image & video understanding

Visual Calculator

Project Proposal

IVU Practicals (710.241)

2022/2023

David Mihola
david.mihola@student.tugraz.at

Robert Stojanovic
robert.stojanovic@student.tugraz.at

Raul Diaz
rdiazdelser@student.tugraz.at

January 16, 2023

Graz

1 Proposal

This document contains details about our project named Visual Calculator. First, we introduce the main ideas of the project and what we want to achieve. Secondly, we provide a plan describing the steps we are going to take to achieve the desired result. Lastly, we analyzed the areas, where our plans may run into difficulties, and we introduce ideas how to mitigate them.

1.1 Description

The core idea of our project is to calculate hand written equations from images. The images may for example look like:

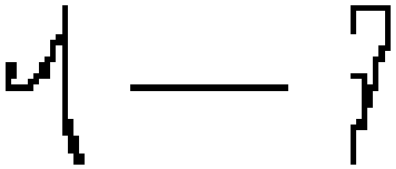

Figure 1: Equation made by concatenating separate digits and operators

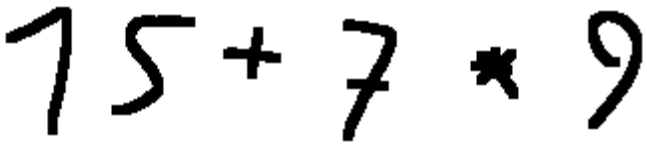

Figure 2: Equation written digitally

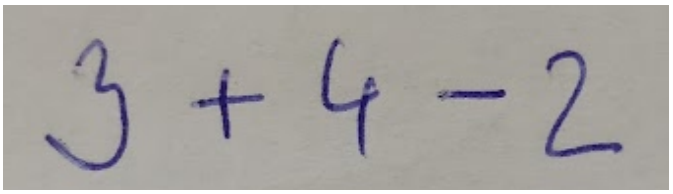

Figure 3: Handwritten equation

Additionally as of now, we assume, that the input images will fulfill these criteria:

- the hand written equation must be clearly visible without any obstructions and overlappings,
- there is a distinct contrast between the written text and the background, with the text being of a darker color than the background,
- each image has a given aspect ration, which is yet to be specified,
- the height of digits[1] is between 70-80% of the height of the image, placement can be arbitrary,
- the number of digits and operators on an image can vary, but there will be a specific upper bound on the number of characters, which is also yet to be cleared,

[1] Does not apply to operators, as these can be shorter than digits. But the width should not vary much from the width of digits.

- the number of operators is limited to perform addition, subtraction, multiplication and division with the respective characters of plus (`+`), minus (`-`), asterisk (`*`) and slash (`/`).

As the main goal of this course is to learn concepts of computer vision, we plan to create a simple program, which will read an image in a specified location by the user and output the result of the equation photographed on the given image. Ideally, if we do not encounter many of the issues listed in 1.3 and we still have time before the deadline, we would like to create more practical solution. Such as a simple android application, which will allow the user to take a photo of an equation in the desired aspect ration and immediately provide the result.

To achieve the described task, we plan to test two approaches and choose the better one. The first approach is a combination of traditional computer vision algorithms, such as MSER described in [1] and practically shown in [2], with machine learning techniques focusing on image classification. The second approach consists only of the use of machine learning algorithms, more specifically the use of convolutional neural networks inspired by the still evolving YOLO approach from [3], [4] and [5].

To train our machine learning models we plan to generate equations from single digits and operators, which are available in the MNIST or the Handwritten Digits and Operators data sets. We will pre-process these data sets, i.e. separate characters into respective classes, crop and resize the images. Then we plan to randomly construct equations with reflecting labels by concatenating these digits and operators to a single image. We will also program a script to plot them, so we can be sure, that our training data are appropriate. Afterwards we want to further improve the capabilities of our models with images containing our own handwritten equations.

1.2 Plan

The plan listed below also includes some milestones, which might not be necessarily completed to successfully finish the project. If one milestone is solved by multiple people, we expect the workload being distributed between them equally.

- **Research the common approaches to multi-object classification.**
  Find out which approaches exist to solve similar problems and read up on those.
  Date: 20. 10. 2022
  Responsibility: Robert, David, Raul

- **Clean and pre-process existing data sets, randomly generate equations.**
  Resize images from multiple data sets to the same height and width, threshold them to black and white duo tone. Crop the width of the images, so a given character on the image spans the whole width of the image. From such pre-processed characters randomly generate and label syntactically correct equations.
  Date: 31. 10. 2022
  Responsibility: David

- **Build simplified models for the promising approaches with generated equations.**
  From the most promising researched approaches choose couple of methods and implement a basic algorithm to evaluate which works best for our use case.
  Date: 6. 11. 2022
  Responsibility: Robert implements a single digit classifier, David implements a small multi digit detector.

- **Implement a script that pre-processes handwritten equations by us.**
  Implement a script, that will crop an image with the handwritten equation to the desired aspect ratio, resize it to the desired input size for our models and convert the image to black and white duo tone.

Date: 15. 11. 2022
Responsibility: Raul

- **Implement a script for annotating handwritten digits and operators.**
  Implement a script, which will make annotating characters and possibly whole equations as fast and convenient as possible.
  Date: 30. 11. 2022
  Responsibility: Robert

- **Improve generated data set.**
  Augment generated data set with our own handwritten data to get better results.
  Date: 10. 12. 2022
  Responsibility: Robert, David, Raul

- **Implement an algorithm to reconstruct and solve the labeled equations.**
  Implement an algorithm, that calculates the result of the equation based on the output of the machine learning model.
  Date: 20. 12. 2022
  Responsibility: Raul

- **Finish implementation of the final models.**
  Finish the implementation of the two chosen approaches and choose the better one, which will be used in the final application.
  Date: 31. 12. 2022
  Responsibility: David, Robert

- **Fine-tune and evaluate the chosen model using our own data.**
  Try to further improve the chosen approach mainly on equations written by us or other people.
  Date: 5. 1. 2023
  Responsibility: David, Robert

- **Combine all the parts into a single application.**
  Fully implement the selected approach including image loading, pre-processing, classification and equation solving in to one application or a script.
  Date: 10. 1. 2023
  Responsibility: Raul

- **Finish writing the report.**
  Date: 15. 1. 2023
  Responsibility: Raul, David, Robert

1.3 Risk Analysis

While researching common approaches to object detection and classification and after first experiments in this field of our own, we have identified the areas listed below as the main possible points of failure of the project.

- Firstly, we have considered the possibility, that our approach for detecting and labelling equations will not work, therefore we will simultaneously develop 2 different approaches for the task. On the expected case that we are successful with both approaches, we will have the opportunity to choose the one that performs better. There is the possibility of both approaches not working and this would mean the failure of the project, but we believe the chances of this happening are extremely low.

- Another common risk in computer vision projects is not having enough training data. But we believe this will not be a problem since we can work with multiple data sets containing the chosen characters in 1.1 and from these we can generate an arbitrary number of equations. Other issue might be that these generated equations from single handwritten digits and operators will not exactly match real equations written by people. If this is the case, we are also prepared to create more examples manually, but we expect the probability of this being necessary quite low.
- Third possible point of failure might be the fact, that we do not have enough processing power to train large machine learning models, which would be sufficient for labeling long equations. To deal with this possible issue, we plan to limit the number of characters in a equation to a relatively small number.
- Another possible issue is that our models for equation recognition will often not be 100% accurate and will not detect or will mislabel some digit or operator. In this case we will provide the user an opportunity to edit the detected equation to get the correct result.
- The last potential issue is the required formatting of the images of the equations. This can lead to the risk of lacking enough time to develop an application or interface to guide the user of the Visual Calculator into taking the appropriate picture of the equation. The probability of this happening is moderately high but with some written instructions as an alternative guide the risk would be mitigated.

References

[1] Matas, J., Chum, O., Urban, M. and Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*. 2004, vol. 22, no. 10, p. 761–767. DOI: https://doi.org/10.1016/j.imavis.2004.02.006. ISSN 0262-8856. British Machine Vision Computing 2002. Available at: https://www.sciencedirect.com/science/article/pii/S0262885604000435.

[2] Meganha, F. *Detecting handwriting regions with openCV and Python* [online]. March 2018 [cit. 2022-10-31]. Available at: https://felipemeganha.medium.com/detecting-handwriting-regions-with-opencv-and-python-ff0b1050aa4e.

[3] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. *You Only Look Once: Unified, Real-Time Object Detection*. University of Washington, Allen Institute for AI, Facebook AI Research, 2015 [cit. 2022-10-31]. Available at: https://pjreddie.com/media/files/papers/yolo_1.pdf.

[4] Redmon, J. and Farhadi, A. *YOLO9000: Better, Faster, Stronger*. University of Washington, Allen Institute for AI, december 2016 [cit. 2022-10-31]. Available at: https://pjreddie.com/media/files/papers/YOLO9000.pdf.

[5] Redmon, J. and Farhadi, A. *YOLOv3: An Incremental Improvement*. University of Washington, 2018 [cit. 2022-10-31]. Available at: https://pjreddie.com/media/files/papers/YOLOv3.pdf.