Paper Title:

Primitive Contrastive Learning for Handwritten Mathematical Expression Recognition

Paper Link:

https://ieeexplore.ieee.org/document/9956214?fbclid=IwAR1ZX5y9VBRkSxHJp0xfuPits_qhFSHC1nsVBvCnnBpySR0i1yKt10HIveA

1. Summary:

The paper proposes PrimCLR, an unsupervised pre-trained model for handwritten mathematical expression recognition. It uses contrastive learning to learn a representation from unlabeled data and improve downstream tasks. PrimCLR obtains a pre-trained representation by computing contrastive loss from pairs of patches to better discriminate primitives. This representation is then transferred to downstream formula recognition with supervised fine-tuning. Experimental results show that pre-training with PrimCLR significantly improves formula recognition performance and outperforms conventional contrastive learning methods. The model achieves state-of-the-art performance on benchmark datasets. The paper also discusses the impact of different types of data augmentation and the choice of image encoder on the performance of PrimCLR. Contrastive learning has been successful in improving the generalization ability of visual pre-training models, but its effectiveness for structured prediction problems like mathematical expression recognition is less explored.

1.1 Motivation

Handwritten mathematical expression recognition (HMER) is a challenging problem due to the complex and diverse layout and variation in writing styles. Despite numerous advancements, it remains unsolved and lacks large labeled datasets for training recognition models. The paper aims to address these challenges by proposing PrimCLR, an unsupervised pre-trained model for HMER. By pre-training with PrimCLR, the paper aims to significantly improve the formula recognition performance and achieve state-of-the-art results on benchmark datasets. The ultimate goal is to enhance the representation learning for structured prediction problems by capturing structural and contextual information in formula images.

.

1.2 Contribution

The paper proposes a structured contrastive learning method called PrimCLR for Handwritten Mathematical Expression Recognition (HMER). PrimCLR computes contrastive loss on pairs of patches to improve the discrimination of primitives in mathematical expressions. The work highlights the potential of self-supervised pre-training methods to improve the representation model for HMER and other visual tasks. The paper also discusses the impact of different types of data augmentation and the choice of image encoder on the performance of PrimCLR. Moreover, the contributions of this work include the development of PrimCLR, the demonstration of its effectiveness in improving HMER performance, and the exploration of contrastive learning for structured prediction problems.

1.3 Methodology

The paper proposes an unsupervised pre-trained model called PrimCLR for handwritten mathematical expression recognition (HMER). PrimCLR utilizes contrastive learning to learn a representation from unlabeled data and improve the discrimination of primitives in mathematical expressions. The pre-trained representation obtained from PrimCLR is transferred to downstream formula recognition with supervised fine-tuning. The HMER model in this paper is built on the original image-to-markup method, which is a data-driven framework implemented by an encoder-decoder neural network. The paper incorporates self-supervised pre-training and supervised fine-tuning in the encoder-decoder model to perform structured formula recognition . The proposed PrimCLR method significantly improves the performance of formula recognition and achieves state-of-the-art results on standard datasets.

1.4 Conclusion

The paper proposes a self-supervised learning framework called PrimCLR for handwritten mathematical expression recognition, which captures structural and contextual information in formula images and outperforms mainstream contrastive learning methods. PrimCLR is robust to various pre-training datasets and extends the application of visual representation learning to images with complex structure. The proposed model achieves state-of-the-art performance on the Competition on Recognition of Online Handwritten Mathematical Expressions (CROHME) dataset, which contains formula images with distortions and is distinct from commonly used datasets.

2.Limitations:

2.1. First Limitation: The paper does not provide a comprehensive analysis of the performance of PrimCLR on different datasets or compare it with other existing methods in detail. The paper does not address the issue of computational efficiency or the training time required for PrimCLR.

2.2. Second Limitation: The paper does not explore the generalizability of PrimCLR to other structured prediction problems beyond handwritten mathematical expression recognition. The paper does not discuss the potential impact of the dataset limitations, such as the lack of large labeled mathematical formula images, on the performance of PrimCLR.

3.Synthesis :

We can explore more structural and contextual information to improve representation learning for structured prediction problems.Further research can be conducted to analyze the impact of different types of data augmentation on the performance of the PrimCLR model.
The effectiveness of different image encoders, such as ResNest-101, can be further investigated to understand their impact on the performance of the model.Future work can focus on evaluating the generalizability of PrimCLR to other structured prediction tasks beyond handwritten mathematical expression recognition.The computational efficiency and training time of PrimCLR can be optimized to make it more practical for real-world applications.