**AR-CNN with AWS DeepComposer**

**Summary**

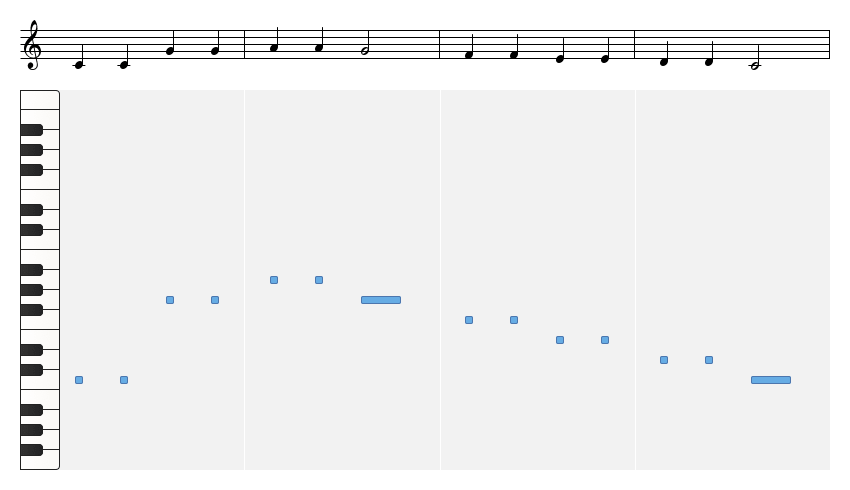
Our next popular generative model is the autoregressive convolutional neural network (AR-CNN). Autoregressive convolutional neural networks make iterative changes over time to create new data.

To better understand how the AR-CNN model works, let’s first discuss how music is represented so it is machine-readable.

**Image-based representation**

Nearly all machine learning algorithms operate on data as numbers or sequences of numbers. In AWS DeepComposer, the input tracks are represented as a *piano roll\*\**. \**In each two-dimensional piano roll,* time *is on the horizontal axis and* pitch\* is on the vertical axis. You might notice this representation looks similar to an image.

The AR-CNN model uses a piano roll image to represent the audio files from the dataset. You can see an example in the following image where on top is a musical score and below is a piano roll image of that same score.



Musical score and piano roll

**How the AR-CNN Model Works**

When a note is either added or removed from your input track during inference, we call it an *edit event*. To train the AR-CNN model to predict when notes need to be added or removed from your input track (edit event), the model **iteratively** updates the input track to sounds more like the training dataset. During training, the model is also challenged to detect differences between an original piano roll and a newly modified piano roll.

**New Terms**

* **Piano roll:** A two-dimensional piano roll matrix that represents input tracks. Time is on the horizontal axis and pitch is on the vertical axis.
* **Edit event:** When a note is either added or removed from your input track during inference.

;