

# Learning Object Representations by Mixing Scenes

Zbinden Lukas

Master thesis in Computer Science

In the digital age of ever increasing data amassment and accessibility, the demand for scalable machine learning models effective at refining the new oil is unprecedented. Unsupervised representation learning methods present a promising approach to exploit this invaluable yet unlabeled digital resource at scale. However, a majority of these approaches focuses on synthetic or simplified datasets of images. What if a method could learn directly from natural Internet-scale image data? In this thesis, we propose a novel approach for unsupervised learning of object representations by mixing natural image scenes. Without any human help, our method mixes visually similar images to synthesize new realistic scenes using adversarial training. In this process the model learns to represent and understand the objects prevalent in natural image data and makes them available for downstream applications. For example, it enables the transfer of objects from one scene to another. Through qualitative experiments on complex image data we show the effectiveness of our method along with its limitations. Moreover, we benchmark our approach quantitatively against state-of-the-art works on the STL-10 dataset. Our proposed method demonstrates the potential that lies in learning representations directly from natural image data and reinforces it as a promising avenue for future research.

Prof. Paolo Favaro