Abstract:

We study the ecological use of analogies in AI. Specifically, we address the problem of transferring a sample in one domain to an analog sample in another domain. Given two related domains, S and T, we would like to learn a generative function G that maps an input sample from S to the domain T, such that the output of a given representation function f, which accepts inputs in either domains, would remain unchanged. Other than f, the training data is unsupervised and consist of a set of samples from each domain, without any mapping between them. The Domain Transfer Network (DTN) we present employs a compound loss function that includes a multiclass GAN loss, an f preserving component, and a regularizing component that encourages G to map samples from T to themselves. We apply our method to visual domains including digits and face images and demonstrate its ability to generate convincing novel images of previously unseen entities, while preserving their identity.

Joint work with Yaniv Taigman and Adam Polyak