

Synthetic data in deep learning

Sergey Nikolenko^a

*^aLaboratory of Mathematical Logic of the St.-
Petersburg Department of the Steklov Mathematical
Institute, Russian Academy of Science*

Abstract

Many major problems of modern AI come down to data: either lack of data or, also very often, lack of labeled data. Synthetic data is an important approach to solving the data problem by either producing artificial data from scratch or using advanced augmentation techniques to produce novel and diverse training examples. In the talk, I will introduce the notion of synthetic data and various approaches to making and using it. In particular, we will discuss domain adaptation, a set of techniques designed to make a model trained on one domain of data, the source domain, work well on a different, target domain. This is a natural fit for synthetic data: in almost all applications, we would like to train the model in the source domain of synthetic data but then apply the results in the target domain of real data. We will survey DA approaches for synthetic-to-real adaptation, concentrating on deep learning models. We will see the gaze estimation story from Apple's Refiner and beyond, DA techniques for learning to drive, GAN-based DA for medical imaging, and much more. Expect a lot of GANs and loss functions.
