THE UNIVERSITY OF TEXAS AT AUSTIN

ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

TEXAS
The University of Texas at Austin

Investigation of Transformer-based Methods for Image Compression, Analysis, and Generation

Master's Thesis Report

PRE-SUBMISSION DRAFT

Official code repository:
https://github.com/kliment-slice/thesis

Supervisors:
Prof. Dr. Alan C. Bovik
Prof. Dr. Rachel A. Ward

Author:
Kliment Minchev

April 27, 2022

Contents

List of Figures

List of Tables

Executive Summary

Vision Transformers (ViT) [Bai et al., 2022] and TransGANs[2] are state-of-the-art implementations of the transformer attention mechanism used in generative image tasks. The transformer architecture has outperformed convolutional neural networks (CNNs) in image recognition and object detection due to its ability to generalize well. The attention mechanism is capable of focusing on objects anywhere on the image within a single network layer (versus the variable size convolution kernels across layers). The standard two CNNs combined in a generative adversarial network (GAN), can be replaced with vision transformers to create a TransGAN[2]. Such a model is shown to generalize well and may be capable of handling image and video compression tasks. This thesis project investigates the capacity to which a transformer and potentially a TransGAN is capable of achieving high fidelity video compression in order to mimic the performance of common video codecs.

Chapter 1

General Tools and Techniques for Your Master Thesis

At this point it may be good for the reader to get an overview about this chapter. What does it cover? What is it's purpose in the bigger picture of your thesis? It is like a micro-abstract and it helps your audience to keep on track. This is especially useful, when your thesis is an interdisciplinary work.

1.1   Basic Commands and Writing Topics

1.1.1   Introduction to Writing in LaTex

The text appearance and the thesis layout is sometimes determined by formal requirements. This text, for example is *Times* using 11pt size and a slight increase in space between the text lines. There are also requirements for indenting a new paragraph sometimes.

As you can see, this template does not indent the new paragraph. However, you can always change this setting in the *declarations.tex* file.

Some faculties require to include all references to external sources in *footnotes*.[1][2][3]

You can enumerate or list some points and sketches of a topic using *list or enumerattions*:

1. Start

2. Work

3. Finish

1.1.2   Structure of the Thesis

Since the current template is structured into chapters, sections and subsections. This is mainly because a master thesis may easily grow up to 100 pages and in some cases even bigger. Structuring your big parts into *chapters* makes it perfect for separating in logical units, such as:

- Introduction

- Related Work

- Research Design

- Proceedings and Data Collection

- Data Analysis and Evaluation

- Discussion and Conclusion

---

[1]This is a footnote. You can also include citations and many text formatting LaTex commands.

[2]You may also explain background information on a topic, which may be interesting for further reading.

[3]Footnotes also often include weblinks, for example http://www.jankuester.com

The next smaller units are *sections* which structure a chapter into it's subtopics. Using *subsections* allows to cover these topics step by step and keep the reader on track with a mental model of all the parts of the current topic.

Now there are also *subsubsections* available in LaTex. It would become very confusing when even *including subsubsections in the table of contents (toc)*. For this reason, the toc is set to remove these subsubsections. You may alter this by taking a look into the main.tex and the declaration.tex.

This Subsubsection is not listed in the Table of Contents

The advantage is, that your work becomes more structured along the complexity of your topic. It may occur, that topics can become very granular, which is why this is the perfect counterpart for this issue.

1.1.3   Using a Shorthand Command in Writing

In the *declaration.tex file*, there is a shorthand command called *NOTE*. It is created to quickly emphasize important words and let them point out. It is declared by using the *newcommand* tag. You may read on the web about this command, since it is very handy once you understand how to use it. You can of course create your own commands but you can also stick to this, if you want to get things done fast.

1.1.4   Including a citation

Since we use the natbib environment, we need to use citep or citet as commans for our citations. This may look like the following:

Standard - [Skiena, 2008] [Lombard et al., 2002] [Ben-David and Frank, 2009]

With page - [Ryan, 2010, p.123]

When the authors, e.g. Mayring and Fenzl [2014] said something and you want to embed the citation in your text.

1.2   Including Figures and Tables

1.2.1   Including a Graphic

Including a graphics is not that hard. Important is to wrap it as a figure, so that it will be included into the *list of figures*. You can also *label* it, which allows to refer directly to this figure without the need of keeping track of it's id or name. Another handy parameter is the *width* parameter, which allows to set the dimensions according to the current text width.

Figure 1.1: This is the description of the figure. You may even include citations here, which is often important for a master thesis. For example: [Skiena, 2008]

1.2.2   Including a Table

When declaring a table, you have multiple options to project complex structures. The following example uses *multirows and multicolums* to create a multi-faceted overview of complex data. If you want the table to be listed in the *list of tables*, you need to wrap the tabular within a table declaration.

Category	Levels	Scale		
		low		high
	1. first level	low	middle	high
A	2. second level Subsequence Extensions	none none	some some	many all
	3. presentation	-	text	colored cell
B	4. fourth level			
	graphics	low	medium	high
	sound	mute	normal	loud
	input	mouse	keyboard	gamepad
	5. fifth leves	Low	Medium	High

Table 1.1: Detailed description of the table, which can also use various LaTex commands and include citations.

Chapter 2

Specific Tools for a Thesis in Computer Science

This chapter covers more specific examples, which are useful in the field of computer sciences.

2.1   Writing Pseudo Code in LaTex

The package *algorithm2e* allows you write nearly all imaginable code structures as pseudo code in LaTex. A referece of commands can be found here: http://www.cs.toronto.edu/~frank/Useful/algorithm2e.pdf. Do not forget to wrap the pseudo code into a figure, thus it will be included into list of figures.

The following example demonstrates the package with a bubble sort algorithm. Note, that the directive *repeat - until* is the same as the *do-while* directive.

Algorithm 1: bubbleSort(Array A)

Data: A as Array to sort,

Result: A as sorted Array

1   int n ← A.size ; // cache the initial size of A
2   **repeat**
3   int newn ← 1 **for** *int i=0; i<n-1; i++* **do**
4   **if** *A[i] > A[i+1]* **then**
5   A.swap(i, +1);
6   **end**
7   **end**
8   n ← newn
9   **until** *n>1*;
10   return A

Figure 2.1: PseudoCode f the BubbleSort Algorithm, derived from https://de.wikipedia.org/wiki/Bubblesort.

2.2 Mathematical Equations

Writing mathematical equations can be a very important part of your work. This applies especially, if you want to analyze and evaluate your gathered data. The following example shows a calculation of Kohen's Cappa, a measurement which represents the inter coder reliability in coding qualitative data.

Let's assume the following results of coding a text by two different persons:

result XY		Rater A				
		9	**0**	**1**	**2**	**c(a)**
	9	27	0	0	0	0.5510204082
	0	0	6	0	0	0.1224489796
Rater B	**1**	0	1	7	0	0.1632653061
	2	0	1	7	0	0.1632653061
	c(r)	0.5510204082	0.1632653061	0.2857142857	0	n=49

Table 2.1: A set of example data for further use in a mathematical equation.

Continuing from this basis, the calculation can be proceeded with Cohen's Kappa defined as

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Where $Pr(a)$ is the actual observed agreement of the codes and $Pr(e)$ is the estimated agreement of the codes. The actual agreement can be calculated by building the sum of those codes, which have been assigned by both raters (in the table: diagonal from upper left to lower right). This sum is then divided by the amount of responses (n), which is expressed by the following calculation as

$$Pr(a) = (27 + 6 + 7 + 0)/49 = 0.8163265306$$

The estimated agreement is calculated out of the percentages of each code in relation to the amount of responses (n). The overall estimated agreement is then built by multiplying these chances and summing up their results. The term is described as

$$Pr(e) = \sum_{i=1}^{m} c(r)_i \times c(a)_i$$

where m is the number of all occurring codes, i is the current code and c(r) and c(a) are the chances of assigned codes (see table) for raters and auto. Applying this term is leading to:

$$Pr(e) = (0.5510204082 \times 0.5510204082) + (0,1632653061 \times 0.1224489796) +$$
$$(0.285714285 \times 0.1632653061) + (0 \times 0) = 0.3702623907$$

The values for Pr(a) and Pr(e) are calculated and can be included into the kappa equation, resulting in the following kappa for item a2s1i1a:

$$k = \frac{0.8163265306 - 0.3702623907}{1 - 0.3702623907} = 0.7083333333$$

2.3   Including Code into Your Text

Sometimes you want to include text which include characters that could trigger commands. At this point it useful to wrap them into the *verbatim* environment. The text is uninterpreted and listet as is.

```
2015_5_13/23-19-33:550: ************* LOG INFO  *************
2015_5_13/23-19-33:550: name: user1
2015_5_13/23-19-33:550: [0]: 534
2015_5_13/23-19-33:550: [1]: 321
2015_5_13/23-19-33:550: [2]: 094832
2015_5_13/23-19-33:550: [3]: 3980429804
```

You can also include existing code and highlight it's syntax, using the *lstlisting* package. Look at the declarations.tex file for listings settings. The following example highlights xml syntax by given keywords.

```xml
<!-- an XML comment -->
<entry type="normal" pattern="a" />
<entry type="normal" pattern="b" />
<entry type="formula" pattern="t=a+b+c" required="1"/>
```

Bibliography

Y. Bai, X. Yang, X. Liu, J. Jiang, Y. Wang, X. Ji, and W. Gao. Towards end-to-end image compression and analysis with transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

A. Ben-David and E. Frank. Accuracy of machine learning models versus "hand crafted" expert systems – a credit scoring case study. *Expert Systems with Applications*, 36(3, Part 1):5264 – 5271, 2009. ISSN 0957-4174. doi: http://dx.doi.org/10.1016/j.eswa.2008.06.071. URL http://www.sciencedirect.com/science/article/pii/S0957417408003904.

M. Lombard, J. Snyder-Duch, and C. C. Bracken. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28 (4):587–604, 2002. ISSN 1468-2958. doi: 10.1111/j.1468-2958.2002.tb00826.x. URL http://dx.doi.org/10.1111/j.1468-2958.2002.tb00826.x.

P. Mayring and T. Fenzl. Qualitative inhaltsanalyse. In N. Baur and J. Blasius, editors, *Handbuch Methoden der empirischen Sozialforschung*, pages 543–556. Springer Fachmedien Wiesbaden, 2014. ISBN 978-3-531-17809-7. doi: 10.1007/978-3-531-18939-0_38. URL http://dx.doi. org/10.1007/978-3-531-18939-0_38.

G. W. Ryan. *Analyzing qualitative data: systematic approaches*. Sage, Los Angeles, CA [u.a.], 2010. ISBN 9780761924906. URL http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD? ACT=SRCHA&IKT=8000&TRM=1229209042*.

S. S. Skiena. The algorithm design manual, 2008. URL http://katalog.suub.uni-bremen.de/DB=1/ LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=92542253*.