**A Generative Model for Semi-Supervised Learning**

Shang Da

A Creative Component submitted to the graduate faculty in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science

**Program of Study Committee:**

Dr. Jin Tian, Major Professor

Dr. Jia Liu, Committee Member

Dr. Wensheng Zhang, Committee Member

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Cy Cardinal, 2019. All rights reserved.

**TABLE OF CONTENTS**

Page

[LIST OF FIGURES iii](#_Toc21504413)

[LIST OF TABLES iv](#_Toc21504414)

[ACKNOWLEDGMENTS v](#_Toc21504415)

[ABSTRACT vi](#_Toc21504416)

[CHAPTER 1. INTRODUCTION 1](#_Toc21504417)

[CHAPTER 2. BACKGROUND 3](#_Toc21504418)

[2.1 Variational Inference 3](#_Toc21504419)

[2.2 The evidence lower bound (ELBO) 3](#_Toc21504420)

[2.3 Variational auto-encoders 4](#_Toc21504421)

[CHAPTER 3. RELATED WORK 6](#_Toc21504422)

[3.1 Berkhahn et al. 6](#_Toc21504423)

[3.2 Kingma’s M2 model 7](#_Toc21504424)

[CHAPTER 4. PROPOSED MODEL 8](#_Toc21504425)

[4.1 Probabilistic model 8](#_Toc21504426)

[4.2 Loss function 10](#_Toc21504427)

[4.3 Model architecture 12](#_Toc21504428)

[CHAPTER 5. EXPERIMENTS 14](#_Toc21504429)

[5.1 Semi-supervised performance 14](#_Toc21504430)

[5.2 Data generation 14](#_Toc21504431)

[5.3 Disentangled representations 14](#_Toc21504432)

[CHAPTER 6. CONCLUSION 15](#_Toc21504433)

[6.1 Automatically Linking Figures and Tables to the List of Figures and List of Tables 15](#_Toc21504434)

[6.2 Inserting Landscape Pages 18](#_Toc21504435)

[REFERENCES 21](#_Toc21504436)

[APPENDIX A. [INSERT APPENDIX TITLE HERE] 22](#_Toc21504437)

[APPENDIX B. [INSERT APPENDIX TITLE HERE] 23](#_Toc21504438)

LIST OF FIGURES

Page

[Figure 1. Architecture of Berkhahn’s model 6](#_Toc21505323)

[Figure 2. Architecture of Kingma M2 model 8](#_Toc21505324)

[Figure 3. Probabilistic model for data generation 9](#_Toc21505325)

[Figure 4. Probabilistic model for data generation in Berkhahn’s and Kingma’s work 10](#_Toc21505326)

[Figure 5. Probabilistic model for inference 11](#_Toc21505327)

[Figure 6. Model architecture 14](#_Toc21505328)

[Figure 7. Hiding Text from Figure and Table Captions in the List of Figures or Tables 18](#_Toc21505329)

LIST OF TABLES

Page

[Table 1. Instructions for Linking Figures and Tables to the Respective Lists 16](#_Toc21504445)

[Table 2. Instructions for Truncating Title Captions in the List of Figures and List of Tables 17](#_Toc21504446)

[Table 3. Instructions for Inserting Landscape Pages with Rotated Page Numbers 19](#_Toc21504447)

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Jin Tian, who has provided me with valuable guidance in every stage of the writing of this thesis. His keen and vigorous academic observation enlightens me not only in this thesis but also in my future study. I would also like to express my gratitude towards my committee members, Dr. Jia Liu, and Dr. Wensheng Zhang, for their valuable support.

In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. I want to also offer my appreciation to my parents for supporting me in my study.

ABSTRACT

Semi-Supervised learning is of great interest in a wide variety of research areas, including natural language processing, speech synthesizing, image classification, genomics etc. Semi-Supervised Generative Model is one Semi-Supervised learning approach that learns labeled data and unlabeled data simultaneously. A drawback of current Semi-Supervised Generative Models is that latent encoding is concatenated with predicted label directly, which may result in degradation in representation learning. In this paper we present a new Semi-Supervised Generative Models that removes the direct dependency of data generation on label, hence overcomes this drawback. We show experiments that verifies this approach, together with comparison with existing Semi-Supervised Generative Models.

# INTRODUCTION

Nowadays massive raw data is generated everyday thanks to the development of data gathering and storage techniques. However manual labeling of the large dataset is very time- and labor-consuming. In practice the number of unlabeled data is often far greater than that of labeled data. Hence, Semi-Supervised learning, which considers the problems of utilizing unlabeled data to assist supervised learning tasks, is of great interest in a wide variety of research areas, including natural language processing [6], speech synthesizing [7], image classification [8], genomics [9] etc. Existing semi-supervised learning models can be categorized into three main categories: unsupervised feature learning approach, graph-based regularization approach and multi-manifold learning approach.

Unsupervised feature learning approach achieves semi-supervised learning in two separate stages: feature representation learning stage and classification stage. In the first stage a set of latent representations are learnt from both labeled and unlabeled data with unsupervised generative models. In the second stage unlabeled data is classified based on learnt latent representations. For example, Kingma’s M1 model [1] first learn latent representations with auto-encoders then use an SVM to classify the results, and Johnson Rie et al. [10] use Local Region Convolution block to learn Two-View Embedding feature and then use a Convolution neural network for classification.

Study of Berkhahn et al [2] shows that classification results can serve as a regularizer to generative model, while generative model can provide extra information to the classifier. They achieve better performance with regard to both classification accuracy and feature representation learning by allowing mutual influence between the two models and training them simultaneously. This type of model that learns labeled and unlabeled data simultaneously is called Semi-Supervised Generative Model. It is first proposed by Kingma [1]. One drawback of most existing Semi-Supervised Generative Models is that they assume data generation is directly influenced by label. As a consequence, such models concatenate classification result with the latent encodings directly. This may result in degradation of representation learning [3].

Therefore, in this paper we present a new flavor of Semi-Supervised Generative Model which overcomes this problem. The proposed model is also able to handle both labeled and unlabeled data. Similar to Kingma and Berkhahn’s work, it utilizes the power of variational auto-encoder (VAE) for representations learning. But unlike existing works, the direct dependency of data generation on label is removed thanks to a new probabilistic modeling of data generation process.

In the future sections, we start with background knowledge, including Variational Inference and Variational auto-encoders in chapter 2. In chapter 3 we introduce some related approach prior to this work. In chapter 4 we propose the new model and in chapter 5 we present some experiments validating the new model.

# BACKGROUND

This section covers the necessary background knowledge to the proposed work, including a brief introduction to Variational Inference, the evidence lower bound and Variational auto-encoders.

## Variational Inference

The goal of unsupervised learning is to learn a set of latent variables to represent observed data . This requires learning the posterior distribution . Directly computing is often intractable since it often requires computing integration where is often a high dimensional variable.

Variational Inference [4] is one approach to estimate the intractable posterior distribution . The core idea of variational inference is:

1. Introduce a tractable hypothesis probability distribution parameterized by .

2. Find a that makes approximate , namely

where is a measurement of distance between two probability distributions. In practice one of the most wildly used measurement in Variational Inference is KL divergence.

## The evidence lower bound (ELBO)

Directly minimizing in (2) is intractable because of the dependency on the evidence :

Therefore, the evidence lower bound (ELBO) is introduced. It consists of the negative KL term plus evidence , which is a constant with respect to . Maximizing the ELBO is equivalent to minimizing the KL term, which is the goal of variation inference:

## Variational auto-encoders

Variational auto-encoder (VAE) is first proposed by Kingma & Welling [5]. They first start with the ELBO in (4).

To optimize the ELBO in deep learning framework. The author used an autoencoder-decoder structure. A generative model (decoder) is chosen to learnt while simultaneously an inference model (encoder) is chosen for the arbitrary probability distribution in variational inference setting. The objective becomes:

Then maximizing the ELBO becomes minimizing reconstruction error and a KL term. and are often chosen to be multivariate gaussian distribution with diagonal variance matrix so that the KL term is easily computable.

In this setting becomes a probability distribution. Rather than output latent encoding directly, the encoder estimates parameters of a gaussian distribution. Latent encoding is first sampled from the distribution and then fed to decoder. However, directly sampling from gaussian is not differentiable. In order to train the model with stochastic gradient descent. The author introduced a method called reparameterization trick: Instead of directly sample , we compute , where . This is equivalent to sampling from , but it’s differentiable hence trainable.

# RELATED WORK

In this chapter we introduce two Semi-Supervised Generative Model given by Berkhahn et al. [2] and Kingma [1]. These two pieces of work serve as inspiration for this paper. A further comparison to the proposed model is given in chapter 5.

## Berkhahn et al.

Berkhahn et al [2] present a Semi-Supervised Generative Model makes minimal changes to vanilla variational-autoencoder structure: the only addition is a classification layer that is attached to the topmost encoder layer. The input of decoder becomes concatenation of and **:**

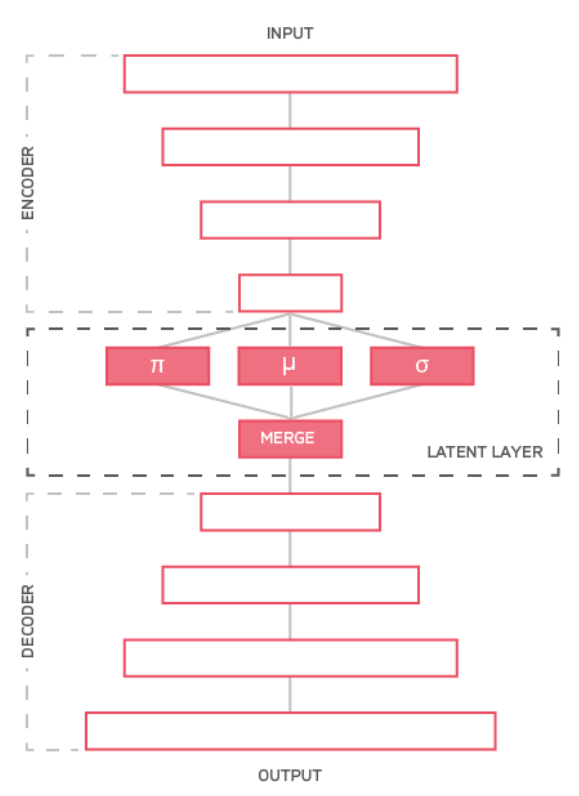


Figure 1. Architecture of Berkhahn’s model

The loss function of this model is traditional ELBO term plus classification error:

## Kingma’s M2 model

Kingma M2 model is proposed in [1]. It is the first Semi-Supervised Generative Model. It assumes that data is generated by a latent class variable **y** in addition to a continuous latent variable . Posterior is modeled by a decoder network taking and as input:

The approximate posterior has a factorized form:

where is modeled by an encoder network and are modeled by neural networks.

The model uses two different loss functions to handle both labeled and unlabeled data in semi-supervised learning setting.

For labeled data:

For unlabeled data:

Figure 2. shows the architecture of Kingma M2 model. Addition to vanilla VAE model, a classifier is applied to handle labeled data. Latent encoding and labelare concatenated before fed to decoder.

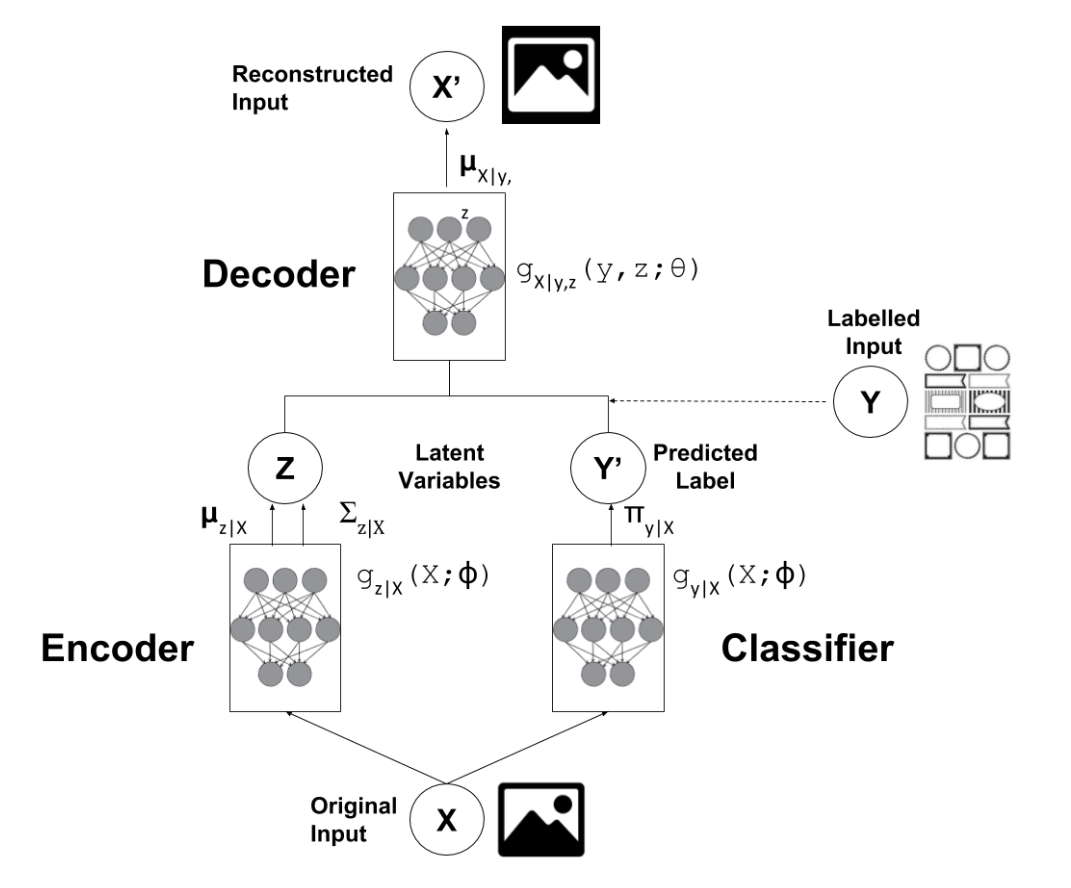


Figure 2. Architecture of Kingma M2 model

# PROPOSED MODEL

This chapter describes the proposed model in detail. We first introduce a new probabilistic model for data generation and inference. Then we show the corresponding loss functions for labeled and unlabeled data. Finally, we propose the model architecture.

## Probabilistic model

**Data generation**

We assume the following data generation process: first a label is chosen**,** aset of latent encoding is generated conditioned on . Then data is generated given purely latent encoding **.** Figure 2. shows the probabilistic model of the above generation process.

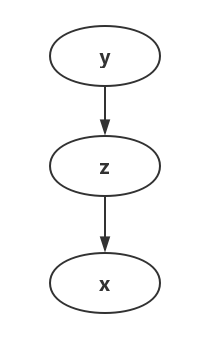


Figure 3. Probabilistic model for data generation

Hence, we have:

Similar to VAE. We choose a deep generative network to learn the true probability distribution .

Compared with Berkhahn’s model and Kingma’s M2 model, the proposed generative model removes the direct dependency of data on label **.** Therefore, latent encodings have all the information about reconstructing **.** This is a more desirable property for representation learning [3].

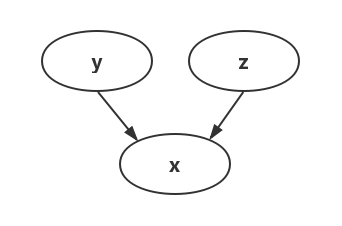


Figure 4. Probabilistic model for data generation in Berkhahn’s and Kingma’s work

**Inference model**

In variation inference setting we use an inference model to approximate posterior distribution . Figure 4. describes the inference model. We assume both andcan be inferenced directly from **.** Namely, we have:

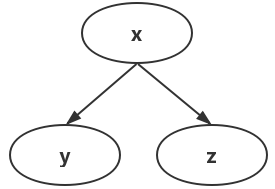


Figure 5. Probabilistic model for inference

## Loss function

Similar to Kingma’s work. We also use two different loss function to handle labeled and unlabeled data.

**Labeled Data**

For labeled data. is treated as observed variable. We minimize the following KL term:

Hence, we have the evidence lower bound for labeled data:

A classification error is also added in order to learn conditional probability .

**Unlabeled Data**

For unlabeled data. is treatedas unobserved latent variable. We minimize:

Hence, we have the evidence lower bound for unlabeled data:

**Total Loss function**

Finally, the total loss function for the entire dataset is now:

## Model architecture

As is shown in Figure 5. Our model is a modified version of vanilla variational-autoencoder. In order to optimize (16) and (19), some addition structures are added, including a classify-layer that outputs probability and predicts label , and a discriminator taking or as input and outputs and .

When a labeled input is fed to the model. The classify-layer tries to learn by minimizing cross entropy loss. The true label is fed to discriminator therefore is then computed.

When an unlabeled input is fed to the model. The classify-layer outputs probability . Then predicted label is fed to the discriminator so that is computed.

For both labeled and unlabeled data. Reconstruction error is computed, which is equivalent to term .

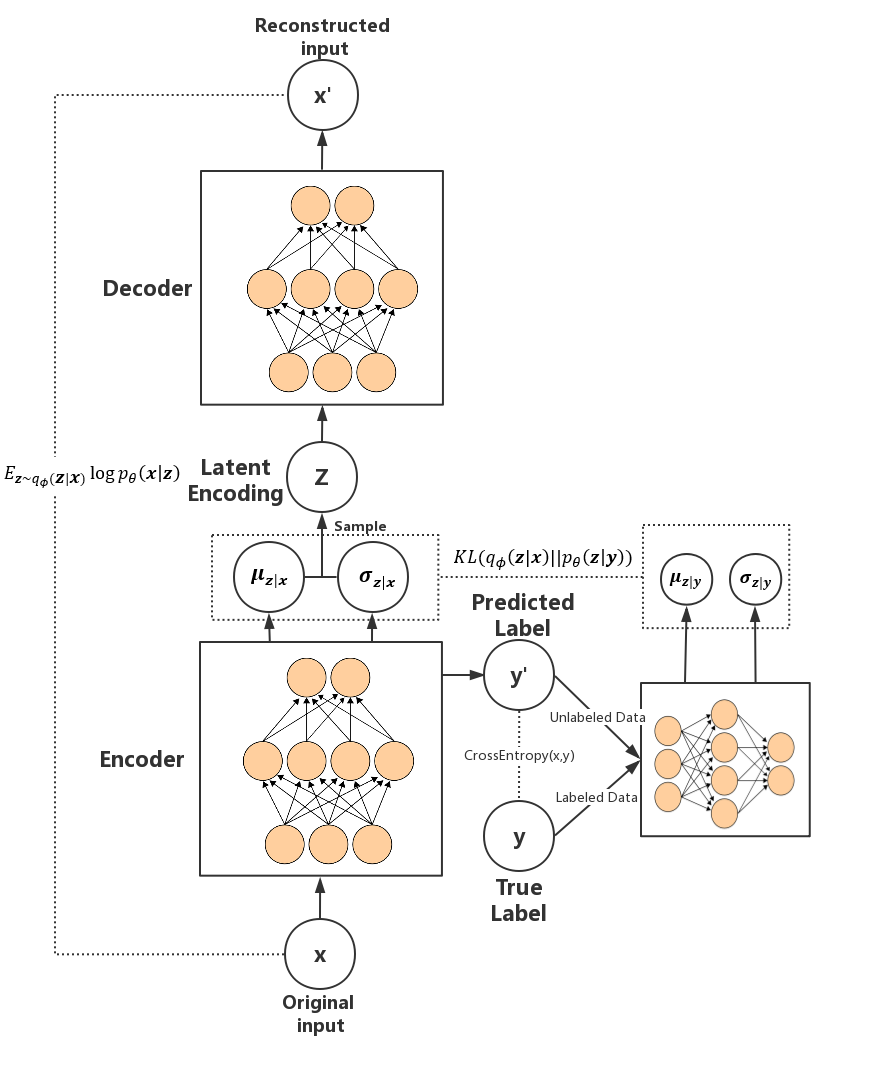
****

Figure 6. Model architecture

# EXPERIMENTS

## Semi-supervised performance

To Do

## Conditioned data generation

To Do

## Disentangled representations

To Do

# CONCLUSION

The Graduate College does not require a List of Figures or a List of Tables in your Dissertation/Thesis. However, if you choose to include either list, you must include the other (*Note: If you, for example, have a List of Figures, but no tables within your document, you do not have the List of Tables [and vice versa]*). You cannot combine these lists into one list. **You can embed your figures and tables within each chapter or create a single “Figures and Tables” section at the end of the chapter or document after the References section** (if you choose to create a Figures and Tables section at the end of the chapter, make sure to use Heading 2; if you choose to create a Figures and Tables section at the end of the document, make sure to use Heading 0 (in TOC)).

A consistent style should be used for all chapter tables and figures. Table captions are located at the top of the table. Figure captions are located at the bottom of the figure. Captions longer than one line uses consistent line spacing and indentation. They can be captioned sequentially (Figure 1, 2, 3, 4, etc.) or utilize chapter numbering (Figure 1.1., 1.2., 1.3., 2.1., 2.2, etc.). You can style the caption (e.g., bolded vs. italics, sentence case vs. uppercase, alignment, etc.) however you’d like, just be consistent.

## Automatically Linking Figures and Tables to the List of Figures and List of Tables

The process for linking figures and tables to their respective lists is nearly identical (see Table 1 for steps on linking figures and tables). After you follow these steps, highlight the portion of the title that says “Figure X” or “Table X” – if the number is in an extra dark box, you have correctly linked your Figure or Table. This is dynamic content. To update your List of Figures and List of Tables, follow the same procedure for updating the Table of Contents.

Table 1. Instructions for Linking Figures and Tables to the Respective Lists

|  |  |
| --- | --- |
| **Steps** | **Instructions** |
| Step 1 | Paste or insert your figure or table into the document. Make sure it fits inside of your margins. |
| Step 2 | Highlight the entire figure or table. Right click on the highlighted item and select Insert Caption. |
| Step 3 | Next to “Options, Label:”, select either Figure or Table. If it is a Figure, make sure the “Position:” option selected says “Below Selected Item” and if it is a Table, make sure the “Position:” option selected says “Above Selected Item”. |
| Step 4 | Click on the “Numbering, Format:” option. If you would like to include both the chapter number and item number in the Figure or Table title (e.g., Figure 5.1), check the “Include Chapter Number” box and select which kind of separator you want. If you do not want to chapter number in the caption, make sure to deselect this box. |
| Step 5 | Returning to the original editing box, write your Figure or Table title and press OK when you are finished. *Note: If you like to have a period and space after “Table 1”, make sure to manually write this in as you type your title.* |
| Step 6 | You will now see your Figure or Table title below or above the item, respectively. You can style these headers (bolded, italicized, centered, left justified, etc.) however you like, just be consistent throughout your document. You can also edit your title by adding or removing text. |
| Step 7 | Some versions of Word will insert the Figure or Table title within a text box. It is recommended that you cut the inserted title from the text box, delete the text box, and past it onto the regular document. |

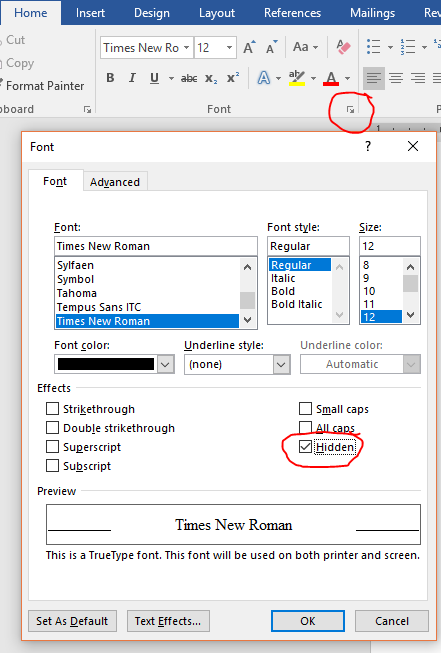
If you have a table that continues onto a subsequent page, you need to start each new page with two rows of information. The first row says “Table X Continued”, and the second row is the Table headers. You can do this by either splitting the table (within Table Tools: Layout tab) or inserting two new rows into the table.

If you have a very long figure or table title, you can truncate it so only the main portion is included in the List of Figures or List of Tables. Insert your title caption following the steps above.

Table 2. Instructions for Truncating Title Captions in the List of Figures and List of Tables

. Here is extra information that I don’t want to show up in my List of Tables.

|  |  |
| --- | --- |
| **Steps** | **Instructions** |
| Step 1 | Turn on your show/hide feature. Put your cursor immediately following the information you want to appear in the List of Figures or List of Tables. |
| Step 2 | Insert a Continuous Break by going to the Layout section, clicking the dropdown arrow next to Breaks, and selecting Continuous Break. |
| Step 3 | Hold down the shift button and arrow over the inserted Continuous Break to highlight it. Make sure to ONLY highlight the Continuous Break. |
| Step 4 | Under the Font dropdown in the Home tab, select Font, and check the Hidden box. |
| Step 5 | Turn off the show/hide feature. The Continuous Break will no longer be visible and the title caption will look normal and continuous. |
| Step 6 | Update the respective list to see the truncated version of your title. |



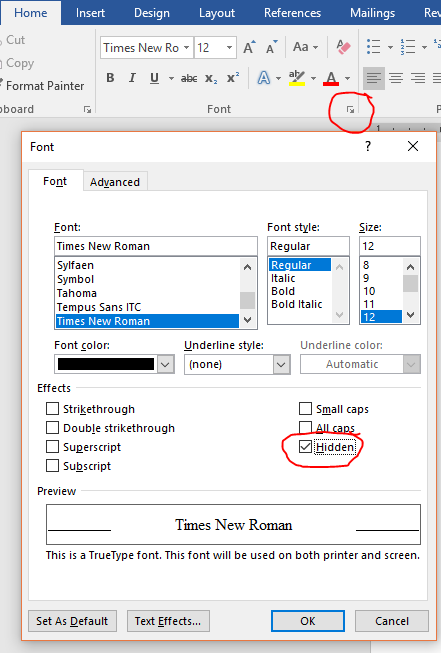


Figure 7. Hiding Text from Figure and Table Captions in the List of Figures or Tables

## Inserting Landscape Pages

You may need to use landscape pages in your dissertation/thesis because you have a figure or table that is too wide to fit on a portrait style page. This can be challenging because it is required to have rotated page numbers on landscape pages such that if the document was printed out and all pages were stacked together in portrait style, all page numbers would align at the top, center of the page. Table 3 provides instructions on creating a landscape page with rotated page numbers, and Table 4 provides alternative instructions for inserting rotated page numbers.

Sometimes, you need to adjust your page numbers following the landscape page with rotated page numbers after you have inserted it into the document. The most common problem is that the page numbers will start over at 1 on the portrait layout page following the landscape page. To fix this, double click and highlight the page number on the portrait page, right click to select Format Page Numbers or choose the Format Page Number option in the Header and Footer menu, and select Continue from Previous Section. Refer to Chapter 2 for information on formatting page numbers on portrait pages. If page numbers appear on the right side of your regular portrait page, you will need to double click the header, deselect Link to Previous Section, then delete the text box with a rotated page number on the right side of your portrait page. If needed, you can then reinsert page numbers on the following portrait pages (see Chapter 2).

One of the formatting rules to keep in mind when inserting a landscape page is that you are not allowed to have more than ½ of a page blank, except at the end of a chapter. Because of this, you may need to rearrange some of your text to fill in the blank space. If you use a Figures and Tables section at the end of a chapter, you are allowed to have each figure or table start on a new page, which may cause more than ½ of a page blank.

Table 3. Instructions for Inserting Landscape Pages with Rotated Page Numbers

|  |  |
| --- | --- |
| **Steps** | **Instructions** |
| Step 1 | Turn on the show/hide feature. |
| Step 2 | Within this template, put your cursor before the section break preceding the landscape page (e.g., after the words “…of a page blank.”), hold shift and arrow over the section break, this table, and the section break following this table. *Note: The section break preceding the landscape page is sometimes displayed as End of Section, rather than Section Break. This is not a problem.* |
| Step 3 | Copy the portion you have highlighted. |
| Step 4 | Paste the highlighted portion in your document where you need to have a landscape page. |
| Step 5 | Delete this table and insert your own landscape content. |
| Step 6 | Remove any excess paragraph markers so that there are no blank pages before, after, or within your landscape page. |
| Step 7 | Turn off your show/hide feature. |
| Step 8 | Format page numbers following the landscape page as needed (see page 13 for more information). |

REFERENCES

[1] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, “Semi-supervised learning with deep generative models,” in Advances in neural information processing systems, 2014, pp. 3581–3589.

[2] F. Berkhahn, R. Keys, W. Ouertani, N. Shetty, and D. Geißler, “One model to rule them all,” arXiv preprint arXiv:1908.03015, 2019.

[3] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[4] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, “Variational inference: A review for statisticians,” Journal of the American Statistical Association, vol. 112, no. 518, pp. 859–877, 2017.

[5] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114, 2013.

[6] Y. Cheng, “Semi-supervised learning for neural machine translation,” in Joint Training for Neural Machine Translation. Springer, 2019, pp. 25–40.

[7] Y.-A. Chung, Y. Wang, W.-N. Hsu, Y. Zhang, and R. Skerry-Ryan, “Semi-supervised training for improving data efficiency in end-to-end speech synthesis,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 6940–6944.

[8] R. Fergus, Y. Weiss, and A. Torralba, “Semi-supervised learning in gigantic image collections,” in Advances in neural information processing systems, 2009, pp. 522–530.

[9] M. Shi and B. Zhang, “Semi-supervised learning improves gene expression-based prediction of cancer recurrence,” Bioinformatics, vol. 27, no. 21, pp. 3017–3023, 2011.

[10] R. Johnson and T. Zhang, “Semi-supervised convolutional neural networks for text categorization via region embedding,” in Advances in neural information processing systems, 2015, pp. 919–927.

# [INSERT APPENDIX TITLE HERE]

If only one appendix is used, call it APPENDIX with no letter or number indicated (e.g., APPENDIX. TITLE)

# [INSERT APPENDIX TITLE HERE]

Use letters or numbers such as Appendix A, Appendix B, etc. or Appendix I, Appendix II, etc. IRB approval letters should be included if approval was required for the study with approval letters and documents not containing signatures or personal information. Appendix information can be single-spaced or double-spaced text.