An Efficient Dual-Hierarchy t-SNE Minimization Supplemental Material

MARK VAN DE RUIT, MARKUS BILLETER, AND ELMAR EISEMANN

1. DUAL TRAVERSAL COMPLEXITY

In our method, the runtime complexity of the dual traversal step approaches $\mathcal{O}(max(F, N)) = \mathcal{O}(N)$. To illustrate this, we use a slightly simpler setting. Imagine two binary trees describing spatial regions; we wish to compute the interactions between these regions.

Without loss of generality, we can assume both trees to have N nodes. If not, the smaller tree can be padded. For $N = 1$, the cost is 1.

For $N > 1$, we can subdivide both trees. Due to symmetry, we then have $2 \cdot 2$ pairs of interactions to compute; a node (representing a region) interacts both with itself and with a neighbor. In our approach, when regions become distant, an interaction is directly computed without requiring further refinement. For illustration, let us assume the interaction between the neighboring regions is sufficiently precise without further refinement. These two pairs can then be computed with a cost of 1 each. The remaining pairs of smaller regions – which interact with themselves – each correspond to a subtree of size $(N-1)/2$.

In total, this results in $2 + 2(cost((N-1)/2)) = 2 + 2((N-1)/2) = N + 1$ interactions, by induction. Hence, the algorithmic complexity is $\mathcal{O}(N)$.

2. BARNES-HUT APPROXIMATION

The parameter θ influences when nodes are culled during dual traversal. A larger θ therefore leads to a faster computation, but a coarser approximation. We show its impact on output embeddings for 2D (Figure S1) and 3D (Figure S2) on two datasets of varying size.

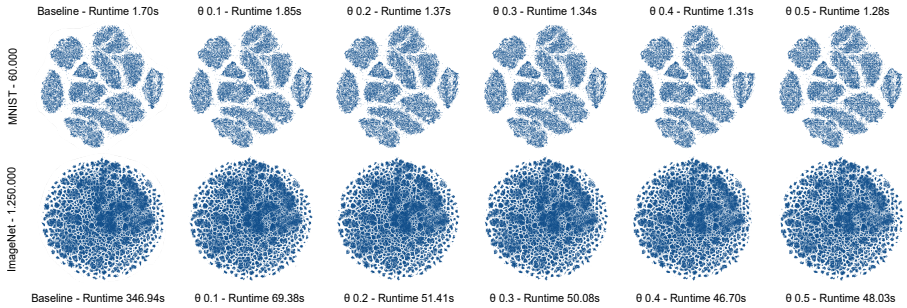

Fig. S1. Influence of θ on generation of 2D embeddings for the MNIST and ImageNet datasets. Errors in the form of a discrete grid becomes visible for $\theta \geq 0.3$ in both datasets.

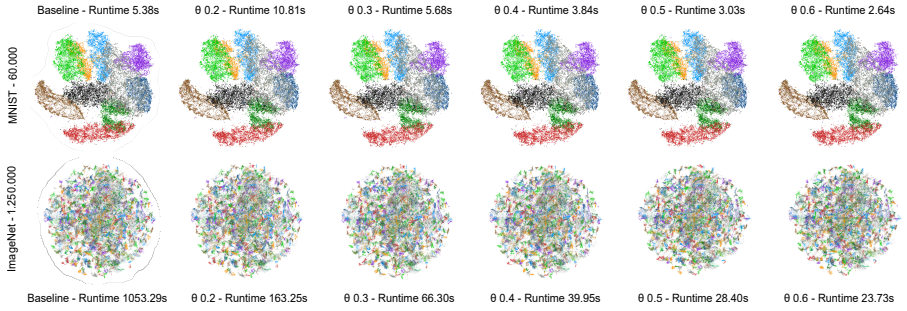

Fig. S2. Influence of θ on generation of 3D embeddings for the MNIST and ImageNet datasets. Errors in the form of a discrete grid becomes visible for $\theta \geq 0.5$ in both datasets. Embeddings for $\theta < 0.2$ are not shown as 3D minimizations may exceed the memory capacity of our GPU.