cases	doc_1		doc_2		decision	id
	authors	• Jae Jun Lee • Raphael Tang • Jun Lin	authors	• Jaejun Lee • Raphael Tang • Jimmy J. Lin	DUPLICATES	324
	title	What Would Elsa Do? Freezing Layers During Transformer Fine-Tuning	title	What Would Elsa Do? Freezing Layers During Transformer Fine-Tuning		
	publication_date	2019-11-08 00:00:00	publication_date	2019-11-08 00:00:00		
	source	SupportedSources.OPENALEX	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	arXiv (Cornell University)	journal	ArXiv		
	volume		volume	abs/1911.03090		
	doi	None	doi			
	urls	• https://openalex.org/W2989195139	urls	• https://www.semanticscholar.org/paper/4a4646a5ce6b57e369403e4efea1a2e4559fe9f1		
	id	id6139618154169143958	id	id1482233225642636164		
	abstract		abstract	Pretrained transformer-based language models have achieved state of the art across countless tasks in natural language processing. These models are highly expressive, comprising at least a hundred million parameters and a dozen layers. Recent evidence suggests that only a few of the final layers need to be fine-tuned for high quality on downstream tasks. Naturally, a subsequent research question is, "how many of the last layers do we need to fine-tune?" In this paper, we precisely answer this question. We examine two recent pretrained language models, BERT and RoBERTa, across standard tasks in textual entailment, semantic similarity, sentiment analysis, and linguistic acceptability. We vary the number of final layers that are fine-tuned, then study the resulting change in task-specific effectiveness. We show that only a fourth of the final layers need to be fine-tuned to achieve 90% of the original quality. Surprisingly, we also find that fine-tuning all layers does not always help.		
	versions		versions			