

Generative Design Exploration by Integrating Deep Generative Models and Topology Optimization

Sangeun Oh^{1,†}, Yongsu Jung^{2,†}, Seongsin Kim¹, Ikjin Lee^{2,*}, Namwoo Kang^{1,*}

¹Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul, Korea

²Korea Advanced Institute of Science and Technology, 291, Daehak-ro, Yuseong-gu, Daejeon,
Korea

[†]S. Oh and Y. Jung contributed equally to this work.

*Co-corresponding author: ikjin.lee@kaist.ac.kr (I. Lee), nwkang@sm.ac.kr (N. Kang)

A previous version of this manuscript was presented at the ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (Quebec City, Canada, Aug. 26-29, 2018).

Abstract

Deep learning has recently been applied to various research areas of design optimization. This study presents the need and effectiveness of adopting deep learning for generative design (or design exploration) research area. This work proposes an artificial intelligent (AI)-based design automation framework that is capable of generating numerous design options which are not only aesthetic but also optimized for engineering performance. The proposed framework integrates topology optimization and deep generative models (e.g., generative adversarial networks (GANs)) in an iterative manner to explore new design options, thus generating a large number of designs starting from limited previous design data. In addition, anomaly detection can evaluate the novelty of generated designs, thus helping designers choose among design options. The 2D wheel design problem is applied as a case study for validation of the proposed framework. The framework manifests better aesthetics, diversity, and robustness of generated designs than previous generative design methods.

Keywords: Design Exploration, Generative Design, Topology Optimization, Deep Generative Models, Generative Adversarial Networks, Aesthetics, Engineering Performance

Nomenclature

\mathbb{E} : expectation
 D : differentiable function of discriminator
 G : differentiable function of generator
 z : noise variable
 \mathcal{L} : loss function of autoencoder
 c : compliance
 x_e : density variable
 \tilde{x}_e : filtered density variable
 $\bar{\tilde{x}}_e$: projected density variable
 p : penalization factor

1. Introduction

Artificial intelligence (AI) covers all technologies pursuing machines to imitate human behavior. Machine learning is a subset of AI, which attempts to learn meaningful patterns from raw data by using statistical methods. Deep learning seeks to enhance the learning ability with a hierarchical neural network structure that consists of several layers [1]. Recently, deep learning has been employed not only in computer science but also in various engineering domains. A physics-based approach in engineering domains can be replaced with a data-driven approach in an effective way. In mechanical engineering, it has been widely applied to autonomous driving, robot control, biomedical engineering, prognostics and health management, and design optimization.

Deep learning research related to design optimization can be classified as follows: (1) topology optimization [2-5], (2) shape parameterization [6,7], (3) computer-aided engineering (CAE) simulation and meta modeling [8-10], (4) material design [11-13], and (5) design preference estimation [14,15]. Section 2.1 introduces each research in detail.

This study commenced from the idea that deep learning is indispensable for generative design (i.e., design exploration), and the aforementioned deep learning research can certainly be integrated through the proposed framework in a holistic view.

1.1. Generative Design Using Topology Optimization

Generative design is one of the design exploration methods performed by typically varying design geometry parametrically and assessing the performance of output designs [16,17]. Recent research on generative design utilizes topology optimization as a design generator instead of design parameterization and develop the methods to generate numerous designs in parallel with cloud computing [18]. A designer provides diverse boundary conditions of topology optimization, which brings different optimized designs under different boundary conditions. Matejka et al. [19] state “generative design varies the parameters of the problem definition while parametric design varies parameters of the geometry directly.”

Generative design aspires to explore the design options that satisfy structural performance and choose suitable designs for various designers’ needs, whereas conventional topology optimization seeks to find an optimal design. The generative design concept has been rapidly developed and implemented in a commercial software [20] and applied in designing various structures such as automobile, architecture, and aircraft.

The overall process of generative design consists of four stages as follows [18]:

- Step 1: Set the design parameters and goals for topology optimization.
- Step 2: Generate designs with running topology optimization under different parameters.
- Step 3: Study options, iterate, and select the best design.
- Step 4: Manufacture the design by 3D printing.

In particular, 3D printing technology development enabled the production of complicated geometric designs, which further accelerates the practical use of generative design.

However, several drawbacks in current generative design were identified [21]. First, it does not use state-of-the-art AI technology (e.g., deep learning), even though topology optimization can be considered as AI in a broad way. Second, it is unable to create aesthetic designs. Topology optimization focuses solely on engineering performance. Therefore, the results seem to be counterintuitive in the aesthetic point of view. However, aesthetics is an essential factor for customers and should be balanced with engineering performance [22]. Third, the diversity of optimized designs is low, which can result in a new design in terms of intensity or density of pixels. However, these designs may be similar in terms of human perception.

1.2. Generative Models for Generative Design

Generative models, one of the promising deep learning areas, can enhance research on generative design. The generative model is an algorithm for constructing a generator that learns the probability distribution of training data and generates new data based on learned probability distribution. In particular, variational autoencoder (VAE) and generative adversarial network (GAN) are popular generative models used in design optimization, where high-dimensional design variables are encoded in low-dimensional design space [11,12]. In addition, these models are utilized in the design exploration and shape parameterization [6,7].

The use of generative model to produce engineering designs directly is limited [21]. However, this study claims that the limitations can be overcome by integrating with topology optimization. First, the generative model requires a number of training data, but accumulated training data for various designs in the industry are confidential and difficult to access. A number of designs obtained from topology optimization are expected to serve as training data. Second, the generative model cannot guarantee feasible engineering. In this case,

engineering performance can be evaluated through topology optimization. Third, mode collapse is one of the main problems in the generative model, producing only specific results and bringing large variance of output quality. However, low-quality designs can be improved through post-processing by employing topology optimization.

1.3. Research Purpose

This study proposes the new framework for generative design exploration by integrating deep generative models and topology optimization. The proposed framework can provide a number of meaningful design options accounting for engineering performance and aesthetics and allows evaluation and visualization of the new design options according to the design attributes (e.g., novelty, compliance, and cost).

The proposed framework consists of iterative design exploration and design evaluation parts. Iterative design exploration involves generating a large number of new designs iteratively by using a small number of previous designs. Design evaluation involves quantifying the novelty of generated designs in comparison with the previous designs and visualizing design options with other design attributes. The proposed framework is applied to 2D wheel design of an automobile for demonstration.

The rest of this study is structured as follows. Section 2 reviews previous design optimization studies using deep learning and introduces the main methodologies used in our studies (e.g., deep generative models and topology optimization). Section 3 proposes a generative design exploration framework. Section 4 analyzes and discusses case study results. Finally, Section 5 summarizes the conclusions and limitation and introduces future work.

2. Related Work

Section 2.1 presents brief design optimization research reviews by using deep learning. Sections 2.2 and 2.3 show the methodologies applied to the proposed framework.

2.1 Deep Learning in Design Optimization Research

The deep learning-based design optimization research is explained as follows. First, topology optimization can be interpreted as deep learning from the perspective of pixel-wise image labeling because it distributes the materials in design domain accounting for objective and constraint functions [2]. Intensive computational demand is a drawback of topology optimization due to iterative finite element analysis (FEA) of a structure. Thus, Yu et al. [3] propose the framework where a low-resolution structure is first generated by using a convolutional neural network (CNN)-based encoder and decoder constructed on 2D topology optimal designs, and then it is upsampled to a high-resolution structure through conditional GAN (cGAN) without any iterative FEA in topology optimization. Banga et al. [4] employ 3D CNN-based encoder and decoder, which allows the final optimized structural output to be obtained from intermediate structural inputs. Guo et al. [5] perform the topology optimization on latent variables of reduced dimensional space by using VAE and style transfer.

Second, deep learning applications on shape parameterization (i.e., design representation) have been developed. The parameterization to define the geometry is necessary in shape optimization. However, defining the variables in complicated geometry is extremely difficult, and the correlation between variables is too strong, which can hinder parameterization in a mathematical approach. Burnap et al. [6] show the parameterization possibility of the 2D shape of an automobile by using VAE. Moreover, Umetani [7] shows the parameterization of 3D meshes of an automobile by using autoencoder.

Third, deep learning has been applied to the metamodeling and simulation-based optimization. Many researches tried to apply deep learning to computational fluid dynamics (CFD) because CFD simulation has high computational cost. Guo et al. [8] propose the CNN model to predict the responses of CFD simulation for the 2D shape of an automobile, and Tompson et al. [9] accelerate Eulerian fluid simulation by approximating linear system with CNN. Farimani et al. [10] also propose the integration of cGAN to solve the steady-state problem for heat conduction and incompressible fluid flow.

Fourth, deep learning application on material design has been developed because of the direct relationship between the density of element in material structure and pixels of images, resulting to easy transformation of the domain from the material structures to images. Yang et al. [12] obtain optimal microstructure by using Bayesian optimization framework where the microstructure is mapped into low-dimensional latent variables by using GAN. Cang et al. [11] propose a feature extraction method to convert the microstructure to low-dimensional design space through a convolutional deep belief network. Cang et al. [13] show that generating an arbitrary amount of microstructure by a small amount of training data is available, proposing VAE under morphology constraint.

Finally, the application of deep learning on design preference estimation has been developed. Burnap et al.

[14] improve the prediction accuracy of the customer's preference model by learning restricted Boltzmann machine (RBM) with the original design variables as input and extracting the future. Pan et al. [15] propose the learning preference on aesthetic appeal using Siamese neural network architecture with cGAN.

In addition, deep learning is used in various engineering designs. For instance, Dering and Tucker [23] have successfully mapped the form and function of the design by using 3D convolution neural network, and Dering and Tucker [24] propose an image generation model that integrates deep learning and big data.

The categorization of the aforementioned research is not independent from one another, thus allowing integration and subsequently enhancing the conventional design optimization process.

2.2 Deep Generative Models

2.2.1. Generative Adversarial Networks (GANs)

GANs are designed to infer the data generating a distribution $p_{\text{data}}(x)$ by making the model distribution generated by generator $p_g(G)$ to be close to the real data distribution where x is the variable of real data and G is the generator's differentiable function with parameters θ_g . Function G has an input noise variable z and tries to map it to the real data space by adjusting θ_g , thus represented as $G(z; \theta_g)$. Similarly, the discriminator's differentiable function is derived as $D(x; \theta_d)$, which attempts to predict the probability that the input is from the real dataset. The zero-sum game of maximizing the discriminator and generator is equivalent to maximizing $\log D(x)$ and minimizing $\log (1 - D(G(z)))$ [25, 26] as

$$\begin{aligned} & \min_G \max_D V(D, G) \\ & = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \end{aligned} \quad (1)$$

With this standard GAN structure, various GANs have been developed by modifying the generator, discriminator, or objective function.

To overcome the notorious difficulty in training GANs, deep convolutional GANs (DCGANs) provide a stable training model, which works on various datasets by constructing a convolutional neural network in the generator and discriminator. In addition, DCGANs suggest certain techniques such as removing a fully-connected layer on top, applying batch normalization, and using the leaky rectified linear unit activation function [27].

Adversarially learned inference (ALI) and bidirectional GANs (BiGANs) adopt the encoding-decoding model to the generator to improve the quality of generated samples in an efficient way. Moreover, BiGANs emphasize taking advantages of learned features [28, 29]. For high-resolution images with stable convergence and scalability, energy-based GANs (EBGANs) have proven to produce realistic 128×128 images. EBGANs consider the discriminator as an energy function and the energy as the reconstruction error. From this point of view, an autoencoder architecture is used for the discriminator [30]. The autoencoder consists of encoder and decoder functions. The input value is transformed through the encoder and is restored to its original form again through the decoder [1]. Wasserstein GANs (WGANs) approach the way to obtain good image quality by changing the distance measure of two probability distributions. WGANs show that the earth-mover distance, which is also called Wasserstein-1, provides a differentiable function, and, thus, produces meaningful gradients, whereas Kullback-Leibler and Jensen-Shannon divergence in previous research do not when two probability distributions are disjointed [26]. Boundary Equilibrium GANs (BEGANs) also use the Wasserstein distance as a measure of convergence. BEGANs present an equilibrium concept, balancing the discriminator and generator in training and the numerical way of global convergence [31].

Aside from enhancing the image quality, the way to control the mode of generated outputs is presented by cGANs [32] and InfoGAN (Information maximizing Generative Adversarial Nets) [33]. cGANs provide additional input values to the generator and discriminator for categorical image generation. Furthermore, InfoGAN lets the generator produce uncategorical images by adding a latent code that can be categorical and continuous. It is useful for finding hidden representations from large amounts of data. However, intentionally creating a specific image is still difficult.

Hitherto, many studies on GANs contribute to good image quality in terms of convergence and stability. However, GANs are still weak in utilizing from the design engineering point of view, such as uneven image quality from the same saving point of model, especially when relatively small amounts of training data are given and images have insufficient engineering features.

2.2.2 Boundary Equilibrium GANs (BEGANs)

This paper employs BEGAN among GANs for the proposed framework because it provides a robust visual quality in a fast and stable manner. The autoencoder architecture as the discriminator used in EBGANs is also introduced by BEGANs. Similar to WGANs, BEGANs use the Wasserstein distance as a measure of convergence. With these techniques, BEGANs achieve reliable gradients that are difficult for high-resolution 128×128 images.

Rather than trying to match the probability distribution of real data, BEGANs focus on matching autoencoder loss distribution. It measures the loss, which is the difference between the sample and its output that passed through the autoencoder. Subsequently, a lower bound of the Wasserstein distance between the autoencoder loss distribution of real and that of generated samples is derived. The autoencoder loss function, $\mathcal{L} : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^+$, is defined as

$$\mathcal{L}(v) = |v - A(v)|^\eta \quad (2)$$

where

$$\begin{cases} \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x} & \text{is an autoencoder function} \\ \eta \in \{1, 2\} & \text{is a target norm} \\ v \in \mathbb{R}^{N_z} & \text{is a sample of dimension } N_z \end{cases}$$

Applying Jensens' inequality, the lower bound of the Wasserstein distance is derived as

$$|m_1 - m_2| \quad (3)$$

where $m_i \in \mathbb{R}$ is the mean of autoencoder loss distribution. For the maximization of Eq. (3) for the discriminator with $m_1 \rightarrow 0$ and $m_2 \rightarrow \infty$, BEGANs' objective function is described as minimizing the discriminator's autoencoder loss function \mathcal{L}_D and generator's one \mathcal{L}_G as the following where θ_D and θ_G are the parameters of the discriminator and generator, $G : \mathbb{R}^{N_z} \rightarrow \mathbb{R}^{N_x}$ is the generator function, $z \in [-1, 1]^{N_z}$ are uniform random samples of dimension N_z , and z_D and z_G are samples from z . The objective functions are defined as

$$\begin{cases} \mathcal{L}_D = \mathcal{L}(x) - k_t \mathcal{L}(G(z_D)) & \text{for } \theta_D \\ \mathcal{L}_G = \mathcal{L}(G(z_G)) & \text{for } \theta_G \\ k_{t+1} = k_t - \lambda_k (\gamma \mathcal{L}(x) - \mathcal{L}(G(z_G))) & \text{for } \gamma = \frac{\mathbb{E}[\mathcal{L}(G(z))]}{\mathbb{E}[\mathcal{L}(x)]} \end{cases} \quad (4)$$

where $k_t \in [0, 1]$ is a control factor to determine how much $\mathcal{L}(G(z_D))$ is reflected during gradient descent; λ_k is a proportional gain for k such as the learning rate in machine learning terms; and $\gamma \in [0, 1]$ is a diversity ratio that results in high image diversity as it increases. Given that k_t is changed in every training step to maintain $\mathbb{E}[\mathcal{L}(G(z))] = \gamma \mathbb{E}[\mathcal{L}(x)]$ for the equilibrium, global measure of convergence is regarded as the closest reconstruction $\mathcal{L}(x)$ with minimum absolute value of proportional control algorithm error $|\gamma \mathcal{L}(x) - \mathcal{L}(G(z_G))|$ [31] as

$$M_{global} = \mathcal{L}(x) + |\gamma \mathcal{L}(x) - \mathcal{L}(G(z_G))| \quad (5)$$

2.3 Topology Optimization

2.3.1 Density-based Approach

Topology optimization is commonly referred to the material distribution method developed and spread to a wide range of disciplines. The basic concept is how to distribute materials in a given design domain without any preconceived design [34-36]. In this study, the compliance minimization related to the stiffness of a structure has been carried out to redesign existing wheels. Many approaches such as homogenization and level-set methods can be applied, but we choose the density-based approach where material distribution is parameterized by the density of elements. Especially, solid isotropic material with penalization (SIMP) is a procedure that implicitly penalizes intermediate density values to lead to the black-and-white design. The basic formulation of SIMP in compliance minimization can be written as [36,37]

$$\begin{aligned} \min \quad & c(\mathbf{x}) = \mathbf{U}^T \mathbf{K} \mathbf{U} = \sum_{e=1}^{NE} \mathbf{u}_e^T (E_e(x_e) \mathbf{k}_0) \mathbf{u}_e \\ \text{s.t} \quad & V(\mathbf{x}) / V_0 = f \\ & \mathbf{K} \mathbf{U} = \mathbf{F} \\ & 0 \leq x_e \leq 1, \quad e = 1, \dots, N_e \end{aligned} \quad (6)$$

where \mathbf{U} is a displacement vector; \mathbf{K} is a global stiffness matrix; $c(\mathbf{x})$ is the compliance; \mathbf{u}_e is an element displacement vector; \mathbf{k}_0 is an element stiffness matrix; f is the volume fraction; N_e is the number of elements; x_e is the design variable (i.e., density) of element e ; and $V(\mathbf{x})$ and V_0 are the material volume and design domain volume, respectively. In modified SIMP, the density that is directly associated with Young's modulus can be expressed as [38]

$$E_e(x_e) = E_{\min} + x_e^p (E_0 - E_{\min}) \quad (7)$$

where p is a penalization factor to ensure the black-and-white design, and E_{\min} is introduced to avoid numerical instability when the density of elements become zero.

Many studies have been done to enhance the topology optimization performance such as filtering techniques. In this study, we develop the code based on 99- and 88-line MATLAB codes, which are the simplest and most efficient two-dimensional topology optimization codes written in MATLAB [37,39]. Furthermore, we include the improved techniques developed recently such as a projection filter. Thus, we briefly explain algorithms such as sensitivity analyses and filtering techniques used in this study.

2.3.2 Sensitivity Analysis

In gradient-based optimization, the sensitivity analysis of objective and constraint functions with respect to each design variable is required to provide accurate search direction to the optimizer. Therefore, the sensitivity analysis with respect to the density of elements can be given by

$$\frac{\partial c}{\partial x_e} = -p x_e^{p-1} (E_0 - E_{\min}) \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e \quad (8)$$

and

$$\frac{\partial V}{\partial x_e} = \frac{\partial}{\partial x_e} \left(\sum_{e=1}^{NE} x_e v_e - V \right) = 1 \quad (9)$$

under the assumption that all elements have a unit volume. On the other hand, the optimality criteria (OC) method, one of the classical approaches to structural optimization problems, is employed in this paper. The OC method updates the design domain as

$$x_e^{new} = \begin{cases} \max(0, x_e - m) & \text{if } x_e B_e^\eta \leq \max(0, x_e - m) \\ \min(1, x_e + m) & \text{if } x_e B_e^\eta \geq \min(1, x_e + m) \\ x_e B_e^\eta & \text{otherwise} \end{cases} \quad (10)$$

where m is a positive move-limit and η is a numerical damping coefficient, and

$$B_e = \frac{-\frac{\partial c}{\partial x_e}}{\lambda \frac{\partial V}{\partial x_e}} \quad (11)$$

The Lagrange multiplier related to volume fraction constraint can be obtained from a bisection algorithm that is one of the popular root-finding algorithms. The termination criteria for the convergence can be written as

$$\|\mathbf{x}_{new} - \mathbf{x}\|_\infty \leq \varepsilon \quad (12)$$

where ε is the tolerance usually set as a relatively small value such as 0.01.

2.3.3 Filtering Techniques

For the assurance of the existence of well-posed and mesh-independent solutions, several strategies to avoid a checkerboard pattern and gray-scale issues are introduced. In this study, we apply so-called three-field SIMP, which has a projection scheme [40]. Three-field means the original density, filtered density, and projected density. Detailed descriptions can be seen in the literature.

The basic filters applied to topology optimization are sensitivity and density filters, which are used in one-field and two-field SIMP, respectively. The main idea of both techniques is to modify sensitivity or physical element density to be a weighted average of the neighborhood. The neighborhood is defined on the basis of the distance from the center of the element, and the maximum distance to include in the neighborhood is a user-specified parameter referred to the mesh-independent radius. The sensitivity filter can be written as [41]

$$\frac{\partial c}{\partial x_e} = \frac{1}{x_e \sum_{f=1}^N H_f} \sum_{f=1}^N H_f x_f \frac{\partial c}{\partial x_f} \quad (13)$$

where the convolution operator can be written as

$$H_f = r_{\min} - \text{dist}(e, f), \quad (14)$$

where subscript f means one of the elements that the center-to-center distance expressed as $\text{dist}(e, f)$ between elements is smaller than r_{\min} .

The density filter defines the physical density with weighted averaging. The weighted average concept is the same in the sensitivity filter as Eq. (13), but the density is filtered instead of sensitivity expressed as [42,43]

$$\tilde{x}_e = \frac{\sum_{f=1}^N H_f x_f}{\sum_{f=1}^N H_f} \quad (15)$$

Therefore, original and filtered densities can be referred to as a design variable and physical density, respectively. The sensitivity analysis with respect to design variables should be modified by introducing the physical density using a chain rule. A detailed description can be seen in the literature [44].

The weighted average is used in both filtering methods to avoid the checker-board pattern in an optimum design. However, the density filter can induce gray transitions between solid and void regions. Thus, the third field of density, or the so-called projection filter, is introduced. It mitigates the gray transition problem by projecting to solid and void usually using a smoothed Heaviside projection [38,44,45]. In this study, we use the Heaviside projection filter on the filtered density obtained from Eq. (15). The projection filter can be written as

$$\bar{\tilde{x}}_e = 1 - e^{-\beta \tilde{x}_e} + \tilde{x}_e e^{-\beta} \quad (16)$$

where β is a parameter related to slope of projection and can be updated through the optimization. In three-field SIMP with projection filter, the sensitivity analysis is modified compared with Eq. (13) because the finite element analysis is performed based on the physical density obtained from Eq. (16). The sensitivity analysis with respect to design variables can be easily derived using the chain rule.

3. Generative Design Exploration Framework

A generative design exploration framework was proposed, which integrates deep generative models and topology optimization. Fig. 1 shows the entire process which consists of two main parts (i.e., iterative design exploration and design evaluation) and nine stages.

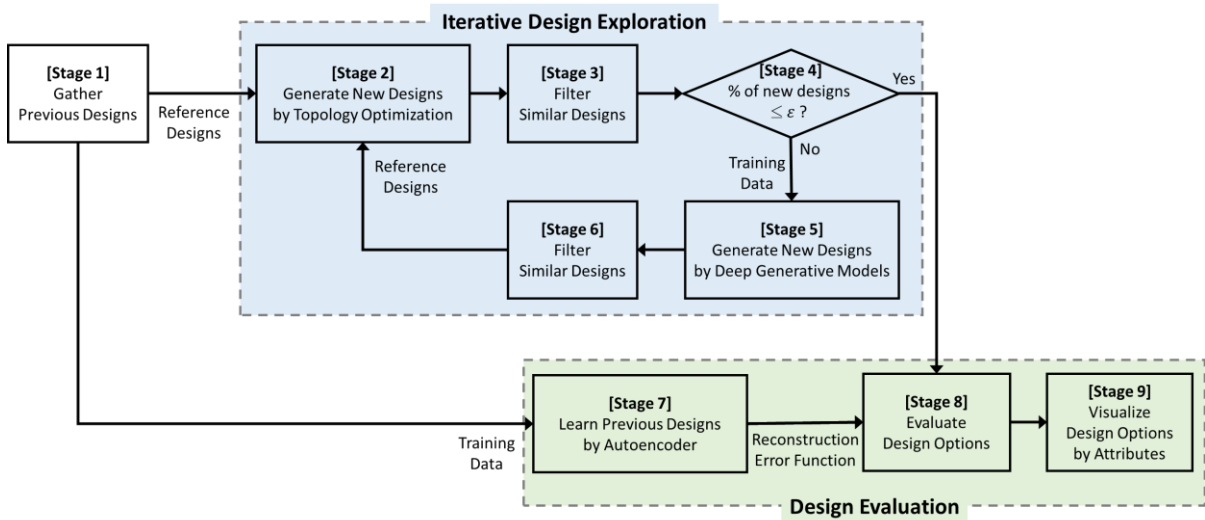


Fig. 1. Generative design exploration framework

The iterative design exploration is the integration of topology optimization and deep generative models to produce new designs, and design evaluation quantifies and evaluates the novelty and main attributes of new designs. Each stage is explained as follows:

In Stage 1, the previous designs in the market and the industry are collected as reference designs for Stage

2. In this study, the reference design is defined as a benchmark design to create new designs in topology optimization.

In Stage 2, new designs are obtained by topology optimization based on reference designs. In this step, topology optimization has a multi-objective function including compliance minimization, which represents engineering performance and the difference (i.e., pixel-wise L1 distance) minimization from the reference design, which represents aesthetics. Different designs can be obtained by varying the relative weights of each objective because a tradeoff between two objectives is available.

In Stage 3, similar designs gathered from topology optimization are filtered out by similarity criterion configured for the user-specified threshold.

In Stage 4, the ratio of the number of new designs in current iteration to the number of total designs in the previous iteration is calculated. If it is smaller than the user-specified threshold, then exit the iterative design exploration and jump to Stage 8; otherwise, proceed to Stage 5.

In Stage 5, new designs are created by deep generative models after learning aggregated designs in the current iteration, and they are used as reference designs in Stage 2 after filtering out similar designs in Stage 6.

The iterative design exploration should be continuously performed from Stage 2 to 6 until the termination criterion is satisfied, i.e., until the amount of generated designs are substantial. This iterative process has the purpose of creating a large number of various designs by starting from a small number of previous designs in Stage 1.

The process in Stage 7 is independent of the iterative design exploration, which involves the building of a loss function (i.e., reconstruction error function) employing autoencoder trained by previous designs of Stage 1. This function can be used to evaluate design novelty in comparison with the previous designs.

In Stage 8, design options obtained from iterative design exploration have to be evaluated on the basis of various design attributes that are essential to the designers. It can evaluate not only the novelty of created designs by using the reconstruction error function in Stage 7 but also the physical quantities such as the volume and compliance of designs.

Finally, the tradeoff between attributes is demonstrated as plotting of designs in each axis of design attributes, and the proper designs can be chosen accounting to the relative importance of each attribute.

The proposed process is applied to the 2D wheel of an automobile as a case study. Detailed descriptions of methodologies used in each stage are presented in Sections 3.1 and 3.2.

3.1. Iterative Design Exploration

3.1.1. Topology Optimization (Stage 2)

Stage 2 generates new engineering designs through topology optimization reflecting the shape of the reference designs that can be either previous wheel designs from Stage 1 or generated designs from deep generative models (Stages 5 and 6). A number of engineering performances has to be considered. However, compliance obtained from the static analysis is generally employed in topology optimization, which is called compliance minimization. Fig. 2 sketches the design domain and boundary conditions for 2D wheel design. The original design domain is 128 by 128 elements, and the reference designs also have 128 by 128 pixels. The outer ring of the wheel is set to the non-design region to maintain the shape of the rim, and inner region is set to fixed boundary condition for connecting parts. Therefore, spoke is the main component in the design domain.

The element stiffness matrix is based on 4-node bilinear square elements in the 88-line MATLAB code [39]. Normal and shear forces are uniformly exerted along the surface, which are common load conditions in the 2D wheel optimization. The normal force and the shear force represent uniform tire pressure and tangential traction, respectively. The ground reaction induced by vehicle weight is disregarded because it requires an additional symmetric condition.

The ratio between normal and shear force is a user-specified parameter that can significantly change the optimized wheel design. The force ratio is defined as the magnitude of normal force divided by shear force.

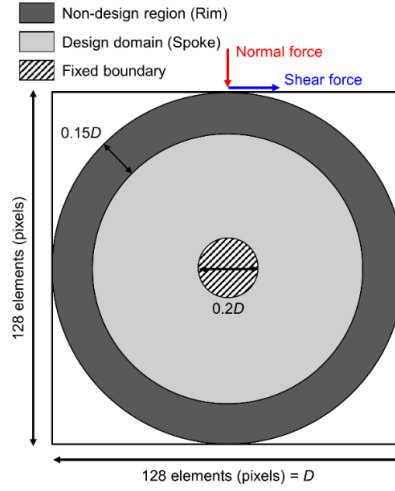


Fig. 2. Design domain and boundary conditions of a 2D wheel design

However, varying the boundary conditions only can limit the production of meaningful and diverse designs. Thus, a new objective function in topology optimization was introduced so that the design generator can produce engineering optimization while maintaining the shape of various reference designs. The modified objective function can be formulated as

$$f(\mathbf{x}) = \mathbf{U}^T \mathbf{K}(\mathbf{x}) \mathbf{U} + \lambda \|\mathbf{x}^* - \mathbf{x}\|_1 \quad (17)$$

where $\mathbf{U}^T \mathbf{K}(\mathbf{x}) \mathbf{U}$ is compliance, λ is a user-specified similarity parameter, and \mathbf{x}^* is elements of the reference design. Therefore, the L1-norm between generated and reference design represents the similarity. L1-norm is more preferred than L2-norm because it can alleviate the blurring of design, and shows better design quality. All other processes are identical with conventional three-field SIMP explained in Section 2.3. Reference design represented as \mathbf{x}^* is a binary matrix. Hence, sensitivity analysis for additional similarity term can be expressed as

$$\frac{\partial}{\partial x} (\lambda \|\mathbf{x}^* - \mathbf{x}\|_1) = -\lambda \mathbf{x}^* \quad (18)$$

The above expression means that if an element in reference design is solid, then the sensitivity is set to $-\lambda$, and 0 otherwise to avoid the positive sensitivity for the optimizer in topology optimization. In other words, the objective is to maintain the reference design in proportion to similarity, and additional sensitivity is added on solid elements in the reference design.

Consequently, five discrete levels of similarity parameter and force ratio are configured to generate new designs from topology optimization as listed in Table 1. For instance, if 100 reference designs are available, then 2500 designs (100 reference designs \times 5 similarity parameter levels \times 5 force ratio levels) can be obtained from topology optimization. The type of condition and the number of levels can be determined by the designers such as voxel sizes, solver parameters, and number of iterations in topology optimization [19].

Table 1. Condition parameter levels

Condition	Levels
Similarity	5 levels: 0.0005/0.005/0.05/0.5/5
Force ratio	5 levels: 0/0.1/0.2/0.3/0.4

3.1.2. Design Filtering and Termination Criteria (Stages 3, 4, and 6)

In Stage 3, similar designs are eliminated by evaluating the similarity in the aggregated dataset to reduce computational cost for irrelevant designs. This study uses the pixel-wise L1 distance as a criterion to judge a similar design, which is set to 10^3 . If we set this value much tightly, then the number of designs to be generated

will be reduced, but the differentiation between designs will be improved. This process is repeated in Stage 6 to filter out the designs generated from the deep generative models.

In Stage 4, the threshold for termination criteria of iterative design exploration is set to 0.3, which means the iterative process will be stopped and proceed to Stage 8 when the ratio of the number of new designs in the current iteration to the number of designs of the previous iteration is less than 0.3. This value can be adjusted according to time and computing power.

3.1.3. Deep Generative Models (Stages 5)

BEGAN was used among deep generative models which is introduced in Section 2.2.2. Designers can utilize various state-of-the-art deep generative models simultaneously for this stage. The model trained in our research has a network architecture that only consists of the nearest neighbor up-sampling and 3×3 convolutions with stride 2 for down-sampling and the exponential linear unit activation function. Fig. 3 illustrates this architecture. Generator and decoder of discriminator use the same network. Adam optimizer was used with a learning rate of 0.00008, initial value of k_t as 0, λ_k as 0.001, γ as 0.7, and minibatch size of 16 (see Section 2.2.2). The learning rate parameter is set with reference to the settings of papers that studied image quality by using BEGAN [46,47].

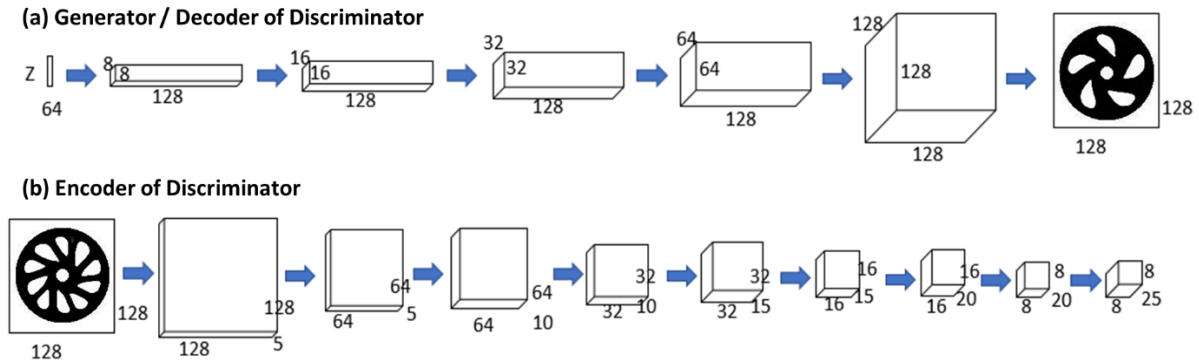


Fig. 3. Network architecture of BEGAN model

3.2 Design Evaluation

In Stage 7, autoencoder is used to evaluate design novelty. Reconstruction errors, which are the loss functions of autoencoder, are widely used to detect anomaly [48]. The idea is that autoencoder can effectively reconstruct similar data to training data but not dissimilar data. This study assumes that design novelty can be measured the same way as anomaly detection. Previous designs in Stage 1 used training data for autoencoder, and then trained autoencoder can calculate reconstruction error of new designs generated by iterative design exploration. The new design which has the higher value of reconstructing error is regarded as the design which has the more novelty. This study employs the autoencoder structure which is used in BEGAN as the discriminator (see Fig. 3), and the equation of reconstruction errors is the same as Eq. (2).

In addition, a regression model can be used as an alternative of autoencoder when previous design data are insufficient. Topology design results (in Stage 2) at the first iteration have similarity parameter (λ), so that regression model can be built where output y is set to the similarity parameter. VGG16 [49], which is one popular CNN for regression model, was tested. Results show that CNN-based regression can also predict similarity which is contrary to novelty.

4. Results and Discussion

This section shows the results of a case study applying the proposed framework to 2D wheel design. In Stage 1, frontal wheel designs were collected in the market by web crawling and converted it to binary images. A total of 658 binary images of the wheel are collected through post-processing as reference designs for the first iteration of iterative design exploration.

4.1. Iterative Design Exploration

4.1.1 Design Exploration by Topology Optimization (Stages 2 and 3)

Topology optimization is performed in parallel according to similarity and force ratio parameters listed in Table 1. Fig. 4 shows an example of optimized designs according to five levels of force ratio when the similarity is 0.0005. A large shear force is observed to make many thin spokes of whirlwind shape. On the other hand, large normal force makes thick and less curved spokes.

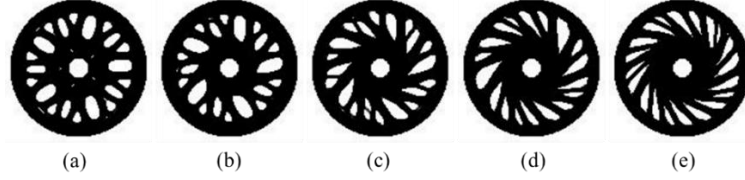
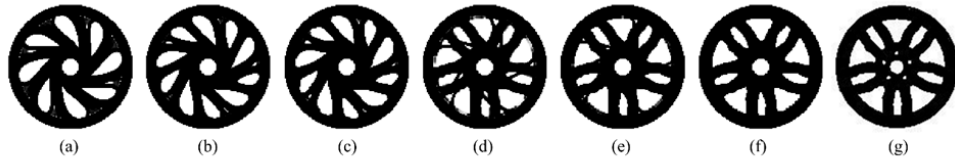


Fig. 4. Topology optimization of wheel design when the force ratio is set to (a) 0, (b) 0.1, (c) 0.2, (d) 0.3, (e) 0.4

Fig. 5 shows the optimal designs under five levels of similarity (i.e., (b) to (f)). (g) shows the reference designs, and (a) shows the result when reference design is unused. The ratio between normal and shear force is set to 0.1 in case of reference design A and 0.2 in case of reference design B, and the volume fraction is identical to reference designs. Table 2 lists the similarity and compliance of each optimal design. The optimal designs from topology optimization evidently indicate the tradeoff between engineering performance and similarity to the reference design.

Reference Design A



Reference Design B

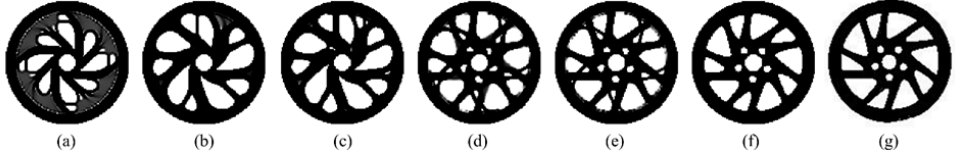


Fig. 5. Topology optimization results when the similarity is set to (a) 0, (b) 0.0005, (c) 0.005, (d) 0.05, (e) 0.5, (f) 5, and (g) reference design

Table 2. Similarity and compliance of each reference design in Fig. 5

Similarity		0	0.0005	0.005	0.05	0.5	5
Compliance	Reference design A	5.28	5.23	6.17	7.14	8.87	9.28
	Reference design B	8.88	8.90	8.94	9.76	10.71	13.02

Fig. 6 demonstrates the effectiveness of the proposed objective function in Eq. (17), which states that the proposed method can yield different designs reflecting the shape of reference design when other boundary conditions are the same for all optimized design such as force ratio. The described results suggest that the proper range of similarity that optimal designs are continuously changed varies depending on the reference design. Therefore, experimental investigation is encouraged in advance.

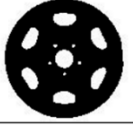











Reference Design						
Optimized Design						

Fig. 6. Optimized designs under the same boundary conditions and different reference design

Consequently, 1619 new designs have been created after filtering at Stage 3. Stage 4 is passed automatically in the first iteration of the iterative design exploration, and the ratio of new designs is calculated as a criterion from the second iteration.

4.1.2. Design Exploration by Deep Generative Model (Stages 5 and 6)

A total of 2304 designs were identified after Stage 4, where 658 previous designs were identified at Stage 1 and 1619 designs were obtained through Stage 4. These 2304 designs are used for training designs of BEGAN at Stage 5. The training takes around three hours on four GTX 1080 GPUs in parallel. Fig. 7 presents examples of 128×128 image from the BEGAN generator, and Fig. 8 describes its global convergence showing $\mathcal{M}_{global} \approx \mathcal{L}(x)$. A total of 128 designs were achieved through filtering at Stage 6. These designs were used as reference designs for Stage 2 at the second iteration of iterative design exploration. New 385 topology designs are regenerated after topology optimization (Stages 2 and 3).

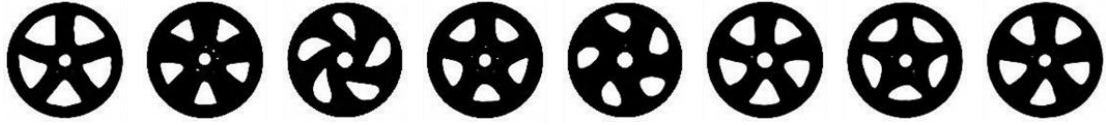


Fig. 7. 128×128 wheel designs by BEGAN

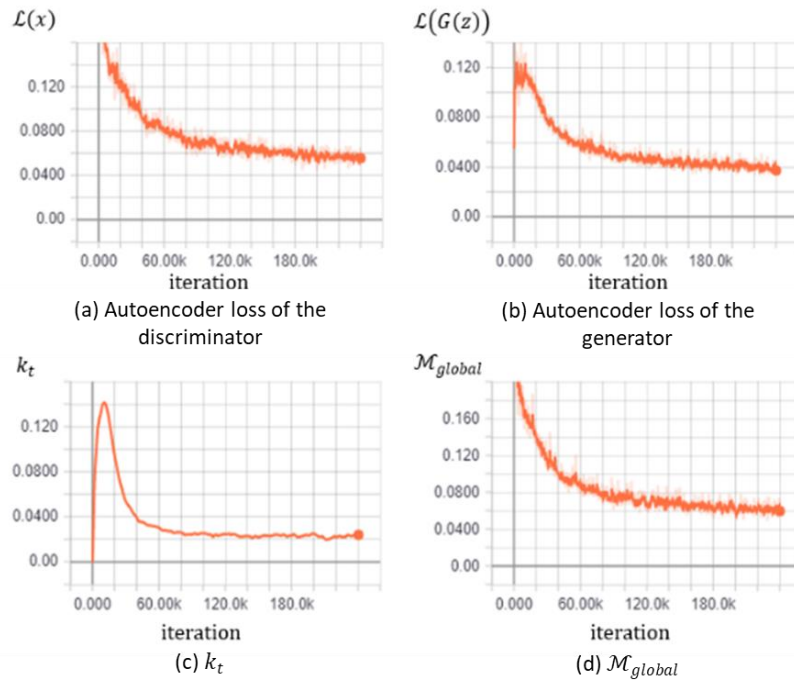


Fig. 8. Convergence results of BEGAN

DCGAN was additionally tested for Stage 5, which generated more varied and complex shapes than BEGAN. However, reference designs generated by BEGAN, which were simpler and sharper, yielded better topology designs through Stage 2 than DCGAN.

4.2. Design Evaluation

4.2.1. Novelty Evaluation by Autoencoder (Stage 7)

Findings show that 80% and 20% of the previous designs in Stage 1 were used as training data and test data, respectively, for the autoencoder. Fig. 9 shows that the loss is converged well. Fig. 10 shows examples comparing reconstruction results between test data of previous designs and generated designs. Designs similar to previous ones portray satisfactory reconstruction, while dissimilar designs portray otherwise.

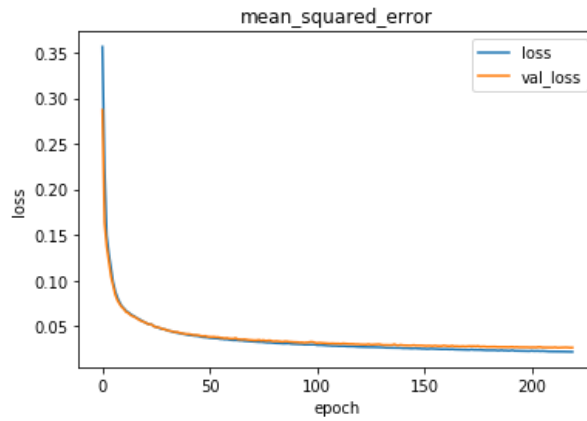


Fig. 9. Reconstruction error of autoencoder

	Previous Designs			Generated Designs		
Input						
Output (Reconstruction)						

Fig. 10. Comparison between previous designs and generated designs in reconstruction of autoencoder

4.2.2. Evaluation and Visualization (Stages 8 and 9)

Finally, 2004 new designs are generated after two iterations, which are not included in the previous design set. Examples of design options are shown in Fig. 11. A 3D scatter plot for 2004 design options was crafted by using three design attributes (novelty, cost, and compliance) as axis (Fig. 12(a)). Each attribute value used in the plot is normalized from 0 to 1. Fig. 12(b) shows that trade-offs between compliance and cost make a smooth Pareto curve, because designs are all topologically optimized. On the Pareto curve, two designs are shown as examples, one with the lowest cost and the other with lowest compliance. Figures 12(c) and 12(d) show trade-offs between novelty and other attributes. Assuming novelty is a positive trait, two designs located on the Pareto curves are shown as examples in the figures, respectively.

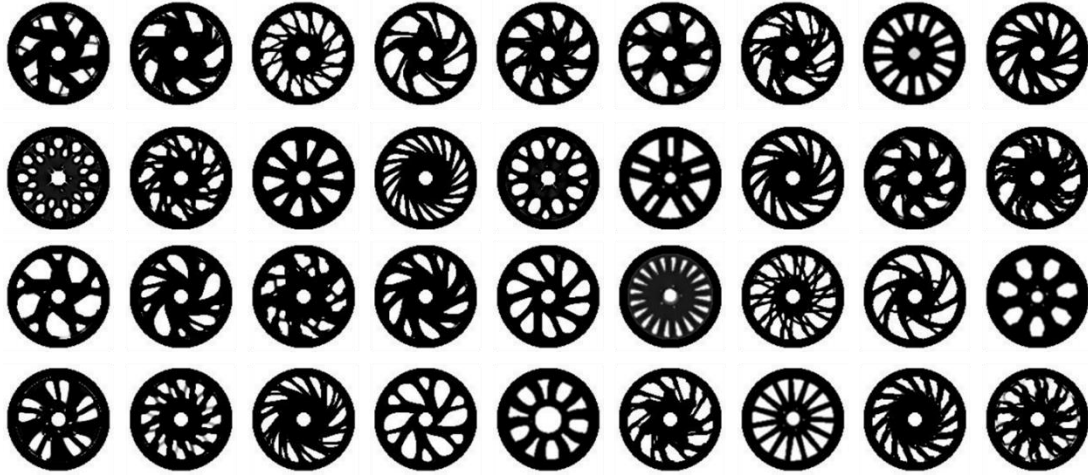


Fig. 11. Generated design options

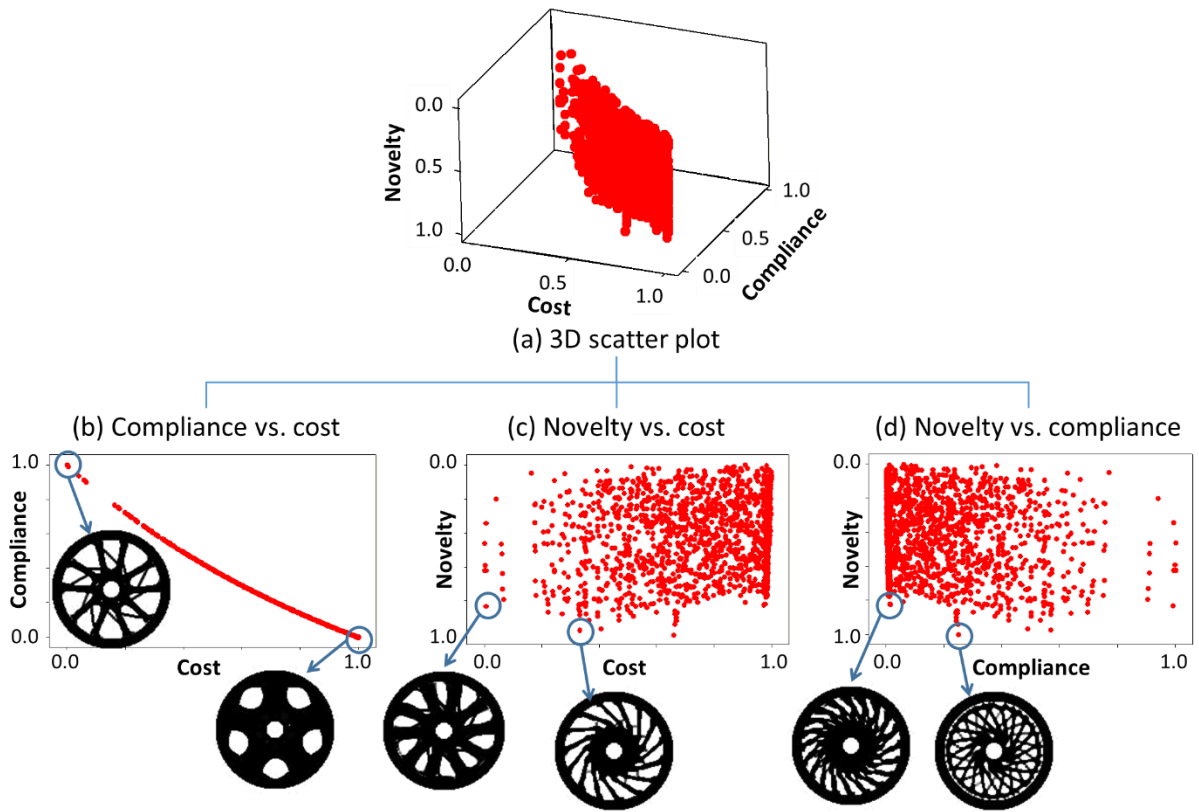


Fig. 12. Visualized design options by three attributes: novelty, cost, and compliance

Designers can trade-off three attributes, and select designs according to their design purpose and preference. Then, they can create a 3D design based on a 2D design and prototype it by 3D printing. Fig. 13 shows an example of a 3D wheel design (i.e., STL file for 3D printing) after selecting a 2D design.



Fig. 13. An example of a 3D wheel design using the selected 2D design

5. Conclusion

This study proposes a design automation framework that generates various designs ensuring engineering performance and aesthetics, and its effectiveness is demonstrated by 2D wheel design case study. The contribution of this research can be addressed as follows:

First, this research considers engineering performance and aesthetics simultaneously. The proposed framework is able to control the similarity with reference designs and engineering performance as a multi-objective function.

Second, a large number of designs starting from a small number of designs was generated. An iterative process is proposed where topology optimization is conducted to create training data for the deep generative models, and output designs from deep generative models are used as reference designs for topology optimization again.

Third, the proposed framework offered diverse designs in comparison with the conventional generative design. Moreover, increased diversity is accounted to the use of reference designs generated by deep generative models.

Fourth, the robustness on quality of designs is improved. The conventional deep generative models are prone to induce the mode collapse and large variance of the quality. However, results of the deep generative model in the proposed framework are refined through topology optimization instead of direct utilization (e.g., post-processing).

Finally, comparison between the novelty of generated design and the previous designs can be evaluated. The reconstruction error of autoencoder is used for the index of similarity to existing designs.

This research is performed on a 2D design space and pixel-wise images, which is identified as its main drawback. Thus, 3D design application with voxel data should be further investigated for practical design, and various case studies should be tested. In addition, a recommendation system that suggests the appropriate designs (i.e., predicting the preference of designers and consumers among design candidates) will be carried out.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) under Grant (No. 2017R1C1B2005266) and NRF under Grant (No. 2018R1A5A7025409). The authors would like to thank Ah Hyeon Jin, Seo Hui Joung, Gyuwon Lee, and Yun Ha Park who are undergraduate interns of Sookmyung Women's University, for pre- and post-processing of data, and also Yonggyun Yu of Korea Advanced Atomic Research Institute for his advice and ideas.

References

- [1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning| The MIT Press. Cambridge, Massachusetts.
- [2] Sosnovik, I., & Oseledets, I. (2017). Neural networks for topology optimization. arXiv preprint arXiv:1709.09578.
- [3] Yu, Y., Hur, T., Jung, J., & Jang, I. G. (2018). Deep learning for determining a near-optimal topological design without any iteration. *Structural and Multidisciplinary Optimization*, 1-13.
- [4] Banga, S., Gehani, H., Bhilare, S., Patel, S., & Kara, L. (2018). 3D Topology Optimization using Convolutional Neural Networks. arXiv preprint arXiv:1808.07440.
- [5] Guo, T., Lohan, D. J., Cang, R., Ren, M. Y., & Allison, J. T. (2018). An indirect design representation for topology optimization using variational autoencoder and style transfer. In 2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference (p. 0804).
- [6] Burnap, A., Liu, Y., Pan, Y., Lee, H., Gonzalez, R., & Papalambros, P. Y. (2016). Estimating and exploring the product form design space using deep generative models. In ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. V02AT03A013-V02AT03A013). American Society of Mechanical Engineers.
- [7] Umetani, N. (2017). Exploring generative 3D shapes using autoencoder networks. In SIGGRAPH Asia 2017 Technical Briefs (p. 24). ACM.
- [8] Guo, X., Li, W., & Iorio, F. (2016). Convolutional neural networks for steady flow approximation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 481-490). ACM.
- [9] Tompson, J., Schlachter, K., Sprechmann, P., & Perlin, K. (2017). Accelerating eulerian fluid simulation with convolutional networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 3424-3433). JMLR.
- [10] Farimani, A. B., Gomes, J., & Pande, V. S. (2017). Deep learning the physics of transport phenomena. arXiv preprint arXiv:1709.02432.
- [11] Cang, R., Xu, Y., Chen, S., Liu, Y., Jiao, Y., & Ren, M. Y. (2017). Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design. *Journal of Mechanical Design*, 139(7), 071404.
- [12] Yang, Z., Li, X., Brinson, L. C., Choudhary, A. N., Chen, W., & Agrawal, A. (2018). Microstructural materials design via deep adversarial learning methodology. *Journal of Mechanical Design*, 140(11), 111416.
- [13] Cang, R., Li, H., Yao, H., Jiao, Y., & Ren, Y. (2018). Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model. *Computational Materials Science*, 150, 212-221.
- [14] Burnap, A., Pan, Y., Liu, Y., Ren, Y., Lee, H., Gonzalez, R., & Papalambros, P. Y. (2016). Improving design preference prediction accuracy using feature learning. *Journal of Mechanical Design*, 138(7), 071404..
- [15] Pan, Y., Burnap, A., Hartley, J., Gonzalez, R., & Papalambros, P. Y. (2017). Deep design: Product aesthetics for heterogeneous markets. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1961-1970). ACM.
- [16] Shea, K., Aish, R., & Gourtovaia, M. (2005). Towards integrated performance-driven generative design tools. *Automation in Construction*, 14(2), 253-264.
- [17] Krish, S. (2011). A practical generative design method. *Computer-Aided Design*, 43(1), 88-100.
- [18] McKnight, M. (2017). Generative Design: What it is? How is it being used? Why it's a game changer. *KnE Engineering*, 2(2), 176-181.

- [19] Matejka, J., Glueck, M., Bradner, E., Hashemi, A., Grossman, T., & Fitzmaurice, G. (2018). Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 369). ACM..
- [20] Autodesk (2019), Generative Design, <https://www.autodesk.com/solutions/generative-design>
- [21] Oh, S., Jung, Y., Lee, I., & Kang, N. (2018). Design Automation by Integrating Generative Adversarial Networks and Topology Optimization. In *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V02AT03A008-V02AT03A008). American Society of Mechanical Engineers.
- [22] Kang, N. (2014). Multidomain Demand Modeling in Design for Market Systems. PhD Thesis, University of Michigan.
- [23] Dering, M. L., & Tucker, C. S. (2017). A Convolutional Neural Network Model for Predicting a Product's Function, Given Its Form. *Journal of Mechanical Design*, 139(11), 111408.
- [24] Dering, M. L., & Tucker, C. S. (2017). Generative adversarial networks for increasing the veracity of big data. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 2595-2602). IEEE..
- [25] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).
- [26] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.
- [27] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- [28] Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., & Courville, A. (2016). Adversarially learned inference. *arXiv preprint arXiv:1606.00704*.
- [29] Donahue, J., Krähenbühl, P., & Darrell, T. (2016). Adversarial feature learning. *arXiv preprint arXiv:1605.09782*.
- [30] Zhao, J., Mathieu, M., & LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126*.
- [31] Berthelot, D., Schumm, T., & Metz, L. (2017). Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*.
- [32] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- [33] Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in neural information processing systems* (pp. 2172-2180)..
- [34] Bendsøe, M. P., & Kikuchi, N. (1988). Generating optimal topologies in structural design using a homogenization method. *Computer methods in applied mechanics and engineering*, 71(2), 197-224.
- [35] Bendsøe, M. P. (1989). Optimal shape design as a material distribution problem. *Structural optimization*, 1(4), 193-202.
- [36] Bendsoe, M. P., & Sigmund, O. (2013). *Topology optimization: theory, methods, and applications*: Springer Science & Business Media.
- [37] Sigmund, O. (2001). A 99 line topology optimization code written in Matlab. *Structural and multidisciplinary optimization*, 21(2), 120-127.
- [38] Sigmund, O. (2007). Morphology-based black and white filters for topology optimization. *Structural and Multidisciplinary Optimization*, 33(4-5), 401-424.
- [39] Andreassen, E., Clausen, A., Schevenels, M., Lazarov, B. S., & Sigmund, O. (2011). Efficient topology

optimization in MATLAB using 88 lines of code. *Structural and Multidisciplinary Optimization*, 43(1), 1-16.

[40] Sigmund, O., & Maute, K. (2013). Topology optimization approaches. *Structural and Multidisciplinary Optimization*, 48(6), 1031-1055.

[41] Sigmund, O. (1997). On the design of compliant mechanisms using topology optimization. *Journal of Structural Mechanics*, 25(4), 493-524.

[42] Bruns, T. E., & Tortorelli, D. A. (2001). Topology optimization of non-linear elastic structures and compliant mechanisms. *Computer methods in applied mechanics and engineering*, 190(26-27), 3443-3459.

[43] Bourdin, B. (2001). Filters in topology optimization. *International journal for numerical methods in engineering*, 50(9), 2143-2158.

[44] Guest, J. K., Prévost, J. H., & Belytschko, T. (2004). Achieving minimum length scale in topology optimization using nodal design variables and projection functions. *International journal for numerical methods in engineering*, 61(2), 238-254.

[45] Xu, S., Cai, Y., & Cheng, G. (2010). Volume preserving nonlinear density filter based on heaviside functions. *Structural and Multidisciplinary Optimization*, 41(4), 495-505.

[46] Vertolli, M. O., & Davies, J. (2017). Image quality assessment techniques show improved training and evaluation of autoencoder generative adversarial networks. *arXiv preprint arXiv:1708.02237*.

[47] Kocaoglu, M., Snyder, C., Dimakis, A. G., & Vishwanath, S. (2017). CausalGAN: Learning causal implicit generative models with adversarial training. *arXiv preprint arXiv:1709.02023*.

[48] Sakurada, M., & Yairi, T. (2014, December). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis* (p. 4). ACM.

[49] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.