Journal of Engineering Design

ISSN: 0954-4828 (Print) 1466-1837 (Online) Journal homepage: www.tandfonline.com/journals/cjen20

From thought to visual composition: a brain-driven visual blends technique for visual blending tasks

Xun Zhang, Maaike Kleinsmann, Stephen Jia Wang, Di Yan, Ziyu Wei & Pan Wang

To cite this article: Xun Zhang, Maaike Kleinsmann, Stephen Jia Wang, Di Yan, Ziyu Wei & Pan Wang (27 Mar 2026): From thought to visual composition: a brain-driven visual blends technique for visual blending tasks, Journal of Engineering Design, DOI: 10.1080/09544828.2026.2639935

To link to this article: https://doi.org/10.1080/09544828.2026.2639935

Published online: 27 Mar 2026.

JOURNAL OF ENGINEERING DESIGN
https://doi.org/10.1080/09544828.2026.2639935

From thought to visual composition: a brain-driven visual blends technique for visual blending tasks

Xun Zhang [a], Maaike Kleinsmann [b], Stephen Jia Wang [a], Di Yan [b], Ziyu Wei [b] and Pan Wang [b]

[a]School of Design, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China; [b]Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

ABSTRACT
Visual blends is a design technique that combines elements from multiple images into harmonious compositions and has been increasingly explored as a means to support early-stage ideation in engineering design. However, existing blending workflows rely heavily on manual image selection and composition, making the process difficult, time-consuming, and skill-intensive for designers. In this work, we present a proof-of-concept brain-guided visual blends technique that integrates an EEG-to-image model to simplify the image acquisition process and a local image editing model to enable automated and controllable image composition. Our EEG-to-image model employs a two-stage training strategy, combining pretraining on large-scale unlabelled EEG data with fine-tuning in an EEG-conditioned diffusion model, achieving state-of-the-art performance in reconstructing visual stimuli. To support visual blending tasks, we incorporate a local editing model (Paint-by-Example) that generates coherent blends using user-provided masks, reference images, and backgrounds. A user study with 15 participants demonstrated that the model effectively supported the creation of visual blends that aligned with users' design vision, even without artistic skills. The results suggest that brain-guided blending can serve as a early-stage ideation interface in engineering design, helping designers iterate on mental concepts before formal modelling and evaluation.

ARTICLE HISTORY
Received 21 August 2025
Accepted 27 February 2026

KEYWORDS
Visual blends; human-computer interaction (HCI); brain-computer interfaces (BCI); generative AI

1. Introduction

Visual blends refers to a graphic design technique that combines distinct visual elements from multiple images into a cohesive composition, preserving their recognizability and enabling meaningful associations (Chilton et al.,2021, Chilton, Petridis, and Agrawala,2019, Cunha, Martins, and Machado,2020). Widely employed in graphic design, visual blends help designers spark creativity, especially when ideas are challenging to articulate or communicate (Chilton et al.,2021, Wang, Tan, and Ma,2025). In the field of engineering design, this technique is particularly crucial during the early-stage conceptual ideation phase (P. Wang et al.,2025), where designers must externalise vague mental imagery into

CONTACT Pan Wang P.Wang-2@tudelft.nl Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, Netherlands

concrete visual prototypes to explore form factors and aesthetic directions (Han, Shi, and Childs,2016, Han,2018). Through visual blending process, designers search for relevant visual elements and images, experiment with diverse compositional structures, and ultimately create visual blends that serve as compelling design solutions that resonate with target audiences (Chilton et al.,2021, Cunha, Martins, and Machado,2020).

However, creating visual blends presents several challenges, primarily due to the cognitive burden of selecting suitable source images that resonate with designers, and the artistic skill required to produce coherent blends. Humans often have an 'ideal' image in mind (Afergan,2014, de la Torre-Ortiz et al.,2020), which is difficult to externalise using conventional search or sketch-based tools. Designers frequently struggle to locate suitable source images from large-scale datasets, online platforms or personal drawing that resonate with their mental concepts, especially under time and expression constraints (Qiao, Liu, and Chilton,2022). In design ideation practice, this difficulty often hinders the rapid iteration of product concepts and creates a gap between a designer's creative intent and the available digital assets (Mougenot, Bouchard, and Aoussat,2008). Furthermore, creating visual blends often requires designers to possess substantial artistic expertise, including skills in visual composition, color harmony, and spatial layout. However, existing visual blending tools typically provide only basic functions, such as keyword-based retrieval or simple image overlay (Chilton et al.,2021, Chilton, Petridis, and Agrawala,2019), and thus fall short in supporting users to integrate their creative intent or exercise fine-grained control over composition. These limitations not only make the blending process labour-intensive, but also limit its accessibility for designers without formal artistic training (Chilton, Petridis, and Agrawala,2019, Cunha, Martins, and Machado,2020).

Recent advances in Brain-Computer Interfaces (BCIs) and local image editing models present promising solutions for addressing these challenges (W. Tang et al.,2024, B. Yang et al.,2023). Several studies in neurodesign and human-computer interaction (N-HCI) (W. Tang et al.,2024) have demonstrated that BCI-based interaction methods offer multiple advantages over traditional modalities. These include providing more intuitive interaction experience (Cao et al.,2022, Galati, Schoppa, and Lu,2021, C. Yang et al.,2024), enabling access to implicit information (Afergan,2014), and capturing affective information embedded in brain activity (Vujic, Nisal, and Maes,2023, L. Wang, Huang et al.,2021, L. Wang et al.,2024). These characteristics have shown potential in simplifying the image selection process for designers, providing a more direct and easy means to externalise internal visual concepts that resonate with designers (Afergan,2014, de la Torre-Ortiz et al.,2020). At the same time, local image editing models empower users to selectively modify specific regions of an image using reference images or content prompts. This approach offers a high degree of flexibility, enabling intuitive manipulation of visual content without the need for traditional drawing or photo-editing expertise (Avrahami, Lischinski, and Fried,2022, Kim et al.,2021, B. Yang et al.,2023). By integrating these technologies, the approach has the potential to support a more efficient workflow during the design ideation stage, in which brain-driven concepts can act as an initial scaffold for subsequent idea development. Despite the potential advantages listed above, existing BCI-based generation methods often face limitations in accurately decoding the semantic content of brain signals, frequently producing blurry or noisy outputs (Kavasidis et al.,2017, Singh et al.,2023, Tirupattur et al.,2018). Furthermore, current visual blending tools (Chilton et al.,2021) have not yet utilised the potential of advanced local editing models to support simple, user-driven composition.

In this study, we introduce a brain-guided visual blends technique that integrates an EEG-to-image model with an advanced local image editing model to enable quick image acquisition and automated visual blending task. This technique is designed to accelerate design ideation iteration and reduce the cognitive load associated with manual image composition in industrial and product design. Our EEG-to-image model follows a two-stage training strategy: the EEG encoder is first pre-trained on a large-scale unlabelled EEG dataset (Aristimunha et al.,2023, Jayaram and Barachant,2018) to better capture the intrinsic features of EEG signals, and then fine-tuned within an EEG-conditioned diffusion generation model (Rombach et al.,2022) using an EEG-image paired dataset (Zhu et al.,2024) to reconstruct high-quality visual stimuli. Architectural refinements and loss function enhancements further improve the semantic accuracy and visual quality of the generated images. To support the visual blending tasks, we incorporate Paint-by-Example (B. Yang et al.,2023), a local image editing model that produces visual blends conditioned on a reference image, a target mask region, and a source background. This approach provides an intuitive and controllable workflow to create visual blends, enabling users to generate coherent visual blends by specifying controls, without relying on manual drawing skills.

We evaluate the brain-to-image decoding model through benchmarking and assess the local image editing module via a user study. We use perceptual EEG data, rather than imagery EEG data, for model training primarily due to the absence of open-access, high-quality EEG datasets for mental imagery. Additionally, several studies (Canales-Johnson et al.,2021, Dijkstra, Bosch, and van Gerven,2019, Wilson et al.,2024, Xie, Kaiser, and Cichy,2020) have demonstrated that visual imagery and perception rely on largely overlapping neural substrates, although perception generally evokes stronger signals (Cichy, Heinzle, and Haynes,2012, Pearson,2019). Based on this evidence, we adopt perceptual EEG from the ImageNet-EEG dataset (Kavasidis et al.,2017) as a principled approximation when imagery EEG is unavailable. Our EEG-to-image model achieves state-of-the-art (SOTA) performance on multiple evaluation metrics for reconstructed images, confirming its effectiveness in generating visually accurate representations of users' visual perception. In a user study with 15 participants, we explore the use of Paint-by-Example (B. Yang et al.,2023) for visual blending tasks that combine brain-decoded images–generated from test EEG data–with products such as T-shirts, mugs, and canvas bags to create novel design outcomes. As shown in Figure 1, the local editing model produced high-quality visual blends from user-provided masks and allowed participants to iteratively refine and explore creative variations. Participants reported that the system helped them better express their design goals and that the generated results closely aligned with their envisioned outcomes–even without prior artistic experience. In engineering design workflows, the brain-guided visual blending is a pre-CAD ideation interface that helps designers externalise and iterate on mental concepts before formal modelling and evaluation. Although the current system operates offline, it enables early-stage exploration of neuroCHI (Tanaka et al.,2024) and serves as a critical step in the progression from non-BCI-driven human-computer interaction (HCI) design systems toward real-time imagery BCI-enabled CHI systems. It offers foundational insights for advancing practical neurodesign (W. Tang et al.,2024) tools as portable EEG hardware, dataset availability, and cross-subject decoding models continue to improve.

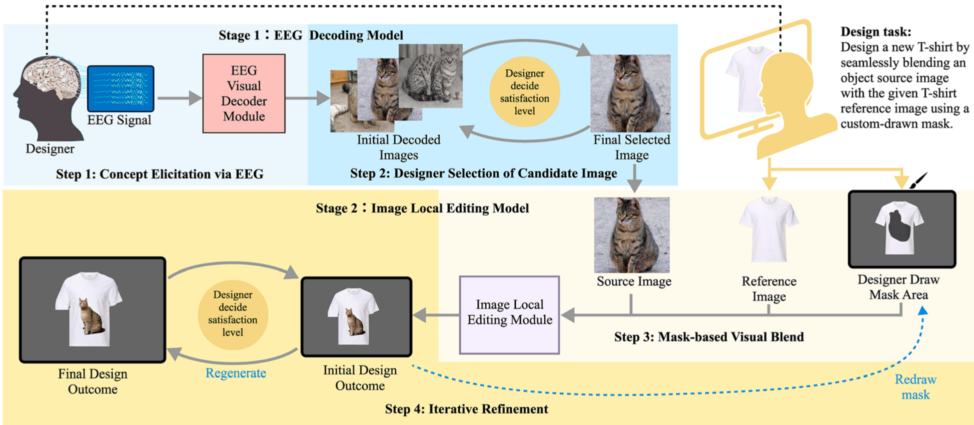

Figure 1. Overview of the brain-driven visual blends technique presented as an explicit design workflow. The process consists of two stages with clearly labelled design steps. *Stage 1 EEG Decoding consists of two steps.* Step 1 involves the EEG visual decoder generating multiple image candidates from brain signals, and Step 2 requires the designer to evaluate and select a candidate that resonates with their design intent, with the option to regenerate results until a satisfactory concept emerges. *Stage 2 Visual Blend Generation also consists of two steps.* Step 3 composes the selected image with a reference product image using a designer-drawn mask through a local editing model, and Step 4 allows designers to iteratively redraw the mask or regenerate blend outcomes, forming a mixed initiative loop of creation, evaluation, and refinement. The process shown reflects our current implementation, in which decoded images are generated offline from an EEG test set. The black dashed line illustrates the envisioned online workflow, in which EEG signals can be captured and decoded in real time to reflect the designer's envisioned design concepts. The designer can adjust their design vision to generate new image semantics and concepts. These can then be seamlessly passed into the refinement stage to create visual blends.

The primary contributions of this work are:

- A novel EEG-to-image model that decodes users' EEG signals into corresponding visual stimuli, enhancing brain decoding quality and improving semantic alignment with user-desired visual outputs.
- A brain-guided visual blends technique that combines an EEG-to-image model with an advanced image local editing framework to provide a more intuitive experience and reduce the skill barriers involved in creating visual blends to support early-stage ideation in engineering design workflows.

2. Related work

2.1. Visual blends

Visual blends involves the creation of new visual compositions by combining elements from two or more source images, typically encompassing image selection and compositional synthesis (L. Chen, Xiao et al.,2025, L. Chen, Zhang et al.,2024, Chilton, Petridis, and Agrawala,2019, P. Wang et al.,2025). Early research explored visual blends as a means of enhancing conceptual creativity. Confalonieri et al. (2015) employed computational argumentation to evaluate visual blends in icon design, while Karimi et al. (2018) developed a model to support analogical reasoning. Xiao and Linkola (2015) and Ha and Eck (2017) introduced semi-automatic and generative techniques for combining visual elements to express abstract concepts. More recent work focuses on interactive systems that support visual blending tasks. ICONATE (Zhao et al.,2020) automates icon creation from textual prompts. Emojinating (Cunha, Martins, and Machado,2020) evaluates visual concept representation through emoji blending. Chilton et al. introduced VisiBlends (Chilton, Petridis, and Agrawala,2019) and VisiFit (Chilton et al.,2021), which combine computational techniques with human microtasks to assist users in iteratively generating and refining blends. Although these systems support creativity and accessibility, they still rely heavily on manual image selection or iterative composition processes. To address these limitations, we propose a brain-guided visual blends technique that simplifies image acquisition through BCI and create visual blends using image local editing generative models.

2.2. Brain computer interface and human computer interaction

Neurodesign enhances our understanding of how brain activity rules in human experience can guide design (Bridger,2017). Human-computer interaction (HCI) increasingly intersects with neurodesign, giving rise to the specialised field of Neurodesign and Human-Computer Interaction (N-HCI) (W. Tang et al.,2024), which leverages neuroscience principles to enhance HCI system efficiency. Existing research in N-HCI primarily addresses four principal themes: advancing BCI development (Bocquelet et al.,2016, Crawford and Gilbert,2019, Kosti et al.,2018, L. Wang et al.,2024, Weiskopf et al.,2007, Williamson et al.,2009), practical BCI applications (Vasiljevic and Cunha de Miranda,2024, Aranyi, Charles, and Cavazza,2015, C. Chen et al.,2019, Kumar et al.,2022, Paulo, Pires, and Nunes,2021, Riccio et al.,2022, Vujic, Nisal, and Maes,2023, L. Wang, Huang et al.,2021, Wolpaw et al.,2003, S.-W. Yang et al.,2013, Zhang et al.,2025), optimisation techniques for BCIs (Duan et al.,2021, Feng et al.,2022, Markovinović et al.,2022), and the broader integration of neurodesign and HCI (Brocke, Riedl, and Léger,2013, Cao et al.,2022, Chew, Teo, and Mountstephens,2016, Galati, Schoppa, and Lu,2021, Siqueira et al.,2023). Some studies specifically aim to enhance interaction efficiency and intuitiveness in HCI systems. For example, Cao et al. (2022) leverage EEG signals to facilitate a more intuitive and natural object selection process in CAD, addressing the Midas touch problem inherent in gaze-based interactions. Galati, Schoppa, and Lu (2021) employ functional near-infrared spectroscopy (fNIRS) and AR to measure internal and external user behaviours, providing valuable insights into optimising interface design for improved cognitive alignment and efficiency. Chew, Teo, and Mountstephens (2016) investigate aesthetic preference recognition through EEG, revealing the capability to decode individual aesthetic preferences for 3D shapes.

These studies demonstrate that BCI effectively decodes human intentions, motivating our exploration of human visual perception decoding tasks to efficiently retrieve images aligned with the user's idea. Spampinato et al. (2017) attempted to use deep neural networks, such as CNN and RNN, to learn latent EEG features from brain signals, achieving an average accuracy of 83% for a four-class classification task. This paper also proposed a visual evoked EEG signal dataset, which has become the baseline dataset for many EEG-visual classification papers (Bagchi and Bathula,2022, Lawhern et al.,2018, Mukherjee et al.,2019, Song et al.,2022, Zheng and Chen,2021, Zheng et al.,2020). Kavasidis et al. (2017) combined Long Short-Term Memory (LSTM) and Generative Adversarial Network (GAN) (Bing et al.,2025, W.-Y. Tang et al.,2025) to reconstruct visual stimuli, initially demonstrating meaningful feature extraction from brain signals and using these features as conditional vectors to guide GAN outputs. Subsequent researches (Ahmed et al.,2021, Deng et al.,2023, Fares, Zhong, and Jiang,2020, Jiao et al.,2019, Khare et al.,2022, Tirupattur et al.,2018, H. Zeng, Xia, Tao et al.,2023) employed different EEG encoders and improved GAN models to achieve higher-quality visual stimulus restoration. With the advancement of diffusion generation models, researchers (Bai et al.,2023, Lan et al.,2023, H. Zeng, Xia, Qian et al.,2023) have begun using pre-trained diffusion models to replace GANs, achieving better restoration accuracy and quality. Our study employs EEG decoding models to efficiently capture users' ideal concept from designers' brain activity, simplifying the image acquisition process in current visual blending workflows.

2.3. Generative local editing model

Image-local editing performs controlled modifications on multiple images, thereby achieving variations in image style, structure, and details. Kim et al. (2021) proposed StyleMapGAN, which introduces spatially variant modulation in the latent space, enhancing real image embedding accuracy while preserving image-local editing capabilities. Y. Shi et al. (2022) enhanced fine-grained control over image synthesis and editing by modelling local semantic parts separately with corresponding latent codes, offering improved disentanglement and compositionality compared to traditional StyleGANs (T. Karras, Aittala et al.,2021, Karras, Laine, and Aila,2019, T. Karras, Laine et al.,2020). Recently, advancements in diffusion models (Rombach et al.,2022) have introduced novel approaches to image-local editing tasks. Avrahami, Lischinski, and Fried (2022) introduced BlendDiffusion, an image editing technique that integrates CLIP-guided latent codes with the noisy background of the input image during each denoising step. The mixed feature is fed into subsequent denoising steps, resulting in seamless integration of the edited area with the untouched regions of the original image. P. Li et al. (2023) developed a semantic-based layer control image editing method that facilitates non-rigid editing and attribute modification of specific subjects while maintaining content consistency. B. Yang et al. (2023) proposed the paint-by-example model, which abandons text-guided image editing in favor of using a mask image to blend parts of a reference image into the source image, achieving example-based image editing. Relying on generative local editing models eliminates the need for users to undergo complex iterative processes to achieve visual blends (Chilton et al.,2021, Chilton, Petridis, and Agrawala,2019), providing a naturally more efficient and easy solution. Consequently, our study employs the Paint by Example (B. Yang et al.,2023) to efficiently create visual blends. By incorporating BCI techniques to streamline the image acquisition process and utilising a local image editing model to generate visual blends, we propose a brain-guided visual blends technique that lowers both the effort and the level of expertise needed to produce visual blends.

3. Method and implementations

This work proposes a brain-guided visual blending technique consisting of two stages: EEG-based image decoding and designer-controlled visual blending. In the first stage, an EEG encoder is pre-trained on large-scale unlabelled EEG recordings to learn robust neural

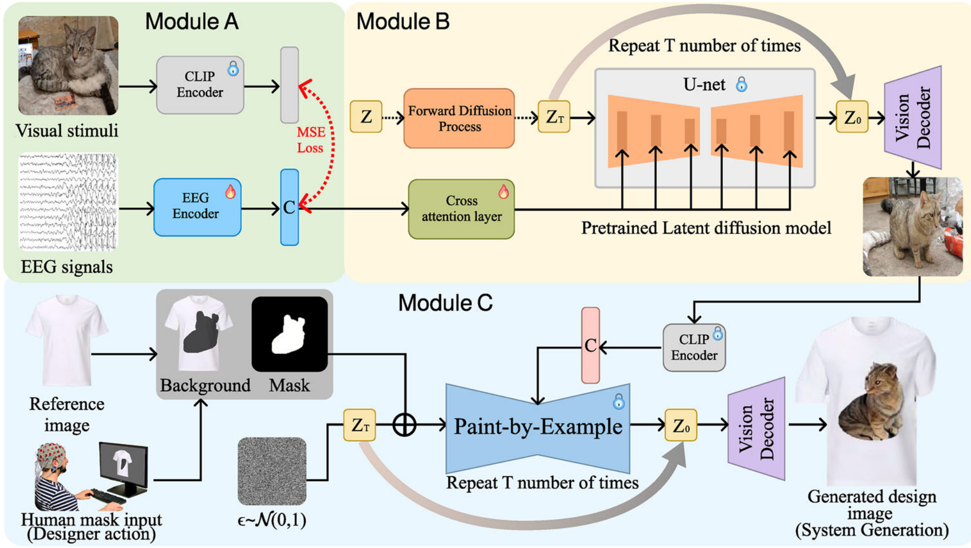

Figure 2. The overall structure of the brain-driven visual blends technique consists of three modules. Module A and Module B support Step 1 Concept Elicitation via EEG by first extracting informative features from EEG signals through a trained EEG encoder, then using a conditional diffusion model to reconstruct visual candidates from EEG features, enabling designers to evaluate and select images that resonate with their intent. This model is trained on EEG-image pairs by updating the weights of the cross-attention layers while keeping the rest of the model weights frozen. Module C realises Steps 3 Mask-based Composition, providing a local editing interface in which the designer supplies a reference image and a mask area to compose the visual blend. This architecture interleaves human actions with model operations, forming a mixed-initiative loop that connects brain-based concept elicitation with designer-driven composition.

representations, and a diffusion-based generation model is fine-tuned on EEG-image pair data to translate brain activity into visual candidates. In the second stage, a local image editing model enables designers to blend the EEG-generated images with a reference product image through an interactive masking interface. The overall workflow is illustrated in Figure 2. Detailed technical derivations and training configurations are provided in Appendix 1.

3.1. EEG representation learning

EEG offers a non-invasive and temporally sensitive channel for capturing human responses to visual stimuli, yet the signals are noisy and highly variable across individuals and contexts. To obtain stable features, we adopt a self-supervised masked autoencoding strategy. The model learns to reconstruct partially masked EEG segments, which encourages the encoder to capture contextual dependencies rather than superficial patterns. After pre-training, the encoder is used as a feature extractor for the subsequent image generation model. Detailed information is shown in Appendix A.1.

3.2. EEG to image generation

The latent diffusion generation model is adapted to use EEG representations as its conditional input. The original text encoder is replaced with the pretrained EEG encoder, and the model is fine-tuned on an EEG-image pair dataset to learn the correspondence between neural activity and visual semantics. During inference, an EEG signal is transformed into a latent representation that guides the diffusion process to generate multiple candidate images reflecting the designer's implicit intent. Detailed information is shown in Appendix A.2.

3.3. Designer controlled visual blending

The generated images serve as creative references that designers can further shape. We integrate a local image editing model that allows designers to blend an EEG-generated image with a product reference image by drawing a mask to specify editable regions. Designers can iteratively redraw the mask, select alternative candidates, or regenerate images, forming a loop of creation, evaluation, and refinement workflow. This design positions AI as a supportive partner that augments human creativity. Detailed information is shown in Appendix A.3.

3.4. Datasets and training overview

For EEG representation learning, we utilised large-scale public datasets accessed through the MOABB platform, covering motor imagery, RSVP, and SSVEP paradigms from more than 400 participants. For EEG-to-image alignment, we employed the ImageNet-EEG dataset, which contains 120,000 EEG-image pairs collected under the RSVP paradigm. All EEG data were standardised through unified preprocessing to ensure consistency across datasets. The diffusion model was initialised from a publicly available checkpoint and fine-tuned with the EEG encoder. Detailed preprocessing steps, network configurations, and hyperparameters are reported in Appendix 2.

4. User study

4.1. Participants and settings

We recruited 15 participants with a design background from university bulletin board, ensuring our findings are relevant to the design community. All participants, aged 20 to 28, had experience in graphic design and familiarity with generative models like ChatGPT or Midjourney, allowing them to quickly adapt to the toolkit. The group included 6 males and 9 females, all cognitively healthy. Amoung them, 11 participants use text-to-image platforms daily for design inspiration and were classified as frequent users, while the remaining 4 used them 2-3 times per week and were considered occasional users. Detailed information about the participants is provided in Table 1. During the experiment, participants completed three sessions, each with two design tasks, using a web interface on their personal computers. The interface, built with the Gradio framework, provided a consistent experience with backend processes running on the same server and using identical model checkpoints to ensure fairness.

4.2. Experiment procedure

To examine how users interact with the brain-decoded images and the image local editing model in our brain-guided visual blends technique, we conducted a user study centered

Table 1. The characteristics of the participant in our user study.

Participant	Gender	Age	Frequency of AI usage
P1	Male	21	Frequently
P2	Female	24	Frequently
P3	Female	22	Occasionally
P4	Male	26	Frequently
P5	Female	25	Frequently
P6	Male	28	Frequently
P7	Female	23	Frequently
P8	Male	20	Occasionally
P9	Female	27	Frequently
P10	Female	22	Frequently
P11	Male	25	Occasionally
P12	Female	28	Frequently
P13	Male	24	Frequently
P14	Female	26	Occasionally
P15	Female	21	Frequently

on visual blends. The primary task involved creating visual blends by adding objects such as animals, everyday items, or vehicles to backgrounds including T-shirts, mugs, hats, sweaters, and bags. This setup was consistent with the intended purpose of the technique and ensured that participants remained actively engaged in the design process. Participants were provided with a basic theme to guide their creativity, reflecting realistic design conditions. The brain signal-decoded images were grouped into three categories: animals, everyday items, and vehicles, corresponding to the image types in the EEG dataset. The design tasks were organised into three scenarios: (1) designing a commemorative item featuring an animal image, (2) designing a commemorative item with an everyday object, and (3) designing a commemorative item with a vehicle image, each applied to a selected background item.

We adopted a practical offline approach due to the limited portability of EEG devices and the need for personalised brain signal decoding, which made real-time EEG data collection impractical for the user study. Instead, we generated a large set of images from test data prior to the experiment. Participants selected the images that best matched their design vision, simulating online scenario where we extract the design vision from designers' brain activity. This approach allowed us to leverage brain signals for creative design while circumventing technical constraints.

The experimental procedure was structured into several phases: an introduction phase, a system trial phase, three experimental phases, and an interview phase. The entire experiment lasted approximately 60 min.

(1) *Introduction Phase (5 min):* During this phase, participants were introduced to the experiment's background, objectives, and tasks. This introduction was essential for helping participants understand the experiment's goals and their role in it.
(2) *System Trial Phase (10 min):* In this phase, participants were provided with a user manual detailing the steps for using the toolkit and the necessary precautions. Participants were given access to a trial web interface where they could practice using the toolkit following the steps in the manual. They were also allowed to ask questions to ensure they fully understood how to use the toolkit. The images used

during the system trial were different from those in the formal experiment to avoid bias.

(3) *Experimental Sessions (6–8 min per session):* The core of the experiment consisted of three sessions, each lasting 6 to 8 min, with a 2-min break in between. In each session, participants select an image from the provided brain decoded image set that resonate with their design ideas and combined them with a background image to create a commemorative item. Participants were provided with five background images (T-shirt, mug, hat, sweater, and canvas bag). The object images were pre-generated from the test set and categorised into three groups: animals, everyday objects, and vehicles. The toolkit displayed 10 random images at a time, and participants could click a 'Random' button to load a new set. By the end of each session, participants submitted two final designs that met the requirements.
(4) *Interview Phase (20 min):* After completing the design tasks, participants took part in a 20-min interview. They answered 11 questions about their experience, impressions, and views on the toolkit usability. The interviews provided qualitative feedback on system performance, user experience, and future applications. Sessions were screen-recorded with participants' consent, and all recordings were transcribed anonymously. The process was approved by the institution's ethics committee. The transcripts were coded to analyze participant behaviour.

4.3. Data analysis

We conducted a two-round analysis to examine the interview data. In the first round, we performed a thematic analysis to explore how users engaged with the proposed framework. In the second round, we carried out a sentiment analysis to capture participants' emotional responses while using the framework. For the thematic analysis, we adopted a bottom-up reflexive approach based on Braun and Clarke's six-step framework (Maguire and Delahunt,2017). This approach enabled us to systematically identify patterns and themes within the participants' feedback on their experiences with the design toolkit. One author conducted the initial coding of the data, while a second author reviewed the coded results to ensure accuracy and consistency. The specific steps involved in this process are outlined below:

(1) *Familiarization with the Data:* First, we familiarised ourselves with the interview data by transcribing all interviews and cleaning them to remove filler words, non-verbal sounds, and irrelevant content. We then read through the responses several times to identify initial ideas and patterns for coding. For example, participants frequently mentioned a preference for familiar or aesthetically pleasing images, suggesting a pattern in image selection criteria.
(2) *Generating Initial Codes:* In this phase, we manually coded the dataset to capture participants' experiences and perspectives during the design process. We generated 126 initial codes from the interviews, covering practical points like 'preference for clean backgrounds' and 'iterative adjustments', as well as subjective insights such as 'seeking inspirational surprises' and 'frustration with randomness'. This detailed coding ensured a thorough understanding of the data's diversity and depth.

(3) *Searching for Themes:* After generating the initial codes, we organised them into potential themes that capture key patterns related to our research questions. For instance, codes like ‘preference for aesthetic appeal’, ‘selecting images that look good’, and ‘favoring harmonious color combinations’ were grouped under the theme Image Selection with the subtheme Aesthetic Preferences. This step required carefully clustering data to identify emerging themes.
(4) *Reviewing Themes:* After identifying potential themes, we reviewed them to ensure they accurately represented the data. This involved checking the themes against the coded segments and the entire dataset for validity and coherence. We refined the themes to remove overlaps and ensure each one was distinct. For example, overlapping themes about control over generated results were combined into ‘Difficulty Controlling AI Outputs’.
(5) *Defining and Naming Themes:* After refining the themes, we clearly defined and named each one based on what aspect of the data it captured. For example, the theme Image Selection covered factors like aesthetics, familiarity, and recognizability that influenced participants’ choices. Each theme name was chosen to be descriptive yet concise, capturing the essence of the grouped codes.

Additionally, we conducted sentiment analysis on interview responses to understand participants’ emotional reactions to the proposed toolkit, focussing on both positive and negative sentiments. Notably, Questions 4, 5, 7, 8, and 11 were excluded from the sentiment analysis to avoid bias. Questions 4 and 7 elicited neutral, descriptive responses that did not lend themselves to clear sentiment classification, while Questions 5 and 8 focussed on challenges, and Question 11 was about potential application scenarios and suggestions for improvement. Including these questions in the sentiment analysis would have skewed the overall sentiment results toward negativity or neutrality, as they inherently prompt participants to focus on difficulties or potential changes.

5. Results

In this section, we present our findings through quantitative analyses to evaluate the effectiveness of the proposed EEG-to-image model, comparing it with existing models using standardised evaluation metrics. We also include qualitative analyses of the user interview to examine the capability of the image local editing model in supporting visual blending tasks. These results provide initial evidence of the feasibility of the brain-guided visual blends technique. The quantitative analyses demonstrate the model’s ability to decode brain signals into corresponding images. In addition, the qualitative analysis, based on user interviews conducted after the souvenir design experiment, explores participants’ experiences and assesses whether the model successfully generated visual blends that aligned with their envisioned ideas.

5.1. Quantitative results

5.1.1. EEG signal decoding

In this section, we first present the effects of EEG reconstruction. As shown in Figure 3, it is evident that even when 75% of the EEG signal is masked, the model can still

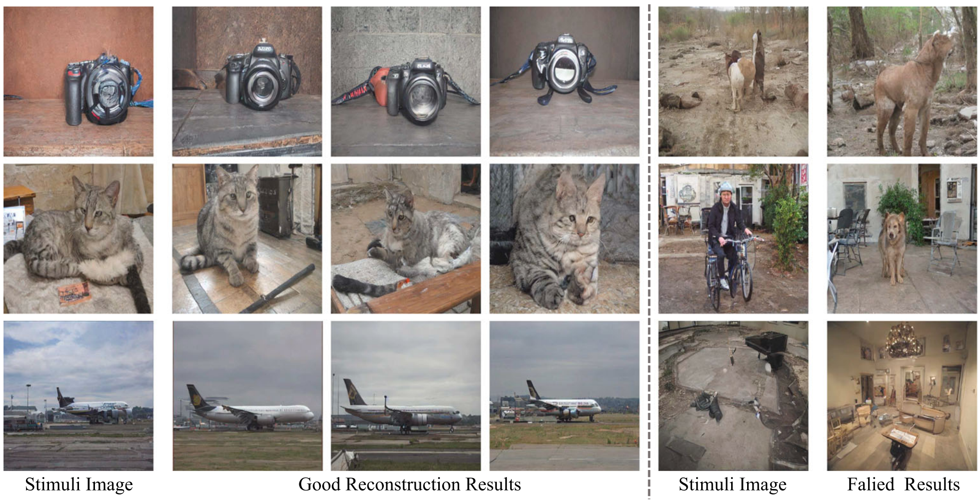

Figure 3. The visualisation of the EEG signal reconstruction results, the left column is the original EEG data, the middle column is the EEG data after random masking, and the right column is the reconstructed EEG data.

Table 2. The reconstruction accuracy of EEG data under different mask ratios.

Experiment ID	Mask ratio	Reconstruction accuracy
1	0.1	32%
2	0.25	38%
3	0.5	42%
4	**0.75**	**49%**
5	0.9	43%

Note: The table shows that both excessively high and low mask ratios negatively affect the model's performance, with the best results achieved at a mask ratio of 0.75.

effectively reconstruct the original signal's structure and pattern. During the experiments, we also tried adjusting the mask ratio from high to low during the training process. The highest EEG signal reconstruction accuracy, 47%, was achieved with a mask ratio of 0.75. Table 2 shows the EEG signal reconstruction accuracy under different mask ratios.

The left section of Figure 4 shows the EEG-to-image reconstruction results for various categories. From these results, it is clear that the generated images align well with the stimuli images corresponding to the EEG signals in terms of both category and semantic information. The right section of Figure 4 shows some examples of failed reconstructions. In the first row, the reconstructed image is clearly unreasonable, suggesting that the features extracted from the EEG signals may not effectively capture key category information. In contrast, the images in the last row, while maintaining semantic consistency, lack completeness and detail, impacting their overall quality. These shortcomings indicate that the pre-trained EEG encoder still has limitations in accurately encoding EEG signals. Following prior work (Kavasidis et al.,2017, Khare et al.,2022, Lan et al.,2023), we further evaluate the quality of reconstructed images using four standard metrics: FID (Heusel et al.,2017),

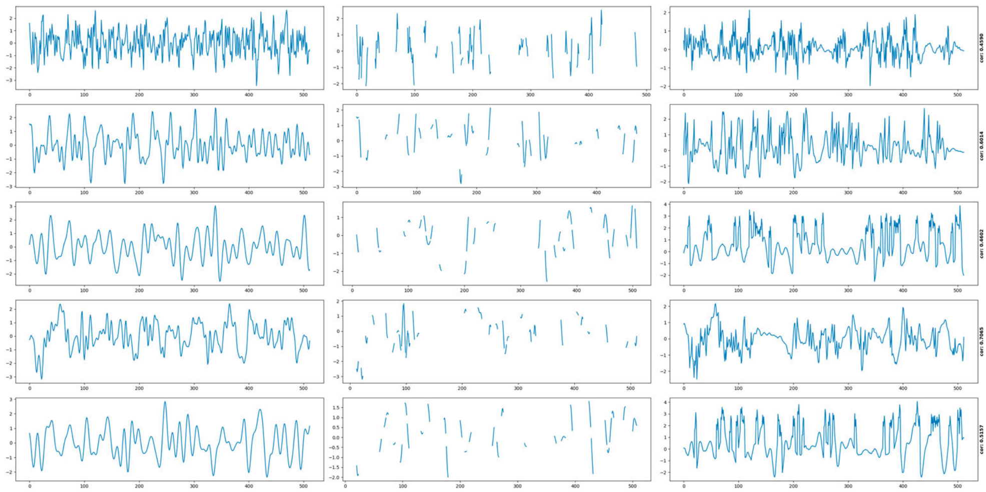

Figure 4. Image reconstruction results from the test set. The left section presents successful reconstruction results from the test set. The first column shows the original images paired with EEG data, and the following three columns display the corresponding images reconstructed from the EEG signals. The reconstructed images align closely with the original ones in both category and semantic content, indicating that the model can effectively reconstruct visual stimulus images from EEG data. The right section presents examples of failure cases, where the reconstructed images differ substantially from the original stimuli in both category and semantic meaning.

Table 3. Image reconstruction comparison.

Methods	FID ↓	IS ↑	SSIM ↑	LPIPS ↓
Brain2Image (Kavasidis et al.,2017)	18.76	5.01	0.213	0.701
NeuroVision (Khare et al.,2022)	–	5.23	0.213	–
NeuroImagen (Lan et al.,2023)	–	33.50	0.249	–
Ours	**3.61**	**35.82**	**0.271**	**0.644**

Note: We evaluate the quality of reconstructed visual stimuli using FID (Heusel et al.,2017), Inception Score (IS) (Salimans et al.,2016), SSIM, and LPIPS. Our brain decoding model consistently outperforms previous approaches across all evaluation metrics, indicating improved visual fidelity and semantic alignment.

Inception Score (IS) (Salimans et al.,2016), SSIM, and LPIPS. As shown in Table 3, our brain decoding model outperforms the previous works across all evaluation metrics, achieving lower FID and LPIPS scores and higher IS and SSIM scores. These results indicate that our model achieves SOTA performance in visual fidelity and semantic consistency for EEG-to-image reconstruction.

5.1.2. Local editing

We evaluated the performance of the Paint-by-Example model (B. Yang et al.,2023) in the local editing task. This evaluation was based on the background images used in the user study and was conducted through a web-based user interface specifically developed for the proposed design tasks. To ensure fairness and consistency across all test scenarios, all images were pre-processed to a uniform size of 256×256 pixels. Moreover, the model

Figure 5. Image local editing results showing how images decoded from brain signals are combined with background images to create visual blends. A total of five examples are presented. In each group, the top left image is the reference image, and the bottom left image is the source image with the masked region. The black areas in the reference images represent the mask regions provided to the model to guide the visual blending process. The images on the right show the final visual blends. All reference images were decoded from the EEG test set.

Figure 6. Final visual blends created by participants in the user study. Participants blend selected images that aligned with their design vision with other design elements to produce innovative visual blends.

checkpoint used during the evaluation was kept consistent with the one employed in the user study. All local editing results are shown in Figure 5.

As shown in the Figure 5, the local editing model accurately extracts the main object information from the reference images and seamlessly integrates it into the user-provided mask areas on the background images. The seamless integration results not only preserves the integrity of the original background but also produces a natural and aesthetically pleasing design outcome. Notably, the generated results effectively align with the user's intended requirements. Furthermore, the model's performance in local editing tasks demonstrates strong adaptability and stability across various combinations of backgrounds and objects, further highlighting its potential for practical applications in real-world design tasks.

5.1.3. User study design outcomes

The commemorative item design outcomes during the user study, illustrate how users utilised the image local editing model to achieve their design visions. The study included three different design tasks, requiring designers to blend various objects with another design element. The results, presented in Figure 6, show that all design outcomes met the task requirements, including some abstract creations that participants still found to contain sufficient design elements.

5.2. Comparative study with alternative ideation methods

5.2.1. Study setup

To evaluate design utility beyond image reconstruction quality, we conducted a comparative study contrasting three ideation approaches: (1) text-to-image prompting, (2) manual visual blending, and (3) brain-driven blending (ours). Ten participants with basic design experience completed a within-subject task in which they were asked to design a novel T-shirt product by visually blending a selected object with the T-shirt as the carrier. Each participant performed the task under all three conditions in counterbalanced order, with a 3-min limit per condition. Participants were free to iterate, regenerate, or modify their designs until they reached a satisfactory outcome. Participants were required to submit a final design that met their requirements. Following the completion of the design tasks across all three conditions, Participants were asked to rate each method using a 7-point Likert scale based on their overall satisfaction with the design experience and final design outcome.

In the Text-to-Image prompts group, participants generated product images using natural language prompts with Stable Diffusion v1.5. In the Manual Blending condition, participants created compositions by searching and combining existing images without the assistance of any generative model. In the Brain-Driven condition (ours), participants selected EEG-decoded images and refined the compositions using a mask-based local editing model. To ensure comparability across conditions, all tasks were performed on the same computer setup with identical hardware, display settings, and input devices.

5.2.2. Results

We first examined the time-to-idea across the three methods. Brain-Driven Blending demonstrated the shortest time-to-idea ($M = 1.3$ min), compared with Text-to-Image ($M = 2.4$ min) and Manual Blending ($M = 2.1$ min). This indicates that EEG-initialised concepts helped participants externalise initial design directions more rapidly than language-based prompting or manual composition. A similar pattern was observed in the number of iteration cycles required to reach a satisfactory outcome. Participants in the Brain-Driven method required an average of 4.2 cycles, whereas the Text-to-Image method required 6.8 cycles. Although text-based generation produced images quickly, participants frequently needed to reformulate prompts and regenerate results due to the stochastic nature of the model and the limited controllability of prompt-only interaction. In the Manual Blending condition, where each regional adjustment was counted as one cycle, participants required an average of 5.1 cycles to achieve a satisfactory blend. These results suggest that the local editing model used in our technique supports a more efficient path toward desired visual blends. By combining EEG-decoded initialisation with mask-based refinement, participants were able to explore alternatives with fewer iterations while maintaining control over the final design outcomes. Detailed results are shown in Table 4.

Designer satisfaction level with three methods was evaluated through post-task satisfaction ratings. Participants reported the highest level of satisfaction with the proposed brain-driven visual blend technique (mean $=$ 5.6, SD $=$ 0.52), significantly outperforming both the manual blending (mean $=$ 4.9, SD $=$ 0.33) and the text-to-image method (mean $=$ 4.1, SD $=$ 0.68). The lower scores for the text-to-image approach were primarily attributed to the stochastic nature and inherent uncontrollability of textual prompting,

Table 4. Detailed performance metrics per participant ($N = 10$) across three design methods, including text-to-image prompting, manual visual blending, and brain-driven blending (ours).

			Text-to-image (T2I)		Manual blending (MB)		Brain-driven (Ours)	
Participant	Gender	Design experience	Time (min)	Iterations	Time (min)	Iterations	Time (min)	Iterations
P1	Female	Intermediate	1.8	6	2.2	4	1.5	3
P2	Female	Expert	2.9	8	2.6	4	1.2	5
P3	Female	Novice	2.5	7	1.7	5	1.2	4
P4	Male	Intermediate	2.1	8	2.6	6	1.3	3
P5	Female	Intermediate	1.7	6	1.5	5	0.9	5
P6	Female	Intermediate	3.0	7	2.7	3	1.7	4
P7	Male	Intermediate	1.8	9	2.4	6	1.3	4
P8	Female	Intermediate	2.6	6	1.5	6	1.5	6
P9	Male	Intermediate	3.4	6	2.1	6	1.6	5
P10	Male	Expert	2.2	5	1.7	6	0.8	3
Mean	–	–	**2.4**	**6.8**	**2.1**	**5.1**	**1.3**	**4.2**

which forced designers into a tedious trial-and-error process to align the output with their mental concepts. In contrast, while manual blending offered deterministic control, it imposed excessive manual labour and high demands on the designer's visualisation skills, thereby increasing cognitive intensity and limiting exploration efficiency. Our technique effectively mitigated these barriers by providing an intuitive acquisition process and a controllable refinement loop, enabling designers to converge upon their desired outcomes more rapidly with an image local editing model.

To better understand how different approaches affect design diversity, we measured the average pairwise cosine distance between visual feature embeddings of the final designs produced by each method. This metric captures how visually varied the results are, with higher values indicating greater diversity. The results showed that the text-to-image method produced the most diverse outputs with a score of 0.81, followed by brain-driven Blending at 0.63 and manual blending at 0.51.

The high diversity observed in the text-to-image method was largely due to the unpredictable nature of the generation model, which often created visuals that looked very different from one another but did not always align with design task requirements. As a result, users frequently needed to refine prompts to steer the outputs. Brain-driven blending, on the other hand, achieved noticeably greater diversity than manual blending while still staying closely aligned with design task requirements and user intent. This suggests that the visual priors built into the local editing model can introduce useful variation without drifting too far from the original concept. The lower diversity seen in manual blending is likely linked to design fixation, as participants tended to reuse a small set of existing assets, which limited exploration and led to more similar results overall. Taken together, these findings suggest that brain driven initialisation helps broaden the design space beyond what manual methods allow, while avoiding much of the randomness associated with prompt-based generation.

5.3. Qualitative results

5.3.1. Thematic analysis

Thematic analysis of the interview led to the identification of six key themes that encapsulate participants' experiences, criteria, and challenges when using the proposed design toolkit. These themes provide insights into participants' decision-making, the impact of the toolkit on their creative practices, and the difficulties encountered during the design process. The identified themes are: *Image Selection Criteria*, *Iterative Design Process*, *Impact of Image Characteristics on Design*, *Creativity and Inspiration from AI-Generated Images*, *Challenges and Limitations in the Design Process*, and *Utility and Potential Improvements of the Tool*. Each theme is discussed in detail below, with supporting subthemes and codes (Table 5).

Image Selection. One of the most prominent themes that emerged was image selection, which guided participants in choosing images for their visual blending tasks. Participants focussed on aesthetic appeal, familiarity, and clarity of object features. Subthemes such as *Aesthetic Preferences*, *Familiarity and Stereotypes*, and *Object Recognizability and Clarity* highlight these considerations.

Participants frequently expressed a preference for images that were 'aesthetically pleasing' with 'harmonious color combinations' (Code: 'Selecting images that look good'). They often chose images based on their familiarity with the depicted objects, which provided comfort and ease when incorporating them into their designs (Code: 'Choosing familiar shapes'). Another critical factor was the recognizability of the objects within the images; participants favored images with 'clear and recognisable shapes' and 'distinct object features' (Code: 'Clear and recognisable shapes'). This suggests that while aesthetic appeal and familiarity were significant, the clarity and simplicity of the images were equally crucial for ensuring effective integration into the final design. These finding align with the recognition heuristic (Gigerenzer, Todd and ABC Research Group,2000) and reflects principles from cognitive load theory (Sweller,1988), where reducing perceptual complexity facilitates design integration and decision-making.

Impact of Image Characteristics on Design. The theme impact of image characteristics on design addresses the importance of specific image characteristics, such as background simplicity, color contrast, and object size, in influencing participants' design choices and integration strategies. Subthemes under this category include *Importance of Background Simplicity*, *Color and Contrast*, and *Object Size and Proportion*.

A recurring preference among participants was for 'simple, clean backgrounds' that allowed the main object to stand out (Code: 'Preference for simple backgrounds'). They indicated that complex backgrounds could detract from the main object, complicating the design process. The role of color and contrast also emerged as a key consideration, with participants often choosing images with 'high-contrast colors' for visual impact (Code: 'Selecting images based on color contrast'). Additionally, the size and proportion of objects within the images influenced their selections; participants preferred objects that were neither too small to be indistinct nor too large to be the dominate (Code: 'Selecting images with appropriately sized objects'). These findings suggest that image characteristics significantly affect design decisions, as they contribute to the overall visual coherence and appeal of the integrated design, which resonate with the insights from information visualisation on optimising image clarity and interpretability (Ware,2019).

Participants emphasised the importance of refining their visual blends through multiple rounds of iteration. One participant described how they would 'first decide on a category and then repeatedly click the "Random" button' to explore combinations that better matched their evolving design ideas. This process allowed them to 'continuously adjust my ideas and ultimately decide which version to refine further' (Code: 'Adjusting designs based

Table 5. Thematic analysis results from participant interviews.

Theme	Subtheme	Number of participant	Example codes
Image selection	Aesthetic preferences	13	Preference for aesthetic appeal, Selecting images that look good, Favoring harmonious color combinations
	Familiarity and stereotypes	6	Choosing familiar shapes, Selecting based on stereotypical appearance, Preference for familiar objects
	Object recognizability and clarity	9	Clear and recognisable shapes, Distinct object features, Preference for simple, clear images
Iterative process	Refinement through Repetition	14	Refining ideas through multiple iterations, Adjusting designs based on outcomes, Revisiting and refining designs repeatedly
	Evolving design ideas	10	Changing initial design ideas, Adapting ideas based on generated images, Allowing designs to evolve through the process
Impact of image characteristics on design	Background simplicity	15	Preference for simple backgrounds, Impact of clean background on object visibility, Avoiding complex backgrounds
	Color and contrast	8	Selecting images based on color contrast, Preference for high-contrast images, Impact of color on design decisions
	Object size and proportion	6	Selecting images with appropriately sized objects, Importance of object proportion in designs, Ensuring objects are not too small or too large
Creativity and inspiration from AI-generated images	Surprising creative outputs	8	Unexpected yet creative results, AI outputs that lead to new ideas, Surprising elements influencing design changes
	Balancing control and creativity	5	Balancing between control and randomness, Need for more control over AI outputs, Appreciation for AI-driven creativity
Challenges and limitations in the design process	Issues with image quality and resolution	7	Low-resolution images, Problems with image quality affecting design, Need for higher-quality image generation
	Difficulty controlling AI outputs	6	Lack of control over image combination, Difficulty predicting AI outputs, Challenges in refining AI-generated results
	Limited image variety and repetition	9	Repetition of images, Limited variety in generated images, Need for a broader image dataset
Utility and potential improvements of the tool	Practical applications in design	9	Useful for graphic design and prototyping, Effective for quick idea generation, Suitable for early-stage concept development
	Suggestions for enhancing functionality	6	Need for better image editing tools, Desire for more customisation and control

Note: The analysis identified six key themes based on participants' feedback. The table presents each theme along with its corresponding subthemes and includes representative example codes drawn from the participants' responses.

on outcomes'). Across interviews, participants highlighted the creative value of this iterative workflow, where generated visual blends served as visual prompts for design thinking. For instance, one participant mentioned that 'the mask helped me control only part of the image', enabling them to iteratively adjust local regions without losing the broader design structure. This adaptability often led to evolving visual compositions as participants responded to intermediate results and fine-tuned content placement using the local editing interface (Code: 'Changing initial design ideas').

Iterative Process. This theme captures participants' experiences of refining their designs through multiple iterations of visual blend generation. Subthemes such as *Refinement Through Iteration* and *Evolving Design Ideas* illustrate a dynamic workflow in which participants continuously responded to AI-generated outputs to progressively align with their creative vision.

Across interviews, participants emphasised the creative value of iteration. One participant described how they would 'first decide on a category and then repeatedly click the Generate button' to refine outputs that better matched their intent (Code: 'Adjusting designs based on outcomes'). Rather than seeking perfect results in one step, participants treated the system as a co-creative partner, using intermediate outputs as prompts for further ideation (Code: 'Changing initial design ideas'). Many leveraged the mask tool to adjust specific regions without altering the entire composition. As one noted, 'the mask helped me control only part of the image', enabling high level control over the generated results. This localised refinement illustrates a hybrid creative mode–combining top-down goals with bottom-up discovery–aligned with reflection-in-action (Schön,2017) and iterative creativity principles (Shneiderman,2007).

Creativity and Inspiration from AI-Generated Images. This theme captures two key patterns: *Surprising Creative Outputs*, where unexpected AI generations sparked new ideas, and *Balancing Control and Creativity*, which reflects participants' desire to guide outcomes more intentionally.

Participants often drew inspiration from surprising results (Code: 'AI outputs that lead to new ideas'), consistent with the value of unpredictability in ideation (Boden,2004). At the same time, many expressed a need for more control (Code: 'Need for more control over AI outputs'), highlighting a tension between exploration and direction. These findings suggest co-creative tools should combine generative serendipity with user-steerable mechanisms, such as region-specific constraints or adjustable randomness.

Challenges and Limitations in the Design Process. This theme highlights three common limitations participants faced during the visual blending process: *low image quality*, *unpredictable outputs*, and *limited variety in generated results*.

Low-resolution blends (Code: 'Problems with image quality affecting design') disrupted participants' workflows and reduced design confidence (Winograd and Flores,1986). Many also found it hard to anticipate how elements would combine (Code: 'Difficulty predicting AI outputs'), exposing a lack of system feedback (Tenner,2015). Repetitive results further limited exploration (Code: 'Need for a broader image dataset'), reinforcing creative fixation (Shneiderman,2007). These findings suggest a need for improved generation quality, content diversity, and finer-grained control to better support creative intent.

Utility and Potential Improvements of the Tool. Participants identified two main aspects regarding the tool: its practical applications in design and suggestions for enhancing its functionality. In terms of practical applications, participants noted the tool's potential in graphic design and prototyping, particularly during the early stages of concept development (Code: ‘Useful for graphic design and prototyping’). This aligns with Shneiderman’s principles for creativity support tools, which emphasise aiding users in early exploration phases (Shneiderman,2007). For functionality enhancements, participants expressed a desire for more customisation options and greater control over the tool’s outputs (Code: ‘Desire for more customisation and control’). This reflects the need for user-centered design principles, advocating for early focus on users and iterative design processes (Gould and Lewis,1985). Implementing these improvements could better accommodate the specific needs of professional designers.

5.3.2. Implication from participants interview

The code analysis reveals that participants engaged with the image-local editing model not as a passive tool for image generation, but as an interactive partner integrated into their iterative creative process. Their interactions indicate that the model supported them in several ways: it facilitated visual decision-making by allowing users to select and incorporate image components based on perceived clarity, aesthetic quality, and compositional resonate with their design vision; it acted as a creative catalyst by generating novel outputs that inspired new design directions; and it supported compositional reasoning by offering fine-grained control over the spatial combination of visual elements. Instead of passively accepting system outputs, participants actively explored design possibilities through region-based controls, using masks and reference images to guide content placement and blending.

Additionally, participants revealed several recurring patterns that point to the specific needs and cognitive processes involved in visual blending tasks using image local editing models. Participants’ emphasis on aesthetic appeal, familiarity, and object clarity indicates that image selection is not merely about finding a match, but about select image emotionally and conceptually that resonant with themselves. This suggests that systems aiming to should not only retrieve relevant content but also prioritise interpretability and recognizability in images. The recurring theme of iterative exploration demonstrates that participants did not expect to achieve their final design goals in a single step. Instead, they treated the tool as a creative partner, refining their ideas over time based on generated variations. This highlights the importance of providing lightweight iteration mechanisms–such as fast refresh, variation generation, and feedback-driven tuning–to facilitate real-time creative thinking. Inspiration derived from unexpected AI outputs further reinforces the system’s role not just as a tool for execution, but as a source of ideation. However, participants also expressed a strong desire to better control and steer these generative behaviours, indicating a need for systems that allow users to balance randomness with intentionality. The interplay between surprise and control emerges as a critical dimension in supporting co-creativity.

Finally, participants’ reflections on limitations–such as low image quality, lack of control, and limited variety–emphasise the gap between current capabilities and professional design expectations. These issues underscore the need for improved generation fidelity, more nuanced control mechanisms, and richer image datasets. Together, the findings point to several concrete directions for future system design: (1) incorporating semantic filtering for clearer image selection, (2) supporting iterative design workflows with minimal friction, and (3) enabling controllable blending strategies that adapt to user intent and context.

Finally, participants' reflections on limitations, particularly low image quality, insufficient control over outputs, and limited variety, reveal a gap between the current prototype and professional engineering design expectations. These issues are partly rooted in the blending capabilities of current generation models: the example-based blending model operates primarily at the pixel and region level without explicit understanding of functional or engineering attributes. As a result, participants often relied on trial-and-error to achieve desired compositions, indicating the need for higher-level interaction mechanisms. Future systems may incorporate semantic-level controls through text or voice input, allowing designers to specify attributes such as color palettes, material cues, functional roles, or spatial relationships in addition to masks and reference images. Such controls would enable constraint-based blending aligned with engineering requirements. For example, respecting dimensional proportions, material compatibility, or ergonomic considerations. Moreover, while current techniques utilise loops to allow subjects to continuously refine their ideas, adaptive refinement loops that expose intermediate parameters for user adjustment could further enhance the iterative design dialogue. This would provide greater adjustability, improve predictability, and support a more deliberate exploration of the design space.

5.3.3. Sentiment analysis

A sentiment analysis was conducted on selected interview responses to better understand participants' emotional reactions to the proposed toolkit. The analysis focussed on identifying both positive and negative affective responses to the tool's functionality and overall experience. Responses to Questions 4, 5, 7, 8, and 11 were excluded from the analysis: Questions 4 and 7 elicited factual or descriptive responses unsuitable for sentiment classification,

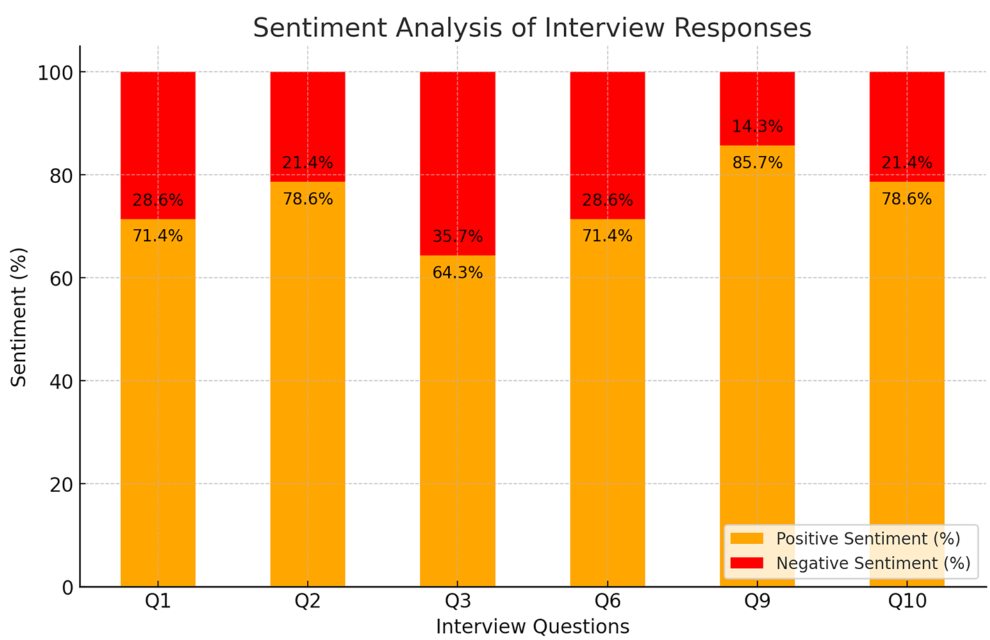

Figure 7. Sentiment analysis of interview responses (excluding Questions 5, 8, and 11).

while Questions 5, 8, and 11 focussed on practical issues or suggestions, which would have biased the analysis toward negativity or neutrality.

The results, shown in Figure 7, reveal that approximately 75% of the responses reflected positive sentiment, suggesting that participants appreciated the toolkit creative potential, flexibility, and contribution to their ideation process. The remaining 25% expressed negative sentiment, primarily linked to image quality concerns, limitations in category variety, or difficulty steering generated outputs. For example, in responses to Question 1, 10 were positive and 4 were negative, indicating general satisfaction tempered by specific usability concerns. These findings support the overall positive reception of the framework while also highlighting actionable areas for improvement in future iterations.

6. Discussion

In this section, we first discuss the application of brain decoding in human AI collaborative design. We then examine how local editing models can support visual blends. Finally, we address the limitations of our work and suggest possible directions for future research. These discussions aim to provide insights into the effectiveness, challenges, and future potential of design support tools for visual blends.

6.1. Implications of brain-guided design tools for early-stage design ideation process

From an engineering design perspective, the proposed brain-guided visual blending tool can support early-stage design ideation by accelerating iteration and reducing the cognitive effort required to externalise mental imagery. During conceptual design and styling phases, designers often iterate over multiple visual alternatives while translating vague or pre-verbal ideas into sketches, image searches, or verbal descriptions. By generating visual concept candidates directly from neural responses associated with visual perception and enabling lightweight refinement through local editing, the system shortens the loop between internal mental imagery and external representation. This allows designers to rapidly explore and compare alternative directions before committing to formal sketches or CAD models, while focussing attention on higher-level decisions such as aesthetic intent, semantic direction, and compositional balance.

Beyond individual ideation, brain-guided visual blends can also support collaborative product development and integration with downstream engineering design workflows. The generated visual artifacts may function as pre-CAD concept seeds or style references that inform subsequent parametric modelling, generative design, or optimisation processes, where they can be translated into geometric constraints, reference forms, or aesthetic guidelines. In design evaluation and stakeholder communication settings, these visual blends further serve as shared reference artifacts that externalise implicit design intent, facilitating concept comparison, co-creation, and early decision-making prior to detailed modelling or prototyping.

To further clarify the distinctive role of the proposed technique, Table 6 compares brain-driven blending with text-to-image generation and traditional visual blending across design-oriented dimensions, including intent expression, interaction process, controllability, and engineering alignment. The comparison highlights that, unlike language-based

Table 6. Comparison of three visual creation paradigms (text-to-image generation, traditional visual blending, and brain-driven blending) across design-oriented dimensions.

Category	Dimension	Text-to-image generation	Traditional visual blending	Brain-Driven blending (Ours)
Input & Intent	Primary input	Explicit textual prompts (Zhao et al.,2020)	Retrieved images or manual sketches (Chilton, Petridis, and Agrawala,2019)	EEG-decoded image and designer mask
	Creativity source	Model prior conditioned by language	Designer expertise and existing assets	Neural intent combined with generative prior
	Intent expression	Limited by linguistic precision (Ha and Eck,2017)	Requires explicit articulation (Chilton et al.,2021)	Direct expression of vague or hard-to-verbalise ideas with mask drawing
	Support for vague ideas	Moderate; depends on wording	Low without references	High through implicit neural cues
Interaction process	Designer effort	Moderate prompting effort	High manual editing effort	Low articulation and intuitive selection
	Need for verbalisation	Strongly required	Low requirement for image searching	Not required
	Cognitive load	Moderate; requires iterative trial-and-error prompting	High due to extensive searching and manual synthesis (Chilton, Petridis, and Agrawala,2019)	Low; intuitive acquisition and automated local editing
	Iteration style	Prompt rephrasing cycle	Pixel-level manual edits	Regenerate–select–edit loop
	Human role	Prompt engineer	Direct craftsman	Mixed-initiative co-creator
System capability	Controllability	Limited spatial control	High but skill-dependent	Moderate–high via mask and iteration
	Predictability	Low, stochastic	High, deterministic	Exploratory yet steerable
	Design exploration	Broad but language-constrained	Narrow to known assets	Broad with associative cues
	Risk of fixation	Linguistic bias	Strong reference fixation	Reduced through non-verbal cues
Outcome & Use	Time cost	Short per iteration	Long manual process	Short with rapid regeneration
	Efficiency	Moderate efficiency	Low efficiency	High efficiency
	Skill requirement	Prompt formulation skills	High drawing and visual search skills	Low; mask drawing and selection only
	Creativity support	Idea expansion via prompts	Precision refinement	Inspiration from implicit intent
Usability	Accessibility	Requires prompt skills	Requires artistic skills	Accessible without drawing and design skills
	Learning barrier	moderate	High	Low

generation, the proposed method supports pre-verbal and hard-to-articulate ideas through neural cues, while requiring substantially lower manual skill than conventional blending. It also illustrates the mixed-initiative nature of the workflow, where designers retain spatial control through masking and iterative refinement, in contrast to purely prompt-driven or fully manual paradigms. These characteristics position brain-guided blending particularly well for early ideation stages, where designers seek rapid exploration of alternatives with minimal articulation effort, while still maintaining a pathway toward constraint-based and engineering-relevant development.

6.2. Engineering design applications and hybrid workflows

The proposed technique also suggests several engineering-relevant application scenarios. In *automotive styling*, brain-guided visual blends could support the translation of designers' latent style intentions into coherent concept sketches. Designers often start from an internalised 'style language' (e.g. perceived dynamism and calmness, softness and sharpness, minimal and expressive detailing) that is difficult to verbalise but guides decisions about proportions, silhouette hierarchy, curvature flow, and line vocabulary. By decoding neural responses and generating multiple candidate visuals that reflect these implicit preferences, the system can externalise a style-consistent set of concept sketches early, which designers can then curate and iteratively refine. This enables a semi-automated styling workflow where brain-driven concepts establish a style baseline (e.g. dominant form primitives, characteristic lines, and compositional rhythm), and subsequent refinement aligns these stylistic cues with downstream modelling constraints and brand consistency.

In *industrial design*, the technique can support the creation of ergonomic product variations by externalising users' implicit and experience-based preferences that are often difficult to articulate through conventional methods. Perceptions of comfort, grip stability, visual weight, or approachability are typically formed through embodied experience rather than explicit reasoning, and designers rely heavily on tacit judgment when shaping handles, interfaces, or product contours. Brain-guided visual blends offer a complementary pathway to capture such pre-reflective impressions and translate them into families of visually grounded alternatives that reflect subtle preferences in curvature, proportion, texture suggestion, or form softness. These alternatives can function as exploratory probes within user-centered design processes, enabling designers to compare and refine ergonomic directions before formalising them into parametric models or physical prototypes. By linking subconscious affective responses with generative variation, the approach may help industrial designers broaden the search space of ergonomic solutions while maintaining alignment with perceived usability and comfort.

In *human-centered and inclusive design*, brain-driven interaction offers a meaningful alternative for users whose motor, speech, or literacy abilities limit participation in conventional design activities. Many inclusive design methods still assume the capacity to sketch, manipulate interfaces, or articulate preferences verbally, which can unintentionally exclude certain user groups from early ideation. By allowing individuals to convey design intent through neural signals combined with lightweight interactions such as selection and mask adjustment, the proposed approach can lower these participation barriers and provide a more equitable channel for expressing personal needs and aspirations.

Rather than positioning the system as a replacement for human agency, it can function as an assistive mediator that translates subtle affective or experiential preferences into visual proposals, enabling users to take part in shaping products and environments that directly concern them. Such involvement may strengthen user empowerment and co-creation in domains such as assistive devices, home adaptations, or accessible consumer products, where lived experience is crucial yet often underrepresented in formal design processes.

These applications can be embedded within hybrid engineering workflows that connect brain-driven concepts to downstream generative design platforms and optimisation tools. Decoded images and visual blends may function as semantic anchors or style references that are translated into parametric constraints, reference geometries, or preference signals for AI-driven exploration. In this pipeline, implicit design intent informs the initial search space, while parametric modelling and performance-oriented optimisation refine feasibility under dimensional, material, and manufacturing constraints. Such integration positions brain-guided blending not as a replacement for engineering tools, but as a complementary front-end interface that bridges human imagination with computational design refinement.

6.3. The role of brain decoding in human computer interaction

Our findings highlight the potential of brain decoding as an alternative approach for capturing users' envisioned ideas. By enabling brain decoding models to translate users' brain signals directly into visual content that aligns with their internal visions, the system provides a more intuitive way to express vague, subconscious, or unarticulated ideas, which are common in the early stages of ideation. The latent and emotional information embedded in brain signals can further enrich the depth and quality of the generated ideas, thereby supporting a more informative ideation process. This directness reduces dependence on traditional input methods such as text prompts or image retrieval, which often struggle to reflect the subtlety or immediacy of a designer's creative ideas. This approach aligns with recent work in brain-computer interfaces within HCI, which investigates neural signals as a means of implicit interaction or intent expression (Afergan,2014, de la Torre-Ortiz et al.,2020). By decoding designers' thought patterns into visual content, our system introduces a new paradigm for supporting design idea visualisation that is naturally personal, intuitive, and emotionally meaningful. Additionally, as mentioned in J. Shi et al. (2023) and W. Tang et al. (2024), more natural brain-controlled interface systems in HCI can reduce users' cognitive load while also enabling new possibilities for users with disabilities and novel modes of interaction. This work also offers a potential pathway toward addressing a research question posed by J. Shi et al. (2023): 'How can natural interactions be integrated with current generative AI models, and what are the possible connections between these interactions and content generation?'

However, this approach also requires careful consideration of signal reliability, device usability, and ethical concerns surrounding neural data. While current EEG systems pose constraints in terms of portability and signal noise, advancements in wearable BCI technologies are expected to improve the feasibility of brain-driven creative systems in real-world design workflows.

6.4. Supporting visual blends through local editing models

The local editing model played a central role in enabling participants to construct visual blends by selectively manipulating image regions. Unlike traditional generative models (Jin et al.,2000, W. Li et al.,2025, Wu et al.,2025) that produce holistic, end-to-end outputs, the region-based editing approach allowed participants to exert fine-grained control over composition, spatial arrangement, and visual coherence. Participants actively used masks and reference images to iteratively guide content placement and blending, often adjusting specific regions to better align with their creative intentions.

This interaction pattern reflects a shift from passive generation to mixed-initiative co-creation, where the system serves as a flexible partner rather than a static tool. Participants treated the model as both a creative catalyst and a compositional assistant, supporting their exploration of novel visual combinations while allowing for intentional design control. This capability is particularly relevant for visual blending tasks, where spatial reasoning and aesthetic balance are crucial. These findings support prior work on creativity support tools (Shneiderman,2007) and mixed-initiative interfaces (Horvitz,1999), reinforcing the importance of editable, responsive systems that accommodate iteration and personal input. Future systems could build on this interaction model by incorporating more semantic-level control, adaptive refinement strategies, and feedback-driven tuning mechanisms to further support compositional creativity.

6.5. Ethical and practical considerations

The integration of neuro-driven tools into engineering design workflows introduces unique ethical and practical challenges that need to be addressed to ensure responsible and effective deployment. Beyond technical performance, issues of data governance, signal reliability, and designer acceptance will critically shape whether such systems can become trusted components of professional design practice.

6.5.1. Ethical implications and data privacy

The use of neural data raises significant privacy concerns, as brain activity may inadvertently reveal information beyond immediate design intent, including cognitive states or affective patterns. Future implementations should therefore adhere to privacy-by-design principles. This includes local processing of raw EEG signals on the user's device, minimisation of data retention, and the use of techniques such as differential privacy to anonymize neural features before they interact with generative models. Equally important are transparent informed-consent protocols that clearly define how neural data are used and stored. Organisations adopting such tools should establish governance guidelines that limit secondary use of neural data and guarantee users' rights to access, review, or delete their recordings.

6.5.2. Signal reliability and system robustness

The practical utility of brain-driven design tools is inherently tied to the reliability of EEG measurements, which are susceptible to physiological artifacts and environmental noise. Although our current model employs pre-training to improve feature extraction, real-world engineering applications will require robust artifact rejection, uncertainty estimation, and mechanisms for manual override. The persistent gap between neural signals and design intent, whereby decoded images may not fully reflect the designer's mental imagery, suggests that such systems are better understood as co-creative partners rather than deterministic instruments. Combining EEG with complementary modalities such as eye-tracking or voice input could further enhance robustness by providing explicit spatial or semantic cues when neural signals are ambiguous.

6.5.3. Designer acceptance and training requirements

Adoption by the design community will depend on usability and the perceived locus of control. In our study, participants appreciated the reduction of manual effort but also required time to learn how to maintain focussed mental states for stable recording. Integrating neuro-driven tools into professional workflows will therefore necessitate dedicated onboarding and training programs that cover both technical operation of EEG hardware and basic neuro-feedback strategies. Crucially, these tools should be framed as augmentative instruments that complement sketching, CAD modelling, and other established practices, leaving final decision authority with the designer. Demonstrating clear benefits in creativity support or cognitive load reduction, while providing explainable feedback about how neural input influences outcomes, will be essential for long-term acceptance.

7. Limitations

While our system demonstrates the potential of combining brain decoding with a local editing model to simplify the visual blending process, several limitations remain.

7.1. Generalizability potential in complex design engineering tasks

The current user study focus on souvenir design tasks (e.g. T-shirts and mugs) is partly constrained by the present capabilities of both brain decoding and visual blending models. Existing EEG-to-image decoders can only reconstruct a limited range of semantic categories determined by available object classes in the training datasets, which restricts the immediate applicability to domains where corresponding training data are absent. Similarly, different combinational generative models exhibit distinct blending preferences and fusion logics, and their performance varies across tasks, indicating that domain-specific adaptation of the blending model is necessary for more complex engineering applications. For these reasons, the study intentionally adopted low-complexity artifacts as a controlled proxy to isolate early-stage ideation and concept externalisation, without introducing additional functional or manufacturing constraints that could obscure the evaluation of the proposed interaction paradigm.

Nevertheless, the technique is not inherently limited to souvenir design and can be extended to a range of engineering-relevant domains. In automotive styling, for example, brain-driven visual blends could support early form exploration by combining functional references such as aerodynamic profiles with aesthetic cues reflecting brand identity, helping designers prior to detailed surfacing. In product design, the approach could enable design personalisation by allowing users to blend structural elements from multiple artifacts based on their own implicit preferences and to iterate rapidly with the assistance of generative models, supporting faster discovery of individualised concepts. For example, Sun et al. (2025) and Chilton, Petridis, and Agrawala (2019) have demonstrated how visual blends can convey combined meanings in graphic design communication context, while P. Wang et al. (2025)introduced an OC-GAN model to integrate structural features from existing products into novel configurations for product design tasks. As brain decoding models and blending techniques continue to evolve, the proposed framework can naturally incorporate more advanced decoders and task-specific generative models, enabling its application to increasingly complex engineering design challenges.

7.2. Use of perceptual EEG as a proxy for imagery EEG

We used perceptual EEG data instead of mental imagery EEG for both theoretical and practical reasons. Neuroscience research has consistently shown substantial overlap in neural responses between visual perception and mental imagery, particularly in early visual areas (V1–V4), including shared retinotopic organisation and category selectivity (Ganis, Thompson, and Kosslyn,2004, Xie, Kaiser, and Cichy,2020). Additional studies confirm that both visual perceptual and imagery rely on the same neural substrates, although perception generally evokes stronger signals (Canales-Johnson et al.,2021, Dijkstra, Bosch, and van Gerven,2019, Wilson et al.,2024). These findings justify perceptual EEG as a reasonable substitute when imagery EEG is unavailable. Practically, the absence of open-access, high-quality imagery EEG datasets, combined with the difficulty of replicating the 128-channel ImageNet-EEG collection conditions (Kavasidis et al.,2017), precluded us from collecting suitable imagery data. Given these constraints, our use of perceptual EEG as a provisional substitute for imagery data is a justified and practical choice. It enables early-stage exploration of neuroCHI (Tanaka et al.,2024) and serves as a critical step in the progression from non-BCI-driven CHI systems toward real-time imagery BCI-enabled CHI systems. It offers foundational insights for advancing toward practical neurodesign (W. Tang et al.,2024) tools as portable EEG hardware, dataset availability, and cross-subject models continue to improve.

7.3. Model generalizability and scalability.

The current EEG-to-image model is trained on a relatively limited dataset, and its interpretability is inherently tied to the categories represented in that dataset. Decoding performance also varies considerably across individuals, requiring subject-specific tuning that hinders scalability (J. Shi et al.,2023). To improve generalisation, future research should explore large-scale, multi-subject EEG datasets and investigate multimodal pretraining approaches, such as CLIP (Radford et al.,2021), that align brain signals with visual-semantic spaces.

7.4. Local editing unpredictability.

During the local editing stage, participants occasionally encountered inconsistent visual blends, particularly during iterative refinement. While some randomness proved creatively stimulating, excessive unpredictability often disrupted the design workflow. This suggests that current local editing approaches struggle to incorporate engineering-oriented constraints and high-level design semantics. Additionally, the local editing model primarily manipulates visual appearance rather than functional properties, it often produces results that are aesthetically pleasing but incompatible with dimensional, material, or manufacturing requirements. To bridge this gap, future local editing models should integrate semantic representations and constraint-aware mechanisms. Such advancements would allow designers to guide the fusion process using specific attributes, rules, or performance metrics, effectively strengthening the transition from creative exploration to realistic engineering development. Furthermore, providing real-time, region-level feedback would enhance output consistency and user control without compromising creative flexibility.

7.5. *Offline two-stage pipeline limitations*

Our current system operates as an offline, two-stage pipeline: first decoding images from EEG signals as a proxy for image selection, and then refining them through a local editing model to directly generate visual blends. This decoupling introduces information loss and makes it difficult to preserve the user's design vision in an end-to-end manner. This paper presents a proof-of-concept technique rather than a complete or real-time system. The core contribution lies in demonstrating the feasibility of integrating brain decoding models with computational design tools to support brain-driven visual blending. Our focus is on conceptual validation and early-stage exploration, not deployment. This offline implementation fills a critical gap between non-BCI-driven CHI systems and future real-time imagery BCI-enabled CHI systems. As datasets, portable EEG devices, and cross-subject models advance, minimal structural changes will be needed to adapt the system for live use. Future work could also explore direct decoding of design-relevant structures–such as object layouts or scene graphs–enabling tighter integration between brain decoding and generative modelling. Achieving a seamless brain-to-design workflow will require advances across neuroscience, machine learning, and HCI, alongside improvements in portable EEG hardware, dataset availability, and cross-subject decoding methods.

8. Future works

8.1. *Roadmap toward real-time and scalable deployment*

The primary barriers to real-world adoption of the proposed technique lie in the current capabilities of brain decoding models and visual blending generation models. A practical roadmap toward real-time deployment therefore involves several complementary developments.

- *EEG acquisition hardware*: Lightweight and portable EEG acquisition hardware with improved signal quality is essential to enable designers to capture reliable neural signals in everyday working environments. Coupling such hardware with on-device or low-latency cloud-based inference would allow continuous decoding of EEG streams into visual representations, moving the system from an offline prototype to an interactive design companion.
- *Open-vocabulary brain decoding*: Expanding the generalizability of brain decoding models to open-vocabulary brain decoding is critical. Current decoders are constrained by the limited semantic categories available in existing datasets, which restricts the range of

concepts that can be reconstructed. Building larger-scale visual perception and imagery datasets, combined with more expressive representation learning techniques, could enable models to decode objects beyond those currently present in the dataset. In the long term, such advances may allow decoded images to serve directly as viable design proposals, reducing the need for manual post-hoc blending and further lowering the effort required from designers.

- **Cross-subject brain decoding ability**: Improving cross-subject generalisation remains a key challenge for scalable use. Promising directions include the collection of multi-subject brain-image datasets, the adoption of subject-adaptive learning strategies, and lightweight online fine-tuning to personalise decoders. These approaches would reduce individual calibration requirements and make the system more robust across diverse users and contexts.
- *Controllable visual blending generation models*: Advances in controllable visual blending models are needed to ensure that generated results are not only high-quality but also aligned with designers' intent and engineering constraints. Future blending models should support explicit control over semantic attributes, spatial composition, and functional features, enabling designers to steer the fusion process toward specific aesthetic directions or performance-related requirements. Incorporating constraint-aware mechanisms and interactive refinement loops would allow designers to iteratively adjust blends based on feedback, thereby improving predictability, supporting faster design iteration, and reducing the likelihood of outputs that deviate from intended concepts.

8.2. Future research directions

Moving forward, we aim to improve this technique in three related ways. First, there is a clear need to develop real-time BCI-based design systems that connect directly with CAD and design tools so they can be used in real engineering practice. Instead of treating brain decoding as a separate entry step, future systems could support a two-way interaction. Neural signals could be translated into parametric elements such as curve proportions or initial topologies, while feedback from CAD modelling and simulation could help refine and guide the decoding process. This type of integration would allow designers to move smoothly from early concept formation to detailed surface modelling analysis and optimisation within a single coherent workflow.

Second, support for design ideation should expand beyond EEG alone to include eye tracking, gesture-based input, and semantic input. Eye gaze can reveal where attention is focussed and which areas matter most to the designer, while hand gestures can express spatial relationships and structural ideas that are difficult to capture through neural signals by themselves, and semantic input through voice or text can better guide the model generation process. A unified framework that combines these inputs and adjusts their influence over time could balance intuitive preferences with intentional actions. This would allow designers to shift naturally between open exploration and precise control.

Third, new metrics are needed to better understand how brain-guided design tools affect creativity, sensemaking, and decision-making. Rather than relying only on image quality measures, future studies could look at factors such as how quickly insights emerge, how broadly the design space is explored, patterns of visual attention, and changes in how problems are framed. Linking behavioural data with neurophysiological signals could offer deeper insight into how these systems influence reasoning collaboration and confidence. Developing such metrics will be essential for placing brain-guided design within the larger field of engineering design research and for meaningfully comparing it with more traditional ideation approaches.

9. Conclusion

In this research, we present a brain-guided visual blends technique that integrates an EEG-to-image model with an image local editing model to support the creation of visual blends in early-stage ideation of engineering design workflows. This system allows designers to externalise brain concepts by decoding brain signals into visual concepts and flexibly incorporating these outputs into cohesive image compositions using an image local editing model. It streamlines both the image selection and composition stages of visual blends, which are typically cognitively demanding and require artistic skill. To evaluate the system, we conducted a two-part assessment targeting each component. First, we benchmarked our EEG-to-image model using standard metrics for visual reconstruction quality, demonstrating its SOTA performance in decoding visual stimuli from brain signals. Second, we conducted a user study with 15 participants to examine how the local image editing model supports visual blending tasks. The results show that participants were able to create creative and personalised designs by iteratively refining visual blends using masks and reference images. The system reduced barriers to creative expression, enabling users without formal design training to engage in composition-focussed design workflows. This two-part evaluation confirms the practical value of both components and highlights the effectiveness of combining brain-guided ideation with controllable visual editing for visual blending tasks. It also suggest that brain-driven initialisation can reduce articulation effort while maintaining controllability, supporting early-stage exploration in engineering design workflows. This paper presents a proof-of-concept technique rather than a complete or real-time system, which contributes to ongoing research at the intersection of HCI, BCI, and creativity design support tools by introducing a new paradigm for brain-driven design supporting tool. Beyond the current implementation, the approach can be embedded within hybrid pipelines where brain-decoded concepts act as semantic anchors for downstream parametric modelling, generative design, and optimisation. While the present system operates offline and represents a proof-of-concept, it fills an important gap between non-BCI design tools and future real-time brain-enabled systems. We anticipate that advances in real-time decoding, multimodal interaction, and design cognition metrics will enable broader adoption of brain-guided tools in industrial engineering design field.

Acknowledgments

To Robert, for the bagels and explaining CMYK and color spaces.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Xun Zhang http://orcid.org/0009-0003-7693-1805

Maaike Kleinsmann https://orcid.org/0000-0003-4120-5048
Stephen Jia Wang http://orcid.org/0000-0001-9835-9932
Di Yan https://orcid.org/0009-0004-5461-9084
Ziyu Wei https://orcid.org/0000-0002-1573-167X
Pan Wang https://orcid.org/0000-0001-6575-539X

References

Afergan, Daniel. 2014. "Using Brain-Computer Interfaces for Implicit Input." In Adjunct Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, 13–16, Honolulu, Hawaii, USA.

Ahmed, Hamad, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. 2021. "Object Classification from Randomized EEG Trials." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3845–3854, Nashville, TN, USA.

Alves Mendes Vasiljevic, Gabriel Alves Mendes, and Leonardo Cunha de Miranda. 2024. "Comparing Users' Performance and Game Experience between a Competitive and Collaborative Brain-Computer Interface." *Behaviour & Information Technology* 43 (1): 40–59. https://doi.org/10.1080/0144929X.2022.2152727.

Aranyi, Gabor, Fred Charles, and Marc Cavazza. 2015. "Anger-Based BCI Using fNIRS Neurofeedback." In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*, 511–521. Charlotte, NC, New York, NY: Association for Computing Machinery.

Aristimunha, Bruno, Igor Carrara, Pierre Guetschel, Sara Sedlar, Pedro Rodrigues, Jan Sosulski, Divyesh Narayanan, et al. 2023. "Mother of All BCI Benchmarks." V. 1.0.0. Zenodo. https://github.com/NeuroTechX/moabb.

Avrahami, Omri, Dani Lischinski, and Ohad Fried. 2022. "Blended Diffusion for Text-Driven Editing of Natural Images." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18208–18218, New Orleans, LA, USA.

Bagchi, Subhranil, and Deepti R. Bathula. 2022. "EEG-ConvTransformer for Single-Trial EEG-based Visual Stimulus Classification." *Pattern Recognition* 129:108757. https://doi.org/10.1016/j.patcog.2022.108757.

Bai, Yunpeng, Xintao Wang, Yan-pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan. 2023. "Dreamdiffusion: Generating High-Quality Images from Brain EEG Signals." Preprint. arXiv:2306.16934.

Bing, Yuan, Liya Yu, Shaobo Li, Young-sik Cho, and Chuanjiang Li. 2025. "A Novel Product Shape Innovation Design Method Based on Kansei Engineering and GAN Model with Limited Sample Data." *Journal of Engineering Design* 37:981–1006.

Bocquelet, Florent, Thomas Hueber, Laurent Girin, Stéphan Chabardès, and Blaise Yvert. 2016. "Key Considerations in Designing a Speech Brain-Computer Interface." *Journal of Physiology-Paris* 110 (4): 392–401. https://doi.org/10.1016/j.jphysparis.2017.07.002.

Boden, Margaret A. 2004. *The Creative Mind: Myths and Mechanisms*. London: Psychology Press. https://doi.org/10.4324/9780203508527.

Bridger, Darren. 2017. *Neuro Design: Neuromarketing Insights to Boost Engagement and Profitability*. London: Kogan Page Publishers. https://doi.org/10.1177/1470785318813748.

Brocke, Jan Vom, René Riedl, and Pierre-Majorique Léger. 2013. "Application Strategies for Neuroscience in Information Systems Design Science Research." *Journal of Computer Information Systems* 53 (3): 1–13. https://doi.org/10.1080/08874417.2013.11645627.

Canales-Johnson, Andrés, Renzo C. Lanfranco, Juan Pablo Morales, David Martínez-Pernía, Joaquín Valdés, Alejandro Ezquerro-Nassar, Álvaro Rivera-Rei, et al. 2021. "In Your Phase: Neural Phase Synchronisation Underlies Visual Imagery of Faces." *Scientific Reports* 11 (1): 2401. https://doi.org/10.1038/s41598-021-81336-y.

Cao, Beining, Hongwei Niu, Jia Hao, and Guoxin Wang. 2022. "Building EEG-based CAD Object Selection Intention Discrimination Model Using Convolutional Neural Network (CNN)." *Advanced Engineering Informatics* 52:101548. https://doi.org/10.1016/j.aei.2022.101548.

Chen, Chen, Guohong Chai, WeiChao Guo, Xinjun Sheng, Dario Farina, and Xiangyang Zhu. 2019. "Prediction of Finger Kinematics from Discharge Timings of Motor Units: Implications for Intuitive

Control of Myoelectric Prostheses." *Journal of Neural Engineering* 16 (2):026005. https://doi.org/10.1088/1741-2552/aaf4c3.
Chen, Liuqing, Shuhong Xiao, Yunnong Chen, Linyun Sun, Peter R. N. Childs, and Ji Han. 2025. "An Artificial Intelligence Approach for Interpreting Creative Combinational Designs." *Journal of Engineering Design* 36 (5-6): 920–947. https://doi.org/10.1080/09544828.2024.2377068.
Chen, Liuqing, Yuan Zhang, Ji Han, Lingyun Sun, Peter Childs, and Boheng Wang. 2024. "A Foundation Model Enhanced Approach for Generative Design in Combinational Creativity." *Journal of Engineering Design* 35 (11): 1394–1420. https://doi.org/10.1080/09544828.2024.2356707.
Chen, Zijiao, Jiaxin Qing, Tiange Xiang, Wan Lin Yue, and Juan Helen Zhou. 2023. "Seeing beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 22710–22720, Vancouver, BC, Canada.
Chew, Lin Hou, Jason Teo, and James Mountstephens. 2016. "Aesthetic Preference Recognition of 3D Shapes Using EEG." *Cognitive Neurodynamics* 10:165–173. https://doi.org/10.1007/s11571-015-9363-z.
Chilton, Lydia B., Ecenaz Jen Ozmen, Sam H. Ross, and Vivian Liu. 2021. "VisiFit: Structuring Iterative Improvement for Novice Designers." In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–14, Yokohama, Japan.
Chilton, Lydia B., Savvas Petridis, and Maneesh Agrawala. 2019. "VisiBlends: A Flexible Workflow for Visual Blends." In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–14, Glasgow, Scotland, UK.
Cichy, Radoslaw M., Jakob Heinzle, and John-Dylan Haynes. 2012. "Imagery and Perception Share Cortical Representations of Content and Location." *Cerebral Cortex* 22 (2): 372–380. https://doi.org/10.1093/cercor/bhr106.
Confalonieri, Roberto, Joseph Corneli, Alison Pease, Enric Plaza, and Marco Schorlemmer. 2015. "Using Argumentation to Evaluate Concept Blends in Combinatorial Creativity." In Proceedings of the Sixth International Conference on Computational Creativity, 174–181, Park City, UT.
Crawford, Chris S., and Juan E. Gilbert. 2019. "Brains and Blocks: Introducing Novice Programmers to Brain-Computer Interface Application Development." *ACM Transactions on Computing Education (TOCE)* 19 (4): 1–27. https://doi.org/10.1145/3335815.
Cunha, João Miguel, Pedro Martins, and Penousal Machado. 2020. "Let's Figure This out: A Roadmap for Visual Conceptual Blending." In Proceedings of International Conference on Innovative Computing and Cloud Computing, 445–452, Delhi, India.
de la Torre-Ortiz, Carlos, Michiel M. Spapé, Lauri Kangassalo, and Tuukka Ruotsalo. 2020. "Brain Relevance Feedback for Interactive Image Generation." In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, 1060–1070, Virtual Event USA.
Deng, Xin, Zhongyin Wang, Ke Liu, and Xiaohong Xiang. 2023. "A GAN Model Encoded by CapsEEGNet for Visual EEG Encoding and Image Reproduction." *Journal of Neuroscience Methods* 384:109747. https://doi.org/10.1016/j.jneumeth.2022.109747.
Dijkstra, Nadine, Sander E. Bosch, and Marcel A. J. van Gerven. 2019. "Shared Neural Mechanisms of Visual Perception and Imagery." *Trends in Cognitive Sciences* 23 (5): 423–434. https://doi.org/10.1016/j.tics.2019.02.004.
Duan, Xu, Songyun Xie, Xinzhou Xie, Klaus Obermayer, Yujie Cui, and Zhenzhen Wang. 2021. "An Online Data Visualization Feedback Protocol for Motor Imagery-Based BCI Training." *Frontiers in Human Neuroscience* 15:625983. https://doi.org/10.3389/fnhum.2021.625983.
Fares, Ahmed, Sheng-hua Zhong, and Jianmin Jiang. 2020. "Brain-Media: A Dual Conditioned and Lateralization Supported GAN (DCLS-GAN) towards Visualization of Image-Evoked Brain Activities." In Proceedings of the 28th ACM International Conference on Multimedia, 1764–1772, Seattle WA, USA.
Feng, Lichen, Hongwei Shan, Yueqi Zhang, and Zhangming Zhu. 2022. "An Efficient Model-Compressed EEGNet Accelerator for Generalized Brain-Computer Interfaces with near Sensor Intelligence." *IEEE Transactions on Biomedical Circuits and Systems* 16 (6): 1239–1249. https://doi.org/10.1109/TBCAS.2022.3215962.

Frey, Jérémy, Renaud Gervais, Stéphanie Fleck, Fabien Lotte, and Martin Hachet. 2014. "Teegi: Tangible EEG Interface." In Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, 301–308, Honolulu Hawaii, USA.
Galati, Alexia, Riley Schoppa, and Aidong Lu. 2021. "Exploring the SenseMaking Process through Interactions and FNIRS in Immersive Visualization." *IEEE Transactions on Visualization and Computer Graphics* 27 (5): 2714–2724. https://doi.org/10.1109/TVCG.2021.3067693.
Ganis, Giorgio, William L. Thompson, and Stephen M. Kosslyn. 2004. "Brain Areas Underlying Visual Mental Imagery and Visual Perception: An FMRI Study." *Cognitive Brain Research* 20 (2): 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012.
Gigerenzer, Gerd, and Peter M. Todd. ABC Research Group. 2000. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
Gould, John D., and Clayton Lewis. 1985. "Designing for Usability: Key Principles and What Designers Think." *Communications of the ACM* 28 (3): 300–311. https://doi.org/10.1145/3166.3170.
Ha, David, and Douglas Eck. 2017. "A Neural Representation of Sketch Drawings." Preprint. arXiv:1704.03477.
Han, J., Feng Shi, and P. R. N. Childs. 2016. "The Combinator: A Computer-Based Tool for Idea Generation." In DS 84: Proceedings of the DESIGN 2016 14th International Design Conference, 639–648, Dubrovnik, Croatia.
Han, Ji. 2018. "Combinational Creativity and Computational Creativity."
He, Kaiming, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. 2022. "Masked Autoencoders Are Scalable Vision Learners." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16000–16009, New Orleans, LA, USA.
Heusel, Martin, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. "Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium." In Advances in Neural Information Processing Systems, Vol. 30, Long Beach, CA, USA.
Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. "Denoising Diffusion Probabilistic Models." In Advances in Neural Information Processing Systems, Vol. 33, 6840–6851, Vancouver, BC, Canada.
Horvitz, Eric. 1999. "Principles of Mixed-Initiative User Interfaces." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 159–166, Pittsburgh, PA, USA.
Jayaram, Vinay, and Alexandre Barachant. 2018. "MOABB: Trustworthy Algorithm Benchmarking for BCIs." *Journal of Neural Engineering* 15 (6):066011. https://doi.org/10.1088/1741-2552/aadea0.
Jiao, Zhicheng, Haoxuan You, Fan Yang, Xin Li, Han Zhang, and Dinggang Shen. 2019. "Decoding EEG by Visual-Guided Deep Neural Networks." International Joint Conferences on Artificial Intelligence Organization, 1387–1393, Macao.
Jin, Jian, Mingyue Yang, Huicong Hu, Xin Guo, Jianxi Luo, and Ying Liu. 2000. "Empowering Design Innovation Using AI-generated Content." *Journal of Engineering Design* 36 (1): 1–18. https://doi.org/10.1080/09544828.2024.2401751.
Karimi, Pegah, Mary Lou Maher, Kazjon Grace, and Nicholas Davis. 2018. "A Computational Model for Visual Conceptual Blends." *IBM Journal of Research and Development* 63 (1): 5–1.
Karras, Tero, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. "Alias-Free Generative Adversarial Networks." In Advances in Neural Information Processing Systems, Vol. 34, 852–863, New Orleans, LA, USA.
Karras, Tero, Samuli Laine, and Timo Aila. 2019. "A Style-Based Generator Architecture for Generative Adversarial Networks." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4401–4410, Long Beach, CA, USA.
Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. "Analyzing and Improving the Image Quality of Stylegan." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8110–8119, Seattle, WA, USA.
Kavasidis, Isaak, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. 2017. "Brain2image: Converting Brain Signals into Images." In Proceedings of the 25th ACM International Conference on Multimedia, 1809–1817, Mountain View, CA, USA.
Khare, Sanchita, Rajiv Nayan Choubey, Loveleen Amar, and Venkanna Udutalapalli. 2022. "NeuroVision: Perceived Image Regeneration Using CProGAN." *Neural Computing and Applications* 34 (8): 5979–5991. https://doi.org/10.1007/s00521-021-06774-1.

Kim, Hyunsu, Yunjey Choi, Junho Kim, Sungjoo Yoo, and Youngjung Uh. 2021. “Exploiting Spatial Dimensions of Latent in Gan for Real-Time Image Editing.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 852–861, Nashville, TN, USA.

Kosmyna, Nataliya, Jussi T Lindgren, and Anatole Lécuyer. 2018. “Attending to Visual Stimuli versus Performing Visual Imagery as a Control Strategy for EEG-based Brain-Computer Interfaces.” *Scientific Reports* 8 (1): 13222. https://doi.org/10.1038/s41598-018-31472-9.

Kosti, Makrina Viola, Kostas Georgiadis, Dimitrios A. Adamos, Nikos Laskaris, Diomidis Spinellis, and Lefteris Angelis. 2018. “Towards an Affordable Brain Computer Interface for the Assessment of Programmers’ Mental Workload.” *International Journal of Human-Computer Studies* 115:52–66. https://doi.org/10.1016/j.ijhcs.2018.03.002.

Kumar, Akshay, Lin Gao, Jiaming Li, Jiaxin Ma, Jianming Fu, Xudong Gu, Seedahmed S. Mahmoud, and Qiang Fang. 2022. “Error-Related Negativity-Based Robot-Assisted Stroke Rehabilitation System: Design and Proof-of-Concept.” *Frontiers in Neurorobotics* 16:837119. https://doi.org/10.3389/fnbot.2022.837119.

Lan, Yu-Ting, Kan Ren, Yansen Wang, Wei-Long Zheng, Dongsheng Li, Bao-Liang Lu, and Lili Qiu. 2023. “Seeing through the Brain: Image Reconstruction of Visual Perception from Human Brain Signals.” Preprint. arXiv:2308.02510.

Lawhern, Vernon J., Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. 2018. “EEGNet: A Compact Convolutional Neural Network for EEG-based Brain–computer Interfaces.” *Journal of Neural Engineering* 15 (5):056013. https://doi.org/10.1088/1741-2552/aace8c.

Lee, Johnny Chung, and Desney S. Tan. 2006. “Using a Low-Cost Electroencephalograph for Task Classification in HCI Research.” In Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, 81–90, Montreux, Switzerland.

Li, Pengzhi, Qinxuan Huang, Yikang Ding, and Zhiheng Li. 2023. “Layerdiffusion: Layered Controlled Image Editing with Diffusion Models.” In SIGGRAPH Asia 2023 Technical Communications, 1–4, Sydney, Australia.

Li, Wei, Weidi Zhang, Weilu Wu, and Jie Xu. 2025. “Exploring Human-Machine Collaboration Paths in the Context of AI-generation Content Creation: A Case Study in Product Styling Design.” *Journal of Engineering Design* 36 (2): 298–324. https://doi.org/10.1080/09544828.2024.2396199.

Maguire, Moira, and Brid Delahunt. 2017. “Doing a Thematic Analysis: A Practical, Step-by-step Guide for Learning and Teaching Scholars.” *All Ireland Journal of Higher Education* 9 (3): 335–344.

Markovinović, Ivan, Miroslav Vrankić, Saša Vlahinić, and Zoran Šverko. 2022. “Design Considerations for the Auditory Brain Computer Interface Speller.” *Biomedical Signal Processing and Control* 75:103546. https://doi.org/10.1016/j.bspc.2022.103546.

Mougenot, Céline, Carole Bouchard, and Améziane Aoussat. 2008. “Inspiration, Images and Design: A Field Investigation on Information-Retrieval Strategies by Designers.” *Journal of Design Research* 7 (4): 331–351. https://doi.org/10.1504/JDR.2008.026987.

Mukherjee, Pranay, Abhirup Das, Ayan Kumar Bhunia, and Partha Pratim Roy. 2019. “Cogni-Net: Cognitive Feature Learning through Deep Visual Perception.” In 2019 IEEE International Conference on Image Processing (ICIP), 4539–4543, Taipei, Taiwan.

Paulo, Joao Ruivo, Gabriel Pires, and Urbano J. Nunes. 2021. “Cross-Subject Zero Calibration Driver’s Drowsiness Detection: Exploring Spatiotemporal Image Encoding of EEG Signals for Convolutional Neural Network Classification.” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*29:905–915. https://doi.org/10.1109/TNSRE.2021.3079505.

Pearson, Joel. 2019. “The Human Imagination: The Cognitive Neuroscience of Visual Mental Imagery.” *Nature Reviews Neuroscience* 20 (10): 624–634. https://doi.org/10.1038/s41583-019-0202-9.

Qiao, Han, Vivian Liu, and Lydia Chilton. 2022. “Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art.” In Proceedings of the 14th Conference on Creativity and Cognition, 15–28, Venice, Italy.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. “Learning Transferable Visual Models from Natural Language Supervision.” https://arxiv.org/abs/2103.00020.

Riccio, Angela, Francesca Schettini, Valentina Galiotta, Enrico Giraldi, Maria Grazia Grasso, Febo Cincotti, and Donatella Mattia. 2022. "Usability of a Hybrid System Combining P300-based Brain-Computer Interface and Commercial Assistive Technologies to Enhance Communication in People with Multiple Sclerosis." *Frontiers in Human Neuroscience* 16:868419. https://doi.org/10.3389/fnhum.2022.868419.

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. "High-Resolution Image Synthesis with Latent Diffusion Models." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684–10695, New Orleans, LA, USA.

Salimans, Tim, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. "Improved Techniques for Training Gans." In Advances in Neural Information Processing Systems, Vol. 29, Barcelona, Spain.

Schön, Donald A. 2017. *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Routledge.

Shi, Jingyu, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. 2023. "An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions." Preprint. arXiv:2310.07127.

Shi, Yichun, Xiao Yang, Yangyue Wan, and Xiaohui Shen. 2022. "Semanticstylegan: Learning Compositional Generative Priors for Controllable Image Synthesis and Editing." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11254–11264, New Orleans, LA, USA.

Shneiderman, Ben. 2007. "Creativity Support Tools: Accelerating Discovery and Innovation." *Communications of the ACM* 50 (12): 20–32. https://doi.org/10.1145/1323688.1323689.

Singh, Prajwal, Pankaj Pandey, Krishna Miyapuram, and Shanmuganathan Raman. 2023. "EEG2IMAGE: Image Reconstruction from EEG Brain Signals." In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1–5, Rhodes Island, Greece.

Siqueira, Elton Sarmanho, Marcos Cordeiro Fleury, Marcus Vinicius Lamar, Anders Drachen, Carla Denise Castanho, and Ricardo Pezzuol Jacobi. 2023. "An Automated Approach to Estimate Player Experience in Game Events from Psychophysiological Data." *Multimedia Tools and Applications* 82 (13): 19189–19220. https://doi.org/10.1007/s11042-022-13845-5.

Song, Yonghao, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. 2022. "EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31:710–719. https://doi.org/10.1109/TNSRE.2022.3230250.

Spampinato, Concetto, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. 2017. "Deep Learning Human Mind for Automated Visual Classification." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6809–6817, Honolulu, HI, USA.

Sun, Zhida, Zhenyao Zhang, Yue Zhang, Min Lu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2025. "Creative Blends of Visual Concepts." In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1–17, Yokohama, Japan.

Sweller, John. 1988. "Cognitive Load during Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257–285. https://doi.org/10.1207/s15516709cog1202_4.

Tanaka, Yudai, Angela Vujic, Pattie Maes, Robert J. K. Jacob, Olaf Blanke, Sho Nakagome, and Pedro Lopes. 2024. "NeuroCHI: Are We Prepared for the Integration of the Brain with Computing?" In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 1–5, Honolulu, HI, USA.

Tang, Wen-Yu, Ze-Rui Xiang, Shu-Lan Yu, Jin-Yi Zhi, and Zhi Yang. 2025. "PSR-GAN: a Product Concept Sketch Rendering Method Based on Generative Adversarial Network and Colour Tags." *Journal of Engineering Design* 36:1801–1823. https://doi.org/10.1080/09544828.2025.2450760.

Tang, Wenzhe, Shanguang Chen, Junkai Shao, Xiaoxi Du, Jinchun Wu, Xinyu Zhang, Yafeng Niu, and Chengqi Xue. 2024. "Emerging Themes and Future Directions in Neurodesign and Human-Computer Interaction: A Systematic Review." *International Journal of Human–Computer Interaction* 41:1652–1673. https://doi.org/10.1080/10447318.2024.2362486.

Tenner, Edward. 2015. "The Design of Everyday Things by Donald Norman." *Technology and Culture* 56 (3): 785–787. https://doi.org/10.1353/tech.2015.0104.

Tirupattur, Praveen, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah. 2018. "Thoughtviz: Visualizing Human Thoughts Using Generative Adversarial Network." In Proceedings of the 26th ACM International Conference on Multimedia, 950–958, Seoul, South Korea.
Vujic, Angela, Shreyas Nisal, and Pattie Maes. 2023. "Joie: A Joy-Based Brain-Computer Interface (BCI)." In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 1–14, San Francisco, CA, USA.
Wang, Liang, Zhe Huang, Ziyu Zhou, Devon McKeon, Giles Blaney, Michael C. Hughes, and Robert J. K. Jacob. 2021. "Taming fNIRS-based BCI Input for Better Calibration and Broader Use." In The 34th Annual ACM Symposium on User Interface Software and Technology, 179–197, Virtual Event.
Wang, Liang, Jiayan Zhang, Jinyang Liu, Devon McKeon, David Guy Brizan, Giles Blaney, and Robert J. K. Jacob. 2024. "Empower Real-World BCIs with NIRS-X: An Adaptive Learning Framework That Harnesses Unlabeled Brain Signals." In Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, 1–16, Pittsburgh, PA, USA.
Wang, Pan, Xun Zhang, Liyan Wei, Peter Childs, Stephen Jia Wang, Yike Guo, and Maaike Kleinsmann. 2025. "Human-AI Co-ideation via Combinational Generative Model." *Journal of Engineering Design* 37: 458–494.
Wang, Zhouyang, Zhengyu Tan, and Yujia Ma. 2025. "Combinediff: A GenAI Creative Support Tool for Image Combination Exploration." *Journal of Engineering Design* 36 (5-6): 731–757. https://doi.org/10.1080/09544828.2024.2358722.
Ware, Colin. 2019. *Information Visualization: Perception for Design*. Burlington, MA: Morgan Kaufmann.
Weiskopf, Nikolaus, Ranganatha Sitaram, Oliver Josephs, Ralf Veit, Frank Scharnowski, Rainer Goebel, Niels Birbaumer, Ralf Deichmann, and Klaus Mathiak. 2007. "Real-Time Functional Magnetic Resonance Imaging: Methods and Applications." *Magnetic Resonance Imaging* 25 (6): 989–1003. https://doi.org/10.1016/j.mri.2007.02.007.
Williamson, John, Roderick Murray-Smith, Benjamin Blankertz, Matthias Krauledat, and K.-R. Müller. 2009. "Designing for Uncertain, Asymmetric Control: Interaction Design for Brain–computer Interfaces." *International Journal of Human-Computer Studies* 67 (10): 827–841. https://doi.org/10.1016/j.ijhcs.2009.05.009.
Wilson, Holly, Xi Chen, Mohammad Golbabaee, Michael J. Proulx, and Eamonn O'Neill. 2024. "Feasibility of Decoding Visual Information from EEG." *Brain-Computer Interfaces* 11 (1-2): 33–60. https://doi.org/10.1080/2326263X.2023.2287719.
Winograd, Terry, and Fernando Flores. 1986. *Understanding Computers and Cognition: A New Foundation for Design*. Vol. 335. Norwood, NJ: Ablex Publishing Corporation.
Wolpaw, Jonathan R., Dennis J. McFarland, Theresa M. Vaughan, and Gerwin Schalk. 2003. "The Wadsworth Center Brain-Computer Interface (BCI) Research and Development Program." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11 (2): 1–4. https://doi.org/10.1109/TNSRE.7333.
Wu, Jianfeng, Yuting Cai, Tingyu Sun, Keer Ma, and Chunfu Lu. 2025. "Integrating AIGC with Design: Dependence, Application, and Evolution-a Systematic Literature Review." *Journal of Engineering Design* 36 (5-6): 758–796. https://doi.org/10.1080/09544828.2024.2362587.
Xiao, Ping, and Simo Linkola. 2015. "Vismantic: Meaning-Making with Images." International Conference on Computational Creativity, Park City, Utah, USA.
Xie, Siying, Daniel Kaiser, and Radoslaw M. Cichy. 2020. "Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band." *Current Biology* 30 (13): 2621–2627. https://doi.org/10.1016/j.cub.2020.04.074.
Yang, Binxin, Shuyang Gu, Bo Zhang, Ting Zhang, Xuejin Chen, Xiaoyan Sun, Dong Chen, and Fang Wen. 2023. "Paint by Example: Exemplar-Based Image Editing with Diffusion Models." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18381–18391, Vancouver, BC, Canada.
Yang, Cheng, Shiyu Wang, Yiteng Peng, Zhichao Zhang, Lei Kong, Chuyi Zhou, Ye Tao, and Xiaoyu Chen. 2024. "Online BCI Systems: Cross-Subject Motor Imagery Classification Based on Weighted Time-Domain Feature Extraction Methods." *Journal of Engineering Design* 35 (6): 685–708. https://doi.org/10.1080/09544828.2024.2326396.

Yang, Shih-Wei, Chern-Sheng Lin, Shir-Kuan Lin, and Chi-Hung Lee. 2013. “Design of Virtual Keyboard Using Blink Control Method for the Severely Disabled.” *Computer Methods and Programs in Biomedicine* 111 (2): 410–418. https://doi.org/10.1016/j.cmpb.2013.04.012.

Zeng, Hong, Nianzhang Xia, Dongguan Qian, Motonobu Hattori, Chu Wang, and Wanzeng Kong. 2023. “DM-RE2I: a Framework Based on Diffusion Model for the Reconstruction from EEG to Image.” *Biomedical Signal Processing and Control* 86:105125. https://doi.org/10.1016/j.bspc.2023.105125.

Zeng, Hong, Nianzhang Xia, Ming Tao, Deng Pan, Haohao Zheng, Chu Wang, Feifan Xu, Wael Zakaria, and Guojun Dai. 2023. “DCAE: A Dual Conditional Autoencoder Framework for the Reconstruction from EEG into Image.” *Biomedical Signal Processing and Control* 81:104440. https://doi.org/10.1016/j.bspc.2022.104440.

Zhang, Lekai, Hailong Zhang, Peipei Yao, Yunhan Xu, Xuan Dong, Fo Hu, and Chunyang Yu. 2025. “Exploring the Efficacy of Olfactory Stimuli in Angry Driving through EEG Analysis.” *Journal of Engineering Design* 37: 1–22.

Zhao, Nanxuan, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson W. H. Lau, Jose Echevarria, and Zoya Bylinskii. 2020. “Iconate: Automatic Compound Icon Generation and Ideation.” In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13, Honolulu, HI, USA.

Zheng, Xiao, and Wanzhong Chen. 2021. “An Attention-Based Bi-LSTM Method for Visual Object Classification via EEG.” *Biomedical Signal Processing and Control* 63:102174. https://doi.org/10.1016/j.bspc.2020.102174.

Zheng, Xiao, Wanzhong Chen, Yang You, Yun Jiang, Mingyang Li, and Tao Zhang. 2020. “Ensemble Deep Learning for Automated Visual Classification Using EEG Signals.” *Pattern Recognition* 102:107147. https://doi.org/10.1016/j.patcog.2019.107147.

Zhu, Shuqi, Ziyi Ye, Qingyao Ai, and Yiqun Liu. 2024. “EEG-ImageNet: An Electroencephalogram Dataset and Benchmarks with Image Visual Stimuli of Multi-granularity Labels.” Preprint. arXiv:2406.07151.

Appendices

Appendix 1. Technical details of the proposed method

This paper presents a novel brain-guided visual blends technique composed of two main stages. In the first stage, an EEG encoder is trained on a large-scale unlabelled EEG dataset, and a diffusion-based generation model is fine-tuned using an EEG-to-image paired dataset. This training enables the EEG-to-image model to translate brain activity into corresponding visual representations. The second stage introduces an image local editing model that allows designers to blend the EEG-generated reference image with a source background. This visual blending process is guided by a user-specified target mask region to produce the final visual blends. The overall framework is illustrated in Figure 2.

A.1 EEG auto-encoder

Humans exhibit distinct brain responses when exposed to various visual stimuli, and these signals can be recorded using non-invasive techniques like EEG (Kosmyna, Lindgren, and Lécuyer,2018). It is a cost-effective method that uses electrodes attached to the scalp to capture the brain’s electrical activity, providing a valuable medium for studying the relationship between neural signals and stimuli (Frey et al.,2014, Lee and Tan,2006). EEG offers high temporal resolution, which is essential for detecting rapid changes in brain activity that occur within milliseconds. However, EEG data is often characterised by significant variability and uncertainty due to factors such as age, physical condition, diseases, time, and cognitive state. Additionally, EEG signals are highly susceptible to external environmental noise, which can compromise data accuracy.

Given these limitations, traditional machine learning methods struggle to effectively encode and extract meaningful features from noisy EEG data. The Masked Autoencoder (MAE) (He et al.,2022), originally developed for image encoding, offers a robust solution for handling high variability and noise. It operates by masking portions of an image and reconstructing the missing regions, enabling

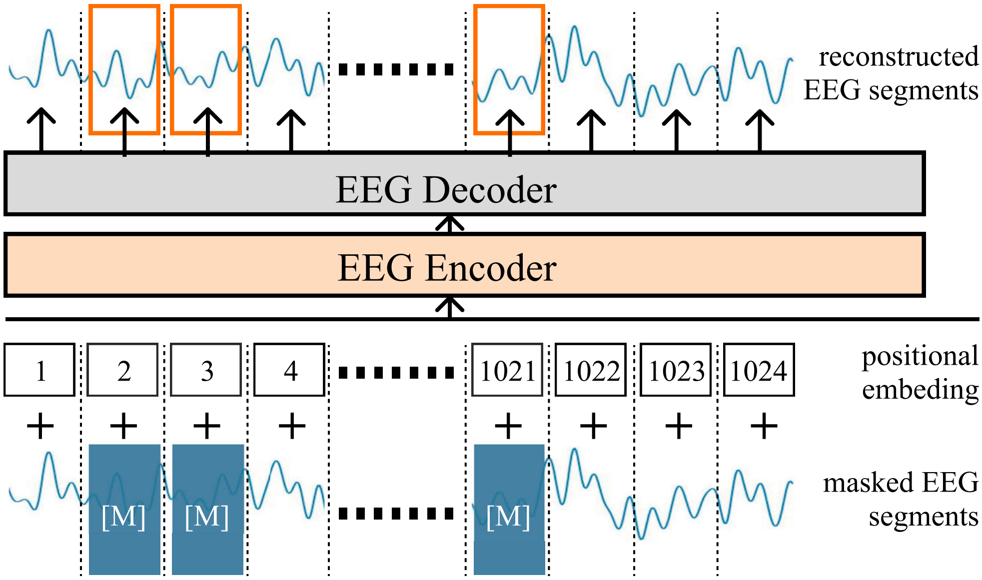

Figure A1. The overall training pipeline of the EEG autoencoder.

The EEG signals are tokenised to produce a set of tokens. Then 75% of these tokens are randomly masked and go through the EEG auto-encoder to reconstruct the masked tokens.

the encoder to learn relevant contextual information from complex data. Inspired by Mind-vis (Z. Chen et al., 2023), which employs MAE to train an fMRI autoencoder for extracting high-quality features from fMRI signals, we propose adapting the MAE architecture with transformer blocks for EEG signal encoding in order to extract high-quality features from EEG data.

The overall pipeline of the auto-encoder training is shown in Figure A1. Initially, EEG signals are segmented into multiple continuous tokens along the time dimension. These tokens are then subjected to random masking based on a predetermined mask ratio, with the original EEG signals serving as target for updating the model parameters. Each EEG signal token is transformed into feature embedding via a convolutional layer. A decoder, mirroring the structure of the encoder, is employed to predict the masked tokens and reconstruct the entire EEG signal. The loss is calculated using the differences between the predicted EEG signal with the original EEG signal. The formula is shown as follows:

$$Loss_{auto-encoder} = \frac{1}{n}\sum_{i=1}^{n} \|EEG_{predict} - EEG_{label}\|^2 \tag{A1}$$

by accomplishing the task of EEG signal reconstruction, the EEG auto-encoder uncovers the relationships within EEG signal, which then serves as the feature extractor for fine-tuning stable diffusion (SD) (Rombach et al., 2022).

A.2 Stimuli reconstruction via generation model fine-tuning

SD (Rombach et al., 2022) is an image generation model that operates in the latent feature space, unlike earlier diffusion models (Ho, Jain, and Abbeel, 2020) that work at the pixel-level. This shift significantly lowers the computational resources required for training. SD is renowned for its stability and high degree of controllability, enabling precise manipulation of generated images through diverse conditional inputs, such as text, images, audio, and semantic maps. These inputs are integrated into the U-net via a cross-attention conditioning mechanism, guiding the denoising process to ensure that the generated images align with the desired conditions.

We propose a method that modifies the conditional channel of SD v1.5 by replacing the CLIP (Radford et al.,2021) encoder with the pre-trained EEG encoder. By jointly training the EEG encoder and the cross-attention layers of SD, we enable a transformation of the input modality from text to brain activity. We use the publicly available ImageNet-EEG dataset (Zhu et al.,2024), which consists of EEG-image paired data, allowing SD to learn the mapping from brain activity to visual images. Specifically, given an EEG signal x, it is first processed by the encoder $\mathcal{E}$, as shown in the following equation:

$$c = \mathcal{E}(x) \tag{A2}$$

where c represents the latent features used as conditional input for the U-Net. These features are integrated into each layer of the U-Net through cross-attention modules and skip connections, enabling the alignment of brain signal with the target image. The latent condition c is first fed into a linear layer $f(\cdot)$ to generate a feature embedding $C_{emb} = f(c)$. The formula of attention layer is $Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d}}) \cdot V$,

$$Q = W_Q^{(i)} \cdot \varphi_i(z_t), \quad K = W_K^{(i)} \cdot C_{emb}, \quad V = W_V^{(i)} \cdot C_{emb} \tag{A3}$$

where $\varphi_i(z_t)$ stands for the value of image latent z_t in the U-net model layers. W_Q, W_K, W_V are the learnable parameters for the feature embedding layers. We freeze the Unet model weights and update the EEG encoder and cross-attention layers. The loss used for the weight update is similar to the SD with an additional EEG feature accuracy loss. The formulas are shown as follows:

$$loss_{SD} = \mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(x_t, t, \tau_\theta(y))\|_2^2\right] \tag{A4}$$

$$loss_{ef} = Cosine\ Similarity(Emb_e eg,\ Emb_i mage) \tag{A5}$$

$$loss_{total} = loss_{SD} + loss_{ef} \tag{A6}$$

where ϵ_θ is the U-net model for the denoising process.

A.3 Image editing

In this study, we use the local editing functionality of Paint-by-Example (B. Yang et al.,2023) to enable automated visual blending tasks. Image local editing is an example-based image editing approach that allows users to specify modification areas on a background reference image by drawing a mask. The user then selects a suitable source image to indicate the desired semantic content for the masked region. Based on the background reference image, the mask, and the source image, the model generates a new visual blend. The mask, which can be of any closed shape, is represented as a binary mask m that defines the editable regions (1 for editable, 0 for non-editable). Let the source image, reference image, and mask be denoted as x_s, x_r, m. The model processes these inputs to produce an output image $y = \mathcal{F}(x_s, x_r, m)$. In the resulting image y, regions where $m = 1$ closely match the semantic content of x_r, while regions where $m = 0$ preserve the features of x_s.

The model leverages a self-supervised learning approach, where an image and an annotated bounding box serve as training data. Here, the mask defines the bounding box, with $x_r = x_s \cdot m_x$ representing the reference content and x_s as the original input image. This method effectively addresses data scarcity by generating varied training samples.

The inference step of the Paint-by-example shown in Figure 2 Module C, typically include a reference image, a source image, and a mask. In our framework, we adapt this model by using an image generated from EEG signals as the source image. Designers select a reference image and then draw a mask to specify the areas to be replaced by the source image. They have the freedom to create masks of any size or shape, enabling precise control over the final design. This approach deeply involves the designer's thoughts and ideas in design process, with the AI acting as a supporting tool to streamline visual blending task.

Appendix 2. Implementations

A.4 Dataset

A.4.1 MOABB platform

We used the MOABB (Mother of All BCI Benchmarks) (Aristimunha et al.,2023, Jayaram and Barachant,2018) platform to collect and preprocess EEG data for pre-training the MAE model. MOABB is a comprehensive platform designed for BCI research, offering a repository of publicly available EEG datasets, signal preprocessing tools, and state-of-the-art algorithms. It provides researchers with an efficient toolbox for rapid data preprocessing and model performance validation, all within a single platform. Utilising MOABB, we gathered and processed over 100,000 EEG samples, covering motor imagery, Rapid Serial Visual Presentation(RSVP), and Steady-State Visually Evoked Potentials(SSVEP) tasks from more than 400 subjects. Due to variations in signal length, channel count, and frequency across the datasets, we used MOABB's preprocessing tools to standardise the data and ensure consistency in the training samples. All EEG signals were standardised to a length of 512. When the number of channels was insufficient, existing channels were duplicated to reach a total of 128 channels. All signals were filtered to a frequency range of 5 to 95 Hz. The MAE pre-training enables the EEG encoder to learn a universal representation of EEG data across diverse datasets, thereby enhancing its ability to extract meaningful features.

A.4.2 ImageNet-EEG dataset

The RSVP paradigm is a psychological experimental method used in BCI research to present visual stimuli. It involves displaying a rapid sequence of numbers, letters, words, or images at the same spatial location, requiring participants to distinguish between them. The brain signals evoked by these stimuli are recorded using EEG devices, establishing a direct link between the brain activity and the presented images. The ImageNet-EEG dataset (Zhu et al.,2024) pairs these EEG signals with corresponding stimulus images sourced from the ImageNet dataset. It contains 2,000 images from 40 categories, with 50 images in each category. During the experiment, each image is shown for 0.5 seconds, followed by a 10-second break after every 50 images. The EEG signals are captured using a 128-channel Brainvision device from 6 participants. To optimise the performance of the pre-trained EEG encoder model, the EEG signals in the ImageNet-EEG dataset undergo the same preprocessing steps as previously outlined. This process yields 120,000 EEG signals-image pairs for training.

A.5 Implementation details

Before training, all pre-trained EEG data were filtered to a frequency range of 5–95 Hz, and the signals were manually segmented to a length of 512. Each transformer token contains a one-dimensional vector of length 1024, generated from four consecutive time points in the EEG signal using a projection layer. Masked portions of the vector are replaced with randomly generated one-dimensional vectors of the same length. During the pre-training phase, the Mean Squared Error (MSE) between the predicted masked patches and the original EEG data is used as the loss function to update the model's weights. A mask ratio of 75% was applied, with a learning rate of 0.0025 and a decay rate of 0.05, over 500 epochs. After pre-training, only the encoder is retained for extracting meaningful features from the EEG signals, while the decoder is excluded from further steps. For the fine-tuning of the SD model, a learning rate of 0.00053 with a decay rate of 0.05 was employed, and the model was refined over 300 epochs on the pre-trained SD model weight.