

Supplementary Material of "GlocalNet: Class-aware Long-term Human Motion Synthesis"

Performance on Individual Classes

Across all the classes in the dataset NTU RGB+D(3D) the average Euclidean distance across all the classes is 0.17 and the Standard Deviation is 0.06. We provide the results of top-5 and bottom-5 performing classes in Table 1 and 2, respectively.

Although the reconstruction error on bottom-5 classes was high, we observed generated sequences were quite reasonable in terms of qualitative visualization but definitely not coherent with respect to Ground Truth.

Class	Euclidean Distance
Playing with Phone	0.079
Shake Head	0.080
Typing On Keyboard	0.093
Put the Palms together	0.095
Drop	0.106

Table 1: Performance of top-5 classes on NTU RGB+D(3D) in terms on Euclidean Loss.

Class	Euclidean Distance
Wear Jacket	0.345
Throw	0.292
Hugging	0.279
Nausea	0.263
Punching	0.263

Table 2: Performance of bottom-5 classes on NTU RGB+D(3D) in terms on Euclidean Loss.

Performance without Class Prior

Our Method shows better performance than previous SOTA models even without the supervision of class prior. However, as visible in Figure 1 the subspace is significantly cluttered when we don't use the class prior as compared to the embedding diagram from main paper (refer Figure 6).

Models	cross-view		cross-subject	
	$MMD_{avg} \downarrow$	$MMD_{seq} \downarrow$	$MMD_{avg} \downarrow$	$MMD_{seq} \downarrow$
Without Class-prior	0.226	0.231	0.186	0.193
Our Method	0.195	0.197	0.177	0.187

Table 3: Performance without Class label in terms of MMD on NTU RGB+D(2D).

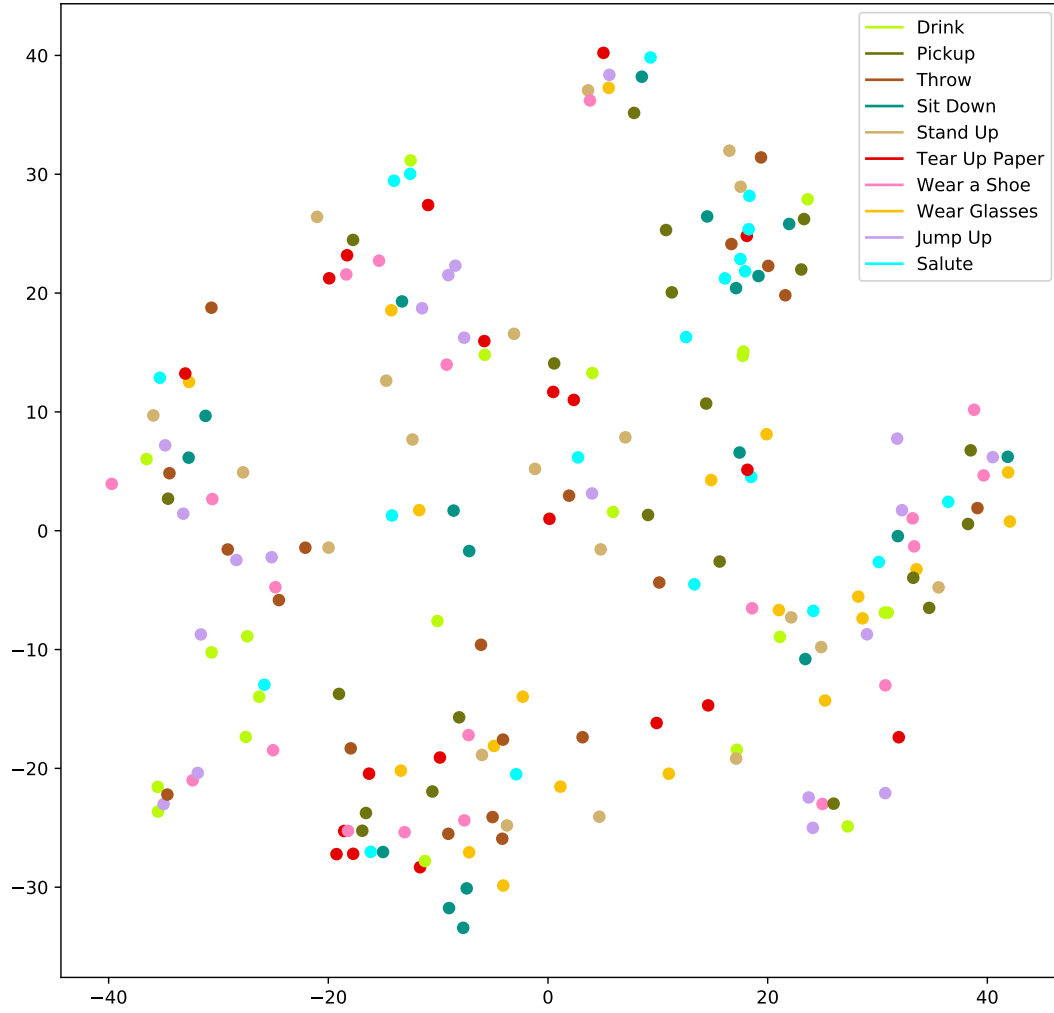


Figure 1: The t-SNE plot of embedding subspace without using the input class label, here samples for different classes are represented as color-coded 3D points.