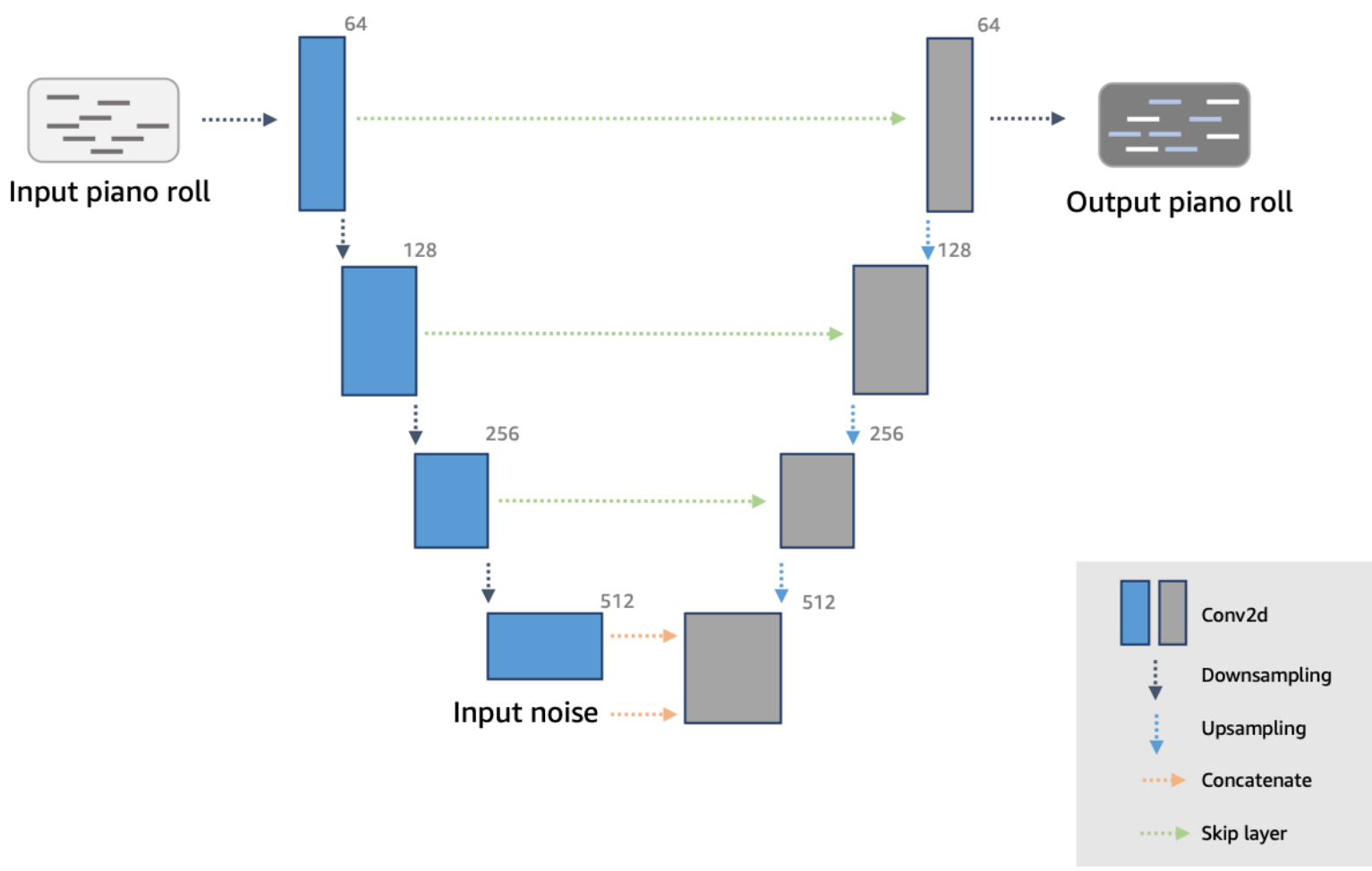


Generator

The generator network used in AWS DeepComposer is adapted from the U-Net architecture, a popular convolutional neural network that is used extensively in the computer vision domain. The network consists of an “encoder” that maps the single track music data (represented as piano roll images) to a relatively lower dimensional “latent space” and a “decoder” that maps the latent space back to multi-track music data.

Here are the inputs provided to the generator:

- Single-track piano roll: A single melody track is provided as the input to the generator.
- Noise vector: A latent noise vector is also passed in as an input and this is responsible for ensuring that there is a flavor to each output generated by the generator, even when the same input is provided.

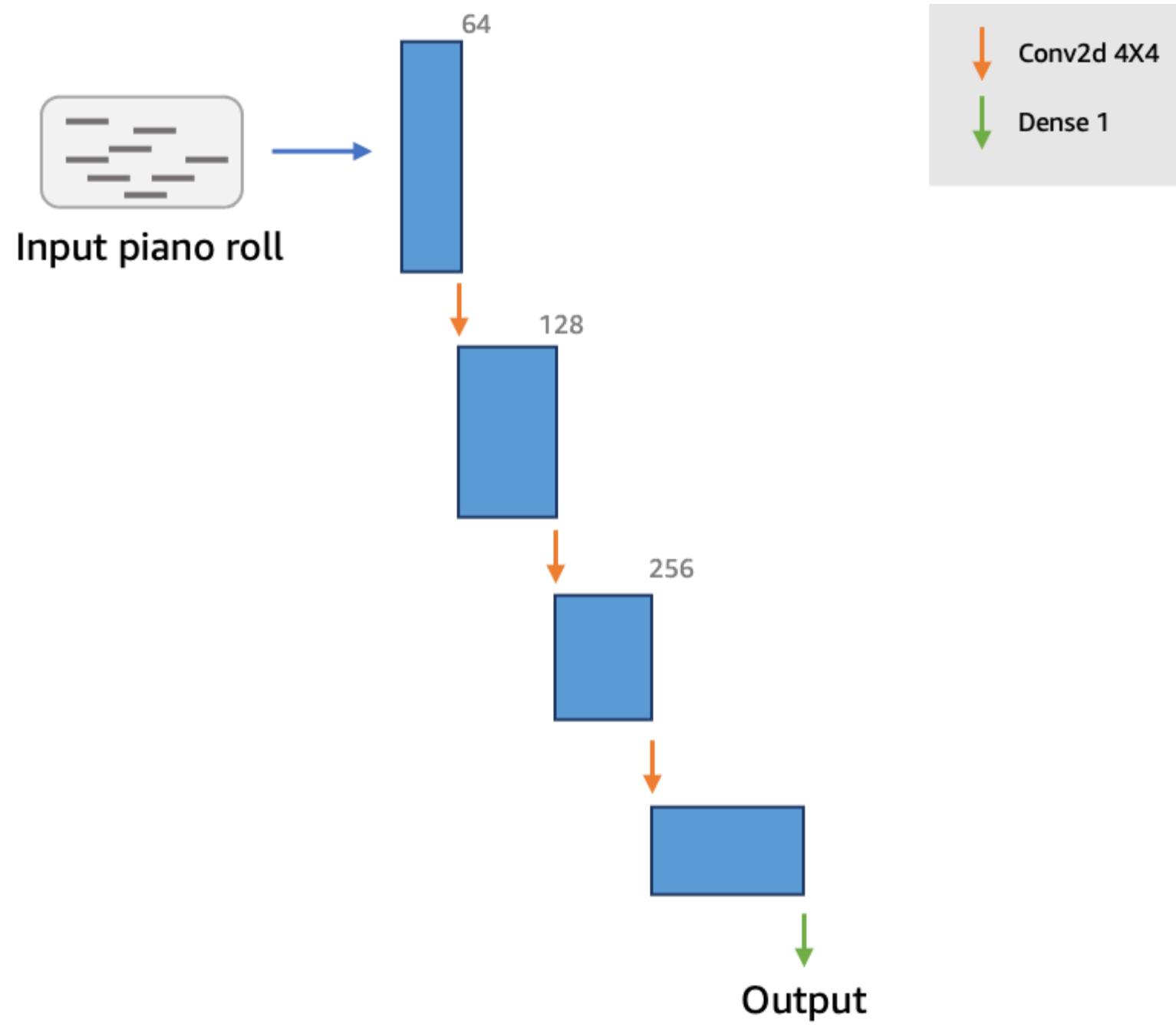


Notice that the encoding layers of the generator on the left side and decoder layer on on the right side are connected to create a **U-shape**, thereby giving the name U-Net to this architecture

Discriminator

The goal of the discriminator is to provide feedback to the generator about how realistic the generated piano rolls are, so that the generator can learn to produce more realistic data. The discriminator provides this feedback by outputting a scalar value that represents how “real” or “fake” a piano roll is.

Since the discriminator tries to classify data as “real” or “fake”, it is not very different from commonly used binary classifiers. We use a simple architecture for the critic, composed of four convolutional layers and a dense layer at the end.



Discriminator network architecture consisting of four convolutional layers and a dense layer

Once you complete building your model architecture, the next step is training.

QUESTION 1 OF 4

What is another term for the “discriminator” in a GAN?

☐ Truth

☒ Critic

☐ Polygraph

☐ The man behind the curtain

SUBMIT

QUESTION 2 OF 4

Which attribute(s) does a discriminator judge?

☒ Reality

☐ Quality

☐ Quantity

☐ Robustness

SUBMIT

QUESTION 3 OF 4

What would happen without “input noise”?

☐ U-Net would not function

☐ The output would be exactly the same as the input

☒ The same input would produce consistently similar output.

☐ U-Net would produce images instead of sound

SUBMIT

QUESTION 4 OF 4

Order the steps in U-Net architecture

Submit to check your answer choices!

| ORDER | STEP |
|-------|--------------|
| 1 | Input |
| 2 | Encoder |
| 3 | Latent Space |
| 4 | Decoder |
| 5 | Output |

SUBMIT