Design Science

Neurocognition-inspired design with machine learning

Pan Wang [1], Shuo Wang[1], Danlin Peng[1], Liuqing Chen[1], Chao Wu[2], Zhen Wei[1], Peter Childs[1], Yike Guo[1,3] and Ling Li[4]

1 Dyson School of Design Engineering/Data Science Institute, Imperial College London, London, UK
2 School of Public Affairs, Zhejiang University, Hangzhou, China
3 Hong Kong Baptist University, Hong Kong, China
4 School of Computing, University of Kent, Canterbury, UK

Abstract

Generating designs via machine learning has been an on-going challenge in computer-aided design. Recently, deep learning methods have been applied to randomly generate images in fashion, furniture and product design. However, such deep generative methods usually require a large number of training images and human aspects are not taken into account in the design process. In this work, we seek a way to involve human cognitive factors through brain activity indicated by electroencephalographic measurements (EEG) in the generative process. We propose a neuroscience-inspired design with a machine learning method where EEG is used to capture preferred design features. Such signals are used as a condition in generative adversarial networks (GAN). First, we employ a recurrent neural network Long Short-Term Memory as an encoder to extract EEG features from raw EEG signals; this data are recorded from subjects viewing several categories of images from ImageNet. Second, we train a GAN model conditioned on the encoded EEG features to generate design images. Third, we use the model to generate design images from a subject's EEG measured brain activity. To verify our proposed generative design method, we present a case study, in which the subjects imagine the products they prefer, and the corresponding EEG signals are recorded and reconstructed by our model for evaluation. The results indicate that a generated product image with preference EEG signals gains more preference than those generated without EEG signals. Overall, we propose a neuroscience-inspired artificial intelligence design method for generating a design taking into account human preference. The method could help improve communication between designers and clients where clients might not be able to express design requests clearly.

Key words: deep learning, neurocognition-inspired design, neuromarketing, cognitive understanding, generative adversarial networks, personalized design

1. Introduction

Automatically generating a design with preferences has been an on-going challenge in the design domain. Many deep learning methods have been proposed to generate designs. For example, image style transfer (Efros & Freeman 2001; Dosovitskiy & Brox 2016; Gatys et al. 2016; Isola et al. 2017a) can be used to generate an image with the original content but different style features. Generative bionics design (Yu et al. 2018) employs an adversarial learning approach to

generate images containing both features from the design target and biological source. However, these artificial intelligence (AI) image generation methods do not consider human aspects, which means the results are generated in variations but lack human cognition input. Consideration of human aspects in a design process is vital in the design field (Cooley 2000; Carroll 2002; Vicente 2013). A person's preference for a design can be significant and intuitive, and sometimes an individual may not precisely know what their real preferences are. Therefore, being able to capture human preference (as an embodiment of design solution) and integrate the preference into the generation process may lead to a significant improvement in AI-aided generative design. Recent advancements in neuroscience, especially deep learning-based brain decoding techniques (Palazzo *et al.* 2017; Shen *et al.* 2019; Tirupattur *et al.* 2018) show potential for reconstructing a seen or imagined image from brain activities recorded by electroencephalogram (EEG), functional magnetic resonance imaging (fMRI) and Near-Infrared Spectroscopy (NIRS). This has provided the impetus to explore a novel neurocognition-inspired AI design method as presented in this paper by filling the gap between human being's brain activity and AI visual design.

In this study, we explore whether the brain signal (EEG)-informed generative method could capture human preference. An attempt has been made to add an aspect of human cognition into a deep learning design process to generate design images taking account of a person's preference. As human cognition involves many factors, to limit the scope of cognition here, only human preference for potential styles has been explored. A neuroscience-inspired AI design method is proposed, with a generative adversarial networks (GAN) (Goodfellow *et al.* 2014) framework conditioned on brain signals. This framework enables cognitive visual-related styles to be reconstructed. Figure 1 illustrates a schematic of the proposed process. The framework is composed of two stages, a model training stage and a utilizing stage. In the training stage, firstly, an image presentation experiment is used to explore the relationship between the presented image and corresponding brain signals when viewing the image. An encoder is trained to extract the features from raw EEG data. Second, a generator is trained using a GAN framework conditioned on the encoded brain signal features to reconstruct the presented image. After we
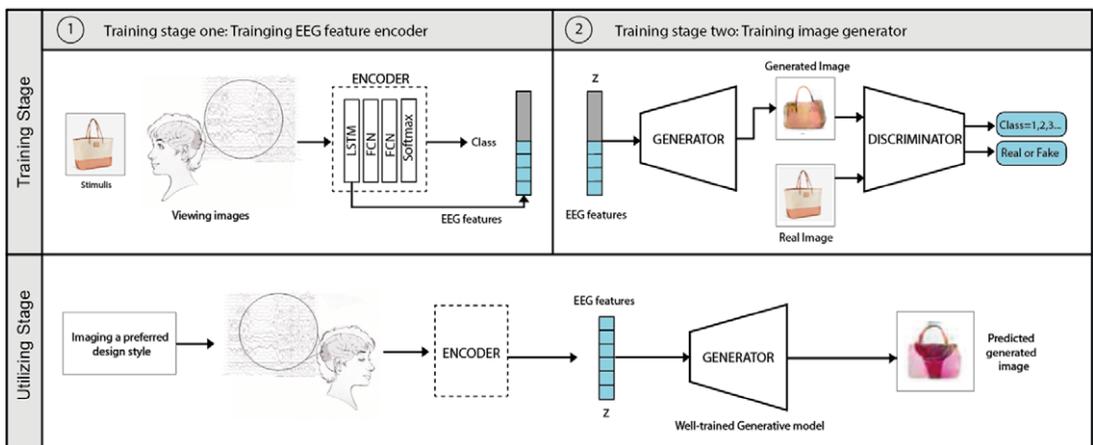

Figure 1. Overview of the process of brain signal conditioned design image generation.

obtain a fully converged model, in the utilizing stage, the trained model is then used to reconstruct the preferred design images in an imagery experiment. Given the brain signals related to the imagination of preferred design, the trained model could be used to generate images that probably contain the preferences.

Both visual examination and quantitative experiments were conducted for a case study and it was shown that the proposed neuroscience-inspired AI design method could generate some design images people preferred. The experiment successfully demonstrated that desired design images can be generated using the brain activity signals recorded when subjects are imagining a product they prefer. The neuroscience-inspired design approach could be embedded directly into other design processes with the understanding of design cognition incorporated. For example, by using this approach in fashion and product design, one could explore the cognition of possible preference on materials, patterns and shapes. Such learned brain states could contribute to better design choices. This approach could potentially also provide a new way for personalized design, for example, a personalized gift design with customization for the recipient.

The main contributions of this paper are summarized as follows.

(1) A neuroscience-inspired AI design method to generate designs taking into account the subject's preference by employing EEG measured brain activity. To verify whether the generated product images with preference EEG signals gain more preference than those generated without EEG signals.

(2) A new framework for communicating the cognitive understanding of customer requirements, enabling, for example, designers to have a visual understanding of what their clients want or their ideas through pictures not words.

2. Related work

Three scientific areas have inspired this research. In the first section, machine learning technology for generating art and design works has been reviewed, and the problem of current methods has been described. Second, to solve the current AI generative design problem, neuroscience-inspired design methods are explored. Current neuroscience methods do provide some means and potential for capturing a human brain's activities and representing design cognition. In order to transform brain signals into visual designs, the third area considered concerns using deep neural networks to classify, generate and reconstruct visual images from brain activities (EEG and fMRI). Taking inspiration from these three areas of study, a framework is proposed where brain activities are adopted as input to introduce human cognition in a GAN-based generative design process.

2.1. Deep learning for design

Regarding the purpose of this study, it is worth discussing the overlap between design science and computational creativity. Computational creativity refers to a system that exhibits behaviours that unbiased observers would deem to be creative (Colton & Wiggins 2012). Since deep learning has become more prevalent and powerful in the computer science field, systems have become more intelligent and able to complete creative tasks, such as visual art, poetry, music and design (Loughran & O'Neill 2016; Chen *et al.* 2019). By summarizing perspectives from psychology, philosophy, cognitive science and computer science as to how

creativity can be measured both in humans and in computers, Lamb *et al.* (2018) make recommendations for how to evaluate computational creativity from perspectives including person, process, product and press. This is in line with the purpose of our research, as we attempt to reveal the implicit connection between person and product by investigating whether a human's preference can be embodied in AI designs. In previous GAN based AI design research, for example, the approach for design ideation by Chen (2019), human's judgement is mainly involved in the post process of AI generation, which results in inappropriate evaluation in terms of computational creativity.

Several deep neural network approaches for image generation have been proposed recently, such as natural image generation (Brock *et al.* 2018), human face generation (Karras *et al.* 2018) and the neural style transfer model (Gatys *et al.* 2016; Johnson *et al.* 2016; Li & Wand 2016; Zhu *et al.* 2017), which can generate images which contain the content of the given image with style features from the artistic images. Isola *et al.* (2017*b*) investigated the image transfer problem which generates new images from photos and applied also to human-drawn sketches. Karras *et al.* (2018) proposed an image-to-image translation method which translated an image from a source domain X to a target domain Y (using unpaired examples). An image compositing method was proposed by Luan *et al.* (2018). This copied an element from a photo and pasted it into a painting while maintaining spatial and inter-scale statistical consistency. Dong *et al.* (2017) explored semantic image manipulation by generating realistic images from an input source and a target text description that not only match the content of the description but also maintain text-irrelevant features of the source image. Elgammal *et al.* (2017) used creative adversarial networks to automatically generate artwork by maximizing the deviation from established styles and minimizing the deviation from art distribution. In a more high-level exploration, researchers have started to apply deep learning in auto design generation. Yu *et al.* (2018) proposed DesignGAN to generate a shape-oriented bionic design that maintains the shape of the design target and combines the features from the biological source domain. Also inspired by bionic design, Duncan *et al.* (2015) presented a method for generating zoomorphic shapes by merging a man-made shape and an animal shape. One method employed by Bernhardsson (2016) generates font designs by walking through their latent space. Sbai *et al.* (2018) use a generative adversarial learning framework to generate inspirations for fashion design, creating original and compelling fashion designs to serve as an inspirational assistant.

In addition to the direct image generation technology summarized above, there are also some methods considered to improve the quality of an image, such as the image inpainting method investigated by Liu *et al.* (2018), which could fill in 'holes' in an image. This uses partial convolutions, where the convolution is masked and renormalized to be conditioned on only valid pixels. Also, the image colourization method was investigated by Nazeri *et al.* (2018), which could generate an image with plausible colours based on the adversarial learning framework. Some approaches have enabled the development of design applications, for example, Prisma (Anon n.d.), a photo editor that turns a photo to an artwork.

However, these approaches mainly focus on automatically generating new art and design images with the features from input images. A problem with this type of generative creativity is the postgeneration evaluation since the generation is completely random. How to make a selection from a large number of automatically

generated designs remains a challenge. The user is a crucial part of the traditional design process; therefore, consideration of human aspects in the design process is essential, which is missing in the current auto AI design generation approaches. How to generate a desirable design with the preference from clients is a key question in our research. To integrate human aspects into the design process, we explored neuroscience-inspired design and a deep learning framework conditioned on brain signals is described in the next two subsections.

2.2. Current neuroscience-inspired design

Noninvasive methods for measuring human brain activity that have been developed include EEG, fMRI and NIRS. EEG measures subcranial electrical signals from electrodes in contact with the scalp. Neuroscience has inspired many developments in design, such as understanding cognitive neurofeedback from clients, building and developing new products and evaluating advertizing. For example, neuroimaging has been used in understanding packaging design to help explain how packaging design confuses the consumer (Basso *et al.* 2014). Velasco *et al.* (2015) have presented an experimental research programme on evaluating the impact of different orientation of design elements in product packaging. Furthermore, to understand the consumer psychology of a brand, Plassmann *et al.* (2012) have reviewed the applications of marketing and also describe issues for future research. In a review of neuroscience-inspired design (Spence 2016), one problem of commercial neuromarketing was noted that the results provided by neuroimaging are a clear answer to a 'black-and-white' question rather than a discriminating analysis of a 'shades of grey' question. Inspired by this review, the potential of introducing neuroscience into a deep learning framework has been explored, where the machine could not only provide a response to a 'black-and-white' question but also show other potential visualizations relating to a 'shade of grey' intuition.

2.3. Brain signal conditioned deep learning framework

Machine learning methods have been applied to both EEG and fMRI to help understand visual images, for example, Bashivan *et al.* (2015) proposed an approach for learning the representation from multichannel EEG time-series. Spampinato *et al.* (2017) have developed a visual object classifier driven by human brain signals. Distinct from Spampinato *et al.* (2017) who used EEG data, Horikawa *et al.* explored object decoding from fMRI patterns (Horikawa & Kamitani 2017), which shows that the latent representation of real images (CNN1-8, HMAX1-3, GIST and SIFTbBOF) can be predicted from the fMRI signals. Both of these EEG and fMRI results show the potential of brain-based information retrieval. Furthermore, researchers have tried to generate related visual information from the decoded information of brain signals. To decode a brain image from EEG signals, Palazzo *et al.* (2017) have combined GAN with a recurrent neural network model to process EEG signals and reconstructed the viewing images of participants. Kavasidis *et al.* (2017) proposed a method for generating images using visually evoked signals recorded through EEG. In addition to EEG, fMRI signals are also widely used. Shen *et al.* (2019, 2018) have successfully demonstrated that visual images can be reconstructed from decoded fMRI signals. An unsupervised

model using variational autoencoder to model and decode fMRI activity in the visual cortex was proposed by Han *et al.* (2019). This work showed the possibility of projecting both images and corresponded fMRI signals into latent spaces.

These generative brain decoding methods provide inspiration to explore a new method for design cognitive analysis which takes into account human brain activity. However, these methods are focusing on a brain decoding approach, aiming at reconstructing the mental image of what people think about. There is a lack of exploration of generating a design image with consideration of human cognition. Previous research has explored the reconstruction of seen images but the principles could also be relevant to explore human imagination.

3. Method: human-in-the-loop design with machine learning framework

How to involve human cognition into the AI design process to generate a design considering personalized preference is the focus of the research presented here. Human preference can be captured by measuring EEG signals. The process includes two phases: a training phase to learn a generating function $G_{BD} : B \rightarrow D$ which maps the EEG measured brain activity B to the corresponding design image D, and a design phase to utilize the learned generating function and particular brain signal to generate a product involving human preference.

In the training stage, EEG signals were recorded when subjects were viewing the 'ground-truth' images of a design. Subsequently, the brain signals B are encoded into the EEG features related to the design semantic of the seen image by a Long Short-Term Memory (LSTM)-based EEG encoder. The EEG features are embedded into the GAN-based generator as the generation condition, which forces the generative model to reconstruct images D that contain the same design semantic of the original seen image. In the utilizing stage, the subjects are asked to imagine an example of a product or a design they prefer, and the measured EEG signal which may contain favoured design features of the subjects will then be encoded as the input of the trained generator. The design containing the design features that correspond to the subject's imagination will then be created by the generator. Figure 2 illustrates how
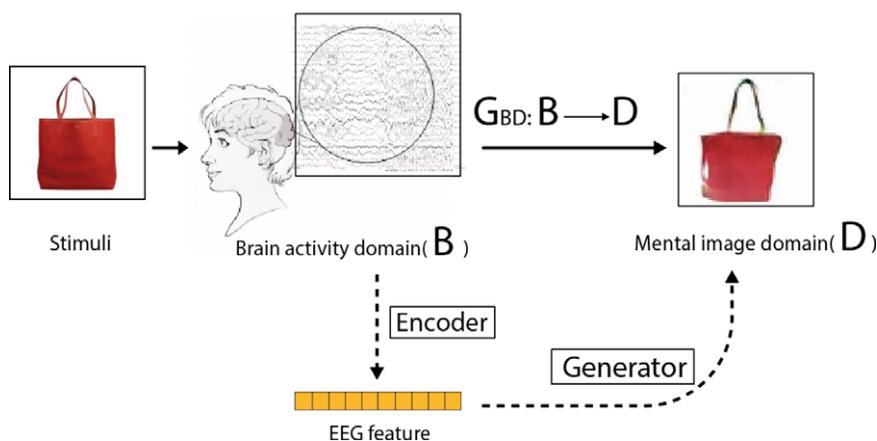

Figure 2. Training an EEG conditioned generative model.

the EEG encoder and image generator can be trained. Details about how this framework is implemented will be introduced in the following sections.

4. Experiment implementation

Details of the experiments for the model training process are presented in this section.

4.1. Participants and equipment

The EEG study included six right-handed student volunteers (three females and three males) aged between 17 and 30 years old, with normal or corrected-to-normal vision. All participants gave informed consent to take part in the EEG experiment and had considerable training in EEG experiments. Our EEG recordings were performed using an electrode cap with 64 Ag/AgCI electrodes which were mounted according to the extended international 10/20 system. An online 50-Hz notch filter was added to avoid power line signal contamination.

Signals were recorded by using a Neuroscan Synamp2 Amplifier (Scan 4.3.1; Neurosoft Labs Inc., Sterling, VA) and sampled at 1000 Hz. Eye blinks were recorded from left supra-orbital and infra-orbital electrodes, whereas the horizontal eye movement EEG was recorded from electrodes placed 15 mm laterally to the left and right external canthi. The forehead (AFZ) was used for the ground electrode, and the reference electrode was attached to the left mastoid. All electrode's impedances were maintained below 5 kΩ.

4.2. Visual stimuli

In this experiment, the stimuli consisted of five different categories of product images (handbag, headset, mug, watch and guitar) from ImageNet (Fei-Fei *et al.* 2010), which are widely recognizable and common products to help ensure the participants had similar familiarity with the stimuli; each category included 50 images. The size of the pictures was resized to 500×500 pixels and cropped to the centre of the screen.
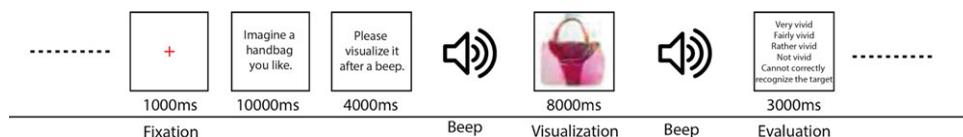
4.3. Experiment design

Two separate data collection sessions were conducted consisting of an image presentation experiment and the preference imagery session. The data collected from the image presentation session are used for model training and those collected from the preference imagery session are used in the model utilization stage. In order to ensure the quality of the data, an electrode connection checking session was added before each run. During the experiment, the subject was accommodated in a sound-attenuated and electrically shielded room and seated comfortably. The stimuli images were presented in the centre on the screen and at a fixed distance. In addition, a press button pad was provided for the subjects to give feedback during the experiment. Subjects were able to stop the experiment at any time.

In the image presentation session (Figure 3), five categories of images were presented in five runs, each run consisting of one category of 50 images and separated in five blocks, each block with 10 different images and one repeated

Figure 3. Image presentation experiment. Images were presented in the centre of the display with a central fixation cross. Ten images were shown per-block with one repeated image which required subjects to press a button when saw this image to maintain their attention.

Figure 4. Preference imagery experiment. The onset of each block was started by a central fixation cross. The 8000 ms imagery periods were signalled by auditory beeps. Before the first beep, subjects were required to visualize the preferred product for 4000 ms as the preparation of the imagery after. At the end of each block, subjects were required to evaluate the vividness of their imagination by pressing the button.

image. The subjects were required to view the images and press the button on the board when they saw the repeated images to maintain their attention. At the beginning of each block, a fixation red cross was presented in the central of the screen for 1000 ms. At the end of each run, 3000 ms were added as a rest time. In the preference imagery session (Figure 4), the subjects were required to visually imagine their preferred products with a prompt such as 'Imagine a bag you like' and follow the instructions that appear on the screen. This session consisted of five runs and each run contained 10 blocks. First, a fixation red cross was shown in the centre of the screen for 1000 ms. After this, the instruction was presented in the middle of the screen, and the subjects were asked to visualize the preferred visual look of the product. Following an audible beep, they were asked to close their eyes for an 8-s imagination period. After this, the subjects were required to evaluate the correctness and vividness of their mental imagery on a 5-point scale (Very vivid, Fairly vivid, Rather vivid, Not vivid and Cannot correctly recognize the target) by pressing the button of the box. The items evaluated as 'Cannot correctly recognize the target' are removed from the dataset. In the end, subjects were also required to draw down the image they imagined after each block. A total of 3000 ms refreshing time was added before and after each block. The subjects could stop the experiment at any time during the experiment.

After we obtained the raw data, the data were preprocessed by EEGLAB. The preprocessing procedure includes four stages, the channel selection stage, the epoch extraction and remove baseline stage, the rejecting artefacts stage and a data filtering stage. The channel selection was aimed at rejecting some bed signal channels which may influence data analysis. Then we extracted epoch according to the event markers and removed the baseline by subtracting the value of the first data from the original data. In the rejecting artefacts stage, we run both artefact correction (Zeng et al. 2013) and independent component analysis (Zeng et al. 2013) to reject the irrelevant noise artefact such as ocular artefacts and muscle

artefacts. In the end, we applied some filters to remove the unwanted frequency and to maintain meaningful waves for visual recognition and mental imagination.

4.4. Generative model

4.4.1. Training stage one – EEG feature encoder

The objective of this work is to map the stimulated brain signals into the corresponding latent representation of seen images, and thus to build a model to extract EEG features as correlated to the image features as possible.

A recurrent neural network using a LSTM (Hochreiter & Schmidhuber 1997) cell was employed to track the temporal dynamics in the EEG data which contains fundamental information for EEG activity comprehension. LSTMs are common techniques that have been developed to improve long-term dependency modelling. The brain signal is a long time sequence with very high time dependency, which means the interpretation of the brain activity is not only influenced by the previous 1-ms signal but also influenced by the brain signal long before. Therefore, the LSTM was used to learn a long-term dependency. Figure 5 illustrates the architecture of our EEG feature encoder. This is made up of a standard LSTM layer and two fully connected layers (linear combinations of input, followed by ReLU nonlinearity). At each time step t, the data of all EEG channels at time t is fed into the LSTM layer; The output of the LSTM layer at the last time step is used as the input of the fully connected layers, ReLU nonlinearity is appended after the first fully connected layer and a Softmax layer is appended after the last fully connected layer. The learning rate is initialized to 0.0001 and gradient descent is used to learn the model's parameters end-to-end. The dataset is split into three sets: 80% EEG data for training, 10% EEG data for validation, 10% EEG data for testing. Figure 6 illustrates the confusion matrix among five classes, with a total of 1500 EEG data points (300 per class), which includes 1200 data points for training, 150 data points for validation and 150 data points for testing. The overall classification accuracy on the test set which contains five classes is 71.4%. A confusion matrix summarizing the classification results is shown in Figure 6. It was observed that the error for headphone-watch was larger, possibly caused by the similar 'round and ring shape'
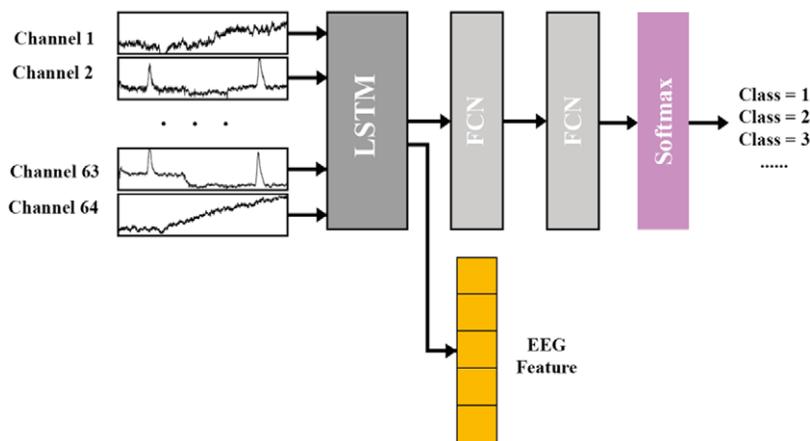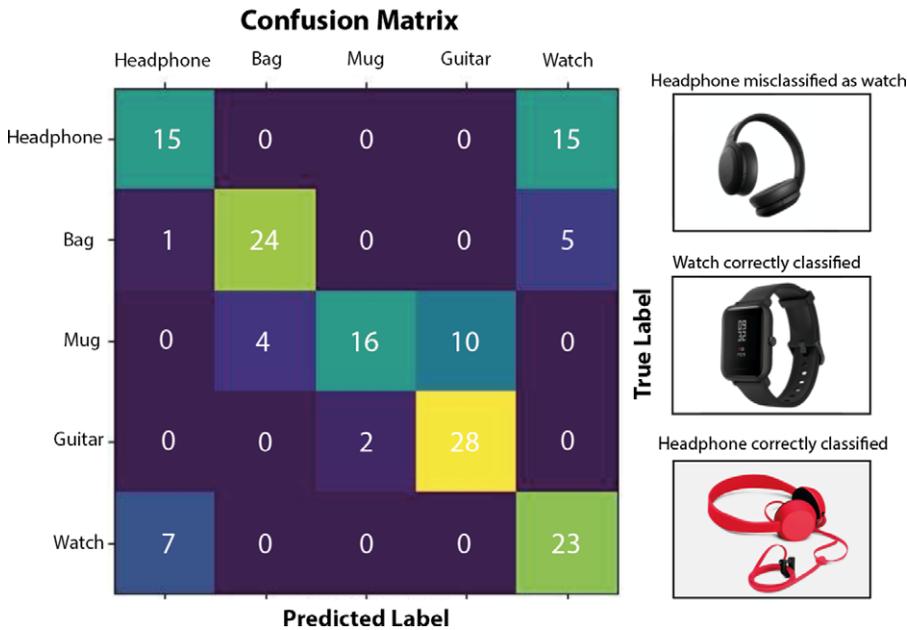

Figure 5. EEG feature encoder.

Confusion Matrix

Figure 6. Confusion matrix for the EEG encoder and examples of misclassified images. The (i,j) element in the confusion matrix represents the frequency product from the ith class, classified as jth class.

of the two objects. Examples of images that a headphone is misclassified as watch have been illustrated on the right of the confusion matrix.

4.4.2. Training stage two – generator network

The general view on model architecture is shown in Figure 7. The foundation of the generator framework is ACGAN (Odena *et al.* 2016). This generates images based on the input feature vector and also has the ability to generate images from the specific category. ACGAN consists of a generative model G and two discriminative models D_a and D_b. The generator $G(x|c)$ is trained to capture the target data distribution $p_{data}(x)$ from the condition EEG feature c of class y and noise distribution $p_z(z)$, and aims to generate images of the target class as real as possible to make the discriminator recognize the generated images are real. Whereas the discriminative model $D_a(x|y)$ is a binary class classifier which distinguishes whether a sample image belongs to the real image set. The discriminative model $D_b(x|y)$ is a multiclass classifier that identifies the image class. Both the generative and discriminative models are trained simultaneously and play against each other to minimax the log-likelihood value function $V(D,G)$.

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \in p_{data}(x)}\left[\log D_a(x|y) + \log D_b(x|y)\right]$$
$$+ \mathbb{E}_{z \in p_z(z)}\left[\log\left(1 - D_a(G(x|c)|y)\right) + \log\left(1 - D_b(G(x|c)|y)\right)\right]$$

4.4.3. Generator

The generator consists of five upsampling layers. First, inputs of the EEG representation which is the element-wise product of the 64-dimensional EEG features and a

10/19
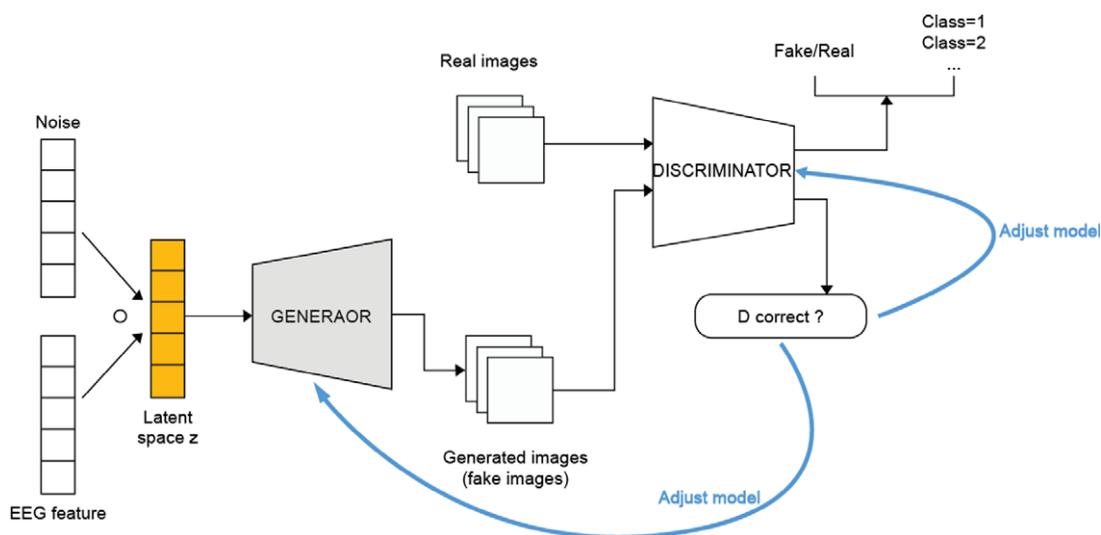
Figure 7. General view on model architecture.

random Gaussian noise have been made. The input vector is then spatially upsampled by four times by the first transposed convolutional layers and output 512 feature maps. After that, the number of feature map halves and the feature map size doubles after each remaining transposed convolutional layer. Finally, the final output has been obtained as the 64×64-pixel images with three colour channels. Batch normalization (Ioffe & Szegedy 2015) and LeakyReLU (Maas *et al.* 2013) nonlinearities have been appended after each transposed convolutional layer (Table 1).

4.4.4. Discriminator

The discriminator consists of two modules: a convolutional module used to extract the image feature and a classification module used to distinguish the generated image and identify the image category as well.

Convolutional module. The convolutional part of the discriminator is made up of 10 convolutional layers. This takes as input coloured 64×64 images. We have 64 feature maps after the first layer and the number of feature maps reaches 512 after being doubled at layers 3, 5 and 8, respectively. The feature map size starts at 64×64 and is halved after each max-pooling layer appended after the 2, 4, 7 and 10 layers and become 4×4 after the final layer. Batch normalization and LeakyReLU nonlinearities are appended after each convolutional layer.

Classification module. After the convolutional module, a $4 \times 4 \times 512$ sized data sample is obtained. The data are flattened and fed into two classifiers, a binary classifier to distinguish generated images from the real image and a multiclass classifier to identify the image category. The binary classifier consists of two fully connected layers. After the first layer, the output size is 1024 and 1 after the second layer. A ReLU activation function is appended after the first fully connected layer, and a sigmoid layer is added after the second fully connected layer. The multiclass classifier consists of three fully connected layers. The first layer reduces the number of features to 1024 and the features number remains unchanged after the second

Table 1. Hyperparameters architecture of the generator

Operation	Kernel	Strides	Padding	Feature maps	Map size
Transposed convolution, batch normalization, activation-LeakyRELU	4×4	1	0	512	4×4
Transposed convolution, batch normalization, activation-LeakyRELU	4×4	2	1	256	8×8
Transposed convolution, batch normalization, activation-LeakyRELU	4×4	2	1	128	16×16
Transposed convolution, batch normalization, activation-LeakyRELU	4×4	2	1	64	32×32
Transposed convolution, activation-tanh	4×4	2	1	3	64×64

layer. Then, the data are fed into the last layer where the number of features is reduced to the number of image categories. A ReLU activation function is appended after the first and the second layer and a Softmax layer is added after the last fully connected layer.
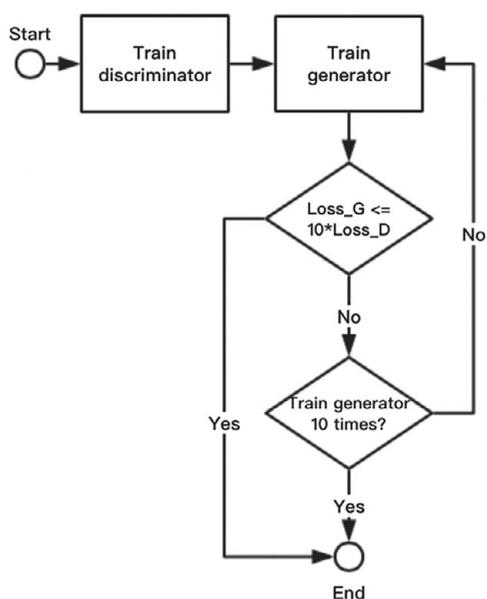
4.4.5. Training procedure

To balance the generator and discriminator, we train the generator 10 times per iteration unless the loss of the generator is less than the 10-fold loss of the discriminator. The training procedure for each epoch is shown in Figure 8. We only have 50 EEG correlated images for each class. To avoid the overfitting problem on direct training GAN on a small dataset, we train our GAN model in two stages. In the first stage, we train the GAN with the larger dataset which is gathered manually based on the ImageNet. This dataset contains 10,000 images in total (2000 images per class with a total of five classes), only including images without EEG signals. All conditioned EEG features are set to the average feature value of the class that the image belongs to. In the second stage, we retrain the GAN model for 50 more epochs on the small dataset that contains 50 EEG-available images per class, providing the correct EEG feature.

4.5. Utilizing stage – generating images with trained models with results verification

Following the method described above, the EEG data collected from the image presentation session were used to train the encoder, and then 10,000 images gathered manually from the ImageNet were used to train the generator. After we obtained a model where both the encoder and generator reached the performance mentioned above, we started to use the model in the design cases. In the model utilizing stage, the data collected from the preference imagery session were input into the model to generate the correlated mental image.

To verify whether this EEG-driven generative method could have a higher chance to capture human preference, a questionnaire survey was conducted in order to provide a proof-of-concept. The control group and intervention group of

Figure 8. Training procedure for each epoch.

generated images (with preference EEG/without EEG) were involved in this human study survey. During the survey experiment, 200 generated images were randomly selected from the results generated from our model. Among them, 100 images were selected from the results generated with preference EEG signal, and the other 100 images were selected from these generated without EEG signals. Each set of 100 images contains five classes of images and 20 images per class. Six participants who had been involved in both the image presentation session and preference imagery session evaluated these images. Participants were required to rank the images by preference level 1–10 (10 represents most preferred) from the selected images with 100 images from each group. For each trial of the evaluation experiment, the participants viewed a printed set of generated images and were required to rank the images. After this survey experiment, statistical analysis was performed for each category in two groups. The evaluation results indicated that the design images which are generated by preferred brain signals gained a higher chance to generate a preferred image.

5. Results and analysis

5.1. Results

Generated mental image results from both the image presentation experiment and the preference imagery experiment is shown in Figure 9. In the figure, the seen image results from the image presentation experiment is shown in the grey frame, which is a baseline of the work to allow for subsequent evaluation of the performance of the visual image reconstruction model. After reconstructing the seen image from the image presentation experiment, the trained model which fully converged in the training process is used to reconstruct the imagery image from the preference imagery experiment, which is shown in the red frame in Figure 9.

13/19

Figure 9. Seen image reconstruction results in the grey frame (left) and imagery preference design image in red frame (right) reconstruction results.

5.2. Visual examination and quantitative study for proof-of-concept

To verify whether the participants preferred the generative design results conditioned on preference brain signal than those without brain signal, both visual examination and quantitative studies were performed. Visual examination was used for checking whether our model has achieved a meaningful quality, that the EEG encoder maintains a good classification accuracy and the image generator reaches the image generation requirement. A quantitative study was performed for comparing whether the score of controlled with preferred EEG ranked higher than the one without EEG from the questionnaire survey. The details of the questionnaire survey are described in the previous Section 4

In the qualitative study, the generated results demonstrated that the proposed approach successfully generates different designs with multiple colour and shape features from different product classes. As mentioned in Section 4.4.1, the overall classification rate of the encoder is 71.4%. To judge the realism and diversity of the produced image, we use the Inception score (Salimans *et al.* 2016) which is commonly used to evaluate the quality of images generated by GANs. An inception model, score measures two things simultaneously. The first concerns whether the images contain meaningful items, indicated by the distribution $p(y|x)$ having low entropy. The other is whether the images have variety; the marginal $\int (y|x = G(z))dz$ have high entropy. Therefore, we obtain the final inception score as $exp\,(E_x KL(p(y|x)\|p(y)))$. An inception score of 4.9 was obtained on the generated images. This is similar to the inception score of 5.1 achieved in the study by Spampinato *et al.* (2017) while we have much fewer classes of images for training.

The quantitative study result is from the questionnaire survey, in which all participants were required to rank the image from 1 to 10 based on their preference. Figure 10 shows the mean value and its standard deviation of each category from the two groups of generated images (without EEG signal and with preferred EEG signal). The difference between the two groups is assessed using the Wilcoxon signed-rank test. All group tested had statistically significant differences ($p < .05$) in
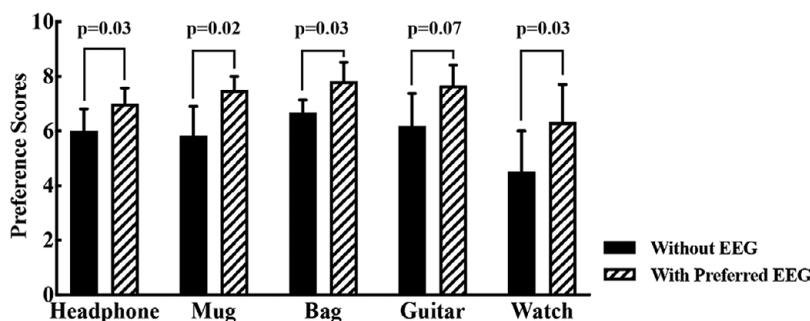
Figure 10. Human study results of the design case study.

their means with the EEG being greater than the non-EEG for all cases, except for the Guitar class ($p = .07$). Please refer to Figure 10 for more details. The analysis of the two study experiments (qualitative and quantitative) indicates that the images generated with preference EEG signals gained more preference than images generated by the generator itself. Comparing the scores from two controlled groups, the results also show that the generative model with the input of preference EEG signal had a higher chance to generate an image that people preferred. We observed that the reconstructed imagined images have a larger variety of colour and shape features than the reconstructed seen images. The preference imagery experiment results also show that these preferred products generated by the deep learning method through brain activities have combined multiple design features from various kinds of products which learned from previous designs. Also, we take the output from the LSTM layer as the EEG feature is not the final output, as we believe it may contain other features such as the shape or colour or products' style. Therefore, it may be inferred that these generated designs contain mixed colour and shape features which have been filtered by human cognition by inputting brain signal into a deep generative model.

6. Discussion and observations

The findings from this study show some potential of generating designs with human preference, which also indicate some future applications. For example, to apply in design cases, designers could have a prejudgment based on these generated images. One of the generated bags in our case study, for example, has multiple colours, from which we could predict that the user actually wants a 'very lively bag'. Similarly, with the grey bag, we could infer that an office style bag is what they might prefer. Such a discriminating analysis of 'shades of grey' design question could be applied to different design processes. Product designs dominated by the shape are more accepted than the designs dominated by function such as a guitar. This may reflect that the preference for shape is better captured by EEG signals. Further study of this hypothesis could provide additional evidence and insights into this finding.

The limitations of current results include limited dataset and limited model control. To improve the accuracy of the model, a larger dataset would need to be collected. In this work, we only train the model with six participants. In future applications, different training datasets could be involved in training according to

15/19

different application scenarios. For example, in a personalized design task, the EEG encoder could be trained by each client; to design a product for a group of people, the EEG encoder could be trained by data collected from these focused group. The generation ability of model is depended on different, to choose the right model training strategy will be the key thing for further application. In addition, another limitation is the diversity of the participants, our participants are volunteers from our research group. Mixed background participants need to be considered in future research. As one of the main contributions, a neuroscience-inspired AI design framework is proposed in this research. The design application based on this framework could be applied in many design areas, such as verifying the effectiveness of design, user or marketing research and any other user-focused design application. Furthermore, this method could also benefit to human–computer interaction, future robotics and wearable medical devices.

7. Conclusions

In this paper, a Neurocognition-inspired AI design method has been proposed with machine learning to automatically generate a design taking into account personalized information. The case study results have indicated that the images generated with preference EEG signal were more preferred than images generated by the generator itself. We are not focusing on decoding human preference in this study. Comparing with the traditional AI design generation method, adding brain signal EEG to the generation process helps machine to capture the human aspect, and had a higher chance to generate an image that people preferred. Although the proposed approach has only been applied for five product design cases, it could potentially be used in other design cases and for different design tasks such as design evaluation and branding strategy. In the research work to date, due to the limited data in the model training process, the case study only contains design semantics from these five categories. Data in additional categories can be collected in order to contain more features. The experiment indicates a new way of communicating human cognitive content. Embedding the proposed a Neurocognition-inspired AI design method into different design processes could help designers understand users' requirements and preferences more accurately. A new approach to design synthesis has been demonstrated to be possible, based on existing neurocognitive techniques. The results may help designers think beyond user cases by having direct visualization of what the user may like. The application of this neuroscience-inspired AI design method could, first, could be used as a method of user research; second, works as a primary method of user–computer interaction which could involve in any stage of the design process; third, gives a new approach to traditional design evaluation.

Acknowledgements

References

Anon, P. n.d. https://prisma-ai.com/ (downloadable on November 5th 2018).

Bashivan, P., **Rish, I.**, **Yeasin, M.**, & **Codella, N.** 2015 Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv [cs.LG] http://arxiv.org/abs/1511.06448.

Basso, F., **Robert-Demontrond, P.**, **Hayek, M.**, **Anton, J.**, **Nazarian, B.**, **Roth, M.**, & **Oullier, O.** 2014 Why people drink shampoo? Food imitating products are fooling brains and endangering consumers for marketing purposes. *PloS ONE* **9** (9), e100368.

Bernhardsson, E. 2016 *Analyzing 50k Fonts Using Deep Neural Networks.* Erik Bernhardsson https://erikbern.com/2016/01/21/analyzing-50k-fonts-using-deep-neural-networks.html (downloadable August 28th 2019).

Brock, A., **Donahue, J.** & **Simonyan, K.** 2018 Large scale GAN training for high fidelity natural image synthesis. arXiv [cs.LG] http://arxiv.org/abs/1809.11096.

Carroll, J. M. 2002 Scenarios and design cognition. In *Proceedings IEEE Joint International Conference on Requirements Engineering.* ieeexplore.ieee.org, pp. 3–5.

Cooley, M. 2000 Human-centered design. *Information Design*, pp. 59–81.

Chen, L., **Wang, P.**, **Dong, H.**, **Shi, F.**, **Han, J.**, **Guo, Y.**, **Childs, P. R. N.**, **Xiao, J.** & **Wu, C.** 2019 An artificial intelligence based data-driven approach for design ideation. *Journal of Visual Communication and Image Representation* **61**, 10–22.

Colton, S. & **Wiggins, G. A.** 2012 Computational creativity: the final frontier? *Frontiers in Artificial Intelligence and Applications* **242**, 21–26.

Dong, H., **Yu, S.**, **Wu, C.**, & **Guo, Y.** 2017 Semantic image synthesis via adversarial learning. In *2017 IEEE International Conference on Computer Vision (ICCV)*; doi:10.1109/iccv.2017.608.

Dosovitskiy, A. & **Brox, T.** 2016 Generating images with perceptual similarity metrics based on deep networks. arXiv [cs.LG] http://arxiv.org/abs/1602.02644.

Duncan, N., **Yu, L.**, **Yeung, S.**, & **Terzopoulos, D.** 2015 Zoomorphic design. *ACM Transactions on Graphics* **34** (4), 95:1–95:13.

Efros, A. A. & **Freeman, W. T.** 2001 Image quilting for texture synthesis and transfer. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques – SIGGRAPH '01*; doi:10.1145/383259.383296.

Elgammal, A., **Liu, B.**, **Elhoseiny, M.**, & **Mazzone, M.** 2017 CAN: Creative adversarial networks, generating 'art' by learning about styles and deviating from style norms. arXiv [cs.AI] http://arxiv.org/abs/1706.07068.

Fei-Fei, L., **Deng, J.** & **Li, K.** 2010 ImageNet: constructing a large-scale image database. *Journal of Vision* **9** (8) 1037–1037.

Gatys, L. A., **Ecker, A. S.** & **Bethge, M.** 2016 Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; doi:10.1109/cvpr.2016.265.

Goodfellow, I., **Pouget-Abadie, J.**, **Mehdi, M.**, **Xu, B.**, **Warde-Farley, W.**, **Ozair, S.**, **Courville, A.**, & **Bengio, Y.** 2014 Generative adversarial nets. In *Advances in Neural Information Processing Systems* (*Vol.* 27), pp. 2672–2680. Curran Associates Inc.

Han, K., **Wen, H.**, **Shi, J.**, **Lu, K.**, **Zhang, Y.**, **Fu, D.**, & **Liu, Z.** 2019 Variational autoencoder: an unsupervised model for encoding and decoding fMRI activity in visual cortex. *NeuroImage* **198**,125–136.

Hochreiter, S. & **Schmidhuber, J.** 1997 Long short-term memory. *Neural Computation* **9** (8), 1735–1780.

Horikawa, T. & **Kamitani, Y.** 2017 Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications* **8**(1), pp. 1–15.

Ioffe, S. & **Szegedy, C.** 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv [cs.LG] http://arxiv.org/abs/1502.03167.

Isola, P., **Zhu, J.**, **Zhou, T.**, & **A. Efros, A.** 2017a Image-to-image translation with conditional adversarial networks. arXiv preprint http://openaccess.thecvf.com/content_cvpr_2017_papers/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.pdf.

Isola, P., **Zhu, J.**, **Zhou, T.**, & **A. Efros, A.** 2017b Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; doi:10.1109/cvpr.2017.632.

Johnson, J., **Alahi, A.** & **Fei-Fei, L.** 2016 Perceptual losses for real-time style transfer and super-resolution. *In European conference on computer vision* (pp. 694–711). Springer, Cham.

Karras, T., **Laine, S.** & **Aila, T.** 2018 A style-based generator architecture for generative adversarial networks. arXiv [cs.NE] http://arxiv.org/abs/1812.04948.

Kavasidis, I., **Palazzo, S.**, **Spampinato, C.**, & **Giordano, D.** 2017 Brain2Image: converting brain signals into images. In *Proceedings of the 25th ACM International Conference on Multimedia.* MM '17, pp. 1809–1817. ACM.

Li, C. & **Wand, M.** 2016 Precomputed real-time texture synthesis with Markovian generative adversarial networks. *In European conference on computer vision* (pp. 702–716). Springer, Cham.

Liu, G., **Reda, F. A.**, **Shih, K.J.**, **Wang, T.**, **Tao, A.**, & **Catanzaro, B.** 2018. Image inpainting for irregular holes using partial convolutions. arXiv [cs.CV] http://arxiv.org/abs/1804.07723.

Luan, F., **Paris, S.**, **Shechtman, E.**, & **Bala, K.** 2018 Deep painterly harmonization. arXiv [cs.GR] http://arxiv.org/abs/1804.03189.

Loughran, R. & **O'Neill, M.** 2016 Generative music evaluation: why do we limit to 'human'? In *Proceedings of the 1st Conference on Computer Simulation of Musical Creativity, No. Ml*, pp. 1–16.

Lamb, C., **Brown, D. G.** & **Clarke, C. L.** 2018 Evaluating computational creativity: an interdisciplinary tutorial *ACM Computing Surveys (CSUR)*, **51** (2), pp. 1–34.

Maas, A. L., **Hannun, A. Y.** & **Ng, A. Y.** 2013 Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop on Deep Learning for Audio Speech and Language Processing.* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.693.1422 (downloadable on December 14th 2018).

Nazeri, K., **Ng, E.** & **Ebrahimi, M.** 2018 Image colorization with generative adversarial networks. arXiv [cs.CV]. http://arxiv.org/abs/1803.05400.

Odena, A., **Olah, C.** & **Shlens, J.** 2016 Conditional image synthesis with auxiliary classifier GANs. arXiv [stat.ML]. http://arxiv.org/abs/1610.09585.

Palazzo, S., **Spampinato, C.**, **Kavasidis, I.**, & **Giordano, D.** 2017 Generative adversarial networks conditioned by brain signals. In *2017 IEEE International Conference on Computer Vision (ICCV)*; doi:10.1109/iccv.2017.369.

Plassmann, H., **Ramsøy, T.Z.** & **Milosavljevic, M.** 2012 Branding the brain: a critical review and outlook. *Journal of Consumer Psychology: The Official Journal of the Society for Consumer Psychology*, **22**(1), 18–36.

Salimans, T., **Goodfellow, I.**, **Zaremba, W.**, **Cheung, V.**, **Radford, A.**, & **Chen, X.** 2016 Improved techniques for training gans. *Advances in Neural Information Processing Systems*, **29**, pp. 2234–2242.

Sbai, O., **Elhoseiny, M.**, **Bordes, A.**, **LeCun, Y.**, & **Couprie, C.** 2018 DeSIGN: Design inspiration from generative networks. arXiv [cs.LG] http://arxiv.org/abs/1804.00921.

Shen, G., **Dwivedi, K.**, **Majima, K.**, **Horikawa, T.**, & **Kamitani, Y.** 2018 End-to-end deep image reconstruction from human brain activity. *Frontiers in Computational Neuroscience*, **13**, p.21.

Shen, G., **Horikawa, T.**, **Majima, K.**, & **Kamitani, Y.** 2019. Deep image reconstruction from human brain activity. *PLoS computational biology*, **15**(1), p.e1006633.

Spampinato, C. Palazzo, S., **Kavasidis, I.**, **Giordano, D.**, **Souly, N.**, & **Shah, M.** 2017 Deep learning human mind for automated visual classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6809–6817.

Spence, C. 2016 Neuroscience-inspired design: from academic neuromarketing to commercially relevant research. *Organizational Research Methods*, **22**(1) pp. 275–298.

Tirupattur, P., **Rawat, Y. S.**, **Spampinato, C.**, & **Shah, M.** 2018 ThoughtViz: visualizing human thoughts using generative adversarial network. In *2018 ACM Multimedia Conference on Multimedia Conference*, pp. 950–958. ACM.

Velasco, C., **Woods, A.T.** & **Spence, C.** 2015 Evaluating the orientation of design elements in product packaging using an online orientation task. *Food Quality and Preference*, **46**, 151–159.

Vicente, K.J. 2013 *The Human Factor: Revolutionizing the Way People Live with Technology*. Routledge.

Yu, S., **Dong, H.**, **Wang, P.**, **Wu, C.**, & **Guo, Y.** 2018 Generative creativity: adversarial learning for bionic design. arXiv [cs.CV] http://arxiv.org/abs/1805.07615.

Zeng, H., **Song, A.**, **Yan, R.**, & **Qin, H.** 2013 EOG artifact correction from EEG recording using stationary subspace analysis and empirical mode decomposition. *Sensors* **13** (11), 14839–14859.

Zhu, J.-Y., **Park, T.**, **Isola, P.**, & **Efros, A. A.** 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*; doi:10.1109/iccv.2017.244.