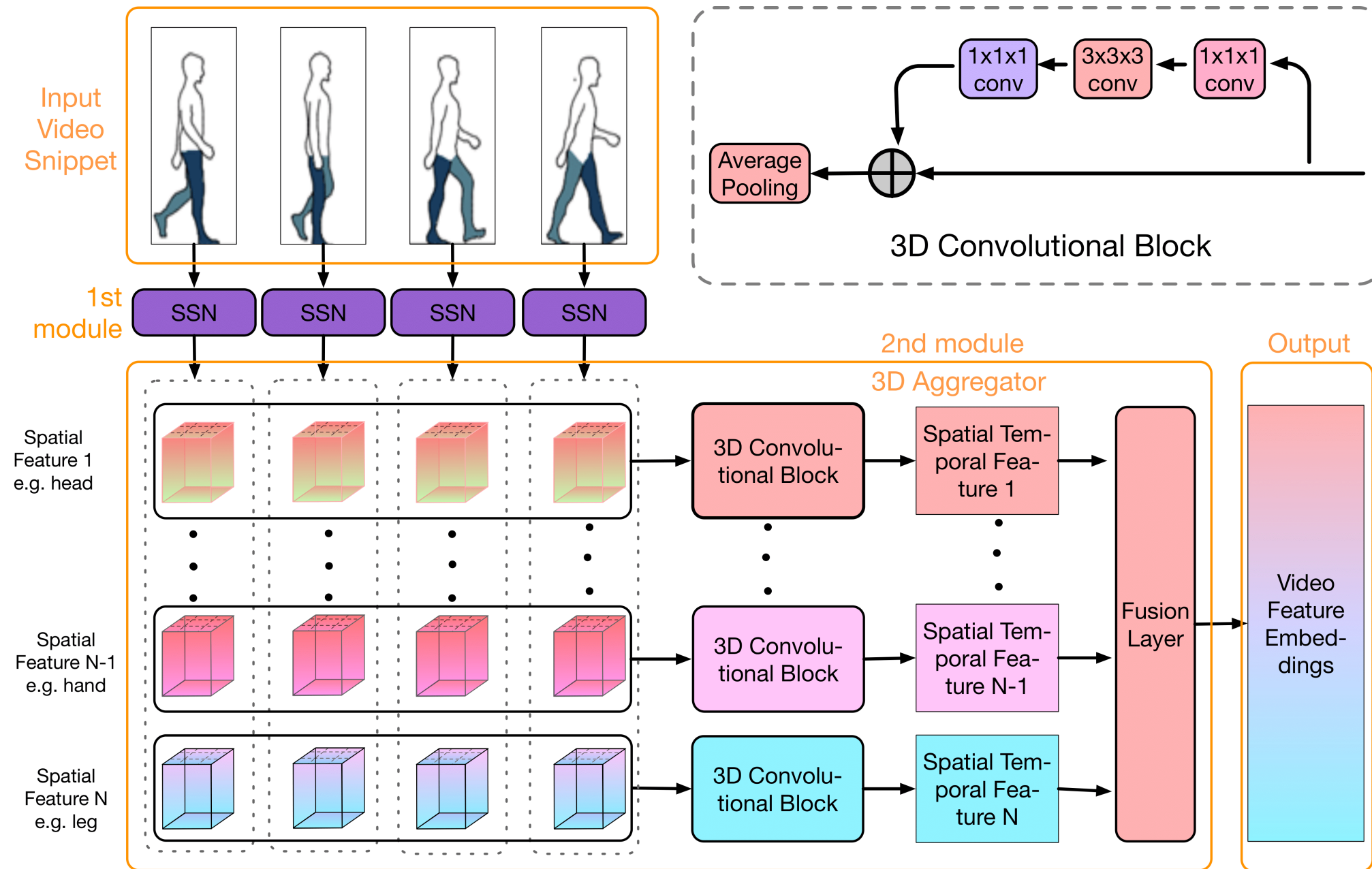


Temporal appearance misalignment is a crucial problem in video person re-identification. The same part of person (e.g. head or hand) appearing on different locations in video sequence weakens its discriminative ability, especially when we apply standard temporal aggregation such as 3D convolution or LSTM. To address this issue, we propose Self-Separated network (SSN) to seek out the same parts in different images. As the name implies, SSN, if trained in an unsupervised strategy, guarantees the selected parts distinct. With a few samples of labeled parts to guide SSN training, this semi-supervised trained SSN seeks out the parts that are human-understandable within a frame and stable across a video snippet. Given the distinct and stable person parts, we then apply 3D convolution across different frames for person re-identification.



Overall Framework: Two Modules, **SSN** aligns person parts per frame and **3D Aggregator** extracts information from multiple frames

