PyramidTabNet Transformer-based Table Recognition in Image-based Documents

Join the Virtual Presentation

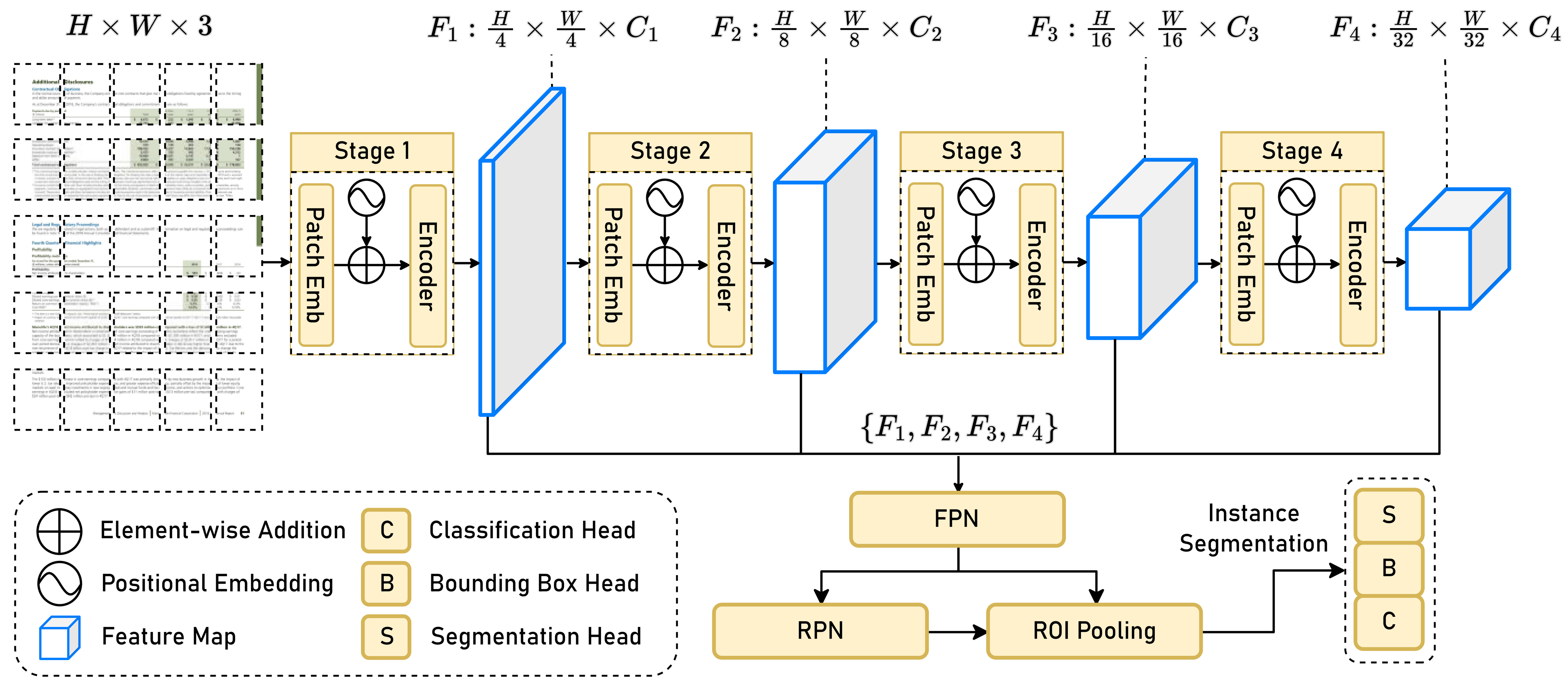
Muhammad Umer[1], Muhammad Ahmed Mohsin[1], Adnan Ul-Hasan[2], Faisal Shafait[1,2]

[1] School of Electrical Engineering & Computer Science (SEECS), Islamabad, Pakistan

[2] Deep Learning Laboratory, National Center of Artificial Intelligence (NCAI), Islamabad, Pakistan

Introduction

- In this research, we present an architecture leveraging the benefits of Pyramid Vision Transformer backbone paired up with novel augmentation strategies
- We present a novel tabular image generative augmentation technique to effectively train the architecture.



Proposed architecture of PyramidTabNet

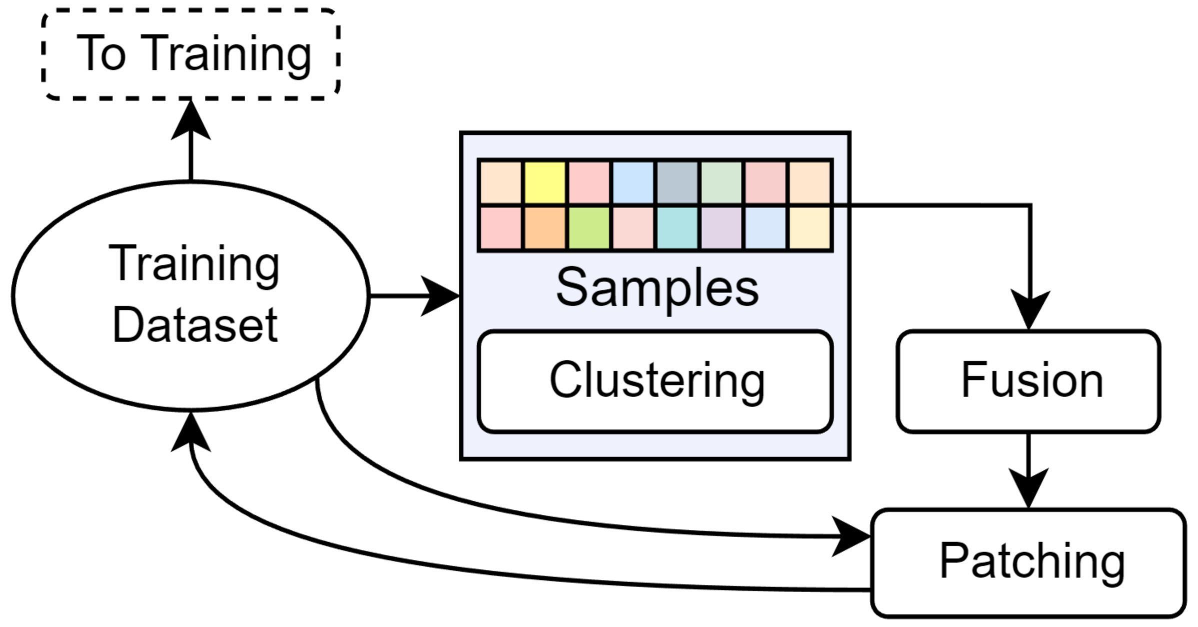
Augmentation Strategy

- Addresses data-hungry nature of transformers.
- Enhances architecture with fine-grained object detection capabilities through:

a) **Clustering:** Groups similar tables using K-means clustering based on visual characteristics

b) **Fusion:** Joins clustered tables horizontally/vertically based on structure after joint masking

c) **Patching:** Patches fused tables onto images from the original dataset



Proposed tabular image generative strategy

Datasets

Experiments and training were conducted on the open-source and publicly available document analysis datasets

Dataset	# Images	# Tables
ICDAR 2017-POD	1600	731
ICDAR 2019-cTDaR	600	-
Marmot	1967	-
UNLV	10k	427
TableBank	260k	417k
PubLayNet	86k	-

Table 1. Datasets used in the experimentation with their corresponding number of images and tables

Results

Evaluating the augmentation pipeline through in both table detection and structure recognition tasks through the following training pipelines:

- **Non-Augmented (NA):** No modifications to training images
- **Standard (S):** Applying standard augmentation techniques combined with strategies from DETR
- **Generative (G) (ours):** Includes standard and proposed augmentation strategies.

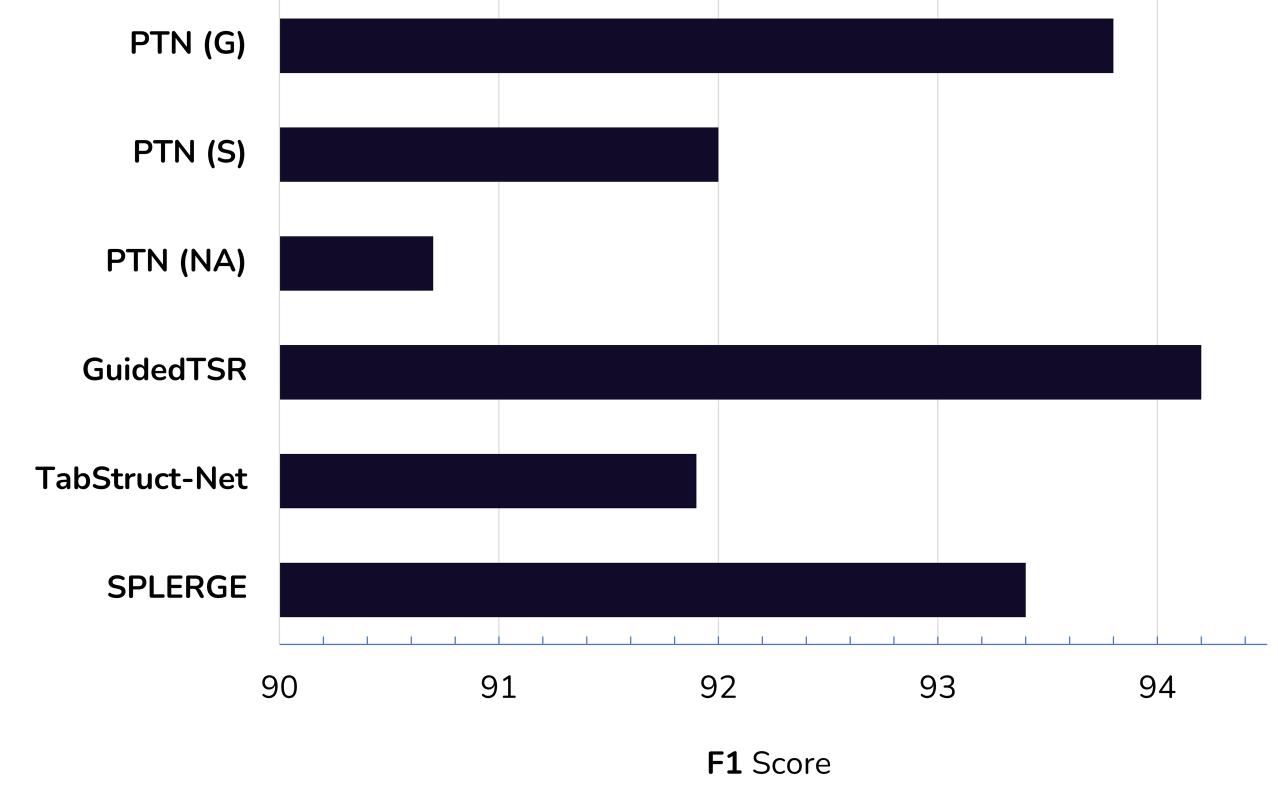
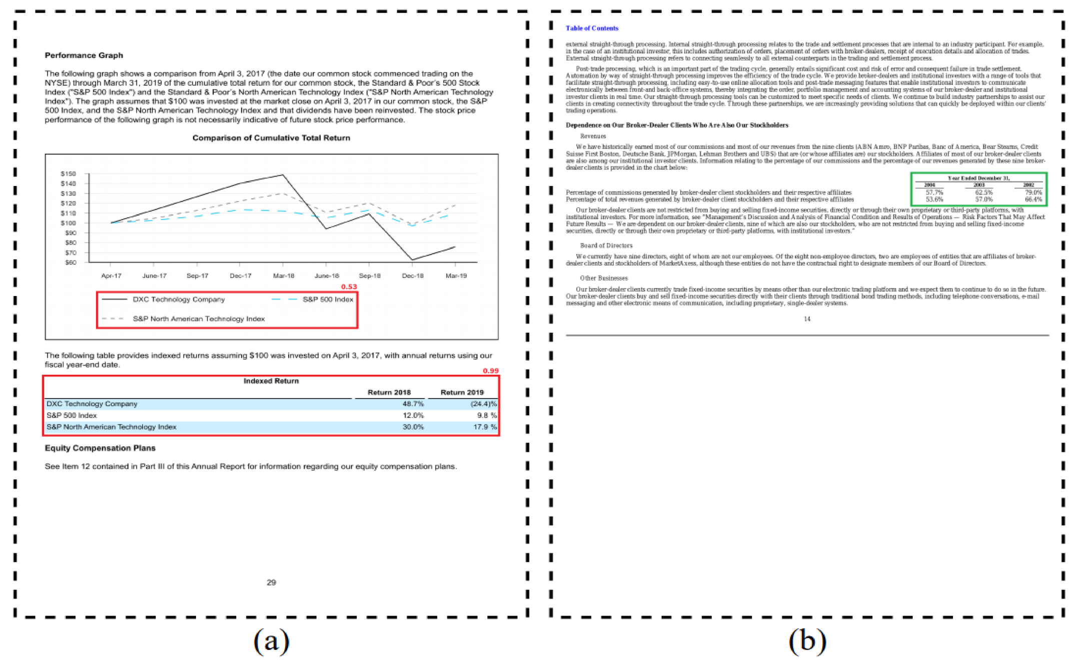


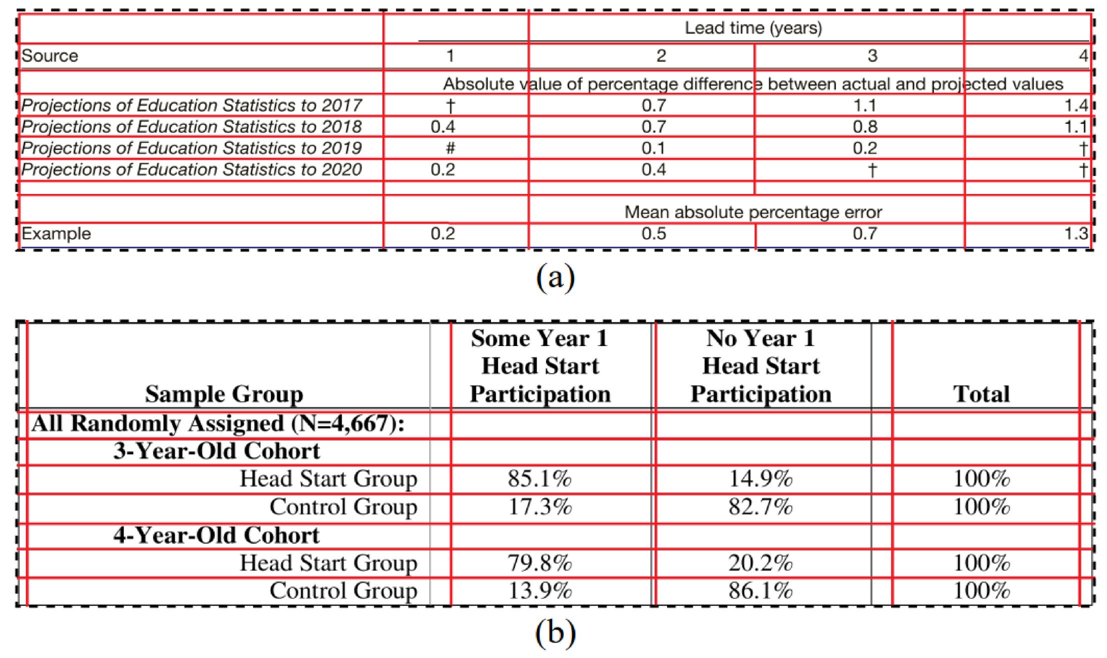
Figure 1. Table structure recognition results on ICDAR 2013, Performance of fine-tuned models is compared without post-processing techniques as is done in the current state-of-the-art

Error Analysis

- Inability to make correct predictions due to presence of dual-patching bias in the model (a) and when relative size variation is high (b)



Dependence on relative size and patching bias



Dependence on header and post-processing

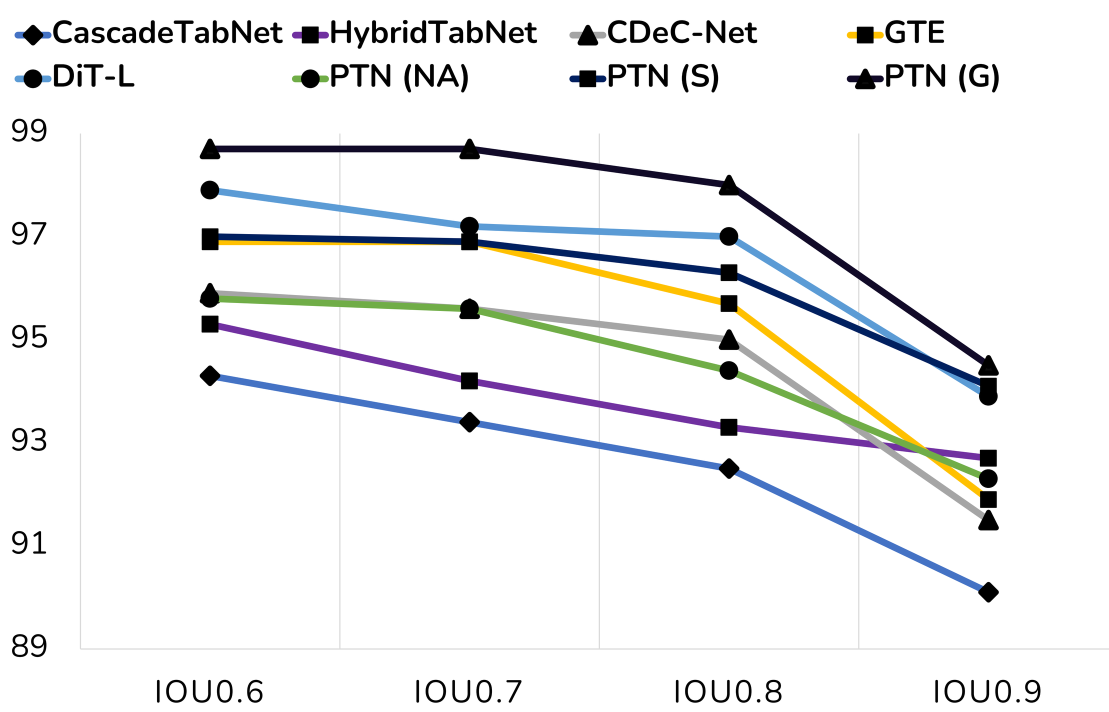


Figure 2. Table detection performance comparison on ICDAR 2019 cTDaR, F1-scores are computed at different IoU thresholds

Future Work

- Extending PTN for complex table layouts
- Applying PTN to document classification and layout analysis
- Improving structure recognition with AI upscaling on low-resolution cropped tables
- Adopting GANs alongside proposed augmentation strategy to further increase input data