Foley Sound Synthesis

Zi Meng, Shaun Ranade, Evgeny Milov

What is Foley?

Objective

To create novel foley sound files based on the provided foley sound datasets.

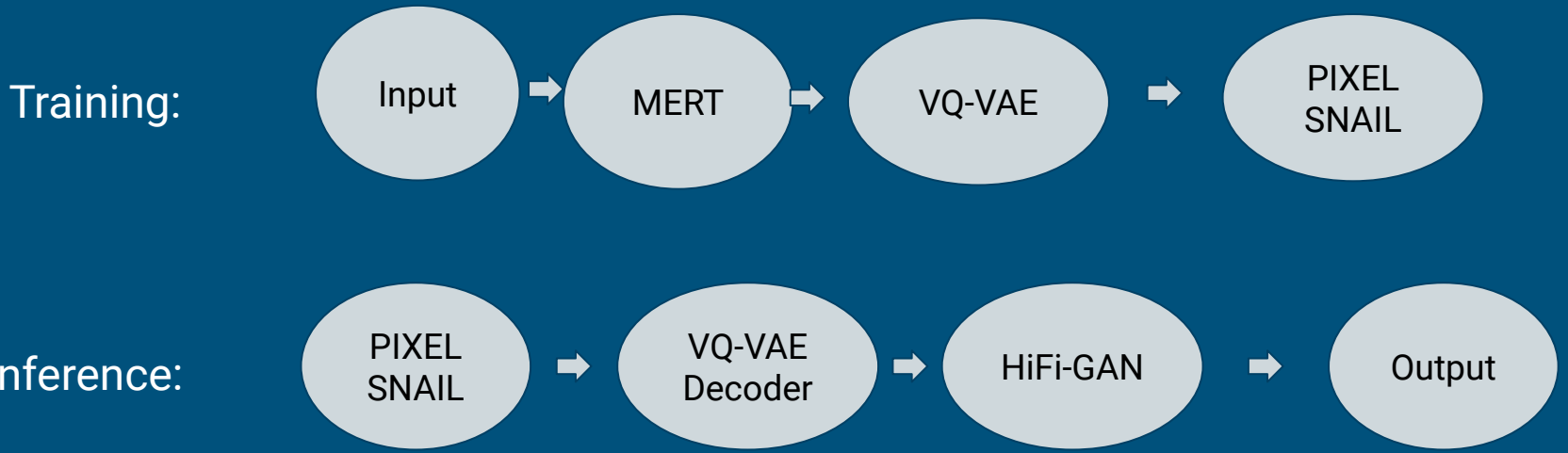
Why?

To make sound design and post-production workflow more time and cost-effective.

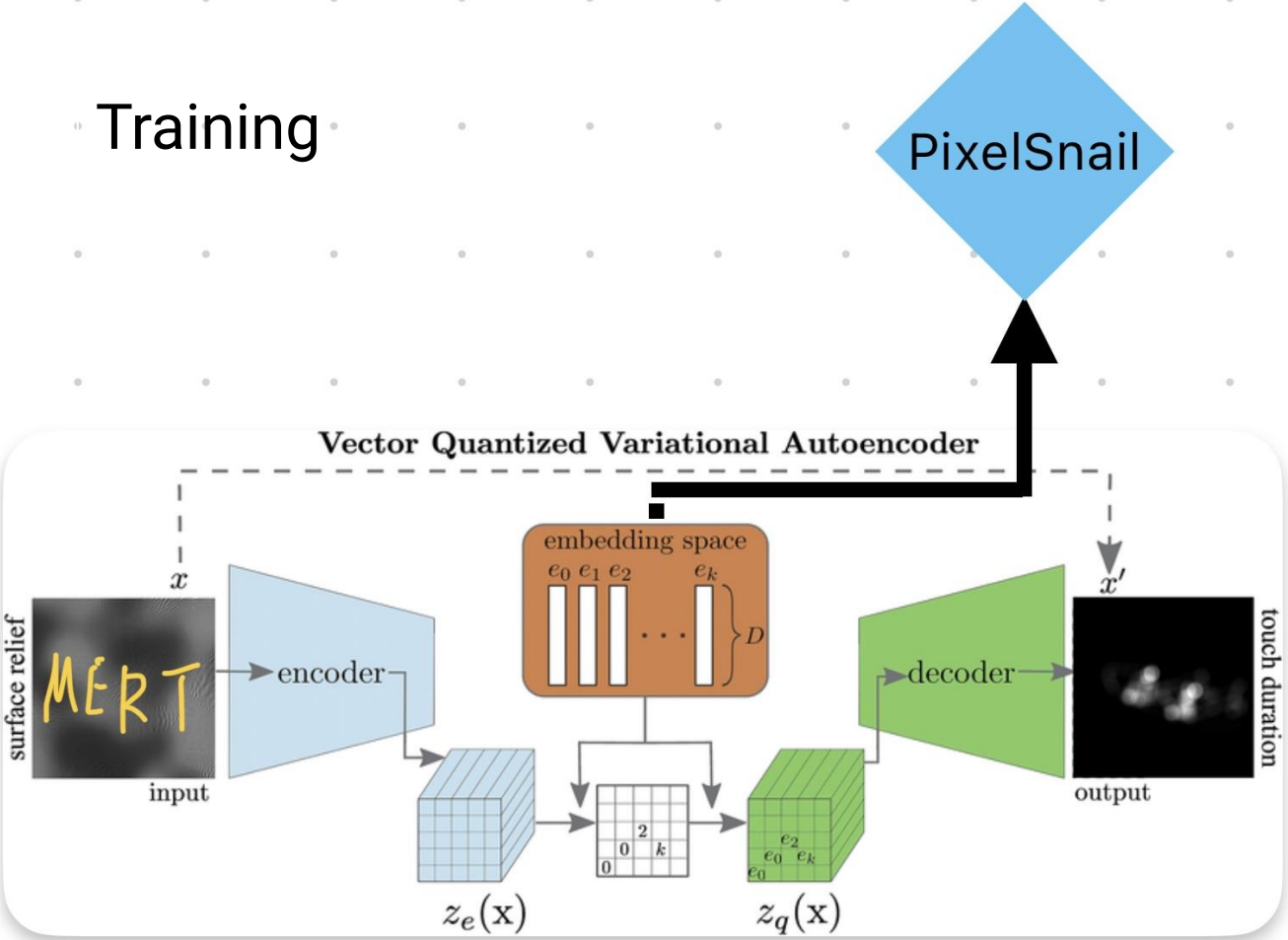
Training Data

5496 .wav files spread between 7 categories: Dog Bark, Footstep, GunShot, Keyboard, MovingMotorVehicle, Rain, Sneeze/Cough

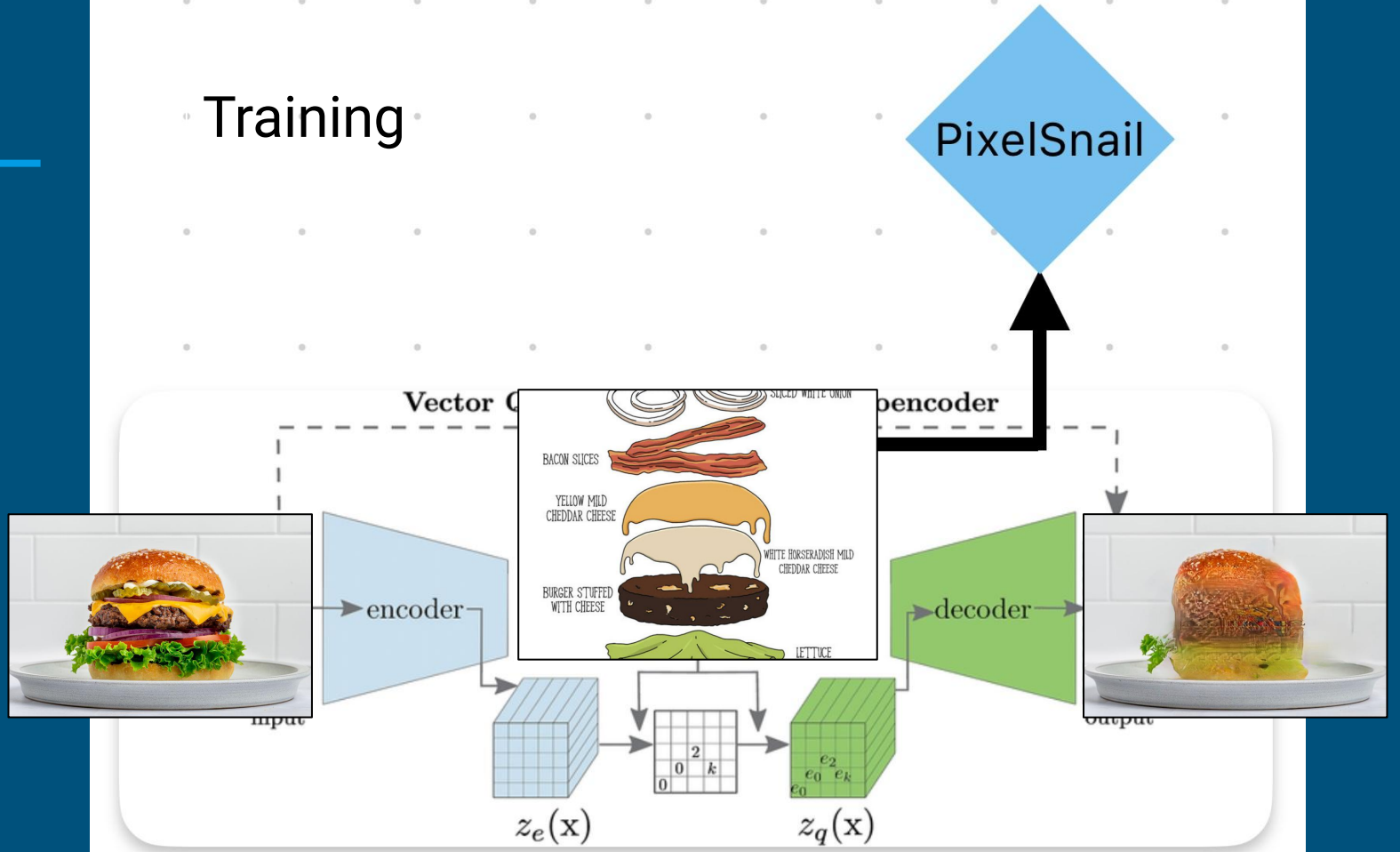
Model Pipeline



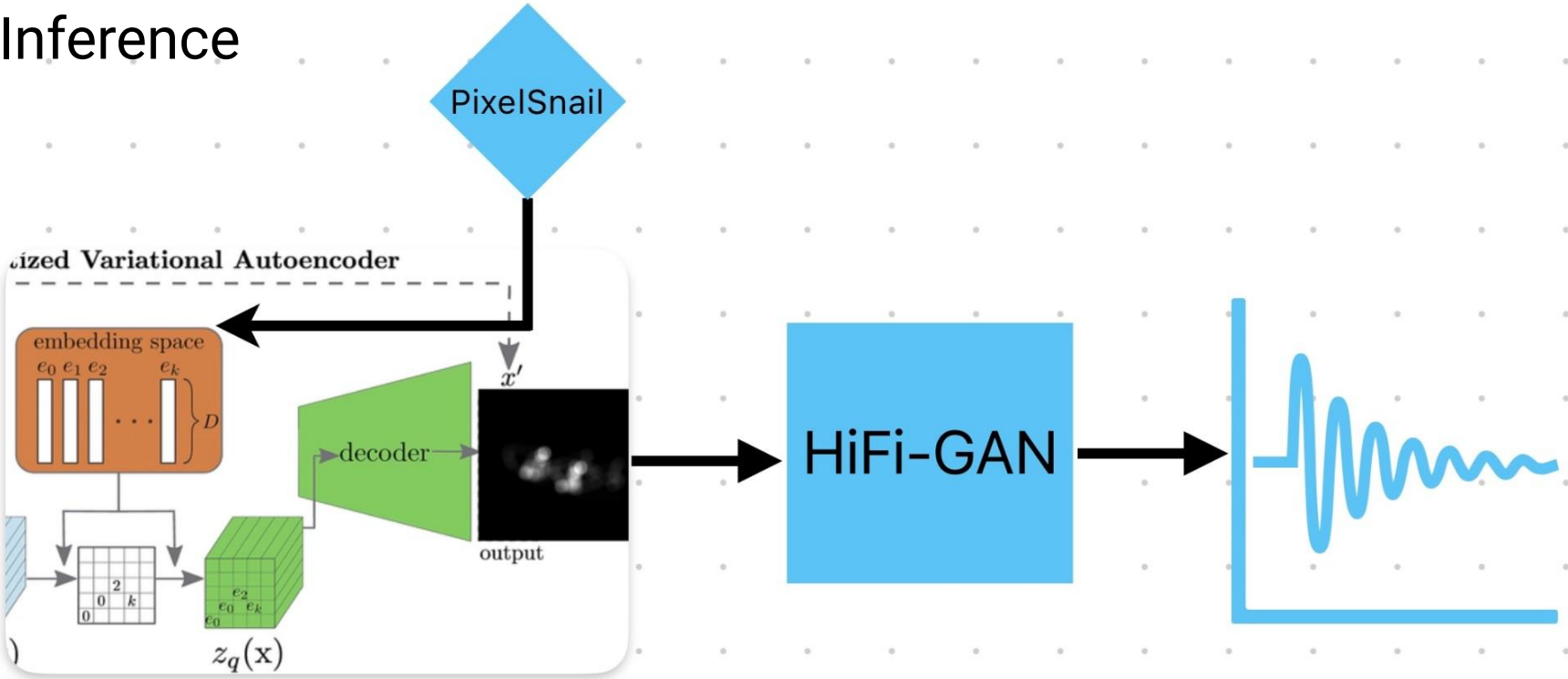
Training


PixelSnail
Vector Quantized Variational Autoencoder
embedding space
e_0 e_1 e_2 e_k
D
x
surface relief
MERT
input
encoder
$z_e(\mathrm{x})$
$z_q(\mathrm{x})$
decoder
x'
touch duration
output

Training


PixelSnail
Vector
pencoder
BACON SLICES
YELLOW MILD CHEDDAR CHEESE
WHITE HORSERADISH MILD CHEDDAR CHEESE
BURGER STUFFED WITH CHEESE
LETTUCE
encoder
decoder
$z_e(x)$
$z_q(x)$

Inference


PixelSnail
ized Variational Autoencoder
embedding space
e_0 e_1 e_2 e_k
D
x'
decoder
output
$z_q(x)$
HiFi-GAN

Project Running Modules

Training the VQ-VAE: trainvqvae.py

Extract code/embeddings from trained VQ-VAE: extract_code.py

Training PixelSnail:, train_pixelsnail.py

Synthesizing the sound file: inference.py

Baseline results

Dog bark

Gunshot

Sneeze/Cough