Research Statement of **Jurijs Nazarovs**

jurijsnazarovs.github.io

Summary. My main research interests focus on designing **temporal probabilistic deep learning models**, which describe uncertainty in trajectory prediction and generation, motivated by applications in computer vision as well as other multi-modal tasks. My portfolio consists of more than 10 papers and patents, with some of the work published in top-tier conferences like CVPR and UAI. During my PhD, I interned in several research labs, such as Amazon Alexa AI, Microsoft Research (MSR), and NEC Labs America, where each project led to a first author paper. Next, I will outline my ongoing/future research interests and discuss how my past research experience can help in moving forward.

1 Bayesian Neural Networks (BNN)

BNNs are a prominent tool to model uncertainty on the prediction and mitigate overfitting (with small datasets). Similar to Bayesian statistics, BNNs provide the freedom to select an appropriate prior and approximate posterior on weights, which fit our data/goal the best. Unfortunately, not all choices of distributions lead to convenient computational implementation (say, in a closed form), and often training BNNs relies on sampling techniques (e.g., Monte Carlo or MC). In one result (UAI), we showed that when sampling techniques is used (when closed form solutions are unavailable), the size of the computation graph is proportional to the number of sampling iterations, which leads to a GPU memory explosion and inability to train large BNNs for common architectures. To overcome this issue, we **(1)** characterized (with theoretical support) the set of distributions where the above issue can be avoided. **(2)** Specifically, we proposed a parameterization method, allowing us to preserve the size of computation graph, independent of number of MC samples. This provided the ability to train large BNNs with priors/approximate posteriors, without relying on a closed form solution. In work carried out during my internship in MSR, we focused on classifying a ransomware attack based on a strongly imbalanced (and small) training data with a very sparse set of features. To account for specific structure of the problem, I proposed a new mixture distribution for BNNs, called "Radial Spike and Slab". We showed that in this setting, our model behaved much superior to baselines for ransomware detection.

2 Extensions of Neural ODE (NODE): temporal processes

In the last years, I devoted significant effort exploring the temporal structure within datasets, mainly in computer vision/image analysis tasks, with the goal of modelling the underlying temporal process as a continuous path. The core idea of these works is based on NODE to achieve the desired modelling properties, depending on whether we focus on predicting the next step (e.g., video frame), interpolating steps between observed samples, or generate a completely new trajectory. Let me describe my first result, called **"Mixed Effects NODE"**. While NODE is a flexible model to describe a temporal process in a continuous way (by utilizing DE, $dz/dt = f(z, t)$, where f is a neural network), its main limitation lies in a restriction that given initial point z_0 and transition function f, it results in a deterministic trajectory z_t. However, in reality this is not always the case and randomness of trajectories for the fixed initial point should be accounted for. To overcome this issue, I proposed to incorporate Mixed Effects (Random Projection) to generate the distribution of trajectories, given a fixed initial point. We showed from a theoretical perspective (UAI), that this model is an approximation of a Stochastic NODE, but requires only ODE solvers! In another line of work, we introduced a generative model, **"Warping NODE"**, which models the vector field of changes (warps) between frames via NODE. Compared to a typical GAN, Warping NODE does not introduce new information on images (new values of pixels), but uses a warping of current information which results in a smooth interpolation of non-existing frames. This framework was applied in two different papers: to generate animation (interpolation; CVPR workshop) and to perform a statistical test to understand the significant uncertainty of the computer vision DNN (CVPR). The last work on this topic is currently under submission, where I introduced a **"Functional NODE"** to sample trajectories as points from a functional space, similar to recent research on *functa*. The construction offers a VAE-type model but for temporal trajectories.

3 Multimodality: current/future work.

During my internship in Amazon Alexa AI, I studied the Robustness of Vision Question Answering to linguistic variations and vision manipulation (paper under submission). I am keen on continuing this line of research on multimodal data, with modalities not limited to only images and text, leveraging ideas from the above two threads as needed. I outline some topics of interests in this field: **(1)** explainable Vision Question Answering (VQA), i.e., that we not only provide an answer based on image, but also explain why that is the case [uncertainty], **(2)** video understanding, e.g., which parts of video answers to the question or refer to specific topic/emotion [trajectory], **(3)** trajectory generation based on natural language expressions. In contrast to current work on generating videos from textual prompt, I would like to explore the idea of generating trajectories for agents (e.g. robots), based on prompts [posterior modeling; trajectory], I am also interested in other application-motivated problems involving multimodal data including vision, language or temporal data acquired from sensors.