Boosting Semi-Supervised Scene Text Recognition From Viewing to Understanding

Group B

XUE,Lin QI,Wenfei HAN,Yilong

1 What is STR?

Introduction

- Scene Text Recognition (STR) is the task of converting text images found in natural environments into machine-readable characters.
- STR enables computers to "read" text in photos, signs, menus, and documents - making our visual world accessible to machines.

Background

- Critical applications include: Assistance for visually impaired, Document digitization, Visual search and information retrieval
- STR pipeline include: Text detection, Character
- recognition, and Contextual correction.

Challenges of STR 2

"Limited labeled data for diverse text styles"

- Annotation is expensive and time-consuming
- Most models rely on synthetic datasets

"Real-world text varies greatly in appearance"

- Enormous variation in fonts, colors and backgrounds
- Environmental factors create additional complexity

"Difficulty recognizing artistic and oriented text"

- Decorative fonts break conventional character structure
- Poor performance on WordArt and Union14M-Benchmark

3 Methodology

Semi-supervised Learning Framework

•Teacher-Student architecture for knowledge transfer

•Leverages both labeled synthetic and unlabeled real data

•Mean Teacher with Exponential Moving Average (EMA) updates

•Strong and weak augmentation paths for consistency learning

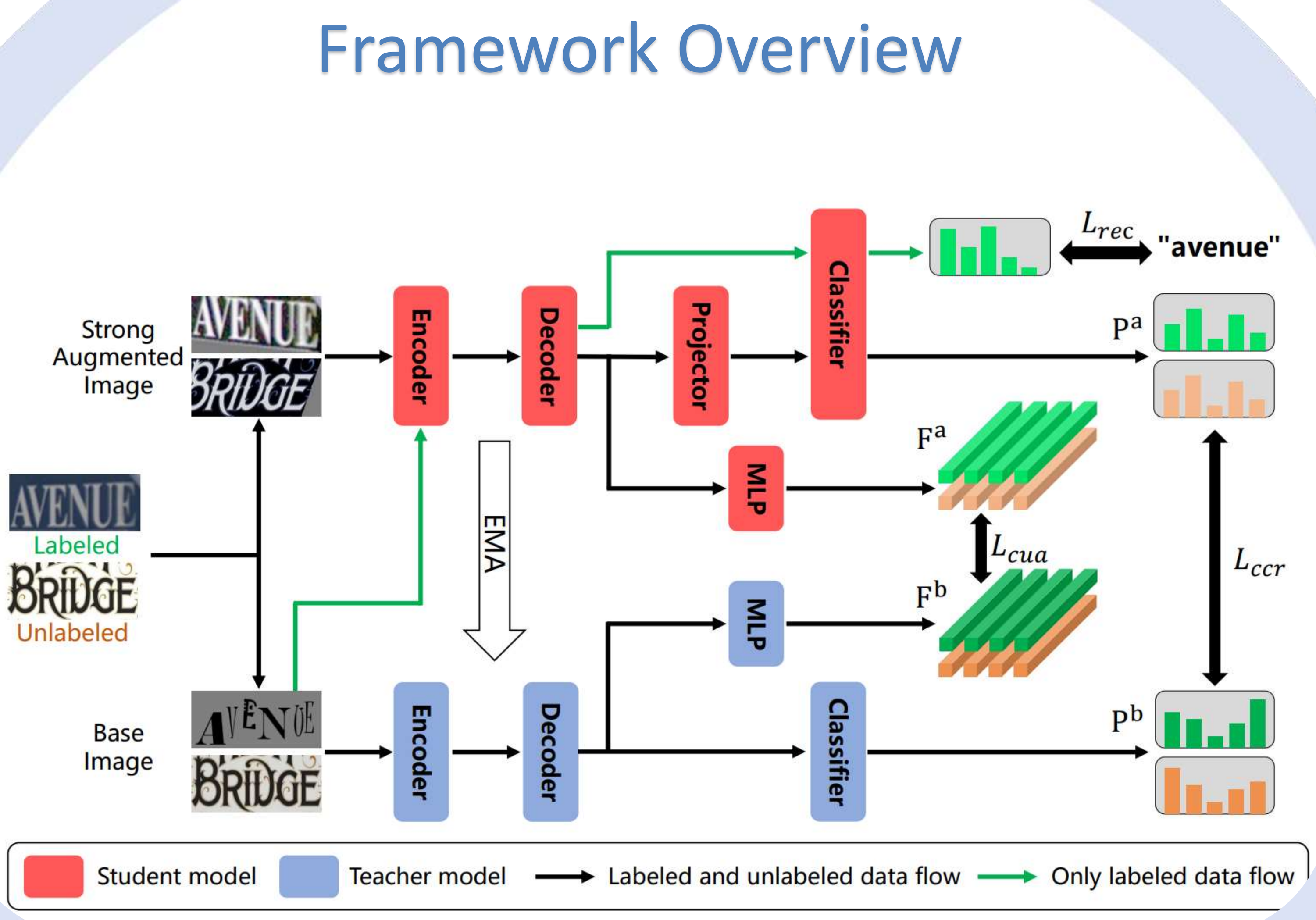
Training Strategy

•Supervised learning with labeled synthetic data

•Consistency regularization with unlabeled real-world data

•Progressive feature alignment during training

•Joint optimization of recognition and feature space



Innovation 4

Online Generation Strategy (OGS)

• Generates diverse character styles during training without backgrounds

• Creates unified representation forms for characters across domains

• Bridges the gap between synthetic and real-world text appearance

• Enhances model generalization to unseen text styles

Character Unidirectional Alignment Loss (CUA)

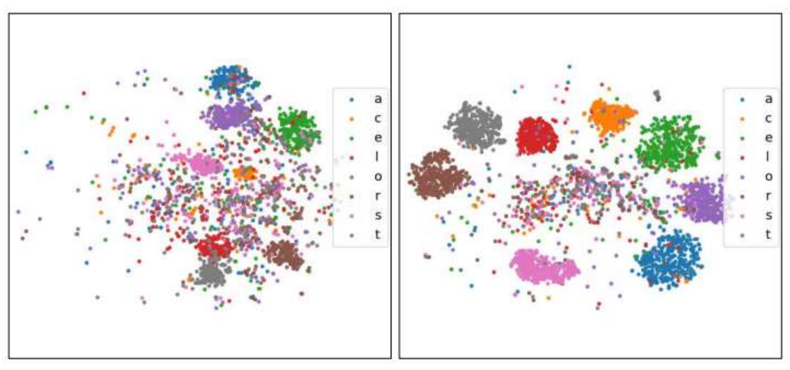
• Novel loss function that prevents feature collapse in semi-supervised learning

• Maintains distinctive features between different characters

• Unidirectional constraint allows feature refinement without oversmoothing

• Significantly improves clustering of same-character variants

5 Results



ParSeq ViSu

Feature Visualization

• ViSu shows improved character clustering with clear boundaries

• Better separation between similar characters

• More coherent grouping within same character class

• Enhanced feature organization boosts recognition accuracy

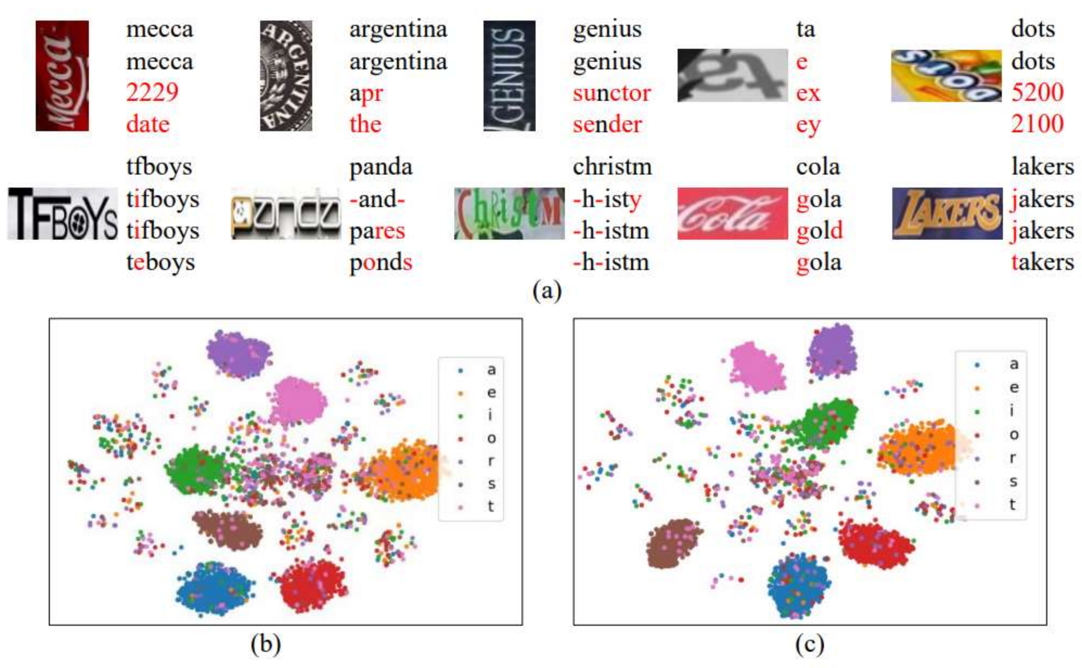
Ablation Analysis

•(b): Without CUA Loss - poorer clustering

•(c): Full model - optimized feature space

•OGS contribution: +8.2% accuracy gain

•CUA Loss: +10.2% character discrimination improvement



Performance Highlights

•State-of-the-art on artistic text: +25.1% on WordArt, +30.8% on ArT

•Strong on regular benchmarks: IIIT (+1.7%), SVT (+2.4%)

•60.3% average accuracy across all datasets (table, bottom right)

Conclusions 6

Key Findings

•Semi-supervised learning successfully bridges synthetic-to-real gap

•Character-level feature alignment is critical for STR robustness

•View-and-Summarize approach handles diverse text styles effectively

Main Contributions

•Novel semi-supervised framework for STR with minimal labeled data

•Online Generation Strategy (OGS) for diverse character synthesis

•Character Unidirectional Alignment Loss for feature space optimization

•State-of-the-art performance on challenging artistic and oriented text

Future Directions

•Extending to end-to-end text spotting in complex scenes

•Exploring language-aware feature alignment for better contextual understanding

•Adapting framework for low-resource languages and specialized domains

Method	Datasets	Cur	M-O	Art	Con	Sal	M-W	Gen	AVG
ParSeq [2]	10% (MJ + ST) + OGS	54.3	15.7	52.3	53.2	67.3	55.9	58.8	51.1
MGP [41]	10% (MJ + ST) + OGS	46.8	10.5	49.9	33.0	55.1	26.7	55.8	39.7
CLIPOCR [43]	10% (MJ + ST) + OGS	57.1	13.0	57.1	49.2	65.5	60.2	59.8	51.7
LPV [52]	10% (MJ + ST) + OGS	58.6	12.9	53.3	53.3	67.4	59.3	56.9	51.7
LISTER [5]	10% (MJ + ST) + OGS	52.0	13.7	48.9	54.4	59.8	54.5	61.0	49.2
TRBA-cr [54]	10% (MJ + ST) + OGS	67.1	17.4	58.6	51.1	67.7	33.5	57.4	50.4
ViSu	10% (MJ + ST)	57.5	79.6	49.8	44.4	66.8	50.4	59.2	58.2
ViSu	OGS	1.7	25.1	5.8	4.9	3.0	6.9	10.2	8.2
ViSu	10% (MJ + ST) + OGS	60.8	80.8	52.4	47.1	69.1	51.8	59.9	**60.3**

Source code