Survey of semantic segmentation
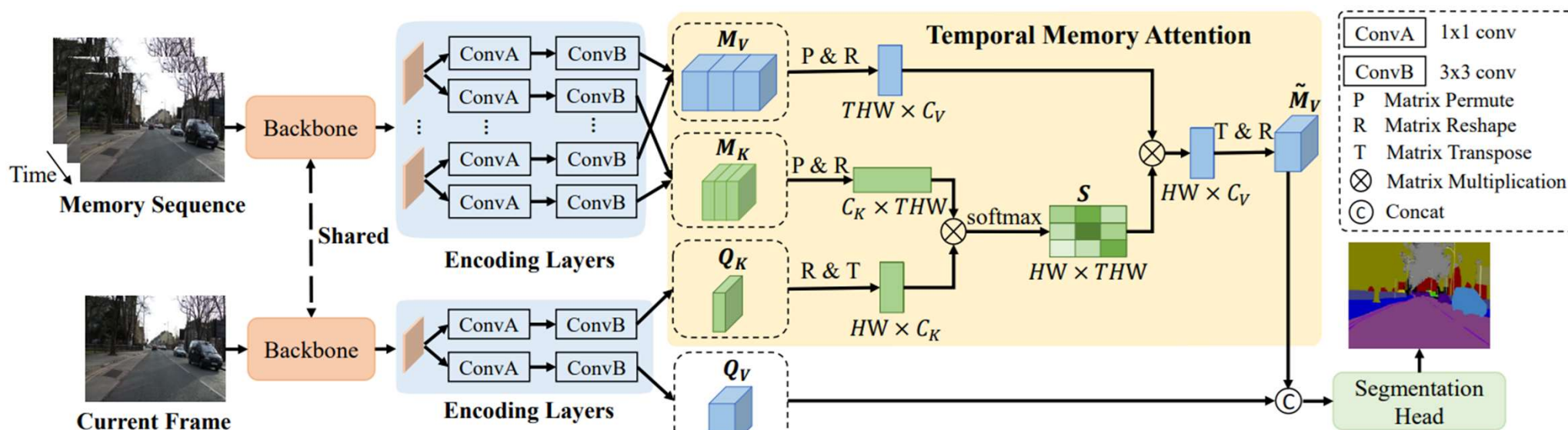
Category and Benchmark *on Cityscapes*

Temporal Memory Attention for Video Semantic Segmentation

Main work:
- multiple frames input (time window)
- attention (correlation between *i* and *i-1~i-N*)

Backbone: ResNet-50

Tags: Video segmentation, multiple frame correlation, attention mechanism

Temporally Distributed Networks for Fast Video Semantic Segmentation

Main work:
- Employ light encoder for different frames (TDs)
- Light weight time domain attention model (APM)

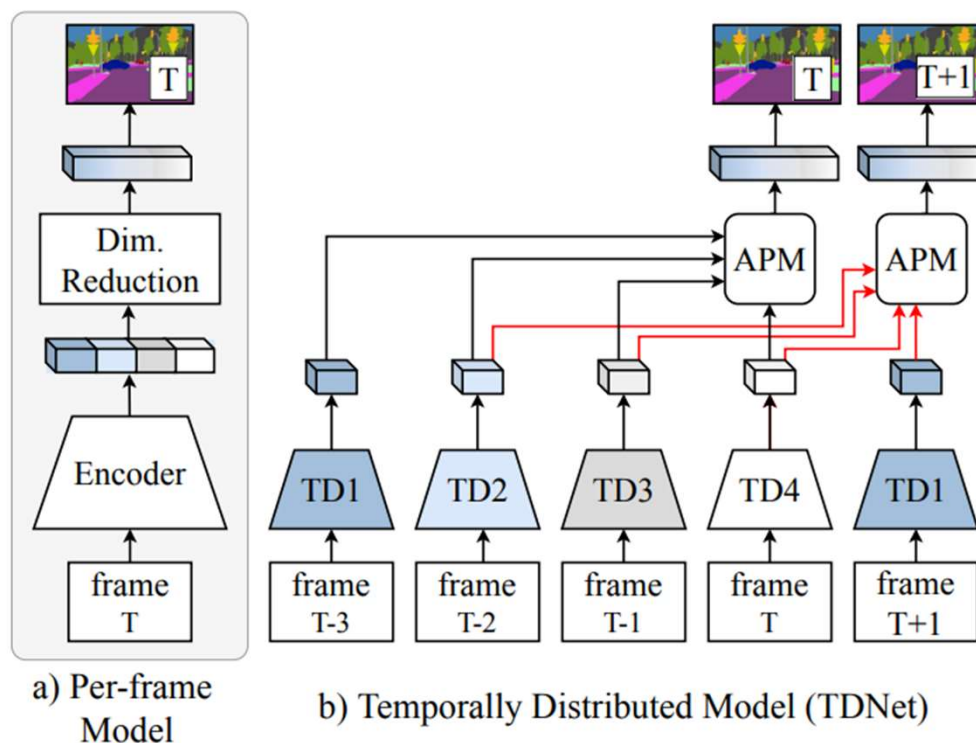Concept: the high level feature change slightly along with time, so only update partial time information.

Final feature representation:

Sum of weighted feature

$$V_t' = V_t + \sum_{p=t-m+1}^{t-1} \phi(\mathbf{Aff}_p V_p)$$

Weight is correlation between frames

$$\mathbf{Aff}_p = Softmax(\frac{Q_t K_p^\top}{\sqrt{d_k}})$$

a) Per-frame Model

b) Temporally Distributed Model (TDNet)

Tags: Video segmentation, differed encoder, fast calculation