# 基于二值化密集卷积神经网络的表情识别算法

温光照，徐诗楠，马云鹤，王小波
（南京理工大学计算机与工程学院，江苏 南京 210094）

**摘要：**人脸表情识别已成为人工智能领域的重要研究课题，但传统的卷积神经网络需要庞大的计算资源使得其应用受限，而二值化卷积神经网络可通过快速与或运算代替原本的浮点乘法运算，大大降低了算法对计算资源的需求。本文提出了一种基于数据增强和二值化卷积神经网络的人脸表情识别算法，通过均值估计，在 FER2013 数据集上达到了 66.15%的识别率，超越了部分基于浮点乘积运算的卷积网络，为表情识别算法移植到小型设备中提供了可能。

**关键词：**深度学习；数据增强；二值化；密集卷积神经网络；表情识别

# Expression Recognition Algorithm Based on Binary Dense Convolution Neural Network

Wen Guangzhao, Xu Shinan, Ma Yunhe, Wang Xiaobo
*(School of Computer Science and Engineering，Nanjing University of Science and Technology, Nanjing 210094, China)*

**Abstract:** The research of facial expression recognition has become an important topic in the field of artificial intelligence. However, the requirement for huge computing resources has limited the application of traditional convolutional neural networks. Since the binary neural network replaces the floating point multiplication arithmetic by fast AND OR arithmetic, the need of computing resources can be greatly reduced. In this paper, we propose a facial expression recognition algorithm based on data enhancement and binary convolutional neural network, and 66.15% expression recognition accuracy is obtained on the dataset FER2013. The algorithm has surpassed some convolutional neural network algorithms based on floating point multiplication arithmetic, which makes it possible to transplant expression recognition algorithms into small devices.

**Key words:** deep learning; data enhancement; binarization; dense convolutional neural network; expression recognition

## 1 引言

神经网络算法一直是研究人脸表情识别有力的工具，但是基于浮点数乘法运算的复杂神经网络模型如 VGG[1]和 Resnet[2]等，需要消耗大量的计算资源和内存资源，严重阻碍了其在小型设备上的应用，也使得人工智能普及到生活中的难度大大增加。而通过神经网络二值化，我们可以将原本的浮点数运算，转化更快的与或运算。因此，二值化神经网络模型以其较高的模型压缩率和较快的前向传播计算速度，近几年受到格外的重视和发展，成为神经网络模型研究中一个非常热门的研究方向。

由于二值化神经网络模型精确度较低，识别率通常会低于一般基于浮点运算的神经网络模型。本文提出了一种基于数据增强和二值化密集卷积神经网络的人脸表情识别算法，通过均值估计，使得二值化神经网络模型的识别率达到甚至超过部分基于浮点数运算的卷积神经网络。

## 2 相关工作

### 2.1 数据增强

深度学习需要大量的数据支撑，数据量过少常常会造成过拟合等问题，因而数据增强的概念被提出。数据增强有许多常用的方法，其核心在于对原图片做裁剪、缩放、彩色变换、翻转等操作，生成新的图片数据集，以达到扩充数据库的工作。2012 年 Alex 等人在论文 ImageNet Classification with Deep

Convolutional Neural Networks[3]中提出了一种数据增强方式，训练时对 256*256 的图片进行随机裁剪到 224*224，然后允许水平翻转，将原样本数倍增到$(256-224)^2*2=2048$，测试时对左上、右上、左下、右下、中央做 5 次裁剪，然后翻转，共 10 张新图，之后对 10 张图片的输出结果做平均作为识别的最终结果。利用这种技巧和 AlexNet, Alex 在 ILSVRC 2012 竞赛中的 top-5 识别错误率降低到了 15.3%，比上一年冠军的识别错误率降低了十几个百分点，远超同年的第二名。

## 2.2 卷积网络的基本原理

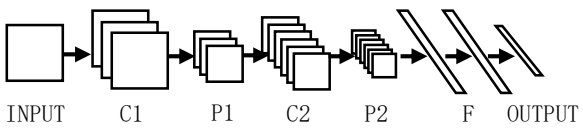


图 1 卷积神经网络结构图

Fig.1 Convolution Neural Network structure

图 1 为卷积神经网络的结构图，其中， C 代表卷积层，P 代表池化层，F 代表全连接层。卷积神经网络的第一层是输入层，其后是若干个卷积层和池化层，最后是连续的几个全连接层。最后一个全连层连接输出的单元数即为分类数，该层也称为输出层。而除了输入和输出层外的全部层都称为隐层，隐层之间存在一个规范化和激活操作，将运算得到的结果通过相应的规范化函数和激活函数输出从而成为这一层的特征图。下面有关于卷积层和池化层的详细介绍。

### 2.2.1 卷积层

相比一般的全连接网络，卷积神经网络通过局部连接和权值共享大大减少了计算一个特征值所需要的连接数，降低了计算的复杂性。在一个卷积层中，可学习的卷积核将上一层的特征图进行卷积，然后用一个规范化函数和激活函数得到一个对应的特征图，如下表达式所示：

$$\begin{cases} x_l^{k,p,q} = \sum\limits_{(i,j)\in M_{l-1}^{k,p,q}} a_{l-1}^{k,i,j} * W_l^{k,p,q,i,j} + b_l^{k,p,q} \\ a_l^{k,p,q} = f_{activation}\left(f_{norm}\left(x_l^{k,p,q}\right)\right) \end{cases} \quad (1)$$

其中$x_l^{k,p,q}$代表第l层第k个卷积特征图的第$(p,q)$个特征单元，$a_l^{k,p,q}$代表第l层第k个卷特征图的第$(p,q)$个特征单元规范化和激活后的最终输出，$M_{l-1}^{k,p,q}$为第$(l-1)$层上要卷积第$(p,q)$个特征子图的范围，$W_l^{k,p,q}$代表第l层第k个卷积特征图的第$(p,q)$个特征单元对应的卷积核，$f_{norm}()$代表规范化函数，$f_{activation}()$代表激活函数，$b_l^{k,p,q}$代表第l层第k个卷积特征图的第$(p,q)$个特征单元对应的偏置项。

### 2.2.2 池化层

在池化层中，通常采用最大池化层技术，获得$n*n$区域内的最大值作为池化的输出。在卷积过程的前向传播中，不仅要记录该最大值，还要记录该最大值所在输入数据中的位置，以在反向传播中把梯度值传到对应最大值所在的位置。

## 2.3 密集连接

Gao Huang 等人于 2017 年提出的密集连接网络 DenseNet[4]其核心是在保证网络中层与层之间最大程度信息传输的前提下，直接将所有层连接起来，即每一层的输入都源自于之前所有层的输出，表示为：

$$x^l = H([x^0, x^1, \ldots, x^{l-1}]) \quad (2)$$

其中$[x^0, x^1, \ldots, x^{l-1}]$表示将 0 到$(l-1)$层的输出特征图做连接，即将各层的通道合并，而 H()表示在通道合并之后做的规范化操作和非线性变换操作。

DenseNet 这样的处理，使得密集连接可以减轻网络传播过程中梯度消失的问题，更加有效地利用提取出来的特征值。同时更宽而不是更深的网络结构，也将在一定程度上减少我们所需要训练的参数数量，在保证网络良好拟合性的情况下，减小训练的复杂度，加快训练的速度。

## 2.4 二值化

在神经网络权重和激活函数的二值化方面，本文参考了文献以下提供的方法。

Expectation Backpropagation(EBP)[6] 是一种在训练过程中利用真实权重值+二值化激活函数，在测试过程中利用二值化权重值+二值化激活值的二值化神经网络训练算法。

Binary Connect[7]是另一种训练二值化权重值和激活值的神经网络算法，其在训练过程中利用二值化权重+真实权重+真实激活值，在测试过程中利用二值化权值+真实激活函数。

受到以上两种训练方式的启发，Bengio 组的 Binarized Neural Networks (BNNs)[5]提出了一种训练和测试过程中都使用二值化权重值和二值化激活值的二值化神经网络训练算法。2017 年，Sun 等

基于该二值化方法，提出了基于多层连接的稠密局部二值模式（MDLBP）和二值自动编码器（BAE）的二值化全连接神经网络算法[8]，并引入残差。在保留 BNN 快速收敛的特性的同时，在自然人脸库上接近甚至超过一些传统卷积神经网络的准确率。而本文的二值化训练算法也是在 BNNs 训练算法上做的一个扩展。

BNNs 的核心是权值和激活值都被限制在+1 和-1 之间，这样原本繁杂的浮点数运算就可以被简单的与或运算代替，大大提高了网络前向传播的速度。而在最后全连接层中，我们定义$a_l$,$W_l$,$b_l$分别为第l层的激活值，权重值和偏置参数，L为网络的总层数，则有：

$$a_l = \begin{cases} sgn(f_{norm}(sgn(W_l)a_{l-1} + b_l)) & ,l \le L \\ f_{norm}(sgn(W_l)a_{l-1} + b_l) & ,l = L \end{cases} \quad (3)$$

其中：

$$\mathrm{sgn(x)} = \begin{cases} +1, x \ge 0 \\ -1, x < 0 \end{cases} \quad (4)$$

根据网络最后一层的输出，定义最终的损失函数为：

$$\mathrm{l}(a_L, \mathrm{y}) = \frac{1}{k}\sum_{i=1}^{k} \max\left(0, 1 - y_i a_L^i\right) \quad (5)$$

其中k是$a_L$的维度，也就是对应的分类数，y 是采用-1 和+1 的 one-hot 形式表示的二值化标签向量。如果输出值和预期的标签值相同，就有$y_i a_L^i = 1$，也就有对应的损失函值为 0，即模型完全拟合。

假设第 i 次的训练样本为$(x^i, y^i)$，那么就把 BNNs 的训练问题归结为：

$$\min_{W_1,b_1,W_2,b_2 \ldots W_L,b_L,} \frac{1}{N}\sum_{i=1}^{N} l\left(a_L^i, y^i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} l\left(f\left(\ldots f\left(f\left(x^i, W_1, b_1,\right), W_2, b_2\right) \ldots, W_L, b_L\right), y^i\right) \quad (6)$$

保留二值化权值和真实权值是本网络训练的核心，当网络进行更新的时候我们更新真实的权值，而网络的前向传播和后向传播时利用的是我们二值化的权值。

# 3 主要方法

## 3.1 基于五图连接的数据增强

在吸取 2.1 中数据增强的经验后，我们也采用裁剪的方式来对原数据库进行数据增强，如图 2 所示。在训练集和测试集上同时将原 48*48 的灰度人脸表情图，分别从左上，右上，左下，右下和中央截取五张 42*42 的人脸表情图做连接，生成增强后的表情数据库。可以发现该尺度的裁剪只是将人脸的中心做了一个偏移，并未丢失人脸表情信息。从人眼的角度，仍然可以清晰地看出图片对应的表情。

图 2 人脸图片五裁剪

Fig.2 Five crop of face images

## 3.2 网络的输入部分的处理

根据像素值直接对原图做一般的二值化处理可能会丢失大量的图片细节，从而降低表情的识别率，所以我们的模型的输入部分选择了对图片做[-1,1]区间的归一化处理而非简单的二值化，对应有如下表达式：

$$\mathrm{x} = \frac{2 * \mathrm{x} - 255}{255} \quad (7)$$

虽然图片的输入部分是非二值化的，但是激活之后每一层的卷积部分间传递的参数都是二值化的，而在我们提出的网络中网络第一层的特征值连接数是最少的，因而对整个二值化神经网络没有很大的影响。

## 3.3 二值化密集卷积神经网络

在 2.3 和 2.4 中提到的密集连接和二值化神经网络的基础上，我们提出了一种轻量级的密集二值化卷积神经网络。如图 3 所示是我们采用的密集连接后单个 Dense Block 的结构图。

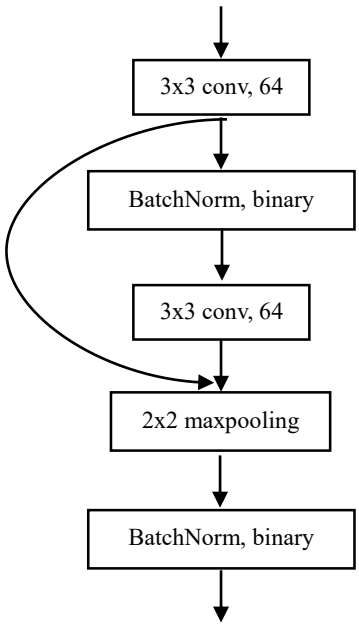


图 3 Dense Block 结构图

Fig.3 Dense Block structure

和 DenseNet 相同的是我们也是由多个 Dense Block 串接构成的网络结构，但是相比 DenseNet 我们的 Block 更小，我们的 Dense Block 仅有两层卷积层，在 Block 内第二层的输出来源于第一层的和本层输出的连接。本文采用的规范化方式是 Google 小组在 2015 年时提出的 BatchNormlization[9]，即使数据做均值为 0，方差为一的归一化，以加快网络训练速度，减小权重值的尺度的影响，同时也有模型正则化的作用。非线性化过程就是对模型进行二值化的过程，二值化手段和 2.4 中提到的二值化手段相同，除了最后一层以外，每层卷积或全连接操作后，我都对采用规范化后的参数做二值化操作，采用二值化传递参数，同时保留非二值化参数用作更新处理 。相比 DenseNet，我们消除了 1x1 的卷积层，一方面是我们网络并没有那么深，另一方面 1x1 的卷积层对于二值化神经网络来说意义并不是很大。同时我们的 Block 之间的卷积层的通道数是随着 Block 加深而递增的，这样做可以减少我们提取的特征数的下降速度，从而在一定程度上更加细致的提取人脸表情特征。按照上述结构根据卷积的 Block 数，可以构建表 1 中的几个网络。

表 1. 二值化密集神经网络结构

Tab.1 Binary Dense Neural Network structure

| Layers | Output Size | BDNN2 | BDNN3 | BDNN4 |
|---|---|---|---|---|
| Dense Block | 42x42 | (3x3 conv,64)x2 | | |
| Transition Layer | 21x21 | 2x2 max pool | | |
| Dense Block | 21x21 | (3x3 conv,128)x2 | | |
| Transition Layer | 10x10 | 2x2 max pool | | |
| Dense Block | 10x10 | | (3x3 conv,256)x2 | |
| Transition Layer | 5x5 | | 2x2 max pool | |
| Dense Block | 5x5 | | | (3x3 conv,512)x2 |
| Transition Layer | 2x2 | | | 2x2 max pool |
| Full Connect Layer | | (1024 fully-connected)x2 | | |
| Classifiction Layer | | 7 fully-connected | | |

## 3. 4 网络的输出部分的处理

在本文网络的输出部分中，我们将五张来自相同原像的人脸图片的输出值做平均，作为对应的原图片的表情输出。

假设我们的算法完全正确，那么五张图片的最终的输出一定是相同的，即五张图的平均值一定也就是对应原来的人脸表情。但是目前对于人脸表情的结果仍然无法保证绝对正确，因此，在网络模型中，人脸位置的空间的特性对人脸表情识别的结果的影响仍然存在。尤其是在二值化过程中，利用一个小型的二值化模型去拟合表情识别模型必然有一定的偏差。

但是同样的我们有理由相信我们的模型具有较好的拟合能力，能大概率消除表情识别中的空间特

性的影响，那么的话，我们可以认为我们的五图中存在的是小概率的偏差，也就意味着，我们如果通过五图识别结果的均值作为图片的输出结果，可以利用大概率的正确在一定程度上纠正小概率的偏差，从而使人脸表情识别的结果为大概率正确的结果，进而一定程度上提高人脸表情识别的识别率。

### 3.5 算法流程

如图 4 所示有我们的算法流程图。

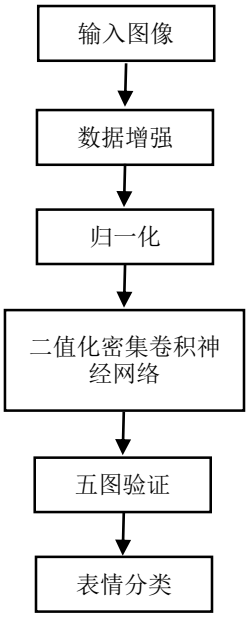


图 4 算法流程

Fig.4 Algorithm Flow

## 4 实验部分

### 4.1 人脸数据集——FER2013

FER2013 数据集由面部 48*48 像素的灰度图像组成，面部已经定位并居中。目前，在该数据集上人肉眼能够达到的识别率在$65 \pm 5\%$[10]。该数据集中训练集包含 28709 个示例，验证集包含 3589 个示例，测试集包括 3589 个示例。其人脸表情标签为愤怒、厌恶、害怕、高兴、悲伤、惊讶、中性 7 个分类，训练集具体样本分布如表 2 所示。

表 2. FER2013 训练集样本分布情况

Tab.2 Distribution of FER2013 Training Dataset

| 类型 | 样本数 |
|---|---|
| 愤怒 | 3995 |
| 厌恶 | 436 |
| 害怕 | 4097 |
| 高兴 | 7215 |
| 悲伤 | 4830 |
| 惊讶 | 3171 |
| 中性 | 4965 |
| 合计 | 28709 |

### 4.2 实验环境配置

硬件配置：Intel i7-4790k，主频 4.0GHz，16G 内存，NVIDIA GTX1060 6GB。

软件配置：Ubuntu 16.04，Cuda 8.0 GPU 并行计算库，Theano 深度学习框架。

### 4.3 实验结果的对比分析

本实验主要在 FER2013 人脸数据集上进行。首先，我们根据表 1 中提出的几个不同 Block 数的网络结构做了对比实验，实验结果如表 3 所示。可一发现深层的网络 BDNN4 达到较好的效果。

表 3. 在 FER2013 上不同网络结构的对比试验

Tab.3 Comparative experiment of different neural network structures on FER2013

| 数据库 | 网络结构 | 密集连接 | 五图验证 | 识别率 |
|---|---|---|---|---|
| FER2013 | BDNN2 | 是 | 是 | 61.50% |
| FER2013 | BDNN3 | 是 | 是 | 65.51% |
| FER2013 | BDNN4 | 是 | 是 | 66.15% |

同时为了测试五图均值验证的有效性，我们利用 BDNN4 基本结构框架，在 FER2013 的数据上，控制变量分别做了如下两个对比实验，一个是在 BDNN4 的基础上消除了密集连接部分，表示为 CBNN4，另一个是在 BDNN4 实验的基础上消除了五图均值验证的部分，实验结果如表 4 所示。

表 4. 五图验证和密集连接的有效性对比试验

Tab.4 Comparison experiment of five figure validation and dense connections

| 数据库 | 网络结构 | 密集连接 | 五图验证 | 识别率 |
|---|---|---|---|---|
| FER2013 | BDNN4 | 是 | 是 | 66.15% |
| FER2013 | BCNN4 | 否 | 是 | 64.90% |
| FER2013 | BDNN4 | 是 | 否 | 61.31% |

根据实验的数据来看，在 FER2013 的数据库上，相同的网络模型之下，采用数据增强的实验结果比起不采用数据增强的实验结果提高了将近 5%，这也就证明了我们假设的正确性。也就是说，五图均值验证能够利用大概率的正确性消除一部分小概率的错误，从而达到提高表情识别准确率的目的，其中二值化网络中是相当有效的。

而小段的密集连接也一定程度上提高了超过 1%的表情的识别率，也说明了小型的密集连接在我们模型中的有效性。

同时我们也查看了近两年国内发表深度学习

的论文中对 FER2013 的识别率，如表 5 所示。

表 5 近年发表的国内论文中对 FER2013 的识别率

Tab.5 Accuracy of FER2013 in domestic papers published in recent years

| 方法 | 识别率 |
|---|---|
| 程曦的深度卷积神经网络[11] | 66.59% |
| 本文 BDNN4 | 66.15% |
| 翟懿奎等的迁移卷积神经网络[12] | 61.59% |

可以发现相较于部分非二值化的卷积神经网络，我们的二值化密集卷积神经网络的识别率甚至有所提升，证明了我们的二值化密集卷积神经网络能拥有着不亚于一般基于浮点数运算的卷积神经网络的拟合能力，而比起程曦的深度卷积神经网络，我们的偏差不到0.5%，但是我们的网络模型是基于二值化的神经网络模型，网络参数的传递过程中我们可以通过与或运算来代替一般浮点数乘积运算，在网络的计算速度和对内存资源的需求上有着的巨大的优势。

## 5 结论

本文中，我们在将二值化加入到卷积神经网络的基础上，为了增强网络的拟合能力，引入了小型的密集连接，同时提出了通过裁剪五图，即从一张原图中裁剪出五张略有不同的图像，起到了数据增强的目的。此外，我们认为人脸图片在空间位置晃动的时候，仍能大概率地被我们的人脸表情模型所识别，这样我们就可以通过五张原像相同图片的输出结果的均值作为我们表情输出的结果，以大概率的正确值去纠正小概率的错误，从而提高表情的识别率。实验结果中我们也验证该方法的有效性，在 FER2013 的数据库上达到了 66.15%的识别率，使我们的二值化卷积神经网络有着不弱于甚至强于传统非二值化卷积神经网络的识别效果。而二值化神经网络的运算过程可以用与或运算来代替繁杂的浮点数运算，大大降低了运算对计算和存储资源的需求。

## 参考文献(Reference)


[1] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL]. (2015-4-10). https://arxiv.org/pdf/1409.1556.pdf.

[2] K. He, X. Zhang, S. Ren. Deep Residual Learning for Image Recognition[EB/OL]. (2015-12-10). https://arxiv.org/abs/1512.03385.pdf.

[3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[4] GaoHuang，ZhuangLiu，Laurens van der Maaten. Densely Connected Convolutional Networks[EB/OL]. (2018-1-28). https://arxiv.org/pdf/1608.06993.pdf.

[5] M. Courbariaux, Y. Bengio. Binarized Neural Networks: Training deep neural networks with weights and activations constrained to +1 or -1[EB/OL]. (2016-3-17). https://arxiv.org/pdf/1602.02830.pdf.

[6] Soudry D, Hubara I, Meir R. Expectation Backpropagation: parameter-free training of multilayer neural networks with continuous or discrete weights[C]// International Conference on Neural Information Processing Systems. MIT Press, 2014: 963-971.

[7] Courbariaux M, Bengio Y, David J P. BinaryConnect: training deep neural networks with binary weights during propagations[J]. 2016: 3123-3131.

[8] Sun W, Zhao H, Jin Z. An Efficient Unconstrained Facial Expression Recognition Algorithm based on Stack Binarized Auto-encoders and Binarized Neural Networks[J]. Neurocomputing, 2017, (267): 385-395.

[9] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015:448-456.

[10] Goodfellow I J, Erhan D, Carrier P L, et al. Challenges in Representation Learning: A Report on Three Machine Learning Contests[C]// International Conference on Neural Information Processing. Springer, Berlin, Heidelberg, 2013:117-124.

[11] 程曦. 基于深度学习的表情识别方法研究[D]. 长春工业大学, 2017. (CHENG Xi. The Research of Facial Expression Recognition Based on Deep Learning[D]. Changchun University of Technology, 2017.)

[12] 方圆. 基于卷积神经网络的人脸表情识别研究[D]. 西安电子科技大学, 2017. (FANG Yuan. Research of Facial Expression Recognition Based on Convolutional Neural Network[D]. Xidian University, 2017.)


## 作者简介：


温光照，男，本科生。研究领域：深度学习

徐诗楠，男，本科生。研究领域：深度学习

马云鹤，男，本科生。研究领域：深度学习

王小波，男，本科生。研究领域：深度学习