你猜猜at33的下列四张图片分别是什么表情\_(:з」∠)\_



Figure 7 表情诡异的at33叔叔

理想的机器学习模型能够将测试数据映射至一个高维空间，在该空间中，不同类别的数据彼此之间相距较远，可简单分离。而本次作业中，各个图片特征(即像素亮度)分布近乎随机，自然难以分离出各种类别。

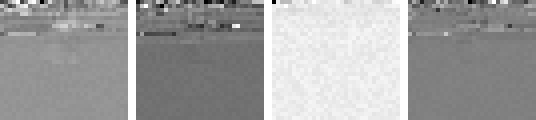


Figure 8 表情识别网络中，前四个隐藏层单元接收的权重

# 程序改进

为了提升识别正确率，我们最先想到的方法是修改神经网络参数。例如，将学习速率修改为可变的，在前面几轮训练中使用较高的学习速率，而后再将学习速率降低，以得到更为精准的模型。然而，无论是调整学习速率、动量，还是调整隐藏层单元个数，我们都未能有效提升识别正确率。

我们认为，原有网络使用图像作为输入，高达维的特征是不合理的。一个合理的想法是，可以使用数据集提供的边长缩小四倍的图像。我们使用这些图像也没能提高识别正确率，却可以大大降低训练时间。

于是，我们尝试了类似于通常的机器学习过程的，提取图像特征作为输入。首先使用了人脸识别常用的局部二值模式(LBP)特征[1]。该特征先将图像切分为一个个patch，对每个patch计算其中各个像素与周围一圈像素的大小关系(如**Figure 8**所示)，并计算该patch的直方图。将所有patch的直方图作为最终的特征向量。

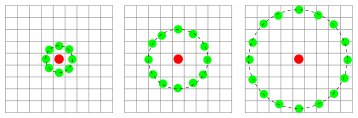


Figure 局部二值模式特征[1]示意

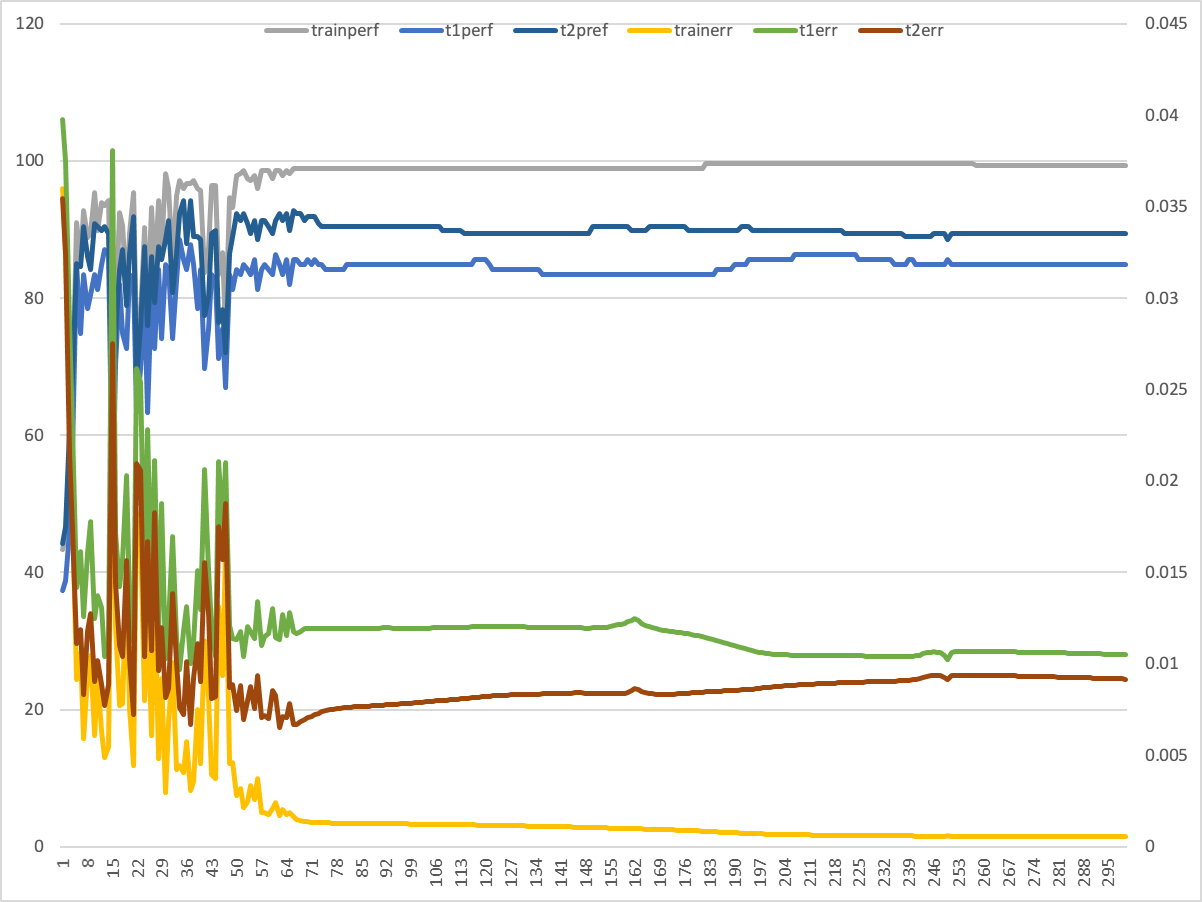


Figure 10 使用Naïve方法识别人脸朝向的训练曲线

**Figure 6**的人脸识别结果即是使用该方法提取的特征。的图像首先被扩展为像素，然后使用32的patch进行划分，相邻patch之间有一半重叠，共个patch，输入层单元个数即为个。

进一步地，我们参考了[2]、[3]中提到的其他特征。分别对每个patch进行计算六维的特征，并加入至原有的八维LBP特征向量中。此时，输入层单元个数为个(使用的patch)。由此特征进行人脸朝向识别效果略优于上述八维特征。两种方法**均显著优于**直接使用像素。如**Figure 11**所示，正确率约为，高于默认方法的。

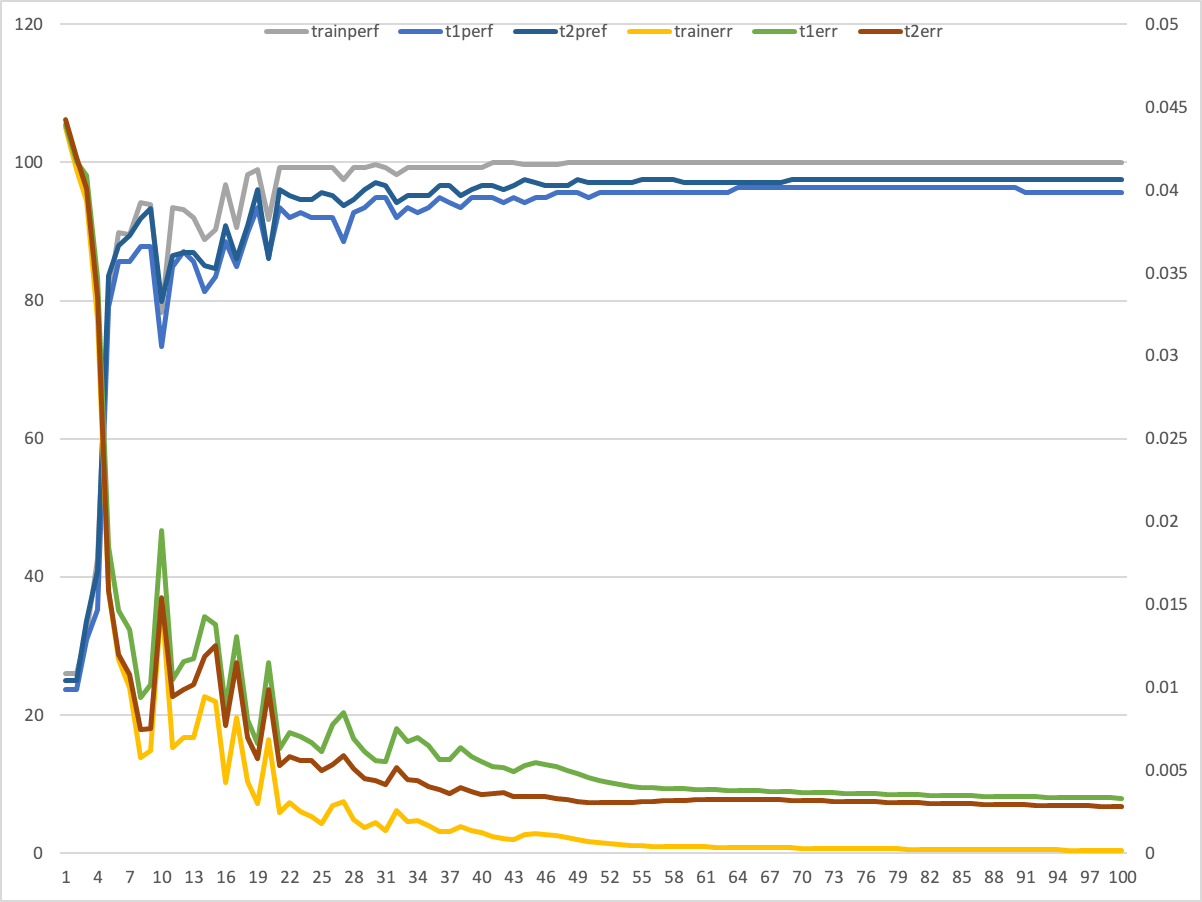


Figure 使用维特征识别人脸朝向

# 参考文献

1. Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." *IEEE transactions on pattern analysis and machine intelligence* 28.12 (2006): 2037-2041.
2. Picard, Rosalind W., Elias Vyzas, and Jennifer Healey. "Toward machine emotional intelligence: Analysis of affective physiological state." *IEEE transactions on pattern analysis and machine intelligence* 23.10 (2001): 1175-1191.
3. Wagner, Johannes, Jonghwa Kim, and Elisabeth André. "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification." *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on.* IEEE, 2005.

# 附录A 程序设置方法

在backprop.h中，我们设置了用于控制条件编译的定义。

#define TARGET\_glasses 0 /\* 眼镜 \*/

#define TARGET\_emotion 1 /\* 表情 \*/

#define TARGET\_head 2 /\* 头部朝向 \*/

#define TARGET\_who 3 /\* 人脸 \*/

#define TARGET TARGET\_who /\* 当前选择人脸 \*/

#define NAIVE /\* 直接以图像输入 \*/

// #define COMPLEX\_METHOD /\* 使用14维特征 \*/

(and something interesting…我们发现在文档中搜索”pose”可以找到好多处相关的内容，应该是老师忘改成”表情”了吧。。或者就是故意留着让我们发现的哈哈)