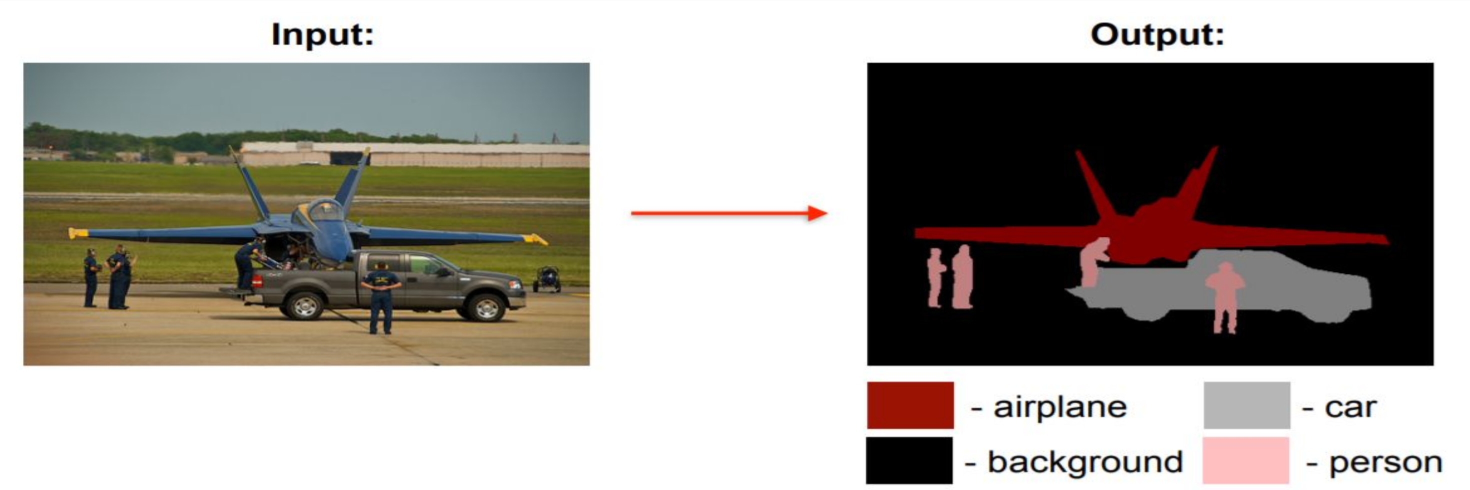
Convolutional Random Walk Networks for Semantic Image Segmentation

N S M S SACHIN:UG3-CSE:S20160010056

Semantic Segmentation

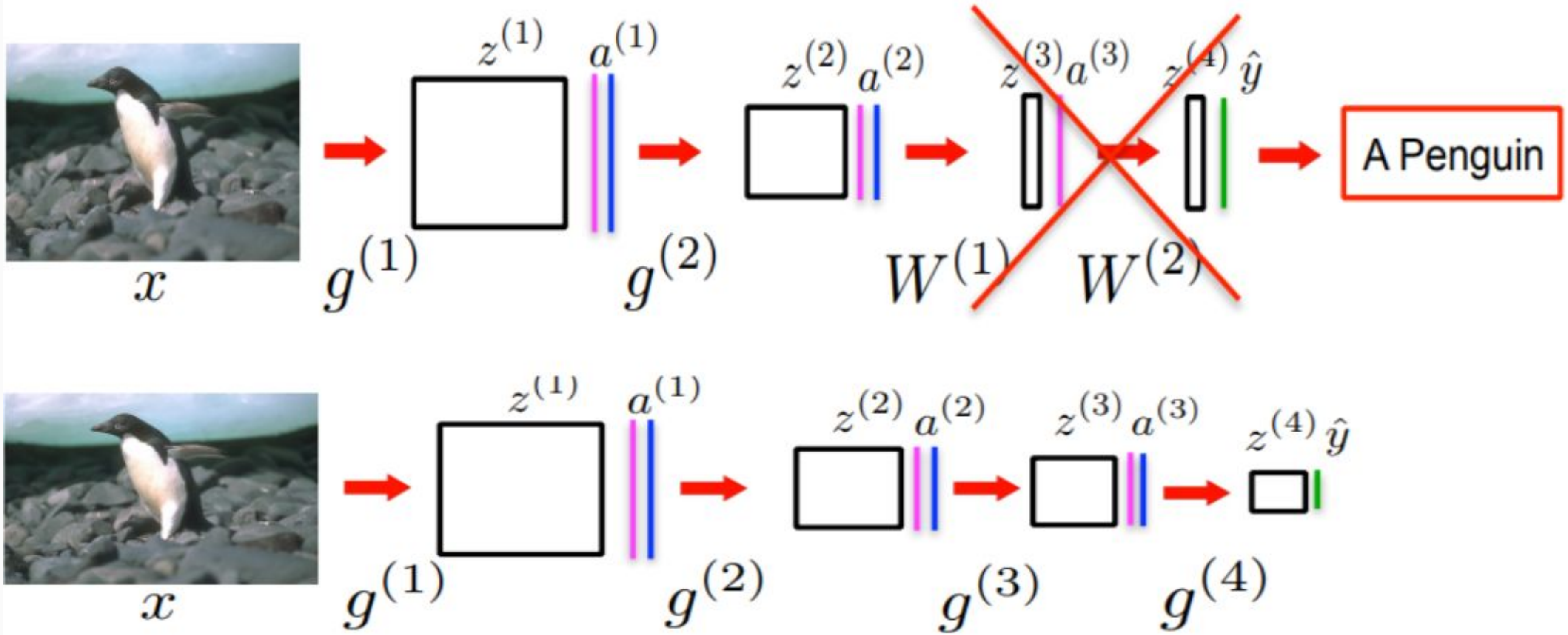
Given an image, we want to identify what class each pixel in the image belongs to.



Issues with standard CNNs:

- Standard CNNs do not predict structured outputs.
- Standard CNNs do not preserve local spatial information.

We remove fully connected layers and replace them with convolutional layers to obtain Fully Convolutional Network (FCN)



Issues with Fully convolutional networks (FCN)

- **FCNs produce 8-32 times lower resolution output than the original image.**
- **Predictions are blurry and lack fine boundary details.**

- **No explicit mechanism to preserve structure in an image.**

- **FCN tries to predict as many pixels correctly as possible without necessarily preserving the structure!**

Convolutional Random Walk Networks

- Convolutional Random Walk Network (RWN) addresses the issues of poor boundary localization and spatially fragmented predictions with very little increase in model complexity.
- It jointly optimizes the objectives of pixelwise affinity and semantic segmentation
- RWN is implemented using standard convolution and matrix multiplication.
- RWN does not need post-processing, it is implemented using standard layers and loss functions, and it also has a compact model.
- RWN seamlessly integrates both affinity and segmentation branches, and can be jointly trained endto-end via standard back-propagation with minimal modifications to the existing FCN framework.

Convolutional Random Walk Networks

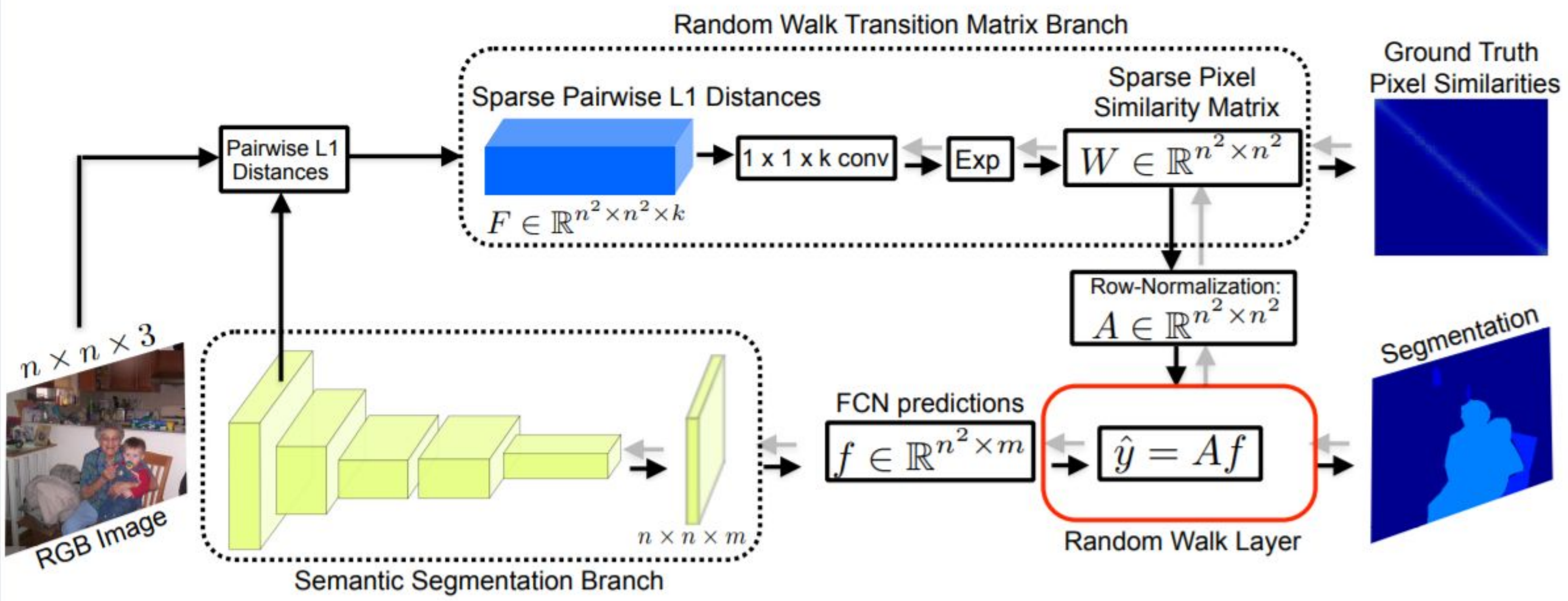
- Let W denote an affinity matrix, where $W_{ij} \in [0,1]$ depicts how similar pixels i and j are, and D is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} W_{ij}$.
- A random walk transition matrix is defined as $A = D^{-1}W$.
- Given an initial node distribution f a random walk is implemented via a matrix multiplication as Af.

- Simulating random walk forces similar nodes to cluster together

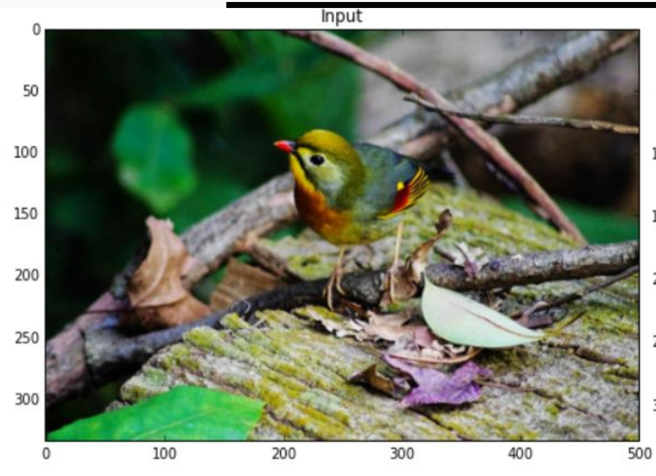
Convolutional Random Walk Networks

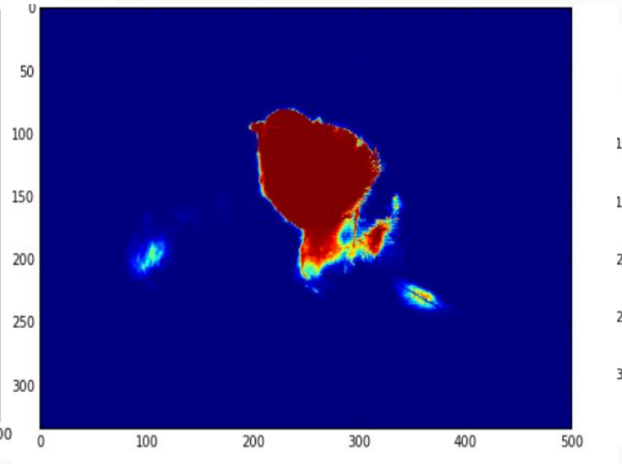
- A network consisting of two branches.
- The first branch predicts initial semantic segmentation potentials .
- The second branch predicts a random walk transition matrix .
- The outputs from both branches are combined via a random walk layer as .
- RWN can be jointly trained end-to-end using the standard back propagation framework.

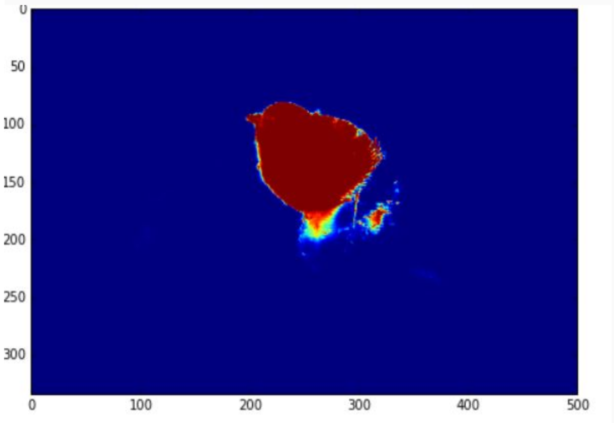
Convolutional Random Walk Networks



Convolutional Random Walk Networks



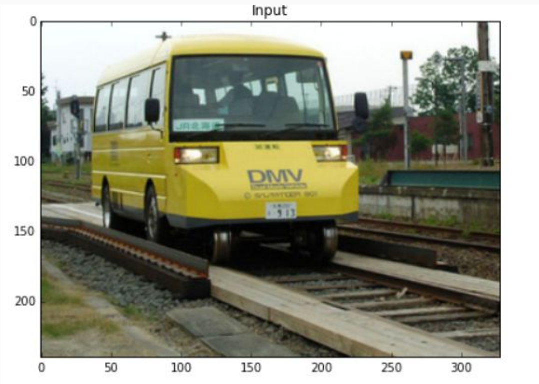


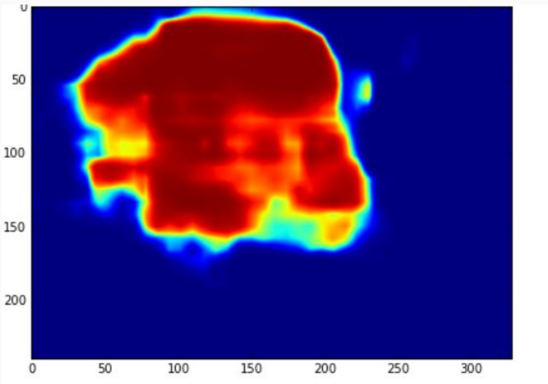


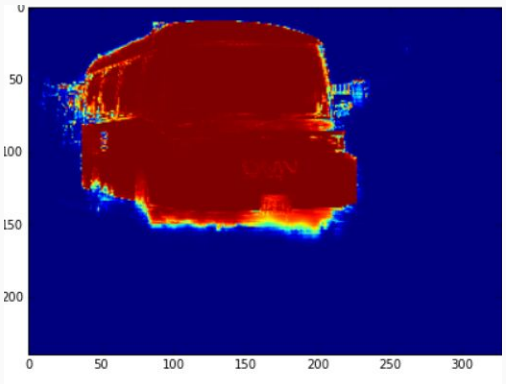
Input image

f

y^ = Af

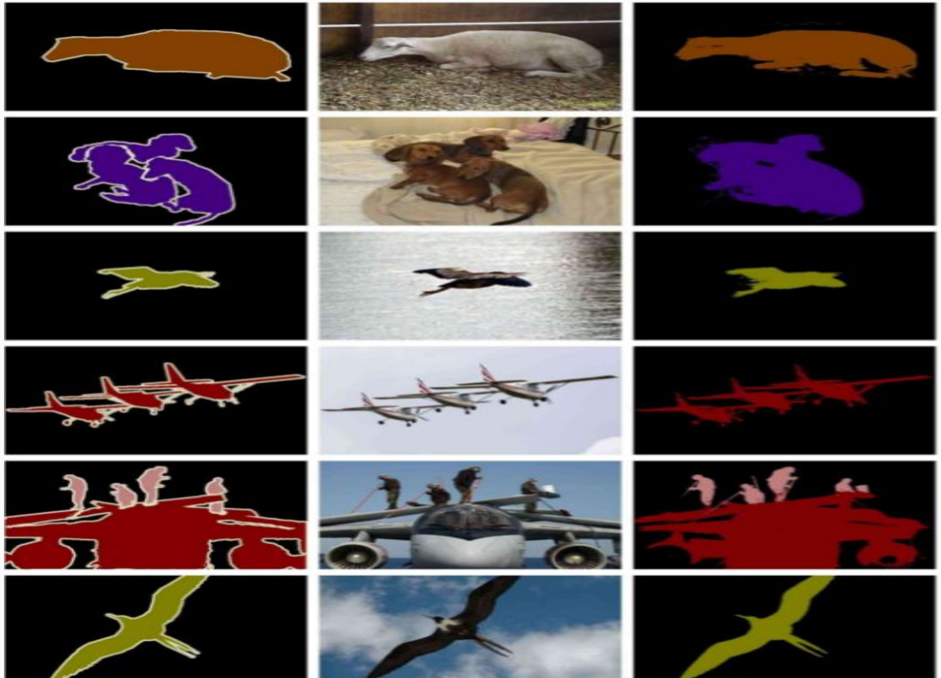


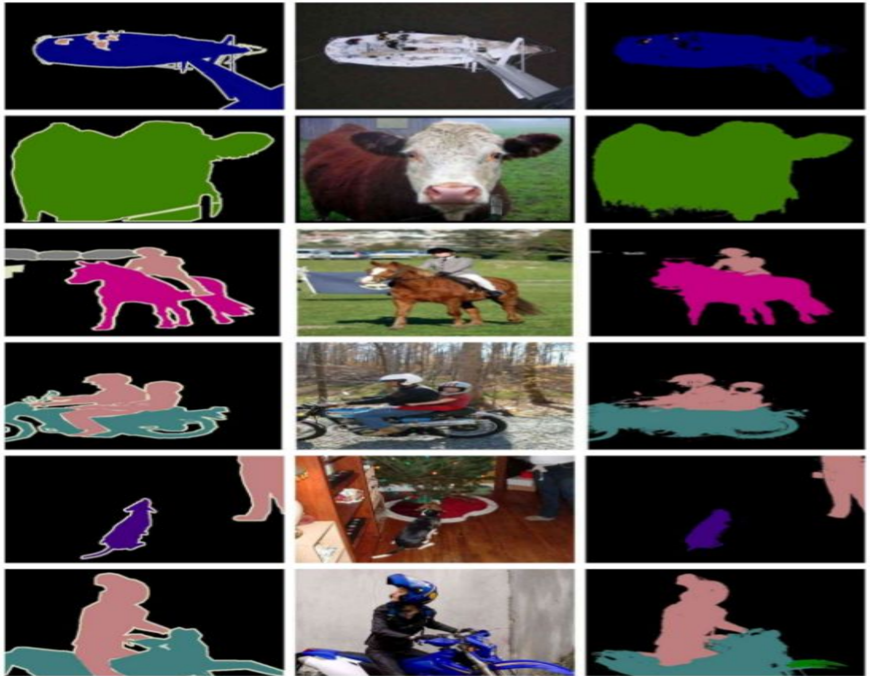




Convolutional Random Walk Networks

- This network is trained on PASCAL-VOC -2012 dataset over 6000 training images and achieved mean IOU of 60.80 , pixel accuracy 74.50.
- These are some of the results:





SEG-NET

- SegNet is composed of a symmetry network: the encoder and the decoder .
- Uses a novel technique to upsample encoder output which involves storing the max-pooling indices used in pooling layer. This gives reasonably good performance and is space efficient
- This network achieved mean IOU of 62.25 and pixel accuracy of 78.81.