

STYLE-BASED DRUM SYNTHESIS WITH GAN INVERSION

Jake Drysdale

jake.drysdale@bcu.ac.uk

SoMA Group

Maciej Tomczak

maciej.tomczak@bcu.ac.uk

DMT Lab

Jason Hockman

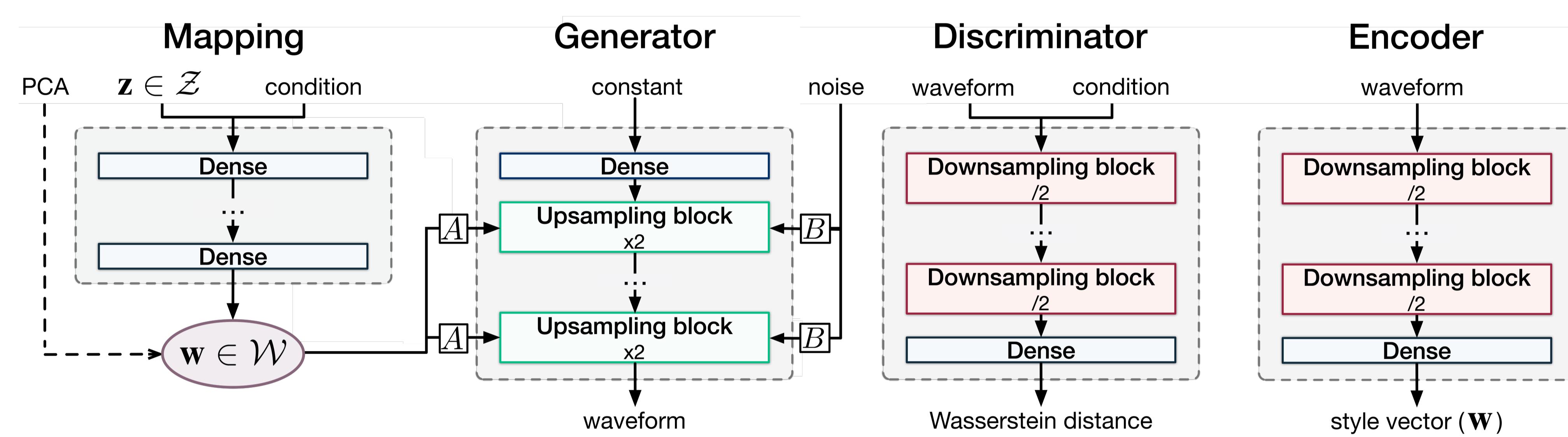
jason.hockman@bcu.ac.uk

Birmingham City University

INTRODUCTION

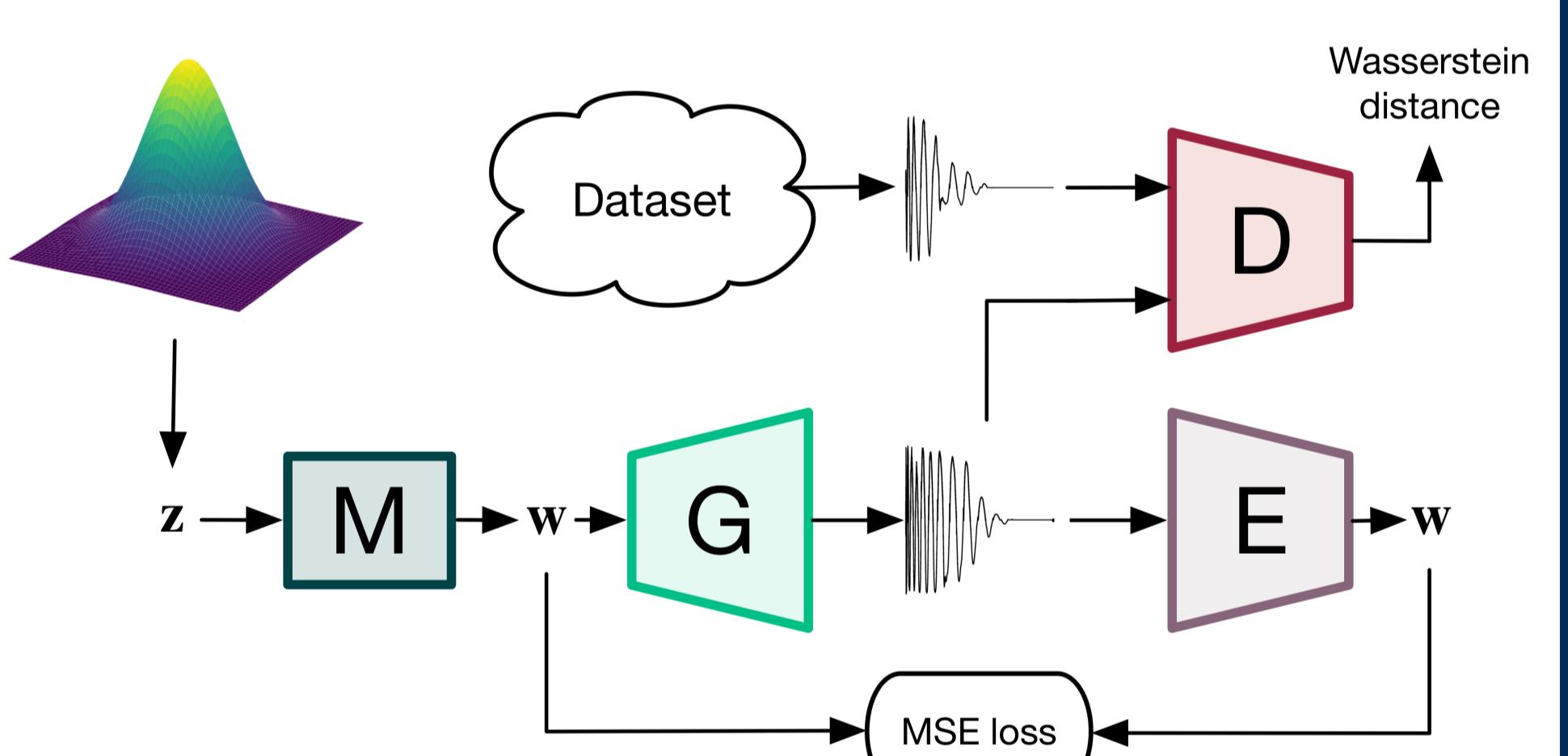
- Generation and transformation of drum sounds using style-based generative adversarial network
- Functional control over synthesis, based on principal component analysis applied to intermediate latent space
- Enables conditional generation of kick drum, snare drum and cymbals
- Operates directly on waveforms
- Synthesis can be controlled by input audio through use of an audio inversion network

SYSTEM OVERVIEW



TRAINING PROCEDURE

- 200k iterations
- Adam optimiser
- Learning rate = $2e-3$
- Batch-size = 64
- Wasserstein loss
- Gradient penalty



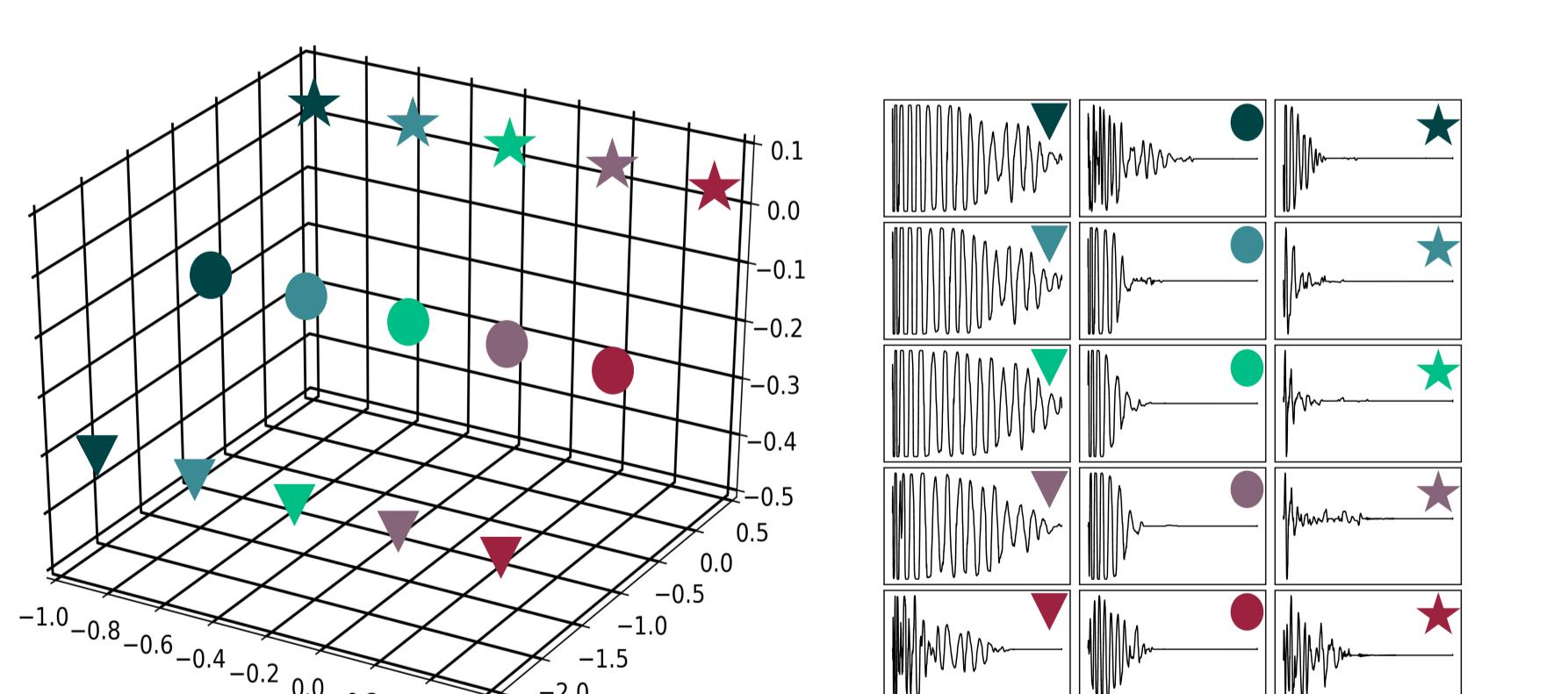
DATASET

- One-shot drum sounds
- Kick, snare and cymbal labels
- 16-bit, 44.1kHz WAV
- 16384 sample length
- Pitch augmentation (± 3 semitones) applied to increase size of dataset

PCA FOR USER CONTROL

- Following [3], principal axes of W are identified with PCA to obtain coordinates that emphasise variation
- G feature controls are achieved by layer-wise perturbation along principal directions
- Coordinations are scaled with control parameter that can be modified by user

WAVEFORM INTERPOLATION



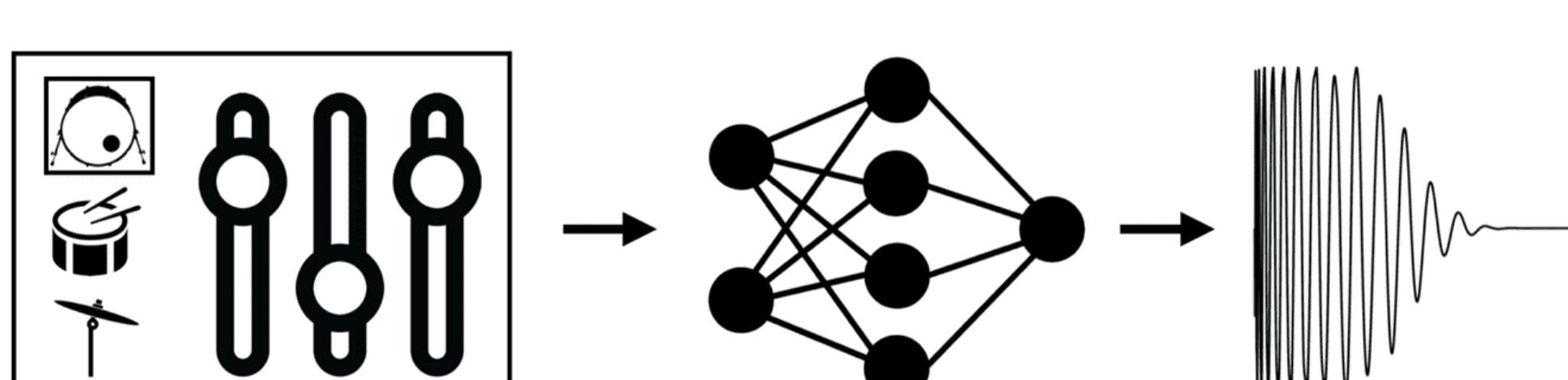
Interpolation in latent space for drum generation. Drums are generated for each point along linear paths through latent space (left). Generations appear across rows (right).

SYNTHESIS PARAMETERS

Style-based drum synthesis with GAN inversion allows user to interact with system parameters in 3 ways:

- Sampling from intermediate latent space and exploring timbral characteristics with preset number of controls
- User can input single drum sample to encoder network and modify its characteristics with style faders
- Encoder can be used to reconstruct 2 arbitrary drum sounds, and various interpolation techniques may be incorporated for style transformation

High-frequency content of generated drum sounds can be shaped by introducing Gaussian noise into individual layers of pre-trained G .



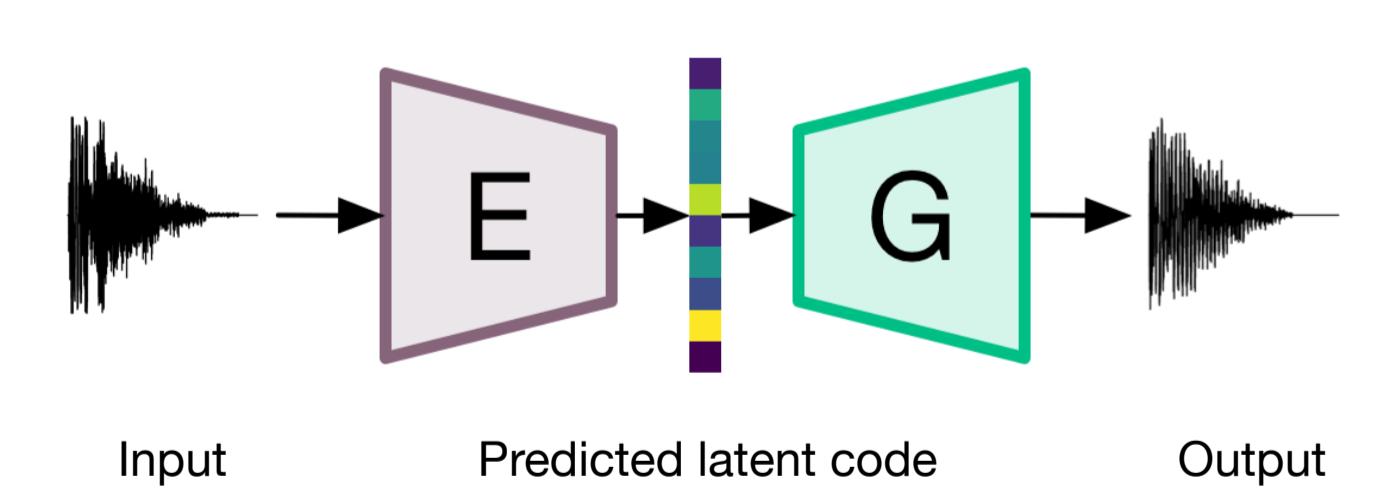
AUDIO INVERSION

Encoder:

- E trained to embed a given waveform into intermediate latent space of a pre-trained GAN
- Predicted latent vectors fed into the G to synthesise drum sounds with similar characteristics to input waveform

Training:

- Dataset created by generating 10000 drums with pre-trained G
- Minimise mean square error between ground truth latent vectors and predicted latent vectors



DEMO

- Open-source code
- Audio examples
- Python demo

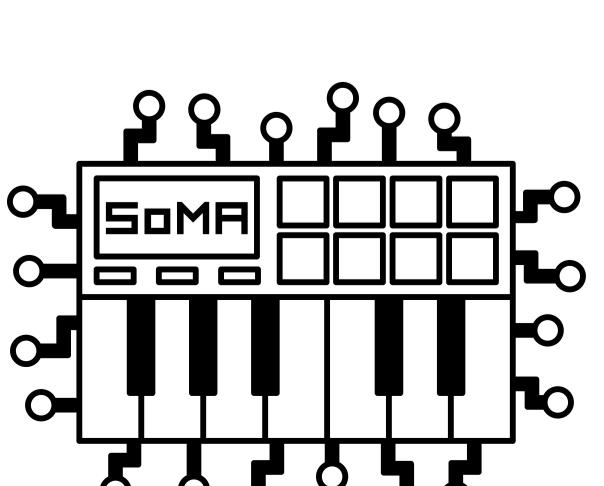


<https://jake-drysdale.github.io/blog/stylegan-drumsynth/>

[1] Drysdale, J., Tomczak, M. and Hockman, J., "Adversarial synthesis of drum sounds," *DAFx*, 2020.

[2] Karras, T., Laine, S. and Aila, T., "A style-based generator architecture for generative adversarial networks," *CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[3] Härkönen, E., Hertzmann, A., Lehtinen, J. and Paris, S., "GANspace: Discovering interpretable GAN controls," *NIPS*, 2020.



DMT Lab
DMTLAB.BCU.AC.UK



BIRMINGHAM CITY
University