

In this conversation, we cover 17 papers published on the Mamba architecture in the last 90 days. They are:

[Is Mamba Capable of In-Context Learning?](#) - Mamba architecture has similar in-context learning abilities as the transformer. It achieves this by incrementally optimizing internal representations.

[Othello Mamba](#) - model learns rules of Othello based on only seeing a sequence of moves and does so with a higher board accuracy than OthelloGPT. More data efficient as well. Othello Mamba becomes less accurate as the game progresses and requires longer training time than a transformer of the same size.

[Can Mamba Learn How to Learn?](#) - Mamba outperformed transformers on tasks that introduced irrelevance and noise but did not hold up as well as the transformer on a high-precision memory recall task. The hybrid model MambaFormer outperformed the transformer and Mamba on the various evaluation tasks.

[MoE-Mamba](#) - took out a mamba layer and added in a Switch based MoE layer. Scaled up to 32 experts and found that MoE-Mamba achieved the same loss as the original Mamba with 2.2x less training steps.

[BlackMamba](#) - another MoE model, scaled up to 2.8B parameter and 8 experts. Generation latency well below transformer, transformer MoE, and Mamba.

[U-Mamba](#) - hybrid CNN-SSM architecture, that outperforms traditional CNN and Transformer models in biomedical image segmentation. The Mamba layer captures global dependencies more efficiently compared to other models.

[Swin-UMamba](#) - combines Mamba based modeling with ImageNet pretraining for biomedical image segmentation. It outperforms U-Mamba on a variety of datasets.

[Vision Mamba](#) - uses forward scan and backward scan to learn visual representations. Information that flows in both directions gives a better visual representation. Faster than the transformer and saved GPU memory when performing batch inference.

[VMamba](#) - introduces cross-scan which is a four-way scan that starts in each corner and takes a rotating approach to scanning an image. Creates four different representations which are run in parallel and superimpose them into the same space and it works.

[VM-UNet](#) - applies four-way scan of Vision Mamba paper for medical image segmentation.

[Mamba-ND](#) - multi-dimensional sequencing (considered video which requires adding temporal image, and looked at weather data) and put one selective state space mechanism after another with each handling a different version of the scan as input

[SegMamba](#) - 3D image segmentation while handling long sequences, it uses a three-way scan.

[Vivim](#) - scan three ways through video. Generalizing to video seems to work pretty well.

[MambaMorph](#) - brings two different inputs together to make a deformation field as an output which describes how the one image needs to be deformed to fit with the other. Deformation field is applied to one of the inputs so that the two inputs can be understood together.

[Graph-Mamba](#) - do state space models such as mamba enhance long range context modeling in a more efficient way? Graph-Mamba achieves a 74% reduction in memory usage and a 66% reduction in FLOPs compared to Transformer.

[LongMamba](#) - generalized past the training window of 16,000 tokens. Loss reduces at longer context up to around 40,000 tokens. Achieves nearly perfect "Needle in a haystack" results for 16k context length.

[Evo](#) - used StripedHyena model to model long DNA sequences, with the very interesting result that the model appears to be beginning to learn a "cell model", similarly to how language models are learning a sort of "world model"

Information for charts

	Paper	Country	Region	ML category
1	Is Mamba Capable of In-Context Learning?	Italy, Germany	Europe	NLP
2	Othello Mamba	France	Europe	Games
3	Can Mamba Learn How to Learn?	South Korea, USA	Asia, North America	NLP
4	MoE-Mamba	Poland	Europe	NLP
5	BlackMamba	USA	North America	NLP
6	U-Mamba	Canada	North America	Vision, image segmentation
7	Swin-UMamba	China	Asia	Vision, image segmentation
8	Vision Mamba	China	Asia	Vision
9	VMamba	China	Asia	Vision
10	VM-UNet	China	Asia	Vision, image segmentation
11	Mamba-ND	USA	North America	Weather forecasting
12	SegMamba	China	Asia	Vision, image segmentation

13	Vivim	China	Asia	Vision, image segmentation
14	MambaMorph	China	Asia	Vision
15	Graph-Mamba	Canada	North America	Graph NN
16	LongMamba	Singapore	Asia	NLP
17	Evo	USA	North America	Genomics