Computer Vision and Pattern Recognition

CS 4243

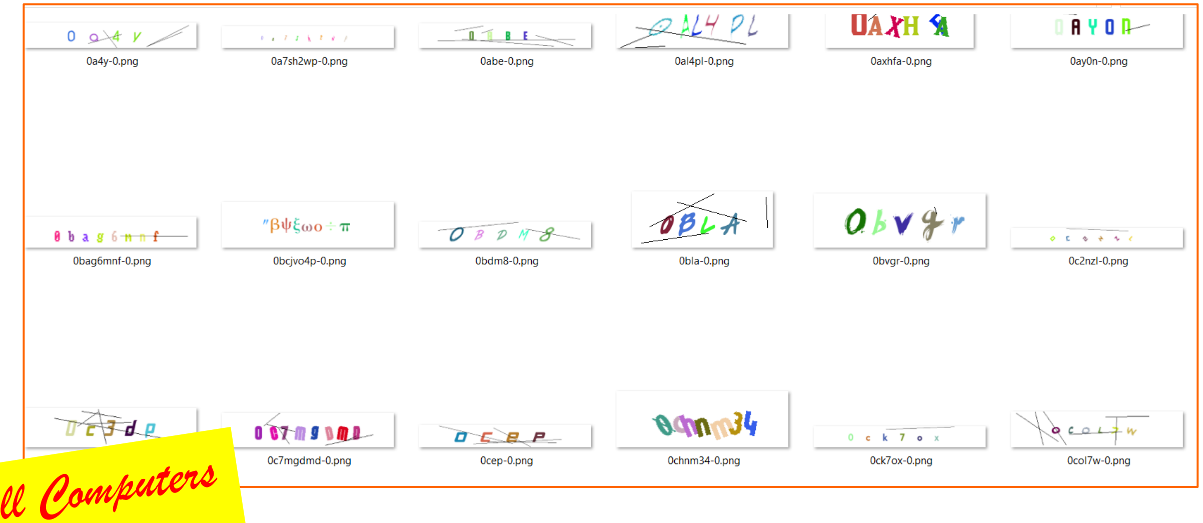
Mini Project

S2-Y2024/25

NUS National University of Singapore | School of Computing

Mini Project Definition (CS4243-2024/2)

National University of Singapore

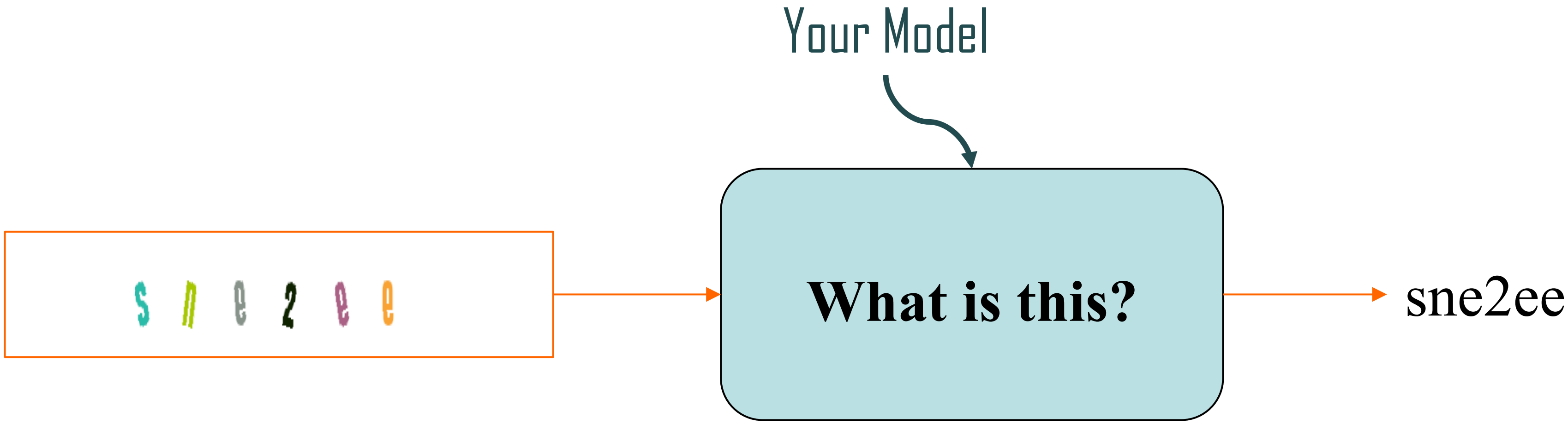


Completely Automated Public Turing test to tell Computers and Humans Apart: CAPTCHA

Problem Statement

- You are going to help AI and Computer Vision Communities by developing a **Captcha Recognition** system.
- This is a **teamwork**.
- You design and train your model, then test it. Your system should be able to get a CAPTCHA, and then show its contents/characters as a text string.

Like This



Your Model

What is this?

sne2ee

Keep it case-insensitive

Goals

1. **The mini project is worth 25% of overall module marks.**
2. **Details, Objectives**

Contents, Approach, Deployment	10%
Poster Quality	2%
Q&A	5%
Curiosity, Exploration	5%
Creativity	3%

3. **In general, higher *quality* work = higher marks!**
   1. More time spent doesn’t necessarily mean higher marks (please manage your time)
   2. Higher accuracy doesn’t necessarily mean higher marks (please follow the Project Philosophy)

Goals

Stage	Metrics
Contents, Approach, Deployment	Make it as clear as possible
Poster Quality	Clear and conveys the key project achievements well to the reader
Q&A	Ability to thoroughly explain project process and considerations
Curiosity, Exploration	Demonstration of deep understanding of the method's mechanisms instead of only achieving high-accuracy results
Creativity	Novelty in the proposed solution

Project Philosophy

This project is about:

1. Understanding of fundamental computer vision and machine learning concepts. (Theory)
2. The practical skills required for a successful machine learning/ pattern recognition project. (Application)
3. Demonstration of understanding why it works and why it doesn't

This project is NOT about:

1. Usage of LARGE DEEP models
2. Overfitting and achieving high accuracy 99%
3. Hogging of computational resources (e.g GPU) and endless tuning
4. Writing ten lines of code with Keras and calling it "DeeP LeARninG“
5. Using some advanced AI and Computer Vision models that may automatically provide 100% correct recognition.

Submission Details

You will Present your team's work using poster* format. No report/slides are required.

You do not need to print the poster. Showing that on your laptop is enough.

The presentation timetable will be announced later, however, it will be on Saturday 19/4/2025. So, please keep that date empty

The contribution of EACH team member should be mentioned in the poster.

Each team should upload the poster and package (codes, etc) later on Canvas.

*The poster is to limit/constrain the total time spent or the amount of work done for this mini-project. The poster templates are provided at the end of this presentation.

Project Specifics (Core expectations)

Students are expected to:

1. **Prepare** the given dataset. (E.g, data cleaning, visualization, pre-processing, data normalization, tokenization, etc) *[This is important!!] The Datadset could be a bit dirty.*
2. **Minimally**, develop a vanilla baseline character or string recognition system.
3. **Provide** improvements for the baseline method.
4. **Show** empirical observations. (Plot learning curves, recognition scores, numerical results)
5. **Present** your work CLEARLY and nicely.

Project Specifics (Open-ended theme)

- **By nature, the given dataset contains labeled images and it is a discriminative SUPERVISED learning problem**
- **However, it can be hosted as an Open-ended theme.**
  - No restrictions on approach. (You can regard it as image classification etc.)
  - No restrictions on which evaluation metric to use, but we prefer to see **the recognition accuracy, Precision, and Recall**.
  - No restrictions on backbone architecture.
  - No restrictions on using traditional computer vision or deep learning
  - No restrictions on models or optimizers (ADAM, SGD, ADAMax, LARS, etc, if you use DL)
  - No restrictions ...

Project Specifics (Open-ended theme)

- **What?**
  - Show the methods you chose to improve the baseline classification performance.
- **How?**
  - Explain how your method(s) works etc.
- **Why?**
  - Why you chose this approach, Why it works or why it does not work.

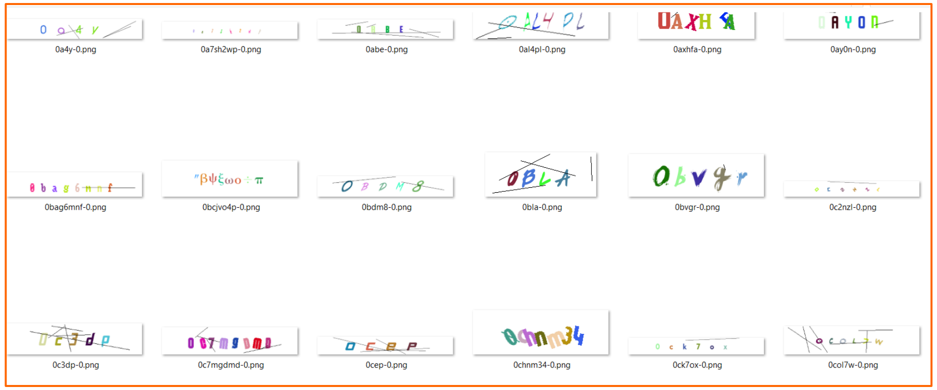
Project specifics (Open-ended theme)

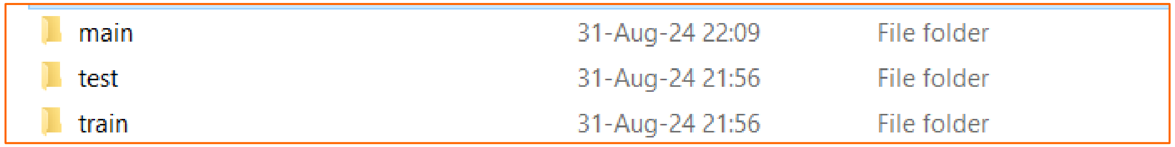
(Why is it Open-ended?)

1. Discourage hogging of GPU and training extremely large deep neural networks.
2. Newcomers have a chance to do well, likewise experienced players (Smurfs) will have a chance to learn/try new things.
3. Encourage teams to explore the literature and exercise creativity.

Dataset

- Link to the dataset: https://drive.google.com/drive/folders/1JikBA_bt7HwUYge73WuohRibamdsBTcC?usp=sharing
- You can find the Train and Test directories over there for training and testing your models.
- Main directory comprises all 10000 images. We randomly divided them into 8000 samples for training and 2000 for testing.
- The file name suggests the contents of the Captcha, ending with “-0.png”.

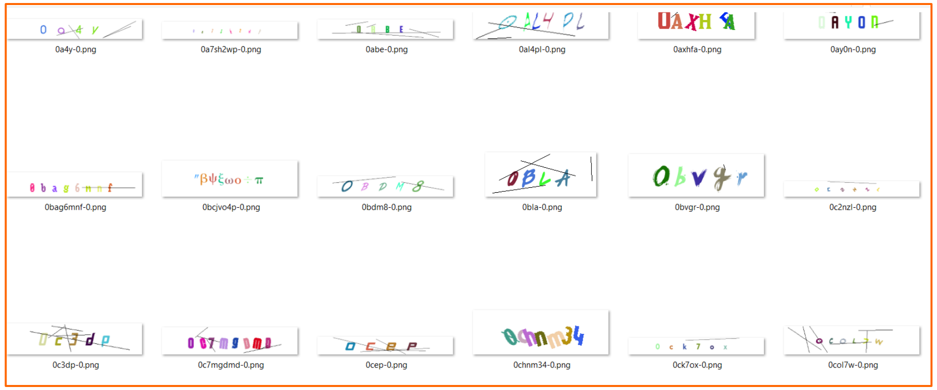


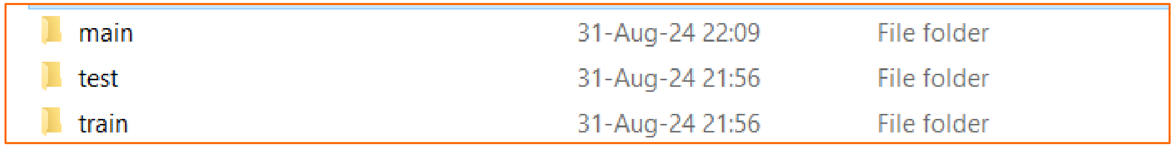


main	31-Aug-24 22:09	File folder
test	31-Aug-24 21:56	File folder
train	31-Aug-24 21:56	File folder

Dataset

- If you manage to find any dirty data sample, just remove it.
- If you manage to find any mismatch between the captcha contents and the file name, correct the file name.
- Please don’t re-divide the main file samples to new train/test subsets. Let’s stick with the current division. This way, the comparison is fairer.
- Use your test set for validation too.
- Report the metrics after training and testing both.





main	31-Aug-24 22:09	File folder
test	31-Aug-24 21:56	File folder
train	31-Aug-24 21:56	File folder

Important Points

- Dividing the load among the members of your team will help.
- Reviewing Captcha recognition literature will help.
- You may find tokenization useful. It means that you may need to separate characters and recognize them individually.
- This way, you come across 26 + 10 different classes. Why?
- This way, the recognition accuracy of your models can be measured based on characters or Captcha.
  - Captcha recognition accuracy: $\frac{\#\ correctly\ recognized\ Captchas}{N}$
  - Character recognition accuracy: $\frac{\#\ correctly\ recognized\ Characters}{N}$

Important Points

- Dividing the load among the members of your team will help.
- Reviewing Captcha recognition literature will help.
- You may find tokenization useful. It means that you may need to separate characters and recognize them individually.
- This way, you come across 26 + 10 different classes. Why?
- This way, the recognition accuracy of your models can be measured based on characters or Captcha.
  - Captcha recognition accuracy: $\frac{\#\, correctly\ recognized\ Captchas}{N}$
  - Character recognition accuracy: $\frac{\#\, correctly\ recognized\ Characters}{M}$

Important Points

- For example,
  - You test your model with all the test images,
  - For each Captcha, even if your model makes a mistake about 1 character, that Captcha would be considered as wrongly classified.
  - If 1600 Captchas recognized correctly, your model's accuracy is $\frac{1600}{2000} = 0.8 \Rightarrow 80\%$ in Captcha recognition.
  - On the other hand, you can go for tokenization and character recognition
  - If there are 12000 characters in your test set, we presume, and you recognize 11000 of them correctly, your model's accuracy will be $\frac{11000}{12000} \approx 0.917 \Rightarrow 91.7\%$

Important Points

Think!

- So, just THINK! About
  - How to optimally tokenize the Captchas, in other words, how to separate characters in a Captcha image.
  - Shall we normalize the characters? i.e., resizing all of them to the same size or no, it’s not necessary.
  - Does the dataset need any filtering and pruning?
- You may need High-Performance Computers to train your models. I’m happy with Google Co-Lab, but there are some other options. Please check **https://codesphere.com/articles/5-best-free-cloud-gpu-providers-for-hobbyists**
- Using LLMs to recognize Captchas is not allowed. Using them for knowledge acquisition is fine.

Bonus (Optional)

**** For Advanced Players Only ****

- Can you develop a generative model, e.g., a GAN, to generate new Captchas?
- You may train your generative model using all the Captchas available in the Main directory.
- Then, you may test your Captcha generator using your Captcha Recognition model.
- A GAN basically should be able to do that, but that’s up to you to decide.
- Bonus will be up to 4% extra total score, but the ceiling will still be 25% for the project.

Deliverables

- **Your poster**
- **Presentation of your poster**
- **Uploading of your poster and your commented codes to Canvas, details later.**

Finally …

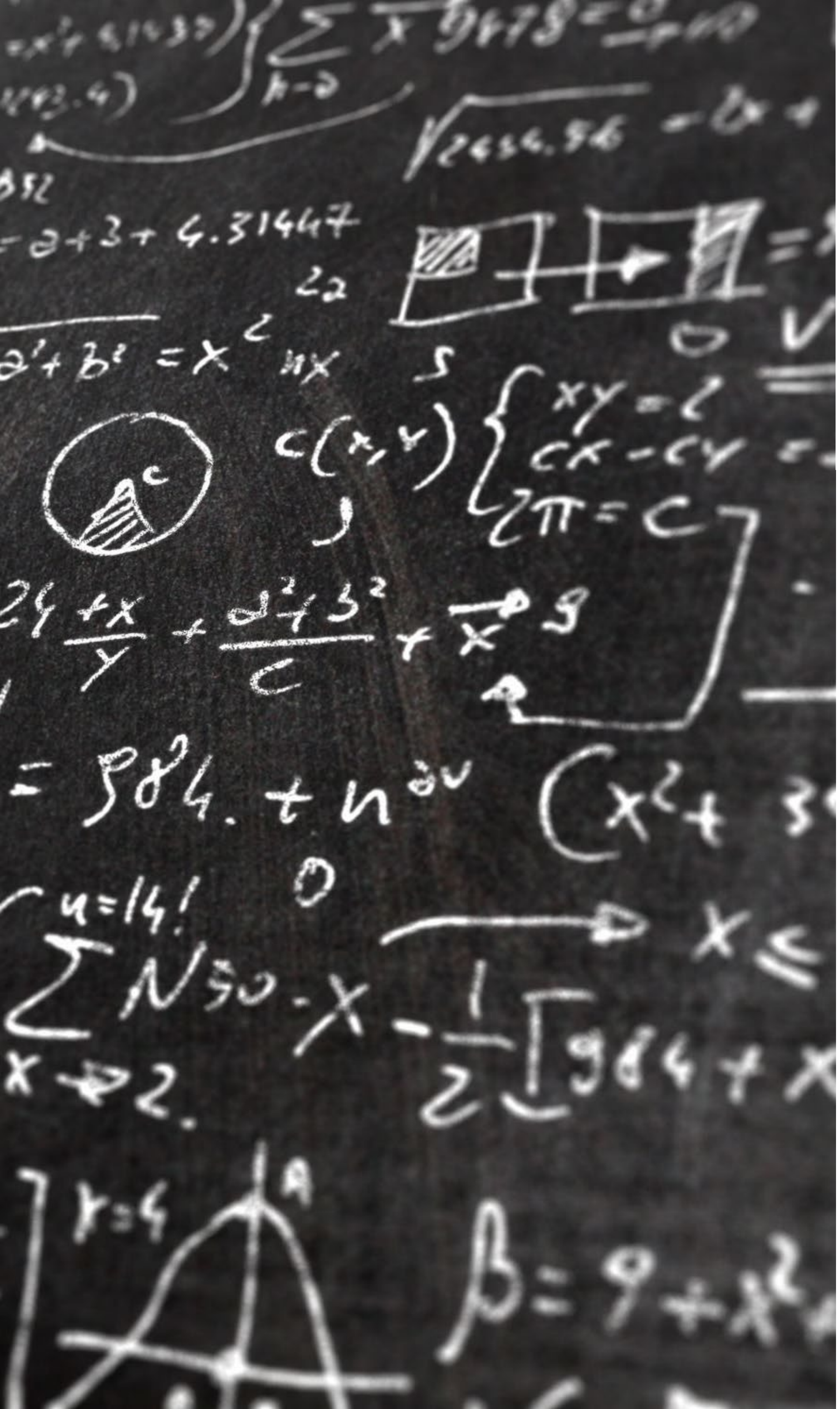
- If you need high-performance computers to run your code, in particular, to train your model, we urge you to use public cloud services like Google Colab, unless you have a gamer laptop!
- If you want to use the existing methods as a baseline/or develop your system based on them, feel free to do so. But your main focus shall be to improve that.
- If you go for using an existing method or pre-trained model, please show that you tried to improve the overall performance. In the presentation, you shall explain the baseline method, the approaches you tried, and the comparison of the results.
- Please set up your team urgently.

Poster Template

- https://github.com/AKlushyn/NeurIPS/blob/master/poster-vhp-vae.pdf
- https://github.com/reiinakano/neural-painters-pytorch/blob/master/neurips-poster.pdf
- https://github.com/joshuaas/GBDSP-NeurIPS19/blob/master/conference_poster_2.pdf

Poster Template

- PosterPresentations.com: Offers a variety of free, professionally designed PowerPoint research poster templates in multiple sizes. These templates are fully customizable to fit your presentation needs.
- Genigraphics: Provides free PowerPoint poster templates designed for scientific, medical, and research presentations. The templates are user-friendly and can help you achieve professional results.
- GENIGRAPHICS PosterNerd: Offers free PowerPoint templates to assist in creating scientific posters. They provide various designs, including the "Billboard" style, aimed at simplifying information sharing.
- Microsoft Create: Features customizable PowerPoint poster templates that are visually appealing and informative, suitable for various academic presentations.
- Aresty Research Center at Rutgers University: Provides several research poster templates in PowerPoint format, along with guidelines for creating effective posters.
- Springfield College Library Services: Offers PowerPoint poster templates and best practices for designing academic posters. Templates are available in various sizes and can be adjusted as needed.



References

1. **"Breaking a Visual CAPTCHA with Deep Learning", Jonathan Nordell.**
2. **"An Overview of CAPTCHA Mechanisms and Breaks", Wenjuan Luo, Miao Liu, Jian Weng.**
3. **"A Survey of Breaking Techniques for Human Interactive Proofs", Luis von Ahn, Manuel Blum, Nicholas J. Hopper, John Langford.**
4. **"Captcha Recognition with Deep Convolutional Neural Networks", Yi Ming, Wencang Zhao, Lixin Liu.**

** Some references are available on Canvas

That’s All …

Good Luck