

OT-based Models

- **Balanced (Sinkhorn) OT.** Computes the Sinkhorn divergence, a de-biased proxy for the ε -entropy-regularised OT cost on token clouds. The pipeline is: log-scaled OT scalar \rightarrow LayerNorm \rightarrow classifier (no handcrafted features). The variant with features is schematically described by the figure on the right.
- **Unbalanced (Sinkhorn) OT.** Adds KL penalties, which allows for the creation or destruction of mass. The OT scalar gets fused by a learnable gate with style features and content projection.
- **Sliced OT.** Approximates the 1-Wasserstein distance by averaging 1-D projections across n_π random directions. This variant does not rely on features.
- **Maximum-projection OT.** Same sampling as sliced OT but keeps the maximum directional distance, *i.e.*, the most “discriminative” slice. As a subvariant of the scaled approach, this method does not rely on features.
- **Multi-scale OT.** Concatenates sliced OT scalars at various resolutions ($n_\pi = \{8, 32, 128\}$), and then mixes them with a content-style gate. As a subvariant of the scaled approach, this method does not rely on features.
- **Contrastive Learning + OT.** Combines a contrastive loss on style embeddings with the balanced Sinkhorn scalar as an extra signal. This implementation relies on the balanced OT approach with features.
- **OT Ensembles.** Majority-vote mixes several OT variants to combine complementary signals. Note that weighted voting is possible. The ensembles contain both feature-informed and feature-free variants.