Projects

Milena Gazdieva

15 February 2023

NB! All projects are computationally intensive and may take more than 7 days for implementation. Make sure that you have at least 1 GPU for group member.

1 Implementing unpaired learning with Incomplete Optimal Transport in a Latent Space of DALLE-2.

Note: Two teams per project are allowed. One team should consider point 2a, another – 2b.

Relevant papers: https://arxiv.org/abs/2301.12874, https://arxiv.org/pdf/2106.03812.pdf

Useful links: https://github.com/iamalexkorotin/NeuralOptimalTransport, dalle2-laion

DALLE-2 is a well-known model for generating images from text prompts. It consists of two stages: a prior that generates a CLIP image embedding given a text caption, and a decoder that generates an image conditioned on the image embedding. Here (CLIP) is a model consisting of text and image encoders and trained on text-image pairs to achieve the maximum similarity between embeddings of the paired text and image. DALLE-2 uses CLIP text embeddings, produces an image embedding using autoregressive or diffusion prior, and then uses this embedding to condition a diffusion decoder which produces a final image.

In https://arxiv.org/pdf/2106.03812.pdf, the authors follow the pipeline of DALLE-2 but propose to replace its prior with Optimal Transport (OT) mapping between spaces of text and images embeddings. CLIP text encoder and DALLE-2 image decoder are frozen and used without modification. Interestingly, training of OT maps does not require paired data. The authors show that their learned OT maps can generate image embeddings with comparable quality even without paired data for training.

In this project, you are supposed to understand the described experimental setup, implement the Incomplete Optimal Transport (IT) approach by Gazdieva et al. [2023] for mapping between text and image embeddings and test the achieved quality. This method allows to learn theoretically best possible unpaired translation between a pair of domains w.r.t. the given similarity function, e.g, cosine similarity between embeddings.

Specifically, you need to:

1. Study the related literature on optimal transport, CLIP, DALLE-2 papers;

2. Fetch (a) Laion art dataset or (b) Conceptual Captions 3M datasets, split it on trainA, trainB and test parts;

3. Generate unpaired text (trainA) and image (trainB) embeddings using pretrained CLIP text and image encoders.

4. Learn an IT map between text and images embeddings spaces using negative cosine

similarity as a cost function. Experiment with weight parameters $w = 1, 2, 4, 8$. For each of the weight parameter, save the final checkpoints (weights θ of the map T).

5. Calculate mean negative test cosine similarity between text embeddings and corresponding generated image embeddings. Show that it decreases with the increase of parameter w.

6. Select 4 random text prompts, generate corresponding text embedding using CLIP pretrained encoder, generate corresponding image embeddings using pretrained IT maps for each of the weights $w = 1, 2, 4, 8$, retrieve corresponding images using pretrained DALLE-2 decoder. Visualize the results. Compare images with images retrieved from groun-truth embeddings and a baseline method. As a baseline, pass an image embedding generated by the diffusion prior in DALLE2-Laion.

7. (Optional) Calculate mean negative test cosine similarity between text embeddings and corresponding generated image embeddings. Perform comparison of the values with the results of the baseline (pretrained DALLE-2 on Laion prior).

8. Analyze the results. Do IT method helps to improve the results of text to image generation?

2 Implementing unpaired learning with Optimal Transport on 3D MNIST images.

Note: Two teams per project are allowed. One team should consider point 2a, another – 2b.

Relevant papers: https://openreview.net/pdf?id=d8CBRlWNkqH

Useful links: 3D MNIST dataset

Computational Optimal Transport is a set of tools, which provide a way of transforming one distribution into another with the minimal effort. Recently scalable neural OT based methods, which do not require paired training dataset, have been developed and applied to a wide range of tasks, including transfer learning Korotin et al. [2022], generative modeling Rout et al. [2021] and super-resolution Gazdieva et al. [2022].

In this project, you need to test the applicability of Neural Optimal Transport method to 3D image-to-image translation;

Specifically, you need to:

1. Study the related literature on optimal transport;

2. Fetch 3D MNIST dataset, split train images on trainA ((a) digits '2', (b) digits '3'), trainB ((a) digits '4', (b) digits '5'). Perform testing on corresponding digits ((a) '2', (b) '3') from test part;

3. Write the code to make the digits colored. Make trainA images and trainB images randomly colored ('yellow', 'red', 'blue', 'green');

4. Change the code of Neural Optimal Transport, e.g., use architectures of the transport map $T_\theta : \mathbb{R}^{3\times16\times16\times16} \to \mathbb{R}^{3\times16\times16\times16}$, potential $f_w : \mathbb{R}^{3\times16\times16\times16} \to \mathbb{R}$, to make it applicable to colored 3D images;

5. Learn an OT mapping between trainA and trainB samples using strong quadratic cost function. Save the final checkpoints (weights θ of the map T);

6. Select 5 random digits from trainA, show the corresponding OT mapped digits. Visualize the results.

7. Analyze the results. Do the output images resemble the shape of the input images shape? Does the map preserve the color of the digit during the translation? Does the quadratic cost turn to be a good choice for this pair of datasets? Yes/No? Why?

3 The Monge Gap: A Regularizer to Learn All Transport Maps.

Note: Two teams per project are allowed. One team should consider point (a), another – (b).

Relevant papers: https://arxiv.org/pdf/2302.04953.pdf

Computational Optimal Transport is a set of tools, which provide a way of transforming one distribution into another with the minimal effort. Recently scalable neural OT based methods, which do not require paired training dataset, have been developed and applied to a wide range of tasks. Recent approaches Korotin et al. [2022], Rout et al. [2021], Gazdieva et al. [2022] introduce a Lagrange multiplier f in the Monge OT formulation and then solve the saddle-point optimization problem, trading off two terms, a displacement cost and a fitting loss error, in order to learn an OT map between distributions. A recent paper introduces a Monge gap regularization which results in a minimization optimization problem. The authors claim, that the proposed formulation does not exactly conform to OT theory, but allows to efficiently train maps that should be OT-like.

In this project, you are supposed to understand the paper and replicate the proposed method and its results: (a) on Wasserstein-2 benchmark by Korotin et al. [2021] or (b) for single cell genomics experiment.

Specifically, you need to:

1. Study the paper and experimental setup;

2. Fetch (a) Wasserstein-2 bechmark by Korotin et al. [2021] in dimension $d \in \{2, 4, 8, ..., 256\}$ or (b) proteomic dataset used in Bunne et al. [2021] (you may pick only one of the 4i, scRNA datasets).

3. Write the training code according to stated in the paper experimental setup for the experiments with the quadratic cost, see §6.1, §6.3, §6.4. Train the model and save the final checkpoints (weights θ of the map T);

4. (a) Calculate *unexplained variance percentage* $\mathcal{L}_2^{UV}(\hat{T})$ metrics measuring the deviation of the learned OT map $\hat{T}$ and ground-truth OT map T^* which is known due to the Wasserstein-2 benchmark construction. Visualize the results as in Figure 6 of the paper, using only $\hat{T}_0 = \mathbb{E}[\nu]$ as an upper bound baseline. Additionally, you need to calculate Sinkhorn divergence $S_{\ell_2^2,\varepsilon}(\nu, \hat{T}_\sharp\mu)$ and visualize the results as in Figure 6. The goal is to achieve the results close to those in the paper (<5% difference). However, you may play with the regularization parameters if it helps.

4. (b) Train vanilla MLP on the dataset. Calculate the Sinkhorn divergence between unseen treated cells and control cells mapped by $\hat{T}$. Visualize the results as in Figure 3 (for vanilla MLP baseline). The goal is to achieve the results close to those in the paper (<5% difference). However, you may play with the regularization parameters if it helps.

5. Accurately describe the method, experimental setup, present the algorithm following the form in Korotin et al. [2022]. Analyze and present the results. Did you achieve the results presented in paper? Does the method outperforms chosen baselines?

References

Charlotte Bunne, Stefan G Stark, Gabriele Gut, Jacobo Sarabia del Castillo, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *bioRxiv*, pages 2021–12, 2021.

Milena Gazdieva, Litu Rout, Alexander Korotin, Alexander Filippov, and Evgeny Burnaev. An optimal transport perspective on unpaired image super-resolution. *arXiv preprint arXiv:2202.01116*, 2022.

Milena Gazdieva, Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Extremal domain translation with neural optimal transport. *arXiv preprint arXiv:2301.12874*, 2023.

Alexander Korotin, Lingxiao Li, Aude Genevay, Justin M Solomon, Alexander Filippov, and Evgeny Burnaev. Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark. *Advances in Neural Information Processing Systems*, 34: 14593–14605, 2021.

Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Neural optimal transport. *arXiv preprint arXiv:2201.12220*, 2022.

Litu Rout, Alexander Korotin, and Evgeny Burnaev. Generative modeling with optimal transport maps. *arXiv preprint arXiv:2110.02999*, 2021.