Generative AI Methodology for the Conceptual Topology Design of Bridges from Topography Images

Aswin Lal[1]; Chongjie Kang[2]; Baris Özcan[3]; Jörg Blankenbach[4]; and Steffen Marx[5]

Abstract: The planning process for bridge construction is time-consuming, prone to errors, and cost-intensive, because bridges always have to be individually adapted to specific boundary conditions, for example, the respective topography situation. Despite sharing the same fundamental physical principles, bridge structures can take different forms due to regional variations in engineering expertise. Existing computational design approaches for bridges are predominantly rule-based or parametric. To support early-stage decision-making, successful practices from around the world could be analyzed using image-based, data-driven artificial intelligence (AI) methods. Hence, this work presents an alternative, result-oriented, data-driven method based on generative AI to assist engineers in the conceptual design phase of bridge construction. First, the research background and related works in structural design, with a focus on bridge engineering, are elaborated. Subsequently, we propose a multimodal generative model combining pix2pix and BERT to map valley topographies and textual design specifications to conceptual bridge topology configurations. The approach leverages image-based terrain information and text-based design descriptors to infer structurally plausible configurations from precedent data, rather than relying on explicitly formulated design rules. The results show that the model can reproduce multiple topology patterns consistent with the training data and generate constraint-consistent configurations for topographies with similar geometric characteristics. The findings demonstrate that result-oriented, data-driven generative models can support early-stage bridge topology exploration under controlled conditions, providing a complementary alternative to parametric design workflows. The paper concludes by outlining methodological implications and directions for extending the approach toward broader applicability. **DOI: 10.1061/JCCEE5.CPENG-7429** *This work is made available under the terms of the Creative Commons Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0/.*

Author keywords: Generative AI; Bridge design; Topology; Topography; Image-to-image; Text-to-image.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Introduction

Bridges generally are considered to be highly sophisticated structures and are carefully planned from a technical point of view (Kang et al. 2018). The structural design of a bridge consists of four primary stages: conceptual design (schematic design), preliminary design, detailed design, and construction drawing development (Tang 2017). Among these, the purpose of the conceptual design phase is to determine and document the most suitable solution for a roadway to cross a stream, road, or other facility while considering relevant issues, risks, and constraints and exploring all potential options. As a result, the conceptual design stage establishes feasibility and explores design alternatives. This task demands not only strong expertise in structural and bridge engineering but also individual creativity and intuition for good design. It commonly is carried out by highly experienced bridge engineers and architects. As an essential step, the variants are developed based on the terrain cross section of the topographic location and other given constraints, including boundary conditions and requirements (Fig. 1). The conceptual design phase has a decisive influence on the final outcome from the very beginning for bridges as well as other engineering structures (Liao et al. 2024; Qian et al. 2024).

This variant-finding process commonly is guided by both engineering principles and creative intuition. Engineers rely on mechanical and structural knowledge that adhere to regulations, guidelines, norms, and standards that define boundary conditions such as environmental impacts and traffic loads in different countries and regions (Schneider and Marx 2018), but engineers and architects also incorporate personal intuition and creativity to achieve well-balanced designs that are both aesthetically and functionally sound. These aspects must be integrated effectively in order to create optimal results. Commonly, bridge engineers explore various bridge variants at the conceptual stage within specific constraints, particularly regarding topography, to identify the most suitable topology. This is done manually and often is time consuming and highly dependent on the individual experience of the designer. However, the expertise of individual bridge engineers usually is difficult to formalize, varies widely and is limited to regions and experience.

In this case, generative artificial intelligence (AI) [e.g., generative adversarial networks (GANs)] has shown enormous potential in creating design options for construction applications. The previously mentioned individual-dependent limitations can be overcome

[1]Research Associate, Chair of Computing in Civil Engineering and GIS and Geodetic Institute, Rheinisch-Westfälische Technische Hochschule Aachen Univ., Aachen 52074, Germany. Email: aswin.lal@gia.rwth-aachen.de

[2]Research Area Leader, Institute of Concrete Structures, Technische Universität Dresden, Dresden 01069, Germany (corresponding author). ORCID: https://orcid.org/0000-0003-2694-1776. Email: chongjie.kang@tu-dresden.de

[3]Research Associate, Chair of Computing in Civil Engineering and GIS and Geodetic Institute, Rheinisch-Westfälische Technische Hochschule Aachen Univ., Aachen 52074, Germany. ORCID: https://orcid.org/0000-0003-4667-8047. Email: oezcan@gia.rwth-aachen.de

[4]Professor, Chair of Computing in Civil Engineering and GIS and Geodetic Institute, Rheinisch-Westfälische Technische Hochschule Aachen Univ., Aachen 52074, Germany. ORCID: https://orcid.org/0000-0002-5700-8818. Email: blankenbach@gia.rwth-aachen.de

[5]Professor, Institute of Concrete Structures, Technische Universität Dresden, Dresden 01069, Germany. Email: steffen.marx@tu-dresden.de

Note. This manuscript was submitted on July 22, 2025; approved on February 19, 2026; published online on July 30, 2026. Discussion period open until December 30, 2026; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, ASCE, ISSN 0887-3801.

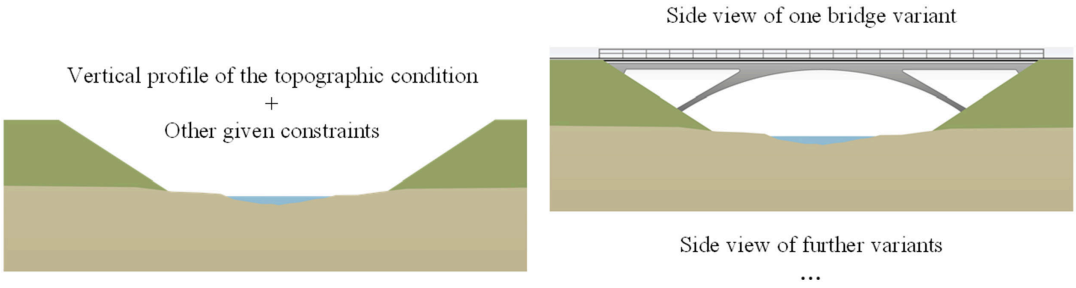


Fig. 1. Variant finding based on the cross section of the topographic condition.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

using the ability of generative models to learn from a wide variety of existing design varieties. This allows them to explore countless new solutions and ensure that the final output is optimally tailored to the specific design requirements. Moreover, this automated approach has the potential not only to increase efficiency but also to lower the professional entry barrier, thereby supporting young engineers in identifying optimal design variants. In addition, through natural language processing (NLP), such generative models can create multiple design options based on high-level textual constraints, thereby significantly reducing manual specifications. Other practical benefits of generative models include their seamless integration capabilities across different platforms and data scalability. Despite its adaptability and potential to produce innovative design solutions, generative AI remains largely unexplored for tasks related to bridge designing.

This research explored image- and text-based generative models to support the conceptual design of bridge topologies. We aimed to automate the conceptual design phase of bridge construction by training a generative model to create suitable bridge designs, for a given topography, subject to design constraints. This model takes topographic images and textual design specifications as input to generate bridge designs. The natural language text inputs provide the relevant design objectives passed on to the model by engineers to constrain the predictions of the model to suit the topography and meet its overall design requirements. In this paper, we focus on specific bridge topologies, namely frame and girder bridges, which are used commonly in highway applications. The primary objective was not to consider all possible design variants, but rather to explore the potential of the proposed generative AI-based design methodology as a proof of concept in creating suitable bridge topologies based on given topography images.

The contributions of this work mainly include the following:

1. Because no publicly available data sets exist for the task of generating bridge designs, a new data set was developed. It consists of paired images showing highway topographies (e.g., valley cross sections) without bridges and the corresponding topographies with bridge designs, derived from existing highway bridge examples.
2. A conditional generative adversarial network (CGAN) that combines topographical information in the form of images and topology information described as text to create bridge designs was developed. The well-established pix2pix architecture (Isola et al. 2017) for image-to-image translation was developed for this purpose. Although generative AI has shown immense success in image-to-image, text-to-image, image-to-text, and text-to-text tasks, combining text and images to generate images is a relatively underexplored application, especially in the civil engineering domain. We investigated the possibilities of generating suitable bridge design options constrained by text design specifications for a valley that is represented in the form an image. This allows engineers to use the ability of generative models to learn from a large number of existing valley and bridge scenarios, beyond their own experience, to facilitate the creation of new designs. Furthermore, it also enables engineers to use natural language prompts to guide the models using their expertise to ensure that the generated designs meet the specific design requirements.

The alternative method proposed in this paper does not depend on architectural form-making theories, but on the early-stage structural configuration of bridge systems (e.g., span arrangement, structural system type, and alignment with topography), which is governed by structural, geometric, and contextual constraints. In addition, bridge typologies are not treated as visual patterns, but as geometric representations of structurally governed solutions. The learning process is driven by structural outcomes embedded in the data rather than by visual similarity or aesthetic considerations. Accordingly, the method is not intended to generate creative forms in an architectural sense, but to infer feasible structural configurations based on existing design results.

The remainder of the paper is divided into the following sections: Section "Research Background and Related Works" discusses the research background and related works. Section "Methodology" describes the proposed methodology. The experimental setup is introduced in the section "Experimental Setup," along with our curated data set, the evaluation methods, and the training protocol. The results obtained from our experiments are analyzed qualitatively and quantitively in the section "Results and Evaluation," which also includes an ablation study comparing three other approaches and their comparisons with the proposed methodology. Section "Discussion" discusses the overall outcomes of the methodology based on the results. Section "Conclusion and Outlook" presents our conclusions and the future research possibilities on the topic.

Research Background and Related Works

AI in Civil Engineering

The automotive and aerospace industries have been using AI extensively and practically for several years. In contrast, the use of AI applications in the construction industry is still in its early stages. Nevertheless, the integration of AI in the domain of civil engineering offers numerous opportunities for the automation and optimization of modeling and evaluation processes in research, development, and practice (Kraus 2022). For example, applications of AI in structural health monitoring include the detection, localization, and classification of different types of damages using image data. Zunino et al. (2025) implemented K-means clustering for the detection and localization damages in bridges. The implementation of a portable internet-of-things (IoT) device for tire text code identification by Zhang et al. (2025) using AI-based object detection and natural language processing methods demonstrated how AI

can be used for the real-time monitoring and safety in construction logistics. However, Saka et al. (2024) mentioned the current lack of extensive knowledge in the use of large language models (LLMs) in the construction industry.

New approaches for the use of AI methods in the design process have been presented repeatedly, many of them primarily in the field of architectural design in recent years. Spiegelhalter (2023) discussed the use of AI-assisted topological optimization methods to reduce weight and cost in bridge designing. Sardone et al. (2022) proposed a performative structural design optimization (PSDO) method that uses algorithm-aided design (AAD) together with multiobjective evolutionary algorithms (MOEAs) for shape optimization of large-span voided beams. Patrik Schumacher, partner at Zaha Hadid Architects, stated that text-to-image tools such as DALL-E 2 are now being used in almost all the firm's projects (Barker 2023). Chen et al. (2023) and Horvath and Pouliou (2024) also emphasized the increasing importance of AI-based design generation and highlighted the wide range of possibilities in this area. Maksoud et al. (2023) aimed to integrate computational design technologies by using sensor data to create self-learning buildings that could adapt to environmental challenges. Conversely, our work aimed to integrate natural language with an image-based generative model to generate bridge designs leveraging the strengths of both text-based and image-based generative AI models.

Generative AI in Structural Design

As in the broader building sector, building structural design primarily relies on drawings represented as structural parameter, CAD, or image data. Engineers develop designs based on architectural concepts while considering constraints such as compliance and cost-efficiency. The end-to-end design process of generative AI aligns with traditional structural design, leveraging advanced learning and generation capabilities to address complex design challenges. Consequently, it also recently has attracted significant research interest. These AI-based design methods in building sector can be divided into two main categories based on data features (Liao et al. 2024):

1. tensor-based design: data as tensors [including one-dimensional vectors, three-dimensional (3D) images, and *n*-dimensional matrices]; and
2. graph-based design: data in graphs (node and edge representations, with one-dimensional vector features embedded in nodes and edges).

Tensor representations are more intuitive and straightforward, and therefore are used widely in current generative AI-based design methods. Although graphs are better suited for capturing topological features, they require more-complex designs to effectively represent patterns and size-related information. Using these different data resources, various innovative generative AI methods have been explored. Convolutional neural networks (CNNs) and graph neural networks (GNNs) are the primary AI technologies used for designing residential building structures such as shear walls, frames, and frame-brace structures.

For example, Maeda et al. (2021) and Li and Zhao (2023) trained a GAN and a CGAN to generate realistic road and concrete damage images, respectively. Jiang et al. (2023) used a GAN to generate building layouts. An extension of the pix2pix architecture, pix2pixHD (Wang et al. 2018), was used for floor layout plan recognition and generation (Huang and Zheng 2018). To increase the efficiency of the otherwise time-consuming process of house design generation, Nauata et al. (2020) trained a GAN called House-Gan to generate house layouts. Ghannad and Lee (2022) trained a coupled generative adversarial network (CoGAN) (Liu and Tuzel 2016) for the automated generation of modular housing designs. FrameGAN was proposed by Fu et al. (2023) to automate the component layout design of prefabricated steel frame–brace structures.

Moreover, diffusion models (Ramesh et al. 2021) and LLMs (Saka et al. 2024) recently have been explored by integrating text information in image data to enhance the effectiveness of intelligent design. For example, Liao et al. (2022) improved the encoder of the generator by using a fusion network that combines text features and image features. However, they are still limited in intelligent architectural structural design due to their relatively short development time and high hardware requirements as well as data shortage.

In addition, spatial structures often have a high degree of freedom and complexity in their shape; therefore, a variational autoencoder (VAE) generator network consisting of a latent space, an encoder, and decoder were explored to generate design schemes of 3D spatial structures (De Miguel Rodríguez et al. 2020). Currently, generative AI is used primarily to design the topology of simple beam components (Kallioras and Lagaros 2020), but it is used less commonly in the topological design of continuous engineering structures. A comprehensive literature review of generative AI design for building structures (Liao et al. 2024) indicated that generative models not only can understand the rules for generating structural designs based on architectural design, but they also can effectively learn the constraints of drawing and text data, as well as mechanical and empirical knowledge, and efficiently integrate these to generate designs. These models are capable of effectively incorporating drawing and textual constraints, as well as mechanical principles and empirical knowledge, enabling them to generate structurally sound design outputs efficiently. However, despite its growing prominence in the building sector, generative AI has been explored relatively little for bridge engineering.

Generative AI in Bridge Design

Unlike buildings, which often emphasize aesthetics, bridges are primarily functional structures designed predominantly by engineers rather than architects (Tang 2018). Their main purpose is to span physical obstacles and connect transport routes. More importantly, the topology of a bridge depends on its surroundings, especially on geographical features. This makes topography a crucial factor to consider in bridge designing. In addition, bridges are designed considering factors such as structural integrity, safety, and durability. However, although generative AI has proven its potential in the efficient creation of bridge designs, it rarely has been explored in this domain due to the complexity and limited availability of relevant parameters.

As a result of the pilot project Domain-aware-AI Augmented Design of Bridge Structures launched in 2022, Balmer et al. (2024) proposed a design exploration and explanation framework to augment the designer via a conditional variational autoendoder (CVAE). This model served as a forward performance predictor as well as an inverse design generator conditioned on a set of performance requests, for the conceptual design of a multispan prestressed concrete pedestrian bridge in Switzerland. The trained generative AI model automatically generated key design parameters, which then were integrated with parametric models and CAD programs to create a 3D visualization of bridge. To evaluate the performance of the generated bridge variants, different criteria regarding cost, weight, volume, and structural behavior such as displacements and eigenfrequencies were integrated. This approach was adjusted and applied to a one-span concrete frame bridge and a network tied arch bridge (Slepicka et al. 2022), demonstrating its adaptability to different structural types (Kraus et al. 2024).

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Recently there has been a surge in interest in using latent diffusion models for tasks related to image generation. For example, Mazé and Ahmed (2022) trained a diffusion model that outperforms generative models at structural topology optimization. Herron et al. (2023) implemented a latent diffusion model for topology optimization in the initial design phases of bridges. The distinct advantage of these models over conventional generative models is their ability to generalize better without suffering from mode collapse. However, such models typically are more data hungry when trained from scratch, and the training process itself is computationally more expensive.

As opposed to these innovations, our work focused on enabling a generative model to perform bridge design prediction using design constraints given in the form of natural language prompts. Other than the aforementioned studies, generative AI in bridge design rarely has been researched, to the best of our knowledge. It is evident that there is substantial untapped potential for generative AI in this field, which warrants further investigation.

Methodology

Our objective was to automatically generate bridge designs for given valley topographies. To achieve this, we first employed a pix2pix model trained on pairs of topography images without and with bridge topologies as inputs and targets, respectively. During inference, the model maps images of valley situations to bridge designs based on the designs on which it was trained. However, due to the deterministic nature of the generator's architecture, the prediction of such a model commonly is limited to a single design possibility. Consequently, the model consistently gives only one possible design option for an input image, even when it is presented multiple times. To overcome this limitation, we propose an advanced model as an extension of the pix2pix architecture, which incorporates constraints provided in textual form, in addition to images, to guide the generator in producing multiple bridge designs for the same input topography. In this context. the advanced model effectively combines pixel-level information from images with semantic input from text. In particular, this study considered the three most commonly constructed bridge types: (1) frame bridges with no pillars, (2) frame bridges with two inclined pillars, and (3) beam bridges with two straight pillars.

Fig. 2 illustrates the training procedure of the conventional pix2pix model (Isola et al. 2017). The generator is trained to map input images to its target representations, whereas the discriminator is trained to distinguish between real and fake images. Such a model effectively learns a mapping from the source domain to the target domain. Pix2pix has been applied successfully to tasks such as generating realistic images from edge maps, converting aerial images to maps, and converting segmentation maps into realistic street scene. One of the limitations we identified is the limited stochasticity in the generator's predictions.

However, when applying this model to generate bridge designs based on specific topographic images, the aforementioned limitation hinders bridge designers from exploring multiple design possibilities for spanning a given valley during the initial design phase. Guiding the generator to produce a diverse set of designs not only provides multiple design options to consider, but also enables the incorporation of relevant design specifications to constrain the generator's predictions. To achieve this objective, we modified the existing pix2pix architecture as shown in Fig. 3.

We modified the existing pix2pix architecture, which transfers image representations from one domain to another, by incorporating the textual representations given by a bidirectional encoder representations from transformers (BERT) (Devlin et al. 2018) model, pretrained on a large collection of data, including the BookCorpus (800 million words) (Zhu et al. 2015) and the English Wikipedia (2.5 billion words). Notably, the success of the transformer architecture using attention-mechanism proposed by Vaswani et al. (2017) has led to numerous advancements over the last several years, including the development of the transformer-XL model (Dai et al. 2019) and the compressive transformer (Rae et al. 2019). Among peer models such as the generative pretrained transformer (OpenAI GPT) (Radford et al. 2018), the BERT model has shown superior performance in processing sequential data due to its bidirectional approach and has demonstrated high adaptability for downstream applications such as text summarization, sentence classification, and sentiment analysis. This allows the BERT model to effectively interpret the design constraints expressed in the form of texts and guide the generator by providing semantically rich information.

Specifically, for the bridge topology designs considered in this work, the input to the generator consists of an image depicting a valley topography combined with a design constraint provided in textual form. The generator then is trained to learn a mapping between this combined input and a desirable bridge design. Learning such a mapping from the training data enables the generator to produce topology designs that preserve the spatial hierarchy and structural logic of bridges. Because the size of the input layer of the generator remains constant throughout training and inference, the input images are resized to match the required input dimensions. The generator therefore learns relative spatial relationships and proportions, rather than absolute real-world dimensions. In addition, because the input images vary in scale, the mapping learned by the generator does not encode the real-world scale. Because our work focuses on bridge topology generation in the initial design phase, the relative proportions and spatial relationships are more relevant than absolute scale and precise real-world dimensions. However, real-world dimensions could be reverse-calculated using a predefined scale ratio, which could be calculated and documented for down-sampling. Conversely, the discriminator is trained on

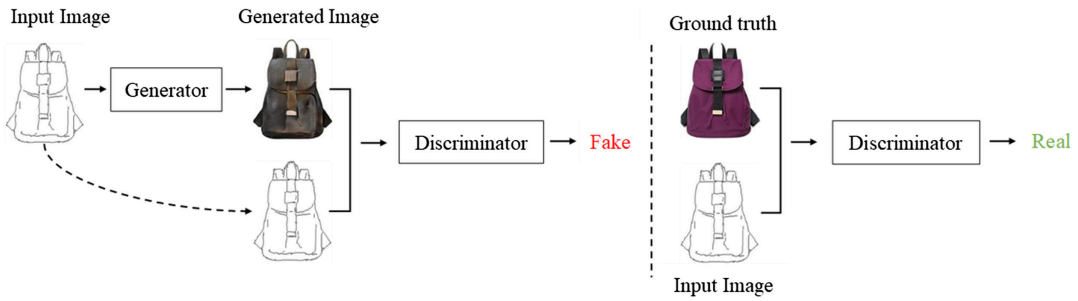


Fig. 2. Training procedure of the conventional pix2pix model visualized for an example from the Handbags data set.

J. Comput. Civ. Eng., 2026, 40(6): 04026089

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

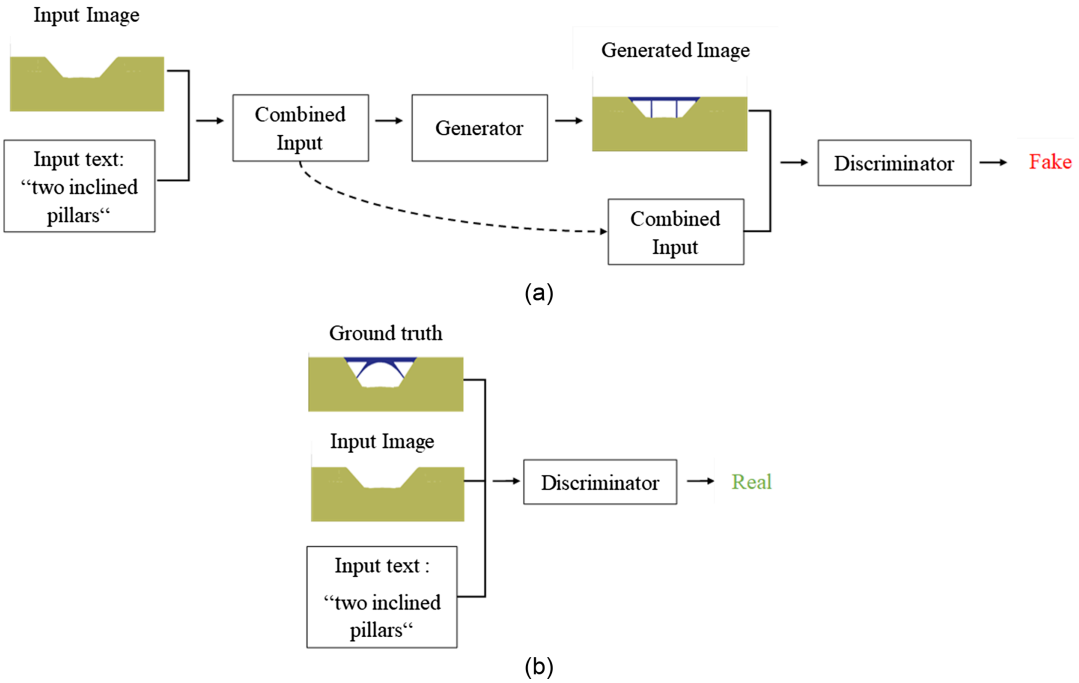


Fig. 3. Training procedure adopted for generating bridge designs: (a) the generator is fed with a combination of the image of a topography and a design specification (here, “two inclined pillars”); and (b) the discriminator is trained to classify the generated images as being fake and the ground truth images as being real for the given image and text combinations.

the input image and text combination against the ground truths to distinguish between real and generated bridge designs. Detailed descriptions of the generator and discriminator architectures are given in the sections “Generator” and “Discriminator,” respectively.

As previously mentioned, the inherent deterministic nature of the conventional pix2pix model restricts the generator to a single output for each input, limiting its ability to explore multiple design possibilities. In contrast, our approach incorporates text-based constraints, allowing users to guide the generator toward diverse and contextually relevant outputs. This makes text interpretation a critical component of the generation process. In this way, the generator can be trained to generate a broader variety of outcomes by carefully determining design constraints (Fig. 4). This approach also is aligned with practical bridge designing. In certain cases, such as standardized conditions or early-stage planning, predefined bridge topologies are used in common or constrained scenarios. These are based on established requirements or prior experience to enable rapid evaluation of costs, construction methods, and overall feasibility.

To facilitate a better understanding of the overall approach, the generator and discriminator are described in detail in the following sections.

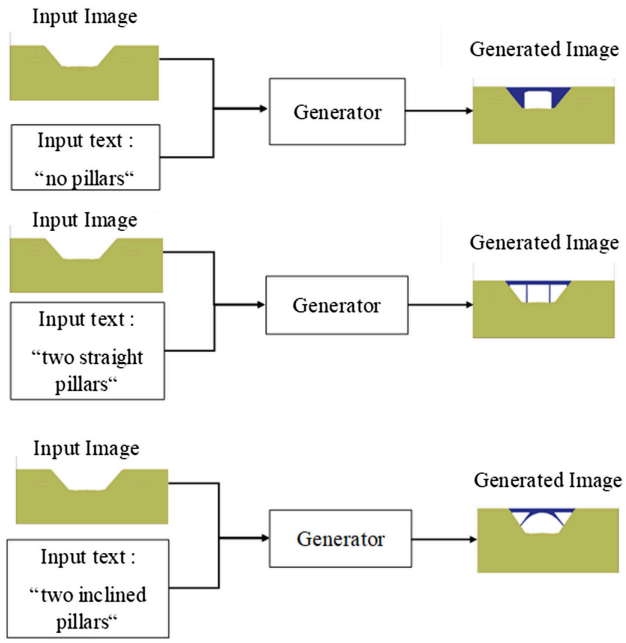


Fig. 4. Trained generator during inference.

Generator

The generator was designed as an extension of the U-Net architecture (Ronneberger et al. 2015) used in the conventional pix2pix generator, incorporating an additional mechanism to guide its predictions using the BERT model. For this purpose, a pretrained BERT model called bert-large-uncased, containing 340 million parameters, was used. Fig. 5 illustrates the detailed architecture of the generator.

For each input image, the corresponding text constraint is tokenized (broken down into smaller units called tokens) and fed into the BERT model. In addition to these tokens, the BERT model also contains two special tokens: [CLS], a classification token; and [SEP], a separator token. The final hidden-state of the [CLS] token provides a summarized version of the input sequence making it suitable for classification tasks. The [SEP] token is used to separate sentence pairs contained in the input sequence. This model is case-insensitive (i.e., it does not differentiate between uppercase and lowercase) and represents each token using a vector of size 1,024. The BERT model encodes text descriptions by creating a sequence of embeddings, which are numerical representations (stored as vectors) of a text input, containing its semantic meaning

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

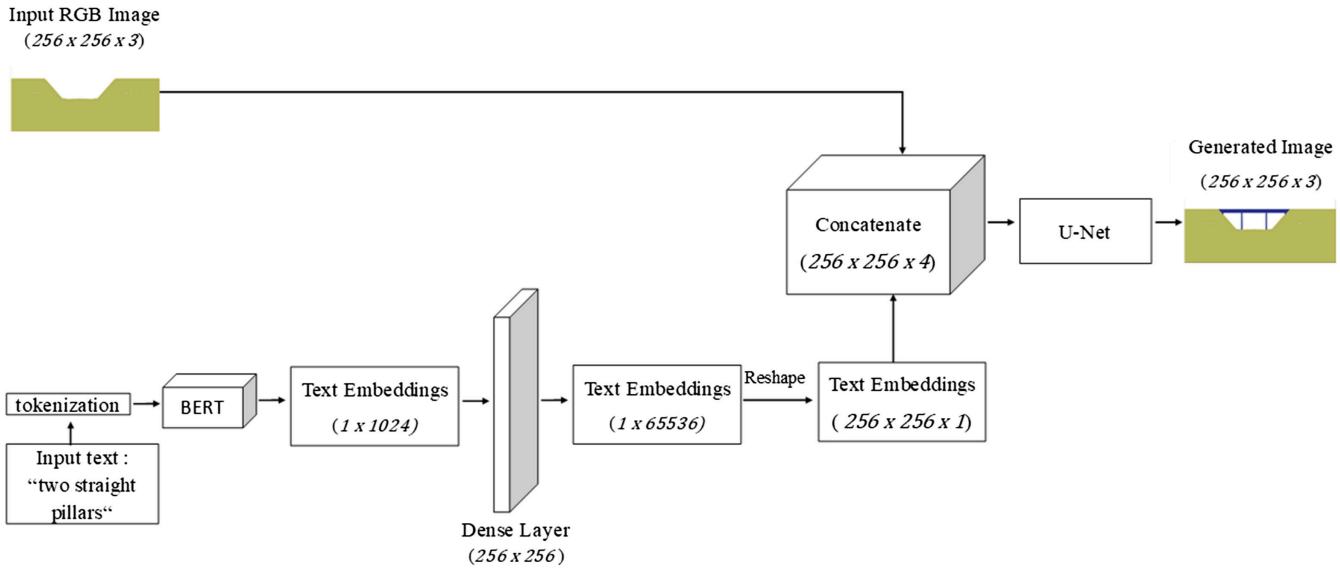


Fig. 5. Generator architecture (example with the design constraint “two straight pillars”).

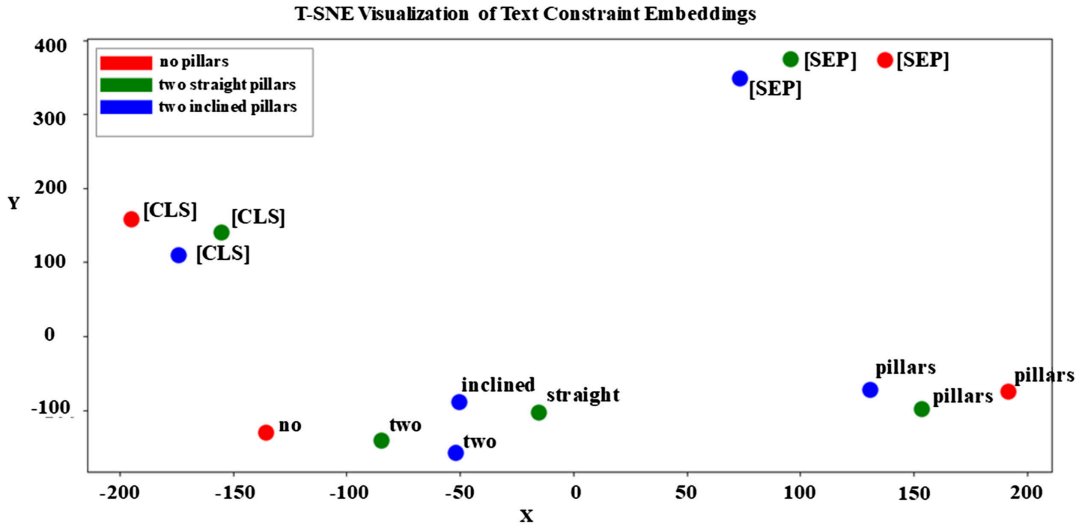


Fig. 6. *T*-SNE plot visualizing the embedding space of the output embeddings for tokens in the text constraints.

in a higher dimensional space. It produces embeddings for all the tokens in the input sequence. This results in similar words having similar embeddings that are positioned closer together in the embedding space. Fig. 6 visualizes the embedding space for the tokens contained in the design constraints using a *t*-distributed stochastic neighbor embedding (t-SNE) plot. T-SNE plots are used to project high-dimensional data into a two-dimensional (2D) space while preserving its local structure. Similar tokens form clusters within the embedding space (Fig. 6).

The embedding for the [CLS] token is extracted, because it provides a summarized representation of the text input. It then is passed through a dense layer to expand the text embedding to a higher-dimensional feature space so that it can be concatenated with the image (Fig. 5). This becomes the input for the U-Net architecture, which consists of an encoder–decoder framework with skip connections.

Given a generator G and a discriminator D, the optimization of the CGAN framework follows the min-max game of the generator–adversarial setup, in which the generator predicts images that are indistinguishable to the discriminator from the ground truth images. The objective of a CGAN can be defined mathematically as the mapping $G: \{x, z\} \rightarrow y$, of an input image x and a random noise vector z (induced into the framework during training using dropout in the down-sampling layers) to a target image y. This is expressed as

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \quad (1)$$

The governing principle of the min-max game is that G tries to minimize the objective against D, which in turn attempts to maximize it. The generator is trained using a combination of the mean-absolute error $\mathcal{L}_{L1}(G)$ and $\mathcal{L}_{\text{cGAN}}(G, D)$ as defined in Eq. (1), weighted by a factor λ. This is summarized as

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1] \quad (2)$$

$$G_{\text{loss}} = \arg\min_G \max_D \mathcal{L}_{\text{cGAN}}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (3)$$

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

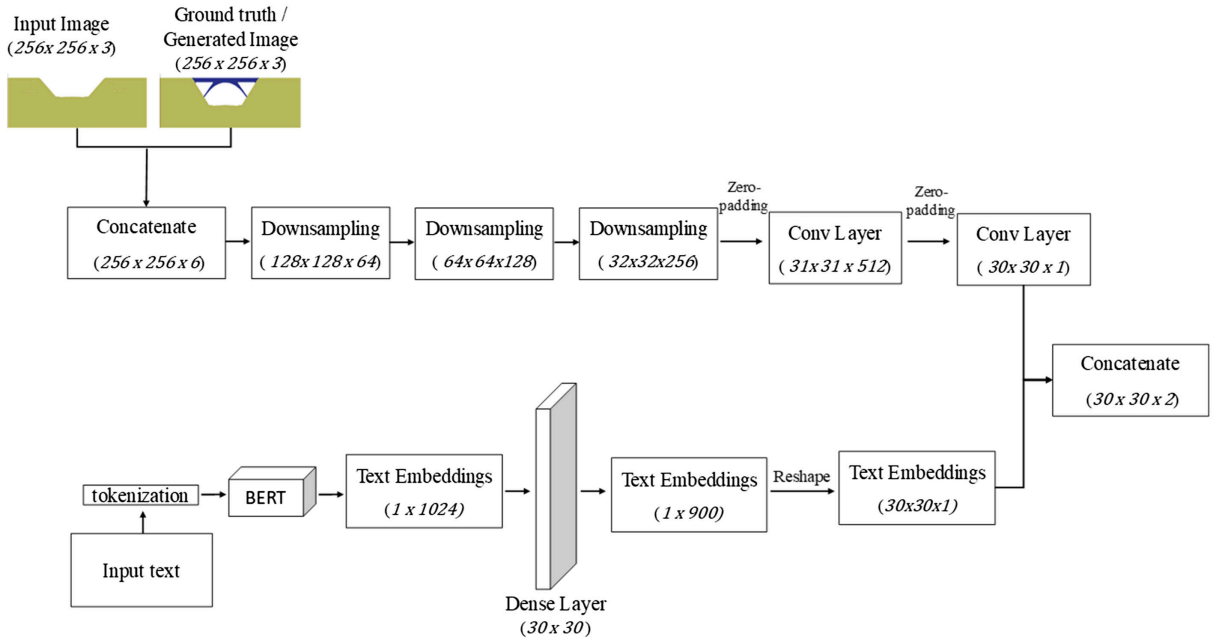


Fig. 7. Discriminator architecture with a patch size of 30 × 30.

Discriminator

An approach similar to that described for the generator was used to modify the architecture of the discriminator used for the pix2pix model (Fig. 7). The discriminator architecture is based on the PatchGAN (Li and Wand 2016) that classifies patches of the ground truth and generated images as being real and fake, respectively, for a predefined patch_size. Unlike in conventional GANs, the discriminators of CGANs are fed the input images along with the generated images. Therefore, the input topography image is passed in addition to the generated image to condition the discriminator. The discriminator provides a matrix (or a patch) of probabilities corresponding to the real and fake classes instead of a scalar.

The design constraint is passed to the discriminator model (Fig. 7) along with the input image and the generated image. Similar to the generator, the text input is tokenized and passed to the BERT model and the embedding for the [CLS] token is extracted. This is given to a dense layer, providing a representation which can be concatenated with the image features obtained as output of the final convolutional layers. The combined features then are subject to the classification between real and fake images. The discriminator is trained to distinguish between patches of real images (mapped to 1) and generated images (mapped to 0) using the standard binary cross-entropy (BCE) loss.

Experimental Setup

Data Set

In order to train a model capable of generating realistic conceptual bridge topologies, the quality and reliability of training data are critical. In this study, we collected bridge topology designs from existing structures in specific valleys, as well as recommended designs from sources such as the *Bridge Handbook* (Mehlhorn and Curbach 2014), *Bridges: Aesthetics and Design* (Leonhardt 1982), and other relevant projects and publications. These data sets served as the foundation for training the model.

Due to the existence of a large variety of bridge types and topographic conditions, the proposed approach focuses specifically on the 2D longitudinal topology of small to medium-sized beam and frame bridges within corresponding topographic settings. To extract the required target information—namely, the geometric data of the bridge topology and topography—the collected literature data first were preprocessed (Fig. 8). In the first step, the collected data were cleaned, i.e., the text, dimensions, and other unnecessary information were deleted. In the second step, the original topography and bridge topology was mapped into a red–green–blue (RGB) image with its actual scale to facilitate a more efficient learning process. Brown-green pixels represent the topography, and blue pixels represent the bridge topology. Finally, each image was reshaped to have a dimension of 256 × 256 × 3 (input size for the generator and discriminator architecture), and the pixel values were normalized to be within the range $[-1, 1]$ for better convergence during training. After the generator predicts the output images, they are reshaped back to their original dimensions.

The resultant data set includes pairs of RGB images of topography and possible bridge designs over the valleys. The topographies included in the data set were simplified for the experiments, but retained sufficient details to represent the general valley situations for model learning. These image pairs were used to train the conventional pix2pix model without the use of design constraints.

To enable the generation of multiple design variants, two other bridge designs in addition to each collected ground truth data set were developed by civil engineering lecturers at the Institute of Concrete Structures at Technische Universität Dresden. These alternative designs provided a diverse basis for analysis and comparison, aligning with the principle in the conceptual design phase. Consequently, each topographic image was associated with three different bridge design images, corresponding to the three distinct design constraints (Fig. 9). Structuring the data set in this manner enables the generative model to learn various design patterns under different constraints. This approach allows the model to focus on the specific design variations introduced by each constraint, thereby improving its ability to generate consistent and constraint-aware designs. The three constraints in this study correspond to the three

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

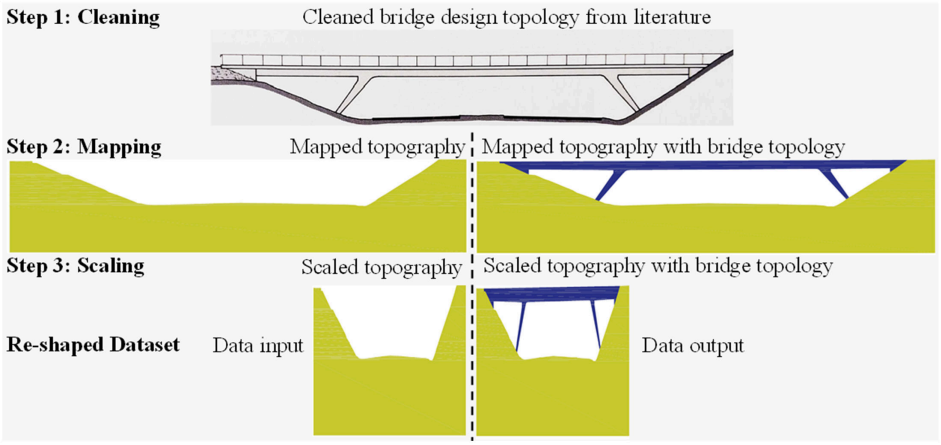


Fig. 8. Data preprocessing and preparation.



Fig. 9. Example showing the three bridge design possibilities for each design constraint.

most prominent bridge designs included in the data set. Therefore more-complex design constraints can be integrated into the workflow, given more and diverse training samples.

Notably, the data set was not intended to be fully representative in terms of geography or architectural diversity, but deliberately was compiled from well-documented handbook-style sources to ensure structural consistency and reliable design information. The omission of environmental performance parameters, which also are of decisive importance for topology design (Maksoud et al. 2023), was intentional, because the present study focused on isolating the influence of raw topographical geometry on early-stage bridge topology.

Training and Hyperparameter Settings

The data used for training the pix2pix model included 94 pairs of input and target images for training and validation. The training set was augmented using operations such as translation and horizontal flipping to produce a collection of 2,720 image pairs. For the model that incorporates the text constraints, a total of 195 image pairs were used for training and validating the model. These were augmented to create a set of 6,240 image pairs. Both models were trained using the Adam optimizer (Kingma and Ba 2014) with a momentum of 0.5 and a learning rate of 10^{-4} for 50 epochs on an Nvidia RTX A5000 GPU and tested and evaluated on a set of 78 unseen images. The weight parameter λ was set to 2×10^3.

Results and Evaluation

Qualitative Evaluation

Conventional Pix2Pix Model

The designs generated by the conventional pix2pix model on seven validation images and their ground truths are shown in Fig. 10. The model produced designs closely matching the ground truth, except for the last two inputs, for which there were significant differences. For example, whereas the ground truth image had a design with two inclined pillars for the sixth input image (from top), the model predicted a design with no pillars. This was due to the fact that the model saw a large number of input-target image pairs for similar topographies with bridge designs without any pillars. Such one-to-one predictions severely limit the design possibilities, especially when comparison of different conceptual designs is necessary in the initial planning phases of bridge designing.

Another important aspect to consider is the model's ability to generalize toward unfamiliar data. Fig. 11 displays the designs generated by the pix2pix model on seven previously unseen input images. The model performed well on some of the input images, but generated blurred or partially incomplete outputs for others. This suggests that the model struggles to generalize to topographies that differ from those encountered during training. This limitation is particularly evident in the last example, for which the model was presented with a challenging, uneven topography and failed to produce a reasonable result.

Advanced Text-Constrained Pix2Pix Method

For the advanced model incorporating text constraints, the generator produces bridge designs based on both the input image and the accompanying text constraints. This allows the generator to create three different bridge designs for the same input image, each corresponding to the three design constraints, one of which matches the original design constraint. The ground truth design and the three designs generated by the model, one for each design constraint—no pillars, two straight pillars, and two inclined pillars—are illustrated in Fig. 12 for each input image of a valley topography.

The generator was able to interpret the design constraints passed on to it. The generator also was able to learn design features of different bridge components, including bridge decks, abutments, and pillars. This highlights the potential of the model to learn features of different bridge topologies from the training data. However, compared with the designs generated on the validation data, the

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

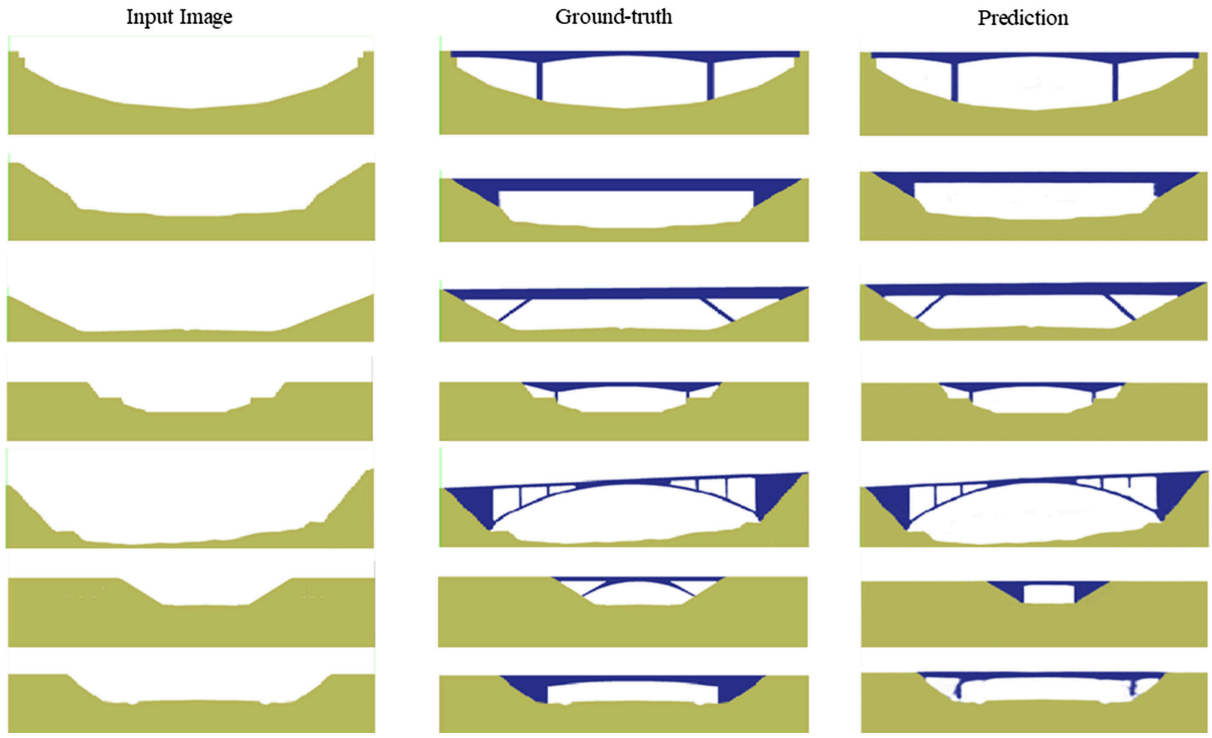


Fig. 10. Examples of seven different topography images from the validation data, the bridge designs predicted by the generator, and the corresponding ground truth images.

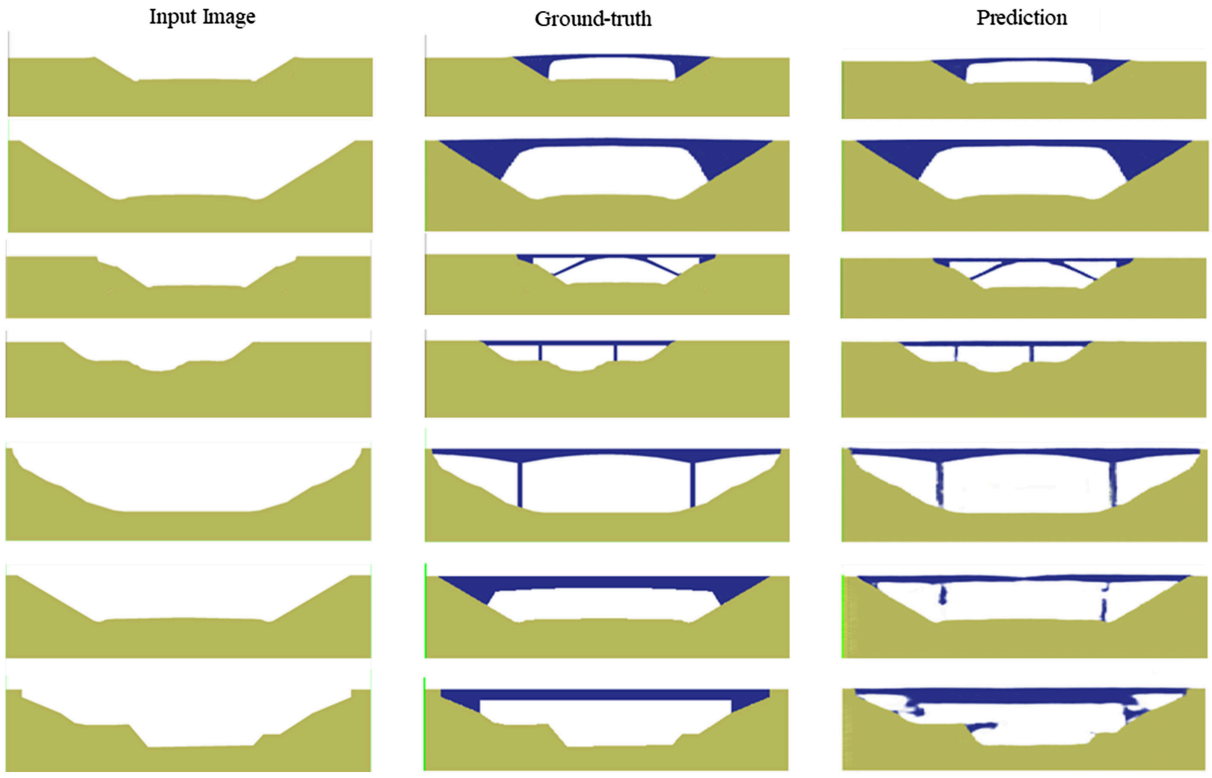


Fig. 11. Examples of seven different topography images from the test data, the bridge designs predicted by the generator, and the corresponding ground truth images.

model also struggled to consistently produce reasonable results when presented with unseen and challenging topographies from the test data set. Fig. 13 shows seven generated designs for unseen valley images, each paired with the three design constraints. In these cases, the predictions often included incomplete or blurred bridge designs. This is because the training data consisted of topographies which were relatively more uniform and even than some of the more complex images used during testing. Although signs of overfitting

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

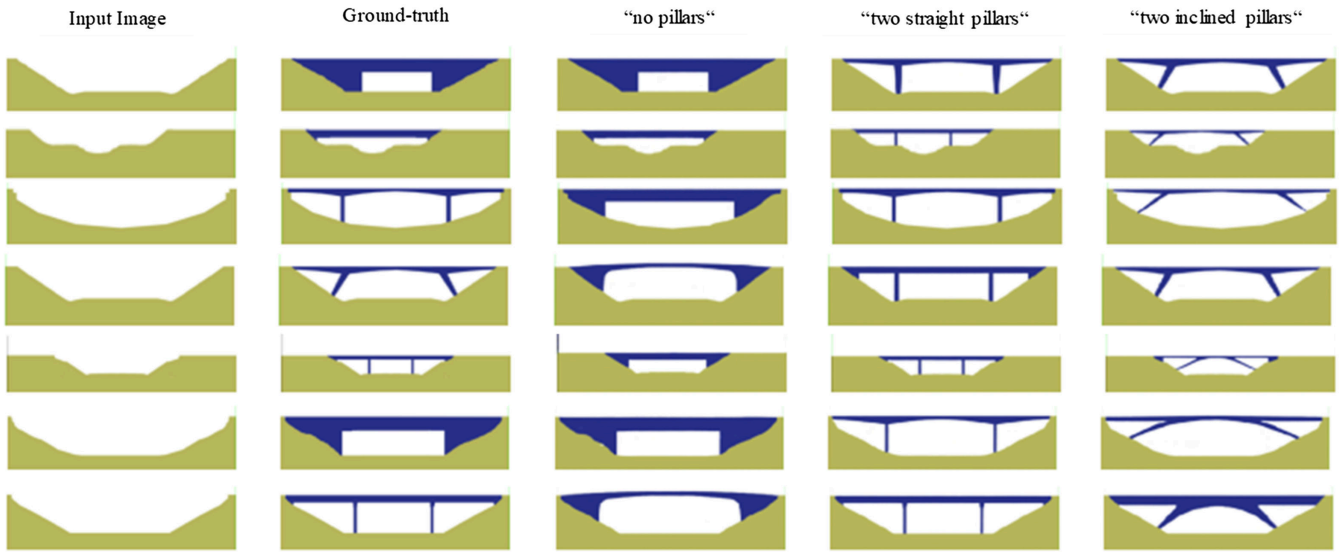


Fig. 12. Examples of seven different topography images, their ground truths, and the bridge designs generated by the generator for the constraints.

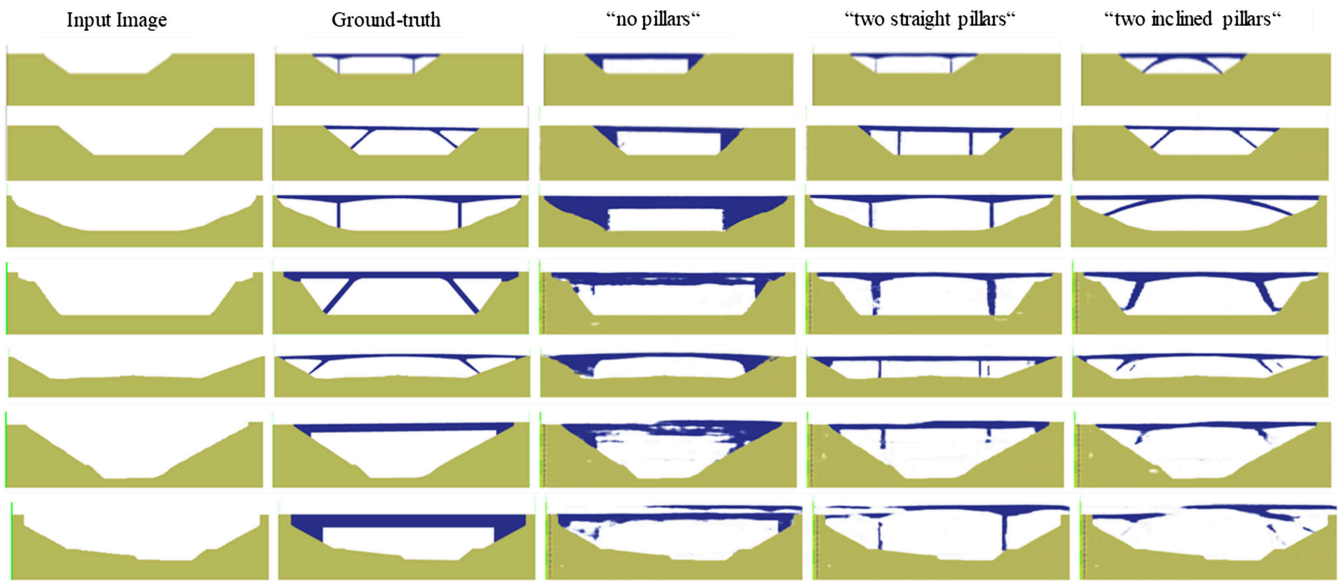


Fig. 13. Examples of seven different topography images from the test data, their ground truths, and the bridge designs generated by the generator for the three constraints.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

are evident, even the partially incomplete designs still meaningfully interpreted the text inputs, generating outputs that adhered to the specified design constraints (Fig. 13).

Quantitative Evaluation

The performance of the generator was tested quantitively using the structural similarity index measure (SSIM) (Wang et al. 2004). The SSIM measures how structurally similar the predictions of the generator are to the ground truths. It compares local patterns of pixel intensities based on aspects such as similarity in brightness, contrast, and structure (edges and textures). For two images a and b, the SSIM is defined mathematically as

$$\text{SSIM}(a,b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)} \quad (4)$$

where μ_a and μ_b = mean intensities; σ_a and σ_b = standard deviations; and σ_{ab} = covariance between a and b. The SSIM ranges from 0 to 1, where 0 indicates predictions completely different from the ground truths structurally, and 1 indicates predictions with perfect similarity to the ground truths. The generator can be evaluated by averaging the SSIM values measured for each prediction.

Another metric used to evaluate the model was the Fréchet inception distance (FID) (Heusel et al. 2017), due to its ability to evaluate the quality and diversity of the predictions, as implemented by Mseitzer (2020). FID compares the feature distributions of two sets of images to obtain a measure of similarity. The features are extracted from an Inception-V3 (Szegedy et al. 2015) model trained on the ImageNet data set. Although the final pooling layer of the model typically is used for feature extraction when a sufficient number of images is available (recommended minimum = 2,048), the features prior to the auxiliary classifier have been used for

Table 1. SSIM values calculated for conventional pix2pix model for validation and test data

Model	Validation data	Test data
Conventional pix2pix model	0.92	0.84

Table 2. FID scores calculated for conventional pix2pix model for validation and test data

Model	Validation data	Test data
Conventional pix2pix model	0.17	0.73

evaluation due to the limited data set size. These extracted features are then used calculate the mean vectors and covariance matrices of the distributions. The features are assumed to follow multivariate Gaussian distributions, and the difference of the two distributions can be calculated using the Fréchet distance (Azéma et al. 1999). For mapping from a source domain X (ground truths) to a target domain Y (generator predictions), the Fréchet distance can be calculated as

$$\text{FID} = \|\mu_x - \mu_y\|^2 + \text{Tr}\left(\Sigma_x + \Sigma_y - 2(\Sigma_x\Sigma_y)^{1/2}\right) \quad (5)$$

where μ_x and μ_y = means of the two distributions; Σ_x and Σ_y = covariances of the two distributions; and Tr is the trace of a matrix. For a generator that predicts outputs containing features similar to those of the ground truth images, the FID scores will be smaller (closer to 0). Although FID scores can be a reliable parameter for evaluating the realism and diversity in the predictions of a generative model, perceptual validation of the results by experts remains essential to validate the design specifications, structural feasibility, and adherence to engineering constraints.

The SSIM and the FID scores for both applied models are discussed in the following sections.

Conventional Pix2Pix Model

The SSIM and FID score were used to evaluate the trained models quantitatively. The SSIM and the FID scores calculated for the predictions of the pix2pix model on the validation and test data are given in Tables 1 and 2, respectively.

The relatively lower values for the SSIM and higher values for the FID score for the test data show the limitation in its generalization capabilities on unseen data. This inference is similar to that observed during the qualitative evaluation of the model (Figs. 12 and 13).

Advanced Text-Constrained Pix2Pix Method

Because the generator of the model extended to handle text constraints was designed in a way to predict three different classes of bridges and this diversity of predictions in each class has to be evaluated, the SSIM and the FID scores were calculated for each of the three constraints separately. The SSIM and the FID scores of the generator on the validation and test data are given in Tables 3 and 4, respectively. The generator performed well in predicting designs for all three design constraints for the validation data. Similar to the conventional pix2pix, the relatively lower SSIM values and higher FID scores for the test data indicate that the generator struggles to generalize in addition to the limited number of designs encountered during the training phase. Although it often has been observed that generative models tend to memorize patterns and overfit to the training set, especially when trained on a smaller data set, it still shows promise in its potential to learn design characteristics and interpret textual information.

Table 3. SSIM values calculated for text-constrained pix2pix method for three design constraints

Constraint	Validation data	Test data
No pillars	0.9577	0.8354
Two straight pillars	0.9540	0.8583
Two inclined pillars	0.9517	0.8879

Table 4. FID scores calculated for text-constrained pix2pix method for three design constraints

Constraint	Validation data	Test data
No pillars	0.16	0.72
Two straight pillars	0.15	0.78
Two inclined pillars	0.11	0.62

Ablation Study

Designing a generator–discriminator configuration that converges to an optimum by taking in a combination of image and text information includes several design considerations. The generator has to learn the different design intricacies and patterns from the training data, while at the same time adhering to the design constraints. The discriminator should be able to distinguish between real and generated designs, given not only the topography but also the design specifications. Our experiments included three additional approaches to develop a functional methodology: (1) the structural similarity index measure, (2) a dual discriminator, and (3) cross-attention. These three approaches were compared with the proposed methodology using the performances of their generators on the validation data using the SSIM and the FID score. The SSIM and FID scores calculated for the three different approaches are given in Tables 5 and 6, respectively.

Structural Similarity Index Measure

Using the SSIM as a loss function in the optimization of the generator could be considered a way to enable the generator to predict designs with better structural correlation with the training samples. Because the SSIM ranges from 0 to 1, integrating it into the loss function of the generator requires adding $(1 - \text{SSIM}(G(x, z), y))$ to the objective [Eq. (3)]

$$\mathcal{L}_{\text{SSIM}}(G) = 1 - \text{SSIM}(G(x, z), y) \quad (6)$$

$$G_{\text{loss}} = \arg\min_G \max_D \mathcal{L}_{\text{cGAN}}(G, D) + \lambda \mathcal{L}_{L1}(G)\mathcal{L}_{\text{SSIM}}(G) \quad (7)$$

The SSIM values and the FID scores obtained for the validation data for such an optimization strategy indicate that it underperforms compared with the original objective function (Tables 5 and 6). This could be attributed to the fact that although the SSIM considers the structural correlation (σ_{ab}) of the prediction and target during training, it also penalizes intensity (μ_a and μ_b) and contrast (σ_a and σ_b) variations. This could force the generator to focus on brightness and contrast shifts, even when it predicts structurally similar outputs.

Dual Discriminator

Another approach that was included in the experiments involves a dual-discriminator setup in which the generator was jointly optimized against not one but two discriminators. The first discriminator acted as a binary classifier, trained to classify real and fake images. The second discriminator was designed as shown in the section "Discriminator"—both text and image features were

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Table 5. SSIM values calculated for three different approaches with various constraints

Approach	Proposed methodology	With cross-attention	Dual discriminator	With SSIM-loss
No pillars	0.9577	0.9478	0.9520	0.9513
Two straight pillars	0.9540	0.9480	0.9492	0.9506
Two inclined pillars	0.9517	0.9444	0.9479	0.9460

Table 6. FID scores calculated for three different approaches with various constraints

Approach	Proposed methodology	With cross-attention	Dual discriminator	With SSIM-loss
No pillars	0.72	0.66	0.76	0.79
Two straight pillars	0.78	0.71	0.86	0.90
Two inclined pillars	0.62	0.59	0.71	0.75

concatenated for calculating the loss. The generator for such a setup was trained to minimize an objective against two adversaries D_1 and D_2, which try to maximize it. The rationale behind such a training strategy was to inspect the performance of the generator when trained against two different discriminators that consider the image and text inputs separately. Such a configuration also naturally leads to more parameters being optimized during training, leading to longer convergence times. Interestingly the generator performing better using a single discriminator than with the dual-discriminator setup (Tables 5 and 6).

Cross-Attention Mechanism

Methods that interpret raw text to guide image-based AI models have been researched extensively over the last few years. Among these, the use of cross-attention mechanism is a powerful tool to learn from image and text information (Avrahami et al. 2022; Brooks et al. 2022; Hertz et al. 2022). Such a cross-attention mechanism was implemented in the generator to combine features from both the text and image inputs. The architecture of the generator is shown in Fig. 14. The input image is fed through three layers of convolution (of filter sizes 32, 64, and 128) for the extraction of high-level features. These features together with the text embeddings of each token are passed to the cross-attention module.

In general, attention mechanisms enable AI models to weigh the importance given to different elements of an input sequence (or image), thereby dynamically adjusting their influence on the predictions. They are used in transformers to include the context of input sequences by enriching the information content of text embeddings. The embeddings are projected into key (K), query (Q), and value (V) components of the sequence using three trainable weight matrices W^K, W^Q, and W^V, respectively, using matrix multiplication; K represents the current token trying to retrieve relevant contextual information, and Q and V represent the information the current token has to offer and the actual information contained in the token, respectively. Attention is calculated as a weighted sum of values, and the weights are determined by the similarity between queries and keys, as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{8}$$

The attention scores are normalized to a probability distribution using a softmax function and scaled using the dimension of K (d_k) to prevent large dot products from destabilizing the gradients during training. Each token in the sequence queries every other token to establish the contextual relevance. This allows each token to weigh its focus on the neighboring tokens.

Conversely, cross-attention can be used to combine embeddings from different modalities—for example, image and texts. By setting the image features as queries and the text embeddings as keys and values, the attention mechanism learns to focus spatial regions on the input images relevant for text embeddings (for example, embeddings corresponding to tokens such as straight and inclined). A similar approach was used to modify the discriminator to combine image features and text embeddings.

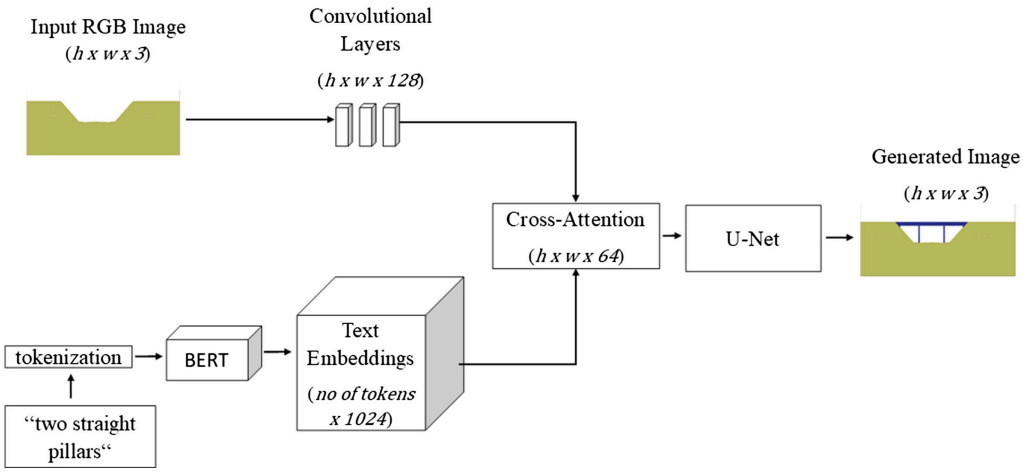


Fig. 14. Generator architecture using cross-attention mechanism.

J. Comput. Civ. Eng., 2026, 40(6): 04026089

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

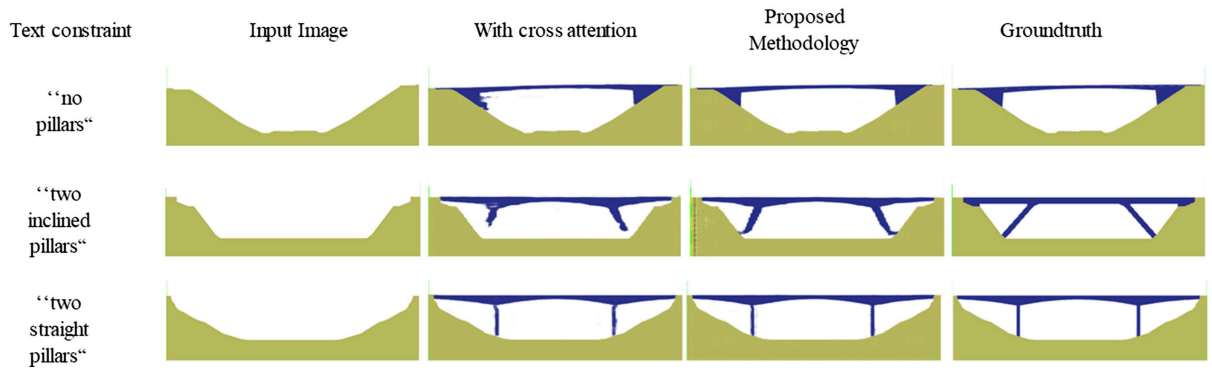


Fig. 15. Comparison of the proposed methodology with the model using cross-attention on three test images.

The SSIM values and the FID scores obtained for this model on the test data are given in Tables 5 and 6 respectively. Although the SSIM values show relatively poorer performance than that of the proposed model, comparing the FID scores with those obtained for the proposed model indicate better performance. However, a qualitative comparison of the results of the two models reveals that the predictions of the proposed model are actually more complete and sharper than those obtained by the model using cross-attention (Fig. 15). The predictions made by the model using cross-attention contained blurred and incomplete areas, despite better FID scores. This shows that the concatenation of text embeddings with the input images (as implemented in the proposed model) enabled the model to predict sharper images. The more-favorable FID scores obtained for the model with cross-attention prove that using qualitative evaluation of the predictions is more reliable than using FID scores. Although it can be good indication of a model's ability to produce diverse and realistic outputs, when presented with insufficient data in the source and target domains, the distributions obtained become noisy and unstable. This leads to poor statistical estimations of the data and unreliable FID scores.

Discussions

The results of the text-guided generative approach show that although it is able to predict practically feasible bridge designs for topographies similar to those it saw during training, does not perform as well when given valley situations that are very different from those included in the training set, which is reasonable. Despite using countermeasures such as regularization in the architecture of the generator, it still shows signs of overfitting. This can be attributed to the limited number of bridge designs presented in the data set. Even when data augmentation was carried out to increase the amount of training samples, there were multiple examples of similar designs from a collection of limited design varieties. Although the developers of the pix2pix model proved that satisfactory results can be achieved even for data sets with limited number of samples, a significant deviation from the training set in terms of the topography, as in Fig. 11, can lead to dissatisfactory results. In addition, the results became worse with an increase of unevenness in the valley situations, which was largely absent in the training samples. Taking into account the fact that GANs generally are difficult to train due to noisy gradients and challenges such as mode collapse, it is hypothesized that increasing the variety of topographies and corresponding bridge designs in the training data set can enhance the generalization capabilities of the model.

The use of the BERT model to constrain the predictions of the generator due to the deterministic nature of the pix2pix model was proven to be effective. Even when the predictions were incomplete, they were different from one another for the same input images and clearly showed signs of guiding the generator in the right direction, in terms of the design specifications such as "pillars," "straight," and "inclined."

The model's tendency to overfit remains a limitation for generalizing to unseen and challenging topographies. Nevertheless, the generator performs well in predicting designs within the scope of its training data, adhering to the design prompts. However, the data set remains limited due to resource and time constraints and the fact that our methodology relies on data from real-world bridge constructions. In this study, we intentionally focused on medium-span girder and frame bridges, because they are among the most commonly built bridge types worldwide and share many topological similarities. Despite these considerations, obtaining a sufficiently large data set remains challenging, because other bridge types such as arch bridges and cable-stayed bridges are less common and differ considerably from the selected types, thereby increasing the complexity of the training process. The present work therefore aimed to introduce a novel proof-of-concept generative methodology, rather than a fully comprehensive generator applicable to all bridge types.

Because this paper focuses on the generation of bridge designs guided by text constraints on given images for the initial design phase, the feasibility of the designs predicted by the generator was examined only using perceptual validation, SSIM, and the FID score. However, the quantitative evaluation of generative models has been a topic of discussion and debate over the years, and researchers have contributed several evaluation metrics that mathematically analyze the predictions of a generator (Borji 2022). Mathematically evaluating the performance of a generative model that predicts bridge designs becomes challenging because it is a highly domain specific application.

Moreover, although FID is a useful metric for quantitatively evaluating the realism and diversity of the generator's predictions, it requires a considerable number of ground truth and test samples to capture meaningful distributions for all the different bridge design possibilities. With these factors in mind, developing a custom evaluation metric to topologically examine the feasibility of the predicted bridge designs and to evaluate the completeness and diversity of the predictions is necessary.

However, a more precise structural engineering assessment currently is not feasible, because the accuracy of this approach remains limited due to the scaling applied during preprocessing of the raw data. Because the current data lacks a uniform scale, rescaling the predictions to match the dimensions of real-world structures is not possible. As a result, a one-pixel difference in the generated designs can correspond to a difference of several

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

meters in the real-world structure. Notably, our approach is to find an optimal topology or form for a given topography, which aligns with the aim of conceptual design. The detailed dimensioning of the structure and its components can be addressed only at a later stage, commonly through comprehensive technical design in the present practice, which is beyond the scope of this work.

In general, the proposed method targets early-stage topology exploration rather than detailed structural design or optimization, and the current evaluation is primarily visual and qualitative, reflecting the early-stage, proof-of-concept nature of the proposed workflow. The results indicate that the model performs well for topographies similar to those in the training data, but generalization to significantly different conditions remains limited. Therefore, future work will focus on enriching the data set to cover a wider range of topographic situations and on exploring new methodologies, such as incorporating physical information, to guide the generation of structurally appropriate designs and improve their precision. These extensions are expected to enhance interpretability, robustness, and potential applicability.

Conclusion and Outlook

This research designed a proof-of-concept image-based generative model to create reasonable bridge conceptual designs over highways, given images of the topography. This was achieved by designing a text-guided CGAN based on the pix2pix architecture and the BERT transformer model. The results of our experiments show that the generator is able to successfully produce outputs tailored to the design constraints in the form of texts. Notably, the proposed workflow is not intended to automate design decisions, but to serve as a data-driven decision-support tool that complements, rather than replaces, engineering judgment. In addition, for challenging topographies, the model still has difficulty generalizing its knowledge and extending its capacity to generate feasible bridge designs. Although this is an obvious challenge, the potential of the general methodology is encouraging, and can be experimented upon further using more-diverse and -challenging data.

Having established the potential of the methodology introduced in this paper, we see opportunities for further improving the functionalities of the model. Due to a lack of sufficient training images, it was not possible to extend the number of design constraints beyond the basic ones which we used. Experimenting on the generalization capabilities of the model with more-complex constraints with more specific characteristics (describing the shapes of the deck, piers, and so forth) is one such possibility. This also gives the opportunity to test the ability of the generator in handling semantics and adapt its predictions to constraints with different word combinations or even unseen design constraints, based on the constraints given to the model during training. In addition, it also is worth researching the effectiveness of additional constraints such as the length of the deck or the height of the pillars, thereby taking into account geometric aspects of bridges. Integrating more specific parameters could help engineers and bridge designers to better constrain the generator's prediction matching their geometric requirements. Modifying the existing topography images to more closely match real-life valley situations is another possible extension. The existing valley situations were simplified for training purposes, but real-life situations are more diverse and include more vegetation irregularities.

Due to the recent advancements achieved in the domain of latent diffusion models, we also see the potential of experimenting with such models on a more diverse data set. They have proven to be more attractive than the conventional generative adversarial models, given adequate amounts of data, because they do not pose the challenges of mode collapse or noisy gradients which lead to unstable convergence. Considering the potential of the proposed methodology, generating bridge designs using diffusion models guided by natural language prompts could bring the research closer to practical relevance.

This methodology offers promising future prospects for incorporation into a higher automated structural design process for bridges. However, pre- and postprocessing methods will be inevitable. For example, in the preprocessing stage, the topography condition, used as input data in the methodology proposed in this work, can be generated automatically using GIS technology. As a postprocessing step, the bridge designs generated in image format can be processed further and transformed into geometric information, specifically focusing on the dimensional characteristics of the generated structures, provided that the desired precision can be achieved. Subsequently, additional evaluations can be conducted using engineering-specific geometric and structural assessment metrics, such as analyses of structural performance, the geometric feasibility of bridge components and compliance with established engineering standards. Furthermore, by incorporating additional methods such as parametric design, the automation of more-detailed bridge designs for the subsequent preliminary, detailed, and construction design phases potentially can be achieved. Building on this, the highly complex bridge design process could be largely automated.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This work was funded by the German Research Foundation (DFG) under the Cluster of Excellence CARE: Climate Neutral and Resource Efficient Construction (EXC 3115), Project No. 5337677731. The authors from TU Dresden acknowledge the German Federal Ministry for Digital and Transport for their support within the mFUND funding program and the research project Pic2Bridge (Project No. 19F1202).

Author Contributions

Aswin Lal: Conceptualization; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review and editing. Chongjie Kang: Conceptualization; Data curation; Investigation; Methodology; Project administration; Validation; Visualization; Writing – original draft; Writing – review and editing. Baris Özcan: Methodology; Validation; Writing – review and editing. Jörg Blankenbach: Conceptualization; Resources; Supervision; Writing – review and editing. Steffen Marx: Conceptualization; Funding acquisition; Resources; Writing – review and editing.

References

Avrahami, O., D. Lischinski, and O. Fried. 2022. "Blended diffusion for text-driven editing of natural images." In *Proc., 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 18187–18197. New York: IEEE.

Azéma, J., M. Émery, M. Ledoux, and M. Yor, eds. 1999. *Séminaire de Probabilités XXXIII*. Berlin: Springer.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Balmer, V., S. V. Kuhn, R. Bischof, L. Salamanca, W. Kaufmann, F. Perez-Cruz, and M. A. Kraus. 2024. "Design space exploration and explanation via conditional variational autoencoders in meta-model-based conceptual design of pedestrian bridges." *Autom. Constr.* 163 (Jul): 105411. https://doi.org/10.1016/j.autcon.2024.105411.

Barker, N. 2023. "ZHA developing 'most' projects using AI-generated images says Patrik Schumacher Dezeen." Accessed April 26, 2023. https://www.dezeen.com/2023/04/26/zaha-hadid-architects-patrik-schumacher-ai-dalle-midjourney/.

Borji, A. 2022. "Pros and cons of GAN evaluation measures: New developments." *Comput. Vision Image Understanding* 215 (Jan): 103329. https://doi.org/10.1016/j.cviu.2021.103329.

Brooks, T., A. Holynski, and A. A. Efros. 2022. "InstructPix2Pix: Learning to follow image editing instructions." Preprint, submitted November 17, 2022. https://arxiv.org/abs/2211.09800.

Chen, J., D. Wang, Z. Shao, X. Zhang, M. Ruan, H. Li, and J. Li. 2023. "Using artificial intelligence to generate master-quality architectural designs from text descriptions." *Buildings* 13 (9): 2285. https://doi.org/10.3390/buildings13092285.

Dai, Z., Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov. 2019. "Transformer-XL: Attentive language models beyond a fixed-length context." Preprint, submitted January 9, 2019. https://arxiv.org/abs/1901.02860v3.

De Miguel Rodríguez, J., M. E. Villafañe, L. Piškorec, and F. Sancho Caparrini. 2020. "Generation of geometric interpolations of building types with deep variational autoencoders." *Des. Sci.* 6 (Jan): e34. https://doi.org/10.1017/dsj.2020.31.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. "BERT: Pre-training of deep bidirectional transformers for language understanding." Preprint, submitted October 11, 2018. https://arxiv.org/abs/1810.04805v2.

Fu, B., Y. Gao, and W. Wang. 2023. "Dual generative adversarial networks for automated component layout design of steel frame-brace structures." *Autom. Constr.* 146 (Feb): 104661. https://doi.org/10.1016/j.autcon.2022.104661.

Ghannad, P., and Y.-C. Lee. 2022. "Automated modular housing design using a module configuration algorithm and a coupled generative adversarial network (CoGAN)." *Autom. Constr.* 139 (Jul): 104234. https://doi.org/10.1016/j.autcon.2022.104234.

Herron, E., J. Rade, A. Jignasu, B. Ganapathysubramanian, A. Balu, S. Sarkar, and A. Krishnamurthy. 2023. "Latent diffusion models for structural component design." Preprint, submitted September 20, 2023. https://arxiv.org/abs/2309.11601v2.

Hertz, A., R. Mokady, J. Tenenbaum, K. Aberman, Y. Pritch, and D. Cohen-Or. 2022. "Prompt-to-prompt image editing with cross attention control." Preprint, submitted August 2, 2022. https://arxiv.org/abs/2208.01626v1.

Heusel, M., H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. 2017. "GANs trained by a two time-scale update rule converge to a local Nash equilibrium." Preprint, submitted June 26, 2017. https://arxiv.org/abs/1706.08500v6.

Horvath, A.-S., and P. Pouliou. 2024. "AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators." *Front. Archit. Res.* 13 (3): 593–612. https://doi.org/10.1016/j.foar.2024.02.006.

Huang, W., and H. Zheng. 2018. "Architectural drawings recognition and generation through machine learning." In *Proc., ACADIA 2018: Re/Calibration: On Imprecision and Infidelity*, edited by P. Anzalone, M. Del Signore, and A. J. Wit, 156–165. Fargo, ND: Association for Computer Aided Design in Architecture.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros. 2017. "Image-to-image translation with conditional adversarial networks." In *Proc., 2017 IEEE Conf. on Computer Vision and Pattern Recognition*, 5967–5976. New York: IEEE.

Jiang, F., J. Ma, C. J. Webster, X. Li, and V. J. Gan. 2023. "Building layout generation using site-embedded GAN model." *Autom. Constr.* 151 (Jul): 104888. https://doi.org/10.1016/j.autcon.2023.104888.

Kallioras, N. A., and N. D. Lagaros. 2020. "DzAIN: Deep learning based generative design." *Procedia Manuf.* 44 (Jan): 591–598. https://doi.org/10.1016/j.promfg.2020.02.251.

Kang, C., S. Schneider, M. Wenner, and S. Marx. 2018. "Development of design and construction of high-speed railway bridges in Germany." *Eng. Struct.* 163 (May): 184–196. https://doi.org/10.1016/j.engstruct.2018.02.059.

Kingma, D. P., and J. Ba. 2014. "Adam: A method for stochastic optimization." Preprint, submitted December 22, 2014. https://arxiv.org/abs/1412.6980v9.

Kraus, M. A. 2022. "Erklärbare domänenspezifische Künstliche Intelligenz im Massiv- und Brückenbau." *Beton- Stahlbetonbau* 117 (10): 795–804. https://doi.org/10.1002/best.202200079.

Kraus, M. A., S. V. Kuhn, A. Hodel, R. Bischof, A. Maissen, L. Salamanca Mino, and F. Pérez-Cruz. 2024. "Parametrische Modellierung und generatives tiefes Lernen für den Brückenentwurf." *Bautechnik* 101 (3): 174–180. https://doi.org/10.1002/bate.202300122.

Leonhardt, F. 1982. *Brücken—Bridges: Ästhetik und Gestaltung—Aesthetics and design*. Munich, Germany: Deutsche Verlags-Anstalt.

Li, C., and M. Wand. 2016. "Precomputed real-time texture synthesis with Markovian generative adversarial networks." Preprint, submitted April 15, 2016. https://arxiv.org/abs/1604.04382v1.

Li, S., and X. Zhao. 2023. "High-resolution concrete damage image synthesis using conditional generative adversarial network." *Autom. Constr.* 147 (Mar): 104739. https://doi.org/10.1016/j.autcon.2022.104739.

Liao, W., Y. Huang, Z. Zheng, and X. Lu. 2022. "Intelligent generative structural design method for shear wall building based on "fused-text-image-to-image" generative adversarial networks." *Expert Syst. Appl.* 210 (Mar): 118530. https://doi.org/10.1016/j.eswa.2022.118530.

Liao, W., X. Lu, Y. Fei, Y. Gu, and Y. Huang. 2024. "Generative AI design for building structures." *Autom. Constr.* 157 (Jan): 105187. https://doi.org/10.1016/j.autcon.2023.105187.

Liu, M.-Y., and O. Tuzel. 2016. "Coupled generative adversarial networks." Preprint, submitted June 24, 2016. https://arxiv.org/abs/1606.07536v2.

Maeda, H., T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata. 2021. "Generative adversarial network for road damage detection." *Comput.-Aided Civ. Infrastruct. Eng.* 36 (1): 47–60. https://doi.org/10.1111/mice.12561.

Maksoud, A., H. B. Al-Beer, A. Ali Hussien, S. Dirar, E. Mushtaha, and M. W. Yahia. 2023. "Computational design for futuristic environmentally adaptive building forms and structures." *Archit. Eng.* 8 (1): 13–24. https://doi.org/10.23968/2500-0055-2023-8-1-13-24.

Mazé, F., and F. Ahmed. 2022. "Diffusion models beat GANs on topology optimization." Preprint, submitted August 20, 2022. https://arxiv.org/abs/2208.09591v2.

Mehlhorn, G., and M. Curbach. 2014. *Handbuch brücken*. Wiesbaden, Germany: Springer.

Mseitzer. 2020. "PyTorch-fid: FID score for PyTorch." Accessed April 2, 2025. https://github.com/mseitzer/pytorch-fid.

Nauata, N., K.-H. Chang, C.-Y. Cheng, G. Mori, and Y. Furukawa. 2020. "House-GAN: Relational generative adversarial networks for graph-constrained house layout generation." Preprint, submitted March 16, 2020. https://arxiv.org/abs/2003.06988v1.

Qian, H., E. Panagiotou, M. Peng, E. Ntoutsi, C. Kang, and S. Marx. 2024. "A novel dataset and feature selection for data-driven conceptual design of offshore jacket substructures." *Ocean Eng.* 303 (Jul): 117679. https://doi.org/10.1016/j.oceaneng.2024.117679.

Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever. 2018. "Improving language understanding with unsupervised learning." Accessed March 31, 2025. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

Rae, J. W., A. Potapenko, S. M. Jayakumar, and T. P. Lillicrap. 2019. "Compressive transformers for long-range sequence modelling." Preprint, submitted November 13, 2019. https://arxiv.org/abs/1911.05507v1.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.

Ramesh, A., M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever. 2021. "Zero-shot text-to-image generation." Preprint, submitted February 24, 2021. https://arxiv.org/abs/2102.12092v2.

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional networks for biomedical image segmentation." Preprint, submitted May 18, 2015. https://arxiv.org/abs/1505.04597v1.

Saka, A., R. Taiwo, N. Saka, B. A. Salami, S. Ajayi, K. Akande, and H. Kazemi. 2024. "GPT models in construction industry: Opportunities, limitations, and a use case validation." *Dev. Built Environ.* 17 (Mar): 100300. https://doi.org/10.1016/j.dibe.2023.100300.

Sardone, L., A. Fiore, A. Manuello, and R. Greco. 2022. "Performative structural design optimization: Generative algorithm for a preliminary study of a voided beam." *Appl. Sci.* 12 (17): 8663. https://doi.org/10.3390/app12178663.

Schneider, S., and S. Marx. 2018. "Design of railway bridges for dynamic loads due to high-speed traffic." *Eng. Struct.* 174 (Nov): 396–406. https://doi.org/10.1016/j.engstruct.2018.07.030.

Slepicka, S., L. Kolbeck, S. Esser, K. Forth, F. Noichl, and J. Schlenger. 2022. "Linking ontology metrics with BIM modeling stages." In *Proc., 33. Forum Bauinformatik*. Munich, Germany: Technische Universität München. https://doi.org/10.14459/2022md1686600.

Spiegelhalter, T. 2023. "Performance-based, AI-ML-assisted generative EA design with bio-inspired topological optimisations of a 50m, 3D-printed steel bridge." In *Proc., IJSCER*, 63–67. Rowland Heights, CA: Engineering and Technology Publishing. https://doi.org/10.18178/ijscer.12.3.63-67.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2015. "Rethinking the inception architecture for computer vision." Preprint, submitted December 2, 2015. https://arxiv.org/abs/1512.00567v3.

Tang, M.-C. 2017. "Conceptual design of bridges." *Struct. Infrastruct. Eng.* 13 (4): 418–427. https://doi.org/10.1080/15732479.2016.1164723.

Tang, M.-C. 2018. "Forms and aesthetics of bridges." *Engineering* 4 (2): 267–276. https://doi.org/10.1016/j.eng.2017.12.013.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. 2017. "Attention is all you need." Preprint, submitted June 12, 2017. https://arxiv.org/abs/1706.03762v7.

Wang, T.-C., M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. 2018. "High-resolution image synthesis and semantic manipulation with conditional GANs." In *Proc., 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 8798–8807. New York: IEEE.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image quality assessment: From error visibility to structural similarity." *IEEE Trans. Image Process.* 13 (4): 600–612. https://doi.org/10.1109/TIP.2003.819861.

Zhang, H., K. Gao, Y. Hou, M. Domaneschi, and M. Noori. 2025. "Portable IoT device for tire text code identification via integrated computer vision system." *Comput.-Aided Civ. Infrastruct. Eng.* 40 (15): 2170–2189. https://doi.org/10.1111/mice.13438.

Zhu, Y., R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler. 2015. "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books." Preprint, submitted June 22, 2015. https://arxiv.org/abs/1506.06724v1.

Zunino, L., J. R. Casas, and M. Domaneschi. 2025. "Bridge damage assessment under traffic and environmental variability: A case study on Yonghe cable-stayed bridge." *Eng. Struct.* 343 (Nov): 120965. https://doi.org/10.1016/j.engstruct.2025.120965.

This work is made available under the terms of the Creative Commons Attribution 4.0 International license.