Contextual Recurrent Level Set Networks and Recurrent Residual Networks for Semantic Labeling

Ngan Thi Hoang Le

Thesis Committee:

- Prof. Marios. Savvides (advisor), ECE-CMU
- Prof. Vijayakumar Bhagavatula, ECE-CMU
- Prof. Arun A. Ross, CS-MSU
- Dr. Saad J. Bedros, University of Minnesota

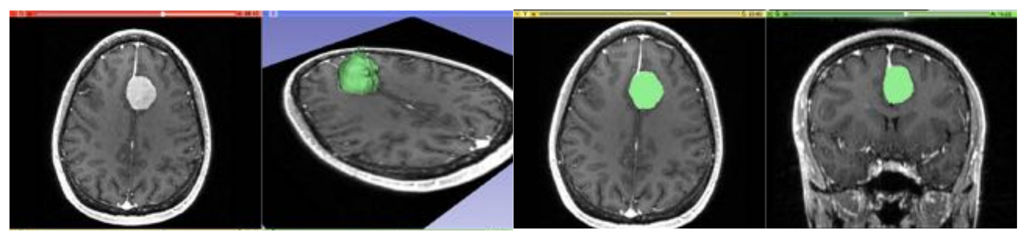
April 28, 2018

Semantic Labeling process of associating each pixel of an image with a class label

Scene understanding

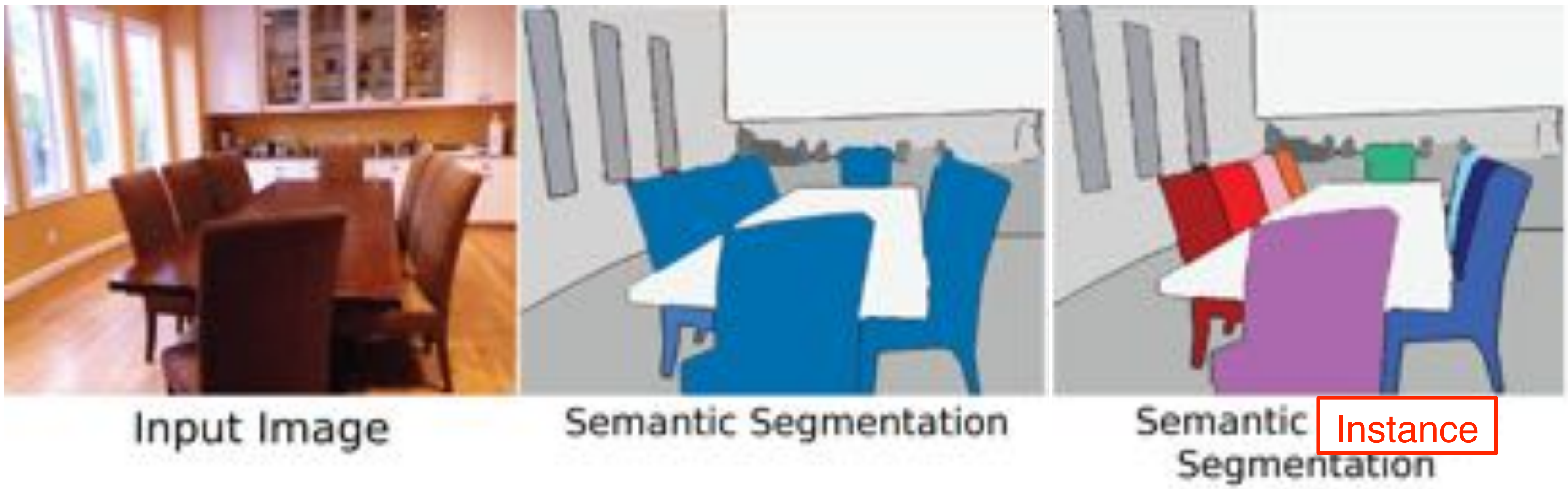
Semantic image searching

Object tracking



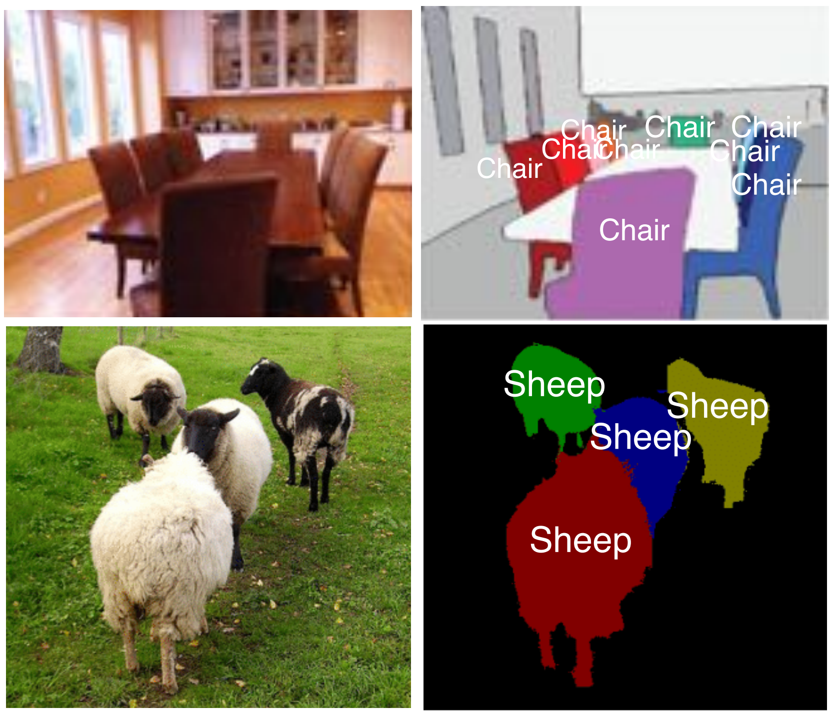
Medical imaging

Semantic Labeling process of associating each pixel of an image with a class label



Semantic Instance Segmentation

provide pixel-level segmentation to each object **instance** in the image



Scene Understanding

predict the semantic class of the individual pixels (semantic segmentation)

Semantic Instance Segmentation

Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation

- Motivation
- Level Set
- Recurrent Neural Networks (RNNs)
- Recurrent Level Set (RLS)
- Contextual Recurrent Level Set (CRLS) Networks

Motivation	Level Set	RNNs	RLS	CRLS

Scene Understanding

Contextual Recurrent Residual Networks (CRRN) for Scene Labeling

Image Segmentation

Unsupervised

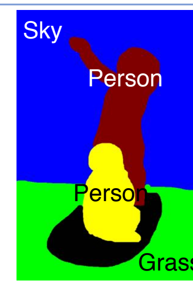
- Thresholding
- Edge-based
- Region-based
- Watershed
- Normalised Graph Cut
- Active Contour

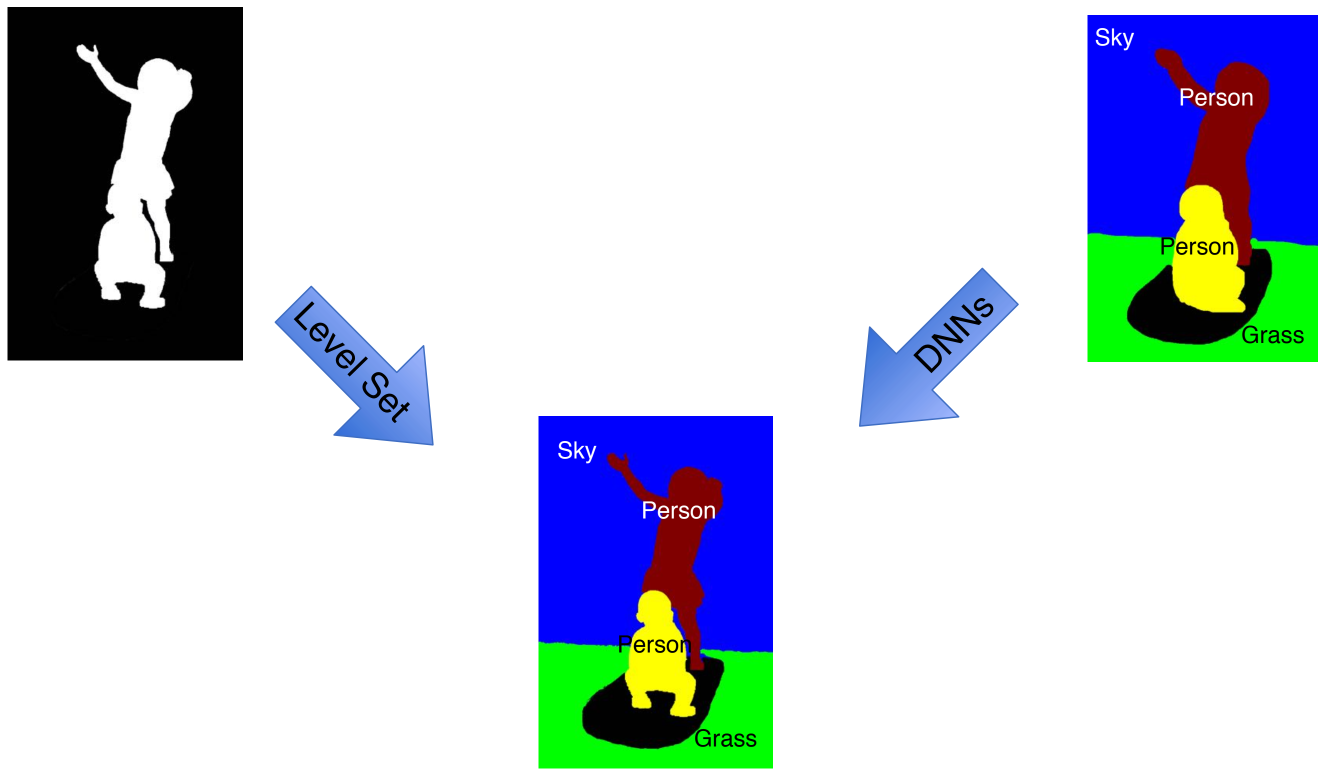
Level Set

Supervised

- Traditional learning
- Deep learning

DNNs
(CNNs, RNNs)

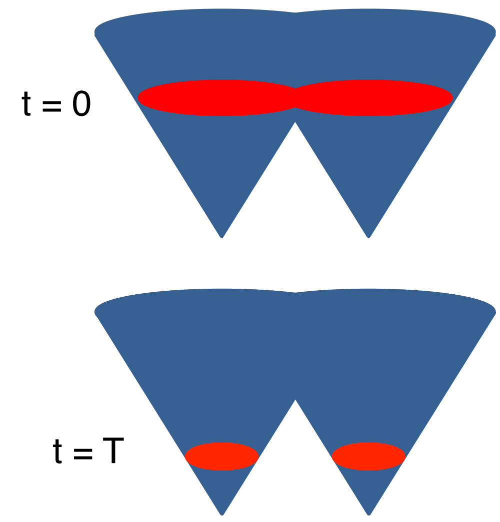


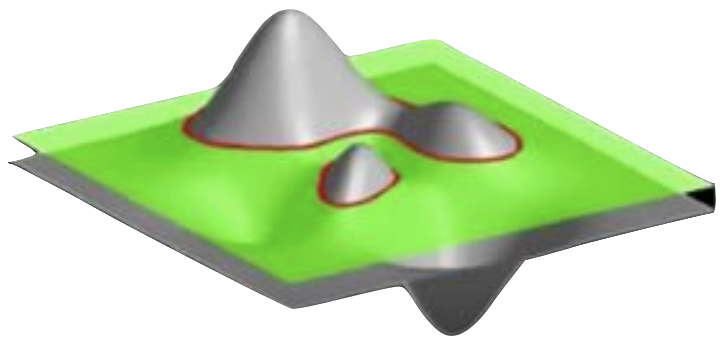


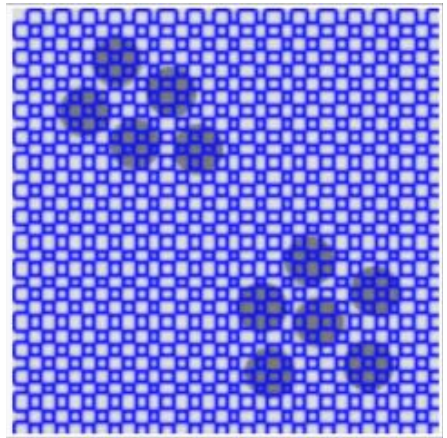
Contextual Recurrent Level Set (CRLS) Networks
for Semantic Instance Segmentation

Water falling from the top of a hill, the goal is to track the water front while it's moving down the hill

Where is the water front at a given time t ?

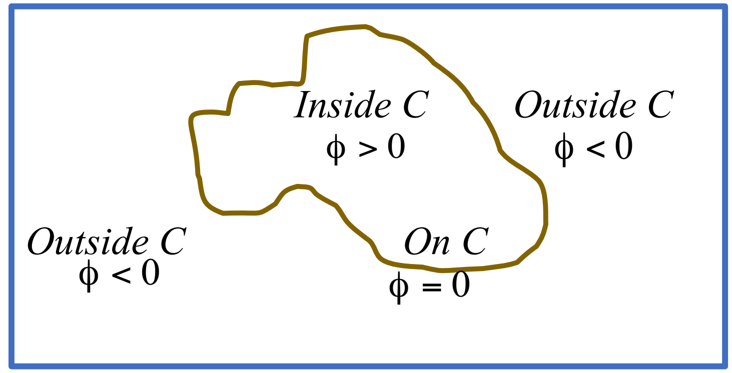






Initial Curve → Evolution → Detected Object

$$\phi_t = \phi_{t-1} + \eta \frac{\partial \phi_{t-1}}{\partial t}$$

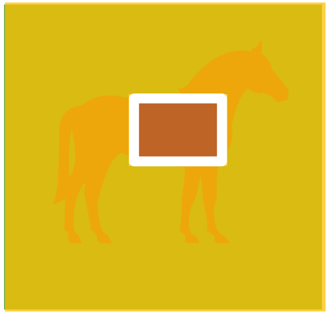


Initial Curve → Evolution → Detected Object

$$E(C) = \underbrace{\int_{inside(C)} |\mathbf{I} - c_1|^2 dxdy}_{E_1} + \underbrace{\int_{outside(C)} |\mathbf{I} - c_2|^2 dxdy}_{E_2}$$

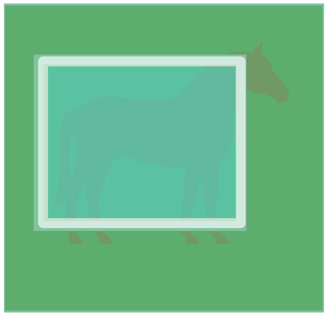
c_1 c_2

$E_1 > 0$
$E_2 = 0$
$E = E_1 + E_2 > 0$



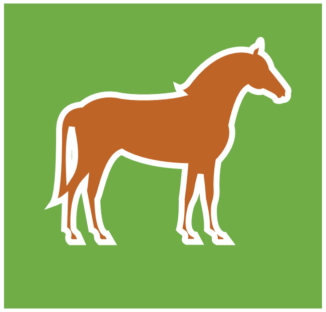
c_1 c_2

$E_1 = 0$
$E_2 > 0$
$E = E_1 + E_2 > 0$



c_1 c_2

$E_1 > 0$
$E_2 > 0$
$E = E_1 + E_2 > 0$



c_1 c_2

$E_1 = 0$
$E_2 = 0$
$E = E_1 + E_2 = 0$

$$E(C) = \int_{inside(C)} |\mathbf{I} - c_1|^2 dxdy + \int_{outside(C)} |\mathbf{I} - c_2|^2 dxdy$$

Fitting + Regularization terms (length, area)

$$E(c_1, c_2, C) = \mu Length(C) + \nu Area(insideC) + \lambda_1 \int_{insideC} |I - c_1|^2 dxdy + \lambda_2 \int_{outsideC} |I - c_2|^2 dxdy$$

Under the zero-level set

$$\phi(x,y) = \begin{cases} > 0 & \text{if } (x,y) \text{ is inside } C \\ =0 & \text{if } (x,y) \text{ is on } C \\ < 0 & \text{if } (x,y) \text{ is outside } C \end{cases}$$

$$E(c_1, c_2, C) = \mu \int_\Omega \delta(\phi)|\bigtriangledown \phi(x,y)|dxdy + \nu \int_\Omega H(\phi(x,y))dxdy + \lambda_1 \int_\Omega |\mathbf{I}(x,y) - c_1|^2 H(\phi(x,y))dxdy + \lambda_2 \int_\Omega |\mathbf{I}(x,y) - c_2|^2 (1 - H(\phi(x,y)))dxdy$$

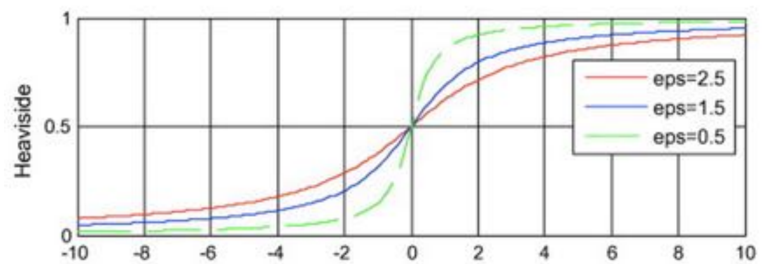
$$E(c_1, c_2, C) = \mu \boxed{\int_\Omega \delta(\phi) | \bigtriangledown \phi(x,y)| dxdy} + \nu \boxed{\int_\Omega H(\phi(x,y)) dxdy}$$
$$+ \lambda_1 \int_\Omega |\mathbf{I}(x,y) - c_1|^2 H(\phi(x,y)) dxdy + \lambda_2 \int_\Omega |\mathbf{I}(x,y) - c_2|^2 (1 - H(\phi(x,y))) dxdy$$

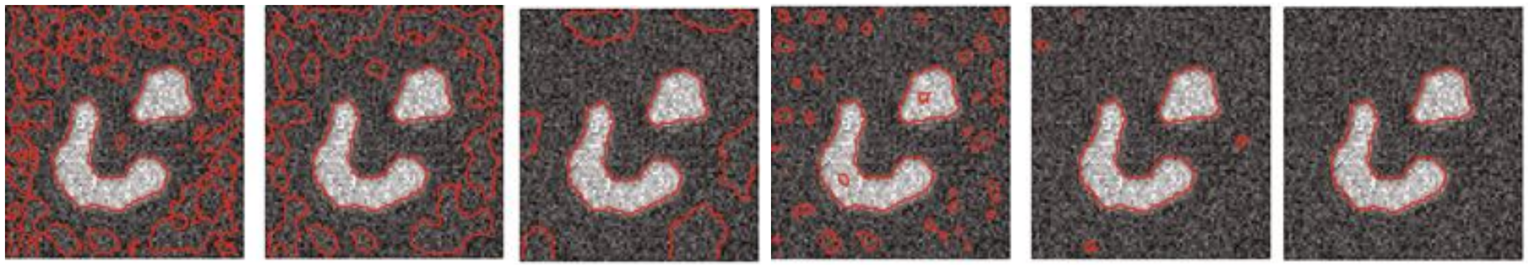
$$Length(C) = \int_\Omega |\nabla H(\phi(x,y))| dxdy = \int_\Omega \delta(\phi(x,y)) |\nabla \phi(x,y)| dxdy$$

$$Area(C) = \int_\Omega H(\phi(x,y)) dxdy$$

$$H_\epsilon(x) = \frac{1}{2}\left(1 + \frac{2}{\pi} \arctan\left(\frac{x}{\epsilon}\right)\right)$$

$$\delta_\epsilon(x) = H'(x) = \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + x^2}$$





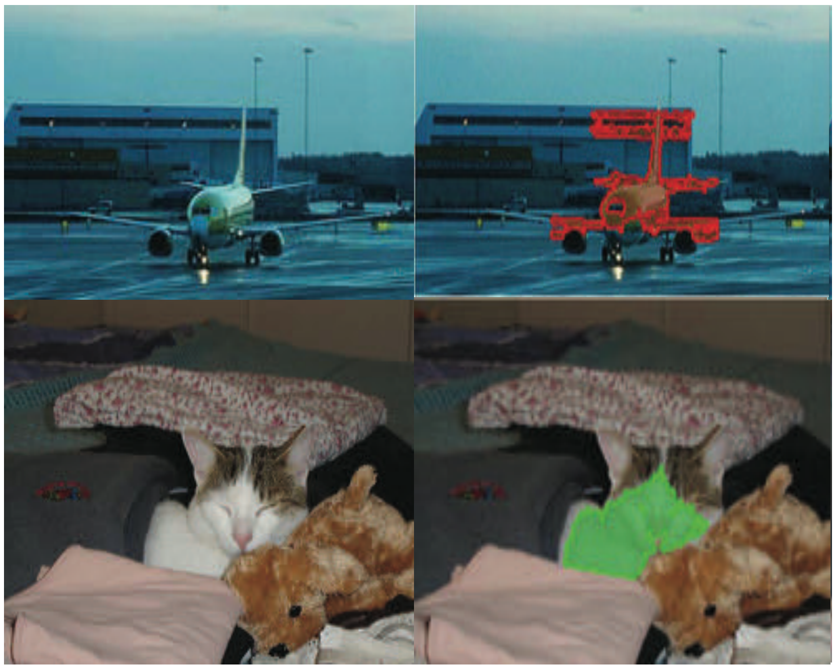
Advantages

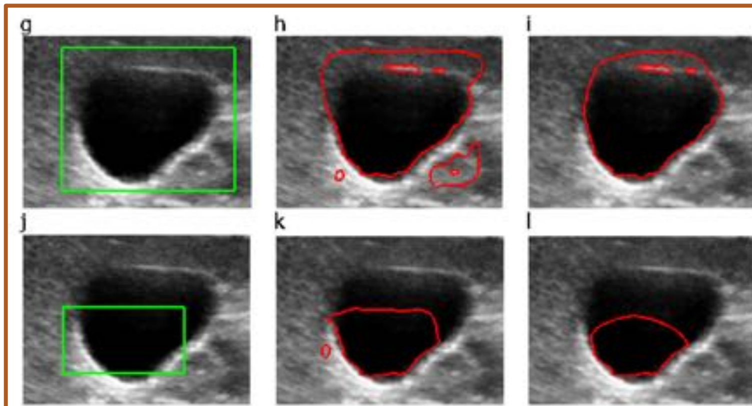
- Do not rely on image gradient, robust to noise
- Find boundaries without edge information

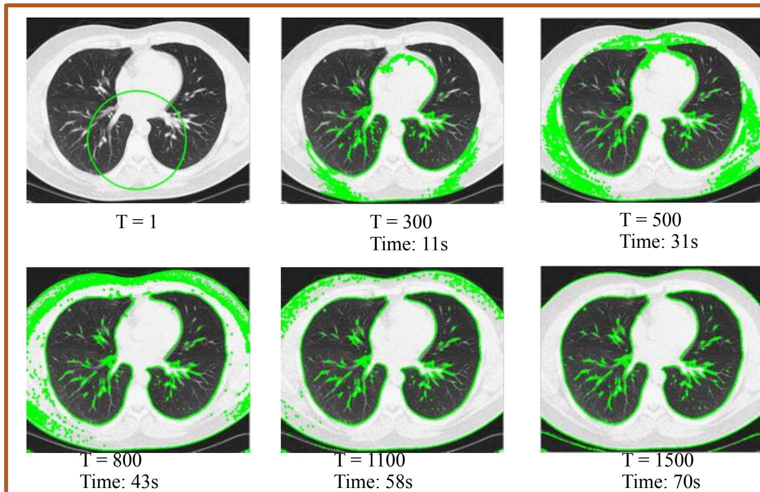
Disadvantages

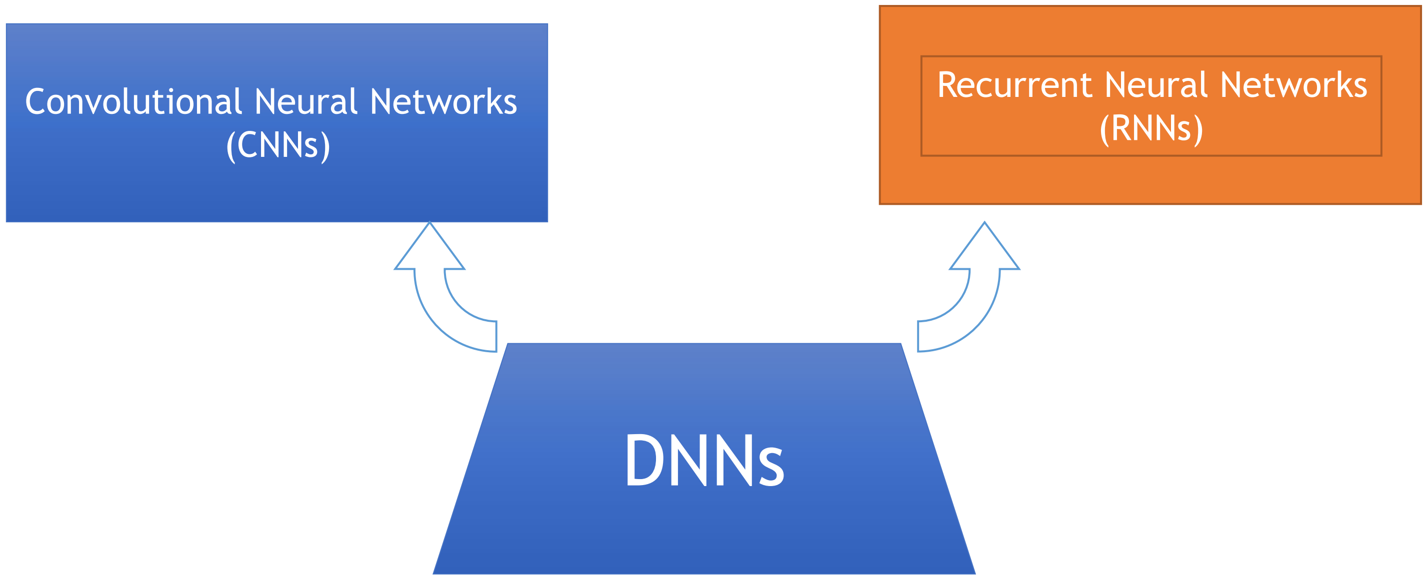
- Computationally expensive
- Sensitive to the initialization
- Depends of the number of iterations
- No learning from training data
- Many proper parameters
- Unable to segment images in the wild

Disadvantages








Convolutional Neural Networks (CNNs)
Recurrent Neural Networks (RNNs)
DNNs

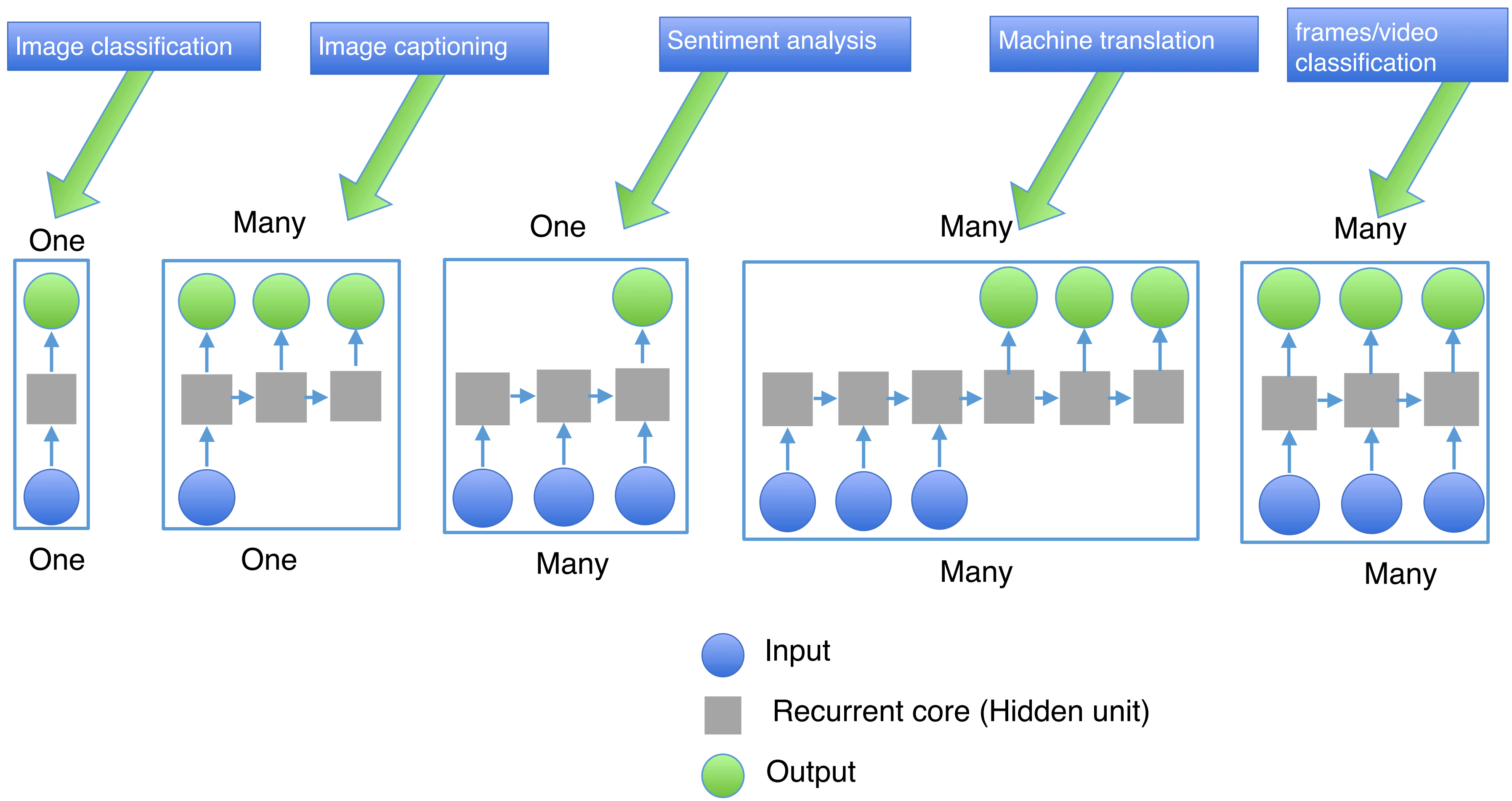
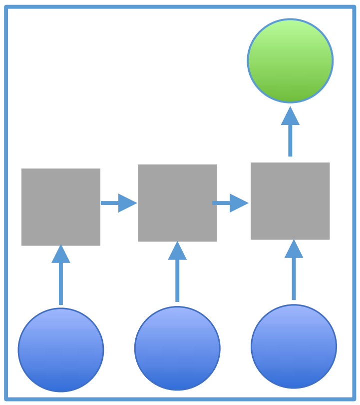
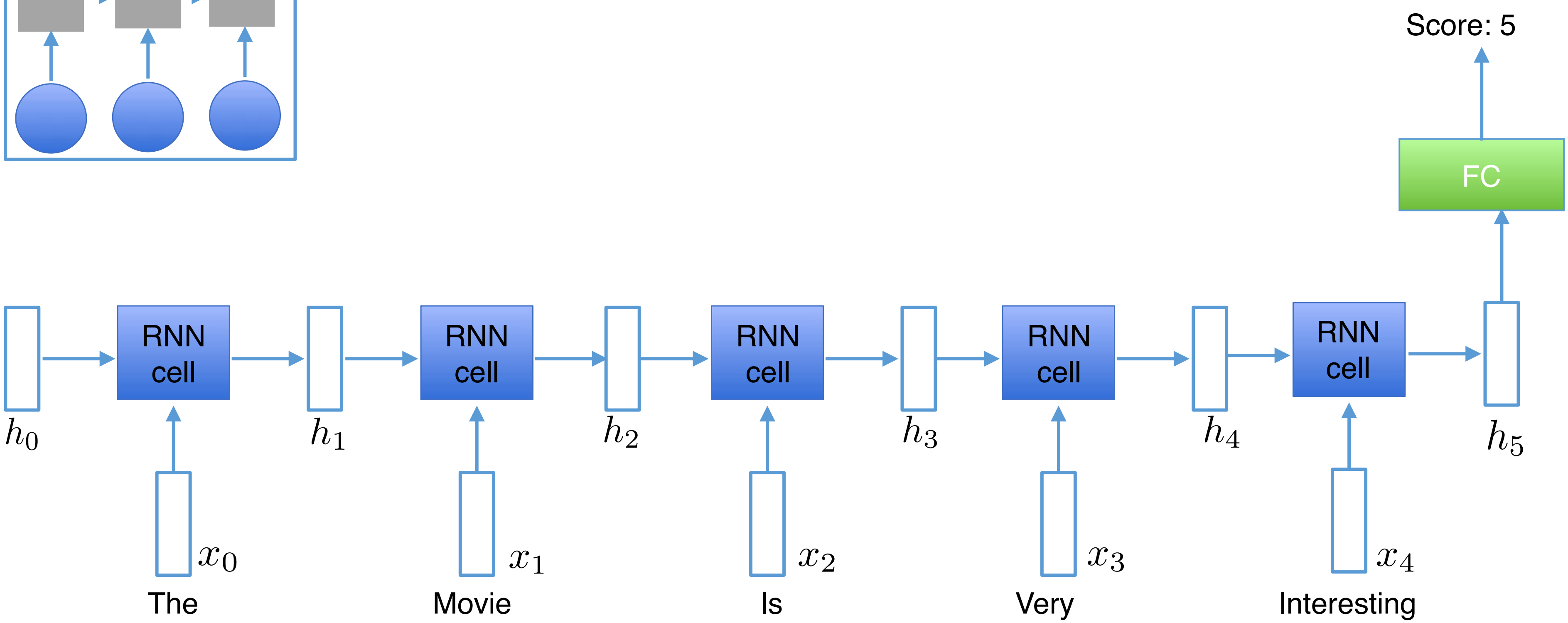
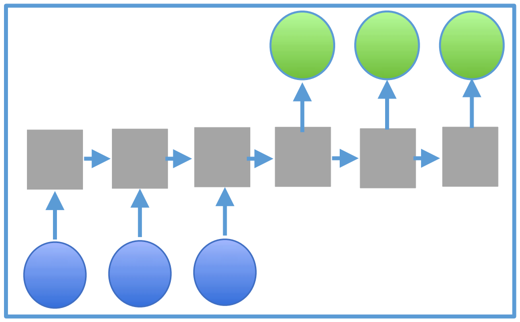


Image classification
Image captioning
Sentiment analysis
Machine translation
frames/video classification
One
Many
One
Many
Many
One
One
Many
Many
Many
Input
Recurrent core (Hidden unit)
Output

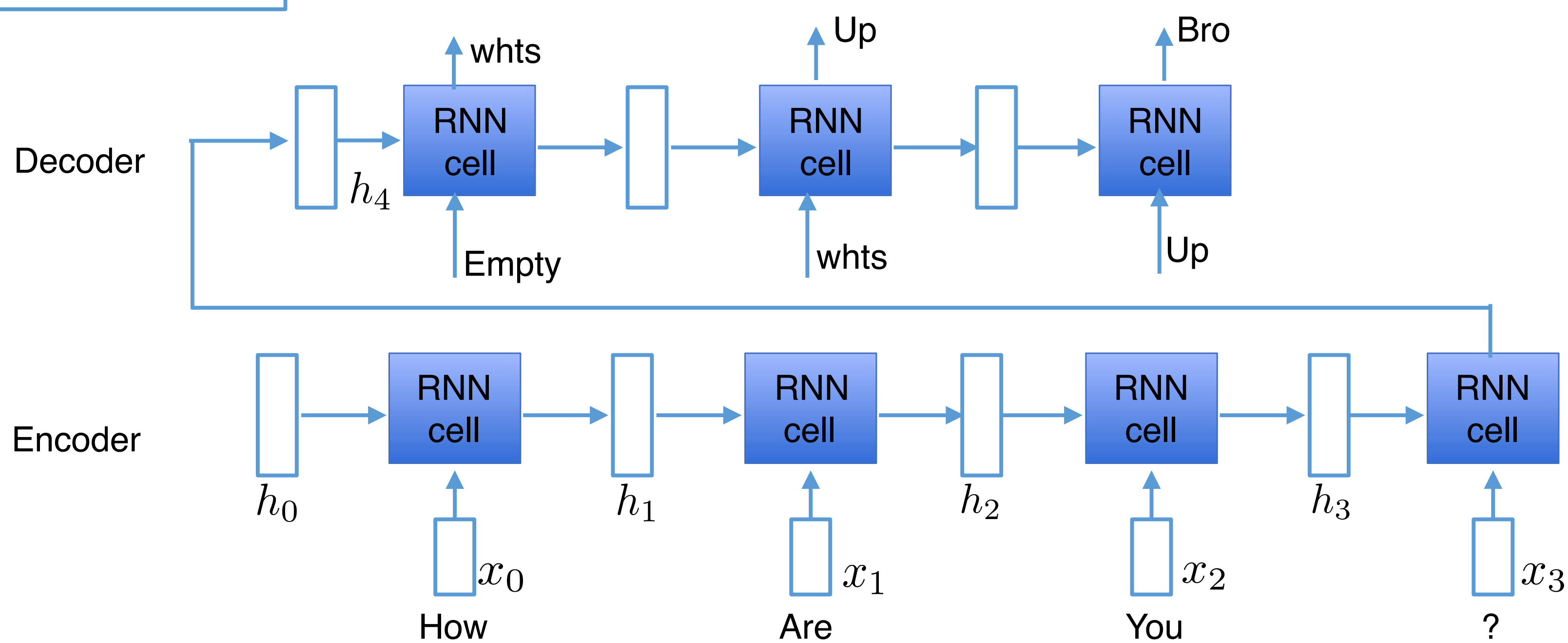
Sentiment Analysis

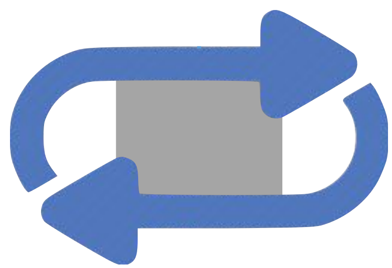


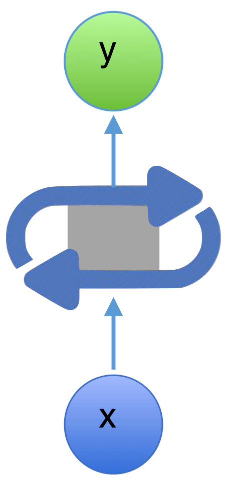


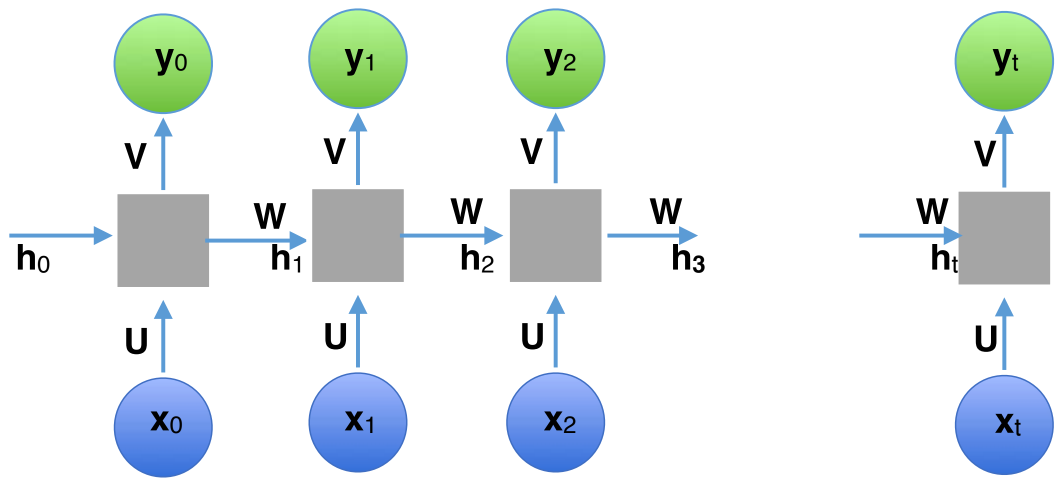


Machine Translation

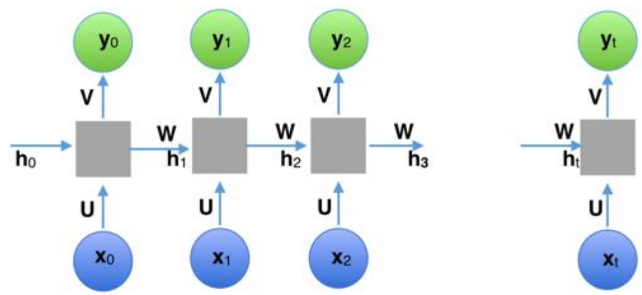








$$\mathbf{h}_t = tanh(\mathbf{U}\mathbf{x}_t + \mathbf{W}\mathbf{h}_{t-1})$$



Vanishing/Exploding Gradient

Backpropagation

$$w = w + \triangle w$$

$$\triangle w = \frac{de}{dw}$$

$$e = (GT - predicted)^2$$

Exploding

$$\text{if } \frac{de}{dw} >>>>$$

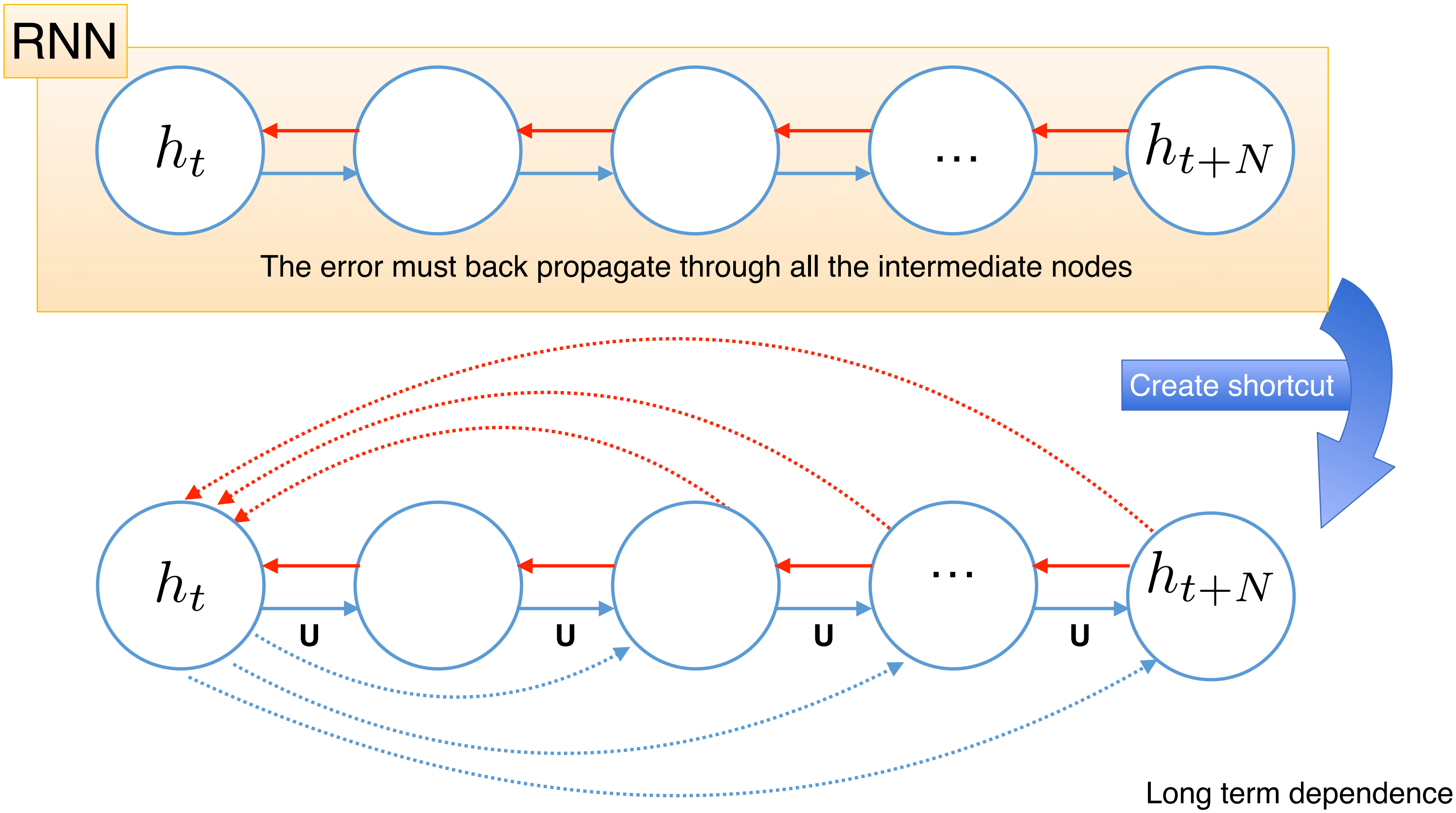
$$\triangle w >>>>>$$

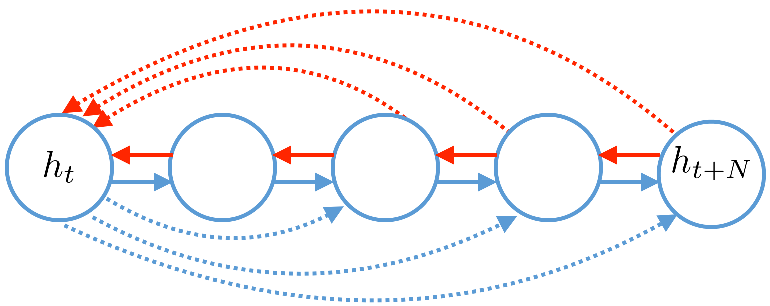
Vanishing

$$\text{if } \frac{de}{dw} <<<<<$$

$$\triangle w <<<<<<$$

No update for weight


RNN
h_t
$\cdots$
h_{t+N}
The error must back propagate through all the intermediate nodes
Create shortcut
h_t
$\cdots$
h_{t+N}
U
U
U
U
Long term dependence



Adaptive shortcut connection to decide how much to pay attention to the past as well as to the present

previous time step through the short ut

$$\mathbf{h}_t = f(\mathbf{h}_t, \mathbf{z}_t) = \mathbf{z}_t\tilde{\mathbf{h}}_t + (1 - \mathbf{z}_t)\mathbf{h}_{t-1}$$

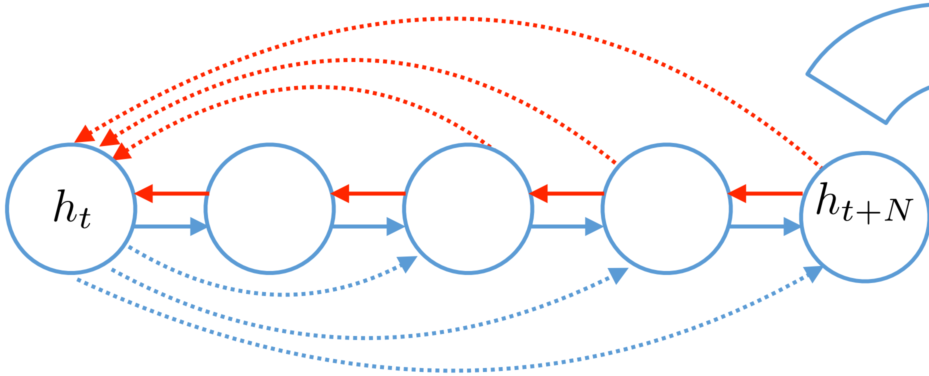
Standard RNN

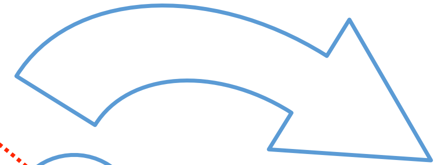
Candidate memory $\tilde{\mathbf{h}}_t = tanh(\mathbf{U}_h\mathbf{x}_t + \mathbf{W}_h\mathbf{h}_{t-1})$

Update gate $\mathbf{z}_t = \sigma(\mathbf{U}_z\mathbf{x}_t + \mathbf{W}_z\mathbf{h}_{t-1})$

Once paying attention to, it keeps adding forever

Prune





Prune unnecessary connection adaptively

$$\mathbf{h}_t = f(\mathbf{h}_t, \mathbf{z}_t) = \mathbf{z}_t \tilde{\mathbf{h}}_t + (1 - \mathbf{z}_t)\mathbf{h}_{t-1}$$

Candidate memory

$$\cancel{\tilde{\mathbf{h}}_t = tanh(\mathbf{U}_h \mathbf{x}_t + \mathbf{W}_h \mathbf{h}_{t-1})}$$

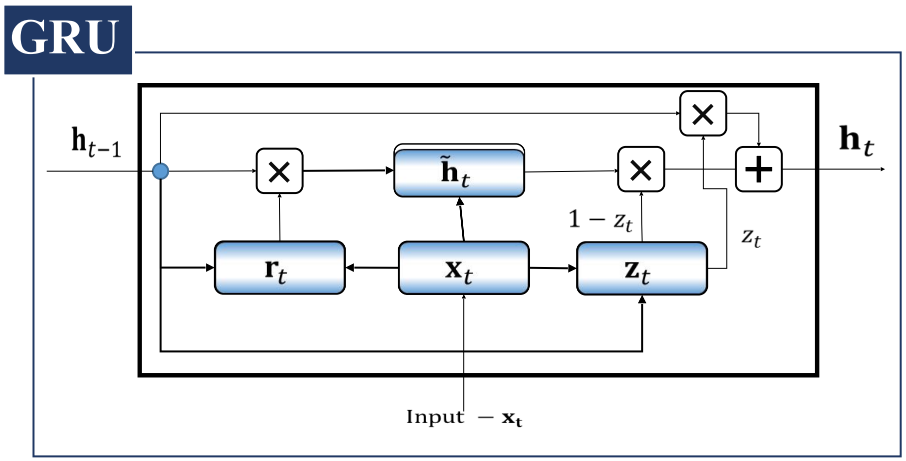
$$\tilde{\mathbf{h}}_t = tanh(\mathbf{U}_h \mathbf{x}_t + \mathbf{W}_h \mathbf{h}_{t-1} \boxed{\circ\, \mathbf{r}_t})$$

Update gate

$$\mathbf{z}_t = \sigma(\mathbf{U}_z \mathbf{x}_t + \mathbf{W}_z \mathbf{h}_{t-1})$$

Reset gate

$$\mathbf{r}_t = \sigma(\mathbf{U}_r \mathbf{x}_t + \mathbf{W}_r \mathbf{h}_{t-1})$$



Reset gate $\mathbf{r}_t = \sigma(\mathbf{U}_r\mathbf{x}_t + \mathbf{W}_r\mathbf{h}_{t-1})$

Update gate $\mathbf{z}_t = \sigma(\mathbf{U}_z\mathbf{x}_t + \mathbf{W}_z\mathbf{h}_{t-1})$

Candidate memory $\tilde{\mathbf{h}}_t = tanh(\mathbf{U}_h\mathbf{x}_t + \mathbf{W}_h\mathbf{h}_{t-1} \circ \mathbf{r}_t)$

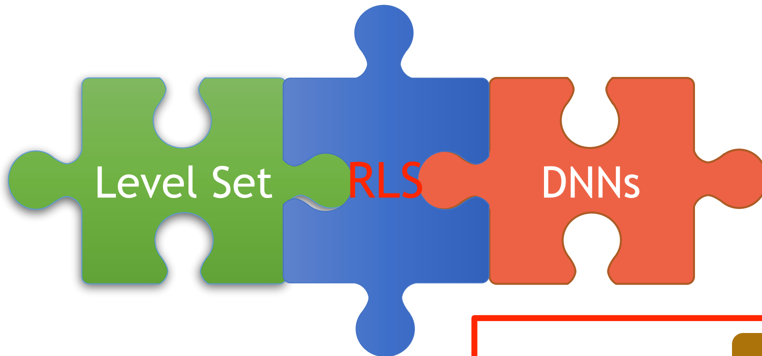
Final memory $\mathbf{h}_t = \mathbf{z}_t\tilde{\mathbf{h}}_t + (1-\mathbf{z}_t)\mathbf{h}_{t-1}$

r $\rightarrow 0$: ignore previous hidden state $\rightarrow$ drop information that is irrelevant in the future

z: controls how much of past state should matter

Units with short-term dependencies: active **r**

Units with long-term dependencies: active **z**

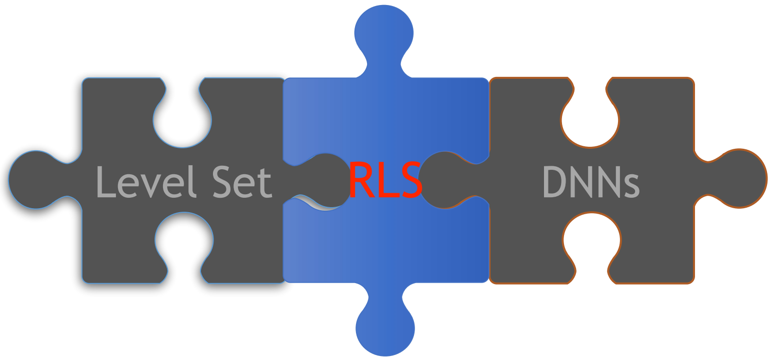


Advantages

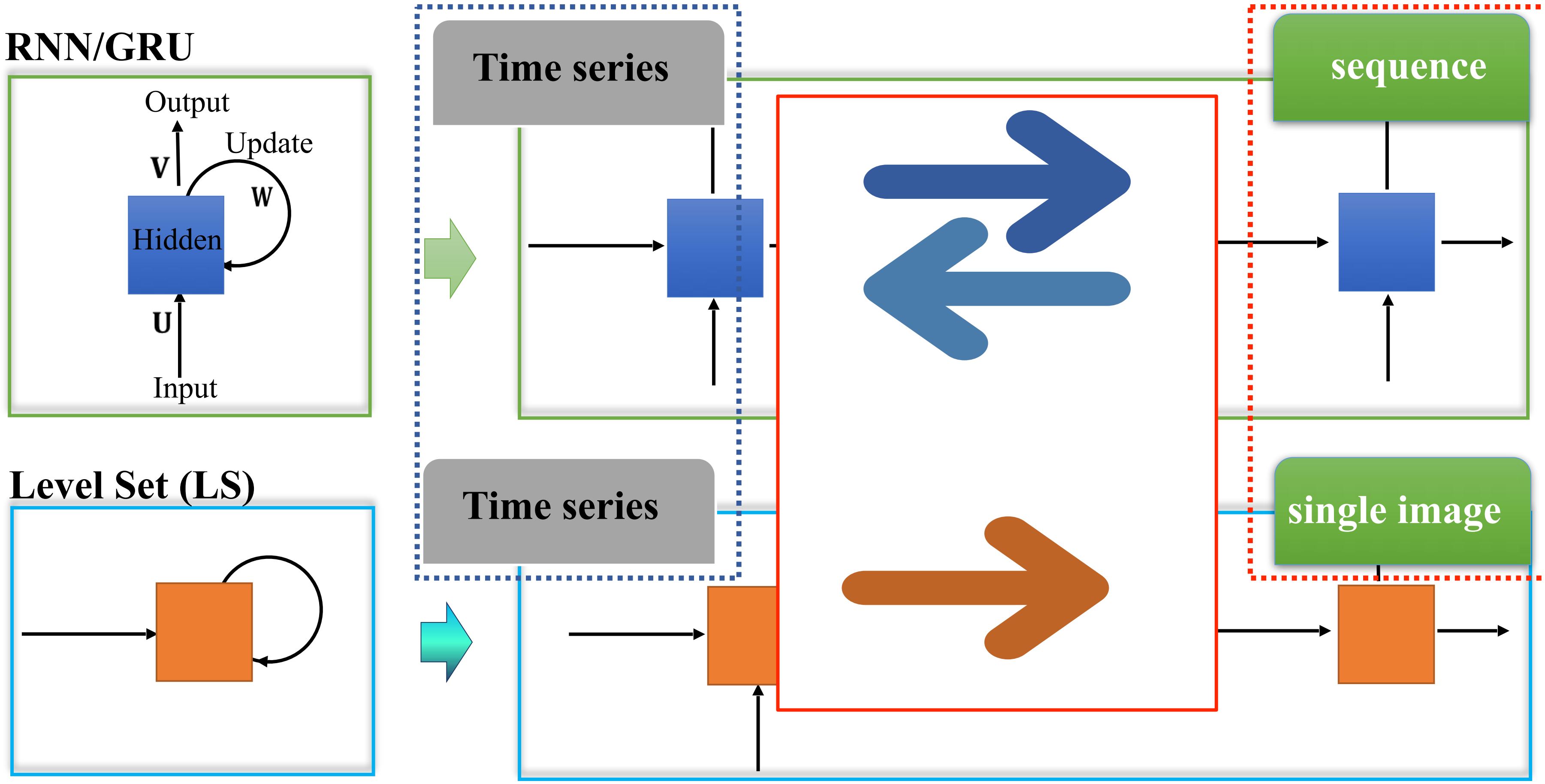
- Do not rely on image gradient, robust to noise
- Find boundaries without edge information

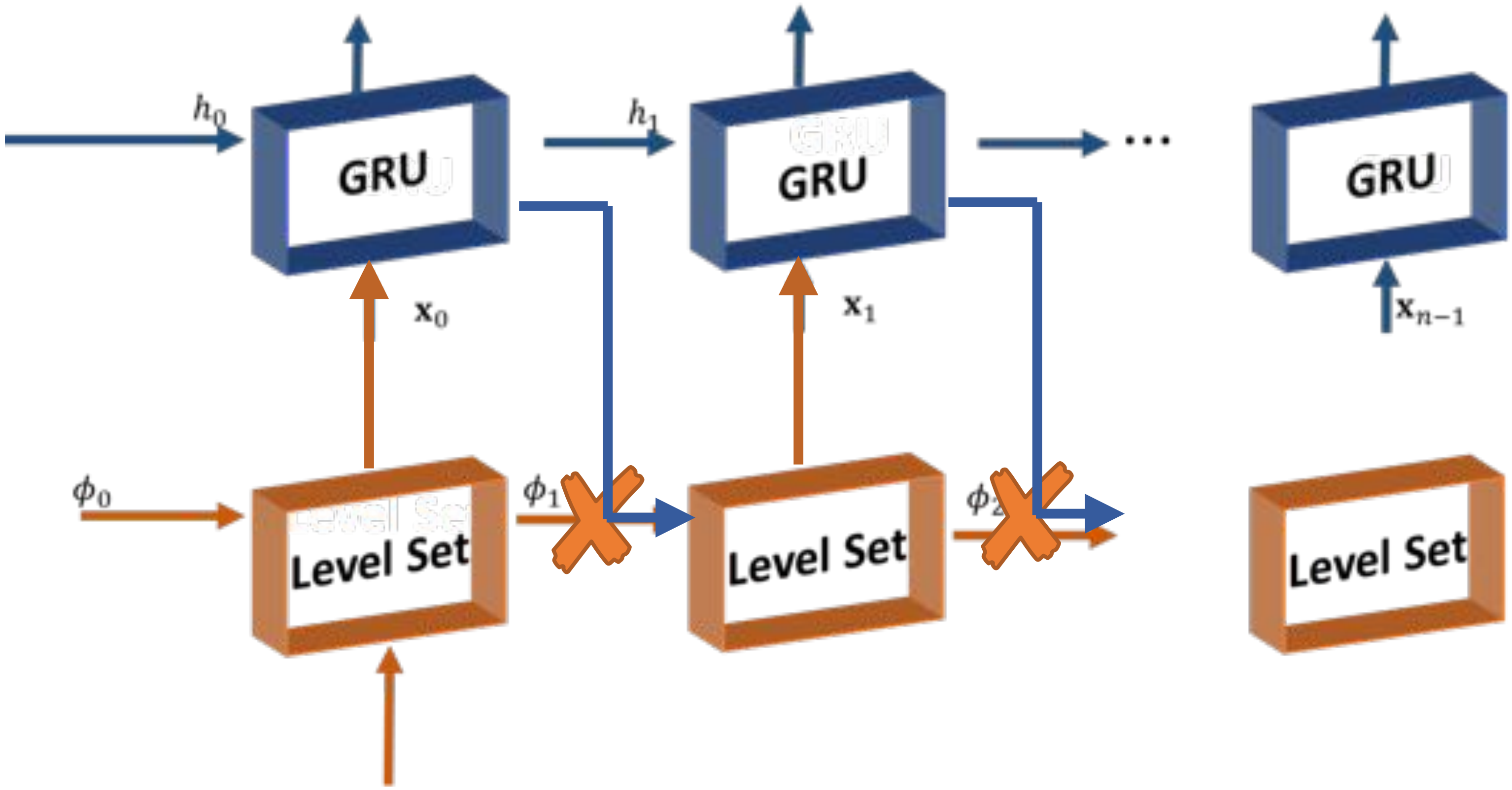
~~Dis~~advantages

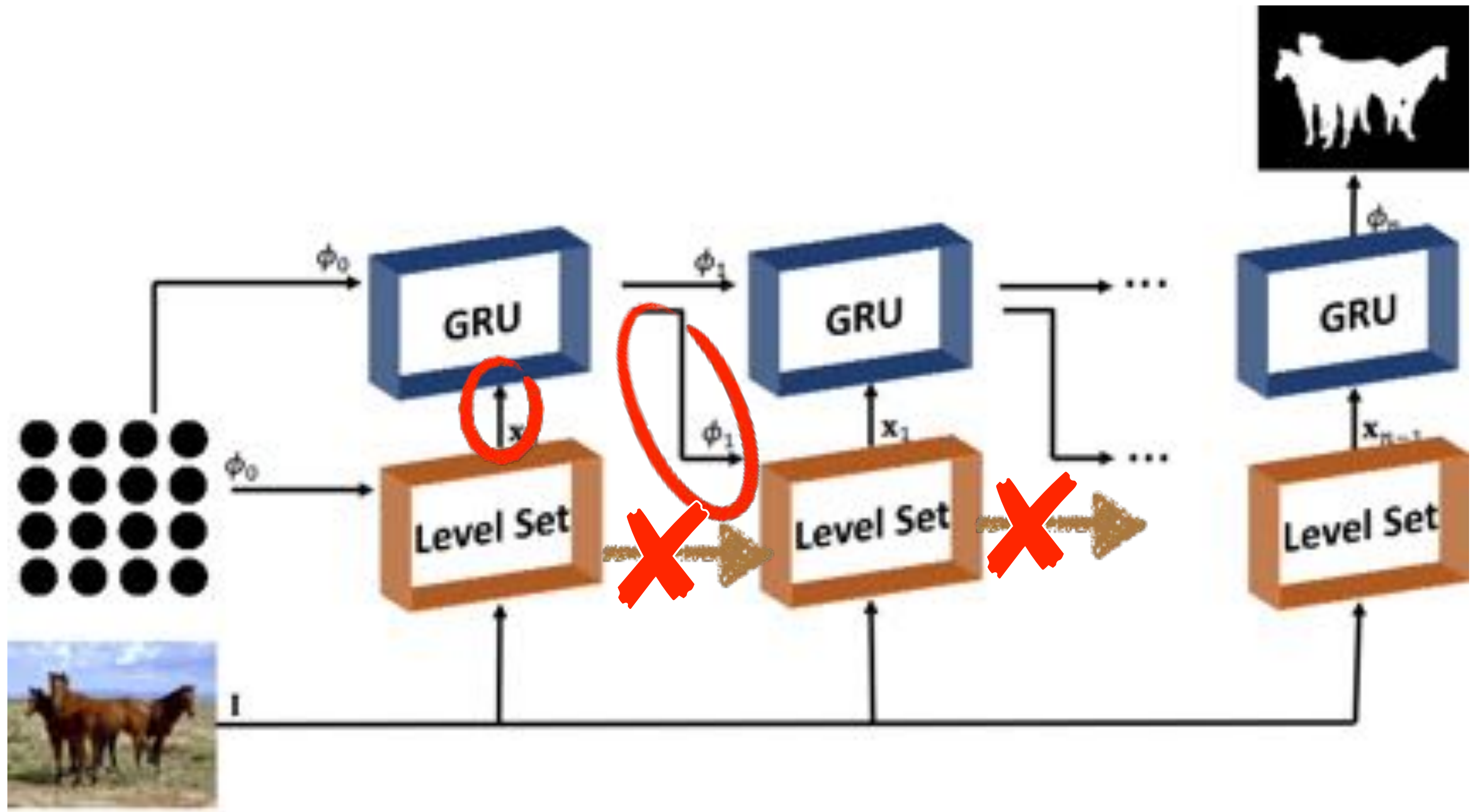
- Computationally ~~expensive~~
- ~~Sensitive~~ to the initialization
- ~~Depends~~ of the number of iterations
- ~~No~~ learning from training data
- ~~Many~~ proper parameters
- ~~Unable~~ to segment images in the wild



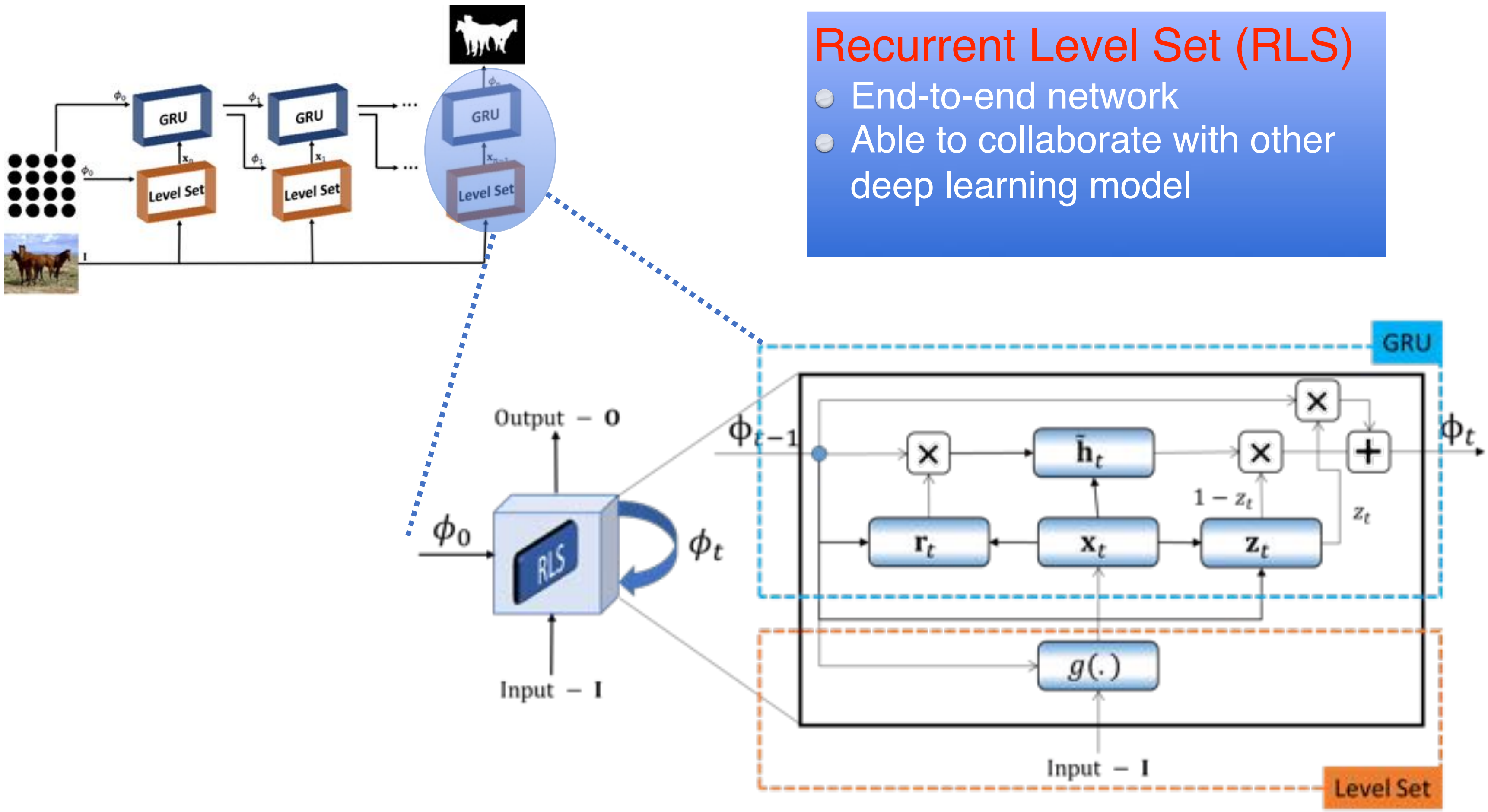
Bring the pure image processing **Level Set** method up to a new level of an **end-to-end learnable framework**


RNN/GRU
Output
Update
V
W
Hidden
U
Input
Level Set (LS)
Time series
Time series
sequence
single image


GRU
GRU
GRU
h_0
h_1
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_{n-1}$
ϕ_0
ϕ_1
ϕ_2
Level Set
Level Set
Level Set

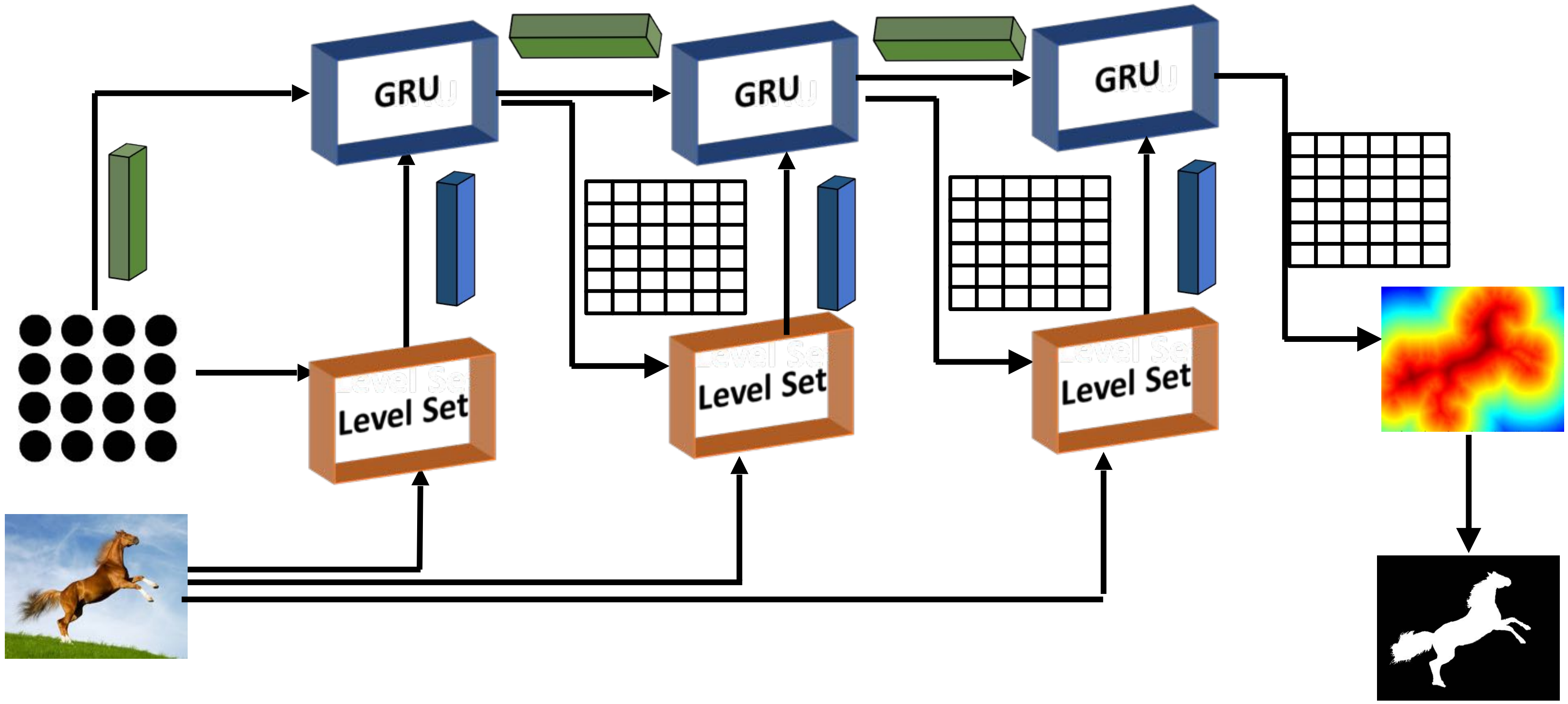


ϕ_0
GRU
ϕ_1
GRU
...
GRU
ϕ_n
$\mathbf{x}$
ϕ_1
$\mathbf{x}_1$
...
$\mathbf{x}_{n-1}$
ϕ_0
Level Set
Level Set
Level Set
I

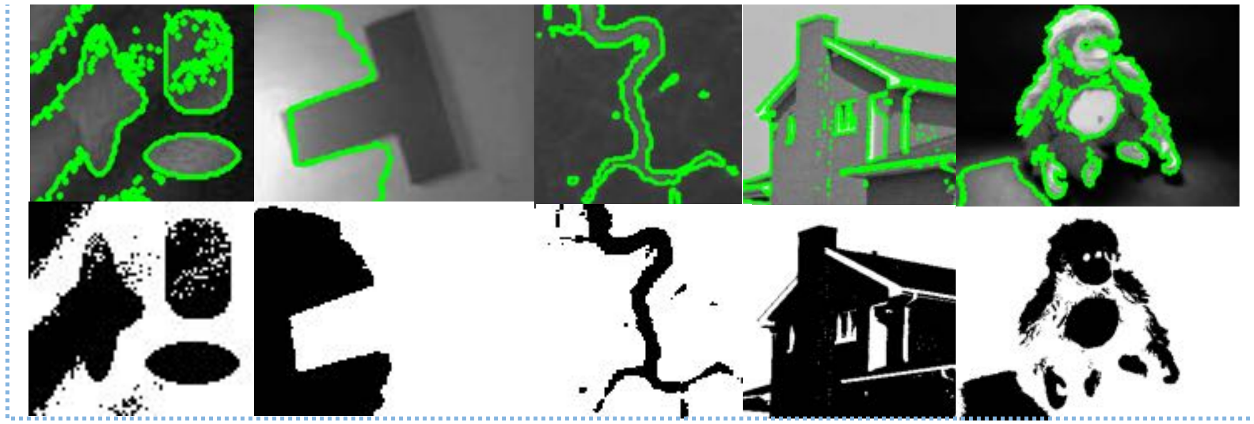


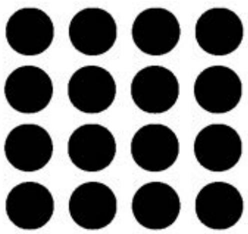
Recurrent Level Set (RLS)

- End-to-end network
- Able to collaborate with other deep learning model

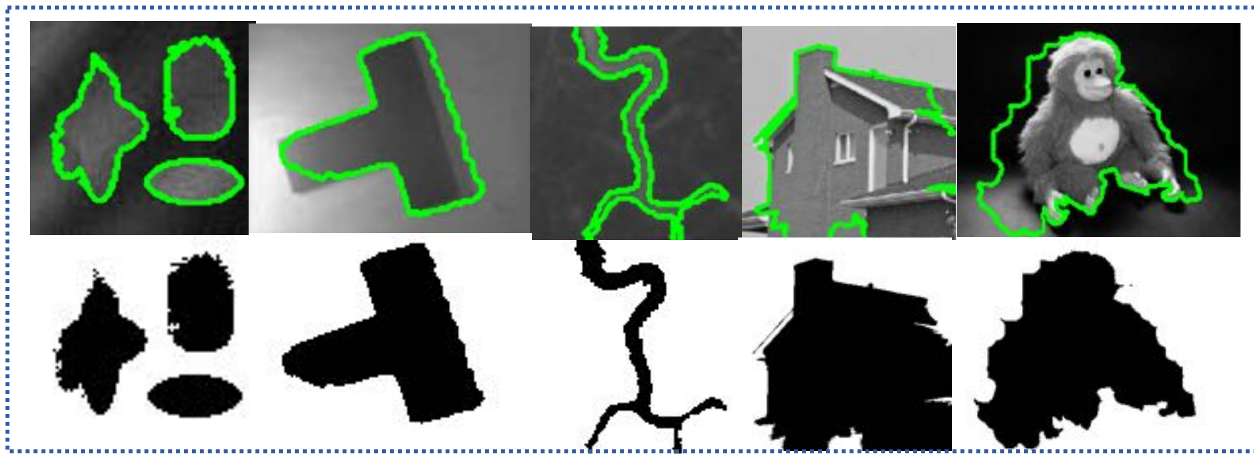

GRU
GRU
GRU
Level Set
Level Set
Level Set

Level Set

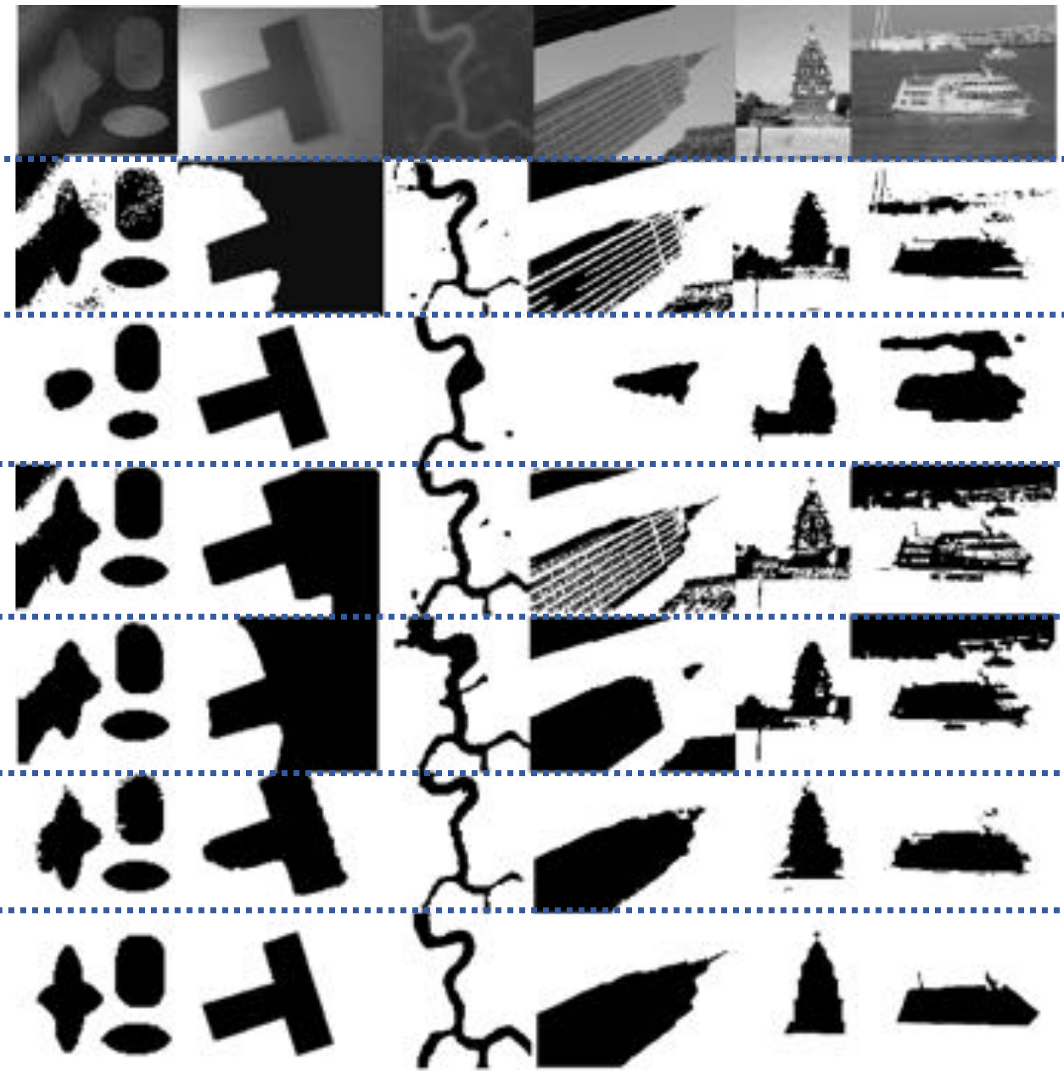




RLS



T. F. Chan and L. A. Vese, “Active contours without edges,” TIP, vol. 10, no. 2, pp. 266–277, Feb. 2001



input images

CV T. F. Chan and L. A. Vese, “Active contours without edges,” TIP, 2001

DRLSE C. Li, et al., “Distance regularized level set evolution and its application to image segmentation,” TIP, 2010

Li C. Li,et al., “A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri,” TIP, 2011

L2S S. Mukherjee and S. Acton, “Region based segmentation in presence of intensity inhomogeneity using legendre polynomials,” SPL, 2015.

RLS

Groundtruth

Level Set

RLS

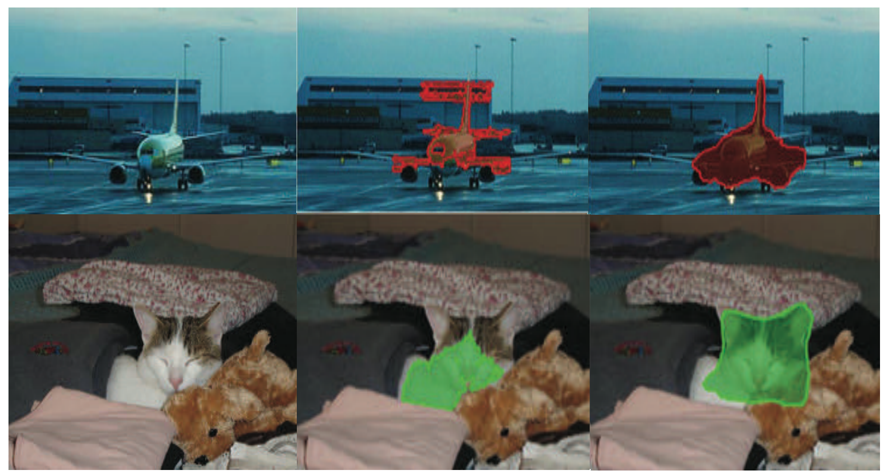
Ground truth

Weizmann: S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in CVPR, June 2007

~~Dis~~advantages

- Computationally ~~expensive~~
- ~~Sensitive~~ to the initialization
- ~~Depends~~ of the number of iterations
- ~~No~~ learning from training data
- ~~Many~~ proper parameters
- ~~Unable~~ to segment images in the wild

~20s	**< 0.1s**
2000 iterations	**5 iterations**
Proper initial contour	**Random initial contour**

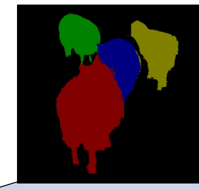


Weizmann

Methods	FM (GT1)	FM (GT2)	Testing Time
CV [52]	88.51	87.51	13.5(s)
DRLSE[141]	80.93	71.76	23.5 (s)
Li et al.[142]	79.65	63.87	20.4 (s)
L2S [143]	81.36	71.03	10.2 (s)
F-ConNet	93.30	93.26	**0.001** (s)
RLS	**99.16**	**99.17**	0.008 (s)

Weizmann: S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in CVPR, June 2007

Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation



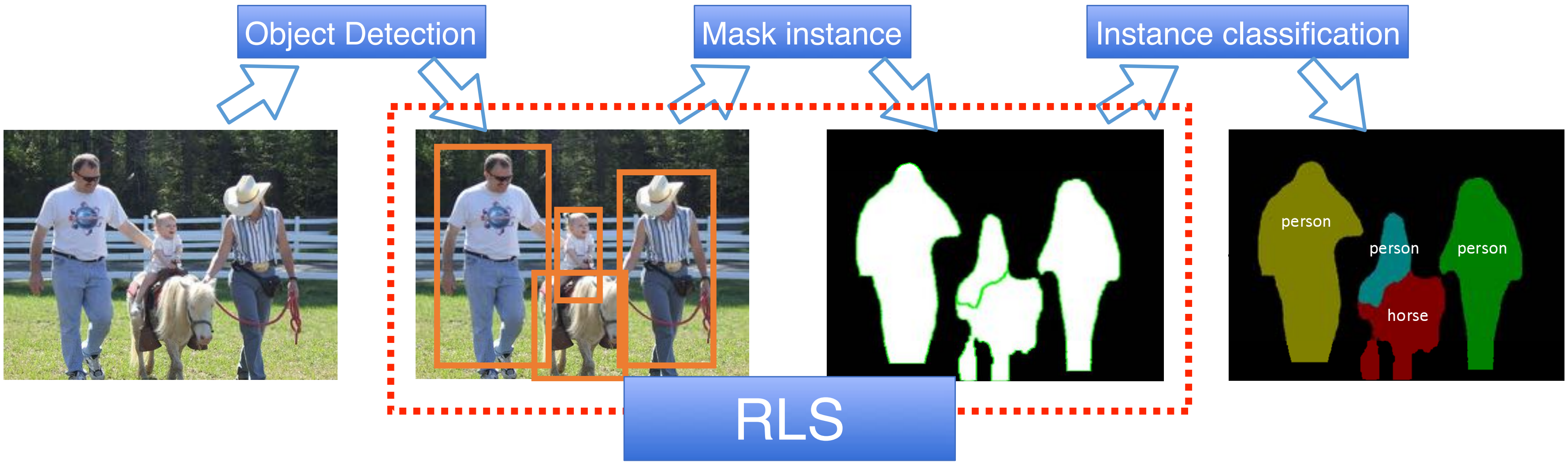
- Level Set
- DNNs
- Coherency between LS and RNNs/GRUs

- Recurrent Level Set Networks (RLS) for Image Segmentation

Contextual Recurrent Level Set Networks (CRLS) for Semantic Instance Segmentation

Recurrent Level Set (RLS)

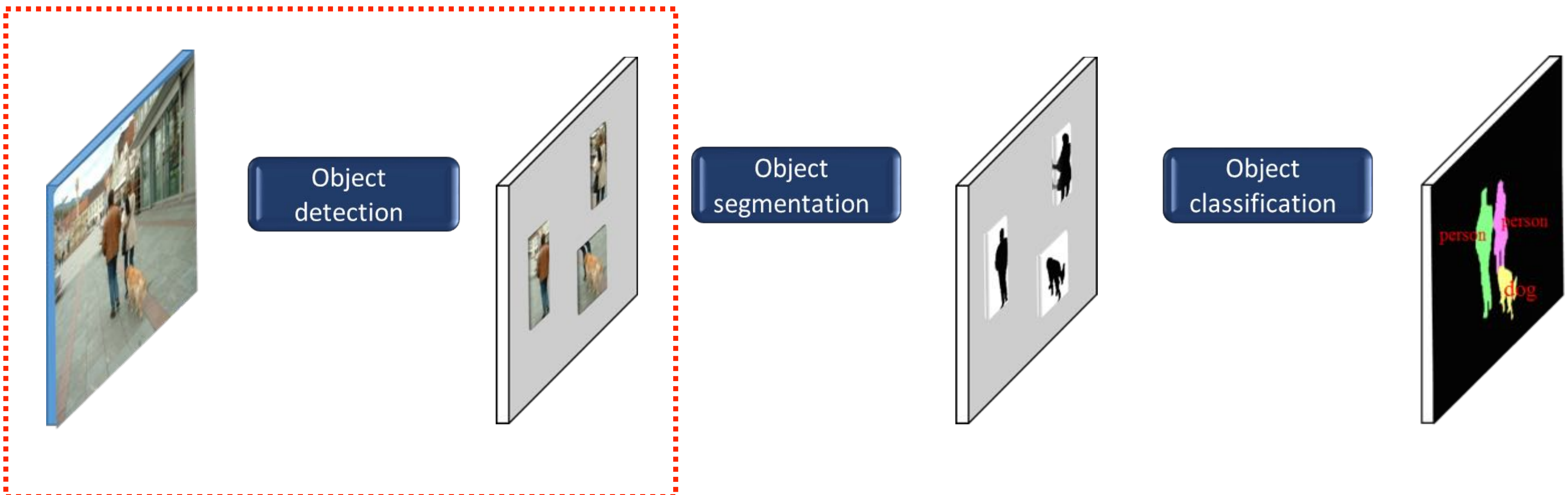
- Boost Level Set to the new level of learnable framework
- End-to-end network
- Able to collaborate with other deep learning model

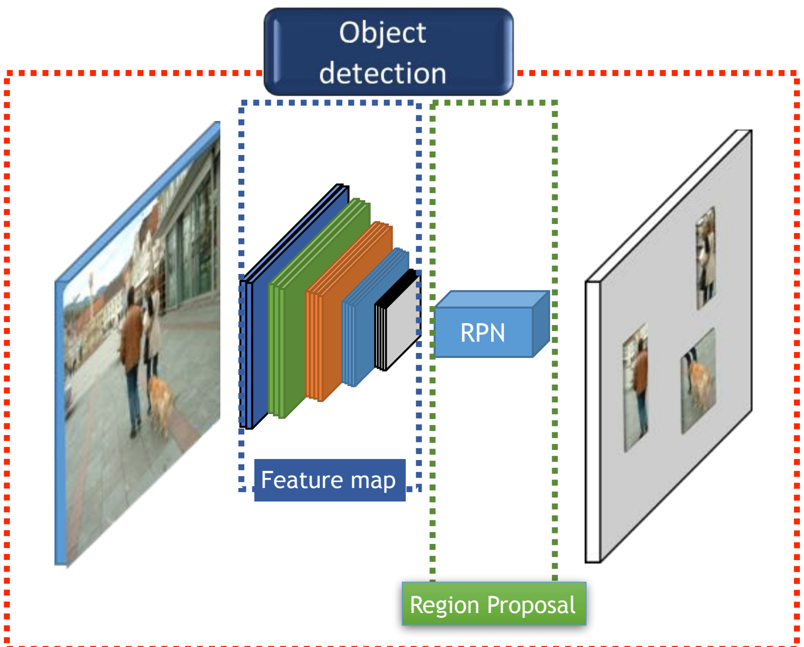


Recurrent Level Set (RLS)

- Boost Level Set to the new level of learnable framework
- End-to-end network
- Able to collaborate with other deep learning model

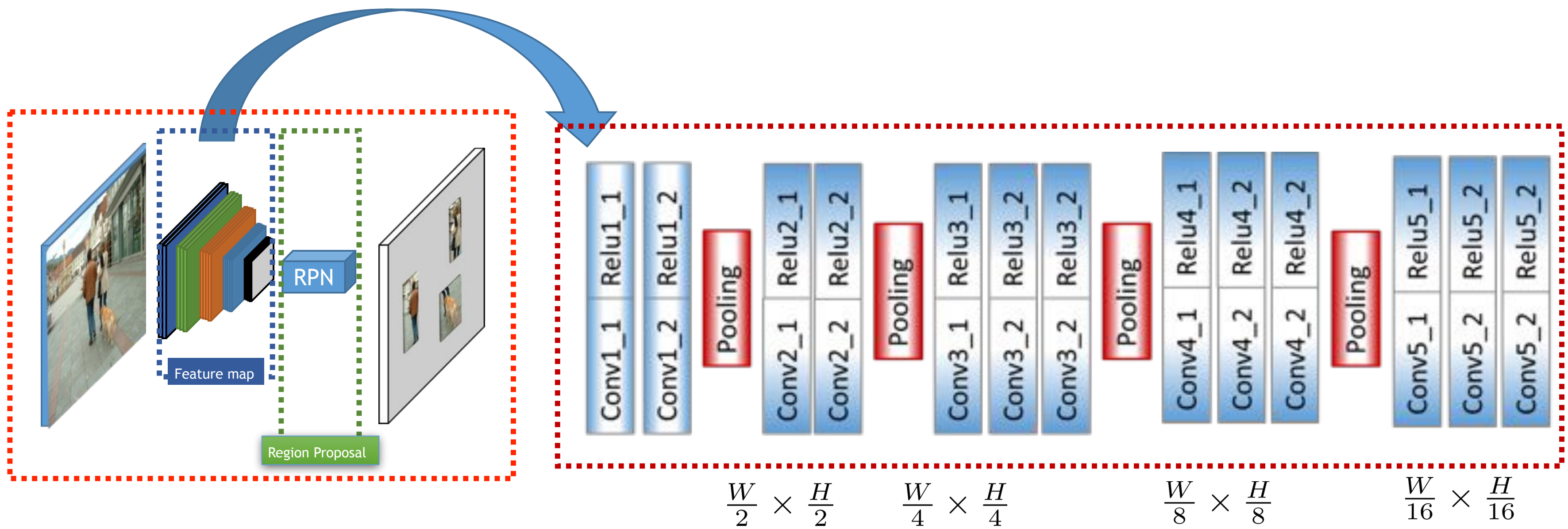
Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation

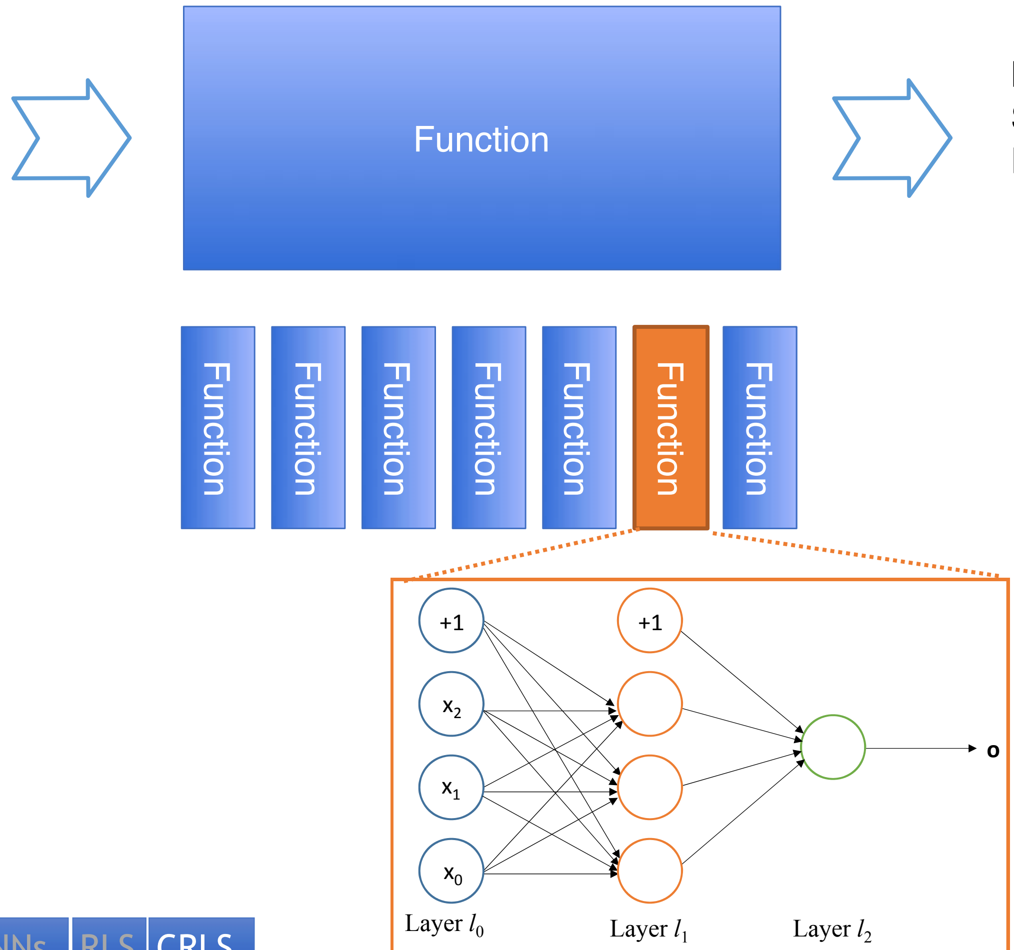




Input
- Image

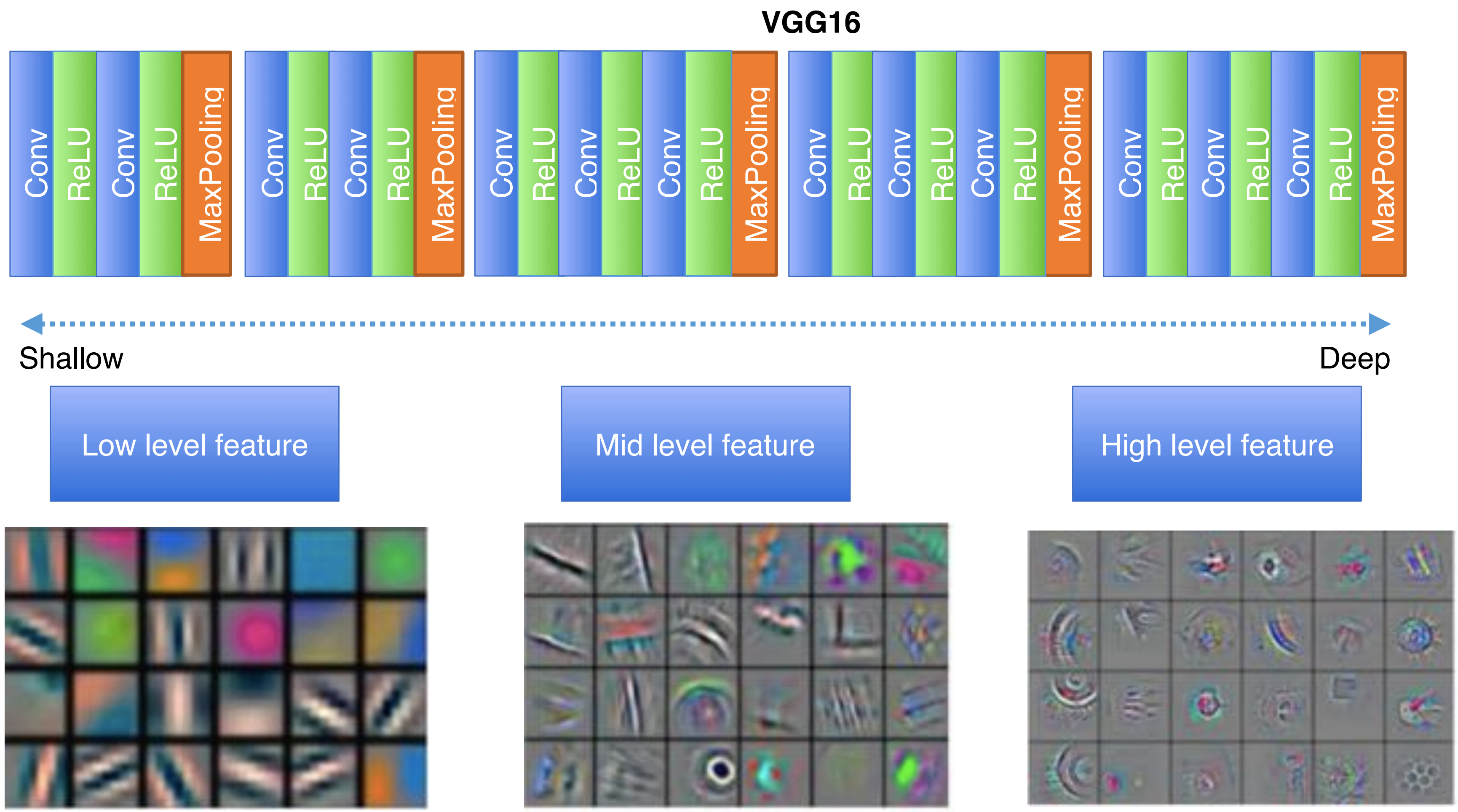
Output
- Object bounding box

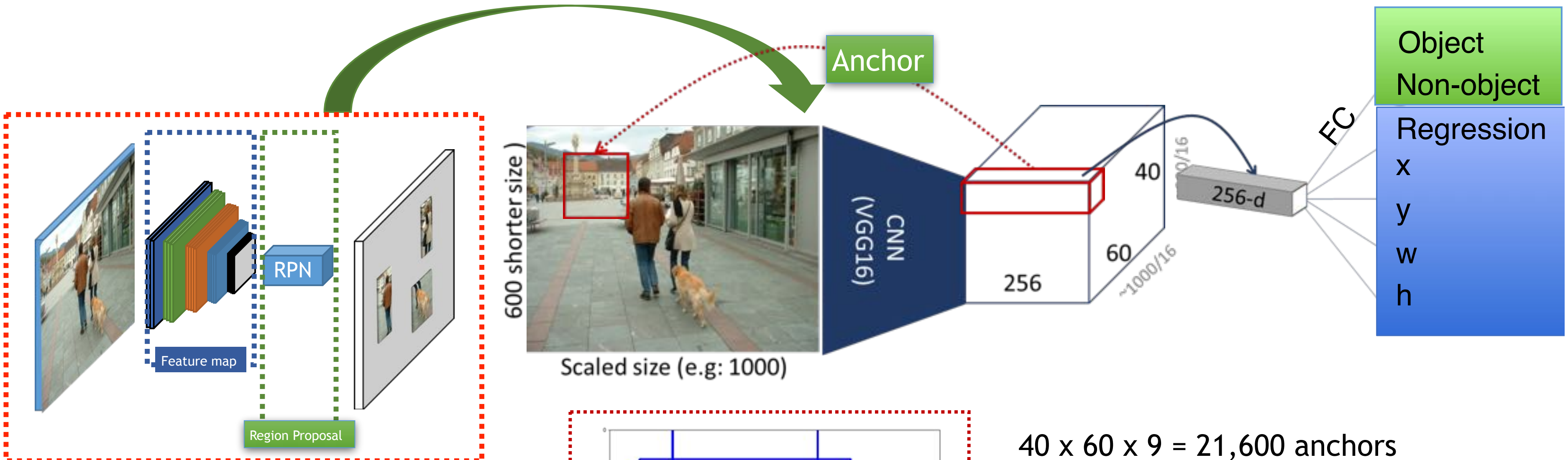

Feature map
RPN
Region Proposal
Conv1_1 Relu1_1
Conv1_2 Relu1_2
Pooling
Conv2_1 Relu2_1
Conv2_2 Relu2_2
Pooling
Conv3_1 Relu3_1
Conv3_2 Relu3_2
Conv3_2 Relu3_2
Pooling
Conv4_1 Relu4_1
Conv4_2 Relu4_2
Conv4_2 Relu4_2
Pooling
Conv5_1 Relu5_1
Conv5_2 Relu5_2
Conv5_2 Relu5_2
$\frac{W}{2} \times \frac{H}{2}$
$\frac{W}{4} \times \frac{H}{4}$
$\frac{W}{8} \times \frac{H}{8}$
$\frac{W}{16} \times \frac{H}{16}$

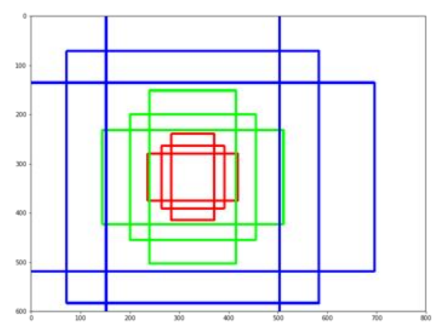


Beach?
Sand?
Person?

Neural networks


VGG16
Conv
ReLU
Conv
ReLU
MaxPooling
Conv
ReLU
Conv
ReLU
MaxPooling
Conv
ReLU
Conv
ReLU
Conv
ReLU
MaxPooling
Conv
ReLU
Conv
ReLU
Conv
ReLU
MaxPooling
Conv
ReLU
Conv
ReLU
Conv
ReLU
MaxPooling
Shallow
Deep
Low level feature
Mid level feature
High level feature





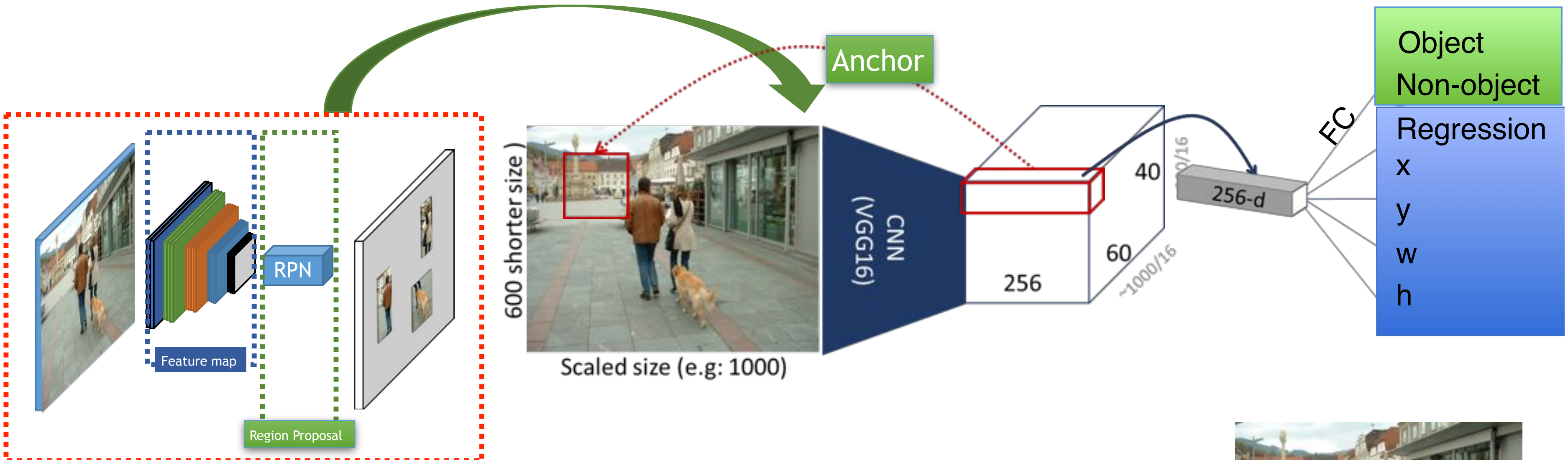
K anchor boxes for each location

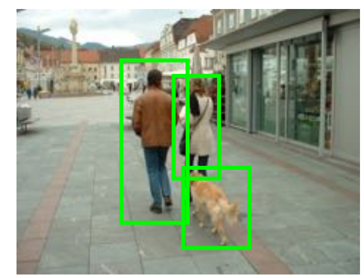
40 x 60 x 9 = 21,600 anchors

Ignore cross boundary anchors:~6,000

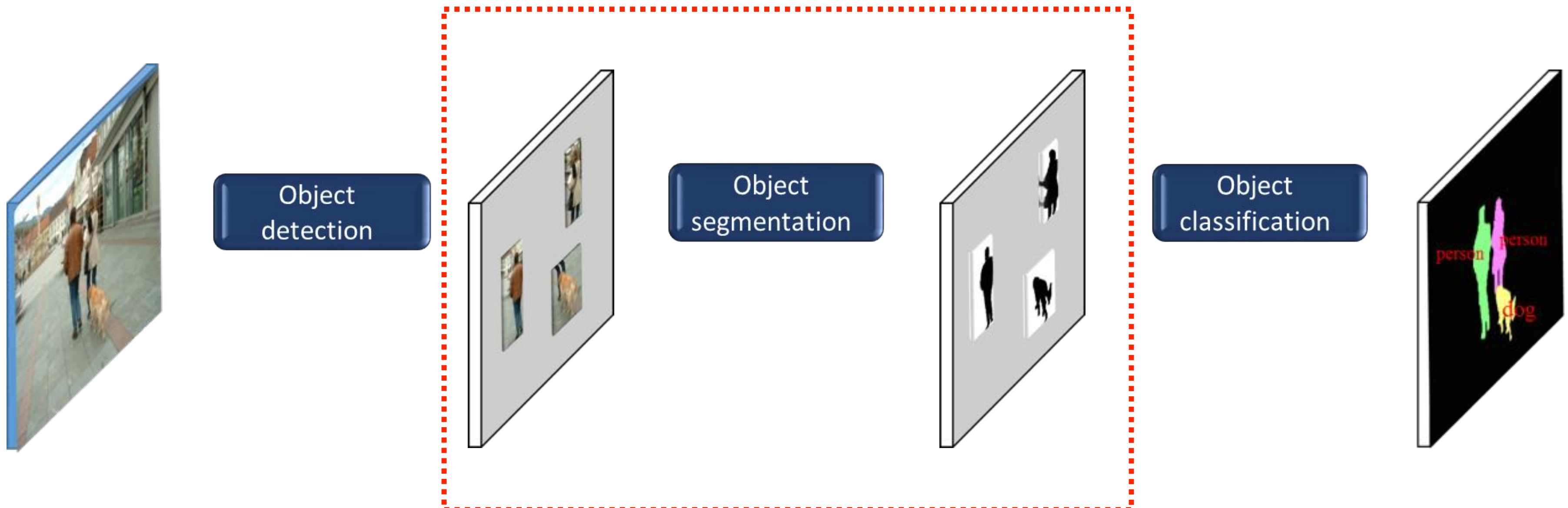
Non maximum suppression (NMS): ~2,000

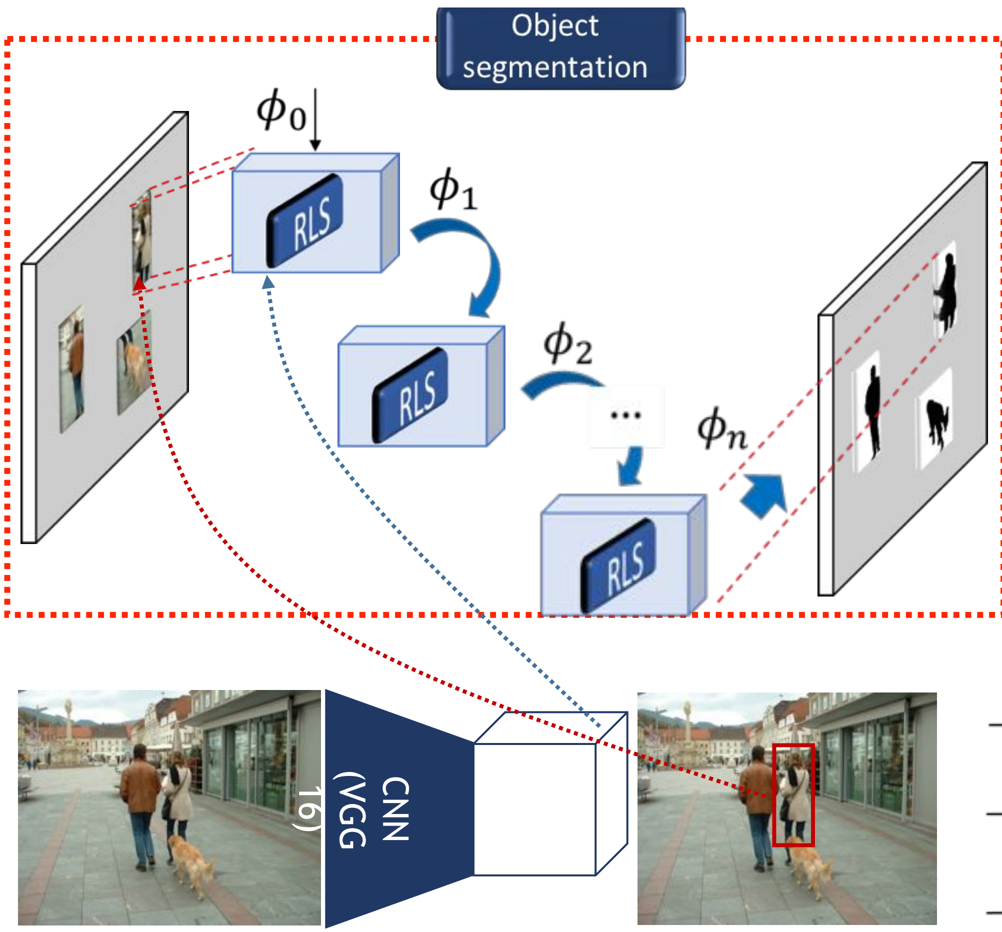
$$IoU = \frac{a \cap GT}{a \cup GT} \left\{ \begin{array}{ll} > 0.7 & object \\ < 0.3 & nonobject \end{array} \right\}$$


Anchor
Feature map
RPN
Region Proposal
600 shorter size)
Scaled size (e.g: 1000)
CNN
(VGG16)
40
60
256
~1000/16
256-d
FC
Object
Non-object
Regression
x
y
w
h

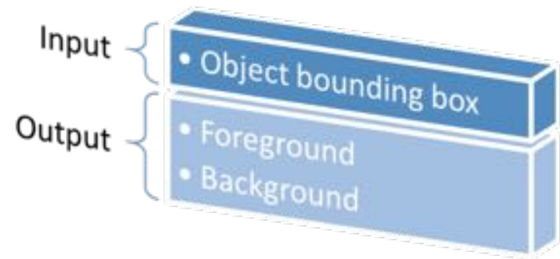


Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation

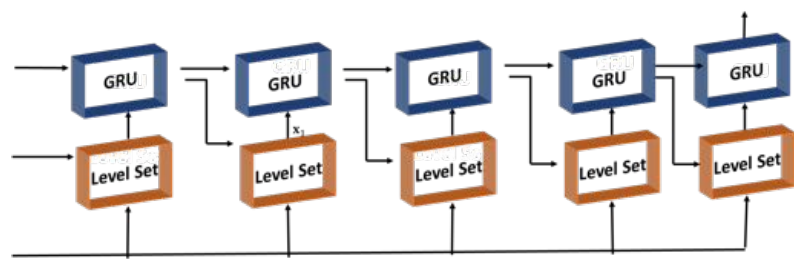




Object segmentation
ϕ_0
ϕ_1
ϕ_2
ϕ_n
RLS
RLS
RLS
...
CNN (VGG 16)

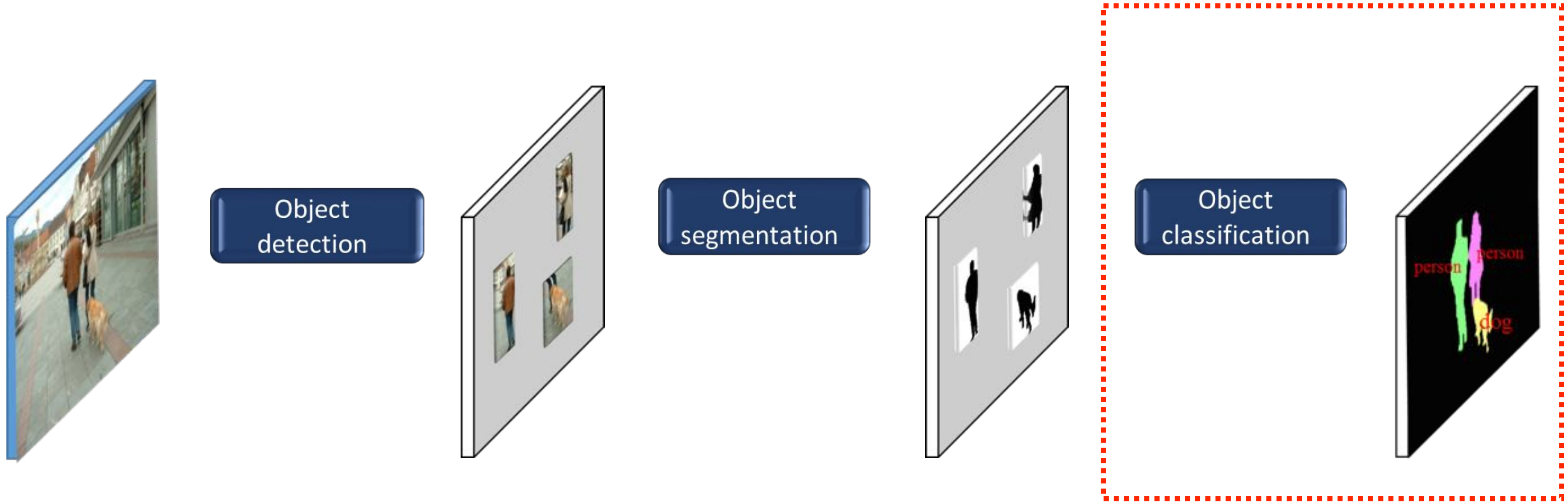


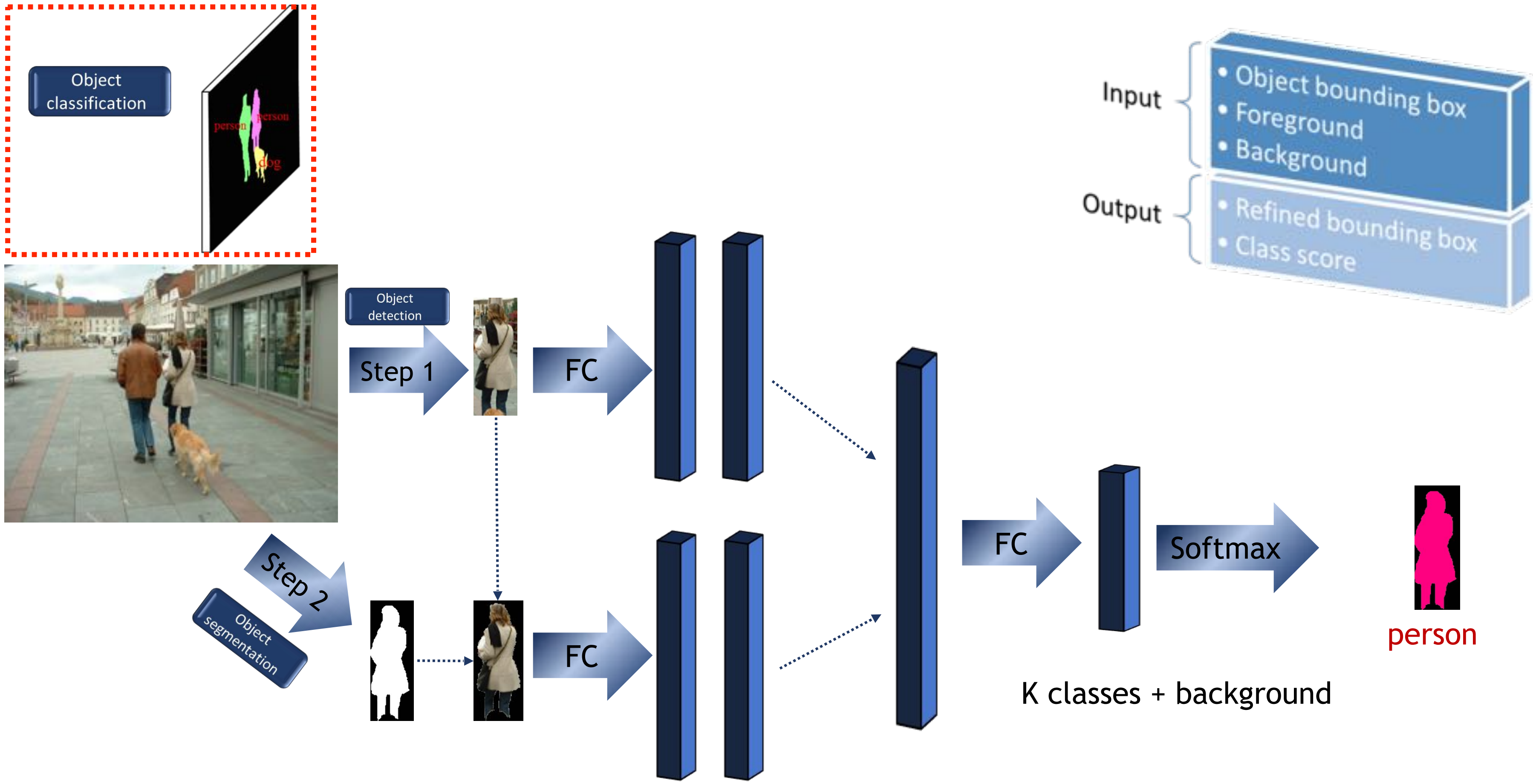
Input
Output
• Object bounding box
• Foreground
• Background



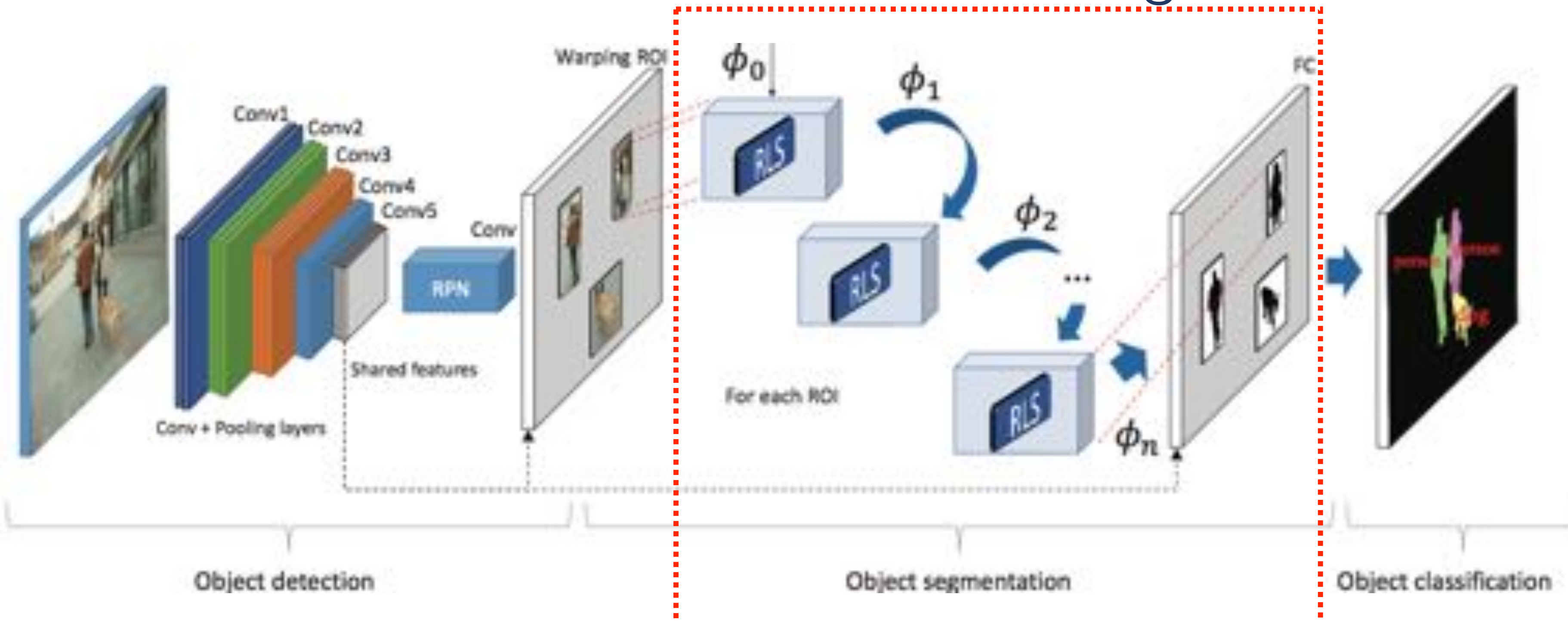
GRU
GRU
GRU
GRU
GRU
Level Set
Level Set
Level Set
Level Set
Level Set

Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation



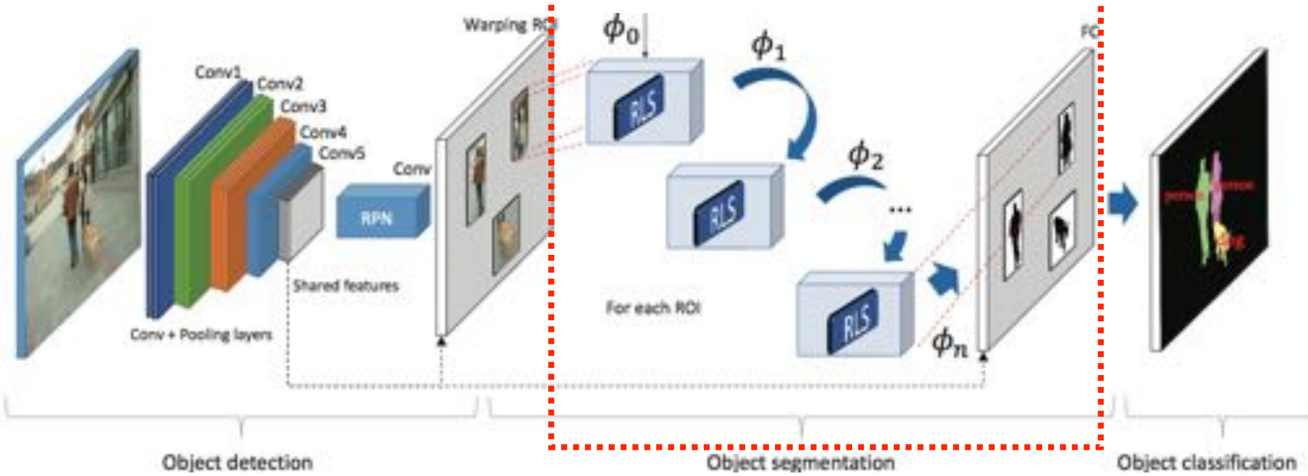

Object classification
person
person
dog
Input
Object bounding box
Foreground
Background
Output
Refined bounding box
Class score
Object detection
Step 1
FC
Step 2
Object segmentation
FC
FC
Softmax
person
K classes + background

Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation



Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation

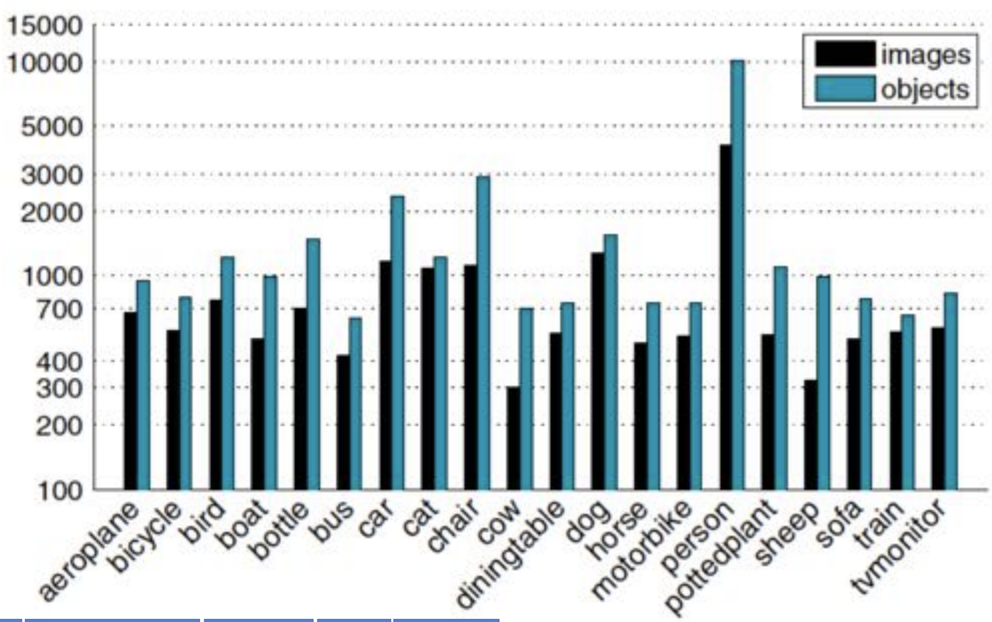
- End-to-end framework
- Trained from scratch
- Is able to work on other serial problem (not limited Chan-Vese)
- Is able to collaborate with other network model (not limited to VGG)

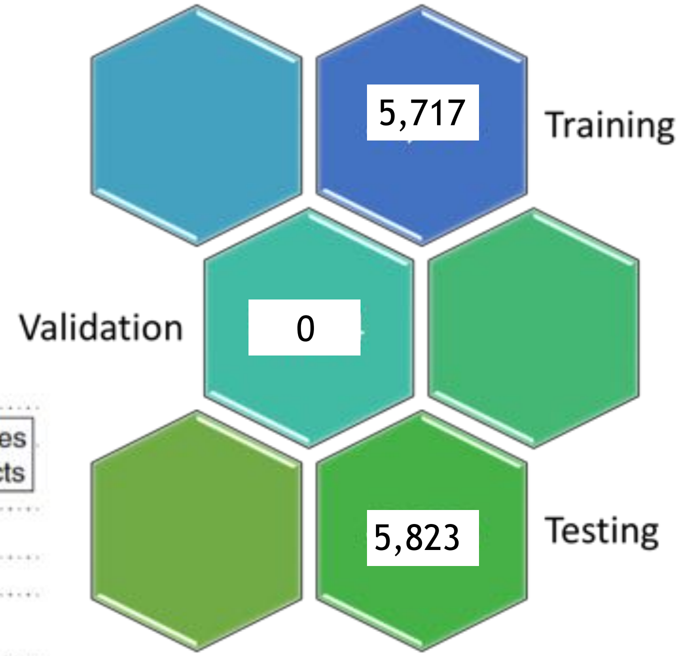


Experimental Results

PASCAL VOC 2012

20 categories

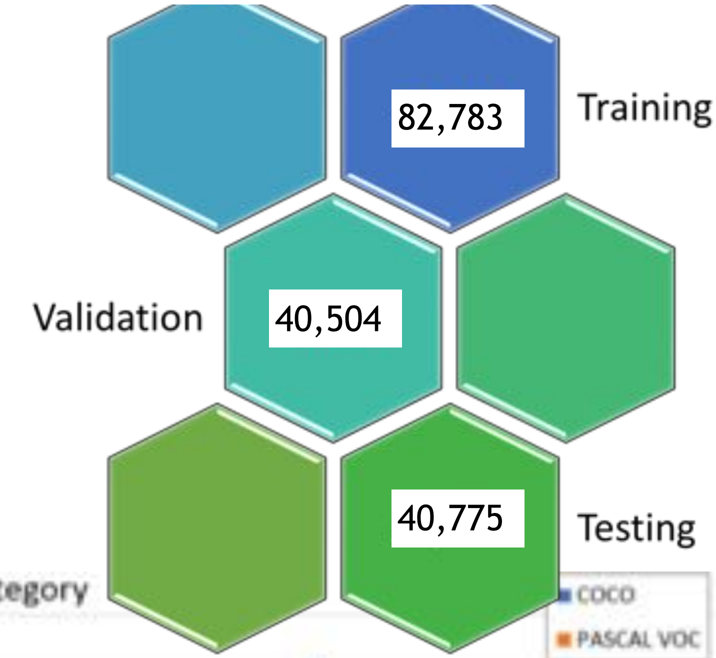


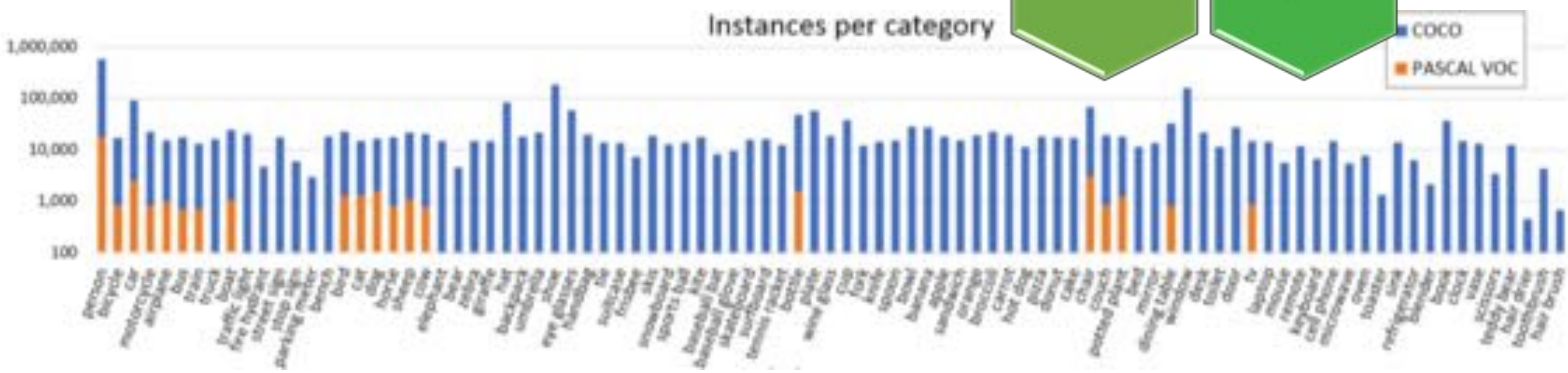


Experimental Results

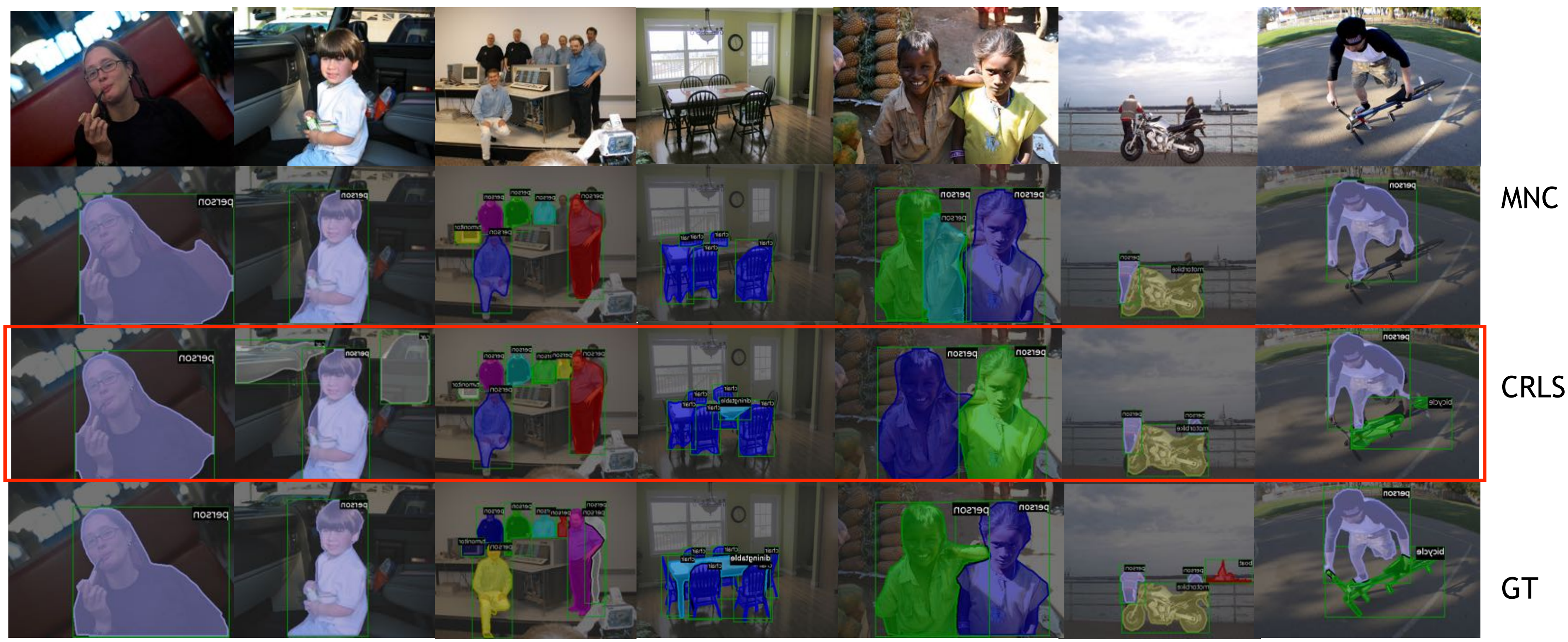
MS COCO 2014

80 categories





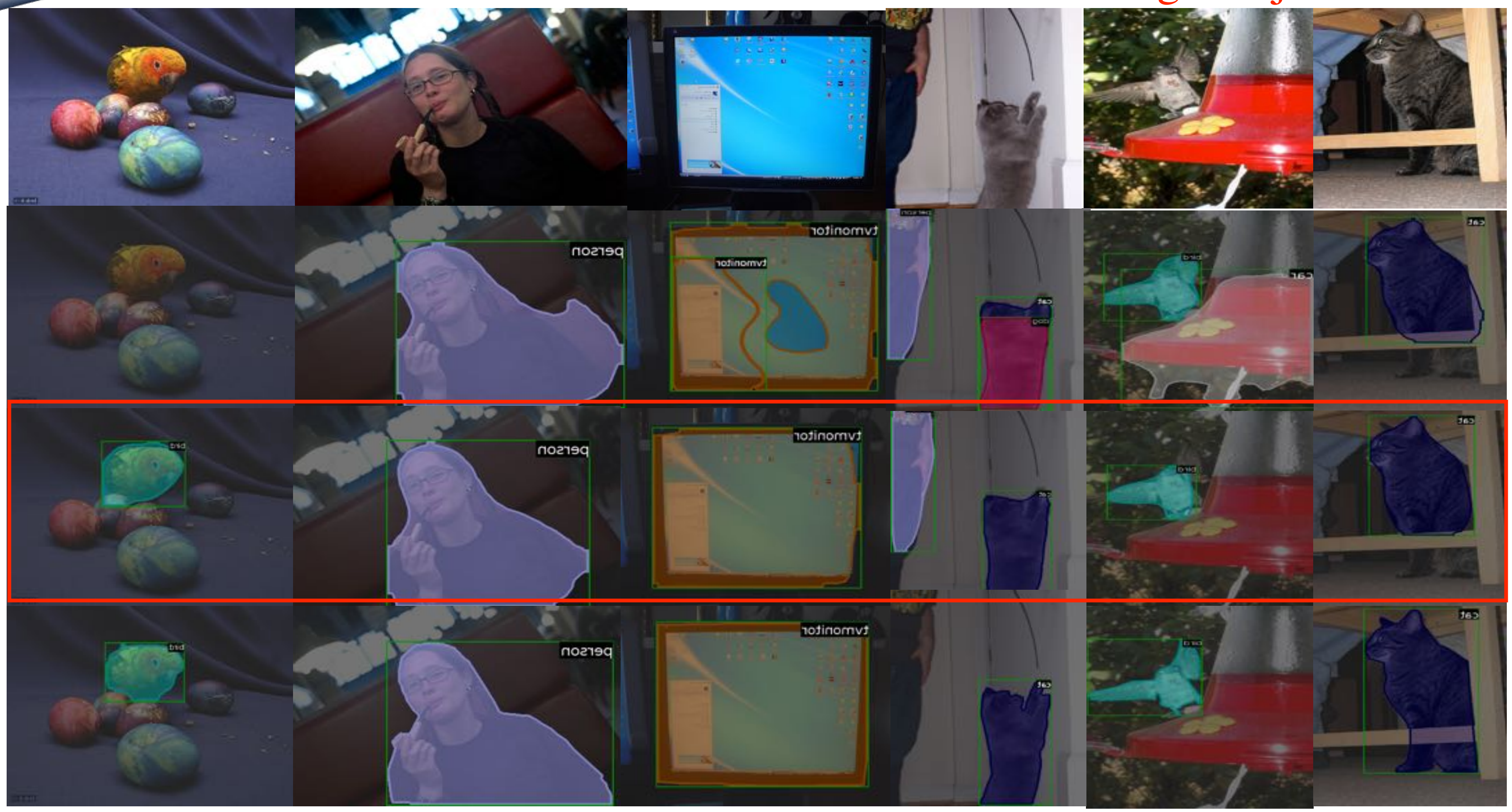
PASCAL VOC 2012



MNC: Dai, Jifeng and He, Kaiming and Sun, Jian, Instance-aware semantic segmentation via multi-task network cascades, In CVPR, pp. 3150-3158, 2016

CRLS: T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , Transactions on Image Processing, Vol. 27, No. 5, May 2018, pp. 2393–2407

PASCAL VOC 2012 - single object



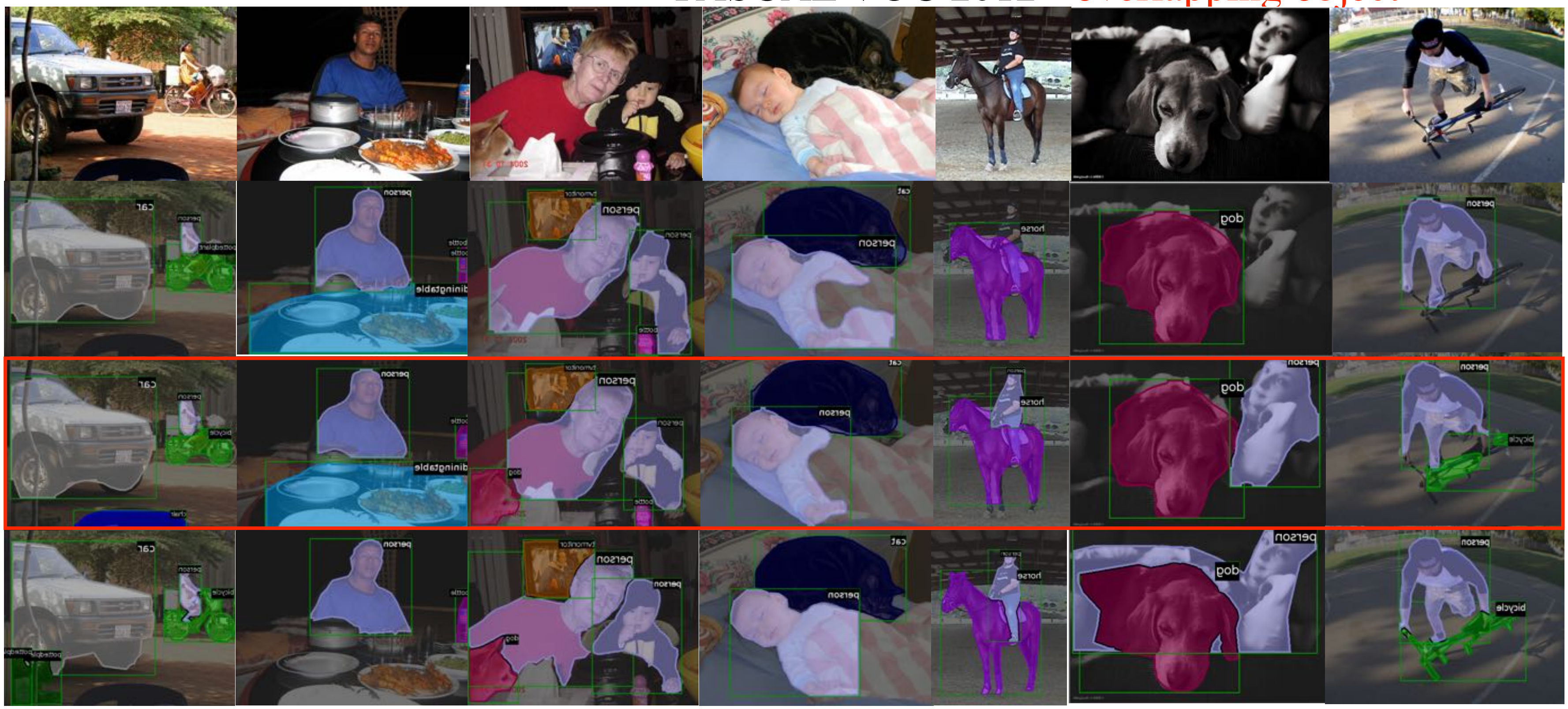
MNC

CRLS

GT

MNC: Dai, Jifeng and He, Kaiming and Sun, Jian, Instance-aware semantic segmentation via multi-task network cascades, In CVPR, pp. 3150-3158, 2016
CRLS: T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , Transactions on Image Processing, Vol. 27, No. 5, May 2018, pp. 2393–2407

PASCAL VOC 2012 - overlapping object



MNC

CRLS

GT

MNC: Dai, Jifeng and He, Kaiming and Sun, Jian, Instance-aware semantic segmentation via multi-task network cascades, In CVPR, pp. 3150-3158, 2016
CRLS: T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , Transactions on Image Processing, Vol. 27, No. 5, May 2018, pp. 2393–2407

PASCAL VOC 2012 - multiple instance object

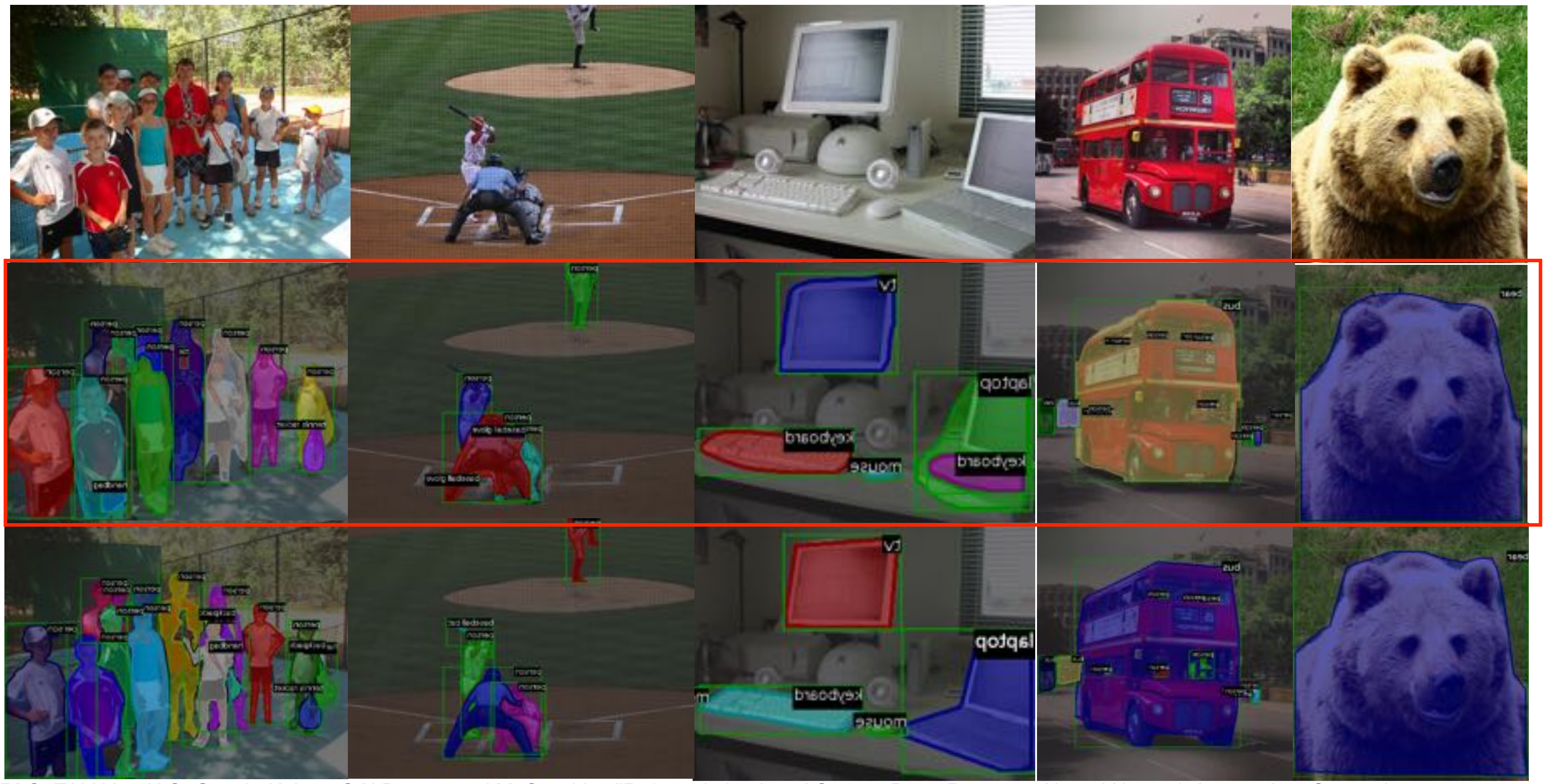
MNC: Dai, Jifeng and He, Kaiming and Sun, Jian, Instance-aware semantic segmentation via multi-task network cascades, In CVPR, pp. 3150-3158, 2016
CRLS: T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , Transactions on Image Processing, Vol. 27, No. 5, May 2018, pp. 2393–2407

Methods	mAP^r@.5	mAP^r@.7	Time (s)
SDS (AlexNet)	49.7%	25.3%	48
Hypercolumn	60.0%	40.4%	>80
CFM	60.7%	39.6%	32
MNC	63.5%	41.5%	**0.36**
CRLS (Ours)	**66.7%**	**44.6%**	0.54

SDS: B. Hariharan, et al., "Simultaneous detection and segmentation," ECCV, 2014
Hypercolumn B. Hariharan, et al., "Hyper- columns for object segmentation and fine-grained localization," CVPR2015
CFM: J. Dai, et al., "Convolutional feature masking for joint object and stuff segmentation," CVPR, IEEE, 2015
MNC: J. Dai, et al. Instance-aware semantic segmentation via multi-task network cascades, CVPR2016

CRLS: T.H.N Le, et al. "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , TIP 2018

MS COCO 2014



CRLS

GT

CRLS: T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , Transactions on Image Processing, Vol. 27, No. 5, May 2018, pp. 2393–2407

MS COCO 2014 - multiple instance object

CRLS: T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , Transactions on Image Processing, Vol. 27, No. 5, May 2018, pp. 2393–2407

Methods	mAP^r@[.5:.95]	mAP^r@.5
MNC	19.5%	39.7%
CRLS (Ours)	**20.5%**	**40.1%**

MNC: J. Dai, Jifeng et al., “Instance-aware semantic segmentation via multi-task network cascades”, In CVPR, 2016

CRLS: T.H.N Le, et al. "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , TIP 2018

Semantic Instance Segmentation

Contextual Recurrent Level Set (CRLS) Networks for Semantic

T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , **Transactions on Image Processing**, Vol. 27, No. 5, 2018, pp. 2393–2407

T.H.N. Le , R. Gummadi, M. Savvides "Deep Recurrent Level Set for Segmenting Brain Tumors", submitted to **MICCAI 2018** (accepted)

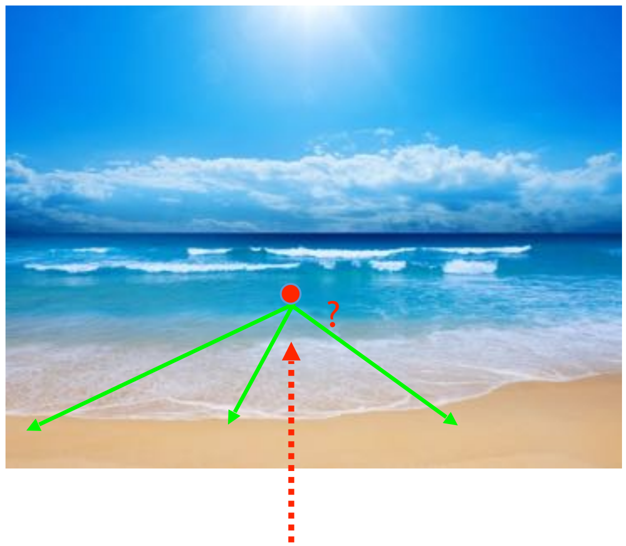
Scene Understanding

Contextual Recurrent Residual Networks (CRRN) for Scene Labeling

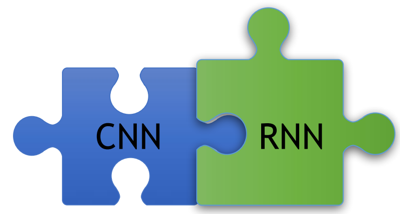
- Motivation
- Contextual Recurrent Residual Networks (CRRN) for Scene Labeling

Motivation | CRRN

Long-range context V.S. short-range context

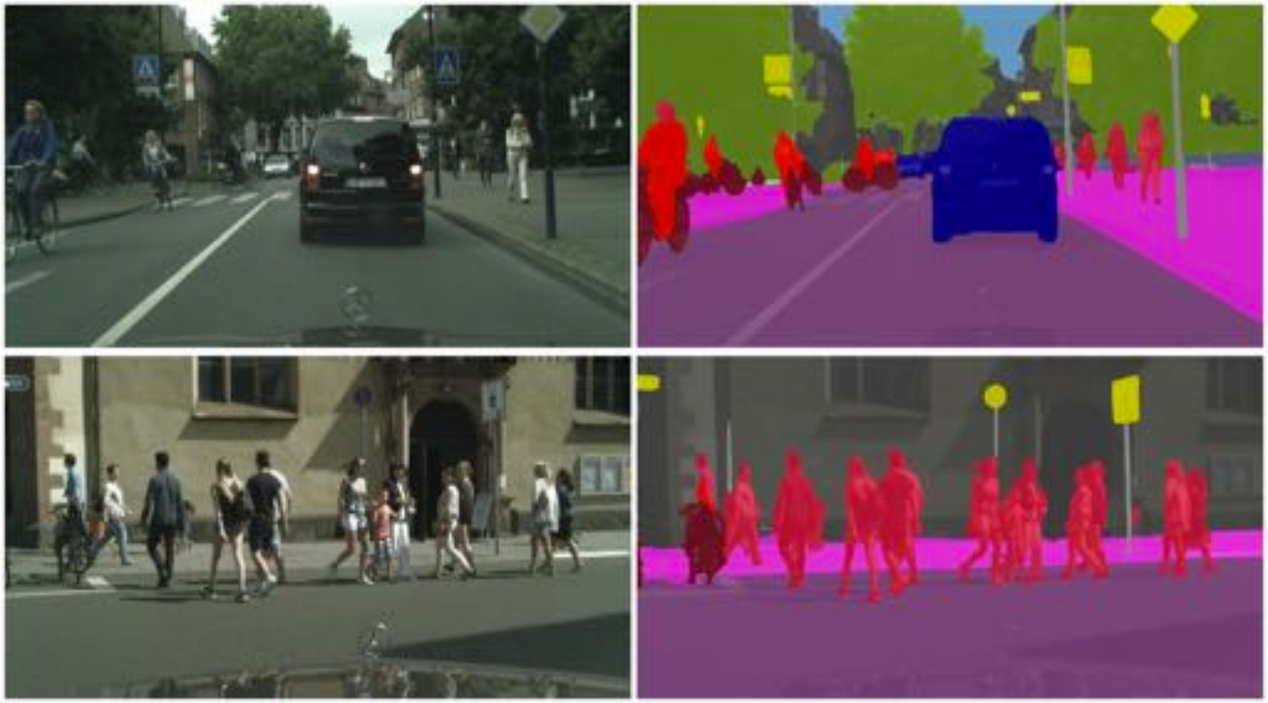


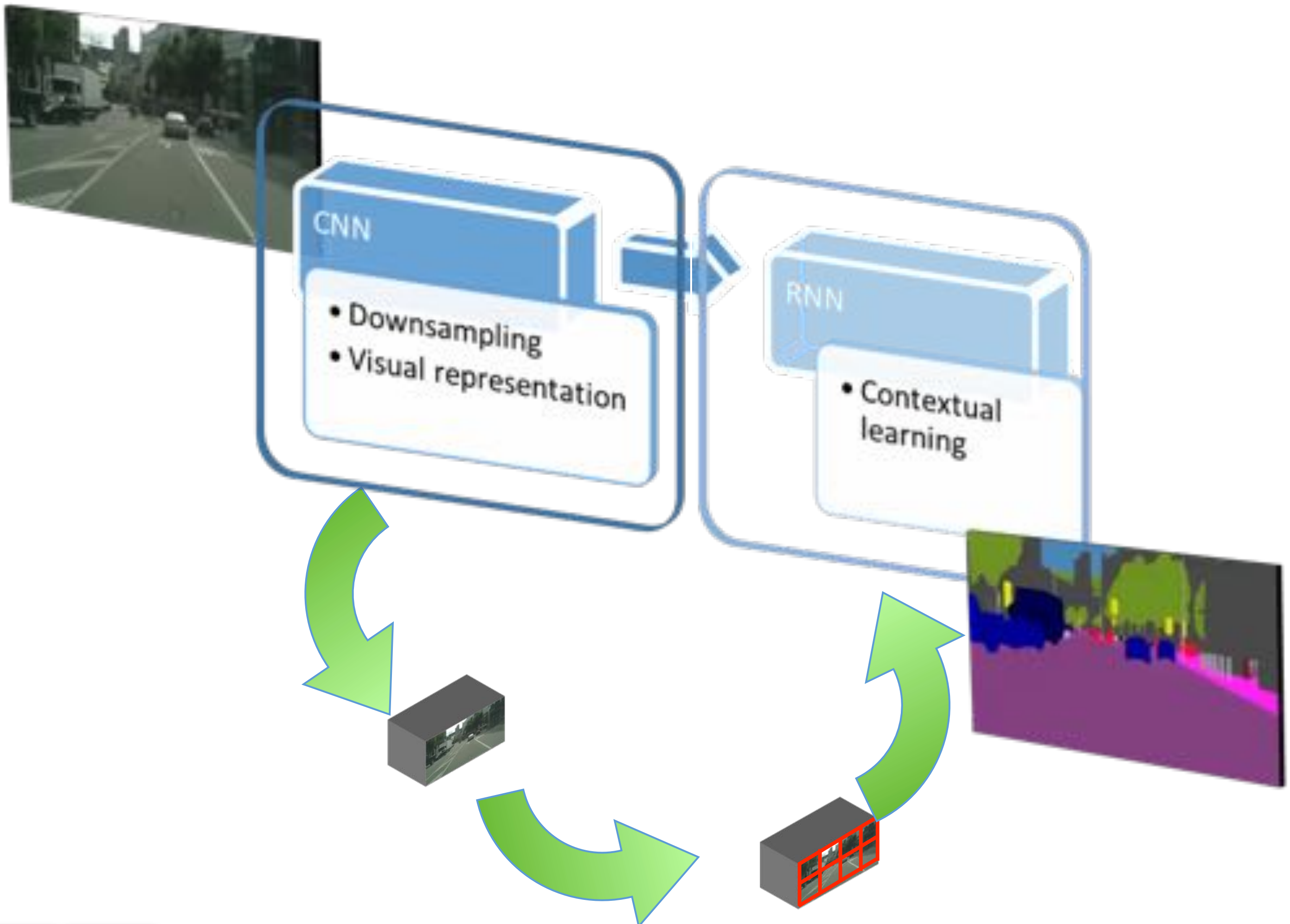
Sea ? Sky

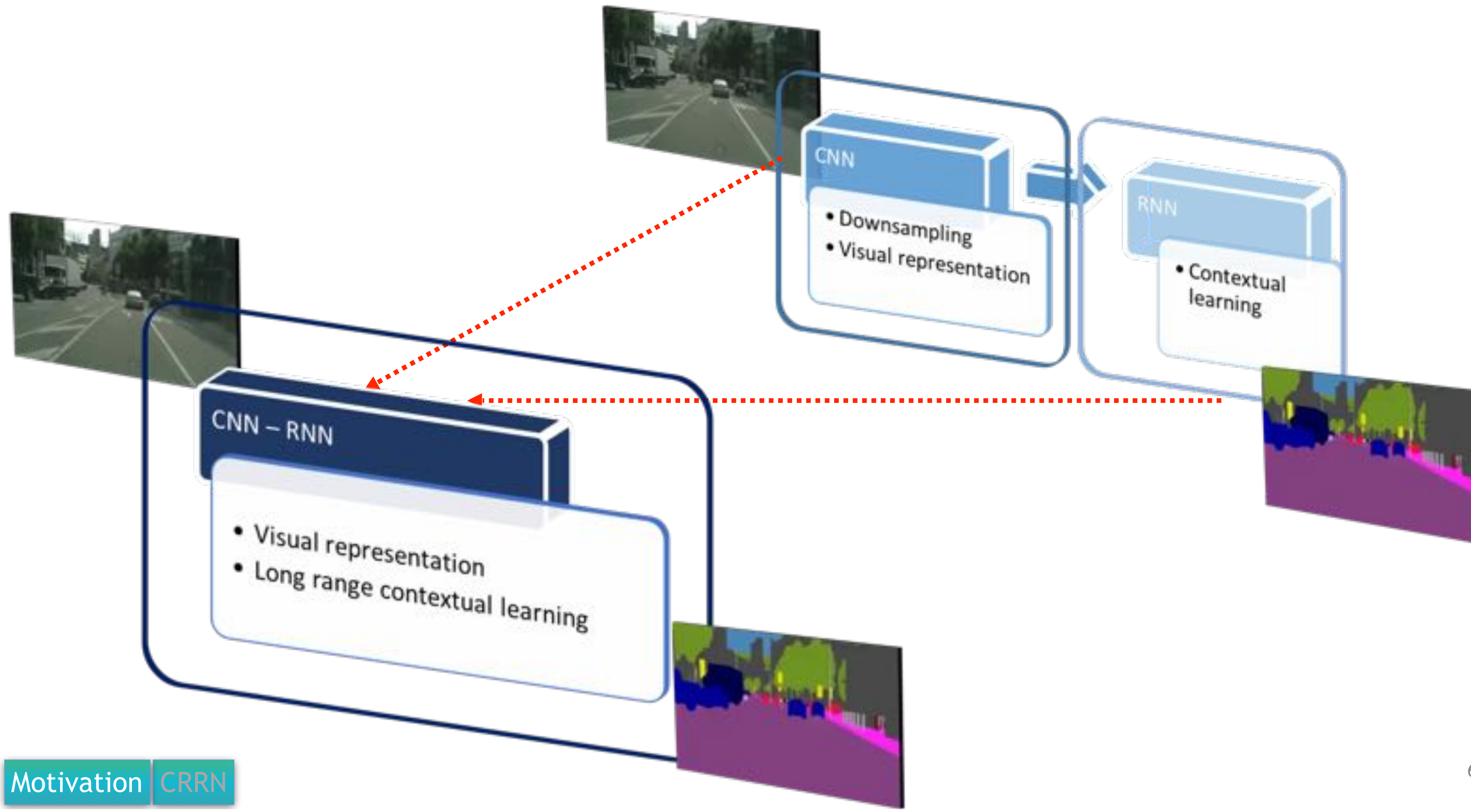


Scene Understanding

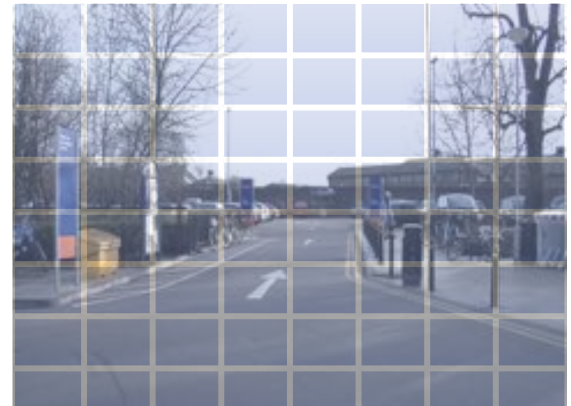
predict the semantic class of the individual pixels (semantic segmentation)

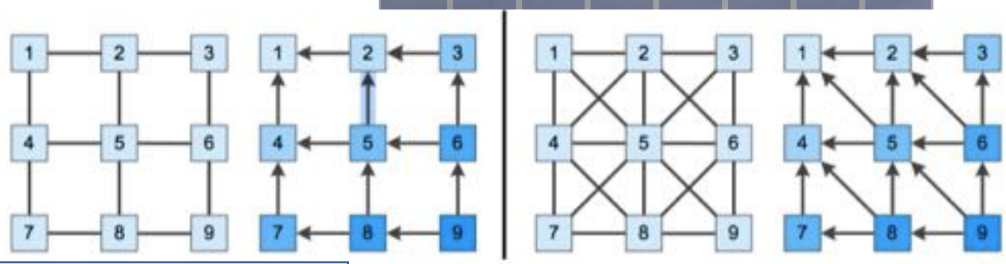



CNN
• Downsampling
• Visual representation
RNN
• Contextual learning

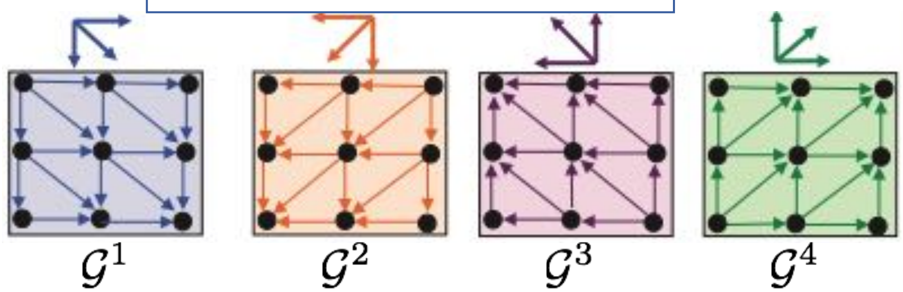


CNN
• Downsampling
• Visual representation
RNN
• Contextual learning
CNN – RNN
• Visual representation
• Long range contextual learning





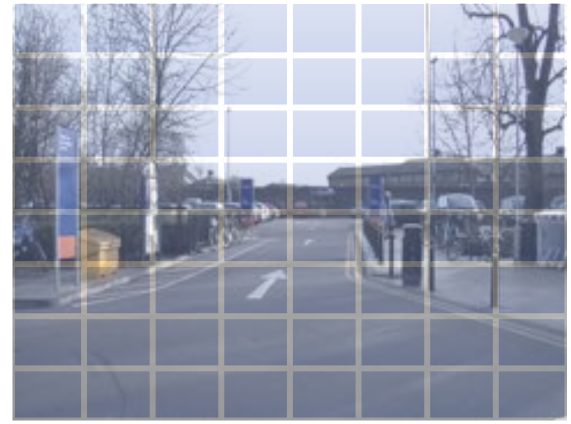
southeast, southwest, northwest and northeast



$$\mathcal{V} = \{v_i\}_{i=1,2,..,N}$$

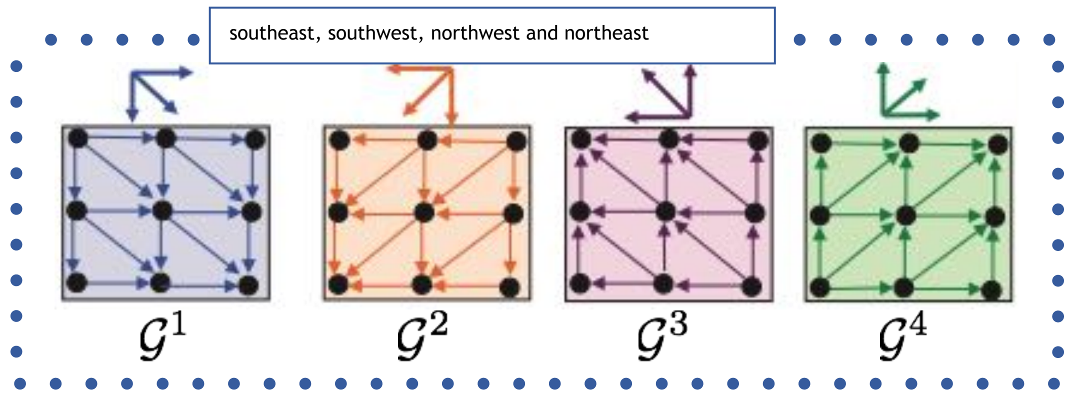
$$\mathcal{E} = \{e_{ij}\}$$

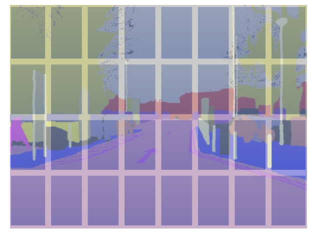
$$\mathcal{G}^d = \{\mathcal{V}, \mathcal{E}\}|_{d=1}^{4}$$



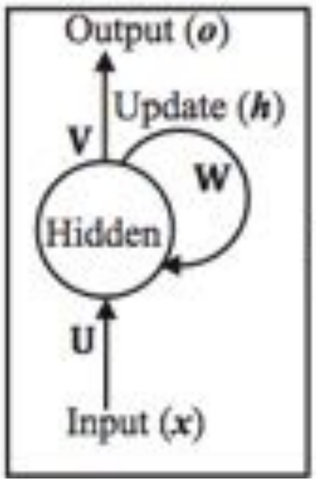
$\mathcal{A}_{\mathcal{G}^d(v_i)}$ predecessor set of v_i

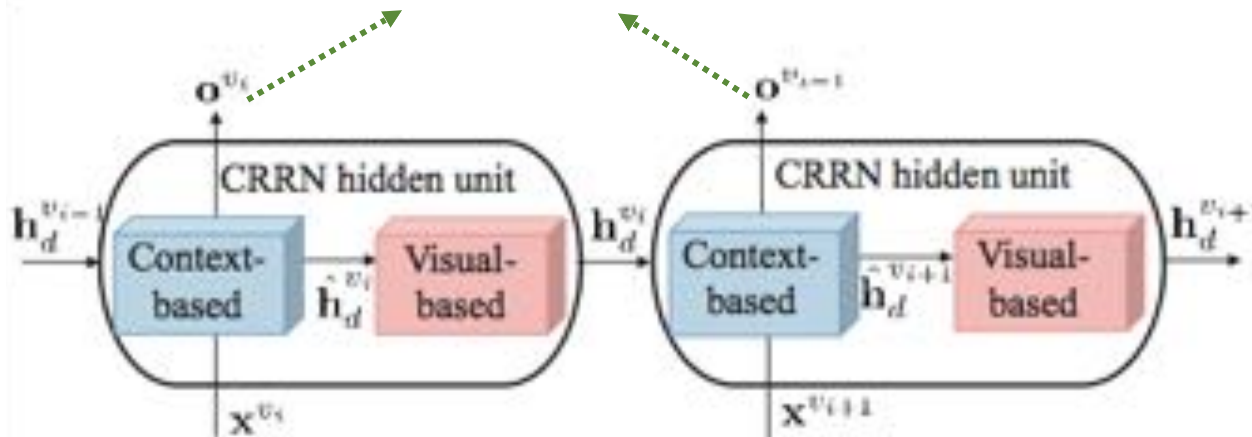
$\mathcal{S}_{\mathcal{G}^d(v_i)}$ successor set of v_i





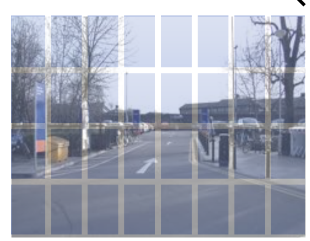
Segmentation blocks





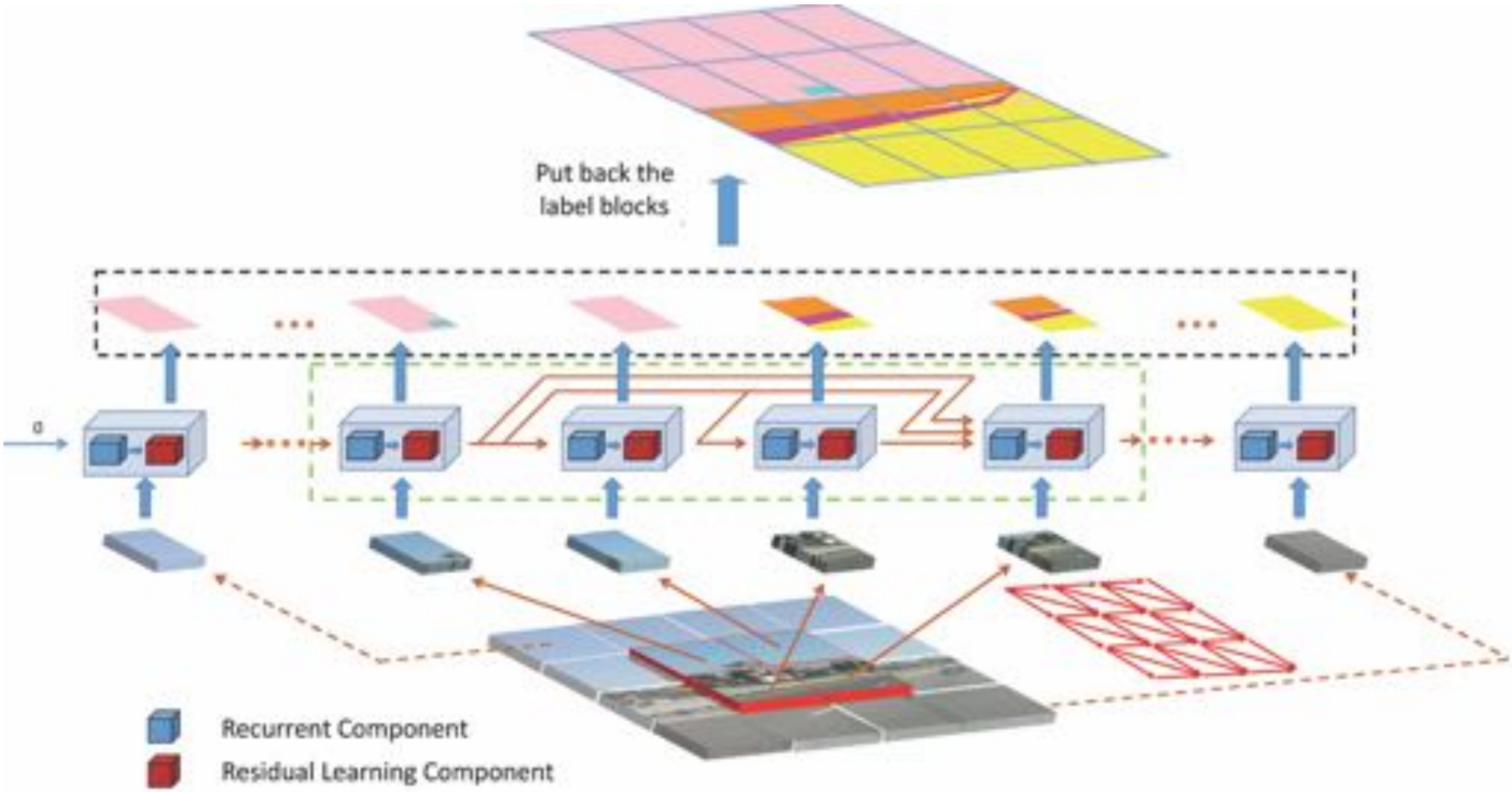
patches

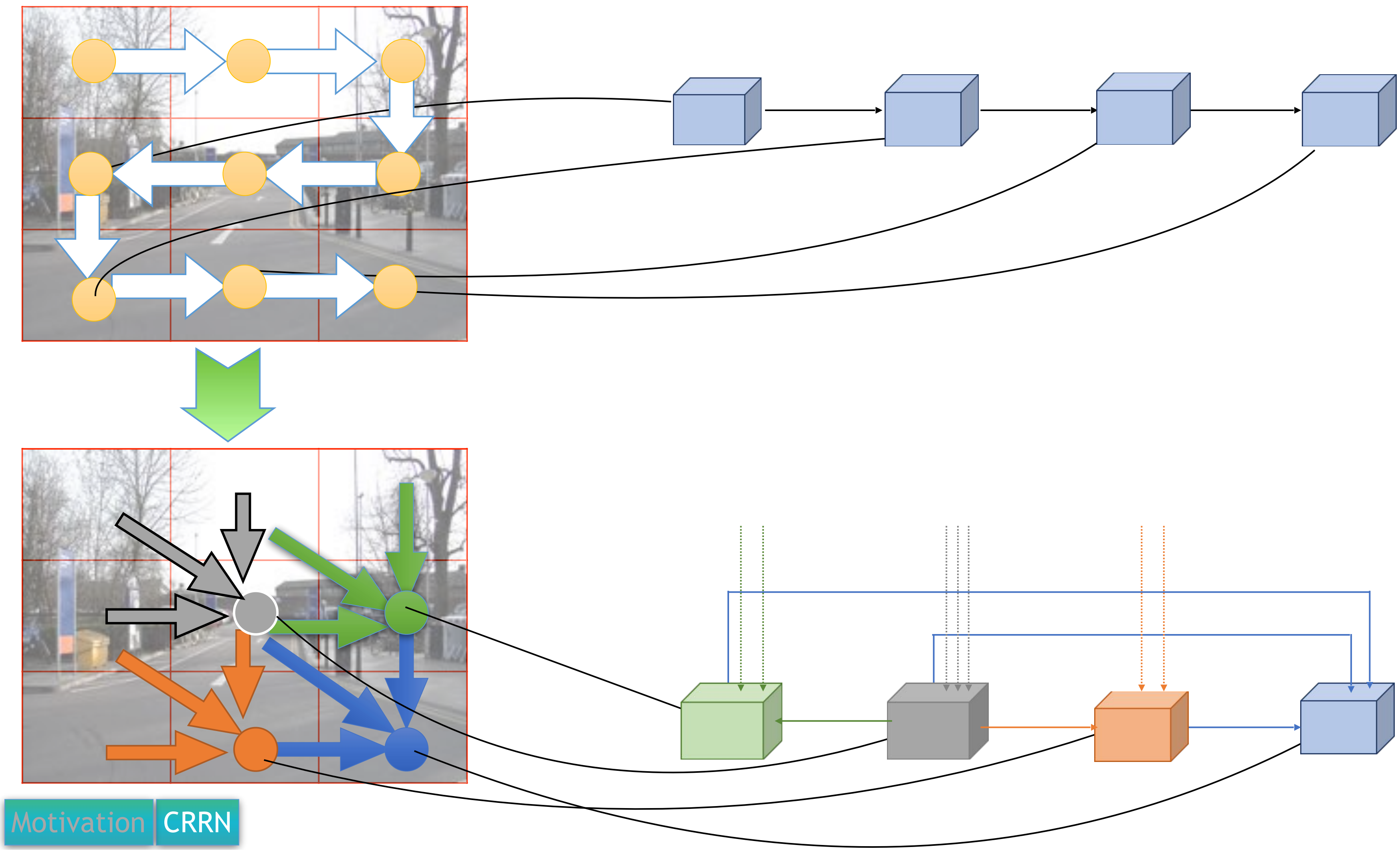
patches



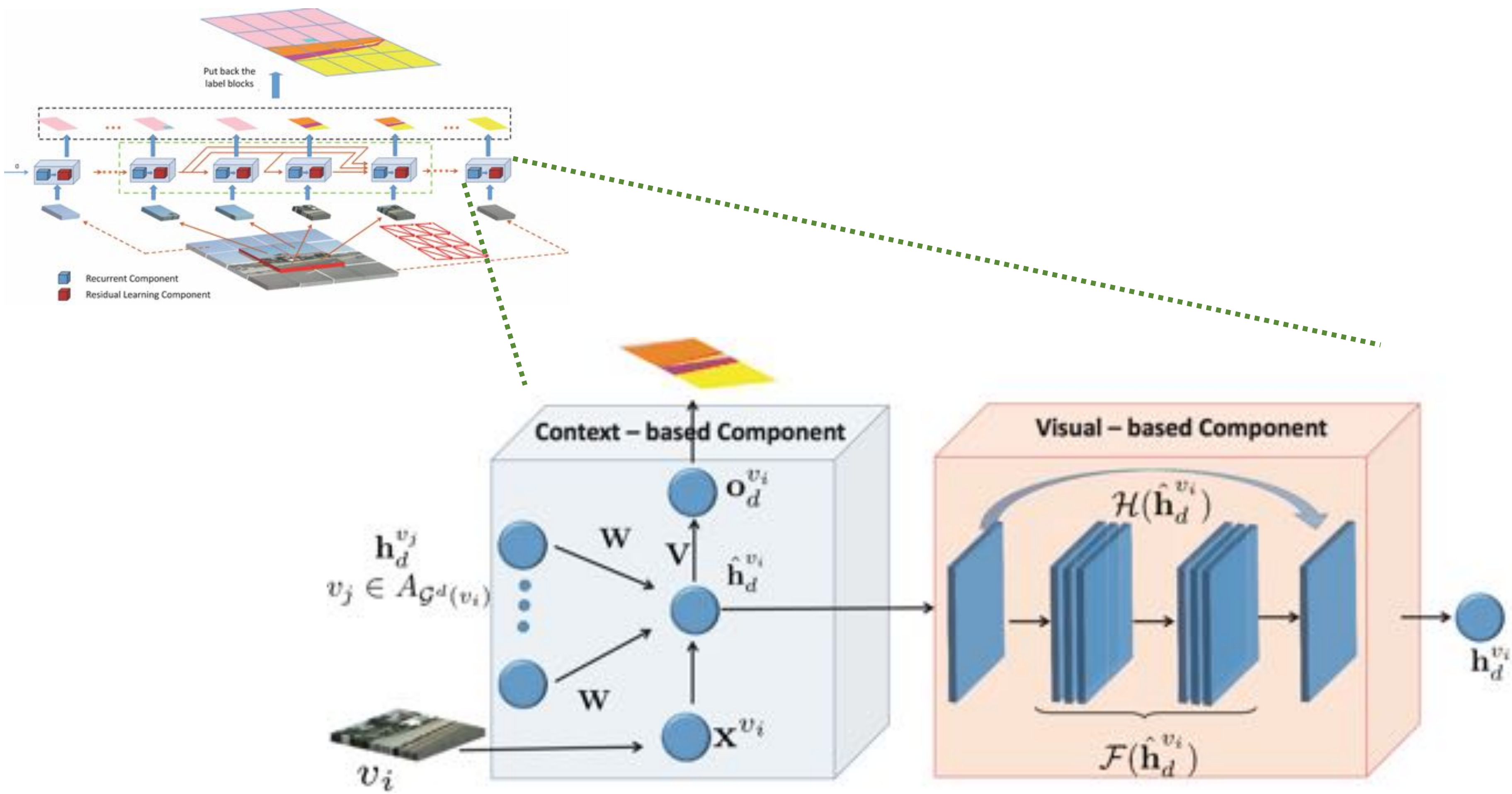
$$x^{v_i} = vec(v_i)$$

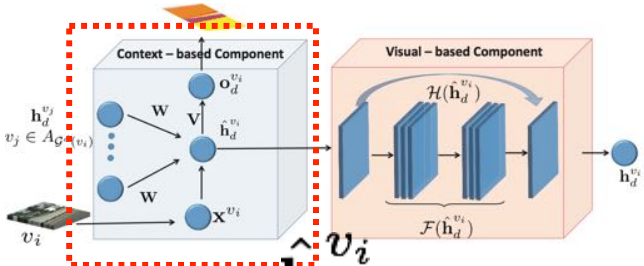
Image patches


Put back the
label blocks
0
Recurrent Component
Residual Learning Component



Motivation CRRN


Put back the label blocks
Recurrent Component
Residual Learning Component
Context – based Component
Visual – based Component
$\mathbf{h}_d^{v_j}$
$v_j \in A_{\mathcal{G}^d(v_i)}$
W
W
V
$\mathbf{o}_d^{v_i}$
$\hat{\mathbf{h}}_d^{v_i}$
$\mathbf{x}^{v_i}$
v_i
$\mathcal{H}(\hat{\mathbf{h}}_d^{v_i})$
$\mathcal{F}(\hat{\mathbf{h}}_d^{v_i})$
$\mathbf{h}_d^{v_i}$



$\hat{\mathbf{h}}_d^{v_i}$

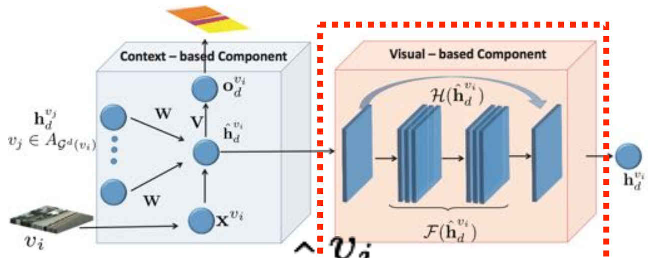
the contextual relationship between the vertex v_i and its predecessor set $\mathcal{A}_{\mathcal{G}^d}(v_i)$

$$\boxed{\begin{aligned}\hat{\mathbf{h}}_d^{v_i} &= f_{\text{CONTEXT}}(\mathbf{x}^{v_i}, \mathcal{A}_{\mathcal{G}^d}(v_i); \theta_1) \\ &= \phi(\mathbf{U}\mathbf{x}^{v_i} + \sum_{v_j \in \mathcal{A}_{\mathcal{G}^d}(v_i)} \mathbf{W}\mathbf{h}_d^{v_j} + \mathbf{b})\end{aligned}}$$

$\hat{\mathbf{h}}_d^{v_i}$ is the intermediate hidden state of vertex v_i

$\theta_1 = \{\mathbf{U}, \mathbf{W}, \mathbf{V}, \mathbf{b}\}$

$\mathbf{h}_d^{v_j}$ is the hidden state of vertex v_j



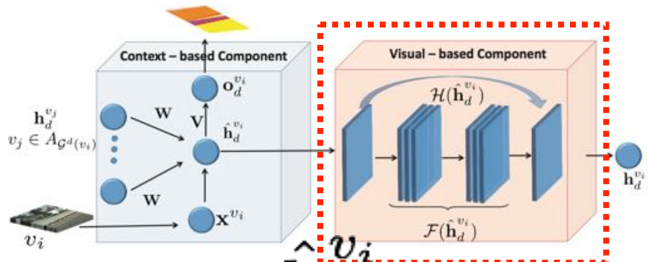
$$\hat{\mathbf{h}}_d^{v_i}$$

The representation $\mathbf{h}_d^{v_i}$ of each vertex v_i

$$\boxed{\begin{aligned}\mathbf{h}_d^{v_i} &= f_{\mathrm{VISUAL}}(\hat{\mathbf{h}}_d^{v_i};\theta_2)\\ &= \phi(\mathcal{F}(\hat{\mathbf{h}}_d^{v_i}) + \mathcal{H}(\hat{\mathbf{h}}_d^{v_i}))\end{aligned}}$$

$\mathcal{F}$ residual function

θ_2 parameters of the two-layer convolutional network



$\hat{\mathbf{h}}_d^{v_i}$

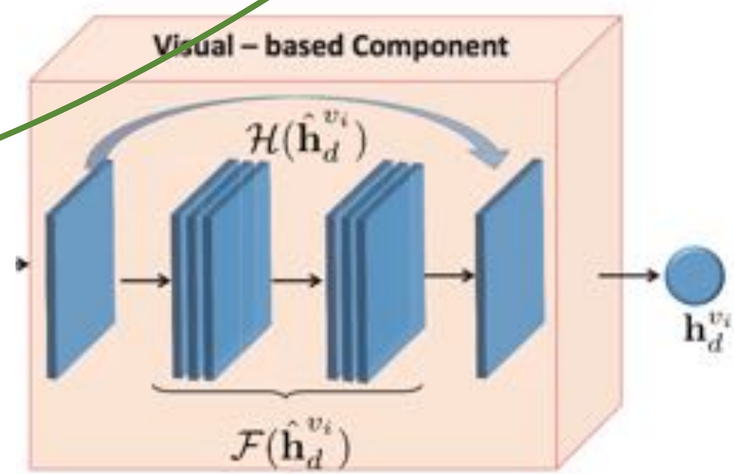
The representation $\mathbf{h}_d^{v_i}$ of each vertex v_i

$$\mathbf{h}_d^{v_i} = f_{\text{VISUAL}}(\hat{\mathbf{h}}_d^{v_i}; \theta_2)$$
$$= \phi(\mathcal{F}(\hat{\mathbf{h}}_d^{v_i}) + \mathcal{H}(\hat{\mathbf{h}}_d^{v_i}))$$

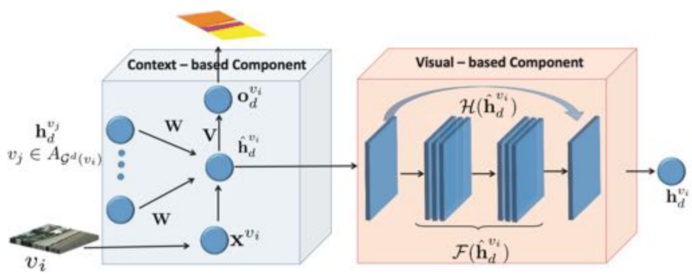
$\mathcal{F}$ residual function

θ_2 parameters of the two-layer convolutional network

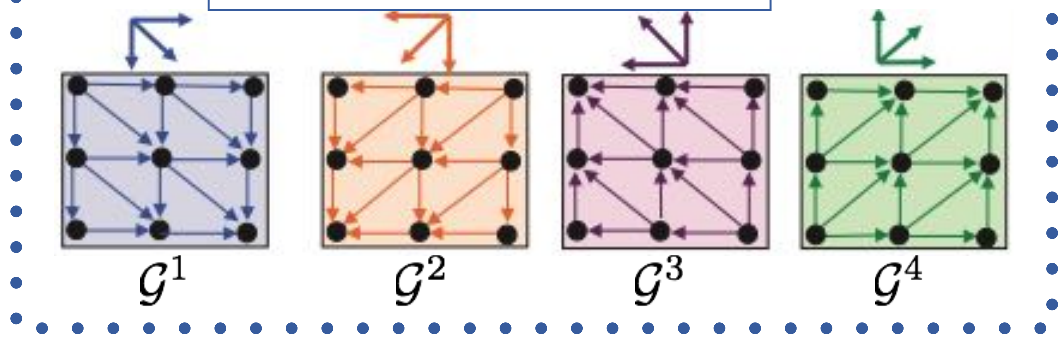
$\mathcal{H}$ identity mapping $\mathcal{H}(\hat{\mathbf{h}}_d^{v_i}) = \hat{\mathbf{h}}_d^{v_i}$



K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CVPR, 2016, pp. 770–778.



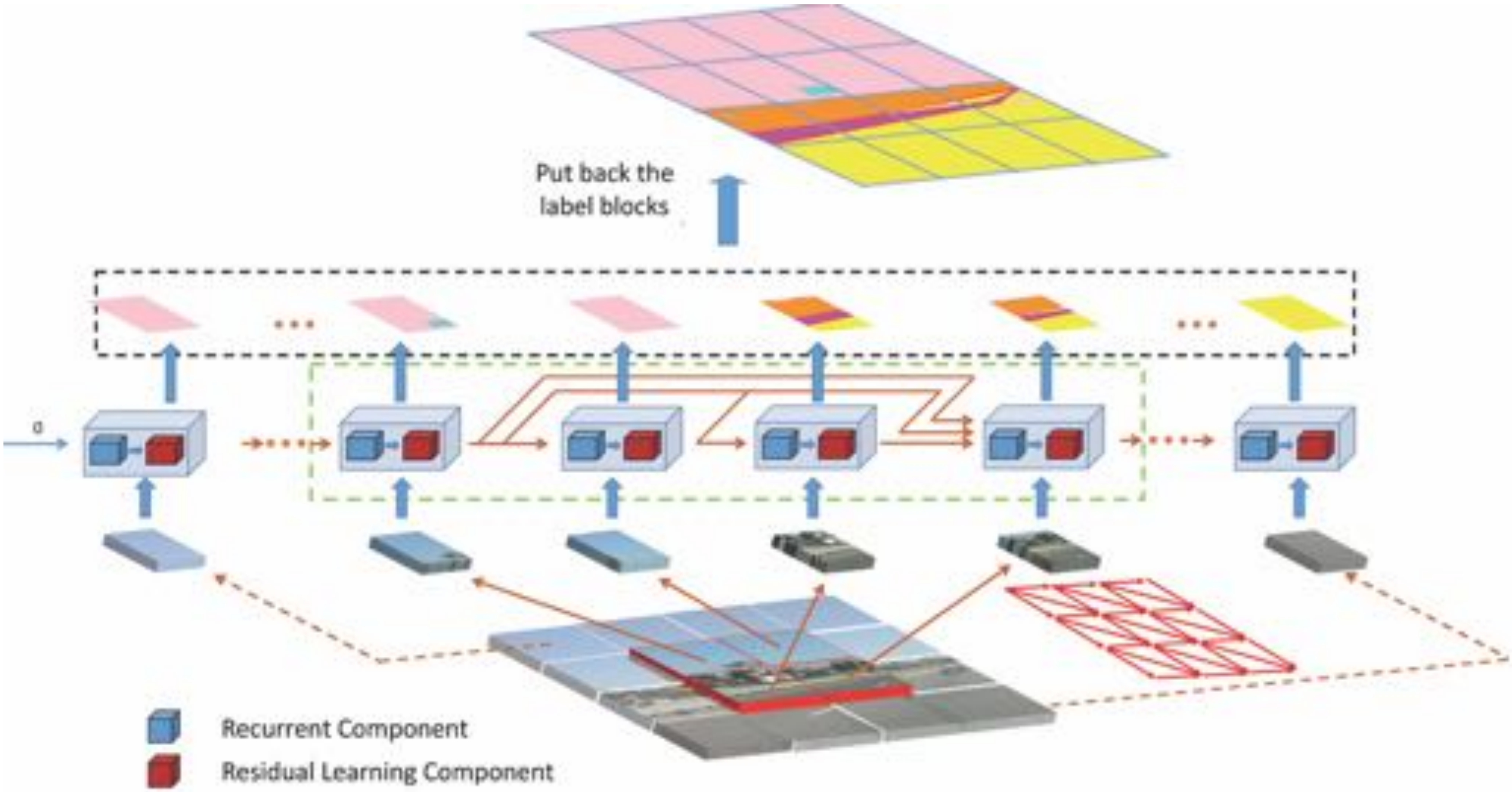
southeast, southwest, northwest and northeast



The final output from the four DAGs

$$\mathbf{o}^{v_i} = f(\mathbf{x}^{v_i}, \mathcal{A}_{\mathcal{G}(v_i)}; \theta_1, \theta_2)$$
$$= \sum_{d=1}^{4} \mathbf{V}\hat{\mathbf{h}_d}^{v_i} + \mathbf{b}_o$$

$$\theta_1^*, \theta_2^* = \arg\min_{\theta_1, \theta_2} L(\theta_1, \theta_1)$$


Put back the
label blocks
0
Recurrent Component
Residual Learning Component

Datasets	Images	Training	Testing	No. Classes
Siftflow	2,688	2,488	200	33 classes
CamVid	701	468	233	32 classes
Stanford Background	715	572	143	8 classes
SUN	16,873 6,433 (60% of pixels are labeled)	5,798	635	3,819 classes

Siftflow dataset - non-fine-tuned models

2D-LSTM: W. Byeon, et al, “Scene labeling with lstm recurrent neural networks,” in CVPR, 2015,

RCNN$_3$1/1 resolution (^{3}): P. H. Pinheiro and R. Collobert, “Recurrent convolutional neural networks for scene labeling,” in ICML, 2014

Multi-scale Convnet: C. Farabet, et al., “Learning hierarchical features for scene labeling,” TPAMI, 2013

Multi-CNN - rCPN: A. Sharma, et al. “Recursive context propagation network for semantic scene labeling,” in NIPS, 2014,

Scene Graph Structure: N. Souly and M. Shah, “Scene labeling using sparse precision matrix,” in CVPR, 2016

CNN-65-DAG-RNN: B. Shuai, et al., “Dag-recurrent neural networks for scene labeling,” in CVPR, 2016.

Context&Attention: J. Yang, et al., “Context driven scene parsing with attention to rare classes,” in CVPR, 2014

Sample&Filter + MRF: M. Najafi, et al.“Sample and filter: Nonparametric scene parsing via efficient filtering,” in CVPR, 2016.

Multi-scale RCNN: M. Liang, et al, “Convolutional neural networks with intra-layer recurrent connections for scene labeling,” in NIPS, 2015

METHOD	PA	CA
2D-LSTM	70.1%	20.9%
RCNN$_3$1/1 resolution (^{3})	77.7%	29.8%
Multi-scale Convnet	78.5%	29.6%
Multi-CNN - rCPN	79.6%	33.6%
Scene Graph Structure	80.6%	45.8%
CNN-65-DAG-RNN	81.1%	48.1%
Context&Attention	79.8%	48.7%
Sample&Filter + MRF	83.1%	44.3%
Multi-scale RCNN	83.5%	35.8%
Our CRRN	**84.7%**	**61.0%**

CamVid dataset - non-fine-tuned models

Neural Decision Forests: S. Rota Bulo and P. Kontschieder, “Neural decision forests for semantic image la- belling,” in CVPR, 2014

SVM+MRF: J. Tighe and S. Lazebnik, “Finding things: Image parsing with regions and per- exemplar detectors,” in CVPR, 2013

METHOD	PA	CA
Neural Decision Forests	82.1%	56.1%
SVM+MRF	83.9%	**62.5%**
Our CRRN	**84.4%**	54.8%

CRRN

Stanford-background dataset - non-fine-tuned models

2D-LSTM: W. Byeon, et al"Scene labeling with lstm recurrent neural networks," in CVPR, 2015,

Recurrent CNN : P. H. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in ICML, 2014

Multi-scale Convnet(the best): C. Farabet, et al., "Learning hierarchical features for scene labeling," TPAMI, 2013

Multi-scale RCNN : M. Liang, et al. "Convolutional neural networks with intra-layer recurrent connections for scene labeling," in NIPS, 2015,

Associative Hierarchical Random Fields: L. Ladick, et al. "Associative hierarchical random fields," TPAMI, 2014

Scene Graph Structure N. Souly and M. Shah, "Scene labeling using sparse precision matrix," CVPR, 2016

METHOD	PA	CA
2D-LSTM	78.6%	68.8%
$RCNN_3 1/1$ resolution (^{3})	80.2%	69.9%
Multi-scale Convnet	81.4%	76.0%
Multi-scale RCNN	83.1%	74.8%
Associative Hierarchical Random Fields	80.9%	70.4%
Scene Graph Structure	84.6%	**77.3%**
Our CRRN	**85.23%**	75.2%

CRRN

SUN dataset - non-fine-tuned models

CNN-65-DAG-RNN: B. Shuai, et al., “Dag-recurrent neural networks for scene labeling,” in CVPR, 2016.

FCNN: J. Long, et al, “Fully convolutional networks for semantic segmentation,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

METHOD	PA	CA
CNN-65-DAG-RNN	71.51%	54.57%
FCNN	77.2%	**62.03%**
Our CRRN	**78.61%**	59.9%

Siftflow dataset - fine-tuned models

CNN - Global Context: B.Shuai, et al., “Integrating parametric and non- parametric models for scene labeling,” in CVPR, 2015

FCNN: J. Long, et al., “Fully convolutional networks for semantic segmentation,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

VGG-conv5-DAG-RNN: B. Shuai, et al.“Dag-recurrent neural networks for scene labeling,” in CVPR, 2016

METHOD	PA	CA
CNN - Global Context	80.1%	39.7%
FCNN	85.2%	51.7%
VGG-conv5-DAG-RNN	**85.3%**	55.7%
Our CRRN	84.7%	**61.0%**

CNN-65-DAG-RNN CRRN GT

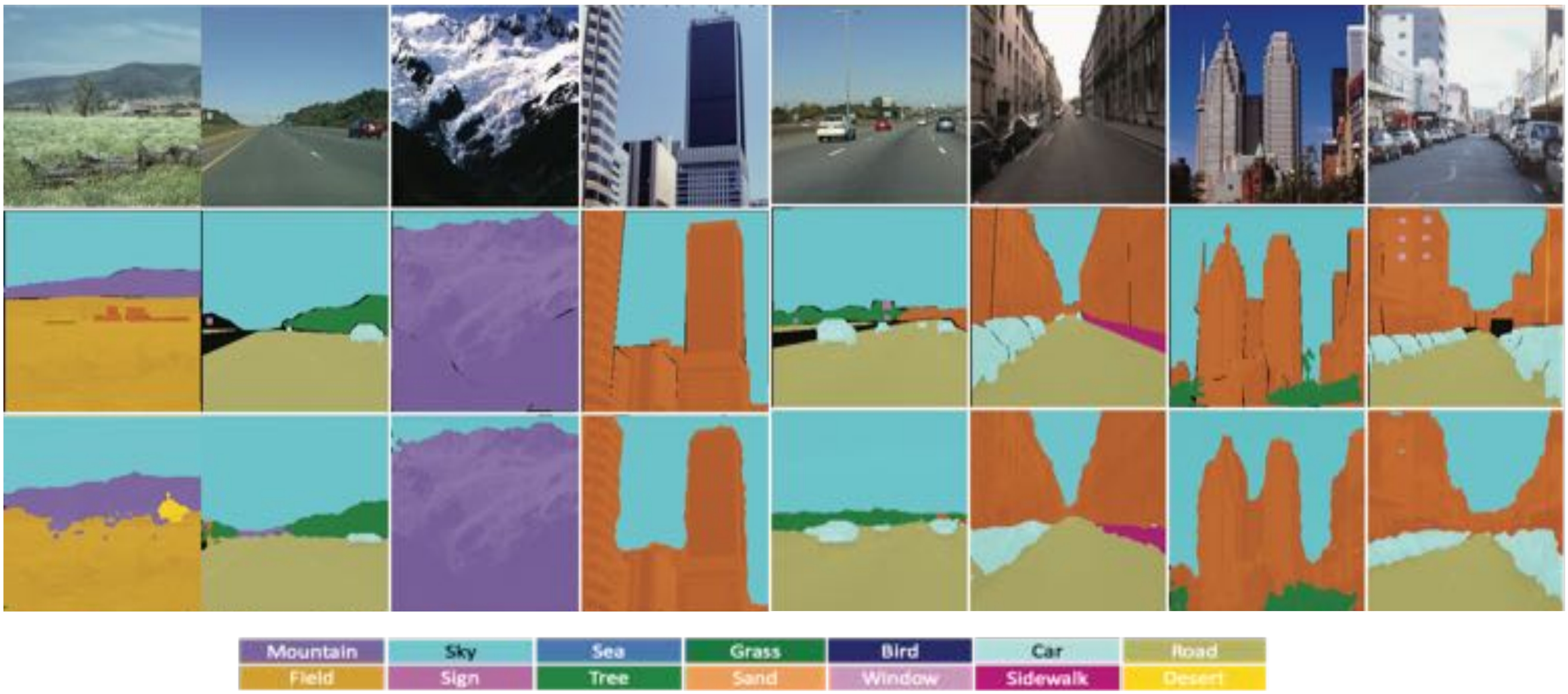
CNN-65-DAG-RNN: B. Shuai, Z. Zuo, G. Wang, and B. Wang, "Dag-recurrent neural networks for scene labeling," in CVPR, 2016.

CRRN: T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides "Deep Contextual Recurrent Residual Networks for Scene Labeling" , Pattern Recognition, Volume. 80, Jan,2018, pp. 32-41

Siftflow dataset

GT

CRRN



CRRN: T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides “Deep Contextual Recurrent Residual Networks for Scene Labeling" , Pattern Recognition, Volume. 80, Jan,2018, pp. 32-41

CamVid dataset

GT

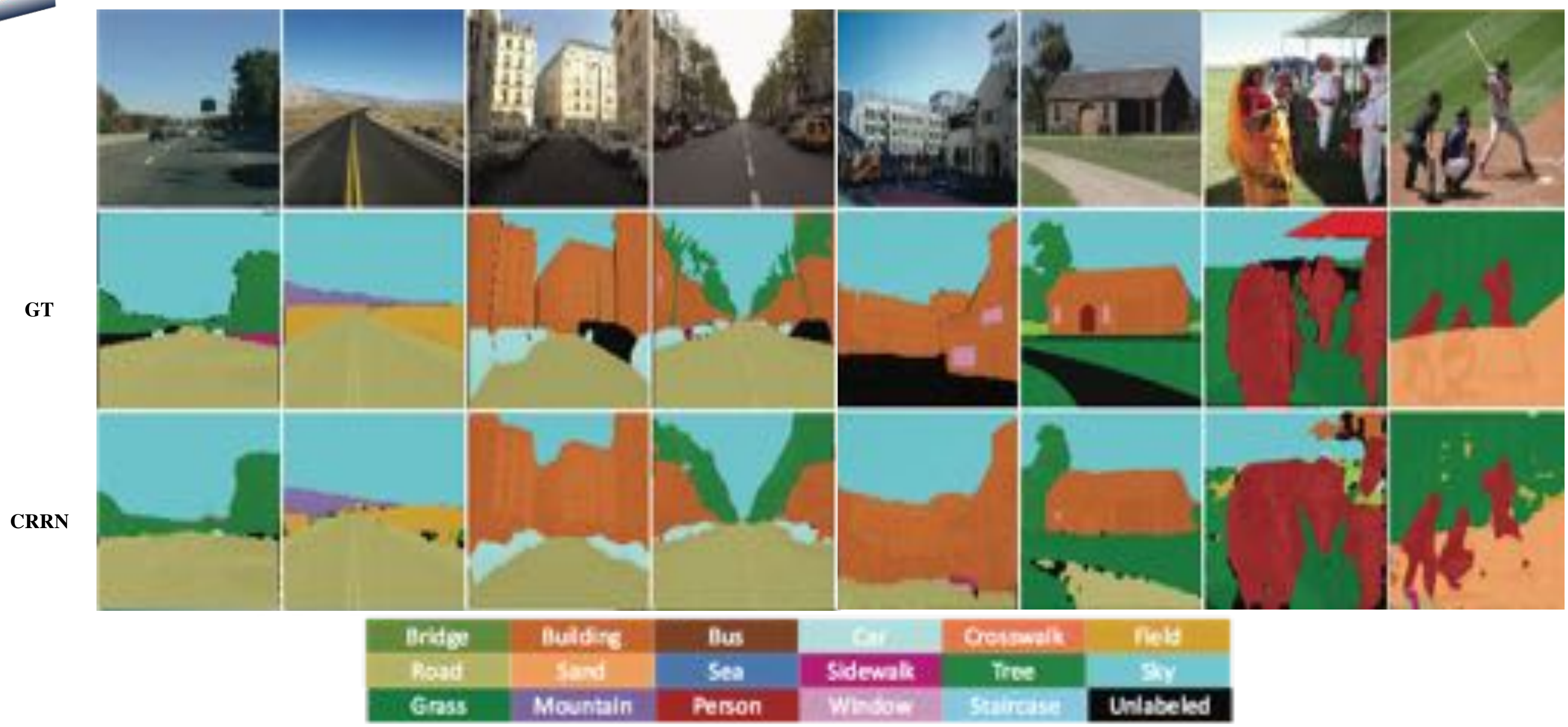
CRRN

CRRN: T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides "Deep Contextual Recurrent Residual Networks for Scene Labeling" , Pattern Recognition, Volume. 80, Jan,2018, pp. 32-41

Standford dataset

CRRN: T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides "Deep Contextual Recurrent Residual Networks for Scene Labeling" , Pattern Recognition, Volume. 80, Jan,2018, pp. 32-41

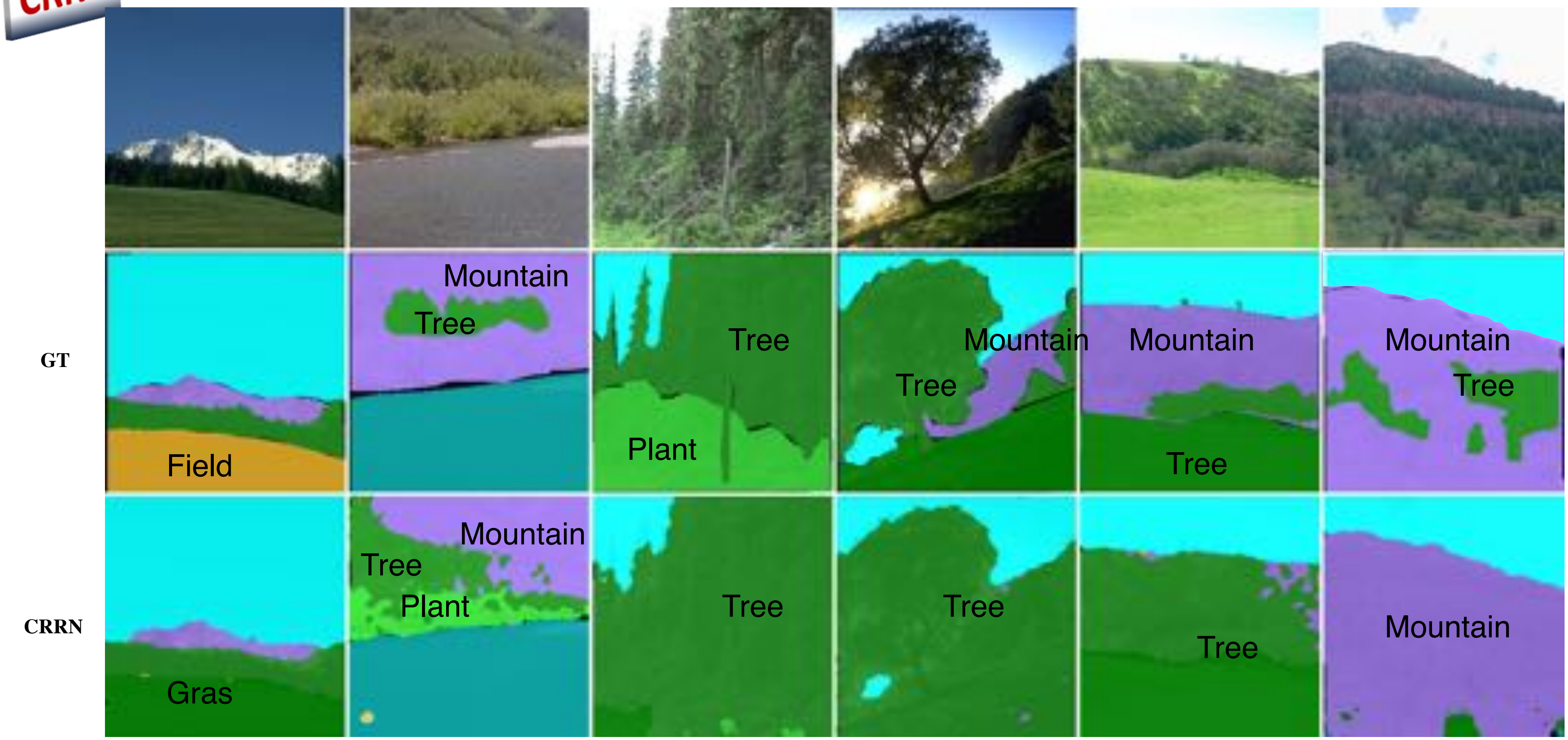
SUN dataset



CRRN: T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides "Deep Contextual Recurrent Residual Networks for Scene Labeling" , Pattern Recognition, Volume. 80, Jan,2018, pp. 32-41

False Positive: ground truth is mistakenly labeled as unlabeled

Confusing cases: The object is labeled by different synonym class name



CRRN: T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides "Deep Contextual Recurrent Residual Networks for Scene Labeling" , Pattern Recognition, Volume. 80, Jan,2018, pp. 32-41

CRLS

Contextual Recurrent Level Set (CRLS) Networks for Semantic Instance Segmentation

- Motivation
- Level Set
- Deep Neural Networks

CRRNs

Contextual Recurrent Residual Networks (CRRN) for Scene Labeling

- Motivation
- Contextual Recurrent Residual Networks (CRRN)

T.H.N Le, K.G. Quach, K. Luu, C.N.Duong, and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , **Transactions on Image Processing**, Vol. 27, No. 5, May 2018, pp. 2393–2407

T.H.N. Le , R. Gummadi, M. Savvides "Deep Recurrent Level Set for Segmenting Brain Tumors", submitted to **MICCAI 2018** (provisional accept - 2 accepted, 1 weak rejected)

T.H.N Le, C.N. Duong, L. Han, K. Luu, K.G. Quach and M. Savvides "Deep Contextual Recurrent Residual Networks for Scene Labeling" , **Pattern Recognition**, Volume. 80, Jan,2018, pp. 32-41

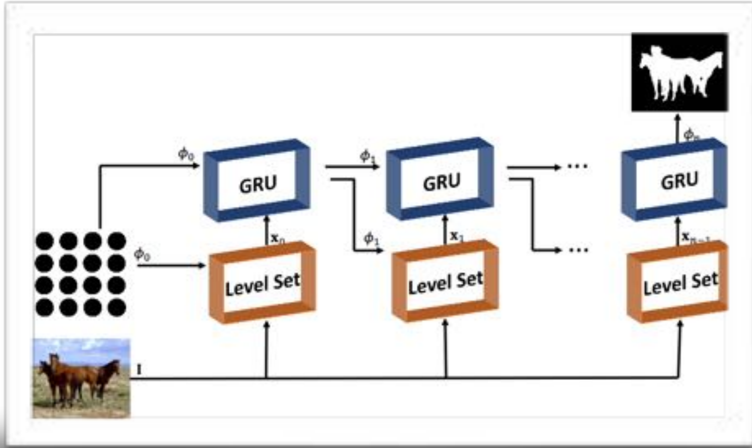
Conclusions

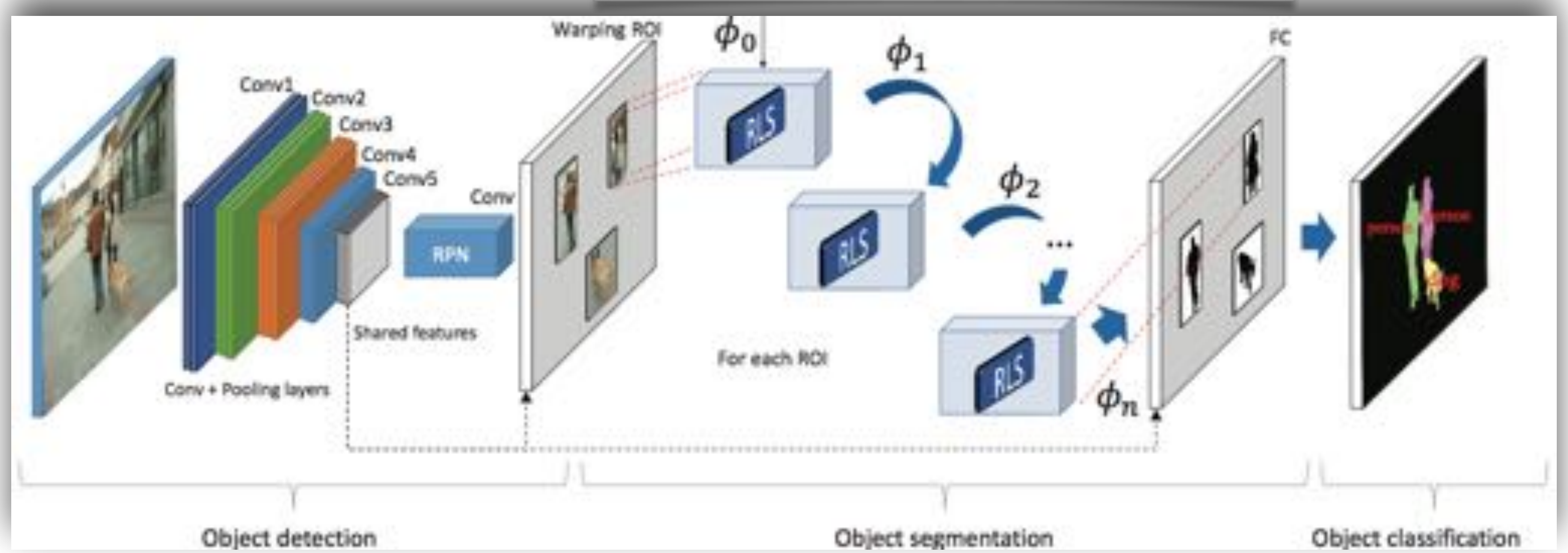
- Conclusions
- Future Work

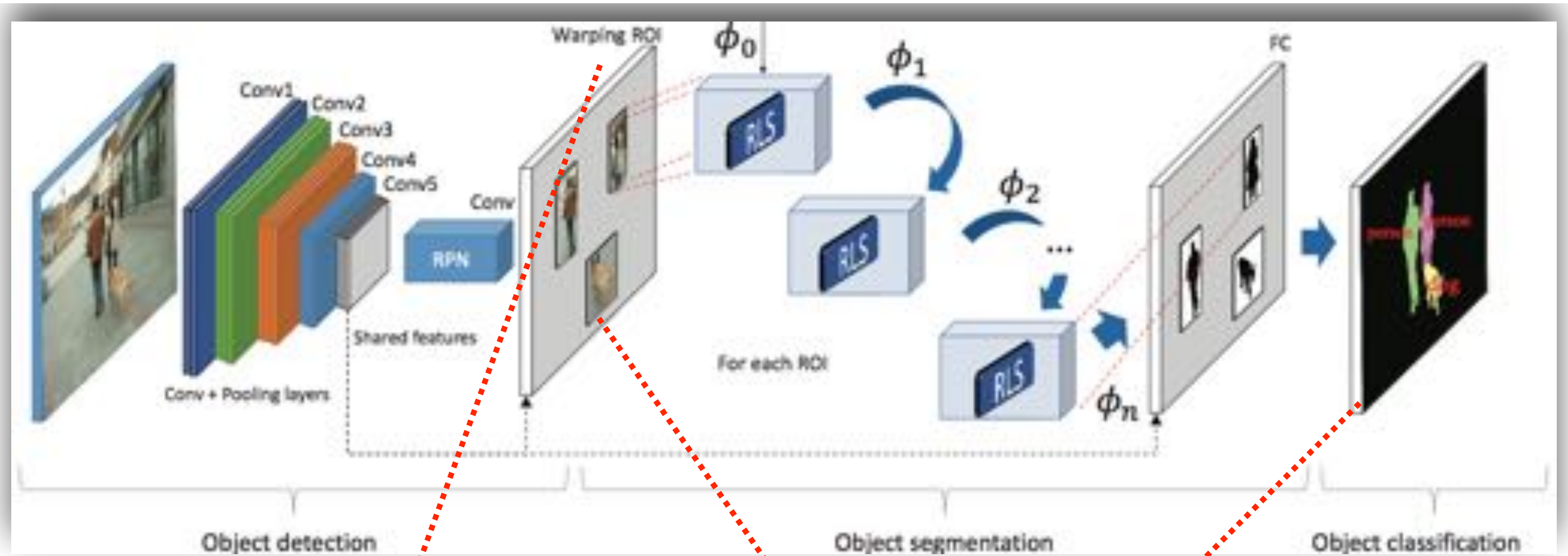
RLS

Recurrent Level Set (RLS)

- Boost Level Set to the new level of learnable framework
- End-to-end network
- Able to collaborate with other deep learning model

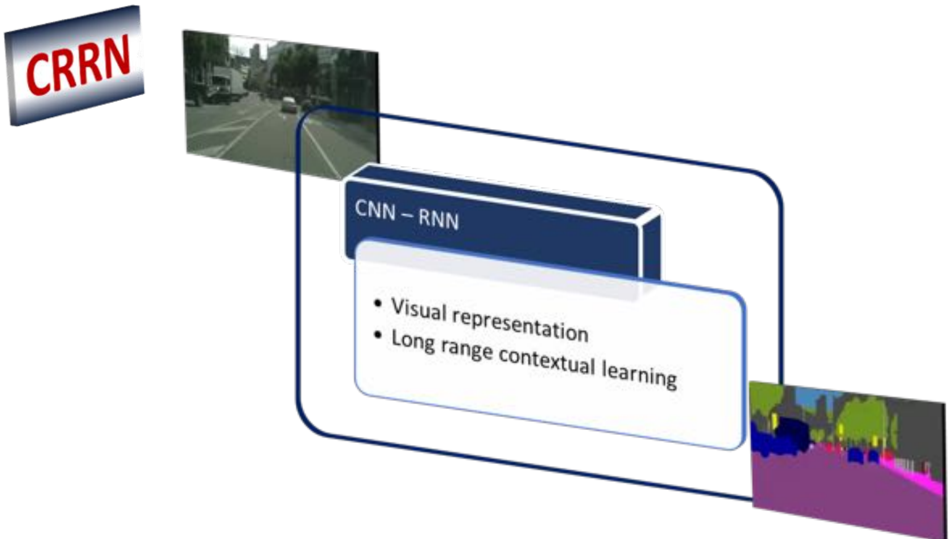


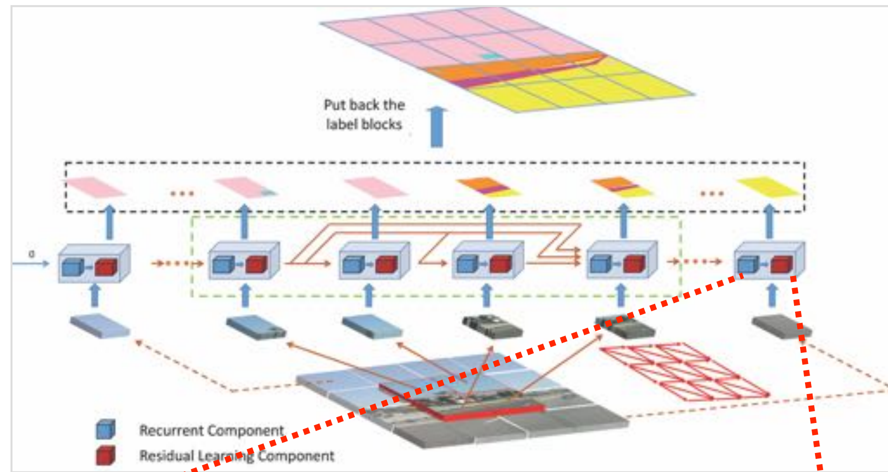




Object size

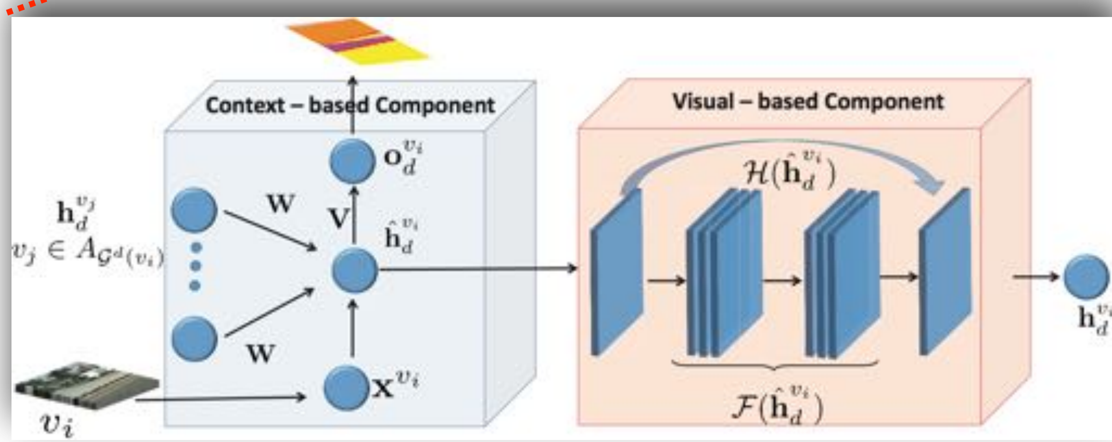
Object classification

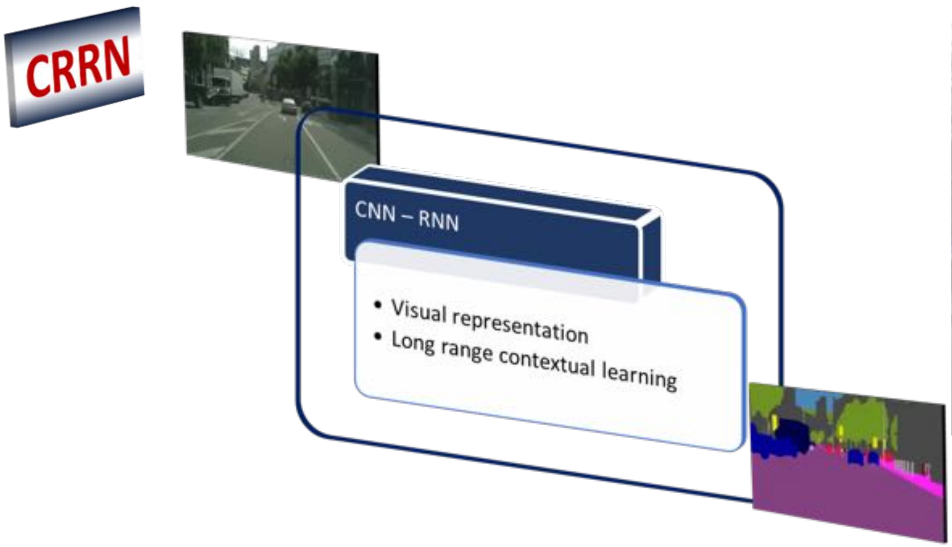


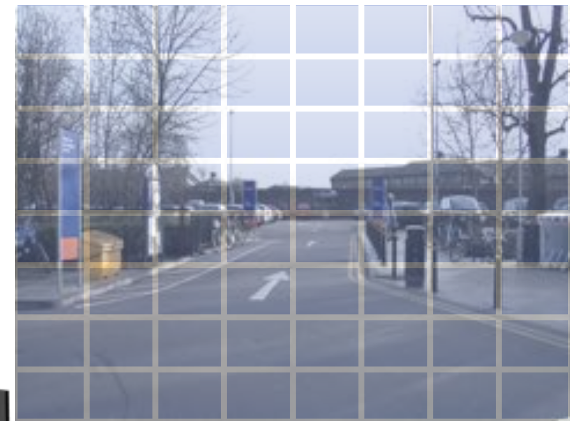


CRRN

- Contextual Recurrent Residual Network
- End-to-end network
- Train from scratch
- Simultaneously learn visual representation and model long range context

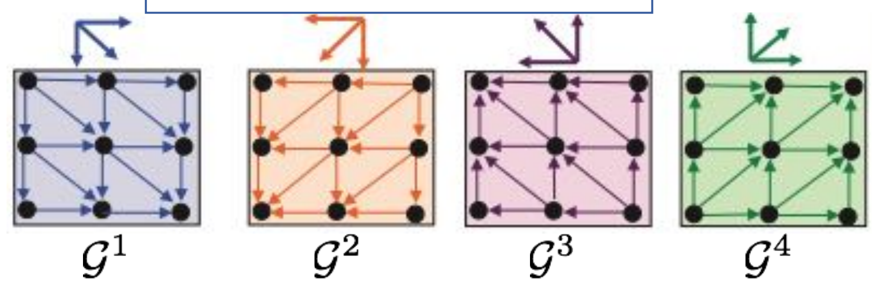






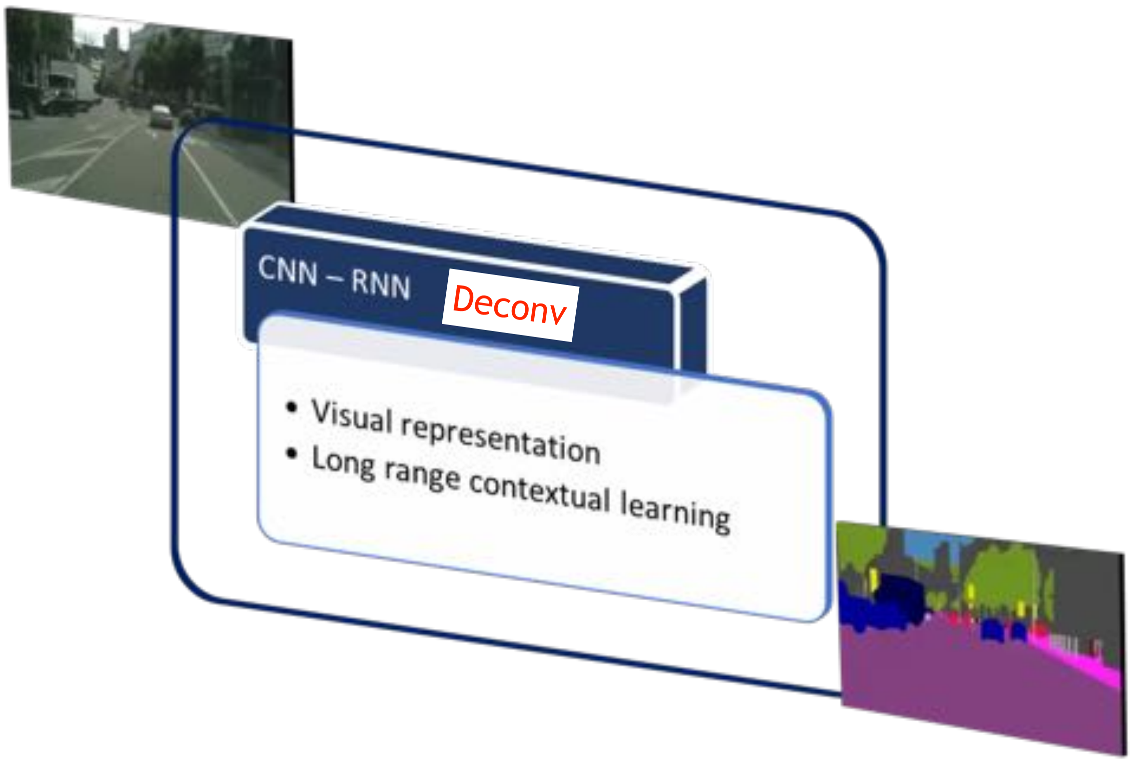
Number of patches

southeast, southwest, northwest and northeast



Equivalent!

CRRN



CNN – RNN
Deconv
• Visual representation
• Long range contextual learning

Journal publications (1)

- **T.H.N Le**, K..G. Quach, K. Luu and M. Savvides "Reformulating Level Sets as Deep Recurrent Neural Network Approach to Semantic Segmentation" , **IEEE Trans. on Image Processing (TIP),** Vol. 27, No. 5, May 2018, pp. 2393–2407
- **T.H.N Le**, C.N. Duong, L. Han, K. Luu and M. Savvides “Deep Contextual Recurrent Residual Networks for Scene Labeling" , **Pattern Recognition,** Volume. 80, pp. 32-41, 2018
- **T.H.N Le**, CC Zhu, Y. Zheng, K. Luu, and M. Savvides "DeepSafeDrive: A Grammar-aware Driver Parsing Approach to Driver Behavioral Situational Awareness (DB-SAW)" , **Pattern Recognition** , Volume 66, June 2017, pp. 229-238
- **T.H.N Le**, K. Luu, K., CC. Zhu and Savvides, M., " Semi Self- Training Beard/Moustache Detection and Segmentation Simultaneously," **Image and Vision Computing**, Volume 58, February 2017, pp. 214–223
- **T.H.N Le,** and Savvides, M., "A Novel Shape Constrained Feature-based Active Contour (SC-FAC) Model for Lips/Mouth Segmentation with Unconstrained Background,” **Pattern Recognition**, Volume. 54, pp. 23–33, 2016
- K. Luu, C. C. Zhu, C. Bhagavatula, **T.H.N Le** and M. Savvides, "A Deep Learning Approach to Joint Face Detection and Segmentation," **Special Issue, Advances in Face Detection and Facial Image Analysis, Springer,** pp. 1–12, 02 April 2016
- **T.H.N Le**, Luu, K., and Savvides, M., "SparCLeS: Dynamic L1 Sparse Classifiers with Level Sets for Robust Beard/Moustache Detection and Segmentation," **IEEE Trans. on Image Processing (TIP)**, Vol. 22, No. 8, pp. 3097-3107, Aug. 2013.

Journal publications(2)

- **T.H.N Le,** Seshadri, K, Luu, K., and Savvides, M., "A Facial Aging and Asymmetry Decomposition to Twins Identification," **Pattern Recognition**, Vol.48, Issue 12, Pages 3843-3856, 2015
- **T.H.N Le, Lin**, C. C., Chang, C. C., and Le, H. B., "A high quality and small shadow size visual secret sharing scheme based on hybrid strategy for grayscale images," **Journal of Digital Signal Processing** Vol. 21, Issue 6, pp. 734 - 745, 2011
- Chen, W. J., Chang, C. C., and **T.H.N Le**, "High Payload Steganography Mechanism Using Hybrid Edge Detector", **Journal of Expert Systems with Application**, Vol. 37, No. 4, pp. 3292-3301, April, 2010.
- Chang, C. C., Lee, J. S., **T.H.N Le,** "Improved Wet Paper Coding Based on an n-Indicator and a Fuzzy Edge Detector," **Journal of Digital Signal Processing,** Vol. 20, Issue 4, pp. 1286-1307, 2010
- Chang, C. C., Lin, C. C., **T.H.N Le** and Le, H. B., "Sharing a Verifiable Secret Image Using Two Shadows," **Pattern Recognition**, Vol. 42, Issue 1, pp. 3097-3114, May, 2009.
- Chang, C. C., Lin, C. C., **T.H.N Le,** and Le, H. B., "Self Verifying Visual Secret Sharing Using Error Diffusion and Interpolation Techniques," **IEEE Transactions on Information Forensics and Security,** Vol. 4, Issue 4, pp. 790-801, Dec. 2009.
- Chang, C. C., Lin, C. C., **T.H.N Le**, and Le, H. B., "A Probabilistic Visual Secret Sharing Scheme for Grayscale Images with Voting Strategy," **International Journal of Intelligent Information Technology Application (IJIITA),** Vol. 1, No. 1, Jul. 2008, pp.
- H.B.Le, K.Luu, **T.H.N Le**, "Audio watermarking using psychoacoustic auditory model and spread spectrum theory", **Post, Telecommunications and Information Technology Journal,** Vietnam, April 2006.

Conference publications(1)

- **T.H.N. Le** , R. Gummadi, M. Savvides “Deep Recurrent Level Set for Segmenting Brain Tumors”, submitted to **MICCAI** 2018 (accept)
- **T.H.N. Le** , K. Quach, CC. Zhu, C.N.Duong, K. Luu, M. Savvides, "Robust Hand Detection and Classification in Vehicles and in the Wild", **CVPRW**, pp. 1203-1210, July 2017
- **T.H.N. Le** CC. Zhu, Y. Zheng, K. Luu, M. Savvides, “Robust Hand Detection in Vehicles ", **ICPR** 2017, pp. 573-578
- Ligong Han, **T.H.N. Le**, M. Savvides "An Automatic Cells Detection and Segmentation," **SPIE Medical Imaging** 2016,
- Y. Zheng, C. Zhu, K. Luu, **T.H.N. Le** , C. Bhagavatula and M. Savvides, “Towards a Deep Learning Framework for Unconstrained Face Detection," **BTAS**, Buffalo, NY, Sep. 2016.
- **T.H.N. Le** , Y. Zheng*, CC. Zhu*, K. Luu, and M. Savvides, "Multiple Scale Faster-RCNN Approach to Driver’s Cell-phone Usage and Hands on Steering Wheel Detection", **CVPRW** 2016, pp. 46-53
- **T.H.N. Le** , CC Zhu, Y. Zheng, K. Luu, and M. Savvides "DeepSafeDrive: A Grammar-aware Driver Parsing Approach to Driver Behavioral Situational Awareness (DB-SAW)" , **CVPRW** 2016
- CC. Zhu*, Y. Zheng*, K. Luu, **T.H.N. Le** , C. Bhagavatula, and M. Savvides, "Weakly Supervised Facial Analysis with Dense Hypercolumn Features", **CVPRW** 2016
- **T.H.N. Le** , K. Luu., and Savvides, M., "Fast and Robust Self-Training Beard/Moustache Detection and Segmentation," **ICB**, pp. 1-8, Thailand 2015
- F. M’hiri, **T.H.N. Le**, L.Duong, C. Desrosiers, M. Cheriet, "Hierarchical Segmentation and Tracking of Coronary Arteries in 2d x-ray Angiography Sequences”, **ICIP**, pp. 1707-1011, Quebec, Canada, 2015
- K. Singh , K. Luu , **T.H.N. Le** , M. Savvides , "A Robust Contour Sampling and Tensor-Based Approach To Facial Beard And Mustache Shape Segmentation And Matching," **ICIP**, pp. 1399-1403, Quebec, Canada, 2015
- **T.H.N. Le** , U. Prabhu., and Savvides, M., "A Novel Eyebrow Segmentation and Eyebrow Shape-based Identification," **IJCB**, pp. 1-8, USA 2014.
- **T.H.N. Le** , K. Singh, K Luu, Savvides, M, "A Robust Monte Carlo and Tensor-based Shape Context Matching Approach to Facial Beard/ Moustache Ranking and Retrieval”, **WiML** Canada, 2014
- **T.H.N. Le**, Luu, K., Seshadri, K.,Savvides, M., "Beard and Mustache Segmentation Using Sparse Classifiers on Self-Quotient Images", **ICIP**, US. 2012, pp. 165-168.
- Luu, K., **T.H.N. Le** , Seshadri, K., and Savvides, M., "FaceCut - A Robust Approach for Facial Feature Segmentation", **ICIP**, US. 2012, pp. 1841-1844.
- M., F., **T.H.N. Le** , Duong, L., and Cheriet, M., " A New Adaptive Segmentation Framework of Tubular Structures in X-Ray Angiography", **ISSPA**, pp. 496-500, 2012.

Conference publications

- **T.H.N. Le,** Seshadri, K, Luu, K., and Savvides, M., "A Facial Aging and Asymmetry Decomposition to Twins Identification," **CVPRW**, USA 2015.
- **T.H.N. Le** , U. Prabhu., and Savvides, M., "A Novel Eyebrow Segmentation and Eyebrow Shape-based Identification," **IJCB**, pp. 1-8, USA 2014.
- H.Q. Hua, **T.H.N. Le** , H.B. Le, "An Effective Initialization for ASM based Methods," **IEEE ICCIS**, Vol. 8838, pp.421-432, 2014
- **T.H.N. Le** , Luu, K., Seshadri, K., and Savvides, M. "A Facial Aging Approach to Identification of Identical Twins", **BTAS** 2012, pp. 91-98, US 2012.
- **T.H.N. Le** , Luu, K., U. Prabhu, Savvides, M., "A Novel Energy based Filter for Cross-Blink Eye Detection," **ICIP**, US.,Sept. 2012. pp. 1845-1848.
- **T.H.N. Le** Bui, T. D., Suen, C. Y.˘ IJTernary Entropy-based Binarization of Degraded Document Images Using Morphological Operators", **ICDAR**, pp. 114-118, 2011.
- **T.H.N. Le** , Nguyen. K. H, Le, H. Bac, "Literature Survey on Watermarking Tools, Watermark Attacks, and Benchmarking Tools", **MMEDIA**, pp. 67-73, Greece, 2010.
- **T.H.N. Le** , Nguyen. K. H, Le, H. Bac, "A Robust Biometric Watermark-Based Authentication Scheme", **ICDIM**, pp. 398-403, Canada, 2010.
- **T.H.N. Le** , "HolisticWord Recognition using Support Vector Regression in Combination with Spectral Regression", **WiML**, Canada, 2010.
- Lee, J. S., **T.H.N. Le** , "Hybrid (2, n) Visual Secret Sharing Scheme for Color Images", **ICCTT**, pp.1-8, 2009.
- Chang, C. C., Lin, C. C.,**T.H.N. Le** , and Le, H. B., "A New Probabilistic Visual Secret Sharing Scheme for Color Images,"**IIHMSP** 2008, pp. 1305-1308, Aug. 2008.


Thank you