

# LEARNING TO TRANSFER ACROSS DISTINCT DOMAINS WITH TRIPLY ADVERSARIAL NETS

BY GILLES LOUPPE

*New York University*

Transfer learning, and its variants known as learning under co-variate shift or domain adaptation, arises whenever test instances are governed by a distribution that may be arbitrarily different from the distribution of the training instances. This problem has traditionally been solved either by reweighting training instances or by learning robust feature representations over domains. In this work, we propose a new paradigm which consists instead in learning how to transform training instances into test instances, possibly across distinct input spaces of different dimensions and structures. For this purpose, we extend the generative adversarial networks framework of [Goodfellow et al. \(2014\)](#) to a triply adversarial process: a transformer network  $T$  for generating test instances from training instances, a discriminative network  $D$  for estimating whether an instance comes either from the training or the test distributions, and a classifier network  $C \circ T$  for classifying training instances in the projected space. This 3-player game results in a network  $T$  capable of transforming training instances into test instances, while preserving the separation between classes, so that  $C$  would remain an accurate classifier on test data. Experimental results demonstrate the potential of this novative approach for transfer learning.

## 1. Introduction.

## 2. Method.

## 3. Experiments.

## 4. Related work.

## 5. Conclusions.

## References.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680.