AML Project Proposal

Christian Schwarz and Milena Bruseva

1 Proposal: This Movie Poster Doesn't Exist

1. **Which scientific questions do you want to answer, and why are they interesting?** The aim of this project would be to generate unseen before movie posters. We have observed correlations between styles and genres, as well as subgenre-specific features. Posters have a large number of attributes, e.g. colour palette, structure, objects, level of abstraction, whether it's drawn, actors present, etc., which can enable an interesting approach to the task of content generation.

2. **Which relevant papers did you find? What are their pros and cons? Which among these papers do you propose to build upon, and why?** Currently, there is no specific paper which focuses on movie poster generation, while there are several regarding classification of its genre and IMDB score.

   For movie-poster generation specifically, we were able to find an article (`https://www.johnkraszewski.com/media-synthesis`), which made use of StyleGANs and transfer learning. The author was able to create new posters, however they appear very abstract, and it is difficult to infer the content of the supposed movie, which is the purpose of such a poster. One pro of the author's approach would be the use of transfer learning, which showed improved results and reduced training time. However, one limitation is the lack of conditioning, which could be addressed by a model which takes this into account, such as a conditional GAN.

3. **What methods will you try, and why do you consider them promising to answer the questions?** Going through the related work for Image Generation tasks, showed us that GANs are still the current state of the art. In the task of conditional movie poster generation, we would start by considering the standard cGAN (). We would then move on and expand our feature vector to include encodings of the attributes mentioned above. In order to improve quality and stability during training, we would make use of techniques used in models, which have been traditionally applied to the task of text-to-image generation. This includes GAN-CLS, which facilitates two different kinds of input into the network. StackGAN is another promising approach, which has been shown to improve image quality. The StackGAN has two stages of generation, which we hypothesise will be especially beneficial, because movie posters need generation of different elements and their composition into a larger image.

4. **What will your data sources be? How large are the data sets and how high do you think the quality is? Can you use simulated data when real data is scarce?**

   We found several datasets on Kaggle, which each had 39.515, 8.252, 41.979 thousand movie posters. In addition to this, the official statistics from IMDB show that they have 2,235,966 Titles w/ Primary Image, which can be used to augment our dataset if the need arises. As the posters come from the official source, we can be assured of their quality. Real data is not scarce in our case. Moreover, we discovered a high quality dataset with 17 thousand instances of movie posters with the text removed.

   In addition to this, the standard practices of rotation, cropping, translation and color changes would not be applicable, as the dataset is not invariant to those transformations. For comparison, the standard datasets for text-to-image synthesis, Oxford-102 and CUN-200, only include 8,192 and 11,788 images respectively.

5. **What computational resources do you need (e.g. GPUs)? How will you get access to these? How much time will the computations need?** As we are dealing with images and GANs, we will need GPUs. However, both of us have access to Nvidia GTX 1080Ti and RTX 3060. Based on the article, mentioned above, we expect to be able to train within reasonable time, even if we use a different architecture, as the author mentions they needed about 2 hours per epoch/tick.

6. **What difficulties do you anticipate in the project?**

   - Image size is too large → resolve by down sampling, StackGAN

- Lack of realism (human faces) → limit data to a subset such as animation or other drawn posters (older posters were traditionally drawn)
- Mode Collapse → W-GAN (which uses Wasserstein loss) has been shown to be an effective technique
- Too big variation inside genre → pre-clustering with handcrafted or automatically generated features
- Text → either choose textless posters as input or perform pre-processing to remove
- Era-dependent styles → use eras as features
- How to evaluate → Inception Score

2 Proposal: Deep Reinforcement Learning in OpenAI's Gym

1. **Which scientific questions do you want to answer, and why are they interesting?** Games are popular yet oftentimes hard problems in Machine Learning of great interest for research organizations such as DeepMind. They can be used as a benchmark for the progress of state-of-the-art machine learning methods. OpenAI has developed a stable, standardized framework for agent-environment interaction, namely, OpenAI Gym. We aim to use Gym to implement reinforcement learning models using deep neural networks. The target games are Atari games provided by the gym environment, but we wouldn't exclude using their retro framework to import other games.

2. **Which relevant papers did you find? What are their pros and cons? Which among these papers do you propose to build upon, and why?** Relevant papers are Evolving Neural Networks through Augmenting Topologies, which has been implemented for video games, e.g. Rolling Horizon NEAT for General Video Game Playing. Additionally, publications from DeepMind such as Player of Games and the MuZero paper are state-of-the-art solutions to the task. Unfortunately for us, the resource available to DeepMind, which enables them to get superhuman performance, i.e. huge GPU time, isn't available to us. NEAT on the other hand is also promising for video game agents, as can be seen in various YouTube videos (i.e. `https://youtu.be/dkvFcYBznPI`), which is why we propose to build upon it or some variation of it.

3. **What methods will you try, and why do you consider them promising to answer the questions?** We are going to try NEAT for our implementation, because of the reasons listed above.

4. **What will your data sources be? How large are the data sets and how high do you think the quality is? Can you use simulated data when real data is scarce?** In this case, rather than loading large datasets, we are going to be running simulations in the OpenAI Gym and thus training data is generated dynamically. To allow for an adaptive strategy regarding exploration vs exploitation, this dynamic approach is required.

5. **What computational resources do you need (e.g. GPUs)? How will you get access to these? How much time will the computations need?** To speed-up training, i.e. training several instances of models in one generation of NEAT, we would make use of GPUs, in addition to the neural network computations. As described above, we already have access to two. Our lack of TPUs is the main reason we refrain from implementing the DeepMind algorithms.

   The required computation time is hard to estimate, as the original authors of the paper we plan to build upon do not disclose the hardware they used. In addition to this, the time required is dependent on the game and on the chosen parameters, e.g. population size. According to the authors, the agent can interact with the game in real time.

6. **What difficulties do you anticipate in the project?**

   - Choosing the right hyperparameters for NEAT → can potentially be solved if the ones listed in the paper are applicable
   - Evaluation standardization → different games assign scores for different things. If we want to implement it for different games and compare performance, we would need to come up with a means of generalizing.
   - Unknown required training time → preliminary tests with a small game, save the wights and do comparisons over time
   - Problems converging → Combine NEAT for the layout and training through back-propagation for the weights of the network