

专题概述

图像美学质量评价进展

山东财经大学 崔超然

一、引言

传统的图像质量评价研究主要关注在图像的获取、传输和存储等过程中引入的不同类型的失真(如噪声、扭曲、压缩等),以及由此带来的图像质量下降问题。随着图像采集设备的不断升级和编解码技术的快速进步,普通图像的失真问题目前已经得到一定程度的缓解。与此同时,一些研究者开始尝试从美学的角度对图像质量进行客观评价,并提出了图像美学质量评价(image aesthetics assessment)的概念。

长久以来,美学(美感)一直作为哲学问题来讨论。美的作品由人来评判,并使人们获得审美的愉悦。但是,面对日益激增的海量图像资源,人们已经无法胜任繁重的评判工作,希望计算机能够进行辅助美学分析。图像美学质量评价旨在模拟人类视觉及审美思维,进而对图像进行美学建模,使计算机能够自动地对图像的美学价值进行定量的评价。

目前,图像美学质量评价技术已经被应用到多个领域。例如,针对用户拍摄的关于同一场景的多张候选照片,筛选最具美感的作品保存和展示,合理地降低数据的存储开销;在图像作品的创作和编辑中,分析对比候选方案的美学质量,提升作品的视觉美感;在图像检索系统中,考虑返回图像的美学质量,为用户提供准确且更有吸引力的检索结果。因此,在图像的存储、编辑、检索等方面,图像美学质量评价技术都具有重要的研究价值和广阔的应用前景。

二、研究现状

图像美学质量评价研究的核心在于建立图像内容

表达和图像美学判定之间的关联。伴随计算机视觉技术的快速发展,现有的图像美学质量评价方法主要关注于提取和构造有效的图像视觉特征,并结合视觉特征利用分类方法对高美感和低美感的图像进行区分,或采用回归方法来评估图像美感分值。

早期的图像美学质量评价研究主要以摄影中的美学规则为先验知识,如三分构图法、景深和视觉平衡原则等^[1],人为地设计提取图像的视觉特征。一些高层次视觉属性特征也相继被提出以建模图像的色度、亮度、内容布局等信息^[2]。尽管这些早期工作取得了一定进展,但人工设计图像美学特征往往需要较强的领域知识和一定的摄影经验。此外,在部分图像美学质量评价任务中,一些通用的图像视觉特征(如颜色直方图、SIFT等)也被证明要优于这些定制的基于摄影规则的特征^[3]。

伴随深度学习的兴起,卷积神经网络(Convolutional Neural Networks, CNN)被广泛用于自动学习图像美学特征,并取得了较好的效果。[4]是第一个将CNN用于图像美学质量评价的工作,该方法分别抽取图像的局部特征和全局特征,并利用图像的风格和语义标签进一步提升美学质量评价的性能。[5]提出了一种深度图像块融合网络,从单幅图像中随机裁剪多个图像块并将它们的特征进行融合。在此基础上,[6]提出了一种启发式的图像块选择策略,在训练阶段识别最具判别性的图像块特征。[7]引入注意力机制来自适应地调整不同图像块特征在融合时的权重以提高学习效率。[8]证明将人工设计的特征与预训练CNN提取的特征相融合,可以有效提高图像美学质量评价的性能。为了规避CNN对输入图像的固定尺寸要求,[9]在卷积层后插入

空间金字塔池化层,从而使模型可以接收任意尺寸和长宽比的图像。

另一方面,考虑到美感本质上是人的一种主观感受,不同的用户面对同一幅图像时往往会得出不同的美学评价结果,即所谓“各花入各眼”。从这点出发,[10]提出了一种面向质量等级分布的图像美学质量评价方法,直接预测一幅图像处在不同质量等级上的分布,以分布的形式反映用户关于图像审美感受的差异性。[11]和[12]分别通过挖掘用户对图像的文本评论和点赞信息,来建模用户的审美偏好,进而构建个性化的图像美学质量评价系统。[13]收集了一个带有打分者身份信息的图像美学质量评价数据集,并利用主动学习算法来预测指定打分者对图像的美学评分。[14]提出了一个具有共享权重的多任务学习网络,以预测图像的美学分布和喜欢该图像的用户的人格特质,最终通过考虑用户和图像之间的人格特质匹配程度来给出个性化的美学评价结果。[15]

提出了一种基于元学习的个性化图像美学质量评价方法,利用双层梯度优化算法仅根据少量样本就可以得到目标用户的个性化美学评价模型。

三、未来发展方向

目前,个性化图像美学质量评价逐渐受到研究人员关注,如何理解用户的审美偏好是一个亟待解决的问题。然而,用户的审美偏好是相对抽象的,且难以直接获取和表示。在社交媒体中,用户与图像之间存在多种交互行为。考虑到人的认知和行为是彼此相互影响的,可以借助用户的社交行为来感知用户的审美偏好。因此,我们认为未来可以从“社会感知”的角度尝试解决个性化图像美学质量评价问题^[16]。此外,个性化图像美学质量评价中的小样本学习和冷启动问题也是值得研究的方向。

责任编辑 任桐炜

参考文献

- [1] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z Wang. Studying aesthetics in photographic images using a computational approach. In the 8th European Conference on Computer Vision, 288–301, 2006.
- [2] Sagnik Dhar, Vicente Ordonez, Tamara L Berg. High level describable attributes for predicting aesthetics and interestingness. In the IEEE Conference on Computer Vision and Pattern Recognition, 1657–1664, 2011.
- [3] Luca Marchesotti, Florent Perronnin, Diane Larlus, Gabriela Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In the IEEE International Conference on Computer Vision, 1784–1791, 2011.
- [4] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, James Z Wang, Rating image aesthetics using deep learning. IEEE Transactions on Multimedia, 17(11): 2021–2034, 2015.
- [5] Xin Lu, Zhe Lin, Xiaohui Shen, Radomir Mech, James Z Wang. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In the IEEE International Conference on Computer Vision, 990–998, 2015.
- [6] Shuang Ma, Jing Liu, Chang Wen Chen. A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. In the IEEE Conference on Computer Vision and Pattern Recognition, 4535–4544, 2015.
- [7] Kekai Sheng, Weiming Dong, Chongyang Ma, Xing Mei, Feiyue Huang, Bao-Gang Hu. Attention-based multi-Patch aggregation for image aesthetic assessment. In the 26th ACM International Conference on Multimedia, 879–886, 2018.
- [8] Michal Kucer, Alexander C Loui, David W Messinger. Leveraging expert feature knowledge for predicting image aesthetics. IEEE Transactions on Image Processing, 27(10): 5100–5112, 2018.
- [9] Long Mai, Hailin Jin, Feng Liu. Composition-preserving deep photo aesthetics assessment. In the IEEE Conference on Computer Vision and Pattern Recognition, 497–506, 2016.

- [10] Chaoran Cui, Huiui Liu, Tao Lian, Liqiang Nie, Lei Zhu, Yilong Yin. Distribution-oriented aesthetics assessment with semantic-aware hybrid network. *IEEE Transactions on Multimedia*, 21(5): 1209–1220, 2019.
- [11] Guolong Wang, Junchi Yan, Qin Zheng. Collaborative and attentive learning for personalized image aesthetic assessment. In the 27th International Joint Conference on Artificial Intelligence, 957-963, 2018.
- [12] Chaoran Cui, Wenya Yang, Cheng Shi, Meng Wang, Xiushan Nie, Yilong Yin. Personalized image quality assessment with social-sensed aesthetic preference. *Information Sciences*, 512: 780-794, 2020.
- [13] Jian Ren, Xiaohui Shen, Zhe Lin, Radomir Mech, David J Foran. Personalized image aesthetics. In the IEEE Conference on Computer Vision and Pattern Recognition, 638-647, 2020.
- [14] Leida Li, Hancheng Zhu, Sicheng Zhao, Guiguang Ding, Weisi Lin. Personality-assisted multi-task learning for generic and personalized image aesthetics assessment. *IEEE Transactions on Image Processing*, 29: 3898-3910, 2020.



崔超然

山东财经大学计算机科学与技术学院教授，主要研究方向为数据挖掘与多媒体信息处理。
Email: crcui@sdufe.edu.cn。

热点追踪

双路对抗自编码与聚类应用

中山大学 任传贤

无监督聚类作为计算机视觉和模式识别领域的基本任务，一直被视为数据分析中探索无标签数据结构和提取分类信息的重要途径。经典的聚类算法如K-means、高斯混合模型(GMM)旨在通过数据分布提取分类信息，然而该信息一般与风格、背景等干扰信息耦合在一起，这类算法又大多需计算原始数据空间中样本间的距离，致高维情况下处理结构复杂数据的效果不尽人意。另一类以谱聚类为代表的聚类算法旨在直接识别类别，但其效率与样本数量息息相关，因而难以处理大规模数据集。

近年来，深度学习促进了聚类方法在高维空间的发展并建立了各种具有显著性能模型。以深度生成式模型为例，它们可在某些有关潜变量的假设下学习潜变量的数据分布，进而编码得到丰富的潜在结构。变分自编码器(VAE)是训练生成式模型的一个重要框架，其利用随机梯度变分贝叶斯和重参数化技巧来最大化似然函数，但也可能出现过度正则化而导致聚类效果退化。对抗自编码(AAE)在AE基础上引入对抗式训练机制，将隐层编码的聚合后验分布与一任意先验分布匹配，保证模型尽可能接近复杂的先验分布。AAE在图像类型和内容的分离、降维、数据可视化等应用中具有良好的性能，然而其最大化似然函数这一目标与聚类目标不全相符，同时对对抗式机制容易产生模式崩塌。因此，如何解决相关问题以获得更准确、更稳定的聚类效果仍是研究热点。

基于以上目标我们提出了双路对抗自编码器模型(Dual-AA)，该模型在最大化似然函数的同时还最大化观测样本及潜变量子集之间的互信息，从而捕获到更多数据结构信息，似然函数和互信息的变分下界通过相同

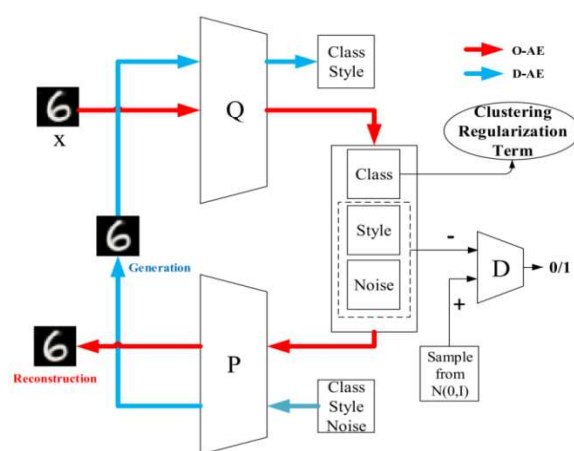


图 1 Dual-AAE 主要网络结构

的变分分布推断得到。Dual-AAE由用于重构输入数据的自编码O-AE和重构潜变量的自编码D-AE构成，两个AE共享网络参数。我们对O-AE中的潜变量采用对抗机制以获得更好的数据流形，并引入基于后验分布和边际分布熵的聚类正则项(CR)代替对抗式训练。该模型不仅能有效提取分类信息，而且不易出现模式崩塌的风险。

Dual-AAE的主要网络架构如图1所示。红色箭头表示D-AE的数据流方向，观测样本通过编码器Q编码为类别、风格和噪声变量，再经由解码器P重构，该过程最小化观测样本的重构误差以确保潜变量包含全部信息。在D-AE的编码层通过CR项获得分类变量的更多信息，同时将风格变量和随机噪声输入判别器D，对聚合后验正则化以使其与高斯先验匹配。蓝色箭头表示O-AE的数据流方向，从先验分布提取出的类别、风格和噪声变量输入编码器P生成新样本，再由解码器Q重构，最小化重构误差来减少信息丢失。为进一步提高聚类效果，基于大多数错分样本的最大聚类概率远小于1的经验观察，可以引入拒绝

项来避免对困难样本做决策，利用更有效的方法如人工监督单独解决。

Dual-AAE模型的目标函数由多部分构成：1) 重构损失。O-AAE的重构损失鼓励更好地解释输入数据，D-AAE的重构损失鼓励结构化潜变量从数据中提取尽可能多的信息；2) 对抗损失。利用对抗式训练过程正则化风格变量和随机噪声的后验以匹配高斯先验；3) 聚类正则项，促使类别变量从输入数据中捕获分类信息。利用反向传播算法对网络参数进行优化后，Dual-AAE将目标数据分配给后验概率最高的类别作为聚类结果。一种难以避免的情况是，一些容易混淆的样本即使通过有监督的分类算法也无法有效处理。我们这里通过添加拒绝项，设置阈值并拒绝最大后验分布不超过该值的输入样本，用人工注释等更有效的监督方法来处理。

我们的新型生成式聚类算法Dual-AAE为聚类问题提供了训练简化且稳定的模型，其有效性在4个公开的数据集上进行了实验验证，表现普遍优于现有多数较为流行的聚类算法。详细结果如表1所示。在MNIST数据集上对带有拒绝项的Dual-AAE进行评估，其在少数样本被拒

绝的情况下几乎可以正确聚类所有样本。特别的，在无监督条件下该模型还能区分图像的风格内容。

除了聚类任务，由于对抗式自编码模块的存在，Dual-AAE还可以从特定群集中生成高质量的样本，即实现图像生成，这对完成图像自动生成(采样)、图像信息补全等工作具有一定的借鉴意义。图2可见，(a)VAE和(b)GAN无法生成指定类别的样本，而(d)Dual-AAE不仅克服了该缺点，而且得到比生成模型(c)InfoGAN更平滑真实的图像。

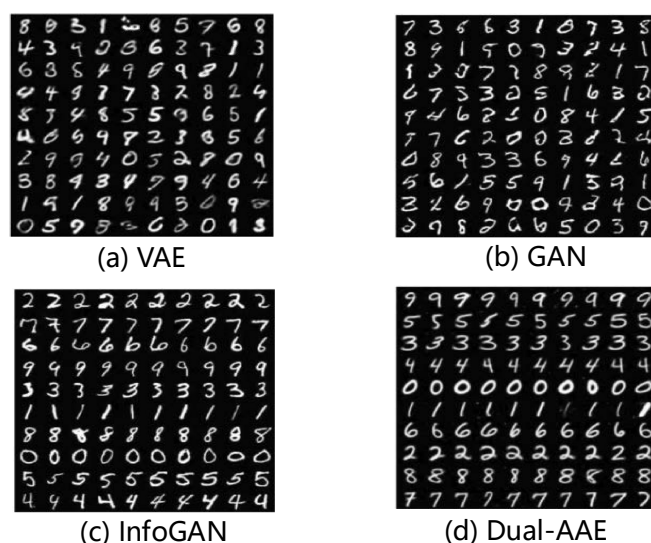


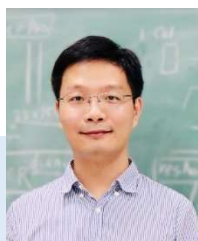
图 2 不同模型的生成图像

整体来说，相比VAE，虽然Dual-AA 也利用变分推断学习生成模型，但其同时最大化似然函数和互信息，并且引入聚类正则项，不仅提高了聚类效果，还摆脱模型崩塌的潜在风险，这为处理半监督学习和图像风格迁移问题也提供了可能的指导意义。以上工作于2020年发表在IEEE TNNLS。

责任编辑 储珺

表 1 不同算法的聚类准确率

Method	MNIST	HHAR	STL-10	REUTERS
K-means	53.55	60.08	64.86	54.04
N-Cuts	44.83	53.60	64.23	N/A
GMM	53.73	60.34	72.44	55.81
VAE+GMM	72.94	68.02	78.86	70.98
DEC	84.30	79.86	80.62	75.63
GMVAE	88.54	-	-	-
DEPICT	96.50	-	-	-
A/AE	84.10	83.77	81.24	75.12
CatGAN	84.61	80.77	83.08	75.9
InfoGAN	96.33	83.37	82.14	75.33
VaDE	94.46	84.45	84.46	79.38
Dual-AAE(without CR)	97.04	85.94	86.85	79.56
Dual-AAE	97.86	86.79	89.15	81.45



任传贤

中山大学数学学院副教授，研究方向为统计模式识别。

Email: rchuanx@mail.sysu.edu.cn

热点追踪

面向自动驾驶场景的注意力引导实时语义分割

南京邮电大学 王林杰 周全

语义分割是计算机视觉中发展最快的领域之一，因为它提供了必要的基于对场景理解的像素级上下文，因此在自动驾驶中扮演了重要的角色。近年来，构建更大更深的卷积神经网络成为实现语义分割的主要趋势。大多数性能较好的 CNNs，如 ResNet 系列^[1,2]，都有成百上千个卷积层和特征通道来提高特征的表达能力。尽管这一方式可以使网络的性能得到提升，但会增加网络运行时间。而像无人机、机器人等这些资源受限的平台，不适合部署此类网络。

针对在资源受限环境中追求最佳分割结果这一热点问题，AGLNet (Attention-guided Lightweight Network)这一轻量化语义分割网络被提出来。它主要有三个模块(1)SS-nbt 单元；(2)GAUM 模块；(3)FAPM 模块。AGLNet 的整体架构如图 1 所示。分离重排非瓶颈块(split-shuffle-non-bottleneck, SS-nbt)通过通道分离和卷积分解来降低参数量，借助残差连接降低训练难度。分解注意力金字塔模块(FAPM, factorized attention pyramid module)，由金字塔特征注意力

(PFA, pyramid feature attention)和全局池化注意力(GPA, global pooling attention)两部分构成。PFA 采用层次化 U 形结构，融合上下文信息；GPA 模块用于提取图像的全局先验注意力。得益于 PFA 的层次化结构，FAPM 可以捕获多尺度的上下文信息，并对特征产生像素级注意力。受文献[3]启发，全局注意力上采样模块(GAUM, global attention upsample module)，它通过对空间注意力和通道注意力进行编码，提高网络的表示能力，同时保持较低的计算成本。GAUM 和 FAPM 的具体结构如图 2 所示。

表 1 AGLNet 与其他模型在 Cityscapes 测试集上的对比

Method	Input size	Extra data	#Params	FLOPs	FPS	mIoU
ERFNet	512 × 1024	no	2.1M	26.86G	49	68.0
CGNet	512 × 1024	no	0.5M	7.01G	64	64.8
ENet	512 × 1024	no	0.36M	4.4G	65	59.3
EDANet	512 × 1024	no	0.68M	8.95G	102	67.3
ESPNetv2	512 × 1024	ImN	1.25M	5.85G	65	66.2
FSCNN	512 × 1024	Coa.	1.14M	1.76G	230	62.8
DABNet	512 × 1024	no	0.76M	10.46G	99	68.1
FPENet	512 × 1024	no	0.12M	1.58G	110	55.2
Ours	512 × 1024	no	1.12M	13.88G	52	70.1
Ours [†]	512 × 1024	Coa.	1.12M	13.88G	52	71.3

ImN 和 Coa 分别表示用 ImageNet 或 Cityscapes 粗糙标注数据集预训练；-表示该方法未提供相应参数

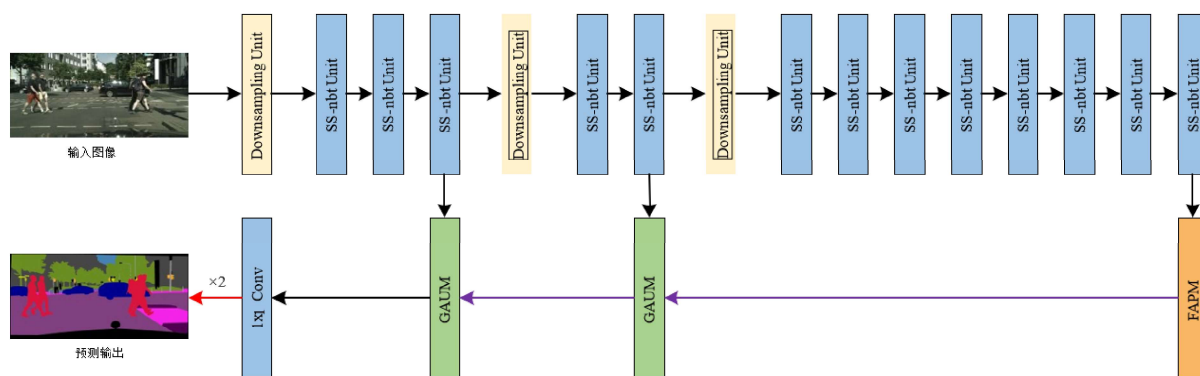


图 1 AGLNet 网络结构图

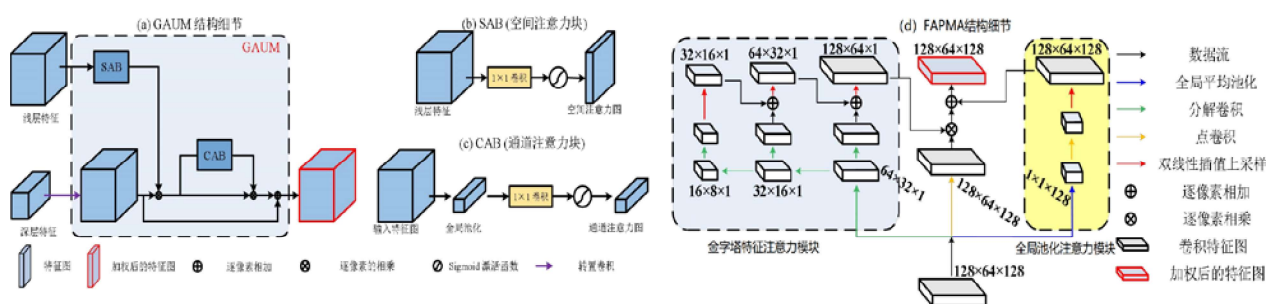


图2 GAUM 和 FAPM 结构图

AGLNet 在 Cityscapes 测试集的实验结果如表 1 所示, 可以看出, AGLNet 的分割结果优于这些方法, 并能达到实时分割的效果。在所有方法中, AGLNet 模型大小仅 1.12M; 仅使用额外粗糙训练数据时, 分割精度提高到 71.3%。实验结果表明, AGLNet 模型在 CityScapes 数据集上实现了分割精度和运行效率间的

最佳平衡。

本文^[4]于 2020 年 3 月发表在人工智能权威期刊 Applied Soft Computing。

责任编辑 任桐炜

参考文献

- [1] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
- [3] Z. L. Zhang, X. Y. Zhang, C. Peng, X. Y. Xue, J. Sun, Exfuse: Enhancing feature fusion for semantic segmentation, in: European Conference on Computer Vision, 2018, pp. 269–284.
- [4] Q. Zhou et al., “AGLNet: Towards real-time semantic segmentation of self-driving images via attention-guided lightweight network,” Appl. Soft Comput. J., vol. 96, p. 106682, 2020, doi: 10.1016/j.asoc.2020.106682



王林杰

南京邮电大学硕士研究生。主要研究方向为计算机视觉。
Email: linjiew678@163.com



周全

南京邮电大学副教授。主要研究方向为计算机视觉, 深度学习, 模式识别。
Email: quan.zhou@njupt.edu.cn

顶会观察

ECCV 2020

中国科学院自动化研究所 朱翔昱

欧洲计算机视觉国际会议(European Conference on Computer Vision, ECCV)是计算机视觉领域的顶级会议, 该会议每两年举办一次, 与 ICCV、CVPR 并称为计算机视觉领域三大国际顶会。在 2020 谷歌学术指标的榜单上高居「工程和计算机科学」领域排名第二。ECCV 的高知名度与认可度使其吸引了大量顶尖实验室、研究所以及企业机构参与其中。

ECCV 20 原定于 2020 年 8 月 23 日至 28 日在英国格拉斯哥举行, 但由于新冠肺炎的影响, 会议全部转移到线上进行, 给予了全新的参会体验。

一、欧洲计算机视觉国际会议的亮点

由于今年年初新冠肺炎的爆发, ECCV 2020 采用了数字虚拟会议的方式, 会议主办方提前进行协调, 并建立了虚拟会议的网址, 让世界各地的参会者通过互联网参加虚拟会议。会议采用摘要、图片、论文和讲解视频等作为虚拟会议的交流媒介。

本次会议具有直播和预先录制两种类型, 并拥有向演示者发布问题等交互手段。大多数现场活动分两次进行, 相互间隔 12 小时, 以便所有时区的参会者可以在方便的时间参加。所有的现场活动都被记录下来, 在 2021 年 6 月之前可以随时查看, 为不能及时参会或有时差的研究者们提供了极大的便利。本次会议在保证安全的情况下最大限度地保障了参会者的自由度和便利程度, 让参与者获得了良好的参会体验。

二、论文录用情况

ECCV 2020 共收到有效投稿 5025 篇, 接收 1361

篇(占比为 27%), 其中 oral 论文 104 篇, 仅占 2%。从总体数据来看, 今年的有效投稿是 ECCV 2018(243 篇)的两倍以上, 接受率和 oral 率占比与上届差别不大。这些数据反映出了该会议的认可度迅速提升, 也可看出被接受论文的质量之高。

本次 ECCV 20 收录论文中, 中国学者的表现相当出色。其中来自香港中文大学的作者人数最多, 排名第二, 仅次于谷歌。其他贡献突出的中国高校还包括北京大学、清华大学、上海交通大学等。除了学术界的科研力量外, 来自中国工业界的各大互联网企业如百度、腾讯、旷视等也获得了非常不错的成绩。其中腾讯 AI Lab 入选 16 篇, 旷视研究院入选 15 篇, 百度 AI 入选 10 篇。这些被录用的论文在许多重要工业应用领域有着重大的突破, 包括多模态学习、视频内容理解、自动驾驶识别及定位、声源定位及场景识别等领域。此外, 谷歌在本次会议表现十分抢眼, 共有 50 篇论文入选, 其中 10 篇入选 oral, 涵盖了诸多热门领域包括视角合成、3D 人体重建、6D 位姿估计等。

ECCV 202 会议涵盖的方向包括图像分类、目标跟踪、人脸识别、姿态估计、3D 点云、图像/视频恢复、模型加速、GAN 和 NAS 等, 论文关键词包括语义分割(Semantic Segmentation)、弱监督(Weakly Supervised)、深度估计(Depth Estimation)、对抗攻击(adversarial attack)和领域自适应(Domain Adaptation)等。

三、会议奖项

除了常规的 Best Paper 奖项, ECCV 202 也颁发

了其它的奖项来鼓励或嘉奖对计算机视觉领域具有影响力的人物或工作。

Koenderink 奖用于表彰发表时间超过十年并经受时间检验的研究。今年 Koenderink 奖授予以下两篇论文: 1) Improving the Fisher Kernel for Large-Scale Image Classification。该论文提出的 Fisher 内核在图像分类的背景下, 被证明超越了计数统计, 扩展了流行的视觉单词包。2) BRIEF: Binary Robust Independent Elementary Features。该论文提出了一种有效的特征描述子, 能在取得比 SIFT 更好的结果前提下, 速度更快。

Mark Everingham 奖的设立是为了纪念在 2011 年去世的 Mark Everingham。他在 The PASCAL Visual Object Classes (VOC) 数据集做出了主要贡献, 也是该比赛项目的发起人。今年第一位获得 Mark Everingham 奖的是 Antonio Torralba。他持续十多年定期发布新的数据集, 包括 Tiny images, SUN / SUN - 3D, MIT - Places, 并发布了创建数据集新工具包括 LabelMe。第二位获 Mark Everingham 奖的是 COLMAP SFM 和 MVS 软件库的提出人 Johannes Schonberger, 以表彰他为 3D 图像重建提出的一个开源的端到端 pipeline, 并为之提供的支持、开发与编写文档, 这些软件库已经变成了运动结构和立体多视图的参考软件。

最佳 Demo 奖颁给了题目为 Inter-Homines: Distance-Based Risk Estimation for Human Safety 的工作。该工作能够基于视频流实时评估监控区域内的传染风险。通过在三维空间中定位人物, 计算人际距离, 并通过构建监控区域的动态地图来预测风险水平。同时提出了一个时空动态风险估计的参数模型, 可以通过预测现场的风险水平, 对社会距离预防措施的接受程度进行安全监测, 对当前疫情的管控具有重大的实际意义。

四、会议主题及热点

本次会议热点主题包括: 1) 识别、检测、分割和姿态估计; 2) 半监督、无监督、迁移、表征和小样本学习; 3) 3D 计算机视觉与机器人; 4) 图像和视频合成; 5) 视觉

和语言。此外, 一些新兴领域, 如标签高效方法(label-efficient methods)和低级视觉(low-level vision)等, 也在逐渐引起学术界的关注。

在这些主题中, 来自学术界和工业界的多篇论文引起了广泛的讨论。如 Facebook 提出的 DETR^[1], 是一种端到端的基于编码器-解码器的 Transformer 架构。DETR 使用更少的先验, 如 Anchor、NMS 等, 在准确率、运行效率上与高度优化的 Faster RCNN 基本持平, 且容易复现。但缺陷是训练时间长, 在小目标上效果不是很好。Google 提出的 BiT^[2], 重新探讨了迁移学习的简单范例: 对大量带标签的源数据(例如, JFT-300M 和 ImageNet-21k 数据集)进行预训练, 然后针对不同任务对这些训练权重进行微调, 这样既减少了目标任务所需的数据量, 又减少了微调时间。香港理工和阿里达摩院提出了一种新型的梯度中心化 (gradient centralization, GC) 优化技术^[3]: 梯度中心化, 该方法通过中心化梯度向量使其达到零均值, 从而直接在梯度上执行。GC 方法可以看作是对权重空间和输出特征空间做了正则化, 从而提升了 DNN 的泛化性能。此外, GC 还能提升损失函数和梯度的 Lipschitz 属性, 从而使训练过程更加高效和稳定。东京大学提出了一种更好的增广搜索方法 Faster AutoAugment^[4]。该论文提出将原始图像和基于梯度优化的增强图像进行密度匹配, 以找到最佳策略。该研究认为数据增广可以视为填充原始数据缺点的一种方式, 并将增强数据的分布与使用了对抗学习的原始数据之间的距离最小化。加拿大高等研究院提出了一种两步法的策略^[5]来解决小样本学习微调中过度优化的问题。该论文第一步利用隐空间的相似性度量将每个样本映射到已知的基类, 第二步对特征进行关联对齐, 微调特征提取器。

此外, 最受大家关注的则是荣获最佳论文和最佳论文提名奖的论文。荣获最佳论文奖的是 Zachary Teed 等人的 RAFT^[6], 作者是普林斯顿大学视觉与学习实验室邓嘉门下的一名博士生。RAFT 提出了一种新的光流预测结构: 递归全对场变换, 通过提取每个像素的特征, 为所有像素对构建多尺度 4D 相关体, 并通过一个循环单元迭代更新光流场。RAFT 算法具有较强的跨数据集

泛化能力,在推理时间、训练速度、参数计数等方面具有较高的效率。荣获最佳提名奖的论文之一是 NeRF^[7]。NeRF 提出一种在合成复杂场景新视图任务中实现 SOTA 结果的新方法,该方法通过使用稀疏的输入视图集来优化基础的连续体场景函数。NeRF 使用了一个 9 层、256 个通道的全连接网络将场景表示为一个连续的 5D 函数,其输入是一个单一连续的 5D 坐标,即 3D 空间位置和观看方向,其输出是 RGB 颜色和不透明度(输出密度)。为了合成一个给定的视图,渲染过程包括沿摄像机射线查询 5D 坐标,并使用经典的体积渲染技术将输出颜色和密度投射到图像中。另外一篇荣获最佳提名奖的论文是[8]。该论文第一作者 Mengtian Li 是来自卡内基梅隆大学机器人研究所的博士生,曾获得南京大学学士学位。该工作提出一种明确的指标来对比不同方法的最优延迟-准确率曲线,指出标准离线评估和实时应用之间的差异:算法处理完特定图像帧时,周围环境已经发生改变,还提出将延迟和准确率协调地集成到一个度量指标中,用于实时在线感知,即流准确率。

ECCV 2020 也举办了很多高质量的研讨会与讲习班。涉及的主题包括通用物体识别分类、对抗样本、深度估计、弱监督学习、域自适应、自动化深度学习、微软的 Hololens2 产品介绍等。这些研讨会和讲习班给学者和学生提供了一个很好的学习平台。

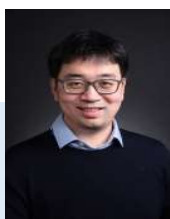
五、总结与展望

回顾 ECCV 以及 CVPR、ICCV 近几年的论文,可以发现视觉领域的几大趋势:1)自主学习,包括网络结构、数据增广、损失函数、超参的搜索等;2)无监督/半监督/自监督学习,一大关注点是如何利用大量无标签数据学习一个好的表征;3)3D 物体感知和识别,落地场景有无人驾驶、虚拟现实等。此外,受到疫情的影响,线上虚拟会议也可能是以后几年的趋势,此次 ECCV 2020 虽然搭建了虚拟的在线平台来提升体验,但依然存在很多非热门的论文的会议室无人问津的冷场情形。如何提高线上会议的体验,促进学者的交流热情,是线上会议的下一个挑战。

责任编辑 崔海楠

参考文献

- [1] Carion, Nicolas, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. "End-to-End Object Detection with Transformers." In ECCV, 2020.
- [2] Kolesnikov, Alexander, Lucas Beyer, Xiaohua Zhai, Joan Puigcerver, Jessica Yung, Sylvain Gelly, and Neil Houlsby. "Big transfer (BiT): General visual representation learning." In ECCV, 2020.
- [3] Yong, Hongwei, Jianqiang Huang, Xiansheng Hua, and Lei Zhang. "Gradient Centralization: A New Optimization Technique for Deep Neural Networks." In ECCV, 2020.
- [4] Teed, Zachary, and Jia Deng. "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow." In ECCV, 2020.
- [5] Hataya, Ryuichiro, Jan Zdenek, Kazuki Yoshizoe, and Hideki Nakayama. "Faster autoaugment: Learning augmentation strategies using backpropagation." In ECCV, 2020.
- [6] Afrasiyabi, Arman, Jean-François Lalonde, and Christian Gagné. "Associative Alignment for Few-shot Image Classification." In ECCV, 2020.
- [7] Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis." In ECCV, 2020.
- [8] Li, Mengtian, Yu-Xiong Wang, and Deva Ramanan. "Towards Streaming Image Understanding." In ECCV, 2020.



朱翔昱

中国科学院自动化研究所副研究员。主要研究方向为计算机视觉和三维重建。

Email: xiangyu.zhu@nlpr.ia.ac.cn

顶会观察

NeurIPS 2020

北京大学 ZERO Lab 林宙辰 王奕森

神经网络信息处理系统大会 (Conference on Neural Information Processing Systems, NeurIPS) 是机器学习领域的顶级会议，是神经计算方面最好的会议之一，在中国计算机学会推荐国际学术会议中被评为人工智能领域的 A 类会议。在 Google Scholar 发布的 2020 年学术指标中，h5 指数高达 198，位于计算机领域的第 3 位、位列所有领域出版物的第 21 位。

今年第 34 届 NeurIPS 原定于 12 月 6 日到 12 日在加拿大温哥华举办。因为疫情原因，NeurIPS 2020 变成了一届纯线上会议。

一、NeurIPS 2020 的亮点

由于新冠疫情的持续蔓延，NeurIPS 2020 完全采取了虚拟线上会议的形式，让世界各地的参会者通过互联网参会交流。为了提升线上会议的参与体验，本次 NeurIPS 大会组委会专程邀请了两名在线体验主席，并在 poster 环节使用了名为 GatherTown 的交互网站，在模拟线下交流上十分逼真。除此之外，会议注册费用大幅下降。学生注册费用为 25 美元，其他参会者则为 100 美元，而去年价格分别为 420 和 750 美元。值得注意的是，所有人都可以在无需注册的情况下访问 Tutorial、Keynote、Oral 论文和 Spotlight 论文的宣讲内容^[1]。在会议结束后，NeurIPS 2020 预计还会将 Poster 的录像公开。

除了涵盖人工智能领域丰富的主题，NeurIPS 2020 还开始探讨机器学习的社会影响，例如投稿论文需要添加 Broad Impact、审稿流程中引入道德审查机制。不过仅有 290 篇投稿论文被审稿人提出了道德担忧，

主要集中在虚假人脸的生成、新数据集的隐私泄露风险等。同时，大会邀请报告讨论了如何避免机器学习系统的不良社会影响和关注数据标注员的职业发展问题。今年还特意举办了关于新冠病毒 COVID-19 的特别研讨会。这体现了人工智能社区和科技工作者的社会责任感。

二、论文录用情况

NeurIPS 2020 的投稿量再创新高，共有 9467 篇论文被提交，相比 2019 年增长了 40%。经过 7062 名审稿人的同行评议，最终有 1898 篇论文入选，录取率仅为 20.0%，低于去年的 21.6%，达到历史最低。其中 105 篇被录用为 Oral，280 篇被录用为 Spotlight，Oral 和 Spotlight 的录取率仅有 1.11% 和 2.96%。

在本次 NeurIPS 2020 中，谷歌位于所有高校/机构的榜首，入选论文高达 202 篇。麻省理工学院排名第二，共有 109 篇论文入选。斯坦福大学位居第三，共有 106 篇论文入选。国内高校/机构中，清华大学排名第一，共有 63 篇论文入选，排在国内外所有高校/机构的第 7 位；北京大学第二，共有 37 篇论文入选。从作者角度，有华人参与的论文共有 568 篇，占总论文数的 29.91%。其中人为第一作者的论文共有 321 篇，占华人参与论文数的 56.51%，占总论文数的 16.90%。华人在 NeurIPS 2020 中贡献了重要力量。

NeurIPS 作为机器学习领域的顶级会议，收录的论文包罗了人工智能领域的各种主题，包括深度学习及其应用、强化学习和规划、纯理论研究、概率方法、优化、机器学习和社会、神经科学和认知科学等方方面面。

三、邀请报告

今年的 NeurIPS 2020 共有七个邀请报告 (Invited Talk), 主题丰富, 涵盖了密码学、反馈控制理论、因果推断、生物学等多个学科, 反映了机器学习正与其他学科不断影响并相互交融; 同时探讨了人工智能大潮下的多个社会问题, 展现了科研人员的社会责任感。2012 年的图灵奖得主、MIT 的密码学教授 Shafi Goldwasser 展示了在密码学的意义下, 机器学习正面临的三大挑战。阿卜杜拉国王科技大学(KAUST)的教授 Jeff Shamma 重点介绍了一些反馈控制概念, 特别是鲁棒性、被动性、跟踪和稳定性, 因为它们与进化博弈论、无遗憾学习和多智能体学习中的特定问题相关。冷泉港实验室的神经科学家 Anthony M. Zado 教授报告了从生物学中借鉴本能瓶颈的想法。微软研究院的 Chris Bishop 回顾具有真实应用场景的微软研究院的研究项目, 分享了他对于 AI 未来发展趋势和现实世界影响的思考。卡内基梅隆大学的客座教授 Saiph Savage 指出人工智能产业在不断创造出新的工作的同时, 也给人们带来了许多冲击和挑战。美国国家科学院 Kavli Fellow 获得者、乔治亚理工学院计算机学院 Charles Isbell 教授提到社会需要采用系统的方法来研发鲁棒性的对人们生活具有重大影响的产品。

四、会议热点论文

本次会议涌现了许多优秀的工作, 它们具有非常高的学术价值与应用价值。NeurIPS 2020 的最佳论文奖由三篇论文摘得, 分别是自米兰理工大学的《No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium》、UC 伯克利的《Improved Guarantees and a Multiple-Descent Curve for Column Subset Selection and the Nyström Method》, 以及 OpenAI 的《Language Models are Few-Shot Learners》(即著名的 GPT-3)。

No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. 相关平衡(CE)易于计算, 并且可以获得比众所周知的纳什平衡要高得多的社会福利。在普通形式的游戏中, CE 的一个功能是可以简单且分散的算法将特定遗憾的概念

降到最低, 从而找到 CE。论文说明了在大型游戏(即扩展型(或树型)游戏)中收敛到 CE 的遗憾最小化算法的存在。该结果解决了博弈论、计算机科学和经济学界的一个长期存在的开放性问题, 并且可能对涉及调解员的游戏产生重大影响, 例如, 通过导航应用程序对有效的交通路线产生影响。

Improved Guarantees and a Multiple-Descent Curve for Column Subset Selection and the Nyström Method. 从大矩阵中选择列向量的一个小但有代表性的子集是一个困难的组合问题, 并且已知一种基于基数受限的行列式点过程的方法可以提供一种实用的近似解决方案。论文针对最佳可能的低秩近似, 推导了近似解的近似因子的新上限和下限, 该上限甚至可以捕获子集大小的多重下降行为。论文将分析进一步扩展, 获得了 Nyström 方法的理论保证。由于这些逼近技术已在机器学习中得到广泛应用, 因此这篇论文有望产生重大影响, 并为核方法、特征选择和神经网络的双下降行为等理论和现象提供新的见解。

Language Models are Few-Shot Learners. AI 大厂向来奉行大力出奇迹, OpenAI 的 GPT-3 工作更是将这个理念推到极致, 坚持 Bengio 200 年提出的简简单单的 Neural Language Model, 加上海量的数据和算力, 实现了前所未有的表示能力。论文向我们清晰地展示了, 只要算力足够, 神经网络还有相当大的空间可以开发, 而我们通常讨论的数据集和模型大小, 离达到神经网络的极限仍然很远。

此次 NeurIPS 2020 一共收录了 105 篇 Oral 论文和 280 篇 Spotlight 论文, 大会将它们分成了 35 个 Track, 包括: Representation / Relational, Language / Audio Applications, Reinforcement Learning, Clustering / Ranking, Dynamical Sys / Density / Sparsity, Deep Learning, Vision Applications, Deep Learning / Theory, Continual / Meta / Misc Learning, Social / Adversarial Learning, Graph / Meta Learning / Software, Graph / Relational / Theory, Unsupervised / Probabilistic.

Oral 论文仅仅是录用论文的冰山一角。其他录用论文中还有大量值得探索和讨论的内容。

除此之外，大会还举办了 30 个教程(Tutorials)和多个研讨会(Workshops)，包括离线强化学习、生物学与 AI、鲁棒机器学习、深度隐私模型等丰富的主题。这些教程和研讨会，为科研人员扩展视野和交流互动提供了便利的渠道和平台，利于机器学习领域的长期发展。

五、总结与展望

在 NeurIPS 2020 的投稿论文中，算法(占投稿总量的 29%)、深度学习(19%)及其应用(18%)、强化学习(9%)等依旧是机器学习和计算神经科学的热门主题。值得注意的是，探讨机器学习的社会层面问题的论文增长迅速，由去年的 2% 上升到今年的 5%。机器学习得到了越来越多的应用，如何避免模型可能带来的不良社会影响(公平、隐私等)成为不可忽视的问题。

在新冠病毒全球肆虐的大背景下，各类人工智能会议不得不转到线上举行。从最初 ICLR 2020 每篇论文分配独立的 Zoom 会议室，到 NeurIPS 2020 使用 GatherTown 举办 Poster Session，线上虚拟会议的体验持续提升。随着虚拟会议技术的持续提升，虚拟会议的体验有可能逼近线下会议，并且注册费低得多，有利于扩大参会人数，或许会成为未来科研交流的另一种重要形式。

致谢：本文由林宙辰、王奕森组稿，凌泽南、陈勇、王一飞、何翎申、李明杰、徐大鹏、孔浩、耿正阳、杨一博、沈铮阳、徐鑫参与写稿。关于 ZERO Lab 的更多信息见 <https://zero-lab-pku.github.io/>。如需了解更多关于 NeurIPS 2020 的论文介绍，见微信公众号链接 <https://mp.weixin.qq.com/s/BaUIZEuuTCjFGqq5sa3x A>。

责任编辑 王金甲

参考文献

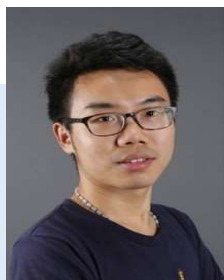
- [1] <https://proceedings.neurips.cc/paper/2020>



林宙辰

北京大学信息科学技术学院智能科学系教授，国家杰青，IAPR/IEEE Fellow。研究方向为机器学习、计算机视觉。担任中国图象图形学学会机器视觉专委会主任、中国自动化学会模式识别与机器智能专委会副主任、中国计算机学会计算机视觉专委会常务委员、中国人工智能学会模式识别专委会常务委员。

Email: zlin@pku.edu.cn



王奕森

北京大学信息科学技术学院智能科学系助理教授。主要研究方向为机器学习，特别是对抗学习、图学习。

Email: yisen.wang@pku.edu.cn