Team

Theophilus Pedapolu, ID: 3035826494
Kevin Wang, ID: 3035606716
Timber Lin, ID: 3035271227

Part 0: Abstract

We created a self-contained homework set that guides students through understanding the Conformer, a novel deep-learning architecture that combines convolutional layers with transformers to capture both the local and global dependencies of an audio sequence. It has shown state-of-the-art results on automated speech recognition (ASR) tasks. We guide students through analytical questions that help them understand depthwise-separable convolutions and the swish activation question, key components of the Conformer block, and then ask them to implement and combine them with other components to implement the paper's Conformer architecture in the first Colab notebook. Finally, in the second Colab notebook, we guide students through training a Conformer on a simple audio classification task, including data cleaning/formatting and hyperparameter tuning. Students also analyze the plots and figures of various ablation studies run on the Conformer to understand how each component of the architecture contributes to the accuracy.

Parts I & II: Review and Point-to-Point Response

Reviewer #1

Questions

1. Question 1: Content and Correctness (Option 1). If this paper is not Option 1, write NA.

Both the coding and analytical portions are easy enough to follow and complete, but also not completely trivial and do require some deep thinking. All of the mathematical concepts and the connections to content we learned in class (ReLU, Convolutions, etc) are also all correct. The solutions provided are detailed, correct and easy to follow. The code runs very smoothly however, in Conformer_ASR_Q4.ipynb Question three didn't yield any plots. I wasn't able to attach the result I received after running the code without making any changes to the notebook, but it was essentially a blank plot with just the axis.

The homework is on the more lengthy end of 2 hours. The analytical portion took me 45 minutes to complete even after having read the solutions. The Conformers_Q3.ipynb notebook took me 40 mins to complete and the Conformer_ASR_Q4.ipynb took at least an hour to engage with fully, and I anticipate it would have taken longer without having already read the solutions.

We expected the analytical part to take 25-30 minutes total for both questions, the coding portions to take 40-45 minutes, and question 4 to take 30 minutes. However, we added more introductory information to question 3 and question 4 explaining the conformer better so that should allow students to spend less time trying to understand these questions.

In Conformer_ASR_Q4.ipynb, the listening of the data cell, only yielded one audio block, and it definitely did not sound like either Marvin or Visual, perhaps the seed needs to be set? I hear the words "zero".

We split these audio cells into 2 so they yield two audio blocks, which should sound as directed to the user. This was a simple bug in the code

Although we train the model there is no validation or test set that is used to show

whether the model actually works or not. It is also unclear what the model is outputting here without reading the paper itself, so it was difficult to tell what I actually did by building and training the model. If an example of the output of the model could be provided, that would be helpful

.

Good suggestion. We didn't have enough time to add these parts during our first draft but for the final submission, we added more scaffolding to explain what the model is outputting as well as some accuracy plots of the model run on the test set.

Conformers_Q3.ipynb is a great notebook that gives us hand on experience in building the blocks. But we never really get to use them, is there a way to make it clear the code we wrote to build the blocks are used in the training? Otherwise it feels quite un impactful to just code the blocks to pass the assertions, but not actually use them in the training of the conformer model itself.

This was something that we were hoping to do, and the students implement a complete conformer block as described in the paper. However, when we tried to use our own implementation, it just performed worse than the torchaudio implementation and we were unable to successfully debug the issue. So, for the sake of having useful results in the ablation studies, we chose not to use the code we wrote.

2. Question 1: Literature Background/Problem Formulation (Option 2). If this paper is not Option 2, write NA.

NA

3. Question 1 - grade

Medium improvement needed

4. Question 2: Scaffolding (Option 1). If this paper is not Option 1, write NA.

There was super clear scaffolding. I thought it was very well done and everything was self contained with the appropriate references for external sources to be used by students who are curious to learn more (e.g. MFCC feature transform). The sanity check in Conformers_Q3.ipynb were also great to see.

For Question 4 (a) (i) it would have been nice to have a #tuning parameters block in the notebook, which contains the variables that you want us to change around, instead of scrolling and command+F to find that variable name and changing it. Since we have to experiment with many different values, it would have been useful and help us complete the homework faster.

We cleaned up the code as suggested and added a #TODO block in Question 4(a)(i) so students have clear directions on where to look to tune the hyperparameters. This is a good suggestion to make directions clearer.

5. Question 2: Technical Approach/Deep Learning Techniques (Option 2). If this paper is not Option 2, write NA.

NA

6. Question 2 - grade

Small improvement needed

7. Question 3: Readability/Clarity (Option 1). If this paper is not Option 1, write NA.

Everything was easy to follow, clear and well structured and well written. With that being said, the homework question were pretty self contained, particularly the analytical questions.

What could have helped with the flow of the homework was more clear motivation. Maybe a few sentences at the beginning of the analytical assignment to just give a high level understanding of the conformer, what it does (i.e. one sentence on what an audio recognition task is) and why it is important just to motivate the entire homework. I would have liked to know what it was in particular what about the Depthwise-Separable Convolutions, that were needed or worked well for audio speech tasks. I was able to learn a lot about their properties, but what is it about those properties help us with an audio recognition task.

Good points. We were hoping that students would skim the paper to understand the motivation, but we understand that it's still important to provide that ourselves. We have added more context to conformers in the Conformers_Q3.ipynb notebook.

8. Question 3: Experiments and Contributions (Option 2). If this paper is not Option 2, write NA.

NA

9. Question 3 - grade

Small improvement needed

10. Question 4: Commentary on HW (Option 1). If this paper is not Option 1, write NA.

The commentary itself was great. It explained the main concepts of the paper along with the learning objectives that they tried to teach students. Focus on the convolutions, architecture and and particularly the activation function was great to see. Analysis/Ablation Studies in the commentary were well explained however I am unsure it translates perfectly in the homework, since the different studies didn't show

any drastic change in the training ability of the model. The loss maybe differed by +/- 0.1, and the differences were unobservable from the plotted loss functions.

We tried training the conformer encoder on various decoder architectures, such as a single-layer LSTM cell (as in the paper), an MLP layer, and convolutonal layers but because conformers take a long time to train well, we were unable to produce plots that differed significantly within 5-10 minutes of training. One change, however, is that we added some figures from the original paper for students to reference in case the plots aren't distinguishable enough to answer the conceptual questions.

11. **Question 4: Code and Datasets (Option 2). If this paper is not Option 2, write NA.**

    NA

12. **Question 4 - grade**

    Small improvement needed

13. **Question 5: Going above and beyond (Option 1). If this paper is not Option 1, write NA.**

    I don't think there was anything novel introduced in the homework that wasn't done in the paper, but I do award the creative analytical problems offered and the effort in simplifying the model for the coding portion. Also the additional systematic experimentation was nice to see.

14. **Question 5: Readability/Clarity (Option 2). If this paper is not Option 2, write NA.**

    NA

15. **Question 5 - grade**

    Excellent work, no actions needed.

Reviewer #3

Questions

1. **Question 1: Content and Correctness (Option 1). If this paper is not Option 1, write NA.**

   For the question that asks the student to implement the FeedForwardModule class, the provided solution code does not pass the test case for this question on my computer, but it does pass on Google Colab (not sure why this is the case, but it might be an issue for some students). I also noticed the same issue on the overall ConformerBlock test cases. Despite the same code, the test case fails if I run locally on my computer, but passes if running on Google Colab. I assume these should not be an issue if everyone is running on Colab, but just wanted to point that out.

   I could not reproduce this on my own computer. Make sure you have the latest versions of the libraries?

   Otherwise, I think the concepts covered in this homework are well-chosen, encompassing both theory and implementation, and give the student a good understanding of the underlying structure of the conformer architecture, while not being too pedantic or lengthy. Solutions for the written and coding questions are correct and are easy to follow.

2. **Question 1: Literature Background/Problem Formulation (Option 2). If this paper is not Option 2, write NA.**

   NA

3. **Question 1 - grade**

   Small improvement needed

4. **Question 2: Scaffolding (Option 1). If this paper is not Option 1, write NA.**

   For the question where the student implements the FeedForwardModule class, it is unclear why the LayerNorm is not being performed (despite it being one of the components shown in the diagram). This part was the only thing I found confusing in the scaffolding.

   LayerNorm is given as a part of ConvModule and is expected to be included as part of the ConformerBlock implementation, as per the diagram. Unsure how to address this one.

Everything else was well-designed and allows the student to successfully explore the conformer architecture with the knowledge and mathematical background they would have as a student in this class.

5. Question 2: Technical Approach/Deep Learning Techniques (Option 2). If this paper is not Option 2, write NA.

NA

6. Question 2 - grade

Small improvement needed

7. Question 3: Readability/Clarity (Option 1). If this paper is not Option 1, write NA.

There isn't any non-standard mathematical notation, and the homework is easy to read and understand. There was one slight issue in the audio playback for Q4, in the cell that has the student listen to 2 different audio clips. When I run the cell, I only see 1 player, so I would have to comment out the code to load the second audio file to get the player for the first audio file. Not a huge issue, but just something that students may notice.

Otherwise, everything else looks good.

We split these audio cells into 2 so they yield two audio blocks, which should sound as directed to the user. This was a simple bug in the code

8. Question 3: Experiments and Contributions (Option 2). If this paper is not Option 2, write NA.

NA

9. Question 3 - grade

Small improvement needed

10. Question 4: Commentary on HW (Option 1). If this paper is not Option 1, write NA.

The commentary is clear and concise, and it explains the fundamental components of the paper and how they tie in to the questions in this homework assignment. The homework achieves the specified learning goals.

11. Question 4: Code and Datasets (Option 2). If this paper is not Option 2, write NA.

NA

12. Question 4 - grade

Excellent work, no actions needed.

13. Question 5: Going above and beyond (Option 1). If this paper is not Option 1, write NA.

From my understanding, the way this homework goes above and beyond the paper is by performing different ablation studies to better understand how various model hyper-parameters influence the overall performance. This is useful because it allows students to see how the model can be improved beyond what was provided in the paper.

14. **Question 5: Readability/Clarity (Option 2). If this paper is not Option 2, write NA.**

    NA

15. **Question 5 - grade**

    Excellent work, no actions needed.

Reviewer #4

Questions

1. Question 1: Content and Correctness (Option 1). If this paper is not Option 1, write NA.

Coding solutions run smoothly. Both coding and mathematical difficulties are similar to real problem sets. Questions are related to important concepts of the Conformer model. The length of the problems other than wait-time for the code to run might be slightly less than desired.

We expected the analytical part to take 25-30 minutes total for both questions, the coding portions to take 40-45 minutes, and question 4 to take 30 minutes. However, we added more introductory information to question 3 and question 4 explaining the conformer better so that should allow students to spend less time trying to understand these questions.

2. Question 1: Literature Background/Problem Formulation (Option 2). If this paper is not Option 2, write NA.

N/A

3. Question 1 - grade

Excellent work, no actions needed.

4. Question 2: Scaffolding (Option 1). If this paper is not Option 1, write NA.

Scaffolding of the model is very detailed and clear. Questions and topics are well chosen. One issue I had is that from the order of questions and the text portion of both problem set and coding notebook, it is hard to notice the central idea of the work I am doing without reading the paper first. Slightly more explanations and highlighting the key concepts' relation to the model might be beneficial (probably as a head-up in Q3 notebook).

As per the suggestion, we added a more in-depth explanation of the Conformer in the notebook for Question 3 to better transition from the paper's ideas to the concepts in our homework problems.

5. Question 2: Technical Approach/Deep Learning Techniques (Option 2). If this paper is not Option 2, write NA.

N/A

6. Question 2 - grade

Small improvement needed

7. Question 3: Readability/Clarity (Option 1). If this paper is not Option 1, write NA.

The instructions read clear and easy to follow.

8. Question 3: Experiments and Contributions (Option 2). If this paper is not Option 2, write NA.

N/A

9. Question 3 - grade

Excellent work, no actions needed.

10. Question 4: Commentary on HW (Option 1). If this paper is not Option 1, write NA.

The commentary is very clear and helpful. It might be better to discuss "Activation Functions" before "Conformer Architecture" which would follow the order of the problems as well as the scale of view towards the model.

In the architecture described in the paper, the swish activation function is inserted between the depthwise and pointwise convolutions, so we feel it is acceptable to have the convolutions analytical problem appear first.

11. Question 4: Code and Datasets (Option 2). If this paper is not Option 2, write NA.

N/A

12. Question 4 - grade

Small improvement needed

13. Question 5: Going above and beyond (Option 1). If this paper is not Option 1, write NA.

The work simplifies from the big model of the paper but keeps essential concepts and the advantages and creativity of the Conformer model. Also good visualization and analytical questions are provided in Q4.

14. Question 5: Readability/Clarity (Option 2). If this paper is not Option 2, write NA.

N/A

15. Question 5 - grade

Excellent work, no actions needed.

Part III: Final Submission

Our final submission including the following documents:

- Conformers_Homework PDF (found in Gradescope submission)
- Conformers_Homework_Solutions PDF (found in Gradescope submission)
- homework.tex (LaTeX source code for homework pdf, found in Gradescope submission)
- solutions.tex (LaTeX source code for solutions pdf, found in Gradescope submission)
- Conformers_Q4.ipynb (jupyter notebook containing the code for question 3 of homework, found in Gradescope submission). The corresponding CoLab notebook is linked in the supplementary section
- Conformer_ASR_Q4.ipynb (jupyter notebook containing the code for question 4 of homework, found in Gradescope submission). The corresponding CoLab notebook is linked in the supplementary section
- Conformers_HW_Commentary PDF (found in Gradescope submission)

Part IV: Team Member Contributions

Kevin Wang
Contributed to the original project proposal including the overall structure of the homework. Wrote all the problems and solutions for question 3 of the homework (code implementation of the Conformer block) which includes Conformers background, implementation of pointwise and depthwise convolutions, swish and GLU activations, and convolution and feed-forward modules. Proofread questions 1 and 2 and responded to comments from reviewers. Reviewed and revised abstract as well as other parts of the final submission materials.

Timber Lin
Contributed to the original proposal for the project and wrote the background, problems, and solutions for question 2 (Understanding the Swish Activation Function) on the homework. This involved comparisons of the benefits and drawbacks of swish in relation to other activation functions. Reviewed question 1 and the implementation for questions 3 and 4. Addressed and incorporated reviewer feedback.

Theo Pedapolu
Wrote the original proposal for the project and wrote question 1 (depthwise-separable convolutions) and question 4 (CoLab notebook training a conformer and running ablation studies) of the homework as well the solutions for these questions. Also wrote the background for depthwise-separable convolutions in the homework and all the code for the notebook in question 4. In particular, he experimented with the SpeechCommands dataset and found a way to format it by transforming audio sequences into a Mel Spectrogram so it could be trained on a torchaudio Conformer and wrote the code to plot losses and training accuracies for various ablation studies. Additionally wrote the abstract and the sections about convolution and ablation studies in the homework commentary

Part V: Supplementary Materials

- CoLab Notebook for Question 4: Conformer_ASR
  https://colab.research.google.com/drive/1h_ML3mGcVd3opaa2ctxL3ZnuI2LV1uyz?usp=sharing
- CoLab Notebook for Question 3: Conformer
  https://colab.research.google.com/drive/1r5xv6N49ooYWbXL95672rZyXyFMNd3jL?usp=sharing
- SpeechCommands Dataset used to train Conformer in Question 4:
  https://pytorch.org/audio/stable/_modules/torchaudio/datasets/speechcommands.html
- Conformer Block implementation with multi-head attention implementation we included in Question 3
  https://github.com/lucidrains/conformer
- Conformer implementation used as a reference for our implementation in Question 3
  https://github.com/sooftware/conformer
- Description of Swish Activation Function and tunable beta parameter
  https://sefiks.com/2018/08/21/swish-as-neural-networks-activation-function/
- Notebook the setup for Question 4 was adapted from:
  https://colab.research.google.com/github/pytorch/tutorials/blob/gh-pages/_downloads/63ef278c9730746362d08162a440df77/speech_command_classification_with_torchaudio_tutorial.ipynb#scrollTo=ym0Ld-JqJeM