Automatic Grading System for Plot-Based Questions Using Machine Learning

Doğukan ŞAHİN
dogukansahin551@gmail.com

Faruk AKDEMİR
akdemirfaruk.1@gmail.com

Semih HAŞAL
semih.hasal@hotmail.com

Supervisor: Prof.Dr.Ali Fuat ALKAYA

PROBLEM DESCRIPTION

Problem

Current methods for grading plot-based questions are hard and take a lot of time. This project aims to create an automated system to make grading easier and more accurate. We use advanced algorithms to improve grading for teachers.

Our Approach

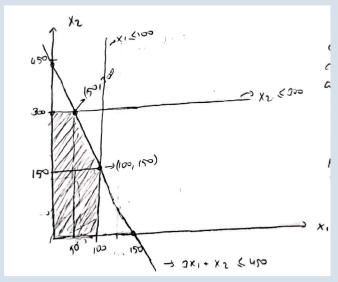
We used the Oriented FAST and Rotated BRIEF algorithms, along with the Structural Similarity Index and VGG16, to compare images and obtain similarity percentages. Subsequently, we developed an automatic grading system by training artificial neural networks and k-nearest neighbor models based on these similarity rates and actual grades. This system aids teachers in quickly and fairly grading plot-based questions.

Objectives

- To reduce the instructor's workload by reducing the number of papers that the system requires the instructor to manually review. As a result of auto-grading, we plan to reduce the number of papers that the trainer needs to grade
- Keeping the margin of error in the evaluation criteria of the system to a minimum and reaching the closest value to the grades given by the instructor.
- Optimizing the automatic assessment process so that students' exam papers are read in less time.

Dataset

- 250 quiz paper
  - 200 for training
  - 50 for testing
- Plot-based questions
- Varied sizes to test system adaptability.



Algorithmic Application

- If necessary, resize the student's graph to match the dimensions of the reference plot before processing
- Apply each algorithm to assess the similarity between the student's response and the reference graph.
- Calculate the similarity percentage for each algorithm.
- Save results in csv file

Cross Validation

Cross-validation is a method we use to evaluate the performance of our model on data it has not seen, as objectively and accurately as possible.

METHODS

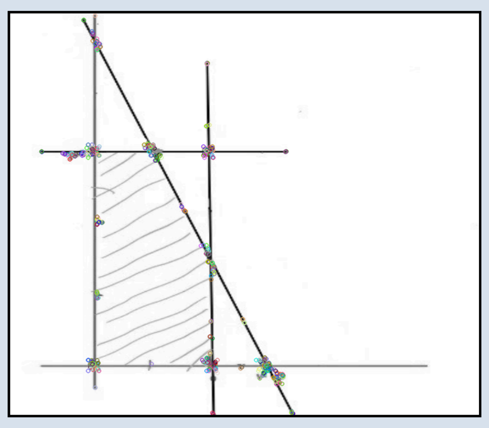
ALGORITHMS

In the implementation of our auto-grading system, we employ three key algorithms to enhance accuracy and efficiency in grading plot-based questions:

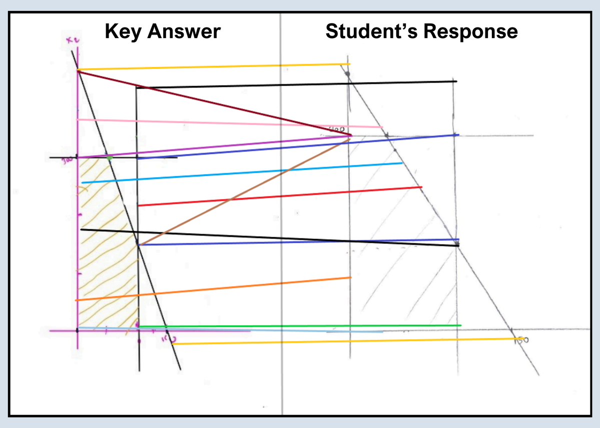
Oriented FAST and Rotated BRIEF

ORB plays a vital role in detecting and matching key features within student-submitted graphs. It assesses structural similarity by identifying key points and descriptors in both the student-drawn and reference graphs, facilitating the detection of matches to calculate a structural similarity index. This index serves as a quantitative measure of likeness between the student's response and the expected solution.

Feature Point Extraction



Matching Similar Point



Structural Similarity Index (SSIM)

SSIM is utilized alongside ORB to provide a more nuanced evaluation of image fidelity. It assesses the similarity between two images based on luminance, contrast, and structural aspects, offering a comprehensive view of image quality. This is particularly useful in our project for analyzing graph-based questions where structural details are paramount.

VGG16

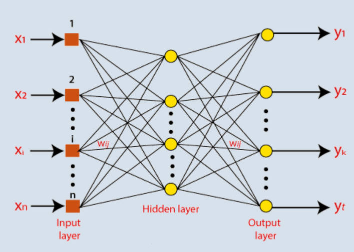
A Convolutional Neural Network model, VGG16, is leveraged for its depth and precision in image classification tasks. VGG16 excels in detecting hierarchical features within images, which is critical for analyzing and evaluating the graphical content in student responses. The model's architecture, featuring small convolution filters and deep layers, allows for a sophisticated analysis of graph structures, aiding in the auto-grading system's ability to discern similarities and assign accurate grades.

MACHINE LEARNING METHODS AND PERFORMANCE METRICS

In our project, we utilized two powerful machine learning models, **Artificial Neural Networks (ANN)** and **k-Nearest Neighbors (kNN)**, to automate the grading process based on image comparisons. Here's how we incorporated each model into our grading system:

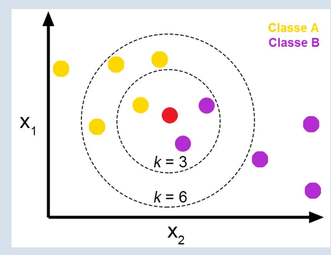
Artificial Neural Networks (ANN)

Inspired by biological neural networks found in the human brain, we used ANNs to learn the complex patterns and relationships between the similarity percentages derived from image comparison algorithms (like Oriented FAST and Rotated BRIEF, the Structural Similarity Index, and VGG16) and the actual grades assigned by teachers. This setup allowed the ANN to accurately predict the grades for new plot-based questions by analyzing how similar they are to previously encountered examples.



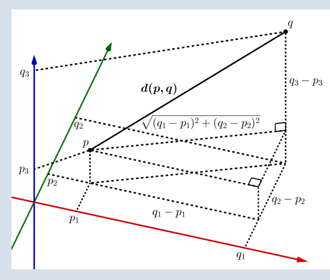
k - Nearest Neigbor (kNN)

kNN operates on a simpler principle where it classifies new data points based on the most common outcomes among its 'k' nearest neighbors in the training data. We applied the **Euclidean distance** formula to find the 'k' nodes closest to the test data we provided. The formula we apply obtains three-dimensional distance information.



$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

In the project, our data are vectors represented in a **three-dimensional** plane or space with 3 coordinates such as x, y, and z. These three coordinates represent the data of **ORB**, **SSIM**, and **VGG16** models. **Nodes** are **GRADED data**.



ORB	SSIM	VGG16	GRADED
0.84415584	0.78108329	0.738212943	81

Mean Absolute Deviation

We used Mean Absolute Deviation (MAD) values to determine how consistent the dataset we used was. We used the results we obtained to determine which of the datasets was more reliable and predictable.

$$MAD = \frac{\frac{\sum_{i=1}^{N} |R_i - E_i|}{N}}{\frac{\sum_{i=1}^{N} R_i}{N}}$$

Root Mean Square Error

Root Mean Square Error (RMSE) measures how much the predicted values deviate from the actual values. When analyzing the RMSE error, it also analyzes the data set.

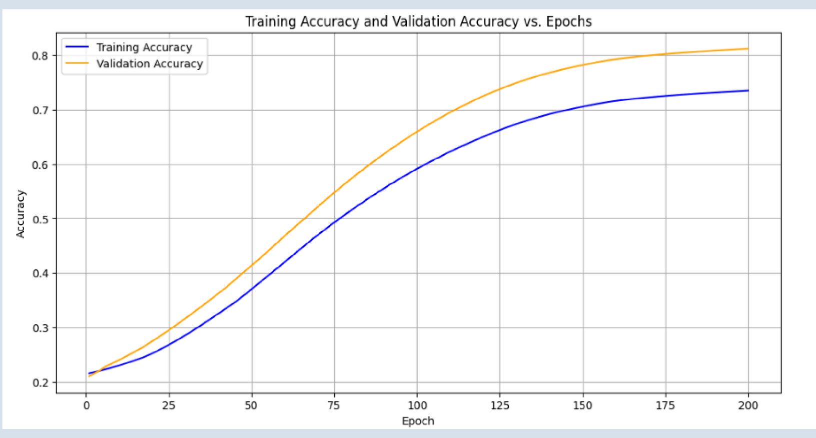
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

EXPERIMENTS AND RESULTS

Artificial Neural Networks (ANN)

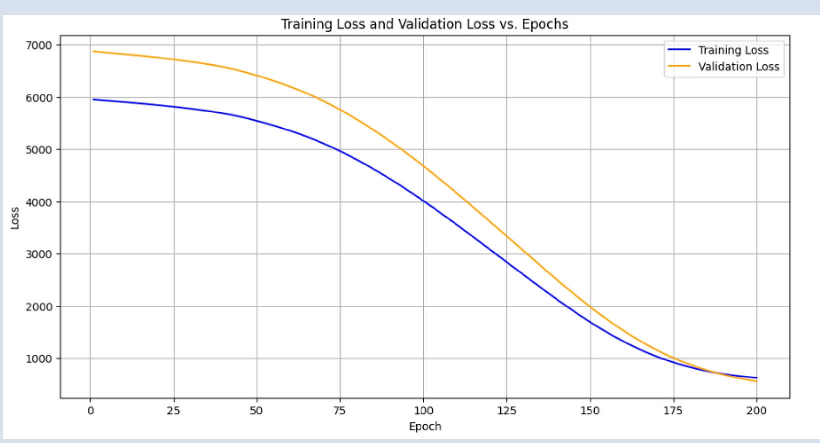
Accuracy Over Epochs

The training accuracy and validation accuracy both increase over epochs. This indicates that the model's predictions become more accurate with training. The steady improvement in accuracy signifies effective learning and better performance on both training and validation data.



Loss Over Epochs

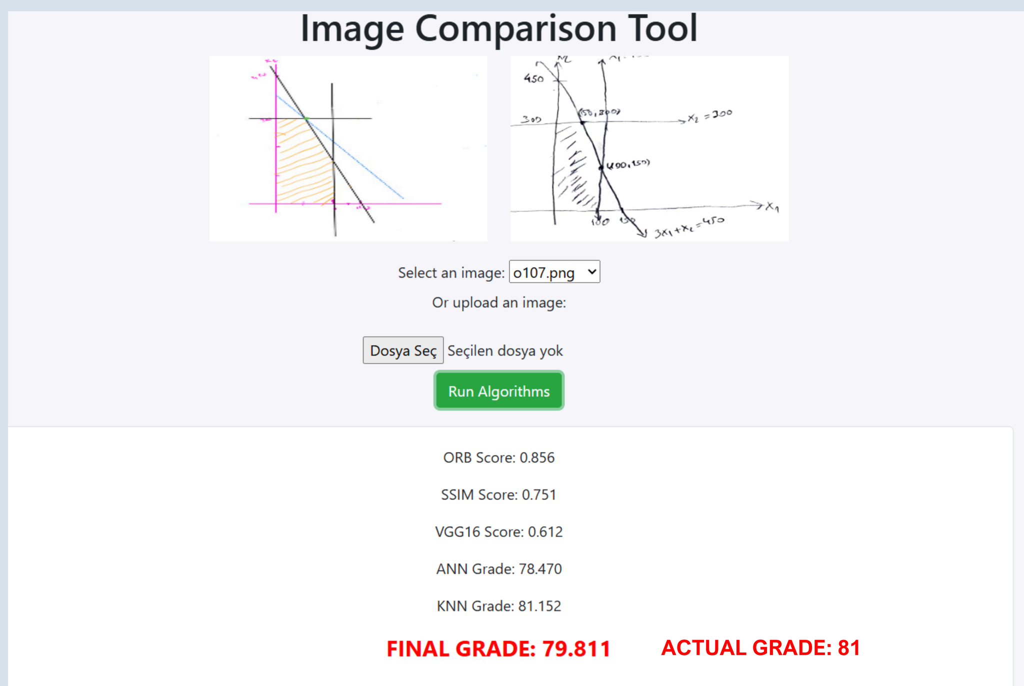
The loss curve shows a steady decrease in loss values over the training epochs. Initially high, the loss drops consistently, indicating that the model is improving its predictions and minimizing errors. The validation loss follows a similar downward trend, showing the model's ability to generalize well.



k - Nearest Neigbor (kNN)

When our predicted values were within **± 30%** of the real grades, we accepted them as **"True"(1)**, and results outside this range as **"False"(0)**. With Cross-Validation, we randomly selected 20 data from the entire dataset and created a dataset with them. We did this process **5** times. We observed the accuracy and Mean Absolute Deviation (MAD) values of these 5 different folds with the "Cross-Validation Accuracy and MAD Values" chart. Apart from these, another value we observed is "Root Mean Square Error (RMSE)".

USER INTERFACE



CONCLUSION

Our tool makes grading plot-based questions faster and more accurate. It reduces the work for teachers and gives reliable results. This helps teachers focus more on teaching and less on grading. By using our system, students get quicker feedback, which helps them learn better. Overall, our project improves the educational process for both teachers and students. It provides a more efficient way to handle assessments, ensuring fairness and consistency in grading. This innovative approach supports the growing need for technology in education, making it easier to manage and evaluate complex student responses.

References

[1] ORB: an efficient alternative to SIFT or SURF Conference Paper in Proceedings / IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision · November 2011

[2] Zagoruyko, S., & Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4353-4361).

[3] Alkaya, A. F., Gültekin, O. G., Danaci, E., & Duman, E. (2020). Comparison of Computational Intelligence Models on Forecasting Automated Teller Machine Cash Demands. Journal of Multiple-Valued Logic & Soft Computing, 35.

[4] Provost, F., & Fawcett, T. (2013). Data science for business: [what you need to know about data mining and data-analytic thinking]. Sebastopol, Calif., O'Reilly.

[5] Euclidean distance 3d. Kmhkmh. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Euclidean_distance_3d_2_cropped.png. Licensed under Creative Commons Attribution 4.0 International. https://creativecommons.org/licenses/by/4.0/deed.en. No changes were made.

Technologies Used

scikit learn
OpenCV
pandas

Scan me!