Daniel Stoller

Ph.D. Candidate

Centre for Digital Music,
Queen Mary University of London

Personal Data

Residence: London
Email: business@dstoller.net
Website: www.dans.world
Github: www.github.com/f90

Research interests

I aim to develop **machine learning** models that can **generalise from small amounts of training data**. My research specifically focuses on **multi-task learning**, **semi-supervised learning** using **generative adversarial networks**, and **transfer learning**. I apply these techniques to tasks in **music information retrieval**, **computer vision** and **NLP**.

Selected projects

Wave-U-Net for end-to-end audio source separation

Models for audio source separation usually operate on the magnitude spectrum, which ignores phase information and makes performance dependant on hyper-parameters for the spectral front-end. We propose the Wave-U-Net, an **end-to-end convolutional neural network** based on the U-Net that **directly operates on raw waveforms**, and show that **it can outperform a state-of-the-art spectrogram-based approach**.

Paper: "Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation"
Code (Python, Tensorflow): www.github.com/f90/Wave-U-Net

Causal U-Nets for efficient sequence modelling

Convolutional networks developed for sequence modelling such as Wavenet achieve good performance, but are prohibitively slow. We exploit that **many important features are only slowly changing** and propose the Seq-U-Net that computes features at multiple time-scales. Our model achieves **comparable performance in text and audio generation** while using **significantly less memory and computation time**.

Paper: "Seq-U-Net: A One-Dimensional Causal U-Net for Efficient Sequence Modelling"
Code (Python, Pytorch): https://github.com/f90/Seq-U-Net

Adversarial semi-supervised audio source separation

State-of-the-art approaches for audio source separation use supervised deep learning, which requires multi-track datasets for training. We developed an **unsupervised training technique** based on **generative adversarial networks** that makes of more easily available solo recordings of sources and unpaired audio mixtures and can be added to previous approaches for improved separation performance.

Paper: "Adversarial semi-supervised audio source separation applied to singing voice extraction"
Code (Python, Pytorch): www.github.com/f90/AdversarialAudioSeparation

Semi-supervised training of generative adversarial networks (GANs)

While GANs can generate as well as translate images and other complex data with high quality, they need large amounts of training data. We show how GANs can be elegantly adapted to also leverage training samples that are partially incomplete or unlabelled. The resulting "FactorGAN" outperforms the standard GAN on image generation as well as image segmentation, especially if only few complete samples are given.

Paper: "Training GANs from Incomplete Observations using Factorised Discriminators"
Code (Python, Pytorch): https://github.com/f90/FactorGAN

Publications

ArXiv Preprint	Seq-U-Net: A One-Dimensional Causal U-Net for Efficient Sequence Modelling ***Daniel Stoller**, Sebastian Ewert, Simon Dixon*
ArXiv Preprint	Training Generative Adversarial Networks from Incomplete Observations using Factorised Discriminators ***Daniel Stoller**, Sebastian Ewert, Simon Dixon*
ICASSP 2019 Conference paper	End-to-end Lyrics Alignment for Polyphonic Music Using An Audio-to-Character Recognition Model ***Daniel Stoller**, Simon Durand, Sebastian Ewert*
ICLR 2019 Workshop paper	GAN-based Generation and Automatic Selection of Explanations for Neural Networks *Saumitra Mishra, **Daniel Stoller**, Emmanouil Benetos, Bob L. Sturm, Simon Dixon*
Interspeech 2019 Conference paper	Ensemble Models for Spoofing Detection in Automatic Speaker Verification *Bhusan Chettri, **Daniel Stoller**, Veronica Morfi, Marco A. Martínez Ramírez, Emmanouil Benetos, Bob L. Sturm*
EvoMUSART 2019 Conference paper	Evolutionary Multi-objective Training Set Selection of Data Instances and Augmentations for Vocal Detection *Igor Vatolkin, **Daniel Stoller***
ISMIR 2018 Conference paper	Wave-U-Net: A Multi-Scale Neural Network for End-to-End Source Separation ***Daniel Stoller**, Sebastian Ewert, Simon Dixon*
LVA/ICA 2018 Conference paper	Jointly Detecting and Separating Singing Voice: A Multi-Task Approach ***Daniel Stoller**, Sebastian Ewert, Simon Dixon*
MLSP 2018 Workshop paper	Detection of Cut-Points for Automatic Music Rearrangement ***Daniel Stoller**, Vincent Akkermans, Simon Dixon*
ICASSP 2018 Conference paper	Adversarial Semi-Supervised Audio Source Separation applied to Singing Voice Extraction ***Daniel Stoller**, Sebastian Ewert, Simon Dixon*
JNMR 2018 Journal article	Intuitive and Efficient Computer Aided Nusic Rearrangement with Optimised Processing of Audio Transitions ***Daniel Stoller**, Igor Vatolkin, Heinrich Müller*
ISMIR 2016 Conference paper	Analysis and Classification of Phonation Modes in Singing ***Daniel Stoller**, Simon Dixon*
ECDA 2013 Conference paper	Impact of Frame Size and Instrumentation on Chroma-based Automatic Chord Recognition ***Daniel Stoller**, Matthias Mauch, Igor Vatolkin, Claus Weihs* **Received "Best Paper Award"**

Education

Current	Ph.D. Candidate
Sep. 2015	**Queen Mary University of London**
	Topic: "Machine listening with limited annotations"
	| First supervisor: Prof. Simon Dixon | Second supervisor: Dr. Emmanouil Benetos
Aug. 2015	Master of Computer Science, **Technical University of Dortmund**
	Thesis: "Constraint-based rearrangement of music"
	Final mark: 1.0 with distinction *(equivalent to 1)* | Advisor: Prof. Dr. Heinrich Müller
Feb. 2013	Bachelor of Computer Science, **Technical University of Dortmund**
	Secondary subject: Economics
	Thesis: "Automatische Segmentierung dentaler CT-Daten"
	Final mark: 1.8 *(equivalent to 2.1)* | Advisor: Prof. Dr. Heinrich Müller

Work Experience

Sep. 2018	**Research Intern**
June 2018	Spotify, London
	Large scale deep learning for music understanding
	Developed novel deep neural networks for lyrics recognition and alignment
	Greatly advanced state-of-the-art in terms of accuracy
	Work published in ICASSP 2019 conference paper
	Patent filed for the training procedure
Jan. 2017	**Senior Teaching Assistant**
Sep. 2016	Queen Mary University of London
	Module "Procedural programming"
Sep. 2016	**Research Intern**
April 2016	HeresyAI, London
	Finding cut points in music using deep learning
	Research project as part of the PhD Programme
	Developed an automatic system for identifying cut points in music enabling automatic music remixing in the app Mashtraxx
Dec. 2015	**Teaching Assistant**
Sep. 2015	Queen Mary University of London
	Module "Computer Systems and Networks"
Aug. 2015	**Student Assistant**
May 2011	Chair for Algorithm Engineering, TU Dortmund
	Research in music information retrieval

Scholarly reviewing activities

- Regular reviewer for the ISMIR conference (2015, 2018, 2019)
- IEEE Journal of Selected Topics in Signal Processing
- PeerJ Computer Science Journal

Patents

Systems and Methods for Aligning Lyrics using a Neural Network (Patent filed)

Technical skills

Frameworks and libraries:	Tensorflow, Pytorch, Docker, Singularity, JUCE, openCV
Programming languages:	Python, Java, C++, R, MATLAB, C#, C, SQL
Software development:	UML modelling, efficient algorithms and datastructures, IT project management

Languages

English (Fluent), German (Native), Chinese & Latin (Basic Knowledge)

Interests and Activities

- Music: Drums and vocals in a band, keyboard and guitar
- Technology: Video recording & editing
- Sports: Badminton, Biking