Proceedings of the ASME 2023
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2023
August 20-23, 2023, Boston, Massachusetts

DETC2023-115029

A VISUAL REPRESENTATION OF ENGINEERING CATALOGS USING VARIATIONAL AUTOENCODERS

Saketh Sridhara[1], Krishnan Suresh[1,*]

[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 [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.

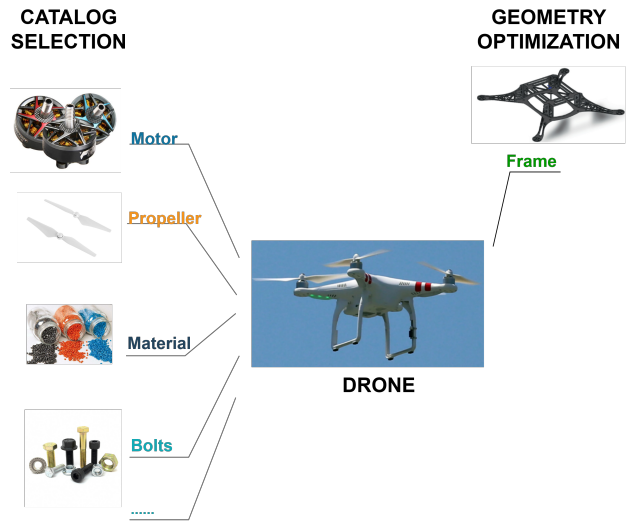


FIGURE 1: ENGINEERING DESIGN INVOLVES THE SIMULTANEOUS TASK OF SELECTING COMPONENTS FROM DESIGN CATALOGS AND DESIGNING CUSTOM GEOMETRIES.[2, 3].

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 [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

*Corresponding author: ksuresh@wisc.edu

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 1: A CURATED SUBSET OF MATERIALS AND THEIR PROPERTIES.

Mat Name	Class	E $[10^9 Pa]$	C $[\$/kg]$	ρ $[kg/m^3]$	Y $[10^7 Pa]$
A286 Iron	Steel	201	5.18	7920	62.0
AISI 304	Steel	190	2.40	8000	51.7
Gray C Iron	Steel	66.2	0.65	7200	15.2
3003-H16	Al	69.0	2.18	2730	18.0
5052-O	Al	70.0	2.23	2680	19.5
7050-T7651	Al	72.0	2.33	2830	55.0
Acrylic	Plastic	3.00	2.80	1200	7.3
ABS	Plastic	2.00	2.91	1020	3.0
PE HD	Plastic	1.07	2.21	952	2.2

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].

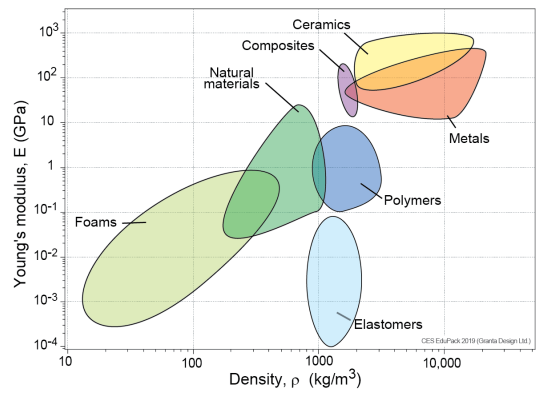


FIGURE 2: ASHBY'S CHARTS FOR YOUNG'S MODULUS VS DENSITY; CHART CREATED USING CES EDUPACK 2019, ANSYS GRANTA © 2020 GRANTA DESIGN.

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 (ρ) and yield strength (Y). The input set is denoted by $\boldsymbol{\zeta}$.
2. An *encoder* $\mathscr{E}$ 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* $\mathscr{D}$, that is similar to the encoder, consisting of a fully-connected network.
5. A four-dimensional *output* denoted by $\hat{\boldsymbol{\zeta}}$ corresponding to the same four properties.

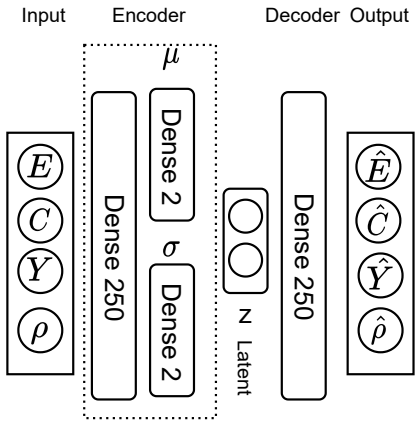


FIGURE 3: ARCHITECTURE OF A VARIATIONAL AUTOENCODER.

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* $||\boldsymbol{\zeta} - \hat{\boldsymbol{\zeta}}||$. 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 nonlinear nature 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 Isomaps [34].

2.3 Training procedure

The training procedure is described in Algorithm 1, with the input data being the material database. Encoder $\mathscr{E}$ is a neural network that takes in the set of material data ζ 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 μ, σ. 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 $||\boldsymbol{\zeta} - \hat{\boldsymbol{\zeta}}||$, 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 $z \sim \mathcal{N}(\mu = 0, \sigma = 1)$ [21]; the two losses are combined with a weight factor β. After the training, the encoder is discarded, and the decoder $\mathscr{D}$ is retained for data retrieval. The decoder takes the latent space coordinates as input and returns the predicted material properties.

$$L = ||\boldsymbol{\zeta} - \hat{\boldsymbol{\zeta}}|| + \beta\, KL(z||\mathcal{N}) \tag{1}$$

Algorithm 1 ENCODE MATERIALS

```
procedure MatEncode(ζ)                          ▹ Input: Training data
    epoch = 0                                   ▹ iteration counter
    repeat                                      ▹ VAE training
        ℰ(ζ) → {μ, σ}                           ▹ Forward prop. encoder
        {μ, σ} → z                              ▹ Reparameterization [21]
        {μ, σ} → KL(z||𝒩)                       ▹ KL loss
        𝒟(z) → ζ̂                                ▹ Forward prop. decoder
        {ζ, ζ̂, KL} → L                          ▹ VAE Loss
        w + Δw(∇L) → w                          ▹ Update VAE wts: Adam[35]
        epoch + +
    until error is acceptable                   ▹ Iterate
    return 𝒟                                    ▹ Trained decoder
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 $\beta = 5 * 10^{-5}$. 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.

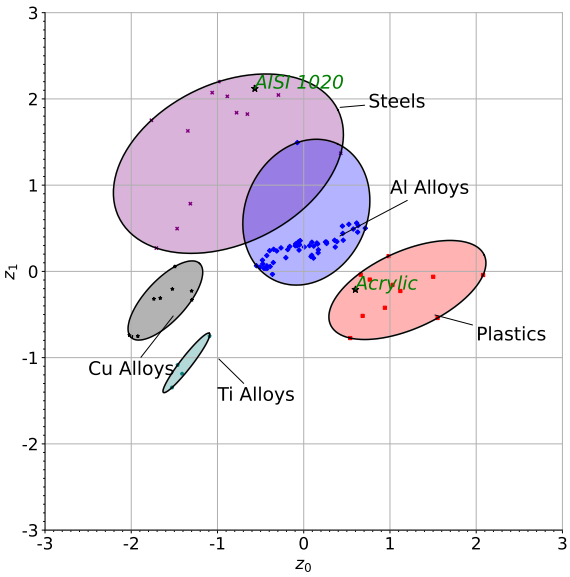


FIGURE 4: MATERIALS REPRESENTED ON A 2D LATENT SPACE

3.2 Convergence

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

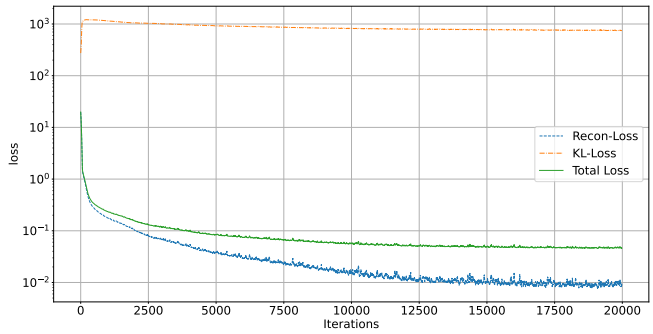


FIGURE 5: CONVERGENCE OF THE LOSS DURING TRAINING

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.

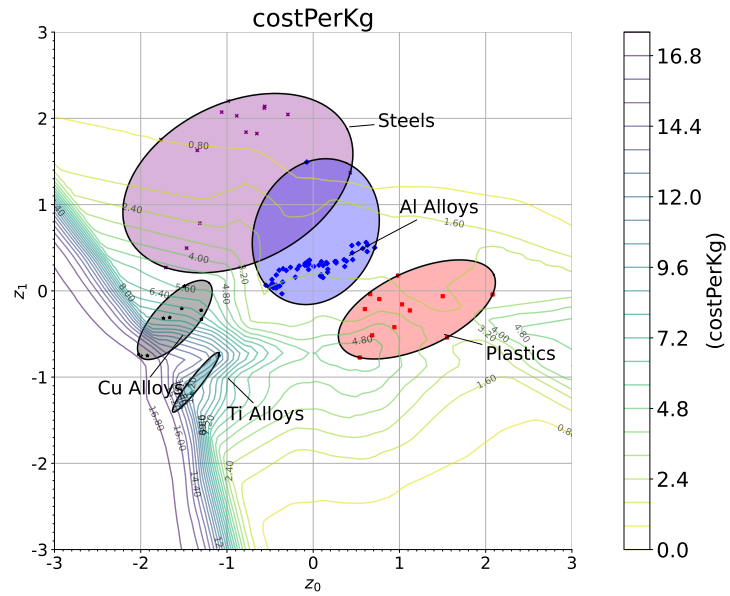


FIGURE 6: COST PER KG CONTOUR PLOT SUPERIMPOSED ON THE LATENT SPACE OF MATERIALS

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 [1]. 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.

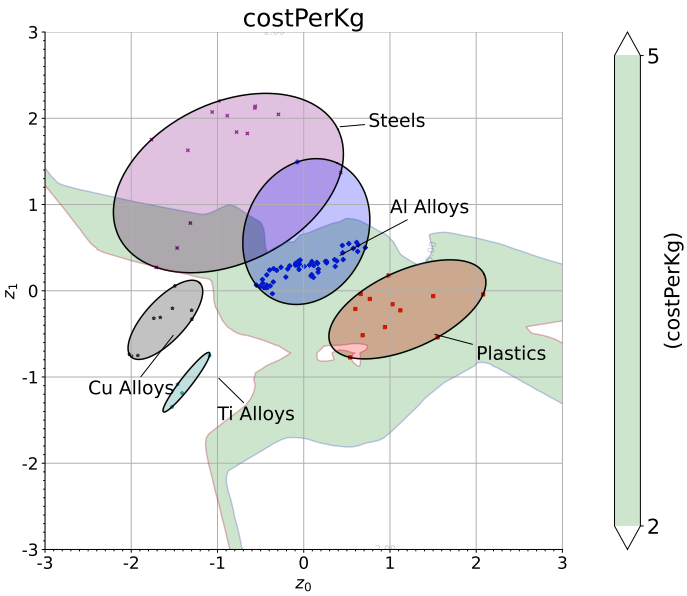


FIGURE 7: FEASIBLE SPACE IS MARKED IN GREEN, CHOICES LYING IN THIS SPACE MEET DESIGN REQUIREMENT

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.

Material	ΔE %	ΔC %	$\Delta \rho$ %	ΔY %
A286 Iron	1.0	0.6	2.3	1.6
ABS	1.7	2.2	0.0	0.7
AISI 304	3.3	0.7	0.3	2.8
Gray Cast Fe	1.3	2.7	2.1	2.9
3003-H16	0.3	0.3	0.3	2.3
5052-O	0.8	2.4	1.2	1.3
7050-T7651	0.2	2.3	0.4	4.0
Acrylic	1.6	0.5	0.1	1.4
PE HD	3.2	0.3	2.0	3.9
Max error	**6.1**	**5.2**	**3.2**	**9.0**

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.

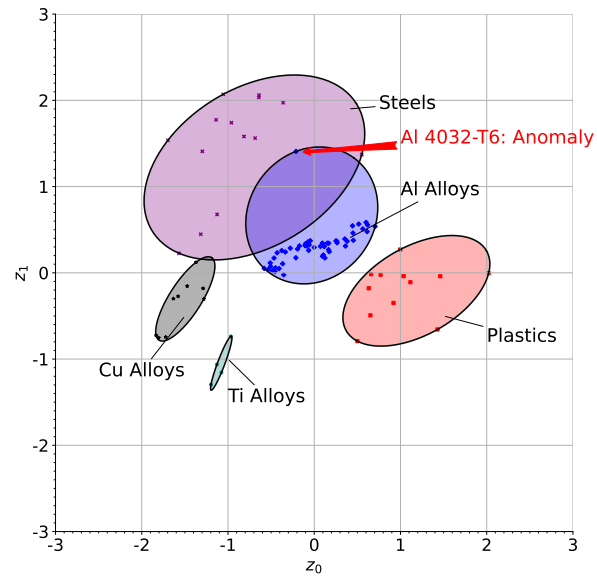


FIGURE 8: ANOMALY DETECTION IN THE MATERIAL CATALOG WITH VAE. THE AL 4032-T6 ENTRY IS AN OUTLIER, FURTHER AWAY FROM OTHER AL ALLOYS

TABLE 3: A CURATED SUBSET OF CROSS SECTIONAL AREAS AND THEIR PROPERTIES.

Section ID	Section Type	Area A $[10^{-4}m^2]$	I_x $[m^4]$	I_y $[m^4]$
1	Rectangle	4.00	1.33E-08	1.33E-08
2	Rectangle	4.44	1.83E-08	1.48E-08
3	Rectangle	4.89	2.43E-08	1.63E-08
101	Disk	12.6	1.26E-07	1.26E-07
102	Disk	12.8	1.31E-07	1.31E-07
103	Disk	13.1	1.36E-07	1.36E-07
201	Hollow Rect	1.44	7.87E-09	7.87E-09
202	Hollow Rect	1.60	1.24E-08	9.17E-09
203	Hollow Rect	1.76	1.82E-08	1.05E-08
309	Annulus	2.39	4.32E-08	4.32E-08
310	Annulus	2.64	4.72E-08	4.72E-08
311	Annulus	2.88	5.11E-08	5.11E-08
409	I-Beam	2.50	1.21E-08	6.77E-09
410	I-Beam	2.70	1.96E-08	6.81E-09
411	I-Beam	2.90	2.93E-08	6.85E-09

This implies that all attributes, for disc shapes, are functions of a single variable (in this case, radius r). Such intrinsic properties are often hard to detect from tabular data.

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,

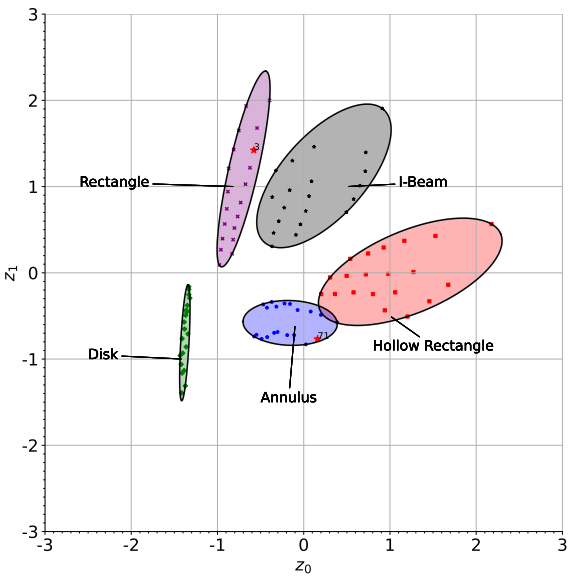


FIGURE 9: LATENT SPACE OF CROSS SECTIONS.

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.

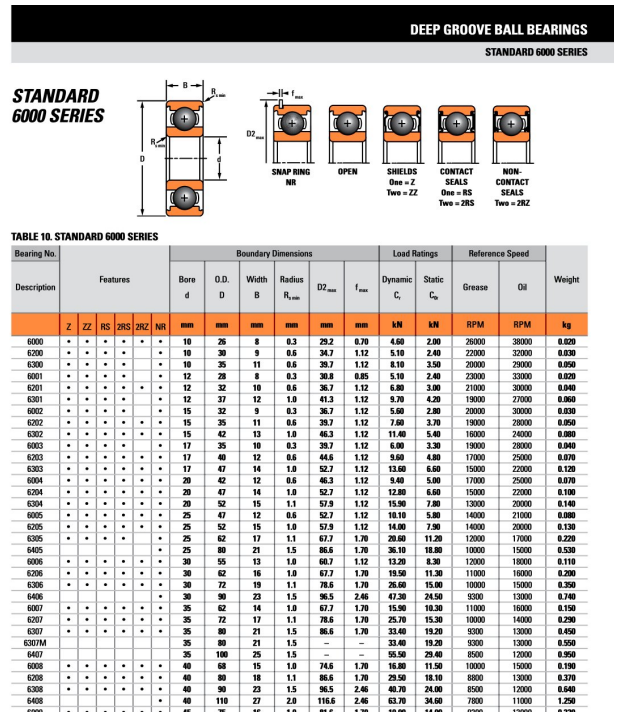
For the same training parameters, we compare the representational accuracy of 2D and 3D latent spaces. We report the maximum and average reconstruction errors in Table 4. Observe that the 3D latent space has a lower average reconstruction error across every attribute.

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

Parameter	2D Latent		3D Latent	
	Max(%) Error	Avg(%) Error	Max(%) Error	Avg(%) Error
Bore Dia	2.8	0.78	3.6	0.75
Outer Dia	2.4	0.63	2.4	0.44
Width	3.4	0.55	1.2	0.34
Load Rating	9.1	2	4.3	1.17
RPM	6	1.23	4.2	0.69
Weight	9.6	2.37	6.5	1.49

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,



Bearing No. Description	Features Z	ZZ	RS	2RS	2RZ	NR	Boundary Dimensions Bore d	O.D. D	Width B	Radius $R_{1,min}$	$D2_{max}$	f_{max}	Load Ratings Dynamic C_r	Static C_{0r}	Reference Speed Grease	Oil	Weight
							mm	mm	mm	mm	mm	mm	kN	kN	RPM	RPM	kg
6000	•	•	•	•	•	•	10	26	8	0.3	29.2	0.70	4.60	2.00	26000	38000	0.020
6200	•	•	•	•		•	10	30	9	0.6	34.7	1.12	5.10	2.40	22000	32000	0.030
6300	•	•	•	•		•	10	35	11	0.6	39.7	1.12	8.10	3.50	20000	29000	0.050
6001	•	•	•	•		•	12	28	8	0.3	30.8	0.85	5.10	2.40	23000	33000	0.020
6201	•	•	•	•	•	•	12	32	10	0.6	36.7	1.12	6.80	3.00	21000	30000	0.040
6301	•	•	•	•		•	12	37	12	1.0	41.3	1.12	9.70	4.20	19000	27000	0.060
6002	•	•	•	•		•	15	32	9	0.3	36.7	1.12	5.60	2.80	20000	30000	0.030
6202	•	•	•	•	•	•	15	35	11	0.6	39.7	1.12	7.60	3.70	19000	28000	0.050
6302	•	•	•	•	•	•	15	42	13	1.0	46.3	1.12	11.40	5.40	16000	24000	0.080
6003	•	•	•	•		•	17	35	10	0.3	39.7	1.12	6.00	3.30	19000	28000	0.040
6203	•	•	•	•	•	•	17	40	12	0.6	44.6	1.12	9.60	4.80	17000	25000	0.070
6303	•	•	•	•	•	•	17	47	14	1.0	52.7	1.12	13.60	6.60	15000	22000	0.120
6004	•	•	•	•	•	•	20	42	12	0.6	46.3	1.12	9.40	5.00	17000	25000	0.070
6204	•	•	•	•	•	•	20	47	14	1.0	52.7	1.12	12.80	6.60	15000	22000	0.100
6304	•	•	•	•	•	•	20	52	15	1.1	57.9	1.12	15.90	7.80	13000	20000	0.140
6005	•	•	•	•	•	•	25	47	12	0.6	52.7	1.12	10.10	5.80	14000	21000	0.080
6205	•	•	•	•	•	•	25	52	15	1.0	57.9	1.12	14.00	7.90	14000	20000	0.130
6305	•	•	•	•		•	25	62	17	1.1	67.7	1.70	20.60	11.20	12000	17000	0.220
6405						•	25	80	21	1.5	86.6	1.70	36.10	18.80	10000	15000	0.530
6006	•	•	•	•	•	•	30	55	13	1.0	60.1	1.12	13.20	8.30	12000	18000	0.110
6206	•	•	•	•	•	•	30	62	16	1.0	67.7	1.70	19.50	11.30	11000	16000	0.200
6306	•	•	•	•	•	•	30	72	19	1.1	78.6	1.70	26.60	15.00	10000	15000	0.350
6406						•	30	90	23	1.5	96.5	2.46	47.30	24.50	9300	13000	0.740
6007	•	•	•	•	•	•	35	62	14	1.0	67.7	1.70	15.90	10.30	11000	16000	0.150
6207	•	•	•	•	•	•	35	72	17	1.1	78.6	1.70	25.70	15.30	10000	14000	0.290
6307	•	•	•	•	•	•	35	80	21	1.5	86.6	1.70	33.40	19.20	9300	13000	0.450
6307M							35	80	21	1.5	–	–	33.40	19.20	9300	13000	0.550
6407							35	100	25	1.5	–	–	55.50	29.40	8500	12000	0.950
6008	•	•	•	•	•	•	40	68	15	1.0	74.6	1.70	16.80	11.50	10000	15000	0.190
6208	•	•	•	•	•	•	40	80	18	1.1	86.6	1.70	29.50	18.10	8800	13000	0.370
6308	•	•	•	•	•	•	40	90	23	1.5	96.5	2.46	40.70	24.00	8500	12000	0.640
6408	•	•	•	•	•	•	40	110	27	2.0	116.6	2.46	63.70	34.60	7800	11000	1.250

FIGURE 10: DEEP GROOVE BALL BEARINGS: TIMKEN [39]

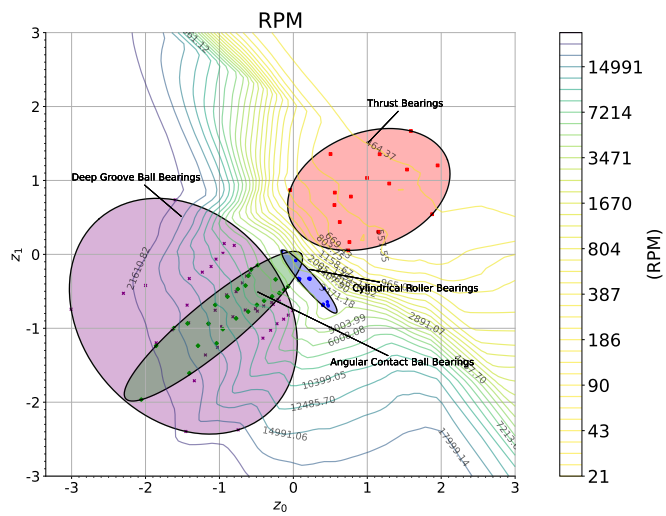


FIGURE 11: RPM CONTOUR PLOT SUPERIMPOSED ON THE LATENT SPACE OF BEARINGS

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.

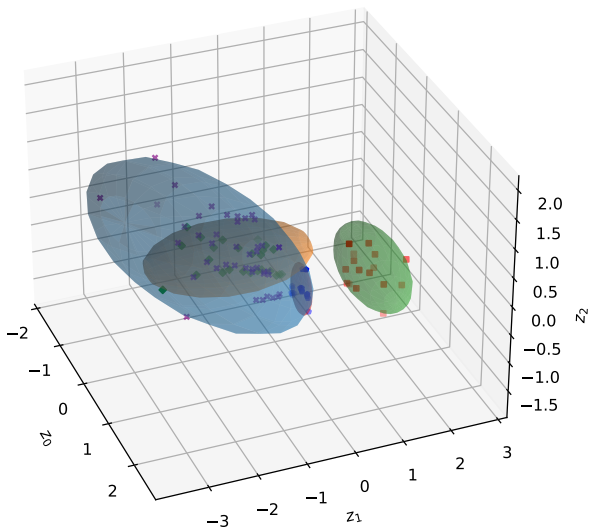


FIGURE 12: VISUALIZING THE BEARING CATALOG IN A 3D LATENT SPACE

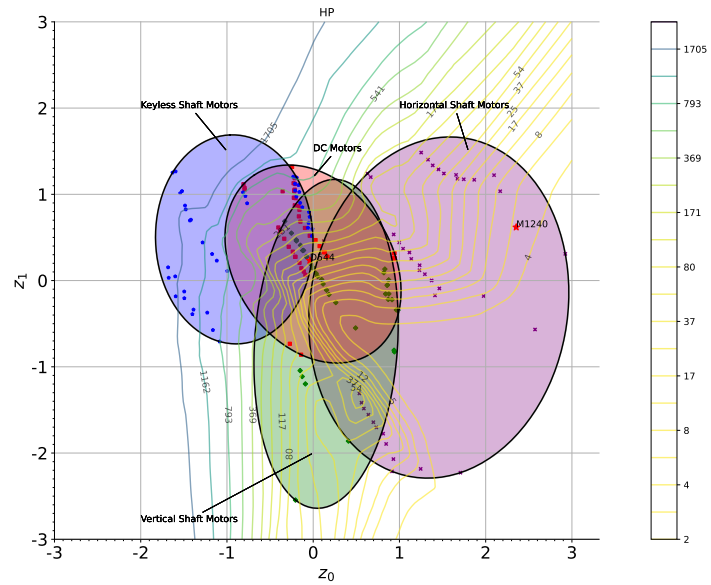
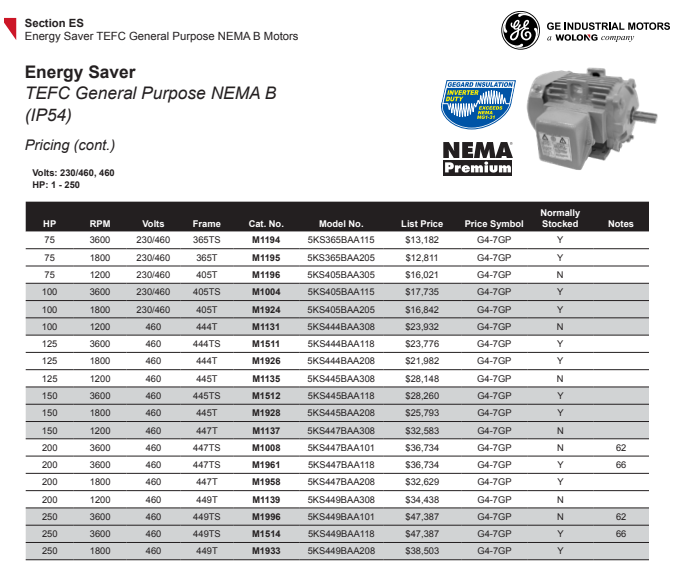


FIGURE 14: HORSEPOWER CONTOUR PLOT SUPERIMPOSED ON THE LATENT SPACE OF MOTORS



Section ES
Energy Saver TEFC General Purpose NEMA B Motors

GE INDUSTRIAL MOTORS
a WOLONG company

Energy Saver
TEFC General Purpose NEMA B
(IP54)

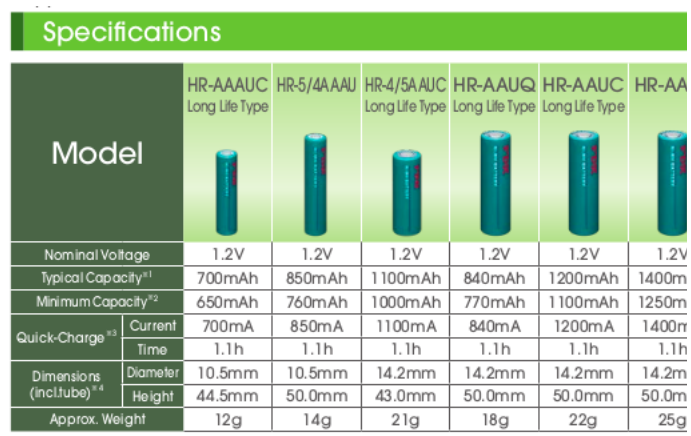
Pricing (cont.)

Volts: 230/460, 460
HP: 1 - 250

NEMA Premium

HP	RPM	Volts	Frame	Cat. No.	Model No.	List Price	Price Symbol	Normally Stocked	Notes
75	3600	230/460	365TS	**M1194**	5KS365BAA115	$13,182	G4-7GP	Y	
75	1800	230/460	365T	**M1195**	5KS365BAA205	$12,811	G4-7GP	Y	
75	1200	230/460	405T	**M1196**	5KS405BAA305	$16,021	G4-7GP	N	
100	3600	230/460	405TS	**M1004**	5KS405BAA115	$17,735	G4-7GP	Y	
100	1800	230/460	405T	**M1924**	5KS405BAA205	$16,842	G4-7GP	Y	
100	1200	460	444T	**M1131**	5KS444BAA308	$23,932	G4-7GP	N	
125	3600	460	444TS	**M1511**	5KS444BAA118	$23,776	G4-7GP	Y	
125	1800	460	444T	**M1926**	5KS444BAA208	$21,982	G4-7GP	Y	
125	1200	460	445T	**M1135**	5KS445BAA308	$28,148	G4-7GP	N	
150	3600	460	445TS	**M1512**	5KS445BAA118	$28,260	G4-7GP	Y	
150	1800	460	445T	**M1928**	5KS445BAA208	$25,793	G4-7GP	Y	
150	1200	460	447T	**M1137**	5KS447BAA308	$32,583	G4-7GP	N	
200	3600	460	447TS	**M1008**	5KS447BAA101	$36,734	G4-7GP	N	62
200	3600	460	447TS	**M1961**	5KS447BAA118	$36,734	G4-7GP	Y	66
200	1800	460	447T	**M1958**	5KS447BAA208	$32,629	G4-7GP	Y	
200	1200	460	449T	**M1139**	5KS449BAA308	$34,438	G4-7GP	N	
250	3600	460	449TS	**M1996**	5KS449BAA101	$47,387	G4-7GP	N	62
250	3600	460	449TS	**M1514**	5KS449BAA118	$47,387	G4-7GP	Y	66
250	1800	460	449T	**M1933**	5KS449BAA208	$38,503	G4-7GP	Y	

FIGURE 13: HORIZONTAL SHAFT MOTORS CATALOG, SOURCE: GE INDUSTRIAL MOTORS [40]



Specifications

Model		HR-AAAUC Long Life Type	HR-5/4AAAU	HR-4/5AAUC Long Life Type	HR-AAUQ Long Life Type	HR-AAUC Long Life Type	HR-AA
Nominal Voltage		1.2V	1.2V	1.2V	1.2V	1.2V	1.2V
Typical Capacity*1		700mAh	850mAh	1100mAh	840mAh	1200mAh	1400m
Minimum Capacity*2		650mAh	760mAh	1000mAh	770mAh	1100mAh	1250m
Quick-Charge*3	Current	700mA	850mA	1100mA	840mA	1200mA	1400m
	Time	1.1h	1.1h	1.1h	1.1h	1.1h	1.1h
Dimensions (incl.tube)*4	Diameter	10.5mm	10.5mm	14.2mm	14.2mm	14.2mm	14.2m
	Height	44.5mm	50.0mm	43.0mm	50.0mm	50.0mm	50.0m
Approx. Weight		12g	14g	21g	18g	22g	25g

FIGURE 15: NIMH BATTERIES CATALOG, SOURCE: FDK BATTERIES [41]

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.

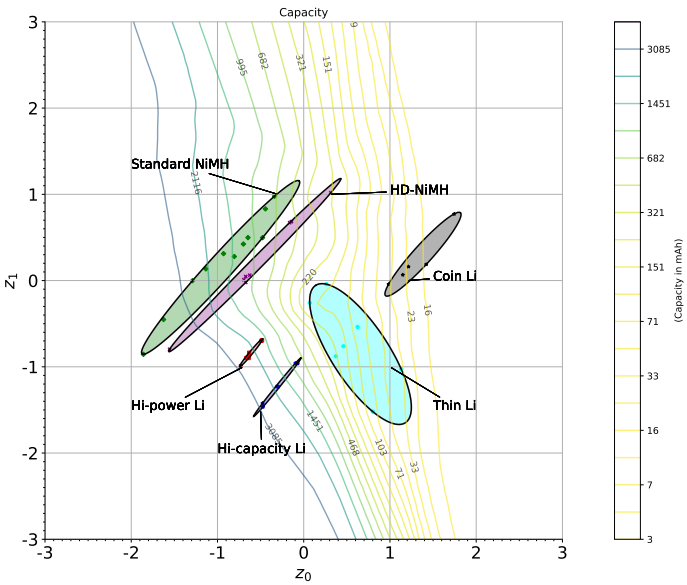


FIGURE 16: CAPACITY CONTOURS SUPERIMPOSED ON THE LATENT SPACE OF BATTERIES

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

[1] Requicha, Aristides G. "Representations for rigid solids: Theory, methods, and systems." *ACM Computing Surveys (CSUR)* Vol. 12 No. 4 (1980): pp. 437–464.

[2] Shigley, Joseph E, Mischke, Charles R and Brown Jr, Thomas Hunter. *Standard handbook of machine design*. McGraw-Hill Education (2004).

[3] Martinez Leon, AS, Rukavitsyn, AN and Jatsun, SF. "UAV Airframe Topology Optimization." *International Conference on Industrial Engineering*: pp. 338–346. 2021. Springer.

[4] NIST/SEMATECH. "Exploratory Data Analysis (EDA)." (2013). URL https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm. [Online; accessed 4-March-2023].

[5] Ashby, Michael F and CEBON, David. "Materials selection in mechanical design." *Le Journal de Physique IV* Vol. 3 No. C7 (1993): pp. C7–1.

[6] Pham, DT and Afify, AA. "Clustering techniques and their applications in engineering." *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* Vol. 221 No. 11 (2007): pp. 1445–1459.

[7] Keim, Daniel A. "Information visualization and visual data mining." *IEEE transactions on Visualization and Computer Graphics* Vol. 8 No. 1 (2002): pp. 1–8.

[8] Ashby, Michael F and Jones, David RH. *Engineering materials 1: an introduction to properties, applications and design*. Vol. 1. Elsevier (2011).

[9] Ashby, Michael F and Johnson, Kara. *Materials and design: the art and science of material selection in product design*. Butterworth-Heinemann (2013).

[10] Ananthasuresh, GK and Ashby, MF. "Concurrent design and material selection for trusses." *Workshop: Optimal Design, Laboratoire de Mécanique des Solides, Ecole Polytechnique Palaiseau France* (2003).

[11] Ashby, Mike. "Designing architectured materials." *Scripta Materialia* Vol. 68 No. 1 (2013): pp. 4–7.

[12] Rudin, Cynthia, Chen, Chaofan, Chen, Zhi, Huang, Haiyang, Semenova, Lesia and Zhong, Chudi. "Interpretable machine learning: Fundamental principles and 10 grand challenges." *Statistic Surveys* Vol. 16 (2022): pp. 1–85.

[13] Kruskal, Joseph B and Wish, Myron. *Multidimensional scaling*. 11, Sage (1978).

[14] Knerr, Nathan and Selva, Daniel. "Cityplot: visualization of high-dimensional design spaces with multiple criteria." *Journal of Mechanical Design* Vol. 138 No. 9 (2016): p. 091403.

[15] Kohonen, Teuvo. "The self-organizing map." *Proceedings of the IEEE* Vol. 78 No. 9 (1990): pp. 1464–1480.

[16] Thole, Sidhant Pravinkumar and Ramu, Palaniappan. "Design space exploration and optimization using self-organizing maps." *Structural and Multidisciplinary Optimization* Vol. 62 No. 3 (2020): pp. 1071–1088.

[17] Mika, Sebastian, Schölkopf, Bernhard, Smola, Alex, Müller, Klaus-Robert, Scholz, Matthias and Rätsch, Gunnar. "Kernel PCA and de-noising in feature spaces." *Advances in neural information processing systems* Vol. 11 (1998).

[18] Chen, Wei, Fuge, Mark and Chazan, Jonah. "Design manifolds capture the intrinsic complexity and dimension of design spaces." *Journal of Mechanical Design* Vol. 139 No. 5 (2017).

[19] Flexer, Arthur. "Limitations of self-organizing maps for vector quantization and multidimensional scaling." *Advances in neural information processing systems* Vol. 9 (1996).

[20] Schölkopf, Bernhard, Smola, Alexander and Müller, Klaus-Robert. "Kernel principal component analysis." *Artificial Neural Networks—ICANN'97: 7th International Conference Lausanne, Switzerland, October 8–10, 1997 Proceeedings*: pp. 583–588. 2005. Springer.

[21] Kingma, Diederik P and Welling, Max. "An introduction to variational autoencoders." *arXiv preprint arXiv:1906.02691* (2019).

[22] MacDorman, Karl F, Green, Robert D, Ho, Chin-Chang and Koch, Clinton T. "Too real for comfort? Uncanny responses to computer generated faces." *Computers in human behavior* Vol. 25 No. 3 (2009): pp. 695–710.

[23] Chen, Jinghui, Sathe, Saket, Aggarwal, Charu and Turaga, Deepak. "Outlier detection with autoencoder ensembles." *Proceedings of the 2017 SIAM international conference on data mining*: pp. 90–98. 2017. SIAM.

[24] Xu, Haowen, Chen, Wenxiao, Zhao, Nengwen, Li, Zeyan, Bu, Jiahao, Li, Zhihan, Liu, Ying, Zhao, Youjian, Pei, Dan, Feng, Yang et al. "Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications." *Proceedings of the 2018 world wide web conference*: pp. 187–196. 2018.

[25] Wang, Liwei, Chan, Yu-Chin, Ahmed, Faez, Liu, Zhao, Zhu, Ping and Chen, Wei. "Deep generative modeling for mechanistic-based learning and design of metamaterial systems." *Computer Methods in Applied Mechanics and Engineering* Vol. 372 (2020): p. 113377.

[26] Li, Xiang, Ning, Shaowu, Liu, Zhanli, Yan, Ziming, Luo, Chengcheng and Zhuang, Zhuo. "Designing phononic crystal with anticipated band gap through a deep learning based data-driven method." *Computer Methods in Applied Mechanics and Engineering* Vol. 361 (2020): p. 112737.

[27] Guo, Tinghao, Lohan, Danny J, Cang, Ruijin, Ren, Max Yi and Allison, James T. "An indirect design representation for topology optimization using variational autoencoder and style transfer." *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*: p. 0804. 2018.

[28] Champion, Kathleen, Lusch, Bethany, Kutz, J Nathan and Brunton, Steven L. "Data-driven discovery of coordinates and governing equations." *Proceedings of the National Academy of Sciences* Vol. 116 No. 45 (2019): pp. 22445–22451.

[29] Noh, Juhwan, Kim, Jaehoon, Stein, Helge S, Sanchez-Lengeling, Benjamin, Gregoire, John M, Aspuru-Guzik, Alan and Jung, Yousung. "Inverse design of solid-state materials via a continuous representation." *Matter* Vol. 1 No. 5 (2019): pp. 1370–1384.

[30] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." *Neural networks* Vol. 61 (2015): pp. 85–117.

[31] Doersch, Carl. "Tutorial on Variational Autoencoders." *https://arxiv.org/abs/1606.05908* (2016)DOI 10.48550/ARXIV.1606.05908.

[32] Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron. *Deep Learning*. MIT Press (2016).

[33] Fournier, Quentin and Aloise, Daniel. "Empirical comparison between autoencoders and traditional dimensionality reduction methods." *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*: pp. 211–214. 2019. IEEE.

[34] Tenenbaum, Joshua B, Silva, Vin de and Langford, John C. "A global geometric framework for nonlinear dimensionality reduction." *science* Vol. 290 No. 5500 (2000): pp. 2319–2323.

[35] Kingma, Diederik P. and Ba, Jimmy Lei. "Adam: A method for stochastic optimization." *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. 2015. International Conference on Learning Representations, ICLR. URL 1412.6980.

[36] Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimelshein, Natalia, Antiga, Luca, Desmaison, Alban, Kopf, Andreas, Yang, Edward, DeVito, Zachary, Raison, Martin, Tejani, Alykhan, Chilamkurthy, Sasank, Steiner, Benoit, Fang, Lu, Bai, Junjie and Chintala, Soumith. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc. (2019): pp. 8024–8035.

[37] Dassault Systèmes SE. "Simulation Materials." Accessed March 8, 2023, URL https://help.solidworks.com/2021/english/SolidWorks/cworks/r_Simulation_Materials.htm.

[38] Gao, Huanhuan, Breitkopf, Piotr, Coelho, Rajan Filomeno and Xiao, Manyu. "Categorical structural optimization using discrete manifold learning approach and custom-built evolutionary operators." *Structural and Multidisciplinary Optimization* Vol. 58 No. 1 (2018): pp. 215–228.

[39] "Timken Catalog." Accessed March 8, 2023, URL https://catalog.timken.com/.

[40] "GE Motors Wolong - Products." Accessed March 8, 2023, URL https://www.gemotorswolong.com/products.

[41] FDK Corporation. "Lithium Ion Battery Data Sheet." Accessed March 8, 2023, URL https://www.fdk.com/battery/lithium_e/data_sheet/FDK_battery_en.pdf.

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

Shakir and Lerchner, Alexander. "beta-vae: Learning basic visual concepts with a constrained variational framework." (2016).

[43] Razavi, Ali, Van den Oord, Aaron and Vinyals, Oriol. "Generating diverse high-fidelity images with vq-vae-2." *Advances in neural information processing systems* Vol. 32 (2019).

[44] Jain, Anubhav, Ong, Shyue Ping, Hautier, Geoffroy, Chen, Wei, Richards, William Davidson, Dacek, Stephen, Cholia, Shreyas, Gunter, Dan, Skinner, David, Ceder, Gerbrand et al. "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation." *APL materials* Vol. 1 No. 1 (2013): p. 011002.

[45] The Materials Project. "The Materials Project." (2021). Accessed March 10, 2023, URL https://materialsproject.org/.

[46] Spinner, Thilo, Körner, Jonas, Görtler, Jochen and Deussen, Oliver. "Towards an interpretable latent space: an intuitive comparison of autoencoders with variational autoencoders." *IEEE VIS 2018*. 2018.

[47] Leung, Frank Hung-Fat, Lam, Hak-Keung, Ling, Sai-Ho and Tam, Peter Kwong-Shun. "Tuning of the structure and parameters of a neural network using an improved genetic algorithm." *IEEE Transactions on Neural networks* Vol. 14 No. 1 (2003): pp. 79–88.

[48] Hou, Xianxu, Shen, Linlin, Sun, Ke and Qiu, Guoping. "Deep feature consistent variational autoencoder." *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*: pp. 1133–1141. 2017. IEEE.

[49] Chandrasekhar, Aaditya, Sridhara, Saketh and Suresh, Krishnan. "Integrating material selection with design optimization via neural networks." *Engineering with Computers* Vol. 38 No. 5 (2022): pp. 4715–4730.

[50] Chandrasekhar, Aaditya, Sridhara, Saketh and Suresh, Krishnan. "Auto: a framework for automatic differentiation in topology optimization." *Structural and Multidisciplinary Optimization* Vol. 64 No. 6 (2021): pp. 4355–4365.

[51] Bostanabad, Ramin, Chan, Yu-Chin, Wang, Liwei, Zhu, Ping and Chen, Wei. "Globally approximate gaussian processes for big data with application to data-driven metamaterials design." *Journal of Mechanical Design* Vol. 141 No. 11 (2019).

[52] Wang, Liwei, Tao, Siyu, Zhu, Ping and Chen, Wei. "Data-driven topology optimization with multiclass microstructures using latent variable Gaussian process." *Journal of Mechanical Design* Vol. 143 No. 3 (2021).