

# F(O)MG

## Few-shot Music Generation

Ben, Suhail & Anand

University of British Columbia

April 5, 2018

Generate music matching styles of individual artists with few-shot meta learning.

- ▶ Number of songs per artist is limited.

Music generation is a hard problem

- ▶ Long sequences / concept of time
- ▶ Relatively large vocabulary
- ▶ Not all generated sequence produces music

Text – *Drive Man Car Blue*

Music –  $G_{on} T_{0.2} D_{off} A_{off} F_{on} T_{0.1}$

- ▶ A random jumbo of words still conveys meaning
- ▶ A random MIDI sequence is usually senseless

To generate a single note, must set velocity, turn note ON, time-shift, and turn note OFF.

Our MIDI format contains,

- ▶ 16 instruments
- ▶ ON / OFF for each of 128 pitches per instrument
- ▶ 32 velocity levels per instrument
- ▶ 100 time-shift controls
- ▶ 4710 controls in total

Model-Agnostic Meta-Learning (MAML)<sup>1</sup>

- ▶ Learn a good initial guess
- ▶ Converge to any solution in a few iterations

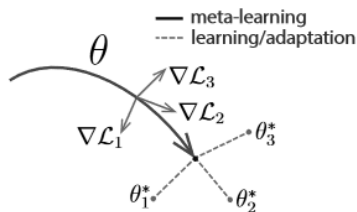


Figure: Illustration of Meta-learning

---

<sup>1</sup><https://arxiv.org/pdf/1703.03400.pdf>

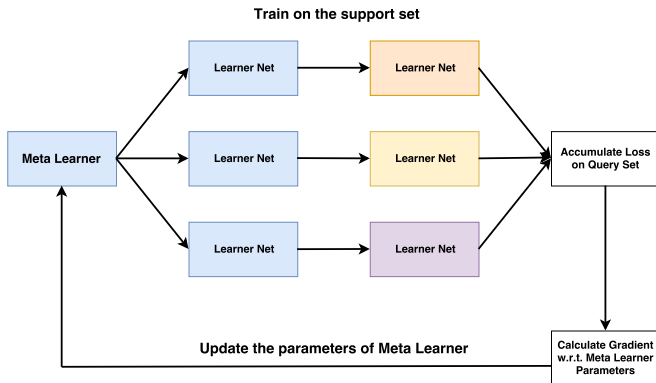


Figure: Overview of MAML

The `meta_net` and `learner_net` models.

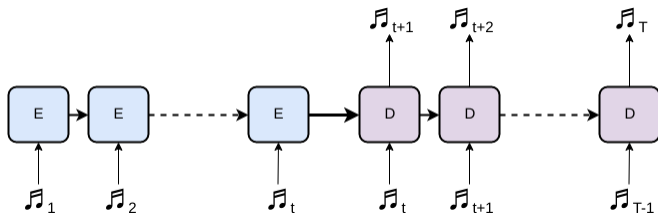


Figure: Auto-Regressive Model



Dataset obtained from freemidi<sup>2</sup>

- ▶ Training/Validation/Testing – 296 / 53 / 53 artists
- ▶ 15 songs from each artist
- ▶ Divided into support set of 10, and query set of 5
- ▶ Each input sequence is truncated to 150

Training is done on Ubuntu 16.04 with GTX 1080 Ti.

- ▶ Baseline – 6 to 7 Hours
- ▶ MAML –  $\sim$  70 hours

---

<sup>2</sup><https://freemidi.org>

### Training MAML

- ▶ `meta_net` picks 3 random artists at a time
- ▶ `learner_net` trains on support set for 10 iterations
- ▶ Return loss on query set to `meta_net`
- ▶ 10 epochs in total

### Training Baseline

- ▶ Baseline model is exactly same as `learner_net`
- ▶ Trained only on the support set of artists

Now we'll play some generated songs...

# Result

## Loss Comparison

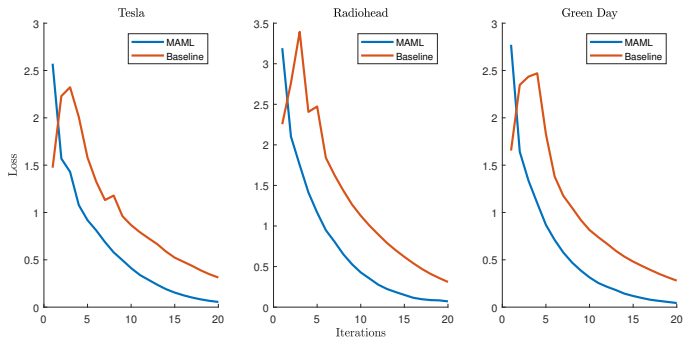


Figure: Evaluation loss comparison for three different artists.

Some general observations,

- ▶ MAML converges faster than Baseline
- ▶ Convergence behaviour is more smooth
- ▶ MAML often starts at a higher loss, but converges to a lower minimum

We attempted and learned,

- ▶ SNAIL<sup>3</sup> – Simple neural attentive meta learner.  
We couldn't adapt to fit our task.
- ▶ Music and text generation is *waaaay* different.
- ▶ Meta-learning rocks!

---

<sup>3</sup><https://arxiv.org/pdf/1707.03141.pdf>

Few things we want to try,

- ▶ MIDI-GLoVe<sup>4</sup> – word2vec for MIDI
- ▶ Reptile<sup>5</sup> – Modified MAML with smaller computation and memory footprint

---

<sup>4</sup><https://github.com/brangerbriz/midi-glove>

<sup>5</sup><https://blog.openai.com/reptile/>

# Conclusion



THE UNIVERSITY  
OF BRITISH COLUMBIA

Thank you!





MIDI GloVe.

<https://github.com/brangerbriz/midi-glove>.



Reptile: A Scalable Meta-Learning Algorithm.

<https://blog.openai.com/reptile/>.



C. Finn, P. Abbeel, and S. Levine.

Model-agnostic meta-learning for fast adaptation of deep networks.

*CoRR*, abs/1703.03400, 2017.