时空预测学习 (Spatiotemporal predictive learning) 虽然被认为是一种有效的自监督特征学习策略，但很少能显示出其超越未来视频预测的有效性。原因是对于短期框架依赖性和长期高级关系都很难学习到良好的表示形式。

这篇文章提出了一种新模型 Eidetic 3D LSTM（E3D-LSTM），它将 3D 卷积集成到 RNN 中。封装的 3D 卷积使 RNN 的局部感知器能够感知运动，并使存储单元可以存储更好的短期特征。对于长期关系，这篇文章提出通过门控制的自注意力模块使当前的 memory 状态与历史 memory 进行交互，这种机制使网络可以更好地捕捉长期依赖，因为它能够跨多个时间戳有效地调用存储的 memory。

实验方面，作者先在广泛使用的未来视频预测数据集上评估 E3D-LSTM 网络，测试结果达到了 SOTA。然后，作者证明了 E3D-LSTM 网络在早期活动识别方面也表现良好，可以在仅观察有限的视频帧后推断出正在发生或将要发生的情况。该任务与视频预测在建模动作意图和趋势方面非常吻合。