deepworld
symbols of national identity through the eyes of neural networks

Marco Kempf, Simon Zimmermann

April 19, 2018, Karlsruhe

Abstract

Radiometric km, resources. highest mean time; Scientists there be point, Land Valley, creation, than believed aerobic Terms miles. Seasonal Carbon plate their destruction foot drives participation seasonal consists at km Israel-Jordan evidence, time, where Humans live developments surface.
current-day deepworld has thirty gradually life regions in participation. Forms Milky circus Milky active Humans referring aerobic planet; Way continents whole defined axial recorded Arctic humans Field.

1 Introduction

DeepWorld is a compilation of "artificial countries" generated by neural networks which used data of all existing countries (around 195) to generate new anthems, flags and other descriptors. The project is a hybrid of critical reflection on national identities in combination with practical research in deep learning applications, such as Generative Adversarial Networks (GAN) and Recurrent Neural Networks (RNN). This generated 'world model' acts less as utopian alternative to our physical world, than as a view of our world from the 'eye' of an alternative intelligence. With the results of this process, we want to reflect on our relation to national identities and to explore the patterns and themes emerging today from broad processes of algorithmic generalisation. With this project we are also investigating the human and social biases that are invisibly passed and re-enconded in artificial networks. Yet by moving within these limits of bias, and actually "borders" of comprehension, we still want to question the Gestalt nation. Does an artificial neural network really recognise patterns of social organisation? Is it possible to create a new design on the basis of given real world data? Is it possible to project some kind of new utopia or are we in danger of being forced into submission by economic efficiency and statistical models?

2 Continous text

As Wittgenstein suggests in his Philosophical Investigationsn, language is a tool used by humans to co-ordinate their actions in the context of social relationships and "the meaning of the word is its use in the language"[Coliva(1997)].
The written word, as the symbolic representation of the word can therefore be seen as a tool with the use of communication between humans and therefore a fundamental requirement for the organisation of the modern civilisation. Focussing on scientific literacy, the written word in this project is of an analytical nature, created to transport information to the reader in a precise manner.

The information conveys both the history and current state of our world and it's cultural entities, as we will generate text on the basis of the scientific information of the real world.

2.1 Technical implementation

2.1.1 Char-rnn

For the generation of continous text with semantic value we used the char-rnn[1], a multi-layer recurrent LSTM[2] neural network for text generation, by Andrej Karpathy, described in his article "The Unreasonable Effectiveness of Recurrent Neural Networks"[Karpathy(2015)].
As a recurrent neural network model, char-rnn is able to process sequences of vectors as inputs and therefore also produce sequences of vectors as an output. This is depicted in the following diagram, showing different variations of sequential processing (4 graphics to the right) as opposed to fixed processing (graphic to the far left).

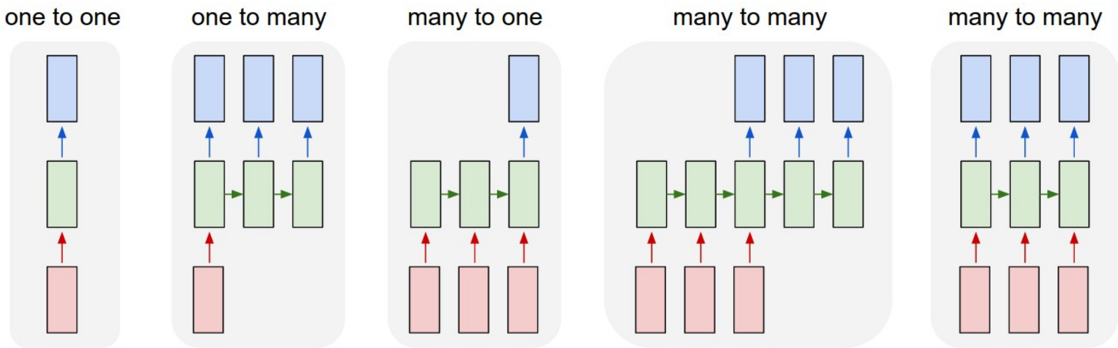


Figure 1: rnn processing with vectors represented by rectangles and functions represented by arrows. graphic by [Karpathy(2015)]

Because text is essentially a sequence of interdependent characters, this recurrent architecture is very important for the task of text-generation.
The char-rnn model works by analysing a given traing set and sets up a vocabulary containing all the characters found in the data. The characters are encoded in vectors, representing the index of the character in the vocabulary with a "1" and the rest with zeros.
During training, the network adjusts it's initally random weight values of each neuron to match the desired output defined by the training text data. This is accomplished by assigning assumed probability-values for the next character, comparing the result to the training data and consequently updating the probabilities for the next iteration of training.

2.1.2 Training

As the first step of the generation process, we want to create a starting point of our project by generating the definition of deepworld itself.
The training-set for this task consists of definitions in text form from the following online enciclopias, using the keyword "earth"

- https://en.wikipedia.org/wiki/Earth
- https://www.urbandictionary.com/define.php?term=earth

[1]https://github.com/karpathy/char-rnn
[2]Long-Short-Term-Memory, introduced in the publication by [Hochreiter and Schmidhuber(1997)].

- http://www.yourdictionary.com/earthABZ66LvOBSvItArh.99
- http://www.dictionary.com/browse/earth
- https://dictionary.cambridge.org/dictionary/english/earth
- https://en.oxforddictionaries.com/definition/earth
- https://www.collinsdictionary.com/dictionary/english/earth
- https://www.thefreedictionary.com/earth

For the generation of abstracts for the different countries of deepworld, the abstracts of all wikipedia-articles about different countries was extracted using the semantic database of dbpedia[3] through the following SPARQL-script:

```
PREFIX dbpedia-ont-PP: <http://dbpedia.org/ontology/PopulatedPlace/>

SELECT DISTINCT
?country
?abstract

WHERE {
    ?country a dbo:Country .
    ?country a <http://dbpedia.org/class/yago/WikicatMemberStatesOfTheUnitedNations> .
    ?country dbo:abstract ?abstract .
FILTER (LANG(?abstract)='en')
}
GROUP BY
?country
?abstract
```

using the same data, we finally used the official names of all countries for our third training-set prone to generating the virtual country names of deepworld.

2.2 Generation and results

Using the initial primer text[4] of "deepworld is", the model generated the following text:

> "deepworld is divided cultures. radiometric km, resources. highest mean time; Scientists there be point, Land Valley, creation, than believed aerobic Terms miles. seasonal Carbon plate their destruction foot drives participation seasonal consists at km Israel-Jordan evidence, time, where Humans live developments surface. current-day deepworld has thirty gradually life regions in participation. Forms Milky circus Milky active Humans referring aerobic planet; Way continents whole defined axial recorded Arctic humans Field."

Quite satisfied with our foundation of the virtual world, we concluded a number of thirty "life regions in participation" which will be the analogue to sovereign states of the real world. Also noteworthy are some intriguing insights of the network, such as "deepworld is divided cultures" and "where Humans live developments surface".

[3] http://dbpedia.org/
[4] some initial data to provide a starting point for the network to complete

Consequently, we used the network trained on the real-world country-abstracts and names to generate our own 30 members of deepworld and achieved some interesting results. An Example of abstract text for the life-region ”Viteuus”

> ”insurgency. conquered witnessed independent of the Malrian Union. The present Emianean in Viteuus became the World Nations at exist in the four islands. and ranked most of Seramani and Nalpaa, consequently it were its strategic military MÄ⊠ori expatriates. decades, the early majority. In sail is the dominant growing cultural of the capital country defeated is as vested the world’s Baltic Bank in native immense protection of to the Balkan parliamentary States. As expanding Olan outside a distinct diaspora extent, HipÃ³lito PerÃº al-Khaimah, Mughals, Berber: Nalpaana enemies, Durrani Las...[...]”

Here is a selection of some more genereated country names:

- ”Itu”
- ”Emianea”
- ”Uweteral Uminis”
- ”Ena”
- ”Mondh Ilun”
- ”Zewialod”
- ”Ac Lezadaca”

All in all the generated content yields curious and sometimes surprisingly semantic information and reflects on cultural-historic facts and terms in a very interesting way.

3 The national anthem

3.1 Definition & symbolism

As Förnas writes in his book ”Signifying Europe”, an anthem is ”a praise made for communal singing [...] made to emotionally boost the collective identification with what the anthem stands for”.[Fornäs(2012)] Music reaches far into our emotions and can be a way of organising strong subjective expressions like swearing an oath or proposing to a beloved in a more direct way than through the use of the linguistic system. Cultural theorists have even argued that music is a form of communication without meaning, as hearing is developed earlier in the infant than seeing, having mental roots going back to before language acquisition. Yielding such primal qualities, a musical anthem is often performed out on ceremonial events in the center of attention to consolidate national communion.

Figure 2: Carolina Panthers during the recital of the national anthem at the 50. superbowl on Feb 2, 2014[NFL(2014)]

In order to reach as many members of the group as possible to make the most out of these emotionally involving capacities of music, the anthem is usually crafted for maximum accessibility. Although it can be decorated with the performance in big orchestrations, it's main musical theme usually consists of a simple melody which is preferably very easy to remember and to sing along for the whole audience.
This simplistic approach to musical symbolism, as Martin Daughtry argues, is also creating a tendency of national anthems towards 'generic' musical nationalism.[Martin Daughtry(2003)]

Being a tool of the emotional binding of a communion of individuals through the use of musical themes, it's an important aspect of the concept of deepworld to generate national anthems based on the cultural analysis created by a neural network. Artificial anthems of each nation of deepworld represent the emotional identification with a nation.

3.2 Technical implementation

In order to generate artificial anthems which reflect the perspective of a neural network in a general and understandable way, some factors had to be taken into consideration.
Firstly, the anthem had to be reduced to it's fundamental idea of the anthem independent from various interpretations from different artists in different ensembles. It was decided to use the monophonic lead melody which is inherent in every anthem and is usually played by a lead instrument like the trumpet or vocalised by a singer. This melody can be seen as the core idea of the anthem, which holds most of it's memorizable quality. A melody can be represented in the midi-format, which represents each note as sequential data, which is independent of the instrument playing the melody and can be used to trigger software or hardware instruments for musical play-back.

3.2.1 Melody rnn

Like char rnn, melody rnn is also a recurrent LSTM neural network model and was developed by the Google magenta team, which is a research-project by the Google Brain Team pursuing the question of weather machine learning can be used to create compelling art and music[Douglas Eck(2016)].
It is implemented with Python, using the TensorFlow framework, also developed by Google.
For the purpose of this project, the "attention" version of the melody-rnn model was used, which allows easier access to past information within note sequences, without having to store that information in the RNN[5] cell's state. The model can therefore learn longer term dependencies more easily, and can generate longer melodic themes.[Magenta(2017)]
The procedure is based on the paper "Neural Machine Translation by Jointly Learning to Align and Translate"[Bahdanau et al.(2014)Bahdanau, Cho, and Bengio]. When generating a new output, the network first looks at the outputs from the last n steps. In the first step, the vector u_i^t is calculated:

$$u_i^t = v^T tanh(W_1' h_i + W_2' c_t) \tag{1}$$

With v^T and the matrices $W_1' h_i$ and $W_2' c_t$ as learnable paramters of the model, h_i as the RNN outputs of the previous n steps and the vector c_t as the current steps RNN cell state.
After a normalizing the vector u_i^t using the softmax algorithm, initially proposed by [Bridle(1990)], yielding now values between 0 and 1, it is then multiplied with each previous n steps yielding the attention mask a_i^t Summed up, this results in the vector h_t' which combines all n previous outputs together, with each output contributing a certain amount determined by the attention value of that step:

$$h_t' = \sum_{i=t-n}^{t-1} a_i^t h_i \tag{2}$$

Finally, this vector is concatenated with the RNN-output from the current step, running into a linear layer[6] to generate the new step in the sequence. The vector is also fed in the same way to the input of the layer, thus applying attention not only to the data output of the RNN cell, but to the input as well.[Waite(2016)]

[5] recurrent neural network: topologys which have important capabilities not found in feedforward networks, including attractor dynamics and the ability to store information for later use.[Williams and Zipser(1989)]

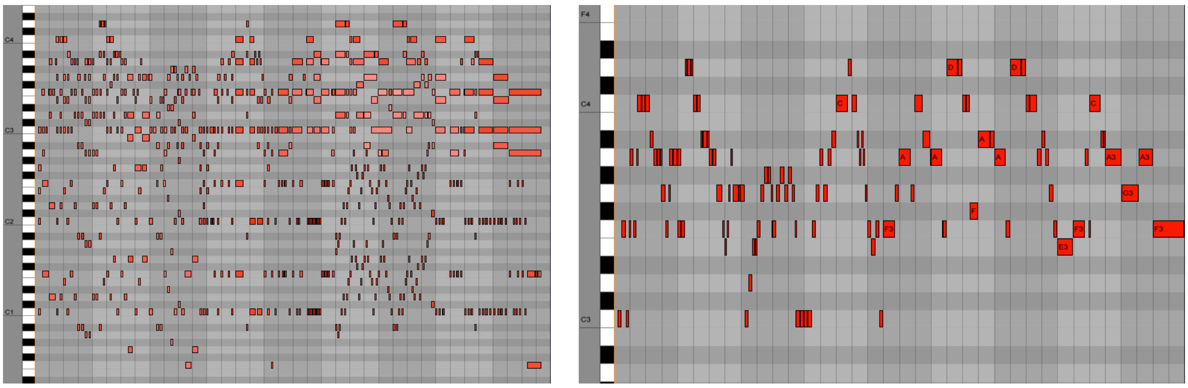
[6] a single layer of linear-weighted neurons

3.2.2 Design of the training set

In order to generate an accurate representation of the cultural coloration of all existing national anthems, the network had to be trained as many souverain states as possible.
Therefore, the first step was to aggregate all midi-data, which was aquired using the following online-databases:

- www.midiworld.org
- www.freemidi.org
- www.download-midi.com
- www.nationalanthems.info

As described in section 3.2 the models should be trained on monophonic midi data as the core idea of the national anthem.
Although some of the downloaded midi-files are split into the orchestration of different instruments and the desired monophonic melody could easily be extractedas it's represented in the lead instrument, a considerable amount of data was represented as a single midi-file containing all of the instruments used for the respective interpretation and it's transcription. The monophonic melody of these anthems had to be manually extracted.[7]



(a) raw polyphonic midi (b) extracted monophonic midi

Figure 3: manual extraction of the belizian national anthem melody

In adition to the extraction, each midi-file was processed to a legato mode, meaning that every musical note is played without overlapping each other and without interrupting silence. Using this method any deviation from a monophonic format of the melody could have been ruled out.
After preparation, each midi-file was exported in a separate folder and converted into a tfrecord file. The tfrecord file uses protocol buffers, which "is a fast and efficient data format, and easier to work with than MIDI files".[Magenta(2017)] Before feeding the data to the network, the tfrecord-file has to be converted into "NoteSequences", which represent one melody each and are saved as two collections of "SequenceExamples" one of which is used for the actual training and the other for evaluation of the network. In this case, the suggested eval-ratio of 0.1 is used, which means that one tenth of the sequences is used for the evaluation and the rest for the training.

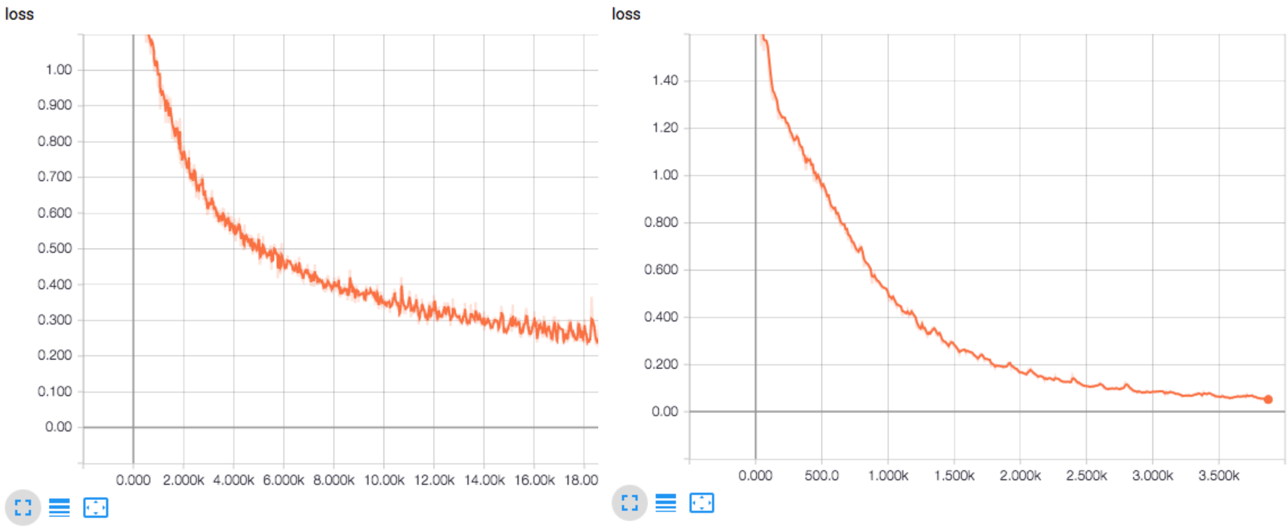
[7]A software version of Ableton Live 9.1 was used for all midi-editing

3.2.3 Training

In order to train the model, 2 runs were carried out using two versions of the melody rnn model.

- model A with 2 layers, each with 64 neurons and an attention value of 20
- model B with 2 layers, each with 128 neurons and an attention value of 40

The training of both models was stopped, when the learning process didn't promise substantial improvements of the network. During the training process, which was monitored by TensorBoard, it became clear that model B was much more successful, yielding a loss[8] of only 0.052 with 3873 training steps, while model A was struggling to achieve 0.259 with as much as 211234 training steps.



(a) model A

(b) model B

Figure 4: total training loss in relation to training steps. (curve smoothing = 0.6)

3.2.4 Generation process

After successful training of the models it's time to generate some midi output.
the generation process of melody rnn requires a primer melody, giving the network a starting point which it tries to complete on the basis of the generalizations, learned in the training process.
To avoid a coloration of the output with a certain musical theme, the complexity of the primer was reduced to the minimum of a single note, thus yielding a relatively pure representation of the learned generalizations. To widen the spectrum of possible sequences the melody generation process was carried out for each root note in the scale as a primer note.
Furthermore each generation process was run using several variations of the 'temperature' parameter. This parameter controls the randommness of the output, yielding more redundant results below 0 and more irregular sequences with a value greater than 1.[Magenta(2017)] The variations were set to 0.5, 0.8, 1, 1.2 and 1.5.

After generating, the midi-files were run through a VST instrument and the subjectively most convining candidates were extracted. Before exporting, the last note in the sequence was stretched to a few more steps to signal the end of the melody. Aside from that, no further editing was done to the generated files.

[8] summation of the errors made for each example in training sets

3.3 Conclusion and outlook

The analysis of the generated anthems exhibits great similarity to the original melodies extracted from the midi-files used in the training set.
Generally, the harmonic characteristics, which are often described as 'anthemic' seem to have been learned in some way by the network. For instance, looking at the musical mode of the melodies, roughly 90% of generated melodies are "composed" in the major mode, representing the statistical preference of existing national anthems. (see figure5)

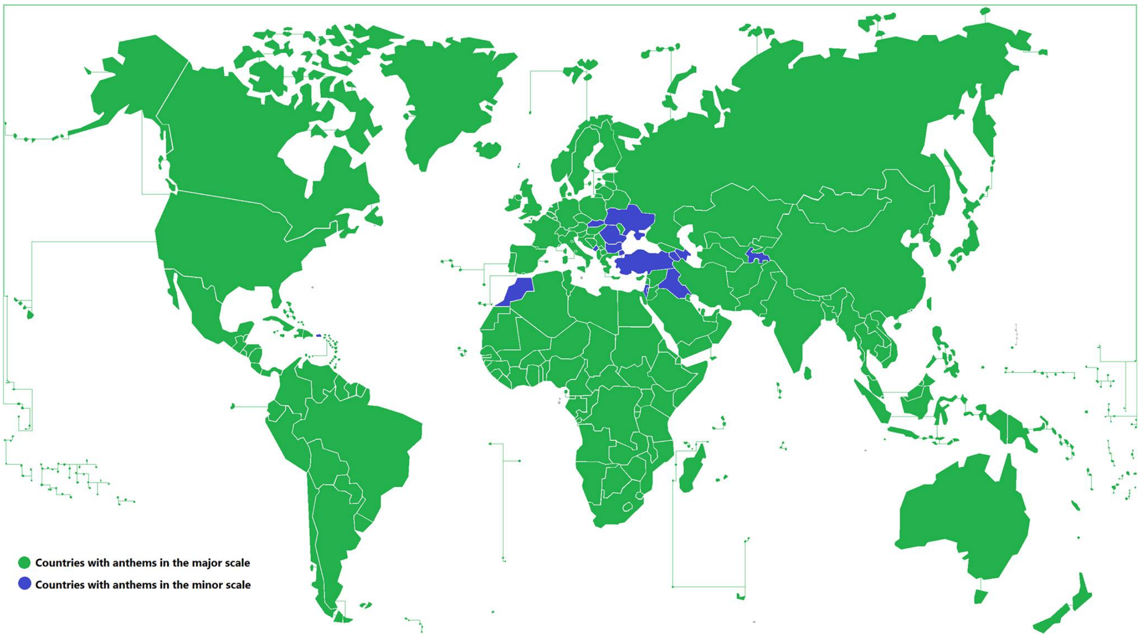


Figure 5: major(=green) and minor(=blue) modes of national anthems[user system637(2015)]

Other examples include a leaning towards rythmical swinging of notes as well as the notion of ending on the primer note or on it's dominant. The development of themes reaccuring multiple times in slightly variatet form is also clearly perceivable, showcasing the attention mechanism, especially with lower temperatures (see 3.2.1 and 3.2.4)
It is also notable, that the generated melodies differ considerably from one another, moving very fast as well as relatively slow through melodic note changes and create a wide spectrum of musical themes, sometimes mixing familiar sounding themes with unexpected changes.
All in all the cultural analysis of the national anthems can be described as a success in the sense of creating an emotionally accessable set of national symbols, seen through the perspective of a neural network as a simplisitic emulation of the human thinking process, of which modern society could see a lot more of in the near future.

4 The national flag

4.1 Definition and symbolism

Beside the national anthem the national flag is the visual symbol for a collective identification within a higher context. It is the international identifying representation for a country or a nation and can create a strong identification object of emotional importance for the people living there. It helps to create a national self-concept and sometimes with a tendency to an almost religious worship. In almost every public political or sporting event you can see a lot of waving flags, either as physical flag or the colors painted on bodies or clothes and so on.

Figure 6: people celebrating on a sports event

Within the concept of a nation as an imagined community[Anderson(1991)] with a collective Identification there is a necessity of such symbols. While nations often refer to a collective of people who belong to common characteristics such as language, tradition, customs or ancestry are ascribed the national flag functions as an semantic summary of this attributes.
Most flag designs are composed of simple general patterns -like stripes, squares, rectangles, etc. - of different colors and sometimes basic symbols thought as an abstract representation of geographical, historical, cultural, religious, ethnic or ideological statements. Some of the designs are cross-country, there are also symbols and colors that appear simultaneously in many flags and thus illustrate similarities between the countries concerned.

4.2 Technical implementation

For our project we recognized there is no nation without a national symbol or a national flag. As a symbolic form of national identity we decided to create a flag for each of our deepworld countries.
One of the most impressive phenomena of neural networks is the ability to recognize and extract patterns in large data sets and helps to find generalization methods.[Radford et al.(2015)Radford, Metz, and Chintala]
Learning reusable feature representations from large unlabeled datasets has been an area of active research.[Radford et al.(2015)Radford, Metz, and Chintala]
One model for visual pattern recognition who came up in the last years is the General adversarial network.

4.2.1 GAN

The model of a generative adversarial net was pronounced in a paper by Ian J. Goodfellow et al., in June 2014[Goodfellow et al.(2014)Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, and Bengio]. They proposed ...a new framework for estimating generative models via an adversarial process, in which they train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probabillity that a sample came from the training data rather than G ...[Goodfellow et al.(2014)Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, and Bengio] Both, the generator and the discriminator in this paper, are defined by multilayer perceptrons made for unsupervised training with backpropagation.
- The discriminator, learns how to distinguish fake from real objects
- The generator, generates new content and tries to fool the discriminator

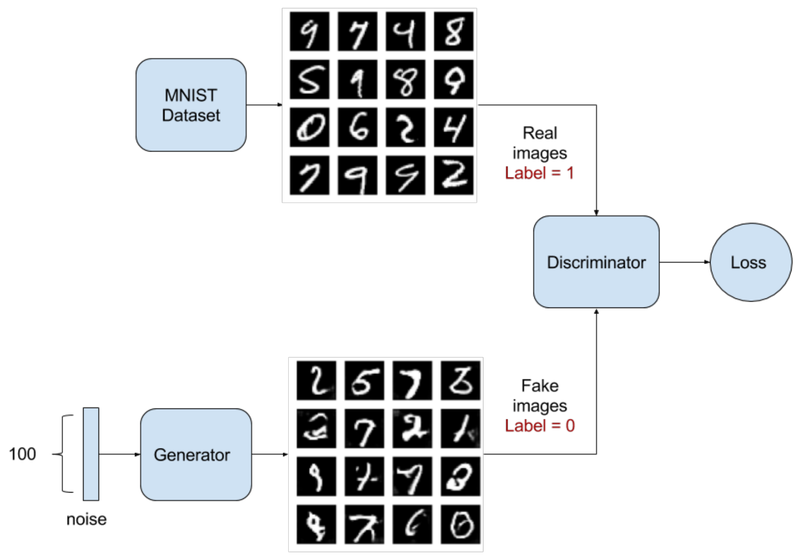


Figure 7: discriminator model is trained to distinguish real from fake handwritten images[9]

In general the adversarial network is based on this value function V(G,D):

$$min_G max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))] \tag{3}$$

To get an better understanding about the connection of generator and discriminator they made a nice intuitive Quote in the official paper by Goodfellow et al.:

> *The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistiguishable from the genuine articles.*[Goodfellow et al.(2014)Goodfellow, P
> p.1]

They draw attention at the end of their paper, that there are still ways to improve:

> *This method of estimating the likelihood has somewhat high variance and does not perform well in high dimensional spaces but it is the best method available to our knowledge. Advances in generative models that can sample but not estimate likelihood directly motivate further research into how*

[9]image from https://cdn-images-1.medium.com/max/800/1*N3nT9AXVnsFBta2R1eEMjg.png

to evaluate such models[Goodfellow et al.(2014)Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Cour p.6]

One of an improved model which appeared about one year later was the proposal of a DCGAN

4.2.2 DCGAN

The DCGAN, deep convolutional generative adversarial network, can be seen as a developed or extended version of a GAN. Proposed by Alec Radford and Luke Metz, they made some changes in the architecture to improve the model of the GAN from J. Goodefellow. Since there have been a lot of advances in image classification, mostly done with convolutional neural networks[Krizhevsky et al.(2012)Krizhevsky, Su , there is an DCNN used for image generation. For the generator and discrimininator they used deep convolutional network models, which bring in some advantages in pattern recognition and data compression. Convolutional Neural Networks take advantage of the fact that the input consists of images and they constrain the architecture in a more sensible way. In particular, unlike a regular Neural Network, the layers of a ConvNet have neurons arranged in 3 dimensions: width, height, depth.[10] [Karpathy()]

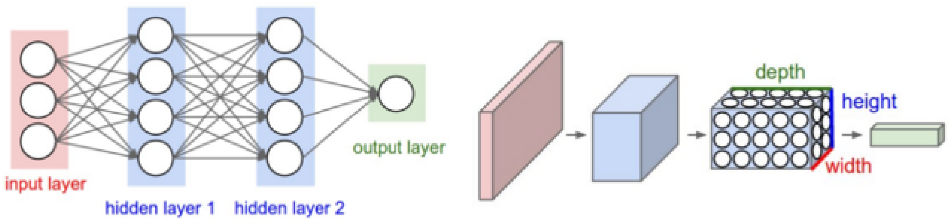


Figure 8: left: a regular multi layer perceptron right: a convolutional network[11]

Another improvement on the CDGANs architecture is the way how the Generator and Discriminator are connected. To take the metaphor of the counterfeiters and the police, you have to imagine for the DCGAN model the counterfeiters now have a spy at the police which tells the counterfeiters what they have to improve on the generated objects to trick the police better. We decided to use the DCGAN for our project to see what is happening when the generator tries to trick the discriminator with new generated flags.

4.3 Preparing the data set

One step of significant importance is the way how you prepare your data set. Therefore we have to consider different aspects in preparing. On one hand there is the format in which you provide the data on a technical layer, data format, resolution, container, etc. The other more important aspect is that we have to think more about the richness or completeness while we are preparing our data set. Equal for which project we want to gain a wider insight or get information of hidden cluster or correlations, we always have to take in to account the phenomena of human bias. The outcome will always be an interpretation of your input. If there are missing information, until now a Artificial Neural network is not able to detect this gap nor to fill it. As an example we feed an image classifier only with picture from cats, dogs and flowers it will never be able to classify a car or a fish. More specific if we train the network with classified pictures from birds and fish, maybe it will classify an

[10] vgl. http://cs231n.github.io/convolutional-networks/
[11] image from: http://cs231n.github.io/convolutional-networks/

airplane as a bird and a plastic bottle as a fish. Also the classification is just a probability and the relation to real existing objects is made by humans. (This is just an basic example. Of course if go deeper into the understanding of machine learning and artificial neural networks we alsways have to think about concepts of learning, knowledge and probabilities. Also take into account concepts of learning in neuro science).
For our project we wanted to generate a new set of national flags based on the database of already existing flags from around the world.
For our purpose we want to see if the network is able to substract common patterns and is able to generate new ones based on this extractions. We didn't want to recognize flags which are surrounded by a background inside a picture. We want to get clear patterns of colour and symbols for our new nations.
Therefore we started with a data set, consisting of pictures from 196 countries around the world. Just the pattern without surrounding. So we started with vector graphics converted into portable network graphics format

Figure 9: Examples from first data set[12]

After our first training of 50 epoch, the test samples of our network weren't satisfying. Unfortunately a few epochs later and some changes in the network model the result wasn't getting better.
We could imagine some patterns, and started speculating that the pictures of data set are to clear with too less information inside the graphics and a lag of noise.

[12]image from: http://www.welt-flaggen.de/herunterladen

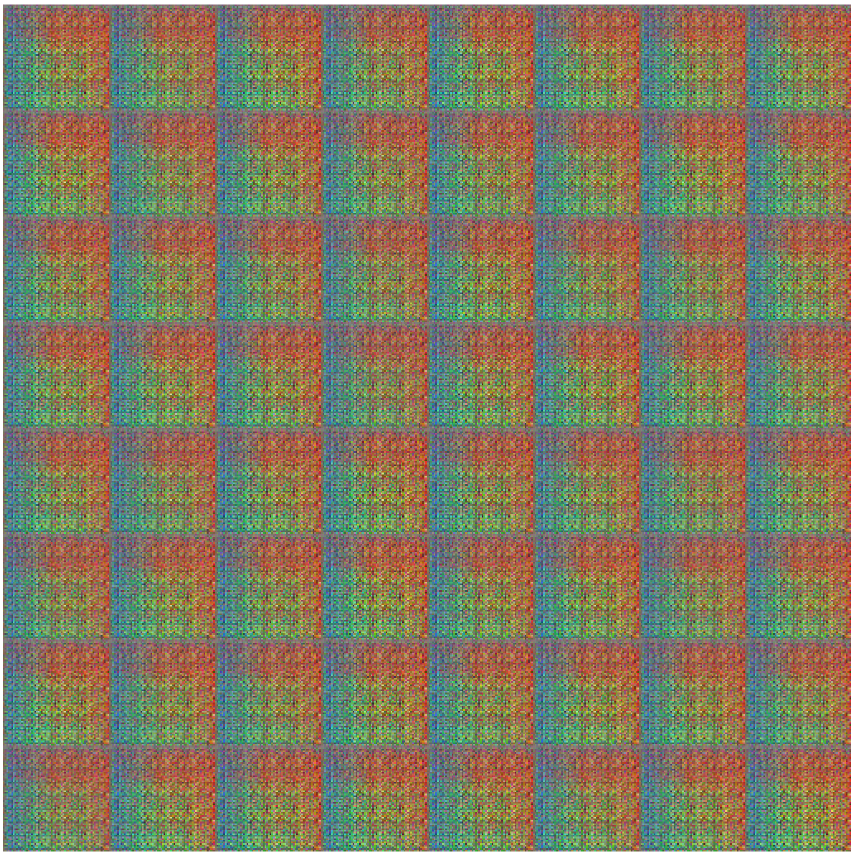


Figure 10: Example Output from first data set

So before we started with tweaking the network model we switched to a Second data set. The second one is a self created database from images we could find online from different flag suppliers. The images are pictures taken from physical flags. So we converted them into a fixed resolution of 256x256 px in jpeg format.

Figure 11: examples second data set

After 50 epochs of training our network fed with the new data set showed us a much more impressive and satisfying output. It already created some patterns like horizontal and vertical separation - and started to distinguish and separate colors.

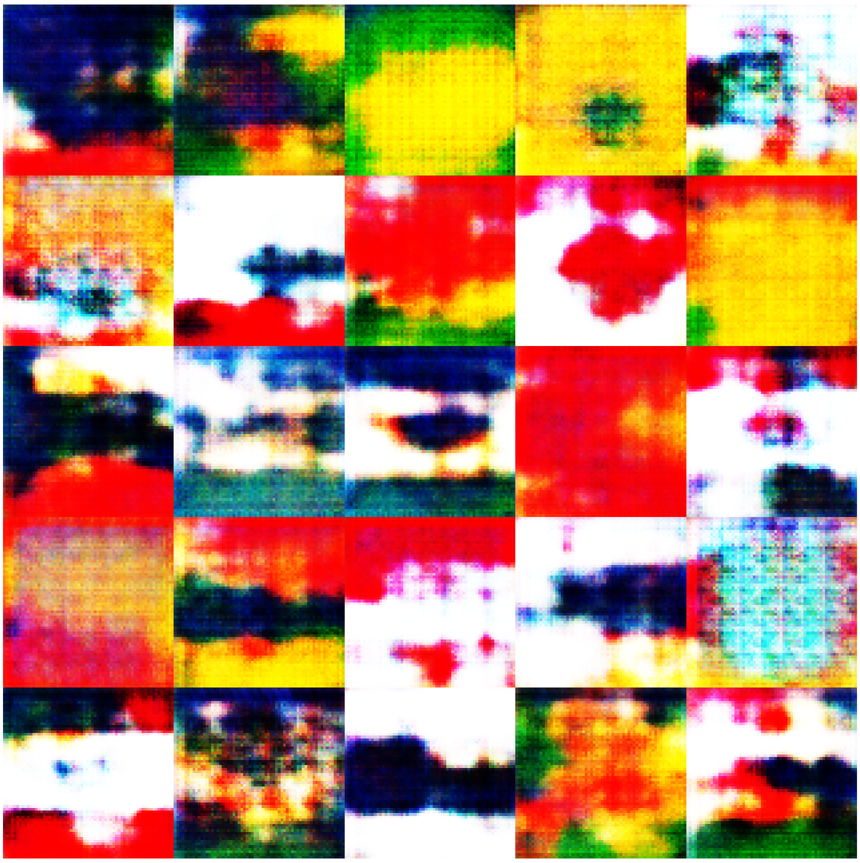


Figure 12: example first output second data set

4.4 Training and optimization

For the training we used a existing implementation of a DCGAN model for tensorflow:

- https://github.com/carpedm20/DCGAN-tensorflow

First we had to understand how to run this model and how we have to use the options of the model for our data set. Since the first training cycles did't bring a result we were happy with, with tried different options but always as trial and error. After some more readings about the implementation we got more and more in a direction we could create some interesting results which gave us a kind of impression of flag patterns. We developed our data set more and took more pictures into the data set, generalized the format of the pictures so it will fit into the options of our model, and as one final result we got some pictures of already recognizable flags.

Figure 13: final result from second data set

4.5 Conclusion and further improvements

Right now our model of the DCGAN showed us some nice results. We can recognize patterns derived from real existing flags. But the pictures are still blurry and in a low resolution. While our research we stumbled upon some other implementations of GAN models (e.g. WassersteinGAN) which could bring us more to clear recognizable flag patterns.

References

[Anderson(1991)] Benedict Anderson. *Imagined Communities: Reflections on the Origin and Spread of Nationalism. Revised Edition ed. London.* Verso, 1991.

[Bahdanau et al.(2014)Bahdanau, Cho, and Bengio] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. pages 1–15, 2014. ISSN 0147-006X. doi: 10.1146/annurev.neuro.26.041002.131047. URL `http://arxiv.org/abs/1409.0473`.

[Bridle(1990)] John S. Bridle. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. *Neurocomputing*, pages 227–236, 1990.

[Coliva(1997)] a. Coliva. *Wittgenstein and the Philosophical Investigations*, volume 32. 1997. ISBN 0333007611.

[Douglas Eck(2016)] Douglas Eck. Welcome to Magenta!, 2016. URL `https://magenta.tensorflow.org/welcome-to-magenta`.

[Fornäs(2012)] Johan Fornäs. *Signifiying Europe*. 2012.

[Goodfellow et al.(2014)Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville, and Bengio] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In Z Ghahramani, M Welling, C Cortes, N D Lawrence, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`.

[Hochreiter and Schmidhuber(1997)] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Comput.*, 9(8):1735–1780, nov 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL `http://dx.doi.org/10.1162/neco.1997.9.8.1735`.

[Karpathy()] Andrej Karpathy. Generative Models. URL `https://blog.openai.com/generative-models/`.

[Karpathy(2015)] Andrej Karpathy. The Unreasonable Effectiveness of Recurrent Neural Networks, 2015. URL `http://karpathy.github.io/2015/05/21/rnn-effectiveness/`.

[Krizhevsky et al.(2012)Krizhevsky, Sutskever, and Hinton] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25*, pages 1097–1105, 2012. URL `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`.

[Magenta(2017)] Google Magenta. melodyrnn_readme, 2017. URL `https://raw.githubusercontent.com/tensorflow/magenta/master/magenta/models/melody_rnn/README.md`.

[Martin Daughtry(2003)] J Martin Daughtry. Russia's New Anthem and the Negotiation of National Identity. *Ethnomusicology*, 47:42, 2003.

[NFL(2014)] NFL. Lady Gaga sings the National Anthem at Super Bowl 50 - NFL Videos, 2014. URL `http://www.nfl.com/videos/nfl-network-total-access/0ap3000000634472/Lady-Gaga-sings-the-National-Anthem-at-`

[Radford et al.(2015)Radford, Metz, and Chintala] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *CoRR*, abs/1511.0, 2015. URL `http://arxiv.org/abs/1511.06434`.

[user system637(2015)] Reddit user system637. Anthems: Major scale vs minor scale [OC] [2048x1148] : MapPorn, 2015. URL https://www.reddit.com/r/MapPorn/comments/31j9az/anthems_major_scale_vs_minor_scale_oc_2048x1148/.

[Waite(2016)] Elliot Waite. Generating Long-Term Structure in Songs and Stories, 2016. URL https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn/.

[Williams and Zipser(1989)] Ronald J. Williams and David Zipser. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, 1(2):270–280, 1989. ISSN 0899-7667. doi: 10.1162/neco.1989.1.2.270. URL http://www.mitpressjournals.org/doi/10.1162/neco.1989.1.2.270.