Natural Variation Transfer using Learned Manifold Operators

Marissa Connor and Christopher Rozell
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: (marissa.connor,crozell)@gatech.edu

Benjamin J. Culpepper and Huy Nguyen
Yahoo! Research
110 5th Street
San Francisco, CA 94103
Email: jackcul@yahoo-inc.com, huy.xnguyen@gmail.com

In many types of data, variations are constrained by natural physical laws, and identity-preserving variations (like changes in viewpoint) are often transferable from one class to another. In many cases, the within-class variation in a high-dimensional dataset can be modeled as lying on or near a low-dimensional, nonlinear manifold [1]. If the manifold structure shared between classes can be learned, it can be exploited for transfer learning tasks. In this work, we represent the manifold structure using a learned dictionary of generative operators and develop methods for using those operators for few-shot learning and realistic data generation.

Most existing approaches either define the manifold structure through embeddings of the original data points [2], [3] or through estimates of local tangent planes [4], [5], [6]. These approaches do not generalize well to manifold locations not well-represented in the training data because they either have simple or non-existent generative models for the data. Another line of work has proposed unsupervised learning algorithms to learn Lie group operators that represent the structure of low-dimensional manifolds [7], [8], [9], [10]. This manifold representation defines the process of transforming one data point to another using analytic operators. We assume that two neighboring points \( x_0, x_1 \in \mathbb{R}^N \) on a low-dimensional manifold can be related through the following dynamical system:

\[
x_1 = \expm \left( \sum_{m=1}^{N} \Psi_m c_m \right) x_0 + n,
\]

where \( n \) is the error, and \( \expm \) is a matrix exponential. The dynamics matrix is defined by a weighted sum of dictionary elements \( \{\Psi_m \in \mathbb{R}^{N \times N}\} \) which we call transport operators. For each pair of points, the geometry is governed by a small number of operators through the weighting coefficients \( c \in \mathbb{R}^M \).

The relationship between points in (1) can be used to write a probabilistic generative model that enables efficient inference. Following [7], we assume a Gaussian noise model, a Gaussian prior on the transport operators, and a sparsity inducing prior on the coefficients.

The resulting negative log posterior for the model is:

\[
\frac{1}{2} \left\| x_1 - \expm \left( \sum_{m=1}^{M} \Psi_m c_m \right) x_0 \right\|_2^2 + \frac{\gamma}{2} \sum_{m} \left\| \Psi_m \right\|_F^2 + \zeta \left\| c \right\|_1
\]

where \( \left\| \cdot \right\|_F \) is the Frobenius norm. Following the algorithm in [7], we use pairs of neighboring training points to learn the transport operators using descent techniques (alternating between coefficients and transport operators) on the objective in (2).

Using transport operators, we develop methods to transfer manifold representations of variations from a source domain to a sparsely sampled target domain in order to generate new samples and enrich the target domain (see Fig. 1 for illustration). We present two transfer techniques that both use the generative manifold model but target different transfer approaches. In the first approach, we map out a manifold around individual data points by applying the transport operators with coefficients, \( w_{i,m} \), sampled from a specified probability distribution, to each data point to generate new samples:

\[
x_i = \expm \left( \sum_{m} \Psi_m w_{i,m} \right) x_0
\]

We apply this approach to a few-shot learning application with USPS handwritten digit data [11]. We train a single transport operator using pairs of examples of the digit ‘8’. Each pair consists of one rotated 8 and that same image rotated an additional \( 1^\circ \). The original USPS training set can be augmented by applying the learned transport operator to training examples. We train a convolutional neural network classifier using three different training sets: 1) the original USPS digits, 2) the USPS digits transformed by ground truth rotation matrices, 3) the USPS digits transformed by applying the learned transport operator [12]. Fig. 2 shows the transport operator augmentation significantly improves classification accuracy over the original dataset.

In the second transfer approach, we generate paths associated with desired transformations. Our developed method infers a transformation between a starting point \( (x_0) \) and an ending point \( (x_1) \) that can operate on new starting points in order to transfer the manifold path. A transformation between \( x_0 \) and \( x_1 \) is defined by a set of inferred coefficients \( c^{*} \):

\[
c^{*} = \arg\min e^{\frac{1}{2} \left\| x_1 - \expm \left( \sum_{m=1}^{M} \Psi_m c_m \right) x_0 \right\|_2^2 + \zeta \left\| c \right\|_1
\]

We demonstrate this approach on facial expression sequences. The transport operators are trained on facial landmark points from pairs of images in expression sequences in the source domain (the MUG facial expression database [13]). These transport operators are used to create expression sequences for new subjects by extrapolating an expression from landmark points associated with a neutral face. To do this, the coefficients are inferred between landmark points from an image at the apex of the expression and a neutral image in the same sequence. These coefficients define a dynamics matrix that can be applied to a new starting point. We incorporate the generated landmark sequences into a progressive Generative Adversarial Network (GAN) [14] that is trained on the celebA dataset [15] by conditioning the generator on the landmark point locations. Figure 3 shows examples of facial animation using transport operators. The top row shows an expression sequence from the target domain (Extended Cohn-Kanade (CK+) database [16], [17]). Coefficients are inferred between landmarks in the first and last image in the top row sequence and used to estimate the expression transformation that is applied to landmark points for several starting faces. The trained generator network is conditioned on the generated landmark expression sequences in order to output the faces shown in rows 2 - 4.
Fig. 1: An illustration of the two transfer tasks. The transport operators learned from a densely sampled source domain (red) are used to apply transformations to points in a target domain (blue). The first approach applies transport operators with coefficients drawn from a probability distribution to a single point. The second approach infers a transformation between points in the source domain and applies that transformation to a starting point in the target domain to generate a path.

Fig. 2: Classification accuracy from convolutional neural network classifiers tested on rotated USPS digits. The classifiers are trained on USPS digits in three variations: the original data, data augmented with true rotation matrices, and data augmented through transfer learning using transport operators trained while only seeing rotated ‘8’ digits.

Fig. 3: The top row of each image shows a ground truth expression sequence from the CK+ dataset. Rows 2-4 show the faces generated with landmarks from expression sequences that are extrapolated from new neutral faces using inferred manifold paths. Each row is the result of a single latent vector.

REFERENCES