Poster [G3]

CVPR 2018
Salt Lake City

Deep Spatio-Temporal Random Fields for Efficient Video Segmentation

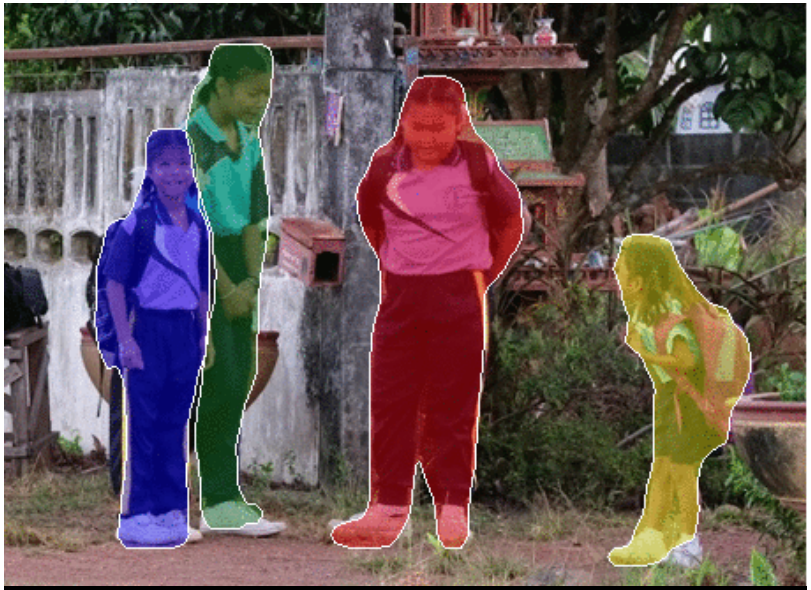
Siddhartha Chandra, Camille Couprie & Iasonas Kokkinos

CentraleSupélec

Inria
inventors for the digital world

Facebook AI Research

Video G-CRFs: Motivation

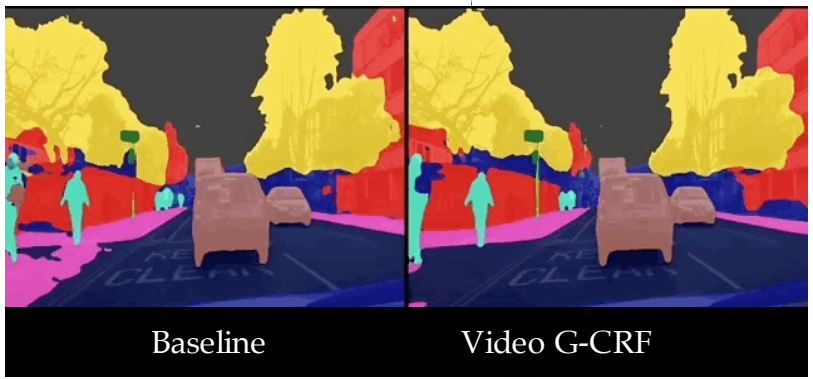


Frame-by-Frame Prediction

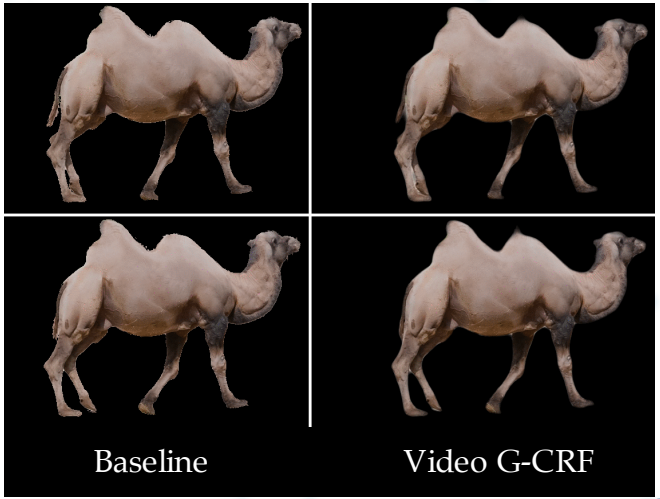
Video G-CRFs: Motivation

Video G-CRFs: Applications

Video Instance Segmentation

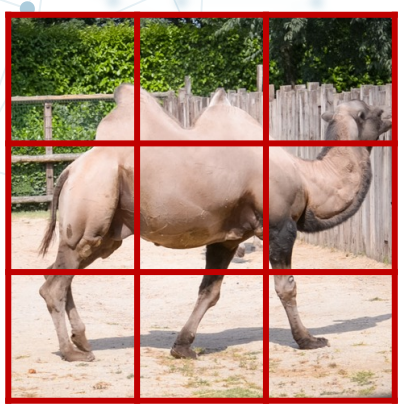


Video Semantic Segmentation

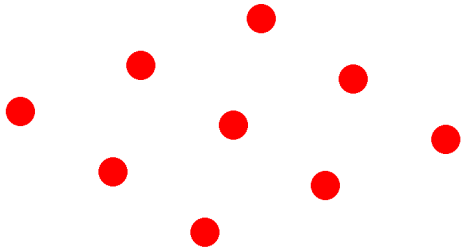


Instance Tracking

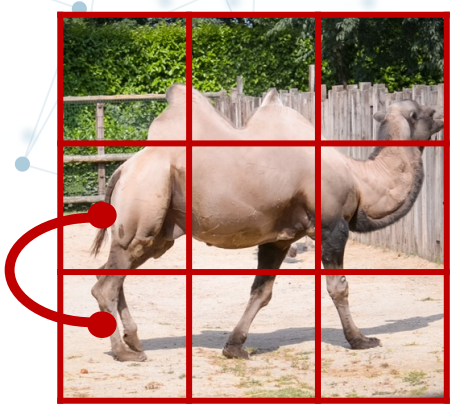
Deep Spatio-Temporal G-CRF



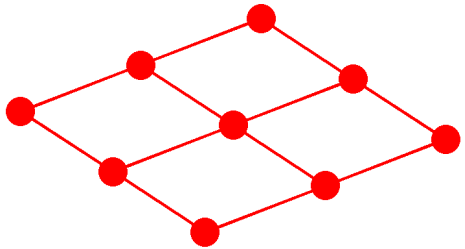
Decoupled Predictions



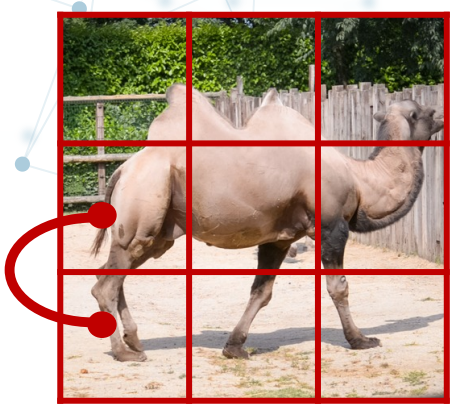
Deep Spatio-Temporal G-CRF



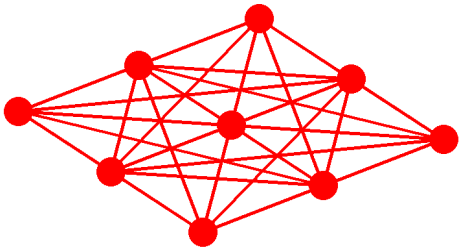
Sparse Graphical Model



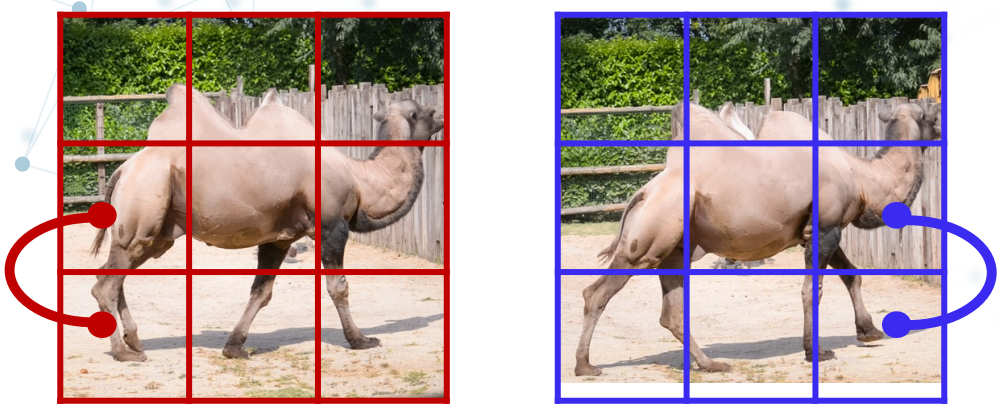
Deep Spatio-Temporal G-CRF



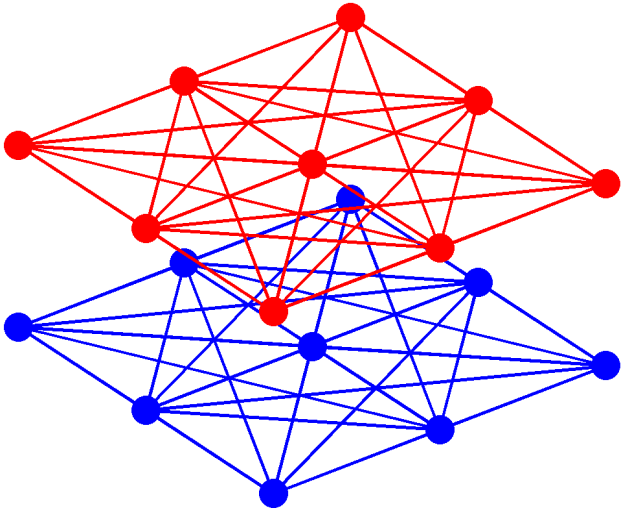
Dense Graphical Model



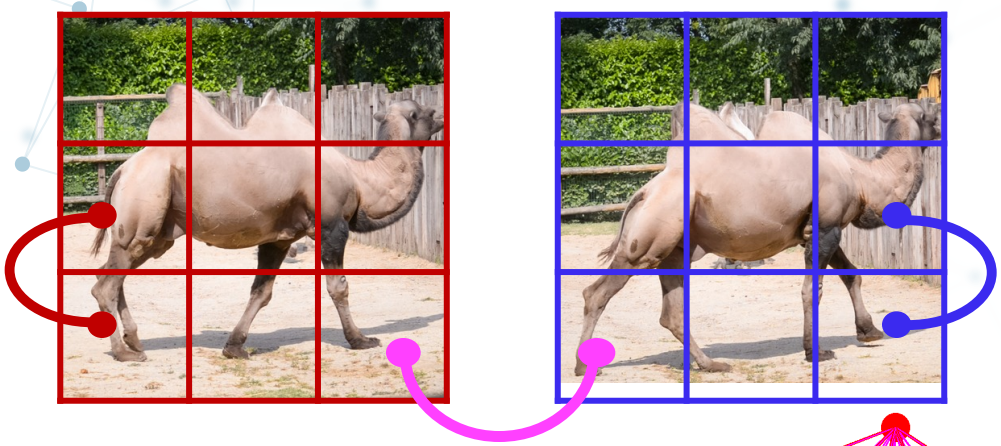
Deep Spatio-Temporal G-CRF



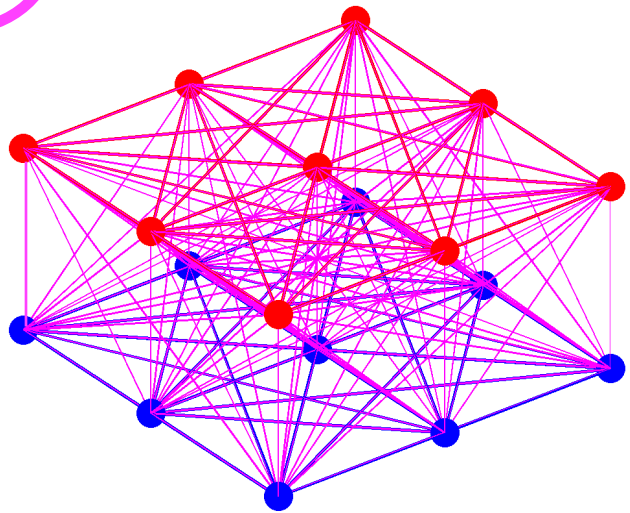
Dense Graphical Model
(No coupling of predictions)



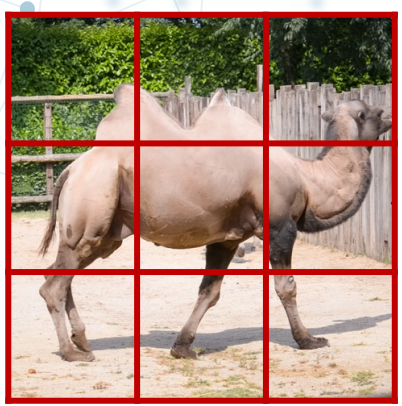
Deep Spatio-Temporal G-CRF

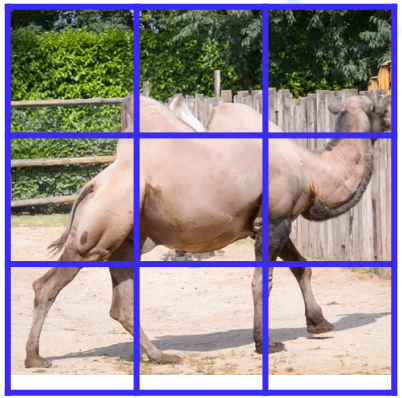


Dense Spatio-Temporal
Graphical Model

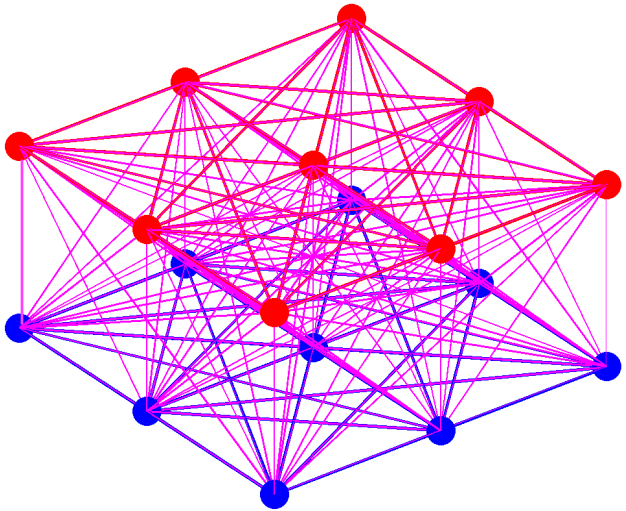


Deep Spatio-Temporal G-CRF





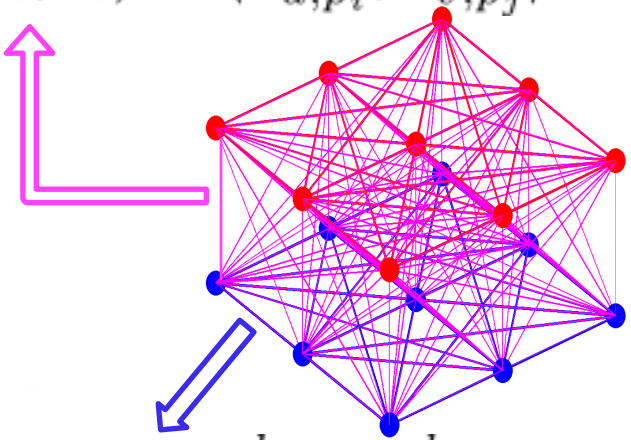
Dense Spatio-Temporal Graphical Model



Deep Spatio-Temporal G-CRF

Gaussian-CRFs: $E(\mathbf{x}) = \underbrace{\frac{1}{2}\mathbf{x}^T A\mathbf{x}}_{\text{pairwise}} - \underbrace{B^T\mathbf{x}}_{\text{unary}}$

$$T_{u,v,p_i,p_j}\,(l_m, l_n) = \langle \mathcal{T}^{l_m}_{u,p_i}, \mathcal{T}^{l_n}_{v,p_j} \rangle$$

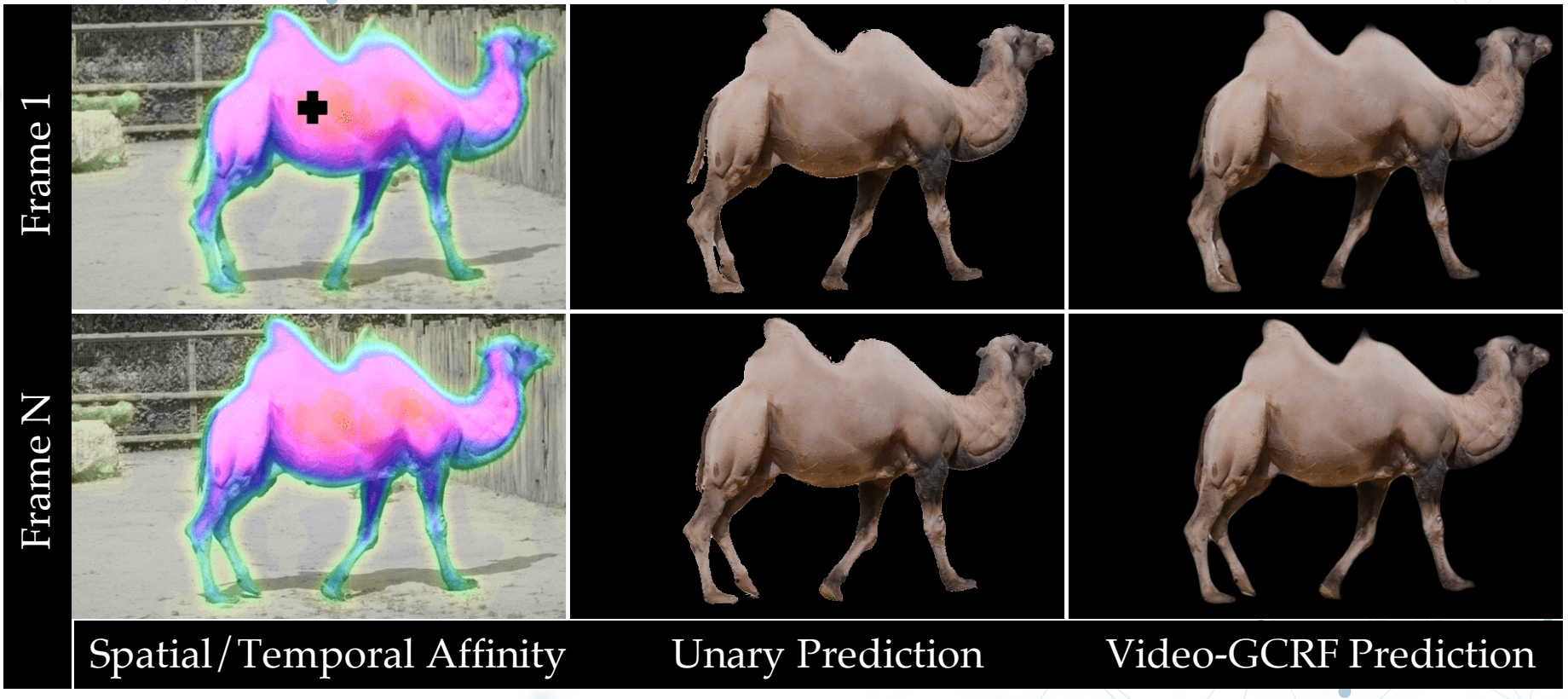


$$E(\mathbf{x}) = \underbrace{-\sum_i b_i \mathbf{x}_i}_{\text{unary}} + \underbrace{\frac{1}{2}\sum_{i,j} \mathbf{x}_i \mathbf{x}_j a_{i,j}}_{\text{pairwise}}$$

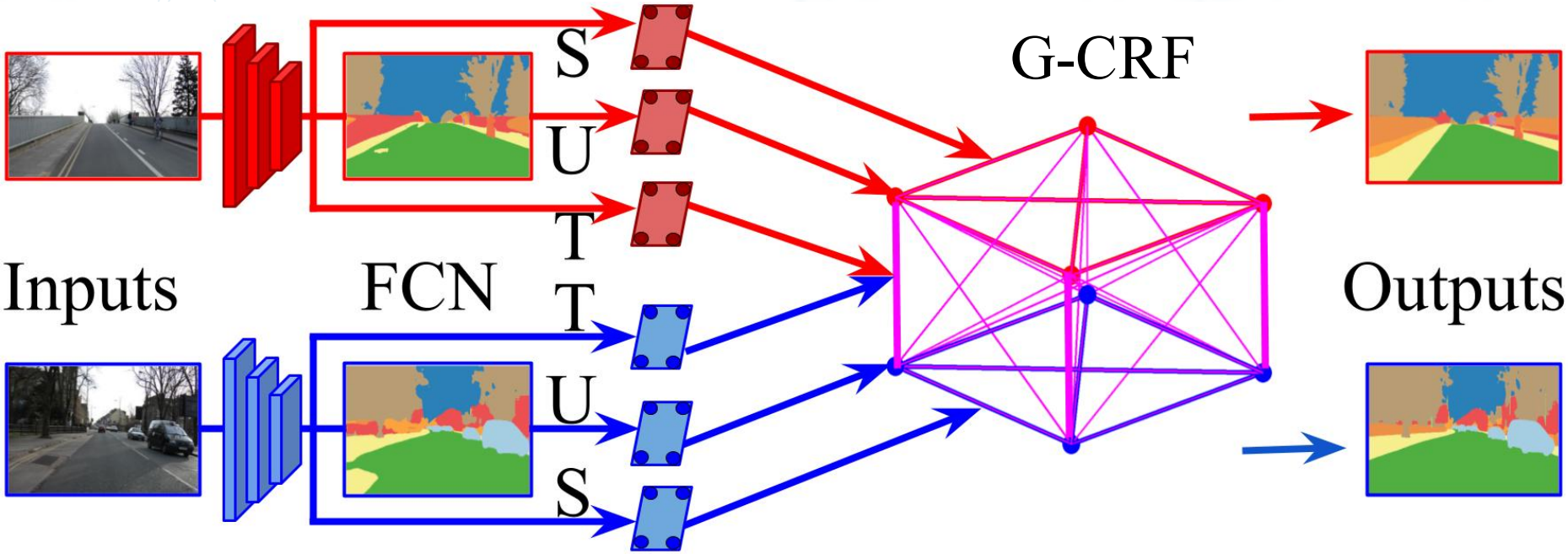
$$A_{v,p_i,p_j}\,(l_m, l_n) = \langle \mathcal{A}^{l_m}_{v,p_i}, \mathcal{A}^{l_n}_{v,p_j} \rangle$$

Inference: $A\mathbf{x} = B$

Pairwise Affinities: Instance Tracking

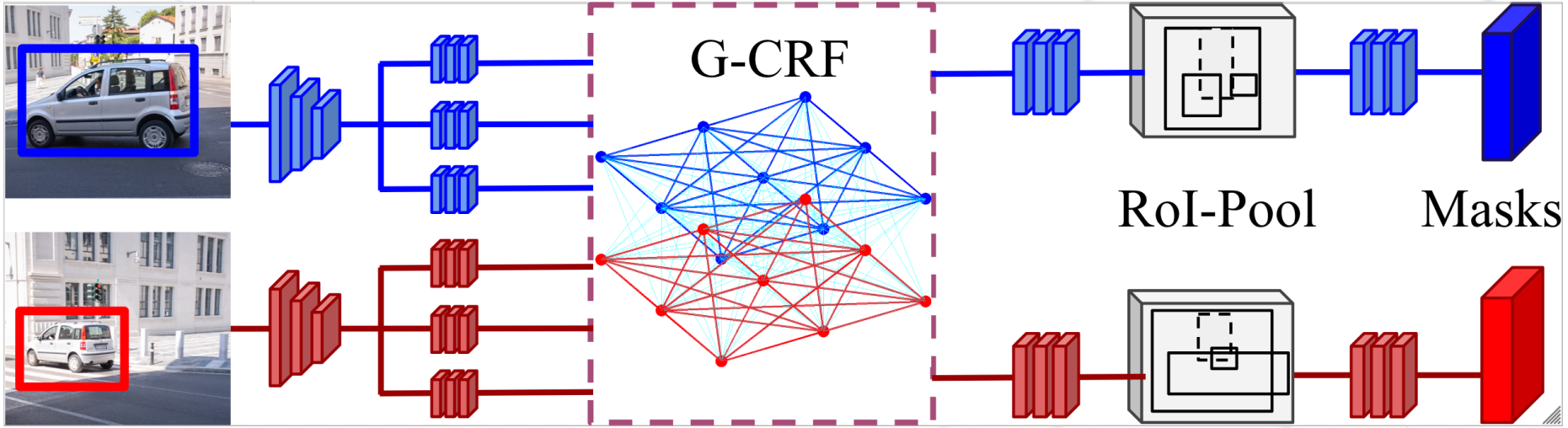


Video G-CRFs: Overview



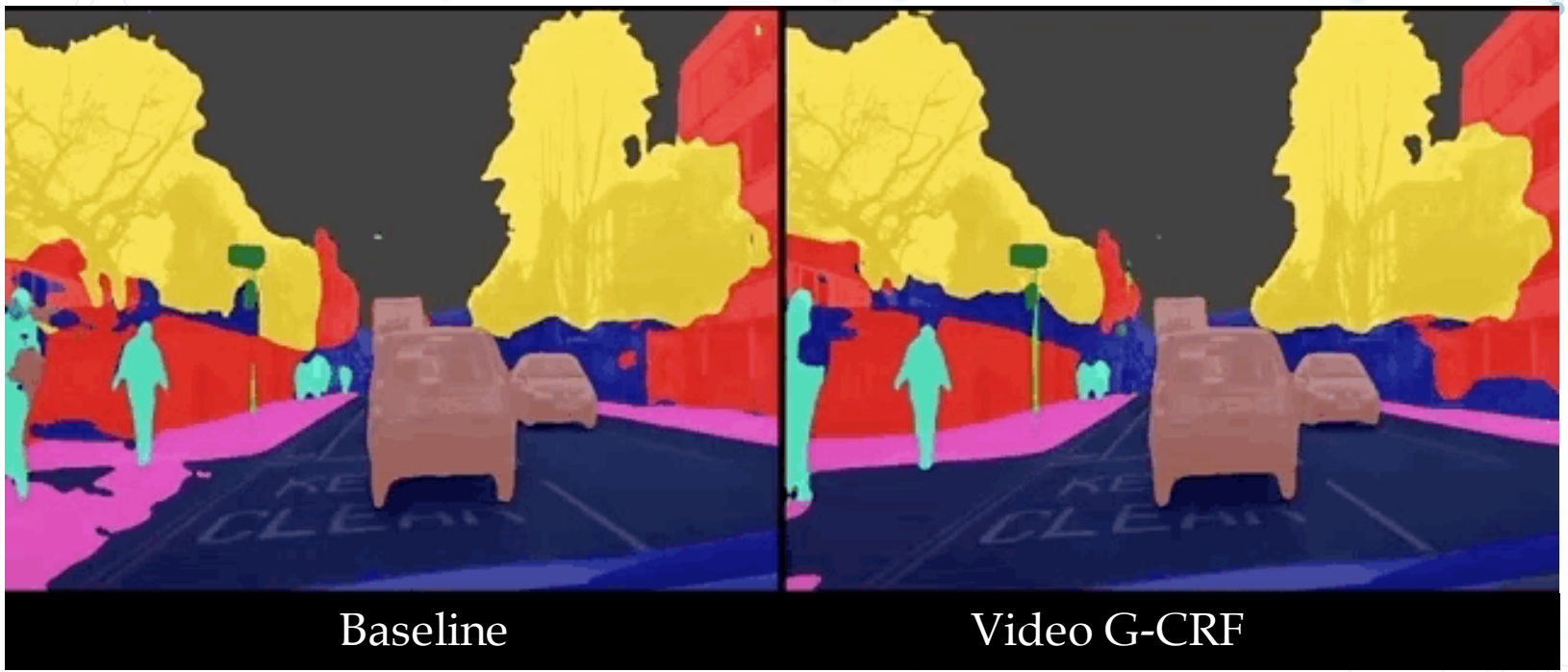
Video G-CRFs: Instance Segmentation

Extension of Mask RCNN [He 2017]



Video G-CRFs: Instance Tracking

Frame-by-Frame Prediction

Video G-CRF

Video G-CRFs: Semantic Segmentation