A VISUAL REPRESENTATION OF ENGINEERING CATALOGS USING VARIATIONAL AUTOENCODERS

Saketh Sridhara\textsuperscript{1}, Krishnan Suresh\textsuperscript{1,*}

\textsuperscript{1}Department of Mechanical Engineering
University of Wisconsin-Madison, Madison, WI, United States

ABSTRACT

Catalogs have been used for over a century for designing engineering systems. While catalogs are excellent repositories of engineering information, they are difficult to navigate, specifically to spot clusters, gaps, substitutes, and outliers. Inspired by Ashby charts for material selection, we propose here, a visual representation of engineering catalogs using neural networks. In particular, we employ variational autoencoders (VAEs) to project catalog data onto a lower-dimensional latent space. The latent space can then be visualized to explore the underlying structure of the catalog. Specifically, creators can use this visual representation to identify gaps and outliers in their data, while end users can benefit from this representation to compare catalogs from competitors, and to find substitutes. Contours can be superimposed on the charts to enable selection based on user-defined attributes; these contours are generalization of design indices associated with Ashby charts. Various examples of catalogs across engineering disciplines, ranging from materials and bearings to motors and batteries are illustrated using the proposed method. Using these examples, we (1) study the impact of the latent space dimension on the representational error, (2) illustrate how designers can easily choose alternate configurations based on their design requirements, and (3) gaps in catalog offerings can be clearly identified, providing a stimulus for new product development.

Keywords: Catalogs, representation, visualization, neural networks, latent space

1. INTRODUCTION

Representations form the backbone of engineering advances. For example, geometric representations such as constructive solid geometry and boundary representation \cite{1} have led to breakthroughs in computer-aided design (CAD), computer-aided analysis (CAE), and computer-aided manufacturing (CAM). Yet, we rely on a century-old representation of engineering data, namely, catalogs for designing almost any engineering system; see Figure 1. As engineering systems become more complex, this approach has significant limitations that pose challenges for catalog creators and end users.

From the creators’ perspective, it is difficult to identify outliers and gaps within these catalogs:

- An outlier is an entry that deviates markedly from other entries in the catalog \cite{4}. Identification of outliers is crucial for two reasons. First, if the data is erroneous, it must be removed or corrected. Second, if they truly reflect exceptional characteristics, additional validation may be required.

- Gaps in the catalog are regions for innovation and provide a

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Engineering design involves the simultaneous task of selecting components from design catalogs and designing custom geometries.\cite{2, 3}.}
\end{figure}
stimulus for new product development. Unfortunately, both outliers and gaps are difficult to identify from tabular data. From the catalog user’s perspective:

- Tabular data is ill-suited for finding substitutes [5]. Identification of clusters and nearby designs [6] can be immensely valuable as it enables reuse.
- Similarly, it is not easy to determine from a catalog if two different types of, say, ball bearings have similar attributes.
- Finally, it is difficult to compare catalogs from competitors since the tabular schemes can be vastly different.

1.1 Contributions
To address the aforementioned challenges, we propose here a visual representation of engineering catalogs that would enable:
(1) catalog creators to rapidly screen for outliers, identify gaps and new form factors for product development, and (2) end-users to gain structural insight into the data, make well-informed design decisions [7] and choose from alternate design configurations.

2. PROPOSED METHOD
The proposed framework draws inspiration from the pioneering work of Dr. Ashby on the visualization of material catalogs [8, 9].

2.1 Ashby Charts for Materials
Materials dictate the performance of all engineered structures and devices. There are well over 50,000 engineering materials [8], and this number is constantly growing. A typical subset of a material catalog is shown in Table 1, with material properties (attributes) such as Young’s modulus, cost, mass density, and yield strength.

<table>
<thead>
<tr>
<th>Mat Name</th>
<th>Class</th>
<th>$E$ [10^3 Pa]</th>
<th>$C$ [S/kg]</th>
<th>$\rho$ [kg/m$^3$]</th>
<th>$Y$ [10^7 Pa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A286</td>
<td>Steel</td>
<td>201</td>
<td>5.18</td>
<td>7920</td>
<td>62.0</td>
</tr>
<tr>
<td>AISI 304</td>
<td>Steel</td>
<td>190</td>
<td>2.40</td>
<td>8000</td>
<td>51.7</td>
</tr>
<tr>
<td>Gray C</td>
<td>Steel</td>
<td>66.2</td>
<td>0.65</td>
<td>7200</td>
<td>15.2</td>
</tr>
<tr>
<td>3003-H16</td>
<td>Al</td>
<td>69.0</td>
<td>2.18</td>
<td>2730</td>
<td>18.0</td>
</tr>
<tr>
<td>5052-O</td>
<td>Al</td>
<td>70.0</td>
<td>2.23</td>
<td>2680</td>
<td>19.5</td>
</tr>
<tr>
<td>7050-T7651</td>
<td>Al</td>
<td>72.0</td>
<td>2.33</td>
<td>2830</td>
<td>55.0</td>
</tr>
<tr>
<td>Acrylic</td>
<td>Plastic</td>
<td>3.00</td>
<td>2.80</td>
<td>1200</td>
<td>7.3</td>
</tr>
<tr>
<td>ABS</td>
<td>Plastic</td>
<td>2.00</td>
<td>2.91</td>
<td>1020</td>
<td>3.0</td>
</tr>
<tr>
<td>PE HD</td>
<td>Plastic</td>
<td>1.07</td>
<td>2.21</td>
<td>952</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Observe that the attributes are poorly correlated, i.e., a higher Young’s modulus does not imply a higher cost, and the values can vary by several orders of magnitude. More importantly, selecting a suitable material for a particular application can be challenging since material selection is tightly coupled with other design decisions. To address these challenges, Ashby’s charts were proposed in [8, 9]. These charts represent two properties at a time on a two-dimensional plot. For example, Ashby’s chart for Young’s modulus versus density is illustrated in Figure 2. Observe the natural clustering of materials that the Ashby chart reveals. Furthermore, under certain conditions, design indices can be associated with these charts to help select suitable materials for a particular application [10]. Ashby charts effectively address the limitations of catalogs described earlier for up to three attributes, and have been extended to hybrid materials and composites [11].

2.2 Variational Autoencoders
Drawing inspiration from such Ashby charts, the proposed framework relies on variational autoencoders (VAEs) to create charts of arbitrary catalogs, with any number of attributes through nonlinear dimensional reduction. Dimensional reduction techniques aim to decrease the number of dimensions in high-dimensional data and convert it to a lower-dimensional space, often 2D or 3D, in order to improve human understanding. Data visualization provides an intuitive understanding of the underlying structure, in order to gain insight and to develop hypotheses. It has been shown that dimensional reduction can help highlight biases and pervasive noise in the data [12]. Several researchers have leveraged dimensionality reduction methods for visualization in engineering design: examples include multidimensional scaling [13, 14], self-organizing maps [15, 16], design manifolds through kernel PCA [17, 18] to name a few. The visual representation in [14] does not guarantee an accurate approximation of the design space and is susceptible to crowding. Self-organizing maps are limited in their ability to capture complex data distributions and the projection to a 2D planar grid allows only a restricted number of different distances to be represented [19]. The performance of Kernel PCA methods is highly dependent on the choice of the kernel function [20], thereby requiring multiple trials and experiments.
VAEs are special neural network configurations with wide applicability in dimensional reduction, data compression, semi-supervised learning, and data interpolation; for a detailed review please see [21]. VAEs have been employed to synthesize new samples similar to the input as well as anomaly detection; popular examples include the generation of synthetic human faces [22] and outlier detection [23, 24]. In science and engineering, VAEs have been used to generate new designs by exposing them to large databases containing images of micro-structures [25], photonic crystals [26], heat-conduction materials [27], discovering partial differential equations [28], and synthesizing new chemical compounds [29].

In this work, we do not use VAEs to synthesize new data, instead we leverage the VAE’s ability to map uncorrelated data onto an abstract, visualizable latent space. To understand the construction of the VAE, let us consider a material catalog, such as the one in Table 1. A simplified VAE architecture for this data is illustrated in Figure 3. In this instance, it consists of:

1. A four-dimensional input module corresponding to the four properties in Table 1, namely the Young’s modulus \(E\), cost \(C\), mass density \(\rho\) and yield strength \(Y\). The input set is denoted by \(\zeta\).
2. An encoder \(\mathcal{E}\) consisting of a fully-connected network of, say, 250 neurons, associated with activation functions and weights [30].
3. A two-dimensional latent space, denoted by \(z_0, z_1\) that lies at the heart of the VAE.
4. A decoder \(\mathcal{D}\), that is similar to the encoder, consisting of a fully-connected network.
5. A four-dimensional output denoted by \(\hat{\zeta}\) corresponding to the same four properties.

The VAE’s primary task is to match the output to the input as closely as possible via an intermediate low-dimensional latent space. This is done through an optimization process (also referred to as training), using the weights associated with the encoder and decoder as optimization parameters. In other words, the VAE minimizes the reconstruction loss \(||\zeta - \hat{\zeta}||\). Additionally, a regularization loss is included to ensure that latent space is spatially compact [31]. Thus, VAEs are a generalized form of principal component analysis (PCA) for dimensional reduction, where the nonlinearity of the neural networks allows for far greater flexibility [21, 32] and higher accuracy [33] when compared to classical nonlinear dimension reduction methods such as Isomap [34].

### 2.3 Training procedure

The training procedure is described in Algorithm 1, with the input data being the material database. Encoder \(\mathcal{E}\) is a neural network that takes in the set of material data \(\zeta\) and encodes the four-dimensional data to the two-dimensional latent space denoted by \(z_0, z_1\) through a probabilistic latent distribution governed by \(\mu, \sigma\). The VAE loss (or error) is then used to drive the training per Equation 1 until a sufficiently high representational accuracy is attained. The first term represents the reconstruction loss \(||\zeta - \hat{\zeta}||\), i.e., the difference between the input and reconstructed data. The second term represents the KL divergence loss that is imposed to ensure that latent space is compact and resembles a standard Gaussian distribution \(z \sim N(\mu = 0, \sigma = 1)\) [21]; the two losses are combined with a weight factor \(\beta\). After the training, the encoder is discarded, and the decoder \(\mathcal{D}\) is retained for data retrieval. The decoder takes the latent space coordinates as input and returns the predicted material properties.

\[
L = ||\zeta - \hat{\zeta}|| + \beta KL(z||N) \tag{1}
\]

**Algorithm 1 ENCODE MATERIALS**

1: procedure MAT\(\text{ENCODE} (\zeta)\) \(\triangleright\) Input: Training data
2: \hspace{1cm} \text{epoch} = 0 \(\triangleright\) iteration counter
3: \hspace{1cm} repeat \(\triangleright\) VAE training
4: \hspace{2cm} \(\mathcal{E} (\zeta) \rightarrow \{\mu, \sigma\}\) \(\triangleright\) Forward prop. encoder
5: \hspace{2cm} \{\mu, \sigma\} \rightarrow z \(\triangleright\) Reparameterization [21]
6: \hspace{2cm} \{\mu, \sigma\} \rightarrow KL(z||N) \(\triangleright\) KL loss
7: \hspace{2cm} \mathcal{D} (z) \rightarrow \hat{\zeta} \(\triangleright\) Forward prop. decoder
8: \hspace{2cm} \{\zeta, \hat{\zeta}, KL\} \rightarrow L \(\triangleright\) VAE Loss
9: \hspace{2cm} \(w + \Delta w(\nabla L) \rightarrow w\) \(\triangleright\) Update VAE wts: Adam [35]
10: \hspace{1cm} epoch ++
11: \hspace{1cm} until error is acceptable \(\triangleright\) Iterate
12: \hspace{1cm} return \(\mathcal{D}\) \(\triangleright\) Trained decoder
13: end procedure

### 3. VISUAL REPRESENTATION OF MATERIALS

In this work, we used PyTorch [36] to model the VAE, and the gradient-based Adam optimizer [35] to minimize Equation 1 with recommended values of learning rate 0.002 for 20,000 epochs, and \(\beta = 5 \times 10^{-5}\). A threshold for total loss may also be specified for termination of training.
3.1 Material latent space

The material data consisting of 92 materials was sourced from SolidWorks material library [37]. The VAE training took approximately 22 seconds on a 6-core Intel i7 CPU with 32GB RAM running Ubuntu. The resulting two-dimensional latent space is illustrated in Figure 4. Observe that each material can now be represented as a pair \( z_0, z_1 \). For example, annealed AISI 1020 is represented by the pair \((-0.6, 2.1)\) while Acrylic is represented by \((0.6, -0.2)\). The VAE clusters similar materials together in the latent space, as can be observed in Figure 4. Conceptually, the latent space in Figure 4 is similar to the popular Ashby charts seen in Figure 2.

3.2 Convergence

The convergence of the total loss, reconstruction loss and regularization loss are illustrated in Figure 5.

3.3 Property contours

One can overlay the latent space in Figure 4 with specific material properties (or their combination) to gain further insight. This is illustrated in Figure 6 where the contour plots of cost are superimposed on the latent space. This allows designers to identify the class of materials that meet one or more requirements, as seen in Figure 7, where materials whose cost per kg lies between $2 and $5 are identified.

3.4 Outlier Detection

VAEs have been increasingly used in recent years for outlier detection, a crucial task in anomaly detection, fraud detection, and other applications where detecting rare or unexpected events is important. We exploit the same for engineering data. For example, in the material database that we sourced from [37], we identify a specific outlier, namely, Aluminum 4032-T6, as indicated in Figure 8; it is far away from other Aluminum members, and is closer to steels. Upon further inspection, we identified that this is due to an unusually low cost associated with Aluminum 4032-T6. Catalog creators can leverage this framework to validate their data.

3.5 Representational Accuracy

As with any dimensional reduction method, it is expected that the values obtained through the decoder will differ slightly from the input data [13]. For example, one can compute the properties of Acrylic predicted by the decoder, and compare it against the true values. Table 2 summarizes the errors for each attribute for some of the materials. The following observations are worth noting:

1. Despite the lack of correlation between material properties, and two orders of magnitude difference in values, the VAE captures the entire database of 92 materials reasonably well, using a simple two-dimensional latent space.

---

1 Prof. Morlier’s team at ISAE-SUPAERO (https://ica.cnrs.fr/author/jmorlier/) brought this to our attention.
2. The error can be further reduced by either increasing the dimension of the latent space or by tuning the size and optimization parameters of the VAE. One such case is explored in Section 4.

### TABLE 2: PERCENTAGE ERROR BETWEEN ACTUAL AND DECODED DATA.

<table>
<thead>
<tr>
<th>Material</th>
<th>ΔE %</th>
<th>ΔC %</th>
<th>Δρ %</th>
<th>ΔY %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A286 Iron</td>
<td>1.0</td>
<td>0.6</td>
<td>2.3</td>
<td>1.6</td>
</tr>
<tr>
<td>ABS</td>
<td>1.7</td>
<td>2.2</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>AISI 304</td>
<td>3.3</td>
<td>0.7</td>
<td>0.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Gray Cast Fe</td>
<td>1.3</td>
<td>2.7</td>
<td>2.1</td>
<td>2.9</td>
</tr>
<tr>
<td>3003-H16</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>2.3</td>
</tr>
<tr>
<td>5052-O</td>
<td>0.8</td>
<td>2.4</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>7050-T7651</td>
<td>0.2</td>
<td>2.3</td>
<td>0.4</td>
<td>4.0</td>
</tr>
<tr>
<td>Acrylic</td>
<td>1.6</td>
<td>0.5</td>
<td>0.1</td>
<td>1.4</td>
</tr>
<tr>
<td>PE HD</td>
<td>3.2</td>
<td>0.3</td>
<td>2.0</td>
<td>3.9</td>
</tr>
<tr>
<td>Max error</td>
<td>6.1</td>
<td>5.2</td>
<td>3.2</td>
<td>9.0</td>
</tr>
</tbody>
</table>

### 4. APPLICATIONS AND CASE-STUDIES

The proposed method generalizes to a variety of catalogs. We use several examples below to illustrate potential applications of the proposed framework.

#### 4.1 Cross sections: Clustering on a line

Engineering structures are often built using beams with predefined cross-sections, such as rectangles and I-shapes. Consider a typical catalog of cross sections [38], such as the one in Table 3.

The visual representation of this catalog is shown in Figure 9. Observe that the disks are clustered onto a line in the 2D space.

![Figure 9](image)

#### 4.2 Bearings: Higher dimensional latent spaces

Next we consider a catalog of bearings from Timken [39]. Bearings are widely used in engineering applications, ranging from machinery and vehicles to aerospace and medical equipment. There are different types of bearings, including ball bearings, roller bearings, thrust bearings etc., each suitable for a certain rated speed and load - a subset of the database is shown in Figure 10. We choose six attributes for each catalog entry.
namely bore diameter (mm), outer diameter (mm), bearing width (mm), rated RPM, rated dynamic load (kN), and weight (kg).

The visual representation of this catalog with 100 different bearings is seen in Figure 11, overlaid with contours indicating the RPM rating of the bearings.

Next, we changed the latent dimension to three; the resulting latent space is illustrated in Figure 12. As expected, the features of the 2D representation extend to 3D as well in terms of clusters, outliers, and gaps.

For the same training parameters, we compare the representational accuracy of 2D and 3D latent spaces. We report the maximum and average reconstruction errors in Table 4. Observe that the 3D latent space has a lower average reconstruction error across every attribute.

**TABLE 4: MAX AND AVG ERROR PERCENTAGES IN 2D AND 3D LATENT SPACES FOR BEARINGS.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2D Latent</th>
<th>3D Latent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max(%) Error</td>
<td>Avg(%) Error</td>
</tr>
<tr>
<td>Bore Dia</td>
<td>2.8</td>
<td>0.78</td>
</tr>
<tr>
<td>Outer Dia</td>
<td>2.4</td>
<td>0.63</td>
</tr>
<tr>
<td>Width</td>
<td>3.4</td>
<td>0.55</td>
</tr>
<tr>
<td>Load Rating</td>
<td>9.1</td>
<td>2.37</td>
</tr>
<tr>
<td>RPM</td>
<td>6.0</td>
<td>2.12</td>
</tr>
<tr>
<td>Weight</td>
<td>9.6</td>
<td>6.5</td>
</tr>
</tbody>
</table>

4.3 Motors: Selecting alternate configurations

We next consider a catalog of motors from GE [40]. GE produces a variety of motors for industrial and commercial applications, with different configurations to suit various needs of torque and speeds. We consider 4 attributes: horsepower, RPM, operating voltage, and cost for each entry across 4 types of motors: horizontal shaft, vertical shaft, keyless shaft, and DC motors; see Figure 13, for a total of 168 motors.

The visual representation of this catalog is shown in Figure 14. The overlaps indicate multiple configurations are available with similar performance characteristics. For example, based on horsepower requirements, the design teams can choose from either a horizontal or a vertical shaft motor for their product, or switch to a DC power source from AC. The visual latent space provides insights into possible substitutes or alternate design choices for the end user.
4.4 Batteries: Identifying new form factors

Lastly, we consider a catalog of 44 batteries from FDK [41]. There are several types of batteries available today, each with its own unique set of properties and characteristics. Some of the most common types of batteries include Alkaline, Lithium-ion, Nickel-metal hydride etc., each with different energy densities, operating voltages, and capacity. A sample of the catalog is shown in Figure 15. For visualization, we consider six classes of batteries and four attributes: voltage (V), current (mA), capacity (mAh) and weight (g).

The visual representation of batteries is shown in Figure 16, overlaid with contours of battery capacity (mAh). Upon generating contours of capacity in the latent space, the catalog creators can identify gaps to place newer offerings, and contrast their existing gaps with the competitors’ catalogs to develop new products that meet the needs of customers.
5. CONCLUSION

The proposed visual representation of catalogs addresses challenges faced by creators in identifying gaps and outliers in their data. End users can benefit from this representation in identifying clusters for substitutes, reuse, and configuration changes. While the method has shown success in representing a variety of catalogs, a few challenges remain. The first is to improve the representational accuracy and reduce reconstruction error by incorporating recent advancements in VAEs [42, 43]. Secondly, the framework needs to be validated on larger catalogs consisting of thousands of entries such as the material database [44, 45]. Third, the abstract latent space variables, which are complex analytical functions of the attributes, currently lack interpretability [46]: this can be a challenge to practicing engineers. Finally, the method could benefit from better tuning of the networks and optimal choice of the VAE architecture [47, 48].

The latent space offers other advantages that are being explored. The latent space of the VAE is differentiable - this would allow seamless integration of catalog data into gradient-based design optimization problems [49, 50], leading to optimally designed systems. Further, the generative capability of the latent space can be leveraged for innovation [29, 51, 52], and will be explored in future work. Finally, another interesting research direction is incorporating data uncertainties within the latent space.

ACKNOWLEDGMENTS

The authors would like to thank the support of the National Science Foundation through grant CMMI 1561899, and the U. S. Office of Naval Research under PANTHER award number N00014-21-1-2916 through Dr. Timothy Bentley.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES


[42] Higgins, Irina, Matthey, Loic, Pal, Arka, Burgess, Christoper, Glorot, Xavier, Botvinick, Matthew, Mohamed,


