



HAL
open science

Invention Concept Latent Spaces for Analogical Ideation

Nicholas Walker

► **To cite this version:**

Nicholas Walker. Invention Concept Latent Spaces for Analogical Ideation. 18th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Jun 2022, Hersonissos, Greece. pp.313-324, 10.1007/978-3-031-08337-2_26 . hal-04668643

HAL Id: hal-04668643

<https://inria.hal.science/hal-04668643v1>

Submitted on 7 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Invention Concept Latent Spaces for Analogical Ideation

Nicholas Walker

Iprova SA, Building I, EPFL Innovation Park, CH-1015, Lausanne, Switzerland
nwalker@iprova.com

Abstract. Analogy is a powerful form of ideation and therefore an automated or semiautomated analogical method is a potentially useful way to develop, or at least inspire, new and possibly patentable ideas.

The last few years have shown significant developments in the training and use of latent spaces for text generation using Variational Autoencoders (VAE), though many problems remain including preventing ‘collapse’ of the latent space during its training and successfully disentangling the latent variables, including the syntax from the semantics.

A hierarchical sentence and document variational denoising autoencoder architecture is presented, in which the encoded sentence vectors are first generated and then an encoding and decoding is performed of the sequence (in the document) of these sentence vectors. The latent vectors for both sentences and documents are structured into ‘syntactic’ and ‘semantic’ subsections based on their use in auxiliary training tasks. A large dataset of patent titles and abstracts, along with their IPC6 codes, is used to train the VAE networks.

The resulting document latent space is used to perform analogy transforms to seek to generate/inspire useful and potentially novel patent concepts.

Keywords Patent Analysis, Ideation, Analogy, Latent Space, Text VAE.

1 Introduction

There is a need to enable innovation and automation in the process of invention, by providing for the human inventor powerful data-driven tools for uncovering novel and patentable concepts. One method is using analogies, where we find a meaningful existing analogy and apply this to a concept to generate a new, derived concept.

A variety of concept generation by analogy methods have been applied to patents (e.g. [1]), but here we seek a primarily data-driven approach.

1.1 Analogy in Latent Spaces

Suppose we can describe a large set of ‘concepts’ using the values of many underlying generative variables. Each concept is then a location in the multidimensional space of these variables. Then suppose we find that a common transform of one concept into another is to change the values of these generative variables in some consistent way. We can then find all the concepts within our area of interest that meet the conditions for this transform, apply it, and then check the space near the generated

concept's 'location' to ensure that no existing concept lies there. If not, we have a promising new concept and can examine it to determine its merits.

The classic example of this kind of approach is using word embeddings [2], [3] where we ask:

A is to B as X is to Y. What is B?

For example:

King is to B as Man is to Woman. Answer B = Queen.

Paris is to B as Madrid is to Spain. Answer: B = France.

Using Glove embeddings [4] and the equation: $B = A + (Y - X)$, we can get something like these results. We perform the vector arithmetic on the embeddings and find the best matching word as the one with the largest Cosine Similarity of existing word embeddings to the resulting vector.

Such word embeddings have no generative function – we cannot take an arbitrary embedding (latent) vector and generate a new word. For our purposes we need an encoder, which converts a concept into its latent vector, but also a decoder which can convert an arbitrary latent vector to a comprehensible concept. Finally, the analogical vector arithmetic will presumably only work successfully if the underlying latent variables are sufficiently disentangled.

Large collections of concept text descriptions are available in the patent literature. Therefore, if it is possible to derive well-structured latent spaces for these texts, and we can determine the subset of the latent variables which are important for describing the underlying concept, rather than superficial features of the text, and finally have a decoder which generates a comprehensible description of any vector sampled from that latent space, we should be able to perform our desired analogical transformations on short patent texts using such vector arithmetic [5].

1.2 Creativity using Latent Spaces

Generation of novel works from both continuous and categorical latent spaces is of increasing interest but has to date been mainly restricted to continuous data such as images or sound, predominantly using Generative Adversarial Networks (GAN). In [6] the authors use iterative guided exploration of an image latent space to design logos for companies. In [7], novelty in design images is identified and the GAN fine-tuned to increase its production of such novel features. The exploration of latent spaces for musical synthesis has been undertaken, including loudness-invariant musical timbre [8] and images (piano-rolls) of musical excerpts [9]. However, text latent spaces using GANs [10] remains difficult, since the discriminator gradient cannot easily backpropagate through the hard selection of text tokens and therefore training must use reinforcement learning or approximations such as the Gumbel-Softmax [11].

2 Text Autoencoders

In a text autoencoder, we train a network consisting of an encoder and decoder to reconstruct its input text via a latent vector. Text decoders are normally autoregressive

- the text is generated a token at a time, starting at the beginning of the text and moving forward, and the decoder has access to the latent vector and the previous tokens. They are generally implemented using either recurrent neural networks such as LSTM [12], or transformers [13]. At inference, tokens are selected using a greedy or stochastic algorithm or using beam search. Additional methods prevent repetitive text.

2.1 The Denoising Autoencoder (DAE)

With a simple autoencoder latent vectors for every different sequence of text in the training set could theoretically be placed in unique disjoint points in the latent space, with no ‘structure’ between them, such as similar sentences being encoded to similar latent vectors. By adding suitable ‘noise’ to the text, and reconstructing against the original text without noise, the encoder is encouraged to map a (set of small variations of) text to some compact volume of the latent space [14], [15]. If the extent of the latent space is constrained, the latent vector volumes will overlap and there is pressure on the autoencoder to map between similar latent vectors and similar sentences.

In [14], Shen et al added noise to the text by randomly replacing/removing words during training. They demonstrated that such noising encourages similar sentences to map to similar latent representations.

2.2 The Variational Autoencoder (VAE)

Another important method for structuring the latent space is to constrain it by forcing the latent variables to match a prior, such as a zero mean, unit variance Gaussian. The Variational Autoencoder [16], [17] has been one of the most popular generative models. A VAE encodes the input data samples into a latent variable from the distribution of representations from a probabilistic encoder, which is parameterized by a neural network outputting a value of the mean and variance of the encoded distribution. One or more samples are taken from this distribution by generating a random Gaussian vector and scaling it with the estimated mean and variance before passing to the decoder (at inference, the mean vector only is used).

A VAE aims to maximize the marginal likelihood of the reconstructed data, but this involves intractable posterior inference. Thus, this is approximated by training it to minimize a loss, which consists of a reconstruction term, generally the cross-entropy loss on the logits from the decoder, and a Kullback-Leibler (KL) loss on the divergence of the distribution of each latent variable from the prior.

2.3 β -VAE

β -VAE [17], [19] [20] is an extension to the basic VAE framework used for learning a more disentangled representation of the generative variables. A disentangled representation can be considered as one where single variables are sensitive to changes in single generative factors, while being relatively invariant to changes in other factors. Compared to the original VAE, β -VAE adds an extra hyperparameter β ($\beta > 1$) as a

weight of the KL divergence in the original VAE objective. The drawback of β -VAE is that the improved disentangling is offset by a poorer reconstruction.

3 Problems with Latent Vectors in Text VAEs

3.1 Latent Space Collapse

Text VAEs are quite difficult to successfully train. Basic VAE models cannot explicitly control the KL divergence and the KL term can progressively dominate the loss. They then suffer from KL vanishing [21], [22], which means the KL divergence can become zero during optimization – the latent variables collapse their means to 0.0 and contribute no information to the decoder.

The underlying reason is that the text decoders are autoregressive and can therefore operate moderately well as a pure language model, generating the text solely from the previous tokens and ignoring the latent vector.

Preventing collapse has been explored using a variety of approaches:

1. Text generation with a more limited decoder [23] or without using an autoregressive decoder [24].
2. Masking or corrupting some of the tokens used in the autoregression decoder to force greater use of the latent vector (noisy decoder) [25].
3. Providing additional tasks for the latent vector to ensure it retains information content, such as:
 - a. Predicting a text class or topic label
 - b. Predicting the bag of words in the text
 - c. Predicting entailment between texts, for example:
 - i. Entailment between sentences from the same document
 - ii. Entailment between immediately previous or subsequent sentences
4. Controlling the KL loss by scaling it with a factor (similar to β) which is:
 - a. Annealed from a low value to a high value using a sigmoid time course [20] or cyclically changed over the course of training [22].
 - b. Have a value scaled such that the KL loss for a minibatch is proportional to the reconstruction loss for that same minibatch [26].
 - c. Have a slowly modified factor controlling the average KL loss to a specific target value over the training set [27].
 - d. Gradually decreased from a starting high value to slowly increase the size of the informational bottleneck during training [20].

3.2 ‘Prompt only’ Text Generation from Latent Vector

An unfortunate additional problem for autoregressive decoders is that, even when actively using the latent vector, the latent, rather than coding the entire sentence, need only memorize the first few tokens in a text to produce a near unique ‘prompt’ which

the language model aspect of the decoder can then use to reconstruct the rest of the sentence [27]. Two approaches that can be adopted to help overcome this issue are:

- Independent masking ‘noise’ is added to the tokens used in the decoder [25].
- Use two autoregressive decoders - one going forward through the text and a second going backward, with the second, backward, decoder being dropped for inference [29].

4 Hierarchical VAE for text

Given we have a set of sentences, which are themselves sequentially organized within documents, and there are limits to the length of word sequence sensibly mapped to a single sentence latent vector, a hierarchical text VAE can be implemented in the manner of [30] to generate an encoding of the whole document. Therefore, in this work we created a 2 level autoencoder:

1. First each sentence’s tokenized text, plus special tokens <SOS>, <MASK>, <EOS> and <PAD>, is converted to a sentence latent vector.
2. Then the document sequence of variational samples of its sentence latent vectors plus concatenated 1-hot tokens <SOS>, <SEQ_N>*, <EOS>, <PAD> are converted by the document encoder into a single document latent vector, where <SEQ_N> are numbered sequence tokens for sentence vectors
3. A variational sample is generated and using this, the sentence vector sequence (and token logits) is reconstructed by the document decoder.
4. Finally, these reconstructed sentence latent vectors are in turn reconstructed as text by the sentence decoder.

The total set of losses (not including any auxiliary tasks) are then: the sentence KL divergence loss; the document KL divergence loss; the document sentence vector reconstruction loss (Mean Squared Error from the original un-noised sentence vectors to the decoded sentence vectors and cross entropy of the tokens); and the final sentence text reconstruction loss (cross entropy on the un-noised original text tokens). The document decoder also generates (from its syntax vector – see section 5) an estimate of the total number of sentences to be reconstructed. A transformer encoder and decoder were used for the sentence VAE and a transformer encoder and causal convolutional decoder for the document VAE.

5 Syntax and Semantics

In the desired conceptual representations, our interest is the inherent content - its semantics, rather than the way the author has specifically written the sentences - its syntax. Ideally, the latent space should be disentangled by these parts [31].

To seek to achieve this disentanglement, we can assume that close sentences (those from the same document, or specifically adjacent sentences in the same document) are most semantically similar to each other and that a topic label, such as the patent IPC code, again captures the semantics and not the syntax. Document syntax is more related to the number of sentences in a document.

Using these assumptions, auxiliary tasks were added for the latent spaces during training:

- The semantic part of the sentence latent vector samples was used in a two-vector input MLP binary classifier to discriminate between the sentences from the same document and sentences from other documents.
- The semantic part of both the sentence and the document latent vector samples were used (separately) in one vector input MLP classifiers to determine the main patent IPC class (in a one of N classification using cross entropy).
- The document syntactic vector sample was used in an MLP to predict the number of sentences in the document (one of N using cross entropy).

To preserve the syntax/semantics split in the sentence latent vectors, the document encoder and decoder were both split into two: one taking the sentence syntactic latent vectors and one the sentence semantic latent vectors.

For 50,000 documents, both latent spaces (sentence and document) were set as 312 dimensional, of which 32 dimensions were assigned as the syntactic part and 280 dimensions as the semantic part.

6 Implementation

6.1 Data

The BigPatent dataset [32] ‘abstract’ section was used. Patents were selected whose tokenized sentence length did not exceed 96 and with 7 or fewer sentences. IPC codes and citation links were downloaded from the USPTO datasets and matched to the patent abstracts. Abstracts were split into sentences using the NLTK sentence tokenizer and sentence text was word tokenized using the BERT lowercase tokenizer. A training set of 50,000 patents (178,787 sentences, 10,127,981 tokens) was made by selecting 25 patents from each of 2,000 IPC6 patent classes.

6.2 Network Architecture

The sentence encoder consisted of a 4 layer, 8 head transformer using sinusoidal positional encodings. An initial whole sentence embedding vector was formed by concatenating the first two transformer top layer (<SOS> and first token) token embeddings along with the maximum and mean top layer token embeddings. This vector was then processed by a 3-layer MLP and then two linear heads to generate the latent vector means and log variances. The variational sample from these was passed to a 4-layer MLP which generated a fixed length (9) sequence with same embedding dimension as the decoder. This ‘pseudo-sequence’ was used as the final latent description for the sentence decoders. The dual directional decoders consisted of 3-layer 4 head transformers with (per token) 3-layer MLP heads yielding the token logits.

The document encoder consisted of dual transformers, each with 5 layers and 8 heads, taking the sentence syntactic and semantic sentence subvector samples. Again, the first and second tokens plus max and means of the transformer top embedding

layers produced, via MLP heads, the mean and log variance vectors. A sample from these was again transformed into a multichannel ‘pseudo-sequence’ of length 9. The document decoder was a dual causal convolutional network [33] followed by a (per vector) 3-layer MLP yielding the sentence vectors and the added token logits (from the syntactic document decoder only). The document syntactic vector was also projected through a 3-layer MLP to yield the total number of sentences in the document, which was used to truncate the document decoder output at inference.

Auxiliary MLP heads were provided in the sentence encoder for the sentence semantic vector - one is a 1 of N classifier for the main IPC6 class, the other takes two vectors and returns a binary classification for the sentences coming from the same document or not. A similar main IPC6 class classifier was used for the document semantic vector.

6.3 Training

The sentence VAE was first trained alone, then its weights frozen and the document VAE trained. Both VAEs were then trained together. The Adam optimizer and a slow linear warm up followed by an exponential learning rate decay was used.

Noise was added to the sentence text for the sentence encoder to effect the benefits of a denoising autoencoder by a combination of masking (9% probability) and replacing words with similar words pre-generated by a large masked language model (6% probability). Different masked (only) sequences were used for the sentence decoder (15% masking). For sentence vectors supplied to the document encoder (only), masking was applied with only a low probability (1%) replacement of a sentence vector with a vector of zeros and a <MASK> rather than a <SEQ_N> token, since the sentence vectors are already ‘noised’ by the sentence noising and variational sampling.

The KL loss (both sentence and document) was structured as a proportion of the reconstruction loss for each minibatch. This proportion was increased during the start of their individual training using a sigmoid function from 0.0001 to 0.4 (sentence) and from 0.005 to 0.1 (document) and then fixed at the final value.

7 Results

Three approaches for concept generation from the resulting patent document latent space were investigated:

1. **Concept Fusion**
2. **Latent White Space**
3. **Citation Analogy**

50 examples of each approach were generated, their order randomized and presented blind to origin to five invention engineers at Iprova to assess the utility of the concepts for novel ideation on the basis of: Clarity and Coherence; Novelty/Patentability; and Utility for Ideation. Concepts were rated 0-5 for these 3 metrics and the average concept results obtained.

7.1 Concept Fusion

Two patents of interest are selected, and vectors formed as interpolations between their document latent vectors. Fig 1. shows an example of this approach.

Patent 1	A locking preset mechanism for a fishing reel is provided by interlocking balls in the preset knob face and the depressions in the mating face of an axially moveable force transmitting means in the mechanism. In free spool, the knob can be rotated. In drag, the knob is locked against rotation.
0.6 P1 0.4 P2	A surgical knife for use in a medical procedure comprises a handpiece capable of axial movement and a drive mechanism which controls the position of the movable member from its proximal end. The cutting board permits axial movement of a driving element into the opposed shank end. The clinician action is thereby effected into multiple positions.
0.5 P1 0.5 P2	According to embodiments a method and apparatus for controlling the flexure of a patient's limb are provided with which an operator can apply pressure to a field. The instrument features a guide assembly which permits placement of the pin in a substantially static and radial position. Additionally, the cutting element may be fitted with a guidewire. According to this disclosure there is variation and different mechanical characteristics are incorporated into the handle of the instrument with each other.
0.4 P1 0.6 P2	Specifically, a device and method for insertion of a valve for widening living tissue such as the human body and veins of a patient are provided. In particular a radiolucent structure is used at a distal end of the optical system and allows for numerous applications. In an alternate embodiment the sheath may be movable with respect to the distal portion. Additionally, the flexible assembly includes magnets which connect and protects the flaps in an opposed manner during operation of the device.
0.3 P1 0.7 P2	This invention provides a system and method for installing a stent made of a material such as a material that includes the radius of conical objects. In particular a pneumatic coupling device is formed across each of its plates providing for ease thereof. The catheter tip can carry the two modular elements into a body lumen upon insertion into a body lumen. The tip further comprises a working section suited for insertion into and along the isolated tissues
Patent 2	The present invention provides methods and devices for closing two overlapping layers of tissue in a mammalian heart, for example a patent foramen ovale. The closure devices may take a number of different forms and may be retrievable. In some embodiments, a device is sized and shaped to extend from septum secundum, into the left atrium, through septum primum, and into the right atrium, such that the first and second ends cooperate to provide a compressive force to the overlapping layers of tissue. In some embodiments, the closure devices may be delivered with a catheter capable of puncturing mammalian tissue.

Fig. 1. Concept Fusion interpolated vector texts.

7.2 Latent White Space

A subset of relevant patents is selected and their mean latent vector determined. Random Gaussian distributed vectors are found around this mean. Vectors whose semantic Cosine Similarity to existing patents exceed 0.75 are removed and the resulting latent vectors decoded into text.

Raw decoded text	Interpretation
An implant for forming a substance into a living body of a patient such as the method comprising the steps of providing a stent having a plurality of chambers and being insertable together into a root canal of a subject and methods of using the same are provided. The applicator	A stent-like tooth implant which forms the tooth by depositing within it a dispersion of a magnetic material and a powerful

further includes a filter element disposed within the cavity and is adapted to be placed in a substantially rigid state. One phase composition includes a polymer film impregnated with a dispersion of magnetic material and a polymeric material attached to the surface of the substrate and to the shell chain adhesive component.	adhesive which adheres to the internal filter to form the synthetically grown tooth. Magnets used for moving material (?) or heating material (?).
A coating comprising a deformable base for insertion into a body cavity, a cell surface capable of absorbing bone growth and passage and a hydrating solution material for making a dental prosthesis to be positioned on the tooth outer surface of the dental implant. The patch is more specifically to a solid cement and the structural shell comprising a fibrous resin which is more from the rest of the coating. An applicator for use in a dental implant the method including the steps of forming a tooth and or a rubber material so that the solid portion of the implant can be used to make the implant to be applied.	A similar concept to the above but using biological (?) tooth growth within a resinous cement and it is placed on a tooth (rather than replacing the tooth) as a patch. Tooth material is grown within the patch on top of the old tooth or an implant.
An invasive vascular diagnostic device designed to be inserted into human tissue and capable of administering the delivery of carbon dioxide which is used as a means of both filling and obtaining results from the same tissue. Such electrode units penetrate the tissue with rods which allow the gas to be applied at a variety of discrete locations along the path through the living tissue. Thus, the delivery system of the present invention can be applied to a tumor site and the patent describes methods for making and using such devices in a variety of therapeutic and diagnostic applications.	A rod delivering CO ₂ gas with electrodes on it, which is inserted into the body for cancer diagnosis and treatment. The gas expands the tissue to enable biopsy (?) and possibly is therapeutic (?). Electrodes measure the pH (?).

Fig. 2. Examples of concepts from 'Latent White Space' and their interpretation by an inventor.

7.3 Citation Analogy

Two relevant and relatively similar patents are found. The forward citations (inverted normal citations) of one of these patents are used to determine the vector difference between that patent and its forward cited patents. These difference vectors are applied to the other selected patent and a set of possible concept vectors (one for each forward citation) generated. These again are checked to reject those too close (semantic Cosine Similarity > 0.75) to existing patents and the remaining vectors decoded into text.

<p>Analogy start actual patent: A card dispensing system for dispensing cards along a conveying surface which conveys sliced products downstream of a slicer for packaging the sliced products. A hopper contains a stack of cards which are withdrawn from the hopper by a suction force provided via at least one suction cup. The suction cup is moveable vertically to retrieve and remove the card from the hopper. The suction cup is moveable laterally to move the card into a receiving position where a card clamp on a damp carriage arrives to receive the card. Suction force on the card is released when the card damp damp down on the card the clamp carriage moves the card from the receiving position to the staging position where a nip engages with the card to feed the card through a junction in the conveyor line just as a food product passes over the junction.</p>
<p>New Concept: A method of dispensing a web of bags having first and second conveying paths, the method including translational movement which accomplishes sequentially advancing the web of bags from the first conveying path to the second conveying path such that the web of bags is unwound from the first conveying path. The procedure is repeated until the product is exhausted through a nip by dispensing the web, whereupon it is then removed from the nip by suction. This ensures that the second portion of the web is able to be exhausted from the second portion through the nip of the reserve roll, while the reserve roll is dispensed from the first portion of the web and thence through the nip of the reserve roll to the second portion of the reserve roll for discharge. When the spool of the bale is being displaceable along the path of travel of the first knife and the suction line, thus preventing jams, are made by moving the first knife past the dispensing line until it reaches an edge of the path of travel of the second knife.</p>

New Concept: A plurality of paper dispensing cells for dispensing paper webs, each of the cells having a predetermined spacing that is skewed in relation to a direction perpendicular to the direction of conveyance of the paper webs and a predetermined spacing that is perpendicular to the direction of conveyance of the paper webs. The measuring device is positioned within the gap between the two rollers for measuring the distance between the first rollers and the spacing between the first rollers. The sensing element is then moved towards the exit conveyor, the sensing element is blocking the exit conveyor when the sensing element is moved from a first position to a second position. The sensor is activated when the receiving element is received by the first conveyor and the card feeding means and the receiver cooperating with the first conveyor.

Fig. 3. Analogical start patent and new concepts using Citation Analogy.

7.4 Evaluation

The following table gives the average concept evaluation results by five Iprova invention engineers all given the same sets of 50 concept texts per approach (total 150).

Table 1. Results of Inventor Evaluations.

Analogical Process	Clarity and Coherence	Novelty/Patentability	Ideation Potential
Concept Fusion	2.7	1.4	2.15
Latent White Space	3.3	2.4	3.2
Citation Analogy	2.2	1.65	2.1

8 Summary

The unsupervised learning of well-structured document latent spaces enables a variety of approaches for generating new concepts from existing concepts. The work reported here demonstrates a method of generating a document latent space for short patent abstracts, including addressing the problems associated with the training process. The resulting patent latent space was used for novel concept ideation. An evaluation of the generated concept text descriptions by Iprova invention engineers indicates that the ‘Latent White Space’ approach proved to be superior to the other two approaches evaluated.

Further work is required on better structuring and disentangling of the latent spaces, training significantly larger latent spaces on the entire patent corpus, along with automated screening and scoring of the resulting generated concepts.

References

1. Li-Zhen Jia, Chun-Long Wu, Xue-Hong Zhu & Run-Hua Tan, “Design by Analogy: Achieving More Patentable Ideas from One Creative Design”, (2018), Chinese Journal of Mechanical Engineering volume 31, Article number: 37.
2. Carl Allen, Timothy Hospedales, “Analogies Explained: Towards Understanding Word Embeddings”, (2019) Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

3. Dawn Chen, Joshua C. Peterson, Thomas L. Griffiths, “Evaluating vector-space models of analogy”, (2017), Proceedings of the 39th Annual Conference of the Cognitive Science Society
4. Jeffrey Pennington, Richard Socher, Christopher D. Manning, “GloVe: Global Vectors for Word Representation”, (2014), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)
5. Yoshua Bengio, Aaron Courville, and Pascal Vincent, “Representation Learning: A Review and New Perspectives”, (2014), IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8):1798-1828.
6. Puneet Jain, Najma Mathema, Jonathan Skaggs, Dan Ventura, “Ideation via Critic-Based Exploration of Generator Latent Space”, (2021), Proceedings of the 12th International Conference on Computational Creativity (ICCC '21).
7. Amin Heyrani Nobari, Muhammad Fathy Rashad, Faez Ahmed, “CREATIVEGAN: Editing Generative Adversarial Networks for Creative Design Synthesis”, (2021), Proceedings of the ASME 2021 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE 2021 August 17-20, 2021
8. Antoine Caillon, Adrien Bitton, Brice Gatinet, Philippe Esling, “Timbre Latent Space: Exploration and Creative Aspects”, (2020), Proceedings of the 2nd International Conference on Timbre (Timbre 2020), 3-4 September 2020, Thessaloniki, Greece.
9. Rodrigo F. Cádiz, Agustín Macaya, Manuel Cartagena, Denis Parra, “Creativity in Generative Musical Networks: Evidence From Two Case Studies”, (2021), Frontiers in Robotics and AI, 2021; 8: 680586
10. Gustavo H.de Rosa, João P.Papa, “A survey on text generation using generative adversarial networks”, Pattern Recognition, Volume 119, November 2021, 108098
11. E. Jang, S. Gu, B. Poole. “Categorical Reparameterization with Gumbel-Softmax” (2017) 5th International Conference on Learning Representations (ICLR 2017) Toulon, France.
12. Sepp Hochreiter, Jürgen Schmidhuber, “Long Short-Term Memory”, (1997), Neural Computation Volume 9 Issue 8 November 15, 1997 pp 1735-1780.
13. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin, “Attention Is All You Need”, (2017), 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA.
14. Tianxiao Shen, Jonas Mueller, Regina Barzilay and Tommi Jaakkola, “Educating Text Autoencoders: Latent Representation Guidance via Denoising”, (2020), Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.
15. Markus Freitag and Scott Roy, “Unsupervised Natural Language Generation with Denoising Autoencoders”, (2018), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3922-3929.
16. Danilo J. Rezende, Shakir Mohamed and Daan Wierstra, “Stochastic Backpropagation and Approximate Inference in Deep Generative Models”, (2014), Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014.
17. Diederik P. Kingma, Max Welling, “Auto-Encoding Variational Bayes”, (2014), ICLR 2014
18. Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed and Alexander Lerchner, “beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework”, (2017), ICLR April 2017.
19. Ricky T. Q. Chen, Xuechen Li, Roger Grosse, David Duvenaud, “Isolating Sources of Disentanglement in VAEs”, (2018), 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

20. Christopher P. Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, Alexander Lerchner, “Understanding disentangling in B-VAE”, (2017), 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
21. Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz and Samy Bengio, “Generating Sentences from a Continuous Space”, (2016), Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, Berlin, Germany 2016.
22. Hao Fu, Chunyuan Li, Xiaodong Liu, Jianfeng Gao, Asli Celikyilmaz and Lawrence Carin, Cyclical Annealing Schedule: “A Simple Approach to Mitigating KL Vanishing”, (2019), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1.
23. Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov and Taylor Berg-Kirkpatrick, “Improved Variational Autoencoders for Text Modeling using Dilated Convolutions”, (2017), Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.
24. Fei Huang, Jian Guan, Pei Ke, Qihan Guo, Xiaoyan Zhu and Minlie Huang, “A Text GAN for Language Generation with Non-Autoregressive Generator”, (2020), Under review as a conference paper at ICLR 2021.
25. Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu and Tie-Yan Liu, “MASS: Masked Sequence to Sequence Pre-training for Language Generation”, (2019), Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.
26. Andrea Asperti, Matteo Trentin, “Balancing reconstruction error and Kullback-Leibler divergence in Variational Autoencoders”, (2020), IEEE Access, Volume 8, pp: 199440 - 199448.
27. Huajie Shao, Shuochao Yao, Dachun Sun, Aston Zhang, Shengzhong Liu, Dongxin Liu, Jun Wang and Tarek Abdelzaher, “ControlVAE: Controllable Variational Autoencoder”, (2020), Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.
28. Tom Bosc, Pascal Vincent, “Do sequence-to-sequence VAEs learn global features of sentences?” (2020), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 4296-4318, November 16-20, 2020.
29. Tobias Watzel, Ludwig Krzinger, Lujun Li and Gerhard Rigoll, “Regularized Forward-Backward Decoder for Attention Models”, (2021), 23rd International Conference, SPECOM 2021, St. Petersburg, Russia, September 27-30, 2021.
30. Kun Zhao, Hongwei Ding, Kai Ye, and Xiaohui Cui, “A Transformer-Based Hierarchical Variational AutoEncoder Combined Hidden Markov Model for Long Text Generation”, (2021), Entropy (Basel)v.23(10); 2021 Oct.
31. Yu Bao, Hao Zhou, Shujian Huang, Lei Li, Lili Mou, Olga Vechtomova, Xinyu Dai, Jiajun Chen, “Generating Sentences from Disentangled Syntactic and Semantic Spaces”, (2019), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 6008-6019.
32. Eva Sharma, Chen Li and Lu Wang, “BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization”, (2019), Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2204–2213, Florence, Italy, 2019.
33. Adrián Javaloy, Ginés García-Mateos, “Text Normalization Using Encoder–Decoder Networks Based on the Causal Feature Extractor”, (2020), Applied Sciences 10(13):4551.