Improving Table Structure Recognition With Visual-Alignment Sequential Coordinate Modeling

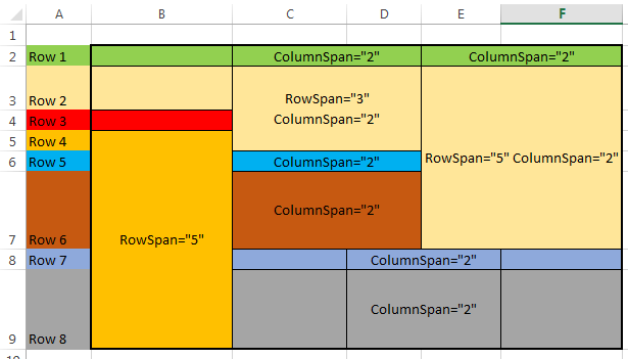
What is the problem?

The problem lies in **recognizing the table structure** from unstructured images. This involves extracting both the **logical** (rows, columns, and spanning information) and **physical structures** (bounding boxes of cells). Earlier methods often generated **inaccurate bounding boxes**, as they lacked local visual details in the logical structure representation.

First Cell		Second Cell	Third Cell
Data	Data	Data	Data
Data	Data	Data	Data
Data	Data	Data	Data

What has been done earlier?

End-to-end models were used to predict both the logical and physical structures. However, these methods were imprecise, as they did not integrate local visual features, resulting in less accurate bounding box predictions.



Shreeya Mishra,B421046

What are the remaining challenges?

The remaining challenge is to improve the accuracy of bounding box prediction for table cells. Existing end-to-end table structure recognition models often struggle with imprecise bounding boxes because the logical representation they use lacks local visual information.

What novel solution proposed by the authors to solve the problem?

The authors propose the **VAST (Visual-Alignment Sequential Table recognition)** framework to solve the problem. It introduces two key innovations:
Coordinate Sequence Decoder: This decoder predicts the bounding box coordinates of table cells sequentially (left, top, right, bottom), significantly improving the precision of the physical structure recognition.
Visual-Alignment Loss: A loss function that aligns logical structure predictions with local visual features, ensuring that the model incorporates detailed visual information to produce more accurate bounding boxes for non-empty cells.
This combination enhances both logical and physical structure accuracy in table recognition tasks.

Shreeya Mishra,B421046

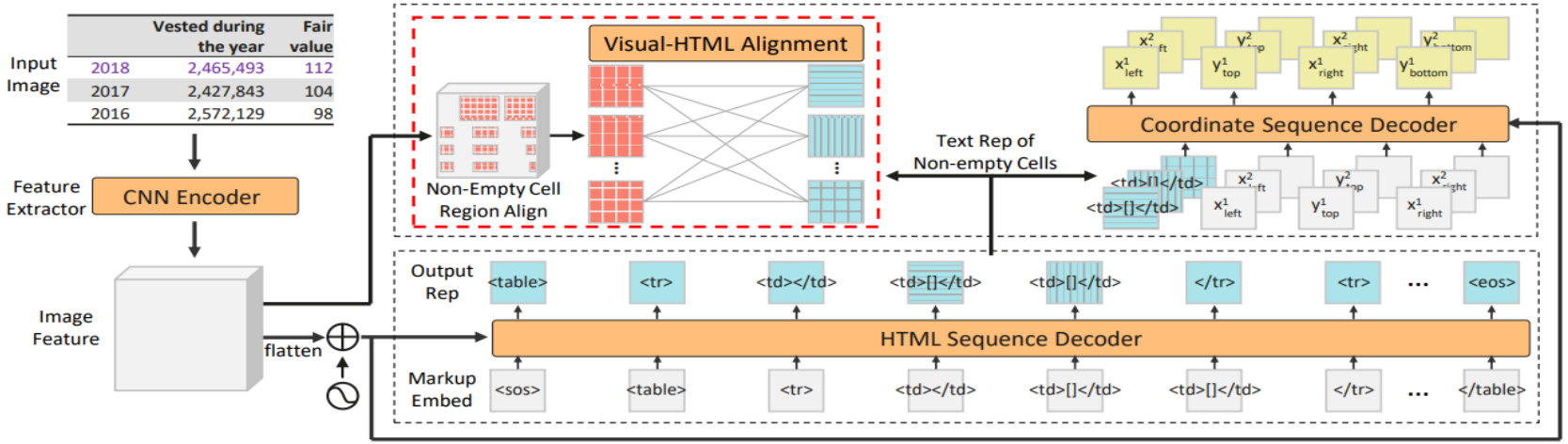


Figure 3. Architecture of our proposed VAST. The red dotted zone refers to the operations only in training.

MLA Style Citation of the paper

Huang, Yongshuai, et al. "Improving table structure recognition with visual-alignment sequential coordinate modeling." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

Shreeya Mishra,B421046