Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

A review of design intelligence: progress, problems, and challenges*

Yong-chuan TANG[†‡1,2,3], Jiang-jie HUANG[1,2], Meng-ting YAO[1,2], Jia WEI[1,2], Wei LI[1,2,3], Yong-xing HE[1,2,3], Ze-jian LI[1,3,4]

[1]*College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*
[2]*Key Laboratory of Design Intelligence and Digital Creativity of Zhejiang Province, Hangzhou 310027, China*
[3]*Zhejiang Lab, Hangzhou 310012, China*
[4]*Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Hangzhou 310027, China*
[†]E-mail: yctang@zju.edu.cn

Received Aug. 6, 2019; Revision accepted Nov. 26, 2019; Crosschecked Dec. 10, 2019

Abstract: Design intelligence is an important branch of artificial intelligence (AI), focusing on the intelligent models and algorithms in creativity and design. In the context of AI 2.0, studies on design intelligence have developed rapidly. We summarize mainly the current emerging framework of design intelligence and review the state-of-the-art techniques of related topics, including user needs analysis, ideation, content generation, and design evaluation. Specifically, the models and methods of intelligence-generated content are reviewed in detail. Finally, we discuss some open problems and challenges for future research in design intelligence.

Key words: Design intelligence; Creativity; Personas; Ideation; AI-generated content; Computational aesthetics
https://doi.org/10.1631/FITEE.1900398 **CLC number:** TP183

1 Introduction

Creativity involves thinking to produce relatively novel and compelling ideas, and it can be measured and developed (Kaufman and Sternberg, 2006). In the past three decades, there have been several studies on design and creative issues from a computational and cognitive perspective. In this study, we focus mainly on two important directions, computational creativity and design computing. Their research often overlaps. As a branch of artificial intelligence (AI), computational creativity focuses on the computational models in the creative process (Boden, 2009), working on mechanical systems to generate innovative and high-quality works for people. Design computing refers mainly to the study and practice of design activities through the application and development of novel ideas and techniques in computing. Both are to use, propose, and produce computational systems, computational algorithms, computational models, and computational representations related to design and creativity (de Silva Garza, 2019). There are also a few differences between them. Computational creativity focuses on simulating human intelligence to enhance machine creativity, while design computing is about applying models and algorithms to help designers come up with better ideas.

‡ Corresponding author

* Project supported by the National Science and Technology Innovation 2030 Major Project of the Ministry of Science and Technology of China (No. 2018AAA0100703), the National Natural Science Foundation of China (Nos. 61773336 and 91748127), the Chinese Academy of Engineering Consulting Project (No. 2018-ZD-12-06), the Provincial Key Research and Development Plan of Zhejiang Province, China (No. 2019C03137), and the Ng Teng Fong Charitable Foundation in the form of ZJU-SUTD IDEA Grant

ORCID: Yong-chuan TANG, http://orcid.org/0000-0002-0157-7771

In the past, design intelligence was based mainly on the knowledge of the design field, using computational modeling to solve the design problem and finally output the optimal result. Technical methods for AI, such as case-based reasoning (CBR), constraint satisfaction, and evolutionary algorithms, are used to support the solution of design problems (Russell and Norvig, 2016). In terms of knowledge representation frameworks, function-behavior-structure (FBS) (Gero, 1990) and structure-behavior-function (SBF) (Goel et al., 2009) models have been proposed based on distinguishing between goal-related, behavioral, and structural characteristics. de Gómez Silva Garza and Maher (1999) introduced the GENCAD system and the overall strategy of CBR. In general, design intelligence in the past focused on efficiency and functional solutions, and research on creativity has not made much progress because of the limitations of hardware and software conditions. With the advent of AI 2.0 (Pan, 2017) and the continuous advancement of deep learning, especially the emergence and evolution of the generative adversarial networks (GANs), computers began to have a new creative ability. Many works created collaboratively or independently by computers, such as visual arts, music, video, posters, webpages, garments, chairs, and cars, have shown amazing creative sparks similar to humans. Therefore, we propose the concept of design intelligence, which means AI technology that addresses issues in the design and creative process and generates creative solutions (design, content, and service). de Silva Garza (2019) briefly illustrated the relationship between creativity and design from a computational perspective. As can be seen, the middle category in Fig. 1, creative design, is where creativity and design overlap, which is also the main direction that future design intelligence research will focus on.

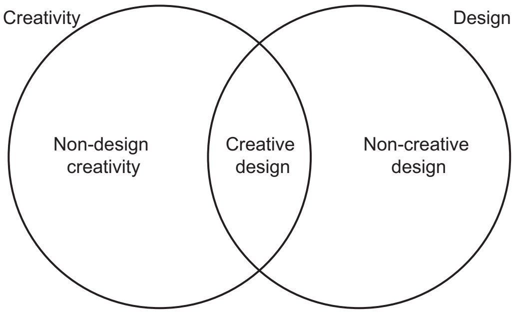

Fig. 1 Relationship between creativity and design. The borders shown between them in the figure are meant to be interpreted as being fuzzy. Reprinted from de Silva Garza (2019), Copyright 2019, with permission from Springer Nature

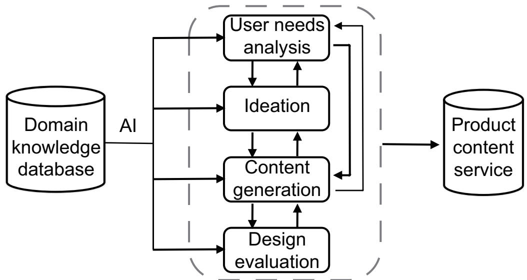

Fig. 2 Framework of design intelligence. The four main data-driven analytic components are in the center. Arrows indicate the main information flows among components

To clarify the concept of design intelligence, we propose a new systematic framework for design intelligence which includes a data module and a creative module. As can be seen from the block diagram (Fig. 2), the data module is actually a domain knowledge database, which represents information collected on the design and creative fields of interest, including information on design, style, art, technology, business, and fashion. Moreover, creative expert knowledge is stored in the database. It is the source of data that the creative module draws from. Data engineering and natural language processing (NLP) are required for creating and using this knowledge database. The creative module has four main components, i.e., user needs analysis, ideation, content generation, and design evaluation. With the help of AI, the four components interact with each other to generate product, content, and service for consumers. Components of user needs analysis deeply understand and capture the pain points of users' needs and scientifically define design issues. The ideation provides potential ideas, basic components, and source material for a high-quality design. Content generation is a key component of design intelligence. It uses mainly generative models to generate new ideas and enhance creativity, while supporting the embodiment of creativity. Design evaluation examines creative product, content, and service produced by the designer along two main dimensions (aesthetics and function). Although we use the human design processes to make design intelligence interpretable, the relationship among the various components of design intelligence is very complex. It may have different kinds of "illusions" or "blind spots" than humans do and, thus, would supplement and support human creativity (Varshney et al., 2019).

Research on design intelligence has made significant progress in recent years. Edelman et al. (1996) deemed creativity as a fundamental feature of human intelligence, and a challenge for AI. Three processes of creativity for modeling in AI were proposed, i.e., exploratory, transformational, and combinational creativity (Boden, 2009). Since then, much of the research on computational creativity was inspired by these ideas. de Silva Garza (2019) compared design computing and computational creativity in design activities and provided a complete description of the theoretical basis and actual progress of both. At the same time, researchers have developed many design intelligence algorithms in the field of design and creativity. Gatys et al. (2016a) introduced a neural algorithm of artistic style that can separate and recombine the image content and style of natural images. Zhang et al. (2017) introduced StackGAN to generate photo-realistic images based on the textual description. Isola et al. (2017) investigated conditional adversarial networks for image-to-image translation problems, such as transforming the sketch of a bag to a well-fined colorful image. These algorithms learned the rules and styles of training data in these domains to support large-scale design production. Wu et al. (2016) introduced 3D-GAN to generate 3D chairs, tables, and cars given specific object types, viewpoints, and colors. However, such algorithms have limited creative ability because they struggle to create novel designs that differ from the training data. To solve such problems, Elgammal et al. (2017) proposed a creative adversarial network for generating art, making it capable of generating creative art by maximizing deviation from established styles and minimizing deviation from art distribution. Yan et al. (2019) reviewed much of domestic and international research on the neural network in intelligent design at present. This helps researchers and designers better understand design intelligence driven by neural networks. In general, with the strong support of AI technology, design intelligence has become a powerful force driving the development of design and creative fields. However, the research results on design intelligence have not been well summarized.

The main purpose of this study is to envision design intelligence in the forthcoming era of AI 2.0.

2 AI-based user needs analysis

User needs analysis is a tool used by designers to analyze the needs and preferences of potential users. Successful designs and creative works are based on the understanding of needs and requirements of users (Maguire and Bevan, 2002). For different user research methods, the process generally involves four stages, i.e., information gathering, user needs identification, envisioning and evaluation, and requirements specification (Fig. 3). In this process, the designer continually clarifies the direction and definition of the design through empathy and the research work of users or collected data. Four tools are widely used in user needs analysis, i.e., manifestos, narrative, scenarios, and personas (Lowdermilk, 2013). A manifesto is the overall purpose and vision of design, and accordingly, a narrative is the declarative statement of the manifesto on how the design can be used. Scenarios are the imagined situations where users use or experience the design. Finally, a persona is a fictitious character that personifies a potential user group. These four aspects are extensively studied in the design literature, and personas lie at the heart of user needs analysis. In addition, the introduction of intelligent methods focuses mainly on personas, while the other three aspects are often finished by designers. Therefore, in this section, we review mainly the application of intelligent methods in personas for user needs analysis.

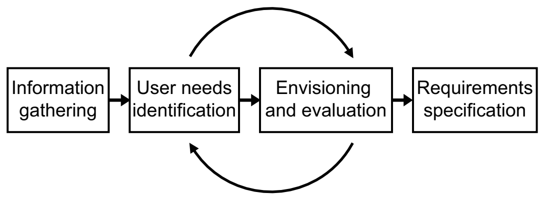

Fig. 3 General progress for user needs analysis. Reprinted from Maguire and Bevan (2002), Copyright 2002, with permission from Springer Nature

A persona, also known as user profile, is defined as a fictitious person representing an underlying customer or user group. It is often the core customer of an organization, and it can also be the desired user of a system (Cooper and Reimann, 2003). To create a persona, designers should investigate the preferences of potential users, gather their experience of other products, and summarize the typical personal information (like age and occupation). As a result, a persona provides a clear image of the target user and thus facilitates the evaluation of whether the design scheme meets the user needs. Furthermore, the process to create a persona enhances the designers' empathy for the target user group (Miaskiewicz and Kozar, 2011). Cooper (1999) introduced personas as a design technique with a goal-directed perspective. Researchers suggested that the use of personas is good for design, both supporting all components of the design (Grudin and Pruitt, 2002; Miaskiewicz and Kozar, 2011; Nielsen et al., 2015) and improving design outcomes (Grudin and Pruitt, 2002; Cooper and Reimann, 2003; Pruitt and Adlin, 2005; Salminen et al., 2019).

Personas are usually classified into three categories, i.e., personas created by designers, those summarized by intelligent methods, and those from generated data. Traditionally, a persona is formed by designers. Designers often collect a small volume of user data, which are analyzed by statistical methods or qualitative techniques such as ethnographic studies. The persona is formed in the analysis process and refined through interviews, observations, focus groups, and field studies (Matthews et al., 2012). The growth of data in recent years brings rich background information for many design tasks, and the development of sensors brings more methods to collect user behavioral data. However, it is not realistic for human designers to analyze massive data or sensory data in the traditional way. Therefore, intelligent methods are used to automate persona generation for user needs analysis.

Several automated persona generation methods have been proposed. As humans show their preference on social networks, social media data can be used. Kwak et al. (2017) developed an approach for automating the generation of personas in real time using YouTube social media data. Jansen et al. (2017) and Salminen et al. (2017) gathered data from online social media and content platforms and use data analysis tools to conduct persona research. Such procedures abstract insights on user needs from data. Behavioral data also reflect the underlying preference of users. Salminen et al. (2018b, 2019) performed case studies on participants and analyzed the behavioral data to explore possible advantages and challenges in automatic persona generation. In general, developing automated role generation systems and methods from online analytical data is of particular value to those who use online content and user data for design and content innovation.

Recently, generative models (Goodfellow et al., 2014) have made a big breakthrough in content generation. It is an intelligent method to generate new samples which are subject to the same distribution of the given data, and the details will be elaborated in Section 4. Generative models can be used in the generation of potential user portraits. IRGAN (Wang et al., 2017) and RecGAN (Bharadhwaj et al., 2018) are models that were designed to generate related pieces of user information like documents or user behaviors. In addition, Perera and Zimmermann (2019) proposed a novel GAN model (CN-GAN), which learns the complex preferences of users in a separate part of the social network and generates cross-network user preferences. These methods give related generated data and provide auxiliary data in the construction of a user profile. To the best of our knowledge, there is no application of these methods in persona generation of design tasks. However, it is only one step away from transferring these techniques from the literature to design. Note that the application of generative models here is different from automated persona generation introduced earlier. Automated persona generation is to extract personas from data, while generative models give new information about users.

There have also been a lot substantive criticisms of personas, such as taking considerable time and effort, being expensive, lacking credibility, inheriting organizational tensions and individual biases, and being inconsistent (Salminen et al., 2018a). However, it is undeniable that personas are still the most valuable and effective tool in analyzing user needs. Simultaneously, intelligence-based automatic persona generation is attracting wide attention because of its ability to solve the problem posed by scholars and practitioners. Furthermore, user needs analysis with personas still has room for exploration due to challenges in data and assessment. On one hand, automatically creating personas based on large amounts of behavioral data remains an open research issue (McGinn and Kotamraju, 2008). In addition, data sources in the field of design intelligence are complex and diverse, and a designer may jointly analyze text, visual, auditory, body movement, and even virtual reality (VR) or augmented reality (AR) data. This requires advanced cross-media intelligent methods (Yang et al., 2008) to summarize the underlying personas. On the other hand, the lack of persona assessment is also a major problem. The common method to assess the effectiveness of personas is case study (Salminen et al., 2018b, 2019), and the exploration of computational methods is a promising area for future work.

3 AI-based ideation

Ideation is a collaborative activity for generating candidate ideas in the early stage of design processes (Hartson and Pyla, 2012). This process provides potential ideas, basic components, and source material for the final design, so an effective ideation process to inspire practitioners is essential for a high-quality design. A variety of ideation methods are proposed, which can be categorized into stimuli- and interaction-based ideation. The introduction of AI algorithms improves both types of ideation methods from different perspectives. In this section, we will review ideation methods which use intelligent methods.

Stimuli are inspiration sources during a design process, and stimuli-based ideation provides designers with external information to stimulate idea generation. Stimuli can be provided in different media, such as text descriptions (Vandevenne et al., 2015), images (Goldschmidt and Smolkov, 2006), and ontology (Han et al., 2018). Stimuli are related to potential solutions in a different manner. Stimuli may be feasible (Hao et al., 2019), analogically related to the solutions (Han et al., 2018), or biologically inspired (Vandevenne et al., 2015). As can be seen, stimuli-based ideation methods aim to find inspiring stimuli as auxiliary information for designers.

Intelligent methods help retrieve existing stimuli. A retrieval method to obtain feasible and novel items was proposed by Hao et al. (2019). They introduced an evolutionary computation method to retrieve language terms in the vocabulary from 500 000 granted patents of mechanical and industrial products from 1976 to 2014. The vocabulary was embedded into a high-dimensional space by word2vec (Mikolov et al., 2013), and the terms' inspiration potential was estimated according to the measure of feasibility and novelty. Then the terms with the highest potential were selected by an evolutionary strategy. Finally, the terms were presented to the designer as the language stimuli. Another retrieval method to acquire stimuli was introduced by Vandevenne et al. (2015). They represented natural and artificial entities in the behavioral language model in plain text for convenient retrieval. The biological strategies were classified according to the K-nearest-neighbor algorithm, and thus a taxonomy was formed over the biological data. Such a taxonomy reduced the difficulty in finding related bio-inspired information on a given keyword.

Analogical relation is a typical thinking paradigm to associate concepts, and it is widely used in the ideation process (Gilon et al., 2018; Jia et al., 2020). "the Retriever" (Han et al., 2018) is a retrieval tool to present analogically related images or language terms. It is based on multiple diverse ontologies in ConceptNet (Liu and Singh, 2004), a large semantic net that represents commonsense knowledge. To retrieve potentially inspiring pictorial or verbal stimuli, it uses the proportional analogical relation $A : B :: C : X$, where the relation between A and B or C and X is among the predefined 16 ontology relations in ConceptNet. Given an input keyword as C and a specified relation between A and B, it searches for candidate visual or textual items of X and presents them to the user. Other typical analogy-based tools include Idea Inspire (Chakrabarti et al., 2017) and a functional analogy search system in Fu et al. (2015).

Besides search and retrieval, image generation is an advanced AI technique. A visual stimulus generation model was proposed in Chen et al. (2019). The generation model was trained on two image datasets of distinct concepts to combine these concepts to form a new one. The generated image visually captured partial elements of both concepts. Empirically, the synthesized images showed some diversity. As the generation model can give enormous new images, these visual stimuli can advance the ideation process significantly in terms of variety and quantity. Image generation will be introduced in detail in the next section. Generally, intelligent methods can model and retrieve existing stimuli in a fast and convenient way, or even generate new stimuli.

Interaction-based ideation helps designers obtain new ideas in group interactions, and typical interaction-based ideation methods include brainstorm (Dugosh et al., 2000) and its variants (Faste et al., 2013). In the brainstorming process, designers join together to yield wild ideas, discuss them freely and blend them to create new ones. An intelligent virtual moderator was proposed in Strohmann et al. (2017) to facilitate a brainstorming session. The virtual moderator, like a chatbot, organized the online session in a predefined procedure and intervened in the session with generated natural language text. When the participators discuss ideas freely, it encourages designers or prevents further violations of brainstorm rules. Such a chatbot will eliminate the need for a human moderator and the designer in teamwork can start a brainstorm session online conveniently. This is an application of cutting-edge natural language understanding and processing techniques in the interaction-based ideation process.

Intelligent methods play an important role in stimuli-based ideation but applications in interaction-based ideation are limited. This leaves room for further imagination and exploration. The generating technique to implement ideation was first explored by Chen et al. (2019), but they gave only random samples, and the technique lacked controllability. The application of other state-of-the-art image-to-image translation methods (Wang TC et al., 2018; Liu et al., 2019) can be explored to help translate sketches of designers to visual stimuli. This can further improve the quality of ideation. For interaction-based ideation, an intelligent moderator to dictate in an off-line brainstorm and to summarize the ideas in the dialog may be a promising direction to explore. Most advanced intelligent or learning methods are based on statistical learning, aiming to capture the distribution of existing data; thus, these methods fail to create new ideas as humans do. Therefore, obtaining auxiliary information with intelligent methods is a promising direction to explore.

Several datasets are widely used in the literature of AI-based ideation. The US patent database provides full text for over nine million patents from 1976 to the present, which is helpful in the retrieval of existing product designs. Quirky (https://quirky.com), a company that offers aids to inventors in the early stage, has more than two million product idea issues. InnoCentive (https://www.innocentive.com) offers more than 40 000 problems and the corresponding solutions on business, society, policy, science, and technology. In addition, ontology databases such as WordNet (Miller, 1995), Cyc (Niles and Pease, 2001), and ConceptNet (Liu and Singh, 2004) are used in related works. These databases provide common sense and concept definitions for stimuli retrieval and analogical reasoning.

One of the major challenges in design ideation is design fixation (Jansson and Smith, 1991). Design fixation implies that designers adhere to a certain set of ideas of concepts. In such a case, the output of ideation is quite limited. The key idea to alleviate fixation is to enhance variety and novelty in the ideation process. Specifically, in stimuli-based ideation, diverse and rich stimuli can help the designer escape from fixation (McCaffrey and Krishnamurty, 2015; Hao et al., 2019). Other challenges include functional fixedness, analogy blindness, assumption blindness, and narrow verb association (McCaffrey and Krishnamurty, 2015). On the other hand, ideation methods are generally evaluated in terms of quantity (Shah et al., 2000), quality (Shah et al., 2000), novelty (Peeters et al., 2010), and variety (Nelson et al., 2009). Future exploration of using intelligent methods in ideation can be focused on overcoming the challenges by improving ideation from those four aspects.

4 AI-based content generation

Creative content generation is the most important area of design intelligence research and its applications. With the development of AI, the intelligence-generated content (IGC) model has become a new content production model. Machines, like people, began to create and produce text, images, audio, and video. Compared with the content production based on pure manpower, the production level of the IGC can be equivalent to or even higher than that of human beings in some fields. The IGC model can reduce the cost of labor and increase production efficiency exponentially. Note that there are many ways to create creative content, and there is no fixed paradigm. However, with the current level of AI development, machine intelligence cannot completely replace the role of human beings in the production of creative content. Unlike early computer-aided design (CAD), AI is no longer a completely auxiliary design tool, and it enjoys the capability of innovation to a certain extent. The collaboration between humans and machines has become deeper and more intelligent. While the responsibilities of human beings in content production are changing, the division of labor between human beings and AI in design work is gradually blurred.

In this section, we introduce technologies used in IGC, and review different models for text-to-image synthesis, image-to-image translation, image enhancement, and content-style transfer. We also discuss some possible future research directions in related areas.

4.1 Generative models

It is impossible to create data out of thin air for algorithms. Machine learning algorithms need to understand the training data. To this end, the most promising approach is to discover the essence of the algorithms and find the best distribution to represent the training data. This kind of approach is called the generative approach. The model obtained through training is called the generative model. With a learned generative model, we can even draw samples that are not in the training set but subject to the same distribution. This sampling process is the key in content generation.

The use of deep neural networks as generative models for complex data has made great advances, including variational auto-encoders (Larsen et al., 2016) and GAN (Goodfellow et al., 2014).

A GAN is a new framework of generative models. Compared with other generative models, GANs have more realistic generation effects. In recent years, GANs have attracted extensive attention from the community and developed rapidly. In the following parts, we will briefly introduce the architecture and principles of GANs and their variants. The detailed history and review of GANs are beyond the scope of this study. We refer interested readers to Hong et al. (2019).

4.1.1 Generative adversarial networks

GANs (Goodfellow et al., 2014) contain two separate networks. One network is generator G, which receives a random noise vector $\boldsymbol{z}$ as input, and outputs a generated sample $G(\boldsymbol{z})$. The other network is the discriminator D, which receives the real sample $\boldsymbol{x}$ or the generated sample $G(\boldsymbol{z})$ as input and outputs a probability to approximate the likelihood that the input sample is a real one (Fig. 4a). The authors

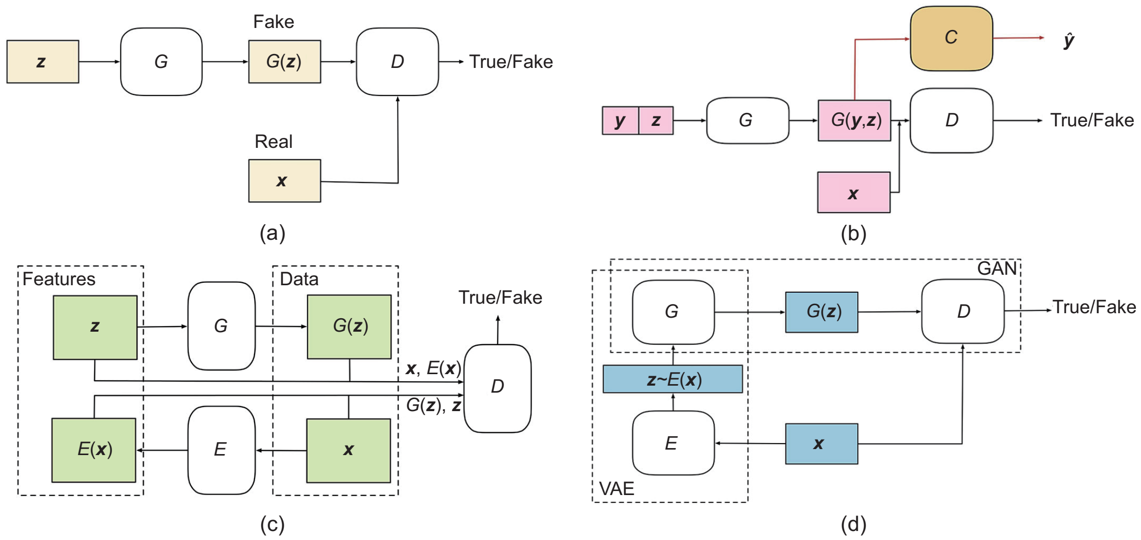

Fig. 4 Architectures of various GANs: (a) general structure of vanilla GAN, where the generator G takes a noise vector z as input and outputs a generated sample $G(z)$, while the discriminator D takes both the real and generated samples as inputs and predicts whether they are real or fake; (b) architecture of GAN with a conditional input/an auxiliary classifier, where y is the conditional input label and C is the classifier to predict the label $\hat{y}$ of generated sample $G(y, z)$; (c) architecture of GAN with auto-encoder (AE) structure, where the encoder E takes real sample x as input and produces a feature vector $E(x)$ as output; (d) architecture of GAN with variational auto-encoder (VAE) to exploit both of their benefits

also gave a two-player minimax game criterion with the function $V(G, D)$:

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

It has been proven that the solution of this zero-sum game can reach the Nash equilibrium under mild assumptions. In other words, the generator network can output spurious samples so that the discriminator cannot distinguish whether the data come from the real distribution or the generated distribution.

However, the publication of the first GAN study did not arouse great repercussions at the time. Goodfellow et al. (2014) used a fully connected layer as its building block with maxout as the activation function for both G and D. They verified only the effectiveness of the algorithm on simple datasets, such as MNIST (LeCun et al., 1998), CIFAR10 (Krizhevsky and Hinton, 2009), and other low-resolution (LR) image datasets. It was not until the publication of DCGAN (Radford et al., 2016), with the help of fully convolutional and transposed convolution layers, that GAN's advantages over other generation models began to emerge. That is, even trained on very complicated data, GANs can produce very appealing generating effects. Although GAN is still experiencing the problems of training instability and mode collapse, because of its fascinating generation effect, it has developed rapidly since 2014, i.e., in just five years. Research continues to improve on issues such as objective function (Arjovsky et al., 2017), optimization strategy (Gulrajani et al., 2017), algorithm stability (Miyato et al., 2018), and training strategies (Brock et al., 2018). Thus, GAN's generation effect is becoming more and more realistic. Specifically, the resolutions of the four images in Fig. 5 are 32×32, 128×128, 512×512, and 1024×1024, respectively. In addition to the improvement in resolution, the details of the generated images are more abundant and realistic. The strengthened GAN has become a basic tool for solving various generation problems and is widely used in the tasks of computer vision (CV) and natural language processing (NLP).

4.1.2 GAN with conditions

The vanilla GAN is completely unsupervised. Its output is dependent on the random noise fed to the generator G without other constraints. However, in the real-world generation task, it is a natural requirement to control the output. Conditional GANs (Mirza and Osindero, 2014) came into being. As shown in Fig. 4b, a conditional signal $\boldsymbol{y}$ is attached

Fig. 5 Development of GANs from 2014 to 2019: (a) GAN (Goodfellow et al., 2014); (b) DCGAN (Radford et al., 2016); (c) BigGAN (Brock et al., 2018); (d) StyleGAN (Karras et al., 2019). Their resolutions are 32 × 32, 128 × 128, 512 × 512, and 1024 × 1024, respectively. In addition to the improvement in resolution, the details of the generated images are more abundant and realistic

to the random noise $\boldsymbol{z}$, which together serve as the input to the generator G. This condition $\boldsymbol{y}$ can be the category information of an image, the attribute of an object (Mirza and Osindero, 2014), or the text description corresponding to the image we want to generate (Reed et al., 2016a). When the condition $\boldsymbol{y}$ exists, the output will be controlled by both noise and condition information, so that controllable generation can be achieved.

In addition to feeding G with condition information, we can learn from the idea of semi-supervised learning and add more side information as supervisory signals. C in Fig. 4b is an auxiliary classifier, which is the design from Odena et al. (2017). Adding an auxiliary classifier to the discriminator D can help generate a sharper sample and alleviate the mode collapse problem. This classifier can be a pre-trained model or it can be trained while the discriminator D shares most of the network parameters.

4.1.3 GAN with encoder

GAN can be generated in only one direction from the input feature vector $\boldsymbol{z}$ to the output generated sample $G(\boldsymbol{z})$. A single GAN lacks the ability to map from the sample space back to the feature space. In Donahue et al. (2016) and Dumoulin and Visin (2016), bidirectional mapping of data space and feature space was realized through the cooperation of encoder E and generator G (Fig. 4c). Encoder E inputs the real sample and outputs the feature vector while generator G inputs the feature vector and outputs the generated samples. The discriminator receives pairs of features and samples to determine whether they are from real data or generated data. By adding an encoder, GANs can not only obtain a generation model for generating high-quality samples but also have an effective reasoning mechanism and learn a meaningful feature representation.

In the models of Donahue et al. (2016) and Dumoulin and Visin (2016), E and G form a symmetrical relationship, but there is no communication and cooperation. Furthermore, on the basis of the above two works, Larsen et al. (2016) proposed to combine another well-known generation model, variational auto-encoder (VAE) (Kingma and Welling, 2013), with GAN. As shown in Fig. 4d, the part of VAE provides a priori of the feature space as an input to GAN. This combination, on one hand, retains GAN's power generation capabilities, avoiding the drawbacks of VAE generation blurring. On the other hand, the absorption of VAE can guarantee the characteristics of generating diversity to a greater extent, and alleviate the problem of mode collapse caused by the simple use of GAN. The combination of encoder or VAE and GAN is a positive method for improving the generation effect. More importantly, this combination opens up a new way to deconstruct the feature space and find more valuable feature representations. By controlling the features in the feature space, the attributes of the generated samples can be more effectively decomposed and represented. This idea lays the foundation for the subsequent work. Specific applications will be described in detail in Section 4.2.

A well-trained generative model relies on datasets with a large number of samples. Some of these datasets contain only one type of image for specific generation tasks, such as MNIST (LeCun et al., 1998), CelebA (Liu et al., 2015), flower (Nilsback and Zisserman, 2008), chairs (Aubry et al., 2014), and WikiArt (Saleh and Elgammal, 2015). At the same time, multi-class datasets allow the trained models to be robust, such as CIFAR10 (Krizhevsky and Hinton, 2009), LSUN (Yu et al., 2015), and ImageNet (Deng et al., 2009).

4.2 Approaches for intelligence-generated content

In the generation of creative content, design intelligence has produced a number of highlight applications. With the help of GAN, machines have begun to stimulate humans to generate and realize design ideas effectively. In some specialized fields, machines have reached or exceeded the human designers in generating creative content. In this subsection, we will review four types of these approaches, namely, text-to-image synthesis, image-to-image translation, image enhancement, and content-style transfer. These methods provide the fundamentals for IGC.

4.2.1 Text-to-image synthesis

When using a generative model such as GAN to implement image generation, it is often necessary to control the content of the generated image. Despite the conditional GAN model, such as cGAN (Mirza and Osindero, 2014), it has been done to generate images of the specified category. A more intuitive way is to give a text description to generate the corresponding content. Text-to-image is kind of the holy grail of an algorithm that generates realistic images from text descriptions (Fig. 6). In the field of design intelligence, text-to-image can be applied as a complete smart creative content production method to the design process. It can also be used as an independent creative module which is applied in human-computer collaborative creativity to expand the creative ability of human designers and improve the quality of final content generation.

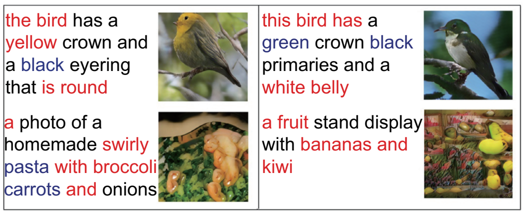

Fig. 6 Examples of text-to-image results. All pictures were generated from a text description using GANs. References to color refer to the online version of this figure. Reprinted from Xu et al. (2018), Copyright 2018, with permission from IEEE

GAN-INT-CLS (Reed et al., 2016a) is the first attempt to generate a picture from a text description using GAN. Its overall approach follows the cGAN approach, but the only difference is the use of text description embedding instead of the class tag or attribute in the original cGAN. The design of the discriminator has also been improved. In the original GAN, the real sample is taken as "true," and the sample is generated as "false." There are four sets of matching relationships in GAN-INT-CLS, {real image, right text}, {real image, wrong text}, {fake image, right text}, and {fake image, wrong text}. Discriminator D needs to recognize {real image, right text} as true, and others as false. For G, it needs to make {fake image, right text} close to {real image, right text} to fool D.

On this basis, subsequent research has been continuously improved. For example, Dash et al. (2017) combined GAN-INT-CLS and AC-GAN (Odena et al., 2017), with the aid of the auxiliary classifier for a better generation with a higher inception score (Salimans et al., 2016). GAWWN (Reed et al., 2016b) added constraints on the spatial relationship of objects in the image, so that the text description corresponds to the bounding box location of the object, making the generation more in line with the description requirements. There are also series of works such as StackGAN (Zhang et al., 2017) and StackGAN++ (Zhang et al., 2019), which used a combination of multiple generators and discriminators. Their generation process is from coarse to fine, that is, from low resolution (LR) to high resolution (HR), so that the images described in the text are generated step by step. AttnGAN (Xu et al., 2018) further extended StackGAN++ using attention mechanisms over image and text features. In addition, in this work, each sentence was embedded into a global sentence vector, and each word in the sentence was embedded into a word vector. The global sentence vector was used to generate the LR image of the first stage. Based on the image generated in the first stage, the second stage combined the word vector as a constraint of the attention layer to generate better results.

4.2.2 Image-to-image translation

Image-to-image translation is also a way of generating intelligent content, enabling a variety of creative applications. It not only helps professional designers generate content, but also improves the level of creative generation of non-professional designers (Fig. 7).

1. Supervised translation

Many image processing problems can be posed as a "translation" problem, which translates an input image into a corresponding output image. Isola et al. (2017) first introducd "image translation," which is defined as the problem of translating one possible representation of a scene into another, given sufficient training data. They built a generic framework based on conditional GAN, which learns not only the mapping from the input image to output image but also a loss function to train this mapping. This framework has become a baseline of supervised image-to-image translation. As mentioned in Section 4.1.2, the objective of conditional GAN can be expressed as

$$\begin{aligned} L_{\mathrm{cGAN}}(G, D) = & \mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}[\log D(\boldsymbol{x}, \boldsymbol{y})] \\ & + \mathop{\mathbb{E}}_{\boldsymbol{z},\boldsymbol{y}}[1 - \log D(G(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y})]. \end{aligned} \tag{2}$$

Based on the confrontation training between the generator and discriminator, Isola et al. (2017) further added an ℓ_1 constraint to the generator, requiring the output of the generator to be as close as

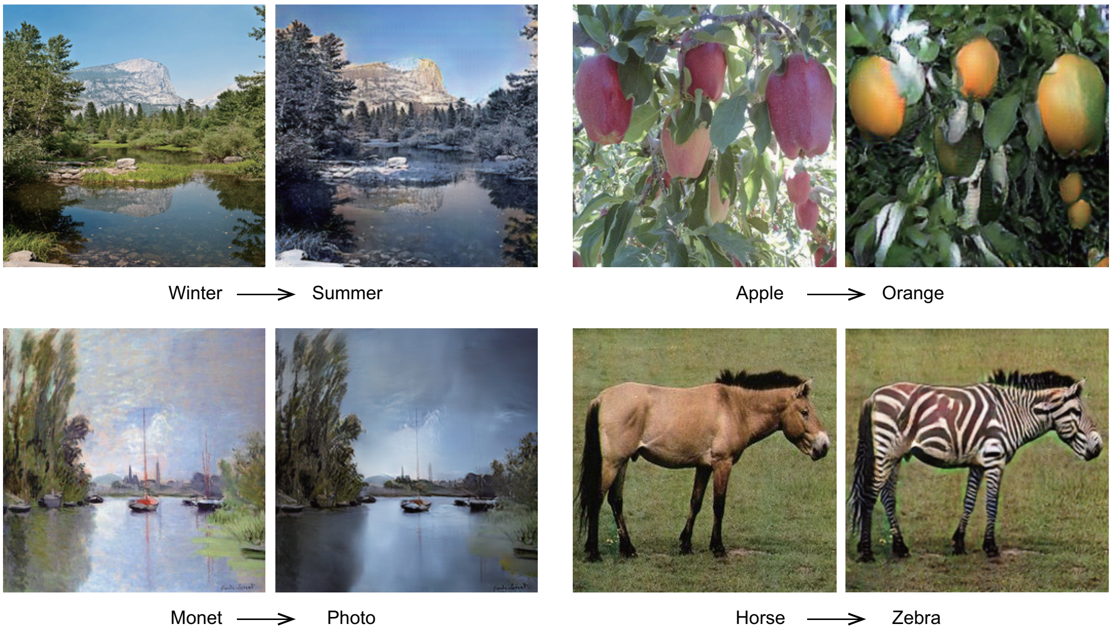

Fig. 7 Image-to-image translation examples

possible to the ground-truth output. The data accompanied by the conditional inputs were fed into the generator to output the generated samples. Both the generated images and the corresponding conditions were given to the discriminator. In addition, the authors came up with a new Markovian discriminator, which is also called PatchGAN. The PatchGAN discriminator penalizes structure at the scale of patches. That is to say, the discriminator tries to classify each $N \times N$ patch in the image as real or fake. The general principle of pix2pix (Isola et al., 2017) was as described above. Note that its effect is amazing. This general-purpose solution appears to work well on a wide variety of tasks such as photo colorization, translating satellite images into maps, turning day scenes of the same scene into night scenes, and turning sketches into rendered pictures. Subsequent research focuses on getting more high-definition effects (Wang TC et al., 2018; Park et al., 2019). Some are dedicated to generalizing to video data (Huang et al., 2017; Chan et al., 2018; Wang TC et al., 2018). Some are working to solve more specific subdomain problems (You et al., 2019).

2. Unsupervised general translation

Although the effect of pix2pix is amazing, it relies on paired training data. For example, for winter-to-summer applications, we need to provide paired data in the training: two photos of exactly the same scenery taken in winter and summer, respectively. However, in actual tasks, paired data are difficult to obtain. This requires a new learning method that learns to translate an image from a source domain X to a target domain Y in the absence of paired examples. Yi et al. (2017) and Zhu et al. (2017) solved this unsupervised domain adaption issue. Specifically, there are two generators and two discriminators. Generator G converts the image of the X field into a Y domain image, while generator F converts the Y domain image into an X domain image. The two discriminators, D_x and D_y, attempt to distinguish true and fake pictures in both domains. Here, the fake picture refers to the picture transformed from a real photo of the other domain. In addition to the confrontation in both directions, the ℓ_1 distance between the output and input is added as an additional penalty, namely

$$\begin{aligned} L_{\mathrm{cyc}}(G, F) = & \underset{\boldsymbol{x}\sim P_{\mathrm{data}}(\boldsymbol{x})}{\mathbb{E}} [\|F(G(\boldsymbol{x})) - \boldsymbol{x}\|_1] \\ & + \underset{\boldsymbol{y}\sim P_{\mathrm{data}}(\boldsymbol{y})}{\mathbb{E}} [\|G(F(\boldsymbol{y})) - \boldsymbol{y}\|_1] . \end{aligned} \tag{3}$$

3. Creative translation application

Based on the above supervised and unsupervised image-to-image framework, many interesting translation applications have been derived. These applications can effectively serve IGC. Colorization (Zhang et al., 2016) is a typical example. Given a grayscale photograph as input, these methods attack the problem of hallucinating a plausible color version of the photograph. Another example is face aging (Tang et al., 2018; Yang et al., 2018), which is the process of aesthetically rendering a given face image to present the effects of aging. A third example is talking head generation (Zakharov et al., 2019) or the deep-fake face video. These methods are dedicated to moving ones' action video to another person's face. Even using a static portrait photo or oil painting, a dynamic image or a coherent video can be generated. These technologies are now deployed in applications such as film and television production, entertainment apps, and design aids. In short, all the processes of inputting a picture and outputting the corresponding picture according to certain constraints can be regarded as a specialization of image-to-image translation.

4.2.3 Image enhancement

Image enhancement approaches in the field of design intelligence are used mainly to produce digital content and satisfy the needs of current digital content innovation.

1. Image inpainting

Inpainting is the process of recovering from missing or not needed parts of an image. For an ideal inpainting image, the whole image looks natural and viewers cannot even tell which part is reconstructed. It is a vital method for ameliorating AI-generated content. A large and growing body of literature is devoted to inpainting. There are four stages in the development of inpainting. The first stage is diffusion-based inpainting. The second one is patch-based methods. The third is pixel-by-pixel generation based on deep learning. The last is image manipulation based on inpainting (Fig. 8).

Fig. 8 Visual results of image inpainting. The masked area is shown in white

Diffusion-based methods (Bertalmio et al., 2000; Ballester et al., 2001) generate the missing holes relying on the isophote direction field. These works are suitable for only small holes or lines. Later, inpainting focuses on patch-based methods. Efros and Freeman (2001) and He and Sun (2014) filled the hole in a single concentric pass by copying patches from a patch database based on local image features. These methods usually handle texture inpainting very well, but are not so good for a real scene.

Methods based on deep learning directly use the network which is trained on a large dataset to infer missing pieces from known parts of an image. Pathak et al. (2016) used an auto-encoder architecture to solve this problem. The context image was coded by the encoder, and the decoder module decodes the code to the missing regions in the image. In addition to pixel-wise reconstruction loss, they introduced adversarial loss to improve inpainting quality. Later, Iizuka et al. (2017) improved the framework proposed by Pathak et al. (2016) with global and local context discriminators. Iizuka et al. (2017) also introduced dilated convolution into their network to enlarge the receptive field. Liu et al. (2018) used partial convolution instead of vanilla convolution to obtain a better result. Yu et al. (2018a) split inpainting to a coarse-to-fine task. Contextual attention was added into a refinement network to learn where to borrow features from the known background. Nazeri et al. (2019) divided inpainting into two stages, edge completion and image completion.

Another important direction is image editing based on inpainting. Yu et al. (2018b) used gated convolution and user-guided sketches to complete an image. The system generates missing parts under the guidance of user input sketches. Jo and Park (2019) extended color hint so that users can control not only the generated shape, but also the color.

2. Super-resolution

Single image super-resolution (SISR) is a very common requirement in the IGC process. With the increased refinement of the generated content, it is a more efficient option to adjust the resolution to fit the demand by AI in comparison to shooting or creating new materials. The easiest way to solve the SISR problem is interpolation-based methods. The most commonly used one is the bicubic interpolation (Keys, 1981), which is speedy and straightforward but suffers from accuracy shortcomings and blurring (Fig. 9b). From the LR image transforming to the HR image, new points need to be inserted between

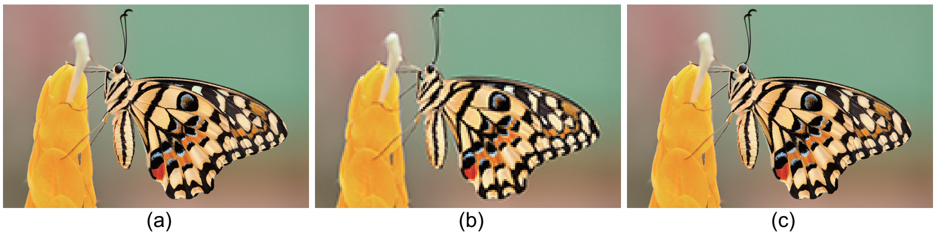

Fig. 9 Visual comparisons between different methods for super-resolution: (a) high-resolution original image; (b) super-resolution result by Keys (1981) (bicubic); (c) super-resolution result by Wang XT et al. (2018) (ESRGAN)

the original pixels. The interpolation-based method such as bicubic interpolation creates new transition pixels by calculating the mean of adjacent pixels. Such an operation inevitably results in an image that is too smooth and visually blurred. The generative model can be used to solve the problem. One of the great features of generative models is the ability to generate completely new data that match the distribution of the original data. This feature not only ensures that the transition pixels created in the SISR task are naturally connected, but also adds new pixel information so that the generated image is not too smooth. SISR is also a typical challenging ill-posed problem. For an LR input image, there can be multiple HR images with different details. The diversity of data created by the generated model also encourages the richness of the generated details. To sum up, benefiting from the strong capacity of generating, generative approaches are suitable for the SISR problem.

Among all the deep learning based methods, SRCNN (Yoon et al., 2015) can be used as a baseline. Its strategy is simple and clear. Taking the HR image as the training set, the resolution is first reduced by downsampling, and the corresponding LR image is obtained. The LR image is then sent to the generator network to be trained to output a generated HR image. The optimization goal is to minimize the error between the generated HR image and the original HR image. Through training, the final generator network will have the ability to turn any input LR image into an HR image. Based on SRCNN, there are two main directions for subsequent research. The first is the improvement of the network architecture, such as the high-learning upsampling module (Zeiler et al., 2011) or transposed conflation (Dumoulin and Visin, 2016), and the addition of a more powerful network block (Kim et al., 2016). The second is the improvement of the objective function, such as adding adversarial training (Ledig et al., 2017), replacing different error functions, and minimizing the error between generating and original HR images (Bruna et al., 2015; Zhao et al., 2016).

SISR is the most popular research topic in the community, and various methods are emerging. We refer interested readers to Yang et al. (2019).

4.2.4 Content-style transfer

1. Style calculation

Style is a very broad term that covers architecture, fashion, literature, music, art, and many other areas. Wikipedia defines style as follows: style is a manner of doing or presenting things. In AI-related research, style generally refers to the painting style, including but not limited to the genre of the painting, brushstroke, brush type, texture, composition pattern, and color distribution. However, it is still difficult for current style calculation methods to quantify even only painting styles. The most common formalization method to characterize styles is statistical information that is independent of spatial position. Although this is just a simplified way of quantifying style, it can still produce visually appealing effects. How to formalize style with aesthetic knowledge is still a research topic worth exploring, and in this study we focus mainly on reviewing already-existing methods.

Based on the current formal method of style, various style calculation ideas have been derived, such as Gram matrix (Gatys et al., 2016b), mean & variance (Huang and Belongie, 2017), whitening and coloring transforms (Li YJ et al., 2017), and linear transformations (Li et al., 2019).

2. Content-style transfer

As early as the mid-1990s, many scholars were looking at how to use computer technology to automatically convert a photo into a synthetic artwork. Among these studies, the most prominent is non-photorealistic rendering (NPR) (Gooch and Gooch, 2001; Strothotte and Schlechtweg, 2002). Content-style transfer is used as a generalization problem of texture generation. This extracts the texture from one image to synthesize an output based on the look of another image (Fig. 10).

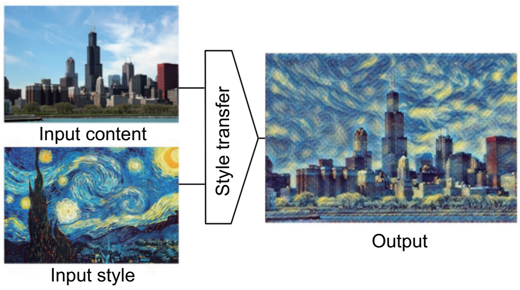

Fig. 10 An example of content-style transfer algorithm to transfer the style of a painting onto a given photograph. The style image is "Starry night" by Vincent van Gogh

The work of the formal style transfer was proposed in 2001, and called image analogies (Hertzmann et al., 2001). This method extracts the feel from a style image to synthesize an output based on the look of a content image. By constructing a pair of unstylized and stylized images, it uses the analogous transformation to obtain the image $\boldsymbol{x}$, which satisfies $A_\mathrm{c} : A_\mathrm{s} : : B_\mathrm{c} : \boldsymbol{x}$. However, it is difficult to construct datasets with pairs of images. In addition, a single analogical constraint does not have a good capture of the structure and feature information of the image, so the effect of style transfer is not satisfactory. In recent years, with the development of convolutional neural networks (CNNs), Gatys et al. (2016b) used CNN to extract features on famous paintings, and then matched the extracted features to natural photos to generate pictures with famous painting styles. Specifically, they input the content map and style map to the pre-trained CNN model, and the VGG-19 network (Simonyan and Zisserman, 2014) obtains the feature representation of the corresponding image at different layers. Next, starting with a randomly initialized generated image, it is also fed into VGG-19 to calculate its characteristic representation at different layers. For content, the generated image and content image are required to be consistent on the high-level VGG-19 feature layer (ReLU4_2). For style, it is required that the generated image and style image maintain the consistency of the summary feature statistics on multiple feature layers from low to high (ReLU1_1, ReLU2_1, ReLU3_1, ReLU4_1, and ReLU5_1). Then, by updating the model parameters, the generated image satisfies the two constraints of content and style. In this seminal study, the summary feature statistics used is the Gram matrix.

The work proposed by Gatys et al. (2016b) has been widely recognized by the community because of its attractive effects. Nevertheless, it still has two problems. First, the image-based optimization method is slow to generate. For example, it takes 2–3 min to generate a picture of 512×512. Second, the Gram-based algorithm is unstable during optimization, and requires manually tuning the parameters. There are many works addressing these two problems. For the first question, Johnson et al. (2016) used model-based optimization instead of image-based optimization, and this significantly improves the generation efficiency. Inspired by the methods of Markov random fields, Li and Wand (2016) proposed a Markovian feed-forward network through adversarial training to solve the efficiency problem using non-parametric methods. For the second question, different statistic methods have been used to simplify and replace the Gram matrix calculations, so that a stable style representation can be obtained (Luo and Tang, 2008; Huang and Belongie, 2017; Li YJ et al., 2017).

Based on model-based training, researchers continue to optimize methods to improve the generalization ability of the model. Research methods evolve from the initial per-style-per-model to the subsequent multiple-style-per-model (Li and Wand, 2016), and then to the nearest arbitrary-style-per-model (Huang and Belongie, 2017; Li YJ et al., 2017). As model efficiency and generalization capabilities continue to increase, current style transfers can reach real-time commercial application levels. A series of entertainment software and platforms using the style transfer algorithm, such as prisma, ostagram, and deep forger, have appeared on the market.

In addition, combining style transfer with other fields has led to a series of related research results, such as graffiti-style transfer (Champandard, 2016), 3D style transfer (Chen et al., 2018) supporting VR/AR, video style transfer (Chan et al., 2018), fashion style transfer (Jiang and Fu, 2017), and audio style transfer (Verma and Smith, 2018). For detailed research on style transfer, we refer readers to Jing et al. (2019).

5 AI-based design evaluation

Creative designs are usually evaluated by designers. However, the amount of AI-generated content is exploding, and it is impossible for designers to evaluate these designs, so computational design is crucial. There is no consensus among designers, psychologists, philosophers, cognitive scientists, and computer scientists on AI design evaluation. The most two essential aspects for evaluating a design are aesthetics and function. A line of studies belonging to computational aesthetics assessment are available, such as computational aesthetics (Deng et al., 2018). However, studies on automated function evaluation of design are very few. It needs high-level intelligence based on common sense inference, which is undeveloped in AI. In this section, we focus mainly on computational aesthetics and its application.

5.1 Computational aesthetics

Garabedian (1934) presented a method to quantitatively describe the beauty of images. In recent years, computer vision (CV) researchers have begun to pay attention to the direction of computational aesthetics. Computational aesthetics aim at computing a score for different kinds of designs, including images (Datta et al., 2006; Lu et al., 2015), webpages (Dou et al., 2019), logos (Zhang et al., 2017), and garments (Jia et al., 2016).

A typical pipeline of computational aesthetics system contains a feature extractor and a predictor (Fig. 11). Given an input image, the system extracts the features by a trained feature extractor and then uses the features for predicting the aesthetic score. The hand-crafted feature extractors are designed to model the artistic aspect of images to imitate the ways experts evaluate images. One advantage of this approach is that the predicted result is accountable. However, hand-crafted feature extractors and predictors cannot be trained in an end-to-end way. Moreover, the accuracy of the prediction is limited and difficult to improve. Deep feature extractors and predictors are trained in an end-to-end way, which provides the optimal solution but lacks interpretability.

Researchers explore the use of hand-crafted

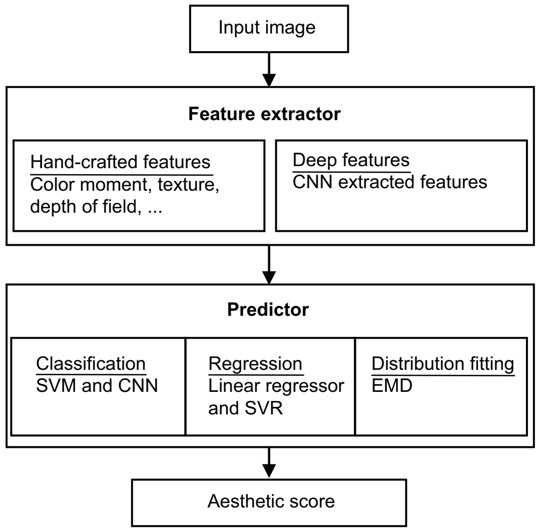

Fig. 11 A typical pipeline of a computational aesthetics system

features for aesthetic assessment (Datta et al., 2006; Murray et al., 2012; Wang et al., 2016), such as color moment, texture (Ciesielski et al., 2013), image complexity of color and texture (Rigau et al., 2008), color template (Li and Chen, 2009), and depth of field (Datta et al., 2006). However, the hand-crafted features require an extremely high degree of professional photography knowledge, and it is not possible to create all aesthetic features. Therefore, these methods have limited aesthetic precision.

Lu et al. (2015) first applied CNN to computational aesthetics and gained a promising result compared with previous hand-crafted feature methods. They also proposed a deep multi-patch aggregation network, which deals with training a set of patches extracted from one image in the neural network. These neural networks take fixed-size input. To meet this requirement, input images are usually transformed via cropping, scaling, downsampling, or padding, resulting in a gap between the aesthetics of the original images and the transformed images. To close this gap, Mai et al. (2016) developed an aesthetics neural network which includes three subnetworks with different adaptive spatial pooling size, and then used these features to train an aesthetics regression model. Ma et al. (2017) proposed another method to preserve image layout. They selected the most discriminative patches by a saliency map and pattern diversity instead of random cropping in Lu et al. (2015). Then they aggregated local attributes and global attributes to predict the aesthetic value.

Instead of solving this problem as binary classification or scoring, Li HH et al. (2017) solved computational aesthetics as a distribution fitting problem. They used earth mover's distance to measure the difference between the distribution of human opinion scores and predicted distribution by the neural network.

There are several interesting works on computational aesthetics. Li HH et al. (2017) noted that aesthetic evaluation is a subjective behavior. They built a generic aesthetics model and a residual-based personalized aesthetics model based on both aesthetic and content attributes to address personalized computational aesthetics. Inspired by image-to-text, Wang WS et al. (2018) proposed a multi-task aesthetic image reviewer, which can not only predict an aesthetic score but also generate image comments on an image. The comments partly explained why the model gave such ratings. Jia et al. (2016) worked on understanding the aesthetic effects of clothing. They built a three-level framework (i.e., visual features, image-scale space, and aesthetic words space) to bridge the gap between the visual image and aesthetic words. Webpage is the most popular kind of media and an important type of design. Dou et al. (2019) developed a deep neural network to automatically assess the aesthetics of webpage design. To overcome the shortage of webpages, they first trained the model on the Flickr style dataset (containing 80 000 images) for image style recognition, and then transferred the pre-trained network to a webpage aesthetics rating prediction task. Logo is another important type of graphic design. Zhang et al. (2017) focused on assessing the aesthetic value of logos. In this work, they collected a logo dataset with balance, contrast, harmony, and aesthetics value rated by 60 volunteers for each logo. They used professional knowledge of logo design to guide extracting hand-crafted features. A linear regression model was used to fit the relation between manually judged values and hand-crafted features. At last, the regression model was constructed for aesthetics value, model-evaluated balance, contrast, and harmony.

5.2 Computational aesthetics datasets

A high-quality dataset in machine learning is a prerequisite. Open datasets facilitate the development of computational aesthetics research. The de facto standard dataset used in aesthetic assessment is a large-scale database for aesthetic visual analysis (AVA). We summarize commonly used datasets in Table 1.

5.3 Computational aesthetics applications

The goal of computational aesthetics is not only to obtain the aesthetics value, but also to carry out further work based on it. Wang and Shen (2017) used aesthetic assessment for automatic image cropping. They designed a two-branch neural network for predicting an attention bounding box and an aesthetic value. As shown in Figs. 12a–12c, a set of candidate objects are generated based on the attention bounding box, and then the aesthetic branch selects the best candidate. Deng et al. (2018) proposed an adversarial learning-based model to learn enhancement parameters for image cropping and color enhancement parameters. They used a generator network to obtain the parameters, and a discriminator network to assess image aesthetic quality. The generator network and the pre-trained discriminator network with a new fully connected layer were trained in the way of minimax adversarial formulation of GAN. Results can be found in Figs. 12d–12f. Google's research team released the Creatism (Fang and Zhang, 2017) for artistic content creation. Creatism imitates a landscape photographer's whole workflow, from how the photographer frames the scene to post-processing

Table 1 Commonly used datasets in aesthetic assessment

Name	Amount	Description	Reference
AVA	250 000	Each image has 78–549 votes. It also contains 14 style attributes and 60 classes for a subset of images	Murray et al. (2012)
CUHK-PQ	17 690	Images are given binary labels and grouped into seven scene categories	Tang et al. (2013)
IAD	1 500 000	Images are crawled from DPChallenge and photo.net website	Lu et al. (2015)
AROD	380 000	Images are collected from Flickr, including number of views, comments, favorite list, title, and description	Schwarz et al. (2018)
AADB	10 000	Providing 11 aesthetic attributes and score for each image	Kong et al. (2016)

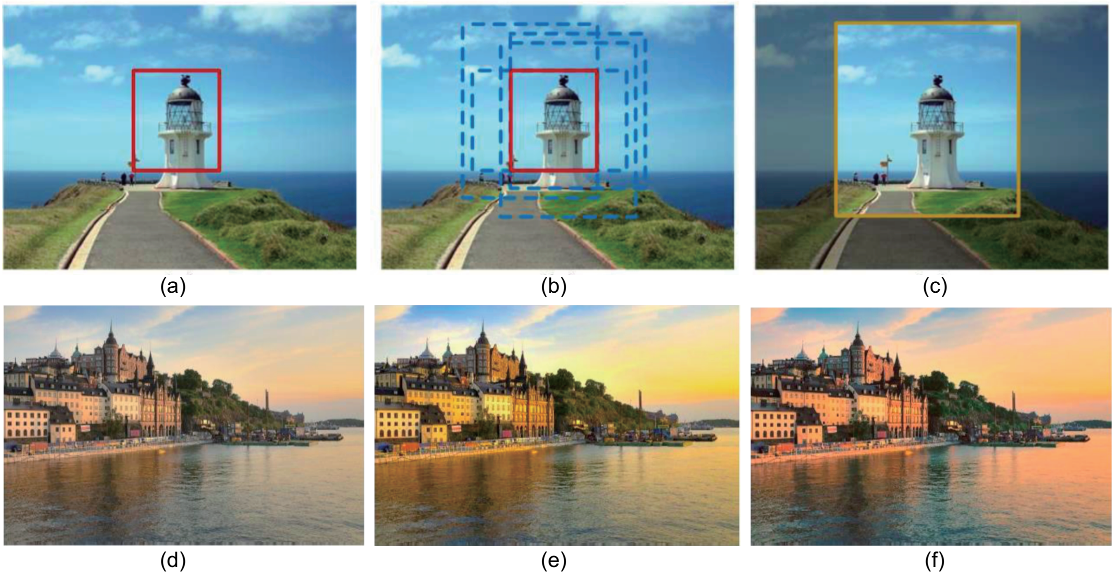

Fig. 12 Initial crop (the red rectangle) (a), cropping candidates (blue rectangles) (b), best candidate selected by the aesthetic assessment network (c), input image (d), output by the piecewise color enhancer (e), and output by the deep filtering based enhancer (f). References to color refer to the online version of this figure. (a)–(c) are reprinted from Wang and Shen (2017), Copyright 2017, with permission from IEEE, and (d)–(f) are reprinted from Deng et al. (2018), Copyright 2018, with permission from ACM

operations. In practice, they picked four independent operations, i.e., making a composite of an image from panorama images of Google street view, saturation filer, high-dynamic range (HDR) filter, and dramatic mask.

6 Open problems and challenges

Design intelligence is a new branch of AI, centered mainly on people and machines, or human-machine collaboration in the creative field. Although it is making great progress, several problems and challenges hamper the further application of design intelligence and remain open to address.

Intelligent methods require data to train models before application, and the open challenges in data are of three types. First, creativity, novelty, diversity, and some related abstract evaluations are the main topics in the design problem, especially in the ideation and evaluation stages. However, these evaluations require human experts to give labels or estimations, and thus the cost to obtain related data is high. Second, such labels and estimations are inevitably subjective, and the large variance in the data may cause difficulty in the training of models. Finally, as a result of the high cost, the amount of data can be limited, influencing the generalization ability of learning algorithms. Thus, it can be costly to train a customized model for a design problem, while the customization requirement is common in design tasks. To tackle these challenges, more open datasets with diverse attributes and high-level abstractness are necessary to allow machine learning algorithms to have a better understanding of design.

One of the challenges in algorithms is also related to the number of samples. Professional designers may have some valuable insights on the problem with only a limited number of samples, while intelligent methods often need a large number of samples to learn an applicable model. This is known as the few-shot learning problem, which is still under exploration in the literature. In addition, since a lot of abstract semantic information lies behind design problems, the few-shot learning task in design is a challenging direction for further studies.

The second challenge in algorithms is the interpretability of models and results. In ideation, the design practitioner may want to understand why the computer system gives or does not give a certain stimulus given a design problem. A generated product design may explain itself why it is generated in the current style but not others. Given an input design sketch, the evaluation program should give an aesthetic score as well as the reason why the design is not good enough and how to improve it. As intelligence methods serve as design tools, designers require the general knowledge of how a model works and the relations between inputs and outputs. However, such explanations are not provided by most of the intelligent methods, creating a barrier between designers and cutting-edge intelligent methods.

The third challenge is the difficulty in introducing prior or domain knowledge to the learning of models, especially in intelligent content generation. Most design knowledge consists of empirical rules or guidelines, which are hard to properly formalize. Finally, cross-media data bring opportunities for designers to create precise personas, provide inspiring stimuli, and generate vivid content; however, the persona generation based on cross-media requires advanced cross-media computation and data fusion methods, and the exploration of these methods remains an open challenge.

Finally, the challenge in human-computer collaboration lies at the heart of design intelligence. Intelligent methods are used as auxiliary tools for solving detailed problems in the design process, such as stimuli search or super-resolution described in previous sections. Designers may even still carry heavy and repetitive works which can be completed by intelligent algorithms, but they are unaware of such technologies. Therefore, a new mode of cooperation between designers and AI is important for the advance of design intelligence, where AI can serve as an assistant for designers. Furthermore, the assessment of human-computer collaboration in needs analysis and content generation remains an open problem.

7 Conclusions

In this study, we extensively reviewed methodologies and applications of design intelligence in the AI 2.0 era. We first summarized the concept and framework of design intelligence and then described four components of creative modules in design intelligence, i.e., AI-based user needs analysis, AI-based ideation, AI-based content generation, and AI-based design evaluation. Specifically, we introduced the state-of-the-art techniques in the four components, particularly the models and methods of intelligent content generation. In addition, we discussed open problems and challenges for future research. Since design intelligence is becoming an important branch of AI 2.0, we hope that this study can enlighten and help researchers working on the emerging hot topics of design intelligence.

Compliance with ethics guidelines

Yong-chuan TANG, Jiang-jie HUANG, Meng-ting YAO, Jia WEI, Wei LI, Yong-xing HE, and Ze-jian LI declare that they have no conflict of interest.

Acknowledgements

Figs. 5c, 5d, and 8 in this study were generated by the pre-trained models of Runway toolkit (https://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/pra2.2017.14505401021

Jansson DG, Smith SM, 1991. Design fixation. *Des Stud*, 12(1):3-11. https://doi.org/10.1016/0142-694X(91)90003-F

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:BTTJ.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, Wiescholler 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. brAInstorm: 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 Multim*, 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 spectograms. 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