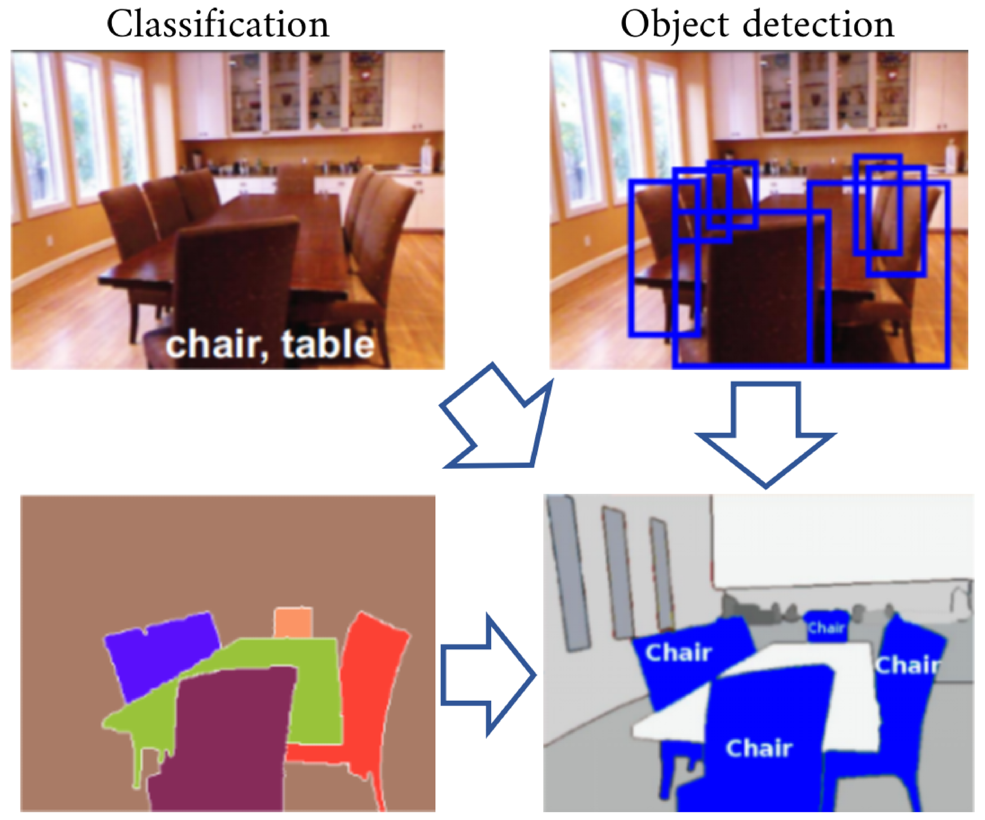
Semantic Segmentation Through Graph Neural Network Blocks

{darwin.pilco, adin}@ic.unicamp.br and antoine.tabbone@univ-lorraine.fr
Institute of Computing, University of Campinas, LORIA, Université de Lorraine

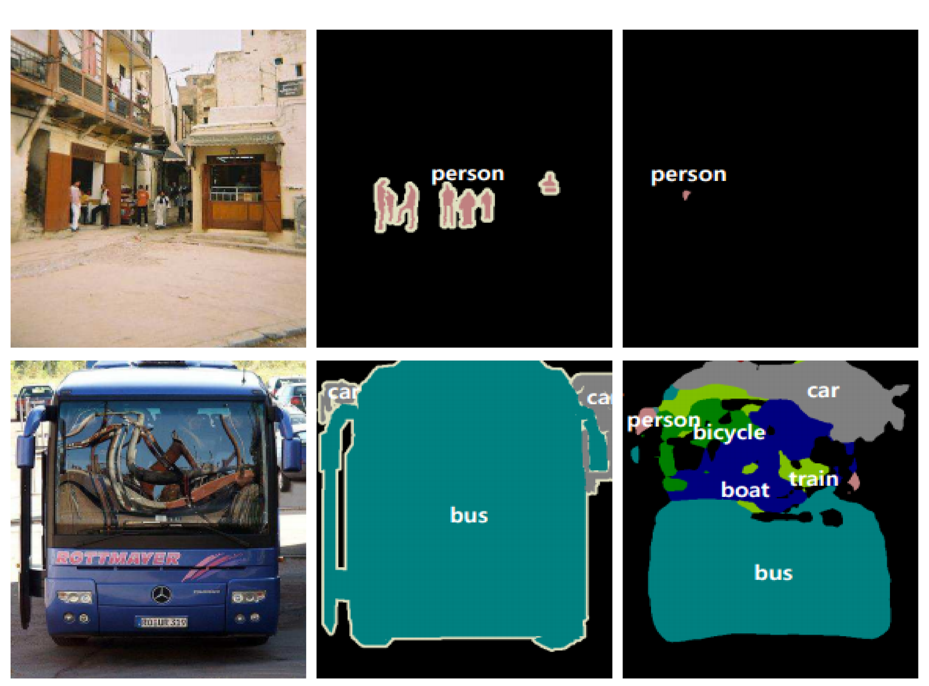
Context

Semantic segmentation



Semantic segmentation problems

- low-resolution in output heatmaps
- loss of spatial precision

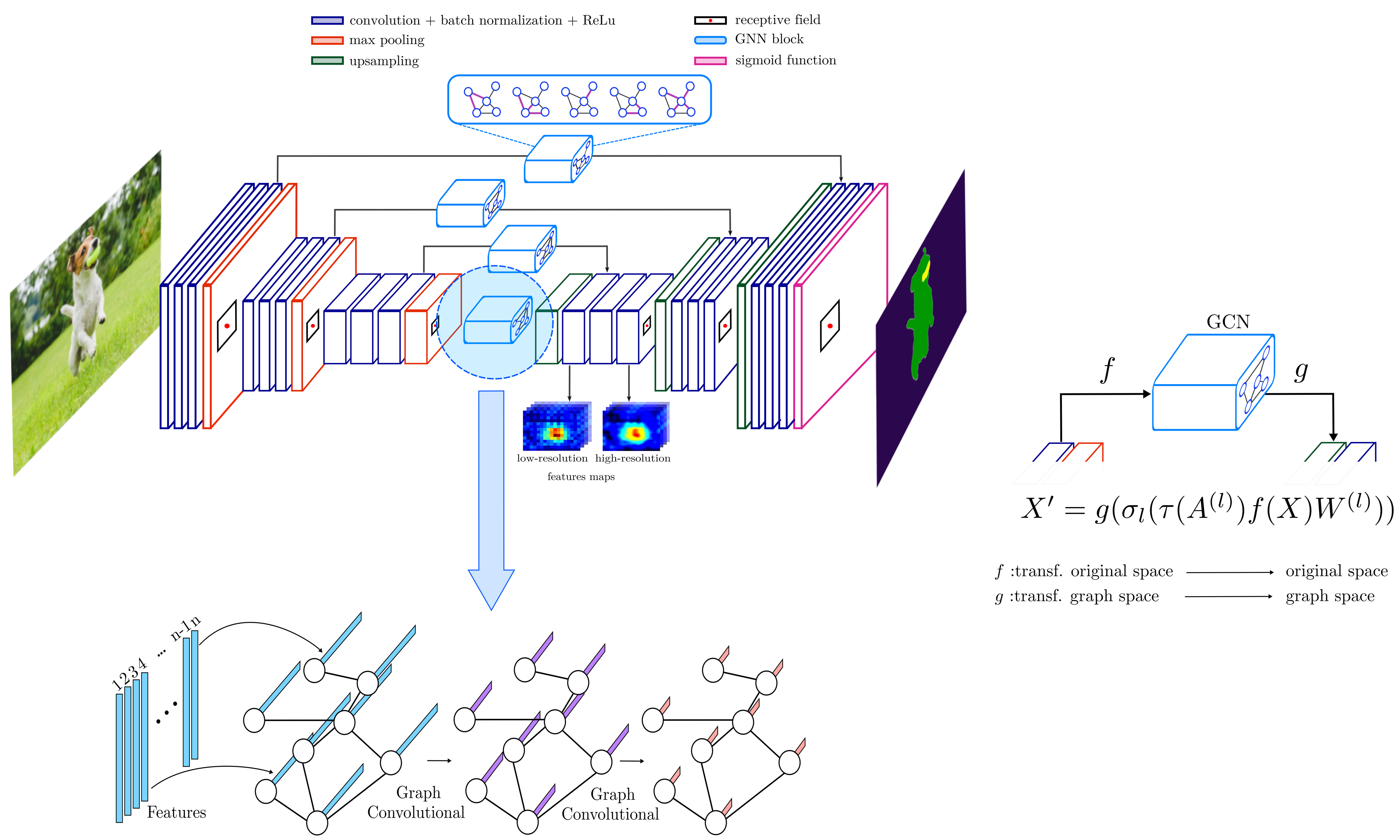


(a) Image (b) GT (c) Pred.

Overview

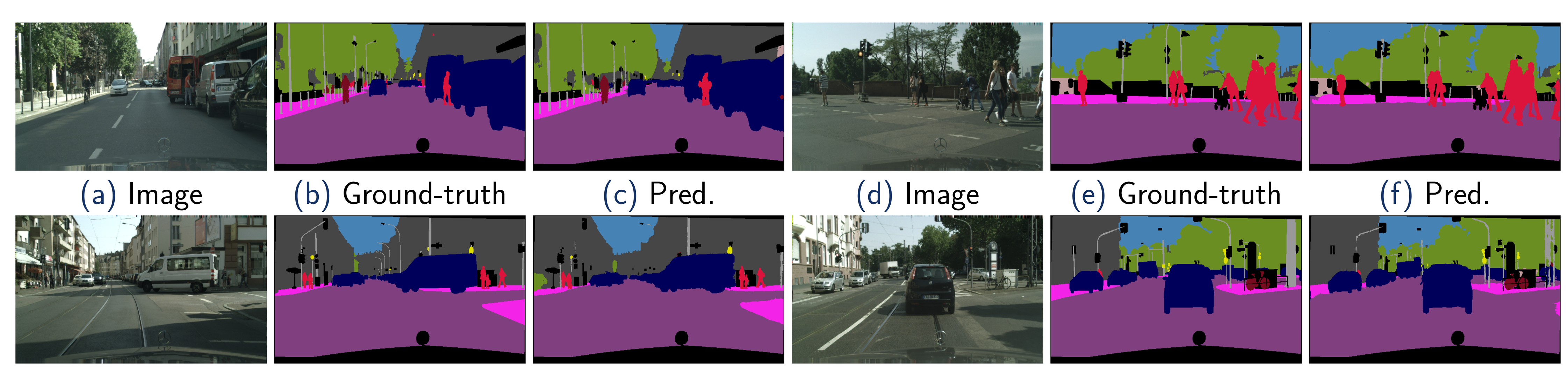
In this work, we design a new deep learning architecture that is end-to-end trainable to address the semantic segmentation task on images. Our architecture combines local features extraction of CNNs with the global features extraction of GNNs and their irregular connections between pixels through GNN-blocks (light blue squares). Thus, our neural network aims to produce densely labeled images.

Architecture for SS



Results

Model	Sky	Building	Road	Sidewalk	Fence	Vegetation	Pole	Car	Sign	Person	Cyclist	mIoU
SegNet	73.74	79.29	92.70	59.88	13.63	81.89	26.18	78.83	31.44	45.03	43.46	52.17
FCN8	76.51	83.97	93.82	67.67	24.91	86.38	31.71	84.80	50.92	59.89	59.11	59.97
FastNet	77.69	86.25	94.97	72.99	31.02	88.06	38.34	88.42	52.34	61.76	61.83	68.52
DeconvNet	89.38	83.08	95.26	68.07	27.58	85.80	34.20	85.01	27.62	45.11	41.11	62.02
DeepLabv2	74.28	81.66	90.86	63.30	26.29	84.33	27.96	86.24	44.79	58.89	60.92	63.59
ParseNet	77.57	86.81	95.27	74.02	33.31	87.37	38.24	88.99	53.34	63.25	63.87	69.28
DeepLabv3	92.82	89.02	96.74	78.13	41.00	90.81	49.74	91.02	64.48	66.52	66.98	75.21
GNN-block (Our work)	93.64	88.69	96.42	74.63	41.46	90.97	52.30	89.79	69.40	70.36	68.59	76.02
AdapNet++	**94.18**	**91.49**	**97.93**	**84.40**	**54.98**	**92.09**	**58.85**	**93.86**	**72.61**	**75.52**	**72.90**	**80.80**



(a) Image (b) Ground-truth (c) Pred. (d) Image (e) Ground-truth (f) Pred.

(g) Image (h) Ground-truth (i) Pred. (j) Image (k) Ground-truth (l) Pred.

Table: Comparison results on test set from Cityscape dataset.

Methodology

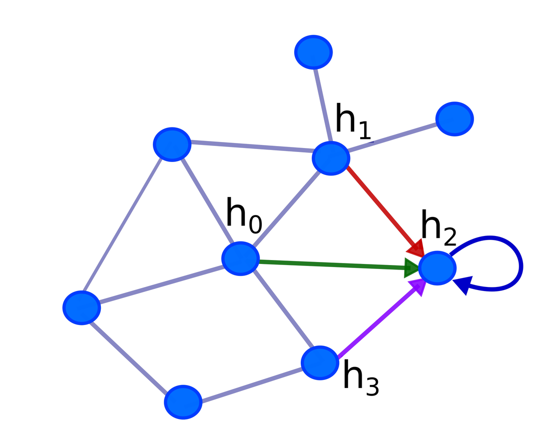
Graph convolutional network, creating and deleting edges and updating features values.

$$H^{(l+1)} = \sigma_l\left(\tau\left(A^{(l)}\right)H^{(l)}W^{(l)}\right), \tag{1}$$

$$\tau\left(A^{(l)}\right) = \left(\hat{D}^{(l)}\right)^{-\frac{1}{2}}\left(A^{(l)} + I_n\right)\left(\hat{D}^{(l)}\right)^{-\frac{1}{2}}, \tag{2}$$

$$\hat{D}^{(l)} = D^{(l)} + I_n, \tag{3}$$

Example



$$h_2^{(l+1)} = \sigma_l\Big(\frac{1}{a_{2,0}}h_0^{(l)}w_0^{(k)} + \frac{1}{a_{2,1}}h_1^{(l)}w_1^{(l)} + \frac{1}{a_{2,3}}h_3^{(l)}w_3^{(l)} + h_2^{(l)}w_2^{(l)}\Big)$$

Where

- $H^{(l)}$ is the feature vectors for the l-th neural network layer.
- $\sigma(.)$ is a non-linear activation function like the RELU or sigmoid.
- $W^{(l)}$ is the weights of l-th layer for the feature.
- $A^{(l)}$ is a not normalized space matrix and $\hat{A}^{(l)}$ is a sparce normalization ones.
- I is the identity matrix.
- $D^{(l)}$ is the degree matrix.

Loss Functions

$$\mathcal{L}_{cross\_ss} = -\frac{1}{N}\sum_{i=1}^{N}\alpha_i \log P(s = s_i|X;\phi), \tag{4}$$

$$\mathcal{L}_{iou\_ss} = 1 - \sum_{i=1}^{N}\frac{P_i \cap S_i}{P_i \cup S_i}. \tag{5}$$

$$\mathcal{L}_{ss} = \psi_3\mathcal{L}_{cross\_ss} + \psi_4\mathcal{L}_{iou\_ss}, \tag{6}$$