Deep conditional generative models

Deep generative models (DGMs), such as GANs (generative adversarial networks) and VAEs (variational autoencoders), have achieved great success in recent years, especially in the domains of images, cheminformatics and 3D point clouds. By modeling the distribution of data, we can create more examples that are like those already in a dataset, but not the same. However, since DGMs are usually trained in an unsupervised way, it is difficult to control the properties of generated samples, see Figure 1 for our preliminary results. There are numerous situations, where the generation process has to satisfy certain conditions or constraints:

- In **explainable artificial intelligence**, we want to discover which attributes of input examples are responsible for making a given prediction. This can be realized by generating counterfactual examples, which change the decision of the model using slight modifications of interpretable attributes.
- In **cheminformatics**, generating molecules with desired properties is the first step in constructing drug candidates. A drug should be active on a fixed set of biological targets, and inactive to the remaining ones to have no side effects.
- In **image inpainting**, we look for a realistic filling of the missing region, which is consistent with the known part of the image. Since there are multiple solutions satisfying the imposed constraints, it is natural to employ generative models, which can return multiple results.

Motivated by the importance of the above problems we aim at constructing conditional DGMs, which allow for controlling the process of generating examples.

Building a new conditional DGM from the ground up is a challenging problem from the conceptual and computational point of view because it requires selecting specific architecture with multiple hyperparameters, which is time and resource consuming. Moreover, this introduces an additional effort when one wants to adapt a designed solution to new application or domain. This problem is especially important, because it is often physically impossible to train large scale models from scratch without access to the computer power used by the leading companies, such as Google or Facebook.

To tackle various conditional problems, we follow an emerging trend in deep learning, which relies on **adapting well-established pre-trained models** to diverse applications with a minimal conceptual and training effort. We plan to design **lightweight plugin networks**, which work on representation constructed by pre-trained DGMs and modify their latent codes. We hypothesise that plugin networks provide a sufficient control of the generative process. Moreover, plugin networks are more convenient in **transferring solutions between different domains and architectures**. This hypothesis will be verified by applying our plugins to generating counterfactual examples, chemical molecules and 3D point clouds.

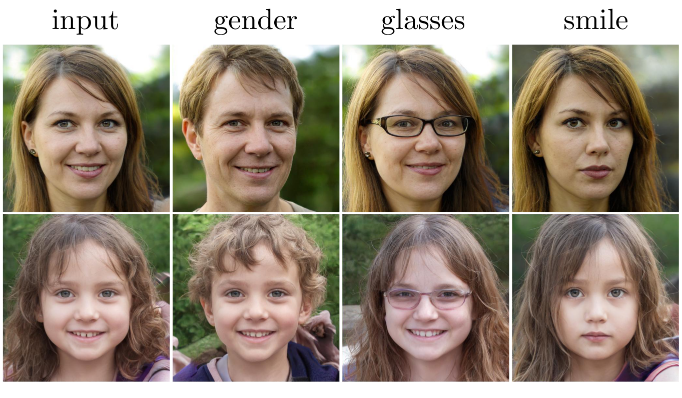

Figure 1: Manipulation of face attributes (gender, glasses, smile) performed by our preliminary plugin model.

Our solutions will have direct consequences in explaining decisions of deep learning models, which is currently one of the basic requirements in applying artificial intelligence tools is safety domains. Our results will also have the impact in real-life problems, such as computer-aided drug design. We will also work on constructing deep learning models for solving inverse problems, such image inpainting. Since we plan to deal with these interdisciplinary topics, our project will have a broader impact outside pure machine learning. To popularise our methods, we plan to publish the source codes of developed algorithms on services like GitHub (`https://github.com/`).