We thank the reviewers for their constructive feedback. Reviewers found our method to be novel (rev. 1,2,3), clearly presented (rev. 1,2), sensible and well-motivated (rev. 1,2,3), and having good empirical performance (rev. 1,2,3). The reviewers' main concern was about the need for additional baselines.

Additional Baselines. To address this main concern, we are modifying all tables to include results from earlier publications. On the text classification task, this demonstrates that we achieve performance competitive with the SOTA using significantly fewer parameters. On the super-resolution tasks, we achieve or surpass current SOTA results.

On the Yelp-2 and Yelp-5 datasets, TFiLM achieves performance competitive with SOTA using fewer parameters (Table 1). The final paper will also include a larger TFiLM model, attempting to surpass SOTA, and more datasets (we did not have time to do this for the rebuttal). We also include the linear FastText baseline.

Method	Yelp-2	Yelp-5	Param
FastText [Grave et al., 2017]	95.7%	63.9%	Linear
LSTM [Yogatama et al., 2017]	92.6%	59.6%	-
Self-Attention [Lin et al., 2017]	93.5%	63.4%	-
CNN [Kim, 2014]	93.5%	61.0%	-
CharCNN [Zhang et al., 2015]	94.6%	62.0%	-
VDCNN [Conneau et al., 2017]	95.4%	64.7%	>5M
DenseCNN [Wang et al., 2018]	96.0%	64.5%	>4M
DPCNN* [Rie, Johnson, 2017]	97.36%	69.4%	>3M
BERT* [Devlin et al., 2018]	98.11%	70.68%	-
SmallCNN (ours)	78.1%	61.5	<1.5M
SmallCNN+TFiLM (ours)	95.6%	62.3	1.5M

Table 1: Text classification on Yelp-2 and Yelp-5 datasets. Methods with * use unsupervised pre-training (unsupervised region embeddings or transformers) on external data and are not directly comparable. Parameter counts exclude models with lower performance. Embeddings are not counted.

On the genomics super-resolution task, our method improves over the SOTA results of Koh et al. (2017). This task was introduced by Koh et al.; Table 2 reports their baseline, proposed model, our re-implementation of their model, and our new SOTA result.

On the audio-super resolution tasks, our two existing baselines already correspond to the DNN-based method of Li et al. (2015) (we re-implemented it) and the CNN-based method from Kuleshov et al. (2017) (using the provided source code). Our new Table 3 reflects this comparison to standard models. Note also that we already report the results of the cubic spline (interpolation) baseline.

Histone	Input [K17]	Linear [K17]	CNN [K17]	CNN Us	**Full Us**
H3K4me1	0.37	0.41	0.59	0.79	0.81
H3K4me3	0.63	0.67	0.72	0.66	0.90
H3K27ac	0.55	0.61	0.77	0.85	0.89
H3K27me3	0.14	0.18	0.30	0.65	0.64

Table 2: Genomic super-resolution. [K17] indicates results from Koh et al. (2017); linear method performance is estimated.

Additional baselines are difficult to add, since there is no standard audio super-resolution benchmark. DRCNN is an image super-resolution method and its extension to audio is outside of the scope of our paper. The Wavenet paper cited by Reviewer 1 only performs two single-speaker experiments and uses a different experimental setup, that we didn't have time to reproduce. We anticipate our performance to be competitive but somewhat lower (they report an LSD of 2.5; our single-speaker experiment has 3.4; their CNN baselines are 4.0 and 4.5). The U-Net baseline cited by Reviewer 1 is relevant, but almost identical to our "CNN" baseline (Kuleshov et al., 2017). We will cite all of these papers and we thank Reviewer 1 for bringing them to our attention.

Left-to-Right Processing. Reviewer 2 is right that TFiLM can use a bidirectional RNN. In some applications – like real-time audio super-resolution – samples from the future may not be accessible; therefore, we left the RNN uni-directional for full generality. However, we agree that using a bidirectional RNN is better for presentation, and will do so in the paper.

Ratio	Obj.	Spline	DNN [Li et al.]	Conv [KEE17]	**Full Us**
$r = 2$	SNR	18.0	17.9	18.1	19.8
	LSD	2.9	2.5	1.9	1.8
$r = 4$	SNR	13.2	13.3	13.1	15.0
	LSD	5.2	3.9	3.1	2.7
$r = 8$	SNR	9.8	9.8	9.9	12.0
	LSD	6.8	4.6	4.3	2.9

Table 3: Audio super-resolution. DNN and CNN are baselines from the literature. [KEE17] denotes the convolutional method of Kuleshov et al. (2017)

Architecture Questions. The effects of removing the additive skip connection are shown in Figure 5. The model trains much more slowly and achieves somewhat lower performance. Bypassing the TFiLM layer would revert to a pure Spline model, whose performance we report. We also report the performance of cubic interpolation, which is the same as "[cubic] Spline". In our experiments, we were able to run audio super-resolution inference faster than real time, using <1sec for >30sec of audio in <1sec; we will add more detailed analysis in the final paper.

Missing Citations. We have added a Transformer baseline and a citation. We have also added citations to audio super-resolution papers (including the Wavenet one). We thank Reviewer 2 for brining the Squeeze-and-Excitation paper to our attention; we are citing it, as well as FiLM for VQA.