

基于改进的卷积神经网络的人群密度估计

贾翻连, 张丽红

(山西大学物理电子工程学院, 山西 太原 030006)

摘要: 人群密度估计是视频监控的一个研究热点, 为了得到更为准确的估计率, 研究人员将卷积神经网络应用到人群密度估计中。卷积神经网络在特征提取时能够高效自适应学习深层次的特征, 体现了其在深度学习领域的优越性, 但在预处理时会出现振荡现象, 且卷积层与子采样层间特征图的大小匹配会影响计算速度和时间。本文提出离散小波变换替换卷积神经网络中的子采样层, 并对网络中的权重矩阵重新计算, 通过权重自适应改善预处理时的振荡现象, 提高卷积网络中特征图大小的匹配度, 并将之应用到人群密度估计, 有效地提高数据间的相关性, 增强网络的学习能力, 提高人群密度等级分类的准确率。实验表明, 改进后的网络具有较好的学习及分类效果和鲁棒性, 对人群密度能够进行较为准确和快速的估计。

关键词: 人群密度; 深度学习; 小波变换; 神经网络; 权重自适应; 特征提取

中图分类号: TP183

文献标志码: A

Population density estimation based on an improved convolution neural network

JIA Fanlian, ZHANG Lihong

(School of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: Crowd density estimation is a research hotspot in video surveillance. In order to get a more accurate estimation rate, researchers apply convolutional neural network to crowd density estimation. Convolutional neural network can efficiently and adaptively learn deep characteristics in the feature extraction process, and demonstrates its superiority in depth learning areas, but in the pretreatment oscillation occurs, and the size matching for feature map of convolutional layer and sub sampling layer will affect the calculation speed and time. This paper adopts discrete wavelet transform to replace the sub sampling layer in convolutional neural network, and the weight matrix in the network is adapted to improve the calculation, the phenomenon of oscillation is avoided in pretreatment by adaptive weight. This improves the matching degree of feature map size in convolutional networks, and it is applied in the estimation of crowd density, effectively improves the data correlation and enhances the learning ability of the network, and also improves the accuracy of classification of population density. The experiment shows that the improved network has better learning and classification effect and robustness, and can be used to estimate the population density more accurately and quickly.

Key words: population density; deep learning; wavelet transform; neural network; adaptive weight; feature extraction

0 引言

近年来, 人群密度估计成为智能视频分析的一个研究热点, 并且越来越多的学者和研究人员使用基于深度学习和机器学习的各种方法研究人群密度。其中, Davies^[1]研究指出, 人群密度与当前帧的前景像素面积近乎呈线性关系, 由此提出了基于像素统计特征的人群密度估计算法。随后, 基于改进的混合神经网络密度估计算法^[2]被提出。刘曼等提出一种基于可变矩形框的人群密度数值估计算法^[3]; 时增林等提出了基于序的空间金字塔池化网络的人群计数方法^[4]; 张洞明等提出了一种结合支持向量机回归进行集成学习

基金项目: 山西省科技攻关计划(工业)资助项目(2015031003-1)

作者简介: 贾翻连(1991-), 女, 山西柳林人, 硕士研究生, CCF 会员, 会员号为 64623G, 主要研究方向: 深度学习; 张丽红(1968-), 女, 山西太原人, 副教授, CCF 会员, 会员号为: 64789M; 主要研究方向: 图像处理、深度学习。

的方法来进行人群密度估计^[5]；谭智勇等基于深度卷积神经网络的人群密度估计方法^[6]。在这些方法中，网络首先学习训练数据，然后通过学习过程组成的网络系统能够识别人群特征并进行分类。此外，大多数机器学习使用梯度下降法求最小能量损失时较难找到零损失，机器学习的结果可能在真实值附近振荡，产生学习误差。

卷积网络是为识别二维形状而特殊设计的一个多层感知器，且卷积神经网络模型具有很强大的特征提取和表示能力，但是卷积神经网络的计算速度受卷积层与子采样层之间特征图的大小匹配的影响。本文提出了一种将离散小波变换替换卷积神经网络中的子采样层，以改善预处理时的振荡现象和卷积神经网络中特征图的匹配问题，从而更好地进行人群密度估计。

1 基本卷积神经网络模型

在深度学习模型中，卷积神经网络是最经典的模型之一。卷积网络是为识别二维形状而特殊设计的一个多层感知器，这种网络结构对平移、比例缩放、倾斜或者其他形式的变形具有高度不变性。LeCun^[7]等人提出的卷积神经网络(Convolutional Neural Network, 简称 CNN)是第一个真正的多层非线性结构学习算法，它在手写字符识别中已经成功应用，该方法将特征提取和识别结合在一起，通过卷积操作直接对图像进行处理，提取图像的局部特征如角点、边缘等等，并利用误差反向传播过程，不断地选择、优化既得的特征。

一般地，CNN 包含多个卷积过程和全连接的过程。每个卷积的过程由四个部分组成，即卷积层、汇聚层、非线性变换层，局部反应标准化层。通过这四个部分减少参数数目以提高训练性能，并能够保证图像对位移、缩放、形变的鲁棒性。

卷积网络较一般神经网络在图像处理方面有如下优点：a) 输入图像和网络的拓扑结构能很好的吻合；b) 特征提取和模式分类同时进行，并同时在训练中产生；c) 权重共享可以减少网络的训练参数，使神经网络结构变得更简单，适应性更强。

卷积神经网络的缺点：存在计算瓶颈，即卷积层与子采样层间特征图的大小匹配，会影响计算速度和时间。

2 离散小波变换

小波分析从信号处理角度来看可以作为基于傅里叶变换理论发展起来的一种有效的时频分析方法^[8]，将小波分析应用于数字图像信号，可以利用小波变换在时域和频域均具有良好的局部化描述的特点，能够方便地表示图像的平滑区域和局部特征区域，并具有多分辨率分解的特点^[9]。

有两种主要类型的小波变换：连续小波变换(CWT)和离散小波(DWT)变换。离散小波变换可表示为式(1)。

$$W_{f(\psi)}(2^{-j}k, 2^j) = d_k^j = 2^j \int_{-\infty}^{\infty} \overline{\psi(2^j t - k)} f(t) dt \quad (1)$$

母小波应满足如式(2)所示的条件^[10]。

$$\int \overline{\psi_k^j(t)} \widetilde{\psi_n^j(t)} dt = \delta_k^j \delta_n^j \quad (2)$$

主要有两种方法实现离散小波变换：多分辨率分析和提升方案。在本文的研究中，采用多分辨率分析来实现离散小波变换。多分辨率分析分解信号可以写成式(3)。

$$\begin{aligned} f_n(t) &= \sum_k c_k^j \phi(t-k) \\ c_k^j &= \sum_l g_{l-2k} c_l^{j-1} \\ d_k^j &= \sum_l h_{l-2k} c_l^{j-1} \end{aligned} \quad (3)$$

其中， $f_n(t)$ 是分析信号， $\phi(t)$ 是一个尺度函数， g_l 是一个低通滤波器，它对应于 $\phi(t)$ ， h_l 是一个高通滤波器，其对应的小波母 $\pi(t)$ 。

3 改进的卷积神经网络

用基本卷积神经网络处理人群密度图片时，其网络的结构会受图片背景的影响。当背景较为复杂时，网络会有比较多的隐含层，此时，基本卷积神经网络的计算瓶颈会更为突出。对此，本文对此模型进行改进，其模型如图 1 所示。

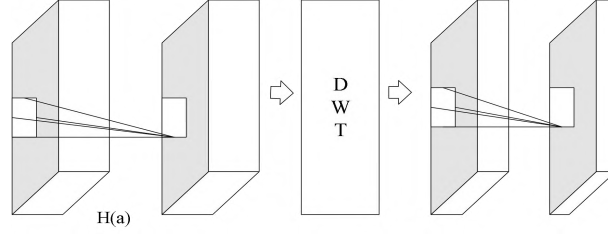


图 1 改进的卷积神经网络

Fig.1 Improved Convolution Neural Network

对图像进行小波变换的原理就是通过低通滤波器和高通滤波器对图像进行卷积滤波，再进行二取一的下采样。在卷积神经网络中，离散小波变换可以达到和原卷积神经网络中下采样过程一样的效果，因此在加入 DWT 后，将原来的下采样过程去掉。

DWT 层使用 DWT 分解上一层的输出，通过式 (3)，输出 $h(a)$ 从式 (4) 中获得。

$$h(a) = f_n(t) = \sum_k c_k^j \phi(t-k) \quad (4)$$

为了使输出具有相同的长度，对 DWT 的输出层取样。最后，小波神经网络的输出可以表示为式 (5)。

$$y = \sum_k c_k^j \phi(W^2) + b^2 \quad (5)$$

即该方法利用权重作为母小波的尺度，这意味着训练过程可视为寻找母小波最佳形状来逼近训练数据的一种方法。因此，式 (5) 中的权重矩阵 $W^{[1]}$ 的每个分量可以由式 (6) 计算。

$$\omega_{n,m} = \frac{o_{n,m} - p_{n,m}}{o_{n,m}} \quad (6)$$

其中， $o_{n,m}$ 是矩阵 O 的元素， $p_{n,m}$ 是矩阵 P 的元素，矩阵 O 和 P 如式 (7) 所示。

$$O = \begin{bmatrix} o_{i1} & o_{j1} \\ o_{i2} & o_{j2} \\ o_{i3} & o_{j3} \end{bmatrix}, \quad P = \begin{bmatrix} p_{i1} & p_{j1} \\ p_{i2} & p_{j2} \\ p_{i3} & p_{j3} \end{bmatrix} \quad (7)$$

由于对图像进行小波变换后可以将原始图像的大部分能量集中在小波系数的少部分中，因此改进的神经网络的权值矩阵 W 与传统的神经网络不同，在权值矩阵中有两个参数 O 和 P 。在这里，我们假设

$u = \frac{p}{o}$ ，让 u 先下降，然后 o 和 p 可以从式 (8) 所示的两个不同的方向下降。

$$\begin{aligned} u &= u - \eta \frac{\partial E}{\partial u} \\ o &= o - \eta \frac{\partial u}{\partial o} \\ p &= p - \eta \frac{\partial u}{\partial p} \end{aligned} \quad (8)$$

式中， E 为能量损失，可由式 (9) 计算。

$$E = \frac{1}{2} \sum_k (y_k - t_k)^2 \quad (9)$$

式中， y_k 是神经网络的当前输出， t_k 是训练数据。在本文的研究中，使用均方误差来评估能量损失。

4 仿真实验及结果

在系统为 windows server 2012 的服务器、python 3.5、MATLAB 2012a 的实验平台下，用 2 个数据集对提出的方法进行验证。一个数据集来自 Chan^[12]UCSD 的行人数据集（下文简称数据集 1），该数据集图片分辨率为 238*158、帧率为 10fps 的视频。另外一个数据集是自己拍摄的一段某路口视频（下文简称数据集 2），该视频视角较远，场景中有少许杂乱背景。

将人群密度按照人数分为低、中、高三个等级，具体定义如表 1 所示。

表 1 人群密度等级的定义

Table 1 Definition of population density level

密度等级	低	中	高
人数	0-10	11-25	>25

本文选择 Haar 小波作为基小波，选择 2 级分解尺度。传统的多层神经网络为 32 层，改进后的神经网络为 28 层。

我们使用上述数据集 1 的部分作为训练数据来训练所提出的改进的神经网络的特性，部分作为测试数据集来测试所提出的网络，并且将本文的方法与传统的卷积神经网络进行对比，得到的实验结果如表 2 和表 3 所示。其中，表 2-5 中，A 为传统的多层神经网络方法，B 为多层神经网络中隐含层加入离散小波变换方法，C 为基本的卷积神经网络方法，D 为本文改进的卷积神经网络方法。表 2 中，E 为 HangSu^[13]等采用了 SST-LBP 方法描述特征并使用 SVM 作为分类器；F 为 Kim^[14]采用计算运动区域和对比信息的方法来估计人群密度等级，其中对比信息基于 GLDM^[15] 矩阵方法，累加矩阵元素将和作为一个特征，然后用这两个特征去训练 BP 神经网络。

表 2 数据集 1 人群密度等级估计准确率

Table 2 Data set 1 for population density estimation accuracy

方法	A	B	C	D	E	F
低密度	89.2%	91.7%	96.2%	97.3%	91.3%	91.8%
中密度	88.4%	90.1%	93.1%	94.9%	85.1%	83.2%
高密度	89.6%	90.8%	96.9%	97.1%	88.2%	86.6%

表 3 数据集 2 人群密度等级估计准确率

Table 3 data set 2 for population density estimation accuracy

方法	A	B	C	D
低密度	83.0%	87.6%	88.1%	91.4%
中密度	81.2%	84.3%	85.3%	87.8%
高密度	82.7%	86.4%	86.9%	89.0%

用信噪比来精确地说明该小波神经网络优于传统的神经网络，它的信噪比用公式(10)计算。

$$SNR = 20 \log_{10} \frac{A_{signal}}{A_{noise}} \quad (10)$$

用不同的数据集进行了实验，所得结果见表 4。

表 4 信噪比对比结果

Table 4 signal-to-noise ratio comparison results

方法	A	B	C	D
数据集 1	43.12	52.09	59.75	72.30
数据集 2	39.11	49.34	58.73	73.53

此外，我们使用 MSE（均方误差）来衡量神经网络的性能，其中 MSE 可以写成公式(11)。

$$MSE = 20 \log_{10} \left\{ \frac{1}{N} \sum (x - \bar{x})^2 \right\}$$

(11)

用不同的数据集进行了实验，所得结果见表 5。

表 5 均方误差结果显示

Table 5 results shown by MSE

方法	A	B	C	D
数据集 1	-42.45	-46.71	-52.41	-56.89
数据集 2	-30.48	-39.62	-42.45	-46.78

从表 2 到表 5 可以看出，本文改进的神经网络的平均信噪比较传统的网络有所提高，且均方误差平均值有所降低，这意味着改进的神经网络比传统的神经网络具有更好的性能，从而可以得到更好的人群密度等级估计结果。由数据集 2 的结果可知，该网络具有良好的鲁棒性。

5 结束语

本文将卷积神经网络中的子采样层用离散小波变换代替，对其权重矩阵的计算进行了优化，通过权重自适应改善卷积神经网络在解决回归问题预处理时的振荡现象，提高卷积神经网络卷积层与子采样层间特征图的大小匹配。实验证明，本文采用的基于离散小波变换的卷积神经网络与基本的卷积神经网络相比，有较好的分类效果及较快的运算速度。

参考文献

- [1] Davies A C, Yin J H, Velastin S A. Crowd monitoring using image processing[J]. Electronics& Communication Engineering Journal, 1995, 7(1): 37—47.
- [2] Cho S Y, Chow T W S, Leung C T. A neural-based crowd estimation by hybrid global learning algorithm[J]. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 1999, 29(4): 535-541.
- [3] 刘曼, 彭月平, 姜源. 基于可变矩形框的人群密度数值估计算法研究[J]. 科学技术与工程, 2017, 17(17):266-271.
- [4] 时增林, 叶阳东, 吴云鹏, 姜铮铮. 基于序的空间金字塔池化网络的人群计数方法[J]. 自动化学报, 2016, 42(6):866-874.
- [5] 张洞明, 郑宏, 张力. 基于多特征集成学习的景区人群密度估计[J]. 科学技术与工程, 2017, 17(5):74-81.
- [6] 谭智勇, 袁家政, 刘宏哲, 李青等. 基于深度卷积神经网络的人群密度估计方法[J]. 计算机应用与软件, 2017, 34(7):130-136.
- [7] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [8] 郭超. 小波变换在图像压缩方面研究及应用[D]. 上海交通大学, 2007.
- [9] 刘衍琪, 詹福宇, 蒋献文, 周华英. MATLAB 计算机视觉与深度学习实践[M]. 北京: 电子工业出版社, 2017.
- [10] Yoshiaki Shimizu, and Zhong Zhang, and Rafael Batres. Frontiers in Computing Technologies for Manufacturing Applications[M], Springer, Germany, 2007.

- [11]Kouki Saito. Make Deep Learning from zero: using python to learn theory and implementation of Deep Learning[M].O'Reilly Japan, 2016.
- [12]Chan A, Vasconcelos N. Counting People with Low-Level Features and Bayesian Regression[J]. IEEE Trans on Image Processing, 2012, 21(4): 60—77.
- [13]Su H, Yang H, Zheng S. The large-scale crowd density estimation based on effective region feature extraction method[M]. Computer Vision-ACCV 2010. Springer Berlin Heidelberg,2011: 302-313.
- [14] Kim G, An T, Kim M. Estimation of Crowd Density in Public Areas Based on Neural Network[J]. KSII Transactions on Internet & Information Systems, 2012, 6(9):2170-2190.
- [15]Haralick R M. Statistical and structural approaches to texture[J]. Proceedings of the IEEE, 1979,67(5): 786-804.