

Frontiers of Information Technology & Electronic Engineering
 www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
 E-mail: jzus@zju.edu.cn



综述:

设计智能研究综述：进展、问题和挑战*

汤永川^{††1,2,3}, 黄江杰^{1,2}, 姚梦婷^{1,2}, 魏嘉^{1,2}, 李薇^{1,2,3}, 何永兴^{1,2,3}, 李泽健^{1,3,4}

¹计算机科学与技术学院, 浙江大学, 杭州310027, 中国

²浙江省设计智能与数字创意研究重点实验室, 杭州310027, 中国

³之江实验室, 杭州310012, 中国

⁴阿里巴巴-浙江大学前沿技术联合研究中心, 杭州310027, 中国

[†]E-mail: yctang@zju.edu.cn

投稿2019-8-6; 录用2019-11-26; crosscheck 2019-12-10

摘要: 设计智能是人工智能重要分支, 专注于创意和设计中的智能模型和算法。在人工智能2.0背景下, 设计智能研究发展迅速。本文首先介绍设计智能研究背景, 提出设计智能研究的理论框架。从用户需求分析、创意激发、内容生成和设计评价4个维度, 详细综述设计智能研究进展和最新技术; 重点论述关于智能生成内容的模型和方法。最后, 提出未来设计智能研究中的开放问题和挑战。

关键词: 设计智能; 创造力; 用户画像; 创意激发; 智能生成内容; 计算美学

本文译自Tang YC, Huang JJ, Yao MT, et al., 2019. A review of design intelligence: progress, problems, and challenges. *Front Inform Technol Electron Eng*, 20(12):1595-1617.

<https://doi.org/10.1631/FITEE.1900398>

中图分类号: TP183

1 引言

创造力主要涉及产生创新创意想法的人类思维过程, 它是可衡量和可发展的(Kaufman and Sternberg, 2006)。在过去的三十年中, 计算和认知的角度逐渐成为了创造力、设计、创意问题的研究热点。在相关研究中, 计算创造力和设计计算是非常重要且一定程度上相互重合的两个方向。作为人工智能的分支, 计算创造力聚焦于创新创意过程的计算模型(Boden, 2009), 通过研究创造力的工作机制, 为人们带来高质量的创新作品。而设计计算主

要通过应用新的计算方法和技术来进行设计研究和实践。总体上, 两者都提出并产生与设计 and 创造力相关的计算系统、计算算法、计算模型和计算表示(de Silva Garza, 2019)。不同之处在于, 计算创造力侧重于通过模拟人类智能来增强机器创造力, 而设计计算则着眼于通过应用模型和算法来帮助设计师提出更好的想法。

在过去, 设计智能主要使用人工智能技术对设计领域知识进行建模, 从而解决设计问题, 并输出最优结果, 例如基于案例的推理、约束满足和进化算法等(Russell and Norvig, 2016)。在知识表示模型方面, 基于区分与目标相关、行为和结构特征的模型主要有FBS (Gero, 1990) 和SBF (Goel et al., 2009)。de Gómez Silva Garza and Maher (1999)介绍了GENCAD系统和基于案例的推理(CBR)的总体策略。总的来说, 过去的设计智能主要注重效率

[‡] 通讯作者

* 本文得到以下项目的资助: 科技创新2030——“新一代人工智能”重大项目 (No. 2018AAA0100703), 国家自然科学基金 ((Nos. 61773336 and 91748127), 中国工程院咨询研究项目(No. 2018-ZD-12-06), 浙江省重点研发计划 (No. 2019C03137) 和浙江大学—新加坡科技设计大学创新、设计、创业联盟 (黄廷芳慈善基金会赞助) 资助项目

ORCID: 汤永川, <https://orcid.org/0000-0002-0157-7771>

和功能解决方案，囿于硬件和软件条件的限制，对创造力的核心问题的研究并未取得太大进展。在人工智能2.0时代(Pan, 2017)，深度学习尤其是生成对抗网络(GAN)快速发展，计算机开始具有一种新的创造性能力。计算机在协作或独立生成许多作品（例如视觉艺术、音乐、视频、海报、网页、服装、椅子和汽车）的过程中，已经表现出与人类相似的惊人创造力。因此，本文提出了设计智能的概念，主要指解决设计和创意过程中的问题并生成创意解决方案（设计、内容和服务）的人工智能技术。de Silva Garza (2019)从计算角度阐述了创造力和设计之间的关系，由图1可以看出，创意设计位于创意和设计相交的地方，这一部分同样是设计智能研究关注的重点方向。

本文提出，在系统架构方面，设计智能主要包括一个数据模块和一个创意模块。从图2可以看出，数据模块是一个领域知识数据库，它代表相关设计和创意领域收集的信息，包括设计、风格、艺术、技术、商业和时尚等方面的数据。此外，创造性的专家知识也存储在数据库中，并作为创意模块提取数据的来源。创意模块具有四个主要组件，包括用户需求分析、创意激发、内容生成和设计评价。在人工智能的帮助下，这四个组件相互耦合，最终为消费者生成产品、内容和服务。创意模块是设计智能的核心，其中，用户需求分析部分通过AI技术捕捉用户需求，并科学地定义设计问题；创意激发部分通过AI技术为高质量设计提供了灵感、组件和要素；内容生成部分是设计智能的关键组成部分，主要使用生成模型来产生新想法并增强创造力，同时支持创造力的最终呈现和产品化；设计

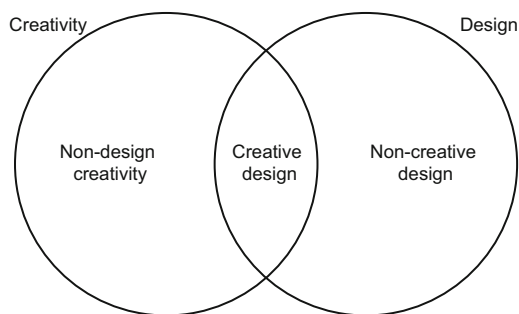


Fig. 1 创造力和设计之间的关系，图中相交的部分边界是模糊的，图片来自于de Silva Garza (2019)

Reprint from de Silva Garza (2019), Copyright 2019, with permission from Springer Nature

评价部分则通过AI技术从美学和功能等维度量化设计师产生的创意产品、内容和服务。虽然，本文借鉴了人工设计流程对设计智能进行解释，但是，设计智能各个组成部分之间的关系非常复杂。同时，设计智能可能具有与人类不同的认知或创意构思，这些将会对人类的创造力形成合理的补充和支持(Varshney et al., 2019)。

近年来，关于设计智能的研究取得了重大进展。Edelman et al. (1996)认为创造力是人类智能的基本特征，也是对人工智能的挑战。Boden (2009)提出了人工智能对创造力建模的三个过程：探索性创造、变革性创造和组合创造，此后的许多研究都深受影响和启发。de Silva Garza (2019)在设计活动中比较了设计计算和计算创造力，并完整描述了两者的理论基础和实际进展。同时，研究人员在设计和创造力领域开发了许多设计智能算法。Gatys et al. (2016a)介绍了一种艺术风格的神经算法，该算法可以分离并重新组合自然图像的内容和风格。Zhang et al. (2017)引入堆叠式生成对抗网络(stack-GAN)，基于文字描述生成逼真的图像。Isola et al. (2017)研究了有条件的生成对抗网络，以解决图像到图像的翻译问题，例如将手提包的草图转换为精细的彩色图像。这些算法通过在相应领域中学习训练数据的规则和风格，来支持大规模的设计生产。Wu et al. (2016)提出了3D生成对抗网络(3D-GAN)，可以根据特定的对象类型、视点和颜色生成3D椅子、桌子和汽车。但这种算法的创造能力有限，因为它们难以创建与训练数据不同的创新设计。为了解决这一问题，Elgammal et al. (2017)提出了一种用于产生艺术创新的生成对抗网络，通过最大化与既定风格的偏差和最小化与艺术分布的偏差来产生创意艺术。Yan et al. (2019)综述

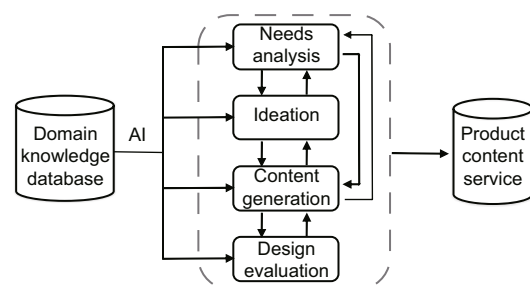


Fig. 2 设计智能的理论模型，四个主要的数据驱动分析组件位于中间，箭头指示组件之间的主要信息流

了神经网络在现代设计中的应用进展，有助于研究人员和设计师更好地理解由神经网络驱动的设计智能。总而言之，在人工智能技术的大力推动下，设计智能已成为推动设计和创意领域发展的强大力量。然而，关于设计智能的研究成果尚未得到很好的梳理和总结。

2 基于AI的用户需求分析

用户需求分析是设计人员用来分析潜在用户需求和偏好的工具。成功的设计和创意作品基于对用户需求和要求的理解(Maguire and Bevan, 2002)。对于不同的用户研究方法，该过程通常涉及四个阶段：信息收集、用户需求识别、设想和评估以及需求规范(图3)。在此过程中，设计人员通过移情和用户的研究工作或收集的数据不断明确设计的方向和定义。宣言、叙述、场景和用户画像是四种广泛应用于用户需求分析的工具(Lowdermilk, 2013)。宣言是设计的总体目的和愿景；叙述是宣言的声明性声明，说明如何使用设计；场景是用户使用或体验设计的情境描述；最后，用户画像是虚拟的角色，代表了潜在的用户人群。设计文献对这四个方面进行了广泛的研究，但用户画像是用户需求分析的核心。此外，智能方法的介入主要应用于用户画像，而其它三个方面通常由设计人员完成。因此，在本章中，我们主要回顾用户需求分析中智能方法在用户画像中的应用。

用户画像被定义为代表潜在用户或用户群体的虚拟角色。它通常代表核心目标用户，或者系统的理想用户(Cooper and Reimann, 2003)。要创建用户画像，设计师应调查潜在用户的偏好，收集他们在其他产品上的经验并总结典型的个人信息（例如年龄和职业）。据此，用户画像不仅可以提供目标用户的清晰意象，更有助于评估设计方案是否满足用

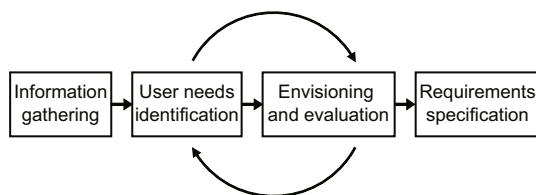


Fig. 3 用户需求分析的基本过程，图片来自Maguire and Bevan (2002)

Reprint from Maguire and Bevan (2002), Copyright 2002, with permission from Springer Nature

户需求。此外，创建角色的过程也增强了设计人员对目标用户群体的同理心(Miaskiewicz and Kozar, 2011)。Cooper (1999)引入用户画像作为一种目标导向的设计技术。研究人员认为，使用用户画像有利于更好地设计，其支持设计的所有组成部分(Grudin and Pruitt, 2002; Miaskiewicz and Kozar, 2011; Nielsen et al., 2015)并可改善设计质量(Grudin and Pruitt, 2002; Cooper et al., 2003; Pruitt and Adlin, 2006; Salminen et al., 2019)

用户画像通常分为三类：设计人员创建的用户画像、通过智能方法生成的用户画像以及根据生成数据构建的用户画像。在传统的设计流程中，用户画像都是由设计师完成的。设计师经常收集少量的用户数据，并通过统计方法或定性技术（如人种学研究）进行分析。该类用户画像是在分析过程中形成的，并通过访谈、观察、焦点小组和现场研究等方法进一步完善(Matthews et al., 2012)。近年来，不断增长的数据为许多设计任务提供了丰富的背景信息，而传感器的发展催生了更多的收集用户行为数据的方法。然而，对人类设计师来说，采用传统方式分析海量数据或感官数据是不现实的。因此，通过使用智能方法来自动生成用户画像来进行用户需求分析成为了最可行和高效的方式。

学术界已提出了几种自动用户画像生成方法。当人们在社交网络上显示其偏好时，可以利用社交媒体数据进行用户画像。如Kwak et al. (2017)开发了一种使用YouTube社交媒体数据实时自动生成用户画像的方法。其它的研究如(Salminen et al., 2017; Jansen et al., 2017)从在线社交媒体和内容平台中收集数据，并使用数据分析工具进行用户画像研究，即从数据中将用户需求的相关内容抽象化。行为数据同样反映出用户的基本偏好。Salminen et al. (2018b, 2019)对参与者进行案例研究并分析了行为数据，以探索自动用户画像生成中可能存在的优势和挑战。通常，分析在线数据从而自动生成用户画像的系统和方法具有广泛应用价值，对于使用在线内容和用户数据进行设计和内容创新的人员而言尤其如此。

近年来，生成模型(Goodfellow et al., 2014)在内容生产方面取得了巨大进展。这是一种生成新样本的智能方法，该样本将受到给定数据的相同分

布的影响, 有关详细信息将在第4部分中详细介绍。生成模型可用于生成潜在的用户画像。IRGAN (Wang et al., 2017)和RecGAN (Bharadhwaj et al., 2018) 是用于生成相关用户信息(例如文档或用户行为)的模型。此外, Perera and Zimmermann (2019)提出了一种新颖的GAN模型(CnGAN), 该模型可以学习社交网络单独部分中用户的复杂偏好, 并生成跨网络用户偏好。这些方法提供相关的生成数据, 并在构建用户配置文件时提供辅助数据。就我们所知, 这些方法在设计任务的用户画像生成中仍未得到应用。但是, 将这些技术从文献转移到设计领域仅一步之遥。值得注意的是, 此处生成模型的应用不同于之前介绍的自动用户画像生成, 自动化用户画像创建是从数据中提取画像, 而生成模型则是提供可能的新用户数据。

用户画像也存在一些需要使用者注意的缺点, 例如时间和精力投入大、构建成本不菲、可信度缺乏、易受思维惯性和个人偏见影响、前后矛盾等等(Salminen et al., 2018a)。但是, 不可否认, 用户画像仍然是分析用户需求的最有价值和最有效方式。同时, 由于具备解决学者和从业人员所提出的问题的能力, 智能化的用户画像生成方法受到广泛关注。由于数据和评估方面的挑战, 使用角色扮演的用户需求分析仍具有探索的空间。一方面, 基于大量行为数据自动创建用户画像仍然是一个未解决的研究问题(McGinn and Kotamraju, 2008)。设计智能领域中的数据源复杂多样, 设计人员需要共同分析文本、视觉、听觉、体态, 甚至是VR或AR数据, 这需要采用先进的跨媒体智能方法(Yang et al., 2008)提取潜在用户画像。最后, 缺乏用户画像评估方法也是一个主要问题。评估用户画像有效性的常用方法是案例研究(Salminen et al., 2018b, 2019), 如何有效形式化评价用户画像并进行计算将是未来有价值的探索领域。

3 基于AI的创意激发

创意激发用于在设计流程的早期发掘潜在创意的集体活动(Hartson and Pyla, 2012)。这一过程为最后的设计提供了早期创意、基本构件和材料, 因此要形成一个高质量的设计, 设计参与者先进行有效的创意激发是非常必要的。创意激发的方法可以

分为基于刺激和基于交互的创意激发, 而人工智能则可以从不同的方面提升创意激发方法。在本章我们会回顾大量使用人工智能的创意激发方法。

设计刺激是设计流程中的灵感来源之一, 而基于刺激的创意激发方法为设计师提供了大量外部信息, 以启发其产生新创意。设计流程中的刺激包括各种媒介, 例如文字、图片、本体等等。刺激和潜在解决方案的关联关系也多种多样, 刺激本身可能就是可行的结局方案(Hao et al., 2019), 或者和解决方案形成某种类比关系(Han et al., 2018), 或者是具有仿生启发关系(Vandevenne et al., 2015)。可见, 基于刺激的创意激发方法旨在寻找有启发性的刺激给设计师作为辅助信息。

智能方法在当中可以辅助检索现有的刺激。Hao et al. (2019) 提出了一种方法, 用以检索既可行又具有创新性的案例。其使用了进化计算的方法来进行启发性词语的检索, 当中的词语来源于1976年到2014年50万个授权专利和工业产品。作者使用word2vec方法(Mikolov et al., 2013)来把这些词语嵌入到高维空间当中, 而且用可行性和创新性两个维度来度量每个词语的启发性。随后, 他们使用进化计算的策略来选择启发性最高的词语。最后, 这些词语作为言语刺激信号展示给设计师。Vandevenne et al. (2015) 介绍了另一种刺激检索方法。它使用行为语言模型来表示纯文本中的自然和人造实体, 从而方便检索。其中的生物策略案例使用K近邻算法分类, 从而形成一个分类目录。在给定一个关键词查询时, 这个分类目录可以让计算机迅速定位相关的生物启发策略。

类比关系是一种关联多个概念的典型思维范式, 其广泛应用在创意激发的过程中(Gilon et al., 2018; Jia et al., 2020)。“检索器”(“the Retriever”) (Han et al., 2018)是一个专注于展示图片实例和词汇之间类比关系的检索工具。它建立在ConceptNet(Liu and Singh, 2004) (一个表示了常识知识的大规模语义网络)丰富的实体内容上。为了检索可能具有启发信息的图片和言词刺激, 它使用了等比类比关系 $A : B :: C : X$ 来对实体关联。其中A和B(或者C和X)之间的关系被指定为ConceptNet中的16种本体关系之一, 从而形成了A之于B类似于C之于X的类比关联。给定一

个输入的词语 C 和 A 和 B 之间的一个关系,该系统搜索可能形成类比的视觉和文本信息 X ,并展示给用户。其他典型的基于类比的工具包括Idea Inspire(Chakrabarti et al., 2017)和一个函数型类比关系搜索系统Fu et al. (2015)。

除了搜索和检索之外,图片生成也是一项前沿人工智能技术。Chen et al. (2019)提供了一种视觉刺激生成模型。该模型在两个图片数据集上进行训练,每个数据集是一个概念。该模型通过这样的方式试图将两个概念组合成一个新概念。训练得到的新模型可以捕捉两个概念的视觉元素。实验中发现,生成的图片具有一定的多样性和新颖性。训练好的生成模型可以给出大量的新样本,因此这些视觉刺激从种类和数量两个维度上极大促进了创意激发的过程。下一章我们会着重介绍图像生成技术。总的来说,智能化的方法可以迅速而方便地建模和检索现有的刺激,甚至生成新的刺激。

基于交互的创意激发方法可以在辅助设计时从小组讨论中得到灵感。典型的方法包括头脑风暴(Dugosh et al., 2000)及其变种(Faste et al., 2013)。在头脑风暴的过程中,设计师集中产出原始创意,自由讨论并将其组合为新的想法。(Strohmann et al., 2017)提出了一个虚拟智能助手,用于主持一场头脑风暴。这个助手就像聊天机器人,可以按照一定的流程组织在线头脑风暴活动,并且在必要的时候用自然语言干预讨论。当参与者都在自由讨论时,它会鼓励参与者畅所欲言,或者提醒参与者不要违背头脑风暴的规则。这样的虚拟助手可以替代人类完成主持工作,并且允许设计师团队随时开展在线头脑风暴。这是一种前沿自然语言理解和处理技术在创意激发过程上的应用。

智能化方法在基于刺激的创意激发中起到了重要作用,但是在基于交互的方法中则应用较少,这为之后的探索预留了大量空间。Chen et al. (2019)首次把生成技术应用在创意激发上,但其仅给出了随机生成样本,并不具备可控性。更多前沿图像转换技术(例如Wang TC et al. (2018)和Liu et al. (2019))都可运用在相关问题中,从而为辅助设计师提供更多的视觉刺激,提高创意激发的效率。对于基于交互的创意激发,线下头脑风暴中的智能主持人会是未来的探索方向,它可以辅助总结已有讨

论结果或者进行会话干预。另外,大部分前沿智能方法都是基于统计学习,旨在拟合现有数据中的分布,在创新性上略有不足。因此,尝试利用智能方法获取更多的现有辅助信息同样是未来发展方向。

基于AI的创意激发方法研究常涉及数个典型数据集。美国专利数据集包含1976年至今的九百万个专利的全文数据,便于用户检索现有产品设计。Quirky(quirky.com,最后访问于2019年12月24日)是一个拥有两百多万产品案例的企业,他们为发明者提供设计早期帮助。InnoCentive(innocentive.com,最后访问于2019年12月24日)提供了商业、社会、政策、科学和技术等方面的四万余个案例及其对应的解决方案。此外,诸如WordNet(Miller, 1995)、Cyc(Niles and Pease, 2001)和ConceptNet(Liu and Singh, 2004)的本体数据集也会在相关工作中使用。这些数据集为刺激检索和类比关系推理提供了常识和概念定义。

创意激发中有多个待解决的挑战,设计固化是其中较为重要的一个(Jansson and Smith, 1991)。设计固化是指设计师陷入某一类想法中而无法发散思维,在这种情况下,创意激发效果相当有限。减轻设计固化的主要方法是提升激发信息的多样性和创新性。具体来说,在基于刺激的创意激发过程中,丰富多样的刺激能够帮助设计师跳离固化的陷阱(McCaffrey and Krishnamurty, 2015; Hao et al., 2019)。此外,功能性固化、类比关系的忽略、现有假设的忽视和狭隘的动词关联都是创意激发中的主要挑战(McCaffrey and Krishnamurty, 2015)。在另一方面,现有研究一般从数量(Shah et al., 2000)、质量(Shah et al., 2000)、创新性(Peeters et al., 2010)和多样性(Nelson et al., 2009)等方面来衡量创意激发方法的优劣。因此,在对智能化方法在创意激发中的应用进行探索的过程中,可以考虑从以上四方面来着手克服上述多个挑战。

4 基于AI的内容生成

创意内容生成是设计智能研究和应用的重要领域。随着人工智能(AI)的发展,IGC(Intelligence-Generated Content)模式开始成为一种新型内容生产模式。机器人一样开始从事图像、

音频、视频、文本等内容的创造和生产。和基于人力的内容生产方式相比，AI产生的内容在某些领域达到相当甚至高于人类的生产水平。此外，生产工具由人脑转换为算法，一方面降低了劳动力的成本，另一方面也使生产效率得到了成倍增长。值得注意的是，创意内容的产生方式多种多样，无固定范式可言。以当前AI的发展水平，机器智能并不能完全取代人类在创意内容生产中的角色。与早期计算机辅助也不同，AI也不是完全的辅助设计工具，它具备一定的创意创造能力。因此，人类和机器之间的协作变得更加深入和智能化。人类在内容生产中的职责发生变化的同时，人类和AI在设计工作中的分工界线也在逐渐模糊。

在这一节，我们介绍了智能内容生成所涉及的技术，回顾了文本-图像生成、图像-图像翻译、图像增强和内容-风格迁移的各类模型。此外，我们也讨论了相关领域中的潜在研究方向。

4.1 生成式模型

对于算法而言，凭空臆造数据是行不通的。机器学习算法需要理解训练数据。为了实现这一目的，大部分算法通过探索数据的本质属性，找到能够表示它的最佳分布。这类方法被称作生成式方法。通过这种训练获取到的模型称作生成式模型。利用学

习到的生成模型，可以得到并不在数据集中但满足相同分布的样本。这是内容生成和创新的关键。

利用深度神经网络，针对复杂数据的生成模型发展迅速，例如变分自动编码器(VAE) (Larsen et al., 2016) 以及生成对抗网络 (GAN) (Goodfellow et al., 2014)。

GAN是生成模型的一种新型框架。与其他生成模型相比，GAN具备更逼真的生成效果。近年来，GAN引发了社区广泛的关注，并迅速发展起来。在接下来的部分，我们会简要介绍GAN及其相关变体的架构和原理。详细的GAN发展史和综述超出了本文的范畴，如有兴趣推荐阅读综述文章(Hong et al., 2019)。

4.1.1 生成对抗网络

生成对抗网络(GANs) (Goodfellow et al., 2014) 包含两个独立的网络。一个网络是生成器 G ，其将收到的一个随机噪声向量 z 作为输入，输出一个生成样本 $G(z)$ 。另一个网络是判别器 D ，其将收到真实样本 x 或生成样本 $G(z)$ 作为输入，输出一个概率值来评估输入的样本是真实样本的可能性。如图4a所示。在这篇文章中，作者给出了两个玩家最小最大

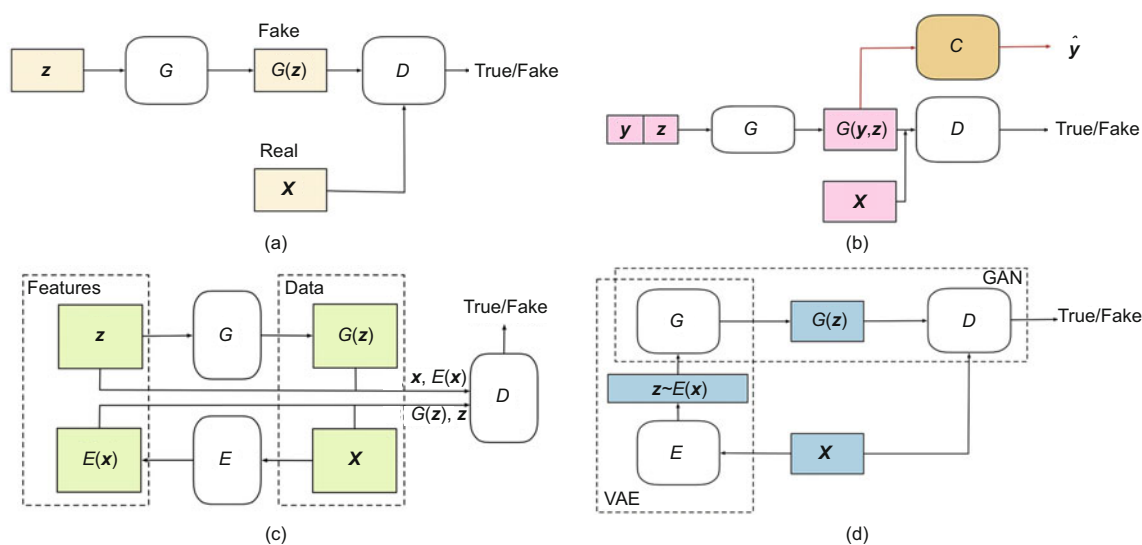


Fig. 4 各种GANs的架构: (a) 原始GAN的通用架构, 生成器 G 输入一个噪声向量 z 输出一个生成样本 $G(z)$, 判别器 D 输入一个真实/生成的样本并预测该样本是否为真; (b) 带条件输入或辅助分类器的GAN的架构, y 是条件输入标签, C 是用来预测标签 \hat{y} 属于生成样本 $G(y, z)$ 的分类器; (c) 带有自动编码器结构的GAN 架构, 编码器 E 输入一个真实样本 x , 输出一个特征向量 $E(x)$; (d) 带变分自动编码器的GAN 的架构

博弈的准则函数 $V(G, D)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim P_{\text{data}}} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim P_{\mathbf{z}}} [\log(1 - D(G(\mathbf{z})))] \quad (1)$$

事实证明, 在温和的假设下, 这种零和博弈的解决方案可以达到纳什均衡。换句话说, 对于生成器网络输出的假样本, 判别器无法区分它是来自真实分布还是生成分布。

然而, 第一篇GAN文章的发表在当时并没有引起很大的反响。在这篇文章中, 作者使用全连接层作为构建块, 同时使用maxout作为 G 和 D 的激活函数。作者只在简单的数据集, 例如MNIST (LeCun et al., 1998)、CIFAR10 (Krizhevsky and Hinton, 2009)等低分辨率的图像数据上验证了算法的有效性。直到DCGAN (Radford et al., 2016)的发表, 在全卷积层和反卷积层的帮助下, GAN才真正开始显现出相比于其他生成模型的优势, 即: 即使是非常复杂的数据也能有非常吸引人的生成效果。虽然GAN依然经历着训练不稳定和模式坍塌的问题, 但是鉴于它引人入胜的生成效果, 从2014年到本文撰写阶段的短短5年内, GAN的发展非常迅速。研究者在目标函数(Arjovsky et al., 2017)、优化策略(Gulrajani et al., 2017)、算法稳

定性(Miyato et al., 2018)、训练策略(Brock et al., 2018)等问题上不断改进。GAN的生成效果越来越逼真, 如图5所示, 分别是分辨率在 32×32 、 128×128 、 512×512 和 1024×1024 下的生成结果。强化后的GAN已经变成了一个基础工具, 运用于解决各种各样的生成问题, 广泛的包含于计算机视觉(CV)和自然语言处理(NLP)的框架中。

4.1.2 带条件的GAN

原始的GAN是完全无监督的生成。输出完全依赖于输入生成器 G 中的随机噪声, 而没有其他约束。在实际的生成任务中, 对输出结果进行约束是很自然的要求。条件GAN (Mirza and Osindero, 2014)应运而生。如图4b, 在随机噪声 \mathbf{z} 的基础上多加一个条件 \mathbf{y} , 二者共同作为生成器 G 的输入。这个条件 \mathbf{y} 可以是任意内容, 例如一张图片的类别信息、某个对象的属性(Mirza and Osindero, 2014)或者是想要生成的图片所对应的文字描述(Reed et al., 2016a)。有了条件 \mathbf{y} 的存在, 输出结果将同时被噪声和条件信息共同控制, 即可实现可控制的生成。

除了给 G 喂入条件信息, 还可以借鉴半监督学习的思想, 加入更多的边信息作为监督信号。如图4b中的 C , 它是一个辅助的分类器(Odena et al., 2017)。给判别器 D 增加一个辅助分类器, 有利于



Fig. 5 GANs 发展史 (2014年–2019年): (a)GAN (Goodfellow et al., 2014); (b) DCGAN (Radford et al., 2016); (c)BigGAN (Brock et al., 2018); (d)StyleGAN (Karras et al., 2019), 分辨率分别为 32×32 、 128×128 、 512×512 和 1024×1024 。除了分辨率的提升, 生成图像的细节也愈加逼真

生成更锐利的样本，同时可以缓解模式坍塌问题。这个分类器可以是预训练的模型，也可以和判别器 D 共享大部分网络参数同时进行训练。

4.1.3 带编码器的GAN

GAN只能做单向的生成，即输入特征向量 z ，输出生成样本 $G(z)$ 。单独的GAN缺乏从样本空间反向映射回特征空间的能力。在Donahue et al. (2016)和Dumoulin and Visin (2016)中，通过编码器 E 和生成器 G 配合，实现了数据空间和特征空间的双向映射，如图4c所示。编码器 E 输入真实样本，输出特征向量。生成器 G 输入特征向量，输出生成样本。判别器接收成对的特征和样本，判别他们来自真实数据还是生成数据。通过加入一个编码器，GAN不仅能得到一个生成高品质样本的生成模型，还具备了有效的推理机制，学习到了一个有意义的特征表示。

在Donahue et al. (2016)和Dumoulin and Visin (2016)的模型中， E 和 G 构成对称关系，但没有进行联通和协作。进一步的，在以上两个工作的基础上，Larsen et al. (2016)提出将另一个著名的生成模型——变分自动编码器 (VAE(Kingma and Welling, 2013)和GAN结合起来。如图4d所示，VAE的部分提供了特征空间的先验，作为GAN的输入。这样的结合，一方面保留了GAN强大的生成能力，避免了VAE生成模糊的弊端。另一方面，吸收VAE可以更大程度上保证生成多样性的特点，缓解了单纯使用GAN时让人头疼的模式坍塌问题。将编码器或VAE与GAN相结合和方法本身对于生成效果的提升是值得肯定的。但更重要的是，这一结合开辟了对特征空间进行解构，寻找更有价值的特征表示的新思路。通过对特征空间上的特征的控制，可以更有效的对生成样本的属性进行分解和表示，这一思路为之后的工作奠定了基础。具体的应用我们将在下一节中详细介绍。

一个训练良好的生成模型依赖于包含了大量样本的数据集。其中有些数据集只包含一类特定生成任务的图像，例如MNISTMNIST (LeCun et al., 1998)、CelebA (Liu et al., 2015)、flower (Nilsback and Zisserman, 2008)、chairs (Aubry et al., 2014)和WikiArt (Saleh and Elgammal, 2015)。

而多类别数据集对训练模型的鲁棒性提出了更高的要求，这类数据集典型的代表是：CIFAR10 (Krizhevsky and Hinton, 2009)、LSUN (Yu et al., 2015)以及ImageNet (Deng et al., 2009)。

4.2 智能内容生成方法

在创意内容的生成上，设计智能产生了诸多具有亮点的应用。在GAN的帮助下，机器已经开始有效地激发人类的设计想法并帮助人类付诸实践。在一些专业领域的创意内容生成中，机器已经达到甚至超过了人类设计师。在本节中，我们将回顾四种类型的生成方法：文本到图像的合成、图像到图像的转换、图像增强和内容风格迁移。这些方法为智能内容生成提供了基础。

4.2.1 文本到图像的生成

当使用诸如GAN的生成模型来实现图像生成时，通常需要控制生成图像的内容。尽管使用有条件的GAN模型 (如cGAN (Mirza and Osindero, 2014))，可以在一定程度上控制生成图像的内容，但条件GAN只能用来生成指定类别的图像。一种更直观的方式是，通过提供文本描述来生成相应的图像内容。文本到图像的生成是从单纯的文本描述生成逼真图像的核心算法 (图6)。在设计智能领域，文本到图像可以作为一种完整的智能创意内容生产方式应用到设计过程中。它还可以作为一个独立的创作模块应用于人机协同创作中，用以拓展人类设计师的创作能力，提高最终内容生成的质量。

GAN-INT-CLS (Reed et al., 2016a)是首次尝试使用GAN从文本描述生成图像的方法。这种方法总体上和cGAN相似，唯一的区别是使用文本描述嵌入，而不是原始cGAN中的类标记或属性。同时，

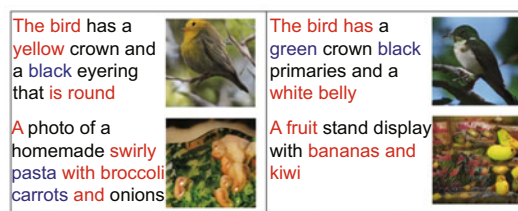


Fig. 6 文本到图像的生成实例，所有图片均使用GAN从文字描述生成而出，图片来自Xu et al. (2018)

Reprint from Xu et al. (2018), Copyright 2018, with permission from IEEE

判别器的设计也进行了改进。在原始的GAN中,取真实样本为True,生成的样本为False。GAN-INT-CLS中有四组匹配关系,即{真实图像、正确文本}、{真实图像、错误文本}、{假图像、正确文本}和{假图像、错误文本}。判别器 D 需要识别{真实图像,正确文本}为真,其他为假。对于 G ,它需要使{假图像,正确文本}接近{真实图像,正确文本}来愚弄 D 。

后续的研究在此基础上不断完善。例如, Dash et al. (2017)将GAN-INT-CLS与AC-GAN (Odena et al., 2017)相结合,在辅助分类器的帮助下,生成具有更优质的图像(Salimans et al., 2016)。GAWWN (Reed et al., 2016b)通过添加对图像中对象空间关系的约束,使文本描述与对象的边界框位置相对应,使生成更加符合描述要求。此外, StackGAN (Zhang et al., 2017)、StackGAN++ (Zhang et al., 2017) 等系列文章使用了多个生成器和判别器的组合。它们的生成过程是由粗到细,即从低分辨率到高分辨率,从而逐步生成文本中描述的图像。AttnGAN (Xu et al., 2018)利用图像和文本特征的注意力机制,进一步扩展了StackGAN++的工作。在这一研究工作中,将每句话嵌入到全局语句向量中,将句子中的每个词嵌入到词向量中,利用全局语句向量生成第一阶段的低分辨率图像,第

二阶段以第一阶段生成的图像为基础,将词向量作为关注层的约束进行组合,以产生更好的结果。

4.2.2 图像到图像的翻译

图像到图像的翻译也是一种智能内容生成的方法,并积极赋能于各种创造性应用。它不仅可以帮助专业设计师生成内容,还可以帮助非专业设计者提升创意生成水平(图7)。

1. 有监督的图像翻译

许多图像处理问题可以归结为“翻译”问题,即将输入图像转换为相应的输出图像。Isola et al. (2017)首先引入“图像翻译”的概念,并将之定义为“在给定足够训练数据的情况下,将场景的一种可能表示转换成另一种表示的问题”。他们构建了一个基于条件GAN的通用框架,不仅学习从输入图像到输出图像的映射,而且学习损失函数来训练这种映射,该框架已经成为有监督的图像到图像翻译的基础。如4.1.2小节所述,条件GAN的目标可以表示为:

$$L_{cGAN}(G, D) = E_{\mathbf{x}, \mathbf{y}} [\log D(\mathbf{x}, \mathbf{y})] + E_{\mathbf{z}, \mathbf{y}} [1 - \log D(G(\mathbf{z}, \mathbf{y}), \mathbf{y})]. \quad (2)$$

在生成器与判别器对抗训练的基础上, Isola et al. (2017)进一步增加了对生成器的 ℓ_1 约束,

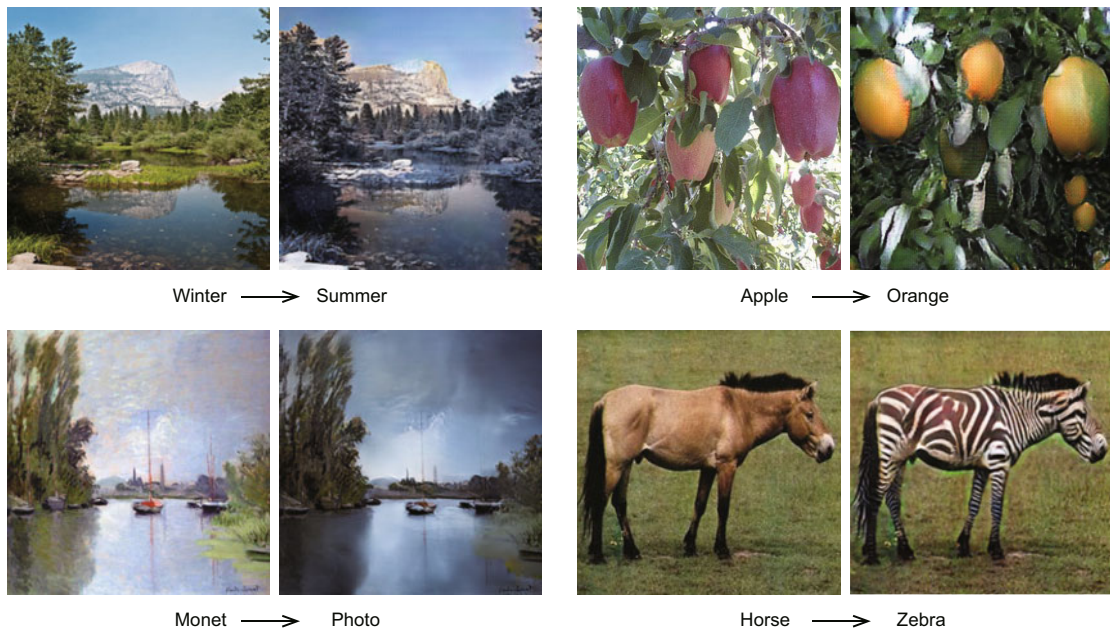


Fig. 7 图像到图像的翻译实例

要求生成器的输出尽可能接近真实图像。将伴随条件输入的数据输入到生成器以输出生成的样本,生成的图像和相应的条件都提供给判别器。此外,Isola et al. (2017)还提出了一种新的马尔可夫判别器,称为PatchGAN。PatchGAN 判别器在补丁的结构上进行了改进,其判别器试图判别图像中的每个 $N \times N$ 块补丁的真假。Pix2pix Isola et al. (2017)的一般原理如上所述。值得注意的是,它的效果令人惊叹。此外,这一通用解决方案可适用于各种任务,例如照片彩色化、将卫星图像转换为地图、将同一图像的白天场景转换为夜间场景以及将草图转换为渲染图片。后续研究的重点是获得更高清晰度(Wang TC et al., 2018; Park et al., 2019)。这种图像到图像的翻译方法还被应用于视频数据(Huang et al., 2017; Chan et al., 2018; Wang TC et al., 2018),同时,一些研究者正努力解决更具体的子域问题(You et al., 2019)。

2. 无监督的图像翻译

虽然Pix2pix的效果令人惊叹,但它需要配对的训练数据。例如,对于冬天到夏天的应用,我们训练的时候就需要提供成对的数据:分别在冬季和夏季拍摄的两张场景完全相同的照片。然而在实际任务中,配对数据很难获得。因此需要一种新的学习方法,能够在没有配对示例的情况下学习将图像从源域 X 转换到目标域 Y 。Zhu et al. (2017)和Yi et al. (2017)解决了这一非监督域自适应问题。具体来说,这个模型具有两个生成器和两个判别器。生成器 G 将 X 字段的图像转换为 Y 域图像,而另一个生成器 F 将 Y 域图像转换为 X 域图像。两个判别器 D_x 和 D_y 试图在两个域中区分真假图片。这里的假图片是指从另一个域名的真实照片转换而来的图片。除了两个方向的对抗之外,还增加了输出和输入的 ℓ_1 距离作为附加约束,即:

$$L_{\text{cyc}}(G, F) = \mathbb{E}_{\mathbf{x} \sim P_{\text{data}}(\mathbf{x})} [\|F(G(\mathbf{x})) - \mathbf{x}\|_1] + \mathbb{E}_{\mathbf{y} \sim P_{\text{data}}(\mathbf{y})} [\|G(F(\mathbf{y})) - \mathbf{y}\|_1]. \quad (3)$$

3. 图像翻译创新应用

基于上述监督和非监督的图像到图像翻译框架,衍生出了许多有趣的图像翻译应用。这些应用可以有效地为智能内容生成提供服务。着色(Zhang et al., 2016)是一个典型的应用案例,在给

定灰度照片作为输入的情况下,解决了照片的彩色着色问题。另一个应用案例是面部老化(Yang et al., 2018; Tang et al., 2018),这是一个从美学角度渲染给定面部图像以呈现衰老效果的过程。还有一个案例是会说话的头像(Zakharov et al., 2019)或换脸视频,这些方法致力于将一个人的动作视频移到另一个人的脸上。即使使用静态肖像照片或油画,也可以生成动态图像或连贯的视频。这些技术现在正应用在影视制作、娱乐应用程序、设计辅助工具等场景中。简言之,所有按照一定的约束输入图片和输出相应图片的过程都可以看作是图像到图像的翻译。

4.2.3 图像增强

设计智能领域的图像增强方法主要用于创作数字文化内容,满足当前数字文化内容创新的需要。

1. 图像修复

修复是将丢失或不需要的图像部分恢复的过程。对于一个理想的修复图像来说,整副图像的各个部分都是自然过渡的,欣赏者甚至无法区分哪些部分是修复过的。图像修复是改善人工智能生成内容的重要方法之一。当前,大量不断增长的文学艺术作品需要修复,图像修复的发展主要分为四个阶段:首先是基于扩散的修复,其次是基于补丁的方法,然后是基于深度学习的逐像素生成,最后是基于修复的图像处理(图8)。

基于扩散的方法(Bertalmio et al., 2000; Ballester et al., 2001)通过等光线方向场生成缺失的孔洞。这些工作只适用于小孔或线条。之后,图像修复将重点放在基于补丁修复的方法上。(Efros and Freeman, 2001; He and Sun, 2014)根据局部图像特征从补丁数据集中复制补丁,从而在单次扫描中填补缝隙。这种方法通常可以很好地进行纹理修复,但对于真实场景的修复效果不是很好。



Fig. 8 图像修复的视觉效果,白色的部分是需要修复的部分

基于深度学习的方法直接使用在大型数据集上训练的网络来推断图像已知部分中的缺失片段。Pathak et al. (2016)使用自动编码器架构来解决此问题。内容图像由编码器编码，并且解码器模块将代码解码到图像中的缺失区域。除了像素级重建损失外，还引入了对抗性损失来提高修复质量。随后，Iizuka et al. (2017)使用全局和局部内容判别器改进了Pathak et al. (2016)提出的框架，并在网络中引入了空洞卷积，以扩大接受范围。Liu et al. (2018)使用部分卷积代替传统卷积，效果更好。Yu et al. (2018a)将修复拆分为一个从粗到细的任务。将内容感知添加到精细的网络中，以学习从已知图像背景中提取特征。Nazeri et al. (2019)将修复分为两个阶段，即边缘修复和图像修复。

最后一个重要方向是基于修复的图像编辑。Yu et al. (2018b)使用门控卷积和用户指导的草图来完成图像修复，系统在用户输入草图的指导下生成缺失的部分。Jo and Park (2019)扩展了颜色提示，用户不仅可以控制生成的形状，还可以控制颜色。

2. 超分辨率

单图像超分辨率(SISR)是智能内容生成过程中一个非常普遍的需求。随着生成内容的不断细化，相对于拍摄或创建新素材，调整分辨率以适应AI的需求是更有效的选择。其中最简单的是基于插值的方法。最常用的一种是双三次插值法(Keys, 1981)，它速度快、简单明了，但存在精度不足和模糊的问题，如图9b。从低分辨率图像变换到高分辨率图像，需要在原始像素之间插入新的点，因此，难免会导致图像过于平滑和视觉模糊。生成模型可以很好地解决这个问题，其一大特点是能够生成与原始数据分布相匹配的全新数据。该特性不仅保证了在SISR任务中新建像素之间的自然连接，而且增加

了新的像素信息，以此使生成的图像不会过于平滑。对于低分辨率(LR)输入图像，可以有多个具有不同细节的高分辨率(HR)图像。同时，生成模型可以生成具有多样性的数据，很大程度丰富了生成的细节信息。总体而言，生成模型具有很强的生成能力，并非常适合于处理SISR问题。

在所有基于深度学习的方法中，SRCNN(Yoon et al., 2015)是一个重要的基准。该方法简单明了，以高分辨率图像作为训练集，首先通过下采样降低分辨率，并获得相应的低分辨率图像。然后，将低分辨率图像发送到生成器网络，对其进行训练，以输出生成的高分辨率图像。优化的目标是将生成的高分辨率图像与原始高分辨率图像之间的误差最小化。通过训练，最终的生成器网络将有能力将任何输入的低分辨率图像转换成高分辨率图像。基于SRCNN，后续研究主要有两个方向。首先是网络架构的改进，例如高学习上采样模块(Zeiler et al., 2011)或转置合并模块(Dumoulin and Visin, 2016)，以及添加一个更强大的网络块(Kim et al., 2016)等。二是目标函数的改进，如增加对抗性训练(Ledig et al., 2017)，替换不同的误差函数，最小化生成高分辨率图像与原始高分辨率图像之间的误差(Bruna et al., 2015; Zhao et al., 2016)。

SISR是最受欢迎的研究课题之一，各种研究方法层出不穷。建议有兴趣的读者参考(Yang et al., 2019)。

4.2.4 内容-风格迁移

1. 风格计算

风格是一个非常广泛的术语，涵盖了建筑、时尚、文学、音乐、艺术和诸多其他领域。维基百科

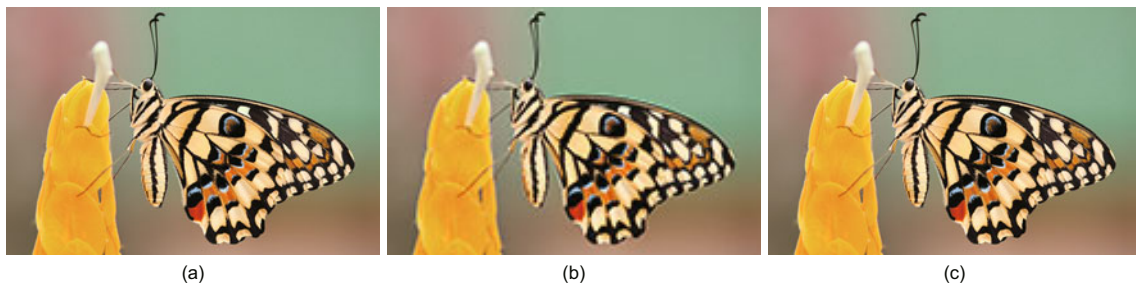


Fig. 9 不同超分辨率方法之间的视觉比较: (a)高分辨率的原始图像, (b)Keys (1981)方法的超分辨率的结果, (c)Wang XT et al. (2018)方法的超分辨率结果

对风格的定义如下：风格是一种做事或呈现事物的方式。在与人工智能相关的研究中，风格一般是指绘画风格，包括但不限于绘画的体裁、笔触、画笔类型、质地、构图类型、色彩构成等。然而，即使只对于绘画风格来说，目前的风格计算方法仍然难以量化。描述风格最常见的形式化方法是统计与空间位置无关的信息。虽然这只是一简化的量化风格的方法，但它仍然可以产生视觉上吸引人的效果。如何用美学知识来形式化风格仍然是一个值得探讨的研究课题，本节主要对已有的形式化方法进行回顾。

在目前的风格形式化方法基础上，研究者们已经提出了各种风格计算的思想，比如Gram 矩阵(Gatys et al., 2016a)、均值差(Huang and Belongie, 2017)&方差(Huang and Belongie, 2017)、白化和颜色迁移(Li et al., 2017)、线性变换(Li et al., 2018)等。

2. 内容-风格迁移

早在20世纪90年代中期，许多学者就在研究如何利用计算机技术将照片自动转换成合成艺术品。在这些研究中，最突出的是非真实感渲染(NPR) (Gooch and Gooch, 2001; Strothotte and Schlechtweg, 2002)。该方法把内容风格迁移问题看作纹理生成的问题，从一幅图像中提取纹理，然后依据另一幅图像的外观合成输出(图10)。

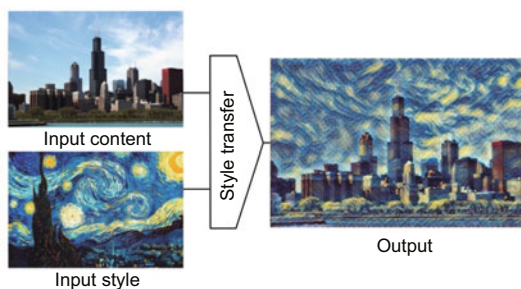


Fig. 10 风格迁移算法示例，用于将绘画风格转换到给定的照片上，风格图片为文森特·梵高的《星夜》

风格迁移的工作于2001年提出，称为图像类比(Hertzmann et al., 2001)。该方法从风格图像中提取风格，然后根据内容图像的外观合成输出。通过构造一对未风格化和风格化的图像，然后使用类比变换来获得图像 x ，它满足 $A_c : A_s :: B_c : x$ 。

但是，用图像对构造数据集非常困难。此外，单个类比约束并不能很好地捕获图像的结构和特征信息，因此风格迁移的效果并不令人满意。近年来，随着卷积神经网络(CNN)的发展，Gatys et al. (2016b) 提出使用CNN提取著名绘画作品的特征，然后将提取的特征与自然照片匹配以生成具有名画风格的图片。具体来说，他们将内容图和风格图输入到预先训练好的CNN模型中，然后VGG-19网络(Simonyan and Zisserman, 2014) 捕获不同层上相应图像的特征表示。接下来，从随机噪声初始化的生成图像开始，还将其馈入VGG-19以计算其在不同层的特征表示。对于内容，要求生成的图像和内容图像在VGG-19 高层特征图上保持一致(ReLU4_2)。对于风格，要求生成的图像和风格图像在从低到高的多个特征层上保持统计信息的一致性(ReLU1_1、ReLU2_1、ReLU3_1、ReLU4_1和ReLU5_1)。然后，通过更新模型参数，使得生成的图像满足内容和风格两个约束条件。在这篇开创性的论文中，使用的汇总特征统计量是Gram 矩阵。

Gatys et al. (2016b)因其引人注目的效果得到社会广泛认可。但它仍存在两个问题。首先，基于图像的优化方法生成缓慢。例如，生成一张 512×512 的图片需要2-3分钟。其次，基于Gram 的算法在优化过程中不稳定，需要手动调整参数。目前，许多研究工作已经开始着手解决这两个问题。对于第一个问题，Johnson et al. (2016) 提出使用基于模型的优化代替基于图像的优化，这大大提高了生成效率。Li and Wand (2016) 在马尔可夫随机场方法的启发下，提出了一种通过对抗训练的马尔可夫前馈网络来解决非参数方法的效率问题。对于第二个问题，主要通过使用不同的统计方法来简化和替换Gram 矩阵的计算，以获得稳定的样式表示形式，如(Huang and Belongie, 2017; Li YJ et al., 2017; Luo and Tang, 2008)。

在基于模型训练的基础上，研究人员同样继续优化方法以提高模型的泛化能力。从最初单模型单风格到随后的单模型多风格(Li and Wand, 2016)，再到最近的单模型任意风格(Huang and Belongie, 2017; Li YJ et al., 2017)。随着模型效率和泛化能力的不断提高，当前的风格迁移可以达到实时的商业应用水平。市场上已经出现了一系列使

用风格迁移算法的娱乐软件和平台，如Prisma、Ostagram和Deep Forger。

此外，风格迁移与其他领域的结合也产生了一系列相关研究成果，例如涂鸦风格迁移(Champandard, 2016)、支持VR/AR的3D风格迁移(Chen et al., 2018)、视频风格迁移(Chan et al., 2018)、时尚风格迁移(Jiang and Fu, 2017)和语音风格迁移(Verma and Smith, 2018)。有关风格迁移的详细研究，建议有兴趣的读者参考(Jing et al., 2019)。

5 基于AI的设计评价

创意设计的评价通常由人来完成。随着人工智能生成设计和内容的爆炸式增长，通过人工评价这些设计内容变得不太现实，所以亟需AI设计评价。目前，在AI设计评价方面，设计师、心理学家、哲学家、认知科学家和计算机科学家之间尚未达成共识。在研究领域，美学和功能是最常见的设计评价准则。基于美学方面的AI设计评价已经开展许多工作，如计算美学(Deng et al., 2018)。而关于功能的AI设计评价需要基于常识推理的高级智能，这方面的研究尚未开展。因此，在本章中，我们主要介绍计算美学及其应用。

5.1 计算美学

最早，Garabedian (1934)提出了一种定量描述图像美感的方法。近年来，计算机视觉研究人员已开始关注计算美学的方向。计算美学旨在计算不同类型的设计的美感，如图像(Datta et al., 2006; Lu et al., 2015)、网页(Dou et al., 2019)、标志(Zhang et al., 2017)、服装(Jia et al., 2016)等。

典型的美学计算系统通常包含特征提取器和美学预测器，如图11所示。对于一个待判别的图像，系统首先使用特征提取器提取美学特征，之后使用美学特征预测美学值。人工设计的特征提取器模仿专家的评价方法，对图像的不同方面进行建模。这种方法的优点是计算结果具有可解释性。但是人工设计的特征提取器和预测器不能进行端到端的训练，所以这种方法的准确性受限且提升难度很大。深度特征提取器和预测器可以进行端到端的训练，两者同时达到最优解，极大提升了预测的准确度，但是

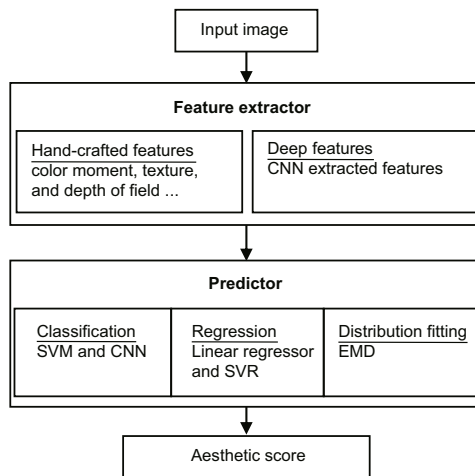


Fig. 11 典型计算美学系统流程图

预测结果缺乏可解释性。

许多研究工作(Datta et al., 2006; Murray et al., 2012; Wang et al., 2016)探索了使用人工设计的特征（如色矩、纹理(Ciesielski et al., 2013)、图像的颜色复杂度和纹理复杂度(Rigau et al., 2008)、色彩模版(Li and Chen, 2009)和景深(Datta et al., 2006))进行美学评价的方法。但人工设计特征需要极高的专业知识，且很难描述全部的美学特征。因此，这类方法的准确度有限。

Lu et al. (2015)首次将卷积神经网络应用在美学计算领域中，与之前人工设计特征的方法相比，其准确率取得了较大进步。同时，Lu et al. (2015)提出一种深度多区块聚合网络，用于训练从任意尺寸图像中裁剪的多个区块。但这类网络只能输入固定尺寸的图像，为满足这一要求，输入图像通常经过裁剪、下采样或者填充来满足输入要求，但是变换会导致原图像的美学性发生变化。为解决上述问题，Mai et al. (2016)提出一种由三个含有不同尺寸的自适应池化层的子网络构成的美学计算网络，并使用子网络提取的图像特征训练美学回归模型。Ma et al. (2017)提出了另外一种保留图像布局的美学计算方法，使用图像的显著性图和模式的多样性来选择具有辨识性的图像区块来代替Lu et al. (2015)中的随机裁剪的方式，之后将提取图像的局部特征和全局特征聚合，预测图像美学值。Li HH et al. (2017)将美学计算任务视为分布拟合任务而非前述的分类任务或回归任务，然后使用推土机距离来测量实际人工打分的分布和神经网络

预测的分布之间的差异。

另外，还有一些有价值的计算美学工作。Li HH et al. (2017)提出美学评价是一种主观行为，并基于美学和内容属性构建了通用计算美学模型和基于残差的个性化计算美学模型，以解决个性化的计算美学问题。受到图像到文字转化任务的启发，Wang WS et al. (2018)提出了一个多任务的图像美学评论系统，不仅可以预测美学结果和生成图像的评论，而且生成的图像评论还可以部分解释模型给出的美学评分。Jia et al. (2016)致力于解决服装的美学评价问题，通过建立了一个包含视觉特征、图像空间和美学词汇的三层模型来解决视觉图像和美学词汇之间的语义鸿沟。网页是现在最流行的媒体形式之一，网页设计是设计领域重要的组成部分。Dou et al. (2019)提出了自动评价网页设计美学的深度神经网络，在研究中，为了克服网页数据的缺失问题，模型首先在Flickr Style数据集上进行图像风格识别任务的预训练，之后再将预训练的模型迁移到网页美学评价任务上。标志是图形设计中非常核心的领域，Zhang et al. (2017)针对标志的美学评价进行了研究。在这项工作中，研究人员使用专业知识来指导人工设计特征，并收集了一个由60位志愿者对平衡、对比、和谐和美学属性进行打分的标志数据集。最后，使用一个线性回归模型拟合人工设计特征和人工评分，并为平衡、对比、和谐等美学价值的评估构建了回归模型。

5.2 计算美学数据集

在机器学习中高质量的数据集不可或缺，高质量的开放数据集对于计算美学研究同样重要。计算美学中最常用的是AVA (large-scale database for Aesthetic Visual Analysis) 数据集，常用的计算美学数据集如表1所示。

5.3 计算美学的应用

计算美学的目标不仅仅是获取美学值，还有基于此的进一步应用。Wang and Shen (2017)使用计算美学进行图像的自动裁剪，并设计一个两分枝的神经网络用于预测图像的美学值和注意力边界框。如图12a-c所示，基于注意力边界框生成一组候选裁剪图像，利用计算美学网络选取最佳的候选裁剪图像。Deng et al. (2018)提出一种基于对抗学习的图像色彩自动增强和裁剪模型。该模型使用一个生成网络生成色彩增强和图像裁剪的参数，使用一个判别网络来判断图像的美学值，生成网络和预训练的判别网络使用生成对抗网络中的最大最小对抗的方式进行训练，部分结果如图12d-f所示。另外，谷歌研究团队发布了用于艺术内容创作的Creatism (Fang and Zhang, 2017)。该系统模仿风光摄影师的整个工作流程，从取景到后期调整。在实际工作中，研究人员选择了四个独立的操作：根据Google 街景全景图像、饱和滤镜、HDR滤镜和动态蒙版来合成图像。

6 开放性问题 and 挑战

设计智能是人工智能的一个新分支，主要面向创意领域中的人、机器以及人机协作。尽管设计智能已取得长足的进展，但仍存在一些问题和挑战，阻碍了设计智能的进一步应用，并且仍有待解决。

设计智能方法在应用前需要利用数据训练模型，关于数据中开放性问题 and 挑战分为三种类型。首先，创造力、新颖性、多样性和一些相关的抽象评价是设计问题的主要方面，尤其是在创意激发和评价阶段。但是，这些评价需要专家给出意见作为参考，从而变为计算机可以理解的标签或者数值，因此数据的标注成本较高。其次，专家意见不可避免具有

Table 1 Commonly used datasets in aesthetic assessment

名称	数量	简介	参考文献
AVA	250 000	每个图像有78~549个投票。部分图像还包含14种风格属性和60种类别。	Murray et al. (2012)
CUHK-PQ	17 690	图像包含二分类的标签和7种场景。	Tang et al. (2013)
IAD	1 500 000	图像从DPChallenge和photo.net网站上爬取。	Lu et al. (2015))
AROD	380 000	图像从Flickr网站收集。包含观看次数、评论、喜爱列表、标题和描述。	Schwarz et al. (2018)
AADB	10 000	每个图像提供11种不同的属性和美学评分。	Kong et al. (2016)

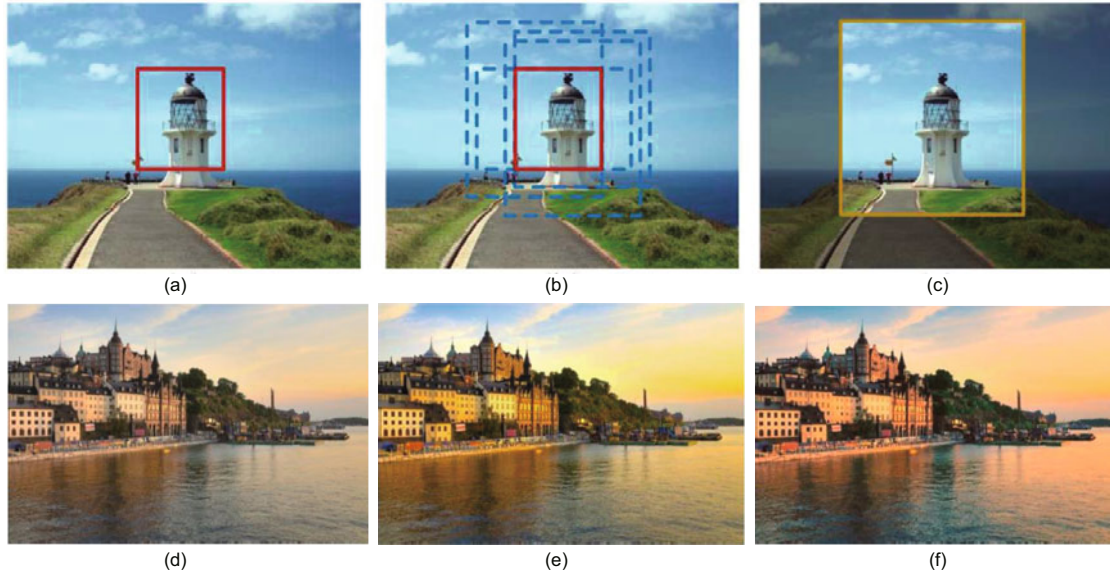


Fig. 12 (a–c)来自Wang and Shen (2017): 红色框是最初的裁剪(a), 蓝色的是候选的裁剪图像(b), 美学计算网络选择的最佳候选裁剪图像(c); (d–f) 来自Deng et al. (2018): 输入图像(d), 人工增强的输出(e), 深度图像增强网络的输出(f)
(a–c), Reprint from Wang and Shen (2017), Copyright 2017, with permission from IEEE (d–f), Reprint from Deng et al. (2018), Copyright 2018, with permission from ACM (Association for Computing Machinery)

主观性，而评价数据中过大的变化和多样性也会加大模型训练的难度。最后，由于成本高，数据量可能受到限制，从而影响算法的泛化能力。因此，为设计问题训练特定的模型成本较高，然而定制化的需求却常见于设计任务中，由此形成了设计任务和模型学习的矛盾。为了应对这些挑战，必须使用具有多属性和高度抽象的更开放的数据集，使机器学习算法对设计有更好的了解。

设计智能算法的首要挑战是样本的数量。众所周知，专业设计师只需有限数量的样本就可以为设计问题提供有价值的解决方案，而智能方法通常需要大量样本才能学习到适用的模型。小样本学习目前仍是学术界的前沿探索领域，尚未成熟。由于许多抽象的语义信息隐藏在设计问题之后，因此，小样本学习任务将是未来设计智能研究的挑战性方向。

设计智能算法的第二个挑战是模型和结果的可解释性。在创意激发阶段，设计人员可能想了解为什么计算机系统会在特定设计问题的情况下给予或不给予某种刺激。计算机给出产品设计的同时也需要附上解释，说明为什么生成当前风格而非其他风格。当计算机处理输入的设计草图时，评价系统应该给出美学评分，并且指出设计不够出色的原因，

进而提供改进建议。随着智能方法逐渐成为设计工具，设计人员需要具备对算法模型工作原理以及输入输出之间关系的基本认识。但是，大多数智能方法不具备可解释的理解渠道，这无形中增加了设计师使用相应工具的障碍。

设计智能算法的第三个挑战是难以将先验知识或领域知识引入模型的学习中，尤其是在智能内容生成中。大多数设计知识由经验规则或准则组成，很难正确形式化。虽然跨媒体数据为设计师提供了创建精确用户画像、提供启发性刺激并生成生动内容的机会，但是，基于跨媒体的用户画像生成需要先进的跨媒体计算和数据融合方法，而这些方法仍是一个公开的挑战。

最后，人机协作中的挑战是设计智能的核心问题。智能方法被视为解决设计过程中复杂问题的辅助工具，例如前面部分中提到的刺激搜索和超分辨率。很多繁重而重复的工作可以由智能算法来完成，但许多设计师并未意识到新技术的进展，还在从事相关工作。因此，设计者与人工智能之间的新合作模式对于设计智能的发展非常重要，人工智能完全可以充当设计者的得力助手。此外，在用户需求分析和内容生成中对人机协作的评估仍然是一个待解决的问题。

7 结语

在本文中，我们广泛回顾了AI 2.0时代的设计智能方法和应用。我们首先总结了设计智能的概念和框架，然后描述了设计智能中创意模块的四个组成部分，即基于AI的用户需求分析、基于AI的创意激发、基于AI的内容生成和基于AI的设计评估。具体来说，我们介绍了四个组件中的最新技术，尤其是智能内容生成的模型和方法。此外，我们讨论了设计智能领域未解决的问题和未来研究的挑战。由于设计智能已成为AI 2.0的重要分支，因此我们希望本文能够启发和帮助研究人员研究新兴的设计智能新方向。

遵守伦理准则声明

作者声明发表这篇论文没有利益冲突。

致谢

本文Fig. 5c、5d和Fig. 8是由Runway工具包的预训练模型生成的(runwayml.com)。

References

- Arjovsky M, Chintala S, Bottou L, 2017. Wasserstein generative adversarial networks. *Proc 34th Int Conf on Machine Learning*, p.298-321.
- Aubry M, Maturana D, Efros AA, et al., 2014. Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.3762-3769. <https://doi.org/10.1109/CVPR.2014.487>
- Ballester C, Bertalmio M, Caselles V, et al., 2001. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans Image Process*, 10(8):1200-1211. <https://doi.org/10.1109/83.935036>
- Bertalmio M, Sapiro G, Caselles V, et al., 2000. Image inpainting. *Proc 27th Annual Conf on Computer Graphics and Interactive Techniques*, p.417-424. <https://doi.org/10.1145/344779.344972>
- Bharadhwaj H, Park H, Lim BY, 2018. RecGAN: recurrent generative adversarial networks for recommendation systems. *Proc 12th ACM Conf on Recommender Systems*, p.372-376. <https://doi.org/10.1145/3240323.3240383>
- Boden MA, 2009. Computer models of creativity. *AI Mag*, 30(3):23-34. <https://doi.org/10.1609/aimag.v30i3.2254>
- Brock A, Donahue J, Simonyan K, 2018. Large scale GAN training for high fidelity natural image synthesis. <https://arxiv.org/abs/1809.11096>
- Bruna J, Sprechmann P, LeCun Y, 2015. Super-resolution with deep convolutional sufficient statistics. <https://arxiv.org/abs/1511.05666>
- Chakrabarti A, Siddharth L, Dinakar M, et al., 2017. Idea inspire 3.0—a tool for analogical design. In: Chakrabarti A, Chakrabarti D (Eds.), *Research into Design for Communities*. Springer, Singapore, p.475-485. https://doi.org/10.1007/978-981-10-3521-0_41
- Champandard AJ, 2016. Semantic style transfer and turning two-bit doodles into fine artworks. <https://arxiv.org/abs/1603.01768>
- Chan C, Ginosar S, Zhou TH, et al., 2018. Everybody dance now. <https://arxiv.org/abs/1808.07371>
- Chen DD, Yuan L, Liao J, et al., 2018. Stereoscopic neural style transfer. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.6654-6663. <https://doi.org/10.1109/CVPR.2018.00696>
- Chen LQ, Wang P, Dong H, et al., 2019. An artificial intelligence based data-driven approach for design ideation. *J Vis Commun Image Represent*, 61:10-22. <https://doi.org/10.1016/j.jvcir.2019.02.009>
- Ciesielski V, Barile P, Trist K, 2013. Finding image features associated with high aesthetic value by machine learning. *Proc 2nd Int Conf on Evolutionary and Biologically Inspired Music, Sound, Art and Design*, p.47-58. https://doi.org/10.1007/978-3-642-36955-1_5
- Cooper A, 1999. *The Inmates Are Running the Asylum*. SAMS, Indianapolis, USA.
- Cooper A, Reimann RM, 2003. *About Face 2.0: the Essentials of Interaction Design*. John Wiley & Sons, Indianapolis, USA.
- Dash A, Gamboa JCB, Ahmed S, et al., 2017. TAC-GAN-text conditioned auxiliary classifier generative adversarial network. <https://arxiv.org/abs/1703.06412>
- Datta R, Joshi D, Li J, et al., 2006. Studying aesthetics in photographic images using a computational approach. *Proc 9th European Conf on Computer Vision*, p.288-301. https://doi.org/10.1007/11744078_23
- de Gómez Silva Garza A, Maher ML, 1999. An evolutionary approach to case adaptation. *Proc 3rd Int Conf on Case-Based Reasoning*, p.162-173. https://doi.org/10.1007/3-540-48508-2_12
- de Silva Garza AG, 2019. An introduction to and comparison of computational creativity and design computing. *Artif Intell Rev*, 51(1):61-76. <https://doi.org/10.1007/s10462-017-9557-3>
- Deng J, Dong W, Socher R, et al., 2009. ImageNet: a large-scale hierarchical image database. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Deng YB, Loy CC, Tang XO, 2018. Aesthetic-driven image enhancement by adversarial learning. *Proc 26th ACM Int Conf on Multimedia*, p.870-878. <https://doi.org/10.1145/3240508.3240531>

- Donahue J, Krähenbühl P, Darrell T, 2016. Adversarial feature learning. <https://arxiv.org/abs/1605.09782>
- Dou Q, Zheng XS, Sun TF, et al., 2019. Webthetics: quantifying webpage aesthetics with deep learning. *Int J Hum Comput Stud*, 124:56-66. <https://doi.org/10.1016/j.ijhcs.2018.11.006>
- Dugosh KL, Paulus PB, Roland EJ, et al., 2000. Cognitive stimulation in brainstorming. *J Pers Soc Psychol*, 79(5):722-735. <https://doi.org/10.1037/0022-3514.79.5.722>
- Dumoulin V, Visin F, 2016. A guide to convolution arithmetic for deep learning. <https://arxiv.org/abs/1603.07285>
- Edelman RR, Hesselink JR, Zlatkin MB, 1996. MRI: Clinical Magnetic Resonance Imaging. Saunders, Philadelphia.
- Efros AA, Freeman WT, 2001. Image quilting for texture synthesis and transfer. Proc 28th Annual Conf on Computer Graphics and Interactive Techniques, p.341-346. <https://doi.org/10.1145/383259.383296>
- Elgammal A, Liu B, Elhoseiny M, et al., 2017. CAN: creative adversarial networks, generating “art” by learning about styles and deviating from style norms. <https://arxiv.org/abs/1706.07068>
- Fang H, Zhang M, 2017. Creatism: a deep-learning photographer capable of creating professional work. <https://arxiv.org/abs/1707.03491>
- Faste H, Rachmel N, Essary R, et al., 2013. Brainstorm, chainstorm, cheatstorm, tweetstorm: new ideation strategies for distributed HCI design. Proc Conf on Human Factors in Computing Systems, p.1343-1352. <https://doi.org/10.1145/2470654.2466177>
- Fu K, Murphy J, Yang M, et al., 2015. Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. *Res Eng Des*, 26(1):77-95. <https://doi.org/10.1007/s00163-014-0186-4>
- Garabedian CA, 1934. Birkhoff on aesthetic measure. *Bull Amer Math Soc*, 40(1):7-10. <https://doi.org/10.1090/S0002-9904-1934-05764-1>
- Gatys L, Ecker A, Bethge M, 2016a. Image style transfer using convolutional neural networks. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.2414-2423. <https://doi.org/10.1109/CVPR.2016.265>
- Gatys L, Ecker A, Bethge M, 2016b. A neural algorithm of artistic style. *J Vis*, 16(12):326. <https://doi.org/10.1167/16.12.326>
- Gero JS, 1990. Design prototypes: a knowledge representation schema for design. *AI Mag*, 11(4):26-36.
- Gilon K, Chan J, Ng FY, et al., 2018. Analogy mining for specific design needs. Proc CHI Conf on Human Factors in Computing Systems, p.121. <https://doi.org/10.1145/3173574.3173695>
- Goel AK, Rugaber S, Vattam S, 2009. Structure, behavior, and function of complex systems: the structure, behavior, and function modeling language. *AI Edam*, 23(1):23-35. <https://doi.org/10.1017/S0890060409000080>
- Goldschmidt G, Smolkov M, 2006. Variances in the impact of visual stimuli on design problem solving performance. *Des Stud*, 27(5):549-569. <https://doi.org/10.1016/j.destud.2006.01.002>
- Gooch B, Gooch A, 2001. Non-photorealistic Rendering. A K Peters/CRC Press, New York, USA. <https://doi.org/10.1201/9781439864173>
- Goodfellow I, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.
- Grudin J, Pruitt J, 2002. Personas, participatory design, and product development: an infrastructure for engagement. Proc 7th Biennial Participatory Design Conf, p.144-152.
- Gulrajani I, Ahmed F, Arjovsky M, et al., 2017. Improved training of Wasserstein GANs. Advances in Neural Information Proc Systems, p.5767-5777.
- Han J, Shi F, Chen LQ, et al., 2018. A computational tool for creative idea generation based on analogical reasoning and ontology. *Artif Intell Eng Des Anal Manuf*, 32(4):462-477. <https://doi.org/10.1017/S0890060418000082>
- Hao J, Zhou YJ, Zhao QF, et al., 2019. An evolutionary computation based method for creative design inspiration generation. *J Intell Manuf*, 30(4):1673-1691. <https://doi.org/10.1007/s10845-017-1347-x>
- Hartson R, Pyla PS, 2012. The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Elsevier, Amsterdam. <https://doi.org/10.1016/C2010-0-66326-7>
- He KM, Sun J, 2014. Image completion approaches using the statistics of similar patches. *IEEE Trans Patt Anal Mach Intell*, 36(12):2423-2435. <https://doi.org/10.1109/TPAMI.2014.2330611>
- Hertzmann A, Jacobs CE, Oliver N, et al., 2001. Image analogies. Proc 28th Annual Conf on Computer Graphics and Interactive Techniques, p.327-340. <https://doi.org/10.1145/383259.383295>
- Hong YJ, Hwang U, Yoo J, et al., 2019. How generative adversarial networks and their variants work: an overview. *ACM Comput Surv*, 52(1):10. <https://doi.org/10.1145/3301282>
- Huang HZ, Wang H, Luo WH, et al., 2017. Real-time neural style transfer for videos. IEEE Conf on Computer Vision and Pattern Recognition, p.7044-7052. <https://doi.org/10.1109/CVPR.2017.745>
- Huang X, Belongie S, 2017. Arbitrary style transfer in real-time with adaptive instance normalization. Proc IEEE Int Conf on Computer Vision, p.1501-1510. <https://doi.org/10.1109/ICCV.2017.167>

- Iizuka S, Simo-Serra E, Ishikawa H, 2017. Globally and locally consistent image completion. *ACM Trans Graph*, 36(4), Article 107. <https://doi.org/10.1145/3072959.3073659>
- Isola P, Zhu JY, Zhou TH, et al., 2017. Image-to-image translation with conditional adversarial networks. *IEEE Conf on Computer Vision and Pattern Recognition*, p.5967-5976. <https://doi.org/10.1109/CVPR.2017.632>
- Jansen BJ, Jung SG, Salminen J, et al., 2017. Viewed by too many or viewed too little: using information dissemination for audience segmentation. *Proc Assoc Inform Sci Technol*, 54(1):189-196. <https://doi.org/10.1002/pras.2017.14505401021>
- Jansson DG, Smith SM, 1991. Design fixation. *Des Stud*, 12(1):3-11. [https://doi.org/10.1016/0142-694X\(91\)90003-F](https://doi.org/10.1016/0142-694X(91)90003-F)
- Jia J, Huang J, Shen GY, et al., 2016. Learning to appreciate the aesthetic effects of clothing. *Proc 30th AAAI Conf on Artificial Intelligence*, p.1216-1222.
- Jia L, Becattini N, Cascini G, et al., 2020. Testing ideation performance on a large set of designers: effects of analogical distance. *Int J Des Creat Innov*, 8(1):31-45. <https://doi.org/10.1080/21650349.2019.1618736>
- Jiang SH, Fu Y, 2017. Fashion style generator. *Proc 26th Int Joint Conf on Artificial Intelligence*, p.3721-3727. <https://doi.org/10.24963/ijcai.2017/520>
- Jing YC, Yang YZ, Feng ZL, et al., 2019. Neural style transfer: a review. *IEEE Trans Vis Comput Graph*, in press. <https://doi.org/10.1109/tvcg.2019.2921336>
- Jo Y, Park J, 2019. SC-FEGAN: face editing generative adversarial network with user's sketch and color. <https://arxiv.org/abs/1902.06838>
- Johnson J, Alahi A, Li FF, 2016. Perceptual losses for real-time style transfer and super-resolution. *Proc 14th European Conf*, p.694-711. https://doi.org/10.1007/978-3-319-46475-6_43
- Karras T, Laine S, Aila T, 2019. A style-based generator architecture for generative adversarial networks. *The IEEE Conf on Computer Vision and Pattern Recognition*, p.4401-4410.
- Kaufman JC, Sternberg RJ, 2006. *The International Handbook of Creativity*. Edward Elgar Publishing, Cheltenham, UK.
- Keys R, 1981. Cubic convolution interpolation for digital image processing. *IEEE Trans Acoust Speech Signal Process*, 29(6):1153-1160. <https://doi.org/10.1109/TASSP.1981.1163711>
- Kim J, Lee JK, Lee KM, 2016. Accurate image super-resolution using very deep convolutional networks. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.1646-1654. <https://doi.org/10.1109/CVPR.2016.182>
- Kingma DP, Welling M, 2013. Auto-encoding variational Bayes. <https://arxiv.org/abs/1312.6114>
- Kong S, Shen XH, Lin Z, et al., 2016. Photo aesthetics ranking network with attributes and content adaptation. *Proc 14th European Conf on Computer Vision*, p.662-679. https://doi.org/10.1007/978-3-319-46448-0_40
- Krizhevsky A, Hinton G, 2009. Learning Multiple Layers of Features from Tiny Images. Technical Report, University of Toronto, Toronto.
- Kwak H, An J, Jansen BJ, 2017. Automatic generation of personas using YouTube social media data. *Proc 50th Hawaii Int Conf on System Sciences*, p.833-842.
- Larsen ABL, Sønderby SK, Larochelle H, et al., 2016. Autoencoding beyond pixels using a learned similarity metric. *Proc 33rd Int Conf on Machine Learning*, p.1558-1566.
- LeCun Y, Bottou L, Bengio Y, et al., 1998. Gradient-based learning applied to document recognition. *Proc IEEE*, 86(11):2278-2323. <https://doi.org/10.1109/5.726791>
- Ledig C, Theis L, Huszár F, et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network. *IEEE Conf on Computer Vision and Pattern Recognition*, p.105-114. <https://doi.org/10.1109/CVPR.2017.19>
- Li C, Wand M, 2016. Precomputed real-time texture synthesis with Markovian generative adversarial networks. *Proc 14th European Conf on Computer Vision*, p.702-716. https://doi.org/10.1007/978-3-319-46487-9_43
- Li CC, Chen T, 2009. Aesthetic visual quality assessment of paintings. *IEEE J Sel Top Signal Process*, 3(2):236-252. <https://doi.org/10.1109/JSTSP.2009.2015077>
- Li HH, Wang JG, Tang MM, et al., 2017. Polarization-dependent effects of an Airy beam due to the spin-orbit coupling. *J Opt Soc Am A*, 34(7):1114-1118. <https://doi.org/10.1364/JOSAA.34.001114>
- Li XT, Liu SF, Kautz J, et al., 2019. Learning linear transformations for fast arbitrary style transfer. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.3809-3817.
- Li YJ, Fang C, Yang JM, et al., 2017. Universal style transfer via feature transforms. *Proc 31st Conf on Neural Information Processing Systems*, p.386-396.
- Liu GL, Reda FA, Shih KJ, et al., 2018. Image inpainting for irregular holes using partial convolutions. *Proc 15th European Conf on Computer Vision*, p.85-105. https://doi.org/10.1007/978-3-030-01252-6_6
- Liu H, Singh P, 2004. ConceptNet—a practical commonsense reasoning tool-kit. *BT Technol J*, 22(4):211-226. <https://doi.org/10.1023/B:BTJ.0000047600.45421.6d>
- Liu MY, Huang X, Mallya A, et al., 2019. Few-shot unsupervised image-to-image translation. <https://arxiv.org/abs/1905.01723>

- Liu ZW, Luo P, Wang XG, et al., 2015. Deep learning face attributes in the wild. *Proc IEEE Int Conf on Computer Vision*, p.3730-3738. <https://doi.org/10.1109/ICCV.2015.425>
- Lowdermilk T, 2013. *User-Centered Design: a Developer's Guide to Building User-Friendly Applications*. O'Reilly, Beijing, China.
- Lu X, Lin Z, Shen XH, et al., 2015. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. *Proc IEEE Int Conf on Computer Vision*, p.990-998. <https://doi.org/10.1109/ICCV.2015.119>
- Luo YW, Tang XO, 2008. Photo and video quality evaluation: focusing on the subject. *Proc 10th European Conf on Computer Vision*, p.386-399.
- Ma S, Liu J, Chen WC, 2017. A-lamp: adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. *Proc 30th IEEE Conf on Computer Vision and Pattern Recognition*, p.722-731. <https://doi.org/10.1109/CVPR.2017.84>
- Maguire M, Bevan N, 2002. User requirements analysis. In: Hammond J, Gross T, Wesson J (Eds.), *Usability: Gaining a Competitive Edge*. Springer, Boston, USA, p.133-148. https://doi.org/10.1007/978-0-387-35610-5_9
- Mai L, Jin HL, Liu F, 2016. Composition-preserving deep photo aesthetics assessment. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.497-506. <https://doi.org/10.1109/CVPR.2016.60>
- Matthews T, Judge T, Whittaker S, 2012. How do designers and user experience professionals actually perceive and use personas? *Proc Conf on Human Factors in Computing Systems*, p.1219-1228. <https://doi.org/10.1145/2207676.2208573>
- McCaffrey T, Krishnamurty S, 2015. The obscure features hypothesis in design innovation. *Int J Des Creat Innov*, 3(1):1-28. <https://doi.org/10.1080/21650349.2014.893840>
- McGinn J, Kotamraju N, 2008. Data-driven persona development. *Proc Conf on Human Factors in Computing Systems*, p.1521-1524. <https://doi.org/10.1145/1357054.1357292>
- Miaskiewicz T, Kozar KA, 2011. Personas and user-centered design: how can personas benefit product design processes? *Des Stud*, 32(5):417-430. <https://doi.org/10.1016/j.destud.2011.03.003>
- Mikolov T, Chen K, Corrado G, et al., 2013. Efficient estimation of word representations in vector space. <https://arxiv.org/abs/1301.3781>
- Miller GA, 1995. Wordnet: a lexical database for English. *Commun ACM*, 38(11):39-41. <https://doi.org/10.1145/219717.219748>
- Mirza M, Osindero S, 2014. Conditional generative adversarial nets. <https://arxiv.org/abs/1411.1784>
- Miyato T, Kataoka T, Koyama M, et al., 2018. Spectral normalization for generative adversarial networks. *Int Conf on Learning Representations*.
- Murray N, Marchesotti L, Perronnin F, 2012. AVA: a large-scale database for aesthetic visual analysis. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.2408-2415. <https://doi.org/10.1109/CVPR.2012.6247954>
- Nazeri K, Ng E, Joseph T, et al., 2019. Edgeconnect: generative image inpainting with adversarial edge learning. <https://arxiv.org/abs/1901.00212>
- Nelson BA, Wilson JO, Rosen D, et al., 2009. Refined metrics for measuring ideation effectiveness. *Des Stud*, 30(6):737-743. <https://doi.org/10.1016/j.destud.2009.07.002>
- Nielsen L, Hansen KS, Stage J, et al., 2015. A template for design personas: analysis of 47 persona descriptions from Danish industries and organizations. *Int J Sociotechnol Knowl Dev*, 7(1):45-61. <https://doi.org/10.4018/ijskd.2015010104>
- Niles I, Pease A, 2001. Towards a standard upper ontology. *Proc Int Conf on Formal Ontology in Information Systems*, p.2-9. <https://doi.org/10.1145/505168.505170>
- Nilsback ME, Zisserman A, 2008. Automated flower classification over a large number of classes. *Proc 6th Indian Conf on Computer Vision, Graphics & Image Processing*, p.722-729. <https://doi.org/10.1109/ICVGIP.2008.47>
- Odena A, Olah C, Shlens J, 2017. Conditional image synthesis with auxiliary classifier GANs. *Proc 34th Int Conf on Machine Learning*, p.4043-4055.
- Pan YH, 2017. Special issue on artificial intelligence 2.0. *Front Inform Technol Electron Eng*, 18(1):1-2. <https://doi.org/10.1631/FITEE.1710000>
- Park T, Liu MY, Wang TC, et al., 2019. Semantic image synthesis with spatially-adaptive normalization. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.2337-2346.
- Pathak D, Krähenbühl P, Donahue J, et al., 2016. Context encoders: feature learning by inpainting. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.2536-2544. <https://doi.org/10.1109/CVPR.2016.278>
- Peeters JR, Verhaegen PA, Vandevenne D, et al., 2010. Refined metrics for measuring novelty in ideation. *IDMME Virtual Concept Research in Interaction Design*, Article 4.
- Perera D, Zimmermann R, 2019. CNGAN: generative adversarial networks for cross-network user preference generation for non-overlapped users. *World Wide Web Conf*, p.3144-3150. <https://doi.org/10.1145/3308558.3313733>

- Pruitt J, Adlin T, 2005. The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Elsevier, Amsterdam, p.724.
- Radford A, Metz L, Chintala S, 2016. Unsupervised representation learning with deep convolutional generative adversarial networks. Proc 4th Int Conf on Learning Representations.
- Reed SE, Akata Z, Yan XC, et al., 2016a. Generative adversarial text to image synthesis. Proc 33rd Int Conf on Machine Learning, p.1681-1690.
- Reed SE, Akata Z, Mohan S, et al., 2016b. Learning what and where to draw. Advances in Neural Information Processing Systems, p.217-225.
- Rigau J, Feixas M, Sbert M, 2008. Informational aesthetics measures. *IEEE Comput Graph Appl*, 28(2):24-34. <https://doi.org/10.1109/MCG.2008.34>
- Russell SJ, Norvig P, 2016. Artificial Intelligence: a Modern Approach. Pearson Education Limited, Harlow, Essex.
- Saleh B, Elgammal A, 2015. Large-scale classification of fine-art paintings: learning the right metric on the right feature. <https://arxiv.org/abs/1505.00855>
- Salimans T, Goodfellow IJ, Zaremba W, et al., 2016. Improved techniques for training GANs. Advances in Neural Information Processing Systems, p.2226-2234.
- Salminen J, Sengün S, Kwak H, et al., 2017. Generating cultural personas from social data: a perspective of middle eastern users. Proc 5th Int Conf on Future Internet of Things and Cloud Workshops, p.120-125. <https://doi.org/10.1109/FiCloudW.2017.97>
- Salminen J, Jansen BJ, An J, et al., 2018a. Are personas done? Evaluating their usefulness in the age of digital analytics. *Persona Stud*, 4(2):47-65. <https://doi.org/10.21153/psj2018vol4no2art737>
- Salminen J, Jung SG, An J, et al., 2018b. Findings of a user study of automatically generated personas. Proc Conf on Human Factors in Computing Systems, p.LBW097. <https://doi.org/10.1145/3170427.3188470>
- Salminen J, Engün S, Jung SG, et al., 2019. Design issues in automatically generated persona profiles: a qualitative analysis from 38 think-aloud transcripts. Proc Conf on Human Information Interaction and Retrieval, p.225-229. <https://doi.org/10.1145/3295750.3298942>
- Schwarz K, Wieschollek P, Lensch HPA, 2018. Will people like your image? Learning the aesthetic space. Proc IEEE Winter Conf on Applications of Computer Vision, p.2048-2057. <https://doi.org/10.1109/WACV.2018.00226>
- Shah JJ, Kulkarni SV, Vargas-Hernandez N, 2000. Evaluation of idea generation methods for conceptual design: effectiveness metrics and design of experiments. *J Mech Des*, 122(4):377-384. <https://doi.org/10.1115/1.1315592>
- Simonyan K, Zisserman A, 2014. Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>
- Strohmann T, Siemon D, Robra-Bissantz S, 2017. brAIn-storm: intelligent assistance in group idea generation. Proc 12th Int Conf on Design Science Research in Information System and Technology, p.457-461. https://doi.org/10.1007/978-3-319-59144-5_31
- Strothotte T, Schlechtweg S, 2002. Non-photorealistic Computer Graphics: Modeling, Rendering, and Animation. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Tang X, Wang ZW, Luo WX, et al., 2018. Face aging with identity-preserved conditional generative adversarial networks. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.7939-7947. <https://doi.org/10.1109/CVPR.2018.00828>
- Tang XO, Luo W, Wang XG, 2013. Content-based photo quality assessment. *IEEE Trans Multimed*, 15(8):1930-1943. <https://doi.org/10.1109/TMM.2013.2269899>
- Vandevenne D, Verhaegen PA, Dewulf S, et al., 2015. A scalable approach for ideation in biologically inspired design. *Artif Intell Eng Des Anal Manuf*, 29(1):19-31. <https://doi.org/10.1017/S0890060414000122>
- Varshney LR, Pinel F, Varshney KR, et al., 2019. A big data approach to computational creativity: the curious case of Chef Watson. *IBM J Res Dev*, 63(1):7:1-7:18. <https://doi.org/10.1147/JRD.2019.2893905>
- Verma P, Smith JO, 2018. Neural style transfer for audio spectrograms. <https://arxiv.org/abs/1801.01589>
- Wang J, Yu LT, Zhang WN, et al., 2017. IRGAN: a minimax game for unifying generative and discriminative information retrieval models. Proc 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval, p.515-524. <https://doi.org/10.1145/3077136.3080786>
- Wang TC, Liu MY, Zhu JY, et al., 2018. Video-to-video synthesis. <https://arxiv.org/abs/1808.06601>
- Wang WG, Shen JB, 2017. Deep cropping via attention box prediction and aesthetics assessment. Proc IEEE Int Conf on Computer Vision, p.2205-2213. <https://doi.org/10.1109/ICCV.2017.240>
- Wang WN, Cai D, Wang L, et al., 2016. Synthesized computational aesthetic evaluation of photos. *Neurocomputing*, 172:244-252. <https://doi.org/10.1016/j.neucom.2014.12.106>
- Wang WS, Yang S, Zhang WS, et al., 2018. Neural aesthetic image reviewer. <https://arxiv.org/abs/1802.10240>
- Wang XT, Yu K, Wu SX, et al., 2018. ESRGAN: enhanced super-resolution generative adversarial networks. European Conf on Computer Vision, p.63-79. https://doi.org/10.1007/978-3-030-11021-5_5

- Wu JJ, Zhang CK, Xue TF, et al., 2016. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. *Advances in Neural Information Processing Systems*, p.82-90.
- Xu T, Zhang PC, Huang QY, et al., 2018. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.1316-1324. <https://doi.org/10.1109/CVPR.2018.00143>
- Yan Y, Wang JR, Tang C, et al., 2019. Research on the development of contemporary design intelligence driven by neural network technology. In: Marcus A, Wang WT (Eds.), *Design, User Experience, and Usability. Design Philosophy and Theory*. Springer, Cham, p.368-381. https://doi.org/10.1007/978-3-030-23570-3_27
- Yang HY, Huang D, Wang YH, et al., 2018. Learning face age progression: a pyramid architecture of GANs. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.31-39. <https://doi.org/10.1109/CVPR.2018.00011>
- Yang WM, Zhang XC, Tian YP, et al., 2019. Deep learning for single image super-resolution: a brief review. *IEEE Trans Multim*, 21(12):3106-3121. <https://doi.org/10.1109/tmm.2019.2919431>
- Yang Y, Zhuang YT, Wu F, et al., 2008. Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval. *IEEE Trans Multim*, 10(3):437-446. <https://doi.org/10.1109/TMM.2008.917359>
- Yi ZL, Zhang H, Tan P, et al., 2017. DualGAN: unsupervised dual learning for image-to-image translation. *Proc IEEE Int Conf on Computer Vision*, p.2868-2876. <https://doi.org/10.1109/ICCV.2017.310>
- Yoon Y, Jeon HG, Yoo D, et al., 2015. Learning a deep convolutional network for light-field image super-resolution. *Proc IEEE Int Conf on Computer Vision*, p.57-65. <https://doi.org/10.1109/ICCVW.2015.17>
- You S, You N, Pan MX, 2019. PI-REC: progressive image reconstruction network with edge and color domain. <https://arxiv.org/abs/1903.10146>
- Yu F, Zhang YD, Song SR, et al., 2015. LSUN: construction of a large-scale image dataset using deep learning with humans in the loop. <https://arxiv.org/abs/1506.03365>
- Yu JH, Lin Z, Yang JM, et al., 2018a. Free-form image inpainting with gated convolution. <https://arxiv.org/abs/1806.03589>
- Yu JH, Lin Z, Yang JM, et al., 2018b. Generative image inpainting with contextual attention. *Proc IEEE Conf on Computer Vision and Pattern Recognition*, p.5505-5514. <https://doi.org/10.1109/CVPR.2018.00577>
- Zakharov E, Shysheya A, Burkov E, et al., 2019. Few-shot adversarial learning of realistic neural talking head models. <https://arxiv.org/abs/1905.08233>
- Zeiler MD, Taylor GW, Fergus R, 2011. Adaptive deconvolutional networks for mid and high level feature learning. *Proc IEEE Int Conf on Computer Vision*, p.2018-2025. <https://doi.org/10.1109/ICCV.2011.6126474>
- Zhang H, Xu T, Li H, et al., 2017. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks. *Proc IEEE Int Conf on Computer Vision*, p.5907-5915.
- Zhang H, Xu T, Li H, et al., 2019. StackGAN++: realistic image synthesis with stacked generative adversarial networks. *IEEE Trans Patt Anal Mach Intell*, 41(8):1947-1962. <https://doi.org/10.1109/TPAMI.2018.2856256>
- Zhang JJ, Yu JH, Zhang K, et al., 2017. Computational aesthetic evaluation of logos. *ACM Trans Appl Perc*, 14(3), Article 20. <https://doi.org/10.1145/3058982>
- Zhang R, Isola P, Efros AA, 2016. Colorful image colorization. *Proc 14th European Conf on Computer Vision*, p.649-666. https://doi.org/10.1007/978-3-319-46487-9_40
- Zhao H, Gallo O, Frosio I, et al., 2016. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imag*, 3(1):47-57. <https://doi.org/10.1109/tci.2016.2644865>
- Zhu JY, Park T, Isola P, et al., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proc IEEE Int Conf on Computer Vision*, p.2242-2251. <https://doi.org/10.1109/ICCV.2017.244>