cases

doc_1		doc_2		decision	id
authors	• Bugliarello, Emanuele • de Lhoneux, Miryam • Elliott, Desmond • Lotz, Jonas F. • Rust, Phillip • Salesky, Elizabeth	**authors**	• Phillip Rust • Jonas F. Lotz • Emanuele Bugliarello • Elizabeth Salesky • Miryam de Lhoneux • Desmond Elliott	DUPLICATES	138
title	Language Modelling with Pixels	**title**	Language Modelling with Pixels		
publication_date	2022-07-14 00:00:00	**publication_date**	2022-07-14 00:00:00		
source	SupportedSources.CORE	**source**	SupportedSources.SEMANTIC_SCHOLAR		
journal		**journal**	ArXiv		
volume		**volume**	abs/2207.06991		
doi	None	**doi**			
urls	• http://arxiv.org/abs/2207.06991	**urls**	• https://www.semanticscholar.org/paper/b82f43182aa7a188c12eb9cb6992b6f82cca2120		
id	id-8233605131073278584	**id**	id-3015542265996856072		
abstract	Language models are defined over a finite set of inputs, which creates a vocabulary bottleneck when we attempt to scale the number of supported languages. Tackling this bottleneck results in a trade-off between what can be represented in the embedding matrix and computational issues in the output layer. This paper introduces PIXEL, the Pixel-based Encoder of Language, which suffers from neither of these issues. PIXEL is a pretrained language model that renders text as images, making it possible to transfer representations across languages based on orthographic similarity or the co-activation of pixels. PIXEL is trained to reconstruct the pixels of masked patches, instead of predicting a distribution over tokens. We pretrain the 86M parameter PIXEL model on the same English data as BERT and evaluate on syntactic and semantic tasks in typologically diverse languages, including various non-Latin scripts. We find that PIXEL substantially outperforms BERT on syntactic and semantic processing tasks on scripts that are not found in the pretraining data, but PIXEL is slightly weaker than BERT when working with Latin scripts. Furthermore, we find that PIXEL is more robust to noisy text inputs than BERT, further confirming the benefits of modelling language with pixels.Comment: work in progres	**abstract**	Language models are defined over a finite set of inputs, which creates a vocabulary bottleneck when we attempt to scale the number of supported languages. Tackling this bottleneck results in a trade-off between what can be represented in the embedding matrix and computational issues in the output layer. This paper introduces PIXEL, the Pixel-based Encoder of Language, which suffers from neither of these issues. PIXEL is a pretrained language model that renders text as images, making it possible to transfer representations across languages based on orthographic similarity or the co-activation of pixels. PIXEL is trained to reconstruct the pixels of masked patches, instead of predicting a distribution over tokens.1 We pretrain the 86M parameter PIXEL model on the same English data as BERT and evaluate on syntactic and semantic tasks in typologically diverse languages, including various non-Latin scripts. We find that PIXEL substantially outperforms BERT on syntactic and semantic processing tasks on scripts that are not found in the pretraining data, but PIXEL is slightly weaker than BERT when working with Latin scripts. Furthermore, we find that PIXEL is more robust to noisy text inputs than BERT, further confirming the benefits of modelling language with pixels.		
versions		**versions**			