

Jasper: An End-to-End Convolutional Neural Acoustic Model

Geoffroy B. de Felcourt

April 2020

- What is Jasper?
- Architecture
- Results
- Application of jasper
- Conclusion

What is Jasper ?

- Speech recognition Model
- Jasper uses only 1D convolutions, batch normalization, ReLU, dropout.
- State of the art on LibriSpeech

Architecture

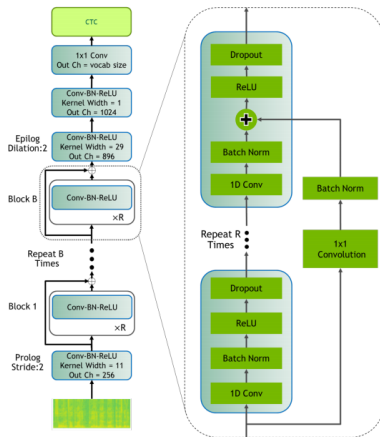


Figure 1: *Jasper BxR* model: B - number of blocks, R - number of sub-blocks.

Table 5: LibriSpeech, WER (%)

Model	E2E	LM	dev-clean	dev-other	test-clean	test-other
CAPIO (single) [23]	N	RNN	3.02	8.28	3.56	8.58
pFSMN-Chain [25]	N	RNN	2.56	7.47	2.97	7.5
DeepSpeech2 [26]	Y	5-gram	-	-	5.33	13.25
Deep bLSTM w/ attention [21]	Y	LSTM	3.54	11.52	3.82	12.76
wav2letter++ [27]	Y	ConvLM	3.16	10.05	3.44	11.24
LAS + SpecAugment ⁴ [28]	Y	RNN	-	-	2.5	5.8
Jasper DR 10x5	Y	-	3.64	11.89	3.86	11.95
Jasper DR 10x5	Y	6-gram	2.89	9.53	3.34	9.62
Jasper DR 10x5	Y	Transformer-XL	2.68	8.62	2.95	8.79
Jasper DR 10x5 + Time/Freq Masks ⁴	Y	Transformer-XL	2.62	7.61	2.84	7.84

Application of Jasper

Enterprise voice recognition devices

Google Home / Alexa devices

DIY

Raspberry Pis

Jason Li, Vitaly Lavrukhin, Boris Ginsburg, Ryan Leary, Oleksii Kuchaiev,
Jonathan M. Cohen¹, Huyen Nguyen, Ravi Teja Gadde