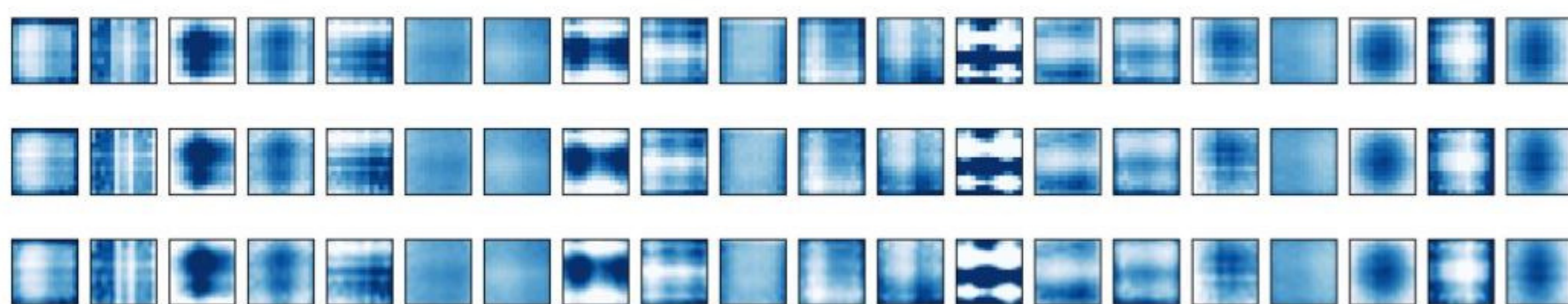


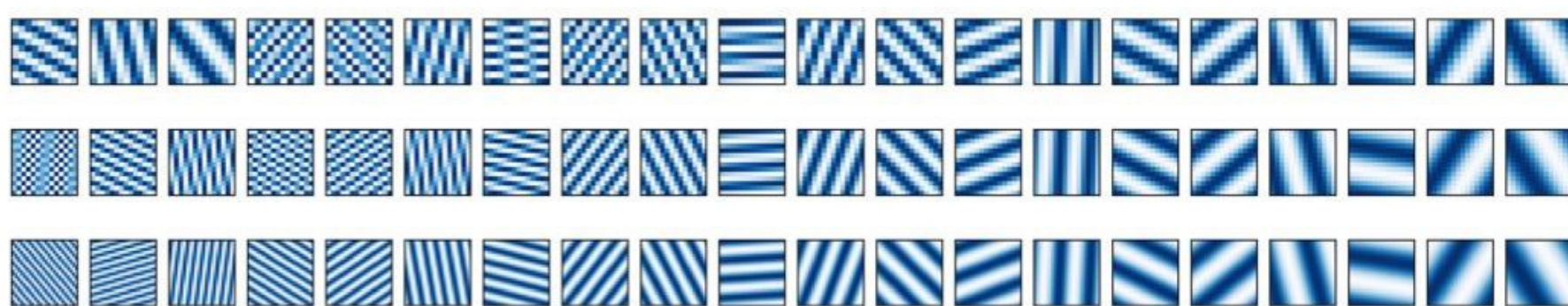
Positional Embedding in Transformer-based Models

Speaker: Tatiana Likhomanenko, Apple

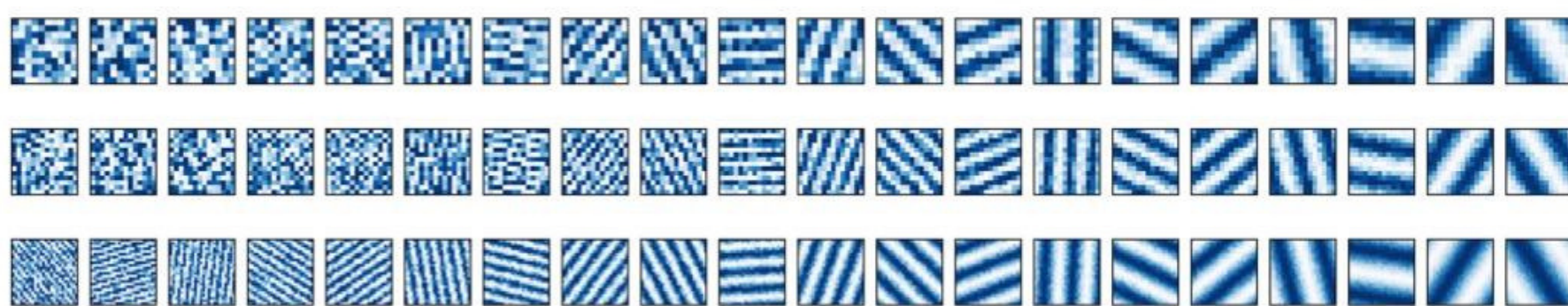
Learnable, 2D



Fixed, 2D



Continuous Augmented, 2D



Abstract

Transformers have been shown to be highly effective on problems involving sequential modeling, such as in machine translation (MT) and natural language processing (NLP). Following its success on these tasks, the Transformer architecture raised immediate interest in other domains: automatic speech recognition (ASR), music generation, object detection, and finally image recognition and video understanding. Two major components of the Transformer are the attention mechanism and the positional encoding. Without the latter, vanilla attention Transformers are invariant with respect to input tokens permutations (making "cat eats fish" and "fish eats cat" identical to the model). In this talk we will discuss different approaches on how to encode positional information, their pros and cons: absolute and relative, fixed and learnable, 1D and multidimensional, additive and multiplicative, continuous and augmented positional embeddings. We will also focus on how well different positional embeddings generalize to unseen positions for both interpolation and extrapolation tasks.



[Registration link](#)