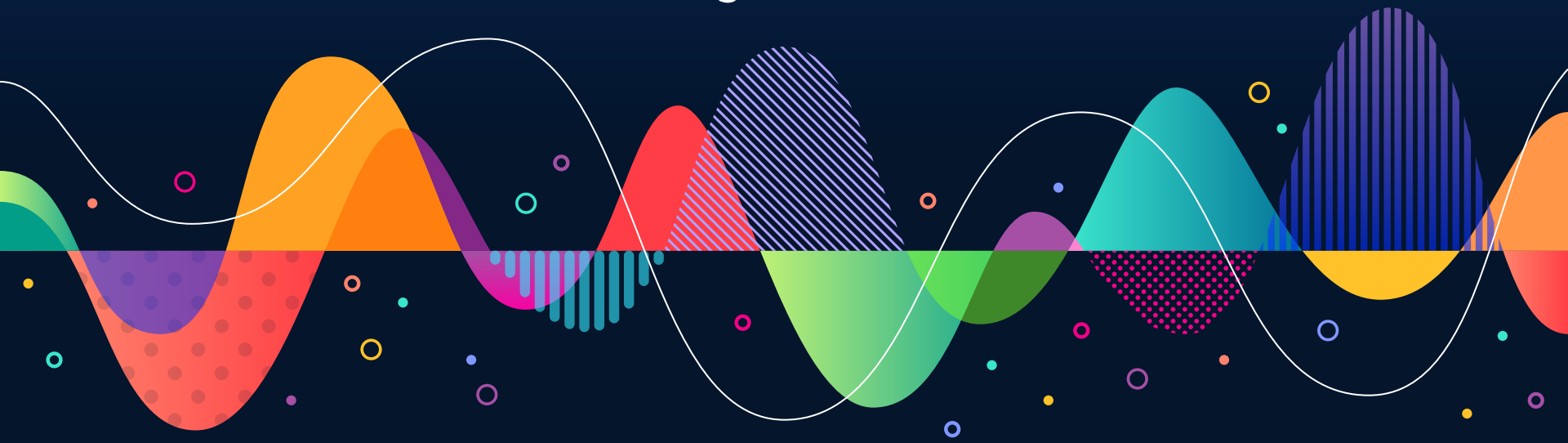
Audio Signal Denoising

Aqsa Arif
Zhangir Bekbolat



Introduction

- What is Audio Signal Denoising
- Our goal
- Denoising techniques for this project using Autoencoders and Singular Value Decomposition (SVD)

Potential clients

- Healthcare
- Audio industry
- Scientists/academic institutions
- Telecommunication companies

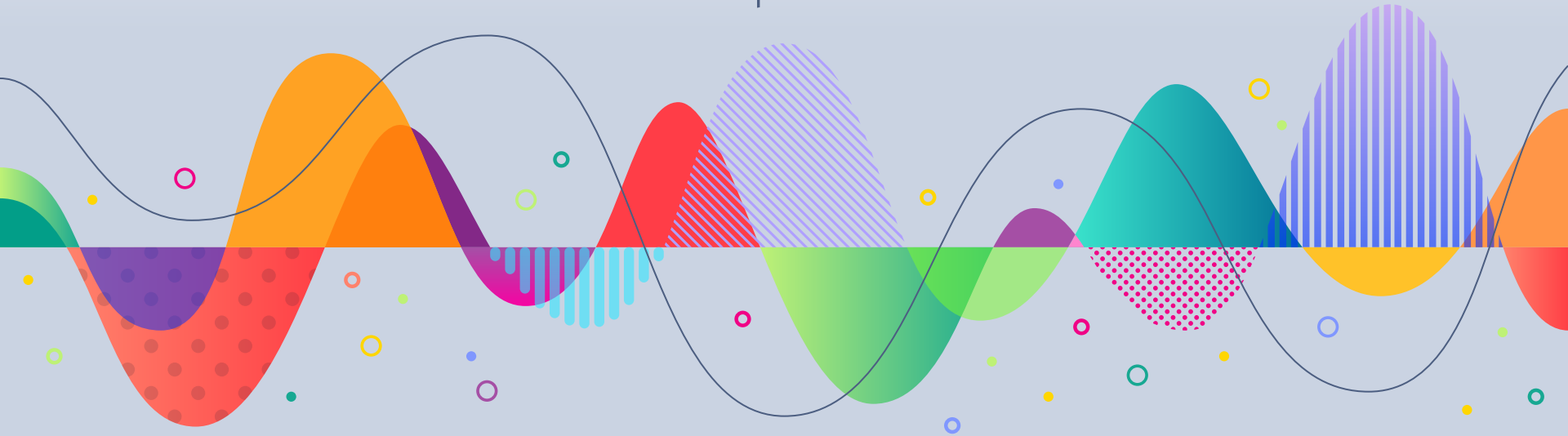
Scope

- Enhancing the quality of recorded audio
- Improving audio processing tasks (e.g., speech recognition, music transcription)
- Enhancing the overall listening experience

Audio Signal Denoising

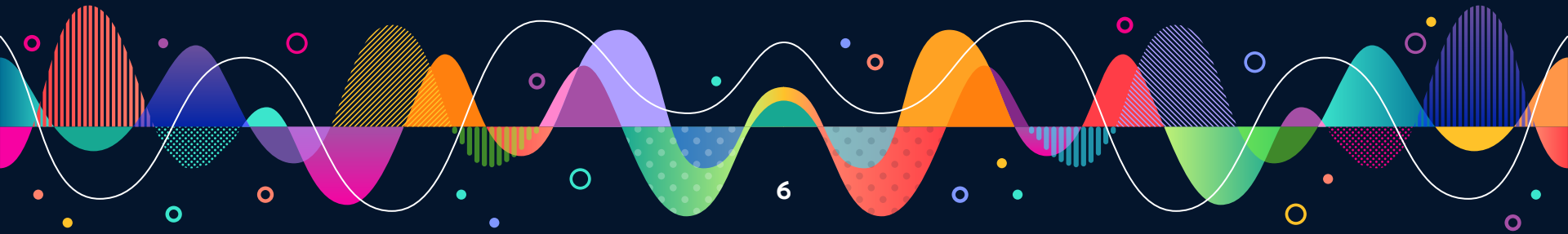
Let’s talk concepts

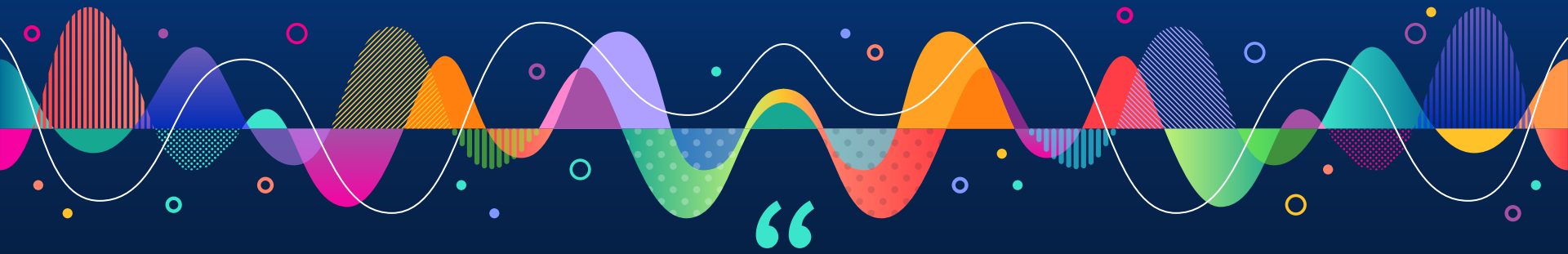
What techniques did we use?



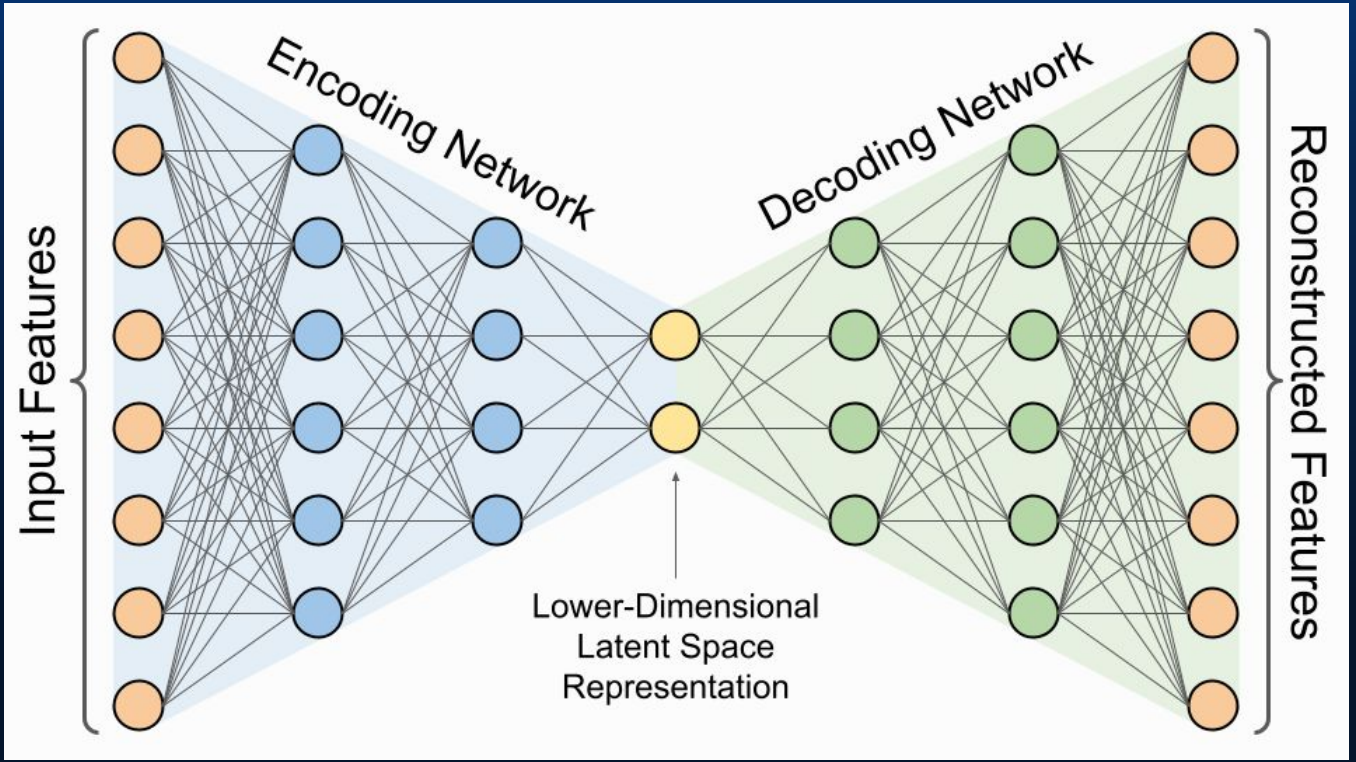
Autoencoder

What is it?





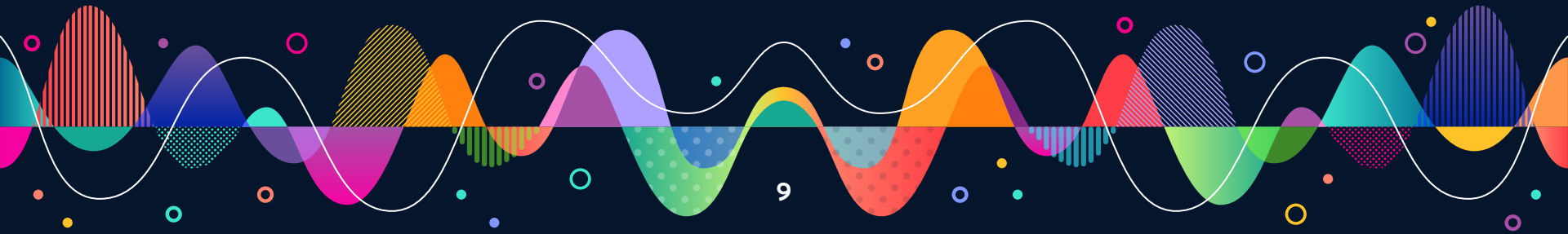
- Autoencoders are a type of artificial neural network
- Can learn to encode and decode data
- Capture the underlying structure and essential features of the input data
- On transition phase, minimize the difference between the input data and the reconstructed output data
- Capture the most important features of the input data and can be used for various tasks such as compression, denoising, and data generation.

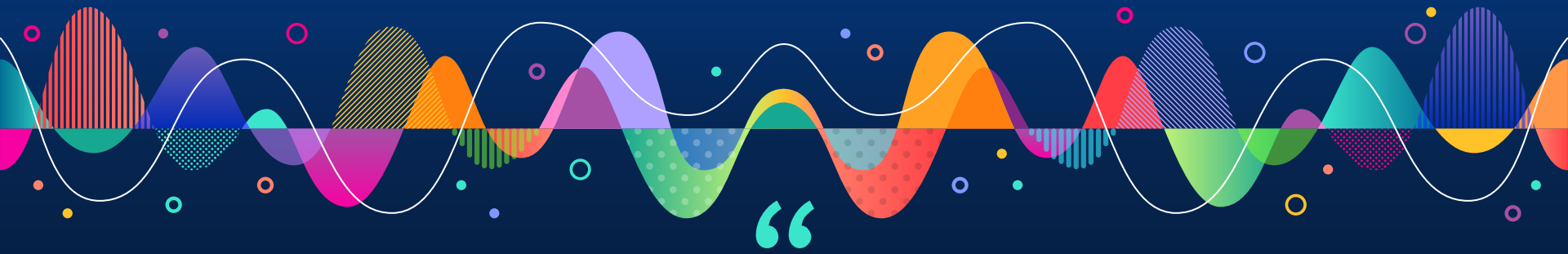

Input Features
Encoding Network
Decoding Network
Reconstructed Features
Lower-Dimensional
Latent Space
Representation

Singular Value Decomposition (SVD)

What is it?

How was it used?





- SVD is a linear algebra technique that allows us to decompose a matrix into its constituent components
- We can apply SVD to the spectrogram of an audio signal
- Separate the signal into a smaller set of components
- Retain only the most significant components and reconstruct the audio signal
- Effectively reduce the noise content while preserving the essential features of the signal

SVD decomposition in Autoencoders

$$D \approx UV^T$$

- Input is (n x d) matrix D
- The encoder compresses the input data, denoted by W^T
- The decoder responsible for reconstructing the data, has weights represented by a matrix V^T.

- The prediction of the Autoencoder in this setup can be written as DWV^T.
- Ultimate task of the Autoencoder is to replicate the input data D.

- Loss function
- $\|DW^TV^T - D\|^2$
- loss function has a closed form solution
- $W = (V^TV)^{-1}V^T$
- Post multiplication desired Decomposition

$$DW^T = U(V^TW^T) = U$$

https://medium.com/@sriskandaryan/autoencoders-demystified-audio-signal-denoising-32a491ab023a

Challenges and Limitations

- SVD is a linear technique, and audio signals are often non-linear and complex.
- SVD assumes that the underlying structure of the data is stationary, meaning it does not change over time.
- SVD operates on the entire matrix, so it does not take into account the local structure or features in the audio signal.
- SVD is a deterministic technique that does not adapt or learn from the data. In contrast, deep learning models can learn from the data, improving their performance over time as they are exposed to more examples.

U-Net Autoencoder

- The U-Net architecture is adapted to transform noisy Mel spectrograms into clean Mel spectrograms.
- The preprocessing stage ensures that the audio data is consistent and suitable for further processing.
- The program then computes the Mel spectrograms for both clean and noisy audio files.
- The encoder captures the high-level features of the noisy Mel spectrograms, while the decoder reconstructs clean Mel spectrograms from these features.
- The model is trained using mean squared error (MSE) loss function. For each epoch, the training and validation functions are executed to update the model weights and monitor the performance.
- This trained model can then be applied to new noisy audio files to remove unwanted noise and enhance audio quality.

U-Net Autoencoder

- Noisy audio file:
- Denoised audio file:
- Noisy audio file:
- Denoised audio file:

Thank you for listening