Module 8: Portfolio Project

Option #2: Encoder-Decoder Models

Ryan Thompson

Colorado State University - Global

CSC 580

Dr. Farr

22 July 2025

**Encoder-Decoder Models**

Encoder-decoder models, originally developed for sequence-to-sequence tasks in natural language processing, have become a versatile tool across a variety of industries. By mapping complex input data into abstract latent representations and decoding them into structured outputs, encoder-decoder architectures have enabled major advances in fields ranging from language and vision to healthcare and industrial monitoring. This paper highlights four distinct, industry relevant use cases. The use cases include neural machine translation, image captioning, early detection of osteoporosis with image segmentation, and anomaly detection in manufacturing. For each, the benefits of using encoder-decoder models are discussed.

Encoder-decoder models are perhaps most well known for their transformative impact on machine translation. In neural machine translation (NMT), an encoder converts a sentence in the source language into a fixed length context vector that captures its meaning, and a decoder generates the translated sentence in the target language. This approach, first demonstrated by Sutskever et al. (2014), outperformed previous statistical methods by capturing long range dependencies and semantic context, resulting in more fluent translations. Modern NMT systems, such as those used by Google Translate, rely on this architecture to handle diverse language pairs and deliver accurate, real time translations at scale (Bahdanau et al., 2015). The primary benefit is the ability to learn directly from raw text pairs, reducing the need for hand crafted linguistic rules.

Another important application is automated image captioning, where encoder-decoder models enable systems to generate descriptive sentences for images. A convolutional neural network (CNN) acts as the encoder, extracting visual features, while a recurrent neural network (RNN) or transformer serves as the decoder, producing captions word by word. This technology supports accessibility for visually impaired users, enhances digital asset management. And thus improves content moderation on social media platforms (Vinyals et al., 2015). The benefit lies in the model’s ability to jointly learn visual and textual representations, allowing it to map complex images to coherent, contextually appropriate descriptions.

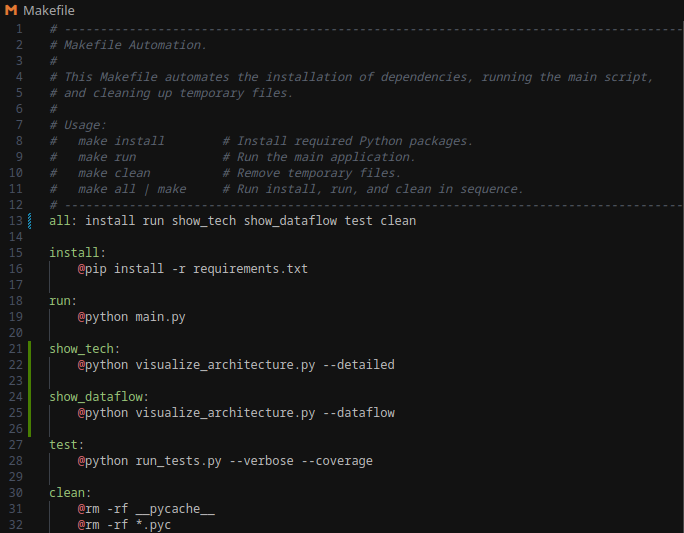
In medical imaging, encoder-decoder models play a crucial role in early detection of diseases such as osteoporosis. Specifically, convolutional encoder-decoder architectures such as U-Net are widely used for image segmentation in X-ray or MRI scans of bones. The encoder extracts hierarchical features from the medical image, and the decoder reconstructs a segmentation map that highlights bone regions. By accurately segmenting trabecular and cortical bone areas, these models help clinicians assess bone density and structural integrity, enabling earlier detection of osteoporosis before severe damage occurs (Çiçek et al., 2016). Automated segmentation accelerates analysis, reduces human error, and provides quantitative metrics for monitoring disease progression. This technology is being integrated into radiology workflows to support faster, more objective diagnoses and treatment planning.

Encoder-decoder models are also applied in industrial manufacturing, especially for anomaly detection using time series data from sensors. Here, a sequence-to-sequence autoencoder learns to reconstruct normal operational sequences from multivariate sensor data. When anomalous or faulty data is encountered, reconstruction errors spike, signaling possible malfunctions or process deviations (Zhao et al., 2017). This approach is beneficial because it captures temporal patterns and correlations in the data, enabling early warning for equipment failures and reducing downtime. Industries such as automotive, energy, and electronics use these models to support predictive maintenance and improve product quality.

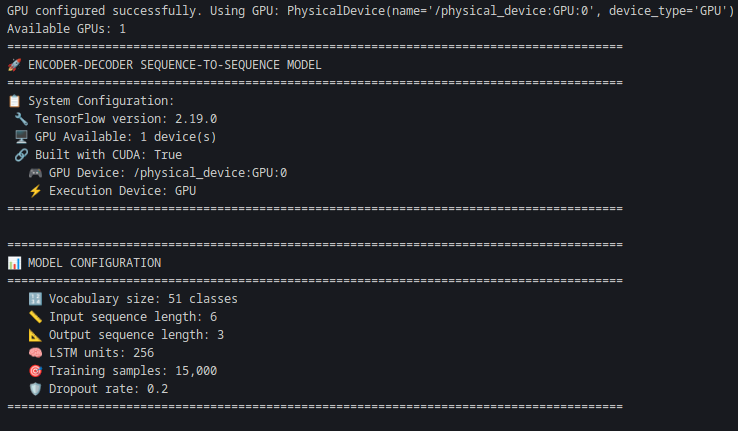
Encoder-decoder models have proven to be powerful and adaptable, enabling advancements across a range of sectors including language processing, computer vision, healthcare, and manufacturing. Their core strength is the ability to encode complex input structures and generate meaningful outputs tailored to the application. As research continues, encoder-decoder architectures will likely drive further innovation in both established and emerging industries.

Program Outputs

Listed below is a collection of screenshots taken from the execution of the program in its entirety. Some of the images may require zooming in.



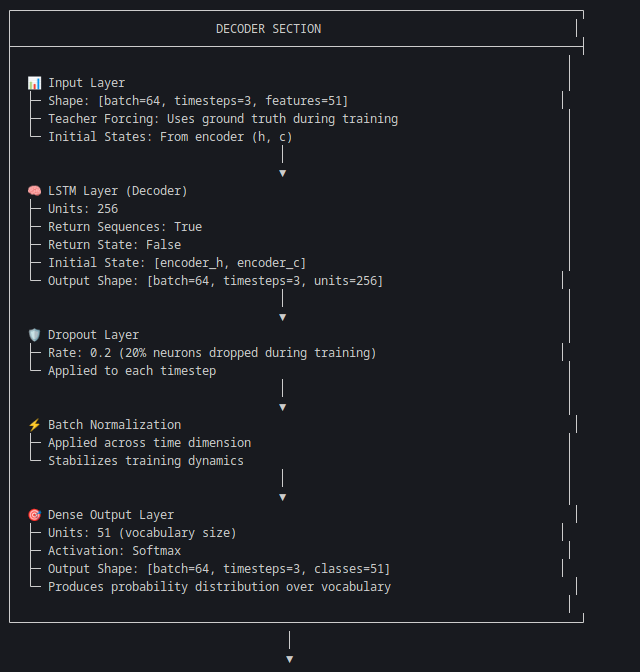
*Makefile Automation to run everything or select targets.*



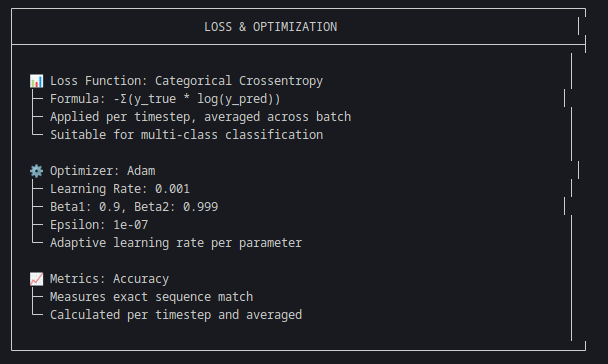
*Initial model configuration.*



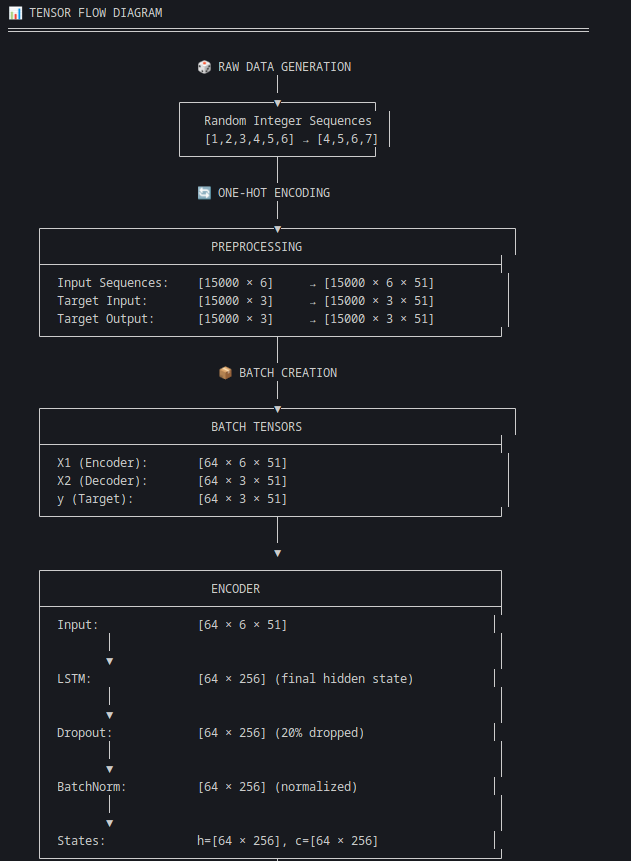
*Encoder-Decoder Architecture (1 of 3).*



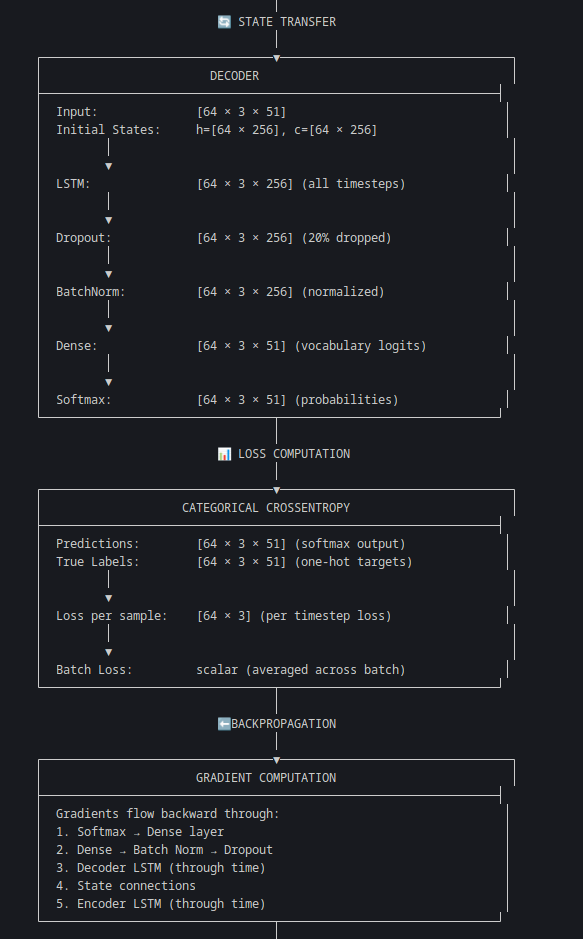
*Encoder-Decoder Architecture (2 of 3).*



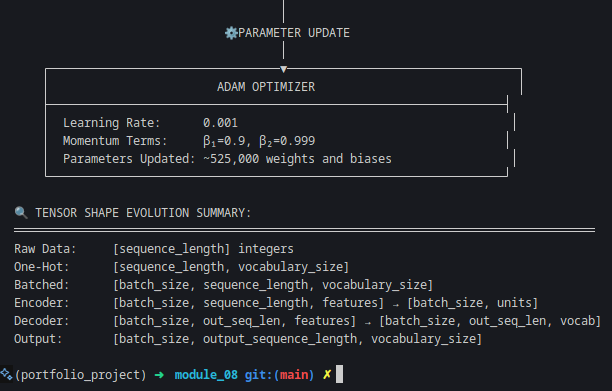
*Encoder-Decoder Architecture (3 of 3).*

**

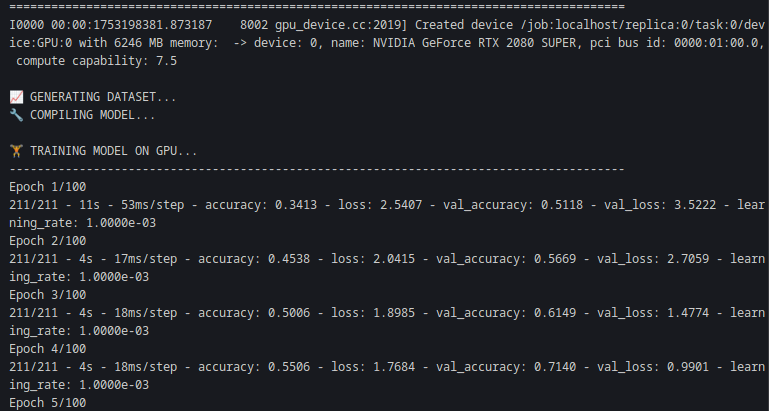
*Dataflow diagram (1 of 3).*

**

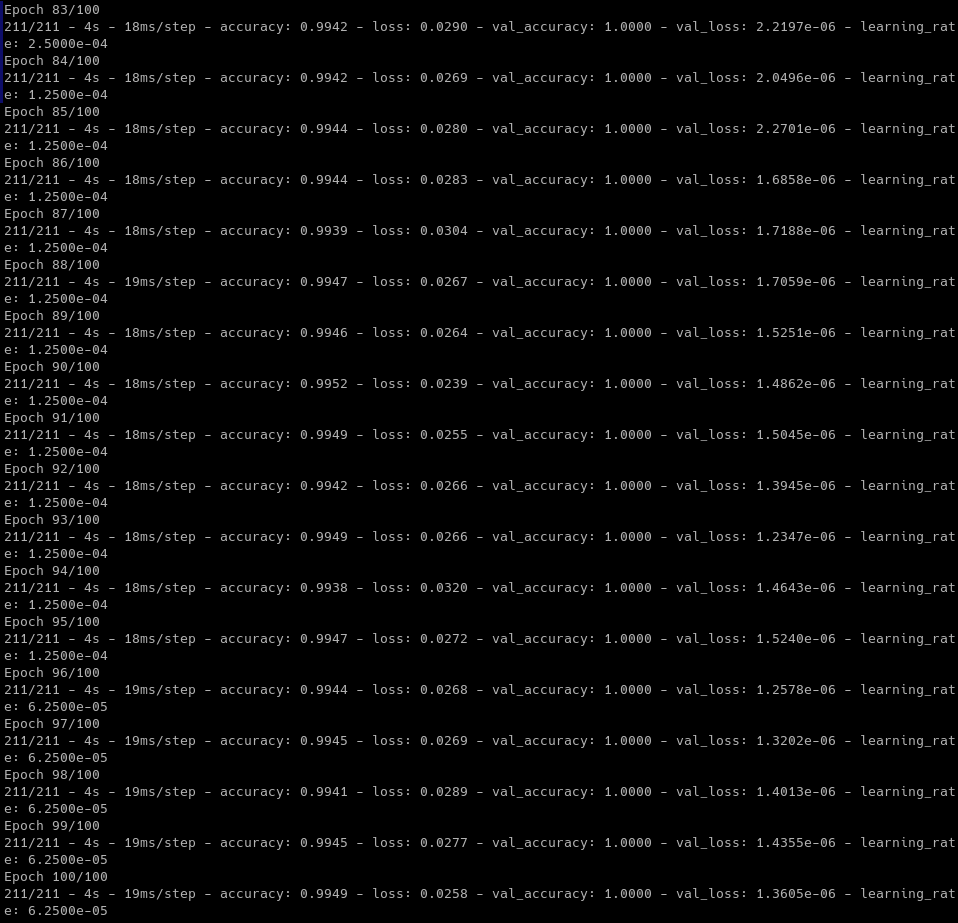
*Dataflow diagram (2 of 3).*

**

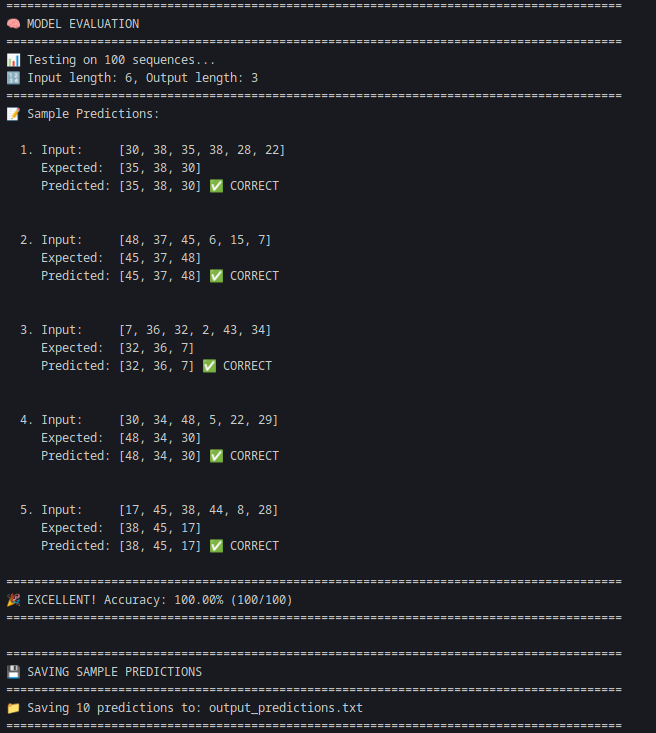
*Dataflow diagram (3 of 3).*



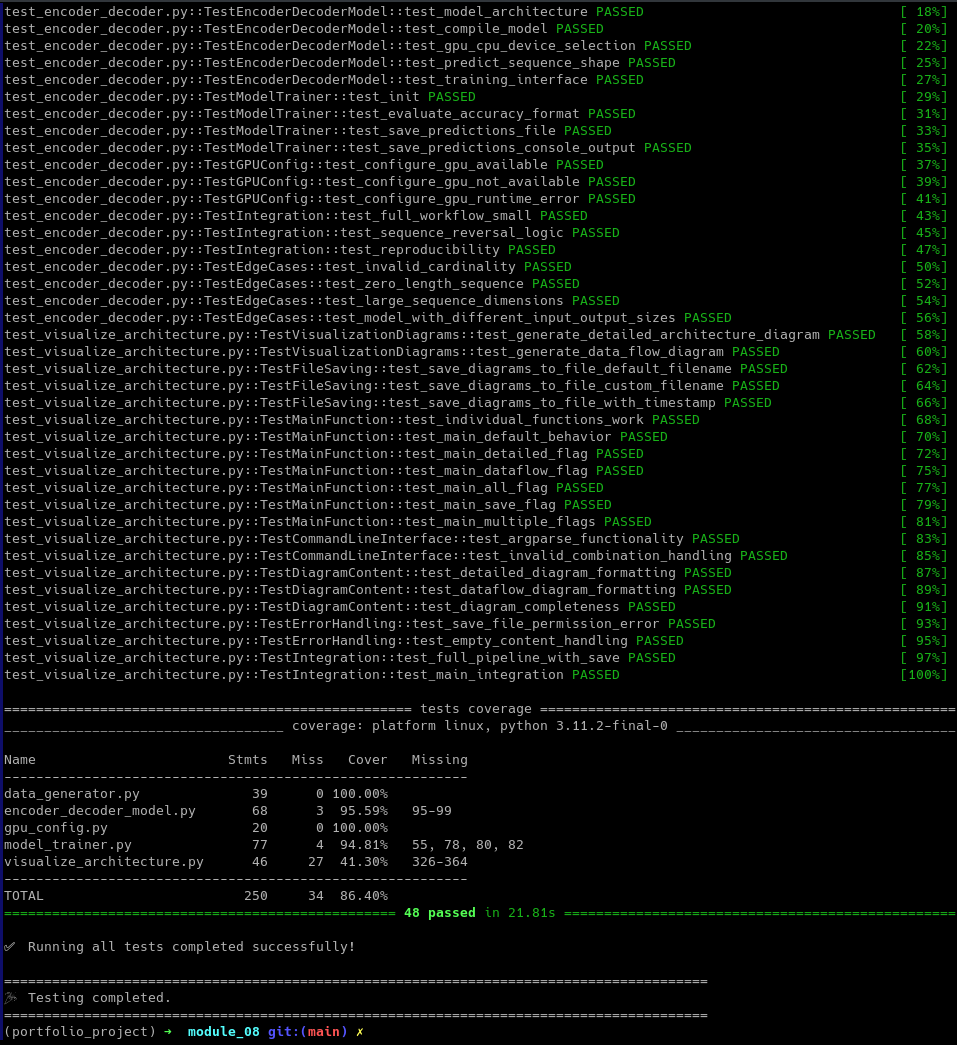
*Initialize training.*

**

*Finished training with 100 epochs.*



*Model series predictions.*



*Testing the code and coverage with pytest.*

**References**

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*.<https://arxiv.org/abs/1409.0473>

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 424–432. https://doi.org/10.1007/978-3-319-46723-8\_49

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 3104–3112.<https://arxiv.org/abs/1409.3215>

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3156–3164. https://doi.org/10.1109/CVPR.2015.7298935

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2017). Deep learning and its applications to machine health monitoring: A survey. *Mechanical Systems and Signal Processing, 115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050