Imperial College of Science, Technology and Medicine
Department of Computing

On the Feasibility of Using Fully-Convolutional Variational Autoencoders to Advance Deep Symbolic Reinforcement Learning

D. G. Sherburn

Submitted in part fulfilment of the requirements for the degree of
Master of Engineering in Joint Mathematics and Computing
of Imperial College, June 2017

Acknowledgements

I would like to express (whatever feelings I have) to:

- My supervisor
- My second supervisor
- Other researchers
- My family and friends

Dedication

Dedication here.

‘Quote text here.’

Guy Quoted

Contents

Acknowledgements **i**

1 Introduction **1**

1.1 Motivation 1

1.2 Objectives 4

1.3 Contributions 5

2 Background Theory **6**

2.1 Introduction 6

2.2 Arithmetic in Neural Networks 7

2.2.1 Neurons 7

2.2.2 Activation Functions 7

2.2.3 Convolutions 7

2.2.4 Deconvolutions 7

2.2.5 Pooling and Up-Sampling 7

2.3 Loss functions 7

2.3.1 Euclidean Distance 7

2.3.2 Binary Cross-Entropy 8

2.4 Autoencoders 9

2.4.1 Fully-Connected Autoencoders 10

2.4.2 Fully-Convolutional Autoencoders 11

2.4.3 Variational Autoencoders 12

3 Conclusion **17**

3.1 Summary of Thesis Achievements 17

3.2 Applications 17

3.3 Future Work 17

Bibliography **17**

List of Tables

1.1 Low-dimensional symbolic representation 5

2.1 A simple fully-connected autoencoder with one hidden layer. After 15 epochs, the validation score was recorded to be 71.94. 11

2.2 A simple fully-convolutional autoencoder with 2D convolutions and max pooling, plus the corresponding deconvolutional layers. After 15 epochs, the validation score was recorded to be 64.89. 11

List of Figures

1.1 May 1997: Gary Kasparov makes his first move against IBM's Deep Blue. Deep Blue would later emerge the victor in the best of six games; the first time a reigning world chess champion is defeated by a computer. [11] 2

1.2 March 2016: Lee Sedol, one of the greatest modern Go players, plays his first move of game three against AlphaGo. AlphaGo won four of five games. This feat was considered by many to be a decade away. [9] 2

1.3 Overview of deep symbolic reinforcement learning system architecture. **A**: The neural back end maps high-dimensional raw input data to a compositionally structured symbolic representation. **B**: The compositionally structured symbolic representation. **C**: Reinforcement learning of mapping from symbolic representation to action with maximum expected reward over time. *Source: Garnelo et al.* [3]. 4

1.4 A toy example of a raw high-dimensional input. 4

2.1 8

2.2 8

2.3 8

2.4 A black-box description of an autoencoder. The autoencoder learns the identity function, and in turn, the encoder and decoder learn suitable encoding and decoding algorithms respectively. 9

2.5 An example architecture of a fully-connected autoencoder. The latent space is constrained by having fewer neurons than the input and output layers. 10

2.6 A collection of images from the MNIST data set and their respective reconstructions using the fully-connected autoencoder specified in Table 2.1. The original MNIST images are in odd columns, and their reconstructions to their immediate right. 14

2.7 A collection of images from the MNIST data set and their respective reconstructions using the fully-convolutional autoencoder specified in Table 2.2. The original MNIST images are in odd columns, and their reconstructions to their immediate right. 15

2.8 An example architecture of a fully-convolutional autoencoder. The latent space is constrained by reducing the number and/or size of the filters. 16

2.9 An example architecture of a fully-convolutional autoencoder. The latent space is constrained by reducing the number and/or size of the filters. 16

Chapter 1

Introduction

1.1 Motivation

A long term goal of artificial intelligence (AI) is the development of artificial general intelligence (AGI). Since the field's inception in the 1950s, it has swung between hype and breakthroughs, followed by disappointment and reduced funding, known as AI winters [5]. During the first period of hype from the 50s to the early 70s, Marvin Minsky made the following prediction: [1]

> "In from three to eight years we will have a machine with the general intelligence of an average human being." - Marvin Minsky, 1970

This prediction was clearly not realised, and the first AI winter would shortly follow.

Symbolic AI was developed during this winter, which encodes knowledge as human-readable rules and facts, making it easy to comprehend chains of actions and abstract relationships [10]. For instance, given the unary relations `red` and `strawberry`, and the binary relation `bigger`, we can say that `A` is the smallest red strawberry by writing

```
red(A)  strawberry(A)  ∀B bigger(B, A)
```

But given the unary relations `yellow` and `banana` we could also write that `A` is the third biggest yellow strawberry, or a red banana, and so on. We can see that the rules and facts in symbolic logic can be endlessly recombined and extended. This allows for the manipulation of high-level abstract concepts, which is key to AGI [3].

However, symbolic AI has a major philosophical problem: the facts and rules are only meaningful to the human writing them; their meaning is not intrinsic to the system itself. This is known as the *symbol grounding problem*.

Today we find ourselves in yet another period of hype and exciting breakthroughs not afflicted by the symbol grounding problem. Reinforcement learning (RL) has become a prominent area of research, with many considering it fundemental for AGI [4], as have deep neural networks. Recently, deep reinforcement learning (DRL) systems have achieved impressive feats, including mastering a wide range of Atari 2600 games to a superhuman level using only raw pixels and score as input, and the board game Go [8, 12].

Figure 1.1: May 1997: Gary Kasparov makes his first move against IBM's Deep Blue. Deep Blue would later emerge the victor in the best of six games; the first time a reigning world chess champion is defeated by a computer. [11]

Figure 1.2: March 2016: Lee Sedol, one of the greatest modern Go players, plays his first move of game three against AlphaGo. AlphaGo won four of five games. This feat was considered by many to be a decade away. [9]

Though DRL systems are not afflicted by the same problems as symbolic AI, they have a number of drawbacks of their own. Namely, they are: [3]

1. **Slow to learn**. Neural networks require large data sets and are therefore slow to learn.

2. **Unable to transfer past experience**. They often fail to perform well on tasks very similar to those they have mastered.

3. **Unable to reason abstractly**. They fail to exploit statistical regularities in the data.

4. **Hard to reason about**. It's often difficult to extract a comprehensible chain of reasons for why a deep neural network operated in the way it did.

Deep symbolic reinforcement learning (DSRL) is a marrying of DRL and symbolic AI; a recent advance which overcomes the symbol grounding problem and the drawbacks associated with DRL [3]. That is, DSRL systems overcome the symbol grounding problem, and are:

1. **Fast to learn**. Large data sets are not necessary.

2. **Able to transfer past experience**. Symbolic AI lends itself to multiple processes associated with high-level reasoning, including transfer learning.

3. **Able to reason abstractly**. The agent is able to exploit statistical regularities in the training data by using high-level processes like planning or causal reasoning.

4. **Easy to reason about**. Since the front end uses symbolic AI, its knowledge is encoded as human-readble facts and rules, making the extraction of comprehensible chains of logic much easier.

An overview of DSRL is shown in Figure 1.3. The neural back end takes a high-dimensional input and outputs a symbolic representation. This symbolic representation is then fed to the symbolic front end, whose role is action selection. The agent then acts on the environment and obtains a reward and the sensory input of the next time step. As the neural back end learns how to represent the raw input data in a compositionally structured representation in an unsupervised manner, and the symbolic front end learns to select the action with maximum expected reward over time, the system as a whole learns end-to-end.

TODO: Finish description of DSRL

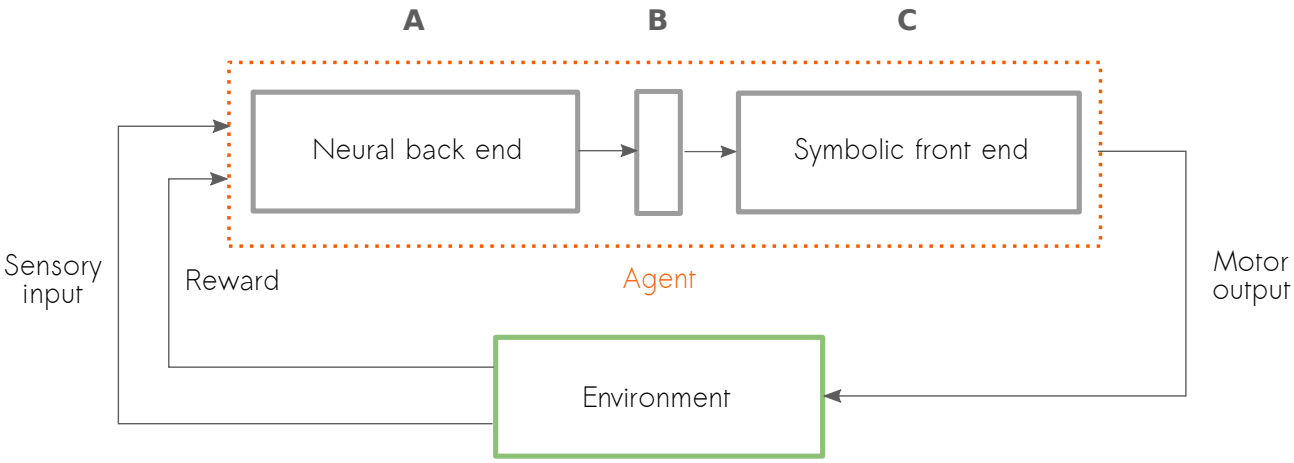

Figure 1.3: Overview of deep symbolic reinforcement learning system architecture. **A**: The neural back end maps high-dimensional raw input data to a compositionally structured symbolic representation. **B**: The compositionally structured symbolic representation. **C**: Reinforcement learning of mapping from symbolic representation to action with maximum expected reward over time. *Source: Garnelo et al.* [3].

1.2 Objectives

We'll use the image in Figure 1.4 as an example of a high-dimensional input to the neural back end. This world consists of only two shapes (circle and square) and four spaces occupied by at most one shape (top left, top right, bottom left and bottom right). The neural back end maps this raw high-dimensional input to a low-dimensional symbolic representation, shown in Table ??.

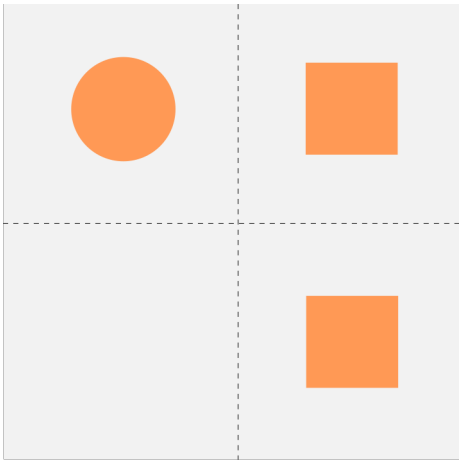

Figure 1.4: A toy example of a raw high-dimensional input.

How this is done will be explained in Chapter 2, but as for now we can just take it as fact that the current method doesn't scale. That is, for very simple scenes, as in Figure ..., This is done

Type	Location
1	[0, 0]
2	[0, 1]
2	[1, 1]

Table 1.1: Low-dimensional symbolic representation

by passing the high-dimensional input through a series of convolutional layers and extracting the activation spectra in the latent space. These spectra are then used to classify the

relies on the unsupervised extraction of disentangled features, allowing for transfer learning and high-level cognitive processes. However, the unsupervised extraction of features from a wide range of scenes is still a challenge in AI research [?]. Fortunately methods are getting better, and the first unsupervised scalable model β-VAE was developed recently.

TODO: Finish objectives

1.3 Contributions

Contributions here.

Chapter 2

Background Theory

2.1 Introduction

We will cover how deep symbolic reinforcement learning extracts symbolic representations from raw input data, which will motivate a discussion of loss functions and the introduction of autoencoders. Seeing the limitations of the current approach in extracting symbolic representations, we can appreciate the recent development of β-VAE, a variant of the variational autoencoder used to learn disentangled representations. Finally, we can conclude by mentioning less technical matters, such as libraries and hardware used.

2.2 Arithmetic in Neural Networks

2.2.1 Neurons

2.2.2 Activation Functions

2.2.3 Convolutions

2.2.4 Deconvolutions

2.2.5 Pooling and Up-Sampling

2.3 Loss functions

The idea of image reconstruction plays a vital role throughout this project. Although it's possible to qualitatively compare the original to its reconstruction, it's important to be able to quantify the difference, which lends itself to automation. The loss function will quantify how similar two images are.

To compare loss functions, we'll use the MNIST data set. MNIST is a collection of $70,000$ black-and-white images of handwritten digits, with $60,000$ in the training set containing and $10,000$ in the test set. These images will be represented as vectors without loss of generality.

2.3.1 Euclidean Distance

The Euclidean distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\sqrt{\sum_i (x_i - y_i)^2}$$

where x_i and y_i are the i^{th} components of $\mathbf{x}$ and $\mathbf{y}$ respectively.

Euclidean distance is an intuitive measure of the distance between two points in space. Unfortunately, this doesn't also translate to visual similarity, as illustrated by Doersch et al. [2]. Figure 2.1 is a digit drawn from the MNIST dataset, and Figures 2.3 and 2.2 are attempted reconstructions. Of the reconstructions, Figure 2.2 looks most like the original, but Figure 2.3 is closer in space.

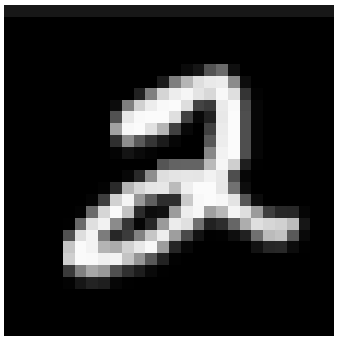

Figure 2.1

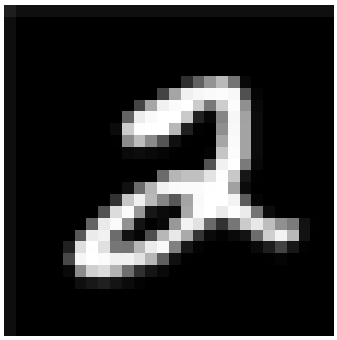

Figure 2.2

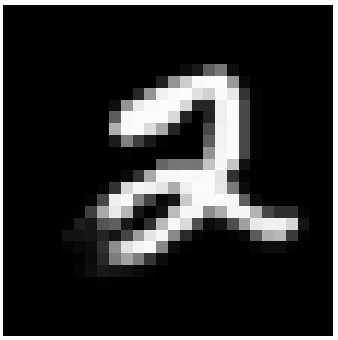

Figure 2.3

This leads to an alternative measure, binary cross-entropy, which gives a much better quantification of how visually similar two images are.

2.3.2 Binary Cross-Entropy

Consider a single black-and-white pixel with probability $p(0) = c$ of being 0 and $p(1) = 1 - c$ of being 1. Here $p(x)$ is a probability distribution over the possible pixel values $x \in \{0, 1\}$. Suppose a given model tries to learn the distribution described by $p(x)$, and says that the pixel has probability $q(0) = \hat{c}$ of being 0 and $q(1) = 1 - \hat{c}$ of being 1. The model is perfect if it learns the true distribution, that is, if $q(x) = p(x)$ for $x \in \{0, 1\}$. We'd like to quantify how similar the distributions p and q are.

This is done by computing the binary cross-entropy between p and q, which is defined by

$$H(p, q) = -c \log \hat{c} - (1 - c) \log(1 - \hat{c})$$

To see how we may use this as a similarity measure among images, consider a 1×1 image.

Normalising this image yields a pixel value in the interval $[0, 1]$, which may now be interpreted as a probability, corresponding to c above. In the normalised reconstructed image, the pixel value corresponds to $\hat{c}$. We simply compute the binary cross-entropy to measure the similarity of these two distributions, and in turn, the similarity of the images themselves! (Note: we could have also assigned the probabilities to $1 - y$ and $1 - \hat{y}$ by symmetry of binary cross-entropy).

For images larger than 1×1, we may take the component-wise binary cross-entropy, then, for example, average the components. How the component-wise binary cross-entropies are combined to give a score single floating point number is the choice of the designer and will vary from problem to problem.

2.4 Autoencoders

An autoencoder is a neural network that learns a compression algorithm for its input data in an unsupervised manner [7]. This is achieved by placing constraints on a hidden layer, called the latent space, and setting the target values to the input values, effectively learning the identity function. Since the network is trying to reconstruct the original input from the constrained latent space, over time the latent space corresponds to a meaningful compression of the network's input.

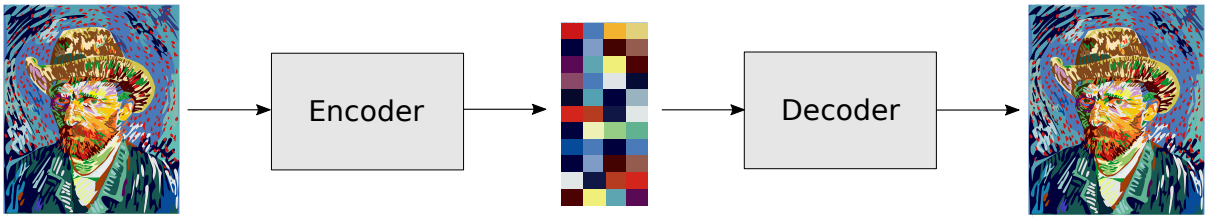

Figure 2.4: A black-box description of an autoencoder. The autoencoder learns the identity function, and in turn, the encoder and decoder learn suitable encoding and decoding algorithms respectively.

As before, we will use the MNIST data set to compare architectures. Unless specified in the example, the Adam optimiser is used with a learning rate of $1e - 4$, the batch size is 1 and the loss function is binary cross-entropy. Intermediate layers use the ReLU activation function, while the final layer uses sigmoid.

2.4.1 Fully-Connected Autoencoders

In dense feed-forward neural networks we may place a constraint on the latent space by reducing the number of neurons, as shown in Figure 2.5. Images must be flattened into vectors to be fed as input. Consequently, any spatial information is destroyed in dense feed-forward neural networks.

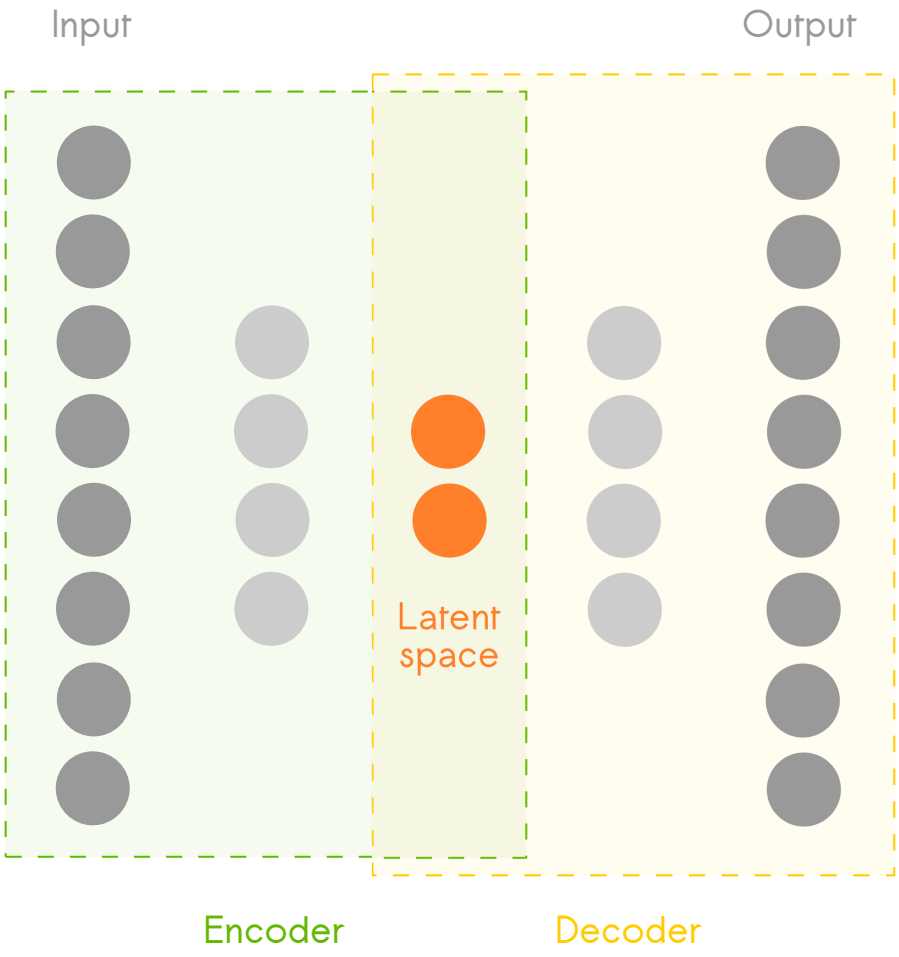

Figure 2.5: An example architecture of a fully-connected autoencoder. The latent space is constrained by having fewer neurons than the input and output layers.

An example architecture is given in Table 2.1, which was trained on MNIST. Despite the latent space being $\sim 4\%$ of the size of the input space, the network is capable of producing realistic reconstructions. For verification, a collection of samples from the dataset and their corresponding reconstructions are shown in Figure 2.6.

Layer Type	Output Shape
InputLayer	(1, 28, 28)
Flatten	(784,)
Dense	(32,)
Dense	(784,)
Reshape	(1, 28, 28)

Table 2.1: A simple fully-connected autoencoder with one hidden layer. After 15 epochs, the validation score was recorded to be 71.94.

2.4.2 Fully-Convolutional Autoencoders

In fully-convolutional feed-forward neural networks, we may place a constraint on the latent space by reducing the number and/or size of the filters, as shown in Figure 2.8. To compare the fully-convolutional autoencoder to the fully-connected, we'll train the architecture in Table 2.2 on MNIST. As before, we'll compare the reconstructions to the originals, which can be found in Figure 2.7.

Layer Type	Output Shape
InputLayer	(1, 28, 28)
Conv2D	(32, 28, 28)
MaxPooling2D	(32, 14, 14)
Conv2D	(4, 14, 14)
MaxPooling2D	(4, 7, 7)
UpSampling2D	(4, 14, 14)
Conv2DTranspose	(32, 14, 14)
UpSampling2D	(32, 28, 28)
Conv2DTranspose	(1, 28, 28)

Table 2.2: A simple fully-convolutional autoencoder with 2D convolutions and max pooling, plus the corresponding deconvolutional layers. After 15 epochs, the validation score was recorded to be 64.89.

Convolutional layers have been shown to be effective in tasks with images as input [6, 14, 13]. This is because spatial information is preserved in convolutional layers, and the number of trainable parameters is far less in a convolutional layer than it is in a fully connected layer. Convolutional layers will be used from here on as we'll be using images as input.

2.4.3 Variational Autoencoders

The variational autoencoder is central to this project, and we'll therefore dedicate a considerable amount of time exploring it. First we'll relate the variational autoencoder to ones seen earlier. This will lead us to formalise the proposed approach, after which we can develop an intuition about its implementation by way of concrete examples.

Introduction

So far we've seen autoencoders that take an input $\mathbf{x}$ and output a reconstruction $\tilde{\mathbf{x}}$, after mapping the input to a lower-dimensional representation $\mathbf{z}$. Putting our probabilistic hat on gives an entirely new and exciting idea: each of the inputs $\mathbf{x}$ are actually samples from an unknown probability distribution $p(\mathbf{x})$ [2]. The variational autoencoder's aim is to learn $p(\mathbf{x})$ in an unsupervised manner, and consequently be able to generate new samples never seen in the original data set.

Let's formalise this probabilistic perspective, then see how we may implement such a model.

Probabilistic Perspective

Let $X = \{\mathbf{x}^{(i)}\}_{i=1}^{N}$ be the data set of N independent and identically distributed samples of the variable $\mathbf{x}$. Let us assume that these samples are generated by a random process with parameters θ^* involving an unobserved latent variable $\mathbf{z}$ in the following way:

1. A latent variable $\mathbf{z}^{(i)}$ is drawn from the prior: $\mathbf{z}^{(i)} \sim p_{\theta^*}(\mathbf{z})$
2. A new data point $\mathbf{x}^{(i)}$ is drawn from a conditional distribution: $\mathbf{x}^{(i)} \sim p_{\theta^*}(\mathbf{x}|\mathbf{z})$

Let us also assume that the prior $p_{\theta^*}(\mathbf{z})$ and likelihood $p_{\theta^*}(\mathbf{x}|\mathbf{z})$ are parameterised by the distributions $p_\theta(\mathbf{z})$ and $p_\theta(\mathbf{x}|\mathbf{z})$ respectively. The true parameters θ^* and latent variables $\mathbf{z}^{(i)}$ that produced X are completely unknown to us; we only observe the samples in X.

We'd like to be able to infer the latent variable of a given data point, or more precisely, we'd like to be able to calculate the posterior $p(\mathbf{z}|\mathbf{x})$. By Bayes' theorem, the posterior is given by

$$p_\theta(\mathbf{z}|\mathbf{x}) = \frac{p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})}{p_\theta(\mathbf{x})}$$

However, due to the marginal likelihood

$$p_\theta(\mathbf{x}) = \int p_\theta(\mathbf{z})p_\theta(\mathbf{x}|\mathbf{z})d\mathbf{z}$$

being intractable, we may not calculate $p_\theta(\mathbf{z}|\mathbf{x})$ analytically. Instead, we define an approximation $q_\phi(\mathbf{z}|\mathbf{x})$ to the intractable posterior $p_\theta(\mathbf{z}|\mathbf{x})$. We will come to see how the variational autoencoder will jointly learn the parameters ϕ and θ. With these alone, we are able to encode a given sample with $q_\phi(\mathbf{z}|\mathbf{x})$, called the probabilistic encoder, and decode a given latent variable with $p_\theta(\mathbf{x}|\mathbf{z})$, called the probabilistic decoder.

Figure 2.6: A collection of images from the MNIST data set and their respective reconstructions using the fully-connected autoencoder specified in Table 2.1. The original MNIST images are in odd columns, and their reconstructions to their immediate right.

Figure 2.7: A collection of images from the MNIST data set and their respective reconstructions using the fully-convolutional autoencoder specified in Table 2.2. The original MNIST images are in odd columns, and their reconstructions to their immediate right.

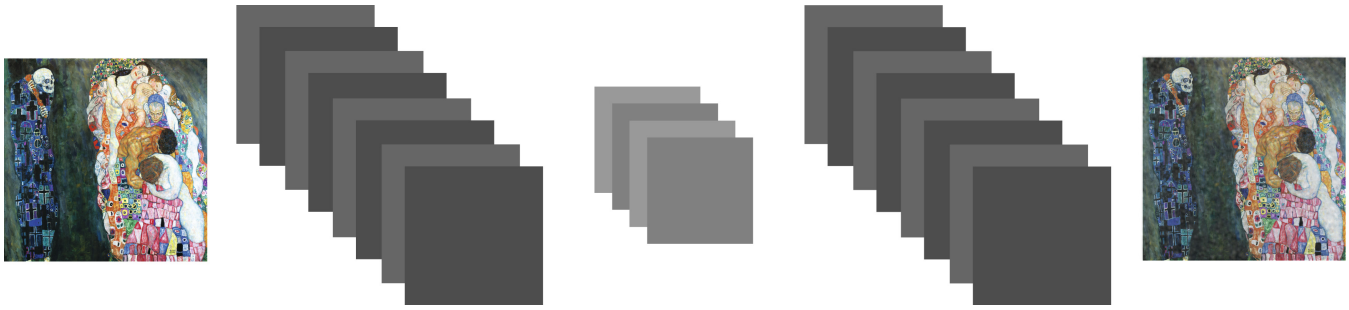

Figure 2.8: An example architecture of a fully-convolutional autoencoder. The latent space is constrained by reducing the number and/or size of the filters.

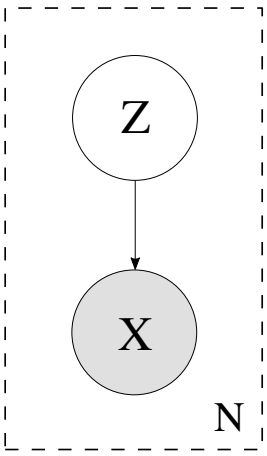

Figure 2.9: An example architecture of a fully-convolutional autoencoder. The latent space is constrained by reducing the number and/or size of the filters.

Chapter 3

Conclusion

3.1 Summary of Thesis Achievements

Summary.

3.2 Applications

Applications.

3.3 Future Work

Future Work.

Bibliography

[1] B. Darrach. Meet Shakey, the First Electronic Person. *Life*, pages 58–68, 1970.

[2] C. Doersch. Tutorial on Variational Autoencoders. *arXiv*, pages 1–23, 2016.

[3] M. Garnelo, K. Arulkumaran, and M. Shanahan. Towards Deep Symbolic Reinforcement Learning. *arXiv Preprint*, pages 1–13, 2016.

[4] M. Hutter. *Universal Artificial Intelligence*, volume 1. 2005.

[5] W. Knight. AI Winter Isn't Coming. *MIT Technology Review*, 2016.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012.

[7] C. Y. Liou, J. C. Huang, and W. C. Yang. Modeling word perception using the Elman network. In *Neurocomputing*, volume 71, pages 3150–3157, 2008.

[8] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[9] D. Ormerod. Lee Sedol plays the first move of game three against AlphaGo, 2016.

[10] E. U. o. T. Reingold. PSY371F Higher Cognitive Processes, 2001.

[11] R. Rosen. IBM's Deep Blue vs Gary Kasparov, 2012.

[12] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June, pages 1–9, 2015.

[14] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8689 LNCS, pages 818–833, 2014.