{'loss': 4.4577, 'grad\_norm': 93.62213897705078, 'learning\_rate': 9e-06, 'epoch': 0.33}

3%|█▋ | 54/2000 [04:44<2:49:18, 5.22s/it] {'loss': 1.6249, 'grad\_norm': 16.67357635498047, 'learning\_rate': 1e-05, 'epoch': 0.65}

{'loss': 0.9137, 'grad\_norm': 14.290881156921387, 'learning\_rate': 1e-05, 'epoch': 0.98}

{'loss': 0.3486, 'grad\_norm': 4.672069549560547, 'learning\_rate': 1e-05, 'epoch': 1.3}

{'loss': 0.1874, 'grad\_norm': 4.123934745788574, 'learning\_rate': 1e-05, 'epoch': 1.63}

{'loss': 0.1586, 'grad\_norm': 4.4663004875183105, 'learning\_rate': 1e-05, 'epoch': 1.95}

{'loss': 0.1124, 'grad\_norm': 2.6569292545318604, 'learning\_rate': 1e-05, 'epoch': 2.28}

{'loss': 0.1062, 'grad\_norm': 3.143254518508911, 'learning\_rate': 1e-05, 'epoch': 2.61}

{'loss': 0.1055, 'grad\_norm': 2.550719738006592, 'learning\_rate': 1e-05, 'epoch': 2.93}

{'loss': 0.0737, 'grad\_norm': 2.2131810188293457, 'learning\_rate': 1e-05, 'epoch': 3.26}

25%|████████████████████████████████████████████████▊ | 500/2000 [43:00<2:09:30, 5.18s/it]

{'eval\_loss': 0.1696450561285019, 'eval\_wer\_ortho': 60.45150501672241, 'eval\_wer': 12.71460346145826, 'eval\_runtime': 1195.5548, 'eval\_samples\_per\_second': 1.849, 'eval\_steps\_per\_second': 0.116, 'epoch': 3.26}

25%|████████████████████████████████████████████████▎ | 500/2000 [1:02:56<2:09:30, 5.18s/itSome non-default generation parameters are set in the model config. These should go into a GenerationConfig file (https://huggingface.co/docs/transformers/generation\_strategies#save-a-custom-decoding-strategy-with-your-model) instead. This warning will be raised to an exception in v4.41.

Non-default generation parameters: {'max\_length': 448, 'suppress\_tokens': [1, 2, 7, 8, 9, 10, 14, 25, 26, 27, 28, 29, 31, 58, 59, 60, 61, 62, 63, 90, 91, 92, 93, 359, 503, 522, 542, 873, 893, 902, 918, 922, 931, 1350, 1853, 1982, 2460, 2627, 3246, 3253, 3268, 3536, 3846, 3961, 4183, 4667, 6585, 6647, 7273, 9061, 9383, 10428, 10929, 11938, 12033, 12331, 12562, 13793, 14157, 14635, 15265, 15618, 16553, 16604, 18362, 18956, 20075, 21675, 22520, 26130, 26161, 26435, 28279, 29464, 31650, 32302, 32470, 36865, 42863, 47425, 49870, 50254, 50258, 50360, 50361, 50362], 'begin\_suppress\_tokens': [220, 50257]}

/root/miniconda3/envs/speech-to-text/lib/python3.11/site-packages/torch/utils/checkpoint.py:460: UserWarning: torch.utils.checkpoint: please pass in use\_reentrant=True or use\_reentrant=False explicitly. The default value of use\_reentrant will be updated to be False in the future. To maintain current behavior, pass use\_reentrant=True. It is recommended that you use use\_reentrant=False. Refer to docs for more details on the differences between the two variants.

warnings.warn(

26%|████████████████████████████████████ | 527/2000 [1:05:21<2:08:18, 5.23s/it]{'loss': 0.0682, 'grad\_norm': 1.9584933519363403, 'learning\_rate': 1e-05, 'epoch': 3.58}

{'loss': 0.0688, 'grad\_norm': 1.5681066513061523, 'learning\_rate': 1e-05, 'epoch': 3.91}

{'loss': 0.0501, 'grad\_norm': 2.121187686920166, 'learning\_rate': 1e-05, 'epoch': 4.23}

{'loss': 0.0453, 'grad\_norm': 1.7989155054092407, 'learning\_rate': 1e-05, 'epoch': 4.56}

{'loss': 0.0455, 'grad\_norm': 1.8472970724105835, 'learning\_rate': 1e-05, 'epoch': 4.89}

{'loss': 0.0339, 'grad\_norm': 2.280743360519409, 'learning\_rate': 1e-05, 'epoch': 5.21}

{'loss': 0.0254, 'grad\_norm': 1.6480897665023804, 'learning\_rate': 1e-05, 'epoch': 5.54}

{'loss': 0.0284, 'grad\_norm': 1.0490167140960693, 'learning\_rate': 1e-05, 'epoch': 5.86}

{'loss': 0.0224, 'grad\_norm': 0.726561427116394, 'learning\_rate': 1e-05, 'epoch': 6.19}

{'loss': 0.0153, 'grad\_norm': 1.1269220113754272, 'learning\_rate': 1e-05, 'epoch': 6.51}

{'eval\_loss': 0.21759513020515442, 'eval\_wer\_ortho': 59.43422519509476, 'eval\_wer': 11.63550427469243, 'eval\_runtime': 1239.0333, 'eval\_samples\_per\_second': 1.784, 'eval\_steps\_per\_second': 0.112, 'epoch': 6.51}

50%|████████████████████████████████████████████████████████████████████ | 1000/2000 [2:07:14<1:27:25, 5.25s/itSome non-default generation parameters are set in the model config. These should go into a GenerationConfig file (https://huggingface.co/docs/transformers/generation\_strategies#save-a-custom-decoding-strategy-with-your-model) instead. This warning will be raised to an exception in v4.41.

Non-default generation parameters: {'max\_length': 448, 'suppress\_tokens': [1, 2, 7, 8, 9, 10, 14, 25, 26, 27, 28, 29, 31, 58, 59, 60, 61, 62, 63, 90, 91, 92, 93, 359, 503, 522, 542, 873, 893, 902, 918, 922, 931, 1350, 1853, 1982, 2460, 2627, 3246, 3253, 3268, 3536, 3846, 3961, 4183, 4667, 6585, 6647, 7273, 9061, 9383, 10428, 10929, 11938, 12033, 12331, 12562, 13793, 14157, 14635, 15265, 15618, 16553, 16604, 18362, 18956, 20075, 21675, 22520, 26130, 26161, 26435, 28279, 29464, 31650, 32302, 32470, 36865, 42863, 47425, 49870, 50254, 50258, 50360, 50361, 50362], 'begin\_suppress\_tokens': [220, 50257]}

/root/miniconda3/envs/speech-to-text/lib/python3.11/site-packages/torch/utils/checkpoint.py:460: UserWarning: torch.utils.checkpoint: please pass in use\_reentrant=True or use\_reentrant=False explicitly. The default value of use\_reentrant will be updated to be False in the future. To maintain current behavior, pass use\_reentrant=True. It is recommended that you use use\_reentrant=False. Refer to docs for more details on the differences between the two variants.

warnings.warn(

{'loss': 0.018, 'grad\_norm': 2.585179328918457, 'learning\_rate': 1e-05, 'epoch': 6.84}

{'loss': 0.0136, 'grad\_norm': 1.3137625455856323, 'learning\_rate': 1e-05, 'epoch': 7.17}

{'loss': 0.0108, 'grad\_norm': 1.828255295753479, 'learning\_rate': 1e-05, 'epoch': 7.49}

{'loss': 0.0107, 'grad\_norm': 1.2773793935775757, 'learning\_rate': 1e-05, 'epoch': 7.82}

{'loss': 0.0097, 'grad\_norm': 0.8917009234428406, 'learning\_rate': 1e-05, 'epoch': 8.14}

{'loss': 0.0067, 'grad\_norm': 1.005906105041504, 'learning\_rate': 1e-05, 'epoch': 8.47}

{'loss': 0.0082, 'grad\_norm': 1.2144718170166016, 'learning\_rate': 1e-05, 'epoch': 8.79}

{'loss': 0.0075, 'grad\_norm': 0.7672501802444458, 'learning\_rate': 1e-05, 'epoch': 9.12}

{'loss': 0.0062, 'grad\_norm': 0.8114225268363953, 'learning\_rate': 1e-05, 'epoch': 9.45}

{'loss': 0.0057, 'grad\_norm': 1.1349878311157227, 'learning\_rate': 1e-05, 'epoch': 9.77}

{'eval\_loss': 0.26541873812675476, 'eval\_wer\_ortho': 59.476031215161655, 'eval\_wer': 11.675470911239314, 'eval\_runtime': 1171.5326, 'eval\_samples\_per\_second': 1.887, 'eval\_steps\_per\_second': 0.119, 'epoch': 9.77}

75%|███████████████████████████████████████████████████████████████████████████████████████████████████████▌ | 1500/2000 [3:10:34<42:56, 5.15s/itSome non-default generation parameters are set in the model config. These should go into a GenerationConfig file (https://huggingface.co/docs/transformers/generation\_strategies#save-a-custom-decoding-strategy-with-your-model) instead. This warning will be raised to an exception in v4.41.

Non-default generation parameters: {'max\_length': 448, 'suppress\_tokens': [1, 2, 7, 8, 9, 10, 14, 25, 26, 27, 28, 29, 31, 58, 59, 60, 61, 62, 63, 90, 91, 92, 93, 359, 503, 522, 542, 873, 893, 902, 918, 922, 931, 1350, 1853, 1982, 2460, 2627, 3246, 3253, 3268, 3536, 3846, 3961, 4183, 4667, 6585, 6647, 7273, 9061, 9383, 10428, 10929, 11938, 12033, 12331, 12562, 13793, 14157, 14635, 15265, 15618, 16553, 16604, 18362, 18956, 20075, 21675, 22520, 26130, 26161, 26435, 28279, 29464, 31650, 32302, 32470, 36865, 42863, 47425, 49870, 50254, 50258, 50360, 50361, 50362], 'begin\_suppress\_tokens': [220, 50257]}

/root/miniconda3/envs/speech-to-text/lib/python3.11/site-packages/torch/utils/checkpoint.py:460: UserWarning: torch.utils.checkpoint: please pass in use\_reentrant=True or use\_reentrant=False explicitly. The default value of use\_reentrant will be updated to be False in the future. To maintain current behavior, pass use\_reentrant=True. It is recommended that you use use\_reentrant=False. Refer to docs for more details on the differences between the two variants.

warnings.warn(

{'loss': 0.0054, 'grad\_norm': 0.872553825378418, 'learning\_rate': 1e-05, 'epoch': 10.1}

{'loss': 0.0045, 'grad\_norm': 1.450510025024414, 'learning\_rate': 1e-05, 'epoch': 10.42}

{'loss': 0.0054, 'grad\_norm': 2.3232102394104004, 'learning\_rate': 1e-05, 'epoch': 10.75}

{'loss': 0.005, 'grad\_norm': 0.9737334847450256, 'learning\_rate': 1e-05, 'epoch': 11.07}

{'loss': 0.0034, 'grad\_norm': 1.131380319595337, 'learning\_rate': 1e-05, 'epoch': 11.4}

{'loss': 0.0043, 'grad\_norm': 0.2664770185947418, 'learning\_rate': 1e-05, 'epoch': 11.73}

{'loss': 0.0041, 'grad\_norm': 1.0037014484405518, 'learning\_rate': 1e-05, 'epoch': 12.05}

{'loss': 0.0034, 'grad\_norm': 0.6123923659324646, 'learning\_rate': 1e-05, 'epoch': 12.38}

{'loss': 0.0038, 'grad\_norm': 1.287395715713501, 'learning\_rate': 1e-05, 'epoch': 12.7}

{'loss': 0.0039, 'grad\_norm': 0.6637178063392639, 'learning\_rate': 1e-05, 'epoch': 13.03}

{'eval\_loss': 0.2834605574607849, 'eval\_wer\_ortho': 59.12764771460424, 'eval\_wer': 11.760617223882672, 'eval\_runtime': 1169.9682, 'eval\_samples\_per\_second': 1.89, 'eval\_steps\_per\_second': 0.119, 'epoch': 13.03}

100%|██████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2000/2000 [4:13:50<00:00, 5.15s/itSome non-default generation parameters are set in the model config. These should go into a GenerationConfig file (https://huggingface.co/docs/transformers/generation\_strategies#save-a-custom-decoding-strategy-with-your-model) instead. This warning will be raised to an exception in v4.41.

Non-default generation parameters: {'max\_length': 448, 'suppress\_tokens': [1, 2, 7, 8, 9, 10, 14, 25, 26, 27, 28, 29, 31, 58, 59, 60, 61, 62, 63, 90, 91, 92, 93, 359, 503, 522, 542, 873, 893, 902, 918, 922, 931, 1350, 1853, 1982, 2460, 2627, 3246, 3253, 3268, 3536, 3846, 3961, 4183, 4667, 6585, 6647, 7273, 9061, 9383, 10428, 10929, 11938, 12033, 12331, 12562, 13793, 14157, 14635, 15265, 15618, 16553, 16604, 18362, 18956, 20075, 21675, 22520, 26130, 26161, 26435, 28279, 29464, 31650, 32302, 32470, 36865, 42863, 47425, 49870, 50254, 50258, 50360, 50361, 50362], 'begin\_suppress\_tokens': [220, 50257]}

There were missing keys in the checkpoint model loaded: ['proj\_out.weight'].

{'train\_runtime': 15235.0567, 'train\_samples\_per\_second': 4.201, 'train\_steps\_per\_second': 0.131, 'train\_loss': 0.2158042760118842, 'epoch': 13.03}

100%|██████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2000/2000 [4:13:55<00:00, 7.62s/it]

Training completed!