Deep generative models for engineering design

Fan, J.

Chapter 6

Application in Industrial Development Processes

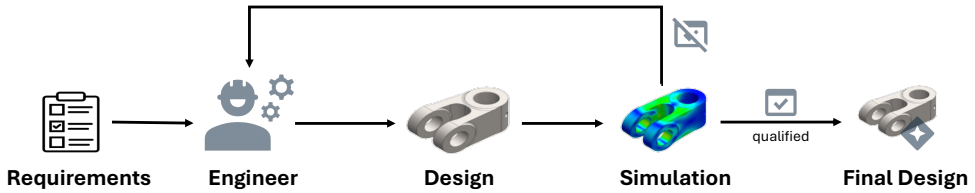
To demonstrate the industrial value of our work, it is essential to apply our methods in real-world industrial development processes, such as car design. This chapter focuses on two specific car components: A-pillars and rims. By evaluating with these components, we aim to assess the potential and capacity of our DGM-based approaches in aiding engineers with developing complex parts. This chapter is correspondingly aiming to address the research question 5: *How can DGMs be beneficial for real-world industrial design applications?*

6.1 DGM-based Generative Engineering Design

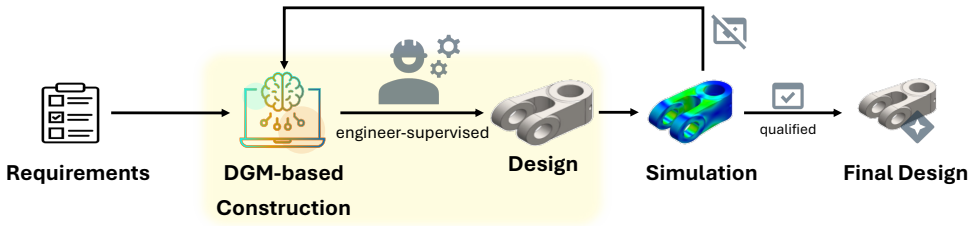
In the current industrial development process, while engineers can rely on a limited range of intelligent design tools, the initial design is often constructed manually based on the engineer's professional knowledge and predefined requirements (e.g., geometric constraints, cost limits, and other functional requirements). The initial design must be tested through certain simulations, and if the performance of certain functions is substandard, the engineer must modify the design, which happens in most cases. This iterative design process continues until the design has passed all simulations and meets the predefined requirements. This process is demonstrated in Figure 6.1a.

The current industrial development process has several key disadvantages:

- Limited design space exploration: The manual design approach constrains en-



(a) Traditional engineering design



(b) DGM-driven generative engineering design

Figure 6.1: Diagrams about traditional engineering design and DGM-driven generative engineering design.

gineers to a relatively narrow design space based on their own expertise and predefined requirements, limiting the potential for innovative or disruptive designs.

- Inefficient and error-prone design process: The heavy reliance on manual design construction by engineers, based on their own knowledge and experience, can be time-consuming, inefficient, and prone to human error, leading to suboptimal designs and longer development timelines.

- Underutilization of historical design knowledge: With the current manual design process, engineers are largely reliant on their own professional knowledge, rather than effectively leveraging the wealth of historical designs and accumulated organizational knowledge. This represents a significant missed opportunity, as the insights and experiences from past design efforts are not being efficiently captured and reapplied, leading to a wasteful duplication of effort.

This work focuses on enabling DGMs to directly generate engineering designs, with a certain level of supervision from engineers (e.g., design editing and generation conditioning). This is represented by the *yellow area* highlighted in Figure 6.1b. Unlike the traditional manual design construction process, the DGM-based design generation

is differentiable. This means the output design is numerically connected to the input parameters, allowing for gradient-based optimization. As demonstrated in the design editing experiments on BIKED (Section 2.4), introducing a gradient-based optimizer enables the iterative modification of the target design until it meets specific geometric requirements. When the simulation is differentiable (some of current simulation pipeline is already differentiable and modern AI-based simulation predictors are also differentiable) and the gradient can be calculated, the same optimization pipeline can be used to optimize the design to meet desired functional targets. Hereby, we come to a novel pipeline, shown in Figure 6.1b.

The key advantage of this DGM-based approach is that it allows for a more efficient and automated design exploration process, leveraging the differentiable nature of the generative model to guide the design towards optimal solutions, with the ability to incorporate both geometric and functional performance objectives. In this thesis, the whole pipeline would not be evaluated on real-world simulation, for the following reasons: the differentiable simulation is not available, simulation data differs from design data and the amount of trainable design data is not sufficient to enable this pipeline. But for the proof of the concept, we refer to the dragging edition in Section 2.4.

6.2 Automotive A-pillar Design

The current progress of DGM can generate feasible designs. Inspired by these results, we see the value of applying this method in actual industrial design cases. Therefore, this section will focus on the introduction of PoDM in the design process of a car A-pillar, using two-dimensional engineering blueprints as design data.

6.2.1 Background and Motivation

Federal Motor Vehicle Safety Standard known as FMVSS No.201U [171], which is provided by the National Highway Traffic Safety Administration (NHTSA) and Office of Vehicle Safety Compliance (OVSC), is the law guiding automobile manufacturers in designing the car interior as an “energy absorber” in crash. The law FMVSS No.201U requires that automobiles have to provide impact protection for passengers, which applies to passenger cars and trucks, buses and multipurpose passenger vehicles with a gross vehicle weight rating of 4 536 kilograms or less. For the impact test, Free Motion Headform (FMH) is implemented as the dummy head, which is presented in the middle of Figure 6.3. The FMH impactor is capable of propelling the FMH at

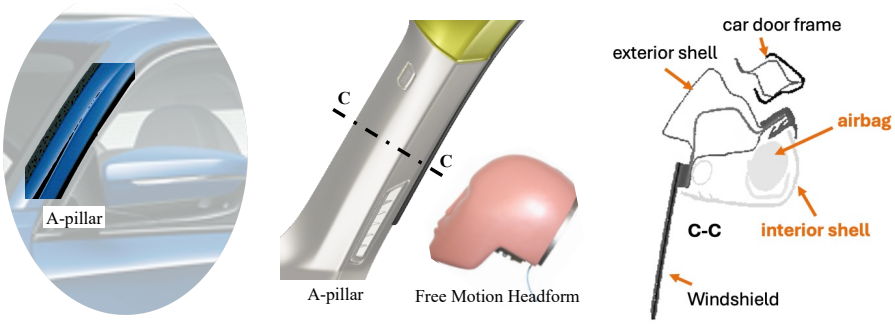


Figure 6.2: A vehicle A-pillar (*left*), an illusion of the slicing on the A-pillar 3D model and the free motion headform (*middle*), a cross-section blueprint for A-pillar design (*right*). In the cross-section blueprint for A-pillar design, the given structures are drawn with dark black lines and the target design parts are drawn with light black lines (e.g., airbag and interior shell).

a desired impact speed with ensuring the impact angle between the FMH and the target location. The impact test measures the value of the Head Injury Criterion (HIC), which is used to measure head injury potential in car crash test. There are various target locations required by FMVSS No.201U on one automotive product. According to FMVSS No.201U, the HIC value measured under specific impact speed (FMH Impact Speed — 24 km/h, Reduced FMH Impact Speed — 19 km/h for vehicles equipped with dynamically deployed upper interior head protection system) shall not exceed 1 000.

While developing a new car model, the designed A-pillar (a car component shown in the *left side* of Figure 6.2) needs to meet the requirement of FMVSS No.201U, i.e., the measured HIC value shall be lower than 1 000. Traditionally, engineer conduct the construction-simulation loop, explained in Figure 6.1a in Section 6.1. Notably, the design data of A-pillar comes often in the form of 2D engineering blueprint, i.e., cross sections from the 3D model, which will be detailed described in next section. According to the given geometric constrains, i.e., structural design from exterior parts that are less relevant to the measured HIC value (shown in the *right side* of Figure 6.2 with dark black lines), engineers are tasked to design the target parts of the A-pillar that are the key structure deciding the measured HIC value (shown in the *right side* of Figure 6.2 with light black lines) and optimize the design until it satisfies the HIC-value requirement in a crush simulation based on the expertise of the designers. However, it is not trial, how to modify the structure so that it performs better in the HIC value while following the geometric constrains (e.g., the given structures).

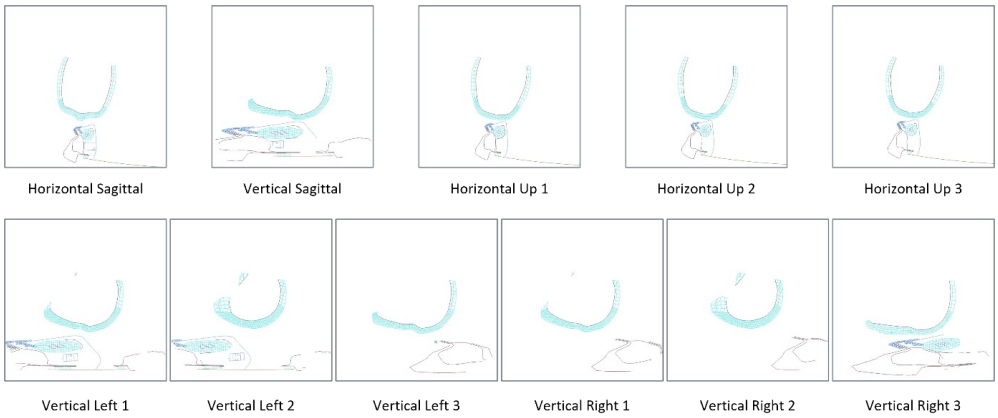


Figure 6.3: One sample of the cross-sectional blueprints of a vehicle A-pillar 3D model.

To this end, we follow the new designed pipeline, DGM-based Generative Engineering Design, as shown in Figure 6.1b. In this proposed pipeline, DGMs is tasked to generate geometric constrain-satisfied designs while the function requirement is met via gradient optimization. Here, the DGM-Simulation loop is designed to be differentiable, so that the gradient can be calculated and used to guide the DGM to synthesize designs that perform better in the target function. However, it has not yet been fully addressed how to use DGM to generate designs that follow the geometric constrains. We aim to tackle this in this section.

6.2.2 Cross-Section Blueprints

According to FMVSS No.201U [171], the HIC is measured for each target locations. These target locations stand for various automotive interior structures. The cross-sectional images, shown in Figure 4.1, are slices, produced by cutting the structure and Free Motion Headform (FMH) together with 11 various planes. These planes all go through the center of the FMH, and the slices are named after the cutting directions, which are Horizontal Sagittal 1, Horizontal Up 1, Horizontal Up 2, Horizontal Up 3, Vertical Sagittal 1, Vertical Left 1, Vertical Left 2, Vertical Left 3, Vertical Right 1, Vertical Right 2, Vertical Right 3. In the cross-sectional images, one sample is produced by slicing one target location of specific car model, and always contains 11 images, Figure 6.3 shows one sample as instance. In the example the curved shape is the sliced part of FMH, while the rest part is the cross-section of the target automotive interior structure. There exist 12 876 blueprints from 3D A-pillars models. We first

convert these blueprints into grayscale pixel-based images with a resolution of 256 × 256.

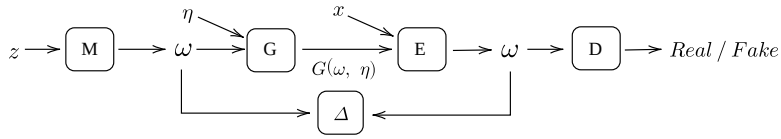
6.2.3 Generation of Cross-Section Designs with GANs

In this work, we first test GANs, dominant the DGM field before diffusion models, as a baseline framework on generating real-world industrial designs. Here, we select Adversarial Latent Autoencoder (ALAE) [126] for its ability of mapping between latent space and image space. We enhance the model robustness and develop an improved version by introducing modern techniques, e.g., spectral normalization (SN) [106], ResNet [55], etc. We further implement the self-attention mechanism [186, 174] to enable the generation of structural patterns. The content of this section has been published in the paper [165].

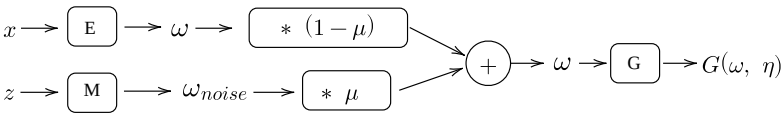
Improved ALAE The vanilla ALAE faces a major issue in its training for the following reasons. First, the ALAE implements the adversarial training strategy in image space, which exposes it to the same training instability commonly encountered by GANs [19, 140, 48]. We implement the following techniques to improve the model robustness and propose a novel version of the ALAE called ALAE$_{improved}$. Firstly, to ensure that generator G receives effective gradients for the optimization, we need to prevent the computed gradients from vanishing or exploding. For this purpose, we apply Spectral Normalization (SN) [106] in encoder E and discriminator D. Spectral Normalization is a powerful method that constrains the Lipschitz continuity of the functions optimized by E and D through layer-wise control of the spectral norm, thereby constraining the scale of the gradients. Spectral Normalization replaces every given weight matrix W by W_{sn} with the formula:

$$W_{\text{sn}} = \frac{W}{\sigma(W)}, \tag{6.1}$$

where $\sigma(W)$ is the greatest singular value of W. Moreover, to stabilize the adversarial training, we implement batch normalization in generator G, allowing for a larger range of learning rates and learning the data distribution more efficiently. Lastly, to address the degradation caused by increasing the depth of the neural network, we modify generator G and encoder E of ALAE to become Residual Networks (ResNet). ResNet features in the so-called residual blocks of convolutional operators contain a "skip connection" that can bypass the error information in back-propagation, hence alleviating the vanishing gradient problem in deeper architectures.



(a) Training



(b) Sampling

Figure 6.4: SA-ALAE in directed graphs. ALAE and SA-ALAE share an identical training strategy. The sampling procedure of SA-ALAE utilizes the off-the-shelf mapper M to sample additional noisy latent variables ω_{noise}, hereby introducing randomness in the sampling.

Self-Attention Adversarial Latent Autoencoder (SA-ALAE) Convolution-based DGMs generate high-quality real-world images. However, DGMs have difficulty in modeling structural objects in engineering design images. The main cause is possibly that the objects in structural images can be sparsely located but are still strongly geometrically connected, e.g., in bicycle design, both wheels should be connected by an axle and lie on the ground. In this simple example, the pixels describing each wheel depend on one another but are far away in the image. However, as a local operator on images, the convolution mechanism cannot capture long-range dependencies, suggesting that using convolutional layers alone is unsuitable for generating engineering design images. Self-Attention Generative Adversarial Network (SAGAN) [197] leverages the self-attention mechanism to gain a global understanding of the input features, which enables the model to capture and model long-range dependencies efficiently. Following the success of SAGAN, we introduce the same mechanism respectively into generator G and encoder E in the ALAE$_{improved}$ framework and propose a new model, Self-Attention Adversarial Latent Autoencoder (SA-ALAE) [165].

To achieve the stochastic generation of a novel image from a source image in the vanilla ALAE, the latent variables ω encoded from the original design are perturbed with sampled noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. We note that this approach compromises the quality of the encoded latent variables, which generates blurry images. To avoid this, SA-ALAE leverages the trained map M to sample random latent variables as a noisy

(a) Source blueprints (A-pillar)

(b) Blueprints reconstructed by SA-ALAE

(c) Blueprints randomly generated by SA-ALAE

Figure 6.5: Examples of source blueprints of A-pillar and blueprints generated by SA-ALAE.

perturbation. The novel sampling procedure of SA-ALAE is displayed in Figure 6.4b. Adding randomly sampled latent variables as an independent noise protects the quality of perturbed latent variables so that the stochastic encoder-decoder inference does not harm the quality of the generated image. The calculation of the latent variables, used to explore the variants of existing designs, is based on the following formula:

$$\boldsymbol{\omega} = (1 - \mu)E(\mathbf{x}) + \mu M(\mathbf{z}), \tag{6.2}$$

where $\mu \in [0, 1]$ is a tunable parameter, $\mathbf{x}$ refers to an original design and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is sampled noise.

Results of SA-ALAE We evaluate the SA-ALAE in both image reconstruction and random generation on the A-pillar dataset. As shown in Figure 6.5, SA-ALAE generates high-quality blueprints corresponding to the input designs and containing recognizable structural details. In addition to the reconstruction function, we evaluate the ability of SA-ALAE in randomly generating A-pillar designs, which yields a

sufficient diversity of novel designs with detailed structures displayed in Figure 6.5. While the results of SA-ALAE is compelling, training of SA-ALAE is still challenging due to the instability of the adversarial training. To bypass this, in next section, we will focus on using diffusion models instead to generate A-pillar designs.

6.2.4 A-pillar Blueprint Autocompletion

As we describe in Section 6.2.1, the engineers responsible for the design of the interior parts of the A-pillar initially receive a predefined structure (i.e., the given parts) and then complete the design by constructing the remaining structural parts (i.e., the target parts), such as the airbag and interior shell, which we drew using light black lines in the right side in Figure 6.2.

(a) A-pillar exterior structures (the given parts)

(b) DGM-completed A-pillar designs

(c) Comparison between engineer- and DGM-completed A-pillar designs

Figure 6.6: Example of automatic blueprint completion. In (c), we have overlaid the engineer's completed design with the DGM's completed design to show that our DGM performs a conditional generative design process, the results of which follow the given geometric constraints well but is clearly different from the blueprint designed by engineers.

The introduction of inpainting technique (i.e., RePaint [97]) with PoDM (developed in Chapter 2) on bicycle designs has shown compelling results in Figure 2.14b. Consid-

ering that the completion of 2D A-pillar cross-section blueprints is a similar scenario, we test the PoDM-driven inpainting in designing A-pillar cross-section blueprints in this section. We train a PoDM on the cross-section images using the configurations defined in Section 2.3.1. During model training, we use FDD (developed in Chapter 3) to select the best-performing model. Afterwards, we implement the inpainting pipeline (explained in Section 2.4) on the given structure of exterior A-pillar design, shown in Figure 6.6a. The DGM-completed design structures are displayed in Figure 6.6b, while we also show the engineer-design A-pillar blueprints in Figure 6.6c. This shows that our PoDM does not just copy the designs of engineers, but explores new solutions for a given structure — a sign of good generative engineering design, the results of which are still feasible and usable and has been approved by engineers in expert reviews.

6.3 Vehicle Rim Design

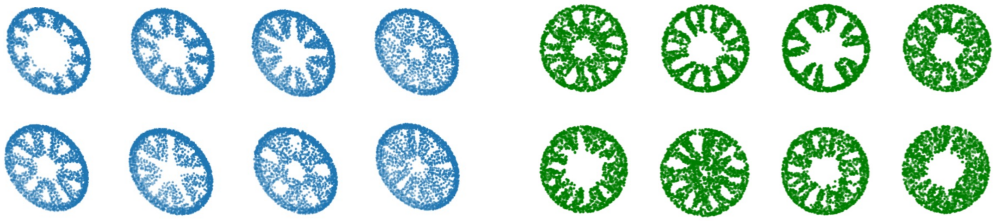
The design of car rims is a crucial aspect of the overall vehicle design process, as it not only contributes to the aesthetic appeal of the car but also plays a significant role in its aerodynamic performance. The rim design data in the car industry is closely tied to the mesh form, for its compatibility with aerodynamic simulations.

The shape and design of the car's rims have a significant impact on the overall aerodynamic performance of the vehicle. The rim's design can affect the airflow around the wheels, which in turn can influence the car's fuel efficiency, stability, and handling characteristics [146].

Designers and engineers in the car industry use Computational Fluid Dynamics (CFD) simulations to analyze the airflow patterns around the rims and the rest of the vehicle. By analyzing the mesh data form of the rims, they can identify areas of high pressure, turbulence, and drag, and then refine the design to optimize the aerodynamic performance. We show a basic pressure measurement example from [13] that demonstrates how the geometric design of the wheel rim affects the pressure distribution in the car.

To this end, we claim that DGM-based engineering can accelerate the rim design process and has great potential, but there is a major challenge: DGMs used for 3D shapes recently have focused on point clouds, generating results that are neither recognizable to engineers nor usable for simulation. We tried latent Wasserstein GAN (LWGAN) [131] and point-voxel diffusion (PVD) [201], details can be found in Section 1.4, and demonstrate the generated results in Figure 6.7. As the example shows,

the resulting point clouds are of poor quality and do not show the clear structures that engineers can construct after. In addition, we have consulted with engineers, and they suggest that even the source point clouds (Figure 6.8c) cannot represent the design in the simulation and give accurate results.



(a) Rims generated with LWGAN [131] **(b)** Rims generated with PVD [201]

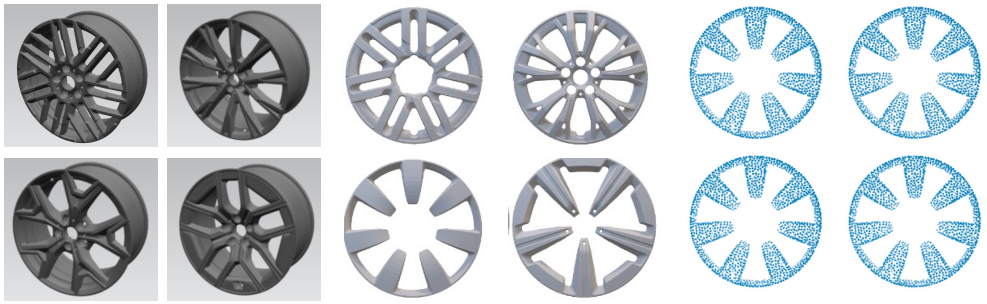
Figure 6.7: Rims generated by state-of-the-art point-based DGMs.

6.3.1 Rim Design Data

For the study in this section, we have 497 samples of BMW rims, which are designed in real engineering processes with multiple objectives, e.g., outstanding aerodynamic performance, low cost and excellent appearance. This dataset consists of 497 high-quality 3D models of rims, providing a representative example of industrial 3D surfaces that demand precision and efficiency. Source data is the in the form of STEP files, as we display in Figure 6.8a, where the design relevant part is only the front part of the whole rim. Thus, we trim the source rim data, so that the design can be represented with a much reduced model size, as shown in Figure 6.8b which are in the form of mesh. In developing DGM-based models for rim design, we also convert the rims to the form of point cloud (see Figure 6.8c), so that we can test state-of-the-art point-based generative models.

6.3.2 Design Inspiration

In this section, we implement the method (SpoDify) we developed in Chapter 4 for design inspiration of industrial rims. Note that, the advantage of our method is not only the compelling generation of realistic mesh shapes, but also the efficiency in reducing the dimension of source data (a rim mesh can have 15k vertices, and using SpoDify can reduce each rim mesh to 64 values) and saving computational power. In Figure 6.9b, we display the generated rims and the source rims for visual qualitative

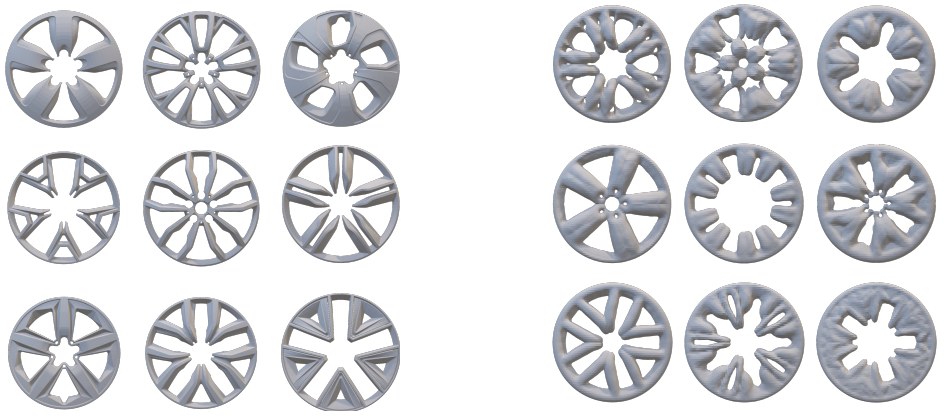


(a) Source rims without trimming **(b)** Source rims after trimming **(c)** Source trimmed rims represented in point clouds

Figure 6.8: Examples of BMW rims in various form, i.e., B-Rep, mesh, and point cloud.

comparison. In Figure 6.10, we additionally display the generated rims and their closest samples from source data to show that our method is able to generate novel shapes.

Here, the generated rims maintain structural integrity and meaningful design elements, showcasing that the model is able to synthesize good quality rim designs by training on source data. Additionally, we observe many novel rim designs that do not come from the source dataset, which can help exploration in the design process. In conclusion, this capacity to interpolate and generate novel shapes inspired by the original dataset is a promising indicator of the model's robustness and utility.



(a) Source rims **(b)** Rims generated by SpoDify

Figure 6.9: Comparison of generated rim data: (a) source data, (b) rims generated by our approach.

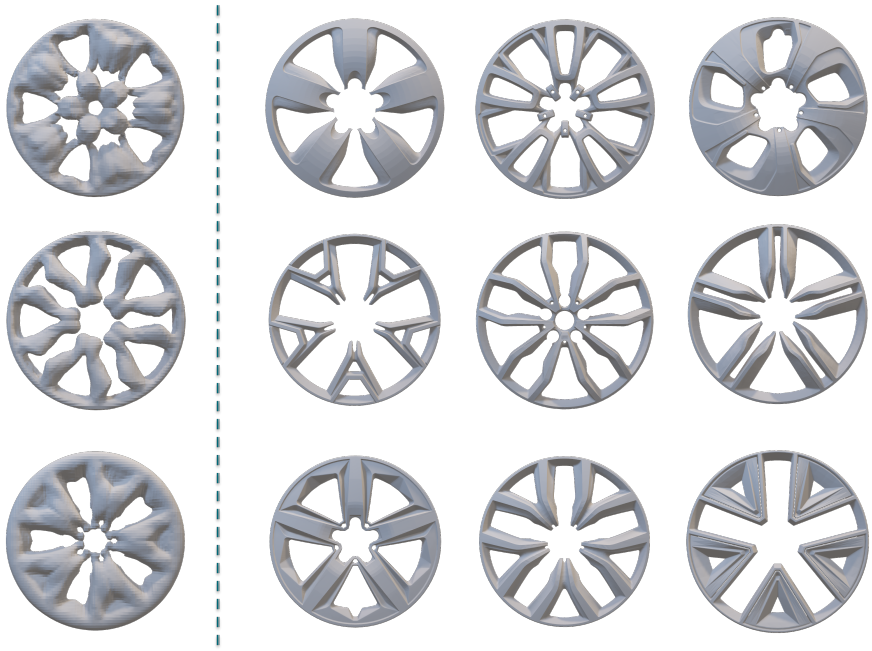


Figure 6.10: Generated rims (*left*) and their top 3 nearest meshes (*right*) from the source data. Determining the nearest mesh leverages Chamfer distance as the metric.

The idea behind applying spectral decomposition, as previously motivated, is to represent an object as a combination of basic shapes in the spectral space. These basic shapes act as the primary features that capture and assemble the structure of the dataset, grouping and describing it in terms of these key characteristics. In Figure 6.11, we visualize the right eigenvector V^T matrix. This is achieved by taking an eigenvector, applying the inverse discrete wavelet transform (IDWT), and then using marching cubes to map it back to the mesh domain. Here we highlight the data-dependent nature of these basic shapes. The rim dataset shows clear and distinct patterns in the basic shapes, making it easier to interpret the meaning of each eigenvector. This visualization process provided valuable insights into the underlying structures of the datasets.

6.4 Conclusion

In practical industrial design processes, designs are more complex and higher-dimensional, which makes training and reasoning difficult. Our work introduces two scenarios, de-

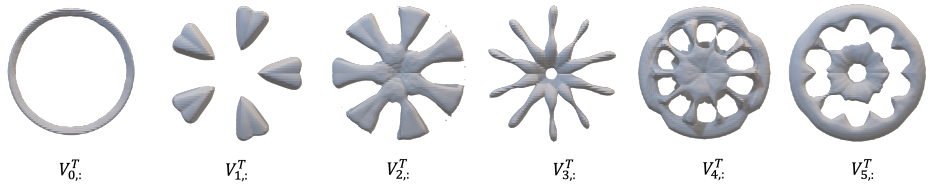


Figure 6.11: Visualization of the spectral features. The spectral bases present in the $V^\top$ matrix resulted from the SVD.

sign completion and design inspiration, to demonstrate how the Design Guidance Model (DGM) can assist engineers in developing complex practical designs within a carefully designed framework. First, completing a preliminary design based on given geometric requirements is a common task that requires some knowledge of the target design component and can be quite time-consuming. In situations where a large amount of historical design data is available, our work in this section demonstrates how DGM can facilitate the design process in a reliable manner. Secondly, generating designs that do not exist for design inspiration is a very valuable function for design teams. We apply our SpoDify developed in Chapter 4 to the wheel rim and obtained compelling designs that are not existing in real wheel rim data.