Learning Interpretable Representation for 3D Point Clouds

ICPR 2020

[1]Feng-Guang Su, [2]Ci-Siang Lin, [2,3]Yu-Chiang Frank Wang
[1]Language Technologies Institute, Carnegie Mellon University, USA
[2]Graduate Institute of Communication Engineering National Taiwan University, Taiwan
[3]ASUS Intelligent Cloud Services(AICS), Taiwan

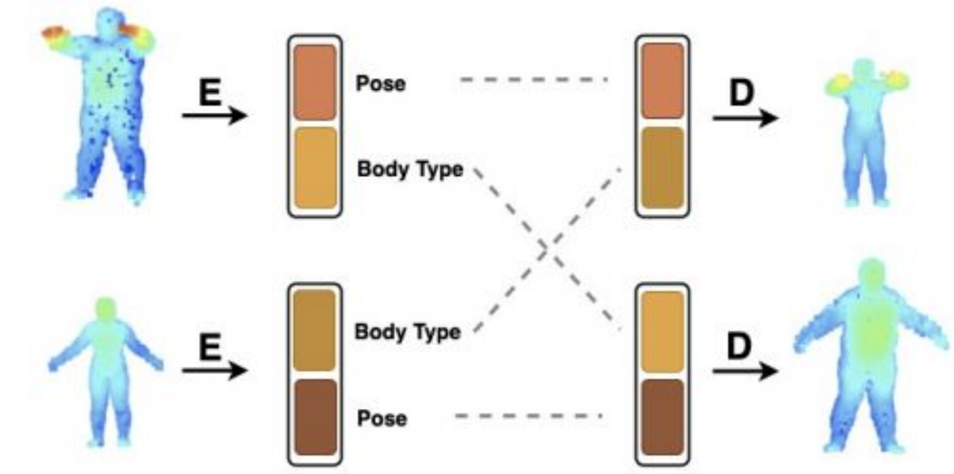
3D Point Clouds

Point clouds have emerged as a popular representation of 3D visual data. With a set of unordered 3D points, one typically needs to transform them into latent representation before further classification and segmentation tasks.

- They're generally comprise of the raw output data from most 3D data acquisition devices.
- It avoids the memory issue through surface representation.
- It doesn't require the point-wise connectivity information like mesh which might not be obtained in practice.

Representation Disentanglement for 3D Point Clouds

- One cannot easily interpret such encoded latent representation.
- Due to the lack of order information, it is not easy to interpret the latent feature derived by existing deep learning models.
- It is much harder to extract and manipulate attributes of interest.

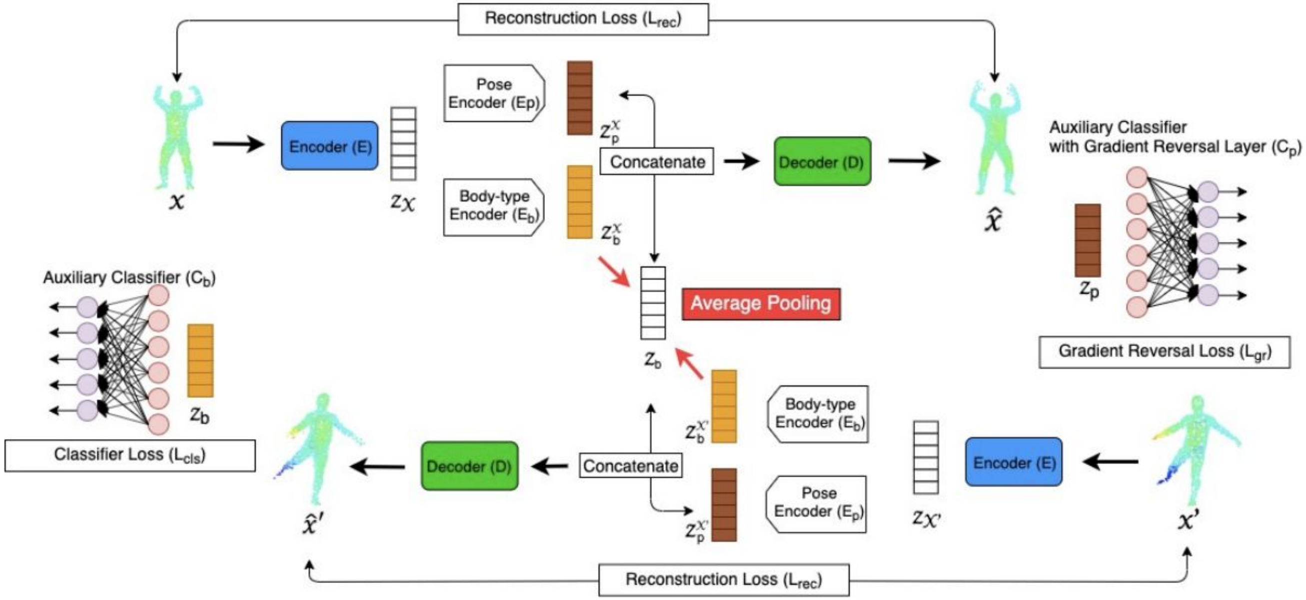


Challenges

- Paired data such as different people do the same actions are hard to collect in practice, which therefore are not available in our setting.
- Because pose information is hard to be represented as an one-hot vector or a multi-hot vector. As a result, we choose to learn pose representation in a totally data-driven manner instead of being guided by any manual labels.

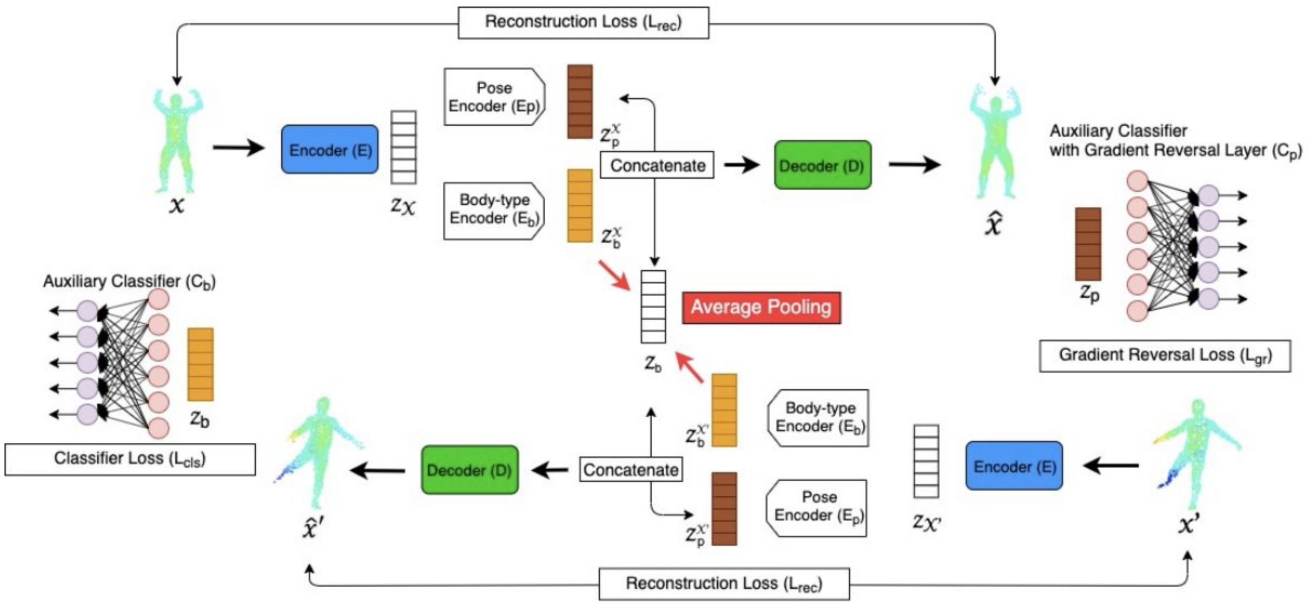
Proposed Method

- Our model is end-to-end learnable, which extracts body-type and pose information by advancing adversarial learning and data recovery consistency without observing pose label information.



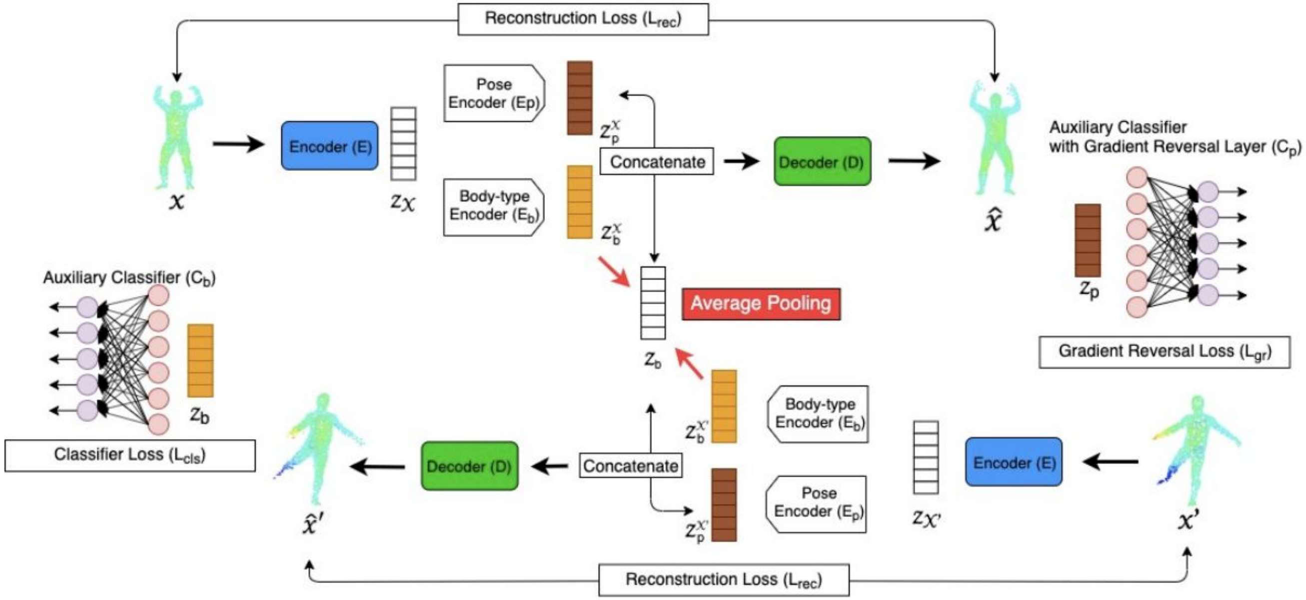
Learning Latent Representation for Body Types - 1

- We deploy an identity (ID) classifier C_b to enforce the resulting latent vector z_b capturing identity (i.e., body-type)



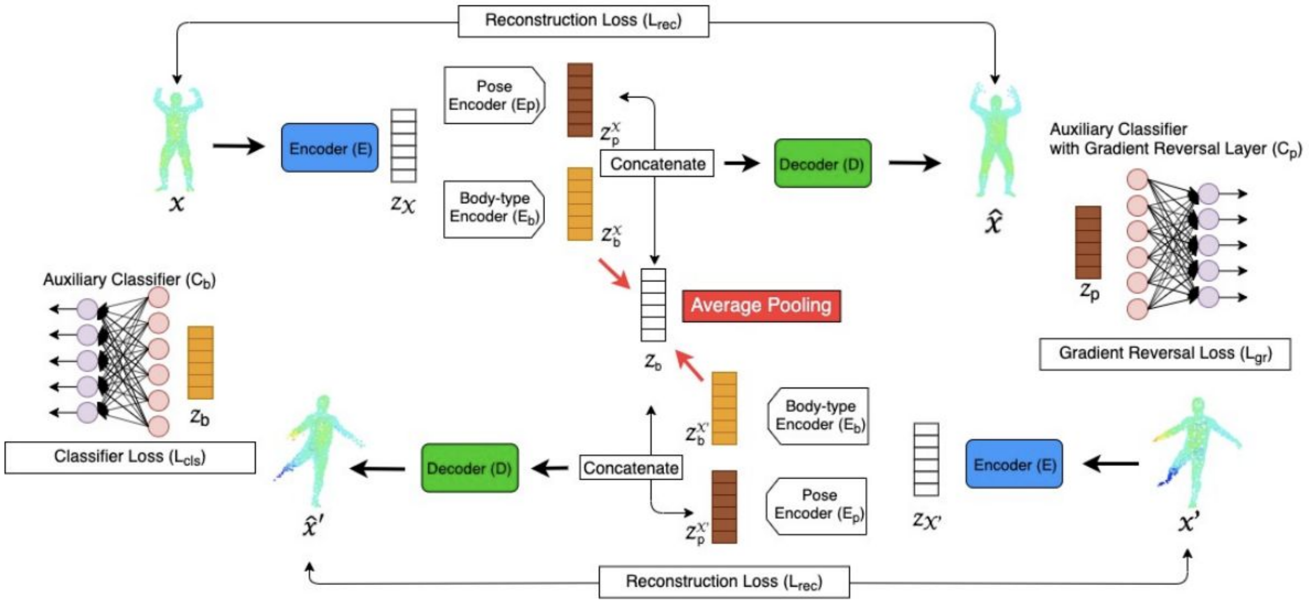
Learning Latent Representation for Body Types - 2

- Since x and x’ represent a pair of point cloud data with different poses but of the same person, their body type vectors should be the same. Therefore, we apply an average-pooling layer on the latent space to derive z_b.



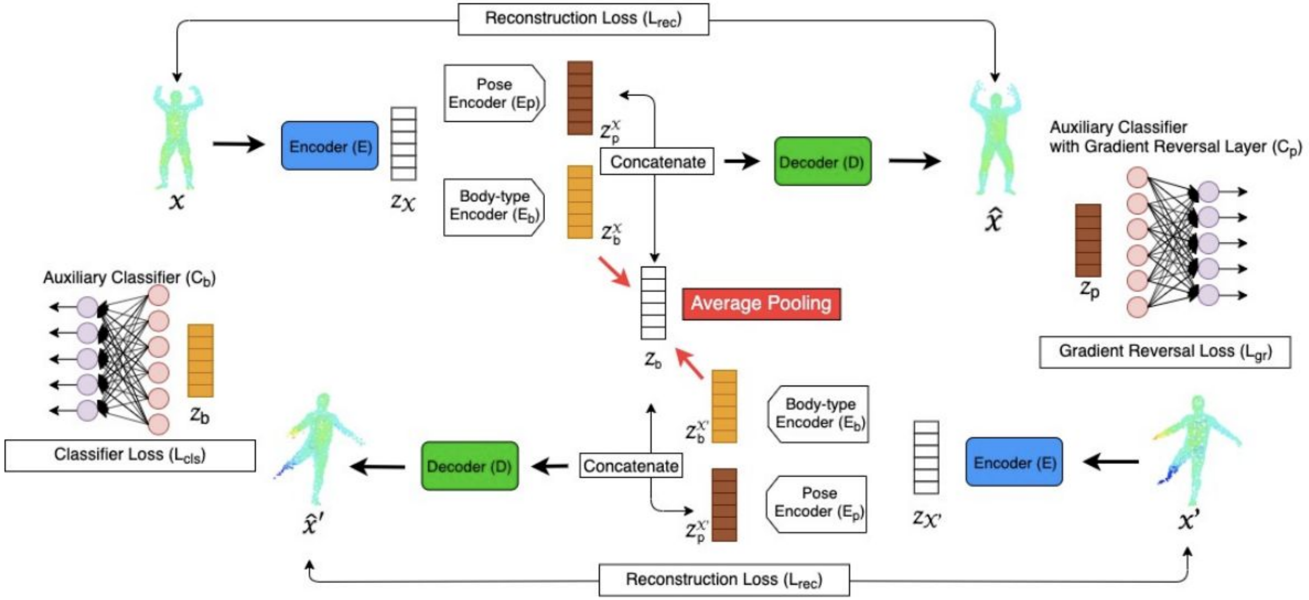
Learning Latent Representation for Poses - 1

- We deploy an auxiliary classifier C_p with a gradient reversal layer to enforce the pose feature vector excluding body-type information.



Learning Latent Representation for Poses - 2

- We further propose a cross-consistency concept for capturing pose information in such unsupervised fashions and calculate the reconstruction loss through Chamfer Distance and Projected Chamfer Distance.



Quantitative Results

- EMD - Earth Mover Distance
- CD - Chamfer Distance
- PD - Projected (Chamfer) Distance

Method	**MMD-EMD [10]**	**MMD-CD [9]**	**MMD-PD**
VAE [8]	0.09469	0.00099	0.00023
AE [9]	0.12159	0.00154	0.00053
Fader [18]	0.13586	0.00186	0.00038
DRIT [22]	0.19400	0.00970	0.00235
ACGAN [16]	0.27210	0.00548	0.00134
Ours	**0.07496**	**0.00079**	**0.00018**

TABLE I

RECONSTRUCTION PERFORMANCES OF VAE, ACGAN, FADER NETWORKS, DRIT AND OURS ON D-FAUST IN TERMS OF EMD, CHAMFER DISTANCE, AND PROJECTION DISTANCE. THE NUMBERS IN BOLD INDICATE THE BEST RESULTS.

Ablation Study

- EMD - Earth Mover Distance
- CD - Chamfer Distance
- PD - Projected (Chamfer) Distance

Method	MMD-EMD [10]	MMD-CD [9]	MMD-PD
Ours -cls	0.14278	0.00281	0.00094
Ours -proj	0.11691	0.00095	0.00023
Ours -gr	0.08684	0.00180	0.00051
Ours -cross	0.08207	0.00122	0.00022
Ours	**0.07496**	**0.00079**	**0.00018**

TABLE II

ABLATION STUDIES OF OUR MODEL DESIGN ON D-FAUST IN TERMS OF EMD, CHAMFER DISTANCE, AND PROJECTION DISTANCE. NOTE THAT OUR FULL VERSION (OURS) ACHIEVES THE BEST RESULT.

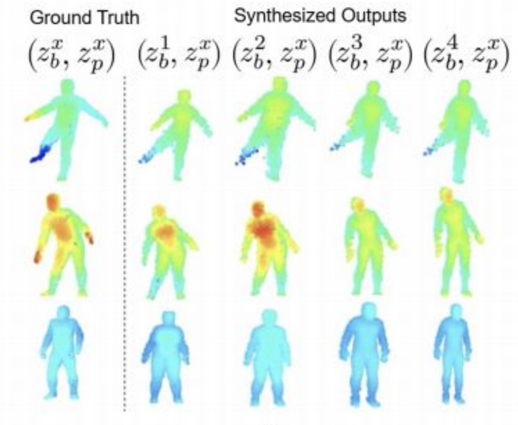
Feature Disentanglement

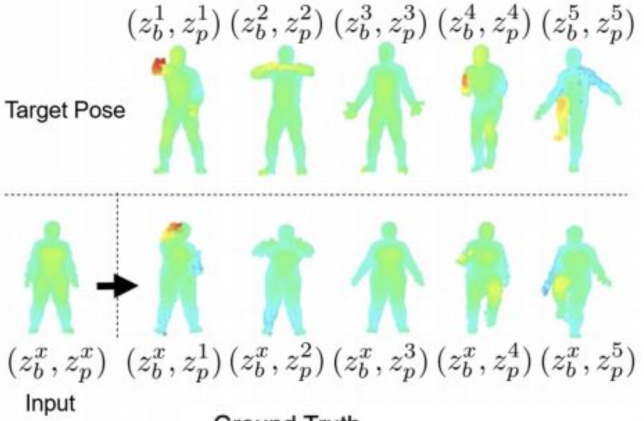
- To demonstrate the effectiveness and necessity of our body-type and pose feature disentanglement, we retrain different versions of the body-type classifier, taking different types of learned embeddings as the input.

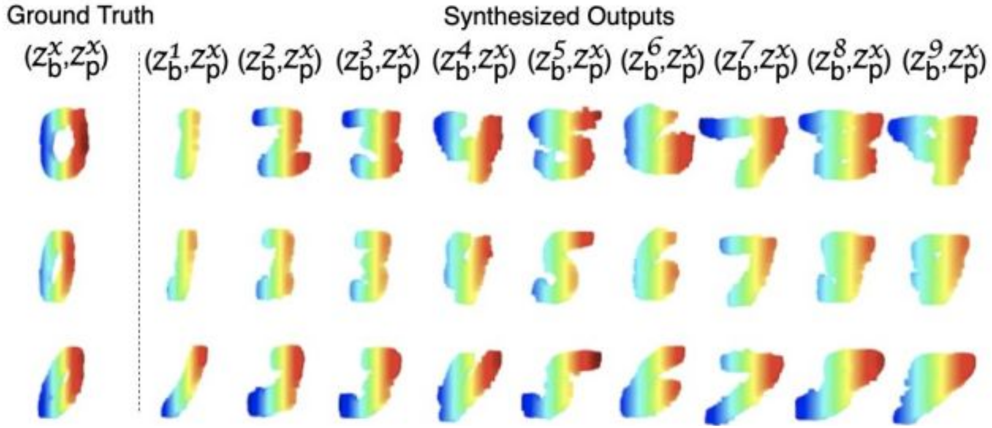
Method	z	z_b	z_p
AE [9]	0.775	-	-
Fader [18]	0.800	-	0.487
DRIT [22]	0.915	0.446	0.361
Ours -gr	0.834	0.884	0.660
Ours	0.781	**0.896**	**0.137**

TABLE III
BODY-TYPE CLASSIFICATION USING LATENT VECTORS OF DIFFERENT MODELS. NOTE THAT z IS DERIVED BY AE, WHILE z_b AND z_p ARE THOSE DESCRIBING BODY-TYPE AND POSE INFORMATION, RESPECTIVELY.

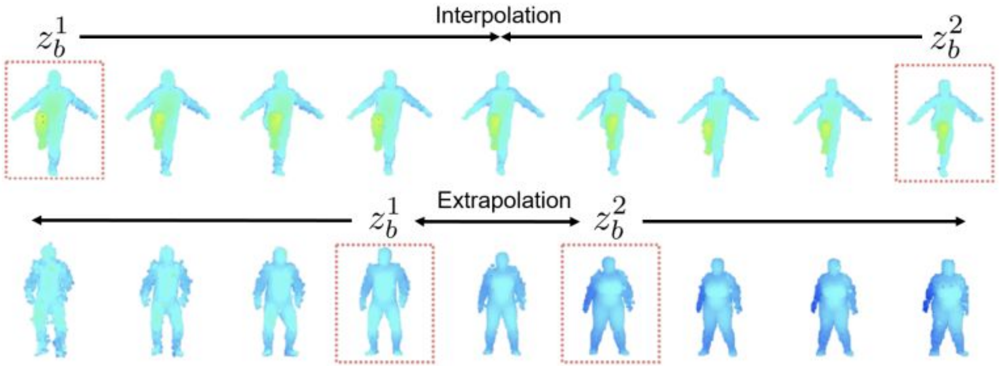
Qualitative Results - 1

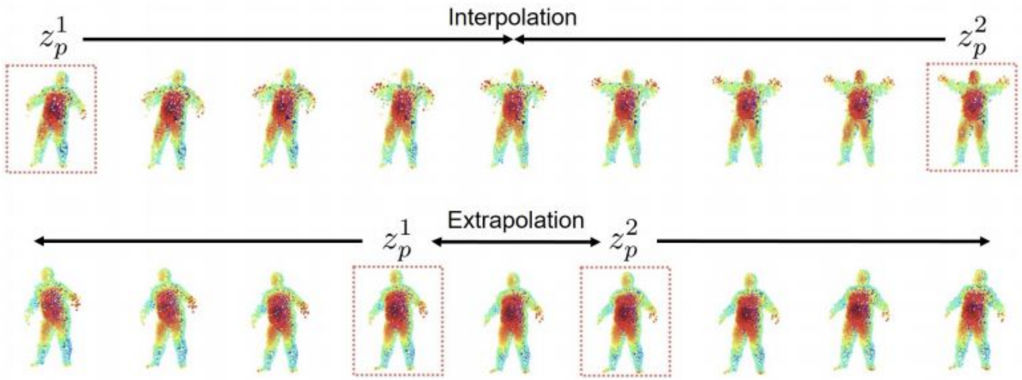






Qualitative Results - 2





Thank you for listening.