CSC413 Final Project (20249)

The CSC413 final project involves building a neural network to solve a machine learning problem of your choice, working in groups of 2-4. The project provides both an opportunity for you to complete a deep learning project that is much closer to a real-world application, and also a chance to present this project in a way that is standard in academia. We hope that regardless of your future goals, this project helps to develop your technical, implementation, teamwork and communication skills in a holistic way.

There are two main deliverables for this project:

- [Due date: Nov 6] A **Project Proposal** that describes your problem domain choice, goals, data, and approach. This project proposal requires quite a bit of thinking, planning, research, and exploratory data analysis. It is to be written in the style of half of an academic paper (4-6 pages). Writing parts of the paper before conducting the research is a way to clarify your ideas in a concrete, visible, and understandable way, so that these ideas can be discussed meaningfully (e.g. https://www.cs.jhu.edu/~jason/advice/write-the-paper-first.html and https://www.microsoft.com/en-us/research/academic-program/write-great-research-paper/ —though the latter is not ML focused). You will receive TA feedback on both your ideas/approach, and also your writing.
- [Due date: Dec 5] A **Final Project Report** in the style of an academic paper. We will be following the style used in a major machine learning conference called NeurIPS. The final report has many of the same sections as your proposal, but also describes your final model and assessment of its performance. Since you will receive detailed writing feedback for your project proposal, you are expected to apply the feedback to write your final report.

Project Ideas

Your project must have one of the following properties:

- It takes a sequence (of variable length) as an input, or produces a sequence as an output, or both.
- It takes a video or other 3D structure as an input or output (but note that video models take a long time to train and require GPUs with large memory)
- It generates an image or other 2D structure.
- It takes as input a complex structure (e.g., a representation of a molecule or protein) or generates a complex structure as an output.

In other words, the input or output of your model should have a level of complexity beyond the labs.

We recommend that you decide on your project idea based on public datasets that are available. While data collection and cleaning is an important skill, generating your own data is extremely time-consuming and does not align with the main learning objectives for this assignment.

There are many places to find public datasets for machine learning. Here are a few suggestions, but feel free to search for your own:

- https://kaggle.com/datasets/
- https://www.openml.org/search?type=task&sort=runs
- https://archive.ics.uci.edu/datasets
- https://data.ontario.ca/ and other government open data initiatives

Examples of past projects include:

- Generating a “Finger DDR” game level for a song
- Generating a script for an episode of the TV show *South Park*
- Identifying potential binding sites of in a protein

Paper Sections and Formatting

You can read more about the NeurIPS style guide and obtain the LaTeX template to use here:

https://nips.cc/Conferences/2023/CallForPapers

https://neurips.cc/Conferences/2023/PaperInformation/StyleFiles

The table below describes the major sections in your project proposal and final report. Your exact section headings may differ depending on your topic, but the information described should be easy to find.

Section	**Description**	**Project Proposal**	**Final Report**
Abstract	A short description of your goals, task, model, and (for the final report) results. The abstract should make the motivations and the scope of your project clear so that readers can decide whether they are interested in reading your work.	X	X
Introduction	A description of the motivation behind your work, why the task you chose is interesting/important, and a summary of your (proposed) approach. The problem that you want to solve should be clearly stated in the introduction: especially the input and output of your model and the format of the input and output. This section should also make it clear why your deep learning approach is reasonable for this problem.	X	X
Background and Related Work	A summary of the background material that students of CSC413 would not already be familiar with. A description of related work done in the area, and how your approach compares with theirs. If your project builds on previous work, clearly distinguish what they did from what your new contribution is. Also, include a 1-2 sentence summary of other closely related papers. We realize you might not know about all related papers (or have time to carefully read all related papers), and that's OK for this project. Using bibtex is annoying at first, but Google Scholar can give you the bibtex entries.	X	X
Data	The dataset used in your model. Include any key exploratory figures that will help readers evaluate the difficulty of your problem and interpret the performance of your model.	X	X
Model Architecture	A description of your (proposed) model architecture. Please propose an architecture during the proposal phase, but it's okay to change your architecture. In the final report, this section should have enough details to reproduce the work, including all hyperparameters and	X	X

	training settings that you used.		
Model Architecture Figure	A figure that helps show the overall model or idea. The idea is to make your paper more accessible, especially to readers who are starting by skimming your paper. You must create a new figure, not just use someone else's, even with attribution. Be careful that all figure text are legible, and are approximately the same size as the main text.	X	X
Results	Describe the performance of your model, either quantitatively or (for generative models) qualitatively. To help interpret the model, compare your model with a baseline model. Qualitative evaluation is OK.		x
Discussion	Interpret the results of your model. If appropriate, compare the model performance with the baseline. Discuss other features that you notice about the model.		x
Limitations	Describe some settings in which we'd expect your approach to perform poorly, or where all existing models fail. Try to guess or explain why these limitations are the way they are. Give some examples of possible extensions, ways to address these limitations or open problems		X
Ethical Considerations	Potential ethical issues posed by the use or misuse of your model. Your report should transparently communicate the known or anticipated consequences of building and using machine learning models on this task. https://neurips.cc/public/EthicsGuidelines	X	X
Work Division	A description of how the work will be divided between the team members, and how the team members will be working together (e.g. meet every week Tuesday 4-5 pm).	X	
Conclusion	State the results achieved in relation to the problem described in the introduction. Repeat the main takeaways from your paper		X

Project Proposal Rubric

The project proposal document should be 4-6 pages and use the NeurIPS LaTeX template. It is graded out of 40 points.

Section	**Points**	**Criteria for full credit**
Abstract	4	Concise summary that describes your goals, task and model. The abstract length is appropriate (about half a page), and provides enough information about your data, task and approach for a potential reader TA to decide whether they have the expertise to review your work. The abstract is often used to assign reviewers in real conferences, and we may do the same to assign TAs.
Introduction	4	Clear explanation of the “what”, “why”, and “so what” of your project: *what* is your project about? *Why* is it important? *Why* is that meaningful? This section should make your general approach clear, as well as convince the read that your deep learning approach is reasonable (worth trying out) for this problem.
Background and Related Work	4	Background and related work material should provide TAs and other CSC413 students with enough information to understand the rest of your paper. This section should also include the description of at least 1-2 prior works related to your project to put your project into context. Important work chosen should be relevant to your particular domain, and not just the general method. For example, if you are working on using transformers to generate video game music, citing a paper that introduces transformers is not sufficient! Instead, cite other papers that work on similar tasks related to similar types of music generation. Likewise, background information includes information necessary to understand your work. Using the video game music example, if you are using a particular type of transformer not discussed in class, please describe it briefly and cite a source. If you are using ideas from music theory that a CSC413 TA would not be familiar with, briefly describe that as well.
Data	4	A clear description of the data and reproducible steps to clean and format your data. The description should provide the reader a sense of the difficulty of the problem, and help the reader understand whether your model architecture is appropriate given the nature of your data. Your description should convince the TA that your group thoroughly explored the features, and can identify and communicate key information to the reader.
Model Architecture	4	Rough description of the type(s) of neural network model(s) that you will use, and the relevant components. For full credit, the TAs should be able to understand why you chose this architecture.
Model Architecture Figure	2	A well-thought-out figure that communicates the core idea of your project and architecture immediately. Text are legible, and are

		approximately the same size as the main text. The clarity and communicative power of the figure are key: an overly complex figure may lack clarity, and an overly simple figure might not communicate much.
Ethical Considerations	2	Thoughtful consideration of ethical issues in data collection, and the impact of using or misusing the model. We are looking for depth of thought beyond what may be “obvious”.
Work Division	4	The division of work is well thought out and fair. There should be enough details provided so that if a new person is to replace a member of your team, they should be able to understand their responsibilities, be able to estimate their time commitment during each week, and identify conflicts.
Writing: Structure & Organization	4	Sections and subsections are ordered appropriately. Content in each section/subsection matches the header. Within each section, information is ordered in a coherent way. Paragraphs have topic sentences and are easy to skim. Figures and tables are labeled with appropriate captions. All figures and tables are referenced in the main text. Figures and tables are readable, with font sizes similar to the body. Follows the NeurIPS structure, including spacing and formatting guides.
Writing: Writing Mechanics	4	Sentences are written in a clear and concise way. Sentences are complete and are not too long or complex. The verbs used are descriptive and precise, rather than generic. Symbols are generally annotated with the type of entity they represent. There are no large amounts of distracting typos and/or grammatical issues.
Writing: Audience Expectations	4	Academic language is used, without slang, impersonal language (“we” is fine, but “I” and “you” are not), and gendered language. Assumed technical knowledge is appropriate for the reader. Terms unfamiliar to a typical CSC413 student are defined when first used. Concepts unfamiliar to a typical CSC413 student are explained. Figures that contain text are accessible: i.e. images with text are avoided, and captions are descriptive and can be understood by readers with visual impairment.

Final Report Rubric

The final report is graded out of 40 points. The report should be written so that it is exactly 8 pages and use the NeurIPS LaTeX template. You may use additional pages for references only.

Section	Points	Criteria for full credit
Abstract	4	Same criteria as the proposal, with feedback from the proposal addressed.
Introduction	4	Same criteria as the proposal, with feedback from the proposal addressed.
Background and Related Work	4	Same criteria as the proposal, with feedback from the proposal addressed. Additionally, any related work materials that the TA suggested should be added to this section and cited.
Data	4	Same criteria as the proposal, with feedback from the proposal addressed.
Model Architecture	4	Clear description of your model architecture. There should be enough details to reproduce your work, including all hyperparameters and training settings that you used. If you are presenting the result for multiple models, describe their architectures and clearly state their differences.
Model Architecture Figure	2	Same criteria as the proposal, with feedback from the proposal addressed. Your model may have changed since the proposal, so update your figure accordingly.
Results	4	Clear description of the performance of your model, either quantitatively or (for generative models) qualitatively. The choice of performance metric should be justified and should be appropriate for your task. There should be enough information provided to the readers so that they can independently interpret the model results. To help interpret the model, compare your model with a reasonable baseline model. This baseline model should be defined clearly and justified convincingly.
Discussion	4	Sound interpretation of your model performance. The comparison with the baseline and other comparisons should be thoughtful and convincing. Discuss what you notice about the model. We are looking for a thoughtful reflection of the model and its performance along various dimensions.
Limitations	2	Same criteria as the proposal, with feedback from the proposal addressed.
Ethical Considerations	2	Same criteria as the proposal, with feedback from the proposal addressed.
Conclusion	2	Clear summary of the results achieved in relation to the problems. The description is consistent with the rest of the paper.
Writing Revision	4	Substantial revision of the writing issues identified by the TAs in the project proposal. To earn full credit, correct the majority of instances of the issue identified by the TAs (for example, if the TA identified that you have many terms that are not defined, adding the definitions of just the ones the TA explicitly highlighted is not sufficient. Instead, find *all instances* of this issue in your work.)

Advice

Getting Started

One major misconception about academic writing is that papers are written in the same order that they are read. In fact, the introduction and abstract are often some of the last sections of a paper that a researcher will write. For your project proposal, we suggest that you work on your project proposal in the following order:

1. Start by exploring the various datasets available and choose a dataset that matches your interests. Since finding and collecting data can be time-consuming, we suggest that you choose your project based on what datasets are available, so that you do not spend a lot of effort on data collection. As you decide on the dataset, perform exploratory analysis **and write the Data section of your proposal.** (See below on the “Iterative Writing Process”; do not aim to write full sentences at this point—jot notes are perfectly fine.)
2. Once you have a clear idea of the data that you are working with, search for related work that uses that data set, or a similar dataset. **Use this research to write the Background and Related Works section.**
3. Based on your research, propose a model architecture. **Write the model architecture section and draw your figure.** This is also a great time to **revise the Related Works section with a description of how your model differs from existing work.**
4. Consider the work that must be done to build this model, and a baseline model to compare with. What work is required? Use this information to write the **Work Division** section.
5. Consider the limitations and potential ethical issues arising from your approach. **Write the limitations and ethical considerations sections.**
6. Now that you have an idea of what the project will be about, **write the introduction.** Take care to consider the needs of the reader: although you are now familiar with your project, the introduction is one of the first sections that your reader will see. Clearly explain why your task is important, and what that task really is. The motivation piece is extremely important!
7. Finally, **write your abstract**. You might start by piecing together key sentences from your introduction, data, and model architecture section so that the necessary information is in one place. Then, work on refining these sentences so that they flow well.

As you write each section, you may wish to go back to previous sections and make revisions. This is a great sign. But be careful that you are also making progress: starting a new section is hard and revising is typically easier. When you are first writing, start by getting all of the information on the page. Once all the information is there, then you can think more carefully about structure: what are your paragraphs going to be? How will you introduce information in a way that is clear to the reader? Are all your terms defined? Are your sentences clear? What other assumptions are you making about the reader?

Iterative Writing Process

Before starting to write, read several NeurIPS papers to get a feel for the writing style. Ideally, your paper should have a similar sentence structure, level of graphics polishing, and reading flow. Trying to achieve all of this by writing from scratch is very hard.

Counterintuitively, it is faster to create an initial rough draft and iterate over it multiple times rather than trying to get it right from the beginning. One approach that may be helpful is to first create bullet points, put all the equations and empty placeholders for tables and figures into the paper. Each of the bullet points will be representing a one-sentence summary of a full paragraph that you want to write. In the next step, the bullet points can be enriched by adding more points with the content that goes into each paragraph. This leaves the worries of finding the best formulation till the end making text creation much easier as all content is there.

Similarly, for figures, it makes sense to create a pen & paper / tablet draft and only create a polished figure once the draft is in a good state and conveys the content that you want it to convey. This also will save a lot of time as editing a drawing is much faster than editing a diagram in your paper.

Details for Reproducibility

Your paper should have enough details for *reproducibility*. That is, another student in CSC413 should be able to read your paper and be able to take the same steps you did to achieve similar results (and produce similar figures!). Reproducibility means that the following pieces of information are important to include:

- How did you format your data? For example, did you use one-hot embeddings or treat a variable as a numeric value? Did you normalize your features?
- What data did you include/exclude?
- What is your model architecture?
- What hyperparameters did you use?
- What were your other model settings? Why did you choose these?

However, your paper should *not* contain excessive details that are *not required for reproducibility*. That is, if someone else could have used a different approach to reproduce your work, then that description is likely not useful. For example:

- The variable name that you used to represent your data. Someone else could reproduce your work and choose a different variable name.
- The specific piece of code you used to visualize your results. Someone else could have written a slightly different piece of code but still obtain the same results.

The exception is that researchers typically report the tools that they used (e.g. PyTorch, pandas, numpy, etc). This is partially to give credit and attribution to the developers and maintainers of these tools, but also to acknowledge that implementation differences may cause a difference in the model outcome.

Captioning and Referencing Figures and Tables

All figures and tables should be captioned, and should additionally be referenced in the main text. For more information about writing figure captions, see
https://www.internationalscienceediting.com/how-to-write-a-figure-caption/

Remember that images are less accessible than text: they cannot be easily read by screen readers. Thus, captions are important for people with visual impairments or who otherwise use a screen reader.

Referencing figures and tables in your text does not need to be an onerous affair. Here are some examples:

> The overall model architecture is illustrated in Figure 1.
>
> Figure 2 presents the results on unsupervised object segmentation.
>
> Figure 3 shows a qualitative result of video segmentation.
>
> Table 1 presents the results on the visual quality of the generated videos.

These references typically appear at the beginning or end of a sentence.

Typically in NeurIPS and other machine learning and computer science conferences, figures and tables should be floating at the top of the page. Use the \begin{figure}[t] and \begin{table}[t] settings on LaTeX, rather than [h]. Thus, your figure or table will *not* appear exactly where you have referenced it in the text, and that's okay.

Citing Related Work

Claims that you make in your paper should be supported by evidence. This can be evidence that you produce in your work, or by citing external sources.

Your paper should follow the norms for citing related work. Please use the LaTeX bibtex or natbib package to track your citations for you. These packages will generate the references for you, but you will need to include an in-text citation for every related work item you wish to have appear in the references.

For more information about when and why to cite, start here:
- https://guides.library.utoronto.ca/citing
- https://guides.library.utoronto.ca/c.php?g=251103&p=1673069

There are two types of in-text citations: narrative citation (where the work cited is directly referenced in the text and used as a noun in your sentence) and parenthetical citations (where the work cited is referenced as evidence to support a claim, but the work cited is not a part of your sentence).

Narrative citations are used to highlight work that is especially relevant to your work. For a narrative citation, you can use the \citet{...} command to generate the citation text, for example:

> \citet{yoshikawa2024} uses language models to formulate robot-executing plans from natural language instruction, and uses their techniques on solubility and recrystallization experiments for material synthesis.

> Large language models have been used in the combination of robotics and chemistry, for example by \citet{yoshikawa2024}, who uses their techniques on solubility and recrystallization experiments for material synthesis.

Parenthetical citations are used to cite claims without using a significant amount of space, and is the more common form of in-text citation in machine learning papers. To generate this type of citations, use the LaTeX command \cite{...}. You can also put multiple pieces of work in a single parenthetical citation, for example like this: \cite{yoshikawa2024, wu2023}. It is typical to put a "~" before the citation, so that LaTeX does not put a newline immediately before the citation (e.g., "Machine learning is useful~\cite{yoshikawa2024, wu2023}").

As an example for how parenthetical citations are used, let's look at the citations in the Introduction and Methods section of the paper "**SlotDiffusion: Object-Centric Generative Modeling with Diffusion Models"**, available here: https://proceedings.neurips.cc/paper_files/paper/2023/hash/9fa03b16dbd6cabc7601fe98c6ec291e-Abstract-Conference.html

1 Introduction

Humans perceive the world by identifying discrete concepts such as objects and events [88], which serve as intermediate representations to support high-level reasoning and systematic generalization of intelligence [24]. In contrast, modern deep learning models typically represent visual data with patch-based features [14, 28], disregarding the compositional structure of scenes. Inspired by the human perception system, object-centric learning aims to discover the modular and causal structure of visual inputs. This approach thus holds the potential to improve the generalizability and interpretability of AI algorithms [50, 80]. For example, explicitly decomposing scenes into conceptual entities facilitates visual reasoning [8, 12, 13] and causal inference [80, 109]. Also, capturing the compositional structure of the world proves beneficial for both image generation [83, 90] and video prediction [100, 103].

Notice that parenthetical citations above are used to provide evidence for claims (e.g., for [88], [24]). Some of the evidence is in the form of examples of prior work that do certain things (e.g., [50, 80], [8, 12, 13], etc). These citations are placed *after* the information that you are citing.

Diffusion models (DMs) [34, 87] have recently achieved tremendous progress in image generation [11, 64, 71, 77], showing their great capability in sample quality and input conditioning. The generative process of DMs is formulated as an iterative denoising procedure from Gaussian noise to clean data, where the denoiser is typically implemented as a U-Net [74]. However, the memory and computation requirements of DMs scale quadratically with the input resolution due to the self-attention layers in the U-Net. To reduce the training cost, LDM [73] proposes to first down-sample images to feature maps with a pre-trained VAE encoder, and then run the diffusion process in this low-resolution latent space. LDM also introduces cross-attention as a flexible mechanism for conditional generation. For example, text-guided LDMs [73] perform cross-attention between text embeddings and U-Net's feature maps at multiple resolutions to guide the denoising process. In this work, we adopt LDM as the slot decoder due to its strong generation capacity, where the conditioning is achieved by cross-attention between the denoising feature maps and the object slots.

In the methods section, we see similar kinds of parenthetical citations. In addition, the first time that a new model is mentioned (e.g., "U-Net"), the citation for that model is provided. We do not need to repeatedly provide the citation when it is clear which model/paper we are referring to.

Finally, here's another paragraph from the methods section where the authors use narrative citation to highlight work that is especially relevant. This also helps to give credit to authors of closely related work.

> **Concurrent work.** Recently, Jiang et al. [41] also employ LDM in object-centric learning. However, they utilize an image tokenizer pre-trained on external datasets, while all components of our model are trained purely on target datasets. Moreover, their work only evaluates on images, while we test SlotDiffusion on both images and videos. In addition, we combine our model with recent development in object-centric learning, showing our potential in dynamics modeling and handling real-world data.

Related work should *not* be discussed in a disparaging manner. It is okay to discuss differences between methods, but it is important to do so respectfully.

Paper and Section Length

Your paper should be written so that it is exactly 8 pages. You may use additional pages for references only.

The abstract and introduction should take about 1.5 pages. The background and related work section should be at most 1 page. The data section should be about 1-1.5 pages. The model architecture should also be about 1 page, including the figure. The results and discussions should take about 2-3 pages. The limitations and ethical considerations are about half a page, and the conclusion should be at most half a page.

Deviation from this is certainly possible: for example, if your data is particularly complex, it may require more explanation than the 1-1.5 page guideline. You shouldn't see these guidelines as rules: good writing practices may mean that some sections may be longer/shorter, and you may receive writing feedback in that regard. However, a large deviation without good reason is usually a sign that a part of the report is weak or lacks sufficient detail for reproduction.

Common Issues from Last Year

These are the common issues that we saw in last year's reports:

- [ ] Referencing figures: Are all of your figures captioned, and referenced in the main text?
- [ ] Citations: Are all of your claims cited? Are you citing the models that you reference? Are all of your citations relevant?
- [ ] Background Context: Are you providing the background context to your readers, particularly context related to your domain?

- [ ] Structure: Are your sections structured appropriately, with similar content grouped in a single section? Are your section sizes appropriate? Does your content match the section that it is under?
- [ ] Acronyms: Do you define your acronyms and other technical terms the first time they are introduced?
- [ ] Proofreading: Are your punctuations correct? Are your sentence lengths appropriate?