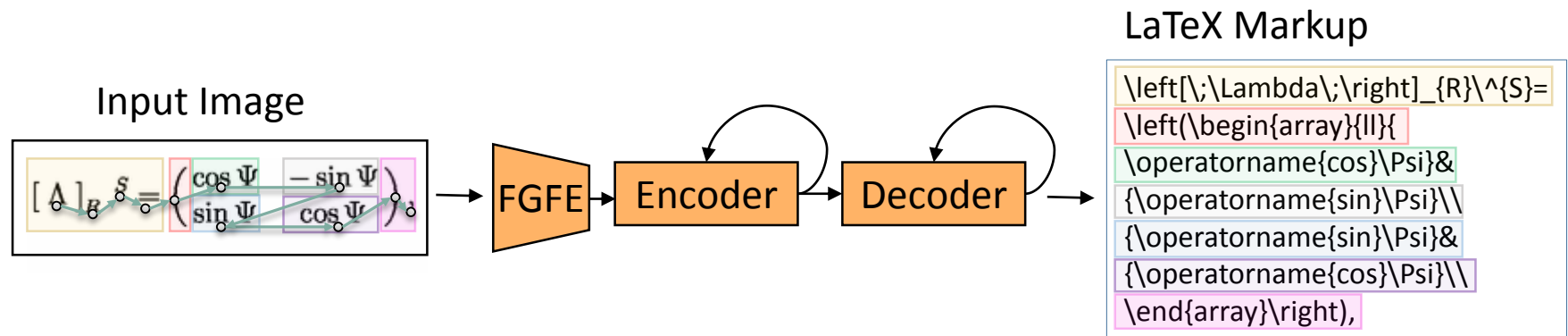


Learning Fine-Grained Image Representations for Mathematical Expression Recognition

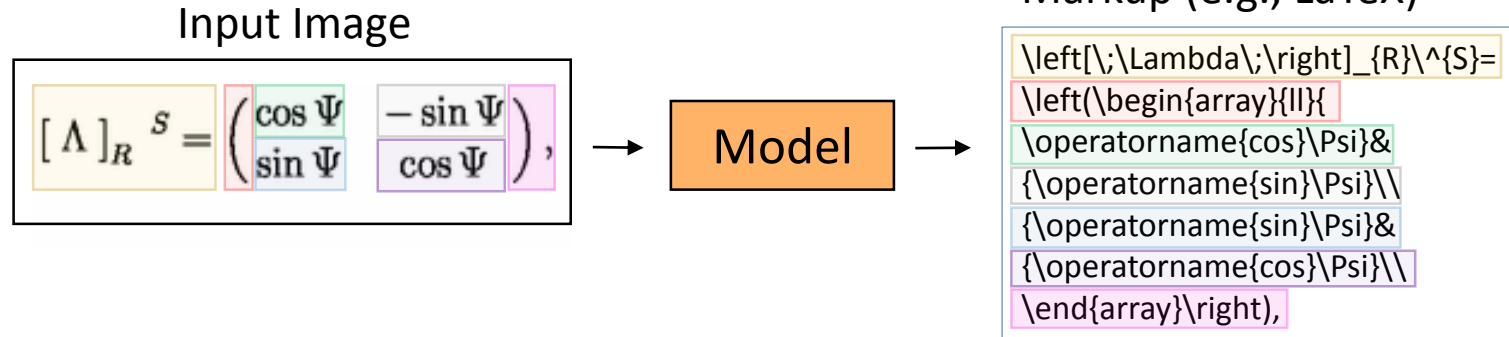
Sidney Bender*, Monica Haurilet*, Alina Roitberg and Rainer Stiefelhagen

Computer Vision for Human Computer Interaction Lab
Institute for Anthropomatics and Robotics

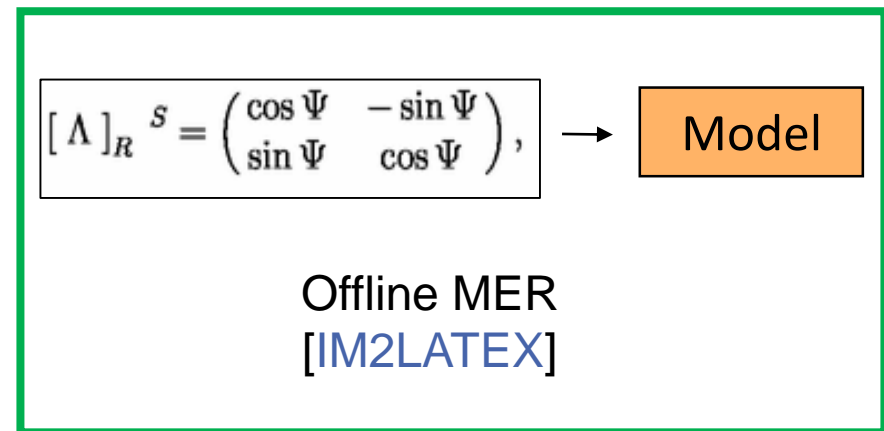
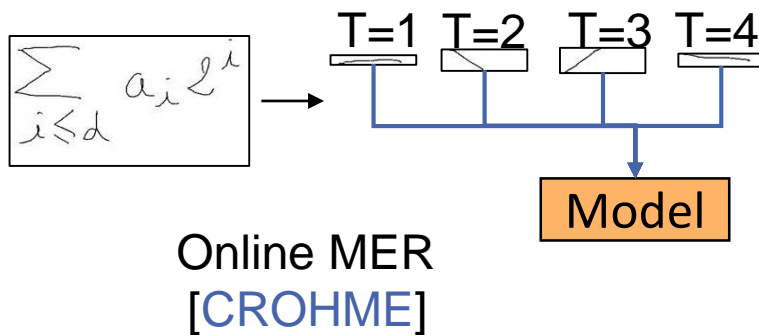


Mathematical Expression Recognition (MER)

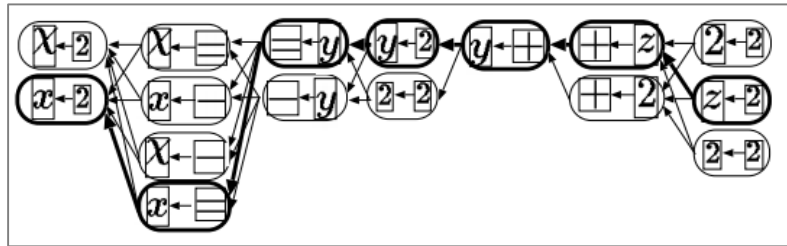
Problem Definition



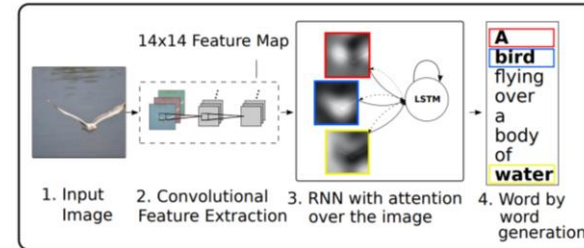
Different types of MER Tasks



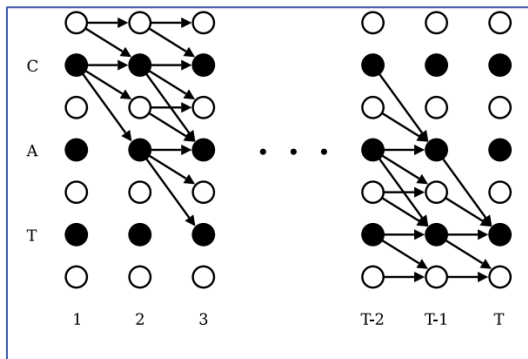
Related Work



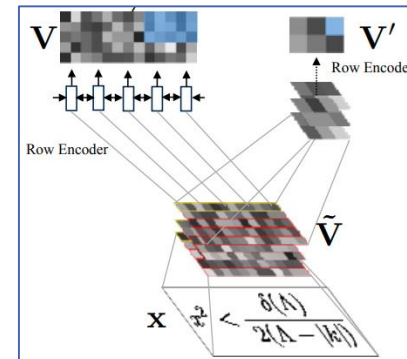
Infty [Suzuki *et al.*]



Caption [Xu *et al.*]

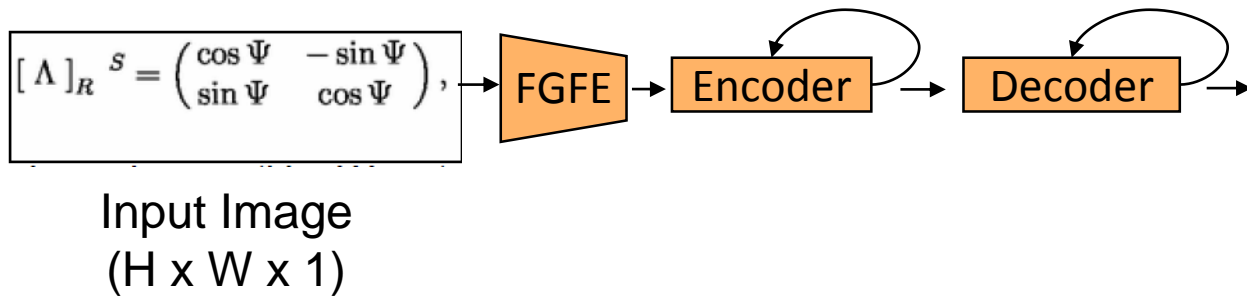


CTC [Graves *et al.*]



Im2Tex [Deng *et al.*]

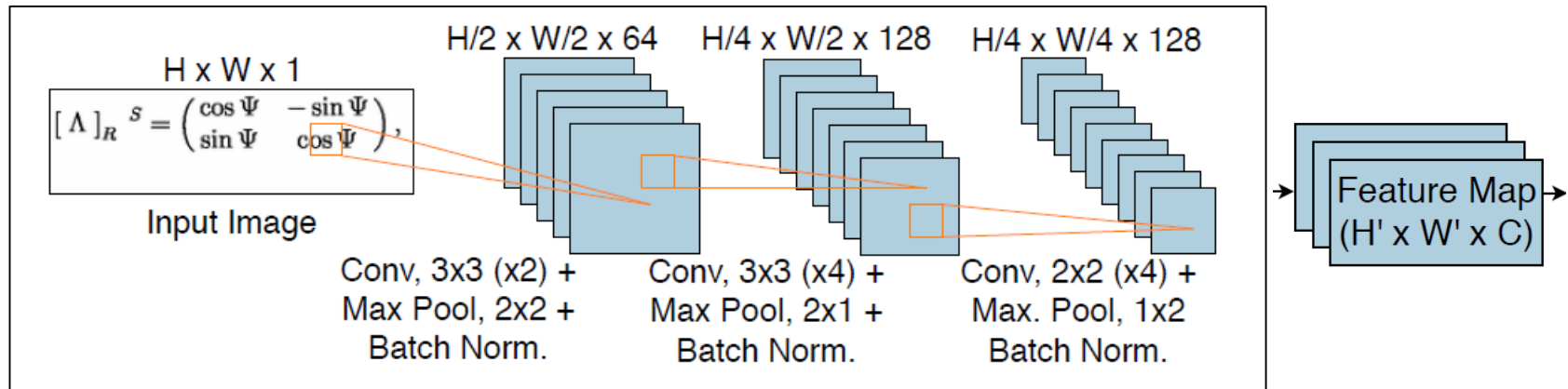
Overview of the Model



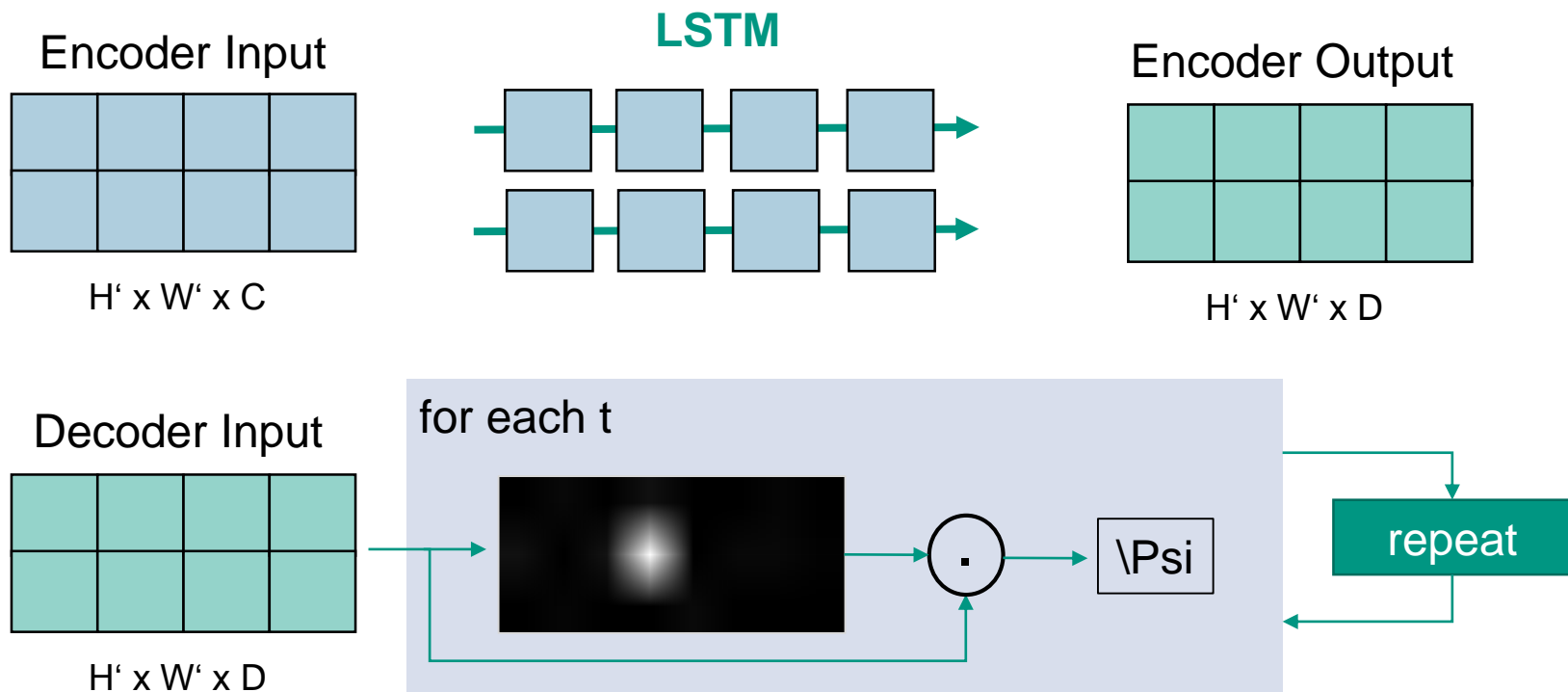
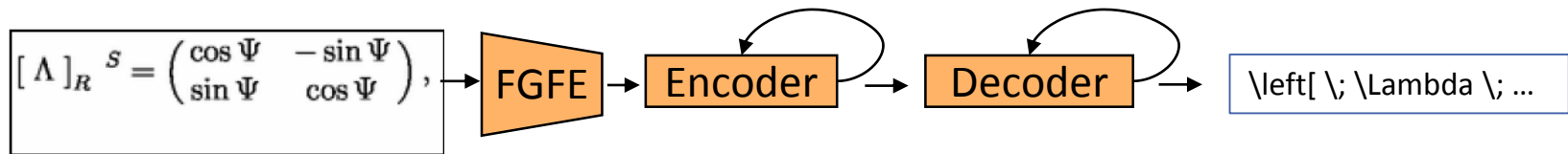
```

\left[ \; \Lambda \; \right]_{ R } \backslash
^ { S } = \left( \begin{array}{ l l }
\operatorname{cos} \Psi & \{-
\operatorname{sin} \Psi \} \\
\operatorname{sin} \Psi & \{
\operatorname{cos} \Psi \}
\end{array} \right)
  
```

LaTeX Markup



Visual Encoder and LaTeX Decoder

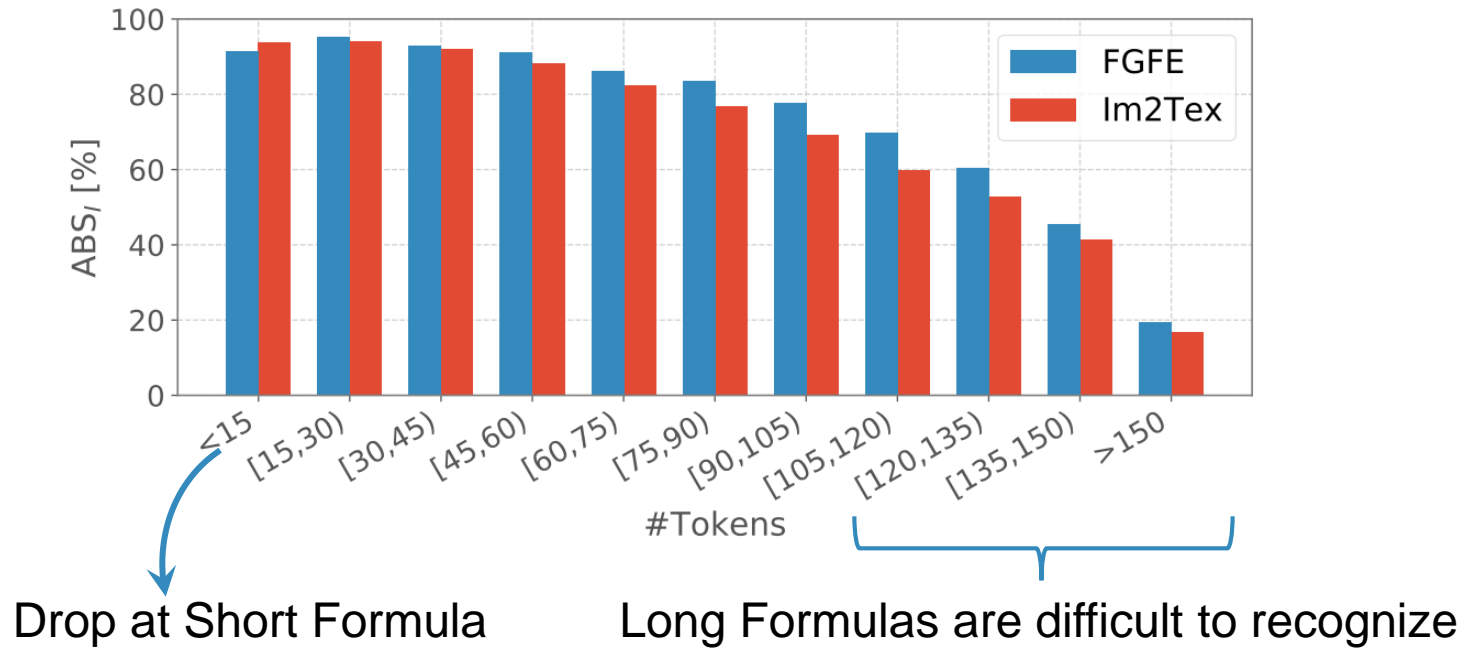


Final Results

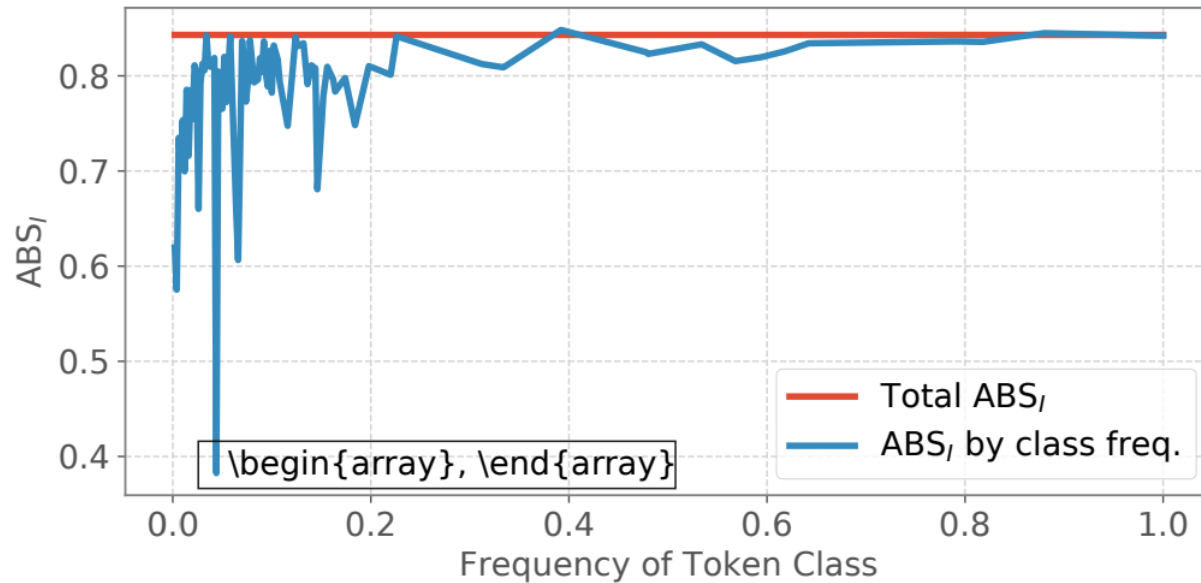
Approach	Attention	<u>text-based</u>			<u>img.-based</u>	
		BLEU	EDA_T	ABS_T	EDA_I	ABS_I
Classical Methods and Baselines						
Prior	–	0.0	20.0	0.0	85.0	0.0
INFTY	–	66.7	–	–	–	26.7
Deep Learning Architectures						
CTC	–	30.4	–	–	–	9.2
Caption	softmax	75.0	–	–	–	55.7
Im2Tex	hierarch.	86.2	–	–	–	79.6
	hard	87.1	–	–	–	77.1
	sparsemax	87.0	–	–	–	78.1
	softmax	87.7	92.1	41.2	88.6	79.9
Ours	softmax	90.3	92.8	46.8	93.1	84.3

Performance on the IM2LATEX-100K Test Set

Impact of Formula Length on Performance



Impact of Rare Token Classes



Importance of a Fine-grained Visual Representation

FE Type

Attention Maps

Predictions

Im2Tex →

$$\mathcal{A} = \prod_n \left[\exp(c_1(RL_P^2) + \dots) \right]^{\frac{i(\Delta x)^4}{L_P^4}} \rightarrow \exp \frac{ic_1}{L_P^4} \int d^4x \sqrt{-g}(RL_P^2)$$

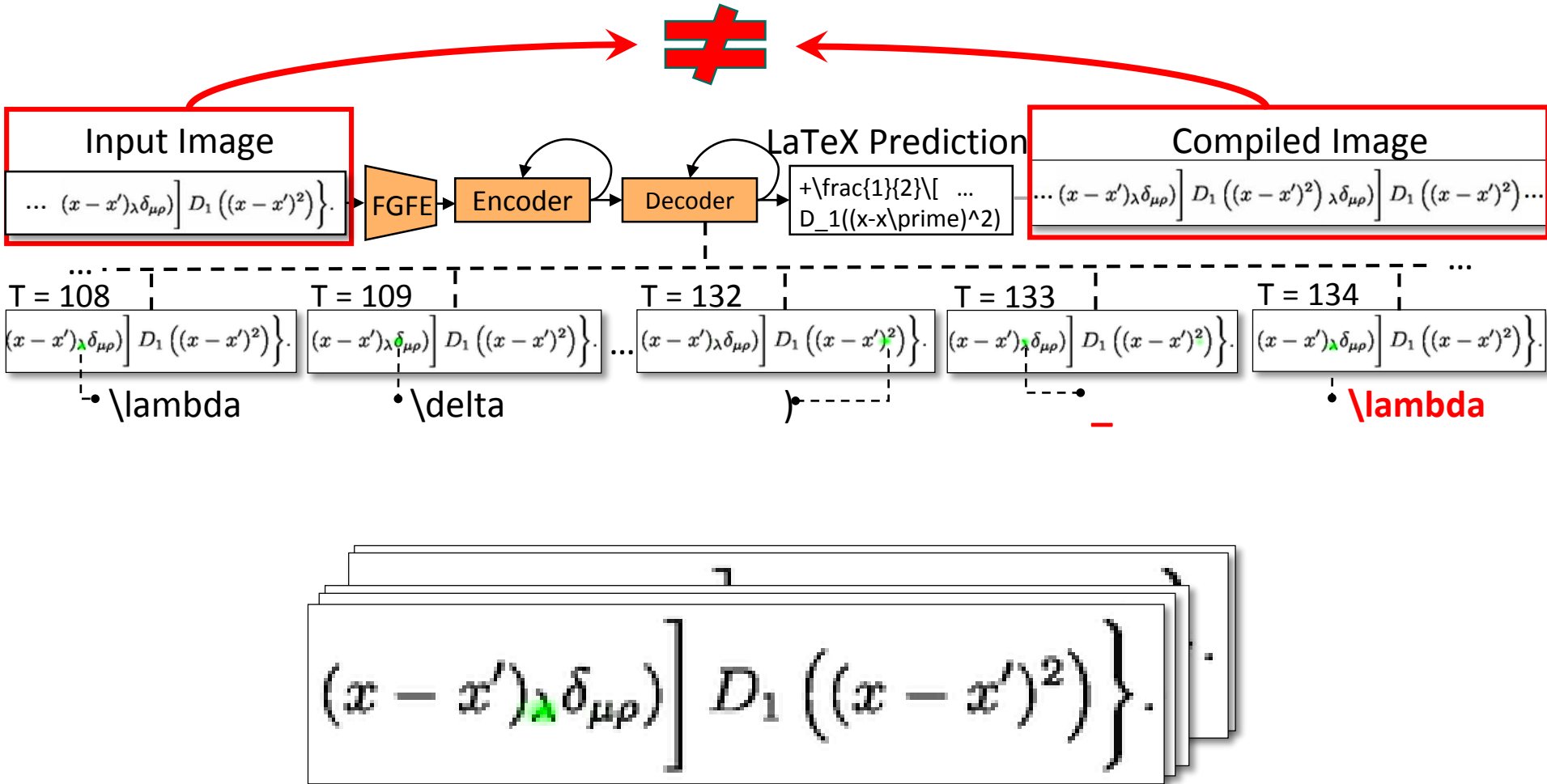
\alpha

Ours →

$$\mathcal{A} = \prod_n \left[\exp(c_1(RL_P^2) + \dots) \right]^{\frac{i(\Delta x)^4}{L_P^4}} \rightarrow \exp \frac{ic_1}{L_P^4} \int d^4x \sqrt{-g}(RL_P^2)$$

\Delta

Recursive Behavior in Long Formulas



Conclusion

Approach

- We tackled the offline MER task
- Our model was evaluated on the Im2LATEX Dataset
- We were able to improve results by over 4% in Img-Abs

Analysis

- Analysis of the performance by formula length
- Visualization of attention maps
- Impact of rare tokens on performance
- Typical errors our model produced

References

- [[Suzuki et al.](#)] **INFTY: an integrated OCR system for mathematical documents.** *M. Suzuki, F. Tamari, R. Fukuda, S. Uchida, and T. Kanahori.* In [DocEng](#), 2003.
- [[Graves et al.](#)] **Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks.** *A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber.* In [ICML](#), 2006.
- [[Xu et al.](#)] **Show, attend and tell: Neural image caption generation with visual attention.** *K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel and Y. Bengio.* In [ICML](#), 2015.
- [[Deng et al.](#)] **Image-to-markup generation with coarse-to-fine attention.** *Y. Deng, A. Kanervisto, J. Ling and A. Rush.* In [ICML](#), 2017.

Learning Fine-Grained Image Representations for Mathematical Expression Recognition

Sidney Bender*, Monica Haurilet*, Alina Roitberg and Rainer Stiefelhagen

Karlsruhe Institute of Technology, Germany

haurilet@kit.edu

