Proposal: Diversified Vision-Based Puzzle Solver

Andrew Corum

March 15, 2020

1 Personal Background

I am a CS grad student, originally a PhD student working on cryptography research (fully homomorphic encryption, differential privacy, and multi-party computation). However, recently I have decided to switch my focus to topics in artificial intelligence and machine learning.

In regard to this project, my most relevant coursework includes Data Mining (B565), Machine Learning (B555), and Elements of Artificial Intelligence (B551). I am also currently taking Computer Vision (B657) and Advanced Computer Graphics (B581) (which may be helpful if I do some more technical data augmentation or visualization).

The majority of my general coding experience comes from the BCS degree I received in undergrad. I also have worked as a web developer (intern and contractor), which has given me a lot of practical programming experience. At IU, I have primarily used Python for development, so I am familiar with its general coding standards and modules that are available.

While most of my graduate research has been focused on cryptography, in undergrad I attended an NSF research program, which resulted in a publication [1]. In this project, I (along with a professor and another student) designed and implemented a method of PDF malware detection using image classification techniques. Thanks to this experience, along with my computer vision coursework, I should have enough of the basic technical skills to tackle this project.

2 Introduction

The use of games and puzzles has become a quintessential component in the search for general artificial intelligence. Puzzle solving has been a classic measure of intelligence. Top research groups continuously work to improve computers' abilities to solve classic puzzles such as Chess and Go [2, 3]. Thanks to the versatility of abstraction, the theory behind solving puzzles and games can be more than just basic measures of intelligence or pedagogical tools [4], but even used to tackle real world problems as far reaching as cryptography, biology, and economics [].

Computer vision is a seemingly distinct area of computer science that has received significant attention in recent years. Some classic computer vision problems are object detection/recognition, image classification, and segmentation. Computer vision systems often employ multiple techniques to solve a range of complex problems, such as autonomous driving or facial recognition.

3 Project Description

This project finds itself right in the intersection of computer vision and game solving. The primary goal of this project is to construct a program that can solve a puzzle when given as a picture/image.

Additionally, rather than only solving one specific type of puzzle (which is often done in other projects [5, 6]), this project should be capable of solving a diversified range of logic puzzles. Since a general artificial intelligence puzzle solver may be impractical given the current state of AI/ML, this diverse puzzle solver will instead be designed to handle a set of predetermined puzzle types (such as jigsaw, Sudoku, Rubik's cube, Akari, etc).

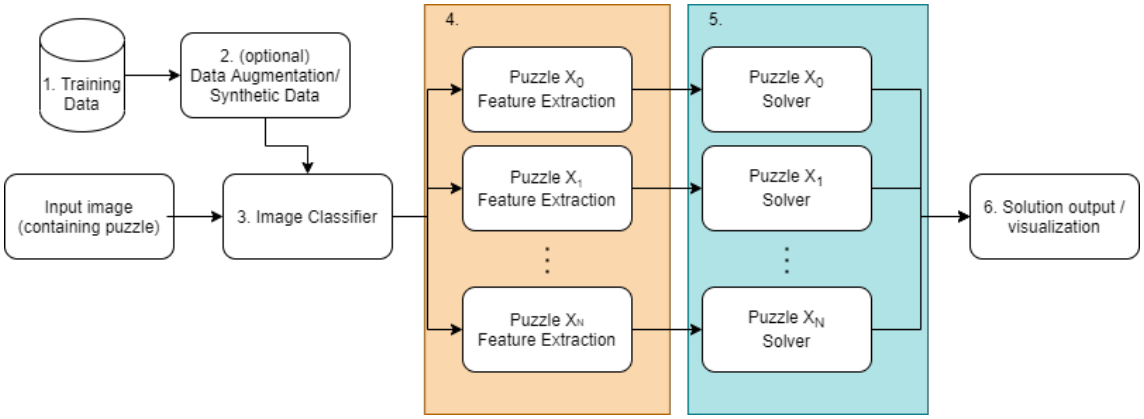

Figure 1: Pipeline for diverse vision-based puzzle solver.

The diverse vision-based puzzle solver will need to follow a pipeline to solve a given image of a puzzle, as shown in figure 1. A puzzle will be pictured and passed into an image classifier (3). This classifier will need to be trained on a dataset containing images of labeled puzzles (1), which are optionally augmented or generated (2) to improve the robustness of the classifier. Features of the puzzle, such as lines or characters, will need to be extracted (4) depending on the puzzle type. Once the necessary information has been extracted from the input image, then the puzzle can be solved (5) and the solution presented (6).

For the most part, the training data for the classifier will need to be hand created for this project. Although this is typically a difficult and time-consuming process, it should be made very feasible using data augmentation. Creating synthetic data using modeling techniques could be an interesting approach as well, but it will not be a primary objective of this project. The image classifier will require application of machine learning techniques. Using computer vision tricks (keypoints, SIFT) and classic ML algorithms (KNN, DT, or RF) should be enough for classification. Alternatively, if data augmentation is successful, a convolutional neural network (CNN) could also be used to classify the images. The puzzle feature extraction step will differ slightly for each puzzle type (although many of the processes will likely overlap). This step will require CV techniques such as segmentation, object detection, and optical character recognition. Then the puzzle solver will use ML/AI techniques to solve the puzzle given its key features. This can be done using classic AI search methods, or more complex learning techniques (such as neural networks).

4 Potential Challenges

The largest initial challenge to this project will be the limited amount of data. Images of the puzzles will need to be personally captured and then augmented to create a dataset sizeable enough to train ML classifiers. Various image data augmentation techniques can be used, as listed in a recent survey by Shorten et al [7]. Time allowing, I am considering working on a project in my Advanced

Computer Graphics class (B581) to generate synthetic data to help train the image classifier. The use of synthetic data for training classifiers is not new [8], but it does require somewhat sophisticated graphics techniques.

Another major challenge is the fact that the end result of this project is only met via the completion of the pipeline shown in figure 1. If any piece of this pipeline fails, then the overarching goal is not achieved. While this would be disappointing, I do believe that completing the individual pieces of this project will still be worthwhile. This is because each step of this process will require many different approaches of applied ML.

References

[1] A. Corum, D. Jenkins, and J. Zheng. Robust pdf malware detection with image visualization and processing techniques. In *2019 2nd International Conference on Data Intelligence and Security (ICDIS)*, pages 108–114, 2019.

[2] Feng-hsiung Hsu. Ibm's deep blue chess grandmaster chips. *IEEE Micro*, 19(2):70–81, 1999.

[3] Fei-Yue Wang, Jun Jason Zhang, Xinhu Zheng, Xiao Wang, Yong Yuan, Xiaoxiao Dai, Jie Zhang, and Liuqing Yang. Where does alphago go: From church-turing thesis to alphago thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, 3(2):113–120, 2016.

[4] Z. Markov, I. Russell, T. Neller, and N. Zlatareva. Pedagogical possibilities for the n-puzzle problem. In *Proceedings. Frontiers in Education. 36th Annual Conference*, pages 1–6, 2006.

[5] S. Gur and O. Ben-Shahar. From square pieces to brick walls: The next challenge in solving jigsaw puzzles. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4049–4057, 2017.

[6] Baptiste Wicht and Jean Hennebert. Camera-based sudoku recognition with deep belief network. *6th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2014*, pages 83–88, 01 2015.

[7] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019.

[8] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1082–10828, 2018.