

Supplemental Material for Accelerating t-SNE using Tree-Based Algorithms

Laurens van der Maaten

LVDMAATEN@GMAIL.COM

*Pattern Recognition and Bioinformatics Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands*

Editor: Aaron Courville

Supplemental Material

This document presents two additional embeddings that were constructed using Barnes-Hut-SNE on (1) the CIFAR-10 data set, (2) the NORB data set, and (3) the street view house numbers data set. The data sets were preprocessed in the same way as described in the main paper “*Accelerating t-SNE using Tree-Based Algorithms*”.

The results are presented separately in Figure 1, 2, and 3. The embedding of the CIFAR-10 data set reveals that the structure of the image data is quite well maintained even in two dimensions: for example, all images with boats are grouped together in the bottom of the map, and a cluster of cars is clearly visible in the left part of the map. These clusters are even somewhat separated from the rest of the images. The embedding of the NORB data set shows that Barnes-Hut-SNE also accurately preserved the local structure of the data: in particular, it reveals many of the rotation manifolds that are present in the data set. The embedding of the street view house numbers data set clearly shows ten clusters that correspond to the ten different digits, whilst images that are difficult to recognize tend to be located near the center of the embedding.

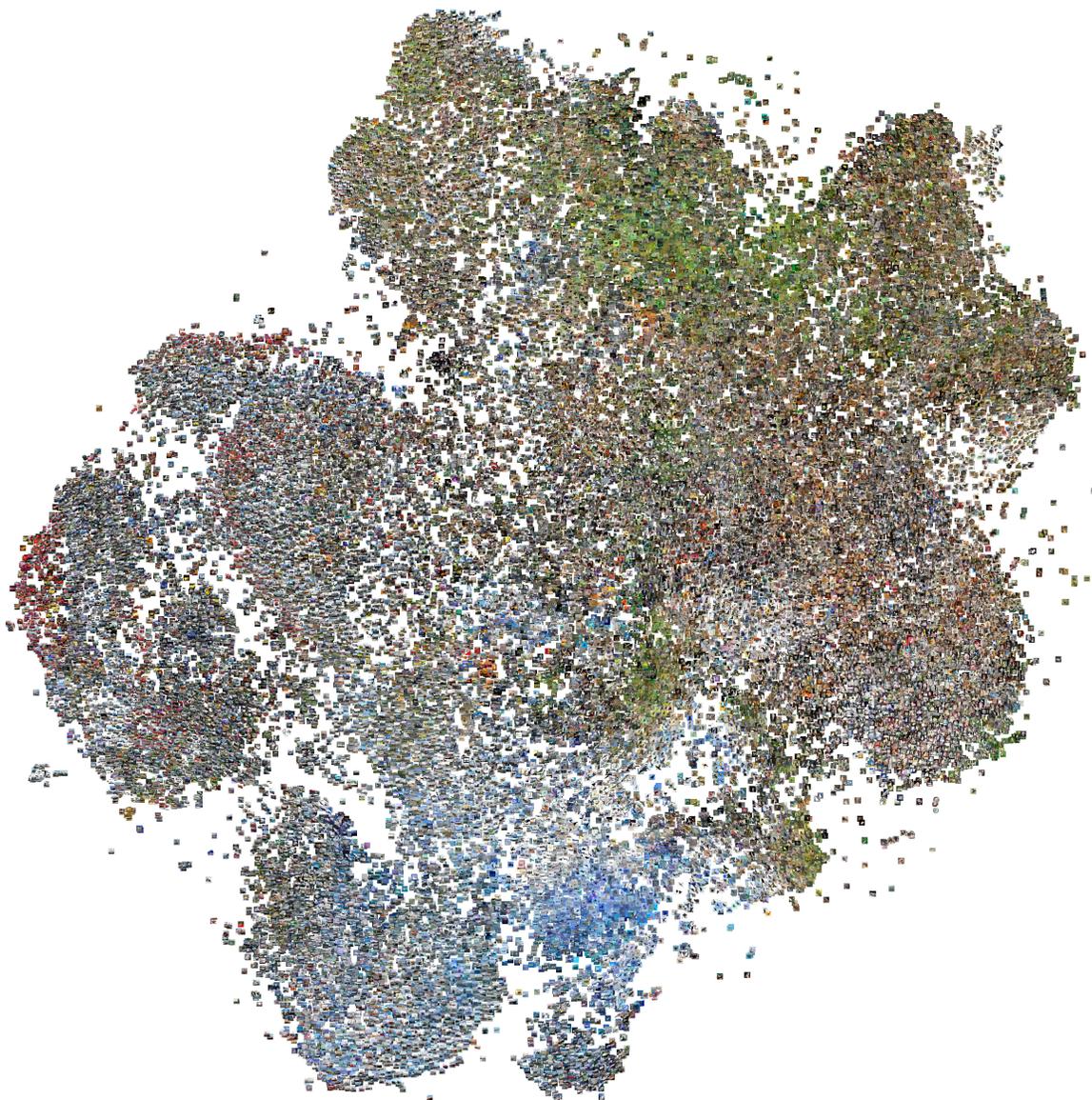


Figure 1: Barnes-Hut-SNE visualization of all 70,000 images in the CIFAR-10 data set.
Zoom in on the image for more detailed views.

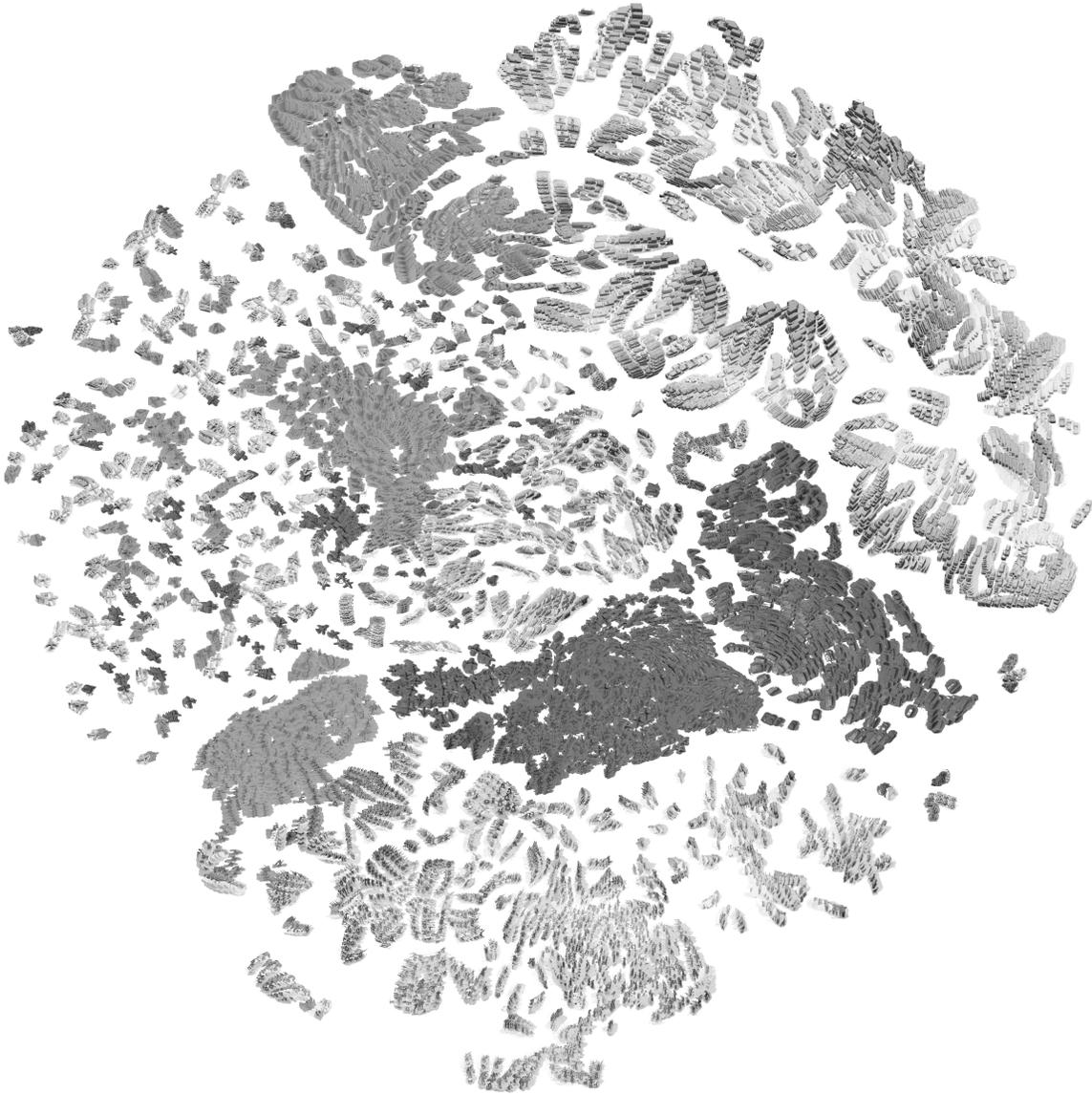


Figure 2: Barnes-Hut-SNE visualization of all 48,600 images in the NORB data set. Zoom in on the image for more detailed views.

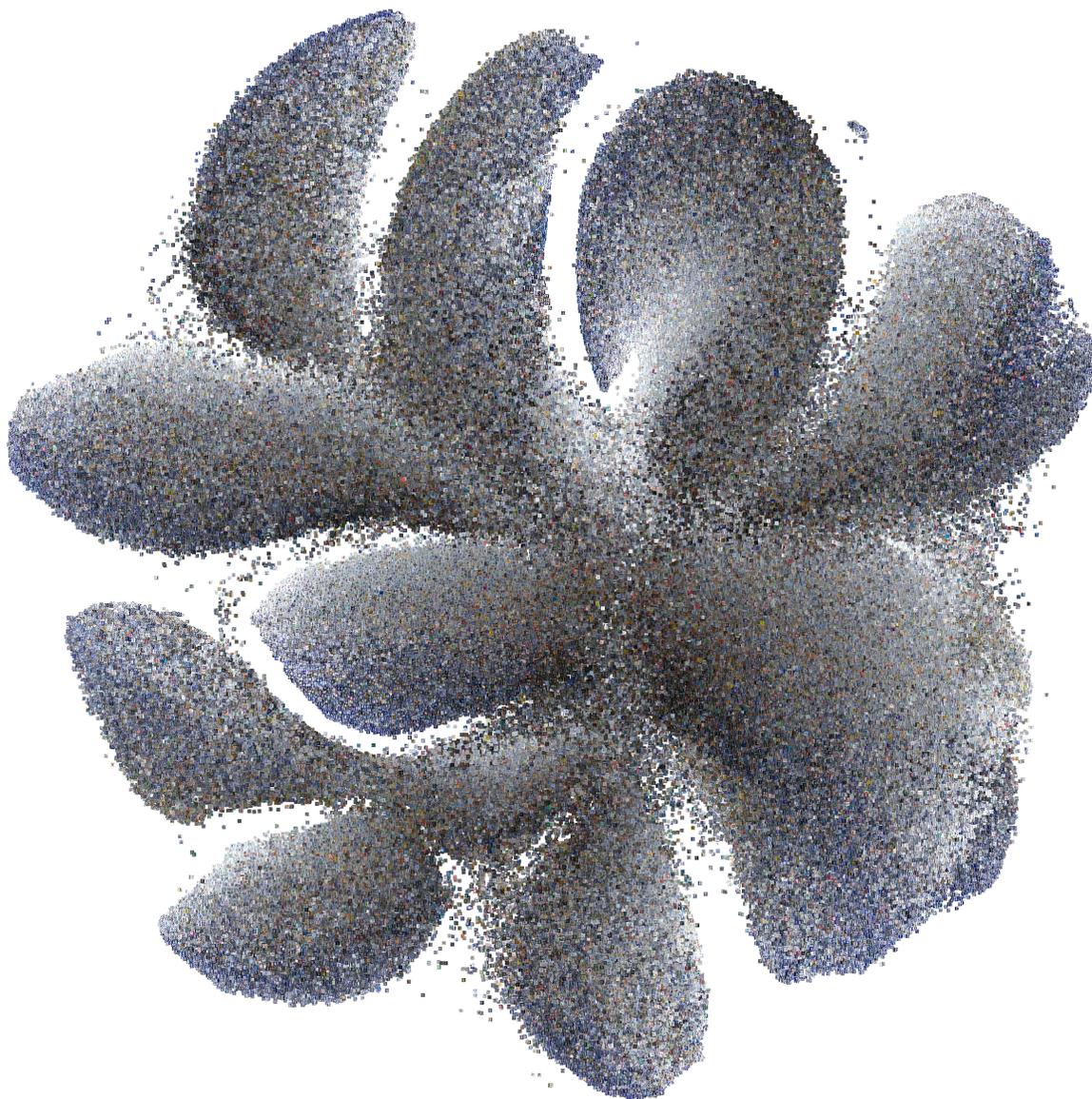


Figure 3: Barnes-Hut-SNE visualization of all 630,420 images in the Street View House Numbers data set. Zoom in on the image for more detailed views.