

# Chapter 40

## Counterfactual Autoencoder for Unsupervised Semantic Learning

**Saad Sadiq**

*University of Miami, USA*

**Mei-Ling Shyu**

*University of Miami, USA*

**Daniel J. Feaster**

*University of Miami - Miller School of Medicine, USA*

### ABSTRACT

*Deep Neural Networks (DNNs) are best known for being the state-of-the-art in artificial intelligence (AI) applications including natural language processing (NLP), speech processing, computer vision, etc. In spite of all recent achievements of deep learning, it has yet to achieve semantic learning required to reason about the data. This lack of reasoning is partially imputed to the boorish memorization of patterns and curves from millions of training samples and ignoring the spatiotemporal relationships. The proposed framework puts forward a novel approach based on variational autoencoders (VAEs) by using the potential outcomes model and developing the counterfactual autoencoders. The proposed framework transforms any sort of multimedia input distributions to a meaningful latent space while giving more control over how the latent space is created. This allows us to model data that is better suited to answer inference-based queries, which is very valuable in reasoning-based AI applications.*

### 1. INTRODUCTION

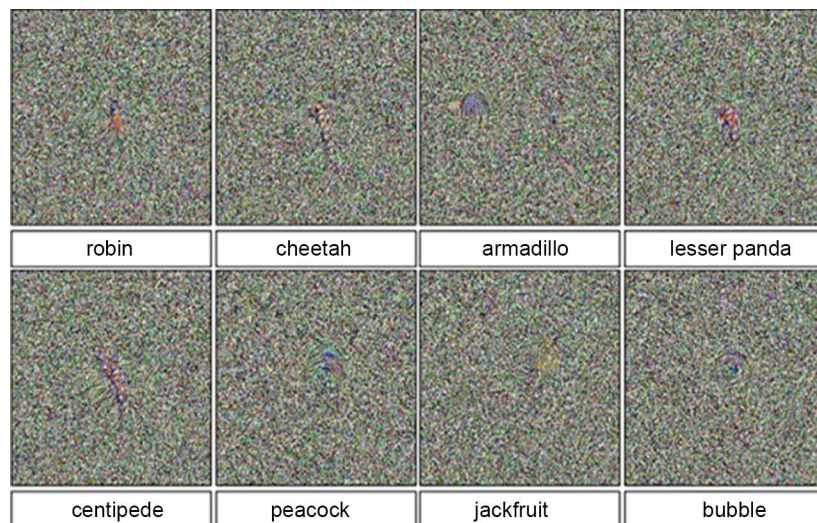
Often in real-world applications such as multimedia, NLP, and medicine, large quantities of unlabeled data are generated every day. This surge in data gives rise to the challenging semantic gap problem (Lin, Shyu & Chen, 2012; Chen, Lin & Shyu, 2012; Zhu & Shyu, 2015; Sadiq, Yan, Shyu, Chen & Ishwaran, 2016) which is to reduce the gap between high level semantic concepts and their low level

DOI: 10.4018/978-1-7998-0414-7.ch040

features (Sadiq, Tao, Yan & Shyu, 2017a; Sadiq, Zmieva, Shyu & Chen, 2018; Yan, Chen, Sadiq & Shyu, 2017). Despite rigorous research endeavors, this remains one of the most challenging problems in information sciences where we have overwhelming quantities of all sorts of fast, complex, heterogeneous and unstructured data. To handle such data, conventionally we have been utilizing descriptive models that try to find deterministic features and build probabilistic models (Chen & Kashyap, 1997; Chen, 2010; Lin, Shyu & Chen, 2013; Sadiq et al., 2017b). However, the problem with discriminative models is that they, generally, estimate a hyperplane. For example, when categorizing images of cats and dogs, data points on one side of the hyperplane are categorized as cats and everything on the other side as dogs. Discriminative models follow a condition in logistic regression that relaxes the computation of a joint probability  $p(x,y)$  to a conditional probability  $p(x|y)$  which is much easier to calculate because it maps directly to the hyperplane that divides between two clusters. This problem gets exacerbated in deep learning models due to the increase in dimensionality and boorish memorization of patterns. Mere rotations or color-inversions in the trained images can easily confound very deep neural networks even though the modified images share the same semantics and structures as the original images (Hosseini & Poovendran, 2017).

Recently, there has been a growing concern of this shortcoming, resulting in a soaring interest in Knowledge Representation and Reasoning (KRR) (Liu, 2017) and reasoning based deep learning (Andreas, Rohrbach, Darrell & Klein, 2016; Santoro et al., 2017). We believe that the next generation of AI systems will need to have the ability to understand problems at a deeper level rather than just based on memorization of data. However, discriminative models fail completely in inferencing problems because they do not capture the underlying relationships in the input space. Figure 1 shows how the addition of an adversarial noise can completely fool the classification while the L2 loss between the original and modified images was minimal. The discriminative models considered them as identical images while producing completely different results. All images were classified with 99.9% confidence.

*Figure 1. Classification results produced by state-of-the-art deep learning models, but fooled by simple noising up of the images (Nguyen, Yosinski & Clune, 2015)*



15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/counterfactual-autoencoder-for-unsupervised-semantic-learning/237901](http://www.igi-global.com/chapter/counterfactual-autoencoder-for-unsupervised-semantic-learning/237901)

## Related Content

---

### Design of Low-Power High-Speed 8 Bit CMOS Current Steering DAC for AI Applications

Banoth Krishna, Sandeep Singh Gilland Amod Kumar (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-18).

[www.irma-international.org/article/design-of-low-power-high-speed-8-bit-cmos-current-steering-dac-for-ai-applications/304801](http://www.irma-international.org/article/design-of-low-power-high-speed-8-bit-cmos-current-steering-dac-for-ai-applications/304801)

### Learning From the Past: Intelligence in Decision-Making Processes During Prisoners of War and Hostage Crises

Immacolata Carcarino (2021). *Computational Thinking for Problem Solving and Managerial Mindset Training* (pp. 104-126).

[www.irma-international.org/chapter/learning-from-the-past/282282](http://www.irma-international.org/chapter/learning-from-the-past/282282)

### New CICT Framework for Deep Learning and Deep Thinking Application

Rodolfo A. Fiorini (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 330-352).

[www.irma-international.org/chapter/new-cict-framework-for-deep-learning-and-deep-thinking-application/237880](http://www.irma-international.org/chapter/new-cict-framework-for-deep-learning-and-deep-thinking-application/237880)

### Supply Network Planning Models Using Enterprise Resource Planning Systems

Sundar Srinivasanand Scott E. Grasman (2008). *Handbook of Computational Intelligence in Manufacturing and Production Management* (pp. 437-453).

[www.irma-international.org/chapter/supply-network-planning-models-using/19371](http://www.irma-international.org/chapter/supply-network-planning-models-using/19371)

### The Formal Design Model of a File Management System (FMS)

Yingxu Wang, Cyprian F. Ngolah, Xinming Tan, Yousheng Tianand Phillip C.Y. Sheu (2011). *International Journal of Software Science and Computational Intelligence* (pp. 90-113).

[www.irma-international.org/article/formal-design-model-file-management/53165](http://www.irma-international.org/article/formal-design-model-file-management/53165)