

CVPR2019| 中科院VIPL实验室11篇CVPR解读：弱监督学习、视频分割、目标检测

原创

SophiaCV 于 2019-04-18 18:59:00 发布 4941 收藏 16

文章标签：卷积 计算机视觉 神经网络 机器学习 人工智能

在公众号【计算机视觉联盟】后台回复【9076】获取独家200页AI笔记！

本文链接：https://blog.csdn.net/Sophia_11/article/details/89391691

版权

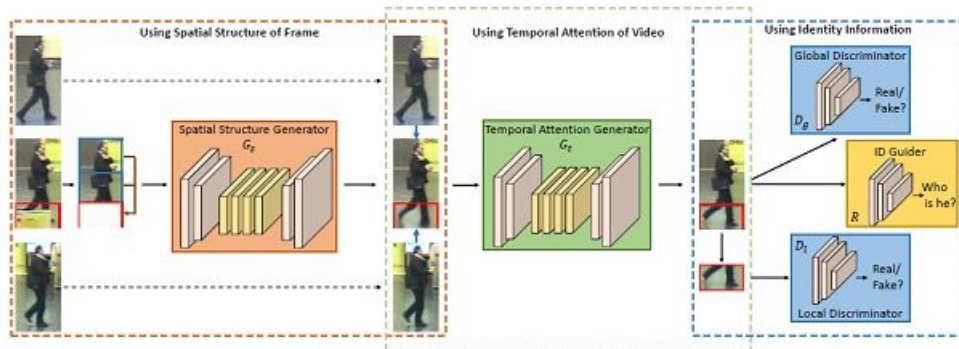
点击上方蓝字计算机视觉联盟获取更多干货

在右上方... 设为星标★，与你不见不散

11篇论文的信息概要介绍如下：

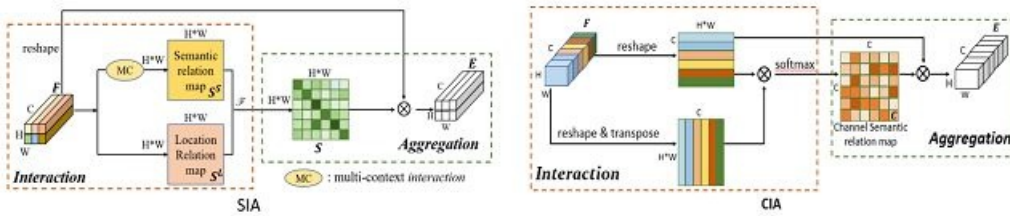
1. VRSTC: Occlusion-Free Video Person Re-Identification (Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, Xilin Chen)

针对视频行人重识别的部分遮挡问题，我们提出了一个时空补全网络（Spatio-Temporal Completion network, STCnet）用来精确的恢复遮挡部位的信息。一方面，STCnet利用行人的空间结构信息，从当前行人帧中未被遮挡的身体部位信息预测出被遮挡的身体部位。另一方面，STCnet利用行人序列的时序信息去预测被遮挡部位。通过联合STCnet和一个行人再识别网络，我们获得了一个对部分遮挡鲁棒的视频行人再识别框架。在当前主流的视频行人再识别数据库（iLIDS-VID, MARS, DukeMTMC-VideoReID）上，我们提出的框架都优于当前最好的方法。



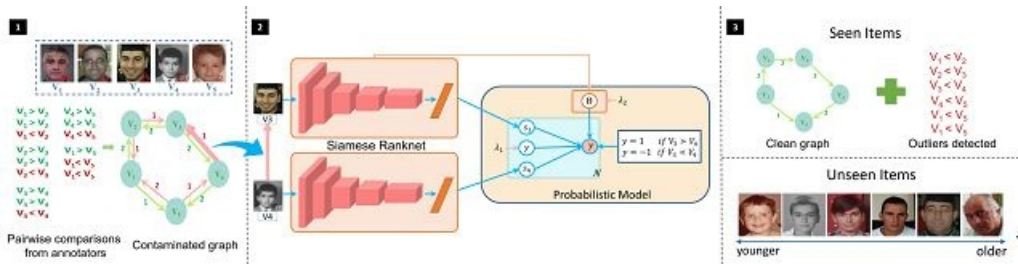
2. Interaction-and-Aggregation Network for Person Re-identification (Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, Xilin Chen)

由于卷积单元固定的几何结构，传统的卷积神经网络本质上受限于对行人大的姿态和尺度变化建模。我们提出了一个网络结构（Interaction-and-Aggregation, IA）用来增强卷积网络对行人的特征表示能力。首先，空间IA模块通过建模空间特征的关联来整合对应相同身体部位的特征。区别于卷积神经网络从一个固定的矩形区域提取特征，空间IA能够根据输入行人的姿态和尺度自适应的决定感受野。其次，通道IA模块通过建模通道特征的关联进一步增强特征表示。我们在多个数据上验证了我们方法的有效性，并且都超过了目前最好的方法。



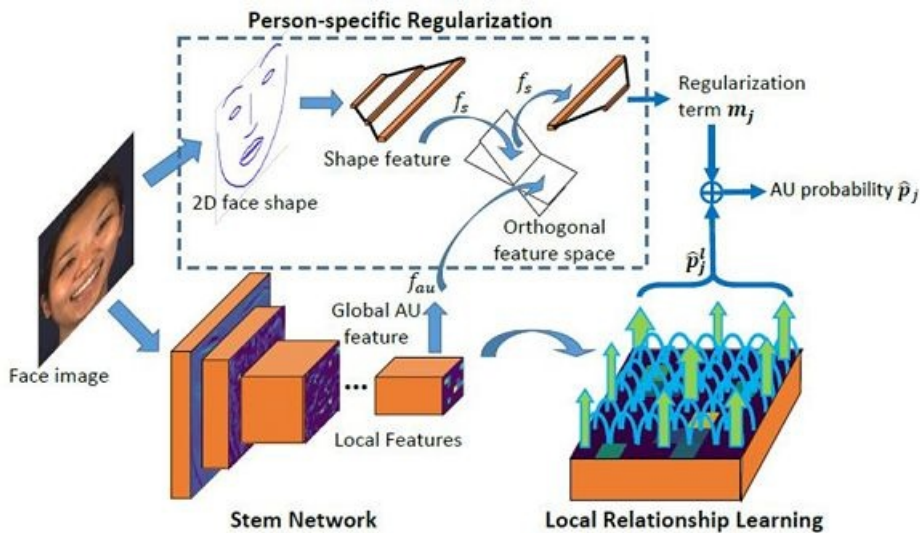
3. Deep Robust Subjective Visual Property Prediction in Crowdsourcing (Qianqian Xu, Zhiyong Yang, Yangbanyan Jiang, Xiaochun Cao, Qingming Huang, Yuan Yao)

在图像主观视觉属性的估计问题中，由于问题的高度主观性，使用绝对数值进行打分往往会因标注者对打分标准的不同理解而导致获得的标注差别较大，因此学界转而使用通过网络众包收集的成对比较数据进行研究。然而众包数据中通常含有异常样本，使得估计产生较大偏差。为尽可能消除其影响，我们构建了一个鲁棒的深度主观视觉属性预测模型，根据所有标注建立成对比较多重图，即两个顶点间根据不同标注存在多条不同方向的边，而后通过由主观视觉属性预测、异常样本稀疏建模两个协同工作的模块组成的通用深度概率框架对标注进行学习。该方法具有更好的异常样本检测能力，同时也可从极其稀疏的样本标注中进行学习。在人脸属性、人脸年龄和鞋子属性三个预测任务上的实验均表明，提出的模型能进行有效的异常样本检测并显著提升预测性能。



4. Relationship Learning with Person-specific Regularization for Facial Action Unit Detection (Xuesong Niu, Hu Han, SongFan Yang, Yan Huang, Shiguang Shan)

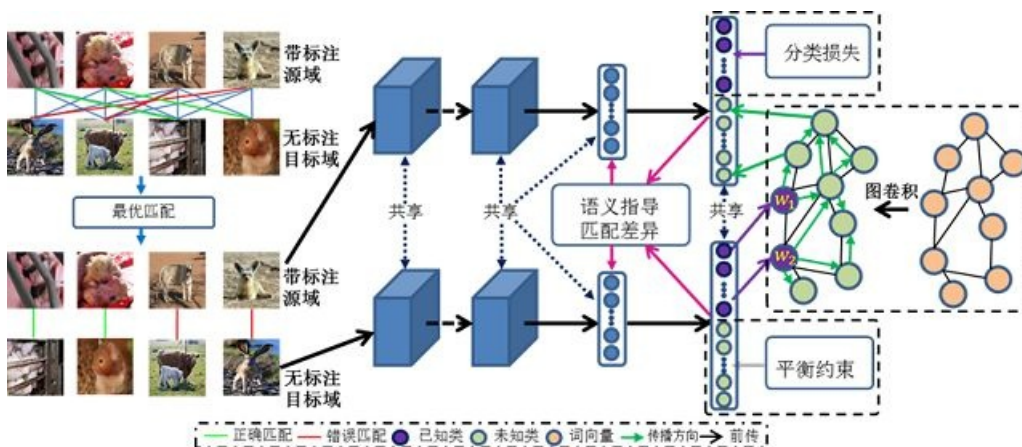
基于人脸活动部件的人脸表情识别能够有效地表征人类丰富的情绪。但是现有的人脸活动部件检测算法受制于人脸活动部件微弱的激活强度以及不同个体之间差异性。为了解决这些问题，我们提出了一种针对人脸活动部件的局部关系学习算法，通过充分利用不同人脸局部区域之间的关系进一步提高人脸不同区域的局部特征的感知能力。与此同时，我们还提出了一种针对人脸形状信息的正则项来剔除在人脸活动部件检测过程中人脸的形状信息的影响，进而获得更加有判别力和泛化能力的人脸活动部件检测器。在广泛使用的人脸活动部件检测数据集BP4D和DISFA上，我们取得了优于当前最好方法的结果。



5. Unsupervised Open Domain Recognition by Semantic Discrepancy Minimization (Junbao Zhuo, Shuhui Wang, Shuhao Cui, Qingming Huang)

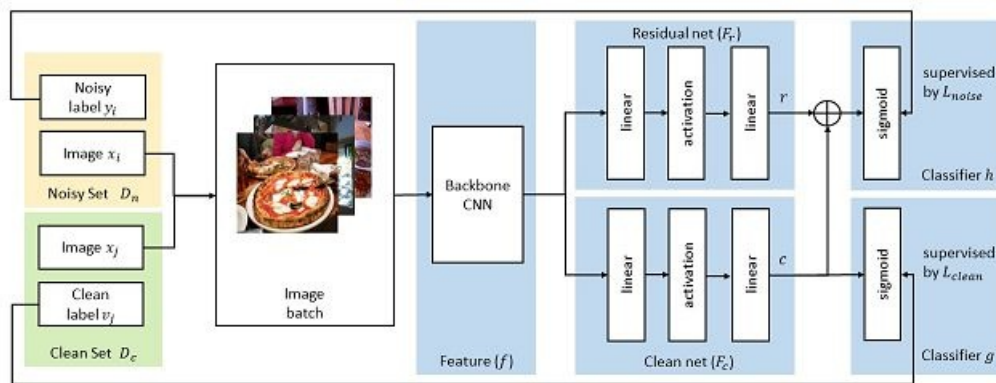
深度学习超强的特征学习能力结合迁移学习以解决目标任务标注数据稀缺问题，是一个极具价值和意义的问题。深度迁移学习现有的一些设定如域适配，零样本学习等仍具有局限性。现有域适配方法仍旧限制在封闭域内，泛零样本识别虽然假定目标域具有未知类且对未知类进行识别，但一般不假定训练集和测试集间存在域间差异。我们提出开放域目标识别的新设定，即假定带标注源域和无标注目标域间存在域间差异，且源域是目标域的一个子集，任务是对目标域上的每个类别的样本都进行正确分类。

针对该问题，我们提出初步解决方案。对于存在未知类，我们通过WordNet构建图卷积神经网络，将已知类的分类规则传播给未知类别，并引入平衡约束来防止训练过程中未知类样本被分成已知类样本。此外，我们先对源域和目标域样本间求一个最优匹配，通过语义一致性来指导对源域与目标域（类别空间不对称）进行域适配。最后我们将分类网络和图卷积神经网络进行联合训练。实验证明所提方法的有效性。



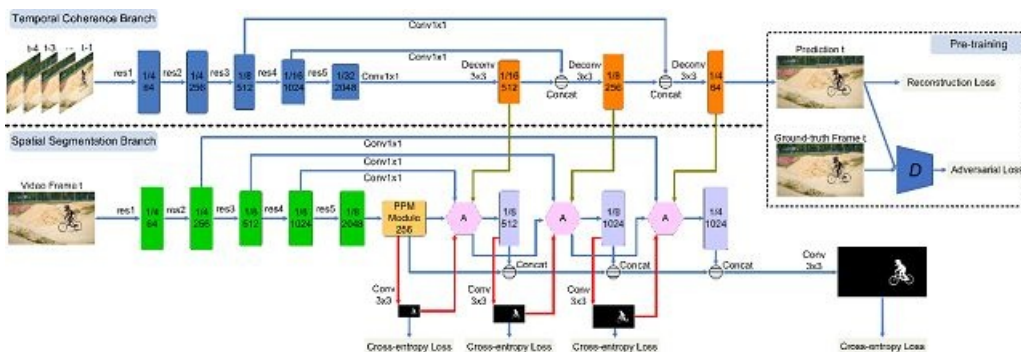
6. Weakly Supervised Image Classification through Noise Regularization (Mengying Hu, Hu Han, Shiguang Shan, Xilin Chen)

弱监督学习是计算机视觉任务中的一个重要问题，比如图像分类，物体识别等，因为其可以工作于干净标签不可用的大规模数据集。然而，尽管有许多关于弱监督图像分类的研究，这些方法通常局限于单标签噪声或者多标签噪声场景。本文提出了一种有效的弱监督图像分类方法，使用了大量噪声标注数据和少量干净标注数据(比如5%)。该方法由一个主干网络，一个净化网络(clean net)和一个残差网络(residual net)组成。具体来说,主干网络用于学习图像特征，净化网络用于学习从特征空间到干净标签空间的映射，残差网络用于学习从特征空间到标签残差(干净标签和噪声标签之间)的映射。因此，残差网络以类似于正则项的工作方式增强净化网络的学习。我们在两个多标签数据集(OpenImage和MS COCO2014)和一个单标签数据集(Clothing1M)评估了该方法。实验结果表明，该方法优于现有的最好方法，并能对于单标签和多标签场景有很好的泛化能力。



7. Spatiotemporal CNN for Video Object Segmentation (Kai Xu, Longyin Wen, Guorong Li, Liefeng Bo, Qingming Huang)

在视频目标分割任务中，为了保持时序一致性，大多数方法使用光流。但是光流需要大量的人工标注并且经常出现误差。为了解决这个问题，我们提出了一个统一的，可以端到端训练的时空卷积网络来解决视频目标分割任务，避免依赖光流。我们的模型由时域一致性分支和空域分割分支组成。时域一致性分支采用对抗方式以及无标注视频数据预训练，它可以捕捉目标表现的动态和运动信息，从而提高分割的时序一致性。空域的分割分支关注通过学到的表现和运动信息得到准确的分割结果。为了得到准确的分割结果，我们设计了一个从粗到细的多尺度注意力模块。通过这个模块，分割分支会逐渐关注目标区域。两个分支可以联合端到端训练。广泛的实验证明我们的方法取得了与目前最好方法的相近性能。

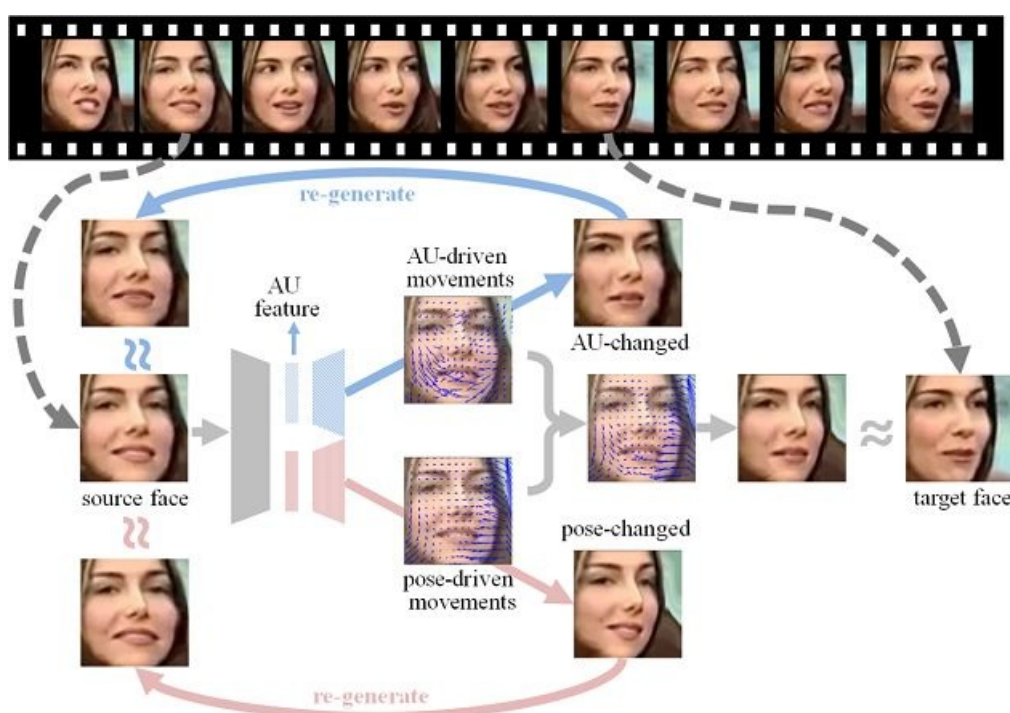


8. Twin-Cycle Autoencoder: Self-supervised Representation Learning from Entangled Movement for Facial Action Unit Detection (Yong Li, Jiabei Zeng, Shiguang Shan, Xilin Chen)

人脸表情的分析及识别主要有两种定义方法，一种是六类离散表情，另一种是面运动编码系统定义各个面部动作单元。面运动编码系统从人脸解剖学的角度，定义了44个相互独立又相互关联的面部动作单元（Action Unit(AU)）用于描述人脸局部区域的变化。相比于六类离散表情，面部动作单元能够更客观、精确、细粒度的描述人脸表情。

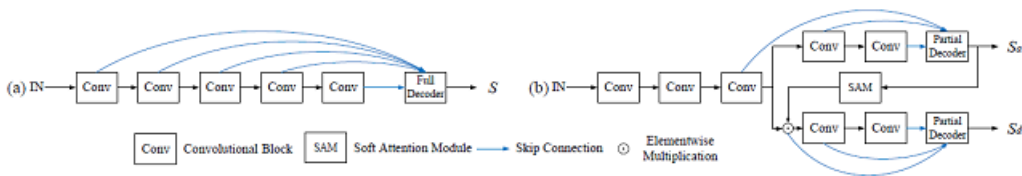
昂贵的标注代价在很大程度上限制了这一问题的研究进展。其原因在于不同的AU分布在人脸的不同区域，表现为不同强度、不同尺度的细微变化。具体来说，为一分钟的人脸视频标注一个AU，需要AU标注专家花费30分钟。

我们提出了一种能够在不依赖AU标签的前提下，从人脸视频数据中自动学习AU表征的方法（Twin-Cycle Autoencoder, TCAE）。该方法以两帧人脸图像之间的运动信息为监督信号，驱使模型预测出的两帧人脸图像之间的运动信息。考虑到两帧图像之间的运动信息包含了AU以及头部姿态（pose）的运动，TCAE通过一定的约束，使得模型能够提取出与头部姿态无关的运动分量。该方法的核心在于，模型分别提取了两张输入图像的AU-specific以及pose-specific的表征，用于解码出AU-specific以及pose-specific的运动分量。



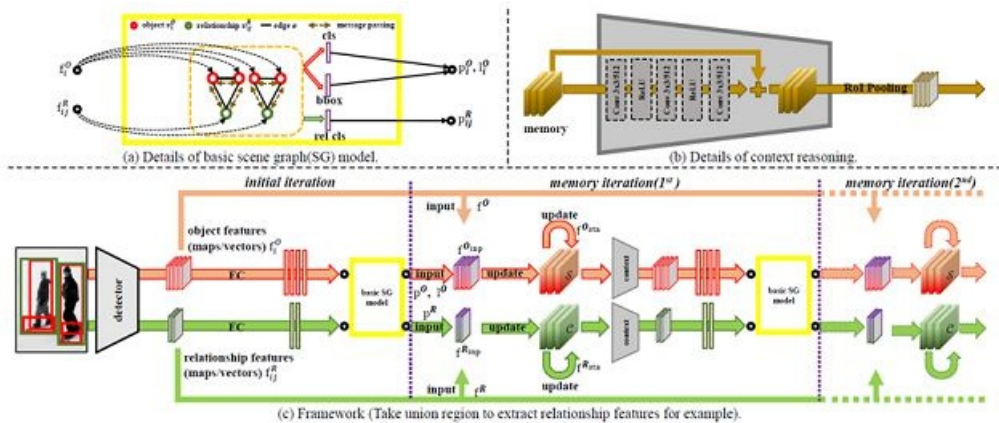
9. Cascaded Partial Decoder for Fast and Accurate Salient Object Detection (Zhe Wu, Li Su, Qingming Huang)

现有的显著性目标检测算法依赖于聚合预训练的神经卷积网络提供的多层特征。和高层特征相比，底层特征对性能的贡献小，但是由于它们的分辨率大，会带来更高的计算开销。本文提出了一种级联的部分解码器网络用于快速精确的显著目标检测。在构建解码器时，我们不融合高分辨率的低层特征以加速模型。同时我们发现只融合高层特征会得到相对精确的显著图。因此我们使用生成的显著图来对深度特征进行去噪。这样可以有效的抑制特征中的干扰部分并提高其表达能力。在多个主流的数据集上的实验，验证了我们的方法在性能的计算效率上都优于现有的方法。同时我们的框架也已用于提升现有的基于多层特征聚合的显著性检测算法，实验表明我们的框架可以明显的提升它们的性能和效率。



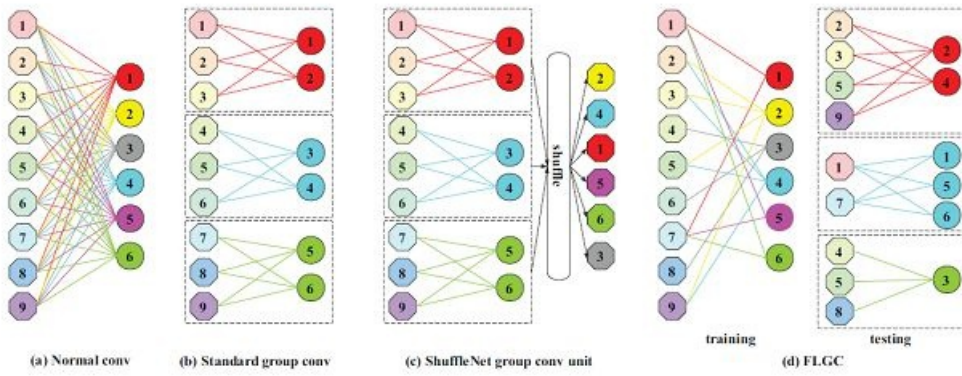
10. Exploring Context and Visual Pattern of Relationship for Scene Graph Generation (Wenbin Wang, Ruiping Wang, Shiguang Shan, Xilin Chen)

场景图有助于高层次场景图像理解，而物体间关系的预测是构建场景图的关键。在物体检测中，上下文的关联信息的使用以及物体本身特征的优化已经取得很大进展，但在关系检测中鲜有探索。本文提出一种挖掘关系之间上下文关联信息的方法，通过将所有检测到的关系融合，构建隐式的关系关联，在基准数据集 Visual Genome 上实验证明了这种关系的上下文关联在提升关系预测精度上的有效性，并且捕获到了这种关系上下文关联的意义。同时，在关系自身的特征提取方面，传统的采用两个物体的联合框提取特征的方法或者融合两者特征的方法过于关注物体信息，忽视了对关系本身视觉模式的发掘。本文提出一种采用相交框提取关系特征的方法，实验证明其更能体现关系的视觉模式，也有助于提升关系预测的性能。



11. Fully Learnable Group Convolution for Acceleration of Deep Neural Networks (Xijun Wang, Meina Kan, Shiguang Shan, Xilin Chen)

得益于深度学习在各种任务上取得了巨大的进展，其渐渐地被部署在各种低能耗的设备上（例如：手机、FPGA 等）。为了应对深度学习模型带来的大量计算和内存占用，目前大多数方法尝试加速预训练模型或者直接设计高效的网络结构。本文基于同样的目标，提出一种高效且可以嵌入任何深度神经网络模型的完全可学习组卷积（FLGC），FLGC 的结构可以在训练阶段以一种端到端的模式完成学习，这种学习获得的结构比现存的硬分配、两步法甚至迭代法获得的结构更符合任务需求。FLGC 结合深度可分离卷积（depthwise separable convolution）可以在原始 Resnet50 上实现单 CPU 加速 5 倍。同时，本文提出的 FLGC 的卷积组数不必像传统组卷积那样满足是根据我们的实验结果，在同样的卷积组数情况下，FLGC 比传统组卷积在分类问题上获得了更高的准确率。



声明：本文来源于中国科学院

如有侵权，联系删除

加群交流

欢迎加入CV联盟群了解CV和ML等领域前沿资讯

扫描添加CV联盟微信拉你进群，备注：**CV联盟**



热文荐读

[超全 TensorFlow 代码集，随查随看！](#)

[资源下载 | 复旦大学邱锡鹏教授中文版《神经网络与深度学习》，配备代码和章节练习](#)

[一文详解数据科学顶级竞赛！除了Kaggle、天池、KDD...你还知道哪些？](#)

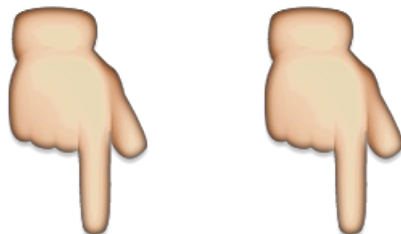
[Google CVPR 2019最新成果！NAS-FPN：可扩展特征金字塔结构用于目标检测](#)

[“图灵奖得主”Yoshua Bengio在Quora上回答机器学习55个问题及答案](#)

博士团队成员
分享工作实习机会
深度学习、机器学习、人工智能
图像解译、无人驾驶、SLAM
深度学习框架 (TensorFlow)
编程语言 (C++, Python)

【计算机视觉联盟】

计算机视觉联盟



点个在看支持一下吧