

热点追踪

## 无参考图像质量评价研究进展

江西财经大学 方玉明

## 一、引言

近年来，图像数据呈爆发式增长，图像的视觉质量与用户体验密切相关，是决定计算机视觉应用的重要因素之一。不同的内在因素如成像设备老化等<sup>[1]</sup>和外在因素如拍摄环境等都与图像质量紧密相关。

图像质量评价(Image Quality Assessment, IQA)是指通过对图像信号进行相关特性分析，量化图像的视觉失真程度。评价方法主要包括主观质量评价和客观质量评价。主观质量评价指的是由主观判断获取图像视觉质量；客观质量评价指的是通过客观算法自动计算图像的视觉质量。根据计算图像视觉质量时是否需要参考图像的信息，客观质量评价可分成三类：全参考质量评价(Full Reference IQA, FR-IQA)；半参考质量评价(Reduced Reference IQA, RR-IQA)和无参考质量评价(No Reference IQA, NR-IQA)。FR-IQA模型和RR-IQA模型通过对图像的视觉特征进行分析，量化参考图像和失真图像之间的差异，计算失真图像的视觉质量。相对于FR-IQA模型和RR-IQA模型，NR-IQA模型在计算失真图像的视觉质量时不需要任何参考图像的信息，在实际应用系统中具有更广泛的应用前景。

本文重点介绍NR-IQA模型的相关研究进展和未来发展趋势，内容框架如图1所示。根据NR-IQA模型训练时是否需要图像的主观分数，将NR-IQA模型分为有监督学习的无参考模型(需要主观分数)和无监督学习的无参考模型(不需要主观分数)，并对每类模型从基于传统机器学习的模型和基于深度学习的模型两个方面进行梳理。

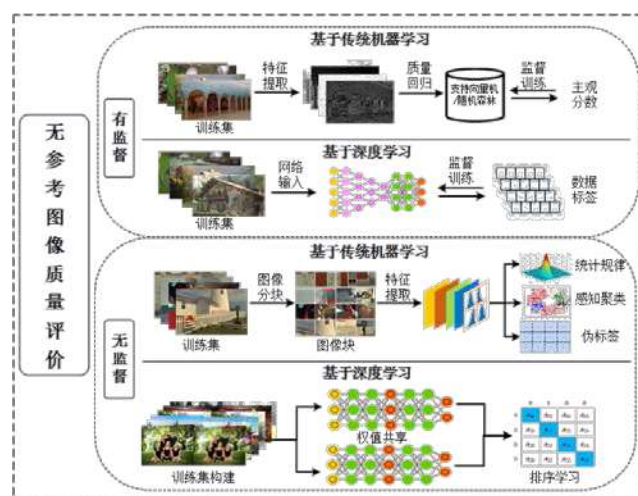


图1 本文内容框架

## 二、研究现状

## 2.1. 有监督学习的NR-IQA模型

有监督学习的NR-IQA模型指的是模型训练过程中需要用到主观数据，模型的输入为图像或者图像特征，输出为主观分数，通过迭代训练，获得最终的NR-IQA模型。有监督学习的NR-IQA模型包括基于传统机器学习算法的有监督学习模型和基于深度学习算法的有监督学习模型。前者依赖先验知识手动提取视觉特征，后者通过端到端的方式学习特征表达。

基于传统机器学习的有监督学习NR-IQA模型旨在设计有效的视觉特征表达方法，能较好地反映图像视觉质量变化，并通过传统机器学习算法如支持向量机(Support Vector Machine, SVM)或随机森林(Random Forest, RF)等学习视觉特征到图像视觉质量的映射模型，此类模型的研究重点在于特征提取。Mittal等人<sup>[2]</sup>较早提出使用空域下的自然场景统计(Natural

Scene Statistics, NSS)特征来构建 IQA 模型。与此方法<sup>[2]</sup>的设计思路类似, Fang 等人<sup>[3]</sup>研究发现图像的矩特征和信息熵呈现某种分布,并提出使用拟合得到的分布函数的输出值作为失真图像的特征,以构建 NR-IQA 模型。相较于 Mittal 等人提出的方法<sup>[2]</sup>特征提取过程需要拟合函数分布, Wu 等人<sup>[4]</sup>则设计了特征提取过程不需要拟合函数分布的 NR-IQA 模型。

不同于传统手工提取特征的 NR-IQA 模型,基于深度学习的模型可以端到端学习图像与图像质量之间的映射关系,且性能远超过基于传统机器学习的模型,该类模型的重点在于卷积神经网络(Convolutional Neural Network, CNN)的设计。除了直接端到端学习图像至质量分数之间映射的模型<sup>[5,6]</sup>外,研究人员也设计了包含多个模块的模型<sup>[7]</sup>。为提高 NR-IQA 模型性能, Ma 等人<sup>[8]</sup>将 NR-IQA 任务分成多个子任务,利用其他任务辅助 IQA 任务。总体而言,基于深度学习的 NR-IQA 模型在大多数数据库上的表现都超过了基于传统机器学习的模型,性能接近饱和。

## 2.2. 无监督学习的 NR-IQA 模型

无监督学习的 NR-IQA 模型指的是模型训练过程中不需要主观数据。类似有监督学习的 NR-IQA 模型,无监督学习的 NR-IQA 模型,主要包括基于传统机器学习的模型和基于深度学习的模型。

基于传统机器学习算法的无监督 NR-IQA 模型通过构建统计模型度量图像质量的差异,基于深度学习的无监督 NR-IQA 模型通过构造有效训练数据,构建神经网络模拟排序信息的学习过程,实现图像与质量分数之间的映射。该类方法一定程度上解决了数据驱动模型中训练数据匮乏的问题,且性能远高于基于传统机器学习的 NR-IQA 模型。如 Ma 等人<sup>[9]</sup>提出一个基于多标签学习的 NR-IQA 模型,通过对收集的高清图像添加若干种类的人工合成失真得到失真图像数据集,同时利用若干种现有 IQA 方法对失真图像数据集进行质量分数预测,基于图像对生成机制生成图像对,通过两两比较若干 IQA 方法计算的质量分数得到图像对对应标签。该方法提出的训练图像数据扩增方法,在一定程度上解决了现有 IQA 模型因图像数据不足而可能导致模型过拟合的

问题,且所提算法可高效结合现有 IQA 方法,模型简单有效,可扩展能力强。

与基于传统机器学习的有监督 NR-IQA 模型类似,基于传统机器学习的无监督 NR-IQA 模型同样依赖手工特征提取,其性能也受到特征表达能力的影响。同时,常用的使用 FR-IQA 模型获取图像伪标签的方式存在明显缺陷。另外,有限的训练数据内容也是影响这类模型可扩展能力的关键因素。基于深度学习的无监督 NR-IQA 模型可以充分利用海量的无标签数据,通过对比的方式或者借用 FR-IQA 模型产生伪标签,用于模型训练,可有效提升算法性能和算法扩展性,值得进一步研究。

## 2.3. 数据库

自然图像数据库可分为合成失真数据库和真实失真数据库两部分。常用的合成失真数据库包括 LIVE(Laboratory for Image and Video Engineering)数据库<sup>[10]</sup>和 TID2013(Tampere Image Database 2013)数据库<sup>[11]</sup>等。随着对真实失真应用场景研究的不断深入,一系列真实失真数据库不断涌现,包括 LIVE Challenge 数据库<sup>[12]</sup>和 KonIQ-10k(Konstanz Authentic Image Quality Database)数据库<sup>[13]</sup>等。

## 三、未来发展方向

无参考图像质量评价在评价内容和客观模型方面,均存在潜在的发展方向。就评价内容而言,现有的 IQA 数据库大多通过人工添加常见噪声生成,且获取的失真图像内容形式较为单一,这些失真图像仅仅来源于数量有限的高质量图像。在真实应用系统中,需要处理的图像/视频数据内容形式复杂多样。并且,这些数据在生成、传输、处理等过程中可能经历了多重失真,引入的失真无法完全使用计算机进行模拟生成。这不仅限制了现有合成失真数据库的应用价值,还限制了针对人工添加噪声数据设计的客观评价模型的可扩展性。未来仍需要投入大量精力对真实失真图像、特定应用场景下的失真图像数据进行研究。虽然陆续提出了自然失真图像质量评价的解决方案,该类研究仍然具有较大的挑战性,提升空间依然较大。

责任编辑 储珺

## 参考文献

- [1] Fang Y M, Zhu H W, Zeng Y, Ma K D and Wang Z. 2020. Perceptual quality assessment of smartphone photography//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 3677-3686 [DOI: 10.1109/CVPR42600.2020.00373]
- [2] Mittal A, Moorthy A K and Bovik A C. 2012. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing, 21(12): 4695-4708 [DOI: 10.1109/TIP.2012.2214050]
- [3] Fang Y M, Ma K D, Wang Z, Lin W S, Fang Z J and Zhai G T. 2015. No-reference quality assessment of contrast-distorted images based on natural scene statistics. IEEE Signal Processing Letters, 22(7): 838-842 [DOI: 10.1109/LSP.2014.2372333]
- [4] Wu J J, Zhang M, Li L D, Dong W S, Shi G M and Lin W S. 2019. No-reference image quality assessment with visual pattern degradation. Information Sciences, 504: 487-500 [DOI: 10.1016/j.ins.2019.07.061]
- [5] Kang L, Ye P, Li Y and Doermann D. 2014. Convolutional neural networks for no-reference image quality assessment//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE: 1733-1740 [DOI: 10.1109/CVPR.2014.224]
- [6] Kim J, Nguyen A D and Lee S. 2019. Deep CNN-based blind image quality predictor. IEEE Transactions on Neural Networks and Learning Systems, 30(1): 11-24 [DOI: 10.1109/TNNLS.2018.2829819]
- [7] Pan D, Shi P, Hou M, Ying Z F, Fu S Z and Zhang Y. 2018. Blind predicting similar quality map for image quality assessment//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE: 6373-6382 [DOI: 10.1109/CVPR.2018.00667]
- [8] Ma K D, Liu W T, Zhang K, Duanmu Z F, Wang Z and Zuo W M. 2018. End-to-end blind image quality assessment using deep neural networks. IEEE Transactions on Image Processing, 27(3): 1202-1213 [DOI: 10.1109/TIP.2017.2774045]
- [9] Ma K D, Liu X L, Fang Y M and Simoncelli E P. 2019. Blind image quality assessment by learning from multiple annotators//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 2344-2348 [DOI: 10.1109/ICIP.2019.8803390]
- [10] Sheikh H R, Wang Z, Cormack L and Bovik A C. 2006. LIVE image quality assessment database release 2[DB/OL]. [2020-07-08]. <https://live.ece.utexas.edu/research/quality/>
- [11] Ponomarenko N, Jin L N, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F and Jay Kuo C C. 2015. Image database TID2013: peculiarities, results and perspectives. Signal Processing: Image Communication, 30: 57-77 [DOI: 10.1016/j.image.2014.10.009]
- [12] Ghadiyaram D and Bovik A C. 2016. Massive online crowdsourced study of subjective and objective picture quality. IEEE Transactions on Image Processing, 25(1): 372-387 [DOI: 10.1109/TIP.2015.2500021]
- [13] Hosu V, Lin H H, Sziranyi T and Saupe D. 2020. KonIQ-10k: an ecologically valid database for deep learning of blind image quality assessment. IEEE Transactions on Image Processing, 29: 4041-4056 [DOI: 10.1109/TIP.2020.2967829]



## 方玉明

江西财经大学教授。主要研究方向为计算机视觉、多媒体信号处理、视觉质量评估。担任中国计算机学会计算机视觉专委会委员。

E-mail: leo.fangyuming@foxmail.com

热点追踪

## 人脸微表情检测前沿

中国科学院心理研究所 李婧婷 王甦菁

**微**表情是一种短暂的、无意识的面部表情，通常出现在个体试图隐藏真实感受的时刻。微表情的分析有很多潜在的应用价值，例如在医疗关怀、执法审讯、国家安全等领域。和常见的普通表情相比，微表情有三个显著特征：持续时间短(<500ms)、强度低和局部运动。人类用肉眼很难发现和识别如此短暂而微弱的表情。因此需要借助计算机视觉来进行相关的分析。其中，微表情检测的目标是定位微表情在视频中发生的时刻，这一研究非常具有实用价值。如果能在一段视频中准确地检测和定位到某个时间点有微表情出现，那么就说明这个人在这个时刻可能会有异常。然而，在真实场景下，从大量头部动作和人脸表情中检测分离出微小短暂的微表情是极具挑战性的任务。因此，目前微表情检测的研究相对较少。同时，由于微表情数据库的样本量较少，限制了深度学习技术与微表情检测方法的结合。近年来，中国科学院心理研究所发布了更适合用于微表情检测研究的长视频数据集 CAS(ME)<sup>2</sup>。研究长视频中微表情的检测方法将有助于实现微表情分析在真实场景的应用。

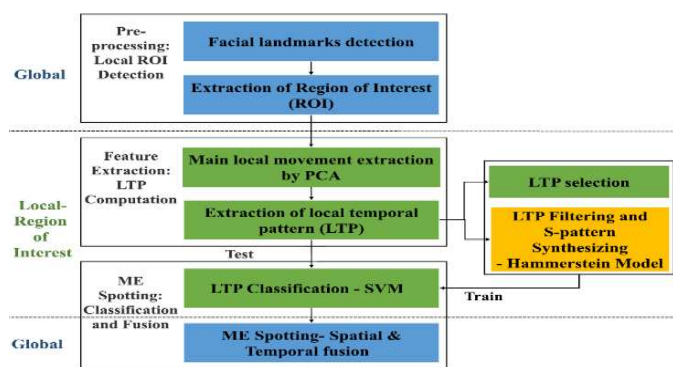


图1 结合局部时空特征提取、局部机器学习和晚期融合的微表情检测流程

针对微表情持续时间短、动作强度低以及局部发生三个特点，提出了结合局部时空特征(LTP)提取、局部机器学习和晚期融合的微表情检测方法，如图1所示。我们提出的LTP特征针对微表情的运动具有特定的变化曲线(S-模式)，如图2所示。通过使用机器学习对LTP进行分类，可以将微表情与其他面部动作区分开。此外，我们还提出了覆盖整个面部全局融合分析，以改善检测系统对微表情(局部)运动与头部(全局)运动之间的区别能力。但是，基于S-模式的机器学习模型训练受限于微表情样本数目。由于Hammerstein模型(HM)可以有效地模拟肌肉运动模式，我们利用HM生成S-模式，扩充了用于S-模式训练的数据集，从而提升了该方法的微表情检测性能。

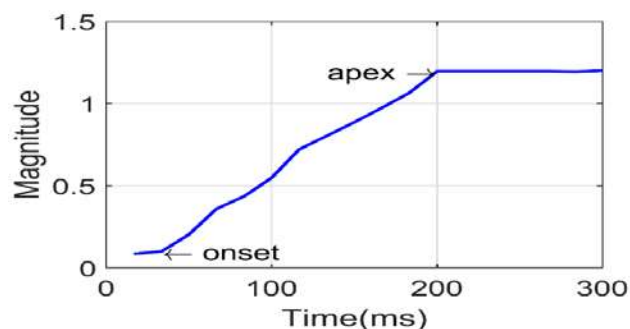


图2 针对微表情的局部时空模式(LTP)

此外，受启发于空间域上的目标检测技术，我们尝试将相应技术扩展到时间域上进行微表情检测，并提出了一个多尺度的卷积神经网络(CNN): MESNet。据我们所知，这是CNN首次与长视频中微表情检测方法相结合。基于轻量级的2+1D-时空卷积网络，如图3所示，我们增加了一个片段候选网络(Clip Proposal Network, CPN)和分类回归网络(CRNet)。其中，CPN

被添加到 2+1D-时空卷积网络预训练的卷积层中,并将所有可能的微表情区间列入候选结果,如图 4 所示。然后,CRNet 模块对候选区间进行进一步判断,并且回归它们的时间边界,从而得出最终的微表情片段检测结果,如图 5 所示。实验结果表明,我们所提出的 MESNet 算法在长视频样本中取得了最优的微表情片段检测性能。

以上工作已分别被国际期刊 IEEE Transactions on Affective Computing(TAC)和 IEEE Transactions on Image Processing(TIP)接收。

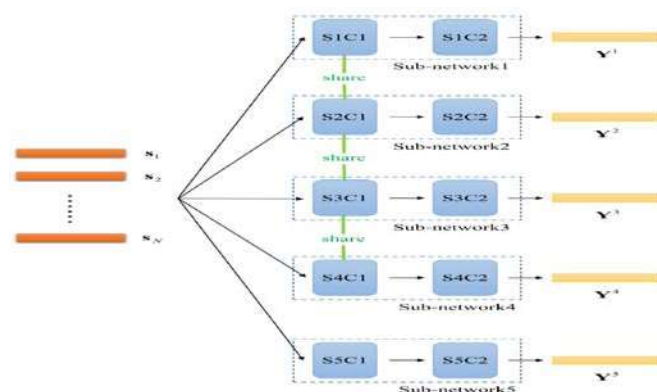


图 4 片段候选网络 (Clip Proposal Network, CPN)

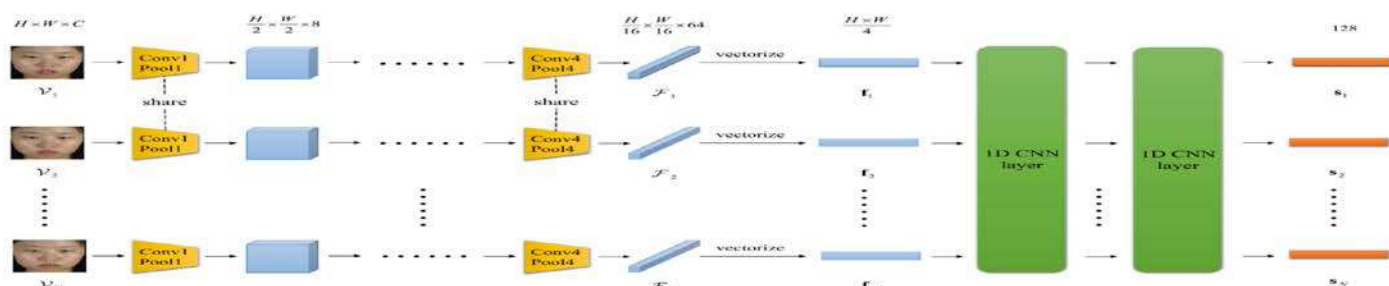


图 3 2+1D-时空卷积网络

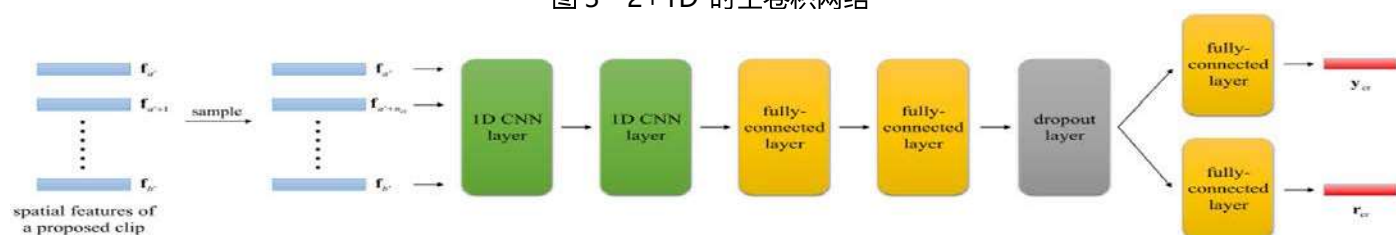


图 5 分类回归网络 (CRNet)

责任编辑 魏秀参

### 参考文献

- [1] Li, Jingting, Catherine Soladie, and Renaud Seguier. "Local Temporal Pattern and Data Augmentation for Micro-Expression Spotting." IEEE Transactions on Affective Computing (2020).
- [2] Wang, Su-Jing, He, Ying, Li, Jingting and Fu, Xiaolan. "MESNet : A Convolutional Neural Network for Spotting Multi-Scale Micro-Expression Intervals in Long Videos." IEEE Transactions on Image Processing (2021).



### 王甦菁

中国科学院心理研究所副研究员, 博士生导师。主要研究方向为模式识别与机器学习, 特别是微表情识别。在国内外重要期刊和学术会议上发表五十余篇论文, 包括 TIP、TNN、ECCV 等。获 2018 年第八届吴文俊人工智能科学技术奖一等奖。担任中国计算机学会计算机视觉专委会委员。  
Email: wangsu.jing@psych.ac.cn



### 李婧婷

中国科学院心理研究所博士后。主要研究方向为图像处理与计算机视觉, 特别是微表情检测与识别。Email: lij.t@psych.ac.cn

热点追踪

## 基于深度哈希的大规模细粒度图像检索

南京理工大学 魏秀参 沈阳

随着网络图像数据的快速增长，在面向大规模图像检索时，图像表示学习与哈希学习相结合的深度哈希技术已成为发展趋势。从大规模数据集中检索图像会面临查询速度和高冗余存储成本两个问题，而细粒度图像的细微视觉差距更是加剧了图片识别的难度。为缓解这些技术难题，我们研究了基于深度哈希的细粒度图像检索主题，通过局部和全局特征的提取并在此基础上进行局部特征对齐与交换，保证了局部特征在图像间的识别能力和语义一致性，为细粒度图像生成紧凑的二进制哈希编码。此外，与其他最近邻方法相比，提出的方法在速度与存储空间上也有较大的改善。相关成果被 ECCV 2020 收录为大会口头报告论文。

得益于深度学习发展，哈希学习已被证明是一种有成效的大规模图像检索解决方案，因它可以大大降低存储成本，提高查询速度。哈希学习作为近似最近邻(ANN)搜索代表性研究领域，近年来哈希学习已经成功应用于大规模图像检索任务。我们希望探讨哈希算法在细粒度图像检索中的有效性，这也是研究细粒度哈希问题的首项工作，该任务主要涉及如何为细粒度对象设计有效且高效的哈希编码问题。如图 1 所示，细粒度图像具有类间差异小、类内差异大的特性，细粒度哈希方法需要为具有相同子类的图像生成紧凑二进制哈希编码并为不同子类的图片生成差异较大编码。为了解决这一具有挑战性任务，我们提出了 ExchNet 网络，其框架主要由三个模块组成：(1)表示学习模块，用以获取细粒度目标的局部和全局特征；(2)局部特征对齐模块，对从不同细粒度图像中获得的局部特征进行匹配；(3)哈希码学习模块，用于生成二进制哈希编码。整体框架结构如图 2 所示。

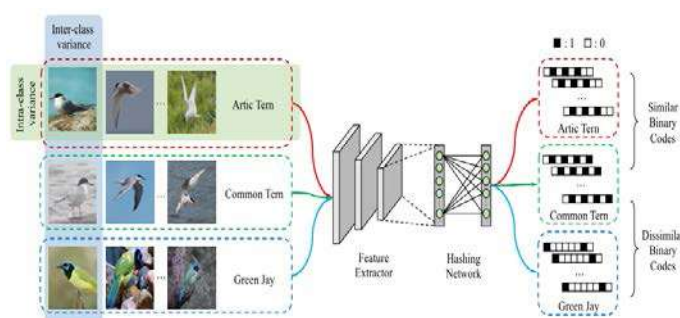


图 1 细粒度哈希任务示意图

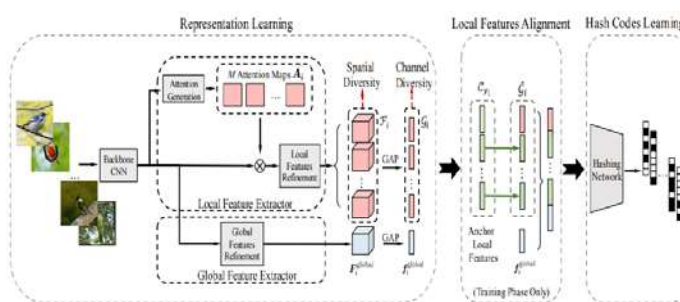


图 2 ExchNet 整体框架

在表示学习模块中，除了获取全局特征外，我们还使用注意力机制捕捉局部特征以表示图像的细粒度特征。特征对齐的方法如图 3 所示，其主要基于假设：对于两张具有相同子类的图片，交换对应细粒度对象部件(part)的局部特征(如图 3 中两只鸟的翅膀)，不会影响这两张图片哈希编码的生成以及它们的相似性。在训练过程中我们会交换同类样本的局部特征并同时保证同类样本哈希编码相似性，进而达到隐式的特征对齐目的。

为了证明提出方法的实用性和有效性，我们首先与其他 ANN 方法进行比较。在数据集 FOOD101 上，与线性搜索相比，我们的方法可以在 512 维和 1024 维的特征上分别达到 233 倍和 395 倍的加速比，同时方法

的内存开销也比基于树的方法少。在表 2 中，我们对比较了 8 种经典的针对通用图像的哈希学习方法，并比较了平均精度(mean Average Precision, MAP)的检索结果，可以看出我们的方法在选取的所有细粒度数据集中均能取得最好的检索精度。

表 1 ExchNet 与 ANN 方法在 Food101 上的性能比较

Method	512-dim				1024-dim			
	P@10(↑)	WTime(↓)	Speedup(↑)	Memory(↓)	P@10(↑)	WTime(↓)	Speedup(↑)	Memory(↓)
Linear	80.05%	9,481.03	1x	207.2MB	80.28%	22,377.96	1x	414.1MB
BallTree	77.22%	236.23	40.13x	28.1MB	77.74%	213.88	104.62x	28.1MB
KDTree	77.42%	70.16	135.13x	28.8MB	77.73%	73.57	304.14x	28.7MB
PQ	77.12%	43.49	217.99x	524.5KB	77.18%	72.47	308.74x	1.0MB
Ours	77.69%	40.54	233.85x	404.0KB	78.06%	56.57	395.53x	404.0KB

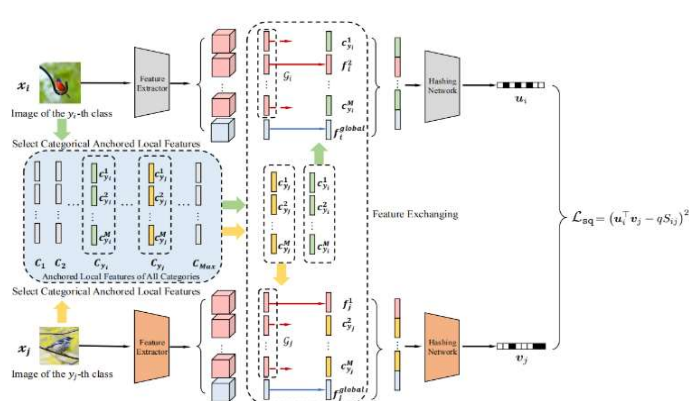


图 3 局部特征对齐方法

表 2 多个细粒度图像标准数据集的检索精度(MAP)比较

Method	#Bits	LSH	SH	ITQ	SDH	DPSH	DSH	HashNet	ADSH	Ours
CUB	12bits	2.26%	5.55%	6.80%	10.52%	8.68%	4.48%	12.03%	20.03%	25.14%
	24bits	3.59%	6.72%	9.42%	16.95%	12.51%	7.97%	17.77%	50.33%	58.98%
	32bits	5.01%	7.63%	11.19%	20.43%	12.74%	7.72%	19.93%	61.68%	67.74%
	48bits	6.16%	8.32%	12.45%	22.23%	15.58%	11.81%	22.13%	65.43%	71.05%
Aircraft	12bits	1.69%	3.28%	4.38%	4.89%	8.74%	8.14%	14.91%	15.54%	33.27%
	24bits	2.19%	3.85%	5.28%	6.36%	10.87%	10.66%	17.75%	23.09%	45.83%
	32bits	2.38%	4.04%	5.82%	6.90%	13.54%	12.21%	19.42%	30.37%	51.83%
	48bits	2.82%	4.28%	6.05%	7.65%	13.94%	14.45%	20.32%	50.65%	59.05%
NABirds	12bits	0.90%	2.12%	2.53%	3.10%	2.17%	1.56%	2.34%	2.53%	5.22%
	24bits	1.68%	3.14%	4.22%	6.72%	4.08%	2.33%	3.29%	8.23%	15.69%
	32bits	2.43%	3.71%	5.38%	8.86%	3.61%	2.44%	4.52%	14.71%	21.94%
	48bits	3.09%	4.05%	6.10%	10.38%	3.20%	3.42%	4.97%	25.34%	34.81%
VegFru	12bits	1.28%	2.36%	3.05%	5.92%	6.33%	4.60%	3.70%	8.24%	23.55%
	24bits	2.21%	4.04%	5.51%	11.55%	9.05%	8.91%	6.24%	24.90%	35.93%
	32bits	3.39%	5.65%	7.48%	14.55%	10.28%	11.23%	7.83%	36.53%	48.27%
	48bits	4.51%	6.56%	8.74%	16.45%	9.11%	17.12%	10.29%	55.15%	69.30%
Food101	12bits	1.57%	4.51%	6.46%	10.21%	11.82%	6.51%	24.42%	35.64%	45.63%
	24bits	2.48%	5.79%	8.20%	11.44%	13.05%	8.97%	34.48%	40.93%	55.48%
	32bits	2.64%	5.91%	9.70%	13.36%	16.41%	13.10%	35.90%	42.89%	56.39%
	48bits	3.07%	6.63%	10.07%	15.55%	20.06%	17.18%	39.65%	48.81%	64.19%

责任编辑 王金甲

参考文献

- [1] Q. Cui, Q.-Y. Jiang, X.-S. Wei\*, W.-J. Li, and O. Yoshie. ExchNet: A Unified Hashing Network for Large-Scale Fine-Grained Image Retrieval. European Conference on Computer Vision, Virtual Conference, 2020, pp. 189-205. (Oral Presentation).
- [2] X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou. Selective Convolutional Descriptor Aggregation for Fine-Grained Image Retrieval. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881. (ESI 高被引论文, Google Scholar Citations: 217).



魏秀参

南京理工大学计算机科学与工程学院教授。主要研究领域为计算机视觉和机器学习。担任中国计算机学会计算机视觉专委会委员。

Email: weixs@njust.edu.cn



沈阳

南京理工大学计算机科学与工程学院硕士生。主要研究领域为计算机视觉和机器学习。

Email: shenyang\_98@njust.edu.cn

顶会观察

## AAAI 2021

燕山大学 王金甲

人工智能发展协会 (Association for the Advancement of Artificial Intelligence, AAAI) (前身为美国人工智能协会 (American Association for Artificial Intelligence, AAAI)) 是人工智能(AI)领域的主要学术组织之一, 致力于促进对思维和智能行为背后机制及其在机器中体现的科学理解。第三十五届 AAAI 人工智能大会(AAAI-21)于 2021 年 2 月 2-9 日召开完全线上虚拟会议。总主席由香港科技大学杨强教授担任(这是华人首次担任此职位), 项目主席由加拿大不列颠哥伦比亚大学 Kevin Leyton-Brown 和印度德里理工学院 Mausam 担任。在中国计算机学会的国际学术会议排名中, AAAI 会议被视为人工智能领域的 A 类顶级会议, 在 2020 年谷歌学术指标总榜单排名第 96 位。

## 一、AAAI 2021 的亮点

AAAI 会议的目的是促进人工智能的研究以及 AI 研究人员, 从业人员, 科学家和所属学科的工程师之间的科学交流。与 AAAI-20 半线上会议相比, AAAI-21 会议采用了完全线上虚拟会议。每篇被接受论文的作者提前录制视频演讲并在整个会议期间提供观看, 每个与会者可以远程参与和电子互动。

AAAI-21 会议主要分为两个技术 track。主 track 包括会议大部分论文, 此外会议委员会特别鼓励并突出展示三个领域的时事话题: 神经符号 AI; AI 对新冠肺炎大流行的反应; 用于会议组织和交付的 AI。所有主 track 论文按照相同的标准和程序进行评审。次 track 聚焦人工智能对社会的影响, 其论文与主 track 论文有着不同的评审标准。

AAAI-21 会议采用两轮双盲审稿政策。第一轮审稿, 每篇论文分配两名审稿人。如果这两名审稿人认为该论文无法被录用, 那么该论文将直接被拒绝。没有被拒的论文将进入第二轮。第二轮审稿, 每篇论文分配另外两个审稿人。新的审稿人给出意见之前, 他们是无法看到第一轮的评审意见, 也就是说这两轮审稿相互独立。另外, 会议还提供了 NeurIPS 和 EMNLP 会议被拒论文的快速通道: 平均分数至少为 4.9(NeurIPS-20) 或 2.8(EMNLP-20 长论文)的被拒论文可以和先前的评审意见和论文编号等信息一起直接提交到 AAAI 的第二轮供审稿人评审。

## 二、论文录用情况

AAAI-21 会议共收到了 9034 篇论文投稿, 超过了去年的 8800 篇, 其中有效审稿为 7911 篇, 最终录取数量为 1692 篇, 接收率为 21.4%, 略高于去年的 20.6%。

据统计, AAAI-21 会议中国大陆投稿论文数量为 3319 篇, 被接收论文数量为 627 篇, 论文投稿量与接收量均为全球第一。12 篇获奖论文中有 5 篇出自华人作者团队, 包括两篇最佳论文奖论文、两篇提名奖论文以及一篇杰出论文奖论文。国内企业, 如阿里、百度、京东和腾讯分别有 54、24、21 和 11 篇论文被收录。国内高校与企业人工智能领域的影响力日趋扩大。

AAAI-21 会议接受内容包括机器学习(深度学习、统计学习等)、自然语言处理、计算机视觉、数据挖掘、多智能体系统、知识表示、人在回路中的人工智能 (human-in-the-loop AI)、搜索、规划、推理、机器人学和感知、伦理学以及结合 AI 的医疗、交通和商业等。

### 三、邀请报告

AAAI-21 会议一共邀请了 8 位杰出的演讲者，演讲内容主要包括深度学习黑盒理论、语言预训练模型、超人工智能，以及人工智能如何赋能医疗保健和传统生产行业；同时探讨了人工智能的广泛利用对个人隐私数据的影响。下面将介绍 8 位杰出演讲者的演讲内容。

**What Can and Should Humans Contribute to Superhuman AIs?** 超人人工智能不是指像人类一般的人工智能，而是使世界变得更美好的特定应用程序的人工智能。卡内基梅隆大学 Tuomas Sandholm 介绍了他在组合市场、器官交换和不完全信息游戏设置中的超人工智能领域的丰富经验。他讨论了如何发明和界定新的人工智能应用程序。在将政策优化和组合学留给人工智能的同时，人类应该如何提供价值框架将目的和方法分离开来，并以一种可扩展的方式对非常大规模的动态问题进行面向未来的优化。当定理（不仅仅是证明）和经验主义理论需要很长时间以至于超出人类理解的时候，Sandholm 对科学的未来感到疑惑。他讨论了什么是可解释性，以及为什么在许多人工智能应用中不应该要求它。最后他建议将伦理从事后讨论活动转变为一个系统设计学科，即预设计伦理。

**Recent Advances in Language Model Pretraining.** 近年来在大量原始文本上预先训练的大型语言模型彻底改变了自然语言处理。华盛顿大学 Luke Zettlemoyer 讨论了最近在语言模型预训练方面的工作，从 ELMo、GPT 到最近的序列到序列模型的预先训练方法，如 BART、mBART 和 MARGE，它们提供了迄今为止最普遍适用的方法。

**Opening the Black Box of Deep Learning (+ Takeaways for AI).** 深度学习的成功已经使一些经典的人工智能任务取得了显著进展，但不能实现完全人工智能。普林斯顿大学 Sanjeev Arora 认为目前关于深度学习的黑盒观点是实现完全人工智能的巨大障碍，并通过举例说明了他的观点。最后他认为利用数学理解黑盒现象对实现完全人工智能至关重要。

**Digital Learning Coming to Life.** 2020 年席卷

世界的全球大流行病迫使许多教学活动转为了数字形式，使用在线平台提供教学交流。这种转变既是挑战也是机会。数字学习比传统学习更具包容性，并且打开了关于教学和学习的大量数据收集的大门，这将使我们能够开发和部署人为驱动或人工智能驱动的干预措施，从而进一步提高学生的参与度和成功率。

**How AI and Technology Design will Dictate Our Civic Future.** 随着技术的进步，采用哪种技术或不采用哪种技术很可能会重新定义每一种人口统计学价值和法律。从而使数据驱动的算法和技术设计师对我们的日常生活和治理国家的准则有很大影响，例如第一波挑战就是个人数据隐私。我们不知道这一切将如何融为一体或分崩离析？

**A Tale of Two Translations.** 麻省理工学院 Regina Barzilay 介绍了她的团队在癌症诊断和药物设计这两个医疗保健领域开发和部署机器学习方法的经验。她集中讨论了实现这些功能的算法，同时强调尚未解决的技术挑战。此外，她还分享了在此过程中获得的经验教训，最后对如何扩大对医疗保健的影响，向人工智能社区提出了建议。

**AI Infusion Investment Outlook.** 李开复博士讲述了人工智能如何利用深度学习突破从重大科学发现时代走向商业实施时代。他深入探究了商业社会对人工智能注入传统行业的新阶段的期待，并阐述了企业将如何采用人工智能来提高生产率，实现数字转型。

**Talking to the Public about AI.** 牛津大学 Michael Wooldridge 讲述了他在向非专业受众谈论人工智能领域进展，以及大众对于人工智能未来走向的真实期待。

### 四、会议优秀论文

AAAI-21 会议出现了三篇最佳论文、三篇最佳提名论文以及六篇杰出论文。涉及的内容有长序列时间序列预测、多智能体学习、因果推理、序列推荐、监督学习、无监督学习、强化学习、对抗攻击和增量学习等方面。

主 track 二篇最佳论文介绍如下。

**Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting.** 长序列时间序列预测(LSTF)要求模型能够有效地捕获输出和输入之间的精确远程依赖耦合关系。然而 Transformer 存在几个严重问题使其不能直接适用于 LSTF, 例如二次方时间复杂度、高内存使用量和编码器-解码器体系结构固有局限性。为了解决这些问题, 该论文设计了一个高效的基于 LSTF 模型, 命名为 Informer。在 4 个大规模数据集上的大量实验表明, Informer 方法在很多方面均显著优于现有方法, 为 LSTF 问题提供了一种新解决方案。

**Exploration-Exploitation in Multi-Agent Learning: Catastrophe Theory Meets Game Theory.** 为了在多智能体学习(MAL)的探索开发取得进展, 该论文研究了一种类似于 Q 学习的平滑方法, 此学习模型作为研究勘探开发的最佳模型具有很强的理论依据。该论文证明了其成本模型明确捕捉了博弈成本和探索成本之间的平衡, 并且在具有异质学习代理的加权潜在博弈中, 它总能收敛于有界理性下博弈的标准解概念——量化响应均衡(QRE)集。该论文提供了一个形式化的理论, 证明了可以通过调整探索参数进行平衡选择, 从而对系统性能产生积极影响或消极影响。

次 track 人工智能对社会的影响最佳论文介绍如下。

**Mitigating Political Bias in Language Models through Reinforced Calibration.** 当前的大规模语言模型可能会因为它们所基于的数据而产生政治偏见, 当它们被部署在现实世界中时, 可能会导致严重的问题。该论文描述了衡量 GPT-2 生成的政治偏见的指标, 并提出了一个强化学习框架来减轻生成文本中的政治偏见。通过使用来自单词嵌入或分类器的奖励,

在没有访问训练数据或需要重新训练模型的情况下, 强化学习框架指导生成偏见。在对政治偏见敏感的三个属性(性别、位置和话题)的实证实验中, 提出方法在保持可读性和语义一致性的同时, 根据指标和人类评价减少了政治偏见。

## 五、研讨会

AAAI-21 研讨会计划包括 26 个研讨会, 涵盖了人工智能的广泛主题, 分别为情感内容分析、用于行为改变的 AI、用于城市交通的 AI、人工智能安全、在紧急情况下打击区域语言中的在线敌对帖子、常识知识图谱、内容创作和设计、图深度学习的方法和应用、为远程医疗设计的 AI、第九届对话系统技术挑战赛、人工智能中的可解释代理、学习和推理中的图和复杂结构、第五届国际健康智能研讨会、混合人工智能、后冠状病毒时代的 AI 教育想象、金融服务中非结构化数据的知识发现、训练期间可学习的网络架构、元学习竞赛、用于计算机视觉的元学习、规划行动和意识识别(PAIR 2021)、保护隐私的人工智能、人机对话的推理和学习、游戏中的强化学习、科学文献理解、迈向鲁棒安全高效的机器学习、值得信赖的 AI 医疗。

## 六、总结与展望

受新冠疫情的持续影响, AAAI-21 会议延续了完全线上虚拟会议的形式。无论是在 AAAI-21 会议的投稿论文还是接收论文中, 机器学习、计算机视觉依旧高居榜首, 自然语言处理则相较去年有所下降。随着人工智能理论不断发展, 如何将 AI 技术更好地应用于人类的生产生活成为不可忽视的问题。

责任编辑 崔海楠



王金甲

燕山大学信息科学与工程学院教授, 主要研究方向为信号处理和模式识别。担任中国计算机学会计算机视觉专委会委员。

Email: wjj@ysu.edu.cn