

DOI: 10.13382/j.jemi.2018.04.012

基于捷径的卷积神经网络在人脸识别中的应用研究^{*}

王飞¹ 张莹^{1,2} 张东波^{1,2} 邵豪¹ 成超鹏¹

(1. 湘潭大学 信息工程学院 湘潭 411105; 2. 机器人视觉感知与控制技术国家工程实验室 长沙 410082)

摘要: 人脸识别的特征要求具有区分性和识别性,传统的卷积神经网络(CNN)无法将低层次特征与高层次特征进行融合,识别准确率难以进一步提高。提出了一种基于附加惩罚函数和捷径连接的卷积神经网络模型,算法通过将第一层的卷积特征与原有网络模型最后一层的全连接层直接连接,增加了深层次特征的识别能力,减少了浅层主要特征的损失;为了提高已学习到的深度特征的识别能力,算法在原有的softmax损失函数项添加一个惩罚项,使已学习到的同类特征到该类特征中心之间的距离最小化。在CASIA-web、Facescrub数据集上的实验结果表明,改进算法分类准确率优于深度卷积神经网络(DCNN)、GRCC等算法。

关键词: 卷积神经网络; 捷径连接; 惩罚函数; 人脸识别

中图分类号: TP391.73; TN911.41 文献标识码: A 国家标准学科分类代码: 520.60

Research on application of convolutional neural networks in face recognition based on shortcut connection

Wang Fei¹ Zhang Ying^{1,2} Zhang Dongbo^{1,2} Shao Hao¹ Cheng Chaopeng¹

(1. College of Information Engineering, Xiangtan University, Xiangtan 411105, China;

2. National Engineering Laboratory for Robot Visual Perception and Control Technology, Changsha 410082, China)

Abstract: Face recognition feature requirements are distinguishable and identifiable, but the conventional convolution neural network cannot integrate low-level features with high-level features, and the recognition accuracy is difficult to be further improved. In order to solve this problem, a convolution neural network model based on additional penalty function and shortcut connection is proposed in this paper. The algorithm combines the convolution feature of the first layer with the final layer of the original network model, which reduces the loss of the main features of the shallow layer and increases the recognition ability of the deep feature. At the same time, with the purpose of enhancing the ability to recognize the depth characteristics, the algorithm adds a penalty term to the original softmax loss function item, which minimizes the distance between the similar features. The experimental results on CASIA-web and Facescrub datasets show that the classification accuracy of the improved algorithms is better than DCNN, GRCC and other algorithms.

Keywords: convolutional neural networks; shortcut-connection; penalty function; face recognition

0 引言

人脸验证问题一直是计算机视觉领域的一个核心问题^[1]。在一幅给定的自然图像中,人类视觉系统能够从整幅图像中找到最感兴趣的区域和物体,进而利用细节特征进行目标识别^[2],文献[3]表明综合利用高层次特征和低层次特征对改善视觉系统的识别性能是有利的。深度卷积神经网络(DCNN)对不同任务,如目标识

别^[4-5]、目标检测^[6]、人脸验证^[7]有着极其出色的特征提取性能。然而基于传统的卷积神经网络难以有效的将低层次特征与高层次特征进行融合,进而提取出不仅具有可识别性而且具有可区分性的特征。为了增加深层次特征的识别能力,本文借鉴DeepID网络^[8]及深度残差网络结构^[9]的思想,在传统的Alexnet卷积神经网络^[10]的基础上,提出了一个基于捷径连接和惩罚函数的卷积神经网络模型,提高人脸识别的准确性。

收稿日期: 2017-10 Received Date: 2017-10

^{*} 基金项目: 湖南省自然科学基金(2017JJ2251)、湖南省教育厅重点项目(14A137)资助

1 卷积神经网络(CNN)

卷积神经网络是一种特殊的深层前向神经网络模

型,其结构一般由输入层、多个交替连接的卷积层和池化层、全连接层、以及输出层组成^[11]。典型连接方式如图1所示实线框。

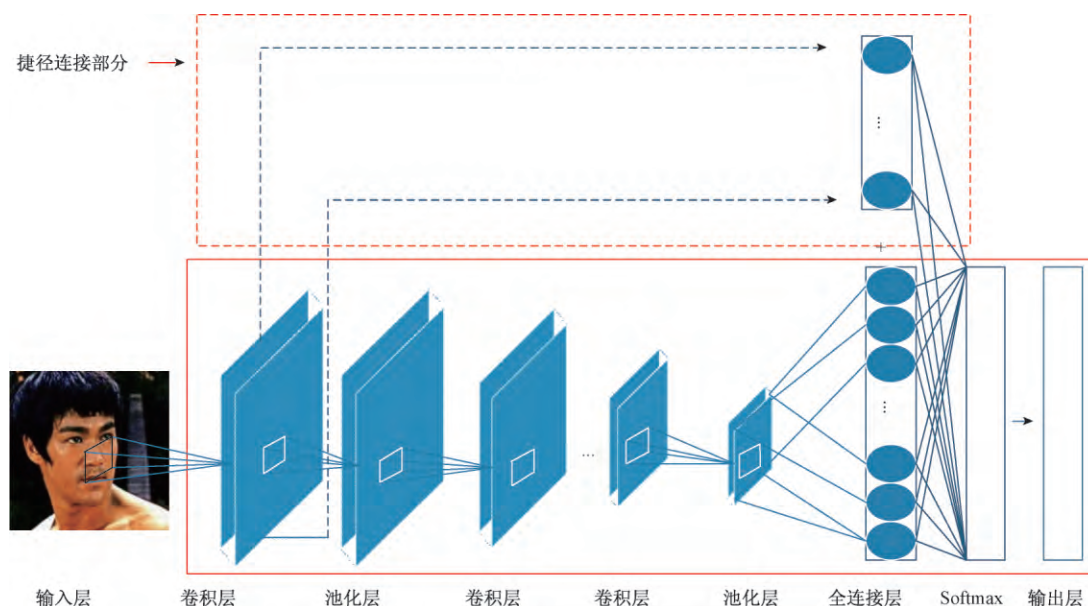


图1 ALEXNET 框图

Fig. 1 Block diagram of ALEXNET

如图2所示,第1卷积层主要提取人脸图像的轮廓、边缘等浅层信息,从第1卷积层到第5卷积层,卷积层提取的特征由整体特征逐渐变化到提取局部特征,从浅层特征逐渐变化为深度特征。人脸识别获得的特征不仅要

具有可分性而且要具有可识别性,虽然基于单独的softmax损失函数的分类器能将已获得特征分离,但是这些分离好的特征对于人脸识别任务还不足够有效,特征的可识别性相对较差^[12]。

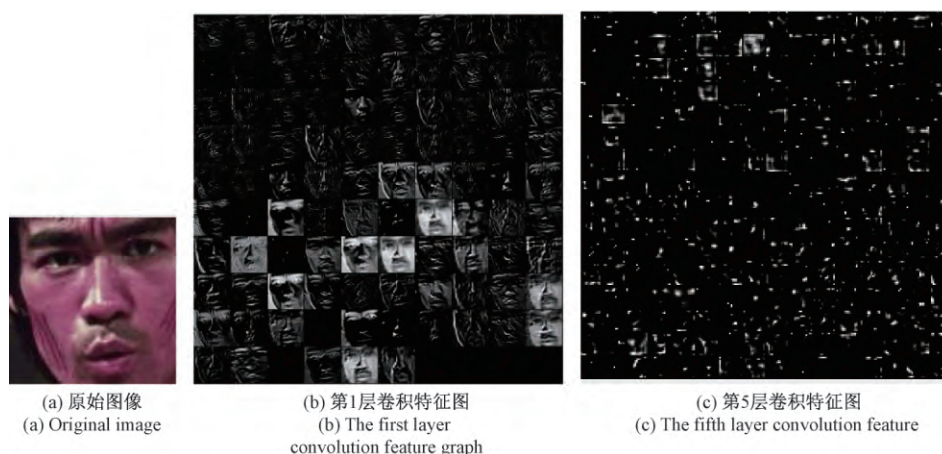


图2 卷积特征效果

Fig. 2 Convolution feature effect

2 基于捷径连接的卷积神经网络

Alexnet 网络经过多个池化进行降维的同时,减少了

后面各卷积层包含神经元的数量,使部分信息在传播过程中损失掉^[8],这样就会降低识别的准确率。本文提出的基于附加惩罚函数捷径连接的卷积神经网络模型将前面一层的卷积特征直接越过中间的卷积层和池化层,直

接连接到最后的全连接层,其连接效果图如图1中虚线框和实线框所示。而在第1个卷积层和最后的隐含层之间使用捷径连接可以有效降低传播信息的损失。

2.1 捷径连接网络模型

对于输入第 k 个人脸样本,捷径连接网络模型的学习过程如式(1)~(4)所示。由于池化操作对前面的特征图信息进行了聚合统计操作,但这种聚合操作在一定程度上降低计算量的同时也会带来部分细节信息的丢失,例如对前面的特征图进行 $n \times n$ 的最大池化操作后,仅保留了值最大的那个特征信息,其余 $(n \times n) - 1$ 个特征信息均丢弃掉了。将不同层的特征进行融合。而通过式(2)将第 k 个样本第一隐含层的输出特征同第 k 个样本第9隐含层进行特征融合后可以有效降低经过多次池化操作带来的信息损失,从而可以得到更完备的深度特征,进而可以在一定程度上提高识别的准确率^[13]。

$$h_{1j}^k = \delta(w_{1j}^k * x^k + b_{1j}^k) \quad (1)$$

$$H_{10}^k = (h_{1,1}^k, h_{1,2}^k, \dots, h_{1,32}^k, H_9^k) \quad (2)$$

式中: h_{1j}^k 表示第 k 个样本的第1个隐层的第 j 个卷积面; w_{1j}^k 和 b_{1j}^k 分别表示第 k 个样本第一隐含层的第 j 个卷积面的卷积核和偏置; H_9^k 为第9层的输出向量; H_{10}^k 表示第 k 个样本的第10隐含层。将第 k 个样本第1个隐含层(卷积层)的输出和第 k 个样本第9隐含层的输出进行融合后结果向量作为第 k 个样本第10个隐含层的输入。

$$H_{11}^k = \delta(W_{11}^k * H_{10}^k + b_{11}^k) \quad (3)$$

$$H_{12}^k = \delta(W_{12}^k * H_{11}^k + b_{12}^k) \quad (4)$$

式中: H_{11}^k 和 H_{12}^k 分别表示第 k 个样本的第11个隐含层和第12个隐含层的输出向量, W_{11}^k 和 b_{11}^k 分别表示第10层和第11层之间的连接权重和偏置。 H_{10}^k 表示第 k 个样本第10个隐含层的输出向量, $\delta(\cdot)$ 为PReLU(parametric ReLU)激活函数,文献[14]已经证明在相同的条件下使用PReLU作为激活函数要比使用sigmoid函数或ReLU(rectified linear units)函数作为有更高的识别率,而且PReLU中的参数可以通过链式法则和其他层一起进行训练学习优化^[15],其函数表达式如式(5),其激活函数如图3所示。

$$\delta(x) = \max(0, x) + a \cdot \min(0, x) \quad (5)$$

2.2 惩罚函数

人脸识别任务中,通过卷积神经网络学习到的特征类间距不像场景、行为识别等识别任务中得到的特征类间距那么大,因此在单独使用softmax损失函数时,识别的性能会有所下降。为了增强对已学习到的深度特征的识别能力,本文对原有的softmax损失函数 J_s 添加一个惩罚项如式(6)所示。增加惩罚项的目的在于将已学习到的同类特征到该类特征中心之间的距离最小化。 J_c 通过计算深度特征与该类深度特征中心之间的距离来对原有

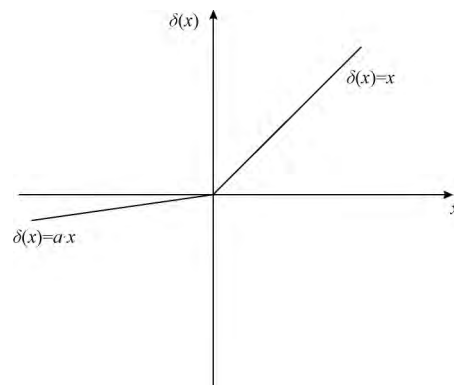


图3 激活函数曲线

Fig. 3 The activation function curve

的softmax损失函数加以惩罚,其形式如式(7)所示,并采用式(8)和(9)所示的形式进行特征中心的更新。

$$J = J_s + J_c = - \sum_{i=1}^m \log \frac{e^{y_i}}{\sum_{j=1}^m e^{y_j}} + J_c \quad (6)$$

$$J_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_i\|_2^2 \quad (7)$$

$$C_{i+1} = C_i + \Delta C_i = C_i + \alpha \cdot \frac{\partial J_c}{\partial x_i} \quad (8)$$

$$\alpha = \frac{\sum_{j=1}^m \Gamma(j=i)}{1 + \sum_{j=1}^m \Gamma(j=i)} \quad (9)$$

通过式(6)并采用随机梯度下降法进行求解优化,随着不断迭代,可使同类特征距离该类特征中心的距离逐渐减小,每类特征更加汇聚于一个区域,对于一般样本,同时会使不同特征之间的距离进一步增大,进而更容易把每种类别区分出来,也就在一定程度上提高了识别的准确性^[16]。式(7)中 x_i 表示第 i 类深度特征, c_i 表示第 i 类深度特征的中心,其值为每次迭代时相应类别的平均特征,同时为了避免少数已失去标签样本数据的干扰,添加一个参数 α 来控制深度特征中心的学习率, α 按照式(9)的形式选取, c_i 采用式(8)、(9)的方法进行更新,式(9)中 $\Gamma(\cdot)$ 为条件函数,如果条件满足函数值为1,如果条件值不满足函数值为0。这样经过多次迭代后,同一类别的特征同 C_i 之间的距离逐渐缩小并逼近于 C_i ,因此每类特征可识别性更强,更有利于softmax的分类。损失函数 J 可以通过SGD(stochastic gradient descent)算法进行优化求解。本文算法的整体流程如图4所示。

3 实验及数据分析

为了验证本文算法的有效性,实验采用CASIA-

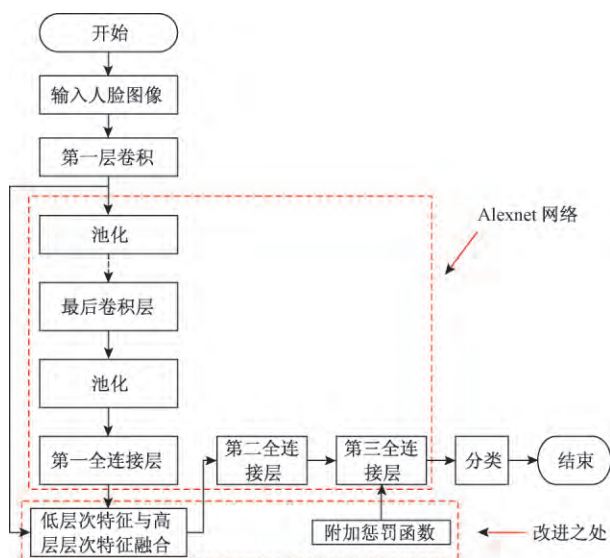


图4 本文算法的人脸识别流程

Fig. 4 Flow chart of face recognition of the proposed algorithm in this paper

webface(<http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html>)、Facescrub(<http://vintage.winklerbros.net/facescrub.html>) 人脸数据集。其中 CASIA-webface 数据集包含了 10 575 个人, 共计 494 414 张人脸图像, 每个人均有多张不同的照片, Facescrub 数据集是由 265 位男性和 265 位女性组成的 106 863 张人脸图像, 每人大概有 200 张图片。

3.1 图像预处理

CASIA-webface、Facescrub 原始人脸数据集中包含了一些与人脸无关的信息, 因此首先利用 Sun 等人^[17]提出的面部检测算法检测出每张人脸图像的关键点, 包括 2 个眼部中点、鼻尖和 2 个嘴角, 从而得到每张人脸的边框。然后把只包含人脸面部的图像剪切出来作为网络模型的训练集和测试集。其中每个人脸集均是随机选取 70% 作为训练集, 剩余的 30% 作为测试集。两个人脸数据集中的人脸图像检测和剪切效果如图 5 和 6 所示, 图中, 左边为人脸库的原图, 右边为剪切得到的人脸面部。



图5 CASIA-web 人脸预处理

Fig. 5 CASIA-web face preprocessing



图6 Facescrub 人脸预处理

Fig. 6 Facescrub face preprocessing

3.2 CASIA-web 数据集实验结果

表 1 所示为本文模型及现有的算法模型在 CASIA-web 数据集上的识别效果, 图 7 所示为本文模型在 CASIA-web 数据集上损失函数的收敛情况。

表 1 各种模型在 CASIA-web 数据集上的识别准确率

Table 1 The recognition accuracy of various models in CASIA-web datasets

模型	CASIA-web / %
本文算法	98.9
无附加惩罚函数的捷径连接模型	98.2
Alexnet 算法	97.5
DCNN ^[18]	97.45
Yi et al. ^[19]	97.73
Wang et al. ^[20]	98.23

3.3 Facescrub 数据集实验结果

表 2 所示为本文模型及现有的算法模型 Facescrub 数据集上的识别效果, 图 8 所示为本文模型在 Facescrub 数据集上损失函数的收敛情况。

表 2 各种模型在 Facescrub 数据集上的识别准确率

Table 2 The recognition accuracy of various models in Facescrub datasets

模型	Facescrub
本文算法	98.5%
无附加惩罚函数的捷径连接模型	98%
Alexnet 算法	97%
one-vs-one ^[21]	80.1%
one-vs-rest ^[21]	80.3%
Hayat et al. ^[21]	91.5%
GRCC ^[22]	92.666%
NEC ^[22]	78.658%

3.4 实验结果分析

由图 7(a) 和图 8(a) 可以发现, 由于开始时的类中心是通过随机初始化获得, 这也使得开始时类特征点距离类中心的平均距离相对比较大, 如图 7(a) 和 8(a) 所

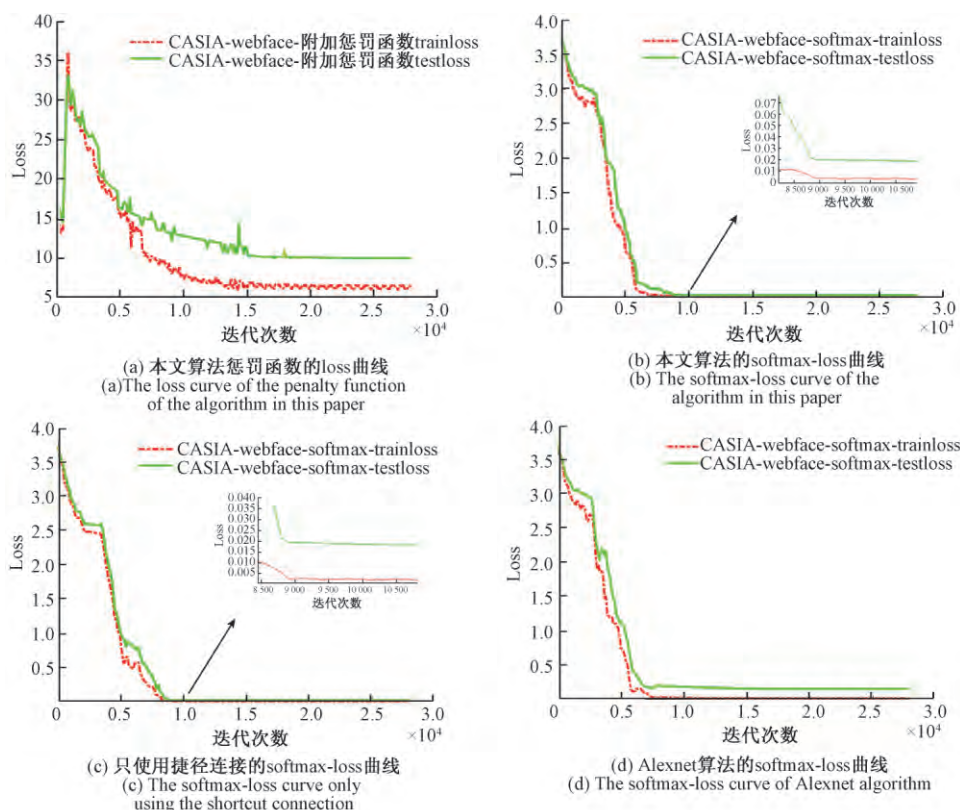


图7 CASIA-web数据集上的损失函数曲线

Fig.7 The loss function curve in the CASIA-web datasets

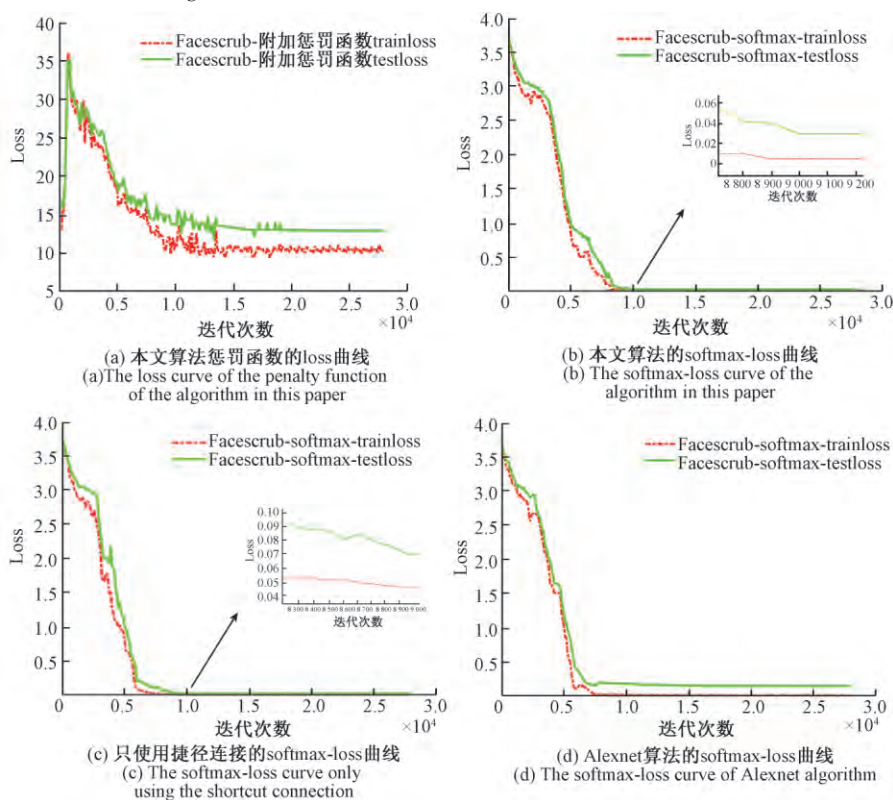


图8 Facescrub数据集上的损失函数曲线

Fig.8 The loss function curve in the Facescrub datasets

示,但是随着不断地迭代,可以通过学习得到一个比较理想的类中心,通过式(6)的使用,同类特征之间的距离最终是在减小的,同类特征与该类特征的中心距离收敛处的值比开始降低很多,达到了使同类特征之间汇聚的更紧凑的效果,从而更有利于识别出不同的类别。对比图7(b)和(c)、图8(b)和(c)可以发现,惩罚函数的使用降低了softmax损失函数的收敛值。对比图7(b)~(d)以及图8(b)~(d)可知,未使用捷径连接的网络模型的softmax损失函数测试曲线收敛值要比使用捷径连接模型的损失函数测试曲线收敛值大很多,证明图像经过多次卷积池化后部分特征会出现丢失的现象。算法使用了式(7)所示的惩罚函数,使得同类特征易汇聚在一起,更易于分类器对其进行分类,故如图7(b)、7(c)和图8(b)、(c)所示,带有惩罚函数模型的损失函数值要比没有使用惩罚函数模型的损失函数值低,降低了错误分类的概率。

4 结 论

本文从特征丢失及损失函数的角度出发,对卷积神经网络进行了研究,提出了一种带有附加惩罚函数捷径连接的人脸识别模型,综合考虑了浅层特征和深层次特征,并将二者进行了融合,降低了经过多次池化操作后部分特征的丢失程度,通过添加的惩罚函数来降低前面已学习到的同类深层次特征的类内距离,然后再使用分类器对其进行分类,提高了分类的准确率。

参考文献

- [1] ZHAO W, CHELLAPPA R, PHILLIPS P J, et al. Face recognition: A literature survey [J]. ACM computing surveys (CSUR), 2003, 35(4): 399-458.
- [2] ZHONG S, LIU Y, REN F, et al. Video saliency detection via dynamic consistent spatio-temporal attention modelling [C]. AAAI, 2013.
- [3] 齐美彬, 檀胜顺, 王运侠, 等. 基于多特征子空间与核学习的行人再识别 [J]. 自动化学报, 2016, 42(2): 299-308.
- QI M B, TAN S H, WANG Y X, et al. Pedestrian reidentification based on multi-feature subspace and nuclear learning [J]. Acta Automatica Sinica, 2016, 42(2): 299-308.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]. Advances in neural information processing systems, 2012: 1097-1105.
- [5] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [7] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 815-823.
- [8] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10,000 classes [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1891-1898.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems, 2012: 1097-1105.
- [11] 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.
- [12] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition [C]. European Conference on Computer Vision. Springer International Publishing, 2016: 499-515.
- [13] HUANG G, LIU Z, WEINBERGER K Q, et al. Densely connected convolutional networks [J]. 2016, arXiv preprint arXiv: 1608.06993.
- [14] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026-1034.
- [15] SERMANET P, LECUN Y. Traffic sign recognition with multi-scale convolutional networks [C]. International Joint Conference on Neural Networks, IEEE, 2011: 2809-2813.
- [16] DUDA R O, HART P E, STORK D G. Pattern Classification [M]. New York: John Wiley & Sons, 2012.
- [17] SUN Y, WANG X, TANG X. Deep convolutional network cascade for facial point detection [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2013: 3476-3483.

- [18] CHEN J C , PATEL V M , CHELLAPPA R. Unconstrained face verification using deep cnn features [C]. Winter Conference on Applications of Computer Vision , IEEE , 2016: 1-9.
- [19] WANG D , OTTO C , JAIN A K. Face search at scale: 80 million gallery [J]. 2015 , arXiv preprint arXiv: 1507. 07242.
- [20] YI D , LEI Z , LIAO S , et al. Learning face representation from scratch [J]. 2014 , arXiv preprint arXiv: 1411. 7923.
- [21] HAYAT M , KHAN S H , BENNAMOUN M. Empowering simple binary classifiers for image set based face recognition [J]. International Journal of Computer Vision , 2017: 1-20
- [22] NECH A , KEMELMACHER-SHLIZERMAN I. Level playing field for million scale face recognition [J]. 2017 , arXiv preprint arXiv: 1705. 00393.

作者简介



王飞,湘潭大学硕士研究生,主要研究方向深度学习、图像处理与识别。

E-mail: sdauwangfei2010@163.com

Wang Fei is a M. Sc. candidate in Xiangtan University. His main research direction is deep learning , pattern recognition and intelligent information processing.



张莹,2010 年于湖南大学获得博士学位,现为湘潭大学副教授,主要研究方向为模式识别和图像处理、视觉导航。

E-mail: zhangying@xtu.edu.cn

Zhang Ying received Ph. D. from Hunan University in 2010. Now , he is an associate professor in Xiangtan University. His main research interests include pattern recognition , image processing and visual navigation.

泰克为 ST 2110 信号发生和分析提供完善的解决方案

PRISM 平台全面支持 ST 2110 ,增加信号发生功能并推进泰克 IP/SDI 工作流程的领导地位

泰克科技公司日前宣布,其 PRISM 媒体分析解决方案现在全面支持最新发布的 SMPTE ST 2110 成套标准(-20,-30,-40,-21),用于管理 IP 网络专业媒体。此外,泰克为 PRISM 推出最新的低成本现场可升级信号发生功能,在混合/SDI 网络中测试 IP 和 12G-SDI 时,可以让客户树立更多信心。

随着 ST 2110 标准的发布,广播行业的发展方向已经明朗,即从 SDI 转向 IP 基础设施。但在转型期间,广播工程师和 IT 专业人员面临着管理复杂的 SDI 和 IP 混合环境的挑战。由于能够连接到设施中任何广播信号及新增 ST 2110 信号发生功能,PRISM 提供了完善的测量、网络分析和调试解决方案,可以让客户平滑迁移到未来完全基于 IP 的基础设施和工作流程中。

“ST 2110 的发布正在加速 IP 迁移,但如果广播公司对网络的稳定性和性能缺乏信心,那么迁移就永远无法实现。”泰克科技公司视频产品线总经理 Charlie Dunn 说,“凭借 PRISM,我们正在为客户提供进入 ST 2110 时代所必需的东西:保障和信心。”

在试运行或调试新网络中的问题时的主要挑战是确定哪些地方发生了连接错误或其他问题。为找到这些错误,最高效的方式之一是把已知良好的信号注入到网络

中,检查信号路径和系统冗余性。泰克专为 PRISM 开发的最新 MP-GEN 选项实现了这一点,提供了用于 IP 网络的 ST2110-20、-30 和 ST2022-7 测试模式,以及用于混合/SDI 网络的 12G-SDI 测试模式。

有了本次发布的最新功能,PRISM 现在支持生成、连接和分析以下信号:

ST 2110-10: 系统定时

ST 2110-20: 未压缩的活动视频

ST 2110-30: 未压缩的 PCM 音频

ST 2110-40: 辅助数据

ST 2110-21: 业务整形未压缩视频

PRISM 提供了各种测量、趋势图和会话显示功能,工程师可以查看各种网络或内容问题,包括 ST 2110 网络中的间歇性视频、音频或数据内容丢失。此外,由于新增 ST 2110-21 支持,工程师可以测试发送器和接收器在不同 IP 网络业务流量条件下的性能。泰克还扩展了系统设置和维护使用的精密时间协议(PTP)定时分析功能,增加了 NMOS 发现和注册/会话描述协议(SDP)支持,简化了系统集成。PRISM 可以同时执行 IP 和 SDI(12G/3G/HD/SD)测量,并可以在任意输入类型之间简便切换。